

Operational Data Store (ODS)

Rosa María Castillo

Div. de Servicios de Redes de Datos
Telefónica I+D
28037 Madrid
rmcc@tid.es

Jesús Morata

Div. DataWareHouse para Telefónica
de España
Telefónica I+D
28037 Madrid
jmmgp@tid.es

Luis del Arbol

Div. DataWareHouse para Telefónica
de España
Telefónica I+D
28037 Madrid
lpap@tid.es

1. Resumen

El ODS es un sistema destinado a liberar a los sistemas operacionales de realizar las labores de query & reporting.

Tiene como objetivo general unificar la plataforma de almacenamiento donde se copien los datos de todas las aplicaciones en un solo sistema.

2. Motivación. ¿Qué es el ODS?

ODS es un sistema que recopila la información existente en los Sistemas Operacionales y la ofrece a los Sistemas Informacionales o directamente a los Usuarios Finales de Negocio.

Para ello existe una plataforma común donde conviven tanto la información procedente de los Sistemas Operacionales (Subsistema ODS Replicación) como la información integrada en Sistemas Informacionales (Subsistema ODS Analítico). Esta plataforma contiene herramientas/procedimientos de extracción, transformación, carga y consulta de la información común para todas las aplicaciones que se encuentran en el entorno del ODS.

Toda aplicación que desee entrar en el Sistema ODS deberá cubrir como mínimo el Subsistema ODS Replicación, es decir, deberá dejar la información para ser consultada por el usuario final y/o para ser fuente de información hacia sistemas informacionales.

Una vez que la información esté almacenada en la parte operacional del ODS, esta deberá ser puesta a disposición de los usuarios para que pueda ser consultada. Para ello tendrá dos alternativas:

- El Subsistema ODS Replicación podrá ser fuente de información hacia usuarios finales a través de una capa de presentación lo menos transformada posible a la información existente en los Sistemas Operacionales, pero intentando adaptar los datos operacionales hacia una información más sencilla de analizar.
- Si la necesidad de información de los usuarios esta orientada a un análisis más complejo, la aplicación podrá optar por ofrecer la información a través de un sistema informacional (Data Mart) implementado dentro del ODS (sobre el Subsistema ODS Analítico) o fuera del entorno ODS.

En cualquier caso, siempre que una aplicación entre en el entorno ODS deberá suministrar como mínimo al Subsistema ODS Replicación la información que necesite de los sistemas operacionales y que aun no se encuentre almacenada en el ODS.

Los sistemas informacionales:

- Se alimentan de datos provenientes de los sistemas de gestión u operacionales, con la periodicidad que se precise. Para esto, es necesario diseñar los procesos de carga de datos de los sistemas al ODS, y configurar las autorizaciones de acceso pertinentes.
- Sirven para consultar datos (Query), generar informes (Reporting) o realizar estudios y análisis de información (OLAP, DSS, EIS, Cuadros de Mando, Data Mining, etc).

Con ODS se solucionan varios de los problemas inherentes de los Data Marts tradicionales:

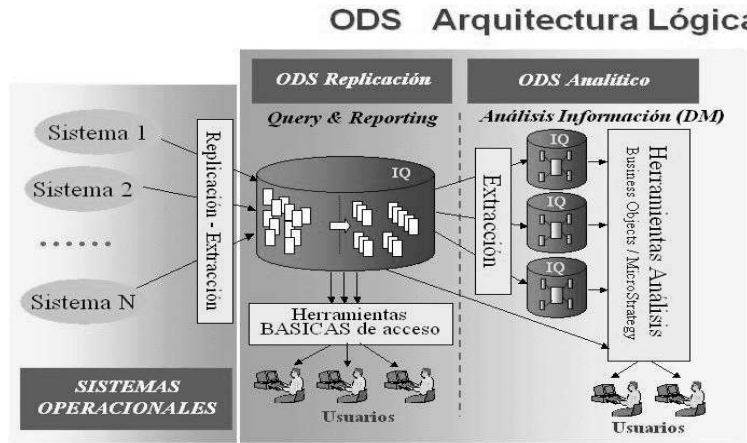


Figura 1. ODS. Arquitectura Lógica

- La extracción de información no afecta a la ejecución de trabajos en los Sistemas Operacionales.
- La información se replica en un sistema paralelo, y no en los propios Sistemas Operacionales.
- Se evita la sobrecarga de los equipos donde se albergan los Sistemas Operacionales con acciones externas al propio sistema propietario.

