



NOVEMBER 19, 2023

DATA WAREHOUSING

DAT701 ASSESSMENT 2

KIRA BYRNE

ID #13509995

NMIT

Contents

| | |
|--|---|
| DISCLAIMER..... | 2 |
| Introduction | 2 |
| FinanceDB vs FinanceDW Entity Relationship Diagrams..... | 2 |
| Rationale | 3 |
| PowerBI Dashboards | 5 |
| Project Summary..... | 5 |
| References..... | 6 |

DISCLAIMER

As described in the marking rubric for this assessment, all SQL scripts and PowerBI documents were to be handed in via the TALOS virtual machines, however, due to an unforeseen technical error, the network connection between my personal virtual machines was not working as expected. It was agreed by my tutor that my work would be completed on my local machine, so my SQL scrips and PowerBI documents will be delivered in a ZIP folder to be assessed as needed.

In regard to section one of the marking rubric, I have successful restored the FianceDB on my TALOS virtual machine that hosts SQL. This can be assessed as the rubric describes.

Introduction

The following is a report documenting the conversion and implementation of the Finance Data Warehouse (FinanceDW) from the Finance Database (FinanceDB). The documentation includes an entity relationship diagram (ERD) for the FinanceDW, rational for each decision made in the design process, a data dictionary, and PowerBI dashboards using the merged data.

FinanceDB vs FinanceDW Entity Relationship Diagrams

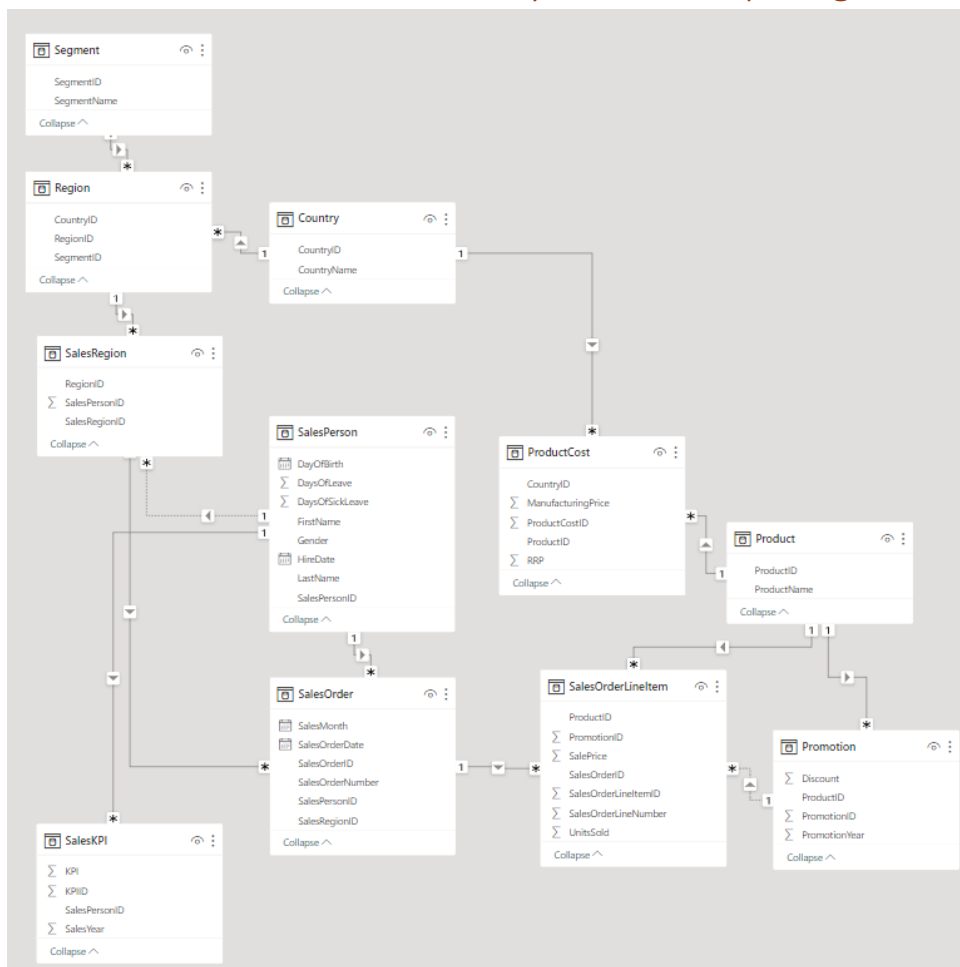


Figure 1 FianceDB (source database)

