TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Economics

Business Intelligence Tools for Supporting Core Web Activities in a Small Company

Master Thesis

Study programme:	N6208 Economics and Management
Study branch:	Business Administration
Author:	**Manomeet Gupta**
Thesis Supervisors:	Ing. Athanasios Podaras, Ph.D. Departmens of Informatics

Liberec 2020

TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Economics

Master Thesis Assignment Form

Business Intelligence Tools for Supporting Core Web Activities in a Small Company

Name and surname: **Manomeet Gupta**
Identification number: E18000609
Study programme: N6208 Economics and Management
Study branch: Business Administration
Assigning department: Departmens of Informatics
Academic year: **2019/2020**

Rules for Elaboration:

- Business Intelligence solutions in modern business.
- Problem definition regarding selected web based processes in a specific company.
- Description of proposed tools and methods.
- Proposal of a business intelligence solution to support the selected processes.
- Discussion, analysis of the results.
- Conclusions and possible further expansion of the inferred output.

Scope of Graphic Work:
Scope of Report: 65 normostran
Thesis Form: printed/electronic
Thesis Language: English

List of Specialised Literature:

DUMAS, M., M.L. ROSA, J. MENDLING, H. A. REIJERS. Fundamentals of Business Process Management. Second Edition. Springer 2018. ISBN 978-3-662-56508-7.
GROSSMANN, W., S. RINDERLE-Ma. Fundamentals of Business Intelligence Data-Centric Systems and Applications. Springer, 2015. ISBN 978-3-662-4653-1-8.
SHERMAN, Rick. Business Intelligence Guidebook: From Data Integration to Analytics . Newnes, 2014. ISBN 978-0-12-411528-6.
MARCEL, Patrick, Esteban ZIMÁNYI. Business Intelligence. Springer, 2017. Lecture Notes in Business Information Processing, 280. ISBN 978-3-319-61163-1
KOHTAMÄKI, Marko. Real-time Strategy and Business Intelligence. Palgrave Macmillan, 2017. ISBN 978-3-319-54845-6.
VAISMAN, A., E. ZIMÁNYI. Data Warehouse Systems – Design and Implementation. Springer, 2014. ISBN 978-3-642-54654-9.
PROQUEST. 2019. Databáze článků ProQuest [online]. Ann Arbor, MI, USA: ProQuest. [cit. 2019-09-27]. Dostupné z: http://knihovna.tul.cz/

Thesis Supervisors: Ing. Athanasios Podaras, Ph.D.
Departmens of Informatics

Date of Thesis Assignment: October 31, 2019
Date of Thesis Submission: August 31, 2021

L.S.

prof. Ing. Miroslav Žižka, Ph.D.
Dean

doc. Ing. Klára Antlová, Ph.D.
Head of Department

Liberec October 31, 2019

Declaration

I hereby certify, I, myself, have written my master thesis as an original and primary work using the literature listed below and consulting it with my thesis supervisor and my thesis counsellor.

I acknowledge that my bachelor master thesis is fully governed by Act No. 121/2000 Coll., the Copyright Act, in particular Article 60 – School Work.

I acknowledge that the Technical University of Liberec does not infringe my copyrights by using my master thesis for internal purposes of the Technical University of Liberec.

I am aware of my obligation to inform the Technical University of Liberec on having used or granted license to use the results of my master thesis; in such a case the Technical University of Liberec may require reimbursement of the costs incurred for creating the result up to their actual amount.

At the same time, I honestly declare that the text of the printed version of my master thesis is identical with the text of the electronic version uploaded into the IS/STAG.

I acknowledge that the Technical University of Liberec will make my master thesis public in accordance with paragraph 47b of Act No. 111/1998 Coll., on Higher Education Institutions and on Amendment to Other Acts (the Higher Education Act), as amended.

I am aware of the consequences which may under the Higher Education Act result from a breach of this declaration.

July 23, 2020 Manomeet Gupta

Anotace

Tato práce se zaměřuje na aplikaci nástrojů obchodní inteligence s cílem optimalizovat webové aktivity a obchodní rozhodnutí na základě webových dat v malé společnosti. Primárním cílem je navrhnout řešení datového skladu, které usnadní ukládání velkých objemů dat, optimalizuje webové aktivity společnosti a usnadňuje obchodní rozhodnutí na základě webových dat.

Teoretická část informuje čtenáře o konceptu obchodní inteligence a objasňuje fungování systému správy databází a jak vyvinout vyhovující architekturu pro systém správy databází. Kromě toho jsou důkladně analyzovány nástroje pro modelování podnikových procesů, které jsou využívány pro dynamické znázornění obchodních rozhodnutí, která mohou být odvozena na základě webových dat.

Praktická část se zabývá návrhem efektivního datového skladu pro ukládání dat a návrhem dotazů pro zobrazení a analýzu dat podle požadavků uživatele. Praktická část také ilustruje několik obchodních procesů používaných pro webové aktivity, které jsou podporovány navrhovaným datového skladu. Omezení jsou analyzována v diskusní části a souvisí s neschopností začlenit realistická data a odhodláním postavit datový sklad místo využití webových analytických řešení.

V závěru jsou vysvětleny dosažené výsledky a důsledky budoucího výzkumu.

Klíčová slova

Databáze, obchodní zpravodajství, datový sklad, webové aktivity, malá restaurace

Annotation

The present thesis focuses on the application of business intelligence tools to optimize web activities and business decisions based on web data in a small company. The primary objective is to propose a data warehouse solution which facilitates the storage of large data volumes, optimizes company's web activities, and facilitates business decisions based on the web data.

The theoretical part informs the reader about the concept of business intelligence and sheds some light on the functioning of database management system and how to develop a suitable architecture for the database management system. Moreover, the business process modelling tools which are utilized for the dynamic representation of the business decisions that may be inferred based on web data are thoroughly analyzed.

The practical part deals with designing an efficient data warehouse for the storage of data and to design queries to display and analyze the data as required by the user. The practical part also illustrates few business processes used for web-based activities which are supported by the proposed data warehouse design. Limitations are analyzed in the discussion section and are related to the inability to incorporate realistic data, and the determination to construct a data warehouse instead of utilizing web analytics solutions. Finally, the achieved results as well as the implication of future research are explained in the conclusions section.

Keywords

Databases, Business Intelligence, Data Warehouse, Web Activities, Small Restaurant

Table of Content

List of Figures

List of Tables

List of Abbreviation & Symbols

BI – Business Intelligence

DBMS – Database Management System

ETL – Extraction, Transformation and Loading

OLAP - Online Analytical Processing

ERD – Entity-Relationship Diagram

SQL - Structured Query Language

BPMN - Business Process Model and Notation

ICT – Information and Communication Technology

RDBMS – Relational Database Management System

MOLAP – multidimensional OLAP

ROLAP – relational OLAP

HOLAP – hybrid OLAP

IMS – Integrated Management System

COBOL – Common Business Oriented Language

INGRES – Interactive Graphics and Retrieval System

OODBMS/ODBMS – Object Oriented DBMS

ODMG – Object Data Management Group

OMG – Object Management Group

PK – Primary Key

BCNF – Boyce-Codd Normal Form

NF – Normal Form

XMLA – XML for Analysis

MDX – MultiDimension eXpressions

API – Application Programming Interface

PHP – Hypertext Pre-processor

RFQ – Request for quote

ERP – Enterprise Resource Planning

UML – Unified Modelling Language

KDD – Knowledge Discovery from Data

SME – Small and Medium Enterprise

Introduction

In the modern society with the globalization and ever-changing economy, Information and communication technology (ICT) tools are used globally by individuals and enterprises for the implementation of routine activities. The technology-enabled basic transactions and communication is undoubtedly one of the main characteristics of the modern times. Individuals are strongly dependent on ICTs for daily activities and enterprises need to incorporate ICTs to increase their competitive advantage. The provision of electronic customer services is an issue of paramount importance for modern enterprises the government and the global economy. The emergence of Electronic commerce has dramatically altered the way of implementing daily tasks like purchasing of goods, financial transactions, online orders, public services and more. Nevertheless, modern technologies have triggered the exchange of big data volumes and the necessity, for companies and organizations, to invest on the purchase or the development of complex and efficient information systems which may support the extraction, storage and processing of large volumes of the daily produced information. Such systems are known as Business Intelligence systems and are aimed to support core business processes, strategic business decisions and complex electronic transactions. (QUADDUS, 2015)

The current diploma thesis is oriented towards the application of Business Intelligence methods to support the core web activities of a small company. Its main goal is the improvement of current as well as the recommendation of new web-based processes of the selected company with the help of business intelligence data warehouse solution, that can serve as the data repository related to the implementation of the aforementioned processes. The achievement of the above stated goal relies on the fulfillment of the following objectives:

- Pinpointing the gaps in existing web activities and services as well as the absence of important web services that are common to other similar companies and their incorporation to ameliorate the web-presence of the selected business,
- Conceptualization of a business intelligence data warehouse solution which enables the storage and analysis of the data which is generated via the implementation of these processes

- Creation of the logical and physical database solution and the use of aggregate queries and analytical reports which demonstrate its practical value for the selected web activities and the decision making based on the available web data

- Illustration of the improved as well as the newly proposed web activities with sophisticated business process modeling tools after the incorporation of the proposed data warehouse schema.

The company is a local restaurant in Czech Republic. The restaurant is mainly focused towards providing wide range of Czech ready meals, fast dishes, sea specialties as well as international gastronomy. The study includes current Business Intelligence theories and solutions that can be applied in this case and motivates to improve the current web-based activities in the company and to optimize the process of data storage and handling.

The thesis is divided into theory and practical applications based on the theories studied. The theoretical part of the thesis canvases the present theories in the domain of Business Intelligence, Database Management Systems, and its types, creating and handling Data Warehouses, OLAP and its operations and few of the data mining concepts.

The practical application is based on the theoretical analysis. It examines the current system architecture of the website and suggests ways to improve the architecture. It also suggests optimal way the restaurant can store and handle the data of customers and the food it provides. The thesis also aims at implementing Data Warehouse solutions and suggest an optimal model to denote data structure hierarchy. Moreover, the thesis intends to suggest processes based on OLAP operations and data analysis to optimize the business of the restaurant and to widen its reach in business.

In the practical part, the implementation of the proposed design is illustrated, and the data warehouse design is shown. It also demonstrates the process of data entry, transformation and loading for the given case.

Motivation and Problem Statement

With the advancement in development of computing and cloud-based networking, it has become especially important for business to increase their online presence. Business Intelligence tools helps the business to obtain a competitive edge in the present fast paced environment. Currently, the restaurant does not use any kind of such ICT tools to add more value to their business. With the current tools and technology used by the business, they have no proper means to store and analyze the data they obtain and turn it into significant profits. The current scenario of the restaurant is based on basic data management without any use of BI tools. With current ever-changing market, any type of competitive edge might prove helpful for a small or medium business enterprises.

The growing complexity in the field of information technology, has made it obligatory for enterprises to be able to handle large volumes of data, thus making tools of business intelligence much more valuable and significant. (Rausch, 2013, s. 5) The proposed model aims to improve the present scenario by introducing business intelligence tools like data warehouse and online processes. These tools will help the enterprise to grow their business and increase their reach to much more further demographics of people. These tools will also result in decreasing the errors in decision making by the owners of the enterprise thus, increasing the efficiency of their business and maximizing profits. The automation of the process of data storage will also support efficient and quicker response to changes in the market rather than manually analyzing the situation and changing the inputs, to obtain the desired outputs.

The current diploma thesis aims at introducing the basics of database, database management and business intelligence concepts with practical examples of the mentioned tools. The study highlights on the collection of data from different sources and showcases best model possible using conceptual model, logical model, and a physical model of the database design. This study also uses different charts such as flowcharts and business process model and notation (BPMN) to emphasize on the web activities that can be supported by the business to maximize their reach and to increase the ability of the enterprise to collect and analyze the data collected.

1. Business Intelligence – Literature Review

This chapter provides an overview on the current available literature works regarding Business Intelligence (BI), Database Management Systems and Data Warehouses. It contains the following sections.

I. Overview – This section gives a brief description about business intelligence, its history of development over the last several decades and highlights some of its characteristics and changes since the beginning of development in the concept of business intelligence.

II. Database Management System – This section outlines the characteristics of database and the types of databases used in the modern business environment. It also focuses on the description of development in the field of databases and database management systems. This section also sheds a light on different data models such as conceptual, logical, and physical data models.

III. Relational Database Management System -This section highlights the relational database management system and few of the concepts to develop such database management system. It also explains the important parts of developing a RDBMS and how to depict these databases using diagrams such as Entity-Relationship diagrams, which is one of the most used diagrams to showcase a database management system.

IV. Data Warehouse and Data Marts – This section of the thesis explains the concept of data warehouses and data marts and steps involved in defining them. It shows an in-depth study of architecture of a data warehouse and also explains the different hierarchies available while constructing a data warehouse or data mart.

V. ETL Process – ETL stands for Extraction, Transformation and Loading processes which are used to prepare and load data in the database. This section highlights how to use these processes to ensure proper allocation of data.

VI. OLAP and OLAP Operations – This section shows the use of OLAP (Online Analytical Processing) and its tools used for processing and analyzing data.

VII. Tools for supporting Web Activities – This section explains the concept of web services and tools that are used in order to support these services. It also shows the steps involved in defining the web databases. This section also defines the processes involved and elements

of the business process involved. There is also a brief introduction to the different modelling tools involved in defining these business processes.

VIII. Data Mining Concepts – This section of the thesis, elaborates on the concept of data mining and its types. It gives a short description about data mining and defines the forms of analysis which are classification and predictions, highlighting the difference between both and where can these methods be used.

1.1. Business Intelligence – Overview

The first reference to Business Intelligence can be found in 1958 in a paper of H.P. Luhn. According to him the terms "Intelligence" means "the ability to apprehend a interrelationships of presented facts in such a way as to guide action towards a desired goal" and "Business" as "a collection of activities carried on for whatever purpose, be it science, technology, commerce, industry, law, government, defense, et cetra", he specifies business intelligence systems as *"[an] automatic system [that] is being developed to disseminate information to the various sections of any industrial, scientific or government organization."* (Grossmann, 2015, s. 1)

This definition did not take effect until 30 years, when in 1989, Howard Dresner coined the term Business Intelligence as an umbrella term for a set of concepts and methods to enhance business decision making, employing systems based on facts. Since then many definitions have been given defining Business Intelligence as a major part of all roles in an organization. (Grossmann, 2015)

Today, we can find many definitions for Business Intelligence, one of which as defined by Dedić N. & Stanier C *"Business intelligence (BI) comprises the strategies and technologies used by enterprises for the data analysis of business information."* (Dedić, 2016, s. 255-236)

According to Talaoui, Kohtamäki, Rabetino (2017) "*A process that transforms internal and external data into knowledge and communicates it to the business user via a set of applications.*" (KOHTAMÄKI, 2017, s. 40)

1.1.1. Big Data and its Characteristics

The extensive use of internet has resulted in large amount of data getting produced and potentially interesting to analyze it. Big Data plays a very crucial role in designing the Business Intelligence architecture and thus is essential to be studied as a part of BI. Big Data as defined by Snijders, C., Matzat, U., Reips, U. (*p.1–5* (2012)) *"data whose sizes are beyond the ability of commonly used software tools to capture, curate, manage, and process them within a tolerable elapsed time."* (MARCEL, 2017, s. 79) Or it can also be defined as any data which includes the properties of the mostly all "**V's**".

The key challenges in handling Big Data are its characteristics (SHERMAN, 2014), i.e.

- Data Volume – Processing Huge amount of Data from all the services,
- Data Velocity – much of the data today is time sensitive, thus increasing the responsibility of decreasing time between the attainment of data and usage of it makes a lot of difference. Thus, data velocity is an important aspect of big data.
- Data Variety – In today's world the source of data is continuously expanding, thus increasing the diversity of data available,
- Data Veracity – trustworthiness of data,

These characteristics are generally defined as the '4Vs' of Big Data. All the issues of handling and working with big data are reliable on these 4 characteristics. They decide how complex the structure of BI must be to handle everyday load of the organization.

1.1.2. BI Architecture

Business Intelligence (BI) is divided into many parts to define the proper structure of it. It consists of ETL process, Data Warehouse, Data Marts, OLAP operations and Data Mining.

