

R.M.K GROUP OF ENGINEERING INSTITUTIONS

R.M.K GROUP OF INSTITUTIONS





Please read this disclaimer before proceeding:

This document is confidential and intended solely for the educational purpose of RMK Group of Educational Institutions. If you have received this document through email in error, please notify the system manager. This document contains proprietary information and is intended only to the respective group / learning community as intended. If you are not the addressee you should not disseminate, distribute or copy through e-mail. Please notify the sender immediately by e-mail if you have received this document by mistake and delete this document from your system. If you are not the intended recipient you are notified that disclosing, copying, distributing or taking any action in reliance on the contents of this information is strictly prohibited.

22AI901

Business Intelligence and Analytics

Batch/Year: 2022-2026/II

Created by:

Ms. S. Subi, Assistant Professor

Ms. J. Bhuvaneswari, Assistant Professor

Table of Contents

S.NO	Topic	Page No.
1.	Contents	5
2.	Course Objectives	6
3.	Pre-Requisites	7
4.	Syllabus	8
5.	Course outcomes	10
6.	CO- PO/PSO Mapping	11
7.	Lecture Plan	12
8.	Activity based learning	13
9.	Lecture notes	23
10.	Assignments	70
11.	Part A Q & A	71
12.	Part B Qs	80
13.	Supportive online Certification courses	81
14.	Real time Applications in day to day life and to Industry	82
15.	Assessment Schedule	84
16.	Prescribed Text Books and Reference Books	85
17.	Mini Project Suggestions	86

2. Course Objectives

- To understand the business intelligence (BI) methodology and concepts.
- To learn about descriptive, inferential statistics and data warehousing operations.
- To analyze wide range of applications of data mining.
- To analyze the various prescriptive analytics methods.
- To develop and deploy Business Analytic Models.

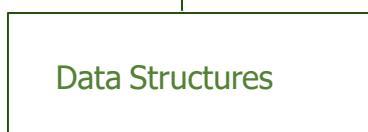


3. Pre-Requisites

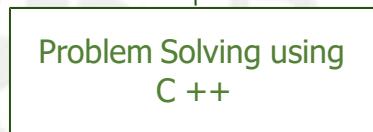
Semester-III



Semester-II



Semester-I



4.SYLLABUS

22AI901

BUSINESS INTELLIGENCE AND ANALYTICS

L T P C 2 0 2 3

UNIT I OVERVIEW OF BUSINESS INTELLIGENCE

6+6

Evolution of Computerized Decision Support to Analytics- A Framework for Business Intelligence - Analytics Overview - Analytics Examples- Introduction to Big Data Analytics- Overview of the Analytics Ecosystem.

List of Exercise/Experiments:

1. Perform Customer Segmentation, Classification using customer data of a certain organization. Analyze the data from the standpoint of paying capacity and purchasing pattern similarities among the company's clients.
2. Build a data model by taking an available data for a certain company and create a series of analysis and visualizations on various metrics related to the products of that company.

UNIT II DESCRIPTIVE ANALYTICS

6+6

The Nature of Data- Data Preprocessing- Statistical Modeling for Business Analytics- Regression Modeling for Inferential Statistics- Business Reporting- Data Visualization- Types of charts and graphs- Visual Analytics- Information Dashboards- Business Intelligence and Data Warehousing- Data Warehousing Process - Data Warehousing architecture - Data Integration and the Extraction, Transformation, and Load (ETL) Processes- Data Warehouse Development.

List of Exercise/Experiments:

1. Consider Groceries dataset for Market Basket Analysis and investigate customer's historical transactions. Focus on descriptive analytics of customer's purchase behavior, revealing interesting combinations of products that are frequently bought together, and creating valuable suggestions for the company.
2. Given Life Expectancy (WHO) dataset that provides information on both life expectancy and GDP per capita by year for different countries and regions, Explore and visualize the data using appropriate plots, and develop meaningful insights.

4.SYLLABUS

UNIT III PREDICTIVE ANALYTICS

6+6

Data Mining Concepts – Data Mining Process – Data Mining Methods - Text Analytics and Text Mining – NLP – Applications – Process – Sentiment Analysis – Web Mining – Search Engines – Web Analytics – Social Analytics.

List of Exercise/Experiments:

- 1.Perform Customer Review Sentiment Analysis with text data extracted from customer reviews of a certain company and explore it using specialized statistical and linguistic tools to identify positive, negative, and neutral experiences and their strength and subjectivity.
- 2.Using Microsoft Stock Data/Amazon Stock Data or INTEL Stock Data, Explore the company's historical stock performance and find insights about the future.

UNIT IV PRESCRIPTIVE ANALYTICS

6+6

Model-based Decision Making – Structure of Mathematical Models for Decision Support – Certainty, Uncertainty and Risk – Decision Modelling – Multiple Goals, Sensitivity Analysis, WhatIf Analysis and Goal Seeking – Decision Analysis – Introduction to Simulation – Location-based Analytics for Organizations – Impacts of Analytics in Organization. Case study: prepare a detailed report on applications of analytics in different industries.

List of Exercise/Experiments:

- 1.Perform Retail Price Optimization using dataset of price data for a retail company containing information such as product names, historical prices, product categories and characteristics, volume of sales, and time and geographic notations. Calculate the optimal selling prices for the products to create efficient, data-driven recommendations for the company.
- 2.Perform Credit Card Fraud Detection using online transactions dataset and analyze it for suspicious operations using statistical methods.

5. COURSE OUTCOMES

At the end of this course, the students will be able to:

COURSE OUTCOMES		K Level
CO1	Understand the business intelligence (BI) methodology and concepts.	K2
CO2	Learn about descriptive, inferential statistics and data warehousing operations.	K2
CO3	Analyze wide range of applications of data mining.	K3
CO4	Analyze the various prescriptive analytics methods.	K3
CO5	Develop and deploy Business Analytic Models.	K4



6.CO – PO /PSO Mapping Matrix

CO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PO O1	PO O2	PO O3	
1	3	2	1		1			1	1	1			1	2	3	3
2	3	2	1		3			1	1	1			1	2	3	3
3	3	2	1		3			3	3	3			3	2	3	3
4	3	3	2		3			3	3	3			3	2	3	3
5	3	2	2		3			3	3	3			3	2	3	3



7. Lecture Plan – Unit 1

Sl. No.	Topic	Number of Periods	Proposed Date	Actual Lecture Date	CO	Taxonomy Level	Mode of Delivery
1	Evolution of Computerized Decision Support to Analytics	1	03.01.2024	03.01.2024	CO1	K3	PPT/Chalk & Talk
2	A Framework for Business Intelligence - Analytics Overview	1	05.01.2024	05.01.2024	CO1	K3	PPT/Chalk & Talk
3	Analytics Examples-	1	06.01.2024	06.01.2024	CO1	K3	PPT/Chalk & Talk
4	Analytics Examples-	1	06.01.2024	06.01.2024	CO1	K3	PPT/Chalk & Talk
5	Introduction to Big Data Analytics	1	10.01.2024	10.01.2024	CO1	K3	PPT/Chalk & Talk
6	Overview of the Analytics Ecosystem	1	12.01.2024	12.01.2024	CO1	K3	PPT/Chalk & Talk

8. ACTIVITY BASED LEARNING

Case Study 1: Sabre Helps Its Clients Through Dashboards and Analytics

Sabre is one of the world leaders in the travel industry, providing both business-to-consumer services as well as business-to-business services. It serves travelers, travel agents, corporations, and travel suppliers through its four main companies: Travelocity, Sabre Travel Network, Sabre Airline Solutions, and Sabre Hospitality Solutions. The current volatile global economic environment poses significant competitive challenges to the airline industry. To stay ahead of the competition, Sabre Airline Solutions recognized that airline executives needed enhanced tools for managing their business decisions by eliminating the traditional, manual, time-consuming process of aggregating financial and other information needed for actionable initiatives. This enables real-time decision support at airlines throughout the world to maximize their (and in turn Sabre's) return on information by driving insights, actionable intelligence, and value for customers from the growing data.

Sabre developed an Enterprise Travel Data Warehouse (ETDW) using Teradata to hold its massive reservations data. ETDW is updated in near real time with batches that run every 15 minutes, gathering data from all of Sabre's businesses. Sabre uses its ETDW to create Sabre Executive Dashboards that provide near real-time executive insights using a Cognos BI platform with Oracle Data Integrator and Oracle Goldengate technology infrastructures. The Executive Dashboards offer their client airlines' top-level managers and decision makers a timely, automated, user-friendly solution, aggregating critical performance metrics in a succinct way and providing at a glance a 360-degree view of the overall health of the airline.

At one airline, Sabre's Executive Dashboards provide senior management with a daily and intraday snapshot of key performance indicators in a single application replacing the once-a-week, 8-hour process of generating the same report from various data sources. The use of dashboards is not limited to the external customers; Sabre also uses them for their assessment of internal operational performance. The dashboards help Sabre's customers to have a clear understanding of the data through the visual displays that incorporate interactive drill-down capabilities. It replaces flat presentations and allows for a more focused review of the data with less effort and time. This facilitates team dialog by making the data/metrics pertaining to sales performance available to many stakeholders, including ticketing, seats sold and flown, operational performance including the data on flight movement and tracking, customer reservations, inventory, and revenue across an airline's multiple distribution channels. The dashboard systems provide scalable infrastructure, graphical user interface support, data integration, and aggregation that empower airline executives to be more proactive in taking actions that lead to positive impacts on the overall health of their airline. With its ETDW, Sabre could also develop other Web-based analytical and reporting solutions that leverage data to gain customer insights through analysis of customer profiles and their sales interactions to calculate customer value. This enables better customer segmentation and insights for value-added services.

Questions for Discussion

1. What is traditional reporting? How is it used in the organization?
2. How can analytics be used to transform the traditional reporting?
3. How can interactive reporting assist organizations in decision making?

CAse Study 2: A Multimedia Exercise in Business Intelligence

TUN includes videos (similar to the television show CSI) to illustrate concepts of analytics in different industries. These are called “BSI Videos (Business Scenario Investigations).” Not only are these entertaining, but they also provide the class with some questions for discussion. For starters, please go to <http://www.teradatauniversitynetwork.com/Library/Items/BSI-The-Case-of-the-Misconnecting-Passengers/> or www.youtube.com/watch?v=NXEL5F4_aKA. Watch the video that appears on YouTube. Essentially, you have to assume the role of a customer service center professional. An incoming flight is running late, and several passengers are likely to miss their connecting flights. There are seats on one outgoing flight that can accommodate two of the four passengers. Which two passengers should be given priority? You are given information about customers' profiles and relationships with the airline. Your decisions might change as you learn more about those customers' profiles. Watch the video, pause it as appropriate, and answer the questions on which passengers should be given priority. Then resume the video to get more information. After the video is complete, you can see the slides related to this video and how the analysis was prepared on a slide set at www.slideshare.net/teradata/bsi-how-we-did-it-the-case-of-the-misconnecting-passengers.

This multimedia excursion provides an example of how additional available information through an enterprise DW can assist in decision making. Although some people equate DSS with BI, these systems are not, at present, the same. It is interesting to note that some people believe that DSS is a part of BI—one of its analytical tools. Others think that BI is a special case of DSS that deals mostly with reporting, communication, and collaboration (a form of data-oriented DSS). Another explanation (Watson, 2005) is that BI is a result of a continuous revolution, and as such, DSS is one of BI's original elements. Further, as noted in the next section onward, in many circles BI has been subsumed by the new terms analytics or data science.

Case Study 3: Siemens Reduces Cost with the Use of Data Visualization

Siemens is a German company headquartered in Berlin, Germany. It is one of the world's largest companies focusing on the areas of electrification, automation, and digitalization. It has an annual revenue of 76 billion euros. The visual analytics group of Siemens is tasked with end-to-end reporting solutions and consulting for all of Siemens internal BI needs. This group was facing the challenge of providing reporting solutions to the entire Siemens organization across different departments while maintaining a balance between governance and self-service capabilities. Siemens needed a platform that could analyze their multiple cases of customer satisfaction surveys, logistic processes, and financial reporting. This platform should

be easy to use for their employees so that they can use this data for analysis and decision making. In addition, the platform should be easily integrated with existing Siemens systems and give employees a seamless user experience. They started using Dundas BI, a leading global provider of BI and data visualization solutions. It allowed Siemens to create highly interactive dashboards that enabled Siemens to detect issues early and thus save a significant amount of money. The dashboards developed by Dundas BI helped Siemens global logistics organization answer questions like how different supply rates at different locations affect the operation, thus helping them to reduce cycle time by 12% and scrap cost by 25%.

Questions for Discussion

1. What challenges were faced by Siemens visual analytics group?
2. How did the data visualization tool Dundas BI help Siemens in reducing cost?

