CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

Big data are quickly all over the place. Everyone seems to be collecting, analyzing, and making money from it. No matter whether we are chatting about analyzing zillions of Google search queries to suppose flu outbreaks, or zillions of phone proceedings to notice signs of terrorist activity, or zillions of airline stats to find out the best time to purchase plane tickets, big data are on the case. By combining the influence of modern computing with the huge data of the digital era, it promises to crack almost any problem like crime, public health, the evolution of grammar, etc.

According to Google Trends, the amount of searches using the keyword big data started to augment dramatically in 2011 and reached its peak this year. While the term big data sounds as if it is connected to the area of data science only, it really plays an significant role in healthcare research, including emergency medicine. Ten years ago, the level of a DVD could be considered large, but now such a volume is common. Furthermore, the size of the data is largely dependent on the data-compression technique adopted. Similarly, a 56K modem used to be measured fast, but broadband Internet speeds over 100M are everywhere in today's society. Another definition of big data is much more realistic. According to this definition, researchers use big data if their data covers the whole population rather than only a sample of the population. However, out-of-date computing infrastructures are not healthy enough to handle big data, making the big data approach not practicable. Moreover, clinical practitioners may not be relaxed with this approach because the bulk of their research involves straight contact with patients. It is hard to imagine a research project involving all the patients who fit the selection criteria. To save resources, clinical practitioners have to take a sample from the population and use inferential statistics to estimate the population's characteristics.

The aim of big data organization is to make sure a high level of data quality and handiness for business intelligence and big data analytics applications. Corporations, government agencies and other organizations utilize big data management strategies to help them compete with fast-
growing pools of data, typically involving a lot of terabytes or even petabytes of information saved in a variety of file formats. Effective big data administration helps companies set valuable information in great sets of formless data and semi-structured data from a variety of sources, including call detail records, system logs and social media sites.

Internet is the main source which has resulted in the tsunami of data in the past few years. Big data is too big, it moves too fast, and doesn’t fit the structures of our presented database architectures. It is like an ocean of data in which we people spin in every day with an attempt to come on the surface, but every day the stage of data increases greatly. Gone are the days when memory was used to be considered in Gigabytes or Terabytes or Petabytes, today it is measured in exabytes, zettabytes or yottabytes. With Big Data solutions, organizations can jump into all facts and gain precious insights that were previously unthinkable. The term “big data” can be striking unformulated, in the same way that the term “cloud” covers varied technologies. Utilizing big data requires transforming information infrastructure into a more flexible, distributed, and open environment.

Big data promises deeper insights that data scientists are extremely involved in exploring this data in such a way that organizations are benefited to its finest with total customer approval. Big data analytics is one of the immense new frontiers of IT. Rising technologies such as the Hadoop framework and Map-Reduce present new and thrilling ways to process and transform big data—defined as compound, unstructured, or big amounts of data—into meaningful insights, but also need IT to organize infrastructure in a diverse way to support the distributed processing requirements and real-time demands of big data analytics.

Big data is a large term used for data sets are huge or hard so as to recognized data processing applications are insufficient. Challenges include investigation, arrest, data duration, discover, allocation, storage space, move, publicity, querying and information division. The word over and over again refers merely to apply analytical or influenced new complex methods to take away significance from data, and infrequently toward a meticulous size of data set. Accuracy in big data could straight to additional certain choice making, and better decisions can attain in better prepared efficiency, cost lessening and condensed possibility.
Big data is data that exceeds the processing ability of traditional database systems. The data is too large, moves too rapid, or doesn’t fit the structures of our database architectures. To gain worth from this data, one must decide an alternative way to process it.

Every day, we make 2.5 quintillion bytes of data— to a great extent that 90% of the data in the world today has been shaped in the last two years alone. As a catch-all term, “big data” can be good-looking imprecise, in the same way that the term “cloud” covers a variety of technologies. Input data to big data systems could be babble from social networks, web server logs, traffic run sensors, satellite imagery, broadcast audio streams, banking transactions, MP3s of rock music, the content of web pages, scans of government documents, GPS trails, telemetry from automobiles, financial market data, the list goes on. This data is “big data”.

