

# CARRERA: INGENIERÍA EN SISTEMAS

INSTITUTO UNIVERSITARIO AERONÁUTICO

*Proyecto: Data Warehouse aplicado en Agricultura de Precisión*

**Año: 2016**



**INSTITUTO UNIVERSITARIO AERONÁUTICO**

Asesor: ING MELONI, BRENDA E.

Equipo de Proyecto

Alumno	DNI	E-mail
Pablo David MALDONADO	31558194	pmaldonado194@alumnos.iua.edu.ar

### **Declaración de derechos de autor**

El presente trabajo fue desarrollado como trabajo de grado para la carrera Ingeniería de Sistemas de la Facultad de Ciencias de la Administración del Instituto Universitario Aeronáutico de la ciudad de Córdoba, Argentina.

El autor, Pablo David Maldonado, deja constancia por la presente la autorización la disposición de este material para la comunidad.

## **Agradecimientos**

- A mi familia
- A Dios,
- A la comunidad del IUA, profesores, docentes y alumnos.
- A mis compañeros de trabajo y compañeros de carrera por su amplia colaboración.
- y a todas aquellas personas que directa o indirectamente prestaron colaboración para permitir alcanzar esta meta en mi vida.

INSTITUTO UNIVERSITARIO AERONAUTICO –  
FACULTAD DE CIENCIAS DE LA ADMINISTRACION

Aprobado por el Departamento de Sistemas en el cumplimiento de los requisitos exigidos para otorgar el título de Ingeniera de Sistemas.

Al Señor: Maldonado Pablo David

DNI: 31.558194

Revisado por:

.....

Tutor del trabajo.

.....

Director Depto. Sistemas.

Tribunal Examinador.

.....

Presidente del Tribunal Examinador.

.....

Vocal del Tribunal Examinador.

## **Resumen**

A lo largo de este trabajo de grado se ha llevado a cabo la realización de una herramienta ad-hoc que permite dar soporte a la toma de decisiones en el proceso de agricultura de precisión, mediante la creación de un data warehouse.

El mismo tiene la función de procesar datos telemétricos capturados por sensores de clima instalados en centrales meteorológicas y crear reportes basados en datos históricos con un margen de diez años atrás.

## INDICE

<b>Declaración de derechos de autor</b> .....	2
<b>Agradecimientos</b> .....	3
<b>Resumen</b> .....	5
INDICE.....	6
I. INTRODUCCIÓN.....	7
II. SITUACIÓN PROBLEMÁTICA.....	8
III. OBJETO DE ESTUDIO.....	9
IV. OBJETIVOS .....	9
• Objetivo General.....	9
• Objetivos específicos.....	9
V. IDEA A DEFENDER / PROPUESTA A JUSTIFICAR / SOLUCIÓN A COMPROBAR.....	10
VI. ALCANCE DEL TRABAJO .....	10
VII. APORTE TEÓRICO .....	10
VIII. APORTE PRÁCTICO .....	11
IX. FACTIBILIDAD .....	12
X. MÉTODOS Y MEDIOS DE INVESTIGACIÓN .....	14
XI. PRIMERA PARTE: MARCO CONTEXTUAL .....	15
XII. SEGUNDA PARTE: MARCO TEÓRICO .....	20
XIII. TERCERA PARTE: MODELO TEÓRICO .....	48
XIV. CUARTA PARTE: CONCRECIÓN DEL MODELO .....	74
XV. CONCLUSIONES.....	80
XVI. BIBLIOGRAFÍA: .....	82
XVII: GLOSARIO .....	83

## **I. INTRODUCCIÓN**

Durante mucho tiempo los agricultores se han sustentado en herramientas muy rudimentarias para la elaboración de un plan de siembra y cosecha, la evaluación de los cultivos, determinación de zonas de siembra, aprovechamiento eficiente de los recursos disponibles con el fin de producir un aumento de la productividad, reducción de insumos empleados, simplicidad en las tareas y obtención de la mayor rentabilidad posible.

En la actualidad, con la necesidad de proyección a largo plazo, la creciente competitividad, el mercado cambiante, la aparición de nuevas tecnologías y nuevos métodos de agricultura, esta actividad requiere de un sustento concreto y fiable para llevar a cabo con mayor precisión y eficiencia el alcance de estos objetivos, eso por ello que surge la agricultura de precisión, denominada como el conjunto de técnicas que permiten la gestión localizada de cultivos que permiten básicamente realizar una toma de decisiones empleando las herramientas que pone a disposición el avance de la tecnología.

Esta metodología se basa en el manejo específico de un área de cultivo, para ello se utilizan herramientas como el posicionamiento global, dispositivos de distribución de riego, fertilizantes y plaguicidas variados, sensores climatológicos y de cultivo. La información es plasmada en mapas digitales sobre los cuales se toman decisiones de administración de recursos.

Esta propuesta pretende brindar una solución al proceso de toma de decisiones que se debe llevar a cabo para responder a interrogantes que surgen del proceso de agricultura de precisión, se realizará foco en los datos capturados por las centrales meteorológicas, y principalmente se implementará un data warehouse que emita reportes personalizados y provea respuesta a las interrogantes antes mencionadas.

Sobre el análisis económico que brinda la implementación del datawarehouse se establecerán métricas de rentabilidad inmediata, razón de beneficio/costo, tipo de licenciamiento y costo de soporte.

Lo que se plantea en esta propuesta es una solución íntegra que es de gran importancia para los ingenieros agrónomos cuya labor principal depende, entre otras cosas, de los datos que obtienen de las centrales meteorológicas.

## **II. SITUACIÓN PROBLEMÁTICA**

La situación problemática radica en la necesidad de investigar, analizar e implementar un sistema de administración de información capaz de manipular la cantidad de datos que se genera, proveniente de los diversos sensores telemétricos pertenecientes a las centrales meteorológicas ubicadas en los diferentes sectores que forman parte de los campos agrícolas, donde se encuentran plantaciones de diversos cultivos, entre ellos soja, trigo, maíz, etc. Actualmente el mercado ofrece sistemas que son capaces de medir las magnitudes meteorológicas de manera que el proceso sólo se limita a la lectura en un momento determinado de uno o varios parámetros específicos (temperatura, humedad, velocidad del viento, etc.) sin llevar a cabo un análisis de los mismos, elemento principal para poder realizar una adecuada toma de decisiones para contribuir a las interrogantes que emergen del proceso de agricultura de precisión.

Este proyecto pretende reducir la brecha que existe entre el componente de adquisición de datos y un sistema que sea capaz de procesarlos para proveer de repuestas concretas a al personal directivo que lleva a cabo la administración de los cultivos, el crecimiento de los

mismos, el estado de la tierra y el modo en que se puede actuar a futuro para el mejor aprovechamiento de los recursos.

### **III. OBJETO DE ESTUDIO**

El objeto de estudio se centra en el desarrollo de una herramienta que permita realizar un proceso ETL y de esta manera aprovechar los datos obtenidos ya sea de una central meteorológica o de registros históricos, para presentarlos como información útil ante el personal que se encarga de realizar la toma de decisiones en el proceso de agricultura de precisión.

### **IV. OBJETIVOS**

- Objetivo General

El objetivo principal de esta tesis es construir una herramienta ad-hoc que sirva de complemento en la toma de decisiones efectuado en el proceso de agricultura de precisión, y aplicándola a un caso real.

- Objetivos específicos

- Realizar un relevamiento de las herramientas disponibles para la elaboración del datawarehouse.
- Seleccionar la herramienta que más se adapte a las necesidades del proyecto
- Modelar el datawarehouse
- Construir el DW con las herramientas elegidas.
- Realizar la implementación de la plataforma de Software del DW

- Realizar el análisis costo / beneficio de esta herramienta.
- Generar conclusiones

## **V. IDEA A DEFENDER / PROPUESTA A JUSTIFICAR / SOLUCIÓN A COMPROBAR**

Se pretende construir una herramienta AD-HOC que brinde datos confiables y permita dar soporte a la toma de decisiones, demostrar la importancia del uso del datawarehouse en el proceso de agricultura de precisión, comprobar su mayor factibilidad frente a productos similares del sector.

## **VI. ALCANCE DEL TRABAJO**

El proyecto se limita a la elaboración de una herramienta AD-HOC que facilite la tarea de toma de decisiones mediante la aplicación de un datawarehouse para el análisis de los datos meteorológicos obtenidos de las centrales o registros históricos.

## **VII. APOORTE TEÓRICO**

En la actualidad existen dispositivos que permiten realizar la captura de datos meteorológicos, permitiendo su visualización, con la limitación de no enfocarse en su análisis, etapa fundamental del procesos de captura y análisis de datos para realizar la toma de decisiones obteniendo un sistema incompleto, y en el caso que cumplan ambas funciones, el costo es elevado.

La novedad de la propuesta presentada consiste en la implementación de una herramienta que cumpla la función de analizar los datos que capturan las centrales meteorológicas, u otro tipo de fuente, para dar soporte al proceso de toma de decisiones a un bajo costo y accesible a los usuarios del sector.

Los aportes teóricos desde el punto de vista de la ingeniería son:

- Implementación de un data warehouse aplicado al proceso de agricultura de precisión con software opensource.
- Establecer una guía para la implementación de esta herramienta.
- Creación de una herramienta AD-HOC para la toma de decisiones para el proceso de agricultura de precisión.

## **VIII. APORTE PRÁCTICO**

Se espera que la implementación de esta herramienta brinde una solución asequible a las empresas del sector agropecuario o las que trabajen con datos provenientes de sensores meteorológicos, impulse la rentabilidad de las empresas del sector.

Los Beneficios que produciría la implementación de este proyecto son:

- Reducir la incertidumbre al momento de tomar decisiones.
- Incrementar la productividad y rentabilidad de la empresa.
- Mejor administración y aprovechamiento de los recursos.
- Elevar la competitividad de la empresa.

Los aportes prácticos que se esperan son:

- Poner a disposición empresas del sector agropecuario una herramienta que colabore con el proceso de toma de decisiones.
- Evaluar inversión y costos correspondientes de la implementación de data warehouse
- Proveer una herramienta que facilite el proceso de agricultura de precisión.
- Elevar el grado de competitividad de las empresas que hacen uso de esta tecnología con respecto a otras del mismo sector.
- Llevar a cabo la implementación de un data warehouse aplicado a una situación específica en un caso real.
- Realizar un uso específico de la información crítica y vital procesada para coadyuvar al proceso de toma de decisiones.

## **IX. FACTIBILIDAD**

Se presenta a continuación la factibilidad del proyecto, haciendo una evaluación técnica, operativa y económica:

- Factibilidad Técnica

Para el desarrollo del data warehouse se tendrá en cuenta la metodología HEFESTO, y para su implementación se realizará una investigación de herramientas opensource disponibles y se elegirá aquella que se adecúe mejor al proyecto de acuerdo a sus prestaciones.

Por último es necesario la implementación de un servidor para alojar el data warehouse y todos sus componentes.

El hardware que se utilizará será provisto por la empresa

El software que se empleará como base del sistema será:

- Máquina virtual alojada en un web hosting.
  - Sistema de base de datos: MySql 6.00
  - Kit de Desarrollo Java: Java JDK 6.
  - Herramienta de elaboración de reportes: Jaspersoft Studio.
  - Herramienta para elaboración de proceso ETL: Jaspersoft ETL.
- 
- Factibilidad Operativa

Dado la facilidad de adquisición de los elementos necesarios y la cordial relación que tengo con la empresa para la implementación del proyecto, se afirma que el mismo es operativamente viable.

- Factibilidad Económica

El proyecto es totalmente realizable, tomando en cuenta que los recursos económicos necesarios para llevar a cabo el mismo son relativamente bajos ya que tengo a mi alcance los materiales para llevarlo a cabo o los medios para adquirirlos.

La empresa en donde se realizara el caso de estudio tendrá que realizar una inversión de muy bajo costo comparado con las capacidades que el proyecto ofrece y las funciones que proporcionan dispositivos similares del sector.

En base a esto podemos concluir que el proyecto es totalmente factible.

## **X. MÉTODOS Y MEDIOS DE INVESTIGACIÓN**

El método de investigación utilizado en el proyecto es el método empírico. Los métodos empíricos de investigación permiten efectuar el análisis preliminar de la información, así como verificar y comprobar las concepciones teóricas.

El fundamento radica en la percepción directa del objeto de investigación y del problema. Su aporte al proceso de investigación es resultado fundamentalmente de la experiencia. Estos métodos posibilitan revelar las relaciones esenciales y las características fundamentales del objeto de estudio a través de procedimientos prácticos con el objeto y diversos medios de estudio.

## **XI. PRIMERA PARTE: MARCO CONTEXTUAL**

### a) Agricultura de Precisión

La agroindustria se ha mecanizando y automatizando para servir mejor a la disminución de las expectativas de precios y el aumento de la productividad. Este es un efecto causado por la globalización de la economía. Actualmente, la agricultura se practica en grandes extensiones de monocultivo, que tienen impactos en la distribución inadecuada de los insumos, que interfiere con la productividad de los cultivos y la conservación de los recursos

naturales<sup>1</sup>. Como aliados, los agricultores tienen las nuevas tecnologías disponibles, tales como satélites, computadoras y sensores que pueden ayudar a identificar mejor cada parte del suelo, por lo que la siembra ocurra de una manera particular en cada área distinta. Estos acontecimientos son la base para el surgimiento de un nuevo sistema de producción llamada Agricultura de precisión (Agricultura de precisión o la agricultura de precisión). La agricultura de precisión (AP) es un paradigma de la agricultura que permite a los agricultores realizar verificaciones de modo espacial y temporal de los factores limitantes de la producción en sus fincas<sup>2</sup>.

