

Business Intelligence per i Big Data

Esercitazione di Laboratorio N. 3

L'esercitazione consiste nel progettare un data warehouse che permetta di gestire la problematica illustrata nei punti seguenti, e costruire di esso opportuni cubi, report e dashboard mediante l'utilizzo della suite Pentaho Community Edition e dei relativi plug-in.

1. Descrizione del problema

Una società che gestisce degli stabilimenti balneari italiani vuole analizzare gli incassi dei suoi stabilimenti nel corso degli anni. Attualmente ogni stabilimento ha una propria base di dati in cui memorizza giornalmente le informazioni relative agli oggetti affittati (ombrelloni, lettini, pedalò, ecc.) e il prezzo a cui sono stati affittati. Lo schema concettuale della base di dati di uno stabilimento balneare (sorgente dati Stab1) è mostrato in Figura 1, mentre lo schema logico è mostrato in Figura 2.

La società vuole aggregare le informazioni provenienti dagli stabilimenti in un proprio data warehouse centralizzato. Il data warehouse deve essere progettato in modo tale da permettere di rispondere velocemente ad alcune interrogazioni effettuate frequentemente dai dirigenti della società.

Alcune delle informazioni a cui i dirigenti della società sono interessati sono le seguenti:

- incassi effettuati su base mensile in ogni provincia e in ogni regione
- incassi effettuati su base annua in ogni provincia e in ogni regione
- incassi effettuati su base trimestrale in ogni provincia e in ogni regione
- incassi effettuati su base quadrimestrale in ogni provincia e in ogni regione
- incassi effettuati su base mensile, trimestrale, quadrimestrale, annuale per ogni categoria di oggetti in funzione della provincia e della regione
- numero di oggetti affittati su base mensile, trimestrale, quadrimestrale, annuale in funzione della categoria di oggetto
- numero di oggetti affittati su base mensile, trimestrale, quadrimestrale, annuale in funzione della provincia e della regione

Ai fini dell'esercitazione viene fornita la soluzione del data warehouse. La versione client-server di Pentaho Community Edition disponibile in laboratorio si appoggia al DBMS Oracle XE 11g R2. Per permettere lo svolgimento dell'esercitazione il caricamento delle tabelle e dei relativi dati è già stato effettuato. Il materiale relativo alla progettazione del data warehouse e gli script per il caricamento dei dati (utili solo nel caso in cui si intenda svolgere l'esercitazione sul proprio PC personale) sono disponibili sul sito del corso <http://dbdmg.polito.it/wordpress/teaching/business-intelligence/>

Durante l'esercitazione dovranno essere eseguite le seguenti attività:

