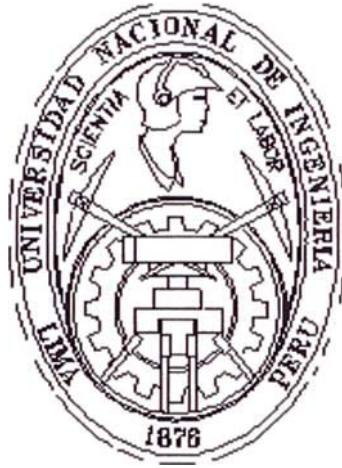


**UNIVERSIDAD NACIONAL DE INGENIERIA  
FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS  
ESPECIALIDAD DE INGENIERIA INDUSTRIAL**



**OPTIMIZACIÓN DEL PROCESO DE EXTRACCIÓN DE  
INFORMACIÓN DEL DATA WAREHOUSE**

**INFORME DE SUFICIENCIA  
PARA OPTAR EL TITULO PROFESIONAL DE:  
INGENIERO INDUSTRIAL  
NILTON CESAR MORI LEÓN**

**LIMA, PERU**

**2006**

## **AGRADECIMIENTO**

A mis padres por inculcar en mí el deseo de superación, por enseñarme a valorar lo mejor de la vida y a luchar por alcanzar las metas propuestas.

A mi esposa Maria Ysabel, por acompañarme en los momentos más difíciles.

A mi pequeña hija Silvana Isabel, por enseñarme a ver lo simple de la vida.

## INDICE

<b>OPTIMIZACIÓN DEL PROCESO DE EXTRACCIÓN DE INFORMACIÓN DEL DATA WAREHOUSE .....</b>	<b>1</b>
<b>AGRADECIMIENTO .....</b>	<b>2</b>
<b>DESCRIPTORES TEMATICOS .....</b>	<b>4</b>
<b>RESUMEN EJECUTIVO .....</b>	<b>5</b>
<b>INTRODUCCION.....</b>	<b>6</b>
<b>CAPITULO I.....</b>	<b>7</b>
OBJETIVOS GENERALES .....	7
PRINCIPIOS ORGANIZACIONALES.....	7
DIAGNÓSTICO ESTRATÉGICO .....	8
ANÁLISIS FODA: .....	8
DIAGNÓSTICO FUNCIONAL .....	10
<b>CAPITULO II MARCO TEORICO.....</b>	<b>13</b>
<b>CAPITULO III METODOLOGIA DE TRABAJO.....</b>	<b>25</b>
1 SITUACIÓN ACTUAL.....	27
1.1 ARQUITECTURA DEL DATA WAREHOUSE .....	27
1.1.1 DIAGRAMA ESTRUCTURAL PLATAFORMA DE BUSINESS INTELLIGENCE ..	27
1.1.2 DISEÑO DE LA ARQUITECTURA DE BUSINESS INTELLIGENCE.....	28
1.2 PROCESO DE OBTENCIÓN DE DATA.....	29
1.2.1 PROCESOS DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS ..	31
1.3 ORÍGENES DE DATOS .....	32
1.3.1 SISTEMAS OLTP Y FUENTES DE DATOS.....	32
1.3.2 MAPA DE BUSINESS INTELLIGENCE .....	33
2. DESARROLLO DEL REQUERIMIENTO .....	36
2.1 LA GENERACIÓN DE VALOR POR MEDIO DEL ANÁLISIS DE DATOS .....	36
2.2 ANÁLISIS GAP .....	36
2.2.1 ANÁLISIS GAP DE ARQUITECTURA .....	36
2.2.2 ANÁLISIS GAP DE MAPA DE BUSINESS INTELLIGENCE.....	38
3 PRESENTACIÓN DE RESULTADOS Y PLAN DE ACCIÓN .....	41
3.1 ARQUITECTURA DE BUSINESS INTELLIGENCE .....	41
3.2 ARQUITECTURA DE DATOS CORPORATIVA PARA CLARO.....	50
3.3 MAPA DE BUSINESS INTELLIGENCE .....	54
3.4 PROCESOS DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS (ETL) ..	55
<b>CAPITULO IV CONCLUSIONES Y RECOMENDACIONES .....</b>	<b>57</b>
<b>GLOSARIO DE TERMINOS.....</b>	<b>60</b>
<b>BIBLIOGRAFIA.....</b>	<b>61</b>

## **DESCRIPTORES TEMATICOS**

- Arquitectura de Data Warehouse, Kimball y Corporate Information Factory.
- Análisis GAP de Arquitectura y Del Mapa de Business Intelligence.
- Business Intelligence, Inteligencia de Negocios.
- Business Data Storage y Data Marts.
- Datacleansing en ambientes de Business Intelligence.
- Mapa de Business Intelligence.
- Procesos de extracción, transformación y carga de datos.
- Principios básicos de la Arquitectura de Business Intelligence.

## **RESUMEN EJECUTIVO**

CLARO es una empresa que brinda servicios de comunicación basados en la tecnología GSM (Sistema Global Comunicaciones Móviles), en el contexto actual en el que CLARO se encuentra se requiere contar con información oportuna y confiable que contribuyan a la toma de decisiones de los directivos así mismo al seguimiento de las acciones de la organización.

Se necesita conocer al cliente en su totalidad, anticiparse a su demanda, conocer todos los procesos a fin de ser optimizados o determinar la rentabilidad de clientes, productos y canales.

CLARO es una empresa que busca potenciar más el uso de la Inteligencia de Negocios en la organización, lo que supone una expansión tecnológica, que consistiría en la correcta explotación del Data Warehouse.

Para lograr este objetivo se requiere la participación e integración de los usuarios finales de las áreas de Administración y Finanzas, Comercial, Marketing y Tecnología de Información (Business Intelligence - BI).

## **INTRODUCCION**

CLARO (América Móvil) es una empresa que brinda servicios de telecomunicaciones basados en tecnología GSM (Sistema Global Comunicaciones Móviles), la fuerte competencia en este sector hace que se necesite contar con información oportuna y confiable que contribuyan a la toma de decisiones y al seguimiento de acciones de la organización.

CLARO es una empresa que busca potenciar el uso de la Inteligencia de Negocios, lo que supone una expansión tecnológica que consistirá en la correcta explotación del Data Warehouse. Actualmente cuenta con un Data Warehouse corporativo que se encarga de proveer datos a diferentes áreas usuarias, información que se genera en cada transacción realizada con los cliente, socios y con el mundo en general. El proyecto realizado se inicia en función a las prioridades del negocio y al objetivo trazado.

La necesidad actual es la de contar con información confiable proveniente del Data Warehouse (DW) sin tener que realizar operaciones adicionales, obtener nuevos reportes generados según criterios particulares de cada área, para esto necesitamos generar mayor valor agregado a la data y así obtener información, esto es uno de los objetivos principales que permitirán maximizar el uso del Data Warehouse.

Este documento a sido desarrollado con las mejores metodologías de implantación de Arquitecturas de Data Warehouse en empresas de telecomunicaciones, estas metodologías se basan en la recolección de información de la situación actual del objeto de estudio, y la fusión de las experiencias encontradas en esta fase con los modelos de arquitectura mundiales, en las industria de telecomunicaciones.

## **CAPITULO I**

### **OBJETIVOS GENERALES**

Nuestro objetivo es garantizar la calidad de la información provista desde el Data Warehouse DW, para llegar a este punto debemos de analizar la distribución de su arquitectura y el proceso actual de extracción de datos, se buscará la optimización de procesos apoyados en las mejores practicas del medio para así generar valor agregado en la obtención de información proveniente del Data Warehouse, esto facilitará la obtención de información, generará confianza y seguridad en la información, permitirá contar con ventajas competitivas, desarrollar profunda y estratégicamente el entorno del Business Intelligence en la empresa.

### **PRINCIPIOS ORGANIZACIONALES**

#### **Nuestra Visión**

Ser la primera operadora de telecomunicaciones móviles en el Perú.

#### **Nuestra Misión**

Alcanzar el liderazgo en el mercado nacional de telefonía móvil con el fin de exceder los objetivos financieros y de crecimiento de nuestros accionistas.

Para lograr estos objetivos debemos anticiparnos a la necesidades de nuestro clientes y superar sus expectativas de servicio. Generando bienestar y desarrollo profesional de quienes lo hacemos posible. Siempre vinculados al desarrollo de nuestra comunidad.

## Nuestros Valores

CLARO se distingue en el medio por los valores que continuación se detallan:

- ✓ Honestidad
- ✓ Manos en la Operación
- ✓ Actitud de Servicio
- ✓ Orden y Disciplina
- ✓ Ahorro

## DIAGNÓSTICO ESTRATÉGICO

América Móvil es el operador de telecomunicaciones líder en Latinoamérica y en el Perú ofrece su servicio a más de 3 millones de clientes. América Móvil se encuentra entre las 10 primeras operadoras más importantes a nivel mundial y cuenta con más de 100 millones de suscriptores siendo la empresa de telefonía móvil líder en los países en donde brinda servicios.

### Análisis FODA:

<b>FORTALEZA</b> <ul style="list-style-type: none"><li>✓ Gran respaldo y reconocimiento latino americano.</li><li>✓ Cuenta con fuerte apoyo económico que le permite invertir en nuevos sectores.</li><li>✓ Cuenta con tecnología de última generación GSM.</li></ul>	<b>DEBILIDAD</b> <ul style="list-style-type: none"><li>× Esta ingresando como una nueva marca a un mercado altamente monopolizado.</li><li>× Tiene cobertura a nivel nacional, pero solo en las principales ciudades.</li></ul>
<b>OPORTUNIDAD</b> <ul style="list-style-type: none"><li>✓ Actual insatisfacción por lo servicios de post venta brindados por la competencia.</li><li>✓ Potencial segmento de mercado que</li></ul>	<b>AMENAZA</b> <ul style="list-style-type: none"><li>× Posible incursión de nuevos operadores.</li><li>× Fuerte competencia de precios y servicios.</li></ul>



aun no ha sido explotado.	* Riesgos políticos relacionados con la libre competencia.
---------------------------	--

**Fortaleza:**

- Empresa líder que brinda servicios de telefonía móvil a nivel mundial.
- Cuenta con una buena imagen corporativa.
- Cuenta con gran respaldo y reconocimiento mundial, tiene un fuerte apoyo económico financiero externo, es parte del grupo América Móvil.
- Empresa que ha establecido el liderazgo en la región latinoamericana LATAM, logrando desplazar a su principal competidor Telefónica.
- El servicio brindado es realizado con tecnología GSM, tecnología de última generación que brinda seguridad y facilidad de uso al cliente.

**Debilidades:**

- Esta ingresando con una nueva marca, en mercado altamente monopolizado, donde las reglas de juego no favorecen a la libre competencia.
- Se encuentra en pleno proceso de captación del mercado, compro una empresa establecida pero con limitaciones de cobertura.
- Actualmente llega a todos los departamentos del Perú pero no cuenta con total penetración del mercado.

**Oportunidades:**

- Insatisfacción de los usuarios de la competencia por el trato recibido en los servicios post venta brindados por su principal competidor.
- Las nuevas normas que permiten una apertura de mercado presenta la oportunidad de extender su segmento de mercado, lo que antes por un tema de costo no han sido cubiertos por los operadores.

**Amenazas:**

- La licitación de una nueva banda de comunicaciones permitiría el ingreso de otro operador potencial.
- Fuerte competencia en precios y servicios de la competencia como estrategia de comercialización.

**DIAGNÓSTICO FUNCIONAL****Productos**

CLARO ofrece servicios de comunicación móvil, y dependiendo de la unidad de negocio se brindan servicios para los diferentes segmentos de mercado.

**CLARO Empresas:** brinda servicios de transmisión de datos:

- ✓ Servicios de telefonía móvil y mensajería, además de servicios de automatización y gestión empresarial tales como:
  - Asignación de ventas o pedidos
  - Asignación de rutas de entrega diaria
  - Consultas de clientes, de stocks, de precios, etc.
- ✓ Servicio de navegación en Internet
- ✓ Monitoreo de equipos a distancia vía GPRS, seguridad satelital.
- ✓ Servicio de Internet, OMA, SMS, mail.

**CLARO:** Ofrece además de servicios adicionales como:

- ✓ Servicio de Localización a través de SMS.
- ✓ Servicios de Internet vía GPRS.

**Clientes**

Claro es una empresa subsidiaria de América Móvil que cuenta con más de 100 millones de abonados a nivel mundial y con más de 3 millones de abonados a nivel nacional.

## **Proveedores**

América móvil tiene entre sus principales proveedores:

A nivel de equipos celulares se tiene

- Nokia
- Motorola
- Siemens
- Sonny Ericsson
- Sagem
- Samsung, Etc.

