**Project Title: "Data Integration and ETL Pipeline for a Retail Company"**

**Project Overview:**

The goal of this project is to design and implement an end-to-end data integration and ETL (Extract, Transform, Load) pipeline for a fictional retail company. The company has various data sources, such as sales data, inventory data, customer data, and product data, which are stored in different formats and locations. The project involves extracting data from these sources, transforming and cleansing it, and loading it into a central data warehouse for further analysis and reporting.

**Project Phases and Tasks:**

**Phase 1: Data Analysis and Requirements Gathering**

* Understand the business requirements of the retail company and define the scope of the project.
* Analyse the existing data sources and identify the data to be extracted for integration.
* Define data quality and validation rules for data cleansing.

**Phase 2: Data Extraction**

* Design and implement data extraction processes using Informatica, Azure Data Factory, or IBM DataStage, depending on your preference and expertise.
* Extract data from various sources, such as CSV files, Excel files, and databases, using appropriate techniques and tools.
* Perform data profiling and validation to ensure data integrity.

**Phase 3: Data Transformation and Cleansing**

* Design and implement data transformation processes using SQL and Python.
* Cleanse and normalize the data, handle missing values, and perform data enrichment as needed.
* Apply data validation rules defined in Phase 1 to ensure data quality.

**Phase 4: Data Loading**

* Design and implement data loading processes to load the transformed data into a central data warehouse or a cloud-based data lake using Snowflake, Azure Synapse Analytics, or other similar technologies.
* Optimize data loading for performance and scalability.
* Monitor and troubleshoot data loading processes for any errors or issues.

**Phase 5: Data Integration and Orchestration**

* Design and implement data integration and orchestration workflows using Informatica, Azure Data Factory, or IBM DataStage, depending on your chosen technology stack.
* Automate data integration processes and schedule them to run at regular intervals.
* Handle error handling, logging, and notifications for data integration workflows.

**Phase 6: Reporting and Analytics**

* Design and implement reporting and analytics solutions using SQL and visualization tools such as Power BI or Tableau.
* Create dashboards and reports to provide insights and actionable information to the retail company's stakeholders.
* Perform data analysis and data mining to derive meaningful insights from the integrated data.

**Project Deliverables:**

* Detailed project plan outlining the phases, tasks, and timelines.
* Documentation of the data integration and ETL pipeline, including data mapping, transformation rules, and data validation rules.
* Complete source code and scripts for data extraction, transformation, loading, and integration processes.
* Documentation on data validation results and data quality metrics.
* Reports and dashboards showcasing the integrated data and insights generated from it.