



A.F. 133014

**ESCUELA SUPERIOR POLITECNICA DEL LITORAL**

**Facultad de Ingeniería en Electricidad y Computación**

**TESIS DE GRADO**

**“Desarrollo de una Solución de Inteligencia de Negocios:  
para Empresas orientadas a la distribución y Ventas de  
Servicios en el Ecuador”**

Previa la obtención del Título de:

**Master en Sistemas de Información Gerencial**

**PRESENTADA POR:**

**Fausto Edison Correa Almazán**

**Guayaquil - Ecuador**

**Año 2010**

**ESCUELA SUPERIOR POLITECNICA DEL LITORAL**  
**Facultad de Ingeniería en Electricidad y Computación**

**TESIS DE GRADO**

**“Desarrollo de una Solución de Inteligencia de Negocios  
para Empresas orientadas a la distribución y Ventas de  
Servicios en el Ecuador”**

**Previa la obtención del Título de:**

**MASTER EN SISTEMAS DE INFORMACIÓN GERENCIAL**

**Presentada por:**

**Fausto Edison Correa Almazán**

**GUAYAQUIL – ECUADOR**

**AÑO 2010**

## AGRADECIMIENTO

A todas las personas que de uno u otro modo colaboraron en la realización de este trabajo y especialmente al Ing. Lenin Freire Director de Tesis, por su invaluable ayuda.

## DEDICATORIA

A Dios, a mi esposa e hijos que en todo momento nos brindaron su comprensión y apoyo y son motivo de esfuerzo y superación.

## TRIBUNAL DE GRADUACIÓN



---

Ing. Lenin Freire

DIRECTOR MSIG

---

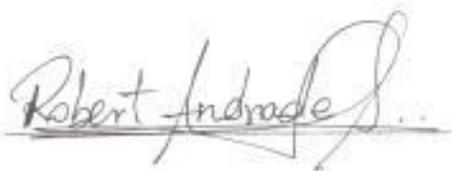
MSIG. Carlos Martín

DIRECTOR DE TESIS

---

MBA. Tania de Pareja

PRINCIPAL



---

MSIG. Robert Andrade

PRINCIPAL



---

MSIG. Fabricio Echeverria

SUPLENTE

---

MBA. Sandra Piedra

SUPLENTE

# ÍNDICE GENERAL

INTRODUCCION .....	1
CAPITULO 1 .....	3
1. La Inteligencia de negocios de los Empresarios en el Ecuador .....	3
1.1. Historia de las Pymes en el Ecuador. ....	3
1.2. Importancia de consolidar a las PYMES en los países en vías de Desarrollo .....	4
1.3. Diversas perspectivas de la inteligencia de negocios de los Empresarios de las Pymes. ....	5
CAPITULO 2 .....	9
2. Plataforma de Inteligencia de Negocios basados en herramientas de código abierto .....	9
2.1. Arquitectura de plataforma de Inteligencia de Negocios a desarrollar. ...	9
2.2. Diagrama y relaciones de los componentes de la solución. ....	11
2.3. Componentes de la Arquitectura.....	13
2.3.1. Servidor Web. ....	13
2.3.2. Motor de generación de Cubos.....	13
2.3.3. Diseñador de cubos. ....	15
2.3.4. Herramienta para Reportes.....	16
2.4. Herramientas de extracción de Datos.....	17
2.5. Herramienta de sincronización.....	19
2.6. Plantillas cuadros de mando integral.....	21
2.7. Almacén de Datos.....	22
CAPITULO 3 .....	23
3. Diseño de procesos de extracción orientados al uso de internet como medio de distribución .....	23
3.1. Diseño de almacén de datos basados en modelos de la industria de importación y distribución de productos.....	23
3.2. Diseño de estructuras especiales particulares para el cliente.....	38

3.3.	Mapeo de datos de sistema transaccional a modelos del almacén de datos.	39
3.4.	Uso de herramienta de sincronización para extracción del sistema transaccional con el almacén de datos.....	67
3.5.	Esquemas de mantenimiento y actualización del almacén de dato para los respectivos ETL.....	68
CAPITULO 4.....		70
4.	Inteligencia de Negocios sobre la plataforma creada.....	70
4.1.	Generación de modelos de datos.....	70
4.2.	Desarrollo de cuadro de mando integral.....	77
4.3.	Implementación del cuadro de mando integral.....	84
4.4.	Cubos y modelos bases para el cuadro de mando integral.....	90
4.5.	Pruebas y control de calidad.....	90
CAPITULO 5.....		97
5.	Modelo de negocio para la venta del servicio sobre la plataforma creada.	97
5.1.	Estudio de mercado.....	97
5.2.	Tácticas de mercadeo (4P).....	98
5.3.	Servicios ofrecidos por la plataforma de inteligencia de negocios.....	101
5.4.	Servicios basados en consultorias de planificación estratégica que se genere por los modelos de inteligencia de negocios creados.....	105
CONCLUSIONES Y RECOMENDACIONES.....		107
BIBLIOGRAFÍA.....		110
ANEXOS.....		111

## ÍNDICE DE FIGURAS

Ilustración 1 Entorno de desarrollo de PYMES .....	5
Ilustración 2 Suite de Pentaho.....	12
Ilustración 3 Capas del sistema OLAP Mondrian. ....	14
Ilustración 4 Editor de Cubos de Pentaho. ....	16
Ilustración 5 Editor de Reportes de Pentaho. ....	17
Ilustración 6 Arquitectura Iconnector .....	21
Ilustración 7 Capa de funciones de pre procesamiento de Datos .....	40
Ilustración 8 Diagrama de Flujo del proceso de extracción.....	45
Ilustración 9 Diagrama de extracción Tablas de Hechos .....	47
Ilustración 10 Flujo de Extracción para DataMart de Ventas.....	48
Ilustración 11 Flujo de Trabajo para DataMart de Compras.....	48
Ilustración 12 Diagrama de Flujo para DataMart de Inventarios.....	49
Ilustración 13 Diagrama de flujo para DataMart de contabilidad .....	50
Ilustración 14 Cuadro de mando integral.....	81
Ilustración 15 Pantalla de Login de Pentaho .....	102
Ilustración 16 Pantalla Front End de Pentaho .....	103

## INTRODUCCION

En la actualidad los conceptos de balance score card, indicadores, dashboard han sido certificados y forman parte de las herramientas de trabajo de las grandes empresas exitosas que le permiten hacer una análisis global de su estrategia operativa, sus costos, ingresos de tal manera que pueden tener control sobre su rentabilidad y alcanzar sus objetivos como organización.

El mercado globalizado, el desarrollo del internet ha dado a lugar que en la actualidad las empresas no solo compitan con otras empresas locales sino también con otras empresas de otros países y continentes.

La toma de decisiones cada vez se vuelve más crítica y debe ser más oportuna, no se puede dar lugar a demoras y esto se ha logrado gracias a la evolución de las herramientas de inteligencia de negocios.

Debido al efecto beneficioso que causan este tipo de herramientas a las organizaciones y la complejidad de construcción de sus modelos altamente integrados hace que consultorías, asesorías, productos y proyectos de este tipo sea sumamente costos. Por lo expuesto es que generalmente las soluciones de inteligencia de negocios solo llegan de manera exitosas a las grandes organizaciones mismas que logran sus objetivos por medio de las buenas decisiones que toman soportada en la información que estas soluciones brindan.

Sin embargo en el mundo no solo existen grandes empresas, existen medianas y pequeñas empresas que por sus limitados recursos se les hace difícil presupuestar soluciones de este tipo y se quedan con herramientas poco funcionales que no le permitan contar con la información de manera oportuna y así escalar en el mercado.

Por esta razón se propone esta solución de crear una plataforma tecnológica en la que se puedan configurar modelos de la industria y brindar servicios de inteligencia de negocios a empresas medianas y pequeñas mediante el cobro de la solución como servicio dejando en los clientes proyectos llave en mano y que permita distribuir los costos de las asesorías y desarrollos del producto en un mercado diverso de clientes, explotando de la mejor manera la inversión hecha en hardware y software para la solución brindada.

Esta solución se la creará inicialmente en Ecuador y se usará un producto de código abierto llamado "Pentaho" y se usará un ERP que sea muy común en nuestro medio y a este ERP se agregará la capa de Inteligencia de negocios, de esta forma todos los clientes que tengan este ERP podrán usar esta solución sin complicación alguna.

## 1.2. Importancia de consolidar a las PYMES en los países en vías de Desarrollo

Las PYMES deben de consolidarse para los países en via de desarrollo por las siguientes razones:

- Aporte al PIB y la capacidad de contribuir en la generación de empleos
- Confrontan mayores problemas que las empresas grandes en cuanto al acceso a mercados, financiamiento y tecnología.
- Flexibilidad o adaptabilidad, ventaja competitiva, pueden más fácilmente identificar y apropiarse de nichos de mercados.

El siguiente gráfico muestra el entorno para el desarrollo de las PYMES en Ecuador.

*EL ENTORNO PARA EL DESARROLLO DE LAS PYMES*



Ilustración 1 Entorno de desarrollo de PYMES

### 1.3. Diversas perspectivas de la inteligencia de negocios de los Empresarios de las Pymes.

En el momento de crear una empresa, los objetivos son, presuntamente, la maximización de beneficios a corto y a largo plazo. A corto plazo el problema se resuelve creando una estructura que haga coincidir el punto más bajo de la curva de costes totales medios con el precio del producto. A largo plazo la empresa deberá mostrar suficiente flexibilidad para, modificando su tamaño y estructura, desplazar su curva de costes de forma que se permanezca el mayor tiempo posible con beneficios máximos.

Sin embargo este modelo está alejado del mundo real, de las razones que declaran los empresarios utilizar en la adopción de decisiones. En la realidad la información disponible por el empresario sobre la evolución futura del mercado es tan escasa que no pueden adoptarse decisiones

maximizadoras. Los riesgos son tan altos que el objetivo principal pasa a ser la seguridad.

Un enfoque teórico alternativo consiste en contemplar a la empresa como un individuo que lucha por su supervivencia en un ambiente ecológico fluctuante. Según este enfoque, el cambio en el mundo económico tendría características similares al mundo biológico: la evolución estaría sometida a las leyes darwinistas según las cuales sobreviviría la empresa mejor adaptada al entorno. Las empresas jóvenes tratarían de emular a las empresas de éxito imitando sus características sobresalientes y contribuyendo así a "subir el listón".

Debemos por tanto considerar quién y cómo adopta las decisiones en las empresas. Está claro que las decisiones las toman los individuos que dirigen las empresas pero ¿Quiénes son estos individuos y cuáles son sus objetivos? La teoría económica tradicional considera, como hemos dicho, que las empresas tratan de maximizar sus beneficios. Pero para que esto sea indiscutible es necesario que, como ocurría antes, los propietarios de las empresas sean los mismos que los administradores.

El empresario del siglo pasado asumía tres tipos de actividades:

- el empresario emprendedor que toma la iniciativa de crear una empresa,
- el empresario capitalista que aporta los medios necesarios para

llevarla a cabo y

- el empresario administrador que gestiona la empresa que ha creado.

Aunque en la pequeña y mediana empresa ese tipo multifacético de empresario sigue existiendo, en las grandes empresas esas tres funciones las realizan individuos que muy frecuentemente no se conocen entre sí. Las iniciativas emprendedoras son estudiadas y analizadas por las divisiones de planificación y marketing, los propietarios de las acciones pueden incluso ignorar cuál es la actividad principal de la empresa, y los administradores pueden haber sido contratados simplemente por tener una brillante trayectoria profesional en otras empresas.

Cada vez es más frecuente, y sobre todo en las grandes empresas, que los propietarios sean un gran número de pequeños accionistas sin ninguna capacidad de decisión. Incluso los consejos de administración, formados por los accionistas más importantes aunque minoritarios, cederán la gestión cotidiana a consejeros delegados, gerentes y a técnicos con un título de Master en Dirección de Empresas.

En los últimos tiempos han salido a la luz pública las diferencias de opinión existentes entre los propietarios y los gerentes de algunas grandes empresas. De hecho es muy posible que el gerente esté más interesado en conseguir poder o prestigio que beneficios. El poder y el prestigio pueden conseguirse haciendo que la empresa aumente de tamaño, gane cuotas de

mercado, aunque para ello se deban sacrificar los dividendos que se reparten.

Se supone que los accionistas o el consejo de administración pueden cambiar en cualquier momento al gerente que no trabaje por sus intereses, y así sucede en ocasiones como sabemos. Incluso una empresa que no esté siendo bien administrada puede ser objeto de absorción por otra. Hay de hecho en la actualidad empresas especializadas en apoderarse de otras mal administradas, cambiar a los gerentes, reflotarlas y revenderlas con impresionantes beneficios. Pero la influencia de los intereses de los gerentes en la adopción de decisiones en la empresa es un hecho que debe ser tenido en cuenta.

## CAPITULO 2

### 2. Plataforma de Inteligencia de Negocios basados en herramientas de código abierto

#### 2.1. Arquitectura de plataforma de Inteligencia de Negocios a desarrollar.

La plataforma Open Source Pentaho Business Intelligence cubre muy amplias necesidades de Análisis de los Datos y de los Informes empresariales. Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Eso hace que Pentaho es una solución muy flexible para cubrir una amplia gama de necesidades empresariales – tanto las típicas como las sofisticadas y específicas al negocio.

Los módulos de la plataforma Pentaho BI son:

**Reporting.** Un módulo de los informes ofrece la solución adecuada a las

necesidades de los usuarios. Pentaho Reporting es una solución basada en el proyecto JFreeReport y permite generar informes ágiles y de gran capacidad. Pentaho Reporting permite la distribución de los resultados del análisis en múltiples formatos - todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto. Los reportes Pentaho permiten también programación de tareas y ejecución automática de informes con una determinada periodicidad.

**Análisis.** Pentaho Análisis suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (pivot tables, crosstabs), generadas por Mondrian y JPivot, el usuario puede navegar por los datos, ajustando la visión de los datos, los filtros de visualización, añadiendo o quitando los campos de agregación. Los datos pueden ser representados en una forma de SVG o Flash, los dashboards widgets, o también integrados con los sistemas de minería de datos y los portales web (portlets). Además, con el Microsoft Excel Analysis Services, se puede analizar los datos dinámicos en Microsoft Excel (usando la conexión a OLAP server Mondrian).

**Dashboards.** Todos los componentes del módulo Pentaho Reporting y Pentaho Análisis pueden formar parte de un Dashboard. En Pentaho Dashboards es muy fácil incorporar una gran variedad en tipos de gráficos, tablas y velocímetros (dashboard widgets) e integrarlos con los Portlets JSP, en donde podrá visualizar informes, gráficos y análisis OLAP.

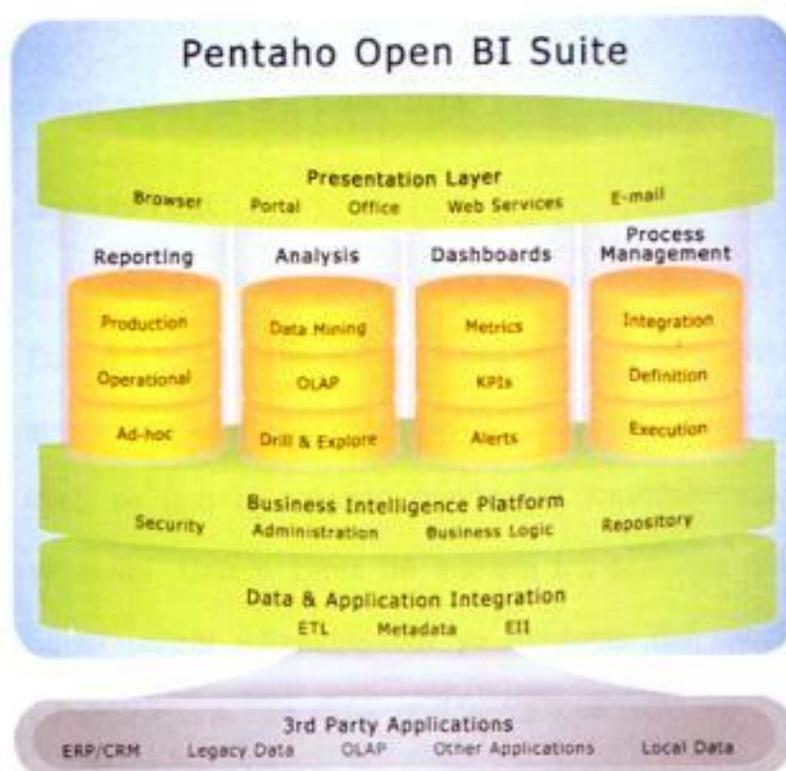
**Data Mining.** Análisis en Pentaho se realiza con una herramienta WeKa.

Integración de Datos - se realiza con una herramienta Kettle ETL (Pentaho

Data Integration) que permite implementar los procesos ETL. Últimamente Pentaho lanzó una nueva versión - PDI 4.0 – que marcó un gran paso adelante en OSBI ETL y que hizo Pentaho Data Integration una alternativa interesante para las herramientas comerciales.

## **2.2. Diagrama y relaciones de los componentes de la solución.**

Pentaho Corporation es el patrocinador principal y líder del proyecto Pentaho BI. El proyecto Pentaho BI es una iniciativa en curso de la comunidad open source que provee a las organizaciones de las mejores soluciones de su clase para sus necesidades de inteligencia de negocios. Al aprovechar la riqueza de las tecnologías de código abierto y las contribuciones de la comunidad de desarrollo de código abierto, Pentaho es capaz de innovar mucho más rápido que los proveedores comerciales. Como resultado, Pentaho ofrece una alternativa de código abierto que supera a las soluciones de Business Intelligence propietarias en muchas áreas como arquitectura, soporte de estándares, funcionalidad y simplicidad de implantación. En otras palabras, no se espera que la gente la adopte sólo porque es de código abierto, se espera que la gente la escoja porque es superior.



