_____

Understanding the Concept of BIG Data Testing, Strategies and Methods

Robby Rahim
Lecturer at Sekolah Tinggi Ilmu Manajemen Sukma, Medan, Indonesia
robbirahim@ieee.org

Abstract: - This paper explains the concept of BIG data and the process of testing BIG data technology. There are various procedures which are used to perform testing of BIG data technology. It is to make sure that all the functions of BIG data work smoothly without getting errors. This paper will explain the concept and strategies used to perform the BIG data testing.

Introduction: - BIG data is the concept which is used to handle large volume of data in advance manner which cannot be retrieved or extracted using simple data extracting techniques. It is the process of creating, saving analysing and extracting the data which is present in high volumes and is complex form of data. BIG data provides easy methods of processing data from a centralised database which is present in its complex form which cannot be easily extracted using the old methods of data extraction. The amount of data present in BIG data database is not limited but it increases from time to time. It has the capacity to handle extremely large amount of data.

Figure 1. BIG Data Concept.

In most of the company's BIG data is used to handle structured data, unstructured data, complex data, simple data all at once place. This usually includes sets of data which have very high volume of information which simultaneously increases over the period of time. BIG data makes is easy to access variety of data from this kind of database which in turn, saves a lot of time which can be utilised to do other productive work. The company gathers big data information from all sorts of platforms whether it is online or offline.

Characteristics of BIG Data: - There are mainly nine major characteristics of BIG data. Each characteristics defines its properties in a specific way which is different from one another and has specific way of handling the data. The main 9 characteristics of BIG data can be seen in the following diagram: -

Figure2. V's characteristics of BIG Data.

Each characteristic of BIG Data is studied in detail in following pointers: - [1]

a. Volume: - The volume of data used in any business depends upon various factors. If the business is small then they might not have large volume of data so the concept of BIG data may not be implemented in such kind of organisations. On the other hand, big business like those who handles social media may have very big amount of data where they need such kind of system where all their data can be collected and uploaded simultaneously whenever it is created. For such organisations the concept of BIG data comes into the picture. Therefore, it is very essential for the business to first decide and have a look about

_____

the volume of data being used by them. This is the first and the basic step for any organisation who are interested to use BIG data concept. Once the quantity of the data which need to be processed and analysed is identified. Now-a-days the measurements of data are not just in gigabytes or terabytes but more than that. For companies like Facebook etc the amount of data being uploaded every minute is larger than traditional measuring methods. Hence, in order to process such large- scale data new technologies like BIG data is used which have better efficiency and accuracy of handling the data which in turn saves time and energy of the data analysts.

b. Velocity: - The speed at which the data is produced is measured by the velocity. For each organisation the speed at which new amount of data is created is identified by the velocity of the data. This also includes the speed of data at the time on which it is created till the time it is processed, uploaded and analysed. So, before knowing the velocity of the data, it is very important to know the volume of data which a company produces or handles. For some organisations, every minute data is being created which can be larger than the other companies. Therefore, it is very important that the business should have technology like BIG data which is very useful and accurate to identify the data, process it and analyse it with a speed faster than the speed of it being created.

c. Validity: - This means to check whether the data being generated is accurate or in other words if the data being generated is matching the intensions for which it was being created. In traditional methods, most of the time being spent by data analyst is to identify the validity of the data than actual analysation of the data being produced. Hence, for handling large amount of data and to identify the accuracy of the data at the faster speed, BIG data comes into the picture.

d. Variety: - This is the main characteristics of BIG data. The type of data depends on the type of business an organisation does. Different organisations will have different varieties of data which will differ from each other. Also, in old days traditional methods of procuring data were used. Everybody uses to have traditional methods of saving data in spreadsheets or excel sheets. But now-a-days, data can be procured from variety of

platforms like online, e-mails, photos. So, there can be vast variety of data which can be handled by using BIG data technology.

e. Visualisation: - As the volume and variety if data is vast and huge, it is necessary that it can be presented in such a way that it can be understood easily without spending much time in it. So, the art of presenting the data with visual effects in such a way is defined by visualisation characteristics of the BIG data. The visual representation of the large amount of data is much easier to analyse as compared to big technical terms and numerical. The person working on the BIG data and data analysts must have sound technical knowledge so that they can make easy graphical representations of the data being analysed in such a manner that it can be used at any point of time.

f. Veracity: - This is one of the main characteristics of BIG data. One may face challenges while handling data veracity. This is the field where one may face challenges. So far it is clear that BIG data has the capacity of handling vast variety and volume of data every second. At the same time, it is very important to know whether the data collected is accurate and valid or not. Here the concept of data veracity comes into the picture. Hence, data veracity means to analyse whether the data being received is accurate or not, where it is valid for the particular organisation or not. Data veracity means to check whether the data is not being duplicated so as to save space.

g. Volatility: - [2] In traditional methods of storing data, the volume and amount of data use to be stored for indefinite period of time. But in today's era of BIG data it is important that the volatility of data should be analysed. Hence, it is very important for the data analysts to decide that a particular data can be retrieved for a certain period of time and not after that.

h. Vulnerability: - As the vast variety and volume of data is available easily at one place so it makes it quite vulnerable which raises the security concerns of the data. So, it is very crucial for the organisations to make sure that the data is secure and cannot be easily accessed without proper authorisation. The security team needs to be more responsible under this scenario.

2

_____

i. Value: -This characteristic is also very important as it is very important for the business to value their client and understand their requirements in detail and make sure that their demand is met. If the business does not value their clients, then nobody will like to invest their valuable time and projects in such businesses.

BIG Data Testing: - BIG data has vast and large volume of data which can be accessed easily. The data procured can be structured, unstructured etc. Once the implementation of BIG data is done, it is necessary to check whether all the requirements are met and all its functions are working according to the specification of the application. Hence, the process of making sure that the functions of each application is working as it is supposed to be, then the process is known as BIG data testing. Like any other testing cycle, BIG data testing also follows procedure. There are different phases of BIG data testing where each stage should first be completed before proceeding to the other stage. Following is the BIG data testing cycle: -[3]

A. Real Time Testing: - The purpose of this type of testing is to test when the application is in use which means it is to test whether the application is functioning properly as per its specification. This step is performed to check whether the application is giving desired output when given proper inputs. If there are any bugs then it should be rectified before going to next stage.

B. Batch Testing: - In this step the testing of the application is done using small batches of data. In this step the tester will keep giving various inputs in different batches to check whether the application is working properly.

C. Interactive Testing: - In this step the interactive nature of the application is tested. The tester will check if different modules of the application is interacting with each other in the manner it is supposed to interact. If any bugs identified then they should be rectified.

the capacity to handle all types of data whether its is structured, unstructured, etc. Like any other testing process the BIG data testing has also its own testing cycle which should be performed to make sure that the application works as it is supposed to.

References: - 1. https://contenteratechspace.com/blogs/8-characteristics-of-big-data/
2.https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx
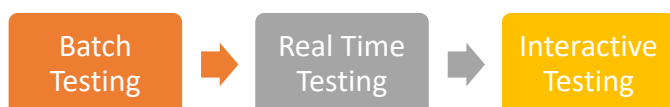3.https://www.edureka.co/blog/big-data-testing/

Figure3 BIG Data Testing

Conclusion: - This paper has explained that BIG data is a new concept which is commonly used in big organisations. It has