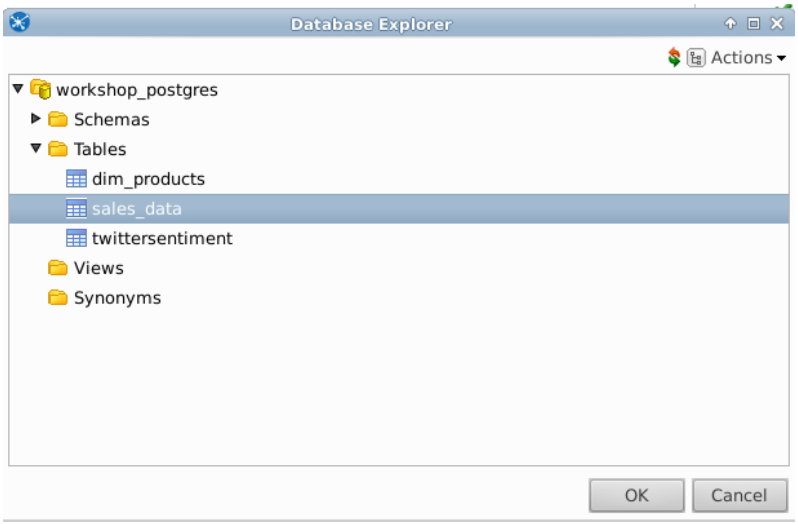
Passing URL Parameters to Kettle Transformations

PDI Remote Invocation

During this exercise we will demonstrate how to pass parameters to kettle jobs and run them via URLs. This is useful for accessing kettle jobs from other applications on a dynamic basis when scheduling does not give enough granularity.

