

BEN 25 – Big Data

Discussion Questions

1. Do you think Big Data will have a positive or negative impact on the future?
2. Are you concerned about your government or companies monitoring and collecting data about you?
3. Is there ten times more value with ten times more data?

Transcript

By now, **odds are** you've heard the term "**Big Data**". The thousands upon thousands of **giga, tera, and peta** bytes that are collected everywhere around us has caused concern for some, and jobs and wealth for others. As Wired magazine suggests, Big Data can help us understand, and predict, things like never before.

Make no mistake: the Big Data movement has done more than just introduce a **tidal wave** of aggressive marketing campaigns. While it's true that the "Big" in Big Data is an increasingly **ambiguous** term, there's no denying that it has **fundamentally changed** the manner by which organizations today make decisions.

Big Data has gotten a bit of a **bad rap** recently, bringing up **privacy concerns** all around the world. While companies and governments have been collecting data for many years, it's still too early to see what Big Data can really do. From the Boston Globe:

It's important to remember all the good things that can be done with Big Data at a time when so many are focused on the bad. Since former **NSA contractor** Edward Snowden **exposed** secret US government programs that collect information about phone calls and websites, **metadata** has become a **dirty word**.

Big Data refers to the **mountains** of information being collected, and is a term that has been **in circulation** since at least the 1990s. A **full** 90% of all the data in the world has been generated over the last two years alone, and as Edd Dumbill points out in his book, "Planning for Big Data", comes from just about everything:

As a **catch-all** term, Big Data can be pretty **nebulous**, in the same way that the term "**cloud**" covers diverse technologies. Input data could be **chatter** from social networks, **traffic flow** sensors, **satellite imagery**, broadcast audio streams, banking transactions, the content of web pages, scans of government documents, **GPS trails**, financial market data, the list goes on.

Whether it's helping politicians win elections, predicting **disease outbreaks**, or increasing a company's sales, the analysis of data can give groups and organizations an important **edge** when it comes to understanding their target customers. From Smart Data Collective:

Wal-Mart is able to take data from your past **buying patterns**, their internal stock information, your mobile phone location data, social media as well as external weather information and analyze all of this in seconds so it can send you a **voucher** for a barbeque cleaner to your phone – but only if you own a **grill**, the weather is nice, and you currently are within a 3-mile **radius** of a Wal-Mart store that has the barbeque cleaner **in stock**.

In London, rubbish bins join a **host** of everyday objects from televisions to toilets that are being manufactured with the ability to send and receive data, **opening up** new potential for interaction - and surveillance. As TIME reports, these **innocent-looking** garbage cans are just a part of a marketing campaign that could become **the new norm**:

If a company could see that a certain smartphone user spent 20 minutes in a McDonald's every day, it could approach Burger King about **airing** an ad on the bin's video display whenever that user walks by at lunchtime. Or it could **target** its commercials in real time by **distinguishing** between people who work in the area and visiting tourists.

With data sets in the multiple **petabyte**-range and up, the manpower required to analyze these **troves** of information is growing just as quickly. IBM is doing its part to create **much-needed** jobs, as Manager of Global Academic Programs Richard Rodts explains:

Leaders in business, education and government must take action to **foster** a new generation of talent with the **technical expertise** to make the most of this **tsunami** of Big Data. To narrow this **skills gap**, IBM is committed to **partnering with** universities around the world to make an **impact** in today's **data-driven** marketplace.

While Big Data may answer the "Why?" of some very complex things, it's important to remember the basics, as Wired concludes:

We still need ideas, **hypotheses** and **theories**. If you don't have good questions, your results can be **silly** and **meaningless**. Having more data won't **substitute** thinking hard, recognizing **anomalies**, and exploring **deep truths**.

Vocabulary

Odds are: the chances or probability is likely that.

Big Data: a collection of data sets so large and complex that it becomes difficult to process.

Giga, Tera, and Peta: prefix modifier/multiplier used to indicate the quantity of something (i.e. storage capacity, distance, or mass. 1,000 x Giga Byte = 1 Tera Byte.)

