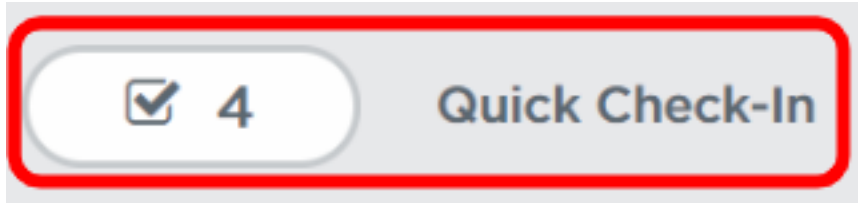


What is Big Data?

Unit 4 Lesson 1 (U4L1)

Activity Guide - Big Data Sleuth Card INB Version

Big Data is Better Data

College Board - Assessment Overview and Performance Task Directions for Students

Warm-Up

Write 2018 in binary. 1 1 1 1 1 1 0 0 0 1 0

2^{10}	2^9	2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
1024	512	256	128	64	32	16	8	4	2	1
1	1	1	1	1	1	0	0	0	1	0

Why did we stop at 2^{10}?

Start with 2018 and subtract the value of the largest power, and repeat….

2018 - 1024 - 512 - 256 - 128 - 64 - 32 - 2

= 994 = 482 = 226 = 98 = 34 = 2 = 0

Unit 4 Lesson 1 (U4L1)
What is Big Data?

Watch the following video and complete the Video Notes in your INB.

Big data is better data (15:51)

Prompt: Based on what you saw in the video, what is big data?

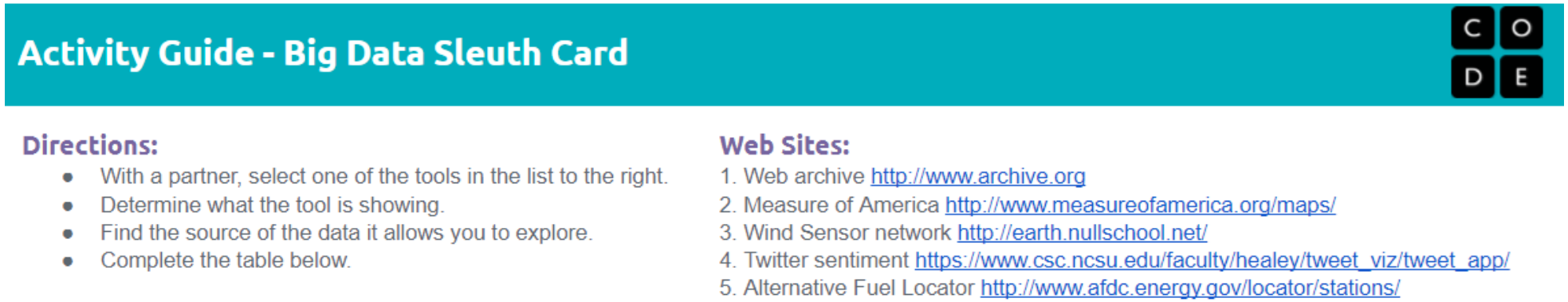
-Share with a neighbor
-Share as a class

Big data means different things, at different times, to different people.

- It can mean devices that are constantly collecting data.
- It can mean digitizing data that's been around for a long time (e.g., every book ever written).
- It can mean machine learning and artificial intelligence.

Big Data Sleuth Card (20 mins)

Big data surrounds us but it is sometimes surprisingly challenging to get access to it, use it, or see it. Much of the data out there is in the "wild." Even when the data is "available," it can sometimes be challenging to figure out where it came from, or how to use it.



Activity Guide - Big Data Sleuth Card

Directions:

- With a partner, select one of the tools in the list to the right.
- Determine what the tool is showing.
- Find the source of the data it allows you to explore.
- Complete the table below.

