

# Big Data Getting Bigger: Catering Challenges

Daksh Mehta

**Abstract:** *The amount of data in world is increasing rapidly. Data is developing as a direct result of utilization of web and increased usage of mobile phones. Big data is an assortment of data sets which is extremely enormous in size. Presently, Big data is one of the most talked theme in IT industry. It will produce significant jobs in future . Big data changes how data is seen and utilized. A portion of the applications are in areas such as healthcare, traffic management, banking, retail, education, etc. organizations are turning out to be more adaptable and more open. As the big data is getting bigger various challenges are arriving in front of it. One of the major challenges in front of big data are information overload . I have discussed an efficient way called DOD (Data on Demand) which helps filter data noise.*

**Keywords:** Big data, Information overload, Data on Demand

## 1. Introduction

Data has been an integral part of our life since ages. We are surrounded by data, from the language we speak to the things we see, everything is data. Technology is growing rapidly in today's world and so is the population. This large number of population produces massive amount of data i.e. more than 2.5 million terabyte per day. In earlier times it was a challenge to store these data however now we have come up with an efficient solution of this problem, i.e. "big data". Applications of "Big data" can be seen all around us in everyday life, from GPS to social media to tech gadgets and tech wearables. However in the process of storing big data a lot of garbage data is collected. In order to save time, efforts and storage space, I have developed a model in big data called D.O.D i.e. Data On Demand.

### WHAT IS BIG DATA?

Big data is a mix of organized, semistructured and unstructured data gathered by associations that can be dug for data and utilized in AI ventures, prescient demonstrating and other investigation applications. Frameworks that procedure and store big data have become a typical segment of data the board structures in associations. Big data is frequently seen to have 3 main characteristics: the enormous volume of data in numerous situations, the wide assortment of data types put away in big data frameworks and the speed at which the data is produced, gathered and handled. These qualities were first recognized by Doug Laney, at that point an examiner at Meta Group Inc., in 2001; After it gained Meta Group in 2005, Gartner further advocated them. All the more as of late, a few different characteristics of big data have been looked into, including veracity, worth and fluctuation.

Albeit big data doesn't compare to a particular volume of data, big data organizations frequently include terabytes (TB), petabytes (PB) and even exabytes (EB) of data caught after some time.

### Big data history

The term Big Data was begat by Roger Mougaldas in 2005. Notwithstanding, the use of big data and the mission to comprehend the accessible data is something that has been in presence for quite a while. Actually, the absolute most ancient records of the utilization of data to dissect and control business exercises date as far back as 7,000 years.

This was in the bookkeeping of in Mesopotamia for the account of harvest development and crowding. The standards kept on developing and improve and John Graunt in 1663 recorded and examined data on the pace of mortality in London. John Graunt did this with an end goal to bring issues to light on the impacts of the bubonic plague that was progressing at that point. In his book 'Regular and Political Observations Made upon the Bills of Mortality', John Graunt gave the world the principal factual investigation of data at any point recorded. The book took its readers into the reasons for death in seventeenth century England. Because of his work, John Graunt is generally viewed as the pioneer of the field of measurements. After crafted by Graunt, bookkeeping standards kept on improving and growing yet there were no innovations in this line up to the point in the twentieth Century when the data period started. The beginning stage of present day data starts in 1889 when registering framework was concocted by Herman Hollerith trying to sort out enumeration data. After Herman Hollerith's info, the following vital data advancement jump occurred in 1937 under Franklin D. Roosevelt's presidential organization in the United States. After the United States congress passed the Social Security Act, the administration was required to monitor a great many Americans.

The legislature contracted IBM to build up a punch card-perusing framework that would be applied in this broad data venture. Be that as it may, the absolute first data-handling machine was named 'Monster' and was created by the British so as to translate Nazi codes in World War II, 1943. This machine worked via scanning for any examples that would show up routinely in the blocked messages. The machine worked at a record pace of 5,000 characters for each second, which diminished work that would take a long time to only a couple of hours. From this turn of events, the National Security Agency (NSA) was made in the United States in 1952. Workers of the NSA were entrusted with unscrambling the messages caught over the span of the Cold War. Machine improvement at this stage had progressed to a level where machines could autonomously and consequently gather and procedure data.

The principal data focus was worked by the United States government in 1965 to store a large number of expense forms and fingerprints. This was accomplished by moving each record onto magnetic tapes that should have been put away deliberately in a focal area. This undertaking,

nonetheless, didn't continue because of dread of treachery or securing. In any case, it is a broadly acknowledged that this activity was the beginning stage of electronic big stockpiling. Tim Berners-Lee a British PC researcher developed the World Wide Web in 1989. Berners-Lee's goal was to empower the sharing of data through a hypertext framework. He had no clue about what sort of effect his innovation would have on the world. As we entered the 1990's, the making of data developed at a very high rate as more gadgets picked up ability to get to the web. The main su- per-PC was worked in 1995. This PC had the ability to deal with work that would take a solitary individual a large number of years very quickly.