- Creazione di cubi
- Creazione di report
- Creazione di dashboard

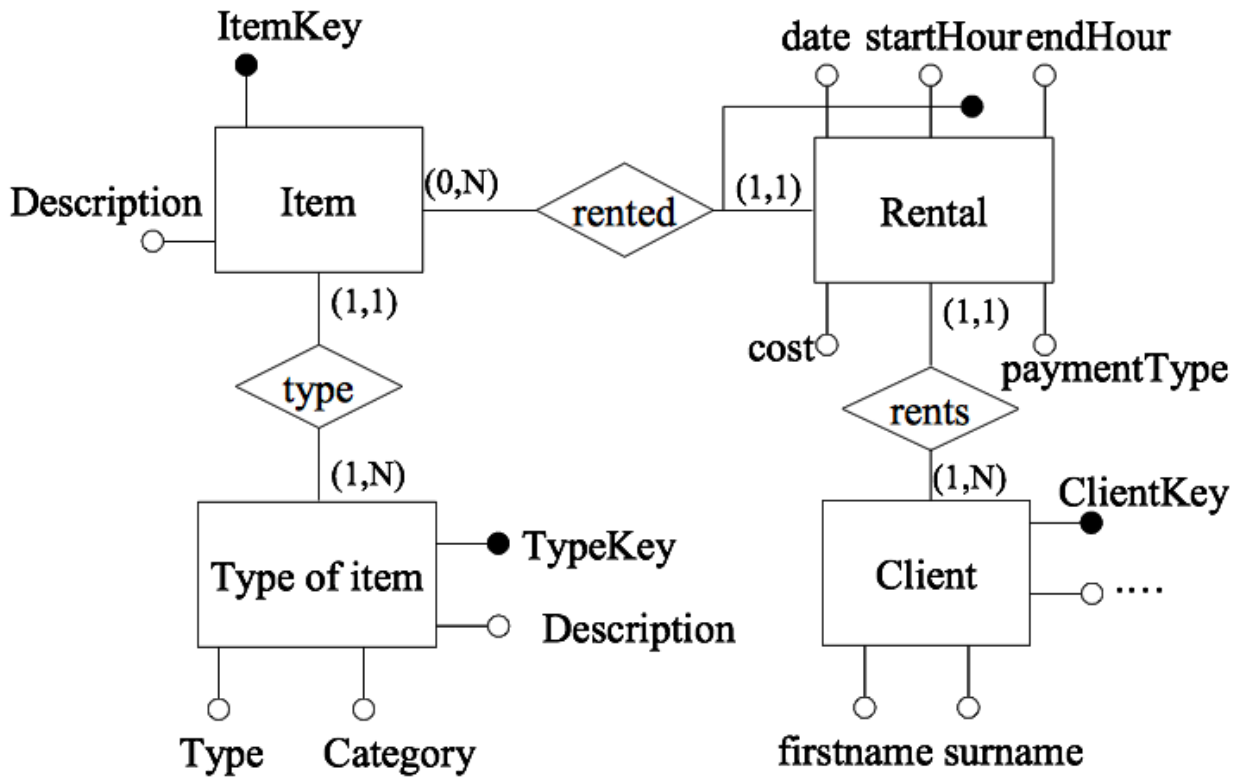


Figura 1. Modello concettuale della sorgente Stab1

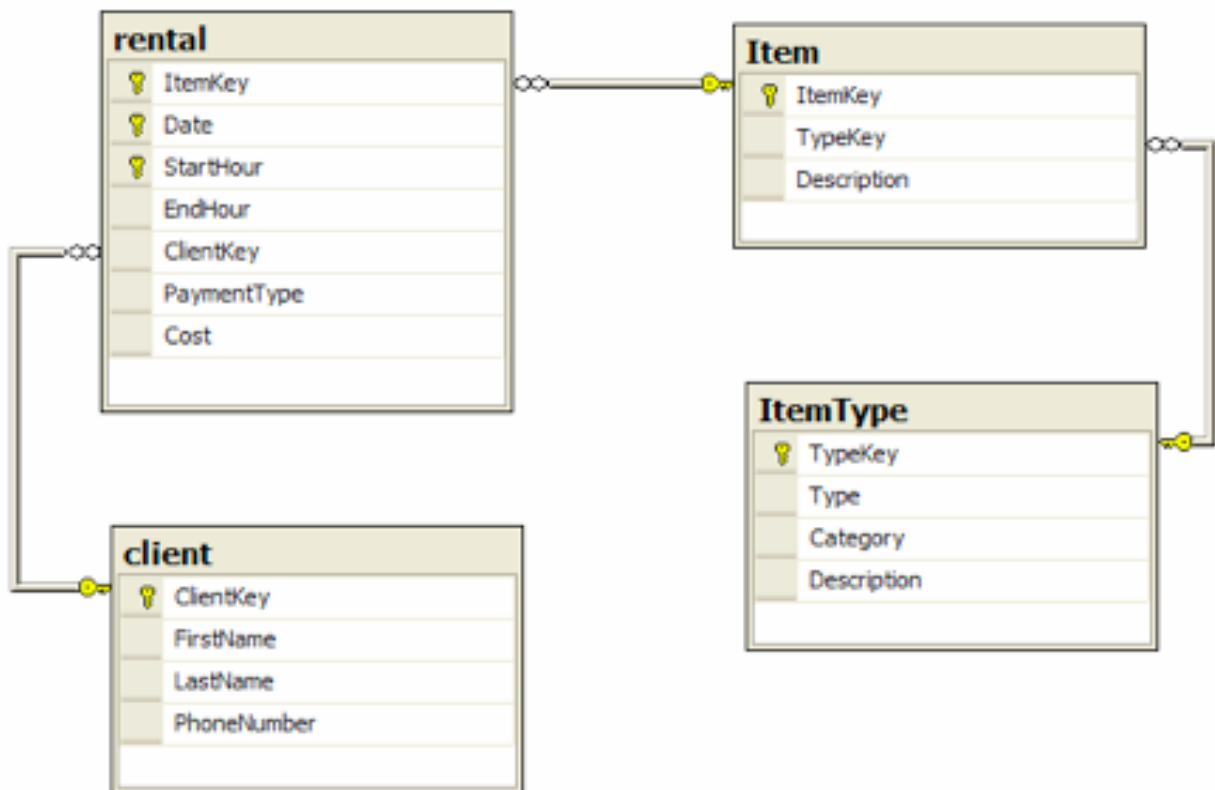
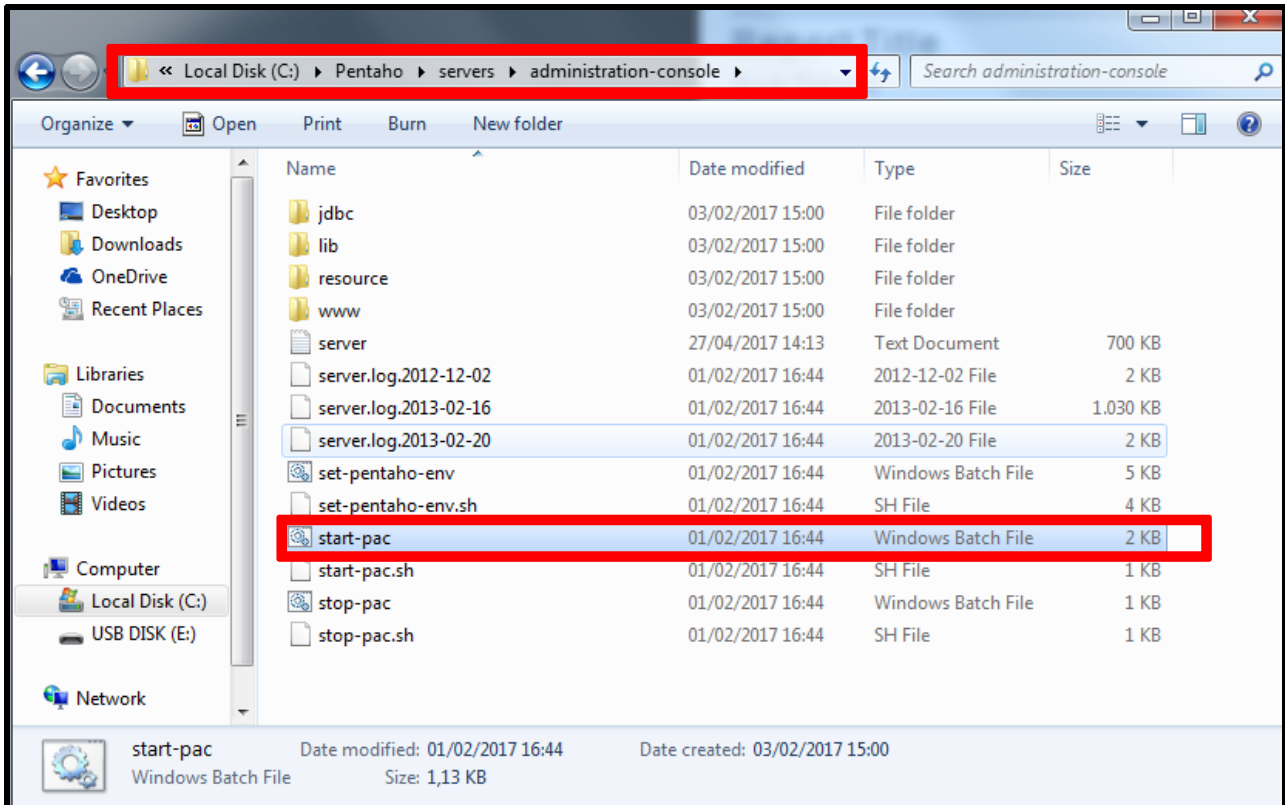