Web Sites:

1. Web archive http://www.archive.org
2. Measure of America http://www.measureofamerica.org/maps/
3. Wind Sensor network http://earth.nullschool.net/
4. Twitter sentiment https://www.csc.ncsu.edu/faculty/healey/tweet_viz/tweet_app/
5. Alternative Fuel Locator http://www.afdc.energy.gov/locator/stations/

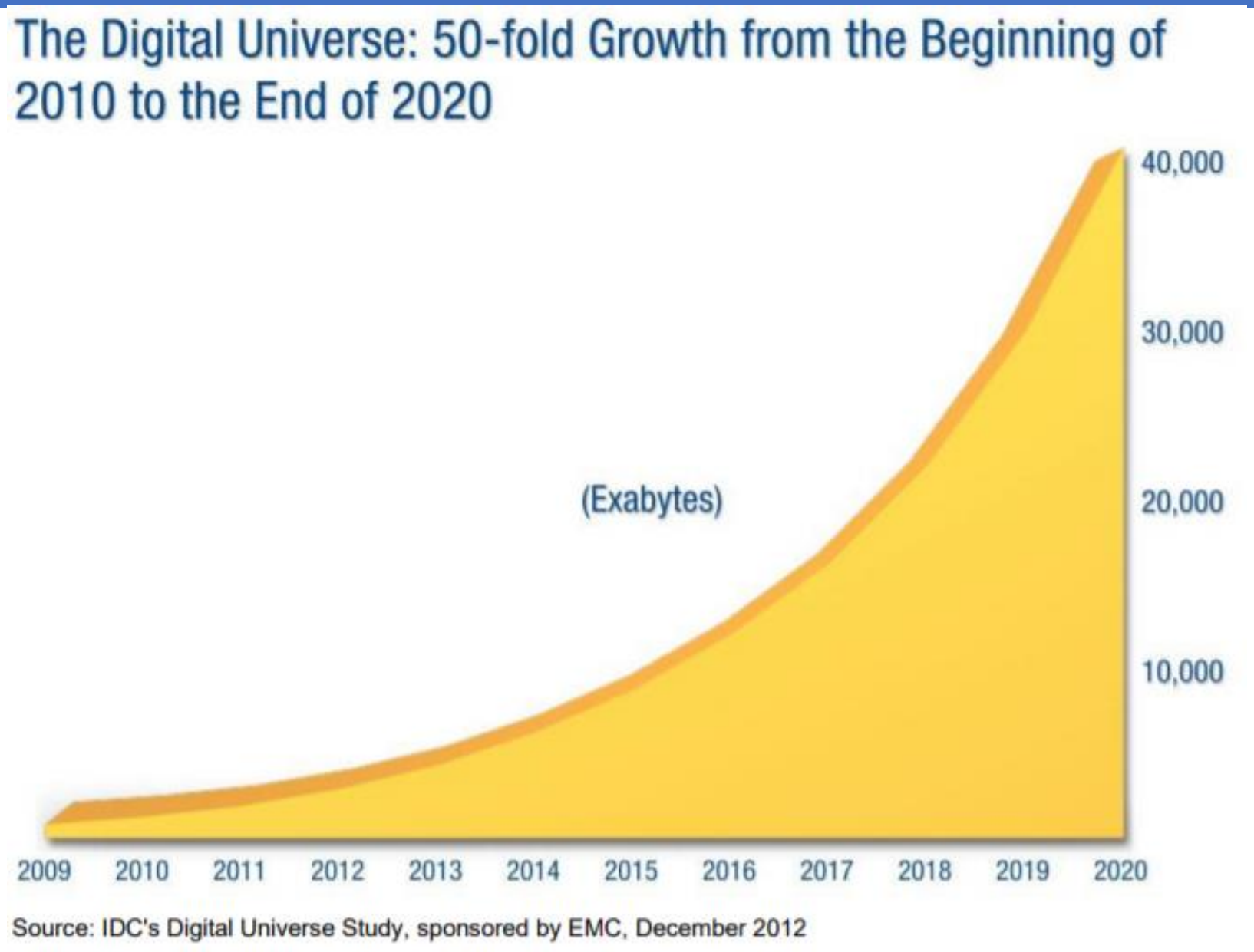
With your elbow partner, use the website assigned to you and your partner. Answer the questions on the paper/INB. Be prepared to share as a class.

