D608 – Data Processing (Task 1)

Morrell J. Parrish

Western Governors University

Table of Contents

D608 – Data Processing (Task 1)

Part I: Business Case Analysis

A. Business Case Overview

1. Problem Statement

Precision Components Inc. is facing a significant data integration challenge due to its acquisition of SmllFirm, Inc. The acquired company uses disparate and outdated data systems, including MS Access databases and Excel spreadsheets. The lack of a centralized data repository leads to inefficiencies, data silos, and inconsistent reporting (Western Governors University, n.d.).

Key business requirements include:

- **Data Integration**: Merging SmallFirm, Inc.'s data into Precision Components Inc.'s on-premises ERP and hosted HR/payroll system.
- **Data Cleansing**: Ensuring consistency, removing duplicates, and standardizing data formats.
- **Daily/Nightly Data Ingestion**: Implementing automated processes for data updates.
- **Unified Sales Data**: Combining data from different departments for a comprehensive view.
- **Scalability**: Preparing for international expansion and increased data volume.
- **Security & Compliance**: Ensuring data governance adheres to industry standards.

2. Source-to-Target Data Mapping

a. Define the attributes

The following datasets must be integrated from SmallFirm, Inc.:

- **Payroll Data (MS Access)**: Employee salaries and pay dates.
- **Personnel Data (MS Access)**: Employee names, positions, and hire dates.

- **Vendor Data (MS Access)**: Vendor names and account representatives.
- **Products Data (Excel)**: Product details, including costs and pricing.
- **Production Batch Data (Excel)**: Batch numbers, production dates, and quantities.
- **Tooling & Raw Materials Inventory (Excel)**: Inventory levels, vendor IDs, and costs.

b. Mapping the attributes

Source (SmallFirm, Inc.)	Target (Precision Components Inc.)	Transformation Required
EmployeeId (payroll)	Emp_ID (HR System)	Data type normalization
Salary (payroll)	Annual_Salary (HR System)	Currency Conversion
PayDate (payroll)	Payment_Date (HR System)	Format Standardization
ProductID (products)	Item_ID (ERP)	No change
Name (products)	Item_Name (ERP)	No change
SalePrice (products)	Retail_Price (ERP)	Rounding adjustments
BatchNumber (production)	Batch_ID (ERP)	Format Standardization
ToolName (tooling)	Equipment_Name (ERP)	No change

