UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF INFORMATION SYSTEMS

FINAL PROJECT REPORT

BUSINESS INTELLIGENCE AND DECISION SUPPORT SYSTEMS

TOPIC: HUMAN RESOURCE

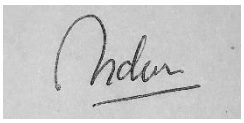
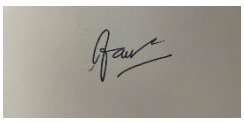
Group: 04

Instructors:

1. Assoc. Prof. Ho Trung Thanh, Ph.D.
2. Nguyen Van Ho, M.A.
3. Le Ba Thien, M.Sc.

Ho Chi Minh City, May 10th, 2024

LIST MEMBER

No.	Student code	Name	Point	Signature
1	K214060409	Tran Thanh Phong	10	
2	K214060390	Huynh Ngọc Dung	10	
3	K214061740	Doan Hoa Ky	10	
4	K214060418	Nguyen Mai Trinh	10	
5	K214061734	Tran Thi Van Anh	10	

ACKNOWLEDGEMENTS

With the completion of this report, our team would like to express our deep gratitude to all who support us. Learning is always a process of acquiring knowledge and skills. The implementation of this report has given us a significant impetus to become more confident in future scientific research papers, and at the same time helped us to make considerable progress in our graduation thesis journey. First, we would like to express our sincere thanks to the Rector of the University of Economics and Law for creating favorable conditions for us to approach the subject "***Business Intelligence And Decision Support Systems***" as well as providing the facilities with a modern library system, a variety of books and documents are convenient for searching and researching information.

We would like to express our deepest gratitude to the subject lecturer - Ho Trung Thanh – Nguyen Van Ho – Le Ba Thien for their resolute and detailed teaching so that we have enough knowledge and apply it to this project.

Furthermore, we would also like to thank our friends Hoa Ky, Mai Trinh, Thanh Phong, Van Anh, Ngoc Dung. The five of us worked together, read many research papers, processed the data, and finished this report as soon as possible.

Team of students doing research

Ho Chi Minh City, May 10th, 2024

COMMITMENT

Our team hereby declares that the research topic “BUSINESS INTELLIGENCE AND DECISION SUPPORT SYSTEMS TOPIC: HUMAN RESOURCE” is our research work under the scientific guidance the subject lecturer of - Ho Trung Thanh – Nguyen Van Ho – Le Ba Thien.

The data and research results in the thesis are collected, handled honestly, and objectively, and have clear origins. The content of the research has references to domestic and foreign studies and the references are fully listed in the list of references. With limited research time, shortcomings are inevitable. Looking forward to your suggestions for a more complete study.

Team of students doing research

Ho Chi Minh City, May 10th, 2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviations	*Defination*
HR	Human Resource
KPI	Key Performance Indicator
BI	Business Intelligence
OLAP	Online Analytical Processing
ETL	Extract, Transform, Load
EDA	Exploratory Data Analysis
DFD	Data flow diagram
FSAs	Flexible Spending Accounts
ROI	Return On Investment
BIDS	Business Intelligence Development Studio
HRIS	Human Resources Information System

CHAPTER 1: PROJECT OVERVIEW

1.1. Bicycle industry and AdventureWork Company overview

1.1.1. Overall bicycle industry in the global market

The global bicycle market size *(Grand View Research, 2023)* is worth USD 70,497.5 million and is expected to expand at a compound annual growth rate (CAGR) of 9.7% from 2023 to 2030.

In particular, the increasing number of people choosing cycling as a form of entertainment is expected to drive market growth. The preference for using bicycles as a convenient form of exercise to ensure healthy living, free from obesity and other disorders, is expected to further boost the market expansion. Currently, the current bicycle market is influenced by several factors such as:

- The trend of sustainable and environmentally friendly appliances is increasing, which is driving the demand for bicycles as an environmentally friendly alternative to traditional vehicles.
- Changing consumer lifestyles. Typically, young customers tend to prioritize personal experiences over material possessions. This makes bicycles relevant to this trend as they are not only a means of transportation but also a platform for outdoor activities and exploration. In addition, the desire to have a healthier and more active lifestyle has made people increasingly interested in cycling as a form of exercise and entertainment.
- Technological advancements such as lightweight materials, electronic components and smart features have played a key role in product innovation and differentiation in the market.
- Finally, the bicycle market is also affected by urbanization trends, especially with the increasing demand for compact and commuter-friendly designs.

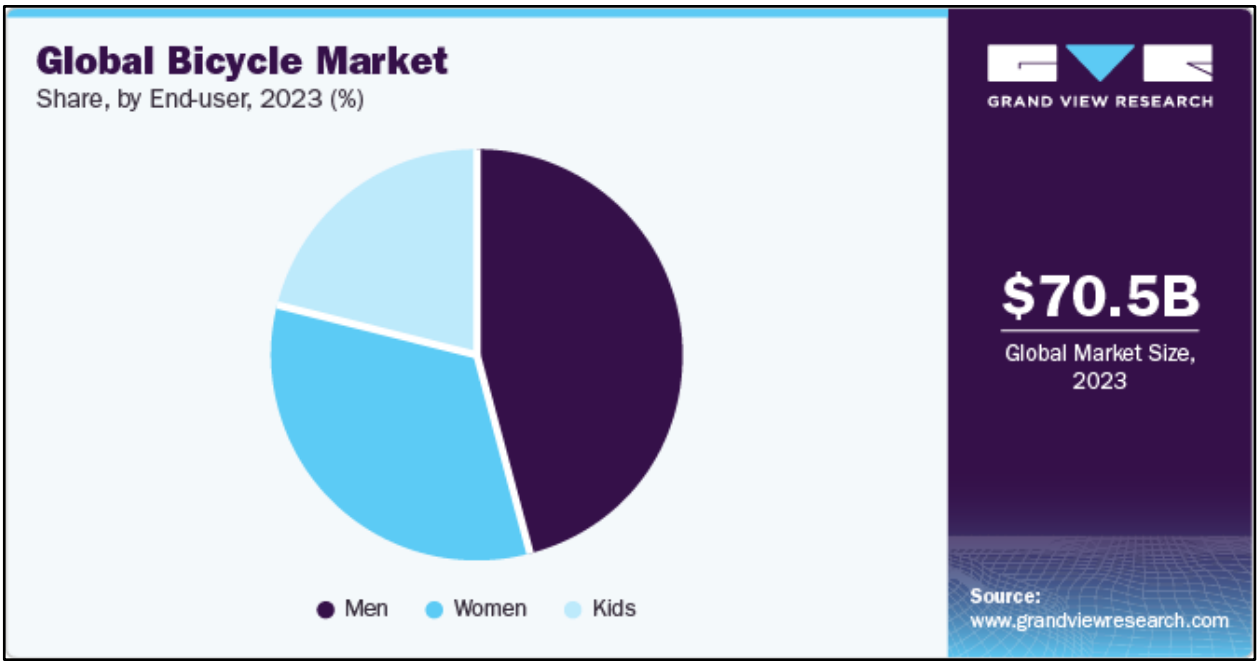


Figure 1.1. Pie charts represent detailed information about end users (Source: Bicycle Market Size, 2024)

The men segment holds the largest market share of over 45.7% by 2022 followed by women and children. According to statistical data published by the UK Department for Transport, in 2019, men in the UK took on average 25 cycling trips in a year compared to 10 cycling trips for women in the same year.

Figure 1.2. Detailed information chart about the bicycle industry area (Source: Bicycle Market Size, 2024)

Asia Pacific emerged as the dominant regional market in 2022 with a revenue share of 34.0%. Countries such as China, Japan and Singapore have also emphasized implementing the necessary infrastructure to encourage and support bicycle travel. Furthermore, Chinese bike sharing companies are actively targeting countries such as India and Australia to expand their operations. Europe is expected to have the highest growth rate of 11.1% during the forecast period. As it is home to several cities considered ideal for commuting by bike.

List of typical bicycle manufacturing companies

Table 1.1. List of typical bicycle manufacturing companies (Source: Bicycle Market Size, 2024)

No.	Company name	No.	Company name
1	Accell Group	6	Giant Bicycles

2	Atlas Cycles (Haryana) Ltd.	7	Merida Industry Co., Ltd
3	Avon Cycles Ltd.	8	Specialized Bicycle Components, Inc.
4	Cervelo	9	SCOTT Sports SA
5	Dorel Industries Inc.	10	Trek Bicycle Corporation

Bicycle market report scope

Table 1.2. Details of report attribute

(Source: Bicycle Market Size, 2024)

Report Attribute	Details
Market size value in 2023	USD 77,012.3 Million
Revenue forecast in 2030	USD 135,020.2 Million
Growth Rate	CAGR of 9.8% from 2023 to 2030
Actual Data	2017 - 2023
Forecast period	2024 - 2030

Quantitative units	Revenue in USD Million and CAGR from 2023 to 2030
Regional scope	North America, Europe, Asia Pacific, Latin America, MEA
Country scope	U.S.; Canada; U.K.; Germany; France; Italy; China; Japan; India; Australia; Brazil; Mexico; Kingdom of Saudi Arabia (KSA); UAE; South Africa
Key companies profiled	Accell Group; Atlas Cycles (Haryana) Ltd.; Avon Cycles Ltd.; Cervelo; Dorel Industries Inc.; Giant Bicycles; Merida Industry Co., Ltd; Specialized Bicycle Components, Inc.; SCOTT Sports SA; Trek Bicycle Corporation

1.1.2. Overview of the bicycle industry in Vietnam

Overview of the bicycle industry in Vietnam *(Statista)*:

- Revenue from the Bicycle market is expected to reach USD 88.85 million by 2024.
- Revenue is expected to have a compound annual growth rate (CAGR 2024-2028) of 2.98%, leading to a projected market value of $99.94 million by 2028.
- The bicycle market unit sales are expected to reach 0.54 million bicycles by 2028.

- The average price by volume of the Bicycle market in 2024 is expected to amount to 159.30 USD.
- From an international perspective, it can be seen that the largest revenue will be generated in China (10,520 million USD by 2024).

When other markets, especially technology products, plummeted due to buyers changing their habits and saving costs, large businesses began to pay more attention to the bicycle market as a place that could help them expand. revenues.

Sales of Bicycle products in September 2023 reached the highest level with 58.6 billion VND and 37.5 thousand in output. The bicycle market size in December 2023 reached 20.2 billion inAccording to a report by Metric.vn, only the children's bicycle segment on e-commerce platforms such as Shopee, Tiki, Lazada also saw total sales reaching 39 billion VND last year, an increase of 11.3% compared to 2021. This shows that the bicycle market in general is attractive, and there is still room for growth. sales and grew 29.2% lower than November 2023.

According to Statista's report, revenue from the bicycle segment in Vietnam is expected to reach 295.8 million USD in 2023. Revenue is expected to record a compound annual growth rate (CAGR) in the period 2023 - 2027 of 5.88%, leading to the Vietnamese bicycle market being able to record revenue reaching 371.8 million USD by 2027.

In terms of sales, Statista predicts that bicycle sales in Vietnam by 2027 *(Nguyễn A, 2023)* will reach 2.49 million units. In addition, annual bicycle sales in Vietnam, according to Statista's predictions, always remain at over 2 million units per year and grow steadily each year in the period 2023 - 2027.

When the covid-19 epidemic broke out again, bicycle sales still increased dramatically. The number of bicycles sold at many stores increased sharply, even being out of stock due to overcrowding. *(ACT Group, 2021).*

1.1.3. Business Case for the project

This project is based on Adventure Works Cycles, a fictitious company. Adventure Works Cycles is a large, multinational manufacturing company that produces and distributes metal and composite bicycles to commercial markets in North America, Europe, and Asia. The headquarters for Adventure Works Cycles is Bothell, Washington, where the company employs 500 workers. Additionally, Adventure Works Cycles employs several regional sales teams throughout its market base.

In recent years, Adventure Works Cycles bought a small manufacturing plant, Importadores Neptuno, which is located in Mexico. Importadores Neptuno manufactures several critical subcomponents for the Adventure Works Cycles product line. These subcomponents are shipped to the Bothell location for final23 product assembly. In 2005, Importadores Neptuno became the sole manufacturer and distributor of the touring bicycle product group.

Following a successful fiscal year, Adventure Works Cycles now wants to broaden its market share by targeting advertising to its best customers, extending product availability through an external Web site, and reducing the cost of sales by reducing production costs

1.2. Business demands/problems

Table 1.3. SWOT Model Adventure Works Cycles

S (Strong)	W (Weakness)
- Product quality guaranteed. - Good customer service, good after-sales service.	- Product price is quite high. - The production system is not really efficient. - High production and advertising costs.

- Wide distribution system, across countries around the world. - Reputable and trustworthy brand.	- The ability to adapt and innovate is still limited. - The data management system is not effective. - There are no diverse designs to suit many types of customers.
O (Opportunities)	**T (Threats)**
- Demand for products increases as more and more people prefer them today. - Market demand is increasing. - Thanks to technological development, the online shopping market is also increasingly promoted.	- Fierce competition from many competitors in the market. - Customer needs are always changing. - Raw material prices are increasing day by day. - Import and export taxes increased.

From the analysis of the above SWOT model, our team has identified the following business demands and business problems:

1.2.1. Business demands

Need for competition and development: As the market is growing and e-commerce is also growing, promoting competition to gain a solid foothold in the market is inevitable. The more diverse customers are, the higher their needs are, so businesses need to always change to adapt to the market, meet consumer needs and especially need new campaigns to attract customers. customers, increasing interactions and revenue while still satisfying them. In

addition, price is also an issue that businesses should consider. It is necessary to have specific strategies to adjust prices to suit the market and consumers' budgets according to each customer file.

The need to manage internal systems, production and distribution effectively: As technology has gradually invaded the market, automating processes to optimize costs is what businesses need to aim for. In addition to integrating software into production, it is also necessary to have predictive analysis and evaluation to help predict customer needs most accurately and effectively, and there is also a need to optimize the transportation process. , diversifying products but still maintaining good product quality to better meet customer needs. In addition, the business's internal system also needs to work more professionally and effectively, but consumes the least amount of time while still bringing good revenue and results.

1.2.2. Business problems

- ***Customer needs:*** Businesses will increasingly have to develop and diversify their products in terms of designs, colors, sizes, types, etc. to better suit customers of all levels. age of need. However, that also means businesses will have to spend higher costs to develop products as well as change features and properties to better suit consumer needs and social development.
- ***Product price:*** This is something businesses should consider, because the product price is quite high and will not be able to reach customer files such as workers and students because this is a very potential customer file that businesses need to exploit if they want to expand their markets and diversify their products.
- ***Competitors:*** As the product develops, it also means there will be more competitors in the market. It could be competing on price, product, raw material cost, promotion, etc. So that will be very harmful for Adventure Works Cycles in reaching and attracting more customers.

- ***Data management system:*** The data management system is not optimized and effective, making it difficult for Adventure Works Cycles to track, monitor and manage production, business, shipping and customer data . This leads to ineffective decision making and wasted resources and resources.

1.2.3. Reference to HR Subsystem

In addressing these exigencies, the HR subsystem emerges as a pivotal player. HR's strategic imperatives encompass talent acquisition, cultivation, and retention, pivotal in mitigating talent management challenges. Moreover, HR's adaptability to align strategies with evolving consumer preferences and technological frontiers is instrumental in ensuring workforce agility and competencies alignment. Additionally, HR serves as a linchpin for ensuring regulatory compliance and fostering a conducive organizational culture conducive to innovation and growth. Leveraging data analytics capabilities within the HR domain augments HR's capacity to decipher workforce dynamics, discern skill gaps, and optimize resource allocations to fulfill overarching business imperatives. Thus, integrating HR data analytics prowess within the broader Business Intelligence framework emerges as imperative to effectively surmount these business exigencies.

1.3. Objectives of the project

1.3.1. General Objective

The project aims to leverage Microsoft's Adventure Works Cycles database for HR analytics and decision-making purposes. Key objectives include understanding the database structure, data cleaning, and preparation, utilizing SQL Server tools for analysis, and implementing pertinent HR-centric key performance indicators (KPIs). Through the use of Power BI, the data will be transformed into visually engaging dashboards and reports, facilitating the identification of HR challenges, trends, and opportunities. The overarching goal is to present clear insights to management, thereby enabling effective

decision-making processes, and fostering continuous improvement in alignment with evolving business requirements and feedback mechanisms.

1.3.2. Specific Objectives

- Integrate Adventure Works Cycles database research into existing business processes.
- Review decision support models and BI systems.
- Design and construct a data warehouse.
- Implement customized ETL processes for financial reporting needs.
- Explore SQL Server Business Intelligence solutions.
- Establish KPIs for HR analytics.
- Identify data sources within the HR department.
- Centralize and store HR data from various sources.
- Emphasize key areas such as employee information, recruitment, training, and compensation.
- Construct an OLAP Cube for multidimensional analysis.
- Develop a user-friendly and interactive visualization dashboard.

1.4. Business Questions

- What is the average salary across different departments? Are there any gender pay gaps within the company?
- How does the workforce demand vary across different departments? Are there any departments that are overstaffed or understaffed based on their current and projected workforce demand?

- How does the time-off usage vary across different departments? Are there any correlations between time-off usage and employee satisfaction or productivity?
- What are the most important KPIs for measuring HR performance and effectiveness and how do they align with organizational goals and objectives?

1.5. Scope of the project

Space scope:

- The database focuses on the production and sale of metal and composite-material bicycles. The company operates in the markets of North America, Europe, and Asia. While the company's main headquarters is located in Bothell, Washington, with a staff of 290 employees (2013), it also has regional sales personnel catering to the respective markets.
- The Business Intelligence (BI) solution is being implemented to support the HR Department in managing employee-related activities and payroll. It is important to note that, at this stage, the scope of the solution is limited to assisting the HR Department exclusively. Other departments, including support functions, are currently not within the scope of this implementation.

Time scope:

- AdventureWorks is a sample database for Microsoft SQL Server 2008 to 2014. This is a documentation of this database created with Dataedo.
- Project time: Project implementation period is from March 2 to May 10, 2024.

1.6. Value and desired outcome of the project

The value of the project:

- Gain insights into HR business processes.

- Enable comprehensive analysis of HR processes, leading to streamlined operations and enhanced efficiency.
- Analyze company data to construct a comprehensive data warehouse for Adventure Works Cycles.
- Provide a unified and intuitive platform for visualizing and interpreting HR metrics, fostering informed decision-making at high levels of the organization.
- Facilitate proactive identification of trends, patterns, and opportunities within HR data, enabling timely interventions and optimizations.
- Drive organizational growth and competitiveness by leveraging data-driven insights to align HR strategies with overarching business goals and objectives.

Desired outcome of the BI Solution project:

- Get a streamlined data warehouse to better track and retrieve HR data.
- Have charts and dashboards accurately reflect the requirements of the project.
- Allows to query information accurately immediately.
- Timely and accurate information sharing between people, every department in the business helps make decisions quickly and accurately
- Save time, human resources and optimize profits.
- Streamlining HR processes through data-driven decision-making, leading to improved efficiency in recruitment, onboarding, performance management, and employee retention.
- Enhancing overall organizational performance by aligning HR strategies with business objectives, leveraging data insights to support strategic workforce planning and talent management initiatives.

1.7. Project Budget

One-time costs refer to expenses that are incurred only once during a specific period, typically related to the initiation or setup of a project or the acquisition of assets. These costs are not expected to recur regularly. Examples of one-time costs include purchasing equipment, software licenses, or hiring consultants for a specific project.

Table 1.4. Project Budget

One-time costs	Estimated Cost
Procurement of advanced data analytics software and licenses	$50,000
Purchase of high-performance servers and hardware infrastructure	$30,000
Development and customization of data visualization tools	$20,000
Training and education programs to enhance employee skills	$10,000
Implementation and configuration of database management systems	$20,000
Investment in cloud computing infrastructure	$50,000
Establishment of a secure data storage and backup system	$10,000
Other incurred costs (10%)	$20,000
Total	**$210,000**

1.8. Models/Process

The team's BI project implementation process will include 7 steps:

Define Project Objectives and Scope: Define the main objectives and the scope of the project, including specific HR data and processes to be analyzed and visualized.

Gather Requirements and Analysis: Collect and analyze available human resources data and determine the information needed to perform the analysis.

Data Exploration and Understanding*:*

- Explore the Adventure Works Cycles dataset to gain a comprehensive understanding of the available tables, fields, and relationships.
- Identify key metrics and dimensions relevant to project goals.

Design Data warehouse and data mart: Design a data model for the Power BI project based on requirements and gathered information. Identify relationships between HR data tables and other entities such as departments, positions, and relevant information.

Extract, Transform, Load (ETL):

- Extract data from HR sources.
- Perform necessary transformations to standardize and clean the data.
- Load transformed data into the Power BI environment.

Develop Reports and Visualization:

- Build reports and visualize HR data, including charts showing workforce trends, historical analysis of employment, performance evaluation, etc.
- Create dashboards to provide an overview of the organization's HR situation.

Testing and Debugging: Conduct data and dashboard testing to ensure system accuracy and performance.

Evaluation and Improvement: Evaluate project effectiveness based on initial goals and requirements. Finally, make suggestions for future development.

1.9. Structure of project

Chapter 1. Overview of the project

The project overview revolves around a review to get specific data from AdventureWork Company. State Business demands/problems, Business Goals, Business Questions, Scopes, Project Budget, Project deliverables, Models/Process, Tools, Methods and Programming languages, Structure of project.

Chapter 2. Defining requirements business/KPIs, data and quality

The content of chapter 1 focuses on Documentation techniques, Business requirements in detail. Analyze KPIs corresponding to each business requirement, Data requirements, Functional requirements, Technical requirements. In addition, BI and Data warehouse Framework or Solutions with the DW and Data mart design approach strategy of the assigned subsystem.

Chapter 3. Data preparation and Data modeling

This chapter 2 focuses on Data preparation (Data collection, Data description, Data understanding, EDA, Pre-processing). Overall description of HR Module's AdventureWorks data. Build bus matrix, design Data model for Data warehouse/Data marts.

Chapter 4. Data Integration

Focus on building strategies (ETL, ELT,...)/diagrams/models for front-end and intermediate integration, integration processes for both dimension tables and facts tables, installing integration processes on SSIS tools (practice) and record/present data integration results for the HR module).

Chapter 5. Multi-dimensional data analysis

In this chapter, we focus on strategy, goals, meaning of multidimensional database, meaning of multidimensional data analysis in the HR module, and practice. Results of data warehouse analysis by practicing OLAP techniques on SSAS, MDX language, analyzing KPIs in SSAS with requirements as well as built KPIs.

Chapter 6. Visualization and Discussion

Scope, goals, diagrams of Dashboards are built based on requirements. Exploit charts and dashboards as well as interpret insights from dashboards in detail and management implications related to KPIs, requirements, questions as well as operations related to HR Module. This section combines visualization and discussion of the results of other analytical models to mine insights from the data warehouse.

Chapter 7. Conclusion and Future Works

Final conclusion in the process of realizing what the project has accomplished, the advantages and disadvantages and future development direction of the implemented project.

CHAPTER 2: DEFINING REQUIREMENT BUSINESS/KPIS, DATA AND QUALITY

The content of chapter 2 focuses on Documentation techniques, Business requirements in detail. Analyze KPIs corresponding to each business requirement, Data requirements, Functional requirements, Technical requirements. In addition, BI and Data warehouse Framework or Solutions with the DW and Data mart design approach strategy of the assigned subsystem.

2.1. Documentation techniques

2.1.1. Meaning and Benefits of Documentation Techniques

Documentation explains how a system works, including the who, what, when, where, why, and how of data entry, data processing, data storage, information output, and system controls. Popular means of documenting a system include diagrams, flowcharts, tables, and other graphical representations of data and information. These are supplemented by a narrative description of the system, a written step-by-step explanation of system components and interactions.

Documentation tools are important on the following levels:

- At a minimum, you must be able to read documentation to determine how a system works.
- You may need to evaluate documentation to identify internal control strengths and weaknesses and recommend improvements as well as to determine if a proposed system meets the company’s needs.
- More skill is needed to prepare documentation that shows how an existing or proposed system operates.

2.1.2. Common Documentation Techniques

Data flow diagram (DFD), a graphical description of data sources, data flows, transformation processes, data storage, and data destinations Flowchart, which is a graphical description of a system. There are several types of flowcharts, including:

- Document flowchart, which shows the flow of documents and information between departments or areas of responsibility.
- System flowchart, which shows the relationship among the input, processing, and output in an information system.
- Program flowchart, which shows the sequence of logical operations a computer performs as it executes a program.
- Business Process diagrams, which is a graphical description of the business processes used by a company.

Accountants use documentation techniques extensively. Auditing standards require that independent auditors understand the automated and manual internal control procedures an entity uses. One good way to gain this understanding is to use business process models or flowcharts to document a system, because such graphic portrayals more readily reveal internal control weaknesses and strengths.

Documentation tools are also used extensively in the systems development process. In addition, the team members who develop information systems applications often change, and documentation tools help the new team members get up to speed quickly.

Documentation is easier to prepare and revise when a software package is used. Once a few basic commands are mastered, users can quickly and easily prepare, store, revise, and print presentation quality documentation.

2.1.3. Application of Documentation Techniques in this BI Project

Within the ambit of the Business Intelligence (BI) project for Adventure Works Cycles, the application of documentation techniques assumes paramount importance in ensuring clarity, precision, and reproducibility throughout the project's lifecycle. Here's how these techniques can be effectively employed:

Narrative Descriptions:

- Purpose: Offer a written explanation of system components, interactions, and procedures.
- Application: Draft narrative descriptions detailing the step-by-step processes involved in data entry, processing, storage, analysis, and reporting within the BI system.
- Benefits: Serve as a comprehensive reference guide for stakeholders, ensuring consistency and accuracy in system operations.

Importance of Documentation Techniques in the BI Project:

- Clarity and Understanding: Documentation techniques facilitate a clear understanding of the BI system's architecture, processes, and functionalities, enabling stakeholders to grasp its operation comprehensively.
- Internal Control Evaluation: Through documentation evaluation, internal control strengths and weaknesses within the BI system can be identified, leading to recommendations for improvement and risk mitigation.
- Onboarding and Knowledge Transfer: Well-documented BI systems expedite the onboarding process for new team members, enabling them to familiarize themselves with the system's intricacies quickly and efficiently.

- Revision and Maintenance: Regularly updated documentation ensures that the BI system remains aligned with evolving business requirements and technological advancements, facilitating seamless revisions and maintenance activities.

2.2. Business requirements

2.2.1. Employee Information Management

Employee information management is a critical aspect of HR operations for Adventure Works Cycles. By establishing clear business requirements for employee information management, Adventure Works Cycles can effectively organize and maintain employee data, ensure compliance with privacy regulations, and streamline HR processes. Implementing robust employee information management processes is essential for Adventure Works Cycles to efficiently manage its workforce and support HR decision-making.

Adventure Works Cycles has defined its business requirements for employee information management. This includes identifying the specific types of employee data that need to be collected and maintained, such as personal information, employment history, performance evaluations, training records, and benefits enrollment details. Adventure Works Cycles also emphasizes the need for accurate and up-to-date employee records.

To streamline employee information management, Adventure Works Cycles requires a centralized HRIS (Human Resources Information System) or employee database. This system should serve as a secure and reliable repository for storing and accessing employee data. It should also have the capability to generate reports and provide relevant HR analytics to support strategic decision-making.

Adventure Works Cycles further specifies the importance of data security and privacy measures. The company requires robust data protection protocols to ensure employee information is safeguarded against unauthorized access or breaches. This includes

implementing access controls, encryption mechanisms, and regular data backups to maintain data integrity and confidentiality.

Additionally, Adventure Works Cycles emphasizes the need for efficient data entry and maintenance processes. The company requires user-friendly interfaces and standardized data entry forms to ensure consistency and accuracy in employee data collection. Regular data audits and updates should also be conducted to keep the employee information repository current and relevant.

In conclusion, Adventure Works Cycles has defined specific business requirements for employee information management. By establishing these requirements, Adventure Works Cycles aims to effectively organize and maintain employee data, ensure compliance with privacy regulations, and streamline HR processes. The company emphasizes the need for a centralized HRIS or employee database, robust data security measures, and efficient data entry and maintenance processes. By implementing these requirements, Adventure Works Cycles can efficiently manage employee information, support HR decision-making, and maintain compliance with data privacy regulations.

2.2.2. Recruitment

Adventure Works Cycles places a strong emphasis on defining specific business requirements to ensure a successful recruitment process. These requirements serve as guidelines for attracting and selecting the most qualified candidates who align with the company's objectives and values. By clearly outlining these requirements, Adventure Works Cycles aims to streamline its recruitment efforts and build a talented workforce.

Adventure Works Cycles has established a need for clearly defined talent requirements and job specifications. Through comprehensive job analyses, the company has identified the essential skills, qualifications, and experience necessary for each position within the organization. These specific requirements enable Adventure Works Cycles to effectively

assess candidates and ensure a strong fit between their capabilities and the job requirements.

In terms of recruitment strategy, Adventure Works Cycles has outlined the desired sourcing channels and methods to reach potential candidates. The company may leverage online job platforms, professional networks, industry-specific events, or partnerships with educational institutions to attract a diverse pool of qualified candidates. By utilizing these channels, Adventure Works Cycles aims to maximize its reach and tap into a wider talent pool.

Moreover, Adventure Works Cycles emphasizes the importance of rigorous selection processes and criteria. The company has established structured interview questions, assessment tests, and reference checks that align with its core values and job requirements. These selection criteria enable Adventure Works Cycles to evaluate candidates objectively and make informed decisions regarding their suitability for the available positions.

In conclusion, Adventure Works Cycles acknowledges the significance of setting clear business requirements in recruitment to attract top talent. By defining talent needs, outlining a targeted recruitment strategy, and establishing rigorous selection processes, Adventure Works Cycles aims to acquire highly qualified individuals who can contribute to the company's growth and success. These requirements serve as a foundation for an effective recruitment process that aligns with Adventure Works Cycles' objectives and fosters a talented workforce.

2.2.3. Leave and Attendance Management

Adventure Works Cycles recognizes the importance of establishing specific business requirements for leave and attendance management. By outlining these requirements, Adventure Works Cycles aims to effectively track and manage employee absences, ensure compliance with labor laws and company policies, and maintain efficient scheduling and resource allocation. Implementing robust leave and attendance management processes is crucial for Adventure Works Cycles to maintain productivity and employee satisfaction.

Adventure Works Cycles has defined its business requirements for leave and attendance management. This includes determining the types of leaves offered, such as annual leave, sick leave, maternity/paternity leave, and other relevant categories. Adventure Works Cycles has also established guidelines for requesting and approving leaves, including the process for submitting leave applications, required documentation, and timelines for approval.

To streamline the leave and attendance management process, Adventure Works Cycles requires a centralized system or software solution. This system should allow employees to easily request leaves, enable managers to review and approve requests efficiently, and provide accurate tracking of leave balances and attendance records. Automation features, such as notifications and reminders, are also necessary to ensure timely communication and adherence to leave policies.

In addition to tracking leave and attendance, Adventure Works Cycles has specified the need for clear policies and procedures to manage attendance-related issues, such as late arrivals, early departures, and unauthorized absences. This includes defining disciplinary actions for excessive absenteeism or repeated violations of attendance policies. Consistent enforcement of these policies will help maintain a fair and productive work environment.

In conclusion, Adventure Works Cycles has identified specific business requirements for leave and attendance management. By establishing these requirements, Adventure Works Cycles aims to effectively manage employee absences, ensure compliance with labor laws and company policies, and optimize scheduling and resource allocation. Defining leave types, establishing guidelines for leave requests and approvals, implementing a centralized system, and enforcing attendance policies are crucial steps for Adventure Works Cycles to successfully manage leave and attendance in the organization.

2.2.4. Reporting and Analytics

Adventure Works Cycles recognizes the importance of defining specific business requirements for reporting and analytics. By outlining these requirements, Adventure Works Cycles aims to gather and analyze relevant data, gain insights into various aspects of the business, and make informed strategic decisions. Implementing robust reporting and analytics processes is crucial for Adventure Works Cycles to drive efficiency, measure performance, and achieve its business objectives.

Adventure Works Cycles has established its business requirements for reporting and analytics. This includes identifying the key metrics, KPIs (Key Performance Indicators), and data points that are essential for monitoring and evaluating the organization's performance. These requirements may vary across different departments and functions within Adventure Works Cycles, such as sales, marketing, finance, and operations.

To streamline the reporting and analytics process, Adventure Works Cycles requires a centralized data repository. This repository should effectively consolidate data from various sources within the organization, ensuring data accuracy and integrity. Adventure Works Cycles also emphasizes the need for robust data analysis tools and software that can handle complex data sets and generate meaningful insights.

Adventure Works Cycles further specifies the importance of real-time reporting capabilities. The company requires timely and up-to-date reports to enable quick decision-making and proactive management. This includes the ability to generate automated reports on a regular basis, as well as ad-hoc reporting capabilities to address specific business needs and emerging trends.

In conclusion, Adventure Works Cycles has defined its specific business requirements for reporting and analytics. By establishing these requirements, Adventure Works Cycles aims to leverage data-driven insights to drive efficiency, measure performance, and make informed strategic decisions. The company emphasizes the need for a centralized data

repository, robust data analysis tools, and real-time reporting capabilities to effectively meet its reporting and analytics needs. By implementing these requirements, Adventure Works Cycles can enhance its ability to monitor, evaluate, and optimize its business operations.

2.3. IT requirements Analysis

The IT requirements for Business Intelligence (BI) solutions in HR turnover management are crucial for effective data management and analysis. Key components include robust ETL tools for data collection, OLAP for efficient querying, and SQL for database interrogation. Advanced Data Mining and varied analytical approaches are essential for uncovering trends and patterns. Visualization tools like Power BI are necessary for translating data into actionable insights. Integration with existing HR systems, scalability, performance optimization, and data security are critical considerations. Comprehensive user training and support are vital for successful BI implementation, enabling organizations to make informed decisions and drive sustainable growth.

2.4. Business Process

Any human resource management task can be converted into an HR process. The most common are: Payroll, Recruitment, Onboarding, Compliance, Retention, Compensation, Training, Employee relations, Performance management, Offboarding. However, in this project we focus on two main processes: Payroll and Recruitment

2.4.1. Payroll process

2.4.1.1. Definition

Payroll is the process of paying a company's employees, which includes tracking hours worked, calculating employees'pay, and distributing payments via direct deposit to employee bank accounts or by check. However, companies must also perform accounting functions to record payroll, taxes withheld, bonuses, overtime pay, sick time, and vacation

pay. Companies must put aside and record the amount to be paid to the government for Medicare, Social Security, and unemployment taxes.

2.4.1.2. Payroll process

Pre-Payroll (Preparation)

Employee Data:

- Verification and Updates: Regularly verify and update employee information including names, addresses, Social Security numbers, tax filing status, bank account details (for direct deposit), emergency contact information, and dependent information (for benefits administration).
- New Hire Onboarding: Establish a smooth onboarding process to collect and verify employee data accurately and efficiently. (Source: WorldatWork)

Wage and Salary Data:

- Pay Structures: Define different pay structures for various employee categories (e.g., hourly, salaried, commissioned).
- Bonuses and Incentives: Establish clear guidelines and criteria for awarding bonuses and incentives to avoid confusion or disputes.
- Cost-of-Living Adjustments (COLA): Implement COLA adjustments based on inflation or changes in living standards, if applicable.

Timekeeping:

- Time Tracking Methods: Explore various time tracking methods like manual timesheets, automated clock-in systems, project management tools, and mobile apps, choosing the most suitable option for your company's needs and workforce.
- Attendance Policies: Establish clear attendance policies outlining expectations, consequences for tardiness or absenteeism, and procedures for reporting absences.

Processing and Payment

Gross Pay Calculation:

- Commission-Based Pay: Calculate commission based on agreed-upon rates and employee performance metrics.
- Piecework Pay: Calculate pay based on the number of units produced, ensuring compliance with minimum wage regulations.

Overtime Pay: Different Overtime Rules: Be aware of different overtime rules for exempt and non-exempt employees, and calculate overtime pay accordingly.

Deductions:

- Flexible Spending Accounts (FSAs): Manage employee contributions to FSAs for pre-tax deductions for qualified medical and dependent care expenses.
- Student Loan Repayment: Offer student loan repayment programs as voluntary deduction options for employees.
- Union Dues and Other Authorized Deductions: Collect and remit union dues and other authorized deductions as per agreements and employee consent.

Net Pay Calculation: Consider Garnishments: Account for court-ordered garnishments for child support, student loans, or other debts in net pay calculations.

Payment and Reporting:

- Direct Deposit Enrollment: Encourage employees to enroll in direct deposit for faster, safer, and more environmentally friendly payments.
- Pay Stubs: Provide clear and detailed pay stubs with all earnings, deductions, and tax information for employee records.

- Internal Reporting: Generate comprehensive reports for payroll expenses, tax liabilities, employee pay history, and benefits administration for informed decision-making.

2.4.2. Recruitment process

Recruiting is the process of finding & attracting capable applicants for employment. The process begins when new recruits are required & ends when their applications are submitted. The result is a pool of applicants from which new employees are selected. (Rahaman, 2016)

In a multinational company, there are typically three types of recruitment strategies employed:

- Recruitment of Fresh Graduates as Knowledge Workers: This type of recruitment focuses on hiring recent graduates who possess the necessary knowledge and skills to contribute to the company's operations. These individuals are typically hired for entry-level positions and are expected to bring fresh perspectives, innovative ideas, and a strong educational foundation to the organization. The company may actively engage with universities and educational institutions to attract and recruit top talent.
- Recruitment of Specialist Experienced Individuals for Specific Purposes: In certain cases, a multinational company may require individuals with specialized expertise and experience to fulfill specific roles or projects. This type of recruitment focuses on identifying and attracting professionals who have demonstrated proficiency in a particular field or industry. These individuals bring in-depth knowledge and a track record of success, enabling the company to address unique challenges or pursue strategic initiatives.
- Recruitment of Skilled and Labor-Based Workers for Events: Multinational companies often engage in large-scale events, such as product launches, trade shows, or promotional campaigns. For these events, the company may require a

temporary workforce comprising skilled and labor-based workers. Recruitment efforts in this area focus on identifying individuals with the necessary skills, such as event management, logistics, customer service, or production. These workers play a crucial role in ensuring the smooth execution of the event and meeting the company's objectives.

Typically, there are two primary types of recruitment channels: internal and external.

- Internal Recruitment Channel: Internal recruitment channels are commonly employed by large organizations to identify and develop existing employees for higher-level positions within the company. This approach aims to nurture talent, promote career growth, and retain valuable employees. There are two primary methods used within the internal recruitment channel:
    + Job Postings allow current employees to apply for open positions, promoting career advancement and identifying qualified candidates.
    + Departing Employees create vacancies that can be filled by existing employees, providing growth opportunities and internal talent transfer. These strategies foster employee development and retention within the organization.
- External Recruitment Channel: involve reaching out to the external community to attract candidates for job openings. This includes advertising positions through various channels, employee referrals, temporary help services, partnerships with schools and universities, engagement with professional organizations, and international recruitment. These methods expand the candidate pool, bring in diverse perspectives, and help organizations find the best candidates for their needs.

2.5. Key Performance Indicators

KPI stands for key performance indicator, a quantifiable measure of performance over time for a specific objective. KPIs provide targets for teams to shoot for, milestones to gauge

progress, and insights that help people across the organization make better decisions. From finance and HR to marketing and sales, key performance indicators help every area of the business move forward at the strategic level.

Key Performance Indicators (KPIs) play a crucial role in ensuring that your teams are aligned with the overall goals of the organization. Here are some of the primary reasons why implementing KPIs is essential:

- Keep your teams aligned: KPIs enable you to measure project success and employee performance, ensuring that all teams are working towards the same objectives and moving in the same direction.
- Provide a health check: KPIs provide a realistic assessment of your organization's health by considering various factors such as risk indicators and financial metrics. They offer valuable insights into the overall well-being of your company.
- Make adjustments: KPIs help you identify both successes and failures, allowing you to make informed decisions about what strategies are effective and which ones need adjustment. This enables you to allocate resources more efficiently and focus on what yields positive results.
- Hold your teams accountable: By implementing KPIs, you can hold your teams accountable for their performance. Employees can track their progress against the defined indicators, and managers can facilitate progress and address any areas that require attention. This ensures that everyone is contributing value and actively working towards achieving the organization's goals.

2.5.1. KPI 1: Pay rate management

Main requirement: calculate the average pay rate for each department in an organization.

Objective: analyze and compare the average pay rates across different departments in order to understand the distribution of salaries within the organization.

Input:

- Sum of Pay Rate: Total salary amount of all employees in the organization.
- Total number of Transactions: The total number of pay rate change records in the organization.

Output:

Average Pay Rate: The average pay rate calculated by dividing the total pay rate amount by the total number of transactions and multiplying by 100.

Business rule: The data is consistent and accurate across all sources.

Key performance indicator: This KPI represents the average pay rate and serves as a measure of pay rate management effectiveness. It can be monitored over time to track changes and trends in employee's pay rate.

$$Average\ Pay\ Rate = \frac{Sum\ of\ Pay\ Rate}{Total\ number\ of\ Transactions} \times 100 \qquad \textbf{(2.1)}$$

2.5.2. KPI 2: Time off hours management

Main requirement: Measure the ratio of days off and sick in relation to actual work hours.

Objective: Measure and monitor the utilization of time off and sick leave compared to actual work hours.

Input:

- Time off Hours: Total number of hours taken as time off or sick leave.
- Actual Work Hours: Total number of hours worked by employees.

Output:

Ratio of days off and sick: The ratio calculated by dividing the time off hours by the actual work hours and multiplying by 100.

Business rule: The data is consistent and accurate across all sources.

Key performance indicator: This KPI represents the utilization of time off and sick leave compared to actual work hours. It provides insights into employee absenteeism and can be used to monitor and manage workforce attendance.

$$Ratio\ of\ days\ off\ and\ sick = \frac{Time\ off\ Hours}{Actual\ Work\ Hours} \times 100 \qquad \textbf{(2.2)}$$

2.5.3. KPI 3: New hire employees management

Main requirement: Calculate the number of new employees hired in the organization in a specific period of time.

Objective: Track and monitor the number of new hires to understand recruitment rates and organizational growth.

Input:

- Departmental employee history data: includes information about the number of employees in each department over time.
- Year: The year for which the KPI is being calculated.

Output:

- The number of new hires each year by department.

Business rule: Data about newly hired employees is consistent and accurate across all information sources. Only employees who have completed the onboarding process and officially started their role will be counted.

Key performance indicator: This index represents the recruitment performance and growth rate of the organization. Tracking the number of new hires can help evaluate the success of recruitment strategies and plan for future staffing needs. To calculate the new hired employees, the following formula is applied:

$$New\ Hire\ Employees = Count\ of\ Employees\ with\ Start\ Date\ per\ year \quad (2.3)$$

2.6. Business Intelligence Solution

Proposing BI solution for the project

Based on the findings from the assessment of Adventure Works Cycles company's HR department, it has been observed that the overall structure and relationships within the department are well managed. The department maintains important information such as employee records, pay history, and potential candidates. However, there are identified shortcomings in specific areas such as management, recruitment, training, and salary. These deficiencies can be attributed, in part, to inadequate information and information management.

In light of this, a proposed solution leveraging Business Intelligence (BI) and BI tools can address these challenges. The solution can be divided into four parts:

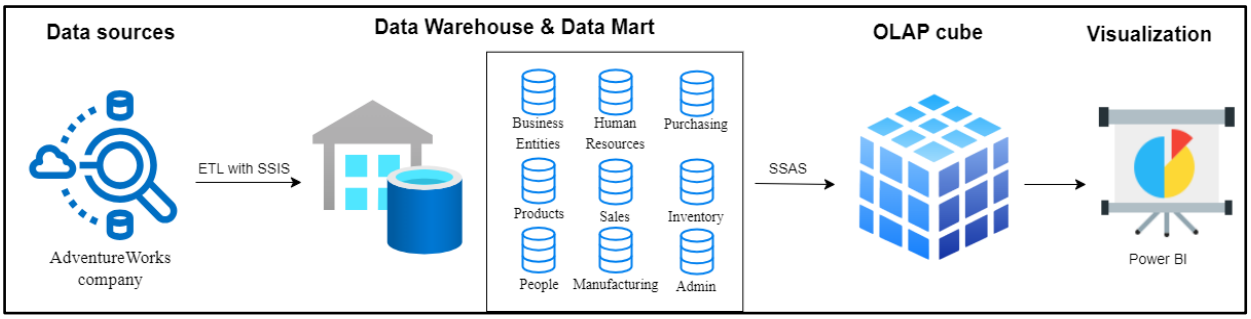


Figure 2.1. BI solution proposal

- ***Metrics and insights:*** Use BI tools such as Tableau, Power BI, or QlikView to create visually appealing charts, graphs, and reports. The dashboard should provide a

coData Sources: Identify the relevant data sources within the HR department, performance management software, payroll system, and employee surveys.

- ***Data Warehouse & Data Mart:*** Establish a data warehouse to centralize and store the HR data from various sources. Design a data mart specifically for HR data, focusing on key areas such as employee information, recruitment, training, and compensation. This data warehousing approach enables efficient data retrieval and analysis.
- ***OLAP Cube:*** Build an OLAP (Online Analytical Processing) cube to facilitate multidimensional analysis of HR data. The cube should include dimensions such as employee attributes, time, department, and performance indicators. This allows HR professionals to slice and dice the data to gain insights from different perspectives.
- ***Visualization Dashboard:*** Develop a user-friendly and interactive visualization dashboard that presents HRmprehensive view of the HR department's performance, recruitment status, training effectiveness, and compensation trends.

