Mais flaviocardoso@gmail.com

Diethard Steiner on Business Intelligence

Open source business intelligence tutorials: Pentaho, Talend, Jasper Reports, BIRT and more.
Topics: Data Integration, Data Warehousing, Data Modeling, BI Server Setup, OLAP, Reporting, Dashboarding, Master Data Management and many more.

About Me

Unknown

View my complete profile

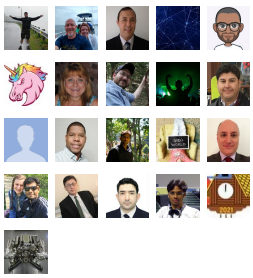
Subscribe

Posts

Comments

Followers

Seguidores (188) Próxima



Seguir

Blog Archive

- 2014 (9)
- 2013 (25)
- 2012 (20)
- 2011 (19)
  - November (3)
    - Jasperwave - the other JasperReports Designer It's...
    - Jaspersoft iReport: How to pass a parameter to a s...
    - Star Schema Modeling with Pentaho Data Integration
  - July (2)
  - June (1)
  - May (4)
  - April (2)
  - March (1)
  - February (3)
  - January (3)
- 2010 (17)
- 2009 (18)

Friday, November 4, 2011

Star Schema Modeling with Pentaho Data Integration

Star Schema Modeling with Pentaho Data Integration

Tutorial Details

- Software (this tutorial was published on 2011-11-04 and download sources were correct on this date but might change in future):
  - Download the latest version of PDI which includes the Star Modeler plugin (download here). If you run pdi-ce-4.3.0-M1, you can download the Star Modeler plug-in from here. Ideally though, just download the latest version of PDI.
  - MySQL or similar database
  - Star Schema file produced in this tutorial (download here)
- Knowledge: Intermediate (To follow this tutorial you should have good knowledge of the software and hence not every single step will be described)

Introduction

Matt Casters, lead developer of PDI, presented a new feature this week for modeling star schemata within PDI. It's in the very early stages, but already showing huge potential and this is the reason why I thought I prepare a quick tutorial about

Now why is this new feature so noteworthy:

- You can create your star schema model within the same interface as you create your ETL process, report models, analyzer models and which allows you to analyze your data as well. So, now it's really like a start-to-finish environment.
- Other features can make use of the metadata you specified within the Star Modeler plug-in! In fact, one button click and PDI automatically generates a job which will run the DDL against the target database. Another click, and a simple ETL transformation gets automatically generated to populate your dimensions. You can use the transformation as a starting point and further improve it if necessary. In future you will be also able to auto-generate the reporting (**Pentaho Metadata**) model and the Mondrian (**Pentaho Analysis**) model. By how much will this speed up your development process? I would say a lot!
- From my point of view Matt also included some really nifty features, in example, when you specify a table as date dimension, you have the possibility to let PDI auto-generate one for you. The source data to target data mapping is also quite nicely integrated.

Update: The Star Modeler was open sourced and is included in the latest versions of PDI/Kettle. Disregard the paragraph below.
*The plug-in architecture of PDI allows to add new features quite easily. The **Star Modeler** can be downloaded from he (status: 2011-11-04). Unzip the file in the PDI plug-ins folder and (re)start Spoon. The Star Modeler will now show up a an additional perspective on the top right hand side.*
***Note**: As said, this feature is still in development, so not everything might work as expected.*

Getting Ready

My Blog List

osbi.fr (Open Source Business Intelligence)
2020
3 weeks ago

Pedro Alves on Business Intelligence
Pentaho 9.1 is available!
3 months ago

Roland Bouman's blog
Building a UI5 Demo for SAP HANA Text Analysis: Part 4
1 year ago

Jens Bleuel about Kettle aka Pentaho Data Integration (PDI) & Pentaho BI
Pentaho Community Contributions (#KCM19)
1 year ago

Matt Casters on Data Integration
Calculate unique values in parallel for Neo4j Node creation
2 years ago

Gretchen and the Pentaho Nation
HAPI FHIR Example Server Using Docker on AWS
2 years ago

Bekwam Data as a Service
Setting the Pivot Point in a JavaFX ScaleTransition
3 years ago

Bens Blog
Pentaho Data Integration - Multi-part Form submission with file upload using the User Defined Java Class Step
3 years ago

Reporting Tales - Inside news on Pentaho Reporting
Your're hot, Thermometer chart!
5 years ago

Reporting Tales
Your're hot, Thermometer chart!
5 years ago

Rama's Free Thoughts
Building a Data Mart with Pentaho Data Integration - Video Course Review
6 years ago

Julian Hyde on Open Source OLAP. And stuff.
Table macros
6 years ago

Michael Tarallo - Open Source BI Guru
My New Blog at http://michaeltarallo.com/
7 years ago

Adventures with Open Source BI
Exploring the sample plugins for PDI
8 years ago

Pentaho Musings
I'm Off!