**Ilustración 2 Suite de Pentaho**

La suite de Pentaho contiene elementos y herramientas que facilitan lo siguiente:

- Extracción de datos para mantenimiento y actualización del almacén de datos.
- Diseño y modelamiento de cubos de información
- Configuración de fuentes de datos
- Administrador de usuarios
- Interfaz del cliente para manejo y acceso a los cubos de información
- Componentes y framework para diseño de dashboards

Permitiendo de esta forma contar con una solución completa para el

desarrollo de Proyectos y soluciones de inteligencia de negocios.

## **2.3. Componentes de la Arquitectura.**

### **2.3.1. Servidor Web.**

La suite de Pentaho funciona sobre un servidor web basado en Tomcat5.5, funciona sobre clases de Java que permiten vía web una aplicación distribuible y de fácil uso y acceso. La ventaja de este servidor Web es que la solución puede ser implementada sobre servidores Windows, Unix ó Linux de manera transparente dando la ventaja de poder utilizar infraestructura ya existente del cliente o sobre nueva infraestructura con bajo costos ya que solo se necesita adquirir el hardware y la asesoría por la instalación del Linux y MySQL que es una base opensource recomendada por Pentaho.

### **2.3.2. Motor de generación de Cubos.**

El Motor de generación de Cubos para la suite de Pentaho es Mondrian, es un Servidor OLAP escrito en Java. Permite analizar interactivamente grandes cantidades de datos almacenados en bases de datos de diversos orígenes sin necesidad de conocer de comandos de SQL.

Este sistema OLAP utiliza sentencias MDX para hacer consultas a los cubos de información.

MDX significa "multi-dimensional expressions". Es el lenguaje de consulta

principal en práctica de Mondrian. MDX fue introducido por Microsoft con Microsoft SQL Server Servicios OLAP en alrededor de 1998, como el componente lingüístico de la OLE DB para OLAP API.

La Arquitectura del sistema OLAP Mondrian consiste en 4 capas:

1. Capa de presentación
2. Capa dimensional
3. Capa estrella
4. Capa de almacenamiento

Estas capas se detallan en la figura #3.

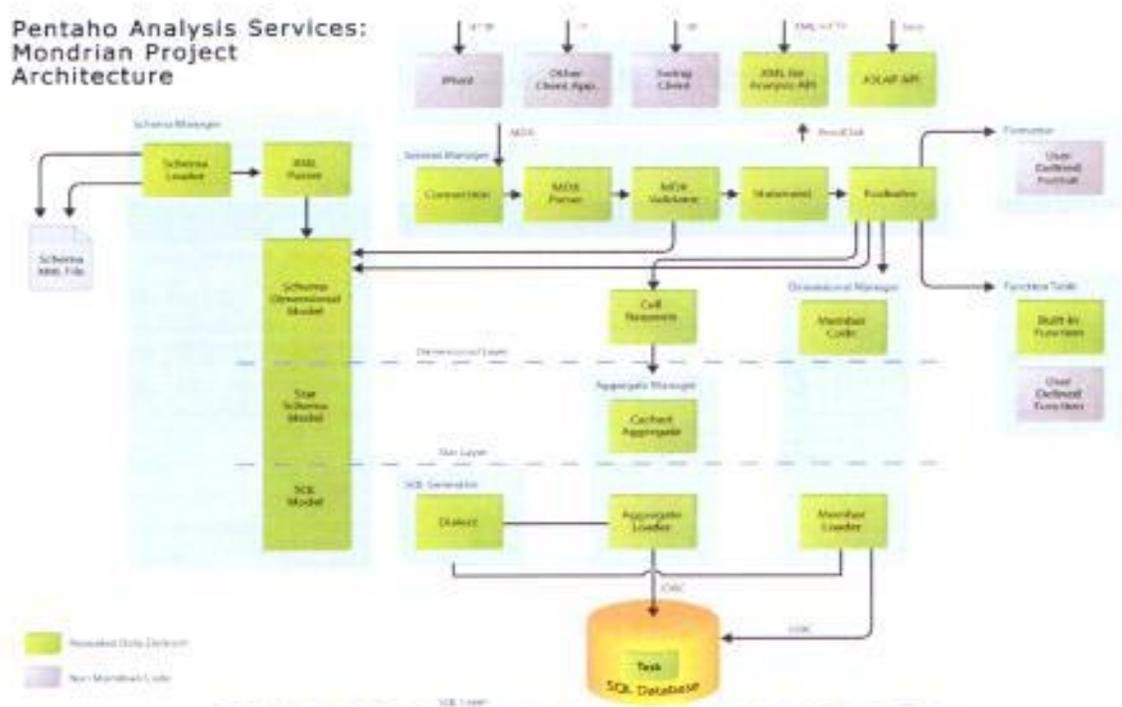
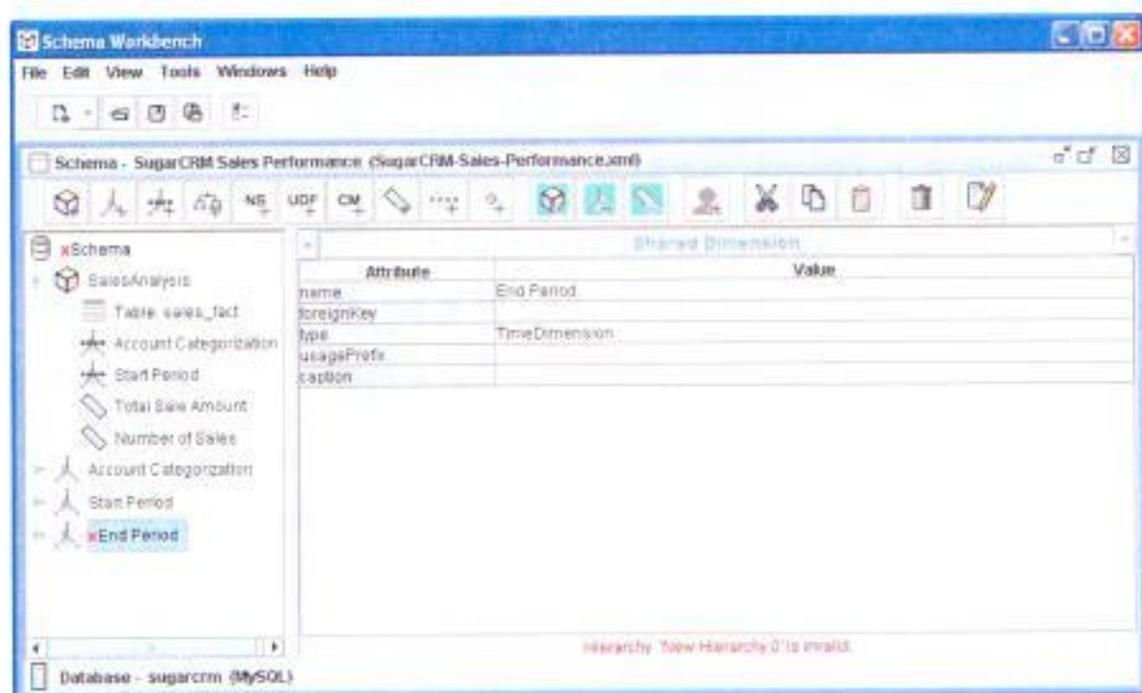


Ilustración 3 Capas del sistema OLAP Mondrian.

### 2.3.3. Diseñador de cubos.

El esquema Mondrian Workbench es una interfaz de diseño que le permite crear y probar esquemas cubo OLAP Mondrian visualmente. El motor de Mondrian procesa las solicitudes de MDX con la ROLAP (OLAP Relacional) esquemas. Estos archivos son los modelos de esquemas XML de metadatos que se crean en una estructura específica utilizada por el motor de Mondrian. Estos modelos XML se puede considerar las estructuras de forma de cubo que utilizan HECHO existentes y tablas de dimensiones que se encuentran en su RDBMS. No requiere que un cubo de física real se construye o mantiene, sólo que el modelo de metadatos se crea.

La siguiente figura #4 muestra la pantalla de la herramienta de diseño de cubos de información sobre Pentaho.



#### Ilustración 4 Editor de Cubos de Pentaho.

Ofrece las siguientes funcionalidades:

1. Editor de esquema integrado con el origen de datos subyacente para su validación. (Véase más arriba)
2. Test consultas MDX contra de esquema y base de datos
3. Examinar las bases de datos estructura subyacente

Consulte la guía de Mondrian técnica para comprender los esquemas.

Una vez que usted tiene el archivo de esquema, puede subir esto en su aplicación.

#### 2.3.4. Herramienta para Reportes.

El Diseñador de Pentaho Report (PRD) es una aplicación de escritorio que proporciona un entorno de diseño visual para crear definiciones de informe. Las definiciones de informe se puede ejecutar y guardar localmente a través del PRD o publicado en un servidor Pentaho BI para permitir a muchas personas a la ejecución de informes de acceso y horario. PRD está dirigido a usuarios experimentados y el poder, que están familiarizados con los conceptos y las fuentes de datos utilizadas, ofrece las siguientes funcionalidades:

1. Un ambiente rico e intuitivo diseño visual
2. componentes reutilizables informe que incluya las fuentes de datos, fórmulas y resúmenes

3. Desarrollo rápido asistente de informes basados en
4. El acceso a cualquier fuente de datos, incluyendo JDBC / ODBC, Web Services, XML, archivos planos, objetos Java y muchos más
5. Soporte para múltiples fuentes de datos en el mismo informe
6. Internacionalización y soporte de localización
7. Múltiples informes en un único modelo
8. El desarrollo ágil está habilitada de diseño iterativo utilizando el Diseñador de JReport y Studio JReport en JReport servidor

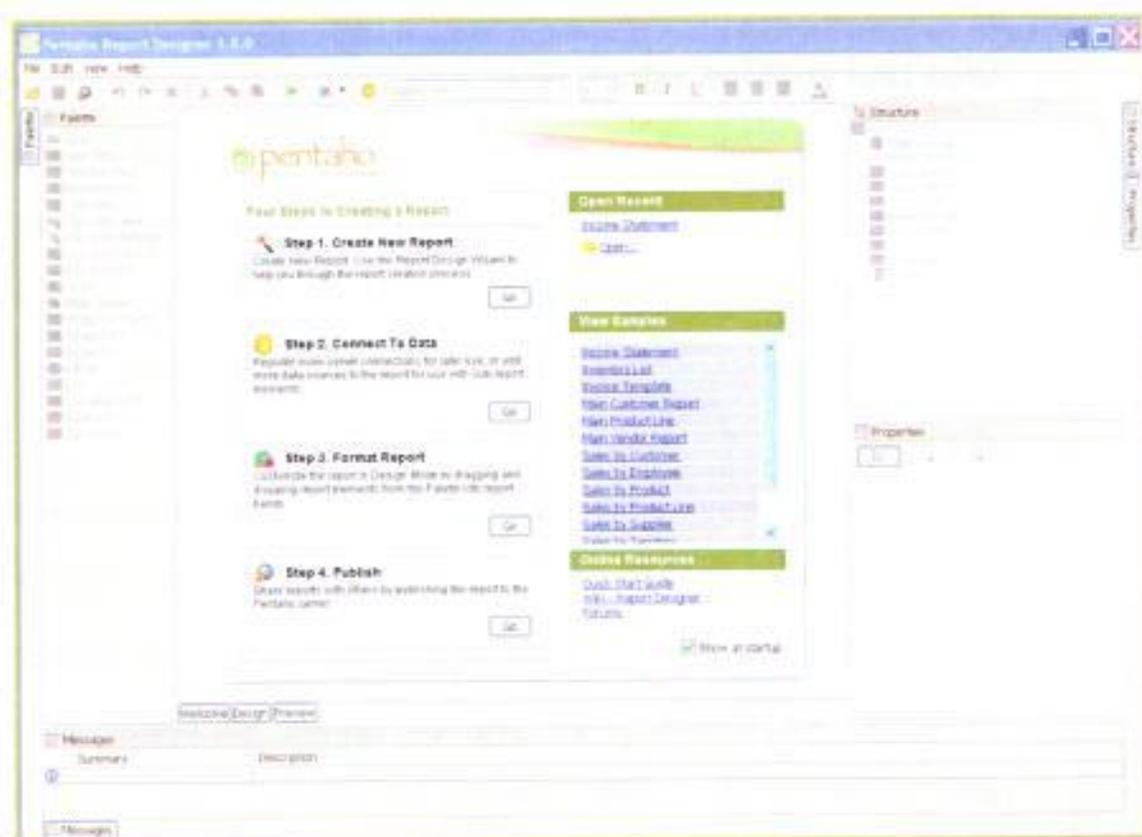


Ilustración 5 Editor de Reportes de Pentaho.

## 2.4. Herramientas de extracción de Datos.

Para la creación del almacén de datos es necesario contar con una

herramienta que permita conectarse a varias fuentes de datos, tales como bases de datos comerciales, de código abierto incluso archivos textos, Excel entre otros.

En el mercado existen muchas herramientas de extracción licenciadas tales como Microsoft SQL Server y de código abierto tal como la herramienta PDI de Pentaho que nos enfocaremos en esta tesis.

Una de las ventajas que tiene esta herramienta es que trae ya en su core componentes de conectividad para varias fuentes de datos tales como, AS400, Oracle, MySql, MS Sql Server, ODBC, entre otros, al ser de código abierto y basarse en Java tiene muchas herramientas de conectividad sin costo en el mercado, esto facilita el poder diseñar procesos de extracción sin invertir en licenciamiento.

Otra ventaja que ofrece esta herramienta es que permite diseñar workflows para definir el proceso de extracción, permitiendo poner lógica a la extracción que permita diseñar el almacén de datos de manera óptima y precisa, incluso se pueden agregar lógica definida en JavaScripts, conectarse vía FTP o HTTP para obtener o subir archivos y definir alertas vía correo electrónico para poder hacer una buena administración y control del proceso de extracción.

Una limitante de esta herramienta es que está orientada a ejecutarse vía

LAN ya que necesita acceder a puertos de las bases de datos que generalmente no se abren, para ello es necesario agregar una capa adicional para poder utilizar un esquema de actualización de datos utilizando como medio a internet, para ello se ha buscado una herramienta del mercado que consiste en un cliente de sincronización y un servidor con la particularidad de que esta herramienta comprime los paquetes de datos y hace una optimización de 10 a 1 del canal.

## 2.5. Herramienta de sincronización.

Como se mencionó anteriormente para hacer una extracción de datos vía internet se tienen que superar los siguientes inconvenientes:

1. Acceder a puertos de las bases de datos que generalmente no se abren en internet.
2. Optimizar el ancho del canal, ya que cuando se accede directamente a una base de datos se requiere de un buen ancho de banda para transferir los datos.

Estas dos limitantes pueden ser mitigadas utilizando una herramienta que no necesite los puertos de las bases de datos directamente abiertos en internet y que comprima los paquetes de datos en una buena relación de compresión.

La herramienta que suple esta necesidad es un producto ecuatoriano llamado Iconnector, este producto consiste en software cliente servidor, en donde el cliente le envía o le solicita datos al servidor de los registros de las tablas que han cambiado o que se necesiten actualizar.

El servidor y el cliente se comunican por un puerto específico configurable y cada uno por su lado se conecta a la fuente de información, de esta forma cada uno es capaz de responder a preguntas de sincronización de datos que cada uno necesite ya que esta herramienta maneja la sincronización en dos vías.

Para nuestro caso de querer utilizarlo como herramienta de ETL la sincronización se hará en una sola vía, desde las fuentes transaccionales hasta el almacén de datos, y aprovecharemos la compresión de datos ya que la estructura del almacén de datos son muy diferentes a las de las fuentes transaccionales por lo que no puede ser una sincronización sino un proceso de extracción y transformación.

La figura #4 muestra la arquitectura para este esquema de extracción y transformación utilizando Iconnector.

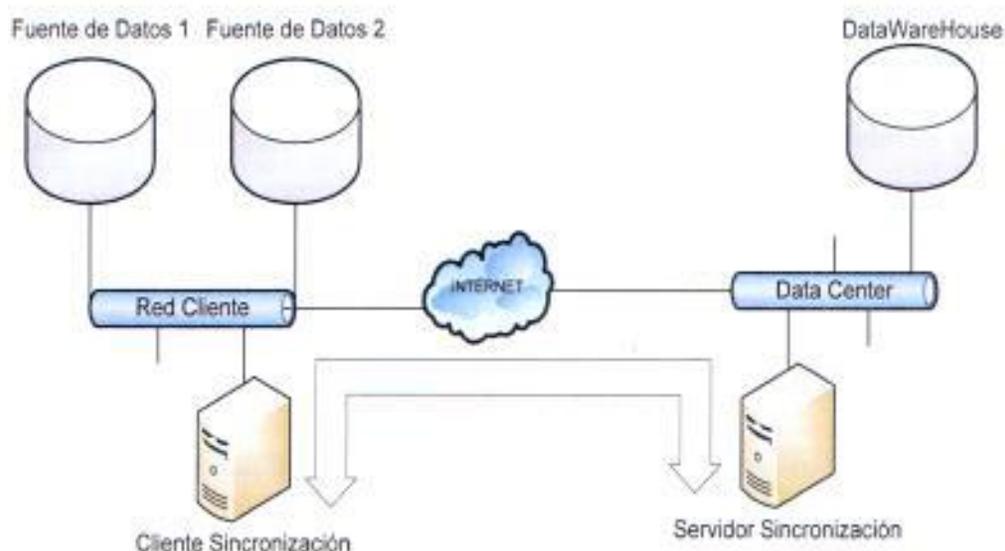


Ilustración 6 Arquitectura Iconnector

## 2.6. Plantillas cuadros de mando integral.

Como se conoce teóricamente el cuadro de mando integral consiste en 4 perspectivas que dependiendo del negocio pueden ser más y estas perspectivas son:

1. Cliente
2. Financiera
3. Interna
4. De Aprendizaje

Como alcance de este producto se ofrecerán dashboards con indicadores para poder enfocarse en estas perspectivas. Con esta línea de base cada negocio podrá extender estos dashboards de acuerdo a su necesidad.

Este DataWareHouse es una solución orientada a una empresa que distribuye productos para insumos agrícolas y aerofumigación y en base a este negocio se han diseñado los cubos, dashboards e indicadores que se detallaran posteriormente.

