



# Data Warehousing



## Esercitazione 2

Pierluigi Del Nostro



# Data Warehousing



### ► Reminder...

- In laboratorio è presente un'installazione Enterprise di DB2.
- Per accedere richiedere un account come specificato sul sito del corso

Pierluigi Del Nostro



## ► Riepilogo

- ❑ Esercitazione 1:
  - Descrizione dello strumento IBM DB2 UDB
    - Centro di controllo
    - Centro Data Warehouse
  - Esempio di realizzazione di un Data Warehouse
    - Traccia d'esempio
    - Star schema
    - Il database di controllo
    - Utenti e gruppi
    - Creazione e registrazione del database Data Warehouse
    - Definizione di origini
    - Definizione di destinazioni
    - Definizione di un'area argomento
    - Costruzione di un processo
      - Il modeler
      - Inserimento di origini e destinazioni
      - Utilizzo dei passi
      - Definizione di sequenze tra passi
      - Scheduling dei passi
  - Risultato: Costruzione di una dimensione dello star schema

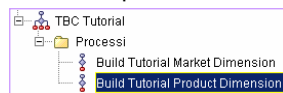
Pierluigi Del Nostro



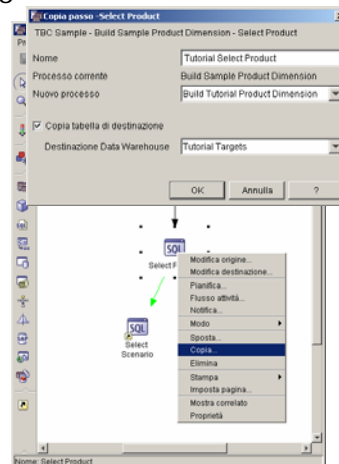
## ► Creazione di uno Star Schema

- ❑ Creare le rimanenti tabelle che definiscono lo star schema
- ❑ Prima tutte le dimensioni e poi il fatto
- ❑ I processi che creano le tabelle sono già implementati
- ❑ Funzione di copia passi

- Creare un processo vuoto



- Aprire il processo **Build Sample Product Dimension**
- Right-click sul passo SQL
- Copia

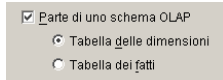


Pierluigi Del Nostro

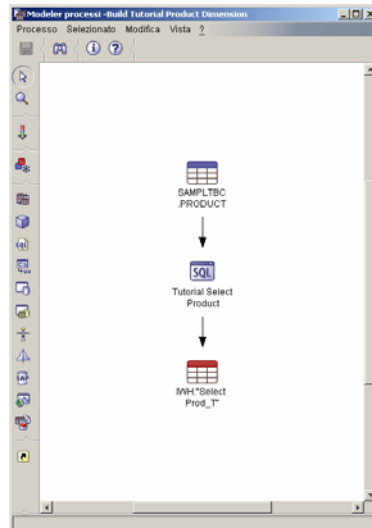


## ► Creazione di uno Star Schema

- ❑ Aprire il processo  
**Build Sample Product Dimension**
- ❑ Rinominare la tabella destinazione  
in: **LOOKUP\_PRODUCT**
- ❑ Verificare nelle proprietà della  
tabella che:



- ❑ Salvare il processo



Pierluigi Del Nostro



## ► Creazione di uno Star Schema

- ❑ Ripetere gli stessi passi per le tabelle dimensione Time e Scenario tenendo conto delle seguenti informazioni

	Time dimension table	Scenario dimension table
Tutorial process:	Build Tutorial Time Dimension	Build Tutorial Scenario Dimension
Sample Process:	Build Sample Time Dimension	Build Sample Scenario Dimension
Sample step to copy:	Select Time	Select Scenario
New tutorial step name:	Tutorial Select Time	Tutorial Select Scenario
Source tables:	TIME	SCENARIO
Target table:	"SelectTime_T"	"SelectScenario_T"
New target table name:	LOOKUP_TIME	LOOKUP_SCENARIO
Warehouse target:	Tutorial Targets	Tutorial Targets

