

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

Bookmark

## Apache Spark Transformations

BERCS1052016-V001300



Couldn't get auth token: undefined

▶ 0:00 / 12:49

▶ 1.0x



Download video

Download transcript

.srt

## User Defined Functions

Note that UDFs in Python are slow, so whenever possible, consider using built-in functions instead. For example, instead of creating a lambda function and using a UDF to subtract one from the values of a column, you should use a select transformation to perform the subtraction.

## Python Documentation

A very useful reference when writing pySpark applications is the Python Documentation site. You can even download the reference documentation for later reference.

## Lazy Evaluation

(1/1 point)

Why is the lazy evaluation of transformations important?

- ☐ Spark can execute the transformations when it has several ready to run
- ☒ The Catalyst Optimizer can optimize which transformations are run and how they are run ✓
- ☐ It is simpler to delay the execution of transformations than run them immediately

**EXPLANATION**

By delaying the evaluation of transformations, Spark can use the Catalyst Optimizer to optimize the execution of transformations, including potentially skipping unnecessary transformations.

*You have used 1 of 2 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®

