Data Integration and Warehousing Assignment

Sanduni Gunasinghe

Data Integration and Warehouse Implementation

1. Daily Data Load

- Developed ETL (Extract, Transform, Load) workflows using SQL Server Integration Services (SSIS) to load daily transactional data into the data warehouse.
- Ensures incremental data loading by processing only new or updated records. packages load daily transactional data into the warehouse through the SQL Server Agent.
- Data will be loaded into the data warehouse every day at **12:00:00 AM**.

Data Flow:

- Source: **Transactional Database (OLTP System): AdventureWorks2019**
- Staging database: **AWN_STG_Demo**
- Destination: **Data Warehouse (OLAP System): AWN_DW_Demo**

2. Data Updates

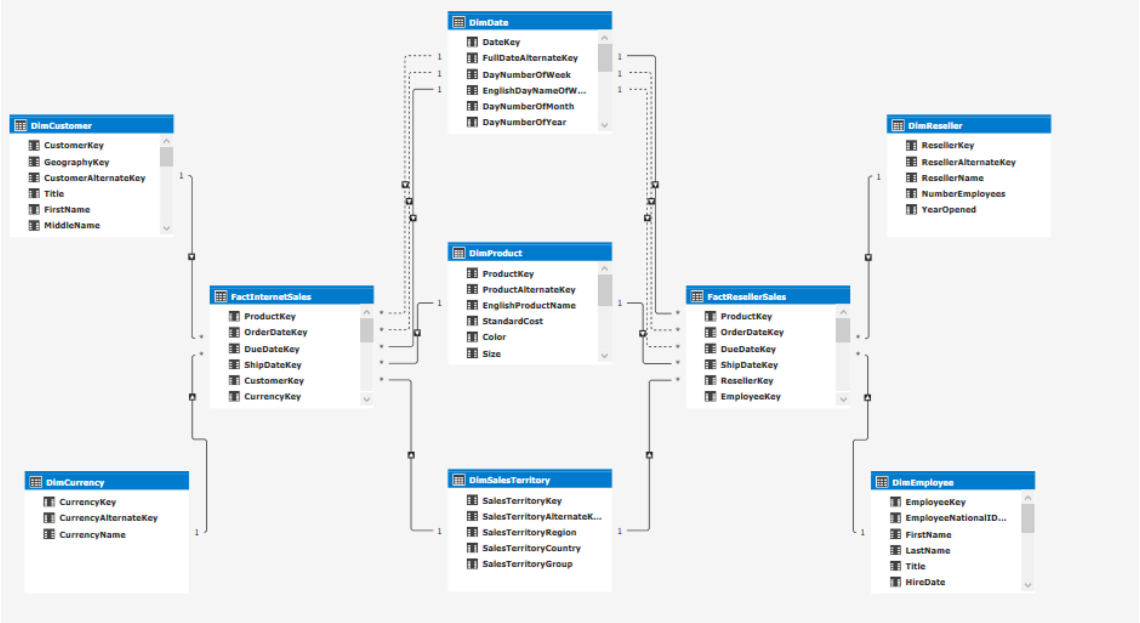
- **Territory Updates**: Implemented using Slowly Changing Dimension only track the updates without tracking the history.
- **Product Updates**: Implemented using **Slowly Changing Dimension Type 2 (SCD2)**.

