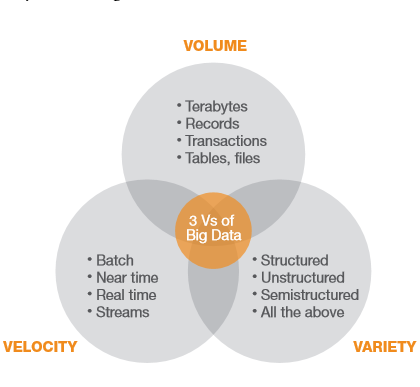
**Bigdata analytics and the role of NoSQL databases**

I. Introduction to Big Data Analytics

Big Data Analytics is a collection of data which is highly being used in business trends by organizations with the help of analytic strategies. The term “Big Data” means a huge collection of data and “Analytics” means analysis of the data systematically.

Firstly, analytics is a process through which a system works on large sets of data and examines it. With the help of these data we can see the changes that had happened in the past and predict what should be done in the future for the benefit of the organization. In terms of business this can better help understand the market trends which further leads to the profit of organization. More and more organization have inclined themselves to using big data as with passing of time.

Secondly, Big Data as the name says it is a collection of huge data. This collection of data is much better as the size of the data increases but there are other important factors contributing to the size of the data. The factors contributing are known as the three Vs of big data which are volume, variety and velocity.

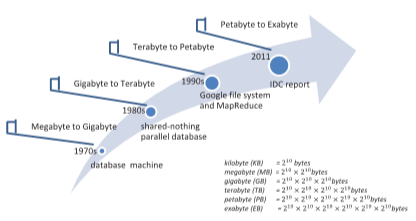
Fig: Vs of Big Data

Among the three Vs the volume is given more priority since the velocity and variety depends upon the volume of data. Volume of data can be defined in many ways such as the storage space it is taking for example, gigabyte, terabytes. Since this way of defining data volume is vague it can also be defined by counting how much quantity of data is collected. The other way of defining the volume is by time. For example, an organization stores the data of past 10 years as the data before that will not be useful to that certain organization.

There are lots of sources through which organization collect big data like websites, smartphones, social media. Every major organization like google, Facebook have their own collection of big data and they also operate through these big data that they collect from the users for purposes like advertisements. Due to the massive collection of data there are variety of data which is one of the factors of three Vs of big data.

Big data is being collected at every second through various methods like clickstream. Whenever you are browsing an online shopping website like amazon you are given suggestions for purchase as you browse through the items. The items the website suggests is relevant to the items you are viewing. This is because the data of your browsing in the website is being collected as big data. Since the velocity of collecting big data being collected and being put to use at such fast rate the velocity contributes to the three Vs of the big data.

So, the term “Big Data” and “Analytic” together means using the Big Data with analytic tools for insights of the business. The higher the amount of big data the better the results from analytic tools. There were big data before 20 years as well but there were no analytic tools back then. But now there have been several advancements in terms of processing speed and storage. Since the analytic tools have gone down in price as well, more business organizations ranging from small to big are using it. The analytic tools as of now can handle very large amount of big data so there is no bottleneck when it comes to processing of big data. The data collected are of different qualities ranging from poor to excellent which is good for the big data analytic as for providing results to the user it requires to process all types of data for more accurate results. To get the insight of everchanging business the big data certainly helps the user. (Russom, 2011)

Fig: Stages of big data over years (Lu et al., 2014)

II. Issues regarding privacy and security with Big Data

According to the survey performed by Intel IT in may 2012 there was a discussion about what standards would be needed for big data analytics and the answers were: data security, technology to keep customers’ data private, data transparency, performance benchmarking, data and system interoperability (SAGIROGLU and SINANC, 2013). The results of the survey were greatly inclined towards the security and privacy of customers. The attacker’s community who can go undetected through the current security systems is quiet a bit alarming. The advancements of security can also be done through big data analytics and can be Artificially Intelligent based too. Since, the placement of big data is in a single place where it can be accessed physically by attackers. Not just physical but since all the data is collected at one place the attackers can get access of all the data which has a risk of exposing of exposing personal information of all users at once. For the physical protection of data there should be access to only limited personnel who actually requires having access for example, admin. The place where the physical hardware is kept should have only trusted employees and should be always monitored. Despite having disadvantages there are some advantages to the security of big data as the data obtained can be useful to get information about tools hackers use which later helps to avoid being hacked from those types of tools. Organizations should not keep the data which are not necessary to them and delete them. When there is a theft of data the hacker will not get the data, which are deleted which lessens the exposure of personal data of the user. There should be security rules to the organizations who keep big data to avoid being hacked as far as possible. The user whose data is being stored should know which data is being stored. The user should also have access to delete the stored data and the organization should maintain transparency with the user.

