Bookmark

Spark RDD Programming Model

BERCS1102016-V001400

0:00 / 3:08 1.0x

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

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?

- [x] RDDs that are reused may be cached ✔
- [ ] Transformations cause parallel computation to be immediately executed
- [x] Actions cause parallel computation to be immediately executed ✔
- [x] Transformations lazily create new RDDs ✔
- [ ] Actions create recipes for peforming 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.