



**SOLID  
QUALITY  
MENTORS**

# Como Construir un DW Relacional

*Javier Loria*

*Mentor*

*[javier@solidq.com](mailto:javier@solidq.com)*

*Miguel Egea Gómez*

*Mentor*

*[megea@solidq.com](mailto:megea@solidq.com)*

# Agenda



## Definiciones

- Visiones de que es un DW
- BI

## Transaccional VS Analitica

- Consideraciones y preguntas

## Modelos

- Multidimensional
- Relacional

## Opciones de arquitectura

- Directa, ODS, DataWarehouse, Staging,...

# Definición de Datawarehouse

Ralph Kimball

- **DataWarehouse**

“ una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis ”

“la unión de todos los data marts de una entidad”

- **DataMart**

“Conjunto flexible de datos, idealmente basados en datos de máxima granularidad (a ser posible) obtenidos desde un origen operacional y presentados de una forma simétrica (Dimensional) resistente a queries de usuario inesperadas”

## Visión Ascendente (Botton-up)

# Definición de DataWarehouse

Bill Inmon

- **Orientado a temas.**- Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- **Variante en el tiempo.**- Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.
- **No volátil.**- La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de *sólo lectura*, y se mantiene para futuras consultas.
- **Integrado.**- La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes

## Visión Descendente o Top-Down

# Business Intelligence

Definición de Forrester :

Proceso de transformar datos en información y hacerla disponible a los usuarios en tiempo y forma para marcar la diferencia.

Es la evolución del data warehousing

- Centrado en entregar información al usuario
- Creado a nivel corporativo o departamental

# Transaccional Vs Analítica

## Preguntas transaccional



- ¿Cuándo se hizo este pedido?
- ¿Cuáles son las líneas de los pedidos de ayer?
- ¿Cual es el stock medio de chicles en toda la vida del producto?



## Preguntas analíticas

- ¿Qué factores favorecen la venta de un producto?
- ¿Cuánto contribuye en porcentaje cada familia de artículos a la venta total? ¿Cuánto ha cambiado comparado con el año pasado
- ¿Qué productos tienen menos margen? ¿son los mismos del año pasado?

# Transaccional VS analítico

- ¿y los usuarios que quieren?
  - Normalmente.. Lo quieren todo
  - Y normalmente lo quieren para ayer
  - Y si quieres saber cada cuanto se actualiza, lo normal es que te digan "en cuanto entre un dato"
  - Pero si les preguntas ¿guardamos histórico? También te dirán que sí...

# Transaccional VS analítico



## Propiedades transaccional

- Datos normalizados
- Escrituras rápidas
- Preparado para operaciones
  - Dar de alta un pedido ha de ser rápido
  - Sacar un listado puede llevarse su tiempo
- Los datos están en tiempo real, tras un commit.



## Propiedades analíticas

- Datos desnormalizados
- Formatos y nombres cercanos al negocio
- Lecturas han de ser rápidas
  - Normalmente no hay escrituras
  - Los listados deben responder muy rápido
- Podrían aceptarse 24 horas de retardo como tiempo real

# En resumen

- Todo es BI
  - No hay verdades o paradigmas que valgan para todas las circunstancias
  - Un datawarehouse bien montado tiene ventajas
    - Pero he visto como no era entendido por usuarios
    - Y ha generado desconfianza (los datos no cuadran con el relacional)
  - Un modelo operacional tiene inconvenientes
    - Pero es más fácil de comprobar
    - Puede usarse en tiempo semi-real
    - No podremos hacer JAMAS una buena integración

# Con datawarehouse

- Utilizar técnicas de modelado multidimensional
- Diseñar interfaces, conectividades de forma que impacten lo mínimo posible en los orígenes de datos
  - Evitar joins complejos
  - Evitar leer todos los datos cuando sea posible
  - Usar Áreas de staging para prevenimos de errores con
    - Páginas de códigos
    - Tipos de datos
    - Ordenaciones
    - .....

