Andrea Marcelli
CIS 2532 NET01
December 10, 2022
Research Big Data

# Big Data

Assignment:

Write a paper explaining what is Big Data? Why is Big Data important? What are the types of Big Data? An industry analyst Doug Laney articulated the now-mainstream definition of Big Data as the three V's. Explain what are the three V's.

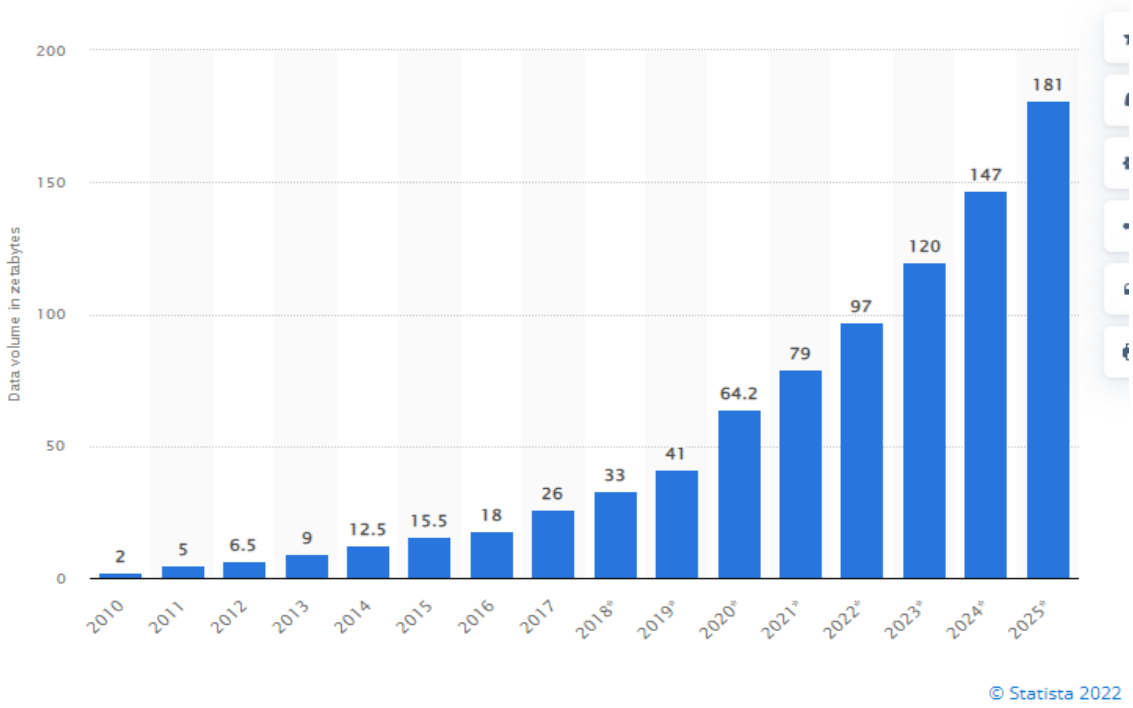
## Research:

What is Big Data: Big Data is referred to as the large quantity of data that has to be stored and managed by large databases around the world. Big Data is more complex and different from traditional data processing and management techniques. Nowadays we have reached certain quantities of data that cannot be structured anymore, most of them are unstructured (examples: Twitter, Facebook, Youtube, and other services). Big Data has existed since the 60s/70s with its first relational databases. However, during the last two decades, we have seen an impressive increase in the capacity of data. Note that data nowadays is calculated in Zettabytes which according to an article on Cisco:

> "A zettabyte is a measure of storage capacity and is 2 to the 70th power bytes, also expressed as 1021 (1,000,000,000,000,000,000,000 bytes) or 1 sextillion bytes. One Zettabyte is approximately equal to a thousand Exabytes, a billion Terabytes, or a trillion Gigabytes. In simpler words for Shruti Jain: If each Terabyte in a Zettabyte were a kilometer, it would be equivalent to 1,300 round trips to the moon and back (768,800 kilometers)."

In the graph below there is the evolution and prediction of the data created between 2010 and 2025 according to statista.com:

Andrea Marcelli
CIS 2532 NET01
December 10, 2022
Research Big Data




Note that the data is counted in zettabytes!

