READING MATERIALS - EDUCATIONAL PURPOSES ONLY

BUSINESS INTELLIGENCE

Table of Contents

Chapter 1 INTRODUCTION TO BUSINESS INTELLIGENCE……3
1.1 AN INTRODUCTION TO BUSINESS INTELLIGENCE 4
1.1.1 Importance of Business Intelligence 5
1.1.2 Value of Business Intelligence 7
1.1.3 History of Business Intelligence 7
1.1.4 The Architecture of Business Intelligence 8
1.1.5 Business Intelligence Tools 9
1.1.6 Power BI Tutorial 10
Chapter 2 DATA WAREHOUSING ……..19
2.1 DATA WAREHOUSING 20
2.1.1 What does Data Warehouse (DW) mean? 23
2.1.2 Application Case 25
2.1.3 Tutorial 1 28
Chapter 3 BUSINESS PERFORMANCE MANAGEMENT……...…40
3.1 BUSINESS PERFORMANCE MANAGEMENT BPM 40
3.1.1 Understanding Business Performance Management 40
3.1.2 BPM and Bl Compared 42
3.1.3 Business Process Performance Management: Case study 44
3.1.3.1 The Control System in Business 45
3.1.3.2 Case Study: Managing a Composite Process “from Customers’ Orders to Payment collection” 46
3.1.3.3 Ways of Resolving a Business Problem: Research Methods 47
3.1.4 Tutorial 1 50
Chapter 4 DATA MINING FOR BUSINESS INTELLIGENCE 61
4.1 DATA MINING FOR BUSINESS INTELLIGENCE 61
4.1.1 How Data Mining is used to Generate Business Intelligence 63
4.1.2 What is Data Science? 64
4.1.3 Why are there so many Different Methods? 65
4.1.3.1 Terminology and Notation 66
4.1.4 Data in data mining 66
4.1.5 Data Mining Techniques 69

4.1.6 Data Visualization 70
4.1.7 Tutorial 1 72
Chapter 5 TEXT AND WEB MINING 95
5.1 TEXT AND WEB MINING 95
5.1.1 Text mining Concepts and Definition 99
5.1.2 Text Mining Lingo 101
5.1.3 Application Case 103
5.1.4 Tutorial 1 104
5.1.5 Tutorial 2 109
Chapter 6 EMERGING TRENDS IN BUSINESS INTELLIGENCE 111
6.1 EMERGING TRENDS IN BUSINESS INTELLIGENCE 111
6.1.1 Implementing Bl: An Overview 113
6.1.2 Tutorial 118
REFERENCES: 132

CHAPTER- 1
INTRODUCTION TO BUSINESS INTELLIGENCE

1.1 AN INTRODUCTION TO BUSINESS INTELLIGENCE [1][2][3]

Business Intelligence (BI) is a terminology representing a collection of processes, tools and technologies helpful in achieving more profit by considerably improving the productivity, sales and service of an enterprise. With the help of BI methods, the corporate data can be organized, analyzed in a better way and then converted into a useful knowledge of information needed to initiate a profitable business action. Thus it's about turning a raw, collected data into intelligent information by analyzing and re-arranging the data according to the relationships between the data items by knowing what data to collect and manage and in what context. [1]

Business intelligence (BI) is a technology-driven process for analyzing data and presenting actionable information to help executives, managers and other corporate end users make informed business decisions. BI encompasses a wide variety of tools, applications and methodologies that enable organizations to collect data from internal systems and external sources; prepare it for analysis; develop and run queries against that data; and creates reports, dashboards and data visualizations to make the analytical results available to corporate decision-makers, as well as operational workers. [2]

Business intelligence (BI) is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies. It is, like DSS, a content-free expression, so it means different things to different people. Part of the confusion about BI lies in the flurry of acronyms and buzzwords that are associated with it (e.g., business performance management [BPM]). BI's major objective is to enable interactive access (sometimes in real time) to data, to enable manipulation of data, and to give business managers and analysts the ability to conduct appropriate analysis. 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.[3]

1.1.1 Importance of Business Intelligence [1][3][4]

A company's collected raw data is an important asset where one can find solutions to many of an organization's critical questions like 'what was the net profit for a particular product last year and what will be sales this year and what are the key factors to be focused this year in order to increase the sales?'. So there arises a necessity of a well planned BI system which can lead to a greater profitability by reducing the operating costs, increasing the sales and thereby improving the customer satisfaction for an enterprise. With the help of a Business Intelligence System, a company may improve its business or rule over its competitors by exploring and exploiting its data to know the customer preferences, nature of customers, supply chains, geographical influences, pricing and how to increase its overall business efficiency. [1]

One example is the Business Pressures–Responses–Support Model, and as its name indicates, has three components: business pressures that result from today's business climate; responses (actions taken) by companies to counter the pressures (or to take advantage of the opportunities available in the environment); and computerized support that facilitates the monitoring of the environment and enhances the response actions taken by organizations. [3]

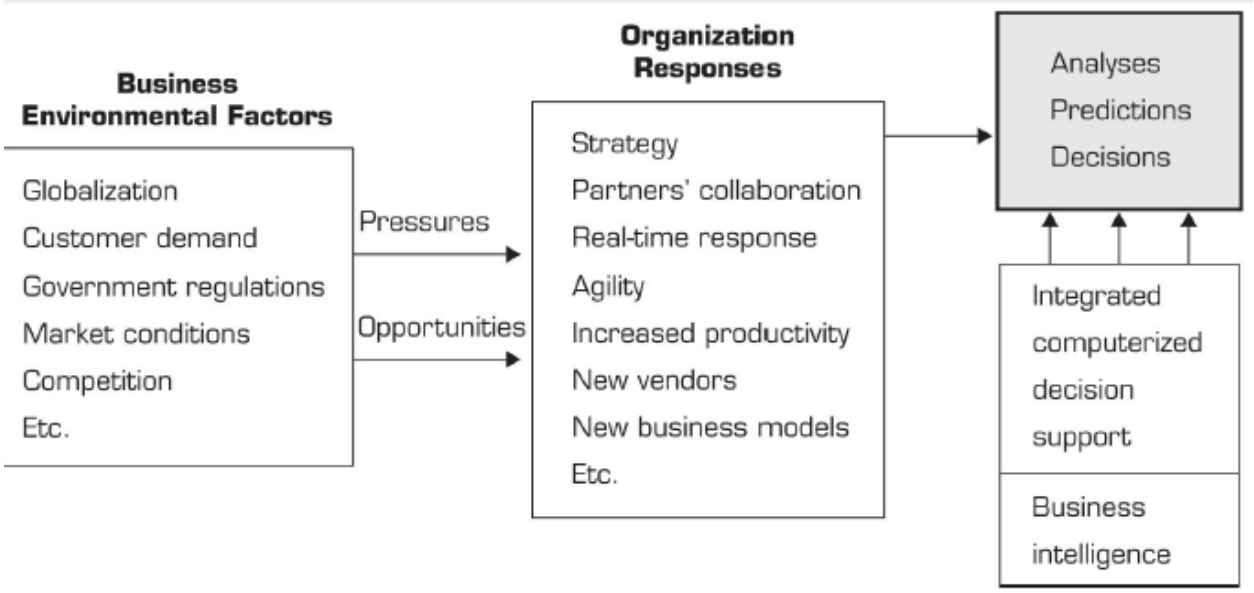


Figure 1.1 The Business Pressures–Responses–Support Model [3]

Business Intelligence is a technology based on customer and profit oriented models that reduces operating costs and provide increased profitability by improving productivity, sales, service and helps to make decision making capabilities at no time. Business Intelligence Models are based on multi dimensional analysis and key performance indicators (KPI) of an enterprise. Business Intelligence applications that are based on Business Intelligence Models are created by Business Intelligence software which provides the aggregated details about suppliers, customers, internal activities, business to business transactions to the managers or whoever needs it to take better corporate decisions. Many business questions or situations need to be analyzed in order to achieve the target of an enterprise with the help of several managers or executives in each cadre.[4]

- Business Intelligence in Finance:

What are the net income, expenses, gross profit, and net profit for this quarter/year?

- Business Intelligence in Accounts:

What is the sales amount this month and what is the outstanding pending payment?

- Business Intelligence in Purchase:

Who is the vendor to be contacted to purchase products?

- Business Intelligence in Production:

How many products are manufactured in each production unit today/weekly/monthly?

- Business Intelligence in Sales:

How many products have been sold in each area today/weekly/monthly?

- Business Intelligence in Quality:

How many products has been defective today/weekly/monthly/quarterly/yearly?

- Business Intelligence in Service:

Are the customers satisfied with the quality?

These business intelligence questions are related with why, what, how, when, and business intelligence reports (olap reports) are useful in providing solutions to the above questions by means of reporting, score cards, balance score cards that are helpful in managerial decisions.[4]

1.1.2 Value of Business Intelligence [1]

Business Intelligence enables us to take some action based on the intelligence acquired using BI strategy. If this knowledge or information is not utilized properly in the right direction, there is no point in analyzing and finding the intelligence. For example, let's assume a company has implemented a BI system to analyze the customer interests and requirements enabling them to promote a particular product in the near future. All the analysis and knowledge management will be pointless and a waste of investment if they don't have a proper plan to approach the right customer at the right time. So Business Intelligence is all about strategies in increasing business efficiency while vastly cutting down the operating costs. Implementing a Business Intelligence system in an organization requires a significant amount of money to be invested in order to build and implement a BI system and its applications. It requires more skilled top level managers to build a ROI (Return on Investment) model to analyze the costs involved in implementing and maintaining these BI models and methods to get the return on investment sooner. A proper business action should be taken based on the strategies derived with the help of these intelligence models. Often an erroneous model and wrong assumptions can bring a loss much greater than building the entire Business Intelligence system itself. Once everything is done more properly in a way an organization want them to be, then the benefit that comes out of it is priceless.

1.1.3 History of Business Intelligence [3]

The term BI was coined by the Gartner Group in the mid-1990s. However, the concept is much older; it has its roots in the MIS reporting systems of the 1970s. During that period, reporting systems were static, two dimensional, and had no analytical capabilities. In the early 1980s, the concept of executive information systems (EIS) 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-basedenterprise 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.

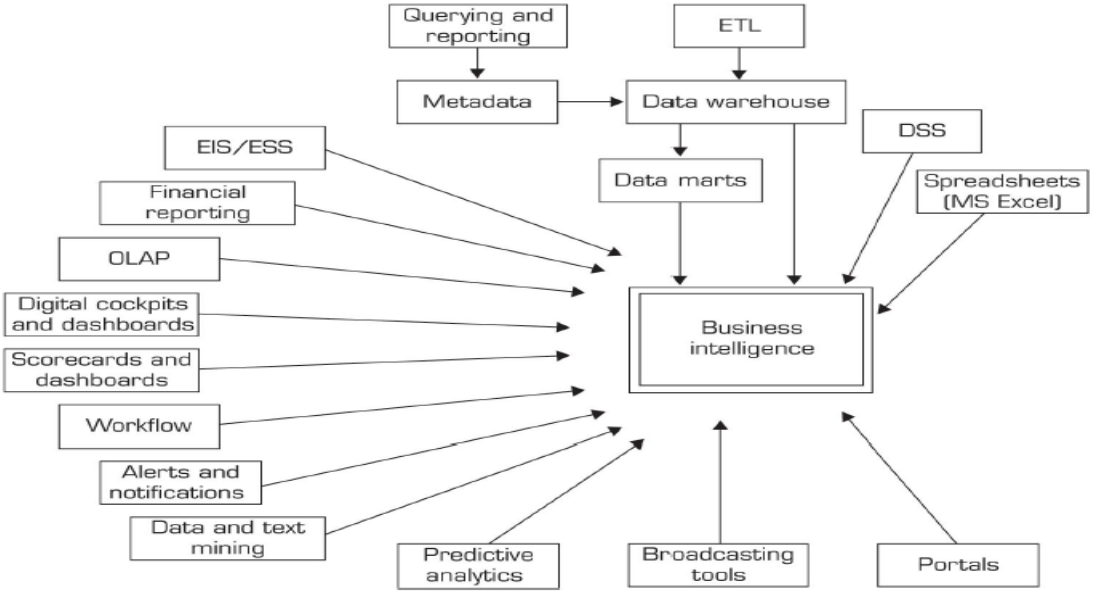


Figure 1.2 Evolution of Business Intelligence (BI) [3]

1.1.4 The Architecture of Business Intelligence [3]

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 data warehouse; business performance management (BPM) for monitoring and analyzing performance; and a user interface (e.g., a dashboard).

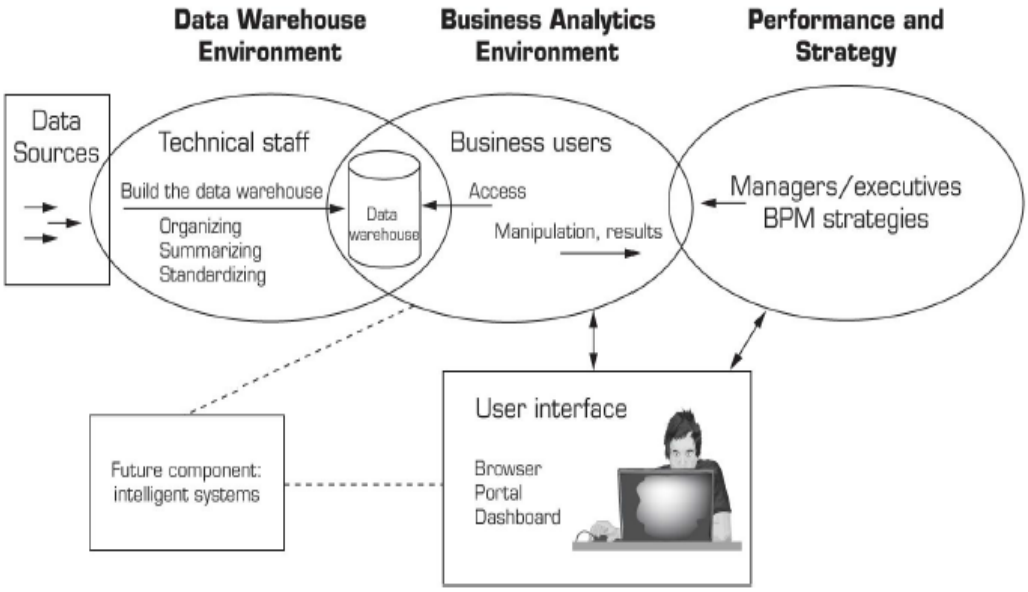


Figure 1.3 A High-Level Architecture of BI [3]

(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)

1.1.5 Business Intelligence Tools [4]

Business Intelligence Tools help to gather, store, access and analyze corporate data to aid in decision-making. Generally these systems will illustrate business intelligence in the areas of customer profiling, customer support, market research, market segmentation, product profitability, statistical analysis, and inventory and distribution analysis. With Business Intelligence Tools, various data like customer related, product related, sales related, time related, location related, employee related etc. are gathered and analyzed based on which important strategies or rules are formed and goals to achieve their target are set. These decisions are very efficient and effective in promoting an organization's growth. Since the collected data can be sliced across almost all the dimensions like time, location, product, promotion etc., valuable statistics like sales profit in one region for the current year can be calculated and compared with the previous year statistics. [4]

- Popular Business Intelligence Tools:

Tool Name	Company Name
Business Objects	Business Objects
Cognos	Cognos
Hyperion	Hyperion
Microstrategy	Microstrategy
Microsoft Reporting Services	Microsoft
Crystal	Business Objects

Figure 1.4 Business Intelligence Tools [4]

1.1.6 Power BI Tutorial [5]

Power BI is a Data Visualization and Business Intelligence tool that converts data from different data sources to interactive dashboards and BI reports. Power BI suite provides multiple software, connector, and services - Power BI desktop, Power BI service based on Saas, and mobile Power BI apps available for different platforms. These set of services are used by business users to consume data and build BI reports. This tutorial covers all the important concepts in Power BI and provides a foundational understanding on how to use Power BI.

Power BI desktop app is used to create reports, while Power BI Services (Software as a Service - SaaS) is used to publish the reports, and Power BI mobile app is used to view the reports and dashboards.

Power BI Desktop is available in both 32-bit and 64-bit versions:

https://powerbi.microsoft.com/en-us/downloads/

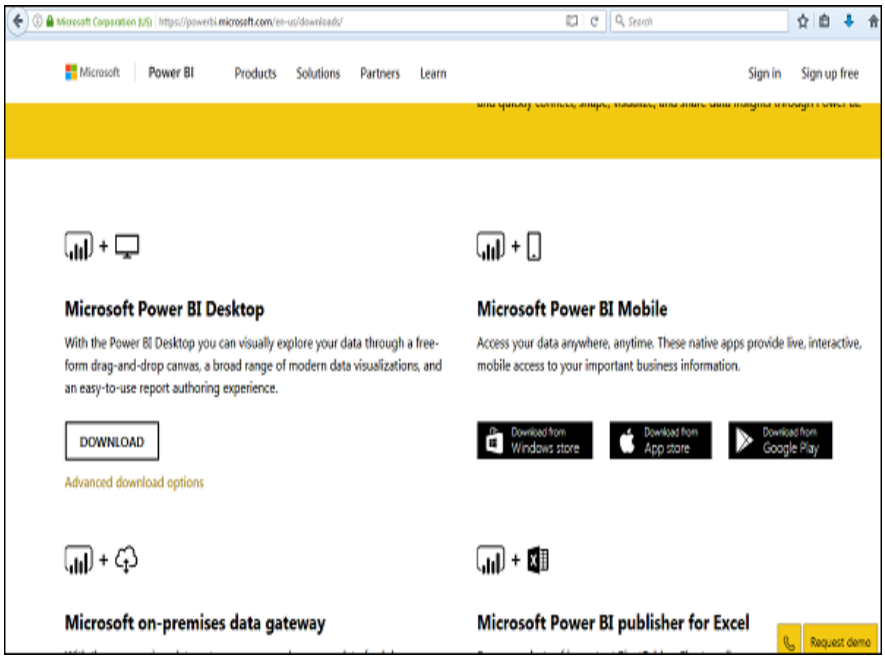


Figure 1.5 Power Bi Desktop download

To check the system requirements, installation files detail, users have to navigate to "Advanced download options". Following are the system requirements to download Power BI tool –

Supported Operating Systems

- Windows 10, Windows 7, Windows 8, Windows 8.1, Windows Server 2008 R2, Windows Server 2012, Windows Server 2012 R2
- Microsoft Power BI Desktop requires Internet Explorer 9 or higher
- Microsoft Power BI Desktop is available for 32-bit (x86) and 64-bit (x64) platforms

Users can select a language in which they want to install Power BI and following files are available for download.

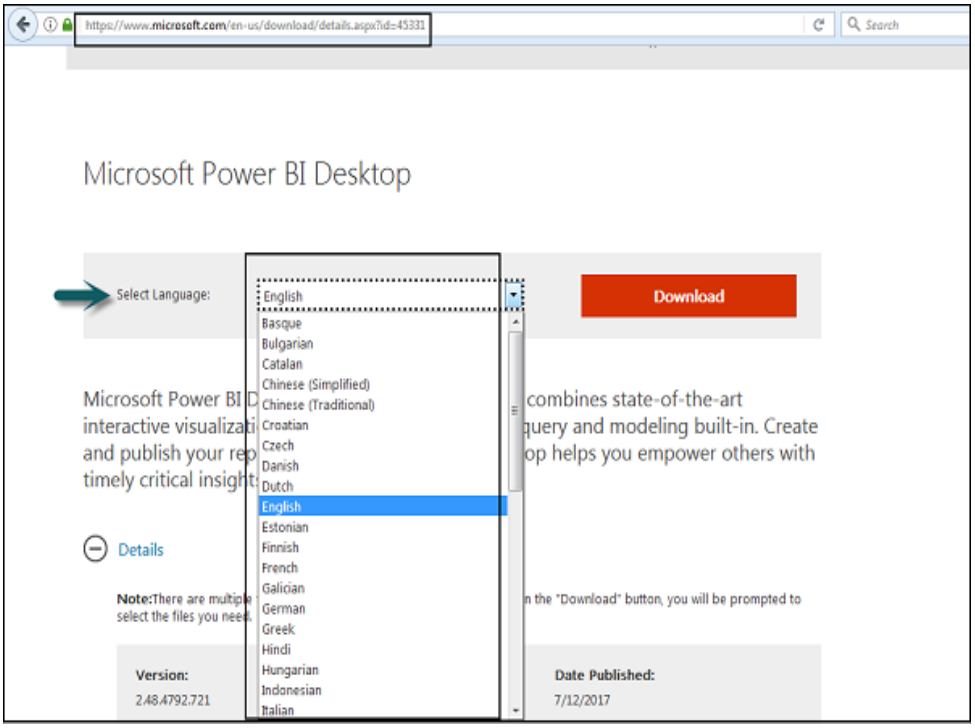


Figure 1.6 Power BI Desktop

This is the link to directly download Power BI files:

https://www.microsoft.com/en-us/download/details.aspx?id=45331

Figure 1.6 Download Power BI files

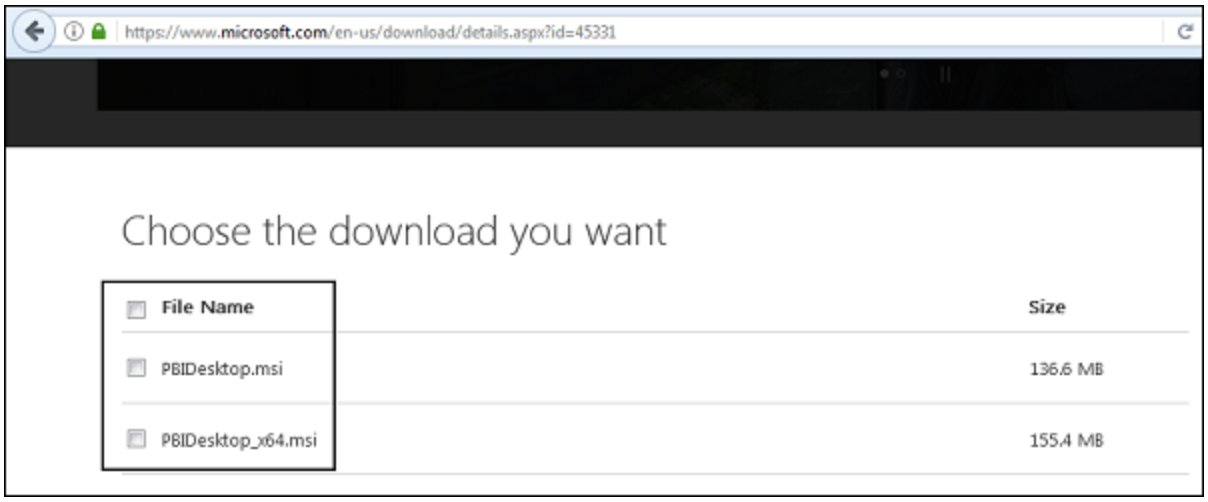


Figure 1.7 Download Power BI files

PBIDesktop_x64.msi shows a 64-bit OS file. The file that you want to be installs should be selected.

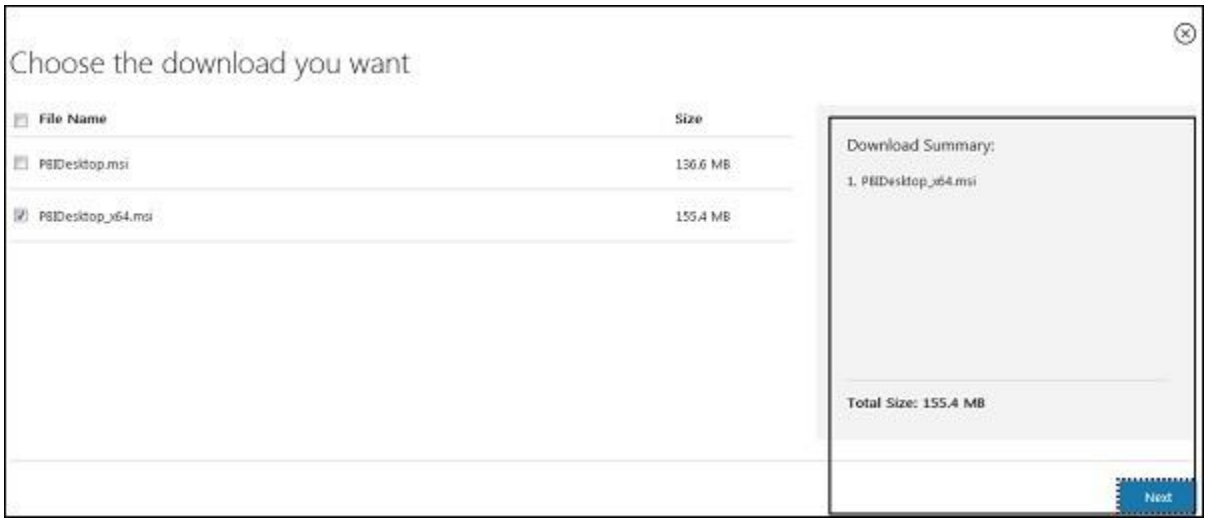


Figure 1.8 Installing Power BI

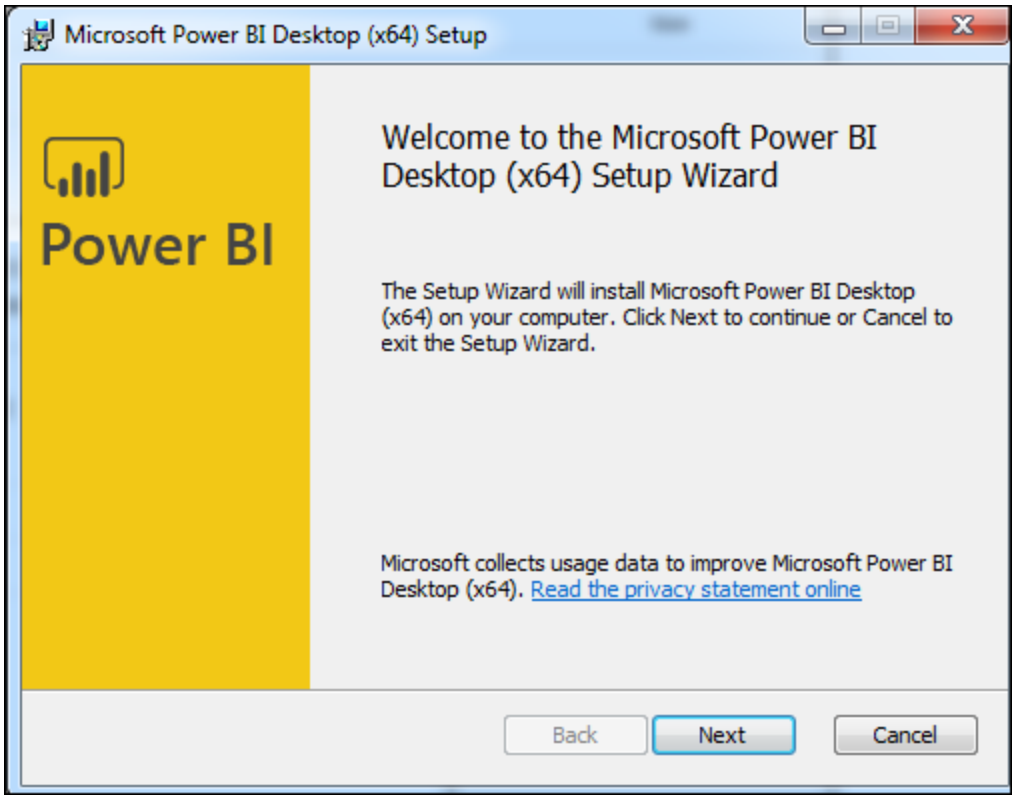


Figure 1.9 Installing Power BI

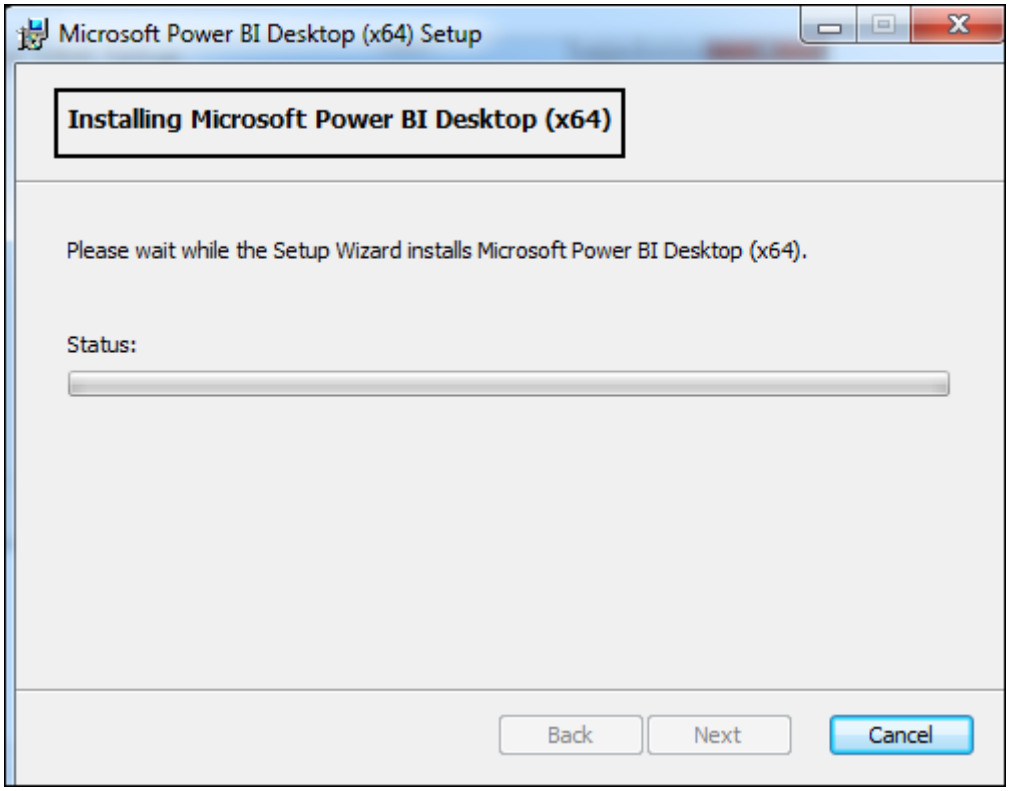


Figure 1.10 Installing Power BI

When Power BI is installed, it launches a welcome screen. This screen is used to launch different options related to get data, enrich the existing data models, create reports as well as publish and share reports.

Power BI – Architecture

Power BI includes the following components:

- Power BI Desktop − This is used to create reports and data visualizations on the dataset.
- Power BI Gateway − You can use Power BI on-premises gateway to keep your data fresh by connecting to your on-premises data sources without the need to move the data. It allows you to query large datasets and benefit from the existing investments.
- Power BI Mobile Apps − Using Power BI mobile apps, you can stay connected to their data from anywhere. Power BI apps are available for Windows, iOS, and Android platform.
- Power BI Service − This is a cloud service and is used to publish Power BI reports and data visualizations.

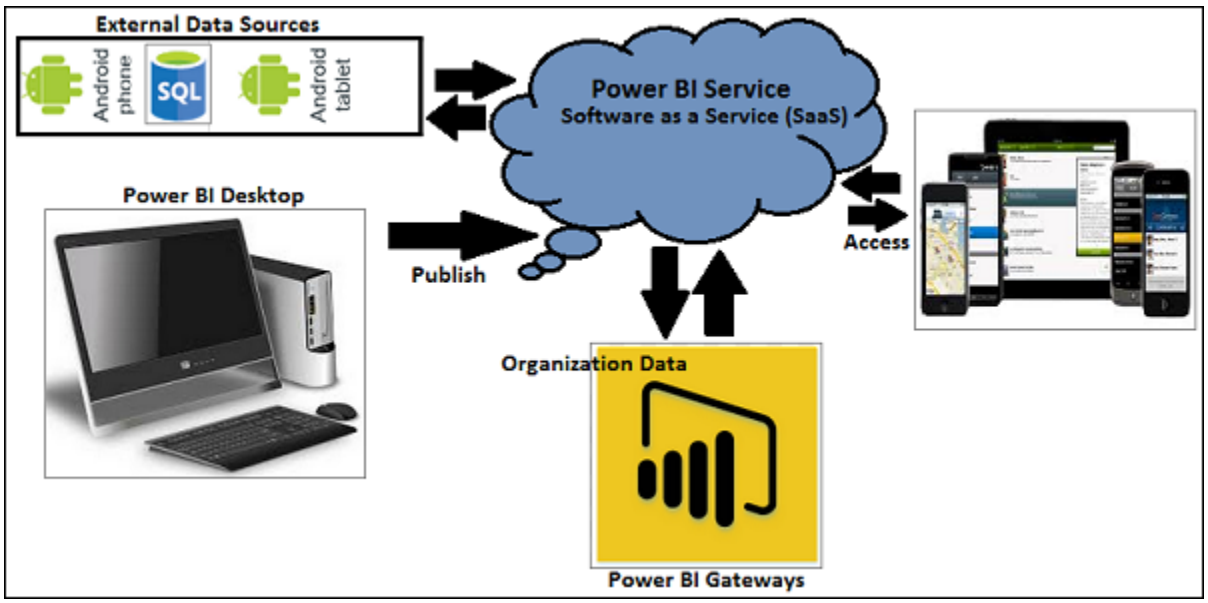


Figure 1.11 Components of Power BI

Power BI - Supported Data Sources

Power BI supports large range of data sources. You can click Get data and it shows you all the available data connections. It allows you to connect to different flat files, SQL database, and

Azure cloud or even web platforms such as Facebook, Google Analytics, and Salesforce objects. It also includes ODBC connection to connect to other ODBC data sources, which are not listed.

Following are the available data sources in Power BI :

- SQL Database
- OData Feed
- Blank Query
- Azure Cloud platform
- Online Services
- Blank Query
- Other data sources such as Hadoop, Exchange, or Active Directo
- Power BI - Comparison with Other BI Tools

Tableau is considered as one of the leading tools in the BI market. Power BI is considered as an emerging tool in close competition with Tableau because of its backend data manipulation features and connectivity with the list of data sources. Tableau is one of the best data visualization tools in the market and is used by medium and large enterprises. Power BI is closely integrated with Office 365 suite, and hence it is compatible other sources such as SharePoint.

