



Data Warehousing



Esercitazione 1

Pierluigi Del Nostro



Data Warehousing



► IBM DB2 UDB

□ DB2 Universal Database

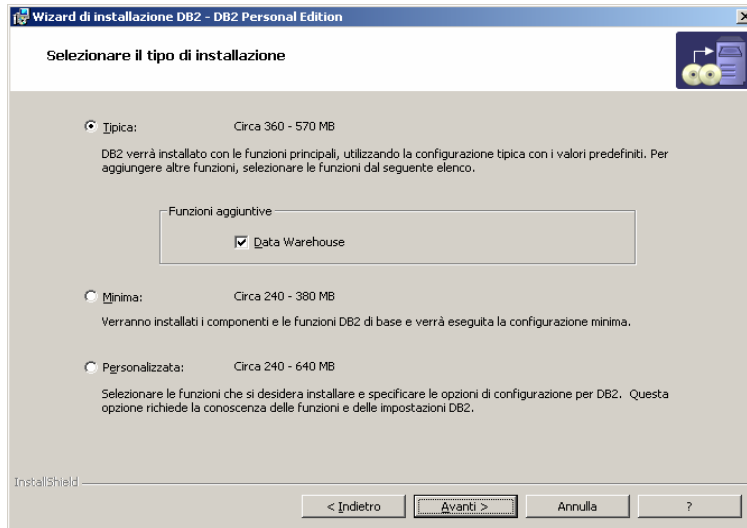
- Suite di strumenti per la gestione dei dati
- Funzioni avanzate per soluzioni Business Intelligence
- Dispone di strumenti di sviluppo del data warehouse e di un On-Line Analytical Processing (OLAP) integrato

□ Componenti (di interesse)

- **Centro di controllo:** Gestisce i sistemi, le istanze, i database e gli oggetti database (tabelle, viste, trigger, stored procedure...)
- **Assistente di configurazione:** Consente la gestione della configurazione e della connettività con i database DB2.
- **Centro Data Warehouse:** Mette a disposizione tutte le funzionalità per implementare e gestire un Data Warehouse (sicurezza, ETL, schemi dimensionali, pianificazione di attività...)
- **Gestione database di controllo Data Warehouse:** Ad ogni Data Warehouse e' associato un database di controllo che contiene tutte le informazioni di configurazione e i metadati.
- **Information Catalogue:** aiuta nelle attività di reperimento, comprensione e accesso alle informazioni disponibili, visualizza i metadati.

