



Facultad de Ciencias de la Administración

Ingeniería de Sistemas

DATAWAREHOUSE MUNICIPALIDAD DE COSQUIN.

Trabajo Final de Grado

Autores: Alison, Rodrigo Martín.
Grigioni Docampo, Pablo José.

Tutor: Ing. Boggio, María Alejandra.

Año 2014



ABSTRACT

El presente proyecto es un trabajo final de grado llevado a cabo, para la Municipalidad de Cosquín, por estudiantes de la carrera de Ingeniería en Sistemas dictada en la Facultad de Ciencias de la Administración del Instituto Universitario Aeronáutico. Surge por la necesidad de la Secretaría de Economía y Finanzas Públicas del Municipio de Cosquín de acceder de manera rápida a información confiable relacionada con el área. Ésta información permitirá realizar diferentes análisis desde diversos puntos de vista y determinar las decisiones y acciones convenientes para mejorar su gestión.

Actualmente los dependientes de la secretaría realizan el procedimiento de obtención de reportes e indicadores de forma manual, para los cuales exportan listados diarios, semanales y mensuales desde el sistema municipal (PGM) hacia Excel. Finalmente estos listados son depurados, ordenados, organizados y clasificados de forma manual para poder realizar sus labores.

Al realizar este procedimiento se incurre en varios problemas que serán posteriormente desarrollados, así como se buscará la solución óptima durante el desarrollo del proyecto.

El objetivo del trabajo es realizar el análisis, diseño e implementación de una Solución de Inteligencia de Negocios (Business Intelligence), de modo que ésta herramienta automatice y permita a los usuarios acceder a información de mejor calidad, más confiable, en menor tiempo y que a su vez posibilite el acceso a información histórica; de esta forma no solo se logra resolver los problemas actuales sino que se brindan diferentes beneficios adicionales.

Para lograr el objetivo del proyecto, se comenzará por describir el contexto de la organización para entender el entorno en el que se sitúa, continuando por presentar distintos conceptos claves (Business Intelligence, Data Warehouse, etc.) con sus descripciones y características. Finalmente se realizará y describirá la implementación del proyecto usando diferentes herramientas de software previamente seleccionadas. La elección del software se hará utilizando el Proceso de Decisión presentado por Kepner y Tregoe en su obra “El Nuevo Directivo Racional”, para lo cual se tendrá en cuenta un análisis costo-beneficio en cuanto a calidad, performance y utilidad para el Municipio, buscando disminuir los costos de mantenimiento y licencia, aprovechando la amplia gama de herramientas de software Open Source desarrolladas para el ámbito del Data Warehouse.



Índice de Contenido

ABSTRACT	2
INTRODUCCION	7
GENERALIDADES.....	10
1.1 Antecedentes y Contexto.....	11
1.2 Situación problemática	12
1.3 Problema	13
1.4 Objeto de estudio y campo de acción	14
1.5 Objetivos	14
1.5.1 Objetivo general	14
1.5.2 Objetivos específicos.....	14
1.6 Solución a comprobar	16
1.7 Alcance y limitaciones	16
1.8 Aporte teórico	17
1.9 Aporte práctico.....	17
1.10 Factibilidad	18
1.10.1 Factibilidad Técnica	19
1.10.2 Factibilidad Operativa	19
1.10.3 Factibilidad Económica	19
1.11 Métodos y medios de investigación	19
MARCO CONTEXTUAL	21
2.1 Análisis de los problemas observados – Contexto del problema	22
MARCO TEÓRICO.....	23
3.1 Abstracto del Marco Teórico.....	24
PRIMERA PARTE	25
3.2 Generalidades y conceptos sobre administración pública.....	26
3.2.1 Entidad Municipal	26
3.2.2 Gestión Municipal	26
3.2.3 Finanzas.....	27



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

3.2.4 Ejecución de Gastos, Ingresos y Presupuestaria 27
3.2.5 Indicadores de Gestión Municipal 28

SEGUNDA PARTE..... 31

3.3 Toma de Decisiones y Business Intelligence 32

3.3.1 ¿Qué es la Toma de decisiones? 32
3.3.2 Tipos de decisiones 32
3.3.3 Proceso de la toma de decisiones 33
3.3.4 El Nuevo Directivo Racional – Análisis de Decisiones 34
3.3.5 Necesidad de información 38
3.3.6 ¿Qué es Business Intelligence (BI)? 39
3.3.7 ¿Por qué implementar Business Intelligence? 40
3.3.8 El proceso de BI 41
3.3.9 Beneficios de BI 42

3.4 Data Warehouse y Data Warehousing. 43

3.4.1 Data Warehousing 43
3.4.2 Data Warehouse (DW) 44
3.4.3 Características de un Data Warehouse. 45
 3.4.3.1 Orientación a temas (Orientado al negocio) 45
 3.4.3.2 Integrado..... 46
 3.4.3.3 Variante en el tiempo 48
 3.4.3.4 No volátil 49
3.4.4 Estructura de un Data Warehouse 51
3.4.5 Flujo de datos en un Data Warehouse 54
3.4.6 Ventajas y Desventajas del Data Warehouse..... 54

3.5 Arquitectura del Data Warehouse 56

3.5.1 Sistema Transaccional (OLTP) 57
3.5.2 Load Manager 58
 3.5.2.1 El proceso de Extracción, Transformación y Carga de datos (ETL) 59
 3.5.2.2 Extracción 60
 3.5.2.3 Transformación 60
 3.5.2.4 Carga 61
 3.5.2.5 Procesamiento Paralelo..... 63
3.5.3 Data Warehouse Manager 63
 3.5.3.1 Data Warehouse y su estructura multidimensional 64
 3.5.3.2 Bases de datos multidimensionales 64
 3.5.3.3 Modelado Multidimensional 65
 3.5.3.4 Tipos de modelamiento de un Data Warehouse 71
 3.5.3.5 OLTP vs DW 73
 3.5.3.6 Tipos de implementación de un Data Warehouse. 74
3.5.4 Query Manager 78
3.5.5 Herramientas de Consulta y Análisis 81
 3.5.5.1 Reportes y Consultas 83
 3.5.5.2 OLAP (On Line Analytic Processing). 83
 3.5.5.3 Dashboards 84
 3.5.5.4 Data Mining o Minería de Datos 84
 3.5.5.5 Cuadros de Mando Integrales (CMI) 85
 3.5.5.6 Sistema de Información Ejecutiva (EIS – Executive Information System) 86
 3.5.5.7 Sistemas de Soporte a la Decisión (DSS) 87
3.5.6 Usuarios 87

3.6 Conceptos Complementarios 88

3.6.1 Data Mart 88



3.6.2	Business Models.....	90
3.6.3	Área de Datos	91
3.6.4	Requisitos que debe cumplir un Data Warehouse	93
3.7	Metodología HEFESTO para la construcción de un Data Warehouse.....	95
3.7.1	Paso 1) ANÁLISIS DE REQUERIMIENTOS.....	97
3.7.1.1	a) Identificar preguntas	97
3.7.1.2	b) Identificar indicadores y perspectivas	98
3.7.1.3	c) Modelo Conceptual.....	98
3.7.2	Paso 2) ANÁLISIS DE LOS OLTP	99
3.7.2.1	a) Conformar indicadores.....	99
3.7.2.2	b) Establecer correspondencias.....	99
3.7.2.3	c) Nivel de granularidad.....	100
3.7.2.4	d) Modelo Conceptual ampliado.....	100
3.7.3	Paso 3) MODELO LÓGICO DEL DW.....	100
3.7.3.1	a) Tipo de Modelo Lógico del DW.....	101
3.7.3.2	b) Tablas de dimensiones.....	101
3.7.3.3	c) Tablas de hechos.....	101
3.7.3.4	d) Uniones.....	102
3.7.4	Paso 4) INTEGRACIÓN DE DATOS	102
3.7.4.1	a) Carga Inicial	102
3.7.4.2	b) Actualización	103
	DESARROLLO PRÁCTICO.....	104
4.1	Abstracto del Desarrollo Práctico	105
4.2	Metodología Propuesta	105
4.3	Análisis general de la organización	107
4.3.1	Descripción de la organización	107
4.3.2	Objetivos	108
4.3.3	Funciones.....	108
4.3.4	Organigrama.....	109
4.3.5	Situación informática actual.....	111
4.3.6	Relación de las metas de la Secretaría con las del DW.....	112
4.3.7	Procesos	112
4.4	Modelado del Data Warehouse	112
4.4.1	Análisis de Requerimientos	113
4.4.1.1	Identificar preguntas.....	113
4.4.1.2	Identificar indicadores y perspectivas	115
4.4.1.3	Modelo Conceptual.....	116
4.4.2	Análisis de los OLTP	117
4.4.2.1	Conformar indicadores	117
4.4.2.2	Establecer correspondencias	118
4.4.2.3	Modelo conceptual ampliado.....	121
4.4.3	Modelo Lógico del DW.....	122
4.4.3.1	Tipo de Modelo Lógico del DW	122
4.4.3.2	Tablas de dimensiones	122
4.4.3.3	Tablas de hechos.....	124
4.4.3.4	Uniones	125
4.5	Selección del conjunto de herramientas	126
4.5.1	Análisis de Decisiones – Elección de la tecnología a utilizar	127
4.5.1.1	Enunciado de la decisión o del propósito	127



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

4.5.1.2	<i>Los objetivos para la decisión</i>	127
4.5.1.3	<i>Generación y evaluación de las alternativas</i>	129
4.5.1.4	<i>Consecuencias de la alternativa</i>	149
4.6	Implementación y Puesta en Marcha	149
4.6.1	Descripción de la Tecnología	150
4.6.1.1	<i>Descripción de las herramientas y elementos de software a utilizar</i>	150
4.6.1.2	<i>Comparación entre tecnología existente y necesaria para el proyecto</i>	152
4.6.2	Definición de la arquitectura del Data Warehouse	154
4.6.3	Integración de Datos	156
4.6.3.1	<i>Carga Inicial</i>	157
4.6.3.2	<i>Actualización</i>	167
4.6.4	Creación de Cubos Multidimensionales	169
4.6.5	Herramientas de Reportes y Consulta	172
4.6.6	Instalación y Puesta en Marcha	183
CONCLUSIONES		190
BIBLIOGRAFÍA		194
ANEXOS		196



INTRODUCCION



Como se indicó anteriormente el presente proyecto tiene por finalidad solucionar algunos de los problemas que ocurren en el área de Economía y Finanzas Públicas de la Municipalidad de Cosquín. El objetivo del trabajo se resume en la posibilidad de acceder a nuevos informes, lo que permitirá contar con información de manera más rápida, confiable y segura. Actualmente en el Área de Economía y Finanzas Públicas de la Municipalidad de Cosquín, los usuarios realizan el procedimiento de obtención de reportes e indicadores de forma manual. Los mismos exportan a Excel listados diarios, semanales y mensuales del sistema municipal (PGM), los que a su vez, son manualmente depurados, ordenados, organizados y clasificados para poder realizar sus labores.

Al realizar este procedimiento incurre en varios problemas. A continuación se exponen algunos de ellos:

- ✓ Cada usuario aplica su criterio para realizar el procedimiento de depurar, ordenar, organizar y clasificar los datos que obtienen, el mismo no está estandarizado a nivel de toda la organización, por lo que la información obtenida no es necesariamente confiable.
- ✓ Al realizar este procedimiento de forma manual se invierte mucho tiempo y los resultados que obtienen no siempre son exactos, debido a que al manipular manualmente una gran cantidad de datos el usuario está propenso a cometer errores.
- ✓ Como resultado del procedimiento se obtienen reportes e indicadores en hojas de cálculo, de modo que es dificultoso y prácticamente imposible contar con información histórica de la Gestión Municipal.
- ✓ Los reportes e indicadores son almacenados en PCs pertenecientes a puestos de trabajo sin respaldo.
- ✓ Se necesita de personal con conocimiento del sistema y del área para poder obtener los informes, generando una dependencia en relación a los mismos.
- ✓ Debido a que los resultados generan archivos Excel no es posible implementar una política de seguridad y confidencialidad de los mismos.
- ✓ No es posible acceder a los resultados de forma rápida, dependiendo de los tiempos y prioridades del área.
- ✓ Existen distintos tipos de reportes que son deseados, pero que no se realizan debido a la complejidad que poseen lo cual hace que se requiera de tiempo excesivo en su construcción.

En la actualidad, la información es considerada uno de los activos más importantes de cualquier organización. La información es una fuente indispensable para realizar el análisis de cualquier ámbito, generando el conocimiento que permite tomar alguna decisión. Pero, como por lo general, las organizaciones manejan grandes cantidades de datos, y posiblemente de diversas fuentes, resulta muy complicado contar con dichos datos adecuadamente organizados para ser usados como una herramienta para la toma de decisiones.

Debido a esto, en el rubro de las tecnologías de información se desarrollan sistemas que son capaces de administrar todo tipo de información y presentarla de forma especialmente estructurada que permita su fácil análisis y explotación.



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

Es por esto, que se ve la necesidad de realizar un proyecto que permita brindar una herramienta que automatice el procedimiento explicado anteriormente solucionando los problemas mencionados.

Por lo tanto, el objetivo principal del presente proyecto es realizar el análisis, diseño e implementación de una Solución de Inteligencia de Negocios para la Secretaría de Economía y Finanzas Públicas, de manera que los usuarios cuenten con información sólida y confiable de alta disponibilidad, que cumpla con las necesidades de información de todos los usuarios, permitiendo resolver las dificultades citadas y obtener beneficios adicionales.



GENERALIDADES



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

El presente capítulo busca presentar y describir el problema con el que cuenta en la actualidad la Secretaría de Economía y Finanzas Públicas de la Municipalidad de Cosquín en cuanto al manejo de la información de su gestión, cuya solución es el objetivo principal del desarrollo de este proyecto de fin de carrera. Asimismo se describe el entorno en el que se presenta dicho inconveniente y las diversas alternativas que existen para dar solución a todo, o parte del problema.

1.1 Antecedentes y Contexto

Para entender el proyecto, es menester comenzar por situarse geográficamente en donde ocurre el mismo. Por lo cual el trabajo comenzará presentando a “Cosquín”, ciudad donde se sitúa el objeto de estudio; esta es una localidad del oeste de la provincia de Córdoba, Argentina, en pleno Valle de Punilla, a 58 km de la capital provincial. La misma es cabecera del departamento de Punilla, al pie de las Sierras Chicas. Se presenta a continuación un mapa de la ciudad.





Cosquín, como toda ciudad del territorio Argentino posee un Municipio, el cual se compone por un número importante de Secretarías como son: Secretaría de Gobierno, Secretaría de Economía y Finanzas Públicas, Secretaría de Planeamiento y Desarrollo Urbano, Secretaría de Salud Ambiente y Desarrollo Sustentable, Secretaría de Desarrollo Social, Secretaría de Educación y Cultura, Secretaría de Turismo y Deportes, Secretaría de Servicios Públicos y Mantenimiento y la Secretaría de Asesoría Letrada.

El presente proyecto se centrará en la Secretaría de Economía y Finanzas Públicas, siendo necesario recolectar información de Tesorería, Contaduría y de la Dirección de Recursos Tributarios para lograr los objetivos planteados.

Le corresponde a dicha Secretaría asistir al Departamento Ejecutivo en todo lo inherente a la elaboración y control de la ejecución del presupuesto municipal, como así también en los niveles del gasto y de los ingresos, conforme a las pautas que fije el Departamento Ejecutivo. A su vez, efectúa la elaboración del proyecto de presupuesto municipal, fijación de la política tributaria, percepción de los recursos municipales, régimen de pagos y el análisis, evaluación y control de la ejecución presupuestaria. Tiene a su cargo la conducción de la Tesorería y el régimen de pagos y la verificación, centralización y conducción de la información sobre el endeudamiento público de la Municipalidad, y las gestiones necesarias para obtener financiamiento y crédito ante instituciones financieras oficiales o privadas del ámbito provincial, nacional o internacional, así como convenir planes de amortización, intereses y demás condiciones relacionadas con la obtención de la financiación respectiva.

El volumen de información que maneja es muy elevado, por lo tanto, para apoyar la toma de decisiones es necesario implementar una solución de Inteligencia de Negocios (B.I. - Business Intelligence). El objetivo es lograr la transformación de los datos que se generan como producto secundario de las transacciones diarias que se realizan y se almacenan en bases de datos relacionales, en información estratégica que apoye la toma de decisiones y contribuya de manera efectiva a la creación de valor, es decir, mediante alguna herramienta de B.I. apoyar a los tomadores de decisiones con la información correcta, en el momento y lugar adecuado, incrementando la efectividad del área, permitiendo tomar mejores decisiones de negocios.

1.2 Situación problemática

Al realizar un análisis de la situación actual es posible darse cuenta que el proyecto tiene actualidad y pertinencia debido a que dentro de las funciones y responsabilidades de la Secretaría de Economía y Finanzas Públicas se encuentran:

- ✓ Formulación periódica de reportes financieros relacionados con la ejecución de los ingresos,
- ✓ Formulación periódica de reportes financieros relacionados con la ejecución de los gastos
- ✓ Elaboración de informes del movimiento rendimiento y control de los valores de la entidad y de los indicadores que permitan llevar un adecuado control de la gestión municipal.

Para llevar a cabo estas tareas, se insume una gran cantidad de horas de personal, exportando reportes del sistema y confeccionando mensualmente una serie de archivos Excel. Éste procedimiento no permite disponer de información online, compartida y



dinámica, además que el mismo es factible de errores y que la información se acumula en diferentes computadoras personales lo cual genera problemas relacionados con la redundancia de información.

1.3 Problema

Como se ha ido mencionando, actualmente los dependientes del área de Economía y Finanzas Públicas realizan el procedimiento de obtención de reportes e indicadores de forma manual. Los mismos exportan a Excel listados diarios, semanales y mensuales del sistema municipal (PGM), los que son depurados, ordenados, organizados y clasificados para poder realizar sus labores.

El realizar este procedimiento se incurre en varios problemas. A continuación se recuerdan algunos de ellos:

- ✓ Cada usuario aplica su criterio para realizar el procedimiento de depurar, ordenar, organizar y clasificar los datos que obtienen, el mismo no está estandarizado a nivel de toda la organización, por lo que la información obtenida no es necesariamente confiable.
- ✓ Al realizar este procedimiento de forma manual invierten mucho tiempo y los resultados que obtienen no siempre son exactos. El hecho de manipular de forma manual una gran cantidad de datos hace que el usuario este propenso a cometer errores.
- ✓ Como resultado del procedimiento se obtienen reportes e indicadores en hojas de cálculo, de modo que es dificultoso y en la práctica imposible contar con información histórica de la Gestión Municipal.
- ✓ Los reportes e indicadores son almacenados en PCs pertenecientes a puestos de trabajo sin respaldo.
- ✓ Se necesita de personal con conocimiento del sistema y del área para poder obtener los informes, generando una dependencia en relación a los mismos.
- ✓ Debido a que los resultados generan archivos Excel no es posible implementar una política de seguridad y confidencialidad de los mismos.
- ✓ No es posible acceder a los resultados de forma rápida, dependiendo de los tiempos y prioridades del área.
- ✓ Existen distintos tipos de reportes que son deseados, pero que no se realizan debido a la complejidad que poseen lo cual hace que se requiera de tiempo excesivo en su construcción.

Analizando los distintos problemas, es posible resumir su solución en la capacidad de contar con una herramienta para la toma de decisiones capaz de extraer, limpiar, filtrar, organizar y presentar de manera sencilla los datos que se encuentran en distintas fuentes de información.

En la actualidad se desarrollan sistemas que son capaces de administrar todo tipo de información y presentarla de forma especialmente estructurada, que permita su fácil análisis y explotación. Por ello uno de los objetivos del proyecto, es el estudio de dichos sistemas, buscando la mejor solución que se adapte al problema planteado.

Es necesario comprender que para poder utilizar alguno de los software que permiten explotar la información y verla de un modo útil para la toma de decisiones analizando distintas variables desde distintas perspectivas según el usuario que lo necesite, se



requiere del uso de una metodología que permita crear la estructura necesaria para transformar los datos operacionales en información para la gerencia.

Éste será el gran objetivo del proyecto, crear una estructura que tomando los datos que se generan diariamente en los sistemas transaccionales de la Secretaría de Economía y Finanzas Públicas de la Municipalidad de Cosquín, transforme los mismos y se le presenten a la gerencia en información directa para la toma de decisiones, solucionando todos los problemas anteriormente mencionados y brindando beneficios adicionales.

1.4 Objeto de estudio y campo de acción

Para poder solucionar el problema anteriormente descripto, se tendrá como objeto principal de estudio los sistemas Data Warehouse, de los cuales se analizará su posible implementación con alguna tecnología Open Source, debido a que existen en el mercado distintos softwares que no requieren el costo de una licencia, y brindan un alto rendimiento y apoyo para la toma de decisiones.

El objeto de estudio se centra en la Ingeniería de Sistemas, dentro de la cual se encuentra lo concerniente al análisis, desarrollo e implementación de una Solución de Inteligencia de Negocios para la Secretaría de Economía y Finanzas Públicas de la municipalidad de Cosquín. Más específicamente el proyecto se enmarca en la Ingeniería de Software.

Para lograr el objetivo del trabajo, se comenzará realizando un relevamiento sobre el estado actual del área de Informática, para poder conocer cuáles son los recursos disponibles. A su vez se relevará el área de Economía y Finanzas Públicas, realizando un acercamiento a los usuarios, para conocer mediante ellos las herramientas que utilizan, cómo se utilizan, para qué se utilizan, el tiempo que les insume realizar las tareas, y como logran los indicadores esperados para la toma de decisiones y con qué margen de error.

El desarrollo de todos los aspectos teóricos que sean necesarios cubrir en el proyecto para trabajar sobre el campo de acción, se hallarán referenciados en la sección “Marco Teórico” del presente documento.

1.5 Objetivos

1.5.1 Objetivo general

El objetivo principal del presente proyecto es realizar el análisis, diseño e implementación de una Solución de Inteligencia de Negocios para la Secretaría de Economía y Finanzas Públicas de la Municipalidad de Cosquín, de manera de poder hacerle contar con información sólida, oportuna, confiable, de alta disponibilidad, y que cumpla con las necesidades del área permitiéndole mejorar su gestión y toma de decisiones.

1.5.2 Objetivos específicos

- ✓ Aplicar un proceso racional de toma de decisiones, para la evaluación, comparación y selección de las herramientas óptimas a utilizar.



- ✓ Implementar un conjunto de herramientas para consultar, analizar y presentar la información. Las herramientas de acceso deben ser simples y fáciles de usar.
- ✓ Obtener información consistente, de buena calidad y útil.
- ✓ Contar con información histórica.
- ✓ Respalidar los informes y eliminar la redundancia.
- ✓ Lograr la implementación de la solución, con la mayor eficiencia posible en la relación costo-calidad en el resultado obtenido.
- ✓ Generar y transmitir el “Know How” necesario a los usuarios del sistema Data Warehouse para que sea utilizado de la mejor manera posible.
- ✓ Permitir que el Municipio de Cosquín pueda reducir los costos en términos de pérdida de tiempo y errores que se producen en la actualidad a partir de la implementación del Data Warehouse.
- ✓ Desarrollar un sistema que aplique un procedimiento estándar e independiente del personal humano, el cual sea capaz de extraer, depurar, ordenar, organizar, clasificar y presentar los datos en forma de informes claros y útiles para la toma de decisiones.
- ✓ Permitir el acceso a la información de manera rápida y oportuna.
- ✓ Crear reportes que en la actualidad son deseados, pero que no se están realizando debido a la complejidad que poseen lo cual hace que se requiera de tiempo excesivo en su construcción.

Con el logro de los objetivos, se pretende ofrecer un conjunto de reportes e indicadores que cumplan con las necesidades de información de la secretaría.

Para verificar el cumplimiento de los objetivos específicos se definieron una serie de indicadores que serán controlados al finalizar el trabajo para comprobar el resultado del proyecto.

- ***Amigabilidad de la herramienta:*** Es medido por el porcentaje del personal que está satisfecho con la interfaz del software implementado.
Calculo: $(\text{nro. de personal satisfecho} / \text{nro. total de personas que usan el sistema}) * 100$.
Total de las personas que usan el sistema: 4 personas.
Cantidad de personas satisfechas: 4 personas.
- ***Verificación de la utilización de un método desarrollado para la elección de las herramientas de software a implementar:*** Se comprueba mediante el uso de algún proceso racional de toma de decisiones.
- ***Calidad de la información brindada por la solución:*** Se mide a través del porcentaje del personal que está satisfecho con la calidad de la información que se obtiene de la solución BI desarrollada. Por calidad de información se entiende que la misma debe ser precisa, oportuna y útil.
Calculo: $(\text{nro. de personal satisfecho} / \text{nro. total de personas que usan el sistema}) * 100$.
Total de las personas que usan el sistema: 4 personas.
Cantidad de personas satisfechas: 4 personas.



- **Capacidad de contar con información histórica:** Se verifica mediante el análisis de datos históricos a través de la solución.
- **Capacidad de resguardo y respaldo de la información:** Se verifica mediante el uso de alguna técnica y medidas de seguridad que permitan asegurar el resguardo y respaldo de la información.
- **Capacidad de uso del sistema por parte de los usuarios:** Se mide mediante el porcentaje del personal que está capacitado para el uso del software implementado.
Total de las personas que usan el sistema: 4 personas.
Cantidad de personas satisfechas: 4 personas.
- **Disminución de los tiempos y errores en la generación de informes:** Se mide comparando los tiempos y errores que se tienen actualmente contra los obtenidos a partir de la implementación del DataWarehouse.

1.6 Solución a comprobar

Con el avance de la informática, la gran mayoría de las organizaciones han ido sistematizando sus procesos, eso significa que se han desarrollado sistemas especialmente orientados a registrar todas las transacciones que se llevan a cabo en la gestión general. Como resultado de esto, se han generado sistemas transaccionales (OLTP) que han almacenado la información en diversos tipos de bases de datos.

Por otro lado la toma de decisiones generalmente es realizada en base a la experiencia o mejor criterio del decidor, pero hoy en día estamos en un mundo competitivo que no se puede dar el lujo de equivocarse en una decisión, por más pequeña que ésta sea. Es por ello que surgen los repositorios de datos (Data Warehouse), donde se almacena toda la información transaccional que generalmente no se aprovecha.

Una Solución de Inteligencia de Negocios permite extraer información que esté disponible en las bases de datos de los sistemas transaccionales que son de vital importancia como apoyo a la toma de decisiones. Si se dispone de un software de Inteligencia de Negocio, las decisiones gerenciales de la organización van a estar mejor fundamentadas y habrá menos posibilidades de equivocación, lo cual hace al proceso más eficaz y eficiente, reduciendo costos y permitiendo una mejor planificación.

Con el proyecto a desarrollar se pretende demostrar que con la implementación del Data Warehouse en el municipio de Cosquín, específicamente del Data Mart en la Secretaría de Economía y Finanzas Públicas se podrán solucionar los errores y problemas que se tienen en la actualidad, así como brindar una solución de alta performance y gran rendimiento para el soporte a la toma de decisiones, permitiendo disponer de la información sé que se necesita de manera rápida, efectiva y fácil de comprender.

1.7 Alcance y limitaciones

Si bien el proyecto se desarrolla dentro del ámbito del Municipio de Cosquín, abarcar todas las secretarías en un mismo proyecto lo haría demasiado amplio. Por lo tanto se ha



limitado el alcance del proyecto a la Secretaria de Economía y Finanzas Públicas, buscando solucionar los problemas que allí se presentan, identificando y seleccionando los indicadores necesarios para una mejor interpretación de los datos y posterior toma de decisiones.

1.8 Aporte teórico

El proyecto brindará distintos aportes teóricos que se pueden resumir en:

- ✓ Presentación de términos propios del negocio.
- ✓ Descripción del proceso de toma de decisiones expuesto por Kepner y Tregoe en “El Nuevo Directivo Racional”.
- ✓ Descripción de términos propios de Business Intelligence y ámbito de los sistemas de Data Warehouse.
- ✓ Exposición de la metodología Hefesto, desarrollada por el Ing. Bernabeu.

Se comenzará con el estudio, reconocimiento, análisis y definición de distintos conceptos propios del negocio Municipal, así como también el estudio de los modelos, metodologías y términos propios de la Inteligencia de Negocios los cuáles son necesarios para comprender el proyecto.

Se realizará una investigación y exposición de las distintas tecnologías Open Source de Business Intelligence disponibles en el mercado, a fin de obtener un conciso conocimiento respecto de las mismas y así poder seleccionar aquella que resulte más calificada. Al finalizar el proyecto se logrará como resultado final una comparación de performance y rendimiento de las distintas tecnologías Open Source disponibles, marcando ventajas y desventajas ofrecidas por cada una.

Para la realización de las comparaciones mencionadas y posterior selección de las herramientas óptimas a implementar, se utilizará el proceso racional de toma de decisiones descrito por Kepner y Tregoe en su obra “El Nuevo Directorio Racional”, así como para la implementación del proyecto se utilizará como base, la reconocida metodología para construcción de un Data Warehouse, “Hefesto”, desarrollada por el Ing. Darío Benabeu.

Por lo tanto los resultados del proyecto podrán ser generalizados y los lectores de esta tesis podrán analizar el contenido de la misma pudiendo aplicar a su negocio tanto el proceso de decisiones, como la metodología mencionada, e inclusive podrá ser utilizada como base y continuar con el estudio y análisis de agregados y otras técnicas para mejorar el desempeño de las tecnologías aquí expuestas.

1.9 Aporte práctico

El tema investigado resulta relevante e interesante, teniendo en cuenta que logrará dos resultados significativos: por una parte, la investigación y conocimiento de las diferentes tecnologías Open Source disponibles para el diseño, desarrollo e implementación de un sistema de Data Warehouse en una organización, y por otra la solución de los problemas mencionados, presentando a los usuarios los informes necesarios de manera útil y oportuna.

De esta forma, el proyecto ayudará a resolver el problema real planteado, presentando una nueva manera de realizar procesos que anteriormente se realizaban de manera



manual, y los cuales no solo se encuentran expuestos a los posibles errores humanos, sino que insumen tiempos y costos, que al ser automatizados, disminuirán notablemente.

A partir de realizar una investigación, análisis, y determinación de la herramienta más conveniente para la implementación del proyecto, éste no solo permitirá solucionar los problemas que hoy se encuentran dentro de la Secretaría, sino que podrá ser aplicado a otras áreas y otros municipios, abriéndose caminos dentro del mercado.

Es importante destacar, que el trabajo no solo se limita a la exploración e implementación de herramientas de software para Business Intelligence, sino que aplica temas teóricos tratados en asignaturas de la carrera de Ingeniería en Sistemas, específicamente “Toma de Decisiones”, aplicando el proceso aprendido en dicha materia para la toma de decisiones, en este caso para el análisis, comparación y selección de la suite de herramientas para el desarrollo e implementación del DataWarehouse.

El aporte práctico se resume en:

- ✓ Aplicación del proceso de toma de decisiones expuesto por Kepner y Tregoe en “El Nuevo Directivo Racional”, para el análisis, comparación y selección de la mejor alternativa de las herramientas Open Source disponibles.
- ✓ Aplicación de la metodología Hefesto, para el diseño y construcción del Data Warehouse de la Secretaría de Economía y Finanzas Públicas del Municipio de Cosquín.

1.10 Factibilidad

Para determinar la factibilidad del proyecto es necesario investigar, analizar y estudiar las diferentes opciones:



- 1.10.1 Factibilidad Técnica:** Se cuenta con la tecnología necesaria para desarrollar el proyecto, tanto en hardware como software. Se deberá realizar una investigación sobre los temas y áreas mencionadas para poder ejecutar la decisión más óptima, examinando y considerando los diferentes aspectos involucrados.
- 1.10.2 Factibilidad Operativa:** El proyecto es factible desde un punto de vista operativo, debido que, los realizadores del trabajo cuentan con total contacto y relación con las personas dentro del municipio que proveerán la información y recursos materiales necesarios para la implementación. Dicha implementación se llevará a cabo utilizando uno de los servidores que dispone el municipio.
- 1.10.3 Factibilidad Económica:** El proyecto es factible desde el punto de vista económico, dado que no genera gastos extras en los tesistas. El mismo no requiere de una inversión para su desarrollo, la información es recolectada directamente desde el municipio, y se cuenta con la tecnología necesaria para el desarrollo del mismo. Se analizarán las opciones de tecnologías y software libre o con licencias educativas que permiten el desarrollo del mismo sin generar erogaciones de importancia.

Es posible concluir que las condiciones técnicas, operativas y económicas actuales del Municipio hacen factible el proyecto.

1.11 Métodos y medios de investigación

Los tipos de investigación a utilizar son:

- **Bibliográfica:** Se realiza en base a libros, revistas e internet. Permite fundamentar científicamente las variables de la presente tesis y se utiliza para elaborar el marco teórico.
- **De Campo:** Con este tipo de investigación se puede valorar la problemática relacionada con la toma de decisiones y ratificarla.

Además serán utilizados métodos de investigación empíricos y lógicos para el desarrollo del presente trabajo.

Por métodos empíricos se refiere a modelos de investigación basados en la experiencia y pruebas sobre el objeto de estudio, permitiendo reconocer características y relaciones esenciales de los elementos considerados en el mismo, lo que dará las bases para estudios descriptivos.

Se aplicarán métodos empíricos para:

- Estudiar el entorno del problema durante la etapa de marco contextual, donde se describe el contexto del proyecto basándose en información obtenida mediante observación directa, reuniones con el personal, búsquedas en Internet, entre otros.
- Efectuar el desarrollo del marco teórico y su correspondiente diagnóstico, en base a la observación directa, la información obtenida mediante libros, blogs,



descripciones documentadas y opiniones de referentes en temáticas relacionadas, así como también el apoyo en la información recopilada del personal afectado.

- Determinar los requerimientos y el desarrollo del modelo teórico, en base al análisis de soluciones implementadas en otros ámbitos y la documentación referente a las tecnologías seleccionadas.
- Ejecutar la concreción del modelo, observando los resultados obtenidos y evaluando si se alcanzaron los objetivos planteados.

Los métodos lógicos son aquellos que utilizan el pensamiento en sus funciones de deducción, análisis y síntesis.

Los métodos lógicos serán utilizados para:

- Analizar la situación problemática y establecer claramente el problema que se pretende solucionar.
- Relacionar adecuadamente el marco teórico a desarrollar con el objeto de estudio y el campo de acción.
- Analizar y deducir el conjunto de fenómenos y hechos que dan surgimiento al problema.
- Realizar la implementación de la solución, considerando lo generado en el modelo teórico y utilizando las tecnologías más adecuadas.



MARCO CONTEXTUAL



2.1 Análisis de los problemas observados – Contexto del problema

Como se ha mencionado previamente, actualmente los dependientes de la Secretaría de Economía y Finanzas Públicas del municipio de Cosquín realizan el procedimiento de obtención de reportes e indicadores de forma manual sin un proceso automatizado que los guíe y asegure el éxito del resultado de la extracción así como de la información obtenida. El hecho de realizar este procedimiento manualmente genera algunos problemas, ya que además de generar información con ciertas falencias para la toma de decisiones, el tiempo y las tareas que demanda generan un costo elevado para la organización.

Es por ello que es considerado necesario desarrollar un sistema que permita de modo rápido y eficiente la recolección, organización, análisis, y proceso de los datos recolectados desde los sistemas operacionales (OLPT), para presentar de modo útil y fácil la información que la alta gerencia requiere para la toma de decisiones.

Es importante destacar que existen en la actualidad herramientas de softwares que permiten desarrollar el sistema mencionado. El conjunto de estas soluciones se enmarcan dentro de lo que se conoce como herramientas de Inteligencia de Negocios y constituyen los denominados sistemas Data Warehouse. Será objetivo de este trabajo comenzar por el estudio de este tipo de sistemas, conocer y comprender su utilidad para finalmente realizar el modelado óptimo para implementar la solución que permita responder a los requerimientos del área de Economía y Finanzas Públicas.

Una vez finalizado el modelado del sistema a implementar se deberán analizar las diferentes alternativas de software disponible en el mercado, buscando el mayor beneficio para la organización, para lo cual se aplicará el proceso racional de toma de decisiones presentado en el libro “El Nuevo Directivo Racional” utilizado en la asignatura de Toma de Decisiones.

Otra realidad que se debe tener en cuenta al momento de afrontar la implementación de un sistema nuevo y desconocido para los trabajadores de la Municipalidad de Cosquín, es la necesidad de programar una capacitación en la cual se le brinde el Know How necesario para que los futuros usuarios puedan utilizar la herramienta de la mejor manera posible sin problema alguno.

Como consecuencia de la situación planteada, durante el desarrollo del marco teórico, en primer lugar se presentarán distintos términos propios del negocio y necesarios para comprender determinados puntos del proyecto, seguido por la descripción del proceso racional de toma de decisiones presentado en el libro “El Nuevo Directivo Racional” de Kepner y Tregoe. Posteriormente el trabajo presentará todos los conceptos y temas relevantes a lo que se conoce en actualidad como Business Intelligence y DataWarehouse, así como la metodología que se utilizará como base para el diseño y construcción de la solución, “Hefesto” desarrollada por el Ingeniero Darío Bernabeu. Una vez comprendidos todos estos puntos, finalmente se exhibirá una investigación de las distintas tecnologías Open Source de Business Intelligence.



MARCO TEÓRICO



3.1 Abstracto del Marco Teórico

El presente es el marco teórico del trabajo final de grado (TFG) “Implementación de un Data Warehouse para la Municipalidad de Cosquín”. En el mismo se describen distintos conceptos teóricos y tecnologías que son necesarios estudiar, analizar y dominar para plantear una solución al problema que ha sido presentado en este proyecto.

Se exponen algunos conceptos que ayudan a dar un entendimiento general de la situación actual, así como de algunas nociones involucradas en la solución que se plantea.

El marco teórico que se expone a continuación está compuesto por dos partes.

En la primer parte se desarrollan los términos relacionados al rubro Municipal fundamentales para conocer el contexto del negocio en el cual se ubica el proyecto.

En la segunda parte son desarrollados todos aquellos conceptos que se deben conocer para tener un sólido dominio del campo de acción, tratando temas como toma de decisiones, Business Intelligence, Data Warehouse, Data Warehousing y arquitectura Data Warehousing. En esta etapa se expone el proceso racional de toma de decisiones, el cual es extraído del libro “El Nuevo Directivo Racional” de Kepner y Tregoe y se presenta el desarrollo teórico de la Metodología Hefesto la cual es utilizada para el modelado e implementación del sistema.

Esta agrupación en partes es realizada para mejorar la organización y permitir un mejor entendimiento de los términos y el trabajo en su conjunto.



PRIMERA PARTE



3.2 Generalidades y conceptos sobre administración pública

Las siguientes definiciones son propias del rubro del negocio de la Municipalidad:

3.2.1 Entidad Municipal

Una municipalidad es la organización que se encarga de la administración local en un pueblo o ciudad. En varios países de América, la municipalidad es el organismo que administra una comuna, cantón o distrito.

Las municipalidades son órganos del gobierno cuya labor es ejercer en las circunscripciones provinciales y distritales de cada una de las regiones del país, con las atribuciones, competencias y funciones que les asigna la Constitución Política.

Está encabezada por un intendente y un concejo, todos elegidos por votación popular. Es también responsable de la administración de la educación y salud pública en su respectivo ámbito territorial.

Para la constitución Argentina, el régimen municipal es: “una persona de derecho público, constituida por una comunidad humana, asentada en un territorio determinado, que administra sus propios intereses y que depende siempre, en mayor o menor grado, de una entidad pública superior, el estado nacional o provincial”.

La constitución tiene escasas referencias a la organización de las municipalidades:

- a. El Art. 5, dispone que cada provincia dictara para si una constitución, bajo ciertas condiciones, una de ellas es que asegure su régimen municipal;
- b. El Art. 123, referido igualmente a las constituciones provinciales, indica que tendrán que sancionarse “asegurando la autonomía municipal y reglando su alcance y contenido en el orden institucional, político, administrativo, económico y financiero”;
- c. El Art. 75 Inc. 30, admite poderes impositivos municipales sobre los establecimientos de utilidad nacional.

3.2.2 Gestión Municipal

La gestión municipal se refiere a la administración de los recursos materiales, económicos y humanos de una determinada municipalidad. Esta labor es ejercida por los funcionarios municipales y su principal objetivo es hacer uso de los recursos disponibles de la mejor forma para lograr el bienestar de la población que conforma dicha municipalidad.

La Gestión Municipal, implica el desarrollo de un conjunto de actividades, estrategias y técnicas idóneas que permitan desarrollarse internamente y proyectarse a la comunidad. Los Principales Instrumentos de Gestión utilizados por una Municipalidad son:

- Reglamento Interno del Consejo Municipal.
- El Organigrama.
- Reglamento de Organización y Funciones (ROF).
- El Manual de Organización y Funciones (MOF).
- El Cuadro de Asignación de Personal (CAP).



- El Presupuesto Analítico de Personal (PAP).
- El Presupuesto Nominativo de Personal (PNP).
- El Texto Único de Procedimientos Administrativos (TUPA).
- El Plan Operativo Institucional (POI).
- El Plan Estratégico Institucional (PEI).
- EL Reglamento Interno de Control y Permanencia de Personal.

3.2.3 Finanzas

Las finanzas son una rama de la economía que se encarga de estudiar la obtención y uso eficaz del dinero a través de un período de tiempo por parte de un individuo, organización o del estado.

3.2.4 Ejecución de Gastos, Ingresos y Presupuestaria

Ejecución de Gastos, se refiere a una fase del proceso presupuestario que se inicia una vez aprobado el presupuesto. La misma consiste en la aplicación de los recursos humanos, materiales y financieros, conjuntamente con la utilización de una serie de técnicas y procedimientos administrativos, contables, de productividad, de control y de manejo financiero para la obtención de las metas y objetivos determinados para el sector público en los planes y programas de mediano y corto plazo.

La Ejecución del Gasto comprende las etapas del Compromiso, Devengado y Pago. El procedimiento de la Ejecución de Gasto es el siguiente:

- a) Calendarización: Previsión de niveles máximos de compromisos presupuestales para un período dado.
- b) Realización de Compromisos: Emisión de documentos sustentatorios de compromisos (Orden de Compra, Orden de Servicio, etc.).
- c) Determinación de Devengados Seguimiento de la obligación de gastar por parte del Pliego.
- d) Autorización de Giro: Nivel máximo de Giro que el Tesoro Público proporciona a cada pliego presupuestal. Difiere de la anterior en la medida que el pliego acumule saldos en su subcuenta.
- e) Autorización de Pago: Tope que el Tesoro Público autoriza al Banco de la Nación a abonar a la subcuenta de cada Pliego Presupuestal.

Ejecución de Ingresos, se produce cuando se recaudan, captan u obtienen los recursos que utiliza un determinado Pliego durante el Año Fiscal. La Ejecución Financiera de los Egresos comprende el registro del devengado, girado y pago de los compromisos realizados.

Ejecución Presupuestaria, es el registro de la información de los recursos captados, recaudados u obtenidos por parte de los Pliegos Presupuestarios. La Ejecución Presupuestaria de los Egresos consiste en el registro de los compromisos durante el Año fiscal. La Ejecución Presupuestaria de las Metas Presupuestarias se verifica cuando se registra el cumplimiento total o parcial de las mismas.

Ejecución Presupuestaria Directa, se produce cuando el Pliego con su personal e infraestructura es el ejecutor presupuestal y financiero de las Actividades y Proyectos así como de sus respectivos Componentes.



Ejecución Presupuestaria Indirecta, se produce cuando la ejecución física y/o financiera de las Actividades y Proyectos así como de sus respectivos Componentes, es realizada por una Entidad distinta al Pliego, sea por efecto de un contrato celebrado con una Entidad Privada a título oneroso, o, con una Entidad Pública, sea a título oneroso o gratuito.

3.2.5 Indicadores de Gestión Municipal

Los indicadores son medidas cuantitativas que se desarrollan recolectando datos y se expresan a través de fórmulas matemáticas, tablas o gráficos. Estos son útiles para medir los resultados obtenidos de un tema de análisis con el fin de obtener un diagnóstico de la situación, comparar características de una población o para evaluar las variaciones de un evento. Existe un consenso bastante generalizado sobre la consideración del indicador como un índice cuantitativo de carácter objetivo, expresado tanto en términos monetarios como en unidades físicas o técnicas, con cuyo diseño se pretende explicar, revelar o medir de forma aproximada la presencia, el grado o la intensidad de un fenómeno, de un sistema o de una variable con la que se asoció en su diseño. Los indicadores desempeñan dos funciones, una descriptiva sobre el conocimiento del estado y evolución del sistema, y otra valorativa, la cual permite apreciar los efectos provocados por una actuación.

En sentido amplio, podría afirmarse que los indicadores de gestión hacen referencia a aquellas variables que tratan de representar y medir las actividades que forman el proceso productivo de una entidad, con el objetivo de mostrar su realidad y los efectos por él provocados.

En este caso los indicadores de Gestión Municipal son las medidas que reflejan el trabajo de la entidad respecto a sus ingresos, gastos y el bienestar de la población.

En el ámbito de la administración pública, se han planteado una gran amplitud de propuestas de indicadores de gestión, muchas de las cuales asignan diferentes denominaciones a un mismo concepto.

Desde el punto de vista de su clasificación, asumiendo la ordenación más generalmente aceptada entre los usuarios de estas herramientas, en una primera fase se diferenciarían dos niveles: indicadores primarios e indicadores secundarios. Los indicadores primarios son asimilables a datos estadísticos cuyo origen reside en la suma de hechos homogéneos y repetitivos. Los indicadores secundarios son combinaciones de indicadores primarios y dependen del efecto o aspecto que se pretenda medir (economía, calidad, etc.) en relación con el control de un factor concreto, motivo por el que se denominan indicadores de productividad parcial.

En un tercer nivel se encontrarían las relaciones funcionales o indicadores terciarios. Estos indicadores alcanzan su significado al ser utilizados de forma complementaria a los anteriores en la medición y análisis de la actuación del sector público y tratan de evaluar dichos aspectos a través de modelos formados por una combinación de indicadores primarios y secundarios que objetivamente explicarán el servicio, programa o área en su totalidad. Presentan la ventaja de superar los problemas que los indicadores de productividad parcial plantean en aquellos organismos que utilizan múltiples inputs para producir varios outputs.



La utilización de un sistema de indicadores debe ir acompañada, posteriormente, por un análisis objetivo de los resultados que permita la obtención de conclusiones útiles para la mejora de la gestión. Para ello es preciso disponer de referencias con las cuales comparar los resultados expresados por los indicadores siendo éste, precisamente, el campo de actuación del benchmarking, la disciplina que se ocupa de establecer comparaciones entre indicadores.

Los indicadores de gestión se han desarrollado en el ámbito de la administración para facilitar la toma de decisiones y la rendición de cuentas, tanto ante los órganos superiores o políticos como ante terceros interesados, dotándola al mismo tiempo de transparencia. Más recientemente, han comenzado a desempeñar un papel estratégico, pasando a convertirse en un elemento que permite la divulgación de las mejores prácticas operativas mediante procedimientos de benchmarking entre entes que prestan idénticos servicios en diferentes zonas geográficas.

El interés por el uso de indicadores de gestión en el sector público es muy reciente, hasta el punto de que se puede afirmar, sin temor a equivocarse, que se ha gestado en el último decenio.

Uno de los documentos más completo elaborado para el sector de las entidades locales es el Sistema de Indicadores de Gestión para los Ayuntamientos, realizado por AECA (2002), en cuyo desarrollo ha participado un numeroso grupo de profesionales entre los que se inscriben profesores universitarios, auditores y funcionarios públicos.

Los indicadores propuestos por AECA se han diseñado en función de áreas de servicios, de acuerdo con la siguiente clasificación:

- Protección civil y seguridad ciudadana.
- Seguridad, protección y promoción social.
- Producción de bienes públicos de carácter social.
- Producción de bienes públicos de carácter económico.
- Actividades generales.
- Gestión financiera, económica y presupuestaria.

Otros indicadores financieros, los cuales serán utilizados en el desarrollo del proyecto, son presentados por Daniel Scandizzo y Cesar Murua en su publicación “Las transferencias financieras a los gobiernos locales en la Provincia de Córdoba, Argentina: La necesidad de un modelo equitativo y transparente”. Los autores indican que el indicador financiero propuesto para el nuevo régimen de coparticipación busca reflejar la evolución en el equilibrio y la eficiencia de la administración de los recursos públicos locales. Este indicador tiene dos dimensiones fundamentales:

- Esfuerzo fiscal: es una medida de desempeño financiero de las finanzas públicas locales estimado a partir del resultado financiero, el ahorro público y la recaudación propia;
- Comportamiento del gasto: refleja nuevas funciones asumidas en la prestación de servicios o inversiones de capital llevadas a cabo por los gobiernos locales.

Esfuerzo Fiscal: Permite precisar el nivel de equilibrio entre Ingresos y Gastos. Por un lado se calcula la **Tasa de Ahorro corriente (AC)**,



$$AC = \frac{(Y_c - G_c)}{Y_c} \times 100$$

Donde “Y_c” es Ingreso corriente y “G_c” Gasto corriente. Cuando el valor supere el 10%, se interpretará como una señal de eficiencia y equilibrio. Asimismo, la evolución a lo largo de los años, en sentido positivo, también será interpretada positivamente. Por otro lado, se calcula el **Grado de Autonomía Financiera (AF)** a través de:

$$AF = \frac{Y_m}{G_c} \times 100$$

Donde “Y_m” representa a los Ingresos de jurisdicción municipal y “G_c” a los Gastos corrientes. En el caso de este indicador, que mide el porcentaje de Gastos corrientes cubiertos con Ingresos propios de la localidad, un valor de 50% es considerado como punto mínimo de referencia general.

Comportamiento del Gasto: Describe el destino de las erogaciones realizadas, para determinar los niveles de eficiencia a la hora de gastar los recursos públicos. Para ello, se calcula el **Gasto en Personal (GP)** a partir de la fórmula:

$$GP = \frac{G_p}{G_t} \times 100$$

Donde “G_p” es el monto destinado por cada localidad para cubrir la partida de Gastos en personal (Sueldos, Jornales y todo tipo de Remuneraciones por trabajos prestados al Municipio o Comuna) y “G_t” es el Gasto total acumulado por la localidad para el período considerado. En general, este indicador no debe superar el 50%. Para estimar la **Inversión Productiva (IP)**, el otro componente del indicador se debe calcular:

$$IP = \frac{I_f}{G_t} \times 100$$

“I_f” es Inversión física, es decir, la porción del Gasto de capital destinado a Obras públicas e inversión en Bienes de capital (maquinarias, equipos, etc.), sin considerar los pagos de Servicios de la deuda, y “G_t” es el Gasto total acumulado por la localidad para el período considerado. Cuando las localidades tienen, al menos, un 25% de Inversión Productiva, puede afirmarse que sus gastos tienen un destino apropiado.

La combinación de disminución en el Gasto de Personal con incremento de Inversión Productiva es una señal de eficiencia en la gestión de los recursos.

Los indicadores mencionados serán utilizados para el desarrollo del Data Warehouse, permitiendo a los usuarios conocer estos datos de suma importancia para comprender la actualidad y futuro del municipio.



SEGUNDA PARTE



3.3 Toma de Decisiones y Business Intelligence

En esta sección se cubrirán conceptos acerca de la toma de decisiones y definiciones propias de la solución de Inteligencia de Negocios.

3.3.1 ¿Qué es la Toma de decisiones?

La toma de decisiones es la selección de un curso de acción entre varias opciones. Un aspecto fundamental en la toma de decisiones es la percepción de la situación por parte del individuo o grupo de personas implicadas, es decir, determinada circunstancia puede ser percibida por una persona como un problema y por otra como una situación normal o hasta favorable.

Este proceso consiste, básicamente, en elegir una opción entre las disponibles, a los efectos de resolver un problema actual o potencial (aun cuando no se evidencie un conflicto latente).

La toma de decisiones a nivel individual se caracteriza por el hecho de que una persona haga uso de su razonamiento y pensamiento para elegir una solución a un problema que se le presente en la vida; es decir, si una persona tiene un problema, deberá ser capaz de resolverlo individualmente tomando decisiones con ese específico motivo.

En la toma de decisiones importa la elección de un camino a seguir, por lo que en un estado anterior deben evaluarse alternativas de acción. Si estas últimas no están presentes, no existirá decisión.

Para tomar una decisión, cualquiera que sea su naturaleza, es necesario conocer, comprender, analizar un problema, para así poder darle solución. En algunos casos, por ser tan simples y cotidianos, este proceso se realiza de forma implícita y se soluciona muy rápidamente, pero existen otros casos en los cuales las consecuencias de una mala o buena elección pueden tener repercusiones en la vida y si es en un contexto laboral en el éxito o fracaso de la organización, para los cuales es necesario realizar un proceso más estructurado que puede dar más seguridad e información para resolver el problema.

El administrador pasa la mayor parte de su tiempo resolviendo problemas y tomando decisiones. Tomar la decisión correcta cada vez es la ambición de quienes practican la gerencia. Hacerlo requiere contar con un profundo conocimiento, y una amplia experiencia en el tema.

La mayor parte de las decisiones gerenciales carecen de estructura y conllevan riesgo, incertidumbre y conflicto.

3.3.2 Tipos de decisiones

La clasificación más general que se realiza de las decisiones, es la siguiente:

Decisiones programadas: como el tiempo es valioso y escaso, deben tenerse identificadas aquellas situaciones que, por su recurrencia o importancia relativa, puedan ser tipificadas de manera tal que, al ocurrir, ya se tenga decidido lo que se debe hacer.

Son aquellas que se toman frecuentemente, es decir son repetitivas y se convierte en una rutina tomarlas; como el tipo de problemas que resuelve y se presentan con cierta



regularidad ya que se tiene un método bien establecido de solución y por lo tanto ya se conocen los pasos para abordar este tipo de problemas, por esta razón, también se las llama decisiones estructuradas. La persona que toma este tipo de decisión no tiene la necesidad de diseñar ninguna solución, sino que simplemente se rige por la que se ha seguido anteriormente.

Las decisiones programadas se toman de acuerdo con políticas, procedimientos o reglas, escritas o no escritas, que facilitan la toma de decisiones en situaciones recurrentes porque limitan o excluyen otras opciones.

Las decisiones programadas se usan para abordar problemas recurrentes. Sean complejos o simples. Si un problema es recurrente y si los elementos que lo componen se pueden definir, pronosticar y analizar, entonces puede ser candidato para una decisión programada. En cierta medida, las decisiones programadas limitan nuestra libertad, porque la persona tiene menos espacio para decidir qué hacer. No obstante, el propósito real de las decisiones programadas es liberarnos. Las políticas, las reglas o los procedimientos que usamos para tomar decisiones programadas nos ahorran tiempo, permitiéndonos con ello dedicar atención a otras actividades más importantes.

Decisiones no programadas: involucran situaciones, imprevistas o muy importantes que requieren una solución específica y particular por parte de los implicados. Las decisiones no programadas por la importancia de la situación que involucran, se reservan a personal de un nivel superior debido al impacto de sus consecuencias en la organización.

También denominadas no estructuradas, son decisiones que se toman ante problemas o situaciones que se presentan con poca frecuencia, o aquellas que necesitan de un modelo o proceso específico de solución. Las decisiones no programadas abordan problemas poco frecuentes o excepcionales. Si un problema no se ha presentado con la frecuencia suficiente como para que lo cubra una política o si resulta tan importante que merece trato especial, deberá ser manejado como una decisión no programada.

3.3.3 Proceso de la toma de decisiones

Cuando se comienza a buscar en la bibliografía disponible tanto impresa como digital, se encuentra con que no existe una única definición de los pasos exactos que se deben seguir en el proceso de toma de decisiones, aunque la mayoría de los autores mantiene una línea central que se debe seguir para implementarlo de modo adecuado y óptimo.

Proceso racional de toma de decisiones: de los procesos existentes para la toma de decisiones, la mayoría coincide en los siguientes pasos.

1.- *Determinar la necesidad de una decisión.* El proceso de toma de decisiones comienza con el reconocimiento de que se necesita tomar una decisión. Ese reconocimiento lo genera la existencia de un problema o una disparidad entre cierto estado deseado y la condición real del momento.

2.- *Identificar los criterios de decisión.* Una vez determinada la necesidad de tomar una decisión, se deben identificar los criterios que sean importantes para la misma.



3.- *Asignar peso a los criterios.* Los criterios enumerados en el paso previo no tienen igual importancia. Es necesario ponderar cada uno de ellos y priorizar su importancia en la decisión.

4.- *Desarrollar todas las alternativas.* Desplegar las alternativas. La persona que debe tomar una decisión tiene que elaborar una lista de todas las alternativas disponibles para la solución de un determinado problema.

5.- *Evaluar las alternativas.* La evaluación de cada alternativa se hace analizándola con respecto al criterio ponderado. Una vez identificadas las alternativas, el tomador de decisiones tiene que evaluar de manera crítica cada una de ellas. Las ventajas y desventajas de cada alternativa resultan evidentes cuando son comparadas.

6.- *Seleccionar la mejor alternativa.* Una vez seleccionada la mejor alternativa se llegó al final del proceso de toma de decisiones. En el proceso racional, esta selección es bastante simple. El tomador de decisiones sólo tiene que escoger la alternativa que tuvo la calificación más alta en el paso número cinco. El paso seis tiene varios supuestos, es importante entenderlos para poder determinar la exactitud con que este proceso describe el proceso real de toma de decisiones administrativas en las organizaciones. El tomador de decisiones debe ser totalmente objetivo y lógico a la hora de tomarlas. Tiene que tener una meta clara y todas las acciones en el proceso de toma de decisiones llevan de manera consistente a la selección de aquella alternativa que maximizará la meta.

El proceso racional de toma de decisiones cuenta con una serie de supuestos, los cuales son:

- ✓ Orientada a un objetivo. Cuando se deben tomar decisiones, no deben existir conflictos acerca del objetivo final. El lograr los fines es lo que motiva que tengamos que decidir la solución que más se ajusta a las necesidades concretas.
- ✓ Todas las opciones son conocidas. El tomador de decisiones tiene que conocer las posibles consecuencias de su determinación. Así mismo tiene claros todos los criterios y puede enumerar todas las alternativas posibles.
- ✓ Las preferencias son claras. Se supone que se pueden asignar valores numéricos y establecer un orden de preferencia para todos los criterios y alternativas posibles.



3.3.4 El Nuevo Directivo Racional – Análisis de Decisiones



Si bien en el punto anterior se expuso el proceso racional de toma de decisiones que puede ser encontrado en distintos materiales bibliográficos digitales o impresos, a continuación se presenta el contenido del libro “EL NUEVO DIRECTIVO RACIONAL” de Kepner y Tregoe, el cual es utilizado como bibliografía obligatoria de la materia Toma de Decisiones y expone un nuevo modelo para el análisis de situaciones, comprensión de problemas y tomar decisiones. Los autores de este material, dividen al proceso de toma de decisiones en 4 etapas: Análisis de Situación, Análisis de Problema, Análisis de Decisión y Análisis de Problema Potenciales.

Será descripta únicamente la etapa de Análisis de Decisión, puesto que será la utilizada en la parte práctica del proyecto (para comparar y seleccionar el conjunto de herramienta a implementar).

Análisis de Decisiones: En todas las organizaciones deben tomarse decisiones y realizarse acciones. Pero aun así, la toma de decisiones organizacional con frecuencia no es tan buena como debería. Al carecer de procedimientos imparciales y aceptados por todos, la toma de decisiones se convierte en una contienda entre los que sostienen distintos puntos de vista. Prevalecen los que detentan más poder. Los demás aceptan las decisiones para no quedar en ridículo y evitar una confrontación directa.

Es por esto que es necesario proporcionar un enfoque común para la toma de decisiones, cuando esto sucede las personas descubren que pueden trabajar como equipo, compartiendo la información pertinente y reconciliando mejor las posturas opuestas porque el proceso de toma de decisiones es menos tendencioso, mejorando la calidad de la toma de decisiones.

El análisis de decisiones es un procedimiento sistemático basado en el patrón de razonamiento utilizando normalmente por cualquier persona. Las técnicas representan ampliaciones y refinamientos de los elementos de este patrón de razonamiento, el cual se describe en:

- ✓ Se debe realizar una elección
- ✓ Se consideran determinados factores específicos que deben ser satisfechos si la elección ha de tener éxito.
- ✓ Se determina qué tipo de acción satisfará mejor dichos factores.
- ✓ Se consideran los riesgos que podrían vincularse a la elección final de acción los cuales podrían poner en peligro su seguridad y éxito.

Sobre cada decisión existe un cierto grado de incertidumbre, pues todas las decisiones acaban subiendo al escenario en algún momento del futuro incierto. Hacer buenas elecciones depende de tres elementos

1. La calidad de la definición de los factores específicos que deben ser satisfechos
2. La calidad de la evaluación de las alternativas disponibles
3. La calidad de la comprensión de lo que pueden producir esas alternativas (para bien o para mal).

El propósito del Análisis de Decisiones es identificar lo que necesita hacerse, desarrollar los criterios específicos para su realización, evaluar las alternativas disponibles con respecto a esos criterios e identificar los riesgos implicados.

Elementos del análisis de decisiones:



Enunciado de la decisión o del propósito: El proceso comienza con un enunciado de la decisión o nombre de ésta, la resolución consistirá en una respuesta a las preguntas ¿Con qué propósito?, ¿Cuál? y ¿Cómo? El enunciado de la decisión da un enfoque para todo lo que sigue y establece los límites de la elección. Como consecuencia de éste se definirán los criterios de selección, las alternativas y la elección final, por ello la manera en que se redacta merece una cuidadosa atención. Siempre indica algún tipo de acción, el resultado esperado y el nivel al que debe tomarse la decisión.

Los objetivos para la decisión: son los criterios o los detalles específicos que debe cumplir la decisión. Se establecen una vez que se ha enunciado el propósito de la decisión y acordado a qué nivel se tomara. Se definen antes de discutir las alternativas y en algunas ocasiones antes de identificarlas. El Análisis de Decisiones es la antítesis de la identificación de un curso de acción elaborando después un argumento que lo justifique, en lugar de ello, parte de lo que necesita realizarse para llegar a la alternativa que mejor pueda realizarlo.

Los objetivos son medidas claras de los fines que se quieren lograr, solamente con medidas claras se podrán hacer elecciones razonadas.

Hay 2 clases de objetivos:

- ✓ **Obligatorios:** son imprescindibles y deben cumplirse para garantizar una decisión exitosa. Estos deben ser cuantificables porque funcionan como un filtro para eliminar las alternativas propensas al fracaso (PASA O NO PASA). Cuando llega el momento de evaluar las alternativas en función de estos objetivos, cualquier alternativa que no satisfaga un objetivo obligatorio queda inmediatamente descartada.
- ✓ **Deseados:** las alternativas se juzgaran con base en sus resultados relativos de desempeño frente a los objetivos deseados, y no en si los satisfacen o no. La función de estos objetivos consiste en dar una idea comparativa de las alternativas, un sentido de cuál podría ser el resultado de cada alternativa en comparación con las demás. Un objetivo “deseado” puede ser imprescindible, pero no puede clasificarse como “obligatorio” por una o dos razones: por no ser cuantificable, por lo cual no puede darnos un juicio absoluto de Si o No sobre los resultados de una alternativa, o porque no se desea un juicio absoluto de Si o No, sino que se prefiere usar ese objetivo como una medida relativa del desempeño. Se debe ponderar los objetivos para saber cuál es la importancia relativa, de 10 (+ importante) a 1 (- importante).

“Los obligatorios deciden quien participa en el juego, pero los deseados deciden quién gana”

Generación y evaluación de alternativas: la alternativa ideal satisface perfectamente todas las condiciones que se le establecen y no presenta dificultades. Pero la realidad no es así, por lo que se debe evaluar cada alternativa disponible contra los objetivos y el riesgo aceptable. Si se debe elegir entre varias alternativas se tendrá que decidir cuál satisfará mejor los objetivos con el menor riesgo aceptable, buscando una elección equilibrada. Una alternativa que mejor satisface los objetivos pero que conlleva serios riesgos no es la mejor elección.



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

Durante este paso del Análisis de Decisiones se pueden dar las siguientes situaciones: la primera ocurre cuando solo se tiene una alternativa, en este caso se debe decidir si es lo suficiente buena para aceptarla, en este caso lo que se realiza es compararla con una alternativa perfecta pero inalcanzable. El segundo caso ocurre cuando se debe elegir entre la línea de acción actual y una propuesta, entonces ambas deben considerarse como alternativas, evaluando sus resultados contra los objetivos de la misma manera en que se haría si ambas alternativas hubieran sido propuestas. Otra ocasión que podría presentarse sería no contar con alguna alternativa, donde se deberá crear algo nuevo, en este caso generalmente se elabora una alternativa con los componentes disponibles y luego se selecciona las mejores y más factibles combinaciones y se la contrasta con la alternativa ideal.

Una vez que se han definido las diferentes alternativas se las debe contrastar contra los objetivos, comenzando por los obligatorios y continuando con los deseados.

Evaluación de las alternativas contra los objetivos OBLIGATORIOS. Aquí se determina si la alternativa cumple o no el objetivo, aquella que lo satisfaga continuará y la que no quedará excluida.

Objetivos Obligatorios	Alternativa 1		Alternativa 2		Alternativa 3	
	Información adicional	Pasa/No pasa	Información adicional	Pasa/No pasa	Información adicional	Pasa/No pasa
Ob. 1						

Evaluación de las alternativas contra los objetivos DESEADOS. Aquí se evalúan las diferentes alternativas que continúan en camino, contra los objetivos deseados obteniendo un puntaje de acuerdo a sus características específicas. A modo de ejemplo se presenta un gráfico con algunos valores para analizar la manera en que se calcula la calificación teniendo en cuenta la ponderación de los objetivos.

