

Spark Essentials

Recap



Takeaways

Spark essential skills

- creating a SparkSession
- the DataFrame API
- filtering, aggregating, grouping, joining
- Datasets
- RDDs

Spark essential concepts

- driver & executors
- lazy evaluation
- transformations & actions
- RDD partitions
- (a plus) stages & tasks
- (a plus) Spark execution planning

Spark rocks

