JUNTA DE ANDALUCIA

Consejería de Agricultura, Pesca y Desarrollo Rural

Sistema BI de Pentaho

Arquitectura de Aplicaciones

Versión: v01r00
Fecha: 19/07/2016

Queda prohibido cualquier tipo de explotación y, en particular, la reproducción, distribución, comunicación pública y/o transformación, total o parcial, por cualquier medio, de este documento sin el previo consentimiento expreso y por escrito de la Junta de Andalucía.

HOJA DE CONTROL

Título	Sistema BI de Pentaho		
Entregable	Arquitectura de Aplicaciones		
Tipo	Procedimiento		
Nombre del Fichero	CAPDER_Arquitectura_Pentaho_0100.odt		
Autor	everis		
Versión	v01r00	**Fecha Versión**	19/07/16
Aprobado por		**Fecha Aprobación**	
		Nº Total Páginas	13

REGISTRO DE CAMBIOS

Versión	Causa del Cambio	Responsable del Cambio	Área	Fecha del Cambio
0100	Creación del documento	everis		19/07/16

CONTROL DE DISTRIBUCIÓN

Nombre/Apellidos	Cargo	Área	Nº Copias

ÍNDICE

1 INTRODUCCIÓN....4
1.1 Objeto....4
2 ARQUITECTURA....5
2.1 Directrices generales de arquitectura de aplicaciones....5
2.2 Requisitos de los nodos del servicio....7
3 PENTAHO....8
4 GLOSARIO....12
5 BIBLIOGRAFÍA....13

1 INTRODUCCIÓN

1.1 Objeto

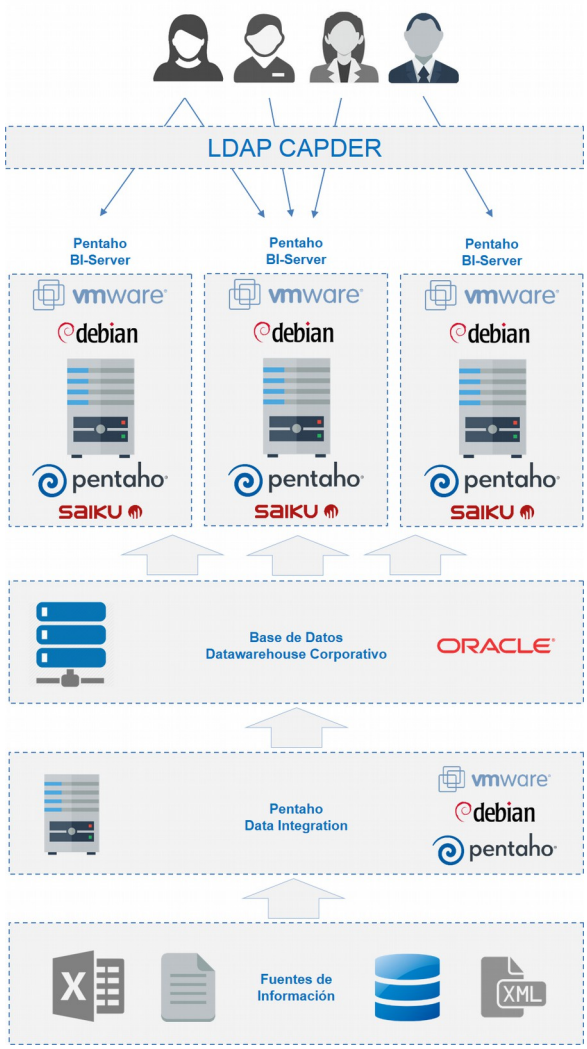
El objetivo de este documento es ofrecer una propuesta de arquitectura común para todas aquella aplicaciones que necesiten usar Pentaho como herramienta de explotación. Cuando hablamos de Pentaho hablamos del uso de cualquiera de las herramientas integradas en la suite, BI-Server, Report Designer, Schema-Workbench, Pentaho Data Integration y de plugins utilizados para una mejor experiencia de usuario como Saiku.

