

## Microsoft DTS

- Se trata de una componente predefinida del RDBMS SQL Server 7.0 de Microsoft.
- Como cliente, esta componente se presenta bajo 3 formas:
  - DTS Designer  
Asistente gráfico para la definición de los procesos (paquetes) encargados de la transformación de datos.
  - DTS Import y Export wizards  
Asistente gráfico para la definición de paquetes más simples.
  - DTS programming interfaces (API)  
Interfaces para ser usadas desde leng. de programación (VBasic, VC++)
- Para la ejecución y planificación
  - Servidor SQL Server 7.0 (incluyendo el servicio Agent)

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

1

## Microsoft DTS (1)

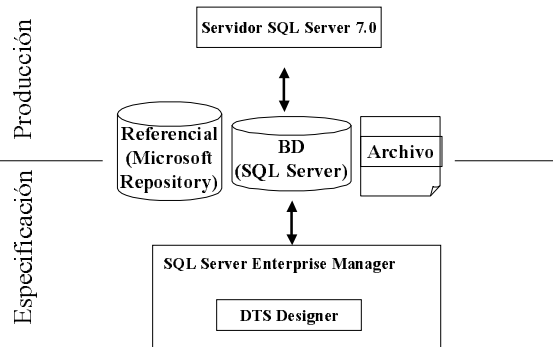
- Permite el uso de un referencial para almacenar todas las definiciones.
  - Necesariamente: Microsoft Repository
  - Otras formas de almacenar las definiciones:
    - » en archivo con formato específico
    - » en SQL Server (dentro de la bd de nombre msdb)

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

2

## DTS / Arquitectura de la herramienta



7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

3

## DTS / Acceso a los datos

- DTS se apoya fuertemente en el acceso y almacenamiento a través de OLE DB.
  - DTS es un consumidor OLE DB
- Provee conexiones específicas para archivos de texto.

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

4

## DTS / Tareas

- DTS se basa en la definición de tareas y un orden parcial entre ellas.
- La tarea básica que permite definir correspondencias y transformaciones entre la fuente de datos origen y la fuente de datos destino:
  - Transform data (data pump)
- Transform data
  - Accede y almacena datos a través de Ole DB o archivos de texto.
  - Transforma datos entre las fuentes.
  - La transformación puede tratarse de (extremos):
    - » una simple copia entre columnas
    - » una invocación de un script (VB Script, JScript)

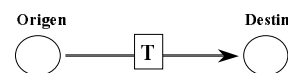
7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

5

## DTS / Transform data

Simplificando,



T:

- copia
- copia más transformaciones incluyendo funciones en un lenguaje script (VB Script, JScript).

se interpreta como:

```
for each o in Origen
    columnas(d) = T ( columnas(o) );
    insert d en Destino;
endfor
```

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

6

## DTS / Otras tareas

- **Data Driven**
  - Permite realizar actualizaciones y borrados además de inserciones.
- **Execute SQL**
  - Permite definir un conjunto de instrucciones SQL.
- **Execute Process**
  - Permite invocar a un ejecutable (.exe, .bat)
- **Send Mail**

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

7

## DTS / Otras tareas (1)

- **Bulk Insert**
  - Método rápido para copiar datos en archivos ascii a una bd SQL Server. No permite definición de transformaciones.
- **Active X Script**
  - Permite invocar un Active X script (VB Script, Perl Script, Java Script)

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

8

## DTS / Paquete

- Un paquete es un "workflow" que define un proceso de transformación.
- Un paquete es un grafo donde:
  - los nodos son *tareas*, y
  - los arcos representan *pasos* que definen el orden en la cual se ejecutarán las tareas.
- Un paso puede tener asociado una *restricción de precedencia* definiendo cómo el resultado de una tarea determina la ejecución de la otra.
  - on success
  - on failure
  - on completion

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

9

## DTS / Paquete (1)

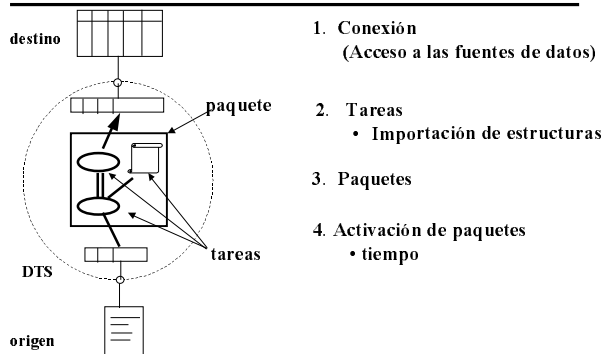
- Posible confusión "gráfica":
  - Transform data es un nodo (tarea) dentro del workflow.

