

Seminario:

*Construyendo una
Solución de BI paso a paso
con SQL Server 2005*

Ing. José Mariano Álvarez
Jose.Mariano.alvarez@sqltotalconsulting.com



La Base de Datos OLAP Analysis Services (SSIS) Avanzado



Agenda

- Proactive Caching
- Particionamiento
- Calculated Members
- Key Performance Indicators
- Actions
- Data Mining

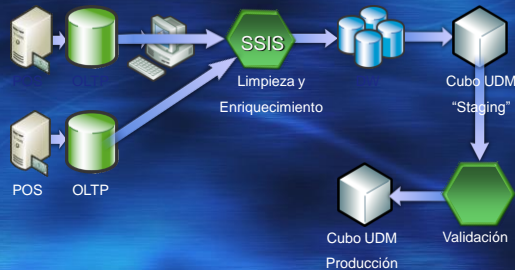


Agenda

- Proactive Caching
- Particionamiento
- Calculated Members
- Key Performance Indicators
- Actions
- Data Mining



Retos de BI en tiempo real



"Empujando" datos hacia el UDM



- Los datos pueden empujarse directamente hacia el UDM
- SQL Server 2005 Integration Services procesa transformaciones
- Incluye tanto a las tablas de hechos como de dimensiones



Actualización con "Trickle Feeds"



- "Trickle feeds" pueden obtener datos directamente dentro del UDM
- Integration Services actualiza el cubo cada pocos minutos

SQL TOTAL CONSULTING

Construyendo el Cubo Directamente



- UDM puede combinar datos desde múltiples fuentes
- Una de las fuentes involucradas debe ser SQL Server
- No se aplica para todos los escenarios

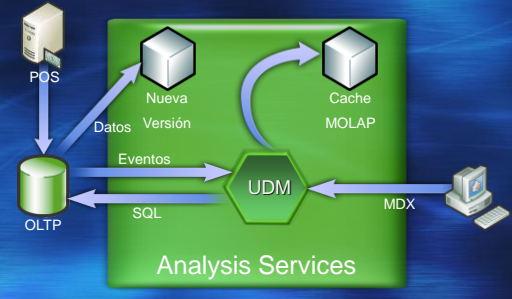
SQL TOTAL CONSULTING

Proactive Caching

- Gestión basada en políticas
 - ¿Ha cambiado la fuente de datos?
 - ¿Cuándo se refresca?
 - Como responder peticiones durante actualización
- Proactive caching combina
 - Desempeño de queries OLAP
 - Acceso a datos en tiempo real "as needed"
- No hay mas procesamiento explícito del cubo

SQL TOTAL CONSULTING

Proactive Caching – Ejemplo



SQL TOTAL CONSULTING

Usando políticas para refrescar el Cache



- Actualización del cubo basado en políticas

SQL TOTAL CONSULTING

Retos del "Proactive Caching"

- Eficiencia
 - Cómo evitar sobrecargar Analysis Services con actualizaciones frecuentes
 - ¿Qué tan rápido pueden estar disponibles los caches?
- Desempeño
 - Cómo equilibrar latencia y desempeño
- Notificaciones
 - ¿El cache se actualiza con los cambios o periódicamente?
 - ¿Cómo sabe AS que el RDBMS ha cambiado?

SQL TOTAL CONSULTING

Configuraciones de políticas

Propiedad	Descripción

SQL TOTAL CONSULTING

Escalando hacia arriba

Problema

Cómo manejar grandes cantidades de datos

Re-creando el cache en su totalidad por cada cambio es costoso

Solución

Usar ROLAP

Usar particiones

Usar actualizaciones incrementales del cache para agregar datos

SQL TOTAL CONSULTING

Agenda

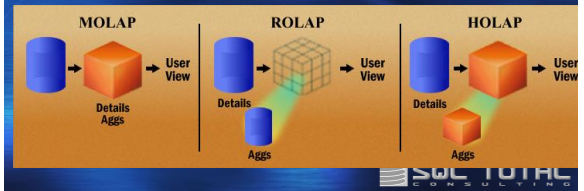
- Proactive Caching
- Particionamiento
- Calculated Members
- Key Performance Indicators
- Actions
- Data Mining

SQL TOTAL CONSULTING

Diseño Servicio de Análisis

Almacenamiento Flexible

- Debate entre los proveedores MOLAP y ROLAP confunden las necesidades de los usuarios
- Servicios de Análisis soportan MOLAP, ROLAP y HOLAP y ofrecen integración en las tres
- Usuarios & aplicaciones trabajan solo cubos



Modos de Almacenamiento

Almacenamiento	MOLAP	HOLAP	ROLAP
Datos base	Cubo	Tablas Relacionales	Tablas Relacionales
Agregaciones	Cubo	Cubo	Tablas Relacionales

Perspectiva del cliente	MOLAP	HOLAP	ROLAP
Rendimiento de las consultas	El más rápido	Más Rápido	Rápido
Consumo de almacenamiento	Alto	Medio	Bajo
Mantenimiento del Cubo	Alto	Medio	Bajo

SQL TOTAL CONSULTING

Por qué particionar datos relacionales?