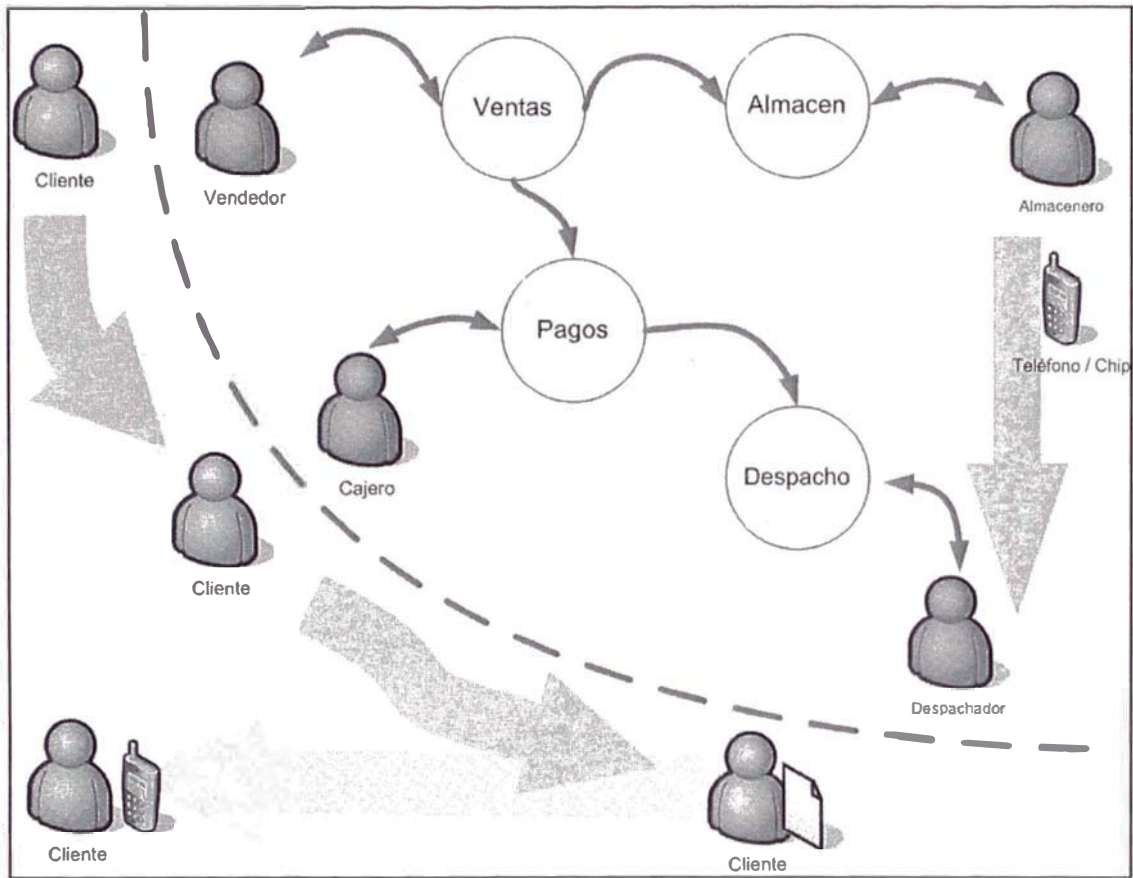
A nivel de seguridad vía GSM se tiene a Segursat (Seguridad Satelital)

## **Procesos**

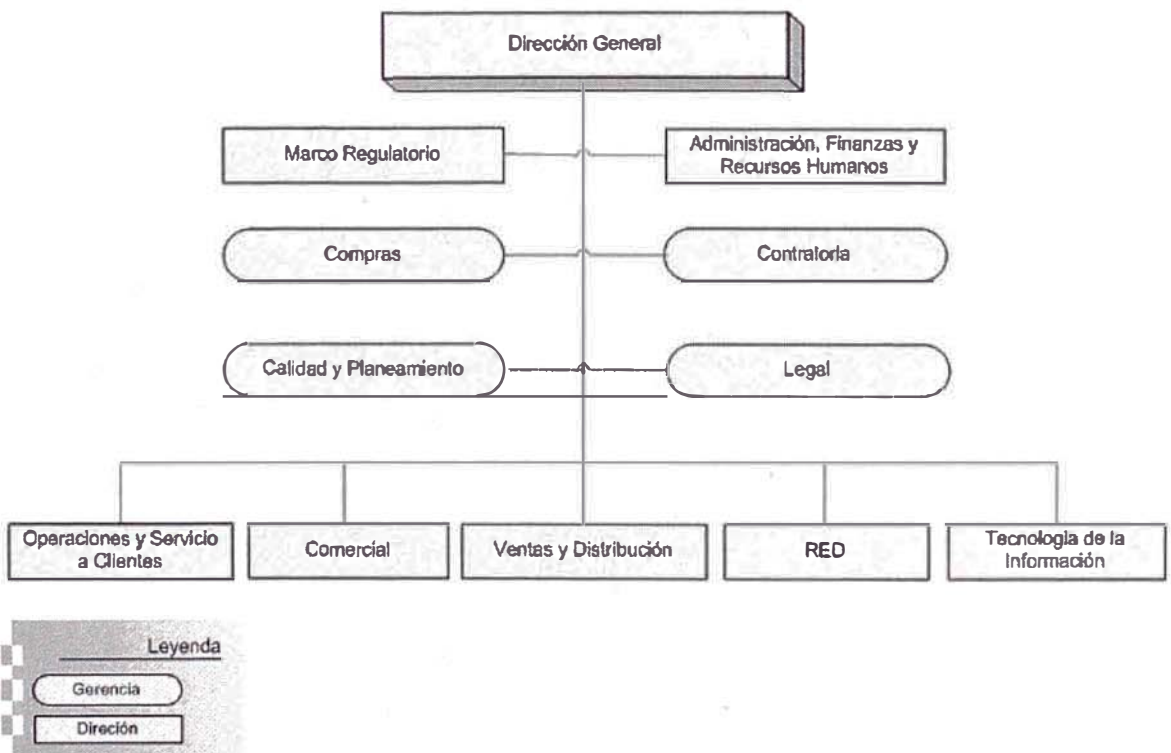
Los procesos que realiza la empresa son:

- Adquisición de bienes y servicios
- Ventas en servicio prepago y postpago
- Servicios de post venta (renovación, reposición y migración).
- Venta de recargas virtuales y físicas en puntos de venta
- Servicio técnico en puntos especializados
- Retención y fidelización en puntos de venta especializados
- Ventas a clientes business
- Lanzamiento productos y servicios
- Medición de la satisfacción del cliente externo

En el siguiente grafico se muestra el procedimiento del proceso de venta, esto es realizado en los centros de atención al cliente (CAC).



### Organización de la empresa



## **CAPITULO II**

### **MARCO TEORICO**

Las empresas de hoy día se mueven bajo un entorno en el que la información juega un rol muy importante, pudiéndose definir incluso como crítico al generar ventajas competitivas, esta situación se presenta más fuerte aun en el sector de las telecomunicaciones.

Para poder generar ventajas competitiva debemos contar con data en tiempo real y altamente confiable, que brinde información de los clientes, los recursos humanos y financieros que se tienen. Debe contribuir en la toma de decisiones y apoyar en la ejecución de acciones estratégicas.

Nos encontramos en un entorno en el que los clientes son muy exigentes, la competencia tradicional es mas agresiva y las no tradicionales empiezan a romper sus paradigmas, generando mayor presión, necesitamos conocer muy bien a los clientes, anticiparse a su demanda, mejorar la calidad de servicio desde la pre venta hasta la post venta, debemos poder determinar la rentabilidad de nuestro clientes, productos y/o canales de venta, conocer toda esta información permitirá generar ventajas competitivas.

La mejor herramienta que nos garantizara contar con esta información es el Data Warehouse (DW), convirtiéndose así en una aplicación estratégica y clave en los procesos de estudio de mercado y toma de decisiones. Para garantizar el correcto funcionamiento del DW debemos revisar los procesos involucrados, desde el modelo de datos, los procesos de extracción, hasta la arquitectura actual del data warehouse.

Vamos a describir los principales procesos a considerar para garantizar la confiabilidad de la obtención de la data, así como el correcto funcionamiento de la arquitectura implantada.

### ✓ **Procesos de Extracción y Carga**

Los procesos de los sistemas OLTP son apoyados en formatos de carga y transformación, que indican para cada dato que cambios van sometidos hasta llegar a la fase final de la Arquitectura.

Las fases de un Proceso de Extracción y Carga son:

#### **1. Selección de Datos**

Este proceso identifica los sujetos, tablas y columnas objetivo en los sistemas OLTP. Son objetivo porque son cargadas una sola vez al momento del desarrollo y luego ocasionalmente.

#### **2. Filtrado**

Es la primera etapa del proceso continuo de extracción y carga. Se selecciona la información identificada como objetivo y elimina la información que no tiene relevancia necesaria para ser cargada.

#### **3. Depuración**

Aquí se corrige o elimina la información errónea según criterios de calidad definida por la organización.

#### **4. Homogenización**

Proceso de normalización de información según códigos y estándares definidos por la organización.

#### **5. Consolidación**

Aquí se unifican las tablas que se estén disociadas en aplicaciones OLTP.

#### **6. Agregación**

Proceso que genera tablas resumen de tablas de hechos detalladas para mejorar la performance.

## 7. Derivación

Precálculo de datos a partir de otros; ejemplo obtener precio total según cantidad y precio unitario.

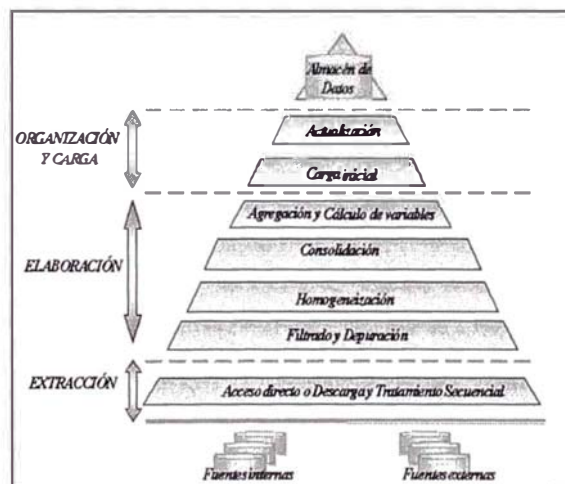
## 8. Modelamiento Temático

Es simplemente la transformación de la información desde el modelo BDS al necesario para el DM.

## 9. Denormalización

Procesos construidos con herramientas especializadas llamadas ETL's (Extract Transform and Load).

Pirámide de los Procesos de Extracción y Carga:



### ✓ ***Datacleansing en ambientes de Business Intelligence BI***

Gran parte de los sistemas OLTP presentan deficiencias de calidad en la información que manejan, esta calidad se mueve en un espectro amplio que va desde la peor calidad hasta una de mejor calidad.

### **Datacleansing**

Conjunto de técnicas que detectan problemas de calidad en la información a cargar en el DW, elimina la información y genera alertas para su corrección.

### **Ámbitos de la corrección de data**

En la corrección de data existen los siguientes ámbitos:

- **Corrección del Valor del Dato:** Corrige contenidos no adecuados o de baja calidad. Es realizada en la Arquitectura de BI o sistemas OLTP, implica recargar el Data Warehouse, se realiza en problemas reiterativos y en algunos casos cuando son opcionales.
- **Corrección de las Causas:** Corrige procesos, programas o módulos de sistemas OLTP que producen data de mala calidad, es un proceso complicado y se aplica a problemas reiterativos.

### Métodos de Datacleansing

Existen tres métodos de datacleansing cada uno ventajas y desventajas, el más recomendable dependerá de las características del ambiente de trabajo y la fiabilidad de los sistemas OLTP.

#### 1. Método Preventivo

Elimina registros de mala calidad, si no cumplen con las reglas son descartados. Durante la extracción y carga existe una etapa llamada de depuración donde se aplica el método de datacleansing.

**Depuración:** Proceso que elimina información sin calidad adecuada para la carga. Los criterios se definen en bandas de calidad o márgenes de tolerancia. Una banda de calidad del 1% será válida si las filas malas no exceden el 1% total, sino se descarta o en ruta al proceso de datacleansing.

#### Ventajas y Desventajas del Método Preventivo

<b>Ventajas</b>	<b>Desventajas</b>
No permite data de mala calidad en el modelo	Hace más lento el Proceso de Extracción y Carga, por los reprocesos de las tablas rechazadas totalmente
Focalizada en los errores de calidad de la data	Puede generar desfase entre los períodos de carga de las tablas (no todas son cargadas al mismo tiempo)



Analiza las bandas de calidad para cada tabla del sistema.	El Proceso de Extracción y Carga es más complejo, por contemplar las bandas de calidad.
--	---

## 2. Método Reactivo

Permite cargar data de mala calidad, con cargo a corregirse en procesos de limpieza posteriores. Para realizar la carga sin afectar el modelo lo primero que se hace es desactivar los constraints de base de datos, excepto los de clave foránea y segundo para las inconsistencias generadas reemplazar el dato errado por uno estándar según la tabla de configuración.

La siguiente figura muestra un ejemplo de ambas variantes:

(A)				(B)			
Antes del Datacleansing				Antes del Datacleansing			
Transacciones		Movimientos		Transacciones		Movimientos	
Cod	Descrip	Trx	Monto	Cod	Descrip	Trx	Monto
D	Depósito	T	89.65	D	Depósito	T	89.65
R	Retiro	M	145.75	R	Retiro	M	145.75
Después del Datacleansing				Después del Datacleansing			
Transacciones		Movimientos		Transacciones		Movimientos	
Cod	Descrip	Trx	Monto	Cod	Descrip	Trx	Monto
D	Depósito	X	89.65	D	Depósito	T	89.65
R	Retiro	X	145.75	R	Retiro	M	145.75
X	Trx. NN			T	Trx. NN		
				M	Trx. NN		

Formas de Asegurar Integridad Referencial en M. Reactivo de Datacleansing

Las filas no se marcan como erradas, es el proceso masivo de limpieza el que identifica y corrige.

### Ventajas y Desventajas del Método Reactivo

Ventajas	Desventajas
El Proceso de Extracción y Carga es rápido.	Ingresa data de mala calidad al Data Warehouse.
La información entra sincronizada en el Data Warehouse.	Los problemas de mala calidad de información pueden quedar subyacentes por mucho tiempo.

