

	Name	Type	Compressed size	Password p...	Size	Ratio	Date modified
★	dim_nivel_trans.ktr	KTR File	3 KB	No	21 KB	87%	2018-02-12 10:18 AM
	dim_rol_trans.ktr	KTR File	3 KB	No	21 KB	87%	2018-02-12 10:22 AM
	dim_zona_trans.ktr	KTR File	3 KB	No	21 KB	87%	2018-02-12 10:24 AM
	dimensions_cuatro.ktr	KTR File	4 KB	No	36 KB	91%	2018-02-14 5:39 AM
	f_salida_TB.ktr	KTR File	3 KB	No	21 KB	88%	2018-02-14 5:35 AM
a	fichero_de_salida.csv	Microsoft Excel Comma S...	20,183 KB	No	142,252 KB	86%	2017-12-22 5:34 PM
	olap_transformation_TB.ktr	KTR File	3 KB	No	26 KB	90%	2018-02-14 8:57 AM
	stg_files_dim_asignatura.sql	SQL File	1 KB	No	2 KB	65%	2018-02-12 7:29 AM
	stg_files_dim_nivel.sql	SQL File	1 KB	No	2 KB	64%	2018-02-12 7:29 AM
	stg_files_dim_rol.sql	SQL File	1 KB	No	2 KB	64%	2018-02-12 7:29 AM
	stg_files_dim_zona.sql	SQL File	1 KB	No	2 KB	64%	2018-02-12 7:30 AM
	stg_files_f_salida.sql	SQL File	1 KB	No	3 KB	69%	2018-02-12 7:30 AM
	stg_files_fact_salida.sql	SQL File	1 KB	No	3 KB	69%	2018-02-12 7:30 AM



BDAW_1. MODELO DE DATOS [STG + DIMENSIONAL ESTRELLA]

EXPLICACION PRUEBA

En base a los datos contenido en el fichero pasado en clase (**fichero_de_salida.csv**) para que todos podamos tener el mismo origen de datos, se pretende:

- Crear un área STG en MYSQL (localhost) que contenga la misma estructura y tipología que el fichero en cuestión.
 - Nombre del STG seguirá la siguiente nomenclatura: **STG_[USUARIO_AULA_VIRTUAL]**
 - Ejemplo: STG_RROMERO
- Crear un área OLAP en MYSQL (localhost) que contenga el modelo dimensional del fichero.
 - Nombre del OLAP seguirá la siguiente nomenclatura: **OLAP_[USUARIO_AULA_VIRTUAL]**
 - Ejemplo: OLAP_RROMERO
 - Crear N dimensiones (MIN: tiene que ser 4)
 - Crear tabla hecho correspondiente

PRECONDICIONES

- Tener el fichero **fichero_de_salida.csv**
- Tener instalado y configurado MySQL Server y MySQL WorkBench.

FICHEROS A ENTREGAR

- Export base de datos MYSQL (.sql) con el modelo de datos con las N dimensiones y la tabla de hechos.
- No tienen que llevar datos. Sólo modelo.

FECHA ENTREGA

La fecha de entrega es el día del 17 de Enero.

PENTAHO KITTLE

DATA INTEGRATION:
import the data from
the CSV

Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

View Design

Steps table

- Input
- Output
 - Delete
 - Microsoft Access Output
 - Table output
 - Update
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse

Welcome! Transformation 1 100%

CSV file input

Table output

CSV Input

Step name: CSV file input

Filename: C:\Users\desdemilab\Desktop\PentahoP1\ fichero_de_salida.csv

Delimiter: Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result: ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim typ
1	user_code	String		36		S/.	,	.	none
2	user	String		21		S/.	,	.	none
3	user_id	Integer	#	15	0	S/.	,	.	none
4	producto	String		9		S/.	,	.	none
5	nota	Integer	#	15	0	S/.	,	.	none
6	texto	Integer	#	15	0	S/.	,	.	none
7	resaltar	Integer	#	15	0	S/.	,	.	none
8	subrayar	Integer	#	15	0	S/.	,	.	none
9	tachar	Integer	#	15	0	S/.	,	.	none
10	manosalzada	Integer	#	15	0	S/.	,	.	none
11	rectangulo	Integer	#	15	0	S/.	,	.	none

Help OK Get Fields Preview Cancel

INSIDE THE CSV INPUT FILE:

GET THE NAMES OF THE FIELDS

CSV Input

Step name: CSV file input

Filename: C:\Users\desdemilab\Desktop\PentahoP1\fichero_de_salida.csv Browse...

