

Seminario:

Construyendo una solución de BI paso a paso con SQL Server 2005

Ing. José Mariano Alvarez
Jose.Mariano.Alvarez@SqlTotalConsulting.com



El Sistema ETL Integration Services (SSIS)



AGENDA

- CONSIDERACIONES DE DISEÑO PARA EL SISTEMA ETL
- EL SQL SERVER BI TOOL SET PARA EL SISTEMA ETL
- CONSIDERACIONES DE DESARROLLO PARA EL SISTEMA ETL



CONSIDERACIONES DE DISEÑO

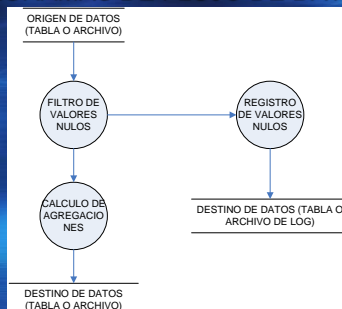
- Un plan de alto nivel
- Diagramas de Flujo de Datos
- Perfilar los Datos
- Mapas Origen – Destino
- Frecuencia de Carga
- Cuanta Historia Debo Almacenar?
- Uso de Particiones
- Carga Histórica e Incremental
- Diseño Físico del Sistema ETL
- Sistema de Auditoría



PLANES DE ALTO NIVEL

DISEÑO

- DIAGRAMAS DE FLUJO DE DATOS



PERFILAR LOS DATOS

DISEÑO

- ENTENDER COMPLETAMENTE LOS ORIGENES DE DATOS
- DETECTAR "ERRORES" DE CAPTURA DE DATOS EN LOS SISTEMAS FUENTE



COMPLETAR EL MAPA: ORIGEN - DESTINO

DISEÑO

- SIMPLE
- CON TRANSFORMACIONES

SQL TOTAL
CONSULTING

FRECUENCIA DE CARGA

DISEÑO

- DIARIA?
- SEMANAL?
- MENSUAL?

SQL TOTAL
CONSULTING

CUANTA HISTORIA DEBO ALMACENAR?

- FRECUENCIA DE PODA DEL DWH
- COMO SE CARGARA LA INFORMACION HISTORICA PARA EL LANZAMIENTO DEL DW
- DEBE SER UNA DECISION DE LOS CONSUMIDORES
 - BASADA EN NECESIDADES DE NEGOCIO
 - EL EQUIPO DE BI SOLO SUGIERE

SQL TOTAL
CONSULTING

USO DE PARTICIONES

- TABLAS DE HECHOS GIGANTES
- DB RELACIONAL
- DB MULTIDIMENSIONAL
- PREPARE LAS NUEVAS PARTICIONES ANTES DE LA CARGA

SQL TOTAL
CONSULTING

CARGAS HISTORICAS E INCREMENTALES

- ESTRATEGIAS PARA EXTRACCION DE DATOS
 - DESDE SISTEMAS EMPAQUETADOS
 - DESDE LAS BD ORIGEN
 - PARA CARGAS INCREMENTALES
 - PARA CARGAS HISTORICAS
- ACTUALIZACION DE LA BASE DE DATOS MULTIDIMENSIONAL

SQL TOTAL
CONSULTING

Diseño Físico del Sistema ETL

- ARQUITECTURA DE SSIS
- STAGING AREA
- ALMACENAMIENTO DE PAQUETES
- CONVENCION EN NOMBRES DE PAQUETES

SQL TOTAL
CONSULTING

CONSTRUCCION DE PAQUETES

DESARROLLO

- PLANTILLA
- PAQUETES PADRE Y PAQUETES HIJOS

SQL TOTAL
CONSULTING

AGENDA

- ▣ CONSIDERACIONES DE DISEÑO PARA EL SISTEMA ETL
- EL SQL SERVER BI TOOL SET PARA EL SISTEMA ETL
- CONSIDERACIONES DE DESARROLLO PARA EL SISTEMA ETL

