

# Managing Information in a Digital World

---

[Big Data](#) | [Information Systems](#) | [Electronic Spreadsheets](#) | [Database Connectivity](#) | [Digital Data & Data Compression](#) | [Electronic Commerce](#) | [Review](#)

---

## What is Big Data?



Can you define "Big Data?" No, data that is big is not a specific definition of *Big Data*. So what is big data? Is that the burning question of 2016 or is it just another technology trend?

**Big data** is a broad term for data sets so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, and information privacy. The term often refers simply to the use of predictive analytics or other certain advanced methods to extract value from data, and seldom to a particular size of data set. Accuracy in big data may lead to more confident decision making. And better decisions can mean greater operational efficiency, cost reductions and reduced risk.

Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big data *size* is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data. Big data is a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale.

Computing and big data are seemingly everywhere in our digital world, but most of the time we are oblivious to how our data is being collected, and for what purpose it is being used. According to Best-Selling Author, Keynote Speaker and Leading Business and Data Expert Bernard Marr, "Big Data is one of those mega trends that will impact everyone in one way or another."

How does Big Data get generated? Where does it come from and who is contributing to the collection and generation?

### **The "V"s of Big Data**

Specifically, Big Data relates to data creation, storage, retrieval and analysis that is remarkable in terms of volume, variety, velocity, and variability:

- **Volume:** The quantity of data that is generated is very important in this context. It is the size of the data which determines the value and potential of the data under consideration and whether it can actually be considered Big Data or not. The name *Big Data* itself contains a term which is related to size and hence the characteristic. A typical PC might have had 10 gigabytes of storage in 2000. Today, Facebook consumes 500+ terabytes of new data every day; a Boeing 737 will generate 240+ terabytes of flight data during a single flight across the United States; the spread of smart phones, the data they create and consume; sensors embedded into ordinary objects will soon result in billions of new, constantly-updated data feeds containing environmental, geolocation, and other information, including video.
- **Variety:** This means that the category to which Big Data belongs to is also a very essential fact that needs to be known by the data analysts. This helps the people, who are closely analyzing the data and are associated with it, to effectively use the data to their advantage and thus upholding the importance of the Big Data. Big Data data is not just

numbers, dates, and strings. Big Data is also geospatial data, 3D data, audio and video, and unstructured text, including log files and social media.

- **Velocity:** The term *velocity* in the context refers to the speed of generation of data or how fast the data is generated and processed to meet the demands and the challenges which lie ahead in the path of growth and development. For example, Clickstreams and ad impressions capture user behavior at millions of events per second; high-frequency stock trading algorithms reflect market changes within microseconds; machine to machine processes exchange data between billions of devices; infrastructure and sensors generate massive log data in real-time; on-line gaming systems support millions of concurrent users, each producing multiple inputs per second.
- **Variability:** This is a factor which can be a problem for those who analyze the data. This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.

*BIG DATA EXPLAINED, MongoDB, Inc., <http://www.mongodb.com/big-data-explained>*

## **The Future of Big Data**

Self-driving cars were just the start. What's the future of big data-driven technology and design? In a thrilling science talk, Kenneth Cukier looks at what's next for machine learning and human knowledge.

There is a mind-boggling amount of data floating around our society. Physicists at CERN have been pondering how to store and share their ever more massive data for decades - stimulating globalization of the Internet along the way, while *solving* their big data problem. Tim Smith plots CERN's involvement with big data from fifty years ago to today.

## **Big Data Review**

[Quiz created by sarush with GoConqr](#)

## The Secret Life of Big Data

One of the concerns about Big Data should be your privacy. With companies like Kroger (Kroger loyalty card) tracking your spending habits and Wal-Mart tracking your purchases (Walmart Savings Catcher) how much information are people readily handing over? I personally find it a little disturbing to see ads in my Facebook newsfeed from Home Depot for the exact item that I just search Amazon for. Here is another example of our privacy not being so private. Read the article: [How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did](#)

Intel's Genevieve Bell shows that we have been dealing with big data for millennia, and that approaching big data problems with the right frame of reference is the key addressing many of the problems we face today from the keynote of Supercomputing 2013.