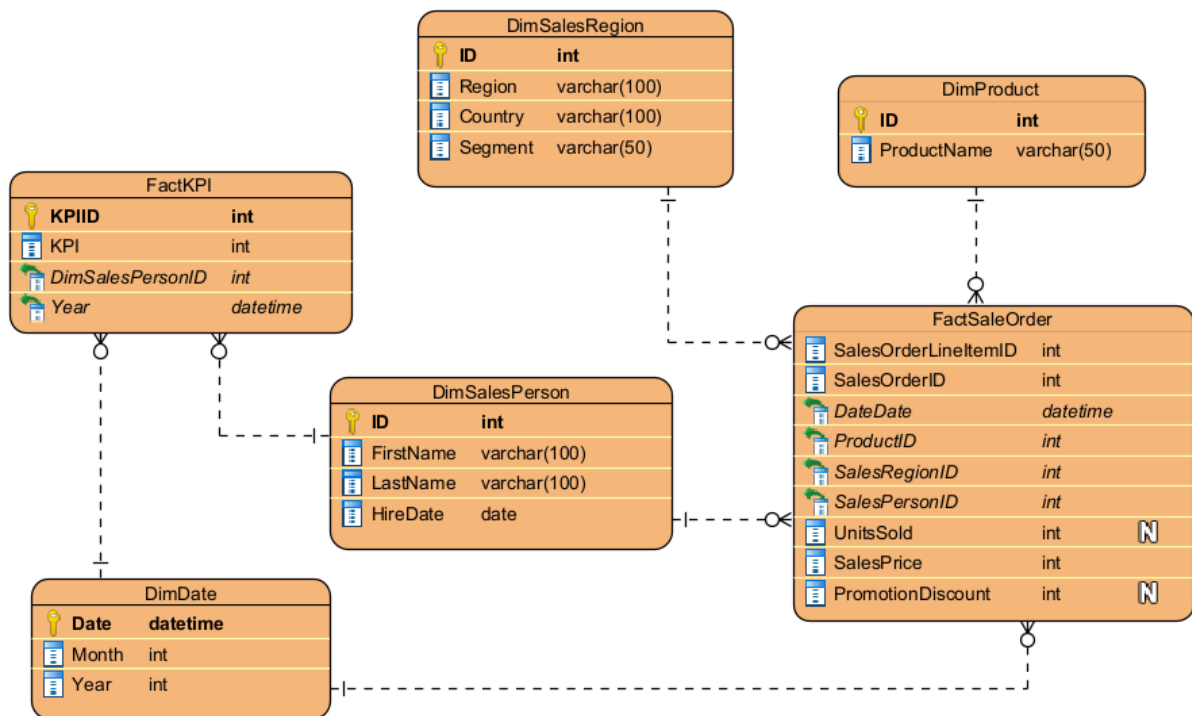
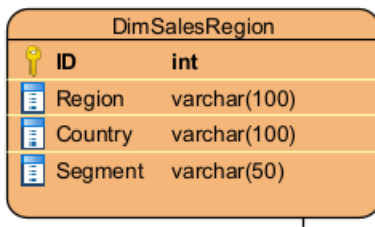


Figure 2 FinanceDW (new data warehouse)

Rationale

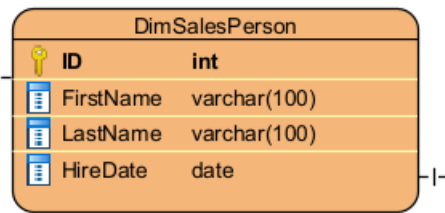
DimSalesRegion



The Sales Region dimension table is a merge of 4 identification tables from the FianceDB: Segment, Country, Region, and SalesRegion.

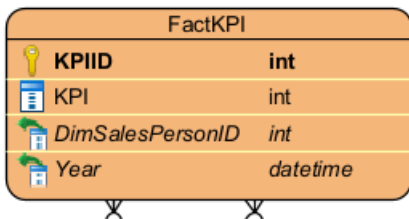
There is only one Sales Person per Sales Region which is identifiable through a foreign key relationship, therefore each sales Region could be identified by a sales person. This includes the profit and KPI for each sales region. The database has a small number of counties US, Mexico, Canada, France, and Germany, each with a unique name. Each region can be identified by its related country, therefore they could be merged with sales region and sales person. Segment, similar to country, is merely a filterable column with each segment name being unique. Each sales person only works for one segment. The primary key is each sales person which each sale for each sales region can be traced back to from the SalesOrder table. Region, Country, and Segment are treated as filters.