ETL (Extraction, Transformation and Loading) defines the sequence of the processes that must be defined while loading the data onto a data warehouse. According to this the BI tools can **Extract** data from multiple internal and external sources as required by the user. The data must be **Transformed** into a suitable format which is standardized by the user via a set of processes for migration of data. The last step is **Loading** the data into the Database or Data Warehouse. While transforming the data remains in a temporary storage facility known as **Staging Area.**

Data Warehouse were introduced in early 1990s as a tool for the businesses to get ahead of the competition in the world. Due to drastic growth in the domain of internet and information systems, the need to handle large volumes of data was required. The traditional operational and transactional databases used in that time did not satisfy the need to handle large volumes of data as they were only designed to handle data related to everyday activities rather than large data ranging from an year old to most recent data. Furthermore, if the user wanted to analyze a body as a whole data from different databases had to be integrated to analyze it. This integration of different data from different sources made it much more complex to analyze these situations. To tackle these problems Data Warehouse was suggested as a solution.

The classical Data Warehouse was defined by Inmon, *"as a collection of subject-oriented, integrated, nonvolatile, and time-varying data to support management decisions."* (VAISMAN, 2014, s. 72)

Data Marts are considered a subset of Data Warehouse. Like Data Warehouse, Data Marts can store data but with lesser volume. The only difference between Warehouse and Mart is Data Marts help a specific part of the company (departments) in terms of decision making.

OLAP (Online Analytical Processing) is an approach used to solve multidimensional analytical queries. It is a tool used to analyze large volumes of data thus generally a viable to analyze the data in data warehouse and data marts. OLAP can be divided into 3 major categories namely, multidimensional OLAP (MOLAP), relational OLAP (ROLAP) and hybrid OLAP (HOLAP). To analyze data, ROLAP can be used to analyze the data rather than segregating it. On the other hand, mapping can also be used to do the same which differentiates data and divides them into more relatable sections. To achieve that schemas are developed to support the activities. The 2 most used are star and snowflake schemas.

Data Mining is a field in computer science and is a process of mining models from large data sets, combining statistical and artificial intelligence methods with database management. Data mining can predict information by using the collected data in data warehouses and can help executives make decisions based on it. Data mining uses various tools to predict data which include regression, clustering, association, classification and many more.

1.2. Database Management System

Database is a collection of data based on their based on their logical relationship. Database management system (DBMS) are applications that allows the users to store and interact with the data stored in databases. The first database management system was developed in 1960s by Charles W. Bachman, calling it Integrated Database System. Later IBM designed their own DBMS called Integrated Management System (IMS).

Later with the development of computing system, many different types of database systems became available to the public, increasing the demand for a standard database system. In mid 1960s, Bachman's Database Task Group designed and standardized the language called, Common Business Oriented Language (COBOL), later known as CODASYL approach. Unsatisfied by the complexity of CODASYL approach and the IMS model, Edgar Codd wrote a series of papers outlying ways to construct databases, later titled as A Relational Model of Data for Large Shared Data Banks, which portrayed new approach for storing data and handling large databases.

In 1973, Michael Stonebraker and Eugene Wong decided to work on the ideas of Edgar Codd and decided to research relational database systems, this project was called INGRES (Interactive Graphics and Retrieval System), which worked with the query language QUEL, which was later replaced by SQL developed by IBM. The RDBMS was unable to handle unstructured data for which NoSQL was developed which was more resource intensive language but was able to handle large volumes of unstructured and structured data. Later in 1985, Object Oriented DBMS (OODBMS) was developed when a lot of articles were published into the research for alternate database management systems. In this type of system, the information is presented in form of an object. According to Harrington (2000) *"An object-oriented program is made up of objects, each of which has attributes and methods. The objects interact by sending messages to one and another."* (Harrington, 2000, s. 24)

1.2.1. Characteristics of DBMS

According to Elmasri and Navathe, *"The DBMS is a general-purpose software system that facilitates the processes of defining, constructing, manipulating, and sharing databases among various users and applications."* (ELMASRI, 2011, s. 5) Defining the database involves specifying the data types, structure, hierarchies, constraints, and relationships

between the data to be stored in the database. Constructing the database is the process of storing the data into the database and defining the relations between the data. Manipulating the database includes functions such as queries to retrieve specific data or to analyze specific set of data stored in the database. This also includes updating the database and reflecting changes in it. Sharing the database involves multiple users and applications to use the database simultaneously.

An application program interacts with database in form of queries. Queries are lines of codes depending upon the language used by the DBMS used to communicate with the database.

Another important aspect of DBMS is security and maintenance. Security is the protection against software or hardware malfunction or protection against unauthorized or malicious access. Maintenance means the real time updating of data as required with changing time.

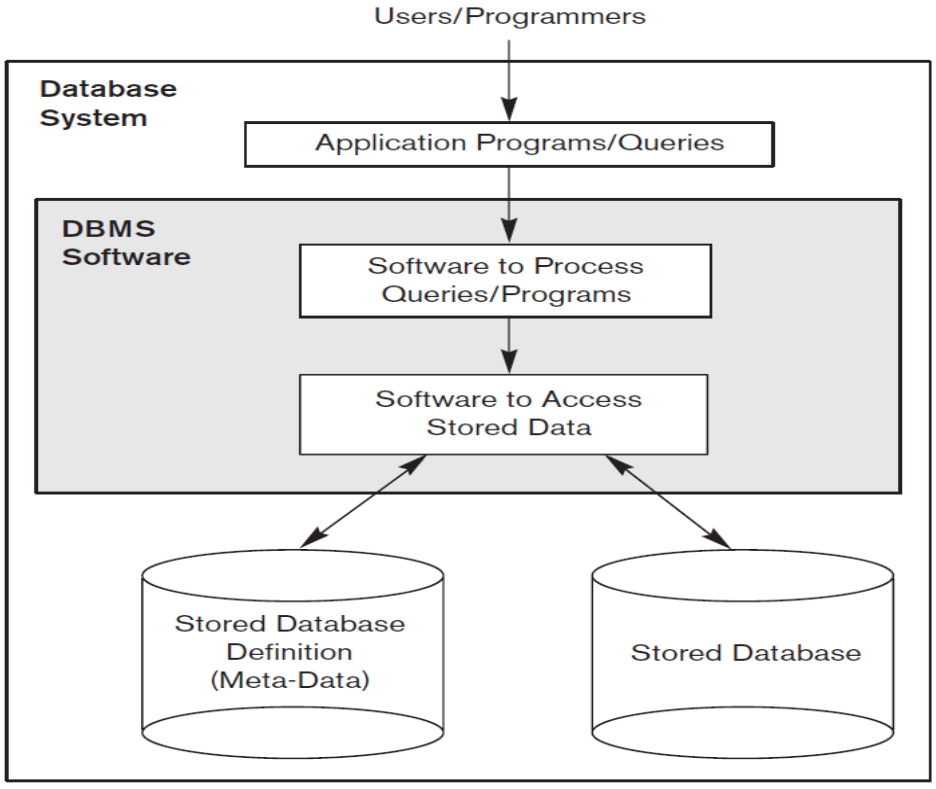


Figure 1. A Simplified Database System Environment

Source; Elmasri and Navathe (*p. 5*, 2011)

The main Characteristics of the Database approach are the following:

- Self-describing nature of a database system – A fundamental characteristic of database is that not only contains the data but also contains the definition of database structure and constraints. This information is stored in a catalogue called **meta-data**, which describes the structure of the primary database.

- Insulation between programs and data, and data abstraction – Generally, in DBMS the structure of data file is stored separately from access of program, which is called program-data independence. Applications can access the data through operations through their names and arguments regardless of the implementation of operations, this is called program-operation independence. These characteristics are made possible due to data abstraction. A user in DBMS works with a conceptual representation which does not include how the data is stored in the database or how the operations are implemented.

- Support of multiple views of the data – A database can be used by multiple users which may require different "views" of data stored in the database. View can be any subset of the database derived from it but is not permanently stored.

- Sharing of data and multiuser transaction processing – A multiuser DBMS, should allow the access to multiple users and applications at the same time, which is essential if the data is to be used and stored in the same database.

1.2.2. Data Models

According to Sumathi and Esakkirajan, "*Data model is collection of conceptual tools for describing data, relationship between data, and consistency constraints.*" (SUMATHI, 2007, s. 13) Data Models depict the structure of data at logical level. A data model is the set of conceptual constructs available for defining a schema.

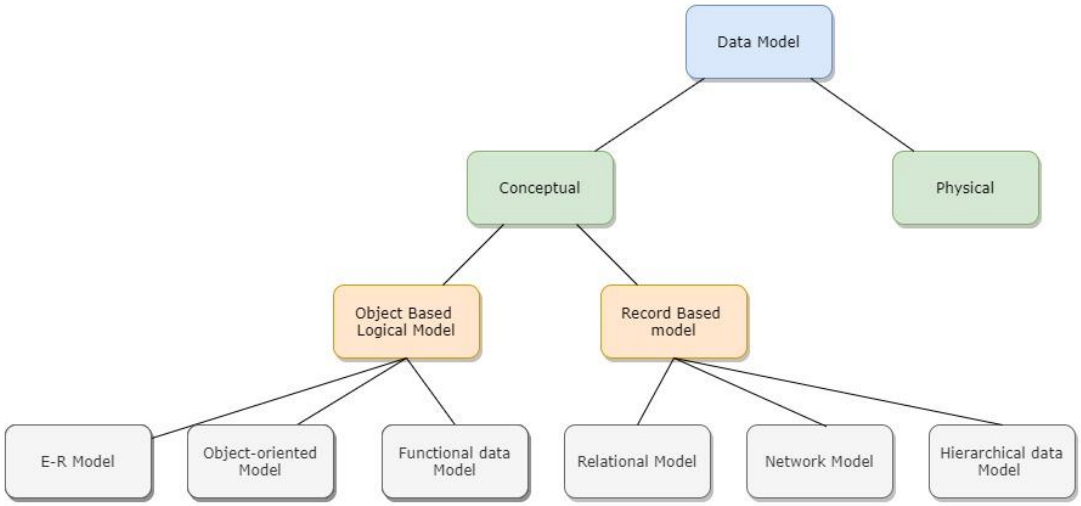


Figure 2. Types of Data Models

Source; Own Contribution

According to Sumathi and Esakkirjan, (2007), the Data Models can be divided into two major types which are conceptual and logical models. There is another type of model which is sometimes referred as physical data models. Historically, there were three important data models used earlier to the development of the models used today, these are **hierarchical, network and relational models**. These data models are especially important as the modern-day data models were developed based on them. The hierarchical and network models were developed in 1960s and 70s and were based on organizing primitive data structures by establishing connections or links between the data stored. (SUMATHI, 2007)

Database schemas are the description of database and are specified during the design of the database and generally are not changed with time. These schemas are displayed as **schema diagrams**, they generally depict the structure of record types stored in the database. The schema diagram is restricted to showing only some details of the schema, such as name of data items and some constraints. It can only display basic information as detailed diagram of the relationship between data and constraints can be difficult to depict. (ELMASRI, 2011).

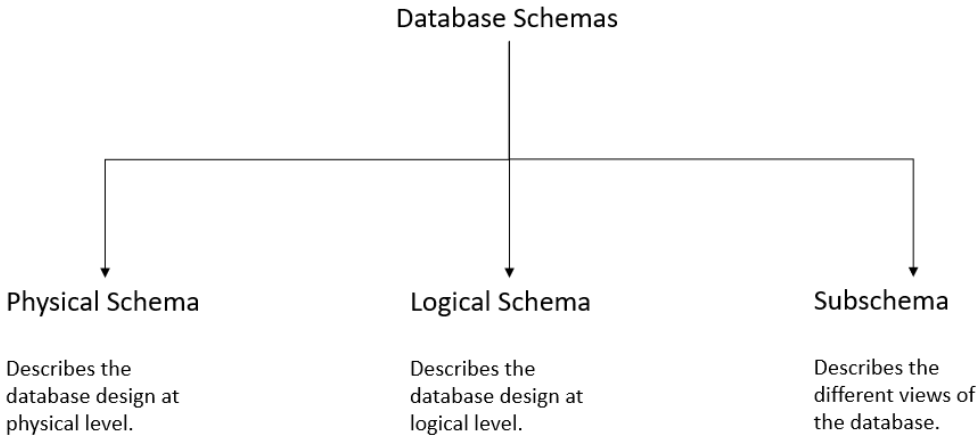


Figure 3. Types of Database Schemas

Source; Own Contribution

According to Sumathi and Esakkirjan (2007), the database schemas can be divided into two types, physical and logical schema. The schema types are closely related to data models. The Physical schema describe the physical level design of the database whereas the logical schema defines the design at logical level. (SUMATHI, 2007)

1.2.3. Conceptual Model

"*Conceptual data models use concepts such as entities, attributes, and relationships.*" (ELMASRI, 2011, s. 31). **Entity** describes a real-world object as a part of the database management system. **Attribute** describes some properties of the entities that are useful for describing the database. **Relationship** describes the connection among the entities and association of them.

The Conceptual Data Model can be categorized into Object based Logical model and Record based model. The Object based model consists of E-R model (Entity-Relationship model), Object Oriented model and Functional data model. The Record based model can be classified into Relational model, Network model and Hierarchical data model. E-R model is one of the popular high-level conceptual data models used today. Object data models are comparatively a newer kind of high-level Conceptual Model, ODMG (Object Data Management Group) is a standard for object databases.

Two approaches used in designing conceptual data model are **Top-down design** and **Bottom-up design**. The decision to use which method depends upon the complexity of the target system and developer's experience. Top-down design is created by summarizing the requirements of the different users of the database, building a unique schema for the purpose. Later the view can be separated as required by the user. Whereas in Bottom-up design, a separate schema is built according to the requirement of each user and later these schemas are merged to form a single conceptual schema for the target database.

1.2.4. Logical Model

Logical data models add another layer of information to the conceptual model. The Logical model defines the structure of the elements and defines the relationship between these elements. **Normalization** of data must be done to create a logical model. Normalization is done to reduce data redundancy and improve data integrity. The most common logical model used is relational model. The design of logical model is independent of the DBMS. Logical Model define the flow of the data in the database.

1.2.5. Physical Model

Physical data model shows how the data is arranged and database is built in the DBMS. Physical design aims to customize the logical model targeted to a database management software (DBMS). It helps the user to visualize the database including keys, constraints, columns, indices, and other RDBMS features. This model also describes the relationship between the tables which addresses cardinality of the relationships. Physical data model also differs for other physical data model.

1.3. Object Oriented Database Management

An object-oriented database system (OODBMS or ODBMS) is a data management system which represents information in form of objects and used object-oriented programming languages. The concept of OODBMS was first developed in 1970s and the first commercial product was offered in 1980s. However, after so much research in the field of object-oriented databases, its users are extremely limited in the market today. OODBMS cannot compete with the growth of RDBMS, which is exceedingly popular in the database management software market. The major strength of these DBMS lies in the ability of it to manage complex relationship between data objects. Unlike RDBMS, which is capable of handling large volumes of data, OODBMS excels in handling reasonable

amount of data but with higher complexity than RDBMS can. (Object-Oriented Database Management System (OODBMS), 2011)

According to Sumathi and Esakkirjan, "*Object oriented databases are a combination of object-oriented programming and database technologies.*" (SUMATHI, 2007, s. 508) They define that the most important characteristics of OODBMS is that it is a combination of object-oriented programming and database technologies which gives it several advantages over relational database systems. These advantages include, the operation applies universally i.e. operations do not depend on the currently running database application, another advantage is that OODBMS supports complex data such as multimedia like audio, video or graphical depictions, polymorphism and dynamic binding allows the users to define the operation for one object and can share it for other objects by calling these operations when needed. A unique characteristic of object is that, they have an identity which is independent to the state of object, this allows the user to share the object in a distributed computing environment.

According to Elmasri and Navathe, "…*key feature of object databases is the power they give the designer to specify both the structure of complex objects and the operations that can be applied to these objects.*" (ELMASRI, 2011, s. 353) Another factor for increasing popularity in object-oriented databases is that they use the object-oriented programming languages. Due to the rising development in the field of OODBMS, there was a need to standardize the object-oriented database, which was done by many organizations like OMG (Object Management Group) and X3H7. Seeing the rise in demand for object-oriented databases, other relational database vendors have started to incorporate the features into existing relational databases, which gave rise to the development of ORDBMS (Object-Relational Database Management Systems).

