

**RDDs,
DataFrames,
Datasets**



Objective

How to make RDDs, DataFrames and Datasets interoperate

Understand performance implications for using each



Lessons

Lesson 1: once you decide on an API level, stay there.

Lesson 2: use DataFrames... most of the time.

Lesson 3: the Datasets' kryptonite is their inability to optimize lambdas.

Spark rocks