### Assignment: Discussion

**Directions:** Why is the new profession of data scientist blossoming now? Data analytics tends to involve hypothesis testing while data mining involves discovery. What are the strengths and weaknesses of each technique when applied to Big Data. Post your response to the Data Scientist discussion board in itslearning and reply to at least two classmates to continue the collaborative discussion.

### Assignment: Question for Thought 1

**Directions:** The key characteristics of Big Data are often called the 3Vs: Volume, Velocity, and Variety. Explain the role each one of these plays in making Big Data hard to handle. The complexity of some data sets leads people to add extra dimensions such as Veracity and Variability to this list. In what ways do these complicate things further? Your essay should be approximately 100 words. Be sure to run spellcheck before submitting your assignment.

---



Your grades for this class are managed with a spreadsheet. I put your individual assignment grades in a spreadsheet and average them to create the unit grade before transferring to the gradebook.

## Information Systems

Computers exist to manage and analyze all types of data. **Data** is defined as raw facts and **information** is data that has been organized to help us answer questions and solve problems. An **information system** helps users organize and analyze data. Three of the most popular general application information systems are:

- electronic spreadsheets
- database management systems
- electronic commerce (e-commerce)

## Assignment: Question for Thought 2

**Directions:** Explain how computational manipulations of information require consideration of (1) representation, (2) storage, (3) security, and (4) transmission. Your essay should be approximately 500 words and must include at least 2 cited sources. Be sure to run spellcheck before submitting your assignment.

---

## Electronic Spreadsheets

Spreadsheets are useful in many situations and they are often designed to manage thousands of data values and calculations. Sorting, querying, and reporting data are just a few of things that spreadsheets are used for. A spreadsheet is an interactive computer application program for organization, analysis and storage of data in tabular form. Spreadsheets developed as computerized simulations of paper accounting worksheets. The program operates on data represented as cells of an array, organized in rows and columns (Wikipedia). The accounting spreadsheet was computerized in 1961 and has come a long way. Microsoft Excel is the most widely used electronic spreadsheet and is considered the market leader.

---

## ***What if...***

*Attendees decreased by 15%?*

*Ticket prices are increased by \$5?*

*I score 70% on the exam?*

*Costs are reduced by 50%?*

---

One reason spreadsheets are so useful is their versatility. The user of the spreadsheet determines what the data represents and how it is related to other data. Therefore, spreadsheet analysis can be applied to just about any topic area. Spreadsheets might be used to:

- track sales
- analyze sport statistics
- maintain student grades
- keep a car maintenance log
- record and summarize travel expenses
- track project activities and schedules
- plan stock purchase
- calculate your grades

The dynamic nature of spreadsheets provides the powerful ability to carry out **what-if analysis**. We can set up spreadsheets that take into account certain assumptions, and then challenge those assumptions by changing the appropriate values. By using formulas within the spreadsheet, we can easily change the data to get quick answers to the **what-if** questions.

## **Practice: Excel 2010**

**Directions:** Excel 2010 is a spreadsheet application in the Microsoft 2010 Office suite that lets you store, manipulate, and analyze data in organized workbooks for home and business tasks. New innovations in Excel 2010

include the enhanced data viewing features of sparklines and slicers. Work through the 21 Excel lessons below. Direct access to the GCF LearnFree.org website is: <http://www.gcflearnfree.org/office2010/excel2010>.

## **Assignment: Excel 2010 Lab 12A**

**Directions:** Download Spreadsheet Lab12A from itsLearning. Read the Lab material and then complete Exercises 1 & 2. You will upload Exercise 1 and Exercise 2 into individual itsLearning assignments.

## **Assignment: Quest Specialty Travel**

**Directions:** Brenda Greene, Finance Director, has asked you to complete the weekly payroll analysis so she can finish the payroll for the week. She will need you to calculate regular pay, overtime pay, and gross pay for employees at Quest Specialty Travel, a marketing agency located in Lewiston, Michigan. Download the assignment sheet here: [Quest Specialty Travel](#).

