



UNIVERSITY OF ECONOMICS AND LAW



Faculty of Information Systems



*Report*

# **BUSINESS INTELLIGENCE SOLUTION FOR HUMAN RESOURCES IN ADVENTURE WORKS**

Instructor: PhD Ho Trung Thanh

MSc Nguyen Van Tuyen

MSc Nguyen Van Ho

Implementation: Group 5

1. Mai Nguyen An Binh - K184060777
2. Hoang Nguyen Quynh Huong – K184060789
3. Truong Thi Tuyet Nga – K184060795
4. Dang Minh Thu – K184060807

*TP. HCM City, May, 2021*

## ACKNOWLEDGEMENTS

During the learning process "DATA ANALYTICS IN BUSINESS ", Group 5 has always received the attention and guidance from PhD Ho Trung Thanh, MSc Nguyen Van Tuyen, MSc Nguyen Van Ho First of all, Group 5 would like to express our deep gratitude to with teachers who have wholeheartedly helped Group 5 during the past time.

The report was made within 2,5 months. Although limited because of limited knowledge, through which the group understood and was able to install, operate and exploit Business Intelligences (BI) solution. Managing data, information and making statistics and reports plays a very important role for all units and organizations. The statistics reported in BI are intended to provide information

Intuitive, complete and timely information will help administrators in assessing, forecasting the situation, planning strategies as well as building policies and plans for business development in the short and long term. .

Group 5 would like to thank!

## **COMMITMENT**

Group 5 hereby declares that the project: “BUSINESS INTELLIGENCE SOLUTION FOR HUMAN RESOURCES IN ADVENTURE WORKS” is an independent research work under the guidance of the instructors of PhD Ho Trung Thanh and MSc Ha Hai Dang. Also do not have any copying by others. The topic and content of the internship report are the products that Group 5 has made efforts to research during their study at the school as well as the subject “Project Management of Information Systems”. The data and results presented in the report are completely honest, Group 5 would like to take full responsibility and discipline of the faculty and the school if there are problems.

# TABLE OF CONTENT

ACKNOWLEDGEMENTS	i
COMMITMENT	ii
LIST OF FIGURES	viii
LIST OF TABLES	x
LIST OF ACRONYMS	xi
CHAPTER 1: INTRODUCTION	1
1.1 Business case for the project	1
1.2 Objectives of the project	2
1.2.1 General Objective	2
1.2.2 Specific Objectives	2
1.3. Research Objects	3
1.4. Scope of the project	3
1.5 Value and desired outcome of the project	3
1.6 Structure of project	4
CHAPTER 2: THEORETICAL BACKGROUND	5
2.1 Overview about BI	5
2.1.1 What is BI?	5
2.1.2 BI Architecture	5
2.1.3 Advantage of BI in enterprises	6
2.1.4 BI Strategy for Business	8
2.2 ETL Process	10

2.2.1	What is ETL?	10
2.2.2	Why do we need ETL?	10
2.2.3	ETL Process	11
2.2	Data warehouse and Data mart	12
2.3.1.	What are Data warehouse and Data mart?	12
2.3.2	Who needs Data warehouse and Data mart?	13
2.3.3.	Advantages and disadvantages of Data warehouse	13
2.3.3.1.	Advantages of Data Warehouse (DWH):	13
2.3.3.2.	Disadvantages of Data Warehouse:	14
2.3.4.	Snowflake and Star schemas	15
2.3.4.1.	Limitations of Star Schema	16
2.3.4.2.	Limitations of the Snowflake Schema	16
2.4	KPIs	17
2.4.1.	KPIs Definition	17
2.4.2.	The advantages and disadvantages of KPIs	17
2.4.2.1.	Advantages of KPIs	17
2.4.2.2.	Disadvantages of KPIs	18
2.4.3.	Categories of KPIs (Purchasing, Production, Sales, or HR)	19
2.5	MDX language for analyzing multi dimension data and OLAP	19
2.5.1.	What is MDX language?	19
2.5.2	OLAP technique	20
2.5.3	MDX method and structure	21
CHAPTER 3: REQUIREMENTS ANALYTICS AND INTRODUCTION TO BI SOLUTION		23

3.1 Business processes HR	23
3.1.1 HR department	23
3.1.2 The structure of HR in 23	
3.1.3 HR process	24
3.1.3.1 Recruiting Process BPMN	25
3.1.3.2 Evaluating Process BPMN	26
3.1.3.3. Off-boarding process BPMN	27
3.2 Data source and challenges	28
3.2.1 Description and meaning of the HR data	28
3.2.1.1 HumanResource.Department	28
3.2.1.2 HumanResource.Employee	29
3.2.1.3 HumanResource.EmployeeDepartmentHistory	31
3.2.1.4 HumanResource.EmployeePayHistory	31
3.2.1.5 HumanResource.JobCandidate	33
3.2.1.6 HumanResource.Shift	33
3.2.1.7 Person.Person	35
3.2.2 Physical Data Model (PDM)	38
3.2.3 Challenges of data HR	38
3.3. Business Requirements Analysis of HR	39
3.4.1. Development process	39
3.4.2. Employee perf the operation	39
3.4.3. The process of timekeeping, payroll at HR	39
3.4.3.1 . Timekeeping	39

3.4.3.2. Calculate salary, salary, bonus and other income	40
3.4.5. Departmental change process	40
3.5 . IT requirements Analysis (IT & Infrastructure)	41
3.6 Comparative Analysis of BI and Data Visualization Tools	42
3.6.1 Surveying and evaluation	42
3.6.2 Proposing BI solution for the project	51
CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA	53
4.1. Designing Data Warehouse	53
4.1.1. Bus matrix	53
4.1.2 Master Data	53
4.1.3 Transaction Data	54
4.1.4. Dimension table	54
4.1.4.1 DimEmployee	54
4.1.4.2 DimDepartment	55
4.1.4.3 DimShift	55
4.1.4.4 DimTime	56
4.1.5 Fact table	57
4.1.5.1 FactPayHistory	57
4.1.5.2 FactDepartmentHistory	58
4.1.6. Data Warehouse model	59
4.2. ETL processes	59
4.2.1. Dimension Table's ETL Process	59
4.2.1.1 DimEmployee	59

4.2.1.2 DimDepartment	60
4.2.1.3 DimShift	61
4.2.1.4 DimTime	62
4.2.2. Fact Table's ETL Process	63
4.2.2.1 FactpPayHistory	63
4.2.2.2 FactDepartmentHistory	64
CHAPTER 5: DATA ANALYTICS	65
5.1. Data analytics with SSAS technology	65
5.1.1. Building the cube	65
5.1.2. Analysis with SSAS	66
5.1.2.1. Report on the situation of personnel in the department in 2009	66
5.1.2.2. List of positions that move departments	67
5.1.3. Building KPIs system	67
5.1.3.1. KPI Number of new hires each year by each department	68
5.1.3.2. KPI Annual departmental average salary	71
Figure 5.1.3.2: Value of Annual departmental average salary	71
5.2. Data analytics with MDX	73
5.2.1. List time-off hours for each department	73
5.2.2. Difference between the time-off hours of each department	74
5.2.3. Top 5 departments with the most time-off hours	75
5.2.4. The department has more than 1200 hours of time-off hours	76
5.2.5 Rate of employee	76
5.2.6 Average Rate of department	77



5.2.7 Top 5 employees with the highest wages at Sales and Marketing	78
5.2.8 Wages of divisions	78
5.2.9 Medium performance wages by sex	79
<i>Figure 6.5.9: Result of Medium performance wages by sex</i>	80
5.2.10 Average break time in each department	80
CHAPTER 6: VISUALIZATION AND FORECASTING	81
6.1. Report and dashboard systems	81
6.2. Data analysis with Power BI	82
6.3. Data analysis with the Pivot Table tool in Microsoft Excel	83
6.4. Data analysis with Tableau	87
6.5. Time series and Forecasting	88
6.5.1. Forecast the total hired employees in the next 7 years	88
6.5.2. Time series plot of recruitment (by Excel)	89
6.5.3. Time series plot of total rate increasing (by Excel)	89
6.6. Evaluation and Discussion	90
CHAPTER 7: CONCLUSION AND FUTURE WORKS	92
7.1. Results	92
7.2. Limitations	92
7.3. Future works	92
REFERENCES	93

## LIST OF FIGURES

Figure 3.1.3.1: Recruiting Process of HR .....	25
Figure 3.1.3.2 : Evaluating Process of HR .....	26
Figure 3.1.3.3: Off-boarding process .....	27
Figure 3.2.2: Physical Data Model (PDM) of HR.....	37
Figure 3.6.2: BI solution.....	50
Figure 4.1.1.1: Table DimEmployee .....	54
Figure 4.1.1.2: Table DimDepartment .....	54
Figure 4.1.1.3: Table DimShift.....	54
Figure 4.1.1.4: Table DimTime .....	55
Figure 4.1.2.1: Table FactPayHistory.....	56
Figure 4.1.5: Data Warehouse model of HR .....	58
Figure 4.2.1.1: SSIS for DimEmployee.....	58
Figure 4.2.1.2: SSIS for DimDepartment .....	59
Figure 4.2.2.2: SSIS for FactDepartmentHistory .....	63
Figure 5.1.1: Cube Model.....	64
Figure 5.1.2.1: Situation of personnel in the department in 2009 .....	65
Figure 5.1.2.2: List of positions.....	66
Figure 5.1.3.1: KPI value of new hires each year by each department .....	67
Figure 5.1.3.1: Goal of new hires each year by each department .....	68
Figure 5.1.3.1: Status of new hires each year by each department .....	68
Figure 5.1.3.1: Trend of new hires each year by each department.....	69
Figure 5.1.3.1: Result of new hires each year by each department.....	69

Figure 5.1.3.2: Goal of Annual departmental average salary .....	72
Figure 5.1.3.2: Status of Annual departmental average salary .....	72
Figure 5.1.3.2: Trend of Annual departmental average salary .....	72
Figure 5.1.3.2: Result of Annual departmental average salary .....	73
Figure 6.5.1: Result of List time-off hours for each department.....	74
Figure 6.5.2: Result of Difference between the time-off hours of each department.....	75
Figure 6.5.4: Result of Top 5 departments with the most time-off hours .....	76
Figure 6.5.5: Result of Rate of employee .....	77
Figure 6.5.7: Result of Top 5 employees with the highest wages at Sales and Marketing	78
Figure 6.5.8: Result of Top 5 employees with the highest wages at Sales and Marketing	79
Figure 6.5.10: Result of Average break time in each department .....	81
Figure 6.1: Structure of Report & Dashboard .....	82
Figure 6.2: Dashboard demographic management in Power BI tool .....	83
Figure 6.3.1: Dashboard salary hourly rate of group department .....	85
Figure 6.3.2: Dashboard the fluctuation of group departments .....	87
Figure 6.4: Dashboard salary management in Tableau .....	88
Figure 6.6: Forecast the hired employees in the next 7 years .....	89
Figure 6.6.2: Time series plot of recruitment .....	90
Figure 6.6.3: Time series plot of total rate increasing .....	91

## LIST OF TABLES

Table 2.3.4: Comparison Star and Snowflake Schema	22
Table 3.2.1.1: HumanResource.Department table	35
Table 3.2.1.2: HumanResource.Employee	36
Table 3.2.1.3: HumanResource.EmployeeDepartmentHistory	38
Table 3.2.1.4: HumanResource.EmployeeDepartmentHistory	38
Table 3.2.1.5: HumanResource.JobCandidate	40
Table 3.2.1.6: HumanResource.Shift	40
Table 3.2.1.7: Person.Person	42
Table 3.6.1: Comparative BI tool	49

## LIST OF ACRONYMS

No.	Abbreviation	Definition
1	BI	Business Intelligence
2	DW	Data Warehouse
3	ETL	Extracts Transforms Load
4	KPI	Key Performance Indicator
5	DB	Database
6	HR	Human Resource
7	CRM	Customer relationship Management
8	OLTP	Online Transaction Processing
10	OLAP	Online Analysis Processing
11	MBO	Management by Objectives
12	MDX	Multi Dimensional eXpression
13	XML	eXtensible Markup Language
15	DDL	Data Definition Language
16	DML	Data Manipulation Language
17	SSIS	SQL Server Integration Services
18	SSRS	SQL Server Reporting Service
19	SSAS	SQL Server Analysis services
20	SQL	Structured Query Language

## **CHAPTER 1: INTRODUCTION**

In this chapter, the topic introduces an overview of the reasons for choosing the research topic and its objectives. From there set out the tasks, objects and methods to implement the topic.

### **1.1 Business case for the project**

An enterprise that wants to organize its business is forced to exchange activities with the market. To do that, businesses can produce goods for exchange, or import goods from other parties, it depends on the type of business of each organization. Therefore, the question of how to operate and manage those activities as well as the employees, the business is carried out according to the right process and effective management is still a difficult challenge. during the operation and development of each business.

Enterprises that want to operate effectively always require the efforts of administrators in managing and exploiting form information, statistical reports. The numbers will provide managers with intuitive, complete, timely and accurate views. From activities such as self-assessment, situation forecasting to strategic planning, final decision making to build development strategies, the system plays an important role in supporting managers.

The strong development of the 4.0 revolution, always sets out requirement of innovation, catching up with the times and applying technology is a problem that any enterprise standing on the threshold of integration must recognize. Therefore, turning the discrete data of the business into useful information to support business decision-making is a very necessary job. Enterprises need to choose the right tools to gather, process and present data logically and consistently.

Business Intelligence Solutions (BI) has been leading the world in recent years. By integrating from various sources into data warehouses, BI provides administrators with an

overview of the business, data transfer. From there turn them into business knowledge and management for businesses.

From the above reasons, group 5 decided to choose the topic "BUSINESS INTELLIGENCE SOLUTION FOR HUMAN RESOURCES IN ADVENTURE WORKS" in order to improve knowledge to serve in the research process of the subject "Data Analysis in Business".

## **1.2 Objectives of the project**

### **1.2.1 General Objective**

Build, deploy, operate and exploit management solutions and use tools, data analysis techniques (Business Intelligences - BI) and decision making, SQL Server Business Intelligence solution suite and metrics Key Performance Indicators assess performance in business and management, then apply these studies to data analysis and build financial analysis reports to support management-level decision-making. is based on Microsoft's Adventure works database.

### **1.2.2 Specific Objectives**

- Microsoft Adventure works database research is integrated with business processes.
- Research on decision support model review and BI system
- Design and build Data warehouse, implement ETL for financial reporting.
- Research and use in detail the SQL Server Business Intelligence solution set
- Key Performance Indicators - Key Performance Indicators.
- Research and apply MDX language to perform queries.
- Research and analyze requirements in the financial reporting system and the significance of each report.
- Research and create dashboard on visualization tools such as Excel, Power BI, Tableau, SSAS and Qlik, ...
- Designing the method of analyzing business results reports, supporting business decision-making based on Microsoft's SQL Server Business Intelligence solution.

### **1.3. Research Objects**

- Adventure works database
- SQL Server Business Intelligence Application
- Microsoft Visual Studio Application
- Visualization tools like Power BI, Excel, Tableau, SSAS and Qlik
- OLAP multidimensional data analysis techniques
- Business process of selling
- KPI performance evaluation solution

### **1.4. Scope of the project**

The topic is limited in understanding the BI solution for the HR module of the company Adventure Work.

- Historical database is the manufacture and sale of bicycles made of metal and synthetic materials. The company's markets include North America, Europe and Asia. While the company's headquarters is located in Bothell, Washington and employs 290 people, the company also has several regional sales staff in its respective markets.
- Reports:
  - Report on employee's salary
  - Reports on staff moving departments
  - Reporting on employee's leave
  - Report on employee's performance
  - Report on employee recruitment process

### **1.5 Value and desired outcome of the project**

- The value of the project
  - Understand the business processes of sales.
  - Understand the BI solution.
  - Understand KPI assessment methodology.
  - Understand the data analysis model and business intelligence system



- The desired outcome of the project: Creating documents with smart business models and data analysis with financial and business reports for the sales module helps administrators to control information effectively as well as make right decisions.

## **1.6 Structure of project**

The report is divided into 7 chapter

- Chapter 1: Introduction
- Chapter 2: Theoretical background
- Chapter 3: Requirements analytics and introduction to bi solution
- Chapter 4: Building data warehouse and integrating data
- Chapter 6: Visualization and forecasting
- Chapter 7: Conclusion and future works

## **CHAPTER 2: THEORETICAL BACKGROUND**

In this chapter, the topic will introduce BI - a new solution for businesses to support decision making for managers. In addition, the concepts and applications for this solution are also mentioned.

### **2.1 Overview about BI**

#### **2.1.1 What is BI?**

Information about orders, inventory, suppliers as well as customers are the most valuable asset of one company. However, a large amount of information everyday does not help the managers make the right decisions. Consolidation and rearrangement of data will give businesses an overview of business situations, helping managers make better decisions, this helps them benefit from a competitive position. Managers and experts need a support solution to make quick decisions over and better. BI will be a decision support solution for those problems.

Business intelligence (BI) (1) combines business analytics, data mining, data visualization, data tools and infrastructure, and best practices to help organizations to make more data-driven decisions. In practice, the managers could have a comprehensive view of their organization's data and use that data to drive change, eliminate inefficiencies, and quickly adapt to market or supply changes.

In other definitions, Business Intelligence leverages software and services to transform data into actionable insights that inform an organization's strategic and tactical business decisions. BI tools access and analyze data sets and present analytical findings in reports, summaries, dashboards, graphs, charts and maps to provide users with detailed intelligence about the state of the business.

#### **2.1.2 BI Architecture**

The key elements of a business intelligence architecture are:

- Data Sources

This component of BI involves various forms of stored data. It's about taking the raw data and using software applications to create meaningful data sources that each

division can use to positively impact business. Organization data is first created in these databases: MySQL, Oracle, MSSQL,...

- Data Warehouse

The data warehouse is the core of the BI system. A data warehouse is a database built for the purpose of data analysis and reporting. This purpose changes the design of this database as well. Warehouse will have data extracted from various operational systems, transformed to make the data consistent, and loaded for analysis. A data warehouse will help in achieving cross-functional analysis, summarized data, and maintaining one version of the truth across the enterprise.

Data Warehouse can only be read, not written or updated and can only be accessed and updated by the ETL package that converts data from Data Sources into Data Warehouse.

- Analysis Server

Responsible for executing cubes designed based on data dimensions and business knowledge. Cube is responsible for receiving input from DWH and executing the predefined business to return the results.

- Reporting Server

Execute reports with the output received from the Analysis Server and a place for administrators to gather web-based reports, these reports can be attached to web applications, or applications.

- Data Mining

It is the process of extracting processed data information (in accordance with the specific requirements of the business) from Data Warehouse and then combined with algorithms to make (or predict) profitable decisions for business. business.

### 2.1.3 Advantage of BI in enterprises

Business intelligence is more than just software. It's a holistic initiative to use data in day-to-day operations. The 7 benefits below translate into real-world success that showcases BI in action (2)

- Faster analysis, intuitive dashboards:

BI platforms are designed to do heavy-duty processing of data in the cloud or on your company's servers. BI tools pull in data from multiple sources into a data warehouse, and then analyzes the data according to user queries, drag-and-drop reports, and dashboards.

- Increased organizational efficiency

BI provides leaders the ability to access data and gain a holistic view of their operations, and the ability to benchmark results against the larger organization. With a holistic view of the organization, leaders can identify areas of opportunity.

- Data-driven business decisions

Having accurate data and faster reporting capability provides for better business decisions. Miller Coors customized mobile dashboards for their sales team so they can view real-time data and sales forecasts before going into meetings with potential clients. They can speak of clients' or prospects' needs confidently and know the data is up-to-date. No longer do leaders have to wait days or weeks for reports and deal with the risk of data that may be outdated.

- Improved customer experience

Business intelligence can directly impact customer experience and customer satisfaction. Verizon deployed BI systems across multiple departments, creating more than 1,500 dashboards for employees. These dashboards pulled data from operations and text data from customer support chat sessions. Using this data, Verizon was able to identify opportunities to improve customer service and reduce support calls by 43 percent.

- Improved employee satisfaction

IT departments and analysts spend less time responding to business user requests. Departments who didn't have access to their own data without contacting analysts or IT can now jump into data analysis with little training. BI is designed to be scalable, providing data solutions to departments who need it and for employees who crave data.

- Trusted and governed data

In traditional data analysis, different departments' data is siloed and users have to access several databases to answer their reporting questions. Now, modern BI platforms can combine all of these internal databases with external data sources such as customer data, social data, and even historical weather data into one data warehouse. Departments across an organization can access the same data at one time.

- Increased competitive advantage

Organizations can be more competitive when they know the market and their performance within the market. With BI, businesses can keep up with changes in the industry, monitor seasonal changes in the market, and anticipate customer needs.

#### 2.1.4 BI Strategy for Business

To build a successful Business Intelligence strategy for your business, you need to follow these steps (3)

- Assess your BI current system and find out where you are

To know where you're heading, you must set a baseline. The first thing to do is talk to all players of the current BI processes: users and the IT team, department managers and stakeholders to understand your problem, what you need and how you can solve it. To understand it clearly, a SWOT analysis can help you.

- Plan your objectives (and map the supporting data)

At this stage, you will need to plan your business goal. It may be an oxymoron, but with so much data out there, the more focused you are in planning your business goals, the better off you're going to be. Essentially, your data needs to address a business question or a business need.

- Define how the data will be shared (and how it will be distributed)

There are two basic strategies that your company can take in this step

- A Decentralized Approach

Some organizations empower its end users with interactive dashboards. This is where the term citizen data scientist comes into play. Citizen data scientists do not need to be data scientists, they just need to know their way around the data, and that begins with giving them access to more than just a dashboard.

