

Implantación de Datawarehouse Open Free



19 de Agosto de 2011

Nicolás Gerolami - Esteban Revello - Germain Venzal

Tutores: Nelson Calero - Lorena Etcheverry - Ariel Sabiguero Yawelak

Agenda

- Introducción
- Objetivos del proyecto
- Evaluación y Comparación de Plataformas
- Calidad de datos
- Proceso de Diseño Multidimensional
- Arquitectura de la Solución Desarrollada
- Implementación
- Resultados obtenidos
- Trabajo a futuro y Conclusiones

Introducción – Conceptos básicos

- Un Data Warehouse (DW) es un conjunto de herramientas de apoyo a la toma de decisiones
 - Componentes básicos:
 - Sistemas de Datos Fuentes
 - Almacenamiento temporal y ETL
 - Herramientas de acceso a datos
 - Presentación de datos
- Business Intelligence usualmente incluye:
 - Integración de datos, calidad de datos
 - Data warehousing
 - Textos y análisis de contenido

Introducción al Proyecto

- Propuesta del área informática de ASSE
- Actualmente la institución dispone de:
 - Reportes gerenciales con información de sistemas heterogéneos
 - Reportes manuales
- Hay necesidades insatisfechas de información
- Actualmente no se dispone de un sistema de información gerencial del tipo Business Intelligence (BI) a nivel organizacional

Objetivos del Proyecto

- Evaluación de herramientas de BI Free/Open Source
- Analisis de la calidad de datos
- Implementación utilizando la plataforma seleccionada
 - Instalación de las herramientas recomendadas
 - Modelado multidimensional
 - Creación de cubos, ETL, reportes
- Traspaso de conocimiento a funcionarios de ASSE
- Interactuar con sistemas de ASSE
 - Sistema de Gestión de Usuarios
 - Sistema AP.SGA (Sistema de prestaciones)

Evaluación y Comparación de Plataformas



Pentaho BI Suite Community



JasperSoft BI Suite Community



SpagoBI Studio



OpenI Suite



Palo Suite

Evaluación y Comparación de Plataformas

- Comparación de funcionalidades:
 - ETL, análisis OLAP, reporting, tableros de mando
 - Seguridad, documentación, usabilidad, otros aspectos
- Prototipado sobre dos plataformas seleccionadas

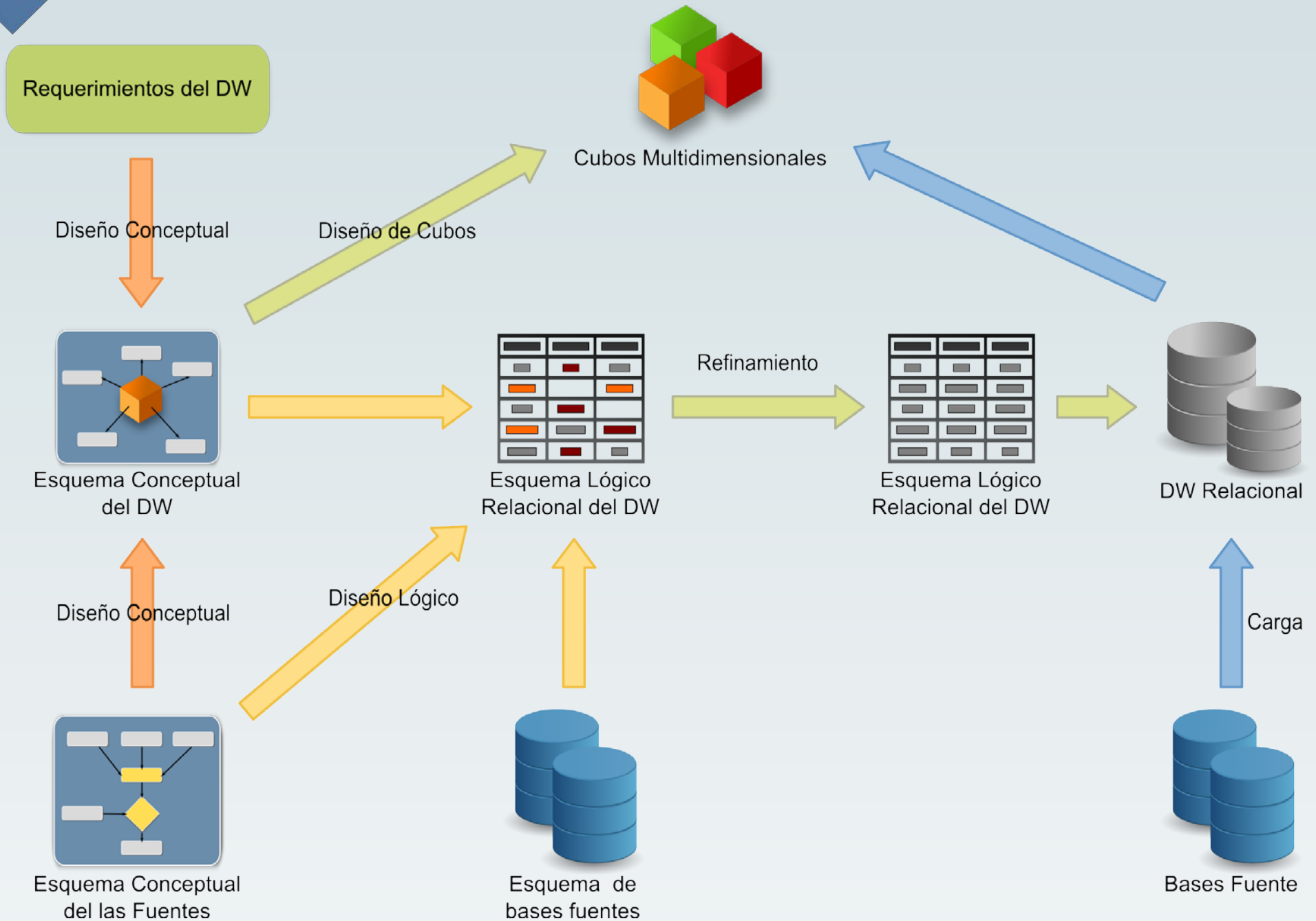
Características	Pentaho	SpagoBI
Instalación	4	5
Configuración	4	2
Documentación brindada por la distribución	4	3
Documentación brindada por fuentes externas	5	2
Accesibilidad y facilidad de uso	5	3
Promedio	4,4	3