<table>
<thead>
<tr>
<th>Memory Unit</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>KiloByte (KB)</td>
<td>$10^3$</td>
</tr>
<tr>
<td>MegaByte (MB)</td>
<td>$10^6$</td>
</tr>
<tr>
<td>GigaByte (GB)</td>
<td>$10^9$</td>
</tr>
<tr>
<td>TeraByte (TB)</td>
<td>$10^{12}$</td>
</tr>
<tr>
<td>PetaByte (PB)</td>
<td>$10^{15}$</td>
</tr>
<tr>
<td>ExaByte (EB)</td>
<td>$10^{18}$</td>
</tr>
<tr>
<td>ZettaByte (ZB)</td>
<td>$10^{21}$</td>
</tr>
<tr>
<td>YottaByte (YB)</td>
<td>$10^{24}$</td>
</tr>
</tbody>
</table>
Study of information sets can find out original correlations to mark commerce trends, avoid diseases and competition offense and so on. Scientists, business executives, practitioners of medicine, marketing and governments similar often meet difficulties with big data sets in areas as well as search of internet, finance and information of business. Scientists meet limitations within science work, including meteorology, genomics, compound physics simulations, biology and environmental study.

Data sets are growing rapidly in division since they are more and more gathered by despicable and frequent information-sensing mobile devices, remote sensing, software logs, cameras, microphones, RFID readers and wireless sensor networks. The world's scientific per head ability to store information has roughly doubled since the 1980s; as of 2012, one question for large enterprises is influential who must have big data initiatives that involve the whole association.

Relational database management systems, desktop figures and apparition packages often have difficulty in handling of big data. The work requires "similar software organization on tens, hundreds, or even thousands of servers". What is measured "big information" vary depending on the capabilities of the users and their tools, and upward capabilities create big data in a moving goal. "For several organizations, facing hundreds of gigabytes of data for the first time require to review data organization options. For others, it could obtain tens or hundreds of terabytes before data size becomes a important reflection.

Mixture of clinical innovation and technology is Data analytics. As the healthcare industry is continuously generating large amounts of data in different forms, it is almost impossible to manage this data over soft or hard copy formats. Digitization of large amounts of data is favoured by this current trend. This promising technique supports a wide range of healthcare functions to improve services and handle problems in healthcare sector. Big Data platform is capable of processing terabytes and petabytes of data, due to which data analysis becomes easier.

Big data analytics has offered a new way for healthcare organizations to develop actionable insights, organize their future vision, boost up the outcomes and reduce time to value. This
approach is also helpful to provide insightful information to the healthcare enterprises regarding their management, planning and the measurements. The evaluated results helps in enhancing the decision making capacity of top management.

Data analytics in healthcare can be used to raise the standards in following fields:-

- **Public Health:** With analytics approach, public health issues can be improved by analyzing disease patterns and recording disease outbreaks. Large amount of data help in determining needs, offers required services, predict and prevent the future crisis to benefit the population.

- **Electronic Medical Record or EMR:** An EMR contains the standard (structured and unstructured) medical data that can be evaluated with the data analytic approach to predict patients at risk and provide him effective care.

- **Patient Profile Analytics:** Advanced analytics can be applied to patients profile for identifying individuals who could benefit from practical approach. This may include lifestyle changes.

- **Genomic Analytics:** The data analytic approach can be effectively included in genomic analytics for making this approach a part of medical care decision process.

- **Fraud Analysis:** This data analytics approach helps examine greater number of claim requests to cut back down fraud cases. An effective analysis helps in reducing fraud.

- **Safety Monitoring:** Data analytics can also be used to examine real time large volumes of brisk data in hospitals. The approach may help in the safety monitoring and negative event prediction.

Data analytics in healthcare has the potential of transforming the ways healthcare organizations offer their sophisticated clinical facilities. This powerful analysis method is ideal to improve the care, save patient’s life and lower the cost of the health facilities charges.
Why Data Analytics in Healthcare

The health data volume is increasing and the graph breeds dramatically in the years ahead. However, the current data managing ways are not comparable to data analytics for evaluation techniques. These data analysis techniques have the capacity to capture, process, distribute and manage the analysis in specific form that makes it easy to get trustworthy information. With this particular evaluation, vast amount of patient-related health data is analyzed in a better way to get a deeper understanding of outcomes, which may be applied at the point of care for better facilities.