- A Centralized Approach

On the flip side, we have organizations that have decided to be in complete control of the data that is distributed, including who sees what, and how much of it they can touch. Take Nasdaq for instance. Their BI strategy took into consideration their sensitive data, huge distribution channels, and the need for better governance to reach one version of the truth. Building on this strategy, Nasdaq provides its customers with dashboards, but it does not provide them with the ability to work directly on the data models. There is a certain amount of drill-down that each customer can perform to see more details, but the data is strictly governed with system level, object level, data level, and row-level security.

- Deliver your solution

Part of your BI strategy needs to be how you'll deliver the solution to your end-users, or your entire company. Some of this may come naturally with the decision on how you will share data. For instance, for a centralized approach (where the user has no room for adjustments) you will need to put more effort into training and documentation. For a decentralized solution, you may want to do short release cycles, asking for feedback on each release, and then incorporate that feedback into the next release.

- Find the roadblocks

Your data can be stored in a database or may even be located with a third party vendor. If you are outsourcing HR services to a company that is managing your hiring pipeline, or if you have cloud-based service providers like Market for managing marketing campaigns or QuickBooks for financial services, then you will

need to plan how to connect to their data and learn about its structure in order to use it properly.

For some third-party applications, there may be a simple API that can be used to import the data. For the IT department, it may be just the manager's permission and a couple of signatures that you will need to find in order to access the data. This all takes time and resources that need to be allocated inside your BI strategy.

## **2.2 ETL Process**

### **2.2.1 What is ETL?**

ETL stands for “Extract, Transform, and Load”. ETL is a process that extracts the data from different source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system.

- Extract: data is extracted from the source system into the staging area.
- Transform: processes data by data cleaning and transforming them into a proper storage format/structure for the purposes of querying and analysis.
- Loading: the insertion of data into the final target database.

In the report, the author uses SQL Server Integration Services (SSIS) tool to design ETL process.

### **2.2.2 Why do we need ETL?**

Businesses have relied on the ETL process for many years to get a consolidated view of the data that drives better business decisions. Today, this method of integrating data from multiple systems and sources is still a core component of an organization's data integration toolbox.

- Offers deep historical context for business.
- Enhances Business Intelligence solutions for decision making.

- Enables context and data aggregations so that business can generate higher revenue and/or save money.
- Enables a common data repository.
- Allows verification of data transformation, aggregation and calculations rules.
- Allows sample data comparison between source and target system.
- Helps to improve productivity as it codifies and reuses without additional technical skills.

### 2.2.3 ETL Process

A viable approach should not only match with your organization's need and business requirements but also perform on all the bellows stages. A basic ETL process can be categorized in three stages: (4)

#### - Step 1: Extraction

In this step of ETL architecture, data is extracted from the source system into the staging area. Transformations if any are done in staging area so that performance of source system is not degraded. Also, if corrupted data is copied directly from the source into Data warehouse database, rollback will be a challenge. Staging area gives an opportunity to validate extracted data before it moves into the Data warehouse.

Hence one needs a logical data map before data is extracted and loaded physically. This data map describes the relationship between sources and target data.

#### - Step 2: Transformation

Data extracted from the source server is raw and not usable in its original form. Therefore it needs to be cleansed, mapped and transformed. In fact, this is the key



step where the ETL process adds value and changes data such that insightful BI reports can be generated.

It is one of the important ETL concepts where you apply a set of functions on extracted data. Data that does not require any transformation is called as direct move or pass through data. In the transformation step, you can perform customized operations on data.

- Step 3: Load

Loading data into the target data warehouse database is the last step of the ETL process. In a typical Data warehouse, a huge volume of data needs to be loaded in a relatively short period (nights). Hence, the load process should be optimized for performance.

In case of load failure, recover mechanisms should be configured to restart from the point of failure without data integrity loss. Data Warehouse admins need to monitor, resume, cancel loads as per prevailing server performance.

## **2.2 Data warehouse and Data mart**

### **2.3.1. What are Data warehouse and Data mart?**

A Data Warehousing (DW) (5) is a process for collecting and managing data from varied sources to provide meaningful business insights. A Data warehouse is typically used to connect and analyze business data from heterogeneous sources. The data warehouse is the core of the BI system which is built for data analysis and reporting.

A Data Mart (6) is focused on a single functional area of an organization and contains a subset of data stored in a Data Warehouse. A Data Mart is a condensed version of Data Warehouse and is designed for use by a specific department, unit or set of users in an organization. E.g., Marketing, Sales, HR or finance. It is often controlled by a single department in an organization.

Data Mart usually draws data from only a few sources compared to a Data warehouse. Data marts are small in size and are more flexible compared to a Data Warehouse.

### 2.3.2 Who needs Data warehouse and Data mart?

DWH (Data warehouse) is needed for all types of users like:

- Decision makers who rely on mass amount of data
- Users who use customized, complex processes to obtain information from multiple data sources.
- It is also used by the people who want simple technology to access the data
- It is also essential for those people who want a systematic approach for making decisions.
- If the user wants fast performance on a huge amount of data which is a necessity for reports, grids or charts, then Data warehouse proves useful.
- Data warehouse is a first step If you want to discover 'hidden patterns' of data-flows and groupings.

Data mart can be used by the same users as Data warehouse. However, because Data mart is specially designed for user needs, so specific people/groups of people belonging to a certain group of people or organization-specific departments can use the respective data marts effectively even though they do not have right to access the central data.

In addition, Data mart also contains historical data, so it is also essential for those who want to identify data trends in a particular area of the organization.

### 2.3.3. Advantages and disadvantages of Data warehouse

#### 2.3.3.1. *Advantages of Data Warehouse (DWH):*

- Data warehouse allows business users to quickly access critical data from some sources all in one place.
- Data warehouse provides consistent information on various cross-functional activities. It is also supporting ad-hoc reporting and query.

- Data Warehouse helps to integrate many sources of data to reduce stress on the production system.
- Data warehouse helps to reduce total turnaround time for analysis and reporting.
- Restructuring and Integration make it easier for the user to use for reporting and analysis.
- Data warehouse allows users to access critical data from the number of sources in a single place. Therefore, it saves user's time retrieving data from multiple sources.
- Data warehouse stores a large amount of historical data. This helps users to analyze different time periods and trends to make future predictions.

#### *2.3.3.2. Disadvantages of Data Warehouse:*

- Not an ideal option for unstructured data.
- Creation and Implementation of Data Warehouse is surely a time confusing affair.
- Data Warehouse can be outdated relatively quickly
- Difficult to make changes in data types and ranges, data source schema, indexes, and queries.
- The data warehouse may seem easy, but actually, it is too complex for the average users.
- Despite best efforts at project management, data warehousing project scope will always increase.
- Sometimes warehouse users will develop different business rules.
- Organizations need to spend lots of their resources for training and Implementation purposes.

### 2.3.4. Snowflake and Star schemas

*Table 2.3.4: Comparison Star and Snowflake Schema*

<b>Star Schema</b>	<b>Snowflake Schema</b>
The Star Schema contains the Fact Tables and the Dimension Tables.	The Snowflake Schema contains the Fact Tables, Dimension Tables, and the Sub-Dimension Tables.
It is a top-down model.	It is a bottom-up model.
Queries execute faster in the Star Schema. This is achieved using the Star Join Query Optimization technique. Tables can be connected with Multiple Dimensions.	Execution of queries takes a longer time in the Snowflake Schema than in the Star Schema. The centralized Fact Table in the Snowflake Schema is less likely connected with Multiple Dimensions. The longer execution time is also caused by the many number of foreign keys.
There is no normalization in the Star Schema.	In the Snowflake Schema, there is both normalization and denormalization.
It has a very simple design.	This Schema has a complex design because of the Sub-Dimension Tables.
The Star Schema has a low query complexity.	The query complexity of the Snowflake Schema is higher than that of the Star Schema.
It only has a few number of foreign keys.	It has a high number of foreign keys.

The Star Schema has a high data redundancy, which makes it a bit difficult to maintain and change.	The Snowflake Schema has a low data redundancy, which makes it easier to maintain and change.
The hierarchies for the Dimensions are kept in the Dimensional Table.	The hierarchies for the Dimensions are divided into separate tables.
In the Star Schema, only a single join creates the relationship between the Fact Table and the Dimension Tables.	In the Snowflake Schema, many joins are needed to fetch the data.
Faster cube processing.	Cube processing may be slow due to the complex join.
It is suitable for data marts that have simple relationships (1:1 or 1:many).	Suitable for use in the Data Warehouse core for simplifying complex relationships (many to many).

#### *2.3.4.1. Limitations of Star Schema*

The following are the challenges of using the Star Schema:

- **Decreased Data Integrity:** It's difficult to impose data integrity effectively in the Star Schema due to its denormalized data structure.
- **Difficulty Handling Diverse and Complex Queries:** The Star Schema can only work well with a narrow set of simple queries due to its denormalized data structure.
- **Doesn't Support Many-to-Many Relationships:** The Star Schema has a simple Dimension Schema, hence, it can't handle "many-to-many data relationships" effectively.

#### *2.3.4.2. Limitations of the Snowflake Schema*

The following are the challenges of using the Snowflake Schema:

- **Complex Data Schemas:** The Snowflake Schema generates many levels of complexity as it normalizes the attributes of the Star Schema. This results in complicated source query JOINS.
- **Slower Cube Data Processing:** This is a result of the complex joins, hence, the Star Schema is better as far as cube data processing is concerned.
- **Lower Data Integrity:** Even after normalization, Snowflake Schemas cannot provide the level of data integrity that is provided by the traditional, highly-normalized database structures.

## 2.4 KPIs

### 2.4.1. KPIs Definition

A Key Performance Indicator (KPIs) (7) is a measurable value that demonstrates how effectively a company is achieving key business objectives. Organizations use KPIs at multiple levels to evaluate their success at reaching targets. High-level KPIs may focus on the overall performance of the business, while low-level KPIs may focus on processes in departments such as sales, marketing, HR, support and others.

### 2.4.2. The advantages and disadvantages of KPIs

#### 2.4.2.1. *Advantages of KPIs*

- Close learning gap

KPIs help you recognize and address learning gaps. It may indicate the present performance of a strategic goal. If you're not reaching an objective or goal, it may indicate that employees need further training (8).

- Empower employees to take action

Empowering employees to take action is another pro of KPIs. Clear KPIs drive employees to action and direct them along the way. If you start to see results, then you're on the right track. If you don't see results, then revisit the goal(s), the KPIs you're measuring on the way to that goal, and make pivots as necessary.

- Measure Outcomes and Results

KPIs allow you to measure outcomes and results. A good KPI, by definition, should be measurable and trackable. Without a way to measure progress towards your goals, you can't make improvements or adjustments. Whether you reach a goal or not, it's essential to know "why." One of the major strengths of a KPI is its ability to measure whether certain training decisions paid off or fell flat.

#### *2.4.2.2. Disadvantages of KPIs*

- KPIs Need Time

One con of KPIs is that they don't always offer actionable information immediately. If you invest in a new employee training program, you want to know that it's making a difference. But training KPIs don't show team transformations in a day. It may take a few months before seeing results, depending on your goal(s) and how often you're tracking employee training KPIs.

- KPIs have a high learning curve

One limitation of performance measurement that you may fall into is tracking so many KPIs that you become overwhelmed by all of the data. If you try to implement too many at one time, you'll lose track and end up failing at all of them. This can leave you feeling frustrated and discouraged as a manager.

Start small and work your way up. As your comfort level creating and implementing KPIs increases, you can start adding more.

- KPIs do not meet the SMART criteria

If the goal does not meet Specific criteria, the entrepreneur does not know what to do and how to achieve the desired effect.

The indicators do not meet the Measurable criteria (measurable): Thus, when giving the criteria, it no longer means to measure the performance of the work.

The KPIs do not meet the Achievable and Realistic criteria ... The construction goal is too far from reality, the company cannot achieve the goal despite trying its best. This leads to disappointment.

#### 2.4.3. Categories of KPIs (Purchasing, Production, Sales, or HR)

##### KPIs building for Human Resource module

- Number of new employees per group of department

From the change in the number of employees per year for each different group of departments combined with the turnover rate of the company to planning future recruitment. Performance indicators are set in accordance with the number forecasted in future recruitment planning

- Staff turnover rate in each department

Keeping the talent turnover rate as low as possible, especially for your junior staff, as they will be your most valuable human resources in the future.

- The annual rate of male and female employees

Gender diversity can be present in your company. There is no target rate in particular as it also depends on the availability of the candidates at a certain time. But countless studies prove that a balanced environment – especially in boards of direction – are much more successful than any other in the long run.



## **2.5 MDX language for analyzing multi dimension data and OLAP**

### **2.5.1. What is MDX language?**

MDX is short for Multidimensional Expressions, which is a syntax developed for defining multidimensional objects and querying and manipulating multidimensional data in OLAP databases. The purpose of MDX is to make creating queries to access data from multiple dimensional databases easier and more intuitive. MDX is a query language designed for OLAP databases, as SQL is a query language for relational databases.

MDX is essentially an extension to SQL used for queries and script access to multidimensional data. MDX queries access data stored in a SQL Server Analysis Server cube by bringing back facts related to dimensions. Querying multidimensional data sources can be a complex task, and a well-thought out syntax makes the job much easier. An MDX expression is designed to produce a multi-dimensional result set comprising axis data and cell data.

MDX was first introduced as part of the OLE DB for OLAP specification in 1997 from Microsoft. It was invented by the group of SQL Server engineers including Mosha Pasumansky. The specification was quickly followed by commercial release of Microsoft OLAP Services 7.0 in 1998 and later by Microsoft Analysis Services. The latest version of the OLE DB for OLAP specification was issued by Microsoft in 1999.

### **2.5.2 OLAP technique**

OLAP technology is a part of business intelligence, along with reporting, data mining and other BI activities. It is broadly used for sales and marketing reporting, financial planning, budgeting and forecasting.

The basis of an OLAP system is OLAP cube, where four main operations can be performed with data: consolidation (roll-up), drill-down, slicing/dicing, and pivoting (rotating). Consolidation aggregates the data to accumulate it in one or more dimensions.

Drill-down allows users to navigate through the details. Slicing and dicing take out a specific chunk of the OLAP cube data and view it from different points.

OLAP data consists of two core types of information: measures (numeric data) and dimensions (categories in which measures are organized).

It is worth mentioning that, unlike two-dimensional relational databases, OLAP data is multidimensional, which allows analyzing the information in different ways and from different perspectives.

### 2.5.3 MDX method and structure

In MDX (9), the SELECT statement is used to specify a dataset containing a subset of multidimensional data. To discuss the various syntax elements of the MDX SELECT statement, this topic presents a basic MDX query example and breaks it down into its syntax elements, discussing the purpose and structure of each element.

To specify a dataset, an MDX query must contain information about:

- The number of axes. You can specify up to 128 axes in an MDX query.
- The members from each dimension to include on each axis of the MDX query.
- The name of the cube that sets the context of the MDX query.
- The members from a slicer dimension on which data is sliced for members from the axis dimensions.

This information can be complex. As you will see in this topic, MDX syntax can provide such information in a simple and straightforward manner, using the MDX SELECT statement. In terms of structure: The structure of MDX is like SQL but is more open to work with multi-dimensional database. Usually an MDX query has the following structure:

- With: is used to define a new dimension from an existing dimension

- SELECT statement is used to describe a data set that is a subset of data multidimensional data.
- FROM: used to specify the data block used for retrieval
- WHERE: Defines the cross-section of data

The syntax of the MDX query is similar to the SQL syntax, but there are some differences Special:

- The MDX syntax distinguishes sets by using the {} wrapper braces around ministries or members.
- An MDX query can contain up to 128 dimensions, but only the first 5 dimensions are available Aliase (sub-name). An axis can be referenced by its order in a sentence MDX or by its sub-name if present.
- In an SQL query, the FROM clause can be subject to multiple data tables. However the FROM clause in an MDX query is limited to a single block. Information from other blocks can be referenced by value use the Lookup Cube function.
- The WHERE clause is used to describe the direction of data cut. If one way is not possible mentioned in the WHERE clause, SQL Analysis Service still considers it one dimension cut, but filtered by its default number of members. The WHERE clause is possible changing the filtering for a certain dimension, helping to refine the resulting data.

## **CHAPTER 3: REQUIREMENTS ANALYTICS AND INTRODUCTION TO BI SOLUTION**

In this chapter, the topic will give the requirements that the topic needs to solve and give a way to solve those requirements. From there build a data warehouse for data analysis and decision making.

### **3.1 Business processes HR**

#### **3.1.1 HR department**

Unlike the sales department, in companies, the human resources department is not the department that directly brings in revenue for the business. However, that cannot be denied the important role of the human resources department in maintaining and developing the business.

Human Resource Management includes planning and effectively using human resources to help organizations and individuals achieve common goals. The HR department develops processes and policies for industrial relations management. Department heads use these policies and procedures to manage staff and perform personnel activities.

#### **3.1.2 The structure of HR in Adventure Works Cycle:**

- 1 Human Resources Manager: is the head of the HR department who is responsible for developing and implementing HR strategies consistent with the overall business strategy, helping to resolve inquiries, complaints or other issues related to the Human Resources department.
- 2 Human Resources Administrative Assistants: Responsible for supporting and assisting the head of the human resources department. In addition to the task of building a schedule, supervising employees in the implementation of the general rules of the company, there is also the responsibility to interact and combine with other departments to ensure the company is maintained in the best condition.
- 1 Benefits Specialist: is responsible for the salary and welfare of employees, inspecting the timekeeping and implementing the labor regulations, managing

the working hours, updating the provisions of the Law on regimes, salaries, policies...; answering to employees' resignation proposals or all questions about salary, ...

- 2 Recruiters: Looking for and persuading candidates to become a member of the company. Responsible for analyzing the job to be recruited; attracting, screening and classifying candidates; interview and selection; guiding and helping new employees to integrate.

### 3.1.3 HR process

Human resource management is essential to streamline workflow in every business. It requires many tasks to help the firm run smoothly. Those tasks are:

- Staff requisition
- New-hire onboarding
- Employee off boarding
- Training requisition
- Performance evaluation
- Leave request
- Notice of absence
- Vacations and extraordinary licenses requisition
- Employee change of position/location
- Employee discount request (for company services)
- And so on

However, we could summarize those tasks into 3 vital steps that the HR department has to do everyday in any firm: Recruitment, Evaluation and Off-boarding process. Three BPMN below describe clearly both 3 tasks.

### 3.1.3.1 Recruiting Process BPMN

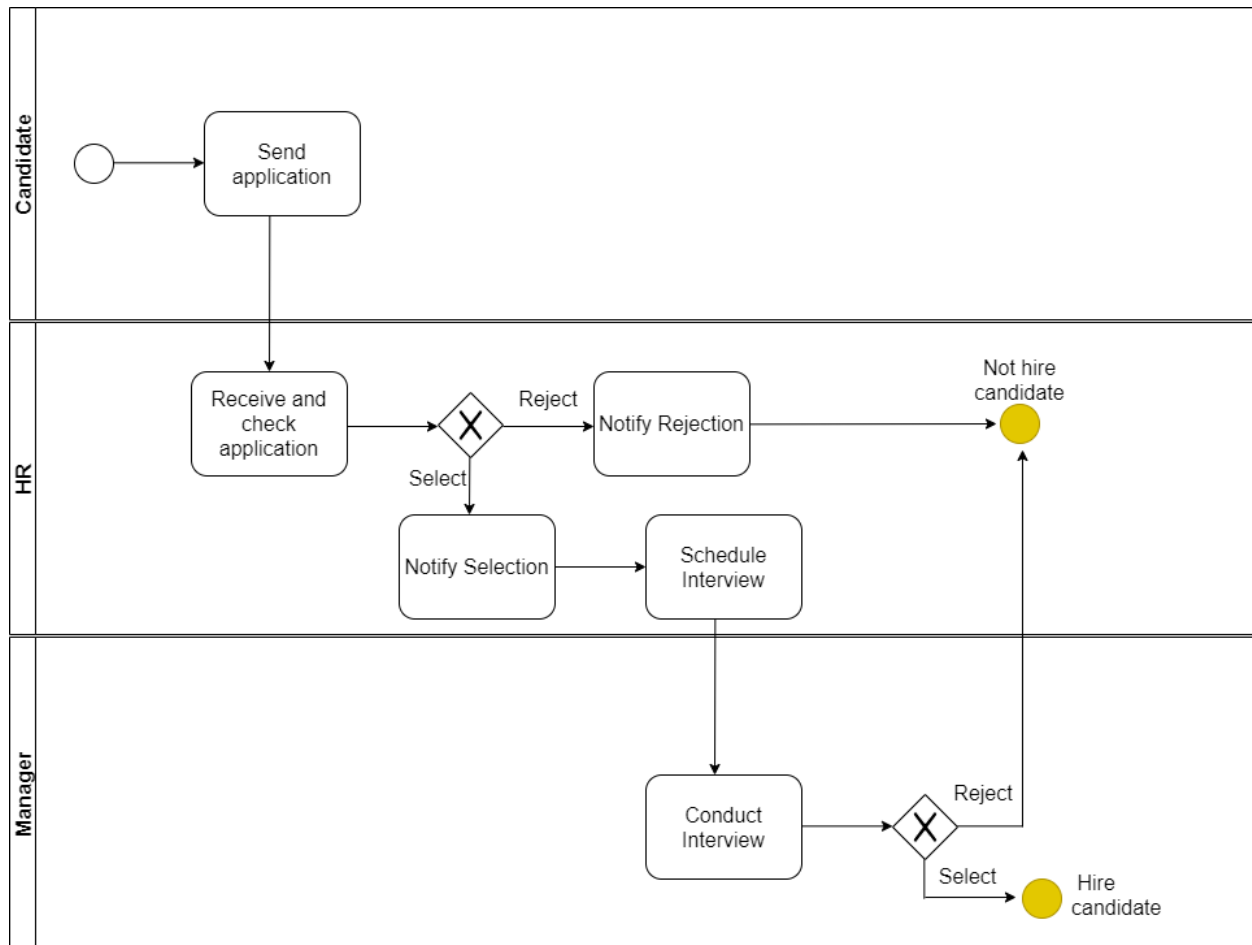


Figure 3.1.3.1: Recruiting Process of HR

Recruitment is an important step to build human resources for the business. Choosing the right person or the right person for the job position is very important. The BPMN below will briefly describe the recruitment process in most firms today.