## 2.7. Almacén de Datos.

El almacén de datos será construido sobre un motor de base de datos de SQL Server 2005, y las fuentes serán también sobre SQL Server 2005 ya que el sistema ERP del cliente está sobre este motor y disponían de licencias.

El almacén de datos se lo diseñó utilizando la técnica del "Modelo Punto" técnica de Chris Todman<sup>1</sup>. Esta técnica ayudó en los talleres con los usuarios para definir la necesidad de información que tenían cada uno de ellos representantes de las áreas comerciales, financieras y logística.

Se usó la plantilla definida en el anexo A, "formato para levantar medidas y dimensiones"

De estos talleres salieron todas las medidas y dimensiones necesitadas en cada modelo de información de este DataWareHouse, levantándose requerimientos particulares en esta solución para el manejo de Fechas que para este negocio no son mensuales sino por periodos definidos por ellos ya que ellos van de acuerdo a los ciclos de producción bananera.

## CAPITULO 3

### 3. Diseño de procesos de extracción orientados al uso de internet como medio de distribución

#### 3.1. Diseño de almacén de datos basados en modelos de la industria de importación y distribución de productos.

El DataWareHouse desarrollado se lo construyó basado en la estrategia de construir pequeños datamarts enfocados a un solo datawarehouse, de esta forma se tomaron módulos a módulos y definiendo por cada módulo en talleres con los usuarios sus necesidades de información.

Del taller con usuarios comercial se definieron las siguientes dimensiones y medidas:

CUBO DE VENTAS	
Dimensiones	Definición
Anio-Periodos-Semana-Dia	Tiempo en Año, periodo, semanas y día
Anio Mes Dia	Tiempo en Año, mes y día
AsistenciaTecnica	Aplica Asistencia Tecnica

Aviones	Nombre de los aviones
Categorias Operativas	Categorias Operativas de los productos
ClienteTipoVenta	Almacenista, cliente final
Cientes	Nombre de los Clientes
Empresas	Unidades de Negocio Aifa o Agritop
EstadoMayorizado	Tiempo de la orden en Anio, mes y dia
Fecha Egreso Anio Mes Dia	Tiempo en Año, mes y dia
Fecha Orden Anio Mes Dia	Tiempo en Año, mes y dia
GrupoClientes	Categoización de los Clientes
Haciendas	Nombre de las Haciendas
LineaNegocio	Categorizacion entre nutrición, especialidades y plagas, enfermedades
Motivos contables	Motivos Contables de la operativa del negocio
Pais-Region-Provincia	Por Geografía
Por Zonas	Zonas de venta (El oro, los rios, guayas)
Proveedores	Nombre de los Proveedores
Tipo Producto	Categorizacion ente producto y servicio
TipoCliente	Categorizacion ente Independiente y Ubesa
TipoCultivos	Categorizacion de los productos por cultivo
TipoDocumento	Devoluciones,Factura,NCreditFac, NCreditoMan
TipoFumigacion	Ciclo Corto, Folear,Plaga,Sigatoka

TipoSIngresos	Categorizacion entre Ingreso variable y otros ingresos
TipoVenta	Categorizacion entre otras ventas y ventas operaciones
Vendedores	Nombre de los Vendedores
<b>Medidas</b>	<b>Definición</b>
ValorBruto	Valor de la venta sin descuentos
NC	Valor de las notas de creditos
Devoluciones	Valor de las devoluciones de ventas
Descuento	Valor de los descuentos de ventas
ValorNeto	Valor neto de la venta sin devoluciones, ni descuentos, ni notas de credito
Cantidad	Cantidad de la venta
CantDev	Cantidad de las devoluciones
Hectareas	Hectáreas fumigadas
DiasFact	Dias en facturarse una orden
Margen Bruto	Margen de la venta entre la venta bruta y el costo
Margen Neto	Margen de la venta entre la venta neta y el corto
MC Bruto	Porcentaje del margen Bruto
MC Neto	Porcentaje del margen neto

Tabla #1 Medidas y dimensiones para modelo de Ventas

Para el modelo de compras se definieron las siguientes dimensiones y

medidas:

<b>CUBO DE COMPRAS</b>	
<b>Dimensiones</b>	<b>Definición</b>
Por Periodo de Compra	Tiempo en Periodos por Fecha de Compra
Pais-Region-Provincia	Categorización Geografica por Pistas
Bodega	Nombre de las Bodegas
Categorías Operativas	Categorías Operativas de los productos
Empresa	Unidades de Negocio Aifa o Agritop
Estados	Facturados, Ordenados, por facturar, por ordenar
Inventariable	Sí o no
Periodo Entrega	Tiempo en Periodos por Fecha de Entrega
Zonas	Zonas de Compra (El oro, los rios, guayas)
Por Zonas Orden	Zonas de Orden de Compra (El oro, los rios, guayas)
Por Zonas Req	Zonas de Requerimiento de la Compra (El oro, los rios, guayas)
Por Periodo de Req.	Tiempo en Periodos por Fecha de Requerimiento
Proveedor	Nombre del Proveedor
Tipo Compra	Categorización Exterior o Local
Tipo Producto	Categorización ente producto y servicio
Tipo Documento	Devoluciones,Factura,NCreditFac, NCreditoMan
<b>Medidas</b>	<b>Definición</b>
Monto Neto de Compra	Valor de la compra

Cantidad	Cantidad de items de la compra
Valor Bruto	Valor Bruto de la compra
Descuento	Descuento ofrecido por el proveedor
Valor IVA	Impuesto IVA
Valor ICE	Impuesto ICE
Costo Unitario	Costo unitario del producto
CostoLiqImp	Costo liquido de la importacion
CostoPesoLiqImp	Costo por peso de la importación
CostoUnitarioCompra	Costo unitario del producto en la compra
CostoUnitarioLiqImp	Costo unitario liquidado en la importación
Cargos	Cargos de importación
PesoLiqImp	Peso liquidado en la importación
TotalProrratedoLiqImp	Valor prorratedo de la liquidación de la importación
DiasEntrega	Dias de entrega de la mercadería
Bonificaciones	Bonificaciones ofrecidas por el proveedor
StockTransito	Inventario en tránsito

Tabla #2 Medidas y dimensiones para el modelo de Compras

Para el modelo de estado de resultados se definieron las siguientes dimensiones y medidas:

<b>CUBO DE ESTADO DE RESULTADOS</b>	
<b>Dimensiones</b>	<b>Definición</b>
Anio Periodo Semana Dia	Tiempo en Año, periodo, semanas y dia
Plan de Cuentas	Plan financiero de la empresa (Ingresos y egresos)
Empresas	Unidades de Negocio Aifa o Agritop
Zonas	Zonas de venta (El oro, los rios, guayas)
S1	Segmento contable S1
S2	Segmento contable S2
S3	Segmento contable S3
S4	Segmento contable S4
S5	Segmento contable S5
S6	Segmento contable S6
S7	Segmento contable S7
S8	Segmento contable S8
S9	Segmento contable S9
S10	Segmento contable S10
<b>Medidas</b>	
Credito	Valor del crédito
Debito	Valor del débito
Saldo	Saldo
SaldoToParent	Porcentaje con respecto a la agrupación

Tabla #3 Medidas y dimensiones para el modelo de Estados de Resultados

La siguiente tabla detalla las medidas y dimensiones para el modelo de saldos de inventarios:

<b>CUBO DE INVENTARIO SALDOS</b>	
<b>Dimensiones</b>	<b>Definición</b>
Periodos	Tiempo en Periodos
Categorias Contables	Categorias Contables de los productos
Bodega	Nombre de las Bodegas
Empresas	Unidades de Negocio Aifa o Agritop
Lineas de Productos	Categorización Por linea de los productos
Pais-Region-Provincia	Categorización Geografica por Pistas
Por Zonas	Zonas de venta (El oro, los rios, guayas)
Proveedor	Nombre de los Proveedores
<b>Medidas</b>	<b>Definición</b>
Existencia	Existencia de inventarios
Monto	Monto del inventario existente
CantidadIngreso	Canitdad de ingreso en el periodo
CantidadIngresoAcum	Cantidad de ingreso acumulado
CantidadEgreso	Cantidad de egreso en el periodo
CantidadEgresoAcum	Cantidad de Egreso acumulado
MontoEgreso	Valor del egreso en el periodo
MontoEgresoAcum	Valor del egreso acumulado

MontoIngreso	Monto del ingreso en el periodo
MontoIngresoAcum	Monto del ingreso acumulado
CostoVenta	Costo de la venta en el periodo
CostoVentaAcum	Costo de la venta acumulado
CostoVentaEgresos	Costo de la venta egresada del inventario
CostoVentaIngresos	Costo de la devolución de ventas

Tabla #4 Medidas y dimensiones para el modelo de SalDOS de Inventarios  
 Para el modelo de cobranzas se definieron las siguientes dimensiones y medidas:

<b>CUBO DE COBRANZA</b>	
<b>Dimensiones</b>	<b>Definición</b>
Por Zonas	Zonas de venta (El oro, los rios, guayas)
Cientes	Nombre de los Clientes
FormasPago	Nombre de las formas de Pago (Autoconsumo, cheque, efectivo, incobrables)
AnPeSeDi Efec.	Tiempo en Año, periodo, semanas y dia por Fecha de Efectivizacion
ApICobro	Aplica Cobro si o No
Empresas	Unidades de Negocio Aifa o Agritop
EstadoEfectivizado	Si esta o no efectivizado el pago
F.Emision. An-Pe-Se-Di	Tiempo en Año, periodo, semanas y dia por Fecha de Emision

F.Venci	Tiempo en Año, periodo, semanas y día por Fecha de Vencimiento
GrupoClientes	Categorización para clientes
TipoCliente	Agrupacion de clientes
Rangos	Por vencer , vence hoy o vencido
<b>Medidas</b>	<b>Definición</b>
Monto	Valor de la venta
Otros Montos	Otros recargos
Iva	Impuesto iva
Cartera	Saldo de la cartera
Cartera Aplic.	Cartera cobrada
Total Efectivizado	Valor efectivizado de la cobranza
No Efectivizado	Valor no efectivizado de la cobranza

Tabla #5 Medidas y dimensiones para el modelo de Cobranzas

Para el modelo de cuentas por cobrar se definieron las siguientes dimensiones y medidas:

<b>CUBO DE CUENTAS POR COBRAR</b>	
<b>Dimensiones</b>	<b>Definición</b>
F. Emision An-Pe-Se-Di	Tiempo en Año, periodo por Fecha de Emision
F. Venci An-Pe-Se-Di	Tiempo en Año, periodo por Fecha de Vencimiento
Pais-Region-Provincia	Categorización Geografica por Pistas
Por Zonas Reg	Zonas de Registro (El oro, los rios, guayas)

Rangos	Categorizacion de los montos Por Vencer, Vence Hoy y Vencido
Tipo Documento	Devoluciones,Factura,NCreditFac, NCreditoMan
Tipo Cultivo	Categorizacion de los productos por cultivo
Vendedores	Nombre de los Vendedores
Por Zonas	Zonas (El oro, los rios, guayas)
Tipo Cliente	Categorizacion ente Independiente y Ubesa
Apl Cobro	Aplica si o no
Clientes	Nombre de los Clientes
Documentos	Id de los documentos
Grupo Clientes	Jerarquia de los Clientes
<b>Medidas</b>	<b>Definición</b>
No. Efectivo	Valor no efectivizado
Monto	Monto original de la deuda
Cartera	Valor de la cartera
Cartera Apl.	Valor aplicado a la cartera
Dias Vencidos	Dias vencidos
Dias Credito Cliente	Dias de credito del cliente
Efectivizado	Valor efectivizado del pago
Otros Pagos	Otros pagos
No Efectivizado	Valor no efectivizado
Otros Montos	Otros Montos
Iva	Valor Impuesto

% Cartera	Porcentaje de la cartera con respecto a la venta
-----------	--

Tabla #6 Medidas y dimensiones para el modelo de Cuentas por cobrar

Para el modelo de cuentas por Pagar se definieron las siguientes dimensiones y medidas:

<b>CUBO DE CUENTAS POR PAGAR</b>	
<b>Dimensiones</b>	<b>Definición</b>
Fecha Final (An-Pr)	Tiempo en Año, periodo por Fecha Final
Fecha Pago(An-Pr)	Tiempo en Año, periodo por Fecha de Pago
Fecha Solicitud (An-Pr)	Tiempo en Año, periodo por Fecha de Solicitud
Fecha Emision (An-Pr)	Tiempo en Año, periodo por Fecha de Emision
Num_Documento	Numero del documento
Pais-Region-Provincia	Categorización Geografica por Pistas
Por Zonas	Zonas (El oro, los rios, guayas)
Rangos	Categorizacion de los montos Por Vencer, Vence Hoy y Vencido
Tipo_Documento	Devoluciones,Factura,NCreditFac, NCreditoMan
Tipo_Participante	
Empresas	Unidades de Negocio Aifa o Agritop
Beneficiarios	Nombre del Beneficiario
<b>Medidas</b>	<b>Definición</b>
Saldo	Saldo de la deuda
Monto Cancelado	Monto cancelado

Monto Total	Monto total de la deuda
Descuentos	Descuentos otorgados por proveedor
Anticipos	Anticipos de los pagos
Retenciones	Valor en retenciones
Notas Credito	Valor en notas de credito
Valor Pagar	Valor a pagar
Dias Credito	Dias credito
Dias Atraso	Dias de atraso
Pago Caja	Pago realizado en caja

Tabla #7 Medidas y dimensiones para el modelo de Cuentas por Pagar

Para el modelo de cuentas por Pagar se definieron las siguientes dimensiones y medidas:

<b>CUBO DE PAGOS PROVEEDORES</b>	
<b>Dimensiones</b>	<b>Definición</b>
Anio Per. Emi.	Tiempo en Año, periodo por Fecha de Emisión
Anio Per. Pago	Tiempo en Año, periodo por Fecha de Pago
Anio. Per Venc	Tiempo en Año, periodo por Fecha de Vencimiento
Anio Per. Final Pago	Tiempo en Año, periodo por Fecha Final
Bancos	Nombre de los Bancos
Forma de Pago	Nombre de las formas de Pago (Autoconsumo, cheque, efectivo, incobrables)
Pais-Region-Provincia	Categorización Geográfica por Pistas

Por Zonas	Zonas (El oro, los ríos, guayas)
Proveedor	Nombre de los Proveedores
Tipo Participante	
Empresas	Unidades de Negocio Aifa o Agritop
<b>Medidas</b>	<b>Definición</b>
Monto	Monto del pago
Días Credito	Días credito
Días Atraso	Días Atraso

Tabla #8 Medidas y dimensiones para el modelo de Pagos Proveedores

Para el modelo de Gastos se definieron las siguientes dimensiones y medidas:

<b>CUBO DE GASTOS</b>	
<b>Dimensiones</b>	<b>Definición</b>
Anio-Periodos-Semana-Dia	Tiempo en Periodos por Fecha de gasto
Empresas	Unidades de Negocio Aifa o Agritop
Por Zonas	Zonas (El oro, los ríos, guayas)
Departamento	Departamento del gasto
Beneficiario	Beneficiario del Gasto
Concepto	Concepto del Gasto
Cargo	Cargo del Beneficiario
Cajas	Caja que emitió el gasto

LineaNegocio	Lineas de negocios
Responsable	Custodio
TipoBeneficiario	Clasificacion para el beneficiario
TipoCaja	Clasificacion para la caja
TipoCultivos	Tipos de cultivos
TipoGasto	Tipo de gasto
TipoMovimiento	Tipo de movimiento
TipoProceso	Proceso que origino el gasto
<b>Medidas</b>	<b>Definición</b>
Valorgasto	Valor del Gasto
ValorDeducible	Valor deducible del gasto
ValorNoDeducible	Valor no deducible del gasto

Tabla #9 Medidas y dimensiones para el modelo de Gastos

Para el modelo de Saldos de Tesoreria se definieron las siguientes dimensiones y medidas:

<b>CUBO DE SALDOS DE TESORERIA</b>	
<b>Dimensiones</b>	<b>Definición</b>
Empresas	Unidades de Negocio Aifa o Agritop
Por Zonas	Zonas (El oro, los rios, guayas)
FormasPago	Formas del pago cheque, tarjeta o efectivo
CuentaBco	Cuenta bancaria

Periodos	Tiempo en Periodos hasta semana
Bancos	Banco de la cuenta del saldo
TipoCajaBanco	Si es caja o banco
<b>Medidas</b>	<b>Definición</b>
Saldo	Saldo de efectivo
MontoEgreso	Egreso de Efectivo
MontoIngreso	Ingreso de Efectivo

Tabla #10 Medidas y dimensiones para el modelo de saldos de Tesorería

Mediante Talleres y sesiones con usuarios del negocio y de tecnología se pudo hacer una lluvia de ideas que fue provechoso para determinar que perspectiva de información tienen los usuarios y con tecnología para mapear que datos era posible obtener de sus sistemas transaccionales.

Hubo muchos datos que no fueron posibles obtener ya que por parte del sistema transaccional no se estaban capturando y el tener esto implicaba llevar implementaciones a nivel de procesos y desarrollo de software quedando pendiente para una segunda versión.

En la siguiente sección se hará un análisis detallado de los datos que soportan este DataWareHouse que se lo empezó desarrollando DataMarts que se enfocaron a un DataWareHouse integral, esta técnica ayudó a

desarrollar un DataWareHouse integral sin necesidad de enfocarse en tener todo en una primera versión y evitar crecer de manera desordenada en el diseño, creación e implementación de los DataMarts.