Pierluigi Del Nostro

► IBM DB2 UDB » Tips installazione



Pierluigi Del Nostro

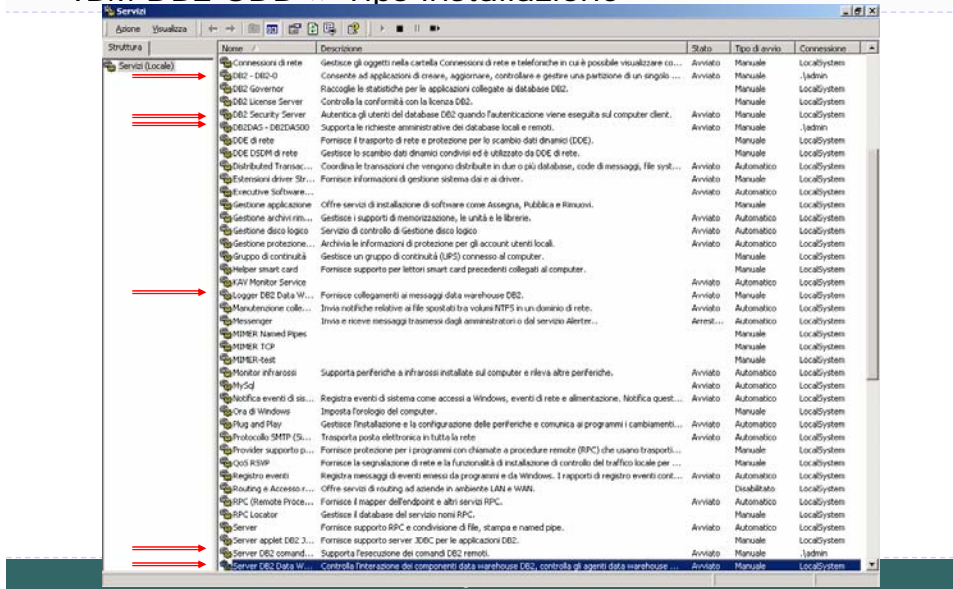
► IBM DB2 UDB » Tips installazione



Pierluigi Del Nostro



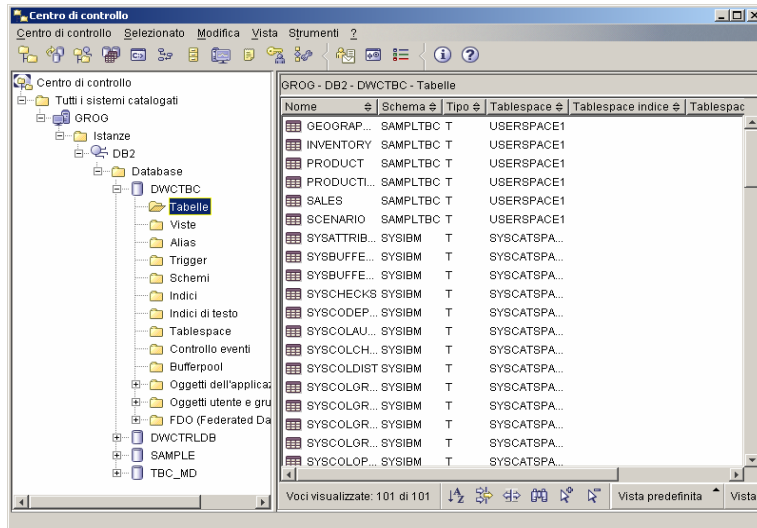
► IBM DB2 UDB » Tips installazione



► IBM DB2 UDB » Database di esempio

- ❑ DWCTB: Fonte informativa relazionale per il data warehouse
- ❑ DWCTRLDB: Database di controllo per data warehouse di default
- ❑ TBC_MD: Database di controllo data warehouse di esempio

► IBM DB2 UDB » Centro di controllo

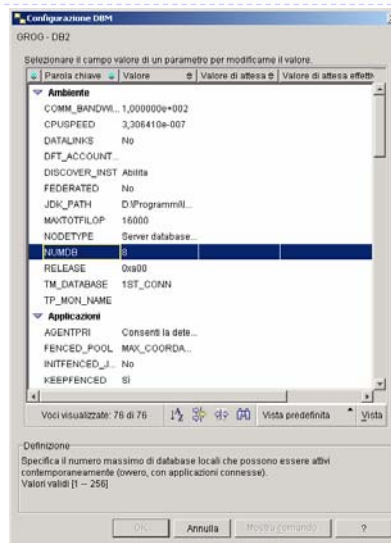


Pierluigi Del Nostro

► IBM DB2 UDB » Centro di controllo

! Possibile errore:
... numero massimo database attivi...

parametri di istanza



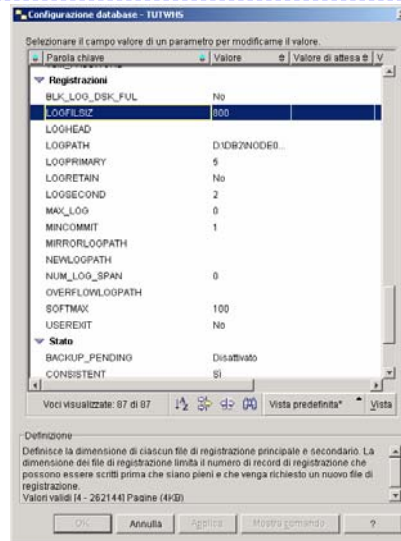
Pierluigi Del Nostro

► IBM DB2 UDB » Centro di controllo

! Possibile errore:

... il file di registrazione è pieno...

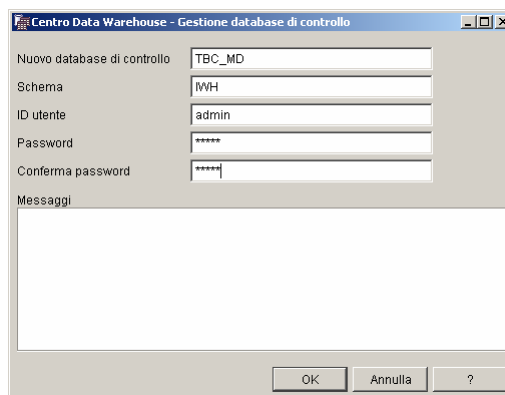
parametri di database



Pierluigi Del Nostro

► IBM DB2 UDB » Database di controllo

- ❑ Start ->Programmi-> IBM DB2 -> Strumenti di configurazione->Gestione database di controllo Data Warehouse



Pierluigi Del Nostro



► IBM DB2 UDB » Centro Data Warehouse

- ❑ Start -> Programmi -> IBM DB2 -> Strumenti Business Intelligence -> Centro Data Warehouse