Delimiter: | Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	user_code	String		36		S/.	,	.	none
2	user	String		21		S/.	,	.	none
3	user_id	Integer	#	15	0	S/.	,	.	none
4	producto	String		9		S/.	,	.	none
5	nota	Integer	#	15	0	S/.	,	.	none
6	texto	Integer	#	15	0	S/.	,	.	none
7	resaltar	Integer	#	15	0	S/.	,	.	none
8	subrayar	Integer	#	15	0	S/.	,	.	none
9	tachar	Integer	#	15	0	S/.	,	.	none
10	manozalada	Integer	#	15	0	S/.	,	.	none
11	rectanulo	Integer	#	15	0	S/.	,	.	none

Help OK Get Fields Preview Cancel

Scan results

Here are the results of the document scan:

Result after scanning 100 lines.

Field nr. 1 :
Field name : user_code
Field type : String
Maximum length : 36
Minimum value : 000007b7-a199-47a4-8288-441880f97a11
Maximum value : 000b3d90-5fa2-411d-a8e1-44684503cc3b
Nr of null values : 1

Field nr. 2 :
Field name : user
Field type : String
Maximum length : 21
Minimum value : Usuario Usuario 55000
Maximum value : Usuario Usuario 55047
Nr of null values : 1

Field nr. 3 :
Field name : user_id
Field type : Integer

Field nr. 4 :
Field name : producto
Field type : String
Maximum length : 9
Minimum value : ESA142695
Maximum value : ESP171925
Nr of null values : 1

Field nr. 5 :
Field name : nota
Field type : Integer

Field nr. 6 :
Field name : texto
Field type : Integer

Field nr. 7 :
Field name : resaltar
Field type : Integer

GO TO MYSQL WORKBENCH

CREATE A NEW SCHEMA

CREATE A NEW TABLE

USE THE NAMES FROM THE FIELDS

FSALIDA = RELATED TO THE TABLE OUTPUT in PENTAHO

Change the the datatype varchar numbers. Allow sufficient data

MySQL Workbench

Local instance MySQL57

File Edit View Query Database Server Tools Scripting Help

Navigator

MANAGEMENT

- Server Status
- Client Connections
- Users and Privileges
- Status and System Variables
- Data Export
- Data Import/Restore