Por lo tanto, aporta las siguientes ventajas:

- Ahorro de costes por la reutilización de recursos.
- Disponibilidad de un dato único, evitando múltiples copias del mismo dato en sistemas dispersos.
- Posibilidad de cruzar la información de diferentes aplicaciones.

La arquitectura del ODS se basa en un gestor de base de datos Sybase Adaptive Server IQ Multiplex desarrollado específicamente para cubrir las necesidades de un sistema de Consulta y Análisis de Información. La modelización física difiere algo de los sistemas tradicionales, por lo que existen diferencias detalladas en los procedimientos al efecto.

3. Descripción funcional

La orientación del Sistema ODS es hacia un sistema informacional, no hacia un sistema transaccional, ya que la funcionalidad principal es el ser fuente de información para usuarios finales directamente o a través de sistemas informacionales. Por tanto, la disponibilidad y criticidad del sistema deben ser similares a la de un sistema informacional (como es el caso de un Data Warehouse) teniendo en cuenta que los tiempos de respuesta y la disponibilidad de la información son lo más óptimos posible pero sin tener la necesidad de llegar al mismo nivel que un sistema transaccional.

Debido a la posible dependencia de información que podría existir entre los distintos sistemas operacionales, el ODS cubre la funcionalidad que estos ofrecen pero teniendo en cuenta las premisas anteriores, es decir, que terceros sistemas que accedan al Sistema ODS para recoger información deben tener una necesidad de dicha información similar a la de un sistema informacional, y que la disponibilidad de la información no va ser como si esta se encontrara en un sistema transaccional.

Con esto se consigue que el Sistema ODS no sea el pivote sobre el que giren todos los demás sistemas, sino únicamente los sistemas informacionales, con lo que una inactividad del

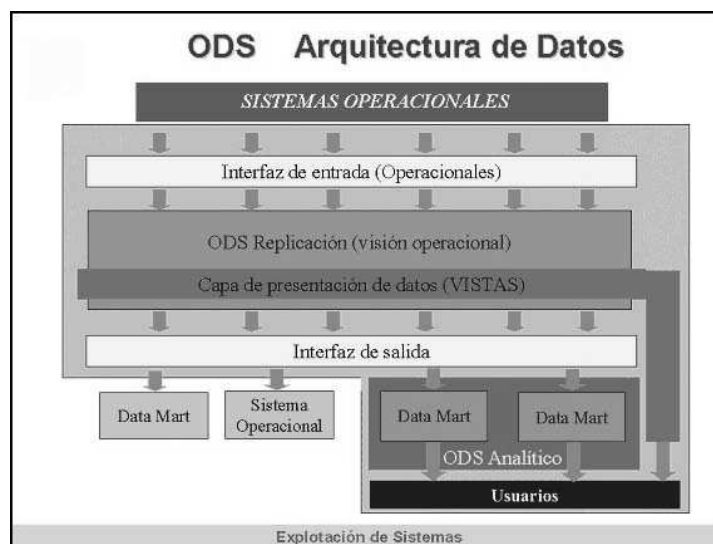


Figura 2. ODS. Arquitectura de Datos

ODS no provocaría dependencias en posibles sistemas críticos.

4. Arquitectura de Datos

El Sistema ODS mantendrá una interfaz de datos con los sistemas y usuarios con los que interactúa:

- **Interfaz de entrada (Operacionales):** Interfaz que se establece con los sistemas operacionales mediante la cual se recoge la información de estos para almacenarla en el Subsistema ODS Replicación.
- **Interfaz de salida:** Interfaz que el Subsistema ODS Replicación mantiene con:
 - El Subsistema ODS Analítico para suministrar la información a los sistemas informacionales que se encuentran dentro del entorno ODS.
 - Los sistemas informacionales que se encuentran fuera del alcance del entorno ODS.
 - Ciertos sistemas operacionales que necesitan información con una baja criticidad.
- **Capa de presentación:** Interfaz que el Subsistema ODS Replicación mantiene con los usuarios finales. La información de los sistemas operacionales, que se encuentra almacenada en el Subsistema ODS Replicación es accesible para los usuarios a través de una capa de presentación. Esta capa

se crea mediante Vistas a partir de las Tablas de la Base de Datos.

