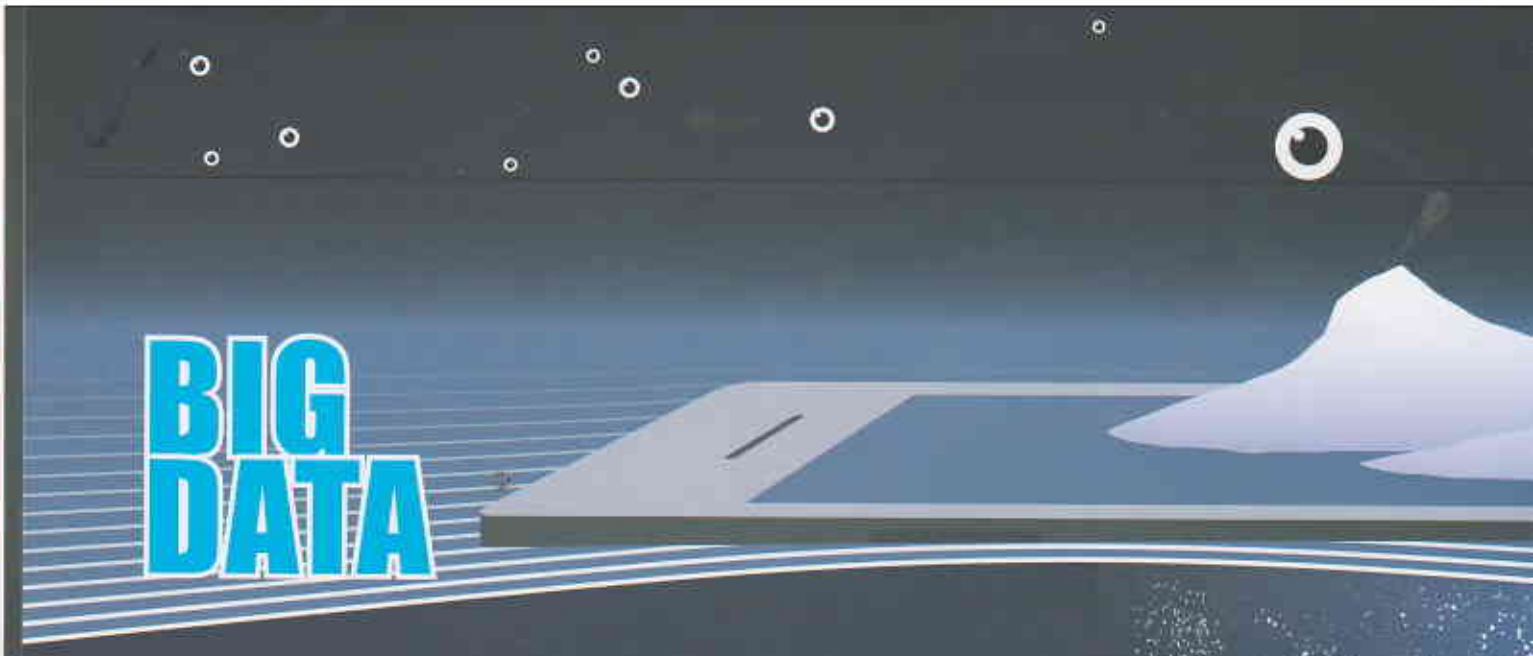


# BIG DATA

Big data is one of the widely (mis)used word in recent times. What is it? In simple English, big data is lot of data. Here it refers to lot of data on a specific process. Data was being collected for analysis and understanding from time immemorial. Due to rise in volume of processes, collection/storage requirements also were growing. Human memory and its analytical ability are limited. Data collections hitherto were limited by various constraints like modes of collection, compilation, storage etc.

With the improvement in ICT(Information and Communication Technologies), it is becoming easier to collect and compile data. Advances in statistics like factor analysis, cluster analysis etc are paving way for analyzing data and find out hitherto unknown relations between various parameters.

Thus big data is broadly compilation of lot of data on any item/process and analyzing it to get useful inferences which can be leveraged for business decisions.