CHAPTER 3: DATA PREPARATION AND DATA MODELING

This chapter 3 focuses on Data preparation (Data collection, Data description, Data understanding, EDA, Pre-processing). Overall description of HR Module's AdventureWorks data. Build bus matrix, design Data model for Data warehouse/Data marts.

3.1. Data warehouse

The AdventureWorks database is an extensive source of data that provides detailed information on the key transactions required by a bicycle manufacturing business. This comprehensive database includes data on various business aspects such as Business Entities, Human Resources, Manufacturing, Sales, and Purchasing.

AdventureWorks database includes 9 modules and 71 tables, the specific meanings of the modules are as follows:

Table 3.1. Meanings of the modules

Modules	Description
Business Entities	Vendors, customers, and employees have common tables for addresses and contacts.
People	Names and addresses of individual customers, suppliers and employees.
Human Resources	Employees of Adventure Works Cycles.

Products	Products manufactured and sold by Adventure Works Cycles.
Manufacturing	Manufacturing process management module.
Sales	Customers and sales related data.
Purchasing	Vendors from whom parts and products are purchased.
Inventory	Inventory management.
Admin	Contains information about versions, errors, and changes of the AdventureWorks dataset.

3.2. Data mart of HR module

3.2.1. Overview data source HR

In our project, we focused on HR processes to respond to requests and questions to achieve the project goal of improving decision making. The HR process of Adventure Works Cycles comprises six different tables and is described as follows:

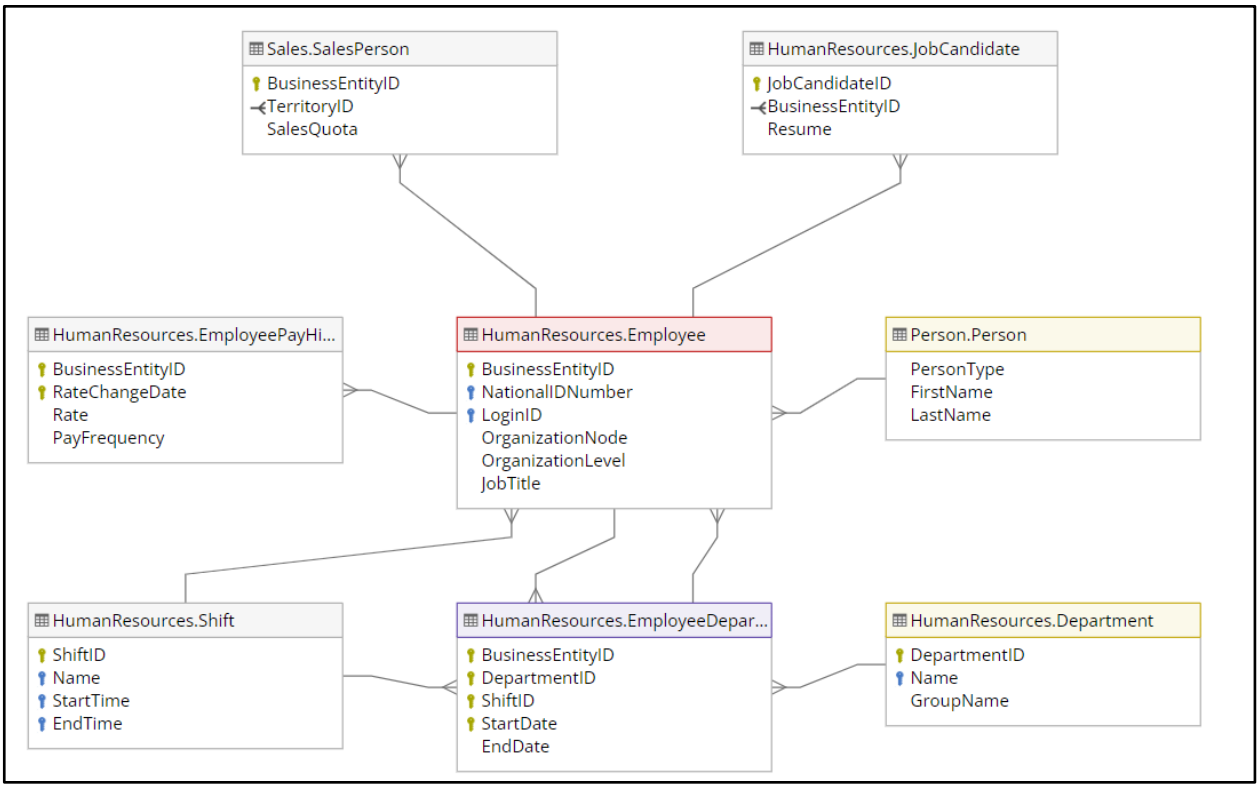


Figure 3.1. Schema data source

Table 3.2. Description of database human resource

No.	Table	Descriptive
1	HumanResources.JobCandidate	This table contains information about the resumes of job applicants submitted to the human resources department.
2	HumanResources.Department	Lookup table containing the departments within the Adventure Works Cycles company.

3	HumanResources.EmployeeDepartmentHistory	Employee department transfers.
4	HumanResources.Shift	Work shift lookup table.
5	HumanResources.EmployeePayHistory	Employee pay history.
6	HumanResources.Employee	Employee information such as salary, department, and title.

3.2.2. Details of database HR module

- ***HumanResources.Department:***

Table 3.3. HumanResources.Department

Key	Name	Data Type	Null	Attributes	References
PK	DepartmentID	smallint		Identity	
Description: Primary key for Department records.					
UK	Name	nvarchar(50)			
Description: Name of the department.					
	GroupName	nvarchar(50)			
Description: Name of the group to which the department belongs.					
	ModifiedDate	datetime		Default: getdate()	
Description: Date and time the record was last updated.					

- ***HumanResources.Employee:***

Table 3.4. HumanResources.Employee

Key	Name	Data Type	Null	Attributes	References

PK	BusinessEntityID	int			Person.Person
Description: Primary key for Employee records. Foreign key to BusinessEntity.BusinessEntityID.					
UK	NationalIDNumber	nvarchar(15)			
Description: Unique national identification number such as a social security number.					
UK	LoginID	nvarchar(256)			
Description: Network login.					
	OrganizationNode	hierarchyid	X		
Description: Where the employee is located in corporate hierarchy.					
	OrganizationLevel	smallint	X	Computed: [OrganizationNode].[GetLevel]()	

Description: The depth of the employee in the corporate hierarchy.					
	JobTitle	nvarchar(50)			
Description: Work title such as Buyer or Sales Representative.					
PK (user-defined)	BirthDate	date			
Description: Date of birth.					
	MaritalStatus	nchar(1)			
Description: M = Married, S = Single					
	Gender	nchar(1)			
Description: M = Male, F = Female					
	HireDate	date			
Description: Employee hired on this date.					
	SalariedFlag	bit		Default: 1	

Description: Job classification. 0 = Hourly, not exempt from collective bargaining. 1 = Salaried, exempt from collective bargaining.					
	VacationHours	smallint		Default: 0	
Description: Number of available vacation hours.					
	SickLeaveHours	smallint		Default: 0	
Description: Number of available sick leave hours.					
	CurrentFlag	bit		Default: 1	
Description: 0 = Inactive, 1 = Active.					
UK	rowguid	uniqueidentifier		Default: newid()	
Description: ROWGUIDCOL number uniquely identifying the record. Used to support a merge replication sample.					
	ModifiedDate	datetime		Default: getdate()	

Description: Date and time the record was last updated.

- ***HumanResources.EmployeeDepartmentHistory:***

Table 3.5. HumanResources.EmployeeDepartmentHistory

Key	Name	Data Type	Null	Attributes	References
PK	BusinessEntityID	int			HumanResources.Employee
Description: Employee identification number. Foreign key to Employee.BusinessEntityID.					
PK	DepartmentID	smallint			HumanResources.Department
Description: Department in which the employee worked including currently. Foreign key to Department.DepartmentID.					
PK	ShiftID	tinyint			HumanResources.Shift
Description: Identifies which 8-hour shift the employee works. Foreign key to					

Shift.Shift.ID.					
PK	StartDate	date			
Description: Date the employee started work in the department.					
	EndDate	date	X		
Description: Date the employee left the department. NULL = Current department.					
	ModifiedDate	datetime		Default: getdate()	
Description: Date and time the record was last updated.					

- ***HumanResources.EmployeePayHistory:***

Table 3.6. HumanResources.EmployeePayHistory

Key	Name	Data Type	Null	Attributes	References
PK	BusinessEntityID	int			HumanResources.Employee
Description: Employee identification number. Foreign key to					

Employee.BusinessEntityID.					
PK	RateChangeDate	datetime			
Description: Date the change in pay is effective					
	Rate	money			
Description: Salary hourly rate.					
	PayFrequency	tinyint			
Description: 1 = Salary received monthly, 2 = Salary received biweekly.					
	ModifiedDate	datetime		Default: getdate()	
Description: Date and time the record was last updated.					

- ***HumanResources.JobCandidate:***

Table 3.7. HumanResources.JobCandidate

Key	Name	Data Type	Null	Attributes	References

PK	JobCandidateID	int		Identity	
Description: Primary key for JobCandidate records.					
	BusinessEntityID	int			HumanResources.Employee
Description: Employee identification number. Foreign key to Employee.BusinessEntityID.					
	Resume	xml			
Description: Résumé in XML format.					
	ModifiedDate	datetime		Default: getdate()	
Description: Date and time the record was last updated.					

- ***HumanResources.Shift:***

Table 3.8. HumanResources.Shift

Key	Name	Data Type	Null	Attributes	References

PK	ShiftID	tinyint		Identity	
Description: Primary key for Shift records.					
UK	Name	nvarchar(50)			
Description: Shift description.					
UK	StartTime	time(7)			
Description: Shift start time.					
UK	EndTime	time(7)			
Description: Shift end time.					
	ModifiedDate	datetime		Default: getdate()	
Description: Date and time the record was last updated.					

3.2.3. Challenges of data HR

AdventureWorks is a sample database provided by Microsoft that represents a fictitious bicycle manufacturer. Some of the challenges associated with data sources for the human resources module of AdventureWorks include:

- Poor data quality and not enough information needed.
- Data retrieval is also difficult due to many data duplications and omissions in many tables, impeding efficient analysis.
- Many time attributes lack clarity, leading to confusion among users and potentially skewed results.
- Lack of information about employees' reasons for leaving. This absence inhibits the formulation of strategies to retain talent and establish effective retention policies.

3.3. EDA database

3.3.1. Descriptive statistics

3.3.1.1. HumanResources.Department

Table 3.9. EDA results of the "HumanResources.Department" dataset

Column	Missing Values	Duplicate Rows
DepartmentID	0	0
Name	0	0
GroupName	0	0

Total Rows in table	16
Total Missing Values	0
Total Duplicate Rows	0

Table 3.10. Quartile table of "HumanResources Department" dataset

Quartile table of "HumanResources.Department" dataset	
	DepartmentID
count	16.000000
mean	8.500000
std	4.760952
min	1.000000
25%	4.750000
50%	8.500000
75%	12.250000
max	16.000000

Based on the EDA (Exploratory Data Analysis) results of the "HumanResources Department" dataset, we can make the following observations:

- Missing Values: There are no missing or empty values in the 'DepartmentID', 'Name', and 'GroupName' columns. This indicates that the data has been collected completely and does not require any handling of missing data.
- Duplicate Rows: There are no duplicate rows in the dataset.
- Distribution of DepartmentID: The dataset contains a total of 16 rows, and the values of 'DepartmentID' range from 1 to 16. The mean value of 'DepartmentID' is 8.5, with a standard deviation of approximately 4.76. This suggests that the distribution of 'DepartmentID' is quite scattered but still close to the mean.
- Descriptive Statistics of DepartmentID: The descriptive statistics provide an overview of the distribution of 'DepartmentID' in the dataset. The mean, median, and percentiles offer detailed information about the distribution of 'DepartmentID'.

Based on these observations, we can conclude that the "HumanResources Department" dataset is complete and does not present any significant issues regarding missing or duplicate data.

3.3.1.2. HumanResources.Employee

Table 3.11. EDA results of the "HumanResources.Employee" dataset

Column	Missing Values	Duplicate Rows
BusinessEntityID	0	0
NationalIDNumber	0	0

LoginID	0	0
JobTitle	0	0
BirthDate	0	0
MaritalStatus	0	0
Gender	0	0
HireDate	0	0
OrganizationNode	1	0
OrganizationLevel	1	0
VacationHours	0	0
SickLeaveHours	0	0
Total Rows in table		290
Total Missing Values		2

Total Duplicate Rows	0

Table 3.12. Quartile table of "HumanResources.Employee" dataset

Quartile table of "HumanResources.Employee" dataset			
	OrganizationLevel	**VacationHours**	**SickLeaveHours**
count	289.000000	290.000000	290.000000
mean	3.522491	50.613793	45.306897
std	0.750240	28.786215	14.540444
min	1.000000	0.000000	20.000000
25%	3.000000	26.250000	33.000000
50%	4.000000	51.000000	46.000000
75%	4.000000	75.000000	58.000000
max	4.000000	99.000000	80.000000

Based on the EDA (Exploratory Data Analysis) results of the "HumanResources Employee" dataset, we can make the following observations:

- Missing Values: The dataset has missing values in the ‘OrganizationNode’ and ‘OrganizationLevel’ columns, with 1 missing value each. This suggests potential data quality issues in these columns that may need further investigation or imputation.
- Duplicate Rows: There are no duplicate rows in the dataset, indicating that each record is unique.
- Quartile Table Analysis:
    + OrganizationLevel: The ‘OrganizationLevel’ column has a mean of approximately 3.52, with a standard deviation of around 0.75. The minimum value is 1, and the maximum value is 4. This suggests that the majority of employees belong to an organization level between 3 and 4.
    + VacationHours: Employees have a mean of approximately 50.61 vacation hours, with a standard deviation of about 28.79. The minimum and maximum values are 0 and 99, respectively. The quartile table indicates that 75% of employees have 75 or fewer vacation hours.
    + SickLeaveHours: The mean sick leave hours are approximately 45.31, with a standard deviation of around 14.54. The minimum and maximum values are 20 and 80, respectively. The quartile table suggests that 75% of employees have 58 or fewer sick leave hours.

Overall, the dataset seems relatively clean with minimal missing values and no duplicate records. However, further investigation into the missing values in the ‘OrganizationNode’ and ‘OrganizationLevel’ columns may be necessary to understand their impact on the analysis.

3.3.1.3. HumanResources.EmployeeDepartmentHistory

Table 3.13. EDA results of the "HumanResources.EmployeeDepartmentHistory" dataset

Column	Missing Values	Duplicate Rows
BusinessEntityID	0	0
DepartmentID	0	0
ShiftID	0	0
StartDate	0	0
EndDate	290	0
Total Rows in table		296
Total Missing Values		290
Total Duplicate Rows		0

Table 3.14. Quartile table of "HumanResources.EmployeeDepartmentHistory"

Quartile table of "HumanResources.EmployeeDepartmentHistory"				
	DepartmentID	**ShiftID**	**StartDate**	**EndDate**

count	296.000000	296.000000	296	6
mean	7.266892	1.560811	2009-06-04 16:17:50.270270208	2011-08-06 20:00:00
std	2.804304	0.774390	NaN	NaN
min	1.000000	1.000000	2006-06-30 00:00:00	2009-07-14 00:00:00
25%	7.000000	1.000000	2008-12-27 00:00:00	2010-09-13 12:00:00
50%	7.000000	1.000000	2009-02-03 00:00:00	2011-08-15 00:00:00
75%	7.000000	2.000000	2009-12-11 00:00:00	2012-04-25 12:00:00
max	16.000000	3.000000	2013-11-14 00:00:00	2013-11-13 00:00:00

Based on the EDA (Exploratory Data Analysis) results of the "HumanResources.EmployeeDepartmentHistory" dataset, we can make the following observations:

- The “HumanResources.EmployeeDepartmentHistory” dataset records the historical departmental assignments of employees. It comprises 296 rows and includes columns such as BusinessEntityID, DepartmentID, ShiftID, StartDate, and EndDate
- Missing Values: The ‘EndDate’ column has 290 missing values, indicating that most department assignments do not have an end date recorded. No other columns exhibit missing values.
- Duplicate Rows: No duplicate rows are present in the dataset.
- Quartile Table Analysis:
    + DepartmentID: The DepartmentID ranges from 1 to 16. The mean DepartmentID is approximately 7.27, with a standard deviation of 2.80.
    + ShiftID: ShiftID ranges from 1 to 3. The mean ShiftID is about 1.56, with a standard deviation of 0.77.
    + StartDate: The StartDate ranges from June 30, 2006, to November 14, 2013. No quartile analysis provided for this column.
    + EndDate: Most department assignments lack recorded end dates, requiring further investigation into the data recording process.

The dataset indicates that the majority of department assignments lack recorded end dates, which may require further investigation. No duplicate rows were found, ensuring data integrity. Further analysis may be necessary to explore the distribution of StartDate and its potential impact on departmental trends over time.

3.3.1.4. HumanResources.JobCandidate

Table 3.15. EDA results of the “HumanResources.JobCandidate” dataset

Column	Missing Values	Duplicate Rows

JobCandidateID	0	0
BusinessEntityID	11	0
Total Rows in table		13
Total Missing Values		11
Total Duplicate Rows		0

Table 3.16. Quartile table of "HumanResources.JobCandidate" dataset

Quartile table of "HumanResources.JobCandidate" dataset		
	JobCandidateID	**BusinessEntityID**
count	13.00000	2.00000
mean	7.00000	243.00000
std	3.89444	43.84062
min	1.000000	212.00000
25%	4.00000	227.50000

50%	7.00000	243.00000
75%	10.00000	258.50000
max	13.00000	274.00000

Based on the EDA (Exploratory Data Analysis) results of the "HumanResources JobCandidate" dataset, we can make the following observations:

- The "HumanResources.JobCandidate" dataset contains information about job candidates. It consists of 13 rows and includes columns such as JobCandidateID and BusinessEntityID.
- Missing Values: The 'BusinessEntityID' column has 11 missing values out of 13 total rows. No other columns exhibit missing values.
- Duplicate Rows: No duplicate rows are present in the dataset.
- Quartile Table Analysis:
    + JobCandidateID: Ranges from 1 to 13. The mean JobCandidateID is 7, with a standard deviation of approximately 3.89.
    + BusinessEntityID: Ranges from 212 to 274. The mean BusinessEntityID is 243, with a standard deviation of approximately 43.84.

The dataset shows a significant number of missing values in the BusinessEntityID column, which may require further investigation or imputation. No duplicate rows were found, indicating unique candidate records. Further analysis could explore relationships between JobCandidateID, BusinessEntityID, and other relevant attributes to gain insights into candidate profiles and recruitment trends.

3.3.1.5. HumanResources.EmployeePayHistoy

Table 3.17. EDA results of the "HumanResources.EmployeePayHistoy" dataset

Column	Missing Values	Duplicate Rows
BusinessEntityID	0	0
RateChangeDate	0	0
Rate	0	0
PayFrequency	0	0
Total Rows in table		316
Total Missing Values		0
Total Duplicate Rows		0

Table 3.18. Quartile table of "HumanResources.EmployeePayHistoy" dataset

Quartile table of "HumanResources.EmployeePayHistoy" dataset		
	Rate	**PayFrequency**
count	316.000000	316.000000

mean	17.758804	1.430380
std	12.276485	0.495915
min	6.500000	1.000000
25%	11.000000	1.000000
50%	14.000000	1.000000
75%	23.076900	2.000000
max	125.500000	2.000000

Based on the EDA (Exploratory Data Analysis) results of the "HumanResources.EmployeePayHistoy" dataset, we can make the following observations:

- The "HumanResources.EmployeePayHistory" dataset contains information about employee pay history. It comprises 316 rows and includes columns such as BusinessEntityID, RateChangeDate, Rate, and PayFrequency.
- Missing Values: All columns in the dataset have no missing values, indicating complete data for each attribute.
- Duplicate Rows: No duplicate rows are present in the dataset, ensuring uniqueness of employee pay history records.
- Quartile Table Analysis:

+ Rate: The Rate column represents employee pay rates, ranging from 6.5 to 125.5. The mean pay rate is approximately 17.76, with a standard deviation of about 12.28.

+ PayFrequency: PayFrequency indicates the frequency of pay, with values of 1 or 2. The mean pay frequency is approximately 1.43, indicating most employees are paid on a monthly basis.

The dataset appears to be complete, with no missing values or duplicate records. The pay rates vary widely, with a mean rate of around 17.76 units. Most employees are paid on a monthly basis, as indicated by the mean pay frequency of approximately 1.43. Further analysis could explore trends in pay rates over time, differences in pay frequency across departments, or relationships between pay rates and employee performance metrics.

3.3.1.6. HumanResources.Shift

Table 3.19. EDA results of the "HumanResources.Shift" dataset

Column	Missing Values	Duplicate Rows
ShiftID	0	0
Name	0	0
StartTime	0	0
EndTime	0	0
Total Rows in table		3

Total Missing Values	0
Total Duplicate Rows	0

Table 3.20. Quartile table of "HumanResources.Shift" dataset

Table. Quartile table of "HumanResources.Shift" dataset		
	StartTime	**EndTime**
count	3	3
unique	3	3
top	07:00:00	15:00:00
freq	1	1

Based on the EDA (Exploratory Data Analysis) results of the "HumanResources.Shift" dataset, we can make the following observations:

- The "HumanResources.Shift" dataset contains information about shifts within the organization. It consists of 3 rows and includes columns such as ShiftID, Name, StartTime, and EndTime.
- Missing Values: All columns in the dataset have no missing values, indicating complete data for each attribute.

- Duplicate Rows: No duplicate rows are present in the dataset, ensuring uniqueness of shift records.
- Quartile Table Analysis:
    + StartTime: There are 3 unique start times recorded in the dataset. The most frequent start time is 07:00:00.
    + EndTime: There are 3 unique end times recorded in the dataset. The most frequent end time is 15:00:00.

The dataset is small but complete, with no missing values or duplicate records. Shifts in the organization have distinct start and end times, with common patterns observed in the recorded times. Further analysis could explore the distribution of shifts across different departments or employee groups, as well as any patterns or trends in shift start and end times.

3.3.2. Employee Analysis

3.3.2.1. Overview

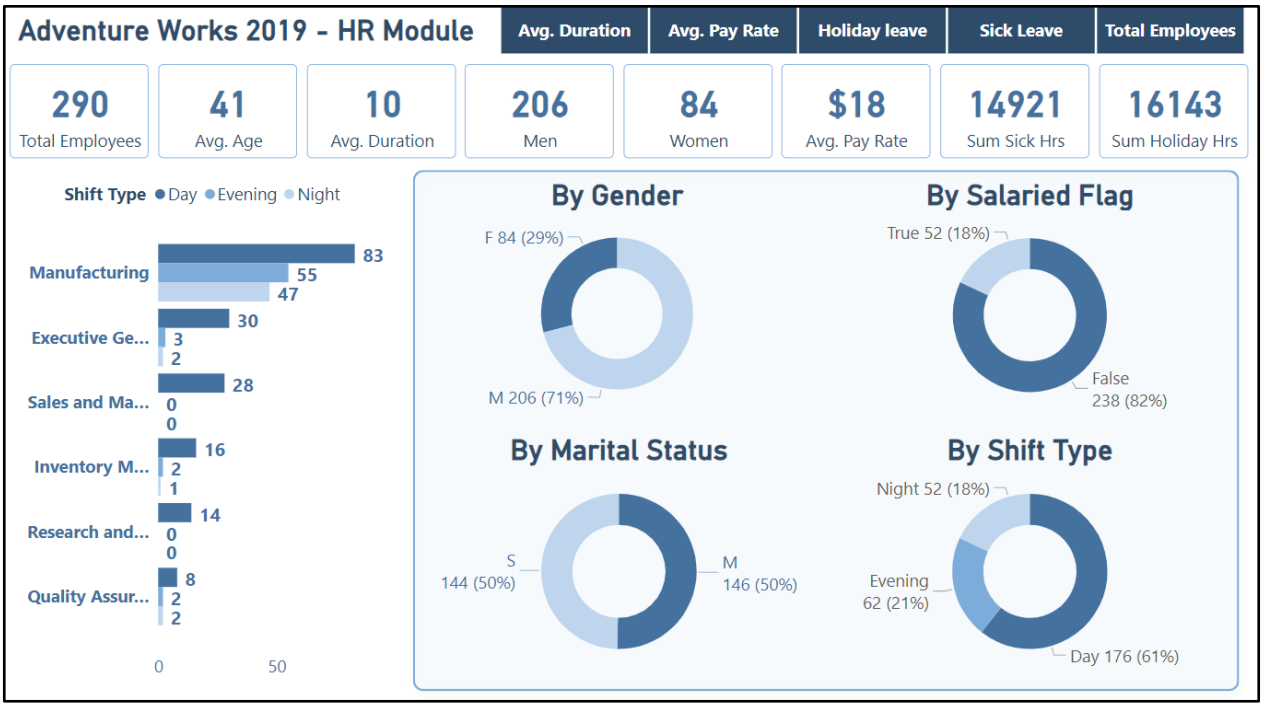


Figure 3.2. HR Overview

- Gender: In 2019, Adventure Works Cycles had a total of 290 employees, including 206 male employees (71%) and 84 female employees (29%).

In the bicycle manufacturing and sales industry like Adventure Works Cycles, achieving gender balance among employees is beneficial. Having equal representation of men and women brings diverse perspectives, fosters innovation, and enhances customer service. However, there is a significant gender gap that needs to be addressed. Companies should implement measures to increase the representation of women in the manufacturing industry, such as diverse recruitment programs and promoting women's participation. Developing policies that support equity for both male and female employees is also important.

- Marital Status: 50% of employees are married

Married employees often have family responsibilities and often need to consider the balance between work and family life. Therefore, companies can provide policies and

support such as flexible working, childcare, or financial support to help employees achieve this balance.

- Salaried Flag: 82% of employees are paid hourly.

Having a lot of employees paid hourly in a bicycle manufacturing and distribution business carries several benefits. Firstly, it allows for effective management of labor costs by paying employees based on actual hours worked. This flexibility is particularly useful in handling fluctuations in production demand and optimizing workforce size accordingly. Additionally, paying hourly wages accommodates non-regular tasks and enables the business to hire workers on-demand for specific assignments. It also provides flexibility in workforce management, allowing for efficient scheduling and task allocation. However, it is crucial to ensure compliance with labor regulations and uphold fair treatment and rights for all employees.

- Shift: Day shift type accounts for the majority with 61%, followed by Evening with 21% and finally Night with 18%

Only the Sales & Marketing department and Research & Development department have no employees working the evening and night shifts.

The company has 6 main department groups and is divided into 13 departments. Each department undertakes a different function.

Department	Group
Human Resources	Executive General and Administration
Finance	Executive General and Administration
Information Services	Executive General and Administration
Facilities and Maintenance	Executive General and Administration
Executive	Executive General and Administration
Shipping and Receiving	Inventory Management
Purchasing	Inventory Management
Production	Manufacturing
Production Control	Manufacturing
Document Control	Quality Assurance
Quality Assurance	Quality Assurance
Research and Development	Research and Development
Engineering	Research and Development
Tool Design	Research and Development
Sales	Sales and Marketing
Marketing	Sales and Marketing

Figure 3.3. HR Overview

3.3.2.2. Analysis

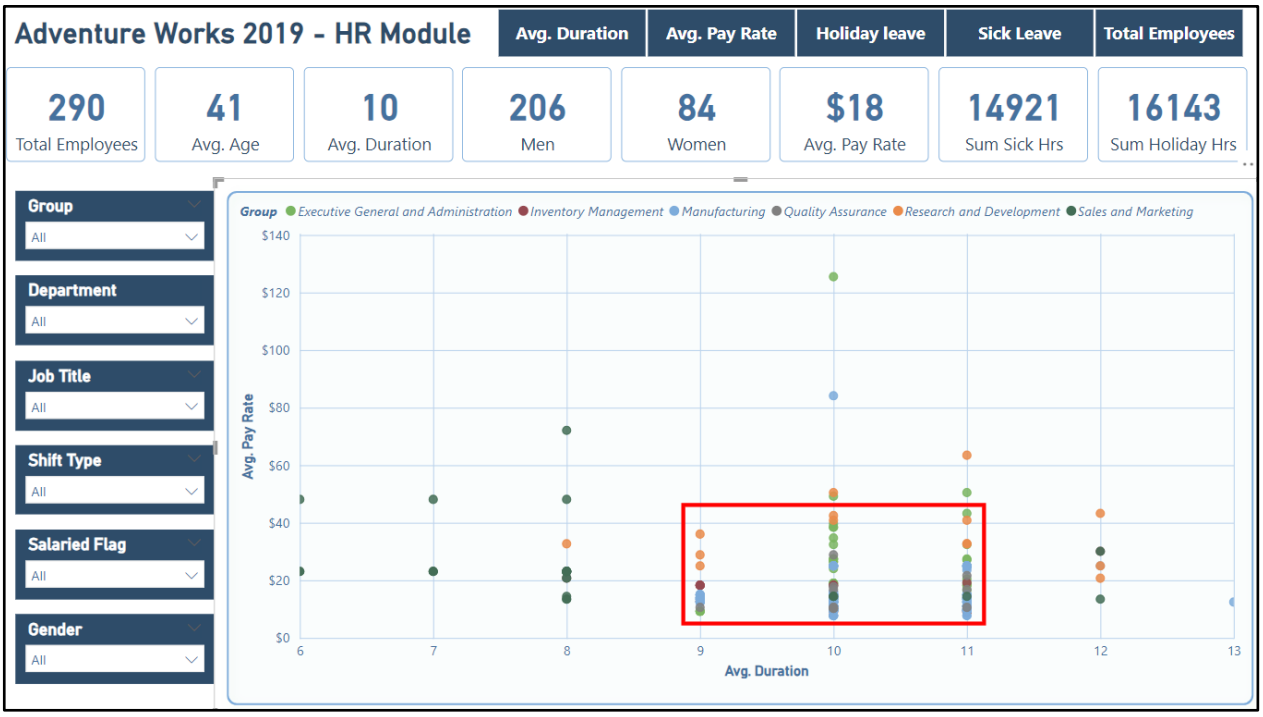


Figure 3.4. Duration Overview

In general, the pay rate is concentrated from 7$ to 40$ with the number of working years ranging from 9 to 11 years.

- Pay Rate:

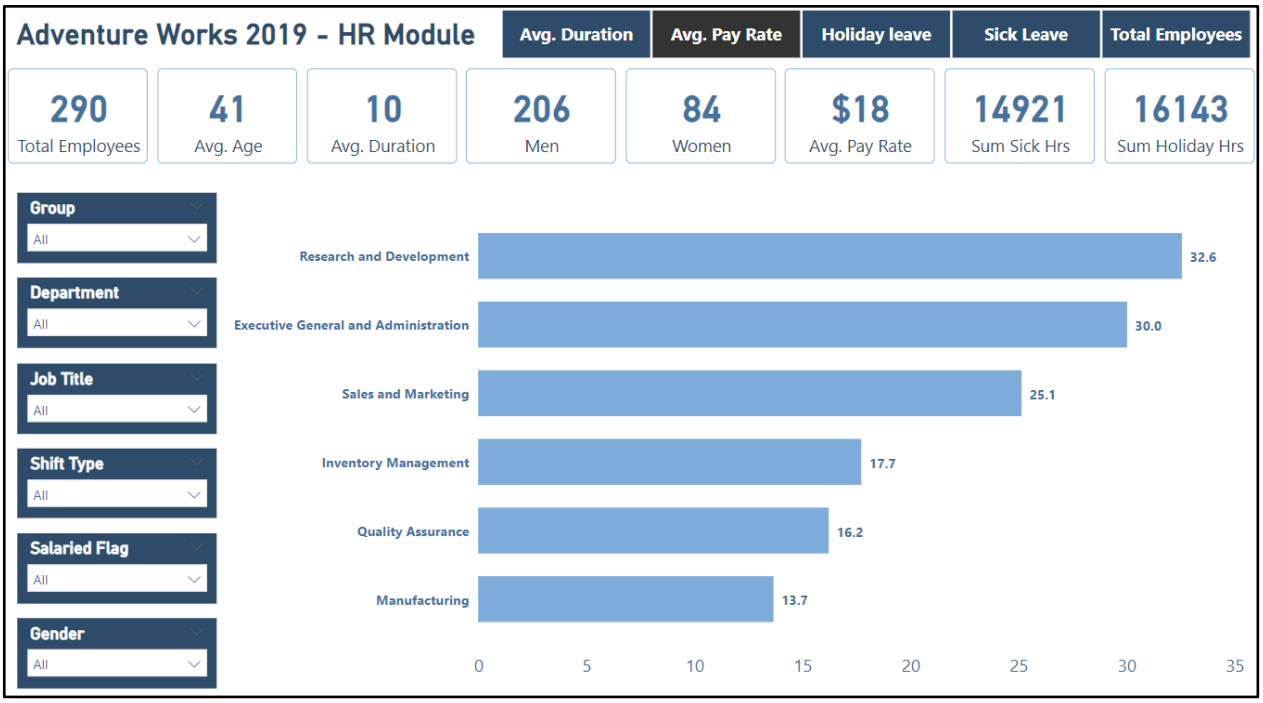


Figure 3.5. Pay rate Analysis

In particular, the pay rate in the Research and Development department has the highest average rate ($32.6), followed by Executive General and Administration ($30). This shows that the company is investing costs to research and develop processes and production. Meanwhile, Manufacturing has the lowest pay rate ($13.7), this may be because this is a manual labor group and needs a lot of manpower (185 employees)

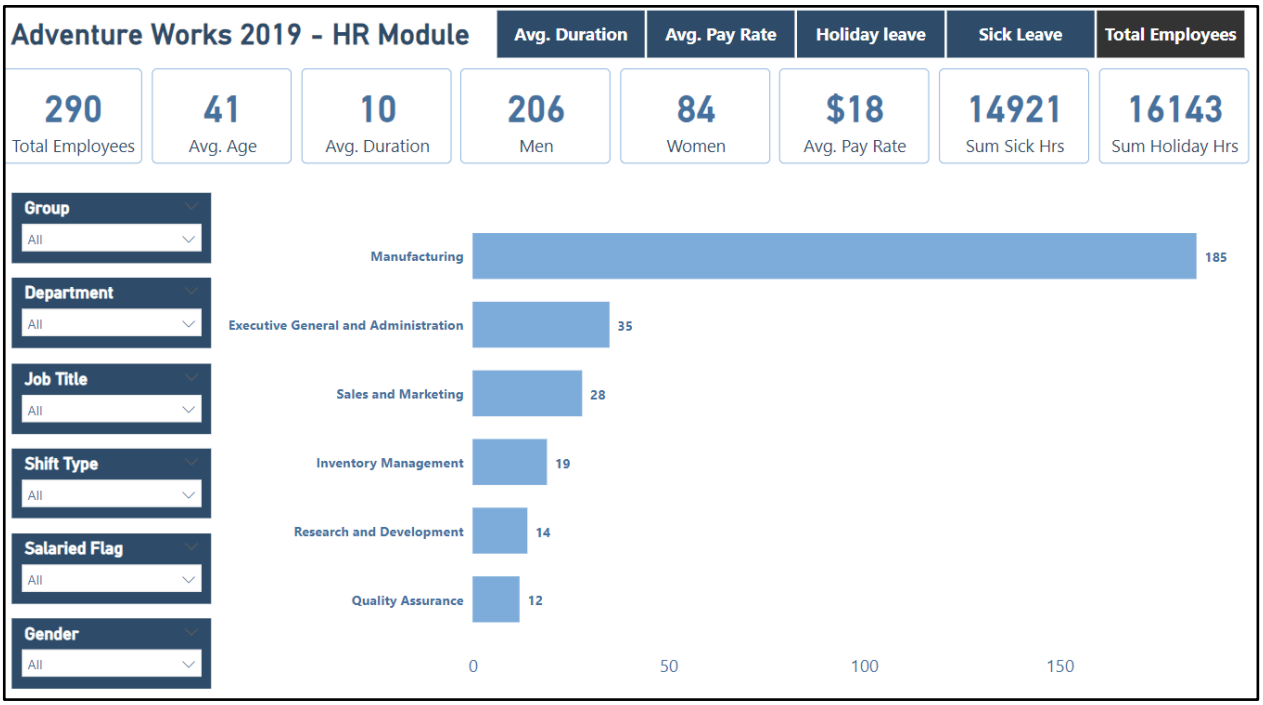


Figure 3.6. Pay rate Analysis

The manufacturing department has the highest employee count with a total of 185 employees, while the executive department has 35 employees. Given the nature of the business as a bicycle manufacturing company, human resources place significant emphasis on ensuring the effectiveness and efficient management of the manufacturing department.

- Duration:

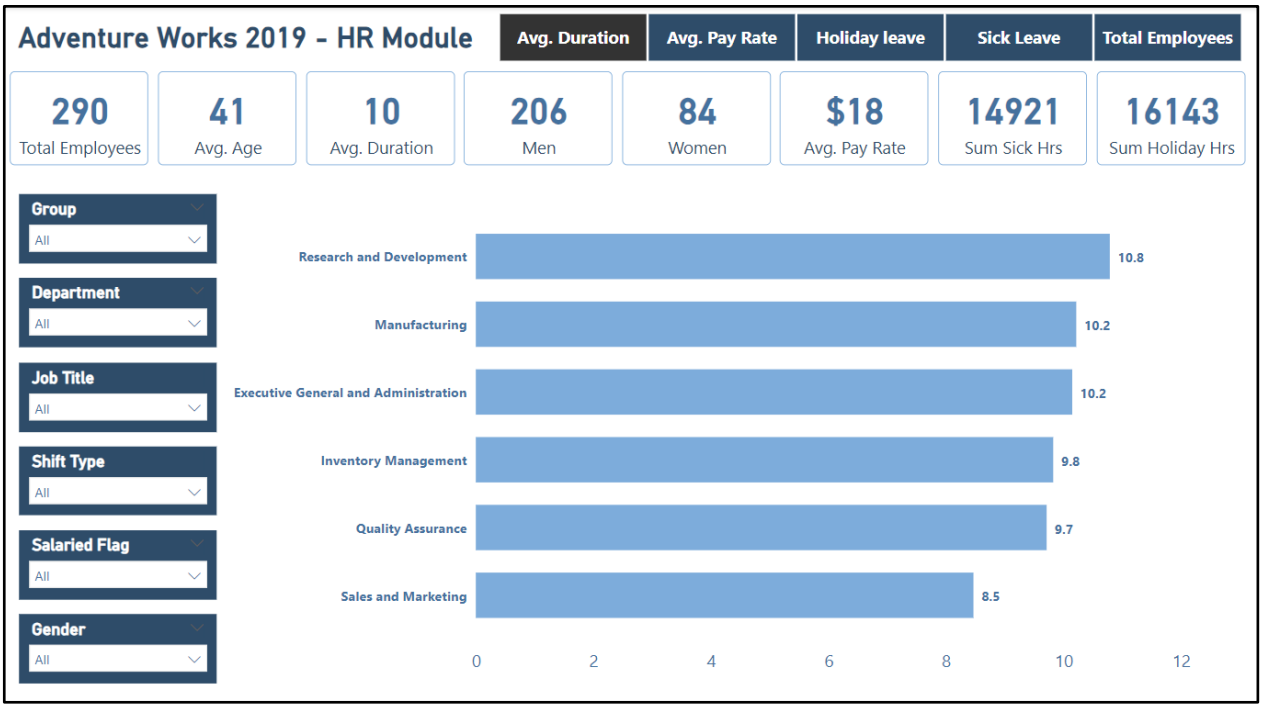


Figure 3.7. Duration of each department

On average, the seniority of company employees ranges from 8 years for the Sales and Marketing department to 11 years for the Research and Development department. Having experienced employees in a manufacturing company brings valuable benefits. They have accumulated knowledge and experience, leading to higher productivity and better problem-solving. Their understanding of company processes and customer relationships ensures stability and loyalty. Additionally, their expertise contributes to innovation and competitiveness.

However, it's important to balance experience with fresh perspectives for long-term success, especially in the Sales and Marketing department, where creativity is needed.

Figure 3.8. Duration of Sales and Marketing Department

Recruiting younger personnel in the Sales and Marketing department can enhance creativity and reduce costs. Their fresh perspectives and innovative ideas can drive marketing strategies, while their lower salary expectations contribute to cost savings. However, it's important to maintain a balanced workforce by leveraging the experience of senior employees.

3.4. Designing Data Warehouse

3.4.1. Bus Matrix

In the data analysis and data warehouse building environment, data structure arrangement is an important and complex part. In particular, when building a data warehouse for human resource management systems, the reasonable organization of data tables is a key factor to ensure consistency and efficiency in analysis and reporting.

A popular tool used to support this process is ***"Bus Matrix"***, an interactive framework developed by Ralph Kimball. Bus Matrix is not just a diagram but also a planning model,

describing the association between physical and dimension data tables in the data warehouse. For human resources projects, Bus Matrix is an important tool that helps organize data in a logical way and ensures consistency and ease of access for analysis and reporting.

Table 3.21. Bus matrix of HR

Business process/ Business Requirements	Common Dimension						
	Date	**Department**	**Pay history**	**Manager**	**Rate**	**Shift**	**Time off hours**
Recruitment	x	x			x	x	x
Training requisition	x	x		x		x	x
Work time management	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	x
Changing departments for employees	x	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	x	

Evaluation from the bus matrix table indicates that developing dimension tables for Department, Shift, Time and Employee is necessary. These dimension tables will contain detailed and important information for managing and analyzing key aspects of human resources in the organization.

3.4.2. Master Data

Master data is defined by a few central characteristics that distinguish it from transaction data in particular: It has a high level of statistics, which means that it rarely changes and is mostly valid in the long term. Since master data is used by several divisions of a company, it is highly relevant for all business processes. This is underlined by the fact that master data is often used as a criterion in statistics and data evaluations. In addition, transaction data is dependent on master data – no transaction data without master data. For these reasons, master data is generally kept long-term.

Table 3.22. Master Data of HR

Object	Description
Employee name	The information of the employee.

Department	The information of the department.
Hire date	Date the employee was hired.
Shift	The information of shift.
Job title	Name of the job in the company.

3.4.3. Transaction Data

Transaction data, on the other hand, is dynamic data that is used by specific departments. Its relevance is limited to a certain period of time. For the mentioned statistics it provides the facts behind the master data criteria. As suggested by the term, master data is key to the organization. Its core and structure is represented by master data. Transactional data though is renewed and replaced regularly.

Table 3.23. Transaction Data of HR

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

TimeOffHours	Number of break hours per employee.

3.5. Dimension tables

Dimension tables in a data warehouse are fundamental repositories of descriptive attributes that provide context to transactional data stored in fact tables. In the realm of human resources (HR), these dimension tables capture vital details about employees, departments, shifts, and time, facilitating analysis and decision-making processes. They play a crucial role in enhancing the understanding of transactional data and enable efficient querying and reporting within the organization. By establishing key relationships with fact tables, these dimension tables offer valuable insights into HR processes such as payroll, recruitment, and training. This structured approach ensures data consistency and supports various analyses across the organization. The following section will delve into a detailed description and analysis of the dimension tables within the data warehouse.

3.5.1. DimEmployee

Meaning: This dimension table contains detailed information about employees.

Create Table DimEmployee in SQL:

	Column Name	Data Type	Allow Nulls
🔑	EmployeeKey	int	☐
	BusinessEntityID	int	☐
	FirstName	nvarchar(50)	☐
	LastName	nvarchar(50)	☐
	PhoneNumber	nvarchar(25)	☐
	BirthDate	date	☐
	Gender	nvarchar(1)	☐
	MaritalStatus	nvarchar(1)	☐
	EmailAddress	nvarchar(50)	☑
	AddressLine1	nvarchar(60)	☐
	City	nvarchar(30)	☐
	StateProvince	nvarchar(50)	☐
	CountryRegion	nvarchar(50)	☐
	OrganizationLevel	smallint	☑
	JobTitle	nvarchar(50)	☐
	NationalIDNumber	nvarchar(15)	☐
	HireDate	date	☐
	VacationHours	smallint	☐
	SickLeaveHours	smallint	☐

Figure 3.9. Table DimEmploye

Table 3.24. Describe the properties of the DimEmployee table

Column Name	Describe
EmployeeKey	This is the primary key of the table, representing a unique identifier for each employee.
BusinessEntityID	The business entity ID of the employee.

FirstName	The first name of the employee.
LastName	The last name of the employee.
PhoneNumber	The phone number of the employee.
BirthDate	The birth date of the employee.
Gender	The gender of the employee.
MaritalStatus	The marital status of the employee.
EmailAddress	The email address of the employee.
AddressLine1	The first line of the employee’s address.
City	The city where the employee resides.
StateProvince	The state or province where the employee resides.
CountryRegion	The country where the employee resides.
OrganizationLevel	The organizational level of the employee within the company.

JobTitle	The job title of the employee.
NationalIDNumber	The national ID number of the employee.
HireDate	The date when the employee was hired.
VacationHours	The number of vacation hours available to the employee.
SickLeaveHours	The number of sick leave hours available to the employee.

Data mapping: Data mapping for DimEmployee involves integrating data from various sources within the AdventureWorks dataset, including employee records from the Human Resources module. The attributes in DimEmployee are populated based on information extracted from these sources to provide a comprehensive overview of employee details for analysis and reporting purposes.

