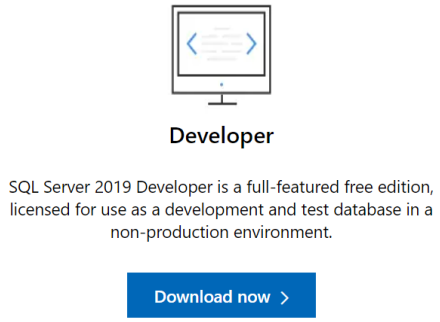
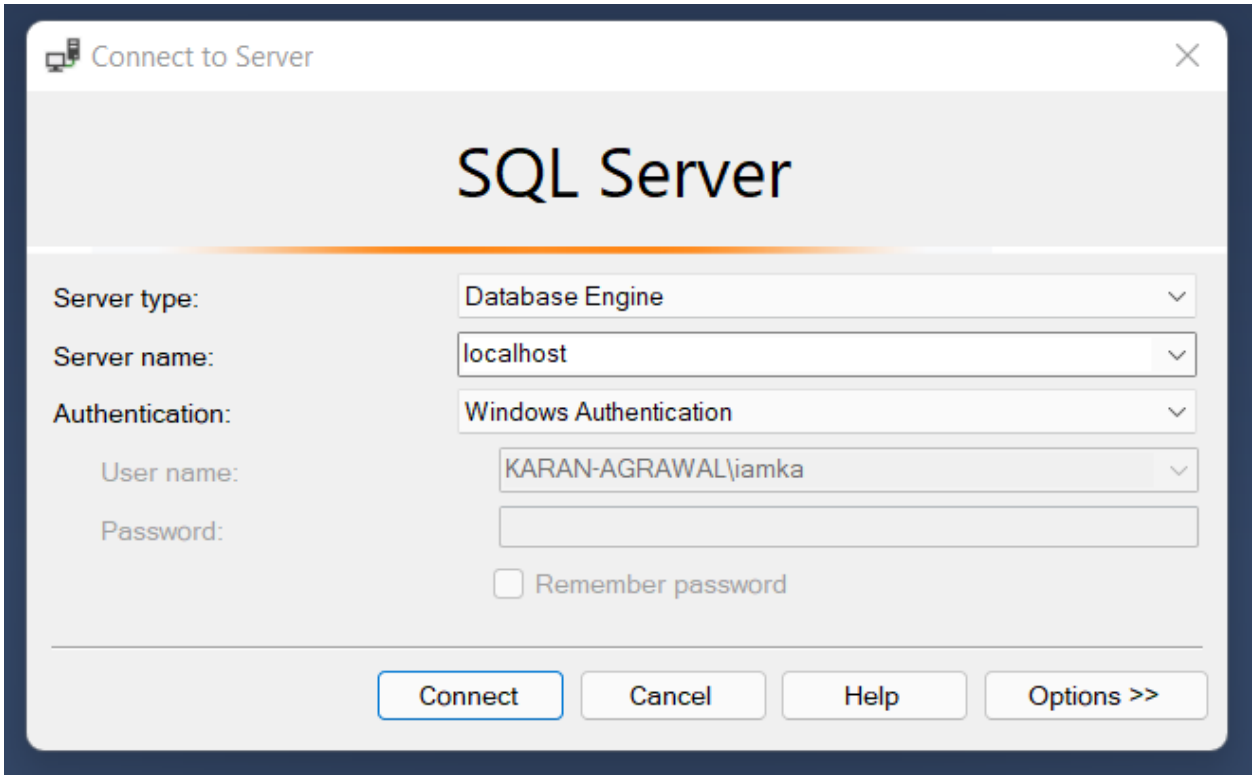
Big Data Class Project - Application Manual

Setup of SSMS and Neo4j

1. Download the zip folder of the project and extract the folder
2. Download and Install Neo4j Desktop - Download Neo4j Desktop
3. Download VSCode - Download Visual Studio Code - Mac, Linux, Windows
4. Download and Install SQL Server Development Edition - SQL Server Downloads



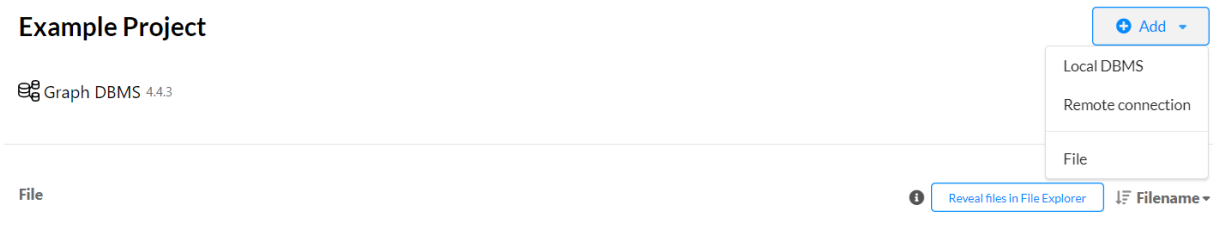
5. Download and Install SQL Server Management Studio (SSMS) - Free Download for SQL Server Management Studio (SSMS) 18.11.1
6. Connect SSMS to the SQL Server -



