Big Data Analytics

Data, which is available in abundance, can be streamlined for growth and expansion in technology as well as business. When data is analyzed successfully, it can become the answer to one of the most important questions: how can businesses acquire more customers and gain business insight? The key to this problem lies in being able to source, link, understand, and analyze data.

Figure 1.6 highlights the proportion of business areas that have benefited by using Big Data:

Big Data Analytics Benefit	Proportion of Businesses Reporting Benefit (%)
Better social influences marketing	61%
More accurate business insights	45%
Segmentation of customer base	41%
Identifying sales and market opportunities	38%
Automated decisions for real-time processes	37%
Detection of fraud	33%
Quantification of risks	30%
Better planning and forecasting	29%
Identifying cost drivers	29%

Source: TDWI July 2013

Figure 1.6: Big Data Benefit Areas

Let us understand some common analytical approaches that businesses apply to use Big Data.

Table 1.6 describes various analytical approaches typically associated with Big Data:

Table 1.6: Analytical Approaches

Approach	Possible Evaluations
Predictive Analysis	❑ How can a business use the available data for predictive and real-time analysis across its different domains? ❑ How can a business avail benefits from the unstructured enterprise data? ❑ How can a business leverage upon new types of data such as sentiment data, social media, clickstream, and multimedia?
Behavioral Analysis	❑ How will a business leverage complex data in order to create new models for: • Driving business outcomes • Decreasing business costs • Driving innovation in business strategy • Improving overall customer satisfaction • Converting an audience to a customer
Data Interpretation	❑ What new business analyses can be estimated from the available data? ❑ Which data should be analyzed for new product innovation?

Advantages of Big Data Analytics

According to Atul Butte, Stanford, "Hiding within those mounds of data is knowledge that could change the life of a patient, or change the world." So, the real power of Big Data lies in its analysis. Processing, studying, and implementing the conclusions derived from the analysis of Big Data help you to collect accurate data, take timely and more informed strategic decisions, target the right set of audience and customers, increase benefits, and reduce wastage and costs.

The right analysis of the available data can improve major business processes in various ways. For example, in a manufacturing unit, data analytics can improve the functioning of the following processes:

❑ **Procurement**—To find out which suppliers are more efficient and cost-effective in delivering products on time

❑ **Product Development**—To draw insights on innovative product and service formats and designs for enhancing the development process and coming up with demanded products

❑ **Manufacturing**—To identify machinery and process variations that may be indicators of quality problems

❑ **Distribution**—To enhance supply chain activities and standardize optimal inventory levels vis-à-vis various external factors such as weather, holidays, economy, etc

❑ **Marketing**—To identify which marketing campaigns will be the most effective in driving and engaging customers and understanding customer behaviors and channel behaviours

❑ **Price Management**—To optimize prices based on the analysis of external factors

- **Merchandising**—To improve merchandise breakdown on the basis of current buying patterns and increase inventory levels and product interest insights on the basis of the basis of the analysis of various customer behaviors

- **Sales**—To optimize assignment of sales resources and accounts, product mix, and other operations

- **Store Operations**—To adjust inventory levels on the basis of predicted buying patterns, study of demographics, weather, key events, and other factors

- **Human Resources**—To find out the characteristics and behaviors of successful and effective employees, as well as other employee insights for managing talent better

Every business and industry today is affected by and benefitted from Big Data analytics in multiple ways.

A closer look at some specific industries will help you to understand the application of Big Data in these sectors.

Transportation

Big Data has greatly improved transportation services. The data containing traffic information is analyzed to identify traffic jam areas. Suitable steps can then be taken, on the basis of this analysis, to keep the traffic moving in such areas. Distributed sensors are installed in handheld devices, on the roads and on vehicles to provide real-time traffic information. This information is analyzed and disseminated to commuters and also to the traffic control authority.

