



#START014

IBM BusinessConnect

El arte de lo posible

Uso de IBM DB2 Analytics Accelerator en Bankia

Borja Argüelles Monforte

Madrid, 20 noviembre 2013
PALACIO MUNICIPAL DE CONGRESOS

INDICE

- Información General de Bankia.
- Antiguo Sistema Informacional.
- PureData System for Analytics (PDA) e IBM DB2 Analytics Accelerator en Bankia.
- Nuevo Sistema Informacional.
- Ahorro en la ejecución de queries.
- Conclusiones & Nuevos Retos.

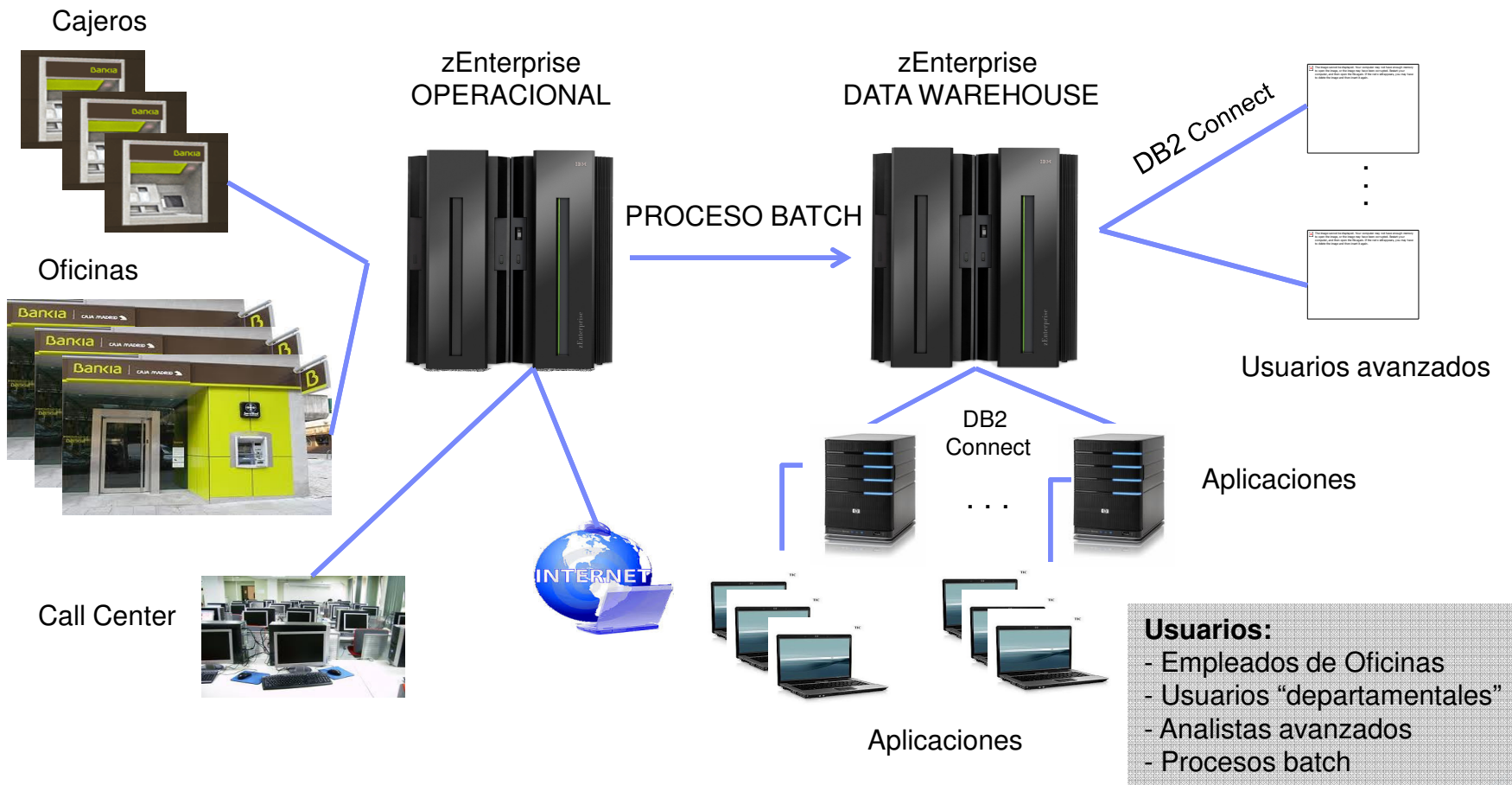


Información General de Bankia

- Más de 7 millones de clientes en todos los segmentos: clientes individuales, pequeñas y medianas empresas, grandes corporaciones e instituciones públicas y privadas.
- Un balance con más de 282.000 millones de euros en activos totales.
- Ingresos de más de 313,000 millones de euros.
- Más de 3,000 oficinas y 15.000 empleados.
- Modelo de banca universal basado en la gestión multicanal.
- Actividad concentrada en España con un modelo de negocio basado en la diversificación geográfica.
- Presencia internacional en los Estados Unidos, China, Irlanda y Portugal.