1.4. Relational Database Management System

Relational database represents databases as a collection of relations. A relation can be described as a table of values. The data are stored in tables in this system, however physical storage of data is independent of the way this data is logically organized. The table and column name help the user to interpret the data and understand the relations. Each row depicts a group of related data. A row is called **tuple.** The column head are called **attributes** and the table is called **relation.**

1.4.1. Concept

Relational databases are defined by some basic properties, which can be attributes, tables, tuple, relational schema, degree, cardinality, relation key, attribute domain.

- Domain – A domain is a set of atomic values. Atomic means that each value in the domain is generally invisible as far as formal relational models are considered. Generally, the data type for the domain is specified, from which data values can be derived. The name of domain helps the users to interpret the data stored. Format or data type is also specified in the domain. These datatypes as the name suggests show the type of data stored in the domain. Example, char, varchar, string etc. These datatypes specify the restrictions for data that can be entered, and the space allotted for it.

- Relation Schema – According to Elmasri and Navathe, Relation Schema "*...R, denoted by R(A1, A2, ...,An), is made up of a relation name R and a list of attributes, A1, A2, ..., An.*" (ELMASRI, 2011, s. 62) A relation schema is used to define a relation between two or more attributes.

- Attributes – Each column in the relational model is called attribute, they are the properties that define the relations.

- Tables – In relational model, the information or data is always stored in tables. A table holds two properties rows and columns. Rows represent the records and columns represents the attributes.

- Tuple – Tuple is a single row in the table, it contains a set of single record.

- Degree – Total number of attributes (columns) in a relation is called the degree of relation. Types of degree are

- Cardinality – Total number of rows in a table. Each relationship in this model has associated with it a pair of cardinalities which describe minimum and maximum times an entity may participate in a relationship.

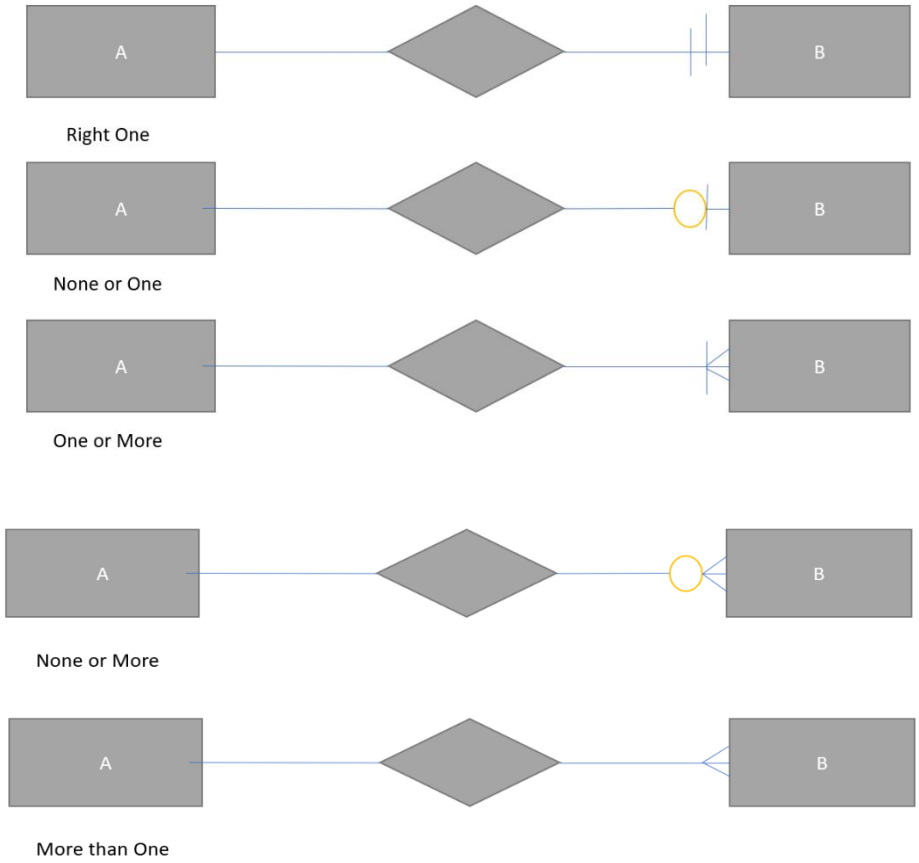


Figure 4: Types of Cardinality

Source; Own Contribution

- Relation key – Each row has one, two or multiple attributes, which is called relation key. These relation keys are essential part of the relational model, it makes sure referential and entity integrity is maintained.
- Primary Key – In a table, primary key (PK) is a field that has a unique value. The field can be identified through this value.
- Foreign Key – A foreign key is a field in another table (Table B) that is associated with the primary key in the first table (Table A).
- Entity Integrity – When a primary key is defined in a table, it enforces entity integrity. It means each row in the table can be identified through the primary key.

- Referential Integrity – It means that the relationships between tables should be consistent, i.e. the foreign key in one table should always agree with the primary key of another table.

Types of Cardinality in Relational Model are:

- One to One – Occurrence of one entity relates to just one occurrence in another entity.

One to One

Figure 5: One to One Cardinality

Source; Own Contribution

- One to Many – Occurrence of one entity relates to more than one occurrence in another entity.

One to Many

Figure 6: One to Many Cardinality

Source; Own Contribution

1.4.2. E-R Model and Relationship types

The Entity-Relationship Model (E-R Model) is constructed using top-down approach of database designing, starting with defining the entities, relationships and entity types and then designing the database into more details by adding constraints and attributes. Entity-relationship model depicts how the model is constructed in conceptual framework of database designing. E-R model is very commonly used by designers as a conceptual model for designing the database. Entity Relationship models are generally designed using **Entity-Relationship Diagrams (E-R Diagram)**.

The core components of an E-R diagram are the entity type, relationship type, attributes, primary key and cardinalities and roles. Entity is an object which exists and is differentiable from other objects. Entity type is a collection of similar entities i.e. entities

which has similar properties. Relationships show the association between different entities. Attributes display different properties of each entity. (SUMATHI, 2007, s. 60)

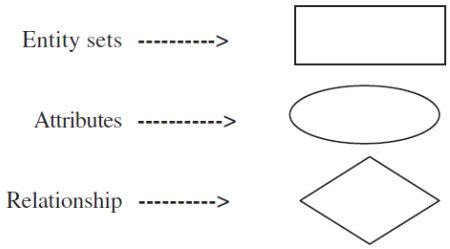


Figure 7: Shapes used in Entity-Relationship Diagram

Source; Sumathi and Esakkirjan (2007)

- To depict entity sets rectangle is used.
- An ellipse is used to depict attributes.
- Diamond is used to show relationships between different entities.
- Lines are drawn to show the link between attributes and entity sets and relationship sets.

Entity-Relationships many exist as unary, binary, ternary, n-nary relationship.

Figure 8: Unary Relationship

Source; Own Contribution

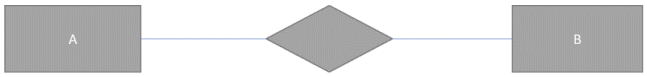


Figure 9: Binary Relationship

Source; Own Contribution

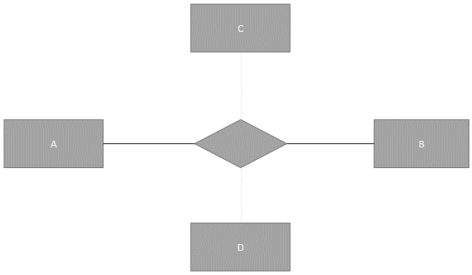


Figure 10: Quaternary Relationship

Source; Own Contribution

1.4.3. Normalization

"Normalization of data can be considered a process of analyzing the given relation schemas based on their FDs and primary keys to achieve the desirable properties of (1) minimizing redundancy and (2) minimizing the insertion, deletion, and update anomalies" (ELMASRI, 2011, s. 517) Normalization process was first introduced by Codd (1972), which takes a relation schema through a series of tests to certify it to be in a normal form. Normalization is a top-down approach process, in which each relationship is tested against criteria of normal form. Generally, the relations can be separated till 3rd normal form. Although 4th and 5th normal forms also exist when needed. Normalization can be considered as a filtering process designed to maintain the quality of data. Another definition of 3rd normal form was proposed by Codd and Boyce and is called Boyce-Codd normal form (BCNF). Normal forms are a standardized form for data to be arranged. Most of the databases are designed based on legacy models, so to achieve highest quality of design to meet the requirements, data is normalized. The main objective of normalization is to eradicate redundancy in the data.

1st Normal Form is defined as *"1NF disallows relations within relations or relations as attribute values within tuples. The only attribute values permitted by 1NF are single atomic (or indivisible) values."* (ELMASRI, 2011, s. 519) 1NF is the most basic normal form and contains only single atomic values. The only rule for designing 1NF is that there should be no repeating groups in the data.

2nd Normal Form is defined as *"A relation schema R is in 2NF if every nonprime attribute A in R is fully functionally dependent on the primary key of R."* by Elmasri and Navathe (ELMASRI, 2011, s. 519) Fully functionally dependent means all nonfunctional

key attributes are dependent on the primary key and no other key. No partial dependencies are allowed.

3rd Normal Form as defined by Codd is *"...a relation schema R is in 3NF if it satisfies 2NF and no nonprime attribute of R is transitively dependent on the primary key."* (ELMASRI, 2011, s. 519). The rule for defining 3NF is that, it should satisfy the 2NF rules and no transitive dependencies are allowed. Transitive dependency exists if there is an attribute in relationship model which neither a candidate key nor a subset of any other key. Elmasri and Navathe (ELMASRI, 2011, s. 519)

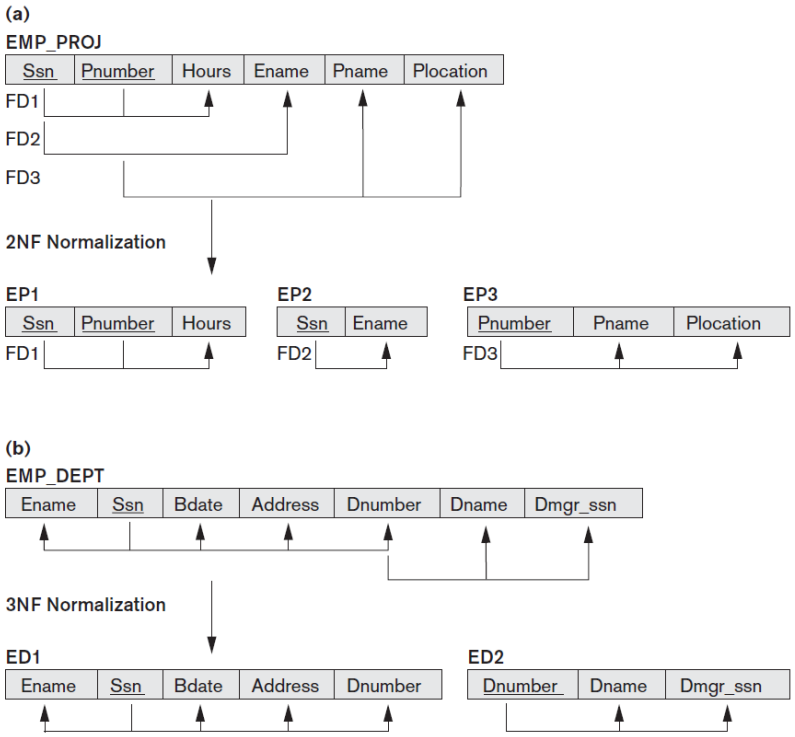


Figure 11: Example of 2nd and 3rd NF

Source; Elmasri and Navathe (*p. 524*, 2011)

Steps to normalize data is given by the figure below. This figure summarizes the process of normalization of data to normal forms.

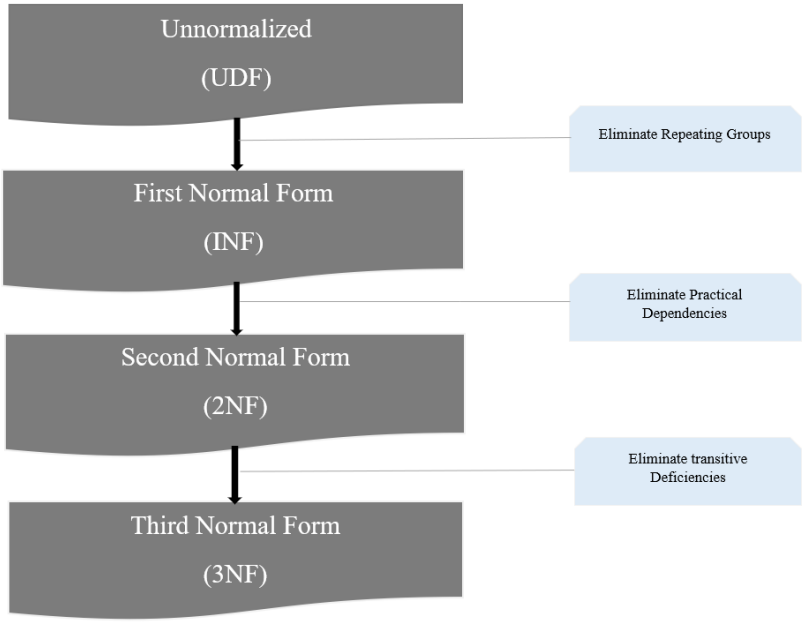


Figure 12: Steps for Normalization of Data

Source; Own Contribution

1.5. Data Warehouse and Data Marts

Data Warehouse is defined as *"…are (usually) large repositories that consolidate data from different sources (internal and external to the organization), are updated off-line (although as we will see, this is not always the case in modern data warehouse systems), and follow the multidimensional data model."* [6] (VAISMAN, 2014, s. 54) Data Warehouse and OLAP systems work on the principle of multidimensional model. In this model data is viewed from n-dimensional space, which is usually called a data cube, which is characterized by dimensions and facts. Dimensions are views used to analyze the data. Each dimension level represents granularity which means level of details. Data Warehouse can also be defined as a collection of subject-oriented, integrated, non-volatile and time-varying data to support management decisions. (VAISMAN, 2014)

- Subject-oriented means that data warehouse focuses on the analytical needs of different areas of department. These needs depend upon the type of work done by the organization.
- Integrated means that the data obtained from different sources (external or internal) must be combined.

- Non-volatile means the durability of data by not allowing data removal or modification, thus increasing the life span of data available.
- Time-varying shows the possibility of maintaining different values of the same information, also time when these changes took place.

Data Marts are a subset of data warehouses. Often in an organization, many departments exist. These departments do not necessarily require access to the whole database itself and rather requires a part of it. Data marts solve this problem by creating a smaller database from the data obtained from the data warehouse which is specialized for that specific department's needs. These data marts may be shared depending on the interest and needs of the other departments. Data marts generally share the same properties to the data warehouse from which it is derived from, thus a set or collection of data marts can be called data warehouse, which is called the **bottom-up approach**. In this design, smaller data marts are merged to create a large data warehouse. On the other hand, data marts can be obtained from a preexisting data warehouse, this approach is called the **top-down approach**.

1.5.1. Architecture

The architecture of a data warehouse is divided into many tiers which are, **the back-end tier, data warehouse tier, OLAP tier and front-end tier**.

- The back-end tier comprises of **extraction, transformation, and loading (ETL)** tools. These tools are used to feed the data into the various from different sources (the source maybe external or internal) of data. The data is extracted to a **staging area** where the data is transformed and then loaded into the warehouse.
- The data warehouse tier comprises of the central enterprise data warehouse and some smaller data marts for specific departments or divisions. Another important component of this tier is the **metadata** which means data about data. Metadata can be classified into two types namely, business, and technical metadata. The business metadata describes the meaning of data, organizational rules, policies, and constraints related to the data. The technical metadata describes the way the data is structured and stored in the system and the process to manipulate the data.
- The OLAP tier comprises of an OLAP server, which provides the user with a multidimensional view of the data stored in the data warehouse, regardless of the way the data is stored in the warehouse.

- The front-end tier is used to analyzing data and generating reports or in any form the data can be visualized. It contains client tools such as OLAP tools, statistical tools, and data mining tools.