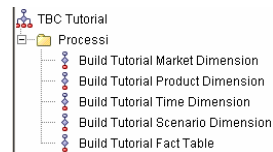
Pierluigi Del Nostro



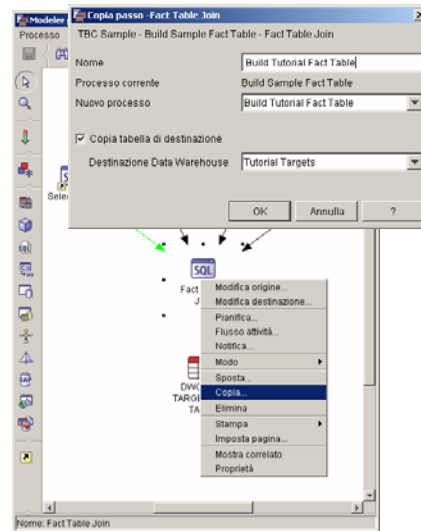
## ► Creazione di uno Star Schema

### ❑ Creazione della tabella fatto:

- Creare il processo:  
Build Tutorial Fact Table  
Nell'area argomento:  
TBC Tutorial



- Procedere come per le  
tabelle dimensione

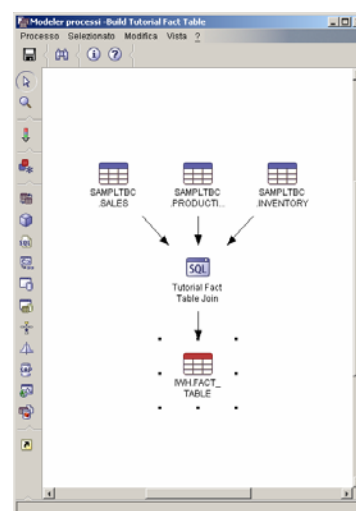


Pierluigi Del Nostro



## ► Creazione di uno Star Schema

- ❑ Aprire il processo:  
Build Tutorial Fact Table
- ❑ Rinominare la tabella destinazione  
in: FACT\_TABLE



Pierluigi Del Nostro



## ► Creazione di uno Star Schema

- ❑ Creazione delle tabelle fisiche di destinazione
- ❑ Promuovere i passi in modo *Verifica*
  1. Aprire il processo: Build Tutorial Fact Table
  2. Right-click sul passo: Tutorial Fact Table Join
  3. Modo → Verifica
  4. Right-click sul passo: Tutorial Fact Table Join → **Verifica**
- ❑ Ripetere i passi per le tabelle dimensioni seguendo le informazioni:

Process	Step
Build Tutorial Product Dimension	Tutorial Select Product
Build Tutorial Scenario Dimension	Tutorial Select Scenario
Build Tutorial Time Dimension	Tutorial Select Time
Build Tutorial Market Dimension	Load Demographics Data, Select Geographies Data, Join Market Data. If these steps are in production mode, demote them to test.

Pierluigi Del Nostro



## ► Creazione di uno Star Schema

- ❑ Definizione delle chiavi primarie nelle tabelle destinazione:
  - Avviare il centro di controllo
  - Vista -> Aggiorna
  - Per aggiungere una chiave primaria:
    - Right-click sulla tabella -> Modifica
    - Chiavi -> Aggiungi chiave primaria
    - Selezionare il campo e cliccare su >
  - Definire le chiavi per:

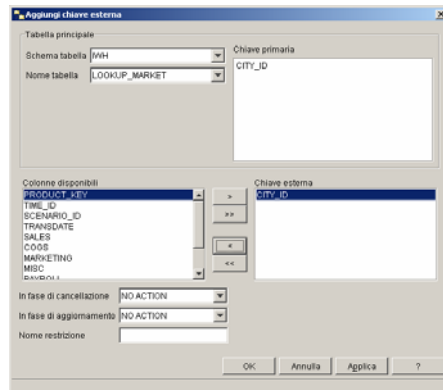
Table	Primary key
LOOKUP_TIME	TIME_ID
LOOKUP_PRODUCT	PRODUCT_KEY
LOOKUP_SCENARIO	SCENARIO_ID