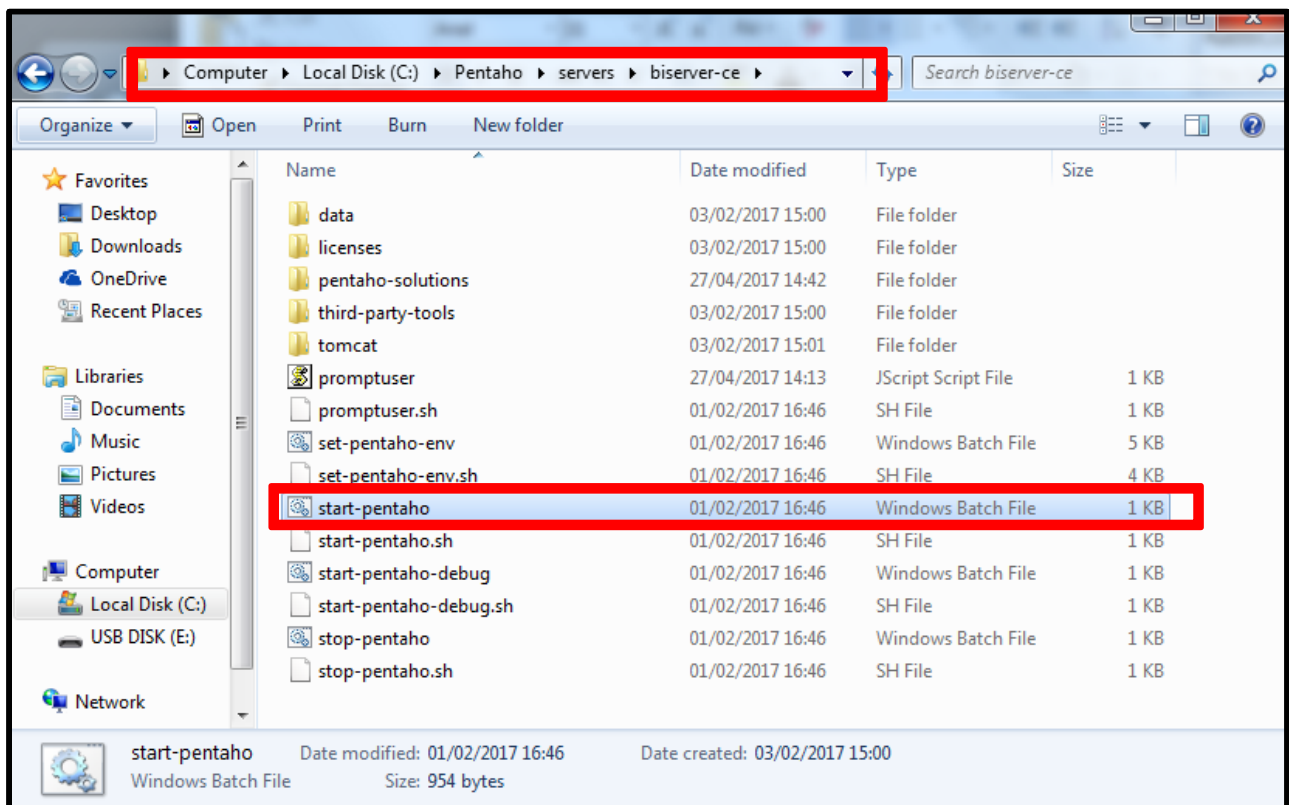
Figura 2 – Schema logico della sorgente Stab1

2. Passi Preliminari

- Posizionarsi nella cartella **C:\Pentaho\servers\administration-console**
- Lanciare l'applicazione **start-pac** (lasciare aperta la schermata per tutta la durata dell'esercitazione)



- Posizionarsi nella cartella **C:\Pentaho\servers\biserver-ce**
- Lanciare l'applicazione **start-pentaho** (lasciare aperta la schermata per tutta la durata dell'esercitazione)



3. Creazione delle tabelle e caricamento dei dati

Per il caricamento delle tabelle del data warehouse e dei relativi dati è disponibile uno script sul portale del corso. Eseguire i passi seguenti:

- **Scaricare** gli script per il caricamento del data warehouse
- **Scomattare** il file **SeasideResortDW**
- **Lanciare** il file **“Load.bat”**

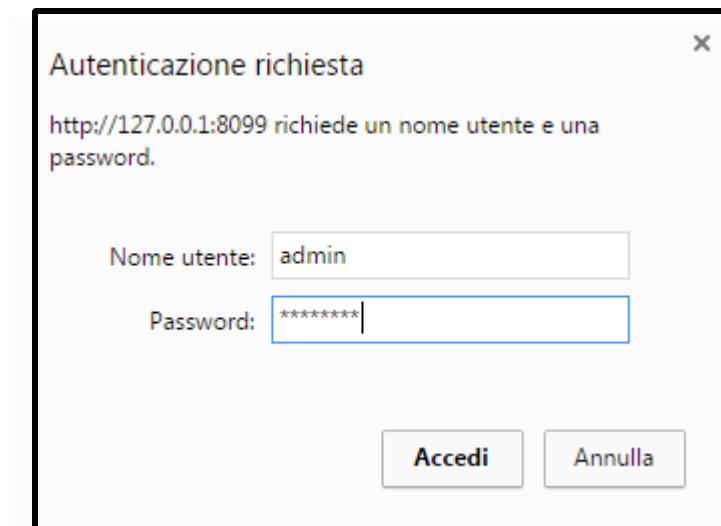
4. Connessione di Pentaho al database Oracle

Per connettere Pentaho alla base di dati in Oracle XE è necessario eseguire i seguenti passi:
Lanciare la Pentaho BI Administration Console tramite un browser Web (Mozilla Firefox o Google Chrome) collegandosi all'indirizzo <http://127.0.0.1:8099>

Login -->

Nome utente: admin

Password: password



Autenticazione richiesta

http://127.0.0.1:8099 richiede un nome utente e una password.