Education

Big Data has transformed the modern day education processes through innovative approaches, such as e-learning for teachers to analyze the students' ability to comprehend and thus impart education effectively in accordance with each student's needs. The analysis is done by studying the responses to questions, recording the time consumed in attempting those questions, and analyzing other behavioral signals of the students. Big Data also assists in analyzing the requirements and finding easy and innovative ways of imparting education, especially distance learning over vast geographical areas.

Travel

The travel industry also uses Big Data to conduct business. It maintains complete details of all the customer records that are then analyzed to determine certain behavioral patterns in customers. For example, in the airline industry, Big Data is analyzed for identifying personal preferences or spotting which passengers like to have window seats for short-haul flights and aisle seats for long-haul flights. This helps airlines to offer the similar seats to customers when they make a fresh booking with the airways.

Big Data also helps airlines to track customers who regularly fly between specific routes so that they can make the right cross-sell and up-sell offers. Some airlines also apply analytics to pricing, inventory, and advertising for improving customer experiences, leading to more customer satisfaction, and hence, more business. Some airlines even go to the length of evaluating customers who tend to miss their flights. They try to help such customers by delaying the flights or booking them on another flight.

Government

Big Data has come to play an important role in almost all the undertaking and processes of government.

According to the UK free market, "the UK government could save up to £33 billion a year by using public Big Data more effectively." Analysis of Big Data promotes clarity and transparency in various government processes and helps in:

- Taking timely and informed decisions about various issues
- Identifying flaws and loopholes in processes and taking preventive or corrective measures on time
- Assessing the areas of improvement in various sectors such as education, health, defense, and research
- Using budgets more judiciously and reducing unnecessary wastage and costs
- Preventing fraudulent practices in various sectors

Healthcare

In healthcare, the pharmacy and medical device companies use Big Data to improve their research and development practices, while health insurance companies use it to determine patient-specific treatment therapy modes that promise the best results. Big Data also helps researchers to work towards eliminating healthcare-related challenges before they become real problems. Big Data helps doctors to analyze the requirement and medical history of every patient and provide individualistic services to them, depending on their medical condition.

Telecom

The mobile revolution and the Internet usage on mobile phones have led to a tremendous increase in the amount of data generated in the telecom sector. Managing this huge pool of data has almost become a challenge for the telecom industry. For example, in Europe, there is a compulsion on the telecom companies to keep data of their customers for at least six months and maximum up to two years. Now, all this collection, storage, and maintenance of data would just be a waste of time and resources unless we could derive any significant benefits from this data. Big Data analytics allows telecom industries to utilize this data for extracting meaningful information that could be used to gain crucial business insights that help industries in enhancing their performance, improving customer services, maintaining their hold on the market, and generating more business opportunities.

Consumer Goods Industry

Consumer goods companies generate huge volumes of data in varied formats from different sources, such as transactions, billing details, feedback forms, etc. This data needs to be organized and analyzed in a systemic manner in order to derive any meaningful information from it. For example, the data generated from the Point-of-Sale (POS) systems provides significant real-time information about customers' preferences, current market trends, the increase and decrease in demand of different products at different regions, etc. This information helps organizations to predict any possible fluctuations in prices of goods and make purchases accordingly. It also helps marketing teams in taking suitable actions rapidly if there is a deviation in the expected sales of a product, thus,

17

preventing any further losses to the company. Therefore, we can say that Big Data analytics allows organizations to gain better business insights and take informed and timely decisions.

Aviation Industry

Big Data analytics also plays a significant role in the commercial aviation industry. Like other industries, the aviation industry also maintains a detailed record of all their customers that includes their personal information, flying preferences, and other trends and patterns. The organization analyzes this data to improve their customer services, and thus brand image. In addition, every aircraft generates a significant amount of data during operation. This data is then analyzed for enhancing operational efficiencies, identifying the parts that require repairs, and taking any necessary constructive or preventive measures on time.

