TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Economics

# Business Intelligence Tools for Process Optimization in a Public Electricity Company

## Master Thesis

*Study programme:* N0413A050008 Business Administration

*Study branch:* Management of Business Processes

*Author:* **Kishan Jagadeesh**

*Thesis Supervisors:* Ing. Athanasios Podaras, Ph.D.

Department of Informatics

Liberec 2021

TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Economics

## Master Thesis Assignment Form

# Business Intelligence Tools for Process Optimization in a Public Electricity Company

| | |
|---|---|
| *Name and surname:* | **Kishan Jagadeesh** |
| *Identification number:* | E19000548 |
| *Study programme:* | N0413A050008 Business Administration |
| *Study branch:* | Management of Business Processes |
| *Assigning department:* | Department of Informatics |
| *Academic year:* | **2020/2021** |

**Rules for Elaboration:**

1. Business Intelligence tools and business process implementation
2. Goal, objectives, methodology
3. Data description and process delineation in a selected public electricity company
4. Process optimization with business intelligence tools
5. Presentation, analysis of the results and discussion
6. Conclusions and future tasks

*Scope of Graphic Work:*

*Scope of Report:* 73 normostran

*Thesis Form:* printed/electronic

*Thesis Language:* English

**List of Specialised Literature:**


- GROSSMANN, W., S. RINDERLE-Ma. Fundamentals of Business Intelligence. Data-Centric Systems and Applications. Springer, 2015. ISBN 978-3-662-4653-1-8.
- SHERMAN, R. Business Intelligence Guidebook: From Data Integration to Analytics. Newnes, 2014. ISBN 978-0-12-411528-6.
- MARCEL, P. & ZIMÁNYI, E. Business Intelligence. Springer, 2017. Lecture Notes in Business Information Processing, 280. ISBN 978-3-319-61163-1
- KOHTAMÄKI, M. Real-time Strategy and Business Intelligence. Palgrave Macmillan, 2017. ISBN 978-3-319-54845-6.
- VAISMAN, A. & ZIMÁNYI, E. Data Warehouse Systems – Design and Implementation. Springer, 2014. ISBN 978-3-642-54654-9.
- PROQUEST. 2020. *Databáze článků ProQuest* [online]. Ann Arbor, MI, USA: ProQuest. [cit. 2020-09-18]. Dostupné z: http://knihovna.tul.cz/


*Consultant:* Ramesha Babu B G, Manager / Finance, BHEL-EPD

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

*Date of Thesis Assignment:* November 1, 2020

*Date of Thesis Submission:* August 31, 2022

L.S.

Ing. Aleš Kocourek, Ph.D.
Dean

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

Liberec November 1, 2020

# 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 up-loaded 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.

April 17, 2021 Kishan Jagadeesh

# Anotace

Tato diplomová práce se zaměřuje na aplikaci principů business intelligence za účelem optimalizace správy záznamů zaměstnanců a dalších souvisejících činností v elektroenergetické společnosti. Tato optimalizace je prováděna především prostřednictvím informačního systému s relační databází, čímž je dosaženo zajištění efektivního záznamu, extrakce a analýzy enormního množství dat. Získaná data organizace dále používá k analýze a správě zaměstnanců.

Na začátku práce je vysvětlena nutná a nezbytná koncepce týkající se systémů business intelligence. Poskytuje stručné vysvětlení o architektuře BI, datech a vlastnostech dat, datovém skladu a dalších. K dispozici jsou také popisy struktur modelování dat. Všechny teoretické znalosti jsou použity v další části práce, kde je vysvětlena praktická implementace systému BI. Data získaná od společnosti jsou analyzována pomocí MS Access prostřednictvím různých tabulek. V praktické části je rovněž popsáno vytvoření vztahu mezi tabulkami pomocí technik modelování dat. Analýza dat a vytváření reportingových sestav jsou prováděny pomocí různých datových dotazů. Práce je zakončena shrnutím výsledků a potenciálních omezení použitých technik. V závěru jsou nastíněny další možné budoucí směry v rámci realizovaného tématu.

# Klíčová slova

Systém Business Intelligence, optimalizace procesů, datový sklad, systém správy databází, modelování dat.

# Annotation

The current thesis focuses on the application of Business Intelligence system to optimize the process of employee record management and other related activities in an electricity company. It is mostly done by introducing a relational database management system hence ensuring efficient recording, extraction and analysis of the enormous amount of data. The data extracted is further used by the organization for employee analysis and management.

The thesis in the beginning explains required and necessary concepts regarding business intelligence systems. It provides brief explanations about BI Architecture, Data and Data properties, Data Warehouse and more. There are also descriptions about data modelling structures. All the informative knowledge is used in the next part of the thesis where the practical implementation of the BI system is explained. The data collected from the company is analysed using MS Access with the help of different tables. It also shows the establishment of relationship between tables using data modelling techniques. There are different queries created for data analysis and reporting. The thesis is concluded with work carried out and its limitations. In the end it shows brief points on future works related to the existing thesis.

# Keywords

Business Intelligence system, Process optimization, Data Warehouse, Database Management system, Data Modelling.

# Contents

# List of Figures

# Abbreviation

BI – Business Intelligence

ETL- Extract Transform Load

ER – Entity Relationship Diagram

RPD - Rapid File Database

DBMS – Database management system.

RDBMS – Relational Database Management System

SQL – Structured Query Language

BPMN – Business Process Model and Notation

PK – Primary Key

FK- Foreign Key

UML – Unified Modelling Language

# Introduction

In general, the process of managing a business organization may seem simple consisting of a small number of activities in order to accomplish the intended goal in accordance with the kind of business or organization being run. But, as we get into the crux of the process, we realize that there are innumerable smaller activities and as many entities involved. There is a need for analysing and simplifying the process and help the employees to be well organized and efficient which is provided by Business Intelligence.

People encounter many problems, obstructions and challenges as a part of an organization or a society. It is important to work efficiently in order to solve these problems, and to optimize the activities in order to maintain effectiveness and competence. Last few decades has seen an exponential growth in the usage of electronic and digital platform for performing various organizational activities, like recording, processing, collection and maintenance of enormous amount of data. The massiveness of these data made it difficult to extract, realize or analyse them and hence brought about an emergence of Business Intelligence systems. (Dmitry Anoshin, 2016)

Business Intelligence systems helps in effective decision making based on the recorded data. It helps the users to possess an idea about the company standings, competitive comparison etc., and allows them to progress accordingly. It is most favourable to use because it provides quick, easily perceived, and segregated form of the enormous amount of data, making it easier to take decisions for the future. (Mary K. Pratt, 2019). Many firms consider BI activities as a process which focuses on supervising and controlling the competitive environment of an organization. It also helps in anticipating actions of customers and the competitors. (Hannula & Pirttimaki, March 2003)

This diploma thesis is aimed at solving a practical and internal problem in an electrical company, using Business Intelligence system. It involves optimizing the process of evaluating and recording data regarding salary details, time of entry and exit of the employee, personal data of the employee etc. This enormous amount of data recorded in a normal spreadsheet is positioned at an inefficient condition, making it difficult to extract and evaluate. Hence Business Intelligence solution is used in order to simplify and improve the process.

# 1. Aim and Objectives

The main goal of the thesis is to improve the method of internal data recording and management in an electricity company. In line with this, there is an effort made to use Business Intelligence tools and make this activity much easier and faster in its real time application. It also aims at structuring the data in an organized manner. This thesis brings about an integration of MS Access and Excel spreadsheets for recording and evaluating the employee records. The amalgamation of these two software helps in taking advantage of the benefits of each other.

The following procedures in the form of a basic structure are to be met in order to achieve the goal of the thesis.

i. Listing the present methods and processes of the company, and the drawbacks with respect to process efficiency are listed.
ii. Research to find out the alternatives to improve the activities using Business Intelligence system tools.
iii. Identification and solution to optimize the process.
iv. Designing the conceptual and logical model of the required data warehouse solution
v. Executing the design in a data base management system software used for viewing, updating and reporting big data approach

# THEORETICAL SECTION

# 2 Business intelligence

## 2.1 Business intelligence definition

Business Intelligence is defined in the paper of H.P. Luhn. The term 'Intelligence' is 'the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal' and 'Business' is "a collection of activities carried on for whatever purpose, be it science, technology, commerce, industry, law, government, defence, etc.'. (Wilfried Grossmann, 2015).

Business Intelligence consists of a collection of methodologies, processes, architectures and technologies that transform raw data into useful information for decision making. It provides assistance in managing various organizational levels of hierarchy for analysing strategic information. The main objective is to simplify complex data into an abridged form. The conversion process of the vast amount of data into an organized and simplified form comprises of a set of activities, namely extraction of the data from a specified source, transformation of the data, integration of the data, filtering or cleansing and storing of the data into a particular repository. This repository is known as Data Warehouse. It will be explained in detail in the subsequent chapters. (Alejandro Vaisman, 2014)]

Business Intelligence is defined in the paper of H.P. Luhn. The term 'Intelligence' is 'the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal' and 'Business' is "a collection of activities carried on for whatever purpose, be it science, technology, commerce, industry, law, government, defence, etc.'. (Wilfried Grossmann, 2015).

Business Intelligence consists of a collection of methodologies, processes, architectures and technologies that transform raw data into useful information for decision making. It provides assistance in managing various organizational levels of hierarchy for analysing strategic information. The main objective is to simplify complex data into an abridged form. The conversion process of the vast amount of data into an organized and simplified form comprises of a set of activities, namely extraction of the data from a specified source, transformation of the data, integration of the data, filtering or cleansing and storing of the data into a particular

repository. This repository is known as Data Warehouse. It will be explained in detail in the subsequent chapters.
(Alejandro Vaisman, 2014)

## 2.2 Business Intelligence Methodology

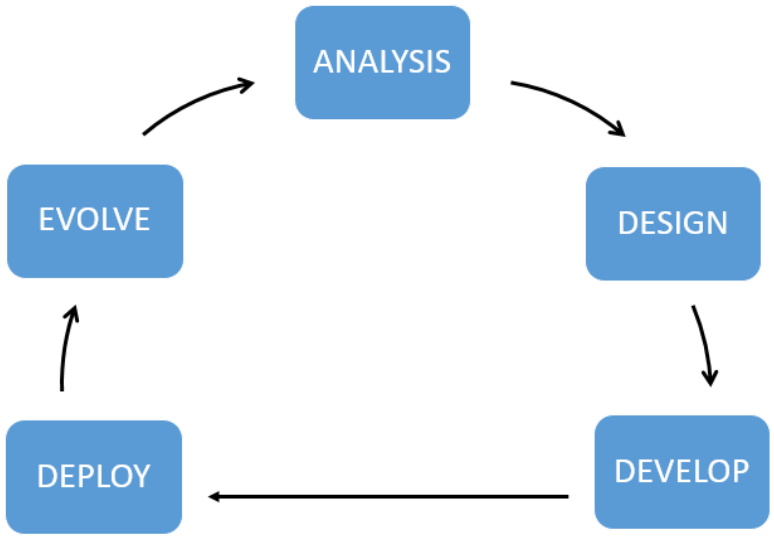



*Figure 1 Life Cycle of Business Intelligence System*

Source; Own Contribution

(Gangadharan & Swami, 2004)

### 2.2.1 Analyse

The project should start of by validating the advantages and disadvantages of optimizing the business process. The main objectives of the process are realised and stated and the required information is collected from many resources and relevant solutions are amassed. The organizational requirements are recognized and the necessary processes are established in order to fulfil it.

### 2.2.2 Design

According to the level of process optimization and its solving methodology, the suitable solutions are selected from the data gathered previously. Initially a prototype design has to be made as to what our solution would resemble. Based on this design the appropriate resources and other elementary parts required for the process are decided.

### 2.2.3 Develop

The process of designing conceptual and logical data are prepared in this particular stage. The analysis and development of frames, rules and constraints which are generally referred to as meta-models are designated and created.

The database for storing the data is designed based on the constraints regarding meta-model. The data undergoes ETL process in the database system where the data is filtered and transformed into a required form with an appropriate output.

### 2.2.4 Deploy

As soon as the process is completed, the data is made ready for the users with an efficient transformation. The end users interact with the data and based on their reaction to the user interface there is development in the reports and analysis. This helps in future advancement in the process of data optimization.

### 2.2.5 Evolve

This stage measures and level of success the process has received and analyses future paths of development in a more efficient manner. It takes feedback from the end users and tries to improve the application of the data system.

(Gangadharan & Swami, 2004)

## 2.3 Business intelligence architecture

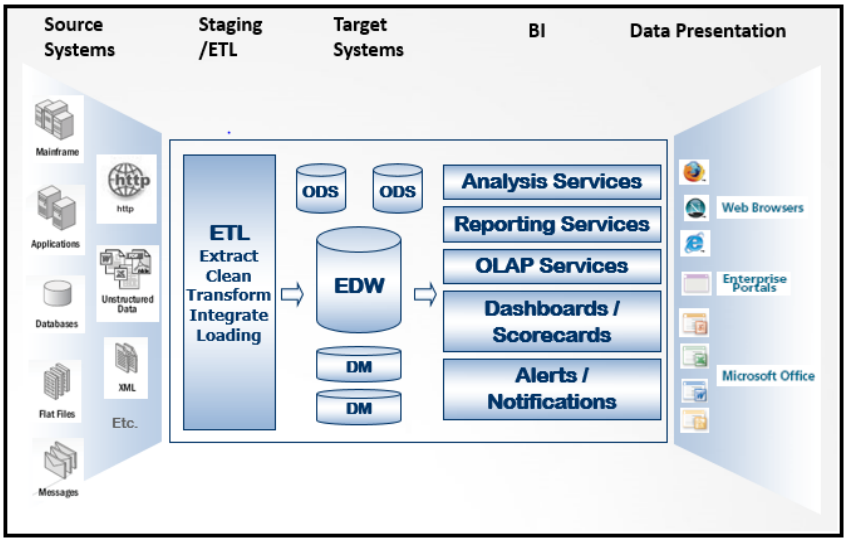



*Figure 2 BI ARCHITECTURE*

Source; (technologies, March 2014)

The business architecture chart comprises of the following:

**1 Source view layer:**

- the information stored, captured, and managed by a system is presented in this layer. All the data required for data processing is extracted and executed here.
- This layer stores many hierarchies of data in a combined or separated form.
- It consists of diversified and varied dataset.

**2 Data staging layer:**

- In this particular layer the data extracted is transmitted and passed through a process of removal of discrepancies and inefficiencies.
- The diversified layer is merged into one homogenous standard representation.
- The process of merging, extracting, transforming, filtering and loading is done using ETL tools

**3 Data warehouse layer:**

- This table consists of fact tables and dimension tables.