Pierluigi Del Nostro



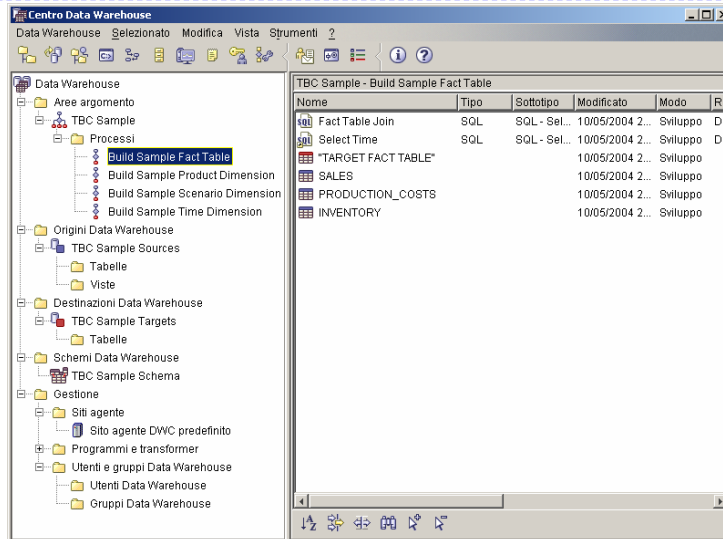
► IBM DB2 UDB » Centro Data Warehouse

- ❑ Struttura organizzativa
 - Origini: file, tabelle
 - Destinazioni: tabelle
 - Aree argomento: raggruppamento logico di processi inerente un determinato argomento o funzione
 - Processi: insieme di passi che effettuano una determinata operazione complessa
 - Passi: singole operazioni sui dati
 - > SQL (select-insert, select-update)
 - > Trasformer (pulizia, statistiche)
 - > Programmi
 - Schema data warehouse: Star schema
 - Sito agente: Si occupa della gestione del flusso di dati da una sorgente ad una destinazione e dell'eventuale esecuzione di programmi di trasformazione.
 - Utenti e gruppi: Politiche d'accesso ai singoli oggetti del DW

Pierluigi Del Nostro



► IBM DB2 UDB » Centro Data Warehouse



Pierluigi Del Nostro



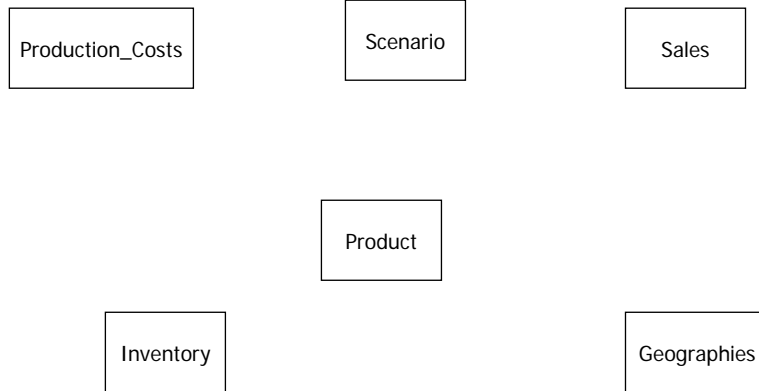
► Problema

- ❑ Siete amministratori del database di una compagnia di produzione di bevande TBC. Il dipartimento finanziario vuole tracciare e analizzare l'andamento delle vendite attraverso diverse regioni geografiche, su base periodica di tutti i prodotti venduti.

Pierluigi Del Nostro



► Schema ER della base di dati operazionele



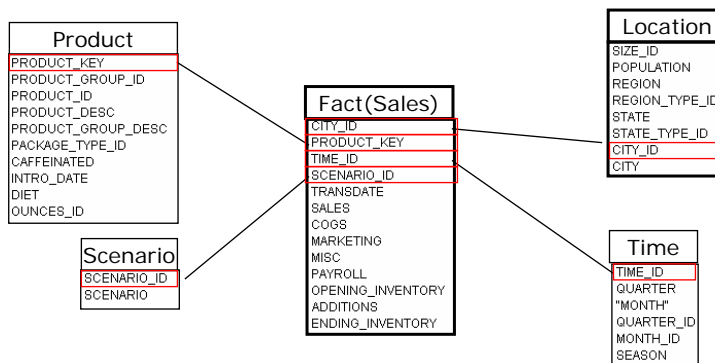
Pierluigi Del Nostro



► Star schema

□ Tabella dei fatti

- Fatto: vendita
- Misure: prodotti venduti, costi di produzione e inventario
- Dimensioni: prodotto, scenario, locazione, tempo

