

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



“IMPLEMENTACIÓN DE UNA HERRAMIENTA DE EXTRACCIÓN,  
TRANSFORMACIÓN Y CARGA (ETL) PARA MEJORAR EL TIEMPO  
DE ATENCIÓN DE LOS REQUERIMIENTOS DE INTELIGENCIA DE  
NEGOCIOS”

**INFORME DE SUFICIENCIA**  
**PARA OPTAR EL TITULO PROFESIONAL DE:**  
**INGENIERO DE SISTEMAS**

RODRIGUEZ LAU, LUIS ALBERTO

**LIMA – PERU**

**2008**

*"Dedico este trabajo a mi padre y a mi madre, quienes en todo momento me brindaron su apoyo incondicional".*

*Agradezco a Dios, a mi familia y amigos que en todo momento me apoyaron para seguir adelante y a mi alma mater la UNI por haberme formado como profesional*

## INDICE

<b>DESCRIPTORES TEMATICOS .....</b>	<b>1</b>
<b>RESUMEN EJECUTIVO.....</b>	<b>2</b>
<b>INTRODUCCIÓN .....</b>	<b>4</b>
<b>CAPÍTULO I: ANTECEDENTES .....</b>	<b>5</b>
<b>I.1 DIAGNÓSTICO ESTRATÉGICO.....</b>	<b>5</b>
I.1.1 SITUACIÓN ACTUAL.....	5
I.1.2 MISION.....	7
I.1.3 VISION .....	7
I.1.4 VALORES.....	7
I.1.5 FORTALEZAS Y DEBILIDADES .....	7
I.1.6 OPORTUNIDADES Y AMENAZAS .....	8
I.1.7 OBJETIVOS ESTRATEGICOS .....	9
<b>I.2 DIAGNÓSTICO FUNCIONAL .....</b>	<b>11</b>
I.2.1 PRODUCTOS.....	11
I.2.2 CLIENTES .....	11
I.2.3 PROCESOS .....	12
I.2.4 ORGANIZACIÓN .....	15
I.2.5 UNIDADES ORGANIZATIVAS .....	18
<b>CAPITULO II: MARCO TEÓRICO .....</b>	<b>27</b>
<b>II.1 INTELIGENCIA DE NEGOCIOS.....</b>	<b>27</b>
<b>II.2 DATAWAREHOUSE &amp; DATAWAREHOUSING .....</b>	<b>28</b>
<b>II.3 EXTRACCIÓN, TRANSFORMACIÓN y CARGA (ETL).....</b>	<b>32</b>
<b>II.4 HERRAMIENTA ETL.....</b>	<b>38</b>
<b>II.5 PROCESO DE ATENCIÓN DE REQUERIMIENTOS (PAR) .....</b>	<b>39</b>
<b>II.6 GERENCIA DE PROYECTOS.....</b>	<b>41</b>
<b>CAPITULO III: PROCESO DE TOMA DE DECISIONES.....</b>	<b>44</b>
<b>III.1 PLANTEAMIENTO DEL PROBLEMA.....</b>	<b>44</b>
III.1.1 SITUACIÓN ACTUAL DEL PROCESO DE DESARROLLO ETL ...	45
III.1.2 PRINCIPALES PROBLEMAS EN EL PROCESO ETL.....	46
<b>III.2 ALTERNATIVAS DE SOLUCIÓN.....</b>	<b>47</b>

<b>III.3</b>	<b>METODOLOGÍA DE EVALUACIÓN DE SOLUCIONES</b> .....	<b>48</b>	
III.3.1	CRITERIOS DE DECISIÓN.....	48	
III.3.2	DEFINICIÓN DE MÉTRICAS PARA LA TOMA DE DECISIONES	51	
<b>III.4</b>	<b>TOMA DE DECISIONES</b> .....	<b>58</b>	
<b>CAPITULO IV:</b>			
<b>ESTRATEGIAS ADOPTADAS Y EVALUACIÓN DE RESULTADOS</b> .....			<b>59</b>
<b>IV.1</b>	<b>METODOLOGIA DE GESTIÓN DE PROYECTOS</b> .....	<b>59</b>	
IV.1.1	INICIO .....	61	
IV.1.2	PLANIFICACIÓN .....	66	
IV.1.4	EJECUCIÓN .....	70	
IV.1.5	CONTROL .....	76	
IV.1.6	CIERRE.....	76	
<b>IV.2</b>	<b>EVALUACIÓN DE RESULTADOS</b> .....	<b>77</b>	
<b>CONCLUSIONES Y RECOMENDACIONES</b> .....			<b>78</b>
<b>GLOSARIO DE TÉRMINOS</b> .....			<b>80</b>
<b>BIBLIOGRAFÍA</b> .....			<b>84</b>
<b>ANEXOS</b> .....			<b>85</b>

## DESCRIPTORES TEMATICOS

- INTELIGENCIA DE NEGOCIOS
- DATAWAREHOUSE & DATAWAREHOUSING
- EXTRACCIÓN, TRANSFORMACIÓN Y CARGA (ETL)
- HERRAMIENTA ETL
- PROCESO DE ATENCIÓN DE REQUERIMIENTOS (PAR)
- GERENCIA DE PROYECTOS

## RESUMEN EJECUTIVO

“En este nuevo mundo, la información reina...” afirma Geoffrey A. Moore, Director de Chasm Group <sup>1</sup>. Vivimos en una época en que la información es la clave para obtener una ventaja competitiva en el mundo de los negocios. Para mantenerse competitiva una empresa, los gerentes y tomadores de decisiones requieren de un acceso rápido y fácil a información útil y valiosa de la empresa. Una forma de solucionar este problema es por medio del uso de Business Intelligence o Inteligencia de Negocios.

El rápido avance de los negocios, el mercado cambiante y el rápido avance de la competencia, exige al negocio a moverse y evolucionar de manera cada vez más dinámica y más rápidamente, lo que obliga a que se deban tomar decisiones rápidas y acciones cada vez más ágiles. Para esto es necesario contar con el conocimiento e información oportuna, consistente y en el tiempo requerido.

El presente informe, muestra la situación actual del proceso de desarrollo de estas soluciones de Inteligencia de negocios y de cómo los objetivos estratégicos de la organización conllevan a realizar una mejora sustancial en la productividad de estas soluciones, de manera que se pueda ofrecer a la organización, soluciones más rápidas, ágiles sin descuidar la calidad en la entrega de información.

---

<sup>1</sup> Net Library. [http://www.netlibraryebooktoolkit.com/content/library\\_patron\\_support/3](http://www.netlibraryebooktoolkit.com/content/library_patron_support/3)

Actualmente las soluciones de Inteligencia de Negocios abarcan varios procesos dentro de la organización, desde la extracción, transformación y carga de la información al Datawarehouse organizacional, pasando por el diseño de Datamarts específicos de información hasta la explotación con herramientas analíticas del negocio. En este informe, hablaremos específicamente del proceso de Extracción, Transformación y Carga de la información, proceso que la organización decide automatizar mediante el uso de herramientas de última generación. Se utilizará la metodología de gestión de proyectos, que es estándar en la organización, para llevar a cabo esta tarea.

Se considerarán dos alternativas, el desarrollo a medida de una herramienta de extracción, transformación y carga o la adquisición de una herramienta de clase mundial que haga esta labor, se harán los respectivos análisis para determinar cual solución es más idónea. Una vez elegida la solución se inicia el proceso de implementación

En este proceso de mejora y automatización se toman en cuenta varias interrogantes, los cuales se irán resolviendo a lo largo de este informe, estas son ¿Qué herramienta es la idónea?, ¿Cómo impacta esta implementación a mis procesos actuales?, ¿Cuál debería ser la estrategia de implementación? ¿Cómo institucionalizo el proceso? ¿Cómo aporta esta solución a la mejora en el proceso de desarrollo?

Finalmente veremos la estrategia de solución y la puesta en marcha de la herramienta y como se ven cubiertos los objetivos de la organización con este proyecto.

## INTRODUCCIÓN

En los últimos años la economía del país está creciendo de manera considerable, por lo cual muchos bancos extranjeros están poniendo sus ojos en el Perú como un país atractivo para la inversión lo que obliga a las instituciones financieras tengan una respuesta adecuada a la competencia y al ritmo cambiante de los mercados. En este entorno, el Banco de Crédito del Perú, busca incrementar su competitividad en todos los aspectos, y para ello tiene como brazo tecnológico a la División de Sistemas y Organización para contar con soluciones eficientes y oportunas para poder optimizar sus procesos y obtener información en el momento preciso para lograr cumplir los objetivos trazados por la organización. En este entorno, las necesidades de los usuarios se incrementan, siendo necesario un tiempo de respuesta adecuado a la demanda de requerimientos.

El área de Gestión de Procesos e Información, que se encarga de proveer la información de manera oportuna y exacta a la organización con la finalidad de apoyar a la toma de decisiones, no es ajena a este entorno, por lo cual busca incrementar su tiempo de respuesta ante las necesidades de los usuarios, para lo cual busca mejorar el tiempo de entrega de los proyectos de Inteligencia de Negocios de manera que pueda satisfacer las necesidades de los usuarios, de esta manera surge la necesidad de implementar nuevas herramientas y metodologías que permitan optimizar el desarrollo de dichos proyectos.

# **CAPÍTULO I**

## **ANTECEDENTES**

El desarrollo de este proyecto, está basado en la experiencia de una de las más importantes Entidades Financieras del Perú, el Banco de Crédito del Perú, específicamente en el Área de Gestión de Procesos e Información, que es la encargada de los procesos de Inteligencia de Negocios de la Organización. La metodología y procesos involucrados para el desarrollo de las soluciones de Inteligencia de Negocios, son un estándar, utilizado en muchas organizaciones a nivel mundial.

### **I.1 DIAGNÓSTICO ESTRATÉGICO**

#### **I.1.1 SITUACIÓN ACTUAL**

En el Banco de Crédito del Perú, el año 2007 al igual que el 2006 han sido extraordinarios, por los resultados obtenidos y por la evolución positiva de los negocios. La utilidad neta del año, aplicando principios internacionales de contabilidad, ha sido de US\$ 294.5 millones, 23.3% mayor que la utilidad comparable del 2006, logrando una alta rentabilidad sobre el patrimonio

promedio, rentabilidad que es muy satisfactoria y la más alta obtenida de la cual se tenga recuerdo.

Actualmente los canales del Banco crean los medios de atención al cliente, por ejemplo el nuevo canal de atención Agente BCP. En ellos se puede efectuar un número determinado de operaciones por montos limitados. Esto es una alternativa para ampliar rápidamente, con bajo costo, la red de atención a los clientes en zonas donde hoy se registran pocas transacciones. En el año 2008 se instalarán 170 nuevos Agentes y la meta fue tener 720 al finalizar el 2007.

A su vez el Banco de Crédito está lanzando muchos productos innovadores, como por ejemplo la Cuenta Sueldo, la Tarjeta de Crédito Exacta, entre otros, de manera que pueda cada vez mas acercarse a su público objetivo y lograr más participación en el mercado cada año.

Como podemos observar, el negocio está creciendo a ritmo acelerado, pero a su vez, la competencia se hace cada vez más fuerte, muchos bancos están realizando estrategias muy agresivas, de manera que es necesario, ante este alto nivel de competitividad, tener las herramientas de gestión y de toma de decisiones eficaces y oportunas que permitan a la organización moverse al ritmo del mercado y la competencia.

### **I.1.2 MISION**

“Servir al cliente”.

El BCP se debe a sus clientes para conseguir su preferencia, fidelidad y permanencia.

### **I.1.3 VISION**

“Ser un banco simple, transaccional, rentable y con personal altamente capacitado y motivado.”

### **I.1.4 VALORES**

- **El cliente:** Nos debemos a nuestros clientes.
- **La Ética:** Somos una institución con integridad, con gente honesta y responsable.
- **Nuestra gente:** Contamos con los mejores profesionales, incentivamos su desarrollo y potencial emprendedor.
- **La Innovación:** Innovamos continuamente para responder a los requerimientos del mercado.

### **I.1.5 FORTALEZAS Y DEBILIDADES**

#### **FORTALEZAS**

- Liderazgo en el Sistema Financiero Peruano.
- Plana Gerencial Profesional y Experimentada.
- Sólida Estructura Financiera, importante respaldo patrimonial del Grupo Credicorp.
- Conservadora Gestión de Riesgo crediticio.
- Diversificación de negocios, productos y servicios bancarios.

- Capacidad para gestionar información de manera integrada
- Bajos niveles de intermediación financiera

### **DEBILIDADES**

- Niveles de eficiencia por debajo del promedio del sistema.
- Riesgo devaluatorio asociado al alto nivel de dolarización de la cartera.
- Reducida automatización de determinadas áreas de negocios.
- Procesos de toma de decisiones lento y burocrático

## **I.1.6 OPORTUNIDADES Y AMENAZAS**

### **OPORTUNIDADES**

- Alianzas estratégicas con laboratorios.
- Expansión de servicios a través del uso intensivo de los canales de distribución y venta cruzada de productos.
- Favorables perspectivas macroeconómicas.

### **AMENAZAS**

- Mayor competencia entre bancos grandes.
- Márgenes financieros sujetos a variaciones en las tasas de interés.
- Altos niveles de dolarización en el sistema bancario.
- Bajo dinamismo del crédito bancario.

## I.1.7 OBJETIVOS ESTRATEGICOS

Las estrategias que guían la gestión de la empresa son las siguientes:

- Fortalecer el balance y reducir la volatilidad de las utilidades, mejorando la calidad de los activos administrados y elevando la capitalización de nuestra empresa.
- Mitigar los riesgos de mercado, además de cubrir adecuadamente el riesgo crediticio y el de seguros.
- Mejorar la eficacia en la gestión de las operaciones, reduciendo nuestros gastos operativos teniendo como meta 50% de los ingresos.
- Privilegiar los negocios con las pequeñas empresas y personas, incorporando a sectores poco bancarizados.
- Impulsar los servicios bancarios.
- Otro lineamiento estratégico que está implícito en la gestión, es el Buen Gobierno Corporativo y la Responsabilidad Social. En este aspecto se aspira a que:
  - Los clientes reciban la mejor calidad de servicio.
  - Los accionistas obtengan una alta rentabilidad y gocen de la más amplia y transparente información.
  - Los empleados reciban una adecuada compensación y desarrollo profesional.
  - El compromiso con el país de seguir contribuyendo para lograr mayores niveles de bienestar económico y social.

En la Figura 1 que se presenta a continuación se muestra el Mapa Estratégico del Banco.

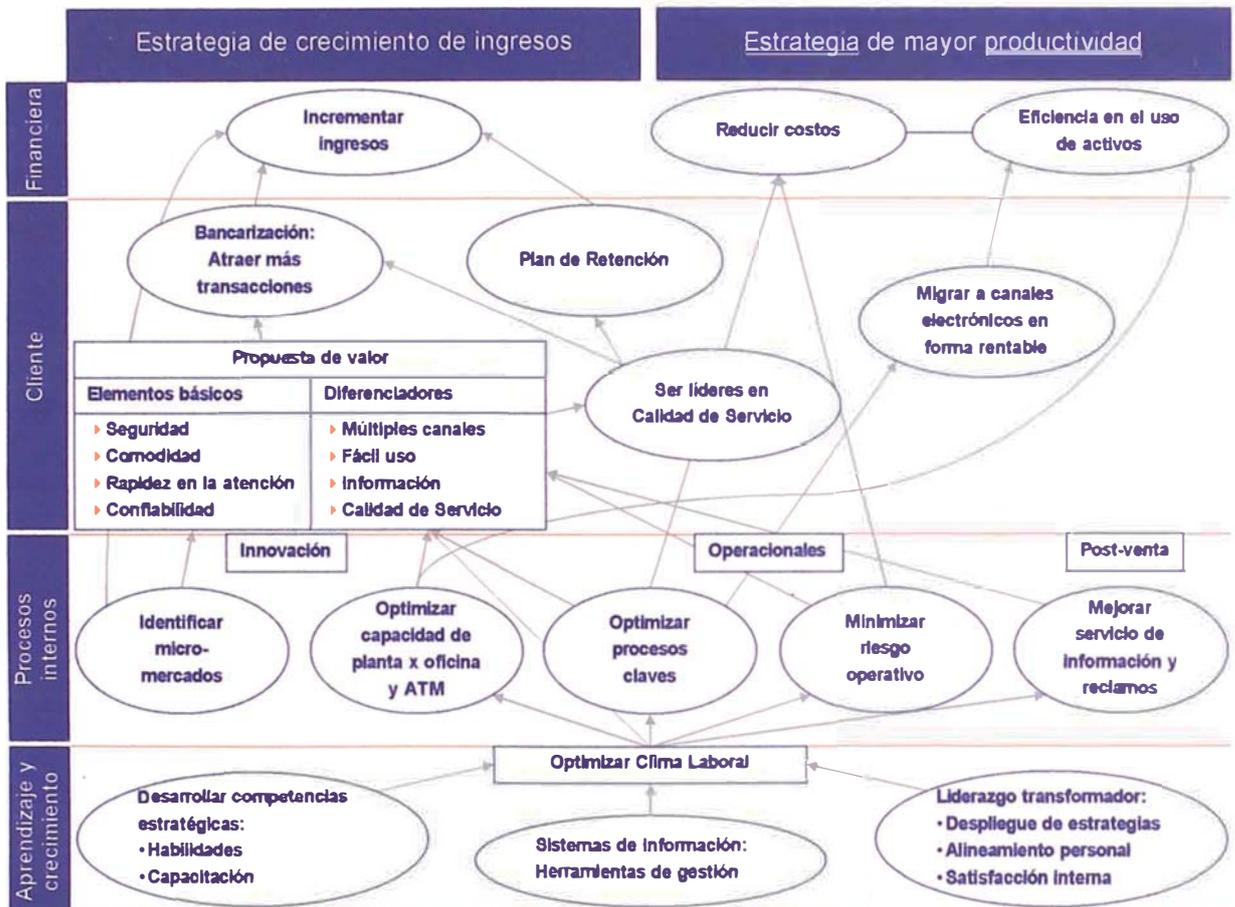


Figura 1. Mapa estratégico

Los objetivos principales mostrados en el mapa estratégico para este año son incrementar los ingresos y reducir los costos utilizando eficientemente los activos. El incremento de los ingresos se logrará a través de la bancarización y la retención de los clientes más rentables basados en una propuesta de valor que tiene como elementos básicos: la seguridad, confiabilidad y rapidez de atención; y como diferenciadores los múltiples canales con los que se dispone que nos coloquen como líderes en calidad y

servicio. Asimismo para lograr la reducción de los costos se buscará migrar a canales electrónicos en forma rentable.

## **I.2 DIAGNÓSTICO FUNCIONAL**

### **I.2.1 PRODUCTOS**

Los productos de la empresa están divididos en 5 categorías:

- Cuenta de ahorros, corriente e inversiones
- Tarjetas de crédito y débito
- Créditos (Personales. Hipotecarios)
- Seguros
- Envíos y transferencias de dinero, servicios diversos

### **I.2.2 CLIENTES**

#### **1. Banca de Consumo**

Es un segmento de clientes con rentas medias y bajas que por sus características y hábitos financieros requieren de productos simples y servicios estándares.

