

**BerkeleyX: CS105x Introduction to Apache Spark**

Bookmarks

**▼ Week 1 - Apache Spark  
Programming Model****Lecture 1: Apache Spark  
Architecture and  
Programming Model**

Quizzes

**Setting up the Course  
Software Environment (Due  
September 10, 2016 at 23:59  
UTC)**

Setup

**(Optional) Survey about your  
machine and setup  
experience**

- ▶ Week 2 - The Structured Query Language and Spark SQL
- ▶ Week 3 - Analyzing Semi-Structured Data with Apache Spark






Week 1 - Apache Spark Programming Model &gt; Lecture 1: Apache Spark Architecture and Programming Model &gt; Analysis, Big Data, and Apache Spark

Bookmark

## Analysis, Big Data, and Apache Spark

BERCS1052016-V002600



 0:00 / 8:10 1.0x    [Download video](#)[Download transcript](#)[.srt](#)

Note that it is possible to run iPython with pySpark, in which case you do not need to create the SparkContext.

You can read a blog post about my research group's ADAM open source genomics pipeline see and run your own ADAM genomics analysis on data from the 1000 Genomes project using this notebook.

Capital One also uses Spark and graph analysis to detect fraudulent applications.

Another example of Big Data processing is IBM's jStart team's work with the USA Cycling Women's Team Pursuit team. The jStart team is using Spark Streaming to analyze streaming data from the riders' power meter, heart rate monitor and wearable muscle oxygen sensor, and provide coaches and riders with real-time performance metrics, such as W-prime depletion and matches burned.

---

## Spark DataFrame References

You can find more information about Spark DataFrames in the Spark online documentation for DataFrames. The DataFrame API is also available online.

## Components of Apache Spark

(1/1 point)

Which of the following are components of Apache Spark?

☐ R

☒ The core Apache Spark runtime ✓

☒ Spark SQL ✓

☐ Python

☒ MLlib ✓

☐ Java

☐ Scala

☒ ML ✓

☒ GraphX ✓

✓

Note: Make sure you select all of the correct options—there may be more than one!

#### EXPLANATION

Python, Scala, Java, and R are languages that are fully supported by Apache Spark, however they are not part of Apache Spark. The core Apache Spark runtime, Spark SQL, MLlib, ML, and GraphX are all components of Apache Spark.

*You have used 2 of 4 submissions*

CC BY-NC-SA Some Rights Reserved



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

POWERED BY  
OPENedX®