Nome utente:

Password:

Selezionare dal pannello di sinistra *Administration*

Selezionare *Database Connection* dal pannello

Selezionare il simbolo + dal pannello



Creare una *nuova connessione*:

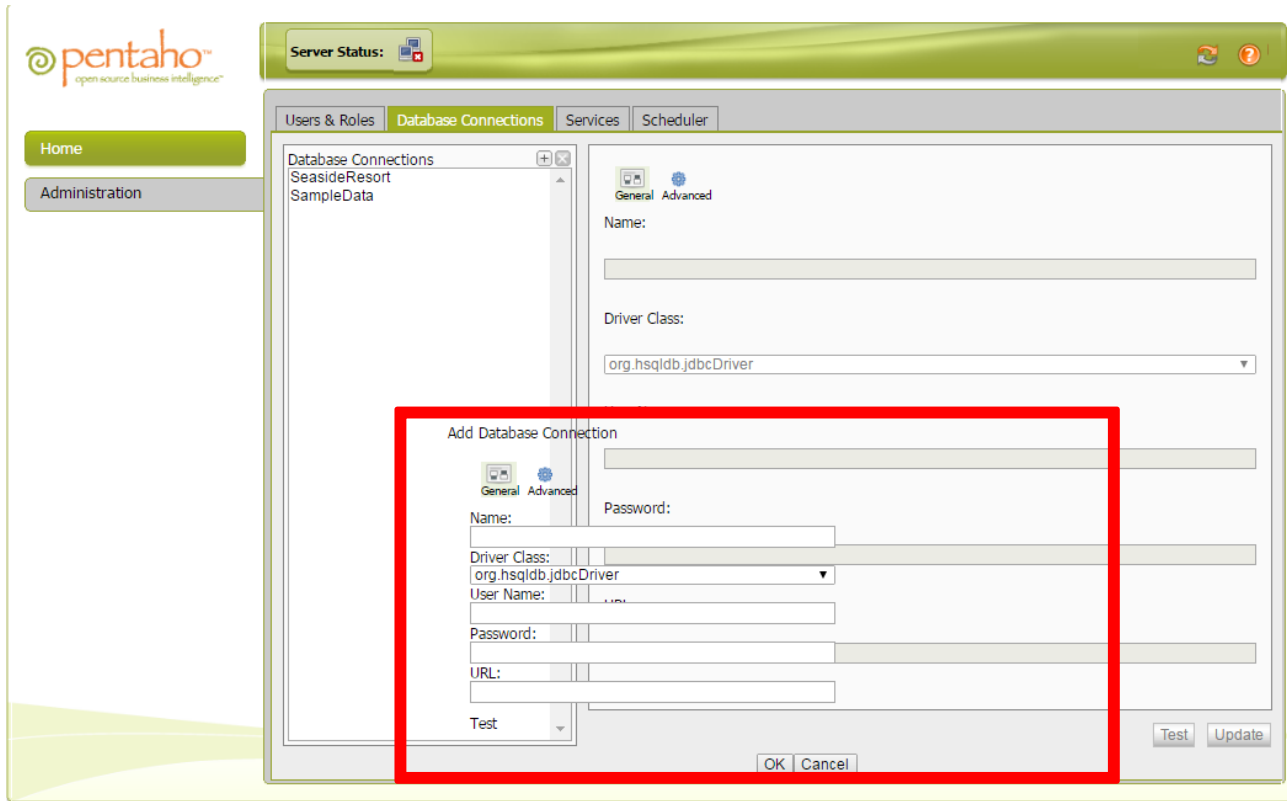
Name: SeasideResort

Driver Class: Oracle.Jdbc.Driver.OracleDriver

Username: seasideDW

Password: password

Url: jdbc:oracle:thin:@localhost:1521:xe



Verificare la connessione cliccando **Test**. Premere **OK** e successivamente **Update**.

2. Creazione di cubi

Partendo dal data warehouse già creato e alimentato e tenendo in considerazione le query più frequentemente eseguite:

- creare uno o più cubi
- pubblicare i cubi creati su Pentaho BI User Console
- visualizzare e analizzare i cubi

Sono di seguito elencate alcune delle interrogazioni che vengono frequentemente eseguite:

- Calcolare per ogni coppia **provincia, trimestre** l'**incasso totale**. Calcolare inoltre l'incasso totale per ogni provincia e l'incasso totale per ogni trimestre.
- Calcolare l'**incasso totale** per ogni coppia **categoria oggetto affittato, mese**. Calcolare inoltre l'incasso totale per categoria di oggetto (indipendentemente dal mese) e l'incasso totale per ogni mese (indipendentemente dalla categoria di oggetti affittati).

Durante l'esercitazione dovranno essere eseguite le seguenti attività:

- Utilizzare Pentaho schema workbench per costruire i cubi
- Utilizzare Pentaho BI User Console e Saiku Analytics plug-in per visualizzare ed interagire con i cubi