- Information stored in the data warehouse is presented in this layer.
- Information is saved to a single central storehouse
- It is accessed directly and also through sub divided sections called DataMart.

**4 Data presentation layer:**

- This layer provides data representation from the viewpoint of the end users.
- This provides efficient and flexible access of the data.
- It is used to issue testimonies, final results, analysis reports, information query and user friendly interfaces.

## 2.4 Business process modelling.

In a workspace or an organization there are innumerable yet smaller processes involved and hence the employees have to be informed and acquainted with the principles and end goals of the organization.

These workflows in the management system is achieved through the business process modelling.

### 2.4.1 Definition of Business process modelling

The business process modelling refers to creating a graphical representation or an illustration of different workflows used in a company to get a clear understanding of their business processes and also analyse new strategies to improve the current processes. (Anon., 2019)

Business process modelling helps in representing the process in a digitized manner which is later transferred into an automated process. It is a representation of connected activities of the product or service in order to improve it.

### 2.4.2 Business Process modelling tools

#### a) Business Process modelling notation

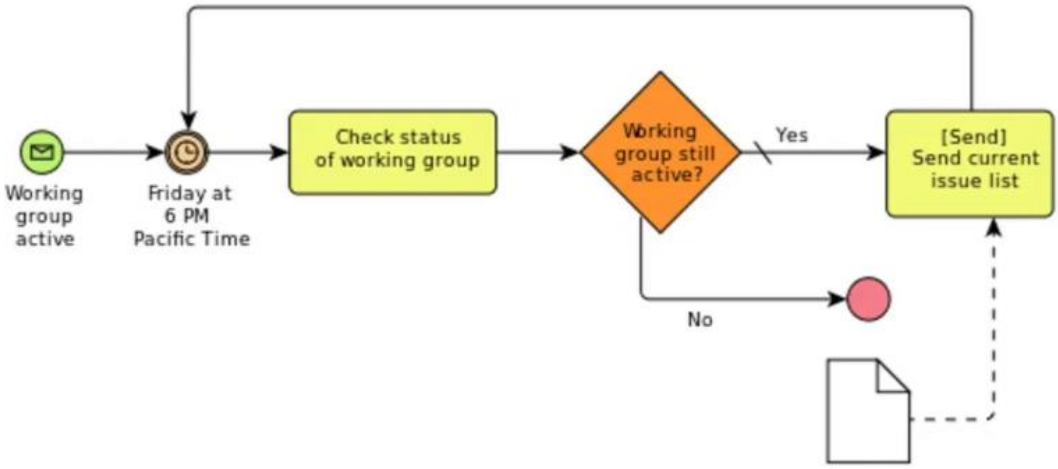



*Figure 3 BPMN Model*

Source; (Anon., 2019)

Business process modelling notation is the graphical representation using standardized symbols which are recognizable by all professionals all over the world.

BPMN is considered as the most accurate and widely accepted method for analysing and processing workflows through pictorial representation.

Some of the components used for the representation are:

- **Flow objects**

i. **Activities** – As the name suggests, activities are the work performed within the organizational process.

There are two types of activities: task and sub-process.

a. Task – when the work or activity to be performed need not be divided into smaller groups the task activity is used.


*Figure 4 Example of tasks*

Source; (Anon., 2012)

b. Sub-groups – in situations of accomplishing complex activities, there is a division of the work process into many smaller sub-groups hence enabling the organization to simplify the process by sharing responsibilities to different employees or departments.

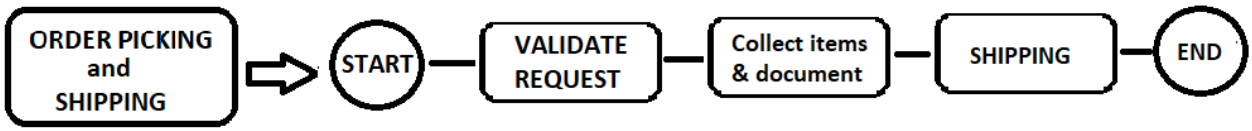



*Figure 5 Example of Sub-groups*

Source; (Anon., 2012)

ii. **Control points** – Control points are used to control the flow of business process flows represented by diamond shapes. It is a point of process where the required conditions are assessed and the choices are made.

**Some types of control points are:**

a. Data based Exclusive Gateway

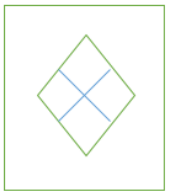


*Figure 6 Exclusive control point*

Source; (Lewis, 2020)

In this particular control point, data which flows through undergoes a conditional function. There will be more than one outgoing flow and each outgoing flow corresponds to a condition. Only one condition gets through to the output. All the outgoing sequences are evaluated and the sequence which matches the required condition is chosen for the process flow

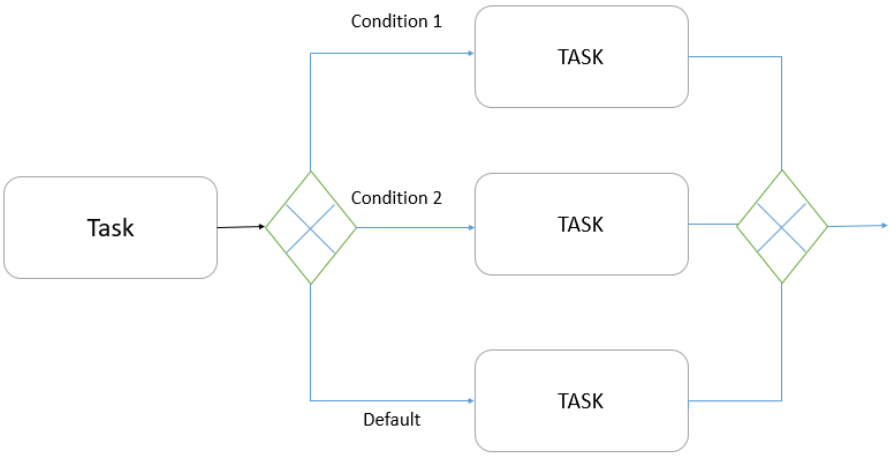



*Figure 7 example for Exclusive gateway*

Source; (Anon., 2018)

b. Inclusive control point

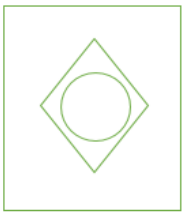


*Figure 8 Inclusive control point*

Source; (Lewis, 2020)

As the name suggests, inclusive control points has more one outcomes. Unlike exclusive control points, it allows more outcomes if the condition is satisfied. All the conditions which analyses the given sequence of process true is directed to the outflow.

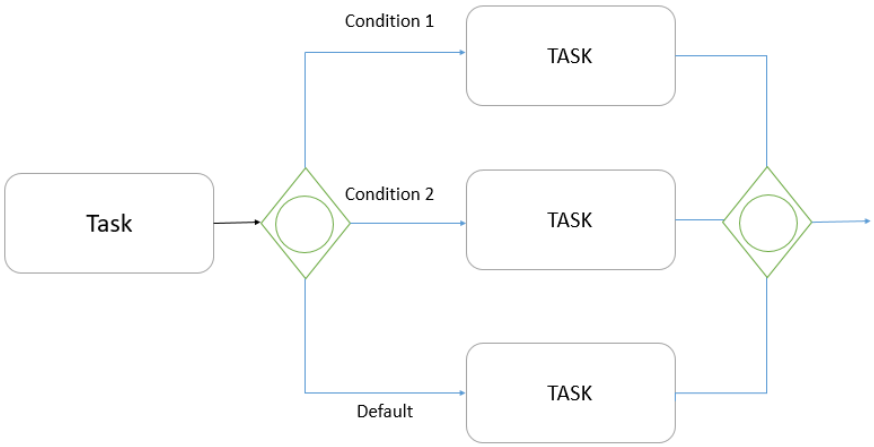



*Figure 9 Example for inclusive control point*

Source; (Anon., 2018)

c. Parallel control point:

*Figure 10 Parallel control point*

Source; (Lewis, 2020)

Parallel control points represent two tasks that occur in the business flow. In this gateway, there exist two types of processes, splitting and joining. Splitting creates a parallel path to the sequence which are outgoing. Joining process is a point where the parallel incoming paths of sequences are in a waiting position before the next process occurs.

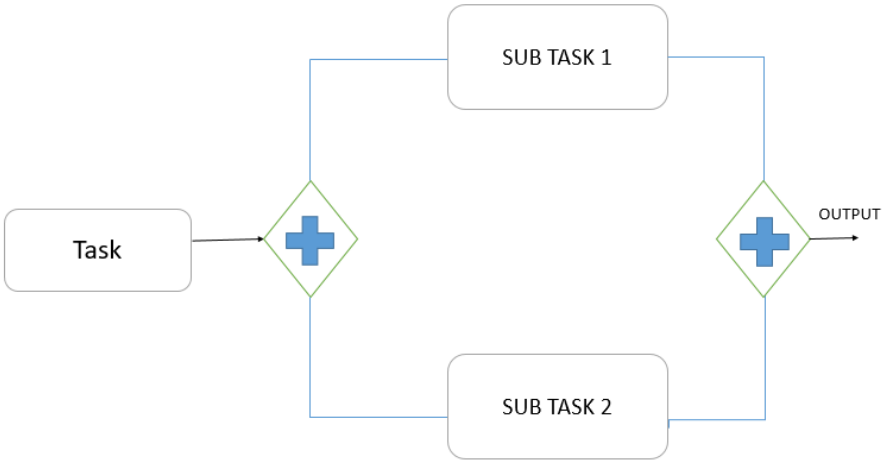



*Figure 11 Example for parallel control point*

Source; (Anon., 2018)

iii. Events - Events are represented by a circle, which shows that which occurs in the process unlike activities which are done. Different sets of events are represented by different objects inside the circle.


*Figure 12 Symbols for events*

Source; (Lewis, 2020)

- Start – Start events are events which act like an initiator or process starter which triggers the process the to begin the process. It is represented by a single narrow border.

- Intermediate – intermediate events occur between the process of start and end events and are represented by a double border. This either collects or throws the message.
- End – the end outcomes are indicated in these events. They are represented by a single bold border and it provides the results at the culmination of the process.

- Connecting objects – The flow objects are not in a separate or isolated state, but connected with each other with help of connectors. The connectors which connect the flow objects are known as connecting objects.
  Some types of connecting objects are:

- **Sequence flow connector**

This represents the task which has been transferred from one point to with the help of a long line. It connects all the flow objects in given sequential order. It is represented by a long line with an arrow at the end.

- **Message flow connector**

It represents the messages which are transferred from one point to another. All the information flow sent from one participant of the process to another is shown through these dashed arrow symbol which has a circle at the start. It will connect events or activities within a pool but represents those which flow across the pools.

- **Association connector**



This connector represents the relationship between artefacts and flow objects. It is indicated by dotted lines. It associates are connects artefacts with an event, activity or gateway.

- **Swim lanes** – the division of work between two or more people is represented by swim lanes. They are used in organizing and managing certain aspects of a process in a BPMN illustration. Swim lanes groups objects into different lanes, where different process features are entered into different lanes. They not only group the activities in different lanes but also shows inefficiencies and the people who are in charge for every step of the process. There exists lanes and pools where lanes represent the task and pools represent the people.

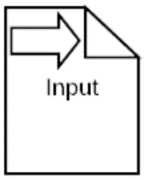

*Figure 13 Pool with two lanes*

Source; (Cimino, 2009)

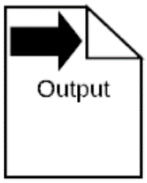
- Pool – It represents the main participants of the processes occurring in a business activity. These groups can be in different organization or a different department in the same company.

- Lanes – They are groups which are represented under different pools. It indicates the activities performed by a certain participant, hence describing people who are responsible for certain parts of the process.

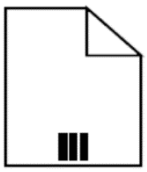
- **Artefacts -** The additional information required to bring in another level of detail to the diagram is filled in by the artefacts. It represents data objects, group and annotations. \
- Data objects – represent different kinds of data like input data, output data, data that are collected and data stored.




a. **Data input** – the data required for the functioning of the task is called input data. This data has its occurrence before the process begins. For every task there is a requirement of data for which the task is done. This data is known as input data.



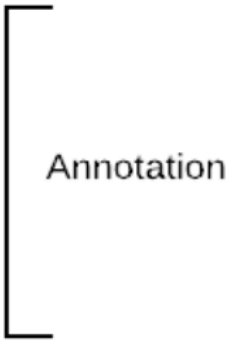

b. **Data output** – after the processing occurs, there is a separate set of data produced as output. This data is represented as output data



c. **Data collection –** it is not the same as the data requirement for the process to begin. For instance, the data collection process during survey is represented by the given shape.


d. **Data storage –** when there is an output data produced during the process, it is necessary to store that data in a virtual storehouse. This helps us store and access data whenever required.

- **Annotations**




It helps in describing the business process and flow objects in depth. Hence annotations are added to make BPMN more readable and understandable. It enables the modeller creator to describe or provide additional explanation to different parts of the notation.

- **Groups**


They are used to arrange the tasks and processes in a given process in order to make it easier and efficient for the company. They collect different activities, but doesn't affect the activity flow of the objects

(Anon., 2019) (Manning, 2020) (Cimino, 2009) (Anon., 2018)

# 3 Data and Information

## 3.1 Data properties and attributes

In the data warehouse, the data possesses its own characteristics which is unique to its function.

**a. Subject oriented**

- This property defines data and categorizes it according to a particular subject
- This helps to analyse the data more efficiently as it is defined more precisely
- It also helps in quicker retrieval of the required data.

**b. Integrated**

- The data warehouse is injected with a large number of data from different sources.
- These data might be in different formats and structures which cannot be processes directly to the end user.
- Hence all data is transformed into a standard format in order to help the user perform complex operations and processes in an efficient way.
- The data is transformed with the help of ETL process and sent to the data warehouse.

**c. Time Variant**

- Data is organized according to different time period making it easier to locate and explore.
- The property also defines about the storage of all historical data helping in prediction of required analysis of real world problems.
- The historical data is useful in the requirement analysis of data, as it can compare previous data with the present data and provide appropriate results.
- This characteristic helps data warehouse to have both present updated data and also historical value of that data.

**d. Non –Volatile**

- The data has a characteristic of permanency, as data is not erased or deleted when a new data is introduced.
- Data warehouse is updated only via the uploading of the data, protecting it from the
- The data is ‘read-only’ and hence cannot be erased.

- This helps in analysis of historical data and in understanding the range of operations.
- ‘Delete’, ‘update’ and similar operations facilitates to loss of data.

(Gupta, 2020)

## 3.2 Database management system

A database management system (DBMS) is a functional software used to define, control, retrieve and manage data. The control is applied to the data, data format, structure of the recorded data and data size. The previous versions of the database management system could only handle an individual dataset, whereas the current systems are much more advanced. They can handle huge quantities of datasets, less efficient and complicated information and expound it in a much better approach.