Let's create our source tables, so that we have some data to work with. Run the following SQL statements in your favour SQL client:

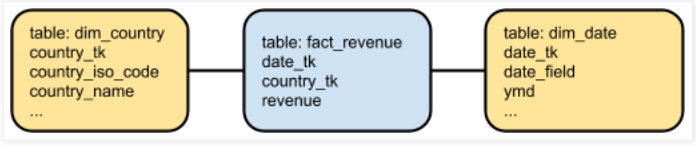
```
CREATE SCHEMA
source_db
;
USE
source_db
;
CREATE TABLE
revenue
(
date DATE,
country_iso_code VARCHAR(3),
revenue DECIMAL
)
;
INSERT INTO
revenue
VALUES
('2011-11-01','GB',22314),
('2011-11-02','GB',23411),
('2011-11-03','GB',22325),
('2011-11-04','GB',22233),
('2011-11-01','US',32423),
('2011-11-02','US',25325),
('2011-11-03','US',43523),
('2011-11-04','US',23453)
;


CREATE TABLE
Countries
(
country_iso_code VARCHAR(3),
country_name VARCHAR(100)
)
;
INSERT INTO
Countries
VALUES
('GB','United Kingdom'),
('US','United States of America')
;
```

Now let's create our target database:

```
CREATE SCHEMA
target_db
;
```

Our aim is to create this rather simple star schema:



Defining Database Connection Details

Start **Spoon**. Currently the Star Modeler sources database information from the **shared.xml** file. So if you haven't locally share any of your database details yet, create an empty transformation and specify two database connection:

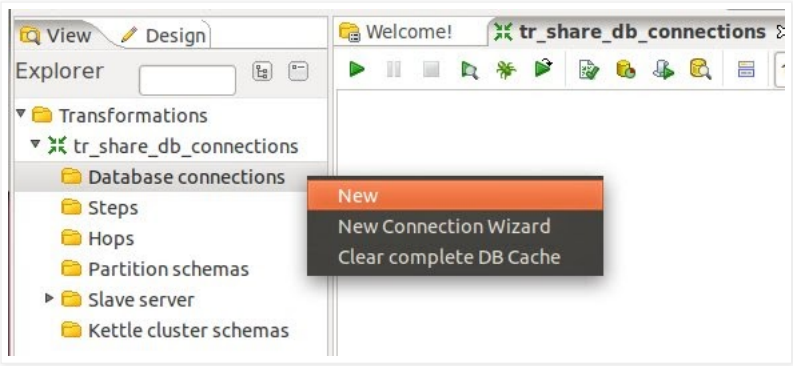
- **source_db**
- **target_db**

9 years ago

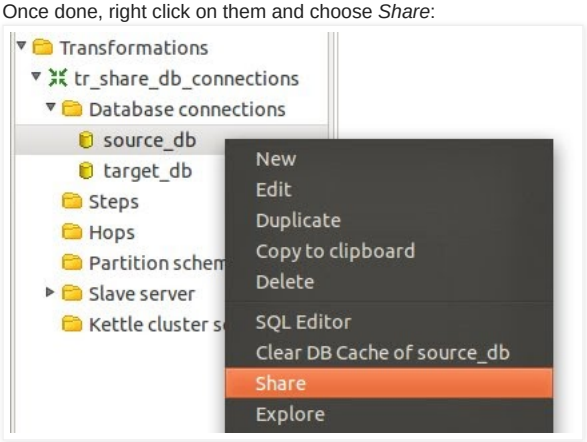
(OS)BI
Pentaho Kettle - BIRT Output Step
9 years ago

Analyse This

Click on the *View* tab and right click on *Database Connections*. Click on *New* then. Fill out all the details for both connections.



Once done, right click on them and choose *Share*:

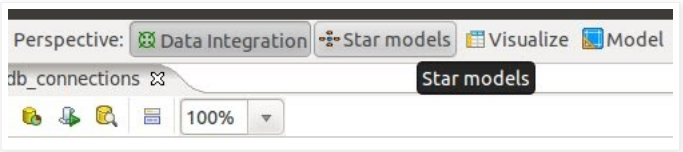


Note: There will be a better way to define database connection details to be used with the **Star Modeler** in future.

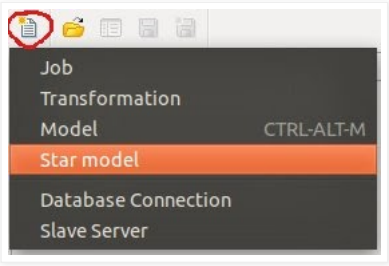
Save your transformation.

