

Business intelligence

Tutorial 1 – Schema Workbench



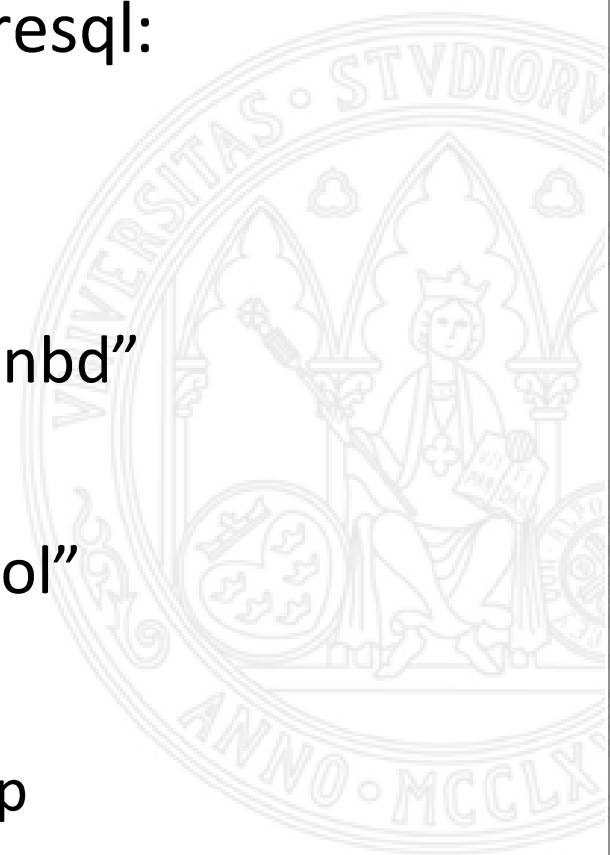
UNIVERSIDAD DE
MURCIA

- Tools
 - Mondrian: OLAP: Already installed in Pentaho biserver.
 - Schema Workbench: for designing cubes
 - Aggregation Designer: for designing aggregated tables
- Download:
 - Schema workbench: psw-ce-3.14.0.0-12.zip
 - <https://sourceforge.net/projects/mondrian/files/schema%20workbench/3.14.0/>
 - JDBC Driver for postgresql: postgresql-9.42.1.4.jar
 - Aula virtual or ...
 - <http://jdbc.postgresql.org/download.html>
- Tutorial: lets map the cube “dw_prof” given.

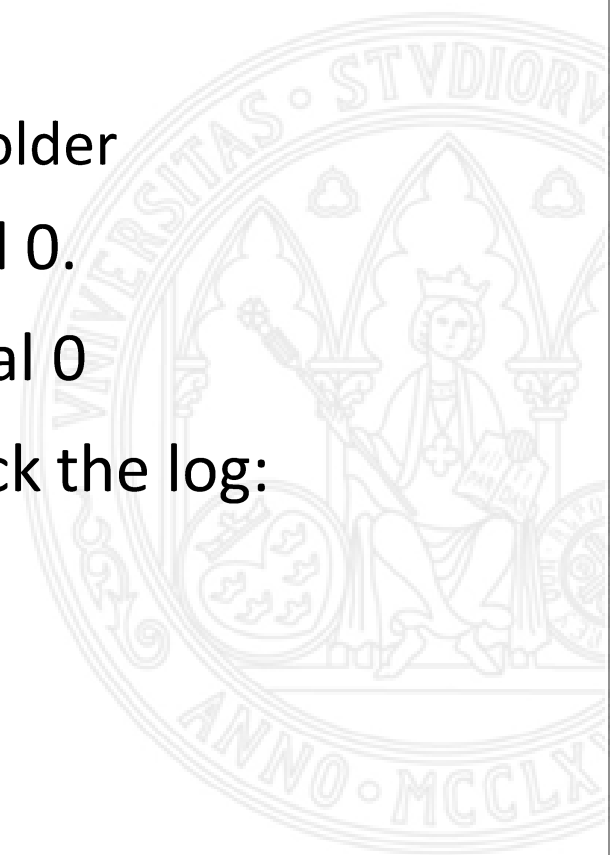
- En el laboratorio he instalado la versión 9.6 de postgres y la versión 7.1 de Pentaho
- Para poder hacer el tutorial en casa, en postgres:
 - Crear grupo de rol “inrol”
 - (No modificar permisos)
 - Crear usuario “inbd” con clave “inbd”
 - Asignarle rol “inrol”
 - Crear base de datos jjj y ponerle como owner “inrol”
 - Restaurar en esa base de datos el backup subido al aula virtual
 - in_2017_t1_dw_cargado_rol_inrol-ok.backup
 - Nota: postgres por defecto usa el puerto 5432.

- Para poder hacer el tutorial en casa, en pentaho:
 - Cambiar driver de postgresql: copia el driver 9.42 en
 - Biserver-ce/tomcat/lib
 - y elimina el anterior postgresql-9.3 de esa misma carpeta
- Prueba primero a hacer en Pentaho una conexión según hicimos en el tutorial 0
- Después prueba a hacer una fuente de datos según hicimos en el tutorial 0 (usa la consulta)
- Si da un error de validación de consulta:
 - biserver-ce/tomcat/logs/pentaho.log
 - Comprueba los permisos de acceso en la base de datos