#### **2. Banca Exclusiva**

Que agrupa a las personas naturales con mayores saldos en el Banco. Sus características, hábitos y necesidades financieras indican que debe brindárseles un servicio preferente y diferenciado y ofrecerles innovadores productos, canales de atención y precios competitivos a la altura de sus exigencias.

### **3. Banca Pequeña Empresa**

Este segmento está conformado por las personas naturales con negocio y personas jurídicas con niveles de ventas entre 10 mil y 300 mil dólares anuales.

El Banco, conocedor de la variedad de negocios existentes en el país incursionó hace algunos años en el segmento de la pequeña empresa, sector con gran potencial de crecimiento y desarrollo, buscando atender cada vez a más clientes.

### **4. Banca de Negocios**

Cuya función es atender a las Personas Jurídicas o Personas Naturales de Negocio cuya facturación anual sea menor a un millón de dólares.

Investigaciones han determinado que en este segmento del mercado existen diversos tipos de clientes, a los cuales la Banca de Negocios los ha denominado como segmentos consolidado y segmento pequeña empresa.

## **I.2.3 PROCESOS**

El Banco cuenta con un Sistema de Gestión de Procesos, que cuenta con las siguientes características:

- a) Definición: Está considerado como Gestión de Procesos, la administración, control y organización de los procesos del Banco con la finalidad de garantizar su revisión y mejora continua.

b) El Área Soluciones de Negocios - Servicio Mejora de Procesos es responsable de la administración del Sistema, el que consta de los siguientes elementos:

- Mapa de Procesos.
- Modelo de trabajo.

### **MAPA DE PROCESOS**

a) Tiene por finalidad identificar y ordenar los procesos del Banco, el cual está agrupado y subdividido en grupos de procesos que interactúan directamente con Clientes (procesos de creación de valor) y los que son de Soporte. El Mapa de Procesos se encuentra graficado en la página siguiente, Figura2.

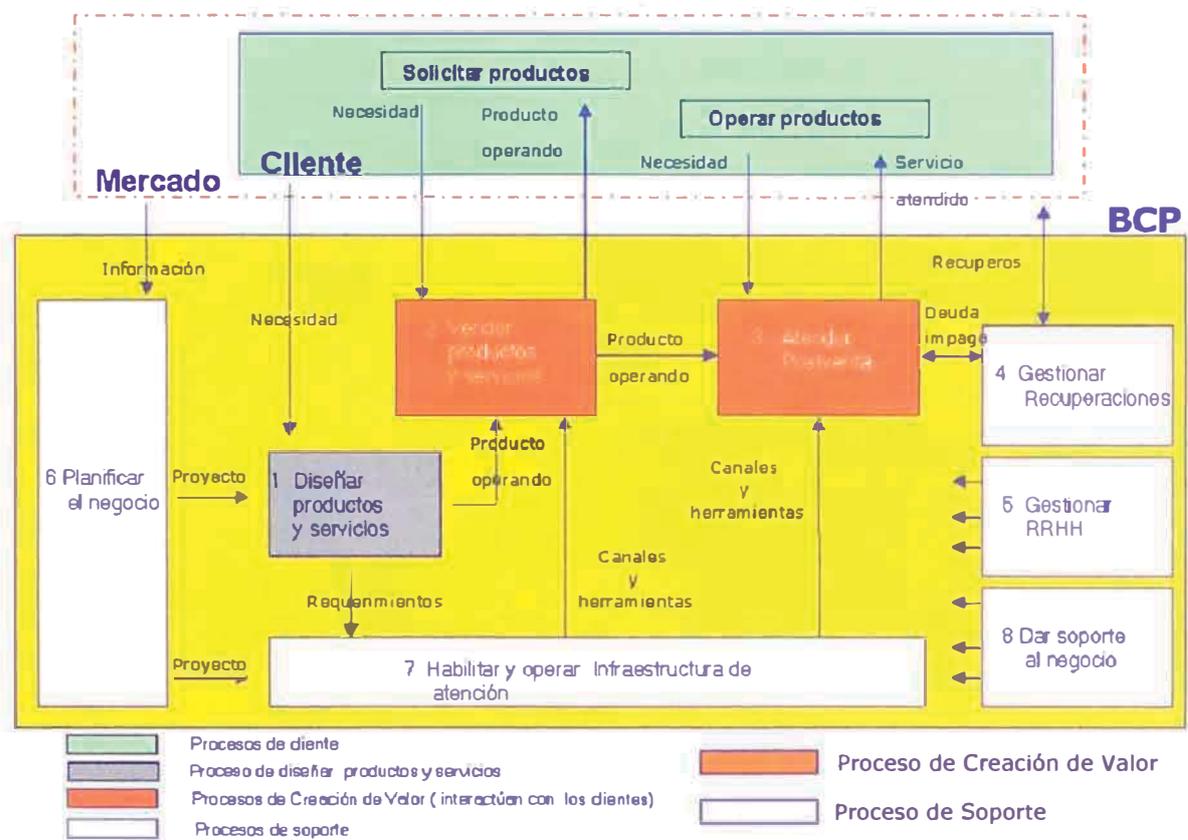


FIGURA 2. Mapa de Procesos

## MODELO DE TRABAJO

a) Tiene la finalidad que los Dueños de Procesos tengan un esquema de trabajo definido para su óptima gestión basado en equipos multidisciplinarios con personal que participa en el proceso. El Modelo de trabajo se encuentra definido en la Figura 3, teniendo los siguientes componentes:

- Marco de trabajo de diseño de procesos:
- Marco de Trabajo de Mejora Continua:
- Marco de trabajo de rediseño de procesos:

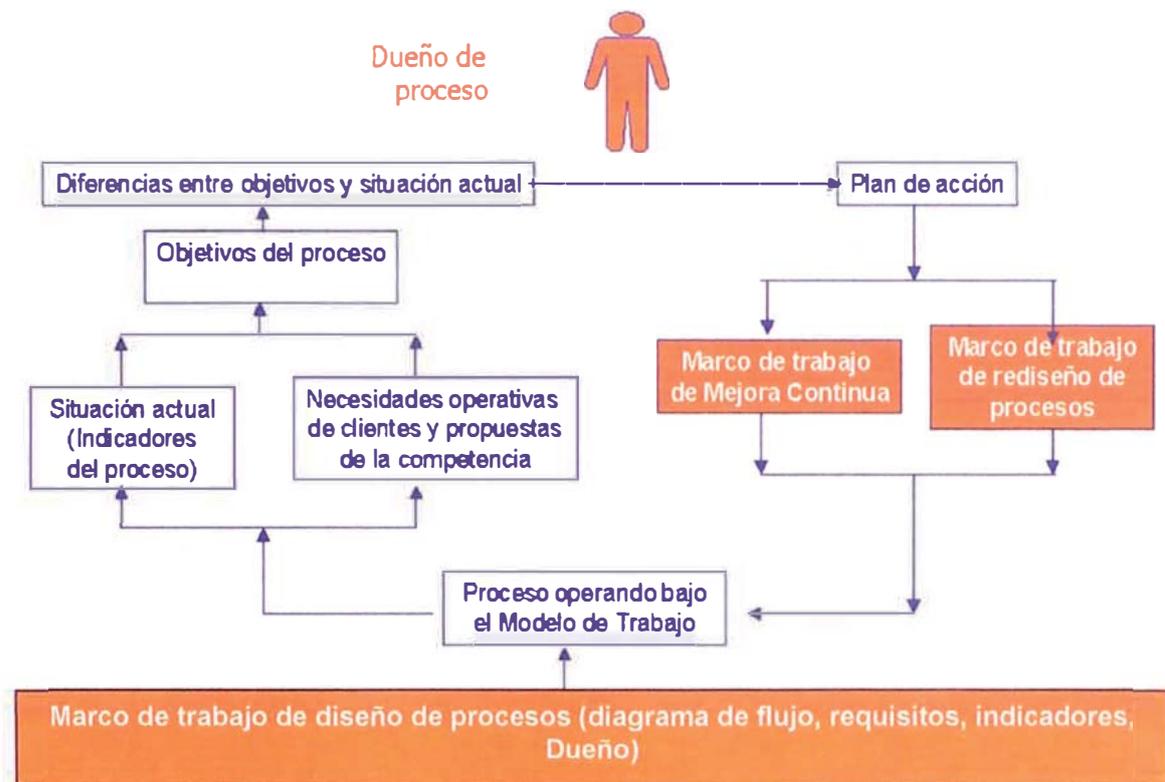


FIGURA 3. Modelo del Trabajo.

## I.2.4 ORGANIZACIÓN

Debido a la envergadura de la empresa está dividida en los siguientes niveles organizativos.

- a) **Directorio:** Unidad de mayor jerarquía y tendrá la responsabilidad de representar los intereses del banco.
- b) **Gerencia General:** Unidad de gestión de mayor responsabilidad y tendrá la responsabilidad de establecer los objetivos y lineamientos de políticas generales del Banco.
- c) **Gerencia General Adjunta:** Unidad de gestión que apoyará y garantizará una gestión eficiente y eficaz a nivel banco de la Gerencia General.

- d) **Gerencia Central:** Unidad que reunirá bajo una misma administración, un conjunto amplio de unidades de menor nivel organizativo.
- e) **División:** Unidad de amplia cobertura funcional, que tendrá la responsabilidad de impartir políticas específicas que afectan un mismo ámbito de negocios, operacional o técnico.
- f) **Área:** Unidad que tendrá la responsabilidad de la ejecución y soporte de los negocios, procesos operativos o un determinado aspecto técnico.
- g) **Gerencia:** Unidad de cobertura funcional específica, que será responsable de definir políticas directamente referidas a un ámbito de negocios, operacional o técnico.
- h) **Servicio:** Unidad de ejecución en un contexto geográfico, segmento, proceso operativo o aspecto técnico específico dentro del ámbito de autoridad de un Área.
- i) **Departamento:** Unidad de ejecución y soporte en tareas y actividades específicas dentro del ámbito de autoridad de una unidad.
- j) **Sección:** Unidad básica de ejecución, constituida por un grupo de personas de línea, apoyo o técnico dentro del ámbito de autoridad de una unidad de ejecución y que puede tener un funcionamiento permanente, periódico o coyuntural.

En la figura 4 que se presenta en la siguiente página aparece el organigrama del Banco del cual se muestran sus unidades más importantes.

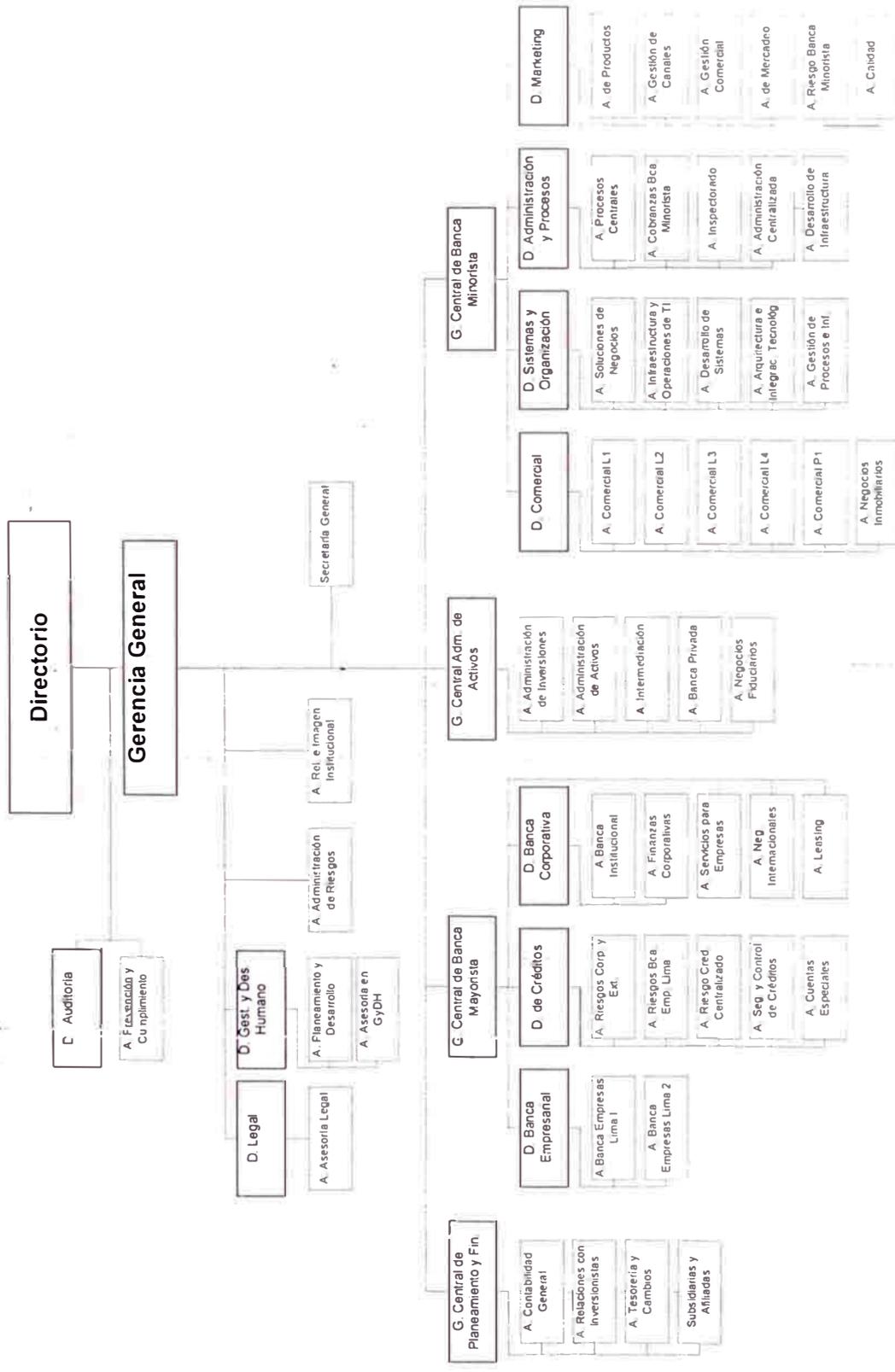


Figura 4: Organigrama

## **I.2.5 UNIDADES ORGANIZATIVAS**

El presente proyecto, se realizó en la División de Sistemas y Organización, específicamente en el Área de Gestión de Procesos e Información, Departamento de Extracción, Transformación y Carga.

A continuación mostramos las unidades organizativas correspondientes que formaron parte del proyecto:

### **División Sistemas y Organización**

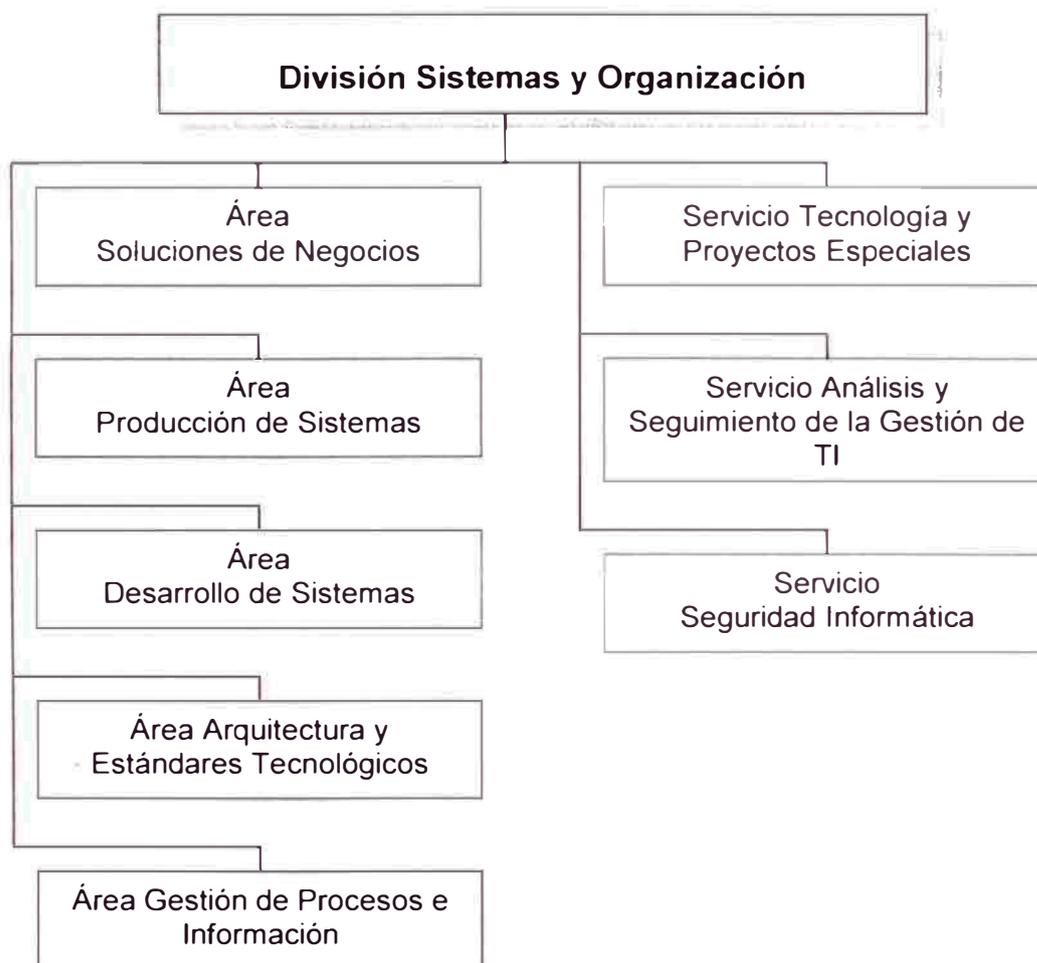
Cuya misión es *“Ser el socio tecnológico y habilitador estratégico de los usuarios de la Organización”*.

La División Sistemas y Organización reporta a la Gerencia Central de Banca Minorista.

Reportan a la Gerencia de la División Sistemas y Organización, las siguientes Áreas y Servicios:

- Área Soluciones de Negocios.
- Área Producción de Sistemas.
- Área Desarrollo de Sistemas.
- Área Arquitectura e Integración Tecnológica.
- Área Gestión de Procesos e Información.
- Servicio Tecnología y Proyectos Especiales.
- Servicio Análisis y Seguimiento de la Gestión de TI.
- Servicio de Seguridad Informática.

En la figura 5 se muestra el organigrama de la División Sistemas y Organización.



**FIGURA 5. ORGANIGRAMA DE LA DIVISION SISTEMAS Y ORGANIZACION**

### Área Gestión de Procesos e Información

Cuya misión es apoyar las necesidades del negocio desde un punto de vista operativo y comercial:

- Facilitando una organización orientada a procesos simples

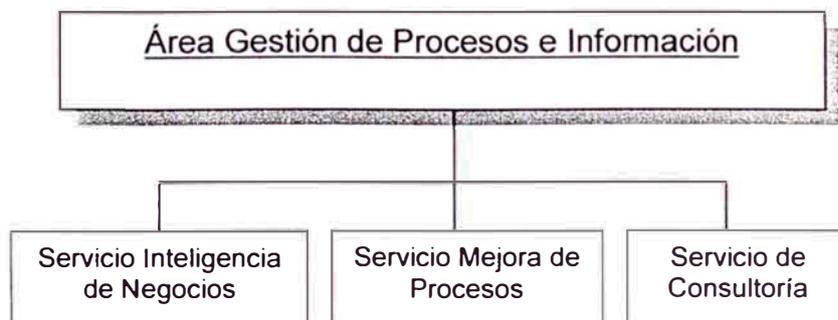
- Gerenciando los activos de información, promoviendo el uso de información altamente relevante, integrada y confiable, identificando oportunamente las necesidades de información de la organización y aplicando de manera eficiente y efectiva los fundamentos de Inteligencia de Negocios.