Case Study 4: Analyzing Athletic Injuries

Any athletic activity is prone to injuries. If the injuries are not handled properly, then the team suffers. Using analytics to understand injuries can help in deriving valuable insights that would enable coaches and team doctors to manage the team composition, understand player profiles, and ultimately aid in better decision making concerning which players might be available to play at any given time. In an exploratory study, Oklahoma State University analyzed American football-related sports injuries by using reporting and predictive analytics. The project followed the CRISP-DM methodology to understand the problem of making recommendations on managing injuries, understanding the various data elements collected about injuries, cleaning the data, developing visualizations to draw various inferences, building PMs to analyze the injury healing time period, and drawing sequence rules to predict the relationships among the injuries and the various body part parts afflicted with injuries. The injury data set consisted of more than 560 football injury records, which were categorized into injury-specific variables—body part/site/laterality, action taken, severity, injury type, injury start and healing dates—and player/sport-specific variables—player ID, position played, activity, onset, and game location. Healing time was calculated for each record, which was classified into different sets of time periods: 0–1 month, 1–2 months, 2–4 months, 4–6 months, and 6–24 months. Various visualizations were built to draw inferences from injury data set information depicting the healing time period associated with players' positions, severity of injuries and the healing time period, treatment offered and the associated healing time period, major injuries afflicting body parts, and so forth. Neural network models were built to predict each of the healing categories using IBM SPSS Modeler. Some of the predictor variables were current status of injury, severity, body part, body site, type of injury, activity, event location, action taken, and position played.

The success of classifying the healing category was quite good: Accuracy was 79.6% percent. Based on the analysis, many busi_x0002_ness recommendations were suggested, including employing more specialists' input from injury onset instead of letting the training room staff screen the injured players; training players at defensive posi_x0002_tions to avoid being injured; and holding practice to thoroughly safety-check mechanisms.

Questions for Discussion

- 1.What types of analytics are applied in the injury analysis?
- 2.How do visualizations aid in understanding the data and delivering insights into the data?
3. What is a classification problem?
4. What can be derived by performing sequence analysis?



Case Study 5: Silvaris Increases Business with Visual Analysis and Real-Time Reporting Capabilities

Silvaris Corporation was founded in 2000 by a team of forest industry professionals to provide technological advancement in the lumber and building material sector. Silvaris is the first e-commerce platform in the United States specifically for forest products and is headquartered in Seattle, Washington. It is a leading wholesale provider of industrial wood products and surplus building materials. Silvaris sells its products and provides international logistics services to more than 3,500 customers. To manage various processes that are involved in a transaction, they created a proprietary online trading platform to track information flow related to transactions between traders, accounting, credit, and logistics. This allowed Silvaris to share its real-time information with its customers and partners. But due to the rapidly changing prices of materials, it became necessary for Silvaris to get a real-time view of data without moving data into a separate reporting format. Silvaris started using Tableau because of its ability to connect with and visualize live data. Due to dashboards created by Tableau that are easy to understand and explain, Silvaris started using Tableau for reporting purposes. This helped Silvaris in pulling out information quickly from the data and identifying issues that impact their business. Silvaris succeeded in managing online versus offline orders with the help of reports generated by Tableau. Now, Silvaris keeps track of online orders placed by customers and knows when to send renew pushes to which customers to keep them purchasing online. Also, analysts of Silvaris can save time by generating dashboards instead of writing hundreds of pages of reports by using Tableau.

Questions for Discussion

1. What was the challenge faced by Silvaris?
2. How did Silvaris solve its problem using data visualization with Tableau?

Case Study 6: A Specialty Steel Bar Company Uses Analytics to Determine Available-to-Promise Dates

This application case is based on a project that includes one of us. A company that does not wish to disclose its name (or even the precise industry) was facing a major problem of making decisions on which inventory of raw materials to use to satisfy which customers. This company supplies custom configured steel bars to its customers. These bars may be cut into specific shapes or sizes and may have unique material and finishing requirements. The company procures raw materials from around the world and stores them in its warehouse. When a prospective customer calls the company to request a quote for the specialty bars meeting specific material requirements (composition, origin of the metal,

quality, shapes, sizes, etc.), the salesperson usually has just a little bit of time to submit such a quote including the date when the product can be delivered and, of course, prices, and so on. It must make available- to- promise (ATP) decisions, which determine in real time the dates when it can promise delivery of products that customers requested during the quotation stage. Previously, a salesperson had to make such decisions by analyzing reports on available inventory of raw materials. Some of the available raw material may have already been committed to another customer's order. Thus the inventory in stock may not really be the free inventory available. On the other hand, there may be raw material that is expected to be delivered in the near future that could also be used for satisfying the order from this prospective customer. Finally, there might

even be an opportunity to charge a premium for a new order by repurposing previously committed inventory to satisfy this new order while delaying an already committed order. Of course, such decisions should be based on the cost–benefit analyses of delaying a previous order. The system should thus be able to pull real-time data about inventory, committed orders, incoming raw material, production constraints, and so on. To support these ATP decisions, a real-time DSS was developed to find an optimal assignment of the

available inventory and to support additional what-if analysis. The DSS uses a suite of mixed-integer programming models that are solved using commercial software.

The company has incorporated the DSS into its enterprise resource planning system to seamlessly facilitate its use of business analytics.

Questions for Discussion

1. Why would reallocation of inventory from one customer to another be a major issue for discussion?
2. How could a DSS help make these decisions?



Case Study 7: CenterPoint Energy Uses Real-Time Big Data Analytics to Improve Customer Service

CenterPoint Energy is a Fortune 500 energy delivery company based in Houston, Texas. Its primary business includes electric transmission and distribution, natural gas distribution, and natural gas sales and service. It has over five million metered customers in the United States. CenterPoint Energy uses smart grids to collect real-time information about the health of various aspects of the grid like meters, transformers, and switches that are used in providing electricity. This real-time power usage information is analyzed with Big Data analytics and allows for a much quicker

diagnosis and solution. For example, the data can predict and potentially help prevent a power outage. In addition, the tool collects weather information allowing historical data to help predict the magnitude of an outage from a storm. This insight will act as a guide for putting the right resources out before a storm occurs to avoid an outage. Second, to better understand their customers, CenterPoint Energy utilizes sentiment analysis, which examines a customer's opinion by way of emotion (happiness, anger, sadness, etc.). The company segments their customers based on the sentiment and is able to market to these groups in a more personalized way, providing a more valuable customer service experience. As a result of using Big Data analytics, CenterPoint Energy has saved 600,000 gallons of fuel in the last 2 years by resolving six million service requests remotely. In addition, they have saved \$24 million for their customers in this process.

Questions for Discussion

1. How can electric companies predict a possible outage at a location?
2. What is customer sentiment analysis?
3. How does customer sentiment analysis help companies provide a personalized service to their customers?

9. LECTURE NOTES UNIT - I

OVERVIEW OF BUSINESS INTELLIGENCE

The business environment (climate) is constantly changing, and it is becoming more and more complex. Organizations, both private and public, are under pressures that force them to respond quickly to changing conditions and to be innovative in the way they operate. Such activities require organizations to be agile and to make frequent and quick strategic, tactical, and operational decisions, some of which are very complex. Making such decisions may require considerable amounts of relevant data, information, and knowledge. Processing these, in the framework of the needed decisions, must be done quickly, frequently in real time, and usually requires some computerized support.

Evolution of Computerized Decision Support to Analytics/Data Science

The timeline in Figure 1.8 shows the terminology used to describe analytics since the 1970s. During the 1970s, the primary focus of information systems support for decision making focused on providing structured, periodic reports that a manager could use for decision making (or ignore them).

Businesses began to create routine reports to inform decision makers (managers) about what had happened in the previous period (e.g., day, week, month, quarter). Although it was useful to know what had happened in the past, managers needed more than this: They needed a variety of reports at different levels of granularity to better understand and address changing needs and challenges of the business.

These were usually called management information systems (MIS). In the early 1970s, Scott-Morton first articulated the major concepts of DSS. He defined DSSs as “interactive computer-based systems, which help decision makers utilize data and models to solve unstructured problems” (Gorry and

Scott-Morton, 1971). The following is another classic DSS definition, provided by Keen and Scott-Morton (1978):

Decision support systems couple the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions. It is a computer-based support system for management decision makers who deal with semistructured problems.

In the late 1970s and early 1980s, in addition to the mature OR models that were being used in many industries and government systems, a new and exciting line of models had emerged: rule-based expert systems.

These systems promised to capture experts' knowledge in a format that computers could process (via a collection of if–then–else rules or heuristics) so that these could be used for consultation much the same way that one would use domain experts to identify a structured problem and to prescribe the most probable solution. ESs allowed scarce expertise to be made available where and when needed, using an "intelligent" DSS.

The 1980s saw a significant change in the way organizations captured business related data. The old practice had been to have multiple disjointed information systems tailored to capture transactional data of different organizational units or functions (e.g., accounting, marketing and sales, finance, manufacturing). In the 1980s, these systems were integrated as enterprise-level information systems that we now commonly call enterprise resource planning (ERP) systems.

The old mostly sequential and nonstandardized data representation schemas were replaced by relational database management (RDBM) systems. These systems made it possible to improve the capture and storage of data, as well as the relationships between organizational data fields while significantly reducing the replication of information. The need for RDBM and ERP systems emerged when data integrity and consistency became an issue, significantly hindering the effectiveness of business practices.

With ERP, all the data from every corner of the enterprise is collected and integrated into a consistent schema so that every part of the organization has

access to the single version of the truth when and where needed. In addition to the emergence of ERP systems, or perhaps because of these systems, business reporting became an on demand, as-needed business practice. Decision makers could decide when they needed to or wanted to create specialized reports to investigate organizational problems and opportunities.

In the 1990s, the need for more versatile reporting led to the development of executive information systems (EISs; DSSs designed and developed specifically for executives and their decision-making needs). These systems were designed as graphical dashboards and scorecards so that they could serve as visually appealing displays while focusing on the most important factors for decision makers to keep track of the key performance indicators.

To make this highly versatile reporting possible while keeping the transactional integrity of the business information systems intact, it was necessary to create a middle data tier known as a DW as a repository to specifically support business reporting and decision making. In a very short time, most large to medium-sized businesses adopted data warehousing as their platform for enterprise-wide decision making.

The dashboards and scorecards got their data from a DW, and by doing so, they were not hindering the efficiency of the business transaction systems mostly referred to as (ERP) systems.

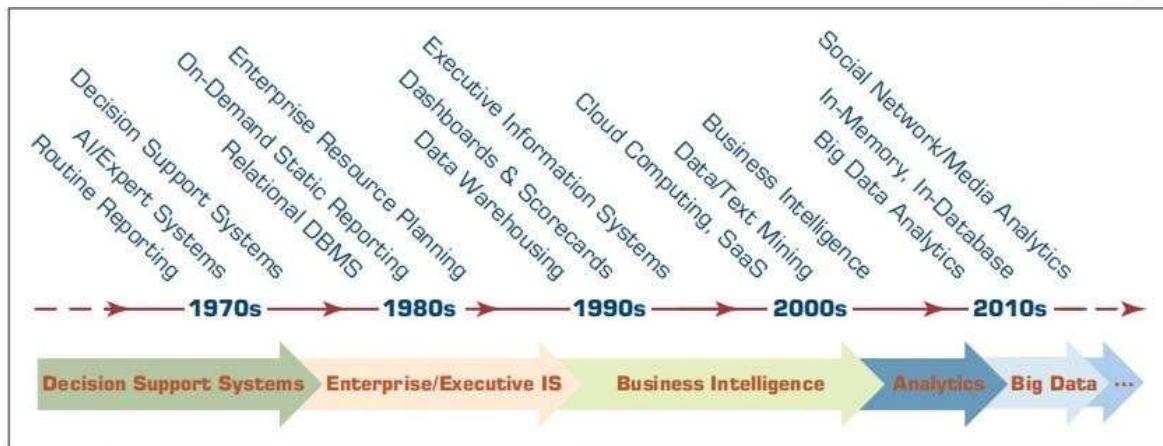


FIGURE 1.8 Evolution of Decision Support, Business Intelligence, and Analytics.

In the 2000s, the DW-driven DSSs began to be called BI systems. As the amount of longitudinal data accumulated in the DWs increased, so did the capabilities of hardware and software to keep up with the rapidly changing and evolving needs of the decision makers. Because of the globalized competitive marketplace, decision makers needed current information in a very digestible format to address business problems and to take advantage of market opportunities in a timely manner. Because the data in a DW is updated periodically, it does not reflect the latest information. To elevate this information latency problem, DW vendors developed a system to update the data more frequently, which led to the terms real-time data warehousing and, more realistically, right-time data warehousing, which differs from the former by adopting a data-refreshing policy based on the needed freshness of the data items (i.e., not all data items need to be refreshed in real time). DWs are very large and feature rich, and it became necessary to “mine” the corporate data to “discover” new and useful knowledge nuggets to improve business processes and practices, hence the terms data mining and text mining. With the increasing volumes and varieties of data, the needs for more storage and more processing power emerged.

Although large corporations had the means to tackle this problem, small to medium sized companies needed more financially manageable business models. This need led to service-oriented architecture and software and infrastructure-as-a-service analytics business models. Smaller companies, therefore, gained access to analytics capabilities on an as-needed basis and paid only for what they used, as opposed to investing in financially prohibitive hardware and software resources.

In the 2010s, we are seeing yet another paradigm shift in the way that data is captured and used. Largely because of the widespread use of the Internet, new data generation mediums have emerged. Of all the new data sources (e.g., radio-frequency identification [RFID] tags, digital energy meters, clickstream Web logs, smart home devices, wearable health monitoring equipment), perhaps the most interesting and challenging is social networking/social media. This unstructured data is rich in information content, but analysis of such data sources poses significant challenges to computational systems, from both software and hardware perspectives. Recently, the term Big Data has been coined to highlight the challenges that these new data streams have brought on us.

Many advancements in both hardware (e.g., massively parallel processing with very large computational memory and highly parallel multiprocessor computing systems) and software/algorithms (e.g., Hadoop with MapReduce and NoSQL) have been developed to address the challenges of Big Data.