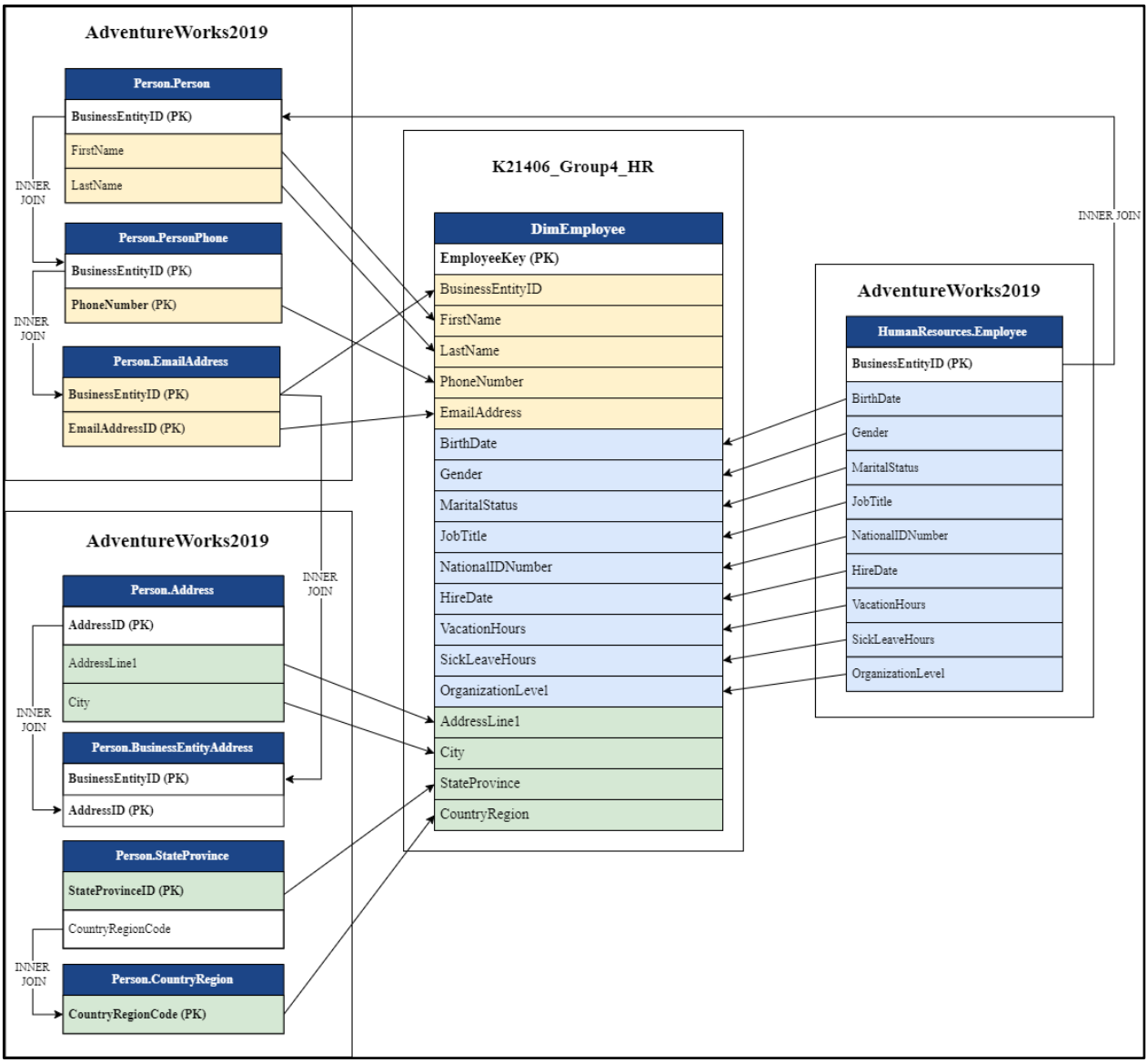


Figure 3.10. DimEmployee table data mapping diagram

DimEmployee query code in SQL:

Table 3.25. Code query SQL for table DimEmployee in database AdventureWorks2019

```
SELECT B.BusinessEntityID, A.FirstName, 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.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
```

3.5.2. DimDepartment

Meaning: This table provides a standardized representation of department-related attributes, facilitating analysis and reporting based on the organizational structure of departments.

Create Table DimDepartment in SQL:

	Column Name	Data Type	Allow Nulls
🔑	DepartmentKey	int	☐
	DepartmentID	smallint	☐
	Name	nvarchar(50)	☐
	GroupName	nvarchar(50)	☐

Figure 3.11. Table DimDepartment

Table 3.26. Describe the properties of the DimDepartment table

Column Name	Describe
DepartmentKey	This is the primary key of the table, representing a unique identifier for each department.
DepartmentID	An identifier for the department, often used internally for referencing.
Name	The name of the department.
GroupName	The name of the group or category to which the department belongs.

Data mapping: Data mapping for the DimDepartment table involves consolidating department-related attributes from various sources within the AdventureWorks dataset. The DepartmentKey uniquely identifies each department, while DepartmentID serves as an internal reference. Name captures the department's descriptive name, and GroupName specifies its category. Data is sourced from tables like HumanResources.Department, ensuring consistency for effective analysis and reporting.

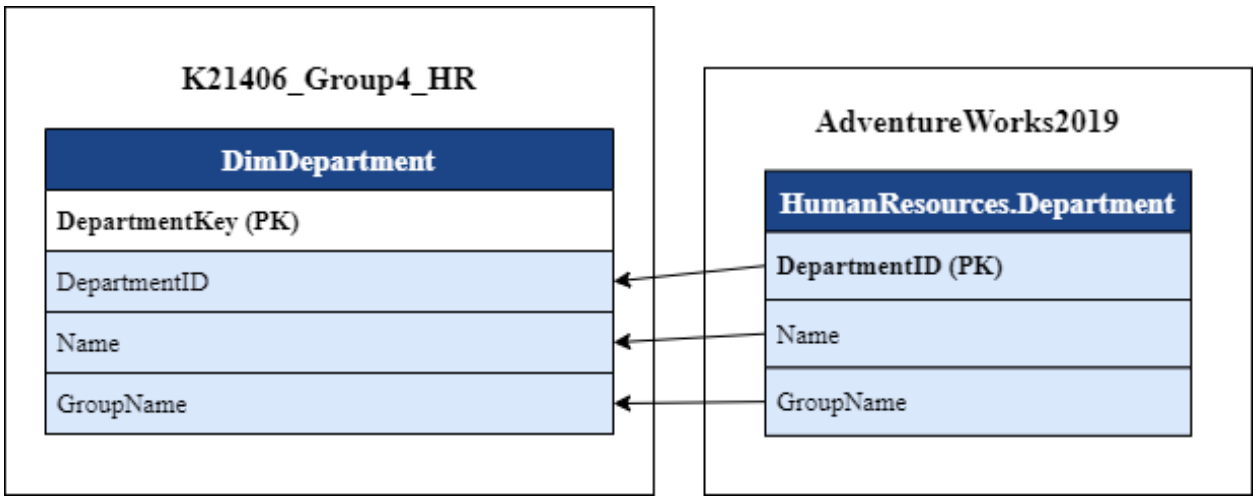


Figure 3.12. DimDepartment table data mapping diagram

DimDepartment query code in SQL:

Table 3.27. Code query SQL for table DimDepartment in database AdventureWorks2019

```
SELECT A.DepartmentID, A.Name, A.GroupName

From HumanResources.Department A
```

3.5.3. DimShift

Meaning: The "DimShift" table allows for the standardized representation of shift-related attributes, facilitating analysis and reporting on activities or events occurring during specific shifts.

Create Table DimShift in SQL:

	Column Name	Data Type	Allow Nulls
🔑	ShiftKey	int	☐
	ShiftID	tinyint	☐
	Name	nvarchar(50)	☐
	StartTime	time(7)	☐
	EndTime	time(7)	☐

Figure 3.13. Table Dim Shift

Table 3.28. Describe the properties of the DimShift table

Column Name	Describe
ShiftKey	This column serves as the primary key of the table, uniquely identifying each record representing a shift.
ShiftID	Represents the identifier of the shift.
Name	Describes or names the shift
StartTime	Represents the time when the shift starts.
EndTime	Indicates the time when the shift ends.

Data mapping: Data mapping for the DimShift table involves gathering shift-related attributes from relevant sources in the AdventureWorks dataset. The ShiftKey uniquely identifies each shift record, while ShiftID serves as its identifier. Name provides a descriptive label for the shift, while StartTime and EndTime specify the shift's start and end times, respectively. Data is sourced from tables HumanResources.Shift, ensuring consistency for analysis and reporting purposes.

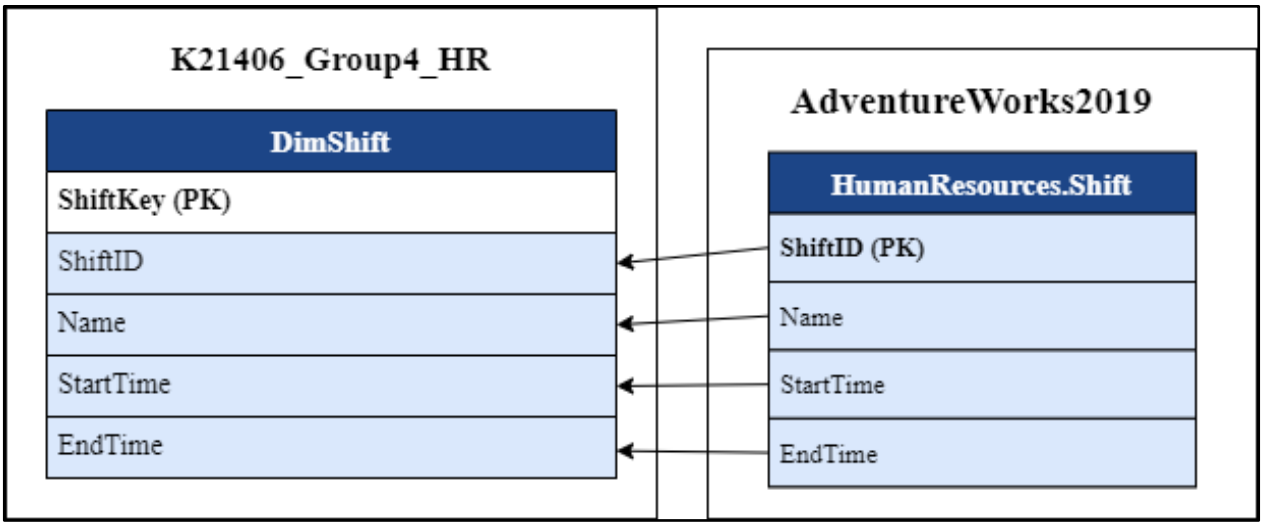


Figure 3.14. DimShift table data mapping diagram

DimShift query code in SQL:

Table 3.29. Code query SQL for table DimShift in database AdventureWorks2019

```
SELECT A.ShiftID, A.Name, A.StartTime, A.EndTime

From HumanResources.Shift A
```

3.5.4. DimTime

Meaning: The "DimTime" table provides a standardized and consistent representation of time-related attributes, enabling efficient time-based analysis across various dimensions and measures in a data warehouse or BI solution.

Create Table DimTime in SQL:

	Column Name	Data Type	Allow Nulls
🔑	TimeKey	int	☐
	Year	int	☑
	Month	int	☑
	Day	int	☑
	DateName	nvarchar(20)	☑
	DayNumberOfWeek	int	☑
	DayNumberOfMonth	int	☑
	DayNumberOfYear	int	☑
	WeekNumberOfYear	int	☑
	Quarter	int	☑

Figure 3.15. Table DimTime

Table 3.30. Describe the properties of the DimTime table

Column Name	Describe
TimeKey	This is the primary key representing a unique identifier for each time record. It appears to be an integer data type.
Year	Represents the year component of the date.
Month	Represents the month component of the date.
Day	Represents the day component of the date.
DateName	Provides the name of the day of the week corresponding to the date.Indicates the day of the week as a number

	(e.g., 1 for Sunday, 2 for Monday, etc.).
DayNumberOfMonth	Represents the day of the month.
DayNumberOfYear	Indicates the day of the year.
WeekNumberOfYear	Represents the week number within the year.
Quarter	Indicates the quarter of the year

Data mapping: Data mapping for the DimTime table involves extracting time-related attributes from various sources within the AdventureWorks dataset. The TimeKey column serves as the primary key, uniquely identifying each time record. Other attributes such as Year, Month, and Day capture temporal information at different granularities. These attributes are sourced from transactional data tables such as EmployeeDepartmentHistory and EmployeePayHistory, ensuring consistent and accurate representation of time dimensions for analytical purposes.

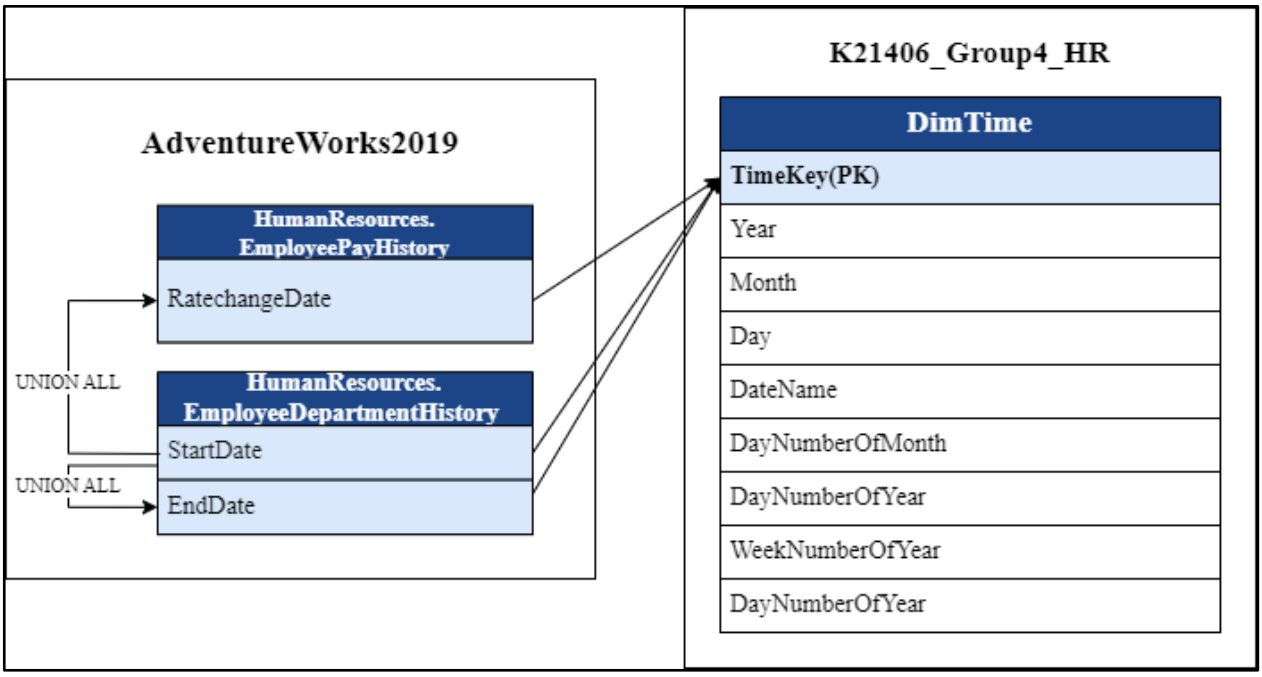


Figure 3.16. Dim table data mapping diagram

DimTime query code in SQL:

Table 3.31. Code query SQL for table DimTime in database AdventureWorks2019

```
SELECT DISTINCT
 CAST(YEAR(DateValue) AS nvarchar(4)) + RIGHT('0' +
CAST(MONTH(DateValue) AS nvarchar(2)), 2) + RIGHT('0' +
CAST(DAY(DateValue) AS nvarchar(2)), 2) AS TimeKey,
    YEAR(DateValue) AS Year,
    MONTH(DateValue) AS Month,
    DAY(DateValue) AS Day,
    DATENAME(WEEKDAY, DateValue) AS DateName,
    DATEPART(WEEKDAY, DateValue) AS DayNumberOfWeek,
```

```
    DATEPART(DAY, DateValue) AS DayNumberOfMonth,
    DATEPART(DAYOFYEAR, DateValue) AS DayNumberOfYear,
    DATEPART(WEEK, DateValue) AS WeekNumberOfYear,
    DATEPART(QUARTER, DateValue) AS Quarter
FROM (
  SELECT StartDate AS DateValue FROM
AdventureWorks2019.HumanResources.EmployeeDepartmentHistory WHERE
StartDate IS NOT NULL
  UNION ALL
  SELECT EndDate AS DateValue FROM
AdventureWorks2019.HumanResources.EmployeeDepartmentHistory WHERE
EndDate IS NOT NULL
  UNION ALL
  SELECT RateChangeDate AS DateValue FROM
AdventureWorks2019.HumanResources.EmployeePayHistory WHERE
RateChangeDate IS NOT NULL
) AS AllDates
```

3.6. Fact tables

In addition to dimension tables containing detailed information about factors such as Department, Shift, Time and Employee, human resource management projects often need to use fact tables to store data. business figures and events. These tables often contain data summarized or calculated from the detailed information in the dimension tables, making analysis and reporting easier and more efficient.

By using two event tables, FactPayHistory and FactDepartmentHistory, organizations can track and analyze important events related to salary payments and departmental operations in a detailed and effective manner.

3.6.1. FactPayHistory

Meaning: This is the salary calculation table for employees. This table will calculate employee salaries according to the following criteria: norms, sick leave hours, vacation hours,...

Create Table FactPayHistory in SQL:

	Column Name	Data Type	Allow Nulls
🔑	FactPayHistoryKey	int	☐
	EmployeeKey	int	☐
	DepartmentKey	int	☐
	ShiftKey	int	☐
	RateChangeDate	int	☑
	Rate	money	☐
	PayFrequency	tinyint	☐

Figure 3.17. Table FactPayHistory

Table 3.32. Describe the properties of the FactPayHistory table

Column Name	Describe
FactPayHistoryKey	Primary key for FactPayHistory records.
EmployeeKey	Primary key for Employee records.
DepartmentKey	Primary key for Department records.

ShiftKey	Identifies which 8-hour shift the employee works.
RateChangeDate	Date the change in pay is effective.
Rate	Salary hourly rate.
PayFrequency	1 = Salary received monthly, 2 = Salary received biweekly.

Data mapping: In the process of constructing and developing the FactPayHistory table, it is essential to integrate data from various sources to ensure accuracy and completeness of information. In this case, the FactPayHistory table is created by combining data from three main tables within the Human Resources (HR) module of the AdventureWorks dataset. These tables include Department, EmployeeDepartmentHistory, and EmployeePayHistory. Below is an overview diagram illustrating the data lineage for the FactPayHistory table:

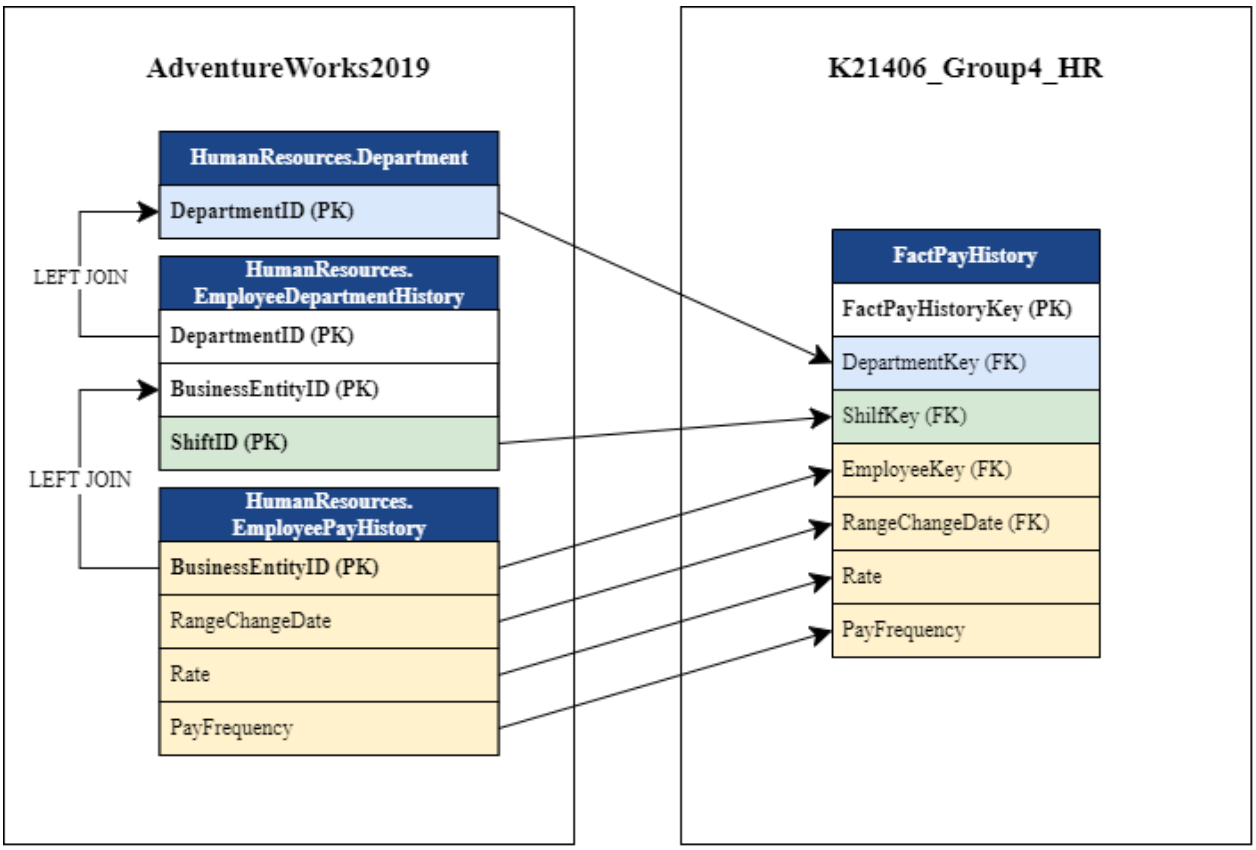


Figure 3.18. FactPayHistory table data mapping diagram

FactPayHistory query code in SQL:

Table 3.33. Code query SQL for table FactPayHistory in database AdventureWorks2019

```
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

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)
```

3.6.2. FactDepartmentHistory

Meaning: Store information and department exchange history of employees.

Create Table FactDepartmentHistory in SQL:

	Column Name	Data Type	Allow Nulls
🔑	FactDepartmentHistoryKey	int	☐
	EmployeeKey	int	☐
	DepartmentKey	int	☐
	ShiftKey	int	☐
	TimeOffHours	smallint	☐
	StartDate	int	☐
	EndDate	int	☑

Figure 3.19. Table FactDepartmentHistory

Table 3.34. Describe the properties of the FactDepartmentHistory table

Column Name	Describe
FactDepartmentHistoryKey	Primary key for FactDepartmentHistory records.
EmployeeKey	Primary key for Employee records.
DepartmentKey	Department in which the employee worked including currently
ShiftKey	Primary key for Shift records.
TimeOffHours	Number of break hours per employee.
StartDate	Date the employee started work in the department
EndDate	Date the employee finished work in the department

Data mapping: Data mapping is an important part of the process of building the FactDepartmentHistory table in the Human Resources module of the AdventureWorks dataset. This table integrates data from three main sources: Department Table, Department History Table, and Shift Table. The goal of the table is to ensure the accuracy and completeness of the information in the FactDepartmentHistory table, through mapping and combining data from these sources.

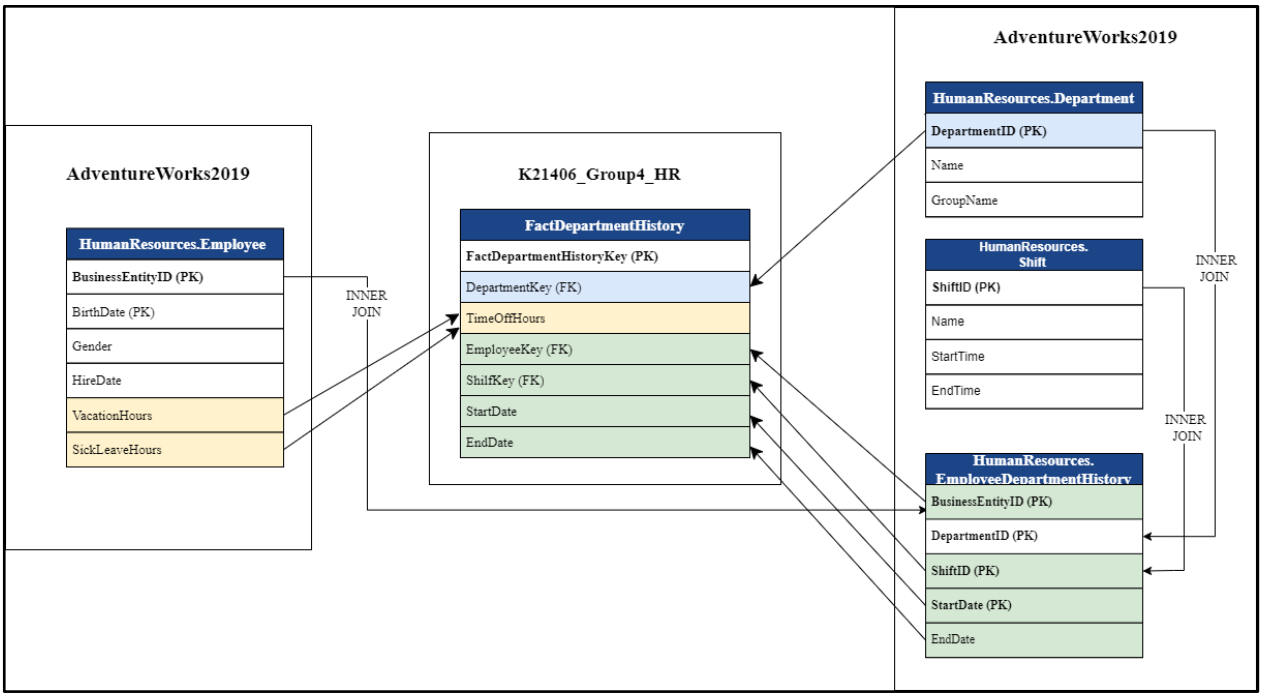


Figure 3.20. FactDepartmentHistory table data mapping diagram

FactDepartmentHistory query code in SQL:

Table 3.35. Code query SQL for table FactDepartmentHistory in database AdventureWorks2019

```
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
```

3.7. Data Warehouse model

The data warehouse modeling project in this case is employing a top-down approach, prioritizing the establishment of the overall structure and high-level concepts before delving into detailed sub-models. This approach began by identifying the business requirements and goals for the data warehouse. The development of data models and submodels followed to achieve a comprehensive structure for the data warehouse.

Data warehouse modeling is a crucial step in building a data warehouse as it involves designing schemas to capture detailed and summarized information. The primary goal is to create schemas that accurately represent the reality or specific aspects of the data required to support the warehouse.

There are two key reasons why data warehouse modeling is essential. Firstly, well-defined schemas enable clients or users of the data warehouse to easily understand the relationships among the data, enhancing their ability to utilize it effectively. Secondly, a carefully designed schema facilitates the creation of an efficient data warehouse structure, resulting in cost savings during implementation and improved overall efficiency in using the warehouse.

In this context, the Galaxy Schema, a widely adopted and intuitive modeling approach, is chosen to organize the data within the AdventureWorks dataset's HR module. This schema

eliminates redundancy and ensures high data quality and accuracy. It consists of interconnected fact tables surrounded by normalized dimension tables, creating a constellation-like structure.

The Galaxy Schema enhances the data warehouse's effectiveness by providing a multidimensional view and facilitating advanced reporting and analytics capabilities. Its strong design considerations for complex database systems ensure data integrity while enabling comprehensive analysis of HR-related data.

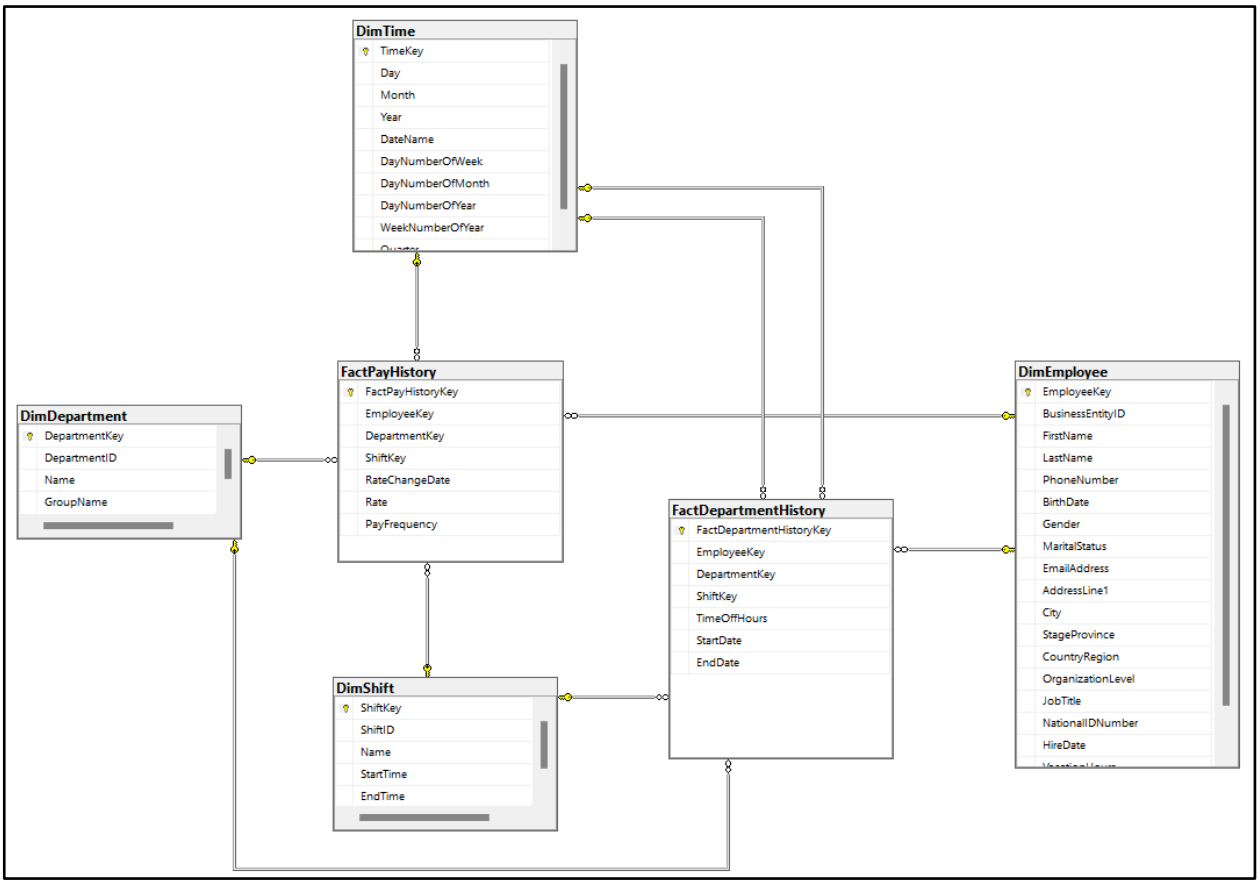


Figure 3.21. Data Warehouse model

Table 3.36. Describes the relationships between tables in the Data Warehouse model

No.	Relationship	Relationship type	Describe
R1	DimDepartment - FactPayHistory	1-n	- Each department in the "DimDepartment" table can have one or more payments in the "FactPayHistory" table. - Each payment in the "FactPayHistory" table belongs to only one department in the "DimDepartment" table.
R2	DimTime - FactPayHistory	1-n	- Each time point in the "DimTime" table can have one or more payments in the "FactPayHistory" table. - Each payment in the "FactPayHistory" table can only have one time point in the "DimTime" table.
R3	DimShift - FactPayHistory	1-n	- Each shift in the "DimShift" table can have one or more payments in the

			"FactPayHistory" table. - Each payment in the "FactPayHistory" table belongs to only one shift in the "DimShift" table.
R4	DimEmployee - FactPayHistory	1-n	- Each employee in the "DimEmployee" table can have one or more payments in the "FactPayHistory" table. - Each payment in the "FactPayHistory" table belongs to only one employee in the "DimEmployee" table.
R5	DimDepartment - FactDepartmentHistory	1-n	- Each department in the "DimDepartment" table can have one or more histories stored in the "FactDepartmentHistory" table. - Each history stored in the "FactDepartmentHistory" table belongs to only one department in the "DimDepartment" table.

R6	DimTime - FactDepartmentHistory	1-n	- Each time point in the “DimTime” table can have one or more histories stored in the “FactDepartmentHistory” table. - Each history stored in the “FactDepartmentHistory” table belongs to only one time point in the “DimTime” table.
R7	DimShift - FactDepartmentHistory	1-n	- Each shift in the “DimShift” table can have one or more histories stored in the “FactDepartmentHistory” table. - Each history stored in the “FactDepartmentHistory” table belongs to only one shift in the “DimShift” table.
R8	DimEmployee - FactDepartmentHistory	1-n	- Each employee in the “DimEmployee” table can have one or more histories stored in the “FactDepartmentHistory” table. - Each history stored in the “FactDepartmentHistory” table

			belongs to only one employee in the “DimEmployee” table.

CHAPTER 4: DATA INTEGRATION

This chapter focuses on building strategies (ETL, ELT,...)/diagrams/models for front-end and intermediate integration, integration processes for both dimension tables and facts tables, installing integration processes on SSIS tools (practice) and record/present data integration results for the HR module).

4.1. ETL pipeline for data mart Human resources

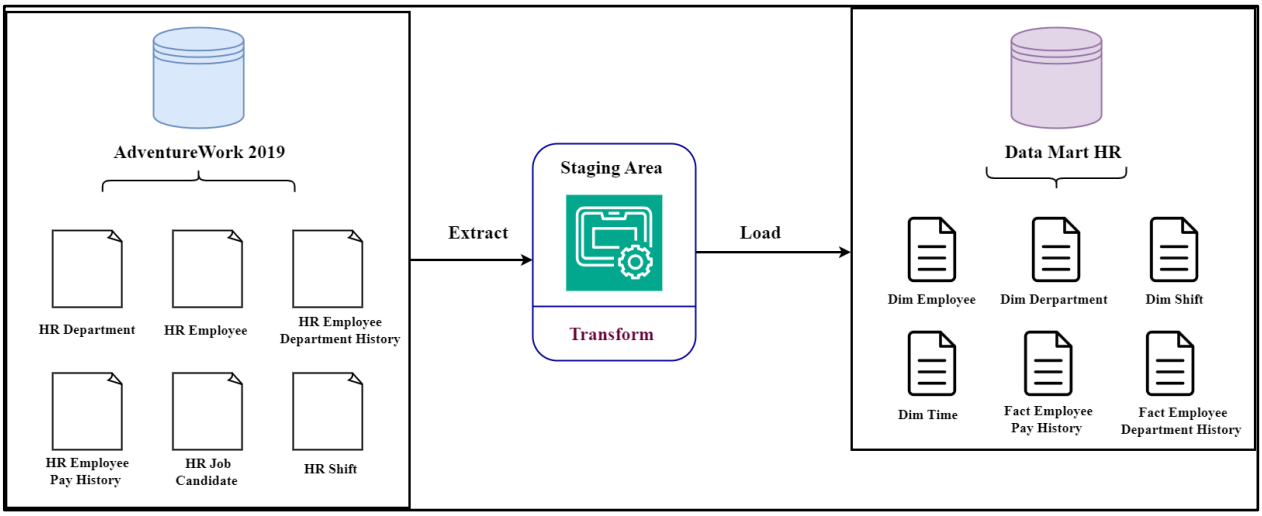


Figure 4.1. ETL pipeline for data mart Human resources

ETL pipeline (Extract, Transform, Load) is an important process in the field of data and analytics. The ETL pipeline plays a key role in transforming data from its origin to data warehouses, such as data marts, to provide critical information for business decisions.

- ***Extract:*** This process includes extracting data from the AdventureWorks 2019 database. Main data tables include: Employee, Department, Job Title,...
- ***Transform:*** After the data is extracted, it needs to be transformed to fit the structure of the Human Resources data mart. Transformation steps may include:
    + Handling missing or inaccurate data.
    + Combine tables to create new data sets.

+ Apply business rules to normalize data.

+ Calculate new metrics or statistics based on existing data.

+ Apply rules to remove or process duplicate or invalid data.

- ***Load:*** is the final process, it will load the transformed data into the Human Resources data mart. Data can be stored in separate data tables corresponding to topics or using star schema or snowflake schema depending on the structure of the data.

4.2. Dimension Table's ETL Process

This is the process for ETL data by SSIS tool of DimDepartment table, which is similar for all other dim tables:

- ***Step 1:*** Create a new Integration Services Project in Visual Studio.

- ***Step 2:*** Create a connection manager (AdventureWorks2019 and my database K21406_Group4_HR).

Right-mouse on "Connection Managers" → Click "New Connection Managers" → Double clicks on "OLEDB" → Click "New" → Choose Provider: "Microsoft OLE DB Provider for SQL Server" → Choose server ".\" and the database is "AdventureWorks2019" → Click "OK".

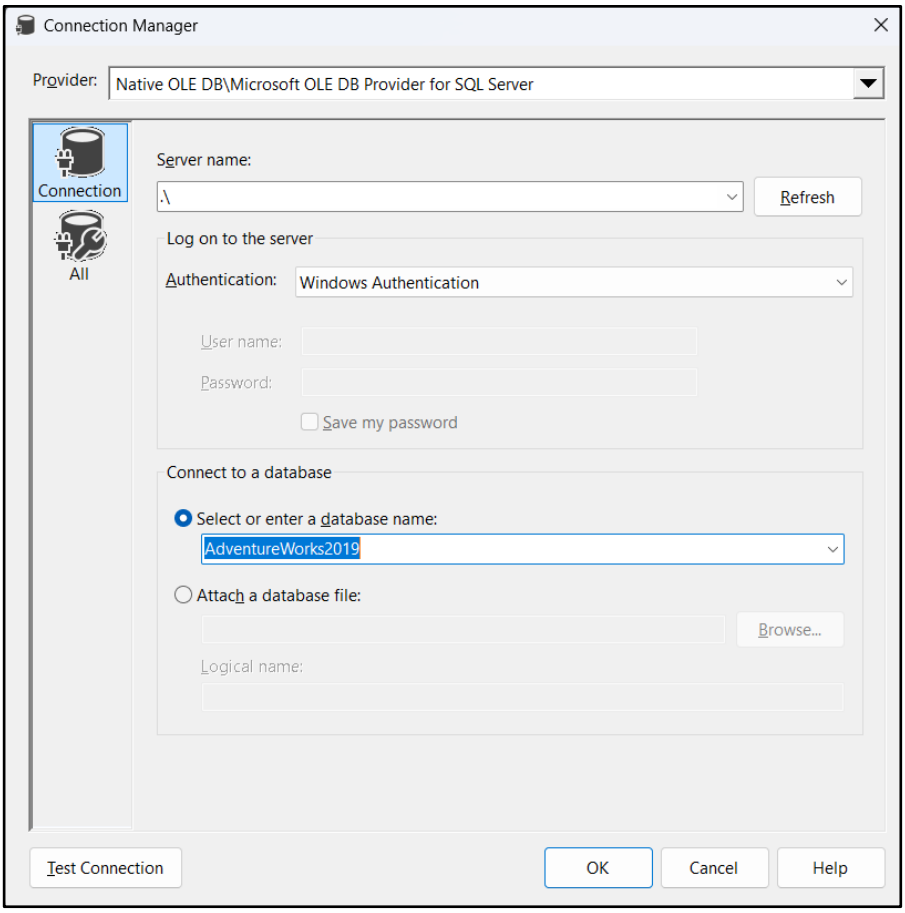


Figure 4.2. Connection Manager (Database AdventureWorks2019)

Then right-mouse on “Connection Managers” → Click “New Connection Managers” → Double clicks on “OLEDB” → Add data connections “K21406_Group4_HR” → Click “OK”.

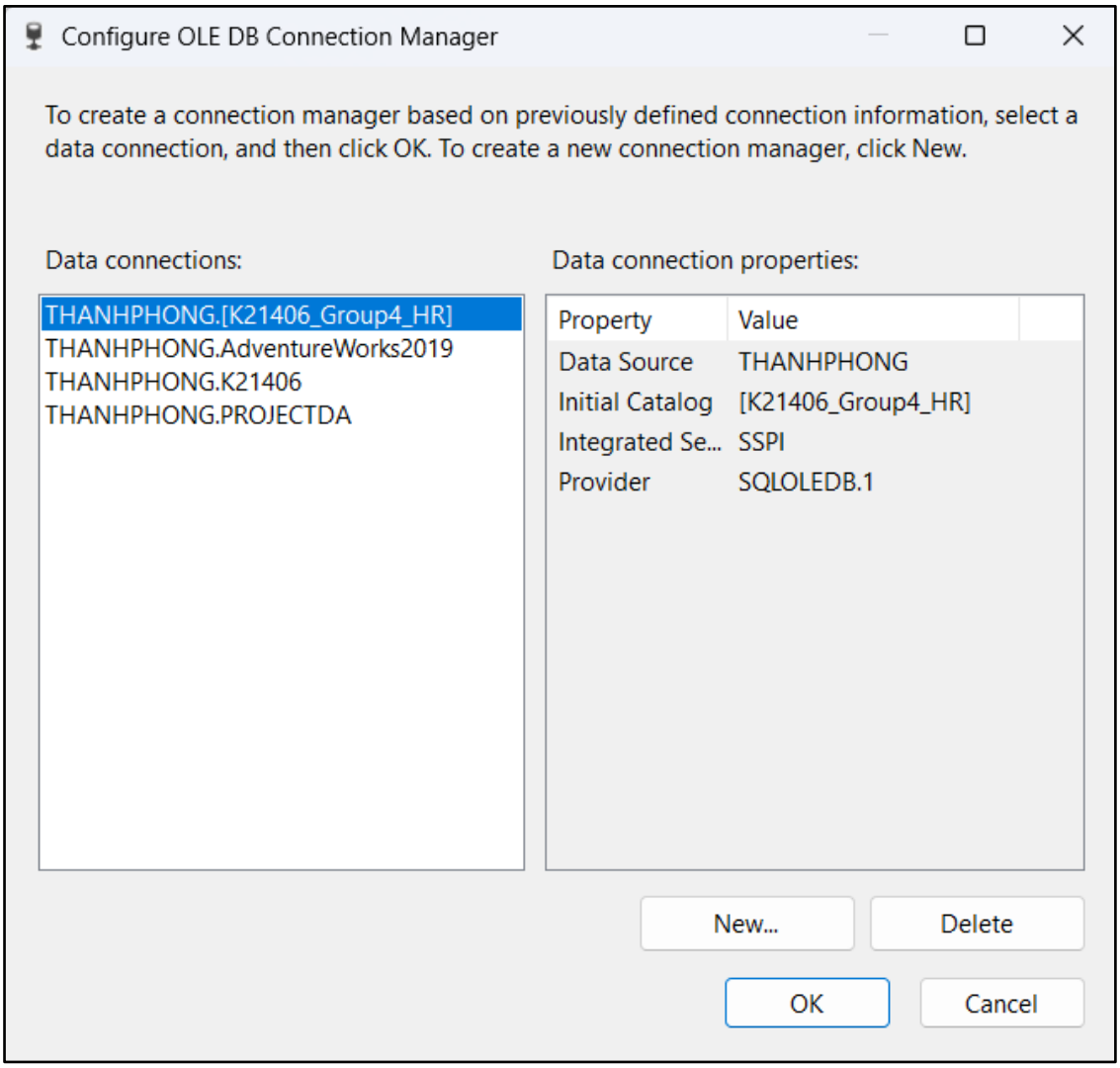


Figure. Create connection manager PROJECTDA

- ***Step 3:*** Create an Overview Diagram for the ETL Data Process of the Corresponding Dim Table.

This diagram will provide an overview of the data integration process from the AdventureWork database to the Dim table we are working on, as well as a clear understanding of the origin of each entity we create.

Specifically, for the DimDepartment table, the ETL process will consist of 4 steps: First, searching for necessary entities from tables in the AdventureWorks database. Next, importing that data into a "Source_Table" within SSIS. Then, generating a new key for the

corresponding Dim table (in this case, DepartmentKey). Finally, transforming the data from the SSIS table to the corresponding table we are working on (here, the DimDepartment table in the K21406_Group4_HR database).

