www.datascienceacademy.com.br

Engenharia de Dados com Hadoop e Spark

Bibliografia, Referências e Links úteis

Real-Time Streaming Data Pipelines with Apache APIs: Kafka, Spark Streaming, and HBase
https://www.mapr.com/blog/real-time-streaming-data-pipelines-apache-apis-kafka-spark-streaming-and-hbase

Hot to Analyze Machine and Sensor Data
https://hortonworks.com/hadoop-tutorial/how-to-analyze-machine-and-sensor-data/

Parquet
http://parquet.apache.org/

OCR File
https://cwiki.apache.org/confluence/display/Hive/LanguageManual+ORC

Scaling the Facebook data warehouse to 300 PB
https://code.facebook.com/posts/229861827208629/scaling-the-facebook-data-warehouse-to-300-pb/

Backup Hadoop – DistCp
https://hadoop.apache.org/docs/r3.2.0/hadoop-distcp/DistCp.html

Hadoop Performance Monitoring Tools: A Comparison Guide
http://www.driven.io/hadoop-performance-monitoring-tools/

Métricas de Monitoramento do Hadoop
https://hadoop.apache.org/docs/r3.2.0/hadoop-project-dist/hadoop-common/Metrics.html