This technique is equally effective in healthcare organizations and used to manage:

- Specific individual and population health issues
- Fraud management in an effective way
- Clinical innovations to determine affordable ways to diagnose and treat patients
- Effective R&D methods for the drugs and devices
- New treatments that are introduced into the market
- Analyze clinical trials prior to a product (medicine or device) reach to the market

1.2 CORES OF BIG DATA

- **Data Volume** – It refers to the huge amount of data that is been created each second, minute and hour of the day. 571 websites are created in a single minute and total of 625000 GB of data is transferred from one end to another in single internet minute, may be terms of mails, pictures, posts etc. If we burn the amount of data present on planet earth today on DVDs and pile them in the form of a stack one upon another, the
A pile will be such huge that one can climb it and touch the moon, come back to earth and again repeat this process once.

- **Data Velocity** – Data is being created at such high velocity that it acts as difficulty to the companies in coping up with such high speeds. They have to establish their infrastructure in such a manner that it is capable of handling such generated data.

- **Data Variety** – Totally diverse data in the form of raw, structured, semi structured and even unstructured form is generated, which is difficult to be handled by the existing traditional analytic systems. Mismatched data formats and data structures represent significant challenges that can lead to analytic collapse.

- **Data Value** – There is a huge gap in between the business leaders and the IT professionals. The main concern of business leaders is to just add value to their business and increase their profit. On the other hand, IT personals provide techniques for handling and processing data.

- **Data Complexity** – Biggest challenge while running big data using relational databases is that they require parallel software for running on hundreds of servers and data scientists have to match and transform data across systems coming from various sources.

- **Data Veracity** - Veracity refers to the delicacy of data or how much faith one can have on data. The data on internet is not always accurate or precise. For example, if some male pretends himself as a female on his facebook profile, there is no authenticity check in such cases.

### 1.3 Big Data Characteristics

Big data analysis has gotten a lot of publicity recently, and for good reason. To become the part of this movement one has to know about big data analysis. This pushing the envelope on analysis is an exciting aspect of the big data analysis movement.

Companies are excited to be able to access and analyze data that they’ve been collecting or want to gain insight from, but have not been able to manage effectively. It might involve visualizing huge amounts of dissimilar data, or it might involve advanced analyzed streaming at you in real time. It is both evolutionary and revolutionary [6].
There exists two views of big data:

- Decision-oriented
- Action-oriented

**Decision-oriented analysis** is more similar to traditional business intelligence. Look at selective subsets and representations of larger data sources and try to apply the results to the process of making business decisions. Certainly these decisions might result in some kind of action or process change, but the purpose of the analysis is to augment decision making.

**Action-oriented analysis** is used for rapid response, when a pattern emerges or specific kinds of data are detected and action is required. Taking advantage of big data through analysis and causing proactive or reactive behavior changes offer great potential for early adopters.

Finding and utilizing big data by creating analysis applications can hold the key to extracting value sooner rather than later. To accomplish this task, it is more effective to build these custom applications from scratch or by leveraging platforms and/or components.

First, look at some of the additional characteristics of big data analysis that make it different from traditional kinds of analysis aside from the three Vs of volume, velocity, and variety:

**It can be programmatic.** One of the biggest changes in analysis is that in the past you were dealing with data sets you could manually load into an application and explore. With big data analysis, you may be faced with a situation where you might start with raw data that often needs to be handled programmatically to do any kind of exploration because of the scale of the data.

**It can be data driven.** While many data scientists use a hypothesis-driven approach to data analysis (develop a premise and collect data to see whether that premise is correct), you can also use the data to drive the analysis — especially if you’ve collected huge amounts of it. For example, you can use a machine-learning algorithm to do this kind of hypothesis-free analysis.
**It can use a lot of attributes.** In the past, you might have been dealing with hundreds of attributes or characteristics of that data source. Now you might be dealing with hundreds of gigabytes of data that consist of thousands of attributes and millions of observations. Everything is now happening on a larger scale.