## **3.2. Diseño de estructuras especiales particulares para el cliente.**

Para el diseño de este DataWareHouse fueron necesarias las siguientes estructuras:

1. Dimensión tiempo: Para esta dimensión fue necesario considerar el hecho de que sus cortes de transacciones no son mensuales sino por periodos de acuerdo al ciclo de la producción bananera, esto hace que los cortes no sean mensual sino dentro de un periodo determinado, dando una jerarquía que tiene 13 periodos al año.
2. Saldo de inventarios: Para esta transacción fue necesario considerar que el valor de saldos para el fin del periodo es el saldo de la última semana y para el año es el saldo del último periodo, esto implicó considerar una técnica en el desarrollo del modelo llamado agregaciones.
3. Saldo de Tesorería: Para esta transacción fue necesario considerar que el valor de saldos para el fin del periodo para bancos y cajas es el saldo de la última semana y para el año es el saldo del último periodo, esto implicó considerar una técnica en el desarrollo del modelo llamado agregaciones.
4. Modelo de Cobros y cartera: Para este modelo se tuvo la complicación

de que no se podía unir las transacciones de pagos de cartera y la cartera ya que la cartera era una cabecera resumida y los cobros eran transacciones aplicadas a cada documento de cuentas por cobrar, para ello se manejaron dos tablas de hechos una para cuentas por cobrar y otra para cobranza y una vez diseñado cada cubo se los tuvo que unir mediante una técnica que tiene el Pentaho que se llama cubos virtuales. Para que esto sean consistente las dimensiones de unos de los modelos deben ser las mismas que el otro y de esta forma por cada documento con saldo de cartera se pueden ver los respectivos pagos aplicados.

### **3.3. Mapeo de datos de sistema transaccional a modelos del almacén de datos.**

El sistema transaccional utilizado para el desarrollo de este DataWareHouse es un sistema ERP que tiene bastante parametrización y tiene casi todas sus tablas normalizadas, esto a nivel transaccional es óptimo pero a nivel de lectura no ya que para sacar una vista de información alineada al negocio se tienen que hacer muchos Joins entre las tablas.

Para mejorar la extracción y hacerlo a futuro un proceso fácil de mantener se creó una capa en la base de datos transaccional que a manera de funciones con tipo de datos de salida tablas se podía procesar toda la

lógica del negocio y entregar una vista con datos consistentes, con las respectivas operaciones del negocio y con una estructura para poder crear un modelo de copo de nieve con sus respectivas tablas de hechos y dimensiones.

Por citar un ejemplo para el modelo de ventas las tablas transaccionales son cerca de 8 pero en el DataWareHouse solo una tabla de hechos de ventas que con la respectiva lógica se crearon dimensiones que permitieron a nivel de cubos contar con filtros que le daban diversos enfoques a la información, enfoques tales como financiero, operativo sin necesidad de crear tablas de hechos o datamarts aislados.

La siguiente figura muestra conceptualmente la capa generada para simplificar el proceso de extracción.

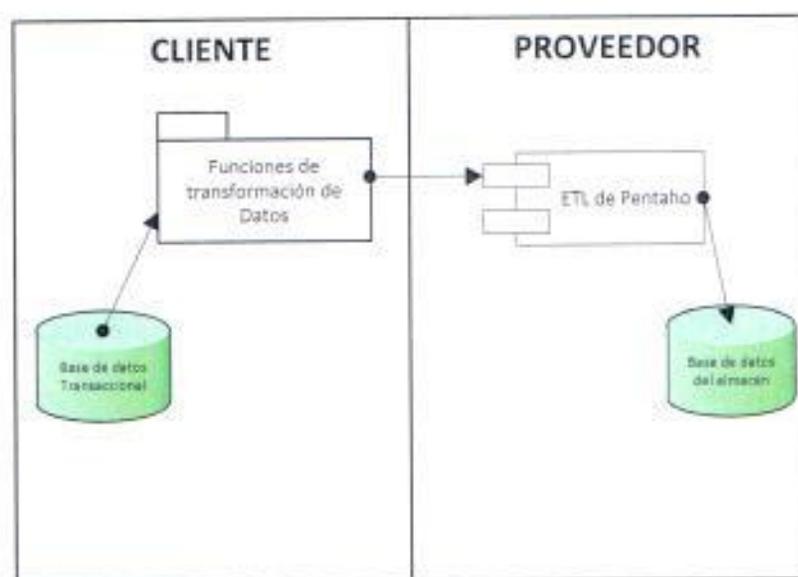


Ilustración 7 Capa de funciones de pre procesamiento de Datos

Esta capa tiene además otras funciones y ventajas tal como la de poder ocultar el procesamiento de datos del cliente, de esta forma ellos entregan la función que arroje los campos para la tabla de hechos y la lógica de sus sistema se la puede ocultar, ya que estas funciones pueden ser encriptados.

Generalmente el cliente conoce todas sus estructuras y son celosos en mostrar información de la misma este mecanismo ayuda a vencer esa barrera e incluso dejar el trabajo de generar sentencias sql en manos del cliente que es quien conoce sus estructuras y reglas del negocio.

De esta forma el mapeo de datos de cada una de las tablas de dimensiones y hechos quedaría de la siguiente manera:

Modulo	Cubo	Nombre de Cubo	Tabla Hechos	ETL
Comercial	<a href="#">Cubo de Ventas Comercial</a>	VentasC	HEC_Ventas	ETL_Facturacion
Comercial	<a href="#">Cubo de Ventas Operativa</a>	VentasOperativa	HEC_Venta_Operativa	ETL_Facturacion
Comercial	<a href="#">Cubo de Ventas Rentabilidad</a>	VentasRentabilidad	HEC_Ventas_Margen	ETL_Facturacion
Logistica	<a href="#">Cubo de Inventario</a>	Inventario_Saldos	HEC_Inventario_S	ETL_Inventarios

	<u>Saldos</u>		aldos	
Logistica	<u>Cubo de Compras</u>	Compras	HEC_Compras	ETL_Compras
Financiero	<u>Cubo Estado de Resultados</u>	Contabilidad	HEC_Contabilidad	ETL_Cont
Financiero	<u>Cubo Cheques Protestados</u>	Protestos	HEC_Cheques_Protestados	ETL_Hechos
Financiero	<u>Cubo Cobranza CXC</u>	Cobranzacxc	HEC_Cartera y HEC_CobranzaEfectivo	ETL_CXC
Financiero	<u>Cubo Cuentas Por Pagar</u>	CxP	HEC_Cxp_Saldos	ETL_Hechos
Financiero	<u>Cubo Cuentas Por Cobrar</u>	CxCCon	HEC_Cartera	ETL_CXC
Financiero	<u>Detalle de Pagos</u>	Pagos	HEC_Detalle_Pagos	ETL_Hechos
Financiero	<u>Cubo de Gastos</u>	GastosCajaChica	HEC_Gastos	ETL_Tesoreria
Financiero	<u>Cubo de Saldos de Tesoreria</u>	SaldosTesoreria	HEC_Saldos_Tes	ETL_Tesoreria
Financiero	<u>Cubo de movimientos de Tesoreria</u>	Tesoreria	HEC_Saldos_Tes_Det	ETL_Tesoreria

Tabla #11 Mapeo de Cubos, tablas de Hechos y Procesos ETL

Para las dimensiones del DataWareHouse se usaron las siguientes tablas que se crearon a partir de diversas tablas de maestros del sistema transaccional.

Tabla DWH
DIM_Unidad_Negocio
DIM_Categoria_UNegocio
DIM_Oficinas
DIM_Clientes
DIM_Bodegas
DIM_Empleados
DIM_Productos
DIM_Pais
DIM_Provincias
DIM_Regiones
DIM_TipoCultivos
DIM_TipoDocCxc
DIM_TipoDocContable
DIM_UnidadMedidad
DIM_Tipo_Documento
DIM_CategoriaArticulos
DIM_TipoFumigacion
DIM_Tiempo

DIM_Ciudades
DIM_Fincas
DIM_MotivosContables
DIM_LineasNegocios
DIM_ArticuloProveedor
DIM_Zonas
DIM_Proveedores

Es importante notar en este mapeo que se pueden generar diversas vistas desde un solo maestro del sistema transaccional, por ejemplo si se observa las dimensiones de pilotos, vendedores, clientes, proveedores todas se alimentan de una tabla que se llama participantes.

En el proceso de extracción se diseñó el diagrama de flujo de la siguiente manera mostrada en el grafico #8.

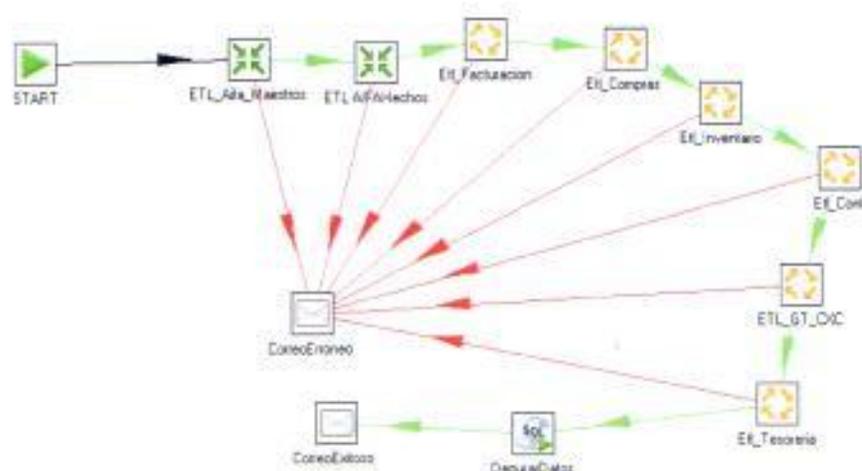


Ilustración 8 Diagrama de Flujo del proceso de extracción

Este diagrama de flujo define el proceso por módulos, en donde cada módulo tiene las sentencias para extraer y transformar las respectivas tablas de hechos y dimensiones necesitadas por el respectivo modelo de datos.

Cada componente de extracción contiene los siguientes pasos:

1. Eliminación de datos a procesar: Antes de traer la data al DataWarehouse elimina para traer la data de nuevo, esto se hace con el fin de que si en el sistema transaccional hubo algún ajuste este cambio se vea reflejado en el DataWarehouse.
2. Extracción de Datos desde el sistema transaccional: Por rango de fechas configurados en tablas de trabajo se pone el rango de fechas a procesar que debe ser consistente con el rango de fechas eliminados

para traer incrementalmente la data al DataWareHouse. Este proceso de extracción contiene todas las tablas de hechos y dimensiones correspondientes al módulo.

3. Sentencias SQL se ajustes: Una vez traída la data de las tablas de hechos y dimensiones es necesario validar para aplicar técnicas de diseño de DataWareHouse, ya sean para hacer consistentes y válidas las dimensiones o generar tablas de agregaciones a partir de las tablas de hechos, es por esto que cada proceso de extracción tiene al final un módulo de ajustes basados en sentencias SQL.

Considerando esto los flujos de extracción para los diversos datamarts del DataWareHouse son:

Paso 1. Inicio

Paso 2. ETL \_Maestros. Este paso extrae todos los maestros compartidos por todo el AD

Paso 3. ETL \_Hechos. Este paso extrae ciertas tablas de hechos que básicamente corresponde al módulo de CXP.

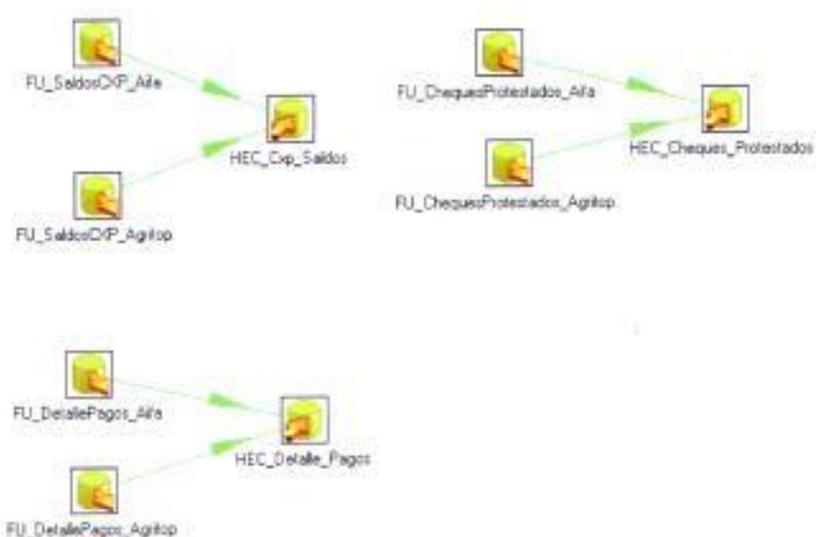


Ilustración 9 Diagrama de extracción Tablas de Hechos

Paso 4. Etl\_Facturación. Subproceso de extracción enfocado exclusivamente al módulo de ventas.

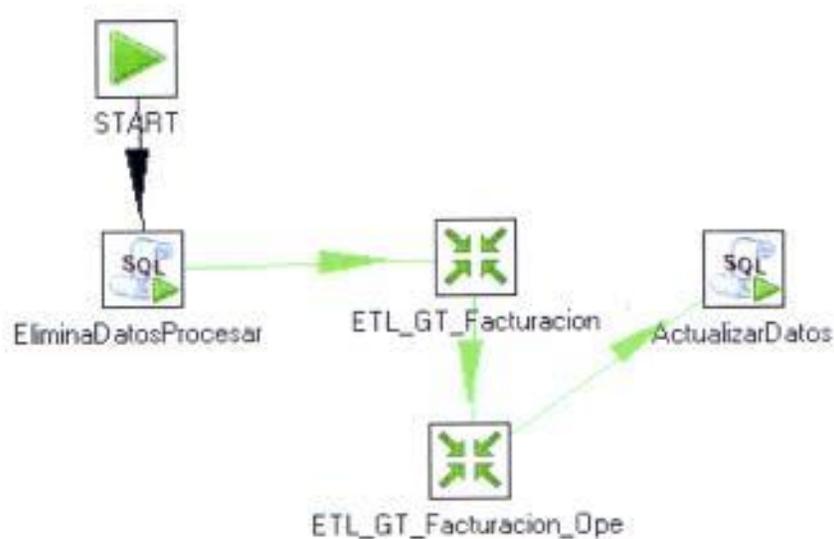


Ilustración 10 Flujo de Extracción para DataMart de Ventas

Paso 5. Etl\_Compras. Subproceso de extracción enfocado exclusivamente al módulo de compras

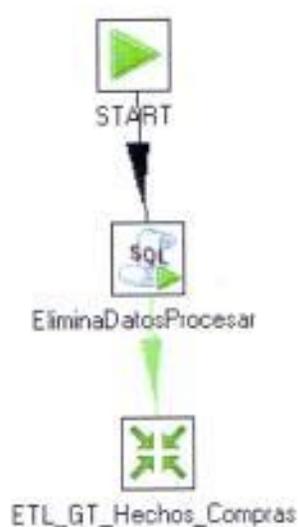


Ilustración 11 Flujo de Trabajo para DataMart de Compras

Paso 6. Etl\_Inventarios. Subproceso de extracción enfocado exclusivamente al módulo de Inventarios

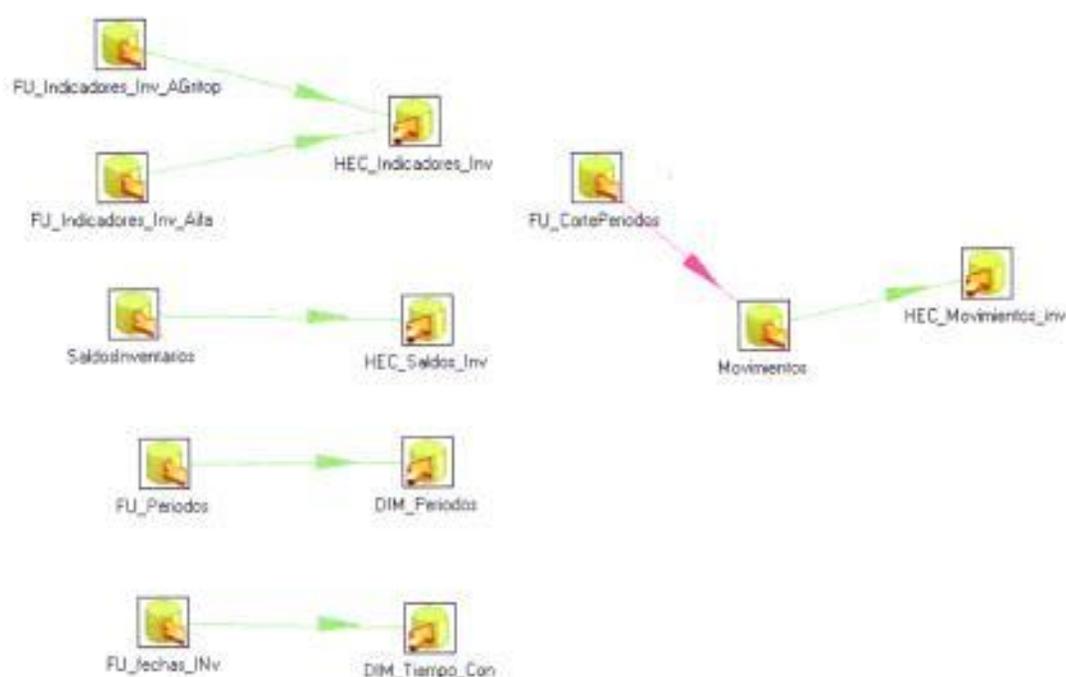


Ilustración 12 Diagrama de Flujo para DataMart de Inventarios.