Feature	**Tableau**	**Power BI**
Data Visualization	Tableau provides strong data visualization and is one of the main data visualization tool in the market.	Power BI provides a strong backend data manipulation feature with access to simple visualizations.
Size of Dataset	Tableau can connect much larger datasets as compared to Power BI.	Power BI has a limit of 1GB data in free version.
Data Sources	Tableau covers a vast range of data sources to connect with for data	Power BI covers most of the data sources available in

	visualization. In Tableau, you select the dataset first and visualizations are used on the fly.	Tableau. It is closely integrated with Office 365, hence provides connectivity to SharePoint. Power BI online version also supports direct visualization on Search Engine, though, only Bling is supported at this point.
Costing	Tableau is expensive as compared to Power BI but still under budget for small and medium enterprise.	Power BI provides a free version with 1GB limit on dataset. Power BI Pro is also a cheaper solution when compared with any other BI tool.
License and Pricing	Tableau Desktop Profession: USD70/user/month and it can connect to hundreds of data sources. Tableau Desktop Personal: USD35/user/month and it can connect to data sources such as Google Sheets and Excel files. Tableau Server: Minimum 10 users with the cost of USD35/user/month Tableau Online with private cloud: USD 42/user/month	Power BI: Free 1 GB storage 10k rows/hour data streaming Power BI Pro: USD9.99/user/month 10 GB storage 1 million rows/hour

Implementation	Tableau provides different implementation types as per organizational needs panning from few hours to few weeks.	Power BI uses cloud storage and includes simple implementation process.

Table 1.1 Power BI vs Tableau [5]

Feature	**SSRS**	**Power BI**
Data Visualization	SSRS is mostly used for Pixel perfect reporting and average dash-boarding features.	Power BI provides a strong backend data manipulation feature with access to simple visualizations.
Size of Dataset	No such limit in SSRS. It can connect to much larger datasets as compared to Power BI.	Power BI has a limit of 1GB data in free version.
Data Sources	SSRS covers a vast range of data sources to connect with for BI reporting.	Power BI covers most of the data sources available in Tableau. It is closely integrated with Office 365, hence provides connectivity to SharePoint. Power BI online version also supports direct visualization on Search Engine, though, only Bling is supported at this point.
Costing	SSRS pricing details are available only upon request.	Power BI provides a free version with 1GB limit on dataset. Power BI Pro is also a cheaper solution when compared

		with any other BI tool.
License and Pricing	SQL Server Enterprise License. It is available on cloud - AWS, Azure, and other providers.	Power BI: Free 1 GB storage 10k rows/hour data streaming Power BI Pro: USD9.99/user/month 10 GB storage 1 million rows/hour
Implementation	SSRS implementation is complex as compared with Power BI.	Power BI uses cloud storage and includes simple implementation process.

Table 1.2 Power BI vs SSRS [5]

CHAPTER- 2
DATA WAREHOUSING

2.1 DATA WAREHOUSING [6][7][8]

Using real-time data warehousing (RDW) in conjunction with decision support system (DSS) and BI tools is an important way to conduct business processes. The opening vignette demonstrates a scenario in which a real-time active data warehouse supported decision making by analyzing large amounts of data from various sources to provide rapid results to support critical processes. The single version of the truth stored in the data warehouse and provided in an easily digestible form expands the boundaries of DirecTV's innovative business processes. \Xfith real-time data flows, DirecTV can view the current state of its business and quickly identify problems, which is the first and foremost step toward solving them analytically. In addition, customers can obtain real-time information on their subscriptions, TV se1vices, and other account information, so the system also provides a significant competitive advantage over competitors. Decision makers require concise, dependable information about current operations, trends, and changes. Data are often fragmented in distinct operational systems, so managers often make decisions with partial information, at best. Data warehousing cuts through this obstacle by accessing, integrating, and organizing key operational data in a form that is consistent, reliable, timely, and readily available, wherever and whenever needed.
Data warehouse databases provide a decision support system (DSS) environment in which you can evaluate the performance of an entire enterprise over time.

In the broadest sense, the term data warehouse is used to refer to a database that contains very large stores of historical data. The data is stored as a series of snapshots, in which each record represents data at a specific time. By analyzing these snapshots you can make comparisons between different time periods. You can then use these comparisons to help make important business decisions.

Data warehouse databases are optimized for data retrieval. The duplication or grouping of data, referred to as database denormalization, increases query performance and is a natural outcome of the dimensional design of the data warehouse. By contrast, traditional online

transaction processing (OLTP) databases automate day-to-day transactional operations. OLTP databases are optimized for data storage and strive to eliminate data duplication. Databases that achieve this goal are referred to as normalized databases.

An enterprise data warehouse (EDW) is a data warehouse that services the entire enterprise. An enterprise data warehousing environment can consist of an EDW, an operational data store (ODS), and physical and virtual data marts.

A data warehouse can be implemented in several different ways. You can use a single data management system, such as Informix®, for both transaction processing and business analytics. Or, depending on your system workload requirements, you can build a data warehousing environment that is separate from your transactional processing environment.

Informix uses the umbrella terms data warehousing and data warehousing environment to encompass any of the following forms that you might use to store your data:

Data warehouse: A database that is optimized for data retrieval to facilitate reporting and analysis. A data warehouse incorporates information about many subject areas, often the entire enterprise. Typically you use a dimensional data model to design a data warehouse. The data is organized into dimension tables and fact tables using star and snowflake schemas. The data is denormalized to improve query performance. The design of a data warehouse often starts from an analysis of what data already exists and how to collected in such a way that the data can later be used. Instead of loading transactional data directly into a warehouse, the data is often integrated and transformed before it is loaded into the warehouse.

The primary advantage of a data warehouse is that it provides easy access to and analysis of vast stores of information on many subject areas.

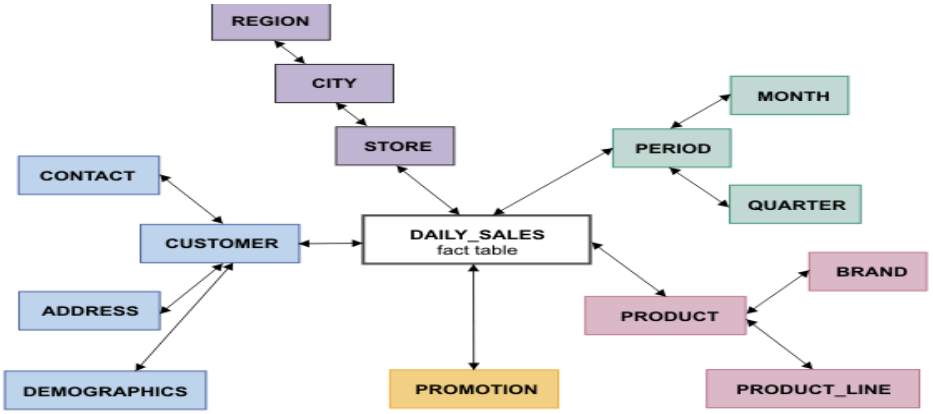


Figure 2.1 A sample snowflake schema which has the DAILY_SALES table as the fact table [8]

Data mart: A database that is oriented towards one or more specific subject areas of a business, such as tracking inventories or transactions, rather than an entire enterprise. A data mart is used by individual departments or groups. Like a data warehouse, you typically use a dimensional data model to build a data mart. For example the data mart might use a single star schema comprised of one fact table and several dimension tables. The design of a data mart often starts with an analysis of what data the user needs rather than focusing on the data that already exists.

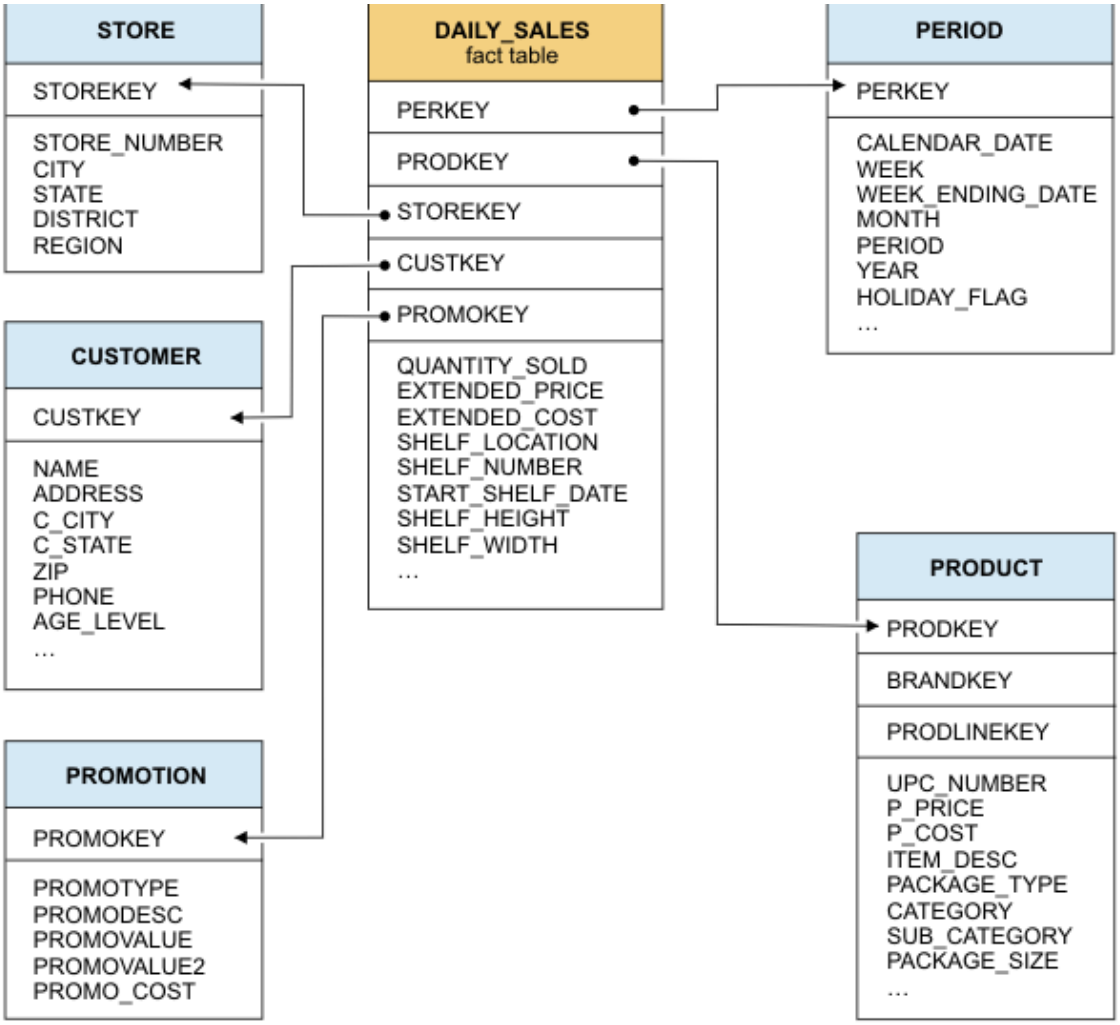


Figure 2.2 A data mart with the DAILY_SALES fact table [8]

Operational data store

A subject-oriented system that is optimized for looking up one or two records at a time for decision making. An operational data store (ODS) is a hybrid form of data warehouse that

contains timely, current, integrated information. Including the ODS in the data warehousing environment enables access to more current data more quickly, particularly if the data warehouse is updated by one or more batch processes rather than updated continuously. The data typically is of a higher level granularity than the transaction. You can use an ODS for clerical, day-to-day decision making. This data can serve as the common source of data for data warehouses.

2.1.1 What does Data Warehouse (DW) mean? [7]

A data warehouse (DW) is a collection of corporate information and data derived from operational systems and external data sources. A data warehouse is designed to support business decisions by allowing data consolidation, analysis and reporting at different aggregate levels. Data is populated into the DW through the processes of extraction, transformation and loading.

The data warehouse architecture was born in the 1980s as an architectural model designed to support the flow of data from operational systems to decision support systems. These systems require analysis of large amounts of heterogeneous data accumulated by companies over time.

In a data warehouse, data from many heterogeneous sources is extracted into a single area, transformed according to the decision support system needs and stored into the warehouse. For example, a company stores information pertaining to its employees, their salaries, developed products, customer information, sales and invoices. The CEO might want to ask a question pertaining to the latest cost-reduction measures; the answers will involve analysis of all of this data. This is a main service of the data warehouse, i.e., allowing executives to reach business decisions based on all these disparate raw data items.

Thus, a data warehouse contributes to future decision making. As in the above example, a firm administrator can query warehouse data to find out the market demand of a particular product, sales data by geographical region or answers other inquiries. This provides insight about required steps to more effectively market a particular product. Unlike an operational data store, a data warehouse contains aggregate historical data, which may be analyzed to reach critical business decisions. Despite associated costs and effort, most major corporations today use data warehouses.

In simple terms, a data warehouse (DW) is a pool of data produced to support decision making; it is also a reposito1y of current and historical data of potential interest to managers throughout the organization. Data are usually structured to be available in a form ready for analytical processing activities (i.e. , online analytical processing [OLAP], data mining, querying, reporting, and other decision support applications). A data warehouse is a subject-oriented, integrated, time-va riant, nonvolatile collection of data in support of management's decision-making process. Characteristics of Data Warehousing A common way of introducing data warehousing is to refer to its fundamental characteristics (Inmon, 2005):

• *Subject oriented*. Data are organized by detailed subject, such as sales, products, or customers, containing only information relevant for decision support. Subject orientation enables users to determine not only how their business is performing, but why. A data warehouse differs from an operational database in that most operational databases have a product orientation and are tuned to handle transactions that update the database. Subject orientation provides a more comprehensive view of the organization.

• *Integrated*. Integration is closely related to subject orientation. Data warehouses must place data from different sources into a consistent format. To do so, they must deal with naming conflicts and discrepancies among units of measure. A data warehouse is presumed to be totally integrated.

• *Time variant (time series)*. A warehouse maintains historical data. The data do not necessarily provide current status (except in real-time systems). They detect trends, deviations, and long-term relationships for forecasting and comparisons, leading to decision making. Eve1y data warehouse has a temporal quality. Time is the one important dimension that all data warehouses must suppo1t. Data for analysis from multiple sources contains multiple time points (e.g., daily, weekly, monthly views).

• *Novolatile*. After data are entered into a data warehouse, users cannot change or update the data. Obsolete data are discarded, and changes are recorded as new data. These characteristics enable data warehouses to be tuned almost exclusively for data access.

Some additional characteristics may include the following:

• *Web based*. Data warehouses are typically designed to provide an efficient computing environment for Web-based applications.

• A data warehouse uses either a relational structure or a multidimensional structure. A recent su1vey on multidimensional structures can be found in Romero and Abell6 (2009).

• *Client/server*. A data warehouse uses the client/server architecture to provide easy access for end users .

• *Real time*. Newer data warehouses provide real-time, or active, data access and analysis capabilities (Basu, 2003; and Bonde and Kuckuk, 2004).

• *Metadata*. A data warehouse contains metaclata (data about data) about how thedata are organized and how to effectively use them.

Whereas a data warehouse is a repository of data, data warehousing is literally the entire process (Watson, 2002). Data warehousing is a discipline that results in applications that provide decision support capability, allows ready access to business information, and creates business insight. The three main types of data warehouses are data marts, operational data stores (ODSs), and enterprise data warehouses (EDWs). In addition to discussing these three types of warehouses next, we also discuss metadata.

2.1.2 Application Case [7]

Enterprise Data Warehouse (EDW) Delivers Cost Savings and Process Efficiencies Founded in 1884 in Dayton, Ohio, the NCR Corporation is now a $5.6 billon NYSE-listed company providing technology solutions worldwide in the retail, financial, insurance, communications, manufacturing, and travel and transportation industries. NCR's technology solutions include store automation and automated teller machines, consulting services, media products, and hardware technology. When acquired by AT&T in 1991, NCR operated on an autonomous country- and product-centric structure, in which each country made its own decisions about product and service offerings, marketing, and pricing and developed its own processes and reporting norms. Under the country model, dozens of different financial and operational applications were required to capture the total results of the company, by no means an enterprise solution. In 1997, when NCR was spun off on its own again, company operations were losing substantial amounts of money every clay. The spin-off provided NCR with the much-needed funds to engage in the deep process changes required to maintain and strengthen its competitive position in the global market and to undertake the transformation to a truly

global enterprise. The goal was to move from a primarily hardware focused integrated, solution-oriented business structure with a global focus. To do this, NCR needed to globalize, centralize, and integrate its vast store of information resources. Only then could it gain effective contribution over the necessary business changes. NCR's EDW initiative was critical to the company's successful transformation and would be vital to the successful deployment of a new worldwide, single-instance, enterprise resource planning (ERP) system planned for several years later. NCR Finance and Worldwide Customer Se1vices (WCS) led the drive for implementation of the EDW. Business teams from NCR Finance and WCS, Financial Information Delivery and Global Information Systems (GIS), respectively, worked closely with the EDW team to ensure that IT understood the business requirements for the new structure. The Teradata system was chosen for its scalability, its flexibility to support to queries and high numbers of concurrent users, and its relatively low maintenance costs. The enormous potential of the EDW spread throughout the company, driving organizational and process changes in NCR Finance, where the financial close cycle was reduced from 14 days to 6 and worldwide reporting integrity standards were established; in \VCS , where individual customer profitability profiles and improvement plans were made possible; and in sales and marketing, operations and inventory management, and human resources. ERP operational standardization and a dramatic improvement in the business of serving its customers mean that NCR is poised for the future . Internally and externally, NCR has become a global solution provider, supported by global business processes. The returns have already been superb. Not only bas the EDW project proved to be more than self funding at the project-cost level, but revenue generation is around the corner. Some of the benefits include $100 million in annual savings in inventory-canying costs, a $200 million sustainable reduction in accounts receivable, a $50 million reduction in annual finance costs, and $22 million in cost savings over the first 5 years of the EDW implementation for WCS. There is still much to be clone and significant value to be realized by the project. Beyond cost savings and process efficiencies, the strategy going forward is to use the EDW to drive growth. Although the EDW project was not undertaken as a profit-producing opportunity, it was self-funding. The cost savings far exceeded the expense of implementation. As the EDW matures, growth-focused goals are developing, and the EDW will drive profits in the future. The quantified benefits of the EDW speak for themselves. There are many more benefits of a qualitative nature. A sampling of both follows.

Qualitative Benefits

• Reduced financial close cycle from 14 days to 6

• Heightened reporting integrity to corporate standards

• Created individual customer profitability profiles and improvement plans

• Provided consistent worldwide reporting processes improved on-time delive1y

• Decreased obsolescence due to enhanced inventory management

Quantified Benefits

• $50 million reduction in annual finance controllership costs

• $200 million sustainable reduction in accounts receivable, which translates into $20 million per year savings in accounts receivable carrying costs

• $100 million sustainable reduction in finished inventory, which, in turn, equals a $10 million per year savings in invento1y carrying costs

• $22 million cost savings over the first 5 years of the EDW implementation for WCS, including

automation of the se1vice level agreement (SLA) reporting to customers, headcount savings, and lower customer maintenance costs

• $10 million for improved supply chain management

• $6.1 million net present value (NPV) of cost reductions over 5 years as a result of reducing headcount from the financial and accounting reporting function

• $3.5 million reduction in telecommunications costs

• $3 million in savings through the reduction of ERP transition costs

• $1. 7 million saved on report development costs in the rollout from Oracle 10.7 and 11 to 11i, for reports that do not have to be custom written for Oracle Source. Teradata, "Enterprise Data Warehouse Delivers Cost Savings and Process Efficiencies," teradata.com/t/resources/case-studies/NCR-Corporation-eb4455/ (accessed June 2009).

2.1.3 Tutorial 1 [5]

Power BI - Data Modeling

Data Modeling is one of the features used to connect multiple data sources in BI tool using a relationship. A relationship defines how data sources are connected with each other and you can create interesting data visualizations on multiple data sources.

With the modeling feature, you can build custom calculations on the existing tables and these columns can be directly presented into Power BI visualizations. This allows businesses to define new metrics and to perform custom calculations for those metrics.

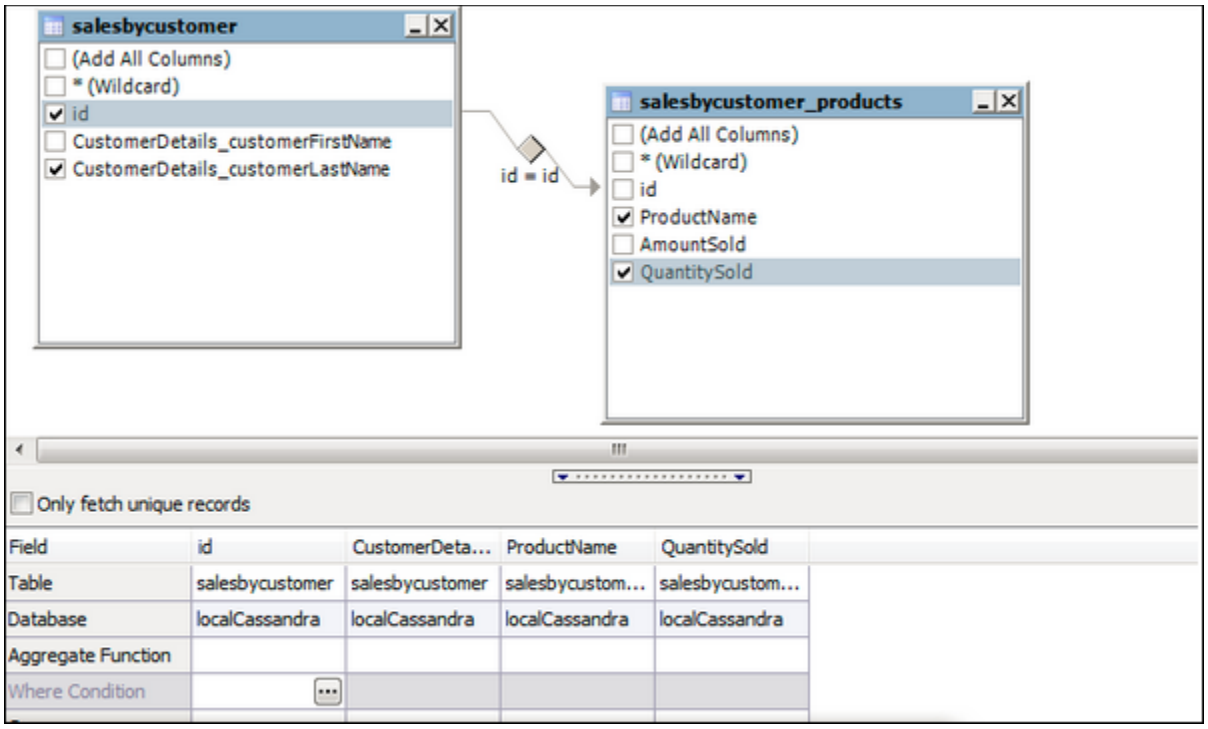


Figure 2.3 Common data model

In the above image, you can see a common data model, which shows a relationship between two tables. Both tables are joined using a column name “Id”.Similarly, in Power BI, you set the relationship between two objects. To set the relationship, you have to drag a line between the common columns. You can also view the “Relationship” in a data model in Power BI. To create data model in Power BI, you need to add all data sources in Power BI new report option. To add

a data source, go to the Get data option. Then, select the data source you want to connect and click the Connect button.

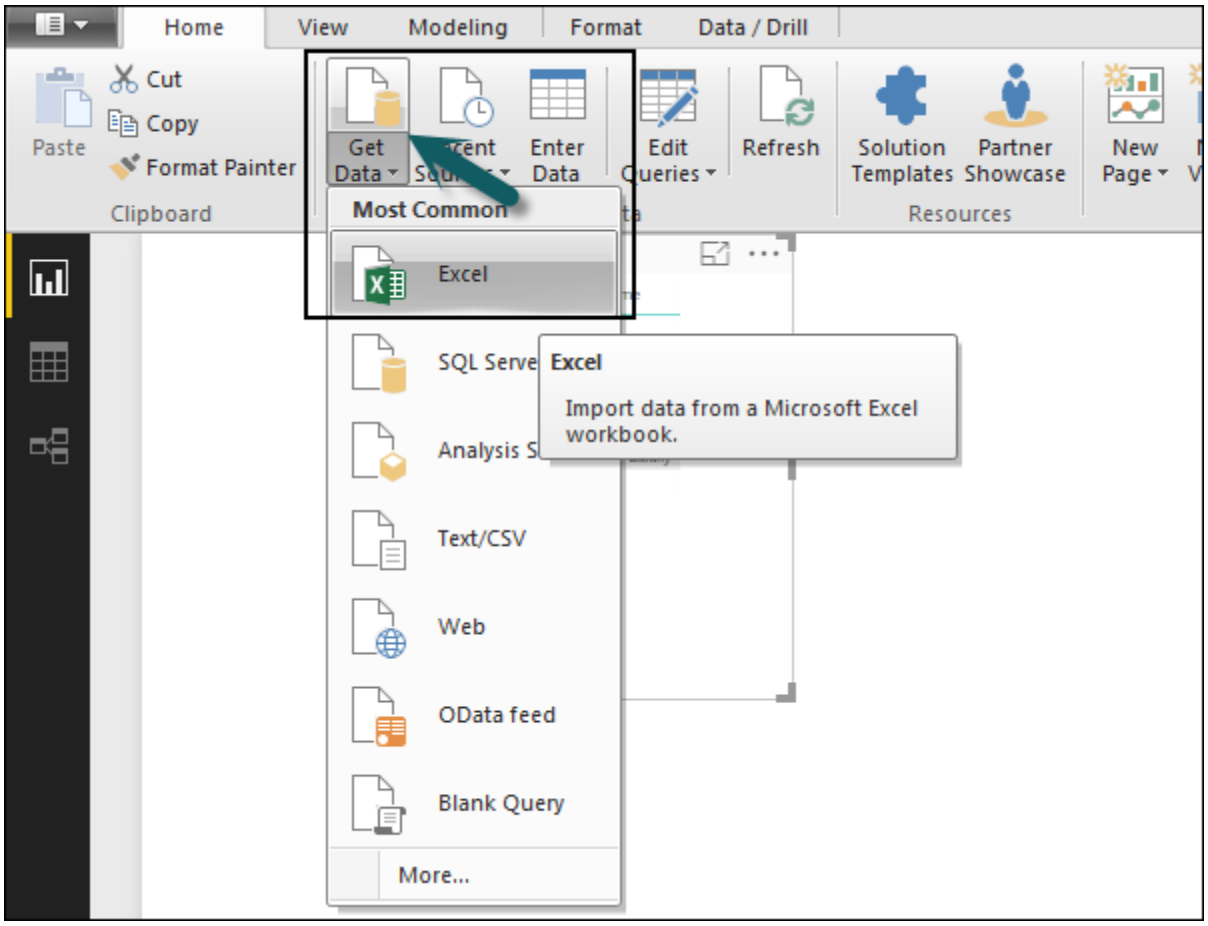


Figure 2.4 Get Data

Once you add a data source, it is presented on the right side bar. In the following image, we have used 2 xls file to import data - Customer and Product.

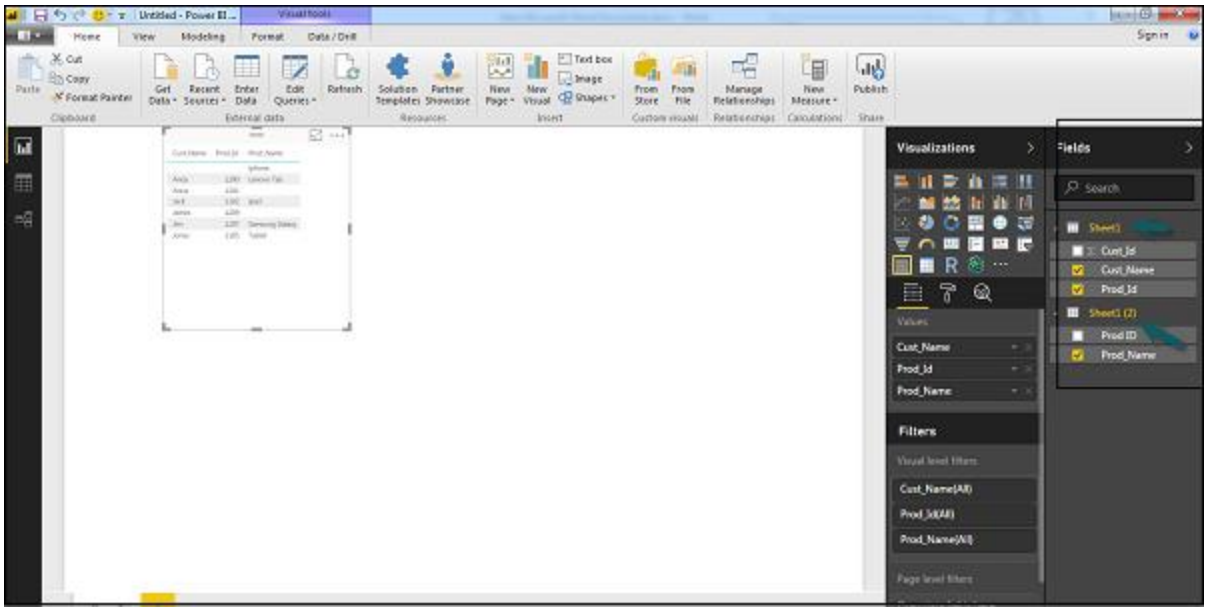


Figure 2.5 Visualization

In Power BI on the left side of the screen, you have the following three tabs :

Report

Data

Relationships

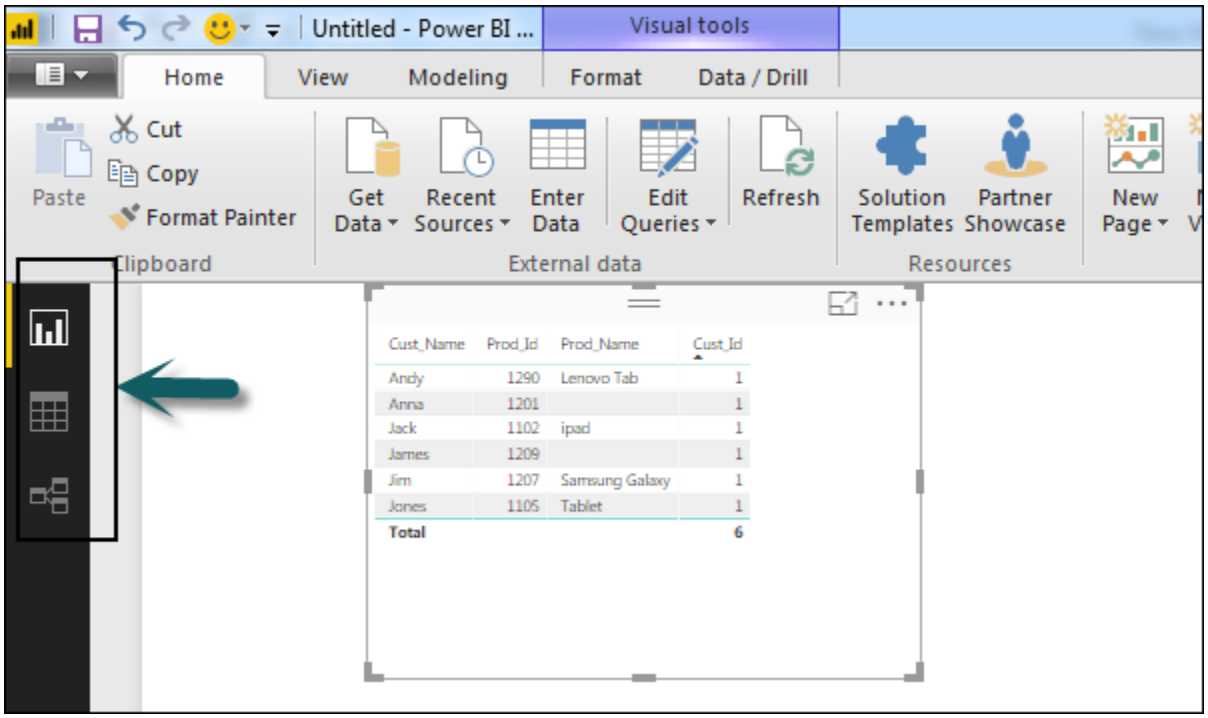


Figure 2.6 Visual tools

When you navigate to the Report tab, you can see a dashboard and a chart selected for data visualization. You can select different chart types as per your need. In our example, we have selected a Table type from available Visualizations.

