UNIT – 1 INTRODUCTION TO BIG DATA

1.1 Introduction to Big Data

1.2 Types of Digital Data

1.3 Big Data Analytics

1.4 Application of Big Data

❖ Introduction to Big Data

➢What is Data?

The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.

➢What is Big Data?

Big Data is also **data** but with a **huge size**. Big Data is a term used to describe a collection of data that is huge in volume and yet growing exponentially with time. In short such data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently.

- "Extremely large data sets that may be analyzed computationally to reveal patterns , trends and association, especially relating to human behavior and interaction are known as Big Data."

➢ Examples Of Big Data

Following are some the examples of Big Data-

- ❑ The **New York Stock Exchange** generates about ***one terabyte*** of new trade data per day.

❑ Social Media

The statistic shows that ***500+terabytes*** of new data get ingested into the databases of social media site **Facebook**, every day. This data is mainly generated in terms of photo and video uploads, message exchanges, putting comments etc.

❑ A single **Jet engine** can generate ***10+terabytes*** of data in ***30 minutes*** of flight time. With many thousand flights per day, generation of data reaches up to many ***Petabytes.***

Tabular Representation of various Memory Sizes

Name	Equal To	Size(In Bytes)
Bit	1 bit	1/8
Nibble	4 bits	1/2 (rare)
Byte	8 bits	1
Kilobyte	1024 bytes	1024
Megabyte	1, 024kilobytes	1, 048, 576
Gigabyte	1, 024 megabytes	1, 073, 741, 824
Terrabyte	1, 024 gigabytes	1, 099, 511, 627, 776
Petabyte	1, 024 terrabytes	1, 125, 899, 906, 842, 624
Exabyte	1, 024 petabytes	1, 152, 921, 504, 606, 846, 976
Zettabyte	1, 024 exabytes	1, 180, 591, 620, 717, 411, 303, 424
Yottabyte	1, 024 zettabytes	1, 208, 925, 819, 614, 629, 174, 706, 176

❑ Characteristics Of Big Data

- The following are known as “Big Data Characteristics”.
    1. Volume
    2. Velocity
    3. Variety
    4. Veracity

1. Volume:

Volume means “How much Data is generated”. Now-a-days, Organizations or Human Beings or Systems are generating or getting very vast amount of Data say TB(Tera Bytes) to PB(Peta Bytes) to Exa Byte(EB) and more.

2. Velocity:

Velocity means “How fast produce Data”. Now-a-days, Organizations or Human Beings or Systems are generating huge amounts of Data at very fast rate.

VELOCITY = Produce Data at Very Fast Rate

3. Variety:

Variety means “Different forms of Data”. Now-a-days, Organizations or Human Beings or Systems are generating very huge amount of data at very fast rate in different formats. We will discuss in details about different formats of Data soon.

VARIETY = Produce Data in Different Formats

4. Veracity

Veracity means “The Quality or Correctness or Accuracy of Captured Data”. Out of 4Vs, it is most important V for any Big Data Solutions. Because without Correct Information or Data, there is no use of storing large amount of data at fast rate and different formats. That data should give correct business value.

❖ Types of Digital Data

1. **Structured**
2. **Unstructured**
3. **Semi-structured**

❑ Structured

- Any data that can be stored, accessed and processed in the form of fixed format is termed as a 'structured' data.
- Over the period of time, talent in computer science has achieved greater success in developing techniques for working with such kind of data (where the format is well known in advance) and also deriving value out of it.
- However, nowadays, we are foreseeing issues when a size of such data grows to a huge extent, typical sizes are being in the range of multiple zettabytes.

➢ ***Do you know? 10^{21} bytes*** equal to ***1 zettabyte*** or ***one billion terabytes*** forms ***a zettabyte***. Looking at these figures one can easily understand why the name Big Data is given and imagine the challenges involved in its storage and processing.

- ***Do you know?*** Data stored in a relational database management system is one example of a **'structured'** data.