INSTANCE

- Startup / Shutdown
- Server Logs
- Options File

PERFORMANCE

- Dashboard
- Performance Reports
- Performance Schema Setup

SCHEMAS

Filter objects

stg_files

- Tables
- Views
- Stored Procedures
- Functions

stg_files - Schema f_salida - Table

Table Name: f_salida Schema: stg_files

Collation: utf8 - default collati Engine: InnoDB

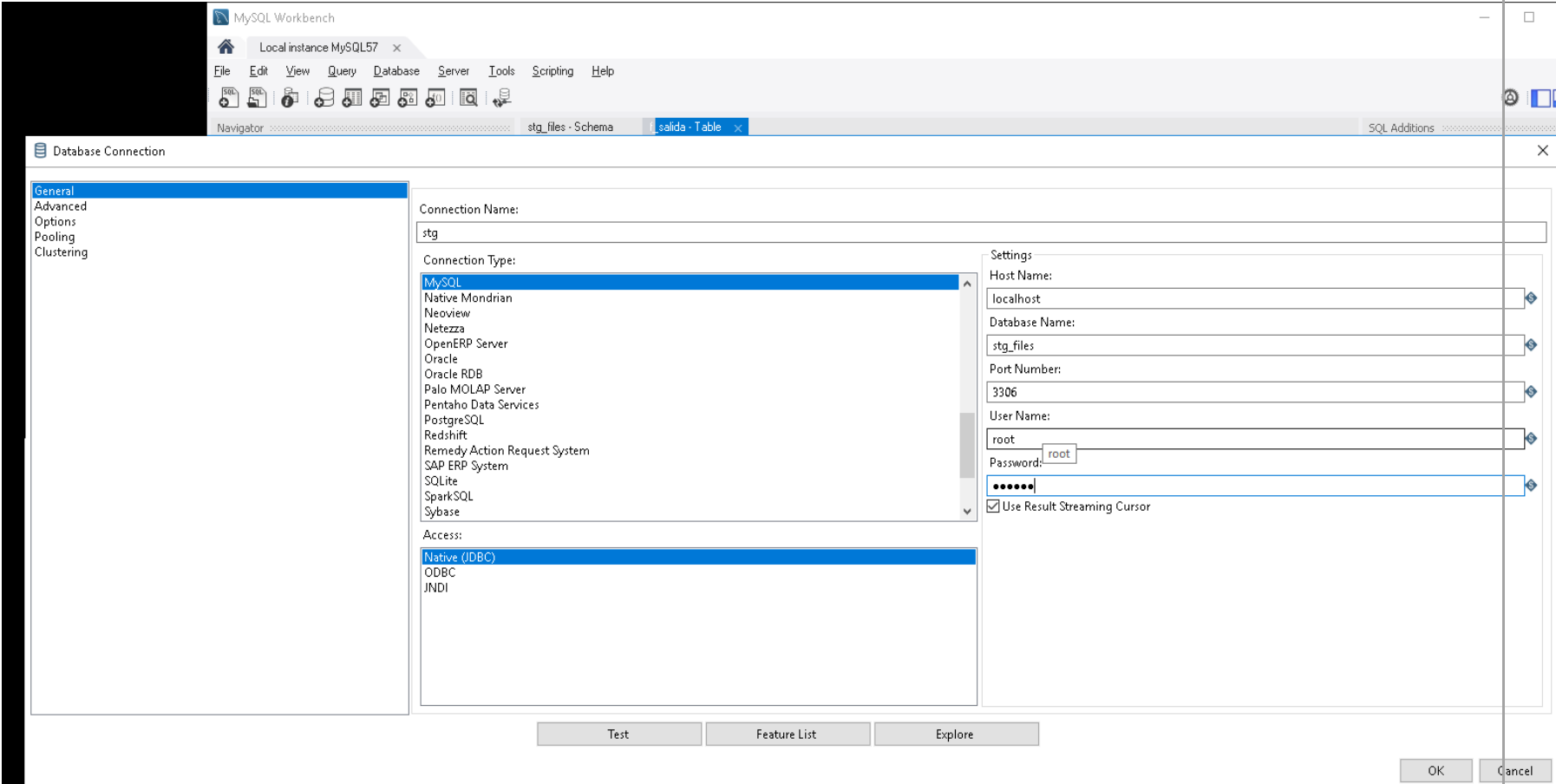
Comments:

Column Name	Datatype	PK	NN
user_id	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>
producto	VARCHAR(100)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
nota	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>
texto	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>
resaltar	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>
subrayar	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>
tachar	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>
manozada	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>
rectangulo	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>
elipse	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>
nivel	VARCHAR(45)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
curso	VARCHAR(45)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
asignatura	VARCHAR(45)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
titulo	VARCHAR(100)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
rol	VARCHAR(45)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
zona	VARCHAR(45)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
user_code	VARCHAR(45)	<input type="checkbox"/>	<input checked="" type="checkbox"/>

CONNECT TO A
DATABASE WITH
PENTAHO TABLE
OUTPUT

Then run the file

IF YOU HAVE ANA
ERROR IN MYSQL
CHECK THAT YOU
ARE CONNECTED TO
THE DATABASE BY
DEFAULT.



DOUBLE CLICK ON
STG

Navigator

MANAGEMENT

Server Status

Client Connections

Users and Privileges

Status and System Variables

Data Export

Data Import/Restore

INSTANCE

Startup / Shutdown

Server Logs

Options File

PERFORMANCE

Dashboard

Performance Reports

Performance Schema Setup

SCHEMAS

Filter objects

stg_files

Tables

Views

Stored Procedures

Functions

Information

Schema: stg_files

stg_files - Schema

f_salida - Table

SQL File 5* x

Limit to 1000 rows

1 x

SELECT * FROM f_salida;

Result Grid

Filter Rows:

Export: Wp.


	user_id	producto	nota	texto	resaltar	subrayar
▶	55000	ESA172025	1	0	0	0
	55001	ESA146820	0	0	0	0
	55002	ESA142695	1	0	0	0
	55003	ESA146714	0	2	6	1
	55004	ESA149230	0	0	0	0
	55004	ESA149232	0	0	0	0
	55004	ESA149340	0	0	0	1
	55004	ESA152759	0	0	0	0

CREATE THE OLAP AREA WITH THE FOUR DIMENSIONS: dimasignatura, dimrol, dimzona, dimnivel