Paso 7. Etl\_Cont. Subproceso de extracción enfocado exclusivamente al módulo de contabilidad

Este diagrama de flujo realiza:

1. Elimina datos a actualizar
2. Trae Maestros para dimensiones propias del módulo
3. Extrae tablas de Hechos para Contabilidad
4. Inserta en cuentas contables los cálculos para rentabilidad
5. Realiza ajustes de dimensiones y cálculos para mantener consistencias

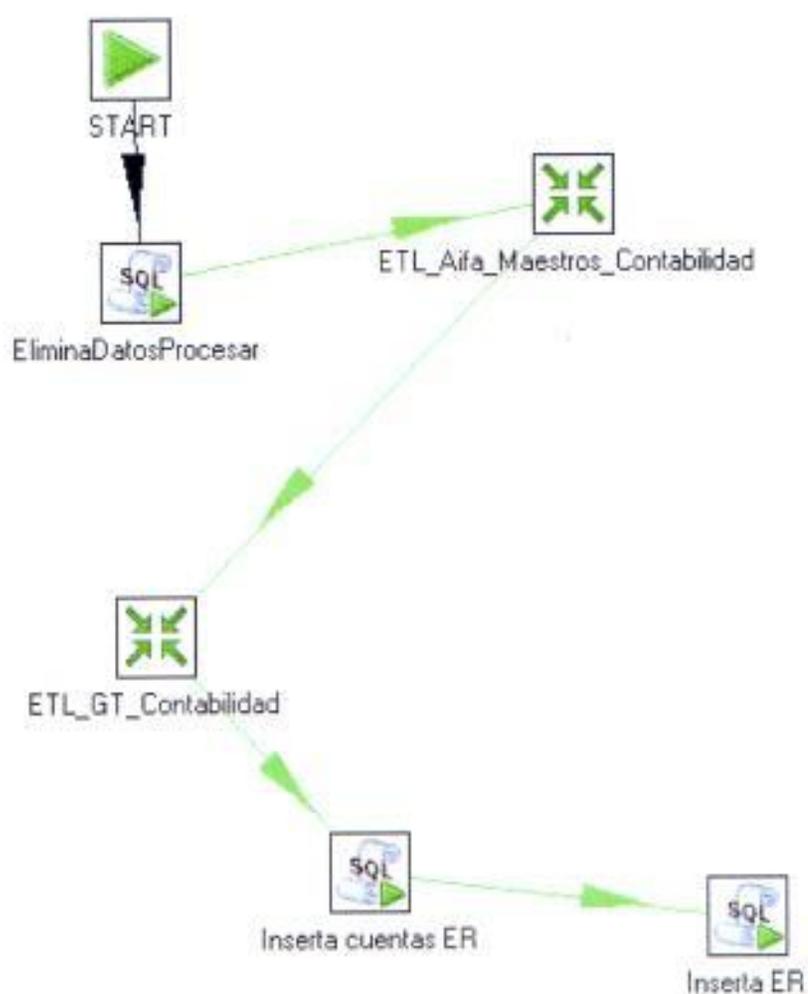


Ilustración 13 Diagrama de flujo para DataMart de contabilidad

Paso 8. Etl\_CXC. Subproceso de extracción enfocado exclusivamente al módulo de Cuentas por cobrar.

Paso 9. Etl\_Tesoreria. Subproceso de extracción enfocado exclusivamente al módulo de Tesoreria.

Paso 10. Depuración de Datos. Proceso dedicado a validar dimensiones y aplicar parametrizaciones a tipos de documentos contables para el proceso de ventas.

Este proceso de extracción y transformación llena y mantiene actualizada las siguientes estructuras:

agg_2_HEC_Inventario_Saldos_anio	
PK	<u>Id_Unidad_Negocio</u>
PK	<u>Anio</u>
PK	<u>Periodo</u>
PK	<u>Id_Oficina</u>
PK	<u>Id_Bodega</u>
PK	<u>Id_Proveedor_comp</u>
PK	<u>Id_Producto</u>
	CantidadIngreso
	CantidadIngresoAcum
	CantidadEgreso
	CantidadEgresoAcum
	Existencia
	MontoIngreso
	MontoIngresoAcum
	MontoEgreso
	MontoEgresoAcum
	Saldo
	contador
	CostoVentaAcum
	CostoVentaIngresosAcum
	CostoVentaEgresosAcum
	CostoVenta
	IdCategoriaRaiz
	rotacion

HEC_Contabilidad	
	Id_CuentaCtable
	Id_Unidad_Negocio
	Id_Oficina
	IdModulo
	IdTransaccion
	IdMotivo
	IdDocumento
	IdLote
	IdDiario
	IdAsiento
	IdSecuencia
	IdCuenta
	IdCosto
	S1
	S2
	S3
	S4
	S5
	S6
	S7
	S8
	S9
	S10
	Fecha
	TotDebito
	TotCredito
	Saldo
	TipoValor
	modulo

ETL_Performance	
	ID_BATCH
	TRANSNAME
	STATUS
	LINES_READ
	LINES_WRITTEN
	LINES_UPDATED
	LINES_INPUT
	LINES_OUTPUT
	ERRORS
	STARTDATE
	ENDDATE
	LOGDATE
	DEPDATE
	REPLAYDATE

agg1_HEC_Saldos_Tes_mes	
PK	<u>Id_Unidad_Negocio</u>
PK	<u>Id_Oficina</u>
PK	<u>Id_Banco</u>
PK	<u>Banco</u>
PK	<u>Id_Cuenta</u>
PK	<u>IdFormaPago</u>
PK	<u>TipoSaldo</u>
PK	<u>anio</u>
PK	<u>Periodo</u>
PK	<u>Semana</u>
	FactCount
	MontoEgreso
	MontoIngreso
	Saldo

DIM_Fabricante	
	Id_Fabricante Id_Fabricante_org Fabricante

DIM_Tipos_Segmentos	
	Id_Tipo_Segmento Tipo_Segmento

HEC_Cxp_Saldos	
	Id_Unidad_Negocio Id_Oficina IdOrdenChq TipoParticipante Participante IdTipoDocumento NumDocumento NumeroFactura FechaSolicitud FechaEmision FechaPago DiasCredito fechapagofinal DiasAtraso TipoDocAsociado NumDocAsociado Estado TotalDocumento Descuentos Anticipos TotalRetenciones TotalNC ValorPagar pagoEnCaja Saldo ReferenciaPago Documento idrange

HEC_Saldos_Tes_Det	
	Id_Unidad_Negocio Id_Oficina Id_Banco Banco Id_cuenta Id_Beneficiario Participante IdFormaPago IdTipoMovimiento FechaContable Tipo Monto Motivo DocumentoContable

PRO_Fechas	
	desde hasta modulo pdesde phasta

AUD_TablaLog	
	ID_JOB JOBNAME STATUS LINES_READ LINES_WRITTEN LINES_UPDATED LINES_INPUT LINES_OUTPUT ERRORS STARTDATE ENDDATE LOGDATE DEPDATE REPLAYDATE LOG_FIELD

DIM_Tiempo	
PK	Fecha
	año mes día díaSemana nombreDía nombreMes AñoPeriodo Periodo SemanaPeriodo NombrePeriodo NombreSemanaP NombreDíaMes

HEC_Venta_Operativa	
	NumeroFactura Id_Unidad_Negocio Id_Productos Id_Proveedor Id_Vendedor Id_Bodega Id_Cliente IdGrupoCliente Id_Oficina Id_Caja IdMotivo IdDocContable TipoFumigacion TipoCultivo IdHacienda LineaNegocio GrupoCliente TipoCliente IdAvioneta IdPiloto FechaEmision SubTotal Recargo TotalDescuentos TotalIva TotalFactura Cantidad Costo Precio Iva Descuento DescuentoIva ValorBruto CantidadDevoluta ValorDevolucion TipoDoc FechaHoja IdTipoFumigacion Id_CuentaCtable hectareas AsistenciaTecnica

HEC_Inventario_Saldos	
PK	Id_Unidad_Negocio
PK	Id_Periodo
PK	Id_Oficina
PK	Id_Bodega
PK	Id_Proveedor_comp
PK	Id_Producto
	CantidadIngreso CantidadIngresoAcum CantidadEgreso CantidadEgresoAcum Existencia MontoIngreso MontoIngresoAcum MontoEgreso MontoEgresoAcum Saldo CostoVentaAcum CostoVentaIngresosAcum CostoVentaEgresosAcum EsBodegaConsignacion CostoVenta IdCategoriaRaiz rotacion

HEC_GastosCajachica	
	Id_Unidad_Negocio Id_Oficina Id_caja Custodio Beneficiario Departamento Cargo Concepto fecha ValorVale ValorDeducible ValorNoDeducible detalle

DIM_Vendedores	
PK	<u>Id_Vendedor Comp</u>
	Id_Vendedor_Org Id_Unidad_Negocio Id_Oficina Id_Supervisor Id_Tipo_Vendedor Id_Estado Identificacion Nombre

HEC_Cartera	
	FechaEmision FechaVencimiento Id_Unidad_Negocio Id_Cliente Id_Estado Id_Cultivo Id_Tipo_Documento Id_Oficina_Registro Id_Oficina Numero_Documento Id_Vendedor Dias_Credito_Cliente Dias_vencido Monto_Original Monto_Cobranza Cartera ApliEfecti Saldo Montos_otros Monto_Cobrado Apl_Cartera NoEfectivizado Porc OtrosMonto Iva IdGrupoCliente Id_TipoCliente Vencido

HEC_Detalle_Pagos	
	Id_Unidad_Negocio Id_Oficina Id_Proveedor_comp TipoParticipante IdOrdenChq IdBanco Banco IdCuenta Cuenta IdFormaPago FormaPago Egreso Refer_NumChq BeneficiarioFinal Fecha FechaEmision FechaVencimiento Monto IdBeneficiarioInicial DiasPago FechaFinal Diasatraso

DIM_MotivosContables	
	Id_Motivo descripcion

DIM_Clientes	
PK	<u>Id_Cliente_comp</u>
	Id_Cliente_org Id_Unidad_Negocio Id_Tipo_Identificacion Identificacion Nombres Id_estado_cliente Id_tipo_persona fecha_nac Id_Genero Id_Nacionalidad Id_Tipo_Cliente Id_Categoria Id_Segmento_venta Id_Segmento_rent Id_Segmento_cob fecha_ingr fecha_ult_compra Id_Oficina Id_Oficina_gest

HEC_CostoVentas	
	Id_Unidad_Negocio Id_Oficina Id_Bodega IdMotivo Id_Cliente IdOrdenVenta IdGrupoCliente Id_Proveedor Id_Vendedor IdDocumentoContable Id_Productos FechaContable FechaTransaccion FechaGrabacion Cantidad Costo Monto Descuento CostoVenta NumerFactura

HEC_Cheques_Protestados	
	IdEmpresa IdBanco IdCuenta NumeroCuenta IdNotaDebito FechaDebito Id_Cliente Concepto NumeroCheque IdAvisoCobro MontoProtestado Id_Oficina IdTipoDocCobro IdDocCobro

DIM_Productos	
PK	<u>Id_Producto_comp</u>
	Id_Producto_org Id_Categoria Id_Unidad_Negocio Descripcion Id_Unidad_Medida Costo_Unitario Id_Fabricante Id_Marca Id_Segmento_Margen Id_Segmento_Rotacion Indice_Rotacion Id_Tipo_Producto Id_Categoria_cont TipoVenta PermitirInventariar AliasArticulo

HEC_Cobranza	
	Id_Unidad_Negocio Id_Oficina IdTipoDocumento IdDocumento Id_Oficina_Cobro IdRecibo Id_Cliente Cobranza Efectivizado FechaCobro NumeroFactura SaldoInicial PlazoPago FechaPago EstadoEfectivizado FechaFactura FechaEfectivizacion TipoAplicacion FechaCmpRte IdTipoDoc Id_Vendedor IdTipoCultivo FechaVencimiento DiasPago TipoDoc TotalFactura

HEC_Movimientos_Inv	
	Id_Unidad_Negocio Id_Oficina Id_Bodega Id_Proveedor_comp Id_Producto Id_tipo_mov_inv TipoMovimiento FechaContable FechaMovimiento Unidades Costo Monto CantAnterior CantActual

HEC_Compras	
	Id_Unidad_Negocio Id_Oficina Id_Oficina_Req Id_Oficina_Orden Id_Bodega Id_Proveedor_comp IdArticulo IdTipoCompra IdTipoDocumento IdCompra IdRequerimiento FechaRequerimiento FechaCompra FechaEntregaMercad DiasEntrega NumOC Cantidad CostoUnitarioCompra ValorDescuento Cargos ValorIVA ValorICE Bonificaciones CostoLiqImp TotalCompra ValorBruto IdEstado IdDepartamento IdLiqImp StockTransito CostoUnitarioLiqImp PesoLiqImp CostoPesoLiqImp TotalProrateadoLiqImp Departamento

HEC_Venta_Operativa_E
NumeroFactura
Id_Unidad_Negocio
Id_Productos
Id_Proveedor
Id_Vendedor
Id_Bodega
Id_Cliente
IdGrupoCliente
Id_Oficina
Id_Caja
IdMotivo
IdDocContable
TipoFumigacion
TipoCultivo
IdHacienda
LineaNegocio
GrupoCliente
TipoCliente
IdAvioneta
IdPiloto
FechaEmision
SubTotal
Recargo
TotalDescuentos
TotalIva
TotalFactura
Cantidad
Costo
Precio
Iva
Descuento
Descuentovta
ValorBruto
CantidadDevolucion
ValorDevolucion
TipoDoc
FechaHoja
IdTipoFumigacion
Id_CuentaCtable
hectareas

DIM_CtaCtable	
PK	<u>Id_CuentaCtable</u>
	IdCategoria1
	Categoria1
	Orden1
	IdCategoria2
	Categoria2
	Orden2
	IdCategoria3
	Categoria3
	Sigla
	Orden3
	OrdenResumen
	IdCategoria4
	Categoria4
	Orden4
	IdCategoria5
	Categoria5
	Orden5
	IdCuenta
	IdCCosto
	Nombre
	Naturaleza
	Operacion
	IdEmpresa

DIM_Unidad_Negocio	
PK	<u>Id_UNegocio_comp</u>
	Id_UNegocio_org
	Nombre
	Id_Categoria_UNegocio

DIM_Tipo_Producto	
PK	<u>Id_Tipo_Producto</u>
	Tipo_Producto

HEC_Indicadores_Inv	
	Id_Unidad_Negocio
	Id_Oficina
	Id_Bodega
	Id_Proveedor
	Id_Producto
	IdModelo
	Modelo
	Abreviacion
	Cantidad
	StockMinimo
	StockMaximo
	StockTransito
	CantidadLogica
	CantidadSugerida
	RequiereComprar
	FechaCorte

DIM_RangosCartera	
PK	<u>Id_Rangos</u>
	Desde
	hasta
	Descripcion
	Tipo
	Orden

DIM_Tipo_Persona	
	Id_Tipo_Persona
	Tipo_Persona

DIM_Estado_Compra	
	IdEstado
	Estado

DIM_Proveedores	
PK	<u>Id_Proveedor_comp</u>
	Id_Proveedor_org
	Id_Unidad_Negocio
	IdTipoidentificacion
	Identificacion
	TipoParticipante
	Nombre
	Id_Estado
	Ruc
	IdTipoNacionalidad
	TipoProveedor

DIM_estado_Vendedores	
	Id_estado Estado

DIM_Periodos	
PK	<u>Id_Periodo</u>
	año Periodo semana NombrePeriodo NombreSemana

HEC_Costo_semana	
PK	<u>Id_Unidad_Negocio</u>
PK	<u>Id_Oficina</u>
PK	<u>Id_Vendedor</u>
PK	<u>Id_Cliente</u>
PK	<u>Id_Periodo</u>
PK	<u>Id_Productos</u>
	Id_Proveedor CostoVenta

agg2_HEC_Saldos_Tes_año	
PK	<u>Id_Unidad_Negocio</u>
PK	<u>Id_Oficina</u>
PK	<u>Id_Banco</u>
PK	<u>Banco</u>
PK	<u>Id_Cuenta</u>
PK	<u>IdFormaPago</u>
PK	<u>TipoSaldo</u>
PK	<u>año</u>
PK	<u>Periodo</u>
	FactCount MontoEgreso MontoIngreso Saldo

DIM_ArticuloProveedor	
	Id_Productos Id_Proveedor nombre TipoProveedor

DIM_Tipo_Segmento_Producto	
	Id_Tipo_Segmento_Producto Tipo_Segmento_Producto

DIM_Bodegas	
PK	<u>IdBodega</u>
	IdEmpresa IdOficina Bodega Tipo

DIM_Pos_Cta	
	sigla posicion

DIM_Tipo_Documento	
	Id_Tipo_Documento IdEmpresa IdDocumento IdFormulario DescAlterno

DIM_Tipo_Oficina	
	Id_Tipo_Oficina Tipo_Oficina

DIM_Bancos	
	Id_Banco Banco

DIM_Categoria_Productos	
PK	<u>IdCategoria</u>
	IdCategoriaPadreId IdEmpresa Nombre NombrePadre

DIM_Tipo_Compra	
PK	<u>idtipocompra</u>
	tipocompra

DIM_TipoCajaBanco	
	TipoSaldo

DIM_TipoCliente	
PK	<u>Id_TipoCliente</u>
	TipoCliente

DIM_Tipo_Movimiento_tes	
	IdTipoMovimiento Descripcion

DIM_Ciudades	
PK	<u>Id_Ciudad</u>
	Id_Provincia Ciudad

DIM_Tipo_Movimiento_inv	
PK	<u>Id_Tipo_mov_inv</u>
	Tipo_mov_inv

DIM_Regiones	
	Id_Region Id_Pais Region

HEC_Ventas	
	NumeroFactura Id_Unidad_Negocio Id_Productos Id_Vendedor Id_Bodega Id_Cliente IdGrupoCliente Id_Oficina Id_Caja IdMotivo TipoFumigacion TipoCultivo IdHacienda FechaEmision SubTotal Recargo TotalDescuentos TotalIva TotalFactura Cantidad Costo Precio Iva Descuento ValorBruto CantidadDevolucion LineaNegocio ValorDevolucion IdAvioneta IdPiloto Descuentovta GrupoCliente TipoCliente IdDocContable TipoDoc IdTipoIngreso FechaHoja IdTipoFumigacion Id_CuentaCtable Id_Proveedor Id_Tipo_venta hectareas Refacturar AsistenciaTecnica IdOrden Mayorizado diasfact FechaEgreso IdClienteTipoVenta ClienteTipoVenta

HEC_Gastos	
	Id_Unidad_Negocio Id_Oficina Id_caja Custodio TipoBeneficiario Beneficiario Departamento Cargo IdConcepto Concepto IdTipoGasto TipoGasto fecha detalle TipoCaja TipoProceso Valorgasto ValorDeducible ValorNoDeducible EsGasto IdBeneficiario TipoMovimiento IdCultivo Id_LineaNegocio