- Evaluación y selección final: Pentaho Community

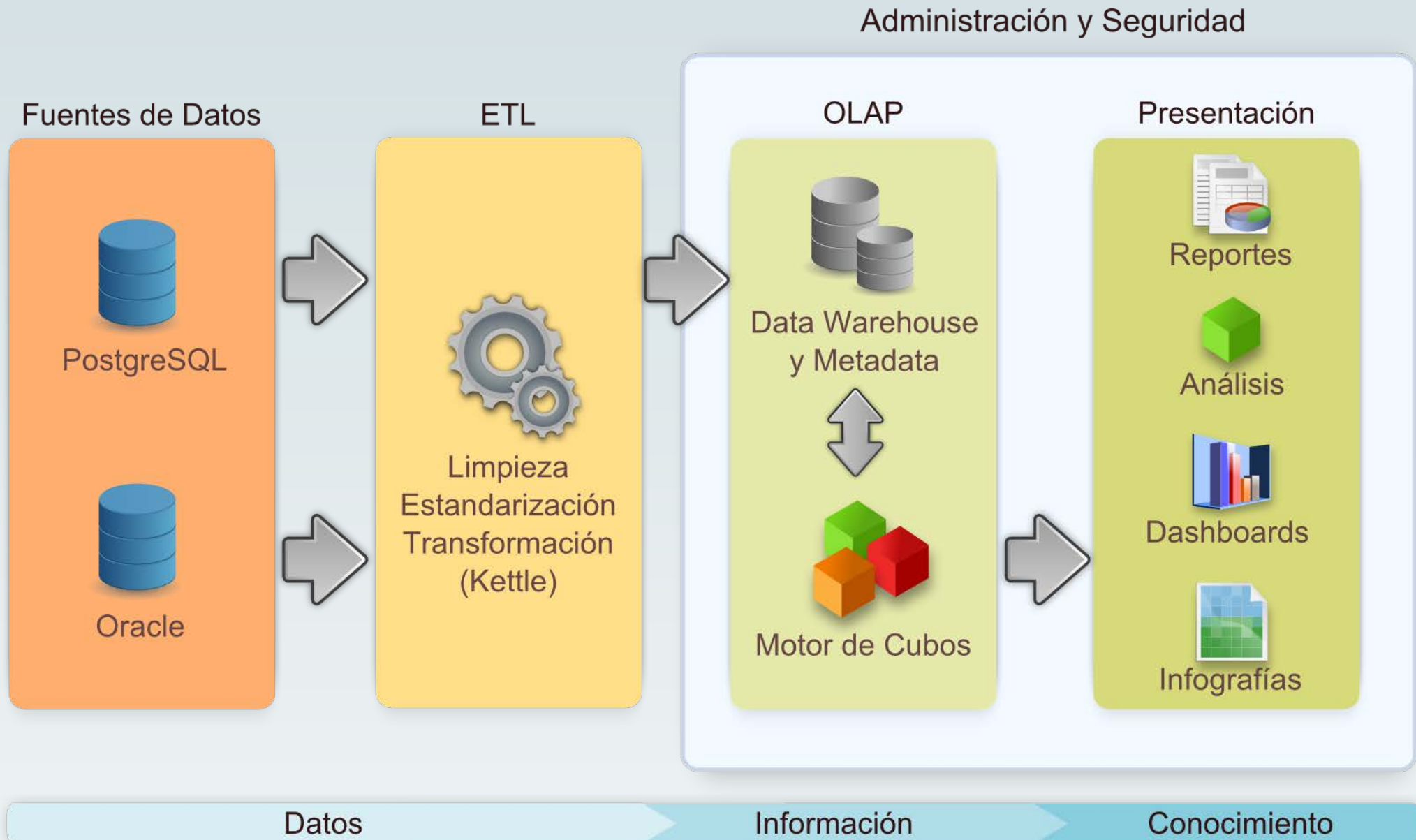
Calidad de datos

- Inexistencia de claves foráneas donde debían existir
- Se encontraron ciertos datos inconsistentes en las bases de datos fuentes
 - Ejemplos:
 - Afiliados con más de 125 años
 - Usuarios asociados a localidades inexistentes