Reporta a la Gerencia de la División Sistemas y Organización.

Reportan al Área Gestión de Procesos e Información las siguientes unidades:

- Servicio de Inteligencia de Negocios.
- Servicio Mejora de Procesos.
- Servicio de Consultoría.
- Departamento Procesos de la División Sistemas y Organización.

En la figura 6 se muestra el organigrama del Área Gestión de Procesos e Información.



**FIGURA 6.** ORGANIGRAMA DEL AREA GESTIÓN DE PROCESOS E INFORMACIÓN

### Servicio de Inteligencia de Negocios

Las principales funciones del servicio son:

- a) Definir y difundir las políticas de Governance de Inteligencia de Negocios, para orientar los recursos del Servicio de modo eficiente de acuerdo a la estrategia de la Organización.
- b) Definir los lineamientos y políticas que deben cumplir los proyectos de inteligencia de negocios.
- c) Identificar y proponer estándares que permitan integrar eficientemente la información de la Organización.
- d) Integrar la información con visión corporativa..
- e) Identificar necesidades actuales y futuras de información, asesorando a los usuarios para el mejor uso de la información.
- f) Definir los procesos de información y las herramientas y tecnologías adecuadas que aseguren la calidad y oportunidad de la información, así como la eficiencia del Banco.
- g) Velar por la correcta aplicación de las políticas, arquitectura, seguridad y estándares de integración de información corporativa de la Organización.
- h) Diseñar los mecanismos que aseguren un adecuado nivel de servicios de información a los usuarios.
- i) Proporcionar una visión común a los diferentes equipos y miembros del Servicio, alineando sus actividades a la estrategia de la Organización.
- j) Autorizar el acceso sobre las tablas del "Datawarehouse".

El Servicio de Inteligencia de Negocios está organizado como muestra la Figura 7:



**FIGURA 7. ORGANIGRAMA DEL SERVICIO DE INTELIGENCIA DE NEGOCIOS**

*Departamento de Extracción, Transformación y Carga (ETL)*

Principales funciones:

- a) Atender los requerimientos de incorporación de nueva información al Datawarehouse, cumpliendo con la metodología, métricas, herramientas y estándares definidos.
- b) Proveer de información sobre el costo, esfuerzo y recursos necesarios asociados al diseño funcional, diseño técnico, desarrollo y puesta en producción de soluciones de información.
- c) Efectuar el análisis y diseño funcional y técnico de los proyectos a su cargo.

- d) Desarrollar, probar unitariamente, implantar y documentar los desarrollos realizados dentro del tiempo y presupuesto pactado asegurando una óptima calidad.
- e) Gestionar el proceso de carga de información al Datawarehouse según la metodología definida, manteniendo informada a la gerencia y a las unidades involucradas sobre los avances y cumplimientos de los compromisos establecidos.
- f) Apoyar la implementación de los estándares de arquitectura de los sistemas.
- g) Definir procesos eficientes de carga de información al Datawarehouse.

#### Departamento de Explotación de Información

Principales funciones:

- a) Recepcionar los requerimientos de información del Datawarehouse y canalizar su correcta y eficiente atención.
- b) Efectuar el análisis y diseño funcional y técnico de las aplicaciones a su cargo.
- c) Desarrollar, probar unitariamente, implantar y documentar las aplicaciones a su cargo, dentro del tiempo y presupuesto pactado asegurando una óptima calidad.
- d) Analizar los requerimientos de información del Negocio y ofrecer soluciones integrales de información.

- e) Dotar a las áreas de negocios de información múltiple (estadística, financiera, estratégica, etc.), de fácil acceso y homogénea, relacionada a sus clientes y necesaria para la toma de decisiones.
- f) Facilitar y promover el uso de información operativa, táctica y estratégica para la toma de decisiones y para la generación de acciones de negocio dentro de la organización.
- g) Proveer soporte a los usuarios en referencia al uso adecuado y a la explotación de las facilidades de los sistemas.
- h) Capacitar en el uso de nuevas herramientas a todos los potenciales analistas de información del Banco.
- i) Difundir el uso de modelos estadísticos dentro de la Organización.
- j) Difundir el uso de la información integrada y corporativa dentro de la Organización.
- k) Monitorear continuamente el uso de información y proponer acciones de mejora.

#### Departamento de Actualización Datawarehouse

Principales funciones:

- a) Asegurar el funcionamiento óptimo del Datawarehouse y el cumplimiento de los niveles de servicio acordados.
- b) Monitorear continuamente el uso de los recursos asignados al Datawarehouse.

- c) Atender oportunamente cualquier problema con los procesos de información del Datawarehouse.
- d) Velar por el cumplimiento de los estándares establecidos.
- e) Definir estándares que aseguren la continuidad operativa del Datawarehouse.
- f) Asegurar la integridad del modelo de datos del Datawarehouse, con el objetivo de asegurar el uso eficiente de los recursos, considerando los diferentes requerimientos de información de las unidades.
- g) Realizar labores de actualización de información en el Datawarehouse que se consideren impostergables.
- h) Gestionar y asegurar la correcta ejecución de los planes de contingencia de recuperación de información en caso se presenten problemas en el Datawarehouse.

#### Departamento de Calidad de Datos

Principales funciones:

- a) Liderar iniciativas para mejorar la calidad de la información del Datawarehouse, comunicando a los responsables los errores o inconsistencias identificados en el análisis de la información y actuando como facilitador para su pronta solución.
- b) Definir procedimientos y mejores prácticas que permitan lograr mejoras en la calidad de la información del Datawarehouse.
- c) Realizar el análisis de datos de la información del Datawarehouse.

- d) Realizar análisis de calidad de información de áreas específicas para conocer su situación y establecer planes de acción.
- e) Liderar iniciativas de mejoras de la calidad de información en la organización, involucrando a las unidades impactadas, cuantificando los beneficios y estableciendo planes de acción.
- f) Definir, proponer e implementar métricas para la medición de la calidad de datos.
- g) Evaluar la calidad de los datos y programar el monitoreo continuo mediante métodos de muestreo, análisis de relaciones, análisis estadístico y otros que se determinen.
- h) Definir y administrar el uso de herramientas de calidad de datos que considere necesaria para sus labores.
- i) Definir y difundir las políticas de Governance de Datos de la Organización.

## **CAPITULO II**

### **MARCO TEORICO**

#### **II.1 INTELIGENCIA DE NEGOCIOS**

Es un conjunto de conceptos y herramientas para facilitar el uso de la información por parte de las organizaciones con el objetivo de tomar mejores decisiones y alcanzar así mayores eficiencias y beneficios financieros.

No es un sistema, no es un producto. Es una disciplina cross-organizacional y una arquitectura empresarial para integrar operaciones y dar soporte a la toma de decisiones de las aplicaciones y base de datos, las cuales proveen a la comunidad del negocio un fácil acceso a la data de la empresa y permite realizar buenas decisiones.

En la figura 8 podemos ver cómo los procesos de Inteligencia de negocios se orientan a entregar información en todos los niveles de la organización.



**FIGURA 8. NIVELES DE INFORMACIÓN EN EL PROCESO DE INTELIGENCIA DE NEGOCIOS**

## II.2 DATAWAREHOUSE & DATAWAREHOUSING

### DATAWAREHOUSE

Un Datawarehouse (DWH) es un repositorio central que contiene toda la información valiosa de la empresa. Los datos que aquí se almacenan han pasado por un proceso de calidad que asegura su consistencia. Además, el repositorio está construido de tal manera que el acceso sea lo más rápido posible.

*Un Datawarehouse es “un conjunto de datos integrados orientados a temas, que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración”*

*W.H.Inmon<sup>2</sup>*

Un Datawarehouse es una colección estratégica de todos los tipos de datos para soportar el proceso de toma de decisiones en todos los niveles de la organización.

*Un datawarehouse es “una copia de los datos transaccionales específicamente estructurados para consultas u análisis”.*

*Ralph Kimball<sup>3</sup>*

Kimball dice que su construcción se va haciendo por etapas que normalmente corresponden a las principales área operativas de la empresa. Por ejemplo: área de Ventas, área Financiero Contable, área de Recursos Humanos, etc.

Los datawarehouses (base de datos OLAP, online analytical processing) son diseñados para cumplir con un conjunto de metas, las cuales son muy diferentes de los objetivos de un sistema transaccional (OLTP, online transaction processing). Por ejemplo, una meta de los OLTP es maximizar la concurrencia mediante el uso de locks, dicho objetivo no es pertinente en el

---

<sup>2</sup> DataPrix: <http://www.dataprix.com/es/data-warehouse>

<sup>3</sup> Kimball Group: <http://www.kimballgroup.com/>

diseño de DW donde las consultas son sólo del tipo SELECT del SQL, es decir, para leer datos.

Ralph Kimball es un referente de la metodología dimensional para diseñar grandes datawarehouses, fue el que realmente explotó al máximo el tema de datawarehousing.

## **DATAWAREHOUSING**

Un datawarehouse es un lugar, el datawarehousing es un proceso.

Existe una gran cantidad de procesos detrás de una arquitectura de datawarehouse de suma importancia para el mismo. Estos comprenden desde procesos de extracción que estudian y seleccionan los datos fuente adecuados para el datawarehouse hasta procesos de consulta y análisis de datos que despliegan información de una forma fácil de interpretar y analizar.

A continuación explicaremos los procesos básicos de un datawarehouse:

**Extracción:** consiste en estudiar y entender los datos fuente, tomando aquellos que son de utilidad para el datawarehouse.

**Transformación:** una vez que los datos son extraídos, éstos se transforman. Este proceso incluye corrección de errores, resolución de problemas de dominio, borrado de campos que no son de interés, generación de claves, agregación de información, etc.

**Carga:** Al terminar el proceso de transformación, se cargan los datos en el datawarehouse.

**Calidad de datos:** una vez ingresada la información al datawarehouse, se realizan controles de calidad para asegurar que la misma sea correcta.

**Publicación:** Cuando la información se encuentra disponible, se les informa a los usuarios. Es importante publicar todo cambio que se haya realizado.

**Consulta:** el usuario final debe disponer de herramientas de consulta y procesamiento de datos. Este proceso incluye consultas ad hoc, reportes, aplicaciones DSS, datamining, etc.

**Retroalimentación:** muchas veces es aconsejable seguir el camino inverso de carga. Por ejemplo, puede alimentarse los sistemas legales con información depurada del datawarehouse o almacenar en el mismo alguna consulta generada por el usuario que sea de interés.

**Auditoria:** los procesos de auditoría permiten conocer de donde proviene la información así como también qué cálculos la generaron.

**Seguridad:** una vez construido el datawarehouse, es de interés para la organización que la información llegue a la mayor cantidad de usuarios pero, por otro lado, se tiene sumo cuidado de protegerlo contra posible hackers, snoopers o espías. El desarrollo de Internet ha incrementado este dilema.

**Respaldo y Recuperación:** se deben realizar actividades de respaldo y restauración de la información, tanto la almacenada en el datawarehouse, como la que circula desde los sistemas fuente al datawarehouse.

## **DATAMART**

Un Datamart es un datawarehouse más pequeño. Normalmente la información contenida en un Datamart es un subconjunto de un datawarehouse.

Entre las principales características de un Datamart tenemos:

- Orientado a un área o departamento dentro de la organización.
- Puede ser implementado como una solución para problemas inmediatos.
- No es necesario para construir un datawarehouse.
- Los beneficios de un Datamart son los siguientes:
  - a. Rápida y sencilla implementación.
  - b. Menor costo de implementación que un datawarehouse.
  - c. Cubre necesidades específicas del negocio.
  - d. Rápidas respuestas por el menor volumen de información.
  - e. Asegura la consistencia de datos.
- Las desventajas principales de un Datamart son:
  - a. Inadvertidamente se pueden usar datos no compatibles con otros Datamarts que luego alargue el tiempo de unificación.
  - b. Si el datawarehouse es construido primero, se requiere de hardware adicional para soportar datamarts individuales.
  - c. Datos centralizados.

### **II.3 EXTRACCIÓN, TRANSFORMACIÓN y CARGA (ETL)**

ETL son las siglas en inglés de Extraer, Transformar y Cargar (Extract, Transform and Load). Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, datamart, o Datawarehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio.

Los procesos ETL también se pueden utilizar para la integración con sistemas heredados.

### **Extraer**

La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. La mayoría de los proyectos de almacenamiento de datos fusionan datos provenientes de diferentes sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos distintos. Los formatos de las fuentes normalmente se encuentran en bases de datos relacionales o ficheros planos, pero pueden incluir bases de datos no relacionales u otras estructuras diferentes. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

Una parte intrínseca del proceso de extracción es la de analizar los datos extraídos, de lo que resulta un chequeo que verifica si los datos cumplen la pauta o estructura que se esperaba. De no ser así los datos son rechazados.

Un requerimiento importante que se debe exigir a la tarea de extracción es que ésta cause un impacto mínimo en el sistema origen. Si los datos a extraer son muchos, el sistema de origen se podría sobrecargar e incluso colapsar, provocando que éste no pueda utilizarse con normalidad para su uso cotidiano. Por esta razón, en sistemas grandes las operaciones de extracción suelen programarse en horarios o días donde este impacto sea nulo o mínimo.

## Transformar

La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas fuentes de datos requerirán alguna pequeña manipulación de los datos. No obstante en otros casos pueden ser necesarias aplicar algunas de las siguientes transformaciones:

- Seleccionar sólo ciertas columnas para su carga (Ej. que las columnas con valores nulos no se carguen).
- Traducir códigos (Ej. Si la fuente almacena una "H" para Hombre y "M" para Mujer pero el destino tiene que guardar "1" para Hombre y "2" para Mujer).
- Codificar valores libres (Ej. convertir "Hombre" en "H" o "Sr" en "1").
- Obtener nuevos valores calculados (Ej.  $\text{total\_venta} = \text{cantidad} * \text{precio}$ ).
- Unir datos de múltiples fuentes (Ej. búsquedas, combinaciones, etc).
- Calcular totales de múltiples filas de datos (Ej. ventas totales de cada región).
- Generación de campos clave en el destino.
- Transponer o pivotar (girando múltiples columnas en filas o viceversa).
- Dividir una columna en varias (Ej. columna "Nombre: García, Miguel"; pasar a dos columnas "Nombre: Miguel" y "Apellido: García").

La aplicación de cualquier forma, simple o compleja, de validación de datos, y la consiguiente aplicación de la acción que en cada caso se requiera:

- Datos OK: Entregar datos a la siguiente etapa (Carga).
- Datos erróneos: Ejecutar políticas de tratamiento de excepciones (Ej. Rechazar el registro completo, dar al campo erróneo un valor nulo o un valor centinela).

### **Carga**

La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con nuevos datos. Los Datawarehouse mantienen un historial de los registros de manera que se pueda hacer una auditoría de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo.

Existen dos formas básicas de desarrollar el proceso de carga:

**Acumulación simple:** La acumulación simple es la más sencilla y común, y consiste en realizar un resumen de todas las transacciones comprendidas en el período de tiempo seleccionado y transportar el resultado como una única transacción hacia el Datawarehouse, almacenando un valor calculado que

consistirá típicamente en un sumatorio o un promedio de la magnitud considerada.

Rolling: El proceso de Rolling, se aplica en los casos en que se opta por mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna o varias de las dimensiones de la magnitud almacenada (Ej. Totales diarios, totales semanales, totales mensuales, etc.).

La fase de carga interactúa directamente con la base de datos de destino. Al realizar esta operación se aplicarán todas las restricciones y triggers (disparadores) que se hayan definido en ésta (Ej. valores únicos, integridad referencial, campos obligatorios, rangos de valores). Estas restricciones y trigger (si están bien definidos) contribuyen a que se garantice la calidad de los datos en el proceso ETL, y deben ser tenidos en cuenta.

En la figura 9 podemos observar cómo el proceso de ETL interviene dentro del proceso de Inteligencia de Negocios en todas sus etapas:

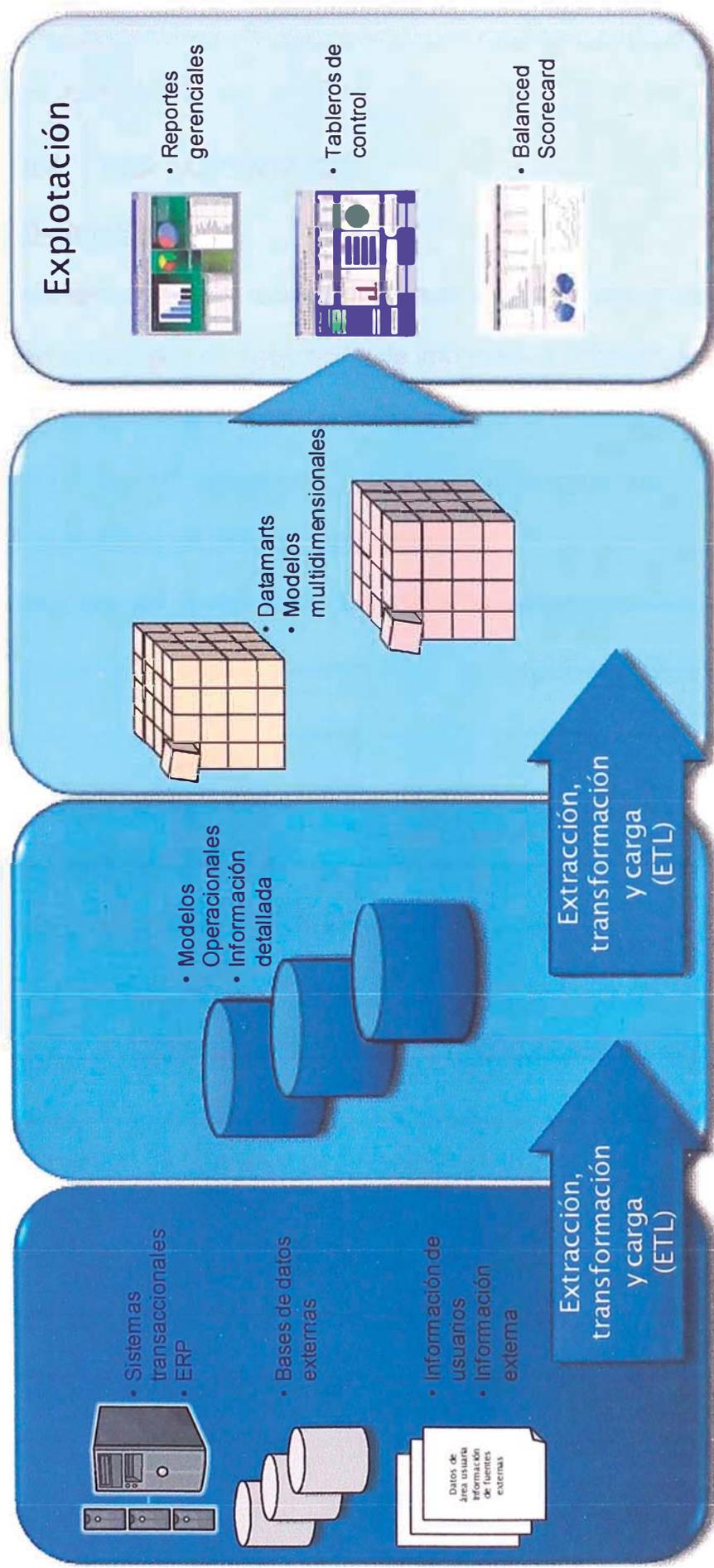


FIGURA 9. PROCESO DE INTELIGENCIA DE NEGOCIOS

## II.4 HERRAMIENTA ETL

### Definición

Herramientas que facilitan las tareas de extracción, transformación y carga en escenarios de integración de información sobre diferentes infraestructuras.