Firstly, when the HR department receives recruitment demand for some vacancies, they need to plan carefully before hiring. After creating a detailed job description which describes the requirements of those positions clearly, they publicize it on social media or any channels. Secondly, HR will check carefully all applications they received to choose the suitable CVs to interview. At this step, in case they reject that candidate they will send a rejection letter for them to say thank you for their attention or a selection letter to announce them to prepare for the next phase. Then, those approved candidates

will go to the interview round with the managers of a department which they apply for. And finally the loop will be reduplicated as they reject or approve the applicants.

### 3.1.3.2 Evaluating Process BPMN

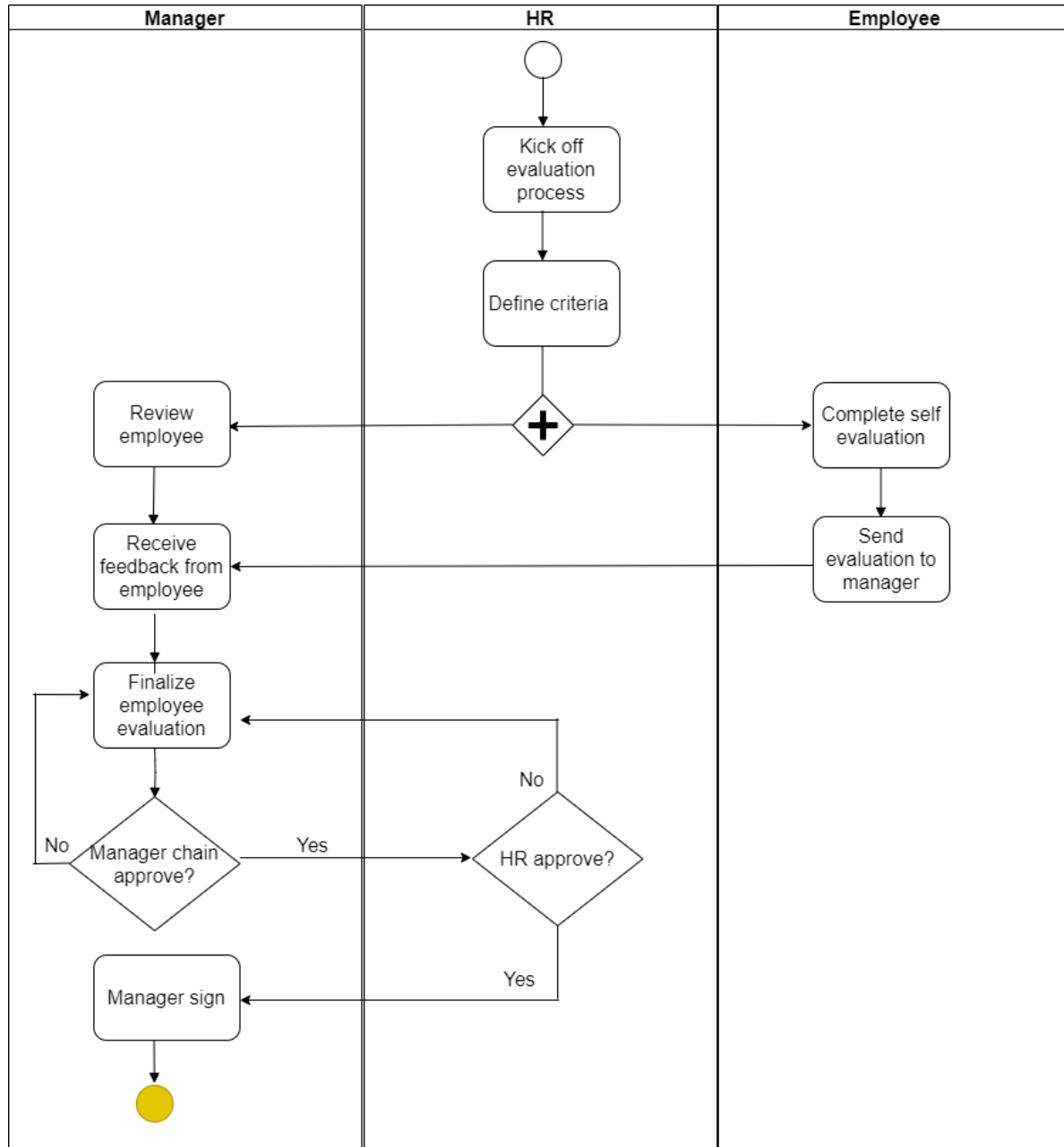


Figure 3.1.3.2 : Evaluating Process of HR

Evaluation or review annually is a vital part of the development process that the HR department has to manage. Because the evaluating process each year will help

employees and managers reflect on themselves every year and try hard next year to increase their income and company's revenue. The BPMN below will briefly describe the evaluating process.

Firstly, the HR department will plan for evaluation by defining suitable criteria. After sending an evaluation form to both managers and employees, managers and employees will self-evaluate the form. Then, the employee will send it to the manager to get approval and the manager will receive feedback from the employee simultaneously. Next, as the manager's chain approves all of the evaluation and feedback, they will be sent to the HR department to review again. Finally, HR approves them and sends again to the manager to sign and the process finishes here.

### 3.1.3.3. Off-boarding process BPMN

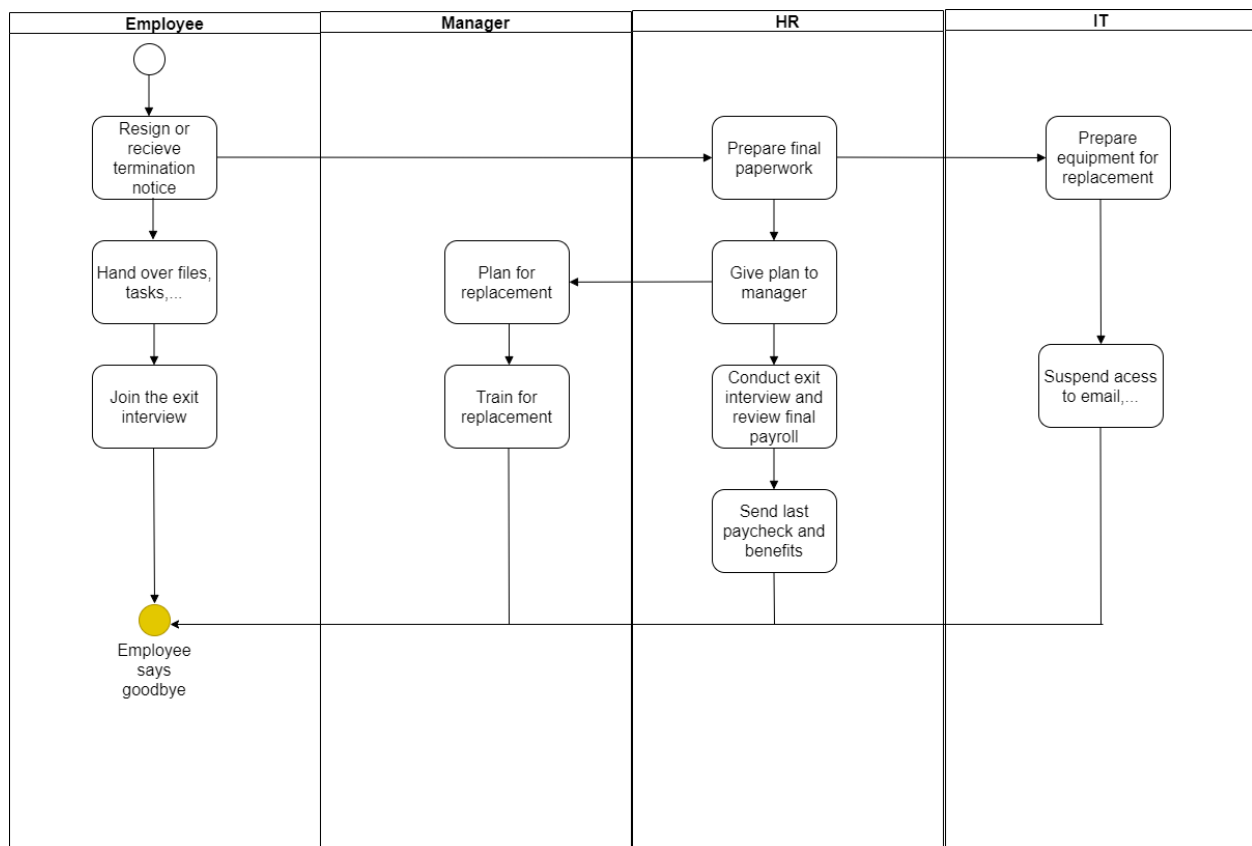


Figure 3.1.3.3: Off-boarding process



In addition to the human resource management, off-boarding process will occur when an employee wants to quit their job or when a termination notice sent from manager. The BPMN below will describe it clearly in relation to 3 departments.

First, as the HR department receives a resignation letter or termination notice from the employee, HR will prepare all paperwork related to that employee such as tax document, final paycheck,... After giving a successful plan to the manager for prepare a replacement plan, an exit interview will be conducted between HR, manager and employee to know the reason why they resigned and review the final paycheck. After confirmation, the HR department will send them their last paycheck and other benefits before they leave. Whereas, the IT department will prepare equipment for replacement and suspend their access to internal email or service. Finally, as employees receive the final paycheck and other benefits, the process completes here.

## 3.2 Data source and challenges

### 3.2.1 Description and meaning of the HR data

There are 7 tables and is described as follows:

#### 3.2.1.1 *HumanResource.Department*

Meaning: Lookup table containing the departments within the Adventure Works Cycles company.

*Table 3.2.1.1: HumanResource.Department table*

Entity	Data Type	Description
DepartmentID	smallint	Primary key for Department records. Identity, auto increment column
Name	nvarchar(50)	Name of department
GroupName	nvarchar(50)	Name of the group to which the department belongs.

ModifiedDate	datetime	Date and time the record was last updated.
--------------	----------	--

### 3.2.1.2 HumanResource.Employee

Meaning: A place to store employee information such as salary, department, and title.

Table 3.2.1.2: HumanResource.Employee

Entity	Data Type	Description
BusinessEntityID	int	Employee information such as salary, department, and title
NationalIDNumber	nvarchar(15)	Unique national identification number such as a social security number
LoginID	nvarchar(256)	Network login of employee
OrganizationNode	hierarchyid	Where the employee is located in corporate hierarchy
OrganizationLevel		The depth of the employee in the corporate hierarchy Computed: ([OrganizationNode].[GetLevel]())
JobTitle	nvarchar(50)	Work title such as Buyer or Sales Representative
BirthDate	date	Date of birth
MaritalStatus	nchar(1)	Marital status of employee M = Married, S = Single

Gender	nchar(1)	Gender of employee M = Male, F = Female
HireDate	date	The date the employee was contracted Employee hired on this date.
SalariedFlag	Flag:bit	Job classification. 0 = Hourly, not exempt from collective bargaining. 1 = Salaried, exempt from collective bargaining. Default: 1
VacationHours	smallint	Number of available vacation hours. Default:0
SickLeaveHours	smallint	Number of available vacation hours. Default:0
CurrentFlag	Flag:bit	0 = Inactive, 1 = Active Default:1
rowguid	uniqueidentifier	Having a role in administration - the project and speed of database exploitation ROWGUIDCOL number uniquely identifying the record. Used to support a merge replication sample. Default: newid()
ModifiedDate	datetime	Date and time the record was last updated

### 3.2.1.3 HumanResource.EmployeeDepartmentHistory

Meaning: Employee department transfers, employee's departmental rotation.

Table 3.2.1.3: HumanResource.EmployeeDepartmentHistory

Entity	Data Type	Description
BusinessEntityID	int	Employee identification number. Foreign key to Employee.BusinessEntityID
DepartmentID	smallint	Department in which the employee worked including currently. Foreign key to Department.DepartmentID.
ShiftID	tinyint	Identifies which 8-hour shift the employee works. Foreign key to Shift.Shift.ID.It is divided into 3 types (1-day; 2-evening; 3-night)
StartDate	date	Date the employee started work in the department
EndDate	date	Date the employee left the department. NULL = Current department
ModifiedDate	datetime	Date and time the record was last updated. Default: getdate()

### 3.2.1.4 HumanResource.EmployeePayHistory

Meaning: Employee pay history, change in employee's salary.

Table 3.2.1.4: HumanResource.EmployeeDepartmentHistory

Entity	Data Type	Description
--------	-----------	-------------

BusinessEntityID	int	Employee identification number. Foreign key to Employee.BusinessEntityID
RateChangeDate	datetime	Date the change in pay is effective
Rate	money	Salary hourly rate
PayFrequency	tinyint	A number of payments in a payment period 1 = Salary received monthly, 2 = Salary received biweekly
ModifiedDate	datetime	Date and time the record was last updated.

### 3.2.1.5 HumanResource.JobCandidate

Meaning: A place to store résumés submitted to Human Resources by job applicants.

Table 3.2.1.5: HumanResource.JobCandidate

Entity	Data Type	Description
JobCandidateID	int	Primary key for JobCandidate records. Identity / Auto increment column
BusinessEntityID	int	Employee identification number if applicant was hired. Foreign key to Employee.BusinessEntityID.
Resume	xml	Résumé in XML format.,...
ModifiedDate	datetime	Date and time the record was last updated.

### 3.2.1.6 HumanResource.Shift

Meaning: A place to store work shift lookup table.

Table 3.2.1.6: HumanResource.Shift

Entity	Data Types	Description
ShiftID	tinyint	Primary key for Shift records. Identity / Auto increment column
Name	nvarchar(50)	Shift description.
StartDate	time(7)	The time to start the shift
EndDate	time(7)	The time to end the shift
ModifiedDate	datetime	Date and time the record was last updated

### 3.2.1.7 Person.Person

Meaning: Human beings involved with AdventureWorks: employees, customer contacts, and vendor contacts.

Table 3.2.1.7: Person.Person

Entity	Data Types	Description
BusinessEntityID	int	Primary key for Person records.
PersonType	nchar(2)	Primary type of person: SC = Store Contact, IN = Individual (retail) customer, SP = Sales person, EM = Employee (non-sales), VC = Vendor contact, GC = General contact
NameStyle	NameStyle:bit	0 = The data in FirstName and LastName are stored in western style (first name, last name) order. 1 = Eastern style (last name, first name) order. Default: 0
Title	nvarchar(8)	A courtesy title. For example, Mr. or Ms
FirstName	Name:nvarchar(50) )	First Name of person
MiddleName	Name:nvarchar(50) )	Middle name or middle initial person

LastName	Name:nvarchar(50) )	Last name of person
Suffix	nvarchar(10)	Surname suffix. For example, Sr. or Jr
EmailPromotion	int	0 = Contact does not wish to receive e-mail promotions, 1 = Contact does wish to receive e-mail promotions from AdventureWorks, 2 = Contact does wish to receive e-mail promotions from AdventureWorks and selected partners. Default: 0
AdditionalContactInfo	xml	Additional contact information about the person stored in xml format
Demographics	xml	Personal information such as hobbies, and income collected from online shoppers. Used for sales analysis.
rowguid	uniqueidentifier	Having a role in administration - the project and speed of database exploitation  ROWGUIDCOL number uniquely identifying the record.



		Used to support a merge replication sample.
ModifiedDate	datetime	Date and time the record was last updated

### 3.2.2 Physical Data Model (PDM)

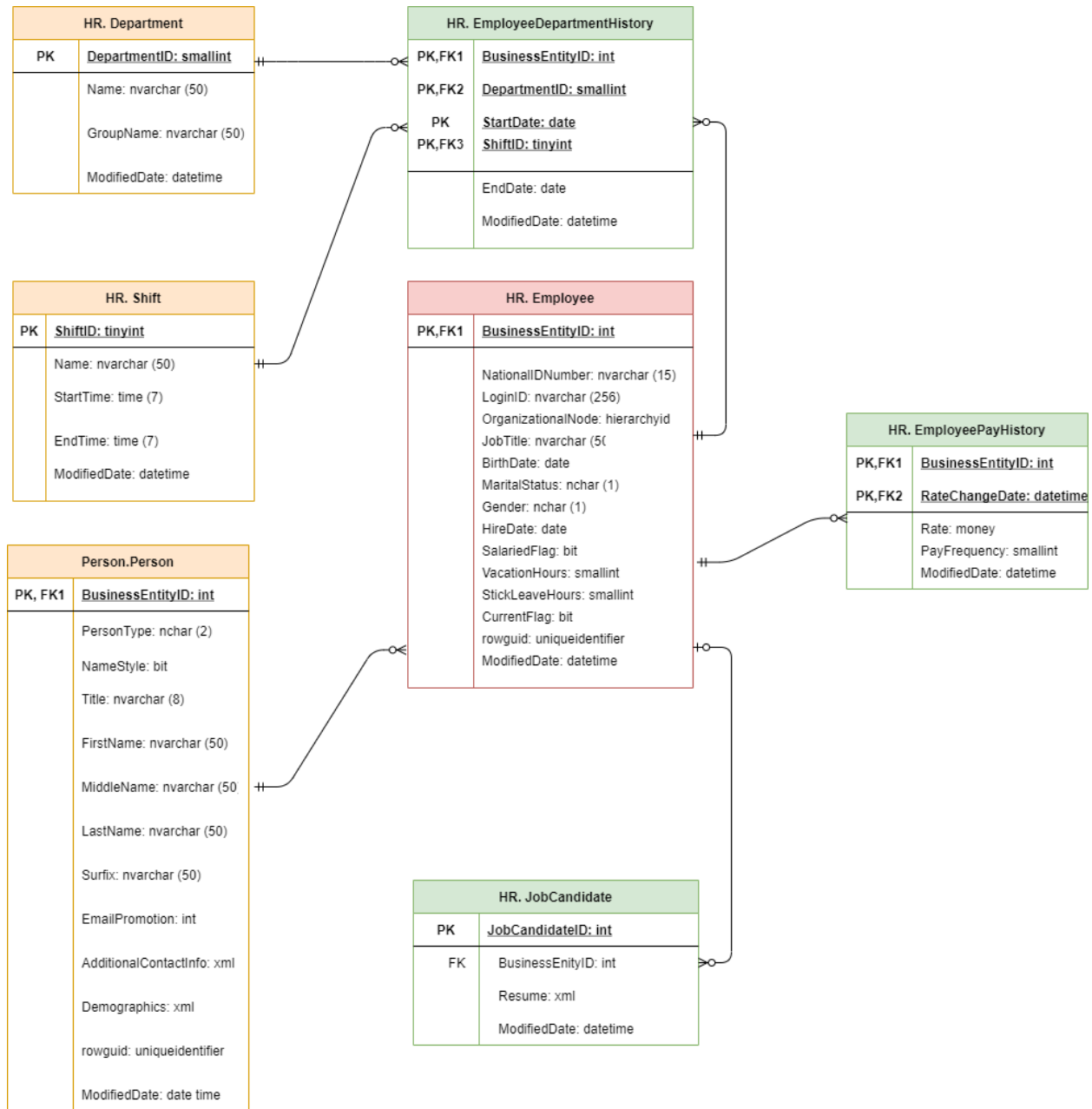


Figure 3.2.2: Physical Data Model (PDM) of HR

### 3.2.3 Challenges of data HR

Data Source is not enough data to the recruitment profession:

- Information about salary of employees to easily calculate detailed pay.
- Information about commission of employees to easily calculate detailed pay.

- Information person about candidates to easily find human resources when needed.
- Information about reasons for leaving employees, thereby giving measures to retain good people or having policies to retain good people.

### **3.3. Business Requirements Analysis of HR**

#### **3.4.1. Development process**

Managing and developing employees is the most important part in Human Resource roles. In addition to recruitment, the HR department also needs to find solutions to retain potential employees by some good policies or interesting events and so on. A report about workers' demographics such as Age group, Job Title, Gender split, Area, Marital status,... can help the company tackle this problem.

Because the Workers Demographics can help the company solve many problems to develop and retain employees such as:

- Celebrating birthday parties.
- Finding our employees' niche to increase hiring in those niches.
- Finding more insights about our employees to have suitable plans for each of them.
- Controlling the retirement process to prepare for a new hiring campaign.

#### **3.4.2. Employee perf the operation**

Employer Performance Reporting: This is an extremely important type of report that helps leadership see how HR performance impacts business performance, while also answering questions about structural positioning, salary structure, employee function structure of the unit is appropriate or not.

#### **3.4.3. The process of timekeeping, payroll at HR**

##### **3.4.3.1 . Timekeeping**

- Quick and accurate time attendance management for payroll and statistics.  
Compatible with all types of timekeeper. Timekeeping by the hour, by day or by month: control of absence, entry / exit, late / early departure, days off, shift, overtime ...

- Support timekeeping in many forms: time (shift work), products, overtime ...
- Detailed management of the number of days off, sick, ... of employees.
- Can use a time attendance calculator by employee code.
- If the company has shift schedules that need to be swapped, the system provides a work shift tool, an alternate schedule if there is a shift change, the shift control becomes automatic because it has a rule and Support tools are software.
- Import information from timekeeper, transfer from Excel file from timekeeping department, to help HR department without re-input, avoid errors and no time to redo timekeeping.