Pierluigi Del Nostro



## ► Creazione di uno Star Schema

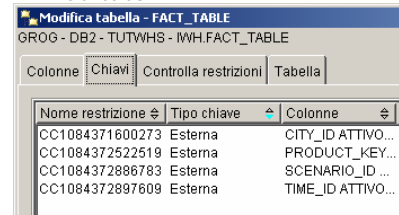
- ❑ Definizione delle chiavi esterne:
  - Stabiliscono le relazioni tra la tabella fatti e le tabelle dimensioni
- ❑ Chiave esterna tra FACT\_TABLE e LOOKUP\_MARKET
- ❑ Sempre nel centro di controllo...



### • Ripetere per le tabelle:

Table	Foreign key
LOOKUP_TIME	TIME_ID
LOOKUP_PRODUCT	PRODUCT_KEY
LOOKUP_SCENARIO	SCENARIO_ID

### • Risultato



Pierluigi Del Nostro



## ► Creazione di uno Star Schema

- ❑ Definizione delle chiavi esterne nel Centro Data Warehouse:
  - Dalla destinazione Tutorial Target -> Tabelle
  - Right-click su FACT\_TABLE -> Proprietà
  - Chiavi esterne di Data Warehouse
  - Eliminare quelle presenti (portate in fase di copiatura del passo)
  - Right-click -> Definisci
  - Impostare lo schema IWH
  - Selezionare la tabella LOOKUP\_MARKET
  - Dalle colonne disponibili selezionare CITY\_ID e cliccare su >
  - Nel nome restrizione: **Whse Market FK**
- ❑ Ripetere per:

Table	Column	Constraint name
LOOKUP_TIME	TIME_ID	Whse Time FK
LOOKUP_PRODUCT	PRODUCT_KEY	Whse Product FK
LOOKUP_SCENARIO	SCENARIO_ID	Whse Scenario FK

Pierluigi Del Nostro

**Data Warehousing** Università Roma Tre  
Dipartimento di Informatica e Automazione

► Creazione di uno Star Schema

Definisci chiave esterna di Data Warehouse

Oggetto principale  
 Schema oggetto: MWH  
 Nome oggetto: LOOKUP\_MARKET

Chiave primaria di Data Warehouse  
 Nome colonna: CITY\_ID

Colonne disponibili				Colonne chiave esterna di Data Warehouse			
Nome colonna	Tipo di dati	Lunghezza	Pre	Nome colonna	Tipo di dati	Lunghezza	Pre
PRODUCT_KEY	INTEGER	-	0	CITY_ID	INTEGER	-	0
TIME_ID	INTEGER	-	0				
SCENARIO_ID	INTEGER	-	0				

Nome restrizione: Whse Market FK

Pierluigi Del Nostro

**Data Warehousing** Università Roma Tre  
Dipartimento di Informatica e Automazione

► Creazione di uno Star Schema

- ❑ Definire lo star schema
  - Dal Centro Data Warehouse, right-click sulla cartella Schemi Data Warehouse → Definisci
  - Inserire il nome dello schema e le altre informazioni
  - Se si usa solo un database spuntare la casella Utilizza solo un database e selezionare il database
  - Salva i cambiamenti

Pierluigi Del Nostro



## ► Creazione di uno Star Schema

Proprietà schema Data Warehouse - Tutorial Schema

Tutorial Schema

Schema Data Warehouse

Nome: Tutorial Schema

Responsabile: Utente predefinito DWC

Descrizione: Questo e' lo star chema per TBC

Note:

Utilizza solo un database

Database di destinazione Data Warehouse: TUTWHS




OK Annulla ?