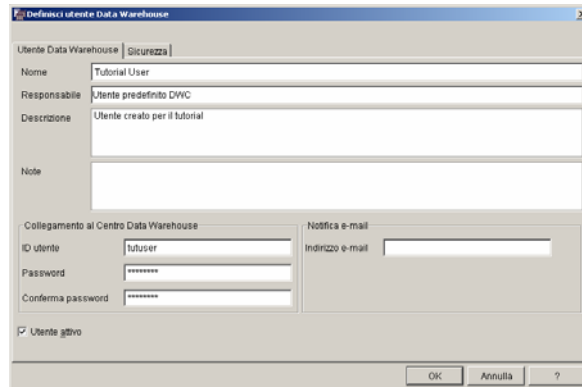
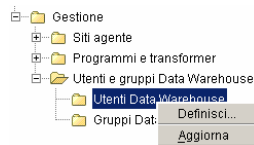


Pierluigi Del Nostro

► Utenti e gruppi

□ Sicurezza

- Autenticazione all'apertura del Centro Data Warehouse
- Definizione di utenze e gruppi per l'accesso agli oggetti del DW

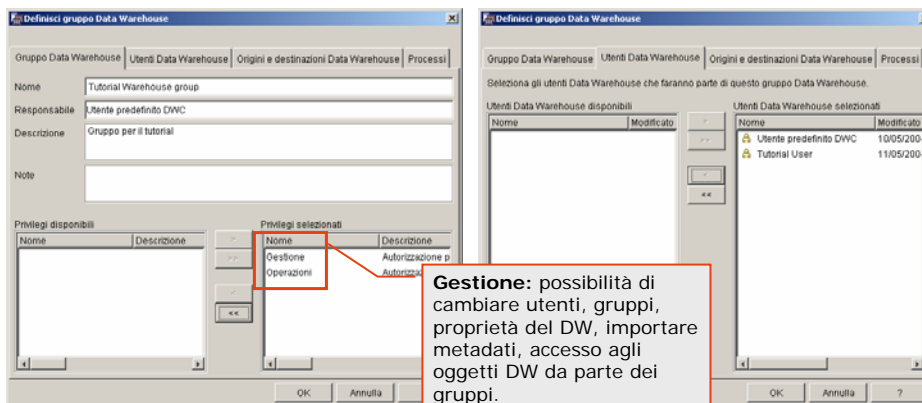


Pierluigi Del Nostro

► Utenti e gruppi

□ Definizione di un gruppo

□ Inserimento dell'utente nel gruppo



Gestione: possibilità di cambiare utenti, gruppi, proprietà del DW, importare metadati, accesso agli oggetti DW da parte dei gruppi.
Operazioni: monitoraggio dello stato dei processi

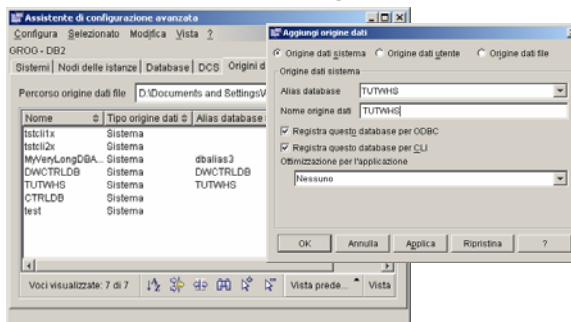
Pierluigi Del Nostro



► Creare un warehouse database

- ❑ Start -> Programmi -> IBM DB2 Strumenti riga comandi -> CLP (Command Line Processor)
 - Digitare: `create db TUTWHS`
 - Attendere messaggio esito
- ❑ Ogni database DB2 per essere accessibile deve essere registrato, utilizzare l'Assistente di configurazione

per visualizzare il DB:
Centro di controllo,
tasto dx su database,
Aggiorna



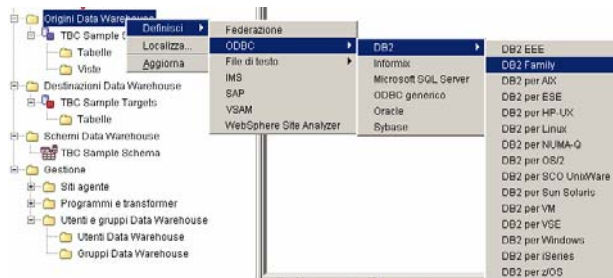
Pierluigi Del Nostro



► Definire un'origine

- ❑ Definiamo due origini:
 - **Relazionale:** La tabella GEOGRAPHIES del database DWCTBC
 - **File:** Un file di informazioni demografiche