---

<sup>1</sup>[Barbosa et al. 2006]

<sup>2</sup>[Mondal et al. 2011]

Dada la gran cantidad de datos obtenidos por la agricultura mecanizada, es imperativo el uso de sistemas informáticos para apoyar el proceso de toma de decisiones, descartando los procesos manuales de adquisición, almacenamiento y análisis utilizados en la agricultura tradicional <sup>3</sup>. La estrategia más adecuada para el almacenamiento de datos para apoyar la toma de decisiones es a través de los almacenes de datos. Data Warehouse es el nombre dado al concepto utilizado para describir conjuntos de datos, sujeto orientado, integrado, no volátil y variable en el tiempo<sup>4</sup>.

En segundo lugar<sup>5</sup> el DW puede contener datos consolidados de múltiples fuentes, ampliadas con información de resumen que cubre un largo período de tiempo.

El concepto sobre el que se basa la agricultura de precisión es aplicar la cantidad correcta de insumos, en el momento adecuado y en el lugar exacto. Es el uso de la tecnología de la información para adecuar el manejo de suelos y cultivos a la variabilidad presente dentro de un lote. La agricultura de precisión (AP) involucra el uso de sistemas de adquisición de datos telemétricos (humedad el suelo, humedad ambiente, temperatura, etc.) y de otros medios electrónicos para obtener datos del cultivo como el posicionamiento global (GPS).

---

<sup>3</sup>[Saraiva 1998]

<sup>4</sup> [Kimball et al. 2002]

<sup>5</sup> [Inmon 2005]

La AP permiten satisfacer una de las exigencias de la agricultura moderna: el manejo óptimo de grandes extensiones. Se presenta como principal ventaja que el análisis de resultados de los ensayos se puede realizar por sectores diferentes dentro de un mismo lote, y de esta manera ajustar el manejo diferencial dentro de los mismos.

Por ejemplo, los rendimientos de dos cultivos pueden ser idénticos si se usan los promedios, pero diametralmente opuestos en una situación de loma y en una de bajo en un determinado lote. Este dato sólo podrá obtenerse mediante la realización de un mapa de rendimiento. Así mismo, con el uso de sensores meteorológicos capturando datos de temperatura, humedad relativa, humedad del suelo, etc. y acoplado a sistemas de control, se convierte en una herramienta importante que permiten realizar una proyección a largo plazo sobre el manejo de recursos, optimizando su uso a lo largo de toda la cosecha. Del mismo modo, podrán analizarse, el tipo y la dosis de fertilizante a aplicar, la densidad de semilla, la fecha de siembra, el espaciamiento entre hileras, etc. El uso de las tecnologías de la agricultura de precisión, como los sistemas de control, contribuye a mejorar los márgenes, a través de un aumento del valor del rendimiento (cantidad o calidad), de una reducción en la cantidad de insumos, o de ambos simultáneamente.

### Data warehouse

A los fines de este proyecto, con el objeto de generar información objetiva, oportuna, de calidad y confiable, a fin de ponerla a disposición de la alta gerencia y que la información obtenida sea útil para el proceso de toma de decisiones, el uso de un data warehouse es propicio para el fin último de los datos.

A través del Data warehouse es posible:

- Visualizar información en diferentes periodos de tiempo.

- Hacer filtros dinámicos.
- Manejar escenarios (filtros por dimensión, etc.) y moverse entre ellos, así como imprimir los o exportarlos a formato JPEG, PNG, XLS, HTML, Acrobat (.pdf).
- Crear consultas personalizadas, mediante la agregación de columnas calculadas, cortes por dimensión, etc.
- Definir operaciones en una columna de indicador.
- Crear plantillas de análisis (tendencias, máximos y mínimos, tendencia incremental, etc.).
- Realizar proyecciones.

Data warehousing es el centro de la arquitectura para los sistemas de información en la década de los '90. Soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis. Facilita la integración de sistemas de aplicación no integrados. Organiza y almacena los datos que se necesitan para el procesamiento analítico, informático sobre una amplia perspectiva de tiempo. Un Data warehouse o Depósito de Datos es una colección de datos orientado a temas, integrado, no volátil, de tiempo variante, que se usa para el soporte del proceso de toma de decisiones gerenciales.

Se puede caracterizar un data warehouse haciendo un contraste de cómo los datos de un negocio almacenados en un data warehouse, difieren de los datos operacionales usados por las aplicaciones de producción.

Data warehouse para la Agricultura de Precisión:

La primera fase consiste en la selección de la fuente de los datos, seguida por el análisis y adecuación de los metadatos al dominio de aplicación específico del proyecto. Finalizadas las fases iniciales, se da comienzo al proyecto del data warehouse propiamente dicho, que consiste en la selección de los procesos de negocio, definición de la granularidad de los datos, el modelado de las dimensiones y la identificación de la tabla de hechos, concluido esto, inicia la fase de implementación y la carga de los datos en el data warehouse desde los sensores ubicados en la centrales meteorológicas u otras fuentes de datos, como los registros históricos.

## **XII. SEGUNDA PARTE: MARCO TEÓRICO**

### **METODOLOGÍA HEFESTO**

Para el desarrollo del data warehouse se seguirá la metodología HEFESTO<sup>6</sup>, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes, experiencias propias en procesos de confección de almacenes de datos. Cabe destacar que HEFESTO está 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 idea principal, es comprender cada paso que se realizará, para no caer en el tedio de tener que seguir un método al pie de la letra sin saber exactamente qué se está haciendo, ni por qué.

La construcción e implementación de un DW puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software. Lo que se busca, es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del DW y motivar a los usuarios.

---

<sup>6</sup> [Hefesto, Data Warehousing]

## DESCRIPCIÓN

La metodología HEFESTO puede resumirse a través del siguiente gráfico:



GRÁFICO 1: METODOLOGÍA HEFESTO

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, señalar las correspondencias con los datos fuentes y para 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.

## **EMPRESA ANALIZADA**

Antes de comenzar con el primer paso, es menester describir las características principales de la empresa a la cual se le aplicará la metodología HEFESTO, así se podrá tener como base un ámbito predefinido y se comprenderá mejor cada decisión que se tome con respecto a la implementación y diseño del DW.

Además, este análisis ayudará a conocer el funcionamiento y accionar de la empresa, lo que permitirá examinar e interpretar de forma óptima las necesidades de información de la misma, como así también apoyará a una mejor construcción y adaptación del depósito de datos.

Campos del sur es una organización con fines de lucro, ubicada en la calle Ituzaingó 87 de la ciudad de Córdoba constituida en el año 2000. Además se desempeña en actividades de Agricultura y Agro negocios en diferentes provincias teniendo a su cargo la siembra, cosecha y comercialización de cultivos.

Más detalle se desarrolla en el Anexo 1.

## **PASOS Y APLICACIÓN METODOLÓGICA**

### **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 a la vez, es decir, si se requiere construir por ejemplo dos Data Marts, se deberá aplicar la metodología dos veces, una por cada Data Mart. 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.

#### **a) Identificar preguntas**

El primer paso comienza con la recopilación de las necesidades de información, la 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.

## **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 comúnmente una perspectiva.

### c) Modelo Conceptual

En esta etapa, se construirá un modelo conceptual 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:

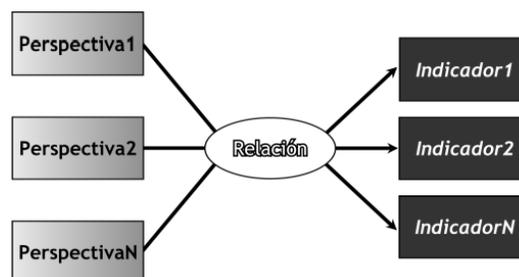


GRÁFICO 2: MODELO CONCEPTUAL DEL DW <sup>6</sup>

---

<sup>6</sup> [HEFESTO]

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.

## **PASO 2) ANÁLISIS DE LOS OLTP**

Seguidamente, 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.

### **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 sumarización que se utilizará para su agregación. Por ejemplo: SUM, AVG, COUNT, etc.

## **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.

## **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.

#### 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:

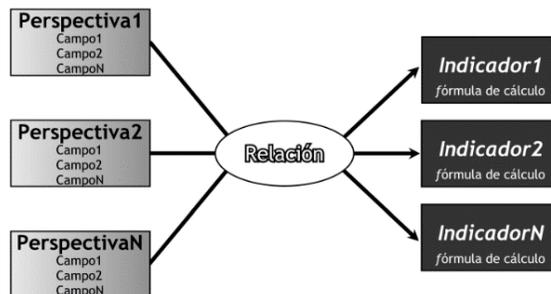


GRÁFICO 3: MODELO CONCEPTUAL AMPLIADO <sup>7</sup>

---

<sup>7</sup> [HEFESTO]

### **PASO 3) MODELO LÓGICO DEL DW**

A continuación, se confeccionará el modelo lógico de la estructura del DW, teniendo como base el modelo conceptual que ya ha sido creado. Para ello, primero se definirá 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 realizarán las uniones pertinentes entre estas tablas.

#### **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.

#### **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:

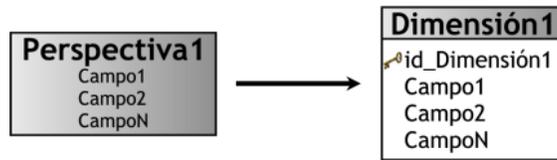


GRÁFICO 4: TABLAS DE DIMENSIONES

Para los esquemas copo de nieve, cuando existan jerarquías dentro de una tabla de dimensión, esta tabla deberá ser normalizada. Por ejemplo, se tomará como referencia la siguiente tabla de dimensión y sus respectivas relaciones padre-hijo entre sus campos:

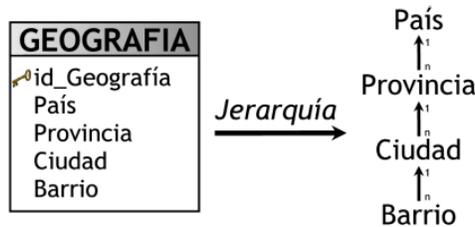


GRÁFICO 5: JERARQUÍAS EN TABLA DE DIMENSIÓN

Entonces, al normalizar esta tabla se obtendrá:

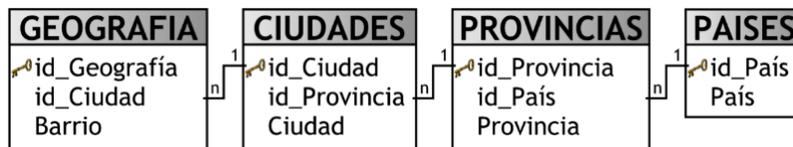


GRÁFICO 6: TABLA DE DIMENSIÓN NORMALIZADA

### 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 realizará lo siguiente:

- 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:

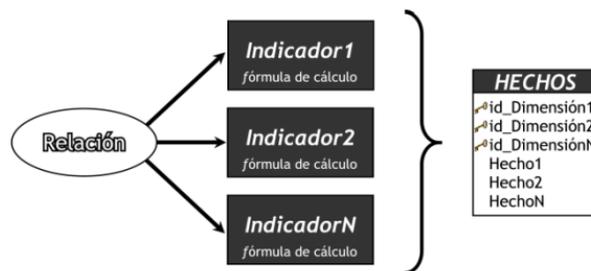


GRÁFICO 7: TABLA DE HECHOS

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. Por ejemplo:

"Analizar el **Indicador1** por **Perspectiva1** y por **Perspectiva2**".

"Analizar el **Indicador1** por **Perspectiva2** y por **Perspectiva3**".

Entonces se obtendrá:

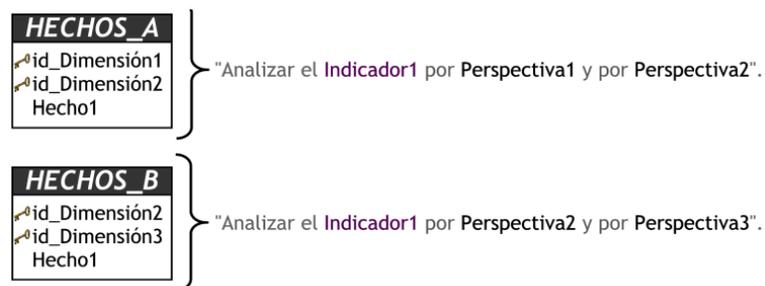


GRÁFICO 8: DISEÑO DE TABLA DE HECHOS

- 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. Por ejemplo:

"Analizar el **Indicador1** por **Perspectiva1** y por **Perspectiva2**".

"Analizar el **Indicador2** por **Perspectiva1** y por **Perspectiva2**".

Se unificará en:

"Analizar el **Indicador1** y el **Indicador2** por **Perspectiva1** y por **Perspectiva2**".

#### **d) Uniones**

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

### **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.

## **a) Carga Inicial**

Debemos en este paso realizar la Carga Inicial al DW, poblando el modelo de datos que hemos construido anteriormente. Para lo cual debemos 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, y que nos facilitarán el trabajo.

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. Por ejemplo, es muy común que sistemas ETL trabajen con "pasos" y "relaciones", en donde cada "paso" realiza una tarea en particular del proceso ETL y cada "relación" indica hacia donde debe dirigirse el flujo de datos. En este caso lo que se debe hacer es explicar que hace el proceso en general y luego que hace cada "paso"

y/o "relación". Es decir, se partirá de lo más general y se irá a lo más específico, para obtener de esta manera una visión general y detallada de todo el proceso.

Es importante tener presente, que al cargar los datos en las tablas de hechos pueden utilizarse preagregaciones, ya sea al nivel de granularidad de la misma o a otros niveles diferentes.