It's hard to predict what the next decade will bring and what the new analytics related terms will be. The time between new paradigm shifts in information systems and particularly in analytics has been shrinking, and this trend will continue for the foreseeable future.

Even though analytics is not new, the explosion in its popularity is very new. Thanks to the recent explosion in Big Data, ways to collect and store this data, and intuitive software tools, data-driven insights are more accessible to business professionals than ever before.

Therefore, in the midst of global competition, there is a huge opportunity to make better managerial decisions by using data and analytics to increase revenue while decreasing costs by building better products, improving customer experience, and catching fraud before it happens, improving customer engagement through targeting and customization all with the power of analytics and data.

More and more companies are now preparing their employees with the know-how of business analytics to drive effectiveness and efficiency in their day-to-day decision-making processes.

A Framework for Business Intelligence

Definitions of BI

Business intelligence (BI) is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies.

Objectives of BI

- To enable **interactive access** to data
- To enable **manipulation of data**, and
- To give business managers and analysts the **ability to conduct appropriate analyses**. **To conduct appropriate analysis by experts**

By analyzing historical and current data, situations, and performances, decision makers get valuable insights that enable them to make more informed and better decisions.

The process of BI is based on the transformation of data to information, then to decisions, and finally to actions.

A Brief History of BI

The term BI was coined by the **Gartner Group** in the mid-1990s. However, as the history in the previous section points out, the concept is much older; it has its roots in the MIS reporting systems of the 1970s. During that period, reporting systems were **static, were two dimensional, and had no**

analytical capabilities. In the early 1980s, the concept of EISs emerged. This concept expanded the computerized support to top-level managers and executives.

Some of the capabilities introduced were dynamic multidimensional (ad hoc or on-demand) reporting, forecasting and prediction, trend analysis, drill-down to details, status access, and critical success factors. These features appeared in dozens of commercial products until the mid-1990s. Then the same capabilities and some new ones appeared under the name BI. Today, a good BI-based enterprise information system contains all the information executives need. So, the original concept of EIS was transformed into BI. By 2005, BI systems started to include artificial intelligence capabilities as well as powerful analytical capabilities. Figure 1.9 illustrates the various tools and techniques that may be included in a BI system. It illustrates the evolution of BI as well. The tools shown in Figure 1.9 provide the capabilities of BI. The most sophisticated BI products include most of these capabilities; others specialize in only some of them.

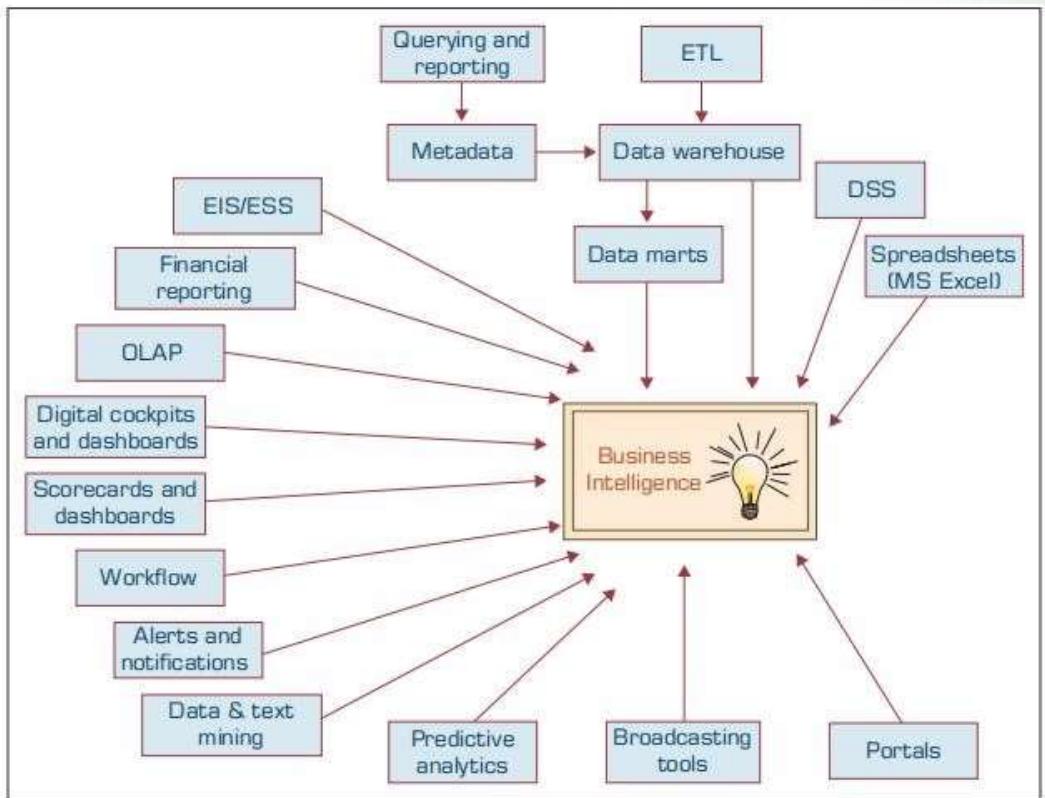


FIGURE 1.9 Evolution of Business Intelligence (BI).

The Architecture of BI

A BI system has four major components: a Data Warehouse, with its source data; business analytics, a collection of tools for manipulating, mining, and analyzing the data in the DW; BPM (Business Process Management) for monitoring and analyzing performance; and a user interface (e.g., a dashboard). Business Process Management (BPM) strategy involves defining, designing, executing, monitoring, and optimizing business processes to increase efficiency, effectiveness, and agility. The relationship among these components is illustrated in Figure 1.10.

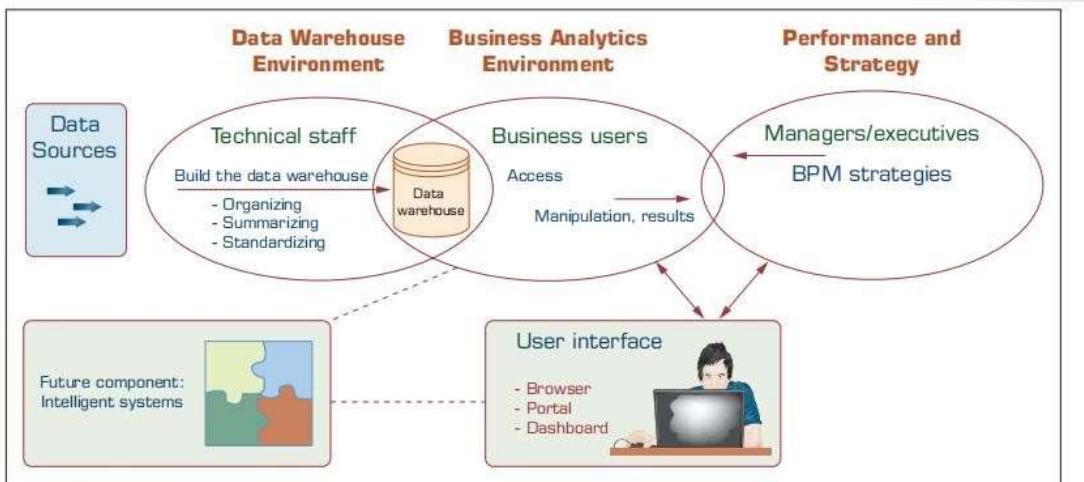


FIGURE I.10 A High-Level Architecture of BI. (Source: Based on W. Eckerson, *Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligent Solutions*. The Data Warehousing Institute, Seattle, WA, 2003, p. 32, Illustration 5.)

The Origins and Drivers of BI

Where did modern approaches to data warehousing and BI come from? What are their roots, and how do those roots affect the way organizations are managing these initiatives today?

- Organizations are being compelled to capture, understand, and harness their data to support decision making to improve business operations.
- Legislation and regulation now require business leaders to document their business processes and to sign off on the legitimacy of the information they rely on and report to stakeholders.
- Business cycle times are now extremely compressed; faster, more informed, and better decision making is, therefore, a competitive imperative.
- Managers need the right information at the right time and in the right place. This is the mantra for modern approaches to BI.

Transaction Processing versus Analytic Processing

These transaction processing systems are constantly involved in handling updates to what we might call operational databases.

For example, in an ATM withdrawal transaction, we need to reduce our bank balance accordingly; a bank deposit adds to an account; and a grocery store purchase is likely reflected in the store's calculation of total sales for the day, and it should reflect an appropriate reduction in the store's inventory for the items we bought, and so on. These online transaction processing (OLTP) systems handle a company's routine ongoing business.

In contrast, a DW is typically a distinct system that provides storage for data that will be used for analysis. The intent of that analysis is to give management the ability to scour data for information about the business, and it can be used to provide tactical or operational decision support.

DWs are intended to work with informational data used for online analytical processing (OLAP) systems. Most operational data in enterprise resources planning (ERP) systems—and in its complementary siblings like supply chain management (SCM) or CRM—are stored in an OLTP system, which is a type of computer processing where the computer responds immediately to user requests. Each request is considered to be a transaction, which is a computerized record of a discrete event, such as the receipt of inventory or a customer order. In other words, a transaction requires a set of two or more database updates that must be completed in an all-or-nothing fashion.

The very design that makes an OLTP system efficient for transaction processing makes it inefficient for end-user ad hoc reports, queries, and analysis.

All requests for reports had to be programmed by the IT staff, whereas only "precanned" reports could be generated on a scheduled basis, and ad hoc real-time querying was virtually impossible.

DWs contain a wide variety of data that present a coherent picture of business conditions at a single point in time. The idea was to create a database infrastructure that was always online and contained all the information from the OLTP systems, including historical data, but reorganized and structured in such a way that it was fast and efficient for querying, analysis, and decision support.

Appropriate Planning and Alignment with the Business Strategy

First and foremost, the fundamental reasons for investing in BI must be aligned with the company's business strategy.

BI has to serve as a way to change the manner in which the company conducts business by improving its business processes and transforming decision-making processes to be more data driven.

Many BI consultants and practitioners involved in successful BI initiatives advise that a framework for planning is a necessary precondition. One framework, developed by Gartner, Inc. (2004), decomposes planning and execution into business, organization, functionality, and infrastructure components.

At the business and organizational levels, strategic and operational objectives must be defined while considering the available organizational skills to achieve those objectives. Issues of organizational culture surrounding BI initiatives and building enthusiasm for those initiatives and procedures for the intra-organizational sharing of BI best practices must be considered by upper management—with plans in place to prepare the organization for change.

BI Competency Center

The center could serve some or all of the following functions (Gartner, 2004): The center can demonstrate how BI is clearly linked to strategy and execution of strategy.

- A center can serve to encourage interaction between the potential business user communities and the organization.

- The center can serve as a repository and disseminator of best BI practices between and among the different lines of business.
- Standards of excellence in BI practices can be advocated and encouraged throughout the company.
- The organization can learn a great deal through interaction with the user communities, such as knowledge about the variety of types of analytical tools that are needed.
- The business user community and organization can better understand why the DW platform must be flexible enough to provide for changing business requirements.
- It can help important stakeholders like high-level executives see how BI can play an important role.

Real-Time, On-Demand BI Is Attainable

The demand for instant, on-demand access to dispersed information has grown as the need to close the gap between the operational data and strategic objectives has become more pressing. As a result, a category of products called real-time BI applications has emerged.

The introduction of new data-generating technologies, such as RFID and other sensors is only accelerating this growth and the subsequent need for real-time BI.

Traditional BI systems use a large volume of static data that has been extracted, cleansed, and loaded into a DW to produce reports and analyses. However, the need is not just reporting because users need business monitoring, performance analysis, and an understanding of why things are happening.

In addition, business applications can be programmed to act on what these real-time BI systems discover.

For example, an SCM application might automatically place an order for more “widgets” when real-time inventory falls below a certain threshold or when a CRM (Customer Relationship Management) application automatically triggers a customer service representative and credit control clerk to check a

customer who has placed an online order larger than \$10,000. One approach to real-time BI uses the DW model of traditional BI systems.

A second approach, commonly called business activity management (BAM). It bypasses the DW entirely and uses Web services or other monitoring means to discover key business events. These software monitors (or intelligent agents) can be placed on a separate server in the network or on the transactional application databases themselves, and they can use event- and process-based approaches to proactively and intelligently measure and monitor operational processes.

Developing or Acquiring BI Systems

Today, many vendors offer diversified tools, some of which are completely preprogrammed (called shells); These tools can be purchased or leased. The issue that companies face is which alternative to select: purchase, lease, or build. Each of these alternatives has several options. One of the major criteria for making the decision is justification and cost–benefit analysis.

As the number of potential BI applications increases, the need to justify and prioritize them arises. This is not an easy task due to the large number of intangible benefits. Both direct and intangible benefits need to be identified. Of course, this is where the knowledge of similar applications in other organizations and case studies is extremely useful.

For example, The Data Warehousing Institute (tdwi.org) provides a wealth of information about products and innovative applications and implementations. Such information can be useful in estimating direct and indirect benefits.

Security and Protection of Privacy

This is an extremely important issue in the development of any computerized system, especially BI that contains data that may possess strategic value. Also, the privacy of employees and customers needs to be protected.

Integration of Systems and Applications

With the exception of some small applications, all BI applications must be integrated with other systems such as databases, legacy systems, enterprise systems (particularly ERP and CRM), e-commerce (sell side, buy side), and many more. In addition, BI applications are usually connected to the Internet and many times to information systems of business partners.