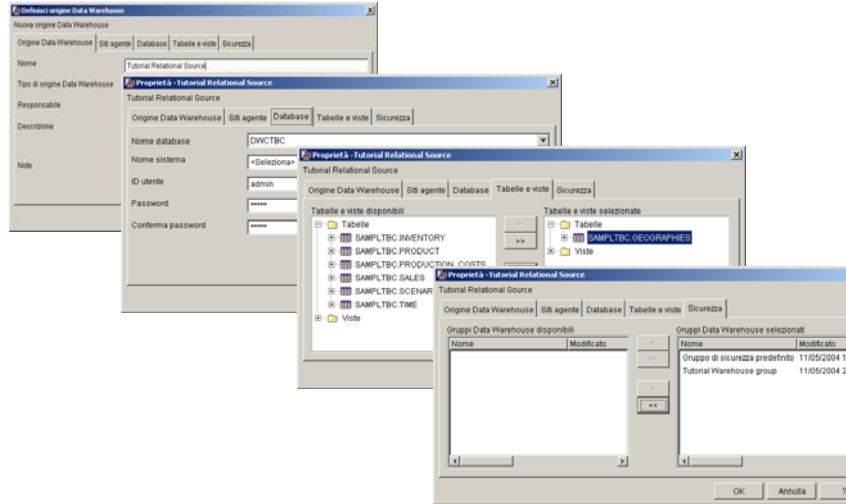
❑ Origine relazionale:



Pierluigi Del Nostro



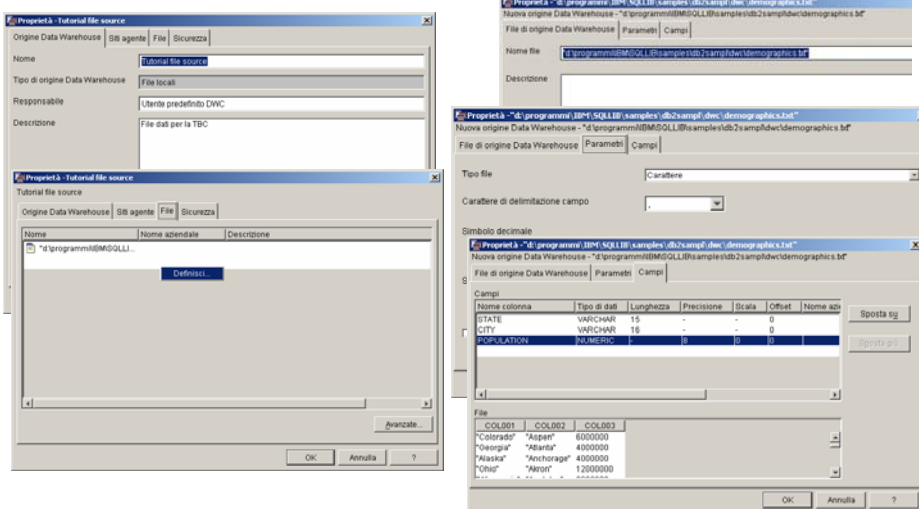
► Definire un'origine relazionale



Pierluigi Del Nostro

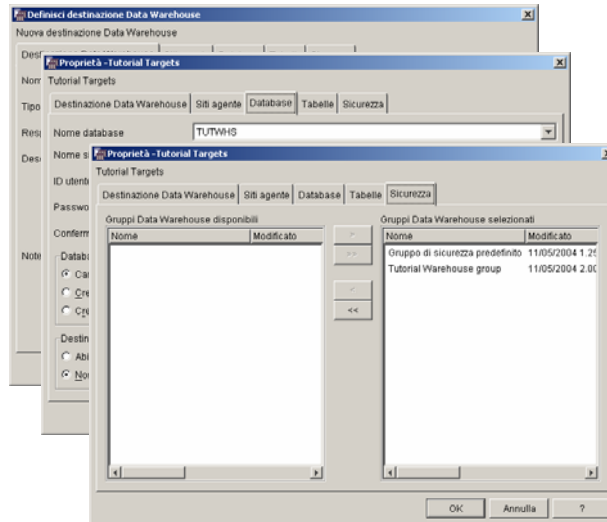


► Definire un'origine file



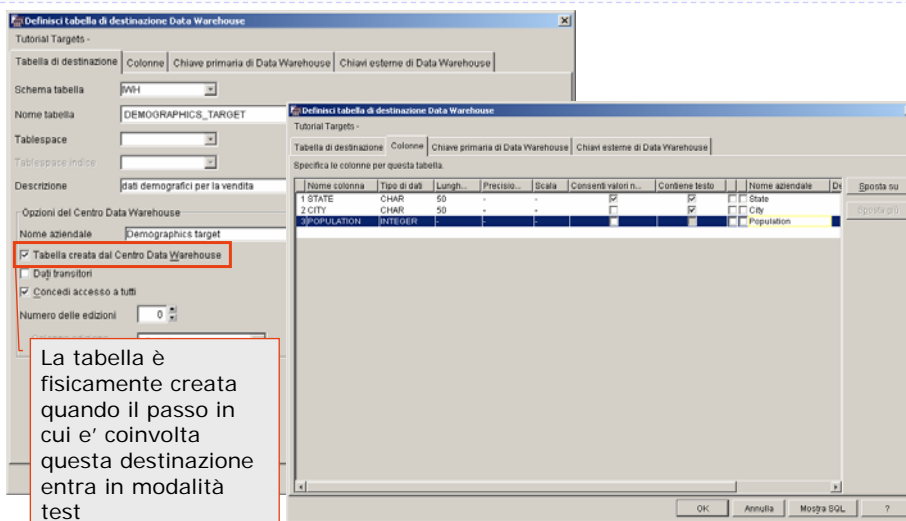
Pierluigi Del Nostro

► Definire una destinazione



Pierluigi Del Nostro

► Definire una destinazione



La tabella è fisicamente creata quando il passo in cui e' coinvolta questa destinazione entra in modalita test

Pierluigi Del Nostro



► Definire un'area argomento

- ❑ Area argomento: raggruppamento logico di processi inerenti uno stesso argomento o funzione

Pierluigi Del Nostro

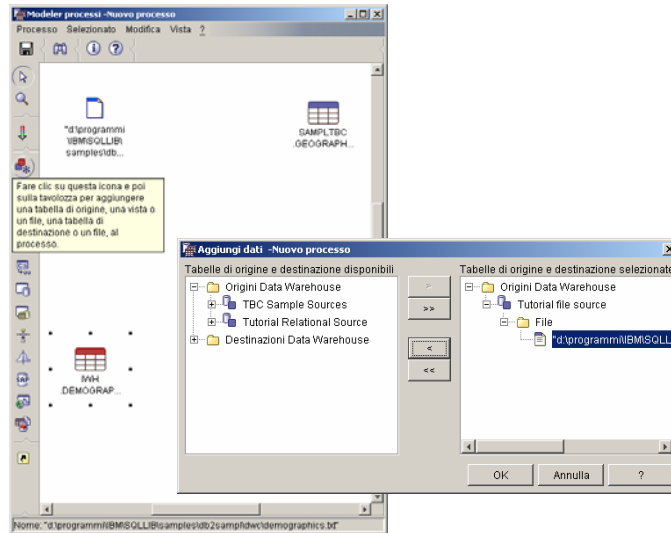


► Definire un processo

- ❑ Processo: serie di passi che definiscono lo spostamento e la trasformazione di dati.