### **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.

## Creación de Cubos Multidimensionales

A continuación se creará un cubo multidimensional de ejemplo, que será llamado “Cubo de Ventas” y que estará basado en el modelo lógico diseñado en el caso práctico de la metodología Hefesto:

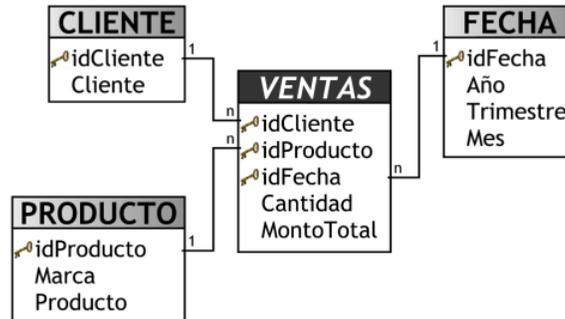


GRÁFICO 9: CUBO MULTIDIMENSIONAL

La creación de este cubo tiene las siguientes finalidades:

- Ejemplificar la creación de cubos multidimensionales.
- Propiciar la correcta distinción entre hechos de una tabla de hechos e indicadores de un cubo.
- Propiciar la correcta distinción entre campos de una tabla de dimensión y atributos de un cubo.

## Creación de Indicadores

En este momento se crearán dos indicadores que serán incluidos en el cubo “Cubo de Ventas”:

- De la tabla de hechos “VENTAS”, se sumará el hecho “Cantidad” para crear el indicador denominado:
  - “Unidades Vendidas”.

La fórmula utilizada para crear este indicador es la siguiente:

- “Unidades Vendidas” =  $SUM(VENTAS.Cantidad)$ .

De la tabla de hechos “VENTAS”, se sumará el hecho “MontoTotal” para crear el indicador denominado:

- “Monto Total de Ventas”.

La fórmula utilizada para crear este indicador es la siguiente:

- “Monto Total de Ventas” =  $SUM(VENTAS.MontoTotal)$ .

Entonces, el cubo quedaría conformado de la siguiente manera:

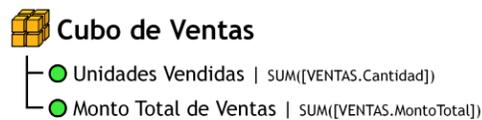


GRÁFICO 10: CUBO MULTIDIMENSIONAL DE VENTAS

### Creación de Atributos

Ahora se crearán y agregarán al cubo seis atributos:

De la tabla de dimensión “CLIENTE”, se tomará el campo “Cliente” para la creación del atributo denominado:

- “Clientes”.

De la tabla de dimensión “PRODUCTO”, se tomará el campo “Marca” para la creación del atributo denominado:

- “Marcas”.

De la tabla de dimensión “PRODUCTO”, se tomará el campo “Producto” para la creación del atributo denominado:

- “Productos”.

De la tabla de dimensión “FECHA”, se tomará el campo “Año” para la creación del atributo denominado:

- “Años”.

De la tabla de dimensión “FECHA”, se tomará el campo “Trimestre” para la creación del atributo denominado:

- “Trimestres”.

De la tabla de dimensión “FECHA”, se tomará el campo “Mes” para la creación del atributo denominado:

- “Mes”.

Entonces, el cubo quedaría conformado de la siguiente manera:

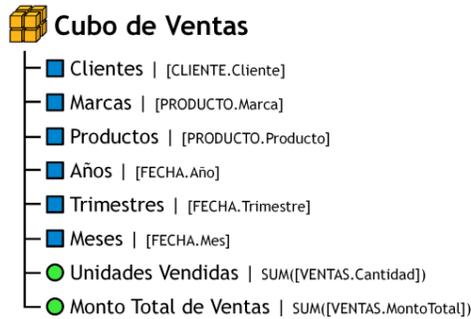


GRÁFICO 11: CUBO MULTIDIMENSIONAL DE VENTAS

### Creación de Jerarquías

Finalmente se crearán y agregarán al cubo dos jerarquías:

- Se definió la jerarquía “Jerarquía Productos”, que se aplicará sobre los atributos recientemente creados, “Marcas” y “Productos”, en donde:
- Un producto en especial pertenece solo a una marca. Una marca puede tener uno o más productos.

Gráficamente:

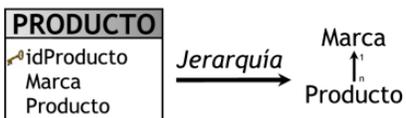


GRÁFICO 12: JERARQUÍAS DE UN CUBO

- Se definió la jerarquía “Jerarquía Fechas”, que se aplicará sobre los atributos recientemente creados, “Años”, “Trimestres” y “Meses”, en donde:
  - Un mes del año pertenece solo a un trimestre del año. Un trimestre del año tiene uno o más meses del año.
  - Un trimestre del año pertenece solo a un año. Un año tiene uno o más trimestres del año.

Gráficamente:

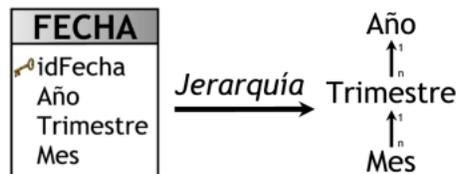


GRÁFICO 13: JERARQUÍAS DE UN CUBO

Entonces, el cubo quedaría conformado de la siguiente manera:

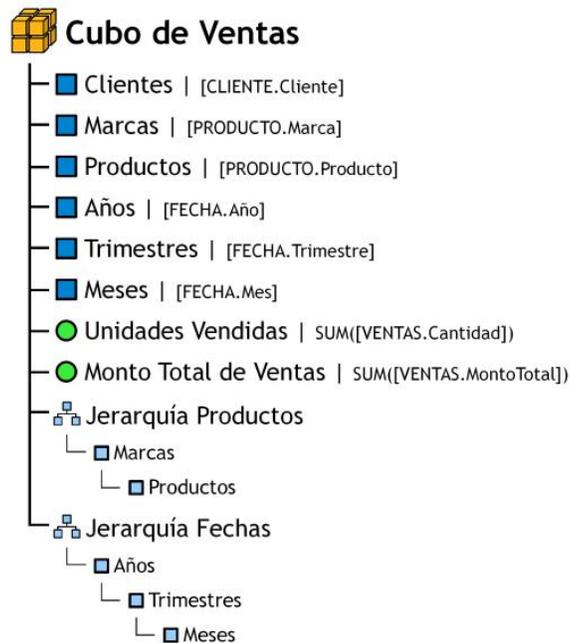


GRÁFICO 14: JERARQUÍA DETALLADA DE UN CUBO

### Otros ejemplos de cubos multidimensionales

A partir del modelo lógico planteado, podrían haberse creado una gran cantidad de cubos, cada uno de los cuales estaría orientado a un tipo de análisis en particular. Tal y como se explicó antes, estos cubos pueden coexistir sin ningún inconveniente.

A continuación se expondrán una serie de cubos de ejemplo:

### Cubo 1:

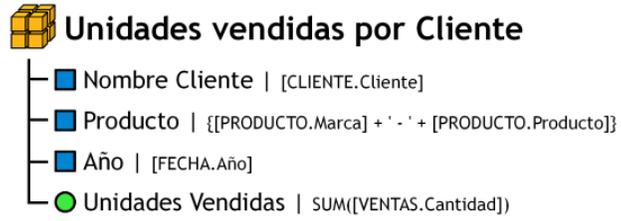


GRÁFICO 15

### Cubo 2:

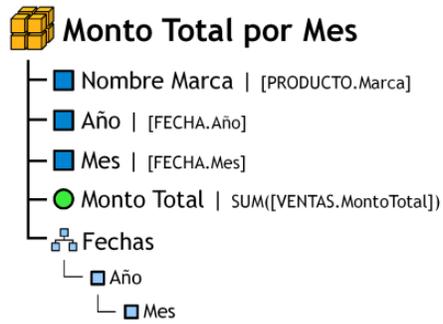


GRÁFICO 16

### Cubo 3:

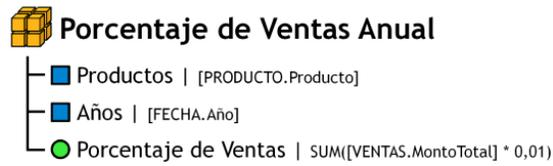


GRÁFICO 17

### **Consideraciones de diseño:**

Dependiendo del negocio, el volumen de datos y el alcance del proyecto, el tamaño del DW puede variar considerablemente, por lo cual, es una buena práctica tener esto en cuenta al momento de diseñar el depósito y al determinar los recursos físicos, los tiempos de desarrollo y los respectivos costos inherentes.

De acuerdo al tamaño del depósito de datos, se lo puede clasificar como:

- Personal: si su tamaño es menor a 1 Gigabyte.
- Pequeño: si su tamaño es mayor a 1 Gigabyte y menor a 50 Gigabyte.
- Mediano: si su tamaño es mayor a 50 Gigabyte y menor a 100 Gigabyte.
- Grande: si su tamaño es mayor a 100 Gigabyte y menor a 1 Terabyte.
- Muy grande: si su tamaño es mayor a 1 Terabyte.

### **Tiempo de construcción**

Diversos autores resaltan la importancia del factor tiempo en la construcción de un DW, por lo cual se ha considerado interesante exponer tres frases seleccionadas al respecto:

“El 70 % del tiempo total dedicado al proyecto se insume en definir el problema y en preparar la tabla de datos”.

“Estime el tiempo necesario, multiplíquelo por dos y agregue una semana de resguardo”.

”Regla 90 – 90”: el primer 90 % de la construcción de un sistema absorbe el 90 % del tiempo y esfuerzo asignados; el último 10 % se lleva el otro 90 % del tiempo y esfuerzo asignado.

### **Implementación**

Las implementaciones de los depósitos de datos varían entre sí de forma considerable, teniendo en cuenta las herramientas de software que se empleen, los modelos que se utilicen, recursos disponibles, SGBD que lo soporten, herramientas de análisis y consulta, entre otros.

## **Performance**

Cuando se diseñan los ETLs, es muy importante que los mismos sean lo más eficiente posible, ya que una vez que se tenga un gran volumen de datos, el espacio en disco se volverá fundamental y los tiempos incurridos en el procesamiento y acceso a la información serán esenciales, y más aún si el data warehouse es considerado o tomado como un sistema de misión crítica.

También es muy importante configurar correctamente el sistema de gestión de base de datos en el que se almacene y mantenga el data warehouse, así como lo es elegir las mejores estrategias para modelar las diferentes estructuras de datos que se utilizarán.

Para mejorar la performance del DWH, se pueden llevar a cabo las siguientes acciones sobre el DW y las estructuras de datos (cubos multidimensionales, Business Models, etc):

Prestar especial atención a los tipos de datos utilizados, por ejemplo, para valores enteros pequeños conviene utilizar tinyint o smallint en lugar de int, con el fin de no asignar tamaños de datos mayores a los necesarios. Esto toma vital importancia cuando se aplica en las claves primarias, debido a que formarán parte de la tabla de hechos que es la que contiene el volumen del almacén de datos.

- Utilizar Claves Subrogadas.
- Utilizar técnicas de indexación.
- Utilizar técnicas de particionamiento.
- Crear diferentes niveles de sumalización.

- Crear vistas materializadas.
- Utilizar técnicas de administración de datos en memoria caché.
- Utilizar técnicas de multiprocesamiento, con el objetivo de agilizar la obtención de resultados, a través de la realización de procesos en forma concurrente.

### **Mantenimiento**

Un punto muy importante es mantener en correcto funcionamiento al DW, ya que a medida que pase el tiempo, este tenderá a crecer significativamente, y surgirán cambios, tanto en los requerimientos como en las fuentes de datos.

### **Impactos**

Al implementar un DWH, es fundamental que los usuarios del mismo participen activamente durante todo su desarrollo, debido a que son ellos los que conocen en profundidad su negocio y saben cuáles son los resultados que se desean obtener. Además, es precisamente en base a la utilización que se le dé, que el depósito de datos madurará y se adaptará a las situaciones cambiantes por las que atraviese la empresa. Los usuarios, al trabajar junto a los desarrolladores y analistas podrán comprender más en profundidad sus propios sistemas operacionales, con todo lo que esto implica.

Con la implementación del data warehouse, los procesos de toma de decisiones serán optimizados, al obtener información correcta al instante en que se necesita, evitando pérdidas de tiempo y anomalías en los datos. Al contar con esta información, los usuarios tendrán más confianza en las decisiones que tomarán y en adición a ello, poseerán una base sustentable para justificarlas.

Usualmente, los data warehouse integrarán fuentes de datos de diversas áreas y sectores de la empresa, esto tendrá como beneficio contar con una sola fuente de información centralizada y común para todos los usuarios. Esto posibilitará que en las diferentes áreas se compartan los mismos datos, lo cual conducirá a un mayor entendimiento, comunicación, confianza y cooperación entre las mismas.

El data warehouse introducirá nuevos conceptos tecnológicos y de inteligencia de negocios, lo cual requerirá que se aprendan nuevas técnicas, herramientas, métodos, destrezas, formas de trabajar, etc.

### **Data Mart como sub proyectos**

Al diseñar e implementar un data mart como partes de un proyecto de data warehouse, se debe tener en cuenta que el análisis que se efectuará, los modelos que intervendrán y el alcance, deben ser globales, con el fin de determinar, por ejemplo, tablas de dimensiones comunes entre las diferentes áreas de trabajo. Esto evitará que se realicen tareas repetidas, ahorrando tiempos y enfocándose en la consolidación, unificación y centralización de la información de los diferentes sectores.

### **XIII. TERCERA PARTE: MODELO TEÓRICO**

En este capítulo se expondrán algunas de las alternativas de software disponible para la implementación de un data warehouse, una breve descripción del proceso de selección y su desarrollo según la metodología planteada.