---

# **Database Connectivity**

Most Web sites do more than simply present text, a few image files and a couple of documents. They are also collecting data and using that data for a variety of purposes. The amount of data being generated, stored, and processed is growing by leaps and bounds. According to a McKinsey Global Institute reports, it is estimated that in 2010 alone global enterprises stored more than 7 exabytes (a billion gigabytes) of data while consumers stored more than 6 exabytes of new data on devices such as PCs, smartphones, tablets, and notebooks. That is a lot of data! Can you image how much data we are generating and storing now?

Fully functional Web sites also include database connectivity. Databases provide the ability to:

- Present stored information to customers, and allow them to search and retrieve.
- Receive information from customers and save it for later retrieval.

Almost all sophisticated data management situations rely on an underlying database and the support structure that allows the user (either human or a program) to interact with it. A **database** can simply be defined as a structured set of data. A **database management system** is a combination of software

# SQL

and data.

Programs, like SQL, are put in place by businesses and other organizations as a way to access and manipulate the information and data that is stored in their databases.

There are several database types:

- Flat file: information is stored in a single table, often in a simple text file. The Windows registry is an example of a flat-file database.
- Non-relational: information is store statically. Information can be searched, but cannot be reorganized or placed into another database.
- Relational: information can be sorted, altered, and placed into other databases for retrieval.
- Object-oriented: a newer form with the capabilities of a relationship database, plus greater storage and search efficiency.

A database must be carefully designed from the outset if it hopes to fulfill its role. Poor planning in the early stages can lead to a database that does not support the required relationships.

Database programs are less standard and generally much more expensive than spreadsheets. Microsoft Access, which is a part of the Microsoft Office Suite, is one of the most popular database programs. SQL is another popular database especially for big data.

## Did you know?

Computers and storage devices are full of data, and there are many different forms of data, depending on how often the data are accessed or modified. Persistent data are those that are typically not accessed and rarely modified. Database persistent data are typically stored on a server and are more commonly accessed than archived data. With archived data, or those stored on disks or tapes, the information is very rarely opened or used. Aside from archiving the data, this allows researchers to go through old or stored information to find past trends that may apply to present situations.

**Persistent data** are very rarely modified; this means the information stored within the database, disk or tape is not changed, except for special occasions. The information being accessed is more common than the information being modified, but it is still rarely done. These data also exist from one session to the next, unlike data types that only exist for one session and are then discarded or bound to that single session.

With database persistent data, an entire database or a section of a database is created to hold the archived data. This can be done locally, on a database stored on the computer's hard drive, or it can be placed on a server. This persistent information is more commonly accessed than the tape and disk variant, because the information is readily available. At the same time, this database will typically exist untouched for months or years. <http://www.wisegeek.com/what-is-persistent-data.htm>

## Structured Query Language

Structured Query Language (SQL: pronounced "ess-que-el") is a language used to create and maintain professional, high-performance corporate databases. SQL is at the heart of all relational databases, including IBM's DB2, Oracle, Microsoft's SQL Server, and open source database MySQL. SQL programs are put in place by businesses and other organizations as a way to access and manipulate the information and data that is stored in their databases, as well as for creating and altering new tables. SQL was devised for manipulating data in relational database tables. According to Chad Brooks, BusinessNewsDaily, "Currently, many of the world's largest and most well-known brands rely on MySQL to make their websites function properly, including Facebook, Google, Adobe, Alcatel Lucent and Zappos."

SQL is a comprehensive database language for managing relational databases. It includes statements that specify database schemas as well as statements that add, modify, and delete database content. SQL also provides the ability to query the database to retrieve specific data. SQL is not case sensitive, so keywords, table names, and attribute names can be uppercase, lowercase, or mixed case. Spaces are used as separators in a statement. SQL has emerged as the de facto language for big data:

- SQL is a natural language for data analysis: The concept of SQL is underpinned by the relational algebra - a consistent framework for organizing and manipulating sets of data - and the SQL syntax concisely and intuitively expresses this mathematical system.
- SQL is a productive language for writing queries: When a developer writes a SQL query, he or she simply describes the results that they want. The developer does not have to get into any of the nitty-gritty of describing how to get the results.
- SQL queries can be optimized: The fact that SQL is a declarative language not only shields the developer from the complexities of the underlying query techniques, but also gives the underlying SQL engine has a lot of flexibility in how to optimize any given query.
- SQL is extensible: SQL provides a robust framework that adapts to new requirements.