<b>Ventajas</b>	<b>Desventajas</b>
El Proceso de Extracción y Carga es simple.	No se define ni se ven de manera integrada los rangos de tolerancia de data de mala calidad.

### 3. Método Mix

Combina los métodos anteriores, los datos que ingresan a la Arquitectura de BI son marcados con un flag a nivel de registro que los indica como dato erróneo. Con el método de reporting y monitoreo se realiza el seguimiento a la data de mala calidad.

En la carga se utiliza el método Reactivo, se deshabilitan los constraints excepto los de clave foránea y se colocan marcas en los registros.

Existen dos métodos, **el global** que enciende un flag en las filas que presenten errores y **el de precisión** que escribe en un campo texto la posición del dato errado seguido de puntos y comas.

La siguiente figura muestra un ejemplo de colocación de marcas.

**Dada la tabla Movimientos donde:**

- TRX debe existir en la tabla Transacciones
- Monto no puede ser nulo
- Estado sólo puede tener los valores "A" e "I".
- Hora de Inicio no puede ser mayor que Hora de Fin.
- \* QA es la columna de marca para calidad de data.

**Antes del Datacleansing**

Transacciones		Movimientos					
Cod	Descrip	Trx	Monto	Est	Hr Ini	Hr Fin	QA
D	Depósito	R	89.65	A	08:50	08:45	
R	Retiro	M		Z	15:50	15:53	

**Después del Datacleansing**

Transacciones		Movimientos					
Cod	Descrip	Trx	Monto	Est	Hr Ini	Hr Fin	QA
D	Depósito	R	89.65	G	08:50	08:45	3 ; 4
R	Retiro	X		Z	15:50	15:53	1 ; 2 ; 3
X	Trx NN						

Es recomendable tener un Tablero de Control dentro de la Arquitectura de BI. Donde se registren las tablas con su banda de calidad, luego por proceso

de conteo se actualiza el número de marcas de QA y por comparación se encienden semáforos (Rojo donde se supere la banda de calidad).

<b>Tablero de Comando Datacleasing</b>					
Elemento	Banda de Tolerancia	Nro Total de Regs	Nro Máx de Regs para Banda	Nro. de Regs con QA en Tabla	Semáforo
Data Warehouse	5%	19,835,460	991,773	1,586,987	
<b>Tablero de Comando Datacleasing: Nivel Data Warehouse</b>					
Elemento	Banda de Tolerancia	Nro Total de Regs	Nro Máx de Regs para Banda	Nro. de Regs con QA en Tabla	Semáforo
DM de Clientes	5%	19,835,460	991,773	1,586,987	
DM de Ctrl de Gestión	2%	5,365,493	107,310	5,690	
DM de Riesgos	10%	10,630,500	1,063,050	1,062,998	z
<b>Tablero de Comando Datacleasing: Nivel Datamart – DM de Clientes</b>					
Elemento	Banda de Tolerancia	Nro Total de Regs	Nro Máx de Regs para Banda	Nro. De Regs con QA en Tabla	Semáforo
Clientes	0%	65,891	0	0	
Convenios	0%	23	0	5	
Solicitudes	1%	87,368	874	1,200	
Grupos Familiares	10%	1,500	150	20	
Vtas_Mes	5%	7,638,994	381,950	382,000	
Solicitudes_Mes	3%	12,361	371	1,560	

#### ✓ **Homologación Contable**

Este proceso no es tratado regularmente, se mantiene a nivel teórico, pero se considera fundamental para el éxito de una Arquitectura de BI, es valido sólo por la calidad y utilidad de la información que aporta.

En toda organización lo que define a un dato como correcto o no, es su homologación con la información contable, por ello es fundamental desarrollar procesos de homologación de información con la contabilidad, lo cual de por sí puede representar complejidad extrema dependiendo de la empresa y el grado de sincronización entre los sistemas transaccionales y la contabilidad

#### ✓ **Principios Básicos de la Arquitectura de Business Intelligence**

Al realizar el diseño de toda arquitectura de Business Intelligence se debe respetar el cumplir con cuatro principios básicos, como son:

**1. Estratificación.** Basado en capas en las que la data “viaja” según el grado de depuración presentado.

**2. Independencia.** Cada una de las capas del modelo es independiente de la anterior y posterior. Puede fungir como un modelo autónomo.

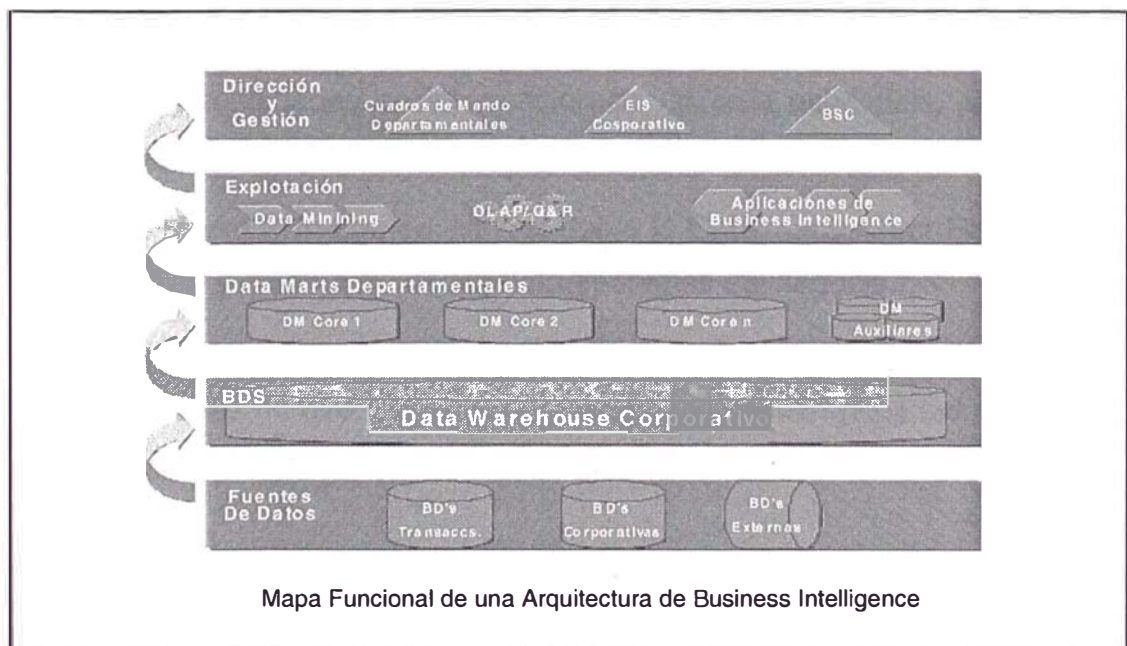
**3. Integridad.** El diseño debe ser coherente en definición y significado de datos. Un dato de una capa inferior debe ser encontrado con el mismo valor en la otra o existir una regla definida para derivarlo.

**4. Especialización.** Cada capa tiene una finalidad definida dentro del modelo, si bien puede cumplir funciones complementarias su objetivo es único y específico.

Además las arquitecturas están fuertemente basadas en dos frentes, la Estructura de datos y el Proceso de Extracción y Carga.

✓ **Arquitectura de Desarrollo de Business Intelligence**

La Arquitectura de desarrollo viene constituido por cinco niveles como se muestra en la siguiente grafica:



Revisando cada nivel del mapa analizado se define lo siguiente:

### **Fuentes de Datos**

Desde aquí se alimenta de datos al Business Intelligence, se tiene tres tipos de fuentes de datos:

- ✓ **Bases de Datos Transaccionales**, Propias de los Sistemas Operacionales (OLTP) del negocio y donde se realizan las operaciones de la empresa.
- ✓ **Bases de Datos Corporativas**, Compartidas para enriquecer la información del negocio.
- ✓ **Base de Datos Externa**, Contiene información externa a la empresa y sirven para complementar información, establecer benchmarkings de la competencia y/o medidas del mercado en general.

### **Business Development Storage BDS**

Aquí se aloja el Data Warehouse Corporativo con las definiciones conceptuales homologadas, con las dimensiones de negocio definidas y con los contenidos de información depurados.

### **Data Marts Departamentales DM**

Soportan necesidades de información de un área específica, se tienen dos categorías de Data Marts:

- ✓ **Data Marts Principales**, Vinculados a la Cadena de Valor Real de la empresa y/o con actividades fundamentales del negocio.
- ✓ **Data Marts Auxiliares**, son aquellos formulados para apoyar actividades no directamente vinculadas con las centrales del negocio, aunque importantes para su desempeño.

### **Explotación**

Aquí se realiza la explotación de la información y compondrán la capa de Decisión Support Systems, aquí reside toda la potencia de análisis de información. Se tiene tres tipos de aplicaciones:



- ✓ **Aplicaciones de Business Intelligence BI**, Tienen un proceso de negocio incorporado y desarrollado. Similares a sistemas operacionales, pero con mayor capacidad de análisis.
- ✓ **Aplicaciones OLAP**, Son de análisis de datos, permiten explorar la información desde distintos niveles de granularidad y con distintas dimensiones de análisis.
- ✓ **Modelos de Data Mining**, Estadísticas que permiten encontrar patrones ocultos en grandes masas de datos. Aquí se utilizan modelos de clusterización, ranking, extrapolación y previsión.

### **Dirección y Gestión**

Administran los indicadores de gestión y los umbrales de tolerancia. Dirigida a los administradores de la organización, su principal característica es tener un único valor y ser significativos para establecer el nivel de cumplimiento del proceso. Existen tres tipos de aplicaciones:

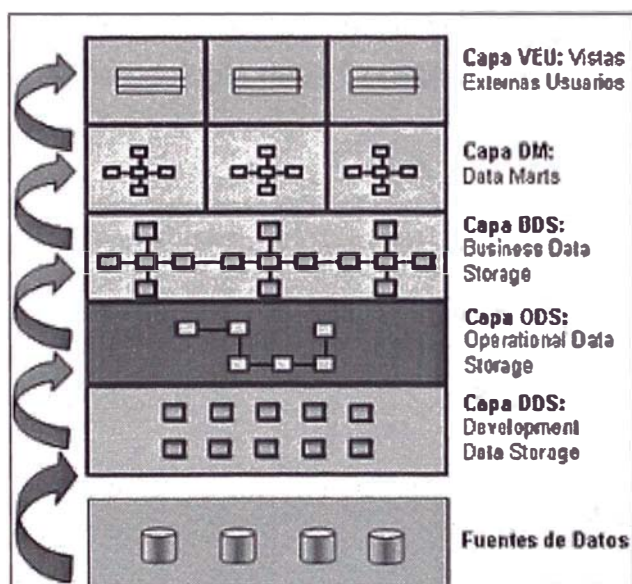
- **Tableros de Control o Cuadros de Mando Departamentales**, Son de seguimiento de indicadores pero que están circunscritas a un área organizacional o proceso de negocio específico.
  - **Executive Information System (EIS) Corporativo**, Relacionada con el juego de indicadores de desempeño de la empresa, recoge parte de los indicadores del Tablero de Control Departamental.
  - **Balanced Scorecard (BSC)**, Recoge el subconjunto de indicadores del EIS relacionados directamente con estrategias del negocio, se encarga de seguir su implantación en cuatro perspectivas: Financiera, Clientes, Procesos Internos e Innovación y Crecimiento.
- ✓ **Arquitectura del Data Warehouse**

Este concepto es fundamental ya que la Arquitectura es la que soportara todos los procesos del DW, es por ello que una arquitectura debe ser correctamente definida para garantizar su confiabilidad.

En el medio se tienen varias metodologías desarrolladas para las empresas de telecomunicaciones, siendo la más importante la **Arquitectura Kimball** formulada por Ralph Kimball, y la **Corporate Information Factory** de Hill Inmon con menor aceptación en el mercado.

**LA ARQUITECTURA KIMBALL**, Es la arquitectura predominante, ya que tiene más años en el mercado y muchas implantaciones exitosas.

En el grafico adjunto se observa la estructura de desarrollo de una Arquitectura Kimball.



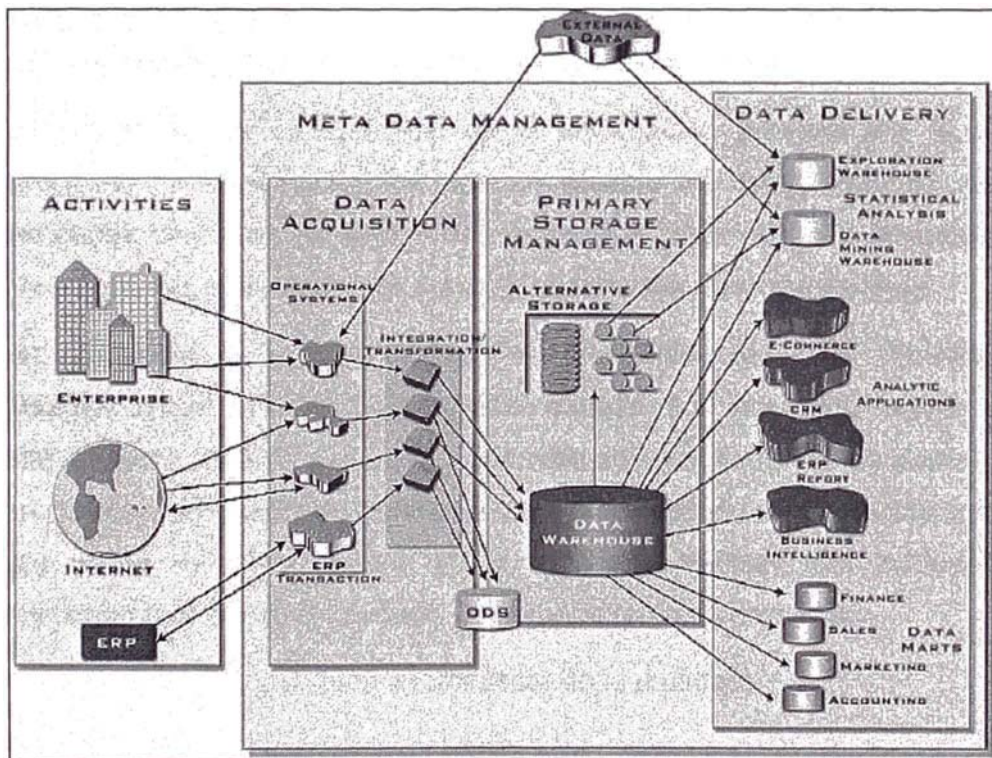
Veamos una relación de las principales ventajas y desventajas de la Arquitectura Kimball.