DimSalesPerson



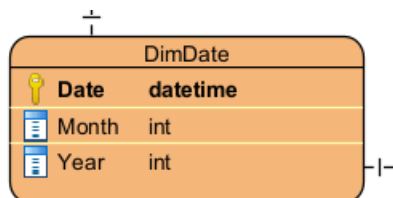
The Sales Person dimension table only stores necessary information about each sales person used either to identify them in both unique (primary key ID) or meaningful ways (first and last name). The hire date can be used for querying such as finding out the length of time each employee has been employed or their matching their yearly KPI quota from their hire date.

FactKPI



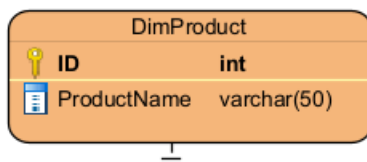
The Key Performance Indicator (KPI) fact table stores the KPI per SalesPersonID, filterable by (sales) year, a foreign key from the DimDate table.

DimDate



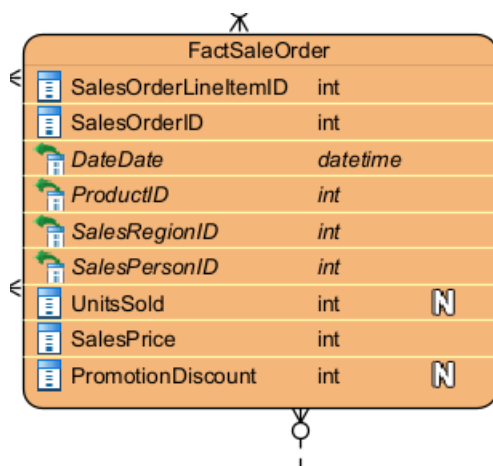
The Date dimension table holds all date columns used in the two fact tables, FactKPI, and FactSalesOrder. The date time acts as a foreign key in each table which is then filterable by month and year.

DimProduct



Each product is uniquely identifiable by a key and known by a meaningful product name. The sales price and units sold are added to the FactSalesOrder table, identified by the productid foreign key which allows filtering by DimSalesPerson to calculate total profits.

FactSalesOrder



The Sales Order fact table merges information from the SalesOrder, SalesOrderLineItem, Product, and Promotion table in the FinanceDB. The dates of sale are identified through the foreign key of the date table where the information can be queried. Each order can be traced back to both sales person and sales region. Sales price and units sold are stored in one table for ease of query when calculating total profit, filterable by sales person, and sales region. Because DimSalesRegion stores all location based data, profits can be filtered against this table by region, country, and segment to save joining in the query.

PowerBI Dashboards

Was not able to complete in timeframe.

Project Summary

The FinanceDB was a simple exemplar of a poorly made database with a large set of data that is spread across many tables that are difficult to effectively query. The FinanceDB could be optimised through conversion into a data warehouse, where the tables could be denormalised into a suitable format where the larger sets of consistently updating data could be stored in fact tables and filtered by the essential information within the dimension tables. The optimisation provided in the FinanceDW creates a better connection for dashboard development in PowerBI with fewer tables and connections to work with.

Roles

Data warehouse developer

The role of a data warehouse developer has its importance compared to a database developer. Data warehouse development is a complex process that deals with large sets of preexisting data. A DWD should be skilled in database creation as a foundation for their core skills. These include using ETL tools for merging data, normalisation, and denormalisation of data (Campana, 2022).

Business Analyst

A business analyst is an integral role of any business venture that is dealing with data. They are responsible for understanding both the information technology and the quality improvement of the business structure to provide meaningful insights from data analysis (Pratt et al., 2023). To simplify,

they are able to manipulate large sets of data to produce understandable reports and presentations to improve a business or project.

In this project I have learned that data science is truly a world of its own, separate to what is commonly known as 'IT'. It is so complex that a general title cannot be made for it, however, the key roles within data science are more interlinked than they are separate. To be a business analyst, one should have a basic level of skill in database development. As they are responsible for working with the data to produce meaningful reports, their skill set should extend to both the IT realm and the business realm to effectively work as communication pathway between the two (BestAccreditedColleges, 2022).

References

BestAccreditedColleges. (2022). *Business intelligence analyst vs Database administrator*.

<https://bestaccreditedcolleges.org/articles/business-intelligence-analyst-vs-database-administrator.html>

Campana, N. (2022, September 20). *What does a data warehouse developer do?* Freelancer Blog.

<https://www.freelancermap.com/blog/what-does-data-warehouse-developer-do/>

Pratt, M. K., & White, S. K. (2023, October 6). *What is a business analyst? A key role for business*.

CIO. <https://www.cio.com/article/276798/project-management-what-do-business-analysts-actually-do-for-software-implementation-projects.html>