#### *3.4.3.2. Calculate salary, salary, bonus and other income*

- Each business has its own way of calculating salary, each department may also have a different way of calculating salary. Every month, HR department just presses the payroll button, the system will automatically calculate and give the correct payroll for each employee, and easily print out payroll, personal pay stubs, and send email notifications. Salary or view salary information on the web.
- Salary support in many forms such as time, product, sales ...• Import payroll information from excel file.
- The system will also automatically retrieve data from the payroll to calculate personal income tax with the salary taxable salary defined by the Enterprise and will also automatically output the combined payroll and detailed payroll for each employee.
- Assisting staff in making requests for advances and paying advances online.
- Automatically warn employees who are overdue but have not paid their travel expenses.

#### *3.4.5. Departmental change process*

Filter out which employees change departments. Find common ground and make predictions with employees about to change departments. Besides, giving the employee's

satisfaction with the job they are doing, the salary, the position is based on (org level, job title, rate, time off hours, ...)

### **3.5. IT requirements Analysis (IT & Infrastructure)**

In enterprises, these BI solutions will be essential to the HR department in solving problems when employees quit their jobs. The application of these solutions will help businesses save a lot of costs, resources as well as effectively solve the problem of resignation of employees. It is expected that the project will be completed after 4 months of implementation.

It can be said that: Data management, Data query and Data analysis which technical requirements are very necessary to provide warning about the situation.

Data management allows users to monitor, ensure that all analytical processes and insights are in compliance with business policies and procedures, ensure data integrity and minimize risks. or has many real sources. It is impossible not to mention "Extract, Transform, Load (ETL) Tools", OLAP and Multidimensional Analysis.

Structured Query Language (SQL), from the database extracts information and formats it for consumption and analysis. Data queries can perform calculations, automate tasks, or dig deeper through data mining, helping to uncover hidden trends and relationships between data points.

Data analysis turns raw information into useful insights, helping businesses maximize the value of their data for better business decisions. Data analytics helps users extract value from operational information and gives them insight into their business. There are four main types of analysis: descriptive, predictive, indicative and diagnostic. It includes Ad-Hoc analysis, Analysis by groups and segments, Cluster analysis, Analyze the scenario and what if it happens, Statistical analysis and regression, Time series analysis and prediction,....

This report helps management see how HR performance impacts business, answers the question of hierarchy of positions in the business, salary structure, qualifications and

capabilities of employees. the member has responded or not. This report is often based on the following criteria:

- Number of products / receipt
- Number of products / salary costs

To do this, the HR department needs to share the production and business results of each unit, department and the whole company.

### 3.6 Comparative Analysis of BI and Data Visualization Tools

#### 3.6.1 Surveying and evaluation

*Table 3.6.1: Comparative BI tool*

Tools	Price ( based on user)	Functions	Time	Cloud
Excel	-Starting at \$1.99/user/month. A free version does not offer a free trial. -Office 365 Personal: \$6.99/month or \$69.99/year -Office 365 Home: \$9.99/month or \$99.99/year -Office Home & Student 2019: \$149.99 one-time Office 365 Business: \$8.25/user/month (annual commitment required) Office 365 Business Premium:	Designed for businesses of all sizes, it is a spreadsheet solution that helps manage and organize data using templates, perform calculations, create analytical charts, and more. -Count and Sum: The most used functions in Excel are the functions that count and sum. You can count and sum based on one criteria or multiple criteria. -Logical: Learn how to use Excel's logical functions, such as IF, AND, OR and NOT. -Cell References: Cell references in Excel are very important. Understand the difference between relative,	Easy to set up Can use immediately	-Office 365 is a cloud-based product suite containing Microsoft programs for office use that can be run locally and synchronized to cloud storage. With the help of Office 365, you can work from anywhere and share the work documents with your colleagues worldwide. It helps individuals as

	<p>\$12.50/user/month (annual commitment required) Office 365 Business Essentials: \$5/month (annual commitment required)</p>	<p>absolute and mixed reference, and you are on your way to success.</p> <p>-Date &amp; Time: To enter a date, use the "/" or "-" characters. To enter a time, use the ":" (colon).</p> <p>-Text: Excel has many functions to offer when it comes to manipulating text strings.</p> <p>-Lookup &amp; Reference: Learn all about Excel's lookup &amp; reference functions, such as VLOOKUP, HLOOKUP, MATCH, INDEX and CHOOSE.</p> <p>-Financial: This chapter illustrates Excel's most popular financial functions.</p> <p>-Statistical: An overview of some very useful statistical functions in Excel.</p> <p>-Round: This chapter illustrates three functions to round numbers in Excel. ROUND, ROUNDUP and ROUNDDOWN.</p> <p>-Formula Errors: This chapter teaches you how to deal with some common formula errors.</p>	<p>well as businesses to easily work on different documents.</p> <p>-Excel is also part of the programs that are shipped with Office 365. You can use the powerful features found in desktop versions of Office 365 Excel in the cloud based version too.</p>
--	---	--	---

		-Array Formulas: This chapter helps you understand array formulas in Excel. Single cell array formulas perform multiple calculations in one cell.		
Tableau	<p>-Personal use: Tableau Creator's price is US \$ 840 / Year (including license for Tableau Desktop + Tableau Prep Builder + License for Tableau Online or Tableau Server) Organization, business</p> <p>-On-Premise : Tableau Creator (\$ 840 / User / Year) Tableau Explorer price (\$ 420 / User / Year) and you need to buy at least 5 User</p> <p>Tableau Viewer (144 \$ US / User / Year) and you need to purchase at least 100 Users.</p> <p>-On-Line : Tableau Creator (\$ 840 / User / Year) Tableau Explore Online (\$ 504 US /</p>	<p>Tableau offers robust visual analytics and data drilling tools across all of its products. We've listed the main features that make this possible:</p> <ul style="list-style-type: none"> <li>- Numerous data connections</li> <li>- Live and in-memory data</li> <li>- Secure collaboration</li> <li>- Mobile optimized designs</li> <li>- Embedded dashboards</li> <li>- Drag-and-drop integration</li> <li>- Unlimited data connectors - Relationships</li> <li>- Metrics</li> <li>- Set control</li> </ul>	Ideally you will want your project to take between one and three months, depending on the level of complexity. After that, it's a case of expanding internally using the lessons learned from your pilot project.	Tableau can connect to any data source on premise or in the cloud



	<p>User / Year) and you also need to buy at least 5 users</p> <p>Tableau Viewer Online (\$ 180 US / User / Year) and you also need to purchase at least 100 Users. --&gt; This price will be suitable for large businesses instead of small and medium businesses"</p>			
Qlik	<p>Qlik Sense Business: \$30 / user / month (billed annually).</p> <p>Qlik Sense Enterprise - a multi cloud edition: price is no longer available (contact sales the vendor to know more).</p>	<p>Qlik has many dynamic features, some of them are:</p> <ul style="list-style-type: none"> <li>- Insight Advisor: provide chart suggestion, recommendations to streamline the data preparation process.</li> <li>- Self-service creation: Drag-and-drop dashboard, charts without query.</li> <li>- Centralize Management - Data Integration: not "hide" data, just "grayed out" instead of discarding.</li> <li>- Interactive Visualization.</li> <li>- Responsive Design: adjust automatically depending on your device. For each type, there are some difference in addition to the shared-features above:</li> </ul>	<p>-Qlik Sense has both cloud-hosted and on-premise solutions, depending on the company's needs.</p> <p>-With its cloud-based solution, users can simply sign up for a free trial.</p> <p>-Qlik offers free and paid training programs,</p>	

		<p>- Qlik Sense Business: + Up to 5 shared spaces. + 1.25GB in-memory app size limit.</p> <p>- Qlik Sense Enterprise: + Unlimited shared spaces. + 5GB in-memory app size limit.</p>	including self-paced learning and instructor-led training.	
Power BI	<p>-Power BI Pro (included in Microsoft 365 E5): \$ 9.99 (monthly price per user).</p> <p>-Power BI Premium (per capacity): \$ 4,995 (monthly price for dedicated cloud computing and storage resources with an annual subscription).</p> <p>-Power BI Premium per user: \$ 20.00 user / month (annual commitment)</p> <p>-Power BI free: \$ 0.00 per year</p>	<p>There are 3 feature categories:</p> <p>Collaboration&amp;Analytics; Data prep, modeling&amp;visualization; Governance&amp;Administration that depend on each type of Power BI you use:</p> <p>- Power BI free: Includes the ability to use the Power BI Service for personal use. No collaboration or sharing options.</p> <p>- Power BI pro: Includes Power BI Free features and ability to share and collaborate with colleagues in the Power BI Service. Includes additional features such as usage metrics.</p> <p>- Power BI premium: Complements Power BI Pro features by supporting higher data limits, enhanced performance, and ability to distribute to report consumers with Power BI Free licensing. On-premises report publishing. APIs for</p>	It will be completed around 30 days depending on your employees' ability and budget and your urgency.	Power BI is now available for three separate national clouds. These clouds offer the same levels of security, privacy, compliance, and transparency as the global version of Power BI, combined with a unique model for local regulations on service delivery, data residency, access, and control. They are:

- US Government
- Germany
- China

		<p>custom application integration.</p> <ul style="list-style-type: none"> <li>- However, Power BI free intended to be individual only, so we just discuss 3 types clearly.</li> <li>- In particular: <ul style="list-style-type: none"> <li>+ Model size limit: Power BI pro (1 GB), Power BI Premium per user (100GB), Power BI Premium per capacity (400GB).</li> <li>+ Refresh rate: Power BI pro (8/day), Power BI Premium per user (48/day), Power BI premium per capacity (48/day).</li> <li>+ Maximum storage: Power BI pro (10GB/user), Power BI Premium per user and per capacity(100TB).</li> </ul> </li> </ul>		
SSAS	<p>"SSAS is a part of SQL Server.</p> <ul style="list-style-type: none"> <li>- SQL Server Enterprise: \$5,434/ year</li> <li>- SQL Server Standard: \$1,418/ year</li> </ul> <p>But you can download a free full-feature</p>	<ul style="list-style-type: none"> <li>- Querying cubes through multidimensional expressions</li> <li>- Adaptable creation and management of data models</li> <li>- Works perfectly with different wizards and designers</li> <li>- Faster response to queries owing to the aggregation of facts</li> </ul>		Cannot publish to the cloud

	evaluation version"	<ul style="list-style-type: none"> <li>- Competency to create a robust data model for enhanced reporting</li> <li>- All-time accessibility, at any place via the Internet</li> <li>- Assists in creating MOLAP, ROLAP and HOLAP architecture</li> </ul>		
SAP Data Analysis	<ul style="list-style-type: none"> <li>- 30-DAY TRIAL (\$0 per user):</li> <li>+ Business Intelligence functionality only</li> <li>+ Basic data source imports (excel, csv, and Google Drive)</li> <li>+ 1 on 1 training with a product coach</li> <li>- Business (up to 20 user ~ \$21 per user/month):</li> <li>+ Business Intelligence functionality only</li> <li>+ Basic data source connections (excel, csv, and Google Sheets)</li> <li>+ Enterprise-level data source connections (SAP HANA, SQL, OData Services, and more)</li> <li>+ 1 on 1 training with a product coach + Team</li> </ul>	<ul style="list-style-type: none"> <li>- <b>Data Warehousing:</b> SAP data warehouse cloud combines data from disparate sources in real time and processes it into a single secure, accessible database for analysis. It simplifies data access by storing and delivering rich semantic information via the cloud.</li> <li>- <b>Augmented Insights:</b> Users can discover reasons for past results by identifying patterns in the data and use these insights to accelerate growth. AI and machine learning simulate future scenarios and let them predict outcomes based on specified inputs in data modeling environments.</li> <li>- <b>Customized Environment:</b> SAP HANA provides multiple tenant databases that ensure completely unique instances of the software, even within an organization. It integrates data from a variety of sources and offers advanced customization options to meet organizations' unique</li> </ul>		<p>The SAP Analytics Cloud solution combines BI, augmented and predictive analytics, and planning capabilities into one cloud environment.</p> <p>As the analytics layer of SAP's Business Technology Platform, it supports advanced analytics enterprise-wide</p>

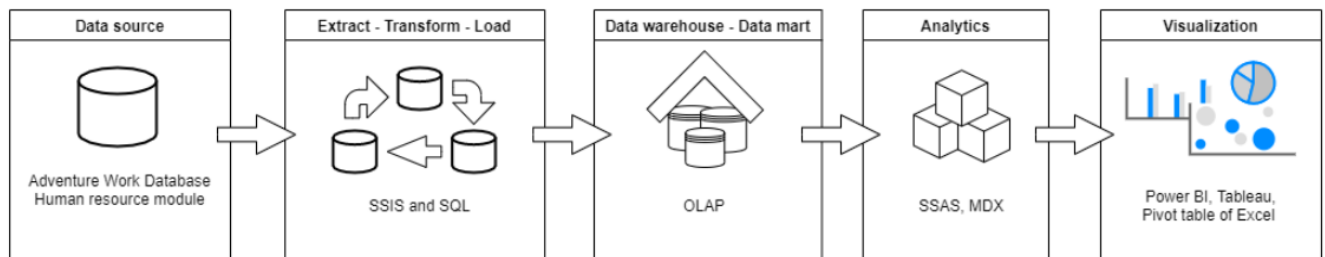
	<p>collaboration features</p> <p>+ 24/7 SAP Support</p> <p>- Enterprises (Custom): Contact for details</p> <p>+ All Business plan features</p> <p>+ Unlimited users</p> <p>+ Optional upgrade for planning and predictive analytics</p> <p>+ Personal customer success manager</p>	<p>needs.</p> <p>- <b>Real-Time Processing:</b> SAP BW/4HANA ensures real-time processing of high-volume data via an intuitive user interface. Users can perform simplified data modeling to uncover insights and make predictions.</p> <p>- <b>Centralized Planning:</b> SAP Analytics Cloud and SAP integrated business planning offer real-time scenario simulations, predictive analytics, embedded analytics and reporting functions that connect business planning processed across locations in the cloud.</p>		
Oracle Data Analysis	<p>-Small business (less than 75 people): varies from \$ 25,000 - \$100,000 depending on customization levels and number of users.</p> <p>-Big business: over \$ 1 million for ERP software.</p>	<p><b>-Scalability and Performance:</b> Features like Real Application Clustering and Portability make an Oracle database scalable according to the usage</p> <p><b>-Availability:</b> Real-time applications require high data availability. High performing computing environments are configured to provide all-time data availability.</p> <p><b>-Backup and Recovery:</b> Its layout complete recovery features to recover data from almost all kinds of failures. In case of failure, the</p>	<p>Oracle has the most stable project deployment times. Oracle will take about 2 months to deploy with a deployment cost of \$ 1 million.</p>	<p>-Providing an advanced infrastructure that allows businesses to scale their IT platforms as they grow. Safe integration and simultaneous operation between different services.</p> <p>-Performing operations faster to bring</p>

		<p>database needs to be recovered within no time for high availability.</p> <p><b>-Security:</b> Securing the data is always the top priority. Oracle provides mechanisms to control data access and usage.</p>		<p>high work efficiency. Users can access anytime 24/7 and anywhere, just have a means (laptop, phone) with network connection, helping to improve productivity significantly.</p> <p><b>-Technology</b> is strong, multi-layered defense spanning IaaS, PaaS and SaaS services, extending security to the network, hardware, chip, operating system, storage, and application layers.</p> <p><b>-Data centers</b> are built around multi-layered physical defense systems</p>
--	--	---	--	--

				designed to allow authorized people to enter and avoid those without access.
--	--	--	--	--

### 3.6.2 Proposing BI solution for the project

After surveying the current state of AdventureWorks company, especially the current situation in the company's HR department, we found that in terms of general structure and relationships among HR departments are quite good, well managed the list of employees who are working or retired, pay history, list of potential candidates,... However, when going into each activity of the department, there are still shortcomings in the management, recruitment, training, salary, ... One of the factors causing the inadequacy is information and information management. Therefore, based on knowledge about Business Intelligence and BI tools, we propose a BI solution for the project like the following 5 parts in the picture below.



*Figure 3.6.2: BI solution*

## CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA

In this chapter, the topic will present techniques and methods to build a database on a data analysis system to support business performance evaluation.

### 4.1. Designing Data Warehouse

#### 4.1.1. Bus matrix

Conformed dimensions are common, standardized, master dimensions that are managed once in the extract, transformation, and load (ETL) system and then reused by multiple fact tables. Conformed dimensions deliver consistent descriptive attributes across dimensional models. They support the ability to drill across and integrate data from multiple business processes. Finally, reusing conformed dimensions shortens the time-to-market by eliminating redundant design and development efforts.

The associated Enterprise Data Warehouse Bus Matrix, shown below, is a key design tool representing the organization's core business processes and associated dimensionality. It's the architectural blueprint providing the top-down strategic perspective to ensure data in the DW/BI environment can be integrated across the enterprise, while agile bottom-up delivery occurs by focusing on a single business process at a time.

Business Process	Common dimensions						
	Date	Department History	Pay History	Manager	Rate	Shift	TimeOff Hours
Recruitment	x	x		x			
Work time management	x	x	x	x	x	x	x
Absence day management	x	x	x	x	x	x	x
Payroll for staff	x	x	x		x	x	x
Changing departments for employees	x	x	x	x	x		
Employee leave management	x	x	x	x	x	x	x
Salary forecast	x	x	x	x	x	x	x
Forecast of employees about to leave work	x	x	x	x	x	x	x
Forecast of the number of employees to hire	x	x	x	x	x		

*Figure 4.1: Bus Matrix of HR*



#### 4.1.2 Master Data

Object	Description
Employee name	The information of employee
Department	The information of department
Job title	Work title such as Buyer or Sales Representative.
Hire date	Employee hired on this date.
Shift	The information of shift

#### 4.1.3 Transaction Data

Object	Description
RateChangeDate	Date the change in pay is effective
Rate	Salary hourly rate
TimeOffHours	Total
StartDate	Date the employee started work in the department.
EndDate	Date the employee finished work in the department.

#### 4.1.4. Dimension table

##### 4.1.4.1 DimEmployee

Meaning: The table describes detailed information of each employee in the company: name, date of birth, telephone number, address, number of days off, ...

```

SELECT B.BusinessEntityID, A.FirstName, A.MiddleName, A.LastName, C.PhoneNumber, B.BirthDate, B.Gender,
B.MaritalStatus,D.EmailAddress, F.AddressLine1, F.City, G.Name StateProvince, H.Name CountryRegion, B.OrganizationLevel,
B.JobTitle, B.NationalIDNumber, B.HireDate, B.VacationHours,B.SickLeaveHours

FROM Person.Person A INNER JOIN HumanResources.Employee B on A.BusinessEntityID = B.BusinessEntityID

INNER JOIN Person.PersonPhone C on B.BusinessEntityID = C.BusinessEntityID

INNER JOIN person.EmailAddress D on C.BusinessEntityID =
D.BusinessEntityID

INNER JOIN Person.BusinessEntityAddress E on E.BusinessEntityID =
D.BusinessEntityID

INNER JOIN Person.Address F on F.AddressID = E.AddressID

INNER JOIN Person.StateProvince G ON
G.StateProvinceID=F.StateProvinceID

INNER JOIN Person.CountryRegion H ON
H.CountryRegionCode=G.CountryRegionCode

GROUP BY B.BusinessEntityID, A.FirstName, A.MiddleName, A.LastName, C.PhoneNumber, B.BirthDate, B.Gender,
B.MaritalStatus,D.EmailAddress, F.AddressLine1, F.City, G.Name, H.Name, B.OrganizationLevel, B.JobTitle,
B.NationalIDNumber, B.HireDate, B.VacationHours,B.SickLeaveHours

```

## Create Table DimEmployee


	Column Name	Data Type	Allow Nulls
	EmployeeKey	int	<input type="checkbox"/>
	BusinessEntityID	int	<input type="checkbox"/>
	FirstName	nvarchar(50)	<input type="checkbox"/>
	MiddleName	nvarchar(50)	<input checked="" type="checkbox"/>
	LastName	nvarchar(50)	<input type="checkbox"/>
	PhoneNumber	nvarchar(25)	<input type="checkbox"/>
	EmailAddress	nvarchar(50)	<input checked="" type="checkbox"/>
	AddressLine1	nvarchar(60)	<input type="checkbox"/>
	AddressLine2	nvarchar(60)	<input checked="" type="checkbox"/>
	City	nvarchar(30)	<input type="checkbox"/>
	JobTitle	nvarchar(50)	<input type="checkbox"/>
	BirthDate	date	<input type="checkbox"/>
	Gender	nchar(1)	<input type="checkbox"/>
	MaritalStatus	nchar(1)	<input type="checkbox"/>
	NationalIDNumber	nvarchar(15)	<input type="checkbox"/>
	HireDate	date	<input type="checkbox"/>
	VacationHours	smallint	<input type="checkbox"/>
	SickLeaveHours	smallint	<input type="checkbox"/>
	ModifiedDate	datetime	<input type="checkbox"/>

Figure 4.1.1.1: Table DimEmployee

#### 4.1.4.2 DimDepartment

Meaning: The table describes detailed information of each department.

```
SELECT A.DepartmentID, A.Name, A.GroupName, A.ModifiedDate  
FROM HumanResources.Department A
```

Create Table DimDepartment


	Column Name	Data Type	Allow Nulls
	DepartmentKey	int	<input type="checkbox"/>
	DepartmentID	smallint	<input type="checkbox"/>
	Name	nvarchar(50)	<input type="checkbox"/>
	GroupName	nvarchar(50)	<input type="checkbox"/>
	ModifiedDate	datetime	<input type="checkbox"/>

Figure 4.1.1.2: Table DimDepartment

#### 4.1.4.3 DimShift

Meaning: The table describes detailed information of shift.

```
SELECT A.ShiftID, A.Name, A.StartTime, A.EndTime, A.ModifiedDate  
FROM HumanResources.Shift A
```

Create Table DimShift


	Column Name	Data Type	Allow Nulls
	ShiftKey	int	<input type="checkbox"/>
	ShiftID	tinyint	<input type="checkbox"/>
	Name	nvarchar(50)	<input type="checkbox"/>
	StartTime	time(7)	<input type="checkbox"/>
	EndTime	time(7)	<input type="checkbox"/>
	ModifiedDate	datetime	<input type="checkbox"/>