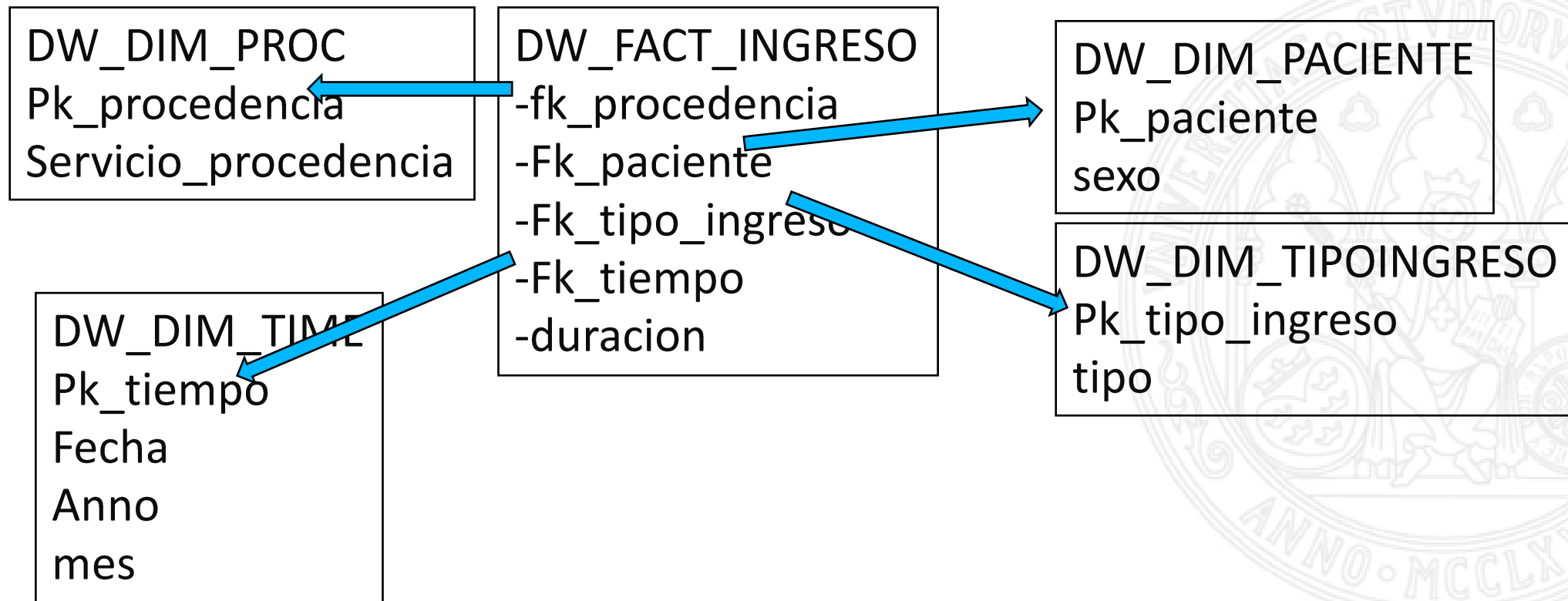
- In laboratory the versions are 9.6 for postgres and 7.1 for Pentaho
- To reproduce the tutorial at home, in postgresql:
 - Create the group role “inrol”
 - (Do not modify grants)
 - Create the user role “inbd” with password “inbd”
 - Membership to the role “inrol”
 - Create a new database “in” with owner “inrol”
 - Restaure in that database the backup:
 - in_2017_t1_dw_cargado_rol_inrol-ok.backup
 - Note: postgres default port is 5432.



- And in pentaho:
 - Change the postgresql jdbc driver: copy new driver in
 - Pentaho-server/tomcat/lib
 - Delete the old postgresql-9.3 driver in that folder
 - Try to create a new connection as in Tutorial 0.
 - Try to create a new Data Source as in Tutorial 0
 - If you get a query validation error, then check the log:
 - biserver-ce/tomcat/logs/pentaho.log
 - Check the database permission