*"In God we trust. All others must bring data."* – W. Edwards Deming

With the advances data analysis techniques, the utility/value of data was being realized. Statistical quality control(SQC) started taking roots in 1920s. Deming popularized use of SQC and helped catapult Japan to become Quality driven economy.(remember made in Japan of 70s and 80s). SQC is simply to analyze process conditions and results with a view to continuously improve the process. From then on, various new tools were developed to monitor and control processes using data.

*"Data really powers everything that we do."* – Jeff Weiner

*"Information is the oil of the 21st century, and analytics is the combustion engine."* – Peter Sondergaard

In last decade and half, two things led to spectacular rise of utilization of big data. They are continuous increase of processing power of computer and continuous reduction in storage costs. It initially started with retail customer data analysis to find hidden trends between different groups of items/events which could then be leveraged to increase sales. This was further supported by multifold increase in online transactions. People are doing many more transactions online leading to electronic trail which could be leveraged by businesses for upselling and cross selling. Hence the saying data is new oil.

Businesses have software systems to collect and store transaction data. Transaction data is stored in huge data bases which are classified and segregated process wise. By design each process has a separate data base and the debases of two processes do not communicate with each other by design. Thus someone needs to put in effort to compile cross functional data to find relationships between variables across data bases. That issue was overcome by introduction of business intelligence(BI) systems. These are big computer data bases where data from disparate data sources in an organization can be combined to get new insights. This data on BI platform works as an ideal base for applying data analytics.

*"Torture the data, and it will confess to anything."* – Ronald Coase

*"You can use all the quantitative data you can get, but you still have to distrust it and use your own intelligence and judgment."* – Alvin Toffler

Due to popularity of the word big data, many started using and dropping the word in conversations. However, all data is not big data. Big data is characterized by 5 Vs(Volume, Velocity, Veracity, Variety and Value). Unless tested processes are followed in collation, compilation and analysis of data, the inferences may not be useful and sometimes can be erroneous. Hence...

*"Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom."* – Clifford Stoll

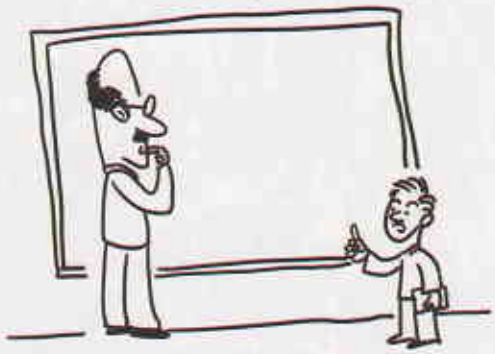
*"If we have data, let's look at data. If all we have are opinions, let's go with mine."* – Jim Barksdale

Over a period, businessmen realized that data can be leveraged profitably. For this, first data sources are to be identified. In our organization, there are two major processes

which are continuously monitored. Manufacturing and marketing. In refining, process parameters are continuously monitored and recorded for business decisions. This data can further be leveraged to identify process specific and parameter specific relations which can be used profitably. Similarly, marketing has huge data bases like inventory and sales data. It is to be understood that unless available data is brought on same platform, analysis is not possible. If these data bases are brought on to same plat form, cross process inference can be obtained using statistical techniques. This inferences can then be profitably leveraged.

*"Getting information off the Internet is like taking a drink from a firehose."* – Mitchell Kapor.

**Big Data: "Hello guys, meet my friend No Privacy."**



**"I will write my name on the board only if you tell me how you plan to use that data."**

Electronic trails are both useful and troublesome at the same time. They are useful to businesses which try to understand customer behavior as explained above. They can be analyzed by evil minds to harm individuals. When ever one logs on internet, the trail may be recorded and it is not easy to remove it. Also many are lured by free apps/free ads on net. Many of these apps may not have been tested thoroughly and may not have good data privacy systems in place. Hence one needs to be cautious while using internet.

In summary, big data, adds context to data, to understand situation in new light and take necessary action. Big data is a big opportunity for business today and is here to stay. Businesses can use it as a force multiplier in their processes to improve process efficiencies and thereby profits.

Chavali Srinivasa Rao
DGM – IS, HPCL ERP Centre