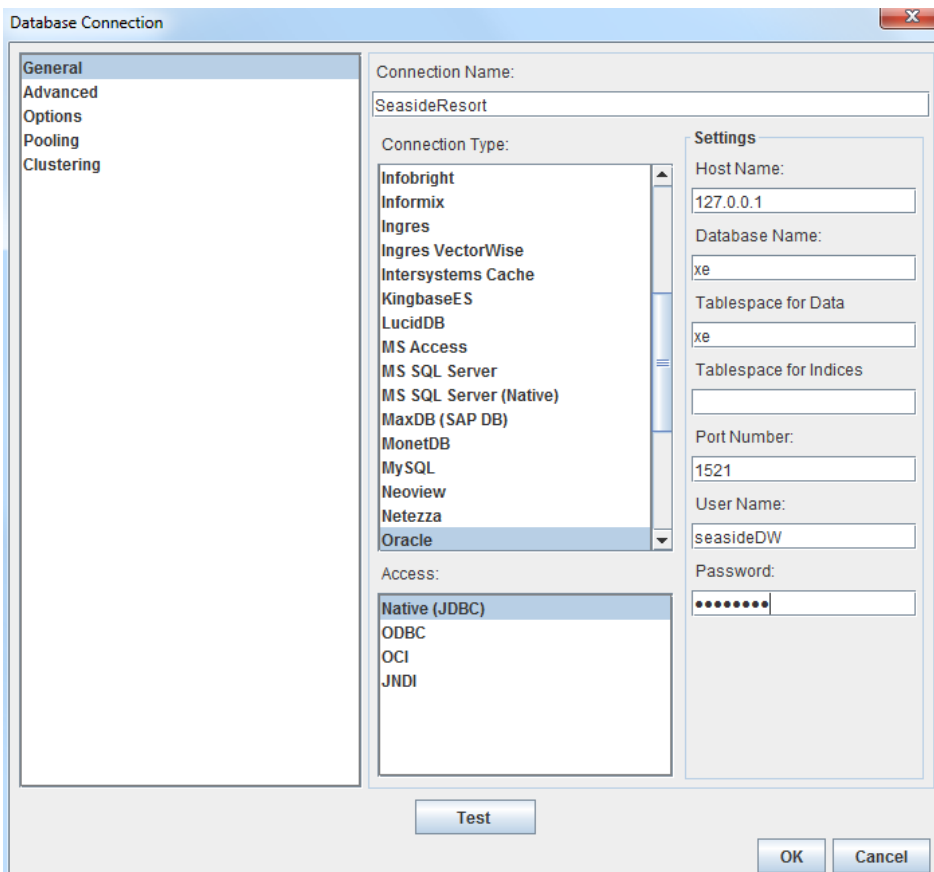
2.1 Istruzioni

Per la *creazione e pubblicazione dei cubi* in Pentaho Community Edition si utilizza il client **Schema WorkBench**.

Passi per la creazione dei cubi in Schema WorkBench:

- Lanciare l'applicazione Java SchemaWorkBench
 - *C:\Pentaho\clients\schema-workbench\workbench.bat*
- Stabilire una connessione con il data warehouse
 - Selezionare *Options -> Connections* dal menu a tendina
 - Inserire i seguenti dati nel pannello *General*
 - Connection name: *SeasideResort*
 - Connection type: *Oracle*
 - Access: *Native (JDBC)*
 - Hostname: *127.0.0.1*
 - Database Name: *xe*
 - Tablespace for Data: *xe*
 - Port number: *1521*
 - Username: *seasideDW*
 - Password: *password*

Selezionare **Test** e poi **Ok**.



- Creare un nuovo schema
 - Selezionare *File* -> *New* -> *Schema* dal menu a tendina
 - Schema name: *SeasideSchema[numero di matricola dello studente]* (ad es. *SeasideSchema190000* per lo studente con matricola 190000)
 - Creare un nuovo cubo
 - Pannello di sinistra -> click tasto destro sull'icona *Schema* -> *Add cube*
 - Cube name: *SeasideCube*
 - Creare un collegamento alla tabella dei fatti
 - Pannello di sinistra -> click tasto destro sull'icona *SeasideCube* -> *Add Table*
 - Schema: selezionare *SeasideDW* dal menu a tendina
 - Table name: *RENTAL*
 - Creare le dimensioni
 - Pannello di sinistra -> click tasto destro sull'icona *SeasideCube* -> *Add Dimension*
 - Esempio di dimensione temporale
 - Name: *TimeDim*
 - Foreign key (dalla fact table): *COD_T*
 - Type: *TimeDimension*
 - Creare le gerarchie sulle dimensioni
 - Pannello di sinistra -> click tasto destro sull'icona rappresentante la dimensione appena creata (*TimeDim* nell'esempio precedente) -> *Add Hierarchy*
 - Esempio di gerarchia per una dimensione temporale
 - Name: *TimeDimHier*
 - Primary key (della tabella della dimensione (*primaryKey*)): *COD_T***
- ** NB: visibile dopo aver creato la tabella sotto**

- Creare le tabelle delle dimensioni
- Pannello di sinistra -> click tasto destro sull'icona rappresentante la gerarchia appena creata (*TimeDimHier* nell'esempio precedente) -> *Add Table*
- Esempio di tabella per una dimensione temporale
 - Schema: *SeasideDW*
 - Name: *TIMEDIM*
- Creare i livelli della gerarchia
- Pannello di sinistra -> click tasto destro sull'icona rappresentante la gerarchia appena creata (*TimeDimHier* nell'esempio precedente) -> *Add Level*
- Esempio di livello per una dimensione temporale
 - Name: *YEAR*
 - Column: *YEAR*
 - nameColumn: *YEAR*
 - Type: *String*
 - levelType: *TimeYears*
 - HideMemberIf: *Never*
- N.B. Creare i livelli in ordine *decrescente* di granularità (ad es., prima *YEAR*, poi *QUARTER*)*
- Creare le misure
- Pannello di sinistra -> click tasto destro sull'icona *SeasideCube* -> *Add Measure*
- Esempio di misura
 - Name: *NetIncomeMeasure*
 - Aggregator: *sum*
 - Column: *NETINCOME*
 - Datatype: *Integer*

***NOTA: Creare istanze multiple per la dimensione temporale aventi UNA sola gerarchia ciascuna anziché una singola dimensione con gerarchie multiple.**

Il cubo finale avrà una struttura analoga a quella rappresentata in Figura 1.

POTETE SCARICARE DAL SITO IL CUBO COMPLETO *SchemaPracticeSolution.xml* E CARICARLO DA FILE – OPEN.

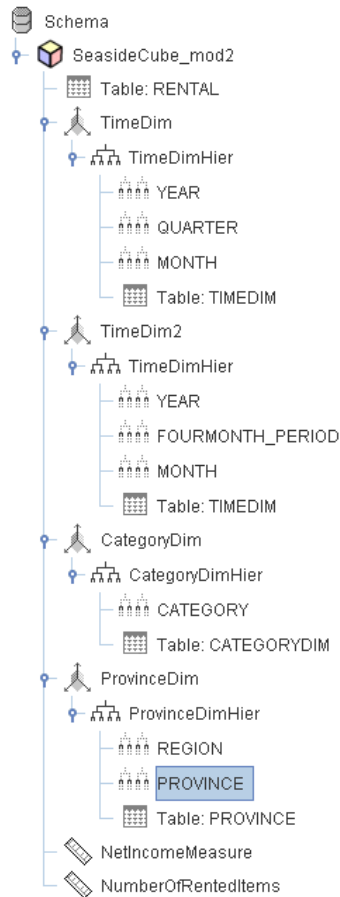


Figura 1. Struttura finale del cubo
 **NB: *NumberOfRentedItems* è di tipo sum

Attribute	Value
name	NumberOfRentedItems
description	
aggregator	sum
column	NUMBEROFRENTEDITEMS
formatString	
datatype	Integer
formatter	
caption	
visible	<input checked="" type="checkbox"/>

