

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

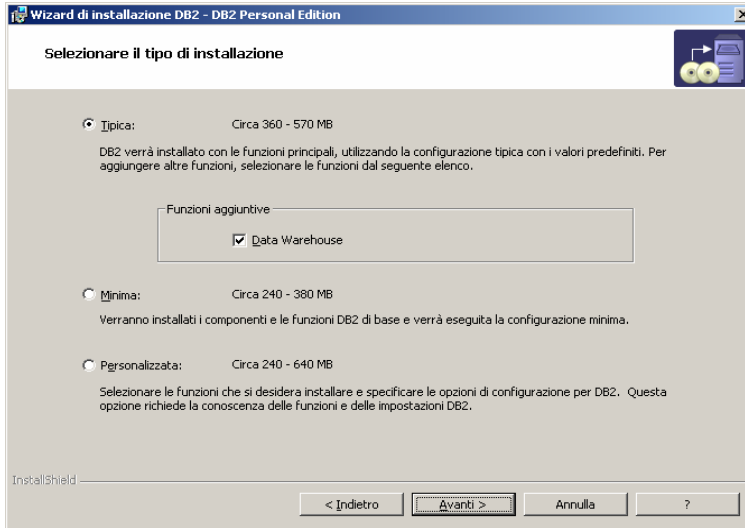
□ **Componenti(di interesse)**

- **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

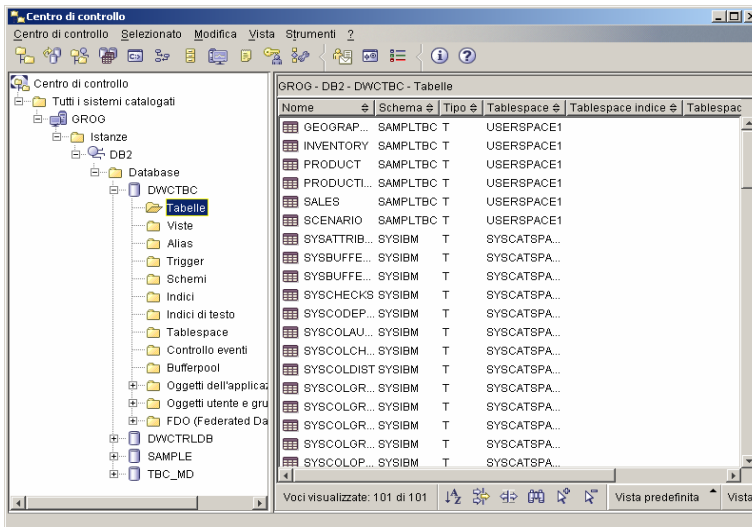


- ▶ IBM DB2 UDB » Database di esempio
 - ❑ DWCTB: Fonte informativa relazionale per il data warehouse
 - ❑ DWCTRLDB: Database di controllo per data warehouse di default
 - ❑ TBC_MD: Database di controllo data warehouse di esempio

Pierluigi Del Nostro



- ▶ IBM DB2 UDB » Centro di controllo



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

The screenshot shows the IBM DB2 UDB Centro Data Warehouse interface. The left pane displays a tree view of the data warehouse structure, including 'Data Warehouse', 'Aree argomento', 'TBC Sample', 'Processi', 'Origini Data Warehouse', 'Destinazioni Data Warehouse', 'Schemi Data Warehouse', and 'Gestione'. The right pane shows a table titled 'TBC Sample - Build Sample Fact Table' with the following data:

Nome	Tipo	Sottotipo	Modificato	Modo	Re
Fact Table Join	SQL	SQL - Sel...	10/05/2004 2...	Sviluppo	De
SelectTime	SQL	SQL - Sel...	10/05/2004 2...	Sviluppo	De
"TARGET FACT TABLE"			10/05/2004 2...	Sviluppo	
SALES			10/05/2004 2...	Sviluppo	
PRODUCTION_COSTS			10/05/2004 2...	Sviluppo	
INVENTORY			10/05/2004 2...	Sviluppo	