Objetivos Deseados	Calificación									
	Peso	Alternativa 1			Alternativa 2			Alternativa 3		
		Información adicional	calificación	Calificación ponderada	Información adicional	calificación	Calificación ponderada	Información adicional	calificación	Calificación ponderada
Ob. 1	10		7	70		10	100		6	60
Ob. 2	8		1	8		8	64		10	80
Totales			-	78		-	164		-	140

Calificación ponderada es igual a: peso * calificación.

ADVERTENCIA: si se observa que ninguna alternativa satisface de modo correcto los objetivos, puede estar ocurriendo que los objetivos son irreales o que ninguna alternativa es real y alcanzable. Por otro lado si todas las alterativas califican en todos los objetivos de manera muy alta, quiere decir que los objetivos son muy vagos.

En este punto, se deberá realizar una elección provisional, es decir se debe elegir la alternativa con mayor puntuación.



Consecuencias de la alternativa: en el último paso se buscan las posibles consecuencias adversas de todas las alternativas factibles y serán evaluadas exhaustivamente, por separado, para tomar medidas en el presente que reduzcan su efecto en el futuro (se debe evaluar el riesgo de cada alternativa). Las consecuencias negativas de cualquier acción son tan tangibles como sus beneficios. Una vez que una decisión es implementada cualquiera de sus efectos negativos puede llegar a convertirse en un verdadero problema, por ello es necesario explorar y evaluar exhaustivamente las posibles consecuencias adversas de cualquier alternativa antes de tomar una decisión definitiva. Se deben identificar las posibles consecuencias adversas, habiéndolas reconocido y evaluado, pudiendo hacer que se eviten en su todo o tomar medidas en el presente que reduzcan su efecto en el futuro. El riesgo vinculado a una alternativa no es necesariamente un factor totalmente condenatorio, siempre que se detecte cuando todavía es tiempo de remediarlo.

Para encontrar consecuencias se deben realizar preguntas como las siguientes:

- ✓ ¿Qué requisitos para tener éxito han sido pasados por alto en las etapas anteriores de este análisis?
- ✓ ¿Qué factores dentro de la organización, con base en la experiencia, podrían perjudicar su aceptación o su implementación?
- ✓ ¿Qué tipos de cambios dentro de la organización podrían perjudicar su éxito a largo plazo?
- ✓ ¿Qué tipos de cambios externos (como actividades de la competencia y reglamentos del gobierno) podrían perjudicar su éxito a largo plazo?
- ✓ ¿Qué tipos de cosas tienden a causar problemas en la implantación de este tipo de decisión?

El siguiente paso consiste en calificar las consecuencias adversas de una alternativa con base en la probabilidad y la gravedad.

Consecuencias adversas	Alternativa 1		Alternativa 2		Alternativa 3	
	¿Probabilidad?	¿Gravedad si ocurre?	¿Probabilidad?	¿Gravedad si ocurre?	¿Probabilidad?	¿Gravedad si ocurre?

Una vez analizadas las posibles consecuencias, se procederá a elegir la alternativa definitiva.

3.3.5 Necesidad de información

El proceso de toma de decisiones utiliza como materia prima a la información. Ésta es fundamental, ya que sin ella no resultaría posible evaluar las opciones existentes o desarrollar opciones nuevas.

En las organizaciones, que se encuentran sometidas constantemente a la toma de decisiones, la información adquiere un rol fundamental, y por ello un valor inigualable.

A medida que la competencia y la velocidad de los negocios han aumentado, también ha incrementado la necesidad de obtener información de manera más rápida y eficiente. La cantidad de datos disponible para cada empresa crece de manera exponencial,



presentando un reto en su manejo coherente, y la producción de los reportes necesarios para apoyar el proceso de toma de decisiones.

Actualmente, en las actividades diarias de cualquier organización, se generan datos como producto secundario, que son el resultado de todas las transacciones que se realizan. Es muy común, que los mismos se almacenen y administren a través de sistemas transaccionales en bases de datos relacionales, pero la idea es que estos dejen de ser solo simples datos, para convertirse en información que enriquezca las decisiones de los usuarios, por ello es que se han desarrollados sistemas especialmente diseñados para ayudar a transitar el proceso de toma de decisiones, que se conocen como sistemas de soporte a decisiones o sistemas de apoyo a la decisión.

Las herramientas clásicas de reportes, las cuales usualmente acceden a bases de datos normalizadas (OLTP) suelen tomar mucho tiempo en ejecutar consultas, requieren personal diestro en su construcción y carecen de flexibilidad. Para suplir estas necesidades satisfactoriamente se introduce el concepto de Data Warehouse, junto con herramientas poderosas en explorar el contenido de la información. A esta combinación de herramientas y filosofías de desarrollo se le conoce como inteligencia de negocio (business intelligence - BI), tema que se desarrollará a continuación.

3.3.6 ¿Qué es Business Intelligence (BI)?

De acuerdo a la definición del DataWarehouse Institute se define como: “La combinación de tecnologías, herramientas y procesos que permiten transformar los datos almacenados en información y la información en conocimiento, destinado a optimizar el proceso de toma de decisiones en los negocios”.



Es decir, si reunimos, depuramos y transformamos los datos de los sistemas transaccionales en información, luego podemos analizar esa información y convertirla en conocimiento que de soporte a la toma de decisiones.

En otras palabras la inteligencia de negocios es el conjunto de tecnologías, aplicaciones y herramientas enfocadas al procesamiento de datos en una empresa para poder ayudar a la toma de decisiones y hacer análisis de los datos. Con Business Intelligence podemos realizar gestión de datos, consultar a estos y generar informes/reportes; es decir, todo el procesamiento para traducir la lógica de negocios a la lógica de sistemas empresariales.

A su vez, mientras mayor conocimiento tenga una organización respecto a un negocio, mayor será su ventaja competitiva y mejores decisiones podrá tomar. Esto se debe simplemente a que, por ejemplo, cuanto más se conoce a los clientes, mejor se logra satisfacer sus necesidades.



BI apoya a los tomadores de decisiones brindando la información correcta, en el momento y lugar oportuno, lo cual incrementa la efectividad de cualquier empresa, generando una potencial ventaja competitiva, que surge como consecuencia de disponer de dicha información para responder a los problemas de negocio.

La inteligencia de negocios (Business Intelligence - BI), permite que el proceso de toma de decisiones esté fundamentado sobre un amplio conocimiento de sí mismo y del entorno, minimizando de esta manera el riesgo y la incertidumbre. Además, propicia que las organizaciones puedan traducir sus objetivos en indicadores de estudio, y que estos puedan ser analizados desde diferentes perspectivas, con el fin de encontrar información que no solo se encargue de responder a preguntas de lo que está sucediendo o ya sucedió, sino también, que posibilite la construcción de modelos, mediante los cuales se podrán predecir eventos futuros.

La aplicación de soluciones BI no es solo para grandes-medianas empresas, sino para quien desee tomar decisiones a través del análisis de sus datos.

Se describe a la BI, como un concepto que integra por un lado el almacenamiento y por el otro el procesamiento de grandes cantidades de datos, con el principal objetivo de transformarlos en conocimiento y en decisiones en tiempo real, a través de un sencillo análisis y exploración. Este conocimiento debe ser oportuno, relevante, útil y debe estar adaptado al contexto de la organización.

Al contar con la información exacta y en tiempo real, es posible, además de los beneficios que han sido mencionados, identificar y corregir situaciones antes de que se conviertan en problemas y en potenciales pérdidas de control de la empresa, pudiendo conseguir nuevas oportunidades o readaptarse frente a la ocurrencia de sucesos inesperados.

La Inteligencia de Negocios tiene sus raíces en los Sistemas de Información Ejecutiva (Executive Information Systems – EIS) y en los Sistemas para la Toma de Decisiones (Decision Support Systems – DSS), pero ha evolucionado y se ha transformado en todo un conjunto de tecnologías capaces de satisfacer a una gran gama de usuarios junto a sus necesidades específicas en cuanto al análisis de información.

En la actualidad, los principales componentes de orígenes de datos en Business Intelligence son:

- Datamart
- Data Warehouse

3.3.7 ¿Por qué implementar Business Intelligence?

Para que una organización alcance el éxito debe ser capaz de tomar decisiones de negocio precisas y de forma rápida, pero en muchas ocasiones ocurre lo siguiente:

- ✓ **Se poseen los datos pero se carece de la información respectiva para tomar las decisiones.** Por ejemplo, se almacenan los datos de los contribuyentes, pero no es posible distinguir patrones de comportamientos entre ellos, faltando información para plantear una estrategia referente a cobranzas.
- ✓ **Falta integración.** Existen múltiples sistemas transaccionales en las diversas áreas, pero carecen de una visión global de la organización. Sus bases de datos

no suelen estar integradas, lo que implica la existencia de islas de información, lo cual dificulta la toma de decisiones.

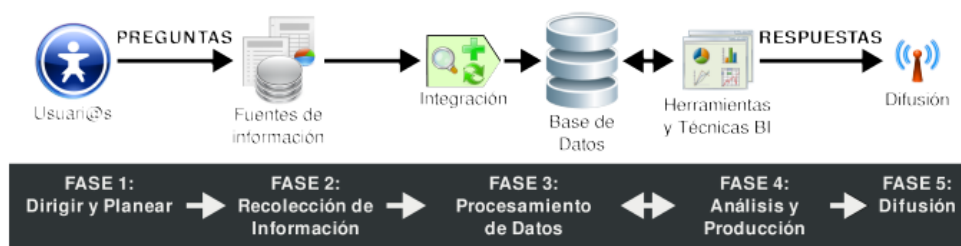
- ✓ **Existen largos tiempos de respuesta y poca agilidad.** A medida que las preguntas que se buscan responder se vuelven más complejas, y el volumen de datos aumenta, los tiempos de respuesta se traducen en una incómoda espera que obstaculiza la fluidez del trabajo y problemas en el funcionamiento normal del sistema operacional.
- ✓ **Escasa flexibilidad.** En muchas ocasiones el usuario tiene que amoldarse a los informes predefinidos que se configuraron en el momento de la implementación de un sistema, los cuales no siempre responden a sus dudas reales

Business Intelligence se apoya en un conjunto de herramientas que facilitan la extracción, transformación, el análisis y almacenamiento de los datos generados en una organización, a fin de generar conocimiento y apoyar la toma de decisiones de sus respectivos usuarios, solucionando los inconvenientes recientemente mencionados.

3.3.8 El proceso de BI

El proceso esta dividido en cinco fases:

- **FASE 1: Dirigir y Planear.** En esta fase inicial es donde se deberán recolectar los requerimientos de información específicos de los diferentes usuarios, así como entender sus diversas necesidades, para que luego en conjunto con ellos se generen las preguntas que ayudarán a alcanzar sus objetivos.
- **FASE 2: Recolección de Información.** Es aquí en donde se realiza el proceso de extraer desde las diferentes fuentes de información de la empresa, tanto internas como externas, los datos que serán necesarios para encontrar las respuestas a las preguntas planteadas en el paso anterior.
- **FASE 3: Procesamiento de Datos.** En esta fase es donde se integran y cargan los datos en crudo en un formato utilizable para el análisis. Esta actividad puede realizarse mediante la creación de una nueva base de datos, agregando datos a una base de datos ya existente o bien consolidando la información.
- **FASE 4: Análisis y Producción.** Ahora, se procederá a trabajar sobre los datos extraídos e integrados, utilizando herramientas y técnicas propias de la tecnología BI, para crear inteligencia. Como resultado final de esta fase se obtendrán las respuestas a las preguntas, mediante la creación de reportes, indicadores de rendimiento, cuadros de mando, gráficos estadísticos, etc.
- **FASE 5: Difusión.** Finalmente, se les entregará a los usuarios que lo requieran las herramientas necesarias, que les permitirán explorar los datos de manera sencilla e intuitiva





3.3.9 Beneficios de BI

Dentro de los beneficios que ofrece una solución BI es importante destacar que la misma permite:

- ✓ Conseguir las respuestas de aquellas preguntas que son claves para el desarrollo y éxito de una organización.
- ✓ Generar reportes e informes de manera dinámica, ofreciendo alta flexibilidad y brindando la información en el momento oportuno.
- ✓ Reunir, normalizar y centralizar toda la información de la empresa, en un almacén de datos, logrando que todas las decisiones estratégicas se basen en la misma información, y eliminando así las islas que suelen producirse por la falta de integración.
- ✓ Realizar un seguimiento del plan estratégico de una organización, permitiendo crear, manejar y monitorizar las métricas y los objetivos estratégicos propuestos en ese plan, para poder detectar a tiempo las desviaciones, adoptando las acciones oportunas para corregirlas.
- ✓ Analizar la raíz de los problemas explorando la información de múltiples perspectivas en varios niveles de detalles.
- ✓ Obtener una visión global de la organización, a fin de dirigir la misma en la dirección correcta.
- ✓ Acceder y analizar directamente los indicadores de éxito.
- ✓ Establecer pautas de comportamiento, tendencias, evoluciones del mercado, cambios en el consumo o en la producción, que resulta prácticamente imposible reconocer sin el software adecuado.
- ✓ Predecir el comportamiento futuro con un alto porcentaje de certeza, basado en el entendimiento del pasado.
- ✓ Aprender de errores pasados, brindando datos históricos relevantes, a fin de que una organización aprenda de su historia y de sus mejores prácticas, y que pueda evitar tropezarse de nuevo con los mismos errores del pasado.
- ✓ Reduce el tiempo mínimo que se requiere para recoger toda la información relevante de un tema en particular, ya que la misma se encontrará integrada en una fuente única de fácil acceso.
- ✓ Automatiza la asimilación de la información, debido a que la extracción y carga de los datos necesarios se realizará a través de procesos predefinidos.
- ✓ Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones.
- ✓ Cierra el círculo que hace pasar de la decisión a la acción.
- ✓ Permite a los usuarios no depender de reportes o informes programados, porque los mismos serán generados de manera dinámica.
- ✓ Posibilita la formulación y respuesta de preguntas que son claves para el desempeño de la organización.
- ✓ Permite acceder y analizar directamente los indicadores de éxito.
- ✓ Se pueden identificar cuáles son los factores que inciden en el buen o mal funcionamiento de la organización.
- ✓ Se podrán detectar situaciones fuera de lo normal.
- ✓ Los usuarios podrán consultar y analizar los datos de manera sencilla e intuitiva.



3.4 Data Warehouse y Data Warehousing.

El nivel competitivo actual en las empresas, les exige desarrollar nuevas estrategias de gestión de uno de sus recursos más valiosos, la información.

Las organizaciones almacenan, en diferentes fuentes, datos tanto internos como externos de clientes, productos, servicios, estructura organizativa, canales de distribución, operaciones, personal, proveedores, competencia, mercado, etc.; para convertir estos datos en información y, mediante ella, generar conocimiento para apoyar la toma de decisiones, será necesario gestionarlos, depurarlos, integrarlos y almacenarlos en un solo destino que permita su posterior análisis y exploración. Por esto es fundamental contar con un proceso que satisfaga todas estas necesidades. Este proceso se conoce como Data Warehousing.

3.4.1 Data Warehousing

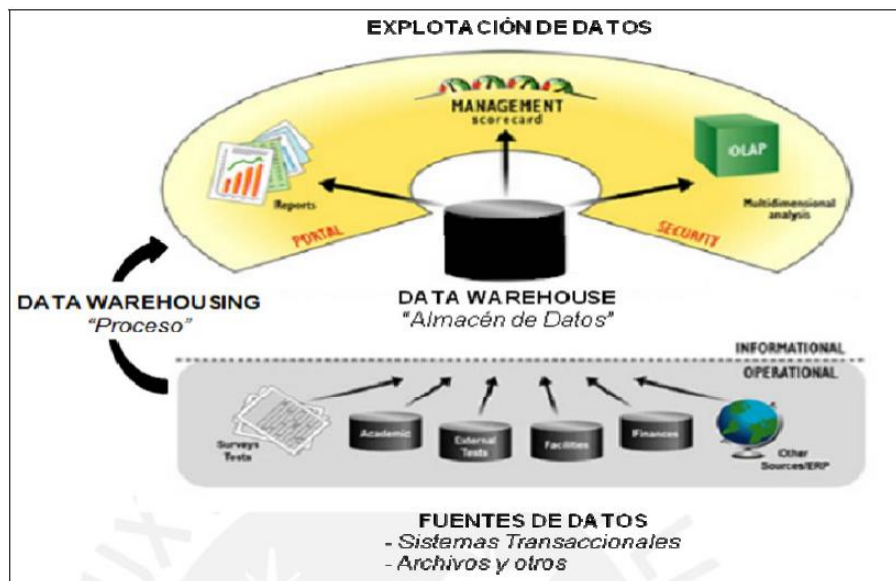
El término Data Warehousing se refiere al proceso de convertir la lógica de negocios (OLTP – Sistema Transaccional) a la lógica de sistemas empresariales (DSS – Sistema de Soporte de Decisiones). Este proceso se refiere a trabajar con los datos generados por las operaciones diarias de una organización, los cuales provienen de sistemas de información transaccionales o de diversas fuentes de información. Estos datos son transformados, integrados y almacenados en un almacén de datos (Data Warehouse) con la finalidad de acceder a ellos para mostrar información confiable y estructurada que permita realizar análisis para tomar decisiones en una organización.

El Data Warehousing posibilita la extracción de datos de sistemas operacionales y fuentes externas, permite la integración y homogeneización de los datos de toda la empresa, provee información que ha sido transformada y resumida, para que ayude en el proceso de toma de decisiones estratégicas y tácticas.

El Data Warehousing, convertirá los datos operacionales de la empresa en una herramienta competitiva, debido a que pondrá a disposición de los usuarios indicados la información pertinente, correcta e integrada, en el momento que se necesita.

Pero para que el Data Warehousing pueda cumplir con sus objetivos, es necesario que la información que se extrae, transforma y consolida, sea almacenada de manera centralizada en una base de datos con estructura multidimensional denominada Data Warehouse (DW).

Es posible resumir el término Data Warehousing como el proceso encargado de extraer, transformar, consolidar, integrar y centralizar los datos que la empresa genera en todos los ámbitos de su actividad diaria de negocios y la información externa relacionada. Permitiendo de esta manera el acceso, la manipulación flexible y exploración de la información requerida, a través de diferentes herramientas tecnológicas.



3.4.2 Data Warehouse (DW)

Como se ha mencionado para que el Data Warehousing pueda cumplir con sus objetivos, es necesario que la información que se extrae, transforma y consolida, sea almacenada de manera centralizada en una base de datos con estructura multidimensional denominada Data Warehouse.

Por esto es que se introduce el término Data Warehouse para referirse al repositorio central, y se utiliza el término Data Warehousing para referirse a la captura, recolección, filtrado, limpieza, depuración, carga, consolidación y establecimiento de relaciones entre la información proveniente de distintas fuentes, sobre la base de un modelo de información al servicio del negocio. Tal como se observa en la figura anteriormente expuesta.

Debido a las dificultades de los sistemas tradicionales de satisfacer las necesidades de información de las organizaciones actuales, surge el concepto de Data Warehouse, como solución a las necesidades de información globales de la empresa. Este término se traduce literalmente como Almacén de Datos. Sin embargo, si el Data Warehouse fuese exclusivamente un almacén de datos, los problemas seguirían siendo los mismos que en los Centros de Información.

Un Data Warehouse es un repositorio central de datos, donde están almacenados grandes volúmenes de información, provenientes tanto de transacciones detalladas como datos agregados de fuentes de distinta naturaleza y capturada en un periodo de tiempo significativo.

Comúnmente se dice que los DW son fuentes secundarias de información pues no generan datos por sí mismos, sino que son alimentados desde sistemas existentes internamente en la organización o desde datos externos. Típicamente los usuarios del DW tienen sólo permisos de lectura sobre este repositorio de datos. Los DW o bases de datos para procesamiento analítico (OLAP) están específicamente estructurados o diseñados para cumplir con un conjunto de metas bien diferentes a los objetivos de un



sistema operacional OLTP. Por ejemplo, una meta de los OLTP es maximizar la concurrencia de actualizaciones; dicho objetivo no es pertinente en el diseño de DW donde las consultas son sólo de lectura. En contrapartida, las metas de un diseñador de DW deben focalizarse en entregar un análisis multidimensional y capacidades de reportes ad hoc y brindarlos de manera eficiente.

La definición más famosa en el ámbito de los Data Warehouse es la propuesta por la persona que muchos consideran como el “padre” de los depósitos de datos, William Inmon, quien define a este concepto en función de las características de éstos.

“Un depósito de datos es una colección de base de datos integradas y orientadas al usuario, diseñadas con el objeto de apoyar a la función de decisión de una organización, en el cual, cada unidad de datos es relevante en un momento dado”¹.

Los datos que entran al depósito provienen en la mayoría de los casos del ambiente operacional. El depósito de datos es siempre un almacenamiento físicamente separado de datos transformados, que provienen de los datos de las aplicaciones que se encuentran en el ambiente operacional.

3.4.3 Características de un Data Warehouse.



Depósito de datos

Un depósito de datos es una colección de datos:

- * orientada a temas
- * integrada
- * variante en el tiempo
- * no volátil

que brinda soporte al proceso de toma de decisiones

3.4.3.1 Orientación a temas (Orientado al negocio)

La primera de las características de un depósito de datos es que está orientado en torno a los principales temas de la empresa, la información se clasifica en base a los aspectos que son de interés para la organización. Esta clasificación afecta el diseño y la implementación de los datos encontrados en el almacén de datos, debido a que la estructura del mismo difiere considerablemente a la de los clásicos procesos operacionales orientados a las aplicaciones. Si se realiza un análisis comparativo de los dos tipos de orientaciones sería posible ver que:

- ✓ Con respecto al nivel de detalle de los datos, el DW excluye la información que no será utilizada exclusivamente en el proceso de toma de decisiones; mientras que en los procesos orientados a las aplicaciones, se incluyen todos aquellos datos que son necesarios para satisfacer de manera inmediata los requerimientos funcionales de la actividad que soporten.
- ✓ Otra diferencia importante está en la interrelación de la información, los datos operacionales mantienen una relación continua entre dos o más tablas basadas en una regla comercial que está vigente; en cambio las relaciones encontradas en los datos residentes del data Warehouse son muchas, debido a que por lo general

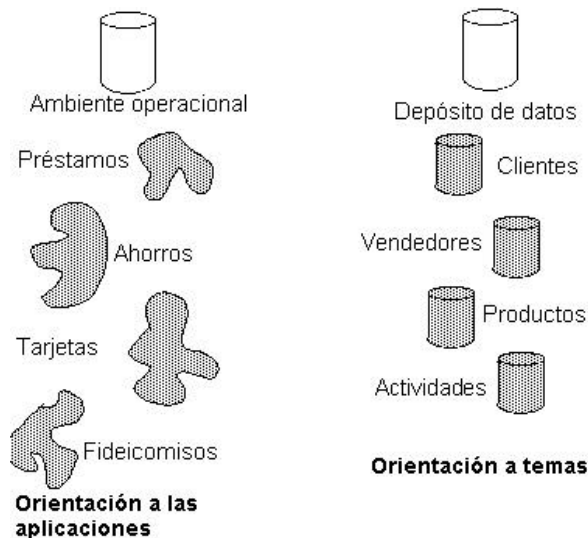
¹ Bill Inmon, reconocido mundialmente como el padre del Data Warehouse



cada tabla del mismo estará conformada por la integración de varias tablas u otras fuentes del ambiente operacional, cada una con sus propias reglas de negocios inherentes.

El origen de este contraste es totalmente lógico, ya que el ambiente operacional se diseña alrededor de las aplicaciones u programas que necesite la organización para llevar a cabo sus actividades diarias y funciones específicas; mientras que en contraposición el ambiente DW se organizará alrededor de entidades de alto nivel. Esto se debe a que el depósito de datos se diseña para realizar consultas e investigaciones sobre las actividades de la organización y no para soportar los procesos que se realizan en ella.

En síntesis, la ventaja de contar con procesos orientados a la aplicación, esta fundamentada en la alta accesibilidad de los datos, lo que implica un elevado desempeño y velocidad en la ejecución de consultas, ya que las mismas están predeterminadas; mientras que en el DW para satisfacer esta ventaja se requiere que la información este desnormalizada, es decir, con redundancia y que la misma esté dimensionada, para evitar tener que recorrer toda la base de datos cuando se necesite realizar algún análisis determinado, sino que simplemente la consulta sea enfocada por variables de análisis que permitan localizar los datos de manera rápida y eficaz, para poder de esta manera satisfacer una alta demanda de complejos exámenes en un mínimo tiempo de respuesta.



3.4.3.2 Integrado

Fácilmente se puede entender que el aspecto más importante de un ambiente de depósito de datos es que los datos que se encuentran dentro de él están integrados. La integración de datos se muestra de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de estructuras consistentes, en atributos físicos de los datos consistentes, fuentes múltiples y otros.

La integración implica que todos los datos de diversas fuentes que son producidos por distintos departamentos, secciones y aplicaciones, tanto internos como externos, deben ser consolidados en una instancia antes de ser agregados al DW, y deben por lo tanto ser



Cualquiera que sea la forma del diseño, el resultado debe ser el mismo, la información necesita ser almacenada en el Data Warehouse en un modelo globalmente aceptable y singular, aun cuando los sistemas operacionales subyacentes almacenen los datos de manera diferente.

3.4.3.3 Variante en el tiempo

Todos los datos del depósito de datos son exactos en algún momento en el tiempo. Esta característica básica de los datos en el depósito es muy diferente de los datos que se encuentran en el ambiente operacional. En el ambiente operacional los datos son exactos en el momento en que se acceden. En otras palabras, en el ambiente operacional, cuando se accede a una unidad de datos, se espera que ello refleje sus valores exactos al momento del acceso. Puesto que los datos en el depósito de datos son exactos en cualquier momento en el tiempo (es decir, no ahora mismo), los datos que se encuentran en éste se dice que son “variantes en el tiempo”.

Los datos históricos son de poco uso en el procesamiento operacional. La información del depósito por el contraste, debe incluir los datos históricos para usarse en la identificación y evaluación de tendencias.

La variación en el tiempo de los datos del depósito se muestra de diferentes formas. La primera y más simple de ellas es que los datos en el depósito representan datos sobre un horizonte de largo plazo (de entre cinco y diez años). El horizonte de tiempo que representa el ambiente operacional es mucho más corto (desde el valor actual hasta sesenta o noventa días). Las aplicaciones deben exhibir un buen desempeño y deben ser capaces de procesar transacciones que requieran una mínima cantidad de datos, si es que tienen algún grado de flexibilidad para ello. Por lo tanto, las aplicaciones operacionales tienen un corto horizonte de tiempo, como una cuestión propia de su etapa de diseño.

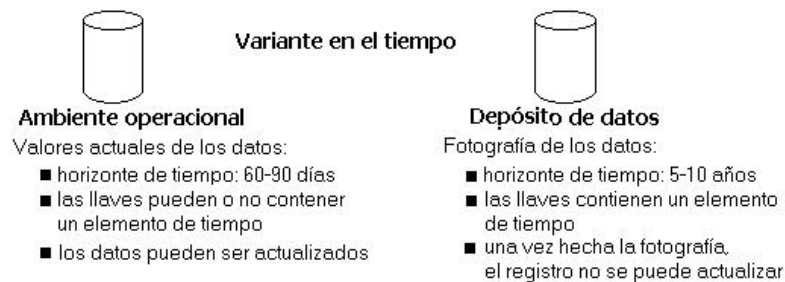
La segunda forma en que se muestra la variación en el tiempo en el depósito de datos es en la estructura clave. Cualquier estructura clave o información en el depósito de datos contiene (implícita o explícitamente) un elemento de tiempo, tal como día, semana, mes, entre otros. El elemento de tiempo está casi siempre al inicio de la llave concatenada que se encuentra en el depósito de datos. En ocasiones, el elemento de tiempo podría existir implícitamente, tal es el caso en que un archivo completo se duplique al final del mes o del trimestre.

Esto contribuye a una de las principales ventajas del almacén de datos: los datos son almacenados junto a sus respectivos históricos. Esta cualidad que no se encuentra en fuentes de datos operacionales, garantiza poder desarrollar análisis de la dinámica de la información, pues ella es procesada como una serie de instantáneas, cada una representando un periodo de tiempo. Es decir, que gracias al sello de tiempo se podrá tener acceso a diferentes versiones de la misma información.





La tercera forma como aparece la variación en el tiempo es que los datos de un depósito de datos, una vez que se graban correctamente, no pueden ser actualizados. Por lo tanto, los datos de un depósito de datos son, para todos los efectos, una gran serie de fotografías. De hecho, si la fotografía de los datos se ha tomado incorrectamente, entonces la fotografía puede ser cambiada. Pero se parte de que las fotografías fueron hechas apropiadamente, entonces éstas no pueden ser cambiadas una vez que fueron tomadas. En algunos casos puede ser falta de ética y hasta ilegal que las fotografías en el depósito de datos se alteren. Los datos operacionales, en procura de ser exactos al momento del acceso pueden ser actualizados conforme surjan las necesidades.



Es importante tener en cuenta la granularidad de los datos, así como también la intensidad de cambio natural del comportamiento de los fenómenos de la actividad que se desarrolle, para evitar crecimientos incontrolables y desbordamientos de la base de datos.

El intervalo de tiempo y periodicidad de los datos debe definirse de acuerdo a la necesidad y requisitos de los usuarios. Es elemental aclarar, que el almacenamiento de datos históricos, es lo que permite al DW desarrollar pronósticos y análisis de tendencias y patrones, a partir de una base estadística de información.

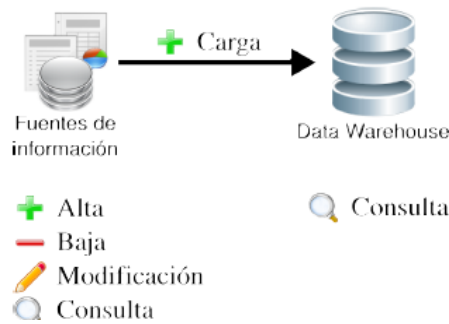
3.4.3.4 No volátil

La cuarta característica que define un depósito de datos es que no es volátil, y esto es importante porque la información es útil para el análisis y la toma de decisiones solo cuando es estable.

En contraste con el ambiente operacional, en el que las actualizaciones (sean éstas inserciones, borrados o cambios) son hechas regularmente sobre una base de registro por registro, en el depósito la manipulación básica de los datos es mucho más simple. Existen tan sólo dos clases de operaciones que pueden ocurrir en el depósito de datos: la carga inicial de los datos y el acceso a ellos. No existe actualización de los datos (en el sentido general de actualización) almacenados en el depósito de datos como parte normal del procesamiento. Los datos operacionales varían momento a momento, en cambio, los datos una vez que entran en el DW no cambian.



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



Existen algunas importantes consecuencias de las diferencias básicas entre el procesamiento operacional y el procesamiento de los depósitos de datos. Por ejemplo, en la etapa de diseño no es un factor preocupante la necesidad de tomar precauciones para que la anomalía de las actualizaciones afecte al depósito de datos, ya que como se ha mencionado no se hace actualización de ningún tipo. Esto significa que a nivel del diseño físico se pueden tener suficientes libertades como para optimizar el acceso a los datos, particularmente en el tratamiento de temas como la normalización y desnormalización física.

Otra consecuencia de la simplicidad de la operación de un depósito de datos está en la tecnología fundamental utilizada para correr el ambiente de depósitos de datos. Tener que soportar la actualización registro por registro en un modo de procesamiento en línea (que es el caso más frecuente del procesamiento operacional) hace que la tecnología deba tener un fundamento complejo por debajo de una fachada de simplicidad. La tecnología que ha de soportar respaldos y recuperaciones, transacciones e integridad de datos, así como la detección y resolución de entramamientos (deadlock) es bastante compleja e innecesaria para el procesamiento de un depósito de datos.



Las características mencionadas de los depósitos de datos apuntan a un ambiente que es muy, pero muy diferente de un clásico ambiente operacional.

La fuente de datos de casi todos los depósitos de datos es el ambiente operacional.

Es una tentación pensar que existe una masiva redundancia de datos entre los dos ambientes. En efecto, la primera impresión que muchas personas tienen es que efectivamente existe una gran redundancia entre el ambiente operacional y el ambiente de depósitos de datos. Entender eso así es ser superficial y demuestra una falta de entendimiento de lo que está ocurriendo en el depósito de datos.



De hecho, existe un mínimo de redundancia de datos entre el ambiente operacional y el ambiente de depósito de datos, ya que:

- ✓ Los datos son filtrados conforme se pasan del ambiente operacional al ambiente de depósito de datos. Por lo tanto, nunca salen muchos datos del ambiente operacional, sólo aquellos datos que se requieren en el procesamiento de los sistemas de soporte a la toma de decisiones encuentran cabida en el ambiente de depósito de datos. Existen muchos datos que nunca ingresarán, ya que no conforman información necesaria o suficientemente relevante para la toma de decisiones.
- ✓ El horizonte de tiempo de los datos es bastante diferente entre un ambiente y el otro. Los datos en el ambiente operacional son mucho más frescos. Los datos en el depósito de datos son mucho más viejos. Desde tan sólo la perspectiva de los horizontes de tiempo, existe un muy pequeño traslape entre los ambientes operacional y de depósito de datos.
- ✓ El depósito de datos contiene información resumida que se encuentra en el ambiente operacional.
- ✓ Los datos sufren una transformación fundamental conforme son pasados al depósito de datos. De hecho la mayoría de los datos son significativamente alterados al ser seleccionados y puestos dentro del depósito de datos. Dicho de otra forma, la mayoría de los datos son física y radicalmente alterados conformes se introducen al depósito. Desde el punto de vista de integración, esto no es lo mismo que le sucede a los datos que residen en un ambiente operacional.

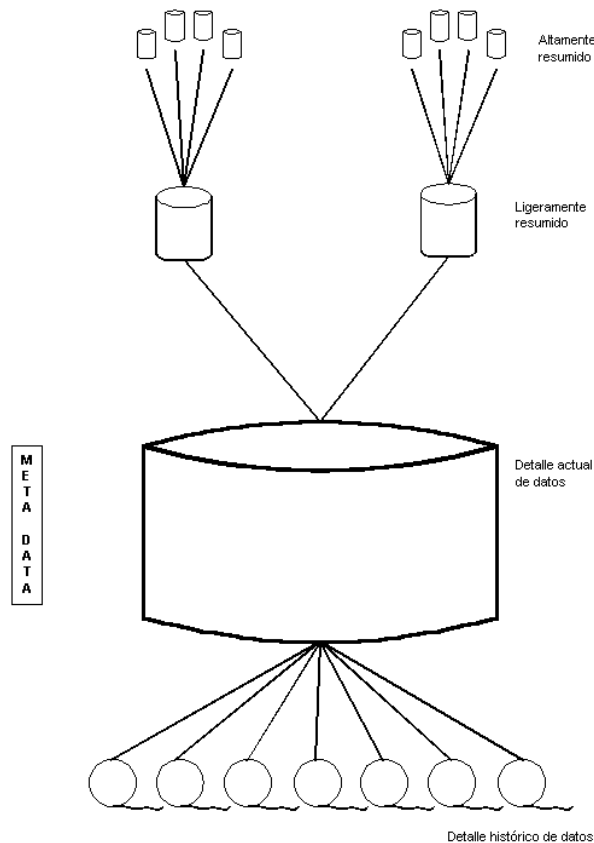
A la luz de estos factores, la redundancia entre los dos ambientes es una rara ocurrencia, resultando en menos de un 1% entre ambas.

3.4.4 Estructura de un Data Warehouse

Los Data Warehouse tienen una estructura distinta a las bases de datos operacionales. Hay niveles diferentes de esquematización y detalle que delimitan el Data Warehouse.

Los diferentes componentes del Data Warehouse son:

- Detalle de datos actuales.
- Detalle de datos antiguos.
- Datos ligeramente resumidos.
- Datos completamente resumidos.
- Meta data.

Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

Como se observa en la figura, los almacenes de datos están compuestos por diversos tipos de datos, que se organizan y dividen de acuerdo al nivel de detalle o granularidad que posean.

Detalle de datos actuales: En gran parte, el interés más importante radica en el detalle de los datos actuales, debido a que:

- ✓ Refleja las ocurrencias más recientes, las cuales son de gran interés
- ✓ Es voluminoso, ya que se almacena al más bajo nivel de granularidad.
- ✓ Casi siempre se almacena en disco, el cual es de fácil acceso, aunque su administración sea costosa y compleja.

Detalle de datos antiguos o históricos: la data antigua es aquella que se almacena sobre alguna forma de almacenamiento masivo. No es frecuentemente accedida y se almacena a un nivel de detalle, consistente con los datos detallados actuales. Mientras no sea prioritario el almacenamiento en un medio de almacenaje alterno, a causa del gran volumen de datos unido al acceso no frecuente de los mismos, es poco usual utilizar el disco como medio de almacenamiento.

Dependiendo del volumen de los datos, la frecuencia de acceso, el costo del medio y el tipo de acceso, es que se puede escoger el medio de almacenamiento que cumpla con las necesidades del viejo nivel de detalle en el depósito de datos.

Algunos de estos medios de almacenamiento son:



- almacenamiento de fotos ópticas
- RAID
- microfichas
- cintas magnéticas
- almacenamiento masivo

Datos ligeramente resumidos: la data ligeramente resumida es aquella que proviene desde un bajo nivel de detalle y sumanizan o agrupan los datos bajo algún criterio o condición de análisis, un claro ejemplo es el almacenamiento de la sumanización del detalle de las ventas realizadas en casa mes. Este nivel del Data Warehouse casi siempre se almacena en disco. Las preocupaciones que debe enfrentar el arquitecto de datos al construir este nivel del depósito de datos son:

- ✓ Qué unidad del tiempo vuelve a hacer el resumen, y
- ✓ Qué contenido (atributos) incluyen los “datos ligeramente resumidos”.

Datos completamente resumidos: estos datos son compactos y fácilmente accesibles. Pueden alojarse en el ambiente de Data Warehouse o fuera del límite de la tecnología que lo ampara (de todos modos, estos datos son parte del Data Warehouse sin considerar donde se alojan los físicamente).

Estos datos compactan aún más a los datos ligeramente resumidos y generalmente se guardan en disco y son muy fáciles de acceder.

Metadatos: es el componente final del Data Warehouse. Representan la información acerca de los datos. De muchas maneras la metadata se sitúa en una dimensión diferente al de otros datos del Data Warehouse debido a que su contenido no es tomado directamente desde el ambiente operacional.

La metadata juega un rol especial y muy importante, ya que es usada como:

- ✓ Un directorio para ayudar al analista a ubicar los contenidos del Data Warehouse.
- ✓ Una guía para el mapping de datos de cómo se transforma, del ambiente operacional al de Data Warehouse.
- ✓ Una guía de los algoritmos usados para la esquematización entre el detalle de datos actual, con los datos ligeramente resumidos y éstos, con los datos completamente resumidos, etc.

Los metadatos desempeñan un papel mucho más importante del que jamás se hizo en el sistema operacional clásico. Los metadatos son datos sobre los datos. Toda la información que se almacena en el Data Warehouse proviene de diferentes fuentes de datos por lo tanto se debe guardar el origen, la naturaleza y el significado de los datos. En los datos operacionales es opcional el uso de los metadatos, mientras que en el DW es obligatorio.

Se pueden distinguir tres diferentes tipos de Metadatos:

- Los metadatos de los procesos ETL, referidos a las diversas fuentes utilizadas, reglas de extracción, transformación, limpieza, depuración y carga de los datos al depósito.

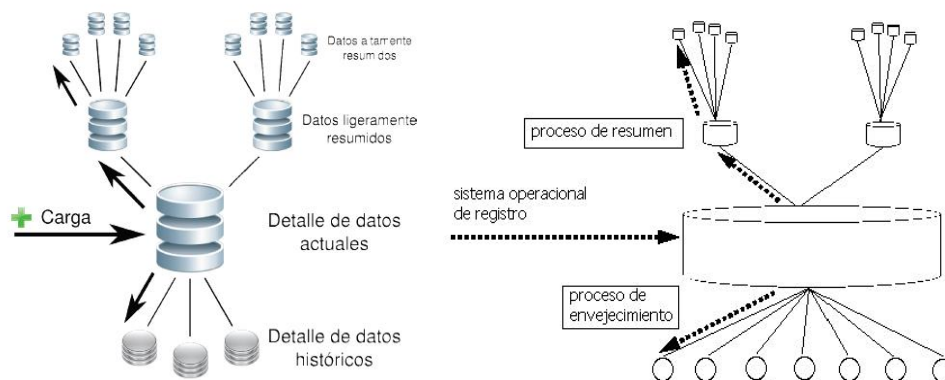
- Los metadatos operacionales, que son los que básicamente almacenan todos los contenidos del DW, para que este pueda desempeñar sus tareas.
- Los metadatos de consulta, que contienen las reglas para analizar y explotar la información del almacén, tales como drill-up y drill-down. Son estos metadatos los que las herramientas de análisis y consulta emplearán para realizar documentaciones y para navegar por los datos.

3.4.5 Flujo de datos en un Data Warehouse

Existe un flujo normal y predecible de los datos dentro del depósito de datos. El flujo de datos en el DW es estándar y generalizado, cuando la información ingresa al depósito de datos se almacena a nivel de detalle de los datos actuales. Los datos permanecerán allí hasta que ocurra alguno de los tres eventos siguientes:

- ✓ Sean borrados del depósito de datos.
- ✓ Sean resumidos, ya sea a nivel de datos ligeramente resumidos o a nivel de datos altamente resumidos.
- ✓ Sean archivados a nivel de detalle de datos históricos

Gráficamente es posible representar al flujo de datos dentro del DW, de la siguiente manera:



3.4.6 Ventajas y Desventajas del Data Warehouse

Las ventajas que se encuentran en un DW son:

- ✓ Proporciona un gran poder de procesamiento de información.
- ✓ Permite una mayor flexibilidad y rapidez en el acceso a la información.
- ✓ Facilita la toma de decisiones en los negocios.
- ✓ Las empresas obtienen un aumento de la productividad.
- ✓ Proporciona una comunicación fiable entre todos los departamentos de la empresa.
- ✓ Mejora las relaciones con los proveedores y los clientes.
- ✓ Permite conocer qué está pasando en el negocio, es decir, estar siempre enterado de los buenos y malos resultados.
- ✓ Transforma los datos en información y la información en conocimiento.
- ✓ Permite hacer planes de forma más efectiva.
- ✓ Reduce los tiempos de respuesta y los costes de operación.



- ✓ Transforma datos orientados a las aplicaciones en información orientada a la toma de decisiones.
- ✓ Integra y consolida diferentes fuentes de datos (internas y/o externas) y departamentos empresariales, que anteriormente formaban islas, en una única plataforma sólida y centralizada.
- ✓ Provee la capacidad de analizar y explotar las diferentes áreas de trabajo y de realizar un análisis inmediato de las mismas.
- ✓ Permite reaccionar rápidamente a los cambios del mercado.
- ✓ Aumenta la competitividad en el mercado.
- ✓ Elimina la producción y el procesamiento de datos que no son utilizados ni necesarios, producto de aplicaciones mal diseñadas o ya no utilizadas.
- ✓ Mejora la entrega de información, es decir, información completa, correcta, consistente, oportuna y accesible. Información que los usuarios necesitan, en el momento adecuado y en el formato apropiado.
- ✓ Logra un impacto positivo sobre los procesos de toma de decisiones. Cuando los usuarios tienen acceso a una mejor calidad de información, la empresa puede lograr, por sí misma, diferentes aspectos tales como: aprovechar el enorme valor potencial de sus recursos de información y transformarlo en valor verdadero; eliminar los retardos de los procesos que resultan de información incorrecta, inconsistente y/o inexistente; integrar y optimizar procesos a través del uso compartido e integrado de las fuentes de información; permitir a los usuarios adquirir mayor confianza acerca de sus propias decisiones y de las del resto, y lograr así, un mayor entendimiento de los impactos ocasionados.
- ✓ Aumento de la eficiencia de los encargados de tomar decisiones.
- ✓ Los usuarios pueden acceder directamente a la información en línea, lo que contribuye a su capacidad para operar con mayor efectividad en las tareas rutinarias o no. Además, pueden tener a su disposición una gran cantidad de valiosa información multidimensional, presentada coherentemente como fuente única, confiable y disponible en sus estaciones de trabajo. Así mismo, los usuarios tienen la facilidad de contar con herramientas que les son familiares para manipular y evaluar la información obtenida en el DW, tales como: hojas de cálculo, procesadores de texto, software de análisis de datos, software de análisis estadístico, reportes, tableros, etc.
- ✓ Permite la toma de decisiones estratégicas y tácticas.

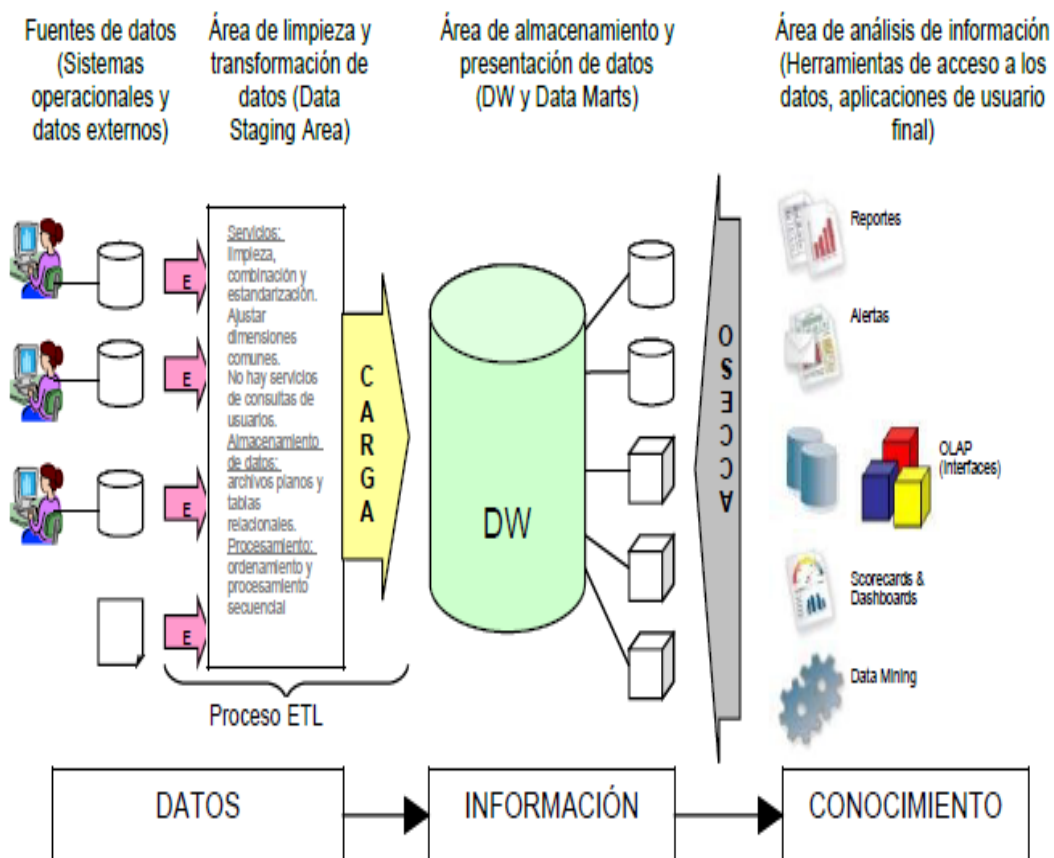
Desventajas:

- ✓ Requieren una revisión del modelo de datos, objetos, transacciones y además del almacenamiento.
- ✓ Requiere de mantenimiento, lo que implica continua limpieza, transformación e integración de datos.
- ✓ En un proceso de implantación puede encontrarse dificultades ante los diferentes objetivos que pretende una organización.
- ✓ Una vez implementado puede ser complicado añadir nuevas fuentes de datos.
- ✓ Requiere una gran inversión, debido a que su correcta construcción no es tarea sencilla y consume muchos recursos, además, su misma implementación implica desde la adquisición de herramientas de consulta y análisis, hasta la capacitación de los usuarios.
- ✓ Existe resistencia al cambio por parte de los usuarios.

- ✓ Los beneficios del almacén de datos son apreciados en el mediano y largo plazo. Este punto deriva del anterior, y básicamente se refiere a que no todos los usuarios confiarán en el DW en una primera instancia, pero sí lo harán una vez que comprueben su efectividad y ventajas. Además, su correcta utilización surge de la propia experiencia.
- ✓ Si se incluyen datos propios y confidenciales de clientes, proveedores, etc., el depósito de datos atentará contra la privacidad de los mismos, ya que cualquier usuario podrá tener acceso a ellos.
- ✓ Infravaloración de los recursos necesarios para la captura, carga y almacenamiento de los datos.
- ✓ Infravaloración del esfuerzo necesario para su diseño y creación.
- ✓ Incremento continuo de los requerimientos de los usuarios.
- ✓ Subestimación de las capacidades que puede brindar la correcta utilización del Data Warehousing y de las herramientas de BI en general.

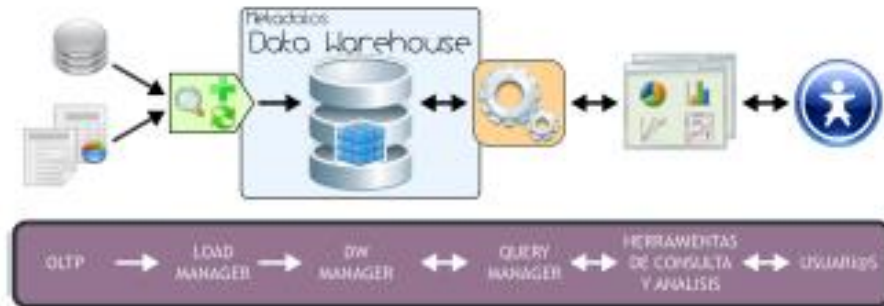
3.5 Arquitectura del Data Warehouse

Una Arquitectura Data Warehouse (DWA) es una forma de representar la estructura total de los datos, la comunicación, el procesamiento y la presentación de los mismos, ante los usuarios finales que disponen de una computadora dentro de la empresa.



Como se observa en el gráfico, el DW es el núcleo de toda la arquitectura de un Sistema de Soporte de Decisión (DSS).

A continuación se expone otro gráfico, muy similar al anterior, el cual permite definir y describir los componentes que intervienen en la arquitectura o ambiente del Data Warehousing.



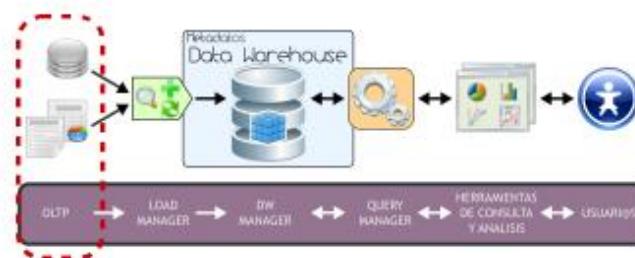
El ambiente está formado por diversos elementos que interactúan entre sí y que cumplen una función específica dentro del sistema. Por ello es que al abordar la exposición de cada elemento se lo hará en forma ordenada y teniendo en cuenta su relación con las demás partes.

Básicamente, la forma de operar del esquema superior se resume de la siguiente manera:

- 1) Los datos son extraídos desde aplicaciones, bases de datos, archivos, etc. esta información generalmente reside en diferentes tipos de sistemas, orígenes y arquitecturas y tienen formatos muy variados.
- 2) Los datos son integrados, transformados y limpiados, para luego ser cargados en el DW.
- 3) Principalmente, la información del DW se estructura en cubos multidimensionales, ya que estos preparan esta información para responder a consultas dinámicas con una buena performance. Pero también pueden utilizarse otros tipos de estructuras de datos para representar la información del DW, como por ejemplo Business Models.
- 4) Los usuarios acceden a los cubos multidimensionales, Business Models (u otro tipo de estructura de datos) del DW utilizando diversas herramientas de consulta, exploración, análisis, reportes, etc.

A continuación se detallará cada uno de los componentes de la arquitectura del Data Warehousing, teniendo como referencia siempre el gráfico antes expuesto, pero resaltando el tema que se tratará.

3.5.1 Sistema Transaccional (OLTP)



OLTP (On Line Transaction Process) esta conformado por todos aquellos medios que contienen información relativa a las actividades cotidianas de una empresa, es decir, aquellos que soportan las operaciones diarias de la organización. Representa toda aquella información transaccional que genera la empresa en su accionar diario, además, de las fuentes externas con las que puede llegar a disponer.

Los OLTP soportan la operatividad del negocio. Es por eso que manejan las transacciones en línea de cierto proceso en ejecución.

Los sistemas OLTP tienen como objetivo guardar la información necesaria para operar un negocio de la manera más eficiente, pero no han sido diseñados para proporcionar funciones potentes de síntesis, análisis y consolidación de los datos.

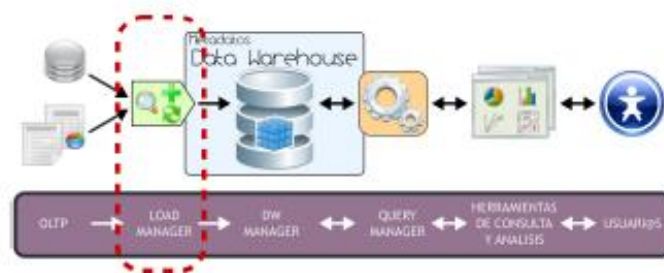
Actualmente, es posible encontrar distintos tipos de OLTP que son muy disimiles entre sí de acuerdo a sus características (formato, procedencia, función, etc.), algunos de los más habituales son:

- ✓ Las bases de datos transaccionales.
- ✓ Los archivos de texto plano.
- ✓ Archivos XML.
- ✓ Archivos de Word.
- ✓ Hipertextos
- ✓ Planillas de cálculos.
- ✓ Informes semanales, mensuales, anuales, etc.
- ✓ Imágenes.

Los sistemas operacionales son las fuentes de datos principales donde se registra toda la gestión de la organización. Son los sistemas centrales que soportan las operaciones diarias del negocio y a los cuales se accede a través de las interfaces de las aplicaciones (APIs). Los datos que recopilan estos sistemas son la principal fuente de información del DW. El éxito o fracaso de toda la solución de DW depende fuertemente de que los sistemas operacionales provean los datos necesarios para entender el negocio y la historia necesaria para evaluar su evolución. La calidad de los datos de estos sistemas es fundamental.

Otras fuentes pueden ser datos externos a la organización (información demográfica, crediticia, financiera, etc. por ejemplo la cotización de alguna moneda), planillas de cálculos, etc.

3.5.2 Load Manager



Para poder extraer los datos desde los OLTP, para luego manipularlos, integrarlos, transformarlos y finalmente cargar los resultados obtenidos en el DW es necesario

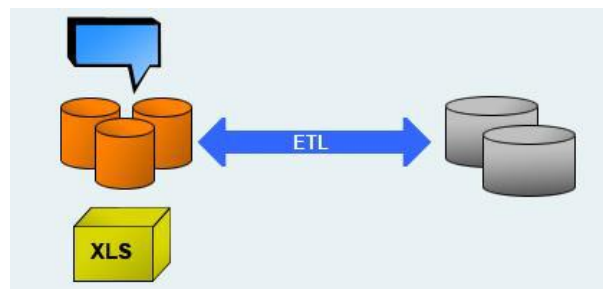
contar con algún sistema que se encargue de ello. Precisamente, la Integración de Datos tiene esta finalidad.

La Integración de Datos agrupa una serie de técnicas y subprocesos que se encargan de llevar a cabo todas las tareas relacionadas con la extracción, manipulación, control, integración y depuración de datos, así como de la carga y actualización del DW. Es decir, todas las tareas que se realizarán desde que se toman los datos de los diferentes OLTP hasta que se cargan en el DW.

Si bien los procesos ETL (Extracción, Transformación y Carga) son solo una de las muchas técnicas de la Integración de Datos, el resto de estas técnicas puede agruparse muy bien en sus diferentes etapas. Es decir, en el proceso de Extracción se tendrá un grupo de técnicas enfocadas en tomar solo los datos indicados y mantenerlos en un almacenamiento intermedio; en el proceso de Transformación estarán aquellas técnicas que analizarán los datos para verificar que sean correctos y válidos; en el proceso de Carga de Datos se agruparán técnicas propias de la carga y actualización del DW.

3.5.2.1 El proceso de Extracción, Transformación y Carga de datos (ETL)

El término popular ETL corresponde a la sigla en inglés de “Extract-Transform-Load” y significa extraer, transformar y cargar.



El proceso ETL implica las tareas de capturar, integrar, transformar, limpiar, reestructurar, validar, filtrar, analizar la calidad y cargar datos en el DW. Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes OLTP, reformatearlos y limpiarlos, para luego cargarlos al Data Warehouse, a fin de apoyar los procesos de negocio. Este proceso organiza el flujo de los datos entre los diferentes sistemas operacionales principales de una organización y el área de almacenamiento y presentación de datos. En ese momento los datos quedan disponibles para ser analizados por los usuarios. Otros nombres que recibe este proceso son “gestión de los datos”, “adquisición de datos” y en inglés “data staging” o “data cleansing”.

Cabe destacar que la integración y transformación de datos es uno de los procesos más importantes de todo el entorno del DW. Tiene la tarea crítica de convertir el caos de datos del mundo operacional en un mundo ordenado de información. Este proceso asimila datos procedentes de tecnologías heterogéneas dentro de un entorno integrado y consistente, apto para ser consumido por los procesos de soporte de decisiones.

El nivel de esfuerzo necesario para integrar y transformar datos está fundamentalmente afectado por el nivel de conocimiento que se tenga sobre estos. Cuanto más familiar resulten los datos operacionales y los procesos que los producen más fácil resultarán estas tareas. El proceso ETL en general tiende a ser subestimado, sin embargo es



altamente demandante y puede abarcar la mayor parte del tiempo de desarrollo de un DW, ocupando hasta el 80% del tiempo en un proyecto de gran magnitud.

Dicho proceso está compuesto por 3 etapas: Extracción, Transformación y Carga.

3.5.2.2 Extracción

La primera parte del proceso ETL consiste en extraer los datos relevantes de diversas fuentes OLTP, como pueden ser las bases de datos y/o archivos operacionales mediante herramientas de gestión de datos. Generalmente en este punto se fusionan datos provenientes de diferentes sistemas de origen.

Es aquí, en donde, basándose en las necesidades y requisitos de los usuarios, se exploran las diversas fuentes OLTP que se tengan a disposición, y se extrae la información que se considere relevante al caso. Si los datos operacionales residen en un SGBD Relacional, el proceso de extracción se puede reducir a, por ejemplo, consultas en SQL o rutinas programadas. En cambio, si se encuentran en un sistema no convencional o fuentes externas, ya sean textuales, hipertextuales, hojas de cálculos, etc., la obtención de los mismos puede ser un tanto más dificultoso, debido a que, por ejemplo, se tendrán que realizar cambios de formato y/o volcado de información a partir de alguna herramienta específica.

Un requerimiento importante que se debe exigir a la hora de realizar la extracción es que ésta cause un impacto mínimo en los sistemas de origen. Si los datos a extraer son muchos, los sistemas de origen se podrían colapsar y quedar fuera de funcionamiento. Por esta razón, en sistemas grandes las operaciones de extracción suelen programarse en horarios o días donde este impacto sea nulo o mínimo.

A su vez se recomienda almacenar los datos extraídos en repositorios intermedios (también conocidos como Staging Área) cuando existe heterogeneidad en las fuentes de datos y complejidad en los mismos.

El Staging Área es aquel sistema que permanece entre las fuentes de datos y el Data Warehouse con el objetivo de:

- ✓ Facilitar la extracción de datos desde las fuentes de origen de carácter múltiple realizando un pre-tratado.
- ✓ Manipular los datos sin interrumpir ni paralizar los OLTP, ni tampoco el DW.
- ✓ Realizar la limpieza de datos.
- ✓ Mejorar la calidad de datos.
- ✓ No depender de la disponibilidad de los OLTP.
- ✓ Almacenar y gestionar los metadatos que se generarán en los procesos ETL.
- ✓ Facilitar la integración de las diversas fuentes, internas y externas.

El almacenamiento intermedio constituye en la mayoría de los casos una base de datos en donde la información puede ser almacenada en tablas auxiliares, tablas temporales, etc. Los datos de estas tablas serán los que finalmente (luego de su correspondiente transformación) poblarán el DW.

3.5.2.3 Transformación



La segunda parte del proceso ETL y uno de los mayores desafíos de cualquier implementación de Data Warehouse, es transformar los datos.

Esta función es la encargada de convertir aquellos datos inconsistentes en un conjunto de datos compatibles y congruentes, para que puedan ser cargados en el DW. Estas acciones se llevan a cabo, debido a que pueden existir diferentes fuentes de información, y es vital conciliar un formato y forma única, definiendo estándares, para que todos los datos que ingresarán al DW estén integrados.

A partir de la transformación se logra consistencia e integración, tanto en lo que respecta al formato de los datos como a su contenido. Las bases de datos operacionales, por ejemplo, diseñadas para el soporte de varias aplicaciones transaccionales, frecuentemente difieren en el formato y hasta puede darse que dentro de una misma base de datos exista inconsistencia en cuanto a los datos almacenados.

La transformación consiste en aplicar una serie de reglas de negocio o funciones sobre los datos extraídos.

Algunas de las transformaciones más comunes son:

- ✓ Seleccionar sólo ciertas columnas para la carga.
- ✓ Codificar valores (por ejemplo, convertir “Femenino” en “F”).
- ✓ Traducir los valores codificados en la fuente (por ejemplo, si en el OLTP se almacena una “H” para “Hijo”, es posible guardar en el destino el valor “Hijo”).
- ✓ Adquirir nuevos valores calculados (por ejemplo, obtener el total de ventas de un producto).
- ✓ Generar campos clave en el destino.
- ✓ Estandarizar unidades de medidas.
- ✓ Definir convenciones de nombramientos.
- ✓ Unir datos de múltiples fuentes (por ejemplo, búsquedas, combinaciones, etc.).
- ✓ Transponer múltiples columnas en filas o viceversa.
- ✓ Dividir una columna en varias (por ejemplo, columna “Nombre: García, Miguel”; pasar a dos columnas “Nombre: Miguel” y “Apellido: García”), etc.

Dentro de la etapa de transformación se realiza la función de *Limpieza de Datos* (Data Cleansing), cuyo principal objetivo es el de realizar distintos tipos de acciones contra el mayor número de datos erróneos, inconsistentes e irrelevantes, a los cuales se los denomina Datos Anómalos (Outliers).

3.5.2.4 Carga

Esta fase corresponde al momento en el cual los datos son cargados en el Data Warehouse.

Los mismos pueden ser datos transformados que residen en repositorios temporales, o bien datos de OLTP que tienen una correspondencia directa.

En esta etapa se realizan las tareas relacionadas con:

- ✓ Carga Inicial (Initial Load).
- ✓ Actualización o mantenimiento periódico (siempre teniendo en cuenta un intervalo de tiempo predefinido para tal operación).

La carga inicial, se refiere precisamente a la primera carga de datos que se le realizará al DW. Por lo general, esta tarea consume un tiempo bastante considerable, ya que se deben insertar registros que han sido generados aproximadamente, y en casos ideales, durante más de cinco años.

Los mantenimientos periódicos mueven pequeños volúmenes de datos, y su frecuencia está dada en función de la granularidad del DW y los requerimientos de los usuarios. El objetivo de esta tarea es añadir al depósito aquellos datos nuevos que se fueron generando desde el último refresco.

Antes de realizar una nueva actualización, es necesario identificar si se han producido cambios en las fuentes originales de los datos recogidos, desde la fecha del último mantenimiento, a fin de no atentar contra la consistencia del DW.

Si este control consume demasiado tiempo y esfuerzo, o simplemente no puede llevarse a cabo por algún motivo en particular, existe la posibilidad de cargar el DW desde cero, este proceso se denomina Carga Total (Full Load).

Se debe tener en cuenta, que los datos antes de moverse al almacén de datos, deben ser analizados con el propósito de asegurar su calidad, ya que este es un factor clave, que no debe dejarse de lado. Por otra parte, el proceso de Carga tiene la tarea de mantener la estructura del DW.

Existen dos formas básicas de desarrollar el proceso de carga:

- ✓ *Acumulación simple*: consiste en realizar un resumen de todas las transacciones comprendidas en el período de tiempo seleccionado y transportar el resultado como una única transacción hacia el Data Warehouse, almacenando un valor calculado que consistirá típicamente en una sumatoria o promedio de una magnitud.
- ✓ *Rolling*: se aplica en los casos en que se opta por mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna de las dimensiones de la magnitud almacenada (por ejemplo, totales diarios, totales semanales, totales mensuales, etc.).

En síntesis los pasos y actividades que se realizan en el proceso ETL son:

- Se extraen los datos relevantes desde los OLTP y se depositan en un almacenamiento intermedio.
- Se integran y transforman los datos, para evitar inconsistencias.
- Se cargan los datos desde el almacenamiento intermedio hasta el DW.



El costo de todas las tareas descriptas depende en gran medida de la calidad de los datos en los sistemas fuentes y apuntan a garantizar la calidad de los datos en el DW. La calidad debe ser una línea conductora en todo el proceso de análisis de información y toma de decisiones.

3.5.2.5 Procesamiento Paralelo

Para efectuar la extracción, transformación y carga es posible utilizar lo que se conoce como procesamiento paralelo. Esto permite mejorar la performance general de los procesos ETL cuando se trabaja con grandes volúmenes de datos. El paralelismo puede ser:

- **De datos:** Consiste en dividir un único archivo secuencial en pequeños archivos de datos para facilitar su acceso.
- **De segmentación (pipeline):** Este tipo de paralelismo permite que varios componentes funcionen simultáneamente, actuando sobre el mismo flujo de datos.
- **De componente:** Se da cuando múltiples procesos que actúan sobre diferentes flujos de datos, funcionan simultáneamente, en el mismo lugar de trabajo.

Estos tres tipos de paralelismo se pueden combinar en una misma operación ETL, es decir, no son excluyentes entre sí.

