Bookmarks

- Week 1 - Big Data and Data Science
- **Week 2 - Performing Data Science**

Lecture 2: Performing Data Science and Preparing Data
Quizzes

Lab 2 - Movie Rating Prediction using Alternating Least Squares
Lab due Sep 13, 2016 at 04:30 IST

Lab 2 Quiz Questions
Quizzes

Bookmark

Data Acquisition and Usage

0:00 / 7:18 1.0x

For example, if your hardware can't keep up

with the input data stream, it's going to drop some of the data.

You may also have no uniform standards

for content and formats.

There may also be parallel data entry,

which can lead to duplicates.

And there may be measurement or sensor errors.

Some of the potential solutions fall into two categories.

There are preemptive solutions, where

you design a process architecture that

builds in integrity checks.

Download video Download transcript .srt

Data Gathering

(1/1 point)
Why is Data Gathering an important part of the data cleaning process?

- [x] 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.