How to Create a Star Model

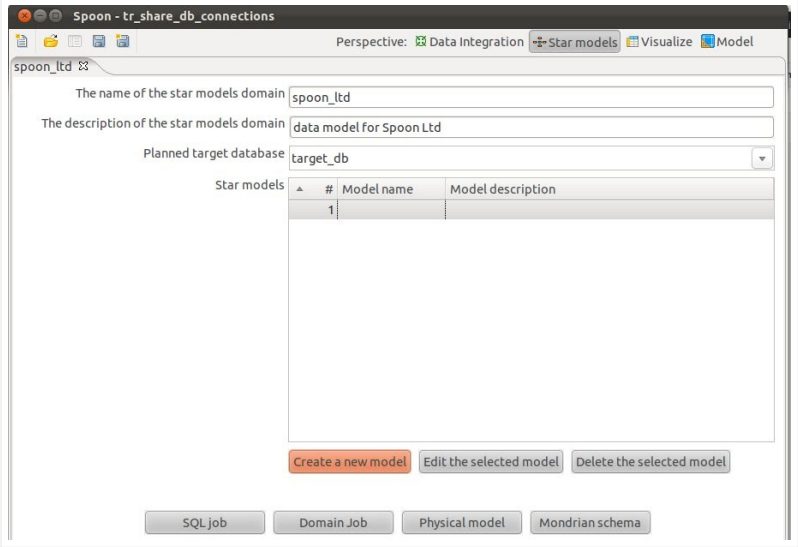
Now that we have the connection details defined, let's click on the *Star Models* perspective in the right top hand side corner:



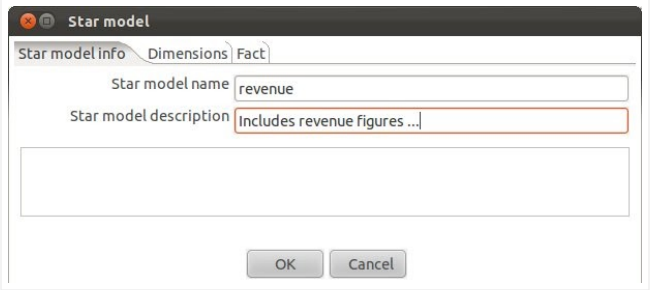
You will be greeted by a blank screen. Now click on the *New* icon and choose *Star Model*:



Give the star model a name and description. Then choose *target_db* as our target database.

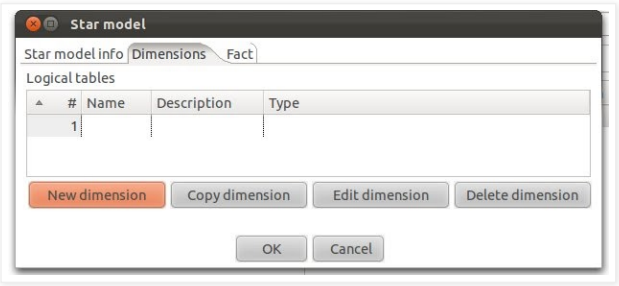


Click on *Create New Model* and fill out the form as shown below:

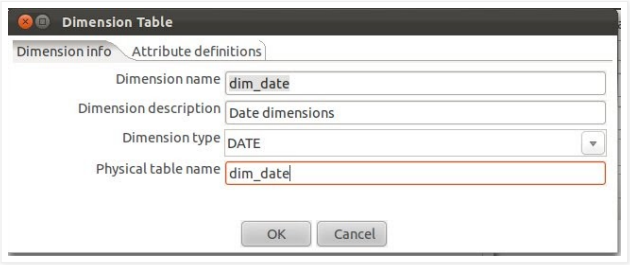


Creating Dimensions

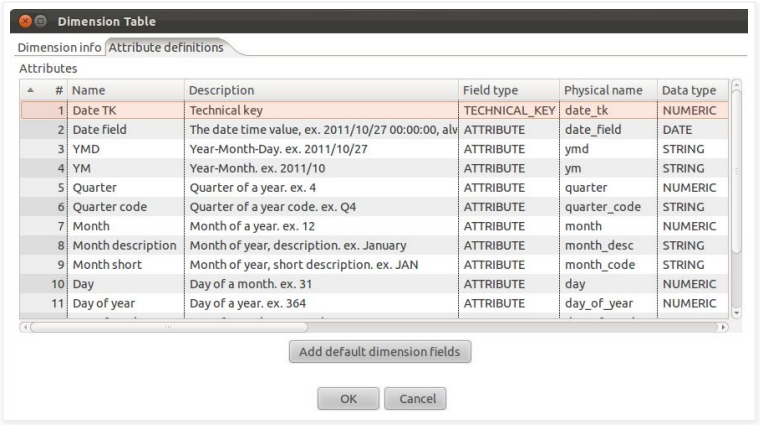
Next click on the *Dimensions* tab and click on *New Dimension*:



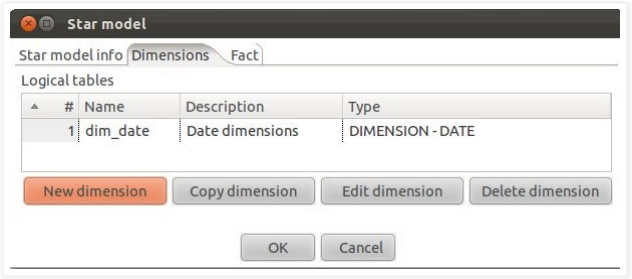
Let's create our date dimension: Fill out as shown below. Make sure that you choose **DATE** as *Dimension Type*.



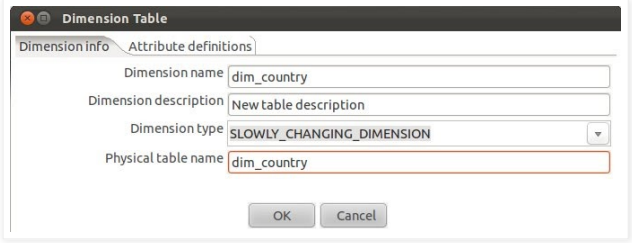
Now click on the *Attributes definitions* tab and then on *Add default dimension fields*:



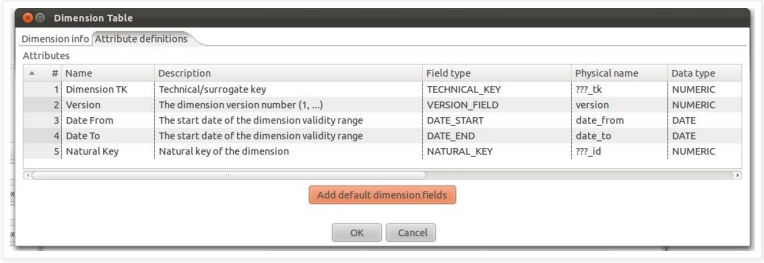
You will realize that PDI proposes a quite complete date dimension structure out-of-the-box. What a time saver! We are f
now quite happy with this (if you want, you can change this), and click on *OK*.
So we have now defined our date dimension. Let's go ahead and work on our country dimension:
Click on *New Dimension*:



Define the following (Make sure you choose **SLOWLY_CHANGING_DIMENSION** for *Dimension Type*):



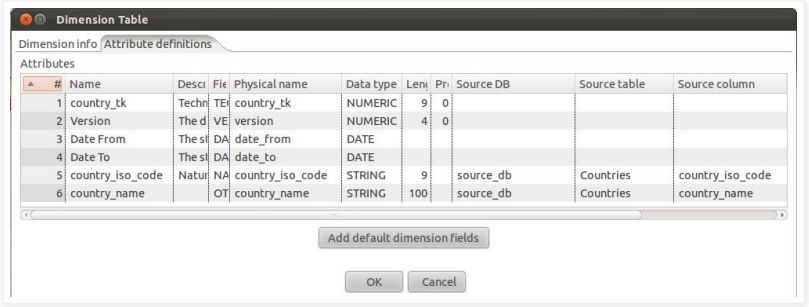
Next click on the *Attributes definitions* tab and click on *Add default dimension fields*:



If you are familiar with **Ralph Kimball**'s slowly changing dimensions, you will realize that PDI makes really good suggestions. Just change:

- the *Physical name*s (replace the ??? with real names)
- the natural key to country_iso_code (data type: string) and add country_name as additional attributes.
- specify the *Source DB*, *Source table* and *Source column* for country_iso_code and country_name.

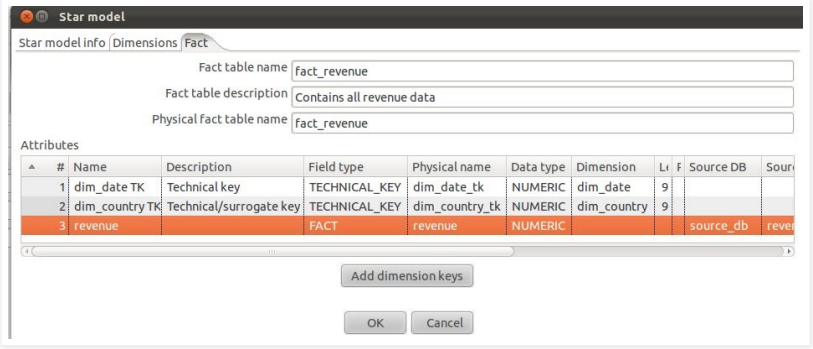
It should look then like this:



Click *OK*.