• Examples Of Structured Data

An 'Employee' table in a database is an example of Structured Data

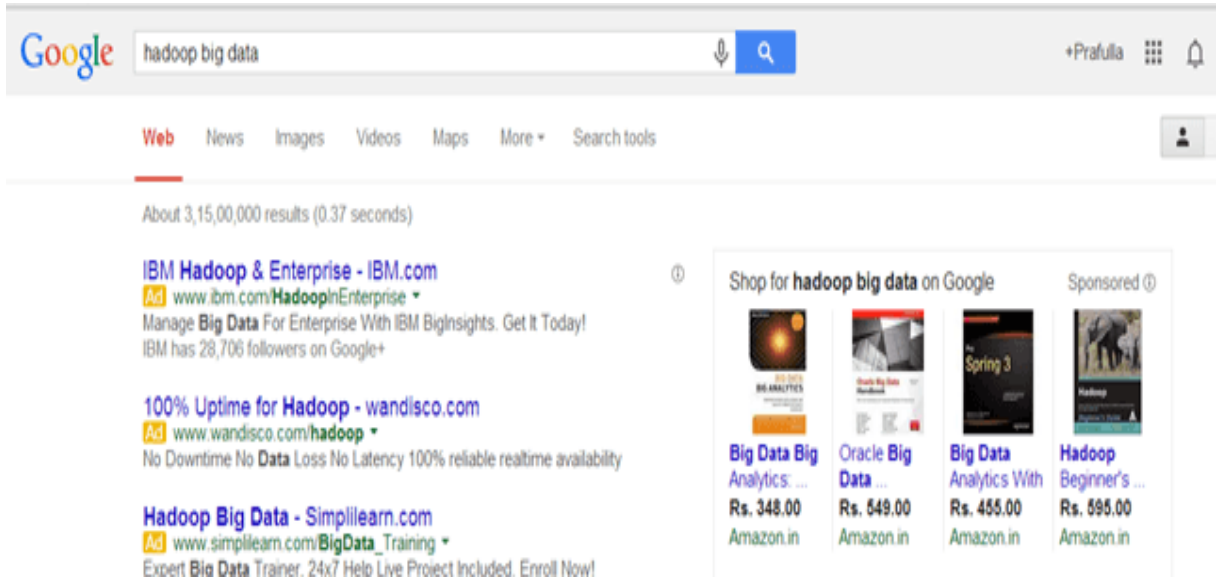
Employee_ID	Employee_Name	Gender	Department	Salary_In_lacs
2365	Rajesh Kulkarni	Male	Finance	650000
3398	Pratibha Joshi	Female	Admin	650000
7465	Shushil Roy	Male	Admin	500000
7500	Shubhojit Das	Male	Finance	500000
7699	Priya Sane	Female	Finance	550000

❑ Unstructured

- Any data with unknown form or the structure is classified as unstructured data.
- In addition to the size being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it.
- A typical example of unstructured data is a heterogeneous data source containing a combination of simple text files, images, videos etc.
- Now day organizations have wealth of data available with them but unfortunately, they don't know how to derive value out of it since this data is in its raw form or unstructured format.

• Examples Of Un-structured Data

The output returned by 'Google Search'



❑ Semi-structured

- Semi-structured data can contain both the forms of data.
- We can see semi-structured data as a structured in form but it is actually not defined with e.g. a table definition in relational DBMS.
- Example of semi-structured data is a data represented in an XML file.

➢ Examples Of Semi-structured Data

Personal data stored in an XML file-

```
<rec><name>Prashant Rao</name><sex>Male</sex><age>35</age></rec>
<rec><name>Seema R.</name><sex>Female</sex><age>41</age></rec>
<rec><name>Satish Mane</name><sex>Male</sex><age>29</age></rec>
<rec><name>Subrato Roy</name><sex>Male</sex><age>26</age></rec>
<rec><name>Jeremiah J.</name><sex>Male</sex><age>35</age></rec>
```

❖ Big Data Analytics

➢ Big Data Analytics:

- Big Data analytics is the process of collecting, organizing and analyzing large sets of data (*called* Big Data) to discover patterns and other useful information.
- Big Data analytics can help organizations to better understand the information contained within the data and will also help identify the data that is most important to the business and future business decisions. Analysts working with Big Data typically want the *knowledge* that comes from analyzing the data.