3.5.3 Data Warehouse Manager



El DW Manager presenta las siguientes características y funciones principales:

- Se constituye típicamente al combinar un SGBD con software y aplicaciones dedicadas.
- Almacena los datos de forma multidimensional, es decir, a través de tablas de hechos y tablas de dimensiones.
- Gestiona las diferentes estructuras de datos que se construyan o describan sobre el DW, como Cubos Multidimensionales, Business Models, etc.
- Gestiona y mantiene metadatos.

Además, el DW Manager se encarga de:

- Transformar e integrar los datos fuentes y del almacenamiento intermedio en un modelo adecuado para la toma de decisiones.
- Realizar todas las funciones de definición y manipulación del depósito de datos, para poder soportar todos los procesos de gestión del mismo.
- Ejecutar y definir las políticas de particionamiento. El objetivo de realizar esto, es conseguir una mayor eficiencia y performance en las consultas al no tener que



manejar todo el grueso de los datos. Esta política debe aplicarse sobre la tabla de hechos que es aquella en la que se almacena toda la información que será analizada.

- Realizar copias de resguardo incrementales o totales de los datos del DW.

3.5.3.1 Data Warehouse y su estructura multidimensional

A lo largo de este capítulo, se han ido exponiendo las diferencias de ambientes y objetivos entre los sistemas OLAP y OLTP. Estas diferencias hacen que las técnicas de diseño y el modelo del ciclo de vida clásico utilizados para desarrollar un sistema operacional no resulten apropiados para el entorno analítico. Es por ello que se han desarrollado técnicas y metodologías específicas para garantizar el éxito de un proyecto de Data Warehousing.

La tecnología Data Warehousing tiene una orientación analítica, ya que su objetivo es permitir el análisis de la información para la toma de decisiones, y esto impone un procesamiento y pensamiento distinto, el cual se sustenta por un modelado de bases de datos propio, conocido como Modelado Multidimensional, que ofrece al usuario la visión respecto a la operación del negocio.

Las bases de datos multidimensionales están orientadas a la realización de consultas complejas y ofrecen alto rendimiento. Dichas bases de datos pretenden superar las limitaciones de los sistemas de gestión de bases de datos relacionales en los procesos de análisis.

Las capacidades de las bases de datos relacionales para la sumarización y agregación (cláusulas *sum*, *avg*, *group by*, etc., de SQL) resultan muy limitadas si se quieren obtener resultados globales. La normalización dificulta una visión general de los datos, obligando a generar más y más tablas virtuales para proceder a un análisis medianamente complejo.

Una base de datos multidimensional dispone la información en base a dimensiones que determinan el entorno de trabajo de la organización, junto a tablas de hechos, cuyos atributos serán combinaciones de las dimensiones anteriores y registrarán las transacciones corrientes de la empresa. Esto permite obtener una visión conceptual de los datos, y hace implícita la posibilidad de sumarizarlos y efectuar distintos tipos de análisis.

3.5.3.2 Bases de datos multidimensionales

Una base de datos multidimensional es una base de datos en donde su información se almacena en forma multidimensional, es decir, a través de tablas de hechos y tablas de dimensiones.

Proveen una organización que permite, a través de la creación y consulta a una estructura de datos determinada (cubo multidimensional, Business Model, etc), tener acceso flexible a los datos, para explorar y analizar sus relaciones, y consiguientes resultados.

Estas bases de datos se basan en el modelado multidimensional, el cual puede ser de tipo:



- ✓ Esquema Estrella (Star Scheme).
- ✓ Esquema Copo de nieve (Snowflake Scheme).
- ✓ Esquema Constelación (Starflake Scheme).

Estos temas serán explicados en detalle posteriormente.

Los modelos antes mencionados suelen implementarse de tres maneras distintas y esto define la arquitectura de modelado, tomando la denominación de:

- ✓ ROLAP (Relacional): cuando se implementa de manera relacional,
- ✓ MOLAP (Multidimensional): cuando se implementa de manera multidimensional,
- ✓ HOLAP (Híbrido): es un híbrido entre las dos implementaciones anteriores.

Independiente de la arquitectura, este tipo de base de datos requiere que todo el modelo este desnormalizado o semi desnormalizado, a fin de no desarrollar complejas uniones para acceder a los datos y agilizar la complejidad de las consultas. A su vez, los cálculos que se efectúan sobre la misma son matriciales (debido a su naturaleza multidimensional), lo que permite un rápido procesamiento de la información.

3.5.3.3 *Modelado Multidimensional*

El modelado multidimensional es una técnica de diseño lógico que busca presentar la información en un marco estándar e intuitivo permitiendo un acceso de alto rendimiento. Permite visualizar a la base de datos como si fuera un cubo de tres, cuatro, cinco o más dimensiones. Cuando esto sucede, los usuarios finales pueden imaginarse “navegando” la información contenida, realizando cortes (slicing and dicing) a ese cubo a través de cada una de sus dimensiones. Los modelos multidimensionales brindan la habilidad de visualizar datos en una forma concreta y tangible que facilita la comprensión de estos. Existen algunos conceptos básicos para comprender la filosofía de este tipo de modelado, los cuales son presentados a continuación:

✓ *Áreas temáticas*

Identifican y agrupan los intereses de la organización. Por ejemplo, el departamento de Comercialización de una empresa puede estar interesado en las áreas temáticas de Pedidos, Promociones, Mercados y Ventas. Para especificar las áreas temáticas se deben identificar los hechos y las medidas.

✓ *Hechos*

Un hecho es un patrón de interés o evento dentro de la empresa que se desea analizar, por ejemplo, una venta es un hecho, y se utiliza para crear medidas o indicadores, a través de sumalizaciones o acumulaciones preestablecidas. Los hechos están implícitamente definidos por la combinación de los valores de las dimensiones. Son datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones.

Algunos autores también denominan a los hechos como datos agregados.

Los hechos son aquellos datos que residen en una tabla de hechos y que son utilizados para crear indicadores, a través de sumalizaciones preestablecidas al momento de crear un cubo multidimensional, Business Model, etc. Debido a que una tabla de hechos se



encuentra interrelacionada con sus respectivas tablas de dimensiones, permite que los hechos puedan ser accedidos, filtrados y explorados por los valores de los campos de estas tablas de dimensiones, obteniendo de este modo una gran capacidad analítica.

Las sumalizaciones no están referidas solo a sumas, sino también a promedios, mínimos, máximos, totales por sector, porcentajes, fórmulas predefinidas, etc., dependiendo de los requerimientos de información del negocio.

Existen dos tipos de hechos, los básicos y los derivados.

- ❖ Hechos básicos: son los que se encuentran representados por un campo de una tabla de hechos.
- ❖ Hechos derivados: son los que se forman al combinar uno o más hechos con alguna operación matemática o lógica y que también residen en una tabla de hechos. Estos poseen la ventaja de almacenarse previamente calculados, por lo cual pueden ser accedidos a través de consultas SQL sencillas y devolver resultados rápidamente, pero requieren más espacio físico en el DW, además de necesitar más tiempo de proceso en los ETL que los calculan.

Los hechos son gestionados con el principal objetivo de que se construyan indicadores basados en ellos, a través de la creación de un cubo multidimensional, Business Model, u otra estructura de datos.

✓ *Medidas o Indicadores del negocio*

Son cuantificadores del desempeño de un hecho. Dicha medida o indicador se obtiene realizando sumalizaciones o acumulaciones sobre un hecho perteneciente a una tabla de hechos.

Los indicadores serán incluidos en algún cubo multidimensional, con el fin de analizar los datos almacenados en el DW. El valor que estos adopten estará condicionado por los atributos/jerarquías que se utilicen para analizarlos. Los indicadores, además de hechos, pueden estar compuestos por otros indicadores, pero no ambos simultáneamente. Pueden utilizarse para su creación funciones de sumalización (suma, conteo, promedio, etc.), funciones matemáticas, estadísticas, operadores matemáticos y lógicos.

La información que brinda una medida es usada por los usuarios en sus consultas para evaluar el desempeño de un área temática.

Una medida contiene una propiedad numérica y una fórmula, existiendo tres clases de medidas:

- ❖ Medidas aditivas: Pueden ser combinadas a lo largo de cualquier dimensión.
- ❖ Medidas semi-aditivas: No pueden ser combinadas a lo largo de una o más dimensiones.
- ❖ Medidas no aditivas: No pueden ser combinadas a lo largo de ninguna dimensión.

✓ *Dimensiones*

El Data Warehouse organiza los datos mediante múltiples dimensiones. Una dimensión es una colección de entidades del mismo tipo, y constituye un calificador mediante el cual el usuario podrá filtrar y manipular la información almacenada, más precisamente, el usuario podrá analizar los hechos. Las dimensiones son los puntos de vistas a través de los cuales se desean analizar los hechos, estas califican a los hechos, los ponen en



perspectiva. Organizan la información de manera que se pueda preguntar qué, cuándo, donde, etc. Las dimensiones son almacenadas en tablas junto con los elementos y atributos de dimensión.

✓ *Elementos de Dimensión*

Son componentes conceptuales de una dimensión organizados jerárquicamente, es decir que el usuario puede navegar por los distintos niveles de la jerarquía de la información. Son conjuntos o grupos de atributos del negocio.

✓ *Atributos*

Una dimensión se identifica por poseer más de un atributo o miembro, donde un atributo es una característica de una dimensión, tal como color, altura, tamaño, etc. Por ejemplo, en la dimensión Tiempo los atributos son: Año, mes, día, etc. Los atributos de una dimensión son las descripciones e información relacionada a cada elemento de dimensión. Un elemento puede contener más de un atributo. Los atributos ayudan al usuario final para realizar consultas utilizando términos del negocio.

Los atributos constituyen los criterios de análisis que se utilizan para analizar los indicadores dentro de un cubo multidimensional. Los mismos se basan, en su gran mayoría, en los campos de las tablas de dimensiones y/o expresiones.

Dentro de un cubo multidimensional, los atributos son los ejes del mismo.

✓ *Jerarquías*

Una jerarquía representa una relación lógica entre dos o más atributos pertenecientes a un cubo multidimensional; siempre y cuando posean su correspondiente relación “padre-hijo”.

Entre los atributos de una misma dimensión pueden existir jerarquías, que se definen por la posición relativa que toma un atributo con respecto a los otros miembros de la misma dimensión, formando en su totalidad una estructura de árbol. Partiendo de la raíz del árbol, los atributos o miembros son progresivamente más detallados hasta llegar a las hojas, donde se obtiene el mayor nivel de detalle.

Por ejemplo: En la dimensión Tiempo, es normal tener la siguiente jerarquía: Año-Mes-Día. Puede darse también que una dimensión no necesite jerarquizarse debido a que ninguno de sus atributos posee una posición relativa con respecto a los otros.

Las jerarquías poseen las siguientes características:

- Pueden existir varias en un mismo cubo.
- Están compuestas por dos o más niveles.
- Se tiene una relación “1-n” o “padre-hijo” entre atributos consecutivos de un nivel superior y uno inferior.

Por lo general, las jerarquías pueden identificarse fácilmente, debido a que existen relaciones “1-n” o “padre-hijo” entre los propios atributos de un cubo.

La principal ventaja de manejar jerarquías, reside en poder analizar los datos desde su nivel más general al más detallado y viceversa, al desplazarse por los diferentes niveles.

✓ *Relación*



Representa la forma en que dos atributos interactúan dentro de una jerarquía. Las relaciones pueden ser explícitas o implícitas.

- Las relaciones explícitas son las más comunes y se pueden modelar a partir de atributos directos y están en línea continua de una jerarquía, por ejemplo, un país posee una o más provincias y una provincia pertenece a un solo país.
- Las relaciones implícitas son las que ocurren en la vida real pero no se pueden ver directamente, por ejemplo, una provincia tiene muchos ríos, pero un río a su vez pasa por muchas provincias.

Una relación define la cardinalidad entre atributos, y puede ser uno a muchos, muchos a muchos, etc.

✓ *Granularidad*

Representa el nivel de detalle al que se desea almacenar la información sobre el negocio que se esté analizando. Por ejemplo, los datos referentes a ventas o compras realizadas por una empresa, pueden registrarse día a día, en cambio, los datos pertinentes a pagos de sueldos o cuotas de socios, podrán almacenarse a nivel de mes.

Mientras mayor sea el nivel de detalle de los datos, se tendrán mayores posibilidades analíticas, ya que los mismos podrán ser resumidos o sumariados. Es decir, los datos que posean granularidad fina (nivel alto de detalle) podrán ser resumidos hasta obtener una granularidad media o gruesa. No sucede lo mismo en sentido contrario, ya que por ejemplo, los datos almacenados con granularidad media podrán resumirse, pero no tendrán la facultad de ser analizados a nivel de detalle. O sea, si la granularidad con que se guardan los registros es a nivel de día, estos datos podrán sumariarse por semana, mes, semestre y año, en cambio, si estos registros se almacenan a nivel de mes, podrán sumariarse por semestre y año, pero no lo podrán hacer por día y semana.

✓ *Tablas del Data Warehouse*

Como cualquier base de datos, un Data Warehouse se compone de tablas. Hay dos tipos:

Tablas de dimensiones: Las dimensiones califican y manejan las consultas que podrá efectuar el usuario. Estas tablas definen como los datos están organizados lógicamente y proveen el medio para analizar el contexto del negocio. Las tablas de dimensiones contienen datos cualitativos.

Estas tablas representan los aspectos de interés, mediante los cuales los usuarios podrán filtrar y manipular la información almacenada en la tabla de hechos. Las tablas de dimensiones están compuestas de elementos y atributos para cada nivel de jerarquía. El nivel más bajo de detalle que es requerido para el análisis determina el nivel más bajo de la jerarquía y el más alto es la base de la jerarquía.

Una tabla de dimensión puede contener los siguientes campos:

- Una clave principal.
- Clave foráneas, si existe.
- Datos de referencia primarios: datos que identifican o describen a la dimensión.
- Datos de referencia secundarios: datos que complementan la descripción de la dimensión.



Usualmente la cantidad de tablas de dimensiones, aplicadas a un tema de interés en particular, varían entre tres y quince.

En un Data Warehouse la creación y mantenimiento de una tabla de dimensión tiempo es obligatorio, y la definición de la granularidad y estructuración de la misma depende de la dinámica del negocio. Toda la información dentro de la bodega posee su propio sello de tiempo que determina la ocurrencia y ubicación de un hecho específico con respecto a otros elementos de iguales condiciones, representando de esta manera diferentes versiones de una misma situación.

Es importante tener en cuenta que la dimensión tiempo no es solo una secuencia cronológica representada de forma numérica, sino que mantiene niveles jerárquicos especiales que inciden notablemente en las actividades de la organización. Esto se debe a que los usuarios podrán por ejemplo analizar los hechos teniendo en cuenta el día de la semana en que se produjeron, quincena, mes, trimestre, semestre, año, estación, etc.

Las tablas de dimensión son un conjunto de tablas compañeras o guías para una tabla de hechos. Cada dimensión es definida por su clave primaria que sirve como base para la integridad referencial con cualquier tabla de hechos a la cual se une. La mayoría de las tablas de dimensión contienen muchos atributos o campos de texto que sirven para restringir y agrupar las consultas del DW y determinan los niveles de análisis.

Hay tablas de dimensión que son íntegramente descriptivas. Otras, llamadas tablas de catálogo o de búsqueda, además de contener las descripciones asociadas a la clave primaria, se utilizan para representar relaciones entre dimensiones relacionadas. De esta forma se evita la redundancia del dato en la tabla de hechos cuando el valor de una dimensión depende del valor de otra (a estos casos se los denomina subdimensiones). Es una forma primaria de normalizar algunos datos para evitar redundancia y reducir el tamaño de la tabla de hechos.

Tablas de Hechos: Contienen, precisamente, los hechos que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. Estas tablas contienen datos cuantitativos.

Cada tabla de hechos contiene un registro por cada combinación de claves de las dimensiones.

Los hechos definen el volumen del DW, y las tablas de hechos pueden estar compuestas por millones de registros, dependiendo de la granularidad de los datos y el intervalo de tiempo de los mismos.

Recordemos que los hechos son accedidos por las dimensiones, por lo que un hecho es un dato instantáneo en el tiempo, que son filtrados, agrupados y explorados bajo condiciones definidas por las dimensiones. Los registros de los hechos siempre son insertados, jamás actualizados y el ciclo de refresco depende de la resolución temporal de los datos.

Dichos registros de hechos poseen una clave primaria compuesta, que está definida por las llaves primarias de las tablas de dimensiones relacionadas a este.

Una tabla de hechos es una tabla primaria en cada modelo (o submodelo) dimensional que contiene medidas del negocio. Cada medida es interpretada como la intersección de todas las dimensiones que la definen. Toda tabla de hechos representa una relación muchos a muchos entre las dimensiones y contiene un conjunto de dos o más claves

foráneas que la vinculan a sus respectivas tablas de dimensión. Las tablas de hechos representan los sucesos ocurridos. No se debiera intentar llenarlas con ceros representando que nada ha sucedido. El nivel de detalle de cada dimensión que se utiliza en la tabla de hechos define la granularidad de las medidas. Las medidas más útiles son numéricas y aditivas, pero no todas son aditivas, existen medidas semi-aditivas y no aditivas. También es posible que la tabla de hechos no contenga medidas y sólo represente las relaciones entre dimensiones.

✓ *Cubo Multidimensional*

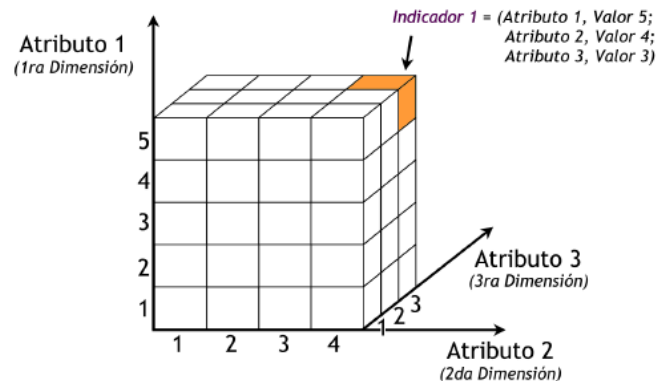
Una de las formas de representar como están organizados los datos en un modelo multidimensional es a través de un cubo. Un cubo multidimensional o hipercubo, representa o convierte los datos planos que se encuentran en filas y columnas, en un matriz de N dimensiones.

Los objetos más importantes que se pueden incluir en un cubo multidimensional son tres de los conceptos presentados en los puntos anteriores:

- ✓ Indicadores: sumalizaciones que se efectúan sobre algún hecho o expresiones basadas en sumalizaciones, pertenecientes a una tabla de hechos.
- ✓ Atributos: campos o criterios de análisis, pertenecientes a tablas de dimensiones.
- ✓ Jerarquías: representa una relación lógica entre dos o más atributos.

De esta manera en un cubo multidimensional, los atributos existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa el valor que tomará el indicador que se está evaluando.

En la siguiente representación matricial se puede ver más claramente lo que se acaba de mencionar.



Para la creación del cubo de la figura anterior, se definieron tres Atributos (“Atributo 1”, “Atributo 2” y “Atributo 3”) y se definió un Indicador (“Indicador 1”). Entonces el cubo quedó compuesto por 3 dimensiones o ejes (una por cada Atributo), cada una con sus respectivos valores asociados. También, se ha seleccionado una intersección al azar para demostrar la correspondencia con los valores de las Atributos. En este caso, el indicador “Indicador 1”, representa el cruce del Valor “5” de “Atributo 1”, con el Valor “4” de “Atributo 2” y con el Valor “3” de “Atributo 3”. Se puede observar, que el resultado del análisis está dado por los cruces matriciales de acuerdo a los valores de las dimensiones seleccionadas.



Más específicamente, para acceder a los datos del DW, se pueden ejecutar consultas sobre algún cubo multidimensional previamente definido. Dicho cubo debe incluir entre otros objetos: indicadores, atributos, jerarquías, etc., basados en los campos de las tablas de dimensiones y de hechos, que se deseen analizar. De esta manera, las consultas son respondidas con gran performance, minimizando al máximo el tiempo que se hubiese incurrido en realizar dicha consulta sobre una base de datos transaccional.

3.5.3.4 *Tipos de modelamiento de un Data Warehouse*

La colección de tablas en el DW se conoce como Esquema. Los esquemas caen dentro de tres categorías:

Esquema Estrella: Este esquema está compuesto por una tabla central, que es la de hechos, y un conjunto de tablas periféricas que son las de dimensiones y se relacionan con la tabla de hechos a través de las claves primarias.

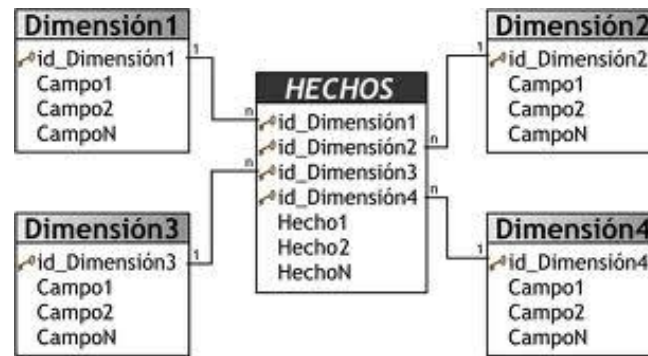
En el modelo estrella las dimensiones no se normalizan, es decir que el esquema está totalmente desnormalizado y por lo tanto no puede presentarse en tercera forma normal. Con ello se logra minimizar el número de uniones e incrementar el rendimiento de las consultas. Cuando se normaliza, se pretende eliminar la redundancia y repetición de datos, así como que las claves sean independientes de las columnas, pero en este tipo de modelos se requiere no evitar precisamente eso.

Como se acaba de decir, las ventajas que trae aparejada la desnormalización, son las de obviar uniones (Join) entre las tablas cuando se realizan consultas, procurando así un mejor tiempo de respuesta y una mayor sencillez con respecto a su utilización. El punto en contra, es que se genera un cierto grado de redundancia, pero el ahorro de espacio no es significativo. El esquema en estrella es el más simple de interpretar y optimiza los tiempos de respuesta ante las consultas de los usuarios. Este modelo es soportado por casi todas las herramientas de consulta y análisis, y los metadatos son fáciles de documentar y mantener, sin embargo es el menos robusto para la carga y es el más lento de construir.

Algunas características de este modelo son:

- ✓ Posee los mejores tiempos de respuesta.
- ✓ Su diseño es fácilmente modificable.
- ✓ Existe paralelismo entre su diseño y la forma en que los usuarios visualizan y manipulan los datos.
- ✓ Simplifica el análisis.
- ✓ Facilita la interacción con herramientas de consulta y análisis.

Este modelo genera que una tabla de hechos está relacionada con numerosas tablas de dimensiones y gráficamente se puede ver de la siguiente forma:

Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

Esquema Copo de Nieve: En este esquema existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, quienes a su vez pueden estar relacionadas o no con una o más tablas de dimensiones. Este modelo es más cercano a un modelo de entidad relación, que al modelo en estrella, debido a que sus tablas de dimensiones están normalizadas.

Una de los motivos principales de utilizar este tipo de modelo, es la posibilidad de segregar los datos de las tablas de dimensiones y proveer un esquema que sustente los requerimientos de diseño, pero hay que recordar que muchas uniones en las consultas puede sacrificar el desempeño del sistema, por lo que hay que definir cuidadosamente la cantidad de niveles del modelo. Otra razón es que es muy flexible y puede implementarse después de que se haya desarrollado un esquema en estrella.

Se pueden definir las siguientes características de este tipo de modelo:

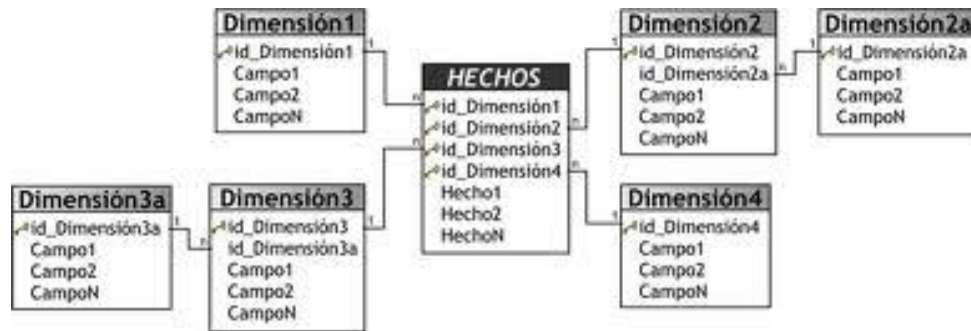
- ✓ Posee mayor complejidad en su estructura.
- ✓ Hace una mejor utilización del espacio.
- ✓ Es muy útil en tablas de dimensiones de muchas tuplas.
- ✓ Las tablas de dimensiones están normalizadas, por lo que requiere menos esfuerzo de diseño.
- ✓ Puede desarrollar clases de jerarquías fuera de las tablas de dimensiones, que permiten realizar análisis de lo general a lo detallado y viceversa.

A pesar de todas las características y ventajas que trae aparejada la implementación del esquema copo de nieve, existen dos grandes inconvenientes de ello:

- ✓ Si se poseen múltiples tablas de dimensiones, cada una de ellas con varias jerarquías, se creará un número de tablas bastante considerable, que pueden llegar al punto de ser inmanejables.
- ✓ Al existir muchas uniones y relaciones entre tablas, el desempeño puede verse reducido.

La existencia de las diferentes jerarquías de dimensiones debe estar bien fundamentada, ya que de otro modo las consultas demorarán más tiempo en devolver los resultados, debido a que se deben realizar las uniones entre las tablas.

Gráficamente se observaría como:

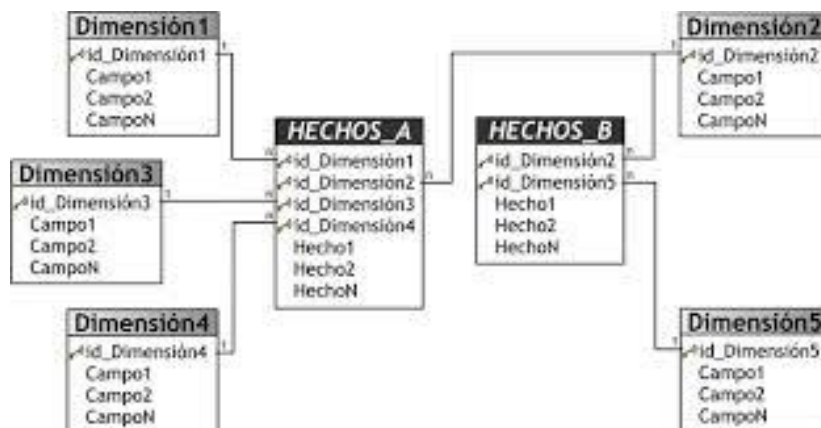
Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

Modelo Constelación: El modelo constelación está compuesto por una serie de modelos estrella, donde existe una tabla de hechos principal y una o más tablas de hechos auxiliares, las cuales pueden ser sumalizaciones de la principal.

Estas tablas de hechos auxiliares no necesariamente están relacionadas con las mismas dimensiones de la principal, por lo que puede ocurrir que estén relacionadas con nuevas dimensiones o con un subconjunto de las dimensiones que referencian a la tabla de hechos principal.

Su diseño y cualidades son muy similares a las del esquema en estrella, pero posee una serie de diferencias con el mismo, que son precisamente las que lo destacan y caracterizan. Entre ellas se pueden mencionar:

- ✓ Permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño.
- ✓ Contribuye a la reutilización de las tablas de dimensiones, ya que una misma tabla de dimensión puede utilizarse para varias tablas de hechos.
- ✓ No es soportado por todas las herramientas de consulta y análisis.



3.5.3.5 OLTP vs DW

Los OLTP son diseñados para soportar el procesamiento de información diaria de las empresas, y el énfasis recae en maximizar la capacidad transaccional de sus datos. Su estructura es altamente normalizada, para brindar mayor eficiencia a sistemas con muchas transacciones que acceden a un pequeño número de registros y están fuertemente condicionadas por los procesos operacionales que deben soportar, para la óptima actualización de sus datos. Esta estructura es ideal para llevar a cabo el proceso transaccional diario, brindar consultas sobre los datos cargados y tomar decisiones



diarias, en cambio los esquemas de DW están diseñados para poder llevar a cabo procesos de consulta y análisis para luego tomar decisiones estratégicas y tácticas de alto nivel.

A continuación se presenta una tabla comparativa entre los dos ambientes, que resume sus principales diferencias:

	OLTP	Data Warehouse
Objetivo	Soportar actividades transaccionales diarias.	Consultar y analizar información estratégica y táctica.
Tipo de datos	Operacionales.	Para la toma de decisiones.
Modelo de datos	Normalizado.	Desnormalizado.
Consulta	SQL.	SQL más extensiones.
Datos consultados	Actuales.	Actuales e históricos.
Horizonte de tiempo	60 - 90 días.	5 - 10 años.
Tipos de consultas	Repetitivas, predefinidas	No previsibles, dinámicas
Nivel de almacenamiento	Nivel de detalle.	Nivel de detalle y diferentes niveles de sumalización.
Acciones disponibles	Alta, baja, modificación y consulta.	Carga y consulta.
Número de transacciones	Elevado	Medio o bajo
Tamaño	Pequeño - Mediano.	Grande.
Tiempo de respuesta	Pequeño (segundos - minutos).	Variable (minutos - horas).
Orientación	Orientado a las aplicaciones.	Orientado al negocio.
Sello de tiempo	La clave puede o no tener un elemento de tiempo.	La clave tiene un elemento de tiempo.
Estructura	Generalmente estable.	Generalmente varía de acuerdo a su propia evolución y utilización.

3.5.3.6 Tipos de implementación de un Data Warehouse.

En sus orígenes el término OLAP (On line Analytic Processing) estuvo asociado exclusivamente a bases de datos propietarias multidimensionales, también denominadas cubos, que fue la arquitectura predominante para DW durante la década del 80. Con la evolución de las bases de datos en cuanto a funcionalidad y rendimiento a principio de los 90's comienzan a desarrollarse DW en motores de base de datos relacionales, dando origen a los conceptos OLAP multidimensional (MOLAP), OLAP relacional (ROLAP) y OLAP Híbrido (HOLAP).

Los sistemas OLAP se refieren a la actividad general de consulta y presentación de datos numéricos y de texto desde el DW, así como también el estilo dimensional de la consulta y presentación de la información. El procesamiento analítico es el uso de los datos para tomar decisiones del negocio, lo cual frecuentemente involucra análisis de tendencias, comparaciones de períodos, y navegación de datos.

El modelo lógico de los datos para OLAP debe seguir el diseño multidimensional, sin embargo la tecnología de la base de datos donde se implementa puede ser multidimensional o relacional, dando origen a tres categorías de productos OLAP: MOLAP, ROLAP y HOLAP.

Una vez modelado el DW, ya sea mediante un esquema estrella, copo de nieve o constelación, es posible implementarlo de tres maneras distintas:



- ✓ **ROLAP** (Relational On Line Analytic Processing).
- ✓ **MOLAP** (Multidimensional On Line Analytic Processing).
- ✓ **HOLAP** (Hybrid On Line Analytic Processing).

ROLAP es una arquitectura multidimensional que se caracteriza por almacenar los datos en una base de datos relacional. En consecuencia, las operaciones multidimensionales se traducen en operaciones relacionales estándar por lo que, cuando los usuarios efectúan consultas multidimensionales, éstas se transforman en consultas SQL ejecutadas sobre una base de datos relacional, devolviendo el resultado a los usuarios finales. Si bien se implemente sobre tecnología relacional dispone de algunas facilidades para mejorar el rendimiento. ROLAP cuenta con todos los beneficios de una SGBD Relacional a los cuales se les provee extensiones y herramientas para poder utilizarlo como un Sistema Gestor de DW.

En los sistemas ROLAP, los cubos multidimensionales se generan dinámicamente al instante de realizar las diferentes consultas, haciendo de esta manera el manejo de cubos transparente a los usuarios. Este proceso se puede resumir a través de los siguientes pasos:

1. Se seleccionan los indicadores, atributos, jerarquías, etc., que compondrán el cubo multidimensional.
2. Se ejecutan las consultas sobre los atributos, indicadores, etc., seleccionados en el paso anterior. Entonces, de manera transparente a los usuarios se crea y calcula dinámicamente el cubo correspondiente, el cual dará respuesta a las consultas que se ejecuten.

Al no tener que intervenir los usuarios en la creación y el mantenimiento explícito de los cubos, ROLAP brinda mucha flexibilidad, ya que dichos cubos son generados dinámicamente al momento de ejecutar las consultas; posibilitando de esta manera la obtención de consultas ad hoc.

La principal desventaja de los sistemas ROLAP, es que los datos de los cubos generalmente se deben calcular cada vez que se ejecuta una consulta sobre ellos (excepto que se guarde alguna query pre calculada). Esto provoca que ROLAP no sea muy eficiente en cuanto a la rapidez de respuesta ante las consultas de los usuarios.

Es importante tener en cuenta que si los datos del DW son almacenados y gestionados a través de un SGBD Relacional, no se requiere de otro software que administre y gestione los datos de manera Multidimensional.

MOLAP, aquí los datos se encuentran almacenados en archivos que poseen una estructura multidimensional, los cuales reservan el espacio necesario para almacenar los resultados pre-calculados de todas las posibles consultas a la base de datos. El objetivo de los sistemas MOLAP es almacenar físicamente los datos en estructuras multidimensionales de manera que la representación externa y la interna coincidan. Para ello, se dispone de estructuras de almacenamiento específicas (Arrays) y técnicas de compactación de datos que favorecen el rendimiento del DW.

La información procedente de los sistemas transaccionales se carga en el sistema MOLAP. Una vez cargados los datos en la base de datos multidimensional, se realiza una serie de cálculos para obtener datos agregados a través de las dimensiones del negocio, poblando la estructura de dicha base de datos.



Luego de llenar esta estructura, se generan índices y se emplean algoritmos de tablas hash para mejorar los tiempos de accesos de las consultas. Una vez que el proceso de poblado ha finalizado, la base de datos está lista para su uso. Los usuarios solicitan informes a través de la interfaz y la lógica de aplicación de la base de datos multidimensional obtiene los datos.

MOLAP requiere que en una instancia previa se generen y calculen los cubos multidimensionales, para que luego puedan ser consultados. Este proceso se puede resumir a través de los siguientes pasos:

1. Se seleccionan los indicadores, atributos, jerarquías, etc., que compondrán el cubo multidimensional.
2. Se precálculan los datos del cubo.
3. Se ejecutan las consultas sobre los datos precálculados del cubo.

El principal motivo de precálculan los datos de los cubos, es que posibilita que las consultas sean respondidas con mucha rapidez, ya que los mismos no deben ser calculados en tiempo de ejecución, obteniendo de esta manera una muy buena performance.

Existen una serie de desventajas que están directamente relacionadas con el hecho de precálculan los datos de los cubos multidimensionales, ellas son:

- ✓ Cada vez que se requiere o es necesario realizar cambios sobre algún cubo, se debe tener que recalcularlo totalmente, para que se reflejen las modificaciones llevadas a cabo. Provocando de esta manera una disminución importante en cuanto a flexibilidad.
- ✓ Se precisa más espacio físico para almacenar dichos datos.

Ésta arquitectura usa Bases de Datos Multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos en múltiples dimensiones; al contrario de la arquitectura ROLAP que cree que las capacidades OLAP están perfectamente implantadas sobre bases de datos relacionales.

MOLAP utiliza una arquitectura de dos niveles: La bases de datos multidimensionales y el motor analítico.

- La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.
- El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través del cual los usuarios finales visualizan los análisis OLAP. Una arquitectura cliente/servidor permite a varios usuarios acceder a la misma base de datos multidimensional.

Una ventaja que presenta MOLAP es que consigue consultas muy rápidas a costa de mayores necesidades de almacenamiento, retardos en las modificaciones (que deberían ocurrir sólo ocasionalmente) y largos procesos de carga y cálculo de acumulados.

En ROLAP, al no haber redundancia, el archivo de base de datos es pequeño, los procesos de carga son rápidos pero, sin embargo, las consultas pueden ser muy lentas, por lo que se suele aplicar la solución de tener, al menos, algunas consultas precalculadas.

Comparación MOLAP – ROLAP



Al comparar las dos arquitecturas, se pueden realizar las siguientes observaciones:

- ✓ ROLAP es una arquitectura flexible y general, que crece para dar soporte a amplios requerimientos OLAP. MOLAP es una solución particular, adecuada para soluciones departamentales con volúmenes de información y número de dimensiones más modestos.
- ✓ ROLAP delega la negociación entre tiempo de respuesta y el proceso por lotes al diseño del sistema. Mientras, el MOLAP, suele requerir que sus bases de datos se precompilen para conseguir un rendimiento aceptable en las consultas, incrementando, por tanto los requerimientos por lotes.
- ✓ Los sistemas con alta volatilidad de los datos (aquellos en los que cambian las reglas de agregación y consolidación), requieren una arquitectura que pueda realizar esta consolidación ad-hoc. Los sistemas ROLAP soportan bien esta consolidación dinámica, mientras que los MOLAP están más orientados hacia consolidaciones por lotes.
- ✓ Los ROLAP pueden crecer hasta un gran número de dimensiones, mientras que los MOLAP generalmente son adecuados para diez o menos dimensiones.
- ✓ Los ROLAP soportan análisis OLAP contra grandes volúmenes de datos elementales.

Resumidamente es posible decir que:

ROLAP	MOLAP
Brinda mucha flexibilidad, ya que los cubos son generados dinámicamente al momento de ejecutar las consultas.	Cada vez que se requiere o es necesario realizar cambios sobre algún cubo, se debe tener que recalcularlo totalmente, para que se reflejen las modificaciones llevadas a cabo. Provocando de esta manera una disminución importante en cuanto a flexibilidad.
Los datos de los cubos se deben calcular cada vez que se ejecuta una consulta sobre ellos. Esto provoca que ROLAP no sea muy eficiente en cuanto a la rapidez de respuesta ante las consultas de los usuarios.	Las consultas son respondidas con mucha rapidez, ya que los mismos no deben ser calculados en tiempo de ejecución, obteniendo de esta manera una muy buena performance.

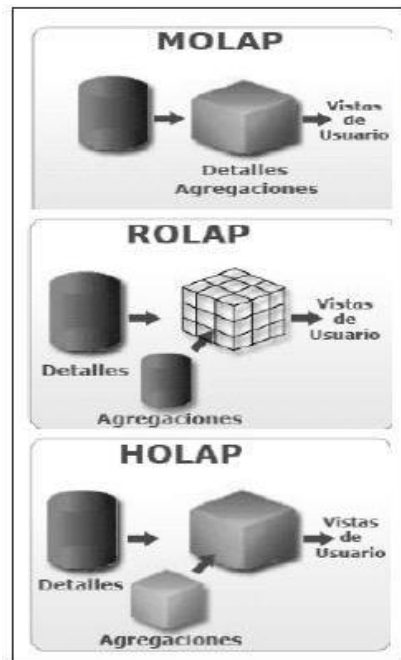
Al momento de optar por una de estas tecnologías deberán evaluarse determinados aspectos tales como: la performance que se necesita, el volumen de datos que se maneja, la escalabilidad y como se adapta la arquitectura propuesta por cada proveedor a la realidad de la organización.

La tendencia del mercado de estas herramientas es hacia los sistemas HOLAP. Tanto los proveedores clásicos de MOLAP como los ROLAP están incluyendo el otro estilo dentro de su solución para abarcar las ventajas de ambos (gran capacidad de análisis y alta performance).

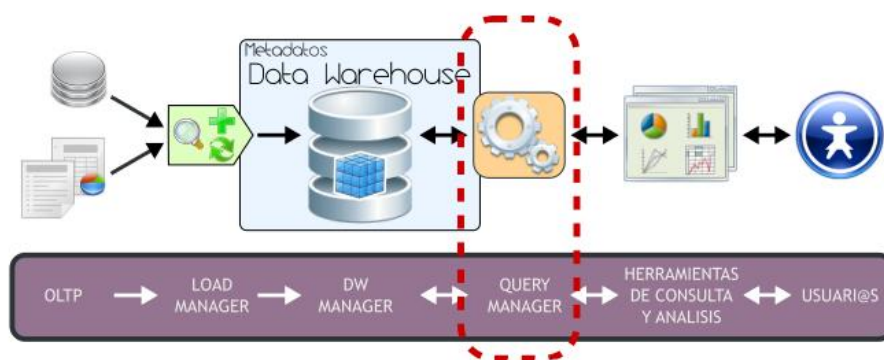
HOLAP, es un sistema híbrido que combina el uso de las arquitecturas ROLAP y MOLAP. HOLAP combina estas dos implementaciones para almacenar algunos datos en un motor relacional y otros en una base de datos multidimensional.

En una solución HOLAP, los volúmenes más grandes de datos, es decir los registros detallados, se mantienen en una base de datos relacional, mientras que los datos agregados y pre-calculados se guardan en un almacén MOLAP independiente. Es decir que se utilizará ROLAP para navegar y explorar los datos y se empleará MOLAP para la realización de tableros.

Una desventaja que presenta HOLAP, es que es necesario realizar un análisis profundo de los distintos tipos de datos, a fin de definir donde se almacenarán.



3.5.4 Query Manager



Como se ha indicado anteriormente, los datos en un Data Warehouse se modelan en estructuras multidimensionales, que pueden representarse como cubos (o hipercubos, en caso de que existan más de tres dimensiones). El Query Manager realiza las operaciones necesarias para soportar los procesos de gestión y ejecución de consultas relacionales tales como Join y agregaciones, y de consultas propias del análisis de datos, como drill-up y drill-down.



Query Manager recibe las consultas de los usuarios, las aplica a la estructura de datos correspondiente (cubo multidimensional, Business Models, etc.) y devuelve los resultados obtenidos.

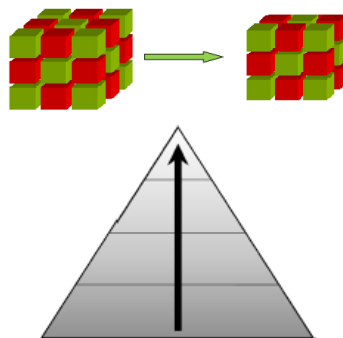
Cabe aclarar que una consulta a un DW generalmente consiste en la obtención de indicadores a partir de los datos (hechos) de una tabla de hechos, restringidas por las propiedades o condiciones de los atributos que hayan sido creados. Las operaciones que se pueden realizar sobre modelos multidimensionales y que son las que verdaderamente les permitirán a los usuarios explorar e investigar los datos en busca de respuestas, son:

- ✓ Drill-down.
- ✓ Drill-up.
- ✓ Drill-across.
- ✓ Roll-across.
- ✓ Pivot.
- ✓ Page.
- ✓ Drill-through.

Dichas operaciones permiten visualizar y analizar los datos de un modelo multidimensional.

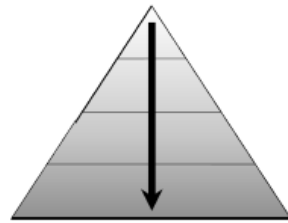
Drill Up: Esta técnica permite que el usuario pueda navegar entre las jerarquías de una dimensión agrupando los datos, recorriéndolos desde su mayor nivel de detalle hacia lo más general. Permite apreciar los datos en menor nivel de detalle, subiendo por una jerarquía definida en un cubo. Esto brinda la posibilidad de quitar un nivel o criterio de agregación en el análisis, agregando los grupos actuales.

Drill-up es ir de lo específico a lo general. Gráficamente:



Drill Down: A contraposición de la operación anterior, Drill-Down permite recorrer las jerarquías de una dimensión desagrupando los datos, pasando de lo general a lo particular. Permite apreciar los datos en un mayor detalle, bajando por una jerarquía definida en un cubo. Esto brinda la posibilidad de introducir un nuevo nivel o criterio de agregación en el análisis, disgregando los grupos actuales. Drill-down es ir de lo general a lo específico. Gráficamente:





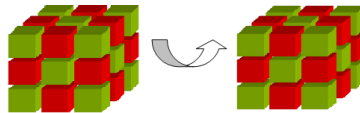
Drill-across: Funciona de forma similar a drill-down, con la diferencia de que drill-across no se realiza sobre una jerarquía, sino que su forma de ir de lo general a lo específico es agregar un atributo a la consulta como nuevo criterio de análisis.

Roll-across: Funciona de forma similar a drill-up, con la diferencia de que roll-across no se hace sobre una jerarquía, sino que su forma de ir de lo específico a lo general es quitar un atributo de la consulta, eliminando de esta manera un criterio de análisis.

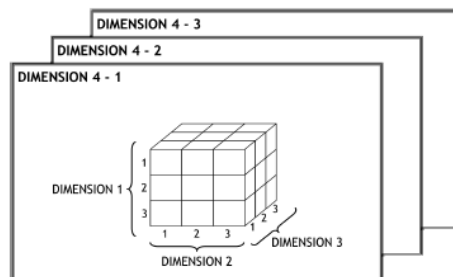
Pivot: Mediante esta técnica el usuario pivotea los ejes del cubo, modificando el orden de visualización de las dimensiones. Permite seleccionar el orden de visualización de los atributos e indicadores, con el objetivo de analizar la información desde diferentes perspectivas. Es decir, Píivot permite rotar o girar el cubo según sus dimensiones.

Pivot permite realizar las siguientes acciones:

- ✓ Mover un atributo o indicador desde el encabezado de fila al encabezado de columna.
- ✓ Mover un atributo o indicador desde el encabezado de columna al encabezado de fila.
- ✓ Cambiar el orden de los atributos o indicadores del encabezado de columna.
- ✓ Cambiar el orden de los atributos o indicadores del encabezado de fila.



Page: Presenta el cubo dividido en secciones, a través de los valores de un atributo, como si se tratase de páginas de un libro. Page es muy útil cuando las consultas devuelven muchos registros y es necesario desplazarse por los datos para poder verlos en su totalidad. Gráficamente:

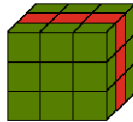


Cuando existe más de un criterio por el cual realizar Page, debe tenerse en cuenta el orden en que estos serán procesados, ya que dependiendo de esto, se podrán obtener diferentes resultados sobre una misma consulta. Es decir, el primer criterio utilizado para realizar Page condiciona los valores disponibles en el segundo, y así sucesivamente.

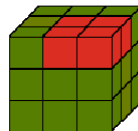


Drill-through: Permite apreciar los datos en su máximo nivel de detalle. Esto brinda la posibilidad de analizar cuáles son los datos relacionados al valor de un Indicador, que se ha sumariado dentro del cubo multidimensional.

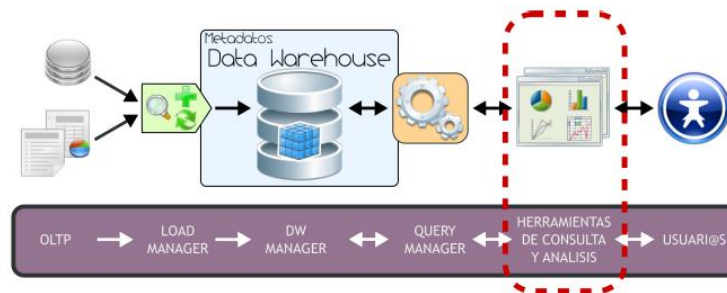
Slice: Tanto la técnica Slice como Dice, son formas particulares de filtrado. Slice consiste en seleccionar un miembro particular de una dimensión, formando así una especie de rebanada (slice) del cubo original.



Dice: En este caso, el usuario selecciona varios miembros de distintas dimensiones, creando un sub-cubo o dado (dice), más pequeño que el original.



3.5.5 Herramientas de Consulta y Análisis



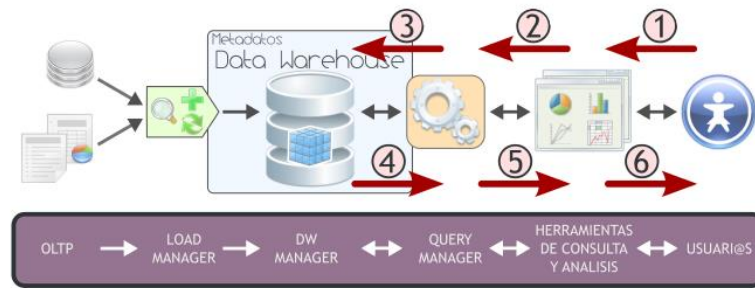
Las herramientas de consulta y análisis son sistemas que permiten a los usuarios realizar la exploración de datos del DW. Básicamente constituyen el nexo entre el depósito de datos y los usuarios. Utilizan la metadata de las estructuras de datos que han sido creadas previamente (cubos multidimensionales, Business Models, etc.) para trasladar a través de consultas SQL los requerimientos de los usuarios, y luego devolver a éstos el resultado obtenido. Estas herramientas suelen emplear simples conexiones a bases de datos (JNDI, JDBC, ODBC), para obtener la información deseada.

A través de una interfaz gráfica y una serie de pasos, los usuarios generan consultas que son enviadas desde la herramienta de consulta y análisis al Query Manager, este a su vez realiza la extracción de información del DW Manager y devuelve los resultados obtenidos a la herramienta que se los solicitó. Luego, estos resultados son expuestos en formatos familiares para los usuarios.

Este proceso se puede comprender mejor al observar la siguiente figura:



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



El mismo, se lleva a cabo a través de seis pasos sucesivos:

1. Los usuarios seleccionan o establecen que datos desean obtener del DW, mediante las interfaces de la herramienta que utilice.
2. La herramienta recibe el pedido de los usuarios, construye la consulta (utilizando la metadata) y la envía al Query Manager.
3. El Query Manager ejecuta la consulta sobre la estructura de datos con la que se esté trabajando (cubo multidimensional, Business Model, etc.).
4. El Query Manager obtiene los resultados de la consulta.
5. El Query Manager envía los datos a la herramienta de consulta y análisis.
6. La herramienta presenta a los usuarios la información requerida.

Una de las principales ventajas de utilizar estas herramientas es que los usuarios no se tienen que preocupar por conocer cuáles son las características y funcionalidades de las estructuras de datos utilizadas, ni por saber emplear el lenguaje SQL, solo se deben enfocar en el análisis.

Las herramientas de consulta y análisis, en general, comparten las siguientes características:

- ✓ *Accesibilidad a la información:* permiten el acceso a la información a través de las diferentes estructuras de datos de forma transparente a los usuarios finales, para que estos solo se enfoquen en el análisis y no en el origen y procedencia de los datos.
- ✓ *Apoyo en la toma de decisiones:* permiten la exploración de los datos, a fin de seleccionar, filtrar y personalizar los mismos, para la obtención de información oportuna, relevante y útil que apoye el proceso de toma de decisiones.
- ✓ *Orientación los usuarios finales:* permiten a través de entornos amigables e intuitivos, que los usuarios puedan realizar análisis y consultas, sin poseer conocimientos técnicos. Si bien lo realmente importante son los datos mismos, que estos puedan ser interpretados y analizados por los usuarios dependerá en gran medida de cómo se presenten y dispongan.

Existen diferentes tipos de herramientas de consulta y análisis, y de acuerdo a la necesidad, tipos de usuarios y requerimientos de información, se deberán seleccionar las más propicias al caso. Entre ellas se destacan las siguientes:

- ✓ Reportes y Consultas.
- ✓ OLAP.
- ✓ Dashboards.
- ✓ Data Mining.



- ✓ Cuadros de Mando Integrales (CMI) o Balanced Scorecard (BSC)
- ✓ Sistemas de Información Ejecutiva (EIS)
- ✓ Sistemas de Soporte a la Decisión (DSS)

3.5.5.1 Reportes y Consultas

Se han desarrollado muchas herramientas para la producción de consultas y reportes que a través de pantallas gráficas intuitivas, ofrecen a los usuarios la posibilidad de generar informes avanzados y detallados del tema de interés que se está analizando, conectándose al Data Warehouse, y otorgándole al usuario la facilidad de no tener que conocer cómo están constituidas sus consultas, ni saber cómo es la estructura de la base de datos multidimensional que está manipulando.

Los usuarios solo deben seguir una serie de simples pasos, como por ejemplo seleccionar opciones de un menú, presionar tal o cual botón para especificar los elementos de datos, sus condiciones, criterios de agrupación y demás atributos que se consideren significativos.

Actualmente las herramientas de generación de reportes y consultas cuentan con muchas prestaciones, las cuales permiten dar variadas formas y formatos a la presentación de la información. Entre las opciones más comunes se encuentran las siguientes:

- Parametrización de los datos devueltos.
- Selección de formatos de salida (planilla de cálculo, HTML, PDF, etc.).
- Inclusión de gráficos de tortas, barras, etc.
- Utilización de plantillas de formatos de fondos.
- Inclusión de imágenes.
- Formatos tipográficos.
- Links a otros reportes.

3.5.5.2 OLAP (*On Line Analytic Processing*).

También conocido como Proceso analítico en línea, es la componente más poderosa del Data Warehousing, ya que es el motor de consultas especializado del depósito de datos. Permite analizar los datos de estructuras multidimensionales de una manera muy intuitiva, otorgando respuestas rápidas a preguntas complejas. Con OLAP se puede ver un conjunto de datos del negocio de muchas y diversas formas sin mucho esfuerzo. Mediante la funcionalidad OLAP un usuario puede explorar de manera minuciosa los subconjuntos de datos interrelacionados o “cubos”, simplemente con un clic. Los usuarios pueden analizar los datos empleando características OLAP estándar, filtros, drill-up/down para obtener distintas vistas de los mismos.

Las herramientas OLAP, son una tecnología de software para análisis en línea, administración y ejecución de consultas, que permiten inferir información del comportamiento del negocio.

Su principal objetivo es el de brindar rápidas respuestas a complejas preguntas, para interpretar la situación del negocio y tomar decisiones. Cabe destacar que lo que es realmente interesante en OLAP, no es la ejecución de simples consultas tradicionales,



sino la posibilidad de utilizar operadores tales como drill-up, drill-down, etc., para explotar profundamente la información.

Además, a través de este tipo de herramientas, se puede analizar el negocio desde diferentes escenarios históricos, y proyectar como se ha venido comportando y evolucionando en un ambiente multidimensional, o sea, mediante la combinación de diferentes perspectivas, temas de interés o dimensiones. Esto permite deducir tendencias, por medio del descubrimiento de relaciones entre las perspectivas que a simple vista serían muy difícil de encontrar.

Las herramientas OLAP requieren que los datos estén organizados dentro del depósito en forma multidimensional, por lo cual se utilizan cubos multidimensionales.

Además de las características descritas, se pueden enumerar las siguientes:

- ✓ Permite recolectar y organizar la información analítica necesaria para los usuarios y disponer de ella en diversos formatos, tales como tablas, gráficos, reportes, tableros de control, etc.
- ✓ Soporta análisis complejos de grandes volúmenes de datos.
- ✓ Complementa las actividades de otras herramientas que requieran procesamiento analítico en línea.
- ✓ Presenta a los usuarios una visión multidimensional de los datos (matricial) para cada tema de interés del negocio.
- ✓ Es transparente al tipo de tecnología que soporta el DW, ya sea ROLAP, MOLAP u HOLAP.
- ✓ No tiene limitaciones con respecto al número máximo de dimensiones permitidas.
- ✓ Permite a los usuarios, analizar la información basándose en más criterios que un análisis de forma tradicional.
- ✓ Al contar con muestras grandes, se pueden explorar mejor los datos en busca de respuestas.
- ✓ Permiten realizar agregaciones y combinaciones de los datos de maneras complejas y específicas, con el fin de realizar análisis más estratégicos.

3.5.5.3 Dashboards

Los Dashboards se pueden entender como una colección de reportes, consultas y análisis interactivos que hacen referencia a un tema en particular y que están relacionados entre sí.

Existen diversas maneras de diseñar un Dashboard, cada una de las cuales tiene sus objetivos particulares. Algunas características de estas herramientas son:

- ✓ Presentan la información altamente resumida.
- ✓ Se componen de consultas, reportes, análisis interactivos, gráficos (de torta, barras, etc.), semáforos, indicadores causa-efecto, etc.
- ✓ Permiten evaluar la situación de la empresa con un solo golpe de vista.
- ✓ Poseen un formato de diseño visual muy llamativo.

3.5.5.4 Data Mining o Minería de Datos



Son tipos de consultas indirectas que buscan encontrar patrones ocultos en los datos. Los resultados más valiosos son agrupamientos, clasificaciones, estimaciones, predicciones y asociaciones de cosas que ocurren juntas. Hay muchos tipos de herramientas para minería de datos. Las principales incluyen árboles de decisión, redes neuronales, razonamiento basado en casos o reglas de asociación, herramientas de visualización, algoritmos genéticos, lógica difusa, y estadística clásica. El Data Mining permite realizar la extracción y análisis de información oculta en grandes bases de datos. Las herramientas de Minería de datos predicen futuras tendencias y comportamientos permitiendo, en los negocios, tomar decisiones proactivas y conducidas por un conocimiento acabado de la información. Los análisis prospectivos automatizados ofrecidos por un producto van más allá de los eventos pasados provistos por herramientas retrospectivas típicas de sistemas de soporte de decisión. Data Mining pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo para poder ser resueltas y a los cuales los usuarios de esta información casi no están dispuestos a aceptar. Estas herramientas exploran las bases de datos en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a hallar porque se encuentra fuera de sus expectativas.

Esta herramienta constituye una poderosa tecnología con un gran potencial que ayuda y brinda soporte a los usuarios, con el fin de permitirles analizar y extraer conocimientos ocultos y predecibles a partir de los datos almacenados en un DW o en un OLTP. Es deseable que la fuente de información sea un DW, por todas las ventajas que aporta.

Una de las principales ventajas del Data Mining es que permite inferir comportamientos, modelos, relaciones y estimaciones de los datos, para poder desarrollar predicciones sobre los mismos, sin la necesidad de contar con patrones o reglas preestablecidas, permitiendo tomar decisiones proactivas basadas en un conocimiento acabado de la información.

Además brinda la posibilidad de dar respuesta a preguntas complicadas sobre los temas de interés, pudiendo ver los resultados en forma de reportes tabulares, matriciales, gráficos, tableros, etc.

Los sistemas Data Mining se desarrollan bajo lenguajes de última generación basados en Inteligencia Artificial y utilizan métodos matemáticos tales como:

- Redes Neuronales.
- Sistemas Expertos.
- Programación Genética.
- Árboles de Decisión.
- Soporta además, sofisticadas operaciones de análisis como los sistemas Scoring, aplicaciones de Detección de Desviación y Detección de Fraude.

3.5.5.5 Cuadros de Mando Integrales (CMI)

También conocidos como Balanced Scorecard (BSC), es una herramienta de control empresarial que permite establecer y monitorizar los objetivos de una empresa y de sus diferentes áreas o unidades. Se pueden considerar como una aplicación que ayuda a una compañía a expresar los objetivos e iniciativas necesarias para cumplir con su estrategia, mostrando de forma continuada cuándo la empresa y los empleados alcanzan los resultados definidos en su plan estratégico. El término Dashboard, anteriormente



definido, se caracteriza por relajar algunas características teóricas del Cuadro de Mando basadas en los principios de Kaplan y Norton. Frecuentemente, el término “Dashboard” se utiliza como sinónimo de “Balanced Scorecard”, sin embargo, si bien ambos comparten algunas similitudes, debe considerárselos como herramientas de gestión del desempeño diferentes. Ambos muestran datos necesarios para tomar decisiones y mejorar el desempeño; y pueden mostrar una selección de medidas financieras y operacionales u indicadores del desempeño clave (KPI “key performance indicators” por su sigla en inglés) que permiten que los usuarios accedan a los indicadores de desempeño.

Pero el Dashboard y el Balanced Scorecard difieren en algunos aspectos importantes, algunos de los cuales son:

- ✓ La metodología que utilizan para seleccionar la información que exhiben.
- ✓ Los usuarios a quienes está dirigida la información.
- ✓ El nivel de detalle de la información.
- ✓ La frecuencia temporal.

3.5.5.6 Sistema de Información Ejecutiva (EIS – Executive Information System)

Es una herramienta de Business Intelligence, orientada a usuarios de nivel gerencial, que permite monitorizar el estado de las variables de un área o unidad de la empresa a partir de información interna y externa a la misma. Se puede considerar que un EIS es un tipo de Sistema de Soporte a la Decisión (DSS) cuya finalidad principal es que el responsable de un departamento o compañía tenga acceso, de manera instantánea, al estado de los indicadores de negocio que le afectan, con la posibilidad de estudiar con detalle aquellos aspectos que no estén cumpliendo con los objetivos establecidos en su plan estratégico u operativo, y así determinar las medidas de contingencia más adecuadas. Una de las características más importantes de un EIS es que permite a usuarios con perfil no técnico construir nuevos informes y navegar por los datos de la compañía, con el objetivo de descubrir información que les resulte relevante. Esto se debe, entre otras cosas, a que la interfaz gráfica de estas aplicaciones suele ser muy atractiva e intuitiva. Proporciona medios sencillos para consultar, analizar y acceder a la información de estado del negocio. Además, pone a disposición facilidades para que los usuarios puedan conseguir los datos buscados rápidamente, empleando el menor tiempo posible para comprender el uso de la herramienta. Usualmente, EIS se utiliza para analizar los indicadores de performance y desempeño del negocio o área de interés, a través de la presentación de vistas con datos simplificados, altamente consolidados, mayormente estáticos y preferentemente gráficos. En algunos casos suele incluir también alertas de negocio, informes históricos comparativos y análisis de tendencias

El concepto principal de esta herramienta, se basa en el simple hecho de que los ejecutivos no poseen tiempo, ni las habilidades necesarias para analizar grandes cantidades de datos.

Al igual que OLAP y Data Mining, los EIS, se pueden aplicar independientemente de la plataforma DW, pero tener como base un depósito de datos para implementar esta herramienta, conlleva todas las ventajas implícitas del mismo.

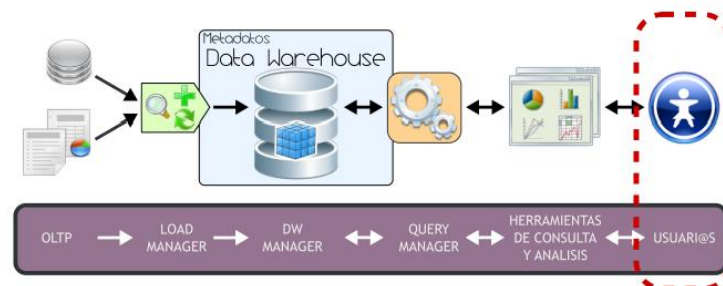
3.5.5.7 *Sistemas de Soporte a la Decisión (DSS)*

Es una herramienta de Business Intelligence enfocada al análisis de los datos de una organización. En principio, puede parecer que el análisis de datos es un proceso sencillo y fácil de conseguir mediante una aplicación hecha a medida. Sin embargo, no es así, estas aplicaciones suelen disponer de una serie de informes predefinidos en los que presentan la información de manera estática, pero no permiten profundizar en los datos, navegar entre ellos, manejarlos desde distintas perspectivas, etc.

El principal objetivo de los Sistemas de Soporte a Decisiones es, a diferencia de otras herramientas como los Cuadros de Mando (CMI) o los Sistemas de Información Ejecutiva (EIS), explotar al máximo la información residente en una base de datos multidimensional, mostrando informes muy dinámicos y con gran potencial de navegación, pero siempre con una interfaz gráfica amigable, vistosa y sencilla. Otra diferencia fundamental radica en que las plataformas DSS están destinadas a usuarios de cualquier nivel gerencial dentro de una organización, tanto para situaciones estructuradas como no estructuradas. (En este sentido, por ejemplo, los CMI están más orientados a la alta dirección).

Por último, es importante destacar que los DSS suelen requerir (aunque no es imprescindible) un motor OLAP subyacente, que facilite el análisis casi ilimitado de los datos para hallar las causas de los problemas de la compañía.

3.5.6 Usuarios



Los usuarios que poseen el DW son aquellos que se encargan de tomar decisiones y de planificar las actividades del negocio, es por ello que se hace tanto énfasis en la integración, limpieza de datos, etc., para poder conseguir que la información posea toda la calidad posible.

Es a través de las herramientas de consulta y análisis, que los usuarios exploran los datos en busca de respuestas para poder tomar decisiones proactivas.

Para comprender mejor a los usuarios del almacén de datos, se hará referencia a las diferencias que estos tienen con respecto a los del OLTP:

- ✓ Los usuarios que acceden al DW concurrentemente son pocos, en cambio los que acceden a los OLTP en un tiempo determinado son muchos más, pueden ser cientos o incluso miles. Esto se debe principalmente al tipo de información que contiene cada fuente.
- ✓ Los usuarios del DW generan por lo general consultas complejas, no predecibles y no anticipadas. Usualmente, cuando se encuentra una respuesta a una consulta



se formulan nuevas preguntas más detalladas y así sucesivamente. Es decir, primero se analiza la información a nivel de datos actual para averiguar el “qué”, luego, para obtener mayor detalle y examinar el “cómo”, se trabajan con los datos ligeramente resumidos, derivados de la consulta anterior, y desde allí se puede explorar los datos altamente resumidos. Teniendo en cuenta siempre la posibilidad de utilizar el detalle de datos histórico. Al contrario, los usuarios de los OLTP solo manejan consultas predefinidas.