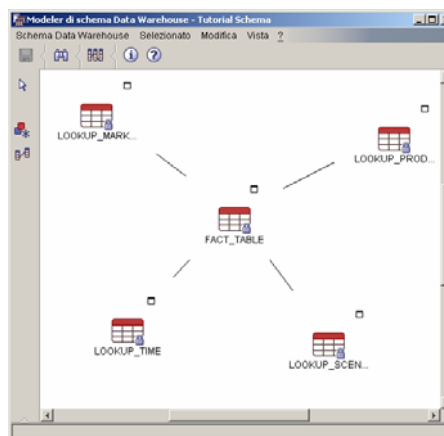
Pierluigi Del Nostro



## ► Creazione di uno Star Schema

### ❑ Aggiungere tabelle ad uno schema

- Aprire lo schema precedentemente inserito
- Si avvia il modeler
-  permette di aggiungere tabelle
-  per inserire relazioni tra tabelle
-  crea automaticamente le relazioni tra le tabelle selezionate

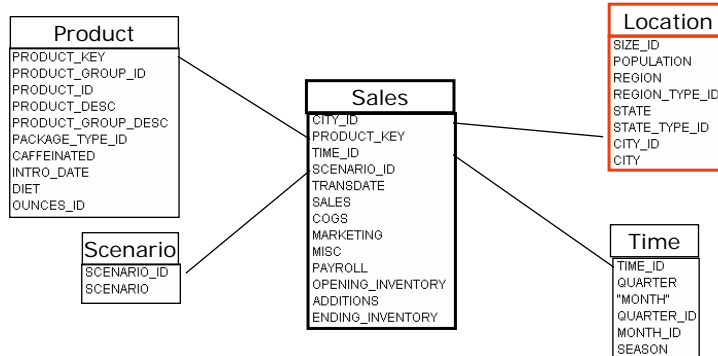


Pierluigi Del Nostro





## ► Creazione di uno Star Schema



Pierluigi Del Nostro



## ► Implementazione → Interrogazione

- Un Data Warehouse è creato per essere interrogato
  - Data Mining
  - OLAP
  
- L'utente finale non è interessato al "motore" che crea ed alimenta il Data Warehouse
  
- Strumenti per l'estrazione e l'analisi dei dati
  - OLAP Center + Office Connect Analytic Edition

Pierluigi Del Nostro



### ► Interrogazione DW » OLAP Center

- A cosa serve OLAP Center?
  - Descrivere la struttura dimensionale delle tabelle relazionali
  - Creare elementi OLAP
  
- Un Data Warehouse deve essere preparato per poter essere utilizzato in DB2:
  - Registrazione della procedura memorizzata di DB2 Multidimensional Metadata Management con il database
  - Creazione delle tabelle di catalogo dei metadati per DB2 Multidimensional Metadata Management

Pierluigi Del Nostro



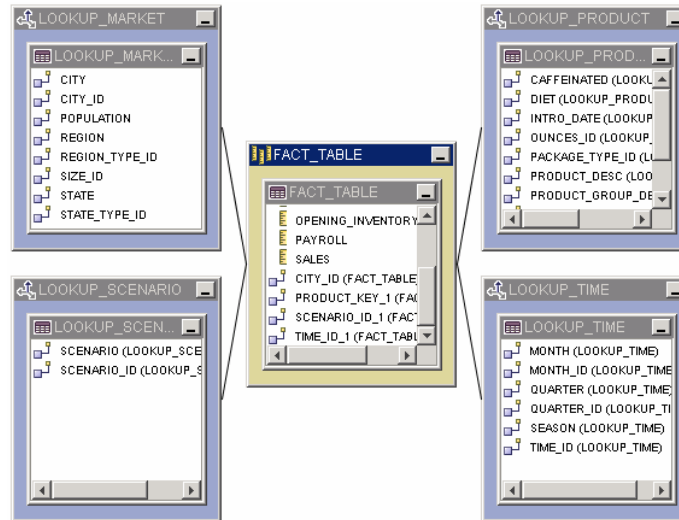
### ► Interrogazione DW » Modello cubo

- rappresentazione di uno schema di tipo star o di tipo snowflake logici.
- ciascuna dimensione può disporre di più gerarchie
- può essere utilizzato da altri strumenti OLAP che supportano il modello cubo e che possono visualizzare le viste multiple di una specifica dimensione (utilizzando più gerarchie).
- componenti obbligatori di un modello cubo in DB2:
  - Un oggetto fatti
  - Almeno una dimensione
  - Almeno una gerarchia corrispondente a ciascuna dimensione
  - Unioni tra l'oggetto fatti esistente e le dimensioni