#### **Pentaho**

Pentaho es una suite de BI open-source que es un producto llamado Kettle para integración de datos, utiliza un enfoque innovador y tiene una GUI robusta y muy fácil de usar. La compañía inició alrededor del 2001 (en el 2002 fue cuando se integró kettle), Esta tiene una comunidad fuerte con 13.500 usuarios registrados. Tiene un motor de Java independiente que procesa los trabajos y tareas para mover datos entre diversas bases de datos y archivos distintos. Tiene la capacidad de agendar tareas (pero es necesario hacerlo con un programador de tareas). Puede ejecutar trabajos remotos en servidores esclavos en otras máquinas.

Tiene características de calidad de datos: desde su propia interfaz gráfica de usuario, permitiendo escribir consultas SQL más personalizadas, Javascript y expresiones regulares.

#### **KETL**

La herramienta ETL Kettle permite a las empresas gestionar una compleja manipulación de los datos, mientras que el aprovechamiento de la capacidad de una plataforma de integración de datos de código abierto. El motor que proporciona la fundación es flexible y probado en producción de extracción, transformación y carga (ETL).

En lugar de desarrollar la lógica personalizada que es difícil de depurar, monitorizar y mantener, Kettle ofrece una solución basada en metadatos que añade robustez y funcionalidad. Con soporte para módulos ensamblables la plataforma permite código abierto y las empresas comerciales para aprovechar la funcionalidad ya existente.

### **Características Kettle™:**

- Escalable, plataforma ETL independientes, motor que permite transformaciones ETL complejas para ser ejecutados de una manera altamente eficiente. Soporta múltiples servidores de 64 bits CPU y.
- La ejecución del trabajo y la programación impulsada-manager-dependencia modelo de ejecución de trabajo soporta múltiples tipos de trabajo, manejo de excepciones condicional, de notificación de correo electrónico y programación basada en el tiempo. Los tipos de trabajo se dividen en tres categorías, con soporte para los ejecutores adicionales a través de la API Kettle.
  - SQL->Ejecuta SQL-declaración predefinido SQL a través de JDBC
  - XML-> Ejecuta trabajos XML definido
  - OS-ejecuta un comando del sistema operativo
- XML de definición de trabajo de lenguaje permite que los trabajos de ETL para ser fácilmente definidos en XML, lo que permite el uso de las herramientas de creación XML ampliamente disponibles y el apoyo asociado para los sistemas de control de versiones.
- Repositorio centralizado que soporta múltiples instancias Kettle para aprovechar los trabajos y definiciones de parámetros.
- Rendimiento monitoreo recoge estadísticas históricas y activas de empleo en el repositorio, lo que permite el análisis exhaustivo de los trabajos problemáticos.

- Integrales de origen de datos de apoyo-soportes de extracción y carga de fuentes relacionales, archivos planos y datos XML, a través de JDBC y API de base de datos de propiedad.
- Programación y ejecución del trabajo orientado a eventos del motor de base de tiempo.

### **Jaspersoft ETL**

Jaspersoft ETL es fácil de implementar y supera a muchos sistemas ETL de fuente propietaria y de código abierto. Se utiliza para extraer los datos de un sistema transaccional para crear un consolidado datawarehouse o datamart para realizar reportes y análisis.

Esta herramienta posee integración con dispositivos móviles para realizar presentación de reportes

Una wiki completa con descripciones detalladas de inicio rápido, instalación y tutoriales básicos

Cuenta con diversas extensiones, algunas de ellas son:

### **RevoConnectR para JasperReports servidor**

RevoConnectR para JasperReports Server es una interfaz de biblioteca de Java entre JasperReports Server y Revolución R Enterprise™ RevoDeployR, una recopilación estandarizada de servicios web que integra seguridad, APIs, scripts y bibliotecas de R en un único servidor. Tablas R y conjuntos de resultados tableros JasperReports Server pueden recuperar de RevoDeployR.

## **JasperServer Repository FTP interface**

Los usuarios pueden utilizar el cliente FTP para acceder a los objetos del repositorio. La interfaz aprovecha plenamente la autenticación JS. Funciona tanto con la versión community y las versiones comerciales. Para entorno de varios clientes, el formato de la cadena de usuario ftp es: nombre de usuario | nombre de la organización. La versión actual limita el acceso sólo a la carpeta / themes.

## DESARROLLO DE LA METODOLOGÍA

### 1) ANALISIS DE REQUERIMIENTOS:

#### Identificar preguntas:

- \* Se desea saber el promedio de precipitaciones por año.
- \* Se desea saber el número de días de precipitaciones por año.
- \* Se desea saber la cantidad de agua total utilizada para riego por mes por sector
- \* Se desea saber la humedad promedio del suelo por mes por sector.
- \* Se desea saber la humedad promedio del ambiente por día, mes o año.
- \* Se desea saber la temperatura máxima por año,mes,día por sensor
- \* Se desea saber la temperatura mínima por mes, sensor.
- \* Se desea saber la velocidad máxima y dirección del viento por periodos fecha (años, meses, semanas).
- \* Se desea saber la variación de la temperatura a través de los años.
- \* Se desea conocer el valor de la humedad del suelo.

#### Indicadores y perspectivas

##### Indicadores:

- \*Promedio de valor de temperatura.
- \*Promedio de valor de humedad ambiente
- \*Promedio de valor de humedad suelo.
- \*Promedio de valor de precipitaciones.
- \*Promedio de agua utilizada para riego.
- \*Cantidad de días de precipitaciones.
- \*Mínimo valor de precipitaciones.
- \*Mínimo valor de temperatura.
- \*Mínimo valor de humedad ambiente.
- \*Mínimo valor de humedad ambiente.
- \*Mínimo valor de agua utilizada para riego.

- \*Máximo valor de precipitaciones.
- \*Máximo valor de temperatura.
- \*Máximo valor de humedad ambiente.
- \*Máximo valor de humedad ambiente.
- \*Máximo valor de agua utilizada para riego.
- \*Velocidad y rumbo del viento.

#### Perspectivas

- \* Por año.
- \* Por trimestre
- \* Por mes.
- \* Por día
- \* Por sector.
- \* Por sensor.
- \* Por equipo.

Modelo conceptual

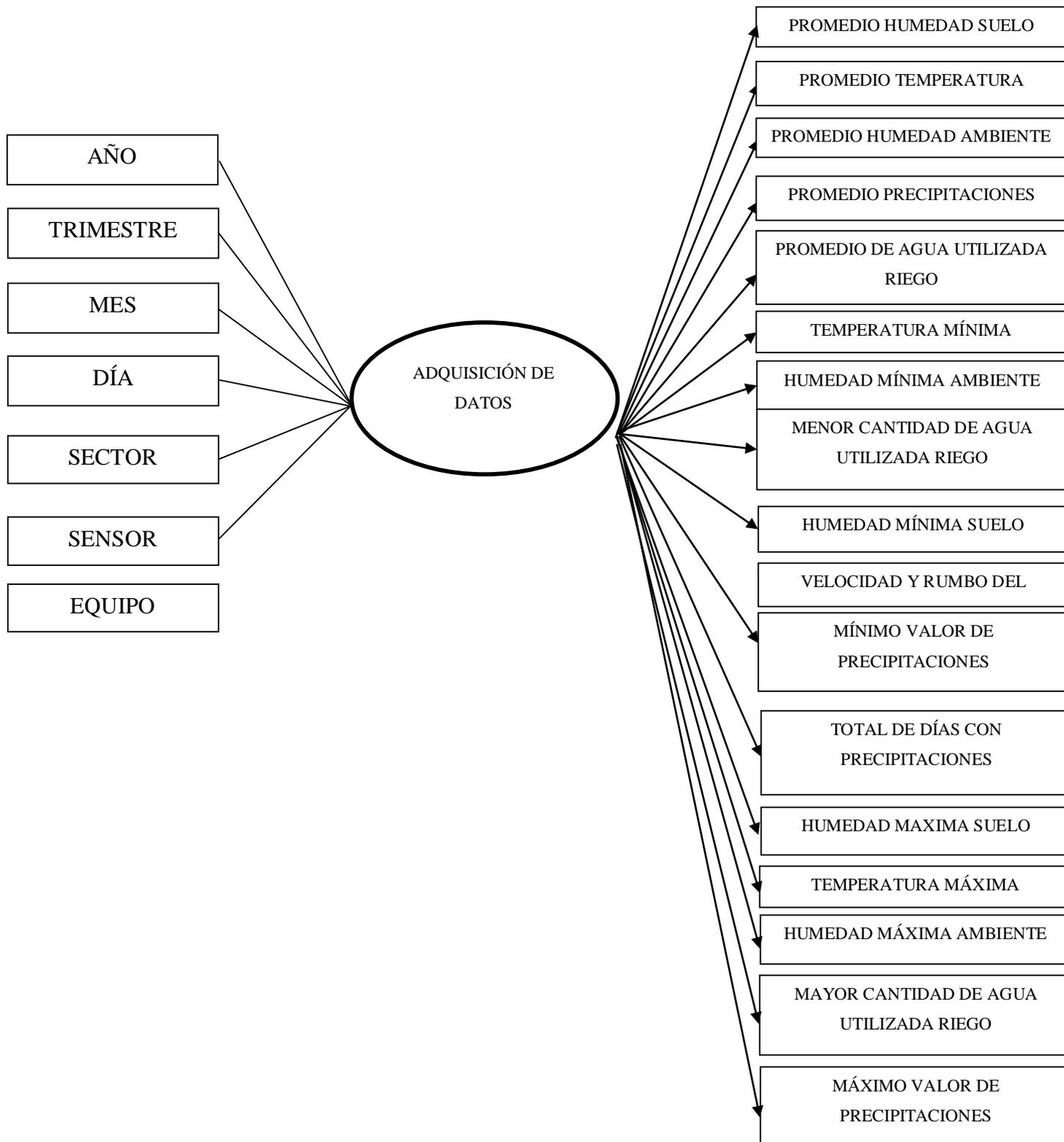


GRÁFICO 18: MODELO CONCEPTUAL

## 2) ANÁLISIS DE LOS OLTP

Conformar indicadores:

### **Valor promedio temperatura**

Hecho: Valor registro

Función: Promedio.

### **Valor promedio humedad ambiente**

Hecho: Valor registro

Función: Promedio.

### **Valor promedio humedad suelo**

Hecho: Valor registro

Función: Promedio.

### **Valor promedio precipitaciones**

Hecho: Valor registro

Función: Promedio.

### **Valor promedio de agua para riego**

Hecho: Valor registro

Función: Promedio.

### **Valor mínimo de temperatura**

Hecho: Valor registro

Función: Mínimo.

### **Valor máximo humedad ambiente**

Hecho: Valor registro

Función: Máximo.

### **Valor velocidad y rumbo viento**

Hecho: Valor registro

Función: valor.

### **Valor total precipitaciones**

Hecho: Valor registro

Función: SUM.

### **Valor mínimo de temperatura**

Hecho: Valor registro

Función: Mínimo.

### **Valor mínimo de humedad ambiente**

Hecho: Valor registro

Función: Mínimo.

### **Valor mínimo de humedad suelo**

Hecho: Valor registro

Función: Mínimo.

### **Valor mínimo de agua para riego**

Hecho: Valor registro

Función: Mínimo.

### **Valor mínimo de precipitaciones**

Hecho: Valor registro

Función: Mínimo.

### **Valor máximo temperatura**

Hecho: Valor registro

Función: Máximo.

### **Valor máximo humedad suelo**

Hecho: Valor registro

Función: Máximo.

### **Valor máximo precipitaciones**

Hecho: Valor registro

Función: MAX.

