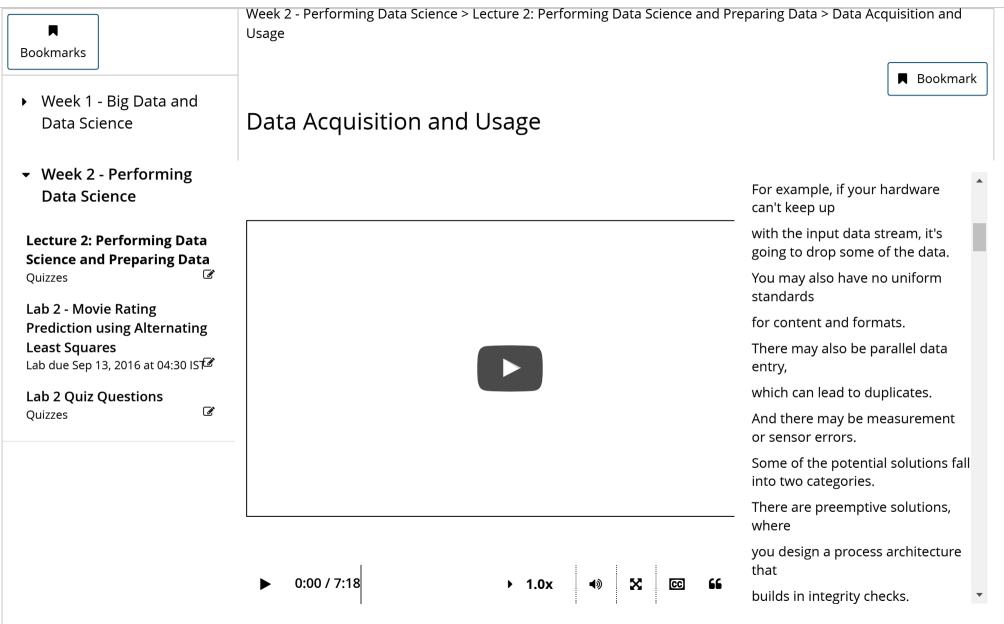


BerkeleyX: CS110x Big Data Analysis with Apache Spark



Download video Download transcript .srt

Data Gathering

(1/1 point)

Why is Data Gathering an important part of the data cleaning process?

- You can build integrity checks into the gathering process
- Data gathering is the process of observing or collecting data
- Data gathering is not a source of data quality issues
- You cannot perform data cleaning during the data gathering process

EXPLANATION

It is important to design a good data gathering architecture that includes integrity checks, as it is better to avoid importing bad data than to have to deal with it in the dataset.

Data Delivery

(1/1 point)

You will often find that your data is provided by a third-party provider. What is not a data quality issue that you have to worry about?

- Data corruption during the delivery process.
- A commitment from the data provider about the quality of the data.
- The amount of data being provided.
- The dependences betwen the data streams and processing steps.

EXPLANATION

Data corruption, data quality, and data interdependencies are all issues that you have to worry about when dealing with a third-party data provider.

© () () ⊙ 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.