- **Interfaz de los Sistemas Informacionales:** Interfaz creada para la accesibilidad de los usuarios finales a los sistemas informacionales del ODS. Esta interfaz se establecerá a partir de las herramientas de usuario final utilizadas por cada uno de los sistemas.

5. Servicios ofrecidos

El Sistema ODS ofrece diferentes tipos de servicios, dependiendo del destino de la información.

5.1. Servicios a usuarios finales

- A través de una **capa de presentación** de la información que se encuentre en el Subsistema ODS Replicación. Esta franja de presentación tendrá como primer nivel una capa de VISTAS de la Base de Datos, no permitiendo en ningún caso que los usuarios puedan acceder directamente a la información que se almacena en las tablas de la Base de Datos, consiguiendo por un lado una mayor seguridad de los datos y por otro ofrecer una visión de la información un poco más cercana a la visión de negocio de los

usuarios finales. Este es un servicio similar al Ofrecido por el Centro de Información.

- A través de **Subsistemas Informacionales** que convivan dentro del Sistema ODS. Estos Subsistemas recogerán la información a partir del Subsistema ODS Replicación y seguirán la Arquitectura específica del entorno ODS. Tendrán herramientas/procedimientos de extracción, carga y consulta de la información, comunes para todos y definidos a medida que el Sistema ODS evolucione. La forma de abastecer de información a estos sistemas se realizará mediante acceso directo desde la Base de Datos del Sistema Informacional a la Base de Datos del Subsistema ODS Replicación.

5.2. Servicios a otros sistemas

- A **Sistemas Informacionales** que no se encuentren dentro de la plataforma definida para el Sistema ODS. Es decir podrán ser Data Marts que no sigan las especificaciones del ODS pero que si necesiten información que se encuentre o pueda ser almacenada en el Subsistema ODS Replicación. La forma de abastecer de información a estos sistemas se realizará de forma similar a como se realiza cuando el Sistema Informacional se encuentra dentro del ODS:
 - Ficheros planos descargados a partir de las tablas de la Base de Datos del Subsistema ODS Replicación.
 - Acceso directo desde el Sistema Informacional a la Base de Datos del Subsistema ODS Replicación.
- A Sistemas **Operacionales** con ciertas necesidades informacionales de otros sistemas para su operativa, teniendo en cuenta que la **criticidad** de la información que solicitan debe ser muy baja. Esto es debido a que las funcionalidades principales del ODS están orientadas a ofrecer información de negocio directamente a usuarios finales o a través de Subsistemas Informacionales que convivan dentro del Sistema ODS. Este servicio pretende cubrir la funcionalidad de ofrecida por los Sistemas Operacionales a terceros sistemas a través de servicios NA. En el ámbito del ODS, esta

III Taller de Minería de Datos y Aprendizaje

funcionalidad se realizará mediante la implementación de procedimientos con interfaz Java para los casos en los que los sistemas que soliciten este servicio se encuentren en un entorno Unix.

6. Arquitectura lógica

El Sistema ODS, está compuesto por los siguientes subsistemas:

- Subsistema de Información.
- Subsistema de Procesamiento
- Subsistema de Acceso
- Subsistema de Transferencia

6.1. Subsistema de Información

Es el subsistema principal, contiene la información que reside en el ODS y que es fuente para los usuarios finales o para otros sistemas.

- **BD ODS Replicación**

Almacena los datos replicados de los sistemas operacionales. Esta información es fuente para otros sistemas o para el usuario final a partir de la capa de presentación –vistas- (funcionalidad de Centro de Información).

Sobre las Tablas de la BD existe una capa de presentación necesaria para el usuario generada a partir de Vistas de la BD.

Provee a los nuevos desarrollos de un amplio catálogo a través de una Interfaz de Acceso, sirviendo como fuente de información para los nuevos desarrollos. Sólo en aquellos casos en donde la información resida en fuentes externas al ODS, será necesario el desarrollo y mantenimiento de procesos de carga por parte del equipo de desarrollo del proyecto. En este caso, el es propio usuario final quien realiza consultas al ODS, pudiendo extraer un máximo de registros en cada query.