Pierluigi Del Nostro



► Problema

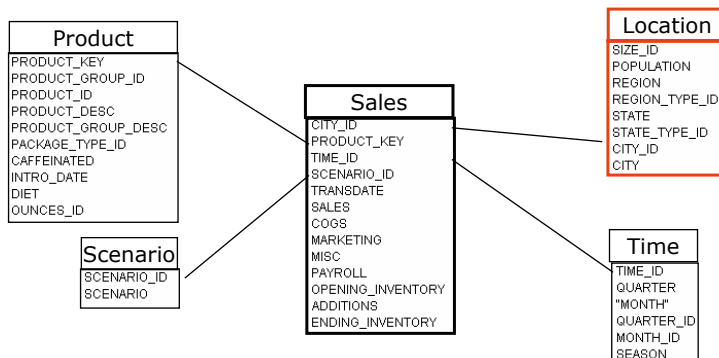
- ❑ Siete amministratori del database di una compagnia di produzione di bevande TBC. Il dipartimento finanziario vuole tracciare e analizzare l'andamento delle vendite attraverso diverse regioni geografiche, su base periodica di tutti i prodotti venduti.

Pierluigi Del Nostro



► Star schema

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



Pierluigi Del Nostro



► IBM DB2 UDB » Database di controllo

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

The screenshot shows a dialog box titled "Centro Data Warehouse - Gestione database di controllo". It contains the following fields and controls:

- Nuovo database di controllo: TBC_MD
- Schema: WWH
- ID utente: admin
- Password: *****
- Conferma password: *****
- Messaggi: (empty text area)
- Buttons: OK, Annulla, ?

Pierluigi Del Nostro



► IBM DB2 UDB » Centro Data Warehouse

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

The screenshot shows two overlapping dialog boxes:

- Collegamento al Centro Data Warehouse:** Fields for ID utente and Password, with buttons for OK, Annulla, Avanzate..., and ?.
- Impostazioni avanzate:** Fields for Database di controllo (TBC_MD), Nome host server (localhost), and Nome servizio server (wkernel), with buttons for OK, Annulla, and ?.

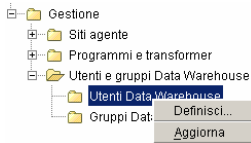
Pierluigi Del Nostro



► Utenti e gruppi

□ Sicurezza

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



Definisci utente Data Warehouse

Utente Data Warehouse | sicurezza

Nome: Tutorial User

Responsabile: Utente predefinito DWC

Descrizione: Utente creato per il tutorial

Note:

Collegamento al Centro Data Warehouse

ID utente: tutuser

Password: *****

Conferma password: *****

Indirizzo e-mail:

Utente attivo

OK Annulla ?

Pierluigi Del Nostro



► Utenti e gruppi

□ Definizione di un gruppo

□ Inserimento dell'utente nel gruppo

Definisci gruppo Data Warehouse

Gruppo Data Warehouse | Utenti Data Warehouse | Origini e destinazioni Data Warehouse | Processi

Nome: Tutorial Warehouse group

Responsabile: Utente predefinito DWC

Descrizione: Gruppo per il tutorial

Note:

Privilegi disponibili

Nome	Descrizione
Operazioni	

Privilegi selezionati

Nome	Descrizione
Operazioni	Autorizzazione p...

Definisci gruppo Data Warehouse

Seleziona gli utenti Data Warehouse che faranno parte di questo gruppo Data Warehouse.

Utenti Data Warehouse disponibili

Nome	Modificato
Utente predefinito DWC	10/05/200...
Tutorial User	11/05/200...

Utenti Data Warehouse selezionati

Nome	Modificato
Utente predefinito DWC	10/05/200...
Tutorial User	11/05/200...

OK Annulla ?