Per la pubblicazione del cubo selezionare *File -> Publish* dal menu a tendina

- URL: <http://127.0.0.1:8090/pentaho/>
- Publish password: *publish*
- Pentaho credential
 - Nome utente: *joe*
 - Password: *password*

Repository Login

Server

URL:

Publish Password:

Pentaho Credentials

User:

Password:

Remember these Settings

OK Cancel

- Creare una nuova cartella: *joe[numero matricola dello studente]*
- Spostarsi nella cartella appena creata *joe[numero matricola dello studente]*
- Cliccare *OK* ed entrare nella cartella appena creata
- Pentaho or JNDI Data Source: *SeasideResort*
- Publish

Publish Schema

Schema Name:
SeasideSchemaPracticeSolution

Schema File:
SchemaPracticeSolution.xml

Location:

Name	Type	Date Modified
------	------	---------------

Publish Settings

Pentaho or JNDI Data Source:

Register XMLA Data Source

Publish Cancel

Per la visualizzazione dei cubi si utilizza invece la *Pentaho BI User Console*

- Per lanciare la Pentaho BI User Console aprire un browser Web (Mozilla Firefox o Google Chrome) e collegarsi all'indirizzo <http://127.0.0.1:8090>

- Login
 - Nome utente: *joe*
 - Password: *password*

Se si presenta un warning nel portale, una volta completato l'inserimento delle credenziali, ignoratelo e proseguite nell'esercitazione.

- New Analysis
 - Schema: *SeasideSchema[numero di matricola dello studente]*
 - Cube: *SeasideCube*
 - Provate a navigare le dimensioni per analizzare le dimensioni del cubo creato.

3. Creazione di report

Utilizzando il data warehouse già creato e alimentato:

- Creare un report che visualizzi per ciascun **anno** una tabella con l'**incasso totale** di ogni **mese** per ciascuna **categoria di oggetto affittato**
- Pubblicare i report realizzati su Pentaho BI User Console
- Visualizzare i report ed esportarli in formato PDF

Durante l'esercitazione dovranno essere eseguite le seguenti attività:

- *Utilizzare Pentaho report designer per realizzare i report*
- *Utilizzare Pentaho BI User Console per visualizzare ed esportare i report*

3.1 Istruzioni

Per la creazione e pubblicazione dei report in Pentaho Community Edition si utilizza il client *Report Designer*

Creazione dei report mediante il Report Designer Wizard:

- Lanciare Pentaho Report Designer
 - *C:\Pentaho\clients\report-designer\report-designer.bat*
- *Ignorare la pagina di welcome chiudendola e cliccare "File->Report Wizard..."*

Passi per completare la creazione mediante il wizard:

STEP "Look and feel"

- Selezionare un template pre-esistente
- Next

STEP "Data source"

- Creare una nuova Data Source per il report
 - Click sull'icona "+" in alto a destra nel pannello *Data Source*
 - Type: *JDBC*

- Creare una nuova connessione
 - Connection name: *SeasideResort*
 - Connection type: *Oracle*
 - Access: *Native (JDBC)*
 - Hostname: *127.0.0.1*
 - Database Name: *xe*
 - Tablespace for Data: *xe*
 - Port number: *1521*
 - Username: *seasideDW*
 - Password: *password*
- Premere **Test** per verificare di avere inserito correttamente i campi, quindi premere **OK**
 - Creare una nuova query nel pannello *Available queries* premendo il "+" verde
 - Esempio di query
 - Query name: *Query1*
 - Esempio di query di prova: *SELECT * FROM RENTAL*
 - N.B. nel codice SQL non va inserito il ";" finale
 - Cliccare su *preview* per visualizzare l'anteprima del risultato
 - Ok e next

STEP "Layout"

- Definizione del layout base di visualizzazione
- Come raggruppare (*Group Items by*) (se necessario)
- Che attributi visualizzare per ciascun gruppo (*Selected Items*)
 - Ad es. *NETINCOME*
- Cliccare su *preview* per visualizzare l'anteprima del risultato

STEP "Format"

- Definizione del formato e del tipo di aggregazione
- Example of details about *NETINCOME*
 - Display name: *NETINCOME*
 - Date format: *NONE*
 - Aggregation: *SUM*
- Cliccare su *preview* per visualizzare l'anteprima del risultato

Terminata la creazione guidata è possibile personalizzare il report manualmente mediante l'interfaccia grafica. Un esempio di risultato è riportato in Figura 2.

Per la pubblicazione del report selezionare *File -> Publish* dal menu a tendina. Vi chiederà di **salvare il file**, salvatelo in locale (va bene anche il desktop).

- URL: <http://127.0.0.1:8090/pentaho/>
- Timeout: 30
- Pentaho credential
 - Nome utente: *joe*
 - Password: *password*
- Location: *joe [matricola studente]*
- Publish password: *publish*
- Output type: *PDF*

A questo punto è possibile accedere al report pubblicato mediante la *Pentaho BI User Console*. In alternativa, cliccando su *File -> Export* dal menu a tendina è possibile esportare il report in diversi formati (ad es., PDF, HTML).

October 30, 2013 @ 12:04

Report Title
Sub Title 1
Sub Title 2

NETINCOME	
	\$70,500.00
	\$39,750.00
	\$9,150.00
	\$63,600.00
	\$70,500.00
	\$39,750.00
	\$9,150.00
	\$63,600.00
	\$89,300.00
	\$50,350.00
	\$11,590.00
	\$80,560.00
	\$70,500.00
	\$39,750.00
	\$9,150.00
	\$63,600.00
	\$70,500.00
	\$39,750.00
	\$9,150.00
	\$63,600.00
	\$89,300.00
	\$50,350.00
	\$11,590.00
	\$80,560.00
	\$131,600.00
	\$74,200.00
	\$17,080.00
	\$118,720.00
	\$70,500.00
\$7,320,000.00	

Wed Oct 30 12:04:10 CET 2013

Figura 2. Esempio di Report Preview

File Name:
Prova

Title:
prova

Report Description:
[Empty text area]

Location:
/joe123456

Title	File Name	Date Modified	Description
[Empty table body]			

Output Type:
[Dropdown menu] Lock

Publish Password:
[Masked password field]