Antiguo Sistema Informacional



El entorno informacional es alimentado por los datos del entorno operacional.



Antiguo Sistema Informacional



De forma periódica los aplicativos operacionales dejan datos en el entorno informacional.

Posteriormente, **los datos se vuelven a procesar** para generar más datos, agrupar históricos y preparar reports. De esta forma, la información queda lista para que el usuario pueda obtenerla de forma rápida.

Con una nueva petición de información por parte del usuario, es necesario **desarrollar nuevos programas** y procesos batch que tomen la información operacional y “pre-cocinen” los nuevos datos.

Este tipo de procesos son grandes consumidores de CPU y tiempo de ejecución, por lo que se corre el peligro de agotar los recursos de la máquina zEnterprise y que se nos quede pequeña la ventana de ejecución.

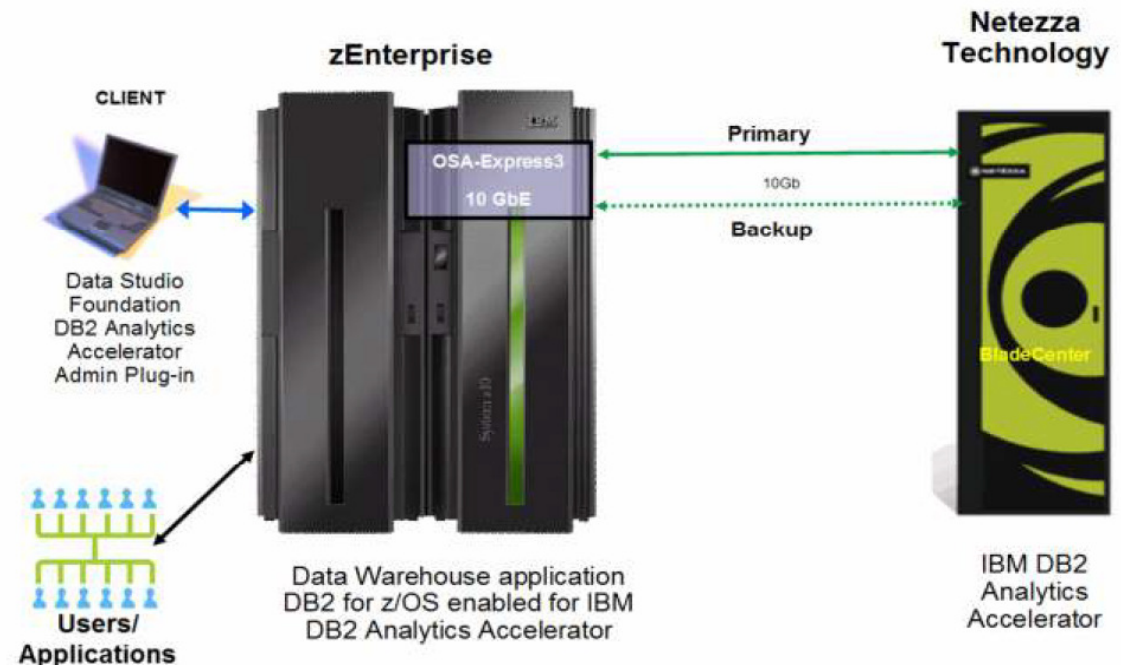
Así pues, el **modelo tradicional de Data Warehouse** es un modelo lento y gran consumidor de recursos, lo que no permite crecer en la medida que demanda el negocio.