Thus came the 21st Century this is the point at which the world was first acquainted with the term Big Data by Roger Mougals. Around the same time (2005), Yahoo made the now open source Hadoop with the goal of ordering the whole World Wide Web. Today, Hadoop is utilized by a great many organizations to experience titanic measures of data. During this period, interpersonal organizations were quickly expanding and a lot of data were being made regularly. Organizations and governments the same started to set up big data ventures. For instance, in 2009 in the biggest biometric database at any point made, the Indian government put away unique mark and iris outputs of the entirety of its residents. Eric Schmidt gave a discourse at the Techonomy meeting in Lake Tahoe, California in 2010. In his discourse he introduced that there were 5 exabytes of data put away since the very beginning up to the year 2003. Eric Schmidt couldn't envision that continuously 2016, a similar measure of data would regularly be made at regular intervals. The rate at which big data is developing doesn't appear to back off either.

Over the past number of years, there have been different associations that have come up trying to manage big data, for instance, HCL. These associations' business is supporting different organizations to see big data. Regular an ever increasing number of organizations are moving towards the acknowledgment and abuse of big data.

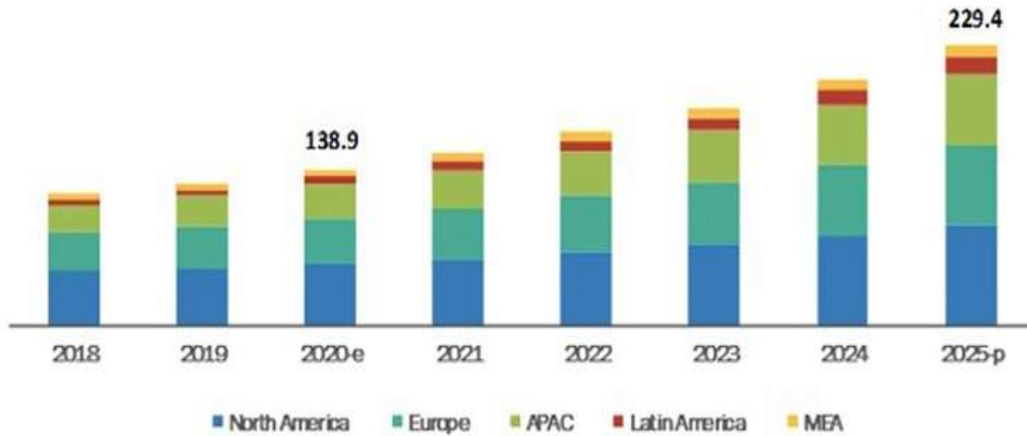
Despite the fact that it appears as though big data has been around for quite a while now and that we are drawing nearer to the apex, big data may simply be at its imposing stages. Big data soon may wind up making big data presently appear to be a poultry sum.

### **Big Data Future**

As far as future of big data is concerned it is for sure that data volumes will proceed to develop and the prime purpose behind that would be the uncommon addition in the quantity of hand held gadgets and web associated gadgets, which is relied upon to develop in an exponential request. SQL will remain as the standard for data analysis and Spark, which is rising, will rise as the complimentary instrument for data examination. Tools for analysis without the presence of an analyst are set to take over, with Microsoft and Salesforce both as of late declaring highlights letting non-coders to make applications for survey business data. According to IDC half of all business investigation programming will incorporate insight where it is required by 2020. At the end of the day it very well may be said that prescriptive examination will be incorporated with business programming. Projects like Kafka and Spark will empower clients to settle on choices continuously. AI will have a far bigger task to carry out for data planning and prescient examination in organizations in the coming days. Protection and security moves identified with big data will develop and by 2018, half of business ethics violations will be related to data.

Chief Data Officer will be a typical sight in organizations in the ongoing future however it is believed that it won't keep going long. Self-governing operators and things like robots, self- ruling vehicles, virtual individual right hand and keen gadgets will be a colossal pattern later on. Big data ability mash as is seen nowadays will decrease in the coming days. The International Institute for Analytics predicts that organizations will utilize selecting and inward preparing to maturing data researchers to complete their own issues. Organizations will before long have the option to purchase calculations instead of program them without anyone else and add their own data to it. Existing administrations like Algorithmia, DataXu, and Kaggle will develop in an enormous scope, that is calculation markets will rise. More organizations will attempt to get their income from their data. The hole in knowledge and activity in big data will decrease and more vitality will be given to acquiring bits of knowledge and execution as opposed to gathering big data. Quick and noteworthy data will supplant big data. Organizations are required to pose the correct inquiries and utilize the data they have, a great part of the big data they have are unused nowadays.