SQL is an immensely popular language today and if anything its popularity is growing as the language is adopted for new data types and new use cases. *Why SQL is becoming the goto language for Big Data analysis, Klaker-Oracle on Sep 26, 2014*

It's time to learn a little about working with SQL. The practice will give you hands on experience of working with SQL statements.

**Practice: [SQLCourse.com](http://SQLCourse.com)**

**Directions:** Welcome to SQLCourse.com! This unique introductory SQL tutorial not only provides easy-to-understand SQL instructions, but it allows you to practice what you learn using the on-line SQL interpreter. You will receive immediate results after submitting your SQL commands. You will be able to create your own unique tables as well as perform selects, inserts, updates, deletes, and drops on your tables. This SQL tutorial currently supports a subset of ANSI SQL. The basics of each SQL command will be covered in this introductory tutorial. Unless otherwise stated, the interpreter will support everything covered in this course. You will follow steps 1-8 below and complete the activities.

## **Assignment: Learn SQL on Code Academy**

**Directions:** Visit [Code Academy](#) and create an account. Complete the *Learn SQL* course. You will learn to manage data with SQL by mastering complex commands to manipulate and query data stored in relational databases.

### **Database Review**

- SQL (relational) databases have a mature data storage and management model. This is crucial for enterprise users.
- SQL databases support the notion of views which allow users to only see data that they are authorized to view. The data that they are not authorized to see is kept hidden from them.
- SQL databases support stored procedures which allow database developers to implement part of the business logic into the database.
- DBMS stands for Database Management System.
- We have four major types of DBMSs namely: Hierarchical, Network, Relational, Object Oriented
- The most widely used DBMS is the relational model that saves data in table formats. It uses SQL as the standard query language
- SQL is standard language used to query a database
- The database approach has many advantages when it comes to storing data compared to the traditional flat file based systems

## **Assignment: Database Lab 12B**

**Directions:** Download Database Lab12B from itsLearning. Read the Lab material and then complete Exercises 1, 2, 3, and 4. You will upload Exercise 1, Exercise 2, Exercise 3, and Exercise 4 into individual itsLearning assignments.

---

# Digital Data & Data Compression

## Compression

Digital data can be compressed and often times, it is necessary to compress images, audio files, and videos for transmission on the Web. Images, video, and sound all contribute to the size of a web page which translates to load time. Google's research has found that a half-second longer load time for search results decreased traffic and ad revenue by 20%. Amazon found that its revenue increased by 1% for every 100 milliseconds faster the site loaded. Back in 2001, a study found that the longest a typical user would wait for a Web page to load is eight seconds. Today, this number is way too long for most Internet users. A much higher percentage of users have high-speed Internet connections. More recent studies have found that most broadband users won't wait four seconds for a page to load.

In digital signal processing, data compression, source coding, or bit-rate reduction involves encoding information using fewer bits than the original representation. Compression can be either lossy or lossless. Lossless compression reduces bits by identifying and eliminating statistical redundancy. No information is lost in lossless compression. Lossy compression reduces bits by identifying unnecessary information and removing it. The process of reducing the size of a data file is referred to as data compression. In the context of data transmission, it is called source coding (encoding done at the source of the data before it is stored or transmitted) in opposition to channel coding.

Two main types of compression are lossy and lossless.

Lossy compression works by discarding aspects of an image that are insignificant. For example, if a photo contains 30 different shades of black, lossy compression will get rid of some of those shades. It is referred to as "lossy" because lossy image compression results in a loss of image fidelity. When used correctly, it's difficult or impossible for most people to detect. Lossy compression:

- method which loses some of the data that is indiscernible to the eye
- saves as a smaller size which equals shorter download time
- repeatedly compressing and decompressing a lossy file will cause it to progressively lose quality

Lossless compression compresses images in such a way that they can be exactly reproduced from the compressed file with no loss of fidelity. Lossless compression works great for icons, clip art, logos, buttons, and the like. The most popular lossless compression formats on the Web are GIF and PNG because of the limitation in colors. Lossless images:

- allows the exact original data to be reconstructed from the compressed data
- used in cases where it is important that the original and decompressed data be identical or where deviations from the original could be damaging

You can read more about image compression in Cameron Chapter's blog, [Everything You Need to Know About Image Compression](#). The blog includes create visual examples and is worth reading.