- Target schema

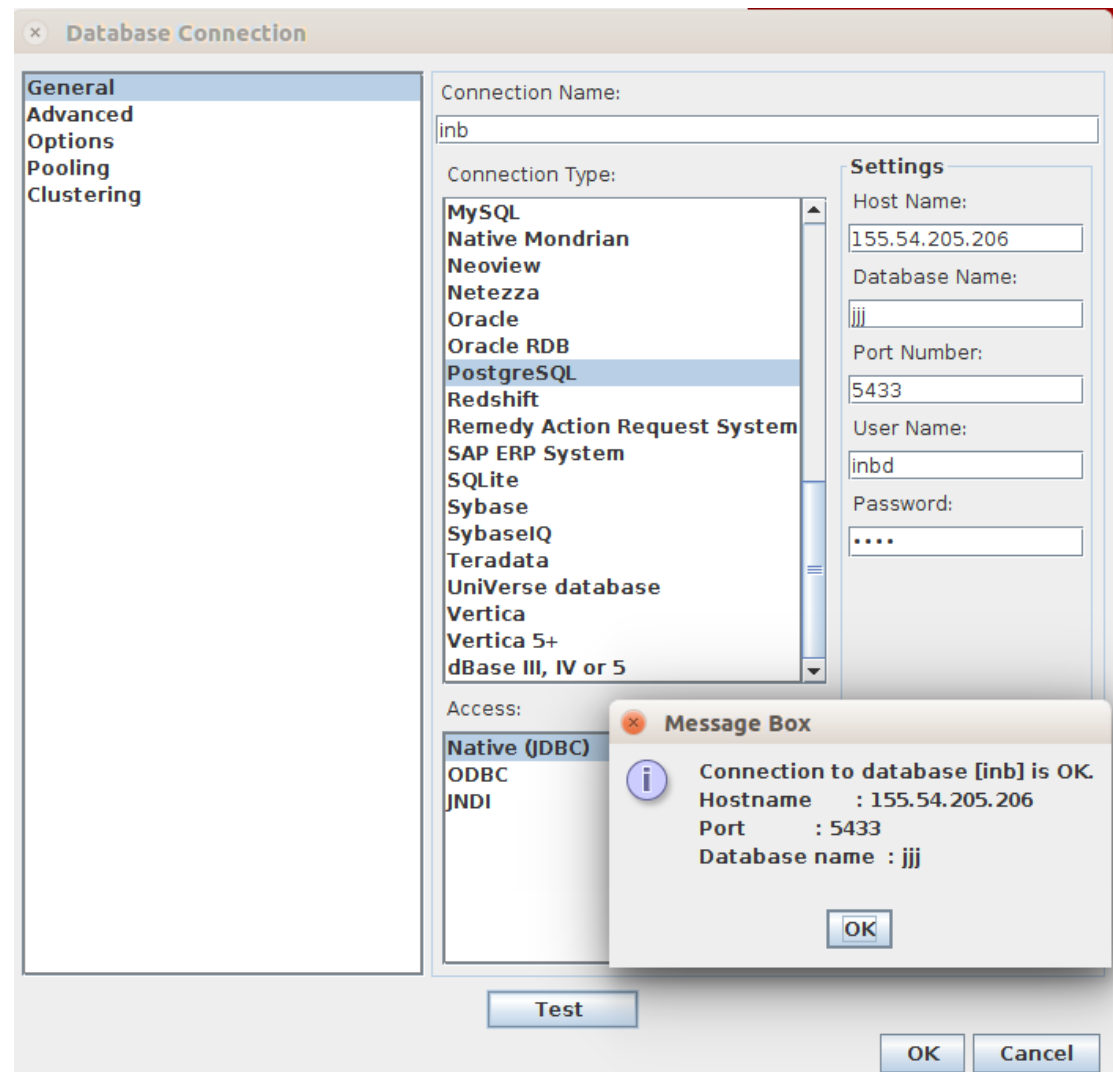


- Steps:
 - Unzip psw-ce-3.14.0.0-12.zip
 - **Copy JDBC driver into the schema-workbench/drivers folder**
 - Launch SchemaWorkbench: workbench.bat or workbench.sh
 - Create database connection
 - Create the cube
 - Defining facts
 - Defining measures
 - Defining dimensions
 - And then we publish the cube in pentaho

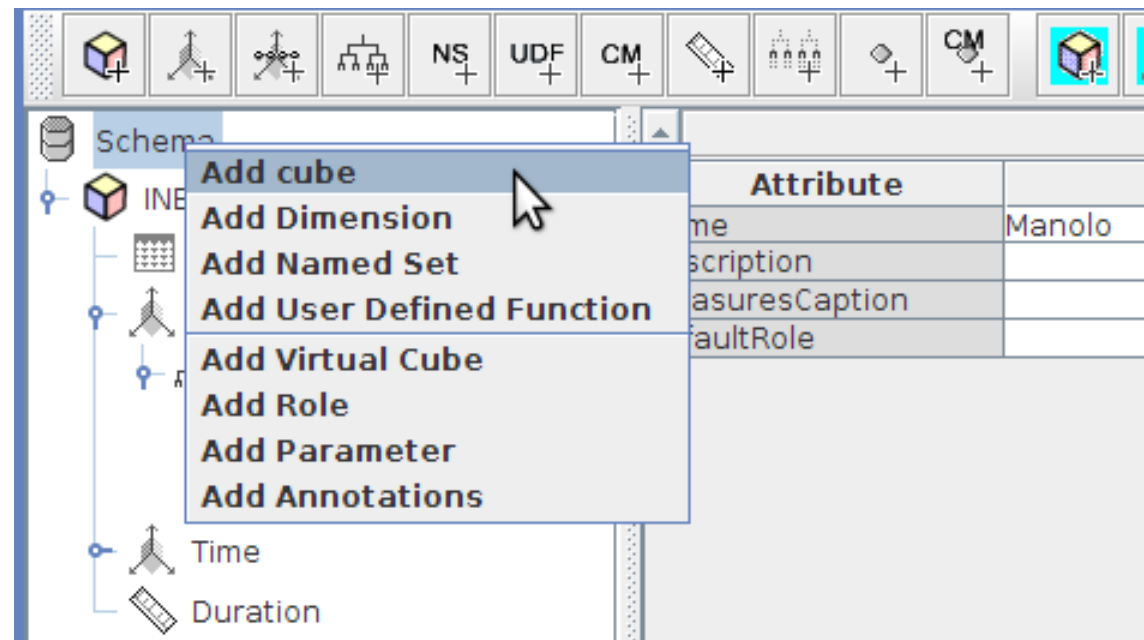


Create database connection

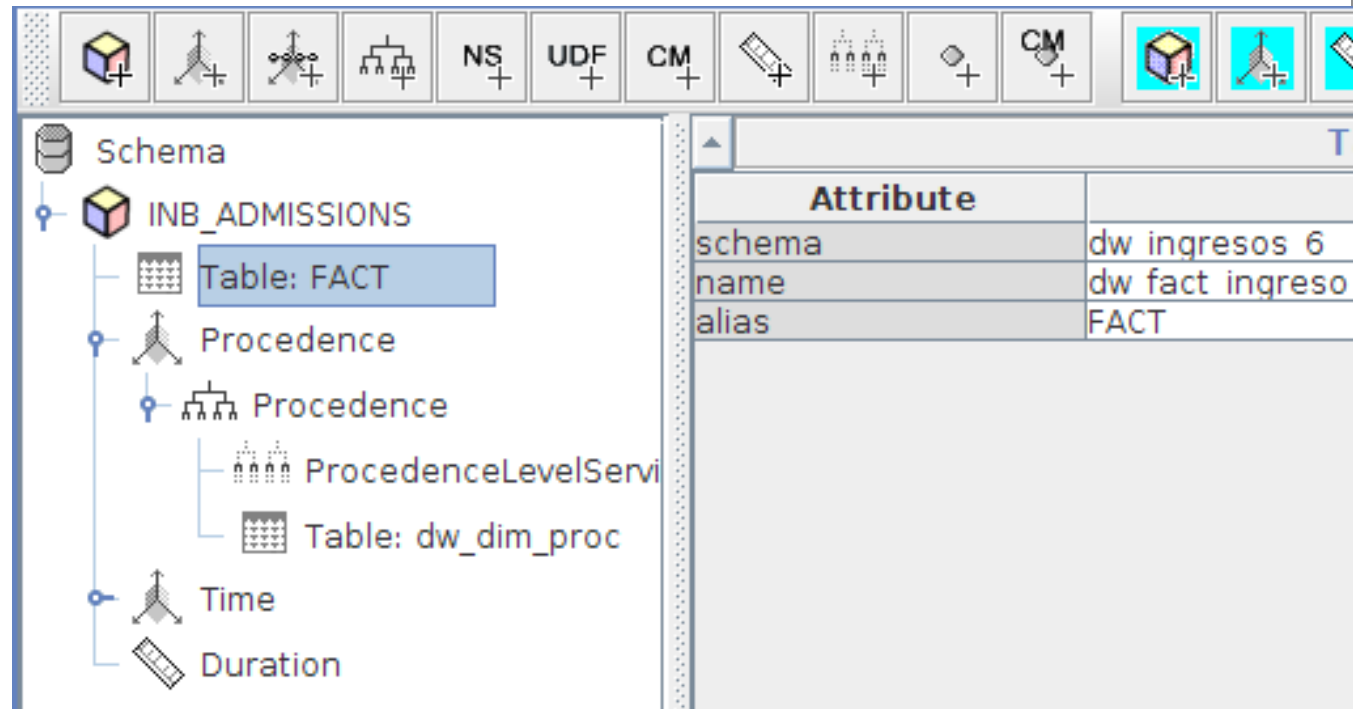
- Create database connection
 - Menu “Option”->Connection
 - Type: Postgresql
 - Hostname: ipgiven
 - Port number: given
 - Database: given
 - User/password: given
- Test it!