Make no mistake: a device for highlighting something that will likely happen; "Make no mistake about how difficult it is to find a job in this economy."

Tidal wave: a huge amount; "Oil production in northern Canada has created an economic tidal wave for the federal government."

Ambiguous: more than one interpretation, or having a double meaning; "Confusing, ambiguous communication won't relay the correct news."

To fundamentally change: to completely alter the properties or structure of an idea or concept.

Bad rap: Unjustified criticism; "These new regulations have been getting a bad rap, but they're actually great for small businesses."

Privacy concern: fear or worry over the issue of privacy.

NSA: National Security Agency; The American government agency responsible for the global monitoring, collection, and analysis of information and data for foreign intelligence and purposes.

Contractor: a worker or company that is either employed on a specific contract (usually for a specified period of time) or via another company; "We're looking for a contractor to help us with our office renovations, do you know of someone reliable?"

To expose: to show or uncover, make known; "Reporters exposed exactly how far the government scandal went."

Metadata: data that provides information about other data.

Dirty word: unpleasant, rude, or inappropriate word.

Mountains: a huge amount; "Even a small accident at work can result in mountains of paperwork if a claim is filed."

(To be) In circulation: in use or available for use; "Police have found a large amount of fake 50euro notes in circulation around Milan."

Full: complete, especially in detail, number, or duration; "A full 3 months went by before I heard anything about my job interview, but they finally offered me the position!"

Catch-all/catchall: something that encompasses a wide variety of items or situations.

Nebulous: lacking definite form or limits, vague; "While customer satisfaction may seem nebulous, it's the most important element to increased sales."

Cloud (computing): the practice of using a network of remote servers hosted on the Internet to store, manage, and process data.

Chatter: talk, discussion, or activity; "The chatter around the coffee machine this morning was all about the new merger."

Traffic flow: interactions between vehicles, drivers, and infrastructure.

Satellite imagery: images of Earth or other planets collected by artificial satellites.

GPS trails: data gathered from a person's GPS (Global Positioning System) movements.

Disease outbreak: a sudden rise in the incidence of a disease.

To give the/an edge over: to give/get an advantage; "Extra tutoring with a qualified native speaker can give you an edge over the other students."

Buying patterns: typical manner in which consumers purchase goods or services in terms of amount, frequency, or timing.

Voucher: coupon, small printed piece of paper that entitles the holder to a discount or that may be exchanged for goods or services; "Our company is offering 20%-off vouchers for any customer who spends more than \$100 on their next purchase."

Grill: Barbeque/BBQ, an outdoor fireplace primarily used for cooking meat; "Jules just bought a new grill, and has invited everyone over to his house for a BBQ this Saturday."

Radius: a line segment that joins the center of a circle with any point on its circumference; "Our marketing department has started to send fliers to every home within a 5km radius of our store."

(To be) In stock: available for sale; "I'm sorry, we don't have that item in stock currently. Would you like me to check at our other locations?"

Host: a large number of something; "There's a whole host of reasons why he didn't get the job."

To open up: to allow or establish; "A lot of new investment opportunities have opened up in China over the last 3 years."

Innocent-looking: having the appearance of being innocent or safe; "Dexter has an innocent-looking appearance, but he's actually a very aggressive manager."

Surveillance: to watch or monitor something/somebody, often in secret; "Government surveillance of private communications has become an increasing concern."

The new norm: a previously unfamiliar situation that has become standard, usual, or expected; "According to a recent survey 27% of Germans say spending less is the new norm."

To air: to broadcast or transmit; "New TV series are usually aired every year in September."

To target: to focus or direct; "This marketing campaign is targeted towards young urban professionals."

To distinguish: to differentiate, recognize, or understand what makes someone or something different; "If you want to be successful, you need to distinguish yourself from your competitors."

Petabyte: 1000 terabytes.

Trove: a collection of objects, especially valuable ones; "LinkedIn has given us troves of candidates for the intern position."