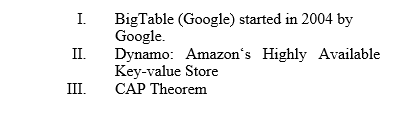
III. Conclusion on Big Data Analytics

With the advancements in software technology and growing age the data keeps increasing every day. Data collection from different sources is being done with various methods and not every data is useful. And the collection of data itself is meaningless if there are not any positive results from it. Drawing out the meaningful information is a challenge being faced.

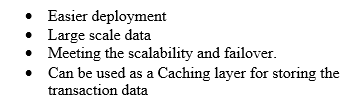
IV. Introduction to NoSQL Databases

NoSQL is a short form of Not Only SQL because some of the NoSQL applications can run the query language of SQL. Since the queries can work in both SQL and NoSQL it is named Not Only SQL to avoid confusion.

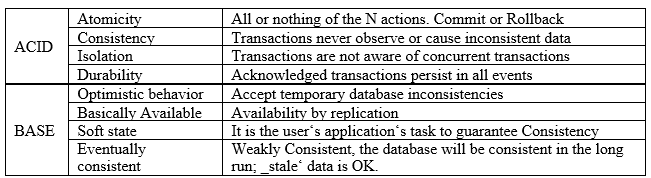
The early adopters of the NoSQL databases are below:

Fig: NoSQL Early adopters

NoSQL are highly being used because of their features. The main reasons for NoSQL being popular in the market are:

Fig: Features of NoSQL

NoSQL uses BASE whereas the SQL uses ACID properties and they both have different purposes. The database of the BASE is handled by the developer and not by the database. ACID is more reliable in financial transaction and the transaction updates should be quick. Due to the properties of ACID if there is failure in transaction the database will not see any change. BASE is more suitable for social media as if there are not any instant updates to the database there will not be any loss and the process can be repeated time and again as stale data is okay in BASE. ACID is slower while performing transactions but secure. ACID needs to be very precise and does not accept any fault whereas BASE not as accurate faults are accepted which makes it harder to work in BASE. Since BASE is suitable large amount of data to work on it is used for Big Data. (Ganesh Chandra, 2015)

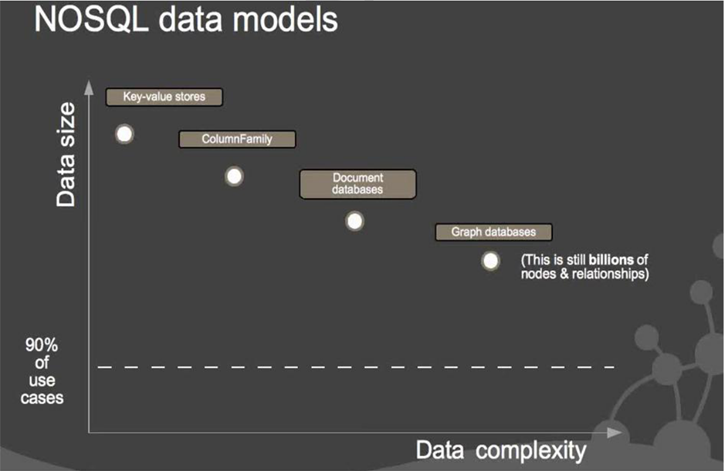
Fig: Features of ACID and Base (Ganesh Chandra, 2015)

NoSQL database overcomes the limitation of SQL databases as it can handle millions of queries thrown at it instantly. This comes handy in websites like Facebook, Google, Amazon etc. as millions and billions of users are using the platform at once and the website has to instantly provide results to their users. Since NoSQL can handle such large amount of data it is suitable for Big Data. Many modern applications use NoSQL database because of the following reasons:

Flexibility: NoSQL database supports fast developing because of its flexibility in schemas. Since NoSQL is flexible it is suitable for both structured and un-structured data.

