AdventureWorks2019 Data Engineering Assessment Report

Introduction

This report documents the process of designing and implementing a data warehouse solution using the AdventureWorks2019 dataset. The primary goal was design and implement Datawarehouse. Additionally, an ETL pipeline was created to transfer data from the source to the staging database and finally to the data warehouse for further analysis.

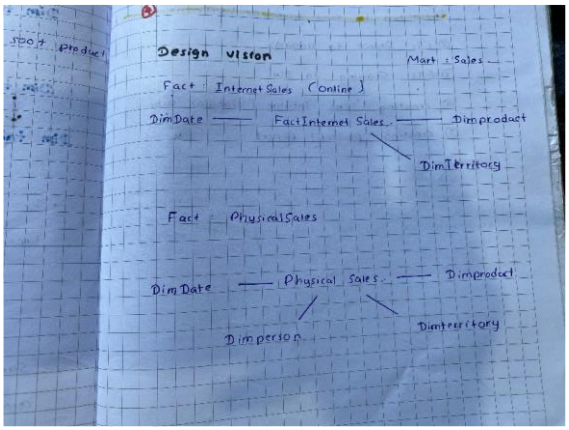
Process

1. **Store OTLP database.**
    - The **AdventureWorks2019 OLTP** database was set up to store transactional data.
    - It serves as the raw data source for further transformation and analysis.

2. **Analyses OTLP Database**.
    - The OLTP database was examined to understand table structures, relationships, and required transformations.

3. **Create AW_STAGE Database.**
    - A staging database (AW_STAGE) was created to hold intermediate data before transforming it into the data warehouse**.**
    - **Link : https://github.com/Yashodara523/AdventureWorks2019/blob/main/AW_STAGE.sql**

4. **Design AW_DW Database**
    - The data warehouse AW_DW was designed using a star schema.



5. **Create AW_DW Database**
    - A separate database for the data warehouse was created**.**

    Link: https://github.com/Yashodara523/AdventureWorks2019/blob/main/AW_Datawarehouse.sql

6. **ETL data source to AW_STAGE**
    - Extracted data from the OLTP database and loaded it into the AW_STAGE database.

    Link: **https://github.com/Yashodara523/AdventureWorks2019/tree/main/Load_Stage**

7. **ETL AW_STAGE to AW_DW**

    Data from AW_STAGE was transformed and loaded into AW_DW.

    Steps involved:

    - Create stg_vw_tables
      Link**:** **https://github.com/Yashodara523/AdventureWorks2019/blob/main/AW_STAGE.sql**
    - Create Procedures and Refresh tables.
      Link:https://github.com/Yashodara523/AdventureWorks2019/blob/main/AW_Datawarehouse.sql

      Link: https://github.com/Yashodara523/AdventureWorks2019/blob/main/ReProcedure.sql

8. **Create cube(SSAS Multidimensional analysis)**
    - Created an SSAS Multidimensional Cube for efficient analysis and reporting.

    Link: https://github.com/Yashodara523/AdventureWorks2019/tree/main/NewDataWareHouse

9. Semantic Model (continue….)
10. PowerBI Report (continue….)

Challenges and Solutions

Challenges	Solutions
Data consistency between staging and warehouse	Used stored procedures and refresh mechanisms
ETL data source to AW_STAGE(Huge tables)	Use Increment_ETL
Error Handling & Debugging while ETL data source to AW_STAGE	Use Full_ETL(Create parallel Dataflow tasks)
Maintenance ETL AW_STAGE to AW_DW	Use staging Views and procedures