

Comparativa de rendiment OLAP per a diferents motors de bases de dades orientades a columnes

Autor: Joan Piferrer Huerta

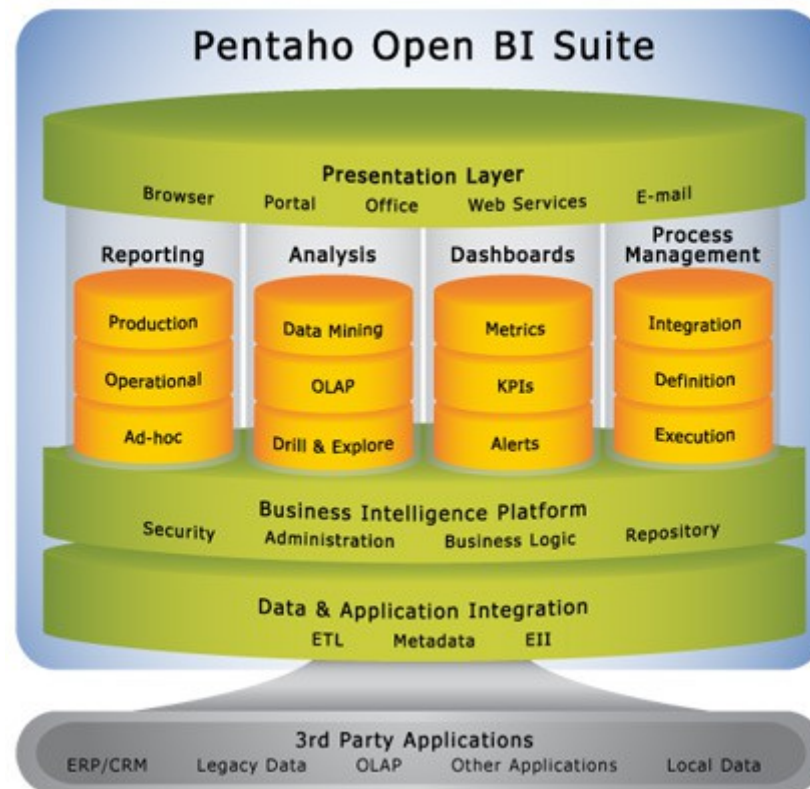
Consultor: Gregorio Robles Martínez

Consultor extern: Juan José Ortilles

Barcelona, Juny 2013



Realitzar una comparativa de rendiment entre els diferents models de bases de dades orientades a columnes mitjançant la construcció i explotació d'un cub OLAP utilitzant la suite de BI Pentaho.



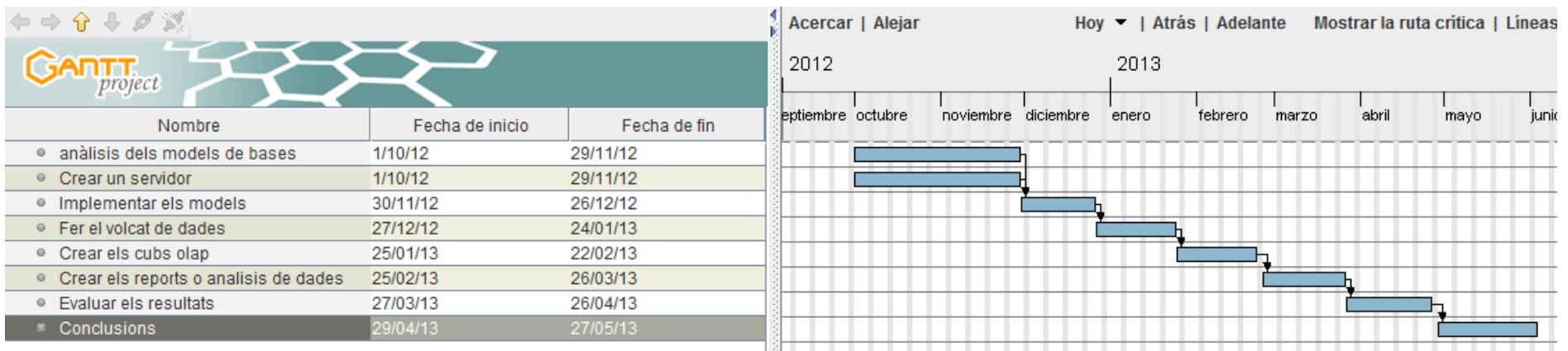
Aquest servidor estarà connectat a un servidor Mysql, que es un sistema gestor de bases de dades relacionals allotjat a la mateixa màquina.