Figure 4.1.1.3: Table DimShift

#### 4.1.4.4 DimTime

Meaning: The table describes detailed information of time.

```
SELECT CAST(YEAR(PK_Date) AS nvarchar)+ RIGHT('0'+CAST(MONTH(PK_Date) AS nvarchar),2)
```

```

+RIGHT('0'+CAST(DAY(PK_Date) AS nvarchar),2) Timekey,

YEAR(PK_Date) Year,

RIGHT('0'+CAST(MONTH(PK_Date) AS nvarchar),2) Month,

RIGHT('0'+CAST(DAY(PK_Date) AS nvarchar),2) Day,

DATENAME(WEEKDAY,PK_Date) DateName,

DATEPART(WEEKDAY,PK_Date) DayNumberOfWeek,

DATEPART(DAY,PK_Date) DayNumberOfMonth,

DATEPART(DAYOFYEAR,PK_Date) DayNumberOfYear,

DATEPART(WEEK,PK_Date) WeekNumberOfYear,


DATEPART(QUARTER,PK_Date) Quarter

INTO DimTime

```

FROM Time

## Create Table DimTime

	Column Name	Data Type	Allow Nulls
	Timekey	int	<input type="checkbox"/>
	Year	int	<input checked="" type="checkbox"/>
	Month	nvarchar(2)	<input checked="" type="checkbox"/>
	Day	nvarchar(2)	<input checked="" type="checkbox"/>
	DateName	nvarchar(30)	<input checked="" type="checkbox"/>
	DayNumberOfWeek	int	<input checked="" type="checkbox"/>
	DayNumberOfMonth	int	<input checked="" type="checkbox"/>
	DayNumberOfYear	int	<input checked="" type="checkbox"/>
	WeekNumberOfYear	int	<input checked="" type="checkbox"/>
	Quarter	int	<input checked="" type="checkbox"/>

*Figure 4.1.1.4: Table DimTime*

## 4.1.5 Fact table

### 4.1.5.1 FactPayHistory

Meaning: Salary calculation table of employees according to the criteria: rate, sick leave hours, vaction hours,.....


```
SELECT A.BusinessEntityID, c.DepartmentID, b.ShiftID,
```

```

        CAST(YEAR(a.RateChangeDate) AS nvarchar)+ RIGHT('0' +CAST(MONTH(a.RateChangeDate) AS
nvarchar),2) + RIGHT('0' +CAST(DAY(a.RateChangeDate) AS nvarchar),2) AS RateChangeDate,
        a.Rate,a.PayFrequency
from HumanResources.EmployeePayHistory a left join HumanResources.EmployeeDepartmentHistory
b on a.BusinessEntityID = b.BusinessEntityID
left join HumanResources.Department c on b.DepartmentID = c.DepartmentID
INNER JOIN HumanResources.Shift D on D.ShiftID = b.ShiftID
WHERE (a.RateChangeDate = b.StartDate and b.EndDate is null)
or (a.RateChangeDate <= b.StartDate and b.EndDate is not null)
or (a.RateChangeDate >= b.StartDate and b.EndDate is null)
or (a.RateChangeDate = b.EndDate)

```

## Create Table FactPayHistory

	Column Name	Data Type	Allow Nulls
	FactPayHistoryKey	int	<input type="checkbox"/>
	BusinessEntityKey	int	<input type="checkbox"/>
	DepartmentKey	int	<input type="checkbox"/>
	ShiftKey	int	<input type="checkbox"/>
	RateChangeDate	int	<input checked="" type="checkbox"/>
	Rate	money	<input type="checkbox"/>
	PayFrequency	tinyint	<input type="checkbox"/>
			<input type="checkbox"/>

*Figure 4.1.2.1: Table FactPayHistory*

## 4.1.5.2 FactDepartmentHistory


Meaning: Store employee's department transfer process.

```

SELECT A.BusinessEntityID, E.DepartmentID, b.ShiftID,
        (A.SickLeaveHours +A.VacationHours) TimeOffHours, CAST(YEAR(B.StartDate) AS
nvarchar)+ RIGHT('0' +CAST(MONTH(B.StartDate) AS nvarchar),2) + RIGHT('0'
+CAST(DAY(B.StartDate) AS nvarchar),2) AS StartDate,
        CAST(YEAR(B.EndDate) AS nvarchar)+ RIGHT('0' +CAST(MONTH(B.EndDate) AS
nvarchar),2) + RIGHT('0' +CAST(DAY(B.EndDate) AS nvarchar),2) AS EndDate
FROM HumanResources.Employee A inner join HumanResources.EmployeeDepartmentHistory B on
A.BusinessEntityID = B.BusinessEntityID
INNER JOIN HumanResources.Shift D on D.ShiftID = B.ShiftID
INNER JOIN HumanResources.Department E on E.DepartmentID = b.DepartmentID

```

## Create Table FactPayHistory

	Column Name	Data Type	Allow Nulls
	FactDepartmentHistoryK...	int	<input type="checkbox"/>
	BusinessEntityKey	int	<input type="checkbox"/>
	DepartmentKey	int	<input type="checkbox"/>
	ShiftKey	int	<input type="checkbox"/>
	TimeOffHours	smallint	<input checked="" type="checkbox"/>
	StartDate	nvarchar(34)	<input checked="" type="checkbox"/>
	EndDate	nvarchar(34)	<input checked="" type="checkbox"/>
			<input type="checkbox"/>

*Figure 4.1.2.2: Table DepartmentHistory*

#### 4.1.6. Data Warehouse model

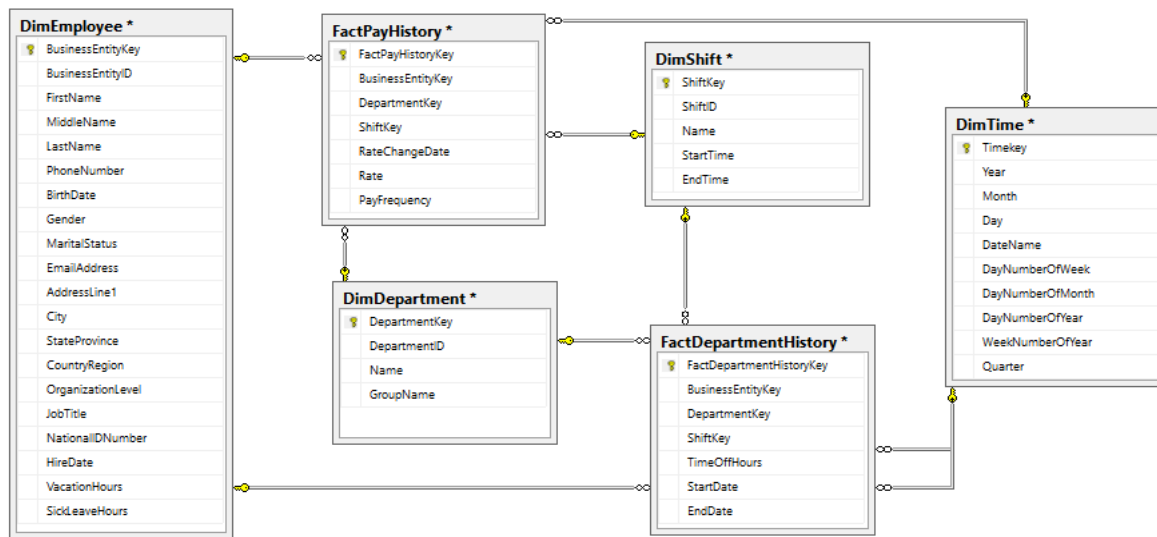


Figure 4.1.5: Data Warehouse model of HR

## 4.2. ETL processes

### 4.2.1. Dimension Table's ETL Process

#### 4.2.1.1 DimEmployee

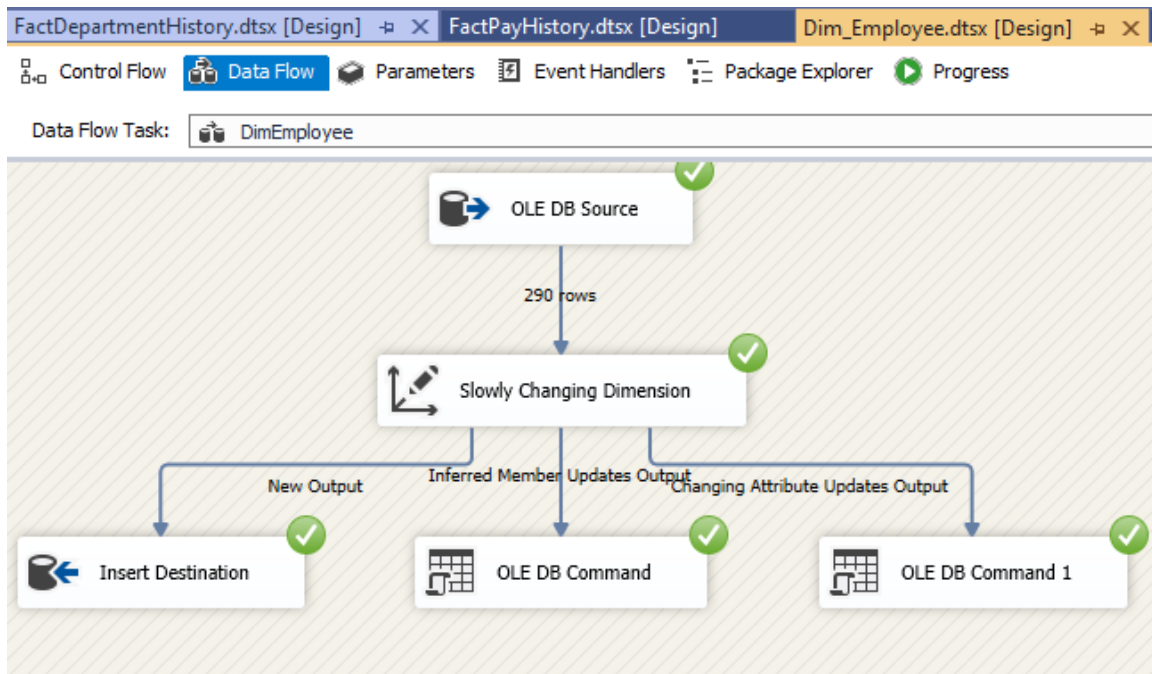


Figure 4.2.1.1: SSIS for DimEmployee

SQLQuery4.sql - D:\644CKH\Home (54) X

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP 1000 [BusinessEntityKey]
, [BusinessEntityID]
, [FirstName]
, [MiddleName]
, [LastName]
, [PhoneNumber]
, [BirthDate]
, [Gender]
, [MaritalStatus]
, [EmailAddress]
, [AddressLine1]
, [City]
, [StateProvince]
, [CountryRegion]
, [OrganizationLevel]
, [JobTitle]

```

100 %

Results Messages

	BusinessEntityKey	BusinessEntityID	FirstName	MiddleName	LastName	PhoneNumber	BirthDate	Gender	MaritalStatus	EmailAddress
1	1	1	Ken	J	Sánchez	697-555-0142	1969-01-29	M	S	ken0@adventure-works.com
2	2	2	Terri	Lee	Duffy	819-555-0175	1971-08-01	F	S	terri0@adventure-works.com
3	3	3	Roberto	NULL	Tamburello	212-555-0187	1974-11-12	M	M	roberto0@adventure-works.com
4	4	4	Rob	NULL	Walters	612-555-0100	1974-12-23	M	S	rob0@adventure-works.com
5	5	5	Gail	A	Erickson	849-555-0139	1952-09-27	F	M	gail0@adventure-works.com
6	6	6	Josief	H	Goldberg	122-555-0189	1959-03-11	M	M	josief0@adventure-works.com
7	7	7	Dylan	A	Miller	181-555-0156	1987-02-24	M	M	dylan0@adventure-works.com
8	8	8	Diane	L	Margheim	815-555-0138	1986-06-05	F	S	diane1@adventure-works.com
9	9	9	Gigi	N	Matthew	185-555-0186	1979-01-21	F	M	gigi0@adventure-works.com
10	10	10	Michael	NULL	Raheem	330-555-2568	1984-11-30	M	M	michael6@adventure-works.com

Query executed successfully. DESKTOP-B644CKH (12.0 SP1) DESKTOP-B644CKH\Home (54) | Nhom5\_HR\_DW | 00:00:00 | 290 rows

#### 4.2.1.2 DimDepartment

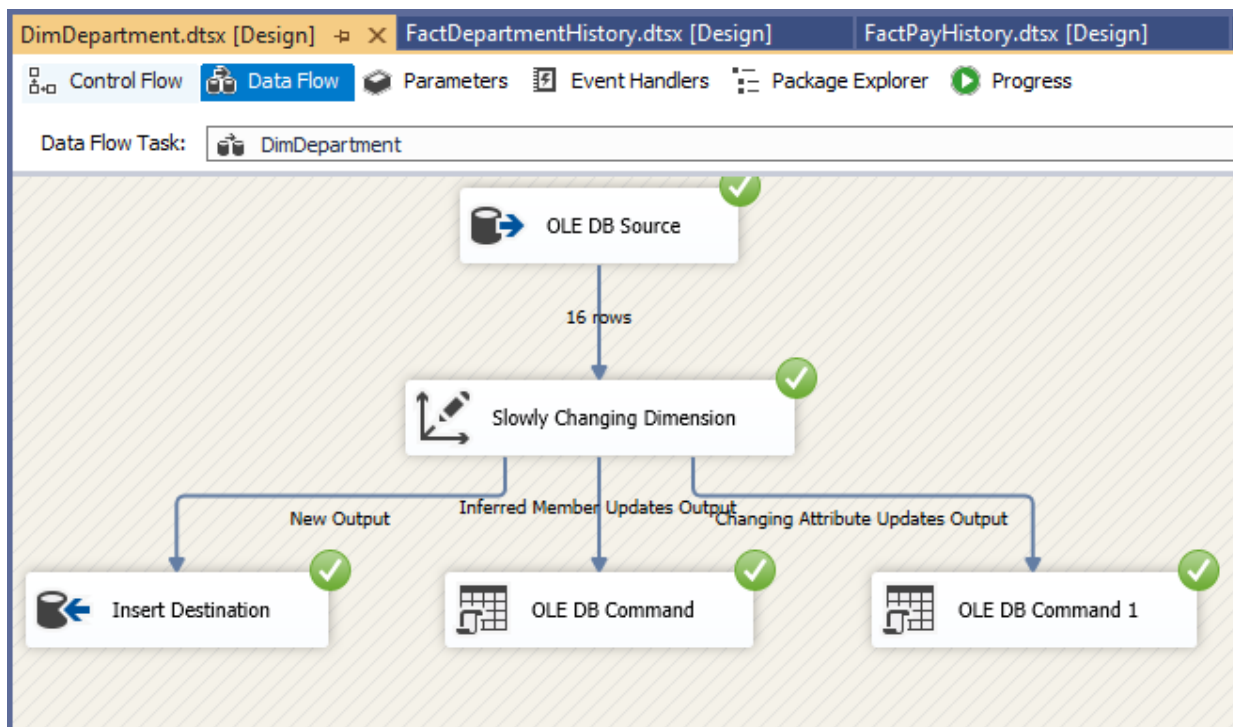


Figure4.2.1.2: SSIS for DimDepartment



SQLQuery3.sql - D:\644CKH\Home (52) X DESKTOP-B644CKH.....FactDemographic SQLQuery1.sql - D:\644CKH\Home (55))\*

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP 1000 [DepartmentKey]
, [DepartmentID]
, [Name]
, [GroupName]
, [ModifiedDate]
FROM [Nhom5_HR_DW].[dbo].[DimDepartment]

```

100 %

Results Messages

	DepartmentKey	DepartmentID	Name	GroupName	ModifiedDate
1	1	1	Engineering	Research and Development	2008-04-30 00:00:00.000
2	2	2	Tool Design	Research and Development	2008-04-30 00:00:00.000
3	3	3	Sales	Sales and Marketing	2008-04-30 00:00:00.000
4	4	4	Marketing	Sales and Marketing	2008-04-30 00:00:00.000
5	5	5	Purchasing	Inventory Management	2008-04-30 00:00:00.000
6	6	6	Research and Development	Research and Development	2008-04-30 00:00:00.000
7	7	7	Production	Manufacturing	2008-04-30 00:00:00.000
8	8	8	Production Control	Manufacturing	2008-04-30 00:00:00.000
9	9	9	Human Resources	Executive General and Administration	2008-04-30 00:00:00.000
10	10	10	Finance	Executive General and Administration	2008-04-30 00:00:00.000
11	11	11	Information Services	Executive General and Administration	2008-04-30 00:00:00.000

Query executed successfully. DESKTOP-B644CKH (12.0 SP1) DESKTOP-B644CKH\Home (52) Nhom5\_HR\_DW 00:00:00 16 rows

#### 4.2.1.3 DimShift

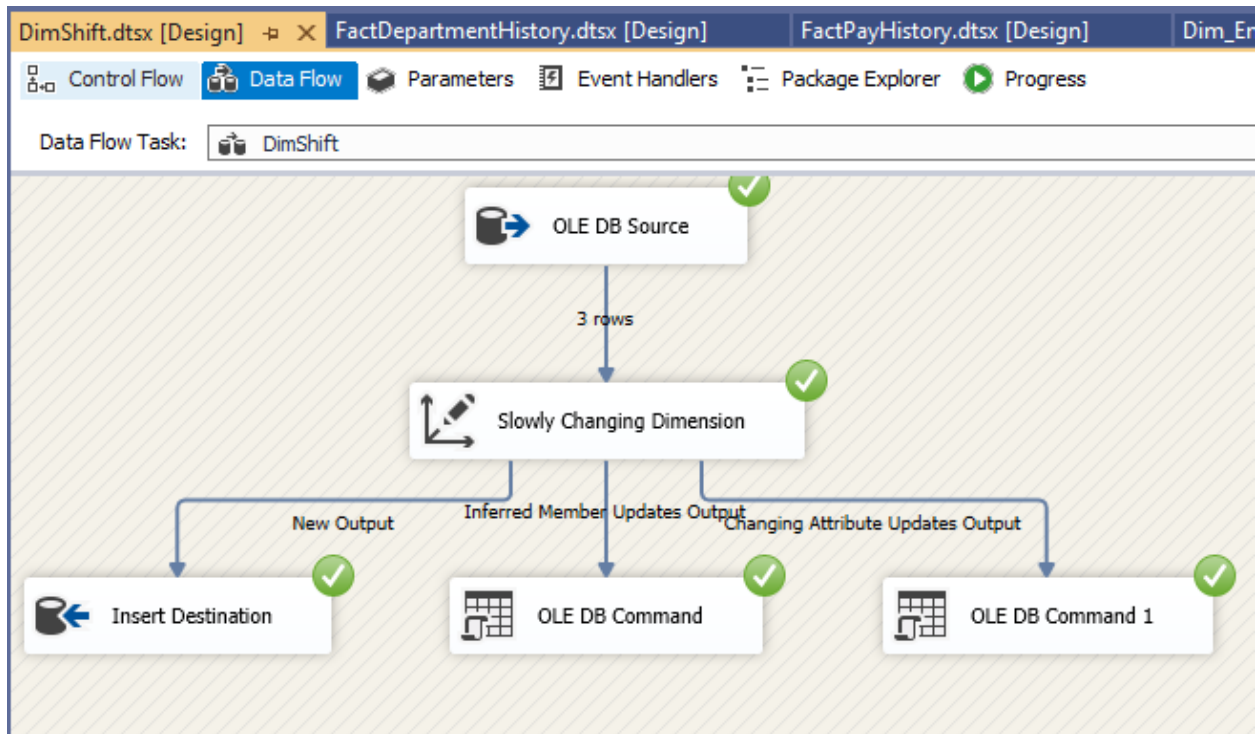


Figure 4.2.1.3: SSIS for DimShift

SQLQuery8.sql - D...644CKH\Home (59) X SQLQuery7.sql - D...644CKH\Home (58) SQLQuery6.sql - D...644CKH\Home (56)

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP 1000 [ShiftKey]
, [ShiftID]
, [Name]
, [StartTime]
, [EndTime]
, [ModifiedDate]
FROM [Nhom5_HR_DW].[dbo].[DimShift]

```

100 %

Results Messages

	ShiftKey	ShiftID	Name	StartTime	EndTime	ModifiedDate
1	1	1	Day	07:00:00.0000000	15:00:00.0000000	2008-04-30 00:00:00.000
2	2	2	Evening	15:00:00.0000000	23:00:00.0000000	2008-04-30 00:00:00.000
3	3	3	Night	23:00:00.0000000	07:00:00.0000000	2008-04-30 00:00:00.000

Query executed successfully. DESKTOP-B644CKH (12.0 SP1) DESKTOP-B644CKH\Home (59) Nhom5\_HR\_DW 00:00:00 3 rows

#### 4.2.1.4 DimTime

SQLQuery9.sql - D...644CKH\Home (60) X SQLQuery8.sql - D...644CKH\Home (59) SQLQuery7.sql - D...644CKH\Home (58)

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP 1000 [Timekey]
, [Year]
, [Month]
, [Day]
, [DateName]
, [DayNumberOfWeek]
, [DayNumberOfMonth]
, [DayNumberOfYear]
, [WeekNumberOfYear]
, [Quarter]
FROM [Nhom5_HR_DW].[dbo].[DimTime]