PDA e IBM DB2 Analytics Accelerator en Bankia

A finales del año 2012 instalamos un sistema PDA - IBM DB2 Analytics Accelerator con las siguientes características:

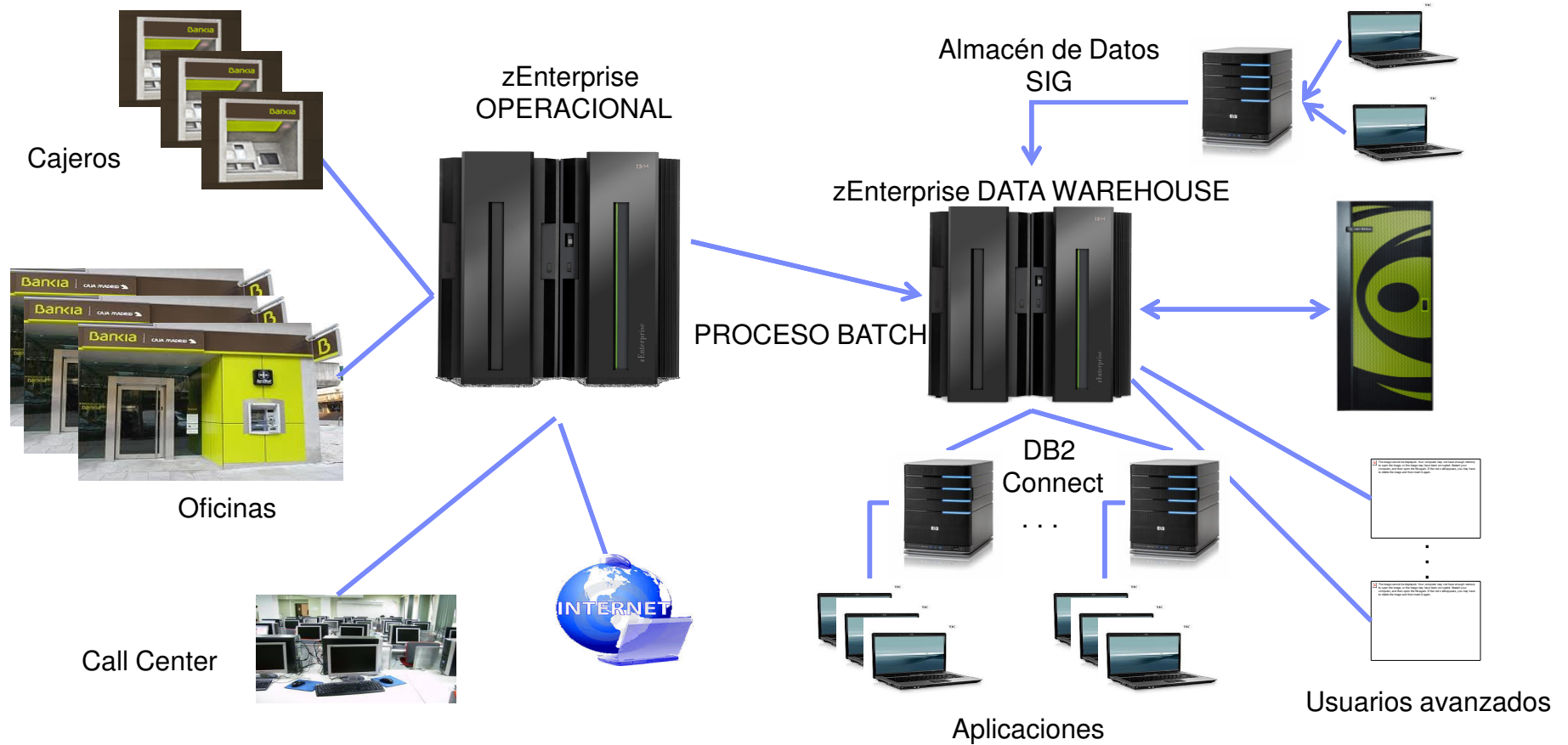
- Appliance Netezza 1000 modelo 3
- 1 quarter rack
- 24 procesadores
- 8 TB de capacidad de datos comprimidos equivalentes a 32 TB sin comprimir aproximadamente.
- 2 tarjetas OSA de 10 GB para conexión con del Z/OS.
- Conectado a un DB2 V9 del sistema informacional.



