

# Big data applications and overview

[Technology](#), [Computer](#)



In the past two decades, extensive progress and generation of data in information technology has led to rise in massive volume of data from different sources such as social networking, online business services, web based applications and mobile devices. The data here is in structured, semi-structured and unstructured format. Since our traditional database systems cannot handle complex unstructured data and the size which it is coming in, Big Data comes into picture. To put in simple words, the volume, velocity, veracity and variety of data is enormous. The reason behind why we are looking at these types of data to process is that it can be used to improve, analyse, develop and relate business solutions through analysis. Big data storage and processing can be achieved through variety of models available in NoSQL databases based on suitable type of data for respective models. Although there are a lot of feasible solutions obtained through data mining in Big Data, issues such as allocation of resources and requirement of storage device arise. Recently, data management systems are dominated by Hadoop based architecture.

<https://www.vormetric.com/data-security-solutions/use-cases/big-data-security>

Online and Offline Big data

Fig. Big Data Model (Goldberg, n. d.)

The data generation possibilities are spread over wide spectrum in Information technology field, it can be classified into two types such as online and offline.

<https://assignbuster.com/big-data-applications-and-overview/>

Online data is a type of data where it is generated continuously through real time systems. For a reference, it could be live video, a banking transaction or stock exchange data. It can be referred as a data which is created, absorbed, processed and transformed in real-time in order to support ongoing applications and online users. As it is flowing in real time data abeyance must be very low and availability of data must be prompt in order to cope up with the expectations of user. (MongoDB, 2016)

Fig. Online Big Data (MongoDB, 2016)

Fig. Offline Big Data (MongoDB, 2016)

Offline data is a type of data where the data is in static form and it can be used in offline environment to analyse but the big data technologies with suitable available tool or technology. Over here the data is not newly created but over the period of time with the help of batch jobs. In this case, latency of data can be high compared with those of online systems and hence these systems can go offline without impacting any of the users or end product. Availability of system can be of low priority, big data technologies can perform complex analysis. Existing examples of offline big data technologies are data warehouse or a storage technology which is used to accommodate bulk data as a static. (MongoDB, 2016)

### Scalability

Although it cannot be purely categorised as failure of the RDBMS systems, it can be addressed as a trait which can be an eventual roadblock for a traditional database trying to scale out in order to handle increasing data

<https://assignbuster.com/big-data-applications-and-overview/>

and performance gains through hardware, storage upgrade. Even if database upgrade is planned it has to go through a time-consuming process while keeping the system offline. A point where upgrading limit of a system reaches to its maximum which is imminent as per the current rate of rising data over the period of time, more flexible systems are needed to store big data in an efficient way. (Allen, 2016)

### Recommendation

Sharding is the method which can be effectively used in RDBMS by dividing data into different tables and treating the tables as lookups. Scaling is not an issue in big data technologies as the databases are created in such a way that they can be expanded with cheap commodity servers. Cassandra, MongoDB, Redis are the common databases used on a high scale.

### Economics & High management

As traditional database systems use proprietary servers in contrast to systems which are divided in form of clusters in big data technologies using low-cost commodity servers, the cost of expansion is much higher than the big data technology which can be replaced with another commodity computer system in case of failure of any one. This allows big data technologies to process and store more data for a much lower price point. (Allen, 2016)

In traditional database systems, management of the database system is highly required and it is carried out by database administrators. Whereas, in big

data technologies things for reference, adding column to table structure, permissions to particular schema are not required. (Allen, 2016)

### Recommendation

Since at this stage of technology and data if we go by the RDBMS systems, we would need to arrange huge data capacity servers and storage in order to cope up with the data. If not, the NoSQL databases can perform complex internal data distribution, auto-correction and very less management is required to maintain the database. Hadoop is dominantly used across big web applications such as Google, Amazon.

### Flexible data model

RDBMS systems are made in such a way where you can have predefined structure for a table and schema. Only data with the respective structure can be dealt while incoming. Whereas in big data technologies it is not mandatory to have data in a particular format as introduced above. (Allen, 2016)

### Recommendation

Since the big data storage bases are categorised by column (Hadoop), document (MongoDB), key-value (Redis), graph (Neo4J) and so on, hence the various data types are accepted across respective open source databases (Allen, 2016)

### T-mobile USA

As the current situation stands in telecommunication industry, data created through each device and region is very dynamic and huge. T-mobile USA has 33 million active users and that is the reason why they chose to put all this big data to its use. The rate at which users were dropping the T-mobile service was brought to half through the big data analysis. Below are few data sources used by them to achieve business objectives.