Analytics Overview

Analytics is viewed as the process of developing actionable decisions or recommendations for actions based on insights generated from historical data.

According to the Institute for Operations Research and Management Science (INFORMS), analytics represents the combination of computer technology, management science techniques, and statistics to solve real problems.

The idea of looking at all the data to understand what is happening, what will happen, and how to make the best of it has also been encapsulated by INFORMS in proposing three levels of analytics. These three levels are identified (informs.org/Community/Analytics) as descriptive, predictive, and prescriptive.

Figure 1.11 presents a graphical view of these three levels of analytics. It suggests that these three are somewhat independent steps and one type of analytics applications leads to another. It also suggests that there is actually some overlap across these three types of analytics. In either case, the interconnected nature of different types of analytics applications is evident.

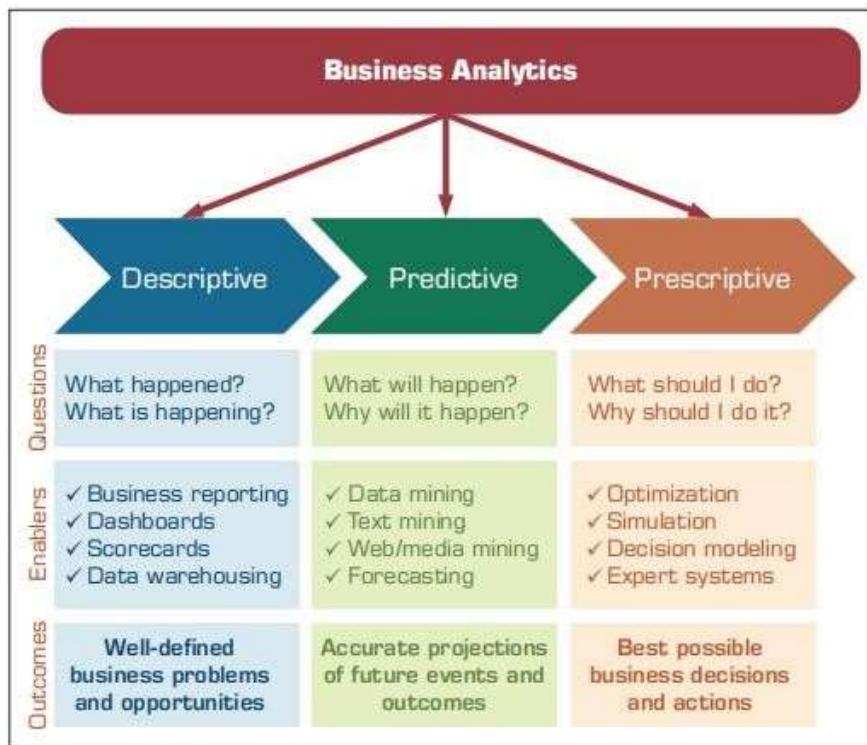


FIGURE 1.11 Three Types of Analytics.

Descriptive Analytics

Descriptive (or reporting) analytics refers to knowing what is happening in the organization and understanding some underlying trends and causes of such occurrences.

First, this involves the consolidation of data sources and availability of all relevant data in a form that enables appropriate reporting and analysis. Usually, the development of this data infrastructure is part of DWs. From this data infrastructure we can develop appropriate reports, queries, alerts, and trends using various reporting tools and techniques.

A significant technology that has become a key player in this area is visualization. Using the latest visualization tools in the marketplace, we can now develop powerful insights in the operations of our organization.

Predictive Analytics

Predictive analytics aims to determine what is likely to happen in the future. This analysis is based on statistical techniques as well as other more recently developed techniques that fall under the general category of data mining.

The goal of these techniques is to be able to predict if the customer is likely to switch to a competitor ("churn"), what the customer would likely buy next and how much, what promotions a customer would respond to, whether this customer is a creditworthy risk, and so forth.

A number of techniques are used in developing predictive analytical applications, including various classification algorithms. we can use classification techniques such as logistic regression, decision tree models, and neural networks. We can also use clustering algorithms for segmenting customers into different clusters to be able to target specific promotions to them.

Finally, we can use association mining techniques to estimate relationships between different purchasing behaviors. That is, if a customer buys one product, what else is the customer likely to purchase? Such analysis can assist a retailer in recommending or promoting related products.

For example, any product search on Amazon.com results in the retailer also suggesting other similar products that a customer may be interested in.

Prescriptive Analytics

The goal of prescriptive analytics is to recognize what is going on as well as the likely forecast and make decisions to achieve the best performance possible. This group of techniques has historically been studied under the umbrella of OR or management sciences and are generally aimed at optimizing the performance of a system.

The goal here is to provide a decision or a recommendation for a specific action. These recommendations can be in the form of a specific yes/no decision for a problem, a specific amount (say, price for a specific item or airfare to charge), or a complete set of production plans. The decisions

may be presented to a decision maker in a report or may be used directly in an automated decision rules system (e.g., in airline pricing systems). Thus, these types of analytics can also be termed decision or normative analytics.

Analytics Applied to Different Domains

Applications of analytics in various industry sectors have spawned many related areas or at least buzzwords. It is almost fashionable to attach the word analytics to any specific industry or type of data. Besides the general category of text analytics—aimed at getting value out of text —or Web analytics—analyzing Web data streams —many industry- or problem-specific analytics professions/ streams have been developed.

Examples of such areas are marketing analytics, retail analytics, fraud analytics, transportation analytics, health analytics, sports analytics, talent analytics, behavioral analytics, and so forth.

Any systematic analysis of data in a specific sector is being labeled as “(fill-in-blanks)” analytics. Although this may result in overselling the concept of analytics, the benefit is that more people in specific industries are aware of the power and potential of analytics.

It also provides a focus to professionals developing and applying the concepts of analytics in a vertical sector. Although many of the techniques to develop analytics applications may be common, there are unique issues within each vertical segment that influence how the data may be collected, processed, analyzed, and the applications implemented. Thus, the differentiation of analytics based on a vertical focus is good for the overall growth of the discipline.

Analytics or Data Science?

Even as the concept of analytics is receiving more attention another term has already been introduced and is becoming popular. The new term is data science. Thus, the practitioners of data science are data scientists. D. J. Patil of LinkedIn is sometimes credited with creating the term data science. There have been some attempts to describe the differences between data

analysts and data scientists (e.g., see emc.com/collateral/about/news/emc-data-science-study-wp.pdf). One view is that data analyst is just another term for professionals who were doing BI in the form of data compilation, cleaning, reporting, and perhaps some visualization.

Their skill sets included Excel, some SQL knowledge, and reporting. In contrast, a data scientist is responsible for predictive analysis, statistical analysis, and more advanced analytical tools and algorithms. They may have a deeper knowledge of algorithms and may recognize them under various labels—data mining, knowledge discovery, or machine learning.

Some of these professionals may also need deeper programming knowledge to be able to write code for data cleaning/analysis in current Web-oriented languages such as Java or Python and statistical languages such as R.

Many analytics professionals also need to build significant expertise in statistical modeling, experimentation, and analysis. Again, our readers should

recognize that these fall under the predictive and prescriptive analytics umbrella.

However, prescriptive analytics also includes more significant expertise in OR including optimization, simulation, decision analysis, and so on. Those who cover these fields are more likely to be called data scientists than analytics professionals. Our view is that the distinction between analytics and data scientist is more of a degree of technical knowledge and skill sets than functions.

It may also be more of a distinction across disciplines. Computer science, statistics, and applied mathematics programs appear to prefer the data science label, reserving the analytics label for more business-oriented professionals. As another example of this, applied physics professionals have proposed using network science as the term for describing analytics that relate to groups of people—social networks, supply chain networks, and so forth.

Aside from a clear difference in the skill sets of professionals who only have to do descriptive/reporting analytics versus those who engage in all three types of analytics, the distinction is fuzzy between the two labels, at best. We

observe that graduates of our analytics programs tend to be responsible for tasks which are more in line with data science professionals (as defined by some circles) than just reporting analytics.

Analytics Examples in Selected Domains

Analytics Applications in Healthcare—Humana Examples

Although healthcare analytics span a wide variety of applications from prevention to diagnosis to efficient operations and fraud prevention, we focus on some applications that have been developed at a major health insurance company, Humana. According to the company Web site, “The company’s strategy integrates care delivery, the member experience, and clinical and consumer insights to encourage engagement, behavior change, proactive clinical outreach and wellness. . . .” Achieving these strategic goals includes significant investments in information technology in general, and analytics in particular.

Brian LeClaire is senior vice president and CIO of Humana, a major health insurance provider in the United States. He has a PhD in MIS from Oklahoma State University. He has championed analytics as a competitive differentiator at Humana—including cosponsoring the creation of a center for excellence in analytics. He described the following projects as examples of Humana’s analytics initiatives, led by Humana’s Chief Clinical Analytics Officer, Vipin Gopal.

Example 1: Preventing Falls in a Senior Population— An Analytic Approach

Accidental falls are a major health risk for adults age 65 years and older with one-third experiencing a fall every year.¹ Falls are also the leading factor for both fatal and nonfatal injuries in older adults, with injurious falls increasing the risk of disability by up to 50%.² The costs of falls pose a significant strain on the U.S. healthcare system, with the direct costs of falls estimated at \$34 billion in 2013 alone.¹ With the percent of seniors in the U.S. population on

the rise, falls and associated costs are anticipated to increase. According to the Centers for Disease Control and Prevention (CDC), "Falls are a public health problem that is largely preventable."

Humana is the nation's second-largest provider of Medicare Advantage benefits with approximately 3.2 million members, most of whom are seniors. Keeping their senior members well and helping them live safely at their homes is a key business objective, of which prevention of falls is an important component. However, no rigorous methodology was available to identify individuals most likely to fall, for whom falls prevention efforts would be beneficial.

Unlike chronic medical conditions such as diabetes and cancer, a fall is not a well-defined medical condition. In addition, falls are usually underreported in claims data as physicians typically tend to code the consequence of a fall such as fractures and dislocations. Although many clinically administered assessments to identify fallers exist, they have limited reach and lack sufficient predictive power. As such, there is a need for a prospective and accurate method to identify individuals at greatest risk of falling, so that they can be proactively managed for fall prevention. The Humana analytics team undertook the development of a Falls Predictive Model in this context. It is the first comprehensive PM reported that utilizes administrative medical and pharmacy claims, clinical data, temporal clinical patterns, consumer information, and other data to identify individuals at high risk of falling over a time horizon.

Today, the Falls PM is central to Humana's ability to identify seniors who could benefit from fall mitigation interventions. An initial proof-of-concept with Humana consumers, representing the top 2% of highest risk of falling, demonstrated that the consumers had increased utilization of physical therapy services, indicating consumers are taking active steps to reduce their risk for falls. A second initiative utilizes the Falls PM to identify high-risk individuals for remote monitoring programs. Using the PM, Humana was able to identify 20,000 consumers at a high risk of falls, who benefited from this program.

Identified consumers wear a device that detects falls and alerts a 24/7 service for immediate assistance. This work was recognized by the Analytics Leadership Award by Indiana University Kelly School of Business in 2015, for innovative adoption of analytics in a business environment.

Example 2: Humana's Bold Goal—Application of Analytics to Define the Right Metrics

In 2014, Humana, Inc. announced its organization's Bold Goal to improve the health of the communities it serves by 20% by 2020 by making it easy for people to achieve their best health. The communities that Humana serves can be defined in many ways, including geographically (state, city, neighborhood), by product (Medicare Advantage, employer-based plans, individually purchased), or by clinical profile (priority conditions including diabetes, hypertension, CHF [congestive heart failure], CAD [coronary artery disease], COPD [chronic obstructive pulmonary disease], or depression). Understanding the health of these communities and how they track over time is critical not only for the evaluation of the goal, but also in crafting strategies to improve the health of the whole membership in its entirety.

A challenge before the analytics organization was to identify a metric that captures the essence of the Bold Goal. Objectively measured traditional health insurance metrics such as hospital admissions or ER visits per 1,000 persons would not capture the spirit of this new mission. The goal was to identify a metric that captures health and its improvement in a community, but was also relevant to Humana as a business. Through rigorous analytic evaluations, Humana eventually selected "Healthy Days," a four-question, quality-of-life questionnaire originally developed by the CDC to track and measure their overall progress toward the Bold Goal. It was critical to make sure that the selected metric was highly correlated to health and business metrics, such that any improvement in Healthy Days resulted in improved

health and better business results. Some examples of how “Healthy Days” is correlated to metrics of interest include the following:

- Individuals with more unhealthy days (UHDs) exhibit higher utilization and cost patterns. For a 5-day increase in UHDs, there is (a) an \$82 increase in average monthly medical and pharmacy costs, (b) an increase of 52 inpatient admits per 1000 patients, and (c) a 0.28-day increase in average length of stay.
- Individuals who exhibit healthy behaviors and have their chronic conditions well managed have fewer UHDs. For example, when we look at individuals with diabetes, UHDs are lower if they obtained an LDL screening (-4.3 UHDs) or a diabetic eye exam (-2.3 UHDs). Likewise, if they have controlled blood sugar levels measured by HbA1C (-1.8 UHDs) or LDL levels (-1.3 UHDs).
- Individuals with chronic conditions have more UHDs than those who do not have: (a) CHF (16.9 UHDs), (b) CAD (14.4 UHDs), (c) hypertension (13.3 UHDs), (d) diabetes (14.7 UHDs), (e) COPD (17.4 UHDs), or (f) depression (22.4 UHDs).