### 3.2.1 Features of database management system

**a.** It is an extended version of human intellect making it more logical in performance and analysis.
**b.** The speed of data retrieval is high.
**c.** It gives precise information and solution to complicated queries regarding the dataset.
**d.** The database presents data in a simplified manner, which makes it easier to grasp without difficulty.
**e.** Due to the efficiency of the system, both time and money is saved in the process of operating DBMS.

### 3.2.2 Database system components

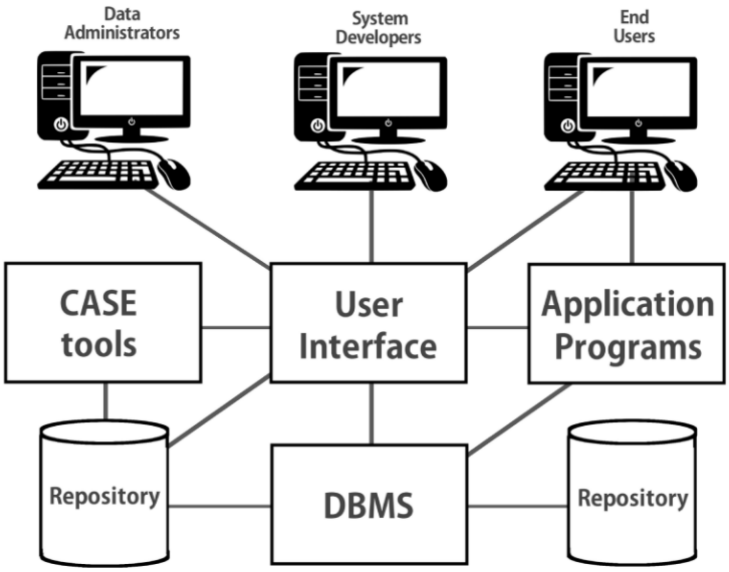



*Figure 14 Database Components*

Source; (Anon., 2019)

**The database environment components are:**

**CASE Tools** – they are used in designing databases and application programs.

**DBMS** – it is used to define, create, analyse and control the usage for the user databases

**Database** – it is the data collected in an organized manner.

**Data administrators** – it controls and looks after the activities of the database system.

**Application programs** – the programs are used to apply all the design in creating and maintaining the database.

**Repository** – it the storage house for all the metadata of the respective database system.

**User interface** – it is the interface viewed by the end users. This stage is where people interact with the components of the system.

**System Developers** – this is a stage where there is constant development of new designs and applications are in process, in order to improve the system.

**End Users** – the people who use and interact with the database systems. They are the subjects for whom the database design is implemented in the first place.

(Anon., 2017)

## 3.3 Relational database management system

E.F. Codd's model of relational database states that an RDBMS allows helps the users in construction, update, management, and interaction with a relational database which amasses data in the form of tables. It can be inferred that RDBMS is an improved and advanced version of data management system through which realizing intricate details about the data is made a lot easier.

The relational database architecture is preferred to most other architectures like flat files or hierarchical database in order to store the data of an organization. Its preference is due to the ability of RDBMS to process an extensive range of different data structures and queries. It also possesses a property of organizing data in the form of tables and interlinking different tables with each other which is useful in retrieving multiple tables in an efficient manner. In case of flat-files, the data is stored in a single table visually representing a two dimensional structure, where the retrieval of data is more time and space consuming, hence making it much less optimal when compared to the RDBMS.

(Anon., 2018)

### 3.3.1 RDBMS components

- **Table –** In a relational database management system. Tables are used in order to store data. It contains a collection of data which are related to each other. The data storage is done in the form of rows and columns.

- **Record –** Every row in a table represents a record. It possesses a particular set of information which are which are specific to every entry provided in the table. It is known as the horizontal entity of the table.

- **Column or Attribute –** it contains all information in the table which are specific to a particular field provided in the table. It is also known as the vertical entity.

- **Null Values –** null value as the name suggests is a blank space or a 'zero' in a table field. The value of the blank is either zero or no value.

- **Keys - Primary key** is a single column or a set of columns in a table that possesses a unique identification of different rows in that particular table. **Foreign key** represents the columns of a table which is directed to the primary key of another table. They possess the property of acting as a cross-reference between different tables.

### 3.3.2 Entity Relationship Diagram

The entity relationship model is helpful in the description of the structure of a database with the help of a pictorial representation called Entity Relationship diagram. It is a high level data model. An ER model is the basic conceptual design of the database which is used in order to practically implement the actual functioning database.

The ER diagram provides the description of relationship among different entity sets.

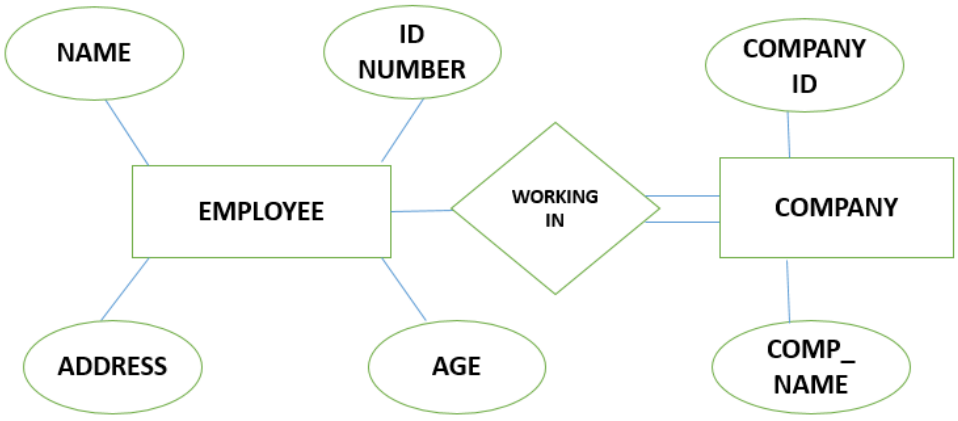



*Figure 15 Example of a basic ER diagram*

Source; (SINGH, 2016)

### 3.3.3 Constituents of ER diagram

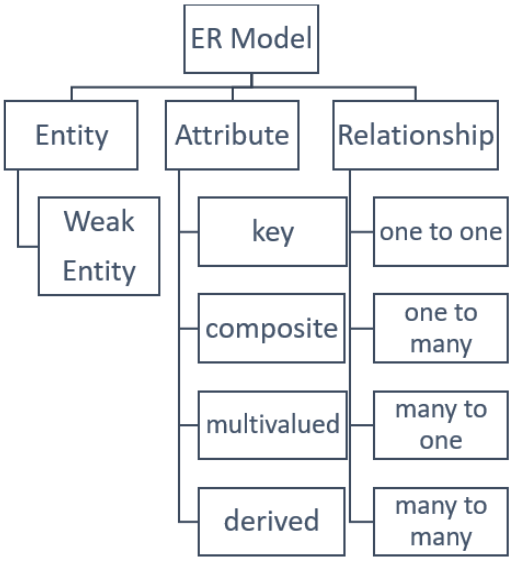



*Figure 16 Constituents of the ER diagram*

Source; (SINGH, 2016)

- **Entity**

An entity is an object, class, person or place. The ER diagram represents entity in the form of rectangles.

Example: For the given example


Source; (SINGH, 2016)

- **Weak Entity**

It is an entity which depends on another strong entity for its existence. It does not possess key attributes on its own. The primary key possesses key attribute of the related Strong entity along with its own attribute. It is represented by a double rectangle.

Example: Company ID and Company name.


Source; (SINGH, 2016)

- **Attributes**

The characteristics, properties and features are described by an attribute. An attribute is represented by an oval.

**The different types of attributes are:**

- **Key attribute** – it describes the main characteristics of the entity. It is the representation of a primary key which is denoted by an ellipse with the text underlined. It is providing a unique identity for a particular entity from a table.

  Example: Employee and Employee ID.

  In the given example Employee ID is the key attribute or the primary key for a given entity set.

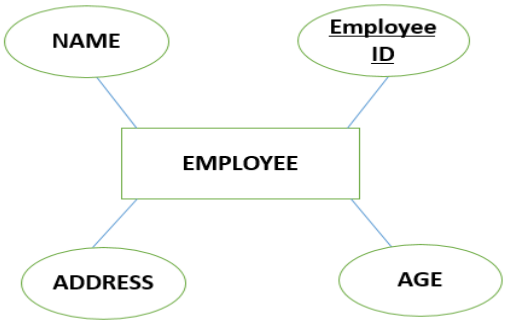



*Figure 17 Example for Key attribute*

Source; (SINGH, 2016)

- **Composite attribute** – it is that attribute which has a combination of 2 or more attributes. It is denoted by an ellipse connected to 2 or more ellipses in a hierarchical structure.

  Example: Employee and Employee details.

- **Multivalued attribute** – this attribute contains more than one value. It is represented by a double oval in the diagram

  Example: Employee and Employee's multiple phone numbers.

- **Derived attribute** – it is an attribute whose value is dynamic and keeps changing according to changing situations. It is derived from another different. It is represented by dashed oval in the diagram.

  Example: Employee and Employee age

- **Relationship** – the relationship among different entities is represented by the ER model. It is denoted by a diamond shape. There exist 4 kinds of relationship:

**One to One**

A relation is said to be one to one, when one instance of an entity possesses a relationship with a single instance of another entity.

Example: a person has only one passport, and a passport is provided for one person.


Source; (SINGH, 2016)

**One to Many**

A relation is said to be one to many, when a single instance of an entity possesses a relationship with more than one instance of another entity it is said to be 'one to many' relationship.

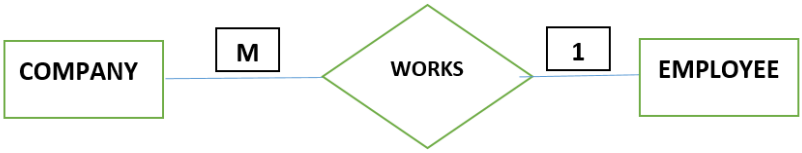
Example: a customer can place many orders, but the same order cannot be placed by many customers.


Source; (SINGH, 2016)

**Many to One**

A relation is said to be Many to One, when more than one instance of an entity is assorted with a single instance of another entity

Example: A company can have many employees, but an employee can work only in one company at a time.

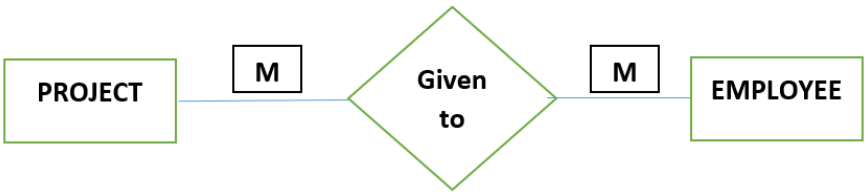



Source; (SINGH, 2016)

**Many to Many**

A relation is said to be Many to Many, when more than one instance of an entity is assorted with any number of instance of another entity.

Example: An Employee can have many projects to work with and at the same time, a project can be assigned to many Employees.




Source; (SINGH, 2016)

### 3.3.4 Normalization

*"Normalization of data can be considered a process of analysing 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). It makes the data more efficient and reliable. A large table is divided into smaller tables and linked to the main table using relations. Generally, the normalization is done up to 3rd level. Even though 4th and 5th level normalization is present, they are not used extensively unless it is necessary.

**1st normalization**

In 1st normalization form the two criterions which must be fulfilled are that every attribute of a table contains a single data values and all the data recorded should be containing unique value. Hence 1st normal form prohibits the use of cells containing multiple values, or repeated attributes.

Given below is an **example** for the normalization process. Here a library is supposed to store the data of students by names, address, study programme and the books they have borrowed from the library.

Table1

| STUDENT NAME | ADDRESS | BOOKS | PROGRAMME |
|---|---|---|---|
| Ayaan Acharya | 6th street 46 | "Economics in One Lesson", "Basic Economics" | MASTERS |
| Erik Christopher | 22nd Avenue | "Freakonomics" ,"Capitalism and Freedom," | PhD |
| Erik Christopher | 3rd street plot 8 | "Doughnut Economics." | MASTERS |

*Figure 18 TABLE WITHOUT NORMALIZATION*

Source; (Anon., 2018)

From the above table there is conversion of data table into 1st Normalization by creating a new cell for every data in the column 'BOOKS'. Hence there are no multivalued cells in the attribute.

| STUDENT NAME | ADDRESS | BOOKS | PROGRAMM |
|---|---|---|---|
| Ayaan Acharya | 6th street 46 | "Economics in One Lesson", | MASTERS |
| Ayaan Acharya | 6th street 46 | "Basic Economics" | MASTERS |
| Erik Christopher | 22nd Avenue | "Freakonomics" | PhD |
| Erik Christopher | 22nd Avenue | "Capitalism and Freedom," | PhD |
| Erik Christopher | 3rd street plot 8 | "Doughnut Economics." | MASTERS |

*Figure 19 1ST NORMALIZATION FORM*

Source; (Anon., 2018)

**2nd normalization**

For a 2nd normalization form to hold good, it should be in 1st normalized form and all non-primary attributes depend on the primary key.

In the given example the 1st normalized form is segregated into two tables. The 1st table has information about the library membership and another table has information of the books borrowed from the library. There is an introduction of a new column 'ID' in the 1st table which

acts as the primary key and all the non-primary records are identified with the help of this primary key in 1st table.

| ID | STUDENT NAME | ADDRESS | BOOKS | PROGRAMME |
|---|---|---|---|---|
| L1 | Ayaan Acharya | 6th street 46 | "Economics in One Lesson", "Basic Economics" | MASTERS |
| L2 | Erik Christopher | 22nd Avenue | "Freakonomics" ,"Capitalism and Freedom," | PhD |
| L3 | Erik Christopher | 3rd street plot 8 | "Doughnut Economics." | MASTERS |

*Figure 20 2ND NORMALIZATION*

Source; (Anon., 2018)

| LIB MEMB ID | BOOKS |
|---|---|
| L1 | "Economics in One Lesson", |
| L1 | "Basic Economics" |
| L2 | "Freakonomics" |
| L2 | "Capitalism and Freedom," |
| L3 | "Doughnut Economics." |

*Figure 21 LIBRARY MEMBERSHIP ID TABLE CONTAINING FOREIGN KEY*

Source; (Anon., 2018)

### 3rd normalization

The 3rd normalization form makes the model further more efficient and simplified. For 3rd normalization form to hold good, it should be in 2nd normalized form and there exists no transitive dependency. One of the primary functionalities of the 3rd normalization form is reduction of duplication of data.