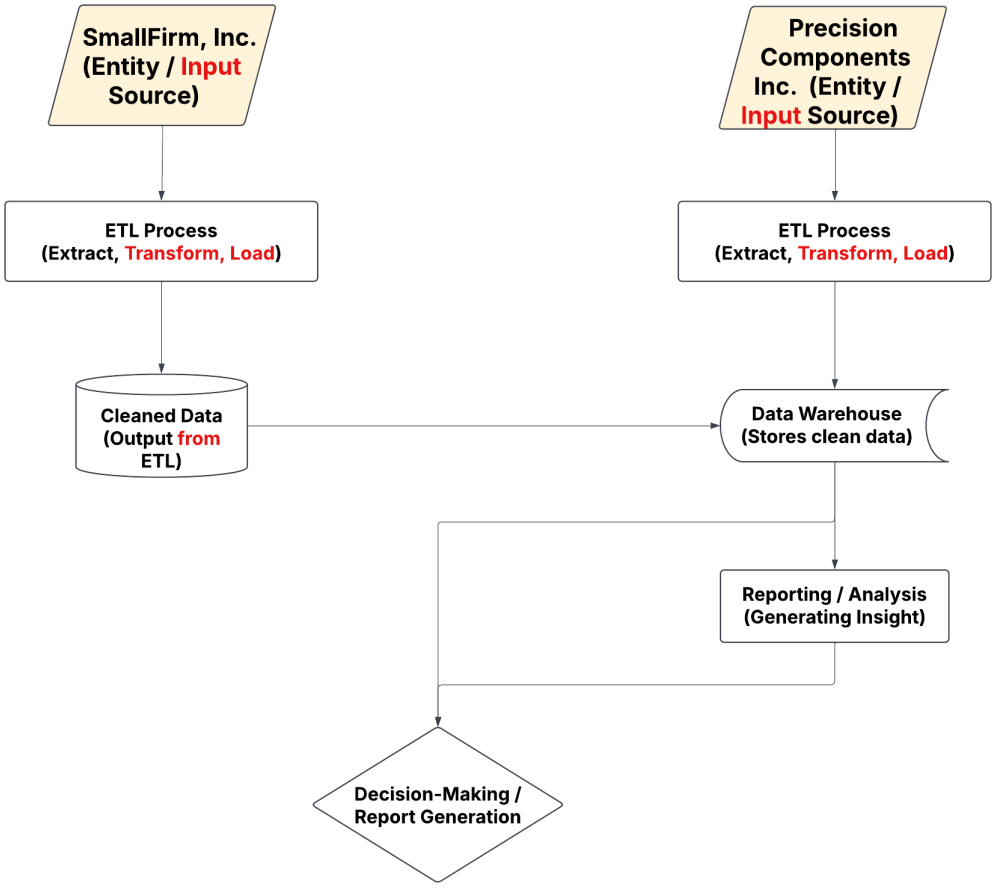
c. Data Transformations

- **Data Cleansing**: Standardizing date formats, removing duplicates, and correcting missing values.
- **Data Normalization**: Ensuring consistent data types across systems.
- **Data Aggregation**: Summarizing sales data for reporting.