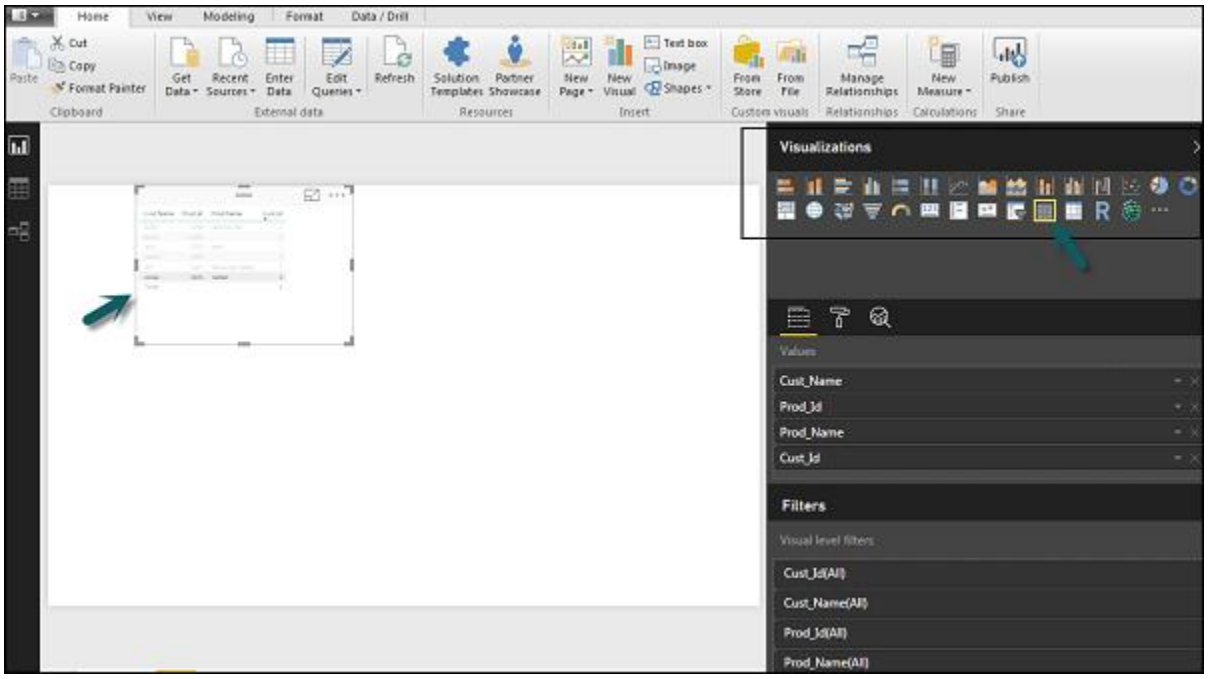
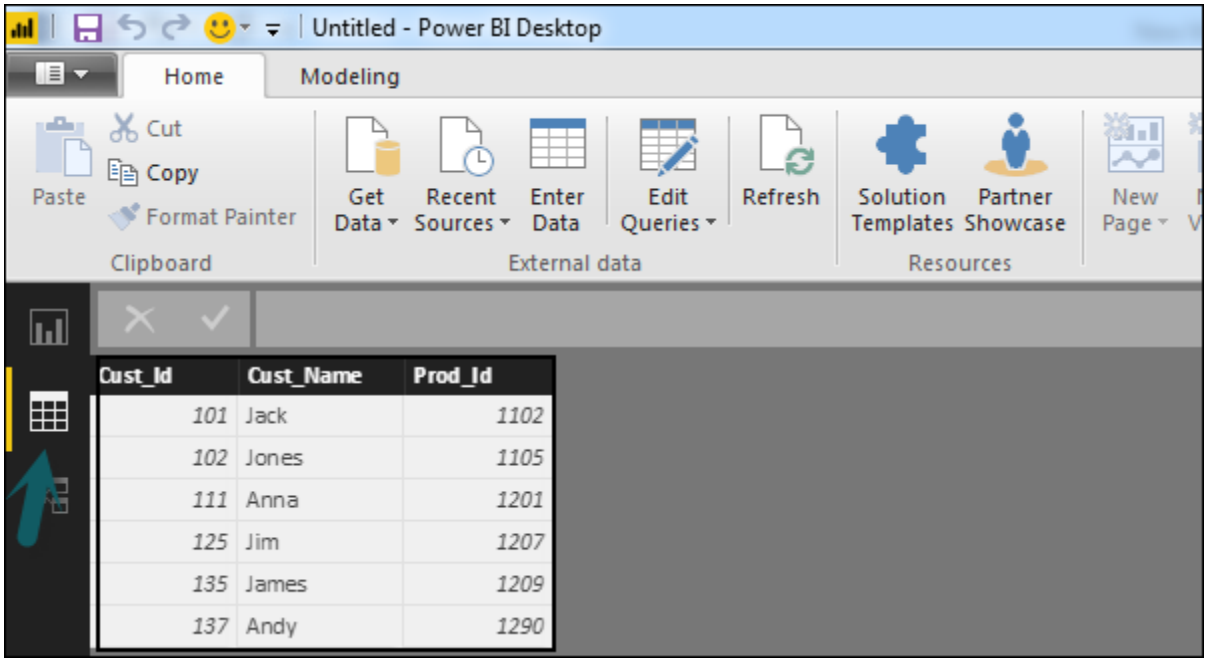


Figure 2.7 Report tab

When you go to the Data tab, you can see all the data as per the defined Relationship from the data sources.



Cust_Id	Cust_Name	Prod_Id
101	Jack	1102
102	Jones	1105
111	Anna	1201
125	Jim	1207
135	James	1209
137	Andy	1290

Figure 2.8 Data source

In the Relationship tab, you can see the relationship between data sources. When you add multiple data sources to Power BI visualization, the tool automatically tries to detect the relationship between the columns. When you navigate to the Relationship tab, you can view the relationship. You can also create a Relationship between the columns using Create Relationships option.

Figure 2.9 Relationship Detection

You can also add and remove relationships in data visualization. To remove a relationship, you have to right-click and select the "Delete" option. To create a new "Relationship", you just need to drag and drop the fields that you want to link between the data sources.

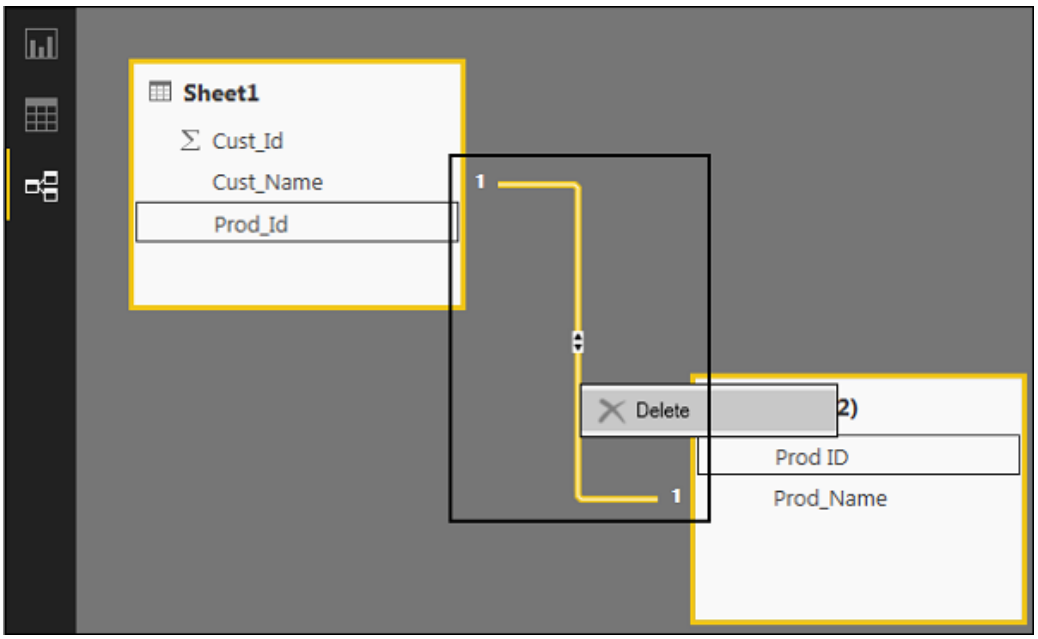


Figure 2.10 Prod_Id

You can also use the Relationship view to hide a particular column in the report. To hide a column, right-click on the column name and select the “Hide in report view” option.

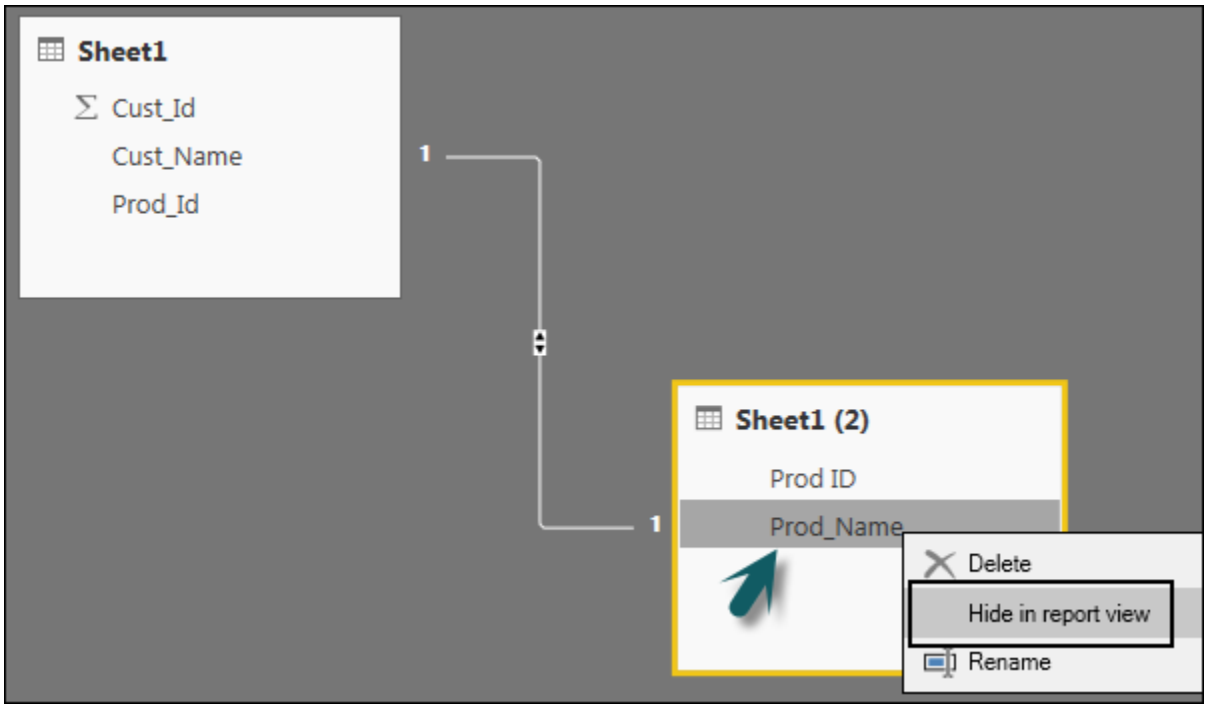


Figure 2.11 Hide in report view

Creating Calculated Columns

You can create calculated columns in Power BI by combining two or more elements of the existing data. You can also apply calculation on an existing column to define a new metric or combine two columns to create one new column.

You can even create a calculated column to establish a relationship between the tables and it can also be used to setup a relationship between two tables.

To create a new calculated column, navigate to Data View tab on the left side of the screen and then click Modeling.

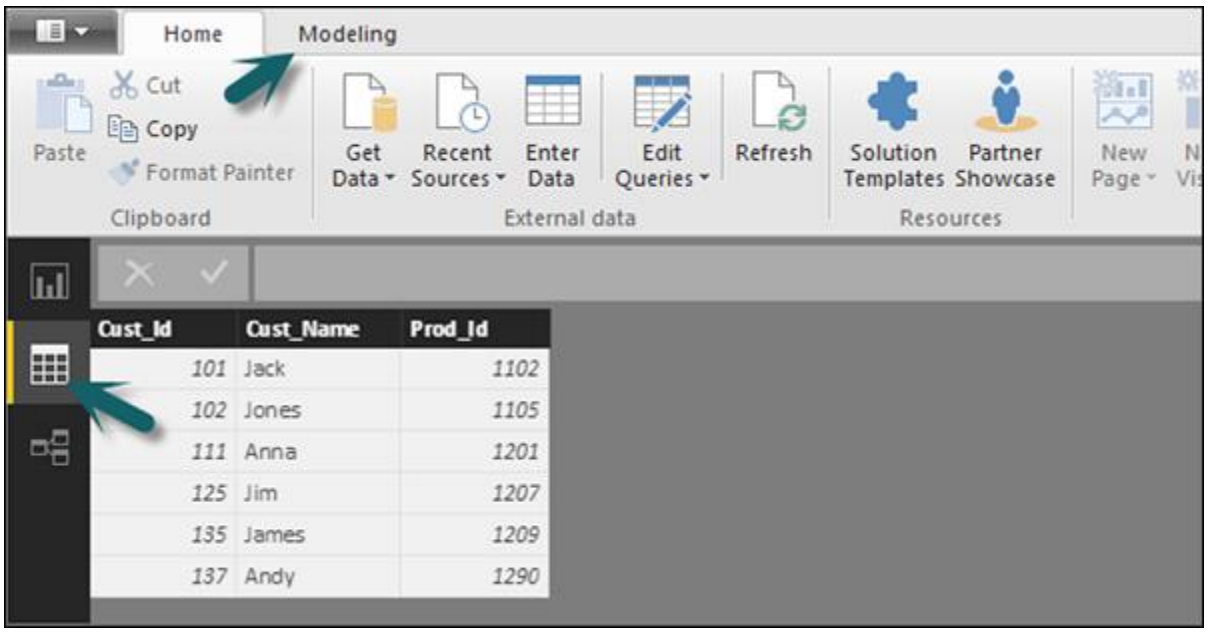


Figure 2.12 Modeling

When you navigate to the Modeling tab, you can see a New Column option at the top of the screen. This also opens the formula bar, where you can enter DAX formula to perform calculation. DAX- Data Analysis Expression is a powerful language also used in Excel to perform calculations. You can also rename the column by changing the Column text in the formula bar.

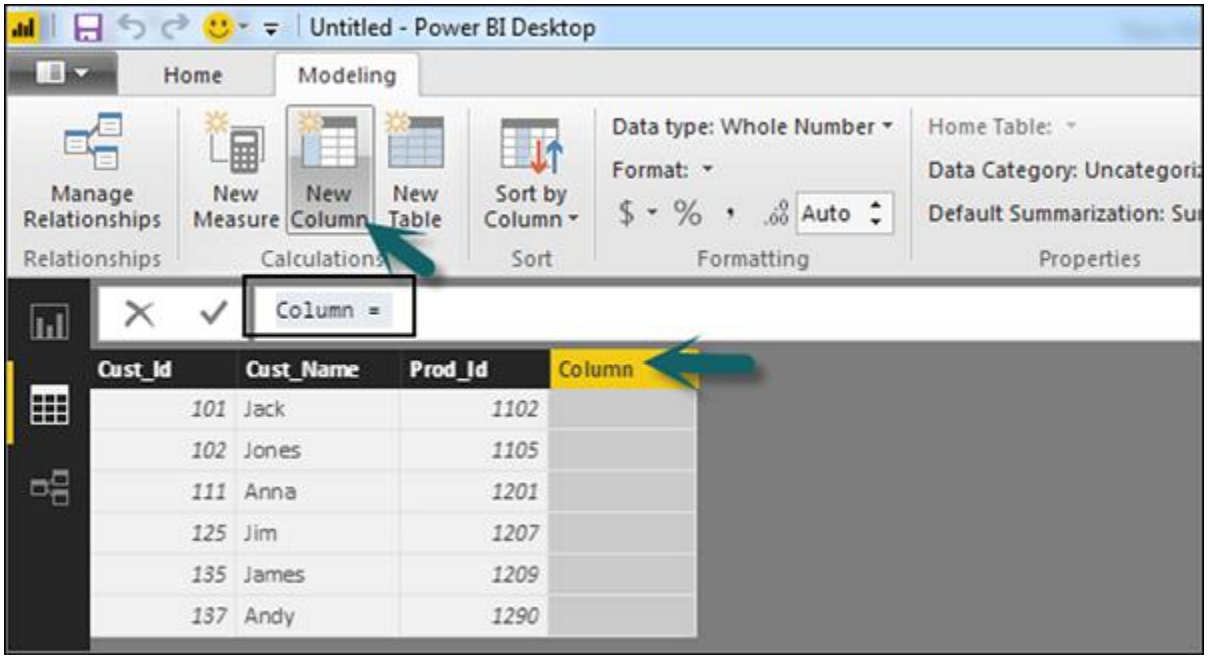


Figure 2.13 New Column

In the following example, let us create a new column: Product Code (Product_C), which is derived from the last three characters of Prod_Id column. Then, write the following formula:

Product_C = RIGHT(Sheet1[Prod_Id],3)

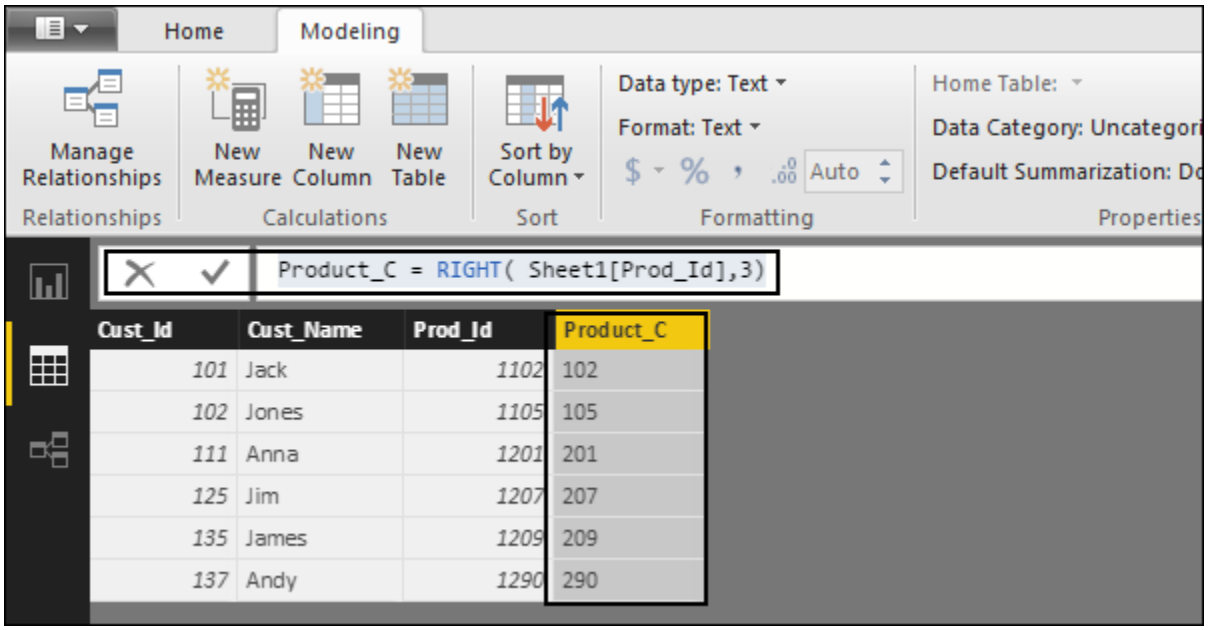


Figure 2.14 Product_C = RIGHT(Sheet1[Prod_Id],3)

A long list of formulas is also provided that you can use for creating calculated columns. You have to enter the first character of formula to be used in calculations as shown in the following screenshot.

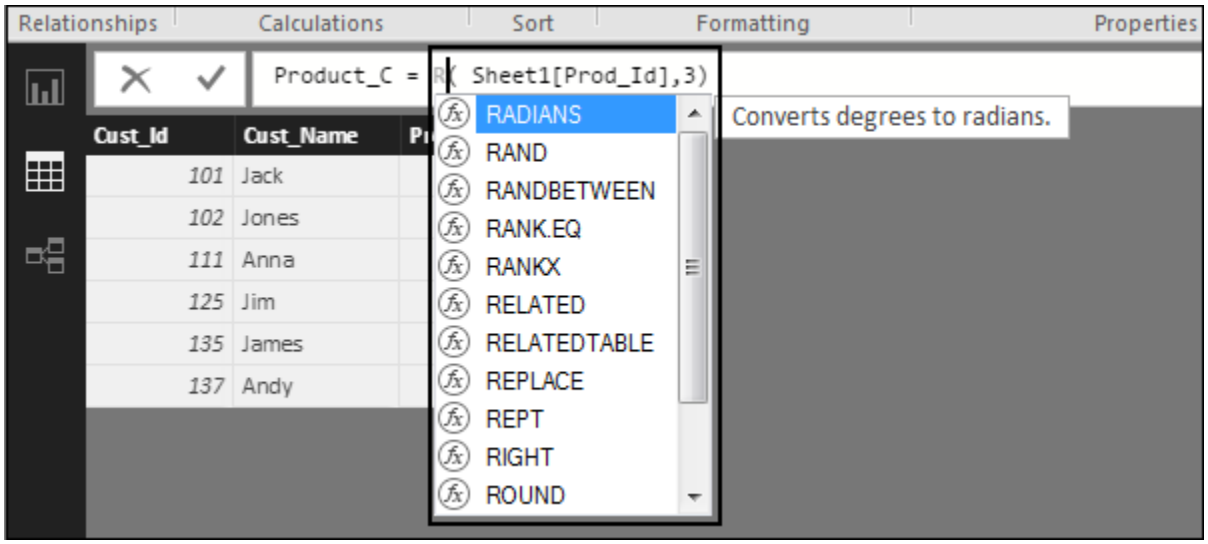


Figure 2.15 Calculations

Creating Calculated Tables

You can also create a new calculated table in data modeling in Power BI. To create a new table, navigate to the Data View tab on the left side of the screen, and then go to the Modeling option at the top of the screen.

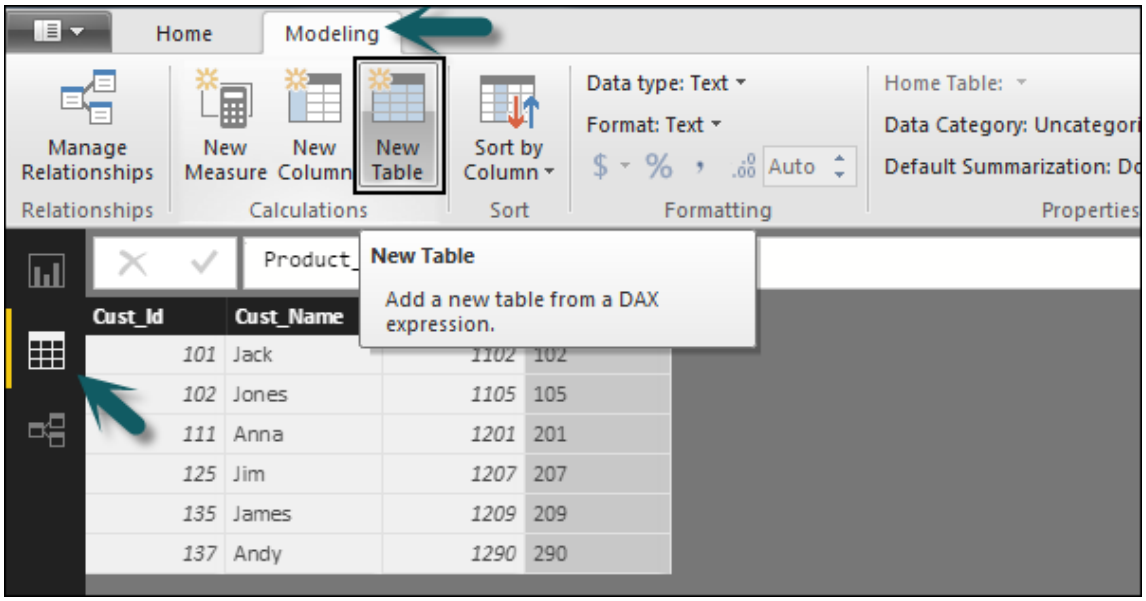


Figure 2.16 Modeling

DAX expression is used to create the new table. You have to enter the name of a new table on the left side of the equal sign and DAX formula to perform the calculation to form that table on the right. When the calculation is complete, the new table appears in the Fields pane in your model.

In the following example, let us define a new table - Table_CustC that returns a one column table containing unique values in a column in another table.

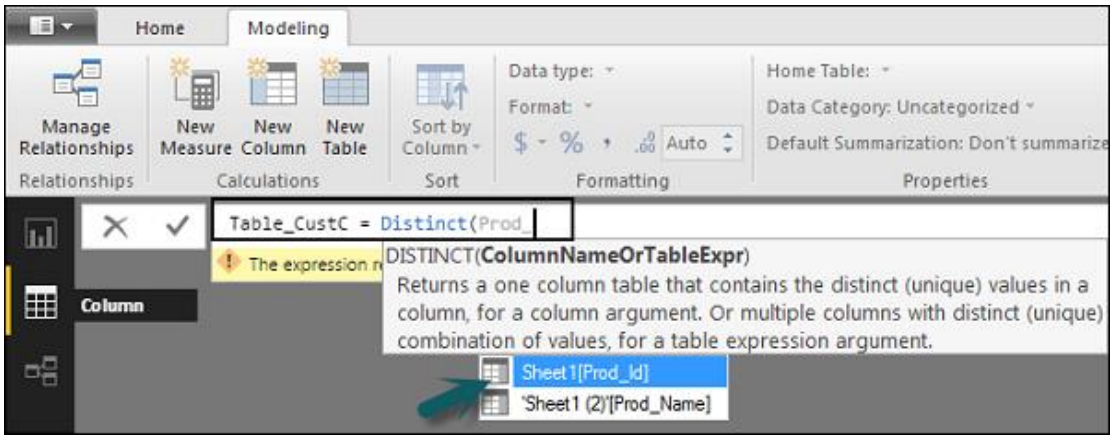


Figure 2.17 Define a new table

A new table is added under the “Fields” section in Power BI screen as shown in the following screenshot. Once the calculated column and calculated tables are created as per your requirement, you can use the fields in the Report tab in Power BI.

To add these objects, you have to select a checkbox and a relationship is automatically detected if possible. If not, then you can drag the columns that you want to connect.

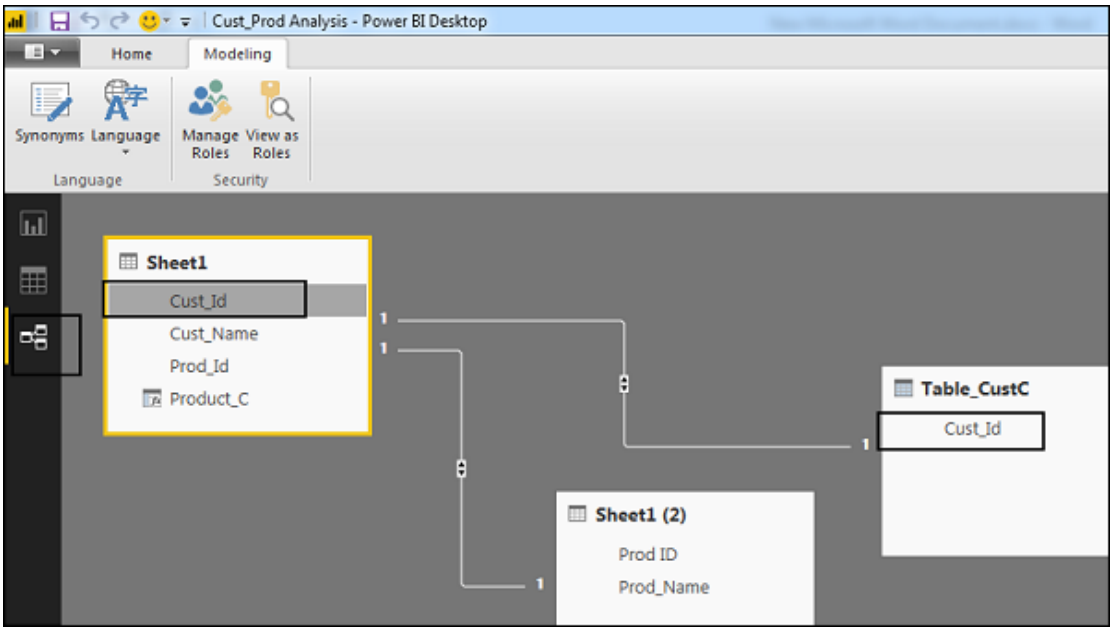


Figure 2.18 Connect columns

To view the report, you navigate to the Report tab and you can see both “Calculated columns” and fields from the new “Calculated table” in the report view.

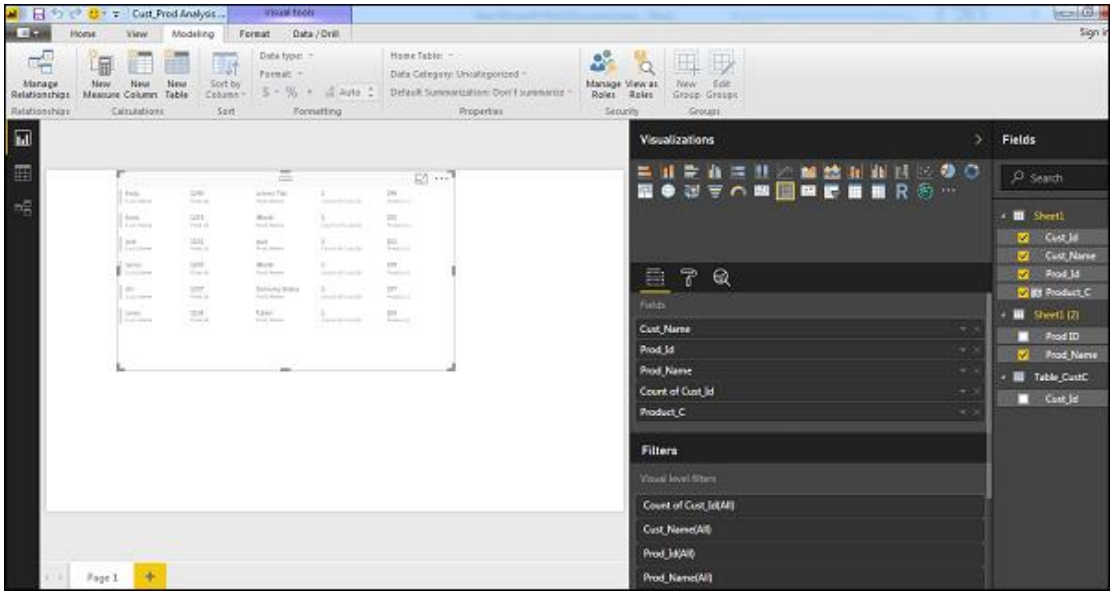
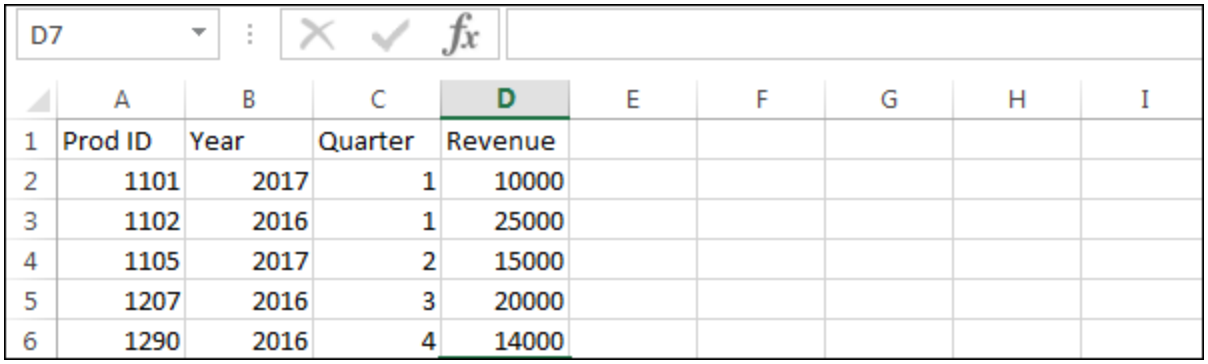


Figure 2.19 Calculated table

Managing Time-Based Data

Power BI allows to drill through time-based data by default. When you add a date field in your analysis and enable drill on your data visualization, it takes you to the next level of time-based data.

Let us consider we have added Time-based table in Power BI visualization. We have added Revenue and Year column in our report.



	A	B	C	D	E	F	G	H	I
1	Prod ID	Year	Quarter	Revenue					
2	1101	2017	1	10000					
3	1102	2016	1	25000					
4	1105	2017	2	15000					
5	1207	2016	3	20000					
6	1290	2016	4	14000					

Figure 2.20 Revenue and Year column

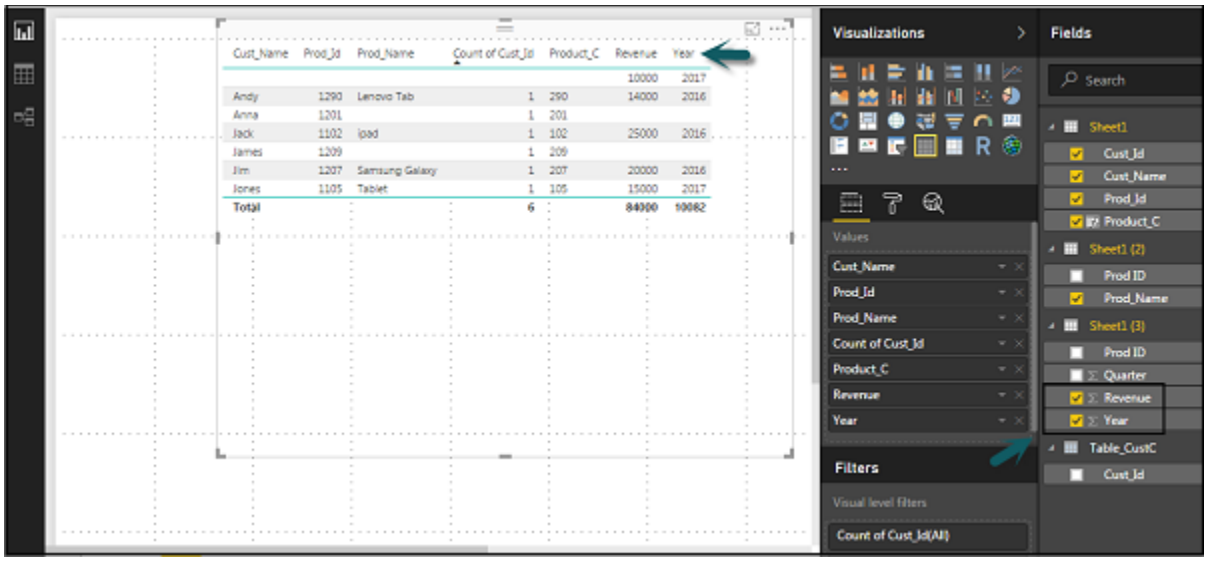


Figure 2.21 Revenue and Year column

We can enable the drill feature in visualizations using the option at the top. Once we enable the drill feature and click the bars or lines in the chart, it drills down to the next level of time hierarchy. Example: Years → Quarters → Months.

We can also use Go to the next level in the hierarchy option to perform a Drill.