In the given example the 2nd normalized table is further normalized, where a 3rd table for the programme of studies of the students is provided a unique ID – 'Programme ID'. This ID acts as a primary key in the new table and in the 1st table this ID acts as a foreign key.

| ID | STUDENT NAME | ADDRESS | BOOKS | PROGRAMME |
|---|---|---|---|---|
| L1 | Ayaan Acharya | 6th street 46 | "Economics in One Lesson", "Basic Economics" | P1 |
| L2 | Erik Christopher | 22nd Avenue | "Freakonomics" ,"Capitalism and Freedom," | P2 |
| L3 | Erik Christopher | 3rd street plot 8 | "Doughnut Economics." | P1 |

*Figure 22 3RD NORMALIZATION*

Source; (Anon., 2018)

| LIB MEMB ID | BOOKS |
|---|---|
| L1 | "Economics in One Lesson", |
| L1 | "Basic Economics" |
| L2 | "Freakonomics" |
| L2 | "Capitalism and Freedom," |
| L3 | "Doughnut Economics." |
| | |

*Figure 23 LIBRARY MEMBERSHIP ID TABLE CONTAINING FOREIGN KEY*

Source; (Anon., 2018)

| PROGRAMME_ID | PROGRAMME |
|---|---|
| P0 | BACHELORS |
| P1 | MASTERS |
| P2 | Ph.D |
| | |

*Figure 24 PROGRAMME ID*

Source; (Anon., 2018)

### 3.3.5 Advantages of RDBMS:

- The data stored is non replicable as it is stored a single time, making it more accurate.
- Several users can access the database making it more mutual and co-operative between users.
- Due to their flexibility carrying out complicated queries is made easier.
- The data provided in the database is well secured, as the access can be limited to specific users

(Biscobing, 2021)

### 3.3.6 Differences between relational database and database management system

| RDBMS | DBMS |
|---|---|
| **1.** Multiple users can operate at the same time<br>**2.** Possesses complex algorithms which is essential for simultaneous access of the database by multiple users.<br>**3.** Utilizes more resources for the required data<br>**4.** Due to its complexity in structure and function they are costlier. | **1.** Only a single user is allowed to operate at a given time.<br>**2.** It lacks the essential algorithms and hence cannot provide simultaneous access for multiple users.<br>**3.** Utilizes less resources for data compared to relational database.<br>**4.** They are simple in structure and hence costs lesser compared to relational database. |

| | |
|---|---|
| **5.** Generally used for large scale purposes | **5.** Used for smaller projects or a subject which are specific for a particular purpose. |
| **6.** Manipulation of data is much simpler with the help of SQL query | **6.** Data modification and changes are difficult to be made. |
| **7.** It is more efficient due its property of allowing multiple user access. | **7.** It is less efficient. |
| **8.** Can handle large amounts of data. | **8.** Can handle small amount of data only. |
| **9.** It specifies relationship with the help of keys and indexes. | **9.** It does not specify the relationship between the data elements. |
| **10.** It is structured and consistent in terms of data processing. | **10.** It is inconsistent in processing the data |
| **11.** Data storing is in the form of tables. | **11.** Data storing is in the form of hierarchical structure. |
| **12.** The process of data fetching is much faster because of SQL. | **12.** The process of fetching data is slow especially in case of large volumes of data. |
| **13.** It supports client server architecture. | **13.** It does not support client server architecture. |

(Biscobing, 2021)

(Naeem, 2021)

# 4 Data modelling

Data modelling is the process of discovering and imposing structure on raw data. (Anon., 2020). The raw data has to be structured in an organized manner in order to build a database or data warehouse.

It identifies what data is required in order to create a database which is further used to create a functioning design for accessing and manipulating data.

There are three major types of data models used for creating a database.

- Conceptual data model layer
- Logical data model layer
- Physical data model layer

## 4.1 Conceptual data model

It provides a brief description of the required data excluding the methodology or procedure details. It helps in identifying and organizing data for the process requirements and possesses required entities and also provides the relationship among different entities. The attributes and primary key for the given database is not specified. It has the nature of that of a synopsis material and hence known as conceptual model. The conceptual model can be designed with the help of two methods which are Top-down and Bottom-up design (Anon., 2018)

For the creation of a typical database the data modelling starts with the conceptualization of the required output model.

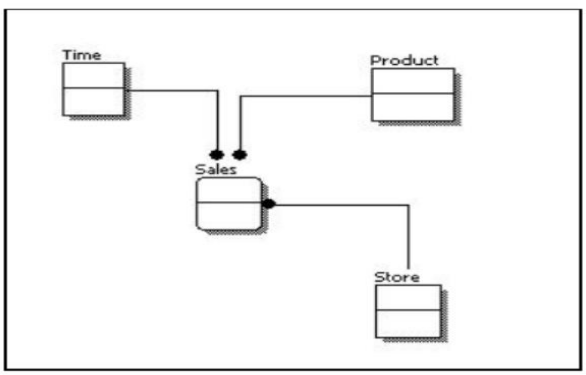



*Figure 25 Example for Conceptual data model*

Source; (Anon., 2019)

## 4.2 Logical data model

The logical data model provides a basic structure of the data information excluding the physical implementation factor. It possesses all entities and relationship among them. Unlike the conceptual model the attributes and the primary key are specified along with the referential integrity. It does not contain the information about the type of data. One of the most important process of normalization occurs during the design of logical data model. Normalization is used to organize data and hence making it more efficient by reducing the unwanted and repeated data. This helps in making the data model more efficient and reliable by the users. (Anon., 2018)

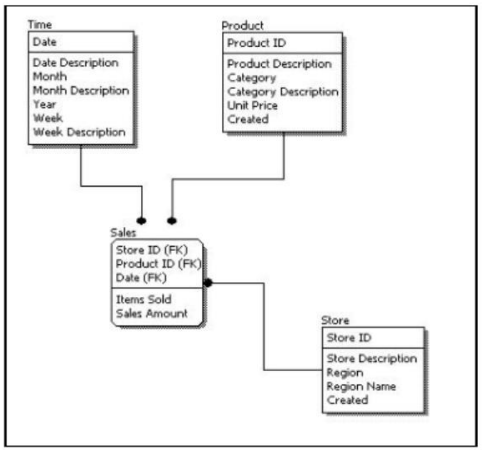



*Figure 26 Example for Logical data model*

Source; (Anon., 2019)

## 4.3 Physical data model

The final output or the presentable data for the end users is provided by the physical data model. It describes how the data is finally perceived at the end of the process. It provides a specification of all entities and attributes along with the primary key, foreign key, referential integrity etc. In this model the entities are converted into tables, the attributes are converted into columns and the relationships are converted into foreign keys. Unlike logical data model, the data

constraints, i.e. the information about the data type for each attribute is provided in the physical data model. (Anon., 2018)

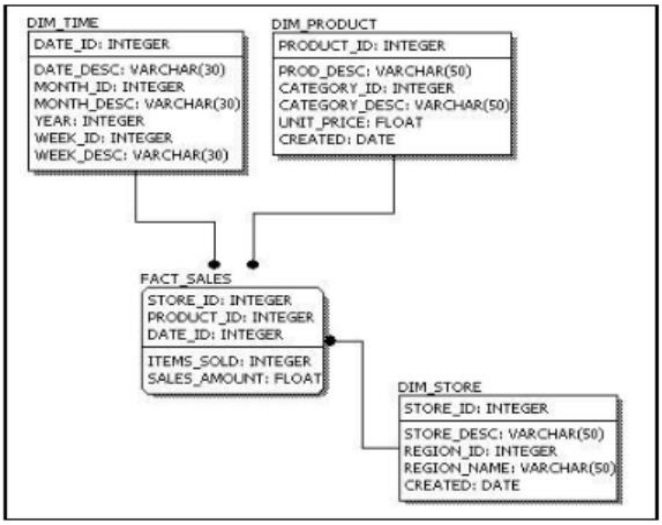



*Figure 27 Example for Physical data model*

Source; (Anon., 2019)

# 5 Data warehouse

It is a single data storage unit where innumerable data from many data sources are merged or joined together in order to assess and evaluate them for various business functions. Since it has to carry a huge amount of data, it has a property of having a large storage capacity, complexity and lengthy which makes it very easily liable to many errors and mistakes. Since business functions and processes are always on the verge of changing from time to time, the data warehouse is also required to change its functionality accordingly. This makes the data warehouse possess a dynamic characteristic where there can be change in the analytical functions as a continuous design process.

## 5.1 ETL process

ETL process is the Data Warehousing process where a large amount of data from innumerable sources are extracted and loaded into the data warehouse with the help of a 3 step process Extraction, Transformation and Loading

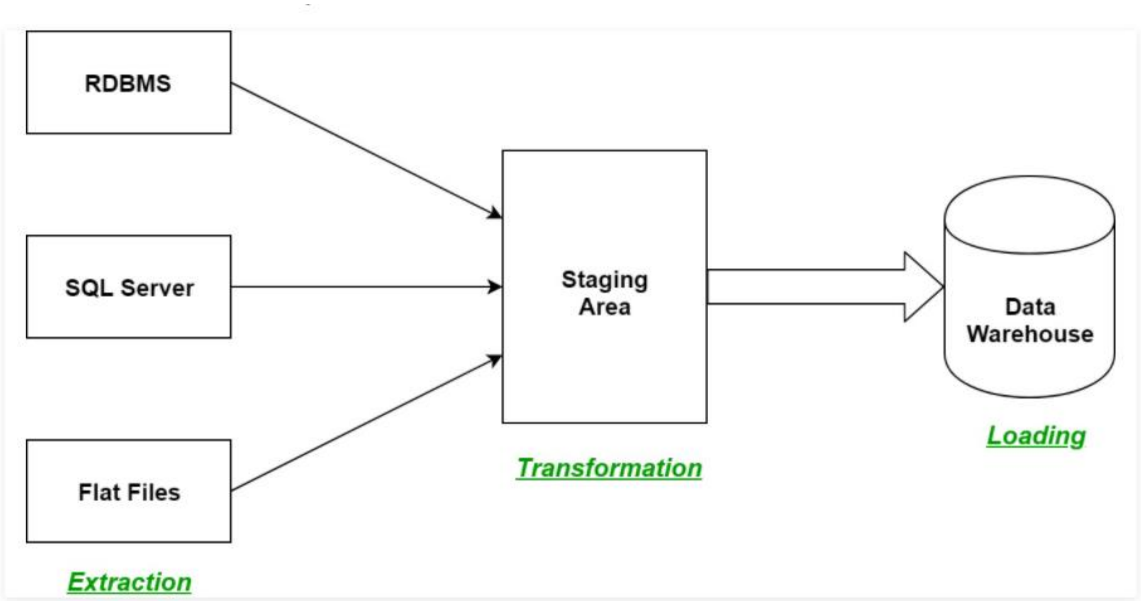



*Figure 28 ETL process*

*Source; (raman_257, 2019)*

### 5.1.1 Extract:

It is the initial step in the data warehouse management. All the data from innumerable sources are extracted for further processing. These data can be in different levels and formats. After extraction the data is transferred to the staging area, where the transformation and cleaning of the data occurs. The data is not loaded into the data warehouse directly as it is important to remove errors and standardize the formats before loading. It is one of the most time consuming processes. Since the data is changed many times, the extraction is also done repeatedly in order to have an updated data in the data warehouse.

### 5.1.2 Transform

The transformation process is a process which makes the large amount data suitable for loading into the Data Warehouse. It consists of many sub-processes which transforms the corrupt and unsuitable data into an appropriate format. Conversion and normalization process makes the all data into a homogenized format. It also sorts out and associates similar attributes which are from different sources, hence making it much easier to find and evaluate. Some of the other processes are that of filtering, cleaning, joining, splitting etc.

### 5.1.3 Load

Loading is the final stage of the ETL process. The data from the preceding process is loaded into the data warehouse for the end users. It is very important to perform this step in an efficient and cost effective way. The loading and updating of the data into the data warehouse is either done very frequently or for a longer period of time with regular intervals. The update is done without modifying or deleting the previously existing data.

(Wilfried Grossmann, 2015)

## 5.2 OLAP

Online analytical processing enables the company managers, business analysts and other decision making entities related to a firm to access information in an efficient and faster approach. It helps the user to visualize the data in a multidimensional view.

OLAP's multidimensional process helps in analysing complex data structures. It helps the users to perform Ad Hoc analysis, which is a single use process enabling the understanding of a complex multidimensional data set in a simpler and faster method.

OLAP has many characteristic features in terms of data information.

- It acts as an intermediate medium between data sources and end users of the system.
- OLAP processes are shared between different data sets, there is occurrence of storage in a layer separate from the sources.
- It distinguishes between zero valued elements and no value elements hence making the cumulative calculations more accurate and precise.
- It enables to perform elaborate calculations and comparisons between different datasets.
- The data representation in OLAP can be obtained in many ways like graphical charts, pictorial representation etc. making it more sophisticated and detailed.
- It can perform various operations like Roll-Up, Drill-Down, Slicing, Dicing and Pivoting which are used when performing query.

## OLAP OPERATIONS

The different OLAP operations helps in obtaining sophisticated information from the multi-dimensional data model regarding the required query analysis.

Some of the OLAP operations are:

**Roll-Up** operation cumulates data with data reduction. It acts like an aggregator on the OLAP cube. This is done by either going up the hierarchy of respective dimensions or by reducing the dimensions.

**Drill-Down** is the opposite of the roll-up operation. The data containing few details are converted into sophisticated form with more details. This process takes place by either going down the hierarchy of a respective dimension or by increasing the dimension.

**Dicing** is an operation where, with the help of two or more dimensions a sub-cube is selected from the actual OLAP cube.

**Slicing** is selection of a single dimension from the OLAP cube which results in a sub-cube.

**Pivot** is an operation where different attributes of the OLAP cube dataset are rotated and placed

in different axis as per requirement. It helps in calculating various query sections of the data which mainly involves swapping of certain rows and columns.
(Fahad, March 2016)

## 5.3 Hierarchies

A dimension is termed as a hierarchical dimension when it consists of a parent-daughter relationship. A basic hierarchy structure has one-to-many relation between the parent daughter structure. This means that a single parent node consists of many child nodes. In this model the data is stored in a tree like structure. The end users use the hierarchies in order to assess the fact data.

### 5.3.1 Hierarchy classification

- Balanced hierarchy – in a balanced hierarchy, every level contains equal number of members in the same level of hierarchy. It has a condition where all parent member has at least one child member and a child member is a part of exactly one parent member.