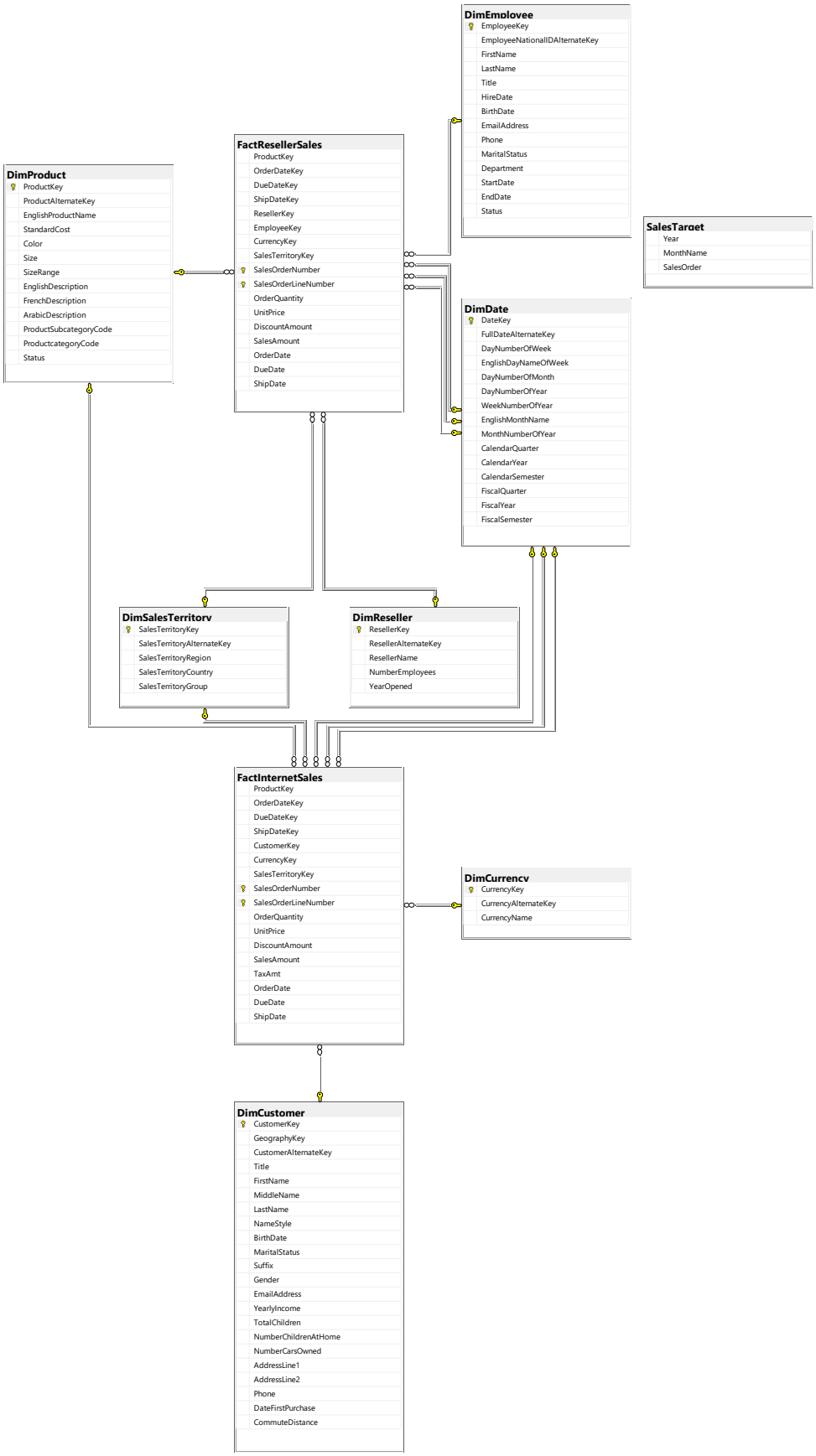
3. Monthly CSV Processing

- The **SalesOrderForecast CSV** file is loaded into the staging database before being uploaded to the data warehouse.
- A SQL Server Agent job is set up to schedule a **monthly refresh**.

4. Data Transformation

- Created the **Star Schema** by combining **Sales Data, Product Data, and Employee Data**.


DimDate
DateKey
FullDateAlternateKey
DayNumberOfWeek
EnglishDayNameOfW...
DayNumberOfMonth
DayNumberOfYear
DimCustomer
CustomerKey
GeographyKey
CustomerAlternateKey
Title
FirstName
MiddleName
DimReseller
ResellerKey
ResellerAlternateKey
ResellerName
NumberEmployees
YearOpened
DimProduct
ProductKey
ProductAlternateKey
EnglishProductName
StandardCost
Color
Size
FactInternetSales
ProductKey
OrderDateKey
DueDateKey
ShipDateKey
CustomerKey
CurrencyKey
FactResellerSales
ProductKey
OrderDateKey
DueDateKey
ShipDateKey
ResellerKey
EmployeeKey
DimCurrency
CurrencyKey
CurrencyAlternateKey
CurrencyName
DimSalesTerritory
SalesTerritoryKey
SalesTerritoryAlternateK...
SalesTerritoryRegion
SalesTerritoryCountry
SalesTerritoryGroup
DimEmployee
EmployeeKey
EmployeeNationalID...
FirstName
LastName
Title
HireDate


DimEmployee
EmployeeKey
EmployeeNationalIDAlternateKey
FirstName
LastName
Title
HireDate
BirthDate
EmailAddress
Phone
MaritalStatus
Department
StartDate
EndDate
Status
SalesTarget
Year
MonthName
SalesOrder
FactResellerSales
ProductKey
OrderDateKey
DueDateKey
ShipDateKey
ResellerKey
EmployeeKey
CurrencyKey
SalesTerritoryKey
SalesOrderNumber
SalesOrderLineNumber
OrderQuantity
UnitPrice
DiscountAmount
SalesAmount
OrderDate
DueDate
ShipDate
DimProduct
ProductKey
ProductAlternateKey
EnglishProductName
StandardCost
Color
Size
SizeRange
EnglishDescription
FrenchDescription
ArabicDescription
ProductSubcategoryCode
ProductcategoryCode
Status
DimDate
DateKey
FullDateAlternateKey
DayNumberOfWeek
EnglishDayNameOfWeek
DayNumberOfMonth
DayNumberOfYear
WeekNumberOfYear
EnglishMonthName
MonthNumberOfYear
CalendarQuarter
CalendarYear
CalendarSemester
FiscalQuarter
FiscalYear
FiscalSemester
DimSalesTerritory
SalesTerritoryKey
SalesTerritoryAlternateKey
SalesTerritoryRegion
SalesTerritoryCountry
SalesTerritoryGroup
DimReseller
ResellerKey
ResellerAlternateKey
ResellerName
NumberEmployees
YearOpened
FactInternetSales
ProductKey
OrderDateKey
DueDateKey
ShipDateKey
CustomerKey
CurrencyKey
SalesTerritoryKey
SalesOrderNumber
SalesOrderLineNumber
OrderQuantity
UnitPrice
DiscountAmount
SalesAmount
TaxAmt
OrderDate
DueDate
ShipDate
DimCurrency
CurrencyKey
CurrencyAlternateKey
CurrencyName
DimCustomer
CustomerKey
GeographyKey
CustomerAlternateKey
Title
FirstName
MiddleName
LastName
NameStyle
BirthDate
MaritalStatus
Suffix
Gender
EmailAddress
YearlyIncome
TotalChildren
NumberChildrenAtHome
NumberCarsOwned
AddressLine1
AddressLine2
Phone
DateFirstPurchase
CommuteDistance

5. Semantic Layer

- ETL packages are deployed using the **SSIS Service Catalogue** for centralized management.
- Data is processed in SSAS Tabular Mode to create an OLAP cube.

6. Visualization and Reporting

- **Power BI** is connected to the data model to retrieve data for further analysis.