### Utilidad de dichas herramientas:

Inteligencia de Negocios (BI) y de almacenamiento de datos (DWH)



FIGURA 10. PROCESO DE ETL PARA INTELIGENCIA DE NEGOCIOS

Creación de Maestros de Almacenamiento Integrados (Hub Clientes)

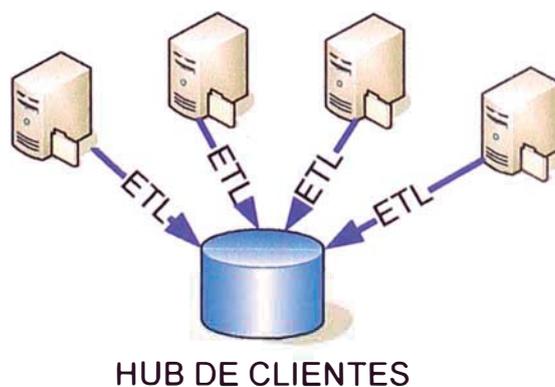


FIGURA 11. ESQUEMA DE HUB DE CLIENTES

## Migración de Aplicaciones

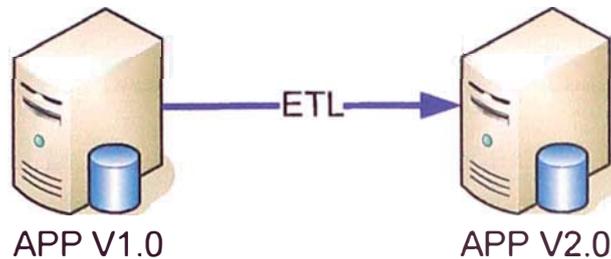


FIGURA 12. ESQUEMA DE MIGRACIÓN DE APLICACIONES

Sincronización de información entre las aplicaciones operacionales  
(Sistemas Transaccionales - Procesos Batch - SOA)

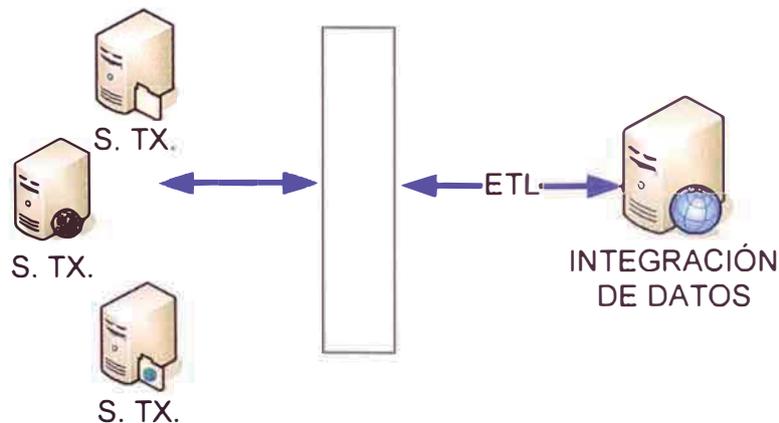


FIGURA 13. ESQUEMA SOA

## II.5 PROCESO DE ATENCIÓN DE REQUERIMIENTOS (PAR)

- Es un método de trabajo de atención de requerimientos de los usuarios utilizado por la División de Sistemas del Banco, fruto de la identificación y formalización de las mejores prácticas de la división basado en CMMI.

- Es una manera de alinear y sincronizar los roles y sus actividades dentro de las distintas fases de la atención de un requerimiento.
- No es un proceso prescriptivo, requiere de una actitud proactiva.
- El PAR es el proceso estándar utilizado por la División de Sistemas que todos los requerimientos deben ejecutar para atender los pedidos de los usuarios.
- El PAR cuenta con guía de adecuación que, a partir del tipo de requerimiento a implementar, especifican las actividades que deben ejecutarse así como los criterios a considerar para esta decisión.
- Debido a que el Banco se encuentra en proceso de obtener la Certificación CMMI Nivel 3, el PAR ha incluido y adaptado las áreas proceso respectivas dentro del mismo.

**Fases:**

El Proceso de Atención de Requerimientos separa los requerimientos en las siguientes fases:

1. Análisis de la Necesidad
2. Planificación Periódica
3. Evaluación de la Metodología del Proyecto
4. Análisis y Diseño Funcional
5. Análisis y Diseño Técnico
6. Construcción
7. Certificación
8. Producción

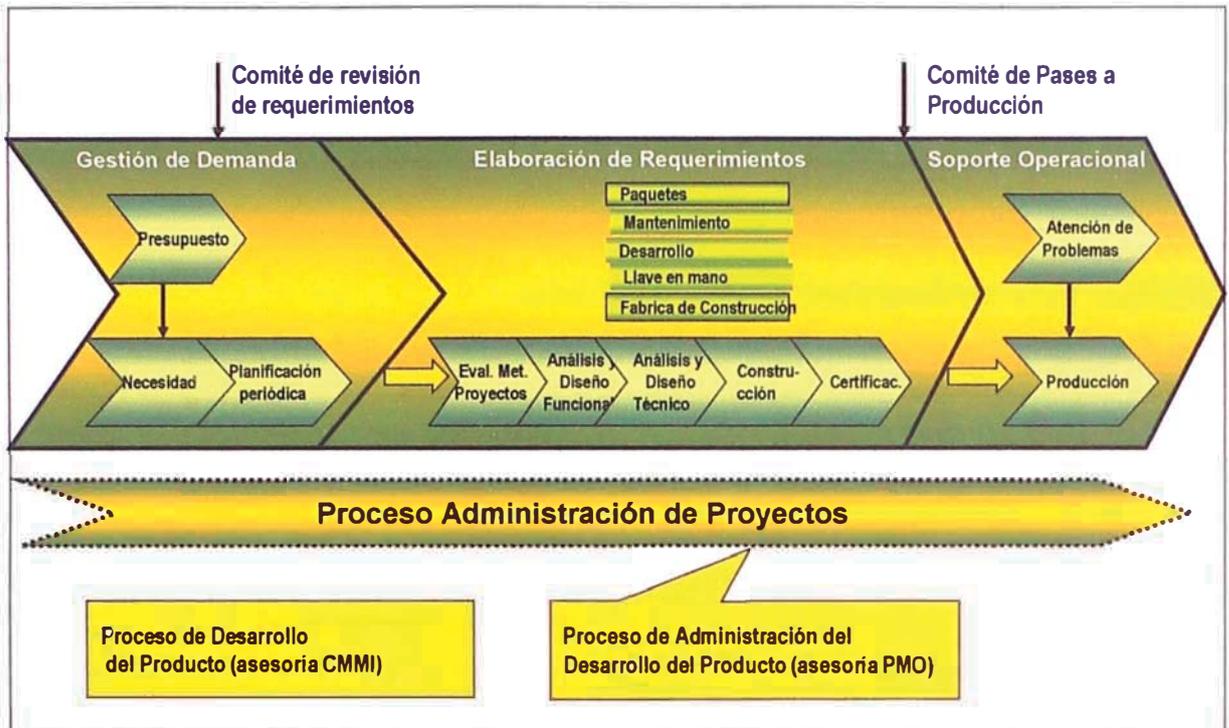


Figura 14: PROCESO DE ATENCION DE REQUERIMIENTOS

## II.6 GERENCIA DE PROYECTOS

La Gerencia de Proyectos es la aplicación de conocimientos, habilidades, herramientas y técnicas a las actividades de un proyecto para atender los requerimientos del mismo. La administración del proyecto se basa en PMO.

La figura 15 muestra como interactúa el proceso de atención de requerimientos (PAR) con el modelo de gestión de proyectos:

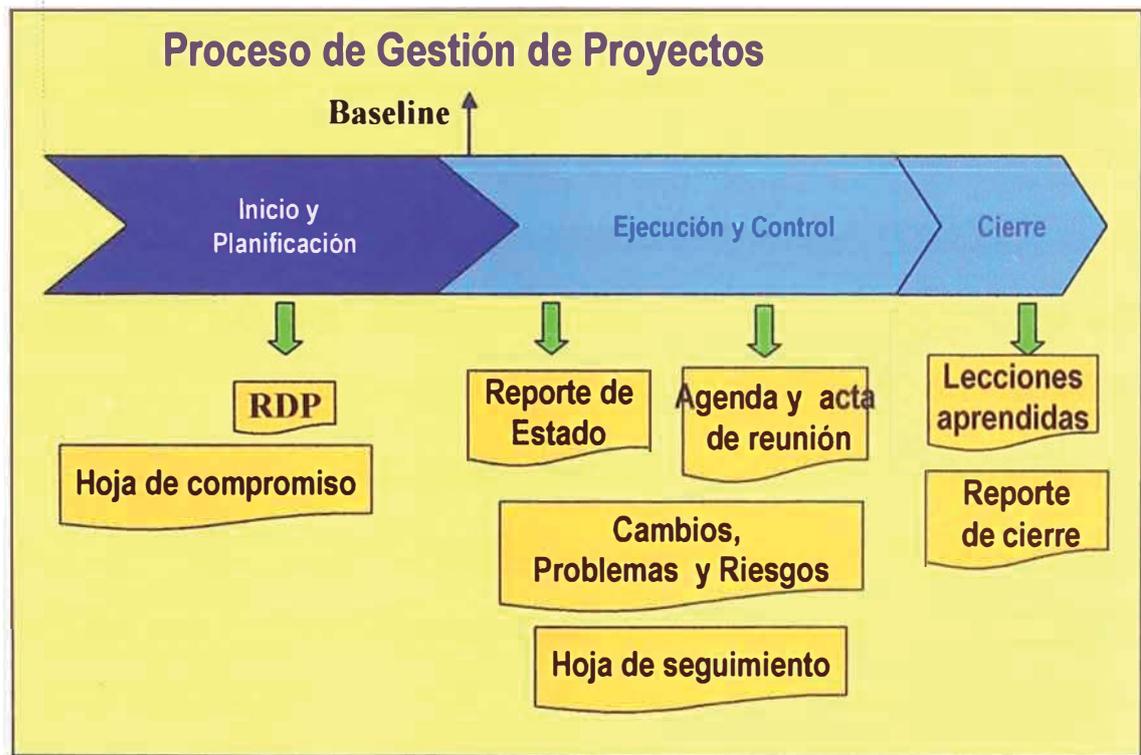


Figura 15: Flujo PAR y de gestión de Proyectos.

El BCP ha desarrollado como adecuar la metodología PMO al desarrollo de requerimiento CMMI, Figura 10.

Proyectos se desarrollan en los sectores de construcción, ingeniería de software, energía y minas, textil, defensa, aeronáutica, naval, mecánica, transporte, telecomunicaciones, educación, etc. y en casi cualquier área del conocimiento humano.

## **Proyecto**

Es un suceso único, no repetitivo, con inicio y fin determinados, formalmente organizado y que congrega y emplea recursos para lograr resultados preestablecidos.

Un proyecto tiene las siguientes características:

- Es de carácter temporal.
- Está formado por una serie de tareas claramente relacionadas.
- Requiere múltiples recursos.
- Está dirigido a la consecución de un objetivo preestablecido y cuantificable.

## **CAPITULO III**

### **PROCESO DE TOMA DE DECISIONES**

#### **III.1 PLANTEAMIENTO DEL PROBLEMA**

Para el planteamiento del problema, primero debemos identificar los objetivos estratégicos del servicio, los cuales nos servirán de base para identificar el problema que vamos a resolver, estos objetivos estratégicos son:

- Optimizar la gestión de BI y dar valor al negocio
- Contar con un modelo de datos corporativo y de clase mundial
- Desarrollar soluciones con vista a las necesidades de la organización
- Reducir/Eliminar inconsistencia de Datos
- Mejorar los tiempos de atención de requerimientos de información, sin perder la consistencia y oportunidad

Sobre el último objetivo estratégico, nos centraremos para el desarrollo del proyecto materia de este informe.

### III.1.1 SITUACIÓN ACTUAL DEL PROCESO DE DESARROLLO ETL

Dentro de la implementación de un proyecto de Inteligencia de Negocios, se tiene varias etapas, de acuerdo a la metodología PAR (Proceso de Atención de Requerimientos). En base a la historia se tiene que la etapa que ocupa el mayor porcentaje dentro del tiempo del proyecto es la construcción (43.2%), tal como se muestra en el cuadro 1:

FASE	% TIEMPO
Evaluación de Metodología de Proyecto	4.8%
Diseño Funcional	9.6%
Diseño Técnico	14.4%
Construcción	43.2%
Certificación	23.1%
Pase a Producción	4.8%

**Cuadro 1:** Porcentaje de tiempo por fase en un requerimiento

Para los proyectos de Inteligencia de Negocios, este 43,2% representa el tiempo de desarrollo de los procesos de Extracción, Transformación y Carga (ETL).

Actualmente, el desarrollo de estos procesos se realiza de forma manual, es decir, los programas son codificados manualmente, de acuerdo a la casuística encontrada para cada fuente en particular. Si bien es cierto que se tienen estándares y mejores prácticas de desarrollo, actualmente este proceso presenta una serie de problemáticas, las que hacen que el proceso

tome demasiado tiempo, lo cual origina que el servicio de Inteligencia de Negocios no pueda atender tantos requerimientos como la organización demanda.

Un factor importante que evidencia esta problemática aun más, es el incremento en la demanda de requerimientos de Inteligencia de negocios, dado que el negocio está creciendo y en constante cambio para adaptarse al entorno competitivo que lo rodea, esto hace un más necesaria la mejora del tiempo de atención de los requerimientos.

### **III.1.2 PRINCIPALES PROBLEMAS EN EL PROCESO ETL**

Entre los principales problemas del proceso de desarrollo actual, tenemos:

- No existe un proceso estandarizado ni automatizado para el desarrollo de los procesos de Extracción, Transformación y Carga (ETL)
- El desarrollo de procesos es de forma manual
- Los tiempos de desarrollo de los procesos ETL son altos
- Proceso de pase a producción no está optimizado, por lo que existen muchos errores de programación
- Tiempos de soporte y mantenimiento de demasiado altos

El problema que se abordará, es el primero: “No existe un proceso estandarizado ni automatizado para el desarrollo de los procesos de Extracción, Transformación y Carga (ETL)”

En base a la solución de este problema se plantea el proyecto:

“Implementación de una Herramienta de Extracción, Transformación y Carga (ETL) para mejorar el tiempo de atención de los requerimientos de Inteligencia de Negocios”

### III.2 ALTERNATIVAS DE SOLUCIÓN

Para la solución del problema seleccionado se plantearon 2 alternativas:

#### **Alternativa 1 (Optimizadora)**

Desarrollar una herramienta ETL a medida, para estandarizar y automatizar el proceso de Extracción, Transformación y Carga.

La herramienta a desarrollar deberá cubrir ciertas funcionalidades necesarias en los diferentes aspectos que contempla la organización como importantes.

<b>Ventajas</b>
Se puede obtener funcionalidades a medida
Se tiene un alto grado de conocimiento de la metodología actual

<b>Desventajas</b>
Altos riesgos asociados por la complejidad del desarrollo requerido
Las necesidades de negocio son altamente cambiantes
Se requiere gran esfuerzo de soporte y mantenimiento

#### **Alternativa 2 (Innovadora)**

Adquirir e implementar una herramienta ETL de clase mundial y aplicar las mejores prácticas que trae la herramienta.

<b>Ventajas</b>
Herramienta de clase mundial, trae mejores prácticas en las organizaciones
Tiempo de implementación de corto plazo
Es flexible y escalable

<b>Desventajas</b>
Es necesario un gran esfuerzo para la migración
Puede presentar resistencia al cambio de forma de trabajo

### III.3 METODOLOGÍA DE EVALUACIÓN DE SOLUCIONES

#### III.3.1 CRITERIOS DE DECISIÓN

Se tomará como criterio de decisión dos evaluaciones, una cuantitativa y otra cualitativa. Estas tendrán un peso como se indica:

Evaluación Cuantitativa (20%)

Evaluación Cualitativa (80%)

A continuación detallamos qué es lo que se tomará en cuenta en las evaluaciones:

#### **Evaluación Cuantitativa**

Los criterios que se tomarán para la evaluación son:

- **Inversión Inicial:** En este punto se considerará la inversión inicial necesaria para la solución propuesta, por ejemplo, adquisición de equipos, personal, licencias.

- Capacitación: En este punto se tomará en cuenta la inversión necesaria para capacitar al personal que utilizara la solución en los diferentes aspectos, como, funcionalidades, operatividad, administración, soporte.
  
- Mantenimiento y Soporte: En este punto se considerará la inversión necesaria para dar soporte y mantenimiento en el tiempo a la herramienta, incluyendo la necesidad de nuevas funcionalidades y el soporte día a día de la solución.

El periodo de tiempo a considerar para la evaluación es de cinco (05) años.

### **Evaluación Cualitativa**

Para la evaluación cualitativa, se tendrán en cuenta los siguientes criterios:

- Funcionalidad: Se tomará en cuenta las capacidades de la solución para satisfacer las necesidades establecidas, así como la facilidad y flexibilidad para hacerlo.
  
- Estándares: Se tomará en cuenta el nivel de cumplimiento de la solución respecto a los estándares y lineamientos de sistemas en la organización.

- Integración: Se cuantificará la capacidad de la solución para interactuar con los demás aplicativos de la organización así como la flexibilidad para interconectarse con otras plataformas.
- Seguridad: Se tomará en cuenta la capacidad de la solución para salvaguardar la información y los niveles de acceso al aplicativo, así como los controles, auditoria y trazabilidad de las actividades de los usuarios.
- Soporte Operativo: Se considerará la rapidez, flexibilidad y eficiencia del soporte de la solución para adaptarse a cualquier entorno cambiante.

Cada uno de estos criterios será evaluado con un puntaje del uno (1) como puntaje más bajo y cinco (5) como puntaje más alto.