Albergarà les dues bases de dades que utilitzarem per la comparativa.



- Instal·lar i configurar un servidor pentaho i Mysql en una màquina.
- Crear les dues bases de dades a comparar.
- Dissenyar els respectius cubs de dimensions
- Extreure la informació necessària per extreure les conclusions



- Fer un anàlisi dels models de bases de dades que s'utilitzaran en la comparativa.
- Crear un servidor i instal·lar un mysql i un pentaho biserver
- Implementar els models que s'han de testejar en una base de dades de mysql
- Fer el bolcat de dades en cadascun dels models.
- Crear els cubs olap que es connectaran als diferents models
- Crear els reports o anàlisis de dades on es presentaran les dades filtrades pels cubs.
- Avaluar els resultats obtinguts amb cadascun dels models
- Conclusions



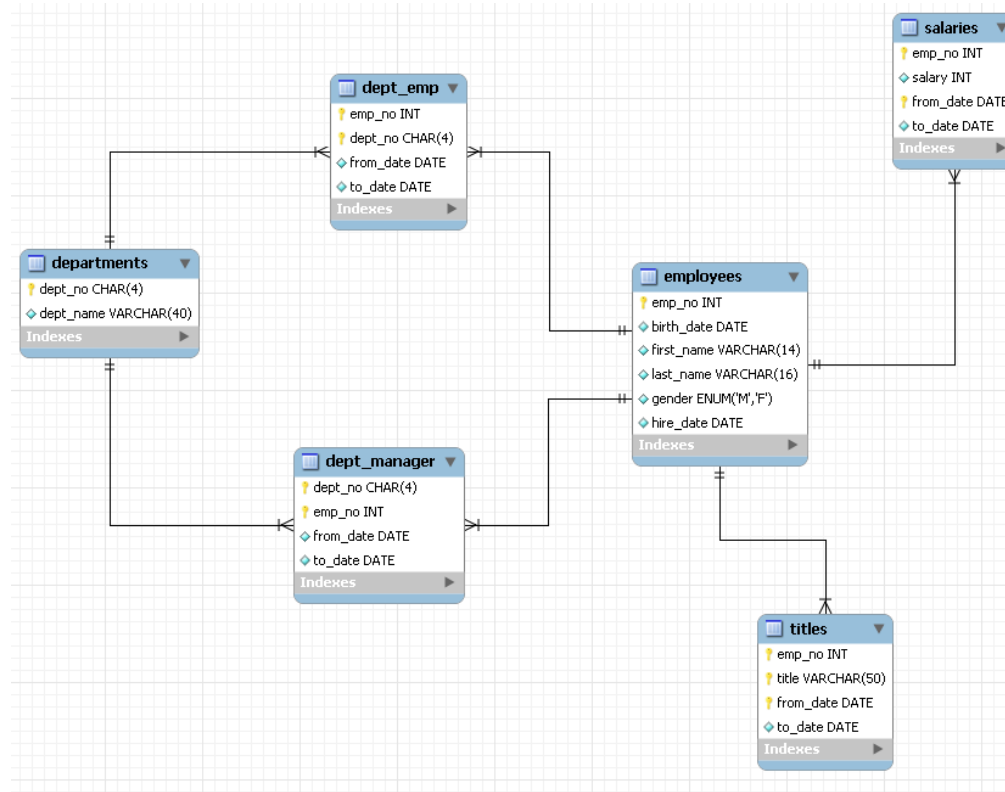
- Servidor Pentaho
- Bases de dades
- Cubs de dimensions
- Anàlisi de dades
- Monitorització del temps de resposta

