





## BerkeleyX: CS110x Big Data Analysis with Apache Spark

  
Bookmarks

- ▶ Week 1 - Big Data and Data Science
- ▶ Week 2 - Performing Data Science
- ▼ Week 3 - Programming with Resilient Distributed Datasets

**Lecture 3: Apache Spark Resilient Distributed Datasets**Quizzes **Lab3a - RDD Tutorial**Lab due Sep 13, 2016 at 04:30 IST **Lab 3b - Text Analysis and Entity Resolution**Lab due Sep 13, 2016 at 04:30 IST **Lab 3b Quiz Questions**Quizzes 

Week 3 - Programming with Resilient Distributed Datasets &gt; Lecture 3: Apache Spark Resilient Distributed Datasets &gt; Spark RDD Programming Model

 Bookmark

## Spark RDD Programming Model

BERCS1102016-V001400

Start of transcript. Skip to the end. 

ANTHONY JOSEPH: Now let's look at an example of Spark's programming model with RDDs.

So we start by reading a file, a text file, into an RDD.

We use `sc.textFile`, passing in the name of the file, and the number of partitions-- in this case, four-- and we

0.00 / 5.00

1.0x



[Download video](#)[Download transcript](#)[.srt](#)

## Spark RDD Program Lifecycle

(1/1 point)

Which of the following are part of a Spark RDD program's lifecycle?

☒ RDDs that are reused may be cached ✓

☐ Transformations cause parallel computation to be immediately executed

☒ Actions cause parallel computation to be immediately executed ✓

☒ Transformations lazily create new RDDs ✓

☐ Actions create recipes for performing parallel computation on datasets



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

### EXPLANATION

Transformations specify how to perform parallel computation in a lazily evaluated manner. Actions cause the transformations to be executed. If you plan to reuse an RDD, you should cache it.



© ⓘ Ⓢ Ⓞ 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®