Careers in Big Data

Now that you know that Big Data is really BIG in today's world, you can well understand that so are the opportunities associated with it. The market today needs plenty of talented and qualified people who can use their expertise to help organizations deal with Big Data.

Qualified and experienced Big Data professionals must have a blend of technical expertise, creative and analytical thinking, and communication skills to be able to effectively collate, clean, analyze, and present information extracted from Big Data.

Most jobs in Big Data are from companies that can be categorized into the following four broad buckets:

❑ Big Data technology drivers, e.g. Google, IBM, Salesforce

❑ Big Data product companies, e.g. Oracle

❑ Big Data services companies, e.g. EMC

❑ Big Data analytics companies, e.g. Splunk

Figure 1.7 shows the logos of some companies that hire Big Data professionals:

TOP COMPANIES HIRING BIG DATA PROFESSIONALS

| MITRE The MITRE Corporation | Google Google | intuit simplify the business of life Intuit | salesforce Salesforce | GE GE |
| Qualcomm Qualcomm | NetApp NetApp | FACTSET FactSet | Adobe Adobe | Apple Apple |

Source: Glassdoor Report October 2011

Figure 1.7: Companies Hiring Big Data Professionals

As shown in Figure 1.7, companies such as Google, Salesforce, and Apple offer various types of opportunities to Big Data professionals. These companies deal into various domains such as retail, manufacturing, information, finance, and consumer electronics. The hiring of Big Data experts in these domains, as per Big Data Analytics 2014 report, is shown in Figure 1.8:

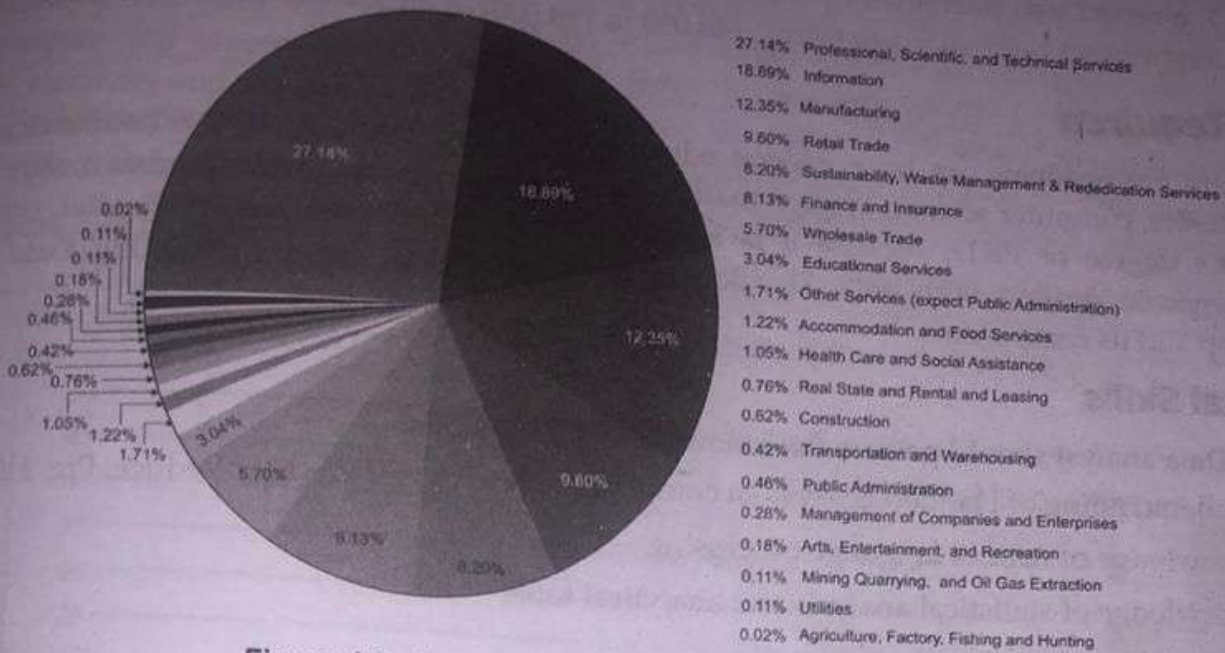Top 20 Industries Hiring Big Data Expertise
Source: Wanted Analytics, 2014