**Audio data compression**, as distinguished from dynamic range compression, has the potential to reduce the transmission bandwidth and storage requirements of audio data. Audio compression algorithms are implemented in software as audio codecs. Lossy audio compression algorithms provide higher compression at the cost of fidelity and are used in numerous audio applications. These algorithms almost all rely on psychoacoustics to eliminate less audible or meaningful sounds, thereby reducing the space required to store or transmit them. You can read more about audio compression in Ian Corbett's article [What Data Compression Does To Your Music](#)

The MP3 audio compression algorithm drastically reduces the size of music files so that we can store more songs on our mobile phones and music players. "This compression scheme has revolutionized the music industry (for better or worse)." Saving data space is "better," certainly. What is "worse" about MP3s?

 recently reformatted most videos away from Flash, which requires a plugin to MP4. MP4 often uses the H.264 video codec, which is native to most browsers that support HTML5, and uses far less processor and battery power.

**Video compression** uses modern coding techniques to reduce redundancy in video data. Most video compression algorithms and codecs combine spatial image compression and temporal motion compensation. Video compression is a practical implementation of source coding in information theory. In practice, most video codecs also use audio compression techniques in parallel to compress the separate, but combined data streams as one package.

The majority of video compression algorithms use lossy compression. Uncompressed video requires a very high data rate. Although lossless video compression codecs perform an average compression of over factor 3, a typical MPEG-4 lossy compression video has a compression factor between 20 and 200. As in all lossy compression, there is a trade-off between video quality, cost of processing the compression and decompression, and system requirements. Highly compressed video may present visible or distracting artifacts.

Some video compression schemes typically operate on square-shaped groups of neighboring pixels, often called macroblocks. These pixel groups or blocks of pixels are compared from one frame to the next, and the video compression codec sends only the differences within those blocks. In areas of video with more motion, the compression must encode more data to keep up with the larger number of pixels that are changing. Commonly during explosions, flames, flocks of animals, and in some panning shots, the high-frequency detail leads to quality decreases or to increases in the variable bitrate.

When digital footage is shot using a camcorder, the files created are often very large. These uncompressed files represent the raw moving image, as the

camera captured it live. Unfortunately, this footage is rarely convenient to upload and store. This is where video compression comes in. So, what exactly is video compression and how does it work? In this video, LockerGnome's Brandon Wirtz explains how video compression works, and why it's so important to get right.

The MP4 video format (MPEG 4 files) is widely used for video format. Mp4 often uses the H.264 video codec. The H.264 codec is also used by Apple's mobile devices and YouTube for video playback.

The WebM and Ogg video formats are also used. The WebM video format often uses the VP8 codec which is an open video compression format owned by Google. Ogg uses the Theora format for HTML5 video, which is a free video compression format that can be distributed without licensing fees.

Compression is useful because it helps reduce resource usage, such as data storage space or transmission capacity. Because compressed data must be decompressed to use, this extra processing imposes computational or other costs through decompression; this situation is far from being a free lunch. Data compression is subject to a space-time complexity trade-off. For instance, a compression scheme for video may require expensive hardware for the video to be decompressed fast enough to be viewed as it is being decompressed, and the option to decompress the video in full before watching it may be inconvenient or require additional storage. The design of data compression schemes involves trade-offs among various factors, including the degree of compression, the amount of distortion introduced (when using lossy data compression), and the computational resources required to compress and uncompress the data.

## **Assignment: Compare/Contrast Formats**

**Directions:** Create a presentation to compare and contrast file formats. Your presentation must compare and contrast two image formats, two audio formats, and two video formats. Download the full assignment from itsLearning.