Ventajas	Desventajas
Arquitectura simple y sólida	Arquitectura con aspectos del Business Intelligence no cubiertos.
De construcción poco costosa	Limitada capacidad de crecimiento.
De administración poco costosa	No tiene una posición clara sobre la relación del Business Intelligence con la mega tendencia (CRM, e-Biz, etc.)
Éxito comprobado	Tiene una visión aislada de la arquitectura de Business

Ventajas	Desventajas
	Intelligence
	Es onerosa en almacenamiento

**La Corporate Information Factory**, es una arquitectura joven en el mercado con pocas implantaciones, algunas muy exitosas, la contundencia de sus definiciones ha permitido que vaya ganando espacio.

En el grafico adjunto se observa la estructura de desarrollo de la arquitectura con la metodología de la Corporate Information Factory.



Corporate Information Factory por Bill Inmon



## **CAPITULO III**

### **METODOLOGIA DE TRABAJO**

La metodología esta basada en la acumulación de información del estado de situación actual del objeto de estudio y la fusión de los hallazgos encontrados hasta esta fase con los Modelos de Clase Mundial, el análisis se realizo con el aporte de un proveedor que cuentan con las mejores prácticas mundiales para la industria de telecomunicaciones.

Para garantizar la confiabilidad de la información provista por el data warehouse se debe de contar con un modelo de datos eficiente, tener adecuadas procesos de extracción, corrección, depuración y limpieza de datos, pero por sobre todo una correcta distribución de la arquitectura.

Para garantizar la confiabilidad de la información del Data Warehouse debemos seguir una metodología de análisis que garantice contar con estándares de análisis y diseño, Esta metodología viene siendo soportada con las mejores prácticas mundiales para empresas de telecomunicaciones, para definir nuestro alcance inicial debemos realizar tres actividades principales que nos darán la base para lo que queremos realizar:

1. Situación Actual (entendimiento de la realidad),
2. Desarrollo del Requerimiento,
3. Presentación de Resultados y plan de acción.

Nuestro primer objetivo es conocer la situación actual o real de la empresa, considerar las necesidades del usuario final y en que porcentaje se tiene actualmente cubierto por el Business Intelligence BI, para esto se requiere revisar dos aspectos básicos: el del usuario y la tecnológica.

En cuanto al *usuario*, se deben realizar entrevistas y/o aplicar cuestionarios a las áreas usuarias involucradas, aplicar encuestas de satisfacción que certifiquen la información obtenida.

Desde el punto de *vista tecnológico*, se revisará la arquitectura del Business Intelligence BI y el Data Warehousing DW, esto involucra modelos de datos, fuentes de información, procesos ETL (Extract, Transform, and Load), Metadatos técnicos y de negocio. Una vez revisado y analizado se realizará un acopio de información, y apoyados en las mejores practicas se validará la arquitectura actual a fin de generar la información final.

Luego de tener en claro la situación actual se plantea la mejor solución a aplicar, y se establecen las pautas que se involucraran en el desarrollo del requerimiento, luego se presentarán los resultados obtenidos para finalmente establecer una serie de recomendaciones y planes de acción par mejorar los procesos y validar la nueva arquitectura propuesta.

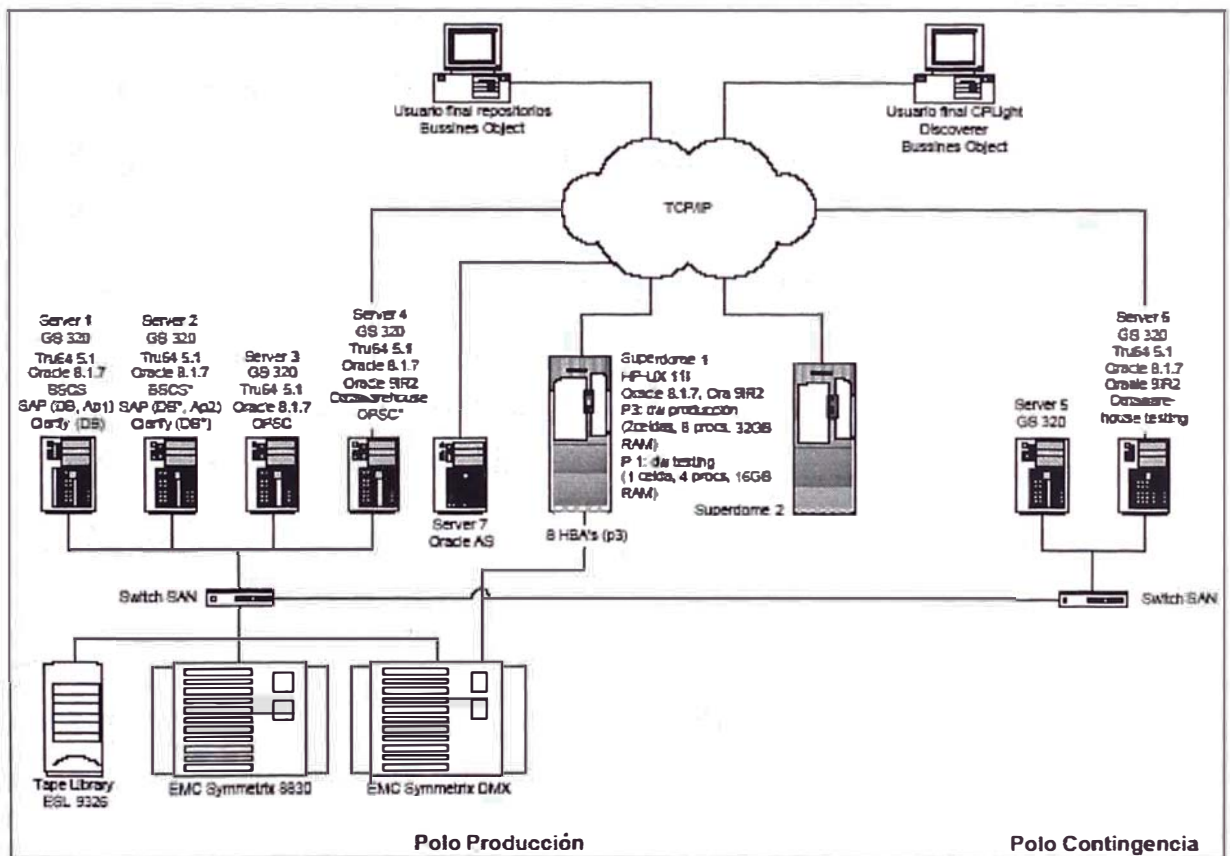
# 1. SITUACIÓN ACTUAL

## 1.1 Arquitectura del Data Warehouse

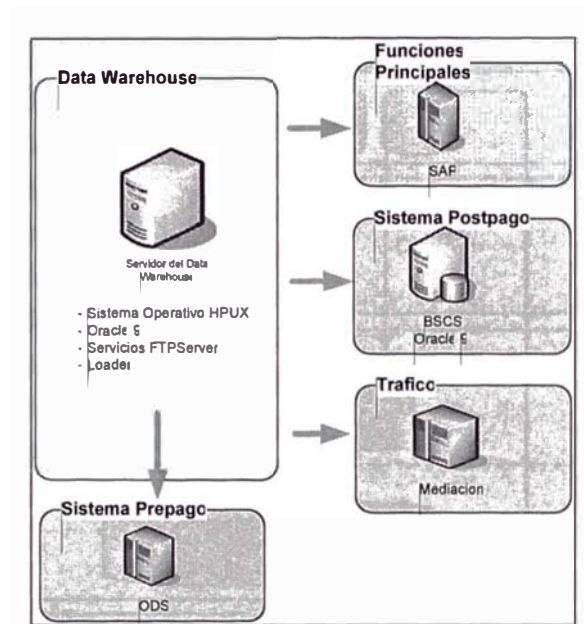
### 1.1.1. Diagrama Estructural de la Plataforma de Business Intelligence

La plataforma donde corre el Data Warehouse viene de ser migrada de la plataforma Alpha/Tru64 hacia la plataforma Superdome/HP-UX.

En el siguiente grafico se puede observar la plataforma actual del data warehouse (Superdome 1) junto con la plataforma antigua (Alpha/Tru64), como proceso previo a la migración de plataformas.



En el siguiente grafico se visualizan las principales interacciones del DW con los sistemas corporativos.

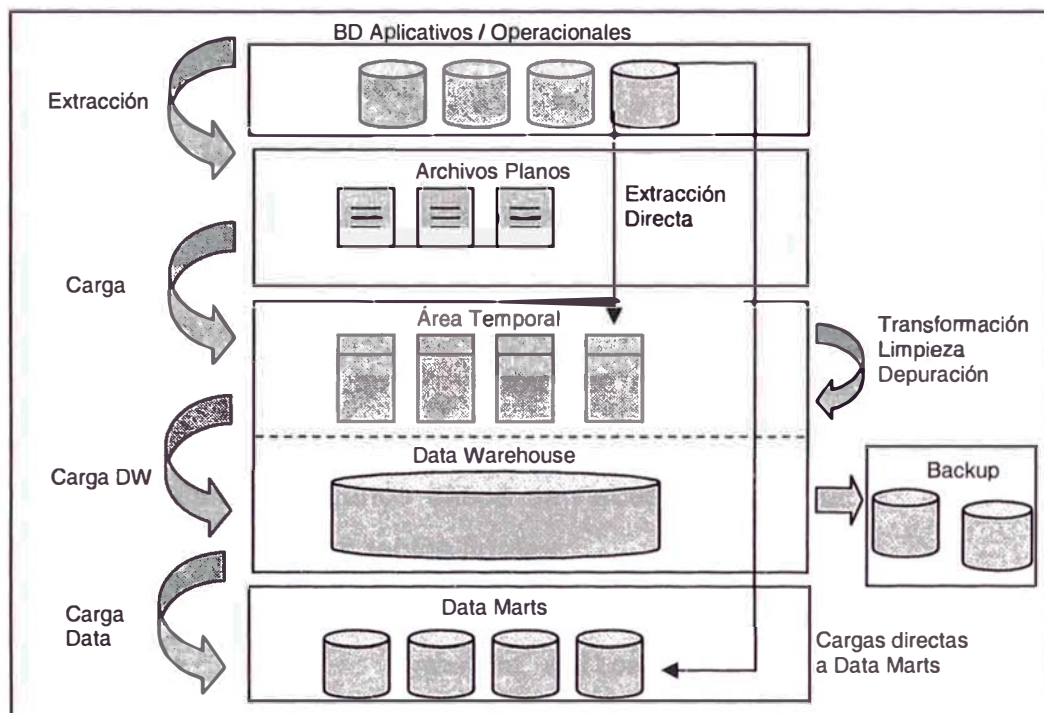


### 1.1.2. Diseño de la Arquitectura de Business Intelligence

CLARO presenta una Arquitectura de Business Intelligence de tres capas de datos

- ✓ La primera capa aloja a los archivos planos extraídos de los sistemas operacionales (OLTP).
- ✓ La segunda capa se ubican las tablas finales y temporales, sobre las que se realizan actividades de transformación, limpieza y depuración.
- ✓ En la tercera capa residen las tablas propias de los distintos Data Marts construidos en CLARO.

Existe también una capa semántica de usuario final construida con la herramienta Business Objects, en el gráfico adjunto se puede apreciar la arquitectura.



## 1.2 Proceso de Obtención de Data

Este proceso presenta grandes problemas, el Data Warehouse actual es mas un proveedor de datos que de información, obligando a las áreas usuarias a realizar procesos adicionales, apoyados en herramientas de oficina, de esta manera generan islas de información, convirtiendo al usuario en un proveedor de resultados para análisis y toma de decisiones y para los reportes de la organización.

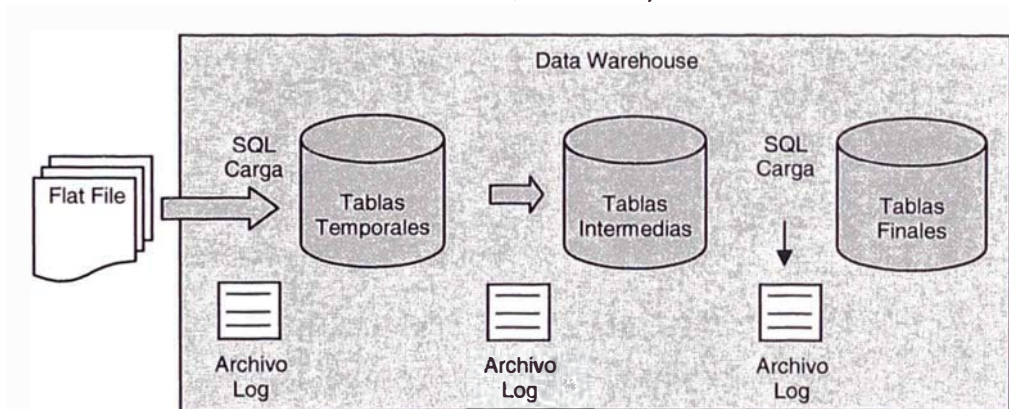
Los procesos manuales presentan problemas de exactitud en la información generada, como resultado de inadecuadas políticas de manejo de datos e información, trayendo como consecuencias:

- × Información costosa, se requiere procesar la información enviada del Data Warehouse.
- × Pérdida de tiempo en la elaboración de los informes y/o reportes.
- × Política centralizada en el procesamiento de datos y no en los procesos de negocio.