**Customer Data Zone:** Every user's likes and dislikes are used to understand and provide services based on the available data created by user.

**Product and Service Zone :** Inspection of services availed and products used by each user is taken into consideration in order to maintain the user base satisfaction.

**Business Operation Zone :** All the accounting and billing information stored is used to maintain (Rijmenam, 2015) (Rijmenam, 2015)

Based on big data analysis done on all the above points such as Sentiment, choices and billing data for each user, churn percentage is reduced.

### McLaren Racing Limited

McLaren is a leading formula one racing constructor. Big data scope is recently widened in this sector due to high competition. The sport's utilization of such data is sophisticated to the point that a few groups are trading their insight to different enterprises where investigating gigantic measures of data in a split second can mean the distinction amongst life and death.

Hundreds of sensors fit into the car body while racing export gigabytes of data during race. The data is live streamed to the team which is monitoring the various aspects of the car at same time such as heat exhaustion, engine diagnosis and track activity. The same data is then used to carry out diagnostics, analysis and strategy. Currently system used to compare and reference is SAP HANA.

Due to strict Formula 1 rules there are very few team members allowed to be on the track during race time. Though that doesn't affect the analysis as the big data through sensors is made available with the delay of milliseconds across international locations for respective team from place to place (Muhammadirvan, 2016)

Tesco

One of the largest retailers in the world right now thriving on the offerings provided by big data. In 1995 they introduced their shopping card called as Clubcard for customers. The shopping done through the card is now used to run analysis on customer's shopping behaviour, likeness for product and management of store sections.

For example, data from the shopping carts offers intuitions where merchandise can be best placed near one another or which merchandise should be placed nearer to the checkouts or doorways. Due to this elaborated client insights with the Clubcard, Tesco's understanding with the customer's choices and liking has become more exclusive. This factors ensures them to provide personal suggestions on the beverages or food items based on data gathered from individual shopping cards.

Big data is used on other few aspects such as food wastage, when we talk about the foods and supplies. Tesco receives local weather forecast data and it is linked with the upcoming food items ought to be supplied to the stores. Through the simulations and analysis, right amount of stock is moved to the stores with adequate optimization.

When you are in food industry, food storage comes into consideration.

Expenditure on storage facility is also a big factor that we need consider.

This is compromised through the data generated by the each refrigerator across storage facility.

Tesco analyses refrigerator data to cut short their bills by \$ 25 million per year. As an example, refrigerator sensors in Ireland measured temperature from every 3 seconds and created 70 million data points over the period of one year. (Rijmenam, tesco-big-data-analytics-recipe-success/665, n. d.)

## References

Allen, M. (2016). *Relational Databases Are Not Designed For Scale* .

Retrieved from Marklogic: <http://www.marklogic.com/blog/relational-databases-scale/>

Goldberg, C. (n. d.). *Big Data Security* . Retrieved from Vormetric:

<https://www.vormetric.com/data-security-solutions/use-cases/big-data-security>

MongoDB. (2016). *Online vs offline big data* . Retrieved from Mongoddb:

<https://www.mongodb.com/scale/online-vs-offline-big-data>

<https://assignbuster.com/big-data-applications-and-overview/>



Muhammadirvan. (2016, September 9). [?/09/12/mhmdirfans/](https://muhammadirvan91.wordpress.com/2016/09/12/mhmdirfans/). Retrieved from <https://muhammadirvan91.wordpress.com/2016/09/12/mhmdirfans/>

Rijmenam, M. v. (2015, February 15). *t-mobile-usa-cuts-downs-churn-rate-with-big-data/512* . Retrieved from [https://datafloq.com](https://datafloq.com/read/t-mobile-usa-cuts-downs-churn-rate-with-big-data/512): <https://datafloq.com/read/t-mobile-usa-cuts-downs-churn-rate-with-big-data/512>

Rijmenam, M. v. (n. d.). *tesco-big-data-analytics-recipe-success/665* . Retrieved from [https://datafloq.com](https://datafloq.com/read/tesco-big-data-analytics-recipe-success/665): <https://datafloq.com/read/tesco-big-data-analytics-recipe-success/665>

*Vormetric* . (n. d.). Retrieved from Thales : <https://www.vormetric.com/data-security-solutions/use-cases/big-data-security>