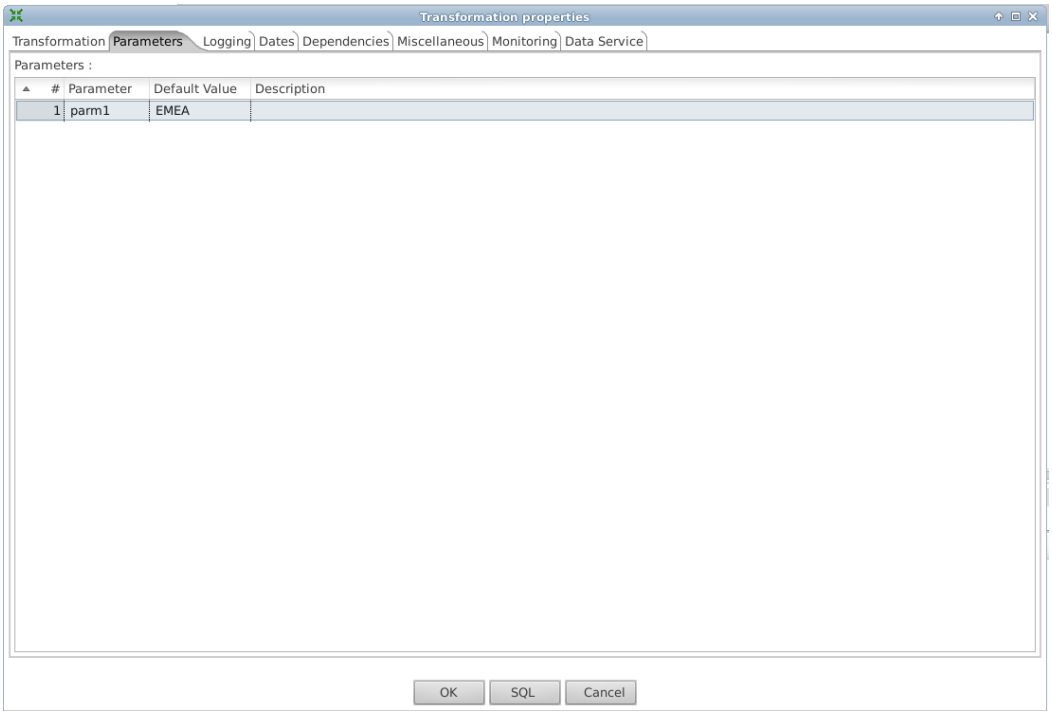
Build the Transformation

1. Open a terminal window and enter: cd ~/pentaho/
    a. Press enter
2. Enter the command: ./ctlscript.sh restart
    a. Press enter
3. This will restart DI server in case it is not currently running.
4. Go to Tools -> Repository -> Connect.
5. Select Workshop and enter “password” as the password. Press ok.
6. Open the pdi_remote_invocation_final located in the public/PDI Workshop for reference purposes.
7. Create a new transformation by selecting File ->New -> Transformation
8. Save your transfrom as pdi_remote_invocation by clicking File -> Save As. Save it into the admin folder.
9. Drag a Table input step onto the canvas from the input folder.
10. Select “workshop_postgres” as the data base Connection and press the Get SQL select statement button.
11. Navigate down to the sales_data table and press OK. Press Yes to include column names.



12. The table input step should look like this. Press OK.

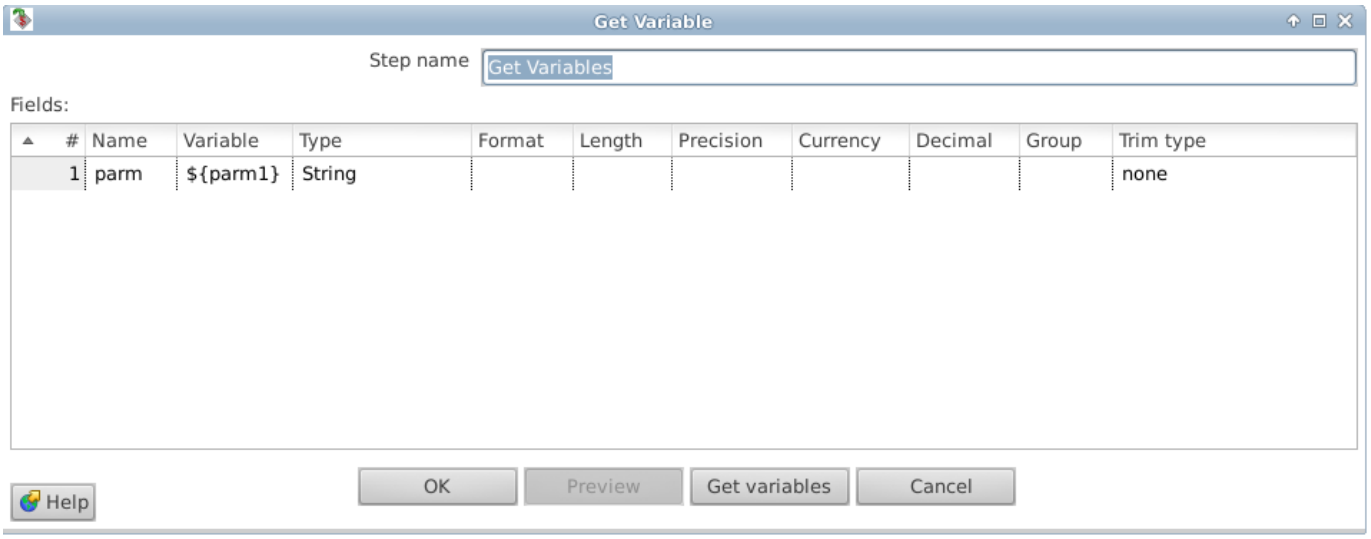
13. Right click anywhere on the canvas and select Transformation Properties.



14. On the Parameters tab, enter “parm1” with the Default Value of “EMEA”.

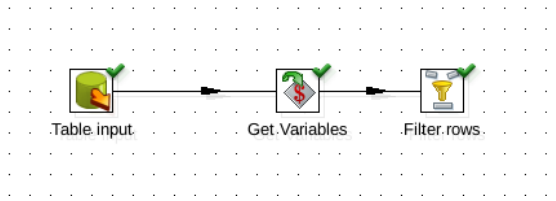
Filter by GET Parameter

1. Drag the Get variables step under the Jobs folder on the canvas and attach it to the Table input step. Enter the following information.