El objetivo de esta arquitectura es maximizar la disponibilidad, fiabilidad y capacidad de reacción ante incidencias y la usabilidad de la plataforma, permitiendo que esté disponible y con unos tiempos de respuesta aceptables en el mayor tiempo posible.

2 ARQUITECTURA

2.1 Directrices generales de arquitectura de aplicaciones

A continuación se presenta una propuesta de arquitectura para los diferentes proyectos que se realicen utilizando la solución de Pentaho en CAPDER.



- **ETL y Pentaho Server:** El escenario de arquitectura de servidores para procesos de carga y para servidor de Pentaho ha sido seleccionado entre las dos opciones que indicamos a continuación enumerando las principales ventajas e inconvenientes. Tras analizar el escenario actual y posibilidades a futuro, la opción más adecuada para CAPDER es la de alojar de manera independiente el servidor de carga y de Pentaho.

- Procesos ETL y Pentaho en una misma máquina: De esta forma tenemos todo integrado en un solo servidor por lo que el reducimos el consumo en mantenimiento e infraestructuras. En contra tendríamos los procesos de carga ejecutados disminuirían el rendimiento del servidor de Pentaho.

- Procesos ETL y Pentaho en máquinas separadas (opción CAPDER). Debido a que el servidor de carga se utilizará para todos los proyectos existentes, no debemos tener servidor de carga y de Pentaho en una misma máquina, ya que reduciría en considerablemente el rendimiento del servidor en caso de necesitar ejecutar cargas mientras se accediendo a la aplicación. Por contra existirán un mayor número de servidor lo que conlleva un mayor mantenimiento y consumo de recursos.

- **Proyectos Pentaho CAPDER**: Para otro tipo de aplicaciones las organizaciones tienen un servidor donde se alojan todas los desarrollos realizados con las mismas herramientas. En el caso de Pentaho es distinto, ya que por ejemplo, en un servidor podrían alojarse contenidos de distintos proyectos pero la gestión de usuarios y personalización de la plataforma sería común. En algunos casos puede ser útil pero en otros casos no. Particularizando en CAPDER, siempre que se trate de proyectos de explotación de datos diferentes se implementarán en servidores diferentes.

  - Utilización de un mismo servidor de Pentaho: En este caso como hemos comentado, habría que reutilizar toda la gestión de usuarios no teniendo porqué coincidir siempre la estructura de los mismos además de que todos los proyectos incluidos en el mismo servidor, compartirían la misma personalización y configuración. En cuanto a los beneficios, reducimos el consumo en mantenimiento e infraestructuras.

  - Utilización de diferentes servidores para cada aplicación: Nos posibilita completa independencia para proyectos que no tengan relación entre si pudiendo aplicar la configuración y personalización que deseemos y tratar de manera particular la gestión de usuarios. Por contra existirán un mayor número de servidor lo que conlleva un mayor mantenimiento y consumo de recursos.

Éstas son algunas de las posibilidades que ofrece la arquitectura de Pentaho que al tratarse de una arquitectura flexible se puede decidir aplicar en unos casos u otros en función de las necesidades de cada proyecto.

2.2 Requisitos de los nodos del servicio

La arquitectura hardware **mínima** para cada entorno que se utilice será la siguiente:

Servidor	Características	Backups
Pentaho Aplicaciones	• Arquitectura de 64 bits • 2 procesadores • 4 GB de RAM (dependerá del proyecto y número de usuarios) • 10 GB de espacio en disco (dependerá del proyecto) • SO Debian (versión determinada por CAPDER) • Java 1.7 o superior (dependiendo versión de Pentaho) • Pentaho BI-Server 5.4 o superior • Plugin Saiku • Visibilidad sobre BD DW	En función de los cambios de configuración. Por defecto quincenal.
Pentaho Data Integration	• Arquitectura de 64 bits • 4 procesadores • 8 GB de RAM • 10 GB de espacio en disco • SO Debian (versión determinada por CAPDER) • Java 1.7 o superior (dependiendo versión de Pentaho) • Pentaho Data Integration • Visibilidad sobre BD DW y Fuentes de Información	En función de los cambios de configuración. Por defecto quincenal.
BBDD DW	• Arquitectura de 64 bits • 4 procesadores • 8 GB de RAM • 40 GB de espacio en disco • SO Determinado por Sistemas • Oracle Database (versión determinada por CAPDER)	En función de la periodicidad de las cargas Por defecto semanal.

