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

Bookmark

Spark Shared Variables

BERCS1102016-V001500

0:00 / 8:38 1.0x

Download video

Broadcast Variables

(1/1 point)

In iterative or repeated computations, broadcast variables avoid the problem of repeatedly sending the same data to workers:

- False
- True ✔

EXPLANATION

Broadcast variables are an efficient way of sending data once that would otherwise be sent multiple times automatically in closures.

Accumulators

(1/1 point)

Accumulators can be used by Spark workers to efficiently read values during distributed computations:

- (•) False ✔
- ( ) True

EXPLANATION

Accumulators can only be written by workers and read by the driver program.

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

Download on the App Store

ANDROID APP ON Google play