Ambas evaluaciones serán realizadas por un equipo experto, el cual estará conformado por los principales roles que intervienen en el desarrollo de los proyectos, estos se muestran en el cuadro 2:

Rol	Área
Arquitecto de sistemas	Arquitectura
Jefe de departamento de Infraestructura de sistemas	Producción
Jefe de departamento de Operaciones de sistemas	Producción
Analista de seguridad informática	Seguridad
Jefe de departamento de ETL	ETL
Jefe de departamento de Actualización DataWarehouse	ETL
Analista Funcional de ETL	ETL
Analista Técnico de ETL	ETL

**Cuadro 2:** Roles participantes en el proceso de evaluación

### III.3.2 DEFINICIÓN DE MÉTRICAS PARA LA TOMA DE DECISIONES

#### ALTERNATIVA 1 (OPTIMIZADORA)

*“Desarrollar una herramienta ETL a medida, para estandarizar y automatizar el proceso de Extracción, Transformación y Carga.”*

#### Evaluación Cuantitativa

Para esta alternativa, se consideró el adquirir el personal necesario para desarrollar la herramienta ETL, proveer la capacitación necesaria y dar el mantenimiento y soporte adecuado. (De acuerdo a los roles que indica el proceso de Atención de requerimientos - PAR),

Se consideró adquirir personal con experiencia para los roles que cumplirían dentro del proyecto, teniendo en cuenta que no se debería impactar la capacidad instalada del Área, ya que actualmente el personal está atendiendo requerimientos.

#### Inversión inicial

En esta estimación se toma en cuenta el personal necesario para el desarrollo, pruebas y pase a producción de la herramienta ETL.

Fase	Rol	Cantidad	Salario mensual promedio (US\$)	Tiempo necesario (Meses)	TOTAL (US\$)
Diseño Funcional	Analista Funcional	3	1,500.00	12	54,000.00
Diseño Técnico	Analista Técnico	3	1,500.00	12	54,000.00
Construcción	Analista Programador I	3	1,200.00	12	43,200.00
	Analista Programador II	4	1,000.00	12	48,000.00
Certificación	Analista de Certificación	2	1,000.00	6	12,000.00
Pase a Producción	DBA	1	1,600.00	12	19,200.00
	Programador	3	1,200.00	6	21,600.00
<b>TOTAL</b>					<b>252,000.00</b>

## Capacitación

En esta estimación se toma en cuenta el personal necesario para realizar la capacitación en todos los aspectos necesarios:

- Capacitación funcional, necesaria para el uso de las funcionalidades del sistema
- Capacitación Técnica, necesaria el detalle de la arquitectura y componentes del sistema.
- Capacitación Operativa, necesaria para conocer el detalle operativo de la herramienta.
- Capacitación para Administración y Soporte, necesaria para conocer la administración y solución de problemas con la herramienta.

Rol	Tipo de Recurso	Cantidad	Salario mensual promedio (US\$)	Tiempo necesario (Meses)	TOTAL (US\$)
Capacitación Funcional	Analista Funcional	1	1,500.00	1	1,500.00
Capacitación Técnica	Analista Técnico	1	1,500.00	1	1,500.00
Capacitación Operativa	Analista Programador I	1	1,200.00	1	1,200.00
	Analista Programador II	1	1,000.00	1	1,000.00
Capacitación para Administración y Soporte	DBA	1	1,600.00	1	1,600.00
	Programador	1	1,200.00	1	1,200.00
<b>TOTAL</b>					<b>8,000.00</b>

## Mantenimiento y soporte

En este tema, se considera el personal necesario para dar mantenimiento y soporte a la herramienta, así como para desarrollar nuevas funcionalidades

requeridas por los usuarios del servicio, de acuerdo con las necesidades cambiantes del negocio. En este aspecto los costos indicados son anuales.

Rol	Tipo de Recurso	Cantidad	Salario mensual promedio (US\$)	Tiempo necesario (Meses)	TOTAL (US\$)
Soporte Funcional	Analista Funcional	1	1,500.00	12	18,000.00
Soporte Técnico	Analista Tecnico	1	1,500.00	12	18,000.00
Soporte Programación	Analista Programador II	2	1,000.00	12	24,000.00
Soporte Producción	DBA	1	1,600.00	12	19,200.00
	Programador	2	1,200.00	12	28,800.00
<b>TOTAL</b>					<b>108,000.00</b>

En resumen, los costos implicados en esta solución son:

Rubro	Total (US\$)
Inversión inicial	252,000.00
Capacitación	8,000.00
Mantenimiento y soporte (Anual)	108,000.00

En base a estos costos, calcularemos la inversión necesaria para la mantenibilidad de la solución en un horizonte de tiempo de 5 años:

Rubro	Costo	Cantidad	TOTAL
Inversión inicial	252,000.00	1	252,000.00
Capacitación	8,000.00	1	8,000.00
Mantenimiento y soporte (Anual)	108,000.00	5	540,000.00
<b>TOTAL</b>			<b>800,000.00</b>

### Evaluación Cualitativa

Tenemos el resultado de la evaluación cualitativa en el siguiente cuadro:

Criterio de Evaluación	Peso	Calificación	Ponderado
Funcionalidad	50%	5	2.5
Estándares	10%	3	0.3
Integración	10%	2	0.2
Seguridad	10%	2	0.2
Soporte Operativo	20%	3	0.6
<b>TOTAL</b>			<b>3.8</b>

## ALTERNATIVA 2 (INNOVADORA)

*“Adquirir e implementar una herramienta ETL de clase mundial y aplicar las mejores prácticas que trae la herramienta.”*

Para esta alternativa, el mercado ofrece una serie de soluciones, entre las cuales tenemos:

Proveedor
1. Data Stage (IBM)
2. Powercenter (Informática)
3. Data Integrator (Business Objects)
4. SAS ETL Enterprise (SAS)
5. Decision Stream (Cognos Inc.)
6. Advantage DT (Computer Associates)
7. Transformation Server (Data Mirror)
8. Co OSystem(AB Initio Software)
9. DB2 Warehouse Manager (IBM)
10. Warehouse Builder (Oracle Corp.)
11. Integration Services (Microsoft)

Para evaluar esta alternativa, se tomó como base las herramientas de clase mundial más reconocidas en el mercado. Para esta elección inicial se consideraron el cuadrante de Gartner y el ranking de Forrester, tal como lo muestran las figuras 11 y 12:



FIGURA 16: Cuadrante de Gartner 2007

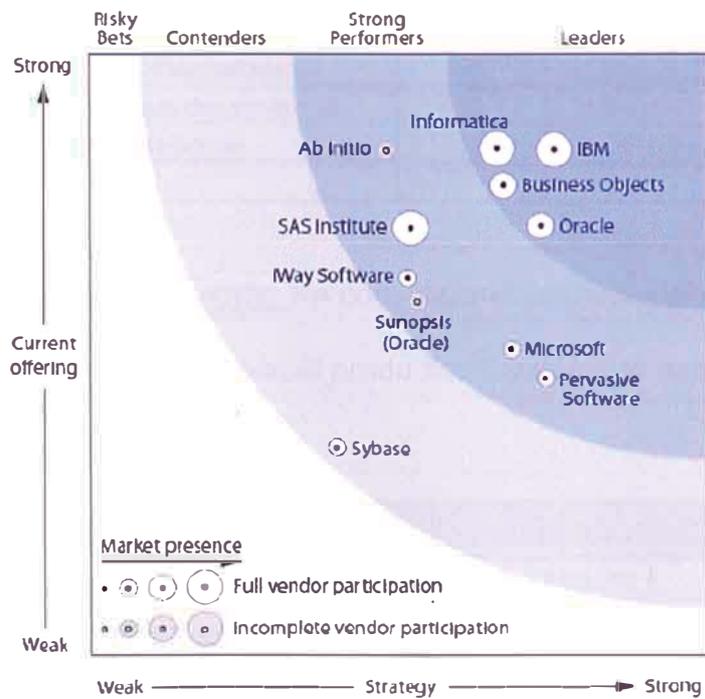


FIGURA 17: Ranking de Forrester 2007

## Evaluación Cuantitativa

### ➤ Inversión Inicial:

Se consideró el costo mayor de las licencias para los servidores y para los clientes:

Concepto	Cantidad	Costo unitario	TOTAL (US\$)
Licencia Servidor de Producción	1	110,000.00	110,000.00
Licencia Servidor de Desarrollo	1	55,000.00	55,000.00
Licencias de aplicativos cliente	30	350.00	24,000.00
<b>TOTAL</b>			<b>189,000.00</b>

➤ Capacitación: Se considero el costo mayor de las capacitaciones tanto básicas como avanzadas de la herramienta

Concepto	Cantidad	Costo unitario	TOTAL (US\$)
Curso básico para desarrolladores	30	1,000.00	30,000.00
Curso avanzado para desarrolladores	30	1,500.00	45,000.00
Curso para administradores	10	1,000.00	10,000.00
<b>TOTAL</b>			<b>85,000.00</b>

➤ Mantenimiento y Soporte: Se considera el costo anual del soporte y mantenimiento que brinda el producto. También se consideró el mayor costo anual.

Concepto	Cantidad	Costo unitario	TOTAL (US\$)
Contrato de mantenimiento y soporte	1	55,000.00	55,000.00
<b>TOTAL</b>			<b>55,000.00</b>

En resumen, los costos implicados en esta solución son:

Rubro	Peso
Inversión inicial	189,000.00
Capacitación	85,000.00
Mantenimiento y soporte (Anual)	55,000.00

En base a estos costos, calcularemos la inversión necesaria para la mantenibilidad de la solución en un horizonte de tiempo de 5 años:

Rubro	Peso	Cantidad	TOTAL
Inversión inicial	189,000.00	1	189,000.00
Capacitación	85,000.00	1	85,000.00
Mantenimiento y soporte (Anual)	55,000.00	5	275,000.00
<b>TOTAL</b>			<b>549,000.00</b>

### Evaluación Cualitativa

Tenemos el resultado de la evaluación cualitativa en el siguiente cuadro:

Criterio de Evaluación	Peso	Calificación	Ponderado
Funcionalidad	50%	5	2.5
Estándares	10%	4	0.4
Integración	10%	4	0.4
Seguridad	10%	4	0.4
Soporte Operativo	20%	4	0.8
<b>TOTAL</b>			<b>4.5</b>

### III.4 TOMA DE DECISIONES

Para la toma de la decisión se tomará el resultado de las evaluaciones tanto cuantitativas como cualitativas, de esta manera se tomara la decisión más adecuada para la solución al problema planteado.

En el siguiente cuadro mostramos los resultados de las evaluaciones de manera comparativa:

Evaluación	Peso	Alternativa 1 (Optimizadora)	Alternativa 2 (Innovadora)
Cuantitativa	20%	3	5
Cualitativa	80%	3.8	4.5
<b>TOTAL</b>		<b>3.64</b>	<b>4.6</b>

Para la Evaluación Cuantitativa, le damos el mayor peso (5) a la Alternativa 2, que es la de menor costo y 3 a la Alternativa 1. A estos pesos lo multiplicamos por 0.2 (20 % de la ponderación total).

A la Evaluación Cualitativa la multiplicamos por 0.8 (80 % de la ponderación total)

Según esta evaluación la alternativa a elegir sería la alternativa 2:

*“Adquirir e implementar una herramienta ETL de clase mundial y aplicar las mejores prácticas que trae la herramienta.”*

Dado que cumple con los requisitos necesarios para la solución al problema planteado.

## **CAPITULO IV**

### **ESTRATEGIAS ADOPTADAS Y EVALUACIÓN DE RESULTADOS**

#### **IV.1 METODOLOGIA DE GESTIÓN DE PROYECTOS**

Una vez establecida la solución a implementar, se dio a inicio al proyecto de “Adquisición e Implementación de una herramienta ETL”

Este proyecto fue llevado a cabo utilizando la metodología PMI. Las fases que se llevaron a cabo son las que se muestran en la figura 18:

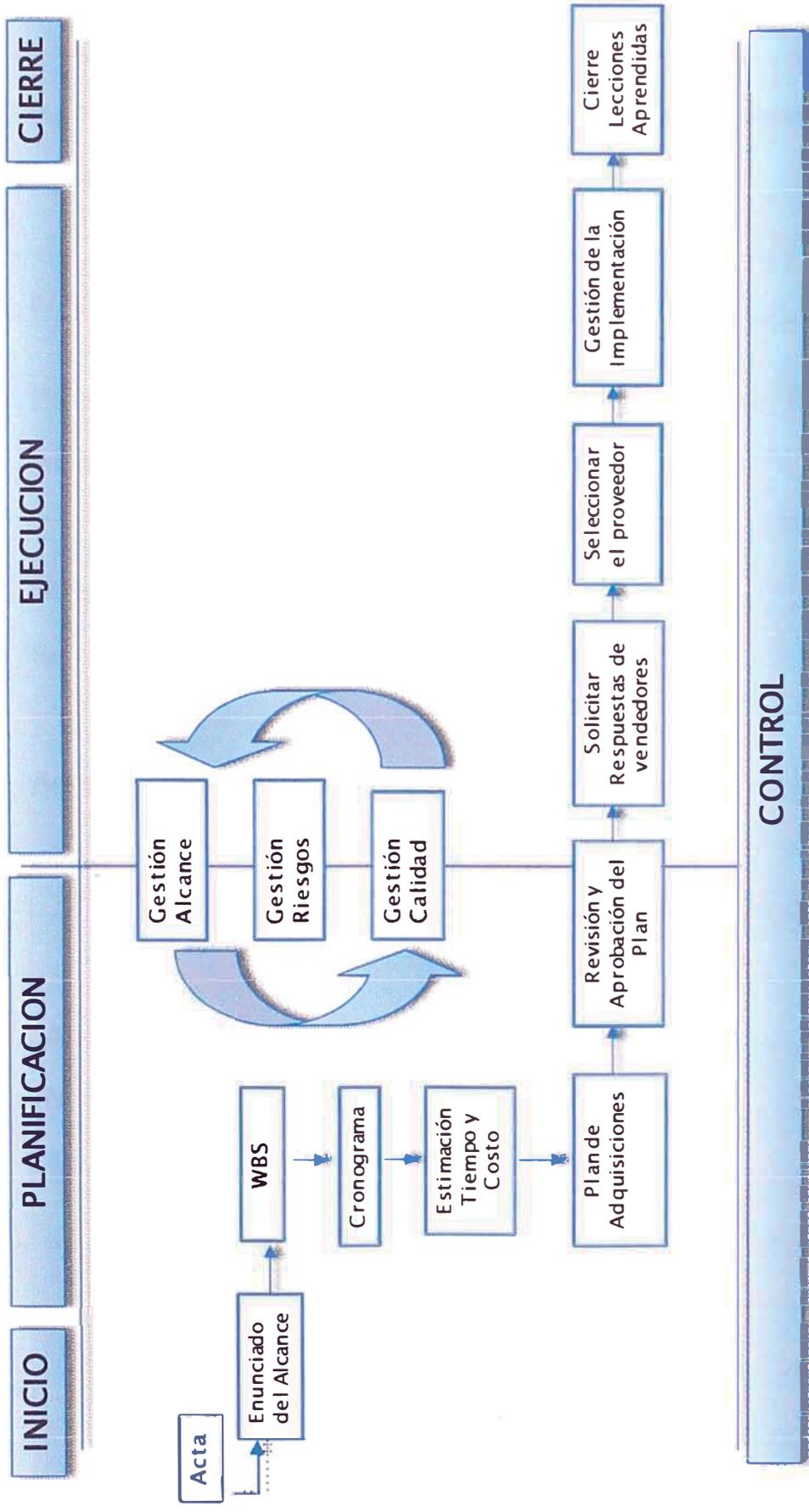


FIGURA 18: Grupos de procesos según PMI

A continuación detallaremos cada una de las fases, y la estrategia adoptada en cada una de ellas.

#### IV.1.1 INICIO

En esta fase se le dio inicio al proyecto, se realizaron dos actividades importantes:

Se elaboró el Acta de constitución del proyecto.

Se elaboró el enunciado del alcance preliminar.

#### Acta de constitución del proyecto

<b>Acta de Constitución del Proyecto (Project Charter)</b>	
<b>A. Antecedentes Generales</b>	
Nombre del Proyecto:	
Adquisición e Implementación de una herramienta ETL	
<b>Cargo</b>	<b>Nombre</b>
Gerente del proyecto	
Líder Técnico	
Equipo Consultivo	

## **B. Propósito del Proyecto**

Adquirir e Implementar una herramienta ETL de clase mundial, para optimizar y automatizar los procesos de extracción, transformación y carga para mejorar el tiempo de atención de los requerimientos de Inteligencia de negocios.

## **D. Descripción del Proyecto**

Este proyecto consiste en la evaluación y selección de una herramienta ETL, para su posterior implementación y puesta en producción para ser utilizada por el departamento de Extracción, Transformación y Carga en el desarrollo de sus procesos.

## **E. Hitos Principales**

<b>Actividad</b>	<b>Fecha Estimada</b>	<b>Duración estimada</b>
Evaluación del proveedor	Noviembre 2007	3 meses
Selección del proveedor	Diciembre 2008	1 mes
Instalación de la solución	Enero 2008	1 mes
Capacitación	Marzo 2008	2 meses
Pilotos y Pruebas	Mayo 2008	2 meses
Despliegue completo	Julio 2008	2 meses

## **Enunciado del alcance Preliminar:**

### **Objetivos del proyecto**

- Mejorar los tiempos de desarrollo en un 40% respecto al proceso actual
- Minimizar los errores de programación y definiciones en los procesos de carga ETL
- Minimizar el tiempo de soporte y reproceso de información
- Mejorar el entendimiento y documentación de los procesos ETL

### **Límites del proyecto**

- El proyecto no contempla la migración de los procesos de carga existentes.

### **Requisitos y características del producto o servicio**

- Tiempo de desarrollo debe ser menor en un 40% al desarrollo actual.
- Herramienta debe ser visual
- Debe cumplir con los requisitos de funcionalidades, estándares, integración, soporte y seguridad establecidos por la organización
- Tiempo de procesamiento debe ser menor o igual al tiempo de procesamiento actual

### Criterios de aceptación del producto

- Cumplir con los requisitos establecidos por la organización en cuanto a funcionalidad, estándares, integración, soporte, seguridad.

### Requisitos y productos entregables del proyecto

- Servidor de Desarrollo implementado
- Servidor de Producción implementado
- Estándares de desarrollo y mejores prácticas
- Definición de flujos de pruebas y pases a producción
- Documentación funcional y técnica del producto
- Manuales de Capacitación
- Curso básico, avanzado y para administradores
- Resultados de pruebas y pilotos satisfactorios.