- ✓ Los usuarios del DW, generan consultas sobre una gran cantidad de registros, en cambio los del OLTP lo hacen sobre un pequeño grupo. Esto se debe a que como se ha mencionado, el depósito contiene información histórica e integra varias fuentes de datos.
- ✓ Las consultas de los usuarios del DW no tienen tiempos de respuesta críticos, aunque sí se espera que se produzcan en el mismo día en que fueron realizadas. Mientras mayor sea el tamaño del depósito y mientras más compleja sea la consulta, mayores serán los tiempos de respuestas. En cambio, las respuestas de las consultas en un OLTP son y deben ser inmediatas.
- ✓ En los OLTP, los usuarios típicamente realizan actualizaciones, tales como agregar, modificar, eliminar y consultar algún registro. En cambio en un DW, la única operación que pueden realizar es la de consulta.
- ✓ El perfil del usuario que interactúa con los sistemas OLTP se encuadra dentro de los empleados operativos de una organización. Los usuarios de un OLTP hacen que la compañía funcione, teniendo como función principal la entrada de datos. También realizan consultas a nivel de un registro por vez. Los usuarios OLTP realizan las mismas tareas una gran cantidad de veces. Por el contrario, dado el objetivo estratégico y el nivel de información que manejan los DW, el perfil del usuario sobre este tipo de sistemas corresponde a la comunidad gerencial, la cual está a cargo de la toma de decisiones. Los usuarios de un DW miran como funciona la organización y definen el rumbo a seguir. Miran qué datos son nuevos, piden que los datos erróneos sean corregidos. Los usuarios de un DW casi nunca consultan por un registro en particular, usualmente requieren que cientos o miles de registros sean buscados y sean comprimidos en un pequeño conjunto de datos de respuesta. Estos usuarios cambian continuamente los tipos de preguntas. Aunque la estructura de las consultas es similar, el impacto en la base de datos varía en ir a buscar de cientos a millones de registros para ser resumidos en el pequeño conjunto de respuesta.

3.6 Conceptos Complementarios

3.6.1 Data Mart

Un Data Mart (DM) es la implementación de un DW con alcance restringido a un área funcional, problema en particular, departamento, tema o grupo de necesidades.

Un Data Mart es definido como un subconjunto de datos del Data Warehouse, el cual está orientado a un tema específico y responde a un grupo de usuarios específicos. Cabe resaltar que el Data Mart puede ser dependiente o independiente de un Data Warehouse. El Data Warehouse y el Data Mart se diferencian básicamente en el alcance que tiene cada uno de ellos. El Data Mart esta orientado a cubrir las necesidades de información de un grupo de trabajo o de un departamento de una organización, mientras que el Data

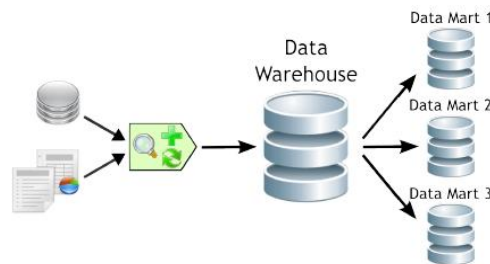
Warehouse tiene como alcance a la organización en su conjunto, analiza datos desde una visión corporativa.

Muchos depósitos de datos comienzan siendo Data Mart, para, entre otros motivos, minimizar riesgos y producir una primera entrega en tiempos razonables. Pero, una vez que estos se han implementado exitosamente, su alcance se irá ampliando paulatinamente.

- ✓ Los Data Marts pueden o no localizarse físicamente en la misma máquina que el DW. Esto permite a los consumidores de la información elegir la mejor tecnología que soporte el estilo de análisis que necesiten.
- ✓ Los Data Marts deben ser implementados como una extensión del DW, no como una alternativa. La estrategia a largo plazo dicta que es necesario contar con la infraestructura completa para tener un DSS saludable.
- ✓ La construcción de Data Marts para soporte de decisiones es ideal. Sin embargo hay que tener en cuenta que la simplicidad de su diseño puede conducir a tener mucha cantidad de ellos implicando así un alto costo para administrarlos.

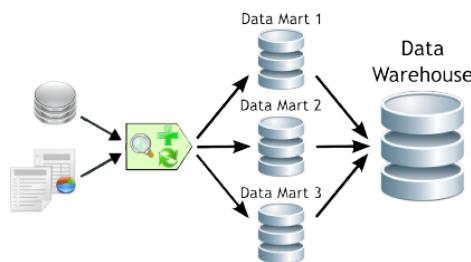
De acuerdo a las operaciones que se deseen o requieran desarrollar, los DM pueden adoptar las siguientes arquitecturas:

- *Top-Down*: primero se define el DW y luego se desarrollan, construyen y cargan los DM a partir del mismo.



Como se observa, el DW es cargado a través de procesos ETL y luego este alimenta a los diferentes DM, cada uno de los cuales recibirá los datos que correspondan al tema o departamento que traten. Esta forma de implementación cuenta con la ventaja de no tener que incurrir en complicadas sincronizaciones de hechos, pero requiere una gran inversión y una gran cantidad de tiempo de construcción.

- *Bottom-Up*: en esta arquitectura, se definen previamente los DM y luego se integran en un DW centralizado.



Los DM se cargan a través de procesos ETL, los cuales suministrarán la información adecuada a cada uno de ellos. En muchas ocasiones, los DM son



implementados sin que exista el DW, ya que tienen sus mismas características pero con la particularidad de que están enfocados en un tema específico. Luego de que hayan sido creados y cargados todos los DM, se procederá a su integración con el depósito. La ventaja que trae aparejada este modelo es que cada DM se crea y pone en funcionamiento en un corto lapso de tiempo permitiendo tener una pequeña solución a un costo no tan elevado. Luego que todos los DM estén puestos en marcha, se puede decidir si construir el DW o no. El mayor inconveniente está dado en tener que sincronizar los hechos al momento de la consolidación en el depósito.

	TOP-DOWN	BOTTOM-UP
VENTAJAS	<ul style="list-style-type: none"> • Requiere de un esfuerzo corporativo. Vista empresarial del negocio. • Se debe definir una arquitectura empresarial. • Se trabaja con un único almacén central, se tiene un control centralizado. • Se puede obtener un resultado rápido si se trabaja con iteraciones. 	<ul style="list-style-type: none"> • Se pueden obtener implementaciones más rápidas porque la unidad de trabajo son los datamarts. • Existen menores riesgos, porque hay menos dependencia entre áreas. • Se obtienen resultados rápidos.

	TOP-DOWN	BOTTOM-UP
DESVENTAJAS	<ul style="list-style-type: none"> • Requiere más tiempo porque se necesita mayor esfuerzo corporativo. • Requiere un conocimiento más profundo del negocio. 	<ul style="list-style-type: none"> • Cada datamart es una vista de una parte de los datos. • Pueden existir datos redundantes o inconsistentes entre los datamarts.

Dentro de las ventajas de aplicar un Data Mart a un negocio, se encuentran:

- ✓ Son simples de implementar.
- ✓ Conllevan poco tiempo de construcción y puesta en marcha.
- ✓ Permiten manejar información confidencial.
- ✓ Reflejan rápidamente sus beneficios y cualidades.
- ✓ Reducen la demanda del depósito de datos.

Considerando la definición de Data Mart y teniendo en cuenta que la solución del trabajo de grado tiene como alcance la Secretaría de Economía y Finanzas Públicas dentro del Municipio de Cosquín es posible concluir que la implementación corresponderá a un DM, pudiendo en un futuro crear el DW de toda la organización.

3.6.2 Business Models

Es una representación de los datos desde una perspectiva empresarial, que permite que se pueda visualizar la información del negocio y su respectiva interrelación.

Se compone de entidades, atributos y relaciones, que están enfocados en dar respuesta a las preguntas de la información que se desea conocer.

El Business Model permite definir el comportamiento que tendrá cada miembro dentro de este, como por ejemplo indicar cuáles campos serán utilizados para realizar sumalizaciones y cuál será el criterio empleado a tal fin además de cuáles serán los campos que se utilizarán para analizar la información.

Pero lo más importante de este tipo de estructura de datos es que el mismo se define a través de reglas de negocio teniendo en cuenta las áreas temáticas que son de interés en la empresa.

Algunas de sus características más sobresalientes son:

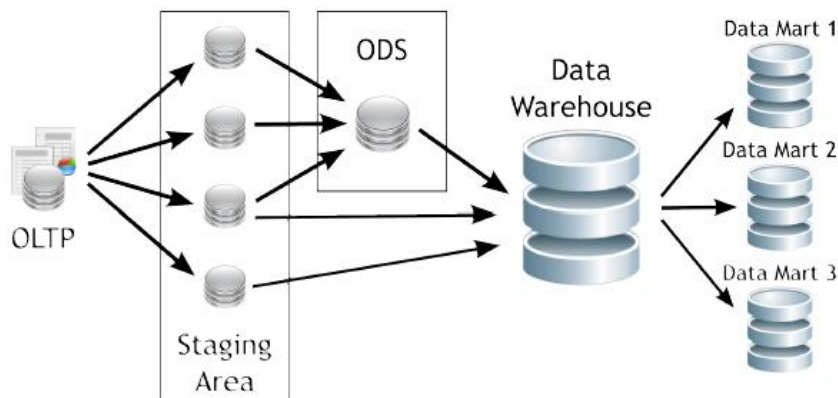
- ✓ Es completamente independiente de las estructuras organizacionales.
- ✓ Plantea la información de la empresa como si fuesen piezas que encajan entre sí.

3.6.3 Área de Datos

Dentro del diseño de la arquitectura de un sistema de Data Warehouse es importante tener en consideración los diferentes entornos por los que han de pasar los datos en su camino hacia el DW o hacia los Data Marts de destino. Dada la cantidad de transformaciones que se han de realizar, y que normalmente el DW, además de cumplir su función de soporte a los requerimientos analíticos, realiza una función de integración de datos que van a conformar el Almacén Corporativo y que van a tener que ser consultados también de la manera tradicional por los sistemas operacionales, es muy recomendable crear diferentes áreas de datos en el camino entre los sistemas origen y las herramientas OLAP.

Cada una de estas áreas se distingue por las funciones que realizan, de qué manera se organizan los datos en la misma, y a qué tipo de necesidad pueden dar servicio. El área que se encuentra “al final del camino” es importante, pero no va a ser la única que almacene los datos que van a explotar las herramientas de reporting.

No hay una convención estándar sobre lo que abarca exactamente cada área y la obligatoriedad de utilizar cada una de ellas. Cada proyecto es diferente e influyen muchos factores como: la complejidad, el volumen de información del mismo, si se quiere utilizar el Data Warehouse como almacén corporativo o Sistema Maestro de Datos, o si existen necesidades reales de soporte al reporting operacional; dependiendo de estos factores se definirán las distintas áreas y sus respectivos alcances.



***Staging Area.***

Es un área temporal donde se escogen los datos que se necesitan de los sistemas origen. Se recogen los datos estrictamente necesarios para las cargas, y se aplica el mínimo de transformaciones a los mismos. No se aplican restricciones de integridad ni se utilizan claves, los datos se tratan como si las tablas fueran ficheros planos. De esta manera se minimiza la afectación a los sistemas origen, la carga es lo más rápida posible para acotar la ventana horaria necesaria, y se reduce también al mínimo la posibilidad de error. Una vez que los datos han sido traspasados, el DW se independiza de los sistemas origen hasta la siguiente carga. Lo único que se suele añadir es algún campo que almacene la fecha de la carga.

Obviamente estos datos no van a dar servicio a ninguna aplicación de reporting, son datos temporales que una vez hayan cumplido su función son eliminados, de hecho en el esquema lógico de la arquitectura muchas veces no aparece, ya que su función es meramente operativa.

Algunos autores consideran que la Staging Area abarca más, o incluso que engloba todo el entorno donde se realizan los procesos de ETL.

Operational Data Store.

Esta área es la que da soporte a los sistemas operacionales. El modelo de datos del Almacén de Datos Operacional (ODS) sigue una estructura relacional y normalizada, para que cualquier herramienta de reporting o sistema operacional pueda consultar sus datos. Está dentro del Data Warehouse porque se aprovecha el esfuerzo de integración que supone la creación del Almacén de Datos Corporativo para poder atender también a necesidades operacionales, pero no es obligatorio. Ni siquiera es algo específico del BI, los ODS ya existían antes de que surgieran los conceptos de Data Warehousing y Business Intelligence.

No almacena datos históricos, muestra la imagen del momento actual, aunque eso no significa que no se puedan registrar los cambios.

Los datos del ODS se recogen de la Staging Area, y en este proceso sí que se realizan transformaciones, limpieza de datos y controles de integridad referencial para que los datos estén perfectamente integrados en el modelo relacional normalizado.

Se debe tener en cuenta que la actualización de los datos del ODS no es instantánea, los cambios en los datos de los sistemas origen no se ven reflejados hasta que finaliza la carga correspondiente. Es decir, que los datos se refrescan cada cierto tiempo, cosa que hay que explicar a los usuarios finales, porque los informes que se lancen contra el ODS siempre devolverán información a fecha de la última carga.

Por esta razón es recomendable definir una mayor frecuencia de carga para el ODS que para el Almacén Corporativo.

Almacén de Datos Corporativo o Data Warehouse.

Este sí contiene datos históricos, y está orientado a la explotación analítica de la información. Las herramientas DSS o de reporting analítico consultan tanto los Data Marts como el Almacén de Datos Corporativo. El DW puede servir consultas en las que se precisa mostrar a la vez información que se encuentre en diferentes Data Marts.



En él se almacenan datos que pueden provenir tanto de la Staging Area como del ODS. Si ya se realizan procesos de transformación e integración en el ODS no se repiten para pasar los mismos datos al Almacén Corporativo. Lo que no se pueda recoger desde el ODS sí que hay que ir a buscarlo a la Staging Area.

El esquema se parece al de un modelo relacional normalizado, pero en él ya se aplican técnicas de desnormalización. No debería contener un número excesivo de tablas ni de relaciones ya que, por ejemplo, muchas relaciones jerárquicas que en un modelo normalizado se implementarían con tablas separadas aquí ya deberían crearse en una misma tabla, que después representarían una dimensión.

Otra particularidad es que la mayoría de las tablas han de incorporar campos de fecha para controlar la fecha de carga, la fecha en que se produce un hecho, o el periodo de validez del registro.

Si el Data Warehouse no es demasiado grande, o el nivel de exigencia no es muy elevado en cuanto a los requerimientos “operacionales”, para simplificar la estructura se puede optar por prescindir del ODS, y si es necesario adecuar el Almacén de Datos Corporativo para servir tanto al reporting operacional como al analítico. En este caso, el área resultante sería el DW Corporativo, pero en ocasiones también se denomina como ODS.

Data Mart

Otra área de datos es el lugar donde se crean los Data Marts. Éstos acostumbran a obtenerse a partir de la información recopilada en el área del Almacén Corporativo, aunque también puede ser a la inversa, como ya se mencionó en el punto 3.6.1. Cada Data Mart es como un subconjunto de este almacén pero orientado a un tema de análisis, normalmente asociado a un departamento de la empresa. El Data Mart se diseña con estructura multidimensional donde cada objeto de análisis es una tabla de hechos enlazada con diversas tablas de dimensiones. Al igual que un DW pueden ser diseñados con el modelo de estrella o copo de nieve, explicados anteriormente.

Este área puede residir en la misma base de datos que las demás si la herramienta de explotación es de tipo ROLAP, o también puede crearse ya fuera de la BD, en la estructura de datos propia que generan las aplicaciones de tipo MOLAP, más conocida como los cubos multidimensionales.

Si se sigue una aproximación Top-down para la creación de los Data Mart, el paso del área de DW a esta ha de ser bastante simple, cosa que además proporciona una cierta independencia sobre el software que se utiliza para el reporting analítico. Si por cualquier razón es necesario cambiar la herramienta de OLAP hay que hacer poco más que redefinir los metadatos y regenerar los cubos, y si el cambio es entre dos de tipo ROLAP ni siquiera esto último sería necesario. En cualquier caso, las áreas anteriores no tienen por qué ser modificadas.

3.6.4 Requisitos que debe cumplir un Data Warehouse

Algunos requerimientos que se establecen para los DW son:

- ✓ El DW debe proveer acceso fácil a la información de una organización. Información accesible significa que administradores y analistas de una organización deben poder conectarse al DW desde sus computadoras personales.



La conexión debe ser inmediata, a demanda y de alta velocidad. No es aceptable que el acceso sea a través de otra persona, sea inseguro o lento. El DW no es sólo los datos, sino también un conjunto de herramientas para consultar, analizar y presentar la información. Las herramientas de acceso deben ser simples y fáciles de usar. Los contenidos del DW deben ser entendibles y navegables, esto quiere decir que deben estar correctamente etiquetados y que los datos pueden ser separados y combinados por el significado de toda posible medida del negocio.

- ✓ Los datos del DW deben ser consistentes y de buena calidad. La información del DW debe ser creíble. Consistencia implica resolver las correspondencias entre la información de diferentes partes de la organización. Si dos medidas de una organización tienen el mismo nombre deben tener el mismo significado. Inversamente si dos medidas difieren en significado deben etiquetarse de manera distinta. Los datos que se publican en el DW deben ser útiles. Como los datos provienen de diferentes fuentes de información, deben ser combinados cuidadosamente, atravesar una etapa de limpieza que debe asegurar la calidad antes de pasar a ser parte del DW. La calidad de los datos del DW debe ser una conductora para la reingeniería del negocio. El DW no puede arreglar la pobre calidad de los datos. Si los datos son opcionales y no están completos no hay nada que el DW pueda hacer. La única forma de mejorar la calidad afecta a las personas que ingresan datos al sistema y a los administradores y consiste en volver al origen del dato con mejores sistemas, mejores administraciones y mejor visibilidad del valor de buen dato. Muchas veces al publicar los datos incompletos la gente ve lo valioso que sería contar con datos de mejor calidad. De esta forma el DW puede jugar un rol clave en los esfuerzos de reingeniería del negocio de una organización.
- ✓ Debe ser una fuente de información adaptable y flexible a los cambios. El DW debe estar diseñado para manejar los cambios continuos de las necesidades de los usuarios, las condiciones del negocio, los datos y las tecnologías. Los datos y aplicaciones existentes en el DW no deben sufrir modificaciones cuando se agregan nuevos datos y/o nuevas preguntas a realizar.
- ✓ El DW debe ser un lugar seguro donde la información se encuentre protegida. El DW contiene información muy valiosa para la organización. Es necesario que existan controles de acceso efectivo a los datos. También se debe permitir a sus dueños visualizar los usos y abusos de los datos, incluso luego de haber abandonado el DW.
- ✓ Debe ser la base para la toma de decisiones. El DW debe contener los datos correctos para soportar la toma de decisiones. Hay sólo una verdadera salida del DW: las decisiones que son tomadas a partir de él. El DW tendrá más valor cuanto más impacte en las decisiones del negocio. Recordar que el DW surge para constituirse como la parte central de un sistema para la toma de decisiones que es lo que finalmente se está tratando de construir.
- ✓ Debe ser aceptado por la comunidad usuaria para considerarse exitoso. No importa si la solución construida es elegante y usa los mejores productos y plataformas si el DW no se utiliza para el propósito que fue concebido. Su uso muchas veces es opcional, a diferencia de los sistemas operacionales donde los usuarios están obligados a utilizarlos. La aceptación de los usuarios de estas herramientas pasa por la simplicidad sobre todas las cosas. Si la comunidad del



negocio no adopta al DW y continua usándolo activamente seis meses después del entrenamiento, entonces habrá fallado el test de aceptación.

De esta lista es fácil ver que para poder implementar exitosamente un DW se requiere mucho más que un buen equipo técnico. Es necesario conocer las reglas del negocio, involucrar a los usuarios y contar con datos de buena calidad. Se requiere conformar un equipo de trabajo con diferentes perfiles para lograr éxito en el proyecto.

3.7 Metodología HEFESTO para la construcción de un Data Warehouse.

La metodología utilizada como referencia para el desarrollo del proyecto será Hefesto. Si bien en el ámbito de la Inteligencia de Negocios existen diferentes metodologías para la construcción de sistemas Data Warehouse, en general, la mayoría poseen un eje en común y coinciden en los puntos clave, variando en algunos pasos o etapas menores. A continuación se procederá a la presentación y descripción de:

- ✓ Hefesto, metodología desarrollada por el Ing. Bernabeu.

HEFESTO, permite la construcción de sistemas Data Warehouse de forma sencilla, ordenada e intuitiva. Su nombre fue inspirado en el dios griego de la construcción y el fuego, y su logotipo es el siguiente:



HEFESTO es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes y experiencias propias de su creador en procesos de confección de almacenes de datos.

Esta metodología fue desarrollada por Darío Bernabeu, Ingeniero en Sistemas graduado del Instituto Universitario Aeronáutico (IUA), cofundador de eGluBI (www.eglubi.com.ar), especialista en el desarrollo e implementación de soluciones OSBI (Open Source Business Intelligence), gestión de proyectos, análisis de requerimientos/necesidades, despliegue y configuración de soluciones BI, confección de procesos de integración de datos, modelado de Data Warehouse, confección de Cubos Multidimensionales y Business Models, desarrollo de reportes ad hoc, reportes avanzados, análisis interactivos, dashboards, etc.

Es importante destacar que el Ing. Bernabeu es también reconocido por algunas publicaciones muy destacadas como:

- ✓ “Data Warehousing: Investigación y Sistematización de Conceptos – HEFESTO: Metodología para la Construcción de un DW”;
- ✓ Artículos en la revista Novatica (<http://www.ati.es/novatica/>).
- ✓ Coescritor del último libro de Pentaho: Pentaho 5.0 Reporting.

Cabe subrayar que HEFESTO esta en continua evolución, y se han tenido en cuenta, como gran valor agregado todos los feedbacks que han aportado quienes han utilizado esta metodología en diversos países y con diversos fines.



La construcción e implementación de un DW puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software, con la salvedad de que para algunas fases en particular, las acciones que se han de realizar serán muy diferentes.

Una de las causas por las cuales fue seleccionada HEFESTO por encima de otras metodologías, es que permite no entrar en la utilización de metodologías que requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue muy largas. Se busca entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del DW y motivar a los usuarios.

HEFESTO puede ser embebida en cualquier ciclo de vida que cumpla con dicha condición. Esta metodología parte de la recolección de requerimientos y necesidades de información de los usuarios, y concluye en la confección de un esquema lógico y sus respectivos procesos de extracción, transformación y carga de datos.

Es posible resumir a HEFESTO en el siguiente gráfico.



Como se puede apreciar, se comienza recolectando las necesidades de información de los usuarios y se obtienen las preguntas claves del negocio. Luego, se deben identificar los indicadores resultantes de los interrogativos y sus respectivas perspectivas de análisis mediante las cuales se construirá el modelo conceptual de datos del DW.



Después, se analizarán los OLTP para determinar cómo se construirán los indicadores, se deberán señalar las correspondencias con los datos fuentes y seleccionar los campos de estudio de cada perspectiva.

Una vez hecho esto, se pasará a la construcción del modelo lógico del depósito, en donde se definirá cuál será el tipo de esquema que se implementará. Seguidamente, se confeccionarán las tablas de dimensiones y las tablas de hechos, para luego efectuar sus respectivas uniones.

Por último, utilizando técnicas de limpieza y calidad de datos, procesos ETL, etc., se definirán políticas y estrategias para la carga inicial del DW y su respectiva actualización.

Características

Esta metodología cuenta con las siguientes características:

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos de los usuarios, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra a los usuarios finales en cada etapa para que tome decisiones respecto al comportamiento y funciones del DW.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el DW y de su respectiva distribución.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- Se aplica tanto para Data Warehouse como para Data Mart.

Pasos y aplicación metodológica

3.7.1 Paso 1) ANÁLISIS DE REQUERIMIENTOS.

Lo primero que se hará será identificar los requerimientos de los usuarios a través de preguntas que expliciten los objetivos de su organización. Luego, se analizarán estas preguntas a fin de identificar cuáles serán los indicadores y perspectivas que serán tomadas en cuenta para la construcción del DW. Finalmente se confeccionará un modelo conceptual en donde se podrá visualizar el resultado obtenido en este primer paso. Es muy importante tener en cuenta que HEFESTO se puede utilizar para construir un Data Warehouse o un Data Mart, por ello es que si se requiere construir dos Data Marts, se deberá aplicar la metodología dos veces, una por cada Data Mart y del mismo modo, si se analizan dos áreas de interés de negocio, como el área de “Ventas” y “Compras”, se deberá aplicar la metodología dos veces.

3.7.1.1 a) Identificar preguntas



El primer paso comienza con el acopio de las necesidades de información, el cual puede llevarse a cabo a través de muy variadas y diferentes técnicas, cada una de las cuales poseen características inherentes y específicas, como por ejemplo entrevistas, cuestionarios, observaciones, etc.

El análisis de los requerimientos de los diferentes usuarios, es el punto de partida de esta metodología, ya que ellos son los que deben, en cierto modo, guiar la investigación hacia un desarrollo que refleje claramente lo que se espera del depósito de datos, en relación a sus funciones y cualidades.

El objetivo principal de esta fase, es la de obtener e identificar las necesidades de información clave de alto nivel, que es esencial para llevar a cabo las metas y estrategias de la empresa, y que facilitará una eficaz y eficiente toma de decisiones.

Debe tenerse en cuenta que dicha información, es la que proveerá el soporte para desarrollar los pasos sucesivos, por lo cual, es muy importante que se preste especial atención al relevar los datos. Una forma de asegurarse de que se ha realizado un buen análisis, es corroborar que el resultado del mismo haga explícitos los objetivos estratégicos planteados por la empresa que se está estudiando.

Otra forma de encaminar el relevamiento, es enfocar las necesidades de información en los procesos principales que desarrolle la empresa en cuestión. La idea central es, que se formulen preguntas complejas sobre el negocio, que incluyan variables de análisis que se consideren relevantes, ya que son estas las que permitirán estudiar la información desde diferentes perspectivas. Un punto importante que debe tenerse muy en cuenta, es que la información debe estar soportada de alguna manera por algún OLTP, ya que de otra forma, no se podrá elaborar el DW.

3.7.1.2 b) Identificar indicadores y perspectivas

Una vez que se han establecido las preguntas de negocio, se debe proceder a su descomposición para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán.

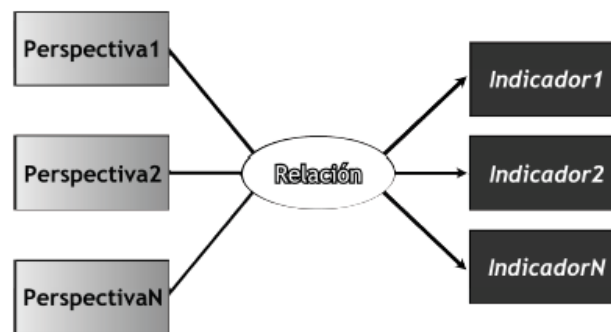
Para ello, se debe tener en cuenta que los indicadores, para que sean realmente efectivos son, en general, valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, fórmulas, etc.

En cambio, las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, países, productos, rubros, etc. Cabe destacar, que el Tiempo es muy comúnmente una perspectiva.

3.7.1.3 c) Modelo Conceptual

En esta etapa, se construirá un modelo conceptual (descripción de alto nivel de la estructura de la base de datos, en la cual la información es representada a través de objetos, relaciones y atributos.) a partir de los indicadores y perspectivas obtenidas en el paso anterior. A través de este modelo, se podrá observar con claridad cuáles son los alcances del proyecto, para luego poder trabajar sobre ellos, además al poseer un alto nivel de definición de los datos, permite que pueda ser presentado ante los usuarios y explicado con facilidad.

La representación gráfica del modelo conceptual es la siguiente:



A la izquierda se colocan las perspectivas seleccionadas, que serán unidas a un óvalo central que representa y lleva el nombre de la relación que existe entre ellas. La relación, constituye el proceso o área de estudio elegida. De dicha relación y entrelazadas con flechas, se desprenden los indicadores, estos se ubican a la derecha del esquema. Como puede apreciarse en la figura anterior, el modelo conceptual permite de un solo vistazo y sin poseer demasiados conocimientos previos, comprender cuáles serán los resultados que se obtendrán, cuáles serán las variables que se utilizarán para analizarlos y cuál es la relación que existe entre ellos.

3.7.2 Paso 2) ANÁLISIS DE LOS OLTP

En este paso se analizarán las fuentes OLTP para determinar cómo serán calculados los indicadores y para establecer las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las fuentes de datos. Luego, se definirán qué campos se incluirán en cada perspectiva. Finalmente, se ampliará el modelo conceptual con la información obtenida en este paso.

3.7.2.1 a) Conformar indicadores

En este paso se deberán explicitar cómo se calcularán los indicadores, definiendo los siguientes conceptos para cada uno de ellos:

- Hecho/s que lo componen, con su respectiva fórmula de cálculo. Por ejemplo: Hecho1 + Hecho2.
- Función de sumariación que se utilizará para su agregación. Por ejemplo: SUM, AVG, COUNT, etc.

3.7.2.2 b) Establecer correspondencias

El objetivo de este paso, es el de examinar los OLTP disponibles que contengan la información requerida, como así también sus características, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos.

La idea es que todos los elementos del modelo conceptual estén correspondidos en los OLTP.

3.7.2.3 c) Nivel de granularidad

Una vez que se han establecido las relaciones con los OLTP, se deben seleccionar los campos que contendrá cada perspectiva, ya que será a través de estos por los que se examinarán y filtrarán los indicadores.

Para ello, basándose en las correspondencias establecidas en el paso anterior, se debe presentar a los usuarios los datos de análisis disponibles para cada perspectiva. Es muy importante conocer en detalle que significa cada campo y/o valor de los datos encontrados en los OLTP, por lo cual, es conveniente investigar su sentido, ya sea a través de diccionarios de datos, reuniones con los encargados del sistema, análisis de los datos propiamente dichos, etc.

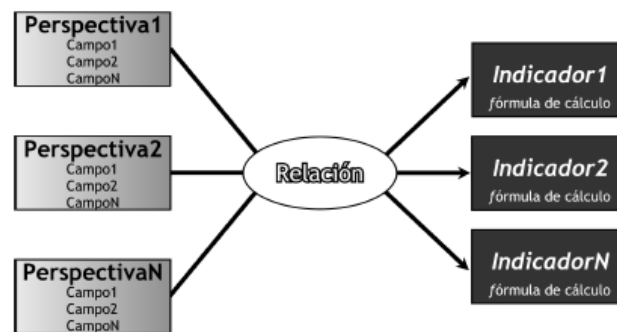
Luego de exponer frente a los usuarios los datos existentes, explicando su significado, valores posibles y características, estos deben decidir cuáles son los que consideran relevantes para consultar los indicadores y cuáles no.

Con respecto a la perspectiva “Tiempo”, es muy importante definir el ámbito mediante el cual se agruparán o sumarán los datos. Sus campos posibles pueden ser: día de la semana, quincena, mes, trimestres, semestre, año, etc.

Al momento de seleccionar los campos que integrarán cada perspectiva, debe prestarse mucha atención, ya que esta acción determinará la granularidad de la información encontrada en el DW.

3.7.2.4 d) Modelo Conceptual ampliado

En este paso, y con el fin de graficar los resultados obtenidos en los pasos anteriores, se ampliará el modelo conceptual, colocando bajo cada perspectiva los campos seleccionados y bajo cada indicador su respectiva fórmula de cálculo. Gráficamente:



3.7.3 Paso 3) MODELO LÓGICO DEL DW

El modelo lógico es la representación de una estructura de datos, que puede procesarse y almacenarse en algún SGBD. Para la confección de un modelo lógico, primero se debe definir el tipo de modelo que se utilizará y luego se llevarán a cabo las acciones propias al caso, para diseñar las tablas de dimensiones y de hechos. Finalmente, se realizan las uniones pertinentes entre estas tablas.



3.7.3.1 a) Tipo de Modelo Lógico del DW

Se debe seleccionar cuál será el tipo de esquema que se utilizará para contener la estructura del depósito de datos, que se adapte mejor a los requerimientos y necesidades de los usuarios. Es muy importante definir objetivamente si se empleará un esquema en estrella, constelación o copo de nieve, ya que esta decisión afectará considerablemente la elaboración del modelo lógico.

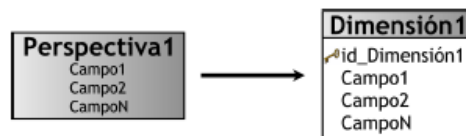
3.7.3.2 b) Tablas de dimensiones

En este paso se deben diseñar las tablas de dimensiones que formaran parte del DW.

Para los tres tipos de esquemas, cada perspectiva definida en el modelo conceptual constituirá una tabla de dimensión. Para ello deberá tomarse cada perspectiva con sus campos relacionados y realizarse el siguiente proceso:

- Se elegirá un nombre que identifique la tabla de dimensión.
- Se añadirá un campo que represente su clave principal.
- Se redefinirán los nombres de los campos si es que no son lo suficientemente intuitivos.

Gráficamente:



Para los esquemas copo de nieve, cuando existan jerarquías dentro de una tabla de dimensión, esta tabla deberá ser normalizada.

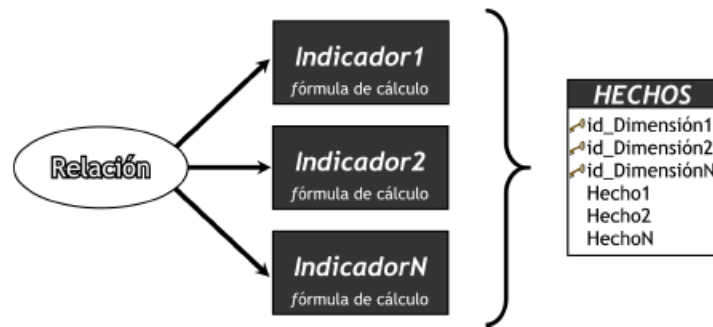
3.7.3.3 c) Tablas de hechos

En este paso, se definirán las tablas de hechos, que son las que contendrán los hechos a través de los cuales se construirán los indicadores de estudio.

Para los esquemas en estrella y copo de nieve, se debe realizar:

- Se le deberá asignar un nombre a la tabla de hechos que represente la información analizada, área de investigación, negocio enfocado, etc.
- Se definirá su clave primaria, que se compone de la combinación de las claves primarias de cada tabla de dimensión relacionada.
- Se crearán tantos campos de hechos como indicadores se hayan definido en el modelo conceptual y se les asignará los mismos nombres que estos. En caso que se prefiera, podrán ser nombrados de cualquier otro modo.

Gráficamente:



Para los esquemas constelación se realizará lo siguiente:

- Las tablas de hechos se deben confeccionar teniendo en cuenta el análisis de las preguntas realizadas por los usuarios en pasos anteriores y sus respectivos indicadores y perspectivas.
- Cada tabla de hechos debe poseer un nombre que la identifique, contener sus hechos correspondientes y su clave debe estar formada por la combinación de las claves de las tablas de dimensiones relacionadas.
- Al diseñar las tablas de hechos, se deberá tener en cuenta:
 - ✓ Caso 1: Si en dos o más preguntas de negocio figuran los mismos indicadores pero con diferentes perspectivas de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición.
 - ✓ Caso 2: Si en dos o más preguntas de negocio figuran diferentes indicadores con diferentes perspectivas de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición.
 - ✓ Caso 3: Si el conjunto de preguntas de negocio cumplen con las condiciones de los dos puntos anteriores se deberán unificar aquellos interrogantes que posean diferentes indicadores pero iguales perspectivas de análisis, para luego reanudar el estudio de las preguntas.

3.7.3.4 d) Uniones

Para los tres tipos de esquemas, se realizarán las uniones correspondientes entre sus tablas de dimensiones y sus tablas de hechos.

3.7.4 Paso 4) INTEGRACIÓN DE DATOS

Una vez construido el modelo lógico, se deberá proceder a poblarlo con datos, utilizando técnicas de limpieza y calidad de datos, procesos ETL, etc.; luego se definirán las reglas y políticas para su respectiva actualización, así como también los procesos que la llevarán a cabo.

3.7.4.1 a) Carga Inicial

En este paso se debe realizar la Carga Inicial al DW, poblando el modelo de datos que se ha construido anteriormente. Para lo cual se deberá llevar adelante una serie de tareas básicas, tales como limpieza de datos, calidad de datos, procesos ETL, etc. La realización de estas tareas puede contener una lógica realmente compleja en algunos



casos. Afortunadamente, en la actualidad existen muchos software que se pueden emplear a tal fin.

Se debe evitar que el DW sea cargado con valores faltantes o anómalos, así como también se deben establecer condiciones y restricciones para asegurar que solo se utilicen los datos de interés.

Cuando se trabaja con un esquema constelación, hay que tener presente que varias tablas de dimensiones serán compartidas con diferentes tablas de hechos, ya que puede darse el caso de que algunas restricciones aplicadas sobre una tabla de dimensión en particular para analizar una tabla de hechos, se puedan contraponer con otras restricciones o condiciones de análisis de otras tablas de hechos.

Primero se cargarán los datos de las dimensiones y luego los de las tablas de hechos, teniendo en cuenta siempre, la correcta correspondencia entre cada elemento. En el caso en que se esté utilizando un esquema copo de nieve, cada vez que existan jerarquías de dimensiones se comenzarán cargando las tablas de dimensiones del nivel más general al más detallado.

Concretamente, en este paso se deberá registrar en detalle las acciones llevadas a cabo con los diferentes softwares.

3.7.4.2 b) Actualización

Cuando se haya cargado en su totalidad el DW, se deben establecer sus políticas y estrategias de actualización o refresco de datos.

Una vez realizado esto, se tendrán que llevar a cabo las siguientes acciones:

- ✓ Especificar las tareas de limpieza de datos, calidad de datos, procesos ETL, etc., que deberán realizarse para actualizar los datos del DW.
- ✓ Especificar de forma general y detallada las acciones que deberá realizar cada software.



DESARROLLO PRÁCTICO



4.1 Abstracto del Desarrollo Práctico

En el presente capítulo se expondrá el desarrollo práctico del trabajo final de grado (TFG) “Implementación de un Data Warehouse para la Municipalidad de Cosquín”. Habiéndose descripto, comentado, analizado y aprendido los diferentes conceptos teóricos relacionados tanto al ámbito municipal, así como a la Inteligencia de Negocios es posible plantear una solución a los problemas presentados en el capítulo 1 y lograr los objetivos del presente proyecto. Es por ello que en las siguientes secciones del trabajo se realizará la aplicación práctica de los conceptos desarrollados durante el marco teórico.

El capítulo ha sido denominado “Desarrollo Práctico” y comenzará presentando la metodología propuesta para el modelado y construcción del Data Warehouse, describiendo todas las etapas y pasos que serán realizados para la correcta implementación del sistema mencionado. Se propone una metodología dividida en 4 secciones, abarcando cada una de ellas una serie de etapas y pasos bien diferenciados que permitirán una mejor comprensión del proyecto.

En la primera etapa se presenta la organización sobre la que realiza el estudio, en este caso el Municipio de Cosquín, sobre el cual se efectuará un análisis general, analizando su contexto y actualidad. De ésta forma es posible situarse en el lugar en donde ocurre el proyecto, y comprender determinadas cuestiones propias de la organización donde se implementara el proyecto.

En la segunda etapa se exponen todas las tareas y actividades relacionadas al modelado del Data Warehouse, definiendo el nivel de información que contendrá, los actores que lo utilizarán y la arquitectura del mismo. Esta etapa se basa en los primeros pasos de la metodología Hefesto, la cual fue presentada y analizada durante al marco teórico del presente trabajo.

En la tercera etapa se aplica parte del Proceso Racional de Toma de Decisiones, específicamente la fase de Análisis de Decisiones, la cual permitirá realizar una selección correcta del conjunto de herramientas de software necesarias para realizar la construcción e implementación del Data Warehouse.

Finalmente en la cuarta y última etapa se lleva a cabo la implementación y puesta en marcha del proyecto en el servidor correspondiente. Así mismo, se realizan los informes adecuados para cada usuario, se definen los roles y se presentan diferentes capturas de pantalla del funcionamiento del sistema en la actualidad.

El presente capítulo se organiza en base a las 4 etapas de la metodología propuesta.

4.2 Metodología Propuesta

Para realizar la implementación, se ha utilizado como metodología de referencia HEFESTO. Si bien se ha tomado como eje esta metodología, la misma ha sido complementada con aportes propios de los autores del trabajo. Dicha complementación surge del cursado de la materia “Toma de Decisiones” dictada por la Ingenieria María Alejandra Boggio, así como con documentos hallados en la web, a fin de adaptarla a las necesidades y características propias del presente proyecto. Se han agregado diferentes



etapas y pasos que incluyen no solo el modelamiento del sistema Data Warehouse, sino actividades previas que permitan tener un conocimiento exacto de la organización donde se implementará el sistema Data Warehouse, así como etapas posteriores que posibiliten realizar una correcta selección del conjunto de herramientas de software y eficaz utilización de las mismas para la implementación del proyecto.

Es importante recordar que la finalidad del trabajo es realizar el análisis, diseño e implementación de una Solución de Inteligencia de Negocios, de manera que permita contar con información sólida y confiable, de alta disponibilidad, y que cumpla con las necesidades de la Secretaría de Economía y Finanzas Públicas del Municipio de Cosquín permitiendo mejorar la gestión y la toma de decisiones.

Finalmente la metodología propuesta para el proyecto quedo establecida de la siguiente manera:

ANALISIS GENERAL DE LA ORGANIZACIÓN

- ✓ Descripción de la organización
- ✓ Funciones
- ✓ Objetivos
- ✓ Organigrama
- ✓ Situación informática actual
- ✓ Relación de las metas de la organización con las del DW
- ✓ Procesos

MODELADO DEL DATA WAREHOUSE

Análisis de Requerimientos

- ✓ Identificar preguntas
- ✓ Identificar indicadores y perspectivas
- ✓ Modelo conceptual

Análisis de los OLTP

- ✓ Conformar indicadores
- ✓ Establecer correspondencias
- ✓ Nivel de granularidad
- ✓ Modelo conceptual ampliado

Modelo Lógico del DW

- ✓ Tipo de Modelo Lógico del DW
- ✓ Tablas de dimensiones
- ✓ Tablas de hechos
- ✓ Uniones

SELECCIÓN DEL CONJUNTO DE HERRAMIENTAS

Análisis de Decisiones

- ✓ Enunciado de la decisión o del propósito
- ✓ Definición de objetivos
- ✓ Generación y evaluación de las alternativas



- ✓ Análisis de las consecuencias

IMPLEMENTACION Y PUESTA EN MARCHA

Descripción de la Tecnología

- ✓ Descripción de las herramientas y elementos de software a utilizar
- ✓ Comparación entre tecnología existente y necesaria para el proyecto

Definición de la arquitectura del Data Warehouse

Integración de Datos

- ✓ Carga Inicial
- ✓ Actualización

Creación de cubos multidimensionales

Herramientas de reportes y consulta.

Instalación y Puesta en marcha

4.3 Análisis general de la organización

4.3.1 Descripción de la organización

El proyecto es realizado para la Municipalidad de la ciudad de Cosquín la cual se encuentra ubicada en la provincia de Córdoba, Argentina. Cosquín es una ciudad del oeste de la provincia de Córdoba, en pleno Valle de Punilla, a 58 km de la capital provincial. La misma es cabecera del departamento de Punilla, al pie de las Sierras Chicas. Dentro de esta ciudad, se encuentra su Municipio, del cual dependen un número importante de Secretarías como son: Secretaría de Gobierno, Secretaría de Economía y Finanzas Públicas, Secretaría de Planeamiento y Desarrollo Humano, Secretaría de Salud Ambiente y Desarrollo Sustentable, Secretaría de Desarrollo Social, Secretaría de Educación y Cultura, Secretaría de Turismo y Deportes, Secretaría de Servicios Públicos y Mantenimiento y la Secretaría de Asesoría Letrada.

En la página web del municipio <http://www.cosquin.gov.ar/> es posible encontrar información al respecto, desde cómo llegar a la ciudad hasta diferentes secciones donde se puede conocer la historia, cultura, educación y más.





Funciones de la Municipalidad de Cosquín:

- ✓ Organización del espacio físico. Uso del suelo.
- ✓ Brindar servicios públicos locales de buena calidad (Educación, Salud, Cultura, Deportes, etc.).
- ✓ Protección y conservación del ambiente.
- ✓ Crecimiento en materia de desarrollo y economía local.
- ✓ Permitir la participación vecinal.
- ✓ Mejora de servicios sociales.
- ✓ Prevención, rehabilitación y lucha contra el consumo de drogas y otras adicciones.
- ✓ Optimizar las instalaciones para el turismo.

Como se definió en los primeros puntos del trabajo, el presente proyecto se centrará en la Secretaría de Economía y Finanzas Públicas, siendo necesario recolectar información de los departamentos de Tesorería, Contaduría, y de la Dirección de Recursos Tributarios, para lograr los objetivos planteados. Debido a esta delimitación en el alcance del proyecto, los puntos siguientes de esta etapa se centraran en la Secretaria de Economía y Finanzas Públicas y no en la Municipalidad, para de este modo tener información más específica del área en cuestión.

4.3.2 Objetivos

Compete a la Secretaría de Economía y Finanzas Públicas asistir al Departamento Ejecutivo Municipal en todo lo inherente a la elaboración y control de la ejecución del presupuesto municipal, como así también de los niveles del gasto y de los ingresos, conforme a las pautas que fije el Departamento Ejecutivo Municipal.

4.3.3 Funciones

- 1) La determinación de los objetivos y la formulación de las políticas del área de su competencia;
- 2) La ejecución de planes, programas y proyectos del área de su competencia, elaborados conforme a las directivas que imparta el Departamento Ejecutivo Municipal;
- 3) La elaboración del proyecto de presupuesto municipal;
- 4) La fijación de la política tributaria, percepción de los recursos municipales y régimen de pagos;
- 5) El análisis, evaluación y control de la ejecución presupuestaria;
- 6) La conducción de la Tesorería y el régimen de pagos;
- 7) Lo referente a la contabilidad pública y a la fiscalización de todo gasto e inversión que se ordene sobre el tesoro de la Municipalidad;
- 8) La distribución de las rentas municipales, conforme con la asignación de presupuesto aprobado por el Concejo Deliberante;
- 9) La elaboración, aplicación, ejecución y fiscalización del régimen de contrataciones y suministros del Gobierno Municipal;
- 10) La elaboración, aplicación, ejecución y fiscalización de la política financiera;
- 11) La gestión de cobro judicial y extrajudicial de tributos y multas impuestas por las distintas reparticiones municipales;
- 12) La organización, dirección, fiscalización, administración y registro de los bienes municipales;



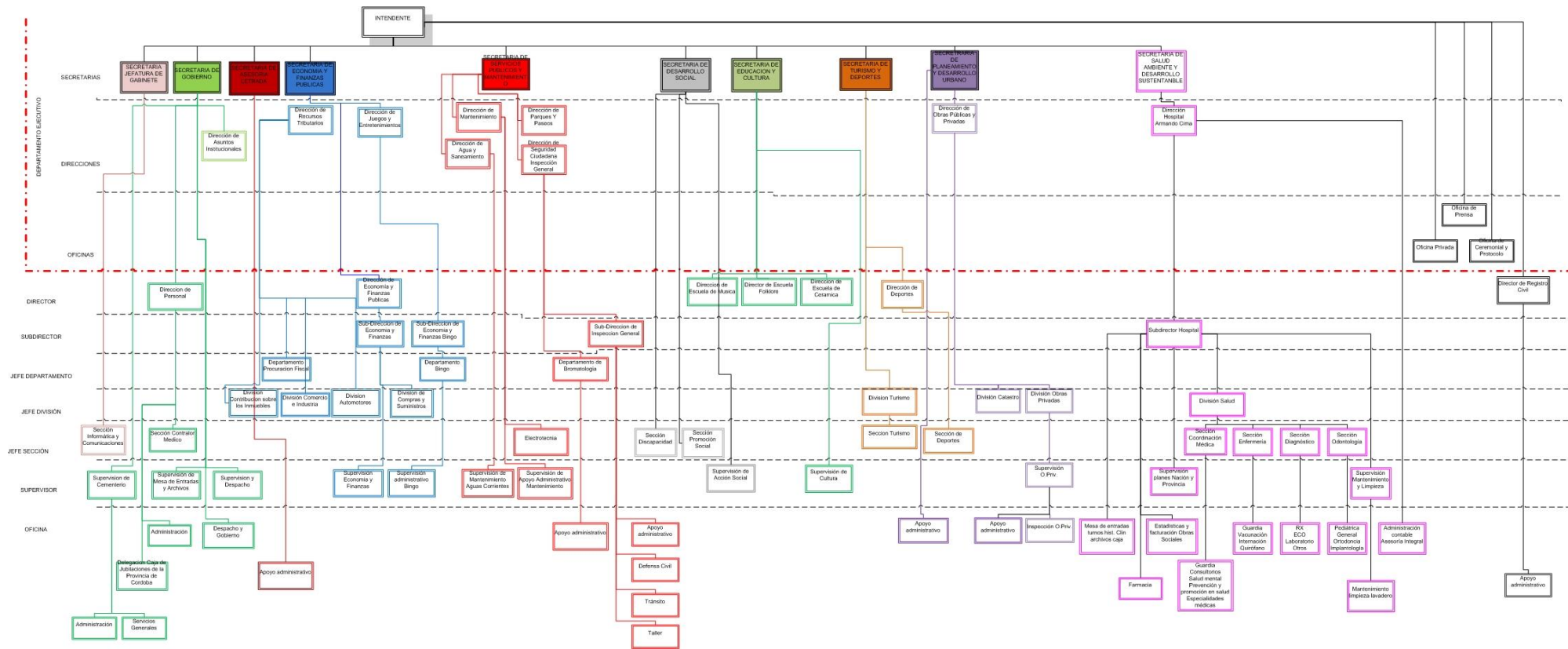
- 13) La elaboración de planes de inversión pública -directos e indirectos- y su posterior ejecución, según las prioridades y directivas que determine el Departamento Ejecutivo Municipal;
- 14) La verificación, centralización y conducción de la información sobre el endeudamiento público de la Municipalidad;
- 15) La recopilación y procesamiento de la información estadística necesaria para la actividad del Gobierno Municipal,
- 16) El control de gestión de las salas municipales de entretenimientos y juegos, y
- 17) Las gestiones necesarias para obtener financiamiento y crédito ante instituciones financieras oficiales o privadas del ámbito provincial, nacional o internacional, así como convenir planes de amortización, intereses y demás condiciones relacionadas con la obtención de la financiación respectiva.

4.3.4 Organigrama

En este paso se presenta el Organigrama de la Municipalidad, en donde es posible observar los departamentos y direcciones que conforman la Secretaría de Economía y Finanzas Públicas.



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI





4.3.5 Situación informática actual

En la actualidad el municipio de Cosquín utiliza para su administración un sistema operacional desarrollado por Program Consultores S.A.

Es un Sistema Integral de Administración Municipal de Aplicaciones con tecnología para Internet “WEB ENABLED”, adaptado a la legislación local, provincial y nacional.

El Sistema está constituido por diferentes módulos, multiusuario y sin límite de licencias. Está desarrollado en ASP y ASP.NET con base de datos Microsoft SQL Server.

Los módulos implementados son los siguientes:

- ✓ Módulo general de integración.
- ✓ Módulo de Tasas, Servicios, Obras e Impuestos.
- ✓ Módulo de Presupuesto, Contabilidad y Tesorería
- ✓ Módulo de administración de sueldos y jornales.
- ✓ Módulo Mesa de Entrada.

Es posible encontrar más información del sistema y de la empresa desarrolladora en su página web. <http://www.municipalidad.com/>



Este sistema está siendo utilizado en más de 160 municipios lo cual brindaría la posibilidad de poder adaptar e implementar la solución de Inteligencia de Negocio de este proyecto de manera muy sencilla en cualquiera de estas jurisdicciones.





4.3.6 Relación de las metas de la Secretaría con las del DW

El DW coincide claramente con las metas de la Secretaría de Economía y Finanzas Públicas ya que ésta necesita mejorar su eficiencia en la toma de decisiones y contar con información detallada y oportuna a tal fin. El DW permitirá solucionar los problemas ocasionados por la falta de un sistema que sea capaz de integrar la información y brindarla de manera organizada, rápida y de fácil acceso a la dirección para la toma de decisiones.

El DW aportará un gran valor al área, entre las principales ventajas que brindará y los inconvenientes que solucionará se pueden mencionar los siguientes:

- Permitirá a los usuarios tener una visión general del negocio.
- Transformará datos operativos en información analítica, enfocada a la toma de decisiones.
- Se podrán generar reportes dinámicos, ya que actualmente son estáticos y no ofrecen ninguna facilidad de análisis.
- Soportará la estrategia de la Secretaría de Economía y Finanzas Públicas.
- Aportará a la mejora continua de la estructura de la organización.
- Los empleados de la Secretaría dejarán de aplicar su criterio individual para realizar el procedimiento de extraer, depurar, ordenar, organizar y clasificar los datos que obtienen desde Excel para desarrollar los informes gerenciales. El DW estandarizará y organizará el proceso mediante el cual la información es recolectada, extraída, transformada y brindada a la gerencia, haciéndolo de manera eficiente, eficaz, confiable y amigable para una fácil lectura que permita una toma de decisiones oportuna.
- Se eliminará el tiempo desperdiciado al realizar el procedimiento actual, debido a que el mismo es efectuado de forma manual generando un costo innecesario en tiempo y recursos.
- Se desarrollarán nuevos informes que anteriormente no se efectuaban por falta de tiempo o capacidad.
- Los resultados obtenidos serán exactos.

4.3.7 Procesos

La Secretaría de Economía y Finanzas Públicas tiene dentro de sus tareas: la elaboración del proyecto de presupuesto municipal, fijación de la política tributaria, percepción de los recursos municipales y régimen de pagos y el análisis, evaluación y control de la ejecución presupuestaria.

Tiene a su cargo la conducción de la Tesorería, el régimen de pagos y la verificación, centralización y conducción de la información sobre el endeudamiento público de la Municipalidad, así como las gestiones necesarias para obtener financiamiento y crédito ante instituciones financieras oficiales o privadas del ámbito provincial, nacional o internacional. Es también función de esta secretaría convenir planes de amortización, intereses y demás condiciones relacionadas con la obtención de la financiación respectiva.

4.4 Modelado del Data Warehouse



Una vez descripta la organización, se continuará con el desarrollo práctico. En ésta etapa basándose en los primeros pasos de la metodología HEFESTO se realizará el modelado de la estructura que contendrá el Data Warehouse.

4.4.1 Análisis de Requerimientos

En esta primera parte se identifican los requerimientos de los usuarios, para lo cual previamente es necesario definir cuáles son los actores intervinientes dentro del ámbito del proyecto. Considerando que el alcance del proyecto se delimitó a la Secretaría de Economía y Finanzas Públicas, los actores principales que abarca el trabajo de grado son:

- Intendente.
- Secretario de Economía y Finanzas Públicas.
- Director de Recursos Tributarios.
- Director de Tesorería y Contaduría.

4.4.1.1 Identificar preguntas

El primer paso comienza con el acopio de las necesidades de información, para el cual existen diferentes formas y técnicas para llevarlo adelante. En el trabajo se han utilizado como fuentes de información, entre otras, la recolección de documentos, entrevistas y observaciones. El resultado de este punto son una serie de preguntas que relevan las necesidades de información de los procesos principales que desarrolla la Secretaria de Economía y Finanzas Públicas. La idea central es que se formulen preguntas complejas sobre el negocio las cuales incluyan variables de análisis que se consideren relevantes, ya que son las que permitirán estudiar la información desde diferentes perspectivas.

El resultado de este primer paso fueron una serie de preguntas las cuales fueron agrupadas utilizando un criterio por proceso:

Caja:

- ¿Cómo está distribuido el ingreso, según la forma de pago (efectivo, tarjeta, cheques, otros) por día, mes, trimestre y año?

Contabilidad:

- ¿Cómo va la recaudación respecto de años anteriores discriminada por concepto, mes, trimestre y año?
- ¿Cómo se comporta el ahorro corriente por mes, trimestre y año?
- ¿Cuál es la autonomía financiera por mes, trimestre y año?
- ¿Cómo se comporta el gasto en personal respecto del gasto total por mes, trimestre y año?
- ¿Cuál es la inversión física (obra pública y bienes de capital) respecto del gasto total acumulado en el periodo, por mes, trimestre y año?

Rentas:

- ¿Cómo se comportaron las promociones (pago contado, cumplidor, etc.) por tipo de bien y año?



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

- ¿Cómo es el estado de deuda tributaria por barrio, zona, mes, trimestre y año?

Una vez identificadas las preguntas, lo que se debe hacer es redactarlas de otro modo para facilitar su descomposición. En este caso:

Caja:

- ¿Cómo está distribuido el ingreso, según la forma de pago (efectivo, tarjeta, cheques, otros) por día, mes, trimestre y año?

Monto total y porcentaje de caja según forma de pago, día, mes, trimestre y año.

Contabilidad:

- ¿Cómo va la recaudación respecto de años anteriores discriminada por concepto, mes, trimestre y año?

Monto total y acumulado de recaudación por concepto, mes, trimestre y año.

- ¿Cómo se comporta el ahorro corriente por mes, trimestre y año?

Tasa de ahorro corriente por mes, trimestre y año.

- ¿Cuál es la autonomía financiera por mes, trimestre y año?

Autonomía financiera por mes, trimestre y año.

- ¿Cómo se comporta el gasto en personal respecto del gasto total por mes, trimestre y año?

Gasto en personal respecto del gasto total por mes, trimestre y año.

- ¿Cuál es la inversión física (obra pública y bienes de capital) respecto del gasto total acumulado en el periodo, por mes, trimestre y año?

Inversión productiva por mes, trimestre y año.

Rentas:

- ¿Cómo se comportaron las promociones (pago contado, cumplidor, etc.) por tipo de bien y año?

Monto total de descuentos por tipo de promoción, tipo de bien y año.

- ¿Cómo es el estado de deuda tributaria por barrio, por zona, por mes, trimestre y año?



Porcentaje de deuda tributaria según barrio, zona, mes, trimestre y año.

4.4.1.2 Identificar indicadores y perspectivas

Los indicadores son, en general, valores numéricos y representan lo que se desea analizar concretamente, en cambio, las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas. Cabe destacar que el tiempo es comúnmente una perspectiva. A continuación se identifican con color celeste los indicadores y con color verde las perspectivas.

Caja:

Monto total y porcentaje de caja según forma de pago, día, mes, trimestre y año.

Contabilidad:

Monto total y acumulado de recaudación por concepto, mes, trimestre y año.

Tasa de ahorro corriente por mes, trimestre y año.

Autonomía financiera por mes, trimestre y año.

Gasto en personal respecto del gasto total por mes, trimestre y año

Inversión productiva por mes, trimestre y año.

Rentas:

Monto total de descuentos por tipo de promoción, tipo de bien y año.

Porcentaje de deuda tributaria según barrio, zona, mes, trimestre y año.

Habiendo marcado con color celeste a los indicadores y verde a las perspectivas, las resumimos como:

Los indicadores son:

- Monto Total caja.
- Porcentaje caja.
- Monto total recaudado.
- Acumulado recaudado.
- Porcentaje tasa de ahorro corriente.
- Porcentaje de autonomía financiera.
- Porcentaje del gasto en personal respecto del gasto total.
- Porcentaje de inversión productiva respecto del gasto total.
- Monto total de descuento en promociones.
- Porcentaje deuda tributaria.

Las perspectivas son:

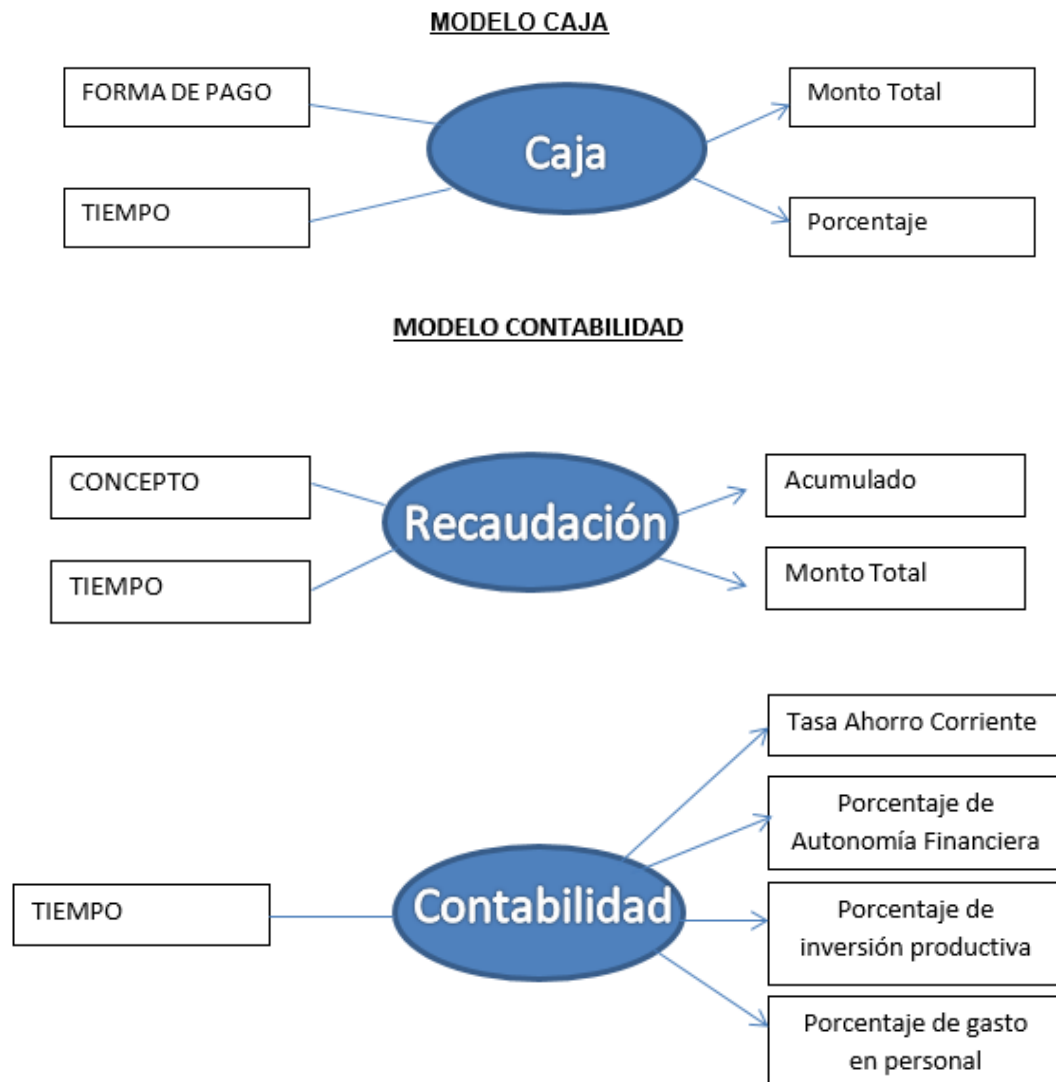
- Tiempo
- Forma de pago
- Concepto

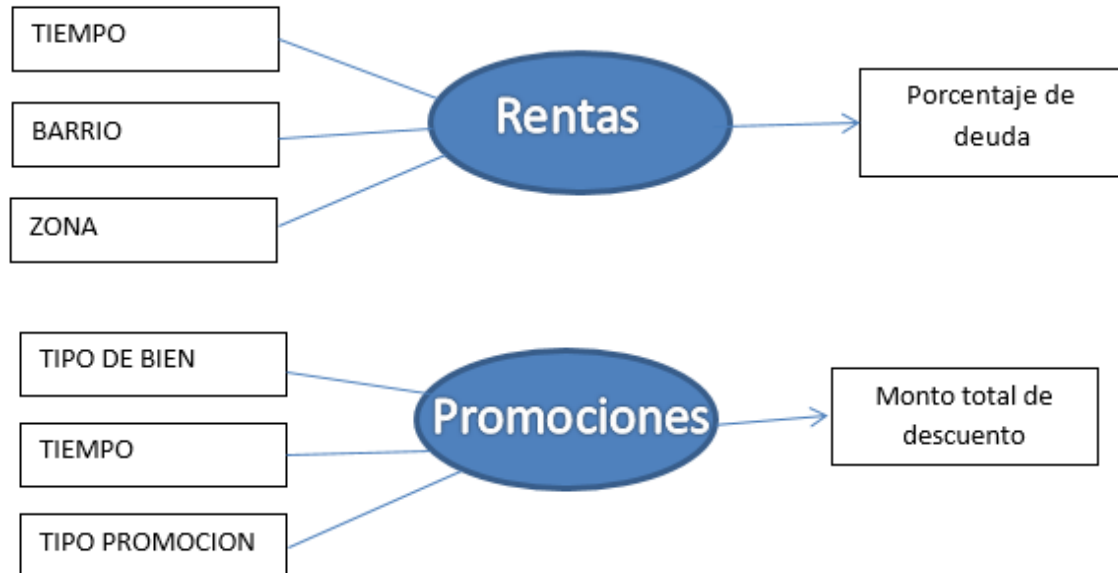


- Barrio
- Zona
- Tipo de Bien
- Tipo Promoción

4.4.1.3 Modelo Conceptual

A partir de los indicadores y perspectivas obtenidas en el paso anterior se creará el modelo conceptual, a partir del cual se podrá observar con claridad cuáles son los alcances del proyecto. En el modelo conceptual se unen las perspectivas con los indicadores a través del proceso o área de estudio. Se utilizará el criterio definido en el paso anterior para unir las perspectivas con los indicadores.



**MODELO RENTAS****4.4.2 Análisis de los OLTP**

Seguidamente son analizadas las fuentes OLTP para determinar cómo serán calculados los indicadores y para establecer las respectivas correspondencias entre el modelo conceptual creado y las fuentes de datos.

En el proyecto se usa como origen de datos la base operacional que está siendo utilizada en la actualidad en la Municipalidad de Cosquín denominada PROGRAM y desarrollada por la empresa Program Consultores. Es necesario considerar que la base de datos abarca todas las áreas administrativas del municipio por lo cual es prácticamente imposible poder mostrar el DER (diagrama entidad-relación) completo, por ello solamente se presentarán diagramas entidad-relación parciales que contienen las tablas de interés.

4.4.2.1 Conformar indicadores

En este paso se deberán explicitar cómo se calcularán los indicadores.