## Establecer correspondencias

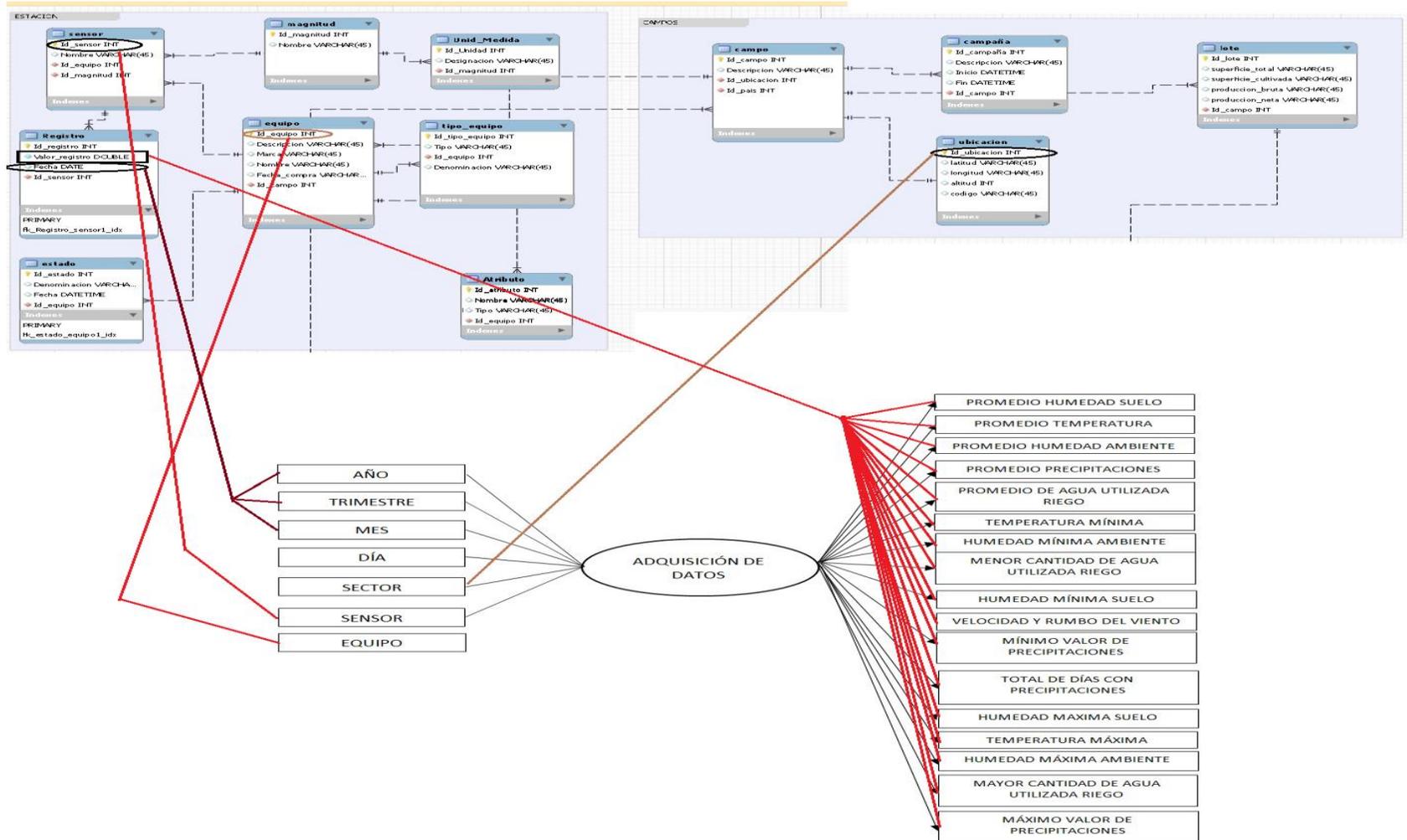


GRÁFICO 19: CORRESPONDENCIAS

- El campo “**Fecha**” de la tabla “**registro**” se relaciona con la perspectiva “**año**”.
- El campo “**Fecha**” de la tabla “**registro**” se relaciona con la perspectiva “**trimestre**”.
- El campo “**Fecha**” de la tabla “**registro**” se relaciona con la perspectiva “**mes**”.
- El campo “**Fecha**” de la tabla “**registro**” se relaciona con la perspectiva “**día**”.
- El campo “**Id\_equipo**” de la tabla “**equipo**” se relaciona con la perspectiva “**equipo**”.
- El campo “**Id\_ubicacion**” de la tabla “**ubicacion**” se relaciona con la perspectiva “**sector**”.
- El campo “**Id\_sensor**” de la tabla “**sensor**” se relaciona con la perspectiva “**sensor**”.
- El promedio del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea HUMEDAD DEL SUELO se relaciona con el **indicador** “PROMEDIO HUMEDAD DEL SUELO”.
- El valor promedio del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea TEMPERATURA se relaciona con el **indicador** “PROMEDIO DE TEMPERATURA”.
- El promedio del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea HUMEDAD RELATIVA AMBIENTE se relaciona con el **indicador** “PROMEDIO HUMEDAD REALTIVA AMBIENTE”.
- El promedio del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea PLUVIÓMETRO se relaciona con el **indicador** “PROMEDIO DE PRECIPITACIONES”.

- El valor promedio del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea AGUA se relaciona con el **indicador** “PROMEDIO DE AGUA UTILIZADA PARA RIEGO”.
- El valor mínimo del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea TEMPERATURA se relaciona con el **indicador** “TEMPERATURA MÍNIMA”.
- El valor mínimo del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea HUMEDAD RELATIVA AMBIENTE se relaciona con el **indicador** “HUMEDAD MÍNIMA REALTIVA AMBIENTE”.
- El valor mínimo del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea AGUA se relaciona con el **indicador** “MENOR CANTIDAD DE AGUA UTILIZADA PARA RIEGO”.
- El valor mínimo del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea HUMEDAD SUELO se relaciona con el **indicador** “HUMEDAD MÍNIMA SUELO”.
- El valor máximo del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea ANEMÓMETRO se relaciona con el **indicador** “VELOCIDAD DEL VIENTO”.
- El valor del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea RUMBO se relaciona con el **indicador** “RUMBO DEL VIENTO”.
- El valor mínimo del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea PRECIPITACIÓN se relaciona con el **indicador** “MÍNIMO VALOR DE PRECIPITACIONES”.
- El valor COUNT del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea PRECIPITACIÓN se relaciona con el **indicador** “TOTAL DE DÍAS CON PRECIPITACIONES”.

- El valor máximo del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea HUMEDAD DEL SUELO se relaciona con el **indicador** “HUMEDAD MÁXIMA DEL SUELO”.
- El valor maximo del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea TEMPERATURA se relaciona con el **indicador** “TEMPERATURA MÁXIMA”.
- El valor máximo del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea HUMEDAD RELATIVA AMBIENTE se relaciona con el **indicador** “HUMEDAD MÁXIMA RELATIVA AMBIENTE”.
- El valor máximo del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea AGUA se relaciona con el **indicador** “MAYOR CANTIDAD DE AGUA UTILIZADA PARA RIEGO”.
- El valor MAX del campo “**valor\_registro**” de la tabla “**registro**”, en conjunto con el campo “Nombre” de la tabla “sensor” cuyo valor sea PLUVIÓMETRO se relaciona con el **indicador** “MÁXIMO VALOR DE PRECIPITACIONES”.

### Nivel de Granularidad

Con respecto a la perspectiva “**TIEMPO**” los datos disponibles son los siguientes:

- Fecha

Del cual se desprenden los siguientes niveles de análisis.

- Año
- Semestre
- Estación del año.
- Número de Mes
- Nombre de Mes
- Quincena
- Semana
- Día de la semana
- Nombre del día de la semana

Con respecto a la perspectiva “**SECTOR**” los datos disponibles son los siguientes:

- Id\_ubicacion
- Latitud
- Longitud
- Altitud
- Codigo\_sector

Del cual se desprenden los siguientes niveles de análisis.

- Id\_sector

Con respecto a la perspectiva “**SENSOR**” los datos disponibles son los siguientes:

- Id\_sensor
- Nombre
- Id\_equipo
- Id\_magnitud

Del cual se desprenden los siguientes niveles de análisis.

- Id\_Sensor
- Nombre

Con respecto a la perspectiva “**EQUIPO**” los datos disponibles son los siguientes:

- Id\_equipo
- Descripcion
- Marca
- Nombre
- Fecha\_compra
- Id\_campo

Del cual se desprenden los siguientes niveles de análisis.

- Id\_equipo
- Nombre

Modelo Conceptual Ampliado

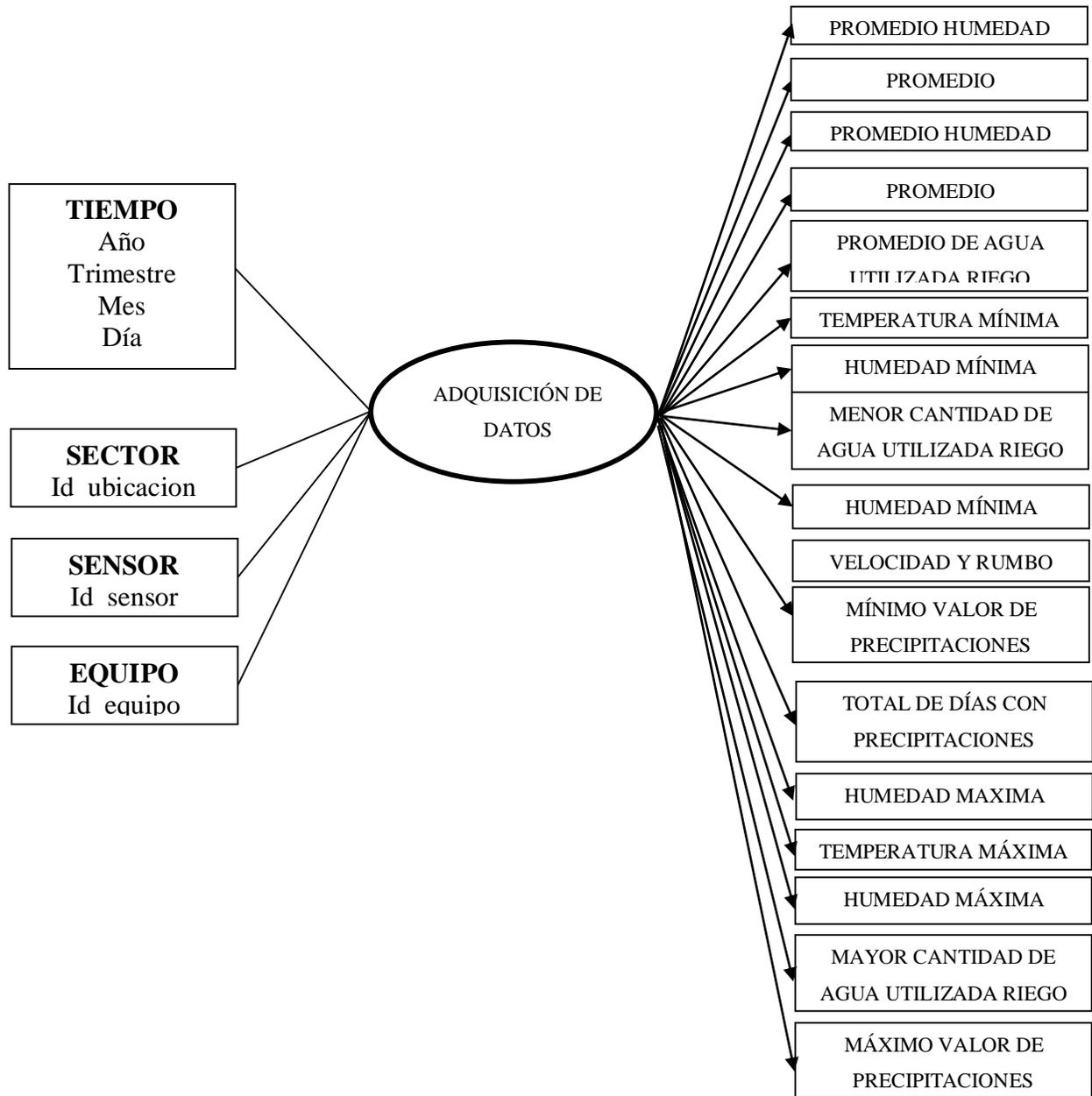


GRÁFICO 20: MODELO CONCEPTUAL AMPLIADO

### 3) MODELO LÓGICO DEL DATA WAREHOUSE

#### Tipo de Modelo Lógico del DW

El modelo lógico que se implementará es constelación ya que:

- Posibilita obtener mejores tiempos de respuesta.
- Permite modificar fácilmente su diseño.
- 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.

#### Tablas de Dimensiones

- Perspectiva **TIEMPO**

La nueva tabla de dimensión tendrá el nombre TIEMPO

Se le agregará una clave principal llamada: Id\_fecha

Los nombres de los campos no serán modificados.

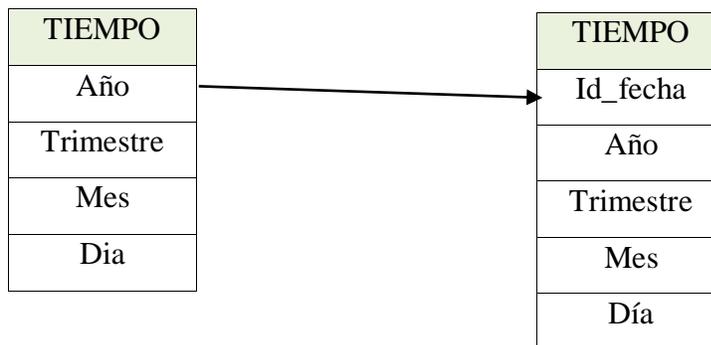


GRÁFICO 21: DIMENSIÓN TIEMPO

- Perspectiva SECTOR

La nueva tabla de dimensión tendrá el nombre SECTOR

Se le agregará una clave principal llamada: Id\_sector

Se modificará el campo “Id\_ubicacion” por “Nombre”

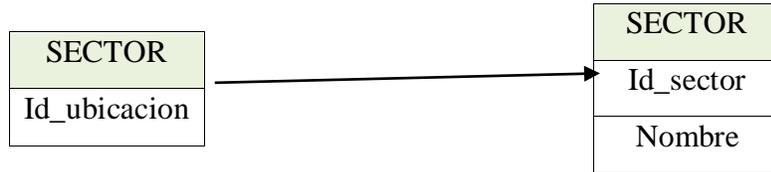


GRÁFICO 22: DIMENSIÓN SECTOR

- Perspectiva SENSOR

La nueva tabla de dimensión tendrá el nombre SENSOR

Se le agregará una clave principal llamada: Id\_sensor

Se modificará el campo “Id\_sensor” por “Nombre”

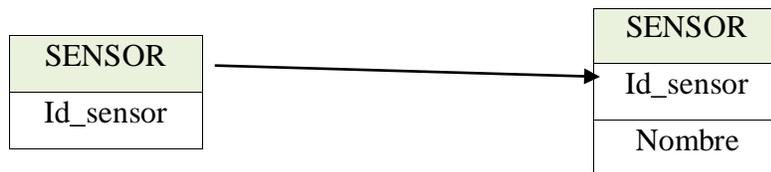


GRÁFICO 23: DIMENSIÓN SENSOR

- Perspectiva EQUIPO

La nueva tabla de dimensión tendrá el nombre EQUIPO

Se le agregará una clave principal llamada: Id\_equipo

Se modificará el campo “Id\_equipo” por “Nombre”