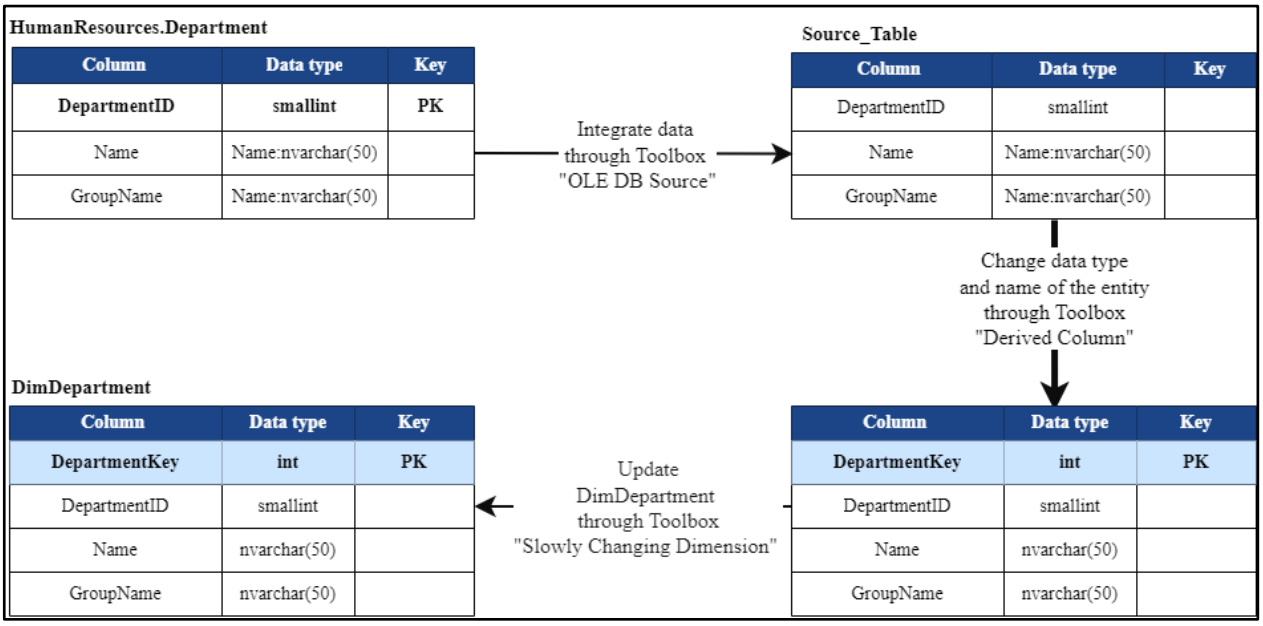


Figure 4.3. Overview of the data ETL process of the DimDepartment table

- ***Step 4:*** Create a new SSIS package.

Right-click “SSIS Packages” > Click “New SSIS Package” > Drag and drop “Data Flow Task” into the design panel → Change the name to “DimDepartment”.

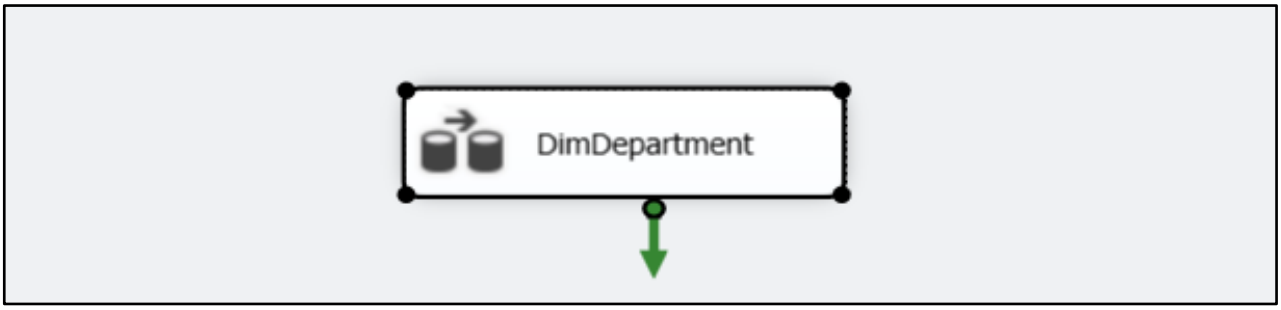


Figure 4.4. Design panel of new SSIS package for Dim table

- ***Step 5:*** Insert database source.

Double click to “DimDepartment” → Drag and drop “OLE DB Source” → Double clicks on “OLE DB Source” → Select “OLE DB connection manager” and data access mode “SQL command”. → Put in the SQL query code in section 3.4 above into “SQL command text”→ Click “OK”.

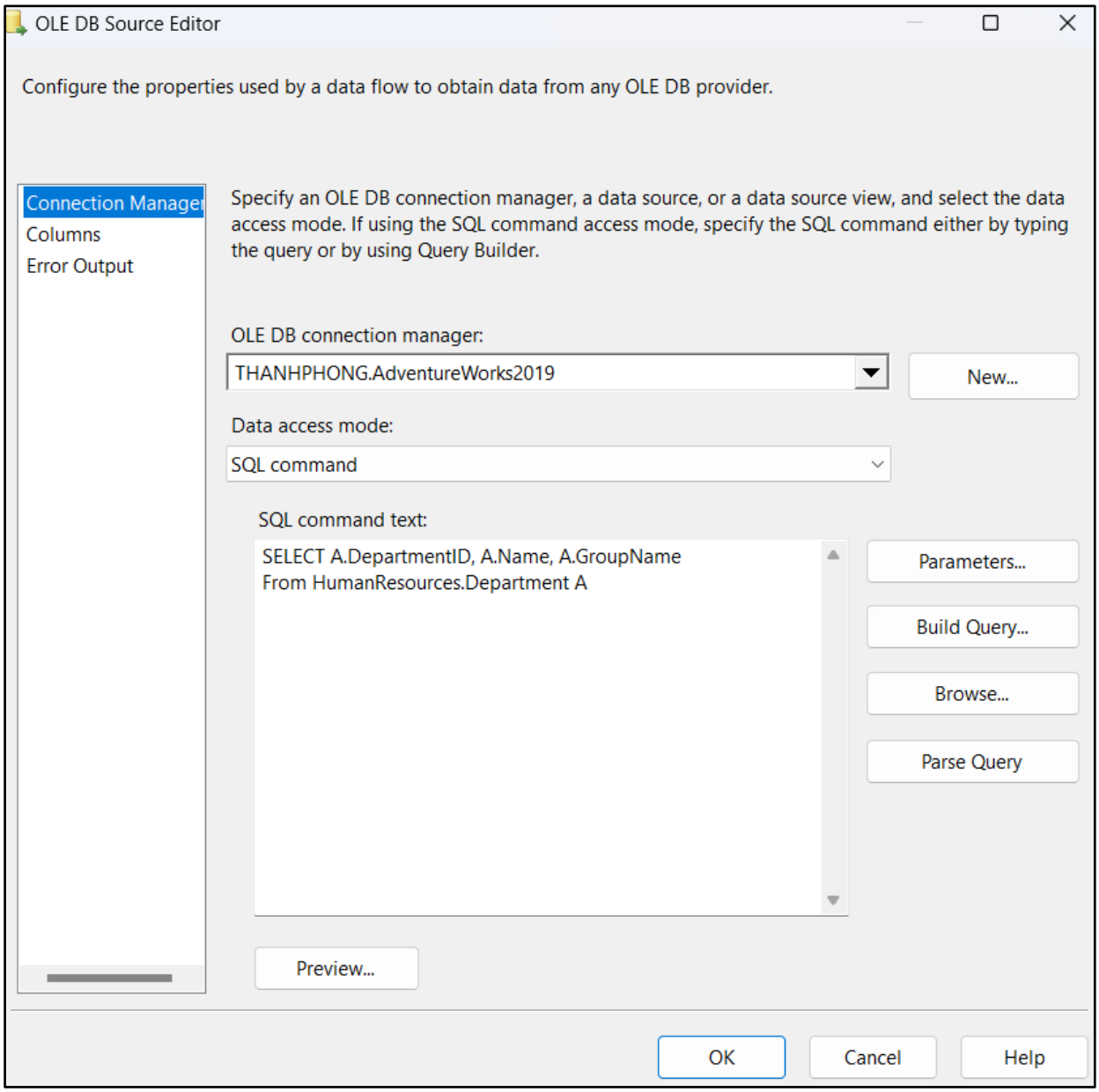


Figure 4.5. Add OLE DB Source by table or view

- ***Step 5:*** Create table keys.

Drag and drop “Derived Column” → Connect the “OLE DB Source” to “Derived Column” → Double click “Derived Column” → Create a new column DepartmentKey as below → Click “OK”.

Note: (DT_I4)DepartmentID is a piece of code intended to insert data from the DepartmentID column to the DepartmentKey column and change the data type back to int (DT_I4).

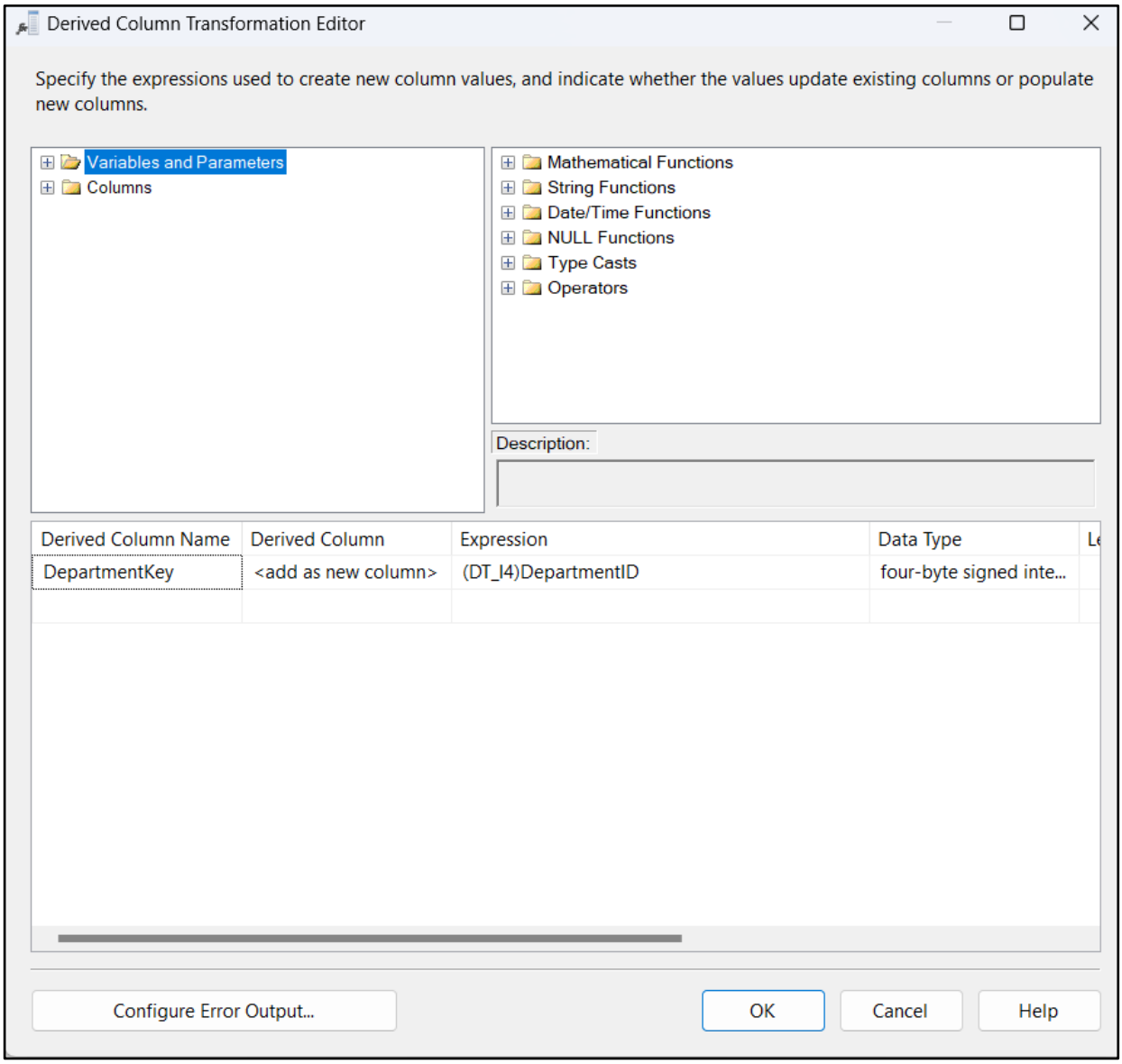


Figure 4.6. Create DepartmentKey in Derived Column

- ***Step 6:*** Configure data destination.

Drag and drop "Slowly Changing Dimension" from toolbox → Choose the destination data warehouse → Select a dimension table to load and map columns → Select at least one business key → Click "Next" → Click "Finish".

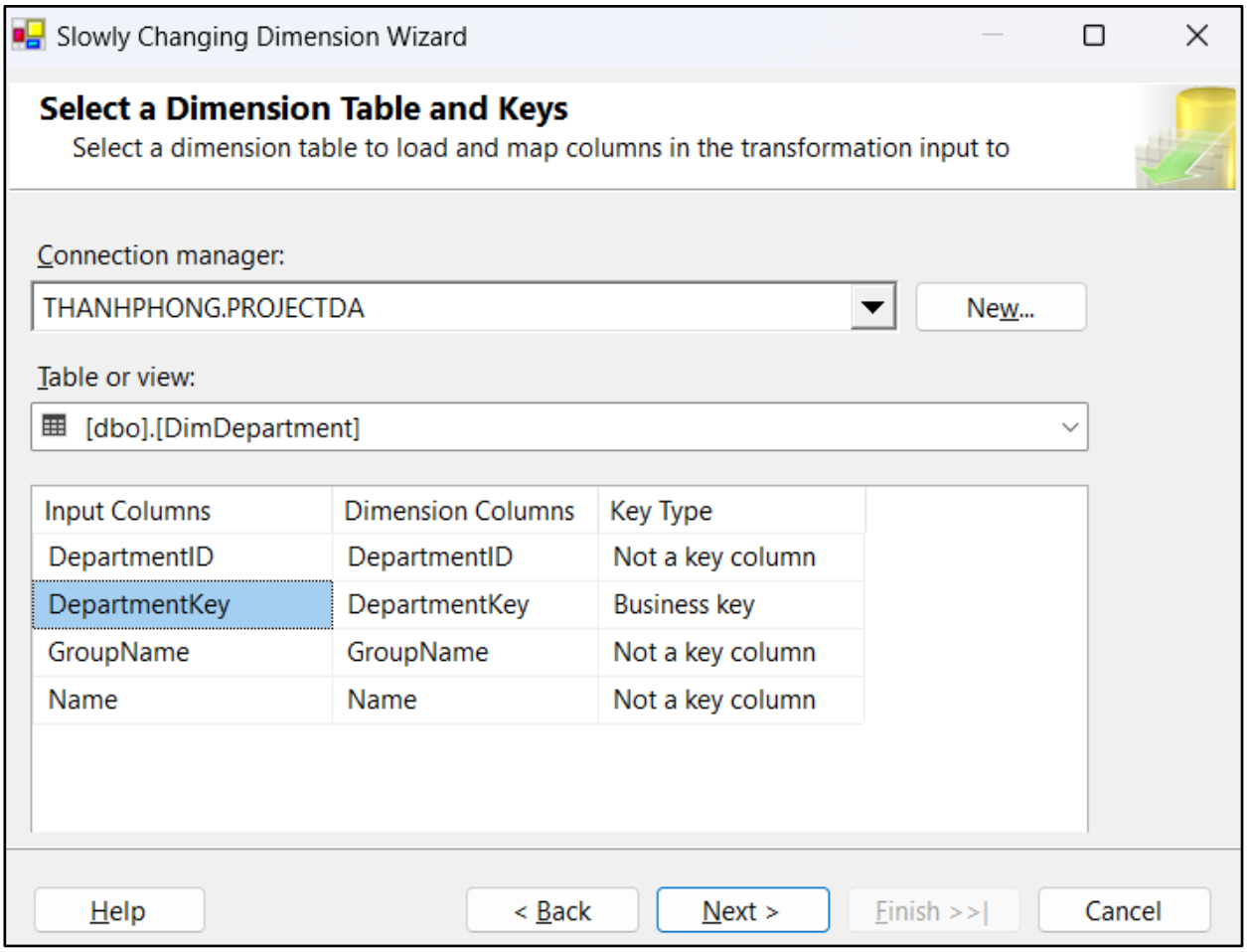
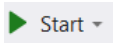


Figure 4.7. Manage changes to dimension columns

- ***Step 7:*** Run package.

Click  to run package → Run package and you will see the result of successfully running the DimDepartment version as shown below. The remaining Dim versions do the same as the DimDepartment version.

4.2.1. DimDepartment

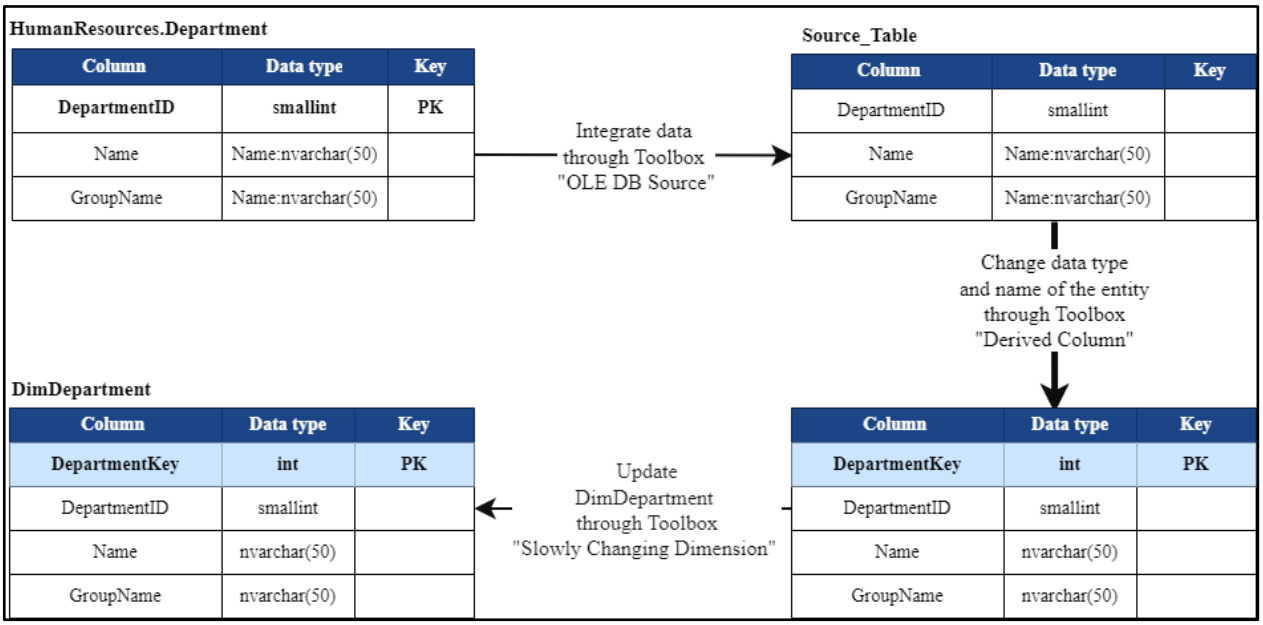


Figure 4.8. The data ETL process of the DimDepartment table

The ETL process for creating the DimDepartment table involves extracting data from the HumanResources.Department table, transforming it to meet schema requirements, integrating it into the destination table, and loading it into the data warehouse. Key steps include extraction of attributes like DepartmentID and Name, transformation to ensure data consistency, integration with slowly changing dimension techniques to handle historical changes, and loading into the DimDepartment table. Data quality and consistency are prioritized throughout the process, with optimization strategies employed for efficient query performance.

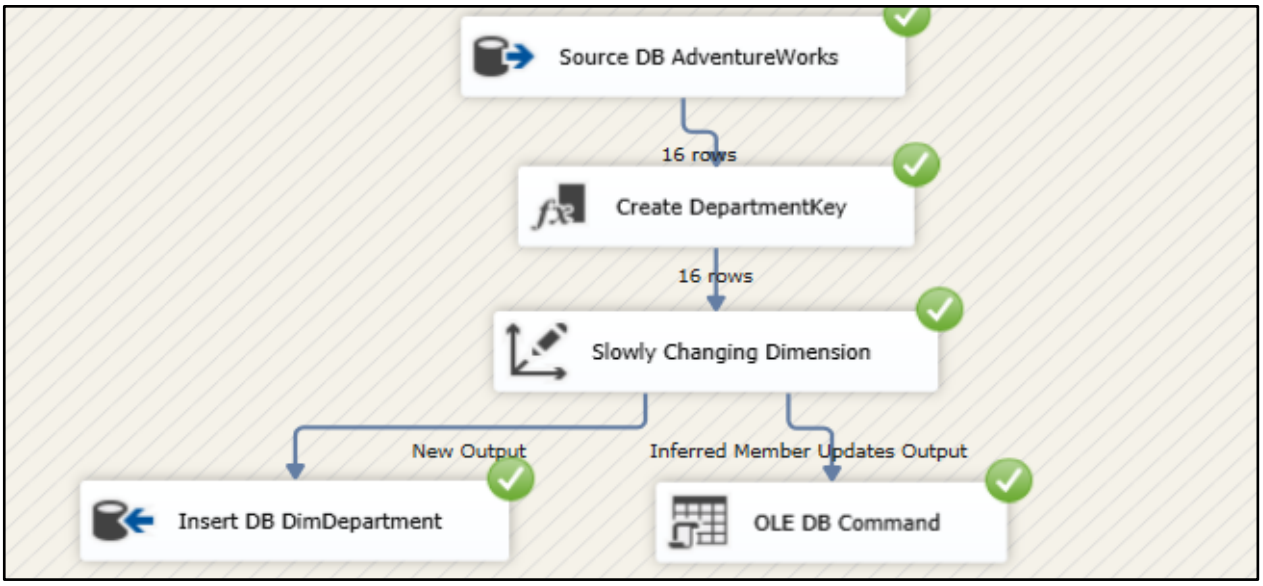


Figure 4.9. SSIS for DimDepartment

```
SELECT TOP 100
    DepartmentKey,
    DepartmentID,
    Name,
    GroupName
FROM [dbo].[DimDepartment];
```

287 %

Results | Messages

	DepartmentKey	DepartmentID	Name	GroupName
1	1	1	Engineering	Research and Development
2	2	2	Tool Design	Research and Development
3	3	3	Sales	Sales and Marketing
4	4	4	Marketing	Sales and Marketing
5	5	5	Purchasing	Inventory Management
6	6	6	Research and Development	Research and Development
7	7	7	Production	Manufacturing
8	8	8	Production Control	Manufacturing
9	9	9	Human Resources	Executive General and Administration
10	10	10	Finance	Executive General and Administration
11	11	11	Information Services	Executive General and Administration
12	12	12	Document Control	Quality Assurance
13	13	13	Quality Assurance	Quality Assurance
14	14	14	Facilities and Maintenance	Executive General and Administration
15	15	15	Shipping and Receiving	Inventory Management
16	16	16	Executive	Executive General and Administration

Figure 4.10. Top 100 rows of DimDepartment table

The query retrieves the top 100 rows from the DimDepartment table, providing a snapshot of department-related information stored in the data warehouse. Each row represents a

department and includes details such as the DepartmentKey (primary key), DepartmentID, Name, and GroupName. This data offers insights into the organizational structure of departments within the organization, allowing users to understand how departments are categorized and named. Analyzing these rows can help stakeholders gain a better understanding of the departmental hierarchy and distribution of departments within the company.

4.2.2. DimEmployee

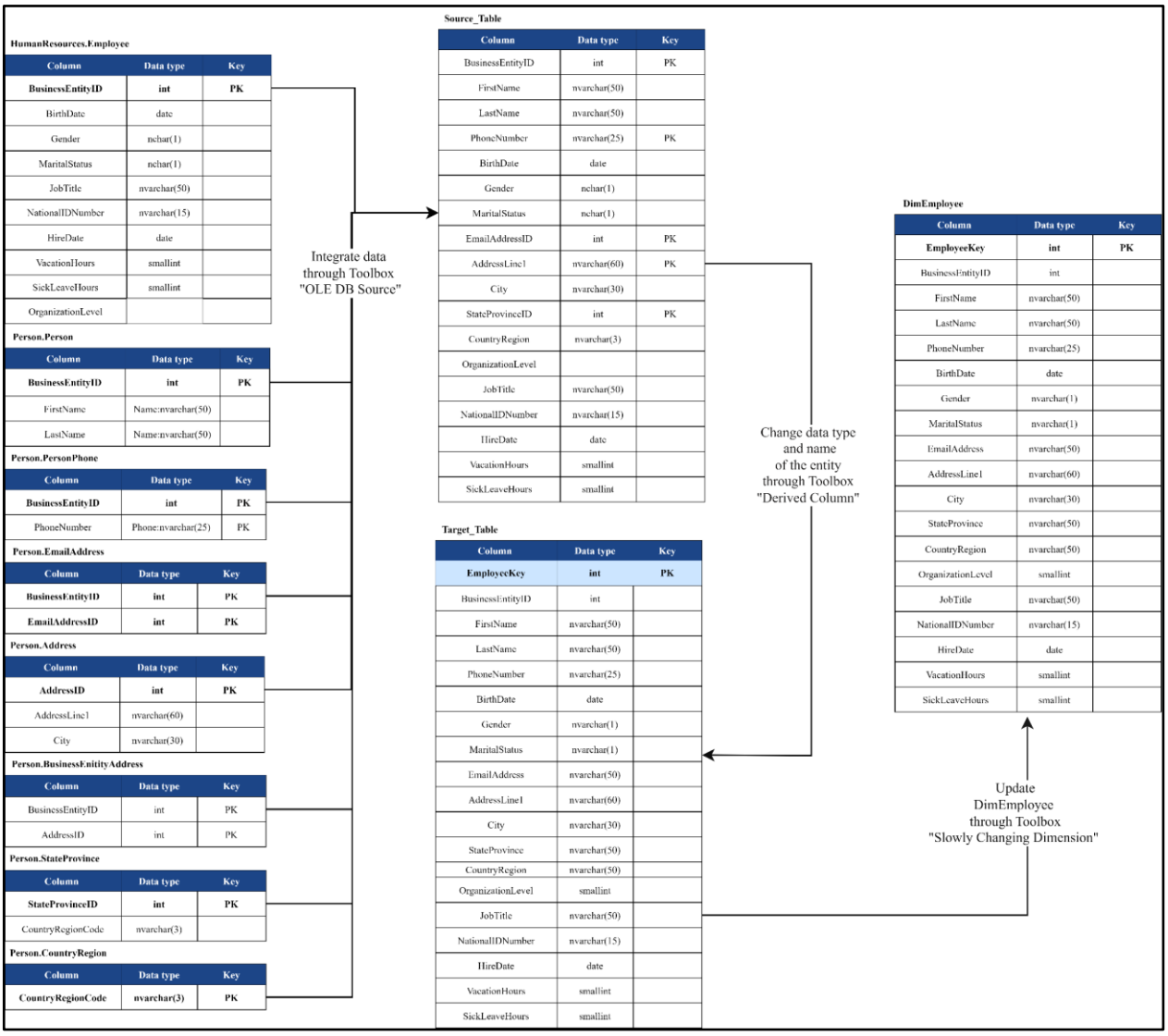


Figure 4.11. The data ETL process of the DimEmployee table

The ETL process for the DimEmployee table involves integrating data from various source tables in the AdventureWorks database, applying transformations such as derived columns and data adjustments, and handling slowly changing dimensions (SCDs) using SCD Type 2 methodology. Integrated data undergoes transformations to ensure consistency and accuracy, with derived columns created as needed. SCD handling involves tracking historical changes to employee attributes by adding effective date and expiration date columns. Once these steps are completed, the DimEmployee table is created in the data

warehouse, providing a comprehensive and accurate representation of employee data for analytical purposes.

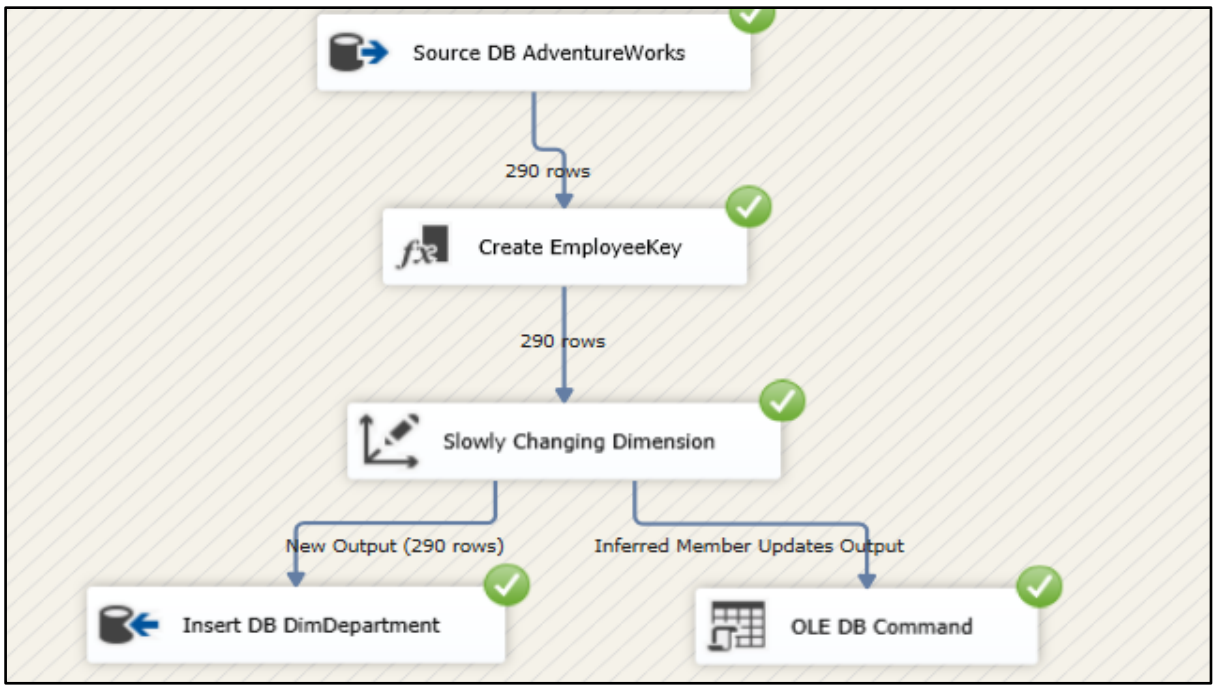
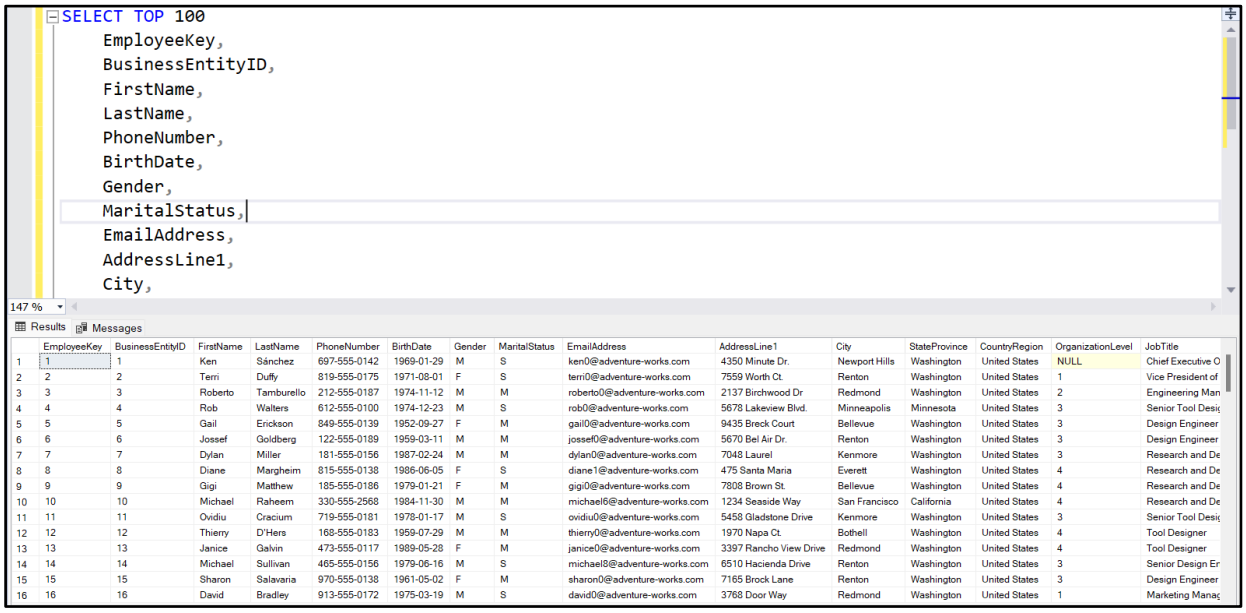


Figure 4.12. SSIS for DimEmployee



```
SELECT TOP 100
    EmployeeKey,
    BusinessEntityID,
    FirstName,
    LastName,
    PhoneNumber,
    BirthDate,
    Gender,
    MaritalStatus,
    EmailAddress,
    AddressLine1,
    City,
```

147 % — Results — Messages

	EmployeeKey	BusinessEntityID	FirstName	LastName	PhoneNumber	BirthDate	Gender	MaritalStatus	EmailAddress	AddressLine1	City	StateProvince	CountryRegion	OrganizationLevel	JobTitle
1	1	1	Ken	Sánchez	697-555-0142	1969-01-29	M	S	ken0@adventure-works.com	4350 Minute Dr.	Newport Hills	Washington	United States	NULL	Chief Executive O
2	2	2	Terri	Duffy	819-555-0175	1971-08-01	F	S	terri0@adventure-works.com	7559 Worth Ct.	Renton	Washington	United States	1	Vice President of
3	3	3	Roberto	Tamburello	212-555-0187	1974-11-12	M	M	roberto0@adventure-works.com	2137 Birchwood Dr	Redmond	Washington	United States	2	Engineering Man
4	4	4	Rob	Walters	612-555-0100	1974-12-23	M	S	rob0@adventure-works.com	5678 Lakeview Blvd.	Minneapolis	Minnesota	United States	3	Senior Tool Desig
5	5	5	Gail	Erickson	849-555-0139	1952-09-27	F	M	gail0@adventure-works.com	9435 Breck Court	Bellevue	Washington	United States	3	Design Engineer
6	6	6	Jossef	Goldberg	122-555-0189	1959-03-11	M	M	jossef0@adventure-works.com	5670 Bel Air Dr.	Renton	Washington	United States	3	Design Engineer
7	7	7	Dylan	Miller	181-555-0156	1987-02-24	M	M	dylan0@adventure-works.com	7048 Laurel	Kenmore	Washington	United States	3	Research and De
8	8	8	Diane	Margheim	815-555-0138	1986-06-05	F	S	diane1@adventure-works.com	475 Santa Maria	Everett	Washington	United States	4	Research and De
9	9	9	Gigi	Matthew	185-555-0186	1979-01-21	F	M	gigi0@adventure-works.com	7808 Brown St.	Bellevue	Washington	United States	4	Research and De
10	10	10	Michael	Raheem	330-555-2568	1984-11-30	M	M	michael6@adventure-works.com	1234 Seaside Way	San Francisco	California	United States	4	Research and De
11	11	11	Ovidiu	Cracium	719-555-0181	1978-01-17	M	S	ovidiu0@adventure-works.com	5458 Gladstone Drive	Kenmore	Washington	United States	3	Senior Tool Desig
12	12	12	Thierry	D'Hers	168-555-0183	1959-07-29	M	M	thierry0@adventure-works.com	1970 Napa Ct.	Bothell	Washington	United States	4	Tool Designer
13	13	13	Janice	Galvin	473-555-0117	1989-05-28	F	M	janice0@adventure-works.com	3397 Rancho View Drive	Redmond	Washington	United States	4	Tool Designer
14	14	14	Michael	Sullivan	465-555-0156	1979-06-16	M	S	michael8@adventure-works.com	6510 Hacienda Drive	Renton	Washington	United States	3	Senior Design Er
15	15	15	Sharon	Salavaria	970-555-0138	1961-05-02	F	M	sharon0@adventure-works.com	7165 Brock Lane	Renton	Washington	United States	3	Design Engineer
16	16	16	David	Bradley	913-555-0172	1975-03-19	M	S	david0@adventure-works.com	3768 Door Way	Redmond	Washington	United States	1	Marketing Manag

Figure 4.13. Top 100 rows of DimEmployee table

The figure displays the top 100 rows of the DimEmployee table, showcasing a subset of employee-related data stored in the data warehouse. Each row represents an employee and includes various attributes such as EmployeeKey (primary key), BusinessEntityID,

FirstName, LastName, PhoneNumber, BirthDate, Gender, MaritalStatus, EmailAddress, AddressLine1, City, StateProvince, CountryRegion, OrganizationLevel, JobTitle, NationalIDNumber, HireDate, VacationHours, and SickLeaveHours. This information provides insights into the demographics, employment details, and contact information of employees within the organization. Analyzing these rows enables stakeholders to understand the workforce composition and demographics, facilitating strategic decision-making and workforce management.

4.2.3. DimShift

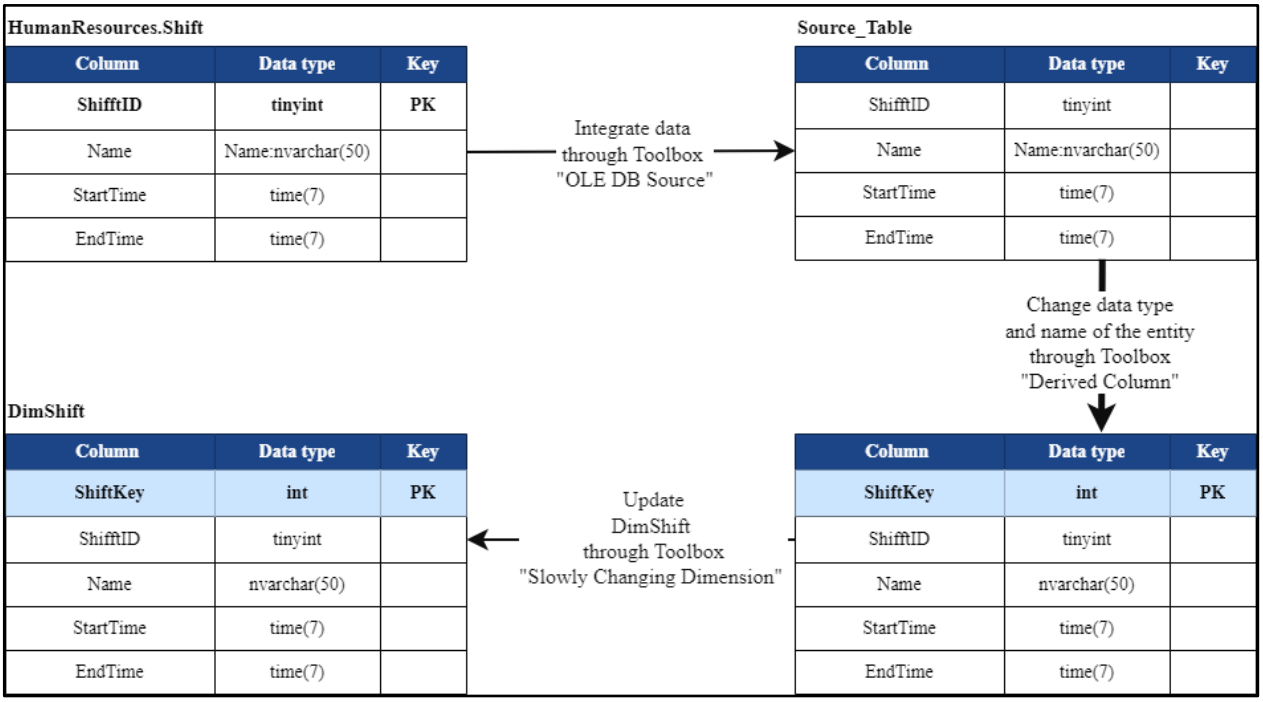


Figure 4.14. The data ETL process of the DimShift table

The ETL process for creating the DimShift table begins with data extraction from the source table "HumanResources.Shift," capturing shift-related attributes. Extracted data undergoes transformation to ensure consistency and alignment with the desired schema, including handling of slowly changing dimensions (SCDs) to track historical changes. The schema design incorporates attributes such as ShiftID, Name, StartTime, EndTime to provide a structured representation of shift information. Transformed data is then loaded

into the DimShift table, following the defined SCD strategy to maintain historical accuracy. Rigorous verification and testing validate the accuracy of the loaded data, ensuring reliability for analytical purposes.

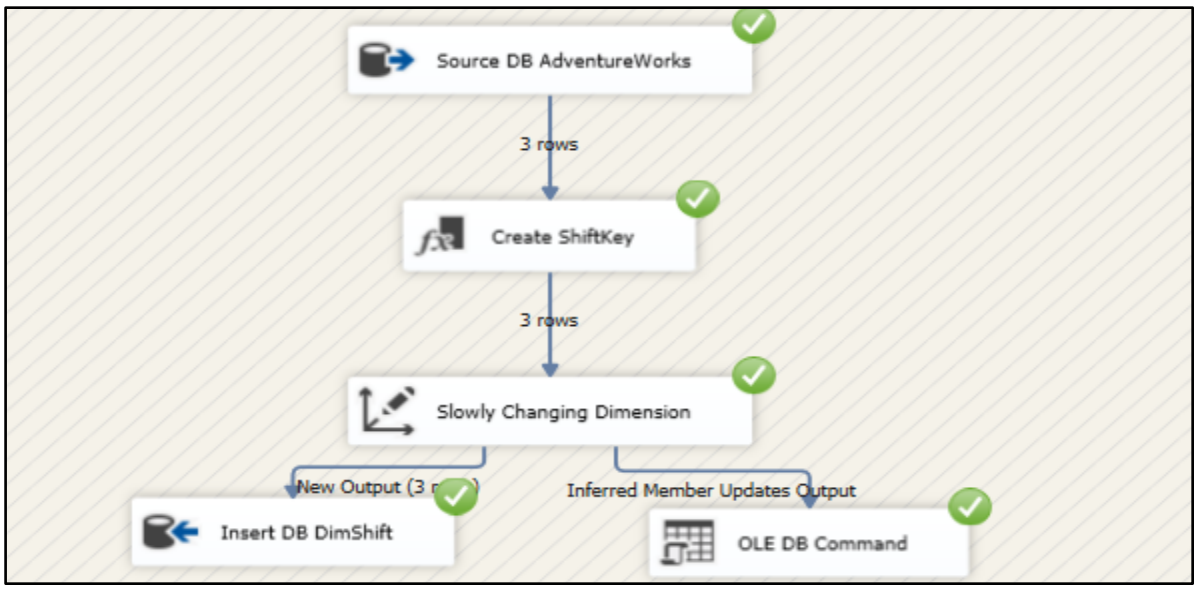


Figure 4.15. SSIS for DimShift

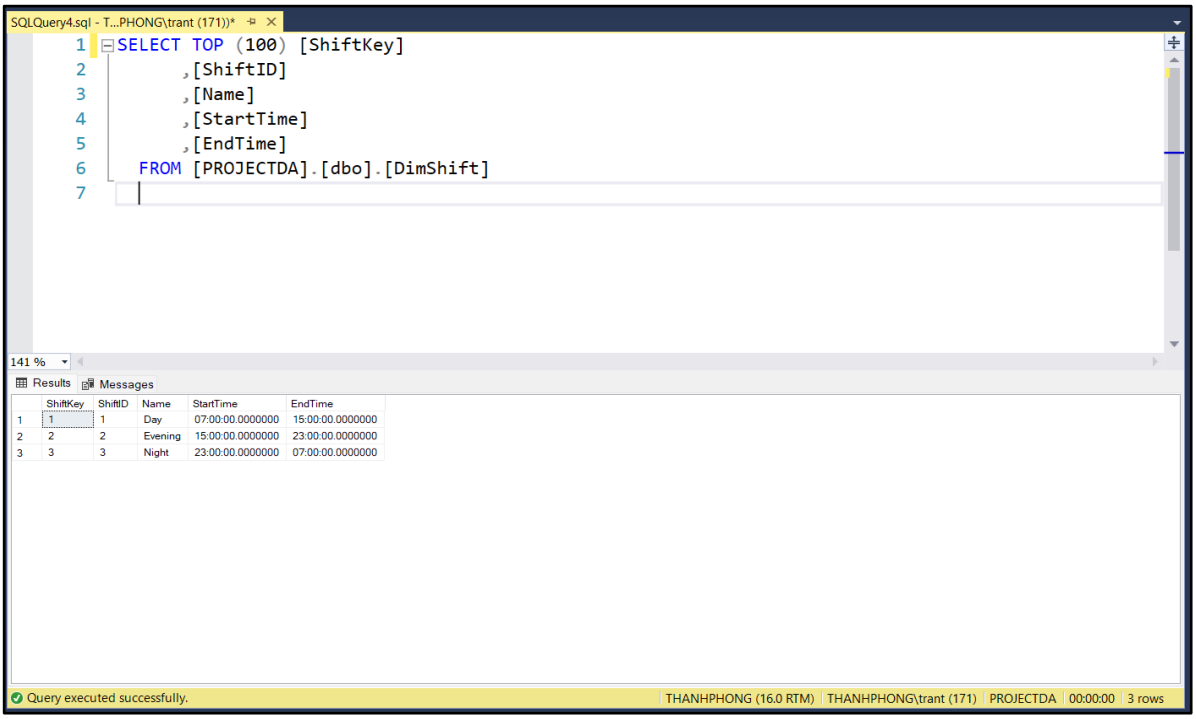


Figure 4.16. Top 100 rows of DimShift table

The figure presents the top 100 rows of the DimShift table, which contains information about different shifts within the organization. Each row represents a shift and includes

attributes such as ShiftKey (primary key), ShiftID, Name, StartTime, and EndTime. These attributes provide details about the shifts' identifiers, names, and start and end times. Analyzing this data can help stakeholders understand the organization's shift schedules, monitor workforce allocation, and optimize resource utilization. Additionally, it facilitates effective workforce planning and scheduling to meet operational demands and ensure smooth business operations.