Humana has since adopted Healthy Days as their metric for the measurement of progress toward Bold Goal.

Example 3: Predictive Models to Identify the Highest Risk Membership in a Health Insurer

The 80/20 rule generally applies in healthcare, that is, roughly 20% of consumers account for 80% of healthcare resources due to their deteriorating health and chronic conditions. Health insurers like Humana have typically enrolled the highest-risk enrollees in clinical and disease management programs to help manage the chronic conditions the members have. Identification of the right members is critical for this exercise, and in the recent years, PMs have been developed to identify enrollees with the high future risk.

Many of these PMs were developed with heavy reliance on medical claims data, which results from the medical services that the enrollees use

Because of the lag that exists in submitting and processing claims data, there is a corresponding lag in identification of high-risk members for clinical program enrollment. This issue is especially relevant when new members join a health insurer, as they would not have a claims history with an insurer.

A claims-based PM could take on average of 9–12 months after enrollment of new members to identify them for referral to clinical programs. In the early part of this decade, Humana attracted large numbers of new members in its Medicare Advantage products and needed a better way to clinically manage this membership. As such, it became extremely important that a different analytic approach be developed to rapidly and accurately identify high-risk new members for clinical management, to keep this group healthy and costs down.

Humana's Clinical Analytics team developed the New Member Predictive Model (NMPM) that would quickly identify at-risk individuals soon after their new plan enrollments with Humana, rather than waiting for sufficient claim history to become available for compiling clinical profiles and predicting future health risk.

Designed to address the unique challenges associated with new members, NMPM developed a novel approach that leveraged and integrated broader data sets beyond medical claims data such as self-reported health risk assessment data and early indicators from pharmacy data, employed advanced data mining techniques for pattern discovery, and scored every MA consumer daily based on the most recent data Humana has to date. The model was deployed with a cross-functional team of analytics, IT, and operations to ensure seamless operational and business integration.

Ever since NMPM was implemented in January 2013, it has been rapidly identifying high-risk new members for enrollment in Humana's clinical programs. The positive outcomes achieved through this model have been highlighted in multiple senior leader communications from Humana.

In the first quarter 2013 earnings release presentation to investors, Bruce Broussard, CEO of Humana, stated the significance of "improvement in

new member PMs and clinical assessment processes," which resulted in 31,000 new members enrolled in clinical programs, compared to 4,000 in the same period a year earlier, a 675% increase. In addition to the increased volume of clinical program enrollments, outcome studies showed that the newly enrolled consumers identified by NMPM were also referred to clinical programs sooner, with over 50% of the referrals identified within the first 3 months after new MA plan enrollments. The consumers identified also participated at a higher rate and had longer tenure in the programs.

Analytics in the Retail Value Chain

The retail sector is where you would perhaps see the most applications of analytics. This is the domain where the volumes are large but the margins are usually thin. Customers' tastes and preferences change frequently. Physical and online stores face many challenges in succeeding. And market dominance at one time does not guarantee continued success.

So investing in learning about your suppliers, customers, employees, and all the stakeholders that enable a retail value chain to succeed and using that information to make better decisions has been a goal of the analytics industry for a long time. Even casual readers of analytics probably know about Amazon's enormous investments in analytics to power their value chain. Similarly, Walmart, Target, and other major retailers have invested millions of dollars in analytics for their supply chains. Most of the analytics technology and service providers have a major presence in retail analytics. Coverage of even a small portion of those applications to achieve our expo sure goal could fill a whole book. So this section just highlights a few potential applications. Most of these have been fielded by many retailers and are available through many technology providers, so in this section we will take a more general view rather than point to specific cases. This general view has been proposed by Abhishek Rathi, CEO of vCreaTek.com. vCreaTek, LLC is a boutique analytics software and service company that has offices in India, the United States, the United

Arab Emirates (UAE), and Belgium. The company develops applications in multiple domains, but retail analytics is one of their key focus areas.

Figure 1.12 highlights selected components of a retail value chain. It starts with suppliers and concludes with customers, but illustrates many intermediate strategic and operational planning decision points where analytics—descriptive, predictive, or prescriptive—can play a role in making better data-driven decisions. Table 1.1 also illustrates some of the important areas of analytics applications, examples of key questions that can be answered through analytics, and of course, the potential business value derived from fielding such analytics.



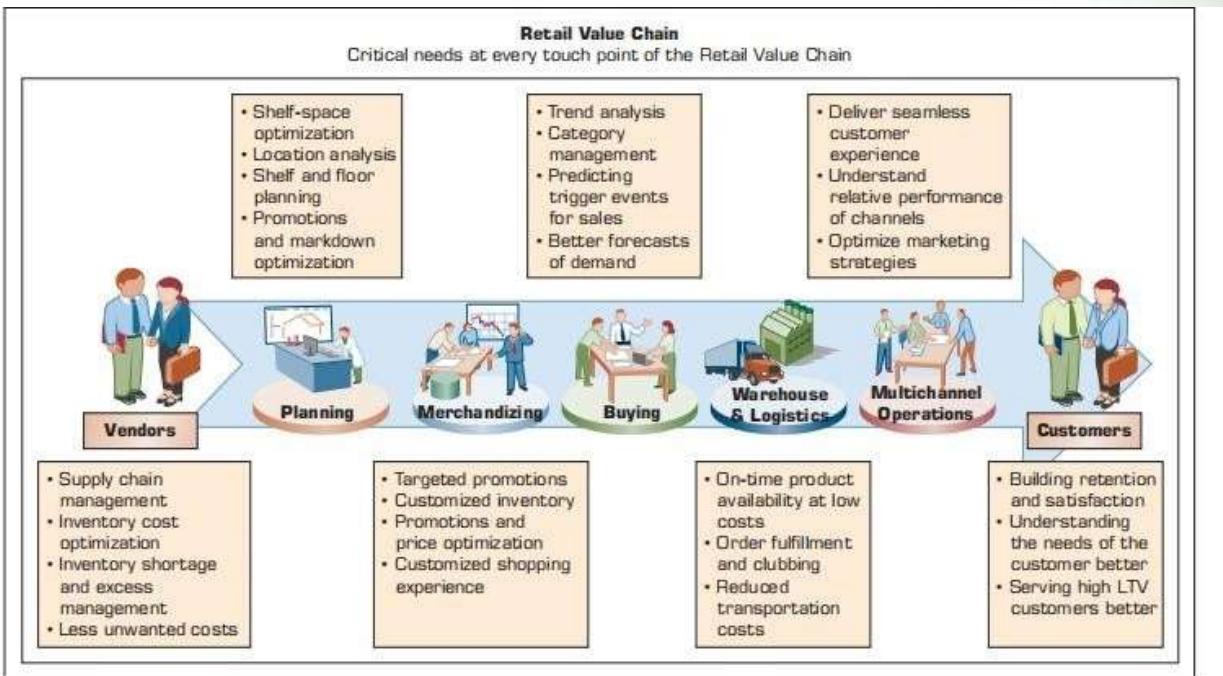


FIGURE I.12 Example of Analytics Applications in a Retail Value Chain. Contributed by Abhishek Rathi, CEO, vCreaTek.com



TABLE 1.1 Examples of Analytics Applications in the Retail Value Chain

Analytic Application	Business Question	Business Value
Inventory Optimization	1. Which products have high demand? 2. Which products are slow moving or becoming obsolete?	1. Forecast the consumption of fast-moving products and order them with sufficient inventory to avoid a stock-out scenario. 2. Perform fast inventory turnover of slow-moving products by combining them with one in high demand.
Price Elasticity	1. How much net margin do I have on the product? 2. How much discount can I give on this product?	1. Markdown prices for each product can be optimized to reduce the margin dollar loss. 2. Optimized price for the bundle of products is identified to save the margin dollar.
Market Basket Analysis	1. What products should I combine to create a bundle offer? 2. Should I combine products based on slow-moving and fast-moving characteristics? 3. Should I create a bundle from the same category or different category line?	1. The affinity analysis identifies the hidden correlations between the products, which can help in following values a) Strategize the product bundle offering based on focus on inventory or margin. b) Increase cross-sell or up-sell by creating bundle from different categories or the same categories, respectively.
Shopper Insight	1. Which customer is buying what product at what location?	1. By customer segmentation, the business owner can create personalized offers resulting in better customer experience and retention of the customer.
Churn Analysis	1. Who will not return? 2. How much business will I lose? 3. How can I retain them? 4. What demography of customer is my loyal customer?	relationships that are not working and show high churn. Thus can have better focus on product quality and reason for that churn. 2. Based on the customer lifetime value (LTV), the business can do targeted marketing resulting in retention of the customer.
Channel Analysis	1. Which channel has lower customer acquisition cost? 2. Which channel has better customer retention? 3. Which channel is more profitable?	1. Marketing budget can be optimized based on insight for better return on investment.
New Store Analysis	1. What location should I open? 2. What and how much opening inventory should I keep?	1. Best practices of other locations and channels can be used to get a jump start. 2. Comparison with competitor data can help to create a differentiator/USP factor to attract the new customers.
Store Layout	1. How should I do store layout for better topline? 2. How can I increase my in-store customer experience?	1. Understand the association of products to decide store layout and better alignment with customer needs. 2. Workforce deployment can be planned for better customer interactivity and thus satisfying
Video Analytics	1. What demography is entering the store during the peak period of sales? 2. How can I identify a customer with high LTV at the store entrance so that a better personalized experience can be provided to this customer?	1. In-store promotions and events can be planned based on the demography of incoming traffic. 2. Targeted customer engagement and instant discount enhances the customer experience resulting in higher retention.

An online retail site usually knows its customer as soon as the customer signs in, and thus they can offer customized pages/offering to enhance the experience. For any retail store, knowing its customer at the store entrance is still a huge challenge. By combining the video analytics and information/badge issued through their loyalty program, the store may be able to identify the customer at the entrance itself and thus enable an extra opportunity for a cross-sell or up-sell. Moreover, a personalized shopping experience can be provided with more customized engagement during the customer's time in the store.

Store retailers invest lots of money in attractive window displays, promotional events, customized graphics, store decorations, printed ads, and banners. To discern the effectiveness of these marketing methods, the team can use shopper analytics by observing closed circuit television (CCTV) images to figure out the demographic details of the in-store foot traffic. The CCTV images can be analyzed using advanced algorithms to derive demographic details such as age, gender, and mood of the person browsing through the store.

Further, the customer's in-store movement data when combined with shelf layout and planogram can give more insight to the store manager to identify the hot-selling/ profitable areas within the store. Moreover, the store manager can use this information to also plan the workforce allocation for those areas for peak periods.

Market basket analysis has commonly been used by the category managers to push the sale of the slowly moving SKUs. By using advanced analytics of data available, the product affinity can be done at the lowest level of SKU to drive better ROIs on the bundle offers. Moreover, by using price elasticity techniques, the markdown or optimum price of the bundle offer can also be deduced, thus reducing any loss in the profit margin.

Thus by using data analytics, a retailer can not only get information on its current operations but can also get further insight to increase the revenue and decrease the operational cost for higher profit. A fairly comprehensive list of current and potential retail analytics applications that a major retailer such as Amazon could use is proposed by a blogger at Data Science Central.

A Brief Introduction to Big Data Analytics

What Is Big Data?

Big Data is data that cannot be stored in a single storage unit. Big Data typically refers to data that comes in many different forms: structured, unstructured, in a stream, and so forth. Major sources of such data are clickstreams from Web sites, postings on social media sites such as Facebook, and data from traffic, sensors, or weather.

A Web search engine like Google needs to search and index billions of Web pages to give you relevant search results in a fraction of a second. Although this is not done in real time, generating an index of all the Web pages on the Internet is not an easy task. Luckily for Google, it was able to solve this problem. Among other tools, it has employed Big Data analytical techniques.

There are two aspects to managing data on this scale: storing and processing. If we could purchase an extremely expensive storage solution to store all this at one place on one unit, making this unit fault tolerant would involve a major expense.

An ingenious solution was proposed that involved storing this data in chunks on different machines connected by a network—putting a copy or two of this chunk in different locations on the network, both logically and physically.

It was originally used at Google (then called the Google File System) and later developed and released as an Apache project as the Hadoop Distributed File System (HDFS). However, storing this data is only half the problem. Data is worthless if it does not provide business value, and for it to provide business value, it has to be analyzed.

How can such vast amounts of data be analyzed? Passing all computation to one powerful computer does not work; this scale would create a huge overhead on such a powerful computer.

Another ingenious solution was proposed: Push computation to the data, instead of pushing data to a computing node. This was a new paradigm and gave rise to a whole new way of processing data.

This is what we know today as the MapReduce programming paradigm, which made processing Big Data a reality. MapReduce was originally developed at Google, and a subsequent version was released by the Apache project called Hadoop MapReduce.

Today, when we talk about storing, processing, or analyzing Big Data, HDFS and MapReduce are involved at some level. Other relevant standards and software solutions have been proposed. Although the major toolkit is available as an open source, several companies have been launched to provide training or specialized analytical hardware or software services in this space. Some examples are HortonWorks, Cloudera, and Teradata Aster.

Over the past few years, what was called Big Data changed more and more as Big Data applications appeared. The need to process data coming in at a rapid rate added velocity to the equation.