Pierluigi Del Nostro



## ► Interrogazione DW » Modello cubo



Pierluigi Del Nostro



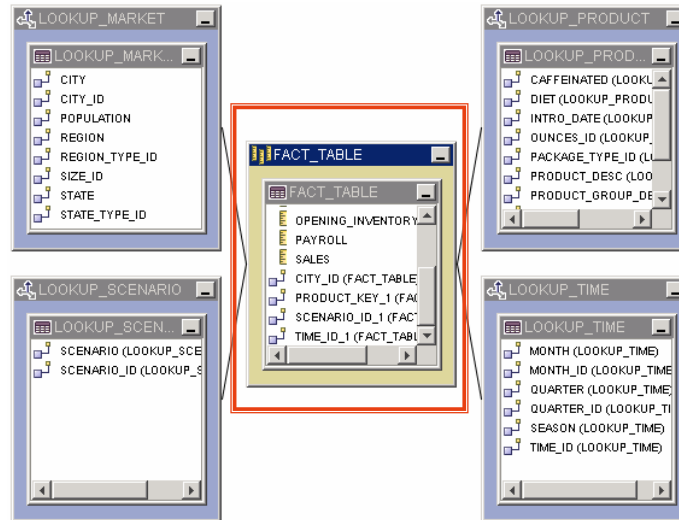
## ► Interrogazione DW » Modello cubo » Fatti

- raggruppa le misure che possono interessare ad una determinata applicazione.
- viene utilizzato nel modello cubo come centro di uno schema di tipo star.
- Proprietà:
  - Insieme di misure: L'insieme di tutte le misure correlate nell'oggetto fatti
  - Insieme di unioni: Insieme di tutte le unioni necessarie ad unire tutte le misure e gli attributi specificati

Pierluigi Del Nostro



## ► Interrogazione DW » Modello cubo » Fatto



Pierluigi Del Nostro



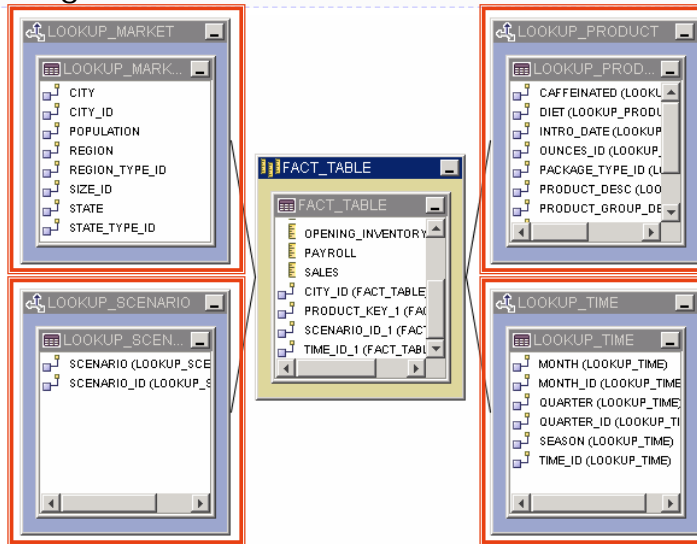
## ► Interrogazione DW » Modello cubo » Dimensioni

- categorizzare un insieme di attributi correlati che insieme descrivono un aspetto della misura.
- utilizzate nei modelli cubo per organizzare i dati nell'oggetto fatti in base a categorie logiche (Luogo, Prodotto, o Tempo).
- Le dimensioni fanno riferimento a una o più gerarchie.
- Proprietà:**
  - Insieme di attributi: Insieme di tutti gli attributi utilizzati nella dimensione.
  - Insieme di unioni: Insieme di tutte le unioni necessarie ad unire tutti gli attributi specificati. Vengono specificate solo le unioni necessarie per unire le tabelle di dimensione.
  - Insieme di gerarchie L'insieme di gerarchie di una dimensione.