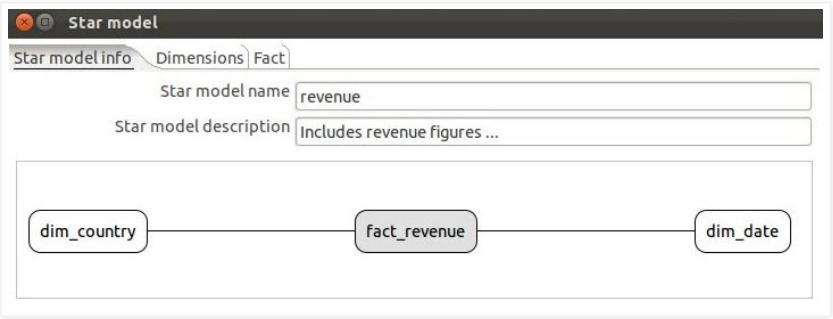
Creating a Fact Table

In the *Star Model* window click on the *Fact* tab and click on *Add dimension keys*. Now PDI automatically insert the techni keys of the **dimensions** we defined before.
Add an additional attribute called **revenue**, provide the details as shown below (make sure you also specify the **source**):



Our **fact table** is now properly defined.

Click on the *Start model info* tab and you will now see a simple graphical representation of our **star model**:



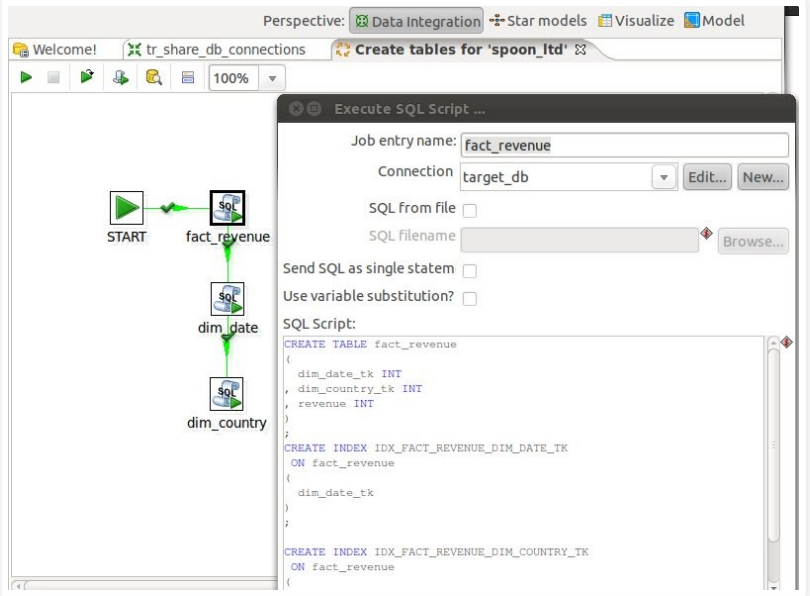
Click *OK*.

Automatic Generation Features

Now that our **star model** is defined, PDI gives us the option to automatically generate the following by just the click on a button:

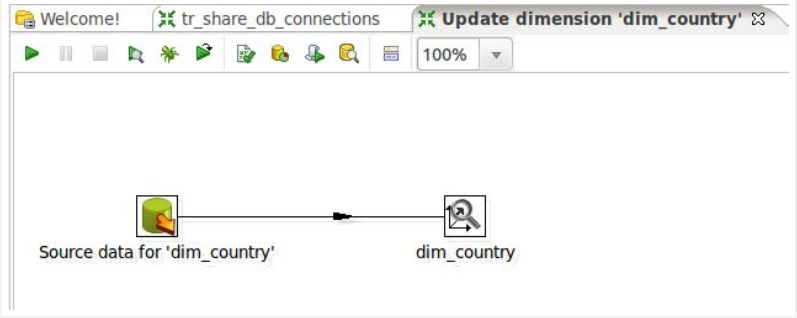
SQL DDL Job

This job will allow you to automatically create the DDL for the target database. Just click on *SQL Job* and a second later you will see a job like this:



Update Dimensions (Domain Job)

By clicking on *Domain Job* PDI will generate a simple transformation to update your dimensions:



Pentaho Metadata Model (Physical Model)

This feature is currently under development.

Pentaho Analyzer Model (Mondrian Schema)

This feature is currently under development.

Documentation

This feature is currently under development.

Conclusion

I hope that this simple tutorial demonstrated the huge potential of this new feature. It will certainly get better and better ov time, so watch the space!