OK Cancel

4. Creazione di dashboard

Utilizzando il data warehouse già creato ed alimentato:

- Creare una dashboard che consenta di selezionare un **anno** di attività dello stabilimento balneare e per l'anno selezionato visualizzare un grafico a barre con il **numero di oggetti affittati** per ciascuna **categoria di oggetto**
 - Durante l'esercitazione dovranno essere eseguite le seguenti attività:
 - Utilizzare *Community Dashboard Editor (CDE)* per realizzare la dashboard
 - Utilizzare *Pentaho BI User Console* per visualizzare la dashboard

4.1 Istruzioni

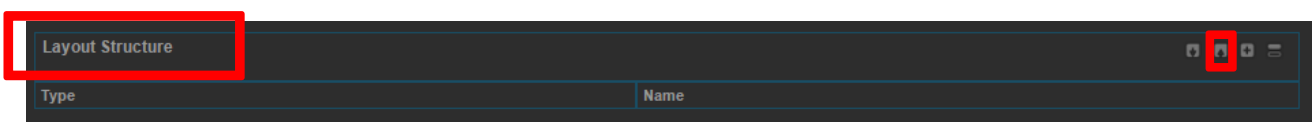
Per la creazione delle dashboard utilizzare il *Community Dashboard Editor (CDE)* integrato all'interno della *Pentaho BI User Console*.

Passi per la creazione delle dashboard mediante CDE:

- Lanciare la *Pentaho BI User Console* aprendo un browser Web (Mozilla Firefox o Google Chrome) e collegandosi all'indirizzo <http://127.0.0.1:8090>
 - Login
 - Nome utente: *joe*
 - Password: *password*
- Lanciare il *Community Dashboard Editor* cliccando sull'icona "CDE" in alto a destra (in alternativa, selezionare *File -> New -> CDE DashBoard*)



- Selezionare uno dei layout pre-esistenti (ad es., *Two Columns Template*) per la dashboard cliccando sull'icona *Apply Template* del pannello *Layout Structure*



- Personalizzare il layout modificando i relativi tag (N.B. righe e colonne della dashboard sono strutturate in modo gerarchico)
- Creare un menu a tendina mediante il quale è possibile settare un parametro

- Selezionare Data Source Wizards -> OLAP Selector Wizard
 - Trascinare i livelli come in Figura 3
 - Assegnare un nome al selettore
 - Selezionare l'elemento di layout ove il selettore verrà visualizzato (ad es., *Html object Panel_2*)
 - Premere Ok

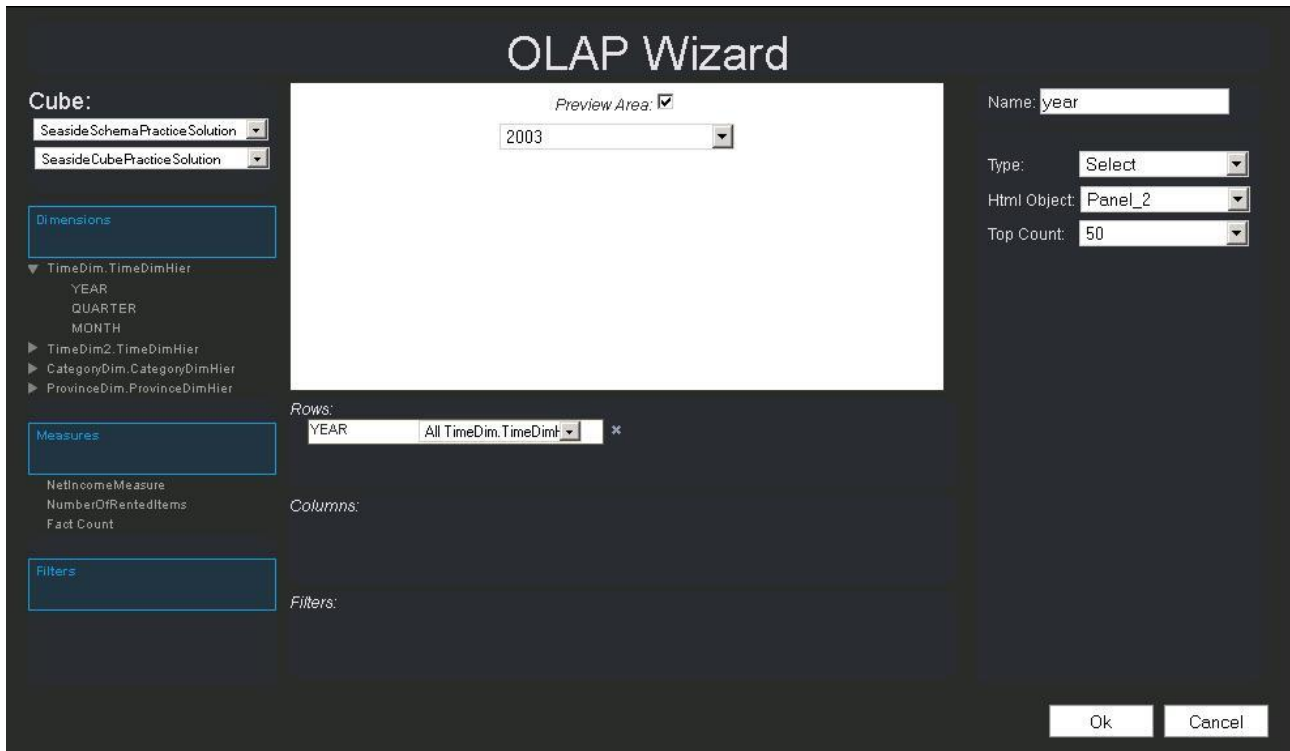


Figura 3. Esempio di selettore per Year

- Creare la data source per il grafico
- Selezionare Data Source Wizards -> OLAP Chart Wizard
 - Trascinare livelli, misure e filtri come in Figura 4
 - Assegnare nome e titolo al grafico
 - Selezionare il tipo di grafico (ad es., *Bar Chart*)
 - Selezionare l'elemento di layout ove il grafico verrà visualizzato (e.g., *Html object Panel_1*)

