

Navigating Through the Pentaho Kettle Interface

David Amadi

What is a Pentaho Kettle Repository?

The Pentaho Kettle Repository is a storage place for the ETL metadata like the transformations, jobs, database connections, etc.

Repository is of two types:

- Kettle Database Repository: A repository is a relational database specially designed for this purpose
- Kettle File Repository: The files method consists of saving jobs and transformations as regular XML files in the file system, with extension .kjb and .ktr respectively.

note:

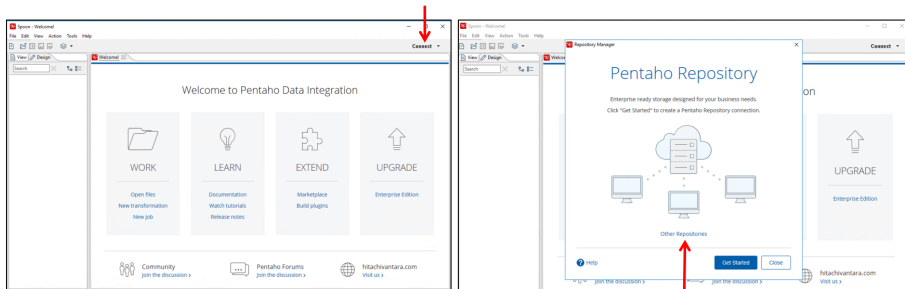
- We shall adopt the “Kettle File Repository” method and will store the transformations and jobs in the local harddrive

Creating a Pentaho Kettle repository

- **Session 1 : Duration 10 minutes**

Step 1

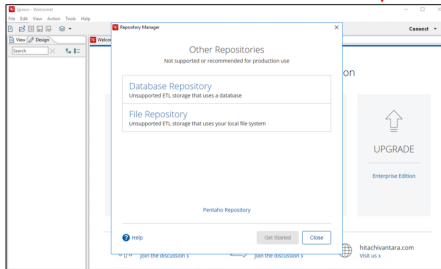
1. Click on Connect



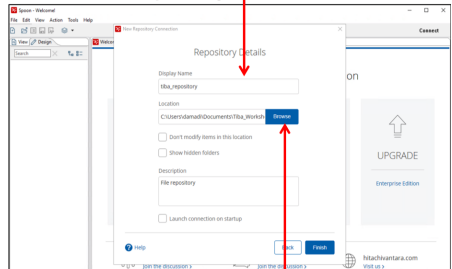
2. Click on Other Repositories

Step 2

3. Click on File Repository



4. Provide the Name for the repository



5. Choose Kettle file location for the repository

Connecting to a database

- **Session 2: Duration 15 minutes**

Step 1

2. Click view tab

1. Open / Start a Transformation or Job

3. Right click Database Connections And select New

Drag & Drop a Step
Also try shift + double-click

Step 2

Connecting to MySQL database server

The screenshot shows the 'Database Connection' dialog box with the following components and annotations:

- 1. Select Connection type:** Points to the 'Connection type' list on the left, where 'MySQL' is selected.
- 2. Select driver:** Points to the 'Access' list on the left, where 'Native (JDBC)' is selected.
- 3. Provide a suitable connection name:** Points to the 'Connection name' field, which contains 'cleaning'.
- 4. Provide the name of the server where database is installed:** Points to the 'Host Name' field, which contains 'localhost'.
- 5. Provide the name of database:** Points to the 'Database Name' field, which contains 'tiba'.
- 6. Default MySQL port:** Points to the 'Port Number' field, which contains '3306'.
- 7. Provide the user name of instance:** Points to the 'Username' field, which contains 'root'.
- 8. Enter password for the instance:** Points to the 'Password' field, which is masked with dots.
- 9. Test the connection:** Points to the 'Test' button at the bottom of the dialog.
- 10. If test successful, click ok to create connection:** Points to the 'OK' button in the 'Connection tested successfully' dialog box.

The 'Settings' section on the right includes a checkbox for 'Use Result Streaming Cursor'.

The 'Connection tested successfully' dialog box displays the following information:

- Hostname: localhost
- Port: 3306
- Database name: tiba

Step 3

Sharing the database connection so that all transformations and jobs can use it

The screenshot shows the Apache Kettle Spoon interface. The title bar reads "Spoon - [Tiba_repository] Transformation 1 (changed)". The left sidebar shows a tree view with "Transformations" expanded, containing "Transformation 1", "Run configurations", and "Database connections". Under "Database connections", a connection named "cleaning" is selected. A right-click context menu is open over this connection, with the "Share" option highlighted. A red arrow points from the text "1. Right click on the specific database connection" to the "cleaning" connection. Another red arrow points from the text "2. Select Share" to the "Share" option in the menu. The main canvas shows a "Design" view with a "Big Data" folder containing "Cassandra" and "Hadoop" icons. A database icon is being dragged from the "Cassandra" folder to a step on the canvas. Below the canvas, the text "Drag & Drop a Step" is displayed, with a sub-note "Also try shift + double-click".

1. Right click on the specific database connection

2. Select Share

Drag & Drop a Step
Also try shift + double-click

Navigating through the Pentaho Kettle Interface

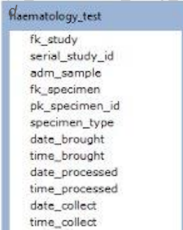
Session 3: Duration 30 minutes

- Creating a Transformation
- Executing a Transformation
- Creating and executing a Job

ER Diagram of Sample Database



pk_serial=serial_study_i



Kettle Transformation and jobs

- **Session 4 : Duration 30 minutes**

Acknowledgements



KEMRI | Wellcome Trust