Posted by Unknown at 3:41 PM

40 comments:

churtado November 4, 2011 at 4:43 PM

Hello Diethard,

One question, in the beginning of the tutorial it says pdi-ce-4.3.0-M1, but following the download link, the latest availa is 4.2.1. Is this the version that we can use?

Reply

Unknown November 5, 2011 at 2:52 AM

Thanks a lot Manolo for pointing this out! I changed the link now so that it points to CI.

Reply

gb December 6, 2011 at 8:02 AM

Can you suggest a build # that works? I have tried the latest good build but it doesn't run for me on os/x

Reply

Unknown December 6, 2011 at 11:18 AM

If you go the ci.pentaho.org you can download the latest PDI release which will have the star modeler already include so no need any more to install a plug-in.

Reply

Sunil January 17, 2012 at 10:42 PM

Hello Diethard,

Happy New Year. This looks to be great feature which will save the development time. Since it is possible to enhance the auto generated transformation this is really helpful.

I would like to ask you one question regarding connection pooling option in PDI. We are using Green Plum as database. Since Green plum has the max no of connections constraint it is observered that for optimal performance i good to enable connection pooling. But I have question regarding how to determine the no:of min and max connect

required. Say if a transformation is having one table I/P step, one table O/P step and one LookpUp step. Is Min of 1 a Max 3 is fine? Can you share your feedback in determining the min and max value for connection pooling.

Reply

Unknown January 18, 2012 at 2:36 AM

Thanks! Happy New Year to you as well! I think you are on the right path. Unless your pool is really limited, I would max a bit higher (just to be on the safe side).

Reply

Giako February 29, 2012 at 1:28 AM

Nice post! It's proving useful for my BI university project! :)

Reply

Unknown February 29, 2012 at 2:00 AM

Thanks a lot for your feedback! Much appreciated!

Reply

Osama Net March 24, 2012 at 5:24 AM

Hello Diethard
I have pdi 4.2.1 and I download the Star Modeler plug-in.
Where should I put Star Modeler plug-in to make pdi see it???

Reply

Replies

Unknown March 24, 2012 at 11:13 AM

You can just copy it in the plugin folder which resides within the PDI folder. I am not too sure though if this plugin is compatible with this version (4.2.1) though. Give it a try ... if it doesn't work, download a later version (ideally a very recent one where the plugin is already included).

Reply

Unknown March 24, 2012 at 11:13 AM

You can just copy it in the plugin folder which resides within the PDI folder. I am not too sure though if this plugin compatible with this version (4.2.1) though. Give it a try ... if it doesn't work, download a later version (ideally a v recent one where the plugin is already included).

Reply

ithackermike May 14, 2012 at 6:19 PM

I just loaded the Star Modeler plugin in PDI 4.2.1 on OSX. For clarity however I had to add the plugin to the spoon fol under plugins to get it to work.

Applications/pentaho/design-tools/data-integration/plugins/spoon/

This statement "Unzip the file in the PDI plug-ins folder and (re)start Spoon." above led to believe that it should go in plugins folder and not the spoon folder.

Hope that helps...

Reply

Unknown May 14, 2012 at 11:54 PM

Thanks a lot for pointing this out!

Reply

Unknown May 30, 2012 at 3:06 AM

Hi Dietmar,

thanks a lot for this tutorial.

Everything works fine until I want to load the fact table. I am using 4.3.0 and the plugin.
When i am trying to add the dimension keys follwoing error occured:
org.pentaho.metadata.model.LogicalRelationship.
(Lorg/pentaho/metadata/model/LogicalTable;Lorg/pentaho/metadata/model/LogicalTable;Lorg/pentaho/metadata/mode
ogicalColumn;Lorg/pentaho/metadata/model/LogicalColumn;)V
java.lang.NoSuchMethodError: org.pentaho.metadata.model.LogicalRelationsh
(Lorg/pentaho/metadata/model/LogicalTable;Lorg/pentaho/metadata/model/LogicalTable;Lorg/pentaho/metadata/mode
ogicalColumn;Lorg/pentaho/metadata/model/LogicalColumn;)V
at org.pentaho.di.starmodeler.StarModelDialog.getRelationshipsFromFact(StarModelDialog.java:591)...

After a while Spoon is crashing down.
I do not know how to fix it and getting crazy.

Reply

Replies

Unknown May 30, 2012 at 3:21 AM

This sounds quite serious. I've never come across this error. Does the same happen in version 4.4? Note, tha
with 4.4 you will not have to install the plugin as the Star Modeler is already included by default. If your problem
does not get resolved in version 4.4 and you do think that you didn't make an error somewhere, it's best to file
a JIRA case on jira.pentaho.com.
Best regards,
Diethard

Unknown May 30, 2012 at 8:10 AM

Thanks for reply.
Your link above is still going to the latest available version 4.2.1.
You know where to find 4.4?