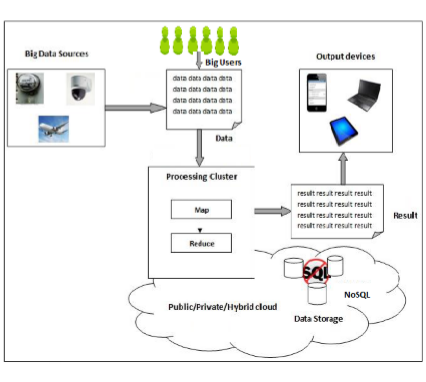
Scalability: NoSQL databases generally are scalable horizontally which means the databases can become more powerful with addition of more hardware and there is no limitation with NoSQL due to its scalability.  
Highly functional: As there are lots of NoSQL databases to pick from the user can pick according to their needs of data models making them useful to the task the user performs. (Amazon Web Services, Inc., 2019)

Because the NoSQL has high scalability huge organizations prefer NoSQL over Relational database.

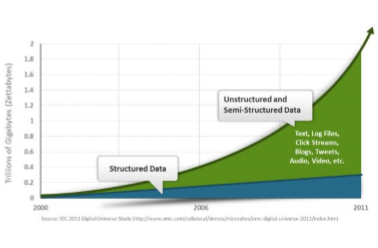
Fig: NoSQL Analysis Graph

V. Role of NoSQL in Big Data Analytics

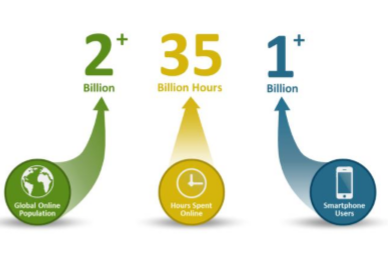
Relation model was fit back when there were limited data but as the time passed and there were more data than ever, and the Relational model could not handle the amount of data. Therefore, different architectures were developed to support all the huge amount of data. Companies like Facebook, Google, Amazon first faced problems with storing data as their data size was growing and they needed alternative ways of storing data. For this reason, the companies worked on developing a database which had no limitations called NoSQL. This was fit for the huge growth of data due to the scalability of NoSQL. In the recent world most companies are shifting their database model toward NoSQL due to the rapid growth of data generated by users, systems etc.

Fig: NoSQL database growth

Big Data can be collected from online based applications like Facebook, twitter etc. Some of the examples of big data are browsing history, sleeping time, location data etc. There are countless of resources of big data and the data collected is increasing rapidly. Since organizations with big data have the resources for the insights of the demands of user the adaptation of the market trends is easier. That is why the startup companies without big data tend to fall behind.

Fig: Growth of Big Data

There are more than two billion people who have access to internet which is a lot of big data that is being collected. This was not the case a few decades ago as just over a thousand of users were considered to be huge. But the technology has improved a lot comparatively to few decades back. People are using internet more and more daily for entertainment as well as work. These users are called big users as they are the sources of big data.

Fig: Big users with an increasing amount of time spent online. Smartphone users increasing daily which leads to increase of smartphone application users.

Due to the huge increase in number of the users there is need for scalable database technology. The users only keep increasing in online platform and the users shift through different applications which are best fit for them. The applications also adapt to the requirements of users with the help of big data to stay relevant. To keep a proper record of all the data floating in the internet the developers are shifting towards the NoSQL database. (Khan, 2014)

VI. Uses of Big Data

There are several uses of big data and some of them are listed below.

Weather Forecast: With the help of big data the prediction of weather can be done. This is done by gathering information which affects weather like ocean currents, temperature, air pressure etc. After analyzing the data, the weather can be predicted which helps to avoid casualties from natural disasters like storm.

Medical field: The big data helps the medical research as the amount of data the health experts can us to learn from are a lot. With the help of big data, the fraud in this field can be prevented which is currently being done by United Health Care.

