Pentaho Data Integration

In the class and the data warehouse project, We are going to use PDI tool for creating star schema in postgres from raw point-of-sale data.

Pentaho Download link

Pentaho version 7.0

https://sourceforge.net/projects/pentaho/files/Data%20Integration/7.0/pdi-ce-7.0.0.0-25.zip/download

Supplementary softwares (required for the class and the Project):

1. Editor:-
   a. Windows: Notepad ++ , Download from here
      https://notepad-plus-plus.org/download/v7.5.9.html
   b. Mac: Textmate / Brackets/others, any editor except the default Textedit will do.

Pentaho Prerequisites

1. Java 8 or above.

   ***Note**: If you do not have java runtime in your system, download and install from here-

   https://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html

Pentaho Installation on Windows

1. Unzip the contents of the zip file into a folder - example: c:\apps\pentaho7
2. In that unzipped folder, there is a folder named "data-integration"
3. Navigate into the folder and locate the following
   a. Spoon.bat
   b. SpoonConsole.bat
   c. SpoonDebug.bat
4. Execute SpoonDebug.bat by double clicking it.

```
C:\WINDOWS\system32\cmd.exe
SpoonDebug is to support you in finding unusual errors and start problems.
-
This starts Spoon with a console output with the following options:
-
Pause after the termination?
(helps in finding start problems and unusual crashes of the JVM)
Pause? (Y=Yes, N=No, C=Cancel) Y
-
Set logging level to Debug? (default: Basic logging)
Debug? (Y=Yes, N=No, C=Cancel) Y
-
Redirect console output to SpoonDebug.txt in the actual Spoon directory?
Redirect to SpoonDebug.txt? (Y=Yes, N=No, C=Cancel)
```

At the prompt "Pause after the termination?" enter Y

At the prompt "Set logging level to Debug?" enter Y

At the prompt "Redirect console output….?" enter N

```
C:\WINDOWS\system32\cmd.exe                                    —    □    ×
SpoonDebug is to support you in finding unusual errors and start problems.
-
This starts Spoon with a console output with the following options:
-
Pause after the termination?
(helps in finding start problems and unusual crashes of the JVM)
Pause? (Y=Yes, N=No, C=Cancel) Y
-
Set logging level to Debug? (default: Basic logging)
Debug? (Y=Yes, N=No, C=Cancel) Y
-
Redirect console output to SpoonDebug.txt in the actual Spoon directory?
Redirect to SpoonDebug.txt? (Y=Yes, N=No, C=Cancel) N
-
Launching Spoon: "C:\apps\pdi-ce-7.0.0.0-25\data-integration\spoon.bat" /level:Debug
DEBUG: Using JAVA_HOME
DEBUG: _PENTAHO_JAVA_HOME=C:\Program Files\Java\jre1.8.0_131
DEBUG: _PENTAHO_JAVA=C:\Program Files\Java\jre1.8.0_131\bin\java.exe

C:\apps\pdi-ce-7.0.0.0-25\data-integration>"C:\Program Files\Java\jre1.8.0_131\bin\java.exe"  "-Xms1024m" "-Xmx2048m" "-
XX:MaxPermSize=256m" "-Dhttps.protocols=TLSv1,TLSv1.1,TLSv1.2" "-Djava.library.path=libswt\win64" "-DKETTLE_HOME=" "-DKE
TTLE_REPOSITORY=" "-DKETTLE_USER=" "-DKETTLE_PASSWORD=" "-DKETTLE_PLUGIN_PACKAGES=" "-DKETTLE_LOG_SIZE_LIMIT=" "-DKETTLE
_JNDI_ROOT=" -jar launcher\pentaho-application-launcher-7.0.0.0-25.jar -lib ..\libswt\win64  /level:Debug
Java HotSpot(TM) 64-Bit Server VM warning: ignoring option MaxPermSize=256m; support was removed in 8.0
```

**** Important/Common Error:

You may encounter the following error message

```
Error occurred during initialization of VM
Could not reserve enough space for 2097152KB object heap
Java HotSpot(TM) Client VM warning: ignoring option MaxPermSize=256m; support was removed in 8.0
```

Solution:-

First close the error terminal. Then, Open Spoon.bat file in Notepad++ (not in windows notepad). Go to line 94. You will see the following line:-

if "%PENTAHO_DI_JAVA_OPTIONS%"=="" set PENTAHO_DI_JAVA_OPTIONS="-Xms1024m" "-Xmx2048m" "-XX:MaxPermSize=256m"

Change it to:-

if "%PENTAHO_DI_JAVA_OPTIONS%"=="" set PENTAHO_DI_JAVA_OPTIONS="-Xms1024m" "-Xmx1024m" "-XX:MaxPermSize=256m"

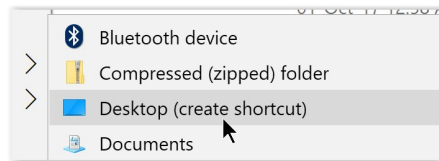Then, Re-run SpoonDebug.bat.