➢ High-Performance Analytics Required:

- To analyze such a large volume of data, Big Data analytics is typically performed using specialized software tools and applications for predictive analytics, data mining, text mining, forecasting and data optimization.
- Collectively these processes are separate but highly integrated functions of high-performance analytics.
- Using Big Data tools and software enables an organization to process extremely large volumes of data that a business has collected to determine which data is relevant and can be analyzed to drive better business decisions in the future.

➢ The Challenges:

- For most organizations, Big Data analysis is a challenge. Consider the sheer volume of data and the different formats of the data(both structured and unstructured data) that is collected across the entire organization and the many different ways different types of data can be combined, contrasted and analyzed to find patterns and other useful business information.
- The first challenge is in breaking down data silos to access all data an organization stores in different places and often in different systems.
- A second challenge is in creating platforms that can pull in unstructured data as easily as structured data.
- This massive volume of data is typically so large that it's difficult to process using traditional database and software methods.

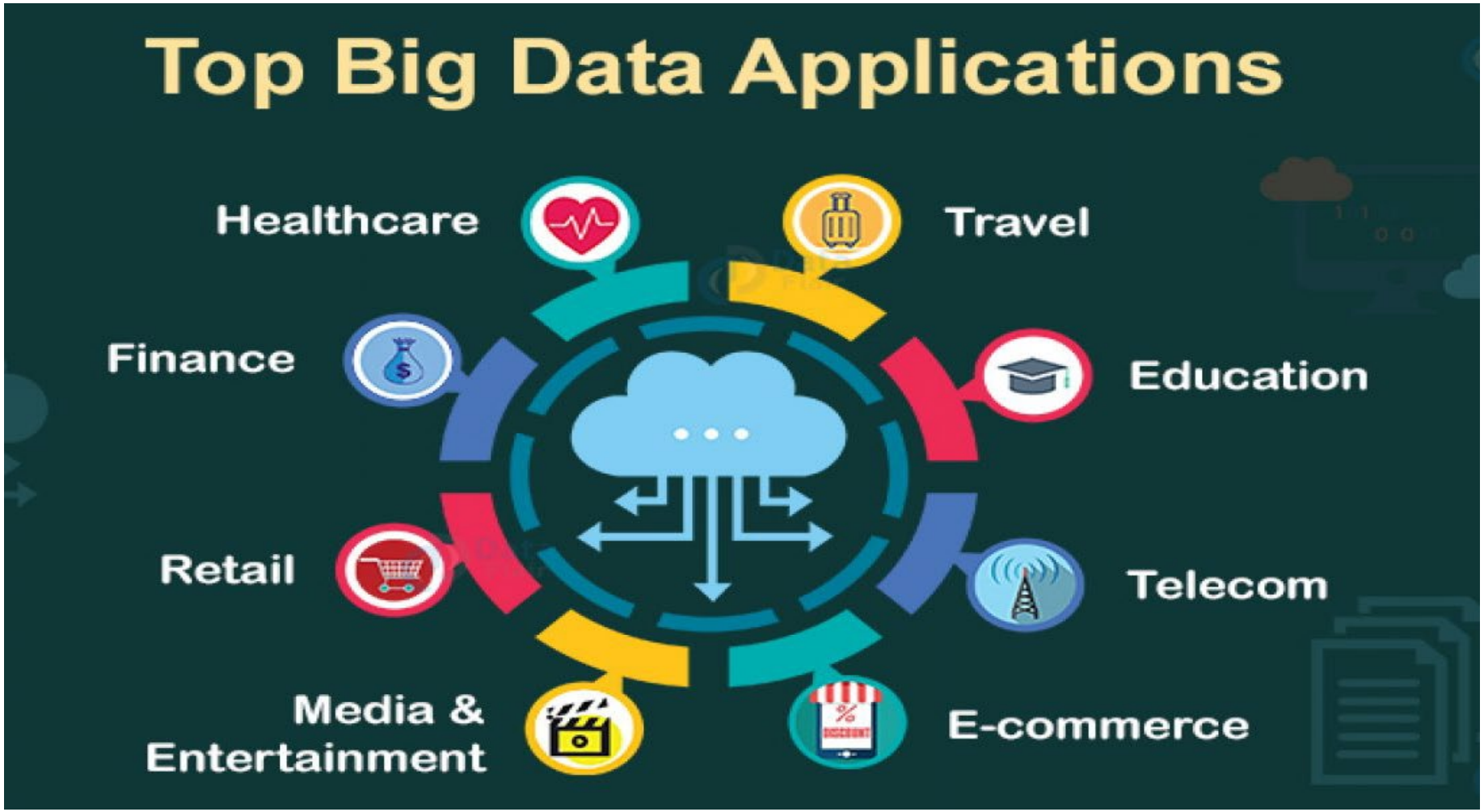
➢ How Big Data Analytics is Used Today:

- As the technology that helps an organization to break down data silos and analyze data improves, business can be transformed in all sorts of ways.
- Today's advances in analyzing big data allow researchers to decode human DNA in minutes, predict where terrorists plan to attack, determine which gene is mostly likely to be responsible for certain diseases and, of course, which ads you are most likely to respond to on Facebook.
- Another example comes from one of the biggest mobile carriers in the world.
- France's Orange launched its Data for Development project by releasing subscriber data for customers in the Ivory Coast.
- The 2.5 billion records, which were made anonymous, included details on calls and text messages exchanged between 5 million users.
- Researchers accessed the data and sent Orange proposals for how the data could serve as the foundation for development projects to improve public health and safety.
- Proposed projects included one that showed how to improve public safety by tracking cell phone data to map where people went after emergencies; another showed how to use cellular data for disease containment. (source)

➢ The Benefits of Big Data Analytics:

- Enterprises are increasingly looking to find actionable insights into their data. Many big data projects originate from the need to answer specific business questions. With the right big data analytics platforms in place, an enterprise can boost sales, increase efficiency, and improve operations, customer service and risk management.
- Webopedia parent company, QuinStreet, surveyed 540 enterprise decision-makers involved in big data purchases to learn which business areas companies plan to use Big Data analytics to improve operations. About half of all respondents said they were applying big data analytics to improve customer retention, help with product development and gain a competitive advantage.
- Notably, the business area getting the most attention relates to increasing efficiency and optimizing operations. Specifically, 62 percent of respondents said that they use big data analytics to improve speed and reduce complexity.

❖ Application of Big Data



✓ Here is the list of top Big Data applications in today's world:

- Big Data in Healthcare
- Big Data in Education
- Big Data in E-commerce
- Big Data in Media and Entertainment
- Big Data in Finance
- Big Data in Travel Industry
- Big Data in Telecom
- Big Data in Automobile

✓ Let's discuss the applications of Big Data in detail.

1. Big Data in Retail

- The retail industry is the one that faces the most fierce competition of all. Retailers constantly hunt for ways that will give them a competitive edge over others. Customers are the real king sounds legit for the retail industry in particular.
- For retailers to thrive in this competitive world, they need to understand their customers in a better way. If they are aware of their customers' needs and how to fulfill those needs in the best possible way, then they know everything.

- ***Check how Big Data act as a weapon for retailers to connect with their customers*** – ***Big Data in Retail*.**
  - Through advanced analysis of their customer's data, retailers are now able to understand them from every angle possible. They gather this data from various sources such as social media, loyalty programs, etc.