# Modelado multidimensional

- Lo que se usa ahora
  - Mezcla de normalización y desnormalización, llamada "Normalización dimensional"
  - A veces para diseñar datawarehouses completos
  - Otras para datamarts
- Fundamentalmente dos tipos de tablas
  - Tablas de dimensión
  - Tablas de hechos

# Modelado multidimensional

- Tablas de dimensión
  - Tienen atributos relacionados con entidades de negocio, como clientes, proveedores
  - Fechas, a veces horas
- Suelen incluir claves subrogadas
  - No son ni las de negocio, ni las alternativas
  - También suelen incluir las de negocio
  - Periodos de vigencia
- No es extraño que sean tablas muy anchas, con muchas columnas

# Modelado multidimensional

- Tablas de hechos
  - Contienen medidas, números que suelen ser acumulados después
  - También contienen las claves de dimensión
  - A veces incluyen información adicional y pueden ser usadas como dimensión jugando el rol de dimensión de hechos
  - Las hay de poquitas columnas pero también de muchas, casi todas tienen muchísimas filas.
  - Si queremos un versionado completo hemos de añadir la fecha de modificación de la fila.

# Demostración

Usar SSIS para crear un DW



# Modelado Relacional

- Generalmente en 3FN
- Muchas tablas, optimización de espacio y transacciones
- Tablas poco orientadas a negocio, orientadas a operación
- Preparadas para escrituras rápidas y accesos por clave.