5. Observe the log output to see no errors are flagged until the Pentaho Splash Screen appears. Allow Network access if it asks to do so.
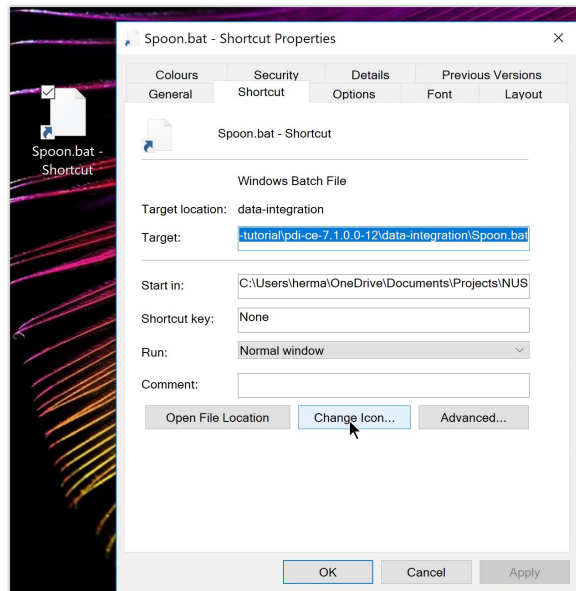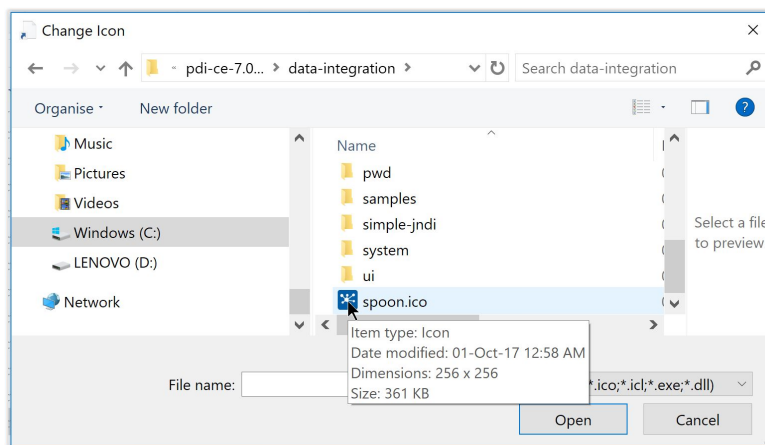

