Lesson 2.3: Caching

DISTRIBUTED COMPUTING WITH SPARK SQL
Caching

Brooke Wenig
Machine Learning Practice Lead
Databricks

UCDAVIS
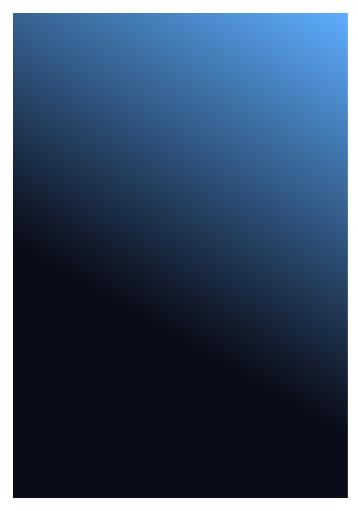Continuing and Professional Education

Slide 2:Welcome Back!

Welcome Back!

Caching partitions

Demonstrate performance impact of caching data

Demonstrate DataFrame optimizations and performance advantages of Tungsten

Slide 3: Learning Objectives

Learning Objectives

Cache your data for increased performance

Read the storage tab of the Spark UI