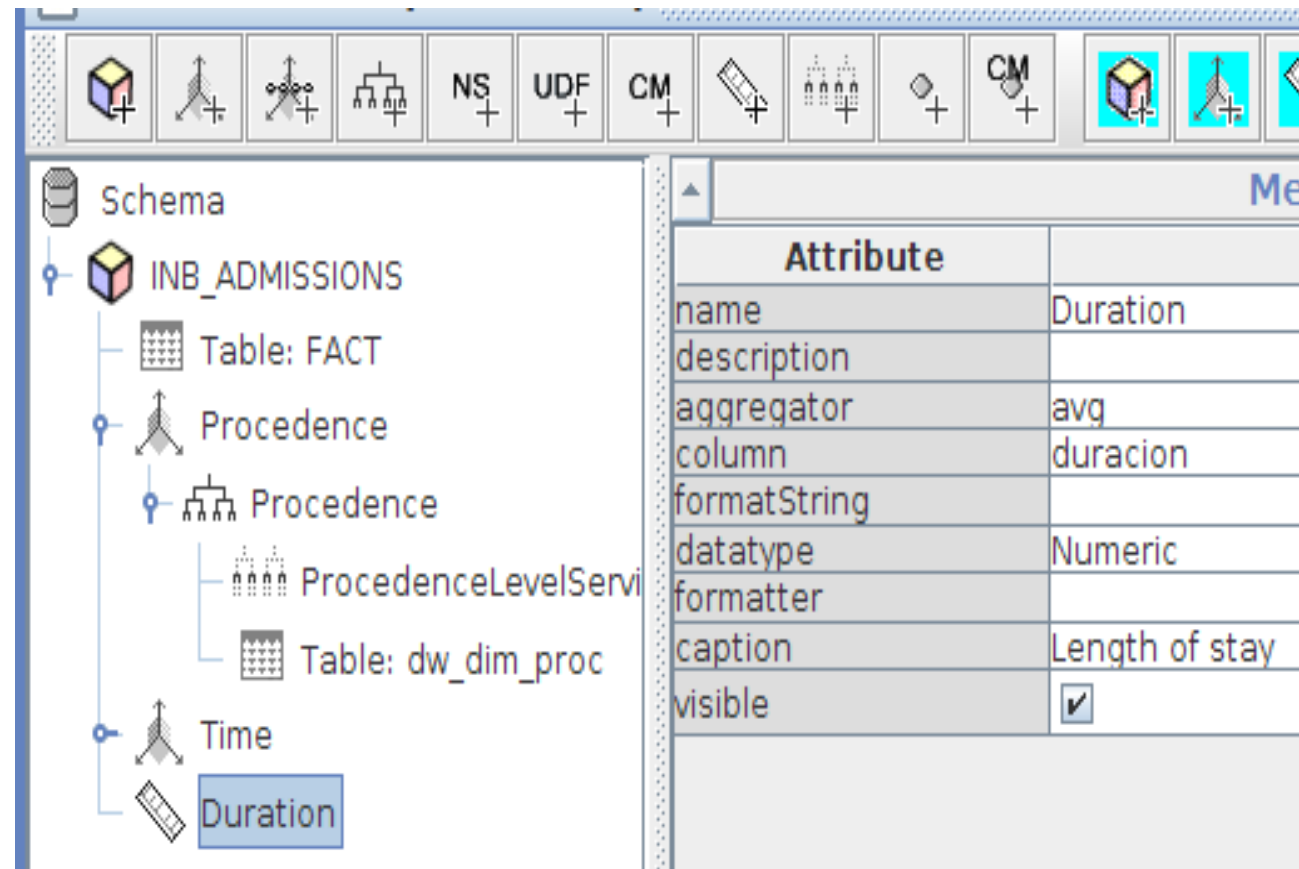
- First of all
 - File -> new -> Schema
 - Set a name for the schema:
YOUR ID
 - R-click on Schema
 - Add a cube
 - Name the cube
 - **SET YOUR NAME**
 - Description
 - Caption
 - Let the other parameters
 - Warning: **red cross tells you something is missing**
 - Ej: **"Fact name must be set"**



