**Data warehouse architecture design**

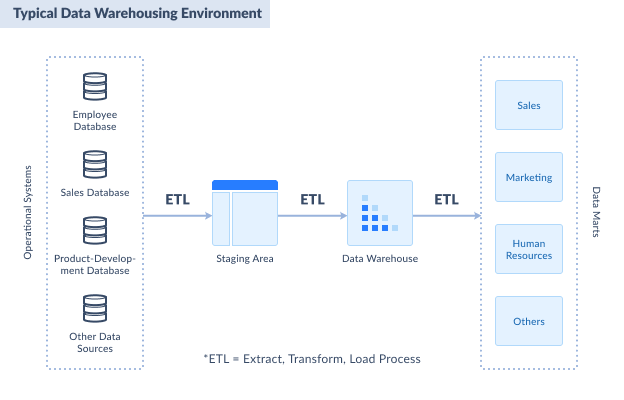


Figure 1.Typical Data Warehousing environment (Sanchez, 2021)

**Design considerations**

Staging Area:

1. Consolidating several databases in one truth source.
2. Consolidate all the needed data from the source into a single format.
   1. Clarify some parts of the data from the stakeholders.
3. Clean the data in the staging area.
   1. Identify the steps to clean the data.
4. Makes precalculation of aggregates, complex calculations, or summarization before reaching the data warehouse.
5. The staging area can be implemented in SQL Server for simplicity.
6. The staging area will have a copy of the transactional database for the tables and columns needed.
   1. Inspect the source transactional database, and identify the tables and columns you need after seeing the source database. You don’t need everything.
7. Create the staging area database.
8. Plan to extract data from the source to the staging area.

Data warehouse:

1. The data warehouse is the SQL Server database with a dimensional model.
2. The data warehouse will use a star schema.

Data marts:

1. A data mart focuses on one aspect of the business, like sales, purchasing, and more.

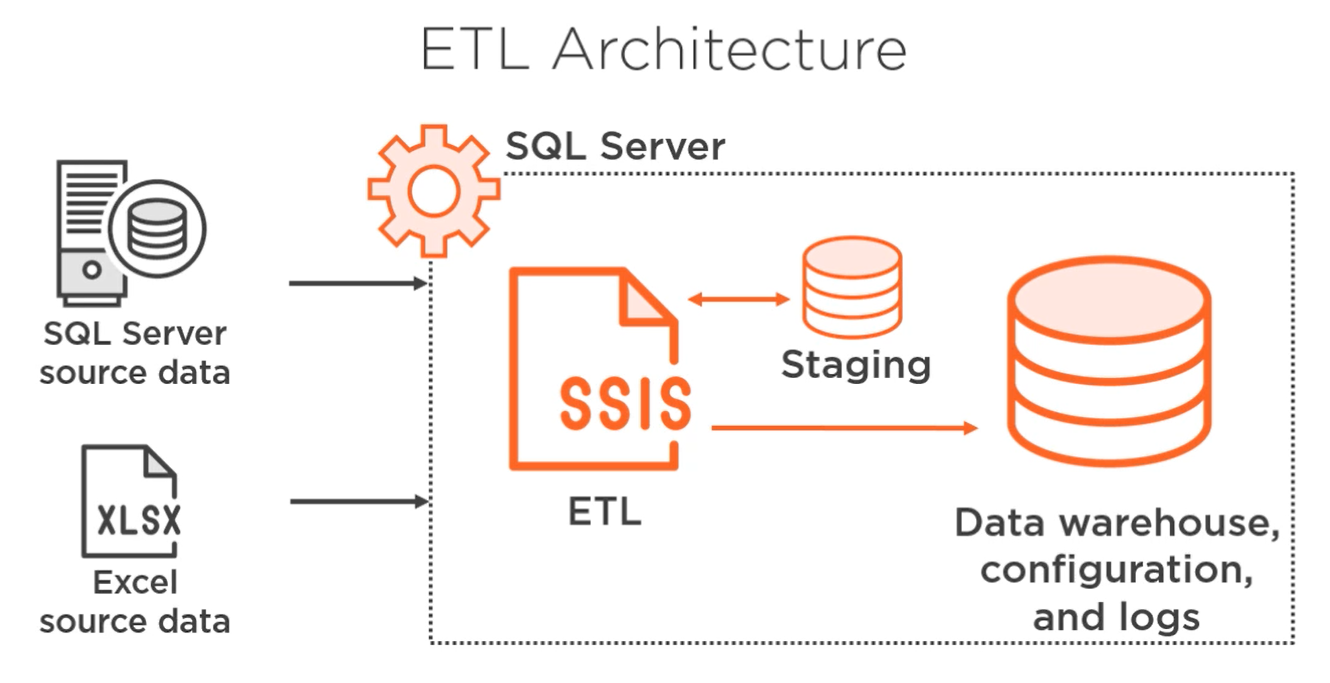
Step 1: Get Business Requirements

**ETL Requirements**

1. Summary

This document describes the ETL requirements for the Adventure Works Data Warehouse project to support sales analysis. The ETL processes for this project iteration must perform one historical extraction and daily incremental extractions of product, customer, and sales data for loading into a dimensional model.

1. Standards
   1. ETL Architecture



Data extraction: document the location of data sources, and how data is stored (databases, flat files, spreadsheets). Also, decide whether the extraction process will run on the servers hosting the data or on a separate server.

Transformation processing: Transformations to clean and restructure the data to match the target schema. Define where those steps happen.

Data load: Target system that needs to be loaded with data.

Also, decide whether the load process runs on the target server or elsewhere.

Process management: Orchestration of the jobs that need to run, configuration management for those jobs, and logging the results of execution.

* 1. Auditing and logging framework
  2. System availability requirements
  3. Process standards
  4. Extraction framework
  5. Slowly changing dimension handling
  6. Business rules validation strategy
  7. Notification requirements

**References**

Sanchez, E. (2021). SQL Server Data Warehouse: The Easy and Practical Guide. Retrieved from <https://blog.skyvia.com/sql-server-data-warehouse-the-easy-and-practical-guide/>