 - Más de 20 departamentos
- Se creó una base de datos intermedia para mejorar la calidad de los datos
 - Depuración de valores no útiles e inconsistentes
 - Restauración de valores (por ej. Edades x f. Nac.)

Proceso de Diseño Multidimensional



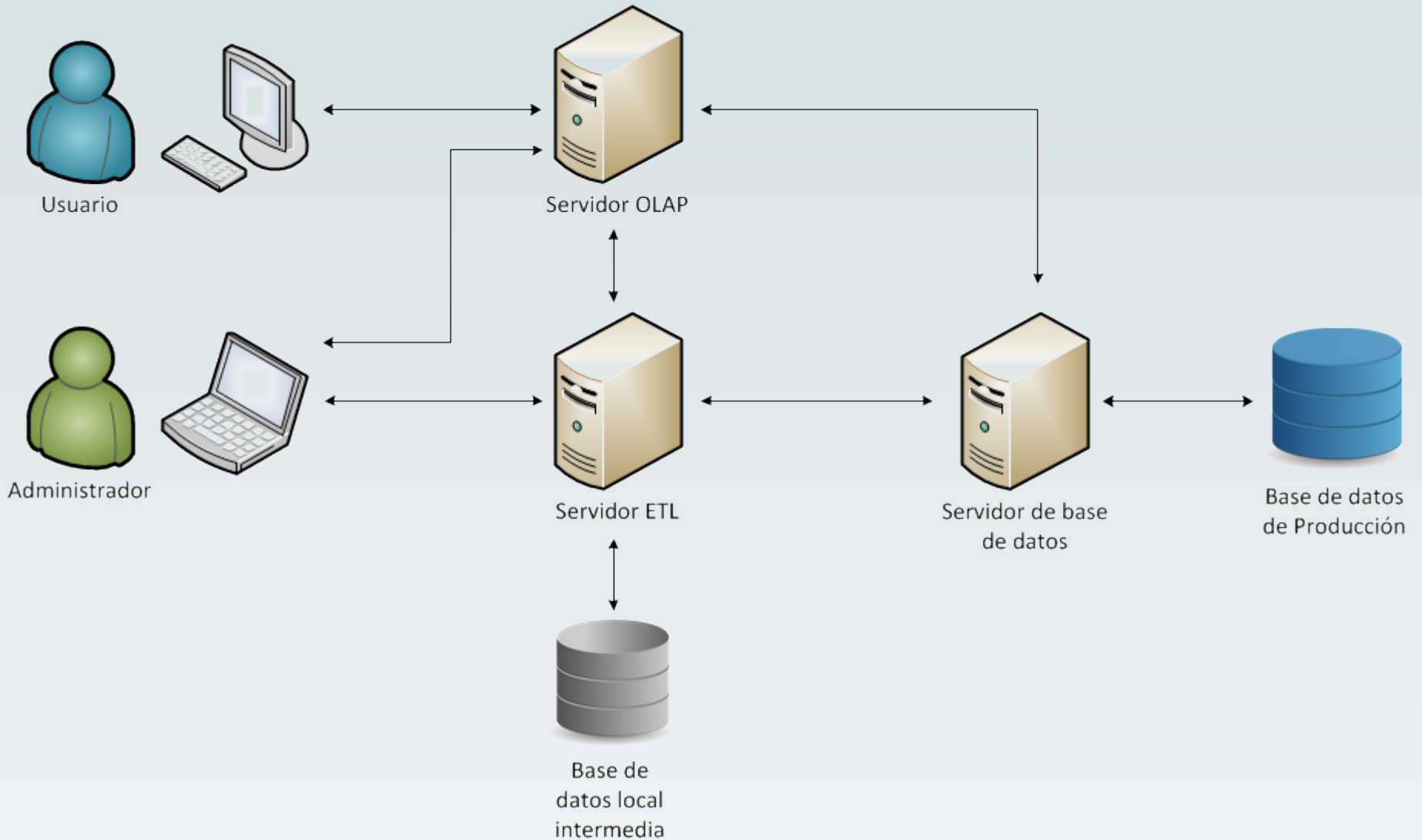
Arquitectura de la Solución Desarrollada