-  Apache Tomcat

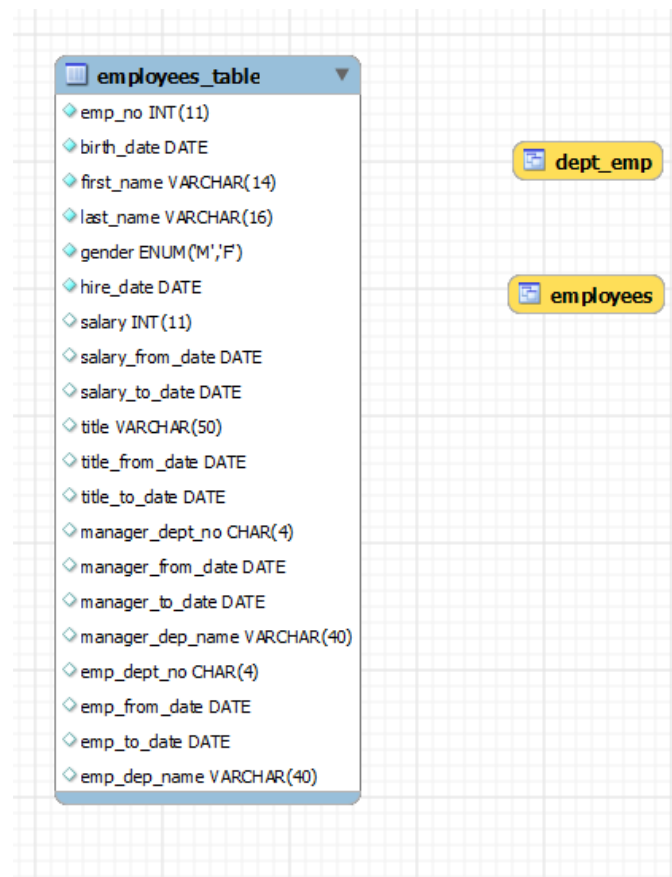
-  **mondrian**[™]
pentaho analysis services[™]

-  → 

- Model estrella vs Model taula única
- Original (estrella)

