Appendix A – Project Management

Mandatory Requirements

* Pentaho EE 7.1 or higher
* PostgreSQL
* pgAdmin III
* Git available through Window’s command prompt

In this module, you are simulating a PDI project for an online Retailer. There are 4 environments:

|  |  |
| --- | --- |
| ENVIRONMENT | REPOSITORY |
| Local | File repository / EE repository |
| Development | EE repository |
| UAT | EE repository |
| Production | EE repository (not configured) |
|  |  |
| Logging | Central logging repository for each environment |

pgAdmin III

Database admin tool for PostgreSQL.

The software can be downloaded from:

<https://www.dropbox.com/sh/6nl31ts10sjimnr/AADFXjTek4f9ANyBivVVAhqFa?dl=0>

or

<https://www.pgadmin.org/download/pgadmin-3-windows/>

You will need to unzip the file.

To install just double-click on the pgadmin3.msi and follow the default settings.

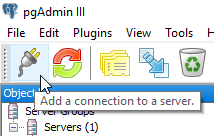
Scripts for creating databases, login roles, schemas and tables can be found in the above Dropbox link:

.\Artefacts for DI 1100\Module 1 - Project Management\Lesson 1 - PDI Project Management\scripts

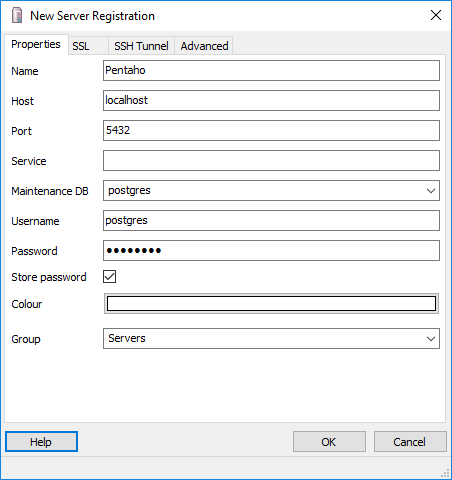
PostgreSQL Databases

Pentaho ships with PostgreSQL. You will first need to create 3 databases for each environment.

1. Highlight the Servers item
2. Click on the connect icon, to define a connection



1. Configure the connection as outlined below:

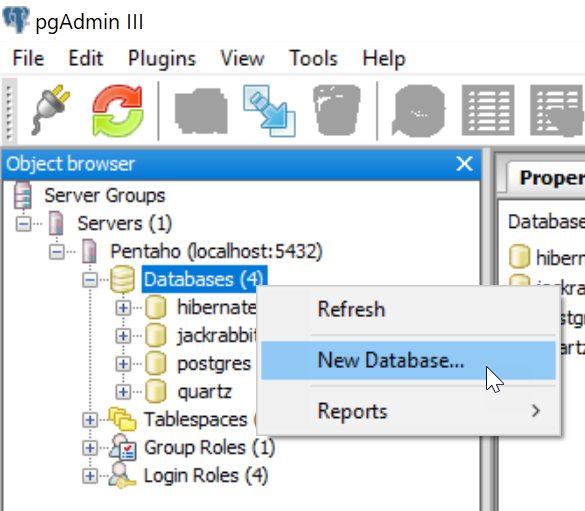


* Password: password

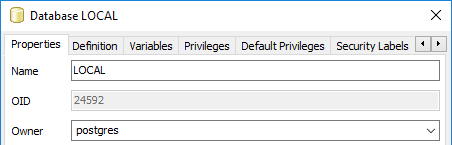
1. Click OK

To create database:

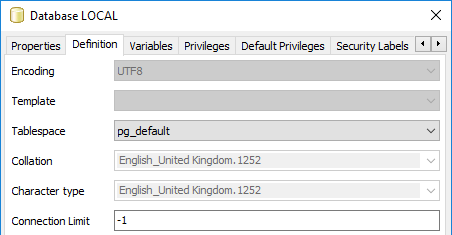
1. Highlight Databases, right mouse click and select the option: New Database



1. Let’s start with Local



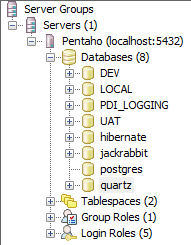
1. Click on the Definition tab and select Tablespace: pg\_default



1. Click OK.

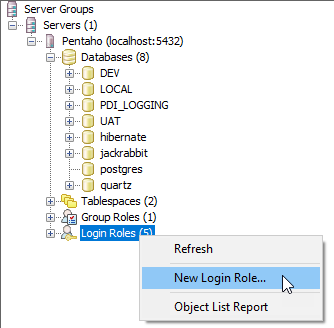
Repeat the workflow to create the other databases:

* DEV
* UAT

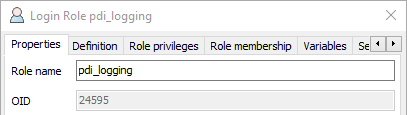


The final database is for logging. First define a new role

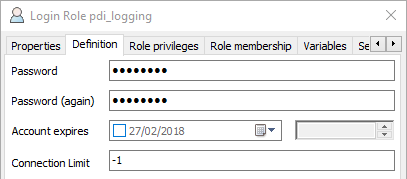
1. Highlight Login Roles and right mouse click, then New Login Role



1. The role is: pdi\_logging

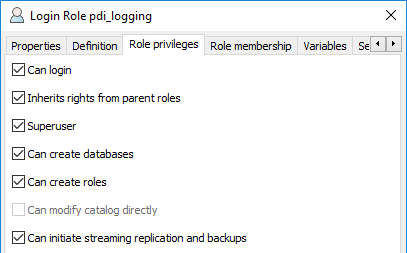


1. Click on Definition tab and Enter the password



* The Password: password

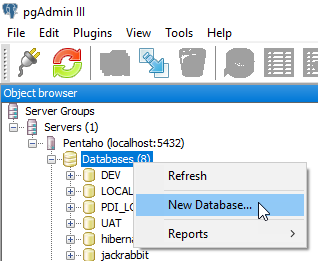
1. Click OK.
2. Click on Role privileges tab



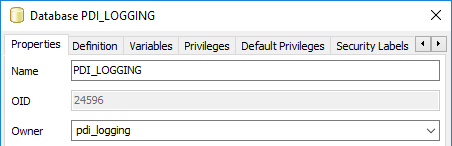
1. Click OK

To create the Logging Database:

1. Highlight Databases, right mouse click and select the option: New Database



1. Define PDI\_LOGGING



1. Click OK

Schemas & Tables

Next is to create the schemas and tables for each of the environments.

Each database will have 2 schemas

* pdi\_control – a table that logs each step of the Jobs with success / fail status
* sales\_dwh – DWH with 1 fact table & 1 dimension table

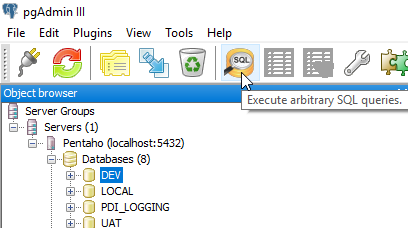
and 3 tables

* job\_control – success / fail status
* fact\_sales – Fact table
* dim\_country – dimension table