### Restricciones del proyecto

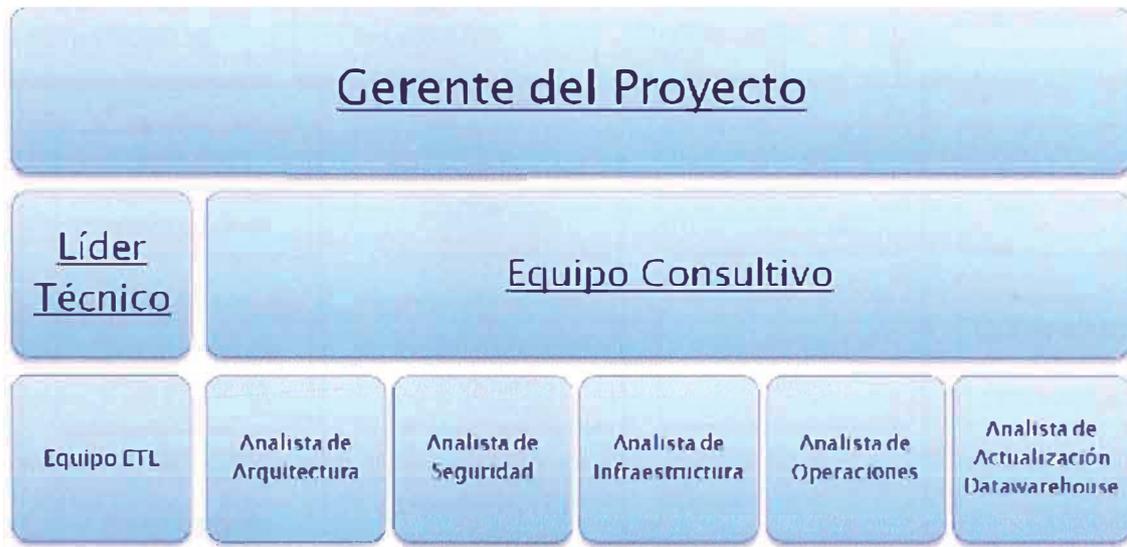
- El proyecto deberá ceñirse a la arquitectura actual de conectividad establecida para el Datawarehouse.
- El tiempo máximo para la implementación es de un (01) año.
- La herramienta ETL deberá estar en producción el semestre 2008-II

### Asunciones del proyecto (supuestos)

- Se contará con el tiempo necesario por parte de las áreas involucradas en la implementación del proyecto.

### Organización inicial del proyecto

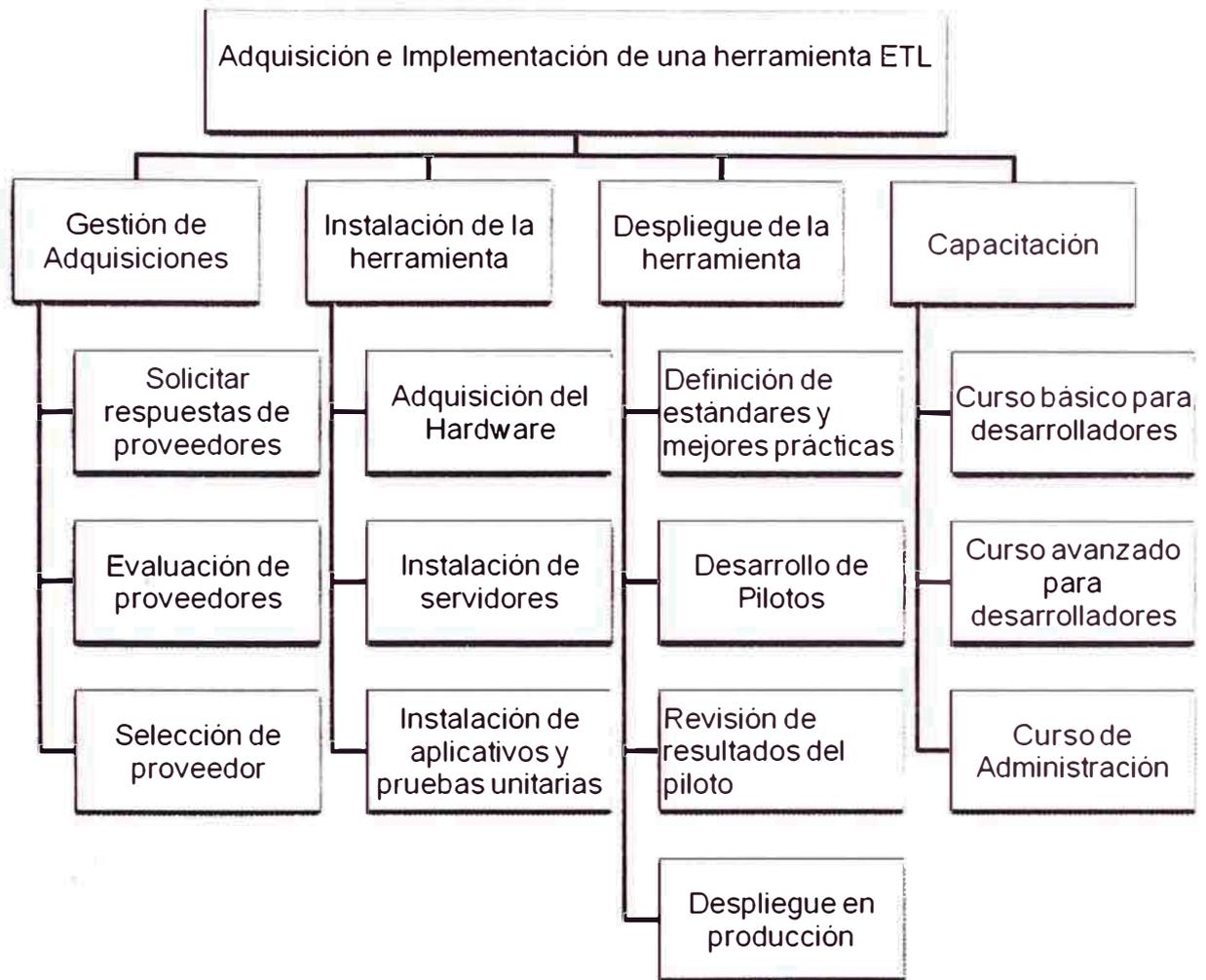
El proyecto presentará la siguiente estructura organizacional, mostrada en la figura 19:



**FIGURA 19:** Organización del proyecto

### EDT inicial

La estructura de trabajo de primer nivel definida par el proyecto, se muestra a continuación en la figura 20:



**FIGURA 20:** EDT inicial

#### IV.1.2 PLANIFICACIÓN

En la fase de planificación se estableció el cronograma inicial del proyecto, este cronograma se dividió en fases, las cuales se muestran en la figura 21:

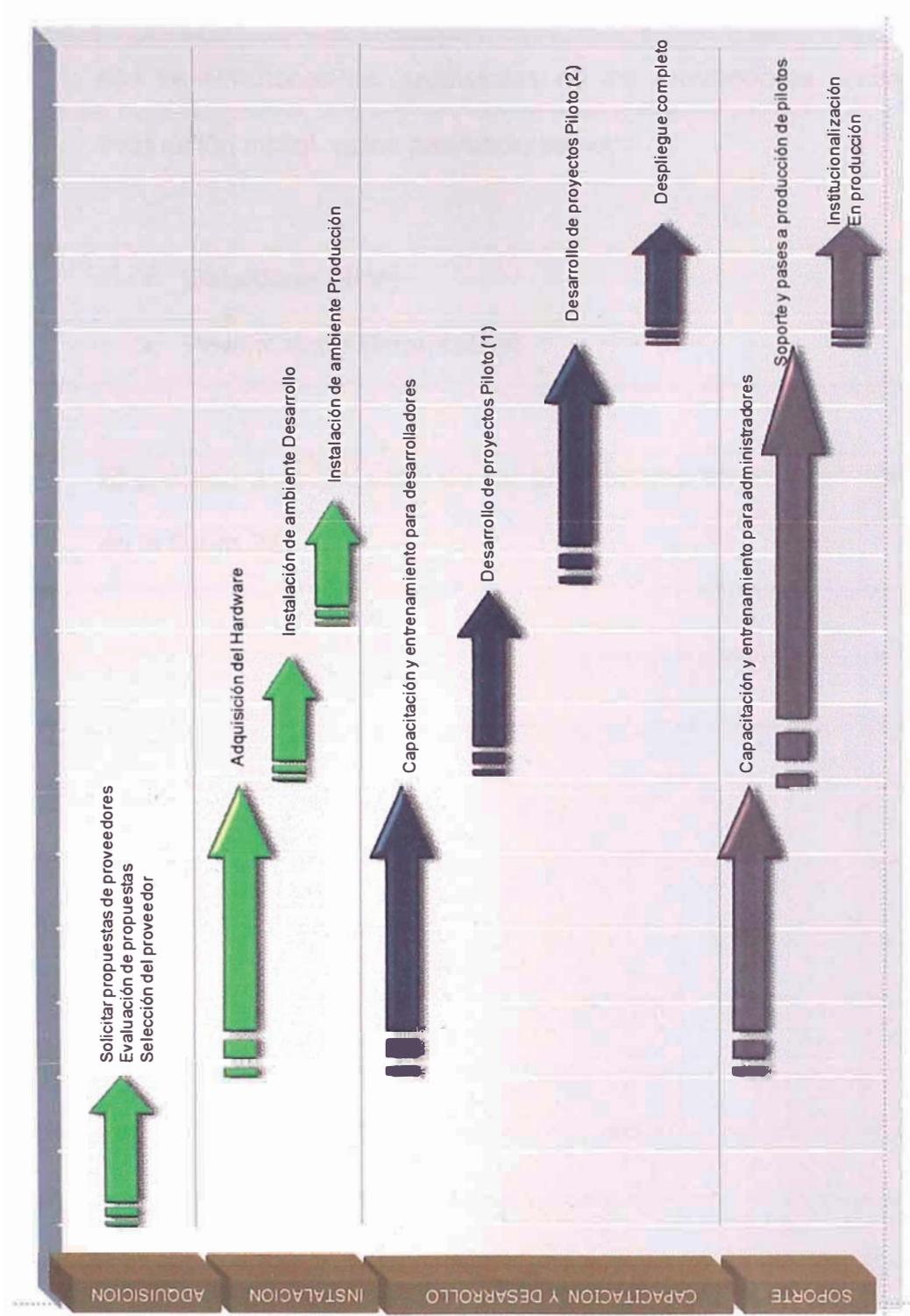
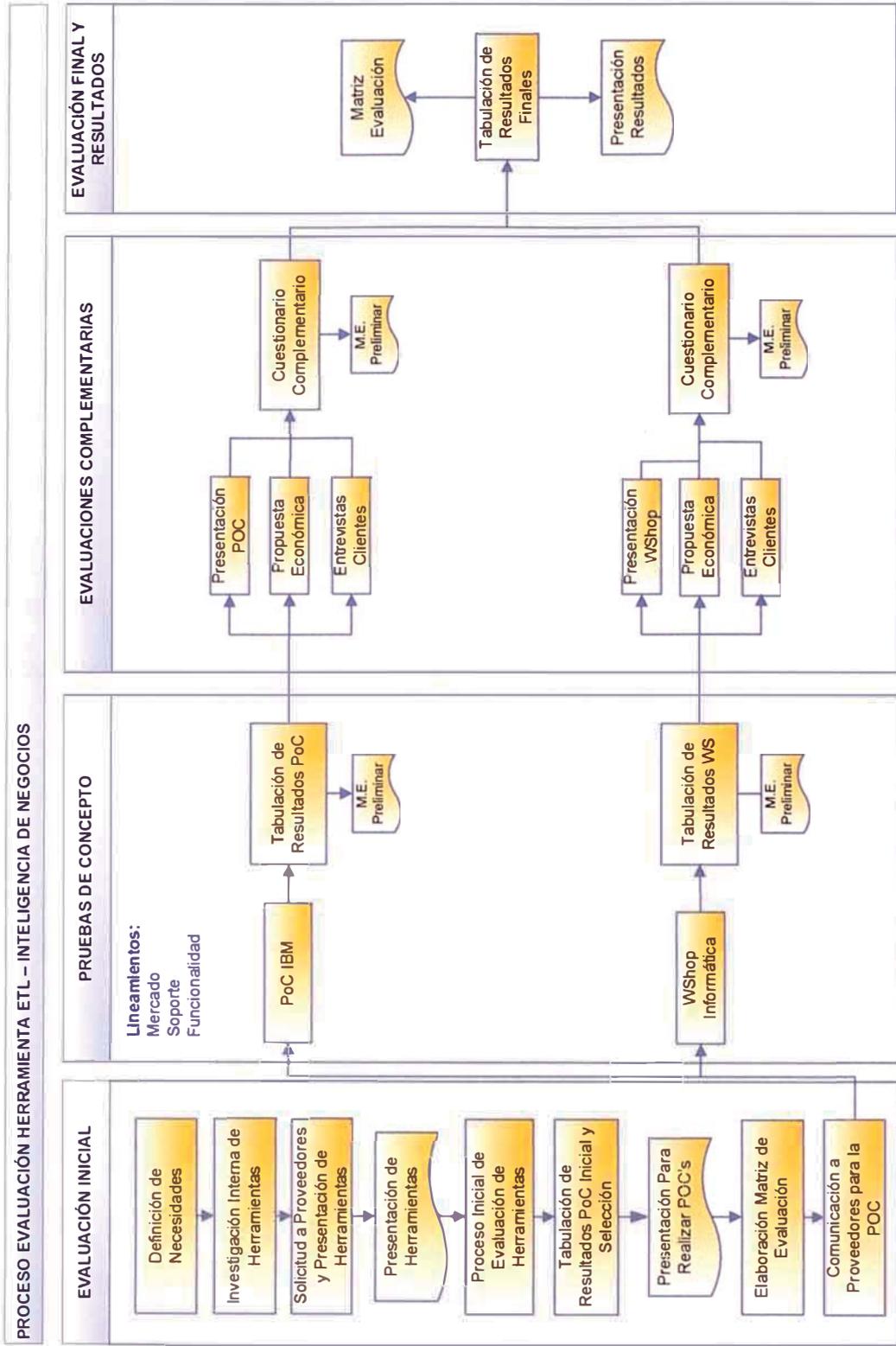


FIGURA 21: Cronograma por fases del proyecto

Durante la etapa de planificación se realizó el plan de adquisiciones, para ello se solicitaron las propuestas de los proveedores considerados en la evaluación inicial, estos proveedores son:

- DataStage (IBM)
- PowerCenter (Informática)

El proceso de evaluación de los proveedores se planificó como se muestra en la figura 22:



**FIGURA 22:** Planificación de la evaluación de la herramienta

La matriz de evaluación se esquematiza en la figura 23:



FIGURA 23: Matriz de evaluación

#### IV.1.4 EJECUCIÓN

Durante la ejecución del proyecto se realizaron las actividades necesarias para cumplir con las fases establecidas en el cronograma. Estas actividades se detallan a continuación:

##### Adquisición

En esta etapa se realizó la evaluación de los proveedores que presentaron sus propuestas, estos fueron:

- DataStage (IBM)
- PowerCenter (Informática)

Se realizó el proceso de evaluación respectivo, según el plan de adquisiciones, utilizando la matriz de evaluación definida, se obtuvo los siguientes resultados para los proveedores que fueron evaluados.

Criterio de Evaluación	Ponderación	Power Center (Informática)	DataStage (IBM)
Funcionalidad	50%	4.65	4.85
Estándares	10%	4.95	5
Integración	10%	5	5
Seguridad	10%	4.85	4.9
Soporte Operativo	20%	4.8	4.8
<b>TOTAL</b>	<b>100%</b>	<b>4.765</b>	<b>4.875</b>

En los anexos se describe a detalle la matriz de evaluación utilizada para la selección.

Dado que los costos de ambas herramientas son similares y según los resultados de la evaluación se tiene que la herramienta a adquirir sería DataStage de IBM

### **Instalación**

Una vez definida la herramienta, se procedió a la adquisición del Hardware necesario para la herramienta, el equipo que se adquirió para la solución fue el siguiente:

#### *Ambiente de Desarrollo*

Sistema Operativo AIX

1 procesador Power 5 de 1.8 GHz

150 Gb de disco para el sistema operativo y aplicaciones

200 Gb de disco para procesamiento de datos

### Ambiente de Producción

Sistema Operativo AIX

2 procesadores Power 5 de 1.8 GHz

150 Gb de disco para el sistema operativo y aplicaciones

500 Gb de disco para procesamiento de datos

Los ambientes que se instalaron a nivel Aplicaciones y el uso que se les dio, se muestran en la figura 24:

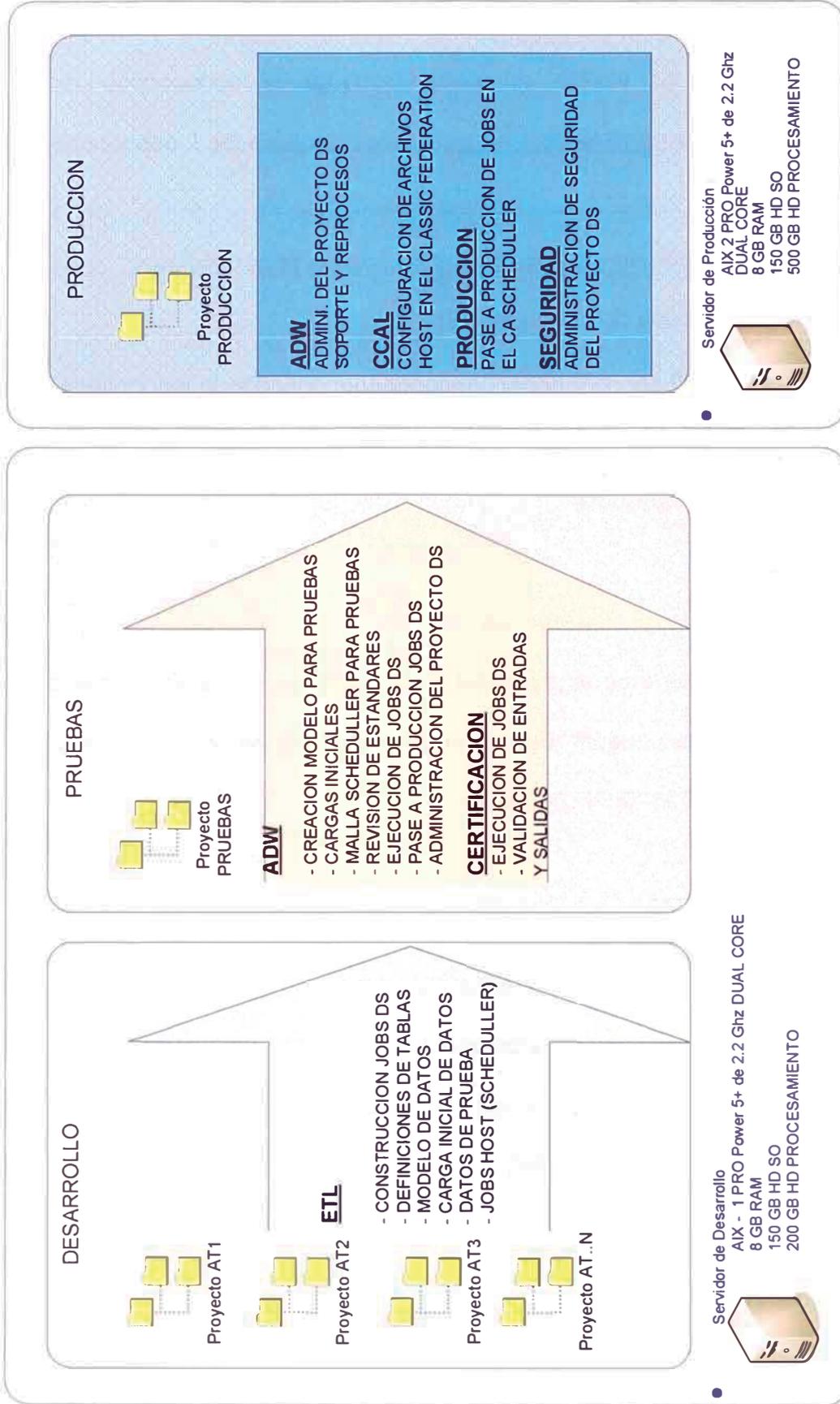


FIGURA 24: Ambientes y funciones de DataStage

## **Capacitación y desarrollo**

Se llevaron a cabo 3 cursos para las personas involucradas en el proceso: Desarrolladores y Administradores.