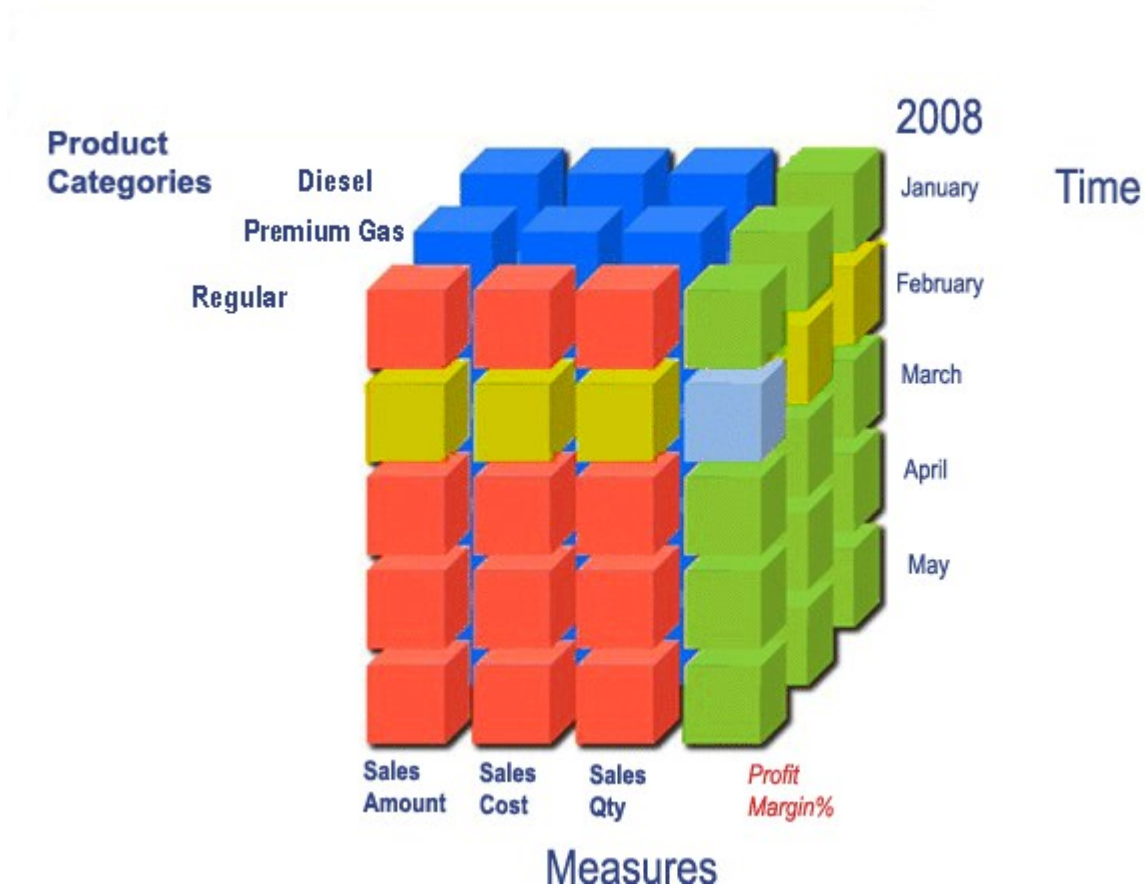


- Model taula única

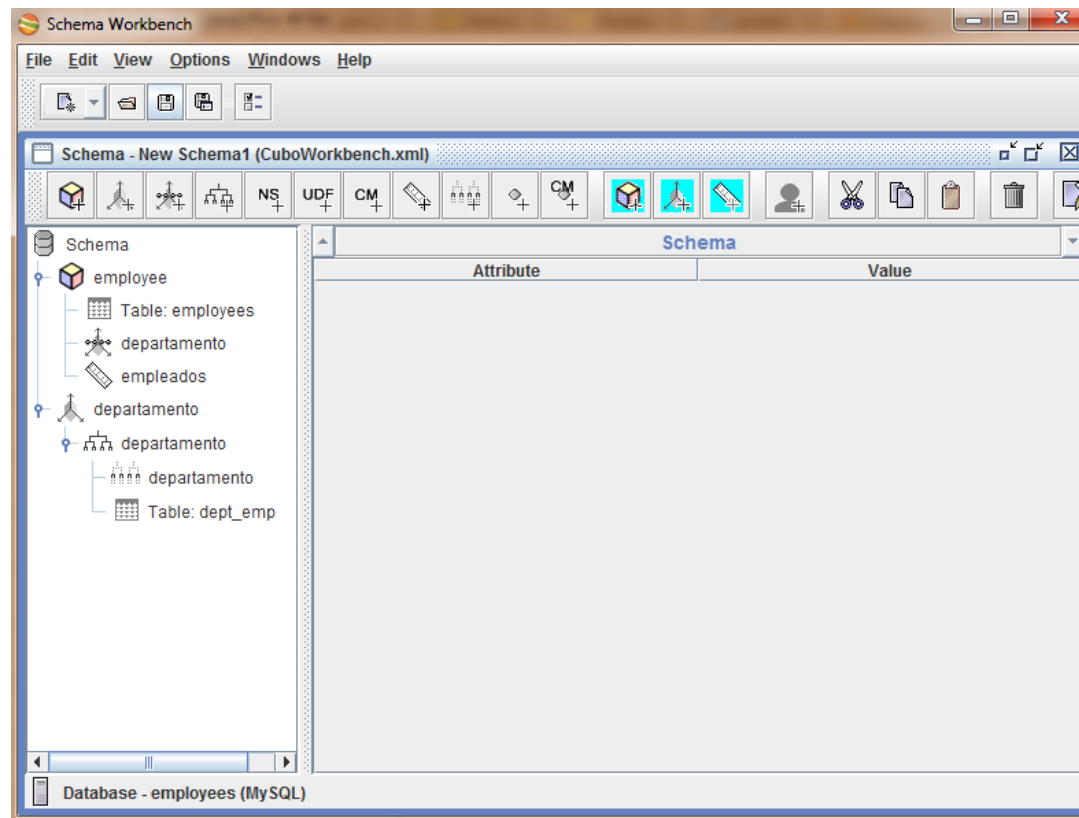


Necessitem:

- Mesures
- Dimensions



Dissenyar els cubs que utilitzarem alhora d'analitzar les dades i fer les proves de rendiment. Dissenyarem un cub que ens permeti comptar empleats separats per departaments.



- Preparar la configuració del pentaho perque faci un log detallat de les consultes.
- Un cop localitzades les peticions d'un model i d'un altre realitzarem una comparativa dels temps contra cadascun dels models.

```
-----  
2009-12-21 12:29:24,078 DEBUG [mondrian.sql] 16: RolapStar.Column.getCardinal  
from "DIM_TIME" as "DIM_TIME"]  
2009-12-21 12:29:24,078 DEBUG [mondrian.sql] 16: , exec 0 ms  
2009-12-21 12:29:24,078 DEBUG [mondrian.sql] 16: , exec+fetch 0 ms, 1 rows  
2009-12-21 12:29:24,093 DEBUG [mondrian.sql] 17: Segment.load: executing sql  
as "c1", "DIM_TIME"."YEAR_ID" as "c2", sum("ORDERFACT"."TOTALPRICE") as "m0"  
"PRODUCTS" as "PRODUCTS", "DIM_TIME" as "DIM_TIME" where "ORDERFACT"."CUSTOME  
"ORDERFACT"."PRODUCTCODE" = "PRODUCTS"."PRODUCTCODE" and "ORDERFACT"."TIME_ID  
"PRODUCTS"."PRODUCTLINE", "DIM_TIME"."YEAR_ID"]  
2009-12-21 12:29:24,375 DEBUG [mondrian.sql] 17: , exec 266 ms  
2009-12-21 12:29:24,390 DEBUG [mondrian.sql] 17: , exec+fetch 281 ms, 79 rows
```

- Autenticar-se via web al pentaho bi server



www.pentaho.com | [Contact Us](#)

Pentaho User Console

The Pentaho User Console provides access to powerful business analytics. Login for access to your reports, analysis views and dashboards.




Login

User Name:

