



www.datascienceacademy.com.br

Big Data Real-Time Analytics com Python e Spark

Bibliografia, Referências e Links úteis



Documentação oficial:

http://spark.apache.org/docs/latest/index.html

Spark Transformations

http://spark.apache.org/docs/latest/programming-guide.html#transformations

Spark Actions

http://spark.apache.org/docs/latest/programming-guide.html#actions

Dados do Uber

https://github.com/fivethirtyeight/uber-tlc-foil-response

Baby Names Dataset

https://www.kaggle.com/kaggle/us-baby-names

Clash of the Titans: MapReduce x Spark for Large Scale Data Analytics

http://www.vldb.org/pvldb/vol8/p2110-shi.pdf

Google Data Center

https://www.youtube.com/watch?v=XZmGGAbHqa0

Facebook Data Center

https://www.youtube.com/watch?v=Y8Rgje94iI0

Exemplos do Spark

http://spark.apache.org/examples.html

Código Fonte do Spark

https://github.com/apache/spark

Design e Implementação do Spark

https://github.com/JerryLead/SparkInternals