### Curso básico para desarrolladores

En este curso se introdujo a los desarrolladores a los conceptos de DataStage, los elementos, flujos de datos y metodología de desarrollo.

También se generaron flujos de datos básicos utilizando los elementos más importantes de la herramienta.

### Curso Avanzado para desarrolladores

En este curso se profundizó más en el desarrollo de los procesos de carga, el uso de elementos más sofisticados y procesos más complejos, así como la personalización de la herramienta.

### Curso para Administradores

En este curso se orientó a la administración de la herramienta, el sistema operativo, aspectos de seguridad, seguimiento, soporte a los procesos de carga, configuraciones de la herramienta.

Curso que complemento las habilidades de uso de la herramienta.

En lo que respecta al desarrollo de los procesos de carga y arquitectura del DataWarehouse, se realizó la redefinición de la arquitectura con la

herramienta ETL, en las figuras 25 y 26 se muestra el esquema actual y el nuevo esquema de la arquitectura de DataWarehouse

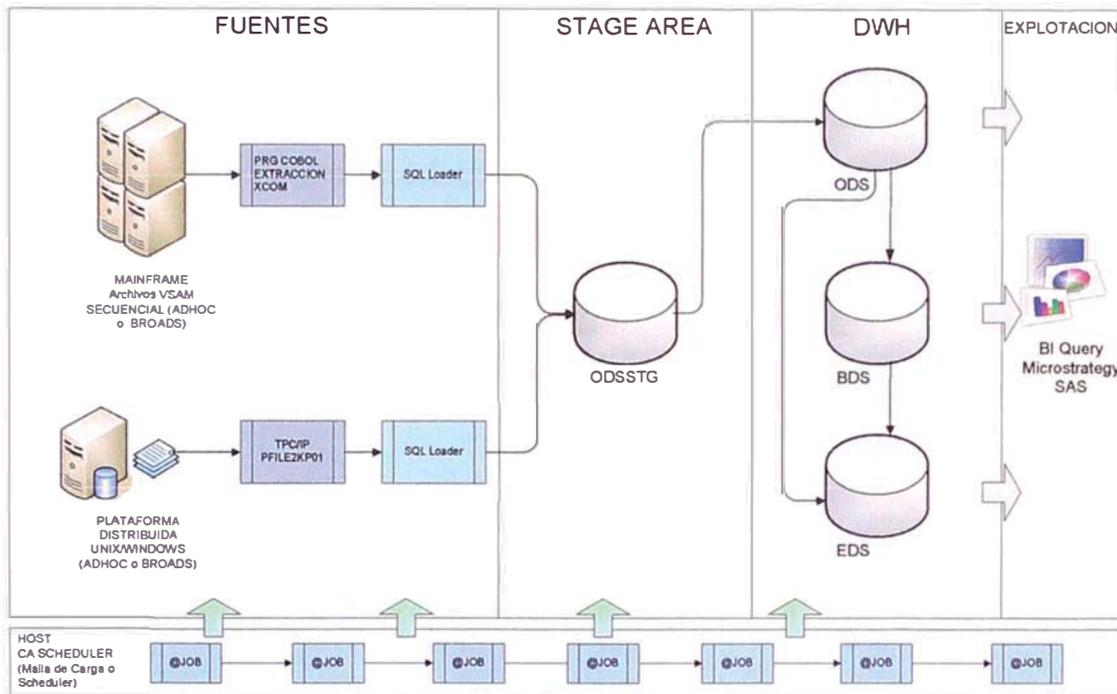


FIGURA 25: Esquema actual de los procesos ETL

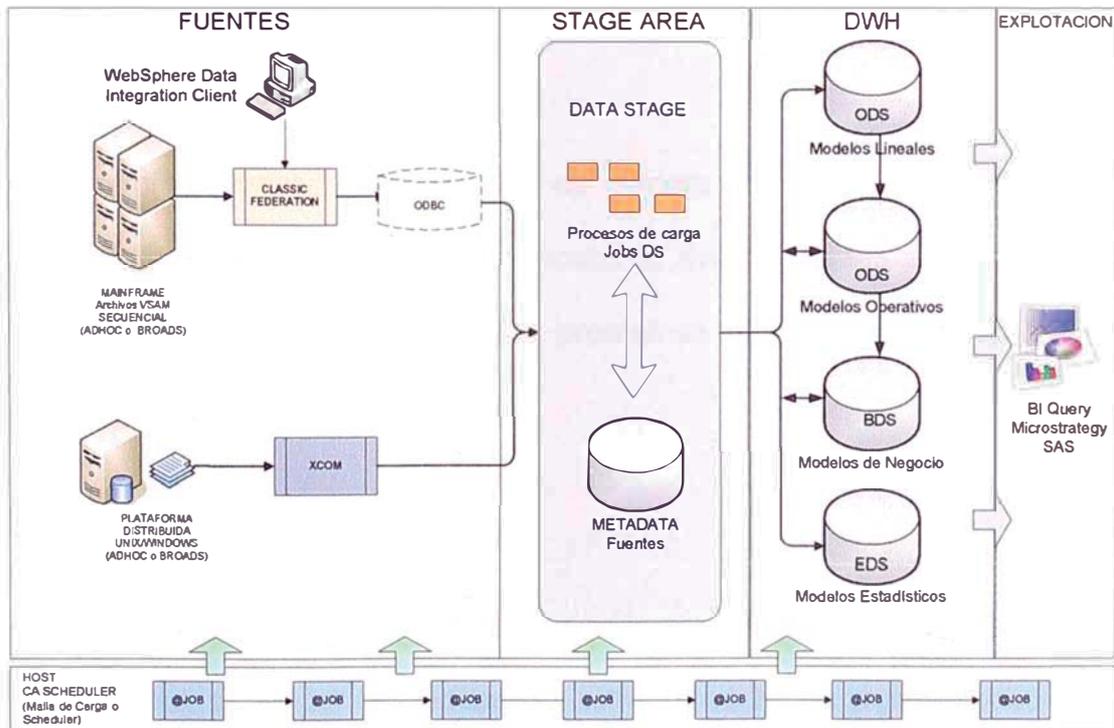
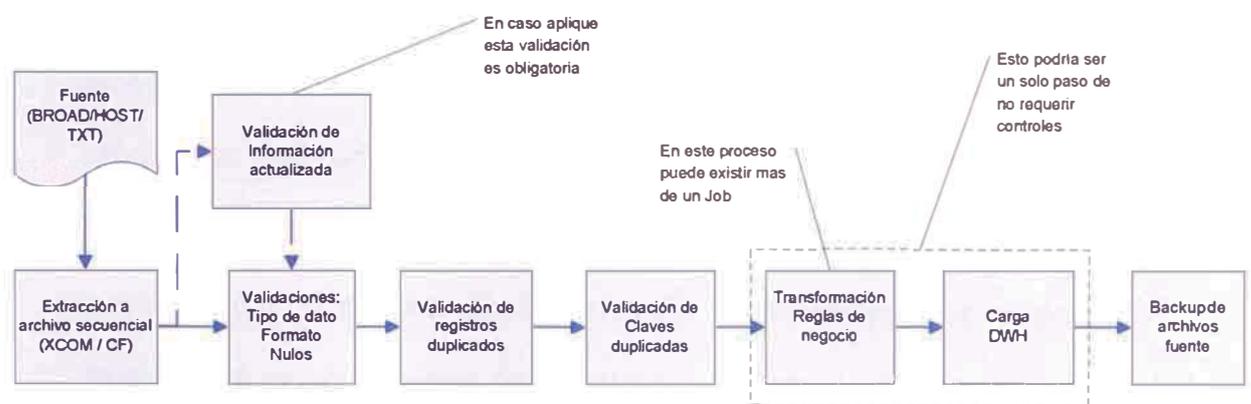


FIGURA 26: Esquema propuesto de los procesos ETL

## Soporte

En el tema de soporte, se realizaron mejoras significativas al flujo de procesos, dado que con la herramienta se incorporaran controles en los procesos de carga, los cuales minimizaran los casos de inconsistencia de información y errores en producción por problemas de datos.

El flujo de los procesos de carga se redefinió de la siguiente manera:



**FIGURA 27:** Esquema propuesto de los procesos de carga

### IV.1.5 CONTROL

El control del proyecto se realizaba de manera quincenal, con una reunión de comité de proyecto, el cual verificaba el avance de las actividades así como la identificación de los riesgos, problemas del proyecto y apoyaban su solución.

### IV.1.6 CIERRE

Una vez finalizadas las actividades y con la conformidad de los entregables del proyecto, se da por finalizadas las actividades del proyecto, procediendo a la liquidación del contrato y al cierre del proyecto.

## IV.2 EVALUACIÓN DE RESULTADOS

Los resultados obtenidos en la implementación de la herramienta fueron los siguientes:

- Se mejoró los tiempos de desarrollo en un 30% respecto al proceso actual, se espera que cuando se estabilice la curva de aprendizaje de la herramienta se supere la meta trazada del 40%
- Gracias a las redefiniciones del flujo de pase a pruebas y producción, se minimizó a un 1% errores de programación respecto a los pases a producción realizados con la herramienta.
- Se minimizó el tiempo de soporte y reproceso, dado que con el nuevo esquema de controles de información establecidos, se pueden detectar las inconsistencias de información existentes en los procesos de carga.
- Al ser una herramienta visual, los procesos de carga ahora son más entendibles y fáciles de documentar.

## CONCLUSIONES Y RECOMENDACIONES

### CONCLUSIONES

- La implementación de la herramienta trajo como resultado la disminución en tiempos de desarrollo en un 30%, se espera que estos tiempos mejoren a medida que la curva de aprendizaje se estabiliza. Este resultado se evidenció en los pilotos realizados.
- Se espera en base a esta mejora, que se puedan atender mayor número de requerimientos y así dar apoyo al cumplimiento de los objetivos de la organización.
- El uso de la herramienta ETL permite tener procesos independientes de la plataforma de la Base de datos, lo cual es favorable para un esquema de migración de plataforma a futuro.
- La implementación de la herramienta ETL trajo como valor agregado la implantación de nuevas metodologías de desarrollo, las cuales favorecen el desarrollo de controles del proceso de carga e integridad de datos, minimizando así los errores en producción y problemas de calidad de información.

## RECOMENDACIONES

- Es necesario iniciar un proyecto de migración de los procesos actuales a la nueva herramienta, para lograr una madurez completa y aprovechar todas las potencialidades de la herramienta en todos los proyectos de Inteligencia de Negocios.
- Se recomienda contar con un servicio de soporte de primer nivel por parte del proveedor IBM durante 1 año posterior al despliegue, mientras la herramienta alcanza un nivel de madurez completa en la organización.
- Se debe tener un proceso de capacitación constante, ya que al no ser una herramienta muy común en nuestro medio, el personal especializado es escaso y los nuevos desarrolladores que ingresen a la organización deberán obtener la habilidad de manejar la herramienta de manera rápida.
- Se debe contar con un repositorio de documentación y mejores prácticas que debe ir enriqueciéndose conforme los proyectos van encontrando nuevas casuísticas y problemáticas.

## GLOSARIO DE TÉRMINOS

### BASE DE DATOS (DATA BASE)

Conjunto de datos no redundantes, almacenados en un soporte informático, organizados de forma independiente de su utilización y accesibles simultáneamente por distintos usuarios y aplicaciones. La diferencia de una BD respecto a otro sistema de almacenamiento de datos es que éstos se almacenan en la BD de forma que cumplen tres requisitos básicos: no redundancia, independencia y concurrencia.

### CLIENTE/SERVIDOR

Arquitectura de sistemas de información en la que los procesos de una aplicación se dividen en componentes que se pueden ejecutar en máquinas diferentes. Modo de funcionamiento de una aplicación en la que se diferencian dos tipos de procesos y su soporte se asigna a plataformas diferentes.

### DICCIONARIO DE DATOS

Descripción lógica de los datos para el usuario. Reúne la información sobre los datos almacenados en la BD (descripciones, significado, estructuras, consideraciones de seguridad, edición y uso de las aplicaciones, etc.).

## **DRILL-DOWN**

Obtención de información más detallada sobre un conjunto de información en el cual se está trabajando. Ejemplo: Si se está mirando el Activo, obtener todas las cuentas del activo.

## **INCONSISTENCIA**

El contenido de una base de datos es inconsistente si dos datos que deberían ser iguales no lo son. Por ejemplo, un empleado aparece en una tabla como activo y en otra como jubilado.

## **INTEGRIDAD**

Condición de seguridad que garantiza que la información es modificada, incluyendo su creación y borrado, sólo por el personal autorizado.

## **REPOSITORIO**

Base de datos central en herramientas de ayuda al desarrollo. El repositorio amplía el concepto de diccionario de datos para incluir toda la información que se va generando a lo largo del ciclo de vida del sistema, como por ejemplo:

componentes de análisis y diseño (diagramas de flujo de datos, diagramas entidad-relación, esquemas de bases de datos, diseños de pantallas, etc.), estructuras de programas, algoritmos, etc. En algunas referencias se le denomina Diccionario de recursos de información.

## **SISTEMA DE GESTION DE BASE DE DATOS**

Software que controla la organización, almacenamiento, recuperación, seguridad e integridad de los datos en una base de datos. Acepta pedidos de datos desde un programa de aplicación y le ordena al sistema operativo transferir los datos apropiados.

Cuando se usa un sistema de gestión de base de datos, SGDB, (en inglés DBMS), los sistemas de información pueden ser cambiados más fácilmente a medida que cambien los requerimientos de la organización. Nuevas categorías de datos pueden agregarse a la base de datos sin dañar el sistema existente.

## **SISTEMA DE INFORMACION (SI)**

Conjunto de elementos físicos, lógicos, de comunicación, datos y personal que, interrelacionados, permiten el almacenamiento, transmisión y proceso de la información.

## **SQL (Structured Query Language)**

Lenguaje de interrogación normalizado para bases de datos relacionales. El SQL es un lenguaje de alto nivel, no procedural, normalizado, que permite la consulta y actualización de los datos de BD relacionales. Se ha convertido en el estándar para acceder a BD relacionales. La primera versión se aprobó como norma ISO en 1987 y la segunda, conocida como SQL2 y vigente actualmente, en 1992.

Actualmente se trabaja en la norma SQL3 que soportará bases de datos orientadas a objeto y bases de datos activas. El SQL facilita un lenguaje de definición de datos y un lenguaje de manipulación de datos. Además, incluye una interfase que permite el acceso y manipulación de la BD a usuarios finales.

## **UNIX**

Sistema operativo multiproceso, multiprograma y multiusuario. Software diseñado por AT&T para ingeniería de telecomunicación. Ha sido el primer sistema operativo concebido con independencia de los fabricantes. Posee una gran facilidad para adaptarse a ordenadores con diferentes arquitecturas, siendo ampliamente autónomo respecto del hardware. Está escrito en lenguaje de alto nivel C.

## BIBLIOGRAFÍA

- Larissa T. Moss, Shaku Atre. Business Intelligence Roadmap. The Complete Project Lifecycle for Decision-Support Applications. Addison – Wesley. 2003
  
- Daniel L. Moody and Mark A. R. Kortink. Artículo: From ER Models to Dimensional Models: Bridging the Gap between OLTP and OLAP Design, Part I. Business Intelligence Journal. Summer 2003.
  
- Daniel L. Moody and Mark A. R. Kortink. Artículo: From ER Models to Dimensional Models: Advanced Design Issues, Part II. Business Intelligence Journal. Fall 2003.

# **ANEXOS**

## ANEXO 1: Matriz de evaluación de proveedores

Para la evaluación de proveedores se toma en cuenta la siguiente matriz:

### Funcionalidades

N°	Rubro	% Pond.	Peso	Nota Máx.	Nota	Nota Pond. Máx.	Nota Pond.
<b>1</b>	<b>Características de transporte, carga y actualización de datos</b>	<b>0.2</b>		<b>75</b>	<b>0</b>	<b>128</b>	<b>0</b>
1.1	Destinos de Carga						
1.1.1	Qué sistemas de destino son soportados		15	5		75	0
1.1.2	Qué utilidades de carga masiva son soportadas		5	5		25	0
1.1.3	Se puede deshabilitar índices y referencias de integridad.		12	5		60	0
1.1.4	Permite obtener fotos de la información después de la carga.		12	5		60	0
1.1.5	Carga particionada de información		4	5		20	0
1.1.6	Generación automática de DDL		6	5		30	0
1.2	Destinos de Tranformaciones						
1.2.1	Qué tipos de transformaciones complejas soporta la herramienta		8	5		40	0
1.2.2	Permite la generación de in-memory surrogate keys		6	5		30	0
1.2.3	Permite agregación y crecimiento incremental de las dimensiones		10	5		50	0
1.2.4	Procesamiento recursivo e iteraciones		8	5		40	0
1.2.5	Permite generación de archivos temporales		6	5		30	0
1.2.6	Permite la fusión o mezcla de registros de múltiples archivos de fuente		15	5		75	0
1.3	Transporte						
1.3.1	Protocolos de red soportados fuera del paquete		9	5		45	0
1.3.2	Puede iniciar programas de transferencia de archivos en servidores remotos		6	5		30	0
1.3.3	Encriptación y compresión de información		6	5		30	0
<b>2</b>	<b>Características de extracción y captura de datos</b>	<b>0.2</b>		<b>65</b>	<b>0</b>	<b>136</b>	<b>0</b>
2.1	Adaptadores de Extracción						
2.1.1	Lista de adaptadores que vienen incluidos en el producto		10	5		50	0
2.1.2	Liste los adaptadores de sistemas fuentes que pueden ser obtenidos por separado		12	5		60	0
2.2	Adaptadores de Desarrollo						
2.2.1	Tipo de SDK que existen para los adaptadores personalizados construidos		9	5		45	0
2.3	Proceso de Extracción						
2.3.1	Proceso de registros en batch		10	5		50	0
2.3.2	Proceso de registros en Near Real Time?		9	5		45	0
2.3.3	Indique los mecanismos para el proceso de registros secuencialmente (ejemplo, cursor de base de datos)		8	5		40	0
2.3.4	Proceso de registros de diferentes fuentes de manera simultánea		12	5		60	0
2.4	Procesamiento Paralelo y Escalabilidad						
2.4.1	Ejecución de múltiples jobs de manera concurrente		15	5		75	0
2.4.2	Ejecución de un jobs en paralelo usando sistema de hebras y pipelining		10	5		50	0
2.4.3	Escalamiento lineal a través de multiples CPU		10	5		50	0