4.2.4. DimTime

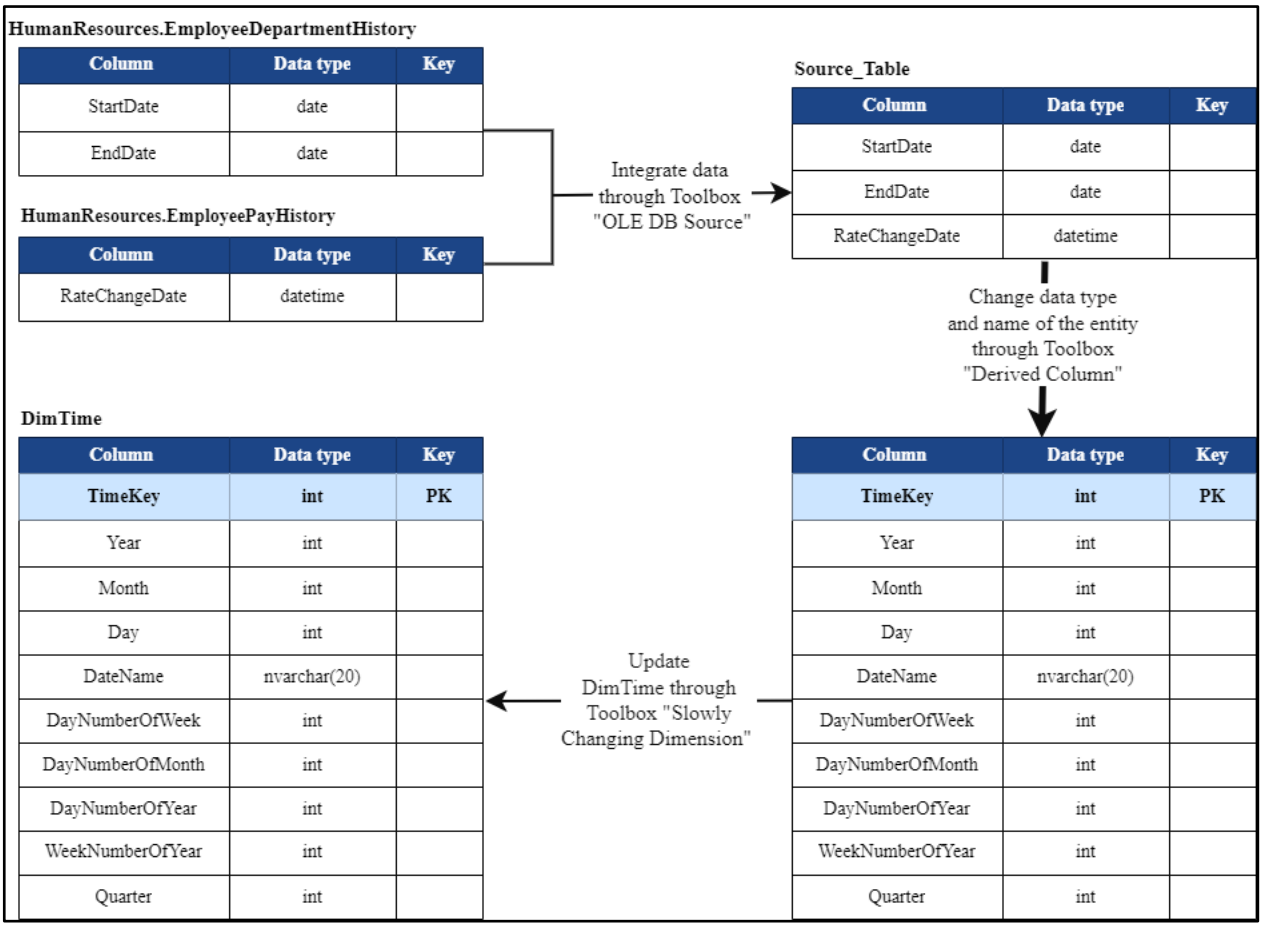


Figure 4.17. The data ETL process of the DimTime table

The data extraction process for the DimTime table involves sourcing data from two main tables within the HumanResources module of the AdventureWorks database: EmployeeDepartmentHistory and EmployeePayHistory. These tables contain crucial information related to employees' department history and pay changes, respectively. By

querying these tables, relevant attributes such as StartDate, EndDate, RateChangeDate, and other pertinent time-related data are extracted. Subsequently, the extracted data undergoes transformation to standardize formats, derive additional attributes, and handle any inconsistencies. Finally, the transformed data is loaded into the DimTime table, ensuring that it accurately represents the time-related dimensions required for analysis and reporting in the data warehouse.

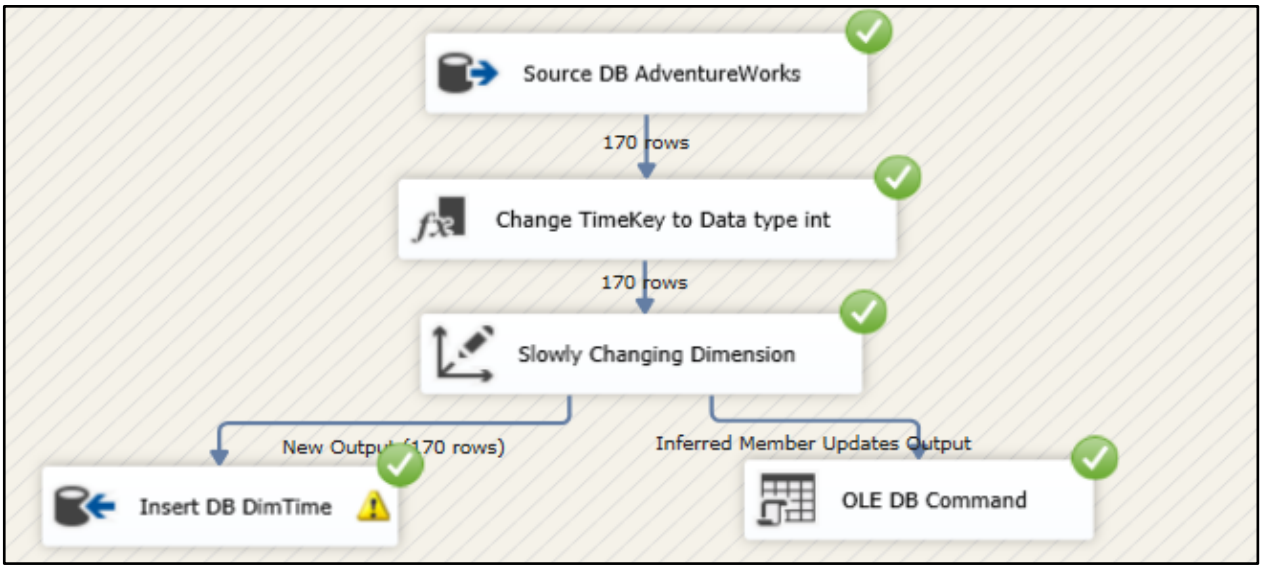
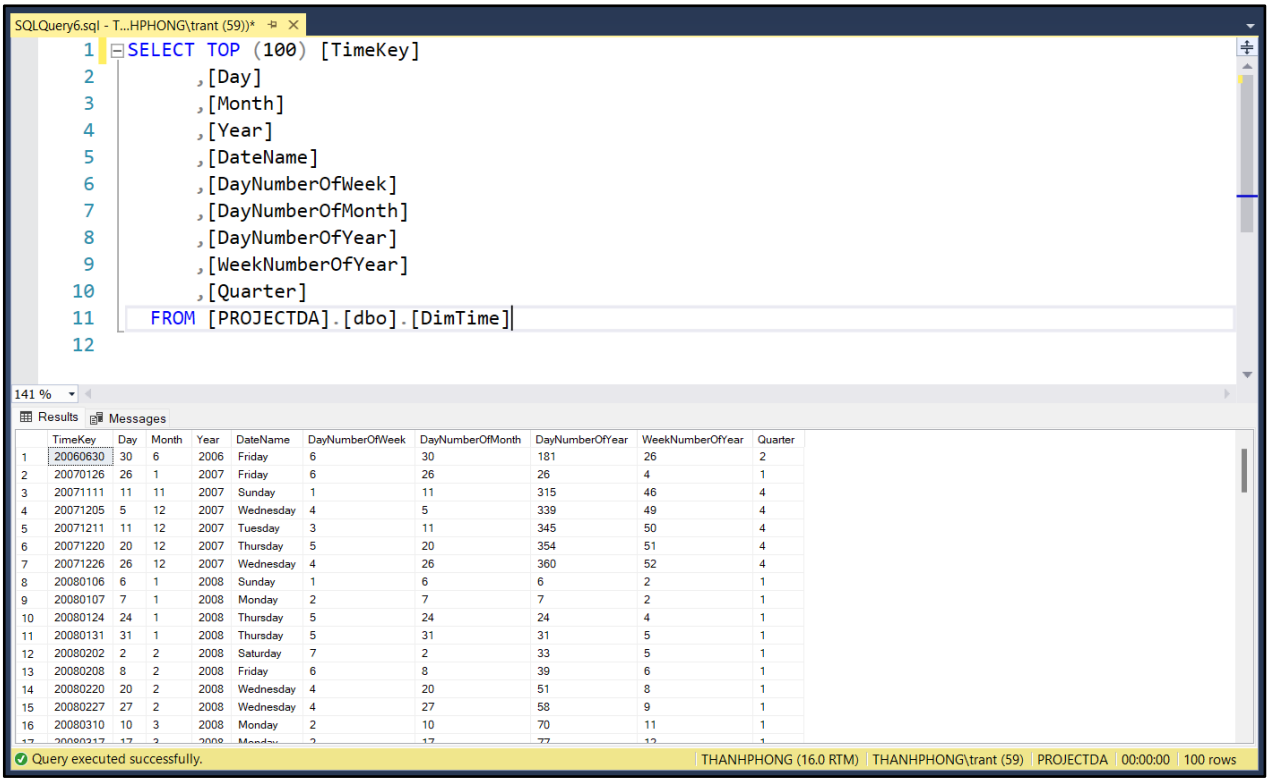


Figure 4.18. SSIS for DimTime



	TimeKey	Day	Month	Year	DateName	DayNumberOfWeek	DayNumberOfMonth	DayNumberOfYear	WeekNumberOfYear	Quarter
1	20060630	30	6	2006	Friday	6	30	181	26	2
2	20070126	26	1	2007	Friday	6	26	26	4	1
3	20071111	11	11	2007	Sunday	1	11	315	46	4
4	20071205	5	12	2007	Wednesday	4	5	339	49	4
5	20071211	11	12	2007	Tuesday	3	11	345	50	4
6	20071220	20	12	2007	Thursday	5	20	354	51	4
7	20071226	26	12	2007	Wednesday	4	26	360	52	4
8	20080106	6	1	2008	Sunday	1	6	6	2	1
9	20080107	7	1	2008	Monday	2	7	7	2	1
10	20080124	24	1	2008	Thursday	5	24	24	4	1
11	20080131	31	1	2008	Thursday	5	31	31	5	1
12	20080202	2	2	2008	Saturday	7	2	33	5	1
13	20080208	8	2	2008	Friday	6	8	39	6	1
14	20080220	20	2	2008	Wednesday	4	20	51	8	1
15	20080227	27	2	2008	Wednesday	4	27	58	9	1
16	20080310	10	3	2008	Monday	2	10	70	11	1

Figure 4.19. Top 100 rows of DimTime table

The figure displays the top 100 rows of the DimTime table, which serves as a standardized representation of time-related attributes in the data warehouse. Each row represents a specific time instance and includes attributes such as TimeKey (primary key), Year, Month,

Day, DateName, DayNumberOfWeek, DayNumberOfMonth, DayNumberOfYear, WeekNumberOfYear, and Quarter. These attributes provide detailed information about the temporal aspects of each record, facilitating time-based analysis and reporting. Analyzing this data can help stakeholders identify trends, patterns, and seasonality in business operations, enabling informed decision-making and strategic planning.

4.3. Fact Table's ETL Process

For the Fact tables, the ETL data processing procedure using SSIS for these tables, although somewhat different, is almost similar to the above Dim tables. Below is the ETL process for the FactPayHistory table, similar to the FactDepartmentHistory table:

- ***Step 1:*** Create an overview diagram for the ETL data process of the Corresponding Fact Table.

Specifically, for the FactPayHistory table, the ETL process will consist of 4 steps: First, searching for necessary entities from tables in the AdventureWorks database. Next, importing that data into a "Source_Table" within SSIS. Then, generating a new key for the corresponding Fact table (in this case, FactPayHistoryKey). Finally, transform the data from the SSIS table to the corresponding table we are working on (here, the FactPayHistory table in the K21406_Group4_HR database).

Note: Unlike the Dim tables, during the ETL process of the Fact table, it is crucial to perform lookups on foreign keys (FK) to verify whether the foreign keys passed to the FactPayHistory table exist in each corresponding Dim table or not. If they do not exist, there is a high probability of encountering errors during execution. In case of errors, it is likely due to inaccuracies in the query code passed to the "Insert database source" process of the corresponding Dim table.

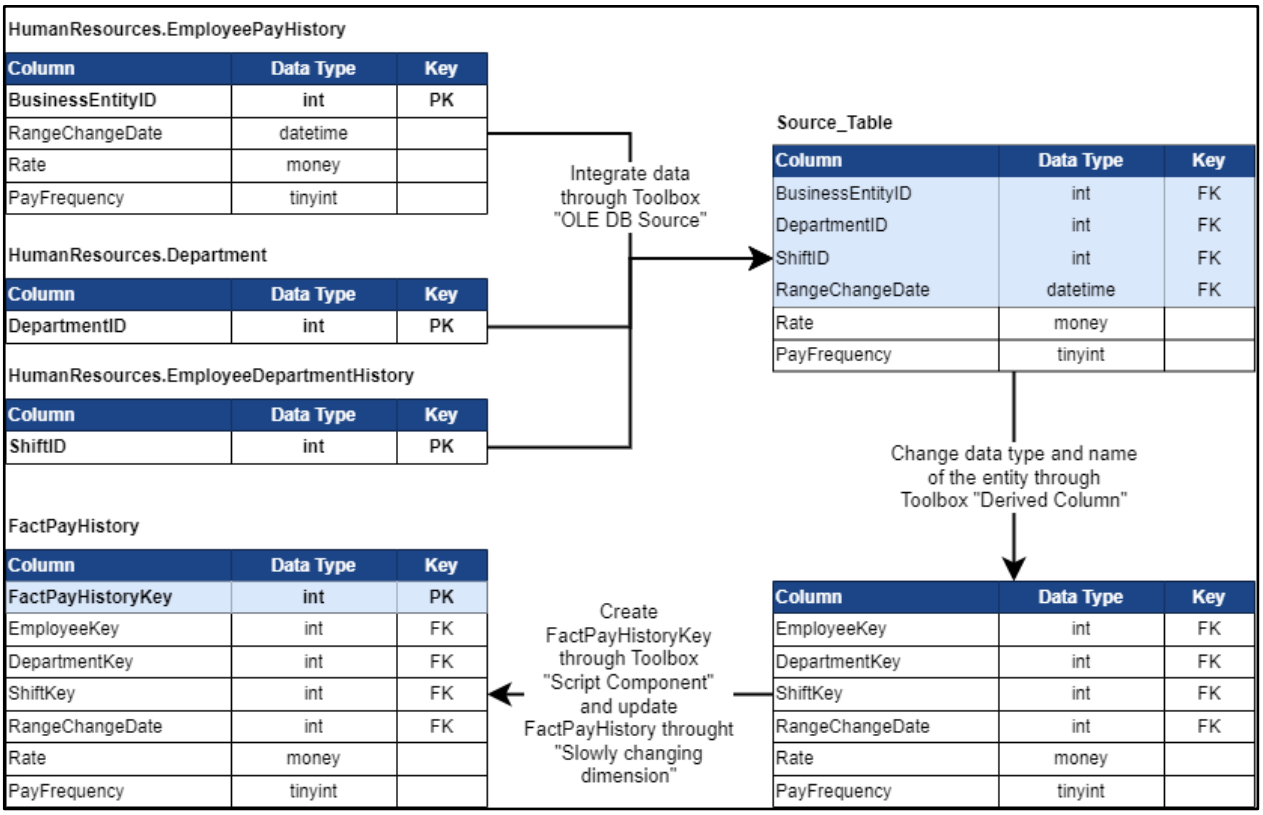


Figure 4.20. Overview of the data ETL process of the FactPayHistory table

- ***Step 2:*** Create new SSIS Packages and rename “FactPayHistory”.
- ***Step 3:*** Drag Data Flow Task from SSIS Tool box.
- ***Step 4:*** Insert database source.

Double click to “FactPayHistory” → Drag and drop “OLE DB Source” → Double clicks on “OLE DB Source” → Select “OLE DB connection manager” and data access mode “SQL command”. → Put in the SQL query code in section 3.5 above into “SQL command text”→ Click “OK”.

- ***Step 5:*** Change the data type of FK columns.

Drag and drop “Derived Column” → Connect the “OLE DB Source” to “Derived Column” → Double click “Derived Column” → Create a new column as below → Click “OK”.

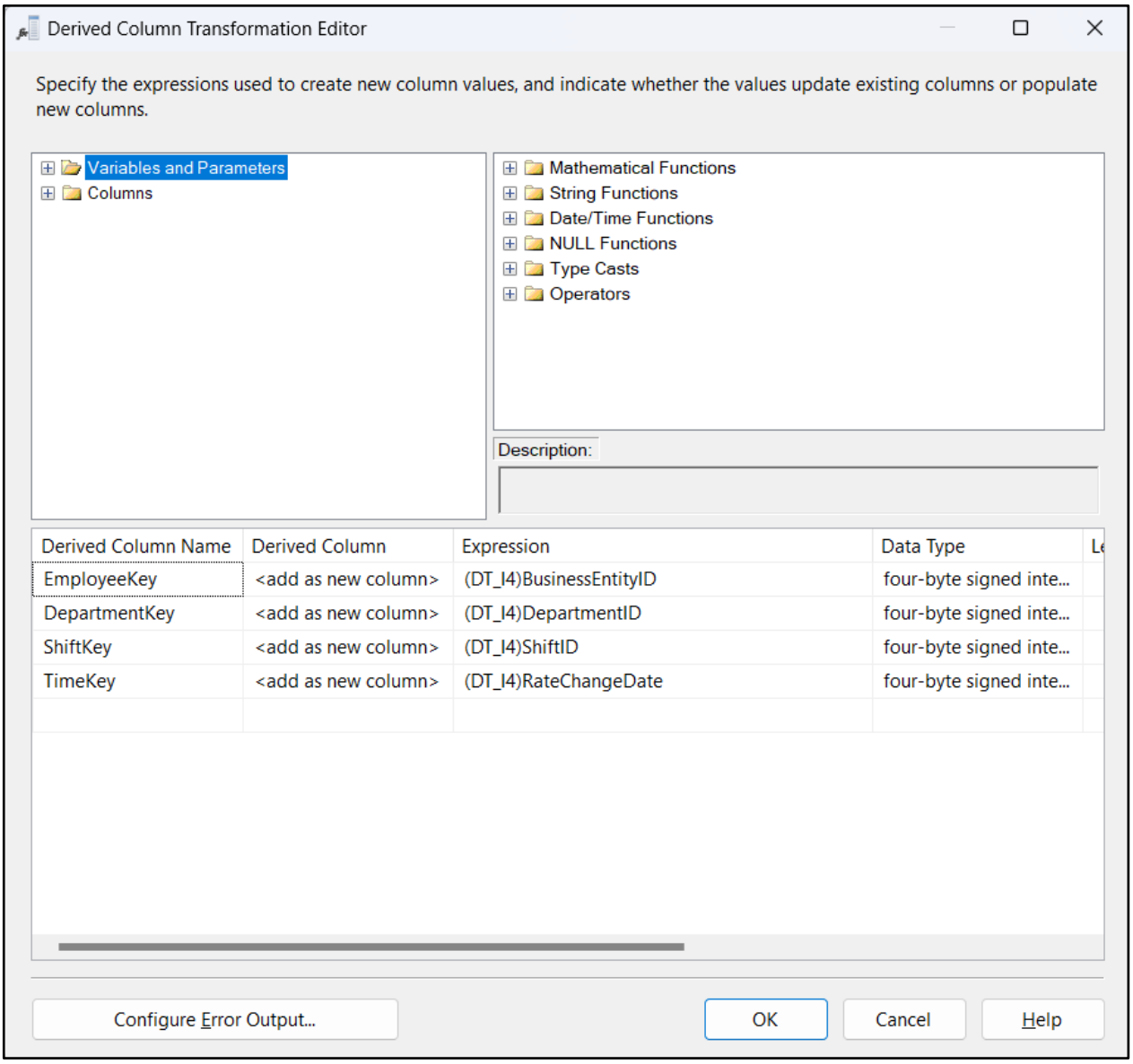


Figure 4.21. Derived column in FactPayHistory table

- ***Step 6:*** Create table keys.

Drag and drop “Script Component” → Rename and Double click “Script Component” → Click “Inputs and Outputs” → Click “Add Column” to create a NumberKey column.

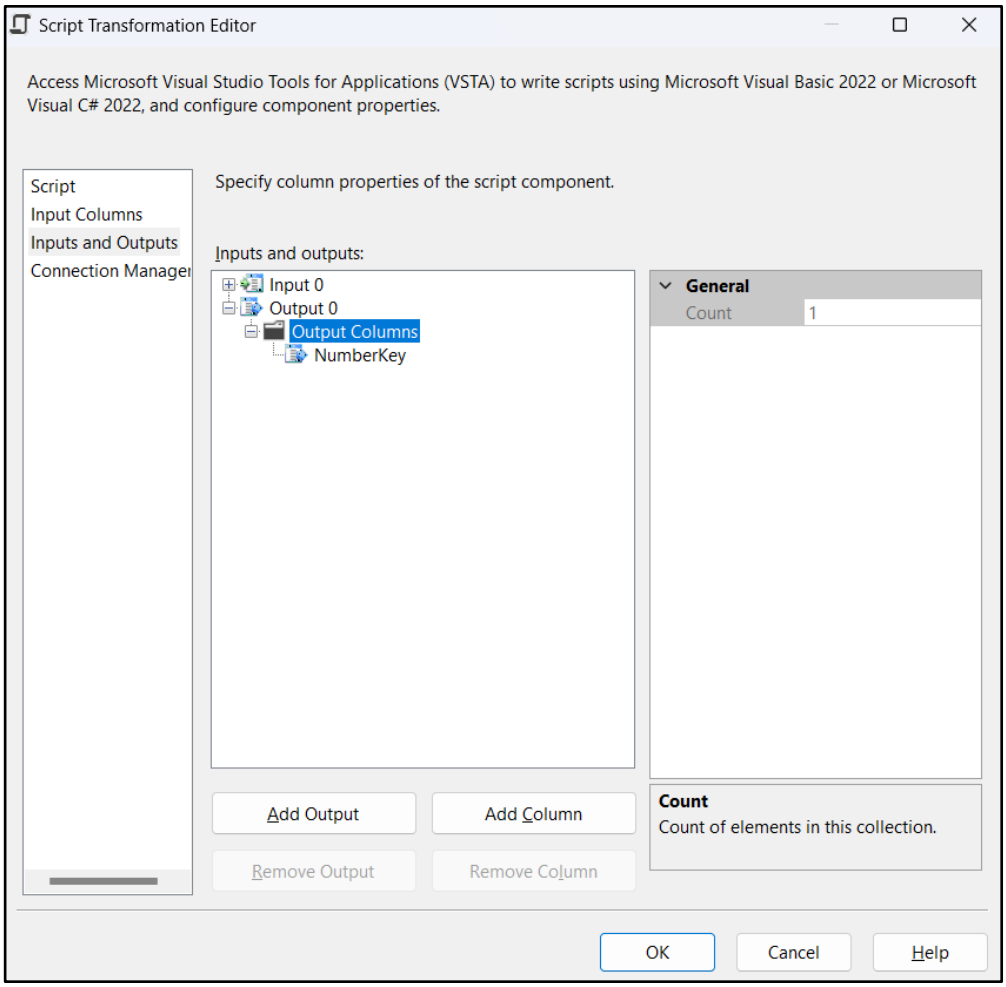


Figure 4.22. Script transformation editor

→ Click “Script” → Click “Edit script” → Edit as shown below → Save the change→ Click “OK” to finish.

```
public class ScriptMain : UserComponent
{
    int row = 1;
    Help:  Using Integration Services variables and parameters

    Help:  Using Integration Services Connnection Managers

    Help:  Firing Integration Services Events
```

```
public override void Input0_ProcessInputRow(Input0Buffer Row)
{
    Row.NumberKey = row;
    row = row + 1;
    /*
     * Add your code here
     */
}
```

Figure 4.23. Code in "edit script" to create automatic incremental key in FactPayHistory table

- ***Step 7:*** Create Lockup to pour data from the corresponding Dim tables.

Drag and drop "Lookup" → Rename to "Lookup EmployeeKey" → Connect the "OLE DB Source" to "Lookup EmployeeKey" → Double click "Lookup EmployeeKey" → Choose the corresponding Dim table (DimEmployee) → Click "Columns" → Drag choose relationship between two tables through corresponding foreign keys.

Here, as mentioned above, we will create a Lookup for the 4 Dim tables: DimDepartment, DimEmployee, DimShift, and DimTime to verify whether the foreign keys passed to the Fact table exist in each corresponding Dim table or not.

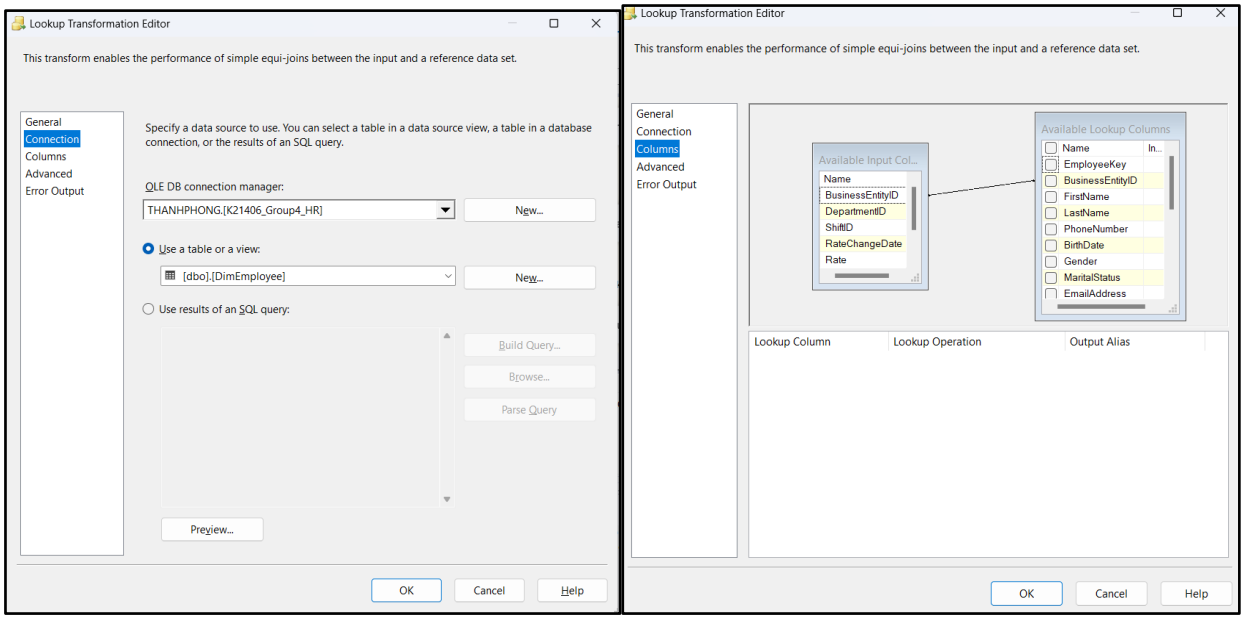
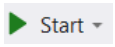


Figure 4.24. Look Transformation Editor and Relationship between two table

- ***Step 8:*** Configure data destination.

Drag and drop “Slowly Changing Dimension” from toolbox → Choose the destination data warehouse → Select a dimension table to load and map columns → Select at least one business key → Click “Next” → Click “Finish”.

- ***Step 9:*** Run package.

Click 

to run package → Run package and you will see the result of successfully running the FactPayHistory version as shown below. Do the same with FactDepartmentHistory.

4.3.1. *FactPayHistory*

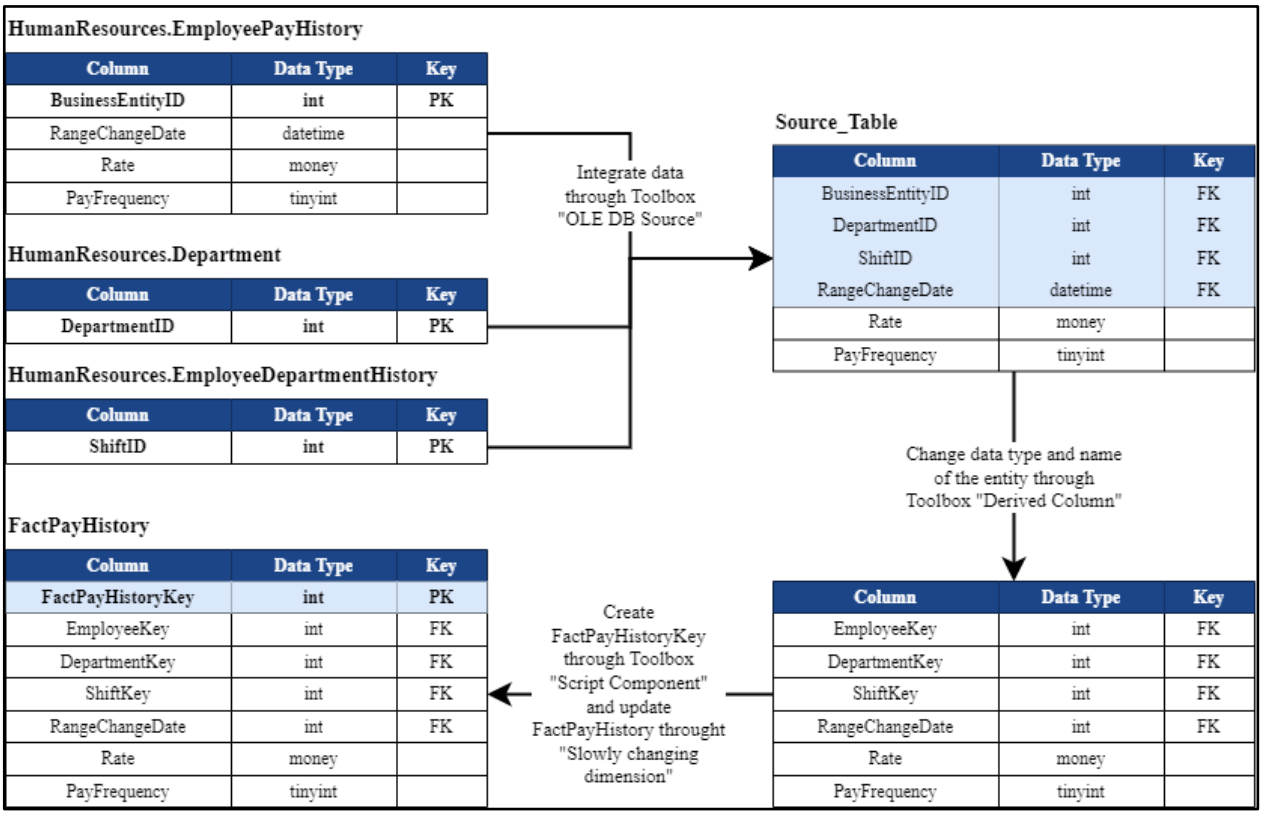


Figure 4.25. The data ETL process of the FactPayHistory table

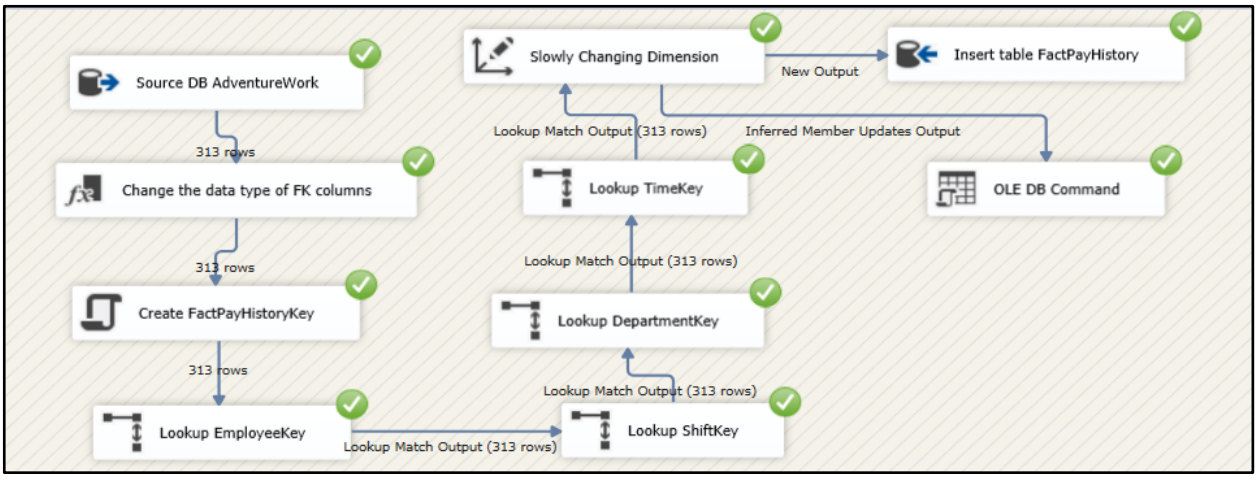
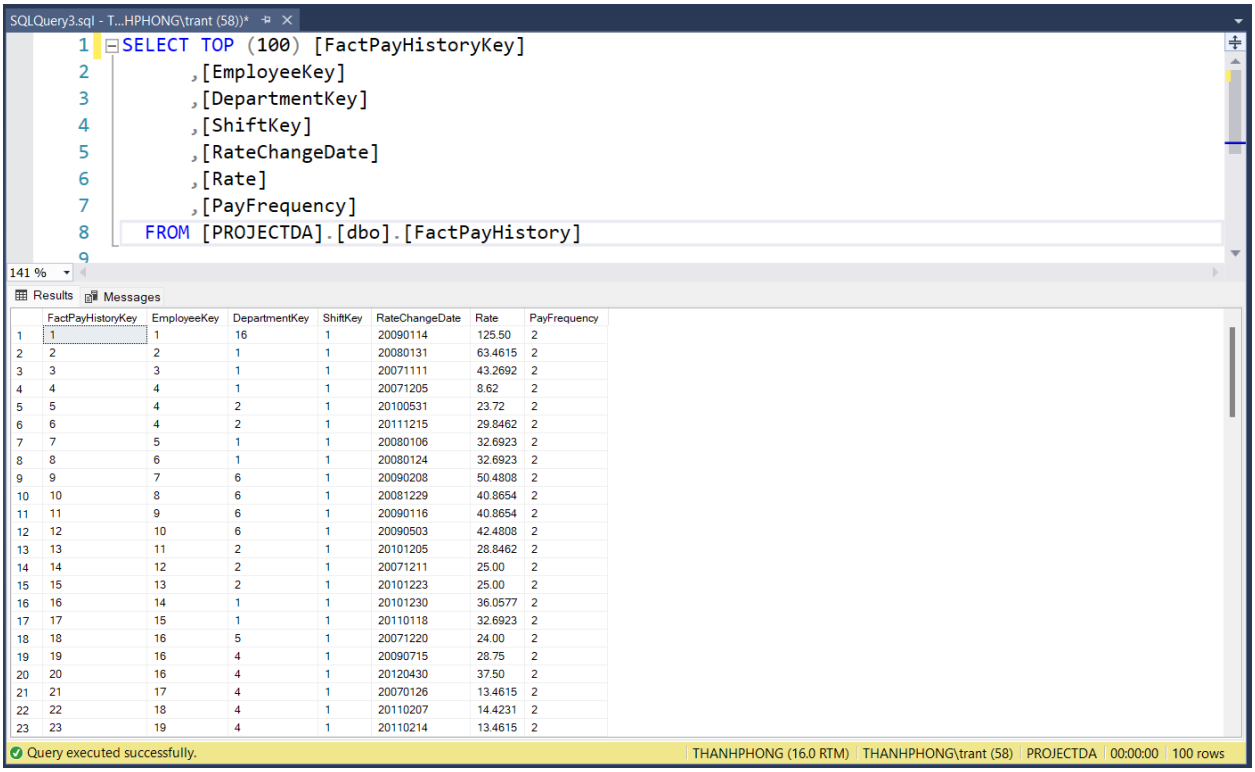


Figure 4.26. SSIS for FactPayHistory



```
SELECT TOP (100) [FactPayHistoryKey]
      ,[EmployeeKey]
      ,[DepartmentKey]
      ,[ShiftKey]
      ,[RateChangeDate]
      ,[Rate]
      ,[PayFrequency]
  FROM [PROJECTDA].[dbo].[FactPayHistory]
```

	FactPayHistoryKey	EmployeeKey	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	2	1	20100531	23.72	2
6	6	4	2	1	20111215	29.8462	2
7	7	5	1	1	20080106	32.6923	2
8	8	6	1	1	20080124	32.6923	2
9	9	7	6	1	20090208	50.4808	2
10	10	8	6	1	20081229	40.8654	2
11	11	9	6	1	20090116	40.8654	2
12	12	10	6	1	20090503	42.4808	2
13	13	11	2	1	20101205	28.8462	2
14	14	12	2	1	20071211	25.00	2
15	15	13	2	1	20101223	25.00	2
16	16	14	1	1	20101230	36.0577	2
17	17	15	1	1	20110118	32.6923	2
18	18	16	5	1	20071220	24.00	2
19	19	16	4	1	20090715	28.75	2
20	20	16	4	1	20120430	37.50	2
21	21	17	4	1	20070126	13.4615	2
22	22	18	4	1	20110207	14.4231	2
23	23	19	4	1	20110214	13.4615	2

Figure 4.27. Top 100 rows of FactPayHistory table

4.3.2. *FactDepartmentHistory*

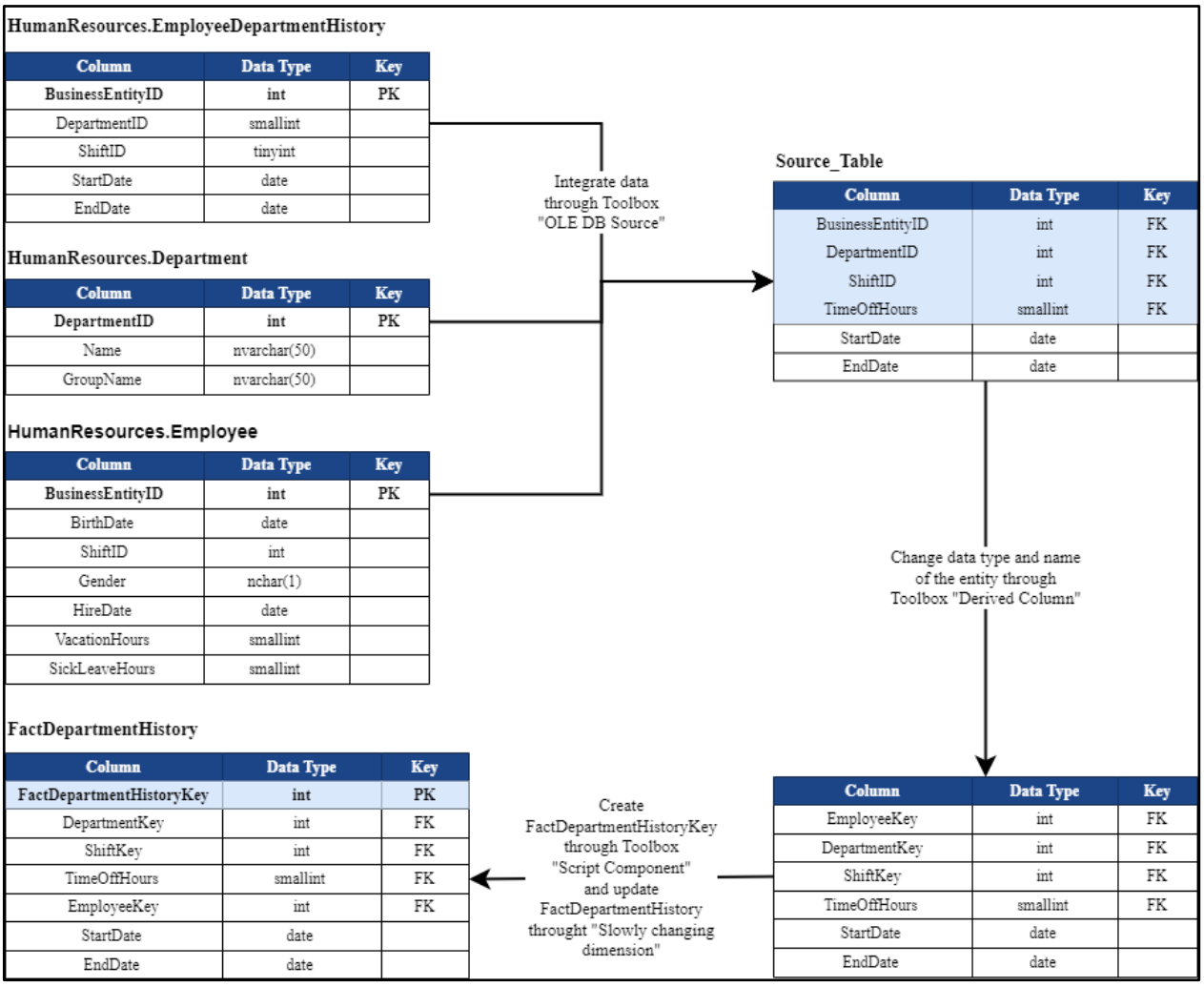


Figure 4.28. The data ETL process of the FactDepartmentHistory table

The ETL process for creating the FactDepartmentHistory table in the AdventureWorks database is formed through the process of extracting data from three main tables in the Human Resources module. Specifically, this process includes collecting information from the EmployeeDepartmentHistory, Employee, and Department tables. These three tables store important information about an employee's department history. By querying these tables, time attributes such as StartDate, EndDate, TimeOffHours and other related information are extracted. Once extracted, data undergoes transformation to standardize formatting, create additional attributes, and handle inconsistencies. The end result is that

the transformed data is loaded into the FactDepartmentHistory table, ensuring that it accurately reflects the time dimensions needed for analysis and reporting in the data warehouse.

