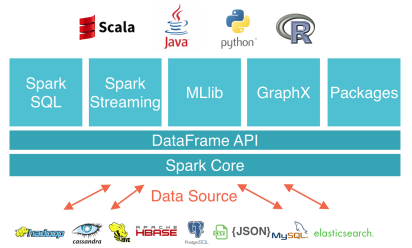
Welcome / Lab 1 - Hello Spark - Instructor_6_0.ipynb

Hello Spark!!!

Apache Spark is a fast and general-purpose cluster computing system. It provides high-level APIs in Java, Scala, Python and R, and an optimized engine that supports general execution graphs. It also supports a rich set of higher-level tools including Spark SQL for SQL and structured data processing, MLlib for machine learning, GraphX for graph processing, and Spark Streaming.



This notebook will show you some basic concepts to start working with Apache Spark including:

- Understanding Spark Context
- Creating Resilient Distributed Datasets (RDD)
- Performing Data Transformations
- Loading Data Files to use with Spark

Tool Tips:

- Notice the navigation and command buttons at top of the notebook. Press Play & Stop buttons to execute code and interupt execution.
- Notice each cell has type. (Markdown, Code, Etc) This cell

Recent Notebooks

- Lab 1 - Hello Spark - Instructor_6_0.ipynb
- NFL Weather Analysis.ipynb
- GB old NFL Weather Analysis.ipynb
- GB - Lab2 - Intro to SparkSQL.ipynb
- Intro to SparkSQL1.ipynb
- Lab 1 - Hello Spark - Instructor_6_.ipynb
- IntroToSparkMLlib0.ipynb
- Tutorial - Spark in Scala.ipynb
- Intro to SparkSQL8.ipynb
- Lab 3 - Machine Learning - Instructor2.ipynb

← See all notebooks

Recent Data

- db.lck
- derby.log
- log1.dat
- c180.dat
- c1a1.dat
- c191.dat
- c1b1.dat
- logmirror.ctrl
- dbex.lck
- log.ctrl

← See all data


jupyter
RS


jupyter
RS