edX   **BerkeleyX: CS105x Introduction to Apache Spark**

Bookmarks

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

🔖 **Bookmark**

▼ **Week 1 - Apache Spark Programming Model**

Apache Spark DataFrames

Lecture 1: Apache Spark Architecture and Programming Model
Quizzes ✎

BERCS1052016-V000100

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

▶   0:00 / 6:11| ▶   1.0x   🔊   ✖   CC   ❝

Download video   Download transcript   .srt

Creating Spark DataFrames

 (1 point possible)
You can create Spark DataFrames from which of the following?

| ☑ | Existing DataFrames |

| ☑ | Data sources, such as files |

| ☑ | pandas DataFrames |

| ☑ | R DataFrames |

| ☑ | Python lists |

✔

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

You have used 2 of 4 submissions

DataFrames

Couldn't get auth token: undefined

Which of the following is not a property of DataFrames?

- ◉ They can be changed after they are constructed ✔

- ◯ They can be created by transformations applied to existing DataFrames

- ◯ They enable parallel operations on collections of distributed data

- ◯ They track lineage information to enable efficient recomputation of lost data

EXPLANATION

DataFrames cannot be changed once they are created - they are immutable. You can create DataFrames by applying transformations to existing DataFrames and Spark automatically tracks how you create and manipulate DataFrames (their lineage) so that it can reconstruct any data that is lost due to slow or failed machine. Operations on DataFrames are performed in parallel.

You have used 1 of 3 submissions

Couldn't get auth token: undefined

POWERED BY
OPENedX®