Arquitectura

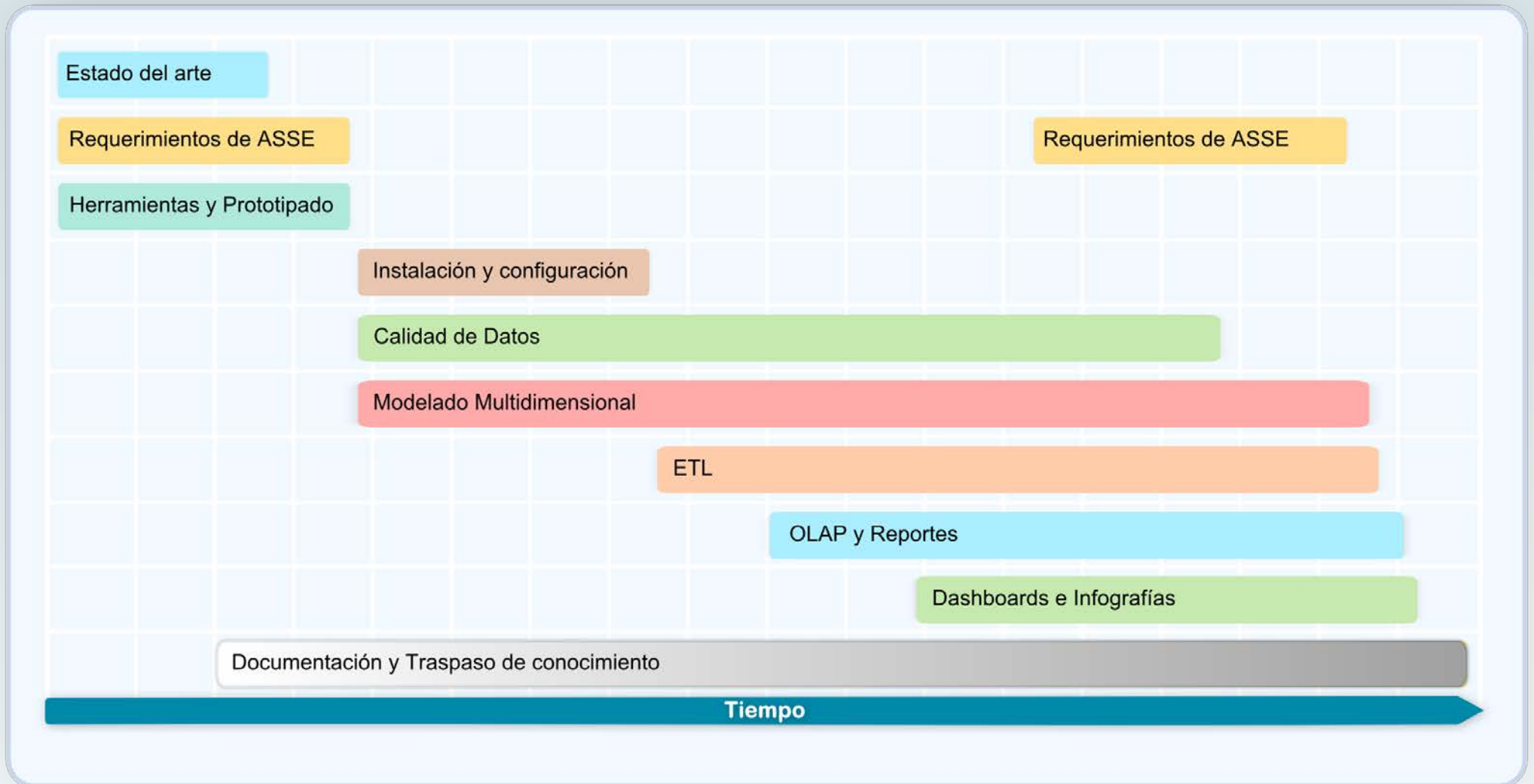
- Software utilizado en la Implementación:
 - SO: Open Suse 11.3 64 bit
 - BD: MySQL 5.5 y PostgreSQL 9.0.3
 - Pentaho BiServer ce 3.7.0 stable
 - Pentaho Data Integration 4.1.0 stable
 - Pentaho Report Designer 3.7.0 stable
 - Design Studio 3.7.0
 - Pentaho Dashboard Editor CDE bundle 1.0 RC3
 - Pentaho Schema Workbench 3.2.1.13885 RC1