Administración



Performance

SQL TOTAL CONSULTING

Guías para particionar datos relacionales

- ✓ Asignar un apropiado partitioning key y partitioning interval
- ✓ Particionar fact tables
- ✓ Usar SWITCH, MERGE, y SPLIT
- ✓ Usar aligned partitioned indexes
- ✓ Share partition functions y schemas

SQL TOTAL
CONSULTING

Porque particionar datos multidimensionales ?



Minimiza tiempo de procesamiento del cubo



Proporciona procesamiento paralelo



Mejora performance de las consultas

SQL TOTAL
CONSULTING

Arquitectura de Particionamiento en Analysis Services

- El Particionamiento ocurre a nivel de measure group
- Las Particiones son procesadas en paralelo para un measure group
- Las Particiones pueden ser procesadas selectivamente
- Los Data slices son opcionales para MOLAP
- Los Data slices son requeridos para ROLAP

SQL TOTAL
CONSULTING

Guías para particionar datos multidimensionales

- ✓ Determinar partition boundaries
- ✓ Crear un template table en el DSV
- ✓ Determinar mode de almacenamiento y nivel de agregamiento
- ✓ Definir plan de sincronizacion de datos

SQL TOTAL
CONSULTING

Agenda

- Proactive Caching
- Particionamiento
- Calculated Members
- Key Performance Indicators
- Actions
- Data Mining

SQL TOTAL
CONSULTING

MDX



Lenguaje de Consulta para datos multidimensionales

```
[Measures].[Reseller Sales-Unit Price]
*
[Measures].[Reseller Sales-Order Quantity]
```

SQL TOTAL
CONSULTING

Calculated Member



Miembro que es calculado a tiempo de ejecución usando una expresión MDX

```
[Measures].[Sales Value]-[Measures].[Cost]
```

SQL TOTAL CONSULTING

Uso de Calculated Member



Cuando un valor es no aditivo



Cuando un valor no está en una dimensión o en un measure group



Cuando se requiere un valor que no se usa tan frecuentemente

SQL TOTAL CONSULTING

Named Set



Una expresión que define un grupo de miembros de una dimensión



Se crean usando expresiones MDX

```
[Sales Territory].[Sales Territory Group].&[Pacific]
```

SQL TOTAL CONSULTING

Uso de Named Set



Si se necesita referenciar a una expresión MDX que es demasiado compleja



Al Encapsular la lógica de negocio



Como un subset de un cubo

SQL TOTAL CONSULTING

MDX Script



Un cálculo aplicado a un definido subcubo

```
SCOPE([Measures].[Reseller Sales-Sales Amount]  
,[Product].[Category].&[1]);  
THIS = [Product].[Category].currentmember.parent/2;
```

SQL TOTAL CONSULTING

Uso de MDX Scripts



Cuando se necesita ejecutar cálculos sobre un subset del cubo



Cuando no se desea afectar los datos relacionales originales

SQL TOTAL CONSULTING

Agenda

- Proactive Caching
- Particionamiento
- Calculated Members
- Key Performance Indicators
- Actions
- Data Mining

KPIs

- Indicadores del Negocio
- Definiciones basadas en el Servidor
- Agrupados en scorecards

Las Ventas
han sido
alcanzadas



El nivel de
ganancia
esta debajo
del target



Consideraciones para implementar KPIs

Soporte de Aplicacion Cliente

Definicion de KPI

- Value
- Goal
- Status
- Indicators
- Trend
- Weight

Actions

Comandos MDX Server-stored:

- URL
- Statement
- Rowset
- Proprietary
- Dataset
- Drillthrough
- Reporting
- Commandline
- HTML

Consideraciones para implementar Actions

	Performance
	Seguridad
	Soporte de App Cliente
	Target
	Formato del Reporte

Agenda





- Proactive Caching
- Particionamiento
- Calculated Members
- Key Performance Indicators
- Actions
- Data Mining


Tipos de Análisis

- Query Reporting Analysis**
 Qué ha sucedido en el pasado
 Reportes simples
 Key Performance Indicators
 Cubos OLAP – slice/dice
- Tiempo Real**
 Qué está sucediendo
 Eventos/triggers
- Minería de Datos**
 Predicciones para el futuro
 Cómo y por qué
 Razones y comportamientos



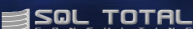
Consideraciones para usar una solución de DM

- 
Tareas que se necesitan ejecutar
- 
Columnas que se deben incluir en el case table
- 
Algoritmo mas apropiado
- 
Número de usuarios que accederan a los resultados










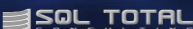
Guía para preparar los datos

- Identificar las columnas apropiadas para la case table
- Determinar si se requiere una nested table
- Identificar key columns, input columns y predictable columns
- Determinar column types
- Identificar el fuente de datos para la training data



Data Mining en SQL 2005

- 
Decision Trees
- 
Clustering
- 
Time Series
- 
Sequence Clustering
- 
Association
- 
Naïve Bayes
- 
Neural Net



Tareas en Data Mining

- Clasificación
- Regresión
- Segmentación
- Asociación
- Pronósticos
- Análisis de Texto

Preguntas típicas de Negocio

- ¿Que tipo de tarjeta de membresía debería ofrecer?
- ¿Que clientes responderán a mi correo directo?
- ¿Es esta transacción fraudulenta?
- ¿Perderé a este cliente?
- ¿Será este producto defectuoso?