Pierluigi Del Nostro



## ► Interrogazione DW » Modello cubo

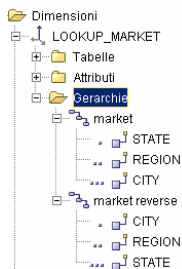


Pierluigi Del Nostro



## ► Interrogazione DW » Modello cubo » Gerarchie

- ❑ Una gerarchia definisce le relazioni tra un insieme di uno o più attributi in una dimensione del modello cubo.
- ❑ Le gerarchie forniscono diverse modalità di navigazione e di calcolo per analizzare le dimensioni.
- ❑ Per una dimensione di un modello cubo è possibile definire più gerarchie.

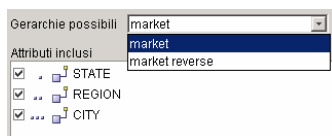


Pierluigi Del Nostro



## ► Interrogazione DW » Modello cubo » Cubo

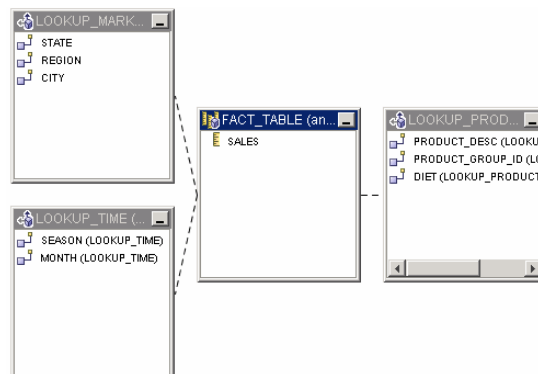
- derivato dal modello cubo
- un cubo è la definizione precisa di un cubo OLAP.
- contiene un sottoinsieme di oggetti appartenente al modello cubo.
- consente di utilizzare solo una gerarchia per dimensione.
- può essere utilizzato da altri strumenti OLAP che possono gestire le viste singole di una specifica dimensione (gerarchia singola).



Pierluigi Del Nostro



## ► Interrogazione DW » Modello cubo » Cubo

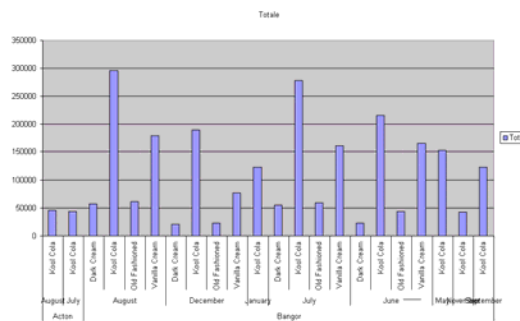


Pierluigi Del Nostro



## Office Connect Analytic Edition

scenario	All		
CITY	MONTH	PRODUCT_DESC	SALES
Acton	August	Kool Cola	46368
		July	43632
Bangor	August	Dark Cream	57526
		Kool Cola	295673
		Old fashioned	62160
		Vanilla Cream	178149
	December	Dark Cream	21384
		Kool Cola	189506
		Old fashioned	22896
	Vanilla Cream	77013	
	January	Kool Cola	122439
		July	Dark Cream
		Kool Cola	278399
		Old fashioned	59784
		Vanilla Cream	160519