SQL TOTAL
CONSULTING

SQL SERVER BI TOOL SET

- De DTS a SSIS
- Integration Services
 - Qué son?
 - SSIS Tools?
 - Data Flow y Control Flow
 - Tareas Disponibles

SQL TOTAL
CONSULTING

SQL Server 2005 BI



SQL TOTAL
CONSULTING

De DTS a SSIS

- DTS - SQL Server 7.0
 - “Visual BCP” – Muy útil
- DTS - SQL Server 2000
 - Fácil workflow & transform engine
- SSIS – SQL Server 2005
 - ETL de clase empresarial
 - Integración Excepcional de BI
 - Riqueza y extensibilidad de APIs

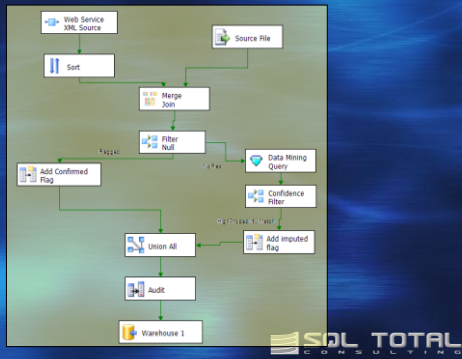
SQL TOTAL
CONSULTING

¿Qué es SQL Server Integration Services?

- Una nueva aplicación de Inteligencia de Negocios
- Sucesor de DTS
- La plataforma para una nueva generación de tecnologías de integradores de datos de alta performance

SQL TOTAL
CONSULTING

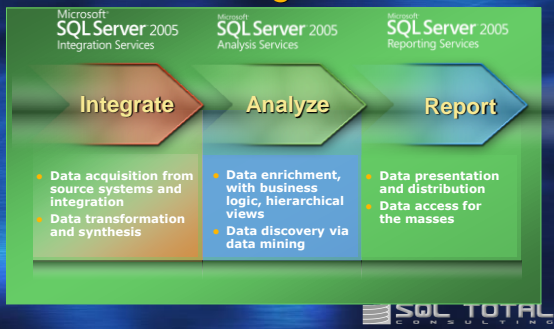
Qué es SQL Server Integration Services? (cont)



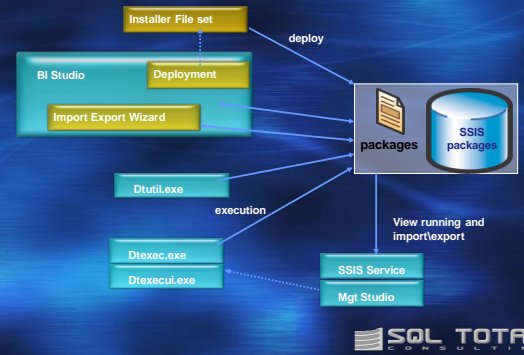
Escenarios de Usos de SSIS

- Mezclar datos de orígenes de datos heterogéneos
- Limpiar y normalizar datos
- Generar Business Intelligence en un proceso de transformación de datos
- Automatizar las funciones administrativas y la carga de datos

SQL Server Business Intelligence

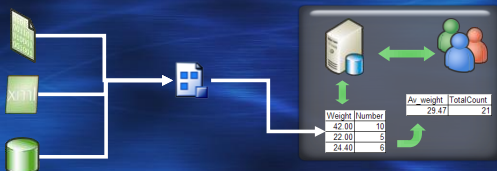


SSIS Tools



Antes de Integration Services...

