Lesson 1: What is Big Data?

Adapted from code.org curriculum

Objectives: You will be able too...

- Identify sources of data produced, used, and consumed by a web application
- Given a tool that provides access to a large dataset, explain the kinds of problems such a tool could solve.
- Use a tool that provides access to "big data" and investigate its sources
- Explain that new techniques are necessary to store, manage, transmit, and process data at the scale it is currently being produced

Video

- Big Data is Better Data – Ted Talk – Video
- Based on what you saw in the video, what is big data?

Big Data Video

- Big data means different things, at different times, to different people
  - It can mean devices that are constantly collecting data
  - It can mean digitalizing data that's been around for a long time (e.g. , every book ever written)
  - It can mean machine learning and artificial intelligence

Activity: Exponential Growth and Moore's Law

The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020
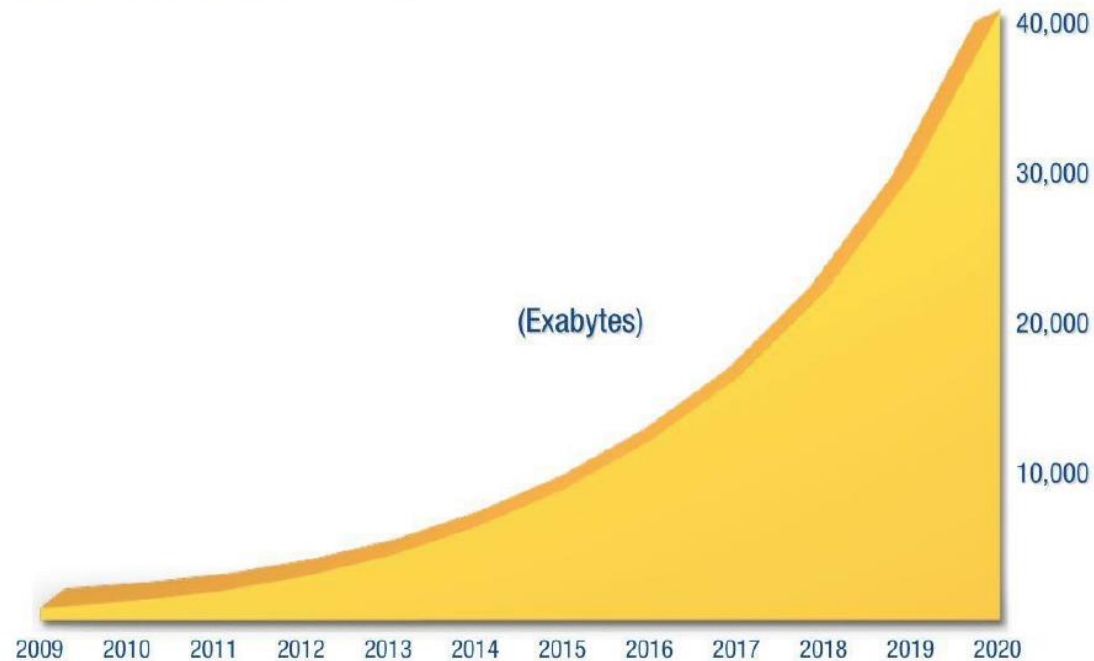
(Exabytes)

This IDC graph predicts exponential growth of data from around 3 zettabytes in 2013 to approximately 40 zettabytes by 2020. An exabyte equals 1,000,000,000,000,000,000 bytes and 1,000 exabytes equals one zettabyte. Source: IDC's Digital Universe Study, December 2012, http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf.

Activity: Exponential Growth and Moore's Law

- Part of what contributes to data being "big" is the sheer growth of the amount of data in the world.
- As you can see from the chart, the amount of data flying around is growing exponentially, doubling every two years or so.
  - The world will produce as much digital data over the next 2 years, as currently existed in all of humanity prior to that.

Moore's Law

- 7.2.1F Moore's law has encouraged industries that use computers to effectively plan future research and development based on anticipated increases in computing power.

What is Moore's Law?

- It is not a law of nature or mathematics but simply a surprisingly accurate prediction that was made a long time ago.
- In 1965, a computer chip designer named Gordon Moore predicted that the number of transistors one could fit on a chip would double every 18 months or so.

What is Moore's Law?

- Amazingly, that prediction has more or less held true to the present day!
- The result is that since about 1970, computers have gotten twice as fast, at half the cost, roughly every 1.5-2 years
- With some small differences, the same is true for data storage capacity
- This is extraordinarily fast growth – exponential growth.

What is Moore's Law?

- This is so fast that it's hard to fathom and even harder to plan for
- For example:
  - If the average hard drive today is 1 TB and you are planning for something 6 years away, you should expect that average hard drives will be 8-10 TB
- Key Takeaway: We need to keep Moore's Law in mind as we plan for the future

Big Data Example:

- Google Maps!

Big Data Sleuth Card – Where does data come from?

- We are going to use the "Activity Guide – Big Data Sleuth Card"
- We will work in pairs:
  - I will assign you one of 5 websites

Wrap-up: What is "Big Data" and where does it come from?

- So after your explorations what do you think "big data" actually means? What makes it "big" as opposed to not?

Wrap-up: What is "Big Data" and where does it come from?

- So after your explorations what do you think "big data" actually means? What makes it "big" as opposed to not?
  - This doesn't have a fixed definition – whether data is "big" often depends on the context of the data itself or how it's trying to be used and even experts might have difficulty pinning it down

Wrap-up: Share out

- What kinds of data are out there?
- What format does it come in?
- Where does it come from?
- Did anyone find a link to an actual data source?
- Did anyone find an API? What's an API?

Wrap-up: What's different about working with big data?

- Big data has a broad impact but requires a whole new set of tools in order to handle its size and scale.
- The potential impacts are great, but there are many challenges involved in using large amounts of data
- What fields and industries are affected by big data?
- What are the challenges with working with big data vs. other data we have seen so far?

Big Data Vocab

- Big data - a broad term for datasets so large or complex that traditional data processing applications are inadequate
- Moore's Law – a prediction made by Gordon Moore in 1965 that computing power will double every 1.5-2 years, it has remained more or less true ever since