Unknown May 30, 2012 at 8:24 AM

I just checked and the link above goes to sourceforge which has Kettle 4.3 stable. If you want the latest versio
- note: not to be used in production! - you can get it from ci.pentaho.org.

Unknown May 31, 2012 at 6:29 AM

Thanks a lot.
With 4.4 it runs without any problems.
Would be interesting where my mistake was...

Unknown May 31, 2012 at 6:35 AM

Good news! Maybe it was just a Kettle bug which got ironed out with version 4.4. When I initially wrote this
tutorial the plugin was only released for testing ... since then there have been several revisions.

simon03 June 14, 2012 at 1:19 AM

Hi Diethard! thank you for this tutorial!
I'm working with 4.4 version but I have a problem when I try to "Generate transformations" (the user interface in
4.4 is a little different, I think that is the same action of the "Domain job" button in this tutorial).
The error that occurred is:

org.pentaho.di.core.exception.KettleXMLException:
Error reading information from input stream
InputStream cannot be null

at org.pentaho.di.core.xml.XMLHandler.loadXMLFile(XMLHandler.java:588)
at org.pentaho.di.trans.TransMeta.(TransMeta.java:2387)
at org.pentaho.di.starmodeler.generator.JobGenerator.generateDateTransformation(JobGenerator.java:390)
at
org.pentaho.di.starmodeler.generator.JobGenerator.generateDimensionTransformations(JobGenerator.java:36
8)
at
org.pentaho.di.starmodeler.StarModelerPerspective.generateDomainJobButton(StarModelerPerspective.java:7
34)
at org.pentaho.di.starmodeler.StarModelerPerspective$11.widgetSelected(StarModelerPerspective.java:631)
at org.eclipse.swt.widgets.TypedListener.handleEvent(Unknown Source)
at org.eclipse.swt.widgets.EventTable.sendEvent(Unknown Source)
at org.eclipse.swt.widgets.Widget.sendEvent(Unknown Source)
at org.eclipse.swt.widgets.Display.runDeferredEvents(Unknown Source)
at org.eclipse.swt.widgets.Display.readAndDispatch(Unknown Source)

at org.pentaho.di.ui.spoon.Spoon.readAndDispatch(Spoon.java:1186)
at org.pentaho.di.ui.spoon.Spoon.start(Spoon.java:7031)
at org.pentaho.di.ui.spoon.Spoon.main(Spoon.java:580)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:39)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:25)
at java.lang.reflect.Method.invoke(Method.java:597)
at org.pentaho.commons.launcher.Launcher.main(Launcher.java:134)
Caused by: java.lang.IllegalArgumentException: InputStream cannot be null
at javax.xml.parsers.DocumentBuilder.parse(DocumentBuilder.java:120)
at org.pentaho.di.core.xml.XMLHandler.loadXMLFile(XMLHandler.java:561)
... 18 more

I double cecked every step but I can't find out the mistake.
Thanks!

Unknown June 14, 2012 at 1:28 AM

I can try to have a look at your file - no promises, I don't know yet when I will find time. Can you make i available online somewhere (Google Drive, Dropbox, etc)?

Reply

simon03 June 14, 2012 at 2:02 AM

Ok! I've shared the star model file at this link:
https://docs.google.com/open?id=0B13bSoiYs-qKeHd1N0ZuaVEwaFU

Thank you!

Reply

Replies

Unknown June 14, 2012 at 2:11 AM

Thanks, I try to take a look at it in the next few days.

Unknown June 17, 2012 at 10:45 AM

Hi Simon,
I tried to open your file in the latest PDI version (pdi-ce-4.4.0-M1-r16933) on Ubuntu and everything is blank. opened your file then in a text editor and can see that there is definitely useful stuff in there. What exact versio are you working with (and which OS)? Maybe there are indeed some bugs with the current version of PDI i which case we should set up some JIRA cases on http://jira.pentaho.com/.
Best regards,
Diethard

simon03 June 18, 2012 at 12:08 AM

Hi Diethard,
I'm working with pdi-ce-4.4.0-M1-r16929 version on Windows7.

Thanks!

Unknown June 18, 2012 at 12:30 AM

I set up this JIRA case: http://jira.pentaho.com/browse/PDI-8018. Once this issue is solved I can look at you file.