2.4.4	Soporte para balanceo de carga y fallas a través de servidores en cluster.		12	5		60	0
2.4.5	Manejo de archivos en memoria para evitar crear ficheros temporales.		10	5		50	0
2.5	Extracciones Basadas en Reglas						
2.5.1	Extracción selectiva de los datos basada en reglas complejas y filtros. Indique las fuentes.		9	5		45	0
2.5.2	Captura de los cambios que ocurrieron desde la carga o la actualización. Indique las fuentes		9	5		45	0
<b>3</b>	<b>Características de transformación</b>	<b>0.2</b>		<b>60</b>	<b>0</b>	<b>99</b>	<b>0</b>
3.1	Codificación						
3.1.1	Que código es generado internamente por la herramienta		8	5		40	0
3.1.2	Se tiene acceso a modificación del código generado internamente		12	5		60	0
3.1.3	Lenguajes scripts soportados para la codificación interna		3	5		15	0
3.2	Características de Tecnología Orientada a Objetos						
3.2.1	Son los elementos de codificación bloques procedurales de código o con orientación a objetos		9	5		45	0
3.2.2	Permite representación de objetos como íconos		9	5		45	0
3.2.3	Permite la creación de objetos personalizados		9	5		45	0
3.3	Librería de Objetos						
3.3.1	Número de objetos en su librería		9	5		45	0
3.3.2	Almacenamiento de objetos personalizados en la librería		9	5		45	0
3.4	Transformación de Objetos						
3.4.1	Objetos adicionales de transformación que provea la herramienta		6	5		30	0
3.5	Registros Rechazados						
3.5.1	Lógica incluida para manejar y controlar registros rechazados		10	5		50	0
3.6	Objetos Externos						
3.6.1	Creación de salidas a objetos externos		9	5		45	0
3.6.2	Indique el lenguaje de programación para la escritura de objetos externos.		6	5		30	0
3.6.3	Documentación de objetos externo a través de la herramienta		6	5		30	0
3.6.4	Indique cuáles son los objetos externos (de terceros) que soporta la herramienta		9	5		45	0
3.7	Control de Versiones						
3.7.1	Existe un repositorio de desarrollo		12	5		60	0
3.7.2	Funciones de Check In - Check Out		12	5		60	0
3.7.3	Rollback a versiones previas		12	5		60	0
3.7.4	Control de accesos granularizados		12	5		60	0
3.7.5	Rastros de modificaciones y actividades realizadas		10	5		50	0
3.7.6	Integración con software de control de versiones de terceros. Mencione cuáles		8	5		40	0
3.8	Depuración						
3.8.1	Debugger visual		15	5		75	0
3.8.2	Checkpoints		15	5		75	0
3.8.3	Verificación de variables en cada checkpoint		15	5		75	0
3.8.4	Capacidad de ejecución de pruebas de datos		15	5		75	0
<b>4</b>	<b>Características del entorno de desarrollo</b>	<b>0.15</b>		<b>125</b>	<b>0</b>	<b>123.825</b>	<b>0</b>
4.1	Entorno de Desarrollo						
4.1.1	Cuenta con interface integrada entre los módulos de desarrollo		15	5		75	0
4.1.2	Entorno gráfico de desarrollo		12	5		60	0
4.1.3	Permite el desarrollo en equipo de flujos comunes		15	5		75	0

4.1.4	Interface de línea de comando		9	5		45	0
4.2	Interface Gráfica						
4.2.1	Cuenta con funciones que pueden ser definidas visualmente		15	5		75	0
4.2.2	Indique las funciones que deben ser codificadas		12	5		60	0
4.2.3	La GUI puede representar un flujo complejo como un icono simple (contenedor)		15	5		75	0
4.3	Asistentes y Wizards						
4.3.1	Asistentes. Indique cuales.		6	5		30	0
4.3.2	Mapeos predefinidos para tareas y funciones comunes		6	5		30	0
4.4	Reusabilidad						
4.4.1	Permite copiado y pegado de objetos a través de diferentes sesiones entre más de un flujo de trabajo		15	5		75	0
4.4.2	Se puede anexar sesiones en otras sesiones		12	5		60	0
4.4.3	Sesiones de contexto independiente (configuración automática de trabajo en otros flujos con diferentes orígenes?)		12	5		60	0
4.4.4	Permite actualizar objetos o sesiones copiados una vez reconfigurado la base u origen		12	5		60	0
<b>5</b>	<b>Administración de Metadatos</b>	<b>0.1</b>		<b>60</b>	<b>0</b>	<b>45.5</b>	<b>0</b>
5.1	Documentación Autogenerada						
5.1.1	Captura de reglas automático por la herramienta		12	5		60	0
5.1.2	Datos operacionales que la herramienta automáticamente captura		12	5		60	0
5.1.3	Metadatos de negocios que la herramienta automáticamente captura		8	5		40	0
5.2	Repositorio de Metadatos						
5.2.1	Motor de almacenamiento (ejemplo: RDBMS o propietario)		9	5		45	0
5.2.2	Repositorio de APIs		9	5		45	0
5.2.3	Repositorio de clientes soportado		9	5		45	0
5.2.4	Estándares soportados		6	5		30	0
5.3	Extensibilidad						
5.3.1	Indique si existe un SDK para extender el modelo de datos del repositorio. Cuales.		0	5			0
5.4	Administración y Gestión Distribuida						
5.4.1	Manejo y sincronización de metadatos en múltiples herramientas ETL sin codificación personalizada		9	5		45	0
5.5	Ingeniería Reversa						
5.5.1	Indique las fuentes de las cuales la herramienta puede hacer ingeniería inversa.		3	5		15	0
5.6	Intercambio de Metadatos						
5.6.1	Importación y exportación de metadatos con herramientas de terceros. Cuales		8	5		40	0
5.6.2	Sincronización manual de metadatos con herramientas de terceros. Cuales		6	5		30	0
5.6.3	Sincronización automática de metadatos con herramientas de terceros sin codificación. Cuales.		6	5		30	0
5.7	Reportes de Análisis de Impacto						
5.7.1	Representación visual de dependencias entre los componentes en múltiples flujos de trabajo ETL		12	5		60	0
5.7.2	Representación visual de dependencias a través de herramientas de terceros. Cuales		0	5			0
5.8	Reportes de Data Lineage						
5.8.1	Generación de reportes visuales que permita a los usuarios de negocio visualizar el origen de los componentes		15	5		75	0
5.9	Otros Reportes						

5.9.1	Que otros reportes puede generar el repositorio adicionales a los enlatados.		8	5		40	0
<b>6</b>	<b>Características de calidad de datos</b>	<b>0.05</b>		<b>120</b>	<b>0</b>	<b>27.375</b>	<b>0</b>
6.1	Validación						
6.1.1	Verificación y validación de rangos de dominio		15	5		75	0
6.2	Integración						
6.2.1	Integración otras herramientas limpieza y calidad de datos. Detalle		6	5		30	0
6.2.2	Integración visual al proceso de validación y/o limpieza de datos en el flujo de procesos ETL		6	5		30	0
6.3	Reglas						
6.3.1	Reglas de generación automática para detectar y corregir defectos en los datos		9	5		45	0
6.4	Seguimiento						
6.4.1	Capacidad de verificar que el contenido del registro concuerde con el tipo de dato a insertar		12	5		60	0
6.5	Reportes						
6.5.1	Generación de reportes de análisis de calidad de datos		8	5		40	0
<b>7</b>	<b>Administración de la herramienta</b>	<b>0.1</b>		<b>70</b>	<b>0</b>	<b>72.5</b>	<b>0</b>
7.1	Consola						
7.1.2	Consola visual para ejecución y administración de jobs		12	5		60	0
7.1.3	Proporciona control a través de línea de comando		9	5		45	0
7.1.4	Permite gestionar múltiples sistemas ETL		9	5		45	0
7.1.5	Se integra con sistemas de gestión de terceros		9	5		45	0
7.2	Monitoreo de Jobs						
7.2.1	Monitoreo en tiempo real		15	5		75	0
7.2.2	Soporta alertas basadas en SMTP		9	5		45	0
7.2.3	Que eventos son monitoreados (Ejem estado de jobs, registros procesados, tiempo transcurrido por tarea, tiempo total transcurrido, consumo de CPU y memoria.)		15	5		75	0
7.3	Scheduler						
7.3.1	Scheduler gráfico para definir dependencia de jobs		12	5		60	0
7.3.2	Scheduler por tiempo, eventos, intervalos o condiciones		8	5		40	0
7.3.3	Integración con scheduler de terceros. Cuales		9	5		45	0
7.3.4	Integración con CA Scheduler HOST		12	5		60	0
7.4	Validación de Jobs						
7.4.1	Puede la herramienta validar los jobs después de su ejecución		9	5		45	0
7.4.2	Puede la herramienta reinicializar el sistema desde el punto de caída		9	5		45	0
7.4.3	Puede la herramienta reiniciar una sesión entera		8	5		40	0
7.5	Reportes						
7.5.1	La herramienta permite la generación de reportes de registro de jobs, estadísticas		12	5		60	0
7.5.2	La herramienta permite la generación de reportes de diagnóstico, detalle de lo sucedido, por qué y qué hacer.		4	5		20	0
7.5.3	Otros reportes		9	5		45	0
7.6	Control de Cambios						
7.6.1	Administración y gestión de QA en los pre pases a producción		12	5		60	0
7.6.2	Permite la gestión de ambientes de desarrollo, QA y producción. Por favor detalle		12	5		60	0
7.7	Seguridad						
7.7.1	Control de accesos y autenticación		8	5		40	0
7.7.2	Integración con LDAP		10	5		50	0

## Estándares

Nº	Rubro	% Pond.	Peso	Nota Máx.	Nota	Nota Pond. Máx.	Nota Pond.
<b>A</b>	<b>Herramientas de Soporte</b>	<b>0.25</b>		<b>20</b>	<b>0</b>	<b>60</b>	<b>0</b>
1	Sistema operativo		15	5		75	0
2	Herramientas de oficina		9	5		45	0
3	Navegador de internet		9	5		45	0
4	Otro software a instalar en estaciones cliente (especificar)		15	5		75	0
<b>B</b>	<b>Servidores Plataforma Unix:</b>	<b>0.25</b>		<b>45</b>	<b>0</b>	<b>150</b>	<b>0</b>
1	Sistema Operativo		15	5		75	0
2	Bases de datos		15	5		75	0
3	Herramientas de Reportes y búsqueda		15	5		75	0
4	Ambiente de desarrollo (entorno, lenguaje de programación, etc.)		10	5		50	0
5	Application Server		15	5		75	0
6	¿Es J2EE compliant?		15	5		75	0
7	Herramientas de explotación de información		15	5		75	0
8	Herramientas para mostrar gráficos e imágenes		10	5		50	0
9	Web Server (IBM HTTP Server, otro)		10	5		50	0
<b>C</b>	<b>Plataforma ZoS390:</b>	<b>0.25</b>		<b>10</b>	<b>0</b>	<b>25</b>	<b>0</b>
1	Cobol para OS390 CICS TS 1.3 (Si / No)		10	5		50	0
2	Estructura de datos VSAM, Datacom (Si / No)		10	5		50	0
<b>D</b>	<b>Sistemas de Explotación de Información</b>	<b>0.25</b>		<b>25</b>	<b>0</b>	<b>93.75</b>	<b>0</b>
1	Sistema Operativo AIX 5.1 (Si / No)		15	5		75	0
2	Base de datos Oracle 9.2 ó 10G (Si / No)		15	5		75	0
3	Herramientas de Reports y Búsqueda / Heramientas olap: Microstrategy, BIQuery, Otros		15	5		75	0
4	Herramientas de distribución de información: Narrow Cast, Microstrategy Office, Otros		15	5		75	0
5	Herramientas de análisis		15	5		75	0

## Integración

Nº	Rubro	% Pond.	Peso	Nota Máx.	Nota	Nota Pond. Máx.	Nota Pond.
<b>A</b>	<b>Aplicación</b>	<b>0.85</b>		<b>65</b>	<b>0</b>	<b>807.5</b>	<b>0</b>
A.1	Arquitectura interna de la aplicación						
1	Capa de presentación (interfaz de usuario). Tipo de interfaz de usuario (Páginas Web, VB, 3270, Java Client, otros)		15	5		75	0
2	Capa de Lógica de Negocios : La lógica de negocios, ¿se encuentra en elementos o programas claramente identificados que no pertenecen a pantallas o páginas?		15	5		75	0
3	Diseño de la aplicación modular, desacoplado, permite cambios focalizados, pases a producción incrementales, etc.		10	5		50	0
4	Reportes y muestra de información estándar, consistente		15	5		75	0
A.2	Integración						
1	Utiliza mecanismos estándares de comunicación: * Transporte: MQ Series, Web Services, otros * Mensaje: SOAP, Trama plana o posicional, IFX, otros		15	5		75	0
A.3	Interfases						
1	¿La solución entrega información en línea que puede ser usada por otras aplicaciones?		15	5		75	0
2	Proceso simple para reutilizar interfaces de aplicaciones existentes		15	5		75	0
3	Formato de entrega de información (string, xml, api, etc)		15	5		75	0
A.4	Información						
1	Mecanismos Sincronización-Replicación (ETL, colas, especificar)		15	5		75	0
2	Tipo de información de gestión que deja (Batch, en Línea, Batch-línea)		15	5		75	0
3	Cuenta con Diccionario de Datos		15	5		75	0
4	Se contemplan validaciones en la captura de datos y procesos internos		15	5		75	0
5	Los datos están comprendidos entre valores pre-determinados (Valores correctos Ej.: dominios)		15	5		75	0
<b>B</b>	<b>Integración</b>	<b>0.15</b>		<b>5</b>		<b>7.5</b>	<b>0</b>
1	Los cambios en los servicios son centralizados, simples y no requieren programación		10	5		50	0

## Seguridad

N°	Rubro	% Pond.	Peso	Nota Máx.	Nota	Nota Pond. Máx.	Nota Pond.
<b>A</b>	<b>Aplicativos en Host</b>	<b>0.35</b>		<b>10</b>	<b>0</b>	<b>35</b>	<b>0</b>
1	¿El manejo de autorización y autenticación se integra con RACF ? De contar con un módulo distinto a RACF describir el módulo de seguridad en la parte Administración de usuarios Generales		10	5		50	0
2	De no ser posible la integración con RACF, es factible manejar la Autenticación y autorización con módulos distintos?		10	5		50	0
<b>B</b>	<b>Aplicativos en ambientes distribuidos</b>	<b>0.3</b>		<b>10</b>	<b>0</b>	<b>45</b>	<b>0</b>
1	¿Se puede integrar con algún directorio LDAP (MS Active directory)?		15	5		75	0
2	De no ser posible la integración con Securinet o Active Directory, ¿es factible manejar la Autenticación y Autorización con módulos distintos? Ejm : Autenticación con Securinet y Autorización con el módulo propio		15	5		75	0
<b>C</b>	<b>Configuraciones Especiales</b>	<b>0.2</b>		<b>50</b>	<b>0</b>	<b>98</b>	<b>0</b>
1	Uso de métodos de encriptación para uso de credenciales de usuarios de conexión.		8	5		40	0
2	No utiliza usuarios con privilegios especiales. Ejm: Administrador del dominio, root, special, sa, dbo, etc		9	5		45	0
3	Protección anti virus		9	5		45	0
4	Mecanismos de encriptación de información sensible (metodos, algoritmos, semillas, llaves, etc.)		12	5		60	0
5	Uso de certificados digitales (Detallar)		6	5		30	0
6	Se permite renombrar y cambiar de contraseña a los usuarios genéricos de procesos y base de datos.		12	5		60	0
7	No utiliza usuarios logueados en los servidores para el funcionamiento de la aplicación.		12	5		60	0
8	No crear logines de usuarios finales directamente a la base de datos.		12	5		60	0
9	No incluir credenciales de usuarios en plano en archivos de parámetros (.ini, txt, config, etc)		9	5		45	0
10	Permitir la instalación de herramientas administrativas en pcs de usuarios.		9	5		45	0
<b>D</b>	<b>Alertas y Auditoría</b>	<b>0.15</b>		<b>45</b>	<b>0</b>	<b>63</b>	<b>0</b>
1	Existe esquema de alertas para eventos de seguridad		9	5		45	0
2	Permite configurar frecuencia de almacenamiento y tamaño de registros de auditoría		9	5		45	0
3	Interacción con eventos del Sistema Operativo		9	5		45	0
4	Registro de log on		6	5		30	0
5	Registro de intentos fallidos		6	5		30	0
6	Registro de bloqueo de cuentas de usuarios		9	5		45	0
7	Registro de modificaciones		12	5		60	0
8	Registro de acciones de administradores		12	5		60	0
9	Cuenta con campos de usuario, fecha y hora, estación, acción, como mínimo.		12	5		60	0

## Soporte

N°	Rubro	% Pond.	Peso	Nota Máx.	Nota	Nota Pond. Máx.	Nota Pond.
<b>A</b>	<b>Infraestructura</b>	<b>0.5</b>		<b>35</b>	<b>0</b>	<b>200</b>	<b>0</b>
1	¿Funciona con las versiones de sw base "estándar" definidas en el BCP para la plataforma que la soporta?		15	5		75	0
2	Se instala en servidores consolidados		15	5		75	0
3	Las distintas capas de la aplicación pueden residir en servidores físicos distintos ? - consolidación		15	5		75	0
4	En caso de ejecutarse sobre .NET debe de ser la version 1.1 o superior		0	5		0	0
5	¿Cuenta con mecanismos de alta disponibilidad ó soporta mecanismos como Sysplex ó Clustering?		5	5		25	0
7	¿Tiene interfaces con MQ?		15	5		75	0
8	El servicio y/o servicios que se ejecuten en el servidor tienen un puerto (puertos) TCP/IP De conectividad definidos		15	5		75	0
<b>B</b>	<b>Monitoreo</b>	<b>0.2</b>		<b>15</b>	<b>0</b>	<b>32</b>	<b>0</b>
1	¿Cuenta con un registro de Errores de la aplicación?		15	5		75	0
2	¿La aplicación genera alertas para su monitoreo?		15	5		75	0
3	¿Soporta ser monitoreada por herramientas estándar del BCP: PEM, Mom, HPOpenview, Quest?		2	5		10	0
<b>C</b>	<b>Herramientas de Operaciones</b>	<b>0.15</b>		<b>20</b>	<b>0</b>	<b>36.75</b>	<b>0</b>
1	¿Cuenta con tareas operativas automatizadas? (Levantar servicios, scripts de depuración de logs, backups, etc.)		15	5		75	0
2	Tiene políticas establecidas de backup-restore que se integren con TSM o BackTrack		4	5		20	0
3	Cuenta con procedimientos de : rollback, recuperación, etc.		15	5		75	0
4	¿Cuenta con procedimientos documentados y manuales para tareas reactivas (restablecimiento automático del Servicio, etc)?		15	5		75	0
<b>D</b>	<b>Contingencia y continuidad operativa</b>	<b>0.15</b>		<b>5</b>		<b>11.25</b>	<b>0</b>
1	¿Cuenta con procesos de contingencia y continuidad operativa?		15	5		75	0