# Running Spark

# Running Spark on Bash Console

- curl -O https://d3kbcqa49mib13.cloudfront.net/spark-2.2.0-bin-hadoop2.7.tgz
- tar zxvf spark-2.2.0-bin-hadoop2.7.tgz
- cd ~
- sudo nano .bashrc
  - alias python=python3.7
  - export SPARK_HOME="/home/<your_username>/spark-2.2.0-bin-hadoop2.7/“
  - export PATH=$SPARK_HOME/bin:$PATH
- source ~/.bashrc
- pyspark

# Configure PySpark on Jupyter Notebook

- import findspark
- findspark.init()
- import pyspark
- from pyspark.sql import SparkSession

- spark = SparkSession.builder.getOrCreate()

- df = spark.sql("select 'spark' as hello ")

- df

# Thanks