Lesson 3.4: File Formats

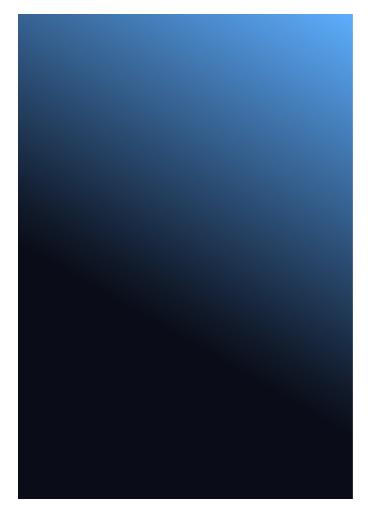
DISTRIBUTED COMPUTING WITH SPARK SQL

File Formats





Slide 2: Welcome Back!



Welcome Back!

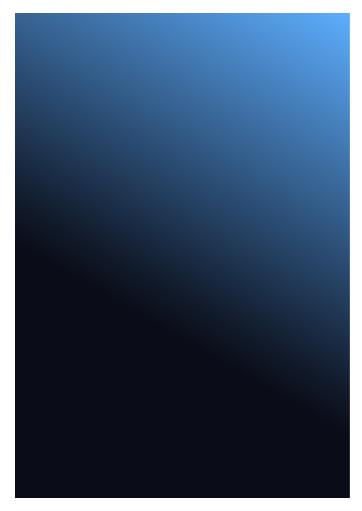
File formats and compression options

CSV files: row based data files

Parquet: column based

This has performance advantages

Slide 3: Learning Objectives



Learning Objectives

Determine the best file formats and compression options to use under a variety of conditions