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

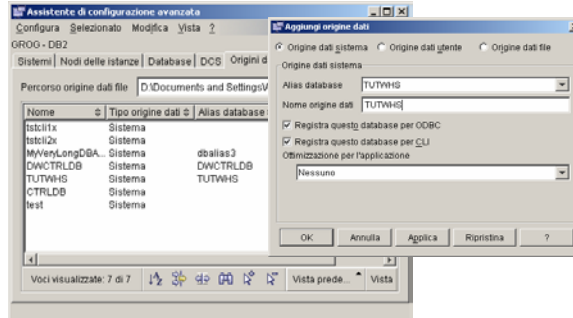
Pierluigi Del Nostro



► 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

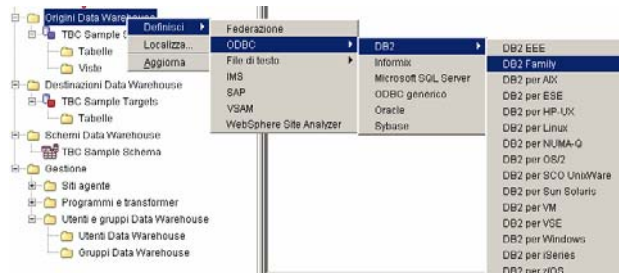


Pierluigi Del Nostro



► Definire un'origine

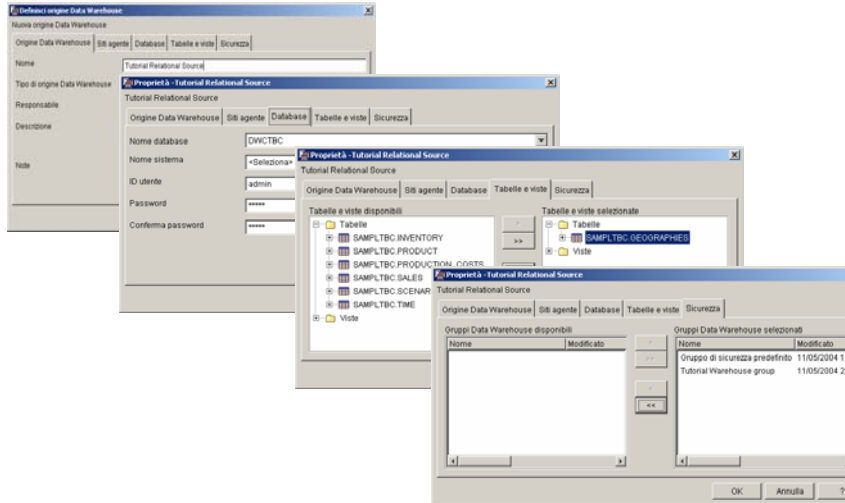
- ❑ Definiamo due origini:
 - **Relazionale:** La tabella GEOGRAPHIES del database DWCTBC
 - **File:** Un file di informazioni demografiche
- ❑ Origine relazionale:



Pierluigi Del Nostro



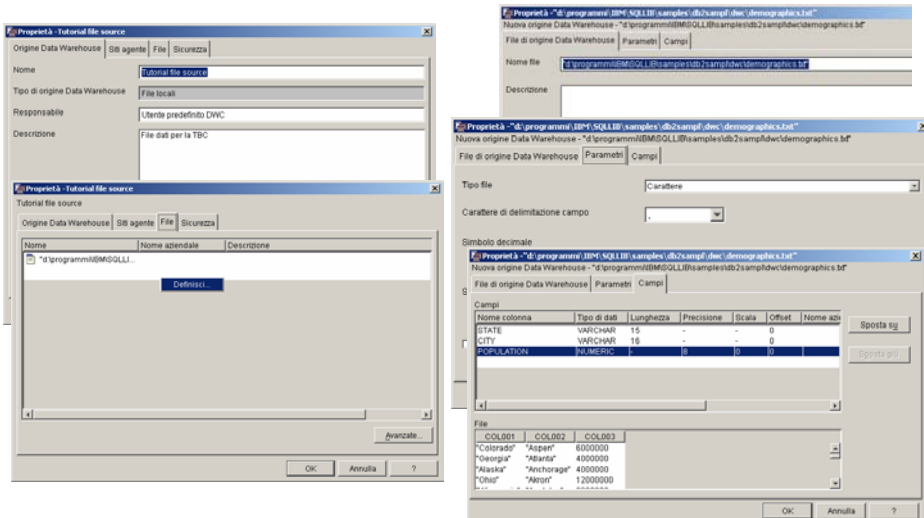
► Definire un'origine relazionale



Pierluigi Del Nostro



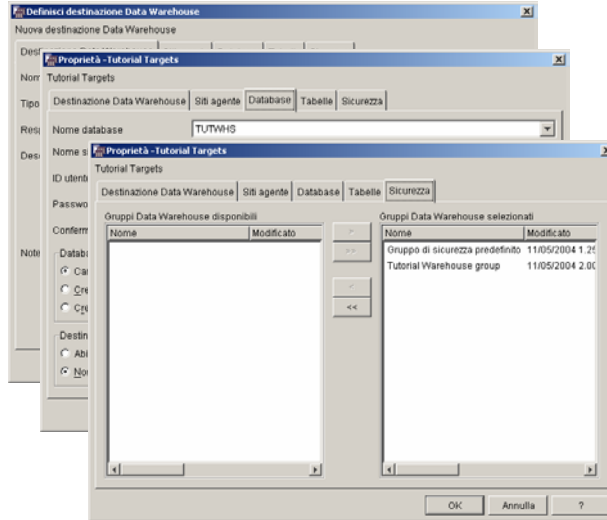
► Definire un'origine file



Pierluigi Del Nostro



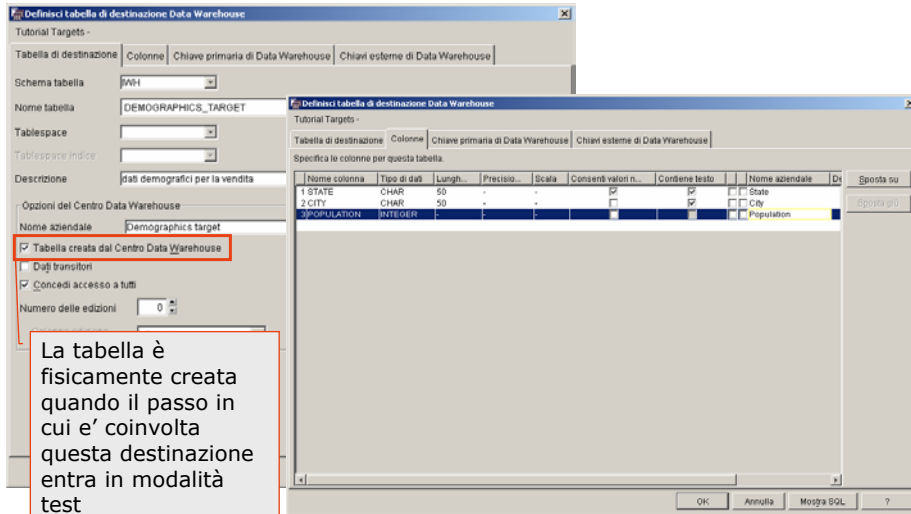
► Definire una destinazione



Pierluigi Del Nostro



► Definire una destinazione



La tabella è fisicamente creata quando il passo in cui è coinvolta questa destinazione entra in modalità test

Pierluigi Del Nostro



► Definire un'area argomento

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

Proprietà - TBC Tutorial

TBC Tutorial

Area argomento

Nome: TBC Tutorial

Responsabile: Utente predefinito DWC

Descrizione: Subject area

Note:

OK Annulla ?

Pierluigi Del Nostro



► Definire un processo

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

Proprietà - Build Tutorial Market Dimension

TBC Tutorial - Build Tutorial Market Dimension

Processo Sicurezza

Nome: Build Tutorial Market Dimension

Responsabile: Utente predefinito DWC

Descrizione: Processo per la costruzione della dimensione mercato

Note:

OK Annulla ?