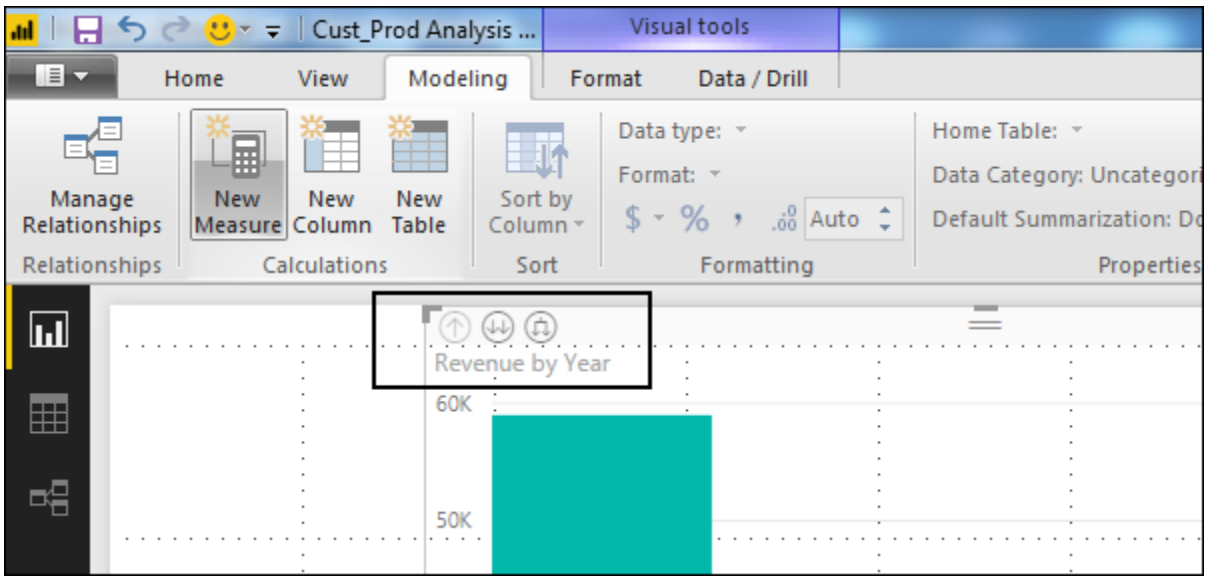


Figure 2.22 Drill

CHAPTER- 3
BUSINESS PERFORMANCE MANAGEMENT

3.1 BUSINESS PERFORMANCE MANAGEMENT BPM [7][10]

Business performance management (BPM) is an outgrowth of decision support system (DSS), enterprise information systems (EIS): and business intelligence (BI). From a market standpoint, it was over 25 years m the making. As with decision support, BPM is more than just a technology. It is an integrated set of processes, methodologies, methods, and applications designed to drive the overall financial and operational performance of an enterprise. It helps enterprises translate their strategies and objectives to plans, monitor performance against those plans, analyze variations between actual results and planned results, and adjust their objectives and actions in response to this analysis.

3.1.1 Understanding Business Performance Management [10]

The business world is constantly evaluating its methods to find business processes that are more efficient in terms of cost and the achievement of goals. The practice of creating metrics to measure performance is just one of the ways that business owners and managers attempt to get a better return on investment for their business processes. Business performance management is a way of monitoring the methods a company uses to reach its goals and then using data to find better methods. The idea of monitoring management procedures to develop effective methods for reaching goals has been around since business first began. Even the great warriors of ancient China understood the need to monitor processes and adjust based on the results. Business performance management was developed as a way to streamline this monitoring process and develop a more efficient way of achieving corporate goals.

What Is Business Performance Management?

Business performance management is the act of setting corporate goals, monitoring the methods used to achieve those goals, and then creating ways for managers to more effectively achieve those goals. By collecting and analyzing data, a company can determine what effects managerial changes had on performance and then alter those changes to help create a more

effective process. The idea of business performance management is a broad concept, but it is best used to analyze specific goals and help a company to save on operating costs, while generating more revenue at the same time. The important thing to remember about business performance management is that it is used to improve the performance of personnel and management. The use of metrics is just a means to an end, with that end being higher profitability.

The Three Main Activities

Each business performance management monitoring program utilizes three primary activities: selection of goals, consolidation, and intervention. Each activity works with the other to help create a more efficient process. This is an extremely dynamic system where each activity affects the other and they are all working together to help develop better business processes.

Selection of Goals

The selection of goals is actually an ongoing process that can be altered by the results that are achieved through intervention. The best place to start with any business performance management program is to set corporate goals and then determine the policies and methods that will be used to achieve those goals. Once the program is underway, the effects that changes have on the process will start to alter the goals. If a managerial decision helped to increase productivity, then it may be necessary to enhance the goal. The point of the goal is to give management a measuring stick to use when it comes to determining success.

Information Monitoring

The second activity involved in business performance management is information monitoring, also known as information consolidation. This is the part of the process where the data is gathered and the pertinent data is analyzed and used to develop a better way to do business. The list of metrics used to create the data varies by company and by project, but the data becomes a critical part of the performance management process.

Managerial Adjustments

Once the data has been reviewed, the management staff decides which measures to take to increase efficiency and profitability. These changes are charted and the effects they have go back into the information monitoring activity. It is critical that the adjustments made reflect the goals of the company. This can be tricky because the goals are not always financial. For example, if the goal is to improve employee job satisfaction by 20 percent, then the actions taken by management would not necessarily require a financial consideration.

The Considerations of Implementation

When a company designs a business performance management program, there are several considerations that have to be made. Not only is company productivity a consideration, but there is also concern for:

- Investors
- Vendors
- Partners
- Competition

Will the goals of the business performance management program weaken the company and allow the competition to start taking more market share? How much risk is the company willing to assume to reach these goals? Before implementing a business performance management plan, a company has to think of how such a plan will affect every aspect of its business.

3.1.2 BPM and Bl Compared [7]

BPM is an outgrowth of BI and incorporates many of its technologies, applications, and techniques. When BPM was first introduced as a separate concept, there was confusion about the differences between BPM and BI. Was it simply a new term for the same concept? Was BPM the next generation of BI, or were there substantial differences between the two? The confusion persists even today for the following reasons:

• BPM is promoted and sold by the same companies that market and sell the BI tools and suites.

• BI has evolved so that many of the original differences between the two no longer exist (e.g., BI used to be focused on departmental rather than on enterprise-wide projects).
• BI is a crucial element of BPM.

The term BI now describes the technology used to access, analyze, and report on data relevant to an enterprise. It encompasses a wide spectrum of software, including ad hoc querying, reporting, online analytical processing (OLAP), dashboards, scorecards, search, visualization, and more. These software products started as stand-alone tools, but BI software vendors have incorporated them into their BI suites.

BPM has been characterized as "BI + Planning," meaning that BPM is the convergence of BI and planning on a unified platform-the cycle of plan, monitor, and analyze (Calumo Group, 2009). The processes that BPM encompasses are not new. Virtually every medium and large organization has processes in place (e .g., budgets, detailed plans, execution, and measurement) that feed back to the overall strategic plan, as well as the operational plans. What BPM adds is a framework for integrating these processes, methodologies, metrics, and systems into a unified solution.

BI practices and software are almost always part of an overall BPM solution. BPM, however, is not just about software. BPM is an enterprise-wide strategy that seeks to prevent organizations from optimizing local business at the expense of overall corporate performance. BPM is not a "one-off' project or departmentally focused. Instead, BPM is an ongoing set of processes that, if done correctly, impacts an organization from top to bottom.

Critical to the success of BPM is alignment throughout an organization. It "helps users take action in pursuit of their 'common cause': achieving performance targets, executing company strategy, and delivering value to stakeholders" (Tucker and Dimon, 2009).

This is not to say that a BI project cannot be strategically oriented, centrally controlled, or impact a substantial part of an organization. For example, the Transportation Security Administration (TSA) uses a BI system called the Performance Information System (PIMS) to track passenger volumes, screen performance (attrition, absenteeism, overtime, and injuries), dangerous items, and total passenger throughput (Henschen, 2008). The system is built on BI software from MicroStrategy (microstt·ategy.com) and is used by over 2,500 "power users" on a daily basis and 9,500 casual users on a weekly basis. The information in PIMS is critical to the operation of the TSA and in some cases is mandated by Congress. It is used by TSA

employees from the top of the agency to the bottom, and it reduced agency costs by approximately $100 million for fiscal year 2007-2008. Clearly, the system has strategic and operational importance. However, it is not a BPM system. The primarly distinction is that a BPM system is strategy driven. It encompasses a closed-loop set of processes that link strategy to execution in order to optimize business performance (see Figure 3.1). The loop implies that optimum performance is achieved by setting goals and objectives (i.e., strategize), establishing initiatives and plans to achieve those goals (i.e., plan), monitoring actual performance against the goals and objectives (i.e ., monitor), and taking corrective action (i.e., act and adjust).

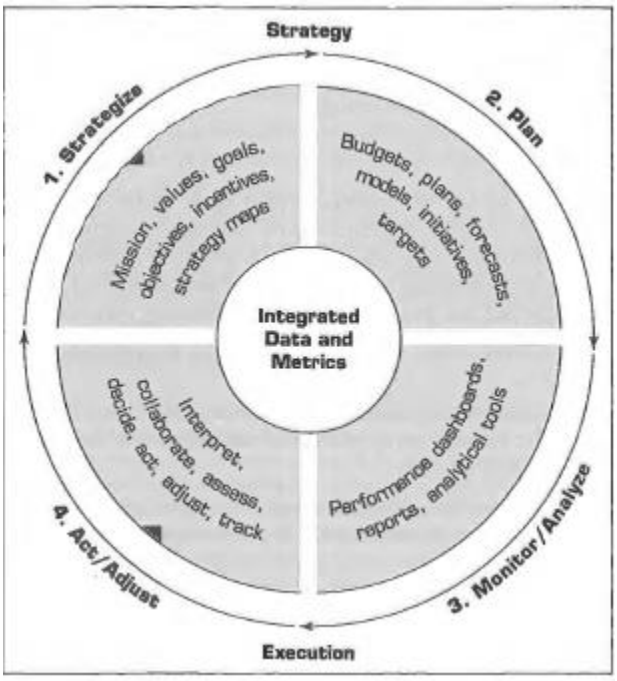


Figure 3.1 BPM Cycle. Source: W. Eckerson, "Performance Management Strategies: How to Create and Deploy Performance Management Strategies." TOW/ Best Practices Report, 2009. [7]

3.1.3 Business Process Performance Management: Case study [11]

At the process level, three perspectives are interpreted: Process targets refer to external users (such as sale, service etc.) and can be derived from the organisation’s targets and other user requirements. Process targets related to internal users (such as planning, budgeting, etc.) can be derived from internal user requirements. Process design is primarily required for achieving the set process target. The process needs to be an appropriately structured, formed, rational, logical,

relevant way of achieving set goals and purposes. According to this concept, process management is the third most important perspective. A process must be managed appropriately.

Process management includes:

1. Target management (including creating functional sub-targets in each critical process stage);
2. Performance management (including obtaining regular feedback on process outputs, monitoring the actual performance by measurement dimensions set in targets, providing feedback, identifying and correcting the process shortcomings, and resetting targets so that they reflect the current user requirement and internal limitations);
3. Resource management (including support to each step in the process of managing equipment, human resources and also budget required for achieving set goals in these process stages;
4. Process interface management (including having interface managed between process stages, and especially, at transfer points between functions) (Harmon, 2003)[11].

3.1.3.1 The Control System in Business [11]

To manage business process performance properly, it is necessary to monitor and control process implementation and execution throughout their lifecycle. Modern control models are highly complex, as they include several aspects of BPM. The model shown in Figure 3.1 illustrates the complexity of the business performance management controlling system, and includes: process modeling and implementation, planning, monitoring, measuring and performance enhancement.

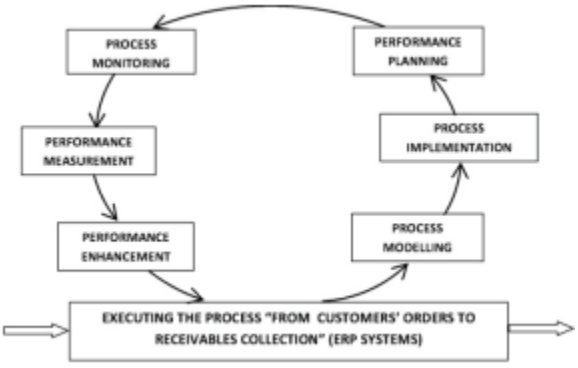


Figure 3.2 Complexity of the business performance management controlling system [11]

Business process management leads to business innovation and optimization through implementing business strategies by way of modeling, developing, deploying and managing business processes throughout the business cycle. Performance planning, monitoring, measuring and enhancing is the essence of performance control. In the performance management process, process monitoring implies constant observing, surveying and tracking an organization's activities, processes and segments, an also the effect of these activities, in order to insight into the scale and rate of progress towards achieving the set objectives and producing specified desired results. The purpose of monitoring is to view a broad range of events in the organization's environment and the organization itself that could make an impact on the course of process performance planning and achieving the planned targets. This requires appropriate performance measurement, performance assessment, comparison between the achieved and the planned, and providing appropriate feedback on achievements. Comparison of achieved performance with planned, i.e. expected performance identifies discrepancies. Analyzing these discrepancies in performance will result in implementing measures contributing to planning performance enhancement. Implementing the performance implementation plan and monitoring enhanced progress leads to continuous business process performance enhancements (Devis & Brabander, 2009). Performance control creates a close integration of operative and analytic environment, corporate and IT environment, and integration of strategies and daily operation. A unified business process management system combines business processes, information and IT resources, coordinating the main parts of the organization's assets: staff, information, technology and processes, with the aim of creating a unified view of real-time intelligence approach.

3.1.3.2 Case Study: Managing a Composite Process "from Customers' Orders to Payment collection" [11]

From the control aspect, performance management is, in fact, monitoring the translation of objectives into results, notably products or services required by internal or external users. Results are normally expressed in terms of quality, quantity, timeliness and costs, and it is of essence to provide appropriate compliance of specified results and the organization's set

objectives. In addition, it is necessary to identify the measures used as a basis for assessing the achieved results (Heß, 2005;Kruppke & Bauer, 2005)[11]. In this research, the defined and set general model of a business problem was applied on a so-called composite process, which we named "from customers' orders to payment collection", integrated from the following constituent sub processes:

- Receiving customers' orders;
- Assessing the ability to fulfil the order;
- Assessing the customers' creditworthiness:
- Issuing delivery orders;
- Preparing delivery;
- Loading goods;
- Despatching and transport;
- Delivery of goods;
- Invoicing;
- Collecting payment.

The process is modeled in ARIS tool, and performance measures were set with determined appropriate performance indicators.

3.1.3.3 Ways of Resolving a Business Problem: Research Methods [11]

Process Modeling

To achieve process measurability in the objective achievement context, all activities in the process and their progress must be unequivocally defined. The starting point in business process modeling is defining the execution flow for a sequence of activities. Process flow modeling also includes modeling the decision-making nodes in processes, and all the branches in the process. Having built the process flow model, it is necessary to make a more detailed definition of all activities comprising the process. Thus, activities are associated with agents/participants in the activity, inputs and outputs of the activity, as well as automated support to the activity (if any). To operationalize the activities, they can have certain attributes defined, which will be the

starting point for determining performance benchmarks. This includes defining the duration and cost of the activities, frequency and number of agents, etc (Kiraka & Manning, 2005). [11]

Setting Performance Indicators

Performance indicators (PIs) are understood and interpreted as values whose measures describe whether and to what extent the defined and modeled process is executed effectively and efficiently, as well as whether performance measures, as qualitatively and/or qualitatively expressed values obtained by measuring indicators, are achieved in compliance, i.e. against the set objectives of business processes. Performance indicators are a set of measures focused on the process performance aspects that are the most critical for current and future success of this process. The following key PIs have been developed and included in the system for managing the performance of the "from customers' orders to payment collection" business process:

- Order execution indicator (OEI) = the number of executed orders / aggregated number of orders * 100;
- Average order execution time in hours (AOET)= SumNi (despatch time – creation time) /number of orders;
- Percentage of changed orders (PCO) = number of changed orders / total number of orders* 100;
- Order growth indicator (OGI) = number of orders in the current month / number of orders in the previous month * 100;
- Average order value 2 (AOV2) = aggregated monthly sales / monthly number of orders
- Percentage of paid orders within due date (PPOwDD) = (number of collected orders within due date / aggregated number of orders * 100;
- Sales performance (SP) = achieved total sales / planned sales * 100;
- Average order collection time (AVCT) = SumNi (collection time – creation time) / number of collected orders;
- Average loading time (ALT) = SumNi (final distribution time – distribution creation time) / number of loaded orders.

Performance indicators were developed with the nominal group technique (NGT). In order to obtain econometrically valid indicators and a feasible performance measurement model, this technique was used for research in the following methodological steps:

1. Defining performance area and level;
2. Determining the group composition for the nominal group technique;
3. Choosing the TNG session leader;
4. Planning, preparing and implementing the TNG session;
5. Compiling a preliminary indicator list;
6. Assessing and final selection of performance indicators;
7. Specifying data gathering methods for chosen indicators;
8. Implementing key performance indicators;
9. Following and perfecting key performance indicators; and
10. Monitoring, assessing and enhancing key performance indicators.

For each of the set PIs, criteria for performance deviation from performance benchmarks were developed, and stored into the ARIS Process Performance manager base. The term "performance deviation" marks the discrepancies between achieved and targeted performance, i.e. set performance benchmarks, obtained by comparing the achieved performance with the set performance benchmark. The performance of the "from customers' orders to payment collection" process is also understood as a multidimensional construct (as it will be shown by the results, their analysis and interpretation), for performance includes both effectiveness and efficiency, qualitative and quantitative aspects, including behavior and behavior outcomes. In brief, performance includes several essential components for a holistic approach to managing the performance of this process Parmenter (2007).[11]

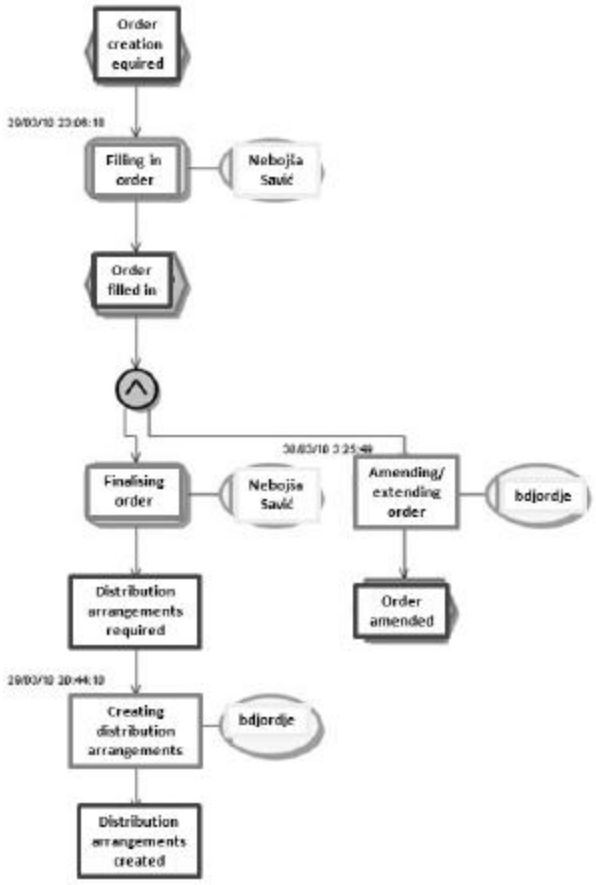


Figure 3.3 "Single instance" of the process in EPC notation [11]

3.1.4 Tutorial 1 [12]

Performance Management Tutorial

Performance management can be defined as a systematic process to improve organizational performance by developing the performance of individuals and teams working with an organization. It is a means of getting better results from the organization, teams and individuals by understanding and managing their performance within a framework of planned goals, standards and competence requirements. In other words, performance management is the process of managing an organization's management strategy. This is how plans are converted into desired outcomes in organizations.

Performance management is a powerful tool

Performance management is a difficult role to play. Some people have difficulty when it comes to performance evaluation. Performance management is about motivation and partnership. When this kind of prospective is shared with your employees and they learn to see in that way, performance management becomes a powerful tool that will help your team to become more successful.

Figure 3.4 Performance management

Performance Management is NOT Human Resource Planning

Performance management is sometimes mistaken for human resources and personnel system, but it is very different when it comes to execution. Performance management comprises of the methodologies, processes, software tools, and systems that manage the performance of an organization, whereas Human Resource Planning only takes care of individual employee's work responsibilities and work delivery.

The benefits of performance management extend to enhancing broad cross-functional involvement in decision-making, and calculated risk-taking by providing greater visibility with accurate and relevant information, to execute an organization's strategy.

Performance management involves many managerial roles, which shows you must be a communicator, a leader and a collaborator as well. Each individual in the team should understand exactly what their responsibilities are and what the expectations from them are, and how to work accordingly to reach the goals.

Scope and Uses

Many organizations jump from one improvement program to another, hoping that one of them will provide that big, elusive result. Most managers would acknowledge that pulling levers for improvement rarely results in a long-term sustained change. The key to improving is integrating and balancing multiple programs sustainably. You cannot break the chain by simply implementing one improvement program and exclude the other programs and initiatives.

There should be a strong bonding between the issues and the strategy of an organization. The manner in which an organization implements performance management can be influenced by its history, goals, mission, vision, strategic priorities, and the various problems it faces in its economic, political, demographic and technological environment.

Performance management is not free floating. If we simplify a little, performance management only exists to help the organization achieve its strategy in the best possible way to help the organization to survive and compete in the market.

Performance management has no end point. Sometimes, for busy, hardworking managers it seems like it is the reason we go through appraisal with staff and get the appraisal process done. Strong and improving performance by individuals and excellent performance management by all managers who are responsible to hold on with their teams are essential to achieving organizational goals.

Research has indicated that a great majority of individuals wants to perform excellently. When managers manage their teams and individual's performance skillfully, this motivates individuals to be proud of what they do. Although this is a big generalization, it does look that most individuals really do want to do a good job, making our leadership in performance management a real-time opportunity.

Performance management is about aligning individual objectives to organizational objectives and ensuring that individuals hold the corporate core values. It provides for expectations to be defined in terms of role responsibilities and accountabilities expected to do, skills expected to have and behavior expected to be.

The overall aim of performance management is to establish a good culture in which individuals and teams take responsibility for the improvement of their own skills and their organizations.

Specifically, performance management is all about achieving the individual objectives according to the organizational objectives and ensuring that every individual is working towards it.

Another aim is to develop the capacity of individuals to meet the expectations of the organization. Mainly, performance management is concerned with the support and guidance for the people who need to develop.

The main points of view towards achieving the aims of performance managements are empowering, motivating and rewarding employees to perform their best for the organization. Focusing on employees' tasks, the right things and make them doing right. Aligning everyone's individual goals towards the goals of the organization. Proactively managing and resourcing performance against objectives of the organizations. Linking job performance to the achievement of the council's corporate strategy and service plans.

The alignment of individual objectives with team, department and corporate plans. The presentation of objectives with clearly defined goals using measures, both soft and numeric. The monitoring of performance and tasking of continuous action as required.

All individuals being clear about what they need to achieve and expected standards, and how that contributes to the overall success of the organization; receiving regular, fair, accurate feedback and coaching to stretch and motivate them to achieve their best.

Performance management is a pre-planned process of which the primary elements are agreement, measurement and feedback.

The following are the characteristics of performance management :

- Measures outputs of delivered performance

It is concerned with measuring outputs of delivered performance compared with expectations expressed as objectives. Its complete focus is on targets, standards and performance measures. It is based on the agreement of role requirements, objectives and performance improvement and personal development plans.

- Concerned with inputs and values

Performance management is also concerned with inputs and values. The inputs are the knowledge, skills and behaviors required to produce the expected results from the individuals.

- Continuous and flexible process

Performance management is a continuous and flexible process that involves managers and those whom they manage acting as partners within a framework that sets out how they can best work together to achieve the required results.

- Based on the principle of management by contract and agreement

It is based on the principle of management by contract and agreement rather than management by command. It relies on consensus and cooperation rather than control or coercion.

- Focuses on future performance planning and improvement

Performance management also focuses on future performance planning and improvement rather than on retrospective performance appraisal. It functions as a continuous and evolutionary process, in which performance improves over the period of time; and provides the basis for regular and frequent dialogues between managers and individuals about performance and development needs.

The following are the main concerns of performance management:

- Concern with outputs, process and inputs

Performance management is concerned with outputs (the achievement of results) and outcomes (the impact made on performance). But it is also concerned with the processes required to achieve these results (competencies) and the inputs in terms of capabilities (knowledge, skill and competence) expected from the teams and individuals involved.

- Concern with planning

Performance management is concerned with planning ahead to achieve success in future. This means defining expectations expressed as objectives and in business plans.

- Concern with measurement and review

If you can't measure it, you can't manage it. Performance management is concerned with the measurement of results and with reviewing progress towards achieving objectives as a basis for action.

- Concern with continuous improvement

Concern with continuous improvement is based on the belief that continuously striving to reach higher standards in every part of the organization will provide a series of incremental gains that will build superior performance.

This means clarifying what organizational, team and individual effectiveness look like and taking steps to ensure that those defined levels of effectiveness are achieved. Establishing a culture in which managers, individuals and groups take responsibility for the continuous improvement of business processes and of their own skills, competencies and contribution.

- Concern with continuous development

Performance management is concerned with creating a culture in which organizational and individual learning and development is a continuous process. It provides means for the integration of learning and work so that everyone learns from the successes and challenges inherent in their day-to-day activities.

- Concern for communication

Performance management is concerned with communication. This is done by creating a climate in which a continuing dialogue between managers and the members of their teams takes place to define expectations and share information on the organization's mission, values and objectives. It establishes mutual understanding of what is to be achieved and a framework for managing and developing people to ensure that it will be achieved.

- Concern for stakeholders

Performance management is concerned with satisfying the needs and expectations of all the organization's stakeholders, management, employees, customers, suppliers and the general public. In particular, employees are treated as partners in the enterprise whose interests are

respected, whose opinions are sought and listened to, and who are encouraged to contribute to the formulation of objectives and plans for their team and for themselves.

- Concern for transparency

Four ethical principles that should govern the operation of the performance management process. These are :

Respect for the individual

Mutual respect

Procedural fairness

Transparency of decision making

What is Performance?

Performance could be defined simply in terms of the achievement of quantified objectives. But performance is not only a matter of what people achieves but also how they are achieving it. A high performance result comes from appropriate behavior and the effective use of required knowledge, skills and competencies.

Performance management must examine how results are attained because this provides the information necessary to consider what needs to be done to improve those results. The concept of performance has been expressed by Brumbrach (1988) as follows: 'Performance means both behaviors and results. Behavior emanates from the performer and transforms performance from abstraction to action.

Not just the instruments for results, behavior is also an outcome in its own right – the product of mental and physical effort applied to tasks – and can be judged apart from results. This definition of performance leads to the conclusion that when managing performance both behavior and results need to be considered.

It is not a question of simply considering the achievement of targets as used to happen in management-by-objectives scheme. Competence factors need to be included in the process. This is the so-called 'mixed model' of performance management, which covers the achievement of expected levels of competence as well as objective setting and review.

- Significance of Performance

Performance is all about the core values of the organization. This is an aspect of behavior but it focuses on what people do to realize core values such as concern for quality, concern for

people, concern for equal opportunity and operating ethically. It means converting espoused values into values in use: ensuring that the rhetoric becomes reality.

- Meaning of Alignment

One of the most important purposes of performance management is to assign individual and organizational objectives. This means what people do at work leads to the achievement of organizational goals.

The real concept of performance is associated with an approach to creating a particular vision of purpose and aims of the organization, which will be helping each employee to understand and recognize their part of responsibilities by the help of which they will manage and enhance the performance of both individuals and the organization.

In an organization, alignment is a flow of objectives from the top to bottom and at each level, team or individual objectives are defined in comparison with higher-level goals. But it also should be a transparent process where individuals and teams are being given the opportunity to set their own goals within the framework defined by the purpose, strategy and values of the organization.

Objectives should be agreed, not set, and this agreement should be reached through the open dialogues that take place between managers and individuals throughout the year. In other words, this needs to be seen as a partnership in which responsibility is shared and mutual expectations are defined.

- Managing Expectations

Performance management is essentially about the management of expectations. It creates a shared understanding of what is required to improve performance and how this will be achieved by clarifying and agreeing what people are expected to do and how they are expected to behave and uses these agreements as the basis for measurement, review and the preparation of plans for performance improvement and development.

- The Significance of Discretionary Behavior

Performance management is concerned with the encouragement of productive discretionary behavior. Discretionary behavior refers to the choices that people make about how they carry out their work and the amount of effort, care, innovation and productive behavior they display. It is the difference between people just doing a job and people doing a great job.

After a research conducted in 2011, researchers found out that the practitioners of performance management were of the following views:

- We are expecting the line managers to recognize performance management as a useful contribution to the management of their teams rather than a chore.
- Managing performance is about coaching, guiding, motivating and rewarding colleagues to help unleash potential and improve organizational performance. Where it works well it is built on excellent leadership and high quality coaching relationships between managers and teams.
- Performance management is designed to ensure that what we do is guided by our values and is relevant to the purposes of the organization.

Guiding Principles of Performance Management

It is necessary to identify any causes that are external to the job and outside the control of either the manager or the individual. Any factors that are within the control of the individual and the manager can then be considered. First, the entire performance management process – coaching, counselling, feedback, tracking, recognition, and so forth – should encourage development. Ideally, team members grow and develop through these interactions. Second, when managers and team members ask what they need — to be able to do to do bigger and better things — they move to strategic development. The researchers also got the following additional views from

practitioners about performance management, a management tool which helps managers to manage. Driven by corporate purpose and values. To obtain solutions that work. Only interested in things you can do something about and get a visible improvement. Focus on changing behavior rather than paperwork. It's about how we manage people – it's not a system. Performance management is what managers do: a natural process of management. Based on accepted principles but operates flexibly. Success depends on what the organization is and needs to be in its performance culture.

Performance Management is NOT Performance Appraisal

It is sometimes assumed that performance appraisal is the same thing as performance management. But there are significant differences. Performance appraisal can be defined as the formal assessment and rating of individuals by their managers at, usually, an annual review meeting. In contrast, performance management is a continuous and much wider, more comprehensive and more natural process of management that clarifies mutual expectations, emphasizes the support role of managers who are expected to act as coaches rather than judges and focuses on the future. Performance appraisal has been discredited because too often, it has been operated as a top-down and largely bureaucratic system owned by the HR department rather than by line managers. It was often backward looking, concentrating on what had gone wrong, rather than looking forward to future development needs. Performance appraisal schemes existed in isolation. There was little or no link between them and the needs of the business. Line managers have frequently rejected performance appraisal schemes as being time consuming and irrelevant. Employees have resented the superficial nature with which appraisals have been conducted by managers who lack the skills required.

Psychological Contract with Performance Management

The concept of psychological contract is a system of beliefs that encompass the actions employees believe are expected of them and what response they expect in return from their employer. It is concerned with assumptions, expectations, promises and mutual obligations. Psychological contracts are 'promissory and reciprocal, offering a commitment to some behavior on the part of the employee, in return for some action on the part of the employer.

A positive psychological contract is one in which both parties – the employee and the employer, the individual and the manager – agree on mutual expectations and pursue courses of action that provide for those expectations to be realized.

A positive psychological contract is worth taking seriously because it is strongly linked to higher commitment to the organization, higher employee satisfaction and better employment relations. Performance management has an important part to play in developing a positive psychological contract.

Performance management processes can help to clarify the psychological contract and make it more positive by providing a basis for the joint agreement and definition of roles.

- Communicating expectations in the form of targets, standards of performance, behavioral requirements (competencies) and upholding core values.
- Obtaining agreement on the contribution both parties have to make to get the results expected.
- Defining the level of support to be exercised by managers.
- Providing rewards that reinforce the messages about expectations.
- Giving employees opportunities at performance review discussions to clarify points about their work.

CHAPTER- 4
DATA MINING FOR BUSINESS INTELLIGENCE

4.1 DATA MINING FOR BUSINESS INTELLIGENCE [13]

Business Intelligence transcends beyond the scope of data, to delve into aspects such as the actual use of insights generated by business leaders. The banner of BI spans across data generation, data aggregation, data analysis, and data visualization techniques, which facilitate business management. In other words, BI entails several processes and procedures to support data collection, sharing, and reporting for better decision-making.

Since the emergence of the first forms of writing until now, people have had the need to collect information; information that, logically, has been accumulated along years, becoming more abundant and profuse over time. Today, the growth of the technology sector has also caused a disproportionate increase in the volume of information data.

Therefore, it is required more sophisticated and complex storage data. Due to the boom of these information technologies, organizations have had to deal new challenges, allowing them to analyze, discover and understand the information beyond this system. And that is where the big data appears.

In this sense, big data is the computerized processing of large amounts of information; data (structured, unstructured and semi structured) that exceeds the capacity of conventional software to be captured, managed and processed within a reasonable time.

It is therefore necessary that the response speed of the system would be as quick as possible, in order to obtain the right information at the right time. These are, precisely, the three main characteristics of a big data project. As there is a wide variety of types of data to analyze, it is necessary to classify this data into different categories, such us Web and Social Media content, Machine-to-Machine (M2M), Big Transaction Data, Biometrics data or Human generated data.

It is important to stand out that the cities are full of data streams. This is where the concept of smart city appears. In it, the innovative use of data helps to provide better and more inventive services to improve people's lives. And it is estimated that about 3000 million people live currently in cities.

In 2050, it is expected that 70% of the population will reside in cities too, by the United Nations. Therefore, people will demand more and more to these cities. Being a project of big data, smart cities would need to capture, store, process and analyze large amounts of data from many different sources, to transform them into useful knowledge.

The extraction of data: Data mining

Taking the necessary data stored, then it is important to consider different techniques of data analysis, such as the association, clustering, text analytics or data mining. Data mining is one of the most important, because it is the process of extracting data, analyzing it from many dimensions or perspectives and producing a summary of the information in a useful form.

Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. There are two types of data mining: descriptive, which gives information about existing data; and predictive, which makes forecasts based on the data. To reach this end, data mining uses statistics and, in some cases, Artificial Intelligence and Neural Networks algorithms.

Basically, data mining arises to try to help understand the content of big data. It is mainly based on:

- Extract, transform, and load transaction data onto the data warehouse system (this one is a process of centralized data management and retrieval).
- Store and manage the data in a multidimensional database system.
- Present the data in a useful format (in a graph or a table).
- Analyze the data by application software.
- Finally, provide data access to business analysts and information technology professionals.

Although data mining is a relatively new concept, the technology is not. Companies have used powerful computers to analyze market research reports for years. However, continuous innovations are dramatically increasing the accuracy of analysis, driving down the cost.

Data mining is presented as an emerging technology, with several advantages: it could be a good meeting point between researchers and business people; on the other hand, it could save

large amounts of money to a company and would open up new business opportunities. That is why data mining is so important. In this point, the next question would be: so, how data mining is used to generate Business Intelligence?

Nowadays, data mining is primarily used by companies with a strong consumer view. Business applications trust on data mining software solutions; due to that, data mining tools are today an integral part of enterprise decision-making and risk management in a company. In this point, acquiring information through data mining alluded to a Business Intelligence (BI).

4.1.1 How Data Mining is used to Generate Business Intelligence [13]

Business Intelligence is the ability to transform data into information and information into knowledge. It is the best way to optimize the decision-making process in business.

