Name: Kaiser Mian Cabatingan Date: December 22, 2021

Subject: Integrative Progamming and Technologies

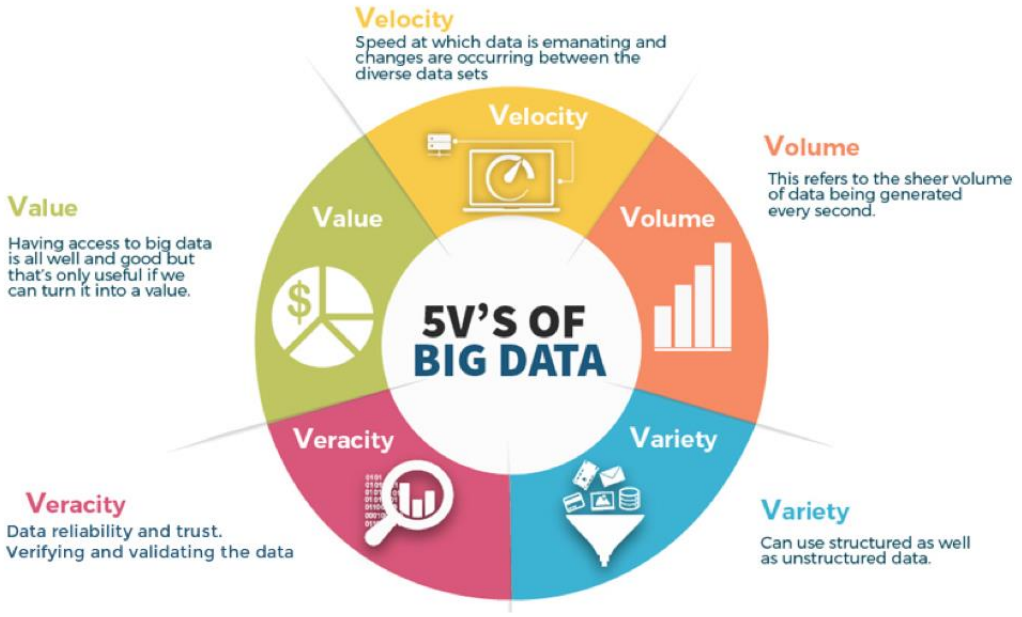
Narrative Report:

Hadoop and HDFS Architecture

Smartphones are widely used all over the world which results in huge volumes of data generated in the form of texts, phone calls, emails, photos, videos, search and music approximately 40 exabytes of data every month and this is only a single smartphone user, now multiply that to 5 billion users, that's a lot for even our mind to process. This huge volume of data is too much for traditional computing systems to handle and this massive amount of data is what is coined as 'Big Data'.

This huge volume of data is too much for traditional computing systems to handle and this massive amount of data is what is coined as ‘Big Data’.

To classify any data as big data, it is possible with the concept of five V’s of Big Data

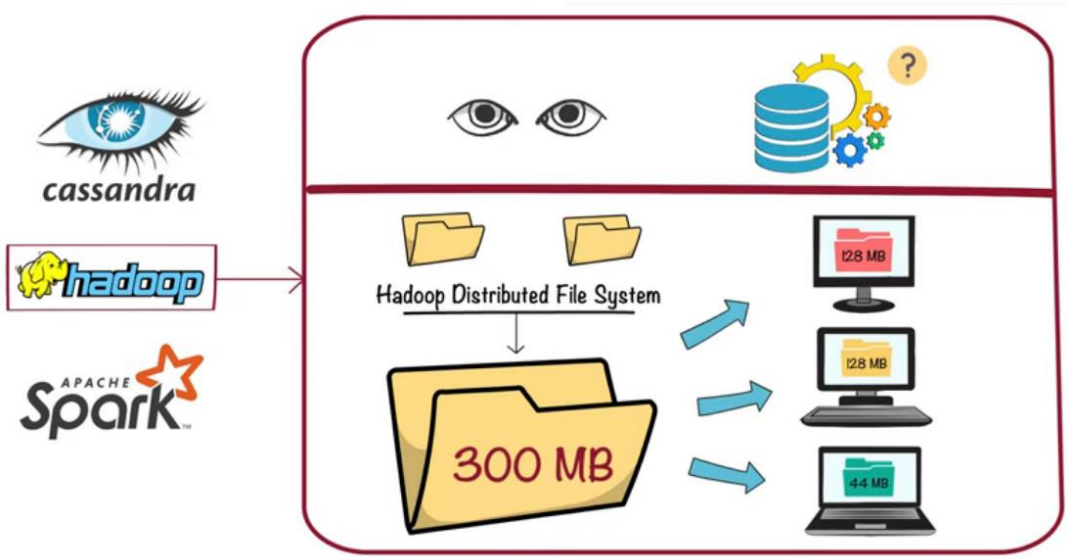


Five V’s of Big Data:

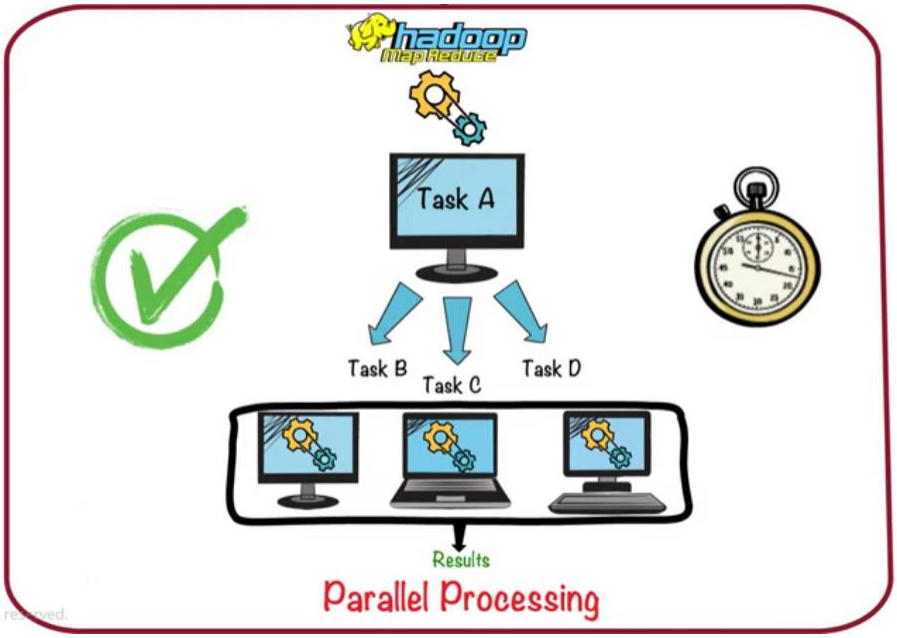
- Velocity - It relates to the rate at which data is generated and sent. This is a crucial consideration for businesses that require their data to flow fast so that they may make the best business decisions possible.
- Volume - refers to the total amount of data available. As the initial size and amount of data acquired, volume is like the foundation of big data. Big data is defined as data with a sufficiently enormous volume.
- Variety - refers to the variety of data kinds available. An organization may get data from a variety of sources, each of which has a different value. Data can come from both inside and outside an organization. The standardization and sharing of all data collected is a difficulty in variety.
- Veracity - relates to the data's quality and accuracy. Data collected could be incomplete, erroneous, or incapable of providing actual, actionable information. Overall, veracity refers to the level of confidence in the data collected.
- Value - refers to the value that big data can provide, and it has to do with what businesses can do with the data they collect. The ability to extract value from big data is a must, as the value of big data rises in direct proportion to the insights that can be gleaned from it.

Analysis of these data benefit the medical sector by enabling faster disease detection, better treatment and reduced costs.

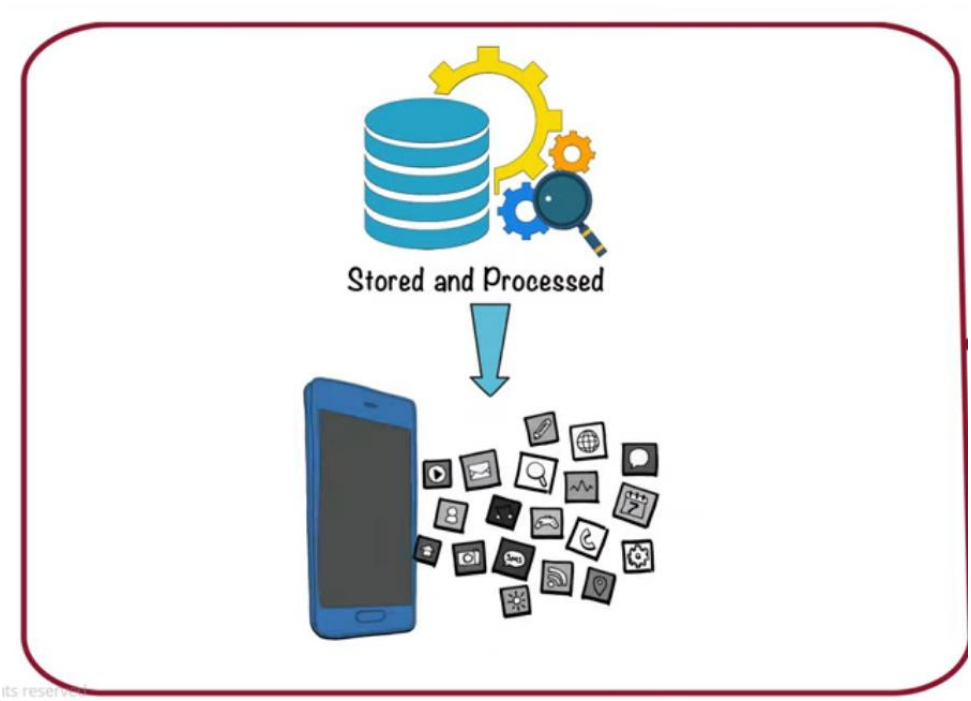
Data in any form can amount to petabytes of data. Storing and processing such large amounts of data is tough, thus specialists utilize frameworks like Hadoop HDFS for storing and executing applications on commodity hardware clusters.



Hadoop uses a distributed file system known as Hadoop distributed file system to store big data if you have a huge file, it will be broken down into smaller chunks and stored in various machines, this benefits considerably in data backup because when the file is broken, multiple copies of it goes into different nodes, this process is called distributed storing and makes sure that even if one machine fails, your data is safe on another.



MapReduce technique is used to process big data, a lengthy task is broken into smaller tasks B, C and D now instead of one machine, three machines take up each task and complete it in a parallel fashion and assemble the results at the end. Processing is easier this way and faster – this is known as parallel processing.



After storing of the data to distributed file systems, analysis can now be done on it for numerous applications such as games, medicine, engineering and much more. In games for example like Halo 3 and Call of Duty, designers analyze user data to understand at which stage most of the users pause, restart or quit playing. This insight can help them rework on the storyline of the game and improve user experience which in turn reduces the churn rate. Similarly, big data also helped with disaster management during Hurricane Sandy in 2012, it was used to gain a better understanding of the storm's effect on the east coast of the US and necessary measures were taken. The model could predict the hurricane's landfall five days in advance which was not possible before. These are some of clear indications of how valuable big data can be once it is accurately processed and analyzed.