There may be variations from the basic architecture depending on the needs of the enterprise. For example, an enterprise may opt to skip the data warehouse tier and instead of having a common enterprise data warehouse, use only small data marts for specific departments as constructing a data mart is much easier and less expensive when compared to designing a data warehouse.

To design a data warehouse the stages followed are requirement specification, conceptual design, logical design, and physical design. These are same as the steps explained for designing a database.

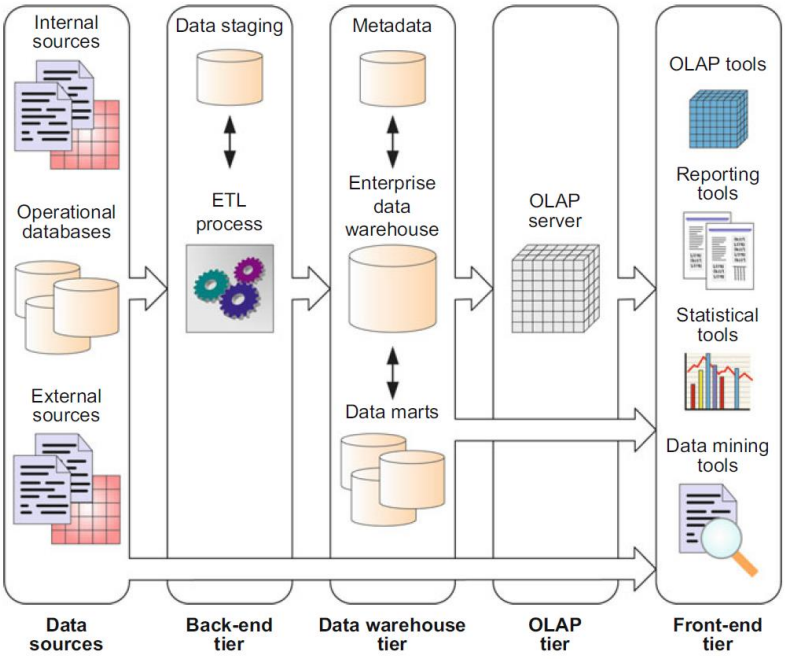


Figure 13: Typical data warehouse architecture

Source; Vaisman and Zimányi, (*p. 77*, 2014)

1.5.2. Hierarchies

Hierarchies are logical structures used to organize data using structured levels. Hierarchies are used to establish family structure; in hierarchies each level is connected to one below and above it logically. These hierarchies can be classified as

Balanced Hierarchies has one path at the scheme level, where all levels are mandatory. All parent members have mandatory one child member and one child member belongs to exactly one parent member.

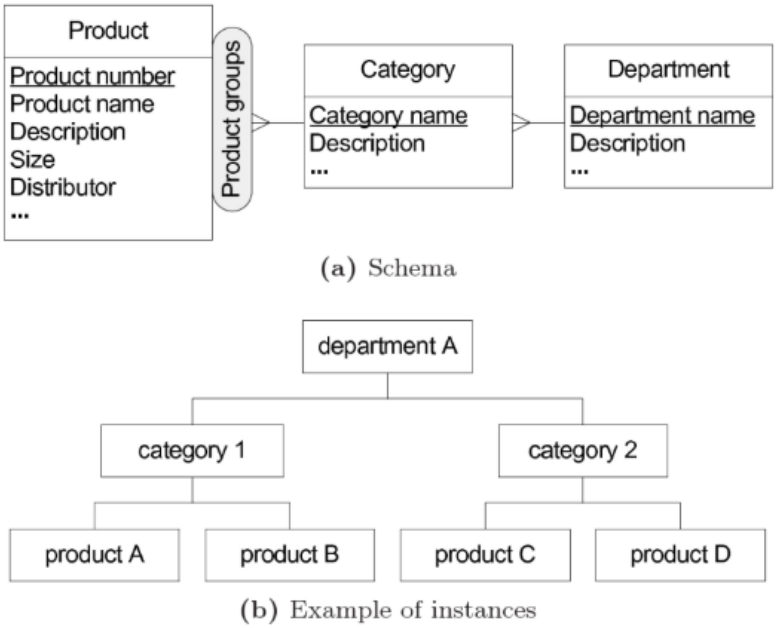


Figure 14: Balanced Hierarchy

Source; Malinowski (*p.* 84, 2008)

Unbalanced Hierarchies has only one level at the scheme level, but unlike balanced hierarchies it is not mandatory to have at least one level. So, there can be a parent level without any child level.

Generalised Hierarchies has members at a same level which belongs to different types. This situation is generally depicted in E-R model using generalized relationship.

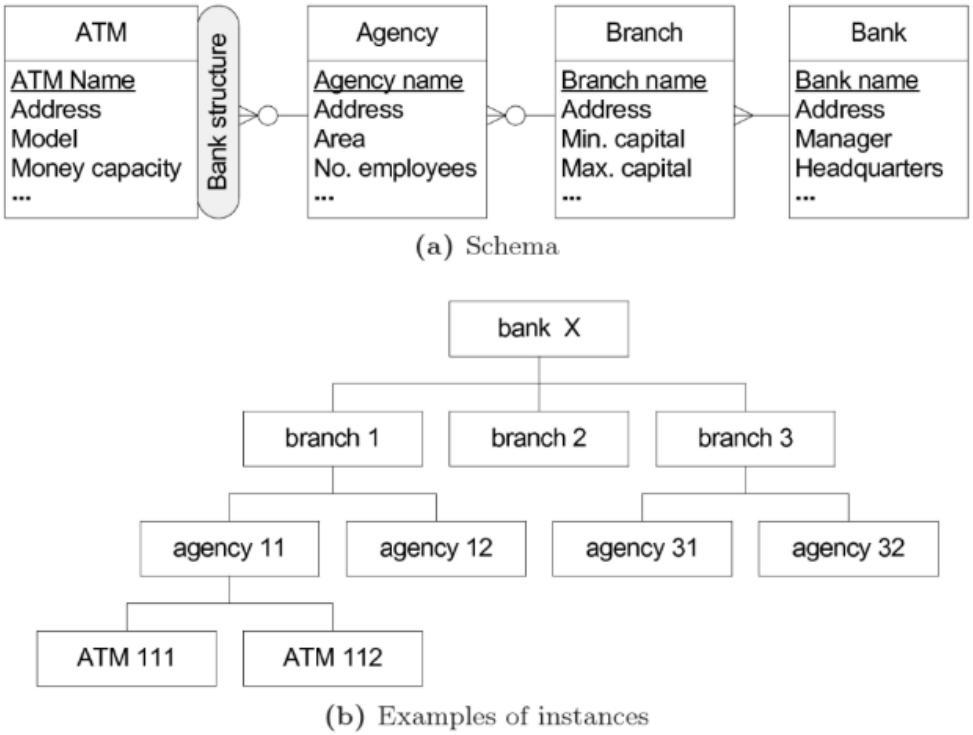


Figure 15: Unbalanced Hierarchy

Source; Malinowski (*p.* 85, 2008)

1.5.3. Schemas

A schema is a collection of database objects like objects, relationships, tables etc. The arrangement of database objects in the schema model can be done in many ways few of which are **star schema**, **snowflake schema**, **star flake schema** or **constellation schema**.

Star schema is the most basic type of the data warehouse schema. It is named star schema because it resembles the shape of star with a fact table in the center and dimension tables radiating from the center. Star schema uses simple queries to keep the response time faster. In this schema the dimension tables are not generally normalized.

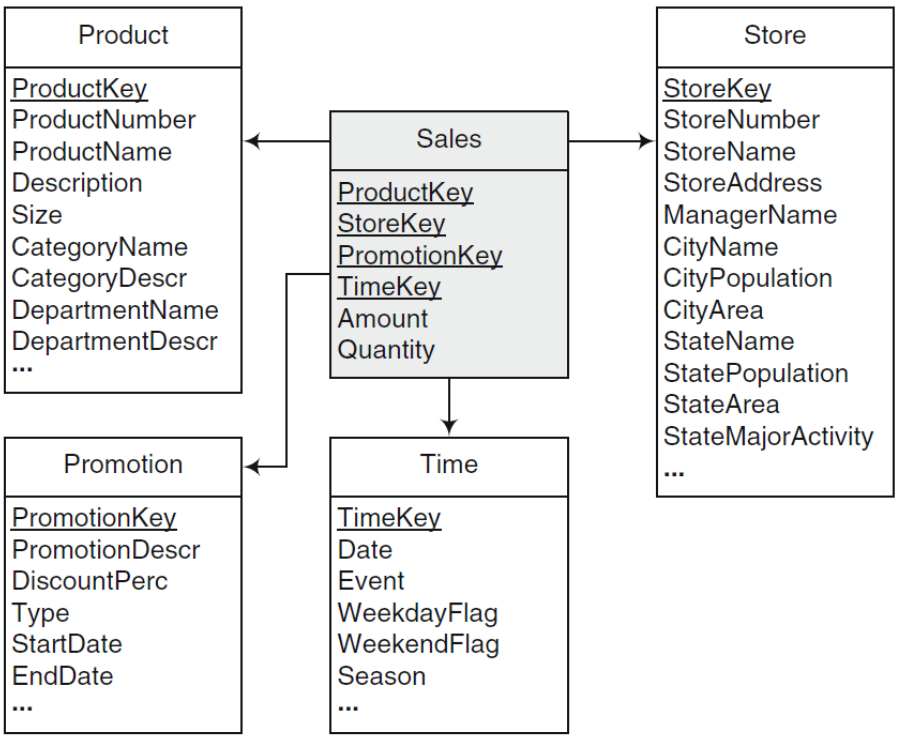


Figure 16: An example of Star Schema

Source; Vaisman and Zimányi, (*p. 123*, 2014)

Snowflake Schema is a modified version of star schema where the dimension tables are normalized too reducing the redundancy of data stored in the data warehouse. Normalized tables save space for storing the data although the response time is a bit slower compared to star schema as the data stored is in a more complex form as compared to the data stored in the star schema.

Star flake schema is a combination of star and snowflake schemas which means some dimensions are normalized whereas other are not. A **constellation schema** has various fact tables that share dimension tables.

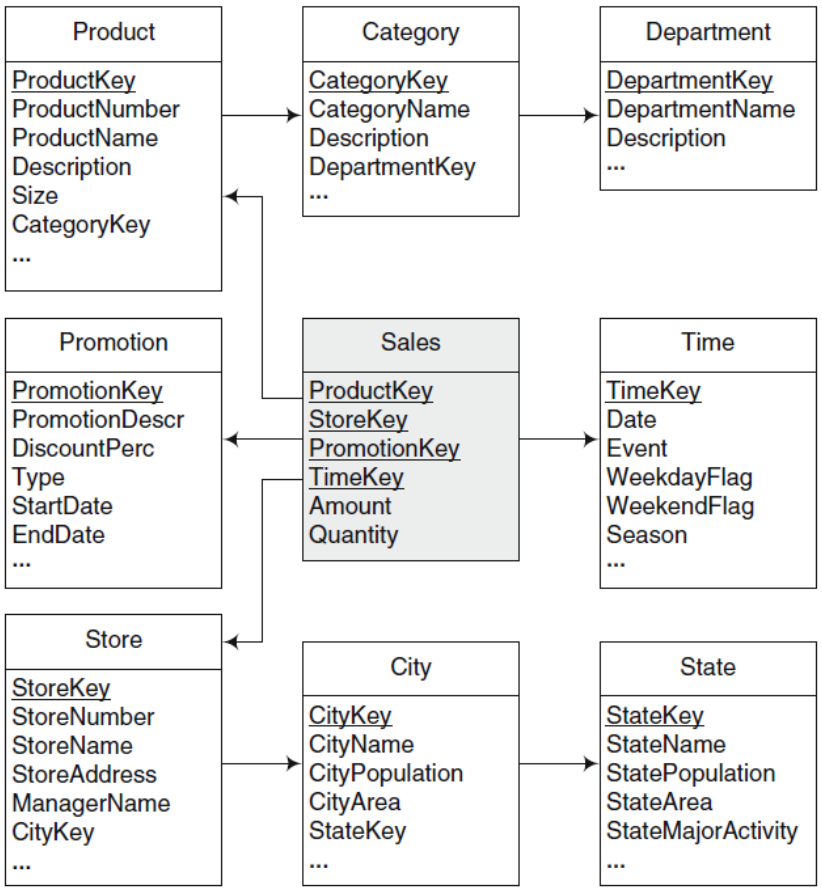


Figure 17: An example of Snowflake Schema

Source; Vaisman and Zimányi, (*p. 124*, 2014)

1.6. ETL Process

Extraction, Transformation and Loading (ETL) processes are used to extract data from the source may it be external or internal and then transforming the data in a standard form and then loading the data into a data warehouse. ETL is a 3-stage process namely,

- Extraction – It refers to gathering data from the multiple sources. These sources can be other databases or may also be files of various formats. It might as well be internal or external.
- Transformation – It refers to modifying the data from the format of original sources to the warehouse format. This step includes various aspects such as *cleaning*, which removes errors and inconsistencies from the data and converts it into a standard format. *Integration*, which collects and merges the data collected from different sources.

Aggregation, which sums up all the data and segregates them based on level of detail, granularity of the database.

- Loading – the data collected and transformed is loaded into the data warehouse in this step. This stage is also linked with updating or refreshing the data warehouse, which means any updating done to the data in the source will also accordingly change the data in the data warehouse. Depending on the policies of organization, this refresh frequency can vary from several months to several times a day.

1.7. OLAP and OLAP Operations

OLAP is called Online Analytical Processing (OLAP) is a process which provides the user with multi-dimensional data from data warehouse and/or data marts. This process consists of OLAP server, which controls the required operation. Most databases nowadays provide an OLAP extension, which allows the users to create and query multi-dimensional cubes. It also makes it possible to navigate, analyze and report about the cubes.

However, there is no standardized language for defining and manipulating the data cubes, and the technology changes with changing systems. Though there are several languages like, XMLA (XML for Analysis) which provides a common ground for exchanging multi-dimensional data between different client applications and OLAP servers. MDX (MultiDimension eXpressions) is another query language used with OLAP databases. SQL standards have also changed to provide analytical capabilities to the users know as SQL/OLAP.

OLAP tools provide a way to interact with the data to manipulate and explore the warehouse data stored. They allow the formulation of complex or higher order queries which may involve large amounts of data. These queries are known as ad hoc queries, as the system has no prior information about them.

Some examples of these tools are –

- Roll-up: Roll-up navigates the data from more detailed to lesser details. The goal of the operation is to reduce the dimensions displayed or to reduce the amount of information available on the system.

- Drill-down: Drill-down navigates the data from less detailed to more detailed. The goal is to descend a concept hierarchy or add more dimensions.

- Slice and Dice: The slice operation focuses on selection of one dimension of the data cube to result in a sub cube. The dice operation creates a sub cube by selecting two or more dimensions.
- Pivot (rotate): Pivot is a visualization operation used to rotate the data axes to provide different presentation of the data cube.

1.8. Tools for supporting Web Services

A Web service is a set of related functionalities that can be programmatically accessed and manipulated through the Web.

There are two types of databases used with web services or web development which are,

- SQL databases
- NoSQL databases

1.8.1. Architecture of Web Database solution

The web-based databases usually have a 3-tier architecture. It consists of a database tier which consists of a database management system which enables storing and management of the data in a database. Over database tier, there exists the middle tier which is the communication link between the database tier and the other tier. The last tier is the client tier which consists of the web browser software which the user uses to access the web database.

A simple representation can be seen in the fig.20.

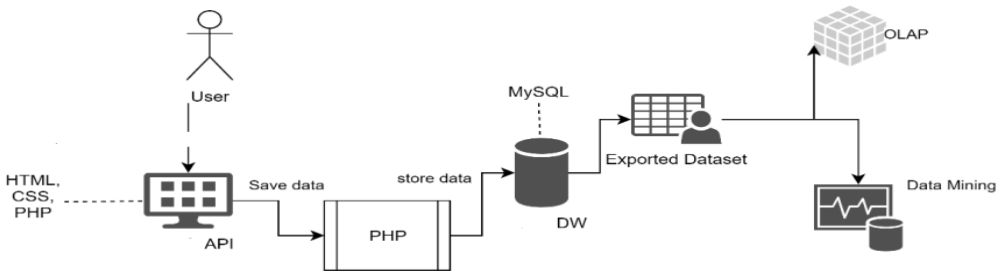


Figure 18: Architecture of a simple web database.