1. Run the following script on each database:

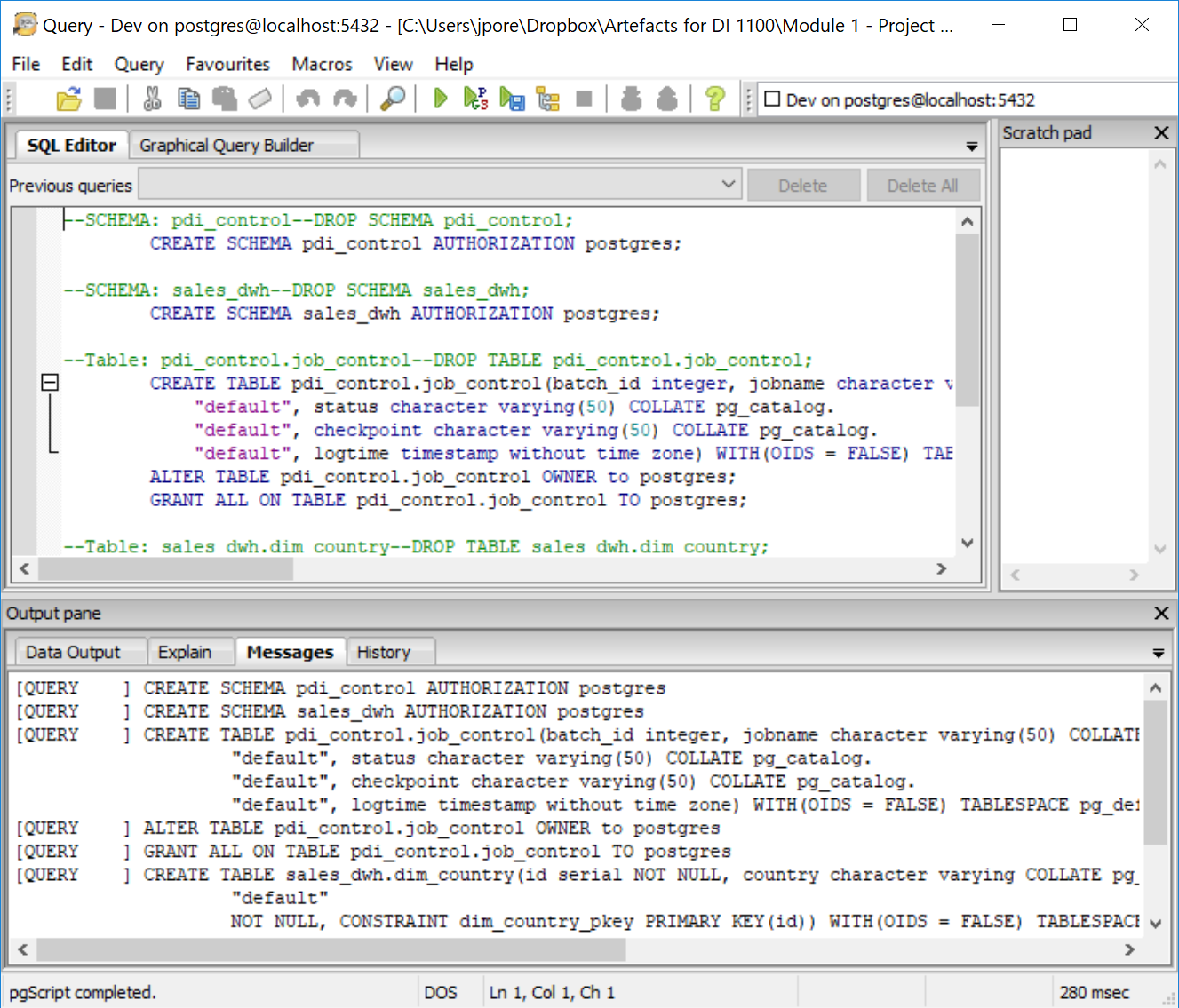
.\Artefacts for DI 1100\Module 1 - Project Management\Lesson 1 - PDI Project Management\scripts\ pentaho\_project\_schemas\_tables.sql

1. Highlight the Dev database and click on the SQL icon in main menu.



1. Browse and open the following script:

.\Artefacts for DI 1100\Module 1 - Project Management\Lesson 1 - PDI Project Management\scripts\pentaho\_project\_schemas\_tables.sql



Execute the script by clicking on the 

1. For the PDI\_Logging table, execute the script:

.\Artefacts for DI 1100\Module 1 - Project Management\Lesson 1 - PDI Project Management\scripts\pentaho\_project\_logging\_tables.sql

1. Check that all the relevant Schemas and Tables have been created.

GIT

To setup the Project folders, you will run a script that connects to a GitHub repository and pulls down the required artefacts.

You will first need to install Git

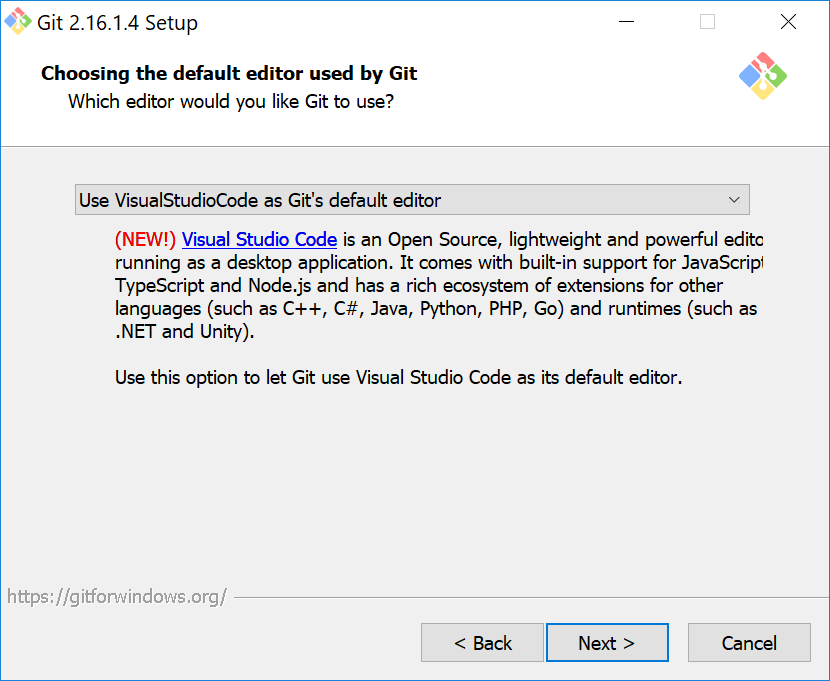
The software can be downloaded from:

<https://www.dropbox.com/sh/6nl31ts10sjimnr/AADFXjTek4f9ANyBivVVAhqFa?dl=0>

or

<https://gitforwindows.org/>

1. To install just double-click on the Git-2.16.1.4-64-bit.exe and follow the default settings.
2. Select Visual Source Code from the dropdown options.