- × Se realizan muchas actividades no productivas involucradas en la obtención del resultado.

Los procesos de extracción son administrados por el área de Gestión de Datos. El método de carga comienza con la recepción de archivos planos desde las fuentes de datos, que son transferidos a tablas temporales dentro del ambiente donde reside el Data Warehouse, luego se realizan actividades de Transformación, Depuración y Limpieza sobre estas tablas temporales, para luego cargar los datos a través de stages desde las tablas temporales a las tablas finales, en algunos casos se necesita utilizar tablas intermedias para realizar la carga a las tablas finales.

El siguiente gráfico muestra el esquema de general de los procesos ETL (Extract, Transform, and Load)



CLARO cuenta con una herramienta de ejecución de procesos automáticos denominados CONTROL M, esta herramienta se encarga de la ejecución de procesos automáticos, existen procesos que se ejecutan en forma diaria (procesos automáticos) y mensual (procesos automáticos y procesos manuales), algunos de estos procesos han sido creados y otros reutilizados e importados a la herramienta Data Stage, los desarrollados con esta herramienta tienen un nivel óptimo de performance.

Dentro de los procesos ETL se tienen controles de carga, que validan: la cantidad máxima de errores que pueden suceder en una carga que provocan que el proceso aborte, integridad referencial, entre otros.

### 1.2.1 Procesos de Extracción, Transformación y Carga de Datos

- Los Procesos ETL que se desarrolla en el Data Warehouse cubre eficientemente las expectativas que se tienen dentro de la Arquitectura de Business Intelligence tanto en tiempo como proceso de carga.
- El objetivo del desarrollo de los procesos ETL es preparase para las necesidades futuras, desde ese punto de vista entendemos que ante un incremento de volúmenes de datos a cargar el actual proceso podría no cumplir con los niveles óptimos.
- Los Procesos ETL vienen siendo enfocados como un conjunto de programas que extraen, transforman y cargan data en el Data Warehouse, pero se debería considerar que el objetivo es convertirse en un Layer de Integración y Transformación de Datos (Layer I/T).
- El objetivo de estos procesos es el ser un Layer I/T que es un área de datos donde la data es integrada y transformada antes de convertirse en data corporativa. Este Layer I/T debe servir a todas las necesidades de la empresa, se basa en tres tareas específicas:
  - **Procesos de transformación conceptual de la data**, aquí se busca homologar conceptos, transformar el esquema de organización de la información para atender necesidades de negocio y unificar criterios de medición a nivel corporativo.
  - **Procesos de transformación técnica de los datos**, este es el punto más parecido a los actuales Procesos ETL, aquí se realiza la selección, filtrado, homogenización (técnica), reformateo, reordenamiento, consolidación y agregación de la información.
  - **Proceso de Datacleansing o Limpieza de Datos**, independizamos este punto de la parte técnica por la relevancia que tiene para la integridad de los datos. El Layer I/T es vital para asegurar la calidad de información que se distribuye. Es responsable de la difusión y no de calidad de data.



### 1.3 Orígenes de Datos

Los datos se extraídos de los principales sistemas corporativos de CLARO, con el fin de garantizar información proveniente de una fuente central de datos, es decir sin tratamiento de sistemas externos lo que podría incurrir en obtener información que no concuerda en concepto ni en fecha de extracción.

El no garantizar la integridad de la información podría genera problemas potenciales, por que la información podría ser obtenida bajo diferentes criterios de transformación de datos. Resultando de esta manera problemas de seguridad de información.

#### 1.3.1 Sistemas OLTP y Fuentes de Datos

Cerca del 90% de la información actual del Data Warehouse proviene de los aplicativos de tráfico (Sistema Prepago, Sistema Postpago y Redes), estos aplicativos tienen como motor de base de datos a Oracle.

Las principales fuentes de información que están siendo cargadas al Data Warehouse son:

Fuentes de Datos	Descripción
GFP	Aplicativo de PREPAGO
PPSTAT	Fuente de Datos de Clientes Prepago
BSCS	Aplicativo de Postpago
CLARIFY	Aplicativo CRM (Contactos, Reclamos, entre otros).
Mediación	Fuente de Datos CDR
Interconexión	Fuente de Datos de Interconexión
Segmentación	Fuente de Datos de Segmentación de Clientes
TOL	Aplicativo de Facturación



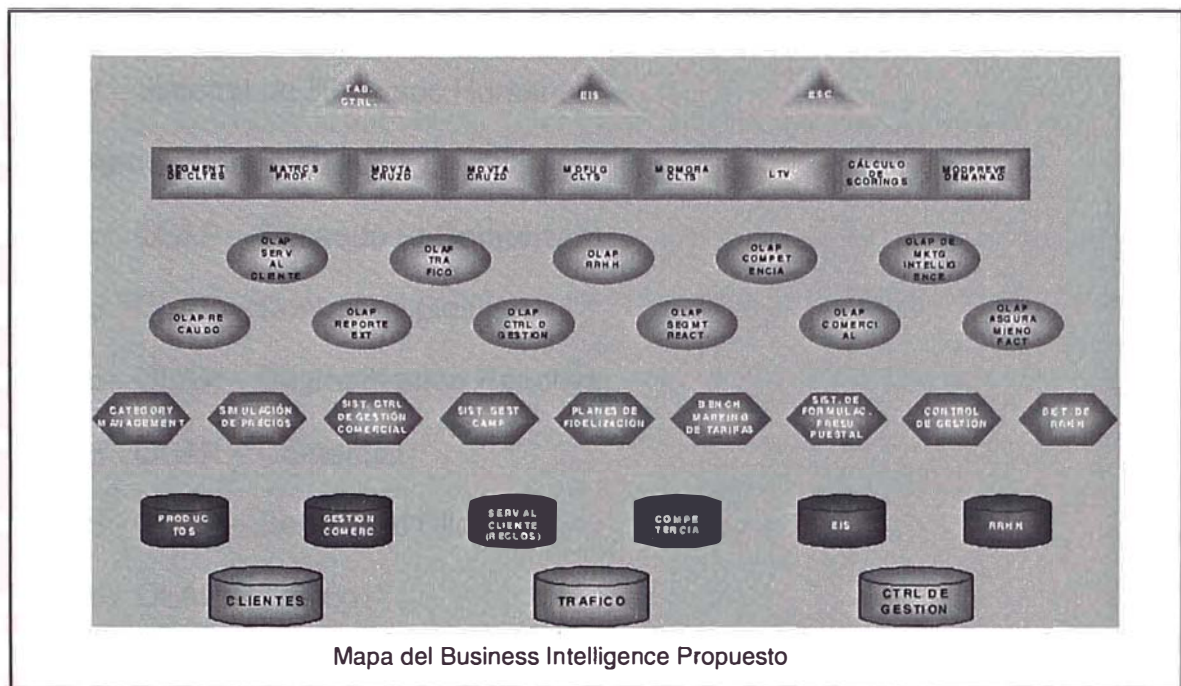
- La información de facturación (SAP) es cargada directamente al Data Mart RAPIDMART sin pasar por el Data Warehouse.
- La información de RED, GFP (PREPAGO) y BSCS (POSTPAGO) es cargada directamente al Data Mart Tráfico/Ventas CO-BI/Parque de Clientes.
- La información de Tráfico para el Data Mart es cargada directamente al Data Warehouse.
- El Data Mart SEGMENTACION recibe cargas directas desde el PPSTAT y de TOL para luego se cargado al Data Warehouse.
- La información del IVR-123, Portal Claro, Aplicaciones Intranet no son cargadas al Data Warehouse.

### **1.3.2 Mapa de Business Intelligence**

Del análisis realizado se describe el Mapa y el diseño actual de Business Intelligence de CLARO, esto nos ayudara a proponer un mapa, el cual se encuentra enfocado en cinco tipos de componentes:

1. Data Marts, que son áreas de datos modeladas para atender a una parte del negocio.
2. Aplicaciones Business Intelligence, desarrolladas explícitamente para procesos de BI.
3. Aplicaciones OLAP, son de consultas libres para usuarios finales.
4. Data Mining, modelos estadísticos utilizados de clasificación, puntuación o predicción de variables.
5. Cuadros de Mando, orientadas al seguimiento de indicadores de performance, se tienen los Tableros de Control, el Executive Information System (EIS) y el Balanced Scorecard (BSC).

A continuación detallamos cada uno de los componentes de la Arquitectura Funcional de Business Intelligence propuesta en la siguiente figura.



Se listan los elementos del Mapa que serán detallados posteriormente:

➤ **Data Marts Principales**

- Data Mart de Clientes
- Data Mart de Tráfico
- Data Mart de Control de Gestión

➤ **Data Marts Auxiliares**

- Data Mart de Productos
- Data Mart de Gestión Comercial
- Data Mart de Servicio al Cliente
- Data Mart de Competencia
- Data Mart de Executive Information System (EIS)

➤ **Aplicaciones de Business Intelligence**

- Category Management
- Sistema de Control de Gestión Comercial
- Sistema Analítico de Planes de Fidelización

- Sistema de Benchmarking de Tarifas
- Sistema de Recursos Humanos
- **Aplicaciones OLAP**
  - OLAP – Recaudo (o Cobranzas)
  - OLAP – Control de Gestión
  - OLAP – Segmentación Reactiva
  - OLAP – Comercial
  - OLAP – Servicio al Cliente
  - OLAP – Tráfico
- **Modelos de Data Mining**
  - Modelo de Segmentación de Clientes
  - Modelo de Productos
  - Modelo de Mora de Clientes
  - Modelo de Previsión de Fraude
  - Modelo de Previsión de la Demanda
- **Tableros de Control**
  - Tableros de Control Departamentales o Cuadros de Mando Departamentales
  - Executive Information System Corporativo
  - Balanced Scorecard (BSC)

## **2. DESARROLLO DEL REQUERIMIENTO**

Una vez revisada la situación actual del Data Warehouse tanto a nivel de arquitectura como de los procesos internos involucrados en la obtención de datos, debemos hacer un análisis de la situación actual que nos permita revisar las mejores alternativas de solución según la metodología definida inicialmente.

### **2.1. La Generación de Valor por medio del Análisis de Datos**

El análisis de esta información comprende lo siguiente

- ✓ El origen correcto de los datos y su actualización.
- ✓ El tratamiento sistemático y organizado de los datos (Modelo de Datos y Data Warehouse).
- ✓ El análisis especializado de los datos (Data Marts y Aplicaciones de Análisis de Datos).
- ✓ Las acciones tomadas a partir del análisis realizado.
- ✓ El valor generado por las acciones tomadas.

### **2.2. Análisis GAP**

Del análisis realizado tanto en la arquitectura como en la distribución (Mapa de Business Intelligence) actual de CLARO, se presentan cuadros comparativos de estos elementos con las aplicaciones desarrolladas respecto a las necesidades del negocio.

#### **2.2.1. Análisis GAP de Arquitectura**

Se muestra un análisis GAP de la arquitectura objetivo con la arquitectura actual para tener un alcance de lo que ya se tiene cubierto y lo que se tendría pendiente de implementar.

Arquitectura Objetivo	Arquitectura Actual	Análisis GAP
Development Data Storage	Existen	Se tiene adecuado manejo de archivos planos extraídos de los sistemas operacionales, son depositados en el mismo servidor de DW. Los archivos son cargados a tablas temporales en el mismo servidor, la cual cumple con la funcionalidad de la capa de datos DDS. Se realizan actividades de transformación, depuración y limpieza. También se encuentra la capa de datos del DW.
Operational Data Storage	No existe	No existe la capa de datos.
Almacenamientos Alternativos	Existen	Se tienen enfoques de almacenamientos alternativos de data con bajo nivel de uso. La política es almacenar data de 4 años en línea y tenerlas en cintas backup.
Business Development Storage	Data Warehouse	Es la más desarrollada en CLARO, aunque la información no es totalmente confiable. Solo los analistas de TI pueden accederla. En el mismo servidor se encuentra la capa de datos del área temporal.
Data Marts	Existen	En el servidor donde reside el DW, también residen los Data Marts, son accedidos por los usuarios finales. Algunos son cargados directamente desde las fuentes de datos operacionales.
Mining WareHouse	No existe	Aún no se ha desarrollado actividades de Data Mining, por lo cual no se cuenta con un área específica para esto.
Vistas Externas de Usuario	No existe	No existen, Con la herramienta Business Objects se han definido Universos que es un concepto

Arquitectura Objetivo	Arquitectura Actual	Análisis GAP
		bastantes cercano a las VEU's, pero no cumple completamente la función por depender de la herramienta.
Fuentes Externas	Existen	Cargan información de entes reguladores directamente a la capa de datos BDS.

### 2.2.2. Análisis GAP de Mapa de Business Intelligence

Se muestra un análisis GAP del Mapa de Business Intelligence actual para revisar lo que ya se tiene cubierto y lo que estaría pendiente de implementar.