IN MYSQL WORKBENCH:

For each of the 4 dimensions, input ID for column name and datatype. ALL IDs of the dimensions must be INT datatype

f_salida - Table olap - Schema **dim_asignatura - Table** dim_nivel - Table dim_rol - Table dim_zona - Table SQL File 6*

 Table Name: Schema: **olap**

Collation: Engine:

Comments:

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
idasignatura	INT(11)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
asignatura	VARCHAR(45)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

PENTAHO:

Input details for each step according to each of the 4 dimensions

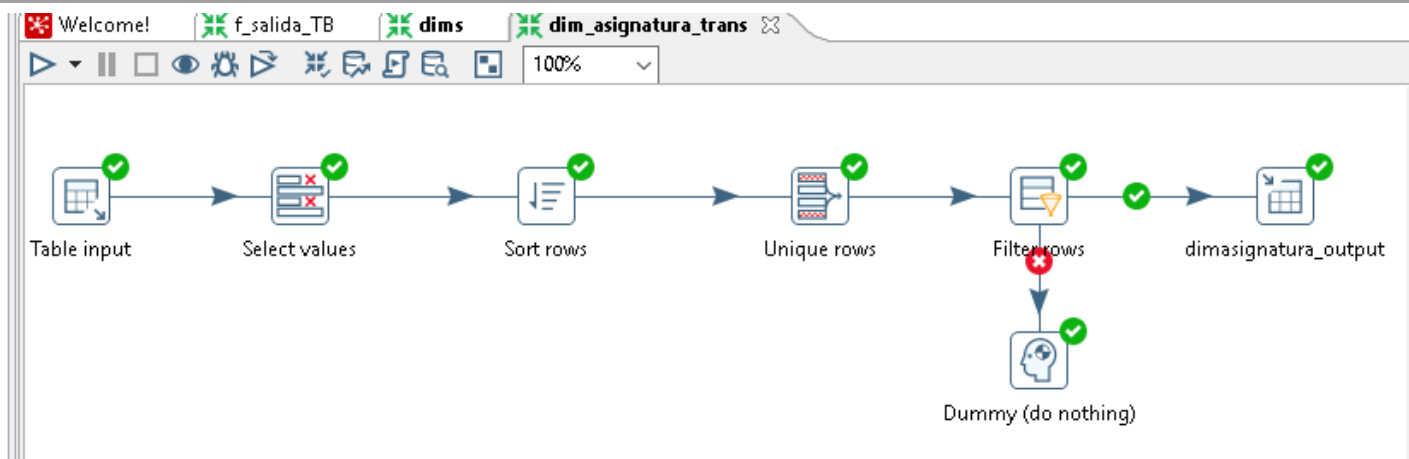


Table Input connection STG

Table input

Step name:

Connection:

SQL

```
SELECT * FROM f_salida;
```

Select/rename values

Select / Rename values

Step name

Select & Alter Remove Meta-data

Fields :

#	Fieldname	Rename to	Length	Precision	
1	asignatura				

Sort rows field: asignatura ONLY

Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

Free memory threshold (in %)

Compress TMP Files? ☐

Only pass unique rows? (verifies keys only) ☐

Fields :

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?	
1	asignatura	Y	N	N	0	N	

Unique rows:
fieldname asignatura
must be UNIQUE, no
doubles

Unique rows

Step name

Unique rows

Settings

Add counter to output?

☐

Counter field

Redirect duplicate row

☐

Error description

Fields to compare on (no entries means: compare complete row)

#	Fieldname	Ignore case
1	asignatura	N

Help

OK

Cancel

Get

Filter rows

Filter rows

Step name

Filter rows

Send 'true' data to step:

dimasignatura_output

Send 'false' data to step:

Dummy (do nothing)

The condition:

asignatura

IS NOT NULL

-

-

+

Assign a new output connection, olap_TB

Database Connection

General

Advanced

Options

Pooling

Clustering

Connection Name:

olap_TB

Connection Type:

MySQL

Native Mondrian

Neoview

Netezza

OpenERP Server

Oracle

Oracle RDB

Palo MOLAP Server

Pentaho Data Services

PostgreSQL

Redshift

Remedy Action Request System

SAP ERP System

SQLite

SparkSQL

Sybase

Access:

Native (JDBC)

ODBC

JNDI

Settings

Host Name:

localhost

Database Name:

olap

Port Number:

3306

User Name:

root

Password:

••••••

☒ Use Result Streaming Cursor

Test

Feature List

Explore

OK

Cancel

Table is now called dimasignatura_output

Table output

Step name

dimasignatura_output

Connection

olap_TB

Edit...