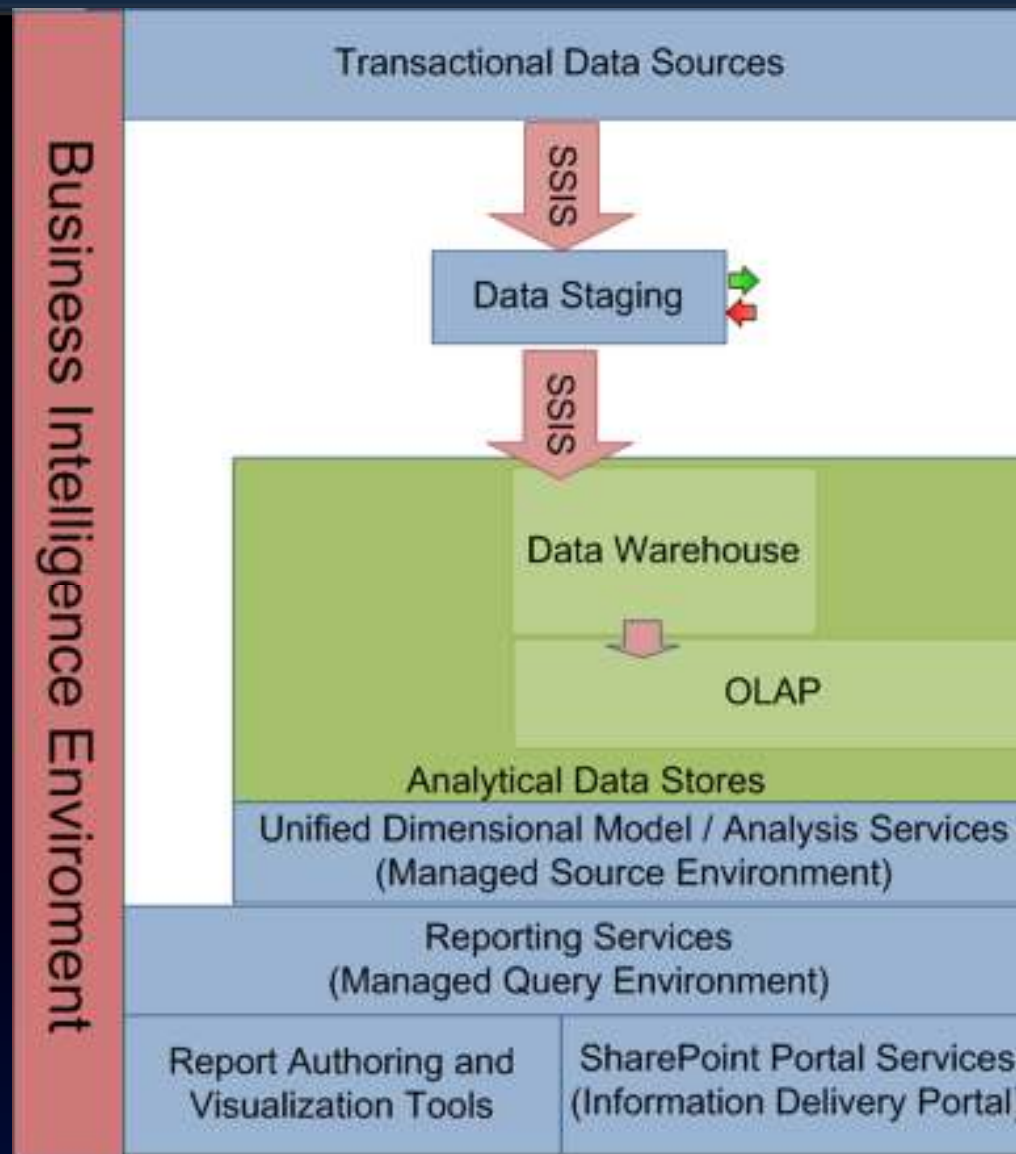
# Modelado relacional

- Problemas comunes
  - Falta de integridad referencial declarativa
    - "eso lo hace la aplicación"
  - Resultado
    - Registros huérfanos, inconsistencias de datos
    - Dependiendo de la consulta, salen unos datos u otros
  - Acciones
    - No tenemos limpieza e integración de datos, así que debemos establecer criterios