**It can be iterative.** More compute power means that you can iterate on your models until you get them how you want them. Here’s an example. Assume you’re building a model that is trying to find the predictors for certain customer behaviors associated. You might start off extracting a logical sample of data or connecting to where the data resides. You might build a model to test an assumption.

In past you don’t need much memory to make your model work effectively, but the thing you need is tremendous amount of physical memory to go through the necessary iterations required to train the algorithm. It may also be necessary to use advanced computing techniques like natural language processing or neural networks that automatically develop the model based on learning.

**It can be rapid to get the compute cycles you need by leveraging a cloud-based infrastructure as a Service.** With Infrastructure as a Service (IaaS) platforms like Amazon Cloud Services (ACS), you can quickly provision a cluster of machines to swallow large data sets and analyze them quickly.

### 1.4 Phases Involved in Big Data

Big data processing involves five different phases

- **Data Acquisition and Recording** – Big data surely have some basis of origin. It is not formed from a vacuum. Different scientific experiments being accepted out in the world today produces petabytes of data per day. Much of this data is of no use and has to be filtered out. The first challenge is to set filtering parameters so such that useful data doesn’t get discarded. For example, suppose one sensor reading differs considerably from the rest: it is likely to be due to the sensor being defective, but how
can we be sure that it is not an artifact that deserves attention? Research that can intelligently process this raw data to a size that its users can handle while not missing the needle in the haystack is needed. The second challenge is related to automatically generating right metadata for illustrating the type of recorded data, method of its recording and measurement. In scientific M experiments, considerable detail regarding specific experimental conditions and measures may be required that are able to interpret the results correctly, and it is important that such metadata be recorded with observational data.

- **Information Extraction and Cleaning** – It is stated here that information collected is not in an analysis ready format. For example, consider the collection of electronic health records in a hospital, comprising transcribed dictations from a number of physicians, structured data from sensors and measurements, and image data such as x-rays. The data in this format cannot be efficiently analyzed. An information extraction process should be applied for such data to extract the required information from the sources under consideration and present it in a structured format suitable for analysis. This is really a big challenge. This data may include images and videos and such extraction is highly application dependent.

- **Data Integration, Aggregation, and Representation** – It is not sufficient to merely collect, record and throw the data into a storehouse. If we have huge data sets in storehouse, then it will be almost not possible for the user to find the desired data when required. But with adequate amount of metadata there is some expect but still challenges persists due to differences in experimental details and in data record structure. Data demanding is much more than simply locating, identifying, understanding and citing data. All this process wants to happen in a complete automatic manner for an effective large scale analysis. Suitable database design is most significant. There are many different ways in which data can be stored. Certain designs will be better than others for convinced purposes and possibly may carry drawbacks for other purposes. Therefore it can be concluded that database design is an art and needs to be carefully executed by trained professionals.
• Query Processing, Data Modeling, and Analysis Methods for querying and mining – There is no doubt in the fact that big data is varied, imprecise and not structured. Even then big data is a large amount of value as compared to small individual comments as general statistics obtained from large sample are more accurate. When it comes to mining, it requires clean and efficiently accessible data. Provision should be there for declarative query and mining interfaces. well-organized mining algorithms and computing environments is another important requirement.

• Interpretation – The analysis of big data remains of no value if users are not able to understand the analysis concept. Decision maker is provided with the result of analysis and is predictable to understand these results. This explanation requires efforts. It involves intensely examining all the assumptions made and retracing the analysis. There are more than a few sources of errors like system may carry bugs and conclusions may be based on error prone data. No responsible user will yield authority to computer system for all this. as an alternative one will try to recognize and verify the results produced by computer system. All this should be made easy by computer system and this is a big challenge with big data due to its density.

1.5 APPLICATIONS OF BIG DATA

• Government

Big Data’s implementation in governmental processes is useful and allows efficiencies in terms of cost, efficiency and originality. A number of divisions of administration are required for data studies and it produces clean and innovative processes to carry out the desired result.