Caja:

- Monto Total = $SUM(IMP_PAGO - IMP_VUELTO)$
- Porcentaje = $MontoTotalFormaPago / Monto\ Total$

Contabilidad:

- Monto Total = $SUM(Monto_Cobrado)$
- Monto Total Acumulado = $SUM(Monto_Cobrado_Acu)$
- Tasa Ahorro Corriente = $(IngresoCorriente - Gasto\ Corriente) / IngresoCorrientes$
- Porcentaje de Autonomía Financiera = $(IngresoJurMun / GastosCorrientes)$



- Porcentaje de inversión productiva = $InversionFisica / GastoTotal$
- Porcentaje de gasto en personal = $GastoPersonal / GastoTotal$

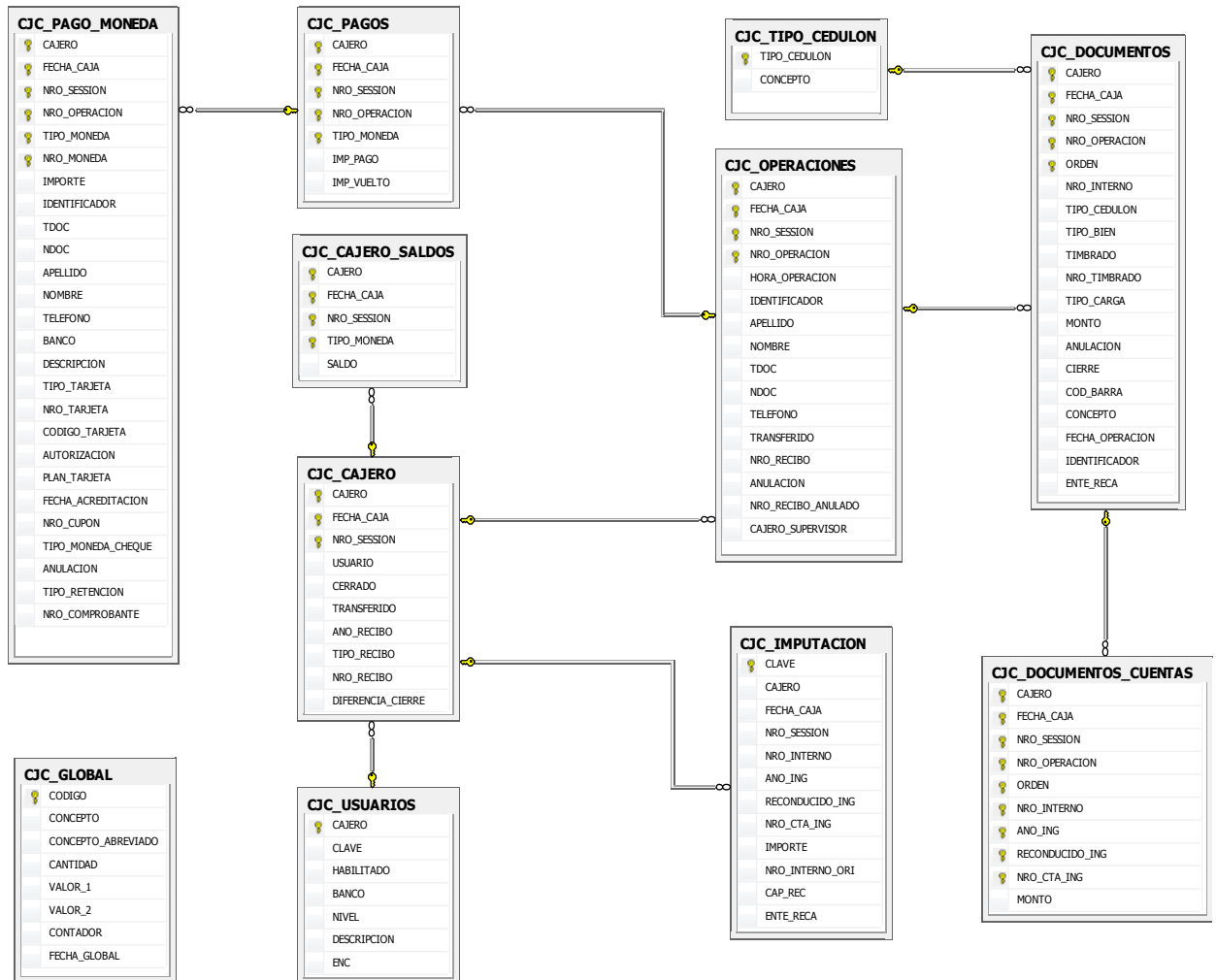
Rentas:

- Porcentaje de deuda = $Count (Cuentasaldía) / Count (Cuentas)$
- Monto total de descuento = $SUM (Capital_Item)$

4.4.2.2 Establecer correspondencias

El objetivo de este paso es el de examinar los OLTP disponibles que contengan la información requerida, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos. Como se indicó, el origen de datos a utilizar es la base de datos PROGRAM, a continuación se irán exponiendo los DER parciales que contienen las tablas relacionadas a los distintos modelos conceptuales.

Para el modelo de Caja, las tablas del OLTP relacionadas son:



Las relaciones identificadas fueron:

La tabla “CJC_Global” se relaciona con la perspectiva “Forma de pago”.

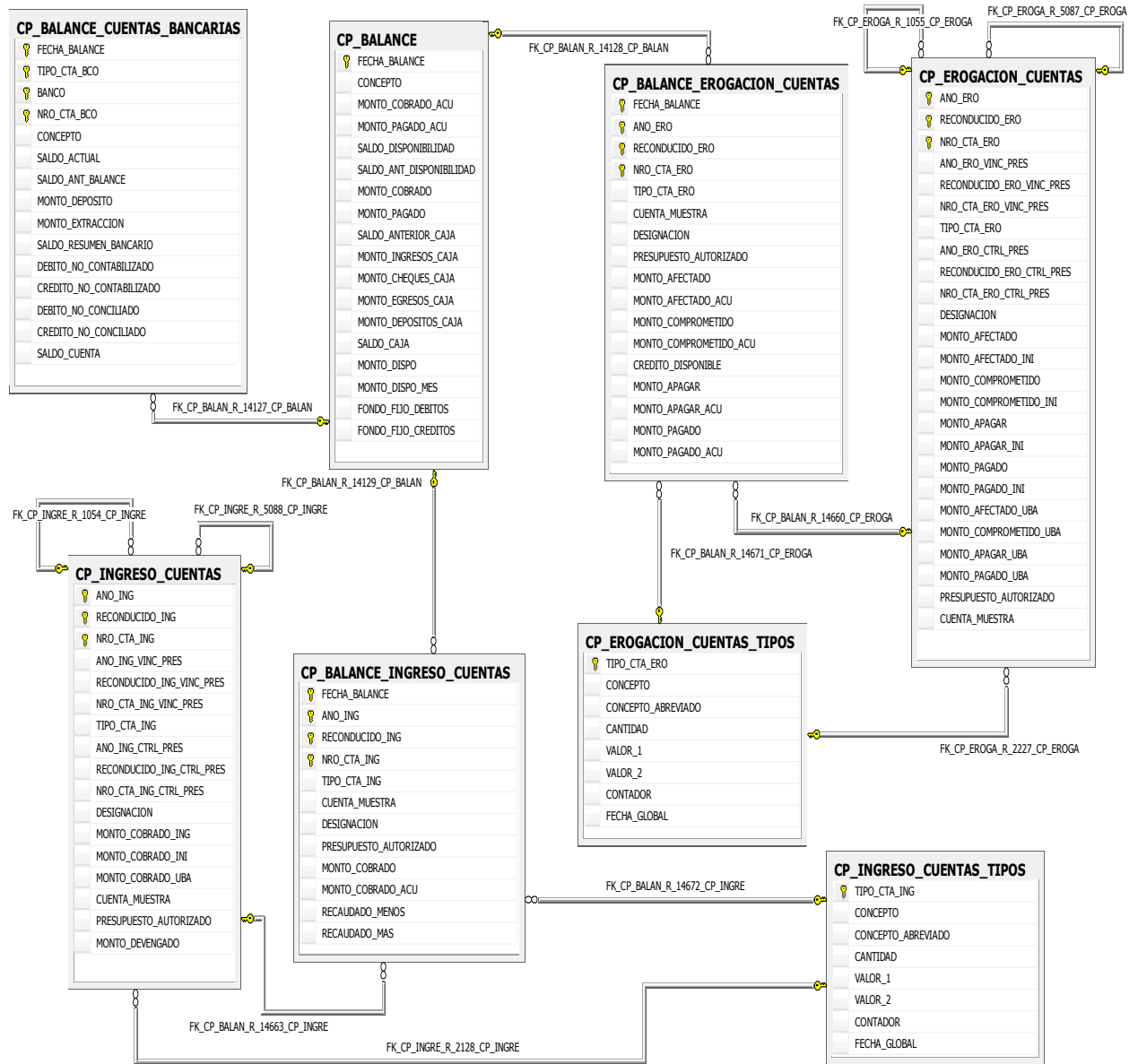


Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

El campo “IMP_PAGO” de la tabla “CJC_PAGOS” menos el campo “IMP_VUELTO” de la misma tabla se relaciona con el indicador “Monto Total”.

El indicador “Monto Total” dividido por la sumatoria de los totales respectivos.

El modelo de Contabilidad tiene como tablas relacionadas:



Las relaciones identificadas fueron:

La tabla “CP_Balance_Ingreso_Cuenta” se relaciona con la perspectiva “Concepto”.

El campo “Monto_Cobrado” de la tabla “CP_Balance_Ingreso_Cuenta” se relaciona con el indicador “Monto Total”.

El campo “Monto_Cobrado_Acu” de la tabla “CP_Balance_Ingreso_Cuenta” se relaciona con el indicador “Acumulado”.

El campo “Monto_Cobrado” de la tabla “CP_Balance_Ingreso_Cuenta” menos el campo “Monto_Pagado” de la tabla “CP_Balance_Erogaciones_Cuenta”, dividido el



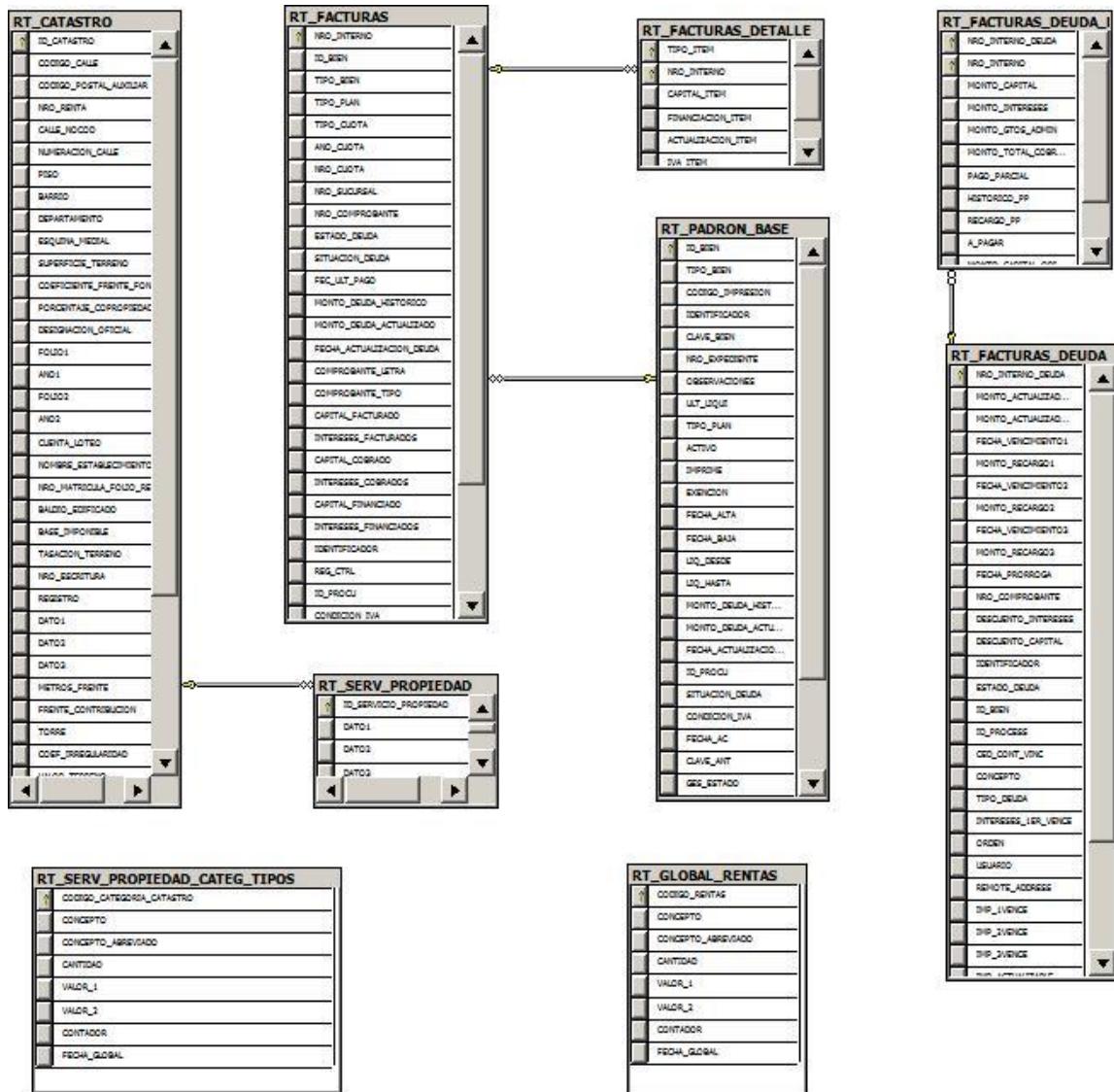
campo “Monto_Cobrado” ya mencionado; se relaciona con el indicador “Tasa de ahorro corriente”.

El campo “Monto_Cobrado” de la tabla “CP_Balance_Ingreso_Cuenta” dividido campo “Monto_Pagado” de la tabla “CP_Balance_Erogaciones_Cuenta” se relaciona con el indicador “Grado de autonomía financiera”.

El campo “Monto_Pagado” de la tabla “CP_Balance_Erogaciones_Cuenta” se relaciona con el indicador “Inversión productiva”.

El campo “Monto_Pagado” de la tabla “CP_Balance_Erogaciones_Cuenta” se relaciona con el indicador “Gasto en personal”.

Finalmente el modelo de Rentas, tiene como las tablas relacionadas:



Las relaciones identificadas fueron:

La tabla “RT_SERV_PROPIEDAD_CATEG_TIPOS” se relaciona con la perspectiva “Zona”.

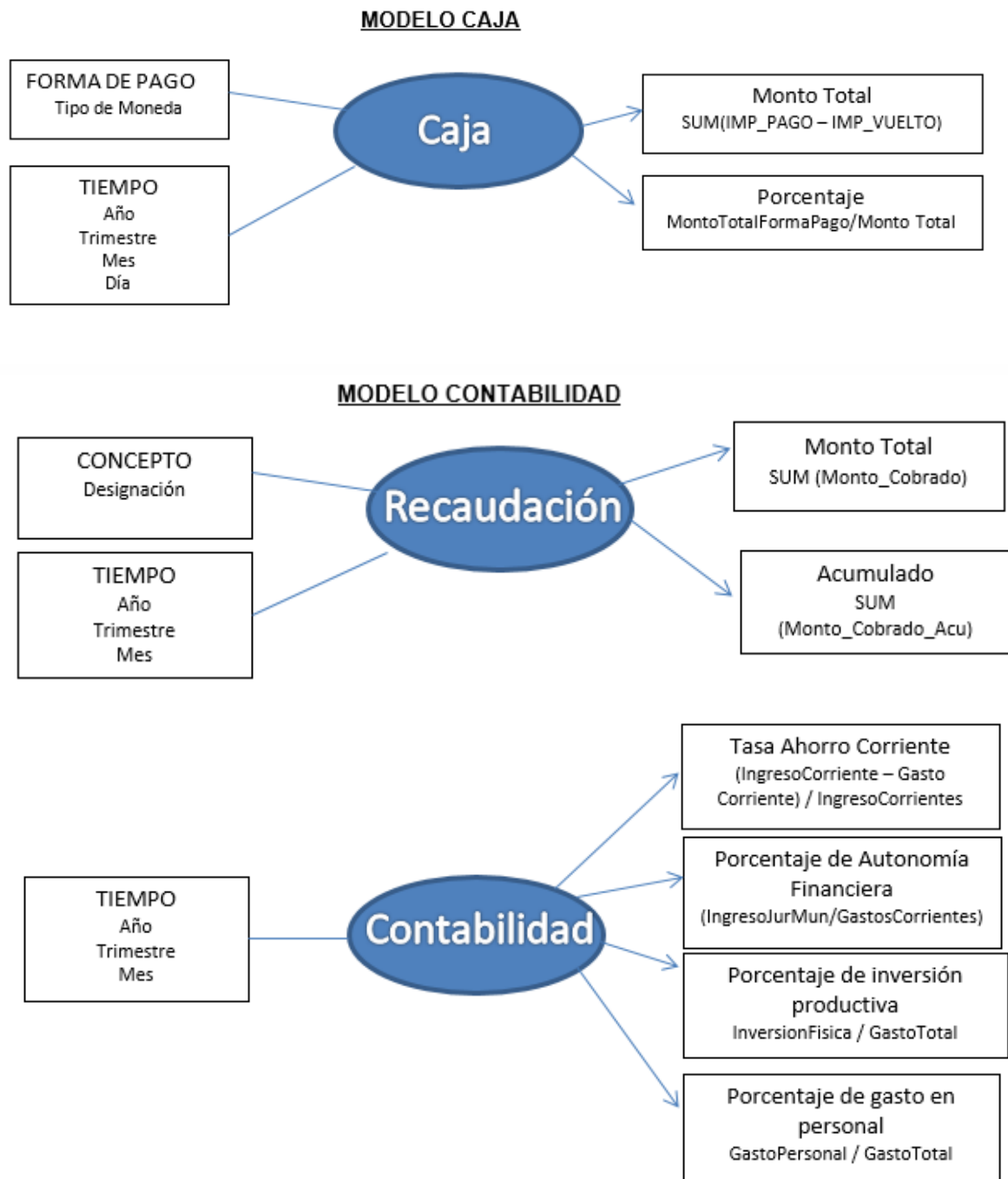
El campo “Barrio” de la tabla “RT_CATASTRO” se relaciona con la perspectiva “Barrio”.



La tabla “RT_GLOBAL_RENTAS” se relaciona con la perspectiva “Tipo_Bien”.
El campo “Monto_Intereses” de la tabla “RT_FACTURAS_DEUDA_DETALLE” se relaciona con el indicador “Monto total de descuento”.
El campo “ID_Bien” de la tabla “RT_FACTURAS” se relaciona con el indicador “Porcentaje de deuda”.

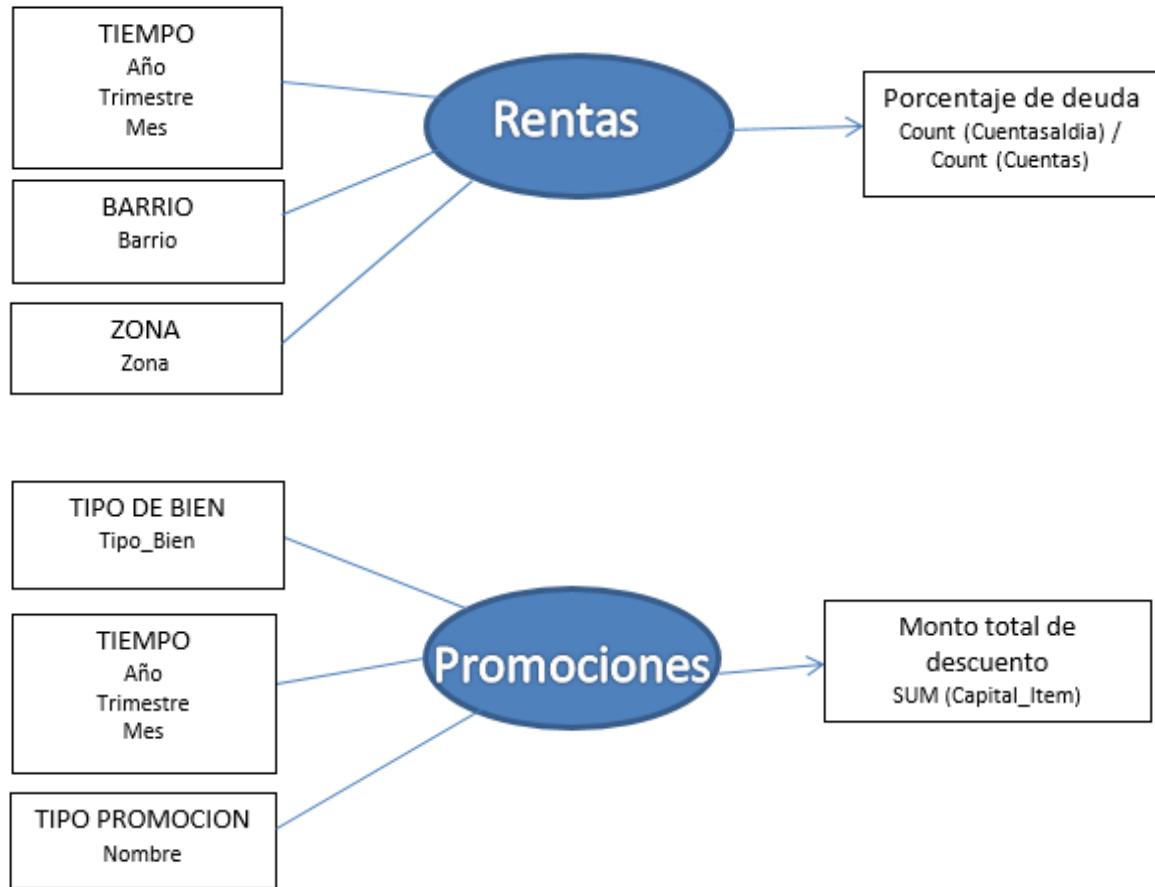
4.4.2.3 Modelo conceptual ampliado

En este paso se busca ampliar el modelo conceptual, colocando bajo cada perspectiva los campos seleccionados y bajo cada indicador su respectiva fórmula de cálculo.





MODELO RENTAS



4.4.3 Modelo Lógico del DW

En ésta etapa se construye el modelo lógico, el cual es una representación de una estructura de datos que puede procesarse y almacenarse en algún SGBD. Tomando como base el modelo conceptual creado en los puntos anteriores, se definirá el modelo a utilizar y luego se diseñaran las tablas de hechos y dimensiones. Por último se realizarán las uniones entre las tablas correspondientes.

4.4.3.1 Tipo de Modelo Lógico del DW

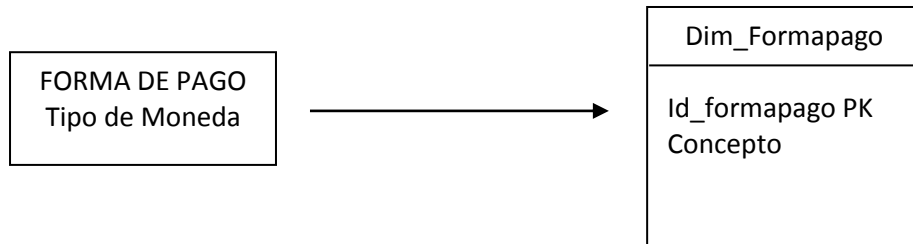
En éste punto se debe seleccionar el esquema que se utilizará para contener la estructura del depósito de datos. En el presente trabajo se ha seleccionado el esquema de tipo CONSTELACIÓN, considerando que se utilizarán varias tablas de hechos y dimensiones. Las características, ventajas y desventajas de dicho esquema fueron expuestas dentro del capítulo 3 punto 3.5.3.4.

4.4.3.2 Tablas de dimensiones

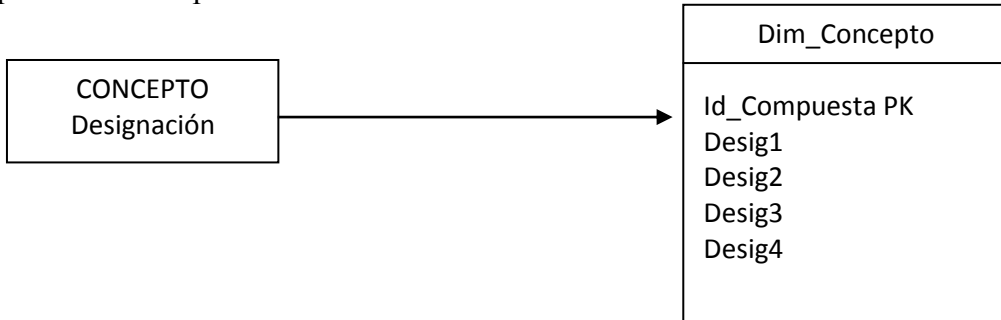
En este paso son diseñadas las tablas de dimensiones que formaran parte del DW. Cada perspectiva definida en el modelo conceptual constituirá una tabla de dimensión. Para ello se debe seleccionar un nombre que identifique la tabla de dimensión y añadir un campo que represente la clave principal.



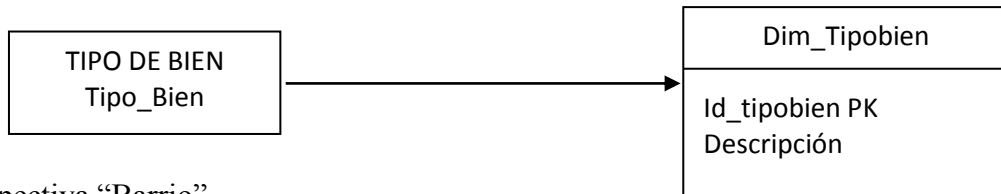
Perspectiva “Forma de Pago”:



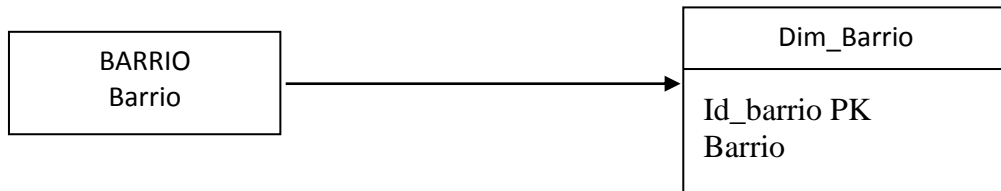
Perspectiva “Concepto”:



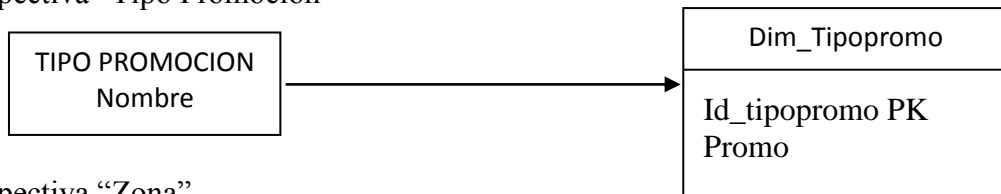
Perspectiva “Tipo de Bien”



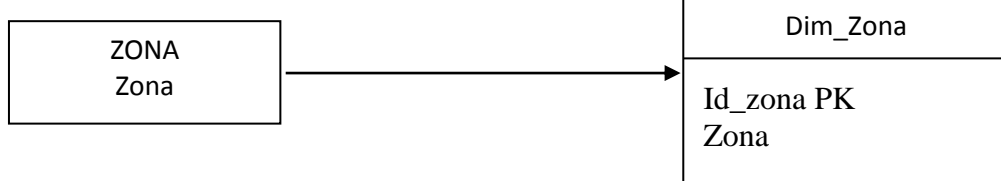
Perspectiva “Barrio”



Perspectiva “Tipo Promoción”

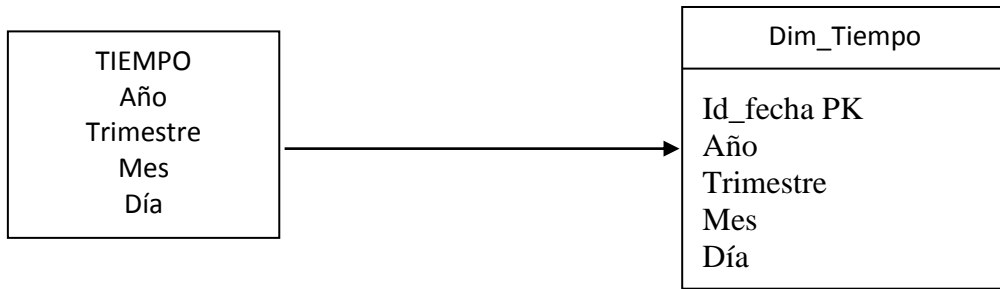


Perspectiva “Zona”





Perspectiva “Tiempo”



4.4.3.3 Tablas de hechos

En este paso se definirán las tablas de hechos, que son las que contendrán los hechos a través de los cuales se construirán los indicadores de estudio.

Tabla de hechos: FAC_CAJA

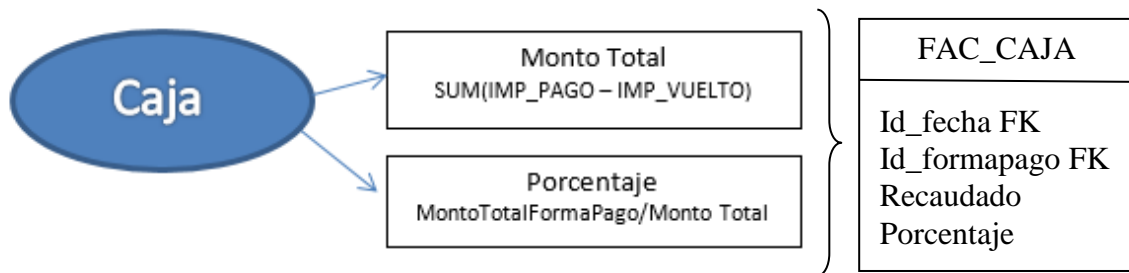


Tabla de hechos: FAC_RECAUDACION

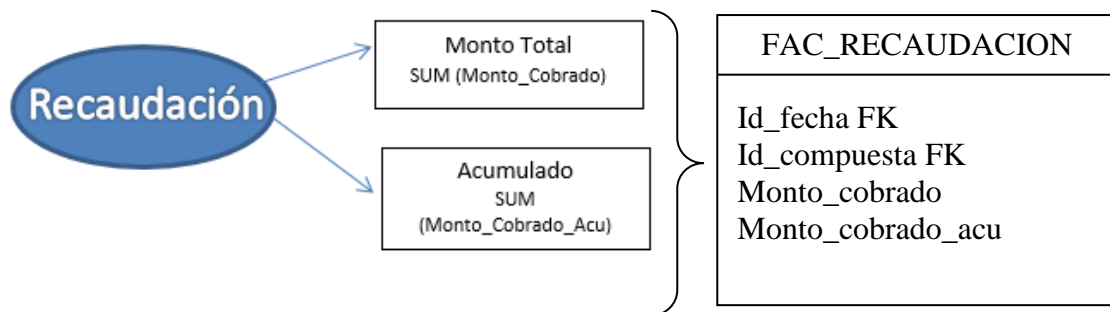




Tabla de hechos: FAC_CONTABILIDAD

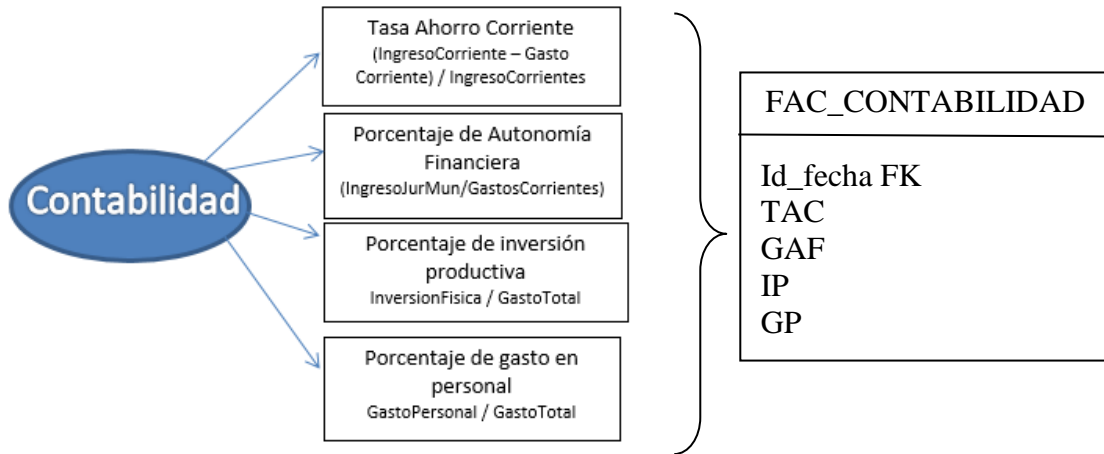


Tabla de hechos: FAC_BARRIOS_ZONAS

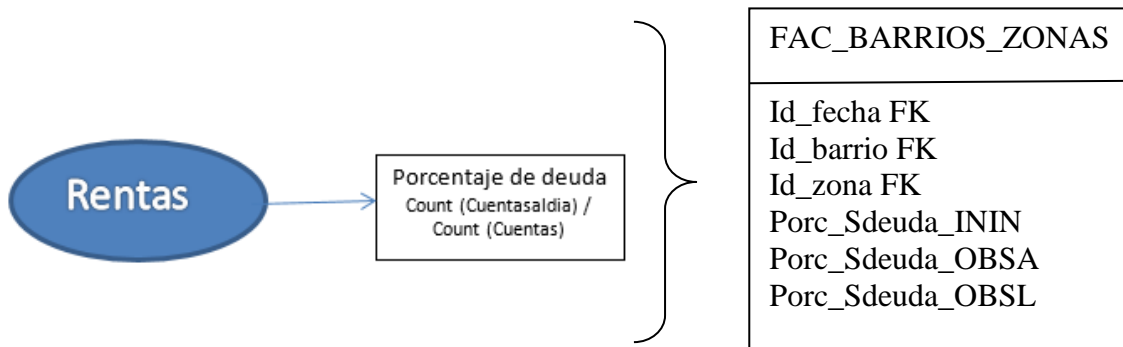
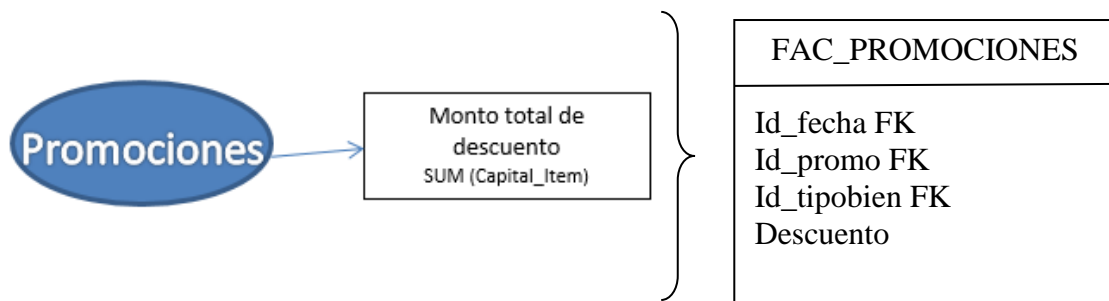


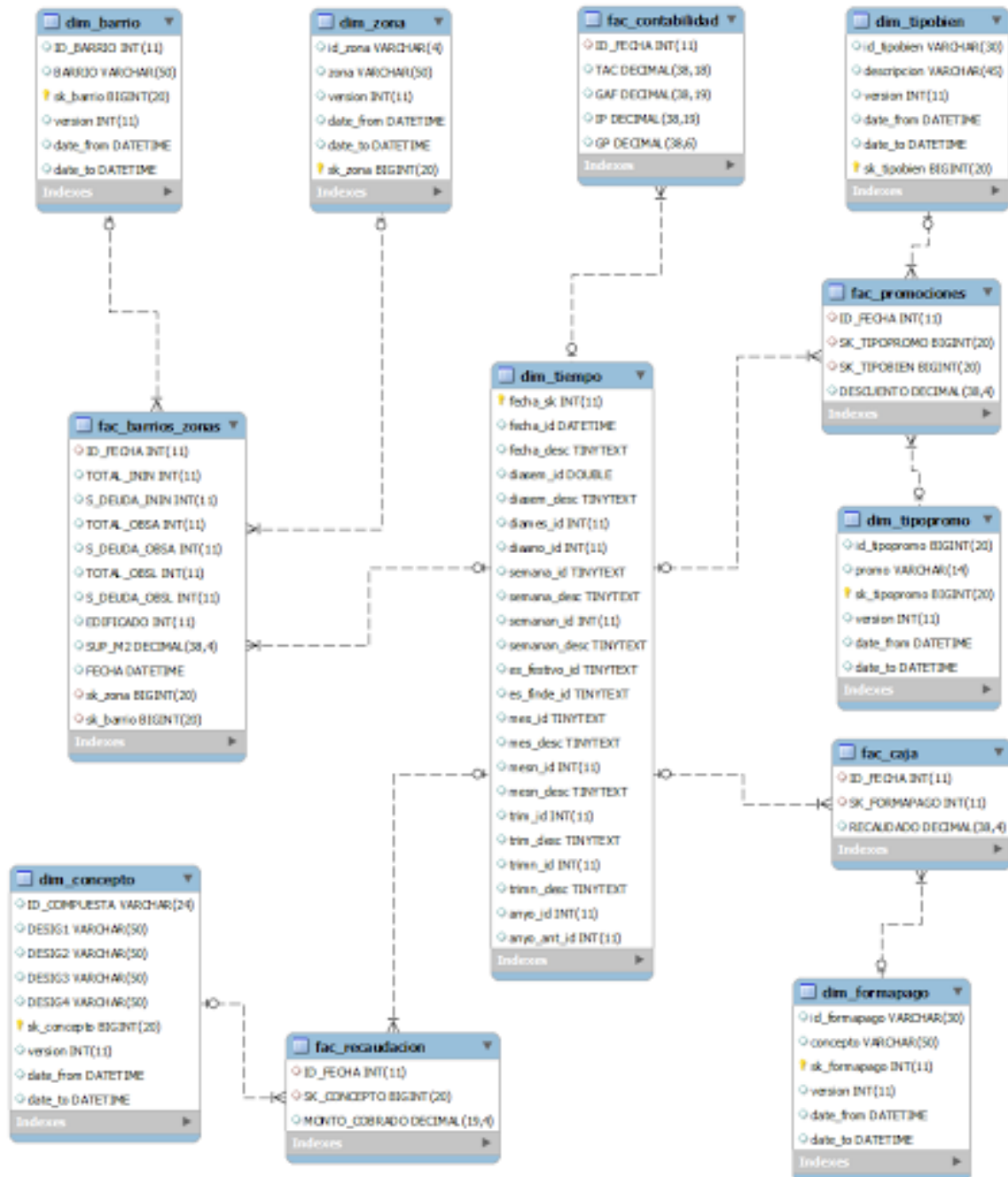
Tabla de hechos: FAC_PROMOCIONES



4.4.3.4 Uniones



En este paso se realizan las uniones entre las tablas de dimensiones y las de hechos. Es importante destacar que debido a que las tablas de hechos FAC_ZONAS y FAC_BARRIOS comparten los mismos indicadores, se ha creado una única tabla de hechos que contiene a las medidas con las claves de las dos dimensiones en su interior. El diagrama lógico quedara de la siguiente manera:



El diagrama lógico constituirá el Diagrama Entidad-Relación que contendrá la base de datos del Data Warehouse.

4.5 Selección del conjunto de herramientas



Una vez que se han completado las etapas y acciones necesarias para el modelado del sistema de Data Warehouse, es posible continuar con aquellas tareas que nos permitan elegir la mejor suite de herramientas de software para implementar el proyecto. Para ello será utilizando el método planteado en el Proceso Racional de Toma de Decisiones desarrollado por Kepner y Tregoe.

4.5.1 Análisis de Decisiones – Elección de la tecnología a utilizar

En esta etapa se aplicaran los pasos del “Análisis de decisión” expuestos por Kepner y Tregoe en el libro “El nuevo directivo racional” para hacer una correcta elección del conjunto de herramientas de software a utilizar para implementar el proyecto.

El proceso se compone de una serie de pasos que se realizarán a continuación.

4.5.1.1 Enunciado de la decisión o del propósito

Este paso brinda enfoque para todo lo que sigue y establece los límites de la elección. Como consecuencia de éste se definirán los criterios de selección, las alternativas y la elección final. Siempre indica algún tipo de acción, el resultado esperado y el nivel al que debe tomarse la decisión. Para el proyecto el enunciado de la decisión es:

“Seleccionar la mejor solución Business Intelligence para implementar el proyecto”

4.5.1.2 Los objetivos para la decisión

Son los criterios o los detalles específicos que debe cumplir la decisión. Hay 2 clases de objetivos:

- ✓ **Obligatorios:** son imprescindibles y deben cumplirse para garantizar una decisión exitosa. Estos deben ser cuantificables porque funcionan como un filtro para eliminar las alternativas propensas al fracaso (PASA O NO PASA).
- ✓ **Deseados:** las alternativas se juzgaran con base en sus resultados relativos de desempeño frente a los objetivos deseados. Se deberá ponderar los objetivos para saber cuál es la importancia relativa, de 10 (+ importante) a 1 (- importante).

“Los obligatorios deciden quien participa en el juego, pero los deseados deciden quién gana”

Objetivo Obligatorio:

- **La solución a implementar debe ser Open Source**

Se decidió que la solución a implementar sea Open Source por una serie de beneficios que se enumeran a continuación:

- ✓ Sin costes de licencias. No genera un gasto para la organización.
- ✓ Posibilidad de copiar, distribuir, modificar el software de acuerdo a nuestras necesidades sin ningún tipo de limitación.



- ✓ Comunidad de usuarios. A diferencia de un software privativo que cuenta con una cierta cantidad de usuarios, la comunidad de usuarios de software libre es ilimitada en cuanto a cantidad, conocimiento y dominio. Se forma por personas que se encuentra en universidades o personas que desarrollan desde su propia casa y desean compartir el conocimiento y proyectos.
- ✓ Corrección de errores y bugs más rápida. Tiene que ver con la comunidad de usuario y la cantidad de miles de personas que trabajan sobre un software y esto hace que los errores se difundan, reporten y corrijan de manera más rápida.
- ✓ Independencia con el proveedor del software.
- ✓ Entorno de desarrollo flexible y descentralizado.
- ✓ Posibilidad de elegir entre múltiples opciones. En general se pueden conseguir muchas versiones de una misma aplicación a las cuales se le puede acceder y leer el código fuente pudiendo seleccionar mejor el software que se necesita para determinado proyecto.
- ✓ Mayor adaptabilidad a entornos.
- ✓ Tendencia a la alta calidad del código. Esto se da porque al existir tantas personas en la comunidad cuando un usuario se decide a compartir su código queda expuesto a ser leído por miles lo cual genera que aquel que quiera que otros descarguen su código se asegurará de que tenga una excelente calidad. En los softwares enlatados no es posible conocer de qué modo este desarrollado el código.
- ✓ Posibilidad de innovación y desarrollo tanto desde la persona que lo crea como así como aquella que lo utiliza.
- ✓ Reduce la brecha digital. Al no existir un coste de licencia permite que muchas empresas de pequeño tamaño accedan a software libre de alta calidad y acorta la brecha de competencia entre aquellas empresas grandes que tienen los recursos para adquirir un software privativo contra las pequeñas que no los tienen.
- ✓ Imagen de la empresa.

Objetivos Deseados:

La herramienta debe brindar la capacidad de:

- ✓ Análisis - Calidad de las herramientas disponibles
- ✓ Reporting- Calidad de las herramientas disponibles
- ✓ Data Integration- Calidad de las herramientas disponibles
- ✓ Dashboard- Calidad de las herramientas disponibles
- ✓ Data Mining- Calidad de las herramientas disponibles

Los criterios serán ponderamos mediante una tabla de 10 puntos:

Criterio	Ponderación
Análisis	10
Reporting	10



Data Integration	10
Dashboard	8
Data Mining	6

4.5.1.3 Generación y evaluación de las alternativas

Teniendo en cuenta que el objetivo obligatorio es que la solución debe ser Open Source se consideran como alternativas para implementar el Data Warehouse únicamente las tecnologías OSBI líderes en el mercado, las cuales son:

Soluciones Open Source

Alternativa 1: PENTAHO



De todas las iniciativas, la más completa es la de Pentaho, que está compuesta por una serie de componentes Open Source muy utilizados y fiables: Mondrian OLAP server, JPivot, Firebird RDBMS, Enhrydra ETL, Shark and JaWE, JBoss, Hibernate, JBoss Portal, Weka Data Mining, Eclipse, BIRT, JOSSO, Mozilla Rhino.

Las características más importantes de Pentaho son:

- ✓ Aplicación web desarrollada en J2EE. Modular.
- ✓ Multiplataforma.
- ✓ Incluye: ETL, Reporting, OLAP, Data mining, Dashboards, etc.
- ✓ Licencia: GNU GPL.

Pentaho es una compañía que ofrece una alternativa Open Source en productos de Inteligencia de Negocios. Está formada por una suite de herramientas que permiten realizar: reportes comprensivos, análisis OLAP, dashboards, integración de datos, minería de datos. Es una plataforma que se ha convertido en la más implementada en el mundo dentro de las suites Open Source.

El modelo de negocios Open Source de Pentaho, elimina los costos de licencia de software, provee soporte, servicios, y una optimización del producto vía una suscripción anual. Desde el inicio de Pentaho como pionero en las herramientas BI Open Source, sus productos han sido descargados más de tres millones de veces con implementación a producción en compañías en el rango desde pequeñas a las globales.

La misión de Pentaho es el crecimiento de la innovación, alta calidad de tecnología y soporte profesional en el mercado de Business Intelligence. Pentaho usa un revolucionario enfoque para el desarrollo, distribución y soporte hecho posible por el modelo de negocios Open Source. Su tecnología soporta un amplio rango de iniciativas de negocios para ventas, análisis de rentabilidad, análisis de clientes, reporte HR,



reportes financieros, herramientas KPI, cadena de suministro de análisis y reportes operacionales.

Con sus herramientas es posible realizar:

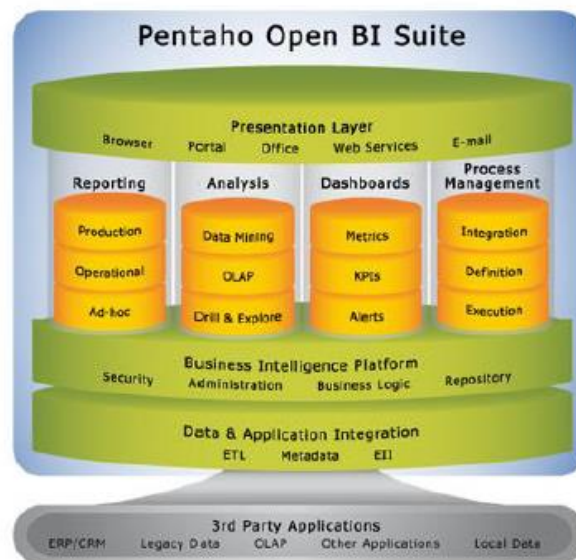
- ✓ Reporting
- ✓ Análisis
- ✓ Dashboards
- ✓ Integración de datos
- ✓ Data Mining
- ✓ Plataforma Business Intelligence

Pentaho Open BI suite, provee un completo espectro de capacidades de la Inteligencia de Negocios que la hacen una opción idónea para las necesidades de cualquier organización. Sus productos son usados por organizaciones líderes como:

- MySQL
- Motorola
- Terra Industries
- DivX, y más.

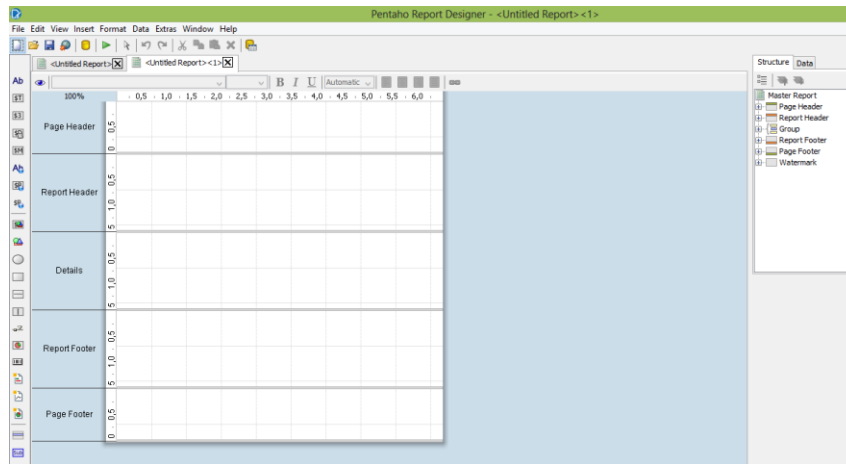
El Proyecto Pentaho BI: La corporación Pentaho es el patrocinador primario y propietario del proyecto Pentaho BI. El Pentaho BI es una iniciativa en curso por la comunidad de Open Source que provee organizaciones con mejores soluciones para las necesidades de BI de su empresa.

El proyecto Pentaho BI abarca las siguientes áreas:



- ✓ Pentaho Reporting

Solución proporcionada por Pentaho e integrada en su suite para el desarrollo de informes



Existen tres productos con diferentes enfoques y dirigidos a diferentes tipos de usuarios.

Pentaho Report Designer: Editor basado en eclipse con prestaciones profesionales y de calidad y con capacidad de personalización de informes a las necesidades de negocio destinado a desarrolladores. Sus características técnicas son:

- Incluye asistentes para facilitar la configuración de propiedades.
- Está estructurado de forma que los desarrolladores pueden acceder a sus prestaciones de forma rápida.
- Incluye un editor de consultas para facilitar la confección de los datos que serán utilizados en un informe.

Pentaho Report Design Wizard: Herramienta de diseño de informes, que facilita el trabajo y permite a los usuarios obtener resultados de forma inmediata. Está destinada a usuarios con menos conocimientos técnicos.

A través de pasos sencillos permite:

- Conectarse a todo tipo de bases relacionales
- Integrar el resultado dentro del portal Pentaho
- Posibilidad de montar codificación semafórica

Web ad-hoc reporting: Es similar a la herramienta anterior pero vía web. Extiende la capacidad de los usuarios finales para la creación de informes a partir de plantillas pre configuradas y siguiendo un asistente de creación.

Las características generales son:

- Proporciona funcionalidad crítica para usuarios finales como:
 - Acceso vía web
 - Informes parametrizados
 - Scheduling
 - Suscripciones
 - Distribución (bursting)
- Proporciona claras ventajas a especialistas en informes:



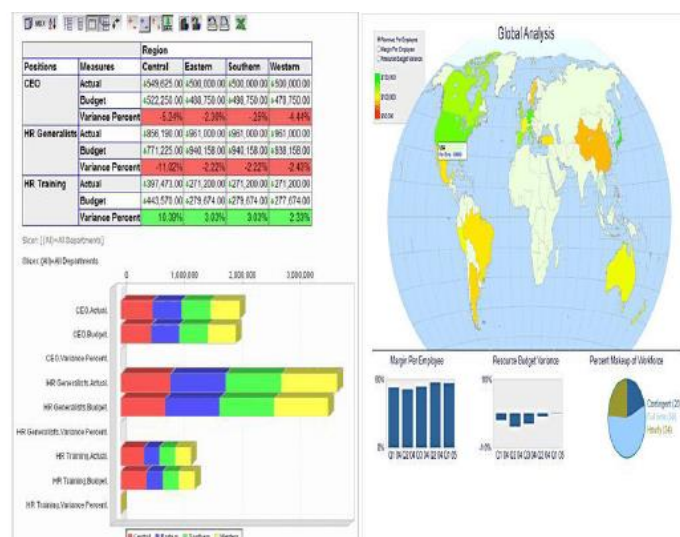
Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

- Acceso a fuentes de datos heterogéneos: relacional (vía jdbc), OLAP, XML, transformaciones de pentaho data integration.
- Capacidad de integración en aplicaciones o portales: jsp, portlet, web service.
- Definición modular de informes (distinción entre presentación y consulta)
- Diseño de informes flexible
 - Entorno de diseño gráfico
 - Capacidad de uso de templates
 - Acceso a datos relacionales, OLAP y XML
- Desarrollado para:
 - Ser embebible
 - Ser fácil de extender
 - No consumir muchos recursos
 - 100% Java: portabilidad, escalabilidad e integración.
- Multiplataforma (tanto a nivel de cliente como servidor): mac, linux/unix y Windows.
- ✓ Pentaho Analysis

Ayuda a operar con máxima efectividad para ganar perspicacia y entender lo necesario para tomar optimas decisiones.

Las características generales son:

- Vista dimensional de datos (por ventas, por periodo).
- Navegar y explorar
 - Análisis Ad Hoc
 - Drill-down
 - Seleccionar un específico miembro para el análisis.
- Interactuar con alto rendimiento
 - Tecnología optimizada para rápida respuesta interactiva



- ✓ Pentaho Dashboards



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

Provee inmediata sagacidad en un rendimiento individual, departamental o empresarial. Para deliberar “key metrics” es una atractiva e intuitiva interfaz visual, Pentaho Dashboards brinda a los usuarios de los negocios información crítica que necesitan para entender y mejorar el rendimiento organizacional.

- Identificación de unas Métricas Clave (KPI's, Key Performance Indicators)
 - Monitoreo/Métricas
- Investiga detalles subyacentes
 - Drill a reportes de soporte
- Seguimiento de excepciones
 - Alertas basadas en reglas del negocio.



✓ Pentaho Analysis Services: Mondrian Project



En la plataforma Open Source de PENTAHO se incluye la herramienta Mondrian, la cual es un servidor OLAP escrito en java que permite crear cubos de información para el análisis multidimensional.

Dichos cubos se componen de archivos XML y en ellos se definen las dimensiones y las conexiones de los datos. Los archivos XML por lo general son complejos de realizar manualmente por lo que es común utilizar herramientas gráficas para realizar la edición de estos.

La arquitectura de Mondrian se ejecuta sobre un servidor web y su tarea es lograr la comunicación entre aplicaciones OLAP con bases de datos. El núcleo del servidor es similar a JDBC pero exclusivo para OLAP. Proporciona la conexión a la base de datos y permite analizar de forma interactiva datos muy grandes en bases de datos SQL sin necesidad de escribir sentencias SQL, ya que Mondrian se encarga de ejecutarlas.

Mondrian es un motor OLAP que contiene 4 capas:

1. Capa de presentación.



2. Capa dimensional.
3. Capa estrella.
4. Capa de Almacenamiento

Capa de presentación: Esta capa es la encargada de determinar cómo los usuarios finales verán la información y de establecer cómo se realizarán las nuevas consultas.

Es con esta capa con la que los usuarios interactuarán para ejecutar nuevas consultas y para visualizar los resultados de las mismas. Es decir, los usuarios a través de las opciones disponibles en esta capa especifican qué quieren visualizar, utilizando dichas especificaciones se genera la consulta MDX la cual es enviada a la capa dimensional. En resumen esta capa presenta la información y permite interactuar con la misma. La información puede ser presentada de diferentes formas.

Capa dimensional: Esta capa es la que recibe las consultas MDX enviadas por la capa de presentación. Mediante la utilización de la metadata descrita en el modelo dimensional, esta capa realiza las siguientes tareas:

- analiza y valida la consulta MDX.
- establece cuáles son las consultas SQL que deben ejecutarse, y envía la consulta a la capa estrella.

Capa estrella: La capa estrella es quien se encarga de mantener la caché multidimensional de agregaciones*1.

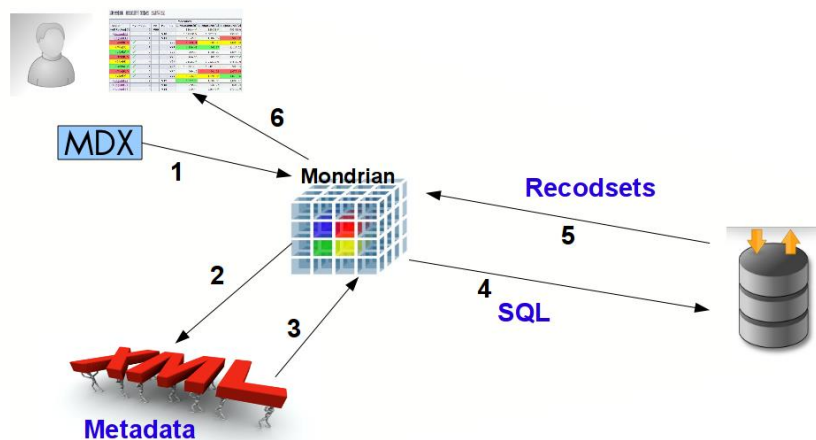
Cuando la capa dimensional le pide datos a esta capa, puede suceder lo siguiente:

- si los datos requeridos se encuentran en caché son devueltos inmediatamente.
- si los datos requeridos no se encuentran en caché, se deriva el pedido a la capa de almacenamiento y luego, al obtener los resultados los guarda en la caché y los retorna a la capa dimensional.

4. Capa de almacenamiento

Esta capa es la que tiene acceso a los SGBDR (SGBD Relacional) y básicamente se encarga de ejecutar consultas SQL y devolver los resultados obtenidos a la capa estrella.

En la siguiente imagen se puede apreciar cuál es el funcionamiento básico del motor Mondrian:





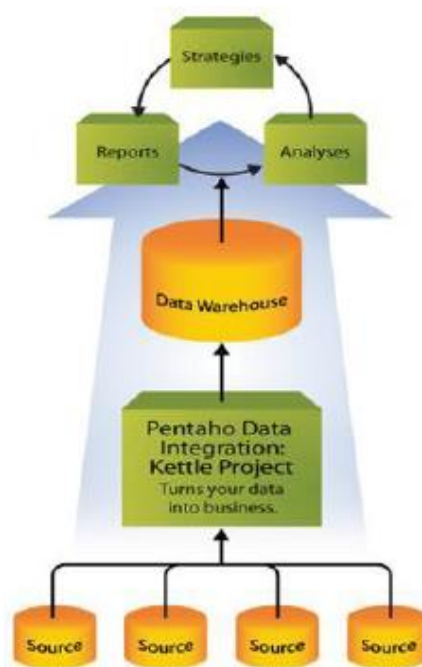
- En el paso 1 las herramientas de consulta y análisis generan una consulta MDX en base a lo que los usuarios desean visualizar.
 - En el paso 2 se utiliza la metadata para analizar esta consulta MDX y para posteriormente, en el paso 3 generar las consultas SQL correspondientes al pedido inicial.
 - Si los datos requeridos se encuentran almacenados en caché, estos son devueltos inmediatamente, en caso contrario se procede al paso 4.
 - El paso 4 consta de enviar las consultas SQL al SGBDR correspondiente.
 - En el paso 5 se devuelven los datos obtenidos luego de la ejecución de las consultas SQL en el SGBDR.
 - Finalmente, en el paso 6, los datos, ya con formato multidimensional, se devuelven a las herramientas de consulta y análisis, para que estas muestren los resultados esperados al usuario final.
- ✓ Pentaho Data Integration



Muchas organizaciones tienen información disponible en aplicaciones y base de datos separados. Pentaho Data Integration abre, limpia e integra esta valiosa información y la pone en manos del usuario. Provee una consistencia, una sola versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones TI hoy en día. Pentaho Data Integration es una poderosa herramienta de ETL (Extracción, Transformación y Carga), aunque no es solo eso sino que permite realizar muchísimas más cosas.

En la actualidad es conocido como PDI o Kettle, y permite evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar.

Su arquitectura se organiza de la siguiente manera:





Además de ser Open Source y sin costes de licencia, las características básicas de esta herramienta son:

- Entorno gráfico de desarrollo
- Uso de tecnologías estándar: Java, XML, JavaScript
- Fácil de instalar y configurar
- Multiplataforma: Windows, Macintosh, Linux
- Basado en dos tipos de objetos: Transformaciones (colección de pasos en un proceso ETL) y trabajos (colección de transformaciones)
- Incluye tres herramientas:
 - Spoon: para diseñar transformaciones y trabajos usando el entorno gráfico
 - PAN: para ejecutar transformaciones diseñadas con spoon desde línea de comandos.
 - Kitchen: para ejecutar trabajos diseñados con spoon desde línea de comandos.
- ✓ Data Mining

Es el proceso de correr datos en algoritmos completamente sofisticados, relevando significantes patrones y correlaciones que pueden estar escondidos. Esto puede ser usado para ayudar a entender lo mejor para el negocio y explotar el rendimiento de este en un futuro prediciendo completamente en el análisis.

Se caracteriza por:

- Descubrir patrones ocultos y correlacionales en los datos
- Prevenir eventos futuros basados en patrones históricos
- Contar con la tecnología de:
 - Poderoso motor de Data Mining
 - Herramientas de Diseño Grafico
 - Seguridad y conformidad
 - Servicios Web, Repositorios y definiciones basadas en XML
 - Rendimiento y escalabilidad.



- ✓ Pentaho BI Platform



La plataforma de Pentaho BI provee servicios críticos incluyendo programación, seguridad, integración, automatización y flujo de trabajo. Proporcionando habilidades a los usuarios finales de Pentaho y proveyendo de un lugar central para administrar y mantener el despliegue de la empresa BI.

- Integración con procesos de negocios
- Administra y programa reportes
- Administra seguridad de usuarios.

Alternativa 2: PALO



PALO (leído al revés OLAP) es un producto perteneciente a la empresa alemana Jedox. Se trata de un motor multidimensional OLAP, que está específicamente diseñado para mostrar información desde Excel, para todo tipo de análisis. También existe una versión sobre Eclipse y vía web.

Los datos quedan almacenados de forma jerárquica (multidimensional), lo que permite realizar las consultas a gran velocidad. Permite hacer “write back”, posibilitando hacer presupuestaciones, simulaciones y todo tipo de inclusión y generación de nuevos escenarios.

Sus características más importantes:

- ✓ Orientada a Planificación y Presupuestos. Basada en AJAX. Multiplataforma.
- ✓ Incluye: ETL, Reporting, OLAP.
- ✓ Licencia Open Source.

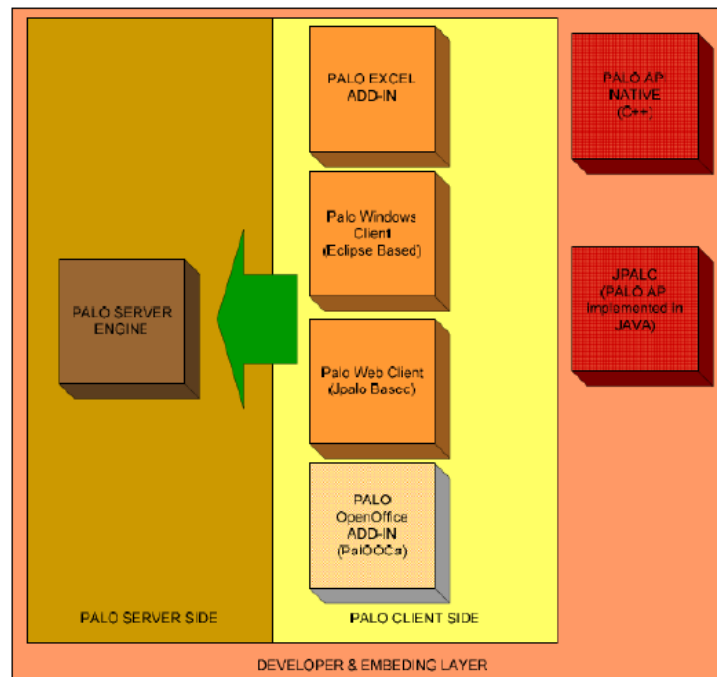
Palo es una herramienta Open Source y el add-in es gratuito, por lo que se trata de una gran utilidad para todos aquellos que manejen gran cantidad de datos en Excel.

Otra de las ventajas que esta tecnología ofrece, es que incorpora reglas que permiten cálculos complejos al estilo MDX con la ventaja que se pueden escribir en el mismo cubo permitiendo así hacer proyecciones de futuro sobre datos del pasado.

- ✓ Arquitectura de PALO

La arquitectura es cliente servidor. En el siguiente diagrama se puede observar los diferentes componentes disponibles para PALO así como las diferentes capas que componen su arquitectura. Palo es una herramienta escrita en C.

No todos los componentes disponibles son nativos de los fabricantes. Ha salido una implementación de la API de PALO en java (JPalo) que ha sentado las bases para futuros desarrollos sobre PALO como por ejemplo el cliente web de PALO para explotación de los cubos.

Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

Palo - Campos de acción

- Planificación
 - Análisis
 - Reporting
 - Dashboards
 - Data Integration
- ✓ Planificación

En la mayoría de las organizaciones la planificación de presupuestos, previsión de ventas, planificación de recursos, o una variedad de diferentes modelos de escenarios, aún se hacen en un entorno de Excel, lo que conlleva mucho tiempo entre los ciclos y las posibles inexactitudes en los modelos.

Una planificación bien fundada es esencial para que un negocio pueda reaccionar y adaptarse a los cambios circunstanciales, la regulación o la competencia que requieren los cambios en las operaciones y las finanzas en la empresa. Palo for Excel and Palo Suite permite la creación de soluciones de planificación flexibles, calcular modelos y escenarios en tiempo real, así como presupuestos y previsiones en todos los niveles.