Pierluigi Del Nostro



## Pubblicazione dei metadati

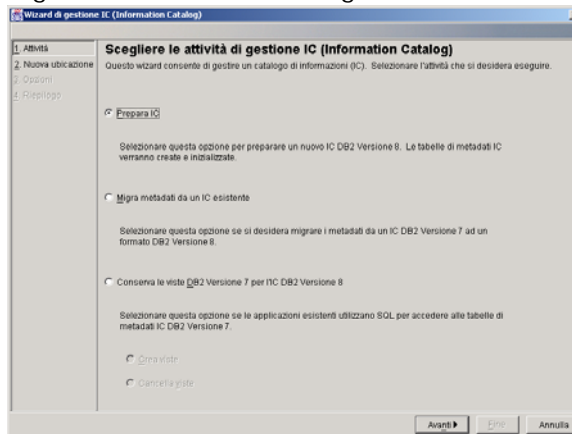
- Trasferimento dei metadati presenti nel data warehouse in un Information Catalog
- Information Catalog: insieme di tabelle gestite dall'Information Catalog Center
  - Metadati di business utili all'utente finale per identificare e localizzare le informazioni e i dati presenti nella propria organizzazione
  - ES: Ricerca della tabella che contiene i dati su cui effettuare un'interrogazione

Pierluigi Del Nostro



## ► Pubblicazione dei metadati

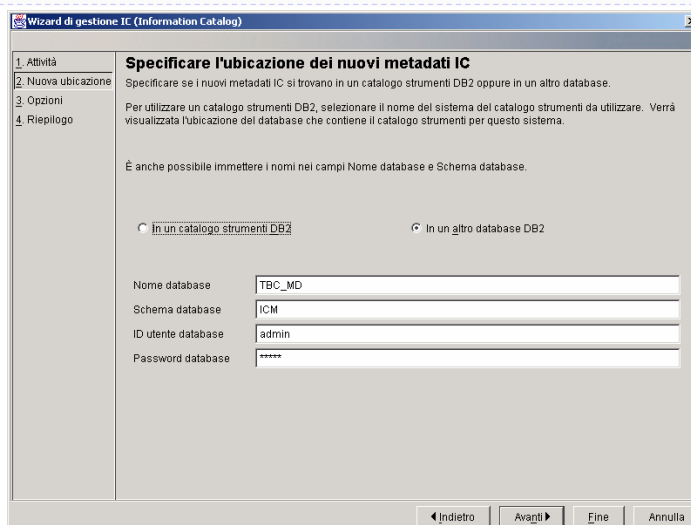
- Il primo passo è creare un catalogo:
  - Start —> Programmi —> IBM DB2 —> Strumenti di configurazione —> Wizard di gestione IC