Arquitectura Física



Implementación

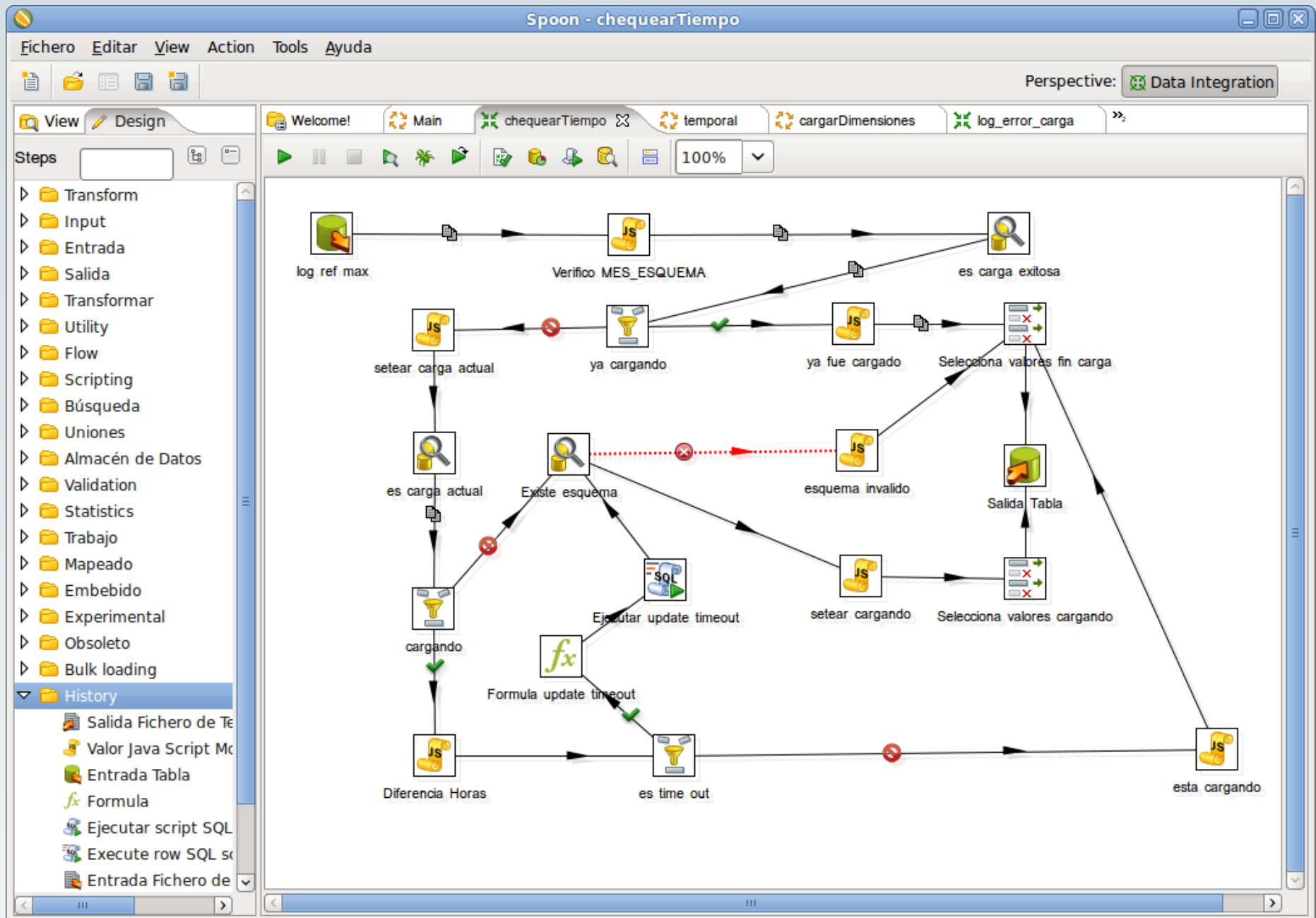
Timeline del proyecto



Resultados obtenidos

- Instalación de las plataformas seleccionadas
- Implementación del DW
 - Diseño de cubos
 - Desarrollo de procesos de ETL
 - Reportes, cuadros de mando e infografías
- Actualización periódica del DW de forma automatizada
 - Log de carga
- Experiencia adquirida a lo largo del proyecto

Resultados obtenidos – ETL



Resultados obtenidos - OLAP

Tiempo	Unidad Asistencial	Tipo de Cobertura											
		Todos	CARNE	Asistencia	Vitalicio	FONASA	Afiliados	De Oficio	CUOTAS	Convenios Colectivos	Individuales	Nucleo Familiar	CONVENIO
Marzo 2011	Todas	111.000	111.000	85.000	111.000	111.000	111.000	111.000	111.000	111.000	111.000	111.000	111.000
	Montevideo	85.000	85.000	65.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000	85.000
	Centro Salud Cruz de Carrasco	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Centro Salud Cerrito	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Centro Salud Cerro	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Centro Salud Giordano	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Centro Salud Jardines del Hipodromo	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Centro Salud Maciel	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Centro Salud Maronas	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Centro Salud Misurraco	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Centro Salud Piedras Blancas	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Centro Salud Sayago	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Centro Salud Union	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	CRAPS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Hospital Pineyro del Campo	