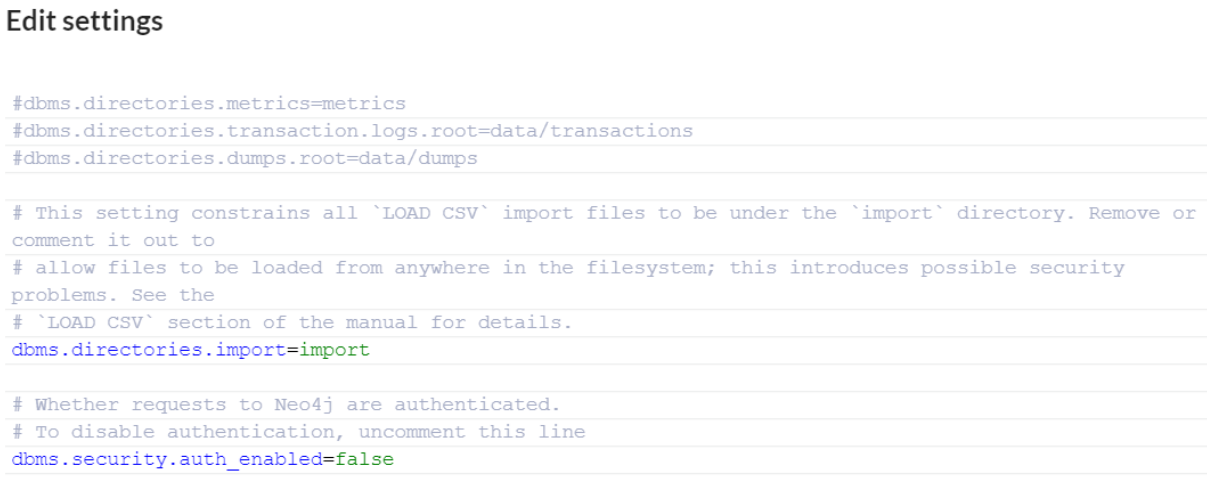
7. Run the SQL script to create the tables
    a. Open the script (SSMS_Script.sql) in VSCode for better readability
    b. Copy the code into SSMS

c. Select and Execute the first line of code :- CREATE DATABASE [Class_MetaData]
d. Select rest of the query and execute it, these query will create all the required tables

8. Now that the destination is all set to accept the data, lets set-up the source. Open Neo4j and create two new projects.
   a. First create a project - Group2_Project



   b. Using the Local DBMS create a new project - Set username and password as Group2_Project and g2 respectively
   c. Go to C drive > Users > Select your User > Select .Neo4jDesktop > Select relate-data > select dbmss > select the recently created database(**remember to select the correct database, as this database does not have its name on file**) > select import > Paste the G2_Data excel sheet in this folder.
   d. Follow the above three steps with the Group 9 project as well - Set username and password as Group9_Project and g9 respectively
   e. Open the settings of Group2_Project, by selecting three dots on the extreme right hand side, and set dbms.security.auth_enabled to false



   f. Apply the settings and follow the same for Group9_Project
   g. Open settings again for Group 2 and paste the command at the end of the settings - dbms.security.procedures.unrestricted=apoc.*

Big Data Class Project - Application Manual

```
#*****************************************************************
# Other Neo4j system properties
#*****************************************************************

dbms.security.procedures.unrestricted=apoc.*
```

   h. Follow the same steps with Group 9 project
9. Copy the file apoc-4.4.0.3-all and navigate to -
   C drive > Users > Select your User > Select .Neo4jDesktop > Select relate-data > select dbmss > select the group 2 project > plugins and paste the apoc-4.4.0.3-all folder here
   Follow the same instructions for Group 9.

Loading Data into Neo4j

1. Start the Group2_Project by pressing start, select continue anyway for the security alert
2. Open Group2_DataIngestion Queries, copy the code
3. Once, the database is active, open the database, past the query infront of the $ sign and execute it. Wait for all data to be loaded into Neo4j.
4. After successful completion, follow the same steps for Group 9, Select the project of group 9, start the project, copy query from Group9_DataIngestion and execute

Now that data is loaded into the source, lets run ETL process to dump data into SSMS

1. First lets install all the required libraries needed to run the ETL, open cmd and run the following commands
   a. pip install neo4j
   b. pip install pyodbc
   c. pip install pandas
   d. Pip install openpyxl
2. Open MetaDataETL.ipynb in VSCode
3. Change the paths according to your system for the following files
   a. Give path of DataBases.xlsx to the databasePath
   b. Give path of AttributeDataTypes.xlsx to the attributeDatatyoePath
   c. Assign the business term path of BusinessTermListG2 in the cell where Group 2 data is being loaded, and assign the business term path of BusinessTermListG9 in the cell where Group 9 data is being loaded.
4. Run the ETL
   a. Remember to run the cell whose database is still running in Neo4j, if you followed the manual, group 9 database must be on. So Execute the cell where the parameters of group 9 are passed to the ETL
   b. Once, the loading of data is successfull, now start the database of group 2 in the Neo4j and execute the cell where parameters of group 2 are passed into tha database.
   c. The data has been successfully loaded into SSMS.
5. Open Inser-Bridge_Table.ipynb in VSCode and execute. This code will inser all the values in the Bridge Table.

Big Data Class Project - Application Manual

The ETL is been successfully executed and now lets make a User Interface using streamlit.

1. Open User-Interface-SSMS.py in VSCode
    a. Download dependencies using
        i. Pip install enum
        ii. Pip install streamlit
2. Execute the file.
3. Open CMD
4. Change the directory to the folder where all the project files are kept
5. Run the command - streamlit run User-Interface-SSMS.py
6. If using for first time it will ask the email and other fields leave them blank
7. The UI will open in a web page.
8. You can see the vital information of all the databases.
    a. Select Database to view the nodes and relationships
    b. Select a node to view its attribute and the business term description
    c. Select attribute to view its information - type, range etc.