Exponential Growth and Moore's Law

(10 mins)

Part of what contributes to data being "big" is the sheer growth of the amount of data in the world. Let’s have a look at a graph that shows us just how large big data is.

Unit 4 Lesson 1 (U4L1)
What is Big Data?



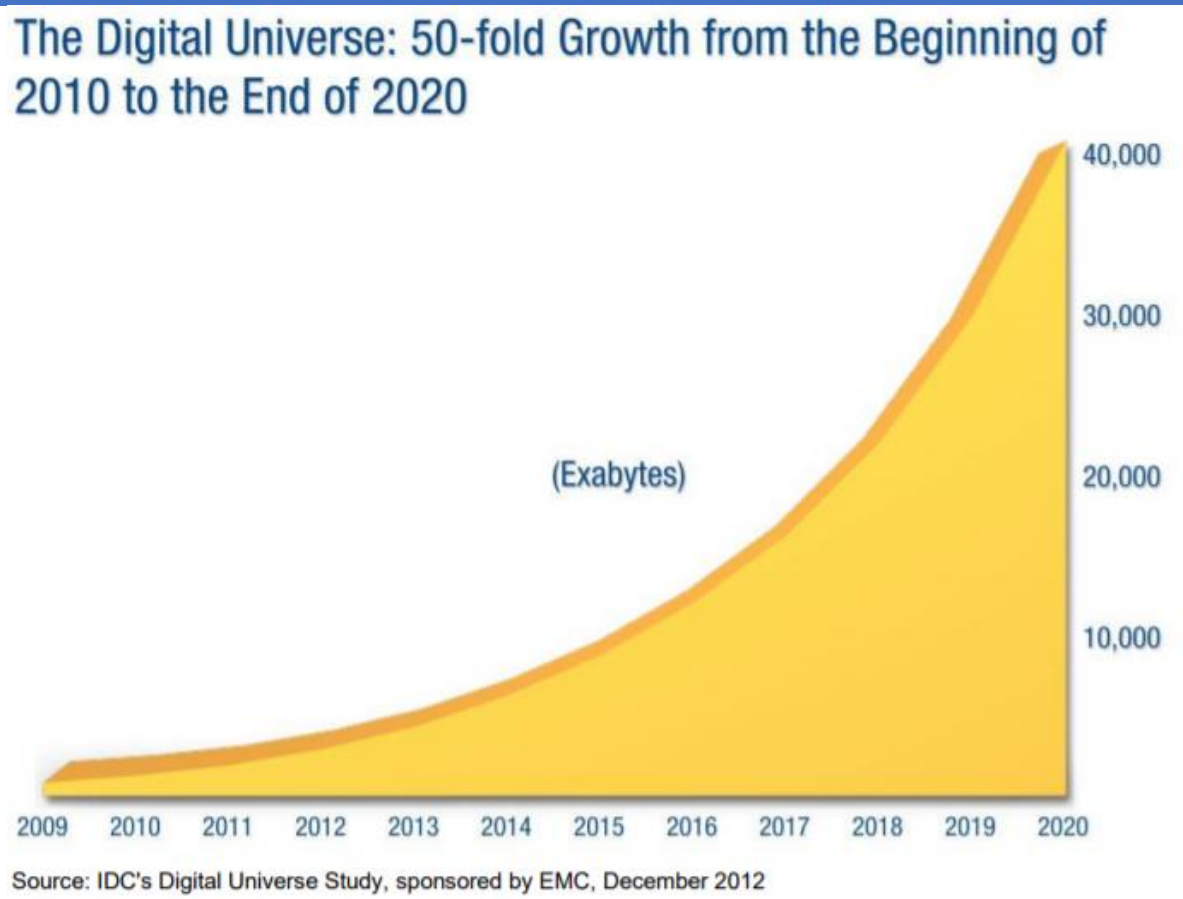
What is Big Data?

As you can see from the chart, the amount of data flying around is growing exponentially, doubling every two years or so.