New...

Wizard...

Target schema

olap

Browse...

Target table

dim_asignatura

Browse...

Commit size

1000

Truncate table

☒

Ignore insert errors

☐

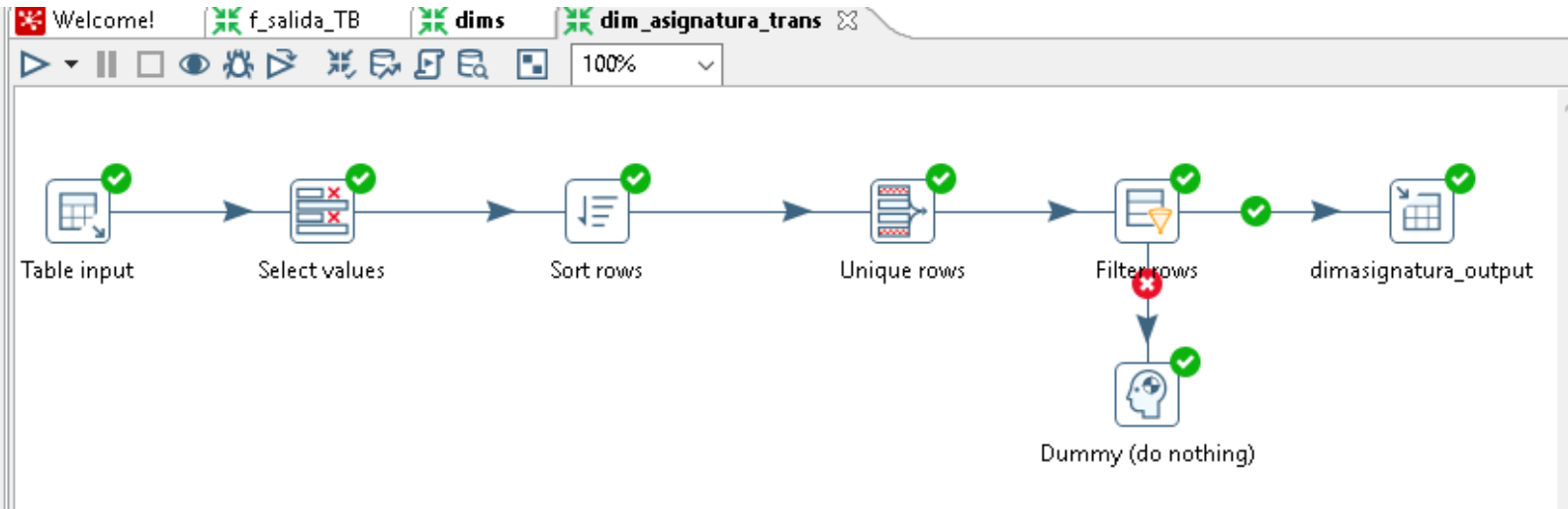
Specify database fields

☐

Main options

Database fields

Run the table



View the output.
Check that it all
worked!!!

MySQL Workbench

Local instance MySQL57 x

File Edit View Query Database Server Tools Scripting Help

Navigator

MANAGEMENT

- Server Status
- Client Connections
- Users and Privileges
- Status and System Variables
- Data Export
- Data Import/Restore

INSTANCE

- Startup / Shutdown
- Server Logs
- Options File

PERFORMANCE

- Dashboard
- Performance Reports
- Performance Schema Setup

SCHEMAS

Filter objects

olap

- Tables
 - dim_asignatura
 - Columns
 - id_asignatura
 - asignatura
 - Indexes

Information

Schema: olap

f_salida - Table olap - Schema dim_asignatura - Table dim_nivel - Table dim_rol - Table dim_zona - Table SQL File 6*

1 SELECT * FROM dim_asignatura;

Result Grid

	id_asignatura	asignatura
34		Lengua autonómica
35		Lengua castellana
36		Lengua y literatura
37		Lenguaje
38		Matemáticas
39		Matemáticas (Ciencias Naturales)
40		Matemáticas (Ciencias Sociales)
41		Matemáticas - Enseñanzas Académicas

asignatura 1 x

Apply Revert