In this sense, Business Intelligence is a set of methodologies, applications and technologies to collect, refine and transform this data from transactional systems and unstructured information (internal and external to the company), in structured information for direct exploitation or for analysis.

Business Intelligence combines data analysis applications, including ad hoc analysis and querying, enterprise reporting, online analytical processing (OLAP), mobile BI, real-time BI, operational BI, cloud and software as a service BI, open source BI, collaborative BI and location intelligence.

BI technology also includes data visualization, tools for building BI dashboards and performance scorecards and key performance indicators. In addition, these tools generate findings that are ultimately used to gain competitive advantage over rivals, better and efficient business operations and better survivability and risk management.

Data mining tools provide better customers relationship management, too, through mining real habits and diverse patterns. In resume, Business Intelligence strategy should be used to apply the knowledge to maximize the benefits of the company.

Thus, Business Intelligence acts as a strategic factor for a business, providing insider information to respond to business problems: entering new markets, financial control, cost optimization, production planning, analysis of customer profiles, profitability… That is how data mining is used to generate Business Intelligence.

For example, the potential benefits of Business Intelligence programs include accelerating and improving decision making; optimizing internal business processes; increasing operational efficiency; driving new revenues; and gaining competitive advantages over business rivals. BI systems can also help companies identify market trends and spot business problems that need to be addressed.

In resume, Business Intelligence (BI) is an increasingly popular term representing the tools and systems that play a key role in the strategic planning process of the corporation by turning knowledge into profit.

Data mining and Business Intelligence have made possible that various industries, such as sales and marketing, healthcare organization or financial institutions, could have a quick analysis of data and thereby, improving the quality of decision making process in their industries.

In addition, data mining technologies have bright future in business applications, making possible new opportunities by automated prediction of trends and behaviours in these businesses. So, how data mining is used to generate Business Intelligence is a concept that we will hear a lot during these years: it is the future.

4.1.2 What is Data Science? [14]

The ubiquity, size, value, and importance of big data have given rise to a new profession: the data scientist. Data science is a mix of skills in the areas of statistics, machine learning, math, programming, business, and IT. The term itself is thus broader than the other concepts we discussed above, and it is a rare individual who combines deep skills in all the constituent areas. Harlan Harris, in Analyzing the Analyzers (with Sean Mnrphy and Marek Vaisman, O'Reilly 2013) describes the skill sets of most data scientists as resembling a 'T' -deep in one area (the vertical bar of the T), and shallower in other areas (the top of the T).

At a large data science conference session (Strata-Hadoop World, October 2014) most attendees felt that programming was an essential skill, though there was a sizable minority who felt otherwise. And, although big data is the motivating power behind the growth of data science, most data scientists do not actually spend most of their time working with terabyte-size or larger data.

Data of the terabyte or larger size would be involved at the deployment stage of a model.

There are manifold challenges at that stage, most of them IT and programming issues related to data handling and tying together different components of a system. Much work must precede that phase. It is that earlier piloting and prototyping phase on which this chapter focuses- developing the statistical and machine learning models that will eventually be plugged into a deployed system. What methods do you use with what sorts of data and problems? How do the methods work? What are their requirements, their strengths, their weaknesses? How do you assess their performances?

4.1.3 Why are there so many Different Methods? [14]

As can be seen in this chapter or any other resource on data mining, there are many different methods for prediction and classification. You might ask yourself why they coexist, and whether some are better than others. The answer is that each method has advantages and disadvantages. The usefulness of a method can depend on factors such as the size of the dataset, the types of patterns that exist in the data, whether the data meet some underlying assumptions of the method, how noisy the data are, and the particular goal of the analysis.

A small illustration is shown in Figure 4.1, where the goal is to find a combination of income level and lot size that separate buyers (solid circles) from nonbuyers (hollow circles) of riding mowers. The first method (left panel) looks only for horizontal lines to separate buyers from nonbuyers, whereas the second method (right panel) looks for a single diagonal line.

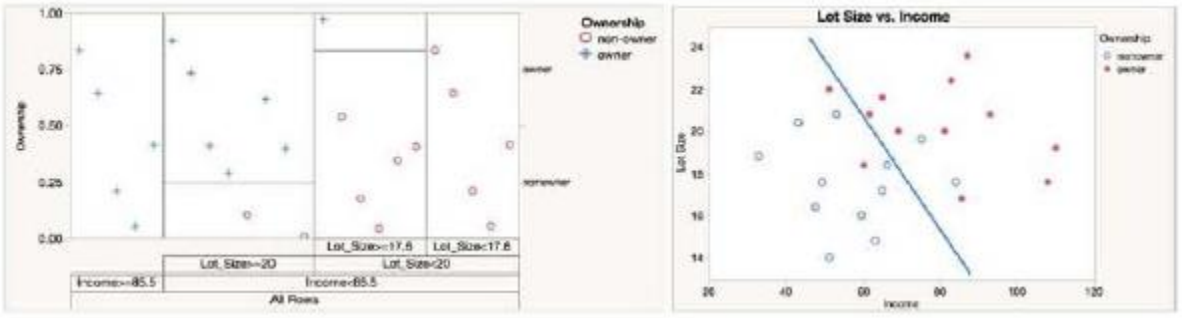


Figure 4.1 Two methods for separating buyers from nonbuyers

Different methods can lead to different results and their performances can vary. It is therefore customary in data mining to apply several different methods and select the one that is most useful for the goal at hand.

4.1.3.1 Terminology and Notation [14]

Because of the hybrid origins of data mining, its practitioners often use multiple terms to refer to the same thing. For example, in the machine learning (artificial intelligence) field the variable being predicted is the output variable or target variable. To a statistician, it is the dependent variable or the response.

Here is a summary of terms used:

Algorithm Refers to a specific procedure used to implement a particular data mining technique: classification tree, discriminant analysis, and the like.
Attribute See Predictor.
Case See Observation.
Confidence Has a broad meaning in statistics (confidence interval), concerning the degree of error in an estimate that results from selecting one sample as opposed to another.
Dependent Variable See **Response**.
Estimation See **Prediction**.
Feature See **Prediction**.
Holdout Sample Is a sample of data not used in fitting a model, used to assess the performance of that model;
Input Variable See **Predictor**.
Model Refers to an algorithm as applied to a dataset, complete with its settings (many of the algorithms have parameters that the user can adjust).

4.1.4 Data in data mining [7]

Data refers to a collection of facts usually obtained as the result of experiences, observations, or experiments. Data may consist of numbers, words, images, and so on as measurements of a set of variables. Data are often viewed as the lowest level of abstraction from which information and knowledge are derived. At the highest level of abstraction, one can classify data as

categorical or numeric. The categorical data can be subdivided into nominal or ordinal data, whereas numeric data can be subdivided into interval or ratio. Figure 4.2 shows a simple taxonomy of data in data mining.

• ***Categorical data*** represent the labels of multiple classes used to divide a variable into specific groups. Examples of categorical variables include race, sex, age group, and educational level. Although the latter two variables may also be considered in a numerical manner by using exact values for age and highest grade completed, it is often more informative to categorize such variables into a relatively small number of ordered classes. The categorical data may also be called discrete data implying that it represents a finite number of values with no continuum between them. Even if the values used for the categorical (or discrete) variables are numeric, these numbers are nothing more than symbols and do not imply the possibility of calculating fractional values.

• ***Nominal data*** contain measurements of simple codes assigned to objects as labels, which are not measurements. For example, the variable marital status can be generally categorized as (1) single, (2) married, and (3) divorced. Nominal data can be represented with binomial values having two possible values (e.g., yes/no, true/fa lse, good/bad), or multinomial values having three or more possible values (e.g., brown/green/blue, white/black/Latino/Asian, single/married/ divorced).

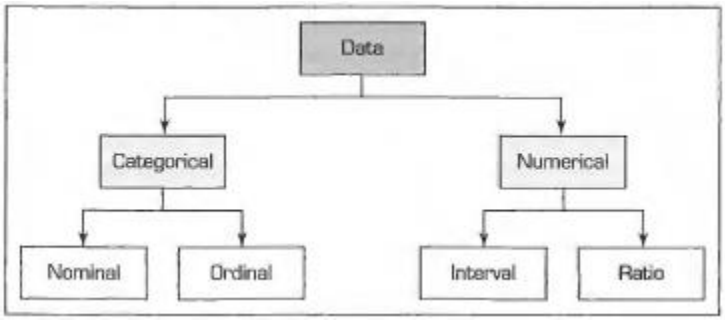


Figure 4.2 A Simple Taxonomy of Data in Data Mining [7]

• ***Ordinal data*** contain codes assigned to objects or events as labels that also represent the rank order among them. For example, the variable credit score can be generally categorized as (1) low, (2) medium, and (3) high. Similar ordered relationships can be seen in variables such as age group (i .e., child , young, middle aged, elderly) and educational level (i.e., high school,

college, graduate school). Some data mining algorithms, such as ordinal multiple logistic regression, take into account this additional rank-order information to build a better classification model.

• ***Numeric data*** represent thee numeric values of specific variables. Examples of numerically valued variables include age, number of children, total household income (in U.S. dollars), travel distance (in miles), and temperature (in Fahrenheit degrees). Numeric values representing a variable can be integer (taking only whole numbers) or real (taking also the fractional number). The numeric data may also be called continuous data, implying the variable contains continuous measures on a specific scale that allows insertion of interim values. Unlike a discrete variable, which represents finite, countable data, a continuous variable represents scalable measurements, and it is possible for me data to contain an infinite number of fractional values.

• ***Interval data*** are variables that can be measured on interval scales. A common example of interval scale measurement is temperature on the Celsius scale. In this particular scale, the unit of measurement is 1/ 100 of the difference between the melting temperature and the boiling tempera of water in atmospheric pressure; there is not an absolute zero value.

• ***Ratio data*** include measurement variables commonly found in me physical sciences and engineering. Mass, lengd1, time, plane angle, energy, and electric charge are examples of physical measures that are ratio scales. The scale type takes its name from the fact that measurement is the estimation of the ratio between a magnitude of a continuous quantity and a unit magnitude of d1e same kind. Informally, the distinguishing feature of a ratio scale is the possession of a no arbitrary zero value. For example, the Kelvin temperature scale has a nonarbitra1y zero point of absolute zero, which is equal to -273.15 degrees Celsius. This zero point is no arbitrary, because the particles comprise matter at temperature have zero kinetic energy.

• ***Other data*** types include elate/ time, unstructured text, image, and audio. These data types need to be converted into some form of categorical or numeric representation before they can be processed by data mining algorithms. Data can also be classified as static or dynamic (i .e., temporal or time series).

4.1.5 Data Mining Techniques [15]

There are many different techniques that can be used to retrieve data, and every technique will garner different results. Which of the various data mining techniques in CRM should be implemented depends on the problem to be solved. Different insights are provided with the application of different data mining techniques, so it's important to first determine what kind of problem you have and how you'd like to solve it.

The objective of data mining is to find information that can be easily understood in order to improve data quality management. Every data management software should be able to provide information from which the user can benefit. Large data sets are impossible to handle manually, so all-around data solutions software is a must.

Important Data Mining Techniques in CRM9

- Anomaly Detection

Searching for information that doesn't match expected behavior or a projected pattern is called anomaly detection (or outliers, surprises, or exceptions). Anomalies can provide actionable information because they deviate from the average in the dataset. An outlier is numerically distant from other data present in the dataset, and indicates that additional analysis is required for data quality improvement to take place.

- Association Rule Learning

This technique enables us to discover relations between data items in huge databases. With association rule learning, hidden patterns are uncovered and the information gained may be used to better understand our customers, learn their habits, and predict their decisions. This is one of the more important data mining techniques in CRM, as it is used in the point-of-sales data.

- Clustering

Clustering is a process that helps us identify similar data sets and allows us to understand both the similarities and differences within the data. Data sets that have similar traits can be used for conversion rate increases. For example, if the buying behavior of one group of customers is similar to that of another group, they can both be targeted with similar services or products.

4.1.6 Data Visualization [16]

Data visualization is a general term that describes any effort to help people understand the significance of data by placing it in a visual context. Patterns, trends and correlations that might go undetected in text-based data can be exposed and recognized easier with data visualization software.

Today's data visualization tools go beyond the standard charts and graphs used in Microsoft Excel spreadsheets, displaying data in more sophisticated ways such as infographics, dials and gauges, geographic maps, sparklines, heat maps, and detailed bar, pie and fever charts. The images may include interactive capabilities, enabling users to manipulate them or drill into the data for querying and analysis. Indicators designed to alert users when data has been updated or predefined conditions occur can also be included.

Importance of data visualization

Data visualization has become the de facto standard for modern business intelligence (BI). The success of the two leading vendors in the BI space, Tableau and Qlik -- both of which heavily emphasize visualization -- has moved other vendors toward a more visual approach in their software. Virtually all BI software has strong data visualization functionality.

Data visualization tools have been important in democratizing data and analytics and making data-driven insights available to workers throughout an organization. They are typically easier to operate than traditional statistical analysis software or earlier versions of BI software. This has led to a rise in lines of business implementing data visualization tools on their own, without support from IT.

Data visualization software also plays an important role in big data and advanced analytics projects. As businesses accumulated massive troves of data during the early years of the big data trend, they needed a way to quickly and easily get an overview of their data. Visualization tools were a natural fit.

Visualization is central to advanced analytics for similar reasons. When a data scientist is writing advanced predictive analytics or machine learning algorithms, it becomes important to visualize the outputs to monitor results and ensure that models are performing as intended. This

is because visualizations of complex algorithms are generally easier to interpret than numerical outputs.

Examples of data visualization

Data visualization tools can be used in a variety of ways. The most common use today is as a BI reporting tool. Users can set up visualization tools to generate automatic dashboards that track company performance across key performance indicators and visually interpret the results.

Many business departments implement data visualization software to track their own initiatives. For example, a marketing team might implement the software to monitor the performance of an email campaign, tracking metrics like open rate, click-through rate and conversion rate.

As data visualization vendors extend the functionality of these tools, they are increasingly being used as front ends for more sophisticated big data environments. In this setting, data visualization software helps data engineers and scientists keep track of data sources and do basic exploratory analysis of data sets prior to or after more detailed advanced analyses.

How data visualization works

Most of today's data visualization tools come with connectors to popular data sources, including the most common relational databases, Hadoop and a variety of cloud storage platforms. The visualization software pulls in data from these sources and applies a graphic type to the data.

Data visualization software allows the user to select the best way of presenting the data, but, increasingly, software automates this step. Some tools automatically interpret the shape of the data and detect correlations between certain variables and then place these discoveries into the chart type that the software determines is optimal.

Typically, data visualization software has a dashboard component that allows users to pull multiple visualizations of analyses into a single interface, generally a web portal.

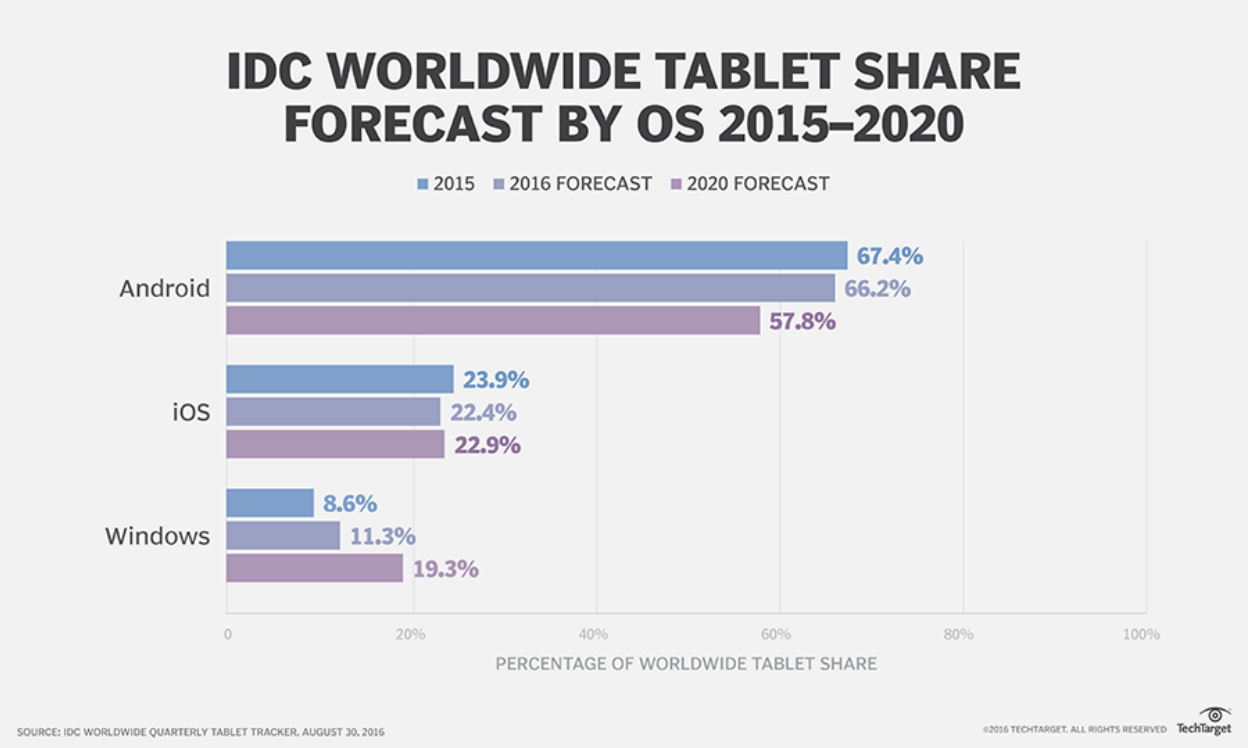


Figure 4.3 Forecasting tablet sales by operating system [16]

4.1.7 Tutorial 1 [17]

Dundas BI Software

Dundas BI, from Dundas Data Visualization, is a browser-based business intelligence and data visualization platform that includes integrated dashboards, reporting tools, and data analytics. It provides end users the ability to create interactive, customizable dashboards, build their own reports, run ad-hoc queries and analyze and drill-down into their data and performance metrics.

The Dundas BI platform is flexible, allowing users to connect and integrate with any data source in real-time, on any device. With the touch-based interface and responsive design, users can create and view dashboards and reports on any device from desktop to mobile.

Dundas BI can support companies of various deployment sizes and across industry verticals. The software was designed to provide a high degree of customization and integration support through their programmable, open API platform, offering .NET, REST, and JavaScript APIs. Dundas BI can be run on premise or hosted in cloud environments such as Microsoft Azure and Amazon Redshift.

Supported Operating System(s):

Mac OS, Web browser (OS agnostic), Windows 8

Figure 4.4 Dundas BI example [17]

Sisense Software

Sisense is an agile business intelligence (BI) solution that provides advanced tools to manage and support business data with analytics, visuals and reporting. The solution allows businesses to analyze big and disparate datasets and generate relevant business trends for them.

Sisense allows businesses to combine data from many disparate sources and club them into a single database. Once done, the solution itself rearranges data into a predefined standard format. Users can then perform slicing and dicing over the complete data set using multiple filters and built-in analytic tools.

Sisense includes functionality for dashboards and scorecards, data warehousing, extract, transform and load (ETL) and a query and report writer. Sisense provides a variety of dashboards to project the data through maps, KPIs, charts, trends, scatter plots and more.

Sisense uses Crowd Accelerated BI technology that provides capabilities for sharing analysis with users, both within and outside the organization. Users can share single analysis or a complete dashboard with their teammates or with the entire organization.

Sisense can be deployed on-premises or hosted in the cloud as a SaaS application.

Supported Operating System(s):

Web browser (OS agnostic), Windows 8, Windows 10

Figure 4.5 Sisence example [17]

Domo Software

Domo is a cloud-based business management suite that integrates with multiple data sources, including spreadsheets, databases, social media and any existing cloud-based or on-premise software solution. It is suitable for company sizes ranging from small businesses to large enterprises, and is compatible with Windows or Mac platforms, iPad tablets, and also works on mobile devices.

Domo provides both micro- and macro-level visibility and analysis; from cash balances and top-selling product lines to forecasted revenue by region, marketing return on investment (ROI) by channel, and more. It offers real-time dashboards with data pertaining to multiple business areas and performance metrics (e.g., key performance indicators, ROI, etc.). Domo also offers

interactive visualization tools and instant access to company-wide data via a centralized dashboard.

Pricing is offered on an annual subscription basis, and depends upon the number of users that need access.

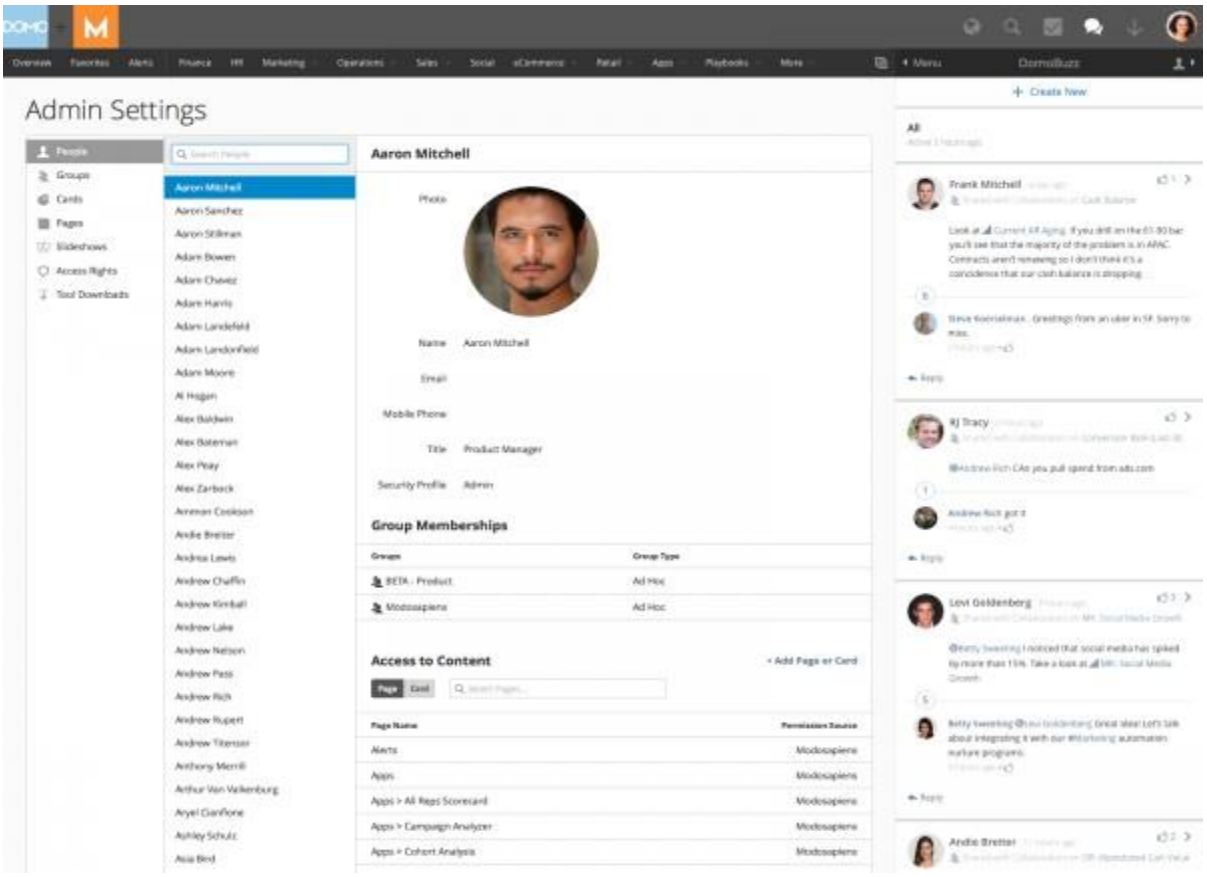


Figure 4.6 Domo example

Yellowfin Software

Yellowfin, a business intelligence platform, is a single integrated solution developed for companies across varying industries and scaling sizes.

This system is customizable to suit businesses in accounting, advertising, agriculture, banking, insurance, manufacturing, media, marketing, food and beverage, engineering, retail, technology and several other fields.

Key functionalities include dashboards and scorecards, predictive analytics, online analytical processing, query and reporting writing and performance management. Yellowfin offers filtering capabilities that can be updated in real time.

With the focus on BI for business users, non-technical staff can be equipped with self-service ad hoc reporting.

Sharing information with colleagues and stakeholders is an offered to facilitate discussions and embed directly into the company intranet. Yellowfin provides many different chart types, and users can toggle between them.

This platform is browser agnostic and is accessible from desktops as well as mobile devices.

Supported Operating System(s):

Windows 7, Windows Vista, Windows XP, Mac OS, Web browser (OS agnostic), Windows 2000, Windows 8, Windows 10

Figure 4.7 Yellofin example

Grow BI Dashboard Software

Grow BI Dashboard is a cloud-based business analytics and reporting solution suitable for small to midsize organizations. The solution allows users to create customizable dashboards for monitoring business workflows and key activities.

Grow pulls business data from a variety of sources and merges it into one single database. It supports connection with Salesforce, Dropbox, Google Analytics, Twitter and MySQL, among others. It supports blending, mashing, cleaning, filtering and comparing of data sets from a single or multiple sources.

Grow brings real-time data into the dashboard, allowing the user to view trends as they occur and share information with their team. The built-in data editor allows users to create customized metrics for their business needs. Users can also access dashboard on their iOS mobile devices and monitor performance metrics on the go. The snapshot feature allows capturing and storing snapshots that analysts can use to view and benchmark values in future.

The solution is available on a 30-day free trial period post which buyers have to pay a per-user subscription.

Supported Operating System(s):

Mac OS, Web browser (OS agnostic), Windows 8, Windows 10

Figure 4.8 Grow BI Dashboard Software example

BOARD Software

Created to combine business intelligence, corporate performance management, and business analytics, BOARD is a full-featured business intelligence system that serves midsize and enterprise-level companies in a variety of different industry segments.

Within the reporting functionality, BOARD allows users to pull from almost any data source, as well as generate full self-service reporting. The reports can be exported into several different formats, if necessary, such as CSV, HTML, and more. The system also features extensive multi-lingual capabilities, making it a great fit for companies that need to deliver reports in another language.

The dashboard application allows BOARD users to create a fully-customizable experience, featuring drill-down and drill-through functionality, as well as several different types of data visualization options. By implementing BOARD's data collecting and analysis functionalities, companies can view data in a relevant way that helps drive intelligent business decisions.

The solution can be either deployed on-premises or hosted in the cloud according to different architecture needs.

Supported Operating System(s):

Windows 7, Windows Vista, Windows XP, Web browser (OS agnostic), Windows 2000, Windows 8

Figure 4.9 BOARD Software example

ClicData Software

ClicData is a business intelligence (BI) dashboard solution designed for use primarily by small and midsized businesses. The tool enables end users to create reports and dashboards.

A drag-and-drop interface designed for ease of use allows users to connect to data located both on-premise and in the cloud. Users can analyze data pulled from a wide variety of sources, including Google Analytics, Facebook, Salesforce, Oracle, MySQL and Dropbox.

In addition to allowing users to create dashboards, ClicData also powers data processing tasks, including data history management, and automated merging and transformation of data. Businesses are thus able to manage their data and keep it centralized.

ClicData is also designed for collaboration: Users can send or share reports and dashboards to teams through its interface, or publish their dashboards and embed them in other Web applications' websites. Dashboards and reports can also be accessed via laptop, tablet or mobile phone.

Supported Operating System(s):

Windows 7, Windows Vista, Windows XP, Mac OS, Web browser (OS agnostic), Windows 2000, Windows 8, Windows 10

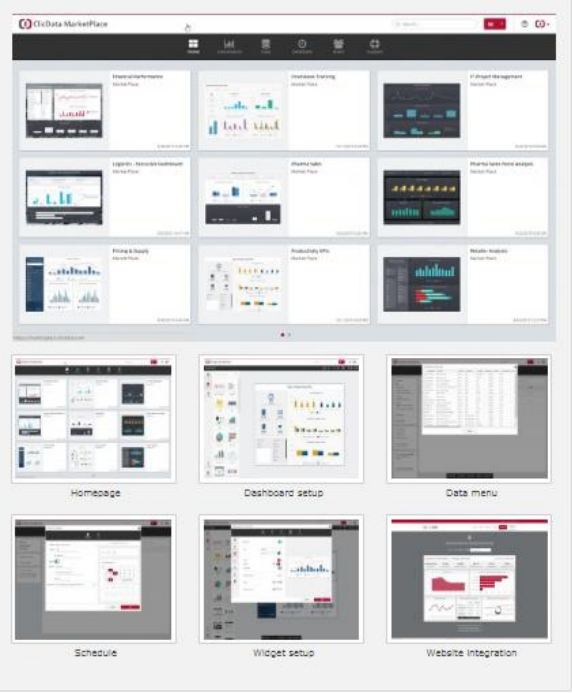


Figure 4.10 ClicData Software example

Style Intelligence Software

InetSoft Style Intelligence is a business intelligence software platform that allows users to create dashboards, visual analyses and reports via a data mashup engine—a tool that integrates data in real time from multiple sources. These sources can include OLAP servers, ERP apps Web services, relational databases and spreadsheets.

Style Intelligence features InetSoft's proprietary Data Block technology, which enables the data mashups to take place in real time. The information produced can then be accessed via interactive dashboards, enterprise reports, scorecards and exception alerts. A wide range of charts is available to users, including custom geographic mapping.

InetSoft's Style Intelligence is available on-premise or in the cloud. If deployed on-premise, the software is designed to have a light footprint on company systems so other critical applications will run unobstructed.

Style Intelligence is also accessible via mobile platforms, comes with an array of customization options and is suited to midsize and global organizations.

Supported Operating System(s):

Windows 7, Windows Vista, Windows XP, Mac OS, Linux, HP-UX, AIX, Solaris, Unix, IBM OS/400, Web browser (OS agnostic), Windows 2000, Windows 8

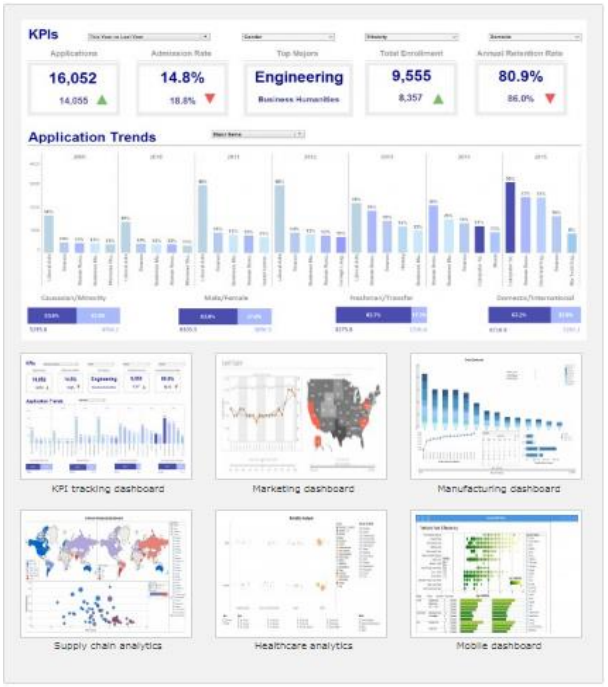


Figure 4.11 Style Intelligence soft were example

Chartio Software

Chartio is a cloud-based business intelligence solution that provides founders, business teams, data analysts and product teams in an organization tools to manage day-to-day business operations.

Chartio's tools provide users functionalities for data exploration, enabling them to represent this data in form of charts. It also provides a central dashboard that displays specific information to the users.

Chartio provides user's connections from Amazon Redshift to CSV files helping them explore data. The Data Pipeline tool lets users blend, layer and transform multiple datasets in a series of steps. Users can also share dashboards via scheduled emails or reports and view historical data and track corporate metrics with the help of the solution's Snapshot feature. The data is SSL-encrypted that is transferred via reverse SSH tunnel providing control over the connection and data to the users.

Services are offered on a monthly subscription basis that includes support via phone and email.

Supported Operating System(s):

Mac OS, Web browser (OS agnostic), Windows 8

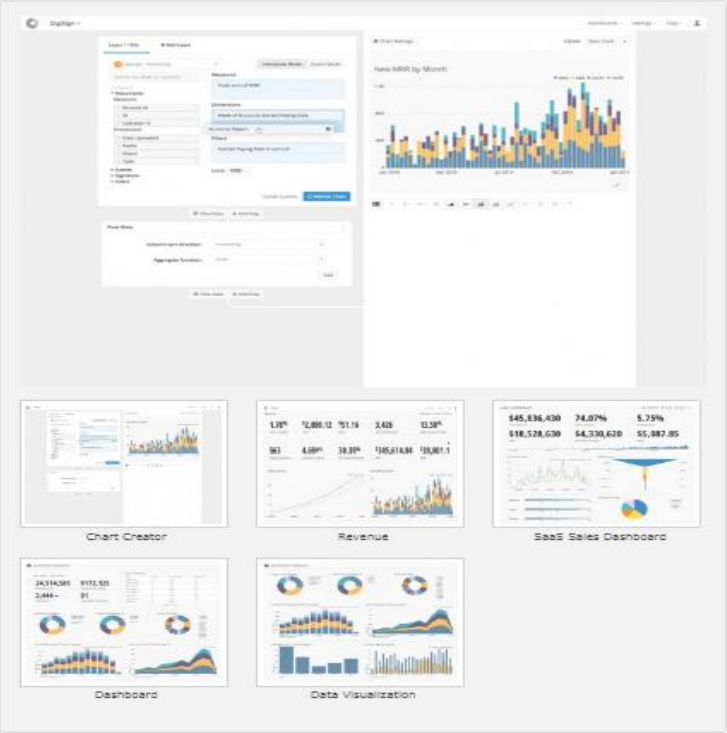


Figure 4.12 Chartio Software example

Looker Software

Looker is cloud-based business intelligence (BI) platform designed to explore and analyze data. The solution helps businesses to capture and analyze data from multiple sources and make data-driven decisions.

Looker provides business teams the ability to analyze supply chains, market digitally, quantify customer value, interpret customer behavior, and evaluate distribution processes. Users can also "view source" to understand how the data they are viewing is being manipulated. The dashboards allow presenting data and insights using customizable charts, graphs and reports. All dashboards and queries can be drilled into, so users can discover information in multiple layers.

With Looker's data modeling language, users can define data metrics and explore relationships between various data sets. The storytelling feature allows users to present data analysis to stakeholders through data-rich visualizations.