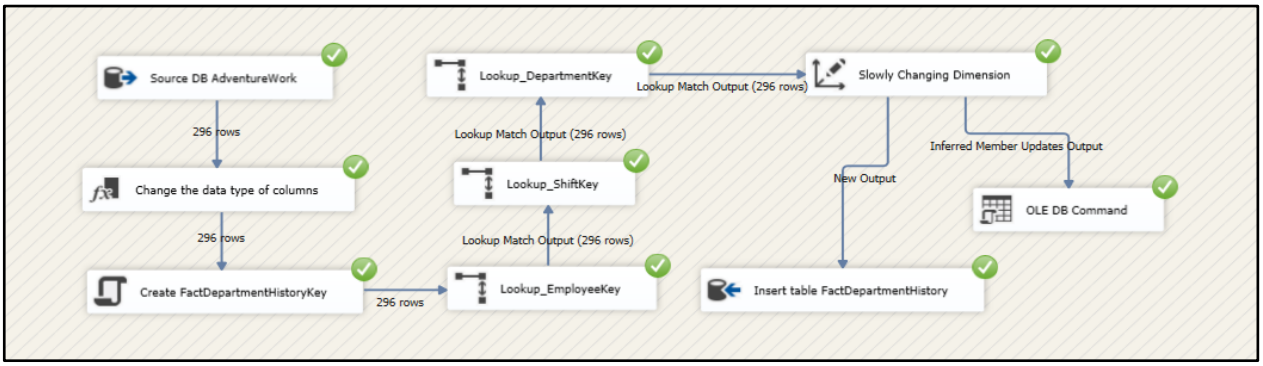
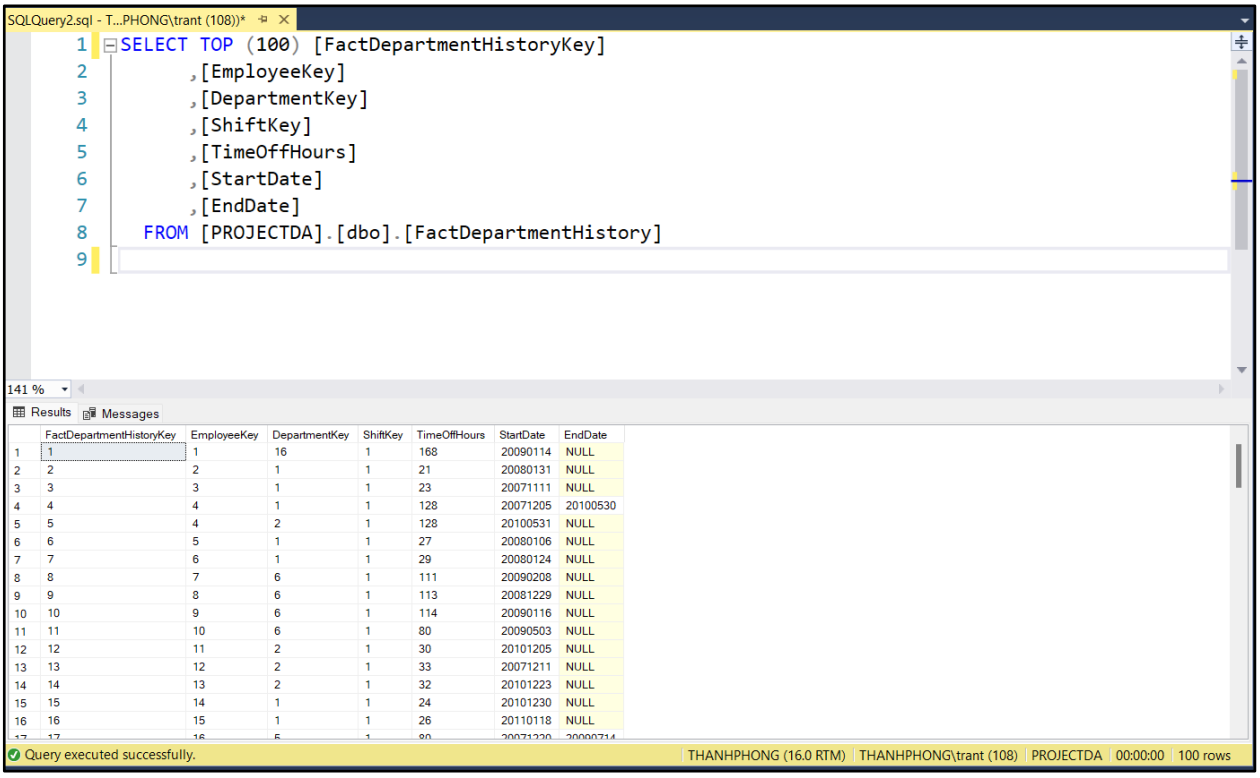


Figure 4.29. SSIS for FactDepartmentHistory



	FactDepartmentHistoryKey	EmployeeKey	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	2	1	128	20100531	NULL
6	6	5	1	1	27	20080106	NULL
7	7	6	1	1	29	20080124	NULL
8	8	7	6	1	111	20090208	NULL
9	9	8	6	1	113	20081229	NULL
10	10	9	6	1	114	20090116	NULL
11	11	10	6	1	80	20090503	NULL
12	12	11	2	1	30	20101205	NULL
13	13	12	2	1	33	20071211	NULL
14	14	13	2	1	32	20101223	NULL
15	15	14	1	1	24	20101230	NULL
16	16	15	1	1	26	20110118	NULL

Figure 4.30. Top 100 rows of FactPayHistory table

4.3.3. DetailPayHistory

FactPayHistory

Column	Data Type	Key
FactPayHistoryKey	int	PK
EmployeeKey	int	FK
DepartmentKey	int	FK
ShiftKey	int	FK
RangeChangeDate	int	FK
Rate	money	
PayFrequency	tinyint	

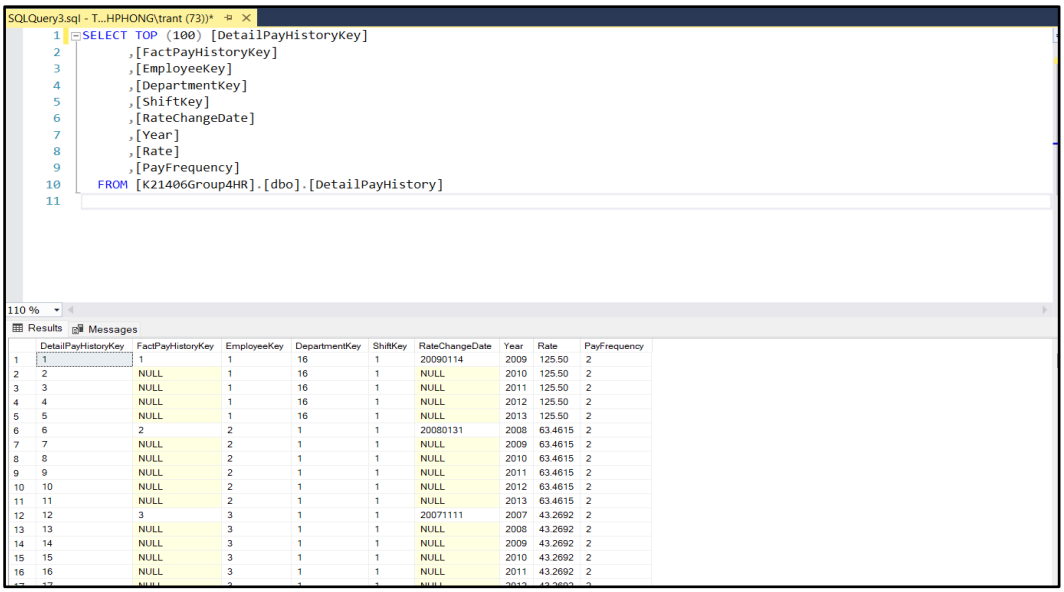
Expanding the database each year →

DetailPayHistory

Column	Data Type	Key
DetailPayHistoryKey	int	PK
FactPayHistoryKey	int	
EmployeeKey	int	FK
DepartmentKey	int	FK
ShiftKey	int	FK
RangeChangeDate	int	
Year	int	FK
Rate	money	
PayFrequency	tinyint	

Figure 4.31. The data ETL process of the DetailPayHistory table

The process to create the DetailPayHistory table in the database is formed through the process of extracting data from the previously created FactPayHistory table using SQL. Specifically, instead of only displaying information about salary changes, DetailPayHistory will display the annual salary of each employee. The end result is that the transformed data is loaded into the DetailPayHistory table, ensuring that it accurately reflects the time dimensions needed for analysis and reporting in the data warehouse.



	DetailPayHistoryKey	FactPayHistoryKey	EmployeeKey	DepartmentKey	ShiftKey	RateChangeDate	Year	Rate	PayFrequency
1	1	1	1	16	1	20090114	2009	125.50	2
2	2	NULL	1	16	1	NULL	2010	125.50	2
3	3	NULL	1	16	1	NULL	2011	125.50	2
4	4	NULL	1	16	1	NULL	2012	125.50	2
5	5	NULL	1	16	1	NULL	2013	125.50	2
6	6	2	2	1	1	20080131	2008	63.4615	2
7	7	NULL	2	1	1	NULL	2009	63.4615	2
8	8	NULL	2	1	1	NULL	2010	63.4615	2
9	9	NULL	2	1	1	NULL	2011	63.4615	2
10	10	NULL	2	1	1	NULL	2012	63.4615	2
11	11	NULL	2	1	1	NULL	2013	63.4615	2
12	12	3	3	1	1	20071111	2007	43.2692	2
13	13	NULL	3	1	1	NULL	2008	43.2692	2
14	14	NULL	3	1	1	NULL	2009	43.2692	2
15	15	NULL	3	1	1	NULL	2010	43.2692	2
16	16	NULL	3	1	1	NULL	2011	43.2692	2

Figure 4.32. Top 100 rows of DetailPayHistory table

CHAPTER 5: MULTI-DIMENSIONAL DATA ANALYSIS

In this chapter, we focus on strategy, goals, meaning of multidimensional database, meaning of multidimensional data analysis in the HR module, and practice. Results of data warehouse analysis by practicing OLAP techniques on SSAS, MDX language, analyzing KPIs in SSAS with requirements as well as built KPIs.

5.1. Multidimensional database

5.1.1. The strategy of multidimensional databases

Analysis strategies for multidimensional databases involve several key steps and techniques to effectively extract insights from complex datasets.

Identify Key HR Metrics: Begin by defining the key performance indicators (KPIs) and metrics relevant to HR management in your organization. These could include employee turnover rates, recruitment effectiveness, training ROI, performance appraisal results, etc.

Understand Dimensional Model: Gain a deep understanding of the dimensional model specific to HR data. Identify dimensions such as employee demographics, job roles, departments, time periods, and measures such as salary, performance ratings, and training hours.

Selection of Analysis Techniques: Choose analysis techniques that align with HR objectives. For instance:

- Slice and dice data to analyze HR metrics by department, location, or time period.
- Drill down into employee performance ratings by job role or department.
- Compare recruitment effectiveness across different sources or channels.
- Conduct trend analysis to identify patterns in employee turnover rates over time.

Query Optimization: Optimize queries to ensure fast retrieval of HR data. Pay attention to indexing, caching, and aggregation techniques to enhance query performance, especially when dealing with large datasets.

Visualization for HR Insights: Utilize visualizations tailored for HR analytics, such as:

- Bar charts to compare employee turnover rates by department.
- Line graphs to track trends in training completion rates over time.

Compliance and Diversity Analysis: Consider specific HR compliance requirements and diversity initiatives in your analysis. For example, analyze workforce diversity metrics such as gender or ethnicity representation across different organizational levels.

Communication of HR Insights: Effectively communicate HR insights to relevant stakeholders using clear and concise visualizations, reports, and presentations. Ensure that insights are actionable and support strategic HR decision-making processes.

5.1.2. The objective of multidimensional databases

Features of multidimensional data models:

- Measures: Measures are numerical data that can be analyzed and compared, such as sales or revenue. They are typically stored in fact tables in a multidimensional data model.
- Dimensions: Dimensions are attributes that describe the measures, such as time, location, or product. They are typically stored in dimension tables in a multidimensional data model.
- Cubes: Cubes are structures that represent the multidimensional relationships between measures and dimensions in a data model. They provide a fast and efficient way to retrieve and analyze data.

- Aggregation: Aggregation is the process of summarizing data across dimensions and levels of detail. This is a key feature of multidimensional data models, as it enables users to quickly analyze data at different levels of granularity.
- Drill-down and roll-up: Drill-down is the process of moving from a higher-level summary of data to a lower level of detail, while roll-up is the opposite process of moving from a lower-level detail to a higher-level summary. These features enable users to explore data in greater detail and gain insights into the underlying patterns.
- Hierarchies: Hierarchies are a way of organizing dimensions into levels of detail. For example, a time dimension might be organized into years, quarters, months, and days. Hierarchies provide a way to navigate the data and perform drill-down and roll-up operations.
- OLAP (Online Analytical Processing): OLAP is a type of multidimensional data model that supports fast and efficient querying of large datasets. OLAP systems are designed to handle complex queries and provide fast response times.

5.1.3. The meaning of multidimensional databases

A multidimensional database is a type of database management system (DBMS) designed to efficiently store, retrieve, and analyze data with multiple dimensions. Unlike traditional relational databases that organize data in tables with rows and columns, multidimensional databases organize data into multidimensional structures known as cubes or hypercubes. It organizes information along multiple dimensions, such as time, geography, or product, allowing for a more nuanced understanding of complex datasets. By structuring data into cubes or hypercubes, multidimensional databases facilitate efficient storage and retrieval, enabling users to perform intricate analytical operations with ease. This multidimensional framework not only streamlines data management but also enhances analytical capabilities, empowering organizations to uncover valuable insights and make informed decisions. With its agility, scalability, and focus on optimizing analytical processes, multidimensional

databases serve as a cornerstone in modern data analysis, driving innovation and enabling organizations to stay competitive in today's dynamic business landscape.

Multidimensional data analysis plays a crucial role in the Human Resources (HR) domain, enabling organizations to gain deeper insights into their workforce and make informed decisions. Here's a breakdown of the significance of multidimensional data analysis in HR:

- Comprehensive Understanding of Workforce Dynamics: Multidimensional data analysis allows HR professionals to analyze various dimensions of employee data such as demographics, skills, performance metrics, and engagement levels. By examining data from multiple perspectives, HR can gain a comprehensive understanding of workforce dynamics, including trends, patterns, and correlations within the data.
- Performance Management Optimization: Through multidimensional analysis, HR can identify key performance indicators (KPIs) and evaluate employee performance across different dimensions such as departments, job roles, and time periods. This enables HR to optimize performance management processes, identify top performers, and address areas for improvement effectively.
- Talent Development and Retention: By analyzing multidimensional data related to employee skills, training, and career progression, HR can tailor talent development programs to individual needs and organizational goals. Moreover, identifying factors contributing to employee turnover through data analysis helps in implementing retention strategies to retain top talent.
- Strategic Workforce Planning: Multidimensional data analysis provides insights into workforce trends, such as hiring patterns, succession planning, and skill gaps. HR can use this information to forecast future workforce needs, develop recruitment strategies, and align talent with organizational objectives effectively.

- Employee Engagement and Satisfaction: Analyzing multidimensional data related to employee feedback, surveys, and sentiment analysis helps HR gauge employee engagement levels and satisfaction. By identifying factors influencing employee morale, HR can implement initiatives to improve workplace culture, employee experience, and overall satisfaction.
- Compliance and Risk Management: Multidimensional data analysis assists HR in monitoring compliance with labor regulations, diversity initiatives, and ethical standards. By analyzing data across various dimensions, HR can identify potential risks, ensure regulatory compliance, and mitigate legal liabilities effectively.
- Cost Optimization and Resource Allocation: Multidimensional data analysis enables HR to assess the cost-effectiveness of various HR initiatives, such as recruitment campaigns, training programs, and benefits packages. By identifying areas of inefficiency or waste, HR can optimize resource allocation and maximize the return on investment in human capital.

In conclusion, multidimensional data analysis empowers HR professionals to make data-driven decisions, optimize HR processes, and align talent strategies with organizational objectives. By leveraging insights from multidimensional data, HR can drive organizational success, enhance employee engagement, and create a competitive advantage in the dynamic business landscape.

5.2. Data analytics with SSAS technology

5.2.1. Cube diagram

To conduct data analysis from the DW data warehouse and generate insightful reports to facilitate decision-making, the thesis utilizes the SSAS tool, which is seamlessly integrated into the BIDS (Business Intelligence Development Studio) toolset provided with SQL Server.

- ***Step 1:*** Create a project in SSAS.

- ***Step 2:*** Create data sources “K21406Group4HR”.

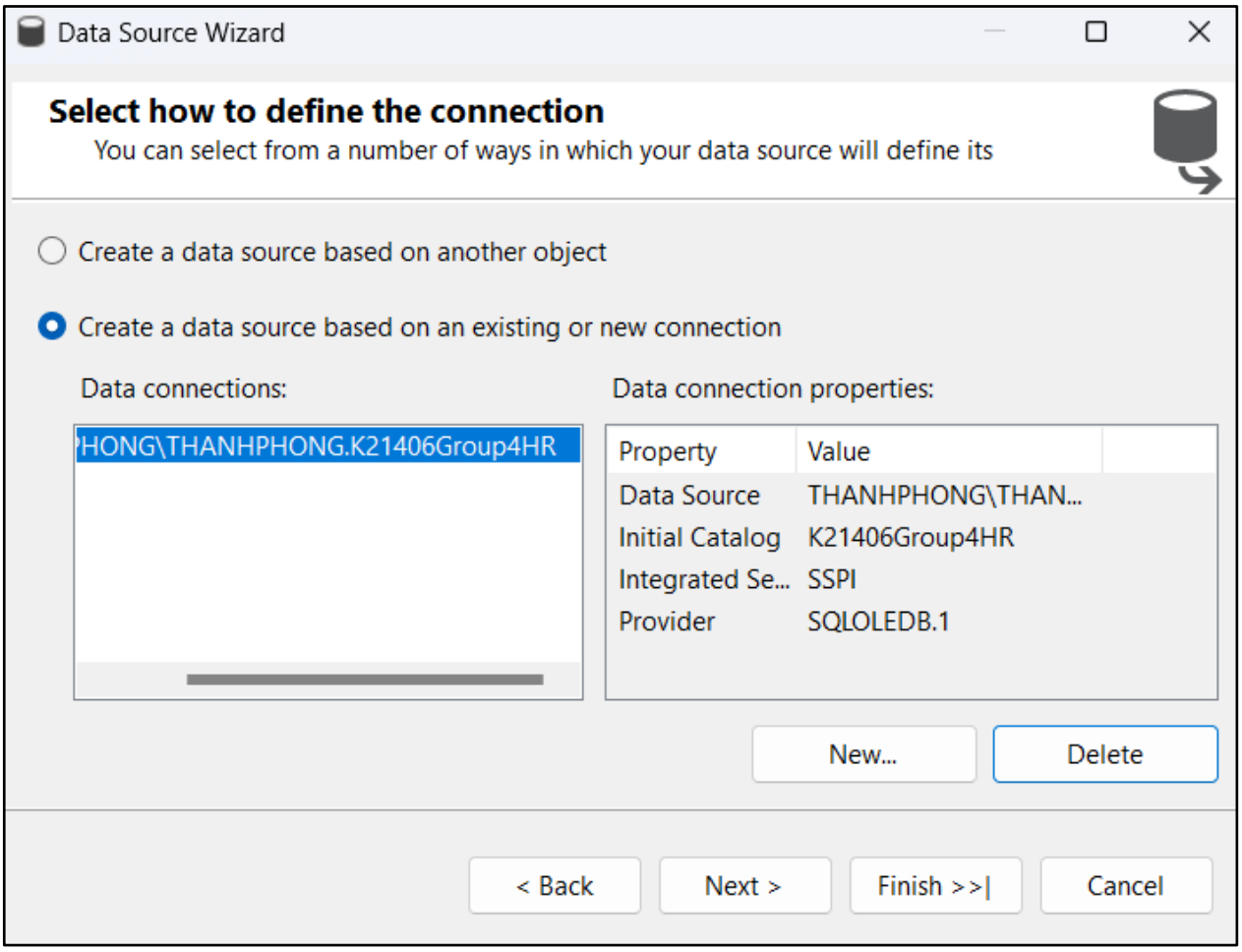


Figure 5.1. Create data sources “K21406_Group4_HR”

- ***Step 3:*** Create Data source Views combined with creating Measures.

- ***Step 4:*** Create Cube combined with creating Dimension.

Right mouse click on “Cubes” → “New Cube” → Select “Use existing tables” and click “Next” → Select the fact table in “Select measure group tables” and click “Next” → Select the dim table in “Select existing dimensions” and click “Next” → Click “Finish”.

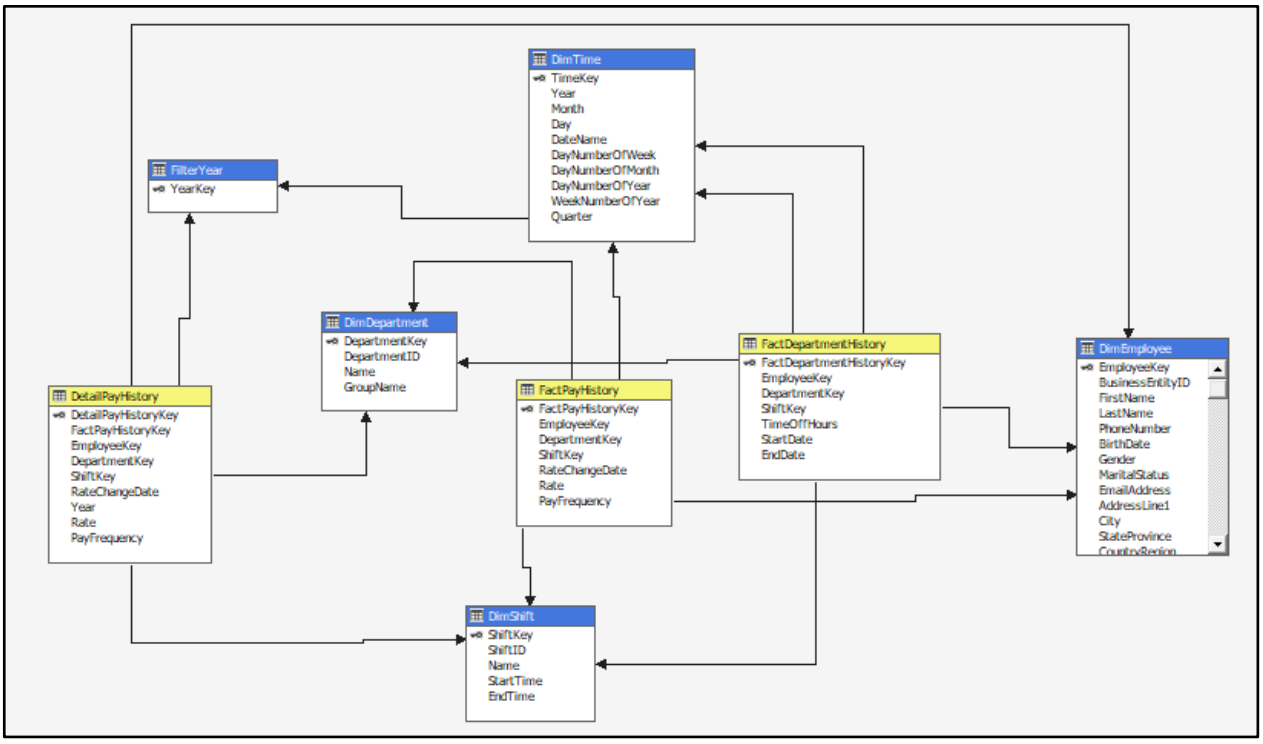


Figure 5.2. Cube model

- ***Step 5:*** Build and Deploy the project.

Right click on your project → Click “Build” → After that, Right click on your project → Click “Deploy”.

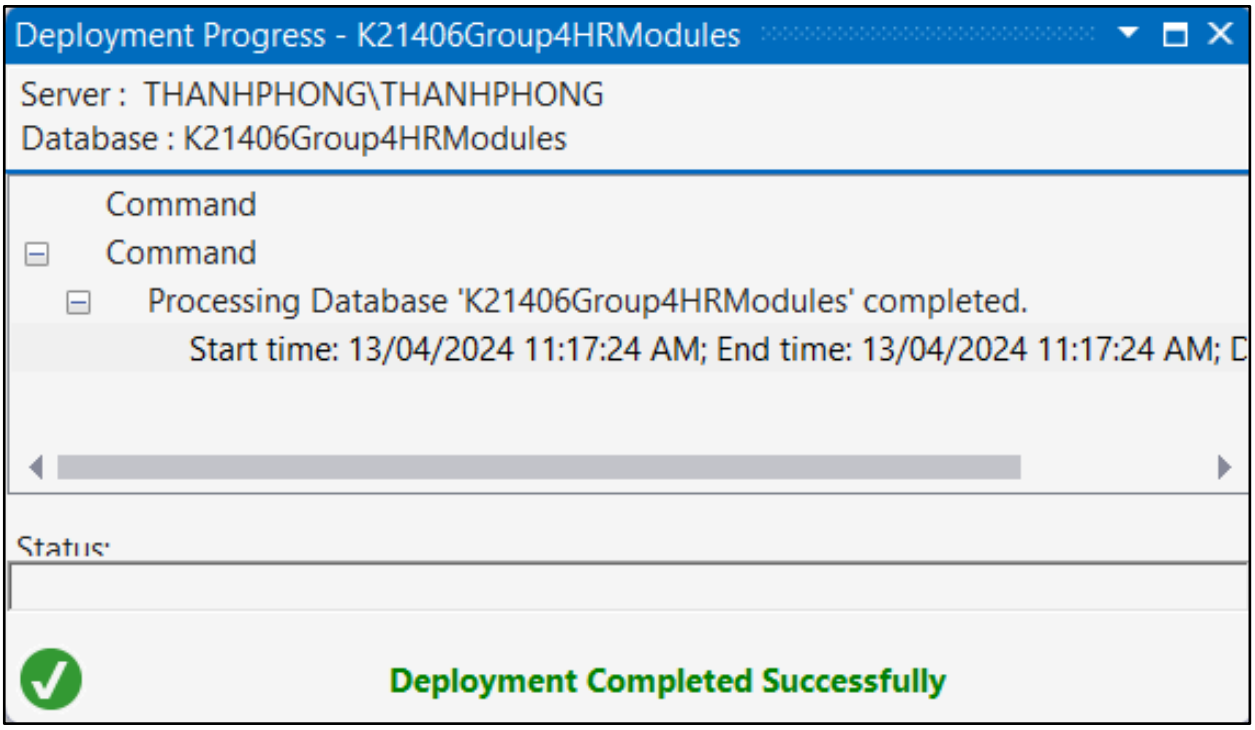


Figure 5.3. Deployment process of project

5.2.2. Analysis with SSAS

Before delving into the analysis of personnel situations, time off hours, and employee transfers between departments, it's essential to understand the context and significance of these factors within an organization. Monitoring and evaluating personnel data provide valuable insights into workforce dynamics, departmental performance, and organizational health. By examining trends and patterns in employee numbers, absenteeism rates, and departmental transfers, HR managers can make informed decisions to optimize workforce efficiency, address staffing needs, and foster a productive work environment. In this analysis, we'll explore the trends and implications of personnel data from 2006 to 2013 to gain a deeper understanding of the organization's dynamics and identify areas for improvement.

5.2.2.1. *Report on the situation of personnel in the department*

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Start Date	Start Date.Year	Equal	{ All, 2006, 2007, 2008, 2009 }	
<Select dimension>				

Year	Group Name	Name	Fact Department History Count
2008	Executive General and Administration	Facilities and Maintenance	1
2008	Executive General and Administration	Finance	2
2008	Executive General and Administration	Human Resources	3
2008	Executive General and Administration	Information Services	3
2008	Inventory Management	Shipping and Receiving	3
2008	Manufacturing	Production	52
2008	Manufacturing	Production Control	2
2008	Quality Assurance	Document Control	1
2008	Quality Assurance	Quality Assurance	2
2008	Research and Development	Engineering	3
2008	Research and Development	Research and Development	1
2008	Sales and Marketing	Marketing	1
2009	Executive General and Administration	Executive	1
2009	Executive General and Administration	Facilities and Maintenance	2
2009	Executive General and Administration	Finance	9
2009	Executive General and Administration	Human Resources	3
2009	Executive General and Administration	Information Services	7
2009	Inventory Management	Purchasing	3
2009	Inventory Management	Shipping and Receiving	3
2009	Manufacturing	Production	104
2009	Manufacturing	Production Control	3
2009	Quality Assurance	Document Control	4
2009	Quality Assurance	Quality Assurance	3
2009	Research and Development	Research and Development	3
2009	Sales and Marketing	Marketing	4

Figure 5.4. Situation of personnel in the department from 2006 to 2009

Dimension	Hierarchy	Operator	Filter Expression
Start Date	Start Date.Year	Equal	{ 2010, 2011, 2012, 2013 }
<Select dimension>			

Year	Group Name	Name	Fact Department History Count
2010	Executive General and Administration	Facilities and Maintenance	4
2010	Inventory Management	Purchasing	8
2010	Manufacturing	Production	22
2010	Quality Assurance	Quality Assurance	1
2010	Research and Development	Engineering	1
2010	Research and Development	Tool Design	3
2011	Manufacturing	Production Control	1
2011	Quality Assurance	Quality Assurance	1
2011	Research and Development	Engineering	1
2011	Sales and Marketing	Marketing	4
2011	Sales and Marketing	Sales	11
2012	Inventory Management	Purchasing	1
2012	Sales and Marketing	Sales	4
2013	Executive General and Administration	Executive	1
2013	Sales and Marketing	Sales	3

Figure 5.5. Situation of personnel in the department from 2010 to 2013

Upon analyzing the departmental personnel data, it becomes apparent that there was a significant upswing in employee numbers across the company's departments in the years

2008 and 2009, indicating a potential expansion in production capacity during this period. Specifically, there was an increase of 52 and 104 employees in 2008 and 2009, respectively. Given the pronounced surge in personnel in 2009 compared to other years, a closer examination of departmental changes during this year is warranted. Notably, the Production department experienced a noteworthy increase of 104 employees, signaling a substantial expansion in production activities. Additionally, departments such as Finance, Information Services, and Marketing bolstered their workforce with highly skilled personnel to support the expansion endeavors.

5.2.2.2. *Report on the situation of time off hours in each department*

Dimension	Hierarchy	Operator	Filter Expression
Start Date	Start Date.Year	Equal	{ 2006, 2007, 2008, 2009 }

Year	Group Name	Time Off Hours
2006	Manufacturing	51
2007	Inventory Management	80
2007	Manufacturing	143
2007	Research and Development	184
2007	Sales and Marketing	83
2008	Executive General and Administration	1028
2008	Inventory Management	494
2008	Manufacturing	5413
2008	Quality Assurance	418
2008	Research and Development	190
2008	Sales and Marketing	87
2009	Executive General and Administration	2529
2009	Inventory Management	806
2009	Manufacturing	9527
2009	Quality Assurance	972
2009	Research and Development	305
2009	Sales and Marketing	345

Figure 5.6. The status of employee absenteeism in each department during the period spanning from 2006 to 2009

Dimension	Hierarchy	Operator	Filter Expression
Start Date	Start Date.Year	Equal	{ 2010, 2011, 2012, 2013 }

Year	Group Name	Time Off Hours
2010	Executive General and Administration	616
2010	Inventory Management	803
2010	Manufacturing	2110
2010	Quality Assurance	147
2010	Research and Development	214
2011	Manufacturing	87
2011	Quality Assurance	93
2011	Research and Development	26
2011	Sales and Marketing	983
2012	Inventory Management	93
2012	Sales and Marketing	275
2013	Executive General and Administration	20
2013	Sales and Marketing	196

Figure 5.7. The status of employee absenteeism in each department during the period spanning from 2010 to 2013

The analysis of absenteeism data reveals a significant increase, primarily concentrated in the years 2008 and 2009, across various departments, notably in manufacturing, executive, and general administration. This surge in absenteeism instances may be linked to organizational changes or relocations within these departments during this period. The data suggests that these departments experienced a higher-than-usual rate of employee time off during this timeframe, which could potentially be attributed to factors such as adjustments in work schedules, changes in job roles, or workplace transitions. Further investigation into the specific causes of absenteeism within each department during this period is warranted

to develop targeted strategies aimed at mitigating absenteeism and improving overall workforce productivity.

5.2.2.3. *List of Job Transfers Between Departments*

Dimension	Hierarchy	Operator	Filter Expression
End Date	End Date.Year	Equal	{ 2013, 2012, 2011, 2006, 2007, 2008, 2009, 2010 }
<Select dimension>			

Year	Month	Group Name	Job Title	Fact Department History Count	Time Off Hours
2009	7	Inventory Management	Marketing Manager	1	80
2010	5	Research and Development	Senior Tool Designer	1	128
2011	7	Sales and Marketing	Purchasing Manager	1	93
2011	8	Manufacturing	Scheduling Assistant	1	87
2012	7	Quality Assurance	Purchasing Manager	1	93
2013	11	Executive General and Administration	Chief Financial Officer	1	20

Figure 5.8. List of Job Transfers Between Departments

Analyzing statistics and reports on employee transfers between departments provides the HR department with valuable insights into the overall situation and performance of each department. By understanding the reasons behind positive or negative changes in departmental transfers, HR can develop appropriate adjustment policies. This analysis enables HR to comprehend employee and departmental dynamics better, facilitating the development of effective strategies for both employees and the company's growth. Despite minimal transfers between departments at Adventure Works Cycles, these movements may signify routine rotations rather than significant organizational changes.

5.2.3. Building the KPIs system

5.2.3.1. KPI 1: Pay rate management

Calculations:

The Average Pay Rate of a department is calculated based on the total Pay Rate divided by the total number of employees in that department. Here, we get the data on the number of employees in each department in the [Measures].[Detail Pay History Count] column because in the salary history data, this data also stores department rotation data.

```
[Measures].[Rate] / [Measures].[Detail Pay History Count]
```

Figure 5.9. Calculation of Average Pay Rate

The KPI of a department's Average Pay Rate is calculated based on the annual average of a department's Average Pay Rate.

```
AVG(
    [Filter Year].[Year Key].Members,
    [Measures].[Rate] / [Measures].[Detail Pay History Count]
)
```

Figure 5.10. Calculation of KPI Average Pay Rate

KPI Value and Goal:

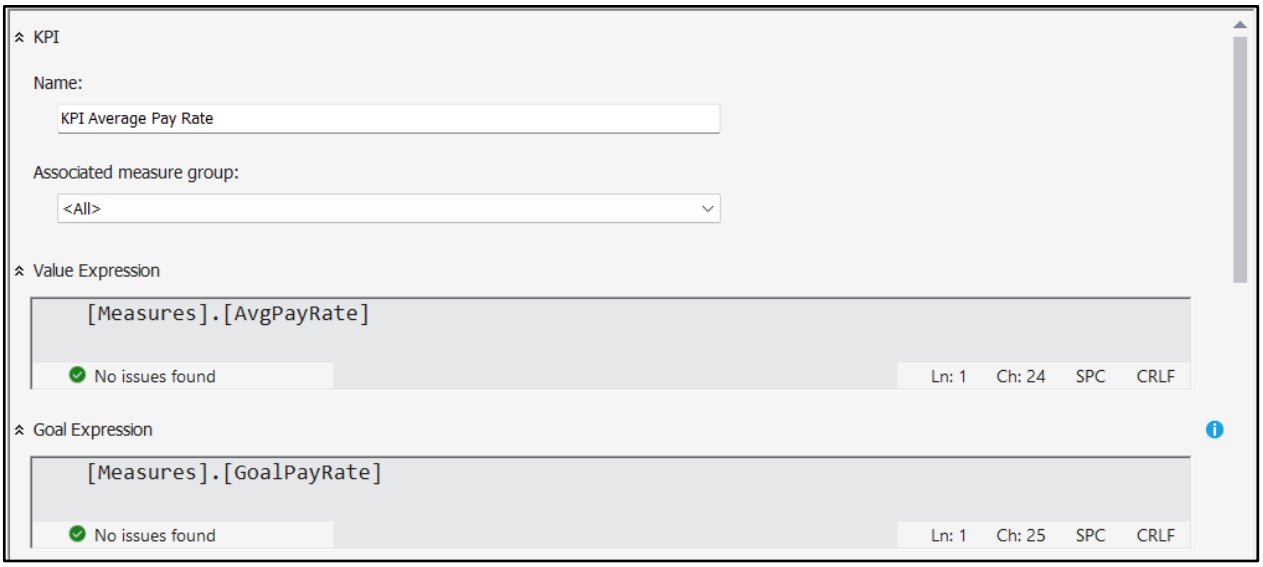


Figure 5.11. KPI Average Pay Rate

The status of the KPI is calculated as a ratio between AvgPayRate and KPIPayRate, then returns the corresponding value based on the value ranges. If the ratio is greater than or equal to 1, the return value is 1. If the ratio is between 0.9 and 1, the return value is 0. In other cases, the return value is -1.

```
CASE
  WHEN [Measures].[AvgPayRate] / [Measures].[GoalPayRate] >= 1 THEN 1
  WHEN [Measures].[AvgPayRate] / [Measures].[GoalPayRate] >= 0.9 AND [Measures].[AvgPayRate] / [Measures].[GoalPayRate] < 1 THEN 0
  ELSE -1
END
```

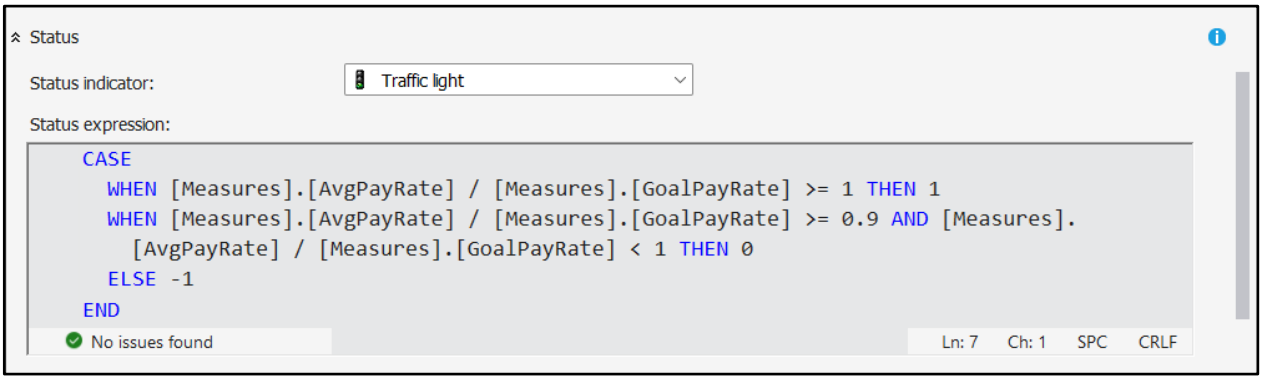


Figure 5.12. Status of KPI Average Pay Rate

Result:

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Filter Year	Filter Year.Year	Equal	{ 2007, 2006, 2008, 2009, 2010, 2012, 2011, 2013 }	
<Select dimension>				

Name	Year	KPI Average Pay Rate Value	KPI Average Pay Rate Goal	KPI Average Pay Rate Status
Document Control	2006	(null)	14.7948666666667	-1
Document Control	2007	(null)	14.7948666666667	-1
Document Control	2008	16.8269	15.0851571428571	1
Document Control	2009	14.38846	14.7368085714286	0
Document Control	2010	14.38846	14.7368085714286	0
Document Control	2011	14.38846	14.7368085714286	0
Document Control	2012	14.38846	14.7368085714286	0
Document Control	2013	14.38846	14.7368085714286	0
Engineering	2006	(null)	37.1865671428571	-1
Engineering	2007	25.9446	35.78132125	-1
Engineering	2008	36.14706	37.05662875	0
Engineering	2009	36.14706	37.05662875	0
Engineering	2010	41.6346	37.74257125	1
Engineering	2011	40.1442166666667	37.5562733333333	1

Figure 5.13. Result of KPI Average Pay Rate

KPI “Average Pay Rate” allows managers to calculate the average salary of employees in the organization. This helps evaluate and compare average salaries across different departments, job positions, levels, or groups of employees. This can help managers understand the distribution and equity of salaries within the organization.

5.2.3.2. KPI 2: Time off hours management

Calculations:

Rate Time Off Hours is calculated by dividing the number of employee's Time Off hours by the employee's actual working hours (total working hours in the year except 2 weekends minus the employee's time off hours).

```
[Measures].[Time Off Hours] / ([Measures].[Employee Key] * 8 * (365 - 2 * 4 * 12) - [Measures].[Time Off Hours])
```

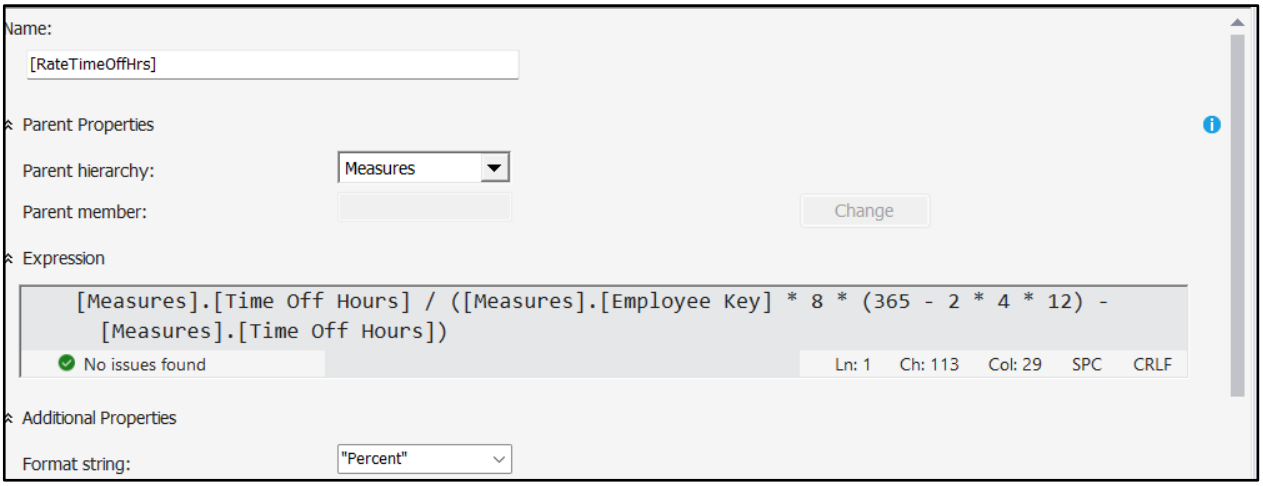


Figure 5.14. Calculation of Rate Time Off Hours

KPI Rate Time Off Hours is calculated using the average time off rate of departments.

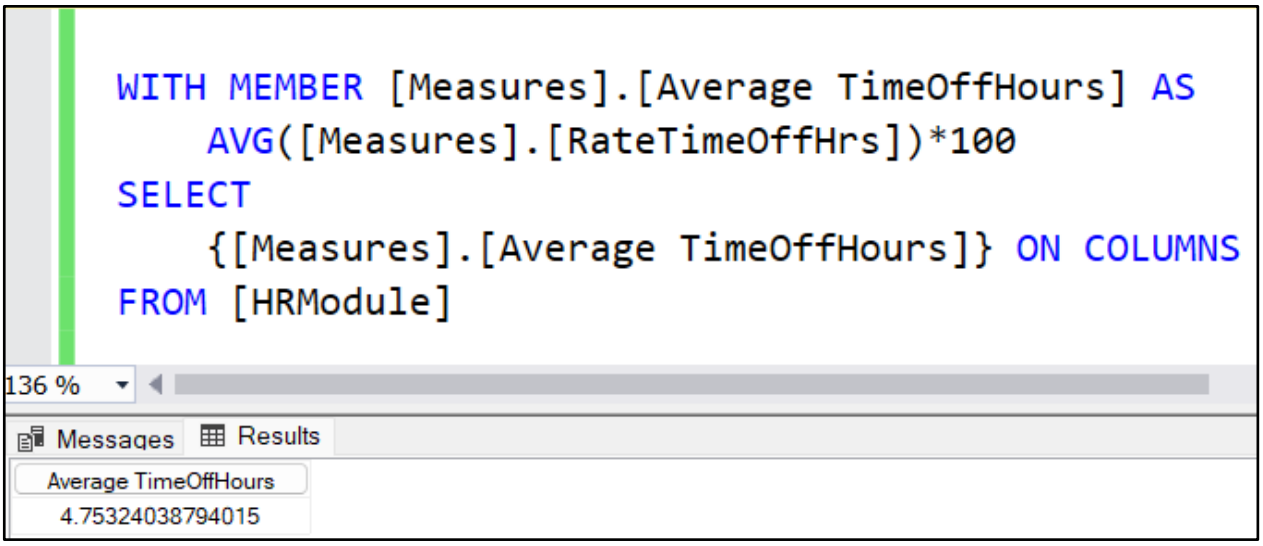
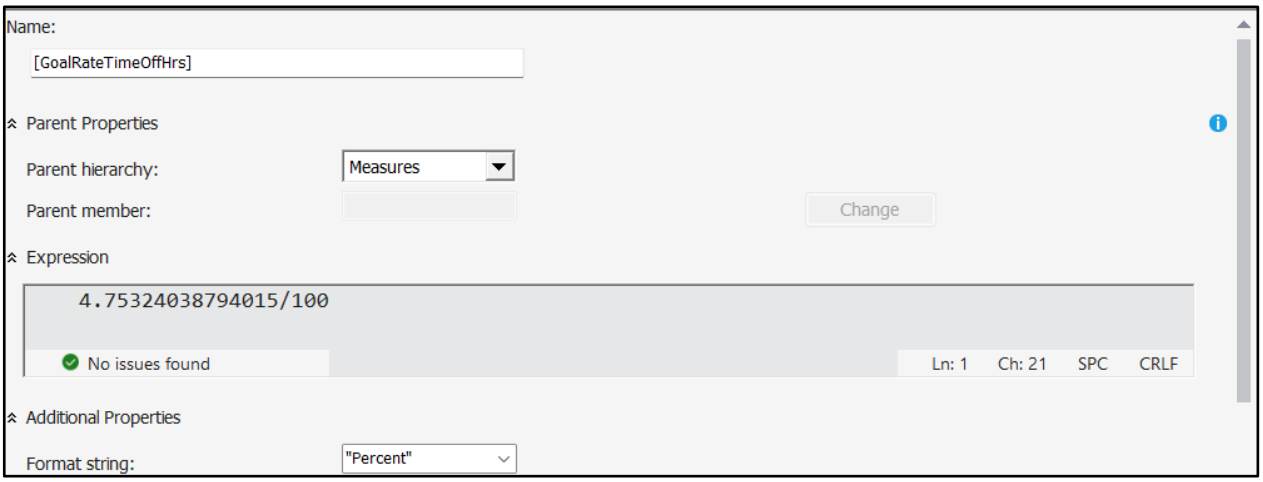


Figure 5.15. Calculation of KPI Rate Time Off Hours

4.75324038794015/100



KPI Value and Goal:

Figure 5.16. KPI Rate Time Off Hours

The return value will be 1 if the number of off hours is smaller than the Goal, and 0 otherwise.

```
IIF([Measures].[RateTimeOffHrs] <= [Measures].[GoalRateTimeOffHrs], 1, 0)
```

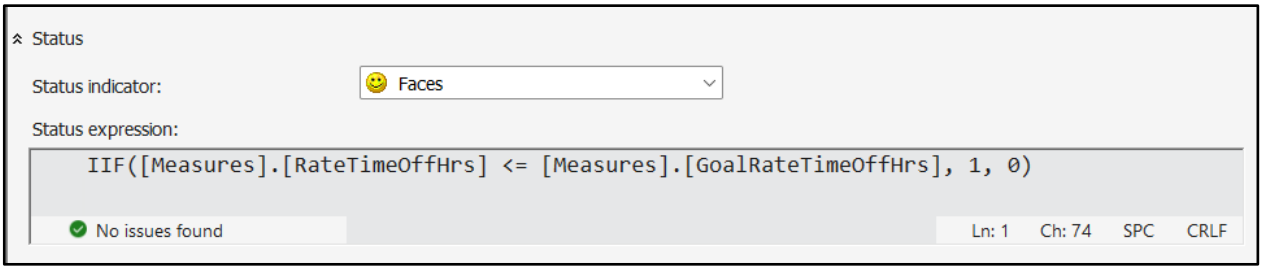


Figure 5.17. Status of KPI Rate Time Off Hours

Result:

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Start Date	Start Date.Year	Equal	{ All }	
<Select dimension>				

Name	KPITimeOffHours Value	KPITimeOffHours Goal	KPITimeOffHours Status
Document Control	0.0670368901229671	0.0475324038794015	0
Engineering	0.0188015690518058	0.0475324038794015	1
Executive	0.0456754130223518	0.0475324038794015	1
Facilities and Maintenance	0.0766921592452291	0.0475324038794015	0
Finance	0.0491977661554827	0.0475324038794015	0
Human Resources	0.0472019464720195	0.0475324038794015	1
Information Services	0.0611439842209073	0.0475324038794015	0
Marketing	0.0418784797869765	0.0475324038794015	1
Production	0.0453452505282588	0.0475324038794015	1
Production Control	0.0426356589147287	0.0475324038794015	1

Figure 5.18. Result of KPI Rate Time Off Hours

Managers can use KPIs on vacation and sick leave rates by department to evaluate and manage the performance and health of departments within the organization.