Proprietà - Build Tutorial Market Dimension

TBC Tutorial - Build Tutorial Market Dimension

Processo Sicurezza

Gruppi Data Warehouse disponibili

Nome	Modificato
------	------------

Gruppi Data Warehouse selezionati

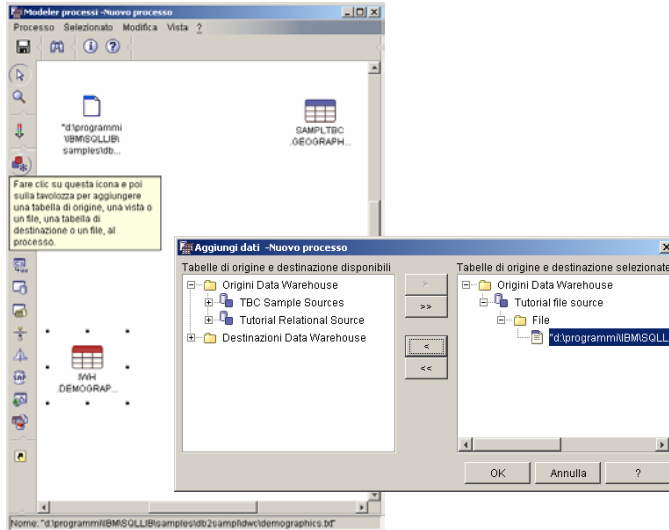
Nome	Modificato
Gruppo di sicurezza predefinito	
Tutorial Warehouse group	

OK Annulla ?

Pierluigi Del Nostro



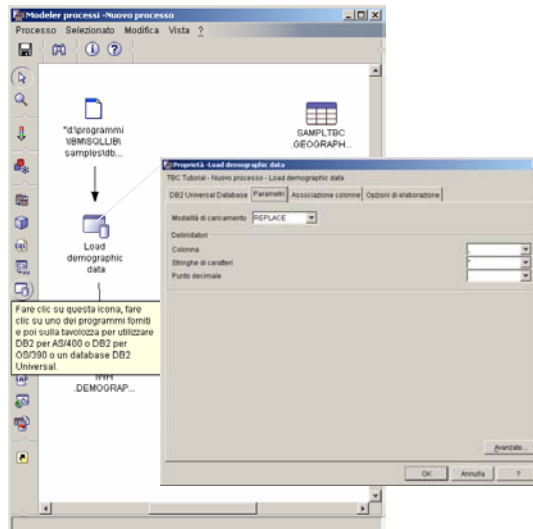
► Aggiungere origini e destinazioni ad un processo



Pierluigi Del Nostro



► Aggiungere un passo LOAD



Pierluigi Del Nostro



► Aggiungere un passo SQL

Pierluigi Del Nostro



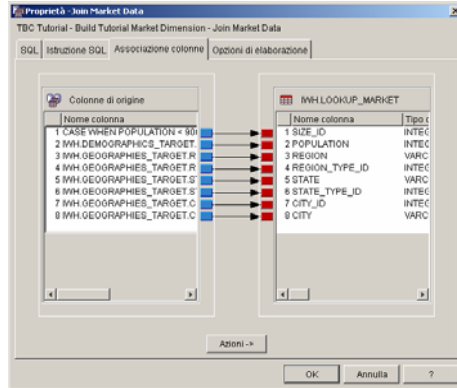
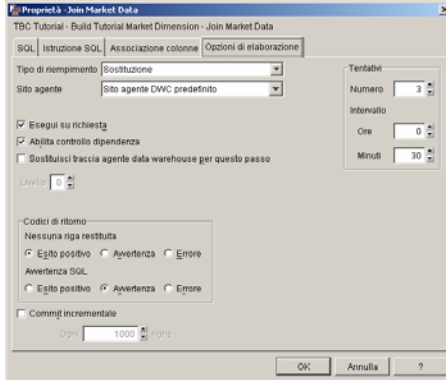
► Effettuare il JOIN tra tabelle

Pierluigi Del Nostro



► Effettuare il JOIN tra tabelle

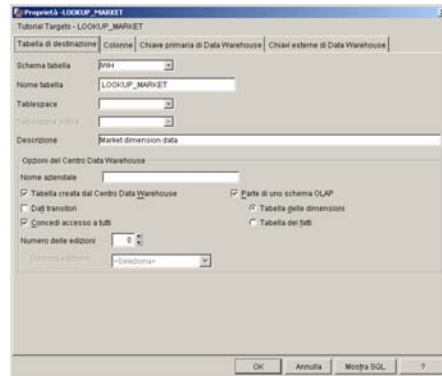
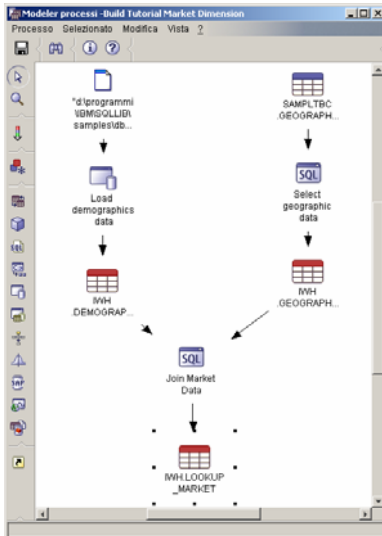
- ❑ E' possibile creare automaticamente la tabella contenente il risultato del JOIN