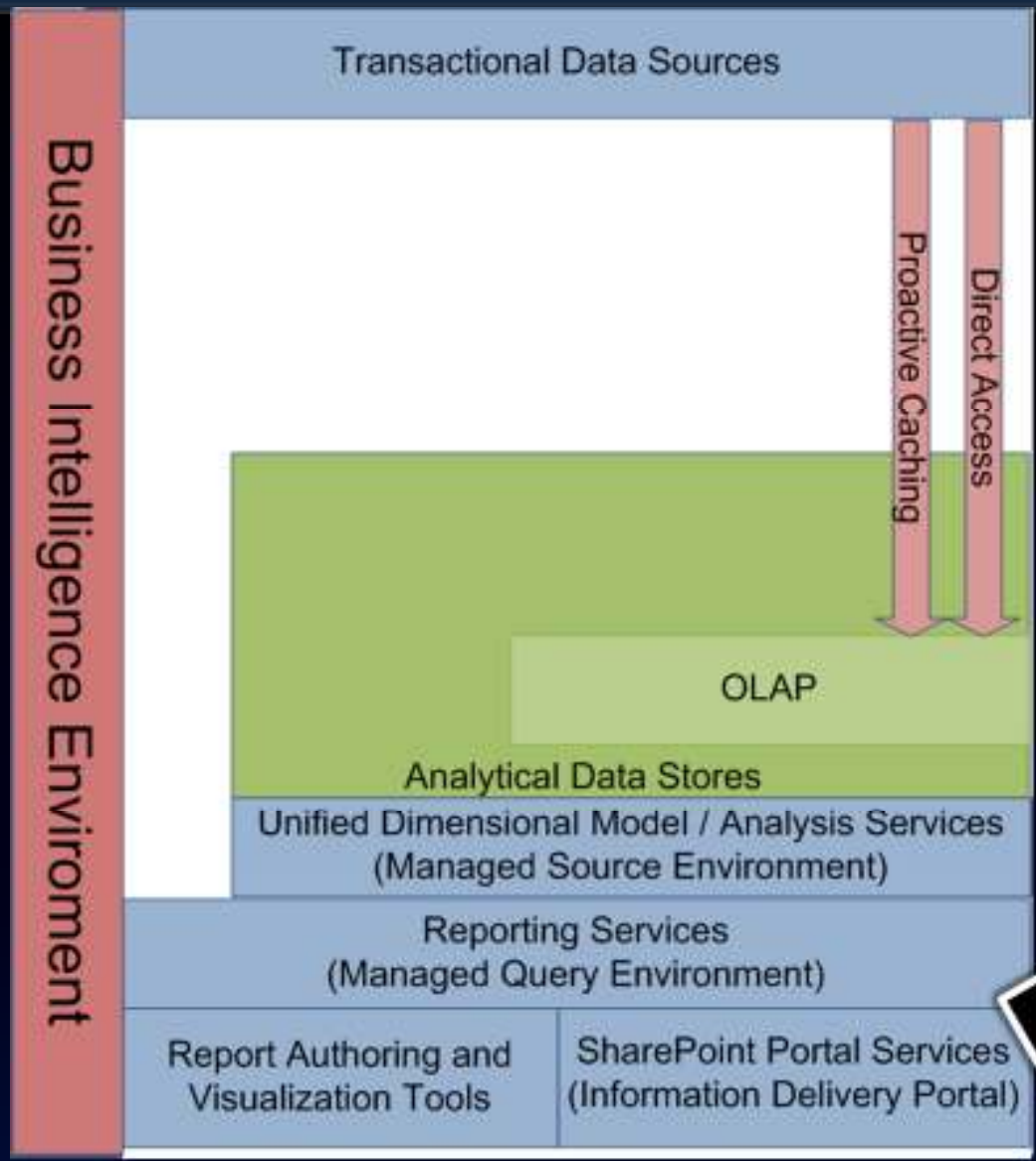
# Opciones de Arquitectura

- Arquitectura Tradicional
- Acceso Directo
- Empuje de Integración
- ODS
- Datamarts/DW
- Eje y Rayo (DW/Datamarts)

