

Reusing Objects



Objective

Avoid recreating JVM objects

Techniques for when performance is critical



To Remember

Reuse JVM objects by making them mutable (oh my!)

Useful for scenarios when

- your aggregations/data structures are not supported by Spark out of the box
- multiple passes over the data are not acceptable
- performance is critical, i.e. even a few seconds count

Beware of

- using tuples – they are (immutable) case classes
- converting between Scala collections – especially implicit conversions

Spark rocks