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Hospital Saint Bois	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Artigas	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Cen. Aux. Bella Union	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Cen. Dep. Artigas	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Centro Salud Baltasar Brum	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Centro Salud Tomas Gomensoro	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Canelones	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Cen. Aux. Las Piedras	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Cen. Aux. Pando	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Cen. Aux. San Ramon	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Cen. Aux. Santa Lucia	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Cen. Aux. Tala	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Cen. de Salud Ciudad de la Costa	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Cen. Dep. Canelones	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000



Columnas ✖

Tipo de Cobertura

Filas

- ▣ ▾ ▾ Tiempo
- ▣ ▾ ▴ Unidad Asistencial

Filtro

- ▣ ▣ Grupos de Edad Fonasa
- ▣ ▣ Grupos de Edad Infamilia
- ▣ ▣ Grupos de Edad Otro
- ▣ ▣ Medidas
- ▣ ▣ Sexo

Resultados obtenidos - Reportes

Archivo Vista Ayuda

Navegar

- ASSE
 - Analysis
 - Indicadores
 - Dashboards
 - Afiliaciones
 - Infografias
 - Reportes
 - Indicadores

FonasaCapitas
FonasaXdepto
FonasaXunAsistenci..
Infamilia

▼ Report Parameters

Mes
Julio 2011

Output Type
HTML (Paginated)

View Report Auto-Submit

Page 1 of 1

Administración de los Servicios de Salud del Estado
 Departamento de Gestión de la Información
 Avda Dr. Luis A. de Herrera 3350 Piso 5 - Tel / Fax: 24816687 int. 523