2. Drag a Filter rows step from the Flow folder onto the canvas and attach it to the Get variables step. Add the following information:

3. At this point the transformation should look something like this:



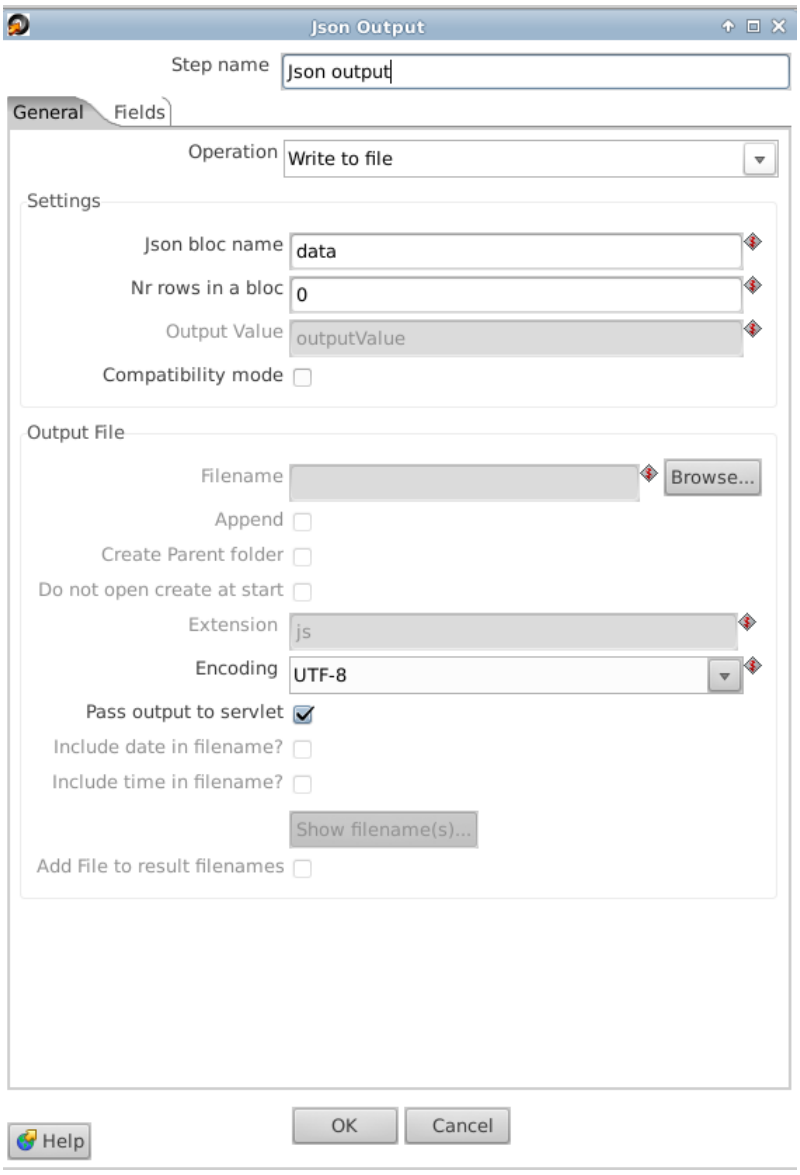
4. Next, add 2 Dummy steps to the canvas, one below the Filter rows step and one to the right of it.

5. Right click on the bottom dummy step and rename it to “No Match”. Right click on the other dummy step and rename it to “Matching Results”.