Las características principales del ODS Replicación son:

- Acceso a datos del operacional.
- Herramientas de acceso de bajo coste, siendo estas las propias herramientas ofimáticas del usuario.
 - Acceso mediante query libre y reporting desde Excel y Access.

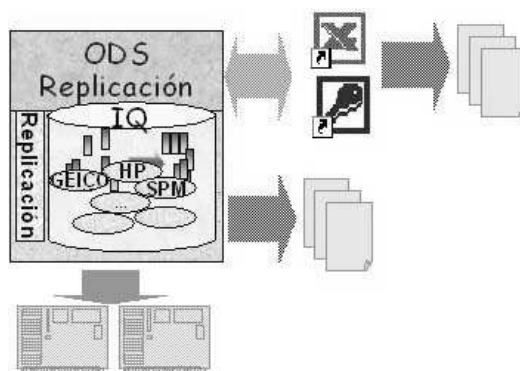


Figura 3. ODS Replicación

- Acceso a la Información desde aplicaciones externas (windows).
- Utilidades de extracción de información a Fichero Plano.
- Disponibilidad de información del operacional, constituida por el conjunto de tablas exportadas por cada Operacional de acuerdo a su propio modelo de información interno, ofreciendo:
 - Facilidades de query y reporting sin sobrecarga sobre los sistemas operacionales.
 - Posibilidad de realizar cruces de información procedente de diferentes sistemas operacionales.
 - Cumplimiento de los estándares de seguridad (integración con LDAPs y cumplimiento requerimientos APD).
- Interfaz ODBC al ODS. El acceso de los usuarios se realiza a través de un interfaz específico desarrollado al efecto, que hace posible tanto el acceso a grandes volúmenes de datos, como el mantenimiento de un control y auditoria centralizados del uso del ODS.
- Ventajas tecnológicas:
 - Reducción considerable en tiempo de respuesta.
 - Optimización del volumen para almacenamiento de información (0,8).
 - Reducción considerable de tiempos de Carga (20 Mill Reg / 5 min).

- **BD DataMart 1, BD DataMart 2, ... BD DataMart n**

Conjunto de Bases de Datos que, cada una de ellas, contiene la información de un Sistema Informacional (Data Mart).

Cuando una aplicación requiera implementar un Sistema Informacional sobre el ODS, deberá generarse una nueva Base de Datos de este tipo, es decir, orientadas a Data Mart.

El subsistema analítico se utiliza en aquellos casos en los que se requiera un sistema de análisis de la información que obligue a una transformación de los datos, la agregación, y la creación de métricas y dimensiones, ampliando el ciclo de desarrollo a la construcción de un DataMart, pero siempre como ampliación del Subsistema de Replicación. De este modo, los datos fuente de información para el Data Mart, deben residir en el subsistema de Replicación.

Dentro de esta Base de Datos existirá:

- Un usuario para el Data Mart que será el propietario de las tablas.
- Un usuario de carga, utilizado en los procesos que cargan las tablas del Data Mart.
- Las herramientas de análisis de usuario final tendrán un usuario en la base de datos del Data Mart.

6.2. Subsistema de Procesamiento

Este subsistema residirá en una máquina diferente a la del Subsistema de Información. En ella se encuentra el software de DataStage (herramienta de extracción-transformación-carga utilizada en el ODS).

-

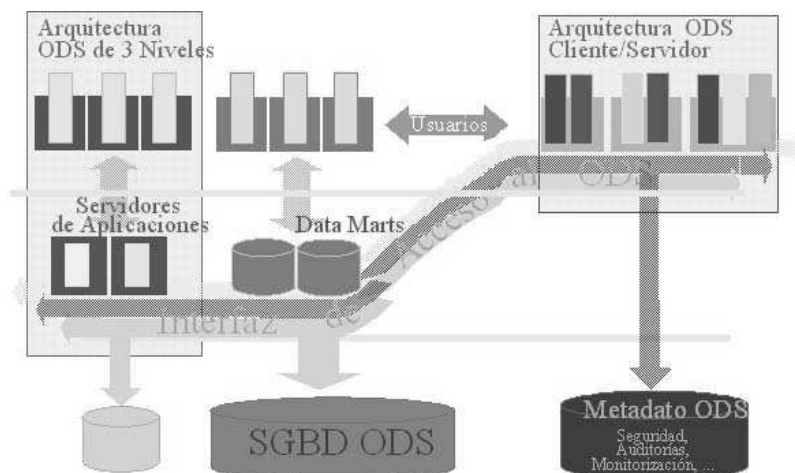


Figura 4. ODS. Interfaz de Acceso

6.3. Subsistema de Acceso

Es el subsistema de acceso de los usuarios. Para los Data Mart (ODS Analítico) se utilizan herramientas de consulta y análisis de la información (Business Objects y MicroStrategy), además del IIS (Internet Information Server) de forma que el acceso a la información se realice de forma Web.