# Aquitectura "Tradicional" de BI



# Acceso Directo



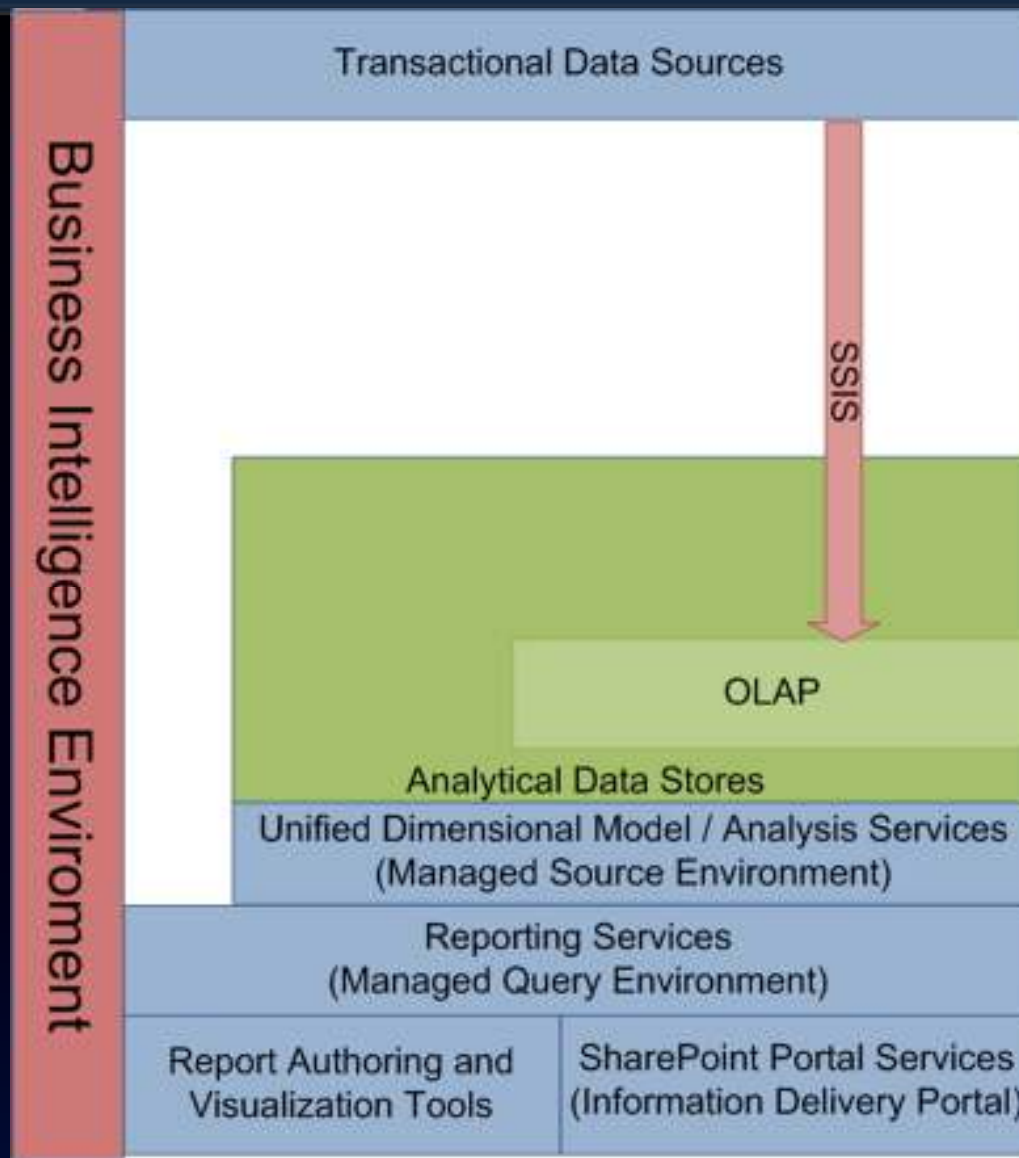
**Look Ma,  
No Hands!!**

# Demostración

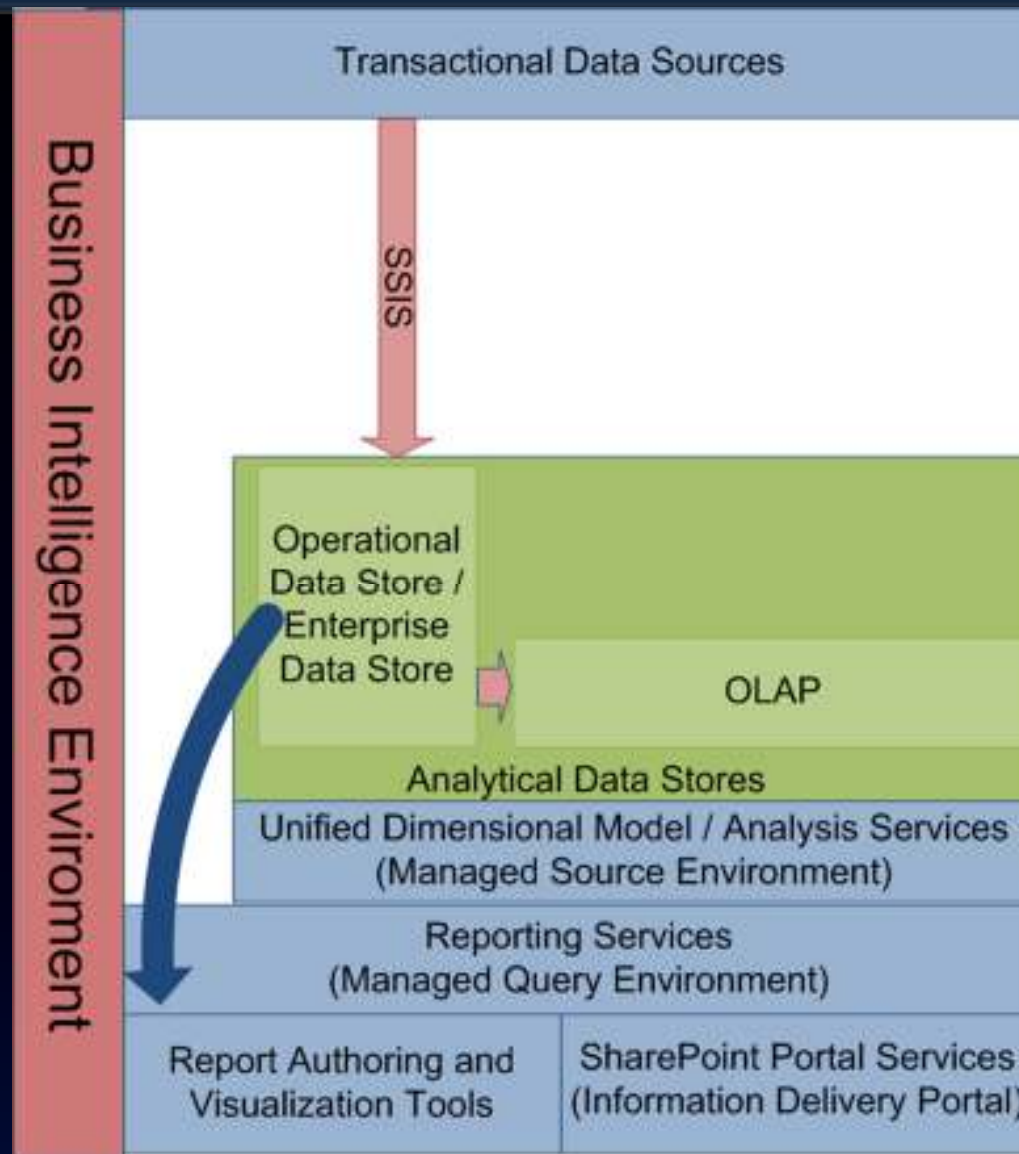
Creación de instantáneas y sinónimos automáticas