3 PENTAHO

Se trata de una plataforma de BI orientada a la solución y centrada en procesos que incluye todos los principales componentes requeridos para implementar soluciones basadas en procesos.

Pentaho se compone de una infraestructura de herramientas de análisis e informes integrado con un motor de workflow de procesos de negocio.

Está formado por los siguientes componentes:

- **Plataforma BI Server:** Se encarga de la visualización de la información en múltiples formatos: Cuadros de mandos, vistas de análisis e informes.
- **Data Integration:** Es la parte encargada de la extracción, transformación y carga de la información en el almacén de datos. Permite la conexión a múltiples orígenes de datos, como pueden ser bases de datos relacionales, bases de datos NoSQL, Hadoop, ficheros, etc.

Tanto la plataforma BI Server como Data Integration, necesitarán de la existencia de una máquina Java para su correcto funcionamiento y existen distintas opciones para la configuración de dicha máquina.

- Instalación de java en la propia carpeta de Pentaho.
- Crear la variable de entorno JAVA_HOME con la ruta de instalación de java.
- Añadir un enlace simbólico a la instalación de java en la carpeta de Pentaho.

El script de arranque de Pentaho buscará la configuración de java teniendo en cuenta los tres casos anteriores.

3.1.1 Plataforma BI Server

Los componentes que se utilizan en la plataforma de Pentaho Business Analytics se pueden dividir en las siguientes categorías:

- BI Server y consola de usuario.
- Plugins y herramientas de diseño.

3.1.1.1 BI Server y consola de usuario

BI Server constituye el núcleo central de la suite de Pentaho, siendo el responsable de procesar los informes, análisis y contenidos de la plataforma. Además es el encargado de controlar la seguridad, compartir los contenidos del repositorio, planificar los trabajos y de la auditoria de los datos.

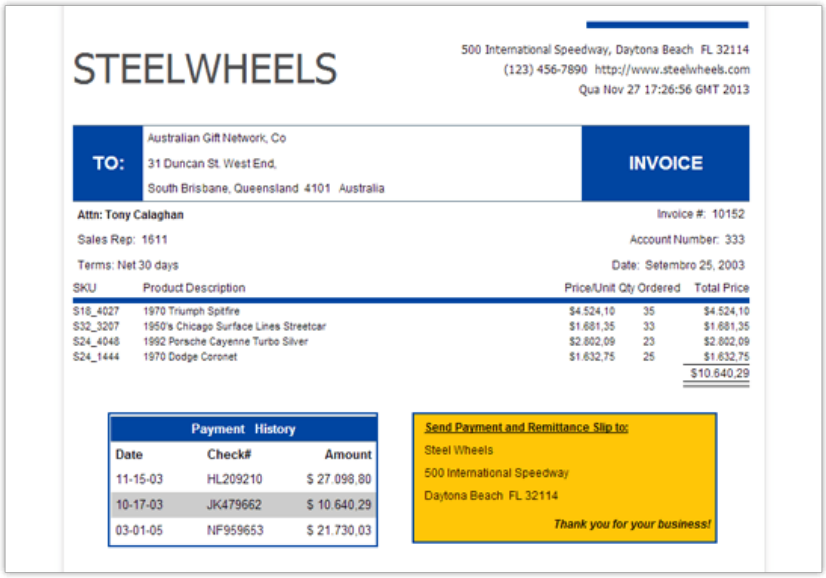
BI Server se gestiona a través de una consola de usuario, a través de la que se crean los contenidos de los cuadros de mandos y de los informes o de administrar la seguridad de Pentaho.

