MSDS 694
Distributed Computing

DIANE WOODBRIDGE, PH.D

Homework

Install Apache Spark on your machine and run a code on jupyter notebook.

Spark Installation

Spark runs on Java 8+, Python 2.7+/3.4+ and R 3.1+. For the Scala API, Spark 2.2.0 uses Scala 2.11. You will need to use a compatible Scala version (2.11.x).

Install JDK

```
brew update
brew tap caskroom/versions
brew cask search java
brew cask install java8
```
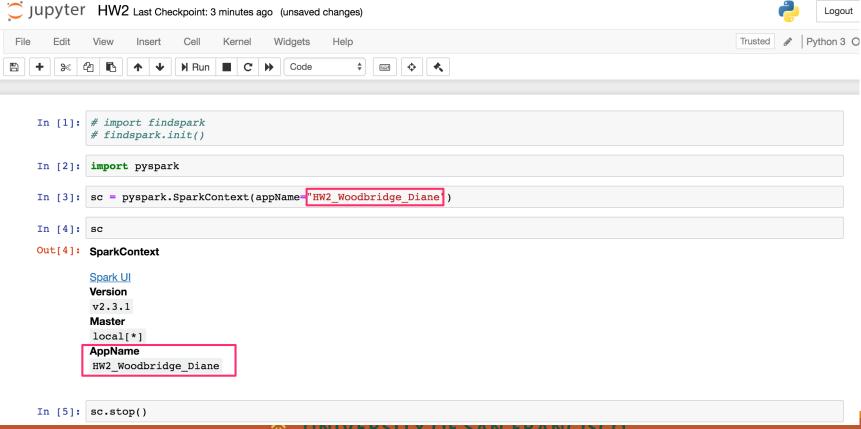
Spark Installation

Install jupyter, scala, pyspark and findspark

```
conda install jupyter

conda install -c anaconda-cluster scala

conda install -c conda-forge pyspark

conda install -c conda-forge findspark
```

HW2

1. Change the appName to "HW2_Lastname_Firstname" and execute the code.

HW2

2. Take a screenshot and submit it as "HW2_Lastname_Firstname.pdf".