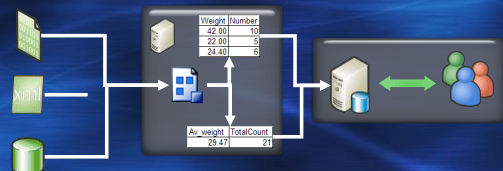
Carga tradicional de un warehouse



- En este escenario tradicional, el proceso de integración carga los datos al servidor de base de datos
- La base de datos ejecuta agregaciones, ordenamiento y otras operaciones
- ...pero tiene contención por demanda por recursos de las consultas de usuarios
- Esta solución no escala para grandes volúmenes de datos y múltiples y complejas agregaciones

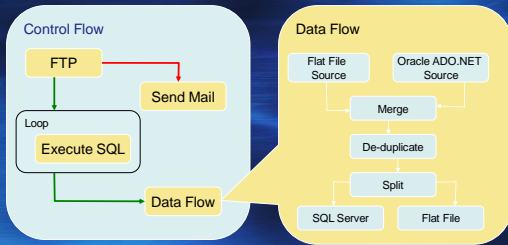
Con Integration Services

Warehouse cargado con SQL Server Integration Services



- Aquí, SQL Server Integration Services trabaja los datos antes...
- ...y también realiza las agregaciones y el ordenamiento y luego carga los datos al servidor de base de datos
- Esto libera al servidor de base de datos para las consultas de los usuarios
- Con 64-bit esta solución escala bien para grandes volúmenes de datos y múltiples y complejas agregaciones
- Con 32-bit, esta arquitectura puede escalar mejor usando un servidor para los procesos de integración

Cómo trabaja esto?



Desde allí pueden ser validados y consistenciados los workflows de tareas.

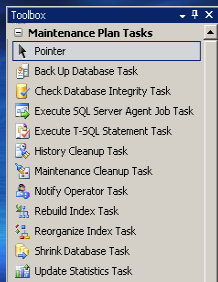


Control Flow Task Groups

- Maintenance Plan Tasks
- Data Preparation Tasks
- Scripting Tasks
- Workflow Tasks
- SQL Server Tasks
- Control Containers



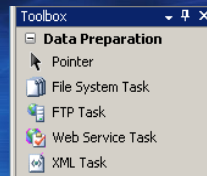
Maintenance Plan Tasks



- Se crea los planes de mantenimiento con mayor control
- Inicia ejecución de SQL Agent Jobs
- Reconstruye o reorganiza (DBCC IndexDefrag) Indices



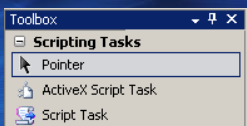
Data Preparation Tasks



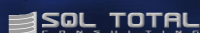
- Crea y administra folders y archivos
- Carga y descarga archivos desde un FTP site
- Comunicación con Web Services para enviar datos a un archivo o variable
- Modificación dinámica de documentos XML
 - Aplica XSLT style sheets
 - Ejecuta expresiones XQuery
 - Combina documentos XML



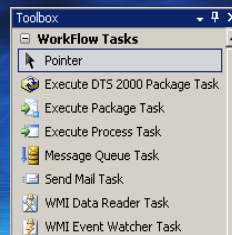
Scripting Tasks



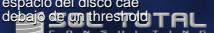
- La tarea de ActiveX Script permite ejecutar DTS 2000 scripts
 - Ejecuta las mismas tareas que en DTS 2000
 - Si el script accesa el modelo de objetos de DTS, ellos deberán ser actualizados
- Script Task reemplaza al ActiveX Script Task
 - Usa Visual Studio for Applications
 - Desarrolla scripts como funciones Visual Basic.NET
 - Es compilado



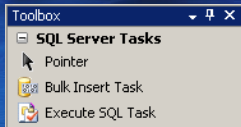
Work Flow Tasks



- Ejecuta paquetes DTS 2000 para compatibilidad y migración
- Ejecuta otros paquetes SSIS
- Las consultas WMI pueden ser enviadas a un archivo o a una variable
 - Lee Windows Event Log
 - Consulta de estado/propiedades de dispositivos
 - Determina que versiones y aplicaciones están instaladas
- Monitorea WMI Events seleccionados
 - Espera por archivos para mostrarlos en un folder
 - Elimina archivos cuando el espacio del disco cae debajo de un threshold