Source; Own Contribution

A user can use an API (Application Programming Interface) to define the interaction between different software and web services. The data is stored into the database by using

any programming language used for web development. PHP being the most common language used for this purpose dude its versatility and flexibility. The data is then stored and managed in the database using any database management language like SQL. This data can then be used for future purpose like data mining or any other OLAP or OLTP operations.

1.8.2. Processes within Organization

- **Order-to-cash**: This process performed by the vendor, starts when the customer submits their order and end with the delivery and payment for the same. The process consists of activities like purchase order verification, shipment, delivery, invoicing, payment receipt and acknowledgement.

- **Quote-to-order**: This type of process generally a prior process to order-to-cash process. The customer requests a Request for quote (RFQ) to the supplier and ends with the supplier providing a quote for the said order. After this point the order-to-cash process is executed.

- **Procure-to-pay**: This process starts when an organization determines a product, or a service is to be purchased. It ends when the product or the service is paid for. It includes processes like obtaining a quote, approving the purchase, selecting a supplier, issuing a purchase order, receiving the goods, and paying the invoice. For each procure-to-pay process there is a consequent quote-to-cash process on the supplier's end.

- **Issue-to-resolution**: This process starts when a customer raises a problem or issue related to the good or service. The process ends until the customer and the supplier agree that the issue is resolved. A variant of this process can be seen in insurance companies that deal with insurance claims. This variant is called claim-to-resolution.

- **Application-to-approval**: This process starts with someone applying for a benefit or privilege and ends with the request being accepted or denied. These processes can be generally found in government agencies.

These processes can be designed and performed by any organization according to their needs which affects the quality of service and the efficiency with which the services are delivered.

1.8.3. Process Categories

- **Core Processes**: These processes cover the essential value creating processes within a company or an organization. These processes may include design and development, marketing sales, delivery etc.
- **Support Processes**: Support processes are the processes which enable the execution of the core processes. Examples of these processes may include finance management, legal services, and others.
- **Management Processes**: These processes set directions, rules, or practices for both the core and support processes. These processes may include strategic planning, budgeting, risk management, etc.

1.8.4. Elements of Business Process

- Events: Events are occurrences within a process, and they trigger the execution of sequence of activities. There is no duration of occurrence.
- Activities: Business processes are composed of set or sequences of activities. They have some duration unlike events.
- Tasks: The selection of activity is composed of a single unit of work which is called task.
- Decision points: These are the points in time when a decision must be made that affects the way a process is executed.
- Actors: Multiple actors are involved in a process such as,
  - Human actors – Customers, users, administrator, etc.
  - Organization and business units – suppliers, institutes, SMEs
  - Software systems – databases, applications, ERP software tools

Actors can be further classified as internal or external actors depending on the place in the organization's system boundary.

- Physical Objects: These can be any item or object involved in the process such as paper or electronic documents, equipment, and many others.
- Outcomes or Process Outputs: Outcome or process output is the result of the executed process. An outcome or output should bring value to the actors involved in the process.

Outcomes can be positive or negative outcome depending on the value the outcome creates for the involved actors.

1.8.5. Business Process Modelling Tools

A business process is a collection of related activities which result in a specific outcome. To design and depict these activities, models are developed. The model consists of a static view (Entity relationship diagram), interaction of involved actors (use case model) and behavioral view (BPMN, Object relation diagram).

Static View or static model is the representation of the static aspect of a process involved in the system. This is depicted using class diagram or UML (Unified Modeling Language) or Entity-Relationship Diagram can also be used. Class diagrams is based on the concept of classes, generalization, and association. Whereas in an ER diagram entity, entity types, attributes, relationship types are depicted. The difference between ER and Class diagrams is that ER diagrams represents the abstract model whereas class diagrams demonstrates the behavior of the system.

Activity diagrams are diagrams used to describe the detailed sequence of tasks and decisions which are involved in a single activity. These diagrams depict the flow of the work or the process and to note the results of the decisions in a process.

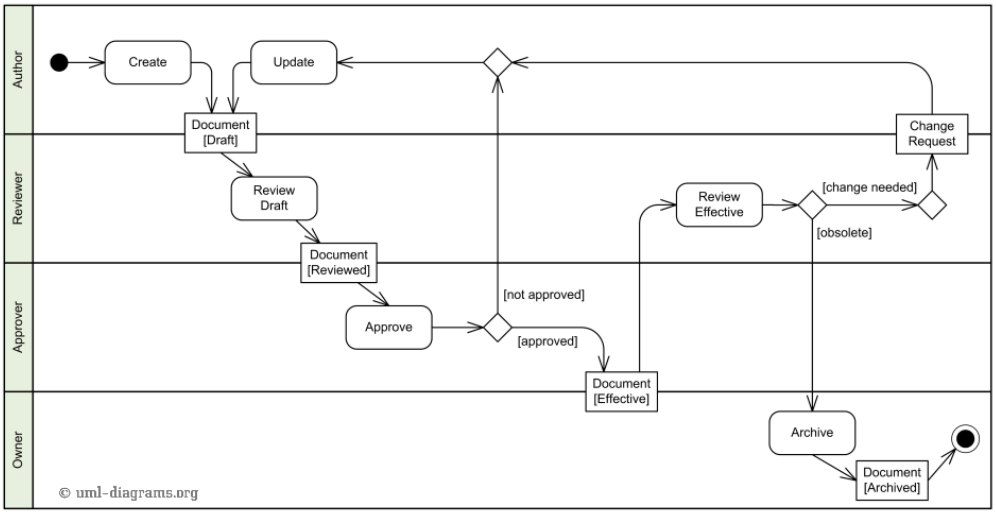
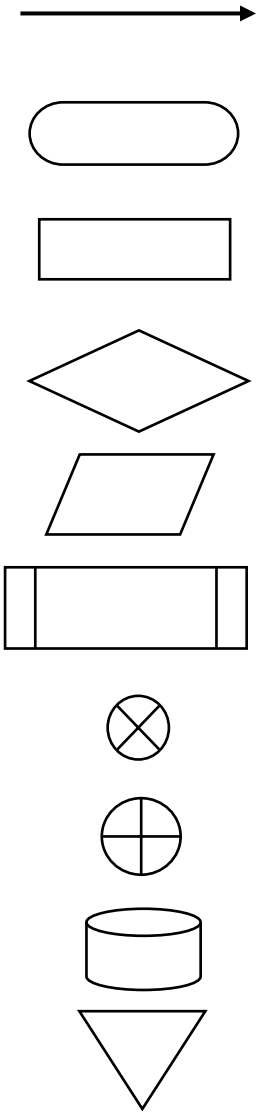


Figure 19: An example of activity diagram

Source: www.uml-diagrams.org

The activities can also be described by many other charts or diagrams such as flowcharts, BPMN etc. Flowcharts were chosen to describe the web activities due to their simplicity and

easy to understand. These flowcharts can easily describe an activity and are easily understood by anyone. These flowcharts consist of many elements to describe an activity, they consist of different shapes and lines to demonstrate the flow of work or an activity. Some of the entities used while constructing a flowchart are-



Flowline – shows the order of process.

Terminal – shows the ending and beginning of a process

Process – shows operations that change values

Decision – indicates a conditional operation to decide which path the operation will take

Input/output – indicates the need to input or get output

Predefined Process – shows processes which are predefined somewhere else.

Summing operator – depicts adding of two or more processes.

Or operator – shows choice between two or more operators

Database – represents database or data file.

Merge – indicates merging of multiple data into one.

These symbols are the most used entities used to define a process in flowcharts, there are other symbols too to depict conditional processes. The fig.24 shows a basic example of using a flowchart to describe a process.

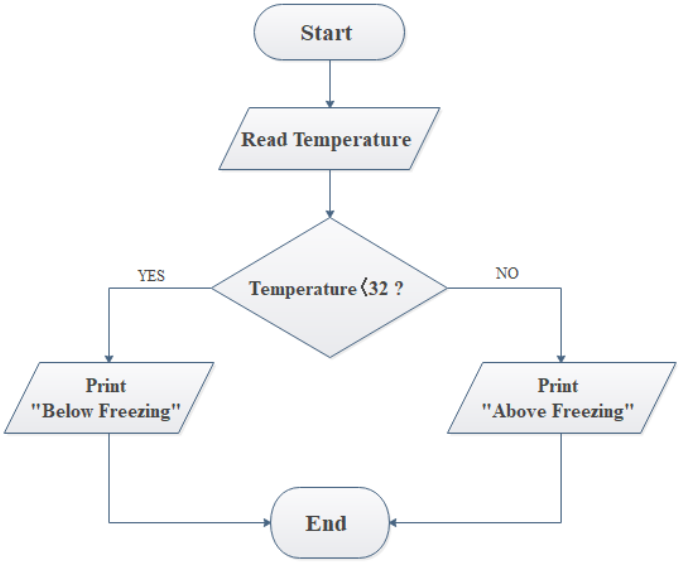


Figure 20:An example of Flowchart

Source: https://www.edrawsoft.com/flowchart-definition.html

Another most used model for representing data models is BPMN (Business Process Model and Notation). It is like flowcharts as it also uses symbols and arrows to demonstrate the flow of an activity or a process. However, BPMN is much more detailed demonstration of the processes and can display multiple processes as a part of one major process. It can also easily show the actors involved in the process and thus, is also preferred as it can show more information when compared to other models. Though BPMN can integrate much more information in the model, it requires the users to have understanding about the model and notations used, which can make it sometimes complex for the users to understand.

The elements involved in a BPMN model can be classified between 4 major subgroups which are, flow objects, connecting objects, swim lanes and artifacts.

(1) Flow Objects

(a) Events – are triggers used to demonstrate start, end, change or completion of a process.

These event symbols can also be modified to represent specific details in a process. For example,

Message symbol – triggers a message or process or finishes the process.

Timer symbol – a time or date or recurring date triggers a process or ends a process.

Error symbol – error occurs at the start, middle or end of the process.

(b) Activities – are tasks or process done by an individual or a system.

Task symbol – shows the most basic of the activities.

Sub-group symbol – shows a group of tasks that fits well with each other.

Transactional symbol – special sub-process that involves payment.

(c) Gateways – are decision points used to define a condition or event in the process path.

Exclusive symbol – based on condition breaks the process into one or more mutually exclusive paths.

Event-based symbol – based on occurrence of an event breaks the process into other paths.

Parallel symbol – used to show two concurrent processes in a business flow.

(2) Connecting Objects

(a) Message Flow – are used to define the path of message in the process.

(b) Sequence Flow – are used to define the sequence of process.

(c) Association – is used to define a relationship or association between two or more entities

(3) Swim Lanes

(a) Pool – shows the major participants in a process.

(b) Lane – shows the activities and flow of specific process of all participants.

(4) Artifacts

(a) Data Input

(b) Annotation

(c) Data Storage

(d) Data Output

1.9. Data Mining

In the recent years, with the increasing development in the field of information technology and data handling, data mining has attracted a great deal of attraction. Mining is the process of finding small pieces of valuable information from huge or large amount of data. According to Hand, Mannila and Smyth, *"Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways are both understandable and useful to the data owner."* (HAND, 2001, s. 1) The information such as relationships between data or summaries of data is referred as models or patterns. Some researchers consider data mining same as Knowledge Discovery from Data (KDD), while others consider data mining as an important step in the KDD process.

While considering database management or data warehousing, data mining can be viewed as an integral part of OLAP. However, it is important to have a large data set for a process to be considered as data mining as any analytical process performed upon smaller data volumes would be a statistical data analysis tool. Data mining processes can be classified as descriptive and predictive. Descriptive process deals with characterizing the properties of data whereas predictive process deals with making conclusions on the present data to make predictions about the future data. The two forms of analysis used to extract data models based on existing data are Classification and Prediction. Classification is used for predicting categorial data (discrete, unordered) whereas prediction is used for continuous value functions. Some of the methods used for classification are building decision tree classifiers, Bayesian classifiers, case-based reasoning, genetic algorithms etc. The most common method used for prediction is regression model or regression analysis. (Han, 2006, s. 24)

The data classification process is a two-step process. These processes include, learning where the classification algorithm builds a classifier by analyzing the training set of data from the database, i.e., the classifier is made more accurate for the data set. Once the accuracy of the classifier reaches an acceptable accuracy it is then applied for the new data in the database which is the next step in the process and is called classification. Similarly, data prediction is also a twostep process the only difference between the two considered is that, the data considered for prediction is continuous or ordered. (Han, 2006, s. 286)

2. Company Overview

The SME under the study is a restaurant in Liberec. It is a small-scale business that has its own website which consists of not much online operations. The restaurant specializes in Czech as well as international food cuisine.

At present, the restaurant has minimum online business activity. The website just shows the menu, opening hours and contact details. There is no service for ordering food online or to make reservations online. With the changing lifestyle and the approach towards business, it is best for the restaurant to establish a better website and to increase their online presence. The modifications suggested in this project will allow the restaurant to take online order or to make reservations for dining in. The restaurant will also be able to keep track of all the orders and the amount of profits on sales. The owners can also check the offers validity dates and times and can also check if the menu needs a change of dishes or some other products.

The objective of this project is to design a database and an infrastructure to support online activities of the restaurant using business intelligence tools. These features can be used to increase the business opportunities and maybe open new business opportunities for the restaurant.

2.1. Pilot Data

The data used in the database is a dataset created to test the features of the database. The data used in the database is not based on realistic facts, however it has been created for the need of the current thesis. The dataset reflects an example of the data stored and used by a small business such as a restaurant and creates a scenario where the tools mentioned in the thesis can be applied.

3. Data Warehouse Model

3.1. Conceptual data model

The data model of the database was created as conceptual, logical, and physical models. The conceptual data model represented in the entity relationship diagram as shown in the fig.27 depicts the fact table and the dimension tables used to store and compute the data. The model was created as top-down design as the model was built first completely depending upon the user's requirements and then different views can be created for different users at different level, such as the owner can check the overall operations of the restaurant whereas the front-line staff can only check the information related to other rather than having access to everything. The dimension tables were determined by the information required to store in the database. The conceptual model depicts a snowflake schema with an unbalanced hierarchy. This model can be further simplified, and it resembles star schema (fig.28). The model contains one fact table which is surrounded by other dimension tables. Most of the dimensions are connected by one-to-many relation with the fact table. The main entity types in the model are the fact table which is orders table and other dimension tables like region, reservation, customer, food items. Other dimension tables were specified to normalize the data in the tables.

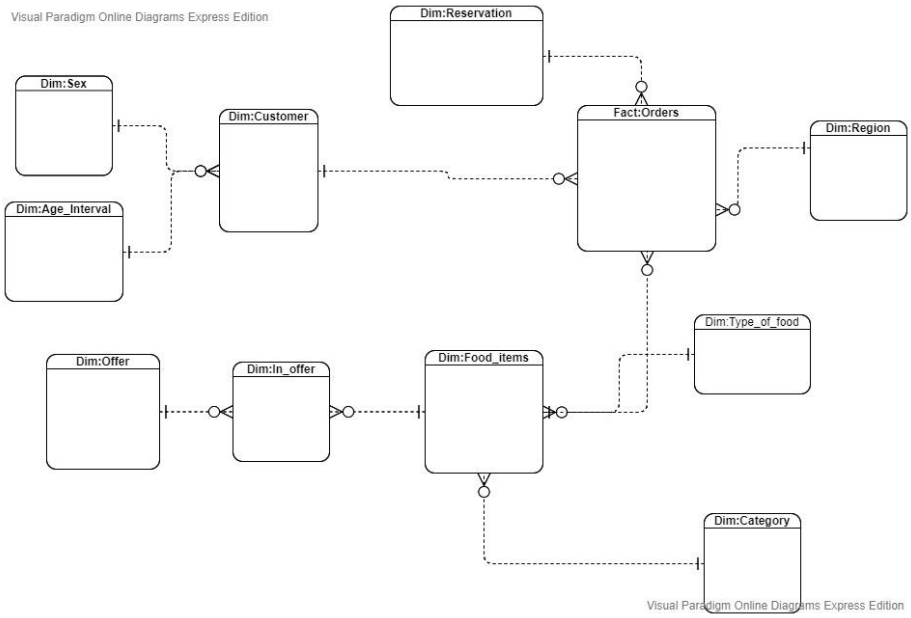


Figure 21:Conceptual Data Model of the proposed ordering system

Source: Own Contribution

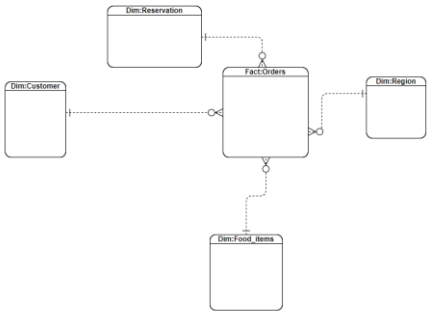


Figure 22: Simplified model as star schema

Source: Own Contribution

The simplified model depicts the conceptual model as star schema. The other dimension tables are merged into the shown tables. This model is not normalised and thus is not efficient in handling large volumes of data. Therefore, after normalizing the data to be stored, the model can be seen in the fig.27.