El DB2 Analytics Accelerator assessment previo realizado por IBM nos arrojó que el **95 %** de nuestras queries actuales podrían ser aceleradas en mayor o menor medida.



Nuevo Sistema Informacional



El nuevo sistema informacional está basado en el antiguo, pero se le ha añadido una máquina PDA, trabajando como acelerador de queries.



Nuevo Sistema Informacional

- De forma periódica los aplicativos operacionales dejan datos en el entorno Informacional, para que puedan ser cargados mediante utilidades en el DB2 zOS.
- Los datos se refrescan en las tablas del PDA.
- La dirección pide qué informes son necesarios.
- Los analistas generan queries necesarias para los informes pedidos, bien mediante las herramientas de reporte, bien mediante “queries libres”.
- Los propios directores generan sus propios informes “ad-hoc” mediante el uso de las herramientas de reporte sin mucho esfuerzo y en un tiempo razonable.



Ahorro en la ejecución de queries

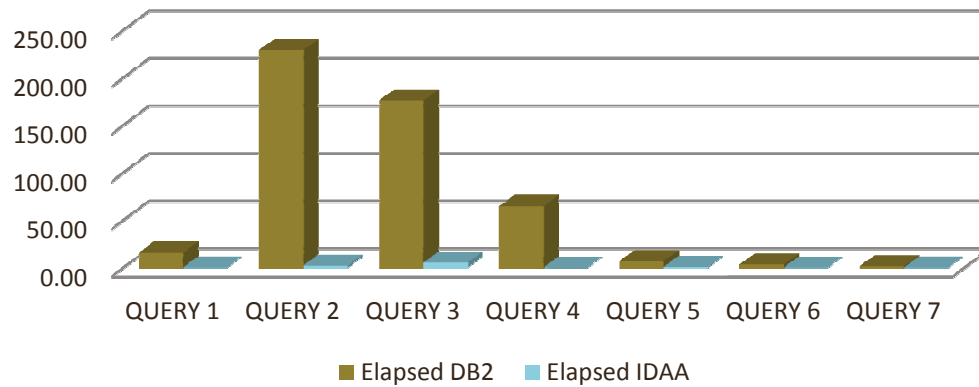
Query	Explicación
QUERY 1	Query no muy compleja. Menos de cinco joins. Tablas no excesivamente grandes.
QUERY 2	Select Distinct, join de tres tablas muy voluminosas con dos subqueries. La query devuelve pocas filas.
QUERY 3	Query muy compleja con más de veinte joins. La query devuelve unos cuatro mil registros.
QUERY 4	Select count de una query no muy compleja, accediendo solo a cinco tablas, pero muy voluminosas.
QUERY 5	Join de dos tablas muy voluminosas. La query devuelve gran cantidad de datos, sobre medio millón de filas y más de 3 MB.
QUERY 6	Query simple que accede solamente a una tabla muy voluminosa.
QUERY 7	Query simple que accede a dos tabas voluminosas con un join. Devuelve ciento cincuenta millones de registros y el resultado pesa más de 5MB