```

100 %

Results Messages

	Timekey	Year	Month	Day	DateName	DayNumberOfWeek	DayNumberOfMonth	DayNumberOfYear	WeekNumberOfYear	Quarter
1	20060101	2006	01	01	Sunday	1	1	1	1	1
2	20060102	2006	01	02	Monday	2	2	2	1	1
3	20060103	2006	01	03	Tuesday	3	3	3	1	1
4	20060104	2006	01	04	Wednesday	4	4	4	1	1
5	20060105	2006	01	05	Thursday	5	5	5	1	1
6	20060106	2006	01	06	Friday	6	6	6	1	1
7	20060107	2006	01	07	Saturday	7	7	7	1	1
8	20060108	2006	01	08	Sunday	1	8	8	2	1
9	20060109	2006	01	09	Monday	2	9	9	2	1
10	20060110	2006	01	10	Tuesday	3	10	10	2	1
11	20060111	2006	01	11	Wednesday	4	11	11	2	1

Query executed successfully. DESKTOP-B644CKH (12.0 SP1) DESKTOP-B644CKH\Home (60) Nhom5\_HR\_DW 00:00:01 1000 rows

Figure 4.2.1.4: SSIS for DimTime

## 4.2.2. Fact Table's ETL Process

### 4.2.2.1 FactPayHistory

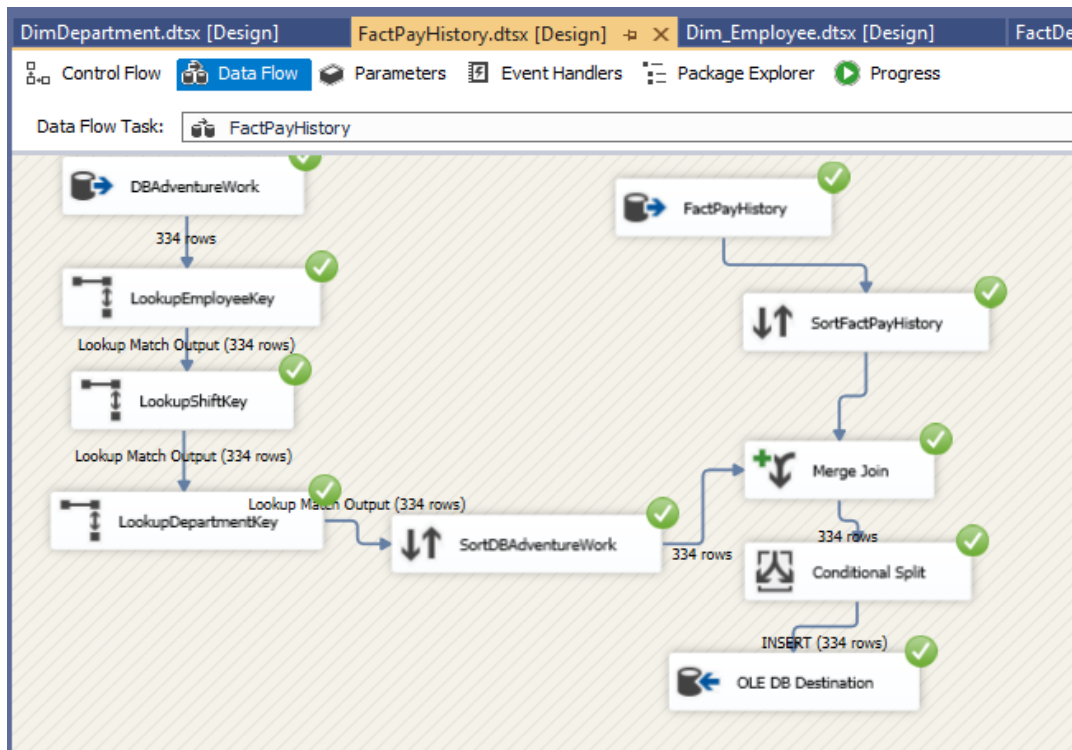


Figure 4.2.2.1: SSIS for FactPayHistory

SQLQuery1.sql - D:\644CKH\Home (58)

```
/****** Script for SelectTopNRows command from SSMS *****/  
SELECT TOP 1000 [FactPayHistoryKey]  
    , [BusinessEntityKey]  
    , [DepartmentKey]  
    , [ShiftKey]  
    , [RateChangeDate]  
    , [Rate]  
    , [PayFrequency]  
FROM [Nhom5_HR_DW].[dbo].[FactPayHistory]
```

100 %

Results Messages

	FactPayHistoryKey	BusinessEntityKey	DepartmentKey	ShiftKey	RateChangeDate	Rate	PayFrequency
1	1	1	16	1	20090114	125.50	2
2	2	2	1	1	20080131	63.4615	2
3	3	3	1	1	20071111	43.2692	2
4	4	4	1	1	20071205	8.62	2
5	5	4	1	1	20100531	23.72	2
6	6	4	1	1	20111215	29.8462	2
7	7	4	2	1	20071205	8.62	2
8	8	4	2	1	20100531	23.72	2
9	9	4	2	1	20111215	29.8462	2
10	10	5	1	1	20080106	32.6923	2
11	11	6	1	1	20080124	32.6923	2

Query executed successfully. DESKTOP-B644CKH (12.0 SP1) DESKTOP-B644CKH\Home (58) | Nhom5\_HR\_DW | 00:00:01 | 334 rows

#### 4.2.2.2 FactDepartmentHistory

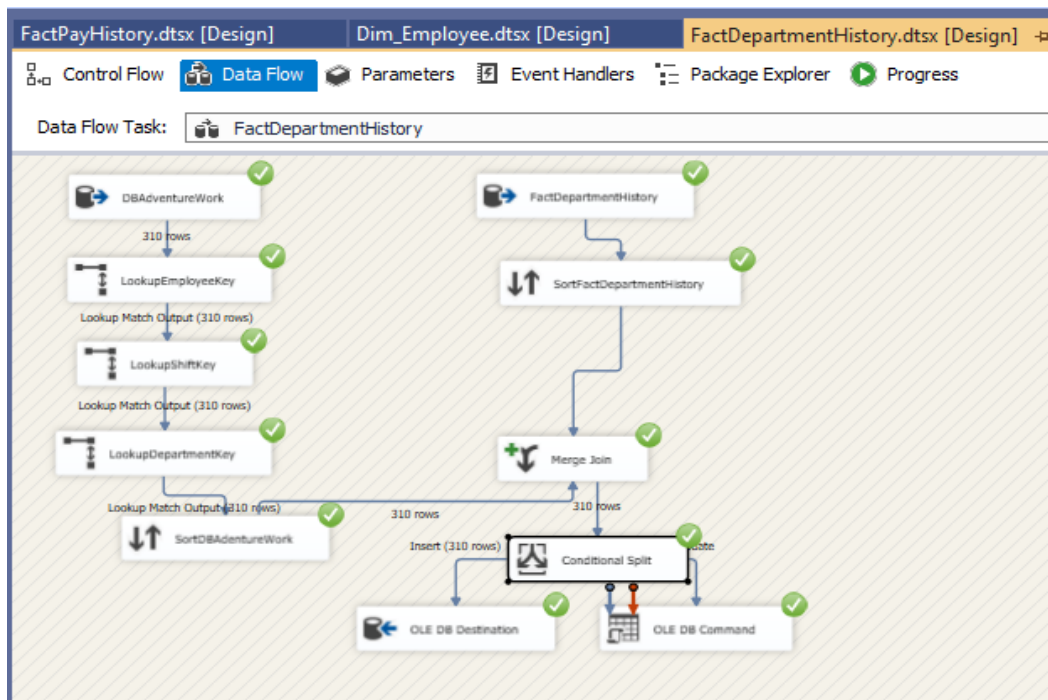


Figure 4.2.2.2: SSIS for FactDepartmentHistory

SQLQuery4.sql - D:\644CKH\Home (55) X SQLQuery3.sql - D:\644CKH\Home (60)\* SQLQuery2.sql - D:\644CKH\Home (52)

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP 1000 [FactDepartmentHistoryKey]
,[BusinessEntityKey]
,[DepartmentKey]
,[ShiftKey]
,[TimeOffHours]
,[StartDate]
,[EndDate]
FROM [Nhom5_HR_DW].[dbo].[FactDepartmentHistory]

```

100 %

Results Messages

	FactDepartmentHistoryKey	BusinessEntityKey	DepartmentKey	ShiftKey	TimeOffHours	StartDate	EndDate
1	1	1	16	1	168	20090114	NULL
2	2	2	1	1	21	20080131	NULL
3	3	3	1	1	23	20071111	NULL
4	4	4	1	1	128	20071205	20100530
5	5	4	1	1	128	20100531	NULL
6	6	4	2	1	128	20071205	20100530
7	7	4	2	1	128	20100531	NULL
8	8	5	1	1	27	20080106	NULL
9	9	6	1	1	29	20080124	NULL
10	10	7	6	1	111	20090208	NULL
11	11	8	6	1	113	20081229	NULL

Query executed successfully. DESKTOP-B644CKH (12.0 SP1) DESKTOP-B644CKH\Home (55) Nhom5\_HR\_DW 00:00:00 310 rows

## CHAPTER 5: DATA ANALYTICS

In this chapter, the topic will present the techniques and methods of data analysis, the method of setting KPIs on the data analysis system to support business performance evaluation and finally construction.

### 5.1. Data analytics with SSAS technology

#### 5.1.1. Building the cube

To analyze data from DW data warehouse and build intelligent reports to support decision making, the thesis uses SSAS tool. SSAS is integrated in the BIDS (Business Intelligence Development Studio) tool that is included with SQL Server.

- Step 1: Create Project in SSAS
- Step 2: Create Data sources
- Step 3: Create Data source Views combined with creating Measures
- Step 4: Create Cube combined with creating Dimension

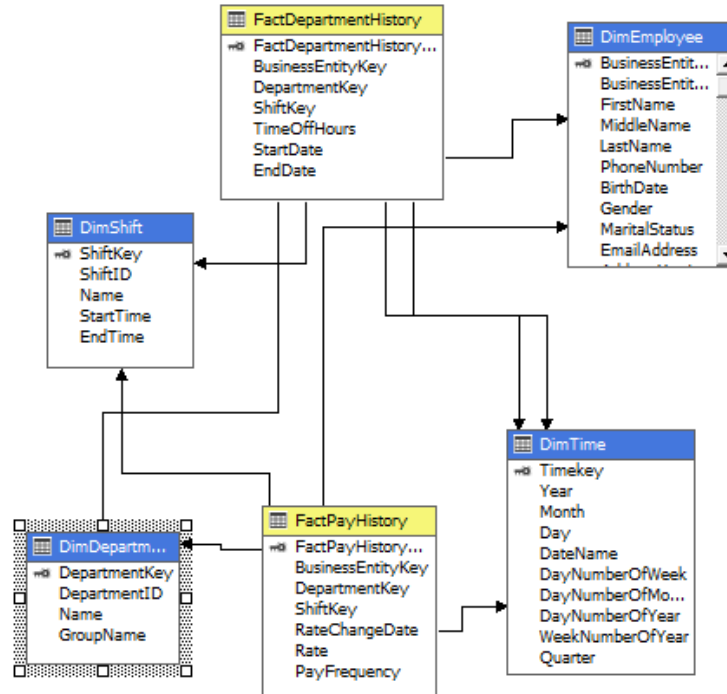


Figure 5.1.1: Cube Model

## 5.1.2. Analysis with SSAS

### 5.1.2.1. Report on the situation of personnel in the department in 2009

Dimension	Hierarchy	Operator	Filter Expression	Param...
Start Date	Start Date.Year	Equal	{ 2009 }	<input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input type="checkbox"/>

Name	Group Name	Fact Department Hist...	AvgTimeOffHour
Document Control	Quality Assurance	4	135.5
Executive	Executive General and Administration	1	168
Facilities and Maintenance	Executive General and Administration	2	149.5
Finance	Executive General and Administration	9	99.777777777...
Human Resources	Executive General and Administration	3	97.333333333...
Information Services	Executive General and Administration	7	124.571428571...
Marketing	Sales and Marketing	4	86.25
Production	Manufacturing	104	89
Production Control	Manufacturing	3	90.333333333...
Purchasing	Inventory Management	3	107.333333333...
Quality Assurance	Quality Assurance	3	143.333333333...
Research and Development	Research and Development	3	101.666666666...
Shipping and Receiving	Inventory Management	3	161.333333333...

*Figure 5.1.2.1: Situation of personnel in the department in 2009*

From the historical data, it is found that the personnel change is much larger than the previous year, so we will take a close look at the personnel change in the departments in 2009. The number of new employees in the production department. This year's production is 104, showing that in 2009 the company expanded production on a large scale, the remaining departments also recruited more employees, especially the number of highly qualified employees to serve the expansion. tissue during the year.

### 5.1.2.2. List of positions that move departments

Dimension	Hierarchy	Operator	Filter Expression	Param...
End Date	End Date.Year	Equal	{ 2006, 2007, 2008, 2009, 2010, 2011, 2...	<input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input type="checkbox"/>

Name	Year	Job Title	Time Off Hours	Fact Department Histo...
Engineering	2010	Senior Tool Designer	128	1
Finance	2013	Chief Financial Officer	20	1
Marketing	2011	Purchasing Manager	93	1
Production	2011	Scheduling Assistant	87	1
Purchasing	2009	Marketing Manager	80	1
Quality Assurance	2012	Purchasing Manager	93	1

Figure 5.1.2.2: List of positions

Human resource managers need to understand this list, find out the reasons for the positive or negative change in departments to have a reasonable adjustment policy. At Adventure Work Cycle, the number of employees moving between departments is quite small and does not change suddenly at any one time, so perhaps this is just a natural department rotation.

### 5.1.3. Building KPIs system

Dimension	Hierarchy	Operator	Filter Expression
<Select dimension>			

Display Structure	Value	Goal	Status	Trend	Weight
KPINewEmployee	296		😊	➡	
KPISalary	36021.51	50000	🚦	😞	

Figure 5.1.3.1: Display KPI of group

### Status light:

- Yellow color shows that the actual value has not reached the target but is still at acceptable level receive.
- Green: meet or exceed the target.
- Red: not acceptable, the actual result is too low compared to the target

### Trends:

- Smileys/ Arrow pointing up: uptrend from last year
- Sad side/ Down arrow: downtrend
- Normal face/ Horizontal arrow: equal to or approximately the previous year

#### 5.1.3.1. KPI Number of new hires each year by each department

##### KPI Value

The screenshot displays a KPI configuration window. It includes a 'Name:' field with the text 'KPINewEmployee'. Below it is an 'Associated measure group:' dropdown menu currently set to '<All>'. A section titled 'Value Expression' is expanded, showing the expression '[Measures].[Fact Department History Count]' in a text area. At the bottom of this section, a status bar indicates 'No issues found' with a green checkmark icon, and on the right, it shows 'Ln: 1 Ch: 5 SPC CRLF'.

*Figure 5.1.3.1: KPI value of new hires each year by each department*



### KPI Goal

Since this is a manufacturing company, it is certain that the number of employees in the production department will account for the majority, and the number of employees in the remaining departments will also vary: some are new hires, others is departmental rotation.

Goal Expression


```
case when [Dim Department].[Name] is [Dim Department].[Name].&[Document
Control] then 3
when [Dim Department].[Name] is [Dim Department].[Name].&[Engineering] then 2
when [Dim Department].[Name] is [Dim Department].[Name].&[Executive] then 2
when [Dim Department].[Name] is [Dim Department].[Name].&[Facilities and
Maintenance] then 3
when [Dim Department].[Name] is [Dim Department].[Name].&[Finance] then 4
when [Dim Department].[Name] is [Dim Department].[Name].&[Human Resources] then
3
when [Dim Department].[Name] is [Dim Department].[Name].&[Information Services]
then 5
when [Dim Department].[Name] is [Dim Department].[Name].&[Marketing] then 3
when [Dim Department].[Name] is [Dim Department].[Name].&[Production] then 30
when [Dim Department].[Name] is [Dim Department].[Name].&[Production Control]
then 3
✓ No issues found Ln: 5 Ch: 25 SPC CRLF
```

Figure 5.1.3.1: Goal of new hires each year by each department

### KPI Status

It shows the level of achievement or failure in the company's recruitment activities.

Status

Status indicator:  Faces

Status expression:

```
case when
KPIVALUE("KPINewEmployee" )/KPIGOAL( "KPINewEmployee" )>0.9 then 1
when KPIVALUE("KPINewEmployee" )/KPIGOAL( "KPINewEmployee" )<=0.9 and
KPIVALUE("KPINewEmployee" )/KPIGOAL( "KPINewEmployee" )>0.7 then 0
else -1
✓ No issues found Ln: 6 Ch: 4 SPC CRLF
```

Figure 5.1.3.1: Status of new hires each year by each department

### KPI Trend

It shows how the number of new employees this year compared to last year.

▲ Trend i

Trend indicator: ↑ Status arrow

Trend expression:

```
case when Isempy
([Start Date].[Year].currentmember.lead(-1)) then 0
when ([Measures].[Fact Department History Count],[Start Date].
[Year].currentmember)>([Measures].[Fact Department History Count],[Start Date].
[Year].currentmember.lead(-1)) then 1
end
```

✓ No issues found Ln: 6 Ch: 4 SPC CRLF

*Figure 5.1.3.1: Trend of new hires each year by each department*

### Result

	A	B	C	D	E
1	Row Labels	KPINewEmployee	KPIGoal	KPIStatus	KPITrend
2	Document Control				
3	2006		3	◆	→
4	2007		3	◆	→
5	2008	1	3	◆	↑
6	2009	4	3	●	↑
7	2010		3	◆	↓
8	2011		3	◆	→
9	2012		3	◆	→
10	2013		3	◆	→
11	Engineering				
12	2006		2	◆	→
13	2007	2	2	●	↑
14	2008	3	2	●	↑
15	2009		2	◆	↓
16	2010	1	2	◆	↑
17	2011	1	2	◆	→
18	2012		2	◆	↓
19	2013		2	◆	→
20	Executive				
21	2006		2	◆	→
22	2007		2	◆	→
23	2008		2	◆	→

*Figure 5.1.3.1: Result of new hires each year by each department*

Managers can apply this type of KPI in evaluating the effectiveness of recruiting personnel in the enterprise. At the same time, it is combined with employee performance evaluation to come up with an effective recruitment plan for the following years. This KPI can also

be used in many different aspects such as job title, gender, .... to discover more insights to support recruitment planning

#### 5.1.3.2. KPI Annual departmental average salary

##### *Calculate*

The average hourly wage

The screenshot shows a KPI configuration window. At the top, there is a 'Name:' label followed by a text box containing 'AvgRateSalary'. Below this, there are two expandable sections. The first section, 'Parent Properties', contains a 'Parent hierarchy:' dropdown menu set to 'Measures' and a 'Parent member:' text box. To the right of these fields is a 'Change' button. The second section, 'Expression', contains a text box with the formula '[Measures].[Rate]/[Measures].[Fact Pay History Count]'. Below the formula text box, there is a status bar that reads 'No issues found' with a green checkmark icon, and on the right, it shows 'Ln: 1 Ch: 54 SPC CRLF'.

##### *KPI Value*

Assume that employees of Adventure Work Cycle work under the following regime: each working day has an 8-hour shift from Monday to Friday. Therefore, the number of working hours in a year is calculated by the following formula:

$$\text{Work Time} = (365 - 2*4*12)*8 \text{ (hours)}$$

⌄ KPI

Name:

KPISalary

Associated measure group:

<All> ▼

⌄ Value Expression

[Measures].[AvgRateSalary]\*(365-2\*4\*12)\*8

✓ No issues found Ln: 1 Ch: 42 SPC CRLF

*Figure 5.1.3.2: Value of Annual departmental average salary*

## KPI Goal

\$50,000 represents the average annual income of people operating in the same industry so that managers can easily compare and evaluate.



Figure 5.1.3.2: Goal of Annual departmental average salary

## KPI Status



Figure 5.1.3.2: Status of Annual departmental average salary

## KPI Trend

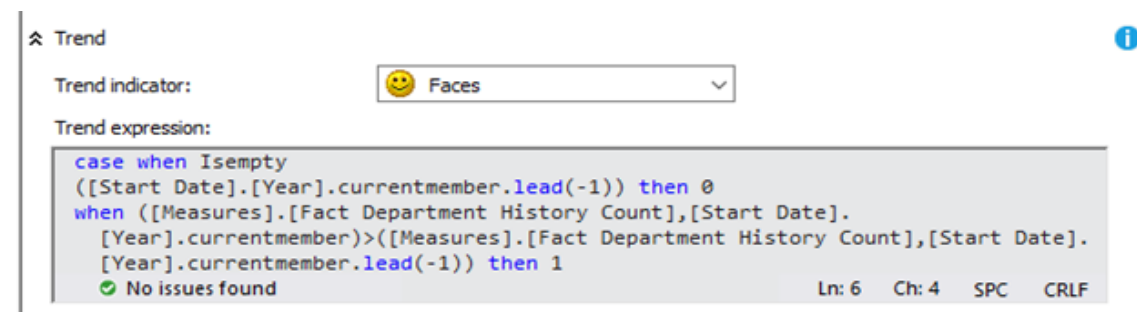


Figure 5.1.3.2: Trend of Annual departmental average salary

## Result


















	A	B	C	D	E
1	Row Labels	KPISalary	KPISalary Goal	KPISalary Status	Time Off Hours
2	Document Control	30963.96592	50000		676
3	Engineering	76698.90937	50000		278
4	Executive	270076	50000		188
5	Facilities and Maintenance	28044.0032	50000		1073
6	Finance	54467.29607	50000		1110
7	Human Resources	38789.40547	50000		582
8	Information Services	73509.37176	50000		1240
9	Marketing	37859.06	50000		865
10	Production	29113.22385	50000		16803
11	Production Control	40198.24813	50000		528
12	Purchasing	38363.4888	50000		1298
13	Quality Assurance	37097.70631	50000		954
14	Research and Development	93984.5112	50000		418
15	Sales	64499.62444	50000		1104
16	Shipping and Receiving	23396.1136	50000		978
17	Tool Design	56990.29696	50000		223
18	Grand Total	37697.37534	50000		28318

Figure 5.1.3.2: Result of Annual departmental average salary

The three groups of departments: Executive General and Administration, Research and Development and Sales and Marketing all achieve salaries greater than \$50,000 per year.

This KPI supports tracking and comparing the average salary between departments and groups of departments, assessing this change over the years. From there, there are appropriate departmental support policies to achieve balance in the organization. For example, it is possible to change the number of days off allowed to satisfy the employee satisfaction level.

## 5.2. Data analytics with MDX

### 5.2.1. List time-off hours for each department