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

10

## DTS / Etapas en la definición



7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

11

## Resumen

DTS	Genio
<ul style="list-style-type: none"> <li>• <b>Paquete</b> <ul style="list-style-type: none"> <li>– Tarea               <ul style="list-style-type: none"> <li>» Transform data</li> <li>» Data driven</li> <li>» Execute SQL</li> <li>» Execute Process</li> <li>» Send Mail</li> <li>» Bulk Insert</li> <li>» Active X Script</li> </ul> </li> <li>– Paso               <ul style="list-style-type: none"> <li>» restricción de precedencia ( on ...)</li> </ul> </li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• <b>Proceso</b> <ul style="list-style-type: none"> <li>– Acción               <ul style="list-style-type: none"> <li>» Módulo + Foreach/Forall + AddRow + datasets</li> <li>» Módulo + Foreach/Forall + AddRow + DeleteRow + UpdateRow + datasets</li> <li>» Módulo + Execute Procedure</li> <li>» Ejecutables</li> <li>» Send</li> <li>» &lt;no hay correspondiente&gt;</li> <li>» &lt;no hay correspondiente&gt;</li> </ul> </li> <li>– (nro, condición, acción 1, acción 2)</li> </ul> </li> </ul>

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

12

### Resumen (1)

- **DTS maneja explícitamente la sincronización**
  - En Genio, se podría “simular” usando vars globales
- **Genio maneja condiciones dentro del proceso**
  - DTS provee las restricciones de precedencia pero no restricciones.
- **Genio provee 3 mecanismos para la activación de procesos.**
  - DTS ofrece activación por tiempo.
- **Genio ofrece un análisis de impacto**
  - DTS ???

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

13

### Observación

- **Las dos últimas transparencias resumen comparativamente las características del "modelo (de WF?)" de Genio y DTS.**

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

14

### Conclusión

- **Un ambiente de data warehousing tiene características diferentes a un ambiente de migración.**
- **Ambiente de data warehousing**
  - Se construye el depósito de datos
  - A intervalos, se agregan cambios al depósito
  - Se archivan o se eliminan datos muy viejos
  - Se ejecutan consultas

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

15

### Conclusión (1)

- **A nuestro conocimiento, no hay una herramienta que realice o ayude a realizar todas las tareas que requiere instanciar (poblar) un data warehouse relacional.**
- **Variedad enfatizando algunos aspectos más que otros**
  - Análisis del estado de los datos origen
  - Limpieza
  - Extracción, transformación y carga
  - Captura de cambios en los datos

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

16

### Conclusion (2)

- **Notar que ETLs no resuelven todas las tareas.**
  - Una solución "comercial hoy" es una combinación de:
    - » herramientas de análisis del estado de los datos (calidad de los datos) a migrar,
    - » herramientas ETL,
    - » herramientas o mecanismos de captura de cambios en los datos.
  - Costo

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

17

### Conclusion (3)

- **La selección de herramientas es una tarea que consume tiempo no despreciable.**
- **Algunas sugerencias**
  - (Como siempre) Importante saber lo que se quiere y se requiere.
  - Fijar un tiempo para evaluar las herramientas y no extenderlo.
  - Entender el diccionario de datos usado y contrastarlo con los ya existentes en su proyecto o los planeados a incorporar.
  - No subestimar el tiempo de analizar que tan "sucios" están los datos origen y la metainformación sobre ellos (documentación, personas).

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

18

#### Conclusión (4)

---

- **Algunas sugerencias (cont.)**
  - No subestimar el volumen y la especificación de las correspondencias.
    - » En una etapa de análisis, *identificar* las correspondencias de 300 atributos a 10 minutos/atr = 50h . 8 días de 8h.
  - << Espacio para compartir experiencias >>

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

19

#### Conclusión (5)

---

- Dificultad para distinguir la problemática que se resuelve bajo los títulos "Limpieza" e "Integración de datos".

7/5/2001

InCo - Aplicación de WF para la definición de procesos de refresco de DWs

20