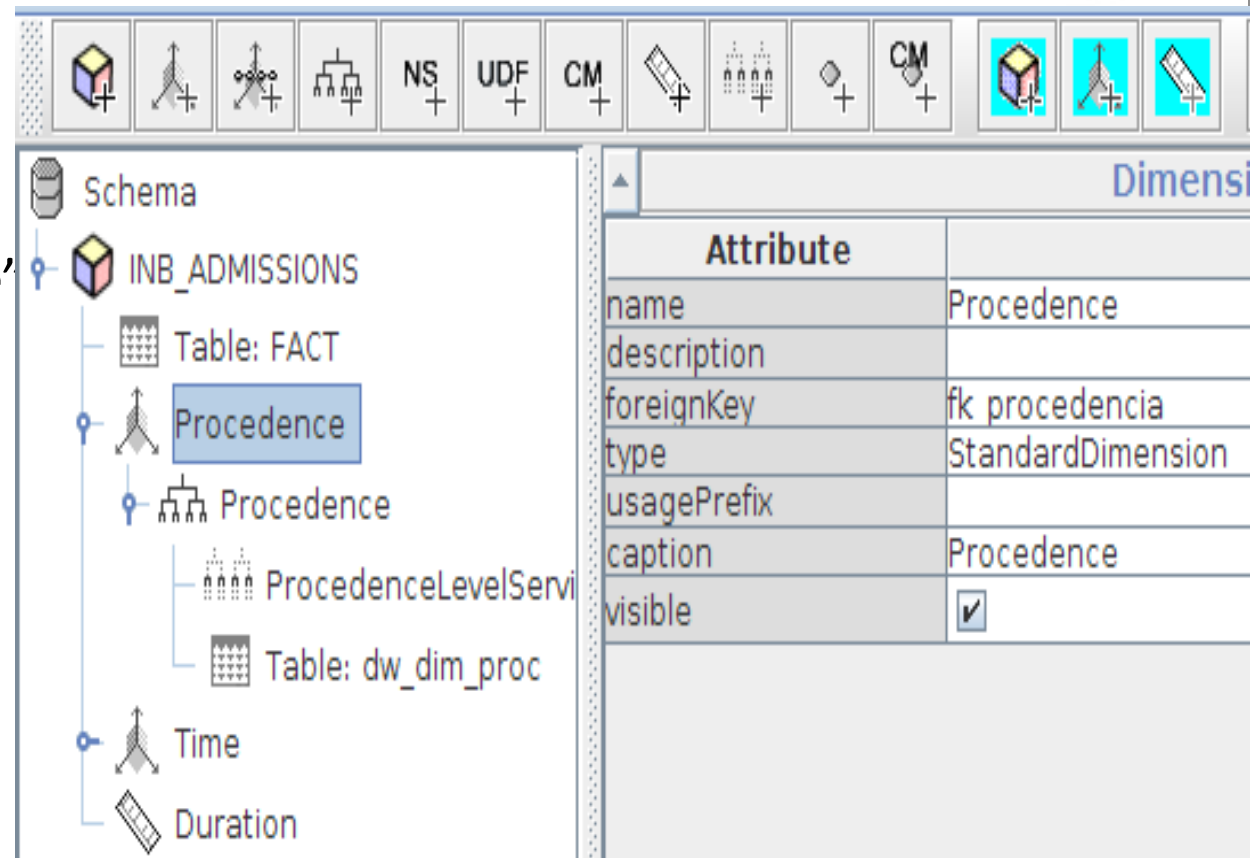
- Enter fact table
 - R-Click on the cube
 - Add Table
- Set *schema*
 - *dw_ingresos*
- Set *table*
 - *dw_fact_ingreso*
- Red message:
 - “*cube must contain dimensions*”



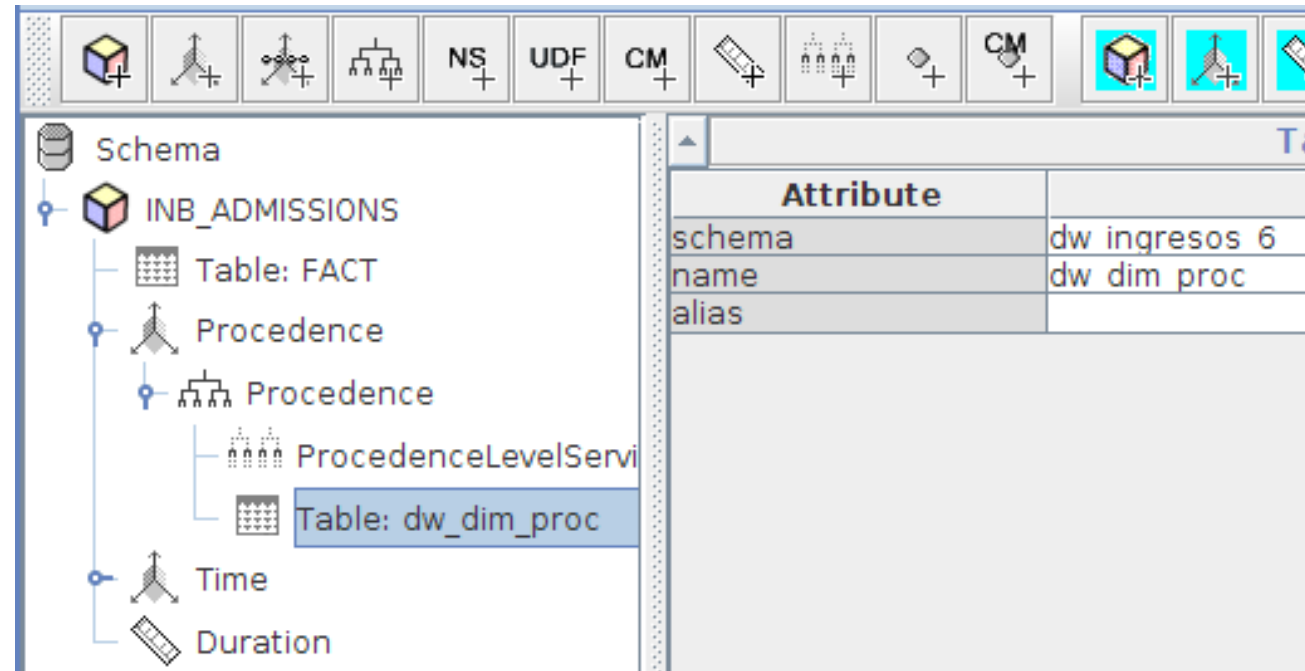
- Enter measures
 - R-Click on the cube
 - Add measure
 - Set *column from the table*
 - “*duracion*” (lenght)
 - Set *aggregator*
 - “*avg*” in our case
 - Set datatype
 - Numeric
 - Set *caption*
 - “*Length of stay*”



- Still error: **Cube must contain dimensions**
- First simple dimensions: Procedence
- Steps
 - R-Click on the cube
 - Add dimension “Procedence”
 - Set
 - Name: “Procedence”
 - Select FK from fact table
 - fk_procedencia
 - Type: StandardDimension
 - **RM: Hierarchy is invalid.**
 - A hierarchy is created by default. We define it now ...