Advertisement: Big data learn the user patterns and figures out what the user is looking for. Suppose you search for a car in google and you will start seeing advertisements relevant to the car.

Logistics: Finding out the route to your destination is easy but finding out the fastest path to your destination is difficult. This is where big data learn from the traffic of the selected route and helps you get to your destination as fast as possible. This is done through the analysis of data like traffic in the past to analyze the fastest route. (Anurag, 2020)

VII. Conclusion

Without NoSQL we would not know if the big data would be possible as NoSQL is the only database which can handle Big Data. No matter how big the data is it can be stored with the help of NoSQL. This is possible due to scalability in nature of NoSQL. With big data there are several advancements in many fields like medical, weather forecast etc. Although Big Data is being collected every day, we cannot say that it is being put use effectively. Collecting big data itself is not much of a challenge but using the collected data to its full potential can be a challenge. There are also issues of big data regarding security. Lots of people use internet in their daily life and their personal data is being stored somewhere. The theft of the persons data can affect a person’s life negatively.

Big Data is no longer held back by hardware limitations as it was before. The future of Big Data is positive as the amount of data being stored is increasing every second.

VIII. References

1. Anurag (2020). 5 Practical Uses of Big Data – NewGenApps, viewed 1st August 2020, NewGenApps The technology company with integrity. Available at: <https://www.newgenapps.com/blog/5-practical-uses-of-big-data/>.
2. Johndcook.com. (2009). ACID versus BASE for database transactions, viewed 1st August 2020, Available at: <https://www.johndcook.com/blog/2009/07/06/brewer-cap-theorem-base/>.
3. Amazon Web Services, Inc. (2019). What is NoSQL? | Nonrelational Databases, Flexible Schema Data Models | AWS, viewed 30th July 2020, Available at: <https://aws.amazon.com/nosql/>.
4. Russom, P. (n.d.). BIG DATA ANALYTICS BIG DATA A N A LY TIC S, viewed 2nd August 2020, Available at: <https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf>.
5. Ganesh Chandra, D. (2015). BASE analysis of NoSQL database. Future Generation Computer Systems, viewed 2nd August 2020, Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0167739X15001788>
6. Khan, A. (2014). NoSQL DATABASES: NEW MILLENNIUM DATABASE FOR BIG DATA, BIG USERS, CLOUD COMPUTING AND ITS SECURITY CHALLENGES, viewed 2nd August 2020, IJRET: International Journal of Research in Engineering and Technology. Available. Available at: <https://pdfs.semanticscholar.org/66f7/35b59c9bc7dcd908e6c4665e70e1a5a8e691.pdf>.
7. Sagiroglu, S. and Sinanc, D. (2013). Big data: A review. 2013 International Conference on Collaboration Technologies and Systems (CTS), viewed 2nd August 2020, Available at: <https://ieeexplore.ieee.org/document/6567202>
8. Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. and Kruschwitz, N. (n.d.). Big Data, Analytics and the Path from Insights to Value WINTER 2011 VOL. 52 NO. 2 REPRINT NUMBER 52205. viewed 3rd August 2020, Available at: <http://foresight.ifmo.ru/ict/shared/files/201309/1_9.pdf>
9. Moniruzzaman, A.B.M. and Hossain, S.A. (2013). NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison. arXiv:1307.0191 [cs], viewed 3rd August 2020, Available at: <https://arxiv.org/abs/1307.0191>
10. Sitalakshmi Venkatraman, Kiran Fahd, Kaspi, S. and Ramanathan Venkatraman (2016). SQL Versus NoSQL Movement with Big Data Analytics, viewed 4th August 2020, undefined. Available at: <https://www.semanticscholar.org/paper/SQL-Versus-NoSQL-Movement-with-Big-Data-Analytics-Venkatraman-Fahd/ff1e0618d9dcad292af5ccf05cd11dfcafecce19>
11. Han Hu, Yonggang Wen, Tat-Seng Chua and Xuelong Li (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial, viewed 4th August 2020, Available at: <https://ieeexplore.ieee.org/document/6842585>
12. Kambatla, K., Kollias, G., Kumar, V. and Grama, A. (2014). Trends in big data analytics. Journal of Parallel and Distributed Computing, viewed 2nd August 2020, Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0743731514000057>