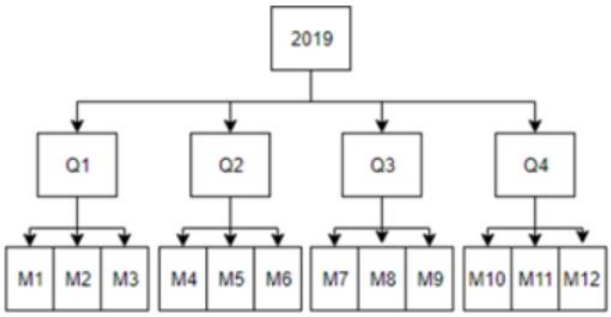



*Figure 29 Example for Balanced Hierarchy*

Source; (Anon., 2018)

- Unbalanced hierarchy – in an unbalanced hierarchy structure it's not an obligation for all parent members to have a child member. The members in the same level of the hierarchy have unequal number of sub members under them.

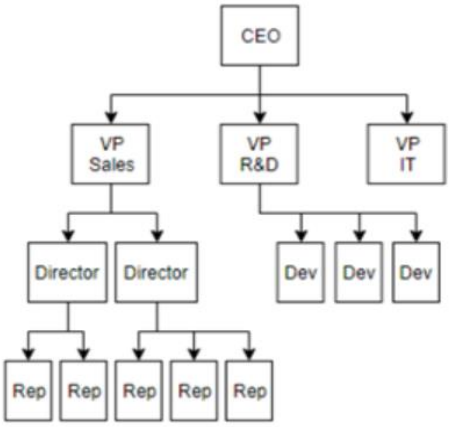



*Figure 30 Example for Unbalanced hierarchy*

Source; (Anon., 2018)

- Ragged hierarchy – in ragged hierarchy the parent member skips one level in the hierarchy. It shows that that some members of the structure are not linked to the intermediate level.

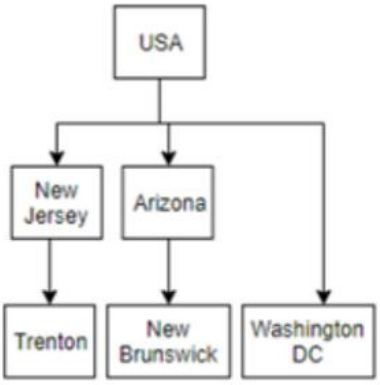



*Figure 31 Example for Ragged hierarchy*

Source; (Anon., 2018)

## 5.4 Schemas

A schema in data warehouse is a type of structure that expresses the elements of the data warehouse like the entities, entity types, relations between different tables, data description etc. The schemas in data warehouse consists of different design types like star, snowflake, and constellation schema. It can be defined as a logical data display structure providing details of the data warehouse records.

**Star schema** is a design where there is one fact table in the centre and a bunch of dimension tables. The fact table consists of foreign keys of its respective dimension tables and has all the primary information in data warehouse. The fact tables are not normalized in this structure. Every primary key in the surrounding dimension table is related to the foreign key in the fact table. In a star schema every dimension is represented by only one-dimensional table. It consists of simpler queried making it easier to measure and evaluate.

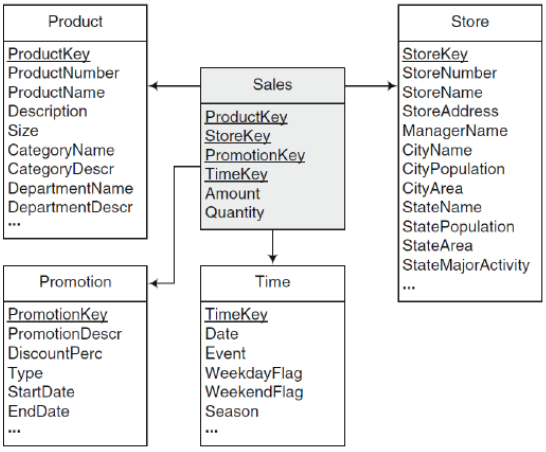



*Figure 32 Example for star schema*

Source; (Alejandro Vaisman, 2014)

**Snowflake schema** is a structure which is a modified version of a star schema. Other than the dimensions present in the star schema there will be additional dimensions added. This schema is known as snowflake schema. In the snowflake schema only the dimensional tables are changed, but the fact tables remain the same. The changes made to the dimensional table are that the tables are in normalized form. This structure more detailed when compared to the star schema. One of the major difference between the star and snowflake schema is that there is reduction in redundancy, making it more efficient in terms of functionality and storage.

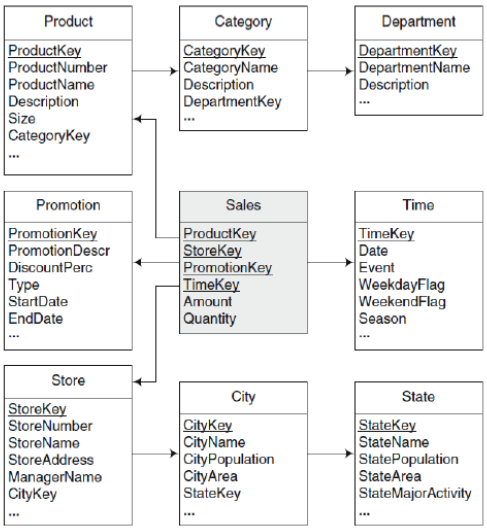



*Figure 33 Example of Snowflake schema*

Source; (Alejandro Vaisman, 2014)

# PRACTICAL SECTION

# 6 Company overview

**Bharat Heavy Electricals Limited (BHEL)** is an Indian engineering and manufacturing company headquartered in Delhi, India established in 1964. It is a power generation equipment manufacturer. The major equipment supplied by BHEL are turbines and boilers. It also supplies electric locomotives to Indian Railways. The following thesis is based on the employee information data acquired from one of the many divisions of the company. The present system uses only MS Excel platform for collecting and recording data of the employees. This data is also used to calculate the payroll of the employees. The analysis of this large volume data is made more efficient by implementing data warehouse and relational database management system. There is also implementation of queries with the help of which, the company will be able to track employee performance in an efficient and time saving manner.

# 7 Analysis of acquired data

The data acquired for processing the data model is a sample data, representing the attendance and salary details of the employees for the month of January 2020. These data refer to the employees of the finance department in the company. This particular department consists of around 460 employees. The datasets are divided into 4 tables consisting of various details concerning the employees.

**Personal** table consists of many personal details of an employee like a unique identification number, name, years of experience, grade, department etc. In the database it acts as one of the fact tables. It consists of many foreign keys. This table consists of the final data produced where along with other details, the final salary details for one month is calculated and displayed.

**Daily Punch** table consists of daily entry details of the employees. It acts as another fact table in the structure. It consists of foreign keys from time and shift dimensional table. It also contains total working hours of every employee for a month.

**Attendance summary** consists of details about the number of days the employee was present, the types of leaves accessed by the employee etc.

**Shift dimension table** consists of shift details according to its unique code which is the primary key. These primary keys have relationship with the foreign keys in the daily punch table. For example, AF stands for general shift which is between 8:00 – 16:00.

# 8 Data model

## 8.1 Conceptual model

The conceptual data model is provided below. It depicts an entity relationship diagram. The model consists of fact tables and dimensional tables with respective relationships between each other. The model is made for the use of the staff member responsible for managing the salary details. In this particular data model there are two fact tables – 'Personal' and 'Daily punch'. It is necessary to note that the two fact tables are not to be connected according to entity relationship diagram rules. These fact tables are connected to various dimensional tables like shift, time etc. The dimensional tables are connected to the fact table mostly in a one-to-many type of relationship. It possesses a star schema as its structure. This model is created as a basic concept for the required data modelling.

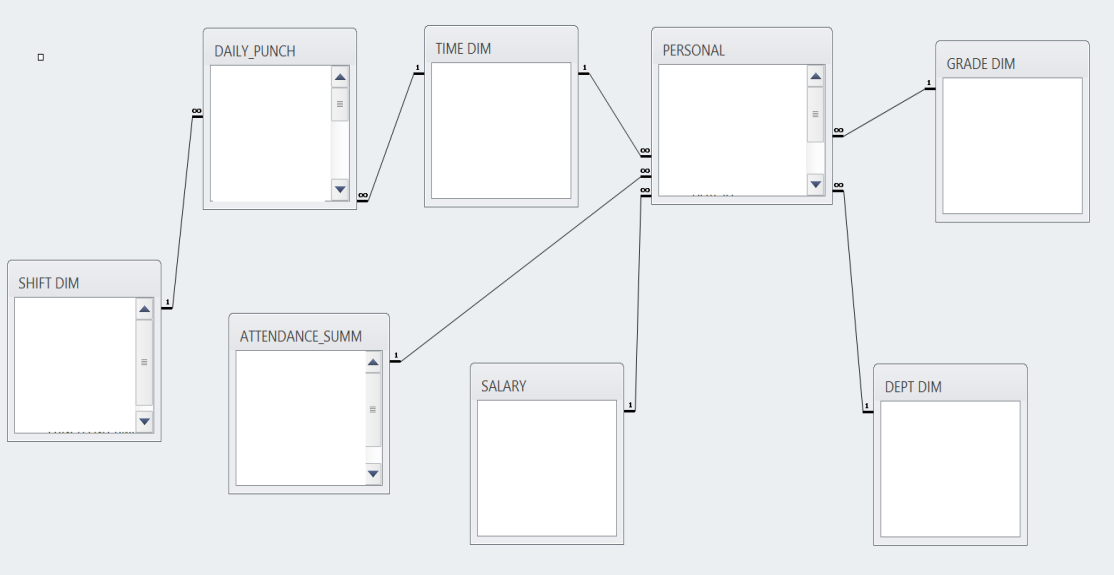



*Figure 34 CONCEPTUAL MODEL FOR THE IMPLEMENTED DATA MODEL*

Source; Own Contribution

## 8.2 Logical model

The intermediate step of the conceptual model is the logical model. This adds further, the required information for the data model. It forms a basic structure for the physical data model which is the next step to be performed. Unlike the conceptual model, the attributes are mentioned in the logical model. It two fact tables 'Daily punch' and 'Personal' consists of many attributes. The 'Daily punch' table has 3 punch in and punch out time attributes, shift

code, daily working hours, personal and company pass etc. The shift code acts as a foreign key. The 'Personal' table consists of personal information of the employees like name, employee code, monthly salary details etc. It also includes foreign keys like 'Grade ID', 'Attendance ID', Department ID etc.

The dimension tables attributes are also provided. The time dimension has a hierarchical structure consisting of year, quarter, month and days. The grade dimension consists of a hierarchical structure having grade name and category. The salary table consists of salary ID and Earning rate of every employee. Every dimension table is connected to the fact tables hence forming a star schema pattern. The primary keys of the dimension table act as a foreign key in the fact table hence linking both the tables. The logical data model also characterizes the primary key and the foreign key. The primary key of every table is denoted by a 'primary key' along with its attribute name.

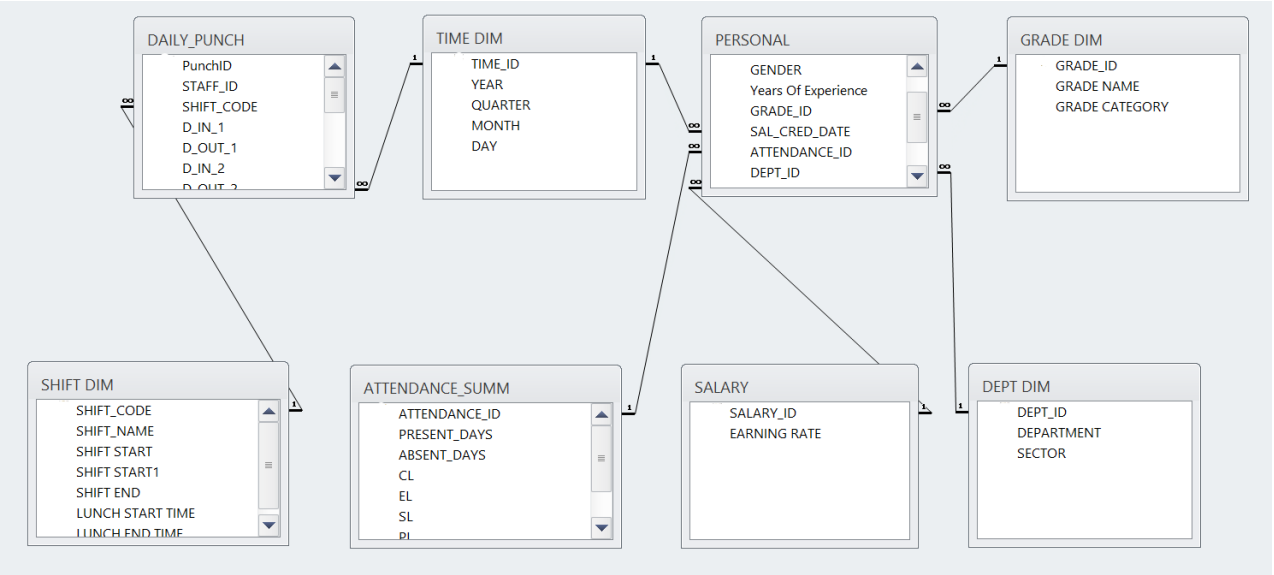



*Figure 35 LOGICAL MODEL FOR THE IMPLEMENTED DATA MODEL*

Source; Own Contribution

## 8.3 Physical model

The physical data model is the intermediate process of the logical data model. It shows the final representation of the model perceived by the end users. The physical data model acts as the final stage of data modelling. Along with the entities and attributes, relationships characterization of primary and foreign key the physical data model shows the data types of

every attribute present in every table of the data model. The data requirements of the user are cumulatively amalgamated in the physical model.

The data lengths are also assigned along with the data types. In this data model some of the data types and length assigned are name of the employee – 'variable character 255', number of days present and absent – 'double integer' etc. This model is then transformed into a database using MS Access which is a relational database management system.