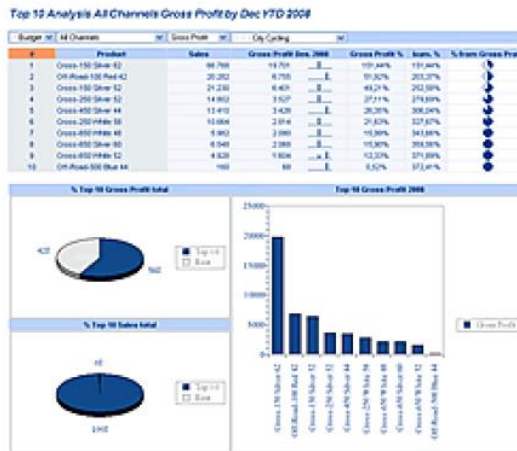
Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

Actual	Budget	Variance	Actual	Budget	Variance	Actual	Budget	Variance	Actual	Budget	Variance
226,702.45	200,000.00	26,702.45	200,000.00	204,139.24	-24,139.24	254,460.75	194,844.47	59,616.28	114,868.45	114,868.45	0.00
108,002.26	89,815.36	18,186.90	143,024.31	131,802.47	11,221.84	110,862.39	578,443.81	567,581.42	10,880.89	10,880.89	0.00
28,208.06	28,348.63	-140.57	36,524.00	36,498.41	25.59	39,887.31	108,479.62	107,992.31	487.31	487.31	0.00
23,478.26	20,862.79	2,615.47	39,653.00	40,787.18	-1,134.18	37,077.65	40,011.62	2,933.97	7,966.26	7,966.26	0.00
8,605.01	12,098.54	-3,493.53	16,642.05	16,642.05	0.00	8,207.28	11,308.63	-3,101.35	43,879.52	43,879.52	0.00
15,728.68	8,432.30	7,296.38	5,158.38	5,281.66	-123.28	7,360.73	28,019.66	20,658.93	6,360.73	6,360.73	0.00
13,641.67	11,471.07	2,170.60	18,108.49	9,688.07	8,420.42	10,555.44	48,667.18	37,133.06	11,534.12	11,534.12	0.00
7,515.47	1,488.72	6,026.75	2,887.86	2,485.24	402.62	1,864.63	7,262.43	5,377.41	1,887.22	1,887.22	0.00
16,302.02	13,004.43	3,297.59	16,274.00	15,784.87	489.13	22,748.87	19,088.75	3,660.12	52,853.47	52,853.47	0.00
8,184.30	4,688.04	3,496.26	5,588.01	5,733.63	-145.62	4,907.68	18,153.96	15,074.43	3,079.53	3,079.53	0.00
30,701.96	36,186.60	-5,484.64	41,114.47	40,865.00	249.47	36,787.66	133,642.23	115,488.27	18,153.96	18,153.96	0.00
5,876.21	5,274.45	601.76	7,465.65	6,271.19	1,194.46	5,162.80	18,708.62	16,142.42	2,566.20	2,566.20	0.00
5,078.22	3,188.87	1,889.35	2,945.19	2,485.30	459.89	2,214.87	9,373.87	8,348.82	1,025.05	1,025.05	0.00
12,268.42	14,669.42	-2,401.00	15,301.61	14,261.46	1,040.15	13,162.01	44,000.65	42,344.02	1,656.63	1,656.63	0.00
18,458.43	15,268.89	3,189.54	15,683.82	17,718.64	-2,034.82	18,208.41	52,698.89	47,282.21	5,416.68	5,416.68	0.00
80,480.46	82,127.21	-1,646.75	89,018.89	89,073.82	-55.93	83,084.38	163,016.04	162,461.49	554.55	554.55	0.00
20,008.00	20,202.17	-194.17	23,071.34	27,008.86	-3,937.52	23,001.71	77,207.16	78,482.27	-1,280.56	-1,280.56	0.00
2,868.06	7,038.22	-4,170.16	9,879.72	8,621.00	1,258.72	8,013.01	25,385.21	28,232.96	-2,847.75	-2,847.75	0.00
25,078.00	25,000.00	78.00	18,487.74	14,814.84	3,672.90	15,044.62	52,883.87	52,521.26	362.61	362.61	0.00
35,023.30	38,887.87	-3,864.57	27,887.65	40,496.18	-12,608.53	38,002.22	188,881.87	183,487.93	5,393.94	5,393.94	0.00
11,997.78	11,938.88	58.90	11,941.73	11,988.88	-47.15	11,417.17	40,429.01	41,504.44	-1,087.27	-1,087.27	0.00

✓ Análisis

Palo ofrece preparación precisa para la toma de decisiones, que hacen una contribución a la seguridad y el aprecio de la empresa. Las herramientas clásicas permiten trabajar con una solución sofisticada pero muy controlable.

A través de Palo se puede crear, sin grandes gastos de TI, un marco funcional, para trabajar de manera creativa y conceptual, con estructuras inteligentes y flexibles. Más allá de la presentación de informes, palo permite planificar los resultados de la información, verificar y consolidar los datos en tiempo real y adaptarlo a las necesidades individuales del momento.



✓ Reporting

Para no tener que solicitar ajustes en su aplicación para ser realizado por su departamento de TI, Palo ofrece una completa plataforma de inteligencia de negocios que incluye informes estandarizados que se pueden crear rápidamente de Microsoft Excel, por los mismos usuarios o para un usuario concreto.



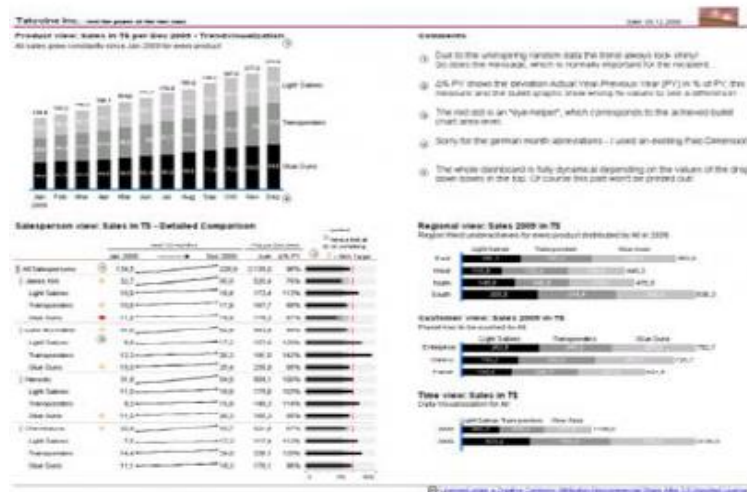
Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



✓ Dashboards

Palo ofrece una infraestructura completa, comenzando con la obtención de datos (ETL), almacenamiento de datos (OLAP), hasta la visualización de estos. Por lo tanto abarca todos los componentes que son necesarios para crear sus cuadros de mando personalizados.

Un tablero de control, es una forma de visualizar grandes cantidades de datos, provenientes de muchas fuentes de información de la organización, en un formato compacto conveniente, transformando dichos datos en indicadores útiles para la toma de decisiones.

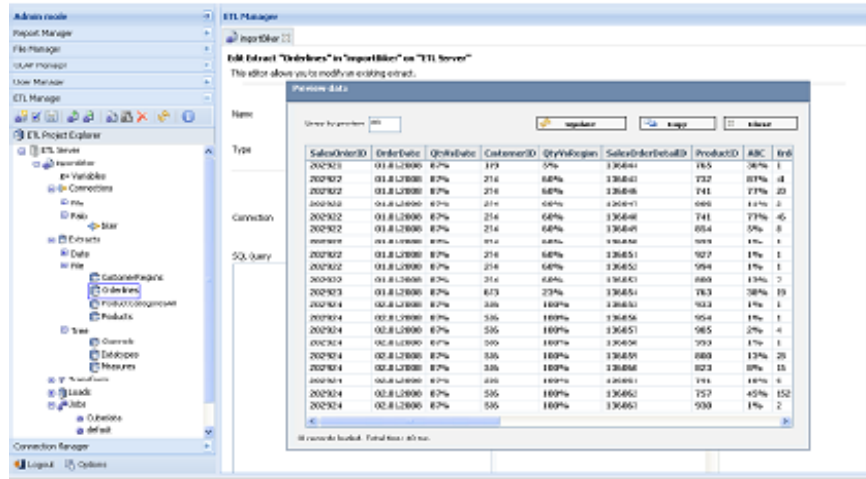


✓ Data Integration (ETL)

La integración de datos (ETL) representa la extracción y transformación de datos dentro de una estructura de TI. Estos procesos consisten en la extracción de datos de diversas fuentes de datos (bases de datos, archivos, aplicaciones, servicios Web, E-Mail, etc.), en el uso de diferentes normas de transformación (de búsqueda, se unen, la limpieza de duplicados, agregación, etc.) y, finalmente, la carga datos en el sistema deseado.



OLAP Palo permite, con el apoyo de una fuente de datos y un destino de datos, la manipulación de los cubos, niveles, atributos y reglas de manera muy simple.



Alternativa 3: JASPERSOFT



Es la denominación del conjunto de herramientas que permiten a las organizaciones generar información basada en sus propios datos de gestión para la evaluación y toma diaria de decisiones, en forma dinámica y on-line. El framework de trabajo de JasperSoft permite integrar fácilmente las diversas fuentes de datos disponibles en la empresa, y por medio de técnicas de análisis multidimensional obtener indicadores que, presentados en tableros de control y reportes dinámicos, proveen de esta sensible información a la alta gerencia.

Las características más importantes:

- ✓ Aplicación de tipo SaaS. Modular. Multiplataforma.
- ✓ Incluye: ETL, Reporting, OLAP, Dashboards.
- ✓ Licencia GNU GPL.

En su versión Open Source ofrece los siguientes beneficios:

- ✓ Permite el análisis in-memory sin la necesidad de un motor OLAP o Data Warehouse. Lo que brinda mayor rendimiento y rapidez en las consultas.
- ✓ Arquitectura SaaS. Prácticamente ninguna herramienta BI puede decir que no tiene un modelo SaaS o Cloud Computing.
- ✓ Permite integrarse con otras aplicaciones SaaS.
- ✓ Mejoras en el rendimiento y escalabilidad para un mayor número de usuarios y de datos.

La Suite de productos Jasper BI se compone de:



- JasperServer



JasperServer es un servidor de informes de alto rendimiento que puede ser desplegado como soporte ya sea en una aplicación independiente o integrado con otras aplicaciones, que proporciona valiosas consulta de usuario, informes, datos y capacidades de análisis.

JasperServer tiene capacidades para ser útil tanto a desarrolladores como administradores.

JasperServer para desarrolladores: es una arquitectura basada en estándares que ofrece una flexibilidad incomparable y escalabilidad para aplicaciones:

- ✓ Basado en una tecnología abierta de web estándar como XML, JDBC, Hibernate, Java y Web Services.
- ✓ Aplicado 100% en aplicaciones Java ofrecerá a los usuarios excelente y escalabilidad de datos.
- ✓ El portal viene incluido con MySQL, Tomcat y Liferay Portal.

JasperServer para administradores: incluye varias funciones integradas de administración:

- ✓ Informe de programación y distribución - incluyendo el calendario basado en las directivas, y versiones de informes históricos
- ✓ Exportación de informes en una amplia gama de formatos, como PDF, HTML, Excel, Flash, Word, RTF, y ODS / ODT
- ✓ Administrar usuarios, roles, y la seguridad a través de un repositorio central
- ✓ Permite construir informes de auditorías para determinar quién y cuándo realizo cada informe, para el cumplimiento, el análisis y la optimización del rendimiento.

JasperServer, permite a los usuarios finales crear informes y consultas ad hoc sin depender de TI. A través de un navegador web, que puede trabajar con una visión empresarial de los datos que no requiere de la comprensión de conceptos de base de datos o lenguajes de consulta, basta tan solo con arrastrar y soltar los elementos en la consulta. Los reportes pueden ser guardados, compartidos, programarse y distribuirse a través de aplicación de administración de JasperServer en el repositorio. Las mismas tecnologías, permiten de modo fácil arrastrar y soltar cuadros de mando. Los usuarios pueden construir y desplegar dashboards que logren vigilar los indicadores claves de rendimiento y métricas en tiempo real. Además, cualquier dirección URL de contenido



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

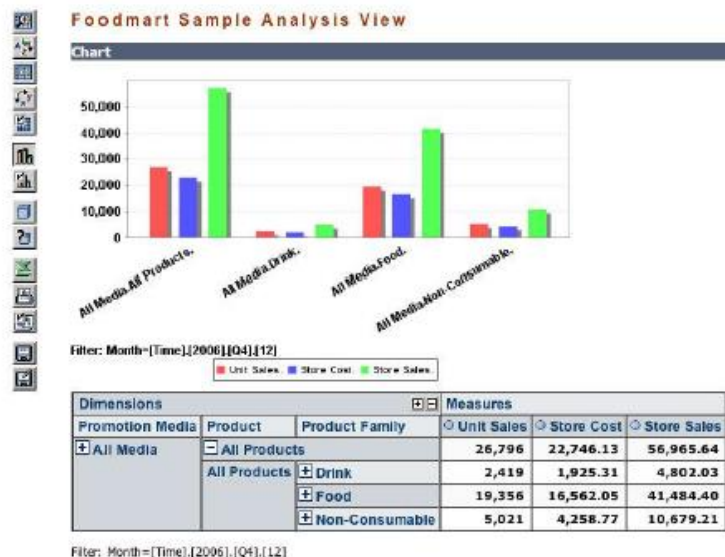
direccionable pueden incluirse en cuadros de mando, lo que permite mash-ups a través de aplicaciones de negocio internos y externos.



- JasperAnalysis



Al final del juego, el software de BI es mejor, pero para lograr eso es necesario ayudar a los usuarios a dar sentido a los datos multidimensionales. JasperAnalysis provee de gran alcance, basado en la web de procesamiento analítico en línea (OLAP) y la capacidad para que los usuarios de negocio pueden explorar los datos para descubrir las causas, detectar tendencias y patrones, identificar las anomalías, y visualizar las correlaciones, realizando todo el trabajo analítico pesado, que sería demasiado difícil o demasiado costosa para llevar a cabo utilizando SQL estándar. OLAP se asocia a menudo con grandes presupuestos, grandes equipos, gran visión, y un montón de tiempo, pero con JasperAnalysis no estás desplegando OLAP. Debido a esto todo lo que se necesita es un presupuesto modesto. Con JasperAnalysis, puede aprovechar las inversiones existentes en relación DBA y habilidades para aplicar gradualmente OLAP.



- JasperReports





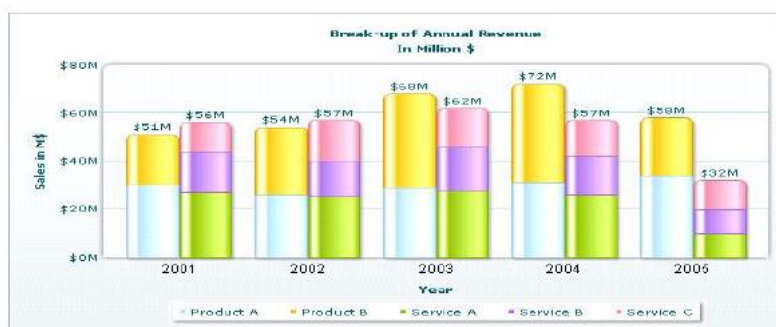
Para los usuarios empresariales, un aspecto muy importante en la inteligencia de negocios es la habilidad de ver la información que se necesita lo más rápido y tan claramente como sea posible, los informes que son demasiado complejos para crear o utilizar pronto pierden valor. Con JasperReports, se puede diseñar, implementar o integrar la potencia de los informes de Jaspersoft en sus aplicaciones o independientemente.

La biblioteca JasperReports ha sido desarrollada 100% en Java, con el fin de integrar sin problemas en su aplicación el uso de estándares y definiciones de informes XML, que permite a los usuarios tomar ventaja de la comunidad más grande y activa de los diseñadores de informes y los desarrolladores.

El corazón de la interfaz de JasperReports es iReport, el cual incluye un diseñador gráfico integrado, dando a los usuarios la flexibilidad para crear dibujos animados, mapas interactivos y mapas que son fáciles de diseñar y poderosas de utilizar. Con iReport, los diseñadores pueden crear, basado en la web, informes listos para imprimir en prácticamente cualquier estilo de cuadros de mando, tablas, tablas cruzadas y medidores.

Las características importantes que permiten a JasperReports brindar presentación de informes de alto rendimiento son:

- ✓ Escalable y útil: Permite a los usuarios generar informes de cualquier longitud y complejidad.
- ✓ Geo-Visualización de Mapas con Flash: Mostrar ubicación basada en la performance a través del mapa base de información.
- ✓ Reproductores Flash: Cientos de mapas flash para todas las geografías, indicadores de rendimiento interactivos, etc.
- ✓ Flash Charting: Algunos de los informes más atractivos se basan en la elegancia y la simplicidad de la presentación. Aquí es donde Flash Charting puede ayudar. JasperReports ahora contiene docenas de suministro de ejes múltiples, de anillos, y gráficos de barras, que pueden dar informes complejos cuando son necesarios cambios de imagen.



- JasperETL



La mayoría de las organizaciones, cuanto más se basan en el poder de la toma de decisiones mediante BI, mayor es su necesidad de integración de datos. JasperETL es la



herramienta de Jaspersoft Business Intelligence Suite que se necesita, cuando es el momento de:

- ✓ Agregar grandes volúmenes de datos de múltiples fuentes de datos.
- ✓ Incluir a la solución BI para Data Marts y Data Warehouses.
- ✓ Aumentar el rendimiento de las cargas fuera de consulta y análisis de los sistemas operativos.

La herramienta JasperETL permite:

- ✓ Acceder a una visión no-técnica de la información del flujo de trabajo, mediante el Business Modeler.
- ✓ Visualizar y modificar el proceso de ETL con el Diseñador de Empleo, una herramienta de edición gráfica.
- ✓ Definir las asignaciones complejas y transformaciones con Mapper y otros componentes de la transformación.
- ✓ Generar Perl portátil o código Java, que puede ejecutarse en cualquier máquina
- ✓ Seguimiento de las estadísticas de ETL, de principio a fin, con la depuración (debugging) en tiempo real.
- ✓ Permitir la salida y la entrada simultánea de múltiples fuentes, incluyendo archivos planos, archivos XML, bases de datos, servicios Web, POP y servidores FTP con cientos de conectores disponibles.
- ✓ Configurar las fuentes de datos heterogéneas y complejas que incluyen formatos de archivo de posición, delimitado, CSV, RegExp, XML, y LDIF con los asistentes de los metadatos.
- ✓ Supervisar eventos de empleo (éxitos, fracasos, advertencias, etc.), plazos de ejecución, y volúmenes de datos, mediante la consola de supervisión de la actividad (AMC).

La integración de datos puede ser uno de los mayores costos iniciales. Se puede reducir el coste total de propiedad y reducir la complejidad de su infraestructura de TI a medida que se implementa la escala de herramientas de Jaspersoft para satisfacer las necesidades de crecimiento de su organización.

Si bien estas 3 son las soluciones líderes en Open Source dentro del ámbito de Business Intelligence es importante mencionar que existen otras alternativas.

OTRAS ALTERNATIVAS:



Se trata de una aplicación Business Intelligence de tipo OLAP construida para acceso web y que permite acceder a datos de SQL Server y Mondrian.

Las características más importantes:

- ✓ Aplicación de tipo SaaS. Desarrollada en J2EE.
- ✓ Multiplataforma.
- ✓ Incluye: OLAP.
- ✓ Licencia Open Source.



Pretende ser una solución completa de Business Intelligence que incluye desde la extracción a la minería, pasando por análisis y reporting. Utiliza componentes muy similares a los de Pentaho.

Características más importantes:

- ✓ Aplicación web desarrollada en J2EE. Modular. Multiplataforma.
- ✓ Incluye: ETL, Reporting, OLAP, Data mining, Dashboards.
- ✓ Licencia GNU LGPL



Eclipse BIRT (“Business Intelligence and Reporting Tools”) es un proyecto de software Open Source dentro de la fundación eclipse, que ofrece un conjunto ampliable de herramientas para desarrollar aplicaciones Java que contengan capacidades de Business Intelligence, utilizando los estándares XML y web. Incluye un diseñador de informes y un componente de ejecución que se puede añadir a su servidor de aplicaciones, es decir, entre sus módulos se destacan:

- *Diseñador de informes Eclipse BIRT*: es una herramienta de escritorio basada en Eclipse que permite a los desarrolladores de informes definirlos usando herramientas visuales. El diseñador aporta un conjunto de controles “ricos” para crear informes que incluyen elementos para mostrar presentaciones complejas de datos, imágenes, realizar cálculos simples o complejos y mostrar representaciones gráficas de los datos. Los diseños de los informes se almacenan en un formato XML abierto y documentado. Las librerías permiten compartir y reutilizar los informes. Las plantillas contienen presentaciones de los informes más comunes, permitiéndole obtener un estilo consistente.
- *Motor de informes Eclipse BIRT*: posibilita que los informes sean usados desde cualquier aplicación Java usando los diseños generados por el diseñador. El motor de informes es el elemento clave del proyecto BIRT en el entorno de ejecución.

Para finalizar la presentación y análisis de las diferentes alternativas se citará una investigación realizada por Gartner S.A. Ésta compañía posee en su interior un área, Gartner Business Intelligence, que se dedica a la investigación, análisis y asesoramiento en este campo.

El 23/12/2009 en la web “TodoBI” se exhibió un artículo donde dicha firma, luego de haber desarrollado una profunda investigación, se atreve a realizar un pronóstico del futuro del Business Intelligence Open Source. Se expone dicho artículo a continuación:

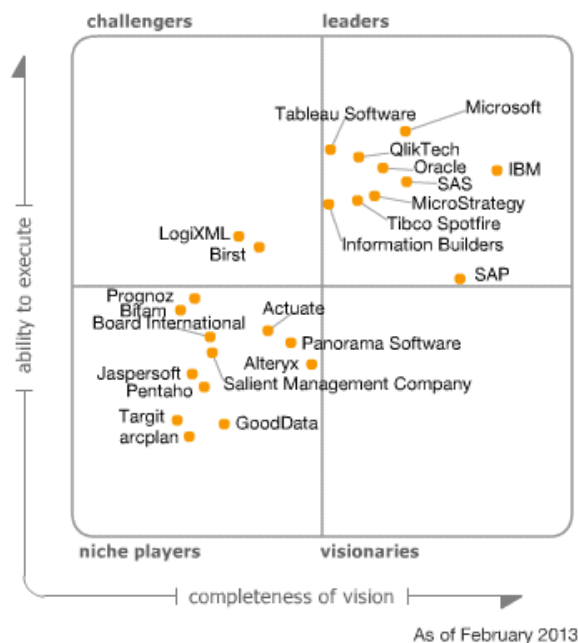
“Aunque con retraso, pero parece que finalmente Gartner pronostica que el Business Intelligence Open Source tendrá un gran crecimiento en los próximos años. Se presenta aquí un resumen de las conclusiones obtenidas en una nota realizada por Andreas Bitterer:



- ✓ El uso y adopción del BI Open Source se dobla cada año
- ✓ Se usa el BI Open Source como cualquier otra solución propietaria, empezando por reporting, análisis y Dashboards.
- ✓ En grandes proyectos sigue siendo muy importante el coste de los consultores y desarrolladores para llevar a cabo las implementaciones (como en proyectos BI tradicionales).
- ✓ Gartner recomienda a las organizaciones comparar las soluciones BI Open Source al mismo nivel que las comerciales.
- ✓ Cuando se evalúe una solución Open Source hay que pensar en un plan a 3 años que incluya también los costes de soporte y mantenimiento.
- ✓ El titular más llamativo es que de aquí al 2012, el Business Intelligence Open Source crecerá en cinco veces.
- ✓ Se comenta el impacto que está teniendo en los fabricantes tradicionales, que están llegando incluso a ofrecer versiones limitadas de sus soluciones de forma gratuita, como el Microstrategy Reporting y Business Objects Crystal Report.
- ✓ El estudio destaca como soluciones Open Source BI más destacadas a Pentaho, Jasper, Palo, SpagoBI, Birt...”

Se puede acceder al post expuesto por la web “TodoBI” mediante el siguiente link:
<http://todobi.blogspot.com/2009/12/gartner-pronostica-un-gran-crecimiento.html>

Artículos más actuales también continúan analizando las herramientas Open Source para BI, por ejemplo Gartner en una nueva investigación (febrero de 2013) publica su “Magic Quadrant” donde posiciona como principales tecnologías de código abierto a Pentaho y Jasper. El artículo completo lo podemos leer desde:
<http://www.zdnet.com/gartner-releases-2013-bi-magic-quadrant-7000011264/>



Dentro del artículo es posible destacar frases como: “... existen jugadores importantes en el ambiente de código abierto de BI, y tres de ellos son Pentaho , Jaspersoft y Actuate...”



Prácticamente todas las publicaciones sobre el tema, mencionan de acuerdo a determinados criterios a estas suite de herramientas como las mejores y líderes. Por ejemplo en la publicación:

<http://todobi.blogspot.com.ar/2011/04/los-63-productos-open-open-source.html> es posible encontrar la lista de principales soluciones Open Source comerciales a nivel empresarial, dentro de la cual encontramos en el rubro Business Intelligence a: Jaspersoft, Palo BI Suite, Pentaho, OpenReports y RapidMiner.

Considerando lo expuesto y considerando que son soluciones de mayor jerarquía y rendimiento se consideran las siguientes alternativas como viables para la implementación del proyecto:

- ✓ Pentaho
- ✓ JasperSoft
- ✓ Palo

Evaluación contra los objetivos OBLIGATORIOS:

Objetivos Obligatorios	Pentaho		Palo		Jaspersoft	
	<i>Información adicional</i>	<i>Pasa/N o pasa</i>	<i>Información adicional</i>	<i>Pasa/N o pasa</i>	<i>Información adicional</i>	<i>Pasa/N o pasa</i>
Solución Open Source	Desarrollado en el punto anterior.	PASA	Desarrollado en el punto anterior.	PASA	Desarrollado en el punto anterior.	PASA

Evaluación de las alternativas contra los objetivos DESEADOS.

Objetivos Deseados	Calificación									
	Peso	Pentaho			PALO			Jaspersoft		
		Información adicional	calificación	Calificación ponderada	Información adicional	calificación	Calificación ponderada	Información adicional	calificación	Calificación ponderada
Análisis	10		10	100		6	60		6	60
Reporting	10		8	80		6	60		10	100
Data Integration	10		8	80		8	80		8	80
Dashboard	8		8	64		6	48		8	64
Data Mining	6		6	36		2	12		2	12
Totales				340			260			314

Puntuación ponderada es: peso * calificación.



Elección provisional: elegimos la alternativa con mayor puntuación, es decir el proyecto se implementará con la suite de herramientas de:

PENTAHO

4.5.1.4 Consecuencias de la alternativa

Al evaluar las posibles causas adversas y futuros riesgos para implementar las diferentes alternativas, es posible concluir en que todas cuentan con los mismos riesgos, debido a que la organización cuenta con el hardware y espacio necesario para la puesta en marcha de cualquiera de las herramientas; pero el posible inconveniente surge por la aceptación o no de esta nueva solución. Algo nuevo implica un cambio y como todo cambio suele encontrar algún nivel de resistencia por parte de los integrantes de la organización. Para disminuir ésta reacción negativa frente a la implementación de la solución, se realizarán capacitaciones como charlas informativas para que además de familiarizarse con la herramienta, los usuarios conozcan los beneficios de este tipo de software.

Al estar todas las alternativas alcanzadas por los mismos riesgos, es posible concluir que la alternativa recomendada es la que obtuvo mayor puntaje, **PENTAHO**.

Buscando e investigando es posible descubrir que otros grandes proyectos y empresas de BI utilizan la suite Pentaho para sus desarrollos lo cual apoya la elección, por ejemplo el Grupo “eGlu BI” es una empresa formada por jóvenes profesionales con amplia experiencia en el mercado local e internacional. Sus miembros han formado parte de grandes compañías locales y multinacionales del sector tecnológico, por lo cual se consolida como una empresa joven con amplios conocimientos en diversas tecnologías.

El grupo eGlu BI utiliza en todos sus pasos la suite Pentaho, junto con la metodología HEFESTO.

Es posible resumir que la selección de Pentaho, se debe a dos grandes razones:

- ✓ Es la plataforma OSBI (Open Source Business Intelligence) líder del mercado y se caracteriza por estar orientada a la solución, centrada en procesos y con una muy buena interfaz gráfica.
- ✓ Incluye todos los principales componentes requeridos para implementar soluciones de Inteligencia de Negocios, tales como: Informes Interactivos, Reportes Ad hoc, Reporting, Cubos OLAP, Procesos ETL, Suscripciones, Data Mining, Alertas, Dashboards, etc.

4.6 Implementación y Puesta en Marcha

Se ha llegado a la última sección de la metodología, habiendo en un primer paso analizado la empresa, conocido su contexto y actualidad. Seguidamente se realizó el modelado del Data Warehouse, para lo cual fue necesario comenzar por distinguir que información era la que se necesitaba plasmar en el mismo y de donde obtenerla. Posteriormente, una vez definida la estructura que debía contener el Warehouse, se procedió a aplicar el método de decisión planteado por Kepner y Tregoe en su libro “El Nuevo Directorio Racional” para tomar una decisión correcta que permita elegir el



mejor conjunto de herramientas de software que posibiliten implementar el proyecto de manera eficaz y eficiente. Finalmente solo resta aplicar el trabajo “duro”, instalar y trabajar con el software para lograr que el Data Warehouse sea real y comience a brindar los beneficios mencionados durante todo el proyecto.

Es importante destacar que ésta etapa se compone de una serie de pasos que harán que el trabajo este correctamente organizado.

4.6.1 Descripción de la Tecnología

Este primer paso tiene por finalidad realizar un relevamiento de la tecnología actual que se encuentra en la organización y compararlo contra los requerimientos que poseen los diferentes software que serán necesarios para la construcción del proyecto. Una vez comparados, se deberá decidir si es necesario hacer una mejora de la tecnología actual o si, por el contrario, es posible la implementación sin la necesidad de realizar una inversión.

4.6.1.1 Descripción de las herramientas y elementos de software a utilizar

En primer lugar se debe hacer un listado de cuáles serán los software que se instalarán y utilizarán para implementar el proyecto.

En el presente trabajo las herramientas que se usarán son:

SGBD:

- ✓ MySQL Server 5.5 y MySQL Workbench 5.2 CE

Herramientas utilizadas de la plataforma Pentaho BI:

- ✓ *Pentaho Data Integration 5.0.1*

Pentaho Data Integration es una herramienta Open Source que permite implementar los procesos ETL. También se la conoce como Kettle, acrónimo recursivo de: “Kettle Extraction, Transformation, Transportation, and Load Environment “).

PDI facilita la construcción, actualización, y mantenimiento de Data Warehouses y se caracteriza por:

- ✓ No tener costes de licencia
- ✓ Ser multiplataforma (se puede utilizar en Windows, Macintosh, Linux),
- ✓ Tener un entorno gráfico de desarrollo
- ✓ Ser fácil de instalar y configurar.

Las soluciones de PDI se basan en dos diferentes tipos de objetos:

Transformaciones: Una transformación es una colección de pasos. Cada paso representa una operación particular sobre los datos. Los pasos se conectan entre sí a través de saltos y trabajan de manera simultánea y asíncrona.

Trabajos: Un trabajo presenta una orientación a la tarea y a los datos. Usualmente un Trabajo se compone de una o más Transformaciones, que serán ejecutadas secuencialmente.

Básicamente PDI incluye cuatro herramientas:



Spoon: para diseñar transformaciones y trabajos utilizando un entorno gráfico

Pan: para ejecutar transformaciones mediante líneas de comando

Kitchen: para ejecutar trabajos mediante líneas de comando

Carte: es un server liviano para ejecutar transformaciones y trabajos en host remotos.

✓ ***Pentaho Schema Workbench 3.5.0***

Es una interfaz de diseño que permite crear y probar esquemas de cubos OLAP Mondrian visualmente. El motor Mondrian procesa las solicitudes de MDX con los esquemas ROLAP (Relational OLAP). Estos archivos de esquema son modelos de metadatos XML los cuales son creados en una estructura específica que utiliza el motor de Mondrian. Estos modelos XML pueden ser considerados estructuras de forma de cubo que utilizan tablas de hechos y de dimensiones existentes que se encuentran en el RDBMS.

✓ ***Pentaho BI Server 4.8.0***

Es un paquete de inteligencia de negocio de código abierto que incluye ETL, análisis, los metadatos y capacidades de reporting. Es un software de código abierto en su totalidad, en su mayoría bajo la Licencia Pública General GNU versión 2, con partes bajo el LGPLv2, la Licencia Pública Común, y de la Licencia Pública de Mozilla. Pentaho optimiza, la plataforma de pruebas y garantías certificadas con la suite de BI; esta versión mejorada del software se empaqueta con una potente herramienta de gestión de servicio denominado Enterprise Console, soporte a usuarios, indemnización IP, y documentación profesional, la cual es vendida por Pentaho como Enterprise Edition.

El BI Server está compuesto por:

- ✓ **Consola de Usuario (biserver-ce):** Permite navegar la solución, ejecutar y crear componentes, administrar calendario, administrar permisos, etc.
- ✓ **Consola Administrativa:** Permite crear conexiones a bases de datos, administrar usuarios y roles, administrar el calendario, refrescos, etc. Esta consola se conecta a la consola de usuario.

Es importante destacar que se utilizará el plugin de Pentaho, *SAIKU*, el mismo se seleccionó por encima de JPIVOT luego de realizar una comparación y prueba de ambos, prefiriendo *SAIKU* por sus ventajas en cuanto a velocidad, flexibilidad y amigabilidad en su interface gráfica.

SAIKU es un excelente visor OLAP que proporciona al usuario final una magnífica herramienta para realizar análisis de forma fácil e intuitiva. Pero Saiku no es sólo eso, es un buen ejemplo de cómo un proyecto Open Source puede ofrecer soluciones de excelente calidad a la vanguardia de la tecnología y delicada experiencia de usuario. Afortunadamente es un proyecto muy joven hecho con paciencia y cabeza lo cual lo dota con una arquitectura técnica que le permite ser muy flexible y versátil. Algunas características de Saiku son:

- Se puede utilizar Saiku como un servidor independiente, permitiendo realizar análisis OLAP.



- Se puede embeber en un servidor Pentaho como un plugin de forma fácil y sencilla. En el proyecto esta es la alternativa utilizada.
- Se puede utilizar como origen de datos. Es un backend, pudiendo, por ejemplo construir nuestra propia interfaz de usuario.
- Implementación de los escenarios de Mondrian. Y la posibilidad de editar celdas.
- Una interfaz fácil limpia e intuitiva que permite a los usuarios construir sus propias vistas arrastrando y soltando campos.
- Permite analizar los datos en forma de tabla pivotante o en modo gráfico.
- Permite visualizar los datos de origen.

En el anexo 4, es incluido un tutorial sobre cómo utilizar Saiku, habiéndolo embebido como un plugin en el BI Server.

La decisión de utilizar Pentaho como solución BI para llevar a cabo el proyecto, fue evaluada en la sección 3 de la metodología propuesta (Selección del conjunto de herramientas), luego de realizar una profunda investigación y comparación de tecnologías OSBI; llegando finalmente a la conclusión de implementar PENTAHO para llevar a cabo el trabajo final de grado, ya que ésta es la herramienta que satisface con mayor amplitud las necesidades de análisis de los datos, desarrollo de informes empresariales e integración de datos; además de contar con buenas herramientas para trabajar con data mining y dashboards, lo cual brinda la posibilidad de continuar y ampliar el proyecto en un futuro.

Además, es importante destacar que la suite de herramientas de Pentaho junto con MySQL conforman una excelente opción para crear soluciones analíticas OLAP (Online Analytical Processing). Esta última, se ha convertido en una tecnología muy popular para los usuarios que necesitan de una alta capacidad y velocidad de acceso a la información que almacenan en sus Bases de Datos.

Pentaho y MySQL ofrecen una combinación para proporcionar soluciones OLAP totalmente en entorno Open Source, con la base de Pentaho Mondrian y de MySQL 5. Es por ello que se ha seleccionado a MySQL como Sistema de Gestión de Base de Datos.

4.6.1.2 Comparación entre tecnología existente y necesaria para el proyecto

Una vez que se han definido cuales son las herramienta de software a utilizar es necesario conocer los requerimientos tecnológicos que estas tienen, para saber si la organización cuenta con los mismos o si por el contrario será inevitable adquirir determinados componentes. En caso de ser necesario una inversión en material tecnológico, se deberá aplicar el mismo proceso racional de toma de decisiones desarrollado en la sección 3 (Análisis de Decisiones del proceso de Kepner y Tregoe) de esta metodología, pero ahora el análisis, comparación y selección de componentes sería de hardware y no de software.

Plataforma tecnológica disponible:

La infraestructura tecnológica con que cuenta la Secretaría de Economía y Finanzas Públicas para el proyecto es:

Servidor:



Constituye la plataforma sobre la que se alojarán las aplicaciones y donde se ejecutarán los procesos, es decir, se alojará aquí la base de datos y las herramientas del software necesarias para el desarrollo y funcionamiento de los cubos. El servidor con que cuenta actualmente la Secretaría de Economía y Finanzas Públicas de la Municipalidad de Cosquín, es un servidor dedicado el cual presenta las siguientes características:

- ✓ Arquitectura: 64 bits
- ✓ Procesador: Intel Pentium Dual CPU E2200. 2.20Ghz
- ✓ Memoria RAM: 2 Gb.
- ✓ Espacio en Disco: 149 GB. Libre actualmente: 20 GB.
- ✓ Sistema Operativo: Windows Server 2008 Standard. Service Pack 2

Cliente:

El cliente accede a las aplicaciones que se encuentran en el servidor a través de un browser de Internet. Actualmente el acceso al servidor por parte de los usuarios de la municipalidad es a través de la red interna (intranet), utilizando computadoras que se encuentran instaladas y funcionando en la Municipalidad, las cuales tienen como características:

- ✓ Arquitectura: 32 bits
- ✓ Procesador: Intel Pentium Dual CPU E2200. 2.20Ghz
- ✓ Memoria RAM: 2 Gb.
- ✓ Espacio en Disco: 320 Gb.
- ✓ Sistema Operativo: Windows 7. Service Pack 1

Además existen ciertos usuarios con permisos para conectarse a través de una VPN, denominada “Muni Cosquín” los cuales pueden ingresar al servidor desde distintos puntos geográficos distantes a la Municipalidad con computadores propios.

Plataforma tecnológica necesaria para el proyecto

Habiendo descrito y detallado la tecnología actual con que cuenta el área de Economía y Finanzas Públicas, se procede a continuación a exponer los requerimientos de las diferentes herramientas que serán utilizadas en el proyecto, para luego comparar si será necesario realizar alguna inversión.

Servidor:

Como se indicó en el primer punto de esta etapa, en el servidor se alojará la base de datos, el software para su manipulación y los componentes de Pentaho necesarios para cumplir con los requerimientos empresariales. Es decir, es necesario ubicar en el servidor los siguientes componentes:

- ✓ MySQL Server 5.5 y MySQL Workbench 5.2 CE
- ✓ Pentaho Data Integration 5.0.1
- ✓ Pentaho Schema Workbench 3.5.0
- ✓ Pentaho BI Server 4.8.0

Los requisitos mínimos de estos componentes son:

Requisitos mínimos para ejecutar MySQL Server 5.5 en Windows:



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

- ✓ Sistema Operativo: Windows de 32 bits, tal como 9x, Me, NT, 2000, XP, o Windows Server 2003.
- ✓ Se recomienda fuertemente el uso de un sistema operativo Windows basado en NT (NT, 2000, XP, 2003) puesto que éstos permiten ejecutar el servidor MySQL como un servicio.
- ✓ Soporte para protocolo TCP/IP.
- ✓ Suficiente espacio en disco rígido para descomprimir, instalar, y crear las bases de datos de acuerdo a sus requisitos. Generalmente se recomienda un mínimo de 200 megabytes.

Requisitos mínimos para Pentaho:

Requirement	BI Server	Data Integration	Schema Workbench	Dashboard Designer
CPU	1.8 GHz Processor	1.8 GHz Processor	1.8 GHz Processor	1.8 GHz Processor
Memory	2 GB RAM	512 MB RAM	256 MB RAM	256 MB RAM
Drive Space	400 MB	180 MB	50 MB	100 MB
Java Virtual Machine	Sun v 1.5 (5.0)	Sun v 1.5 (5.0)	Sun v 1.5 (5.0)	Sun v 1.5 (5.0)

Cliente:

El único requisito que poseen estas herramientas para que el cliente sea capaz de acceder a las aplicaciones y poder consultar la información, es tener instalado un browser de Internet, el cual le permita navegar las herramientas. Se recomienda utilizar Google Chrome.

Comparación de la plataforma tecnológica existente y necesaria para el proyecto

En este punto ya se conoce cuál es la situación tecnológica con que cuenta el área, así como la requerida por las herramientas a utilizar, solo resta compararlas y definir si el proyecto se podrá implementar con la tecnología actual o no.

Servidor:

Comparando las descripciones del servidor dedicado que posee la Secretaría de Economía y Finanzas Públicas con los requisitos que posee el software a instalar, se observa claramente que el mismo satisface todos los requerimientos de manera excelente, por lo cual no será necesario hacer ninguna inversión en esta materia y se utilizará la tecnología disponible.

Cliente:

Al ser el único requisito la disponibilidad de un browser (preferiblemente Google Chrome) en las pc de usuario, es posible confirmar que las máquinas del municipio cumplen ésta necesidad y no es necesaria ninguna inversión.

4.6.2 Definición de la arquitectura del Data Warehouse

En este punto se definirá la arquitectura del proyecto de Data Warehouse teniendo en cuenta y utilizando como referencia lo expuesto en el marco teórico, punto 3.5 “Arquitectura del Data Warehouse”.

La arquitectura del proyecto representa la estructura total de los datos, la comunicación, el procesamiento y la presentación de los mismos, ante los usuarios finales que disponen de una computadora dentro de la empresa.



Es posible realizar un mapeo general de la arquitectura del DWH planteada en etapas anteriores, con los componentes de la suite Pentaho que se van a utilizar. A grandes rasgos este mapeo sería:

- 1) Los datos son extraídos a través de Kettle o Pentaho Data Integration (herramienta que ocupa el lugar de Load Manager descrito en el capítulo 3) desde la base de datos operacional de la Municipalidad de Cosquín: Program. En un primer momento se realiza la extracción únicamente de los datos necesarios y se los coloca en un Staging Area, aplicando el mínimo de transformaciones. En este punto no se aplican restricciones de integridad ni se utilizan claves, los datos se tratan como si las tablas fueran ficheros planos, minimizando la afectación a los sistemas origen. Este almacenamiento intermedio se constituye por la base de datos: *STAGING_AREA_DW_MUNICOSQUIN*.
- 2) Una vez que se encuentran los datos en el Staging Area, se continúa trabajando con PDI. Los datos son integrados, transformados y limpiados, para luego ser cargados en el DW llamado: *DW_MUNICOSQUIN*.
- 3) Utilizando a Pentaho Workbench como Data Warehouse Manager la información del DW se estructura en cubos multidimensionales, ya que éstos preparan la información para responder a consultas dinámicas con una buena performance.
- 4) Pentaho BA Server forma parte del DW Manager, ya que entre otras cosas gestiona los metadatos de negocio y las estructuras de datos (modelos de negocios, cubos multidimensionales, etc). Así mismo, Pentaho BA Server, contiene los componentes necesarios para realizar las tareas del Query Manager, por ejemplo el motor multidimensional Mondrian.



- 5) Ya creados los cubos multidimensionales, los usuarios acceden a éstos utilizando diversas herramientas de consulta, exploración, análisis, reportes, etc.
- 6) Los usuarios seleccionan o establecen que datos desean obtener del DW, mediante las interfaces de la consola de usuario de Pentaho Server, en este caso utilizando el plugin Saiku. En el proyecto los usuarios pueden acceder al servidor donde se encuentra dicha consola a través de la intranet, ya sea desde las PC ubicadas en el edificio de la Municipalidad o vía VPN.
- 7) Pentaho BA Server recibe el pedido de los usuarios, construye la consulta (utilizando la metadata) y la envía al Query Manager. Es importante destacar que Pentaho Server contiene en su interior un administrador de consulta (motor Mondrian del cual ya se explicó su funcionamiento) que permite generar análisis y reportes.
- 8) El Query Manager ejecuta la consulta sobre la estructura de datos con la que se esté trabajando, es decir sobre los cubos multidimensionales creados con Pentaho Workbench.
- 9) El BA Server envía a Saiku los resultados de la consulta para que éste presente la información requerida por los usuarios.

A partir de los siguientes puntos se irán desarrollando los pasos descriptos, para comprender de qué forma se crea toda la arquitectura de Data Warehouse utilizando la suite de herramientas de Pentaho.

Tomando como referencia la arquitectura planteada se resaltaré en cada etapa la parte de la arquitectura trabajada.

4.6.3 Integración de Datos

Se ha llegado a un punto muy importante de la metodología, es el momento de trasladar todo lo modelado y definido durante las 3 primeras secciones a la implementación física, para ello será necesario utilizar las herramientas de software.

A partir de este momento se buscará convertir en realidad la arquitectura presentada en el paso anterior. En éste punto se exponen los trabajos y acciones que se realizan utilizando Kettle o PDI, es decir todo el trabajo de ETL, comenzando con la extracción de datos del OLTP de la Municipalidad (Base de datos Program) para trasladarlo al Staging Área y finalmente crear y poblar el DW.

En este punto se expone lo referido únicamente a la parte de izquierda de la arquitectura (OLTP – STAGING AREA – LOAD MANAGER – DATAWAREHOUSE), como se podrá observar en la imagen aparece un nuevo dibujo haciendo referencia a un archivo de Excel (TIOPROMO), esto es debido a que al momento de crear las dimensiones y cargarlas en el Warehouse se detectó que, para la dimensión TIPO_PROMO será necesario utilizar un archivo desde donde se obtendrán los datos tanto para su creación como actualización.

En este punto se expondrá lo referido a lo que se enmarca en la siguiente imagen.



Más adelante, en los puntos siguientes, se presentarán la manera en que se crean los cubos multidimensionales utilizando Pentaho Workbench, así como la instalación y acceso a las herramientas de consulta que permiten ver, analizar y tomar decisiones a la dirección.

4.6.3.1 Carga Inicial

El primer paso consiste en crear y poblar el modelo de datos modelado durante la sección 3 de la metodología, para ello será necesario realizar una serie de tareas como limpieza de datos, calidad de datos, procesos ETL, etc.

Para el presente proyecto se ha convenido crear un área de datos previa al Warehouse, conocida como Staging Area la cual se constituye mediante una base de datos, y recordando lo expuesto en el marco teórico, la misma es un área temporal donde se escogen los datos que se necesitan de los sistemas origen (OLTP) necesarios para las cargas, aplicando el mínimo de transformaciones a los mismos. No se aplican restricciones de integridad ni se utilizan claves, los datos se tratan como si las tablas fueran ficheros planos. De esta manera se minimiza la afectación a los sistemas origen, la carga es lo más rápida posible para acotar la ventana horaria necesaria, y se reduce también al mínimo la posibilidad de error.

Una vez que los datos han sido traspasados, el DW se independiza de los sistemas origen hasta la siguiente carga.

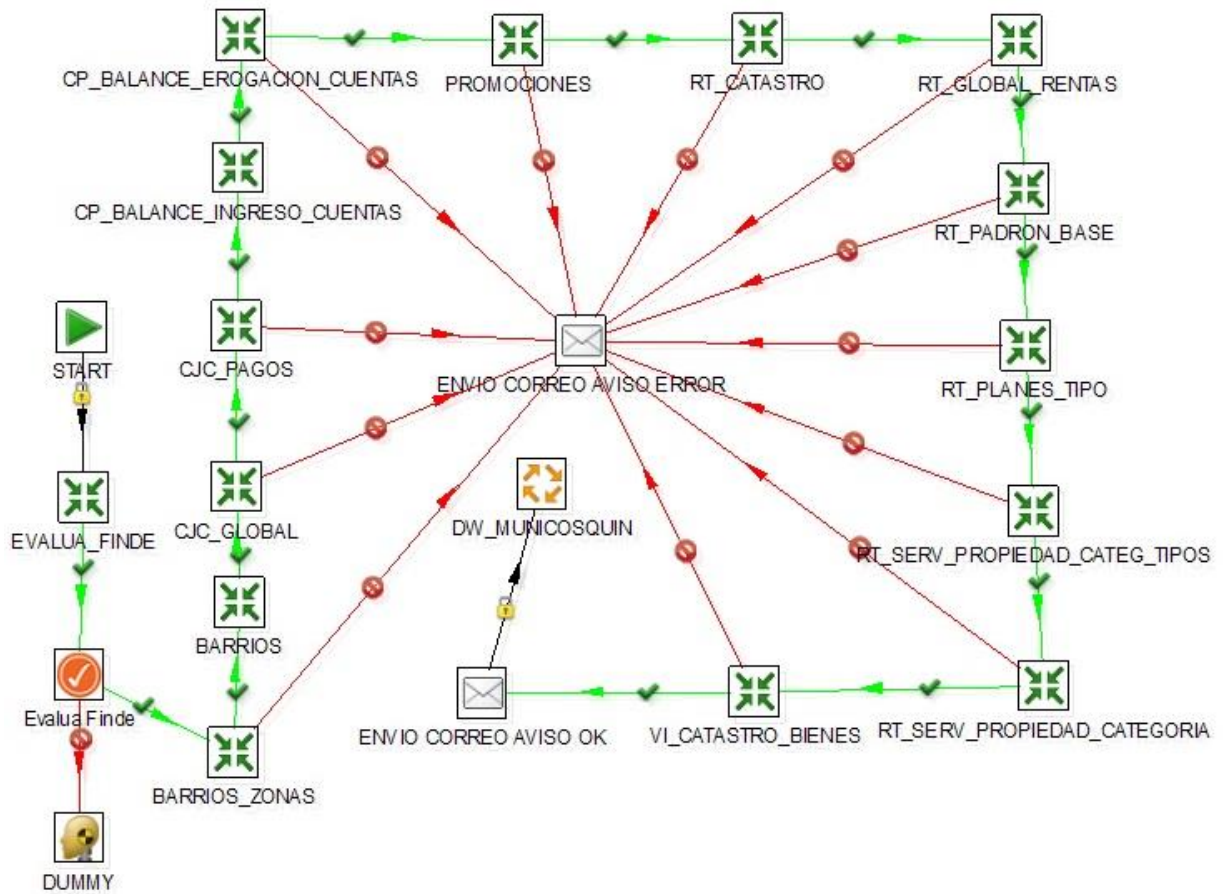
Como se mencionó anteriormente y de acuerdo a la selección de herramientas realizada en la sección 2 de la metodología, se utilizará el software de Pentaho, "Pentaho Data Integration (PDI)" o anteriormente conocido como KETTLE para la extracción de datos del OLTP y creación del Staging Area, como para la extracción, limpieza, transformación y carga de datos en el Warehouse.

Utilizando la aplicación de Pentaho, PDI, se crea el trabajo necesario para constituir el Staging Area.

El trabajo realizado se compone de transformaciones (cada una de las cuales mueve la tabla que lleva su nombre desde el OLTP al Staging Area) y otros pasos los cuales se ejecutan de manera secuencial y finalizan con el envío de un e-mail al responsable de controlar que el proceso se ejecute correctamente. En cualquier caso, éxito o falla, el e-mail informa la situación.



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



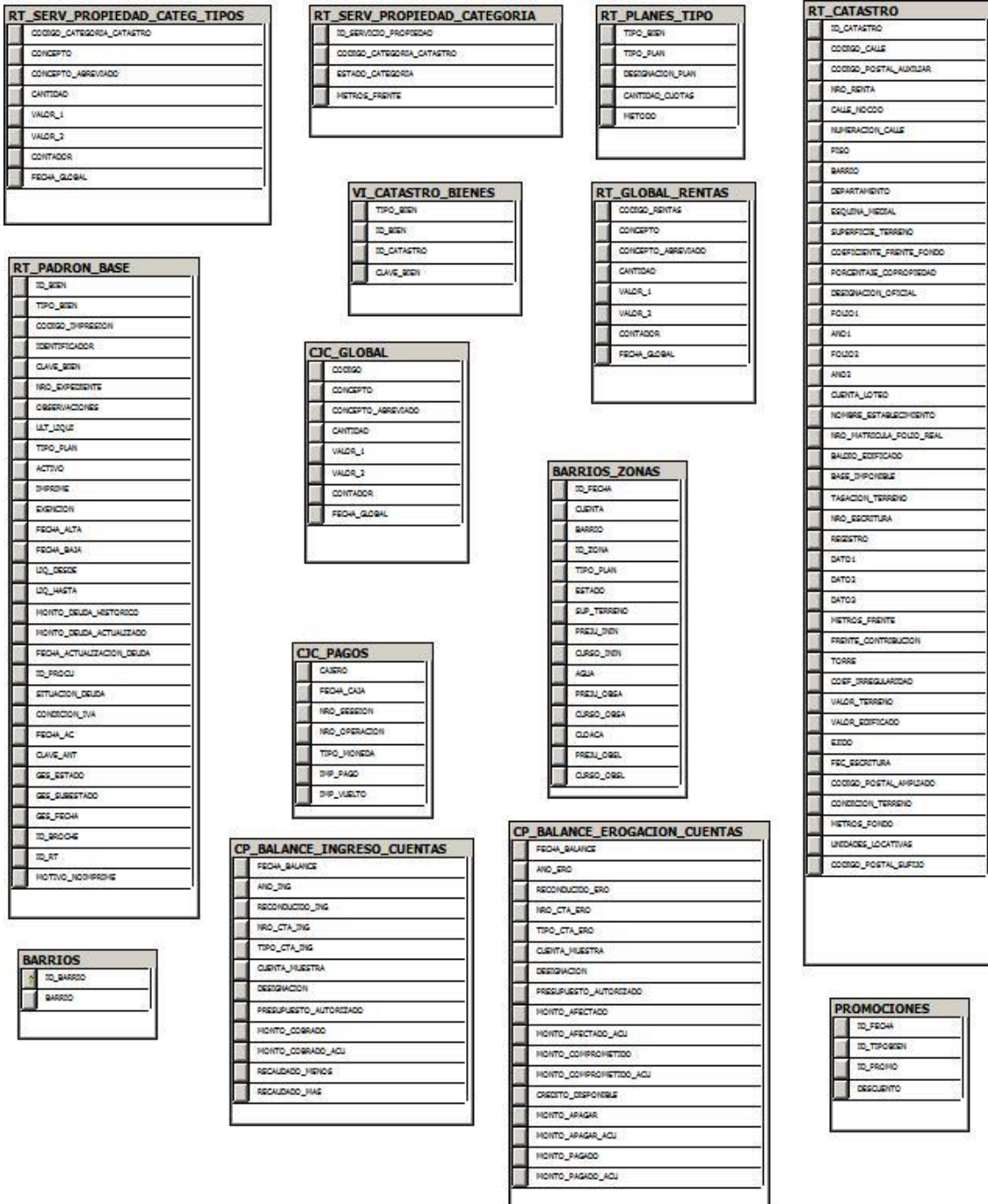
Una vez ejecutado todo el trabajo de manera satisfactoria los datos necesarios del sistema operacional son trasladados al área intermedia, almacenándolos temporalmente en una base de datos con el nombre STAGING_AREA_DW_MUNICOSQUIN.

El trabajo que puebla el Staging Área se configura para ejecutarse en un horario donde no afecte el rendimiento del sistema operacional, en caso de este proyecto en particular se realizará a las 19hs, luego de que las actividades municipales diarias han finalizado.

El Diagrama Entidad-Relación de la base de datos que conforma el Staging Area es presentado a continuación. El mismo no posee relaciones de tablas, puesto que lo que se busca es extraer de manera rápida y eficiente los datos del sistema operacional, afectando lo menos posible al mismo.



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



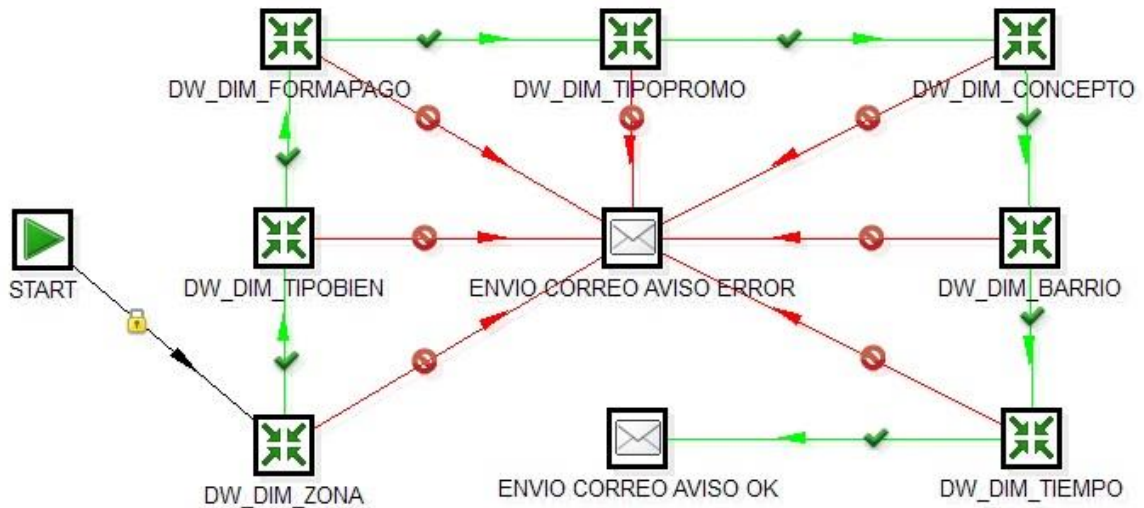
En este momento se dispone únicamente de los datos necesarios para el DW siendo posible manipularlos sin intervenir en el rendimiento de los sistemas operacionales. Por ello se puede ver que el trabajo que crea y puebla el Staging Area en su último paso contiene la llamada a otro trabajo que procederá a poblar las tablas de hechos del DataWarehouse.

Continuando con el uso de Kettle se crea un trabajo que contendrá las distintas transformaciones que, crearán y poblarán las dimensiones con los datos correspondientes. Este trabajo es independiente de otros ya que solo se ejecutará una sola vez para realizar la carga inicial de los datos de las dimensiones, luego para su



actualización se utilizara otro trabajo que estará conectado al trabajo que carga y actualiza los datos de los hechos y tendrá una ejecución periódica.

El trabajo que realiza la carga inicial de las dimensiones se presenta a continuación:

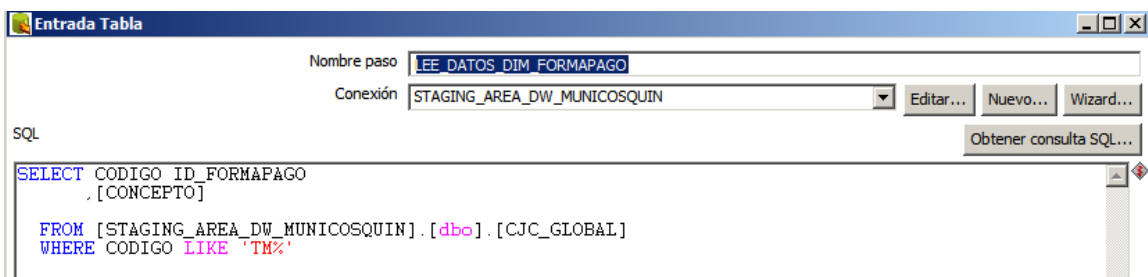


Este trabajo ejecuta de modo secuencial cada una de las transformaciones que se observan. Durante su ejecución las transformaciones crearán las tablas de dimensiones del modelo lógico y cargarán los datos correspondientes en dichas tablas.

Al igual que en el trabajo anterior, finaliza con el envío de un e-mail informando el resultado del proceso.

Las diferentes transformaciones y consultas SQL que se ejecutan en su interior para lograr las tablas de dimensiones son:

Dimensión: "Dim_FormaPago"



Dimensión: "Dim_Concepto"



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



Entrada Tabla

Nombre paso: LEE_DATOS_DIM_CONCEPTO

Conexión: STAGING_AREA_DW_MUNICOSQUIN

SQL

```

SELECT DISTINCT
ANO_ING + [NRO_CTA_ING] AS ID_COMPUESTA
,[DESIGNACION] DESIG1
,CASE WHEN LEN(BIC.NRO_CTA_ING) > 2
THEN
(SELECT DISTINCT (BI1.DESIGNACION) FROM CP_BALANCE_INGRESO_CUENTAS BI1
WHERE BI1.ANO_ING = BIC.ANO_ING
AND BI1.NRO_CTA_ING = SUBSTRING(BIC.NRO_CTA_ING, 1, LEN(BIC.NRO_CTA_ING) - 2))
ELSE 'S/D'
END DESIG2
,CASE WHEN LEN(BIC.NRO_CTA_ING) > 4
THEN
(SELECT DISTINCT (BI1.DESIGNACION) FROM CP_BALANCE_INGRESO_CUENTAS BI1
WHERE BI1.ANO_ING = BIC.ANO_ING
AND BI1.NRO_CTA_ING = SUBSTRING(BIC.NRO_CTA_ING, 1, LEN(BIC.NRO_CTA_ING) - 4))
ELSE 'S/D'
END DESIG3
,CASE WHEN LEN(BIC.NRO_CTA_ING) > 6
THEN
(SELECT DISTINCT (BI1.DESIGNACION) FROM CP_BALANCE_INGRESO_CUENTAS BI1
WHERE BI1.ANO_ING = BIC.ANO_ING
AND BI1.NRO_CTA_ING = SUBSTRING(BIC.NRO_CTA_ING, 1, LEN(BIC.NRO_CTA_ING) - 6))
ELSE 'S/D'
END DESIG4
FROM [STAGING_AREA_DW_MUNICOSQUIN].[dbo].[CP_BALANCE_INGRESO_CUENTAS] BIC
  
```

Dimensión: "Dim_Tipobien"



Entrada Tabla

Nombre paso: LEE DATOS DIM TIPOBIEN

Conexión: STAGING_AREA_DW_MUNICOSQUIN

SQL

```

SELECT [CODIGO_RENTAS] ID_TIPOBIEN
,CONCEPTO DESCRIPCION
FROM [STAGING_AREA_DW_MUNICOSQUIN].[dbo].[RT_GLOBAL_RENTAS]
WHERE LEN(CODIGO_RENTAS) = 4
  
```

Dimensión: "Dim_Barrío"



Entrada Tabla

Nombre paso: LEE_DATOS_DIM_BARRIOS

Conexión: STAGING_AREA_DW_MUNICOSQUIN

SQL

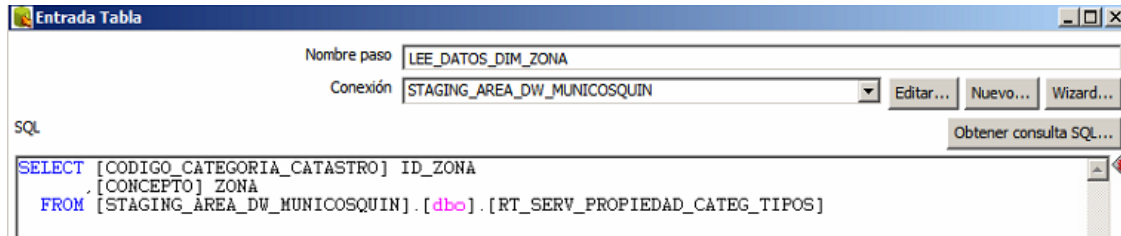
```

SELECT
ID_BARRIO
BARRIO
FROM BARRIOS
  
```

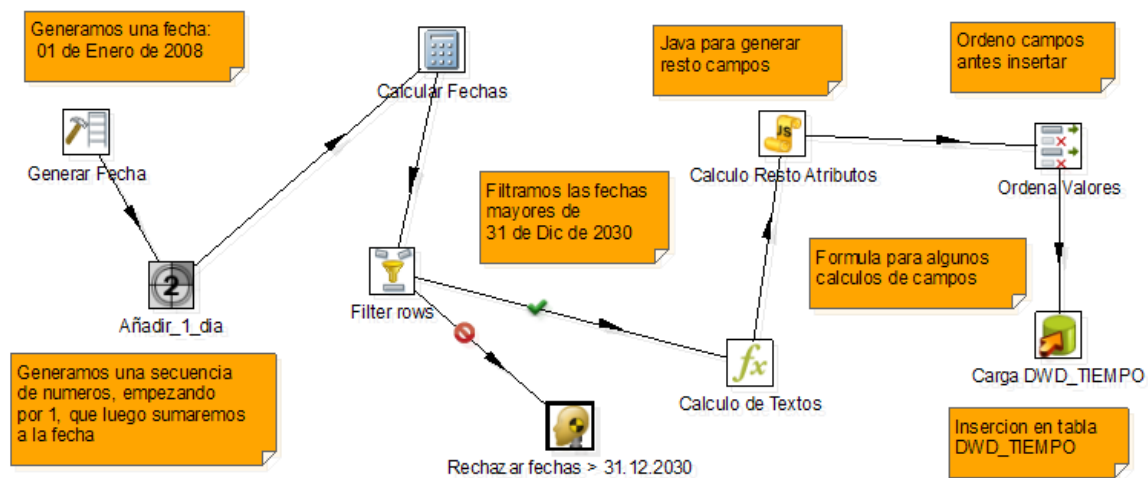
Dimensión: "Dim_Zona"



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



Dimensión: "Dim_Tiempo"



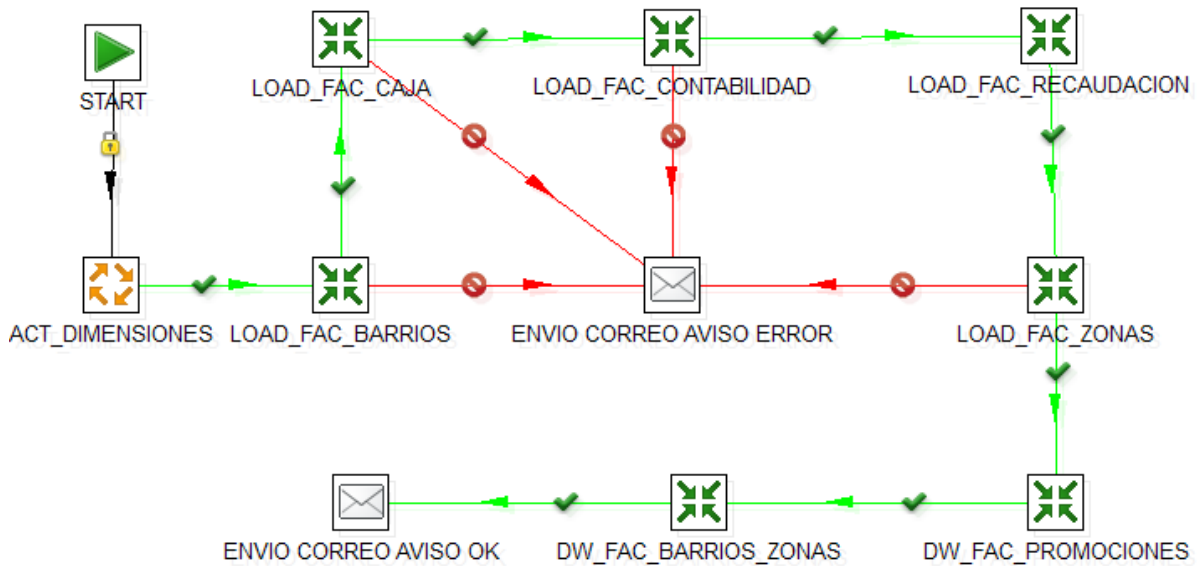
Dimensión: "Tipo_Promo"



Habiendo expuesto todas las transformaciones que crean las dimensiones, y el trabajo que ejecuta a las mismas, se procede a exponer el trabajo que llevara adelante la construcción de las diferentes tablas de hechos.