Here's a way to think about how fast this is:

The world will produce as much digital data over the next 2 years, as currently existed in all of humanity prior to that.

And it will do the same the 2 years after that. And so on. That's a lot!



Unit 4 Lesson 1 (U4L1)
What is Big Data?

There is a principle in computer science known as Moore's Law.
Wikipedia: 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.

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 - we call it exponential growth. With more and more machines that are faster and faster, the amount of data being pushed around, saved, and processed is growing exponentially. 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 Wrap Up (10 mins)

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

Prompt (5 mins):

In your INB, respond to this prompt:

After your explorations what do you think "big data" actually means?

Unit 4 Lesson 1 (U4L1)
What is Big Data?

Here is a general-purpose definition of Big Data (taken from Wikipedia: Big Data):
“Big data is a broad term for datasets so large or complex that traditional data processing applications are inadequate."

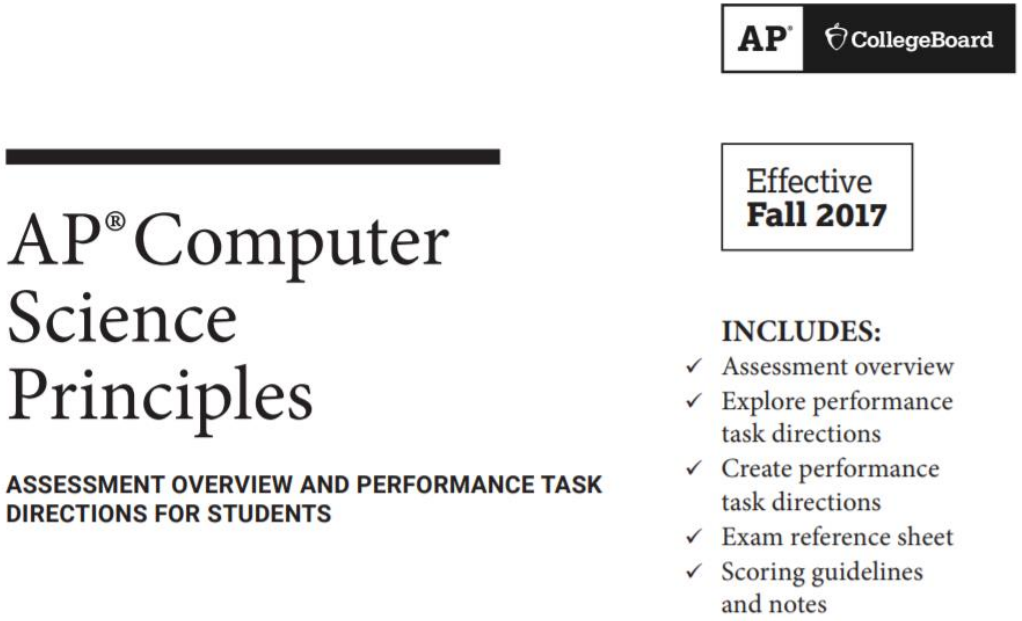
The fact that big data is increasingly important across industries reflects rapid changes in how much data we're collecting, and the ways we're using it.

In this unit we're going to be looking into how growth in data and computing more generally is impacting society. In almost every industry and every aspect of our lives, computing and data is affecting our lives in both positive and negative ways.

This will also be very useful preparation as we begin to look towards the Explore PT.

Introduce Explore PT (10 mins)

At the end of this unit we will be doing the Explore PT. To practice the different components of the PT we'll be practicing them throughout this unit. We're going to quickly review those components now, but we'll have opportunities to review and practice them in the next few lessons as well. For right now you don't need to understand all the details, just the big picture.



Page 4: The Explore PT has 2 major components, 1. computational artifact, 2. written responses

Pages 5-6: Skim the submission requirements and give students time to read prompts 2a - 2e.

Highlight prompts 2c and 2d which references beneficial / harmful effects and the way computing innovations use data.

Questions?

We will have many opportunities to review the Explore PT in this unit.

Quick Check-In