3.2. Logical data model

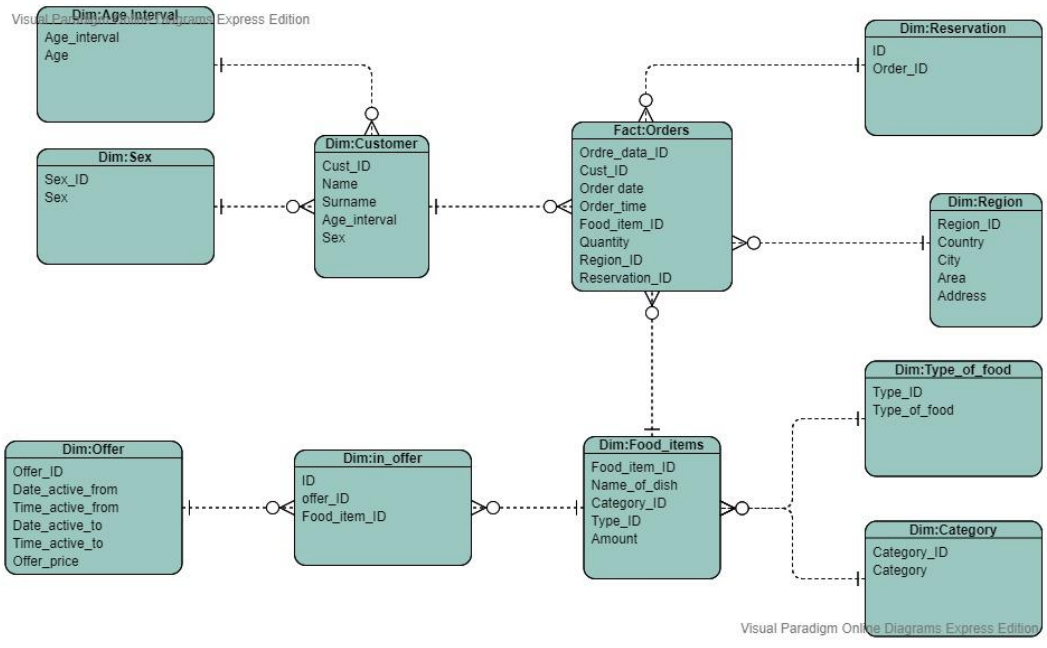


Figure 23:Logical Data Model

Source: Own Contribution

The conceptual model once prepared and finalized; it was then converted to logical model while keeping other restraints in mind like referential integrity and not linking two fact tables to each other. Logical model is like conceptual model but with more complexities. The logical model shows the primary keys and foreign keys in the database. It also shows the columns in the database. The final model created had food_items, customer, reservation, and region tables connected to one fact table named orders. The other tables like sex and age_interval was related by one-to-many relationship to the dimension customer as there can be many customers that belong to same age intervals or same sex groups. Similarly, other tables like type of food and category are also related to the food_items table by one-to-many relationship depicting the relationships between these entities same as the customer table. Another dimension table, time was created before however due to complications with the relationship and calculations, it was later merged into the fact table (orders) rather than allotting it an individual dimension.

The entities food_items, customer, region, reservation have different levels of granularity and follows different hierarchies, thus it was established that the database follows an unbalanced hierarchy to store the data.

The dimensions were connected to the fact table by foreign keys which link the data between the tables. These dimensions store the data about different aspects of business such as the customer table that stores the data about the customers that have logged in before and this can be later used to analyze sales per customer or any other prospect. Another such dimension is Food_items, which stores information about the different types of food offered by the restaurant. This entity is also linked to the offers table through one-to-many relationship, which means that different food items can have different offers connected to them depending on the requirement of the restaurant. Like the customer table the data from this dimension can be used to check which item has an offer applied to it and how can the restaurant change their pricing model to increase the profits. The dimensions were designed so that different aspects of the business can be analyzed at different granularity. The data from all the tables is linked to the fact table by foreign keys so that it can be used by the user anyway they require. These foreign keys link the data using the primary keys in the dimension, making sure that the referential integrity is maintained in the database.

The dimension Food_items can be seen forming its own star schema, which was first designed as a fact table in its own. But, due to designing constraints and the fact that two or more fact tables should not be linked together as it may become too complex or may show errors while performing certain operations, due to these factors decision to keep it as a dimension table was made.

The primary keys were assigned depending on the type of data to be stored. Generally, the primary keys were set to be varchar to make it easier for the user to store the data and to make it less confusing for the user to read the data. The primary keys be bold columns in the figure and the foreign keys are depicted as italics text. The entity food_items also forms a snowflake schema with the other dimensions like category and food_type.

3.3. Physical data model

The Logical model is further developed into a physical model. The physical model is the representation of the database. Physical data model is the represents how the data is structured and related in the database, thus making sure all the conventions and restrictions of developing a relational database are considered. After the logical model, the data types for each column was also decided and assigned to different columns. For example, some of the primary key columns like order_data_ID were set to auto number by the software itself whereas some columns were made to user-defined like cust_ID was defined as varchar.

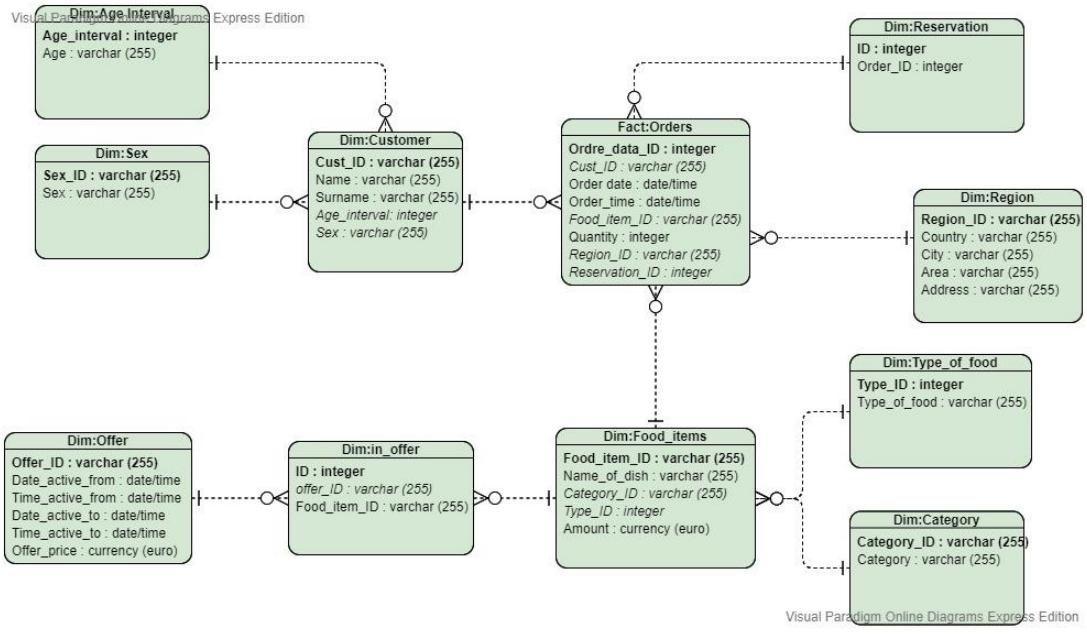


Figure 24:Physical Data Model

Source: Own Contribution

The main objective of modelling process is to make sure that the data objects are represented accurately. The modelling process starts with designing the conceptual model and then modifying the model to get a logical model. As more and more information is added to the model, we can develop a physical model which is the representation of the relational database. Using the physical model, the database can be designed in any DBMS as according to the user's requirement. The physical model was then converted into a multi-dimensional relational database using MS Access. The choice for this DBMS was made due to availability of the software as a part of MS Office package and the ease of access for the user and the developer to create a database and to execute commands as well. Other DBMS like MySQL were also considered however due to the lack of availability and the complex nature of using commands made me decide to use MS Access rather than MySQL or any other DBMS.

4. Queries Performed

4.1. Query for billing summary of a customer

This query was created to show the order summaries of different customers at the restaurant so that the stored data can be later analyzed to find useful information for promotion of business for the restaurant.

Name	Surname	Name_of_Dish	Amount	Quantity	Total	Offer Price	Order_Time
Peter	Parker	French Fries	€ 1.15	1	€ 0.65	€ 0.50	13:00:00
John	Smith	Grilled cup with ice cream and whipped cream	€ 2.10	2	€ 4.20	€ 0.00	16:30:00
Peter	Parker	Croquettes	€ 1.25	1	€ 1.25	€ 0.00	20:15:00
Jana	Dedkova	Chicken fillets, lemon	€ 4.90	1	€ 4.90	€ 0.00	21:15:00
James	Bond	Gordon Bleu, lemon	€ 5.00	1	€ 5.00	€ 0.00	17:55:00
Peter	Parker	Becherovka	€ 1.25	2	€ 2.50	€ 0.00	18:47:00
Jane	Dones	French Fries	€ 1.15	1	€ 1.15	€ 0.00	12:00:00

Figure 25: Billing Summary

Source: Own Contribution

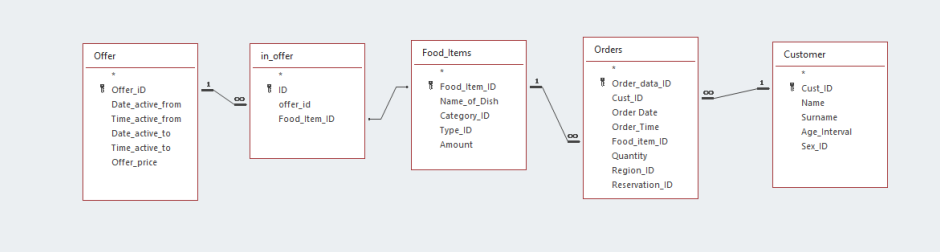


Figure 26: Design view of billing summary

Source: Own Contribution

To obtain the required results in this query expression builder was also used. Conditional statement is used to apply discount to the customer's order depending upon the discount conditions specified by the owners. To apply this condition, If statement was used to ensure that the order date and time is between the specified time and date. This statement checks if the value lies between the conditions specified and returns true or false results depending upon the value entered. For example, the if statement used in this query returns the offer amount if the order date and time is between the mentioned values otherwise it returns 0. Another area where the expression builder was used is for the Total column, where the amount is displayed as the product of quantity and amount reduced by the applicable discount.

```
SELECT Customer.Name, Customer.Surname,
Food_Items.Name_of_Dish, Food_Items.Amount, Orders.Quantity,

IIf([Orders].[Order_Time]>=[Offer].[Time_active_from] And
[Orders].[Order_Time]<=[Offer].[Time_active_to] And
[Orders].[Order Date]>=[Offer].[Date_active_from] And
[Orders].[Order
Date]<=[Offer].[Date_active_to],[Offer].[Offer_price],0) AS
[Offer Price], [Quantity]*[Amount]-[Offer Price] AS Total,
Orders.Order_Time

FROM Offer INNER JOIN ((Food_Items INNER JOIN in_offer ON
Food_Items.Food_Item_ID = in_offer.[Food_Item_ID]) INNER JOIN
(Customer INNER JOIN Orders ON Customer.Cust_ID =
Orders.Cust_ID) ON Food_Items.Food_Item_ID =
Orders.Food_item_ID) ON Offer.Offer_iD = in_offer.offer_id;
```

Figure 27:SQL code for above mentioned query

Source: Own contribution

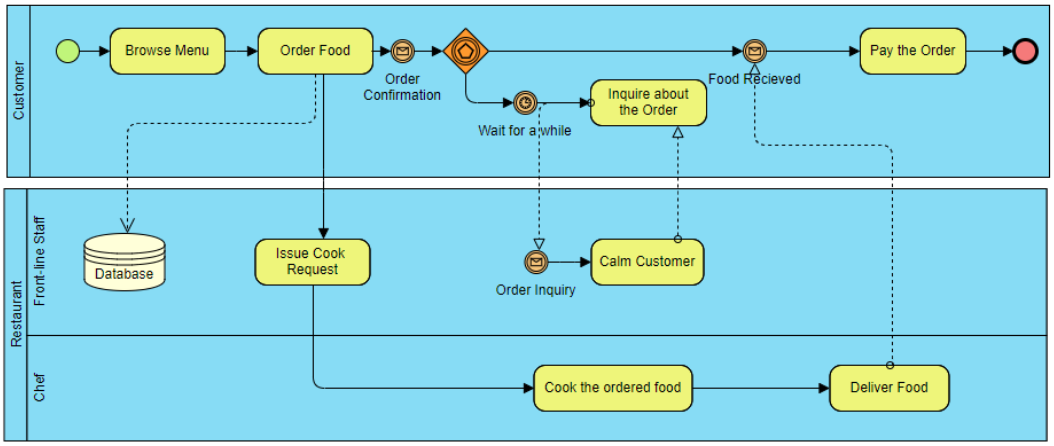


Figure 28:Activity Diagram for online ordering system

Source: Own Contribution

The process of online ordering starts with the customer seeing the online menu and choosing the food they would like to order. This order is then forwarded to the restaurant which confirms the order to the customer. After the order has been confirmed by the restaurant the customer has to wait for the order. The customer may inquire about the order which is then handled by the front-line staff. Once the order is prepared it can be delivered to the customer which is confirmed by the delivery person. The payment can be chosen from multiple methods such as cash on delivery, card on delivery or online payment.

The order is stored in the database created for the restaurant. This data can be later used by the owner to analyse the sales in the restaurant. This information can be used for multiple strategies to increase the sales. For example, this data can be used to create

discounts specific to the customers such as if some food item is ordered more number of times by a single customer, offering them a discount on his next order or suggesting other food items depending on the past orders. This data can also be used to suggest food items based on their previous orders like showing the most repeated order by the same user as recommended to the user. The information can also be used to save more personal data such as address or phone number so that the next time a customer order's, their information is pre-defined rather than then asking the customer to fill their information again. Other means of promotions or advertising can be, providing a recurring customer a coupon for free drinks with certain food. If certain customer has been ordering same type of food on a regular basis, they can be also advertised for other dishes by providing them discounts or free coupons on other types of food items, so that they may try other items as well and may change their ordering patterns if they find something interesting.

4.2. Query for menu item summary

Name_of_Dish	SumOfTotal	SumOfQuan
Becherovka	€ 2.50	2
Chicken fillets, lemon	€ 4.90	1
Croquettes	€ 1.25	1
French Fries	€ 1.80	2
Gordon Bleu, lemon	€ 5.00	1
Grilled cup with ice cream and whipped cream	€ 4.20	2

Figure 29: Menu items summary

Source: Own Contribution

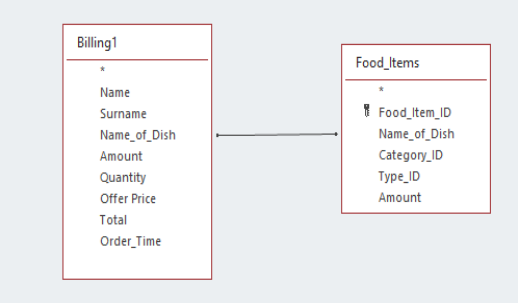


Figure 30:Design view for menu item summary

Source: Own Contribution

```
SELECT Food_Items.Name_of_Dish, Sum(Billing1.Total) AS SumOfTotal,
Sum(Billing1.Quantity) AS SumOfQuantity

FROM Food_Items INNER JOIN Billing1 ON Food_Items.Name_of_Dish =
Billing1.Name_of_Dish

GROUP BY Food_Items.Name_of_Dish;
```

Figure 31:SQL code for the above-mentioned query

Source: Own Contribution

To get the required results from this query, column totals were added into the query results. This button allows the user to display the total of each column in the specified row. The sum operation was selected to display the data. This operator groups the data by the sum of the column for example sum of total and sum of quantity.