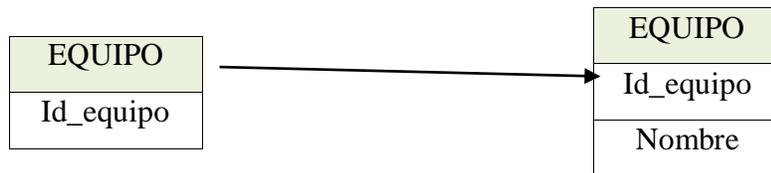


GRÁFICO 24: DIMENSIÓN EQUIPO

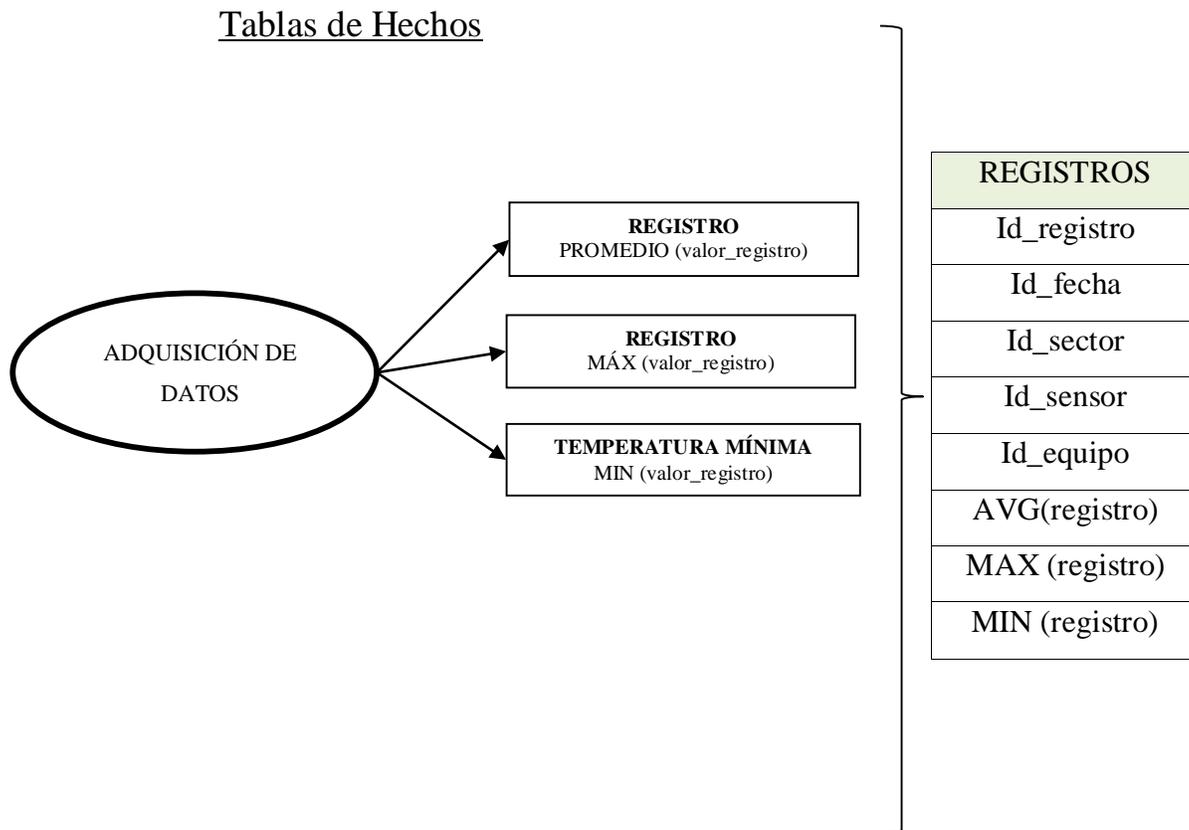


GRÁFICO 25: TABLA DE HECHOS “REGISTROS”

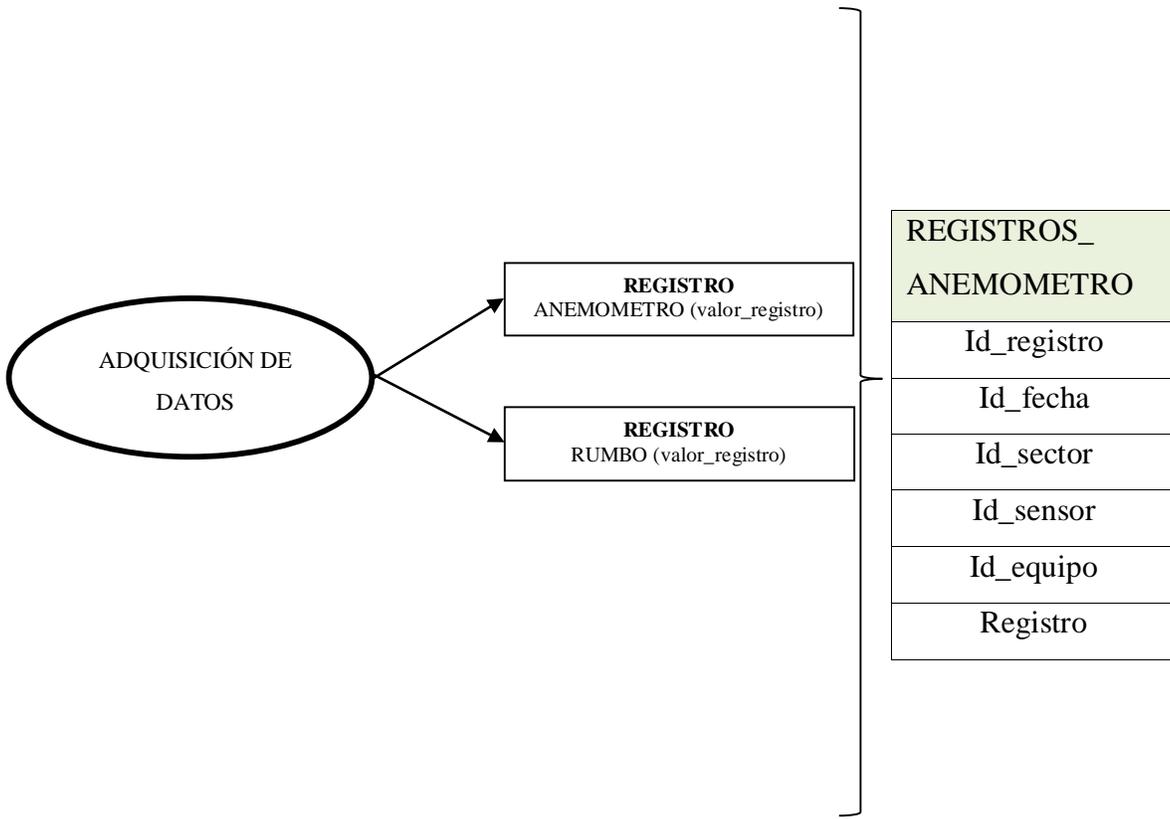


GRÁFICO 26: TABLA DE HECHOS “REGISTROS ANEMOMETRO”

## Uniones

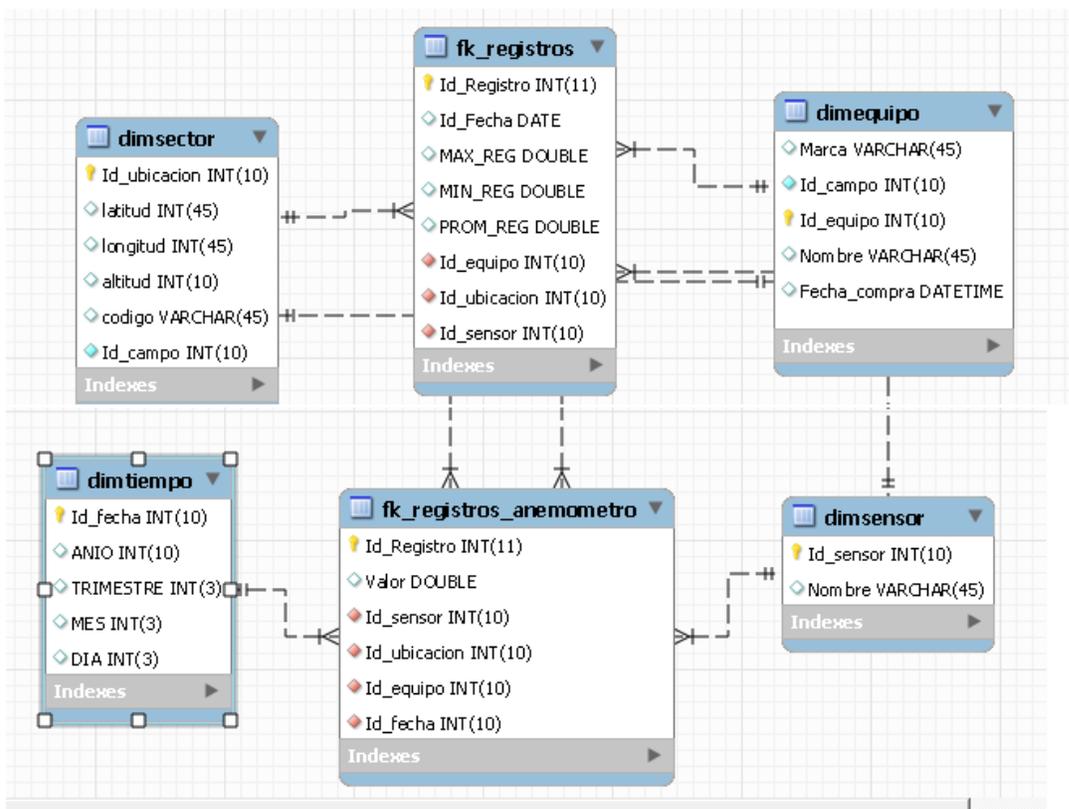


GRÁFICO 27: UNIONES

#### 4) PROCESOS ETL

##### **Tabla dimensión SECTOR:**

```
"SELECT
UBICACION.ID_UBICACION,
CAMPO.ID_CAMPO,
UBICACION.LATITUD,
UBICACION.LONGITUD,
UBICACION.ALTITUD,
UBICACION.CODIGO
FROM UBICACION
INNER JOIN CAMPO
ON CAMPO.ID_UBICACION=UBICACION.ID_UBICACION"
```

##### **Tabla dimensión SENSOR:**

```
"SELECT SENSOR.ID_SENSOR, SENSOR.NOMBRE
FROM SENSOR"
```

##### **Tabla dimensión EQUIPO:**

```
"SELECT
EQUIPO.ID_EQUIPO, EQUIPO.ID_CAMPO, EQUIPO.NOMBRE, EQUIPO.MA
RCA, EQUIPO.FECHA_COMPRA FROM equipo"
SELECT
EQUIPO.DESCRIPCION, EQUIPO.ID_CAMPO, EQUIPO.ID_EQUIPO,
EQUIPO.NOMBRE FROM EQUIPO"
```

**Tabla dimensión TIEMPO:**

```
"SELECT REGISTRO.ID_REGISTRO AS  
ID_FECHA, YEAR(REGISTRO.FECHA) AS  
ANIO, QUARTER(REGISTRO.FECHA) AS TRIMESTRE,  
MONTH(REGISTRO.FECHA) AS MES, DAY(REGISTRO.FECHA) AS DIA  
FROM REGISTRO  
GROUP BY  
YEAR(REGISTRO.FECHA) , MONTH(REGISTRO.FECHA) , DAY(REGISTRO.  
FECHA) "
```

**Tabla de hechos REGISTROS:**

```
" SELECT  
REGISTRO.ID_REGISTRO AS ID_REGISTRO,  
SENSOR.ID_SENSOR AS ID_SENSOR,  
EQUIPO.ID_EQUIPO AS ID_EQUIPO,  
UBICACION.ID_UBICACION AS ID_UBICACION,  
REGISTRO.FECHA AS ID_FECHA,  
MAX(REGISTRO.VALOR_REGISTRO) AS MAX_REG,  
MIN(REGISTRO.VALOR_REGISTRO) AS MIN_REG,  
AVG(REGISTRO.VALOR_REGISTRO) AS PROM_REG  
FROM REGISTRO  
INNER JOIN SENSOR  
ON REGISTRO.ID_SENSOR=SENSOR.ID_SENSOR  
INNER JOIN EQUIPO
```

```

ON SENSOR.ID_EQUIPO=EQUIPO.ID_EQUIPO

INNER JOIN CAMPO

ON EQUIPO.ID_CAMPO=CAMPO.ID_CAMPO

INNER JOIN UBICACION

ON CAMPO.ID_UBICACION=UBICACION.ID_UBICACION

WHERE SENSOR.ID_SENSOR!='5' OR SENSOR.ID_SENSOR!='6'

GROUP BY REGISTRO.FECHA,SENSOR.NOMBRE,EQUIPO.ID_EQUIPO

ORDER BY

YEAR(REGISTRO.FECHA),MONTH(REGISTRO.FECHA),DAY(REGISTRO.
FECHA),EQUIPO.ID_EQUIPO,SENSOR.NOMBRE"

```

**Tabla de hechos REGISTROS ANEMOMETRO:**

```

"SELECT

REGISTRO.ID_REGISTRO AS ID_REGISTRO,

SENSOR.ID_SENSOR AS ID_SENSOR,

EQUIPO.ID_EQUIPO AS ID_EQUIPO,

UBICACION.ID_UBICACION AS ID_UBICACION,

REGISTRO.FECHA AS ID_FECHA,

REGISTRO.VALOR_REGISTRO AS VALOR

FROM REGISTRO

INNER JOIN SENSOR

ON REGISTRO.ID_SENSOR=SENSOR.ID_SENSOR

INNER JOIN EQUIPO

ON SENSOR.ID_EQUIPO=EQUIPO.ID_EQUIPO