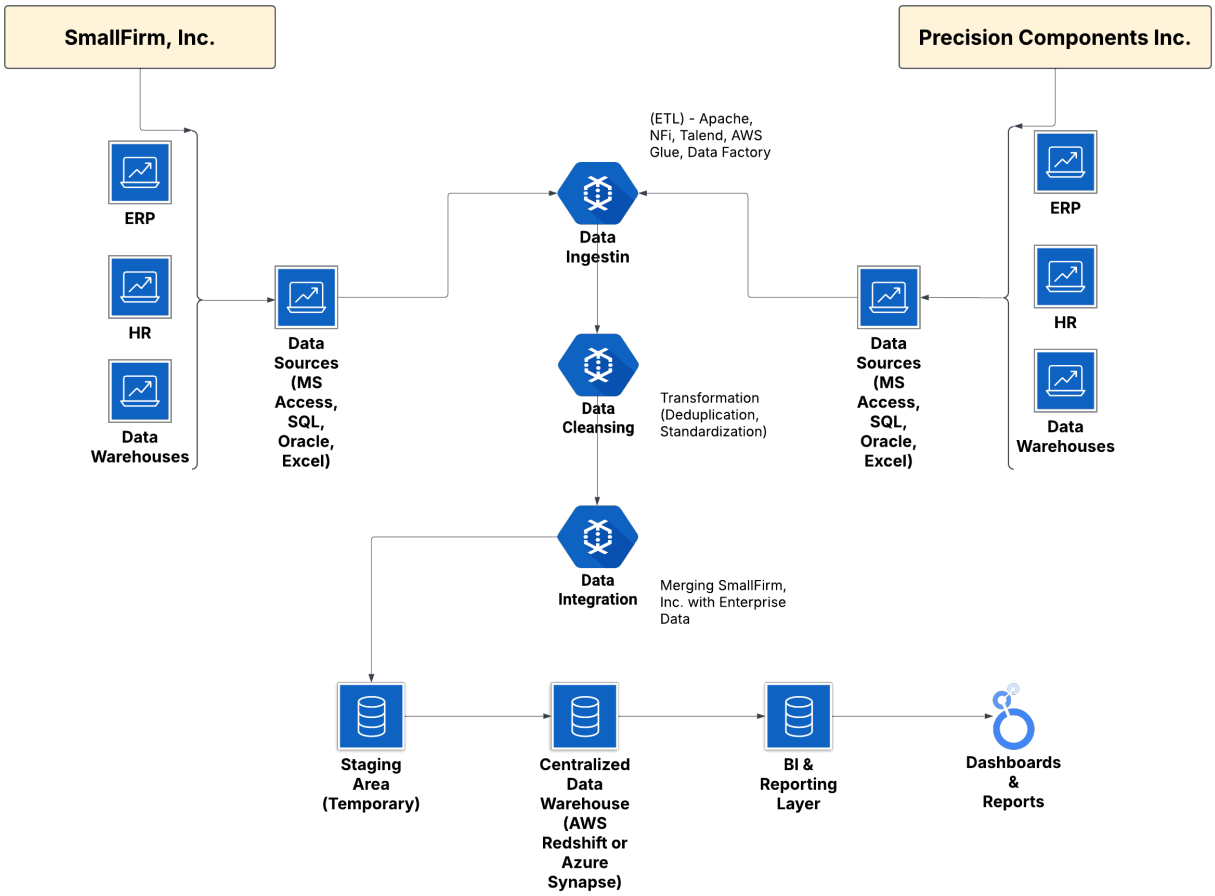
Part II. Recommended Design Solution

B. Data Engineering Design

1. Process Flow Diagram



2. Data Flow Diagram (Level 1 or Higher)



Part III. Processing Design Evaluation

C. Relevance to Stakeholders

1. Advantages of the Proposed Approach

The recommended data engineering design offers several benefits compared to alternative approaches:

- **Scalability & Future Readiness**

  The cloud-based data warehouse (e.g., AWS Redshift, Google BigQuery, or Azure Synapse Analytics) ensures that the system can scale as the company expands

internationally. Unlike on-premises solutions, cloud infrastructure allows for dynamic resource allocation based on business needs.

- **Automated Data Integration & Cleansing**

  The use of ETL tools (such as AWS Glue, Apache NiFi, or Talend) automates the ingestion, transformation, and cleansing of data. This reduces manual effort and enhances data consistency compared to a manual or ad-hoc integration approach.

- **Holistic Business Insights**

  By consolidating data from the ERP, HR/payroll, and home-grown systems into a unified data warehouse, business intelligence (BI) tools (e.g., Power BI, Tableau) can provide real-time analytics and reporting, giving Precision Components Inc. a competitive advantage.

2. Disadvantages of the Proposed Approach

While the proposed solution is robust, there are some potential challenges:

- **Implementation Complexity**

  Migrating legacy MS Access databases and Excel files to a modern data warehouse involves data transformation challenges. Mapping data fields and resolving inconsistencies may require significant effort.

- **Cost Considerations**

  Cloud-based infrastructure, ETL tools, and BI solutions incur ongoing operational expenses. Compared to maintaining an on-premises SQL Server, cloud pricing models can fluctuate based on storage and compute usage.

Part IV. Findings and Recommendations

To address the data integration challenges at Precision Components Inc., the following solution is recommended:

- **Data Warehouse Implementation**

  Deploy a centralized cloud-based data warehouse to unify disparate data sources.

- **ETL Automation**

  Utilize ETL tools to extract, clean, and merge data from SmallFirm, Inc. into the ERP and HR/payroll systems.

- **Dashboard & Reporting Integration**

  Implement BI tools (e.g., Tableau, Power BI) for real-time business insights.

- **Data Governance & Security**

  Establish role-based access controls and compliance measures to safeguard sensitive business data.

In conclusion, the data engineering design outlined in this report provides a comprehensive solution to the challenges faced by Precision Components Inc. through the integration of SmallFirm, Inc.'s data. By leveraging cloud-based infrastructure, automated ETL processes, and business intelligence tools, the solution ensures scalability, operational efficiency, and the capacity for real-time business insights. These advantages align with Precision Components Inc.'s long-term goals for growth and enhanced decision-making. However, attention must be given to the complexity of migrating legacy data systems and the associated costs. Addressing these challenges will provide the company with a robust data environment capable of supporting its evolving business needs. As emphasized by HubSpot, creating a data

flow diagram to visualize this process is an essential first step in understanding and managing data integration, making it a vital part of this solution (HubSpot, n.d.).

References

HubSpot. (n.d.). *Data flow diagram: What it is and how to create one*. HubSpot. Retrieved February 15, 2025, from https://blog.hubspot.com/marketing/data-flow-diagram

Western Governors University. (n.d.). *FDN1 Task 1: Data Processing Solution.* Retrieved February 8, 2025, from https://tasks.wgu.edu/student/000194226/course/38080021/task/4698/overview