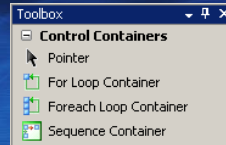
SQL Server Tasks



- Bulk Insert
 - Igual que en DTS 2000
- Execute SQL Task
 - Igual que en DTS 2000
- Excepto
 - Sentencias SQL pueden ser leídas desde un archivo o variable
 - Pueden generar documentos XML



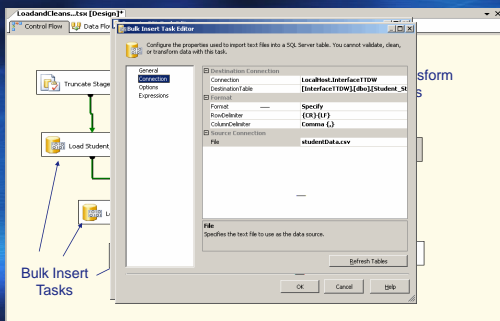
Control Containers



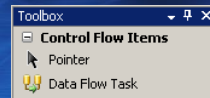
- FOR Loop Container implementa un bucle para una o mas tareas
- FOREACH Loop container realiza un bucle sobre un enumerador
 - Archivos en un Folder
 - Filas en un ADO Rowset
 - Static List
 - ADO Objects
 - SMO Objects
 - Variable
- Sequence Container permite definir un subconjunto de tareas
 - Puede ser deshabilitado
 - Puede definir una transacción



Control Flow Diagram



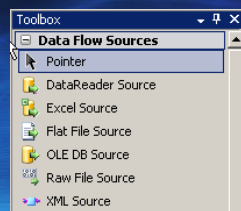
Data Flow Task



- Reemplaza a la DTS 2000 Transform y Data-Driven Query Tasks
- Incluye:
 - 6 Origenes de flujo de datos
 - 11 Destinos de flujo de datos
 - 28 Transformaciones de datos



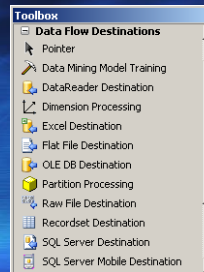
Data Flow Sources



- DataReader configura .NET data provider
- OLE DB configura un OLEDB data provider
- Raw File lee datos creados desde un Raw File destination
 - Mas rapido que un Flat File u Origenes de datos OLE DB
- XML Source lee documentos XML desde un archivo o una variable
 - Puede aplicar un XSD schema



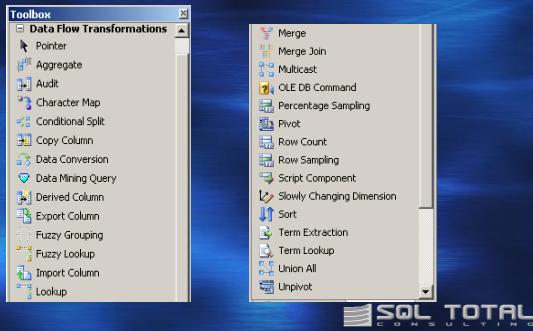
Data Flow Destinations



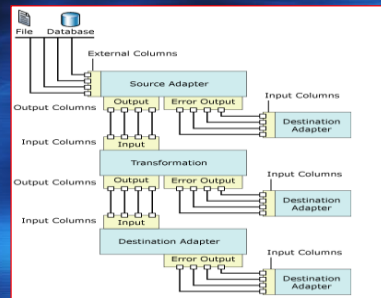
- Data Reader puede pasar salidas de transformaciones a otras aplicaciones
- Raw File Destinations para almacenar archivos intermedios usados en subsecuentes paquetes SSIS
- RecordSet is un in-memory disconnected ADO recordset
 - Consumidos por Scripts en otra partes del paquete
- SQL Server Destination toma salidas de previos pasos e inicia un Bulk Insert en tablas SQL Server
- SQL Server Mobile envia datos a base de datos SQL Server CE en dispositivos portables