Looker is used in many industry sectors across the US and Canada including e-commerce, construction, education, finance, healthcare, media and technology.

Supported Operating System(s):

Mac OS, Web browser (OS agnostic), Windows 8

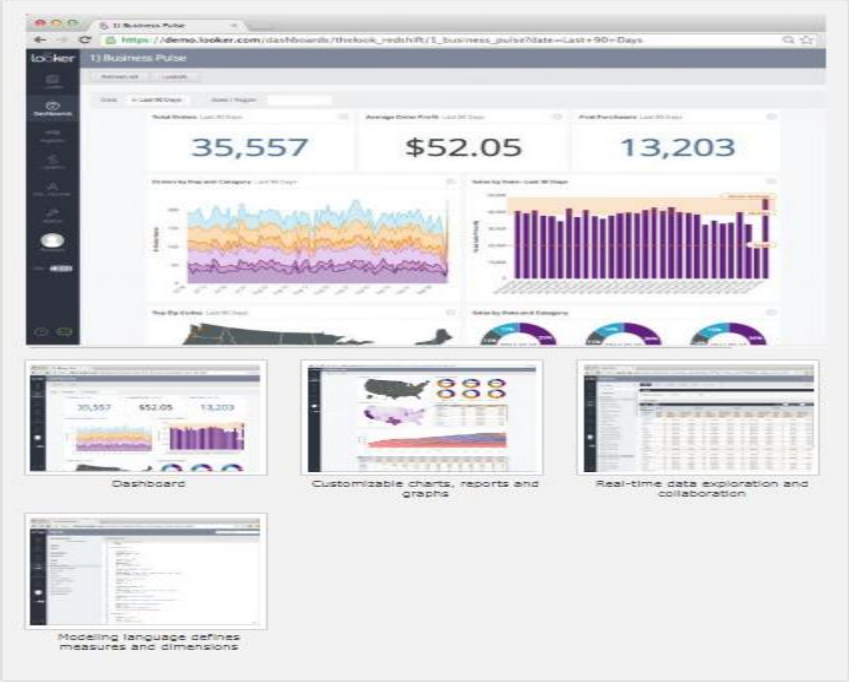


Figure 4.13 Looker Software example

Birst Enterprise BI & Analytics

Birst, an Infor Company, is a web-based networked BI and analytics solution that connects insights from various teams and helps in making informed decisions. The tool enables decentralized users to augment the enterprise data model virtually without compromising data governance. Birst also offers a unified semantic layer that maintains common definitions and key metrics.

Birst's two-tier architecture aligns back-end sources with line-of- business or local data. Birst's Automated Data Refinement extracts data from multiple sources into a unified semantic layer. The solution also supports blending of data in both top-down and bottom-up manner.

Birst allows users to access self-service analytics through executive dashboards, reporting, visual discovery, mobile tools and predictive analytics. Birst Open Client Interface also integrates with Tableau, Excel and R.

Birst's is packaged in three available formats: platform and per-user fee; by department or business unit; by end-customer (for embedded scenarios).

Supported Operating System(s):

Mac OS, Web browser (OS agnostic)

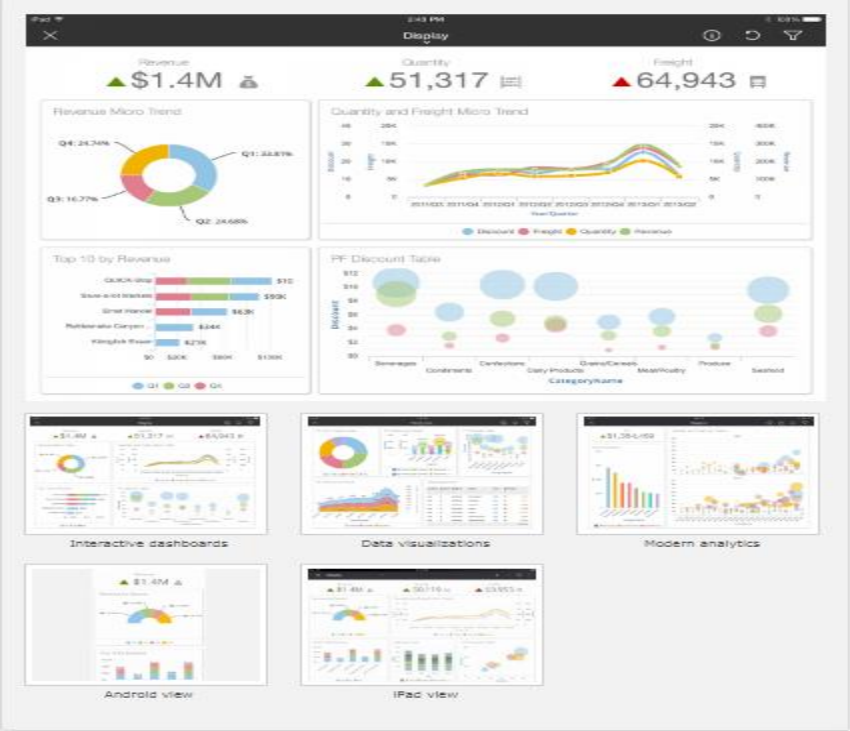


Figure 4.14 Birst Enterprise BI & Analytics example

SAP Business Objects Software

With its cohesive, consolidated approach, industry powerhouse SAP BusinessObjects gives organizations a full set of tools to manage and optimize Business Intelligence. From a centralized portal, companies can handle everything from ETL and data cleansing to predictive dashboards and a variety of reports - Crystal Reports, OLAP, ad hoc and more.

The key advantage to SAP BusinessObjects is that the solution can grow with a company, allowing organizations to employ additional tools as necessary and eliminating the need to stitch together various disparate solutions that can result in a bulky, user-unfriendly system. SAP's intuitive, standardized solution encourages high end-user adoption.

SAP BusinessObjects provides solutions for SMBs to large enterprises, and can be configured for a number of industries, including government entities, manufacturing and distribution, financial services, nonprofits and more.

Users can easily mine, analyze and report information from almost any data source and present in a number of formats both internally and externally. SAP BusinessObjects Mobile gives users access to reports, metrics and other data from smartphones and other mobile devices. SAP BusinessObjects Dashboards help end-users understand complex company data and make quick and accurate decisions based on that knowledge. With drag and drop, dashboards can be personalized to each company's unique data needs.

The BusinessObjects platform uses a common semantic layer, avoiding report duplication and ensuring data is accurate and consistent across the organization. The solution can also be seamlessly integrated with other enterprise solutions, Salesforce and Microsoft Office.

SAP BusinessObjects offers a mature, industry-leading solution for comprehensive Business Intelligence.

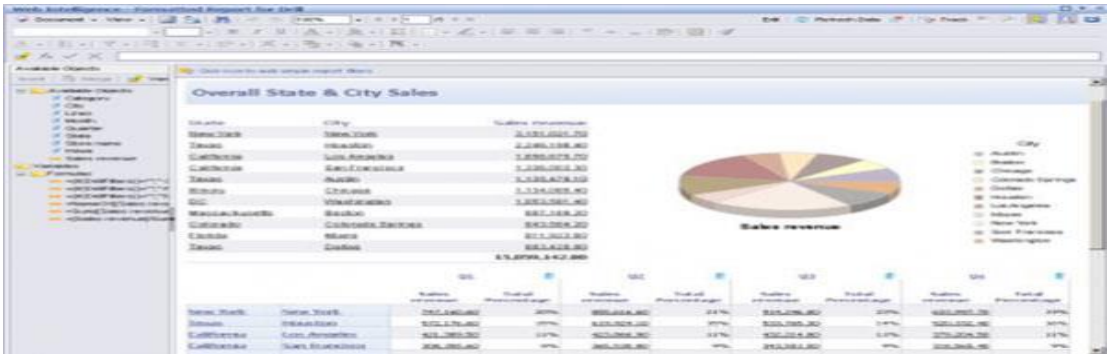


Figure 4.15 SAP Business Objects Software example

ReportPlus Software

ReportPlus is a business intelligence and analytics solution designed for midsize and large enterprises. It helps to analyze sorted and unsorted databases, set and view performance metrics and explore trends. It is available in both on-premise and cloud-based deployment models and also offers native apps for iOS, Windows and Android devices.

ReportPlus allows users to connect to various databases, including online sources like cloud storage drives, social media platforms, SaaS applications as well as offline sources like Excel, SQL database and more. Users can choose from multiple visualization types and create real-time dashboards to monitor and display key performance indicators. Dashboard instances can also be shared across teams. Users can even integrate dashboards into any third-party application.

ReportPlus is used across various industry verticals including finance, insurance and retail. It is available on an annual subscription basis that includes support via phone and email.

Figure 4.16 ReportPlus Software example

Halo Software

Halo is an end-to-end supply chain management and business intelligence platform that helps in business planning and forecasting inventory for supply chain management. The system uses data from all sources - big, small, and in-between - to form a cumulative view of all business information.

By combining the BI data analytics with a social platform, companies can use the data to spark dialogue between employees, helping to generate ideas and guide business decisions. Deployment options for Halo include on-premise, SaaS (Software as a Service) or PaaS (Platform as a Service, which allows users to develop, access and run custom apps in the cloud). This system is suited for mid-size companies in several target verticals, including manufacturing and distribution.

Supported Operating System(s):

Windows 7, Windows Vista, Windows XP, Mac OS, Web browser (OS agnostic), Windows 2000, Windows 8

Figure 4.17 Halo Software example

Pentaho Software

Pentaho is a business intelligence system designed to help companies make data-driven decisions, with a platform for data integration and analytics. The platform includes extract, transform, and load (ETL), big data analytics, visualizations, dashboards, reporting, data mining, and predictive analytics.

Pentaho's data integration functionality allows users to find, manage, and combine data from multiple sources, including native support for analytic databases, Hadoop, and NoSQL. The system can integrate with partners like Melissa Data and Human Inference.

Pentaho is also suitable for embedding or white labeling visual analytics as part of third-party Software as a Service (SaaS)/software applications as it is data agnostic. The application can be rebranded and customized based on open standards and architecture.

The software offers interactive business analytics tools like visual analysis and dashboards, as well as flexible reporting solutions. Predictive analytics offered by the system includes machine learning algorithms, tools for processing data and the capability to import third-party models with PMML.

Pentaho helps users translate big data into insights within a singular platform. Users have access to a complete spectrum of data from different sources with the system's adaptive big data layer, which takes the source of the data into account. The software is built on an open architecture and can be integrated with multiple systems. Pentaho is available for a limited period free trial.

Supported Operating System(s):

Mac OS, Web browser (OS agnostic), Windows 8, Windows 10

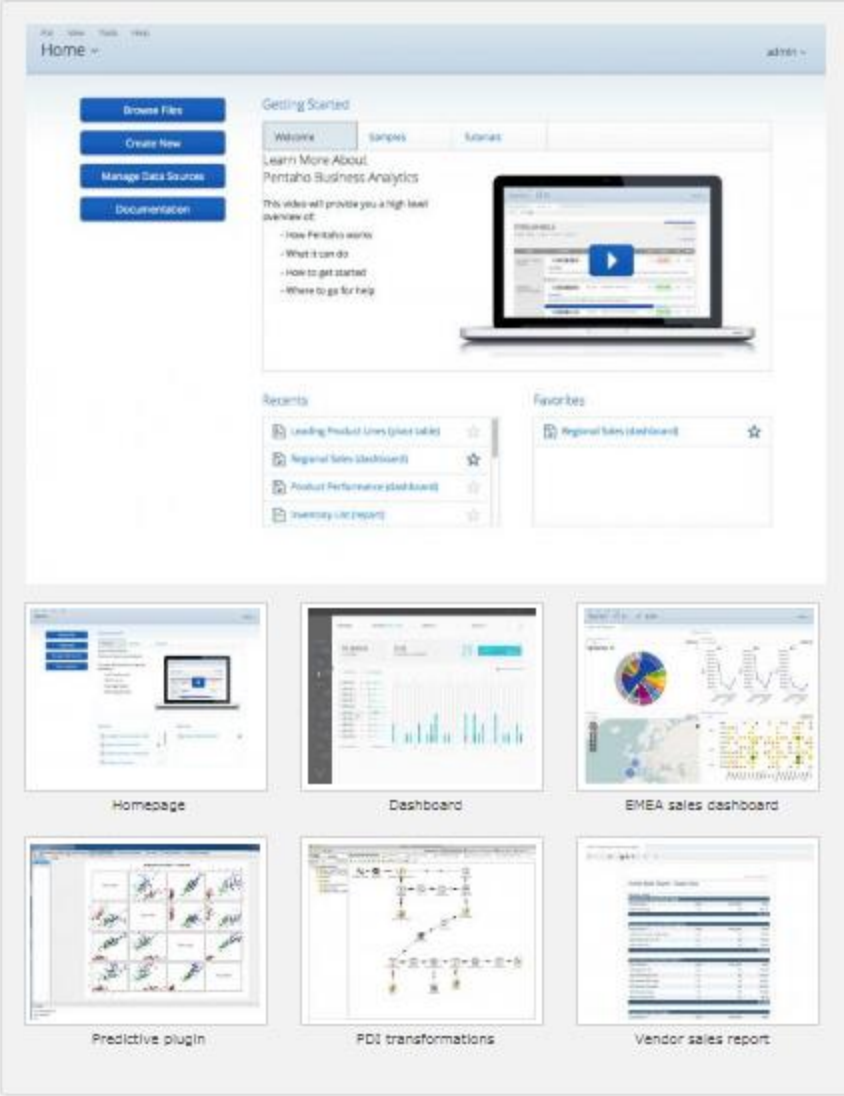


Figure 4.18 Pentaho Software example

Microsoft SharePoint Software

Microsoft SharePoint 2013 helps organization collect and analyze key business data in order to gain an accurate and actionable view of business operations. The software supports both on-premises as well as web-based deployment architecture. The solution can also be combined with existing business management software such as Dynamics CRM, Dynamics ERP and Microsoft Office to provide an intuitive end-user experience.

The solution enables users across organizations to monitor and analyze critical business data from desktops, smartphones or any Internet-enabled device. Using Microsoft SharePoint, users from across the company can collaborate effectively by setting up websites for sharing

information. In addition to creating websites, SharePoint also allows users to manage documents through their entire lifecycle as well as publishes collaborative reports. With its interactive dashboards and scorecards, users within the organization can quickly view and address specific needs of a department or the company as a whole.

Microsoft SharePoint 2013 is suitable for companies looking for a business intelligence solutions that increases collaboration while providing actionable business insights.

SharePoint Community Site

Supported Operating System(s):

Windows 7, Windows Vista, Windows 8

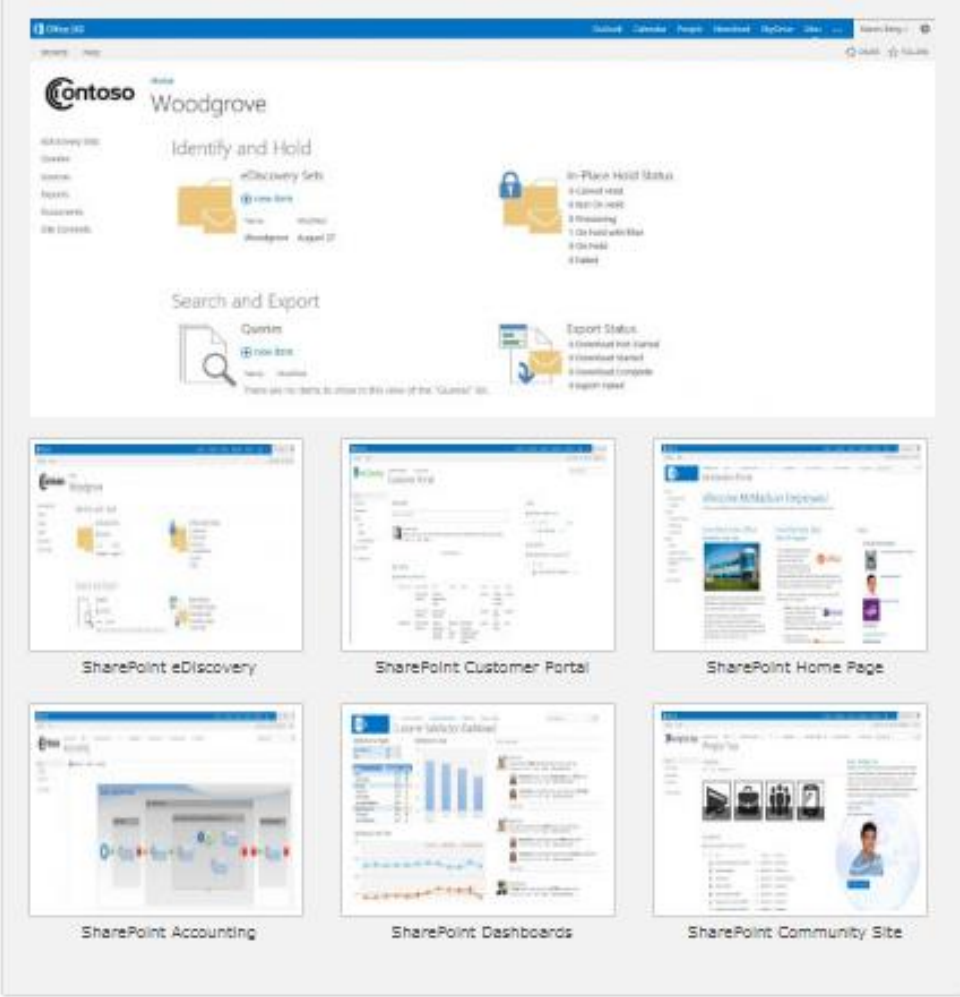


Figure 4.19 Microsoft SharePoint Software example

Izenda Reports Software

Izenda is a business intelligence (BI) platform that enables real-time data exploration and report creation. It is suited for software vendors and development teams in enterprises that need to embed BI and analytics functionality into their existing applications. Izenda can be accessed through web browsers on desktop and mobile devices, and on-premise deployments are also available.

Izenda is built on a 3-tier embedded architecture with an open source front end that enables integration with Ruby, Python, Java, .NET, PHP and other applications. Izenda can run on user infrastructure and utilize an application's existing security. Reporting tools include ad-hoc reports, dashboards, visualizations, tabular summaries, fixed-layout reports, trending comparisons and aggregations.

Users can build charts and grids, add filters and include drill-downs from inside their applications. A GUI-based administrative interface allows system and tenant administration and copy management for code-free deployment and administration. The platform supports a variety of databases, including SQL Server, Oracle and MySQL.

Supported Operating System(s):

Mac OS, Web browser (OS agnostic), Windows 8

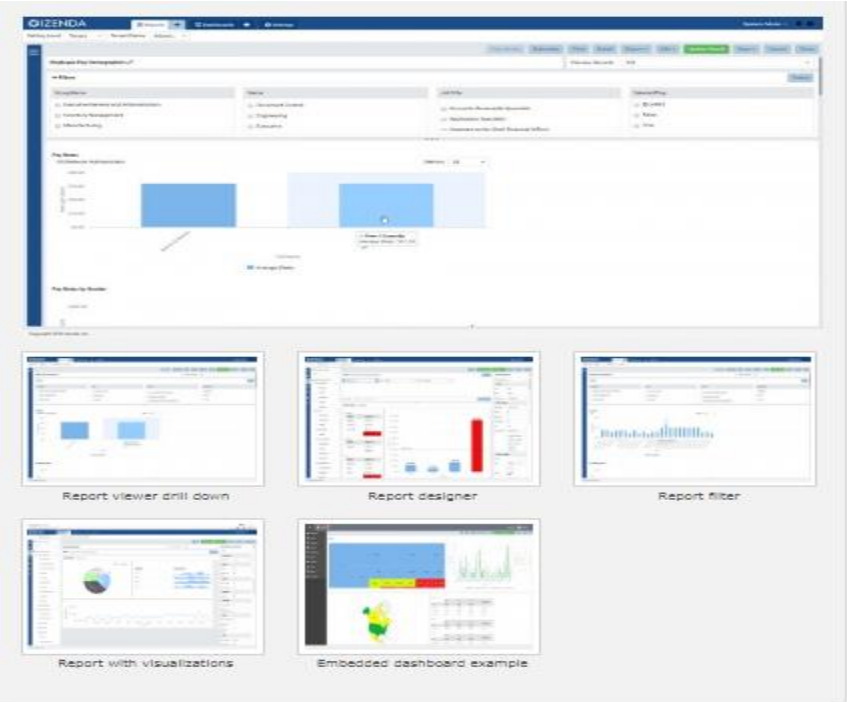


Figure 4.20 Izenda Reports Software example

Exago Software

Exago BI is a web-based solution that's designed to be embedded in web-based applications. Embedding Exago BI allows SaaS companies of all sizes to provide their customers with self-service ad hoc, operational reporting, and interactive dashboard capabilities.

With Exago BI, users can blend data from different sources including SQL Server, MySQL, Oracle, PostgreSQL, DB2, and Informix. Drag-and-drop report designers allow users to build custom reports, visualizations and dashboards from scratch, schedule reports and merge data onto pixel-perfect forms.

Visual and language aspects of the Exago BI interface can be modified to match the host application, including text strings, which can be translated into any language.

Exago BI operates as a single sign-on, and security can be set up by group or user at the database, data object, row or column level.

Supported Operating System(s):

Mac OS, Web browser (OS agnostic), Windows 8

Figure 4.21 Exago Software example

Stratum Software

Stratum by Silvon is a robust business intelligence solution that was designed to meet the unique needs of business professionals working for manufacturing and distribution companies.

Stratum offers a full suite of integrated analytic applications and a number of options for distributing business intelligence data via dashboards, scorecards, scheduled reports, alerts and other methods. With built-in analytics package and forecasting capabilities, Stratum gives intelligent insight and visibility into company's sales, inventory, financials, customer service, and vendor performance.

The solution extracts data from multiple locations and compiles it into one central place so that users across departments are able to analyze the same information in real-time. The reporting metrics allows businesses to compare KPIs, recognize trends, and plan for performance improvement.

Stratum can be deployed on-premise or over the web. Their web-based version is compatible with mobile devices including iPads and Android tablets, giving users access in and out of the office.

Supported Operating System(s):

Web browser (OS agnostic)

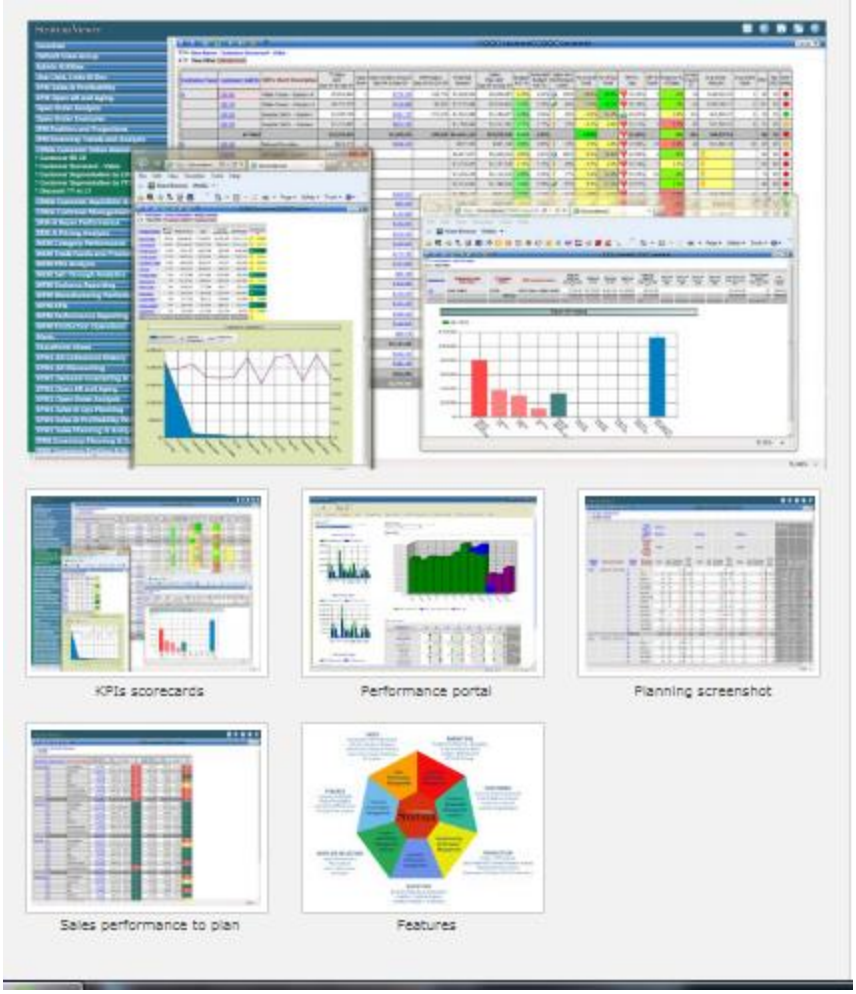


Figure 4.22 Stratum Software example

TARGIT Decision Suite Software

TARGIT Decision Suite is a business intelligence and analytics solution that offers visual data discovery tools, self-service business analytics, reporting and dashboards in a single, integrated solution.

TARGIT combines the control of a centralized BI solution with the flexibility of a decentralized solution. It's self-service capabilities enable business users to create their own reports and analyses, eliminating IT bottlenecks.

Through a vast range of deployment options, TARGIT Decision Suite can be made accessible to every person in the organization through Windows, web interface, mobile client or embedded dashboards in other programs. Users can connect to any and all data sources and keep control of the access with centralized data governance tools.

To aid adoption and successful implementation, TARGIT has a team of consultants, a wide range of training courses offered and an active community of users.

Supported Operating System(s):

Web browser (OS agnostic), Windows 10

Figure 4.23 TARGIT Decision Suite Software example

CHAPTER- 5
TEXT AND WEB MINING

5.1 TEXT AND WEB MINING [18]

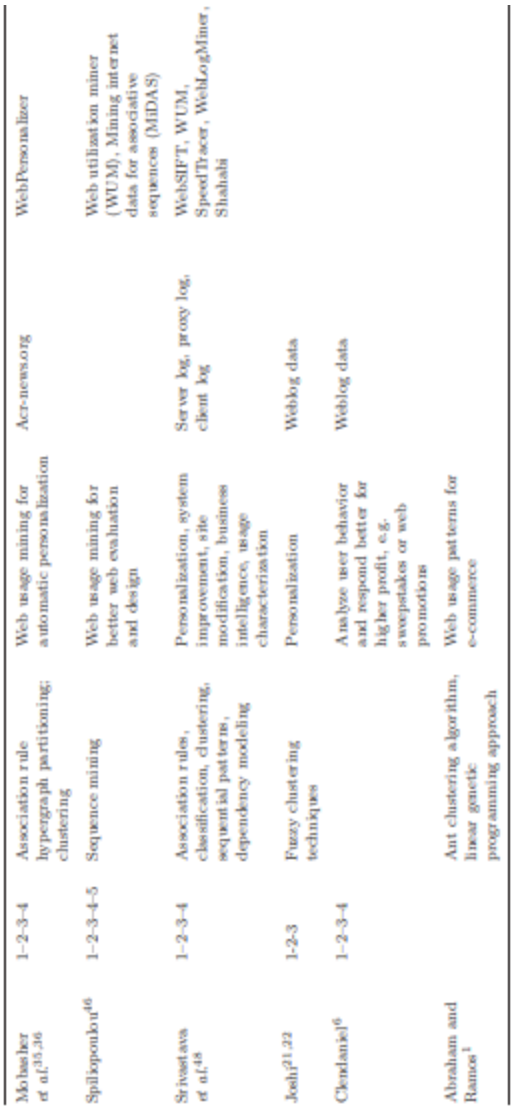
In the data mining communities, there are three types of mining: data mining, web mining, and text mining. There are many challenging problems in data/web/text mining research. Data mining mainly deals with structured data organized in a database (DB) while text mining mainly handles unstructured data/text. Web mining lies in between and copes with semi-structured data and/or unstructured data. Web mining calls for creative use of data mining and/or text mining techniques and its distinctive approaches. Mining the web data is one of the most challenging tasks for the data mining and data management scholars because there are huge heterogeneous, less structured data available on the web and we can easily get overwhelmed with data. In the literature, the terms of web mining, web data mining, and web data extraction mining are used interchangeably. The taxonomy of web mining has grown from that of only web content mining and web usage mining such as considered by Cooley et al.8 to include that of web structure mining as elaborated by Liang.[18] Web content mining is the process of discovering useful information from the content of web pages that may consist of text, image, audio or video data in the web; web usage mining is the application that uses data mining to analyze and discover interesting patterns of user's usage of data on the web; and web structure mining is the process of using graph theory to analyze the node and connection structure of a web site. An example of the latter would be discovering the authorities and hubs of any web document, e.g. identifying the most appropriate web links for a web page. According to Kosala and Blockeel, "In practice, the three web mining tasks above could be used in isolation or combined in an application, especially in web content and structure mining since the web document might also contain links."

For example, Zhong studies the brain informatics (i.e. combination of content and structure) from a web intelligence perspective. Kosala and Blockeel25 present a survey of web mining research for each of the three web mining categories presented above, and distinguish web mining as different from information retrieval (IR) and information extraction (IE). They hold that web mining techniques are not the only tools to solve information overload problems

either directly or indirectly. They claim that "Other techniques and works from different research areas, such as database (DB), information retrieval (IR), natural language processing (NLP), and the web document community, could also be used. . . . By the direct approach we mean that the application of the web mining techniques directly addresses the above problems. . . . By the indirect approach we mean that the web mining techniques are used as a part of a bigger application that addresses the above problem." Kosala and Blockeel[18]also claim that the web mining research area is a converging research area from several research communities, such as DB, IR, and artificial intelligence (AI) with machine learning and natural language processing (NLP) from the latter. The purpose of this paper is to provide a more current evaluation and update of web mining research and techniques available. This paper also presents the comparisons and summaries of selected software for web mining. The web mining software selected for discussion and comparison in this paper are SPSS Clementine, Megaputer PolyAnalyst, ClickTracks by web analytics, and QL2 by QL2 Software Inc.

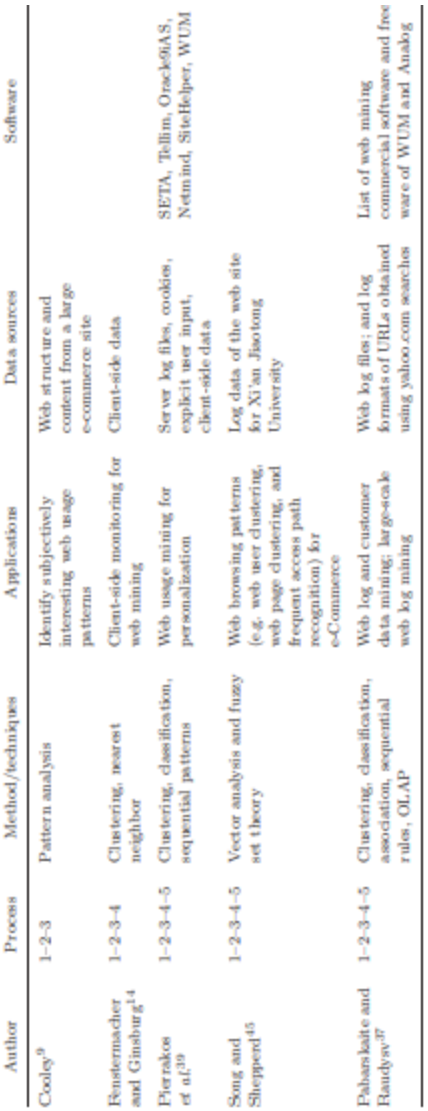
Author	Process	Method/techniques	Applications	Data sources	Software
Zhang *et al.*[55]	1–2–3–4	Relevance feedback algorithm	Contend-based image retrieval		A prototype system of content-based web image search
Lau *et al.*[26]	1–2–3–4	Keywords search	Homepage analysis	6173 students' homepages	Search engine
Chen *et al.*[5]	1–3–5	Clustering, categorization, web structure, and summarization techniques	Semantic Virtual Document (SVD)		Prototype intelligent Search And Review of Cluster Hierarchy (iSEARCH)
Liu[29]	1–2–3–4–5	Correlation mining, clustering, machine learning, partial tree alignment	Web query interface integration (e.g. travelocity.com); opinion mining	Structured and unstructured; CompUSA.com	
Darmont *et al.*[11]	1–2	Transforming multiform data into a unified format	Warehousing web data		Java prototype
Graves *et al.*[16]	1–2	Earth Science Markup Language (ESML)	Satellite imagery	NASA Goddard Earth Sciences Data	Algorithm Development and Mining Toolkit (ADaM)

Figure 5.1 Review table for web content mining [18]



Author	Process	Method/techniques	Applications	Data sources	Software
Mobasher et al.[35,36]	1–2–3–4	Association rule hypergraph partitioning; clustering	Web usage mining for automatic personalization	Acr-news.org	WebPersonalizer
Spiliopoulou[46]	1–2–3–4–5	Sequence mining	Web usage mining for better web evaluation and design		Web utilization miner (WUM), Mining internet data for associative sequences (MIDAS)
Srivastava et al.[48]	1–2–3–4	Association rules, classification, clustering, sequential patterns, dependency modeling	Personalization, system improvement, site modification, business intelligence, usage characterization	Server log, proxy log, client log	WebSIFT, WUM, SpeedTracer, WebLogMiner, Shahabi
Joshi[21,22]	1-2-3	Fuzzy clustering techniques	Personalization	Weblog data	
Cleslanski[6]	1–2–3–4		Analyze user behavior and respond better for higher profit, e.g. sweepstakes or web promotions	Weblog data	
Abraham and Ramos[1]		Ant clustering algorithm, linear genetic programming approach	Web usage patterns for e-commerce		

Figure 5.2 Review table for web usage mining [18]



Author	Process	Method/techniques	Application	Data sources	Software
Cooley[9]	1–2–3	Pattern analysis	Identify subjectively interesting web usage patterns	Web structure and content from a large e-commerce site	
Fürstenmacher and Ginsburg[14]	1–2–3–4	Clustering, nearest neighbor	Client-side monitoring for web mining	Client-side data	
Pierrakos et al.[39]	1–2–3–4–5	Clustering, classification, sequential patterns	Web usage mining for personalization	Server log files, cookies, explicit user input, client-side data	SETA, Tellim, Oracle6iAS, Netmind, SiteHelper, WUM
Song and Shepperd[45]	1–2–3–4–5	Vector analysis and fuzzy set theory	Web browsing patterns (e.g. web user clustering, web page clustering, and frequent access path recognition) for e-Commerce	Log data of the web site for Xi'an Jiaotong University	
Pabarskaite and Raudys[37]	1–2–3–4–5	Clustering, classification, association, sequential rules, OLAP	Web log and customer data mining; large-scale web log mining	Web log files; and log formats of URLs obtained using yahoo.com searches	List of web mining commercial software and free ware of WUM and Analog

Figure 5.3 Continued [18]

5.1.1 Text mining Concepts and Definition [7]

The information age that we are living in is characterized by the rapid growth in the amount of data and information collected, stored, and made available in electronic media. A vast majority of business data are stored in text documents that are virtually unstructured. According to a study by Merrill Lynch and Gartner, 85 to 90 percent of all corporate data are captured and stored in some sort of unstructured form (McKnight, 2005). The same study also stated that this

unstructured data are doubling in size every18 months. Because knowledge is power in today's business world, and knowledge is derived from data and information, businesses that effectively and efficiently tap into their text data sources will have the necessity knowledge to make better decisions, leading to a competitive advantage over those businesses that lag behind. This is where the need for text mining fits into the big picture of today's businesses. Text mining (also known as text data mining or knowledge discovery in textual databases) is the semi automated process of extracting patterns (useful information and knowledge) from large amounts of unstructured data sources. Remember that data mining is the process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data stored in structured databases, where the data are organized in records structured by categorical, ordinal, or continuous variables. Text mining is the same as data mining in that it has the same purpose and uses the same processes; but with text mining, the input to the process is a collection of unstructured (or less structured) data files such as Word documents, PDF files , text excerpts, XML files , and so on. In essence, text mining can be thought of as a process (with two main steps) that starts with imposing structure to the text-based data sources followed by extracting relevant information and knowledge from this structured text-based data using data mining techniques and tools . The benefits of text mining are obvious in the areas where very large amounts of textual data are being generated, such as law (court orders), academic research (research articles), finance (quarterly reports), medicine (discharge summaries), biology (molecular interactions), technology (patent files), and marketing (customer comments). For example, the free-form text-based interactions with customers in the form of complaints (or praises) and warranty claims can be used to objectively identify product and service characteristics that are deemed to be less than perfect and can be used as input to better product development and service allocations. Likewise, market outreach programs and focus groups generating large amounts of data. By not restricting product or service feedback to a codified form, customers can present, in their own words, what they think about a company's products and services. Another area where the automated processing of unstructured text has had a lot of impact is in electronic communications and e-mail. Text mining not only can be used to classify and filter junk e-mail, but it can also be used to automatically prioritize e-mail based on importance level as well as to generate automatic responses (Weng and Liu, 2004). Following are among the most popular application areas of text mining:

• Extraction Information. Identification of key phrases and relationships within text by looking for predefined sequences in text via pattern matching.