```
SELECT [Measures].[Time Off Hours] on 0,[Dim Department].[Name].members on 1
```

```
FROM [Nhom5 HR DW]
```

## RESULT

Messages	Results
	Time Off Hours
All	28318
Document Control	676
Engineering	278
Executive	188
Facilities and Maintenance	1073
Finance	1110
Human Resources	582
Information Services	1240
Marketing	865
Production	16803
Production Control	528
Purchasing	1298
Quality Assurance	954
Research and Development	418
Sales	1104
Shipping and Receiving	978
Tool Design	223

*Figure 6.5.1: Result of List time-off hours for each department*

### 5.2.2. Difference between the time-off hours of each department

```
WITH MEMBER CHENHLECHGIONGHI AS  
  
([Measures].[Time Off Hours],[Dim Department].[Name])  
  
-([Measures].[Time Off Hours],[Dim Department].[Name].LEAD(-1))  
  
SELECT CHENHLECHGIONGHI ON 0,[Dim Department].[Name] .MEMBERS ON 1  
  
FROM [Nhom5 HR DW]
```

## Result

Messages	Results
	CHENHLECHGIONGHI
All	28318
Document Control	676
Engineering	-398
Executive	-90
Facilities and Maintenance	885
Finance	37
Human Resources	-528
Information Services	658
Marketing	-375
Production	15938
Production Control	-16275
Purchasing	770
Quality Assurance	-344
Research and Development	-536
Sales	686
Shipping and Receiving	-126
Tool Design	-755

Figure 6.5.2: Result of Difference between the time-off hours of each department

### 5.2.3. Top 5 departments with the most time-off hours

```
SELECT TOPCOUNT([Dim Department].[Name].children, 5,[Measures].[Time Off Hours]) ON 1,
[Measures].[Time Off Hours] ON 0

FROM [Nhom5 HR DW]
```

### RESULT

Messages	Results
	Time Off Hours
Production	16803
Purchasing	1298
Information Services	1240
Finance	1110
Sales	1104

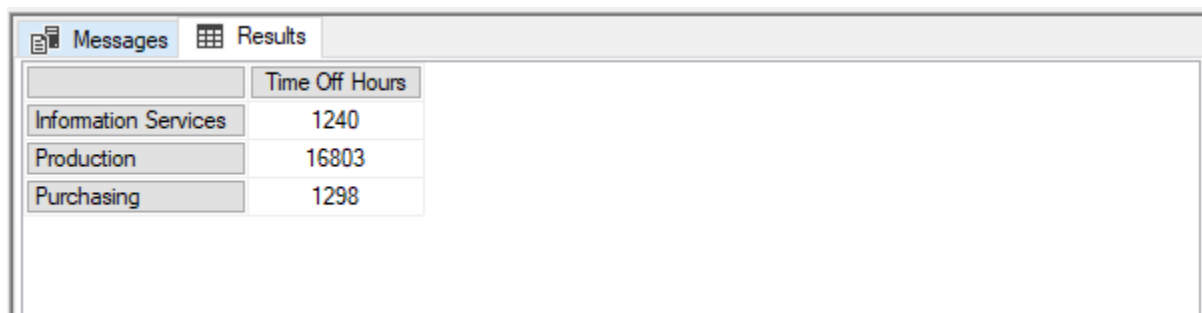
Figure 6.5.3: Result of Top 5 departments with the most time-off hours



#### 5.2.4. The department has more than 1200 hours of time-off hours

```
WITH  
  
SET [DEPARTMENT - HighTimeOff] AS FILTER({[Dim Department].[Name].children},[Measures].[Time  
Off Hours]>1200)  
  
SELECT [Measures].[Time Off Hours] ON 0,  
  
[DEPARTMENT - HighTimeOff] ON 1  
  
FROM [Nhom5 HR DW]
```

#### RESULT



	Time Off Hours
Information Services	1240
Production	16803
Purchasing	1298

*Figure 6.5.4: Result of Top 5 departments with the most time-off hours*

#### 5.2.5 Rate of employee

```
select [Measures].[Rate] on 0,  
  
([Dim Employee].[First Name].children,[Dim Employee].[Last Name].children) on 1  
  
from [Nhom5 HR DW]
```

#### RESULT

Messages		Results
		Rate
A. Scott	Wright	23.5577
Alan	Brewer	16
Alejandro	McGuel	15
Alex	Nayberg	10
Alice	Ciccu	11
Amy	Alberts	48.101
Andreas	Berglund	10.5769
Andrew	Cencini	10
Andrew	Hill	25
Andy	Ruth	9.5
Angela	Barbariol	11
Anibal	Sousa	14
Annette	Hill	12.75
Annik	Stahl	12.45
Arvind	Rao	18.2692
Ashvini	Shama	32.4519

Figure 6.5.5: Result of Rate of employee

### 5.2.6 Average Rate of department

```
with member LuongTBPhongBan as avg([Measures].[Rate])

select LuongTBPhongBan on 0, [Dim Department].[Group Name].members on 1

from [Nhom5 HR DW]
```

### RESULT

Messages		Results
		LuongTBPhongBan
All		5535.4882
Executive General and Administration		944.8673
Inventory Management		279.1536
Manufacturing		2777.1815
Quality Assurance		227.0907
Research and Development		556.5901
Sales and Marketing		750.605

Figure 6.5.6: Result of Average of department

### 5.2.7 Top 5 employees with the highest wages at Sales and Marketing

```
SELECT TOPCOUNT([Dim Employee].[First Name].children, 5, [Measures].[Rate])ON 1,  
[Measures].[Rate] ON 0  
FROM [Nhom5 HR DW]  
WHERE [Dim Department].[Group Name].&[Sales and Marketing]
```

#### RESULT



The screenshot shows a software interface with two tabs: 'Messages' and 'Results'. The 'Results' tab is active and displays a table with two columns: 'First Name' and 'Rate'. The table contains five rows of data, representing the top 5 employees by wage in the Sales and Marketing department.

First Name	Rate
David	89.3269
Brian	72.1154
Amy	48.101
Stephen	48.101
Syed	48.101

*Figure 6.5.7: Result of Top 5 employees with the highest wages at Sales and Marketing*

### 5.2.8 Wages of divisions

```
WITH MEMBER ChenhLechLuongCacPhongBan AS  
([Measures].[Rate],[Dim Department].[Department ID])  
-([Measures].[Rate],[Dim Department].[Department ID].LEAD(-1))  
SELECT ChenhLechLuongCacPhongBan ON 0, [Dim Department].[Department ID].children ON 1  
FROM [Nhom5 HR DW]
```

#### RESULT

Messages		Results
		ChenhLechLuongCacPhongBan
1		249.4853
10		28.9256
11		63.1754
12		-269.644
13		83.2061
14		-63.9272
15		-25.9904
16		60.2692
2		6.912399999999999
3		407.0826
4		-328.385
5		2.812800000000001
6		-39.2304
7		2490.4122
8		-2553.0277
9		-3.928

Figure 6.5.8: Result of Top 5 employees with the highest wages at Sales and Marketing

### 5.2.9 Medium performance wages by sex

```
With member AvgSalary as avg([Measures].[Rate])
select [Dim Time].[Year].children on 1,
(AvgSalary,[Dim Employee].[Gender].children) on 0
from [Nhom5 HR DW]
```

RESULT

Messages Results		
	AvgSalary	AvgSalary
	F	M
2006	(null)	12.45
2007	25	114.3507
2008	477.1989	821.1545
2009	674.5004	1874.2681
2010	141.3192	456.3968
2011	193.566	374.1971
2012	101.1779	99.6538
2013	46.1538	124.101
2014	(null)	(null)

*Figure 6.5.9: Result of Medium performance wages by sex*

#### 5.2.10 Average break time in each department

##### MDX CODE

```
with member TongLuong as avg([Measures].[Time Off Hours])

select [Measures].[Time Off Hours] on 0, ([Dim Department].[Group Name].CHILDREN,[Start
Date].[Quarter].CHILDREN) on 1