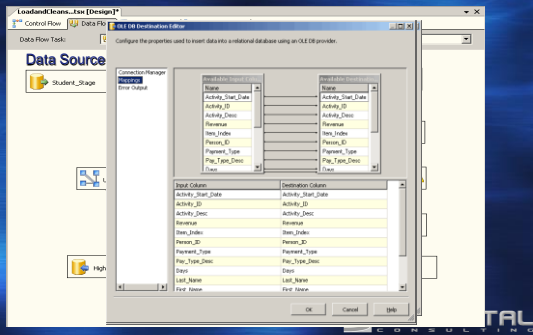
Data Transformations



Data Flow Process



Data Flow Diagram



Connection Managers

- OLEDB
- ODBC
- ADO
- ADO.NET
- Analysis Services
- File
- Flat File
- FTP
- HTTP
- WMI

Reliability y Recovery

- Flujo de errores
 - Manejo de problemas en datos sin hacer fallar el paquete
- Manejo de eventos
 - Responde a eventos en el paquete durante la ejecución
- Integración WMI
 - Realiza consultas de los registros de eventos de Windows
- Restart de paquetes
 - Checkpoints por fallas
- Firma digital
 - Asegura integridad con certificados

Logging y Monitoring

- Diferentes proveedores de registro
 - Archivo de texto (.log)
 - SQL Profiler (.trc)
 - SQL Server
 - Registro de sucesos de Windows
 - Archivo XML (.xml)
- Monitoreo de la Performance
 - SSIS incluye un conjunto de contadores de rendimiento para supervisar el rendimiento del motor de flujo de datos

Limpieza de Datos

- Microsoft research
 - SSIS incluye la "lógica difusa" de última tecnología basada en componentes de limpieza de datos
 - El enfoque elegido es el de dominio independiente y no depende de ningún dato de dominio específico, como los datos de referencia de dirección o código postal.
- Fuzzy lookup
 - Busca equivalencias aproximadas
 - Du Pont = Dupont
 - Proporciona similitud y las mejores equivalencias
- De-duplication
 - Eliminación de duplicados
 - "Windows XP", "WinXP", etc.



AGENDA

- ▣ CONSIDERACIONES DE DISEÑO PARA EL SISTEMA ETL
- ▣ EL SQL SERVER BI TOOL SET PARA EL SISTEMA ETL
- CONSIDERACIONES DE DESARROLLO PARA EL SISTEMA ETL



CONSIDERACIONES PARA DESARROLLO

- Construcción de Paquetes
- Procesamiento de Dimensiones
- Procesamiento de Hechos
- Sistema de Auditoría



CONSTRUCCION DE PAQUETES

DESARROLLO

- PLANTILLA
- PAQUETES PADRE Y PAQUETES HIJOS



PROCESAMIENTO DE DIMENSIONES

DESARROLLO

- TRANSFORMACIONES
- CARGA DE DIMENSIONES
 - FILAS QUE HAN CAMBIADO
 - DIMENSIONES LENTAMENTE CAMBIANTES (SCD)



PROCESAMIENTO DE HECHOS

DESARROLLO

- EXTRACCION DE DATOS
- TRANSFORMACIONES
- CARGA DE TABLA DE HECHOS
 - INCREMENTAL
 - HISTORICA



SISTEMA DE AUDITORIA

DESARROLLO

- Cual Fue el Ultimo Proceso Que Actualizo Los Registros De La Tabla De Hechos?
- Como Puedo Encontrar Los Registros Cargados Hoy?
- Fue Una Fila Cargada A través Del Proceso Estándar?
- Cuantas Filas Fueron Cargadas Hoy? Cuantas Fueron Descartadas Por Errores?
- Qué Procesos Terminaron Exitosamente?
- Cuantos Registros Tenía La Tabla De Hechos Antes De La Carga?, Cuantos Después?



Microsoft®

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

The logo for SQL TOTAL CONSULTING, featuring the text "SQL TOTAL" in a bold, sans-serif font above the word "CONSULTING" in a smaller font, with a stylized graphic element to the left.