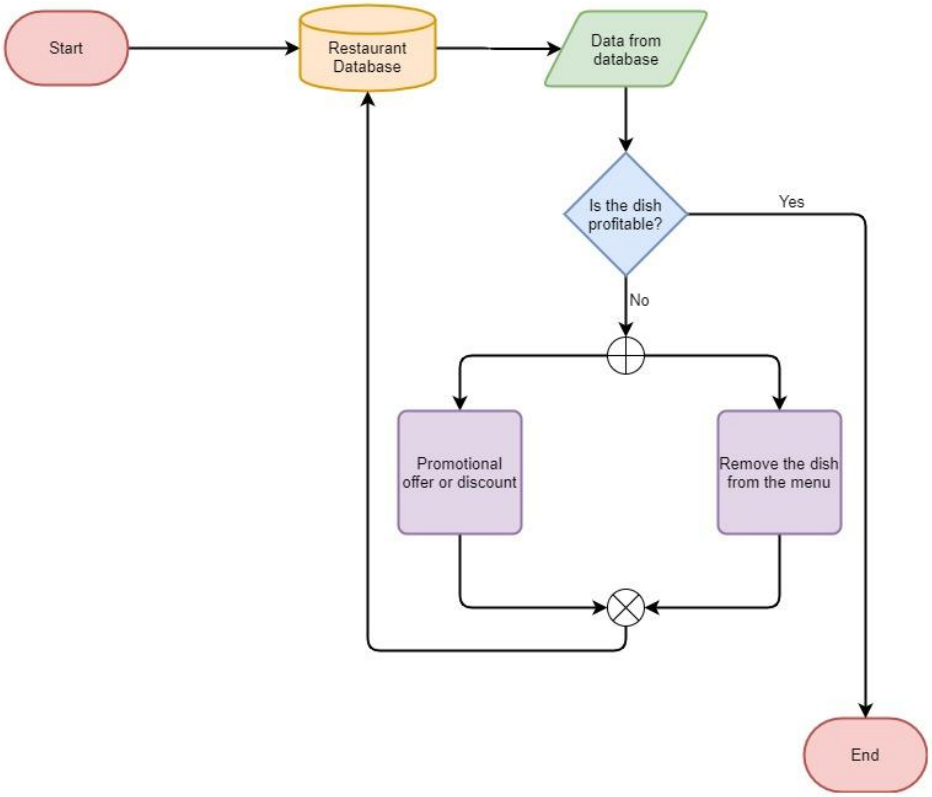


Figure 32:Flowchart displaying the flow of process

Source: Own Contribution

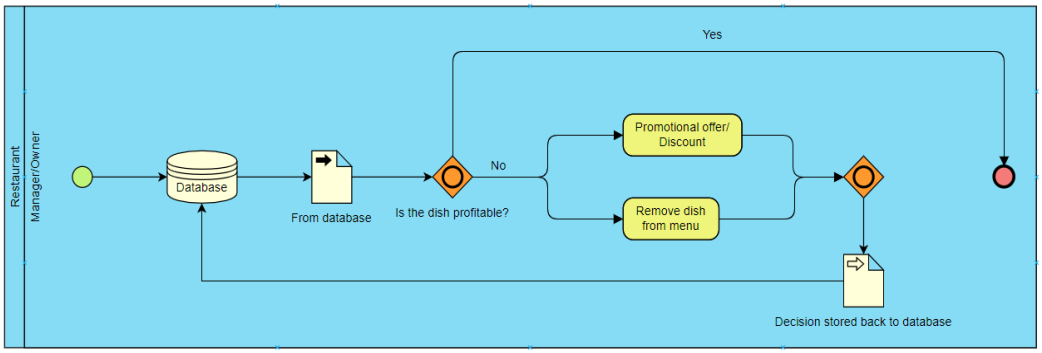


Figure 33:BPMN for optimising the menu

Source: Own Contribution

This query shows the method through which the restaurant menu can be optimised to increase the amount of profit the restaurant makes. There may be some food items which are not ordered too many times by customers, these items can be removed from the menu or the owners may decide to offer discounts or promotional offers for these items.

The data is stored in the database already depending on the orders made by the customers. This data can be then retrieved and can be analysed to calculate the sales per item. This can help owners to decide to either remove the dish from the menu or create a promotional offer for it. The owner of the restaurant can also decide to change the price of the items depending on the sales. Once the decisions are made about the products, it can be again updated in the database, reflecting the changes made to the products. The discounts can be similar to age based discounts, for example providing students with certain discount on finger food or drinks, or it can also be discounts depending upon certain products depending upon the time of the day or year, such as in summers cold drinks are preferably ordered more than the hot ones, therefore giving customers discount on certain hot beverages at specific time of the day.

4.3. Query for showing the reservation of tables

Name	Surname	Name_of_Dish	Order_data_ID
Peter	Parker	French Fries	1
John	Smith	Grilled cup with ice cream and whipped cream	2
Peter	Parker	Croquettes	3
Jana	Dedkova	Chicken fillets, lemon	4
James	Bond	Gordon Bleu, lemon	5
Peter	Parker	Becherovka	6
Jane	Dones	French Fries	7
			(New)

Figure 34:Reservation System

Source: Own Contribution

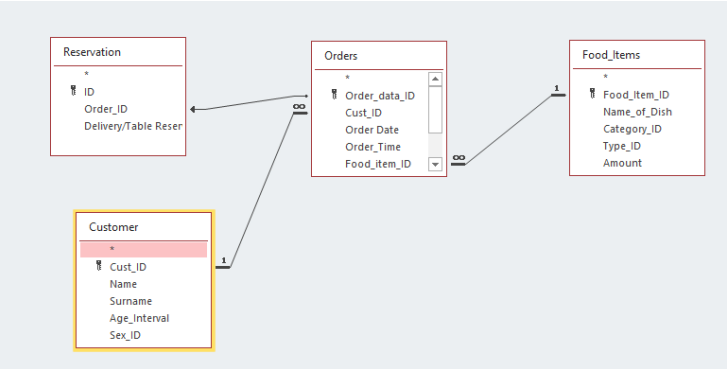


Figure 35:Design view for reservation system

Source: Own Contribution

```
SELECT Customer.Name, Customer.Surname, Food_Items.Name_of_Dish, Orders.Order_data_ID

FROM Food_Items INNER JOIN (Customer INNER JOIN (Reservation RIGHT JOIN Orders ON
Reservation.Order_ID = Orders.Order_data_ID) ON Customer.Cust_ID = Orders.Cust_ID) ON
Food_Items.Food_Item_ID = Orders.Food_item_ID;
```

Figure 36:SQL code for the above-mentioned query

Source: Own Contribution

For displaying the results in this query, different entities were to be selected from different tables. As all the data is associated to each other through some relationship, other facts can be deduced from the table, such as the order number displayed can be checked in the database and the specific items ordered can be seen including the amount paid by the customer and if there were discounts applied to the order can also be checked.

This query shows the process of booking tables at the restaurant. The customer can choose the option to either dine-in or delivery or take away. If the customer chooses to book a table, they will be able to choose the timing of their arrival at the restaurant. When the customer arrives at the restaurant, the front-line staff can check if they have booked seating or not, if they have booked a table, they will be assigned to a table otherwise they will be given the available seat. If there is no empty table, they can be asked to wait for a space to get free. If the customer chooses to pick up their order, they will have to specify their time of arrival, so that the chef can prepare food before their arrival to reduce the waiting time for the customers. Once the customer arrives at the restaurant, they can just show the receipt for the order at the restaurant and can collect their order.

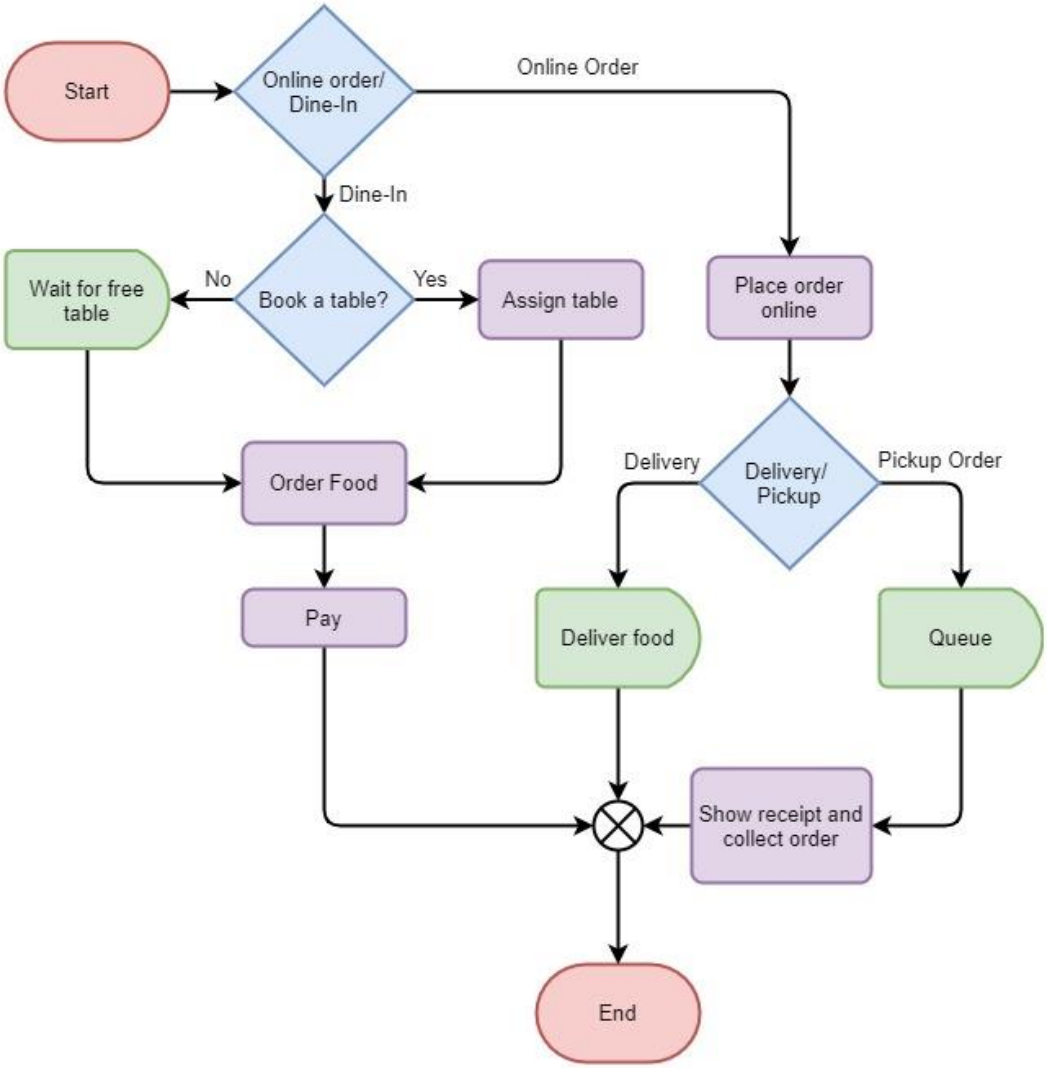


Figure 37:Flowchart to show the flow of process for reservation system

Source: Own Contribution

This data can also be used to promote specific promotions. Like if the restaurant decides to promote people to dine-in more rather than delivery, they can offer free drinks or coupons for some food items to the customers who have been ordering food to be delivered, so that customers can be persuaded to dine-in at the restaurant instead of ordering deliveries. Similarly, the restaurant can also decide to persuade people to order online for delivery in some conditions like recently it was seen in the times of COVID-19. The restaurant can also promote booking the tables rather than people coming directly to the restaurant, by providing some offers to people who book their seats beforehand. The reservation system makes it easier for the restaurant to manage their seating space and their time. This also makes the dining experience of the customers much pleasant as they would not have to wait in queue if the seating is not free. The customers can also

cancel their booking if they want to which will be directly updated in the database in real-time. The restaurant for this matter can keep a time limit, that the customers will have to be at the restaurant in half an hour of their booking or else the booking may cancel if the restaurant is crowded and needs more seating space for customers who are already in a queue waiting for seating space. The restaurant can decide from several options to promote specific method between, dine-in, delivery or for take away. While selecting take away, customers will be able to pay at the restaurant when they collect their order.

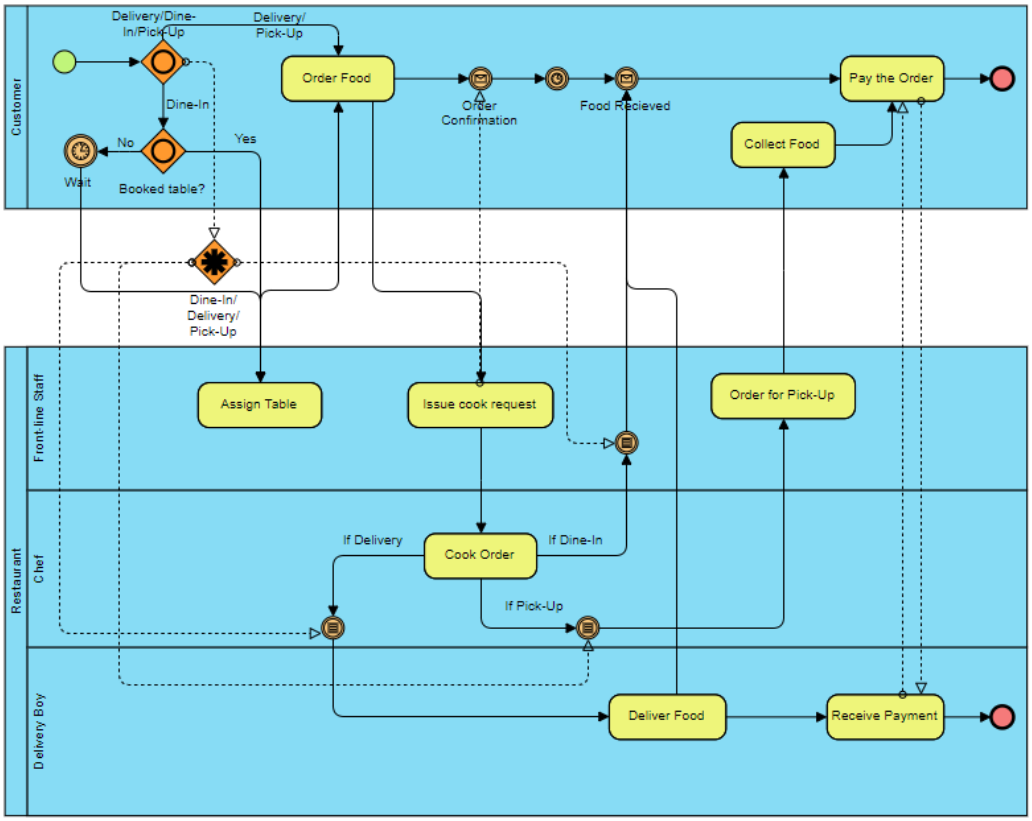


Figure 38:BPMN for reservation system

Source: Own Contribution

5. Discussion of Results

This chapter highlights the result of the thesis and summarizes the work displayed in the study. There were some limitations to the work displayed in the thesis which can be achieved with further research around the subject.

Google Analytics, which is a prominent software, (Google Analytics Usage Statistics) used for analysis of web-based activities is very commonly used by the small and medium enterprises. Due to ease of access and cheaper solution for analysis of online data. It contains metrics and different analytical features like conversion rate, bounce rate etc. which help the owners to make their online processes more efficient and robust. Upon further work in the field of business intelligence, the database and the data warehouses can be designed in such a manner to support these features. As the business grows and they have to deal with larger volumes of data, it would be vital for the company to integrate business intelligence tools in their activities, so rather than developing a totally new data warehouse for the higher volumes of data, the companies can expand their current data warehouse features to support larger operations. This way the enterprise will not have to transfer their data from one place to another and would also avoid the time and money wasted upon creating a new database.