Big Data in Retail Industry

- Even a minute detail about any customer has now become significant for them. They are now closer to their customers than they have ever been. This empowers them to provide customers with more personalized services and predict their demands in advance.
- This helps them in building a loyal customer base. Some of the biggest names in the retail world like Walmart, Sears and Holdings, Costco, Walgreens, and many more now have Big Data as an integral part of their organizations.
- A study by the National Retail Federation estimated that sales in November and December are responsible for as much as 30% of retail annual sales.

2. Big Data in Healthcare

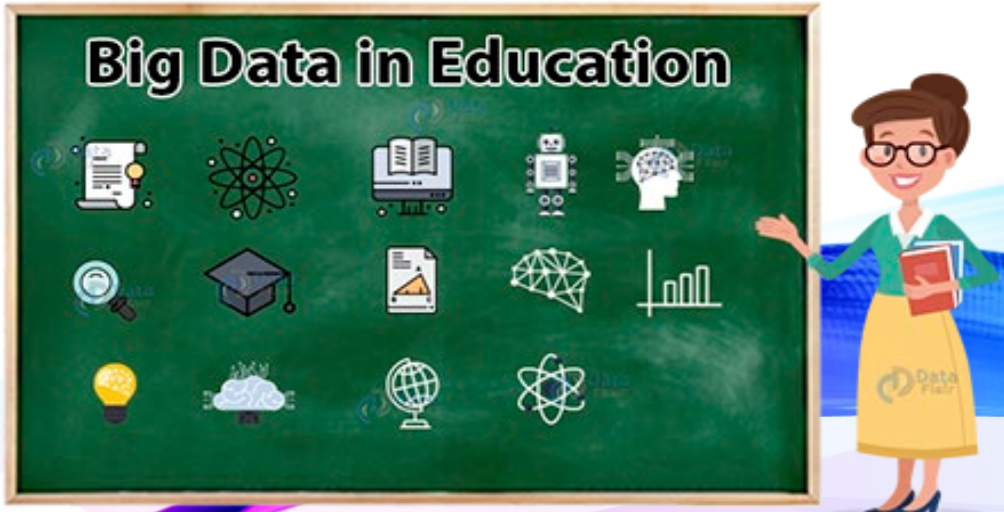
- Big Data and healthcare are an ideal match. It complements the healthcare industry better than anything ever will. The amount of data the healthcare industry has to deal with is unimaginable.
- Gone are the days when healthcare practitioners were incapable of harnessing this data. From finding a cure to cancer to detecting Ebola and much more, Big Data has got it all under its belt and researchers have seen some life-saving outcomes through it.
- Big Data and analytics have given them the license to build more personalized medications. Data analysts are harnessing this data to develop more and more effective treatments. Identifying unusual patterns of certain medicines to discover ways for developing more economical solutions is a common practice these days.

Big Data in Healthcare Industry

- ***Explore how Big Data helps to speed up the treatment process – Big Data in Healthcare.***
- Smart wearables have gradually gained popularity and are the latest trend among people of all age groups. This generates massive amounts of real-time data in the form of alerts which helps in saving the lives of the people.

3. Big Data in Education

- When you ask people about the use of the data that an educational institute gathers, the majority of the people will have the same answer that the institute or the student might need it for future references.
- Even you had the same perception about this data, didn’t you? But the fact is, this data holds enormous importance. Big Data is the key to shaping the future of the people and has the power to transform the education system for better.
- Some of the top universities are using Big Data as a tool to renovate their academic curriculum. Additionally, universities can even track the dropout rates of the students and are taking the required measures to reduce this rate as much as possible.



4. Big Data in E-commerce

- One of the greatest revolutions this generation has seen is that of E-commerce. It is now part and parcel of our routine life. Whenever we need to buy something, the first thought that provokes our mind is E-commerce. And not your surprise, Big Data has been the face of it.
- Some of the biggest E-commerce companies of the world like Amazon, Flipkart, Alibaba, and many more are now bound to Big Data and analytics is itself an evidence of the level of popularity Big Data has gained in recent times.
- Big Data is now as important as anyone else in these organizations. *Amazon, the biggest E-commerce firm in the world and one of the pioneers of Big Data and analytics, has Big Data as the backbone of its system. Flipkart, the biggest E-commerce firm in India, has one of the most robust data platforms in the country.*
- ***See how Flipkart used Big Data to have one of the most robust data platforms.***
- Big Data's recommendation engine is one of the most amazing applications the Big Data world has ever witnessed. It furnishes the companies with a **360-degree** view of its customers.
- Companies then suggest customers accordingly. Customers now experience more personalized services than they have ever had. Big Data has completely redefined people's online shopping experiences.