agosto 10, 2011 @ 09:18

TOTAL DE USUARIOS DE ASSE, POR TIPO DE COBERTURA SEGÚN SEXO Y GRUPOS DE EDAD

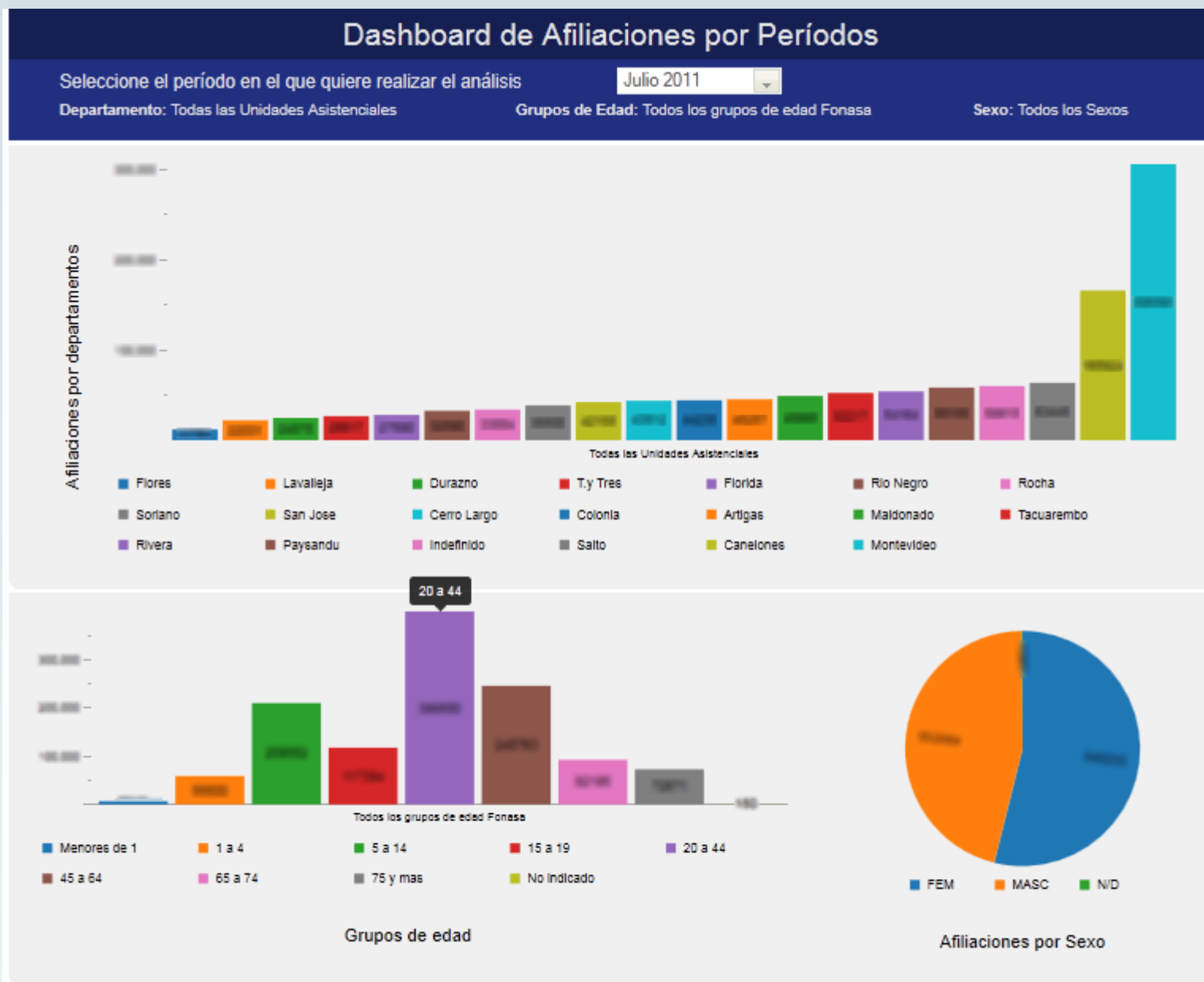
Julio 2011

SEXO	GRUPOS DE EDAD	TOTAL	CARNE Asistencia*	Vitalicio	FONASA	Afiliados	de Oficio	CUOTAS	Convenios Colectivos	Individuales	Núcleo Familiar	CONVENIO(**)
Todos	Todos											
Todos	Menores de 1											
Todos	1 a 4											
Todos	5 a 14											
Todos	15 a 19											
Todos	20 a 44											
Todos	45 a 64											
Todos	65 a 74											
Todos	75 y mas											
Todos	No indicado											
MASC	Todos											
MASC	Menores de 1											
MASC	1 a 4											
MASC	5 a 14											
MASC	15 a 19											
MASC	20 a 44											
MASC	45 a 64											
MASC	65 a 74											
MASC	75 y mas											
MASC	No indicado											
FEM	Todos											
FEM	Menores de 1											
FEM	1 a 4											
FEM	5 a 14											

Archivos

- FonasaCapitas
- FonasaXdepto
- FonasaXunAsistencial
- Infamilia
- InfamiliaXunAsistencial

Resultados obtenidos - Dashboards



Todos los valores mostrados en las gráficas son absolutos

Trabajo a futuro

- Extender el DW creado según nuevas necesidades de ASSE
 - + Cubos
 - + Reportes y Dashboards
 - + Indicadores
 - Data Mining
 - Integración con sistemas de información geográfica y geolocalización
- Optimizar performance en los procesos de carga y en las dimensiones