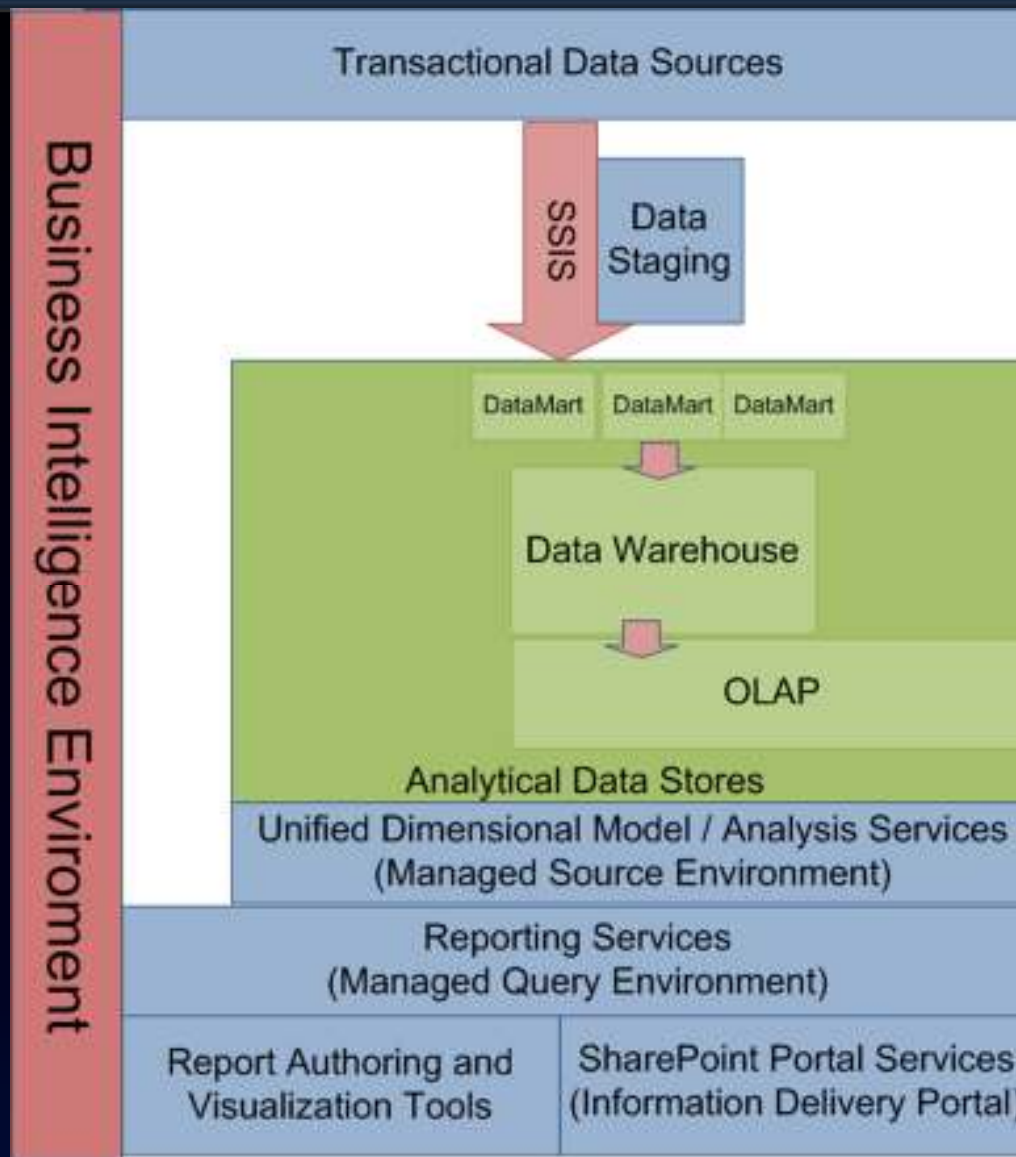
# Empuje de Integración



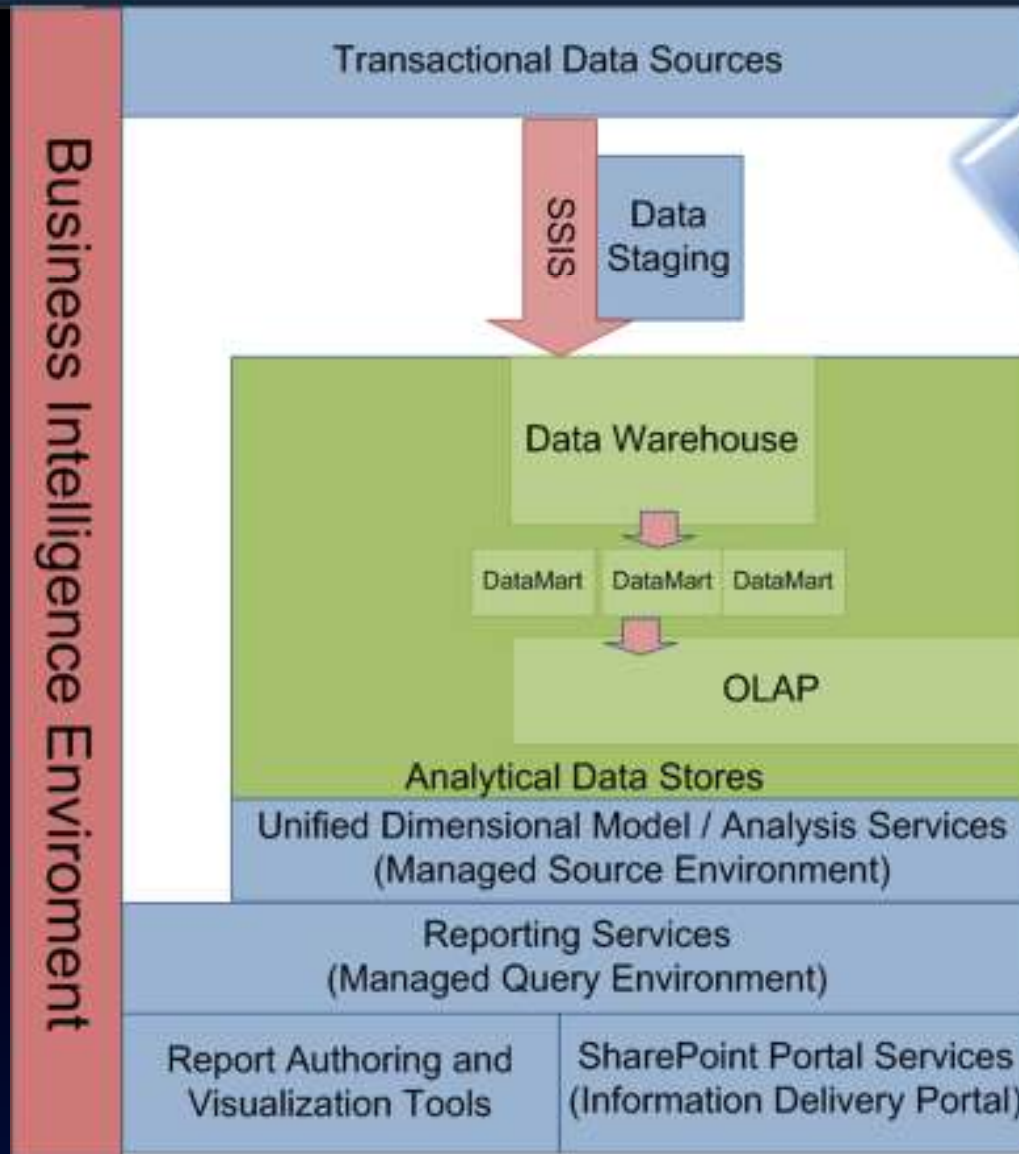
# ODS/EDS



# Data Marts-DataWarehouse Centralizado EDW

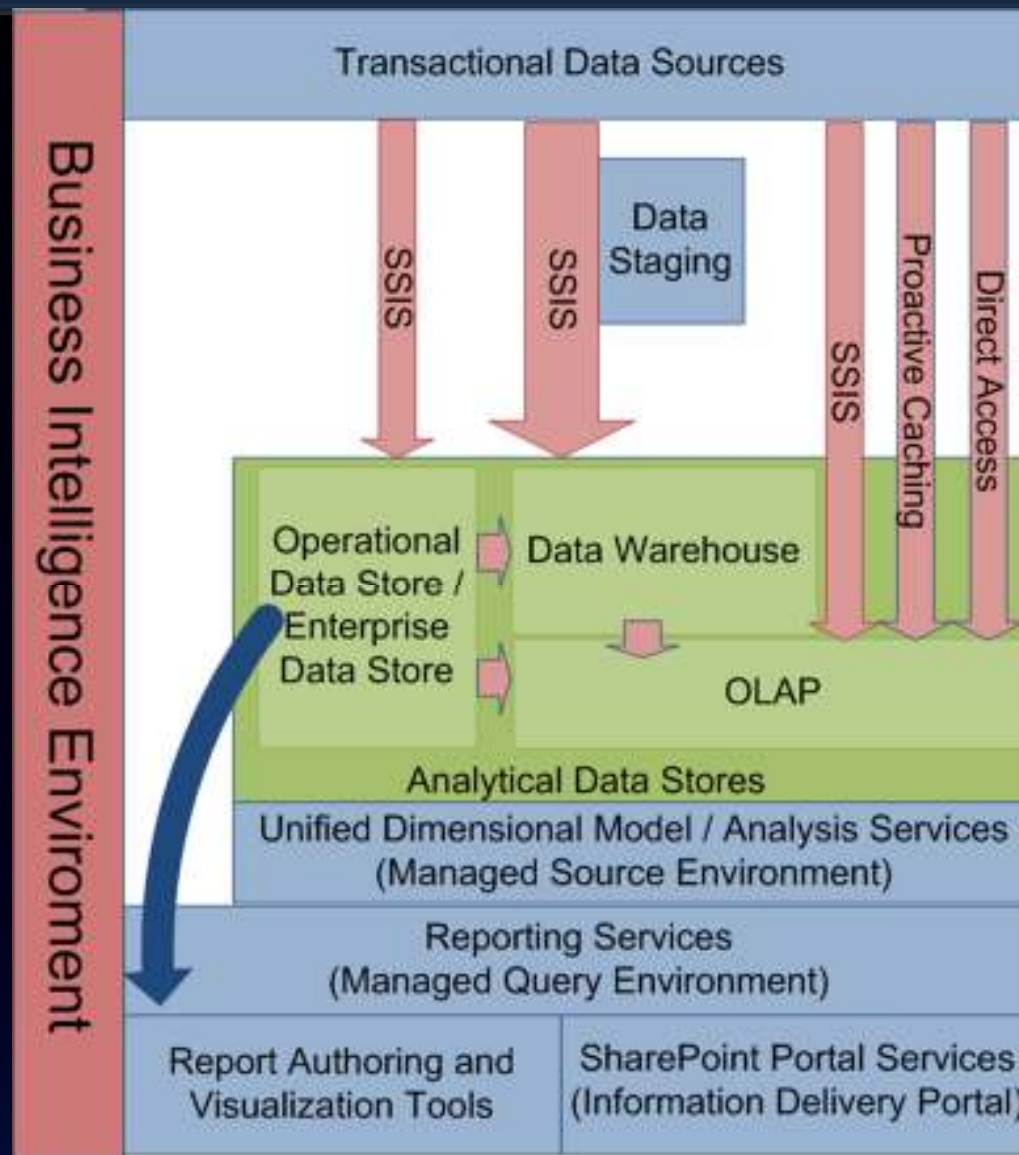


# Eje-y-Rayo (Hub-and-Spoke)



**PDW**

# Combo: Arquitectura Pragmática



# Demostración

Triggers-CDC- y alternativas para mantener históricos



# Link del CAT

- <http://sqlcat.com/msdnmirror/archive/2008/08/05/microsoft-sql-server-database-snapshots-and-synonyms.aspx>



SOLID  
QUALITY  
MENTORS

Como construir un DW Relacional

Gracias!!!