The classification and prediction tools can be used to predict the future trends with the use of regression analysis, so that predictions can be made about the future and decisions can be made depending on those predictions. These decisions might give the enterprise a competitive edge over the competitors. However, for the application of these data mining tools a much larger volume of data needs to be stored and handled, which might prove to be a difficult task, which can be further researched and a data warehouse can be created which is capable of handling such large volumes of data.

The reason of not applying data mining now is that it was not inside the scope of your thesis due to limited data. Data mining requires large data volumes. Moreover, more features are required for accurate prediction patterns. Due to the present scenario, of the pandemic the world is facing it was difficult for us to contact the owners of the restaurant and get access to their data. Due to this situation the work had to be based on non-realistic data rather than a real one, which can be resolved when everything is back to normal and further work can be done to obtain the real data and base the work around it. Also, the

processes elaborated in the thesis are completely new to the restaurant's infrastructure, thus the restaurant has not stored any kind of similar data in the past which makes it unfeasible to get access to real time data from the restaurant. Thus, to demonstrate the feasibility of the exemplified processes, unrealistic data had to be relied upon.

Another issue resolved by the engaging these tools was the ability of the enterprise to be able to store large volumes of data and the ability to characterize the data into different tables according to their characteristics. This data can be easily retrieved to be analyzed by the owners using OLAP operations.

This study utilized many other tools like flowcharts, BPMN, conceptual, logical, and physical models, E-R diagrams to demonstrate different activities involved in the business process and depict the model of the data warehouse. The flowcharts and BPMN diagrams were used to demonstrate the flow of processes in the business. BPMN diagram was also used in combination with flowcharts to display more information about the process such as the involved actors in the process and the flow of messages in the system.

5.1. OLAP Analysis

OLAP analysis was also carried out for the queries mentioned in the last section of the thesis. The analysis was carried out using MS Excel. The data was imported from the Access file. Once the data and the queries were selected different tools from the software were used to analyze the data.

The one of the most useful features used was the Pivot tables. Once the data was imported to the spreadsheet, insert command was used to insert pivot table. Once the data was arranged in the table different entities were selected to display the information. The table created using Pivot table can be seen in the table 1. The table displays the customer's name and the ordered item. It also displays the total quantity of the items ordered and other aggregate data. Based on this data a pie chart was also produced. This table can also be summarized by reducing the amount of data selected which results in a table shown in table 2 and the fig.43 displaying the corresponding data.

Table 1: Pivot Table showing order summary

Row Labels	Sum of Total	Sum of Quantity	Sum of Amount
Jana	**4.9**	**1**	**4.9**
Chicken fillets, lemon	**4.9**	**1**	**4.9**
4.9	4.9	1	4.9
Jane	**1.15**	**1**	**1.15**
French Fries	**1.15**	**1**	**1.15**
1.15	1.15	1	1.15
John	**4.2**	**2**	**2.1**
Grilled cup with ice cream and whipped cream	**4.2**	**2**	**2.1**
4.2	4.2	2	2.1
Peter	**4.4**	**4**	**3.65**
Becherovka	**2.5**	**2**	**1.25**
2.5	2.5	2	1.25
Croquettes	**1.25**	**1**	**1.25**
1.25	1.25	1	1.25
French Fries	**0.65**	**1**	**1.15**
0.65	0.65	1	1.15
Grand Total	**14.65**	**8**	**11.8**

Source: Own Contribution

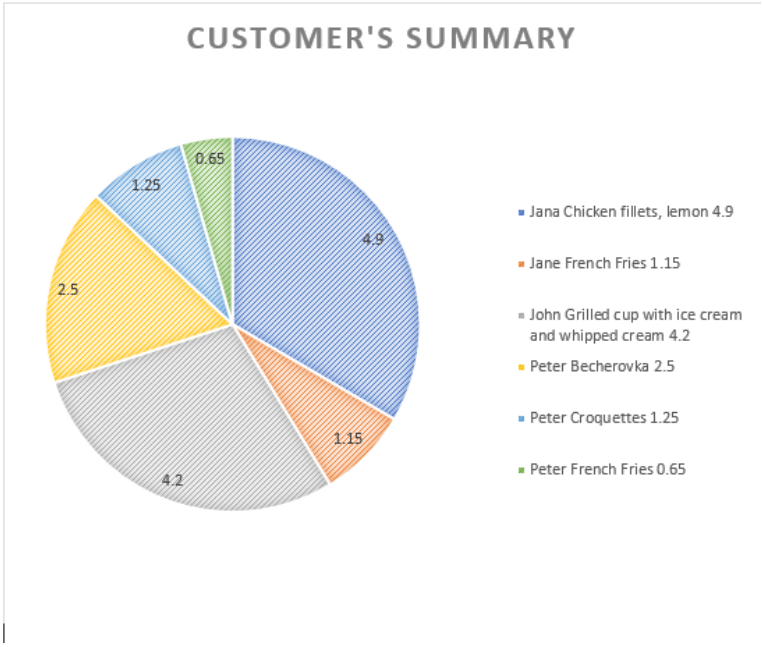


Figure 39:Pie chart showing data from table 1

Source: Own Contribution

Table 2: Pivot table showing summarized order summary

Customers	Sum of Total	Sum of Quantity	Sum of Amount
Jana	**4.9**	**1**	**4.9**
Jane	**1.15**	**1**	**1.15**
John	**4.2**	**2**	**2.1**
Peter	**4.4**	**4**	**3.65**
Grand Total	**14.65**	**8**	**11.8**

Source: Own Contribution

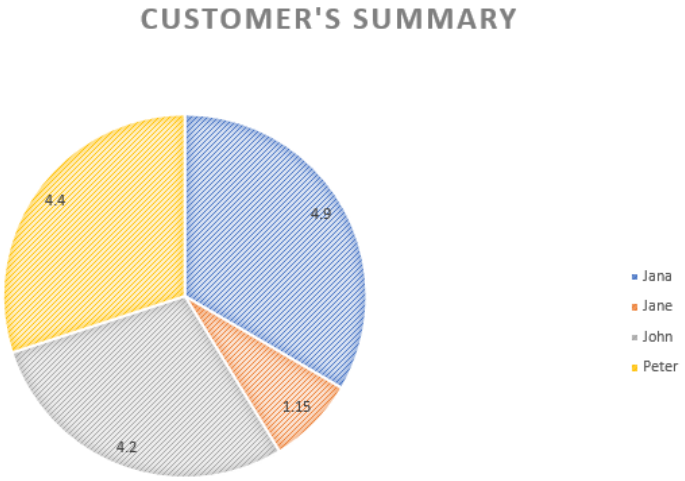


Figure 40:Pie chart showing data from table 2

Source: Own Contribution

Similarly, other tools like slicer can also be used to show a specific set of data for example, if the owner wants to see only orders that belong a single customer that can be done through the slicer tool. Such example can be seen in the fig.44 where the pie chart displays orders by only one customer i.e. Peter.

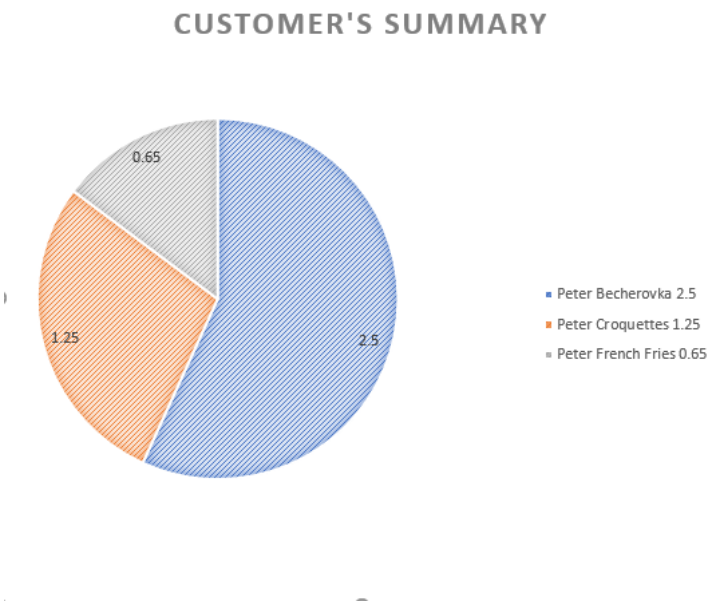


Figure 41: Pie chart showing orders from only Peter

Source; Own Contribution

6. Conclusion – Implication for Future Application

This thesis was based on the study of the application of various business intelligence tools, which can be used to support web activities of an enterprise. In this thesis, we outline the different concepts involved in business intelligence, data warehousing and data analysis. We also discussed about the various data models which can be used to describe the database and steps involved in creating the architecture of a data warehouse. To demonstrate the selected web activities (Online Ordering and Billing Summary, Menu Item Summary and Table Reservation) and their applications, specific business process modeling tools like flowcharts and BPMN was used. These models display the flow of action and the involved actors in the process. A fully functional data model was successfully created which was able to store the data in various tables and relationship between the data was also established. The objective of creating such database was to optimize the daily operations of a restaurant and try to integrate information and communication technology such as business intelligence tools to maximize the profits and to give the restaurant a competitive advantage over other business in the area.

Based on the results of the study it can be concluded that the integration of business intelligence tools to a business can prove to be a significant part to any growing business. These tools also improve the decision-making process in an enterprise as they provide the user with tools to analyze the data and also view the data from a different perspective that helps the owners to make more efficient and reliable decisions about the data.

Though the changes and tools suggested in the study might be successfully applied to the business, there can still be many improvements regarding the business processes. They may include introduction of newer tools of designing the data warehouse and tools to analyze and store the data. These tools may include metrics to measure bounce rate, time spent on website, time required to order food on website and recurring rate. Other than the mentioned ways simpler processes like contacting the enterprise top get real data may also enhance the business process. As suggested before, the field of data handling and processing is an ever-changing space, with time the data may change and the method to handle the data may change as well. Designing a much more robust database and database which can handle much larger volume of data may also be a future expansion possibility. Data Mining concept as mentioned earlier can also be studied and applied to the processes to make predictions regarding categorial and numeric variable.

Literature

1. DUMAS, M., M.L. ROSA, J. MENDLING, H. A. REIJERS. Fundamentals of Business Process Management. Second Edition. Springer 2018. ISBN 978-3-662-56508-7.
2. GROSSMANN, W., S. RINDERLE-Ma. Fundamentals of Business Intelligence Data-Centric Systems and Applications. Springer, 2015. ISBN 978-3-662-4653-1-8.
3. SHERMAN, Rick. Business Intelligence Guidebook: From Data Integration to Analytics. Newnes, 2014. ISBN 978-0-124-11528-6.
4. MARCEL, Patrick, Esteban ZIMÁNYI. Business Intelligence. Springer, 2017. Lecture Notes in Business Information Processing, 280. ISBN 978-3-319-61163-1.
5. KOHTAMÄKI, Marko. Real-time Strategy and Business Intelligence. Palgrave Macmillan, 2017. ISBN 978-3-319-54845-6.
6. VAISMAN, A., E. ZIMÁNYI. Data Warehouse Systems - Design and Implementation. Springer, 2014. ISBN 978-3-642-54654-9.
7. PROQUEST. 2019. Databáze článků ProQuest [online]. Ann Arbor, MI, USA: ProQuest. [cit. 2019-09-27]. Dostupné z: http://knihovna.tul.cz/ .
8. Dedić N. & Stanier C. (2016). Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting (PDF). Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting. Lecture Notes in Business Information Processing. Lecture Notes in Business Information Processing. 268. Springer International Publishing. pp. 225–236. doi:10.1007/978-3-319-49944-4_17. ISBN 978-3-319-49943-7. closed access.
9. ELMASRI, Ramez and Shamkant NAVATHE. Fundamentals of Database Systems.6th ed. Boston, Massachusetts, USA: PEARSON, 2011. ISBN 978-0-136-08620-8.
10. HARRINGTON J. Object-oriented Database Design Clearly Explained, Morgan Kaufmann, 2000. ISBN 978-0-123-26428-2.

11. SUMATHI, S. and S. ESAKKIRAJAN. Studies in Computational Intelligence: Fundamentals of Relational Database Management Systems. Volume 47. Berlin Heidelberg: Springer, 2007. ISBN 978-3-540-48397-7.

12. https://www.techopedia.com/definition/12027/object-oriented-database-management-system-oodbms.

13. QUADDUS, M., Arch G. WOODSIDE. Sustaining Competitive Advantage via Business Intelligence, Knowledge Management, and System Dynamics. Emerald Group Publishing, 2015. ISBN 978-1-784-41763-5.

14. RAUSCH P., Alaa F. SHETA, Aladdin AYESH. Business Intelligence and Performance Management: Theory, Systems and Industrial Applications. Springer Science & Business Media, 2013. ISBN 978144714866115.

15. HAN J., Micheline KAMBER. Data Mining: Concepts and Techniques. 2nd Edition. Morgan Kaufmann. ISBN: 978-1-55860-901-3.

16. HAND D., Heikki MANNILA, Padhraic SMYTH, Principles of Data Mining. MIT Press, 2001. ISBN: 978-0-26208-290-7.

Bibliography

DEDIć, N. a C. STANIER, 2016. *Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting: Lecture Notes in Business Information Processing. Lecture Notes in Business Information Processing.* Springer International Publishing, 225–236 s. DOI: 10.1007/978-3-319-49944-4_17. ISBN ISBN 978-3-319-49943-7.

ELMASRI, Ramez a Shamkant NAVATHE, 2011. *Fundamentals of Database Systems*. 6th edition. Boston, Massachusetts, USA: PEARSON. ISBN 978-0-136-08620-8.

Google Analytics Usage Statistics [online]. builtwith.com [cit. 2020-07-22]. Dostupné z: https://trends.builtwith.com/analytics/Google-Analytics

GROSSMANN, Wilfried a Stefanie RINDERLE-MA, 2015. *Fundamentals of Business Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg. Data-Centric Systems and Applications. ISBN 978-3-662-4653-1-8.

HAND, D., Heikki MANNILA a Padhraic SMYTH, 2001. *Principles of Data Mining*. MIT Press. ISBN 978-0-26208-290-7.

HAN, Jiawei a Micheline KAMBER, 2006. *Data mining: concepts and techniques*. 2nd ed. San Francisco: Morgan Kaufmann. ISBN 978-1-55860-901-3.

HARRINGTON, J, 2000. *Object-oriented Database Design Clearly Explained*. Morgan Kaufmann. ISBN 978-0-123-26428-2.

KOHTAMÄKI, Marko, ed., 2017. *Real-time Strategy and Business Intelligence*. Palgrave Macmillan. ISBN 978-3-319-54845-6.

MARCEL, Patrick a Esteban ZIMÁNYI, 2017. *Business Intelligence: Lecture Notes in Business Information Processing*. Springer. ISBN 978-3-319-61163-1.

Object-Oriented Database Management System (OODBMS) [online], 2011. [cit. 2020-07-18]. Dostupné z: https://www.techopedia.com/definition/12027/object-oriented-database-management-system-oodbms

QUADDUS, M. a Arch G. WOODSIDE, 2015. *Sustaining Competitive Advantage via Business Intelligence, Knowledge Management, and System Dynamics*. Emerald Group Publishing. ISBN 978-1-784-41763-5.

RAUSCH, Peter, Alaa F. SHETA a Aladdin AYESH, 2013. *Business Intelligence and Performance Management: Theory, Systems and Industrial Applications: Advanced Information and Knowledge Processing*. Illustrated. Springer Science & Business Media,. ISBN 9781447148661.

SHERMAN, Rick, 2014. *Business Intelligence Guidebook: From Data Integration to Analytics*. Newnes. ISBN 978-0-124-11528-6.

SUMATHI, S. a S. ESAKKIRAJAN, 2007. *Studies in Computational Intelligence: Fundamentals of Relational Database Management Systems.* 47. Berlin Heidelberg: Springer. ISBN 978-3-540-48397-7.

VAISMAN, A. a E. ZIMÁNYI, 2014. *Data Warehouse Systems - Design and Implementation*. Springer. ISBN 978-3-642-54654-9.