5. Big Data in Media and Entertainment

- Media and Entertainment industry is all about art and employing Big Data in it is a sheer piece of art. Art and science are often considered to be the two completely contrasting domains but when employed together, they do make a deadly duo and Big Data's endeavors in the media industry are a perfect example of it.

- Viewers these days need content according to their choices only. Content that is relatively new to what they saw the previous time. Earlier the companies broadcasted the Ads randomly without any kind of analysis.
- But after the advent of Big Data analytics in the industry, companies now are aware of the kind of Ads that attracts a customer and the most appropriate time to broadcast it for seeking maximum attention.
- Customers are now the real heroes of the Media and entertainment industry - courtesy to Big Data and Analytics.

6. Big Data in Finance

- The functioning of any financial organization depends heavily on its data and to safeguard that data is one of the toughest challenges any financial firm faces. Data has been the second most important commodity for them after money.
- Even before Big Data gained popularity, the finance industry was already conquering the technical field. In addition to it, financial firms were among the earliest adopters of Big Data and Analytics.
- **Digital banking** and **payments** are two of the most trending buzzwords around and Big data has been at the heart of it. Big Data is bossing the key areas of financial firms such as **fraud detection**, **risk analysis**, **algorithmic trading**, and **customer contentment**.
- This has brought much-needed fluency in their systems. They are now empowered to focus more on providing better services to their customers rather than focussing on security issues. Big Data has now enhanced the financial system with answers to its hardest of the challenges.

7. Big Data in Travel Industry

- While Big Data is spreading like wildfire and various industries have been cooking its food with it, the travel industry was a bit late to realize its worth. Better late than never though. Having a stress-free traveling experience is still like a daydream for many.
- And now Big Data's arrival is like a ray of hope, that will mark the departure of all the hindrances in our smooth traveling experience.

- ***See how Big Data is revolutionizing the travel & tourism sector.***
- Through Big Data and analytics, travel companies are now able to offer more customized traveling experience. They are now able to understand their customer's requirements in a much-enhanced way.
- From providing them with the best offers to be able to make suggestions in real-time, Big Data is certainly a perfect guide for any traveler. Big Data is gradually taking the window seat in the travel industry.

8. Big Data in Telecom

- The telecom industry is the soul of every digital revolution that takes place around the world. With the ever-increasing popularity of smartphones, it has flooded the telecom industry with massive amounts of data.
- And this data is like a goldmine, telecom companies just need to know how to dig it properly. Through Big Data and analytics, companies are able to provide the customers with smooth connectivity, thus eradicating all the network barriers that the customers have to deal with.
- Companies now with the help of Big Data and analytics can track the areas with the lowest as well as the highest network traffics and thus doing the needful to ensure hassle-free network connectivity.
- Big Data alike other industries have helped the telecom industry to understand its customers pretty well.
- Telecom industries now provide customers with offers as customized as possible.
- Big Data has been behind the data revolution we are currently experiencing.

9. Big Data in Automobile

- *"A business like an automobile, has to be driven, in order to get results." B.C. Forbes*
- And Big Data has now taken complete control of the automobile industry and is driving it smoothly. Big Data is driving the automobile industry towards some unbelievable and never before results.
- The automobile industry is on a roll and Big Data is its wheels or I must say Big Data has given wings to it. Big Data has helped the automobile industry achieve things that were beyond our imaginations

- From analyzing the trends to understanding the supply chain management, from taking care of its customers to turning our wildest dream of connected cars a reality, Big Data is well and truly driving the automobile industry crazy.