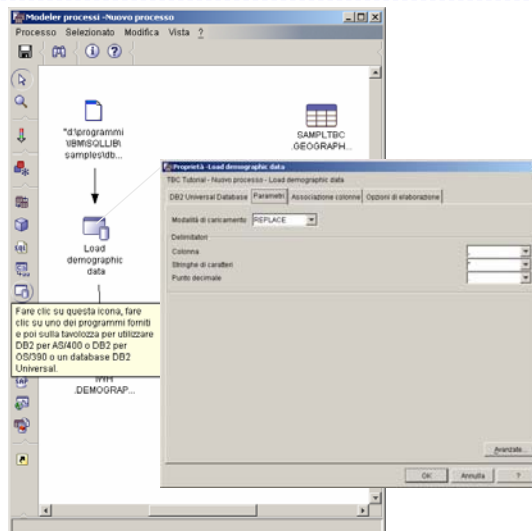
Pierluigi Del Nostro

► Aggiungere origini e destinazioni ad un processo



Pierluigi Del Nostro

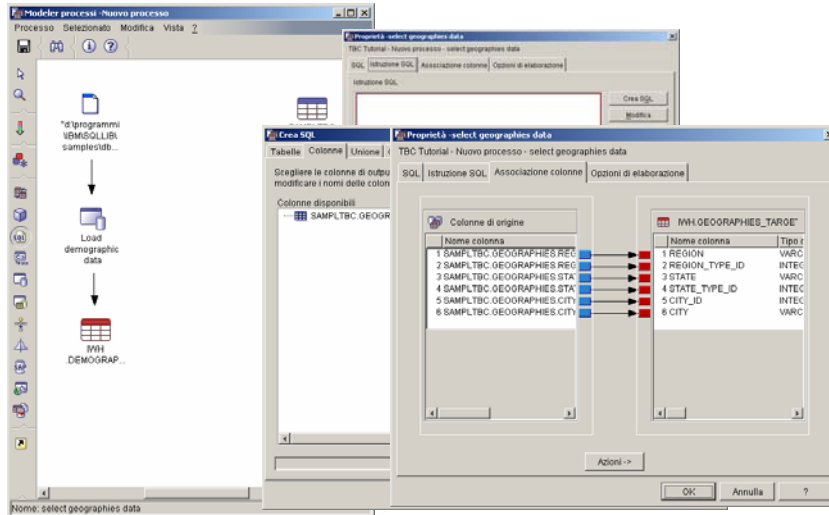
► Aggiungere un passo LOAD



Pierluigi Del Nostro



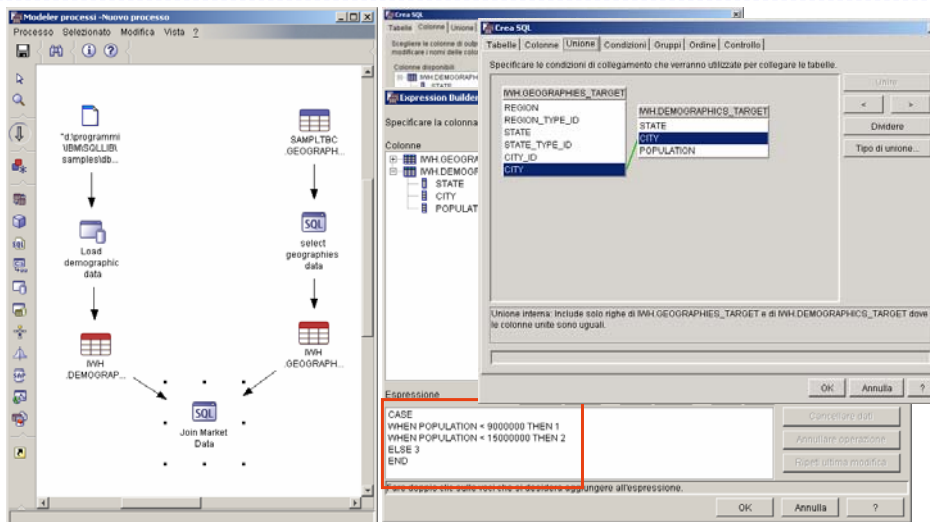
► Aggiungere un passo SQL



Pierluigi Del Nostro



► Effettuare il JOIN tra tabelle

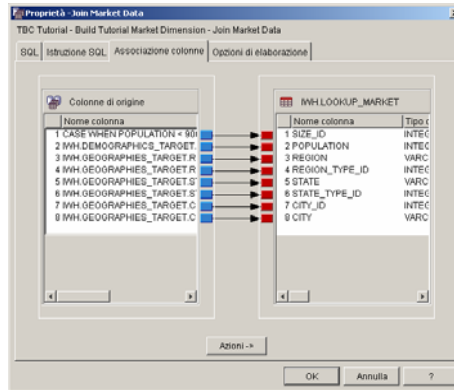
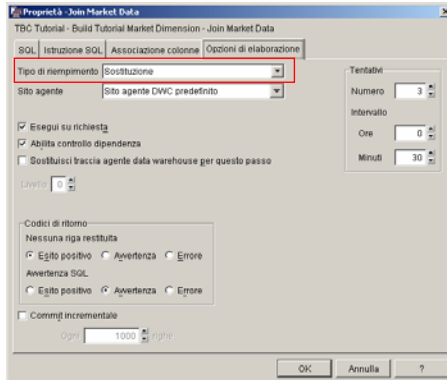


Pierluigi Del Nostro



► Effettuare il JOIN tra tabelle

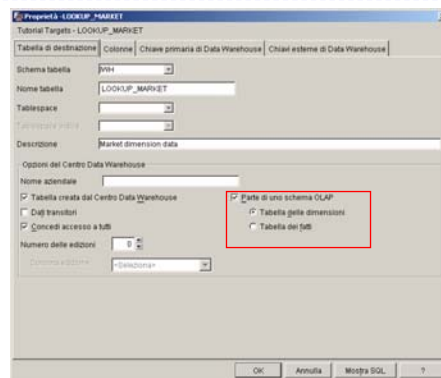
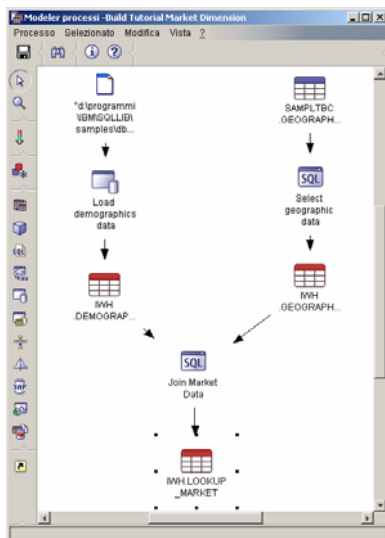
- ❑ E' possibile creare automaticamente la tabella contenente il risultato del JOIN