Mapa de Business Intelligence	Status de Desarrollo	Análisis GAP
<b>DATA MARTS CORE (DM)</b>		
DM de Clientes	No existe.	No se tiene un DM de Clientes, se tiene un DM de Parque de Clientes, relacionado a nuevas altas.
DM de Tráfico	Desarrollado.	Se tiene un DM de Tráfico que cumple con las características principales descritas anteriormente.
DM de Control de Gestión	No existe	Se sugiere formular un proyecto para desarrollarlo.
<b>DATA MARTS AUXILIARES (DM)</b>		
DM de Productos	No existe.	No se tiene un DM como tal, se sugiere aprovechar los datos cargados en el DM de Ventas.
DM de Gestión Comercial	No existe.	Se sugiere formular un proyecto para desarrollarlo.
DM de Servicio al	No existe.	Se sugiere formular un proyecto para desarrollarlo.



<b>Mapa de Business Intelligence</b>	<b>Status de Desarrollo</b>	<b>Análisis GAP</b>
Ciente		
DM de Competencia	No existe.	Se sugiere formular un proyecto para desarrollarlo.
DM del Executive Information System	No existe.	Se sugiere formular un proyecto para desarrollarlo.
<b>APLICACIONES DE BUSINESS INTELLIGENCE</b>		
Category Management	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Sistema de Control de Gestión Comercial	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Sistema Analítico de Planes de Fidelización	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Sistema de Benchmarking de Tarifas	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Sistema de Recursos Humanos	No existe.	Se sugiere formular un proyecto para desarrollarlo.

<b>APLICACIONES OLAP</b>		
Recaudo (o Cobranzas)	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Control de Gestión	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Segmentación Reactiva	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Comercial	No existe.	Se sugiere formular un proyecto para desarrollarlo.

<b>Mapa de Business Intelligence</b>	<b>Status de Desarrollo</b>	<b>Análisis GAP</b>
Servicio al Cliente	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Tráfico	Cubierto parcialmente.	Se sugiere formular un proyecto para desarrollarlo. Se tienen algunas explotaciones sobre esto a partir de lo desarrollado en Tráfico.

<b>MODELOS DE DATA MINING</b>		
Modelo de Segmentación de Clientes	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Modelo de Productos	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Modelo de Mora de Clientes	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Modelo de Prevención de Fraude	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Modelo de Previsión de la Demanda	No existe.	Se sugiere formular un proyecto para desarrollarlo.
<b>CUADROS DE MANDO</b>		
Tablero de Control Departamental/Cuadro de Mando	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Executive Information System Corporativo	No existe.	Se sugiere formular un proyecto para desarrollarlo.
Balanced Scorecard	No existe.	Se sugiere formular un proyecto para desarrollarlo.



### **3. PRESENTACIÓN DE RESULTADOS Y PLAN DE ACCIÓN**

De la información revisada se plantean una serie de mejoras en diferente aspecto las cuales deben realizarse para poder garantizar la confiabilidad de la información generada por el Data Warehouse, se han dividido las recomendaciones según el ámbito en el que se desarrolla los procesos.

Se requiere mejorar la Arquitectura actual del Business Intelligence por influir directamente en la performance, escalabilidad y plan de crecimiento corporativo del Data Warehouse. Se necesita optimizar los procesos de extracción ya que influyen en reprocesos necesarios para garantizar la calidad de la información.

El plan de acción ha sido diseñado con el apoyo del proveedor, sobre la base de las metodologías definidas inicialmente y la que se considera ad-hoc para CLARO.

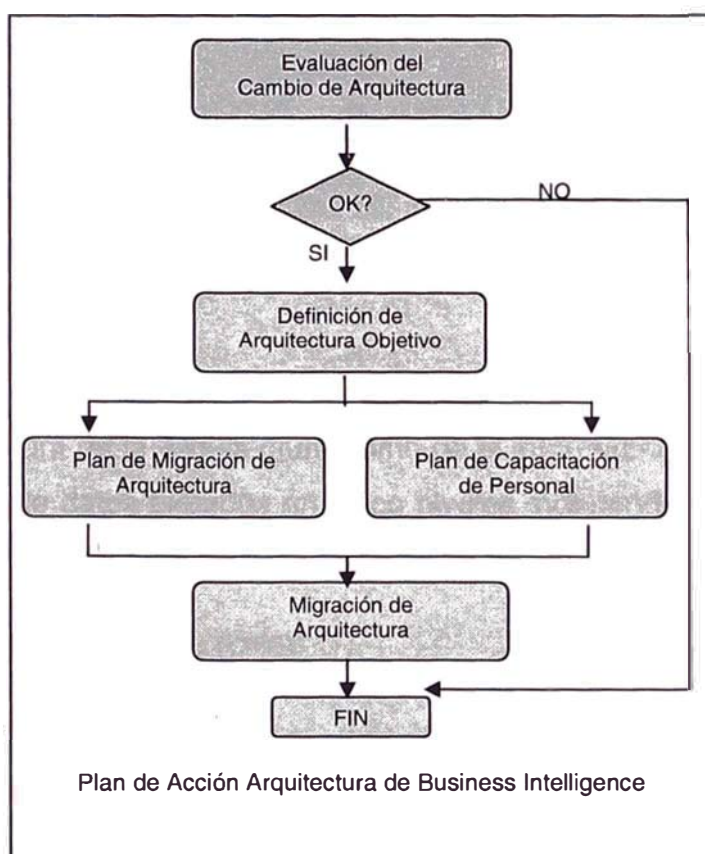
#### **3.1. Arquitectura de Business Intelligence**

Del análisis de los procesos estudiados se listan las siguientes recomendaciones a considerar dentro de la distribución de la arquitectura.

- Se debe definir una Arquitectura ideal para las necesidades, presentes y futuras de CLARO.
- Se recomienda que todas las capas de datos creadas se encuentren físicamente separadas, para evitar dependencia entre capas y problemas mientras algún proceso se este ejecutando, actualmente se ejecutan procesos sobre las mismas capas afectando la performance del DW
- Se debe realizar el desarrollo de una capa de datos ODS (Operational Data Storage) a fin de obtener un modelo relacional "limpio", donde las imperfecciones de data sean corregidas.
- Se debe apostar por ambientes de trabajo donde las iniciativas de Data Mining sean desarrolladas.

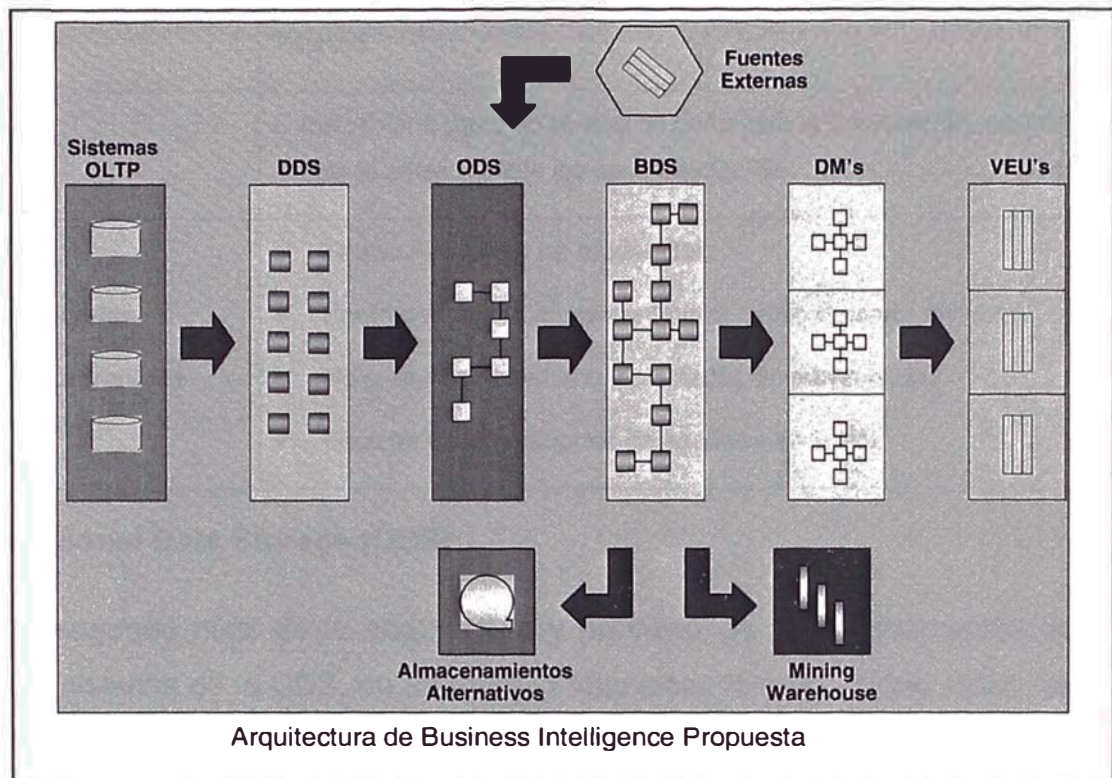
- Se debe dejar de construir capas semánticas dentro de las herramientas de explotación, en caso de depender de estas, se debe disminuir su utilización funcional.
- Se debe generar un Plan de Capacitación de los recursos humanos en el tema de Arquitectura de Business Intelligence Objetivo, capacitar a personas en las competencias adecuadas dentro de la arquitectura.

Se muestra un flujograma con las actividades a realizar en la etapa de diagnóstico de la Arquitectura.



### Arquitectura de Business Intelligence propuesta

La siguiente propuesta de Arquitectura de Business Intelligence BI ha sido adaptada sobre la base de las metodologías mencionadas inicialmente y la que se considera ad-hoc para CLARO.



## Estructura

Tanto la estructura de la Arquitectura cómo de la información son guardadas en estructuras de datos basados en cinco niveles independientes por donde puede pasar la información para su explotación en las distintas etapas.

## Development Data Storage (DDS)

Es el primer nivel de la Arquitectura, esta formado por tablas idénticas a las tablas de los sistemas OLTP, no tienen modelamiento y están creadas como copia simple, recibe los datos y sirve como área de construcción, se emplea para reducir tiempos de procesamiento y no afectar la operatividad.

## Resumen de Development Data Storage

Tópico	Descripción
Nivel	DDS

<b>Layer</b>	Técnico
<b>Objetivos</b>	Capa inicial a partir de la cual se construye la información, permite liberar lo antes posible los sistemas OLTP.
<b>Principales Características</b>	<ul style="list-style-type: none"> <li>✓ Tablas aisladas y no modeladas.</li> <li>✓ Por lo general es implementado a través de archivos texto.</li> <li>✓ No presenta constraints habilitados de darse en BD.</li> <li>✓ Foco en las operaciones de escritura en la BD.</li> </ul>

### Operational Data Storage (ODS)

Es el segundo nivel de la arquitectura y proviene de la transformación de tablas aisladas de la DDS, en un modelo relacional "limpio", donde todas las imperfecciones de data son depuradas y corregidas. La data de mala calidad es eliminada, el resto es homogenizado y consolidado. Solo la data no critica puede ser explotada con técnicas de Querying & Reporting, la data es alojada en Bases de Datos y se recomienda no habilitar los constraints, se garantiza información lista a ser modelada para la BDS.

### Resumen de Operational Data Storage

Tópico	Descripción
<b>Nivel</b>	ODS
<b>Layer</b>	Técnico
<b>Objetivos</b>	Asegurar que la información extraída de los Sistemas OLTP esté depurada, homogenizada y consolidada.
<b>Principales Características</b>	<ul style="list-style-type: none"> <li>✓ Tablas modeladas relacionalmente</li> <li>✓ Data Limpia.</li> <li>✓ Implementado en una Base de Datos.</li> <li>✓ Información ya puede ser explotada.</li> </ul>

## Almacenamientos Alternativos (AA)

Es un área de datos alterna y la menos tratada en la arquitectura, No debe ser tratada como área de respaldo o backup, deben tener diferente tratamiento al ser información de bajo nivel de acceso y que no se requiere frecuentemente (data inactiva), al crecer el volumen de datos crece la data inactiva, los problemas que comúnmente se presentan son:

- × Se están sobrestimando las necesidades de datos históricos.
- × Se esta considerando data que no se necesita para el análisis.
- × Creación de data sumariada que es sólo utilizada una vez.

Implantar AA tiene beneficios para la Arquitectura de Business Intelligence:

- ✓ Ahorro de dinero en costos de espacio y costos de procesamiento.
- ✓ Mejora de la performance de los querys.
- ✓ Mejor manejo de la granularidad.

## Resumen de los Almacenamientos Alternativos

Tópico	Descripción
Nivel	AA
Layer	Técnico
Objetivos	Economizar espacio con el almacenamiento de la data inactiva.
Principales Características	<ul style="list-style-type: none"><li>✓ Tablas modeladas dimensionalmente y tablas planas.</li><li>✓ La información no estará disponible inmediatamente.</li><li>✓ Cuando el modelo de la BDS cambia se desvinculan de los AA.</li><li>✓ Debe diferenciarse de las funciones de seguridad de datos.</li><li>✓ La sobrestimación de necesidades históricas, contribuyen a su importancia.</li><li>✓ Foco en las operaciones de recuperación de datos de la BD.</li></ul>

## Business Data Storage (BDS)

Es el corazón de la Arquitectura de BI y es comúnmente llamado Data Warehouse, aquí encontramos información corporativa modelada relacionamente, La BDS no es accesada por los usuarios finales, provee información para el resto de componentes.

