An Insufficient Introduction to Spark

Part 6: Concluding remarks

Riccardo Murri `<riccardo.murri@gmail.com>`

Concluding Remarks

There is more to Spark than this. . .

The time and scope of this course is quite limited.

Here is an (incomplete) list of Spark features that you might want to look up as you become more experienced in the system:

- ▶ The Streaming API
- ▶ Spark MLlib
- ▶ Data Sources!
- ▶ Support for other languages: Scala, Java, R

Spark packages

There are number of add-on packages and libraries for Spark at: https://spark-packages.org/

Where to go now?

Roškar R.: "'Big' Data analysis using Apache Spark",
https://github.com/rokroskar/spark_workshop

Material from this course

All the material from this course is online:
http://github.com/riccardomurri/pyspark-intro/

ElastiCluster was used to build the Spark + Jupyter cluster
for the course: http://gc3-uzh-ch.github.io/elasticluster

Thanks!

Now go write some code ;-)

Appendix