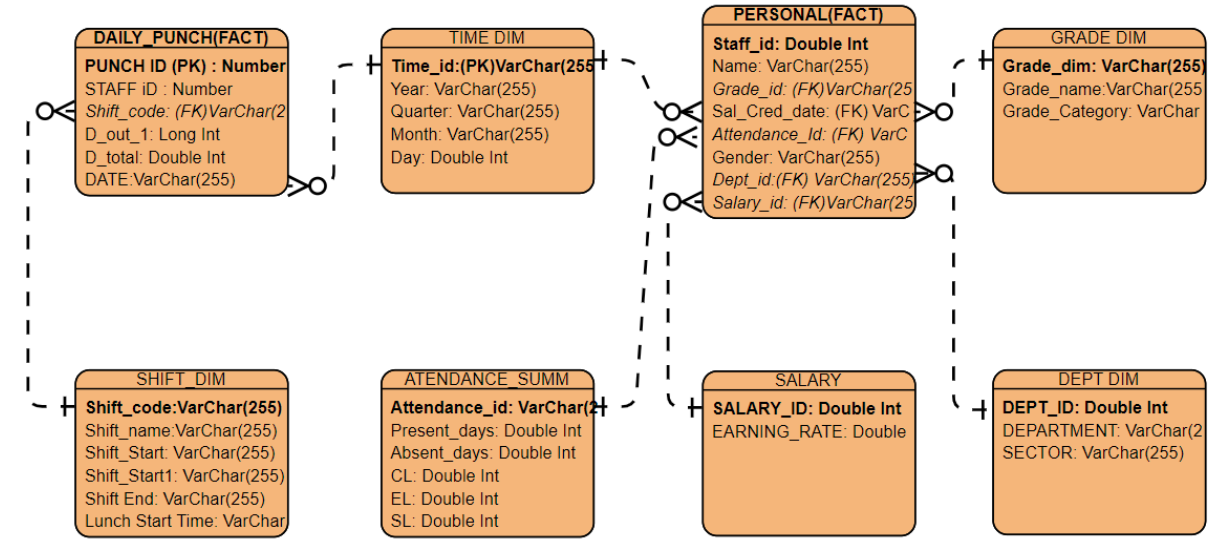



*Figure 36 PHYSICAL MODEL FOR THE IMPLEMENTED DATA MODEL*

Source; Own Contribution

# 9 Queries executed

## 9.1 Average salary calculated based on different Grades of the employee.

The query shows the average salary and average working hours of employees based on different grades. The query is carried out in MS Access. The query then is exported to Excel subsequently undergoing OLAP analysis for calculating the average salary according to different grades. The reason we use two different programmes is because MS Access helps in manipulating the data according to requirements and Excel comes into picture for implementing Pivot analysis for the same requirements.

This query enables the company to find out the range of salaries earned according to an employee's respective grade. It helps the firm to analyse and compare different grades and hence map out an efficient path of growth for an employee. They can also set up corresponding salaries based on the obtained output. One of the main intention to calculate or create the query is to bring about an organized pay structure and easy retrieval of required data. Usually when there are huge amounts of data present, it is difficult to obtain the required result. Hence to make it efficient and cost effective the companies should use appropriate queries in a database management system.

| NAME | STAFF_ID | GRADE_ID | GRADE NAME | GRADE CATEGORY | TOTAL EARNING |
|---|---|---|---|---|---|
| Debasish Mandal | 1285718 | E6 | Deputy General. Manager | EXECUTIVE | 115025.08 |
| S Arunachalam | 2175460 | E6 | Deputy General. Manager | EXECUTIVE | 114855.83 |
| N Anandarao | 3771547 | E6 | Deputy General. Manager | EXECUTIVE | 124426.56 |
| K S Ravishankar | 3818454 | E6 | Deputy General. Manager | EXECUTIVE | 131094.62 |
| S Seshapathi | 3818470 | E6 | Deputy General. Manager | EXECUTIVE | 109908.94 |
| D Chandrashekar | 3818500 | E6 | Deputy General. Manager | EXECUTIVE | 101125.50 |
| Seetharama Hegde | 3819094 | E6 | Deputy General. Manager | EXECUTIVE | 129132.90 |
| L Prabhushankar | 3819140 | E6 | Deputy General. Manager | EXECUTIVE | 107037.08 |
| S John Felix Anto | 3819531 | E6 | Deputy General. Manager | EXECUTIVE | 103160.58 |
| A Ramesh Babu | 3820114 | E6 | Deputy General. Manager | EXECUTIVE | 126808.30 |
| Manish Kumar Pandey | 3825116 | E6 | Deputy General. Manager | EXECUTIVE | 112049.55 |
| James Joy Mathew | 3830055 | E6 | Deputy General. Manager | EXECUTIVE | 76459.93 |
| Rabindra Kumar Verma | 4257081 | E6 | Deputy General. Manager | EXECUTIVE | 115202.01 |

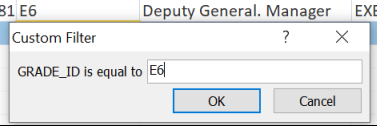



*Figure 37 Query design for the required analysis in MS Access*

Source; Own Contribution

```
SELECT PERSONAL.NAME, PERSONAL.STAFF_ID,
[GRADE DIM].GRADE_ID, [GRADE DIM].[GRADE NAME],
 [GRADE DIM].[GRADE CATEGORY], PERSONAL.[TOTAL EARNING]

FROM [GRADE DIM] INNER JOIN PERSONAL ON [GRADE DIM].GRADE_ID = PERSONAL.GRADE_ID;
```

*Figure 38 SQL code for the performed query*

Source; Own Contribution

The salary division according to different grades helps the company to set the range of salaries for every grade in an organized manner. The average salary of every grade is taken as the midpoint range of salary for that particular grade. For example, the E6 executive grade employees have an average salary earning of 1,12,791.298₹. This value is noted as the midpoint salary reference for the E6 Executive grade employees.

| Grade | Average of TOTAL EARNING |
|---|---|
| **⊞ ARTISANS - Non Supervisory (Technical)** | **45534.52919** |
| **⊟ EXECUTIVE** | **79640.95797** |
| **⊞ Addl. General Manager** | **152489.9538** |
| **⊞ Sr. Deputy Gen. Manager** | **115349.9394** |
| **⊞ Sr. Engineer/Officer / Sr. Executive** | **57380.09779** |
| **⊟ Deputy General. Manager** | **112791.2982** |
| ⊟ E6 | 112791.2982 |
| A Ramesh Babu | 126808.3038 |
| D Chandrashekar | 101125.5048 |
| Debasish Mandal | 115025.0846 |
| James Joy Mathew | 76459.92981 |
| K S Ravishankar | 131094.6154 |
| L Prabhushankar | 107037.075 |
| Manish Kumar Pandey | 112049.55 |
| N Anandarao | 124426.5558 |
| Rabindra Kumar Verma | 115202.0106 |
| S Arunachalam | 114855.8322 |
| S John Felix Anto | 103160.575 |
| S Seshapathi | 109908.9394 |
| Seetharama Hegde | 129132.9 |
| **⊟ Deputy Manager** | **62645.3084** |

*Figure 39 Pivot analysis for average salary of employees in different grades*

Source; Own Contribution

| NAME | STAFF_ID | GRADE_ID | GRADE NAME | GRADE CATEGORY | TOTAL EARNING |
|---|---|---|---|---|---|
| Debasish Mandal | 1285718 | E6 | Deputy General. Manager | EXECUTIVE | 115025.0846 |
| S Arunachalam | 2175460 | E6 | Deputy General. Manager | EXECUTIVE | 114855.8322 |
| N Anandarao | 3771547 | E6 | Deputy General. Manager | EXECUTIVE | 124426.5558 |
| K S Ravishankar | 3818454 | E6 | Deputy General. Manager | EXECUTIVE | 131094.6154 |
| S Seshapathi | 3818470 | E6 | Deputy General. Manager | EXECUTIVE | 109908.9394 |
| D Chandrashekar | 3818500 | E6 | Deputy General. Manager | EXECUTIVE | 101125.5048 |
| Seetharama Hegde | 3819094 | E6 | Deputy General. Manager | EXECUTIVE | 129132.9 |
| L Prabhushankar | 3819140 | E6 | Deputy General. Manager | EXECUTIVE | 107037.075 |
| S John Felix Anto | 3819531 | E6 | Deputy General. Manager | EXECUTIVE | 103160.575 |
| A Ramesh Babu | 3820114 | E6 | Deputy General. Manager | EXECUTIVE | 126808.3038 |
| Manish Kumar Pandey | 3825116 | E6 | Deputy General. Manager | EXECUTIVE | 112049.55 |
| James Joy Mathew | 3830055 | E6 | Deputy General. Manager | EXECUTIVE | 76459.92981 |
| Rabindra Kumar Verma | 4257081 | E6 | Deputy General. Manager | EXECUTIVE | 115202.0106 |
| | | | | | 112791.2982 |

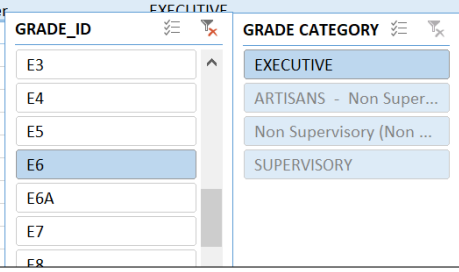



*Figure 40 Slicer operation of the performed query in excel*

Source; Own Contribution

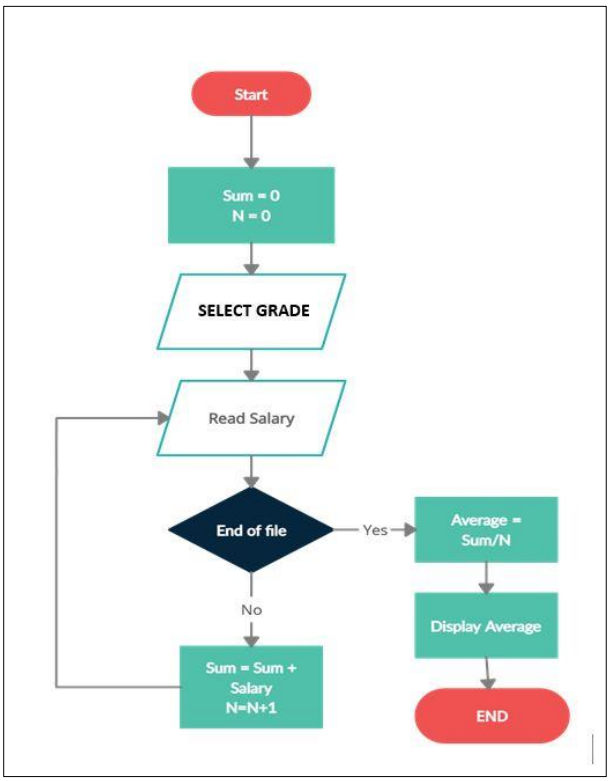
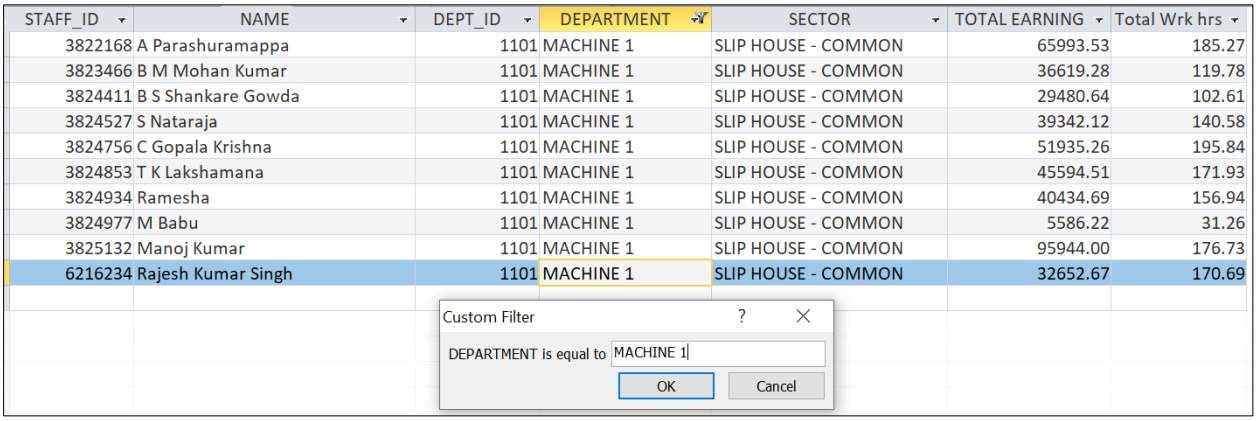



*Figure 41 9.2 Average salary flow chart based on different Grades of the employee.*

Source; Own Contribution

The flow chart explains the process of the given query. The flow of events occurs starting from selecting the grade that the user is seeking. The slicer program helps is filtering the selected grade and shows only data of those grades. Further the average salary is known with the simple arithmetic average function of excel and the required actions are taken.

## 9.2 Average salary and working hours calculated based on different departments of the company.



| STAFF_ID | NAME | DEPT_ID | DEPARTMENT | SECTOR | TOTAL EARNING | Total Wrk hrs |
|---|---|---|---|---|---|---|
| 3822168 | A Parashuramappa | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 65993.53 | 185.27 |
| 3823466 | B M Mohan Kumar | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 36619.28 | 119.78 |
| 3824411 | B S Shankare Gowda | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 29480.64 | 102.61 |
| 3824527 | S Nataraja | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 39342.12 | 140.58 |
| 3824756 | C Gopala Krishna | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 51935.26 | 195.84 |
| 3824853 | T K Lakshamana | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 45594.51 | 171.93 |
| 3824934 | Ramesha | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 40434.69 | 156.94 |
| 3824977 | M Babu | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 5586.22 | 31.26 |
| 3825132 | Manoj Kumar | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 95944.00 | 176.73 |
| 6216234 | Rajesh Kumar Singh | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 32652.67 | 170.69 |



*Figure 42 Query design for the required analysis in MS Access*

```
SELECT PERSONAL.STAFF_ID, PERSONAL.NAME, [DEPT DIM].
DEPT_ID, [DEPT DIM].DEPARTMENT, [DEPT DIM].SECTOR,
PERSONAL.[TOTAL EARNING], PERSONAL.[Total Wrk hrs]

FROM [DEPT DIM] INNER JOIN PERSONAL ON [DEPT DIM].DEPT_ID = PERSONAL.DEPT_ID;
```

*Figure 43 SQL code for the performed query*

This query helps in finding the average salary and average working hours of the employees based on different departments of the company. It is the same as the grade division query. The only difference is that, instead of divisions in grades, the query displays the average salaries of different divisions or departments in the company. This query helps in internal assessment and continuous improvement of the company. Based on the results there are different level of job analysis which can be done.

| Departments | Average of TOTAL EARNING | Average of Total Wrk hrs |
|---|---|---|
| ⊞CERAMIC | 64572.11315 | 160.7766071 |
| ⊞COMMERCIAL | 66233.14223 | 151.7388889 |
| ⊞ENGINEERING | 60584.55616 | 146.7845455 |
| ⊞GM OFFICE | 0 | 0 |
| ⊟MACHINE 1 | 44358.2926 | 145.163 |
| ⊟SLIP HOUSE - COMMON | 44358.2926 | 145.163 |
| A Parashuramappa | 65993.53029 | 185.27 |
| B M Mohan Kumar | 36619.27981 | 119.78 |
| B S Shankare Gowda | 29480.64231 | 102.61 |
| C Gopala Krishna | 51935.26154 | 195.84 |
| M Babu | 5586.222115 | 31.26 |
| Manoj Kumar | 95943.99808 | 176.73 |
| Rajesh Kumar Singh | 32652.66875 | 170.69 |
| Ramesha | 40434.68558 | 156.94 |
| S Nataraja | 39342.12404 | 140.58 |
| T K Lakshamana | 45594.51346 | 171.93 |
| ⊞MACHINE 2 | 50796.56879 | 146.2245714 |
| ⊞MAINTENANCE | 50925.28956 | 154.8882353 |
| ⊞MARKETING | 69370.87479 | 135.3655556 |
| ⊞MM | 68826.91487 | 162.3509091 |
| ⊞P & A | 58210.35668 | 147.0074074 |