Como se mencionó anteriormente, este trabajo estará unido secuencialmente al trabajo que crea el Staging Area, de este modo cada vez que se ejecute el trabajo inicial continuará con la ejecución de éste.

Al igual que con las dimensiones, en primer lugar es necesario crear un trabajo que se ocupará de ejecutar las transformaciones que crearán las tablas de hechos. El trabajo creado recibe el nombre de LOAD_DWMUNICOSQUIN.



La primera vez que se ejecute el trabajo, cada transformación creará la tabla de hechos correspondiente. En las siguientes ejecuciones del trabajo, ya estando creadas las tablas de hechos, cada transformación actualizará los datos.

Es posible observar que al iniciar el trabajo lo primero que realiza es llamar a otro trabajo ACT_DIMENSIONES el cual contendrá las transformaciones que se ocuparán de mantener actualizado los datos de las dimensiones. Este trabajo será expuesto más adelante.

Las diferentes transformaciones y consultas SQL que se ejecutan en el interior del trabajo de hechos que permitirán lograr las tablas de hechos son:

Tabla de Hechos: “FAC_CAJA”

Transformación: LOAD_FAC_CAJA



En el primer se ejecuta la consulta SQL que obtendrá los datos desde la Staging Area y los pasará a la tabla de hechos FAC_CAJA. El paso del medio se ocupa de buscar la clave subrogada para conocer cuál es la versión actual y guardar está en la tabla de hechos.

La consulta SQL que realiza la extracción es:

```

“SELECT CONVERT(INT,CONVERT(VARCHAR, P.FECHA_CAJA, 112))
ID_FECHA, ‘TM’ + P.TIPO_MONEDA ID_FORMAPAGO, SUM(P.IMP_PAGO -
P.IMP_VUELTO) RECAUDADO, SUM(P.IMP_PAGO - P.IMP_VUELTO)/(SELECT
SUM(PG.IMP_PAGO - PG.IMP_VUELTO) FROM CJC_PAGOS PG WHERE
PG.FECHA_CAJA = P.FECHA_CAJA GROUP BY PG.FECHA_CAJA) PORCENTAJE
FROM CJC_PAGOS P GROUP BY P.FECHA_CAJA, P.TIPO_MONEDA ORDER BY
P.FECHA_CAJA”

```



Tabla de Hechos: “FAC_RECAUDACION”

Transformación: *LOAD_FAC_RECAUDACION*



En el primer se ejecuta la consulta SQL que obtendrá los datos desde la Staging Area y los pasará a la tabla de hechos FAC_RECAUDACION. El paso del medio se ocupa de buscar la clave subrogada para conocer cuál es la versión actual y guardar está en la tabla de hechos.

La consulta SQL que realiza la extracción es:

```

    "SELECT CONVERT(INT,CONVERT(VARCHAR, FECHA_BALANCE, 112))
    ID_FECHA ,ANO_ING + NRO_CTA_ING ID_COMPUESTA ,[MONTO_COBRADO]
    ,[MONTO_COBRADO_ACU] FROM [CP_BALANCE_INGRESO_CUENTAS] ORDER
    BY FECHA_BALANCE"
  
```

Tabla de Hechos: “FAC_CONTABILIDAD”

Transformación: *LOAD_FAC_CONTABILIDAD*



En el primer paso se ejecuta la consulta SQL que obtendrá los datos desde la Staging Area y los pasara a la tabla de hechos FAC_CONTABILIDAD. Dicha consulta es:

```

    "SELECT CONVERT(INT,CONVERT(VARCHAR, [FECHA_BALANCE], 112))
    ID_FECHA ,ISNULL((([MONTO_COBRADO] - (SELECT EC.MONTO_PAGADO
    FROM CP_BALANCE_EROGACION_CUENTAS EC WHERE EC.ANO_ERO =
    IC.ANO_ING AND EC.FECHA_BALANCE = IC.FECHA_BALANCE AND
    EC.NRO_CTA_ERO = '201'))/[MONTO_COBRADO]),([MONTO_COBRADO] -
    (SELECT SUM(EC.MONTO_PAGADO) FROM
    P_BALANCE_EROGACION_CUENTAS EC WHERE EC.ANO_ERO = IC.ANO_ING
    AND EC.FECHA_BALANCE = IC.FECHA_BALANCE AND CONVERT(BIGINT,
    EC.NRO_CTA_ERO) > '2000' AND CONVERT(BIGINT, EC.NRO_CTA_ERO) <
    '3000'))/[MONTO_COBRADO]) TAC ,ISNULL((SELECT IC1.MONTO_COBRADO
    FROM CP_BALANCE_INGRESO_CUENTAS IC1 WHERE IC1.ANO_ING =
    IC.ANO_ING AND IC1.FECHA_BALANCE = IC.FECHA_BALANCE AND
    IC1.NRO_CTA_ING = '10101'))/(SELECT EC.MONTO_PAGADO FROM
    CP_BALANCE_EROGACION_CUENTAS EC WHERE EC.ANO_ERO = IC.ANO_ING
    AND EC.FECHA_BALANCE = IC.FECHA_BALANCE AND EC.NRO_CTA_ERO =
    '201'),(SELECT IC1.MONTO_COBRADO FROM
    P_BALANCE_INGRESO_CUENTAS IC1 WHERE IC1.ANO_ING = IC.ANO_ING
    AND IC1.FECHA_BALANCE = IC.FECHA_BALANCE AND IC1.NRO_CTA_ING =
    '10101'))/(SELECT SUM(EC.MONTO_PAGADO) FROM
  
```

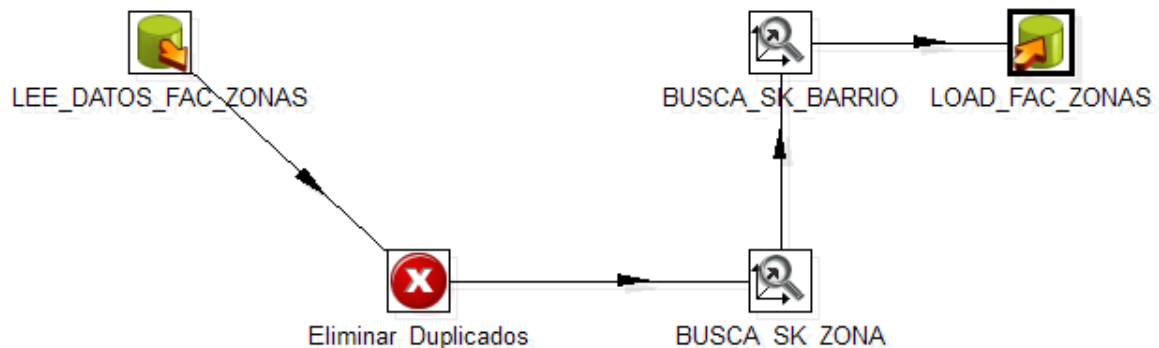


```

CP_BALANCE_EROGACION_CUENTAS EC WHERE EC.ANO_ERO = IC.ANO_ING
AND EC.FECHA_BALANCE = IC.FECHA_BALANCE AND CONVERT(BIGINT,
EC.NRO_CTA_ERO) > '2000' AND CONVERT(BIGINT, EC.NRO_CTA_ERO) <
'3000')) GAF, ISNULL((SELECT EC.MONTO_PAGADO FROM
CP_BALANCE_EROGACION_CUENTAS EC WHERE EC.ANO_ERO = IC.ANO_ING
AND EC.FECHA_BALANCE = IC.FECHA_BALANCE AND EC.NRO_CTA_ERO =
'20201')/(SELECT EC.MONTO_PAGADO FROM
CP_BALANCE_EROGACION_CUENTAS EC WHERE EC.ANO_ERO = IC.ANO_ING
AND EC.FECHA_BALANCE = IC.FECHA_BALANCE AND EC.NRO_CTA_ERO =
'2'),(SELECT SUM(EC.MONTO_PAGADO) FROM
CP_BALANCE_EROGACION_CUENTAS EC WHERE EC.ANO_ERO = IC.ANO_ING
AND EC.FECHA_BALANCE = IC.FECHA_BALANCE AND CONVERT(BIGINT,
EC.NRO_CTA_ERO) >= '200201' AND CONVERT(BIGINT, EC.NRO_CTA_ERO) <=
'290201')/(SELECT SUM(EC.MONTO_PAGADO) FROM
CP_BALANCE_EROGACION_CUENTAS EC WHERE EC.ANO_ERO = IC.ANO_ING
AND EC.FECHA_BALANCE = IC.FECHA_BALANCE AND CONVERT(BIGINT,
EC.NRO_CTA_ERO) >= '20' AND CONVERT(BIGINT, EC.NRO_CTA_ERO) < '30')
) IP FROM [CP_BALANCE_INGRESO_CUENTAS] IC WHERE NRO_CTA_ING =
'101' ORDER BY FECHA_BALANCE"
    
```

Tabla de Hechos: “FAC_BARRIOS_ZONAS”

Transformación: LOAD_FAC_BARRIOS



En el primer paso se ejecuta la consulta SQL que obtendrá los datos desde la Staging Area y los pasara a la tabla de hechos FAC_BARRIOS_ZONAS. Los pasos intermedios se encargan de eliminar los registros duplicados y de buscar las versiones actuales (por medio de claves subrogadas) de los barrios y las zonas.

La consulta SQL que extrae los datos es:

```

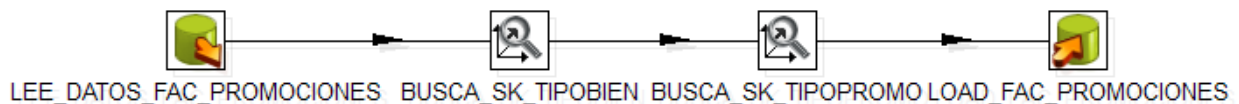
"SELECT ID_FECHA, B.ID_BARRIO, FDB.ID_ZONA, COUNT(0) AS TOTAL_ININ,
(SELECT COUNT(0) AS Expr1 FROM dbo.BARRIOS_ZONAS AS F WHERE
ID_ZONA = FDB.ID_ZONA) AND F.BARRIO = FDB.BARRIO AND (CURSO_ININ =
0) AND (PREJU_ININ = 0)) AS S_DEUDA_ININ, (SELECT COUNT(0) AS Expr1
FROM dbo.BARRIOS_ZONAS AS F WHERE (ID_ZONA = FDB.ID_ZONA) AND
F.BARRIO = FDB.BARRIO AND (AGUA = 1)) AS TOTAL_OBSA, (SELECT
COUNT(0) AS Expr1 FROM dbo.BARRIOS_ZONAS AS F WHERE (ID_ZONA =
FDB.ID_ZONA) AND F.BARRIO = FDB.BARRIO AND (CURSO_OBSA = 0) AND
    
```



```
(PREJU_OBSA = 0) AND (AGUA = 1)) AS S_DEUDA_OBSA, (SELECT COUNT(0)
AS Expr1 FROM dbo.BARRIOS_ZONAS AS F WHERE (ID_ZONA = FDB.ID_ZONA)
AND F.BARRIO = FDB.BARRIO AND (CLOACA = 1)) AS TOTAL_OBSL, (SELECT
COUNT(0) AS Expr1 FROM dbo.BARRIOS_ZONAS AS F WHERE (ID_ZONA =
FDB.ID_ZONA) AND F.BARRIO = FDB.BARRIO AND (CURSO_OBSL = 0) AND
(PREJU_OBSL = 0) AND (CLOACA = 1)) AS S_DEUDA_OBSL, (SELECT COUNT(0)
AS Expr1 FROM dbo.BARRIOS_ZONAS AS F WHERE (F.ID_ZONA =
FDB.ID_ZONA) AND (F.BARRIO = FDB.BARRIO) AND (F.ESTADO NOT LIKE
'BALDIO')) AS EDIFICADO, SUM(FDB.SUP_TERRENO) AS SUP_M2, GETDATE()
FECHA FROM dbo.BARRIOS_ZONAS AS FDB, dbo.BARRIOS AS B WHERE
FDB.BARRIO = B.BARRIO AND FDB.BARRIO NOT LIKE " GROUP BY ID_FECHA,
FDB.BARRIO, B.ID_BARRIO, ID_ZONA ORDER BY ID_FECHA, FDB.BARRIO"
```

Tabla de Hechos: “FAC_PROMOCIONES”

Transformación: LOAD_FAC_PROMOCIONES



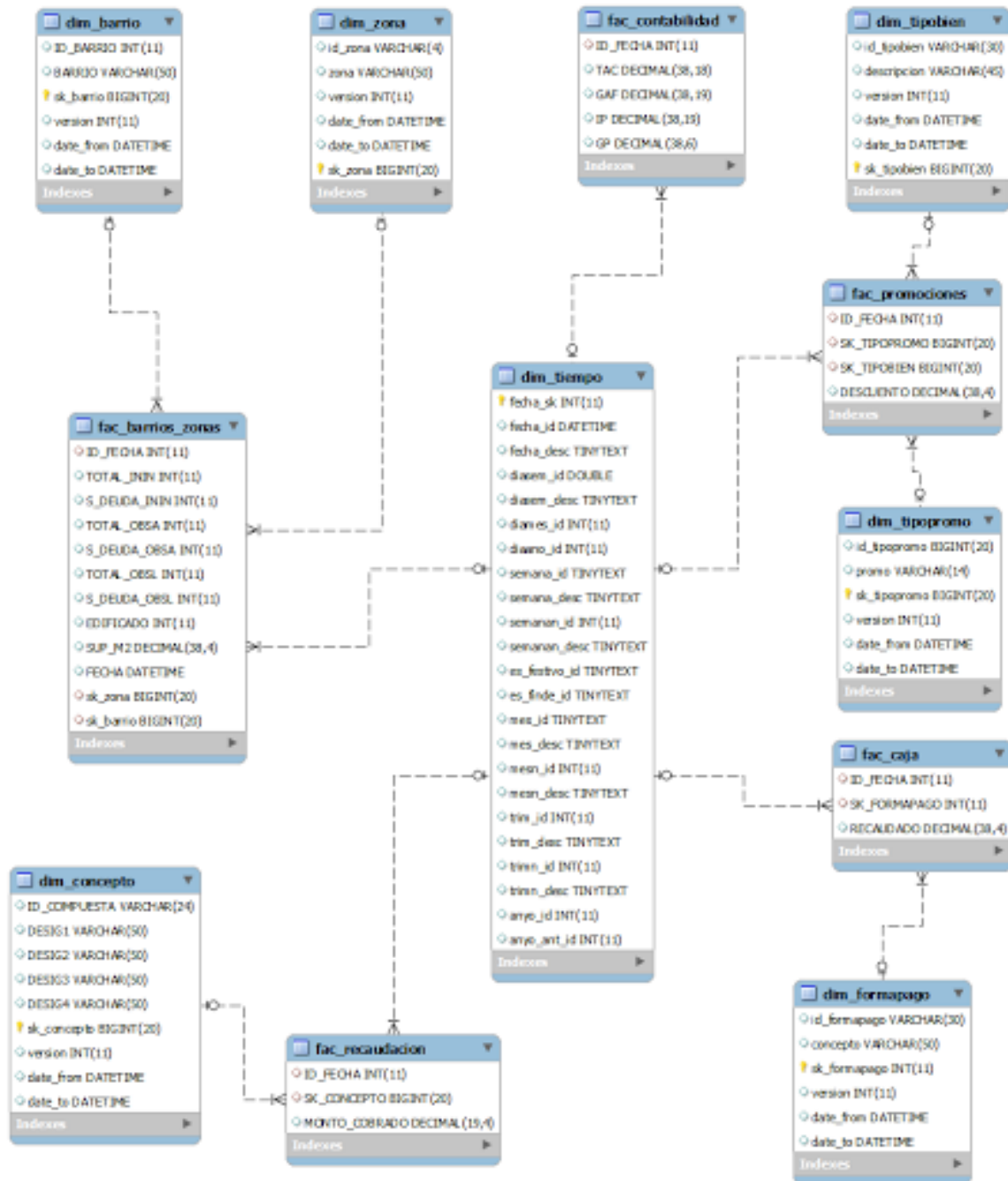
En el primer paso se ejecuta la consulta SQL que obtendrá los datos desde la Staging Area y los pasara a la tabla de hechos FAC_PROMOCIONES.

Los pasos intermedios se encargan de buscar por medio de claves subrogadas las versiones actuales del tipo de bien y tipo de promoción, para que sean estas las cargadas en la tabla de hechos.

La consulta SQL que extrae los datos es:

```
“SELECT ID_FECHA, ID_TIPOBIEN, ID_PROMO, SUM(DESCUENTO)
DESCUENTO FROM PROMOCIONES GROUP BY ID_FECHA, ID_TIPOBIEN,
ID_PROMO ORDER BY ID_FECHA, ID_TIPOBIEN, ID_PROMO”
```

Ahora que se han creado los trabajos y transformaciones necesarios para extraer los datos del OLTP, se han pasado al Staging Area, y nuevamente se han extraído para ser limpiados, transformados y cargados en el Data Warehouse en forma de tablas de dimensiones y hechos, es posible decir que el modelo lógico ha tomado forma física en la base de datos llamada DW_MUNICOSQUIN que posee el siguiente diagrama entidad-relación:



4.6.3.2 Actualización

Si bien el DW ha sido creado y poblado, los datos de las tablas de hechos y en algunos casos de las dimensiones se irán modificando a lo largo del tiempo, es por ello que se deben establecer políticas y estrategias de actualización o refresco de datos. Las políticas de actualización que se han convenido con los usuarios son las siguientes:

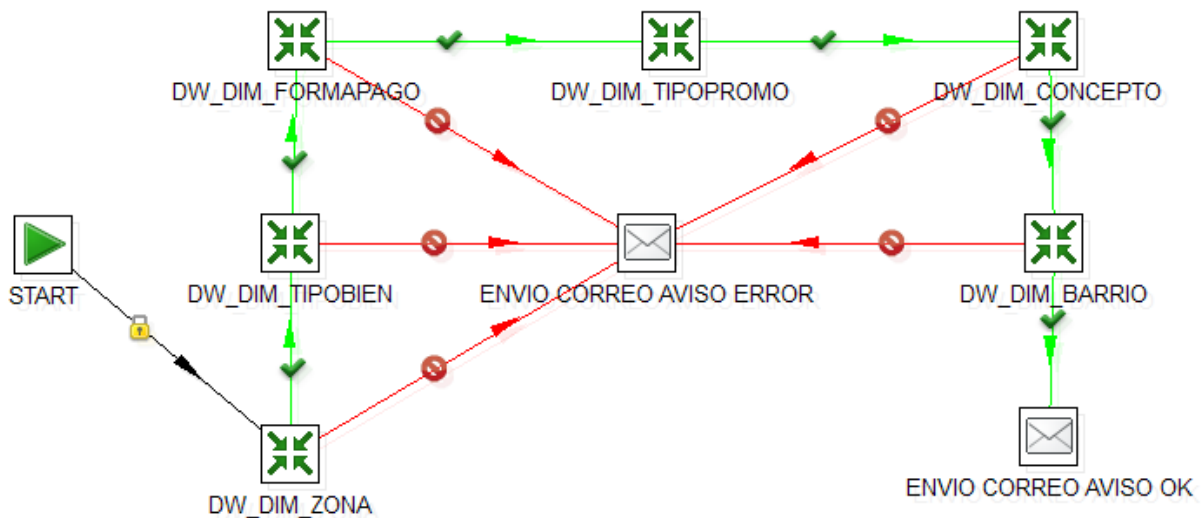
- ✓ La información se refrescara todos los días a las 19hs. Esto se acordó debido a que es el momento en que el sistema operacional no esta operativo, y es posible trasladar los datos a la Staging Area sin afectar el rendimiento.
- ✓ Los datos de la tabla de hechos se cargaran totalmente cada vez.



- ✓ Los datos de las dimensiones se cargaran una vez al iniciar el proceso y se actualizarán, si sufren alguna modificación todos los meses.
- ✓ Las tablas de dimensiones del proyecto, son consideradas tablas de dimensiones lentamente cambiantes o SCD (Slowly Changing Dimensions), esto significa que sus datos tienden a modificarse a través del tiempo, ya sea de forma ocasional o constante, o implique a un solo registro o la tabla completa.
- ✓ Se utilizan claves subrogadas para mantener las versiones de los distintos datos de las dimensiones.

Para mantener actualizadas a las dimensiones, se creó un nuevo trabajo encargado de esta función y se colocó su ejecución previa a la de las transformaciones que cargan los hechos (lo cual fue mostrado en el punto previo).

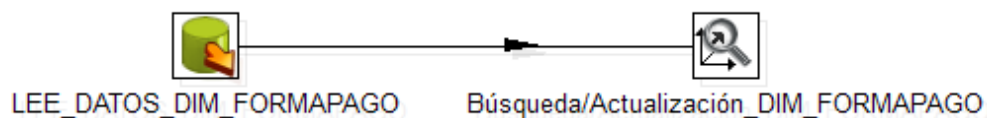
El trabajo encargado de mantener actualizadas a las dimensiones se constituye de la siguiente manera:



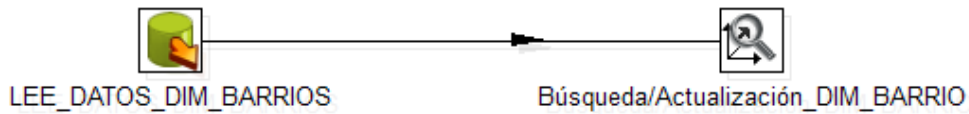
Este trabajo ejecuta de modo secuencial cada una de las transformaciones que se observan. Durante su ejecución las transformaciones actualizarán en caso de ser necesario las tablas de dimensiones del modelo lógico y finaliza con el envío de un e-mail informando el resultado del proceso.

Las diferentes transformaciones que se ejecutan en su interior para lograr la actualización de tablas de dimensiones son:

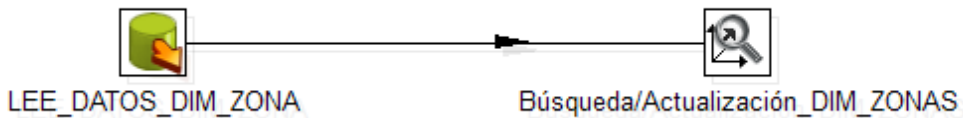
Actualización Dimensión: "Dim_FormaPago"



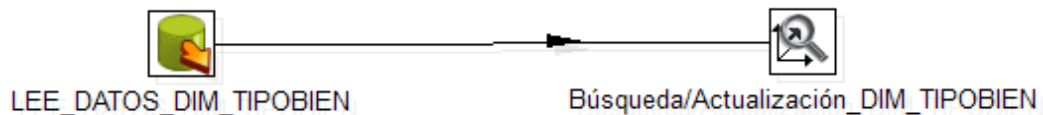
Actualización Dimensión: "Dim_Barríos"



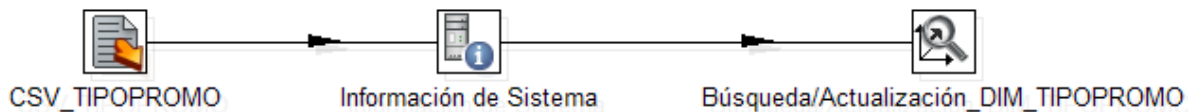
Actualización Dimensión: "Dim_Zonas"



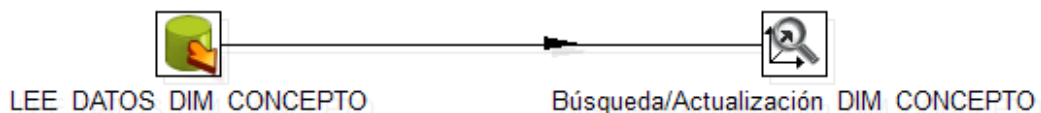
Actualización Dimensión: "Dim_TipoBien"



Actualización Dimensión: "Dim_TipoPromo"



Actualización Dimensión: "Dim_Concepto"



4.6.4 Creación de Cubos Multidimensionales

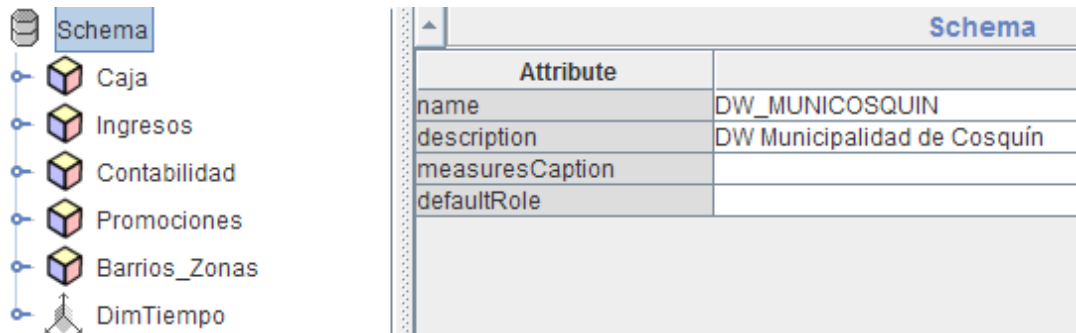
Una vez finalizada la carga y definidas las políticas de actualización del modelo de datos, se cuenta con el DataWarehouse poblado para poder explotar. Pero para poder aprovechar sus grandes ventajas y hacerlo de manera multidimensional es necesario crear justamente estructuras multidimensionales las cuales son denominadas: cubos multidimensionales.

Continuando con la arquitectura, el siguiente paso será crear dichos cubos, los cuales serán utilizados posteriormente para, a través de alguna herramienta de reporte y consulta ver y analizar la información que contiene el DW. En este momento nos situamos en la parte que ha sido denominada durante el marco teórico como: Data Warehouse Manager.



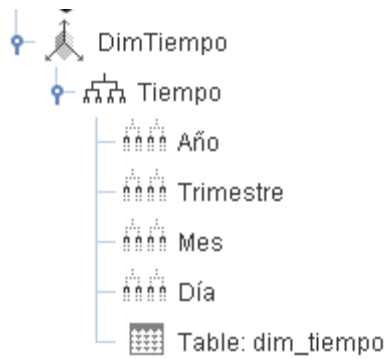
Para crear los cubos la herramienta utilizada como Data Warehouse Manager es Pentaho Schema Workbench.

Se comienza abriendo PSW y creando un nuevo esquema, dentro del cual, se construirán los cubos necesarios para consultar la información deseada.



Se puede observar que se ha creado un esquema con el nombre DW_MUNICOSQUIN, dentro del cual se crearon 5 cubos y una dimensión (DimTiempo) que será compartida dentro de los cubos.

La dimensión Tiempo (DimTiempo) compartida se compone de la siguiente manera:

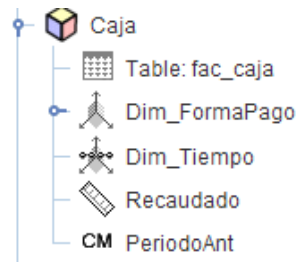


Los cubos creados son:

Caja:



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



Ingresos:



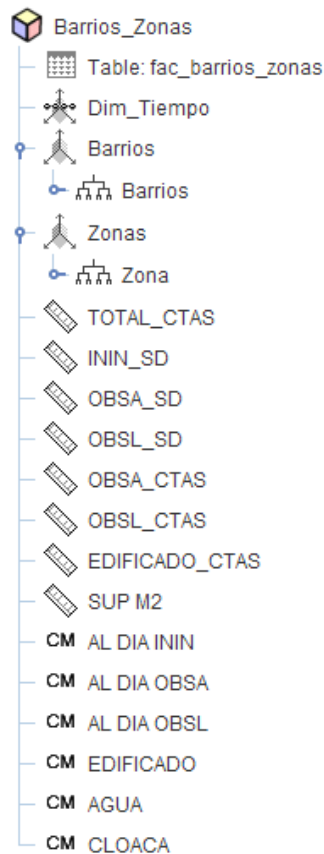
Contabilidad:



Barrios_Zonas:



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



Promociones:



Ahora que se han creado los cubos, los cuales contienen las medidas y dimensiones definidas durante el modelamiento del DW, es posible utilizar alguna herramienta de explotación de los mismos para analizar la información. En el proyecto, como se ha mencionado, se utilizará la aplicación de Pentaho BI Server con el plugin Saiku.

4.6.5 Herramientas de Reportes y Consulta



Ya habiendo creado y poblado el DW, así como definido los cubos multidimensionales que permitan explotar los datos, se ha llegado al punto en donde solo resta definir cuáles serán las herramientas de consultas y reportes mediante las cuales los usuarios observarán los datos que se encuentran en el Warehouse, para de este modo poder tomar las decisiones pertinentes.

Como se ha mencionado en este proyecto, los usuarios del DW se podrán conectar al mismo (teniendo en cuenta que será alojado en el servidor dedicado que posee la Secretaría de Economía y Finanzas Públicas) a través de la intranet de la Municipalidad utilizando los computadores que se encuentran físicamente en el área municipal o mediante acceso por VPN a la red interna.

La aplicación que se usará para conectarse al DW es Pentaho BI Server haciendo uno del plugin SAIKU.

A continuación se irán exponiendo distintas imágenes que muestran de qué manera acceden al sistema los distintos usuarios, con sus diferentes roles y permisos correspondientes. Se podrá observar como son respondidas todas las preguntas que se plantearon en el modelado del DW.

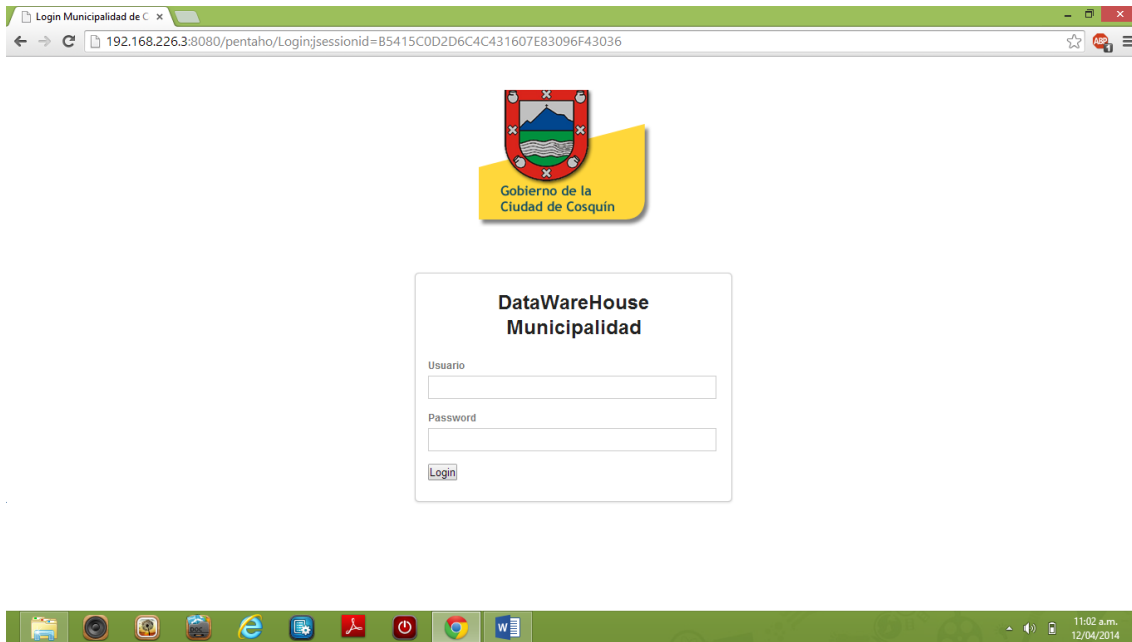
Para ingresar al sistema los usuarios, estando conectados a la red interna mediante cualquiera de las dos opciones mencionadas, deben ingresar en un browser de Internet la dirección: <http://192.168.226.3:8080/pentaho>.

El acceso ha sido testeado tanto con Internet Explorer como con Google Chrome y funciona perfectamente.

Una vez conectados aparecerá la pantalla de ingreso al sistema:



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



De acuerdo a cual sea la combinación de usuario y password ingresada (si es correcta) se habilitarán los permisos asociados al rol del usuario correspondiente. Mediante la consola de administración de Pentaho Server BI se han creado 4 usuarios los cuales tienen diferentes permisos asignados lo que les permitirán ingresar solamente a los reportes que les competen. Los usuarios creados han sido:

- ✓ Usuario: Admin
Rol: Administrador
Acceso Permitido: Creación, ejecución y edición de todos los informes y carpetas.
- ✓ Usuario: Intendente
Rol: Ejecutivo
Acceso Permitido: Ejecución de todos los informes y carpetas.
- ✓ Usuario: Eduardo (Secretario de la Secretaría de Economía y Finanzas Públicas).
Rol: Ejecutivo
Acceso Permitido: Ejecución de todos los informes y carpetas.
- ✓ Usuario: Pablo
Rol: RecursosTributarios
Acceso Permitido: Ejecución de informes del área de recursos tributarios.
- ✓ Usuario: Martin
Rol: Tesoreria-Contaduria
Acceso Permitido: Ejecución de informes del área de tesorería y contaduría.

Accediendo con el usuario Admin se podrán observar todos los informes creados los cuales han sido organizados en diferentes carpetas.

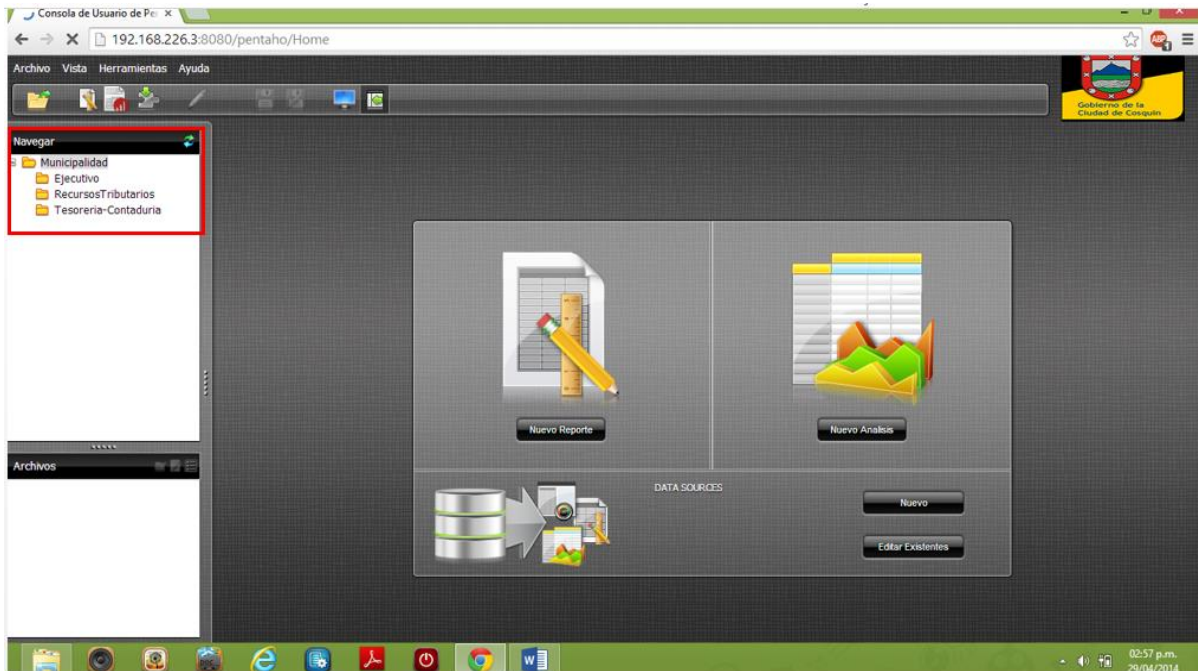


Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

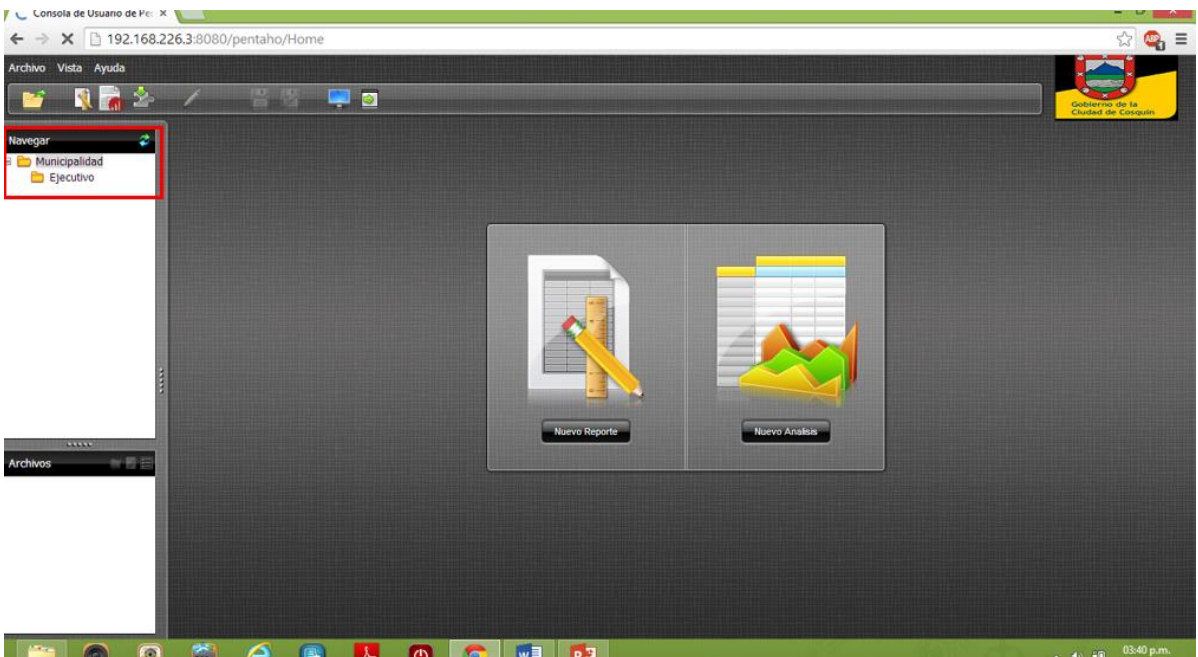
Se puede observar que existe una carpeta raíz la cual contiene otras 3 carpetas (Ejecutivo, Recursos Tributarios y Tesorería-Contaduría).

Si se ingresa con el usuario “Intendente” o “eduardo” se observará únicamente la carpeta “Ejecutivo” la cual contiene todos los informes, pero al hacerlo con el usuario pablo o martin solamente se observarán solamente las carpetas con los informes del área correspondientes. A continuación se exponen éstas imágenes.

Usuario Admin:

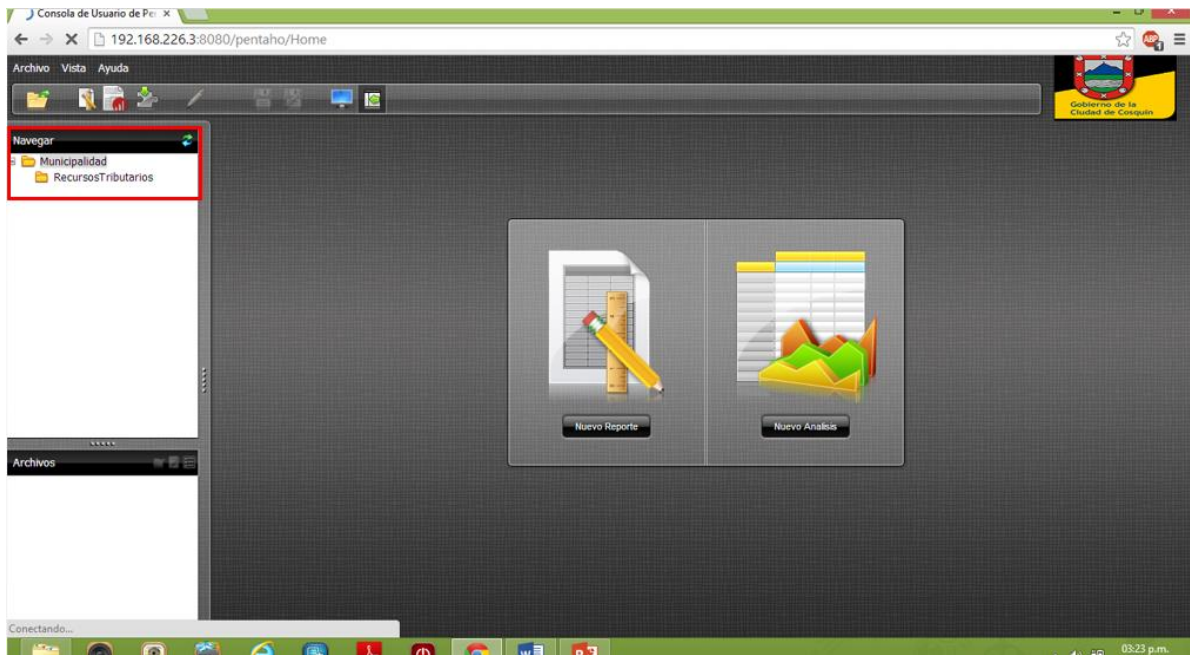


Usuario Intendente o eduardo:

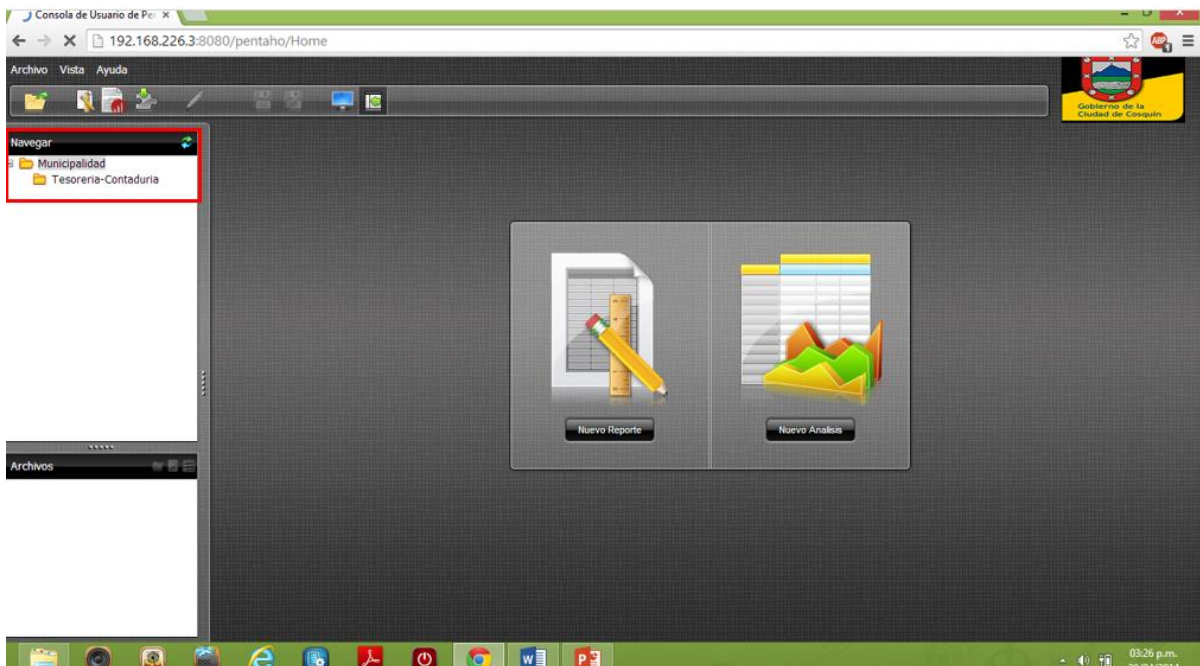




Usuario pablo:



Usuario martin:

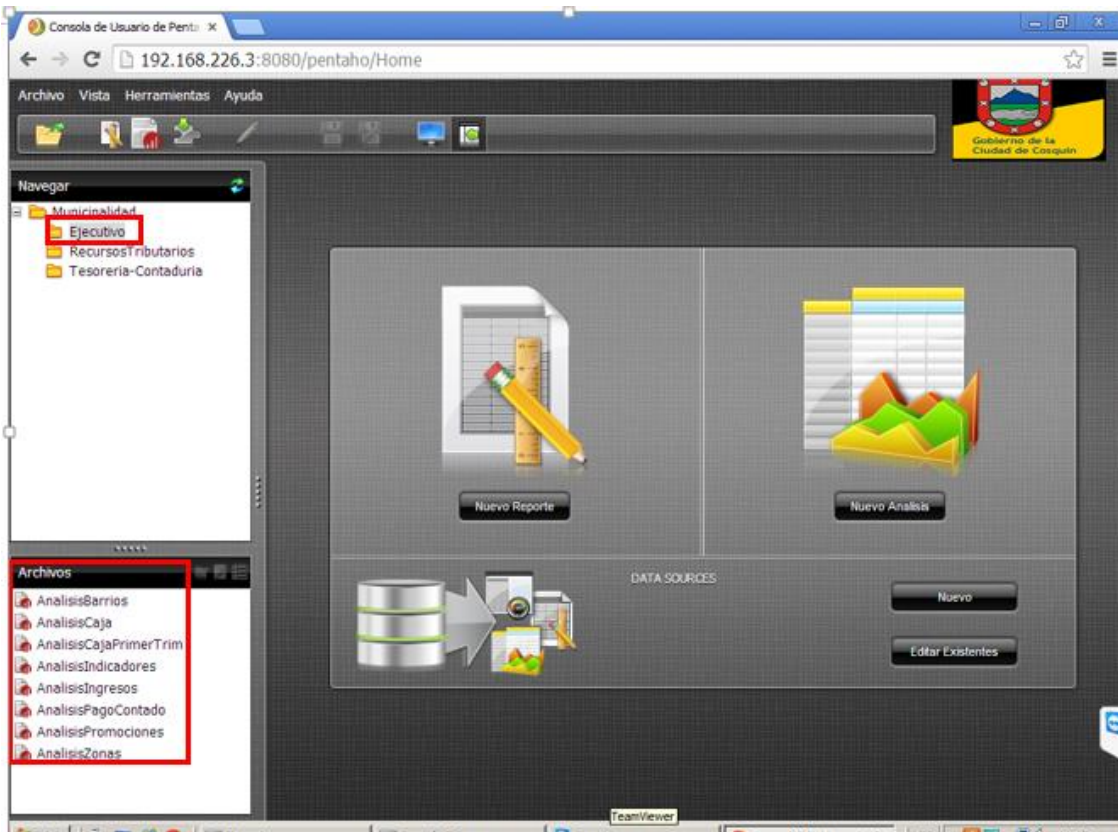


A continuación, habiendo accedido mediante el usuario Admin, se expondrán los informes que contiene cada carpeta demostrando que todos los indicadores que habían sido recolectados como requerimientos durante las primeras etapas de la metodología desarrollada han sido cumplidos.

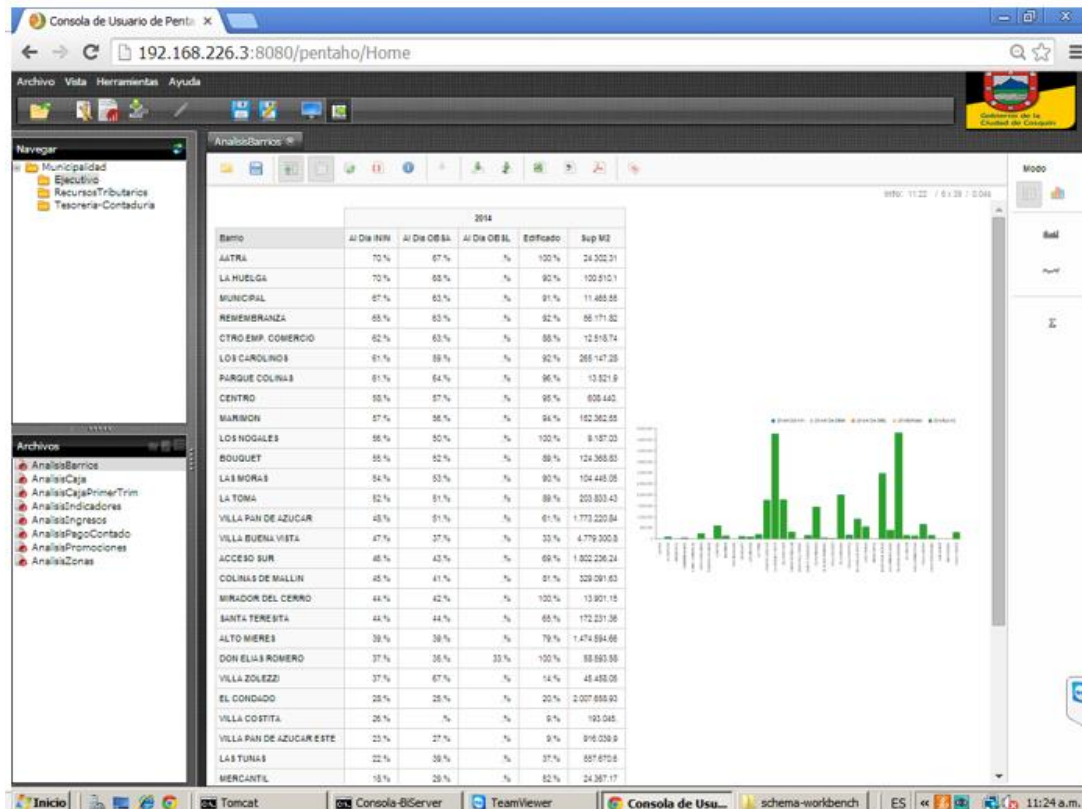
Carpeta Ejecutivo:



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



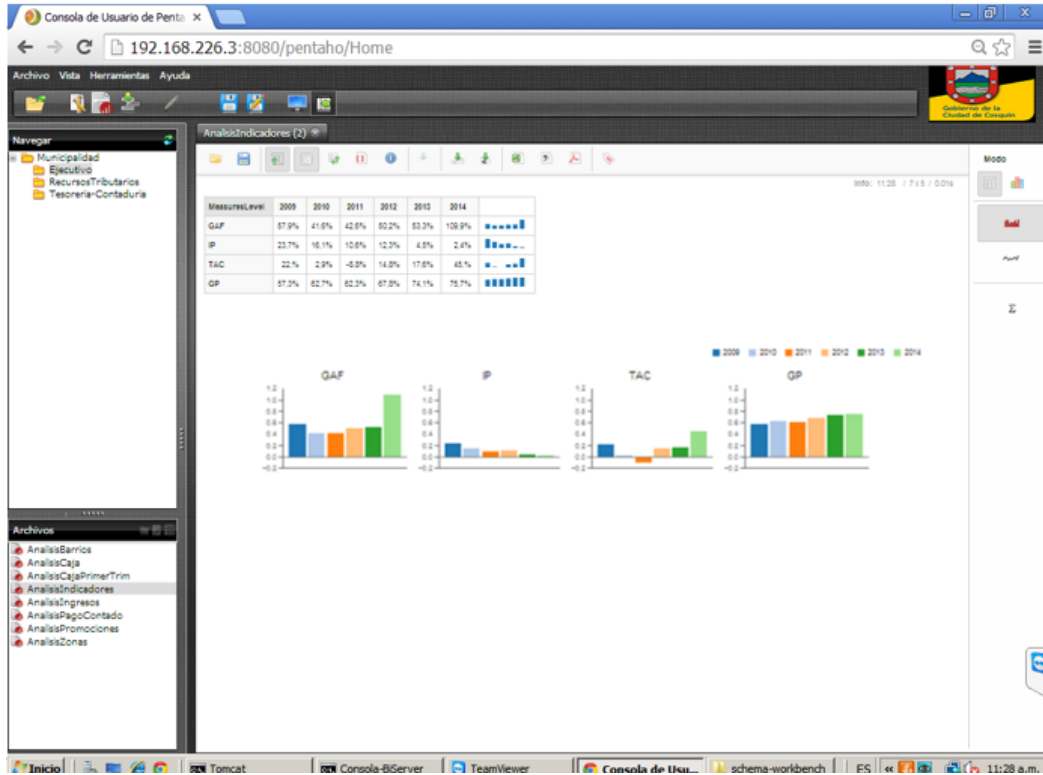
Informe AnalisisBarrios el cual presenta el estado de deuda tributaria discriminada por barrio:



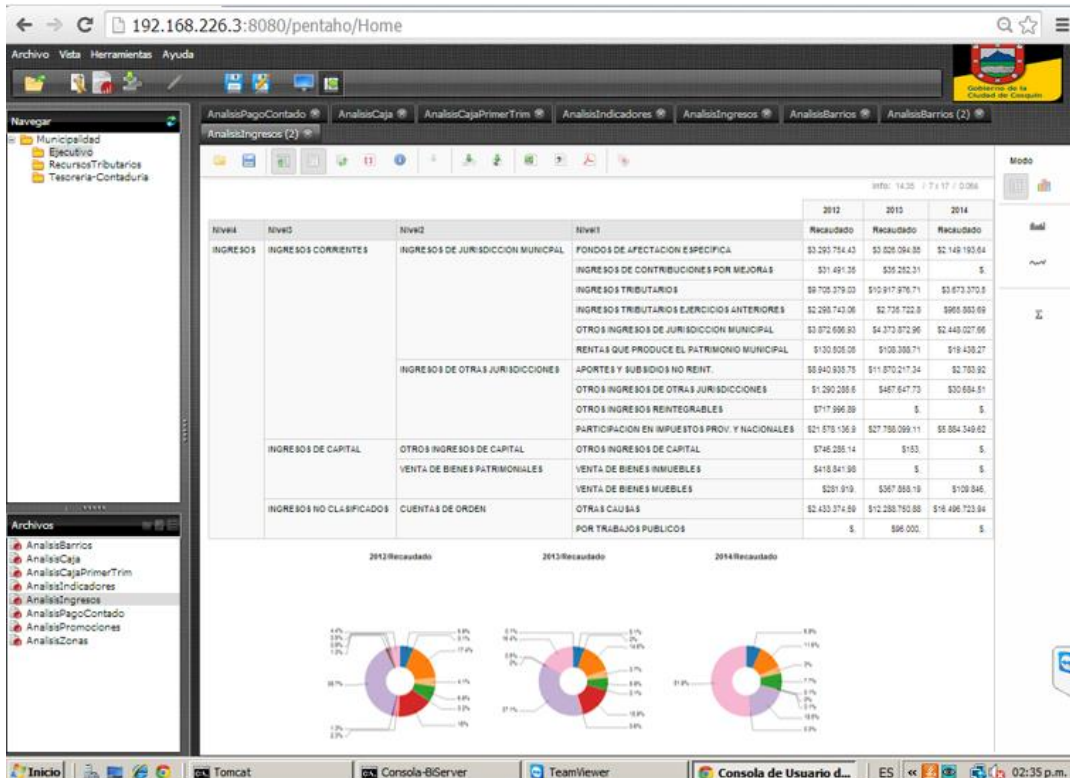


Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

Informe AnalisisIndicadores el cual presenta el análisis de los indicadores de contabilidad (grado de autonomía financiera, tasa de ahorro corriente, inversión física y gasto en personal):



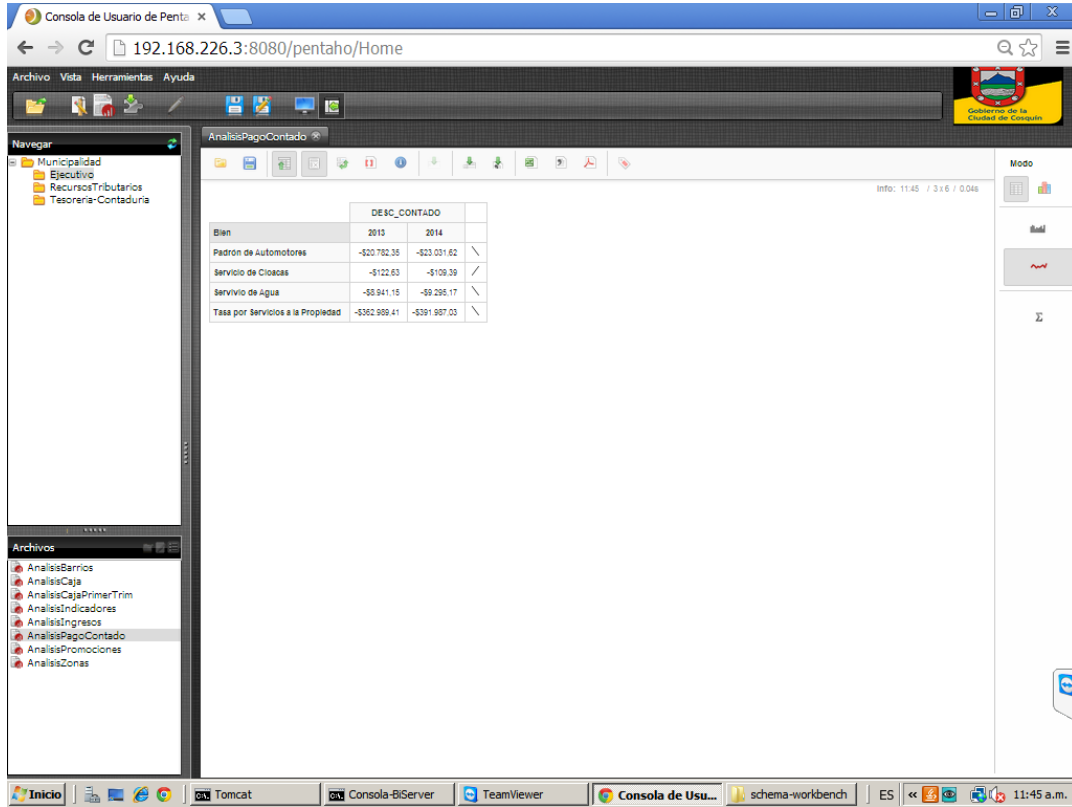
Informe AnalisisIngresos el cual presenta de qué manera se distribuyen las recaudaciones del municipio discriminados por concepto:



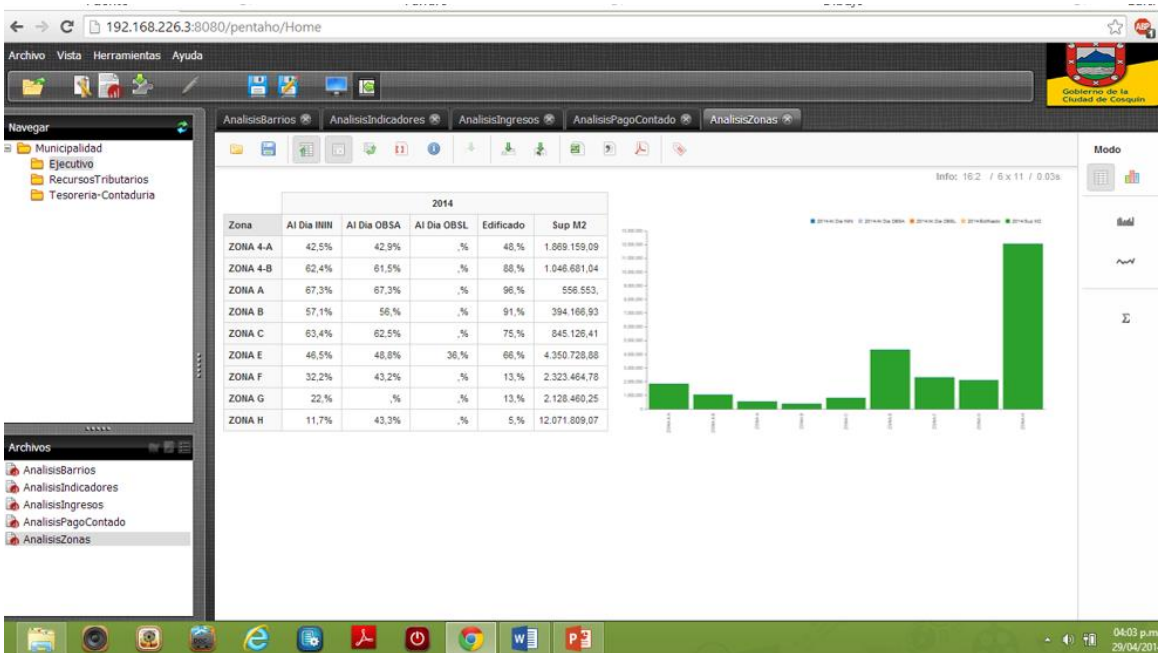


Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

Informe AnalisisPagoContado presenta los descuentos otorgados por la promoción de por pago de contado:



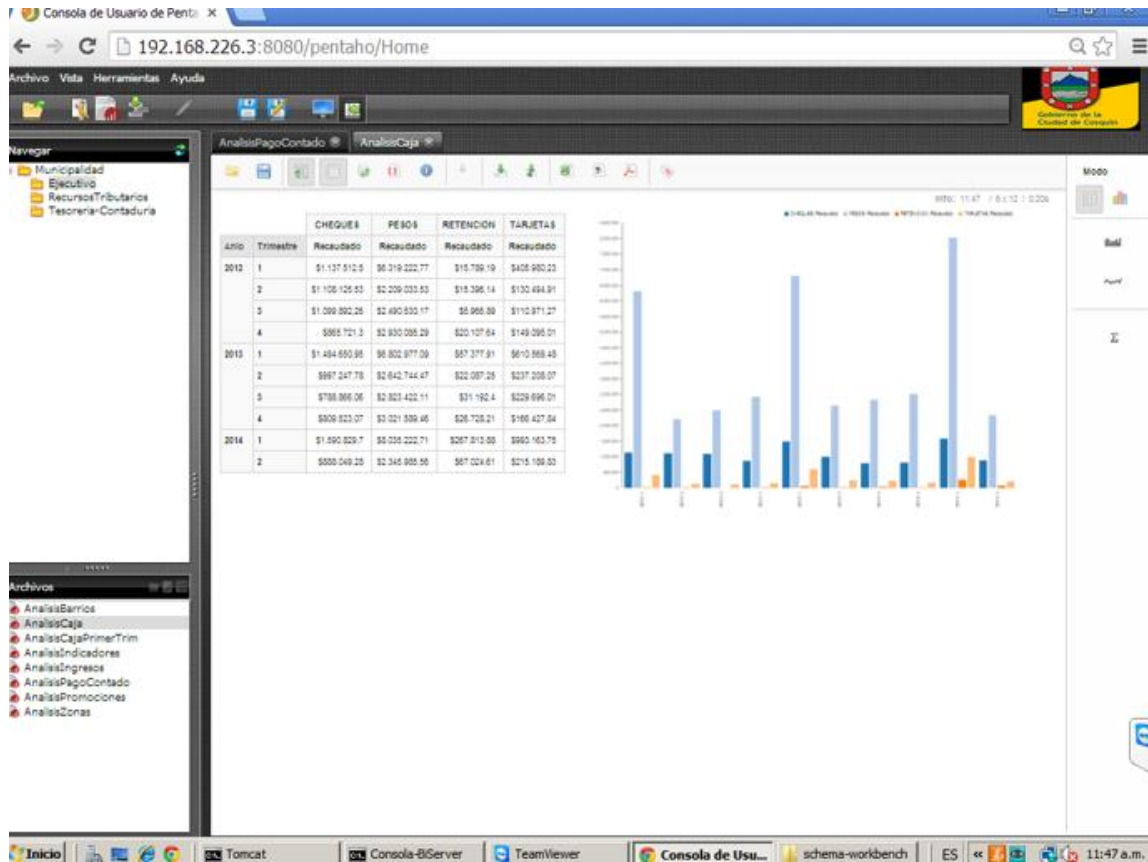
Informe AnalisisZonas el cual presenta el estado de deuda tributaria discriminada por zona:





Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

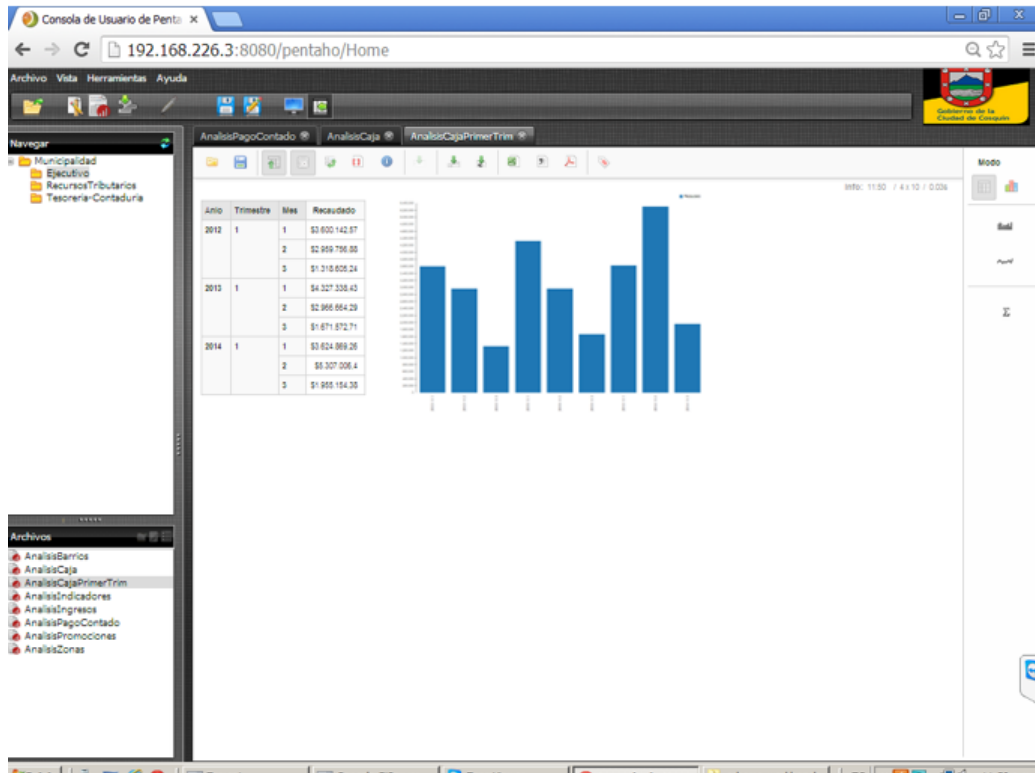
AnalisisCaja es uno de los informes más importantes, debido a que muestra la recaudación que ha tenido la Municipalidad discriminada de acuerdo a la forma de pago, pudiéndola observar por trimestre, mes y año.