DIM_TipoDocContable	
	Id_TipoDoc descripcion TipoTransaccion

DIM_Categorias_Clientes	
	Id_Categoria_Comp Id_Categoria_Org Nivel1 Nivel2 Nivel3 Nivel4 Nivel5

HEC_OrdenesCheque	
	TipoOrdenCheque Id_Unidad_Negocio Id_Oficina Id_caja Fecha TipoDocumento monto

DIM_Segmento_Productos	
	Id_Segmento_Producto Id_Segmento_Producto_org Segmento_Producto

DIM_Oficinas	
<b>PK</b>	<b>Id_Oficina_Comp</b>
	Id_Oficina_Ori Id_Unidad_Negocio Id_Tipo_Oficina Id_Ciudad Id_Zona Estado Identificacion Descripcion

DIM_Cuentas_Bco	
	Id_cuenta NumeroCuenta TipoCuenta

DIM_TipoVenta	
<b>PK</b>	<b>IdTipoVenta</b>
	TipoVenta

agg_1_HEC_Inventario_Saldos_mes	
PK	<u>Id_Unidad_Negocio</u>
PK	<u>Anio</u>
PK	<u>Periodo</u>
PK	<u>Semana</u>
PK	<u>Id_Oficina</u>
PK	<u>Id_Bodega</u>
PK	<u>Id_Proveedor_comp</u>
PK	<u>Id_Producto</u>
	CantidadIngreso CantidadIngresoAcum CantidadEgreso CantidadEgresoAcum Existencia MontoIngreso MontoIngresoAcum MontoEgreso MontoEgresoAcum Saldo contador CostoVentaAcum CostoVentaIngresosAcum CostoVentaEgresosAcum CostoVenta IdCategoriaRaiz rotacion

HEC_Ventas_Margen	
PK	<u>Id_Unidad_Negocio</u>
PK	<u>Id_Oficina</u>
PK	<u>Id_Vendedor</u>
PK	<u>Id_Cliente</u>
PK	<u>Id_Periodo</u>
PK	<u>Id_Productos</u>
	Cantidad Iva Descuento ValorBruto CantidadDevuelta ValorDevolucion Descuentovta hectareas Costo

DIM_Segmentos	
	Id_Segmento Id_Tipo Segmento

DIM_Categoria_UNegocio	
PK	<u>Id_Categoria_UNegocio_Comp</u>
	Id_Categoria_UNegocio_Org Descripcion

PRO_Cubos_Medidas	
	IdCubo Cubo IDMedida Descripcion

DIM_Generos	
	Id_Genero Genero

DIM_TipoDocCxc	
	Id_TipoDoc TipoDocCxc TipoTransaccion

DIM_Marca_Productos	
	Id_Marca Id_Marca_org Marca

DIM_Segmentos_cbles	
PK	<u>Id_Segmento</u>
	IdEmpresa IdSegmento TipoSegmento IdItem Nombre

DIM_Aplicacion	
	TipoAplicacion IdTipoAplicacion

DIM_Tipo_Ingreso	
PK	<u>IdTipoIngreso</u>
	TipoIngreso

DIM_LineasNegocios	
	Id_LineaNegocio descripcion

DIM_Paises	
	Id_Nacionalidad Nacionalidad

DIM_Valores_Logicos	
PK	<u>Id_Valor_Logico</u>
	Valor_Logico

DIM_UnidadMedida	
PK	<u>IdUnidadMedida</u>
	Nombre



### 3.4. Uso de herramienta de sincronización para extracción del sistema transaccional con el almacén de datos.

En la sección anterior se hizo referencia al uso de la herramienta de Pentaho para diseñar el proceso de extracción, pero esta herramienta tiene una limitante, que es que está orientada a redes LAN y accede directamente a los puertos de las bases de datos, circunstancias que si queremos hacer el proceso de carga y extracción vía internet se va a complicar.

Para ello en la sección 2.5 se mencionó una herramienta de sincronización llamada Iconnector. Una vez que se tenga diseñado todo el proceso de Extracción y carga podemos utilizar las mismas sentencias SQL para configurarlas en esta herramienta, la ventaja es que esta herramienta solo necesita que se abra un puerto y por medio del mismo envía la trama de datos comprimidas, aprovechando el canal y haciendo posible una sincronización vía internet.

Una vez configurado todos los scripts podremos hacer sincronización via internet sin abrir puertos de bases de datos y comprimiendo la trama para aprovechar el ancho de banda.

### 3.5. Esquemas de mantenimiento y actualización del almacén de dato para los respectivos ETL.

Para mantener actualizado este DataWareHouse se utilizará actualizaciones incrementales que van sin incluir desde el último día de 3 meses atrás a la fecha, esto se hace con el fin de poder hacer cargas incrementales con un periodo que incluya automáticamente cualquier actualización a periodos anteriores.

Por lo general en las empresas suelen abrir periodos cerrados de meses atrás para corregir transacciones, con este esquema se hará la respectiva actualización en el modelo de datos de manera automática.

En caso de requerirse actualizaciones en fechas anteriores que estén fuera de este rango se ha definido una tabla en el datawarehouse llamada PRO\_Fechas que contiene los siguientes campos:

1. Fecha Desde, fecha inicial del rango
2. Fecha Hasta, fecha final del rango
3. Modulo, hace referencia al DataMart, por ejemplo FAC indica que este registro contiene el rango de fechas para el DataMart de Facturación.
4. PDesde, tiene la misma función que la fecha inicial solo que está en formato de periodos, por ejemplo 201007 es el periodo 7 del año 2010
5. PHasta, tiene la misma función que la fecha final solo que está en formato de periodos, por ejemplo 201007 es el periodo 7 del año 2010

De esta forma la tabla contiene los rangos de fechas de los procesos, cada proceso del ETL llama a este rango de fechas antes de hacer el proceso de extracción.

La siguiente tabla es una muestra de esta tabla:

desde	Hasta	modulo	pdesde	phasta
2009-12-01	2009-12-31	CIN	NULL	NULL
2009-01-01	2009-01-31	CVE	NULL	NULL
2009-12-01	2009-12-31	CCC	NULL	NULL
2009-12-01	2009-12-31	CCP	NULL	NULL
2009-12-01	2009-12-31	CBC	NULL	NULL
2009-12-01	2009-12-31	CCH	NULL	NULL
2010-01-01	2010-12-31	CON	NULL	NULL
2010-01-01	2010-12-31	COM	NULL	NULL
2010-01-01	2010-12-31	FAC	200911	2009131
2010-01-01	2010-12-31	CXC	NULL	NULL
2010-01-01	2010-12-31	INV	200911	2009131
2010-01-01	2010-12-31	TES	200911	2009131

Para hacerlo compatible en internet hay que tener presente de configurar esta tabla en los diversos scripts de extracción que utiliza la herramienta de sincronización.

## CAPITULO 4

### 4. Inteligencia de Negocios sobre la plataforma creada.

#### 4.1. Generación de modelos de datos

##### 4.1.1 Enfoque en Información resumida

Uno de los principales retos en la implementación de proyectos de inteligencia de negocios es transmitir bien el concepto de información detallada asociada con la operación y la información resumida asociada a lo estadístico y gerencial.

El motivo por el cual esto es muy importante es porque ayuda a definir el alcance y delimitar al cliente a que no pida hacer funcionar a la solución como algo operativo que procese reportes transaccionales, sino más bien enfocarse en el soporte a la toma de decisiones.

La facilidad de navegación que ofrecen los cubos de información, los dashboards que se les puede dar la funcionalidad de drill-down hace que el usuario piense en usar a la herramienta para completar falencias de reportes de su sistema transaccional y si caemos en esto la solución presentará muchos problemas y no tendrá la visión adecuada.

Por lo expuesto un sistema de Inteligencia de Negocios debe enfocarse a información resumida, en lo posible diseñar dashboards enfocados en temáticas del negocio que con indicadores muestren la situación de la organización y si se necesita un mayor detalle navegar en los cubos y si se quiere llegar a un enfoque transaccional buscar con mejor criterio el reporte respectivo pero en el sistema transaccional y así delimitar bien los detalles de información requerida.

#### 4.1.2 Enfoque en Información detallada a nivel transaccional

Para hacer simple este enfoque una información detallada a nivel transaccional es cuando en el informe a revisar incluye hasta el detalle de los documentos e incluso el contenido de los documentos. Por citar un ejemplo un detalle a número de documento en un módulo de facturación es la factura, pero se puede llegar a mas detalle ya que la factura puede contener varias órdenes de pedidos y cada orden de pedido productos.

El momento que nuestro análisis de información llega a este nivel de detalle nuestro enfoque es transaccional y generalmente esto sirve para controles y auditorías pero para análisis gerencial sirve mucho más el enfoque resumido.

#### 4.1.3 Análisis Multidimensional de la Información.

Generalmente cuando se revisa un reporte operativo emitido por algún

sistema ERP se observa en las columnas una variable de análisis, en las filas otras variables de análisis y en el cruce valores cuantificables, se puede pensar en un reporte que cruce el tiempo con los productos y se revise el valor de la venta bruta y los descuentos.

Pero esto es un reporte limitado ya que un negocio no tiene solo dos o tres variables sino que muchas más, y diseñar un reporte por cada cruce implica muchas horas de desarrollo y lo más común cuando un usuario revisa la información es que su enfoque cambia a medida que sus preguntas van siendo respondidas.

Por eso cuando consumidor de información revisa un reporte de ventas por zonas y ve una zona baja en ventas va a querer luego saber que vendedores están en esas zonas y luego de identificar al vendedor va a querer saber los clientes, luego los productos e incluso el proveedor del producto.

A esta facilidad, o funcionalidad de tener en un modelo de información muchas variables del negocio, muchos valores cuantificables o medidas y le permita al usuario cambiar por sí mismo los diversos enfoques de información y variables del negocio, pudiendo el mismo llegar desde una vista general a una vista con mas dimensiones se lo conoce como análisis multidimensional.

El Análisis Multidimensional de Datos (numéricos, textuales, simbólicos) es la alternativa a la Estadística tradicional para el tratamiento de datos no convencionales, favoreciendo la utilización conjunta de técnicas cuantitativas y cualitativas.

Esta corriente, ha tomado gran impulso debido al avance de la computación y se nutre en la actualidad con los progresos realizados en la inteligencia artificial, lenguajes y bases de datos orientadas a objetos.

Una aplicación de este tipo de análisis da apoyo a la planificación estratégica, el seguimiento de metas, la definición de indicadores que necesitan alimentarse de las transacciones y aplicarle un modelo multidimensional para facilidad de consulta y seguimiento de resultados.

En un proyecto de Inteligencia de negocios se busca integrar en una sola base de datos, conocida como almacén de datos o DataWareHouse, todos los datos requeridos para hacer un análisis global del negocio que suelen tener enfoques operativo, comercial, financiero y logística. Muchas organizaciones cuentan con sistemas diversos para capturar estos datos y hacen que consolidarlos para hacer reportes estadísticos lleve mucho tiempo.

Esto trabajo puede reducirse substancialmente al implementar un proceso de extracción y carga conocido como ETL (Extraction, Load and

Transformation) que actualice el almacén de datos, luego sobre él se implementan herramientas para sacar provecho del análisis multidimensional.

#### 4.1.4 Modelos de datos para análisis comercial

Para hacer un análisis multidimensional para un enfoque comercial los modelos que generalmente se definen son:

- Ventas
- Cartera
- Mercadeo
- Control de Presupuestos de Ventas
- Servicio al Cliente

Estos modelos nos ayudan a identificar problemáticas tales como:

- Seguimiento detallado
- Comparaciones entre productos
- Lanzamiento de Productos
- Promociones
- Resultados del canal
- Base de clientes
- Efectividad en la comunicación

Con estos modelos se puede analizar

- ✓ Ventas de Productos:

- En Importe Neto, Importe Bruto y Margen.
- Por Cliente, Zona Geográfica, Fuerza de Ventas, Familia de Productos
- Por cualquier unidad de Tiempo
- ✓ Análisis del Canal de Ventas
- ✓ Segmentación de Clientes
- ✓ Análisis de la frecuencia de compra y estacionalidad
- ✓ Análisis de tendencias en descuentos
- ✓ Análisis de Devoluciones
- ✓ Análisis de perfiles de Clientes

#### 4.1.5 Modelos de datos para análisis financiero

Para el análisis financiero se dispone de los siguientes modelos de información:

- ✓ Estado de Resultados:
  - Plan de cuentas
  - Créditos, débitos y saldos
  - Utilidades operacionales, antes impuestos
- ✓ Cuentas por cobrar:
  - Saldo de cartera por empresas, oficinas, clientes, vendedores
  - Detalle de cobranza a clientes, demora en pagos, pagos a tiempo
- ✓ Cuentas por pagar:
  - Saldo de pago a proveedores, descuentos, retenciones

- Pagos detallados a proveedores
- ✓ Gastos
  - Gastos por empresas, departamentos, beneficiarios
  - Detalle de los gastos por RRHH, administración.
- ✓ Tesorería y saldos banco

Con estos modelos de información se podrá hacer un análisis del estado financiero de la organización y poder analizar el flujo de efectivo entre lo que una empresa debe pagar a sus proveedores y el tiempo en que sus clientes pagan por los productos o servicios.

Esto permitirá hacer proyecciones de efectivo y prepararse para los meses en los que la empresa tenga bajo flujo de efectivo, dependiendo del modelo de negocio las empresas tienen meses bajos y altos.

#### 4.1.6 Modelos de datos para análisis de logística

Para un análisis de logística se debe tener presente que la rotación de productos y el mantener los stocks adecuados es primordial para un negocio de distribución. Tener productos en bodega sin rotación no genera ganancia, incluso estos productos deben ser analizados en base a su rentabilidad por proveedor ya que existen productos que se adquieren a diversos proveedores y es importante este indicador de rentabilidad.

Un modelo de logística debe enfocarse en lo siguiente:

- ✓ Saldos de inventarios:
  - Por periodos de tiempo, por empresas, por bodegas, por proveedores, por productos
- ✓ Movimiento de inventarios:
  - Por periodos de tiempos, por empresas, por tipo de transacción, por naturaleza de transacciones

#### 4.2. Desarrollo de cuadro de mando integral.

##### 4.1.7 El Análisis de los Reportes de Información Transaccional – Relacional.

El objetivo del DataWarehouse es incorporar, tanto en metadata como en construcción física de la base de datos, la representación del modelo de información previamente diseñado, el cual responde a las necesidades previas identificadas de análisis e información.

Estas necesidades previas de información se las obtiene mediante un proceso de levantamiento de información que consiste en analizar los reportes de información transaccional. En estos reportes están las diversas variables y medidas o indicadores cuantificables que son objeto del análisis de los usuarios consumidores de la información.

Para hacer este análisis es necesario tener respuestas a estas preguntas:

- ¿Cuál es el negocio?
- ¿Cuáles son las preguntas más relevantes por área de negocios que

los usuarios buscan responder con los reportes?

- ¿Cuáles son los procesos involucrados en la toma de decisiones?

Enfocados en las respuestas a estas preguntas podemos empezar el proceso de levantamiento para definir el alcance del proyecto y en base a los reportes de información transaccional poder identificar que no se podrá satisfacer con este DataWareHouse.

Las actividades para realizar este análisis consiste en:

1. Entrevista directa con las personas para conocer sus perspectivas del negocio. De esta forma se obtiene información individual.
2. Reunión de facilitación que consiga el consenso de los participantes sobre el modelo de información. De esta forma se obtiene información en grupo.

Siguiendo estos pasos podremos llegar de un criterio individual a una propuesta grupal y finalmente a un acuerdo colectivo.

Para las reuniones grupales se hacen talleres con lluvias de ideas que luego son alineadas con el personal técnico de sistemas para determinar que es posible obtener del sistema transaccional y que no va a ser posible incluir.

Los pasos a seguir en los talleres son:

**Introducción:** Los participantes se presentan ellos mismos y se presenta los objetivos de la sesión.

**Educación:** Se relevan los conceptos y los procesos.

**Brainstorming:** Es usado para desarrollar listas potenciales de preguntas, dimensiones y hechos.

**Refinamiento:** La lista potencial es revisada y refinada para arribar a un conjunto específico.

**Conclusión:** Se presentan los resultados para revisar su respuesta

Y el resultado de esto es un documento de visión y alcance que forma parte del acuerdo de la solución a desarrollar. En este documento se especifica que se hace por parte del proveedor y que hace el cliente con su personal interno.

Generalmente por parte del cliente queda la tarea de proporcionar el diccionario de datos de sus aplicaciones y definir el equipo de trabajo para el control de calidad, los sistemas del cliente son de dominio de su personal técnico o de asesores en caso de que el cliente tenga en outsourcing la administración tecnológica. Y para evitar conflictos de intereses y no ser juez y parte en el control de calidad el cliente debe proporcionar el equipo que certifique el cuadro de información.

4.1.8 Diseño del cuadro de mando integral para una empresa de importación y distribución de productos.

El cuadro de mando integral se basa en las siguientes perspectivas del negocio:

1. Financiera. ¿cómo nos ven los accionistas?
2. Cliente. ¿cómo somos percibidos por el cliente?
3. Interna. ¿cómo somos percibidos por el cliente?
4. Aprendizaje y crecimiento. ¿podemos seguir mejorando y creando valor?

Este esquema de cuadro de mando integral que se menciona es el de los autores Kaplan y Norton. Con esta técnica nos ayudan hacer una análisis estratégico que ayuda a las organizaciones a determinar sus fortalezas, oportunidades, amenazas y debilidades y definir un plan de acción que ayude a alcanzar los objetivos planteados.

La base de todo es el aprendizaje interno es lo que ayuda a las organizaciones a determinar cuánto valor se puede agregar con el personal y habilidades actuales, y de allí sube a la perspectiva interna, al cliente y finalmente la parte financiera es una efecto de las estrategias aplicadas, pero en donde se trabaja es en el personal y el cliente.

La siguiente figura nos muestra una plantilla para organizar las metas y objetivos utilizando estas perspectivas.

