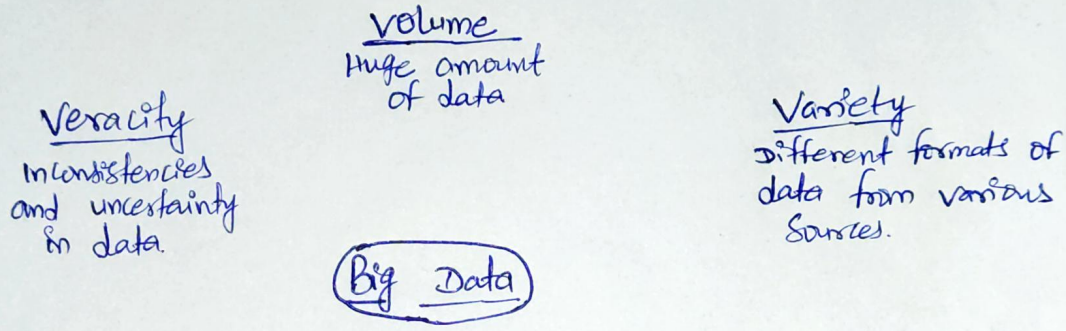
Name- Aniket Kumar
Registration no. - 12108348
Roll no - B62

Q Make a documentation or Presentation on different V's in the field of big data. For all V's the Practical use case should be discussed as example.

Ans Big data, the data that is so large that it cannot be processed using conventional methods, has five V's given by Oscar Herencia, that are following.



1. Volume

- As the name suggests, Big data has to be "big", and Size in this case is measured as volume.
- It refers to the amount of data being collected.
- The real life example can be Netflix, which has over 86 million members globally, streaming over 125 million hours of the content per day.
  This results in a data warehouse which is over 60 Petabytes in size.

- we can also take examples of the emails, twitter messages, Photos, video clips, sensor data, etc. we Produce and share every second.
- on facebook alone, we send 10 billion messages Per day, click the 'like' button 4.5 billion times and upload 350 million new Pictures each and every day.
- If we take all the data generated in the world between the beginning of the time and 2008, the same amount of data will soon be generated every minute!
- This increasing makes data sets too large to store and analyse using traditional database technology.
- with big data technology, we can now store and use these data sets with the help of distributed systems, where Parts of data is stored in different locations and brought together by software.
- If we see big data as a Pyramid, volume is the base.

2. Variety
- It refers to the different formats of data from various sources.
- In the Past, we focused on structured data that neatly fits into tables or relational databases, such as financial data (e.g. Sales by Product or region).
- In fact, 80% of the world's data is now unstructured, and therefore can't easily be Put into tables. for example,

Photos, videos, or social media updates.

- This 'V' describes, the huge diversity of data types.
- The data could be both structured data such as first name or email, or unstructured data, such as a Product review.
- The data must be Processed in order to analyze it.
- for a Product review, this could be Performing a Sentiment analysis to determine whether the review is Positive or negative.
- These data can have many layers, with different values.
- According to Muñoz, "when launching an email marketing campaign, we don't just want to know how many People opened the email, but more importantly, what these Peoples are like.
- with big data technology, we can harness different types of data (structured and unstructured) including messages, social media conversations, Photos, Sensor data, video or voice recordings and bring them together with more traditional, structured data.

- we take the real life example of medicine sector.
- Electronic health records and medical devices collect a different kind of data, which in turn might be interpreted differently by different physicians - or made available to a specialist but not a Primary care Provider.

- with increasing adoption of Population health and big data analytics, we are seeing greater variety of data by combining traditional clinical and administrative data with unstructured notes, socioeconomic data, and even social media data.

3. Value

- value is the worth of data being collected.
- Some Big Data that a business stores may have little or no value in decision making or improving operations, whereas some big data has a larger impact.
- If the data has no value now or in the near future, it may be best to simply stop collecting it.
- Data that has no value can often serve as distraction and only hinder the data analysis process.
- If we are going to invest in the infrastructure required to collect and interpret data on a system-wide scale, it's important to ensure that the insights that are generated are based on accurate data and lead to measurable improvements at the end of the day.
- It is well and good having access to big data, but unless we can turn it into value, it is useless.
- we can say that the 'value' is the most important V of Big Data.
- It is important that the businesses make a business case for any attempt to collect and leverage big Data.

- If we are going to invest in the infrastructure required to collect and interpret data on a system-wide scale, it's important to ensure that the insights that are generated are based on accurate data and lead to measurable improvements at the end of the day.
- For example, every clinician and healthcare system is different, and to provide high quality patient care, organizations have to put the analyzed ~~that~~ data in such a way that it is worth it.

4. velocity

- velocity is the speed at which the Big Data is collected.
- This speed tends to increase every year as network technology and hardware become more powerfull and allow business to capture more data points simultaneously.
- Example - Google recieves over 63,000 searches per second on any given day.
- In addition to managing data, Companies need that information to flow quickly - as close to real-time as possible.
- According to the metlife executive, "velocity can be more important than volume because it can give us a bigger competitive advantage. Sometimes it is better to have limited data in real time than lots of data at a low speed."

- The data have to be available at the right time to make appropriate business decisions.
- Just think of social media messages going viral in seconds, the speed at which credit card transactions are checked for fraudulent activities, or the milliseconds it takes trading systems to analyze social media networks to pick up signals that trigger decisions to buy or sale shares.
- Big data technology allows us to analyze the data while it is being generated, without ever putting it into databases.
- A real life example can be taken of the champions league soccer matches.
On the days of the matches, the food delivery company La Nevera Roja decides whether to buy a Google Adwords campaign based on its sales data 45 minutes after the start of the game. Three hours later, this information is not nearly as important.

5. Veracity
- veracity is the quality or trustworthiness of the data.
- we have all the data but we could be missing something, data might not be clean and accurate, data might not offer anything useful.
- If we are not confident enough that the resulting analyze can be trusted, there is little point to collecting big data.

- for example, if we are piping all order data in but also including fraudulent or cancelled orders, we can not trust the analysis of the e-commerce conversion rate because it will be artificially inflated.
- with many forms of big data, quality and accuracy are less controllable.
- for example, twitter posts with hash tags, abbreviations, typos and colloquial speech as well as the reliability and accuracy of content.
- The volumes often make up for the lack of quality or accuracy.
- The real life use case can be taken of the medicine sector.
  What a clinician reads in the medical literature, where they trained, or the professional opinion of a colleague down the hall, or how a patient expressed herself during her initial exam, all may play a role in what happens next.
  Such variability, means data can only be meaningfully interpreted when care setting and delivery process is taken into context.
  for example, a diagnosis of "CP" may mean chest pain when entered by a cardiologist or primary care physician but may mean "cerebral palsy" when entered by a neurologist or pediatrician.

Thus, these were the five 'V's of the big data, that plays a very important role.