

# ADVENTUREWORKS2022 SSIS ETL DOCUMENTATION

## Project Overview

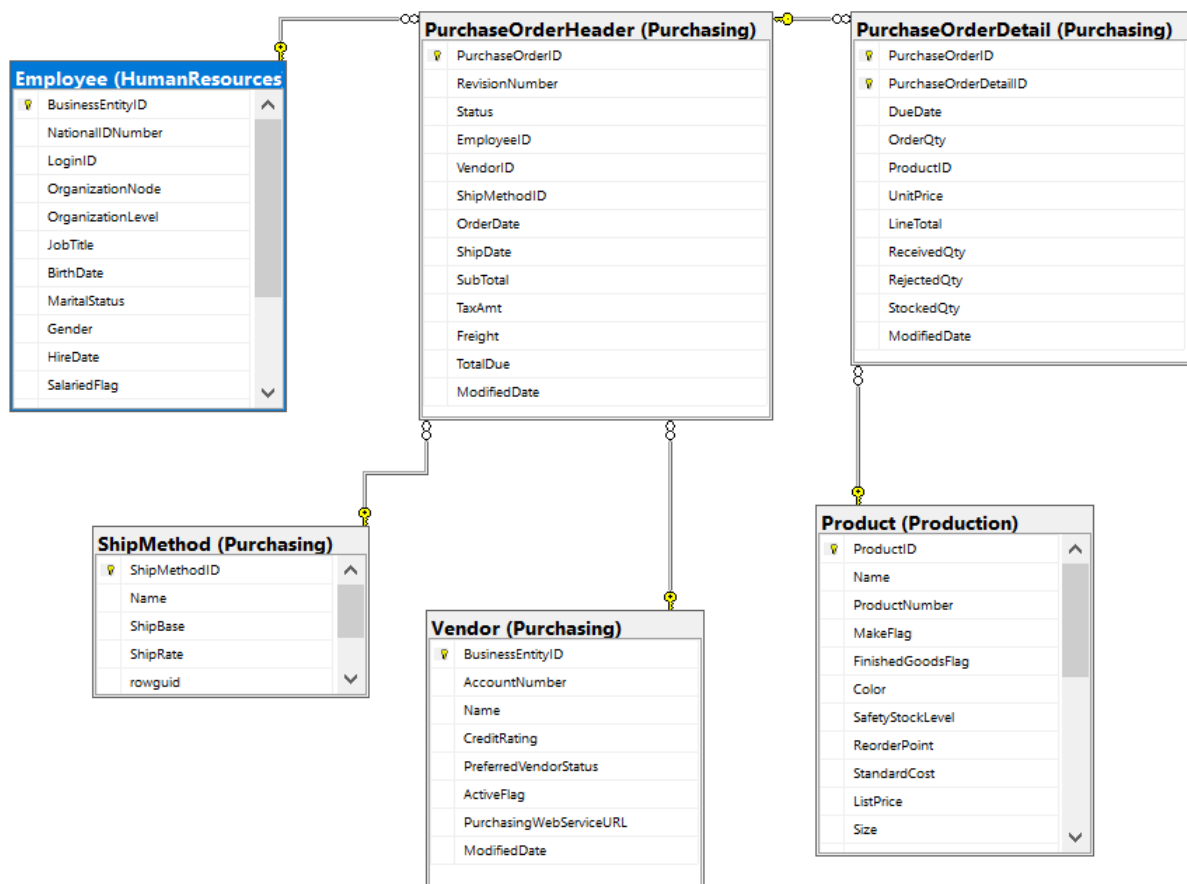
- Project Title: AdventureWorks2022 ETL to Data Warehouse
- Developers: Sahar Marzougui and Chaima Ben Abdallah
- Date of Documentation: 23/12/2023

## Executive Summary

This document articulates the intricacies of the ETL (Extract, Transform, Load) process facilitated through SQL Server Integration Services (SSIS) for the AdventureWorks2022 dataset. The primary objective is to systematically extract, transform, and load data into a Data Warehouse, optimizing it for sophisticated business intelligence reporting.