Del ODS se crean tablas con información corporativa, La diferencia entre el BDS y el ODS es que la ODS se especializa en limpieza de datos y la BDS en provisión al resto del modelo, tener ambas capas juntas generaría problemas de disponibilidad durante las labores de limpieza.

## Resumen de Business Data Storage

Tópico	Descripción
<b>Nivel</b>	BDS
<b>Layer</b>	Negocio
<b>Objetivos</b>	Es el almacén que recoge la información corporativa del negocio modelada relacionamente y proveedora de información para el resto componentes.
<b>Principales Características</b>	<ul style="list-style-type: none"><li>✓ Tablas modeladas relacionamente, Implementado en una BD.</li><li>✓ No presenta constraints habilitados.</li><li>✓ Información no es explotada por usuarios finales.</li><li>✓ Se definen las tablas de conceptos corporativas.</li><li>✓ Foco en las operaciones de lectura y escritura de la BD.</li></ul>

## Data Marts (DM)

Este nivel es considerado el mas polémico y podría estimarse redundante. Un Data Mart es la organización de información extraída desde el Data Warehouse (BDS) a fin de describir un objetivo y que puede ser construido como conjunto de vistas de la BDS o de tablas de otra instancia de BD,



conjunto de tablas y vistas, un cubo en una localización distinta o una combinación de estos.

Los Data Marts junto al MW deben ser los únicos puntos de acceso a datos. Aquí se replica la información de la BDS cada vez que se necesite, según el requerimiento se necesitaran aplicaciones de Business Intelligence.

### Resumen de Data Marts

Tópico	Descripción
<b>Nivel</b>	DM
<b>Layer</b>	Negocio
<b>Objetivos</b>	Punto de acceso al modelo de datos para usuarios finales, resolución de problemas de negocio, para lo cual se cambia el modelamiento de la BDS.
<b>Principales Características</b>	<ul style="list-style-type: none"> <li>✓ Tablas modeladas dimensionalmente.</li> <li>✓ Puede resolver problemas de negocio especializado.</li> <li>✓ Pueden estar en la misma instancia de BD de la BDS.</li> <li>✓ Son necesarios para las Aplicaciones de BI</li> <li>✓ Foco en operaciones de lectura y contención de la BD.</li> </ul>

### Mining WareHouse (MW)

Es un área de datos alterna que almacena información necesaria para la minería de datos, el objetivo es salvaguardar la disponibilidad de la BDS ya que las consultas realizadas son pesadas pudiendo degradar la performance e incluso bloquearla. La extracción de datos es en batch y en horas previamente programadas.

Los datos son tratados para servir de estudios de Data Mining, Conservan su estructura inicial para poder volver a cargar información o generar nuevas corridas. Se almacena en archivos planos que alimentan los modelos



estadísticos desarrollados. Posee dos subdivisiones. *La de Congelamiento* que almacena los resultados, es de sólo lectura y se descongela para actualizar o correr procesos. *El área de estudio*, de libre disponibilidad para manipular datos, luego de validada y aprobada es congelada.

### Resumen de Mining Warehouse

Tópico	Descripción
Nivel	MW
Layer	Negocio
Objetivos	Salvaguardar la disponibilidad de la BDS ante los requerimientos de información de los data mineros.
Principales Características	<ul style="list-style-type: none"> <li>✓ Tablas planas, excepcionalmente se permite modelamiento dimensional.</li> <li>✓ Otorga a los data mineros independencia para modificar los datos.</li> <li>✓ Puede servir de fuente para otros estudios de data mining.</li> <li>✓ Tiene una subdivisión adicional, Área de Congelamiento y Estudio.</li> <li>✓ Se carga en batch de la BDS, pero puede recibir información de AA.</li> </ul>

### Vistas Externas de Usuario (VEU)

Son vistas y/o tablas desnormalizadas construidas a partir de la BDS o los DM, son puestas a disposición de los usuarios finales para explotar la información libremente, Aquí los estándares técnicos de nomenclatura se obvian y se utilizan nombres intuitivos a los usuarios.

### Resumen de Vistas Externas de Usuario

Tópico	Descripción
Nivel	VEU

<b>Layer</b>	Usuario
<b>Objetivos</b>	Simplificar la libre explotación de datos a los usuarios finales, fungiendo como capa semántica mediadora.
<b>Principales Características</b>	<ul style="list-style-type: none"> <li>✓ Vistas y/o tablas construidas a partir de Data Marts o la capa de datos BDS.</li> <li>✓ Debe ser un modelo intuitivo, Usa mucha denormalización.</li> <li>✓ Obvia los estándares de nomenclatura del Área de TI.</li> <li>✓ Se construyen on-demand.</li> </ul>

### Fuentes Externas (FE)

Son áreas de datos alternas que ha sido ignorada por años en las Arquitectura de BI, La crisis de análisis de muchas empresas ha sido el estar viendo la realidad por medio de los datos que sus sistemas transaccionales capturan vía transacciones, pero existe mucha más información que ella, la cual le permite ver posicionamiento versus competidores, tendencias en sus clientes, oportunidades de negocio no detectadas, etc.

La Arquitectura de BI no tiene autoridad en la conformación de los datos, estos llegan con determinado formatos y valor sobre el que se tienen pocas opciones de estandarización.

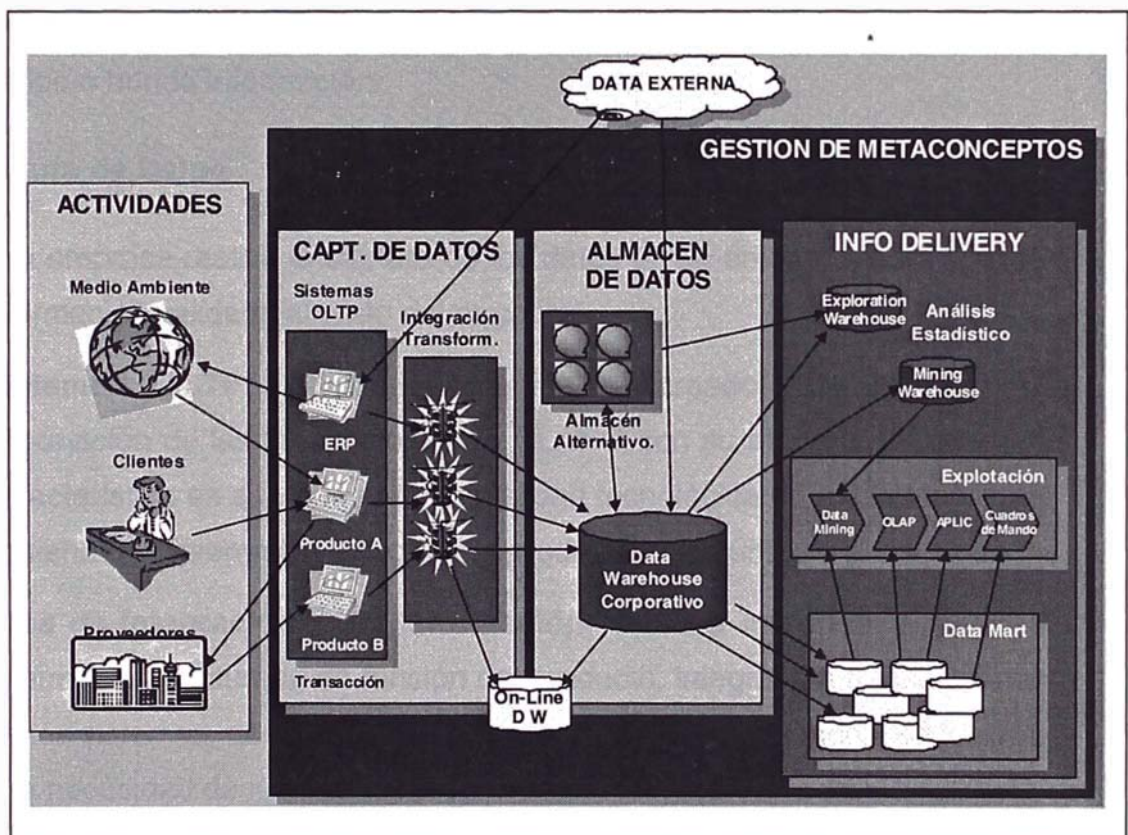
### Resumen de Fuentes Externas

<b>Tópico</b>	<b>Descripción</b>
Nivel	FE
Layer	Externo
Objetivos	Captura información del medio ambiente, no capturada por los sistemas OLTP.
Principales	✓ Ingresa directamente a la BDS

Características	✓ La Arquitectura de BI tiene pocas opciones para formateo de datos.
-----------------	--

### 3.2. Arquitectura de Datos Corporativa para CLARO

Se plantea además la siguiente propuesta de Arquitectura Corporativa de Datos para CLARO, este planteamiento es realizado a nivel teórico y basado en el conocimiento de plataformas de datos corporativos, y con ayuda de las mejores prácticas mundiales en empresas de Telecomunicaciones.



Los componentes que forman parte de la arquitectura planteada son:

#### I. Ambiente de Actividades

Es el mundo en cual CLARO interacciona como empresa, en este ambiente se da el relacionamiento de la organización con los diferentes agentes de su entorno.

- **Clientes:** Son los que interactúan con la organización y generan el mayor flujo de información, el cual generalmente es capturado por los Sistemas OLTP en forma de transacciones.
- **Proveedores:** Son el lado opuesto de la cadena de valor de CLARO, también generan transacciones que son capturadas por los Sistemas OLTP.
- **Medio Ambiente:** Engloba al resto de agentes del entorno: Organismos Reguladores, Tributarios, Gobierno, Sociedad y de los cuales CLARO recibe o brinda información.

## II. Captura de Datos

Aquí la empresa realiza la captura de los datos de su entorno, opera y nutre de información validada al resto del negocio.

- **Sistemas OLTP:** Son los sistemas encargados de capturar la información de las interacciones de CLARO con su entorno. La principal característica es que dado la complejidad tecnológica de estas empresas se tiene un Sistema OLTP por cada producto o conjunto de productos.
- **Área de Integración y Transformación:** Tiene la misión de corregir la información capturada en función del producto, integrar las transacciones de los sistemas, dándole sentido semántico y reacomodándola en función a la necesidad de la información.

## III. Almacenamiento de Datos

Es el responsable de almacenar los datos corporativos y disponerlos con alto grado de confiabilidad e integridad.

- **Data Warehouse Corporativo:** Repositorio central y único de información de la organización encargado de la salvaguarda de la información oficial del negocio.
- **On-Line Data Warehouse:** Área de datos que consolida información del negocio para necesidades operativas, mejor tiempo de respuesta e

información actualizada. Se diferencia del Data Warehouse Corporativo en que no almacena historia y tiene información más actual.

- **Almacenamientos Alternativos:** Áreas de datos donde se guarda data con bajos niveles de acceso y que pasa a conformar la historia del negocio.

#### IV. Info Delivery

Ambiente donde se manipulan los datos para efectos de realizar análisis de información y generación de conocimiento.

- **Data Marts:** Áreas de datos donde se modelan temas específicos de negocio con la finalidad de que las herramientas de explotación accedan de manera simple a los datos.
- **Modelos de Data Mining:** Modelo estadísticos que buscan encontrar patrones ocultos en grandes volúmenes de información.
- **Aplicaciones OLAP:** Aplicaciones donde se explota de manera libre los Data Marts por medio del uso de modelos dimensionales.
- **Aplicaciones de Business Intelligence:** Aplicaciones desarrolladas para análisis de información y generar procesos de creación de valor.
- **Cuadros de Mando:** Modelos de seguimiento a los indicadores y sus desviaciones.
- **Exploration WareHouse:** Ambiente separado del Data Warehouse que permite realizar consultas pesadas para explorar la información.
- **Mining WareHouse:** Ambiente separado del Data Warehouse que permite realizar tareas de minería de datos.

#### V. Data Externa

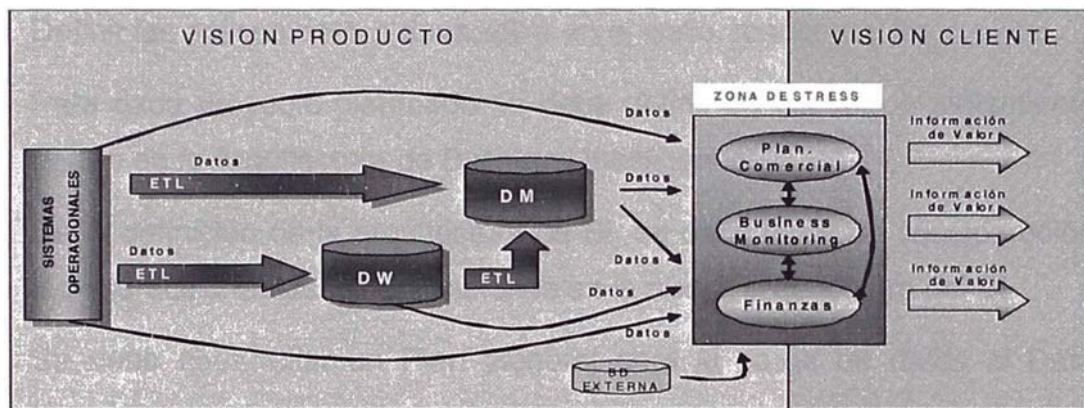
Ambiente que recoge información del entorno con el fin de observar la performance versus su competencia y encontrar información no percibida sobre los clientes.