```

```
INNER JOIN CAMPO
ON EQUIPO.ID_CAMPO=CAMPO.ID_CAMPO
INNER JOIN UBICACION
ON CAMPO.ID_UBICACION=UBICACION.ID_UBICACION
WHERE SENSOR.ID_SENSOR='5' OR SENSOR.ID_SENSOR='6'
GROUP BY REGISTRO.FECHA,REGISTRO.VALOR_REGISTRO
ORDER BY
YEAR(REGISTRO.FECHA) ,MONTH(REGISTRO.FECHA) ,DAY(REGISTRO.
FECHA) ,EQUIPO.ID_EQUIPO ,SENSOR.NOMBRE"
```

## **RELACIÓN BENEFICIO / COSTO DE LA IMPLEMENTACIÓN DE UN DATA WAREHOUSE**

Se han identificado diversos costos y beneficios en la elaboración de un proyecto de construcción de un data warehouse y sus valores de acuerdo a la tabla de honorarios otorgadas por el Consejo Profesional de Ciencias Informáticas de Córdoba vigentes en el año en curso, tales como:

a) Costos

o *Costos preliminares*

Actividad	Hs	Valor por hora	TOTAL
Planificación	28	\$320	\$8960
Diseño del DW	32	\$436	\$13952
Desarrollo del Proceso ETL	38	\$478	18164
Modelamiento/Ingeniería de Información	35	\$395	\$13825
		TOTAL	\$54901

TABLA 1: COSTOS PRELIMINARES

o *Costos iniciales*

Actividad	Valor
Plataforma de hardware (Servicio Cloud)	7.000 al año
Software de base de datos	Licencia free
Herramientas de proceso ETL y generación de reportes.	Licencia free

TABLA 2: COSTOS INICIALES

○ *Costos en procesamiento*

Actividad	Hs	Valor por hora	
Mantenimiento de datos	26	\$290	\$7540
Desarrollo de reportes	29	\$310	\$8990
Capacitación y soporte	20	\$250	\$5000
		TOTAL	\$21530

TABLA 3: COSTOS DE PROCESAMIENTO

b) Beneficios

*Beneficios operativos*

- Reducción en el consumo de agua, de al menos el 20 %, lo que conlleva a reducir el consumo de energía eléctrica y horas hombre para el encendido apagado de equipos de riego.
- Mayor rendimiento con el mismo nivel de insumos
- Reducción de la aplicación de pesticidas y fertilizantes.

*Beneficios Tácticos*

- Impresión y emisión de reportes con datos claros y concisos.
- Demanda reducida para consultas de inversionistas.

- Entrega más rápida de información a los usuarios.

*Beneficios Estratégicos (Potencialidad)*

- Aplicaciones y herramientas de acceso para los usuarios finales.
- Decisiones con mayor certidumbre.
- Toma de decisiones más ágil.
- Capacidad de soporte a la información organizacional.
- Reducir la incertidumbre al momento de elegir período de tiempo y lugar para realizar la siembra.
- Gestión optimizada de las explotaciones.
- Menor impacto medioambiental.
- Productos con mayor valor nutritivo.
- Obtención de información precisa y de trazabilidad, cuya importancia se acrecienta en las zonas con carencia de nitrógeno.

#### **XIV. CUARTA PARTE: CONCRECIÓN DEL MODELO**

Aplicación del data warehouse al caso de estudio. A continuación se describe la forma en que se llevó a cabo la implementación del data warehouse al caso de estudio. Se describirán los pasos en general que se utilizaron durante la aplicación, ya que la metodología ya fue explicada en el apartado anterior, lo cual puede ser útil al lector, como una guía o ayuda para futuras aplicaciones de los procesos que componen la guía propuesta.

##### **Técnicas utilizadas durante el relevamiento de la información**

Para realizar el relevamiento de información del caso de estudio, se utilizaron algunas de las técnicas sugeridas:

- Cuestionarios: se utilizó un cuestionario, con el cual se buscaba recoger un primer conjunto de datos acerca de la organización y de su infraestructura tecnológica.
- Entrevistas: se realizaron entrevistas en las oficinas de la organización, con las cuales se logró recoger un volumen mayor de información, con más detalle, y con varias personas del área involucrada. Durante las entrevistas se logró tener un contacto más directo con la organización, y de esta manera descubrir y recoger información adicional.
- Visitas a instalaciones: se efectuaron visitas a las instalaciones de la organización. Se realizó una visita al centro de cómputo, así como también una visita al departamento de ingeniería de producto.

Resultado de la guía de la Aplicación:

1) Modelo de datos operacionales

En primera instancia fue necesario la integración y homogeneización de los datos operacionales, con el fin de adecuar los datos para el entorno de DW.

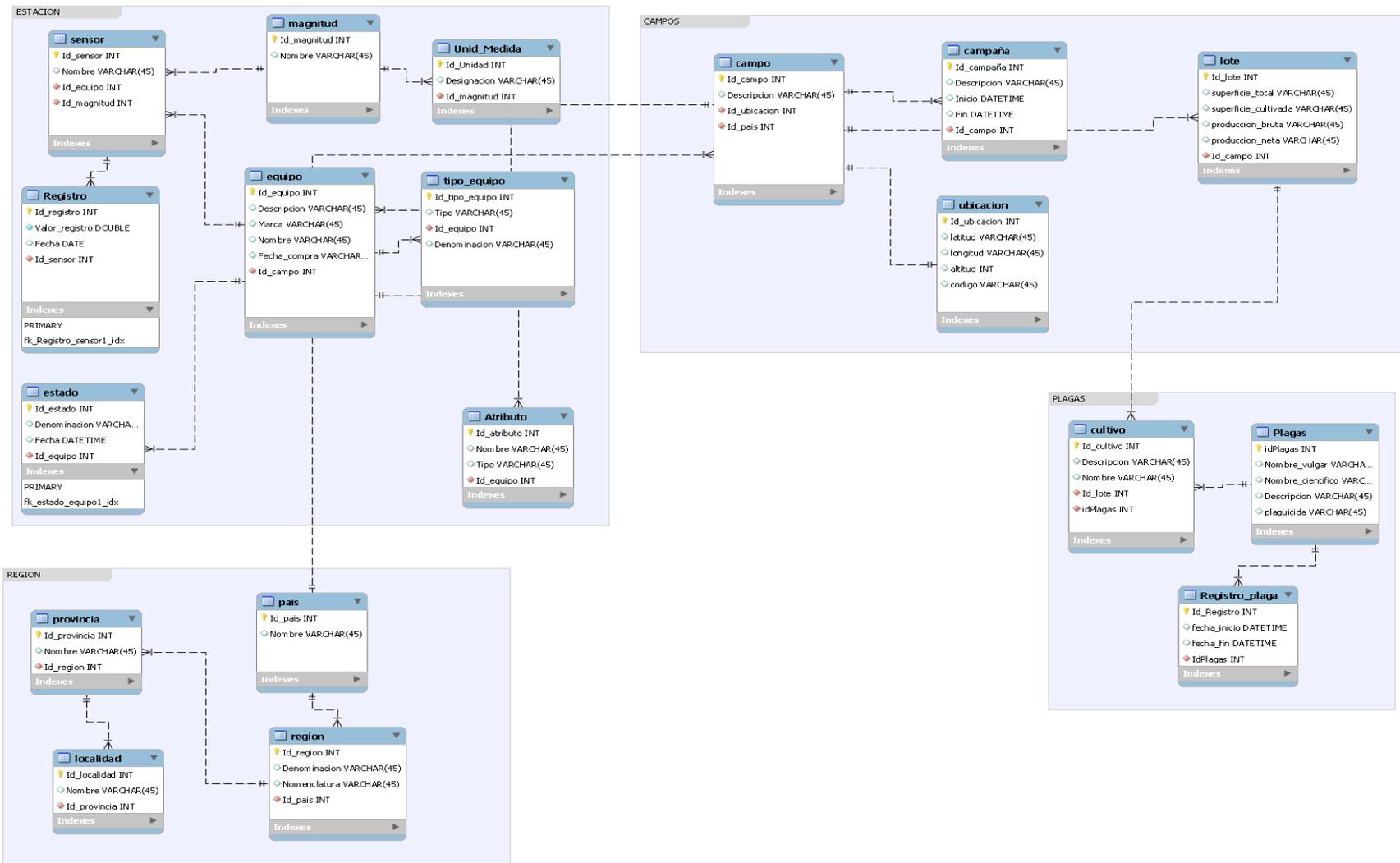


GRAFICO 28: MODELO DE DATOS OPERACIONALES

## 2) Estructura del proceso ETL

Luego, con la herramienta de Jaspersoft ETL, se llevó a cabo el modelado del proceso ETL, en el que mediante consultas SQL, se toman los datos de la base de datos operacional, se extraen los datos necesarios para cumplir con la estructura del DW, se crea la base de datos del DW y en la cual se efectúa el proceso de la carga de los datos.

Se puede apreciar la existencia de la dimensión: TIEMPO, SECTOR, SENSOR Y EQUIPO, correspondiendo el último proceso a la Tabla de Hechos.



GRAFICO 29: MODELO ETL

### 3) Estructura del DW

Luego del proceso ETL, se obtiene finalmente la estructura del data warehouse, sobre la cual se realizarán las consultas SQL que darán cumplimiento a los requerimientos solicitados por los reportes.

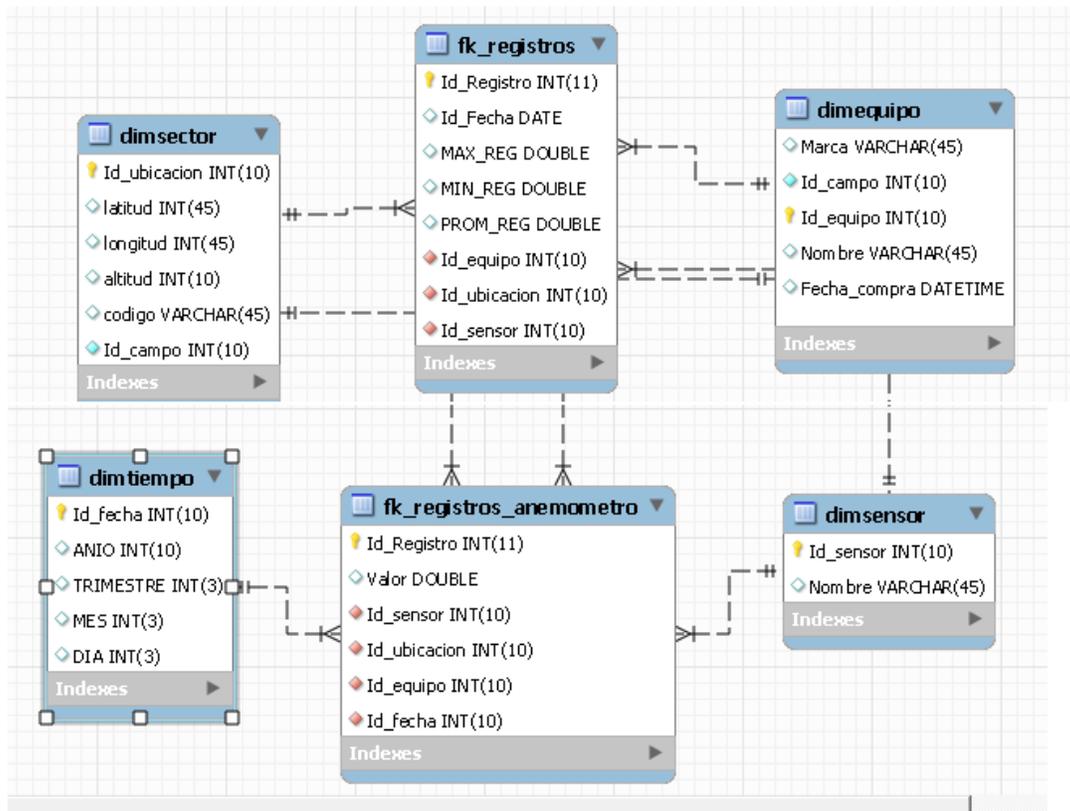


GRAFICO 30: ESTRUCTURA DEL DW

#### 4) Reportes

Finalmente, para satisfacer las necesidades emanadas del proceso de toma de decisiones, con el empleo de la herramienta JasperSoft Studio se realizan los reportes requeridos.



GRAFICO 31: REPORTE DE TEMPERATURA

REPORTE ANUAL  
AÑO ▼ VER REPORTE

REPORTE MENSUAL  
AÑO ▼ MES ▼ VER REPORTE

REPORTE POR PERIODO  
DESDE: dd/mm/aaaa  
HASTA: dd/mm/aaaa  
INTERVALO: AÑO / MES / DIA

GRAFICO 32: FRONT END DEL SISTEMA DE REPORTES

## **XV. CONCLUSIONES**

Luego de la implementación del data warehouse se observaron diversos cambios, algunos en mayor medida que otros. En primer lugar la certeza a la hora de planificar las próximas campañas de siembra fue mucho más sencilla al tener los datos de manera más organizada en los diferentes reportes (gráficos de series, de barra, de columna, de barra horizontal con rango, etc) de los diversos factores climáticos.

El acceso a los datos es más sencillo y ágil, permitiendo a los usuarios hacer sus propias consultas de manera inmediata sin incurrir en demoras.

Luego se realizó un cálculo estimativo de la cantidad de agua utilizada actualmente y la que realmente requerirán los cultivos futuros, evidenciando una reducción del 17%, que conlleva a su vez, a la reducción del consumo de combustible por el uso de los equipos de riego.