Manufacturing

The maximum advantage of Big Data is used for developing improvement for preparation and product quality. Big data provides an infrastructure for simplicity in automated trade, which is the ability to disentangle doubts such as inequitable component performance and convenience
analytical developed while a suitable move toward near-zero downtime and precision require huge amount of data and highly developed prediction tools used for a orderly process of data into useful information. A theoretical structure of logical manufacturing begins with data accomplishment where similar type of sensory data is accessible to achieve such as acoustics, trembling, power, current, electrical energy and checker information. Huge amount of sensory data in addition to past data make up the big data. The generated big data acts as the input into systematic tools and aggressive strategies such as Health organization.

- **Healthcare**

Big data analytics helps healthcare get better by providing modified medicine and firm analytics, clinical risk interference and orderly analytics, waste and care inconsistency reduction, automatic exterior and internal coverage of patient data, standardized medical terms and patient registries and permanent point solution.

- **Sports**

Big data can be used for preparation of training and understanding competitors. Moreover, it is possible to predict winners in a match as well as performance of players might. Therefore, data collected during season helps to predict the value and salary of players.

In Formula One races, race cars with hundreds of sensors create terabytes of data. These sensors can gather the data points from tire pressure to fire burn efficiency. After that, this data is transferred to team through fiber optic cables that could bring data at the speed of light. Based on the data, engineers and data analysts make a choice to win the race. In addition, using big data, race teams try to expect the time they will close the race in advance, based on simulations using data composed over the time.

**1.6 ROLE PLAYED BY BIG DATA IN MEDICAL SCIENCE**

Devices like cell phones which are a part of our daily life give us enormous stream of data on human life. Behavioral data obtained from these devices helps to predict long term health conditions and design better analytic tools, prevent diseases, increase access to health care and reduce cost of health care. Significant application areas are mental health, environmental health,
chronic and infectious disease, health care cost, health care quality, accidents and injury. Along with these promises, there are also certain issues related to big data in medical science like data privacy and possession, burglary of personal data, its maltreatment and new scientific risks.

Medical care and medically related research is becoming superior and complex day by day and results in enormous volumes of data. All this is further propagated with the arrival of new technologies. Processing big data professionally and effectively limits the advancement of medical care and change of science into modern science. The sources of data in medical science ranges from human heredity and pathogen genomics to routine clinical certification, from internal imaging to motion capture, from digital epidemiology to pharmacokinetics and from action pathways to life course estimation.

Fig. 1 Health care goes digital and amounts to huge data

Health care is one of the top social and financial issues in many countries, such as the India, the UK, South Korea, The United States and even middle-income countries. In India, health care
sector suffers from underfunding and awful primacy. There is no uncertainty in this fact that India has made enormous improvements since independence; majority (70%) of these efforts has been led by the private sector. Still India accounts for 21% of the world’s load of disease.

The key challenges faced by big data in medical science are:

- Data accuracy
- Data disintegration
- Data protection for commercial or cultural reasons, or related to personal privacy.
- Data handling (data management, data access, data quality, data querying, data sharing)
- Data privacy and integrity
- Data conceptualization

Formless data forms close to 80% of information in the healthcare industry and is rising exponentially. In receipt of access to this unstructured data such as output from medical devices, doctor’s notes, lab results, imaging reports, medical correspondence, clinical data, and financial data is valuable resource for improving patient care and increasing effectiveness.

In the last few years there has been a move toward evidence-based medicine, which involves making use of all clinical data accessible and factoring that into clinical and advanced analytics. The outcomes of this connection comprise better capability to sense and diagnose diseases in their early stages, assigning more efficient therapies based on a patient’s heritable structure, and adjusting drug doses to reduce side effects and recover efficiency.