## **Assignment: Question for Thought 3**

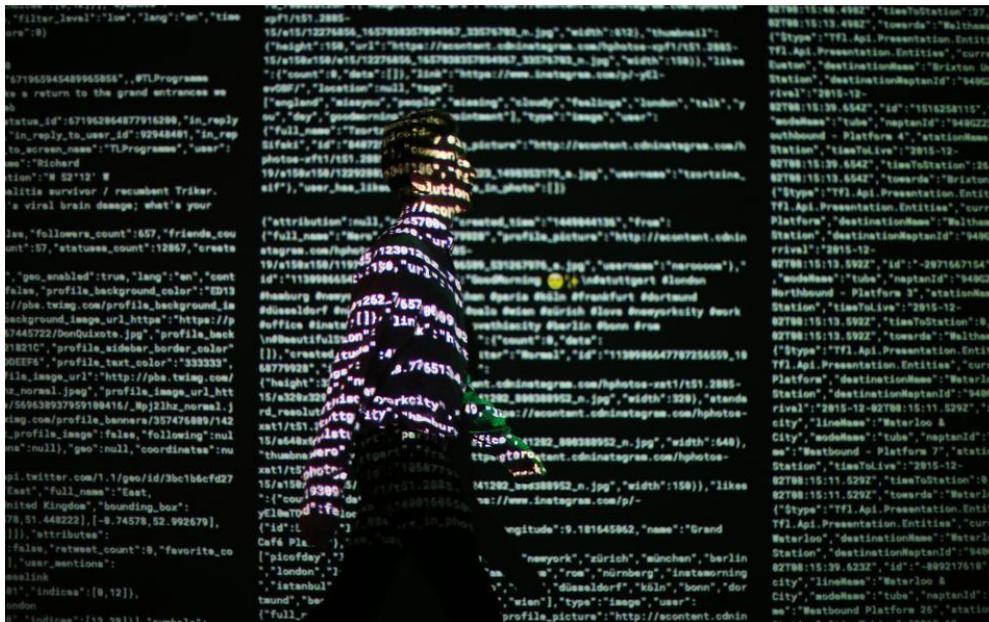
**Directions:** Debate the trade-offs in representing information as digital data. Your essay must include at least three **advantages** and **disadvantages** of

digital data representation. Your essay should be approximately 500 words. Be sure to run spell check before submitting your assignment.

## Electronic Commerce

The Internet and World Wide Web have greatly affected the way most of us live and do business. Frequently named as the fastest-growing sector of the Internet economy, Web-based e-commerce, the act of doing business transactions over the Internet or similar technology, is redefining the way business operate and compete in the 21st century. It has been estimated that over 50,000 US companies make most or all of their money online and Internet-based revenue is expected to exceed over one trillion dollars annually shortly. In addition, the Web influences offline sales, such as the scores of consumers who research purchases they eventually make offline.

**E-commerce** - conducting business transactions - generally financial transactions - online.



Despite the economic slowdown and national security concerns, e-commerce in the United States has continued to grow at a steady pace. According to Forester Research, online sales from even as far back as 2003 were expected to total \$96 billion (a total of 4.5% of all retail sales), up over 48% since 2001. "E-commerce will continue to outgrow traditional retail, as the Internet appears to growing numbers of consumers in search of the best deals, convenience

and breadth of offerings." according to Michelle David Adams, comScore Networks vice president.

E-commerce sales continue to grow rapidly, having topped \$200 billion in 2011. Forrester expects that online sales will grow from 7% of overall retail sales to close to 9% by 2016. Key drivers of this growth include consumers' greater comfort level with purchasing various categories online, broader web shopping capabilities with mobile and tablet devices, innovative new shopping models that divert spend away from physical stores (e.g., flash sales, subscription models), online loyalty programs, and aggressive promotional offers from web retailers. (*Sucharita Mulpuru with Vikram Sehgal, Patti Freeman Evans, Andy Hoar, Douglas Roberge in US Online Retail Forecast, 2011 To 2016*)

In a report released today, Forrester Research Inc. forecasts that business-to-business e-commerce sales in the United States will reach \$780 billion this year—more than twice the most recent figure of \$304.91 billion in U.S. retail e-commerce sales released by the U.S. Department of Commerce, for 2014—and is on course to grow at a compound annual growth rate of 7.7% until it reaches an estimated \$1.13 trillion in 2020.