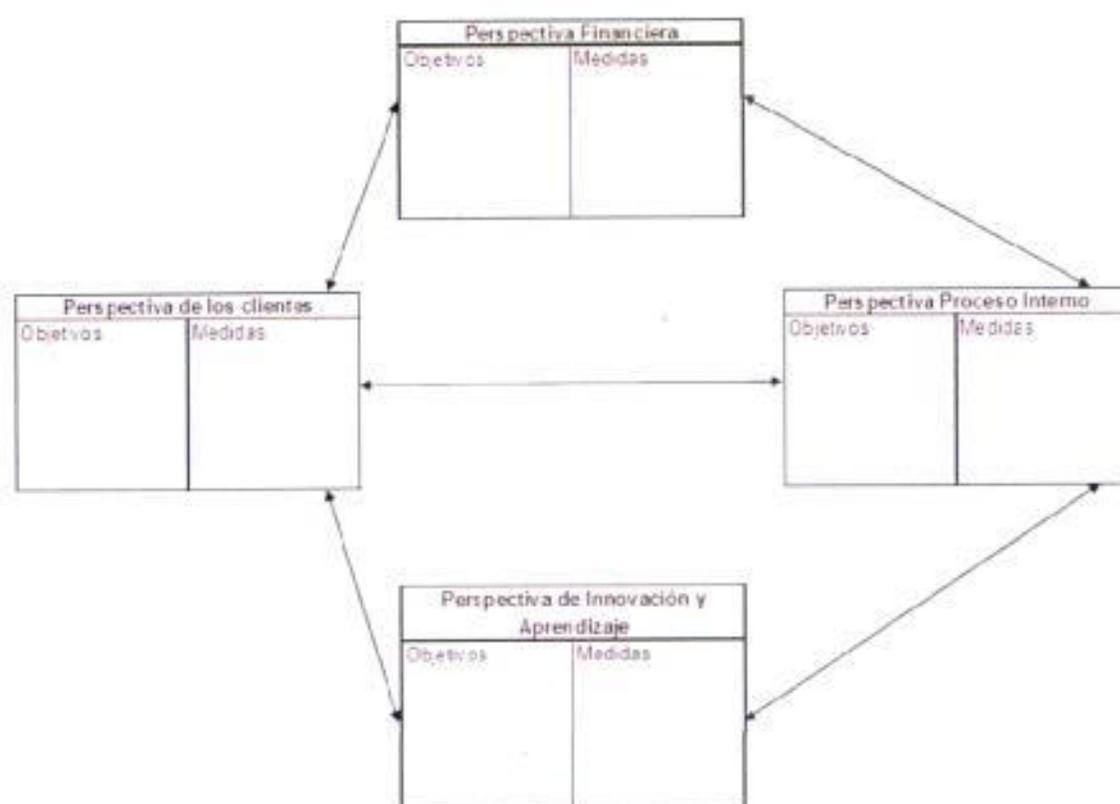


Ilustración 14 Cuadro de mando integral

De esta forma se pueden obtener indicadores que pueden ser diseñados como dashboards sobre la solución de inteligencia de negocios. Pentaho la herramienta que se está utilizando para este proyecto tiene un framework para diseñar dashboards que permite graficar barras, pies, semáforos y varios objetos adicionales que serán útiles para indicadores.

El modelo de cuadro integrar es un modelo causa efecto que busca con acciones sobre el personal interno, la organización causar el efecto de aumentar la rentabilidad y atracción de los clientes.

Los indicadores financieros permiten una gestión reactiva porque explican lo que ha pasado, no son "causas" sino "resultados".

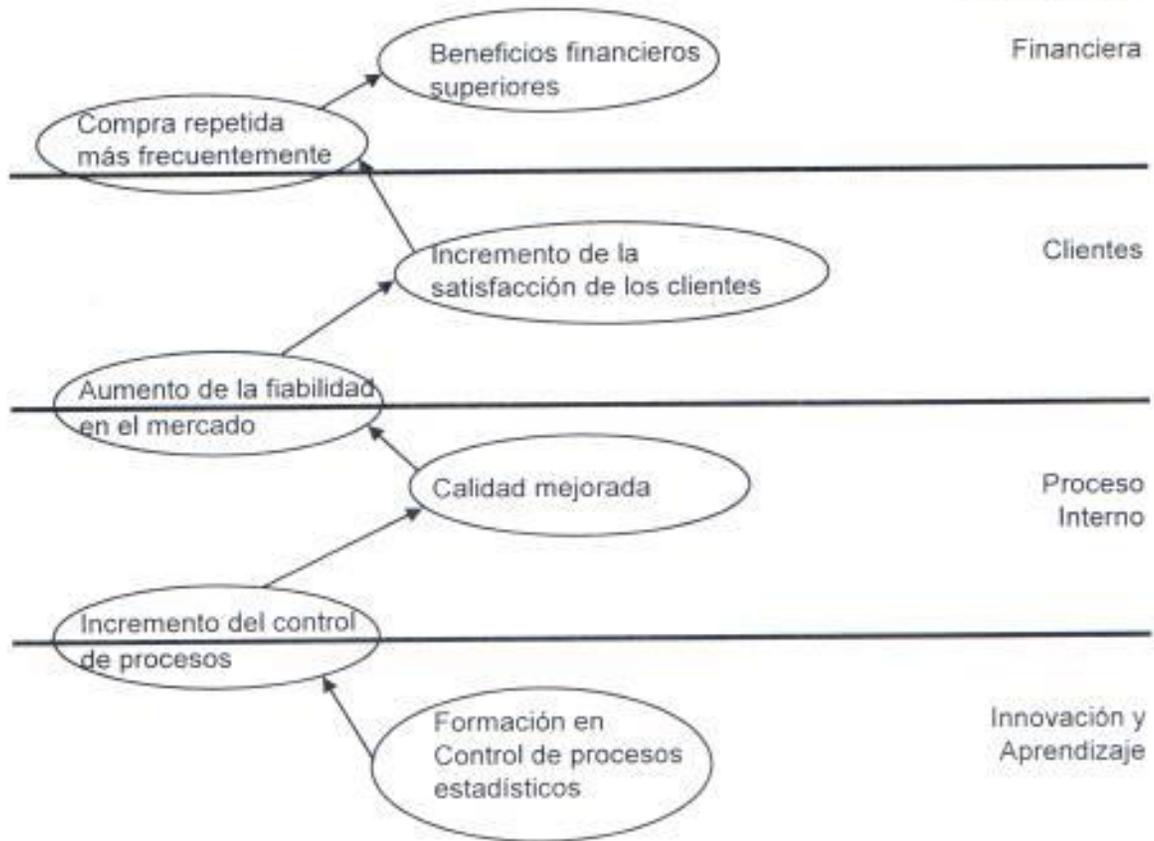
Si la empresa alcanza los objetivos de servicio al cliente, se incrementará su valor, y así lo reflejarán los indicadores financieros.

Una correcta ejecución de los procesos internos, y una capacidad de aprendizaje y crecimiento tienen como consecuencia una optimización del servicio al cliente.

Bajo esta técnica lo que se debe buscar son los impulsores principales del negocio, El enfoque "qué/cómo", ¿qué hay que lograr?, ¿cómo se debería conseguir?, Es para traducir un objetivo empresarial de alto nivel en una serie de sub-objetivos necesarios.

La siguiente figura muestra este esquema de causa y efecto sugerido por esta técnica del cuadro de mando integral de Kaplan y Norton.

## Perspectivas



Para este proyecto los dashboards se enfocaron en la perspectiva interna para analizar la capacidad de entrega, la emisión de facturas y ventas a clientes, la perspectiva del cliente y la perspectiva financiera.

#### 4.1.9 Desarrollo y herramientas del cuadro de mando integral.

Para el desarrollo de los Dashboards se usó el framework que viene en Penaho llamado CDF, "Community Dashboards Framework", este framework se basa en objetos java que son llamados utilizando javascript desde una plantilla html. Existe una herramienta no muy madura pero

funcional para el desarrollo de manera gráfica, pero es flexible hacerlos directamente con html y javacripts.

La desventaja de este esquema es que no se tiene una herramienta gráfica para hacer debug de errores, se tiene que usar sentencias como el "alert()" que envia mensajes al navegador y alli se pone el string que se envia al componente para identificar el problema.

Los Dashboards diseñados consumen cubos de información, no acceden a la base de datos directamente sino a través de un cubo de información. Se usan sentencias MDX para definir los queries a los cubos y mostrar la información requerida.

Estas sentecias MDX son enviadas utilizando javascript al componente CDF y este procesa la información. Estos componentes CDF permite relacionar dashboards para poder usar filtros de información y mostrar información relacionada entre cubos de información.

En el sitio de la comunidad de Pentaho existe información para la programación de dashboards utilizando CDF.

#### **4.3. Implementación del cuadro de mando integral**

##### **4.3.1. Estudio y análisis de habilidades del personal de la empresa**

Como parte de la metodología utilizada para asegurar el resultado del proyecto siempre se parte de identificar a los principales interesados, ya sea que estén a favor o en contra del proyecto, en toda organización hay

siempre resistencia a los cambios y es frecuente encontrarse con barreras para el desarrollo del proyecto.

Para el control de calidad siempre es importante contar con usuarios muy exigentes que nos aporten en la identificación de errores, los errores más comunes en este tipo de proyectos es el descuadre de información que en muchas ocasiones no se produce por la extracción de datos sino por consideraciones en los filtros de datos que se utiliza para consultar la información.

Por citar un ejemplo la empresa tiene registrado en sus ventas todas sus transacciones, pero las tiene clasificadas por ventas operativas y otras ventas, si las ventas son consultadas por un financiero le interesaría ver todas pero para un comercial solo las operativas, si no se hace el filtro adecuado verán descuadres de información que realmente no existe sino que el criterio no es el correcto.

Para estos filtros en los modelos de información se tiene que crear dimensiones que permitan hacer las respectivas selecciones dependiendo del perfil del usuario. En proyecto de inteligencia de negocios es muy sensible a la información si se tienen descuadres el usuario empezará a desconfiar y dejar de utilizarlo. Es muy importante tener muy claro todos los esquemas de cuadro de información y ejecutarlos durante la fase de control de calidad.

Para lograr una aceptación, la comunidad de usuarios debe formar parte del proceso de identificación del problema. Aun cuando la decisión de seguir con la implementación viene de la dirección, necesita la información sobre los requisitos de los usuarios, ya que puede representar la diferencia entre implementar una herramienta que funciona como propuesta de valor e implementar una que puede percibirse como inútil.

Una forma para contribuir al éxito de un proyecto es decidir cuándo se llevará a cabo la capacitación. Las iniciativas de capacitación deben empezar durante o antes de la fase de implementación. Sin embargo, en muchas organizaciones, inician meses antes de la implementación misma. En estos casos, los empleados se entusiasman con el sistema nuevo y todo lo que podrán hacer con él, pero cuando finalmente llega el momento de la implementación –algunas veces meses más tarde-, la excitación inicial y la aceptación han disminuido y, lo que es más importante, los usuarios han olvidado sus nuevas habilidades. Para volver a generar un impulso hay que repetir la capacitación –y gastar dinero y tiempo.

Nunca es fácil lograr que en una organización se acepten los cambios. Los usuarios se apegan a sus procesos actuales, aunque estos no necesariamente sean productivos. La aceptación no es algo que se da inmediatamente al mostrarles a los usuarios el valor inherente de BI, ya que cambiará toda la forma en que hacen negocios. Crear un programa de capacitación y ofrecerla a tiempo ayuda a los usuarios a aplicar sus nuevas habilidades inmediatamente y a que acepten mejor el sistema.

#### 4.3.2. Modelos de participación para guiar al usuario.

Para proyectos de BI es muy necesario utilizar talleres para promover la participación de los usuarios y contar con su criterio y aceptación de la herramienta.

El primer paso para asegurar el éxito de un proyecto es identificar el problema de negocios de BI. Cuando una empresa sabe perfectamente qué es lo que no funciona, puede no sólo empezar a buscar formas para corregirlo, sino identificar los recursos adecuados, generar una aceptación y asignar prioridades a las medidas que tomará para atacar el proyecto. Para generar un retorno de la inversión, la solución de BI debe corresponder con un problema de negocios de la organización; de lo contrario, implementar una herramienta de consultas particulares, un cubo de procesamiento analítico en línea (OLAP) o un tablero de control no será una medida que genere ventajas a largo plazo.

Los usuarios operativos conocen el día y viven la problemática de la extracción de información solicitada por sus mandos superiores, es por esto que es muy importante realizar talleres grupales y sacar las diversas perspectivas de información que tiene los usuarios y así se identificaran con el modelo desde el principio y se irán alineando las expectativas sobre la solución.

#### 4.3.3. Talleres de implementación

Una vez desarrollado los modelos identificados en conjunto con los usuarios se deben de hacer los talleres de implementación, para que revisen los formatos de los modelos y el cuadro respectivo de la información.

Esta revisión inicial debe hacerse con un equipo de usuarios que tendrán la función de hacer el control de calidad y detectar posibles fallas e inconsistencias en los modelos, cada observación debe ser registrada en una matriz y clasificar los diversos incidentes para su posterior revisión.

La matriz de incidentes debe registrar lo siguiente:

1. Tipo de objeto, cubo o dashboard
2. Descripción de la prueba
3. Objetivos de la prueba
4. Descripción del resultado de la prueba
5. Datos de entrada de la prueba
6. Salida esperada de la prueba
7. Fecha del incidente
8. Usuario que reporta
9. Responsable de la solución
10. Estado del incidente, resuelto, pendiente, en proceso, en control de calidad
11. Imagen del error.

Este registro ayudará a revisar cada incidente y se arman los respectivos planes de correcciones.

Una vez que la solución esté libre de errores se hace la entrega por cada módulo a los usuarios respectivos y se les da una capacitación enfocada al uso del modelo, las diversas opciones del navegador entre otros.

De esta forma el proyecto cumple el siguiente ciclo:

- Levantamiento de requerimientos, Talleres con usuarios para definir el alcance
- Análisis de requerimientos, talleres con personal de sistemas del cliente para identificar que es posible y que no ir en los cubos ya que los cubos deben leer un almacén de datos que extrae información del sistema transaccional
- Desarrollo de cubos y dashboards identificados con usuarios y definidos los alcances con el personal de sistemas.
- Control de calidad de los modelos, inicio de los talleres de implementación y control de calidad con los usuarios designados a esta tarea.
- Entrega y cierre, se entregan los modelos y se firman acta de entrega.

En lo posible todos estos talleres deben realizarse con la herramienta, esto ayudará a que el usuario tenga expectativas reales de las herramientas

#### 4.3.4. Capacitación de las herramientas.

Las herramientas a capacitar utilizando a Pentaho como solución de BI son:

1. Capacitar sobre las plantillas y formatos a utilizar durante el proyecto
2. Inducción sobre la herramienta de Extracción llamada PDI (Pentaho

Data Integration) de Pentaho, esto se hace con el personal de sistemas que tiene a cargo la generación de queries y entrega de información, la información de los clientes es confidencial y es mejor manejar esta parte con el personal interno y alineados con la metodología

3. Inducción sobre la herramienta de diseño de cubos PSW (Pentaho Schema Workbench), esta herramienta sirve para el desarrollo de los cubos de información, de igual forma es recomendable hacer los talleres con el personal técnico para que estén claros en los conceptos y diseños aplicados en los diversos cubos desarrollados para que así puedan dar mantenimiento.
4. Inducción sobre la herramienta de diseños de dashboards, que básicamente consiste en cómo utilizar los diversos componentes del framework CDF de Pentaho utilizando javascript y html
5. Inducción de la herramienta front-end, que es el visor de cubos para el usuario final y de dashboards. Con esta herramienta el usuario podrá consultar cubos y crear nuevas vistas sobre los modelos creados.

#### **4.4. Cubos y modelos bases para el cuadro de mando integral**

#### **4.5. Pruebas y control de calidad**

##### **4.5.1. Metodología de pruebas.**

La metodología utilizada consiste en:

1. Talleres con usuarios
2. Registro de preguntas y consultas ya sea de parte del cliente o del proveedor, este registro consiste en grabar lo siguiente:

- a. Campo ID, para identificar la pregunta
  - b. Campo Responsable, para registrar el responsable de la respuesta
  - c. Campo para tipo de pregunta,
  - d. Campo para la pregunta, solicitud o actividad pendiente
  - e. Campo para el seguimiento en donde se registra la última actividad
  - f. Campo para el estado de la pregunta, para ver si está pendiente o no
  - g. Fecha de registro de la pregunta
  - h. Fecha de la última actividad
  - i. Fecha de cierre de la pregunta
3. Registro de incidentes en los que se registra:
- a. Tipo de objeto, cubo o dashboard
  - b. Descripción de la prueba
  - c. Objetivos de la prueba
  - d. Descripción del resultado de la prueba
  - e. Datos de entrada de la prueba
  - f. Salida esperada de la prueba
  - g. Fecha del incidente
  - h. Usuario que reporta
  - i. Responsable de la solución
  - j. Estado del incidente, resuelto, pendiente, en proceso, en control de calidad

k. Imagen del error.

Con estos formatos se facilitará el registro y control de calidad, es importante tener mucha disciplina en esto para no perder el respectivo control y seguimiento.

Con esta información registrada en estos formatos se definen y dimensionan las respectivas correcciones a realizar.

No es necesario esperar al final del proyecto las pruebas y cuadros de datos puede hacerse desde la extracción de datos, como el cliente proporciona la información desde allí es necesario hacer los cuadros para evitar tener una cantidad acumuladas de correcciones al final.

#### 4.5.2. Casos de pruebas

Los casos de pruebas que se definan son obtenidos de los reportes operativos del cliente. Por esta razón se debe contar con un inventario de reportes del cliente para poder cuadrar la información, el datawarehouse debe cuadrar con todos los reportes colectados para certificar la solución.

Hacer bien el proceso de extracción no garantiza que cuadre la información ya que pueden existir reportes que usen criterios diversos para analizar la información y el modelo de datos debe de contener todos estos filtros en dimensiones para poder crear las diversas perspectivas de análisis.