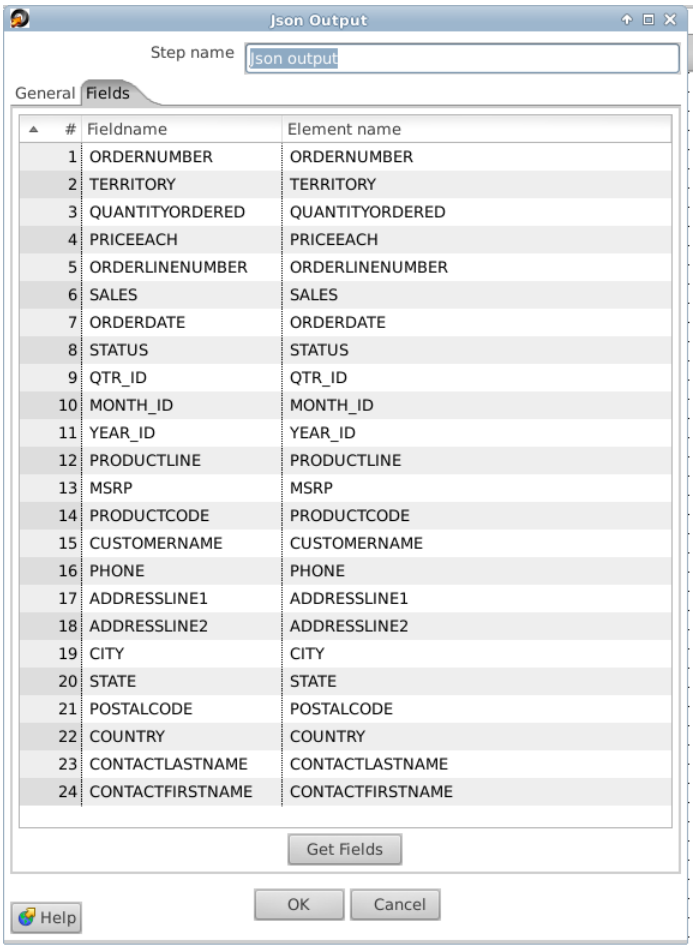
6. Connect the Filter step to No Match when Result is false and connect it to Matching Results when Result is True.

7. Add a JSON output step to the canvas and connect it to the Matching Results dummy step.

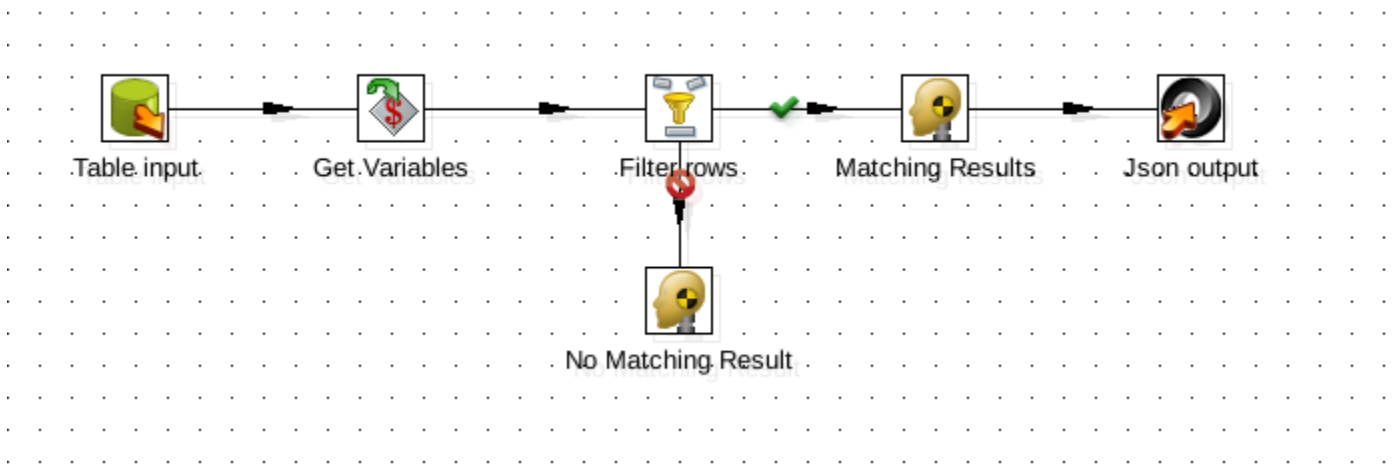
8. Update it as follows, making sure to check the servlet checkbox.



9. Click on the Fields tab of the JSON output step and press Get Fields.



10. The transformation should look like this. Save it and run it to preview data.



11. Navigate to the following URL: http://localhost:9080/pentaho-di/kettle/executeTrans/?rep=Workshop&trans=/public/PDI%20Workshop/pdi_remote_invocation_final&parm1=APAC

    Enter admin/password if prompted for user name and password.

12. You will see JSON output for EMEA data. By changing parm1=NA in the URL, the data will update to show only North America Territory sales data.

Review Exercise #5

What We Covered…

- Created parameterized transformation for running from URL
  - JSON Output
  - Filter on Territory