from [Nhom5 HR DW]
```

##### RESULT

Messages Results		
		Time Off Hours
Executive General and Administration	1	2846
Executive General and Administration	2	(null)
Executive General and Administration	3	(null)
Executive General and Administration	4	1347
Executive General and Administration	Unknown	(null)
Inventory Management	1	1314
Inventory Management	2	(null)
Inventory Management	3	93
Inventory Management	4	869
Inventory Management	Unknown	(null)
Manufacturing	1	11923
Manufacturing	2	51
Manufacturing	3	87
Manufacturing	4	5270
Manufacturing	Unknown	(null)
Quality Assurance	1	1119
Quality Assurance	2	(null)

*Figure 6.5.10: Result of Average break time in each department*

## CHAPTER 6: VISUALIZATION AND FORECASTING

In this chapter, visualize information and insights that are analyzed on smart reports into graphical forms such as graphs, charts, or use different methods and tools to visualize and illustrate data is the best. Thereby helping managers have a multi-dimensional perspective and make more effective decisions.

### 6.1. Report and dashboard systems

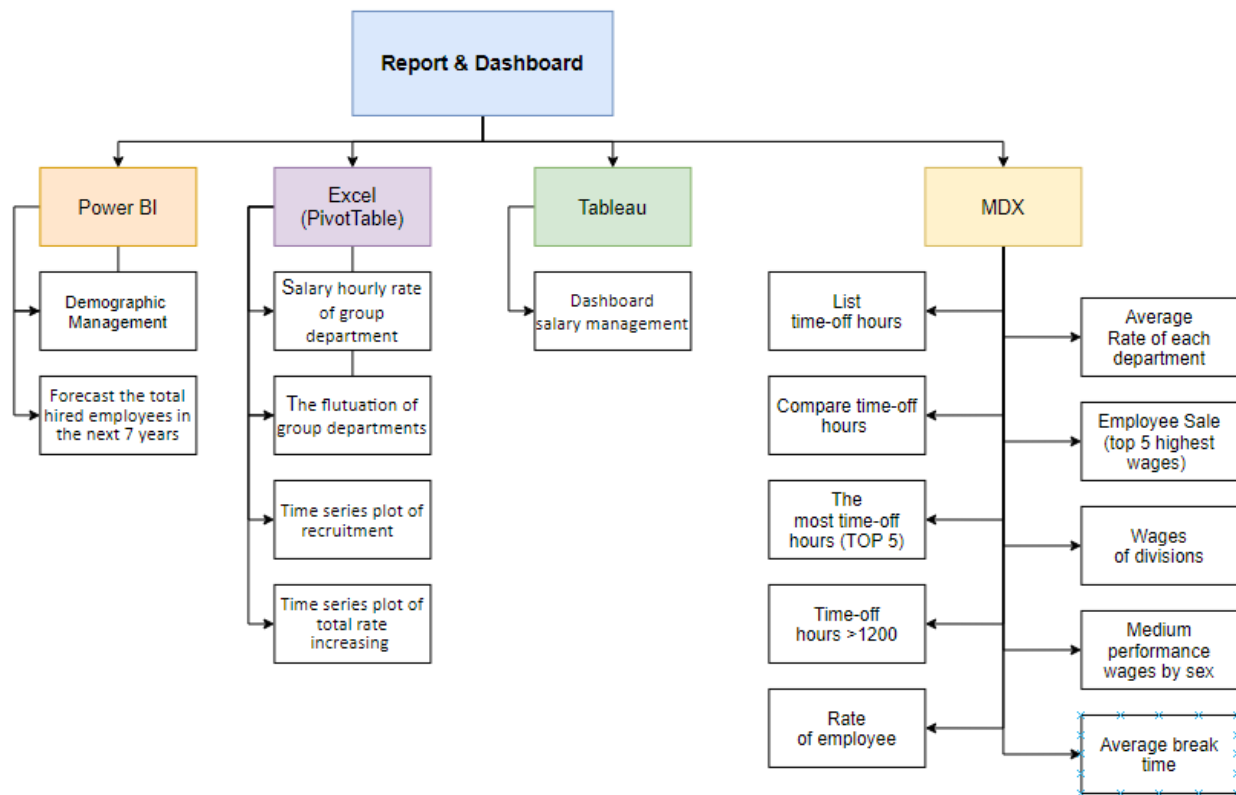


Figure 6.1: Structure of Report & Dashboard

## 6.2. Data analysis with Power BI



Figure 6.2: Dashboard demographic management in Power BI tool

In this dashboard, the manager can use the Department slicer to choose the department they want or all departments in the company. Then, other charts will change according to the department they selected before.

- The donut chart visualizes the percentage between female and male in the area the manager wants to see.
- The pie chart shows the percentage of marital status in the company. There are 2 marital status: Married and Single.

However, for example the manager can also see the percentage of male who are single by clicking the percentage of male in donut chart or single in pie chart to see the highlight.

- The map indicates the distribution of our employees by Country and City.
- Three numbers in the middle of the dashboard are the average rate of employees, the total time-off hours and the total number of employees in our company or each department, respectively.



- The above clustered bar chart presents the Top 5 Job Titles in our company with the highest rate.
- The below clustered bar chart represents the Job Title in every department followed by their rate, which ranks in the suitable order to help the manager easily understand.

For example, the dashboard above is the filter of all information relating to the Quality Assurance department. The manager can see the Average Rate of this department is 16.228\$; 1816 total time-off hours and total 12 employees in the department from the beginning to the present. Moreover, they also can see all positions in the department through the clustered bar chart in the corner and its rates.

Besides that, when the manager clicks one of the job titles in the clustered bar chart in the corner, all information will change. So they can see the number of employees who are in that position, where they live in the map on the left, their total time-off hours and their rate as well.

### **6.3. Data analysis with the Pivot Table tool in Microsoft Excel**

The below dashboard shows the salary hourly rate of the group department. In Adventure Work, the departments are divided into 6 groups: Executive General and Administration, Quality Assurance, Manufacturing, Inventory Management, Sales and Marketing, and Research and Development.

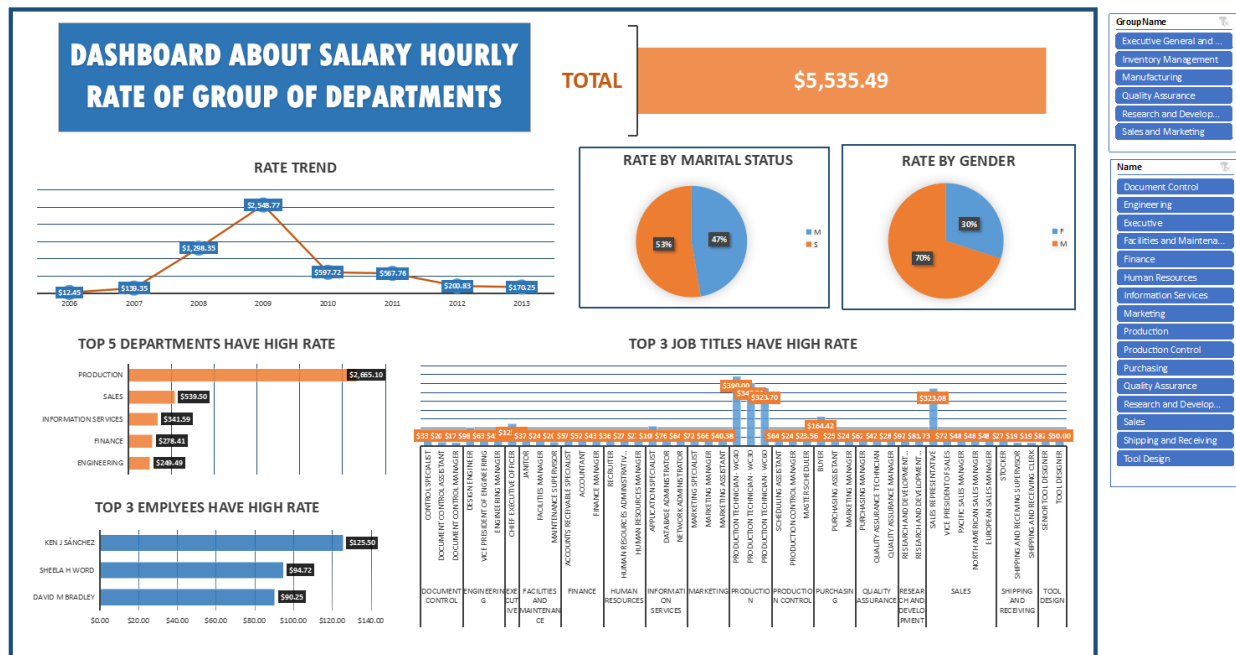
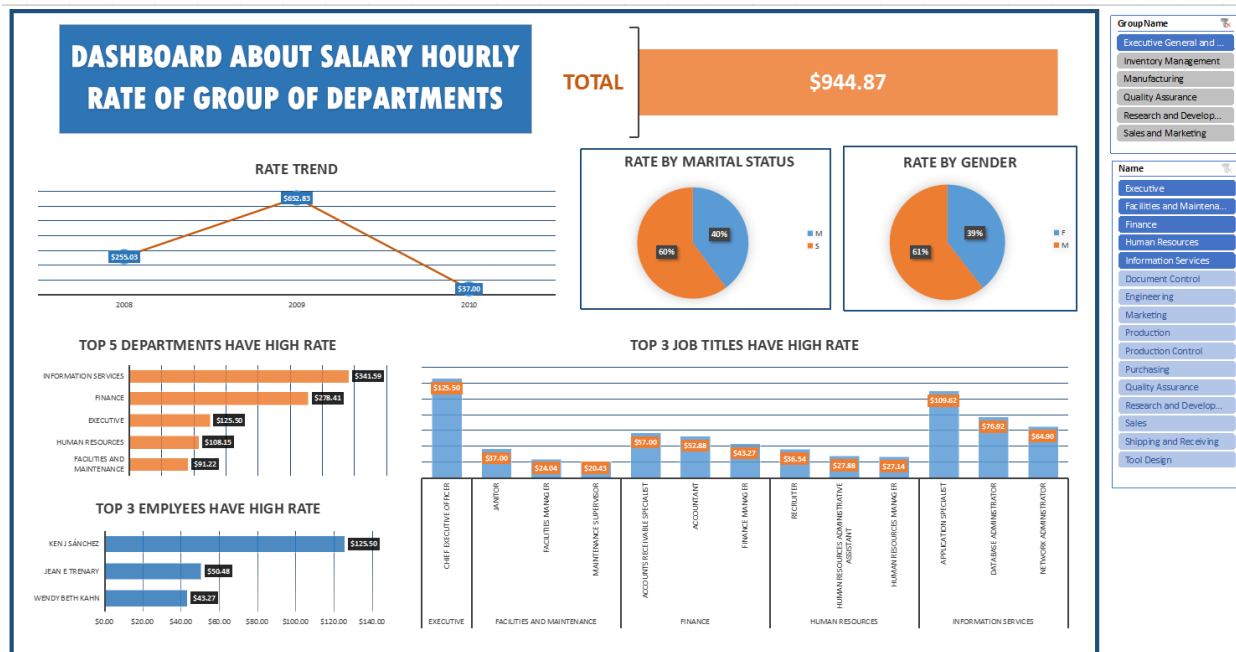


Figure 6.3.1: Dashboard salary hourly rate of group department

To visualize the detail of each group of the department, the manager can choose the name of group departments in the box on the right. Then, it will filter the information related to this group of departments. The below picture is the example for group Executive General and Administration.



In this dashboard, we can see the information about the group Executive General and Administration.

This group includes departments: Executive, Facilities and Maintenance, Information Services, Finance and Human Resources.

- The total salary rate of males is higher than one of females as well as the single is higher than people who are married.
- This group started to recruit staff and had a change in 3 years: 2008, 2009, 2010. In which, the total salary rate hourly in 2009 was the highest.
- Looking at this dashboard, the manager knows that 3 employees and 3 job titles in each department have the highest salary rate hourly.

From here, administrators can track the trends of salary rate hourly of each group of departments as well as each gender or married status to manage wages and have the appropriate policies for each group of department.

The following is the dashboard about the fluctuation of group departments. This dashboard also filters by groups of departments. It will show the figures about the number of department transfers of each group department and each department. In addition, it also describes the relation between time-off hours and salary in each departments.

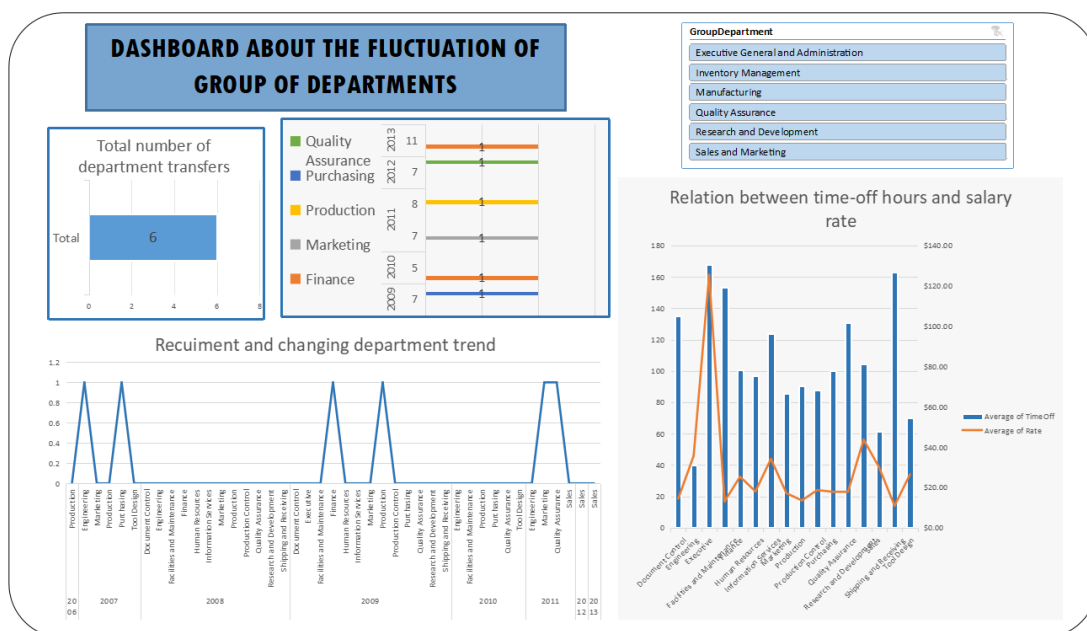
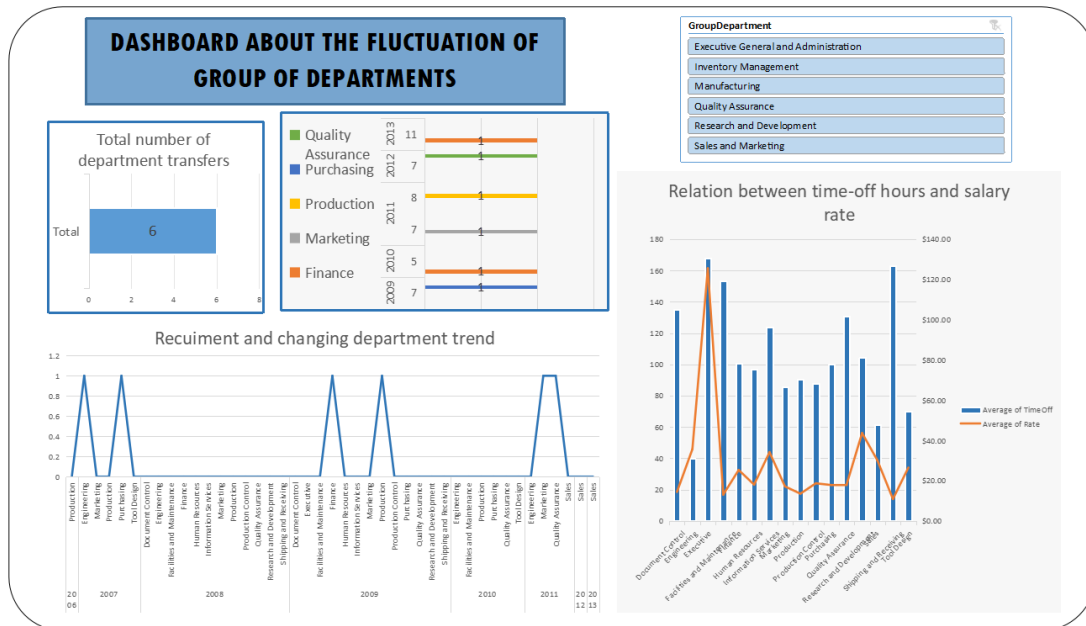
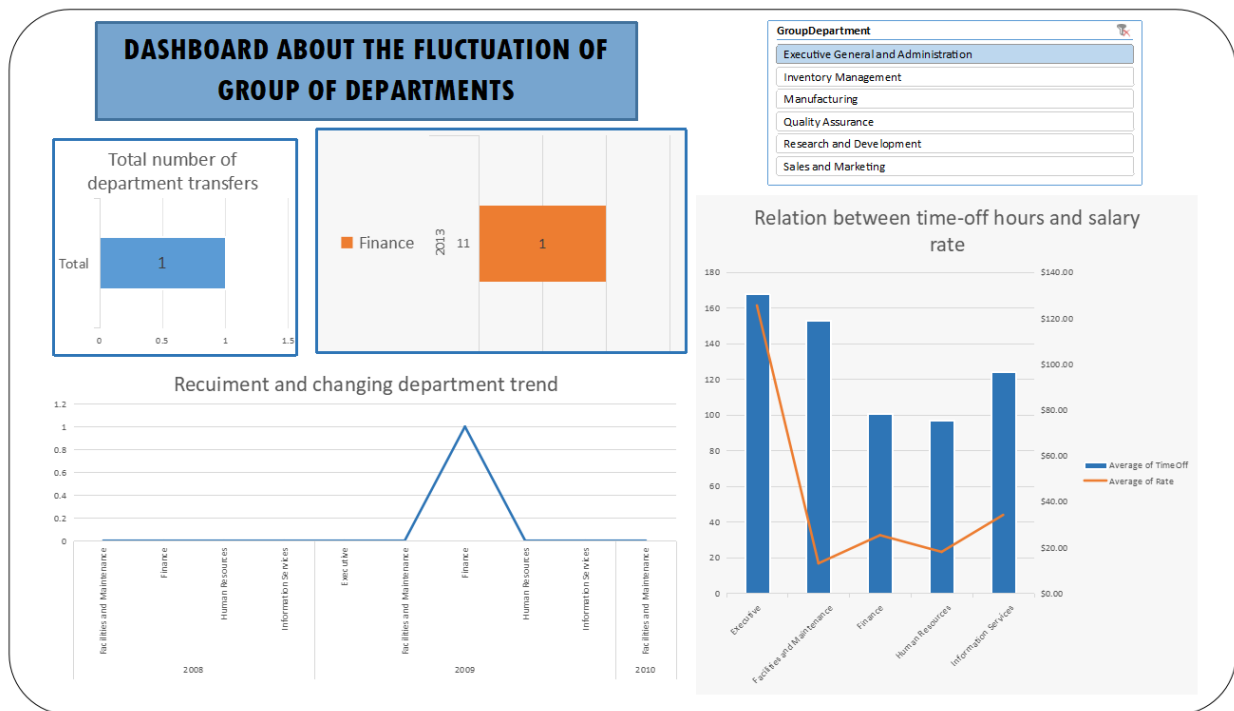


Figure 6.3.2: Dashboard the fluctuation of group departments



The picture is an example when the manager chooses the filter. This report helps the manager overview the fluctuation of each department about hire and transfer. As a result, the company can find the solution to deal with the problem of recruitment.



## 6.4. Data analysis with Tableau

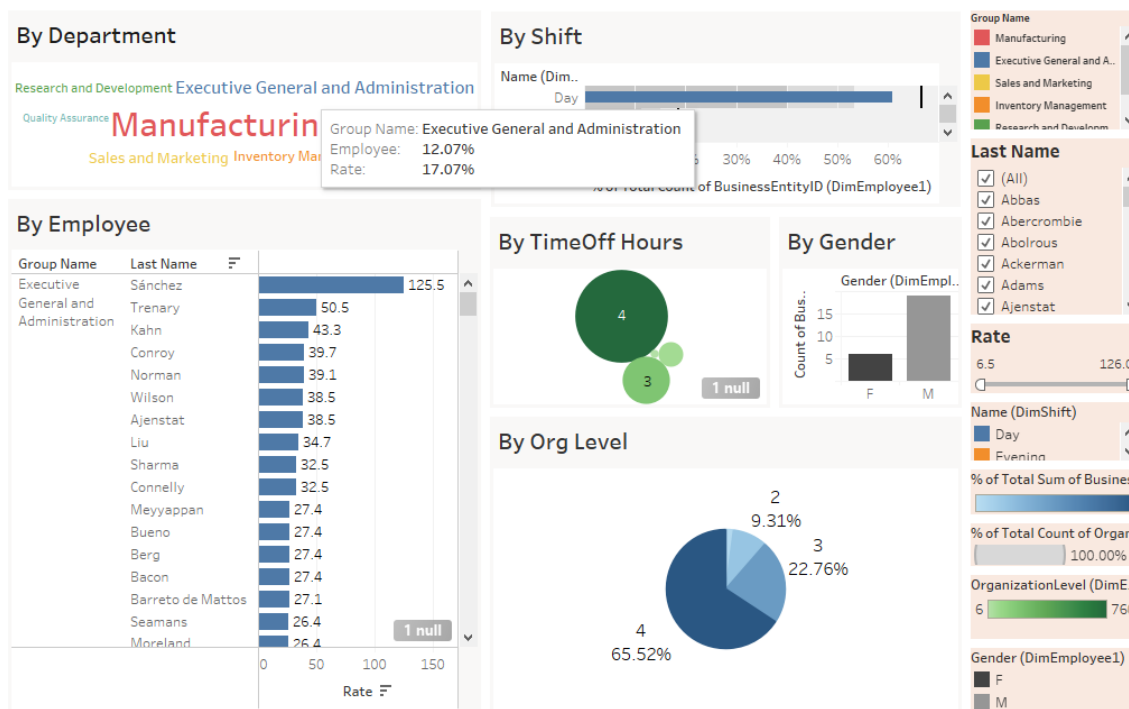


Figure 6.4: Dashboard salary management in Tableau

In the dashboard, the HR department can not only be used to manage the salary in the departments, the salary ratio between departments, but also see the specific salary of any employee in any department.

- The organizational structure of Adventure Work's departments can be easily seen through the "By Department" section. Each size of the department letter represents the size of the employees as well as the salary that the company must pay its employees. Here, we see that the subsystem "Manufacturing" has the largest word size. When we hover our mouse over it, it will show us the number of employees and the total salary of this module.
- For example, when hovering over the "Executive General and Administration" department with percent of employees is 12.07% and percent rate is 19.77%. They are responsible for effectively planning, organizing, leading and controlling everything in the company such as finance, people, and facilities. From there, get

the end goal that is good and beneficial for the company. In order for the company to develop strongly, managers need to pay attention, invest in these two departments.

- In "By Employee" will display a list of employees in that department with each salary. Sanchez, Trenary, and Kahn are the 3 highest-paid employees for the "Executive General and Administration" department.
- In "By Shift" and "By TimeOff Hours" will display specific information about the number of employees working in each shift. This is very important to keep track of the employee's performance, and to come up with effective policies for conducting recruitment.
- In "By Org Level" and "By Gender" will display specific information about the structure of personnel in the field. It helps us to have an overview of the gender balance in the company.

## 6.5. Time series and Forecasting

### 6.5.1. Forecast the total hired employees in the next 7 years

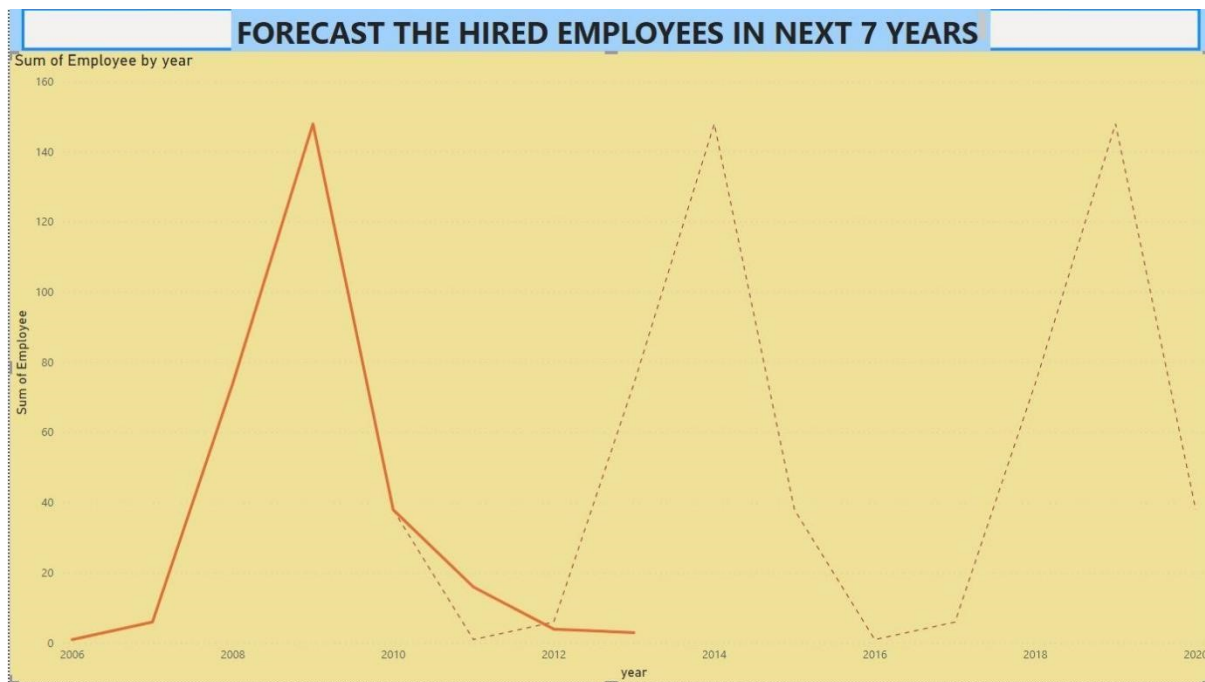


Figure 6.6: Forecast the hired employees in the next 7 years

Based on the total number of hired employees from 2006 to 2013, there is a prediction for the total number in the next 7 years (to 2020). 3 years (from 2010 to 2013) also has a prediction so that we can compare them and understand how it works. However, the prediction is just a prediction, managers can change it based on the business situation. So the real number can be exact or not.

#### 6.5.2. Time series plot of recruitment (by Excel)

The chart below shows the total employees recruited by Adventure Works each quarter from 2003 to 2013 and the prediction for 2014 by using Data Analysis - Regression in Excel. The data shows that the company hiring the most is in quarter 1 of 2009, after that the company almost did not hire any more.

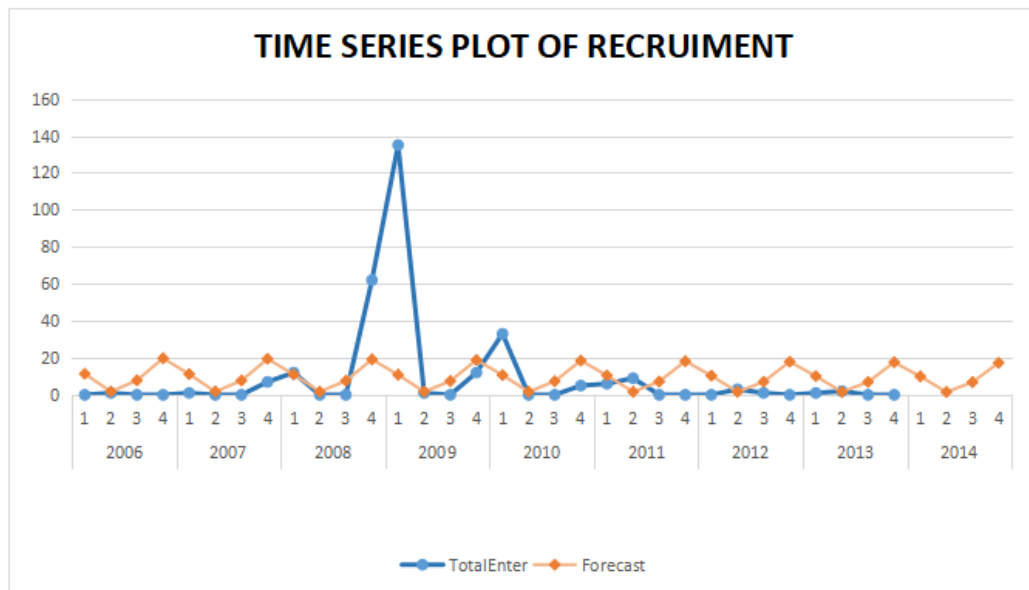
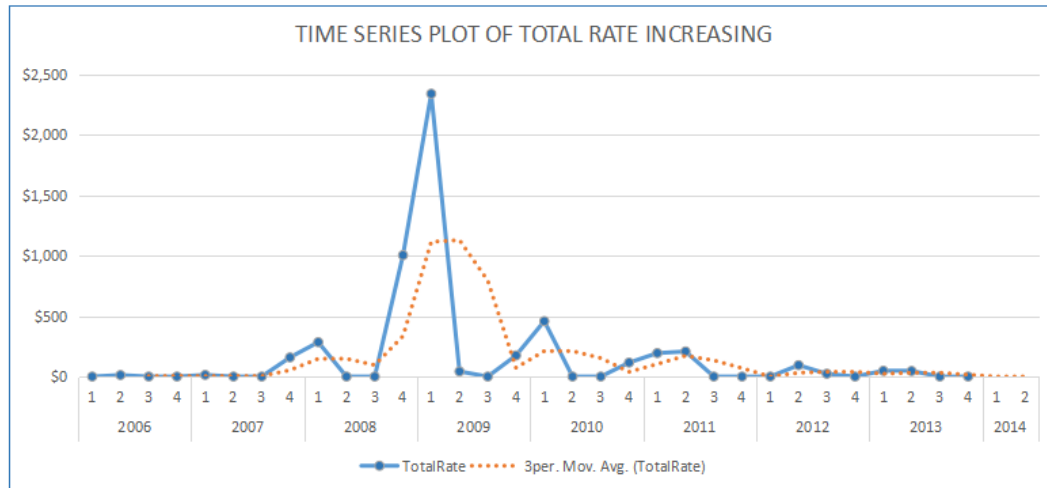


Figure 6.6.2: Time series plot of recruitment

#### 6.5.3. Time series plot of total rate increasing (by Excel)

The chart “TIME SERIES PLOT OF TOTAL RATE INCREASING” shows the increasing of the total salary hourly rate that the company must pay each time it hires more employees each quarter of year. In addition, the chart also gives the prediction about 2 first quarters of 2014 by “moving average” method. Because the prediction line shows the value of 0 so Adventure Works may not increase the total salary hourly rate to be paid to employees.



*Figure 6.6.3: Time series plot of total rate increasing*

## 6.6. Evaluation and Discussion

The dashboards are integrated by many charts to help managers easily overview the business situations and judge them. Moreover, combining many kinds of charts is a great way to make decisions based on the overview.

Besides, using a department slicer is very useful to narrow down the department they want to see. There are many total numbers and average numbers we use to statistically analyze the number of employees, average rate, total time-off hours, employees by gender,... in each department. Furthermore, there are also many lists of top job titles with the highest salary in each department and in all the company, or top departments, employees with the highest salary. Especially, combined chart between line chart and column is a great way to see the relationship between time-off hours and salaries, then suggest some advice to improve their salaries.

Regarding prediction, we use line charts to see clearly the fluctuation of the total employees of each department throughout the periods and also the changeable decision of those employees to find the reasons for that change and solutions to solve them if they are not fine.



## **CHAPTER 7: CONCLUSION AND FUTURE WORKS**

### **7.1. Results**

Firstly, this study is a great opportunity for all members of our team to apply the theory relating to this subject and others such as Management Information System, Introductory Database, Accounting, and Management into reality.

Secondly, this study proposed a data warehouse solution and provided many business facts to support managers to make the right decision.

Thirdly, in addition to the data warehouse proposed, this study also provided some KPI models, Time series and Dashboards in order to correctly evaluate the performance of the business.

### **7.2. Limitations**

Besides the results achieved, the project still has many limitations as follows:

The first phase of the project was delayed because the business case system survey and analysis were not deep and the data warehouse was not well understood.

The data warehouse has not yet integrated data related to sales and purchases, so it has not met all the requirements previously given.

The project is at the level of study and research, so the implementation of BI projects is still limited, mainly data representation, also data analysis, data comparison, and data analysis are not yet deep.

### **7.3. Future works**

From the results achieved and the limitations mentioned above, the thesis can be developed according to the following suggestions:

Build a better data warehouse, integrate sales and purchase data of employees to create monthly reports of salary and individual performance reports of all employees on the Adventure Works database platform.

Applying many other analytical tools and methods such as Qlik, Python, ... to have more opportunities to apply new technologies into the learning and research process as well as building a solid foundation when approaching actual projects at the enterprise.

Build a database of jobs such as recruitment to store application information of candidates. This makes recruitment easy when there is an urgent need to recruit a human resource.

## REFERENCES

1. **Aspin.** *SQL Server Reporting Services as a Business Intelligence*. 2015, viewed date 10/3/2021.
2. **Tableau.** The Top 7 Benefits of Business Intelligence. *TABLEAU*. [Online] 2021. <https://www.tableau.com/learn/articles/business-intelligence/enterprise-business-intelligence/benefits>, viewed date 5/3/2021.
3. **Laudan, Yonatan.** Five Steps for Building a Successful BI Strategy. *Sisense*. [Online] 9 12, 2019. <https://www.sisense.com/blog/five-steps-for-building-a-successful-bi-strategy/>, 9/3/2021.
4. **Guru99.** ETL (Extract, Transform, and Load) Process in Data Warehouse. *guru99*. [Online] 2021. <https://www.guru99.com/etl-extract-load-process.html>, viewed date 15/3/2021.
5. —. What is Data Warehouse? Types, Definition & Example. *Guru99*. [Online] 2021. <https://www.guru99.com/data-warehousing.html>, viewed date 18/3/2021.
6. —. *Guru99. What is Data Mart in Data Warehouse? Types & Example*. [Online] 2021. <https://www.guru99.com/data-mart-tutorial.html>, viewed date 25/3/2021.
7. **Klipfolio.** Key Performance Indicator (KPI) Definition. *Klipfolio*. [Online] 2021. <https://www.klipfolio.com/resources/articles/what-is-a-key-performance-indicator>, viewed date 20/4/2021.
8. **Bahu, Sami.** What Is the Importance of KPIs for Performance Measurement? *Unboxedtechnology*. [Online] 21, 2021. <https://unboxedtechnology.com/advantages-and-disadvantages-of-performance-indicators>, viewed date 15/4/2021.
9. **Idera.** MDX. *Idera*. [Online] 2021. <https://www.idera.com/glossary/mdx>, viewed date 23/4/2021
10. **Hevodata.** Star and Snowflake Schema: A Comprehensive Analysis. *Hevodata*. [Online] 2021. <https://hevodata.com/learn/star-and-snowflake-schema-analysis/>, viewed date 5/4/2021.