Pierluigi Del Nostro



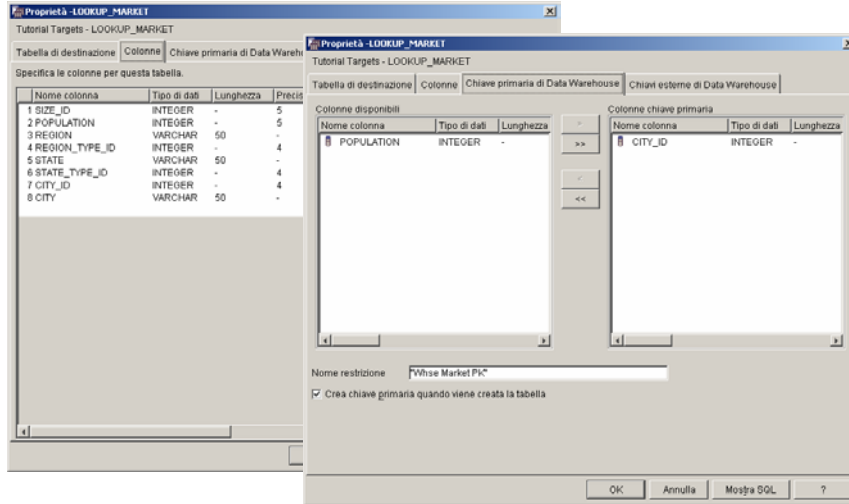
► Effettuare il JOIN tra tabelle



Pierluigi Del Nostro



► Effettuare il JOIN tra tabelle

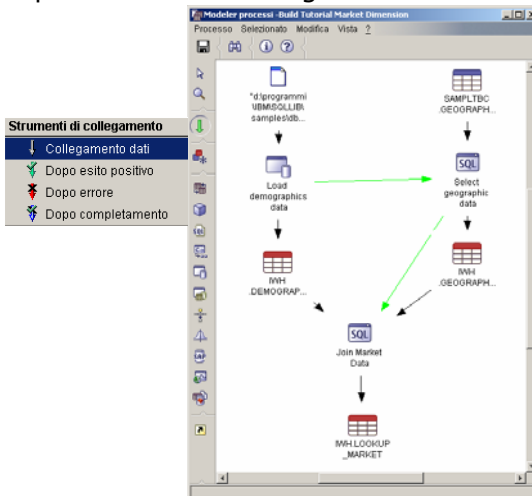


Pierluigi Del Nostro



► Sequenze di passi

- E' possibile stabilire regole di successione tra i passi



Pierluigi Del Nostro



► Scheduling dei passi

The screenshot shows two windows from SQL Enterprise Manager. The left window, titled 'Modelo processi: build Tutorial Market Dimension', displays a process flow diagram. A context menu is open over a step, with 'Pianifica...' selected. The right window, titled 'Pianificazione - Lead demographic data', shows the scheduling configuration for the selected step. It includes fields for recurrence, interval, frequency, start date, and end date, along with a table of planned executions.

Data di inizio	Ora di inizio	Intervallo	Frequenza
11/05/2004	22:00:00	Ogni anno	Ogni anno

Pierluigi Del Nostro



► Cosa manca?

- Visualizzazione dei risultati del processo implementato
- Definizione di uno Star Schema nel Centro di controllo Data Warehouse
- Gestione metadati

Pierluigi Del Nostro

Contacts: pdn@dia.uniroma3.it

Pierluigi Del Nostro