Reply

KRISHNAMOORTHY February 24, 2013 at 10:44 PM

In Pentaho DI 4.4 , if i try to add fact table using star model , I am getting the following error continuously..Kindly help out

An unexpected error occurred in Spoon:
org.pentaho.metadata.model.LogicalRelationship.
(Lorg/pentaho/metadata/model/LogicalTable;Lorg/pentaho/metadata/model/LogicalTable;Lorg/pentaho/metadata/mode ogicalColumn;Lorg/pentaho/metadata/model/LogicalColumn;)V

java.lang.NoSuchMethodError: org.pentaho.metadata.model.LogicalRelationsh
(Lorg/pentaho/metadata/model/LogicalTable;Lorg/pentaho/metadata/model/LogicalTable;Lorg/pentaho/metadata/mode
ogicalColumn;Lorg/pentaho/metadata/model/LogicalColumn;)V
at org.pentaho.di.starmodeler.StarModelDialog.getRelationshipsFromFact(StarModelDialog.java:591)
at org.pentaho.di.starmodeler.StarModelDialog.drawLogicalModel(StarModelDialog.java:645)
at org.pentaho.di.starmodeler.StarModelDialog$6.paintControl(StarModelDialog.java:285)
at org.eclipse.swt.widgets.TypedListener.handleEvent(Unknown Source)
at org.eclipse.swt.widgets.EventTable.sendEvent(Unknown Source)
at org.eclipse.swt.widgets.Widget.sendEvent(Unknown Source)
at org.eclipse.swt.widgets.Widget.sendEvent(Unknown Source)
at org.eclipse.swt.widgets.Widget.sendEvent(Unknown Source)

Reply

Replies

Unknown February 24, 2013 at 11:52 PM

Please create a bug report on jira.pentaho.com

Reply

Nemo May 6, 2013 at 1:46 AM

Hello and thanks for this great tutorial. I want to create a star model but when i download the 4.4.2 version of the PD
doesn't include the starmodeler plugin and when i install it it gives the NoSuchMethod error which is an unresolved bu
was wondering when can i get the exact same PDI tha you worked with in this tutorial. Thanks in advance.

Reply

Replies

Unknown May 6, 2013 at 12:11 PM

Thanks for your feedback! The latest version you can find on ci.pentaho.org ... it's included in the latest PDI ..
just note: do not use this version for production purposes!

Reply

giac_man July 14, 2013 at 9:55 AM

Hi Diethard,

I am a little bit confused. I have PDI version 4.4.1 installed and you mentioned that the latest version (in 2011) sho
have the star modeller already included. I can't find it on this version. Am I missing something?

I have installed the plug in and everything works until I load the fact table. I got the same error as KRISHNAMOORT
and then kettle crash.

Can anyone help?

Many thanks!

Reply

Replies

Unknown July 15, 2013 at 1:32 AM

Thanks a lot for your feedback! In the latest versions the Star Modeler is available as a plugin via the
marketplace. If you get errors, please submit a jira case on jira.pentaho.org.

giac_man July 15, 2013 at 3:48 AM

Hi,

There is already a case in Jiira for this error. Is the one created by KRISHNAMOORTHY.

Marketplace is not present in my Pentaho version 4.4.1 - I am not sure I have the latest version of Sta
Modeler, could you send me a link?

Thanks a lot

Giacomo

Unknown July 15, 2013 at 5:52 AM

The only link I have is the one posted in this blog post. The last thing I heard about the star modeler was that i will be available in version 5 as a plugin via the marketplace and that in future the community will take over the development of the plugin. Maybe you can ask the question in the Pentaho Kettle forum.

Unknown July 15, 2013 at 9:05 AM

Actually, you can find the star modeler in the agile bi version of Kettle which you can download from ci.pentaho.org (please note that this is in active development and should not be used for production purposes)

giac_man July 16, 2013 at 3:01 AM

Thanks a lot!

Reply

Unknown November 25, 2013 at 8:10 AM

Hi, Pentaho website says this plugin is in hold status. Its not available in data integration 5.0.1 build nor in marketplace. However I did download the zip and install in latest version.
while it gives all the errors above...it is still possible to save the file and generate jobs by using 'save as' after closing error panels.
This is a really useful plugin and hope Pentaho incorporates this in future.
Thanks.

Reply

Replies

Unknown November 25, 2013 at 8:16 AM

Yes, this plugin is unfortunately not maintained any more by Pentaho. Now it's down to somebody from the community to continue development on it.

Reply

Unknown February 2, 2014 at 4:01 PM

So, did anybody finally continue development con this plugin? Is pentaho thinking about developing something similar?

Reply

Replies

Unknown February 2, 2014 at 11:54 PM

Good question ... I haven't heard about anything. The idea/hope was that someone from the community would pick up development. I doubt that Pentaho will develop something similar ... they could have just continued working on this plugin then. If you are a Java developer, you might want to have a look yourself at the source code and maybe you can then add some new functionality.

Reply

Enter your comment...

Comment as: Flavio Cardoso Sign out

Publish Preview Notify me

Newer Post Home Older Po

Subscribe to: Post Comments (Atom)

Simple theme. Powered by Blogger.