Set up startup icon

6. Right click on Spoon.bat,

Go to your desktop, right click on Spoon.bat icon

Change the icon by browsing to data-integration folder and select Spoon.ico
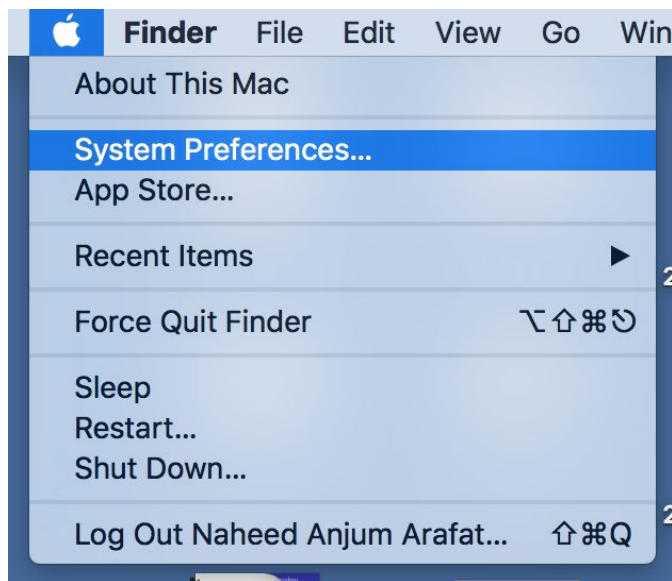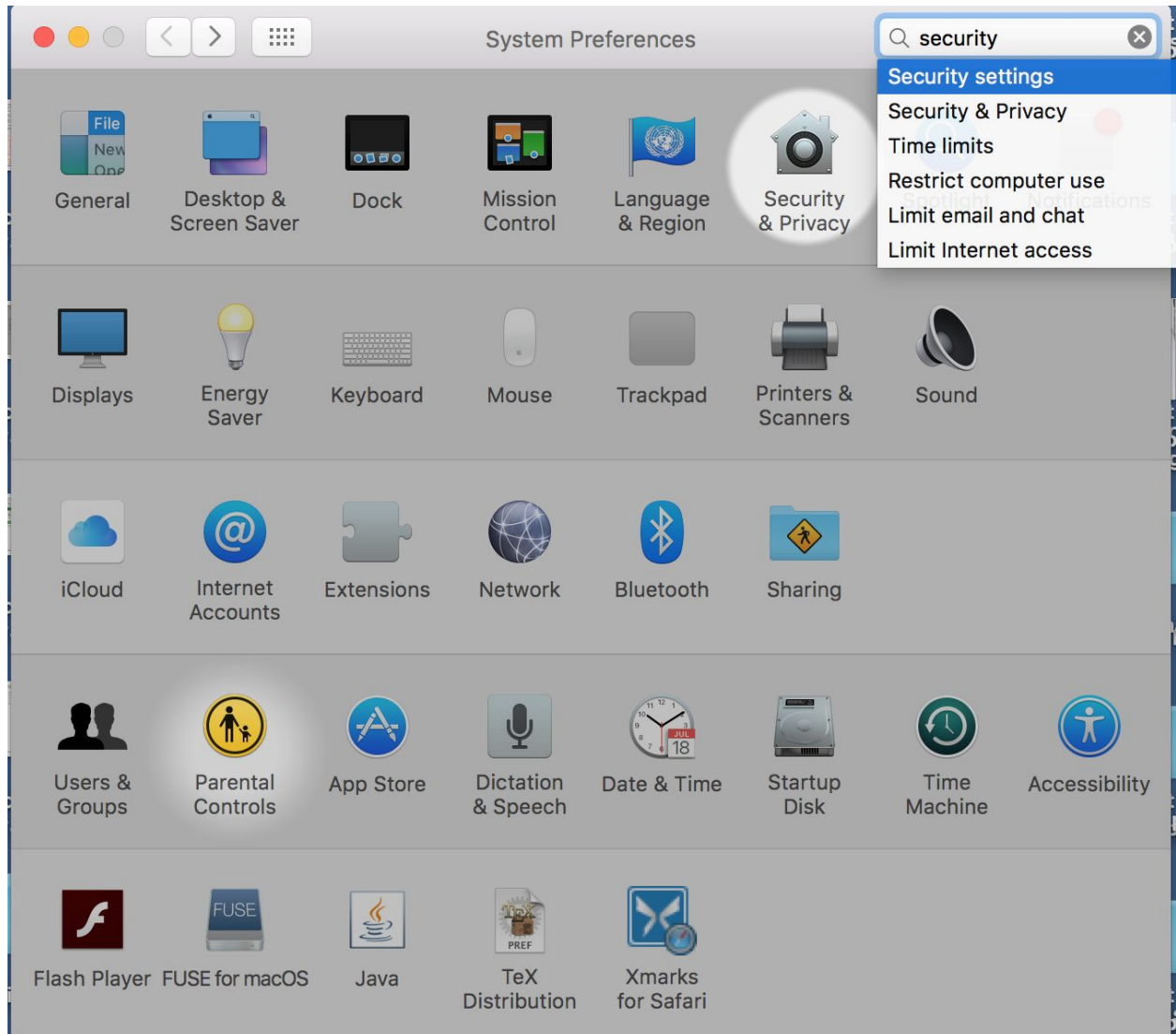
Pentaho Installation on Mac