Con respecto a la seguridad, Pentaho dispone de dos posibilidades para la gestión de usuarios y los permisos de acceso de los mismos. Una de ella es a través de la consola de administración y almacenando los datos en una base de datos propia. La otra posibilidad es un integración con LDAP, cediendo la gestión de dichos usuarios y roles a dicho LDAP.

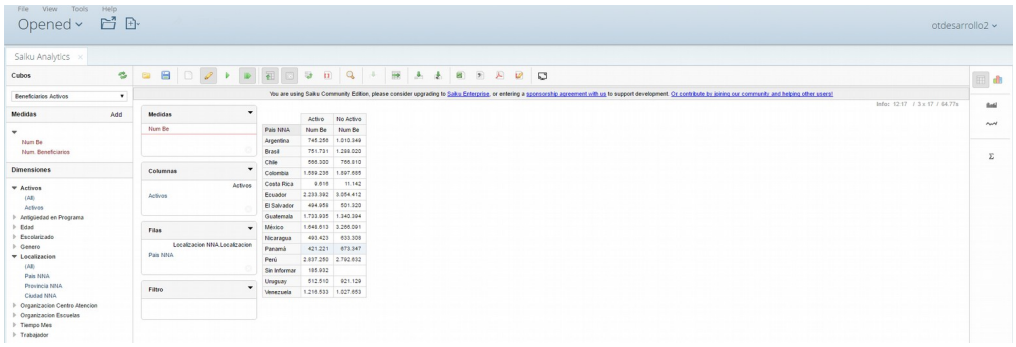
3.1.1.2 *Plugins y herramientas de diseño*

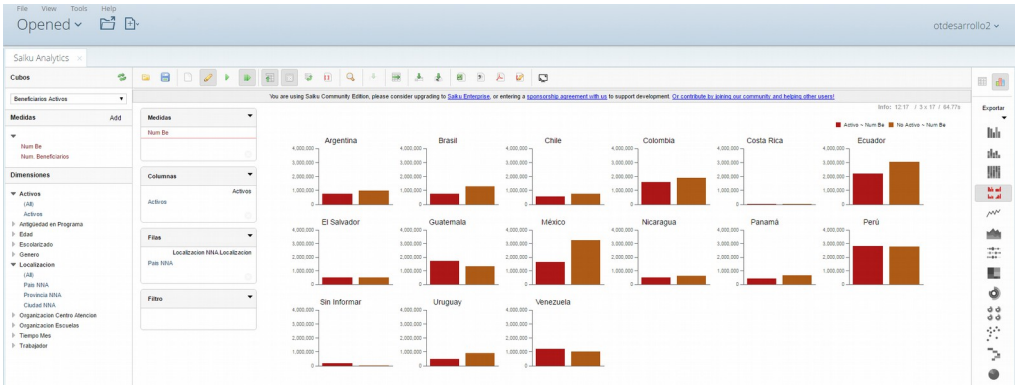
Para la visualización de la información se contemplan una serie de herramientas que se enumeran a continuación:

- **Report Designer:** El Report Designer es una herramienta gráfica que genera informes de datos (pixel perfect) a través del motor de integración de datos, sin la necesitad de crear tablas intermedias. Puede dar salida a los informes en varios formatos, incluyendo PDF, Excel, HTML, etc.