%	Industry
27.14%	Professional, Scientific, and Technical Services
18.89%	Information
12.35%	Manufacturing
9.60%	Retail Trade
8.20%	Sustainability, Waste Management & Rededication Services
8.13%	Finance and Insurance
5.70%	Wholesale Trade
3.04%	Educational Services
1.71%	Other Services (expect Public Administration)
1.22%	Accommodation and Food Services
1.05%	Health Care and Social Assistance
0.76%	Real State and Rental and Leasing
0.62%	Construction
0.42%	Transportation and Warehousing
0.46%	Public Administration
0.28%	Management of Companies and Enterprises
0.18%	Arts, Entertainment, and Recreation
0.11%	Mining Quarrying, and Oil Gas Extraction
0.11%	Utilities
0.02%	Agriculture, Factory, Fishing and Hunting

Figure 1.8: Top 20 Industries Hiring Big Data Experts

The most common job titles in Big Data include:

- Big Data analyst
- Data scientist
- Big Data developer
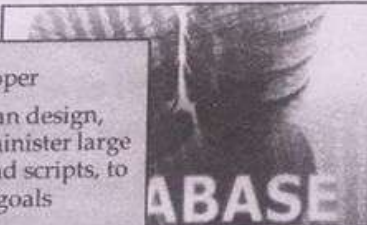- Big Data administrator
- Big Data engineer

Figure 1.9 illustrates the roles of these profiles:

Big Data Analyst
A well-trained professional who is able to collect data from different sources, organize it in a suitable format, and analyze the data to generate desired results.

Big Data Scientist
A thought leader who can collect and analyze data from various angles, determine what it indicates, and recommend ways to apply the data. It needs a combination of insight and technology expertise.

Big Data Developer
A programmer who can design, create, manage, and administer large datasets, custom tools and scripts, to achieve business goals

Big Data Administrator
An administrator who is responsible for planning and coordinating in issues regarding system upgrades, management of data warehouse, allocation of workload, and storage.

Big Data Engineer
A professional who can design and develeop applications using various frameworks and tools.

Figure 1.9: Role of Different Job Titles in Data Analytics

19

In 2011, a report was published by McKinsey & Co which indicated that by 2018, the United States alone might face a huge shortage (about 140,000 to 190,000) of data analytic professionals.

Skills Required

Big Data professionals can have various educational backgrounds, such as econometrics, physics, biostatistics, computer science, applied mathematics, or engineering. Data scientists mostly possess a master's degree or Ph.D. because it is a senior position and often achieved after considerable experience in dealing with data. Developers generally prefer implementing Big Data by using Hadoop and its components.

Technical Skills

A Big Data analyst should possess the following technical skills:

- Understanding of Hadoop ecosystem components, such as HDFS, MapReduce, Pig, Hive, etc.
- Knowledge of natural language processing
- Knowledge of statistical analysis and analytical tools
- Knowledge of machine learning
- Knowledge of conceptual and predictive modeling

A Big Data developer should possess the following skills:

- Programming skills in Java, Hadoop, Hive, HBase, and HQL
- Understanding of HDFS and MapReduce
- Knowledge of ZooKeeper, Flume, and Sqoop

These skills can be acquired with proper training and practice. This book familiarizes you with the technical skills required by a Big Data analyst and Big Data developer.

Soft Skills

Organizations look for professionals who possess good logical and analytical skills, with good communication skills and an affinity toward strategic business thinking.