Estos reportes operativos definirán los casos de pruebas para poder cuadrar la información durante el proceso de control de calidad.

Para citar casos se pueden presentar la situación que un reporte de ventas

para el área financiera le interese solo los que tienen clasificados como cuentas contables de ingreso, sin embargo los datos de ventas pueden contener todas las ventas incluso ventas no operativas que para el comercial le puede ser de interés y al financiero no, y para emitir el informe cada uno emitirá un criterio de selección diferente.

#### 4.5.3. Calidad de la información del almacén de datos

Un sistema de Inteligencia de Negocios es muy sensible a la calidad de información, si el usuario del sistema observa inconsistencias empezará a desconfiar de la solución, el proceso de control de calidad para el almacén de datos debe ser muy preciso, el equipo de trabajo para este proceso debe ser el siguiente:

1. Líder de las pruebas, este rol define los casos de pruebas y se encarga de inventariar los reportes operativos con los cuales se cuadrará la información, adicional identificará inconsistencias en las dimensiones producto de algún mal ingreso de datos en el respectivo sistema transaccional, por ejemplo el dato de nacimiento de un cliente muchos sistemas transaccionales no lo validan y el operario ingresa cualquier valor, pero en el almacén de datos si existe un análisis basado en la edad de los clientes se mostrará información inconsistente en este dato. Adicional a lo descrito este rol se encargará de controlar y priorizar los incidentes registrados para su respectiva solución.
2. Usuario de las pruebas, este rol se encarga de ejecutar y validar cada modelo de información, este rol debe ser personal del cliente que

domine el proceso en el que se basó el modelo de información o cubo de información. Al ser personal del cliente y dominar el respectivo proceso tendrá un criterio que aporte a encontrar las inconsistencias al modelo.

3. Técnico de las pruebas, este rol se encarga de resolver las inconsistencias encontradas en el modelo, dependiendo del incidente podrá ser resuelto por el proveedor o en conjunto con personal técnico del cliente.

Estos roles ayudarán a llevar un proceso de control de calidad preciso y que garantice la calidad de datos del almacén.

El tipo de datawarehouse utilizado es rolap por lo que la solución a incidentes se basa en corregir y emitir las sentencias sql adecuadamente.

#### 4.5.4. Talleres de pruebas con el usuario

Para realizar el proceso de control de calidad para el almacén de datos se debe clasificar las tareas en dos grupos:

- Un grupo para cuadrar los datos y sentencias sql emitidas por el cliente, si esta entrega de información es inconsistente el modelo será inconsistente y es básico pasar a la siguiente fase con usuarios con las sentencias SQL y datos cuadrados a nivel de registros y tablas de la base de datos, es decir que si un ETL que arma un modelo de ventas emite una tabla de hechos de ventas esta tabla de hechos debe de cuadrar con cada una de las tablas fuentes de donde se origina la información, como la información es transformada a pesar de la

transformación los valores cuantitativos deben de cuadrar, para ello se hacen sumas de los registros extraídos y se hace la suma de los valores en los registros fuentes y estos deben de cuadrar. Si se cuadra a este nivel se podrá pasar a la segunda fase de cuadro sin problemas.

- El segundo grupo de tareas consiste en revisar los datos del almacén pero usando los cubos de información y mostrándole a los usuarios, para esto es necesario realizar talleres de capacitación de la herramienta para que el usuario sea capaz de poder interactuar con la solución y realizar el respectivo proceso de cuadro, si garantizamos el primer grupo de tareas en este segundo grupo los descuadre que suelen presentarse se deben a diferentes criterios para emitir los filtros de información.

De esta forma se lleva a cabo el proceso de control de calidad con usuarios mediante talleres, esto ayuda incluso a que el usuario se familiarice con la solución desarrollada y cuando se libere luego de certificar los datos los usuarios ya estarán capacitados.

#### 4.5.5. Certificación de la solución

La solución se la certifica una vez que se tenga listo y cuadrado todos los modelos, es importante llevar un registro de cada incidente y el respectivo control de las tareas realizadas para su corrección.

En el anexo B se muestra un formulario para el proceso de certificación de la solución que consiste en un formulario que detallar el modelo cada una

de sus dimensiones, medidas con un indicador que muestre si la medida o dimensión cumple con lo exigido por el cliente.

Si alguna dimensión o medida no satisface ya sea porque el sistema transaccional no registra ese dato o por cualquier otro motivo se debe registrar el respectivo resultado y una observación para tener presente este incidente.

Por ejemplo el cliente puede solicitar mostrar rentabilidad por proveedor, para ello es necesario contar con la dimensión proveedores y la medida de rentabilidad, a nivel de estructura del cubo de información se lo puede desarrollar pero si el sistema transaccional no lleva control de lotes para indicar a que proveedor se compró determinado producto este valor no tendría sentido a pesar de que el cubo de información lo tenga, y se debe de registrar en el formulario de aceptación al cliente que a pesar de tener el cubo esta información el sistema transaccional no lo registra y ya queda a criterio del cliente mejorar este proceso.

## CAPITULO 5

### **5. Modelo de negocio para la venta del servicio sobre la plataforma creada.**

#### **5.1. Estudio de mercado.**

El mercado sujeto a estudio es el mercado ecuatoriano, las grandes soluciones de Inteligencia de negocios solo han llegado a unas cuantas empresas grandes pero las medianas y pequeñas no.

Sin embargo las medianas empresas y algunas pequeñas dependiendo del nivel de preparación de sus administradores se han convencido que están sobre mucha información y que no saben qué hacer con ella y que sus decisiones las toman sin un verdadero sustento de información.

El costo elevado de las soluciones licencias para inteligencia de negocios hace que las empresas no inviertan en soluciones de este tipo y sigan trabajando en realizar informes de manera manual.

Otra situación que se da es que los administradores saben que necesitan de una solución de este tipo pero que su personal no tiene la cultura para utilizar estas

herramientas y piensan que el comprarlas les resultará costoso para el provecho que le sacarán sus usuarios.

Todas estas situaciones descritas llevan a la conclusión de que existe la necesidad de contar con una herramienta de Inteligencia de negocios cuyo costo inicial no impacte al negocio y que paulatinamente conforme madure la cultura de los usuarios de la organización se justifique conforme a su necesidad los costos respectivos.

Por esto se ve una oportunidad de captar todas esas medianas empresas que quieren soluciones de Inteligencia de negocios y que cuentan con bajo presupuesto, con Pentaho utilizando la versión de la comunidad se evita costos de licencias y se puede generar utilidad de las consultorías para construir el almacén de datos, el diseño de cubos y dashboards.

## **5.2. Tácticas de mercadeo (4P)**

Las 4P se refiere a:

- Producto
- Precio
- Plaza
- Promoción

Para nuestro producto basado en esas tácticas la estrategia es:

Producto, el producto basado en esta solución open source Pentaho se lo puede desarrollar de las siguientes formas:

1. Instalar la solución las instalaciones del cliente, el cliente debe adquirir los equipos para instalar el servicio.
2. Instalar la solución en algún servicio hosting adquirido por el cliente, de esta forma se evita adquirir equipos

Precio, al ser un producto open source se evita pagar licenciamiento haciendo que el costo de la solución baje considerablemente ya que el costo se reduce al trabajo de consultorias y horas de trabajo que lleve la construcción del almacén de datos, el desarrollo de los cubos de información y de los dashboards. Contando con estos beneficios el precio que se puede ofrecer con competidores que ofrecen productos licenciados es considerablemente bajo y rentable por las horas de desarrollo, control de calidad e implementación.

El producto posee un nivel de complejidad que se requiere tener dominio sobre ingeniería que hace que el cliente necesite de soporte y atención.

Plaza, inicialmente debido a las limitaciones del equipo de trabajo la venta que se haga será directa, a futuro conforme la solución gane madurez se hará un producto empaquetado con los modelos y dashboards que contengan análisis estadísticos predefinidos que sean diferenciadores de otras herramientas del mercado, contado con este producto empaquetado se podrá hacer distribuciones utilizando canales y llegar a otros países.

Promoción, la promoción del servicio se lo hará de utilizando los siguientes medios:

- a) Venta personal: definida como la presentación personal de los servicios en una conversación con uno o más futuros compradores con el propósito de hacer ventas.
- b) Relaciones Públicas: definida como la estimulación no personal de demanda para un servicio obteniendo noticias comercialmente importantes acerca de éste, en cualquier medio u obteniendo su presentación favorable en algún medio que no esté pagado por el patrocinador del servicio.
- c) Promoción de ventas: actividades de marketing distintas a la publicidad, venta personal y relaciones públicas que estimulan las compras de los clientes y el uso y mejora de efectividad del distribuidor.

De esta forma se podrá ir desarrollando los diversos servicios de asesoría y a la vez ir madurando una solución que puede desarrollarse diversas versiones para una industria o sector de mercado.

Para no crecer en infraestructura de venta y madurar un producto comercial conforme se desarrollen para diversos sectores o industrias se tiene pensado empezar con pequeñas y medianas empresas que pueda soportarse con una estructura de 4 a 5 personas.

Cada cliente nos deja la experiencia de modelos y estructuras de tablas que sirven para el desarrollo del datawarehouse de futuros clientes. Los negocios sin importar su modelo tienen procesos comunes que pueden ser estandarizados en estructuras o modelos de datos de un datawarehouse.

### 5.3. Servicios ofrecidos por la plataforma de inteligencia de negocios

#### 4.1.1 Diseño, creación y mantenimiento del almacén de datos.

Este servicio consiste en lo siguiente:

1. Levantamiento de información en base a reportes usados por la organización
2. Talleres con usuarios para investigar las diferentes visiones y perspectivas de la organización
3. Diseño de DataWareHouse con información levantada
4. Desarrollo de procesos de extracción y transformación ETL
5. Construcción y validación del DataWareHouse
6. Control de calidad de datos procesados

Con estos servicios el cliente contará con un DataWareHouse para montar sobre el mismo cualquier herramienta de inteligencia de negocios.

El DataWareHouse es lo más crítico en una solución de inteligencia de negocios, ya que este brindará la información para los diversos modelos de datos, dashboards e indicadores que se implementen.

El diseño y desarrollo del DataWarehouse suele llevar aproximadamente el 40% del tiempo de un proyecto de inteligencia de negocios.

#### 4.1.2 Desarrollo del Portal Web.

Sobre el portal web lo que se diseña es la imagen corporativa del cliente, diseño web, logos, misión, visión.

La herramienta trae una pantalla inicial que puede ser diseñada para el

cliente la siguiente figura muestra la pantalla inicial de la herramienta Pentaho.



Ilustración 15 Pantalla de Login de Pentaho

Esta pantalla puede modificarse y aplicarse diseño gráfico para adaptarlo a una imagen corporativa del cliente.

La pantalla front end para navegar en los cubos y dashboards desarrollados no tiene mucho que hacer a más de cambiar logos ya que está orientada a visualizar cubos y dashboards y el diseño gráfico no primar sobre el orden y la facilidad del manejo para navegar en los cubos.

La siguiente figura muestra la pantalla front end de Pentaho para la navegación de cubos.



desempeño de la base de datos y cuando se necesita leer cantidades mucho más grande de datos se deben aplicar ciertas técnicas llamadas agregaciones que harán que las consultas sean rápidas incluso sobre millones de registros.

Una vez afinada la base de datos que contiene el almacén de datos, los cubos de información tendrán buen resultados en su ejecución y será buena su aceptación ante los usuarios.

#### 4.1.4 Diseño y Desarrollo de cuadros de mando integral

El diseño y desarrollo de los cuadros de mando integral es la parte más laboriosa de un proyecto de Inteligencia de negocios.

Esta labor parte de los talleres con usuarios, en donde se toman perspectivas individuales, perspectivas grupales y acuerdos colectivos que definen el comportamiento de la organización y la cultura de información de los usuarios.

El fin de estos talleres es determinar la visión del negocio que tienen los diversos directivos y funcionarios de la organización y esa visión plasmarla en un pantalla que contenga de manera resumida los indicadores más importantes que muestran la situación del negocio.

La forma como los usuarios quieren ver los indicadores hace que no se ajustes a una sola plantilla estándar sino que se tenga que diseñar y desarrollar dashboards a la medida.

Sobre Pentaho estos dashboards se los desarrolla usando los componentes para gráficos que trae la herramienta, pero no es tan sencillo ya que tiene mucho codificación en javascript de por medio, y esto se hace sobre una

programación web en la que el manejo de eventos sobre los objetos de una página web no es tan flexible.

**5.4. Servicios basados en consultorías de planificación estratégica que se genere por los modelos de inteligencia de negocios creados.**

Una vez concluido el proyecto de Inteligencia de negocios y los usuarios se familiaricen con el uso y la explotación de la información para la toma de decisiones se puede brindar servicios adicionales a los clientes que se basen en planificación estratégica.

Madurar la cultura de información hace que se cambie la forma de pensar del personal de la organización y les hará más sentido el enfoque estratégico, direccionar el trabajo en base a indicadores que se los pueda medir con frecuencia y analizar el resultado de las decisiones.

Se puede aportar con este estudio que cuando en las corporaciones se hacen planificaciones estratégicas y la organización no tiene madura la cultura de organización existe una brecha entre los objetivos estratégicos de los directivos y el desarrollo del día a día de los procesos operativos.

Por esta razón el roadmap para el desarrollo estratégico de una corporación puede empezar con modelos básicos que muestren el resultado de la operación y mediante indicadores bien estructurados ir enlazando con la visión estratégica de los directivos.

El desarrollo de la planificación estratégica y el mapeo de los procesos de negocios con la tecnología pueden dar lugar a muchos servicios de asesorías y consultorías que pueden ser muy rentables.

Muchos profesionales que asesoran en planificación estratégica dejan indicadores y dashboards en Excel que son complejos de mantener y utilizar, pero con un portal con cubos, dashboards e indicadores bien estructurados y automatizados puede hacer mucho más fácil el desarrollo estratégico de los clientes.

# CONCLUSIONES Y RECOMENDACIONES

Como conclusiones se aporta con lo siguiente:

1. En el mercado ecuatoriano medianas empresas acercándose a ser grandes no toman decisiones en base a información resumida o estadística basada en indicadores, la mayoría solo se basa en reportes operativos que de manera aislada no presentan una verdadera situación del negocio.
2. Es posible con un proceso de extracción de datos y correctamente automatizado reducir la generación de informes por parte del personal de 4 horas a segundos brindando facilidades a la organización y aliviando de carga de trabajo y reproceso de información innecesaria a sus colaboradores
3. Un sistema de inteligencia de negocios no cubre los vacíos de información de los sistemas transaccionales de la operativa diaria, es un error tratar de cubrir informes operativos con cubos de información, los cubos de información o dashboards están orientados a información estadística resumida que en base de indicadores con información consolidada brindan un análisis más amplio de la información.
4. Se puede aprovechar las funcionalidades disponibles en las herramientas

de open source para desarrollar la cultura de información en las organizaciones a un bajo costo rompiendo la barrera de costos altos que generalmente incurre una solución de inteligencia de negocios del mercado. Al romper esta barrera el desarrollo de modelos y la facilidad que brinda el análisis multidimensional para cambiar dinámicamente los criterios y perspectivas de la información cambian la manera de pensar de sus directivos y funcionarios para llevarlos a otro nivel de análisis que los haga pensar en cómo direccionar el negocio en base al comportamiento histórico de la compañía, comportamiento que está registrado en sus diversas bases de datos o fuentes de información que la organización utiliza.

5. Una planificación estratégica no puede obtener los resultados esperados si existe una disociación entre la parte operativa y la visión de los directivos, si existe esta brecha los objetivos y metas no estarán alineados con los procesos del día a día incluso los sistemas no capturan los datos de control necesarios para sustentar los diversos indicadores de negocio que servirán de guía para direccionar el negocio de manera oportuna, con decisiones correctas que harán aumentar su rentabilidad.

Entre las recomendaciones se tiene:

1. Seguimiento a la comunidad de Pentaho para estar actualizados con los nuevos desarrollos que se hagan sobre la plataforma, cada día se desarrollan nuevas funcionalidades que facilitarán la administración y desarrollo de proyectos sobre esta plataforma.

2. Desarrollar versiones orientadas a diversos sectores o industrias para poder contar con productos empaquetados que puedan ser distribuidos por canales que permitan atender mayor cantidad de clientes no solo en el mercado ecuatoriano sino extenderse a la región.
3. Mejorar constantemente las plantillas necesarias para el registro de información que sirva para documentar correctamente el datawarehouse y todos los cubos y modelos construidos sobre Él.
4. Como recomendación final hay que ir midiendo el desarrollo que tengan los clientes sobre la versión opensource de pentaho, si el cliente madura su cultura de información y necesite análisis más avanzados necesitará herramientas que las tiene la versión empresarial de pentaho, esta versión empresarial es licenciada, a pesar de ser licenciada su costo no está a la altura de otras herramientas comerciales como cognos, Oracle entre otros, la versión empresarial utiliza el mismo motor de cubos pero contiene herramientas que facilitan el análisis y brinda herramientas adicionales que ofrece mayor funcionalidad, la comunidad se encuentra desarrollando alguna de estas herramientas pero aún s nivel de madurez no supera a la versión empresarial.

# BIBLIOGRAFÍA

<http://www.capiq.org.ec/otrosServicios/bibliotecaVirtual/index.html>

<http://www.capiq.org.ec/sdePymes/index.html>

<http://www.pentaho.com>

<http://www.msdn.com>

## ANEXOS

## ANEXO A

LEVANTAMIENTO DE MEDIDAS Y DIMENSIONES PARA MODELO DE INFORMACIÓN		
Nombre del Modelo		
Fecha: _____	Documentado por: _____	
Medidas de Hecho		
Nombre Medida	Descripción	Formula
DIMENSIONES		
Nombre dimensión	DESCRIPCIÓN	
CAMPOS		
Nombre dimensión	DESCRIPCIÓN	
CAMPOS		
Nombre dimensión	CLIENTES	
CAMPOS	DESCRIPCIÓN	
Nombre dimensión	DESCRIPCIÓN	
CAMPOS		
COMENTARIOS		