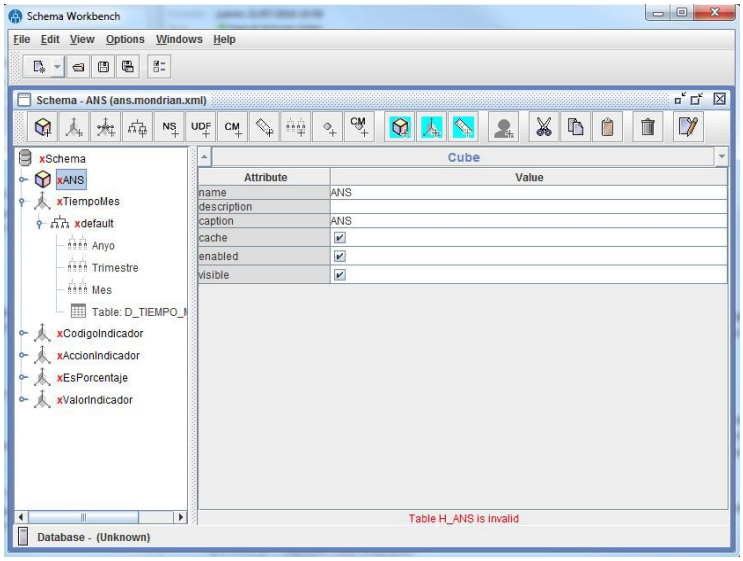


- **Saiku:** Saiku es un excelente visor OLAP que proporciona al usuario una herramienta para realizar análisis de forma fácil e intuitiva. Permite una gran variedad de visualizaciones sobre los datos, georeferenciar la información, filtrar datos, añadir parámetros a las consultas, etc.





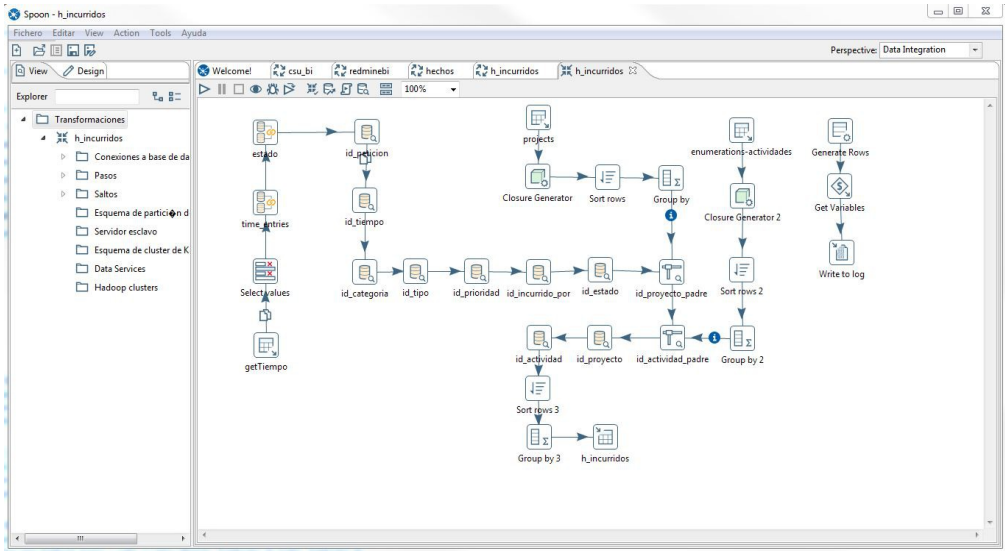
- **Schema Workbench:** El Schema Workbench es una interfaz visual que permite editar y crear modelos mondrian multidimensionales, que constituyen la base para los análisis con Saiku.



3.1.2 Pentaho Data Integration (PDI)

Pentaho Data Integration es la plataforma que permite el tratamiento de la información (extracción, transformación y carga) para la creación de un almacén de datos que permita el análisis posterior de esos datos por parte de los usuarios finales.

El tratamiento de la información se realiza con la herramienta de diseño **Spoon**. Con esta herramienta se diseñan los procesos de carga, a través de una interfaz visual intuitiva de gran riqueza que dispone de más de 200 pasos pre diseñados para realizar las diferentes operaciones implicadas en el proceso de integración de los datos.



4 GLOSARIO

A continuación se indican los términos utilizados, en el presente documento:

Término	Descripción
DW	DataWarehouse
PDI	Pentaho Data Integration
ETL	Extract, Transform and Load (Extracción, Transformación y Carga)
CAPDER	Consejería Agricultura, Pesca y Desarrollo Rural
SO	Sistema Operativo
OLAP	On-Line Analytical Processing
BI	Business Intelligence
BD	Base de Datos

5 BIBLIOGRAFÍA

Se desglosa a continuación el material de soporte, utilizado para la elaboración del presente documento.

Referencia	Título	Código