Ahorro en la ejecución de queries



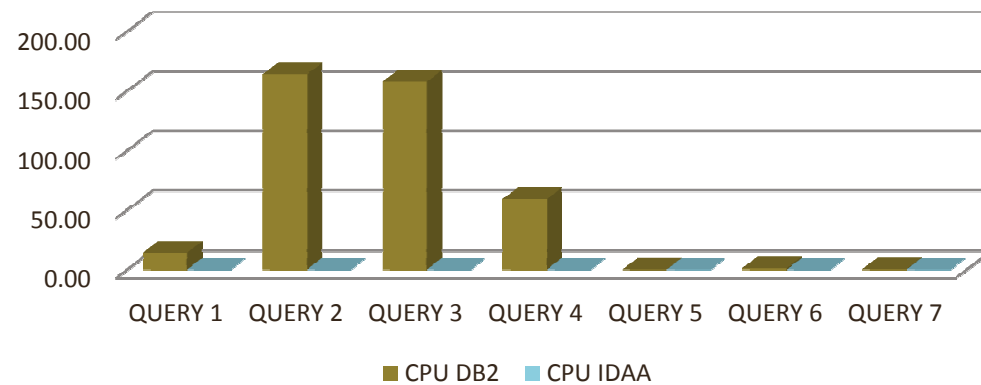
Elapsed Time (Minutos)



Query Number	Elapsed DB2	Elapsed PDA
QUERY 1	15,88	0,21
QUERY 2	229,07	2,02
QUERY 3	176,40	5,21
QUERY 4	65,09	0,02
QUERY 5	7,16	0,87
QUERY 6	3,39	0,06
QUERY 7	1,62	0,15

Query Number	CPU DB2	CPU PDA
QUERY 1	14,45	0,00
QUERY 2	164,09	0,00
QUERY 3	158,15	0,00
QUERY 4	59,33	0,00
QUERY 5	0,21	0,04
QUERY 6	1,23	0,00
QUERY 7	0,53	0,03

CPU Time (Minutos)



Ahorro en la ejecución de queries

Tal como se muestra, **la CPU consumida por las queries es prácticamente cero**, ya que se ejecuta en el PDA, y el tiempo de respuesta se reduce significativamente.

El ahorro de CPU se produce cuando se ejecuta la consulta. Este suele ser el período "Online", donde se producen más picos de CPU. (Cuando la CPU es "más cara").

El proceso de actualización de tablas en PDA (refresh table) tiene un coste de CPU que deben tenerse en cuenta. Sin embargo, este consumo se produce normalmente en períodos "batch". (Cuando la CPU es "más barata").

Con el fin de hacer una revisión más precisa del consumo de la CPU de la consulta con PDA, tendríamos que ponderar el consumo de CPU de la actualización de datos en el PDA en el consumo de consulta. Es obvio que cuantas más queries se ejecuten contra PDA, más se diluye el consumo asociado al refresco de las tablas.

Los ahorros de CPU son muy importantes, pero el ahorro de tiempo de respuesta es, sin duda, lo más interesante.



Conclusiones y nuevos retos

Los sistemas informacionales de Bankia están evolucionando para intentar proporcionar cada vez más información en un período de tiempo menor, esto es, acercando los “tiempos informacionales” a “tiempos transaccionales”.

El primer paso en esta evolución se está dando con el PDA e IBM DB2 Analytics Accelerator:

- Es fácil de instalar y manejar.
- Su implantación es atomizada, pudiendo priorizar el uso de la solución para aquellos casos en los que el beneficio sea mayor.
- Transparente para los usuarios, que no necesitan modificar su forma de trabajar.
- Permite ahorro de CPU en períodos críticos.
- Proporciona una reducción drástica del tiempo de respuesta a usuarios finales.



Nuevos Retos

- **Expandir** el uso de PDA al mayor número de aplicaciones informacionales posibles.
- **Eliminar** ciertos **procesos** batch muy **costosos** por ejecuciones de queries en PDA ahorrando tiempo de desarrollo, CPU y elapsed time.
- Acceso transaccional a datos del PDA.
- **Eliminar del zOS información histórica**, con el consiguiente ahorro de espacio en disco y disminuyendo la gestión de todos esos datos.





Gracias