1. After unzipping the downloaded file copy the unzipped "Data Integration" folder into "Applications".

2. Run "Data Integration" application from the Launcher.
3. You might need to change the security settings to be able to run the file manually. Follow the instructions below only if you need to do so.
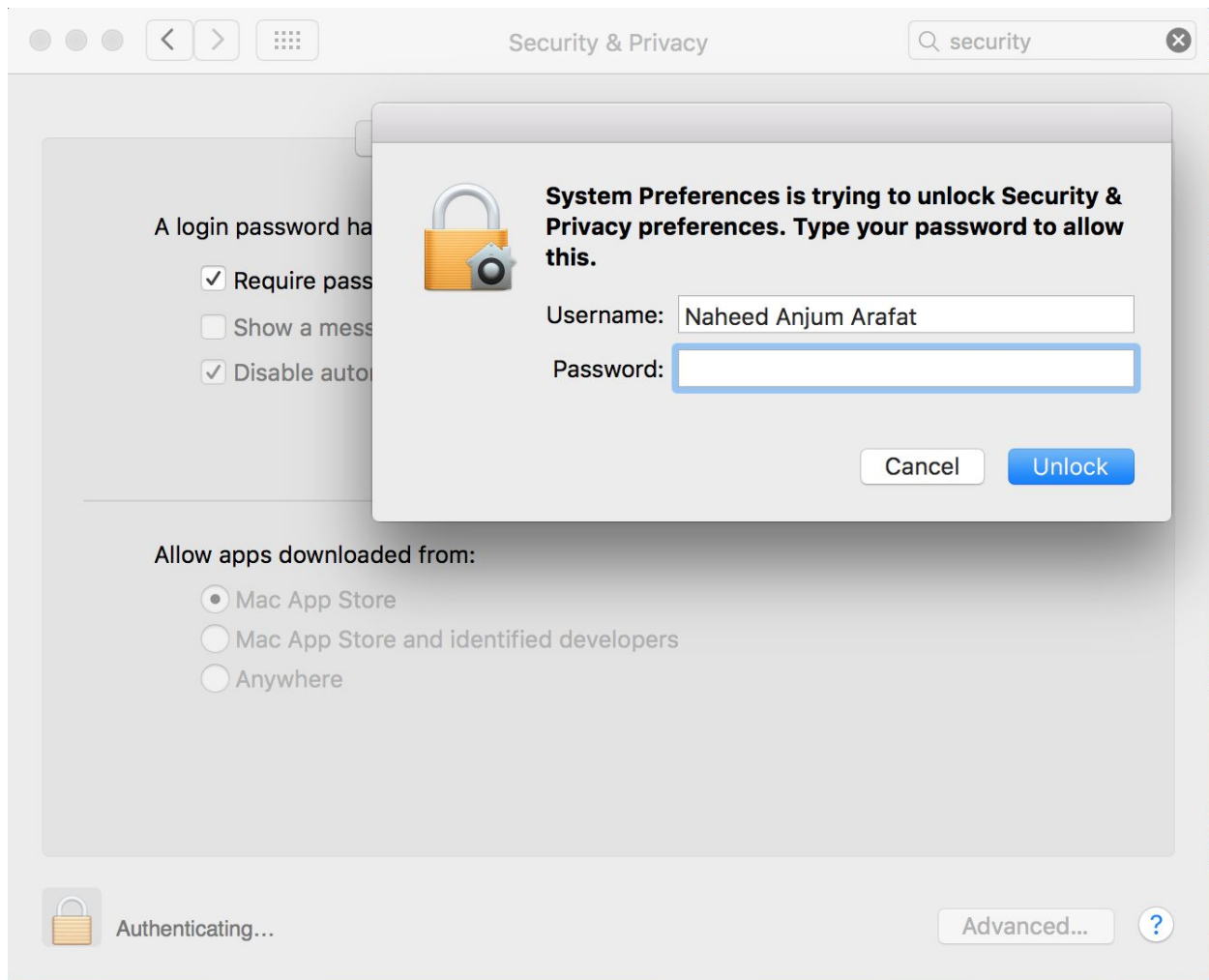
Change Security Settings:
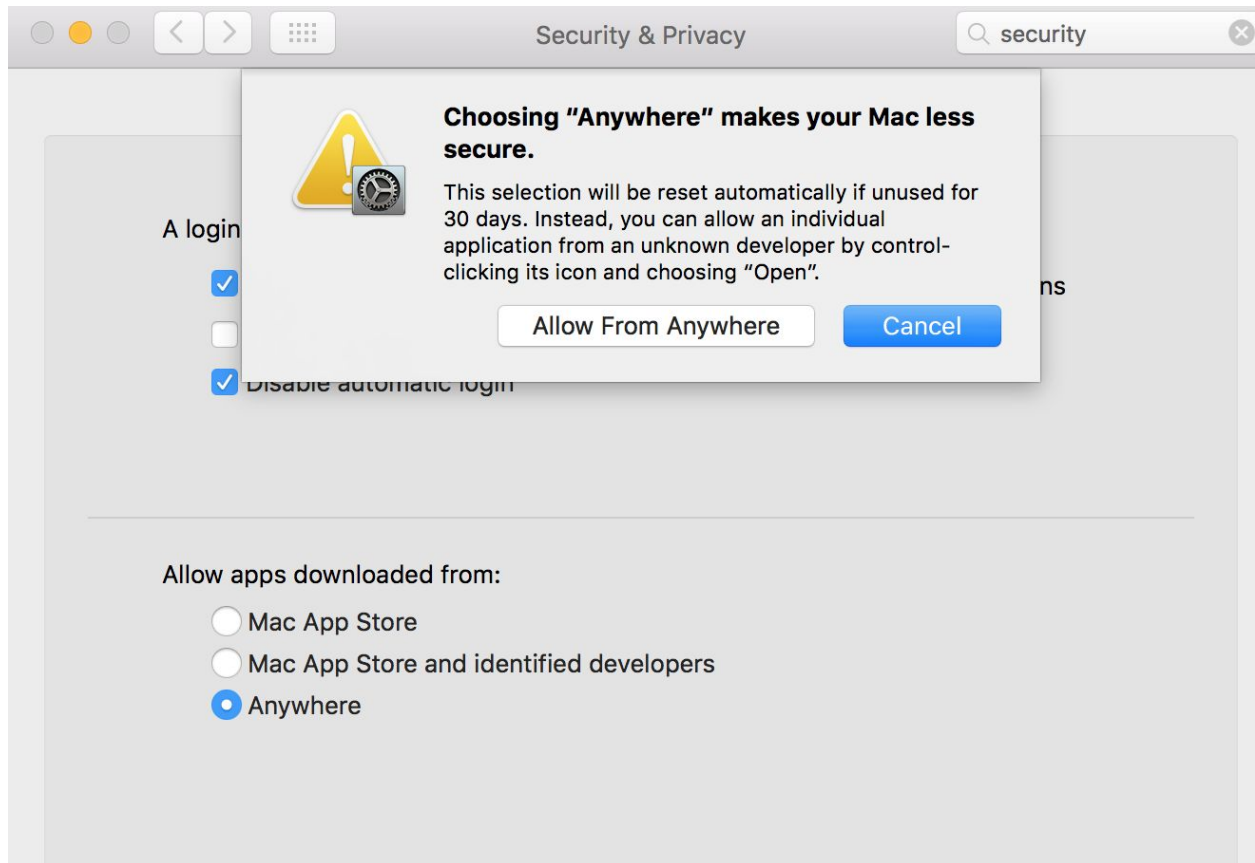
1. Go to System preferences:-