- A dimension can have multiple hierarchies
- Steps
 - L-Click on hierarchy
 - Name the hierarchy
 - “Procedence”
 - R-Click on hierarchy
 - Add table
 - Select your schema
 - dw_ingresos
 - Select table
 - dw_dim_proc



- Go back to the hierarchy
 - Set *primaryKey*
 - pk_procedencia*
 - Set *caption*
 - Description shown*
 - Additionally we could define
 - hasAll*
 - allMemberName*
 - allMemberCaption*
 - allLevelName*
- RM: At least one level must be set for a hierarchy*

The screenshot shows a data modeling tool interface. On the left, a tree view under 'Schema' shows a fact table 'INB_ADMISSIONS' and a dimension 'Procedence'. The 'Procedence' dimension has a hierarchy with levels: 'Procedence' (root), 'ProcedenceLevelServi', and 'Table: dw_dim_proc'. Other dimensions like 'Time' and 'Duration' are also visible. On the right, a 'Hierarchy' table lists attributes for the 'Procedence' dimension.

Attribute	
name	Procedence
description	
hasAll	<input checked="" type="checkbox"/>
allMemberName	
allMemberCaption	AllProcedences
allLevelName	AllLevels
defaultMember	
memberReaderClass	
primaryKeyTable	
primaryKey	pk procedencia
caption	ProcedenceCaption
visible	<input checked="" type="checkbox"/>

Create dimensions – Adding levels to a hierarchy

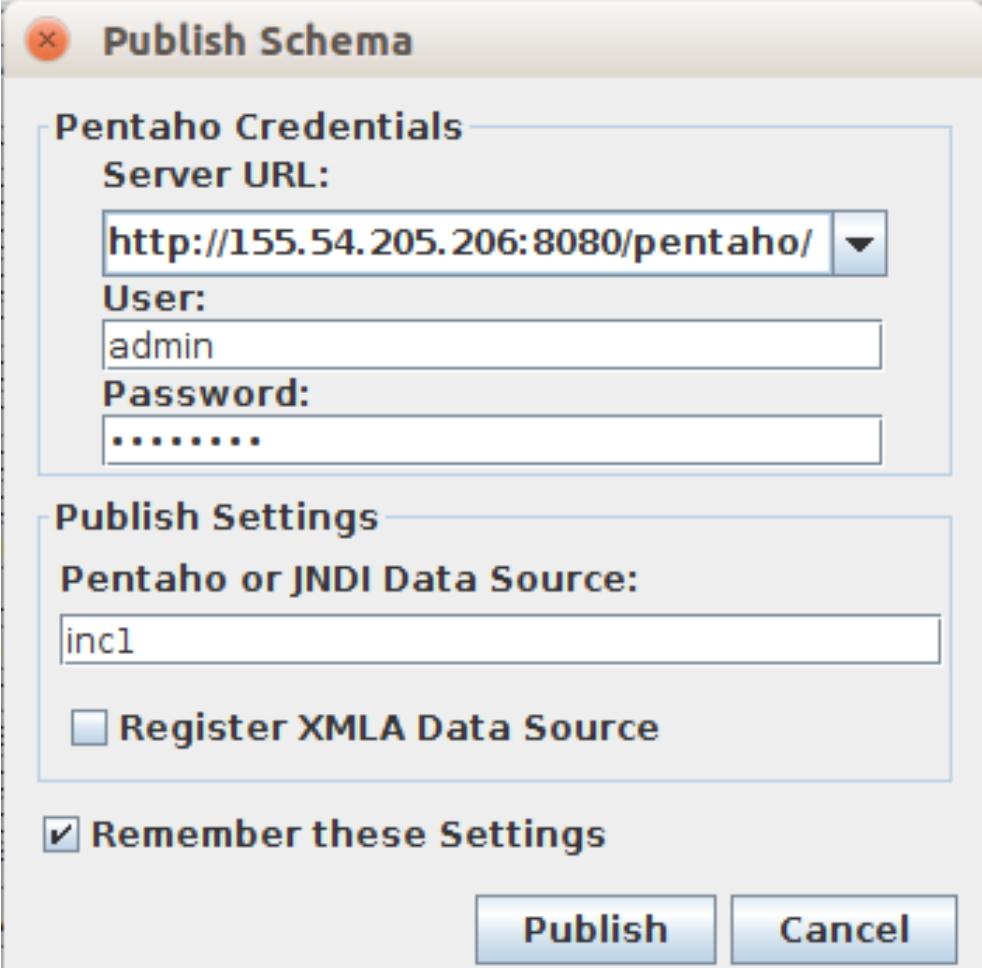
- A hierarchy needs some levels
- R-Click on hierarchy
- Add level
 - Set *name*
 - Set *table*
 - *dw_dim_proc*
 - *important to write it, not to select, or an error occurs*
 - Set *column*
 - *servicio_procedencia*
 - Set *type: String*
 - *LevelType: Regular*

The screenshot shows the SQL Server Enterprise Manager interface. On the left, a hierarchy is displayed under the 'Schema' folder. The hierarchy consists of: 'INB_ADMISSIONS' (cube icon), 'Table: FACT' (table icon), 'Procendence' (tree icon), 'Procendence' (tree icon), 'ProcendenceLevelService' (tree icon, highlighted with a blue selection bar), 'Table: dw_dim_proc' (table icon), 'Time' (calendar icon), and 'Duration' (calendar icon). A red arrow points from the 'ProcendenceLevelService' node to the 'Attribute' table on the right. The 'Attribute' table has the following columns and values:

Attribute	
name	ProcendenceLevelService
description	
table	dw dim proc
column	servicio procedencia
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	
type	String
internalType	
uniqueMembers	<input type="checkbox"/>
levelType	Regular
hideMemberif	Never
approxRowCount	
caption	Nivel Procedencia
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>

Lets try it in pentaho

- **Verify you don't have any redcross**
- Save the schema
- Publishing
 - Menu File -> Publish
 - URL: <http://ipgiven:8080/pentaho>
 - User/pwd: **your own**
 - Pentaho or JNDI Data source: **your own**
 - (PREVIOUSLY JDBC DATA SOURCE from tutorial)
- Publish successful!!



