**CSCI316 Software and Programming Packages Installation Guide
(for Linux/MacOS)**

(1) Java 8 (JDK 1.8)
(2) Python 3
(3) Apache Spark 2.4 (https://spark.apache.org/downloads.html, select the "pre-built for Apache Hadoop 2.7 and later" installation)
(note. Spark requires Scala which requires Java 8.)
The following environment variable need to be set up for Spark:
export JAVA_HOME=<path to java home>
export SPARK_HOME=<path to spark home >
export PYSPARK_PYTHON=<path to python3>
export PYSPARK_DRIVER_PYTHON=<path to ipython3>
(4) Jupyter Notebook (https://jupyter.org/install)
(5) Python libraries (all of which can be installed with pip/pip3):
    I.   IPython3
    II.   PySpark 2.4 (e.g., pip install pyspark==2.4.4)
    III.   TensorFlow 2
    IV.   Scikit-Learn
    V.   Scientific computing libraries: SciPy, NumPy, Pandas, Matplotlib, re, importlib, psutil
    VI.   Some dependence libraries: Py4J, pyarrow, psutil, cairocffi