*Figure 44 Pivot analysis for average salary and average working hours of employees in different departments*

Source; Own Contribution

| STAFF_ID | NAME | GENDER | SAL_CRED_DATE | DEPT ID | DEPARTMENT | SECTOR | TOTAL EARNING | Total Wrk hrs |
|---|---|---|---|---|---|---|---|---|
| 3822168 | A Parashuramappa | M | 20T31 | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 65993.53029 | 185.27 |
| 3823466 | B M Mohan Kumar | M | 20T31 | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 36619.27981 | 119.78 |
| 3824411 | B S Shankare Gowda | M | 20T31 | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 29480.64231 | 102.61 |
| 3824527 | S Nataraja | M | 20T31 | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 39342.12404 | 140.58 |
| 3824756 | C Gopala Krishna | M | 20T31 | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 51935.26154 | 195.84 |
| 3824853 | T K Lakshamana | M | 20T31 | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 45594.51346 | 171.93 |
| 3824934 | Ramesha | M | 20T31 | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 40434.68558 | 156.94 |
| 3824977 | M Babu | M | 20T31 | 1101 | MACHINE 1 | SLIP HOUSE - COMMON | 5586.222115 | 31.26 |
| 3825132 | Manoj Kumar | M | | | CHINE 1 | SLIP HOUSE - COMMON | 95943.99808 | 176.73 |
| 6216234 | Rajesh Kumar Singh | M | | | CHINE 1 | SLIP HOUSE - COMMON | 32652.66875 | 170.69 |
| | | | | | | | 44358.2926 | 145.163 |

| DEPARTM... | DEPT ID |
|---|---|
| CERAMIC | 1101 |
| COMMERCIAL | 100 |
| ENGINEERING | 200 |
| GM OFFICE | 400 |
| MACHINE 1 | 901 |
| MACHINE 2 | 1201 |
| MAINTENAN... | 1202 |

*Figure 45 Slicer operation for average salary and working hours calculated based on different departments of the company*

Source; Own Contribution

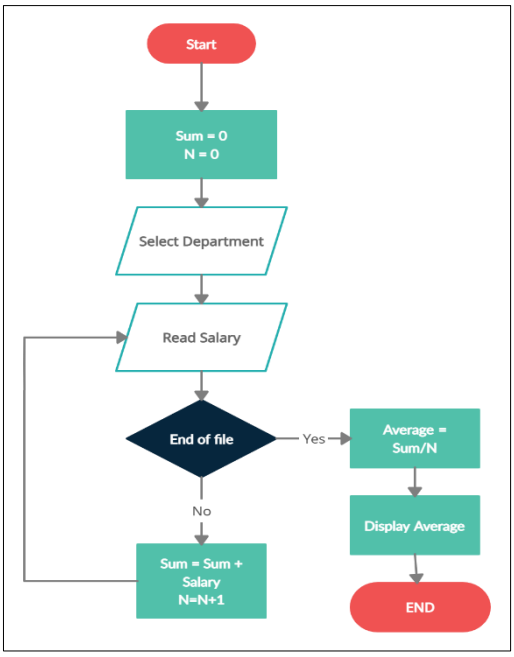



*Figure 46 Flow chart for calculating average salary according to required departments*

Source; Own Contribution

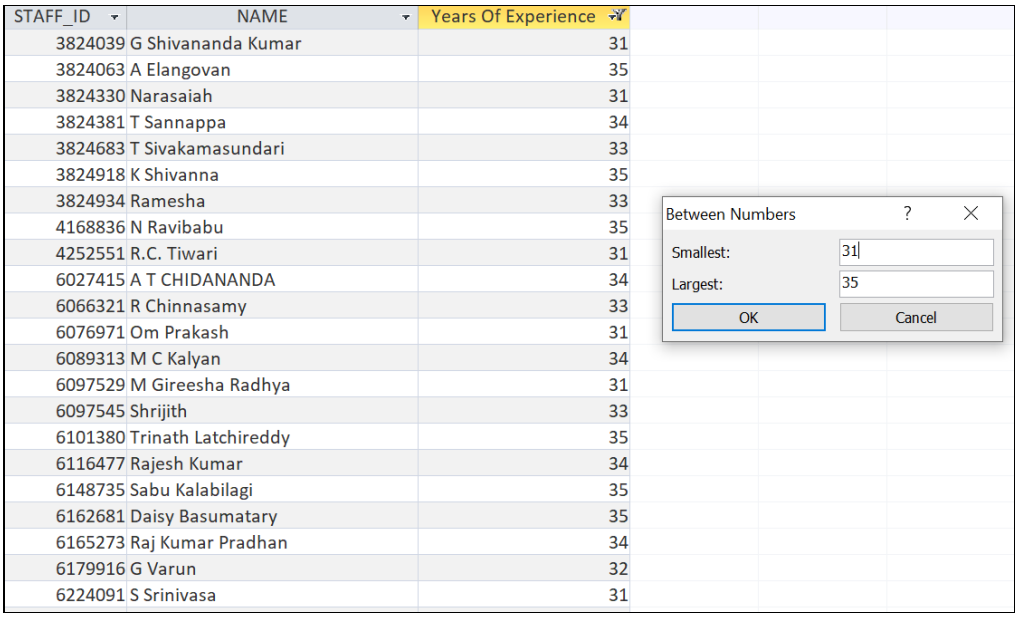
The given flow chart provides a process flow of the query. It is very similar to that of the previous query flow chart. The process starts with the selection of the required department in the slicing programme. This results in the display of the average salary of the corresponding department employees. For example, the average salary of the employees in the machine maintenance department is found to be 44,358.2926₹.

## 9.3 Query on the employees having required range of years of experience in the company

This query helps to find the employees having a required range of years of experience. Like the previous queries, this query is also helpful for internal assessment. It reflects more on the employee management system of the company. The long term experience of an employee in the company shows the level of conducive environment provided by the company. For a new employee there is more training to be provided by the company. But for an experienced employee there is no need to utilize much energy and time as they are already trained. The long

term employees can supervise the new employees and reassure them the confidence in them. The number of highly experienced employees acts as a measure of stability, cost effectiveness, etc.

The statistics show that the employee turnover rate is high for young employees. It means that the youngsters hop from one company to another within a few years of work in a company. This occurs due to both negative and expected normal reasons. The company with the help of this query can realize the employee turnover rate of youngsters and try to change their process or employee relationships. Doing this will help them retain these employees hence maintaining their company processes to flow without any bottlenecks or restrictions.



| STAFF_ID | NAME | Years Of Experience |
|---|---|---|
| 3824039 | G Shivananda Kumar | 31 |
| 3824063 | A Elangovan | 35 |
| 3824330 | Narasaiah | 31 |
| 3824381 | T Sannappa | 34 |
| 3824683 | T Sivakamasundari | 33 |
| 3824918 | K Shivanna | 35 |
| 3824934 | Ramesha | 33 |
| 4168836 | N Ravibabu | 35 |
| 4252551 | R.C. Tiwari | 31 |
| 6027415 | A T CHIDANANDA | 34 |
| 6066321 | R Chinnasamy | 33 |
| 6076971 | Om Prakash | 31 |
| 6089313 | M C Kalyan | 34 |
| 6097529 | M Gireesha Radhya | 31 |
| 6097545 | Shrijith | 33 |
| 6101380 | Trinath Latchireddy | 35 |
| 6116477 | Rajesh Kumar | 34 |
| 6148735 | Sabu Kalabilagi | 35 |
| 6162681 | Daisy Basumatary | 35 |
| 6165273 | Raj Kumar Pradhan | 34 |
| 6179916 | G Varun | 32 |
| 6224091 | S Srinivasa | 31 |



*Figure 47 Query on the employees having required range of years of experience in the company*

Source; Own Contribution

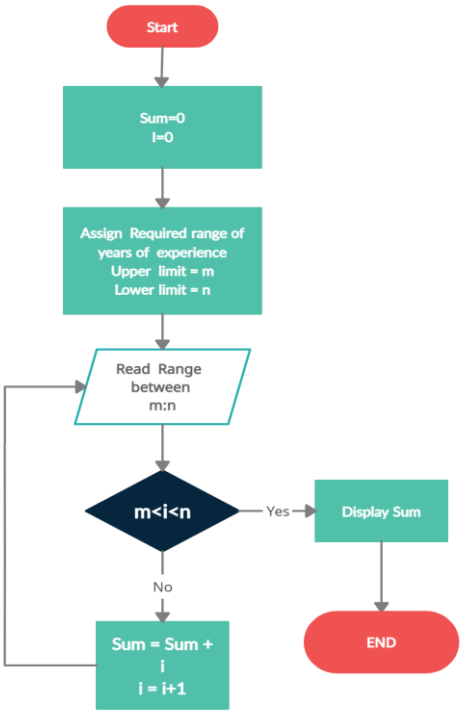



*Figure 48 Flow chart for the employees have required range of years of experience*

Source; Own Contribution

The above flowchart shows the flow of process to find out the employees having a required range of years of experience. It starts of by assigning the value of the upper range of the years of experience and lower range of years of experience. The employees having the value between this range are read and displayed as the output.

## 9.4 Query on number of personal passes by an employee in a month for performance evaluation.

This query shows the number of personal or unofficial passes applied by the employees in a month. In the company there exists two types of passes – '**company/official pass**' and '**personal/unofficial pass**'. This implies that an employee can exit the office or company premises either for official reasons or personal reasons. If the employee exits with an official pass, it is not accounted as an absence or any other factor. But if an employee exits with a personal pass, it is accounted as a negative in the report of the employee performance. This query checks for employees having personal pass count greater than ten in a month and sends them a warning message so as to be careful about the number of mid-day leaves applied by the

employee. Hence it helps the management of the company to have a strict employee performance system.

| Name | STAFF_ID. | Count of Personal_pass |
|---|---|---|
| T C Mohan | 3821323 | 2 |
| D Chandrashekar | 3818500 | 2 |
| T Krishna | 3819183 | 2 |
| B N Devaraju | 3820157 | 3 |
| Duraippandi P | 6144209 | 3 |
| M S Eachalakaranji | 3819582 | 3 |
| N Anandarao | 3771547 | 3 |
| M Babu | 3824977 | 4 |
| Vinay Kumar | 3787796 | 4 |
| Raj Kumar Pradhan | 6165273 | 4 |
| S Seshapathi | 3818470 | 4 |
| K Suresh Bedre | 3820866 | 4 |
| P Manjappa | 3820696 | 5 |
| M Obulesh | 3822842 | 6 |
| J R L Setty | 3820874 | 6 |
| M Saroja | 3825191 | 7 |
| A Sekar | 3820564 | 7 |
| K Selvan | 3822060 | 7 |
| Rachita Garg | 6078710 | 7 |

*Figure 49 Query on number of personal passes by an employee in a month*

Source; Own Contribution

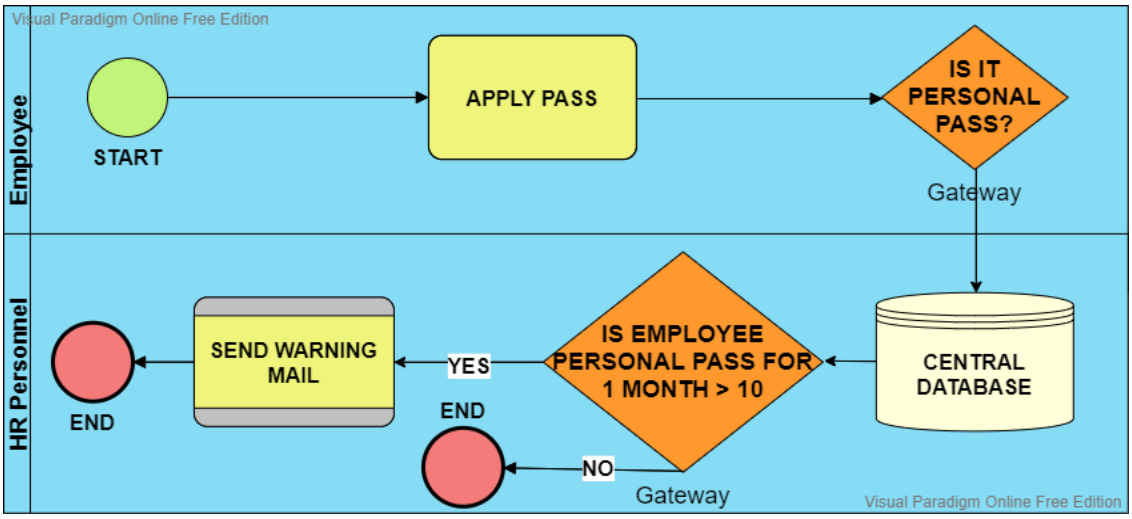



*Figure 50 Business process diagram for personal pass applied by the employee*

Source; Own Contribution

The above business process diagram shows the process of sending a warning mail to the employees having a record of personal pass greater than ten in a month. The applied pass is recorded in the central database through an internal software. This record is then used in the query for counting the personal passes and hence through the obtained result, the employee achieving the condition is sent a warning mail.

# 10 Discussions of results

This thesis mainly aimed at optimizing the process of recording data in a company with the help of business intelligence. This is mainly done with by the amalgamation of MS Access and MS Excel programs, Simplification of data tables, enabling queries etc. which will be explained briefly:

## 10.1 REPORT 1 - Integration of MS Access and MS excel.

This process is one of the main work of the thesis. The integration of MS Access and Excel creates an efficient programme for recording and analysing large amount of data. Excel and Access have their own disadvantages and advantages separately. The usage of both the programmes depend on the aim of the project. Since the integration of the both was the best method for this project, corresponding tasks have been performed.

