Correlation:

In today’s world, almost every organisation has a data warehouse either big or small as per there requirements. Bigger organisations may even have multiple data warehouses. The main purpose of any data warehouse is to unify & combine the data from operational systems in such a manner that the business may get the maximum benefit from this. Data usually inserts into data warehouses from several sources and in diverse formats. Data warehouses are used to collect and combine diverse data and to provide information to support business and reporting intelligence. For this reason, this is very crucial that the quality of data is good in data warehouses. Data quality means that entire data satisfies the requirements & needs of the usage.

**Data is getting increased at a rapid pace and the rate of information growth is very high**. Data generation occurs through many users, businesses and industries. It is very crucial to integrate this data that have been generated via different sources. If it gets wasted, lots of valuable information can be lost.

Data, in today’s technology and business world, is vital. Big Data technologies and initiatives are growing to analyse this data for gaining insights which helps in making planned & strategic decisions. This concept evolved at the beginning of 21st century, and now every big organisation is using Big Data technologies to manage their data. It refers to massive and huge data sets that can be structured or unstructured. This huge amount of data is produced every day by businesses & users. Big Data analytics is the process of examining the large data sets to underline insights and patterns. The Big Data analytics is indeed a revolution in the field of Information Technology.

Previously, skilled analysts were required for processing the data but these days there are tools used for high-speed data and this helps in incorporating the data analytics at the time of making decisions. Social media has enhanced data growth in the organization and thus changes can be made based on that. Now all the decisions made upon data can be very crucial and everything depends on quality of data.

Data quality is important to organizations for many reasons. Bad quality data can harm operation and performance of the organization's daily activities. In worst case scenarios it can disturb making of important decisions or prevent to see the true status of the organization. Data quality is also meaningful cost maker for organizations. Costs are coming from efforts to correct data, lost customers, and wrong decisions.

Data in itself is nowadays very valuable asset for organizations, and the volume is increasing all the time. It is very essential to have correct data to provide correct information. Data is the most important part of all Data Analytics, Machine Learning & Artificial Intelligence. Without data, we can’t train any model and all modern research and automation will go vain. Big Enterprises are spending loads of money just to gather as much certain data as possible.

So, in nutshell it is very important to have such kind of mechanism which can identify wrong data automatically. In current world, most of the organisations rely on manual testing of data warehouses. Manual testing is cumbersome and error prone process of identifying incorrect information which also includes huge costs. Organisations are dependent on manual testing of data warehouses as there is no automatic process available which can help in identifying wrong data in warehouse within time and without any extra cost. Automation testing of Big Data is the need of time, which can be used to extract wrong or invalid data from warehouses automatically within very less time, and can results in saving huge cost.