

Lesson 2.1: Introduction to Spark Core Concepts

DISTRIBUTED COMPUTING WITH SPARK SQL

Introduction to Spark Core Concepts



Brooke Wenig/Conor Murphy
Machine Learning Practice Lead/
Data Scientist, Databricks


UC DAVIS
Continuing and Professional Education

Slide 2: Welcome Back!

Welcome Back!



Slide 3: Looking at Spark Under the Hood



Looking at Spark Under the Hood

Caching data for increased performance

Partitioning -- Modify configurations for decreased query time

Reading the Spark UI in order to analyze performance and identify bottlenecks

Slide 4: Learning Objective



Learning Objective

Test your skills to see how performant you can make your Spark SQL queries