Pierluigi Del Nostro



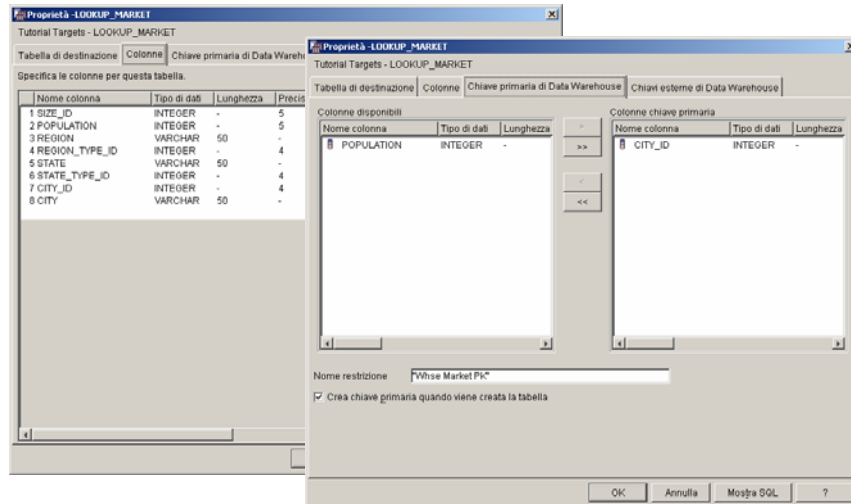
► Effettuare il JOIN tra tabelle



Pierluigi Del Nostro



► Definizione chiave Datawarehouse

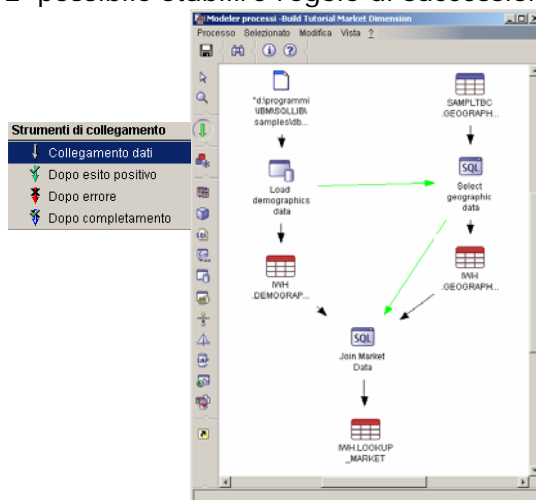


Pierluigi Del Nostro



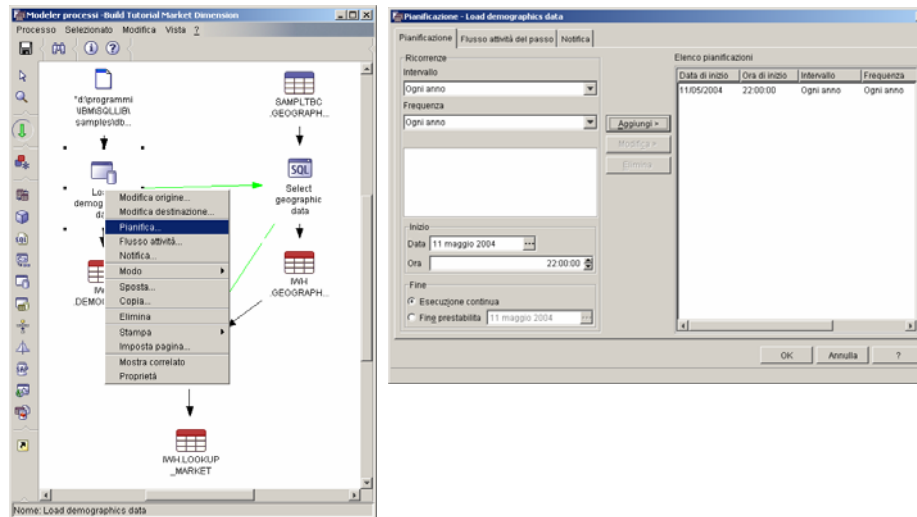
► Sequenze di passi

- ❑ E' possibile stabilire regole di successione tra i passi



Pierluigi Del Nostro

► Scheduling dei passi



Pierluigi Del Nostro

► Cosa manca?

- Definizione dei processi di popolamento delle rimanenti dimensioni e della tabella dei fatti
- Definizione di uno Star Schema nel Centro di controllo Data Warehouse
- Gestione metadati
- Interrogazioni OLAP

Pierluigi Del Nostro