Big Data Market, By Region (USD Billion)

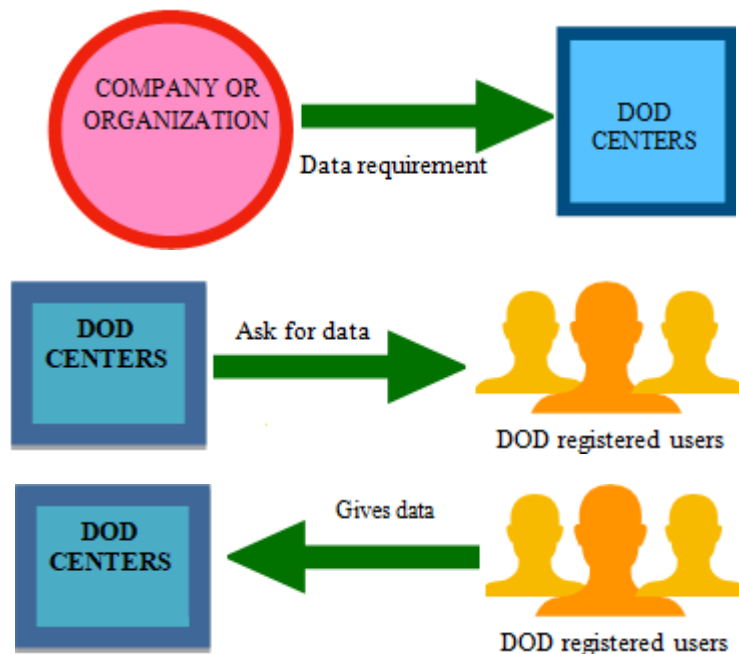


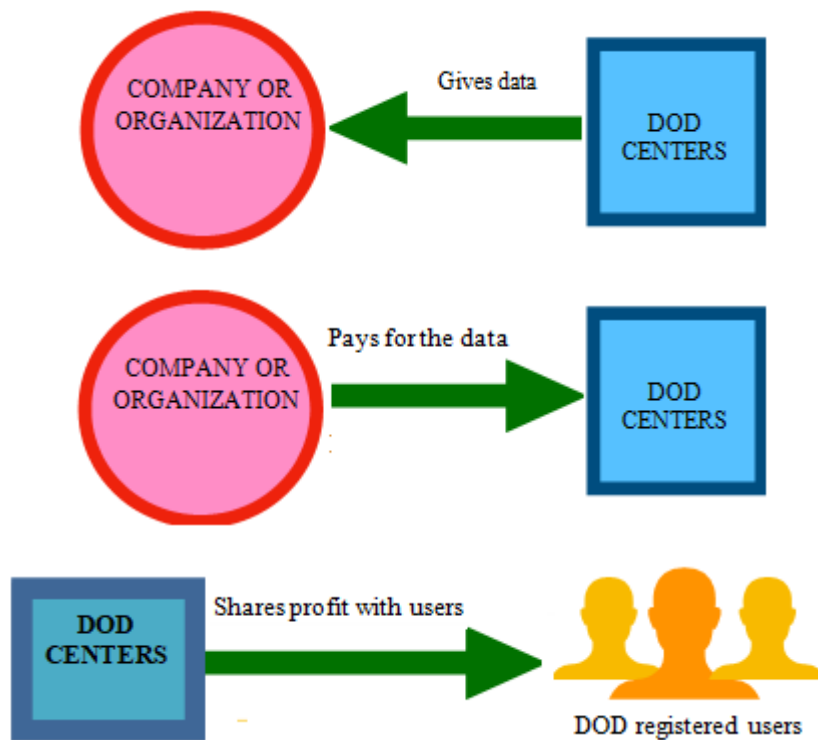
Source: MarketsandMarkets Analysis

## 2. DOD

DOD refers to Data On Demand. One of the major challenges faced in big data is Information overload. As data is continually getting through the servers, organizations and businesses can experience difficulties of finding helpful data that fulfills the company's needs. As mentioned by Tole, "The quantity of data being stored is not always one hundred percent useful" (Tole, 2013, p.33). It's not simply the volume of the data that is the issue, it's likewise the assortment. Big Data can come in different forms ranging from texts to visuals and distinguish- ing them while still trying to apply them to real life situations is a complicated process. The kind of data that gets through the big data workers shifts and before breaking down the valuable data, IT specialist should initially sort out through them (Tole, 2013). I have come up with an efficient way to deal with it. That is Data On Demand, DOD is a way in which there will

be DOD centers where people can create their accounts. While creating their account they would have to provide their identity proofs and one person could only register to one data center. If a company would require any specific data or want some specific information to study market they would contact the DOD center, the DOD center would charge some money from the company. The DOD center would then circulate the information required by the company or organization to its linked people. The linked people will willingly provide the information to the DOD center, which will further give information to the company. In exchange of the information the DOD center would give a percentage of money given by the company or organization for information among the linked people who provided the required information. In this way company would be able to get noise free data and users providing it would have some benefit.





## References

- [1] <https://searchdatamanagement.techtargget.com/definition/big-data>
- [2] <https://www.cleverism.com/brief-history-big-data/>
- [3] Bains ,A.S. , Big Data: The FutureAwaits
- [4] Tole, A. (2013). Big data challenges. Database Systems Journal, (3),31-40.
- [5] <http://www.dataversity.net/common-big-data-management-issues-solutions/> The Most Common Big Data Management Issues (And Their Solutions). By: A.R.Guess. July 152014.
- [6] Patel,B. and Chaudhari,p..(2017,January 01).Future of BigData
- [7] Tarekegn,G.B.and Munaye,Y.Y.(2016,july 04) Big Data: Security Issues, Challenges and Future Scope