The growth will be driven largely by "channel-shifting" B2B (business-to-business) buyers who are buying more online than through phone and other offline channels, and the opportunity for manufacturers, wholesalers and distributors to cut operating costs by processing more sales to customers through self-service e-commerce sites and electronic processing of orders, Forrester says in the report, "US B2B eCommerce Forecast: 2015 to 2020." *Paul Demery, B2B e-commerce sales will top \$1.13 trillion by 2020*

## **Assignment: E-commerce Current Event**

**Directions:** You have read about the growth of e-commerce. Use your research skills and find an article from the most recent Christmas holiday and summarize the impact of the electronic shopping on the economy. Be sure to cite your source to receive full credit.

## **Assignment: Big Data Balance article summary**

**Directions:** Read the article [\*How I Stopped Worrying And Found Balance In Big Data\*](#) (both pages). Note **three** things you found interesting in the article about big data, describe the information, and explain why you choose each

piece of information. Submit your response directly to the itsLearning assignment box. Do not submit a separate file and be sure to run spellcheck prior to submitting.

### **Assignment: Question for Thought 4**

**Directions:** The digital divide puts those who have access to technology on one side and those who do not on the other side. Do you feel that it is the right of everyone to have access to technology? It will cost a great deal of money to erase the digital divide. Who do you think should be responsible for paying the cost? Explain your answers in approximately 150 words.

### **Assignment: How we found the worst place to park in New York City -- Using Big Data**

**Directions:** Watch the video on Big Data in itsLearning. What did you learn from the video? How do you think that Big Data affects your life on a daily basis? Provide at least one additional resource from the World Wide Web to support your analysis. Cite your source. Your submission should be approximately 100 words.

---

## **Review**

[Flash Card Deck created by sarush with GoConqr](#)

## **Resources**

If you are having problems viewing this page, opening videos, or accessing the URLs, the direct links are posted below. All assignments are submitted in itsLearning. If you have having problems, contact Mrs. Rush through the itsLearning email client.

GCFlearningFree.org: <http://www.gcflearnfree.org/office2010/excel2010>.

Information Technology - Electronic Spreadsheet:  
[https://www.youtube.com/watch?v=R\\_F1VzBg1IU&feature=player\\_embedded](https://www.youtube.com/watch?v=R_F1VzBg1IU&feature=player_embedded)

Explaining Big Data: [https://www.youtube.com/watch?v=7D1CQ\\_LOizA](https://www.youtube.com/watch?v=7D1CQ_LOizA)

What is Big Data?: <https://www.youtube.com/watch?v=c4BwefH5Ve8>

Exploration on the Big Data Frontier: <https://www.youtube.com/watch?v=j-0cUmUyb-Y>

Big Data Review: <https://www.goconqr.com/en-US/p/2767372-Big-Data-quizzes>

The Secret Life of Big Data | Itel: <https://www.youtube.com/watch?v=CNoi-XqwJnA>

SQLCourse: <http://www.sqlcourse.com/>

Code Academy: <https://www.codecademy.com/>

The Science and Application of Data Compression Algorithm:  
<https://www.youtube.com/watch?v=ZEQRz7BmGtA>

Understanding Lossy and Lossless Compression:  
<https://www.youtube.com/watch?v=2Qo5prktYNQ>

How Does Video Compression Work?:  
<https://www.youtube.com/watch?v=kzytYavfFMs>

How I Stopped Worrying And Found Balance In Big Data:  
<http://www.forbes.com/sites/chrismyers/2015/12/14/how-i-stopped-worrying-and-found-balance-in-big-data/>

What is a Database & SQL?:  
<https://www.youtube.com/watch?v=FR4QIeZaPeM>

Digital Compression explained by Aloe Blacc:  
<https://www.youtube.com/watch?v=By30SCp-Tsw>

Transcript: <http://mrsrush.net/digitaldesign/state/index.pdf>

---

### Credits

Big Data picture: <https://www.linkedin.com/pulse/illegal-interview-questions-bob-harrington-cpc>

Why SQL is becoming the goto language for Big Data analysis:  
[https://blogs.oracle.com/datawarehousing/entry/why\\_sql\\_is\\_becoming\\_the](https://blogs.oracle.com/datawarehousing/entry/why_sql_is_becoming_the)

Data picture (Photo by Peter Macdiarmid/Getty Images for Somerset House)

---