Conclusiones

- Se puede implementar una solución de DW y Business Intelligence Free/Open
 - en el tiempo de un proyecto de grado
 - en el marco de una institución estatal
 - a pesar de la baja calidad en los datos de origen
- Dificultades:
 - Existencia de bugs en los componentes utilizados
 - Constantes cambios de versión en la herramienta seleccionada
 - Trabajo remoto (por privacidad)

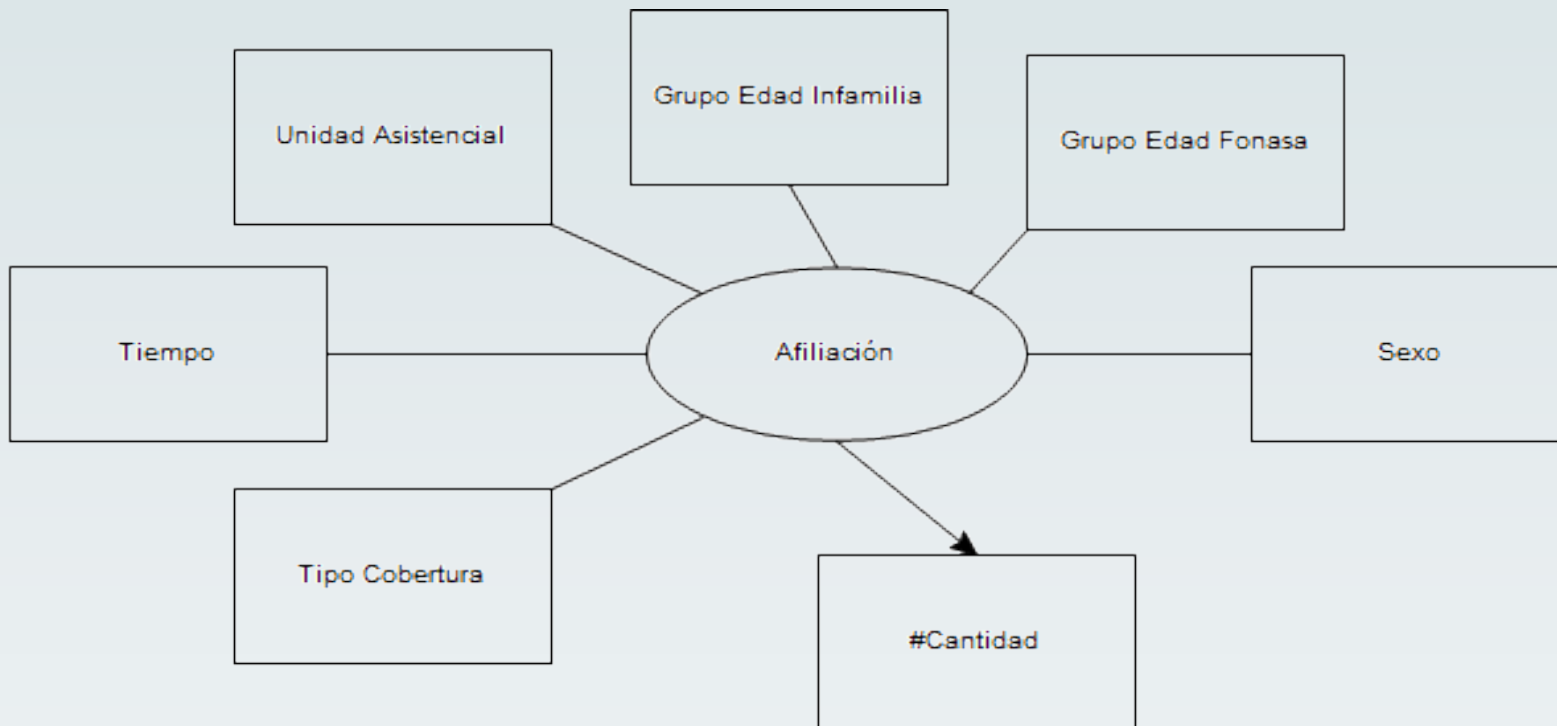
¿Preguntas?



Gracias por su Atención

Implementación

- Relación dimensional



Implementación

- Procesos de extracción, transformación y carga
- Cubos y OLAP
- Consultas Ad Hoc
- Reportes
- Cuadros de Mando
- Infografía
- Consola de Administración
- BI Server

Resultados obtenidos - Infografías