5.2.3.3. KPI 3: New hire employees management

KPI value:

[Measure].[Fact Department History Count]

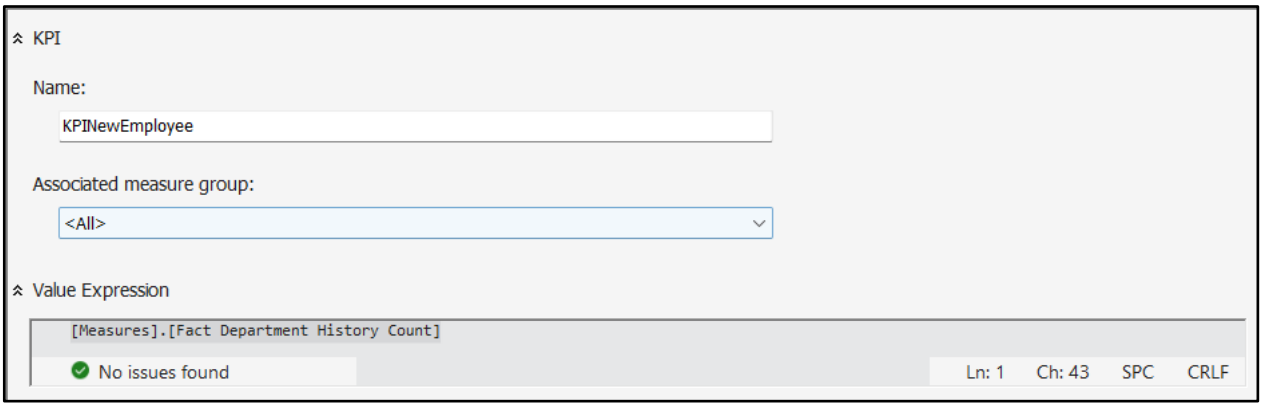


Figure 5.19. KPI value Number of New Hire Employee

KPI Goal:

Each department will have different functions and workloads, so the number of employees in each department will also be different, there will be new employees coming in, and there will also be employees transferred from other departments.

```
CASE
   WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Document Control] THEN 4
   WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Engineering] THEN 8
   WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Executive] THEN 4
   WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Facilities and Maintenance] THEN 9
   WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Finance] THEN 3
   WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Human Resources] THEN 6
   WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Information Services] THEN 10
   WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Marketing] THEN 12
   WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Production] THEN 100
   WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Production Control] THEN 9
   WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Purchasing] THEN 12
   WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Quality Assurance] THEN 12
```

WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Research and Development] THEN 6

WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Sales] THEN 9

WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Shipping and Receiving] THEN 6

WHEN [Dim Department].[Name] is [Dim Department].[Name].&[Tool Design] THEN 6

END

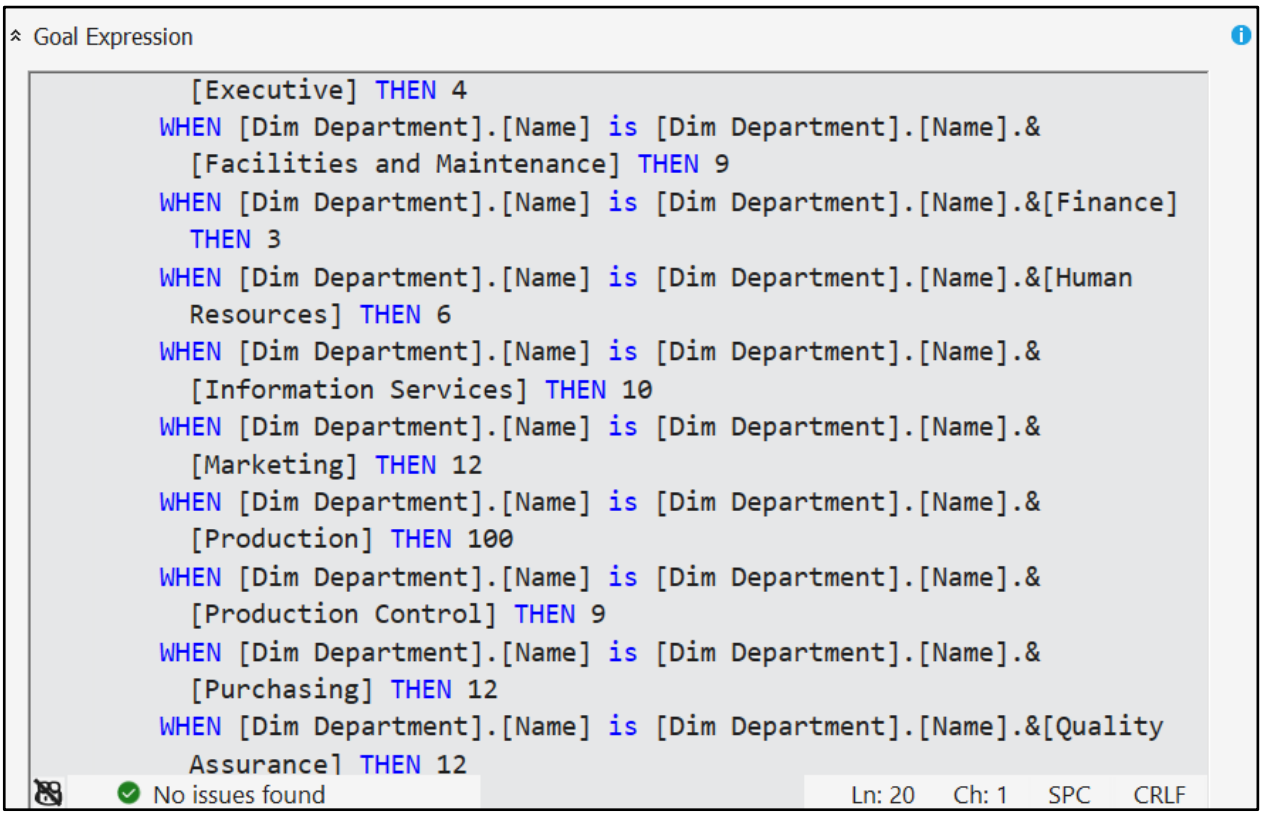


Figure 5.20. KPI Goal Number of New Hire Employee per department

KPI status:

It will evaluate the success of the company's recruitment targets.

```
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
end
```

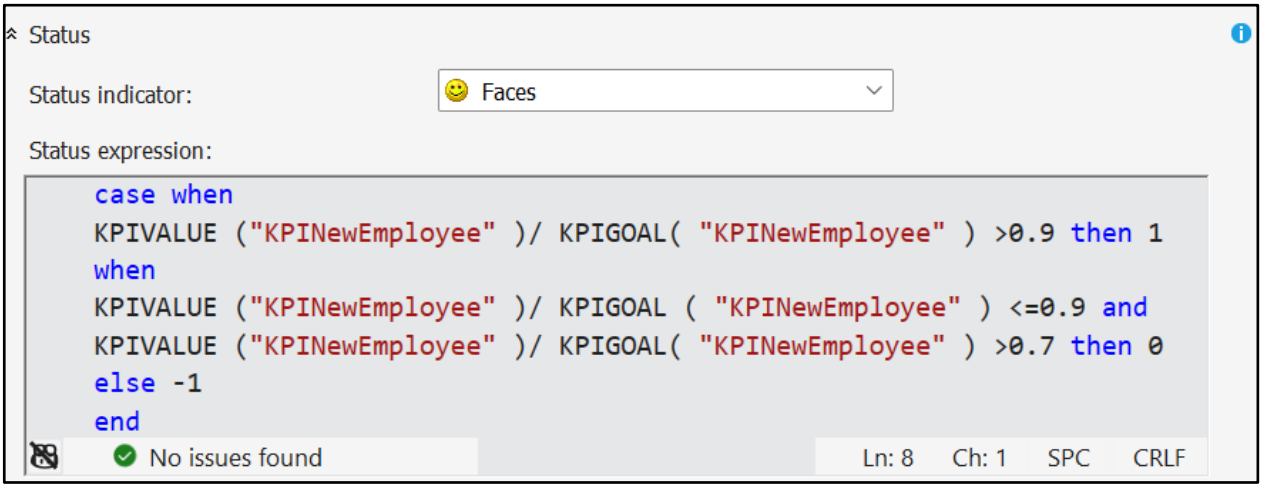


Figure 5.21.KPI Status Number of New Hire Employee

KPI Trend:

It shows how the number of new employees hired this year compared to last year (Increased, decreased or remained the same).

```
CASE
  WHEN Isempty([Start Date].[Year].currentmember.lead(-1)) AND Isempty([Start Date].[Year].currentmember.lead(1)) THEN 0
```

```
  WHEN [Measures].[Fact Department History Count] > ([Measures].[Fact Department History Count], [Start Date].[Year].currentmember.lead(-1)) THEN 1

  WHEN [Measures].[Fact Department History Count] = ([Measures].[Fact Department History Count], [Start Date].[Year].currentmember.lead(-1)) THEN 0

END
```

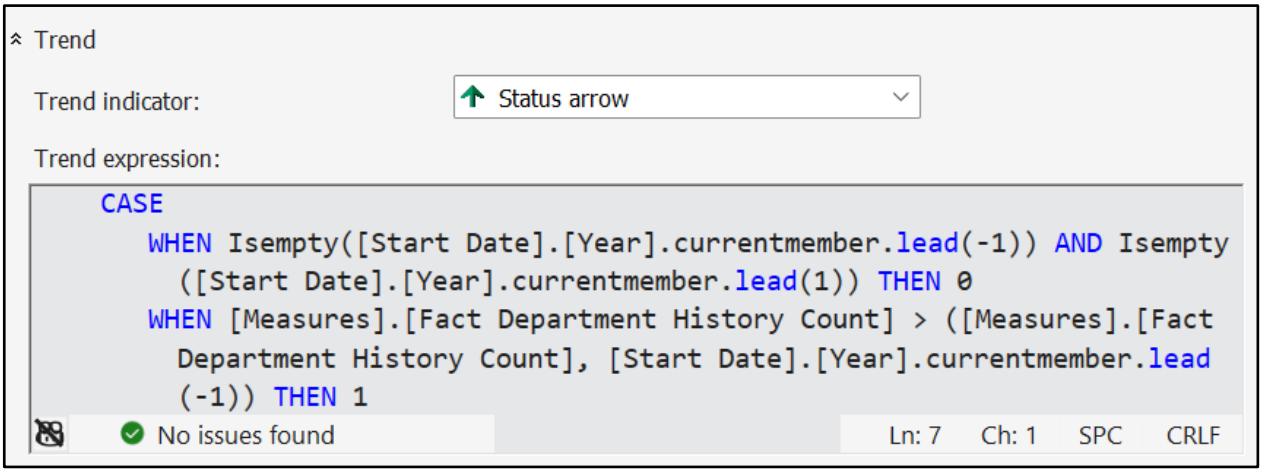


Figure 5.22 . KPI Trend Number of New Hire Employee

Result:

Dimension	Hierarchy	Operator	Filter Expression	Param...
Start Date	Start Date.Year	Equal	{ All }	
<Select dimension>				

Name	KPINewEmployee Val...	KPINewEmployee Goal	KPINewEmploye...	KPINewEmployee Tre...
Document Control	5	4	1	0
Engineering	7	8	0	0
Executive	2	4	-1	0
Facilities and Mainten...	7	9	0	0
Finance	11	3	1	0
Human Resources	6	6	1	0
Information Services	10	10	1	0
Marketing	10	12	0	0
Production	180	100	1	0
Production Control	6	9	-1	0
Purchasing	13	12	1	0
Quality Assurance	7	12	-1	0

Figure 5.23. Result KPI Number of New Hire Employees in Visual Studio

	Fact Department History Count	KPINewEmployee Goal	KPINewEmployee Status	KPINewEmployee Trend
⊞2006			◆	➔
⊞2007			◆	➔
⊞2008			◆	➔
⊟2009				
Executive General and Administration			◆	➔
Inventory Management		1	●	➔
Manufacturing			◆	➔
Quality Assurance			◆	➔
Research and Development			◆	➔
Sales and Marketing			◆	➔
⊟2010				
Executive General and Administration			◆	➔
Inventory Management			◆	➔
Manufacturing			◆	➔
Quality Assurance			◆	➔
Research and Development		1	●	➔
Sales and Marketing			◆	➔
⊟2011				
Executive General and Administration			◆	➔
Inventory Management			◆	➔
Manufacturing		1	●	➔
Quality Assurance			◆	➔
Research and Development			◆	➔
Sales and Marketing		1	●	➔
⊟2012				
Executive General and Administration			◆	➔
Inventory Management			◆	➔
Manufacturing			◆	➔

Figure 5.24. Result KPI Number of New Hire Employees in Excel

This KPI can be applied by managers in evaluating the effectiveness of recruiting new personnel in the business. At the same time, combine with employee performance evaluation to come up with an effective recruitment plan for the following years.

CHAPTER 6: VISUALIZATION AND DISCUSSION

In this chapter, visualize information and insights that are analyzed on smart reports into graphical forms such as graphs, charts, or using 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. Introduce the structure of the Reporting System

The system of reports and dashboards is the result of data that has been transformed into charts that businesses can easily understand and provide useful information for them.

Our team used Power BI to analyze and present the following information:

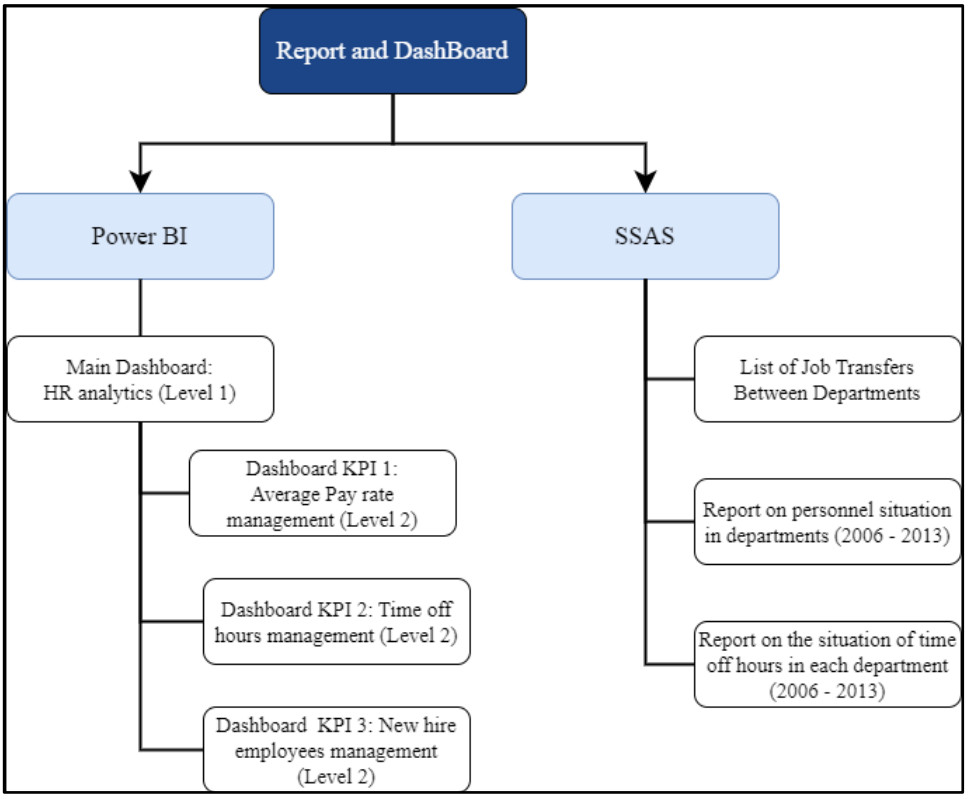


Figure 6.1. Structure of the Report and dashboard

6.2. Data analysis with Power BI

6.2.1. HR analytics

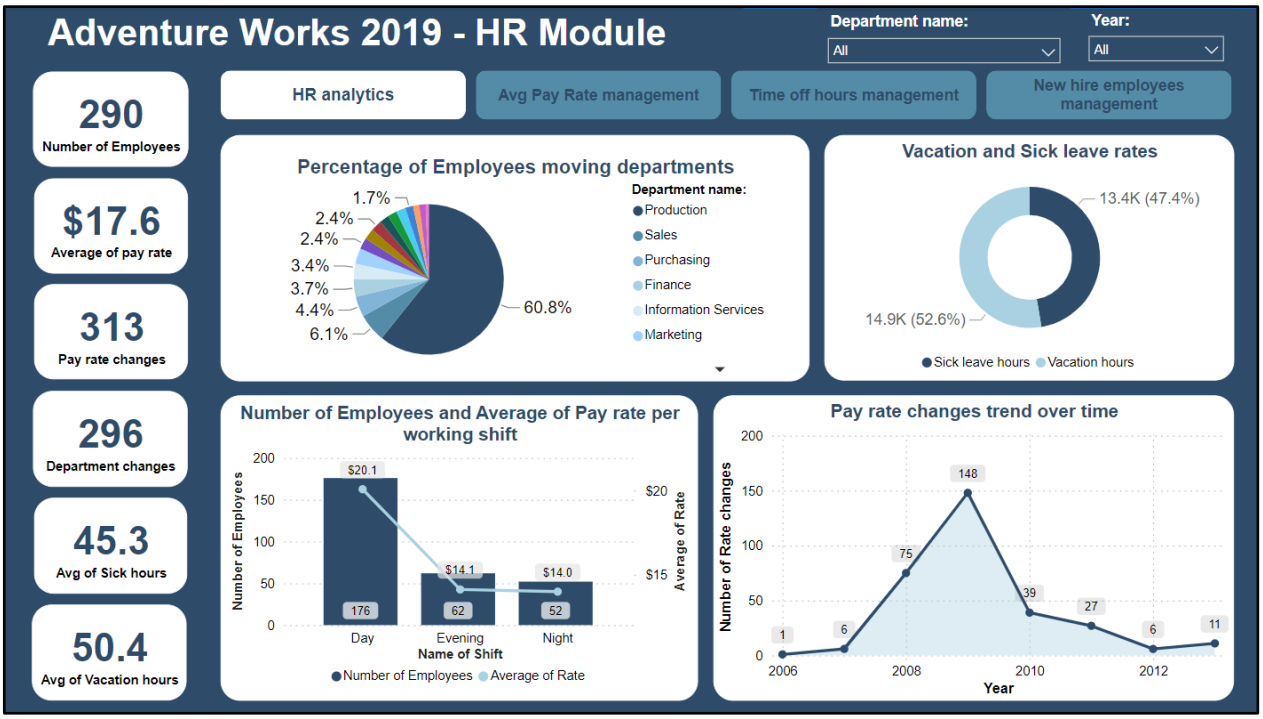


Figure 6.2. HR Analytics Dashboard

6.2.1.1. Purpose

The HR Analytics Dashboard serves as a comprehensive tool designed to provide insights into workforce performance and trends within the organization. Its primary objective is to empower HR professionals and organizational leaders with actionable intelligence for informed decision-making.

6.2.1.2. Description

At the top of the report, users are provided with filters and options to customize their data view. These filters include:

- Department Name: Users can select a specific department to focus their analysis on. This filter enables users to drill down into the data for a particular department, allowing for more targeted insights and decision-making.

- Year: Users can select a specific year to analyze HR metrics and trends over time. This filter allows users to track changes and patterns in workforce data across different years, facilitating historical analysis and forecasting.

In addition to the main HR Analytics Dashboard, users have access to three specific dashboards focused on key performance indicators (KPIs). These additional dashboards provide targeted insights into specific aspects of HR management, enabling HR professionals and organizational leaders to effectively monitor and optimize key performance indicators related to salary management, time off hours management, and new hire employees management.

The left section of the dashboard contains various visualizations presented as slices, providing quick insights into key HR metrics:

- Total Attending Employees Slice: This segment features the total count of currently active employees within the organization, offering valuable insights into the scale of the workforce.

- Average Rate Slice: The Average Rate Slice is typically represented as a numerical value or a line chart showing changes in the average pay rate over time. This visualization provides a clear indication of trends and patterns in compensation levels across different time periods.

- Rate Changes Slice: The Rate Changes Slice focuses on capturing variations or fluctuations in pay rates over specific time intervals or within certain employee groups. It provides insights into the dynamics of pay rate changes and helps identify trends or anomalies that may require attention.

- Departmental Changes Slice: this component delineates the number of employee transfers between departments, aiding management in discerning patterns and trends in workforce mobility.

- Average Sick Leave Hours Slice: this segment provides a visual representation of the average number of sick leave hours utilized per employee, offering insights into employee health and productivity.

- Average Vacation Hours Slice: this section visually showcases the average number of vacation hours available per employee, assisting management in assessing vacation utilization patterns within the organization.

The middle section of the dashboard contains additional visualizations to provide deeper insights:

- Percentage of Employees Moving Departments Pie Chart: This section features a pie chart illustrating the percentage of employees who have transitioned between departments, shedding light on the organization's flexibility and workforce mobility.

- Distribution of Vacation and Sick Leave Hours Donut Chart: Presented as a donut chart, this segment visualizes the distribution of vacation and sick leave hours relative to each other, enabling assessment of leave utilization trends.

- The "Number of Employees and Average of Rate per working shift" chart: illustrates the relationship between the average of rate and the number of employees allocated to each shift. The chart is divided into multiple clusters, with each cluster representing a different shift. The x-axis represents the shifts, while the y-axis is divided into two scales: one for the average rate (line) and the other for the number of employees (columns). This visualization enables a quick comparison between the staffing levels and pay rate distribution across different shifts within the organization.

- Rate changes trend over time Line Chart: Displayed as a line chart, this section depicts the historical trends in rate changes over time, facilitating the identification and prediction of future developments.

6.2.1.3. Data Analysis

- Percentage of Employees Moving Departments Pie Chart:

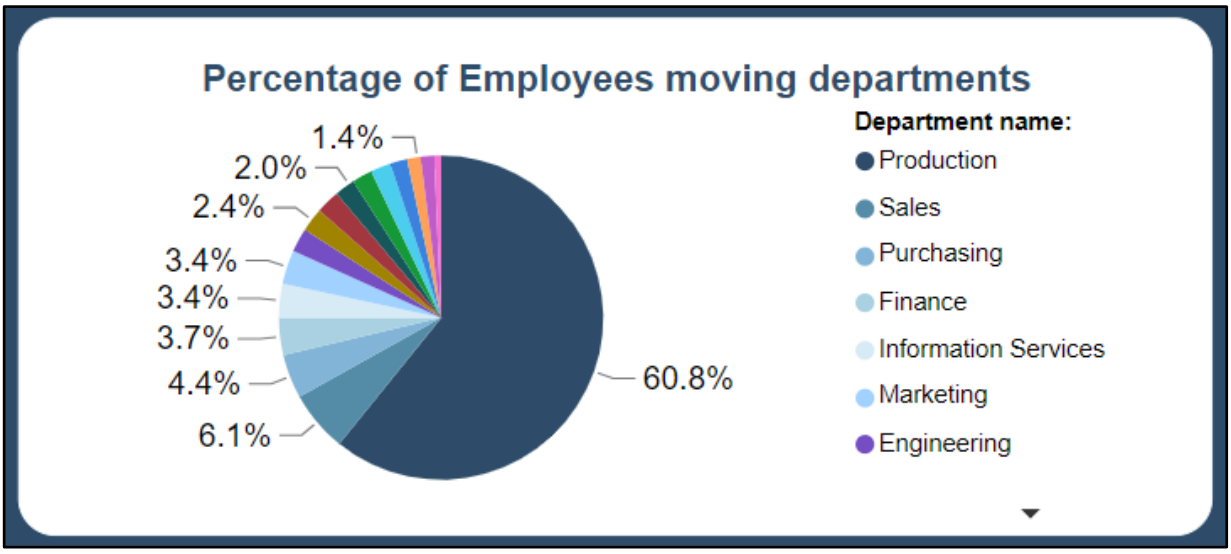


Figure 6.3. Pie Chart of Percentage of Employees Moving Departments

The analysis of the pie chart indicates that the Production department comprises the largest percentage (60.8%) of departmental transfers compared to other departments. This significant proportion suggests a high level of workforce mobility within the Production department, potentially driven by factors such as project requirements, skill development opportunities, or organizational restructuring initiatives.

- Distribution of Vacation and Sick Leave Hours Donut Chart:

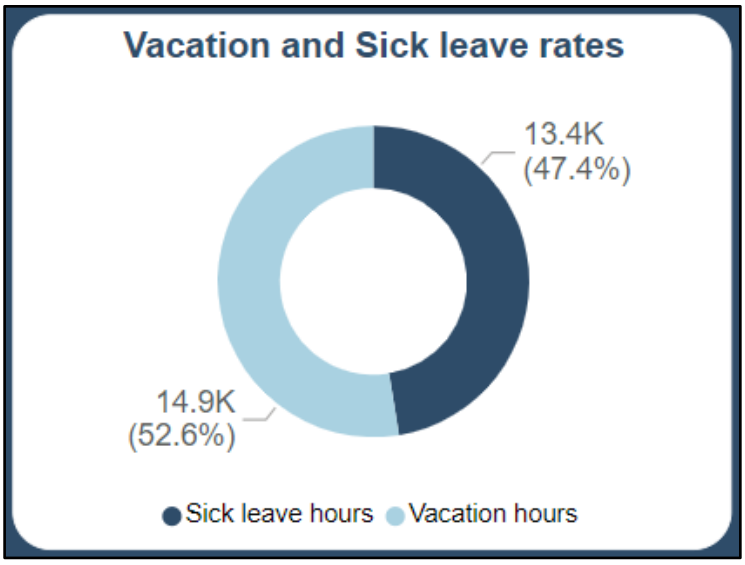


Figure 6.4. Donut Chart of Distribution of Vacation and Sick Leave Hours

The donut chart provides insights into the distribution of vacation and sick leave hours within the organization. With sick leave hours accounting for 52.7% of the total leave hours, and vacation hours representing 47.3%, it suggests that employees are utilizing more sick leave compared to vacation time. This discrepancy may indicate underlying health-related issues, workplace stress factors, or inadequate work-life balance, warranting further investigation and potential interventions from HR and management.

- The “Number of Employees and Average of Rate per working shift” chart:

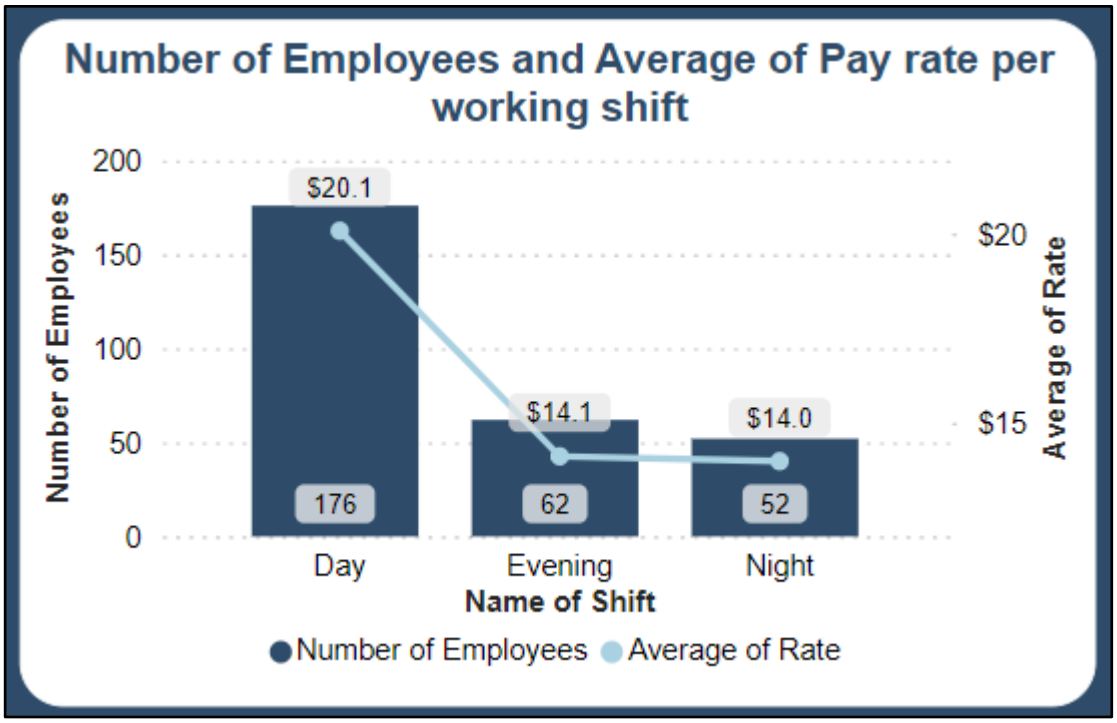


Figure 6.5. Chart of Number of Employees and Average of Rate per working shift

+ The average rate for the Day Shift is significantly higher than that of the Evening and Night Shifts, standing at $20,100. This suggests that the Day Shift may involve roles with higher skill levels or responsibilities, warranting higher compensation.

+ Both the Evening and Night Shifts have the same average rate of $14,000, indicating comparable compensation levels despite the difference in working hours.

+ The Day Shift has the highest number of employees, with 176 staff members allocated to this shift. This indicates a higher demand for labor during the day, possibly due to peak operational hours or business requirements.

+ The Evening Shift follows with 62 employees, indicating a smaller workforce compared to the Day Shift but still substantial enough to meet operational needs during evening hours.

    + The Night Shift has the fewest number of employees, with 52 staff members assigned to this shift. This suggests that there is a lower demand for labor during the night, which is consistent with typical working patterns and business operations.

    + The Day Shift, having the highest average rate, also has the highest number of employees. This correlation suggests that higher compensation may attract more employees to the Day Shift, possibly due to better working conditions or financial incentives.

    + Both the Evening and Night Shifts, with lower average rates, have fewer employees compared to the Day Shift. This indicates that lower compensation levels may influence staffing decisions, resulting in a smaller workforce allocated to these shifts.

- Rate changes trend over time Line Chart:

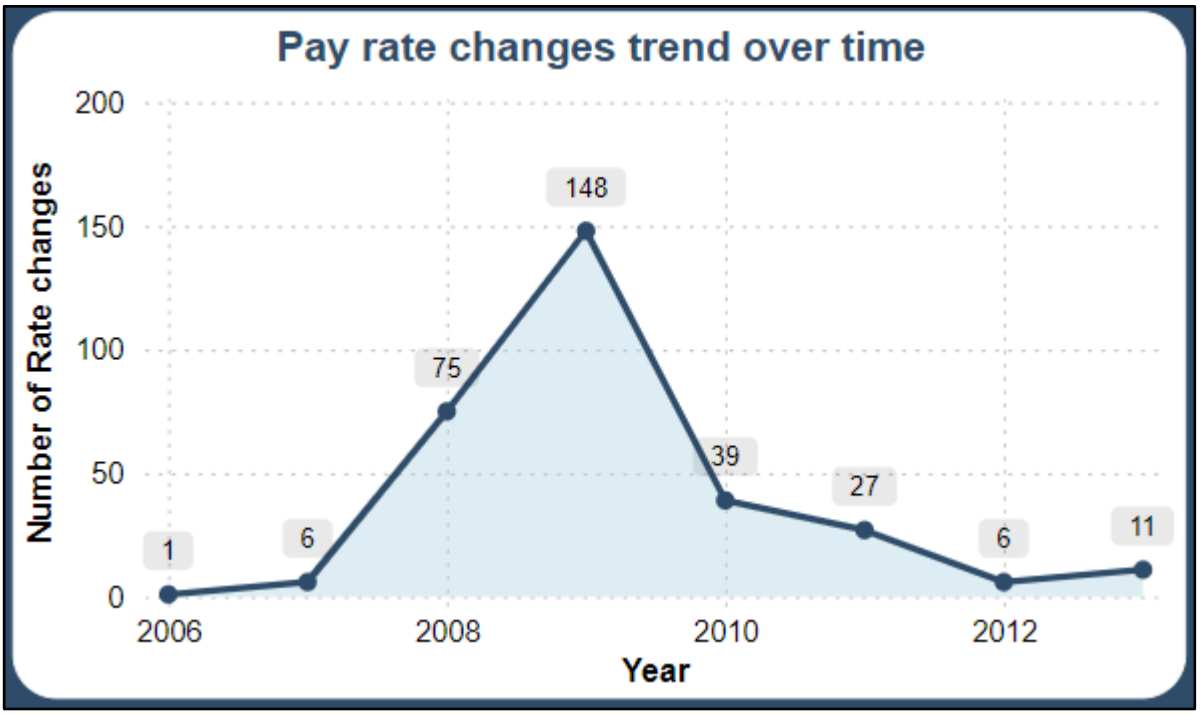


Figure 6.6. Line Chart of Rate changes trend over time

The line chart depicting historical trends in rate changes reveals a notable peak in pay rate adjustments during the year 2009, with approximately 148 instances recorded. This peak suggests a period of significant flux in compensation management within the organization, potentially influenced by external factors such as economic conditions, industry trends, or internal changes in compensation policies. Understanding the drivers behind these fluctuations can provide valuable insights for HR professionals and organizational leaders to develop strategies for salary optimization, talent retention, and competitive positioning in the job market.

6.2.2. KPI 1: Average Pay rate management

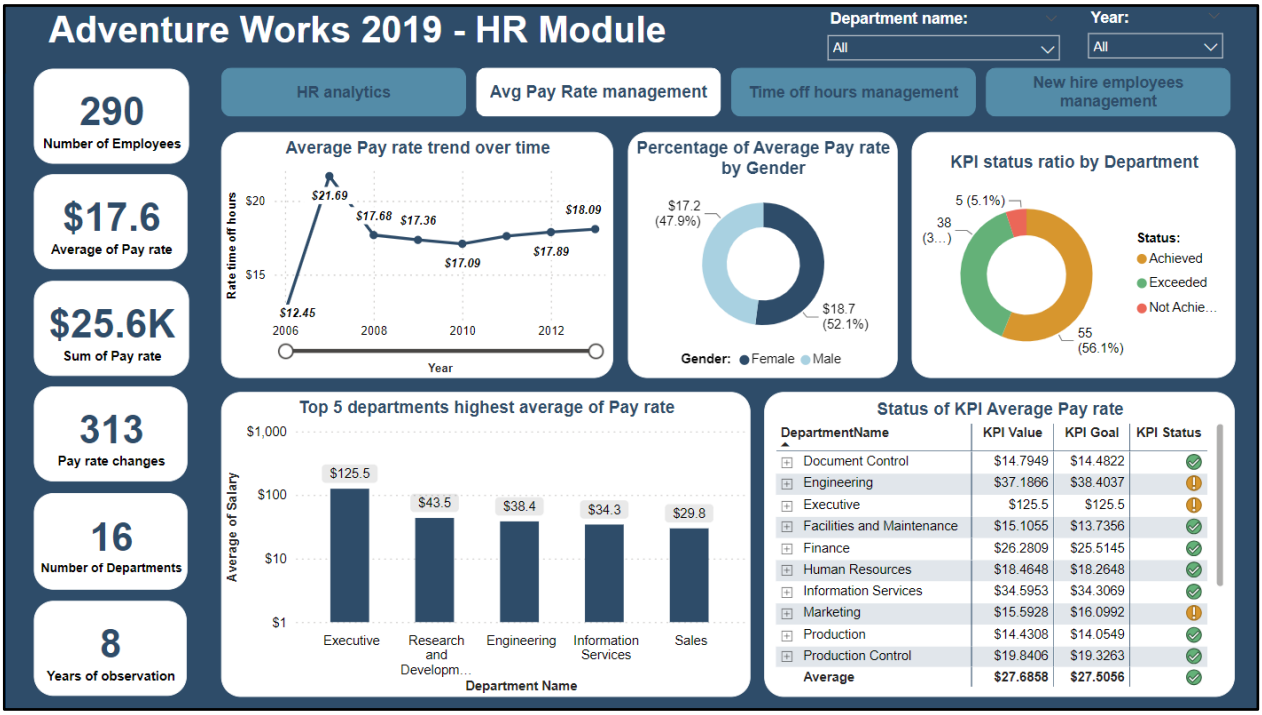


Figure 6.7. Dashboard of Average Pay rate management

6.2.2.1. Purpose

This dashboard aims to analyze the average salary of employees by each department.

6.2.2.2. Description

This is a table containing information about the average salary of 16 departments from 2006 to 2013 of Adventure Works Cycles company.

On the dashboard, users can view the average salary of employees at the company, the total salary the company has paid, the number of salary changes, the average salary ratio by gender or by marital status . Based on this dashboard, you can also know which departments pay the most in the company as well as know the annual salary increase and decrease trend of each department through visualized charts and tables. In particular, users can filter information by department or year depending on their needs.

This provides a comprehensive view of the company's salary situation over the years and from there, management can make reasonable strategic decisions about human resource

and financial management. By analyzing information from this dashboard, companies can optimize compensation strategies, ensure fairness and transparency in the human resources management process, and at the same time create a positive work environment. positive and fair work for all employees.

6.2.2.3. Data Analysis

As mentioned above, by customizing filters per department and year, we can conduct a more detailed analysis of each department's performance in each year. However, within this scope, the team will focus on a comprehensive analysis of the overall performance of the Adventure Works Cycles company across all years in the data set. This will help us better understand the trends and fluctuations of important indicators, thereby making strategic decisions and improving the company's performance more comprehensively.

- Status of KPI average pay rate by department:

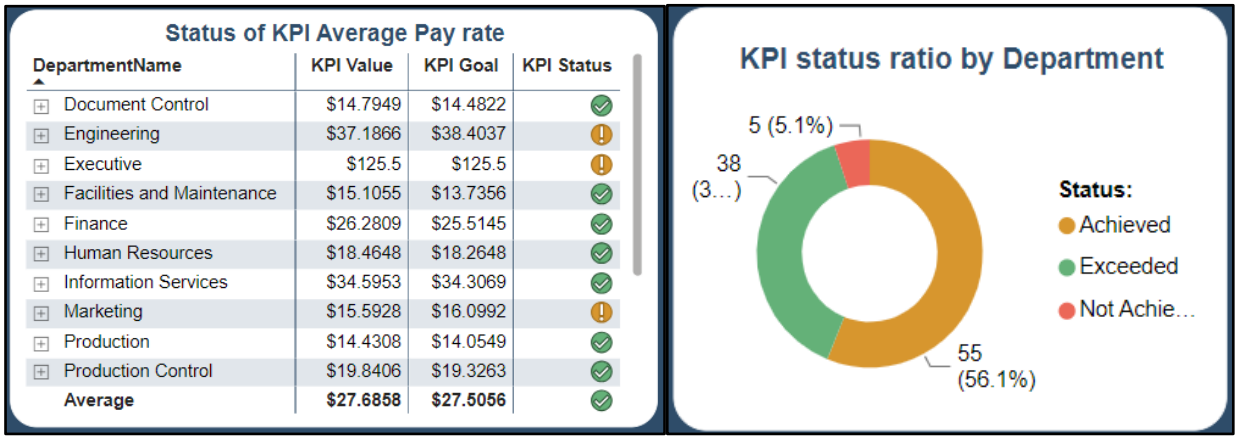


Figure 6.8. Status of KPI average pay rate by department

This dashboard serves as the primary tool for monitoring the Key Performance Indicator (KPI) of Average Pay Rate across departments. Through this chart, we can identify which departments are below, meeting, or exceeding the set KPI. The KPI Value column represents the current value of the KPI, KPI Goal indicates the target value that needs to be achieved, and KPI Status signifies whether the KPI has been met. Additionally, a pie chart

is utilized to visualize the frequency of departments achieving, not achieving, or surpassing the KPI, along with their respective percentages.

Taking the Engineering department as an example, we observe that it did not meet the KPI in 2007. However, starting from 2008 and 2009, it transitioned to meeting the KPI, and from 2010 onwards, it consistently exceeded the KPI. This trend suggests that the Engineering department underwent significant improvements in its performance metrics over the years. Possible factors contributing to this could include strategic realignment, enhanced operational efficiency, or targeted interventions to address previous shortcomings. The department's ability to not only meet but surpass the KPI indicates a commendable level of effectiveness and adaptability within its operations. This underscores the importance of continuous performance monitoring and adjustment to drive organizational success. Additionally, lessons learned from the Engineering department's journey can inform strategies for other departments striving to improve their KPI performance.

- Trend assessment line chart:

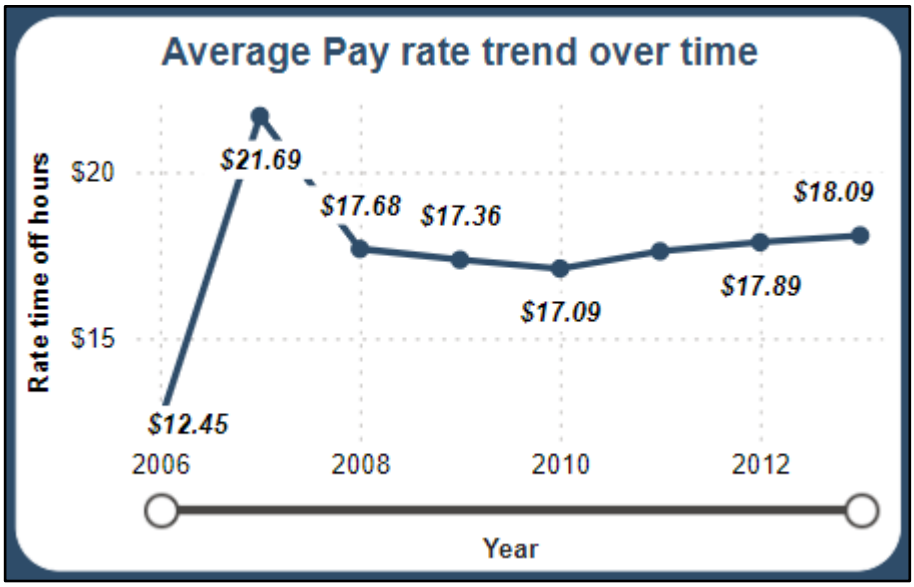


Figure 6.9. Line chart of trend KPI 1

Use collected data to analyze trends in average pay rate over time. Users can compare average pay rate by year to get insights on trends. Based on the chart, it can be seen that the average pay rate increased sharply from 2006 to 2007 and gradually decreased until 2010. Then increased gradually until 2013.

- Pie chart illustrating Average Salary Percentage by Gender:

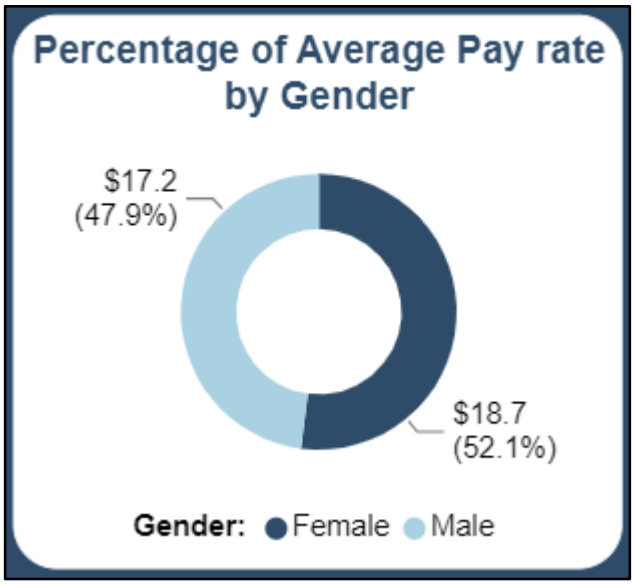


Figure 6.10. Pie chart illustrating the percentage of average pay rate by by gender

This chart provides insight into the distribution of average pay rate based on gender of staff. From the chart, we see that the average pay rate of female employees is higher than that of male employees ($18.7 > $17.2), indicating a potential gender gap in compensation within the company. Further examination of these differences is necessary to ensure pay is equitable and promotes an inclusive work environment. Overall, this analysis highlights the importance of considering gender and other information such as marital status in determining pay to promote fairness and equality in the workplace.