In recent years the Indian government has amplified expenditure in the health care industry. The government tactics to augment it even further by 2.5% of the GDP in the 12th five year plan. As compared to other rising economies the amount of public support that India invests in health care is extremely minute. India ranks among the last 5 countries with 6% of GDP expenditure on health care. Hospital divan thickness in India has been stagnating at 0.9 per 1000 population since 2005 and falls considerably short of WHO laid guidelines of 3.511 per 1000 patients’ population. In addition, there is a huge inequality in utilization of facilities at the
village, district and state levels with state level facilities remaining the most tensed. India is presently accepted to have approximately 600,000 doctors and 1.6 million nurses. This interprets into one doctor for every 1,800 people. The suggested WHO guidelines suggest that there should be 1 doctor for every 600 people. This translates into a resource gap of approximately 1.4 million doctors and 2.8 million nurses. There is also a lucid imbalance in the man power present in the rural and urban areas.

Fig. 2 Difference between working of traditional and modern medical science

These are some examples that illustrate how big data is being used in healthcare, helping to increase efficiency and improve patient care.
A. Improving health care quality while reducing costs

Big data is quite efficient to expose inconsistencies in health care system. If these inconsistencies are eradicated from health care section, it will certainly improve and costs will be tremendously reduced. It is frequently noticed that almost similar health services in different places are having different costs. Costs of these services are not clear and depend much on the bargaining power of the providers. With the help of big data analytics, patients will be able to access information on the doctors who produces the highest cost for exact events. It is up to doctors to know which tests are needless and can be avoided to lower the treatment load on patient’s pocket.

B. Assisted Diagnosis

Big data being able to access a wide combination of knowledge across many data sources aids in the precision of diagnosing patient circumstances. Assisted examination is accomplished using expert systems that hold full knowledge of circumstances, symptoms, medications and side effects.

C. Fraud Detection

Healthcare organizations need to be able to notice fraud based on examination of irregularity in billing data, scientific standard data or patient records. For example, they can study patient records and billing to notice anomalies such as a hospital’s over utilization of services in short time periods, patients receiving healthcare services from different hospitals in different locations concurrently, or the same prescriptions for the same patient packed in numerous locations.

D. Accident and safety

Major source of death and long term disability is accidents. Data linked to accidents like when, where and how accidents occur form a large part of strategy involved in big data. Effectual accident databases exists in developed countries exist because healthcare providers are necessary to case digital accident reports containing correct location, time and other related data. Though, in rural and poor regions, such reporting is usually not possible. But it is possible to report digital health information collected by rural healthcare workers, often on smart phones or
similar devices. In some of the best rural health systems, nurse midwives call each family once or twice a year, and collect a wide variety of digital health information. These data can then be shared with information such as traffic patterns or environmental conditions, and thus help to get better policy and set priorities.

E. Monitor Patient Vital Signs

Healthcare facilities are looking to give more proactive care to their patients by frequently monitoring patient’s essential symbols. The data from these different monitors can be used in real time and send alerts to nurses or care providers so they identify immediately about changes in a patient’s condition.

F. Digitization of Data

Till date most of the data in the health care industry are stored in the form of firm copy, but the present inclination is toward fast digitization of these huge amounts of data. The medical society has accepted big data as a research tool and attitude that the society’s vast amount of miscellaneous health information has possible to assist solve some of medicine’s most worrying problems. By discovering relations and accepting patterns within the data, big data analytics has possible to improve care, save lives and lower costs.

1.7 Big Data security issues

A Big Data key is an exact information system in itself, include application, processing components, network, data storage but with the special feature that it calls on enormous use of data from a wide variety of sources, as well as scattered processing and storage resources. Not suddenly, the security events should necessary for this kind of environment similar to those wanted to protect every information system.

Typically, they fall into three categories:
• Security governance, to depict the protecting actions and controls that are best suited to the commerce concern about Big Data

• Information systems protection, which involves to implement the suitable mechanisms on the correct stage, without magnifying it, but also without being too inexperienced.

• And lastly, information systems management, since caring events only are no longer enough to deal with increasingly compound threats.

However, there are certain specific issues connecting to Big Data in all of these three areas:

1) Governance: data used for new purposes involves adapting safety policies

Business data represents the middle information advantage for about all the organizations. Their security policies built about an expression that needs for data security expressed in conditions of different criteria, as well as accessibility and convenience, consistency and stability, confidentiality, as well as traceability and evidence for the purposes that can be distinct by the business.