2. Click on "Security & Privacy"

3. Click on the Lock icon around the bottom left . Put in your system password when asked :-
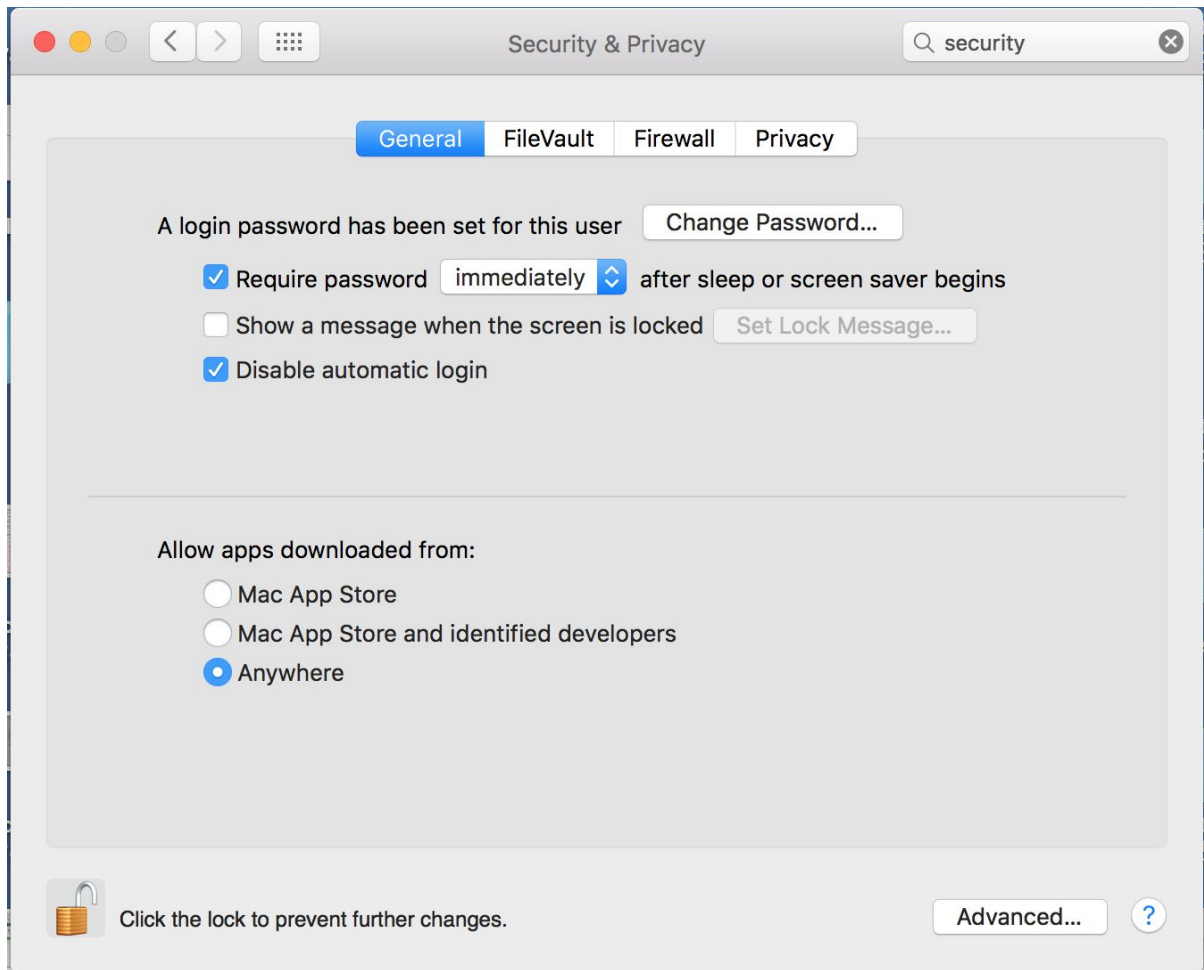
4. Click on "Anywhere" under "Allow apps downloaded from" and Click "Allow from Anywhere" option in the pop up dialog.

5. Click on the Lock icon to lock and save your changed settings. Then, Exit the window.

common mac errors:

1. Can not open Data Integration from Applications/Launcher.

    Solution:- Try running from your mac terminal. Copy-paste the following command:

    sh /Applications/data-integration/spoon.sh

    And Enter.

2. I am using mac os Sierra or newer OS. Can not find "Allow app download from" Anywhere option.

    Solution: Follow instructions in the link below:

    http://osxdaily.com/2016/09/27/allow-apps-from-anywhere-macos-gatekeeper/

3. You may encounter other errors related to java if you have java 9 installed in your system.

Solution:

First, uninstall java 9 following this :
https://gist.github.com/schnell18/bcb9833f725be22f6acd01f94b486392

Then, install java 8 (JDK) from here:-
https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html

Pentaho Installation on Ubuntu

 Follow the instructions in the following link-

http://fcorti.com/2016/12/05/installing-pentaho-data-integration-7/

*** **For troubleshooting**:- Send an email to *naheed_anjum@u.nus.edu* and attach a screenshot of error message or anything that might help me troubleshoot the issue. Thanks.

Tableau

In the project, you are going to use tableau desktop for doing visual analytics on the star-schema. We are going to use the trial version of tableau, which has a validity of **14 days**.

** Important Note for ubuntu users:- Tableau does not support ubuntu operating system. Only Mac and Windows is supported.

Tableau Desktop Download link

https://www.tableau.com/support/releases/desktop/10.5.9

Instructions:

Once you download the installer, Installing is straight forward. Double click the installer and follow the instructions. The first time you open tableau, you click the option trial and enter your name, email, institution etc, if you are asked by the application.

If you need to use tableau after the trial:

If you still need to use tableau after the trial period, you can download a version (say, 2018.1) newer than the version you installed before (say, 10.5) from here https://www.tableau.com/support/releases .

Uninstall the version you installed before (10.5) and install this version (2018.1) , and you can use it for another 14 days. Since *Tableau is backward compatible, the viz you created in older version will work in a newer version.*

*** **For troubleshooting**:- Send an email to *naheed_anjum@u.nus.edu* and attach a screenshot of error message or anything that might help me troubleshoot the issue.