- Average pay rate ranking of each department:

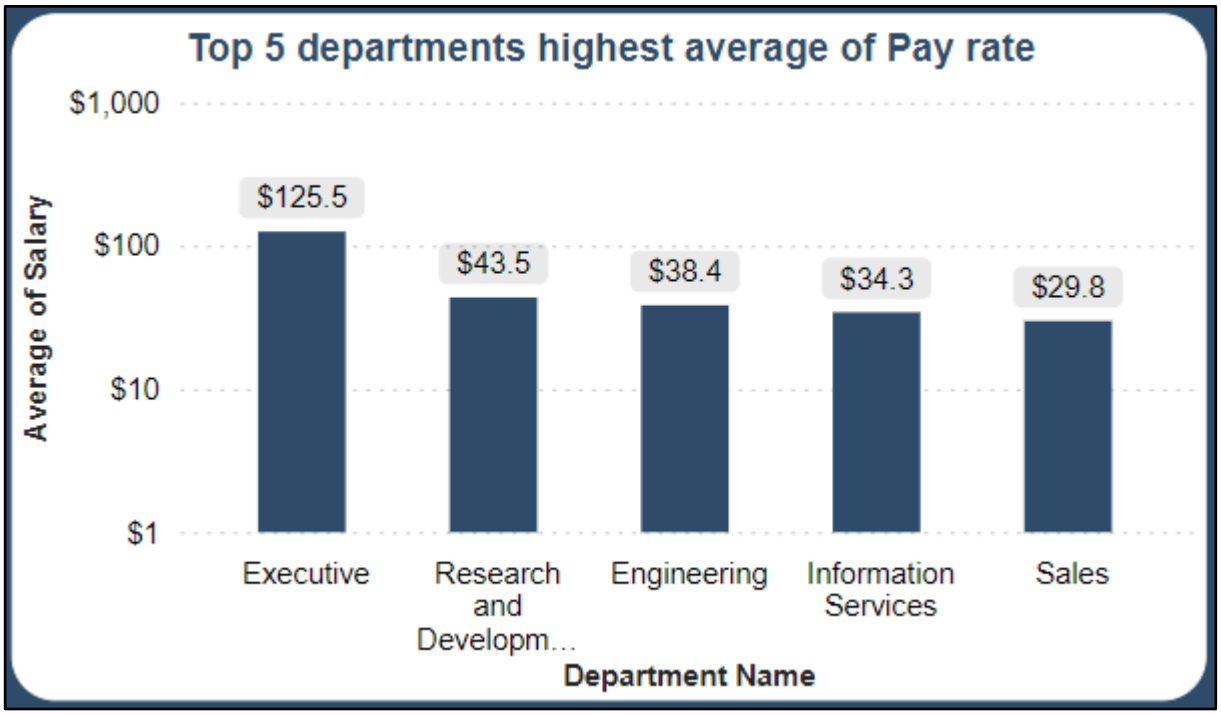


Figure 6.11. Column chart of average pay rate ranking of each department

This chart provides insights into the pay rate distribution across different departments within the company. Notably, it helps identify both the top-earning and lowest-earning departments. From the data, it's evident that Executive, Research and Development, Engineering, Information Services, and Sales are among the top 5 departments with the highest average salaries, with Executive leading at $125.5. This suggests that these departments may have roles that require specialized skills or responsibilities, hence commanding higher compensation. Conversely, departments ranking lower in the salary hierarchy may need further examination to understand factors influencing their compensation structures. Overall, this analysis aids in understanding the company's internal dynamics, informing decisions related to talent management, and ensuring equitable compensation practices across departments.

6.2.2.4. Business Question Resolution

Business question: "What is the average salary across different departments? Are there any gender pay gaps within the company?"

Answer:

Based on the table containing information about the average salary of 16 departments from 2006 to 2013 of Adventure Works Cycles company, one can know which department is paid the most in the company. And the top 5 departments with the highest average salary are concentrated in Executive, Research, Engineering, Information Services and Sales. Additionally, we also see that the average salary of female employees is higher than that of male employees ($18.7 > $17.2), indicating a potential gender gap in compensation within the company.

So there is a need to further examine these differences to ensure pay equity and promote inclusive work environments. And this analysis also highlights the importance of considering gender and other information such as marital status in determining pay to promote fairness and equality in the workplace.

6.2.3. KPI 2: Time off hours management

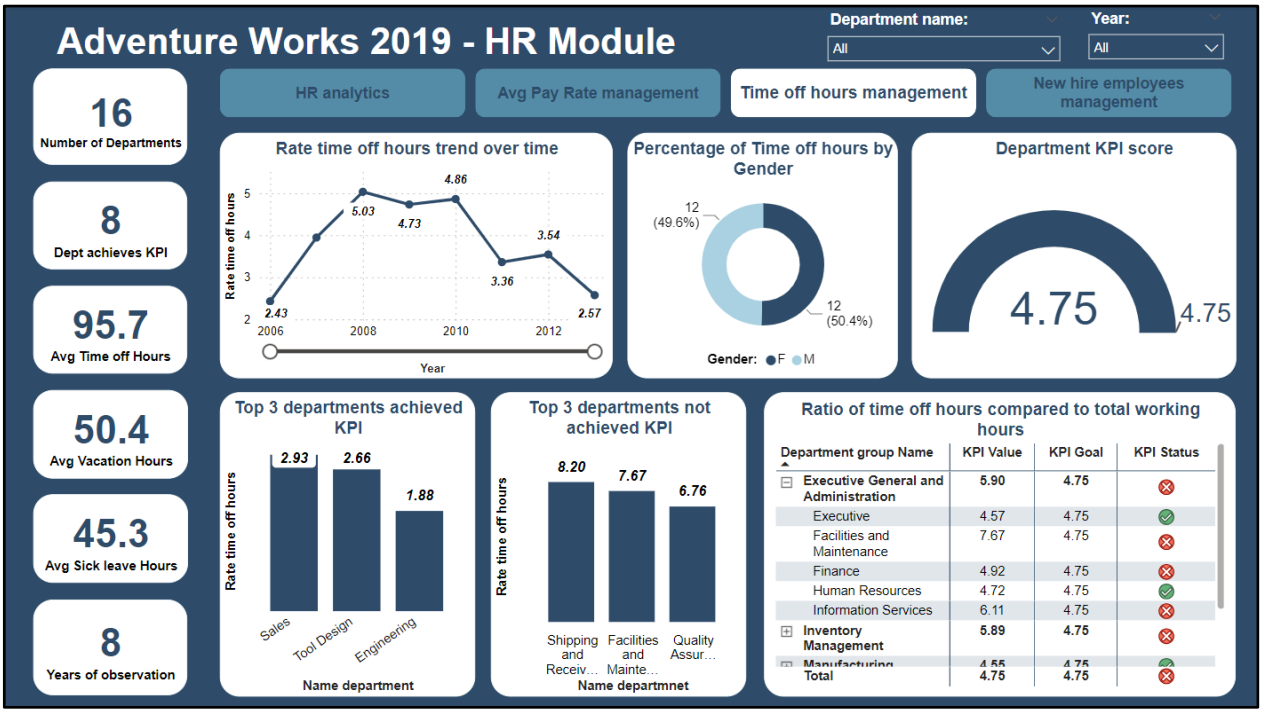


Figure 6.12. Dashboard of time off hours management

6.2.3.1. Purpose: Analyze employee absenteeism by each department

6.2.3.2. Description

On the dashboard, users can see the average time off hours, vacation hours, and sick hours of each department. It provides intuitive charts and dashboards, helping users track and analyze Time off of each department as well as compare between different groups of departments. Data collected from the dashboard also provides users with an overview of turnover trends in the company.

Based on these analyses, we noticed unique trends and patterns related to time off work. These comments provide us with important information to better understand and manage the use of working time of employees in the company as well as the health and mental status of employees.

6.2.3.3. Data Analysis

- KPI status data table

Rate time off hours: ratio of time off hours compared to the employee's total working hours in a year

Ratio of time off hours compared to total working hours

Department group Name	KPI Value	KPI Goal	KPI Status
Executive General and Administration	**5.90**	**4.75**	✗
Executive	4.57	4.75	✓
Facilities and Maintenance	7.67	4.75	✗
Finance	4.92	4.75	✗
Human Resources	4.72	4.75	✓
Information Services	6.11	4.75	✗
Inventory Management	**5.89**	**4.75**	✗
Manufacturing	**4.55**	**4.75**	✓
Total	**4.75**	**4.75**	✗

Figure 6.13. KPI 2 status table

KPI Evaluation Guidelines

+ A department is deemed to have met the Key Performance Indicator (KPI) when the ratio of time off hours is lower than the predetermined KPI threshold. This indicates that the HR department is effectively managing human resources and promoting employee well-being.
+ Conversely, a department is classified as not meeting the KPIs when the ratio of time off hours exceeds the established KPI threshold. This suggests that HR should investigate the underlying issues. Considering that the number of holidays may not be the same for every employee, it is possible that health-related concerns are causing increased sick hours.

By referring to this data table, it is possible to identify which departments are successfully meeting the KPIs. Additionally, if desired, the table allows for further observation by year.

- Trend assessment line chart:

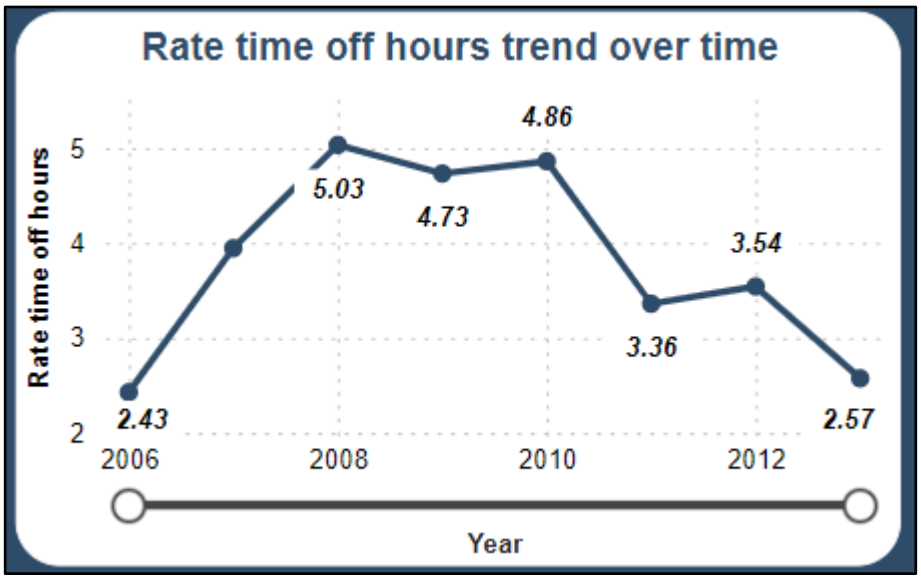


Figure 6.14. Line chart of Trend KPI 2

Use collected data to analyze trends in time off work over time. Users can compare time off hours rate by year to get insights on trends.

During the period between 2008 and 2010, there was a noticeable upward trend in this rate. This could be attributed to the company's expansion during that time, resulting in an increase in personnel recruitment. However, it appears that there were limited initiatives in place to enhance employees' overall well-being, including their living standards, spiritual life, and health.

- Pie chart illustrating Percentage of time off hours by gender:

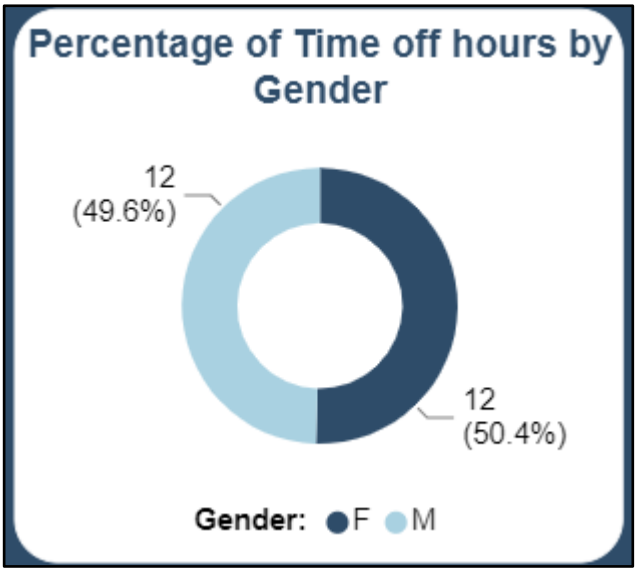


Figure 6.15. Pie chart illustrating the percentage of average pay rate by by gender

Based on the pie chart, which shows that the proportion of male employees' time off hours is 50.4% and the proportion of female employees' time off hours is 49.6%, we can conclude that there is no significant difference in time off hours between genders in the analyzed dataset.

However, to gain a better understanding of the underlying reasons for this small difference, it is necessary to consider other factors such as sample size, the time frame of the time off hours being analyzed, and additional dependent factors such as industry, job positions, or cultural aspects within the organization.

It is important to note that drawing conclusions about gender differences based solely on a single pie chart may be limited and should be examined in conjunction with statistical

analysis and other evidence to provide a more comprehensive view of the time off patterns between male and female employees in the organization.

- Column chart comparing time off hours ratio between departments:

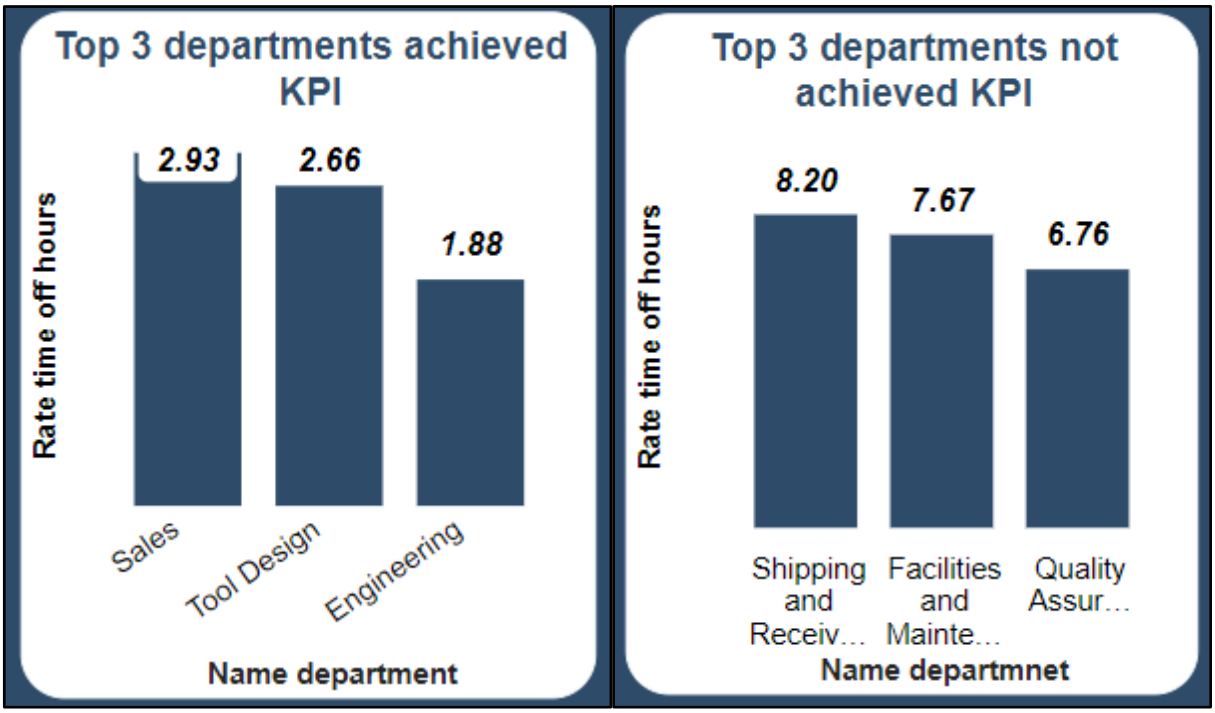


Figure 6.16. Column chart comparing time off hours ratio between departments

This chart provides a clear overview of departments that have met and not met their Key Performance Indicators (KPIs), allowing users to quickly identify high-performing departments and those requiring special attention.

The top-performing departments, with exemplary performance compared to the average, include Engineering (1.88% - 2.5 times lower than the average), Tool Design (2.66% - 1.8 times lower than the average), and Sales (2.93% - 1.6 times lower than the average). These departments have demonstrated strong performance in meeting their KPIs.

On the other hand, there are departments that warrant further analysis and attention. Shipping (8.2% - 1.7 times higher than the average), Facilities (7.67% - 1.6 times higher than the average), and Quality Assurance (6.78% - 1.4 times higher than the average) are

examples of departments that require special attention. These departments have not met their KPIs and need closer examination to identify areas for improvement.

Overall, this chart offers a comprehensive snapshot of departmental performance, highlighting both successful departments and those in need of focused intervention.

6.2.3.4. Business question resolution

Business question: "How does the time-off usage vary across different departments? Are there any correlations between time-off usage and employee satisfaction or productivity?"

Answer:

Through the analysis, it has been identified that there is a significant disparity in Time Off Hours across various departments within Adventure Works Cycles. The Shipping Department exhibits the highest rate of Time Off Hours, accounting for 8.2% of total hours, whereas the Engineering Department has the lowest rate at 1.8%. This indicates a substantial difference with a ratio of 4.5 times between the highest and lowest departments.

This significant variance raises concerns and warrants attention from the management. It suggests that certain departments may experience higher levels of absenteeism or time off requests compared to others. Further investigation is required to understand the underlying factors contributing to this discrepancy.

Regarding the correlation between time off usage and employee satisfaction, the initial hypothesis suggests a potential relationship. However, to validate this hypothesis, Adventure Works Cycles should gather additional data and insights regarding employee perceptions and experiences. This can be done through the implementation of employee satisfaction surveys or conducting interviews to gain a deeper understanding of how time off usage impacts employee satisfaction levels.

Based on the findings, it is recommended that Adventure Works Cycles takes proactive measures to enhance discipline and time off management within the organization. This can

be achieved through clear policies, effective communication, and regular monitoring of time off requests across departments. By maintaining a balance between providing a comfortable work environment and ensuring optimal productivity, the company can strive for improved performance and employee satisfaction.

6.2.4. KPI 3: New hire employees management

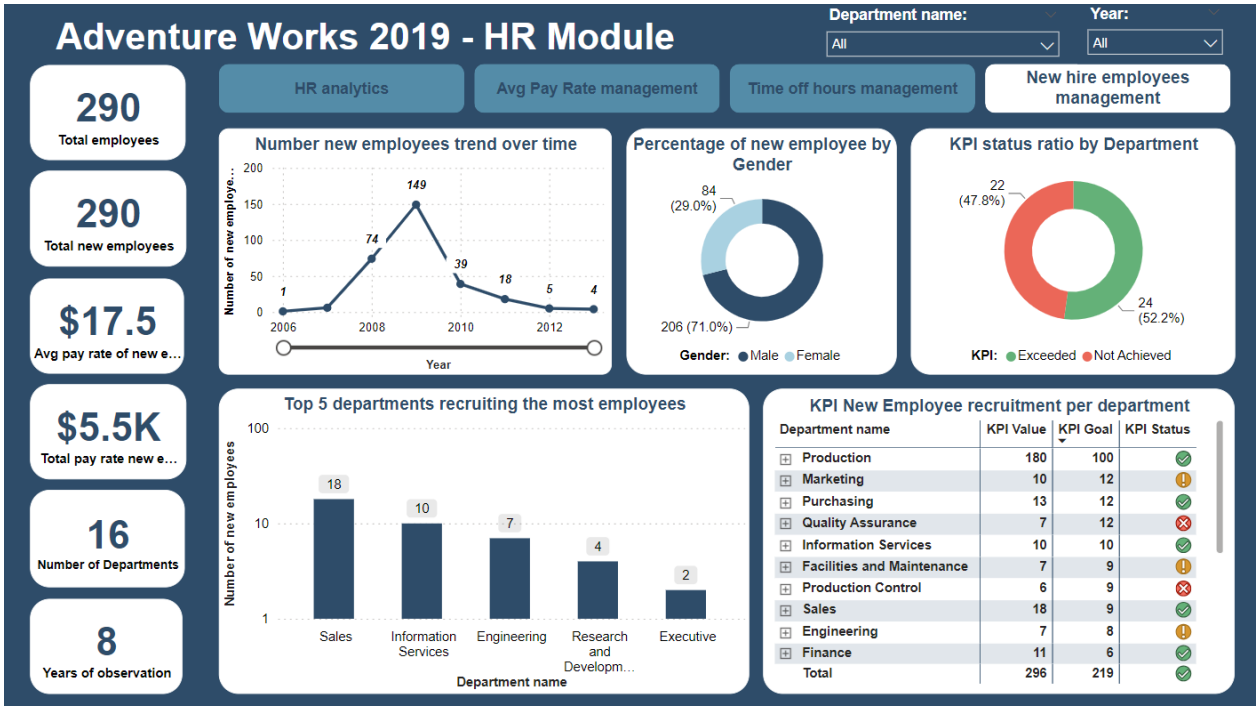


Figure 6.17. Dashboard of New hire employees management

6.2.4.1. Purpose: Analyze new employee recruitment trends of each department each year.

6.2.4.2. Description:

Below is a summary of information about recruiting new employees of 16 departments from 2006 to 2013 at Adventure Works Cycles company.

This dashboard allows users to track each department's annual hiring trends, total number of employees, gender ratio of newly hired employees, etc. through charts and tables. intuitive. Users can also filter information by department or year to meet specific needs.

This way, the dashboard provides an overview of the company's hiring process over the years. Through analyzing information from this table, managers can make strategic decisions about human resource and financial management. They can optimize compensation strategies, ensure fairness and transparency in the human resources management process, as well as create a positive, fair and inspiring working environment for all employees.

6.2.4.3. Data Analysis

- KPI New employee recruitment per department:

KPI New employee recruitment per department is evaluated through the targets set each year and based on the visual statistics table below we can see that:

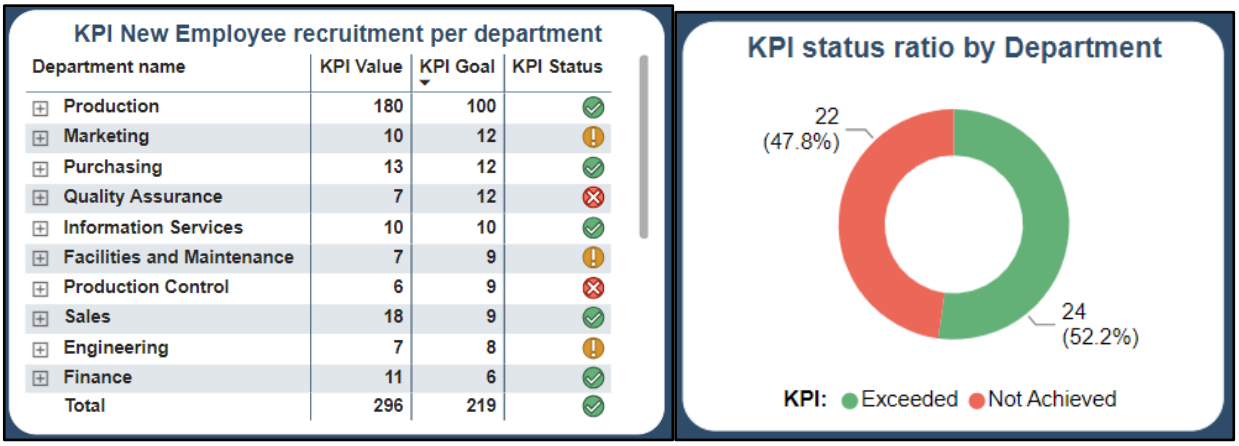


Figure 6.18. Dashboard of KPI New Employee recruitment per department.

For departments that meet the set recruitment goals, there will be an upward arrow, similarly for departments that do not meet the set targets, there will be a downward arrow. And looking at the visual table, User can see that the production department always has very high recruitment needs and targets, because this is the department that needs the most human resources, while the vast majority of other departments will have approximately Spend 6-9 new employees each year.

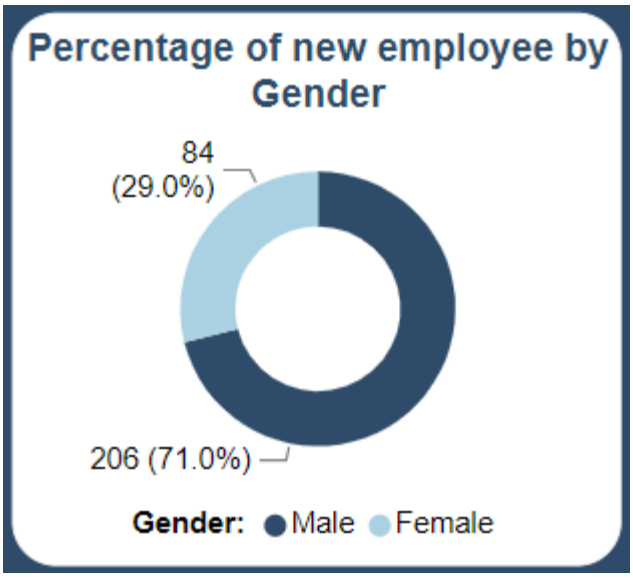


Figure 6.19. Donut chart of Percentage of new employees by gender

Pie chart illustrating the company's ratio of new hires by gender: This chart provides insight into the distribution of employees based on gender. From the chart, we see that the proportion of male employees recruited is higher than that of employees (71% > 29%). This shows that the gender gap is quite high. There is a need to further examine these differences to ensure a balanced gender ratio between men and women and promote inclusive work environments.

- Trend assessment line chart:

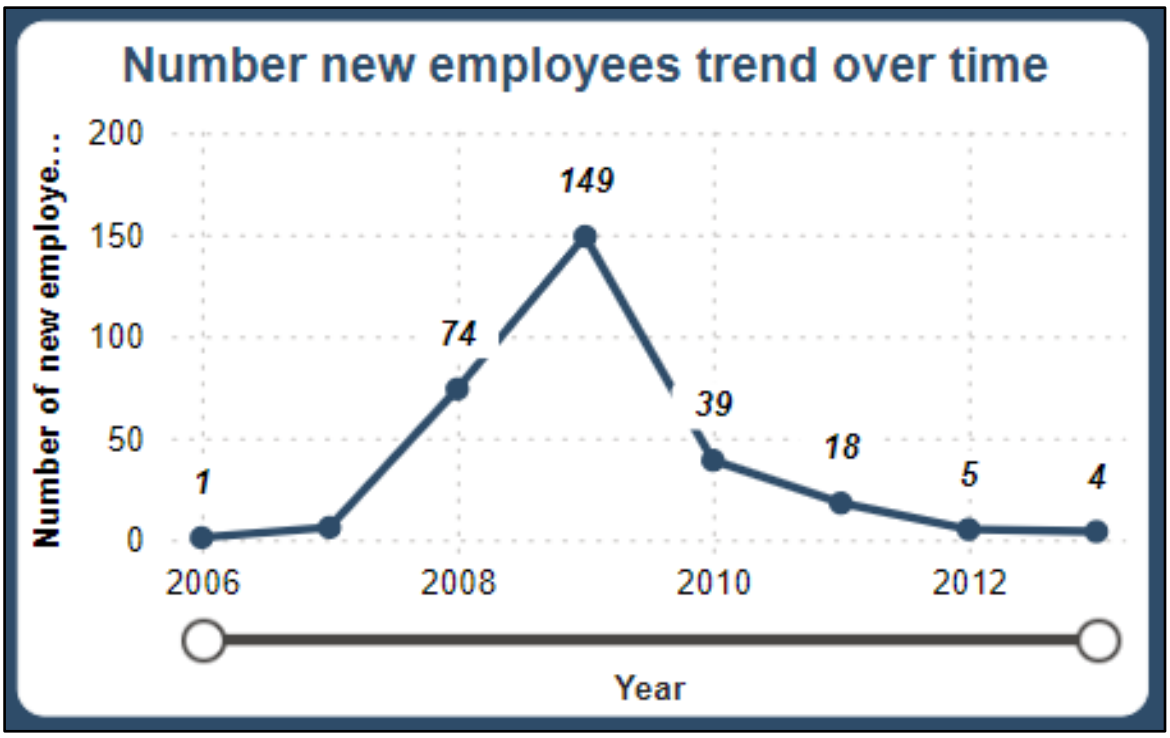


Figure 6.20. Line chart of Number new employees trend over time

Collected data was used to analyze recruitment trends of each department from year to year. Users can use the number of new employees starting work to predict whether each department's targets are met or not. Between 2008 and 2010, this ratio had a clear upward trend. Chart analysis shows that in 2008 there were a total of 74 new employees recruited, in 2010 there were 39 new employees, and in 2009 it peaked with 149 new employees. This can be explained by the company's expansion during that period, which led to increased hiring.

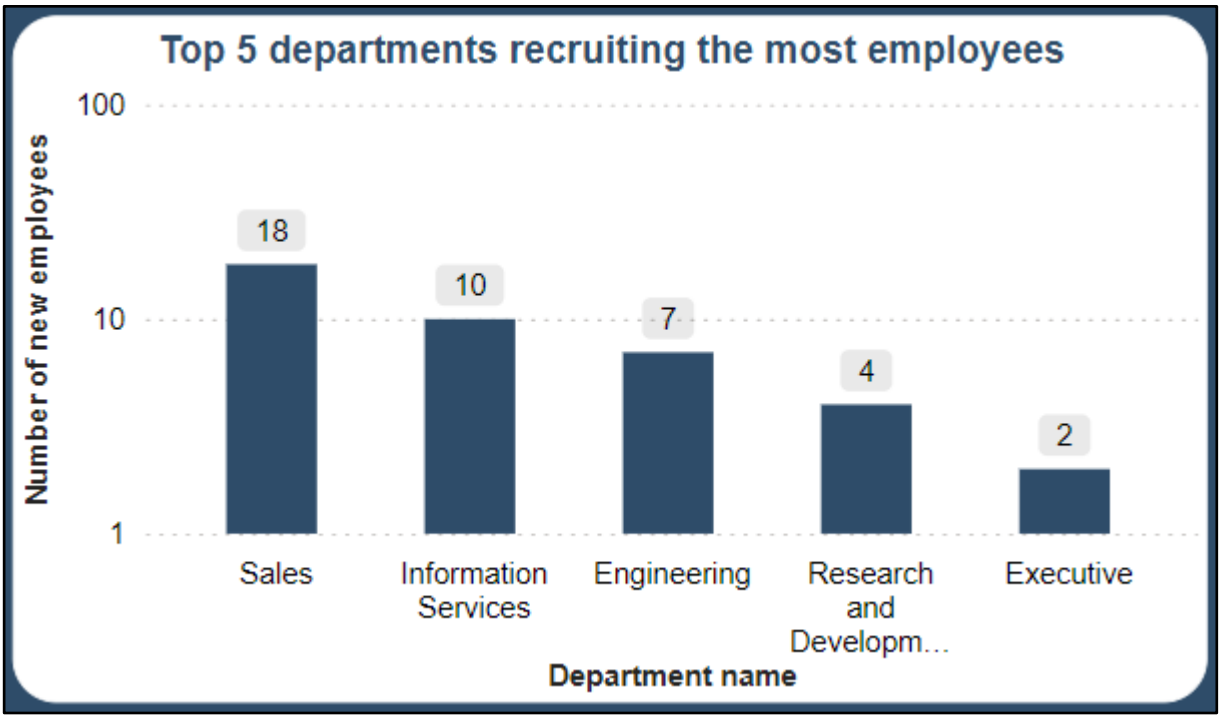


Figure 6.21. Top 5 Department recruiting the most employees.

This chart provides insight into the distribution of new hires across departments within the company. Notably, it helps identify the five departments with the highest total number of new hires. From the data, it is clear that Sales, Information Service, Engineering, Research and Executive are among the top 5 departments with the highest total number of new employees, with Sales being the top department (18 new employees). This suggests that these departments may have roles that require specialized skills or responsibilities, thus requiring more human resources. Conversely, departments ranked lower in the hierarchy may need further examination to understand influencing factors. In general, this analysis helps understand the company's recruitment trends so that appropriate salary policies and regimes can be adjusted to make good use of human resources to serve the company.

6.2.4.4. Business question resolution

Business question: "How does the workforce demand vary across different departments? Are there any departments that are overstaffed or understaffed based on their current and projected workforce demand?"

Answer:

Based on statistics and analysis, there are clear differences in workforce needs in each department. From 2006 to 2013, Group Name Manufacturing accounted for the largest number of employees (104 employees in 2009, 52 employees in 2008 and 22 employees in 2010); Next is the Executive General & Administration department with a total of 35 employees, Sales & Marketing (28 employees),...

The reason for the difference in human resources in each department is first of all because Manufacturing is where the production process is strictly controlled to ensure the quality of the final product. This ensures that the product meets customer standards and requirements. Therefore, a large force of personnel is needed to allocate work properly and closely manage it to ensure efficiency.

Besides, based on statistics, it can be seen that there are currently 6 departments with a turnover of one employee per year (Inventory Management, Research, Sales, Manufacturing, Quality Assurance and Executive) and based on the values KPI goal and KPI value have been analyzed, all departments tend to lack staff but the number is not significant, and it is forecast that the number will increase higher in each department due to the amount of work and customer needs client.

6.3. Evaluation and Discussion

Dashboards serve as the nerve center for informed decision-making, offering managers a lucid and succinct view of business health through integrated data insights presented in a variety of charts. Let’s delve into the key insights provided by each KPI dashboard within Adventure Works Cycles:

The central emphasis is placed on the Pay Rate Management Dashboard, serving as a fundamental tool for understanding and refining the organization’s compensation strategies. This dashboard delivers extensive insights into salary patterns across departments and time periods, offering valuable information to ensure fair and balanced

compensation structures. Through the examination of average salary trends, HR professionals can detect potential discrepancies and effectively tackle compensation-related issues. This proactive approach not only boosts employee morale and retention but also cultivates a more equitable workplace atmosphere. Additionally, the Pay Rate Management Dashboard allows for targeted interventions to promptly address compensation challenges, contributing to operational efficiency and organizational success. Given its capacity to provide detailed insights into compensation practices, this dashboard plays a pivotal role in guiding strategic decisions and promoting fairness and transparency within the organization. Overall, the Pay Rate Management Dashboard emerges as a critical tool for HR professionals aiming to refine compensation strategies, elevate employee satisfaction, and foster organizational advancement in a competitive business landscape.

In addition, the Time Off Hours Management Dashboard offers a granular view of employee absenteeism patterns, shedding light on crucial aspects of workforce well-being and productivity. Through meticulous analysis of time off trends across departments, HR professionals can identify root causes of absenteeism and implement proactive measures to mitigate their impact. This proactive approach not only fosters a healthier work environment but also enhances operational efficiency by addressing absenteeism-related disruptions promptly.

Moreover, the New Hire Employees Management Dashboard provides essential insights into recruitment trends, enabling HR to align talent acquisition strategies with departmental needs and organizational objectives. By tracking recruitment patterns and new hire demographics, HR professionals can optimize recruitment processes, improve candidate quality, and ensure a seamless onboarding experience. This strategic alignment fosters workforce agility and enables the organization to adapt swiftly to changing business dynamics.

Overall, the integration of these data-driven insights into HR decision-making processes empowers organizations to cultivate a culture of continuous improvement and adaptability. By leveraging the power of HR analytics, organizations can optimize resource allocation, enhance employee satisfaction, and ultimately achieve their strategic objectives in a dynamic and competitive business landscape.

However, while the text provides a comprehensive overview of the benefits of each dashboard, it could benefit from deeper analysis and discussion of potential limitations or challenges associated with their implementation. Additionally, further exploration of specific metrics and their alignment with broader organizational goals could enhance the richness and depth of the evaluation.

CHAPTER 7: CONCLUSION AND FUTURE WORKS

7.1. Conclusion

7.1.1. Results

During the implementation of our HR project, our team effectively applied a range of theoretical and practical concepts learned in class to analyze data and demonstrate the significance of data analysis in business operations. Our focus was Adventure Works Cycles, a multinational manufacturing company, chosen to showcase the effectiveness of Business Intelligence (BI) solutions in real-world scenarios.

Throughout this project, our team conducted meticulous analysis of the Adventure Works Cycles sample data, gaining hands-on experience in constructing a Data Warehouse using various data analysis tools and techniques. Additionally, we defined key performance indicator (KPI) criteria to measure business performance based on the analyzed data. By integrating BI solutions into the HR management process, we evaluated the company's workforce productivity and generated in-depth analysis reports that supported decision-making at the management level.

The results of our analysis highlighted the critical role of implementing BI solutions in optimizing productivity, facilitating decision-making, and translating data into actionable insights within business operations. Through this project, our team acquired essential knowledge and skills in data analysis, BI implementation, and KPI evaluation methods. We also had the opportunity to develop comprehensive analysis reports that effectively supported decision-making in HR operations.

We would like to express our sincere gratitude to our instructor, Mr. Nguyen Van Ho, and teaching assistant, Mr. Le Ba Thien, for their enthusiastic support and providing us with valuable opportunities to apply data analysis techniques in the context of HR decision-making.

7.1.2. Limitations

Despite completing the project within the designated timeframe, we encountered several limitations throughout the process. One major challenge our team faced was the considerable amount of time spent on identifying suitable dimensions and fact tables. We also engaged in multiple discussions and made adjustments in collaboration with the teaching assistant to ensure the completeness and alignment of the data warehouse with our initial criteria. Moreover, familiarizing ourselves with data analysis tools and techniques presented some difficulties. Nevertheless, through mutual support and cooperation among team members, we successfully surpassed our initial expectations and accomplished the project.

7.2. Future Work

Based on the attained outcomes and the outlined constraints, the project can progress with the following recommendations:

- Integrating other modules into the research to get a more detailed overview of the company's performance and overall strategy. For example, combining with the Sales module will provide insights into sales data, customer behavior and market trends. Similarly, integrating employee purchasing and sales data can help create monthly sales or individual performance reports for all employees on the Adventure Works Cycles database platform.
- Appling more tools, analysis and visualization methods such as Python, Tableau, VPA Excel,... to the research process. Aims to build a solid foundation when approaching real projects at businesses.
- Finally, 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.

APPENDIX

Table 0.1. Lesson Learned

No	Issue / what did not work well	What should the team have done instead?	Lesson Learned
1	Difficulty in identifying suitable dimensions and fact tables for the Data Warehouse	Prioritize thorough data profiling and analysis to better understand the structure and relationships within the data. Engage in early discussions with stakeholders to define clear objectives and requirements.	Importance of comprehensive data exploration and stakeholder engagement in the initial stages of the project to ensure alignment and efficiency in data modeling.
2	Challenges in familiarizing with data analysis tools and techniques	Allocate sufficient time for team members to undergo training and hands-on practice with selected tools. Seek additional support or resources from instructors or online tutorials to overcome specific technical hurdles.	Recognizing the importance of continuous learning and skill development, and the need for proactive measures to address knowledge gaps.
3	Time constraints and scheduling	Implement effective project management practices, such as	Emphasizing the importance of efficient

	conflicts among team members	setting clear milestones, regular progress updates, and flexible scheduling options. Prioritize effective communication and collaboration to address any scheduling conflicts promptly.	time management, communication, and teamwork in overcoming challenges and meeting project deadlines.
4	Initial ambiguity in defining KPI criteria and performance metrics	Conduct thorough research and benchmarking to establish clear and measurable KPIs aligned with organizational goals. Seek input from stakeholders and subject matter experts to refine KPI definitions and ensure relevance and effectiveness.	Highlighting the significance of stakeholder involvement and iterative refinement in defining meaningful KPIs that drive actionable insights and support decision-making.
5	Overcoming technical hurdles and troubleshooting data analysis challenges	Foster a collaborative and supportive team environment where members can freely exchange ideas, seek assistance, and collectively problem-solve technical issues. Utilize online	Emphasizing the importance of resilience, resourcefulness, and teamwork in overcoming technical obstacles and achieving project objectives.

		forums, tutorials, and peer-reviewed resources to supplement classroom learning and enhance technical proficiency.	

Table 0.2. Lesson Learned - Best Practice

No	What worked Well	Lesson Learned (Best Practice)
1	Effective collaboration among team members, fostering a supportive team dynamic	Emphasize the importance of teamwork and communication in project success. Encourage an environment of mutual support and open dialogue to facilitate problem-solving and knowledge sharing.
2	Comprehensive data analysis and visualization, providing actionable insights	Prioritize thorough data analysis and visualization to derive meaningful insights. Utilize a variety of analytical tools and techniques to uncover trends and patterns in the data.
3	Regular progress updates and milestone tracking, ensuring project alignment	Implement a structured project management approach with clear milestones and regular progress monitoring. Maintain transparency and accountability

		through frequent updates and checkpoints.
4	Engagement with instructors and teaching assistants for guidance and support	Seek guidance and support from instructors and teaching assistants to overcome technical challenges and hurdles. Leverage their expertise and experience to address project-related queries and clarify concepts.
5	Iterative refinement of KPI criteria, ensuring alignment with organizational goals	Highlight the importance of continuous improvement and refinement in defining KPIs. Solicit feedback from stakeholders and subject matter experts to refine KPI definitions and enhance relevance.
6	Flexibility and adaptability in addressing unforeseen challenges	Emphasize the need for resilience and adaptability in navigating project hurdles. Encourage a proactive approach to problem-solving and the willingness to adjust strategies as needed.

REFERENCES

1. ACT Group. (2021). Báo cáo thị trường xe đạp Việt Nam 2020 – Dự báo thị trường 2021. *ACT Group.* Retrieved February 20, 2024, from https://actgroup.com.vn/bao-cao-thi-truong-xe-dap-viet-nam-2020-du-bao-thi-truong-2021/

2. Admin, & Admin. (2023). Data Mart vs. Data Warehouse: Key Differences Explained. *Rishabh Software.* Retrieved March 3, 2024, from https://www.rishabhsoft.com/blog/data-warehouse-vs-data-mart.

3. Amazon. Kho dữ liệu là gì? – Giải thích về Kho dữ liệu – AWS. (n.d.-b). *Amazon Web Services, Inc.* Retrieved March 3, 2024, from https://aws.amazon.com/vi/what-is/data-warehouse/

4. Anh Nguyen. (2023). Sức hút từ thị trường bán lẻ xe đạp khiến Thế Giới Di Động lẫn FPTShop đều phải lao vào cuộc chơi. *Vietnambiz.* Retrieved February 20, 2024, from https://vietnambiz.vn/suc-hut-tu-thi-truong-ban-le-xe-dap-khien-the-gioi-di-dong-lan-fptshop-deu-phai-lao-vao-cuoc-choi-202332893923781.htm

5. Ballejos, L. (2024). What is a Data Warehouse? | Definition & Overview | NinjaOne. *NinjaOne*. Retrieved March 3, 2024, from https://www.ninjaone.com/blog/what-is-a-data-warehouse

6. Grand View Research. Bicycle Market - Industry Analysis, Market Size, Share, Trends, Growth and Forecast, 2021 - 2028. *Grand View Research.* Retrieved February 22, 2024, from https://www.grandviewresearch.com/industry-analysis/bicycle-market

7. Grand View Research. Bicycle Market Size, Share & Trends Analysis Report, 2030. *Grand View Research.* Retrieved February 20, 2024, from https://www.grandviewresearch.com/industry-analysis/bicycle-market

8. Humaans. Retention Rate: How to Calculate and How to Improve It (+ Turnover Cost Calculator). *Humaans.* Retrieved March 3, 2024, from https://humaans.io/hr-glossary/retention-rate.

9. Human resources - AdventureWorks. (n.d.). *Dataedo.* Retrieved February 22, 2024, from https://dataedo.com/samples/html/AdventureWorks/doc/AdventureWorks_2/modules/Human_Resources_9/module.html

10. IBM. What is a data mart? *IBM*. (n.d.). Retrieved March 3, 2024, from https://www.ibm.com/topics/data-mart

11. Kenton, W. (2024). What is payroll, with Step-by-Step calculation of payroll taxes. Investopedia. *Investopedia.* Retrieved February 22, 2024, from https://www.investopedia.com/terms/p/payroll.asp

12. Kfollis. (2022, December 9). Querying Multidimensional Data with MDX. *Microsoft Learn*. Retrieved March 3, 2024, from https://learn.microsoft.com/en-us/analysis-services/multidimensional-models/mdx/querying-multidimensional-data-with-mdx?view=asallproducts-allversions

13. LinkedIn. (2024). Using MDX to Perform Complex Calculations in OLAP. *LinkedIn Advice.* Retrieved March 3, 2024, from https://www.linkedin.com/advice/1/how-can-you-use-mdx-perform-complex-calculations-olap-sl9ge

14. Microsoft. (n.d.). MDX Query Fundamentals in Analysis Services. *Microsoft learn.* Retrieved March 3, 2024, from https://learn.microsoft.com/en-us/analysis-services/multidimensional-models/mdx/mdx-query-fundamentals-analysis-services?view=asallproducts-allversions

15. PriskeyJeronika-Ms. (2023, March 16). *Overview of OLAP cubes for advanced analytics*. Microsoft Learn. Retrieved March 3, 2024, from

https://learn.microsoft.com/en-us/system-center/scsm/olap-cubes-overview?view=sc-sm-2022

16. Rahaman, M. A. (2016). Human Resource Management practices Recruitment and selection process in Multinational Company. *Journal of Economics and Development Studies, 5*(1), 1-12.

17. Resource For Employee. (2023). How to calculate employee turnover rate. *Resource For Employee.* Retrieved March 3, 2024, from https://resources.workable.com/tutorial/calculate-employee-turnover-rate.

18. Rita Rain and Tibor Pál (2024). Pay Raise Calculator. Retrieved March 3, 2024, from https://www.omnicalculator.com/finance/pay-raise.

19. Statista. *Bicycles - Vietnam | Statista market forecast*. *Statista*. Retrieved February 20, 2024, from https://www.statista.com/outlook/mmo/bicycles/vietnam

20. Tableau. Business intelligence: A complete overview. (n.d.). *Tableau.* Retrieved February 22, 2024, from https://www.tableau.com/learn/articles/business-intelligence