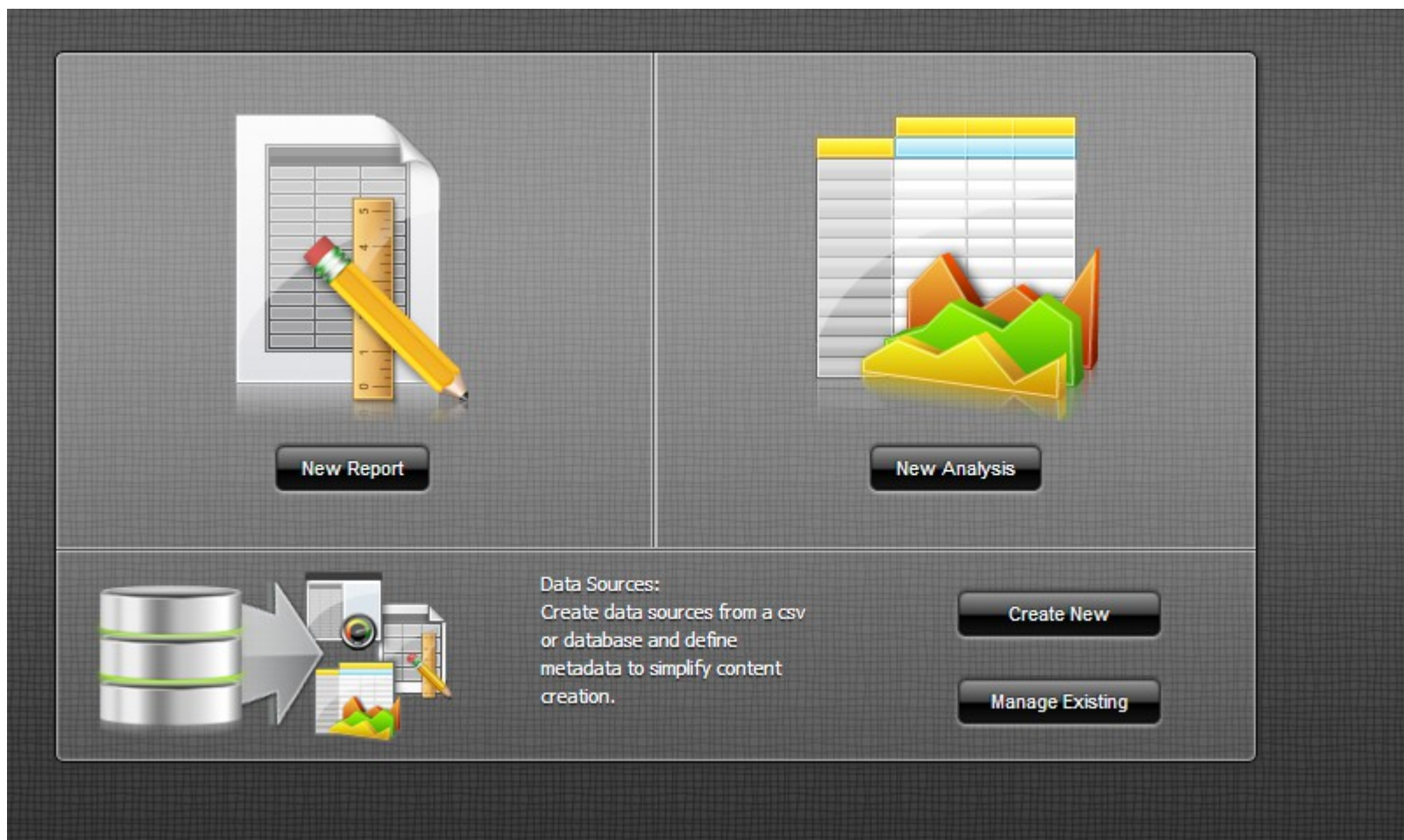
Password:

Launch in new window

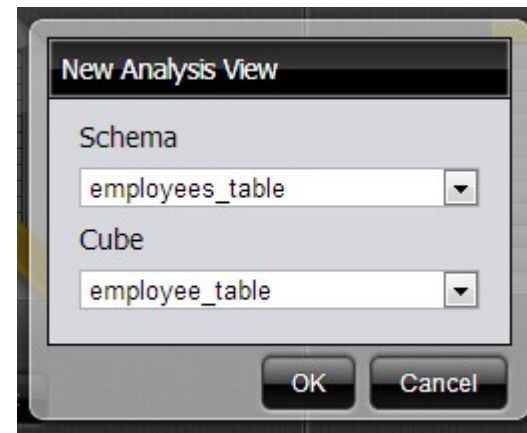
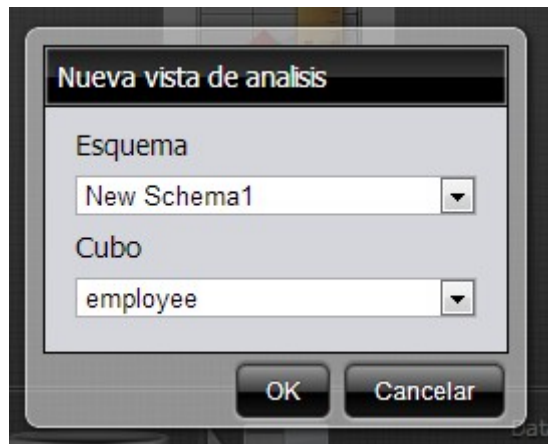
 [Evaluation Login](#)