• Topic tracking. Based on a user profile and documents that a user views, text mining can predict other documents of interest to the user.

• Summarization. Summarizing a document to save time on the part of the reader.

• Categorization. Identifying the main themes of a document and then placing the document into a predefined set of categories based on those themes.

• Clustering. Grouping similar documents without having a predefined set of categories.

• Concept linking. Connects related documents by identifying their shared concepts and, by doing so, helps users find information that they perhaps would not have found using traditional search methods.

• Question answering. Finding the best answer to a given question through knowledge-driven pattern matching.

5.1.2 Text Mining Lingo [7]

The following list describes some commonly used text mining terms:

• ***Unstructured data*** (versus structured data). Structured data have a predetermined format. They are usually organized into records with simple data values (categorical, ordinal, and continuous variables) and stored in databases. In contrast, unstructured data do not have a predetermined format and are stored in the form of textual documents. In essence, the structured data are for the computers to process while the unstructured data are for humans to process and understand.

• ***Corpus***. In linguistics, a corpus (plural cOJpora) is a large and structured set of texts (now usually stored and processed electronically) prepared for the purpose of conducting knowledge discovery.

• ***Terms***. A term is a single word or multiword phrase extracted directly from the corpus of a specific domain by means of natural language processing (NLP) methods.

• ***Concepts***. Concepts are features generated from a collection of documents by means of manual, statistical, rule-based, or hybrid categorization methodology. Compared to terms, concepts are the result of higher level abstraction.

• ***Stemming***. The process of reducing inflected words to their stem (or base or root) form. For instance, stemming, stemmed are all based on the root stem.

• ***Stop words***. Stop words (or noise words) are words that are filtered out prior to or after processing of natural language data (i.e. , text). Even though there is no universally accepted list of stop words, most natural language processing tools use a list that includes articles (a, am, the, of, etc.), verbs (-is, are, was, were, etc.), and context-specific words that are deemed not to have differentiating value.

- ***Synonyms and polysemes***. Synonyms are syntactically different words (i.e., spelled differently) with identical or at least similar meanings (e.g., movie, film, and motion picture). In contrast, polysemes, which are also called homonyms, are syntactically identical words (i.e., spelled exactly the same) with different meanings (e.g., bow can mean "to bend forward," "the front of the ship," "the weapon that shoots arrows," or "a kind of tied ribbon").

  • ***Tokenizing***. A token is a categorized block of text in a sentence. The block of text correspondingto the token is categorized according to tl1e function it performs. This assignment of meaning to blocks of text is known as tokenizing. A token can look like anything; it just needs to be a useful part of tl1e structured text.

  • ***Term dictionary***. A collection of terms specific to a narrow field that can be used to restrict the extracted terms within a corpus.

  • ***Word frequency***. The number of times a word is found in a specific document.

  • ***Part-of-speech tagging***. The process of marking up the words in a text as corresponding to a particular part of speech (such as nouns, verbs, adjectives, and adverbs) based on a word's definition and the context in which it is used.

  • ***Morphology***. A branch of the field of linguistics and a part of natural language processing that studies the internal structure of words (patterns of word formation within a language or across languages).

  • ***Term-by-document matrix*** (occurrence matrix or term-document matrix). A common representation schema of the frequency-based relationship between the terms and documents in tabular format where terms are listed in rows, documents are listed in columns, and the frequency between the terms and documents is listed in cells as integer values.

  • **Singular-value decomposition** (latent semantic indexing). A dimensionality reduction method used to transform the term-by-document: to a manageable size by generating an

intermediate representation of the frequencies using a matrix manipulation method similar to principle component analysis.

5.1.3 Application Case [7]

A patent is a set of exclusive rights granted by account to an inventor for a limited period of time in exchange for a disclosure of an invention (note that the procedure for granting patents, the requirements placed on the patentee, and the extent of the exclusive rights vary widely from country to country).

The disclosure of these inventions is critical to future advancements in science and technology. If carefully analyzed, patent documents can help identify emerging technologies, inspire novel solutions, foster symbiotic partnerships, and enhance overall awareness of business' capabilities and limitations. Patent analysis is the use of analytical techniques to extract valuable knowledge from patent databases. Countries or groups of countries that maintain patent databases (United States, European Union, Japan, etc.) add tens of millions of new patents each year. It is nearly impossible to efficiently process such enormous amounts of semi structured data (patent documents usually contain partial structured and partial textual data). Patent analysis with semi-automated software tools is one way to ease the processing of these very large databases. Eastman Kodak employs more than 5,000 scientists, engineers, and technicians around the world. During the twentieth century, these knowledge workers and their predecessors claimed nearly 20,000 patents, putting the company among the top 10 patent holders in the world. Being in the business of constant change, the company knows that success (or mere survival) depends on its ability to apply more than a century's worth knowledge about imaging science and technology to new uses and to secure those new uses with patents. Appreciating the value of patents, Kodak not only generates new patents but also analyzes those created by others. Using dedicated analysts and state-of-the-art software tools (including specialized text mining tools from ClearForest Corp.), Kodak continuously digs deep into various data sources (patent databases, new release archives, and product announcements) in order to develop a holistic view of the competitive landscape. Proper analysis of patents can bring companies like Kodak a wide range of benefits:

• It enables competitive intelligence. Knowing what competitors are doing can help a company to develop countermeasures.
• It can help the company make critical business decisions, such as what new products, product lines, and/ or technologies to get into or what mergers and acquisitions to pursue.
• It can aid in identifying and recruiting the best and brightest new talent, those whose names appear on the patents that are critical to the company's success.
• It can help the company to identify the unauthorized use of its patents, enabling it to take action to protect its assets.
• It can identify complementary inventions to build symbiotic partnerships or to facilitate mergers and/or acquisitions.
• It prevents competitors from creating similar products, and it can help protect the company from patent infringement lawsuits.
Using patent analysis as a rich source of knowledge and a strategic weapon (both defensive as well as offensive), Kodak not only survives but excels in its market segment defined by innovation and constant change.[7, 23]

5.1.4 Tutorial 1 [19]

SQL Server Business Intelligence Project Tutorial

Overview

Every time we venture into something for the first time, we can be both excited and nervous at the same time. This is especially the case when dealing with something in your career for the first time. One is eager to make it a success by putting all of your energy into it. They say that "Knowledge is Power", and the potential of being successful in a new venture is directly proportional to the informed decisions one makes in day-to-day work. If you are reading this tutorial, this means that you are keen to take your first venture into a Business Intelligence project and make it a success. So let's get started and understand what we are going to study.
To understand a project of any nature, one needs to understand the typical architectures and best practices for the type of project. We will start with an understanding of the Business Intelligence (BI) architecture to develop our knowledge of the key points for a BI project. Once you land on a project, you will be assigned responsibilities that you need to understand to

analyze your work and provide estimates on how soon you can deliver the work - which for many people is the most unnerving part of the job as you don't want to ask for too much or too little time.

In a BI project, data is the currency and analytics is the merchandise. Without having the right type, amount and shape of data, one cannot get the right kind of analytics. In this tutorial, we will look at different types of data sources as well as relational data modeling. Data makes many hops from one layer of the solution to another in any BI solution architecture. During this data movement from source to destination, the shape of the data keeps transforming. This facilitation of data movement and transformation along the way is typically known as Extract, Transform and Load (ETL). We will look at typical ETL needs and the associated topics like data staging and master data management concepts.

Once we gather a huge volume of data from a variety of sources for our required analytics, it needs to be stored in the right manner for better retrieval. To understand it in simple terms, it's similar to how books are organized in a library for easy retrieval. We will discuss data warehousing and developing a data model for high performance.

When huge volumes of data at the scale of GBs / TBs is required for a project, we need an eagle-eye view of the data that goes beyond the reach of an Online Transaction Processing (OLTP) database. For example, let's say a national bank has 10 million transactions per day, and the CEO needs to know an average amount of these transactions for the past five years. These high volume computations are better suited for an Online Analytical Processing (OLAP) database. To meet these needs, we need to understand data marts, cubes, OLAP as well as dimensional modeling.

Finally we reach a stage where we are ready to report on the statistically computed values in a visually appealing way (i.e. reporting and analytics). We need to understand the different kinds of reports and different analytics options.

Through the course of this tutorial, we will associate each topic with corresponding tools / technology from the Microsoft Business Intelligence technology stack with example use-cases of how to implement the solution.

Visual Analytics

Every project should start with a thorough requirements analysis phase. If the project is outsourced, before the project is awarded to a vendor, different pre-sales activities take place which are outside the scope of this tutorial. We will consider it turned over to the technology team once the project has been initiated and at that point in time the development team has been asked to kick-off the technical aspects of the project. Business Analysts are typically positioned closed to the clients to understand the functional requirements and the technical team works with the Business Analysts to translate the needs into technical requirements. In this section of the tutorial we will discuss different aspects of the requirements analysis.

Explanation

Requirements analysis is thought to be the job of only a business analyst, but it reality it is a collective team effort. Everyone needs to collect requirements for their own set of tasks and deliverables. In a Business Intelligence project, one needs to have some high level understanding of a standard BI architecture. A clear understanding of the architecture helps to identify the right areas for requirement analysis. We will consider a business scenario that we will use as our project requirements for the rest of the tutorial, and corresponding to that, we will start looking at a typical BI architecture that addresses the project requirements.

Business Scenario

AdventureWorks is a fictitious company that manufactures Bikes as well as related clothing and accessories. The company conducts business globally and has sales in Europe, North America as well as Asia-Pacific. AdventureWorks wants to move away from paper-based and manual processes, and wants to have an online system that stores the data in a relational database. Each business unit from every geographical location would access the online system and manage sales and order related data that would be stored in a central data repository. This centralized data repository should act as the one-stop-shop for all the data related needs of the company, as the company grows and adds more sources of data in the future.

Different countries have different regulations and compliance requirements. Hence different types of reports have to be developed to enable immediate, short-term and long-term reporting requirements. As the company has many sales executives who travel frequently to conduct

business, the reports should be accessible from smart devices. The CEO of the company is the key sponsor and user of reporting, and is interested in deriving sales analytics from the data. A team of data analysts would be accessing the data to create reports on-demand, as required by the CXOs of the company. They need to be able to slice and dice the data for detailed analysis.

Business Intelligence Architecture

For now, let us keep in mind the business scenario and look at a typical end-to-end BI architecture. We will discuss requirements gathering and time estimations for this scenario in the next section. The below diagram shows a typical BI architecture. Each entity in the diagram represents a different layer and/or function of the solution. Generally a technology / business intelligence architect would develop the solution / technology architecture based on the project requirements.

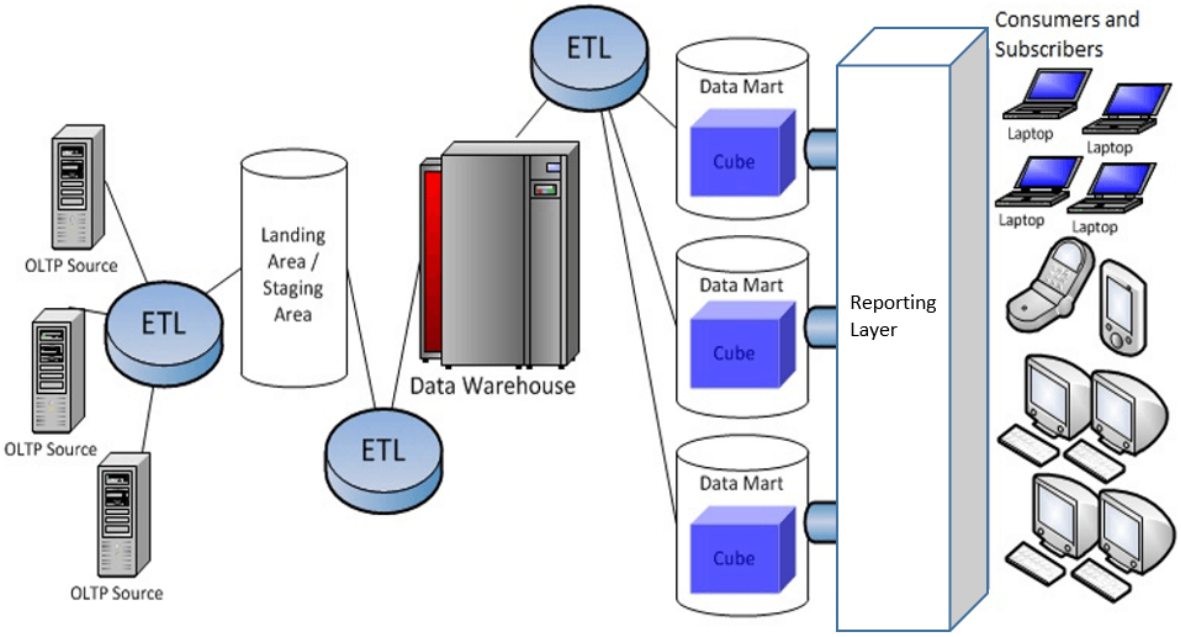


Figure 5.4 Business Intelligence Architecture [19]

Source System - Typically data is extracted from OLTP source systems. Data is stored in these systems using application front-ends. In some cases data is stored in file-based data sources like Excel spreadsheets too.

Staging / Landing Area - Data is extracted from source systems and an identical or transformed copy of the data is gathered on a centralized platform commonly referred to as a staging area. Data collected in staging area may be cleansed and transformed by applying business rules and stored into yet another storage area called the landing zone. Reading data from transactional systems would adversely affect the performance. There are other factors such as requiring de-coupling transactional systems from reporting systems, hence a copy of the data is staged at regular intervals.

ETL - The process of extracting, transforming and loading data from one system to another, popularly known as ETL, is often used as the data movement vehicle in BI projects. SQL Server Integration Services is the Microsoft BI Technology to meet ETL requirements.

Data Warehouse - As data continuously gets accumulated and different kinds of data requirements keep evolving in the enterprise, it becomes necessary to warehouse the data. A data warehouse acts as a central data repository for any data related needs of an enterprise. There can be multiple source systems in an enterprise, but data is ideally stored in a single centralized data warehouse.

Data Mart - Depending upon the data analysis need of any given business department, business-specific data marts would be created. A data mart in simplified terms means a sub-section of the data warehouse that is used for analysis. Typically OLAP (Online Analytical Processing) technologies are employed to create a data mart, like SQL Server Analysis Services. The volume of calculations on the voluminous amount of data mandates the use of OLAP technology as relational databases are not well suited to meet the performance needs of such a solution. The most recognizable implementation of an OLAP is in the form of a cube, similar to databases in the relational world.

Reporting - The flow of data within any application may start with different forms of data collection methods, and mostly ends with some kind of reporting. SQL Server Reporting services are the commonly used reporting technology in the Microsoft BI stack. Reports can be operational, analytical or strategic and may be consumed by desktops, tablets and mobile devices too. Reporting can be as simple as an operational report formed of a few rows and columns, and can be as complex as a dashboard composed of scorecards, geo-spatial analysis, charts, graphs, data matrix with drill-down and drill-through.

5.1.5 Tutorial 2 [20]

Business Intelligence & Metadata Tutorial:

Business Intelligence is a terminology refers to taking advantage of data and converting them into an intelligent information or knowledge by carefully observing data patterns or trends. These findings are key factors in helping any business to improve it's current business processes to gain more on customer satisfaction, increase sales, produce more profit etc. The knowledge observed from several report based analysis may lead to new business changes or improvements thus helping the organization to grow in the targeted direction. Browse through the various topics listed below to know more.

Metadata Tools:

Metadata is data about data and Metadata tools are used for gathering, storing, updating, and for retrieving the business and technical metadata of an organization. Data – The actual data that is stored in the database. Business Metadata – The information of the data related with business and this data is used by functional team (business analysts, smart management experts, business managers etc). Technical Metadata – The information about the data related with technology and this data is used by the technical team (Developers, DBAs).

Tool Name	Company Name
Rochade	ASG
Metatrieve	Metatrieval
Datamapper	Exeros
Metacenter	Data Advantage Group
Meta Data Integration Framework	Info Librarian
SuperGlue	Informatica
Metastage	Ascential

Figure 5.5 Popular Metadata Tools [20]

Business Metadata

In IT, Business Metadata is adding additional text or statement around a particular word that adds value to data. Business Metadata is about creating definitions, business rules. For example, when tables and columns are created the following business metadata would be more useful for generating reports to functional and technical team. The advantage is of this business metadata is whether they are technical or non-technical, everybody would understand what is going on within the organization.

Table's Metadata: While creating a table, metadata for definition of a table, source system name, source entity names, business rules to transform the source table, and the usage of the table in reports should be added in order to make them available for taking metadata reports.

Column's Metadata: Similarly for columns, source column name (mapping), business rules to transform the source column name, and the usage of the column in reports should be added for taking metadata reports.

Entity (Table) Name	Attribute (Column)	Attribute Definition
TARGET AUTO LOAN BY WEB	Auto Loan Identifier	The number that uniquely identifies an AUTO LOAN.
	Auto Loan Amount	The amount of auto loan that has been approved. Mapping: SOURCE_AUTO_LOAN_BY_WEB.AUTO_LOAN_AMOUNT
	Auto Loan Broker Commission Amount	The commission amount that has to be paid to AUTO loan broker. Note: This column is a derived column and not found in the source system. Derivation Rule: Auto Loan Amount * .01
	Auto VIN Identifier	This column identifies the Auto VIN Number
	Borrower Full Name	The full name of the borrower. Note: This column is a derived column and not found in the source system. Derivation Rule: SOURCE_AUTO_LOAN_BY_WEB.(BOR_FST_NAME concatenated with BOR_LAST_NAME)
	DateTimeStamp	The date on which the record has been created or updated.

Figure 5.6 Business Metadata Example [20]

CHAPTER- 6
LOOKING AHEAD - EMERGING TRENDS IN BUSINESS INTELLIGENCE

6.1 EMERGING TRENDS IN BUSINESS INTELLIGENCE [7][21]

The IT industry and the world at large have always been subject to technology and business trends, sometimes undergoing major changes, such as the development of the personal computer, client/server computing, the evolution of the internet, and now cloud computing.

Over the last few years, new trends have emerged that have had an enormous influence on how organizations work, interact, communicate, collaborate and protect themselves.

IT 'meta-trends' influence organizations' strategies, operations and investments in a wide variety of ways: Digitalization and security & privacy seem to make the most headlines now, but agility, cloud, mobile and artificial intelligence are also major technology drivers.

These meta-trends can be considered as the main drivers behind a number of important trends, either related to the use of software and technologies for business intelligence/analytics (BI) and data management or to the way the use of technology is organized in companies.

BARC's BI Trend Monitor 2018 reflects on the business intelligence and data management trends currently driving the BI market from a user perspective.

In order to obtain useful data for the BI Trend Monitor, we asked almost 2,800 users, consultants and vendors for their views on the most important BI trends. Their responses reveal a comprehensive picture of the future of BI as well as regional, company and industry-specific differences, delivering an up-to-date, objective perspective on the business intelligence market.

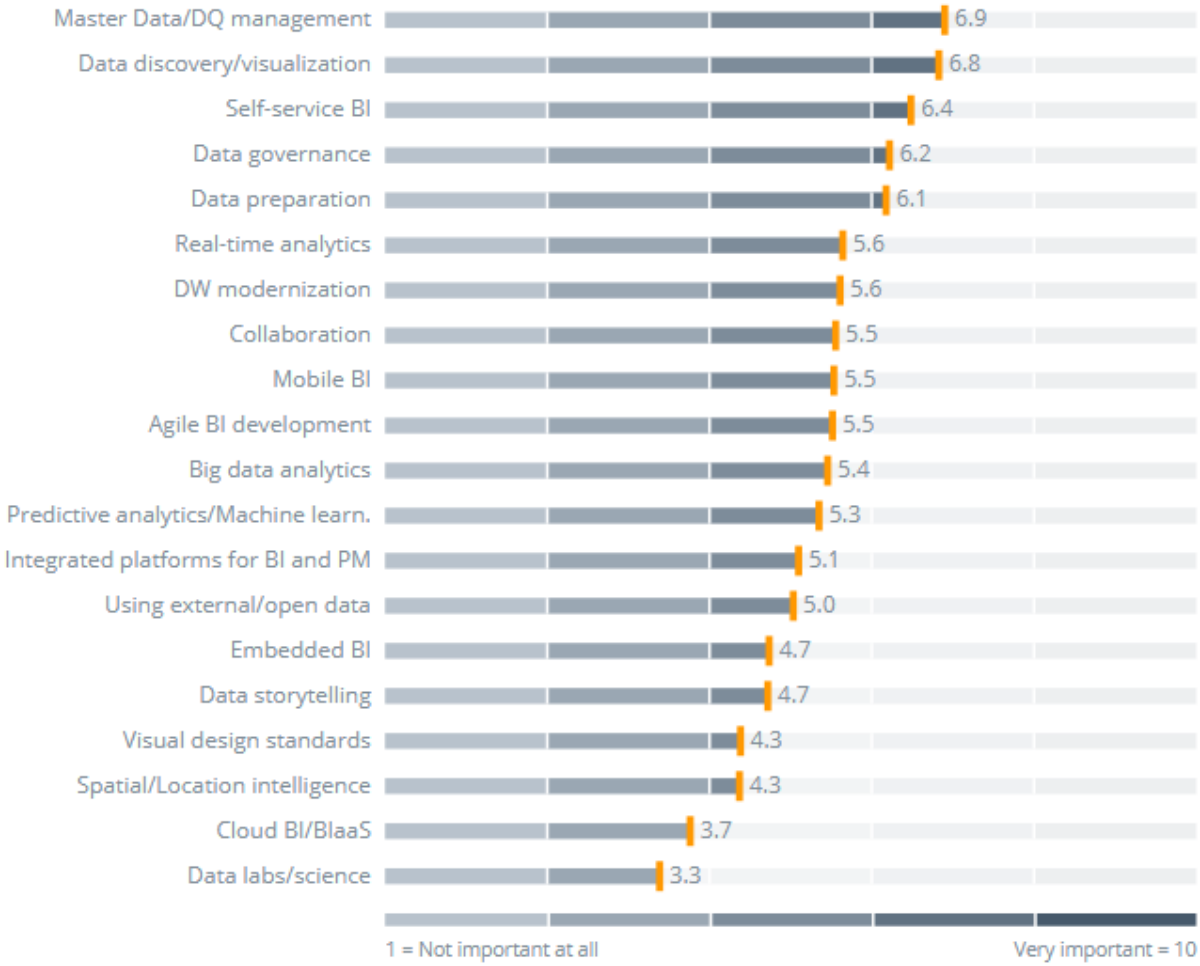


Figure 6.1 Importance of Business Intelligence Trends in 2018 (n=2,770) [21]

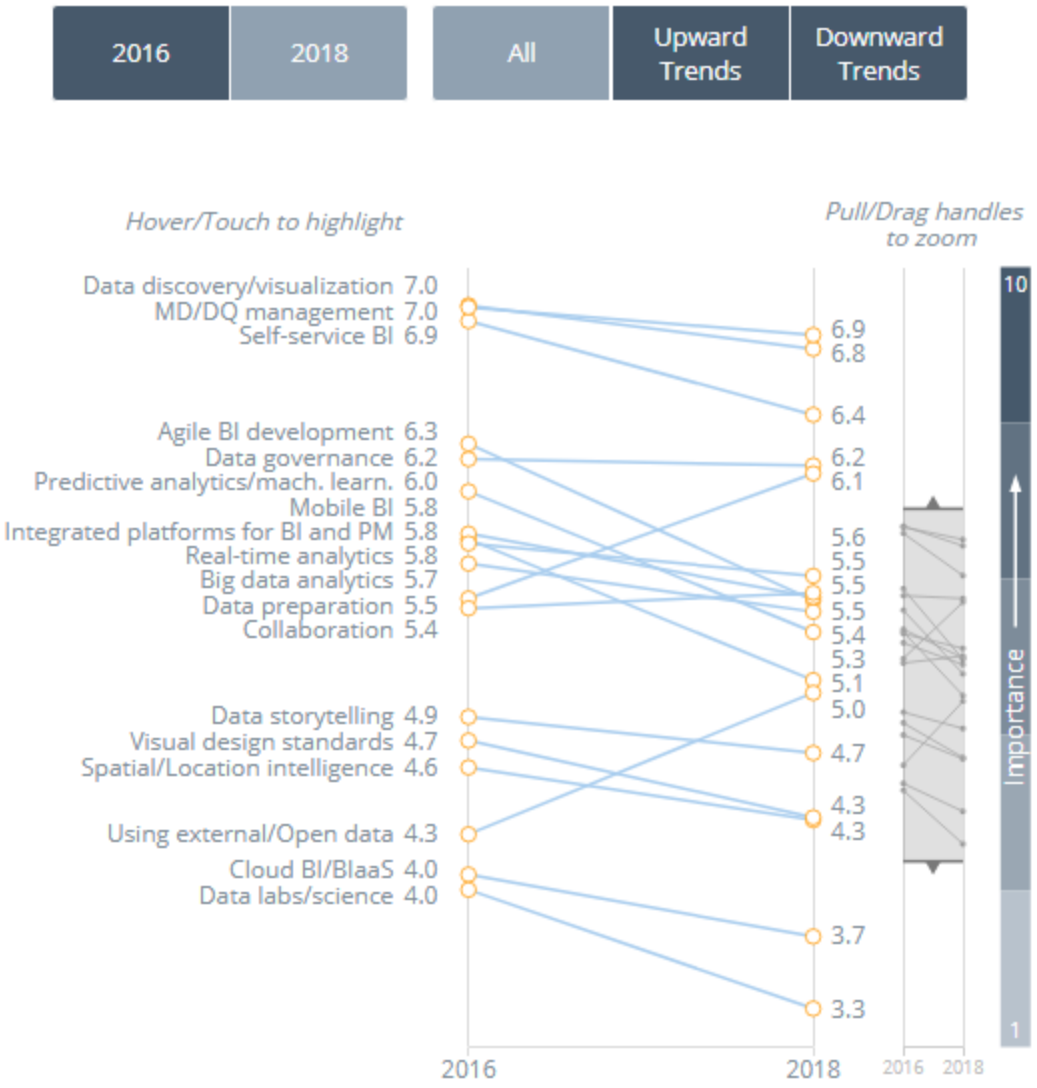


Figure 6.2 Importance of BI Trends (Timeline) [21]

6.1.1 Implementing Bl: An Overview [7]

Implementing BI systems can be very complex. In addition to typical issues in information system implementation, such as conducting appropriate cost-benefit analysis with intangible variables to justify the system and dealing with resistance to change, there are some complex issues related to integration, security, scalability of the system, and handling the construction of a data warehouse, analytics, and dashboards.

Bl Implementations Factors

A large number of factors may influence BI implementation. These factors are technological, administrative, behavioral, and so on. Many of these are generic to most information systems and have been researched extensively in the information systems literature.

The major factors that affect the decision making process of BI implementation.

1. Reporting and Analysis Tools
    a. Features and functionality
    b. Scalability
    c. Usability and manageability
    d. Ability to customize applications
2. Database
    a. Scalability and performance
    b. Manageability and availability
    c. Security and customization
    d. Ability to write back
3. Extraction, Transformation, and Load (ETL) Tools
    a. Ability to read any source
    b. Efficiency and productivity
    c. Cross platform support
4. Costs Involved
    a. Hardware costs (actual or opportunity)
    b. Costs of software (ETL, database, applications, and front end)
    c. Internal development costs
    d. External developments costs
    e. Internal training
    f. Ongoing maintenance
5. Benefits
    a. Time savings and operational efficiencies
    b. Lower cost of operations
    c. Improved customer service and satisfaction

d. Improved operational and strategic decision making

e. Improved employee communications and satisfaction

f. Improved knowledge sharing

These factors need to be analyzed both quantitatively and qualitatively.

- Critical success factors of business intelligence implementation

Although there could be many factors that could affect the implementation process of a BI system, a report from Vodapalli (2009) [7] shows that the following are the critical success factors for a business intelligence implementation.

a. Business driven methodology and project management

b. Clear vision and planning

c. Committed management support and sponsorship

d. Data management and quality issues

e. Mapping the solutions to the user requirements

f. Performance considerations of the BI system

g. Robust and extensible framework

- Managerial Issues Related to Bl Implementation

1. ***System development and the need for integration***. Developing an effective BI application is complex. For this reason, most BI vendors offer highly integrated collections of applications, including connection to enterprise resource planning (ERP) and customer relationship management (CRM; see Section 6.3). Notable are Oracle, Business Objects, MicroStrategy, IBM, and Microsoft. Most BI vendors provide for application integration, usually Web enabled.

2. ***Cost-benefit issues and justification***. Smaller organizations can make the solutions cost effective if they leverage existing databases rather than create new ones. One solution is on-demand BI. Nevertheless a careful cost-benefit analysis must be undertaken before any commitment to BI is made.

3. ***Legal issues and privacy***. BI analysis may suggest that a company send electronic or printed catalogs or promotions to only one age group or one gender.

A man sued Victoria's Secret (a brand of Limitedbrands) because his female neighbor received a mail order catalog with deeply discounted items and he received only the regular catalog (the discount was actually given for volume purchasing).

Settling discrimination charges can be very expensive. Some data mining may result in the invasion of individual privacy. What will companies do to protect individuals? What can individuals do to protect their privacy? These issues have to be kept in mind as BI solutions are implemented.

4. ***BI and BPM today and tomorrow***. The quality and timeliness of business information for an organization is not the choice between profit and loss-it may be a question of survival. No enterprise can deny the inevitable benefits of BI and BPM. Recent lndust1y analyst reports show that in the coming years, millions of people will use BPM dashboards and business analytics (BA) every day. Enterprises are getting more value from BI by extending information to many types of employees, maximizing the use of existing data assets. Visualization tools including dashboards are used by producers, retailers, government, and special agencies. Industry-specific analytical tools will flood the market to support analysis and informed decision making from top level to user level. BI takes advantage of existing IT technologies to help companies leverage their IT investments and use their legacy and real-time data. Thus a planned, careful, proactive approach to BI implementation is becoming a competitive necessity.

5. ***Cost justification***; intangible belle fits. While enterprise systems provide tangible benefits, it is difficult to quantify their intangible benefits. In a down-turned economy with high energy costs, mo1tgage crises, and political unrest, IT investments must be economically justified.

6. ***Documenting and securing support systems***. Many employees develop their own decision support or BI modules to increase their productivity and the quality of their work. It is advisable to have an inventory of these ad hoc systems and make certain that appropriate documentation and security measures exist, so that if the employee is away or leaves the organization, the productivity tool remains. Taking appropriate security measures is a must. End users who build their own BI applications are not professional systems builders. For this reason, there could be problems with data integrity and the security of the systems developed.

7. ***Ethical issues***. BI and predictive analytics can lead to serious ethical issues such as privacy and accountability. In addition, mistakes can cause harm to others as well as the company. For example, a company developed a decision support system (DSS) to help people compute the

financial implications of early retirement. However, the DSS developer did not include the tax implications, which resulted in incorrect retirement decisions. Another important ethical issue is human judgment, which is frequently a key factor in decision making. Human judgment may be subjective or corrupt, and therefore, it may lead to unethical decision making. Companies should provide an ethical code for system builders. Also, the possibility of automating managers' jobs may lead to massive layoffs. There are ethical issues related to the implementation of expert systems and other intelligent systems. The actions performed by an expert system can be unethical, or even illegal. For example, the expert system may advise you to do something that will hurt someone or will invade the privacy of certain individuals. An example is the behavior of robots and the possibility that the robots will not behave the way that they were programmed to. There have been many industrial accidents caused by robots that resulted in injuries and even deaths. The issue is, Should an organization employ productivity-saving devices that are not 100 percent safe? Another ethical issue is the use of knowledge extracted from people. The issue here is, Should a company compensate an employee when knowledge that he or she contributed is used by others? This issue is related to the motivation issue. It is also related to privacy. Should people be informed as to who contributed certain knowledge? A final ethical issue that needs to be addressed is that of dehumanization and the feeling that a machine can be "smarter" than some people. People may have different attitudes toward smart machines, which may be reflected in the manner in which they will work together.