An example of fast data processing is algorithmic trading. This uses electronic platforms based on algorithms for trading shares on the financial market, which operates in microseconds. The need to process different kinds of data added variety to the equation.

Another example of a wide variety of data is sentiment analysis, which uses various forms of data from social media platforms and customer responses to gauge sentiments. Today, Big Data is associated with almost any kind of large data that has the characteristics of volume, velocity, and variety.

An Overview of the Analytics Ecosystem

Objectives:

The objective of this section is to identify various sectors of the analytics industry, provide a classification of different types of industry participants, and illustrate the types of opportunities that exist for analytics professionals. Eleven different types of players are identified in an analytics ecosystem.

A secondary purpose of understanding the analytics ecosystem for the BI professional is also to be aware of organizations and new offerings and opportunities in sectors allied with analytics.

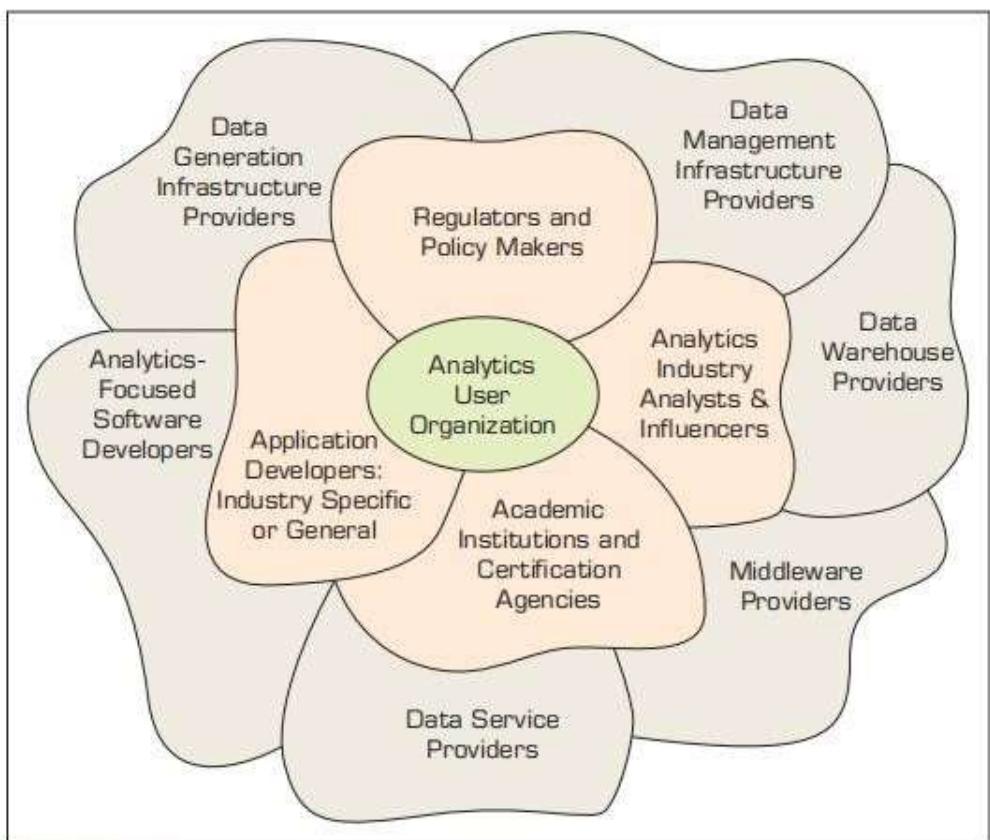


FIGURE I.13 Analytics Ecosystem.

Figure 1.13 illustrates one view of the analytics ecosystem. The components of the ecosystem are represented by the petals of an analytics flower. Eleven key sectors or clusters in the analytics space are identified. The components of the analytics ecosystem are grouped into three categories represented by the inner petals, outer petals, and the seed (middle part) of the flower.

The outer six petals can be broadly termed as the technology providers. Their primary revenue comes from providing technology, solutions, and training to analytics user organizations so they can employ these technologies in the most effective and efficient manner.

The inner petals can be generally defined as the analytics accelerators. The accelerators work with both technology providers and users.

Finally, the core of the ecosystem comprises the analytics user organizations. This is the most important component, as every analytics industry cluster is driven by the user organizations.

The main goal is to focus on the different analytic capabilities within each component of the analytics space. Many companies play in multiple sectors within the analytics industry and thus offer opportunities for movement within the field both horizontally and vertically.

Data Generation Infrastructure Providers

Perhaps the first place to begin identifying the clusters is by noting a new group of companies that enable generating and collection of data that may be used for developing analytical insights.

Although this group could include all the traditional point-of-sale systems, inventory management systems, and technology providers for every step in a company's supply/value chain and operations, we mainly consider new players where the primary focus has been on enabling an organization to develop new insights into its operations as opposed to running its core operations. Thus this group includes companies creating the infrastructure for collecting data from

different sources. This cluster is probably the most technical group in the ecosystem.

One of the emerging components of such an infrastructure is the "sensor." Sensors collect a massive amount of data at a faster rate and have been adopted by various sectors such as healthcare, sports, and energy.

Examples:

Health data collected by the sensors is generally used to track the health status of the users.

Some of the major players manufacturing sensors to collect health information are AliveCor, Google, Shimmer, and Fitbit. Likewise, the sports industry is using sensors to collect data from the players and field to develop strategies and improve team play. Examples of the companies producing sports related sensors include Sports Sensors, Zepp, Shockbox, and others.

Sensors are used for traffic management. These help in taking real-time actions to control traffic. Some of the providers are advantech B+B SmartWorx, Garmin, and Sensys Network.

Sensors play a major role in the Internet of Things and are an essential part of smart objects. These make machine-to-machine communication possible. The leading players in the infrastructure of IoT are Intel, Microsoft, Google, IBM, Cisco, Smartbin, SIKO Products, Omega Engineering, Apple, and SAP.

Data Management Infrastructure Providers

This group includes all of the major organizations that provide hardware and software targeting the basic foundation for all data management solutions.

Obvious examples of these include all major hardware players that provide the infrastructure for database computing—IBM, Dell, HP, Oracle, and so on.

Storage solution providers like EMC (recently bought by Dell) and NetApp;

Companies providing indigenous hardware and software platforms such as IBM, Oracle, and Teradata; and

Data solution providers offering hardware and platform independent database management systems like the SQL Server family of Microsoft and specialized integrated software providers such as SAP fall under this group.

This group also includes other organizations such as database appliance providers, service providers, integrators, developers, and so on, that support each of these companies' ecosystems.

Companies such as Amazon (Amazon Web Services), IBM (Bluemix), and Salesforce.com pioneered to offer full data storage and analytics solutions through the cloud.

Companies in Big data space such as Cloudera, Hortonworks, and many others do not necessarily offer their own hardware but provide infrastructure services and training to create the Big Data platform. This would include Hadoop clusters, MapReduce, NoSQL, Spark, Kafka, Flume, and other related technologies for analytics. Thus they could also be grouped under industry consultants or trainers enabling the basic infrastructure.

Full ecosystems of consultants, software integrators, training providers, and other value-added providers have evolved around many of the large players in the data management infrastructure cluster.

Data Warehouse Providers

This cluster is probably the most technical group in the ecosystem. Companies with a data warehousing focus provide technology and services aimed toward integrating data from multiple sources, thus enabling organizations to derive and deliver value from its data assets.

Many companies in this space include their own hardware to provide efficient data storage, retrieval, and processing. Companies such as IBM, Oracle, and Teradata are major players in this arena.

Recent developments in this space include performing analytics on the data directly in memory. Another major growth sector has been data warehousing in the cloud. Examples of such companies include Snowflake and Redshift.

Companies in this cluster clearly work with all the other sector players in providing DW solutions and services within their ecosystem and hence become the backbone of the analytics industry. It has been a major industry in its own right and, thus, a supplier and consumer of analytics talent.

Middleware Providers

Data warehousing began with a focus on bringing all the data stores into an enterprise wide platform. Making sense of this data has become an industry in itself.

The general goal of the middleware industry is to provide easy-to-use tools for reporting or descriptive analytics, which forms a core part of BI or analytics employed at organizations.

Examples of companies in this space include Microstrategy, Plum, and many others.

A few of the major players that were independent middleware players have been acquired by companies in the first two groups. For example, Hyperion became a part of Oracle, SAP acquired Business Objects, and IBM acquired Cognos.

This sector has been largely synonymous with the BI providers offering dashboarding, reporting, and visualization services to the industry, building on top of the transaction processing data and the database and DW providers.

Thus many companies have moved into this space over the years, including general analytics software vendors such as SAS or new visualization providers such as Tableau, or many niche application providers.

Data Service Providers

Much of the data an organization uses for analytics is generated internally through its operations, but there are many external data sources that play a major role in any organization's decision making. Examples of such data sources include demographic data, weather data, data collected by third parties that could inform an organization's decision making, and so on.

Several companies realized the opportunity to develop specialized data collection, aggregation, and distribution mechanisms. These companies typically focus on a specific industry sector and build on their existing relationships in that industry through their niche platforms and services for data collection.

For example, Nielsen provides data sources to their clients on customer retail purchasing behavior. Another example is Experian, which includes data on each household in the United States.

Omniture has developed technology to collect Web clicks and share such data with their clients. Comscore is another major company in this space. Google compiles data for individual Web sites and makes a summary available through Google Analytics services. Other examples are Equifax, TransUnion, Acxiom, Merkle, Epsilon, and Avention.

This can also include organizations such as ESRI.org, which provides location-oriented data to their customers. There are hundreds of other companies that are developing niche platforms and services to collect, aggregate, and share such data with their clients.

Many industry-specific data aggregators and distributors exist and are moving to offer their own analytics services. Thus this sector is also a growing user

and potential supplier of analytics talent, especially with specific niche expertise

Analytics-Focused Software Developers

Companies in this category have developed analytics software for general use with data that has been collected in a DW or is available through one of the platforms identified earlier (including Big Data).

It can also include inventors and researchers in universities and other organizations that have developed algorithms for specific types of analytics applications.

Major industry players in this space can be identified using the three types of analytics: descriptive, predictive, and prescriptive analytics.

Reporting/Descriptive Analytics

Reporting or descriptive analytics is enabled by the tools available from the middleware industry players identified earlier, or unique capabilities offered by focused providers. For example, Microsoft's SQL Server BI toolkit includes reporting as well as predictive analytics capabilities.

On the other hand, specialized software is available from companies such as Tableau for visualization. SAS also offers a Visual Analytics tool with similar capacity. There are many open source visualization tools as well. Literally hundreds of data visualization tools have been developed around the world, and many such tools focus on visualization of data from a specific industry or domain.

Because visualization is the primary way thus far for exploring analytics in industry, this sector has witnessed the most growth. Many new companies are being formed. For example, Gephi, a free and open source software, focuses on visualizing networks.

Predictive Analytics

Many statistical software companies such as SAS and SPSS embraced predictive analytics early on, and developed software capabilities as well as industry practices to employ data mining techniques and classical statistical techniques for analytics.

IBM-SPSS Modeler from IBM and Enterprise Miner from SAS are some of the examples of tools used for predictive analytics. Other players in this space include KXEN, Statsoft (recently acquired by Dell), Salford Systems, and scores of other companies that may sell their software broadly or use it for their own consulting practices (next group of companies).

Three open source platforms (R, RapidMiner, and KNIME) have also emerged as popular industrial-strength software tools for predictive analytics and have companies that support training and implementation of these open source tools.

Revolution Analytics is an example of a company focused on R development and training. R integration is possible with most analytics software.

A company called Alteryx uses R extensions for reporting and predictive analytics, but its strength is in shared delivery of analytics solutions processes to customers and other users. Similarly, RapidMiner and KNIME are also examples of open source providers.

Companies like Rulequest that sell proprietary variants of Decision Tree software and NeuroDimensions, a Neural Network software company, are examples of companies that have developed specialized software around a specific technique of data mining.

Prescriptive Analytics

Software providers in this category offer modeling tools and algorithms for optimization of operations usually called management science/operations research software.

This field has had its own set of major software providers. IBM, for example, has classic linear and mixed integer programming software. IBM also acquired a company called ILOG, which provides prescriptive analysis software and services to complement their other offerings.

Analytics providers such as SAS have their own OR/MS tools—SAS/OR. FICO acquired another company called XPRESS that offers optimization software. Other major players in this domain include companies such as AIIMS, AMPL, Frontline, GAMS, Gurobi, Lindo Systems, Maximal, NGData, Ayata, and many others.

Of course, there are many techniques that fall under the category of prescriptive analytics, and each has their own set of providers.

For example, simulation software is provided by major companies like Rockwell (ARENA) and Simio. Palisade provides tools that include many software categories. Similarly, Frontline offers tools for optimization with Excel spreadsheets, as well as predictive analytics.

Decision analysis in multiobjective settings can be performed using tools such as Expert Choice. There are also tools from companies such as Exsys, XpertRule, and others for generating rules directly from data or expert inputs.

Some new companies are evolving to combine multiple analytics models in the Big Data space including social network analysis and stream mining.

For example, Teradata Aster includes its own predictive and prescriptive analytics capabilities in processing Big Data streams. Several companies have developed complex event processing (CEP) engines that make decisions using streaming data, such as IBM's Infosphere Streams, Microsoft's StreamInsight, and Oracle's Event Processor.

Other major companies that have CEP products include Apache, Tibco, Informatica, SAP, and Hitachi.

