

Challenges of Big Data

What is Big Data ? Starting from scientists struggling to analyze massive amounts of genetic data, as they did in the Human Genome Project a decade ago, or astronomical data, such as the survey of more than 930,000 galaxies undertaken by the Sloan Digital Sky Survey, Big Data has blossomed into a constellation of computer science approaches to handling, visualizing and blending together “big” sets of data.

However following are the challenges for Big Data

1. **Data storage implications** – People are creating staggering amounts of data just through the course of their daily lives, and the volume is growing even larger from data flowing from “machines and implanted sensors in oceans, in the soil, in pallets of products, in gambling casino chips, in pet collars, and countless other places.” Starting from switching on your cable TV in the morning based on which genre of channel you like, to making a phone call (GPRS location, talk time and calling profile), to filling up gas with your credit card (trending of gas usage based on days of the week and the time of the day) to browsing the net to either google or tweet is generating terabytes or zeta bytes of data. Imagining the world getting smaller and smaller every day and large multinational expanding their presence rapidly across the world expanding their customers in geometric progression, storage of data is a big challenge.
2. **Technology** – The right technologies and methods to store, map and process such huge streams of data are still evolving. There are number of tools like Hadoop for mapping global data and storing data into localized nodes. Languages (like Hive) are also developed or reused (like SQL) to query and fetch the relevant data from file systems. And there are statistical tools like ‘R’ which are used to analyze such data sets. However none of these are matured systems and without certain constraints. It will take some time for the technologies to be seamless and acquiring the capability to do the needful end to end for Big Data.
3. **Cost vs return** – Industry experts are still grappling with the cost benefit analysis of Big Data. Certain practitioners believe that social networking data as was thought to be useful in driving creative marketing and hence impacting the top line revenue has not been that successful in past. However optimists are of the opinion that though the intent was there in the past there was a lack of tools and technology to achieve that. With the IT giants lie IBM and others in the field now, Big Data certainly is very promising and will be bigger wave than either Y2K or even the IT boom that the world has witnessed in the last one and half decade.
4. **Privacy** – The biggest worry is the privacy concern. As an article from CNN noted, “it has become cheap and easy to pry into the lives of others at the same time that protecting our own lives has become time-consuming and expensive.” Andrew Rowland, global head of database engineering at UBS, said there is a privacy and a relevancy element surrounding data: “You might have bought ladies underwear

online once, but do you want to be presented with it every day? Do people want every bit of data assessed for them?"

5. **Sampling** – Big data contradicts the basis of statistical sampling. This is a dichotomy and challenges the premise of age old concept of taking a fraction of homogenous population or stratified portion of the population to predict and infer characteristics and behavior of the population. Big data takes into count the entire data set and analyzes it. Unless the ROI is disproportionately high, Big data will face challenges from the businesses in near future.
6. **Data accuracy** – The popular cliché 'garbage in, garbage out' is valid here also. Unless the data is generated directly by the end user and fed directly into data storage systems there is a huge chance that the data is not correct. And if improper data is used for analysis it can create huge lost opportunity and resource not to mention of any detrimental impact on marketing if any.

Business users need to properly understand and make sense of the data their business produces. The big data paradigm must be made useful to business; rather than being just a corporate conversation turning into another FAD.