edX   **BerkeleyX:** CS110x Big Data Analysis with Apache Spark

🔖 Bookmarks

🔖 Bookmark

▼ **Week 1 - Big Data and Data Science**

Data Science Topics

Lecture 1: Big Data and Data Science
Quizzes ✎

Setting up the Course Software Environment
Setup ✎

Lab 1: Power Plant Machine Learning Pipeline
Lab due Sep 13, 2016 at 04:30 IST ✎

Lab 1 Quiz Questions
Quizzes ✎

▶ ▶

✎

▶   0:00 / 11:01   ▶ 1.0x   🔊   ✕   CC   ❝

Download video   Download transcript   .srt

Overcoming Assumptions

 (1/1 point)

Why is overcoming assumptions hard? Select all the answers that match:

☐   Assumptions are always correct

☑   It is human nature to make assumptions   ✔

☑   Humans have biases based on prior observations   ✔

✔

Correct:
You are correct, we all naturally try to generalize when we see what appears to be a pattern.
You are correct, once we observe a pattern, we tend to see that pattern in future observations.

> **EXPLANATION**
>
> We all naturally try to generalize when we see what appears to be a pattern. Also, once we observe a pattern, we tend to see that pattern in future observations. As data scientists we have to try and overcome these tendencies.

✏

Impediments to Collaboration

 (1/1 point)

Based on what you've seen in the lecture, what are some impediments to collaboration? Select all the answers that match:

☑　One-off analyses　✔

☑　Diverse tools and programming/scripting languages　✔

☑　Poor script, program, and tool library management　✔

☑　Often easier to rewrite a script instead of searching for it　✔

✔

EXPLANATION

All of the answers are correct. Often, an analysis is exploratory and only performed once. The wide variety of tools and programming and scripting languages makes it difficult to share them, and there are few systems for managing all of the software. Also, the lack of good management systems means that it often is easier to reimplement an analysis rather than to search for it.

Extract-Transform-Load

 (1/1 point)

Why is Extract-Transform-Load critical? Select all the answers that match:

☑ ETL prepares data for processing   ✔

☐ It is easier to perform data analyses before performing ETL

☑ ETL provides the data we need for analyses   ✔

✔

EXPLANATION

ETL is the first step in Data Science. We have to acquire and convert data into a canonical format before we can analyze the data. Without ETL, there is no data to analyze.