Los administradores de los cultivos han podido determinar el comportamiento de los mismos a lo largo del período analizado de acuerdo a los cambios meteorológicos. Pudieron eliminar los retardos de los procesos que resultaban de información incorrecta (presentación de reportes, determinación de los insumos necesarios para la próxima campaña, etc), inconsistente y/o ausente. Se integraron y optimizaron los procesos a través del uso compartido e integrado de las fuentes de información

Como objetivo principal de esta investigación, **se desarrolló una herramienta AD-HOC** que realizó una mejora significativa en la toma de decisiones efectuada en el proceso de agricultura de precisión en un caso real, la empresa citada con anterioridad (AUSA).

Dando cumplimiento a los objetivos específicos, se realizó un **“relevamiento de las herramientas disponibles”** para la elaboración del data warehouse especificado en el apartado XIII: modelo teórico, posterior a ello se realizó una **“selección”** de acuerdo a las necesidades del proyecto.

Con respecto al objetivo **“modelar el data warehouse”** se ha concretado siguiendo la metodología HEFESTO.

Con respecto al objetivo **“construir el DW con las herramientas elegidas”** se llevó a cabo con la selección de la metodología HEFESTO y la herramienta ETL de jaspersoft.

Para el objetivo **“realizar la implementación de la plataforma de Software del DW”** se realizó, tomando en cuenta las herramientas y técnicas disponibles de acuerdo al estudio de factibilidad planteado.

Para el objetivo **“Realizar el análisis costo / beneficio de esta herramienta”** se realizó de acuerdo a un análisis de beneficios operativos, tácticos y estratégicos.

En conclusión, se han alcanzado todos los objetivos, resolviendo la problemática planteada al inicio del proyecto

## **XVI. BIBLIOGRAFÍA:**

### LIBROS

- Autor: Ing. Bernabeu Ricardo Dario, Nombre: “HEFESTO: Metodología para la Construcción de un Data Warehouse”; Instituto Universitario Aeronáutico; Año 2010.
- Terry A. Brase : “Precision Agriculture”, Año 2006 - Publicador: Delmar Cengage Learning.

### PAPERS

- Autor: Emiliano García y Fernando Flego, Nombre: “Agricultura de Precisión”; Universidad de Palermo, Año: 2012.
- Fabiano Baldo, Gilmar Araujo e Felipe Carlos dos Rios: “Desafios no Projeto de Data Warehouses - Um Estudo de Caso Aplicado à Agricultura de Precisão”, Universidad del Estado de Santa Catarina, Brasil, Año: 2011.
- Autor: INEI (Instituto Nacional de Estadísticas e Informática), Nombre: “Manual para la construcción de un data warehouse”, Lima, Perú, Año: 1997.
- Mondal, P., Basu, M., Bhadoria, P., Emam, A., Salih, M., Adegbite, A., Ayinde, T., Omolehin, R., and Ibrahim, U. Critical Review of Precision Agriculture Technologies and Its Scope of Adoption in India. American Journal of Experimental Agriculture, 2011.
- Saraiva, A. Um modelo de objetos para sistemas abertos de informações de campo para agricultura de precisão-MOSAICo. Ph.D. thesis, Tese (Doutorado)Escola Politécnica, Universidade de São Paulo. São Paulo, 1998.
- Kimball, R., Ross, M., and Merz, R. The data warehouse toolkit: the complete guide to dimensional modeling. Wiley, 2002.

- Inmon, W. Building the Data Warehouse. Wiley, 2005.

- HEFESTO, DATA WAREHOUSING: Sistematización de Conceptos - HEFESTO: Metodología para la construcción de un Data Warehouse.

[www.businessintelligence.info/docs/hefesto-v2.pdf](http://www.businessintelligence.info/docs/hefesto-v2.pdf)

## REVISTA

-Barbosa, R. Z., Garcia, F. M., and Giati Jr, N. O. Um Estudo de Caso de Aplicação de Técnicas de Agricultura de Precisão no Cultivo de Laranja no Município de Limeira - SP. Revista Científica Eletrônica de Agronomia, 2006.

## XVII: GLOSARIO

**DW:** En el contexto de la informática, un almacén de datos (del inglés **Data Warehouse**) es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.

**ETL:** proviene de la siglas en inglés “**Extract, Transform and Load**” cuya traducción al español significa “Extracción, Transformación y Carga”, es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizarlos, o en otro sistema operacional para apoyar un proceso de negocio.

**OLAP:** es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Es una solución utilizada en el campo de la llamada Inteligencia de negocios (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales (o Cubos OLAP) que contienen datos resumidos de grandes Bases de datos o Sistemas Transaccionales (OLTP).

Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares.

**OLTP:** es la sigla en inglés de Procesamiento de Transacciones En Línea (OnLine Transaction Processing) es un tipo de procesamiento que facilita y administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones (gestor transaccional). Los paquetes de software para OLTP se basan en la arquitectura cliente-servidor ya que suelen ser utilizados por empresas con una red informática distribuida.

**AD HOC:** es una locución latina que significa literalmente «para esto». Generalmente se refiere a una solución específicamente elaborada para un problema o fin preciso y, por tanto, no generalizable ni utilizable para otros propósitos.

Se usa para referirse a algo que es adecuado sólo para un determinado fin o en una determinada situación.

## ANEXO 1

### **1 Descripción y Posicionamiento de la Empresa**

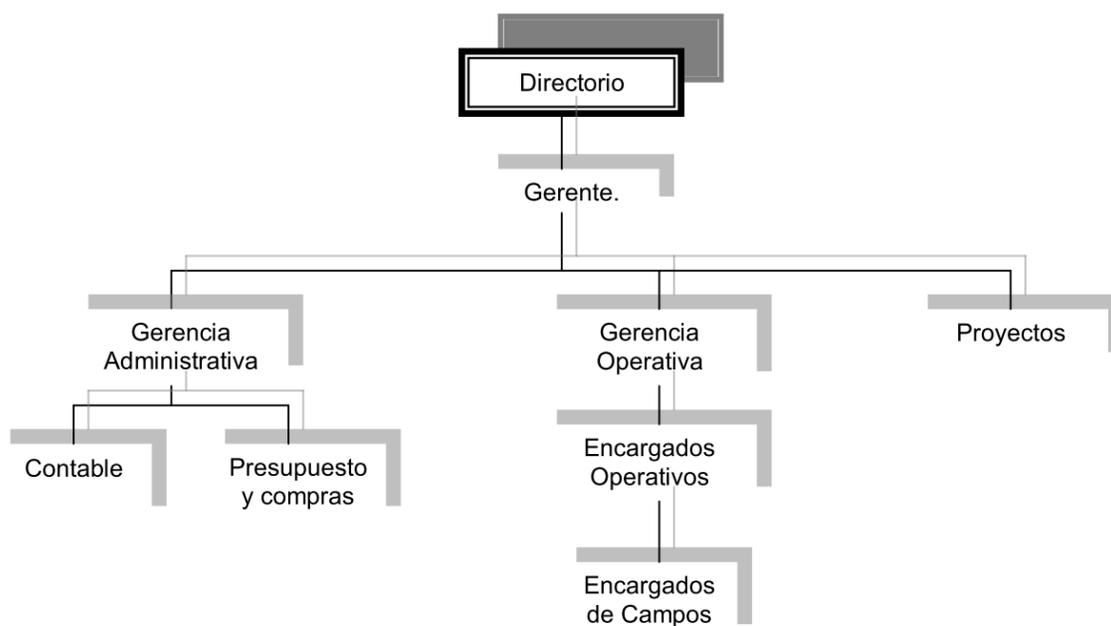
#### **1.1 Nombre de la Empresa: Campos del Sur**

#### **1.2 Descripción de la organización:**

Campos del sur es una organización con fines de lucro, ubicada en la calle Ituzaingo 87 de la ciudad de Córdoba constituida en el año 2000.

Campos del sur se desempeña en actividades de Agricultura y Agro negocios en diferentes provincias teniendo a su cargo la siembra, cosecha y comercialización de cultivos.

#### **1.3 Organigrama:**



#### 1.4 Descripción funcional del organigrama:

Encargado de campo: Los encargados de campo son caseros. Su principal función es la seguridad del campo, monitorear el stock de los insumos utilizados, lleva el registro de lluvias.

Responde al encargado operativo a través de visitas o por teléfono. Brindándole información sobre sus funciones, con ausencia de documentación registrada.

- Encargado operativo: Los encargados operativos se encargan de realizar el presupuesto de siembra donde se reflejan la cantidad de semillas que se van a utilizar por hectáreas, enfocándose en la campaña anterior y aplicando una serie de ajustes, cantidad de labranzas necesarias (si el suelo debe moverse), cantidad de fertilizantes (glifosato, fósforo), entre otros detalles, consultando con la Gerencia Operativa. También prevé la compra de insumos, controla el trabajo de los contratistas, guía y da órdenes a los encargados de campo, y realiza un seguimiento y control del cultivo.

Responde a la gerencia Operativa y a su vez ésta a la Gerencia general, sobre su actividad, a través de reuniones y por teléfono.

Se vincula con el área proyecto donde la experiencia de sus integrantes (ingenieros) permite la elaboración de proyectos más eficientes y eficaces.

- Gerencia Operativa: Gestión de los contratistas.

Realiza (“Alta, Baja y Modificación”) de algunos recursos de la organización  
Ej.: Maquinarias, Lotes, Campos, etc.

Fija el precio consultando con los encargados operativos sobre las tareas que deben realizar los contratistas.

- Área Proyectos: Investiga y lleva adelante los nuevos proyectos ya sean a corto mediano o largo plazo. Ej. (corporación de riego, la posibilidad de un nuevo cultivo, el biodiesel, etc.)
- Gerencia Administrativa: Su función es registrar la actividad diaria de la organización. Ya sea registrar presupuesto, facturas de compras, pagos, etc. También realiza la gestión de proveedores en función al presupuesto evaluado, y con previa consulta con la gerencia Operativa.
- Gerencia General: Gerente Gral. plantean un Plan de Inversión, producción y estratégico para la campaña. Además recibe del directorio la confirmación de la campaña.
- Directorio: Toma la decisión final acerca de realizar o no las inversiones. Dicha área está conformado por 5 personas.

El directorio es quien aprueba el proyecto.

### **1.5 Misión de la empresa:**

Agregar valor al producto final a partir de la cadena agroindustrial.

### **1.6 Políticas de la empresa:**

- Planificar todas las tareas a realizar dentro de la organización
- Lograr eficacia y eficiencia en las tareas a desarrollar
- Seguridad jurídica
- Protección de la ecología, principalmente del suelo
- Lograr calidad en el producto
- Responsabilidad en las relaciones comerciales

### **1.7 Objetivo y metas de la empresa típicos para un plan de siembra:**

- Alcanzar un nivel de económico, seguro, estable y sencillo posible.
- Que tienda a mejorar o por lo menos mantener la productividad del suelo.
- Que la siembra sea fácilmente controlable tanto productiva como económicamente. Partiendo de la base de las inversiones ya realizadas e intentando potenciar el rendimiento de las mismas

### **1.8 Descripción general de actividades:**

Campos del sur desarrolla su actividad de agro negocios en la ciudad de Córdoba.

Su actividad se realiza de la siguiente manera:

La base fundamental de la actividad está en la atracción de potenciales inversores interesados en la compra o alquiler de campos. A partir de allí, y en función de esos intereses, se analizan las posibilidades existentes en el mercado y se le brindan al inversor las alternativas consideradas financiera y económicamente viables.

Si el inversor llega a una decisión positiva, Campos del Sur ofrece su administración, y el inversor pasa a formar parte del Directorio. La administración consiste, a grandes rasgos, en el siguiente procedimiento:

El Gerente Gral. En conjunto con el Gerente Operativo plantean un Plan de Inversión, Producción y Estratégico para la campaña.

Una vez que el Plan es aprobado por el Directorio, se comienza con la ejecución del mismo, para lo cual entran en escena principalmente los Encargados Operativos de cada zona productiva y los Encargados de Campo.

De esta manera, se seleccionan los contratistas y los insumos para empezar a trabajar los suelos, con el criterio de menor costo en función de una buena eficiencia.

Cada insumo utilizado y cada tarea realizada por los contratistas son controlados por el encargado operativo, el cual envía un fax o mail donde posteriormente el Gerente operativo lo registrará en una planilla Excel.

El pago a contratistas y proveedores es realizado por el área contable de la gerencia administrativa.

El seguimiento de cada lote estará a cargo de los encargados operativos quienes verificarán la evolución del cultivo enviando una descripción al gerente operativo. Además se realiza un seguimiento de lluvias, humedad relativa ambiente y de suelo, temperatura, etc. y un

seguimiento de Stock de insumos llevados a cabo por los encargados de campo que informan vía telefónica al encargado operativo.

Una vez finalizada la campaña se comparan y evalúan el presupuesto real con el inicial para verificar y corregir las desviaciones y utilizar dicha información para sus campañas futuras.

El gerente general toma una estadística que se realizó en planillas Excel sobre la campaña.

Dicha estadística y un informe final son presentados al directorio.

### **1.9 Relación de las metas de la organización con las del Data warehouse**

El DWH coincide con la metas de la empresa, ya que esta necesita mejorar su eficiencia en la toma de decisiones y contar con información detallada a tal fin. Esto es vital, ya que es muy importante para procurar una mayor ventaja competitiva conocer cuáles son los factores que inciden directamente sobre su rentabilidad, como así también, analizar su relación con otros factores y sus respectivos por qué.

El DW aportará un gran valor a la empresa, entre las principales ventajas e inconvenientes que solucionará se pueden mencionar los siguientes:

- Permitirá a la gerencia tener una visión general del estado de los campos en base a los datos recolectados.
- 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 empresa.
- Aportará a la mejora continua de la estructura de la empresa.

### **1.10 Áreas involucradas**

El propósito es realizar un sistema que pueda satisfacer las necesidades del área de gerencia administrativa.