Pierluigi Del Nostro



## ► Pubblicazione dei metadati

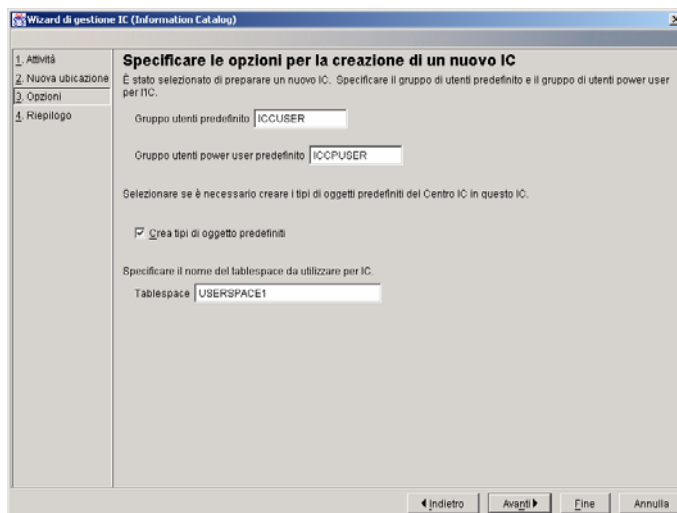


Pierluigi Del Nostro





## ► Pubblicazione dei metadati

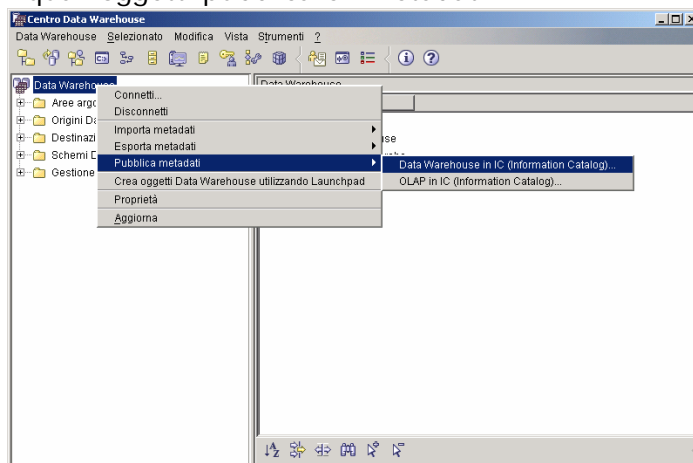


Pierluigi Del Nostro



## ► Pubblicazione dei metadati

- ❑ Dopo aver creato il catalogo metadati è possibile scegliere per quali oggetti pubblicare i metadati.



Pierluigi Del Nostro

**Data Warehousing** Università Roma Tre  
Dipartimento di Informatica e Automazione

### ► Pubblicazione dei metadati

**Proprietà - Published Tutorial Metadata**

Nome catalogo: TBC\_MD  
 Nome schema: ICM  
 ID utente del responsabile: admin  
 Password del responsabile: \*\*\*\*\*  
 Conferma password: \*\*\*\*\*

**Oggetti disponibili**

- Area argomento
- Origini Data Warehouse
- Destinazioni Data Warehouse
- Schemi Data Warehouse

**Oggetti selezionati**

- Area argomento
- TBC Tutorial
  - Build Tutorial Fact Tabl
  - Build Tutorial Market Di
  - Build Tutorial Product C
  - Build Tutorial Scenario
  - Build Tutorial Time Dim

Pierluigi Del Nostro

**Data Warehousing** Università Roma Tre  
Dipartimento di Informatica e Automazione

### ► Information Catalogue

**Centra IC (Information Catalog)**

Nome	Proprietario	Ultimo aggiornamento	Creazione	Autore
Demographics Target	ADMIN	23 maggio 2005 6.54.11	21 maggio 2004 10.32.44	ADMIN
IMHFACT_TABLE	ADMIN	23 maggio 2005 6.54.08	21 maggio 2004 10.32.42	ADMIN
IMH GEOGRAPHIES_TARGET	ADMIN	23 maggio 2005 6.54.11	21 maggio 2004 10.32.44	ADMIN
IMHLOOKUP_MARKET	ADMIN	23 maggio 2005 6.54.11	21 maggio 2004 10.32.44	ADMIN
IMHLOOKUP_PRODUCT	ADMIN	23 maggio 2005 6.54.11	21 maggio 2004 10.32.48	ADMIN
IMHLOOKUP_SCENARIO	ADMIN	23 maggio 2005 6.54.12	21 maggio 2004 10.32.49	ADMIN
IMHLOOKUP_TIME	ADMIN	23 maggio 2005 6.54.12	21 maggio 2004 10.32.49	ADMIN
IMH PRODUCT	ADMIN	23 maggio 2005 6.54.11	23 maggio 2005 6.54.05	ADMIN
SAMPLTBC GEOGRAPHIES	ADMIN	23 maggio 2005 6.54.10	21 maggio 2004 10.32.46	ADMIN
SAMPLTBC INVENTORY	ADMIN	23 maggio 2005 6.54.07	21 maggio 2004 10.32.42	ADMIN
SAMPLTBC PRODUCT	ADMIN	21 maggio 2004 11.09.02	21 maggio 2004 10.32.48	ADMIN
SAMPLTBC PRODUCTION_COSTS	ADMIN	23 maggio 2005 6.54.07	21 maggio 2004 10.32.42	ADMIN
SAMPLTBC SALES	ADMIN	23 maggio 2005 6.54.07	21 maggio 2004 10.32.42	ADMIN
SAMPLTBC SCENARIO	ADMIN	23 maggio 2005 6.54.12	21 maggio 2004 10.32.49	ADMIN
SAMPLTBC TIME	ADMIN	23 maggio 2005 6.54.12	21 maggio 2004 10.32.49	ADMIN

Pierluigi Del Nostro



► Fine