The rational tools based on enormous quantities of data leads:
- On the sole hand, to the use of reachable records for those purposes which was not originally planned. Data that can be used to direct the existing security policies that are being breached, especially in relative to regulatory constraint.

- On the other hand, to the importing of new kind of data into the involvement – for example through social networks – for purposes that have not been obviously define. This kind of data may require security policies to be revised, to take an account which are particularly responsive in nature.

- Lastly, the penalty of this investigation represents, in themselves, new high additional rate business data which it is necessary to protect.

The opening of a data analysis answer requires revision of accessible policies to incorporate new uses of business data and an growth of those policies to fit in issues that are precise to new data.

**Security: maintenance of data can be classified by restricting access to it rather than using encryption**

When it comes to data security, adding to more conventional methods that employ putting up a protective barrier about the data and ensuring necessary safety, specific events may be preferred depending on how responsive the data is. To keep privacy, we advise two courses of action:

- The simplest way involves ensuring confined control more access to the data, using a uniqueness and access organization solution at the very last.

- If compulsory, encrypting is most responsive step in data analytics. But in order to be able, an encryption declaration must provide means of controlling access to data granularity and the association of encryption keys, while maintaining a high stage of performance.

- Confidentiality: first and foremost via access control and, where necessary, via encryption.

2) **Protection: action trails, necessary for safety, have a new business value.**
Two technical developments are likely to mine value from action trails. Firstly, at present that vast quantity of storage are obtainable, action trails for systems and users can now be formed on a enormous level to know which processes or users are importing data, where did they arrive from, which process or user is consulting that data.

The development of logical tools not only means that these trails include new business value, but also makes them working by the most intelligent safety management system.

3) Supervision: extending security management to all mechanism, to create them smarter

In recent years, we have seen the appearance of higher inexorable threats under attacks information systems, with the main plan of extracting data which the attacker can use to make value.

In this circumstance, it is clearly necessary to make sure that security actions are in place about the data. Since new threats that may be able to evade these measures, only monitoring these events is just a substance of hygiene.

Today, to commence these threats efficiently and more proactively, organizations require using a second-generation safety Operation Center. This efficiently allows all trails formed by the system to be monitored, interrelated and analyzed, so as to identify still the weakest of signals which might indicate that a hard attack is in development.

In this area, the analytical technologies enabled by Big Data give resources that total range of security procedures, by manufacturing management smarter and so more well-organized beside specific threats.

As the amount of data being composed continues to grow, more and more companies are building big data repositories to hoard, cumulative and extract meaning from their data. Big data provides a huge spirited advantage for corporations, helping businesses amend their products to customer requirements, be familiar with and reduce corporate inefficiencies and share data with user groups across the enterprise.

Unfortunately, lawful organizations are not the only groups that are going big. Large sets of consolidated data are striking for cyber attackers. Breaching organizations big data stock can
give unlawful groups with bigger payoffs and more gratitude from a single attack. And when attackers set their sights on big data repositories, the effects can be worrying for the affected organizations. Terabytes of data in these repositories may include a company’s crown jewel—customer data, employee data, and trade secrets.

Securing big data comes with its own unique challenges further than being a high-value aim. It is not that big data security is basically different from traditional data safety. Big data security challenges occur because of incremental differences, not elementary ones. The differences between big data environments and traditional data environments include:

- The data collected, aggregated, and analyzed for big data analysis
- The shipping used to hoard and house big data
- The technologies applied to examine structured and unstructured big data

The diversity, velocity and volume of big data augment security administration challenges that are addressed in traditional security organization. Big data repositories will likely include information deposited by various sources across the enterprise. This variety of data makes protected access management a challenge. Each data source will predictable its own access limitations and security policies, making it difficult to balance suitable security for all data sources require combining and extracting meaning from the data. For example, a big data environment may contain a dataset with proprietary research information, a dataset requiring stiff compliance, and a split dataset with personally certain information. In addition, many of the repositories assemble data at high volumes and velocity from a number of different data sources, and they all might have their own data move workflows. These dealings to multiple repositories can increase the attack outside for a challenger. A big data system receiving feeds from different data sources could present an attacker with possible vectors to attempt to gain access to a group.