

Lesson 1: Introduction to the Spark Environment

1.3 Origins of the Framework

What Is Spark?

Apache Spark™ is a fast and general engine for large-scale data processing.



What Is Spark?

- Framework for distributed processing
- In-memory, fault tolerant data structures
- Flexible APIs in Scala, Java, Python, SQL... and now R!
- Open Source



The Recent Past

- Found by AMPLab, UC Berkeley (circa 2009)
 - Created by Matei Zaharia ([PhD Thesis](#))
 - Maintained by Apache Software Foundation
 - Commercial Support by Databricks



The Recent Past

- Found by AMPLab, UC Berkeley (circa 2009)
 - Created by Matei Zaharia ([PhD Thesis](#))
 - Maintained by Apache Software Foundation
 - Commercial Support by Databricks