Publish Schema

Pentaho Credentials

Server URL:

User:

Password:

Publish Settings

Pentaho or JNDI Data Source:

☐ Register XMLA Data Source

☒ Remember these Settings

Publish **Cancel**

Lets try it in pentaho

- You should be able to see your own cube ...
- New **JPivot** View
- Only Procendence is available
- Check “Captions” to see what you wrote



You can save it in Pentaho

- Push “Save” and set a name and folder
- You will have it ready

The screenshot shows the Pentaho BI Desktop interface. A red arrow points to the 'Save' icon (a blue floppy disk) in the top toolbar. Another red arrow points to the 'Analysis1' file in the 'Archivos' (Files) pane on the left. The 'Archivos' pane shows a tree structure with folders like 'Steel Wheels', 'Una Colección de muestras y', 'cde', 'manolo', and 'plugin-samples'. The 'Analysis1' file is highlighted at the bottom of the list.

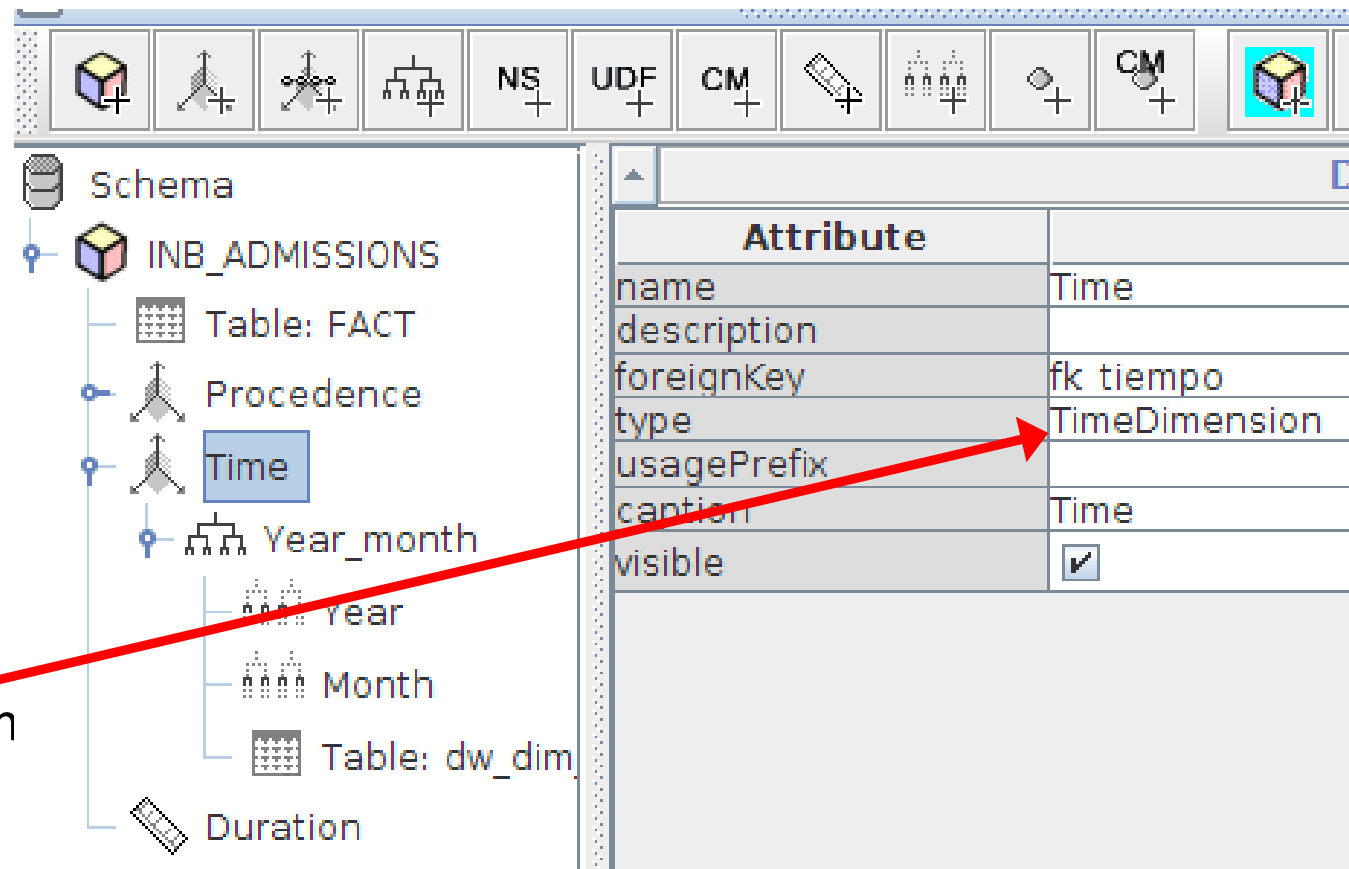
Analysis1

Medidas	
ProcedenceCaption	YearMonths
+	ProcsCap
-	AllCap
-	2003
	December
-	2004
	January
	February

LengOfStay
7,601
116
116
7,262
9,68
10,368

- Time

- Create dimension as before
- Set *name*
- Set *caption*
- Choose *foreign key*
 - *fk_tiempo*
- Set *type*: TimeDimension



- Time
 - Define the hierarchy
 - Set name
 - Check “hasAll”
 - Set *AllMemberName*
 - Set *AllMemberCaption*
- Add table
 - Select YOUR schema
 - **dw_ingresos**
 - Select table: ***dw_dim_time***
- Back to the hierarchy
 - Set *primaryKey*: ***pk_tiempo***

The screenshot shows a data modeling software interface. On the left, a hierarchy tree is displayed under the 'xSchema' node. The tree structure is as follows:

- xSchema
 - xCubeAdmissionsTest
 - Table: fact
 - Procidence
 - Procidence
 - ProcLevel
 - Table: dw_dim
 - xTime
 - xTime (highlighted in blue)
 - Table: dw_dim
 - AvgLeng

On the right, a table displays the properties of the selected 'xTime' node:

Attribute	Value
name	Time
description	
hasAll	<input checked="" type="checkbox"/>
allMemberName	All
allMemberCaption	AllCap
allLevelName	AllLevelName
defaultMember	
memberReaderClass	
primaryKeyTable	
primaryKey	pk_tiempo
caption	TiemCap
visible	<input checked="" type="checkbox"/>