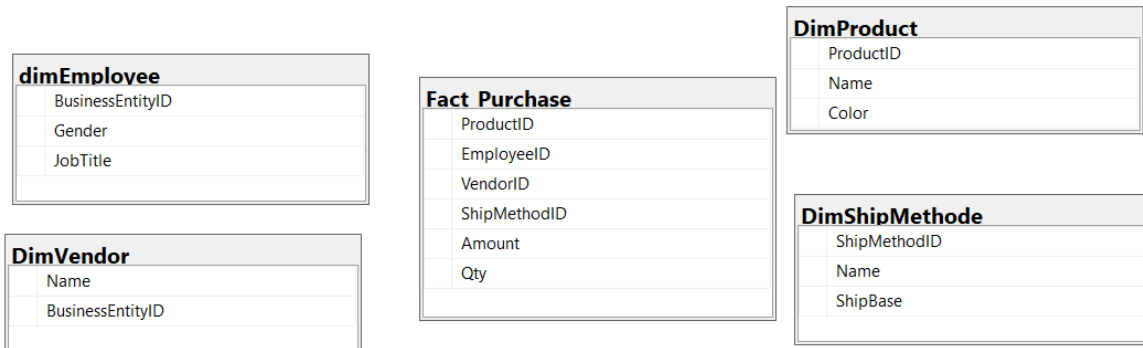
## Data Source Analysis

The source data emanates from the structured **AdventureWorks2022** dataset within Microsoft SQL Server Management Studio. A meticulous analysis was conducted to discern tables relevant to the creation of the Data Warehouse.



## Data Warehouse Design

The Data Warehouse architecture is meticulously structured to provide a resilient framework for comprehensive business intelligence reporting. Central to this design are the Fact Table (FactPurchase) and Dimension Tables (DimEmployee, DimProduct, DimVendor, DimShipMethod).



## SSIS Package Overview

### Purpose

The SSIS package serves as the orchestrator for the ETL process, seamlessly integrating tasks for data extraction, transformation, and loading into the Data Warehouse.

### Components

- Fact Table (FactPurchase): Utilizes advanced transformations including Lookup and Aggregate Transformations for precise data integration.
- Dimension Tables: Independently loaded processes for DimEmployee, DimProduct, DimVendor, and DimShipMethod.

## Fact Table: FactPurchase

### Structure

The FactPurchase table captures essential purchasing information, featuring key columns such as:

- ProductID
- VendorID
- EmployeeID
- ShipMethodID
- Qty
- Amount

### Transformation Approach

The population of the FactPurchase table employs Lookup and Aggregate Transformations, executing complex joins and aggregations aligning with business requirements.

## Dimension Tables

- DimEmployee: BusinessEntityID, JobTitle, Gender
- DimProduct: ProductID, Name, Color
- DimVendor: BusinessEntityID, Name
- DimShipMethod: ShipMethodID, Name, ShipBase

## ETL Process Steps

### 1. Truncate Tables

A singular Execute SQL Task initiates the truncation of dimension tables (DimEmployee, DimProduct, DimVendor, DimShipMethod) and the Fact Table (FactPurchase) at the commencement of the ETL process.

### 2. Fact Table Load

The FactPurchase table is populated with a sophisticated combination of Lookup and Aggregate Transformations, ensuring optimal performance and data integrity.

### 3. Dimension Table Loads

Each dimension table (DimEmployee, DimProduct, DimVendor, DimShipMethod) undergoes an independent loading process, leveraging appropriate transformations for efficiency.

## Error Handling and Logging

### Logging Mechanism

Error handling and logging are seamlessly integrated using SSIS logging features. The package logs errors, warnings, and execution status changes to a dedicated log file (SSIS\_LogFile.txt).

### Testing and Validation

Thorough testing scenarios encompass normal execution, error scenarios, and edge cases. The SSIS package has undergone extensive validation to align with predetermined expectations.

## Deployment

### Deployment Environment

The SSIS package is deployed to the SQL Server Integration Services Catalog, providing a centralized and secure environment for package execution.

## Conclusion

The SSIS ETL process has been meticulously engineered to seamlessly integrate data from the AdventureWorks2022 dataset into a well-architected Data Warehouse. This robust foundation positions the organization for enhanced business intelligence reporting capabilities.