As we can see from the graph the increase in data volume will be massive therefore there will be a lot of challenges in order to handle such data. Just think about the fact that all this data is counting a very small percentage of the IoT devices that are not connected yet. IoT will change how we refer to data, and how to manage data. We can see these events coming as we have smart TVs connected to wifi networks, cars, fridges, air conditioners, doors, windows, webcams, printers, and much more.

So why is Big Data important? Big Data as we learned also from previous chapters about machine learning and deep learning, we understood that having the capacity to manage, learn, and analyze huge quantities of data is very important for different fields and purposes. Big Data is what gives us predictions for economic events or a corporation's financial status.

Andrea Marcelli
CIS 2532 NET01
December 10, 2022
Research Big Data

We can analyze and identify the various opinions that groups of people whether large or small think and the sentiments of data. We can create software that simplifies tasks that would require us years to complete. Overall, the evolution to Big Data associated with machine and deep learning can improve the efficiency and knowledge of organizations and people in general (whether its politics, freelancing, business, non-profit, etc…).

What are the three types of Big Data: The three main types that compose Big Data are structured data, unstructured data, and semi-structured data.

**Structured data:** This is a type of data that is organized in a well-defined manner, typically in a database or spreadsheet, and can be easily searched and analyzed. Structured data is also called relational data as it works with databases with multiple tables connected with relationships. Structured data needs to be organized, modified, and analyzed using a Structured Query Language (SQL). In our case, we used SQLITE3 with python which gave us the possibility to read, write, and modify the structured data given in a database.

**Unstructured data:** As we said most of the data nowadays uses unstructured data, and examples of it are the common data with which we interact the most such as Facebook, Youtube, Twitter, and much more. Unstructured data is a type of data that has no fixed structure and cannot be easily searched or analyzed using traditional methods. Examples include videos, images, and audio files.

**Semi-structured data:** This is a type of data that has a more flexible structure than structured data such as it does not have relational databases or any delimitations by rows and columns. However, it still contains some defined elements that can be searched and analyzed. Examples include emails and social media posts which through a key-value pair can be differentiated by other data. Due to these reasons,

semi-structured data is managed, modified, and analyzed with NoSQL such as XML, JSON, and YAML. All these No Structured Query Languages are used to exchange semi-structured data across systems that may even have a varied underlying infrastructure.

What are the three V's? According to oracle.com the exact definition of Big Data is: "data that contains greater variety, arriving in increasing volumes and with more velocity. This is also known as the three Vs." In 2001, industry analyst Doug Laney articulated the now-mainstream definition of Big Data as the three V's: volume, velocity, and variety. These three V's capture the key characteristics of Big Data:

**Volume:** The number of data matters. With big data, you'll have to process high volumes of low-density, unstructured data. This can be data of unknown value, such as Twitter data feeds, clickstreams on a web page or a mobile app, or sensor-enabled equipment. For some organizations, this might be tens of terabytes of data. For others, it may be hundreds of petabytes. (oracle.com)

**Velocity:** is the fast rate at which data is received and (perhaps) acted on. Normally, the highest velocity of data streams directly into memory versus being written to disk. Some internet-enabled smart products operate in real-time or near real-time and will require real-time evaluation and action. (oracle.com)

**Variety:** Variety refers to the many types of data that are available. Traditional data types were structured and fit neatly into a relational database. With the rise of big data, data comes in new unstructured data types. Unstructured and semistructured data types, such as text, audio, and video, require additional preprocessing to derive meaning and support metadata. (oracle.com)

Andrea Marcelli
CIS 2532 NET01
December 10, 2022
Research Big Data

**References**

What is Big Data? Oracle. (n.d.). Retrieved December 11, 2022, from https://www.oracle.com/big-data/what-is-big-data/#defined

Thomas Barnett, J. (2016, October 11). The Zettabyte era officially begins (how much is that?). Cisco Blogs. Retrieved December 11, 2022, from https://blogs.cisco.com/sp/the-zettabyte-era-officially-begins-how-much-is-that

Taylor, P. (2022, September 8). *Total Data Volume Worldwide 2010-2025*. Statista. Retrieved December 11, 2022, from https://www.statista.com/statistics/871513/worldwide-data-created/

*Types of big data*. GeeksforGeeks. (2022, February 2). Retrieved December 11, 2022, from https://www.geeksforgeeks.org/types-of-big-data/