Algoritmos

- Preferidos
- Árboles de decisión
- Naïve Bayes
- Redes Neuronales

Además

- Clustering
- Sequence Clustering
- Reglas de Asociación



Tareas en Data Mining

- Clasificación
- Regresión
- Segmentación
- Asociación
- Pronósticos
- Análisis de Texto

Preguntas Típicas de Negocio

- ¿Cuántos Ingresos obtendré de este cliente?
- ¿Cuanto tiempo estará este activo en servicio?

Algoritmos

- Preferidos
- Árboles de decisión
- Redes Neuronales

Además

- Clustering
- Sequence Clustering



Tareas en Data Mining

- Clasificación
- Regresión
- Segmentación
- Asociación
- Pronósticos
- Análisis de Texto

- Preguntas Típicas de Negocio
 - + Describo mi Cliente
 - + ¿Cómo puedo diferenciar mis clientes?
 - + ¿Cómo puedo organizar mis datos de tal forma que tenga sentido?
- Algoritmos
 - + Preferidos
 - + Clustering
 - + Sequence Clustering
- Además
 - + Redes Neuronales



Tareas en Data Mining

- Clasificación
- Regresión
- Segmentación
- Asociación
- Pronósticos
- Análisis de Texto

- Preguntas Típicas de Negocio
 - + Análisis de Canasta de mercado/Venta Cruzada
 - + ¿Qué items se compran juntos?
 - + ¿Qué productos debería recomendar a mi cliente?
- Algoritmos
 - + Preferidos
 - + Reglas de Asociación
 - + Árboles de Decisión (Catálogos pasatiempos)
- Además
 - + Clustering, Sequence Clustering, Naive Bayes, Redes Neuronales



Tareas en Data Mining

- Clasificación
- Regresión
- Segmentación
- Asociación
- Pronósticos
- Análisis de Texto

- Preguntas Típicas de Negocio
 - +Cuál es el ingreso proyectado para todos los productos?
 - + ¿Cuáles serán los niveles de inventario para el próximo mes?
- Algoritmos
 - + Series de Tiempo



Tareas en Data Mining

- Clasificación
- Regresión
- Segmentación
- Asociación
- Pronósticos
- Análisis de Texto

- Análisis de Datos no-estructurados
 - + Extracción de términos y/o frases claves
 - + Conversión a datos estructurados
 - + Alimentar otros algoritmos
 - + Clasificación
 - + Segmentación
 - + Asociación
- Preguntas Típicas de Negocio
 - + ¿Cómo manejar los datos del call center?
 - + ¿Cómo clasificar el correo?
 - + ¿Qué puedo hacer con la retroalimentación del Web?





Consideraciones para escoger los algoritmos de DM


	Classification	Regression	Segmentation	Association	Sequence Analysis
Association				✓	
Clustering	✓		✓		
Decision Trees	✓			✓	
Naive Bayes	✓				
Neural Network	✓				
Sequence Clustering	✓		✓		✓
Time Series		✓			
Linear Regression		✓			
Logistic Regression		✓			



Minería de Datos con OLAP

- 

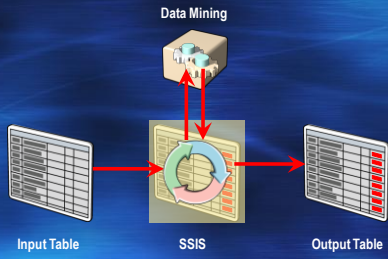
Uso de un cubo como un data source
- 

Uso de dimensiones o measures groups
- 

Se genera nuevas dimensiones en base a los resultados del DM

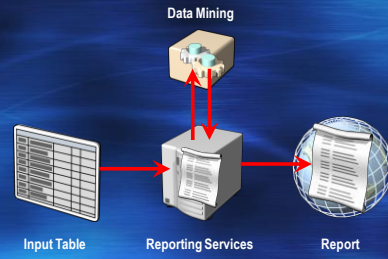


Minería de Datos con SSIS



SQL TOTAL
CONSULTING

Minería de Datos con SSRS



SQL TOTAL
CONSULTING

Microsoft®

© 2005 Microsoft Corporation. All rights reserved.
This presentation is for informational purposes only. Microsoft makes no warranties, expressed or implied, in this presentation.

SQL TOTAL
CONSULTING