It is worthwhile to note again that the provider groups for all three categories of analytics are not mutually exclusive. In most cases, a provider can play in multiple components of analytics.

Application Developers Industry Specific or General

The organizations in this group use their industry knowledge, analytical expertise, solutions available from the data infrastructure, DW, middleware, data aggregators, and analytics software providers to develop custom solutions for a specific industry. Thus, this industry group makes it possible for analytics technology to be used in a specific industry.

Most major analytics technology providers like IBM, SAS, and Teradata clearly recognize the opportunity to connect to a specific industry or client and offer analytic consulting services.

Companies that have traditionally provided application/data solutions to specific sectors are now developing industry-specific analytics offerings.

For example, Cerner provides electronic medical records solutions to medical providers, and their offerings now include many analytics reports and visualizations.

Similarly, IBM offers a fraud detection engine for the health insurance industry, and is working with an insurance company to employ their famous Watson analytics platform in assisting medical providers and insurance companies with diagnosis and disease management. Another example of a vertical application provider is Sabre Technologies, which provides analytical solutions to the travel industry including fare pricing for revenue optimization and dispatch planning.

This cluster also includes companies that have developed their own domain-specific analytics solutions and market them broadly to a client base.

For example, Nike, IBM, and Sportvision develop applications in sports analytics to improve the play and increase the viewership.

Axiom has developed clusters for virtually all households in the United States based on the data they collect about households from many different sources. Credit score and classification reporting companies (FICO, Experian, etc.) also belong in this group. IBM and several other companies offer pricing optimization solutions in the retail industry.

This field represents an entrepreneurial opportunity to develop industry-specific applications. Many emerging in Web/social media/location analytics are trying to profile users for better targeting of promotional campaigns in real time.

Examples of such companies and their activities include: YP.com employs location data for developing user/ group profiles and targeting mobile advertisements, Towerdata profiles users on the basis of e-mail usage, Qualia aims to identify users through all device usage, and Simulmedia targets advertisements on TV on the basis of analysis of a user's TV watching habits.

The growth of smartphones has spawned a complete industry focused on specific analytics applications for consumers as well as organizations.

For example, smartphone apps such as Shazam, Soundhound, or Musixmatch are able to identify a song on the basis of the first few notes and then let the user select it from their song base to play/download /purchase.

Waze uses real-time traffic information shared by users, in addition to the location data, for improving navigation. Voice recognition tools such as Siri on the iPhone, Google Now, and Amazon Alexa are leading to many more specialized analytics applications for very specific purposes in analytics applied to images, videos, audio, and other data that can be captured through smartphones and/or connected sensors.

Smartphones have also elevated the shared economy providers such as Uber, Lyft, Curb, and Ola. Many of these companies are exemplars of analytics leading to new business opportunities.

Online social media is another hot area in this cluster. Undoubtedly, Facebook is the leading player in the space of online social networking followed by Twitter and LinkedIn.

Moreover, the public access to their data has given rise to multiple other companies that analyze their data. For example, Unmetric analyzes Twitter data and provides solutions to their clients. Similarly, there are several other companies that focus on social network analysis.

A trending area in the application development industry is the IoT. Several companies are building applications to make smart objects.

For example, SmartBin has developed intelligent remote monitoring systems for the waste and recycling sectors. Several other organizations are working on building smart meters, smart grids, smart cities, connected cars, smart homes, smart supply chains, connected health, smart retail, and other smart objects.

The application developer sector is perhaps the biggest growth industry within analytics at this point. This cluster provides a unique opportunity for analytics professionals looking for more entrepreneurial career options.

Analytics Industry Analysts and Influencers

The next cluster of the analytics industry includes three types of organizations or professionals. The first group is the set of professional organizations that provide advice to the analytics industry providers and users. Their services include marketing analyses, coverage of new developments, evaluation of specific technologies, and development of training/white papers/ and so on. Examples of such players include organizations such as the Gartner Group, The Data Warehousing Institute, Forrester, McKinsey, and many of the general and technical publications and Web sites that cover the analytics industry.

Gartner Group's Magic Quadrants are highly influential and are based on industry surveys. Similarly, TDWI.org professionals provide excellent industry overviews and are very aware of current and future trends of this industry.

The second group includes professional societies or organizations that also provide some of the same services but are membership based and organized. For example, INFORMS, a professional organization, has now focused on promoting analytics.

Special Interest Group on Decision Support and Analytics, a subgroup of the Association for Information Systems, also focuses on analytics. Most of the major vendors (e.g., Teradata and SAS) also have their own membership-based user groups. These entities promote the use of analytics and enable sharing of the lessons learned through their publications and conferences. They may also provide recruiting services, and are thus good sources for locating talent.

A third group of analytics industry analysts is what we call analytics ambassadors, influencers, or evangelists. These analysts have presented their enthusiasm for analytics through their seminars, books, and other publications.

Illustrative examples include Steve Baker, Tom Davenport, Charles Duhigg, Wayne Eckerson, Bill Franks, Malcolm Gladwell, Claudia Imhoff, Bill Inman, and many others. Again, the list is not inclusive.

All of these ambassadors have written books and/or given many presentations to promote the analytics applications. Perhaps another group of evangelists to include here is the authors of textbooks on BI/analytics who aim to assist the next cluster to produce professionals for the analytics industry. Clearly, it will take some time for an analytics student to become a member of this cluster, but they could be working with members of this cluster as researchers or apprentices.

Academic Institutions and Certification Agencies

In any knowledge-intensive industry such as analytics, the fundamental strength comes from having students who are interested in the technology and choosing that industry as their profession. Universities play a key role in making this possible. This cluster, then, represents the academic programs that prepare professionals for the industry. It includes various components of



business schools such as information systems, marketing, management sciences, and so on.

It also extends far beyond business schools to include computer science, statistics, mathematics, and industrial engineering departments across the world. The cluster also includes graphics developers who design new ways of visualizing information.

Universities are offering undergraduate and graduate programs in analytics in all of these disciplines, though they may be labeled differently.

A major growth frontier has been certificate programs in analytics to enable current professionals to retrain and retool themselves for analytics careers. Certificate programs enable practicing analysts to gain basic proficiency in specific software by taking a few critical courses from schools that offer these programs.

TUN includes a list of analytics programs. It includes almost 150 programs, and there are likely many more such programs, with new ones being added daily.

Another group of players assists with developing competency in analytics. These are certification programs that award a certificate of expertise in specific software. Virtually every major technology provider (IBM, Microsoft, Microstrategy, Oracle, SAS, Tableau, and Teradata) has their own certification programs. These certificates ensure that potential new hires have a certain level of tool skills.

INFORMS offers a Certified Analytics Professional certificate program that is aimed at testing an individual's general analytics competency. Any of these certifications give a college student additional marketable skills.

The growth of academic programs in analytics is staggering. Only time will tell if this cluster is overbuilding the capacity that can be consumed by the other clusters, but at this point, the demand appears to outstrip the supply of qualified analytics graduates, and this is the most obvious place to find at least entry-level analytics hires.

Regulators and Policy Makers

The players in this component are responsible for defining rules and regulations for protecting employees, customers, and shareholders of the analytics organizations. The collection and sharing of the users' data require strict laws for securing privacy. Several organizations in this space regulate the data transfer and protect users' rights.

For example, the Federal Communications Commission (FCC) regulates interstate and international communications. Similarly, the Federal Trade Commission (FTC) is responsible for preventing data-related unfair business practices. The International Telecommunication Union (ITU) regulates the access to information and communication technologies (ICTs) to underserved communities worldwide.

On the other hand, a nonregulatory federal agency named the National Institute of Standards and Technology (NIST), helps advance the technology infrastructure. There are several other organizations across the globe that regulate the data security and accelerate the analytics industry. This is a very important component in the ecosystem so that no one can misuse consumers' information.

For anyone developing or using analytics applications, it is crucial to have someone on the team who is aware of the regulatory framework. These agencies and professionals who work with them clearly offer unique analytics talents and skills.

Analytics User Organizations

Clearly, this is the economic engine of the whole analytics industry, and therefore, we represent this cluster as the core of the analytics flower.

Organizations in every industry, regardless of size, shape, and location, are using or exploring the use of analytics in their operations. These include the private sector, government, education, military, and so on.

Examples of uses of analytics in different industries abound. Others are exploring similar opportunities to try and gain/retain a competitive advantage. Specific companies are not identified in this section; rather, the goal here is to see what type of roles analytics professionals can play within a user organization.

Of course, the top leadership of an organization, especially in the information technology group (chief information officer, etc.), is critically important in applying analytics to its operations. Reportedly, Forrest Mars of the Mars Chocolate Empire said that all management boiled down to applying mathematics to a company's operations and economics.

Although not enough senior managers subscribe to this view, the awareness of applying analytics within an organization is growing everywhere. A health insurance company executive once told us that his boss (the CEO) viewed the company as an IT-enabled organization that collected money from insured members and distributed it to the providers.

Thus efficiency in this process was the premium they could earn over a competitor. This led the company to develop several analytics applications to reduce fraud and overpayment to providers, promote wellness among those insured so they would use the providers less often, generate more efficiency in processing, and thus be more profitable.

Virtually all major organizations in every industry that we are aware of are hiring analytical professionals under various titles. Figure 1.14 is a word cloud of the selected titles of our program graduates at Oklahoma State University from 2013 to 2016. It clearly shows that Analytics and Data Science are popular titles in the organizations hiring graduates of such programs. Other key words appear to include terms such as Risk, Database, Security, Revenue, Marketing, and so on.



FIGURE 1.14 Word Cloud of Job Titles of Analytics Program Graduates.

Of course, user organizations include career paths for analytics professionals moving into management positions. These titles include project managers, senior managers, and directors, all the way up to the chief information officer or chief executive officer.

This suggests that user organizations exist as a key cluster in the analytics ecosystem and thus can be a good source of talent. It is perhaps the first place to find analytics professionals within the vertical industry segment.

Eleven different groups that play a key role in building and fostering this industry were identified. More petals/components can be added over time in the analytics flower/ecosystem. Because data analytics requires a diverse skill set, understanding of this ecosystem provides you with more options than you may have imagined for careers in analytics.

Moreover, it is possible for professionals to move from one industry cluster to another to take advantage of their skills. For example, expert professionals from providers can sometimes move to consulting positions, or directly to user organizations. Overall, there is much to be excited about the analytics industry at this point.

10. Assignment

Toppers: Go to <http://analytics-magazine.org/issues/digital-editions>, and find the January/February 2013 edition titled "Work Social." Read the article "Big Data, Analytics and Elections," and answer the following questions:

- a.What kinds of Big Data were analyzed in the article? Comment on some of the sources of Big Data.
- b.Explain the term integrated system. What is the other technical term that suits an integrated system?
- c.What kinds of data analysis techniques are employed in the project? Comment on some initiatives that resulted from data analysis.
- d. What are the different prediction problems answered by the models?
- e.List some of the actionable decisions taken that were based on the prediction results.
- f.Identify two applications of Big Data analytics that are not listed in the article

Above average Search the Internet for material regarding the work of managers and the role analytics plays. What kinds of references to consulting firms, academic departments, and programs do you find? What major areas are represented? Select five sites that cover one area, and report your findings

Average Go to microstrategy.com. Find information on the five styles of BI. Prepare a summary table for each style.

Below average Go to the TUN site. Explore the Sports Analytics page, and summarize at least two applications of analytics in any sport of your choice.

Slow performers - Search the Internet for material regarding the work of managers and the role analytics plays. What kinds of references to consulting firms, academic departments, and programs do you find? What major areas are represented? Select five sites that cover one area, and report your findings

11. Part - A Question and Answers

1.List three of the terms that have been predecessors of analytics. (CO1,K1)

Analytics have evolved from other systems over time including Data Support Systems (DSS), Operations Research (OR) models, Expert Systems(ES)

2.What was the primary difference between the systems called MIS, DSS, and Executive Support Systems? (CO1,K1)

Many systems have been used in the past and present to provide analytics Management Information system provided reports on various aspects of business functions

3. List and describe the major components of BI. (CO1,K1)

BI systems have four major components: the data warehouse (with its source data), business analytics (a collection of tools for manipulating, mining, and analyzing the data in the data warehouse), business performance management (for monitoring and analysing performance), and the user interface (eg., a dashboard).

4. Define OLTP. (CO1,K1)

OLTP (online transaction processing) is a type of computer processing where the computer responds immediately to user requests. Each request is considered to be a transaction, which is a computerized record of a discrete event, such as the receipt of inventory or a customer order.

5. Define BI. (CO1,K1)

Business Intelligence(BI) is an umbrella term that combines architecture, tools, databases, analytics tools, applications, and methodologies. Its major objective is to enable interactive access (sometimes in real time) to data,

enable manipulation of these data, and provide business managers and analysts the ability to conduct appropriate analysis.

6. Define OLAP. (CO1,K1)

Online analytical processing (OLAP) is a technology that organizes large business databases and supports complex analysis. It can be used to perform complex analytical queries without negatively affecting transactional systems.

7. Define analytics. (CO1,K1)

The term replaces terminology referring to individual components of a decision support system with one broad word referring to business intelligence. More precisely, analytics is the process of developing actionable decisions or recommendations for actions based upon insights generated from historical data.

8.What is descriptive analytics? What are the various tools that are employed in descriptive analytics? (CO1,K1)

Descriptive analytics refers to knowing what is happening in the organization and understanding some underlying trends and causes of such occurrences. Tools used in descriptive analytics include data warehouses and visualization applications.