The preferred soft skills requirements for a Big Data professional are:

- Strong written and verbal communication skills
- Analytical ability
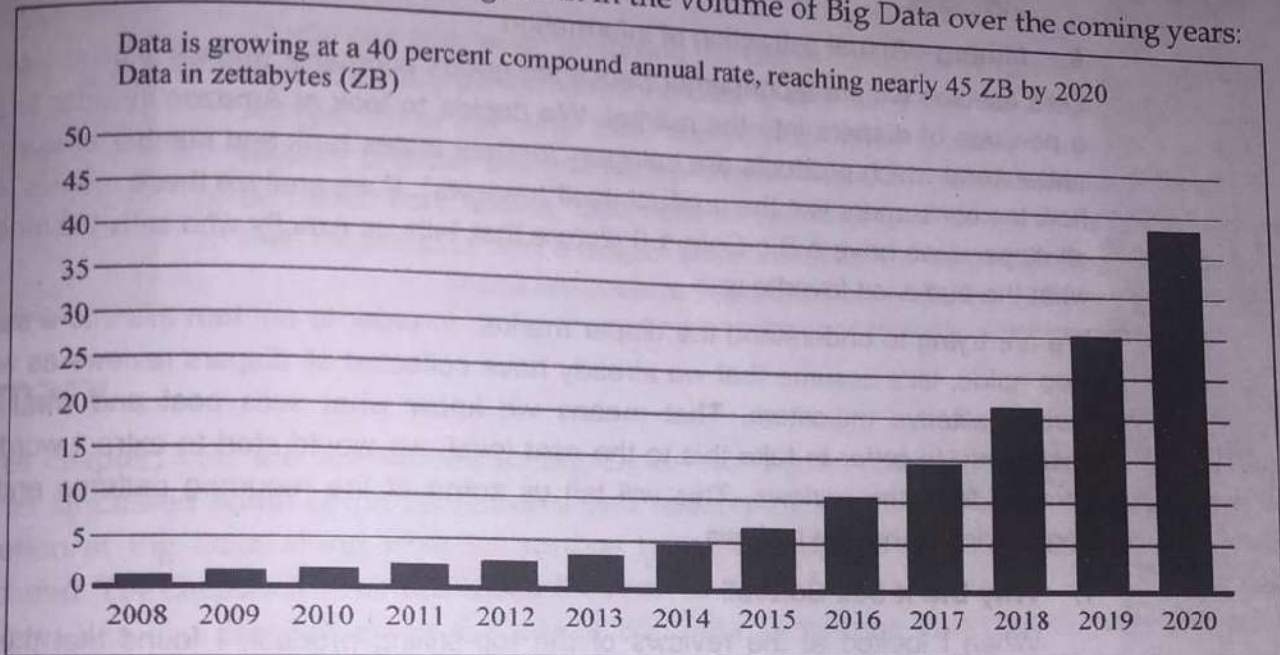- Basic understanding of how a business works

Future of Big Data

The need for Big Data is evident. If leaders and economies want exemplary growth and wish to generate value for all their stakeholders, Big Data has to be embraced and used extensively to:

- Allow the storage and use of transactional data in digital form
- Provide more and specific information
- Refine analytics that can improve decision making
- Classify customers for providing customized products and services based on buying patterns

Most organizations today consider data and information to be their most valuable and differentiated asset. By analyzing this data effectively, organizations worldwide are now finding new ways to compete and emerge as leaders in their fields to improve decision making and enhance their productivity and performance. At the same time, the volume and variety of data is also increasing at an immense rate every day. The global phenomena of using Big Data to gain business value and competitive advantage will only continue to grow as will the opportunities associated with it.

Figure 1.10 depicts the tremendous growth in the volume of Big Data over the coming years:

Data is growing at a 40 percent compound annual rate, reaching nearly 45 ZB by 2020
Data in zettabytes (ZB)

Source: *Oracle, 2012*

Figure 1.10: Growth Pattern of Data

Research conducted by MGI and McKinsey's Business Technology Office suggests that the use of Big Data is most likely to become a key basis of competition for individual firms for success and growth and strengthening consumer surplus, production growth, and innovation.