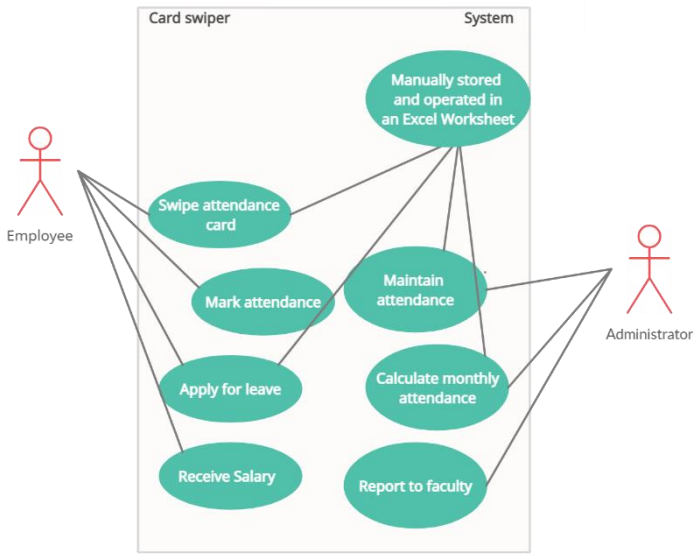



*Figure 51 Use case model of the previous system for employee records*

Source; Own Contribution

Some of the advantages of excel are that it can store enormous amounts of data, has high flexibility, can analyse and calculate simple queries, has pivoting function for manipulating the

data rows and columns etc. However, Excel has its limitations. The flexibility and easy calculation and analysis function of excel becomes more complicated and difficult to perform as the work progresses to higher level of complexity. Data manipulation becomes challenging with the gradual increase in data volume. The errors and mistakes caused, ultimately produces unnecessary or undesirable outputs which in turn leads to bad analysis.

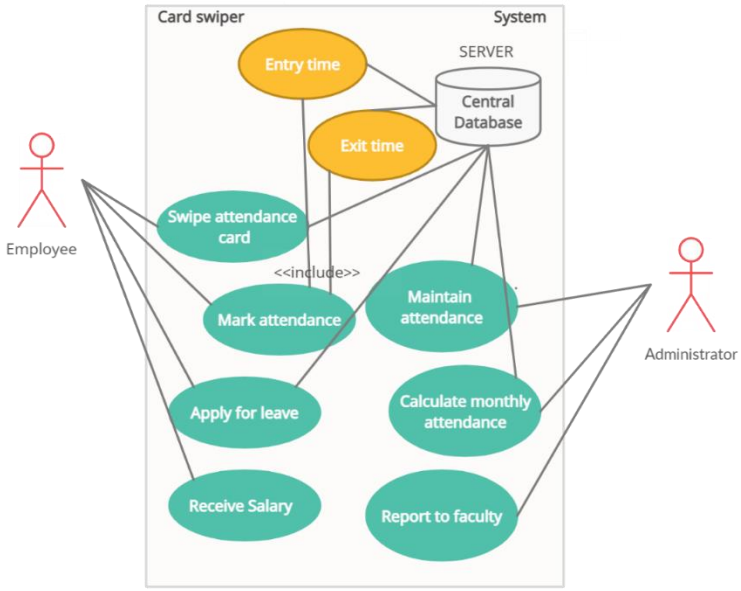



*Figure 52 Use case model of the optimized system for employee records*

Source; Own Contribution

Excel acts like a flat file database where simple database can be stored unlike Access which is a relational database storing multiple tables. Access is an easy database application used for collecting and bringing together complex data files of an organization which are usually stored in excel. It can store data in the form of tables creating corresponding relations, queries, etc. This project takes advantage of both the programmes, hence optimizing the process as better as possible. Initially the data in excel is imported into access file and accessed in the 'datasheet view' form. The advantage is that the data is copied in the same format as the initial data type, making no changes and hence reducing the error of file transfer from one application to another. The excel worksheets are linked to the access database file. This helps in the performance of query and reporting. The query forms were designed without manually typing any SQL codes which makes it very effective and easier.

One more advantage is that the limitation of difficulty of 'access' data manipulation is terminated as the linking makes it possible to edit data in access by changing it in the corresponding excel worksheet. The sorting and filtering functions of access is very similar to access and hence there is no need of learning a new programme all over again.

## 10.2 REPORT 2 – Data analysis efficiency

For the current project, it was realized that access is much better to manage data. It mainly provides the user with easy sorting of complex data, filtering and organization. Excel is an option used for storing a single table with multiple data columns for easy calculation and analysis. But since the employee information records contain multiple tables and complex data structures, it is necessary to use the relational database management system in the form of MS Access. The main intention of converting the single flat file into relational database is to make the data analysis and calculation much more efficient and easier.

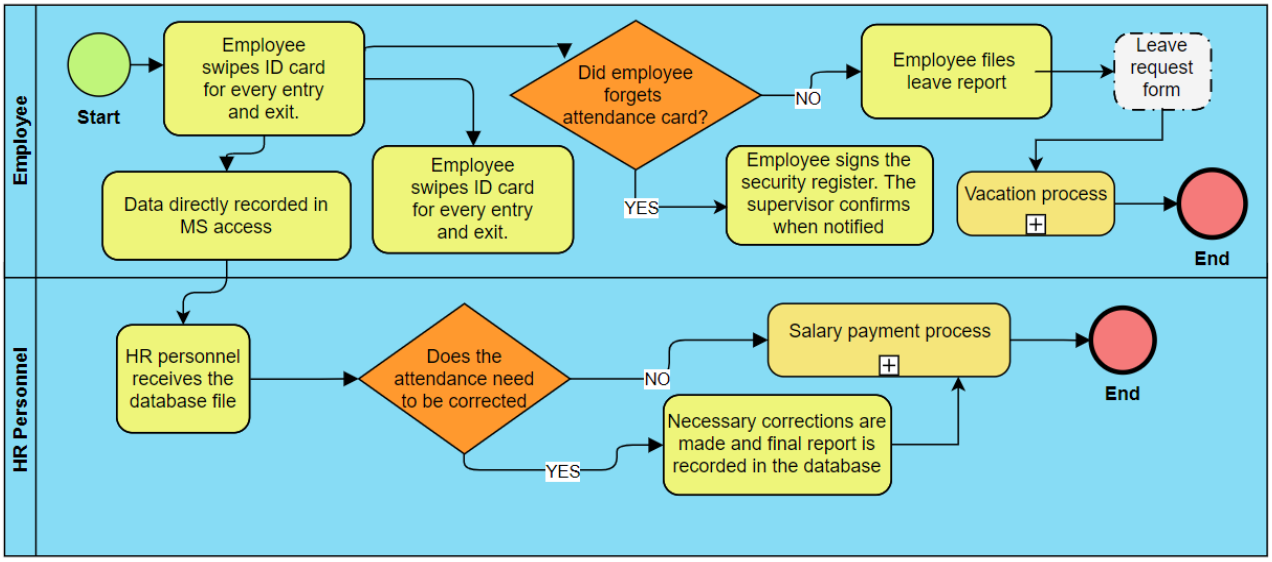



*Figure 53 Business process diagram of the optimized system for employee records*

Source; Own Contribution

In this structure the flat file is converted into various multiple tables thus ensuring that every data of an employee is stored in a separate table. Multiple data tables provide an efficient structure with no repetitive attributes. Due to data integrity, the change in data of one table leads to change in all the related tables. The data integrity constraints applied in the database structure ensures that there are relevant data entered in relevant data type field. For example, date format should be entered in the date type field or attribute in the table, which if not followed displays

the error message. It also displays an error message if there are any duplicate or repeated value in the primary key attribute. This feature ensures that there are no errors while typing or manipulating the corresponding field. It also provides a differentiation in user access to the database file according to company policies hence ensuring security.

## 10.3 REPORT 3 – Implementation of queries.

In this project, both access and excel are used to perform various queries. Access and excel are linked together and the required queries are employed. Queries enable the user to retrieve huge amounts of data and report the required output in a simple and efficient way. This process helps to combine different tables to obtain the relevant information as per user requirement. Excel projects the query output with charts and other sophisticated analysis tools. There is also no need of writing a special code for the queries hence making it more process and time efficient. The pivot table tool is used to explore various data queries of the employee like ‘average salary and working hours of the employees based on different department and grade’, ‘employees having the required years of experience range’ and ‘number of personal passes applied by an employee in a month’.

These queries mainly help in the internal management of the company and hence indirectly affecting their financial position as an organization. It enables the company to evaluate and control the employee performance. This analysis is used by the company to work on their limitations and improve their efficiency. It also provides a strong base for both the managers and the employees to take optimal decisions. (Serbanescu Luminita, 2009)

# 11 Conclusion and suggestion for future work

This project is executed with the aim of optimizing the process of recording employee data records with the help of business intelligence. The most efficient, cost effective and simplest way to carry out this process was to implement and include a central database structure. This project also employs the integration of MS excel and MS Access programmes. The database structure was created as a gradual process starting from designing conceptual, logical and physical models, establishing relevant relationships between different entities etc. This thesis also explains various concepts of the business intelligence, data modelling techniques, business intelligence architecture, relational database management system, hierarchy structures and data schemas. In order to show the process structure and flow, BPMN process diagrams, use case models and flowcharts are used. There were also queries executed in order analyse and calculate the employee data as per requirements. This helps in the evaluation of performance of the employees. Hence in the long run, the optimization of the employee records leads to efficient employee functioning and in turn optimal financial returns.

Along with the current work in optimizing the employee records, there can be further developments employed. Business intelligence can be studied and applied with more sophistication. The process can be made collaborative or teamwork friendly where more number of users can cumulatively work with the database file. There can also be developments in the visual presentation of reports, making it more graphical or pictorial, providing high-level data results (Tanila, 2018). There can also be implementation of cloud based storage system for data recovery and data backup. Many companies have also employed web based data application system where employees can have a real time read only access of daily attendance, vacation passes etc.

Finally, big data volumes can facilitate predictive decisions using data mining techniques. The efficient reporting results of the current thesis can be further used as input for predictions and association rules (Zarmina Jaffara, 2019). It will help in making managerial decisions in organization of performance and talent of the employees (Hurbean, 2005). The prediction analysis of an employee's job satisfaction factor is easily done using data mining, hence helping the organization performance by providing knowledge about employee behaviour. (Oden, 2017)

# Bibliography


1. Alejandro Vaisman, E. Z., 2014. *Data Warehouse Systems: Design and Implementation.* s.l.:s.n.

2. Anon., 2017. *GoLearningBus.* [Online]
   Available at: http://cdn.wagmob.com/subject/G124/html/introduction00to00dbms_1.html

3. Anon., 2018. *Java T Point.* [Online]
   Available at: https://www.javatpoint.com/data-warehouse-modeling

4. Anon., 2018. *Normalization.* [Online]
   Available at: https://www.javatpoint.com/dbms-normalization#:~:text=Normalization%20is%20the%20process%20of,Insertion%2C%20Update%20and%20Deletion%20Anomalies.

5. Anon., 2018. *Processmaker-Best Practices - Process Developmen.* [Online]
   Available at: https://wiki.processmaker.com/Best_Practices/Process_Development

6. Anon., 2018. *Sisense Inc..* [Online]
   Available at: https://documentation.sisense.com/latest/managing-data/working-with-data/ragged-unbalanced-hierarchies.htm#gsc.tab=0

7. Anon., 2018. *Techopedia.* [Online]
   Available at: https://www.techopedia.com/definition/1234/relational-database-rdb

8. Anon., 2019. *Cflow.* [Online]
   Available at: https://www.cflowapps.com/business-process-modeling/#:~:text=The%20term%20Business%20Process%20Modeling,graphical%20view%20of%20your%20workflows.

9. Anon., 2019. *Components Of Database Environment,* s.l.: Teacher Notes.

10. Anon., 2019. *Lean Six Sigma Belgium Trainings.* [Online]
    Available at: https://leansixsigmabelgium.com/blog/bpmn/

11. Anon., 2019. *Stack Oveflow.* [Online]
    Available at: https://stackoverflow.com/questions/4279089/what-is-the-difference-between-logical-data-model-and-conceptual-data-model

12. Anon., 2020. *keboola.* [Online]
Available at: https://www.keboola.com/blog/what-is-data-modeling

13. Biscobing, J., 2021. *searchdatamanagement.* [Online]
Available at: https://searchdatamanagement.techtarget.com/definition/relational-database

14. Cimino, M. G. C., 2009. *Using BPMN and Tracing for Rapid Business Process Prototyping Environments..* s.l.:ResearchGate.

15. Dmitry Anoshin, H. R. N. M., 2016. Mastering Business Intelligence with MicroStrategy.

16. ELMASRI, R. a. S. N., ,2011. *Fundamentals of Database Systems. 6th edition. Boston, Massachusetts,.* Boston, Massachusetts, USA: PEARSON. ISBN 978-0-136-08620-8..

17. Fahad, M., March 2016. *Business Intelligence,* s.l.: s.n.

18. Gangadharan, G. & Swami, S., 2004. *Business intelligence systems: design and implementation strategies,* s.l.: IEEE.

19. Gupta, M., 2020. *Business Intelligence Tools for SupportingCoreWebActivitiesin a SmallCompany,* s.l.: s.n.

20. Hannula, M. & Pirttimaki, V., March 2003. Business intelligence empirical study on the top 50 Finnish companies.

21. Hurbean, L., 2005. *Performance Management with Business Intelligence,* s.l.: SSRN Electronic Journal.

22. Lewis, L., 2020. *Processmaker - Understanding BPMN Diagrams and Symbols.* [Online]
Available at: https://www.processmaker.com/blog/bpmn-diagram-and-symbols/

23. Manning, K., 2020. *Processmaker.* [Online]
Available at: https://www.processmaker.com/blog/bpmn-2-0-tutorial-and-examples/

24. Mary K. Pratt, J. F., 2019. Business intelligence (BI) leverages software and services to transform data into actionable insights that inform an organization's business decisions.. 16 october.

25. Naeem, T., 2021. *Relational Database Management Systems (RDBMS): All You Need to Know,* s.l.: Astera.

26. Oden, C., 2017. *Employee Data Mining Information System,* s.l.: s.n.

27. raman_257, 2019. *ETL Process in Data Warehouse.* [Online]
Available at: https://www.geeksforgeeks.org/etl-process-in-data-warehouse/

28. Serbanescu Luminita, M. R., 2009. *BUSINESS INTELLIGENCE INSTRUMENTS FOR HR MONITORING,* s.l.: s.n.

29. SINGH, C., 2016. *Beginnersbook - Entity Relationship Diagram – ER Diagram in DBMS.* [Online]
Available at: https://beginnersbook.com/2015/04/e-r-model-in-dbms/

30. Tanila, U. ,. A. A., 2018. *Business Intelligence and Databases; An Empirical Study On Its Application Towards Recruitment Process,* s.l.: s.n.

31. technologies, N., March 2014. *Business-Intelligence-Analytics,* s.l.: s.n.

32. Wilfried Grossmann, S. R.-M., 2015. *Fundamentals of Business Intelligence ISBN 978-3-662-46531-8.* s.l.:s.n.

33. Zarmina Jaffara, D. W. N. Z. K., 2019. *Predictive Human Resource Analytics Using Data Mining Classification Techniques,* s.l.: International Journal of Computer (IJC).