

# ANALYTICS SYSTEMS ENGINEERING (MSDS 436)

## EXERCISE 5

---

[shreenidhi.bharadwaj@northwestern.edu](mailto:shreenidhi.bharadwaj@northwestern.edu) | [ChristopherFiore2015@u.northwestern.edu](mailto:ChristopherFiore2015@u.northwestern.edu)

### CONTENTS

1. [Working with AWS](#)
  - a. [Amazon EC2 \(Week2\)](#)
  - b. [Object Storage \(Week2\)](#)
  - c. [PostgreSQL \(Week2\)](#)
  - d. [Tableau & DBeaver \(Week2\)](#)
  - e. [RedShift \(Week2\)](#)
  - f. [Elasticsearch \(Week3\)](#)
  - g. [Neo4j\(Week4\)](#)
  - h. [Apache Spark \(Week5\)](#)
  - i. [Apache Spark \(Week5 & Week6\)](#)

### INSTALLING SPARK ON LAPTOP

1. Installing Pyspark and dependencies on your laptop. Instructions detailed below

Windows: <https://medium.com/@GalarnykMichael/install-spark-on-windows-pyspark-4498a5d8d66c>

MAC: <https://medium.com/@GalarnykMichael/install-spark-on-mac-pyspark-453f395f240b>

Note: In case you do get java related errors while running Pyspark, run the following  
conda install -c cyclus java-jdk

2. Run the scrips in order. Data references is in the data folder.