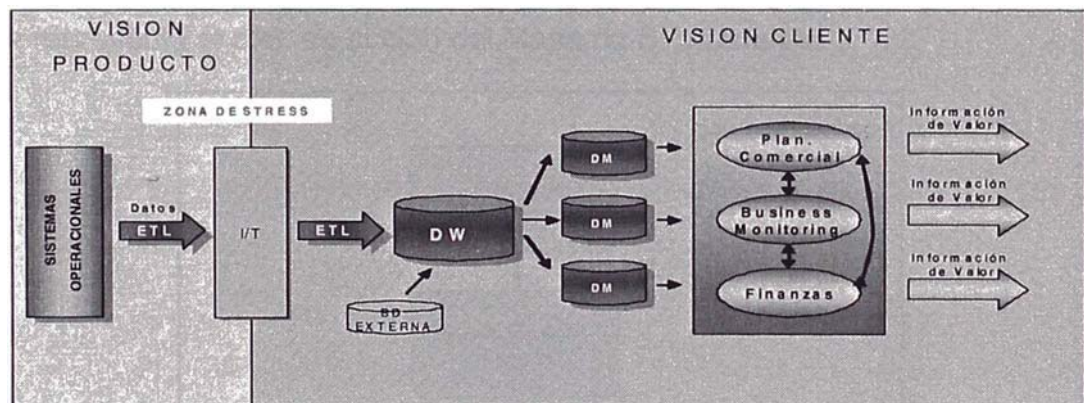
## VI. Gestión de Metaconceptos

Administra el catalogo de conceptos y definiciones para todo el negocio, su principal objetivo es homologar los significados en toda la empresa.

El siguiente grafico muestra la conceptualización del modelo actual y el propuesto:



Modelo Actual de Gestión de Información de Análisis



Modelo Propuesto de Gestión de Información de Análisis

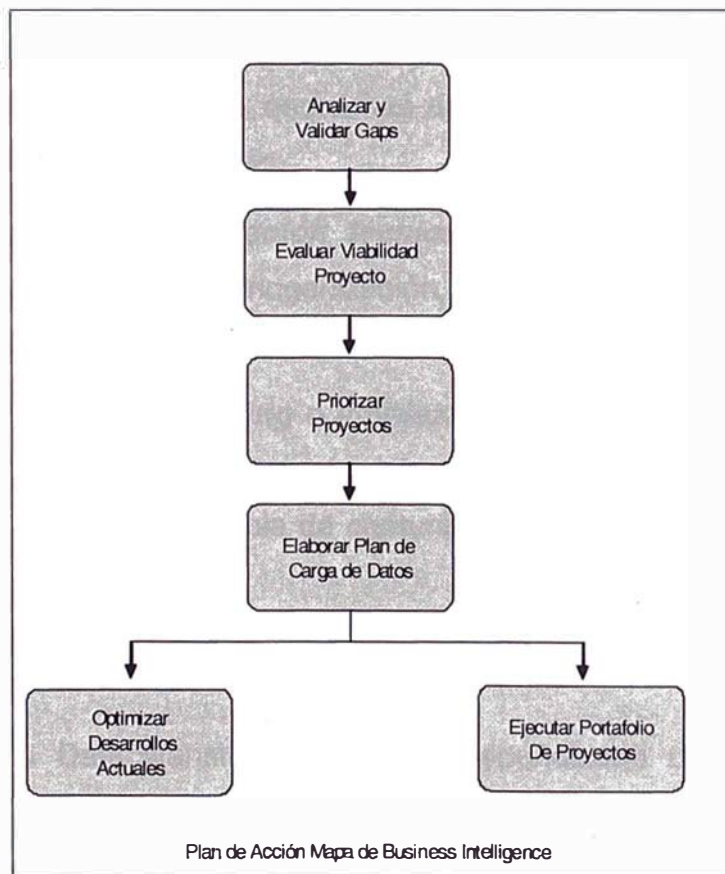
<b>Modelo Actual</b>	<b>Modelo Propuesto</b>
La zona de stress se encuentra del lado usuario	Se asigna una etapa inicial de stress a fin de revisar y validar la información
Los usuarios reciben información del DW, de los DM y de los sistemas corporativos	Los usuario solo recibirán información de los DM especializados
Los usuarios a su vez pueden obtener información de BD externas.	Toda información externa debe ser ingresada al DW.

### 3.3. Mapa de Business Intelligence

Con apoyo de los procesos revisados se listan las siguientes recomendaciones a considerar para optimizar el mapa del BI

- ✓ Analizar cuidadosamente los resultados obtenidos de los GAPs y realizar las validaciones funcionales para determinar la viabilidad de estas.
- ✓ Definir las aplicaciones a desarrollar en el corto, mediano y largo plazo.
- ✓ Para cada proyecto planteado se debe definir su espacio o alojamiento dentro de la Arquitectura de Business Intelligence.
- ✓ De la definición de la arquitectura de BI, se debe tener en consideración las interacciones de las aplicaciones en la nueva arquitectura del BI
- ✓ Se debe desarrollar un Plan General para la carga de datos al Data Warehouse y que permita adelantarse a las necesidades del usuario.

Se representa el plan de acción del Mapa de BI propuesto.





### **3.4. Procesos de Extracción, Transformación y Carga de Datos (ETL)**

#### **A nivel de Extracción de Datos**

Los procesos ETL que se desarrollan en CLARO cubren eficientemente las expectativas que se tienen dentro de la arquitectura de BI tanto en el aspecto de ventanas de tiempo como del proceso mismo de carga de datos.

Estos Procesos son enfocados como un conjunto de programas que extraen, transforman y cargan data en el Data Warehouse de la empresa, esta visión es adecuada y efectiva para los intereses del propio Data Warehouse, pero consideramos que el nuevo rol de los Procesos ETL es el de convertirse en un Layer de Integración y Transformación de Datos (Layer I/T).

La visión de un Layer I/T es la de un área de datos donde la data de las aplicaciones son integradas y transformadas antes de convertirse en data corporativa. El Layer I/T sirve a todas las necesidades de la empresa, no sólo al Data Warehouse, y realiza tres tareas específicas

- ✓ ***Procesos de transformación conceptual de data***, se busca homologar conceptos transformar los esquemas de la organización para atender a las necesidades del negocio.
- ✓ ***Procesos de transformación técnica de data***, donde se realiza la selección, filtrado, homogenización, reformateo, reordenamiento, consolidación y agregación de la información.
- ✓ ***Procesos de datacleansing o limpieza de datos***, este proceso es vital para asegurar la calidad de la información que se distribuye en la empresa, no es responsable de capturar data de buena calidad, sino de difundirla.

#### **Datacleansing en ambientes de Business Intelligence**

El análisis de Datacleansing han sido independizado de los temas de estudio, esto por considerar un proceso fundamental en toda Arquitectura de Business Intelligence, en la actualidad operativa del Data Warehouse no se

a profundizado en que técnicas a utilizar, los procesos ETL tienen implementadas algunas consideraciones, pero no utilizan técnicas formales.

La arquitectura de Business Intelligence tiene valor técnico y funcionalmente por dos características principales:

- ✓ Capacidad para disponer de información de análisis y
- ✓ Calidad de información entregada

Por esto es que las operaciones de datacleansing son fundamentales en los procesos de extracción y carga, ya que garantiza que la información capturada por los sistemas OLTP y que son cargadas en la Arquitectura del Business Intelligence sean de calidad.

### **Definición de Calidad en la Data**

La calidad en la data se define sobre la base de filas buenas o malas de una tabla. Una fila es buena cuando cumple con todos los constraints definidos para las columnas de la tabla.

Un constraint es una regla de negocio escrita en SQL. Se tienen los siguientes tipos de constraint:

- ✓ De Clave Foránea: El valor de la columna debe existir como Primaria en otra tabla.
- ✓ De No Nulidad: El valor de la columna no puede ser nulo.
- ✓ De Valores: El valor de la columna debe encontrarse dentro de un subconjunto de valores definido.
- ✓ De Chequeo Lógico: Indica en general cualquier regla lógica para las columnas. Ejemplo: "fecha inicial menor que fecha final".

Pueden haber columnas en una tabla que no tengan constraints, por ejemplo: nombre, dirección, etc.

## **CAPITULO IV**

### **CONCLUSIONES Y RECOMENDACIONES**

#### **CONCLUSIONES**

1. Hoy en día las empresas buscan tener información en tiempo real, con el objetivo de obtener ventajas en mercados cada vez más competitivos, esta tendencia pone al Data Warehouse en una posición crítica, debido a que no solo provee información estratégica, sino que además sirve para apoyar a todos los procesos que aporten al negocio, tales como administración de los riesgos, retención y fidelización del cliente, cobranzas, fraudes, trafico de llamada, etc.
2. CLARO utiliza las mejores prácticas en el desarrollo y explotación de la información, sin embargo los agentes proveedores de información no se encuentran correctamente involucrados, generando procesos adicionales que afectan la naturaleza del Data Warehouse, por esto se deben aplicar verdaderos cambios tecnológicos que permita optimizar los subsistema input/output para obtener mejores rendimientos.
3. Se apuesta por potenciar el uso de la inteligencia de negocios, lo que para CLARO implica la transición de un modelo más tradicional hacia uno más activo, esto es un gran cambio para la organización, tanto en aplicación de tecnologías como procesos de negocio, no se trata de una implementación que se haga en una noche, sino que es una evolución.
4. Se debe optimizar las decisiones funcionales y técnicas tomadas en el pasado para contribuir a incrementar el retorno de la inversión, Por este motivo, para garantizar los beneficios en los sistemas de información, es

5. necesario que estos se desarrollen de manera efectiva, de forma que resuelvan las necesidades de información del negocio.

## **RECOMENDACIONES**

Del estudio realizado se han obtenido mejoras y recomendaciones a aplicar que determinaran el plan de acción para el desarrollo del entorno del Business Intelligence, las cuales garantizaran la confiabilidad de la información generada, estas recomendaciones están referidas a los procesos que influyen directamente en la obtención de la data.

1. Se deben definir los roles de los agentes participantes en la administración de información, desarrollar culturas de información basada en la “necesidad de saber” según su rol, esto nos lleva a definir modelos que relacione el rol de cada uno con el logro del objetivo individual y corporativo.
2. Reordenar el envío de datos a los usuarios, revisar los datos enviados y realizar las correcciones, establecer procedimientos operativos que garanticen la calidad de información en los procesos que los generan, asegurar la consistencia de la información y cuadro de datos en la fuente.
3. Se debe reestructurar la Arquitectura de Business Intelligence a una ideal planteada según las necesidades, presentes y futuras, para lo cual se recomienda seguir la metodología propuesta.
4. Aprovechar mejor el uso de los recursos informáticos, mejorando los tiempos de producción y centralizando los procesos, esto permitirá reducir costos (usuarios, IT).
5. Desarrollar mecanismos de distribución de información que permitan dosificar la entrega a quien sea útil y necesaria, basados en una distribución acorde a las necesidades.
6. Incrementar la seguridad de la información, reduciendo el riesgo de vulnerabilidad de los repositorios de datos por parte de las áreas usuarias.

7. Se debe realizar el desarrollo de una capa de datos ODS (Operational Data Storage) para obtener un modelo relacional "limpio", donde todas las imperfecciones de data sean corregidas, actualmente se realizan las correcciones sobre la misma área temporal definida anteriormente.

## **GLOSARIO DE TERMINOS**

**Business Intelligence:** Inteligencia de Negocios

**Business Object:** Herramienta de software

**BDS:** Business Development Storage

**Clientes Consumer:** Clientes naturales que poseen cuentas simples

**Clientes Business:** Clientes jurídicos, que cuentan con más de 3 líneas

**DDS:** Development Data Storage

**Análisis GAP:** Comparación de lo que se tiene versus lo que se espera

**GPRS:** Global Packet Radio System

**GSM:** Sistema Global Comunicaciones Móviles

**ETL:** Extract, Transform and Load

**Minería de Datos:** Conjunto de técnicas de explotación de información a partir de una gran masa de datos

**ODS:** Operational Data Storage

**OMA:** Open Mobile Alliance (relacionado con la tecnología WAP)

**OLTP:** OnLine Transaction Processing

**SMS:** Short Message Service

**MMS:** Multimedia MessagING Service

**DATA STAGES:** Es la herramienta de desarrollo de ETL's

**VEU:** Vistas Externas de Usuario

**WAP:** Wireless Application Protocol

## BIBLIOGRAFIA

- ✓ Kimball, Ralph ; Ross, Margy  
"The Data Warehouse Toolkit: The complete Guide to Dimensional Modeling"  
Editorial: Wiley Computer Publishing, Printed in The United State of America, 2002.
- ✓ Barquini, Ramon  
"Planning and Designing the Warehouse"  
Editorial: Prentice-Hall, Printed in New Jersey, 1996.
- ✓ Consultora SOLUZIONE – 2005  
"Consultoria sobre Ásessment del Modelo de Datos - TIM"
- ✓ BCTS - Business Consulting and Technology Services  
Guía practica acerca de data Warehouse, mejores prácticas.
- ✓ Centro de estudios Seminariun  
Separatas sobre programación de ETL's y el manejo de Data Minig
- ✓ Web Sites sobre construcción de un Data Warehouse y programación de ETL's  
Solution for managing, optimization a flexible data architecture, and best practices.  
<http://www.informationbuilders.com>
- ✓ Strategies for custom Data Warehouse and ETL process.  
<http://www.dmreview.com>
- ✓ ETL Architecture in Depth (ETLD), Kimball Group, Data Warehouse Training.  
<http://www.kimballgroup.com>
- ✓ Conceptos básicos sobre Data Warehouse, ETLs, OLAP, etc.  
<http://www.monografias.com/trabajos6/dawa/dawa.shtml>