- Time

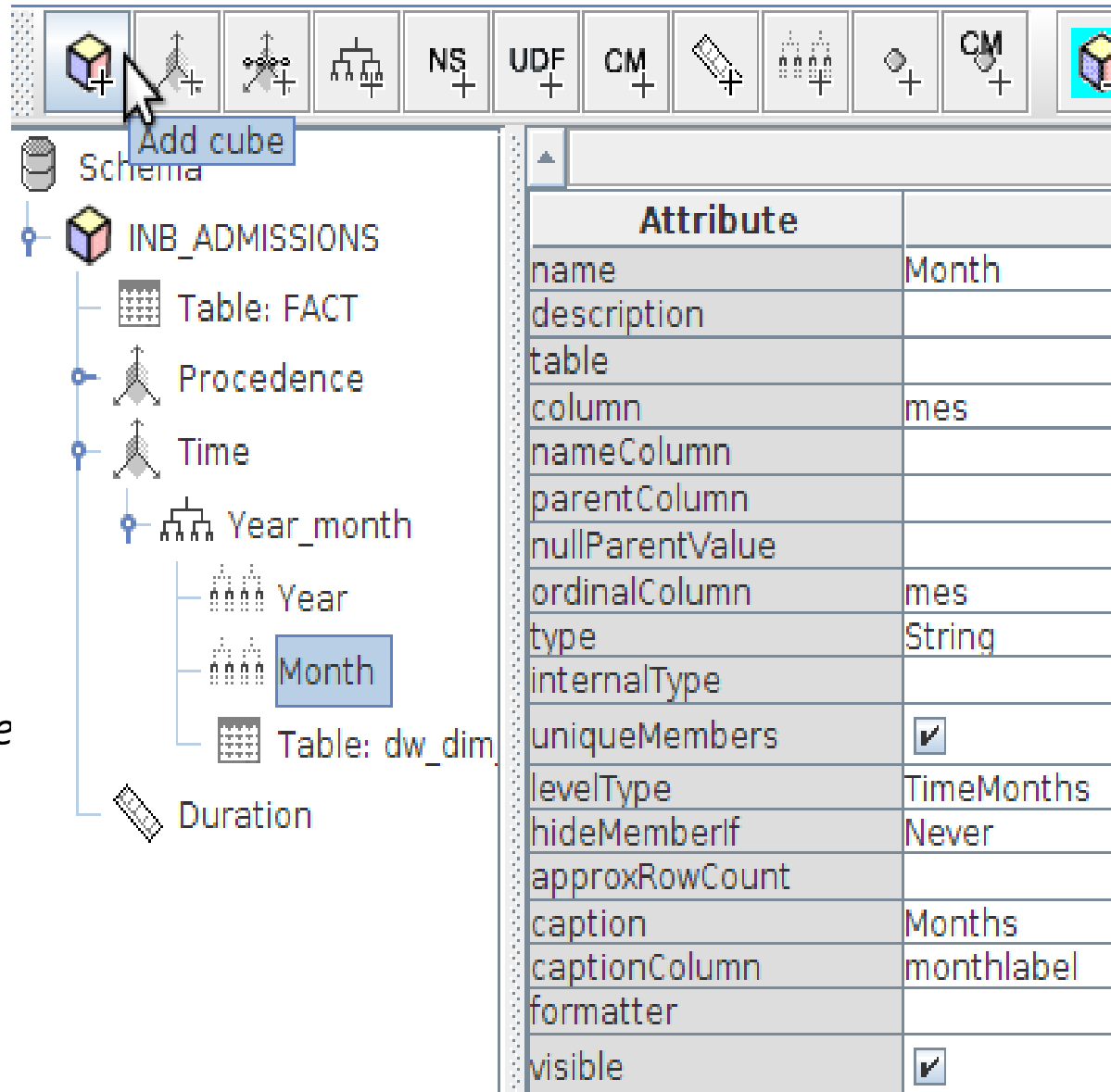
- R-click on hierarchy
- Add level
- Set *name*: Year
- Set *column*: anno
- Set *type*: Integer
- Set *LevelType*
 - TimeYears

The screenshot shows a data modeling tool interface. On the left, a hierarchy tree is displayed under the 'Schema' node. The hierarchy is: INB_ADMISSIONS (Table: FACT) -> Procidence -> Time -> Year_month -> Year (highlighted). Other nodes in the hierarchy include Month and Table: dw_dim. On the right, a table of attributes for the 'Year' level is shown. The table has two columns: 'Attribute' and a value column. The attributes and their values are:

Attribute	
name	Year
description	
table	
column	anno
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	
type	Integer
internalType	
uniqueMembers	<input checked="" type="checkbox"/>
levelType	TimeYears
hideMemberIf	Never
approxRowCount	
caption	Year
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>

• Time

- R-click on hierarchy
- Add level
- Set *name: Month*
- Set *column: mes*
 - *Better: load a name!*
 - *January, february, ...*
 - Set *ordinalColumn: mes*
 - Or set *captionColumn* for the “mes” column (*monthlabel*)
- Set *type: Integer*
- Set *LevelType*
 - TimeMonth
- UnCheck *uniqueMembers*



The screenshot shows a data modeling tool interface. On the left, a hierarchy is displayed under a 'Schema' node. The hierarchy includes a cube icon, a table icon labeled 'INB_ADMISSIONS', a 'Table: FACT' icon, a 'Procedence' icon, a 'Time' icon, a 'Year_month' icon, a 'Year' icon, a 'Month' icon (highlighted with a blue box), a 'Table: dw_dim' icon, and a 'Duration' icon. A tooltip 'Add cube' is visible over the cube icon. On the right, an 'Attribute' table is shown with the following data:

Attribute	
name	Month
description	
table	
column	mes
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	mes
type	String
internalType	
uniqueMembers	<input checked="" type="checkbox"/>
levelType	TimeMonths
hideMemberIf	Never
approxRowCount	
caption	Months
captionColumn	monthlabel
formatter	
visible	<input checked="" type="checkbox"/>

- Upload to the Aula Virtual the XML definition of the cube adding:
 - Dimension for admission type
 - Dimension for patient -> Only sex hierarchy
- Publish the schema in the platform