AnalisisCajaPrimerTrim es un informe muy útil que se obtiene del anterior, éste permite observar y comparar la recaudación del primer trimestre de cada año puesto que es el trimestre que tiene mayor peso y prácticamente determina el ingreso anual del municipio. De esta manera conociendo cual es el ingreso del primer trimestre del corriente año y comparándolo con los anteriores es posible hacer una proyección anual y planificar los gastos de acuerdo al ingreso esperado.



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



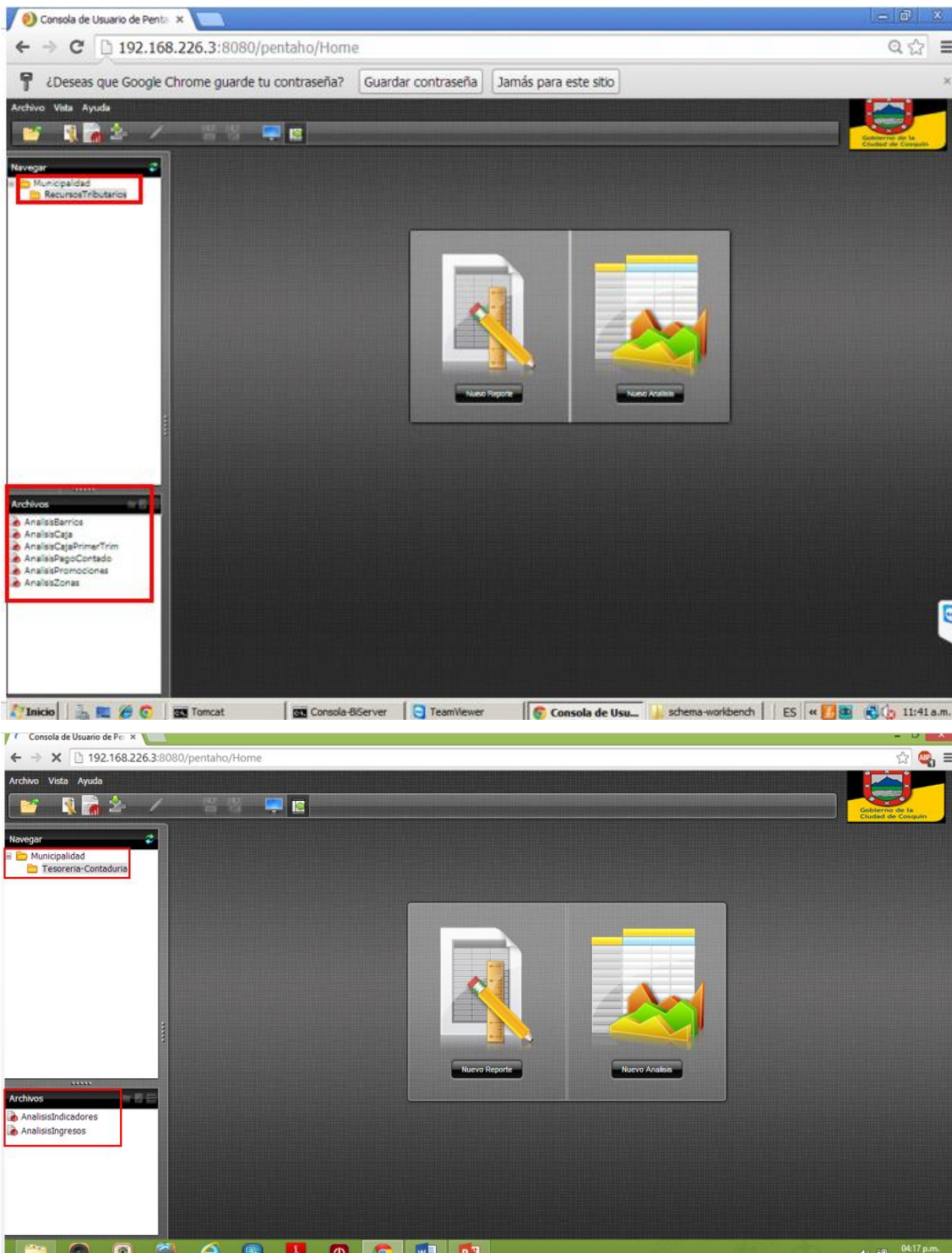
El informe AnalisisPromociones presenta los descuentos otorgados a los contribuyentes por pago en contado o cumplidor.

Año	Bien	Descuento	
		DESC_CONTADO	DESC_CUPLIDOR
2009	Padrón de Automotores	-\$5.247,47	
	Servicio de Cloacas	-\$12,5	\$1,46
	Servicio de Agua	-\$1.195,0	-\$5.675,73
	Tasa por Servicios a la Propiedad	-\$137.493,36	-\$60.199,15
2010	Padrón de Automotores	-\$9.473,36	
	Servicio de Cloacas	-\$42,45	-\$425,0
	Servicio de Agua	-\$4.779,49	-\$48.899,92
	Tasa por Servicios a la Propiedad	-\$197.199,89	-\$251.705,65
2011	Padrón de Automotores	-\$11.922,52	
	Servicio de Cloacas	-\$76,93	-\$1.038,76
	Servicio de Agua	-\$4.653,23	-\$129.758,02
	Tasa por Servicios a la Propiedad	-\$164.733,49	-\$418.087,47
2012	Padrón de Automotores	-\$15.880,64	
	Servicio de Cloacas	-\$106,3	-\$890,5
	Servicio de Agua	-\$7.277,6	-\$78.464,94
	Tasa por Servicios a la Propiedad	-\$345.152,91	-\$435.292,44
2013	Padrón de Automotores	-\$20.782,35	
	Servicio de Cloacas	-\$122,63	-\$777,2
	Servicio de Agua	-\$5.941,15	-\$112.761,69
	Tasa por Servicios a la Propiedad	-\$352.999,41	-\$521.991,26
2014	Padrón de Automotores	-\$23.031,62	
	Servicio de Cloacas	-\$109,39	
	Servicio de Agua	-\$9.295,17	
	Tasa por Servicios a la Propiedad	-\$391.987,03	



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

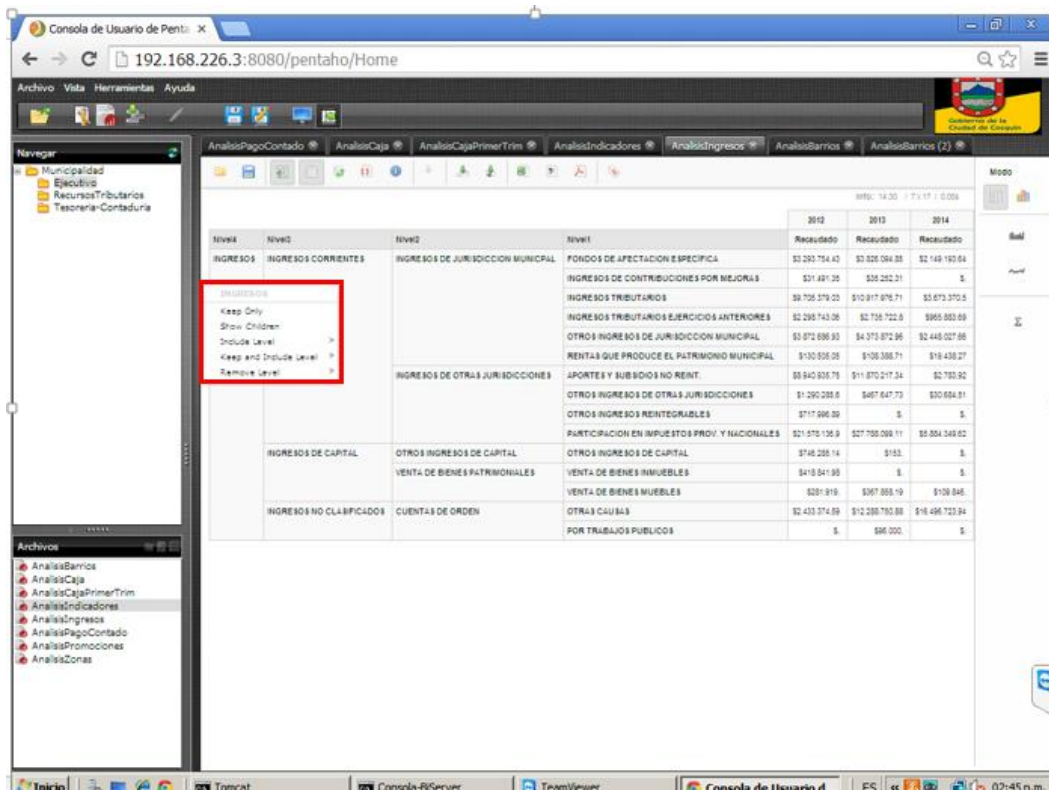
Finalmente se observará que en las carpetas de Recursos Tributarios y Tesorería-Contaduría solo se encuentran los informes correspondientes:



Como último aspecto es importante destacar que el plugin Saiku permite realizar las distintas operaciones multidimensionales que fueron explicadas durante el marco teórico, lo cual nos permite recorrer las jerarquías, agrupar o desagrupar los datos,



mover los ejes, etc. Simplemente haciendo clics derecho sobre algún atributo nos brinda diferentes opciones.



4.6.6 Instalación y Puesta en Marcha

Si bien el proceso ha finalizado y se ha expuesto todo el desarrollo de la arquitectura del DW, desde la extracción de los datos hasta el análisis de estos por parte de los usuarios, se podría suponer que el trabajo ha llegado a su fin, pero se expondrá como último punto, la manera en que se deben instalar las herramientas de Pentaho para que puedan ser accedidas.

Proceso de instalación de herramientas de Pentaho y MySQL:

En el anexo 1 se expone todo lo relativo a la instalación y uso de MySQL Server 5.5 y MySQL WorkBench, las dos herramientas de SGBD utilizadas en nuestro proyecto.

En cuanto a las herramientas de Pentaho utilizadas en el proyecto es importante destacar que el único prerequisite para poder trabajar con ellas es tener instalada la JRE 5.0 o superior. La misma se puede descargar de <http://www.javasoft.com/>

Instalación de Pentaho Data Integration (PDI)

Para su instalación es necesario:

- ✓ Instalar Java Runtime Environment (JRE) versión 1.5 o superior.
- ✓ Descargar Kettle, desde: <http://sourceforge.net/projects/pentaho/files/Data%20Integration/>.

En el proyecto se ha seleccionado la versión 5.0.1 y al ser el sistema operativo Windows se descargó el archivo con extensión “.zip”.



- ✓ Descomprimir el archivo descargado.

En el Anexo 2 se adjuntan documentos acerca de las distintas herramientas de PDI. Los mismos son:

- ❖ “Manual de Usuario de Spoon” desarrollado por el Ing. Darío Bernabeu.
- ❖ “Documentación del usuario de Pan”, “Documentación del usuario de Kitchen” y “Documentación del usuario de Carte” desarrollados por Tomás Di Doménico.
- ❖ “Tutorial: Primeros pasos con Pentaho Data Integration” de María Carina Roldán y las 3 partes del tutorial “Pentaho Data Integration (Kettle)” redactado por Jorge Rivera Ramos. Allí se contemplan diversos temas, desde la instalación y configuración de cada una a la realización de un primer trabajo utilizándolas.

Instalación de Pentaho Schema Workbench (PSW)

Para su instalación, lo primero que se debe hacer es:

- ✓ Descargar Mondrian Schema Workbench (Pentaho Schema Workbench -PSW-) desde: <http://sourceforge.net/projects/mondrian/files/schema%20workbench/>

La versión que ha sido utilizada en el proyecto es 3.5.0

- ✓ Al ser el sistema operativo Windows, se debe descargar el archivo con extensión “.zip”
- ✓ Una vez descargado el archivo, se procede a descomprimirlo en algún lugar apropiado.

En el anexo 3 se adjunta el documento desarrollado por el Ing. Dennis Alba Infante, en el “*PasosparacrearCubosconSchemaWorkbench.pdf*”, el cual hace una breve introducción a la herramienta.

Instalación de Pentaho BI Server

Para la instalación de este software, se debe descargar el archivo .zip o .tar.gz de la web y descomprimirlo (en el presente trabajo se descargó el archivo .zip por ser Windows el sistema operativo del servidor). Esto puede realizarse desde la siguiente URL:

<http://sourceforge.net/projects/pentaho/files/Business%20Intelligence%20Server/>

En la misma dirección web, es posible encontrar el manual de usuario que explica temas acerca de la instalación y usos de las consolas.

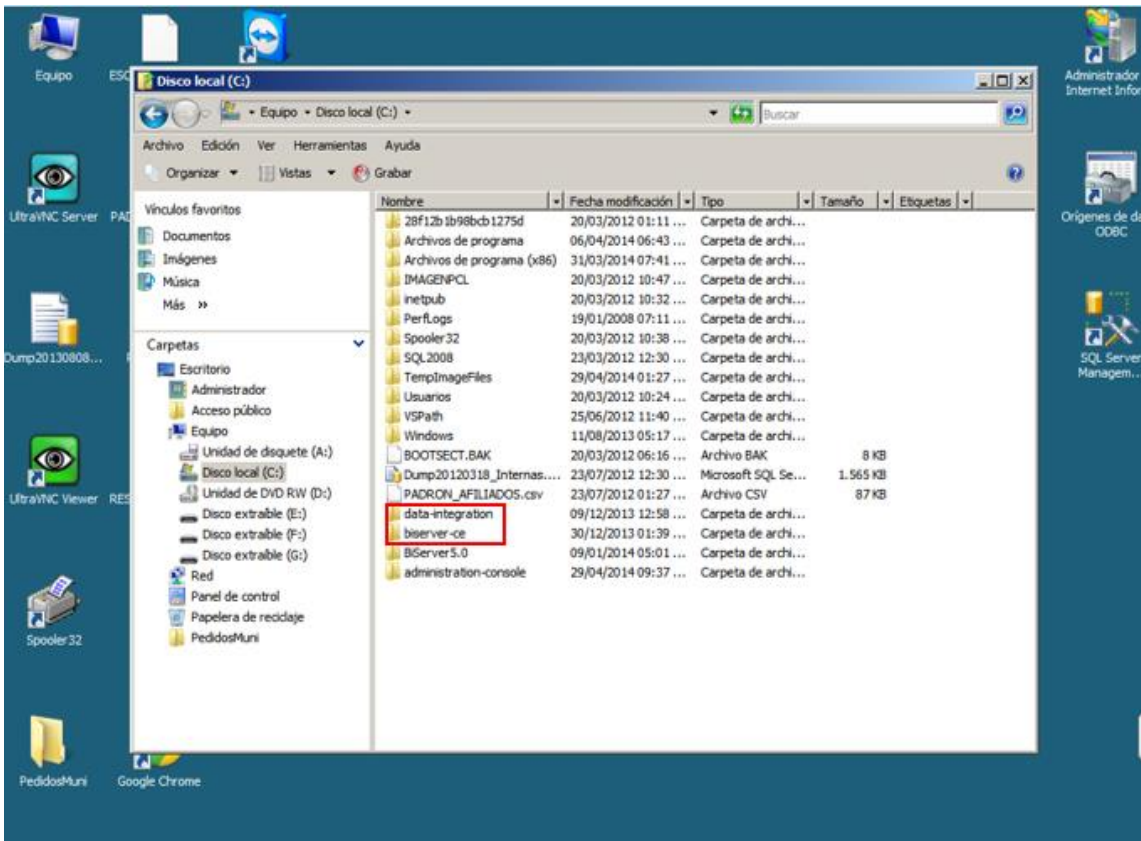
En el proyecto se instalaron las aplicaciones de BI Server y PDI en el servidor, mientras que PSW fue instalada en máquinas locales y desde allí se publicó el Schema de cubos en el Servidor.

A continuación se presenta en imágenes las carpetas creadas en el servidor.

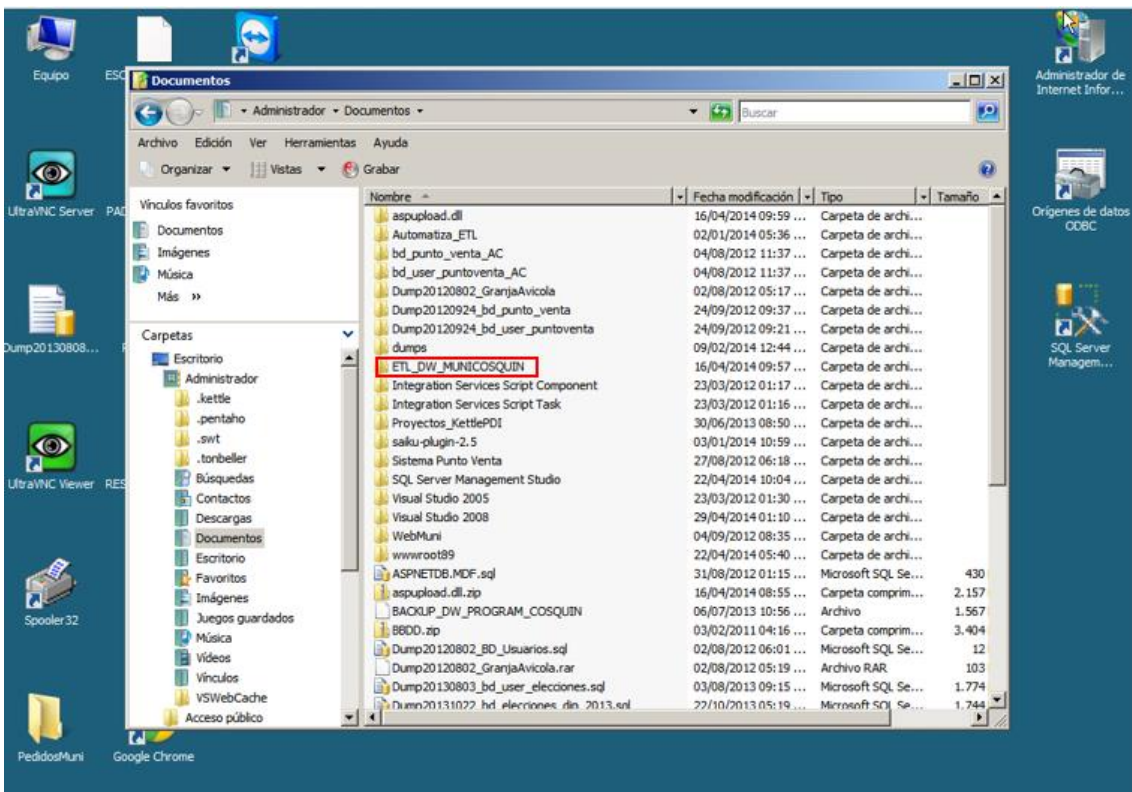
En la carpeta raíz del server se descomprimen los archivos “.zip” que hemos descargado creando la carpeta biserver donde están los archivos y componentes del BI Server y la carpeta data integration que contiene los archivos y componentes de PDI.



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

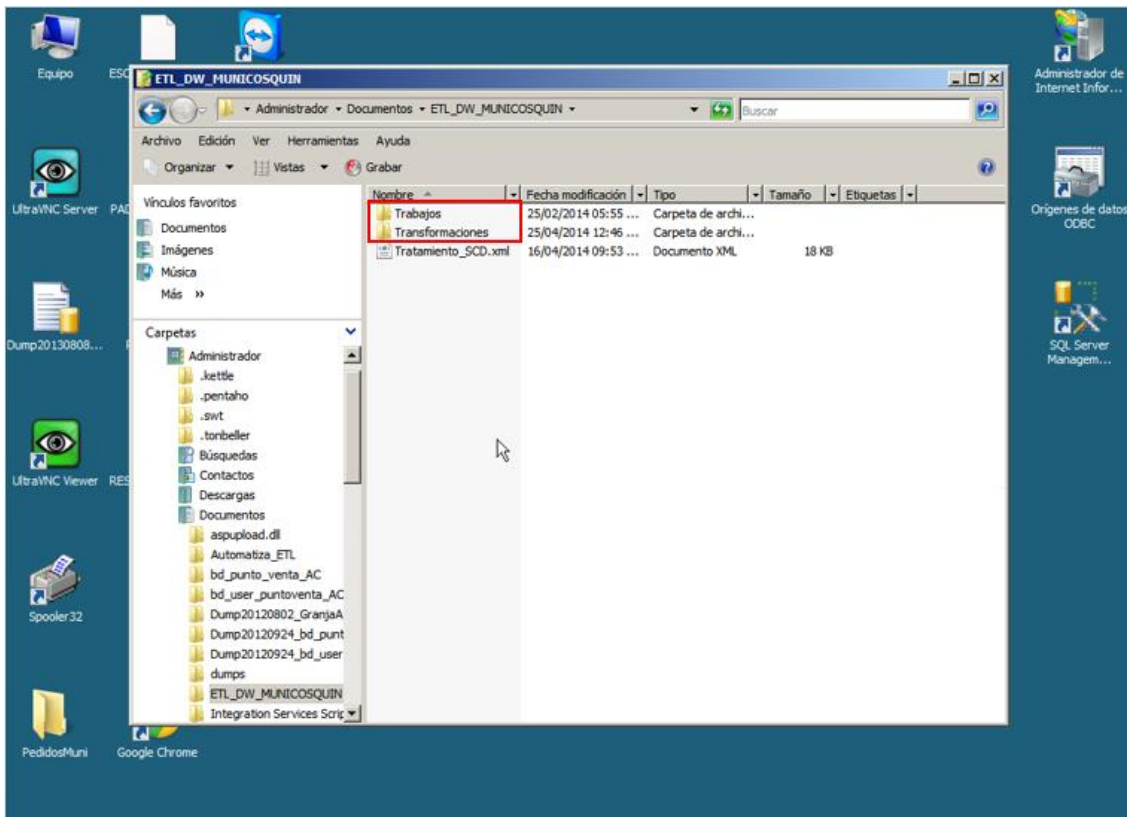


Además es creada otra carpeta denominada “ETL_DW_MUNICOSQUIN” en la carpeta de Documentos, en la cual se crearon dos carpetas para diferenciar los trabajos de las transformaciones.



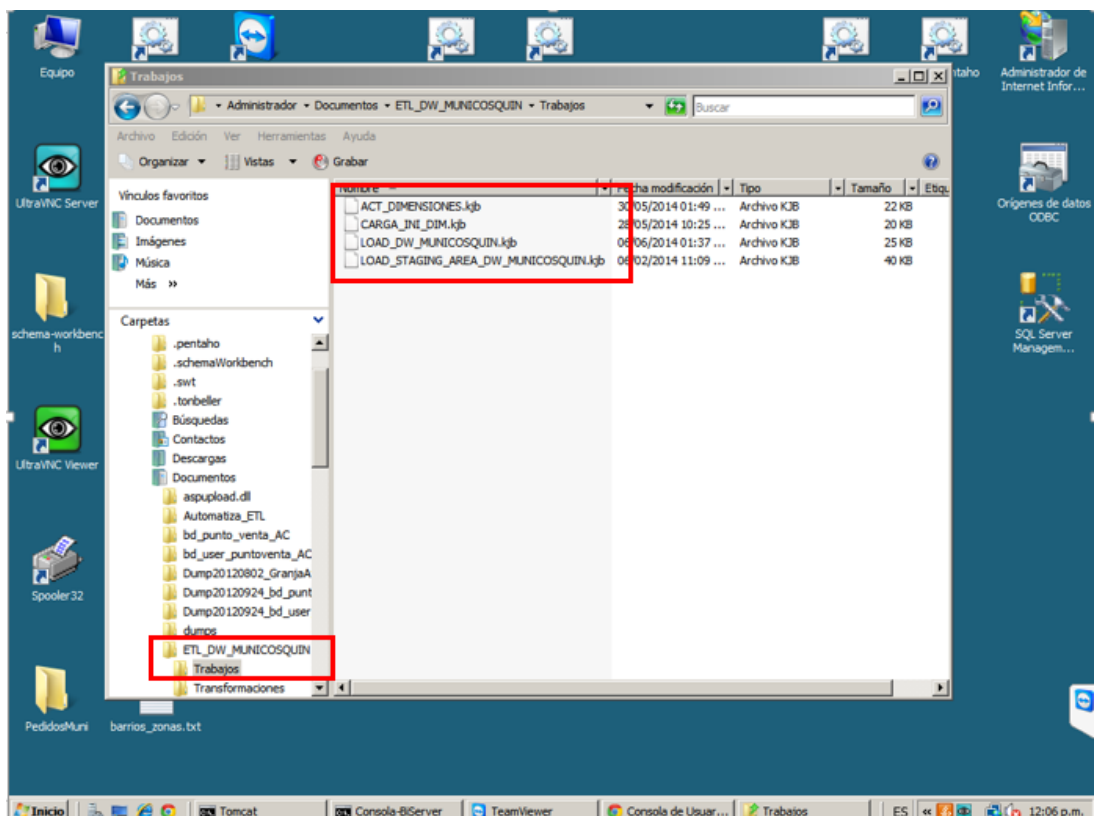


Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



Dentro de cada carpeta se ubican los trabajos y transformaciones creados con PDI.

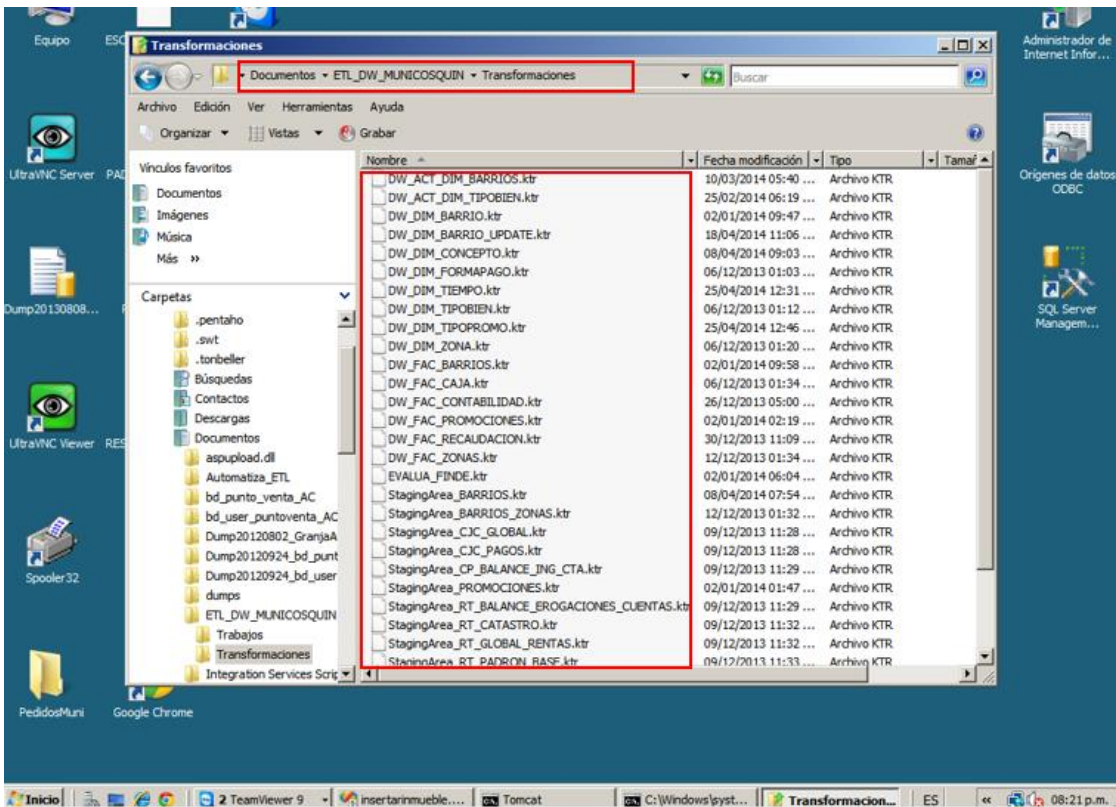
Trabajos:



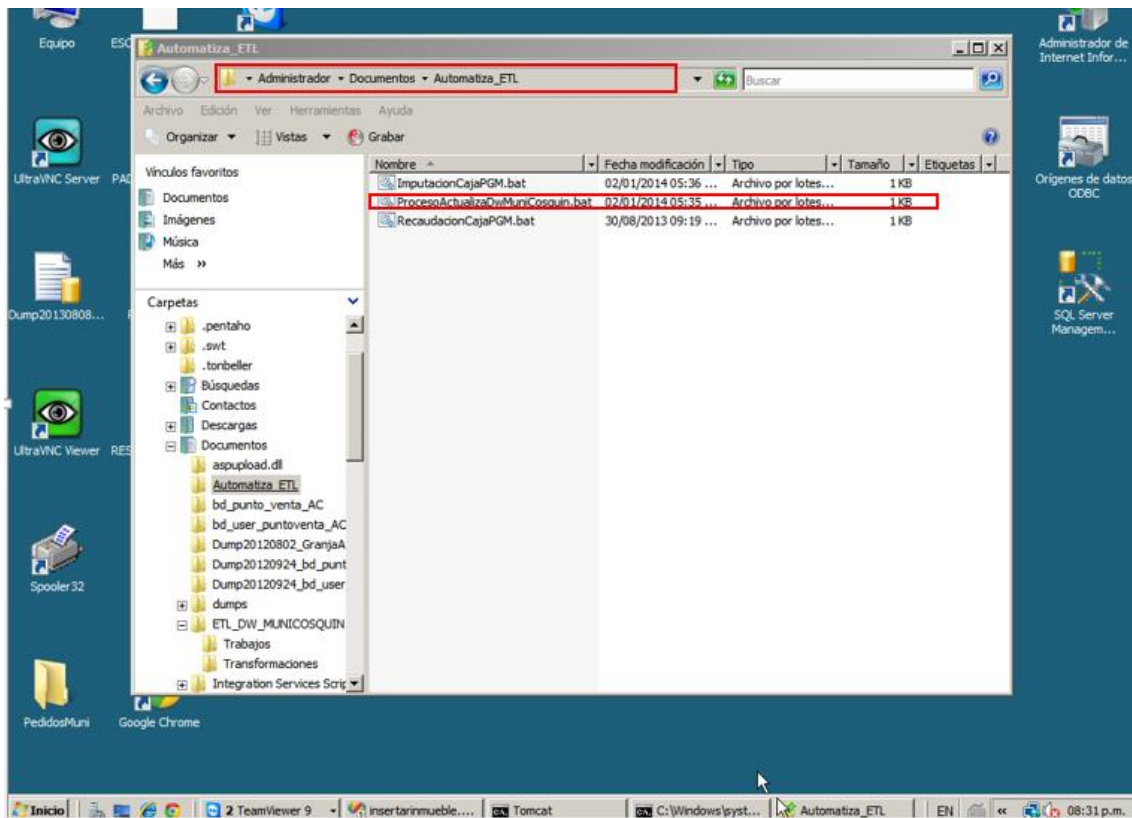
Transformaciones:



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

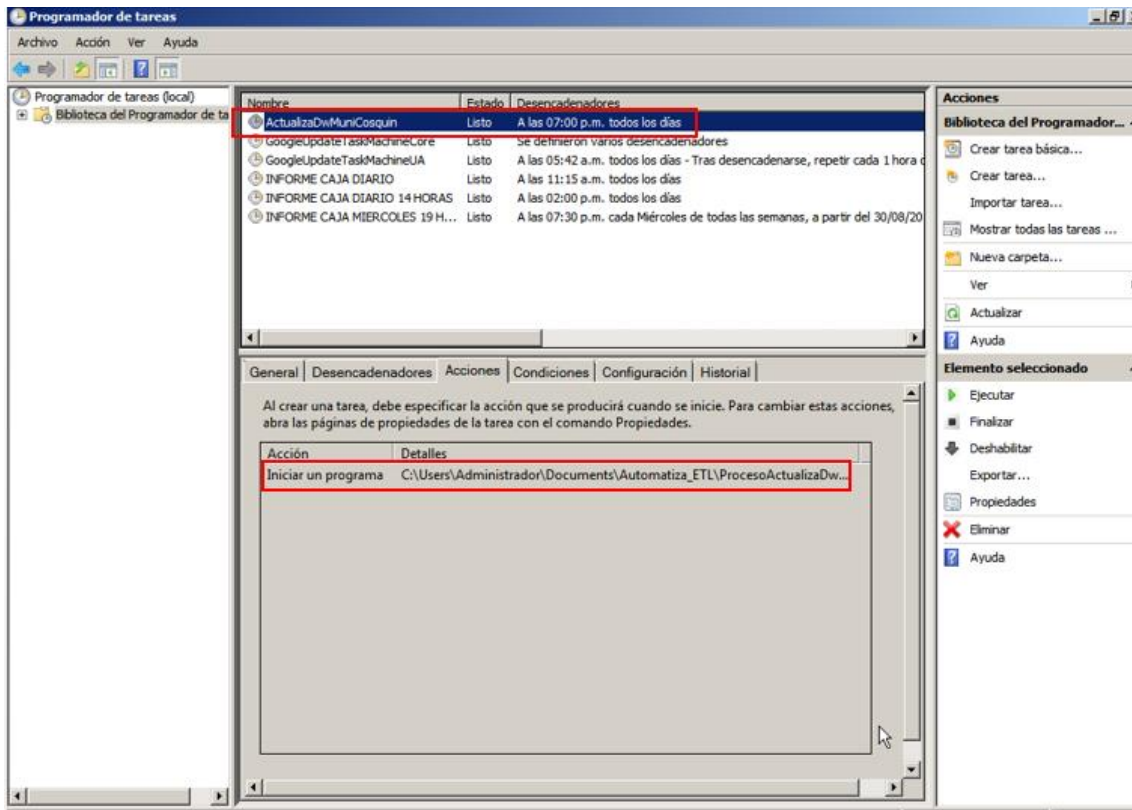


En la carpeta Documentos también se creó una carpeta “Automatiza_ETL” en donde se ubica el archivo .bat “ProcesoActualizaDwMuniCosquin.bat” que se ejecutará automáticamente todos los días a las 19hs para actualizar los datos del DW:

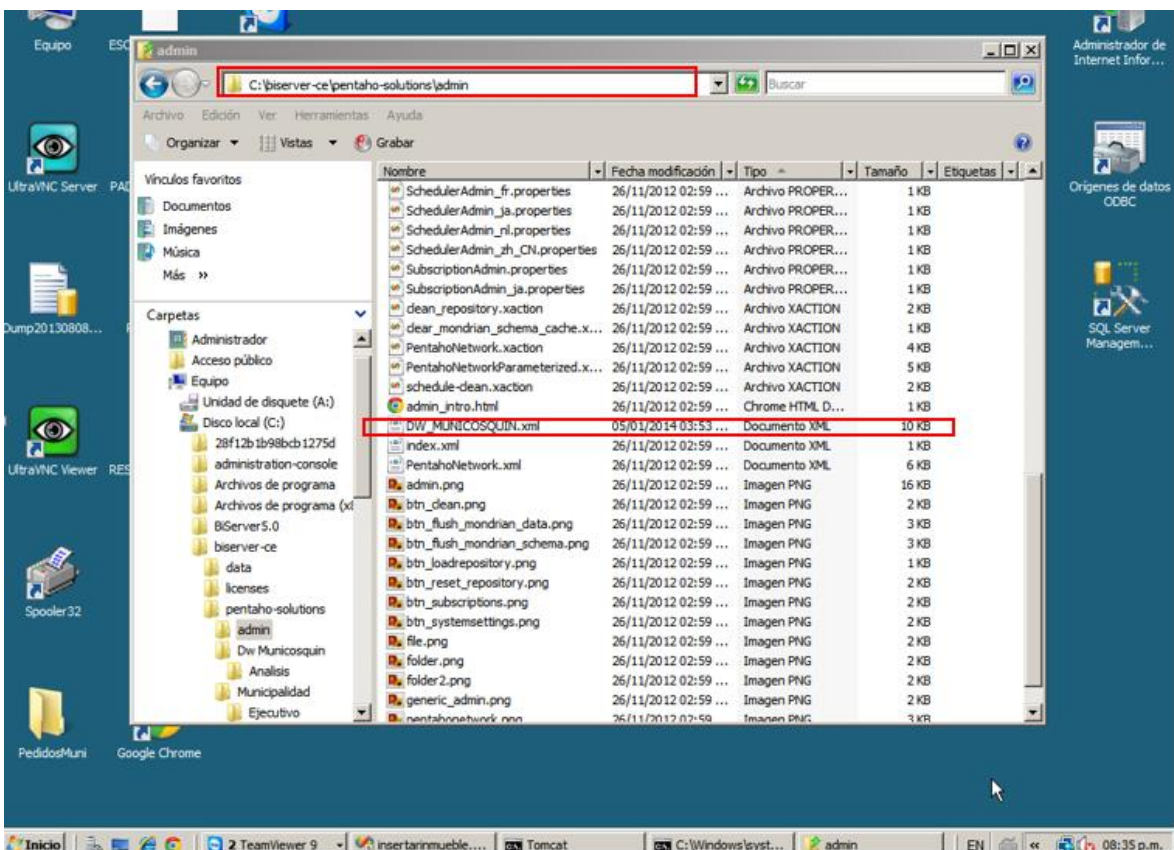




Proyecto Data Warehouse - Municipalidad de Cosquín ALISON - GRIGIONI



Luego utilizando PSW instalado en una máquina local se configura la publicación del Schema “DW_MUNICOSQUIN.xml”, que contiene los cubos desarrollados y que serán leídos por el BI Server, en la ubicación:





De esta forma se puede observar cómo quedan todos los archivos ubicados en las distintas direcciones dentro del servidor de la municipalidad.



CONCLUSIONES



Se ha arribado a la etapa en que es necesario realizar la conclusión del trabajo final de grado. En este momento es menester observar en retrospectiva el camino recorrido durante varios meses, para finalmente poder hacer un balance del proyecto realizado.

Haciendo una rápida lectura es posible evaluar al proyecto muy satisfactorio, no sólo por los resultados obtenidos, principalmente alcanzado los objetivos planteados, sino también por la experiencia que genera en los autores el afrontar un trabajo de estas dimensiones.

Para darle un cierre al proyecto, se realizará un análisis de las etapas transitadas. En una primera instancia se produjo el acercamiento a la organización, obteniendo una apreciación general sobre su realidad actual y la situación problemática a la cual se enfrentaba. A partir de ello, fue posible entender tanto su actualidad como su contexto fijando el objetivo general. Una vez establecido el objetivo general fue viable fijar los objetivos específicos e indicadores de evaluación y planificar un conjunto de tareas, acciones y actividades necesarias para lograrlos.

El trabajo fue estructurado buscando darle una organización que facilite su lectura, así como ser entendido y comprendido por personas no idóneas en la materia, por ello para comenzar, se efectuó el desarrollo de la “Introducción” a fin de brindar la información que permita la comprensión del contenido presente en el proyecto tanto como de los resultados esperados. Seguidamente se encaró el “Marco Contextual” analizando los problemas principales con los que se encontraba la Secretaría de Economía y Finanzas Públicas de la Municipalidad de Cosquín y a partir de analizar sus causas llegar a los efectos o defectos que estos ocasionan. Durante el “Marco Teórico” se describieron teorías, tecnologías, tendencias y otros conceptos técnicos que son necesarios estudiar, analizar y dominar para plantear una solución a la problemática existente. Para lograr esto, primero se describieron conceptos propios del ámbito municipal, continuando por los del perímetro de la toma de decisiones y llegando por último al punto eje del proyecto las soluciones de Business Intelligence, principalmente el DataWarehouse.

Finalmente, durante el capítulo “Desarrollo Práctico” se expone la implementación realizada, desde la presentación de una metodología propia con base en la metodología HEFESTO (presentada y explicada en el marco teórico) hacia la creación de los diferentes reportes que permitan a los usuarios acceder a la información que desean y necesitan consultar.

Es realmente importante subrayar que lo implementado se encuentra funcional en la actualidad, con el compromiso de los autores del trabajo de dedicar esfuerzo personal en el desarrollo y crecimiento futuro del proyecto, continuando la formación en herramientas de consulta y reportes; así como otros aspectos del área de BI.

Para concluir se destacará que como resultados del proyecto, no solo se ha logrado implementar una Solución de Inteligencia de Negocios, la cual brinda grandes beneficios a la Secretaría de la Municipalidad de Cosquín, sino que el desarrollo del trabajo permitió a los autores del proyecto integrar distintos conceptos adquiridos en diferentes asignaturas de la carrera (Toma de Decisiones y Soporte de Decisión) así como investigar y aumentar el mismo, teniendo hoy en día mayores conocimientos acerca de la Inteligencia de Negocios, metodologías, software disponible y más. Estas herramientas aprendidas abren y posibilitan a los autores del trabajo de grado explotar un área de trabajo relativamente nuevo dentro de las Tecnologías Informáticas.



Se concluye el trabajo final de grado recordando el objetivo general así tanto como los indicadores planteados para evaluar los objetivos específicos para de este modo controlar que los resultados del trabajo sean exitosos:

- ***Amigabilidad de la herramienta:*** Es medido por el porcentaje del personal que está satisfecho con la interfaz del software implementado.
Calculo: $(\text{nro. de personal satisfecho} / \text{nro. total de personas que usan el sistema}) * 100$.
Total de las personas que usan el sistema: 4 personas.
Cantidad de personas satisfechas: 4 personas.
✓ **Resultado: 100% - Objetivo Cumplido.**
- ***Verificación de la utilización de un método desarrollado para la elección de las herramientas de software a implementar:*** Se comprueba mediante el uso de algún proceso racional de toma de decisiones.
✓ **Resultado: Durante el proyecto se utilizó el proceso racional de toma de decisiones de Kepner y Tregoe. Objetivo Cumplido.**
- ***Calidad de la información brindada por la solución:*** Se mide a través del porcentaje del personal que está satisfecho con la calidad de la información que se obtiene de la solución BI desarrollada. Por calidad de información se entiende que la misma debe ser precisa, oportuna y útil.
Calculo: $(\text{nro. de personal satisfecho} / \text{nro. total de personas que usan el sistema}) * 100$.
Total de las personas que usan el sistema: 4 personas.
Cantidad de personas satisfechas: 4 personas.
✓ **Resultado: 100% - Objetivo Cumplido.**
- ***Capacidad de contar con información histórica:*** Se verifica mediante el análisis de datos históricos a través de la solución.
✓ **Resultado: La solución permite dicho análisis. Objetivo Cumplido.**
- ***Capacidad de resguardo y respaldo de la información:*** Se verifica mediante el uso de alguna técnica y medidas de seguridad que permitan asegurar el resguardo y respaldo de la información.
✓ **Resultado: Se ha establecido como política de backups la implementación del método de backup diferencial que posibilita SqlServer. Objetivo Cumplido.**
- ***Capacidad de uso del sistema por parte de los usuarios:*** Se mide mediante el porcentaje del personal que está capacitado para el uso del software implementado.
Total de las personas que usan el sistema: 4 personas.
Cantidad de personas satisfechas: 4 personas.
✓ **Resultado: 100% - Objetivo Cumplido.**



- **Disminución de los tiempos y errores en la generación de informes:** Se mide comparando los tiempos y errores que se tienen actualmente contra los obtenidos a partir de la implementación del DataWarehouse.

Comparación de los tiempos:

ACTIVIDADES	ANTES	DESPUÉS
Informe Caja	2 horas	3 seg.
Informe Caja Primer Trimestre	No se realizaba.	3 seg.
Informe Análisis Barrios	No se realizaba.	3 seg.
Informe Análisis Zonas	No se realizaba.	3 seg.
Informe Indicadores	1 semana.	3 seg.
Informe Análisis Ingresos	4 horas.	3 seg.
Informe Pago Contado	No se realizaba.	3 seg.
Informe Análisis Promociones	4 horas.	3 seg.

En cuanto a los errores es posible afirmar que con la implementación de la herramienta se evita realizar el procedimiento de forma manual y de esta forma los errores inherentes.

✓ **Resultado: Objetivo Cumplido.**

A partir del cumplimiento de los objetivos específicos se logra la meta del proyecto que fue definida como objetivo general, y con ello resolver los problemas que tenía la Secretaría de Economía y Finanzas Públicas del Municipio de Cosquín.

“El objetivo principal del presente proyecto es realizar el análisis, diseño e implementación de una Solución de Inteligencia de Negocios, de manera de contar con información sólida y confiable, de alta disponibilidad, y que cumpla con las necesidades del área permitiendo mejorar la gestión y la toma de decisiones.”



BIBLIOGRAFÍA



1. Carme Martin Escofet (n.d). *El lenguaje SQL*.
2. Charles H. Kepner, Benjamin B. Tregoe (1992). *El Nuevo Directivo Racional, Análisis de problemas y toma de decisiones*. Ed. McGraw-Hill.
3. *Comparativa B.I Open Source* (2010). Obtenida el martes 29 de Octubre de 2013 de www.stratebi.com.
4. Daniel Scandibuzzo, Cesar E. Murua (2008, 26 de diciembre). *Las transferencias financieras a los gobiernos locales en la Provincia de Córdoba, Argentina: La necesidad de un modelo equitativo y transparente*. Revista OIKOS año 12, número 26, páginas 49-68.
5. Espinosa, Roberto. *El Rincón del BI (Blog Internet)*. Disponible en <http://churriwifi.wordpress.com/>
6. Ing. Bernabeu (2010), *Hefesto*.
7. Ing. Bernabeu, Darío. *Open Source Business Intelligence (Blog Internet)*. Disponible en <http://tgx-hefesto.blogspot.com.ar/>
8. Ing. Boggio, María Alejandra (2010). *Laboratorios, prácticos, apuntes y bibliografía de la materia TOMA DE DECISIONES*. Instituto Universitario Aeronáutico (IUA).
9. Ing. Boggio, María Alejandra (2010). *Laboratorios, prácticos, apuntes y bibliografía de la materia SOPORTE DE DECISIONES*. Instituto Universitario Aeronáutico (IUA).
10. Ing. Mariano García Mattio. *Programación, Base de datos, BI e IT en general (Blog Internet)*. Disponible en <http://jmagm.blogspot.com.ar>.
11. Ing. Mariano García Mattio (2010). *Laboratorios, prácticos, apuntes y bibliografía de la materia MOTORES DE BASE DE DATOS*. Instituto Universitario Aeronáutico (IUA).
12. Juan Rivera, J. D. *Utilización de Información Histórica para Decisiones Empresariales*.
13. Luis Alberto Casillas Santillán, Marc Gibert Ginesta, Oscar Perez Mora (n.d). *Bases de Datos en MySql*.
14. María Carina Roldan (2010). *Pentaho 3.2 Data Integration: Beginner's Guide*. Packt Publishing
15. MySQL. (s.f.). Obtenido de <http://dev.mysql.com/doc/refman/5.0/es/index.html>
16. *Open Business Intelligence. La red del business intelligence*. Disponible en <http://www.redopenbi.com/>
17. *Pentaho Community*. (s.f.). Obtenido de <http://community.pentaho.com/>
18. *Pentaho Mondrian Documentation*. (s.f.). Obtenido de <http://mondrian.pentaho.com/documentation/schema.php>
19. *Pentaho Wiki*. (s.f.). Obtenido de <http://wiki.pentaho.com/display/COM/Community+Wiki+Home>
20. *Procesos ETL – La base de la inteligencia de negocios*. Obtenida el sábado 2 de noviembre de 2013 de <http://www.powerdata.es/>.
21. *Procesos ETL en profundidad*. Obtenida el miércoles 29 de enero de 2014 de <http://www.powerdata.es/>
22. Rivera, L. R. (1997). *Nociones de Datawarehousing*. Lima.
23. Robert S. Kaplan, David P. Norton (1992). *Cuadro de Mando Integral (The balanced Scorecard)*. Ed. Gestión 2000.
24. *TodoBI. Business Intelligence*. Disponible en <http://todobi.blogspot.com.ar/>



ANEXOS



ANEXO 1: Instalación y configuración de MySQL Server 5.5 y MySQL Workbench

Tanto MySQL Server 5.5 como MySQL Workbench pueden ser descargados de la página oficial de MySQL: <http://dev.mysql.com/>

Descargamos MySQL Server 5.5 desde: <http://dev.mysql.com/downloads/mysql/>

Y MySQL Workbench desde: <http://dev.mysql.com/downloads/tools/workbench/>

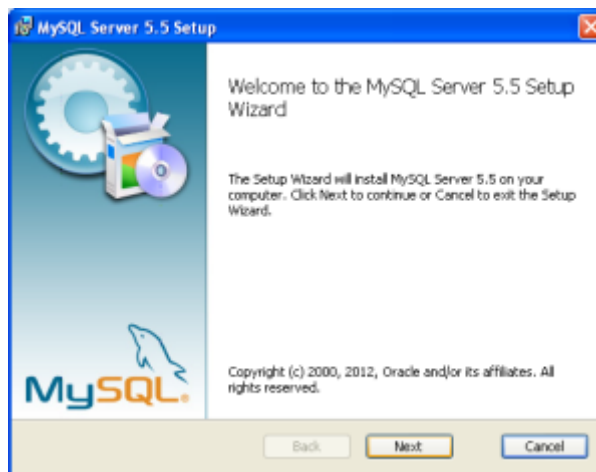
Una vez descargados los instaladores, para llevar a cabo la instalación y uso de MySQL Server 5.5, se dispone de un manual en línea “*MySQL 5.5 Reference Manual*”, el cual puede ser consultado en la siguiente página web:

<http://dev.mysql.com/doc/refman/5.5/en/index.html>

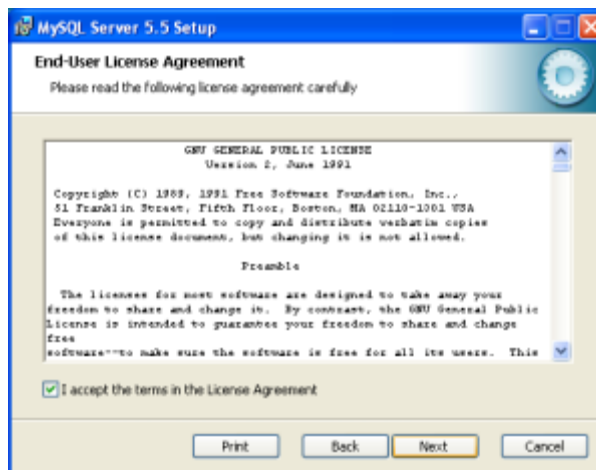
Aquí no solo se pueden encontrar todos los pasos e indicaciones para instalar y hacer uso de esta herramienta, sino que mucha más información de la herramienta como backup y recovery, procedimientos almacenados, vistas, seguridad, tutoriales, etc.

Los pasos para instalar MySQL Server 5.5 se pueden resumir en:

1. Iniciamos el ejecutable, hacemos click en “Next” para continuar.



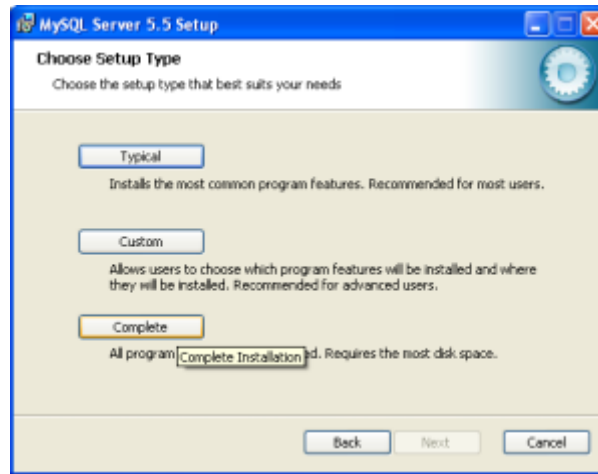
2. Pantalla de los términos de licencia. Click en “Next”



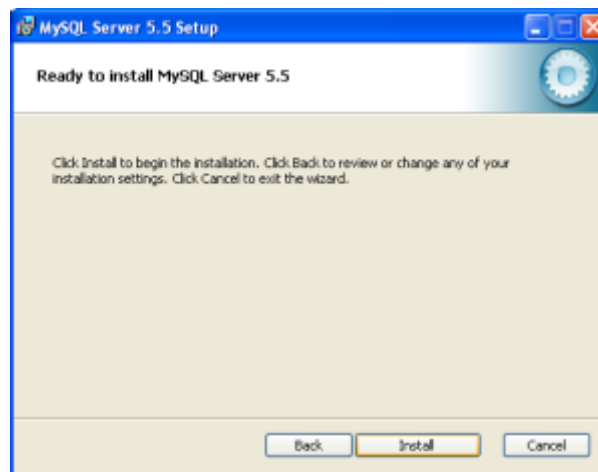


Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI

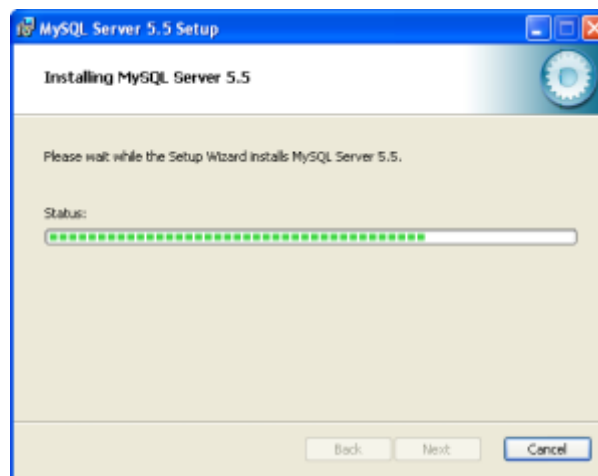
3. Nos pregunta por el tipo de instalación. Elegimos completa.



4. Ya está listo para empezar el proceso de instalación. Click en “Install”.



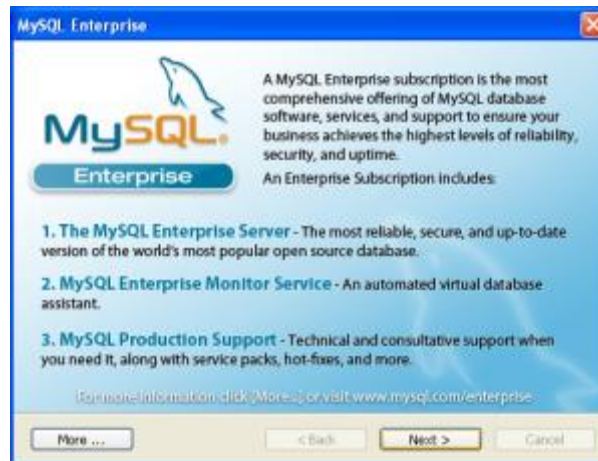
5. Instalación en proceso



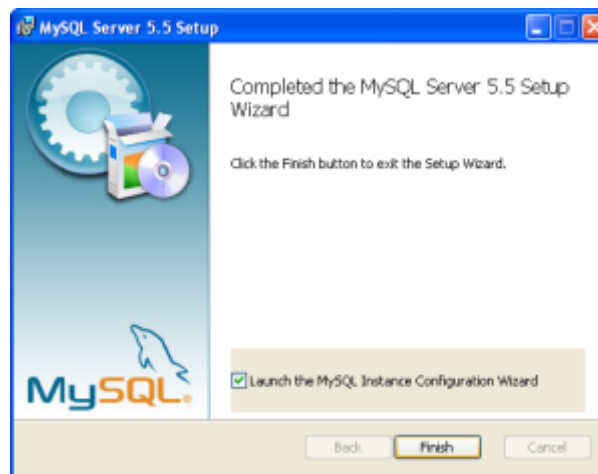
6. La instalación finalizó, algunas pantallas para leer. Click en “Next”.



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



7. Instalación finalizada. Proseguimos con la herramienta de configuración



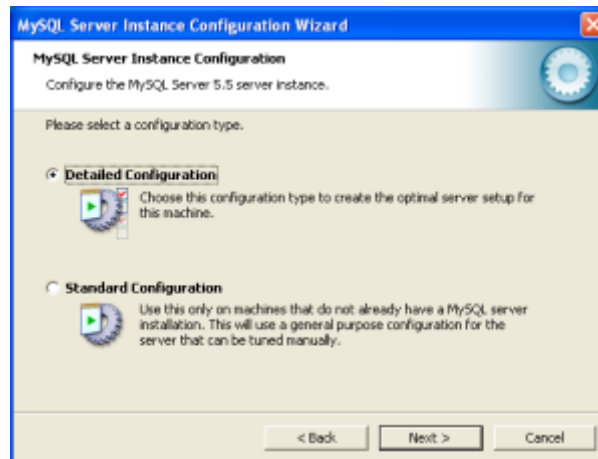
8. Se abre la ventana de la herramienta de configuración. Click en “Next”



9. Seleccionamos un tipo de configuración. Es este caso se elige la configuración detallada.



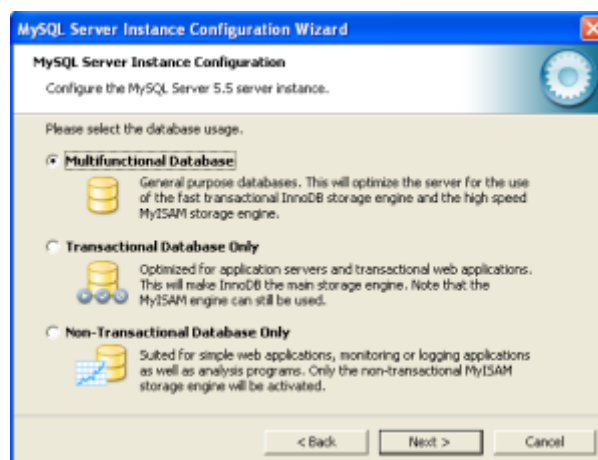
Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



10. Ya que la máquina donde se instala va a servir para desarrollo, elegimos “Developer Machine”



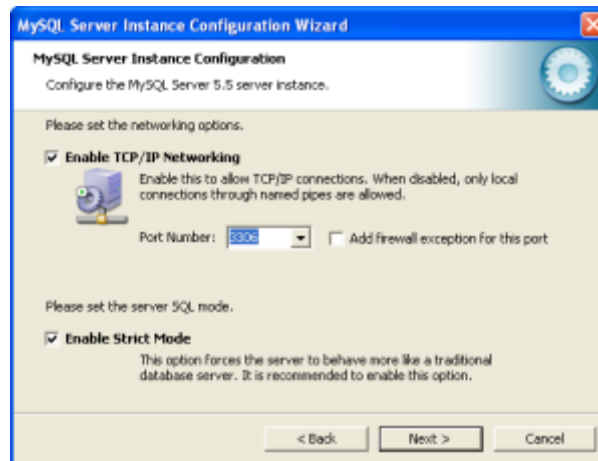
11. Configuramos el soporte de MyISAM y/o InnoDB como tecnología de almacenamiento de datos.



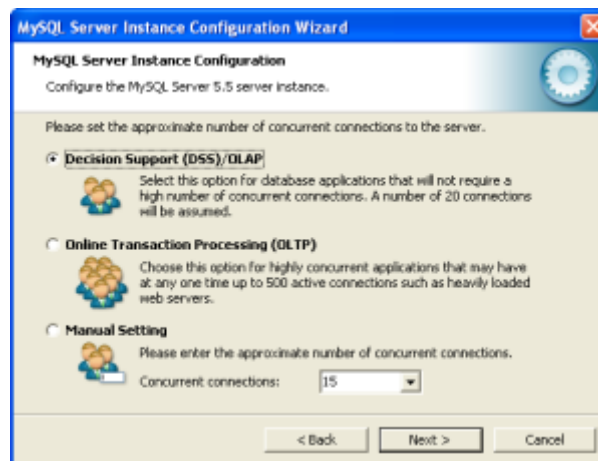
12. Habilitamos la conexión TCP/IP y designamos un puerto para la gestión remota de la base de datos.



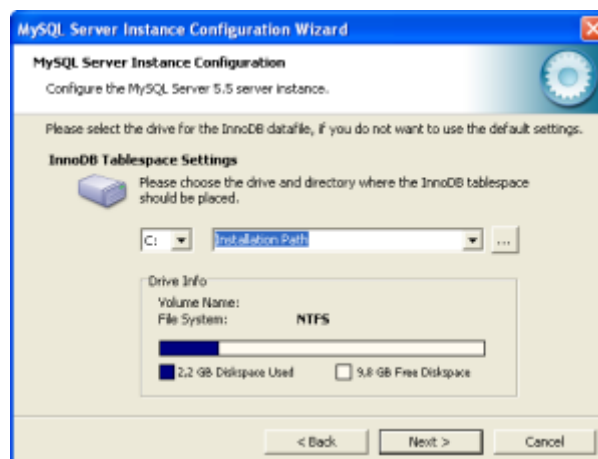
Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



13. Elegimos el número aproximado de conexiones concurrentes a la base de datos.



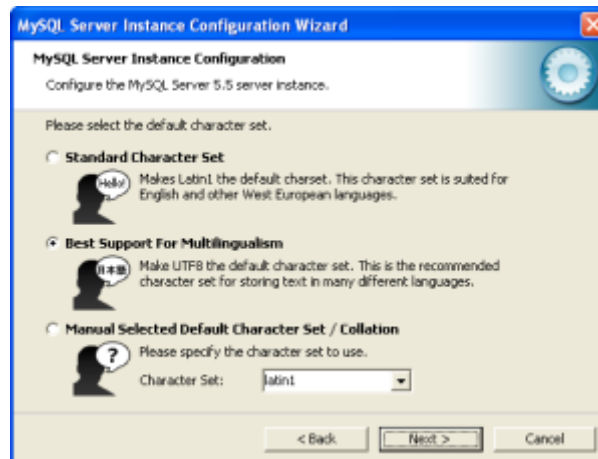
14. Elegimos el lugar de instalación



15. Se elige el conjunto de caracteres que soportara por defecto la base de datos.



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



16. Seleccionamos para instalar MySQL como servicio (recomendado).



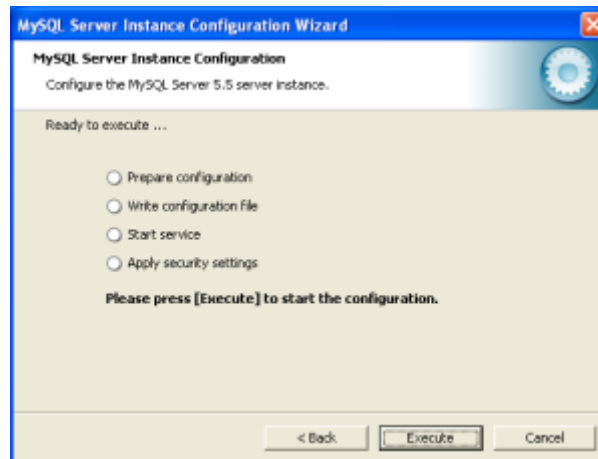
17. Colocamos una contraseña al root (administrador)



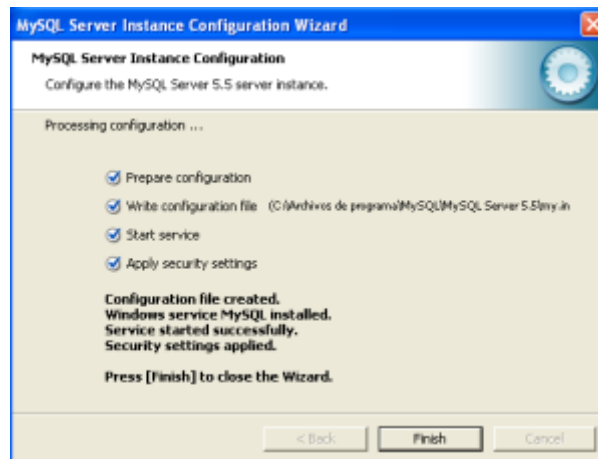
18. Pantalla antes de empezar la configuración. Haga click en “Ejecute” para ejecutar lo previamente elegido.



Proyecto Data Warehouse - Municipalidad de Cosquín
ALISON - GRIGIONI



19. La configuración ha finalizado. Haga click en “Finish”.



También podemos acceder a la página de MySQL Workbench la cual describe la herramienta y nos brinda los enlaces para el manual y el tutorial de instalación de la misma: <http://dev.mysql.com/doc/workbench/en/>

Los pasos para instalar la aplicación pueden ser consultados desde: <http://dev.mysql.com/doc/workbench/en/wb-starting.html>.

Y el manual de usuario en: <http://dev.mysql.com/doc/workbench/en/index.html>