1. Keep all the default options then OK.

Project Folders

1. Browse to the folder:

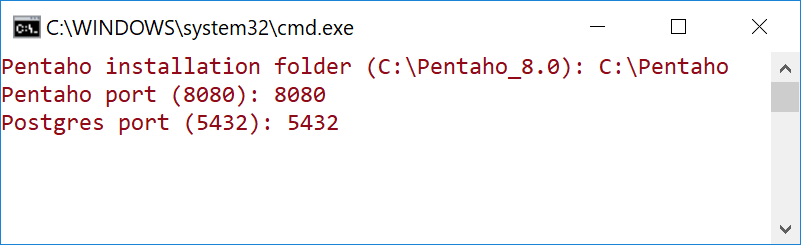
C:\Projects\online\_retailer

1. Execute the script:

C:\Projects\online\_retailer\install\_project\_setup.bat

The script will ask you for:

* The path to your Pentaho implementation
* The port of your Pentaho implementation
* The port of the PostgreSQL database

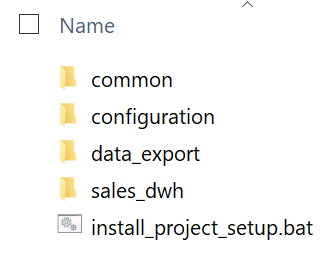


Wait until the script has finished. The command window will disappear.

1. Browse to the folder:
2. C:\Projects\online\_retailer

Four folders have been cloned from the GitHub repository:

|  |  |
| --- | --- |
| FOLDER | DESCRIPTION |
| common | common artefacts like PDI jobs, BA user custom reports, SQL and so on |
| configuration | configuration settings for every environment, e.g. kettle properties, connections, metastore and project properties |
| sales\_dwh | DWH Project |
| data\_export | Data export project |



Ok.. you should be good to go..!

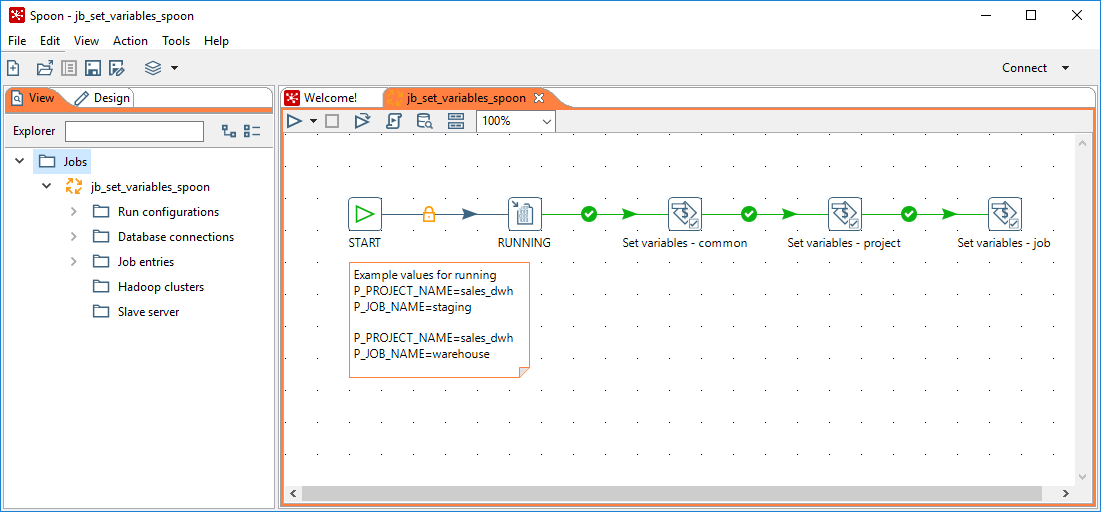
Local Repository

To test the implementation:

1. Browse to the folder:

C:\Projects\online\_retailer\configuration\config-pdi-local\

1. Execute: spoon.bat



The Orange frame indicates that you’re in the local environment. These settings have been set in the Options under Look & Feel tab.