Much-needed: something that is needed badly; "After working 17 consecutive hours, I went home for some much-needed sleep."

To foster: to nurture or promote; "A teacher's primary responsibility is to foster learning."

Technical expertise: superior knowledge related to technical subject matter.

Tsunami: tidal wave, significant amount.

Skills gap: the difference between knowledge and ability.

To partner (up) with: to become partners; "We partnered with an IT company in order to complete the system updates more quickly."

Impact: effect or influence; "The environmental impact of constantly burning fossil fuels has added to the effects of global climate change."

Data-driven: moved forward or progressed through the increase and addition of more data.

Hypothesis: an educated or informed guess.

Theory: a system of rules, procedures, and assumptions used to produce a result; "It's just a theory, but I think the problem we've had increasing sales is directly related to our lack of new products."

Silly: having a lack of wisdom or good sense, not to be trusted; "Silly expenditures are causing us to lose a lot of money."

Meaningless: having no meaning or significance; "In the world of business, there are no meaningless decisions, every one is important."

To substitute: to replace, either permanently or temporarily; "We are looking at how to substitute imported products with local ones."

Anomaly: deviation or departure from the normal or common order, form, or rule.

Deep truth: real facts, data, or values.

Language Review

A. Collocations Table

Match the words in each column to make collocations from today's lesson.

1. Dirty	a. Chatter	
2. Privacy	b. Outbreak	
3. Social Network	c. Between	
4. Make	d. An Edge	
5. Fundamentally	e. Word	
6. Give	f. Concern	
7. Distinguish	g. Change	
8. Disease	h. An Impact	

B. Review Quiz

1. According to the article, which of the following is NOT a problem affecting Big Data?
 - a. Privacy concerns
 - b. Data sets are too small
 - c. Not enough analysts
 - d. Too many sources
2. According to the article, which of the following is NOT a source for collecting data?
 - a. GPS device
 - b. Rubbish bins
 - c. Facebook/Twitter
 - d. Barbeques
3. "Many employees are protesting the _____ policy changes requiring them to work unpaid overtime during the holidays."
 - a. Much-valued
 - b. Much-hated
 - c. Much-admired
 - d. Much-loved
4. _____ creates complexity and means decision making is difficult.
 - a. Ambiguity
 - b. Ambiguously
 - c. Ambiguous
 - d. Ambiguousism

5. "The items that I was looking for were _____, so I had to try a different shop."
a. In stock
b. Out of stock
c. In stocked
d. Out of stocked
6. "I find it very _____ that we haven't been able to fix the supply chain problem we've been having for the last 6 months."
a. Concern
b. Concerned
c. Concerns
d. Concerning
7. Which of the following is closest in meaning to "technical expertise"?
a. Technical know-how
b. Technical background
c. Technical expert
d. Technical requirements
8. Our company hopes to become greener by _____ traditional paper materials with digital alternatives.
a. Substitution
b. Substituting
c. Subs
d. Substitutes
9. "We've received _____ of requests for more information about this Friday's big sale."
a. Mountains
b. Troves
c. A host
d. All of the above
10. When dealing with Big Data, data sets are generally:
a. Gigabytes or smaller
b. Metabytes and up
c. Petabytes or higher
d. No bigger than terabytes

Answer

A. Collocations Table

Match the words in each column to make collocations from today's lesson.

1. Dirty	a. Chatter	Dirty Word
2. Privacy	b. Outbreak	Privacy Concern
3. Social Network	c. Between	Social Network Chatter
4. Make	d. An Edge	Make An Impact
5. Fundamentally	e. Word	Fundamentally Change
6. Give	f. Concern	Give An Edge
7. Distinguish	g. Change	Distinguish Between
8. Disease	h. An Impact	Disease Outbreak

B. Review Quiz

1. C, 2. D, 3. B, 4. A, 5. B, 6. D, 7. A, 8. B, 9. D, 10. C

Online Practice

Click the "Launch" button to open the **online practice**:

Launch Quiz