9.How is descriptive analytics different from traditional reporting? (CO1,K1)

Descriptive analytics gathers more data, often automatically. It makes results available in real time and allows reports to be customized.

10.What is a DW? How can data warehousing technology help to enable analytics? (CO1,K1)

A data warehouse is the component of a BI system that contains the source data. developing a data warehouse usually includes development of the data infrastructure for descriptive analytics that is, consolidation of data sources

and making relevant data available in a form that enables appropriate reporting and analysis. A data warehouse serves as the basis for developing appropriate reports, queries, alerts, and trends.

11. What is predictive analytics? How can organizations employ predictive analytics? (CO1,K1)

Predictive analytics is the use of statistical techniques and data mining to determine what is likely to happen in the future. Businesses use predictive analytics to forecast whether customers are likely to switch to a competitor, what customers are likely to buy, how likely customers are to respond to a promotion, and whether a customer is creditworthy. Sports teams have used predictive analytics to identify the players most likely to contribute to a team's success.

12. What is prescriptive analytics? What kind of problems can be solved by prescriptive analytics? (CO1,K1)

Prescriptive analytics is a set of techniques that use descriptive data and forecasts to identify the decisions most likely to result in the best performance. Usually, an organization uses prescriptive analytics to identify the decisions or actions that will optimize the performance of a system. Organizations have used prescriptive analytics to set prices, create production plans, and identify the best locations for facilities such as bank branches.

13. Define modeling from the analytics perspective. (CO1,K1)

Analytics uses descriptive data to create models of how people, equipment, or other variables operate in the real world. These models can be used in predictive and prescriptive analytics to develop forecasts, recommendations, and decisions.

14. Is it a good idea to follow a hierarchy of descriptive and predictive analytics before applying prescriptive analytics? (CO1,K1)

It is important in any analytics project to understand the business domain and current state of the business problem. This requires analysis of historical data, or descriptive analytics. Although the chapter does not discuss a hierarchy of analytics, students may observe that testing a model with predictive analytics could logically improve prescriptive use of the model.

15. How can analytics aid in objective decision making? (CO1,K1)

Problem solving in organizations has tended to be subjective, and decision makers tend to rely on familiar processes. The result is that future decisions are no better than past decisions. Analytics builds on historical data and takes into account changing conditions to arrive at fact-based solutions that decision makers might not have considered.

16. What is Big Data analytics? (CO1,K1)

The term Big Data refers to data that cannot be stored in a single storage unit. Typically, the data is arriving in many different forms, be they structured, unstructured, or in a stream. Big Data analytics is analytics on a large enough scale, with fast enough processing, to handle this kind of data.

17. What are the sources of Big Data? (CO1,K1)

Major sources include clickstreams from Web sites, postings on social media, and data from traffic, sensors, and the weather.

17. What are the characteristics of Big Data? (CO1,K1)

Today Big Data refers to almost any kind of large data that has the characteristics of volume, velocity, and variety. Examples include data about Web searches, such as the billions of Web pages searched by Google; data about financial trading, which operates in the order of microseconds; and data about consumer opinions measured from postings in social media.

18.How does customer sentiment analysis help companies provide a personalized service to their customers? (CO1,K2)

By understanding customer sentiment, the company is able to customize communications to customers

19.List some of the implementation topics addressed by Gartner's report. (CO1,K1)

Gartner's framework decomposes planning and execution into business, functionality and infrastructure components. At the business and organizational levels, strategic and operational objectives must be defined while considering the available organizational skills to achieve those objectives. Issues of organizational culture surrounding BI initiatives and building enthusiasm for those initiatives and procedures for the intra-organizational sharing BI best practices must be considered by upper management-- with planes in place to prepare the organization for change.

20.What processing technique is applied to process Big Data? (CO1,K1)

One computer, even a powerful one, could not handle the scale of Big Data. The solution is to push computation to the data, using the MapReduce programming paradigm.

21 .List the 11 categories of players in the analytics ecosystem. (CO1,K1)

These categories include:

Data Generation Infrastructure Providers

Data Management Infrastructure Providers

Data Warehouse Providers

Middleware Providers

Data Service Providers

Analytics Focused Software Developers

Application Developers: Industry Specific or General

Analytics Industry Analysts and Influencers

Academic Institutions and Certification Agencies

Regulators and Policy Makers

Analytics User Organizations

22. How does customer sentiment analysis help companies provide a personalized service to their customers? (CO1,K2)

By understanding customer sentiment, the company is able to customize communications to customers

23. What is customer sentiment analysis? (CO1,K2)

Customer sentiment analysis categorizes and evaluates customer opinion by types of emotion.

24. How do visualizations aid in understanding the data and delivering insights into the data?(CO1,K2)

Visualizations provide a great tool for gaining the initial insights into data, which can be further refined based on expert opinions. Visualizations also aid in generating ideas for obscured business problems, which can be pursued in building predictive models.

25. What is a classification problem? (CO1,K2)

Classification is a technique used in developing predictive analytical applications. In this example, various kinds of data, such as severity of injury and healing time, were classified for purposes of making decisions about how to handle injuries and which players might be available to play in the future.

26. Is it better to be the strongest player in one category or be active in multiple categories? (CO1,K2)

It can be argued that cross-discipline strength provides better integration and insight, or that domination in multiple areas reduces completion or innovation.

27. What is traditional reporting? How is it used in the organization?(CO1,K2)

The traditional reporting process is a manual process of collecting and aggregating financial and other information. Organizations have used this time-consuming process as a way to obtain information for making decisions. However, the resulting presentations may be flat, slow to develop, and difficult to apply to specific situations

28. How can analytics be used to transform traditional reporting?(CO1,K2)

Analytics can enable real-time decision support and deliver information to a user-friendly dashboard. Users of a dashboard such as the one provided by Sabre's Enterprise Travel Data Warehouse can see at a glance a 360-degree view of the company's overall health generated from various data sources. Many stakeholders in the organization can request data needed for particular types of decisions, and the graphical user interface makes the information easily understandable.

29. How can interactive reporting assist organizations in decision making?(CO1,K2)

When a system incorporates interactive drill-down capabilities, users can select the data they need for evaluating a specific kind of performance and making decisions in a particular function or situation. For example, the airlines using the Sabre system can focus on data about sales performance (ticketing, seats sold, etc.) and operational performance (flight movement, inventory etc).

This flexibility encourages decision makers to use data in support of their decisions.

Application Case 1.2: Silvaris Increases Business with Visual Analysis and Real-Time Reporting Capabilities

30. Give examples of companies in each of the 11 types of players.

Examples of companies by area include:

Data Generation Infrastructure Providers (Sports Sensors, Zepp, Shockbox, Advantech B+B SmartWorx, Garmin, and Sensys Network, Intel, Microsoft, Google, IBM, Cisco, Smartbin, SIKO Products, Omega Engineering, Apple, and SAP)

Data Management Infrastructure Providers (Dell NetApp, IBM, Oracle, Teradata, Microsoft, Amazon (Amazon Web Services), IBM (Bluemix), Salesforce.com, Hadoop clusters, MapReduce, NoSQL, Spark, Kafka, Flume)

Data Warehouse Providers (IBM, Oracle, Teradata, Snowflake, Redshift, SAS, Tableau)

Middleware Providers (Microstrategy, Plum, Oracle, SAP, IBM, SAS, Tableau)

Data Service Providers (Nielsen, Experian, Omniture, Comscore, Google, Equifax, TransUnion, Acxiom, Merkle, Epsilon, Avention, ESRI.org)

Analytics Focused Software Developers (Microsoft, Tableau, SAS, Gephi, IBM, KXEN, Dell, Salford Systems, Revolution Analytics, Alteryx, RapidMiner, KNIME, Rulequest, NeuroDimensions, FICO, AIIMS, AMPL, Frontline, GAMS, Gurobi, Lindo Systems, Maximal, NGData, Ayata, Rockwell, Simio, Palisade, Frontline, Exsys, XpertRule, Teradata, Apache, Tibco, Informatica, SAP, Hitachi)

Application Developers: Industry Specific or General (IBM, SAS, Tera data, Nike, Sportsvision, Acxiom, FICO, Experian, YP.com, Towerdata, Qualia, Simulmedia, Shazam, Soundhound, Musixmatch, Waze, Apple, Google, Amazon, Uber, Lyft, Curb, Ola, Facebook, Twitter, LinkedIn, Unmetric, Smartbin)

Analytics Industry Analysts and Influencers (Gartner Group, The Data Warehousing Institute, Forrester, McKinsey, INFORMS, AIS, Teradata, SAS)

Academic Institutions and Certification Agencies (IBM, Microsoft, Microstrategy, Oracle, SAS, Tableau, Teradata, INFORMS)

Regulators and Policy Makers (Federal Communications Commission, Federal Trade Commission, International Telecommunication Union, National Institute of Standards and Technology)

Analytics User Organizations (many topic specific and local groups)



12. Part - B Questions

- 1.Explain Evolution of Computerized Decision Support to Analytics in detail (CO1,K1)
2. Explain Big data analytics in detail. (CO1,K1)
3. Explain Analytics ecosystem in detail. (CO1,K1)
- 4.Compare and contrast predictive analytics with prescriptive and descriptive analytics. Use examples.(CO1,K1)
5. Discuss the major issues in implementing BI. (CO1,K1)
6. Give examples of companies in each of the 11 types of players. (CO1,K1)



13. SUPPORTIVE ONLINE CERTIFICATION COURSES

NPTEL : https://onlinecourses.nptel.ac.in/noc20_mg11/preview

Coursera : <https://in.coursera.org/specializations/business-analytics>

Udemy : <https://www.udemy.com/course/businessanalysis/>



14. Real Time Applications

Fraud Detection in Financial Transactions:

Real-time analysis of financial transactions to identify and prevent fraudulent activities.

Dynamic Pricing in E-commerce:

Adjusting product prices in real-time based on market demand, competitor pricing, and other factors.

IoT Analytics for Smart Cities:

Monitoring and analyzing data from sensors, cameras, and other IoT devices to optimize city services, traffic flow, and resource allocation.

Healthcare Monitoring and Alerts:

Real-time monitoring of patient data, allowing for immediate alerts and interventions in critical situations.

Adaptive Streaming in Media Services:

Adjusting the quality of video streams based on network conditions and viewer preferences in real-time.

Supply Chain Optimization:

Tracking and analyzing data throughout the supply chain to optimize inventory levels, reduce delays, and enhance overall efficiency.

Energy Grid Monitoring:

Real-time monitoring of energy consumption and grid performance to optimize distribution and prevent outages.

Social Media Analytics:

Analyzing social media feeds in real-time to understand trends, sentiments, and engage with customers or address issues promptly.

Predictive Maintenance in Manufacturing:

Monitoring equipment sensors to detect anomalies and predict when maintenance is required to prevent downtime.

Real-time Analytics for Online Gaming:

Analyzing player behavior, preferences, and in-game actions to optimize gaming experiences and introduce dynamic challenges.

Network Security Monitoring:

Analyzing network traffic in real-time to detect and respond to security threats and breaches.

Weather Forecasting:

Processing real-time data from weather sensors to provide accurate and timely weather forecasts.

Airline Flight Optimization:

Analyzing real-time data on weather conditions, flight schedules, and airport operations to optimize flight routes and schedules.

Customer Interaction and Personalization:

Providing personalized recommendations and responses to customers in real-time based on their behavior and preferences.

15. Assessment Schedule

Sl. No.	ASSESSMENT	Proposed Date	Actual Date
1	FIRST INTERNAL ASSESSMENT	12-02-2024 to 17-02-2024	
2	SECOND INTERNAL ASSESSMENT	01-04-2024 to 06-04-2024	
3	MODEL EXAMINATION	20-04-2024 to 30-04-2024	
4	END SEMESTER EXAMINATION	11-05-2024	



16. PRESCRIBED TEXT BOOKS & REFERENCE BOOKS

TEXT BOOKS:

Ramesh Sharda, Dursun Delen, Efraim Turban, "Business Intelligence, Analytics, and Data Science: A Managerial Perspective", Pearson, 4th Edition, 2018.

Jesper Thorlund & Gert H.N. Laursen, "Business Analytics for Managers: Taking Business Intelligence beyond Reporting, Wiley, 2010.

REFERENCES:

Shmueli, Patel, and Bruce: Wiley, Data Mining for Business Intelligence, Concepts, Techniques and Applications, Wiley, 2010

R.N.Prasad and Seema Acharya, "Fundamentals of Business Analytics", 2nd Edition, Wiley, 2016.

17. Mini Project

Toppers - Smart City Traffic Management using Real-time Analytics

Above average - Sports Analytics: Performance Prediction and Player Evaluation

Average - Fraud Detection in Financial Transactions using Machine Learning

Below average - Social Media Sentiment Analysis for Brand Reputation Management

Slow performers - Healthcare Analytics: Predicting Patient Readmission Rates

Thank you



Disclaimer:

This document is confidential and intended solely for the educational purpose of RMK Group of Educational Institutions. If you have received this document through email in error, please notify the system manager. This document contains proprietary information and is intended only to the respective group / learning community as intended. If you are not the addressee you should not disseminate, distribute or copy through e-mail. Please notify the sender immediately by e-mail if you have received this document by mistake and delete this document from your system. If you are not the intended recipient you are notified that disclosing, copying, distributing or taking any action in reliance on the contents of this information is strictly prohibited.