Login

- Crearem un nou anàlisi de dades



- Obrirem cadascun dels cubs (primer el d'estrella i despres el de taula)



The screenshot displays the MDX Query Editor interface. On the left, a 'Filtro' (Filter) pane is open, showing a tree structure with 'Columnas' (Columns) containing 'Medidas' and 'Filas' (Rows) containing 'departamento'. Below the tree are 'Aplicar' and 'Cancelar' buttons. The main editor window contains the following MDX query:

```
select NON EMPTY {[Measures].[empleados]} ON COLUMNS,  
NON EMPTY {[departamento].[todos]} ON ROWS  
from [employee]
```

Below the editor is a toolbar with various icons for query execution and visualization. At the bottom left, a pivot table is shown:

	Medidas
departamento	● empleados
+ todos	300.024

Below the pivot table is a 'Slicer:' label. A yellow warning box in the center reads: 'JPivot has been replaced by Pentaho Analyzer. It is provided as a convenience but will no longer be enhanced or officially supported by Pentaho.' In the bottom right corner, there is a 'Reshacer' button.

Model estrella

```
select
  count(`employees`.`emp_no`) as `m0`
from
  `employees` as `employees`]
```

exec 145 ms

```
select
  `dept_emp`.`dept_no` as `c0`
from
  `dept_emp` as `dept_emp`
group by
  `dept_emp`.`dept_no`
order by
  ISNULL(`dept_emp`.`dept_no`) ASC,
  `dept_emp`.`dept_no` ASC]
```

exec 0 ms

```
executing sql [
select
  count(distinct `dept_emp`.`dept_no`) as `c0`
from
  `dept_emp` as `dept_emp`]
```

exec 150 ms

Model taula

```
select
  count(`employees`.`emp_no`) as `m0`
from
  `employees` as `employees`]
```

exec 11576 ms

```
select
  `dept_emp`.`dept_no` as `c0`
from
  `dept_emp` as `dept_emp`
group by
  `dept_emp`.`dept_no`
order by
  ISNULL(`dept_emp`.`dept_no`) ASC,
  `dept_emp`.`dept_no` ASC]
```

exec 6384 ms

```
executing sql [
select
  count(distinct `dept_emp`.`dept_no`) as `c0`
from
  `dept_emp` as `dept_emp`]
```

exec 6238 ms

Com hem vist en l'apartat anterior podem confirmar que el model en estrella és molt més ràpid que el model en taula, però per què?

- Olap necessita que les dimensions i les mesures estiguin representades per taules
- Existència de vistes intermediàries que retrassen l'execució de les peticions
- En el model de taula Les vistes son consultes sobre una taula que conté tota la informació
- Obligació de consultar sobre totes les dades i no només sobre les dades que necessitem

- Posada a punt d'un servidor pentaho bi server amb Mysql
- Disseny de cubs amb pentaho schema workbench
- Control i monitorització dun servidor pentaho bi server

Vies de continuació



Guió d'instal·lació i configuració del pentaho bi server

<http://interestingittips.wordpress.com/2011/04/27/complete-pentaho-installation-on-ubuntu-part-1/>

Com publicar un cub utilitzant l'esquema workbench

<http://wiki.pentaho.com/display/ServerDoc1x/Publishing+an+Analysis+Schema+Using+Schema+Workbench>

Configurar els orígens de dades

<http://wiki.pentaho.com/display/ServerDoc1x/Configuring+JNDI+Datasources+for+use+in+the+Pentaho+BI+Platform>

Com monitoritzar les consultes que fa el pentaho a mysql

<http://pentaho-en.phi-integration.com/bi-server/how-to-trace-sql-generated-by-mondrian>