6.4. Subsistema de Transferencia

En el caso que la interfaz entre el Sistema ODS y los sistemas operacionales necesite un espacio en disco intermedio para la transferencia de la información, este espacio se encontrará en un sistema NAS (*Network Attached Storage*). Éste es un sistema de discos externos dedicado a proveer y compartir datos dentro de una red de área local.

7. Otras consideraciones técnicas

A continuación indicamos algunas consideraciones a tener en cuenta respecto al sistema ODS, relacionadas con el sistema gestor de base de datos SYBASE IQ utilizado, en pos de obtener el mayor rendimiento posible:

- En Sybase IQ se permite la utilización de tablas temporales para la carga de los datos, siempre y cuando la gestión de dichas tablas sea la correcta, es decir, se deben borrar de la

base de datos cuando ya no sean necesarias para el procesamiento de la información.

- La carga total de tablas debe ir precedida del borrado de los datos de dicha tabla, para lo cual se recomienda utilizar el comando *'truncate'*.
- Siempre que la opción para cargar la información en el ODS, sea a través de ficheros, debe tenerse en cuenta que estos no pueden contener caracteres de control debido a restricciones del entorno NAS (por ejemplo el carácter 3F) o de las utilidades de carga del SGBD del ODS (por ejemplo carácter 00 – *nulo*-). En el caso que la información origen contenga este tipo de datos, se debe realizar un tratamiento de la misma para eliminarlos.
- Los nombres de todos los objetos de la base de datos (tablas, usuarios, índices...) deben crearse en mayúsculas.

8. Conclusiones

El ODS es un sistema destinado a liberar a los sistemas operacionales de realizar las labores de query & reporting.

Contiene un repositorio de información compuesto por un conjunto de tablas replicadas de los sistemas operacionales. La cadencia de replicación se define en función de las necesidades de disponibilidad de la información.

También dispone de otro repositorio, con información extraída, tanto agregada como de detalle, del replicación, en el que residirán los

datos utilizados por los sistemas de toma de decisión, cuadros de mando, Data Marts, y en general sistemas Business Intelligence.

El acceso al ODS se realiza a través de:

- Herramientas ofimáticas (Excel y Access)
- Herramientas BI (MicroStrategy, Business Objects, ...)
- Los usuarios dispondrán de sus propias áreas de usuario dentro del ODS, sobre las cuales depositar sus propias tablas, permitiéndose el cruce de esta información con la del ODS.
- Las aplicaciones pueden acceder al ODS, siempre bajo la supervisión y monitorización de los módulos de seguridad y auditoría del sistema.

Agradecimientos

Queremos expresar nuestro sincero agradecimiento a Dña. Paloma Granado García-Bernalt, D. Jorge González Pérez, D. José Miguel Hernández López De Munain y a D. Rafael Hernández Fernández, quienes, con su experiencia, conocimiento organizativo y técnico, nos han facilitado la elaboración de la presente ponencia.

También queremos expresar nuestro agradecimiento a D. Tomás Cruz Serrano y a D. José Antonio García Fernández por el total apoyo ofrecido para la elaboración de la misma.

Referencias

- [1] INMON, W. H.; RUDIN, KEN; BUSS, CHRISTOPHER K.; SOUSA, RYAN. "DATA WAREHOUSE PERFORMANCE". JOHN WILEY & SONS, INC., 1999
- [2] TODMAN, CHRIS. "DESIGNING A DATA WAREHOUSE". PRENTICE HALL PTR, 2001
- [3] ADELMAN, SID ; TERPELUK MOSS, LARISSA. "DATA WAREHOUSE PROJECT MANAGEMENT". ADDISON-WESLEY, 2000