8. ***BI Project failures***. There have been many cases of failures of all types of BI projects. There are multiple reasons for such failures, ranging from human factors to software glitches. Here are some examples:

a. Failure to recognize BI projects as enterprise-wide business initiatives and that they differ from typical stand-alone solutions.

b. Lack of business sponsors with the ability to insure funding

c. Lack of cooperation by business representatives from the functional areas

d. Lack of qualified and available staff

e. No appreciation of the negative impact of "dirty data" on business profitability

f. Too much reliance on vendors

6.1.2 Tutorial [22]

Learn Microsoft Business intelligence step by step

We will start with very basic stuffs like understanding what is mean by Data warehouse and business intelligence and end with creating some complex SSRS Reports.

Agenda

What is Data Warehouse?

In a very simple word it means, "It's a place where we store all of our data".

How it is different from Database?

Usually data warehouse is also a database. The records from multiple data sources (may be some other databases) are collected and stored in denormalized manner. It normally stores months or years of data to support historical analysis.

What is Denormalization?

Normalization is a database designing technique which makes sure that there will not be any redundant data.

It makes our data more reliable (because there will not be any repeating data. Exiting data will be referred wherever required)

It makes database management easy.

It reduces the size of database.

CityId	CityName
1	Mumbai
2	Kolkata
3	Delhi

CustomerId	CustomerName	CityId
1	Sukesh	1
2	Rajesh	1
3	Mahesh	2
4	Ganesh	1

Figure 6. 3 Normalized database

Figure 6.4 Data warehouse summarized [22]

What is the purpose of Data Warehouse?

As you can see, in the above real life example we collect different data from different places like from my bank account, from the Personal notebook (one maintained by mom containing household expenditure), from investment detail data (stored in Access database) etc.

Once we have collected all data, I store them in a separate excel sheet in Denormalized fashion.

This new excel sheet (we may refer it as Data warehouse) can be used to make my decisions about my expenses and incomes.

In simple words Data warehouse make us retrieve calculated data quickly and efficiently (without having too many joins in our select query).

What is Business Intelligence?

There is a saying that Images are better than long description. Human mind understands graphical explanation more than theoretical explanation. In order to make decision, our information need to be displayed with proper presentation in terms of charts, reports, score cards etc. Initially the concept of data warehouse was all about keeping historical data.

Data warehouse is a foundation for the BI. BI is all about leveraging our existing data and converting them into information or we can say Knowledge. We use this knowledge for making decisions in company.

Figure 6.5 Scheme BI

BI or Business Intelligence is simply a solution for:

- Collect information from multiple data sources
- Transform that data into meaningful information
- And finally show data to users with elegant presentation.

Introduction to Business Intelligence development studio / Sql Server Data tools

In the Microsoft world we will use "Business Intelligence development studio" commonly known as BIDS for this purpose.

It's an IDE which will let us develop Data Analysis and Business Intelligence solutions.

It has special project types and tools for developing "Sql server Integration Services", "Sql server Analysis services" and "Sql server reporting services". We will learn each of these in a series of step by step article.

With the release of Sql Server 2012, BIDS was renamed to Sql Server Data tools.

Basic idea on SSIS, SSAS and SSRS

Step 1. Open Sql Server Data Tools

Step 2. Click on File >> New >> Project. A dialog box will popup similar to like this

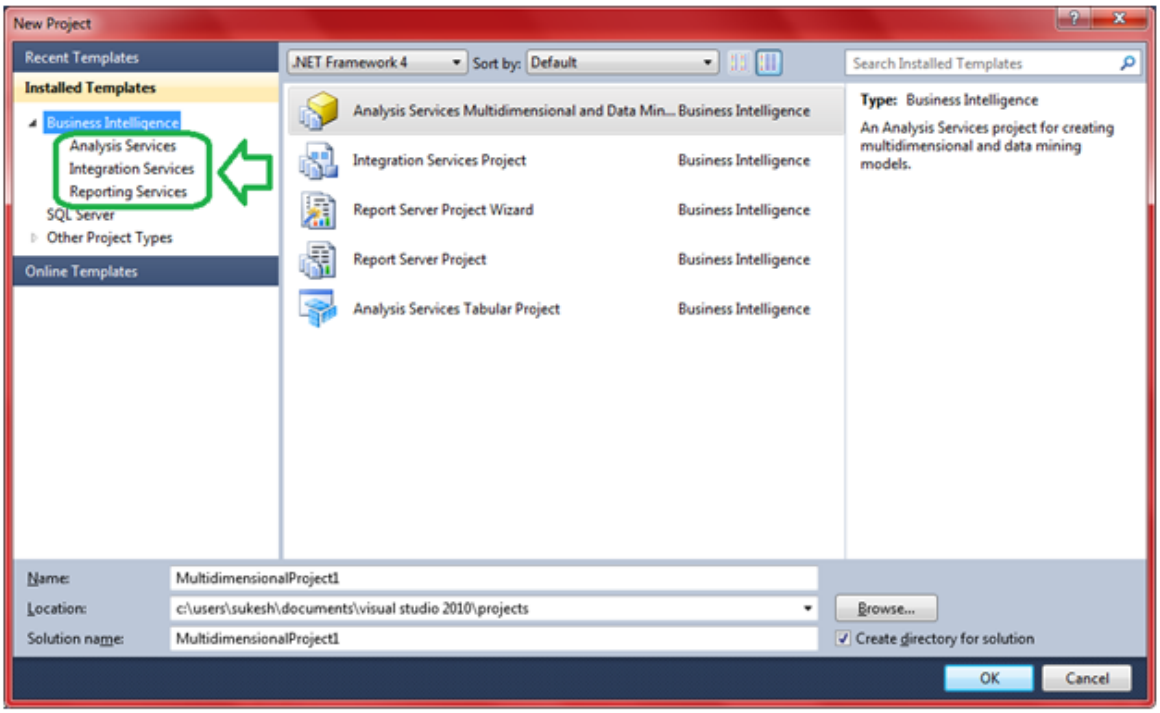


Figure 6.6 In order to perform a complete business intelligence task we need to go up with all these three projects.

1. Integration Services – SSIS – Sql server Integration services

It will let us perform wide range of data migration tasks. It let us collect data from various data sources and store them into central location.

2. Analysis Services – SSAS –Sql Server Analysis services

It will let us analyze the data

3. Reporting Services – SSRS – Sql Server Reporting services

It will let us create reports from analyzed data and present it to end user.

Start with SSIS

SSIS basically performs three basic things,

Collect data from various sources. – We call it Extraction (E)

Data obtained from different sources may or may not be same format. So first convert all of them according to business needs - We call it Transformation (T)

Load them into one big data source (mostly Data Warehouse) – We call it Load (L)

Figure 6.7 Together termed as ETL process [22]

Problem statement

You have Two excel files as follows,

A	B	C	D
Title	**FirstName**	**LastName**	**Address**
Mr.	Kenen	Joshep	USA
Mrs.	Kamala	Kumar	Chennai
Mrs.	Sheetal	Raju	Chennai

	A	B	C
1	**Name**	**Address**	
2			
3			
4			
5			
6			

Figure 6.8

You will collect data from first file (Datafile.xlsx).

Convert that data so that it match to second excel file format (merge Title, FirstName and LastName and call them as Name).

Dump final result to second file (Result.xlsx).

Step by Step Demo

Click File >> New >> Project. Select Integration Services from the group. Specify some nice name. Say Ok

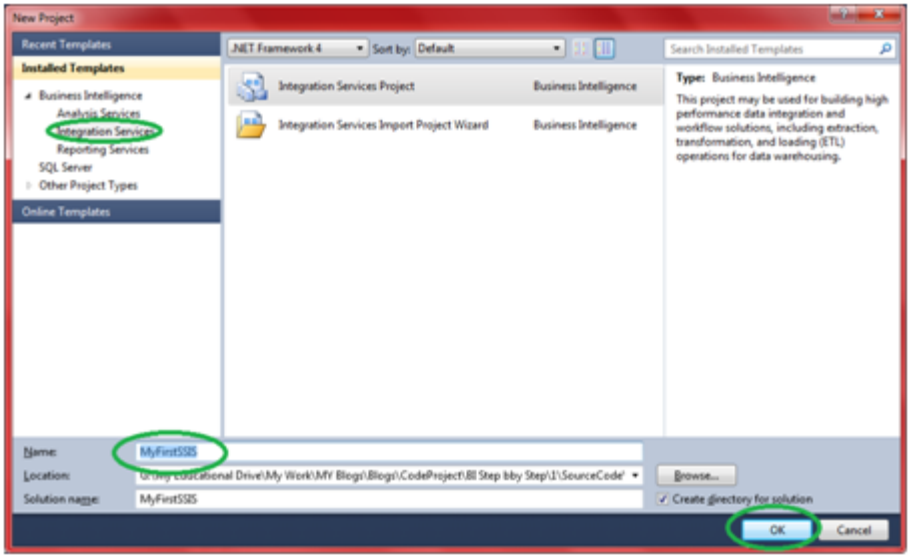


Figure 6.9

It will open up SSIS designer which you will use for creating and maintaining Integration service packages. It looks like follows,

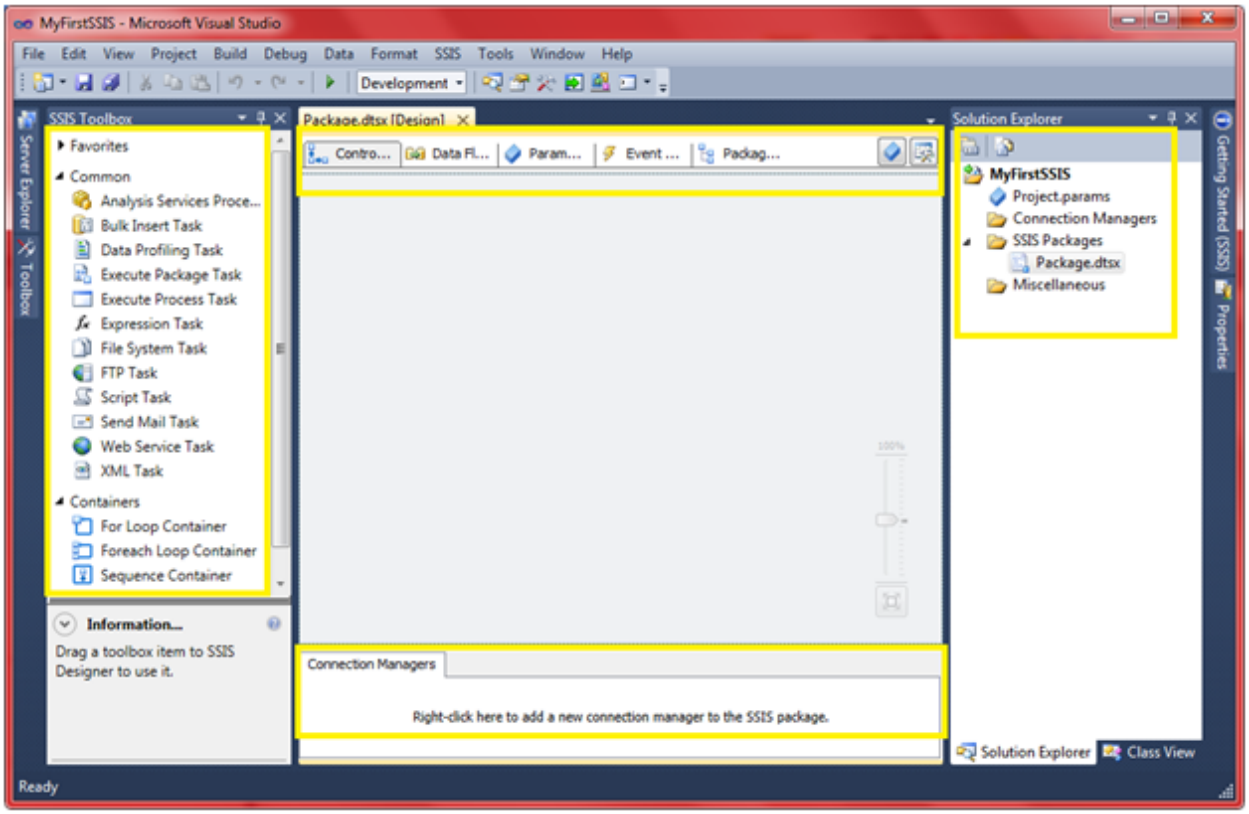


Figure 6.10

In the solution explorer under “SSIS packages” folder you will see one default package created with name “Package,dtsx”. If you want you can simply “rename it or remove it and add new one (right click the folder and say “New SSIS Package”).

Note: Package is simply a collection of connections, control flow elements, data flow elements, event handlers, parameters etc. We will talk about each one of this as move further.

Step 2. Create Connection Manager for Excel File

2.1 Right click Connection Manager and Say New Connection.

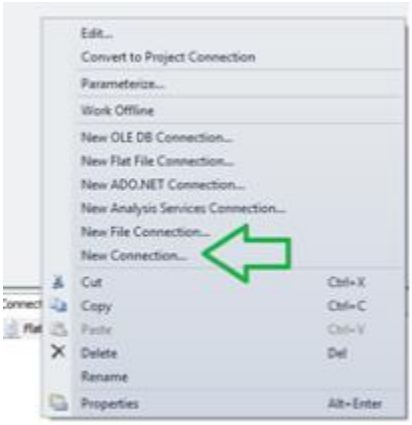


Figure 6.11

2.2 Select Excel from the popup and click on Add.

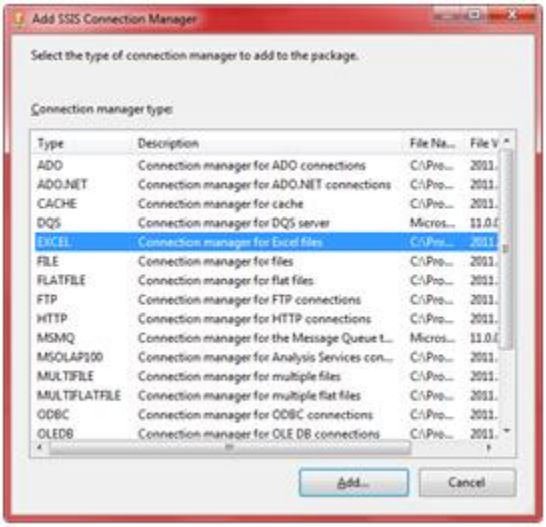


Figure 6.12

2.3 Click the browse button and select the excel file and click on OK.

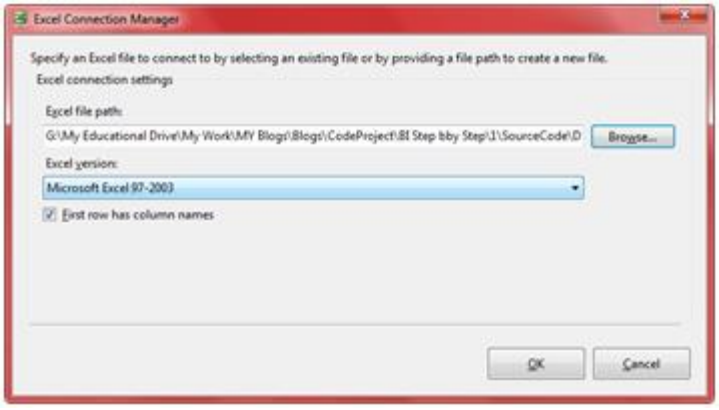


Figure 6.13

Step 3. Rename Connection Managers

3.1 Right click the connection manager just added and rename it to SourceExcelManager.

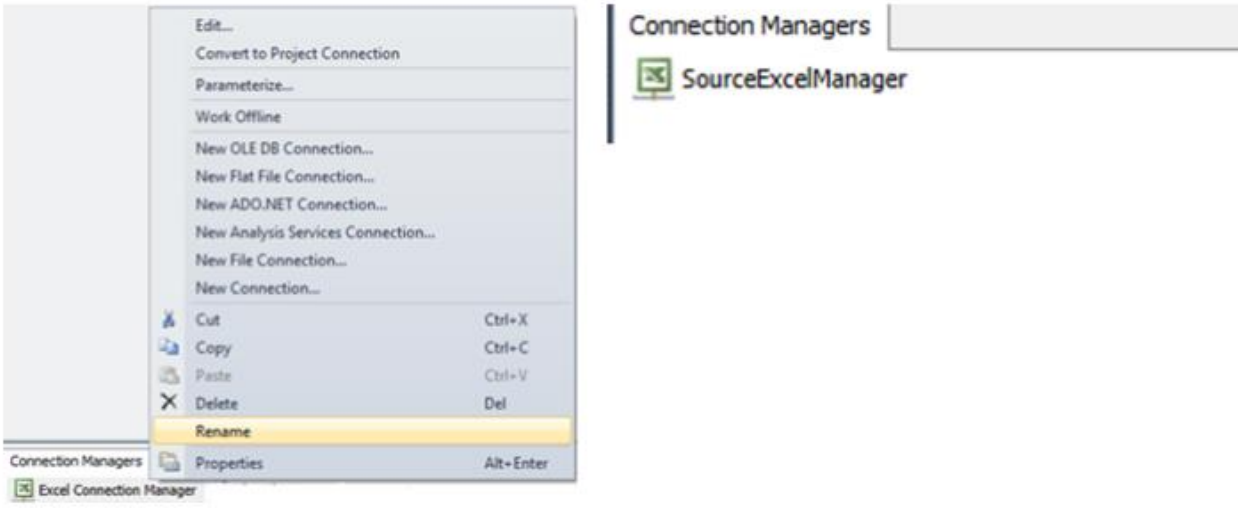


Figure 6.14

Step 4. Create Destination connection manager

4.1 Follow the Step no 3 and create one more connection manager pointing to Result.xlsx file.

4.2 Rename connection manager to ExcelDestinationManager.

Step 5. Create control flow – Pass data from Source Excel to Destination Excel.

5.1 Make sure control flow tab is selected in SSIS designer.

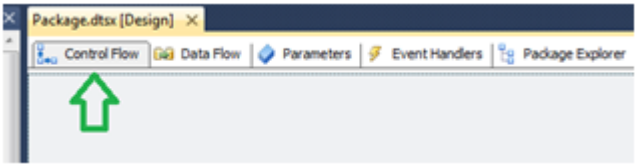


Figure 6.15

5.2 Select data flow task from the toolbox and drag it into designer.

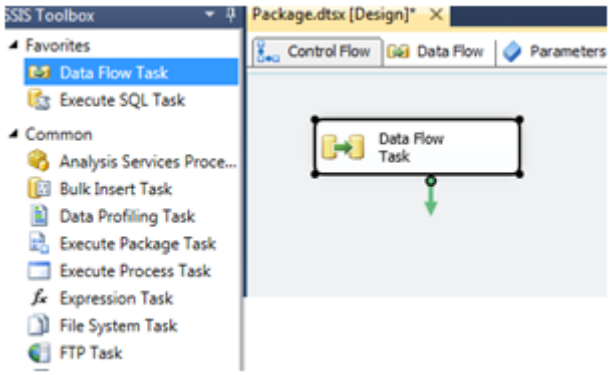


Figure 6.16

5.3 Rename Data Flow Task to “Source excel to Destination excel transfer task”

Control Flow

Control flow will be used to define the workflow. As the name implies it control the flow of execution.

What all tasks need to be executed?

What will be the sequence?

Whether tasks need to be executed in loops or not?

Step 6. CreateData Flow.

Data Flow defines the flow of data between source and destination.

6.1 Double click the control flow created in last step.

6.2 It will take you to second tab that is Data Flow tab.

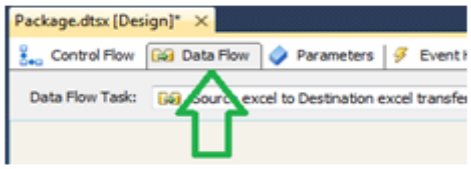


Figure 6.17

Step 7. Create Excel Source

7.1 Now you will see a different SSIS toolbox all together. In toolbox you will see couple of groups defined like Sources, Destinations and Transformations.

Take excel source from Source group and place it in designer.

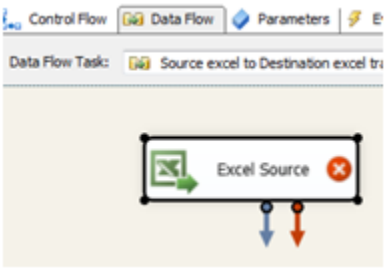


Figure 6.18

Step 8. Configure Excel Source

8.1 Red Cross mark on excel source indicates that, it's not configured yet. Double click the excel source. It will show up a dialog box something like this

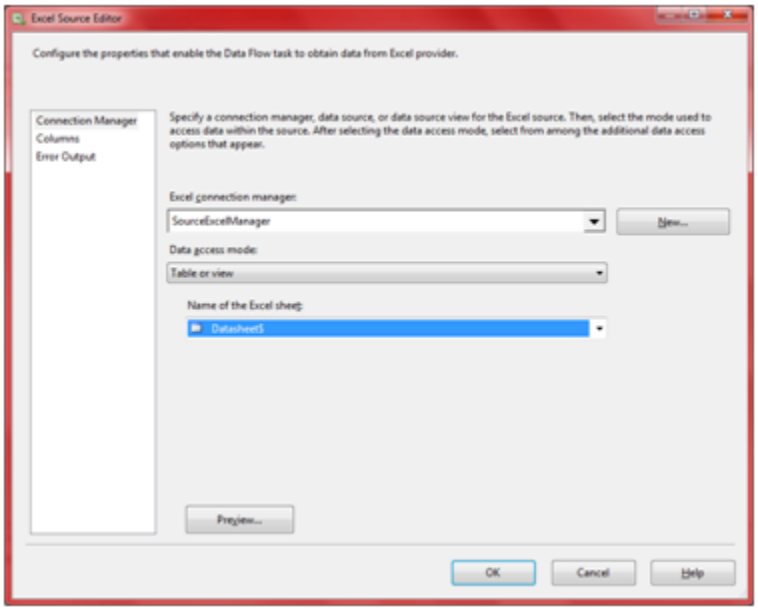


Figure 6.19

8.2 Select Data Source as "SourceExcelManager", Data Access Mode as "Table or View" and Name of the sheet as "DataSheet1" (Name of the sheet in the excel file).

Note: This Excel Source will perform the Extraction Task (E) in the ETL process

Step 9. Create Derived Column

9.1 From the SSIS toolbox from Transformation group drag Derived column to SSIS designer.

Step 10. Connect Source to Derived Column

10.1 Click the Excel source added in prior step.

You can see a small blue arrow attached to the source. We call it “Data Flow Path”.

Figure 6.20

Data Flow Path: It lets you define how data will flow.

Click on the blue arrow and connect it to Derived Column.

Figure 6.21

Note: we will speak about the red arrow in one of the future article in the series.

Step 11. Configure the derived column

11.1 Double click the derived column. Popup looks like follows.

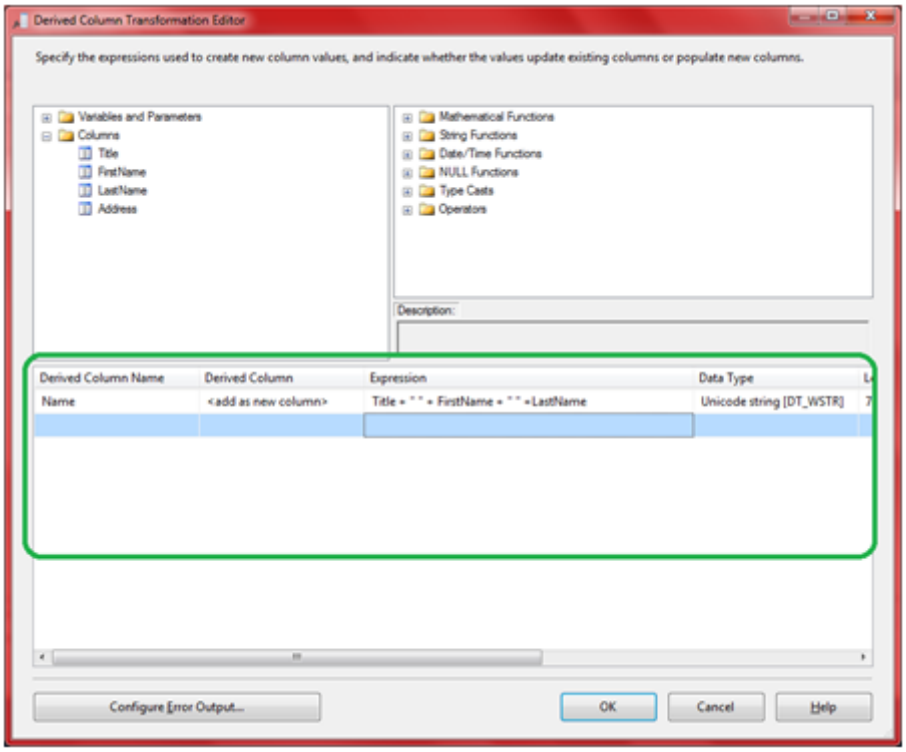


Figure 6.22

11.2 Put down Derived Column Name as Name, Select Derived Column as <add>and expression as Title + " " + FirstName + " " + LastName

11.3 Click Ok.

Note: This DerivedColumn will perform the Transformation Task (T) in the ETL process

Step 12. Create Excel Destination

12.1 Add Excel Destination from the Destination group in SSIS toolbox.

Note: This Excel Destination will perform the Load Task (L) in the ETL process

Step 13. Connect Derived Column to Excel Destination

13.1 Just like step no 10, connect derived column to excel destination.

Step 14. Configure Excel Destination

14.1 Double click the Excel destination, popup looks like follow.

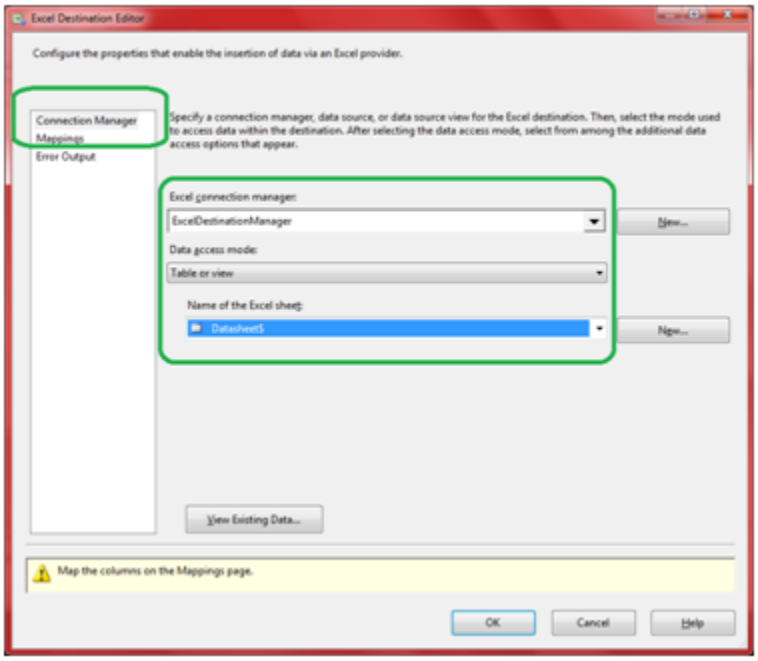


Figure 6.23

14.2 Set connection Manager to "ExcelConnectionManager", Data Access mode to "Table or View" and Name of the excel sheet to "Datasheet1".

14.3 Click on mapping and make sure its proper, if not make sure to do it before proceeding.

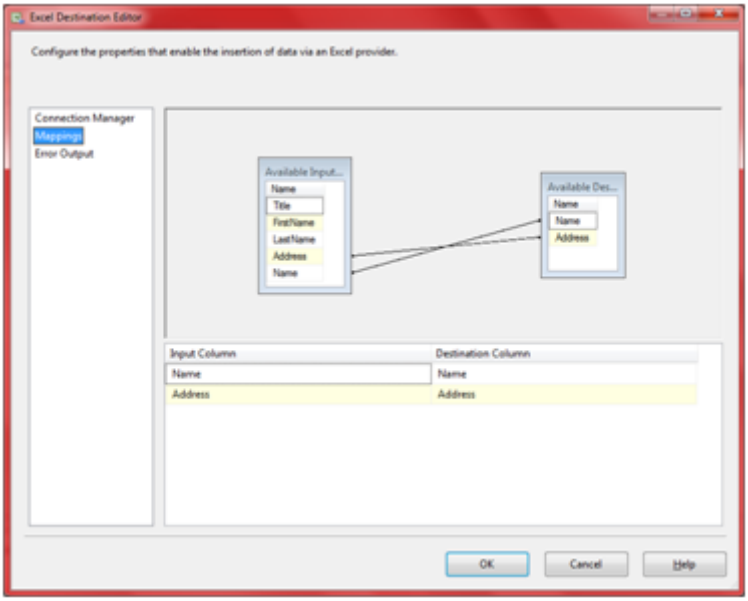


Figure 6.24

Note: In our case, mapping will be already done by the IDE itself (because names of columns are matching).

14.4 Click ok.

Step 15. Execute package

15.1 Press F5.

On successful execution you will get a screen something like this.

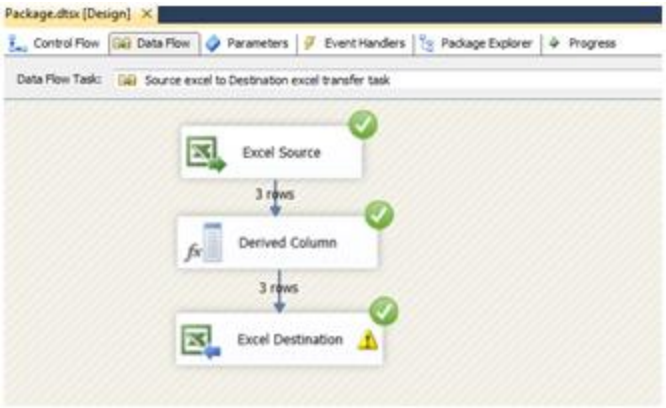


Figure 6.25

16.2 Open the Result.xlsx file and confirm the output.

	A	B	C
1	**Name**	**Address**	
2	Mr. Kenen Joshep	USA	
3	Mrs. Kamala Kumar	Chennai	
4	Mrs. Sheetal Raju	Chennai	
5			

Figure 6.26

REFERENCES:

1. LearnBI [online] Available: < http://www.learnbi.com/blog/category/bi-intro/ > [Accessed: 24-March-2018].
2. SearchBusinessAnalytics [online] Available: <http://searchbusinessanalytics.techtarget.com/definition/business-intelligence-BI> [Accessed: 24-March-2018].
3. Ramesh Sharda, Dursun Delen, Efraim Turban. Business Intelligence: A Managerial Perspective on Analytics (3rd Edition). Pearson; 3 edition (December 16, 2013).
4. LearnDataModeling.com [online] Available: < http://learndatamodeling.com/blog/advantages-of-business-intelligence/> [Accessed: 24-March-2018].
5. tutorialspoint (simplyeasylearning) [online] Available: <https://www.tutorialspoint.com/power_bi/index.htm> [Accessed: 24-March-2018]
6. Technopedia [online] Available: <https://www.techopedia.com/definition/1184/data-warehouse-dw>_ [Accessed: 24-March-2018]
7. Efraim Turban , Ramesh Sharda, Dursun Delen, David King. Business Intelligence (2nd Edition). Prentice Hall; 2 edition (July 28, 2010)
8. IBM®IBM Knowledge Center [online] Available: <https://www.ibm.com/support/knowledgecenter/en/SSGU8G_11.50.0/com.ibm.whse.doc/ids_ddi_344.htm> [Accessed: 24-March-2018]
9. Business Intelligence Tutorial. Version 7. IBM® DB2® Universal Database.
10. cornerstone [online] Available: <https://www.cornerstoneondemand.com/understanding-business-performance-management> [Accessed: 25-March-2018]
11. Neđo Balaban, Katarina Belić, Marko Gudelj. Business Process Performance Management: Theoretical and Methodological Approach and Implementation. Management Information Systems, Vol. 6 (2011), No. 4, pp. 003-009
12. tutorialspoint (simplyeasylearning) [online] Available: < https://www.tutorialspoint.com/performance_management/index.htm > [Accessed: 25-March-2018]
13. geographica [online] Available: < http://www.blog-geographica.com/2016/11/15/how-data-mining-is-used-to-generate-business-intelligence/> [Accessed: 25-March-2018]

14. Galit Shmueli, Peter C. Bruce, Mia L. Stephens, Nitin R. Patel. Data Mining for Business Analytics: Concepts, Techniques, and Applications with JMP Pro. Wiley; 1 edition (June 13, 2016)

15. B2C [online] Available: < https://www.business2community.com/business-intelligence/data-mining-techniques-in-crm-to-improve-data-quality-management-01307372> [Accessed: 25-March-2018]

16. TechTarget [online] Available: <http://searchbusinessanalytics.techtarget.com/definition/data-visualization> [Accessed: 25-March-2018]

17. Software Advice [online] Available: < https://www.softwareadvice.com/bi/#top-products> [Accessed: 25-March-2018]

18. Qingu Zahng, Richards S. Segall. WEB MINING: A SURVEY OF CURRENT RESEARCH, TECHNIQUES, AND SOFTWARE. International Journal of Information Technology & Decision MakingVol. 7, No. 4 (2008) 683–720 c World Scientific Publishing Company

19. mssqltips [online] Available:
< https://www.mssqltips.com/sqlservertutorial/3700/my-first-sql-server-business-intelligence-project-tutorial/> [Accessed: 25-March-2018]

20. learndatamodeling [online] <http://learndatamodeling.com/blog/business-intelligence-tutorial/> [Accessed:25-March-2018]

21. bi-survey [online] < https://bi-survey.com/top-business-intelligence-trends > [Accessed: 25-March-2018]

22. codeproject [online] < https://www.codeproject.com/Articles/751447/Learn-Microsoft-Business-intelligence-step-by-step > [Accessed: 25-March-2018]

23. P. X. Chiem, "Kodak Turns Knowledge Gained About Patents into Competitive Intelligence," Knowledge Management, 2001 , pp. 11-12; Y-H. Tsenga, C-J. Linb, and Y-I. Linc, "Text Mining Techniques for Patent Analysis," Information Processing & Management, Vol. 43, No. 5, 2007, pp. 1216-1245.