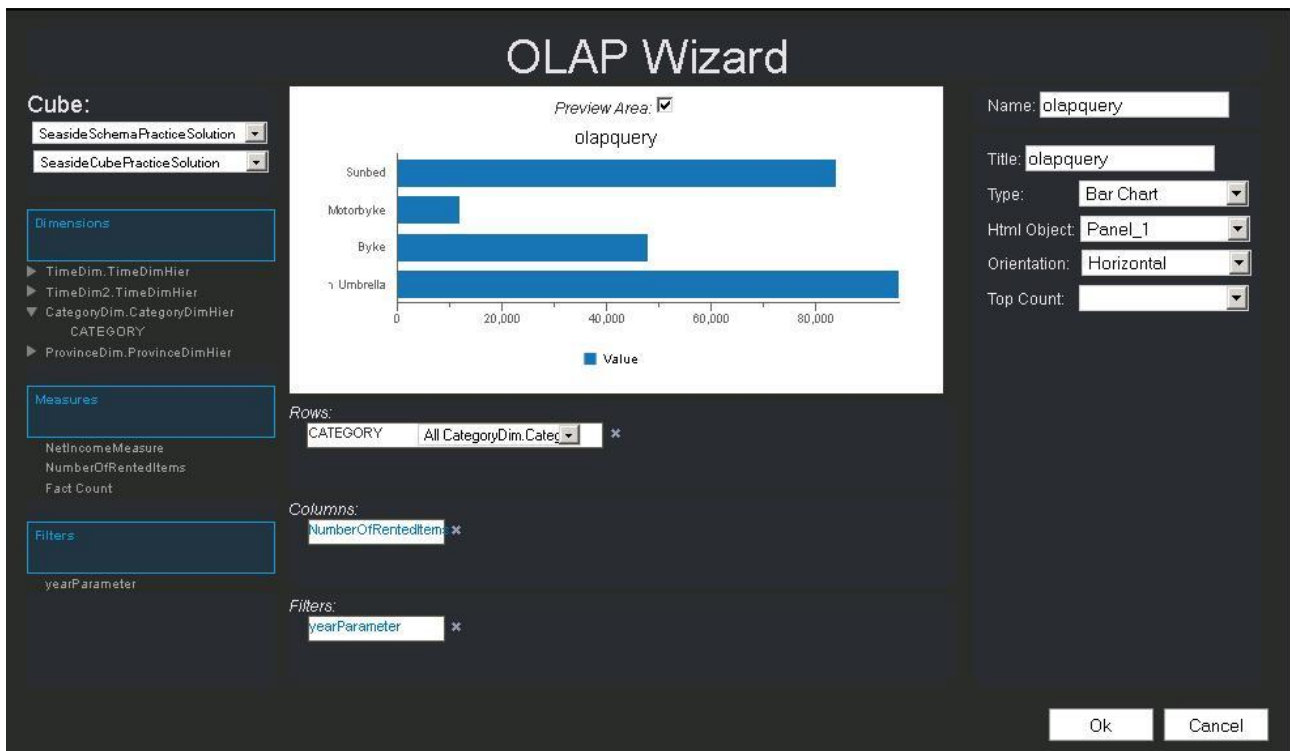


Figure 4. Definizione della Data Source per il grafico

- Verificare che il parametro JINDI delle Data Source sia settato a *SeasideResort*
 - Premere **Ok**. Salvare il risultato nella propria cartella.
 - Selezionare **Preview**
- Un esempio di risultato finale dopo la preview è riportato in Figura 5.

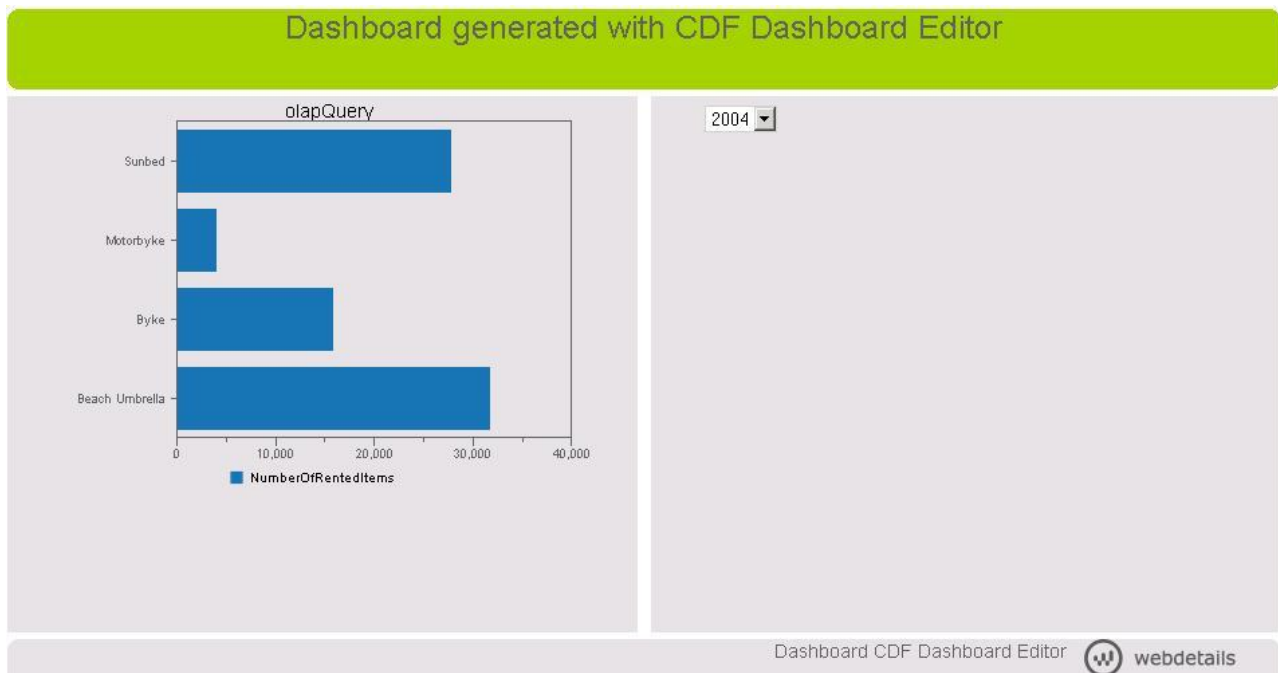


Figure 5. Esempio di Dashboard con Grafico a barre dinamico e selettore con menu a tendina