

Outcomes to Displayed

Outcome: Successfully extracted customer information, order details, and product data from the MySQL database using Sqoop.

Deliverables:

Sqoop command(s) for extracting the data from the MySQL database.

CSV or Parquet files containing the extracted data.

Task: Data Migration from SQL to Hive

Outcome: Successfully transfer the extracted data from SQL to Hive using Sqoop.

Deliverables:

Sqoop command(s) for migrating the data from SQL to Hive.

Hive database and table creation scripts.

Task: Data Processing and Transformation in Hive

Outcome: Perform data processing and transformation on the migrated data in Hive.

Deliverables:

HiveQL queries for data cleansing, preprocessing, and enrichment.

Scripts or commands for executing the HiveQL queries.

Task: Data Quality Checks and Validation

Outcome: Ensure data integrity and perform quality checks on the migrated and processed data.

Deliverables:

Data quality check scripts or queries to validate the data.

Reports or logs indicating the results of the data quality checks.

Task: Timely Execution of the Data Pipeline

Outcome: Set up a scheduled job or workflow to execute the data migration and processing pipeline at specified intervals.

Deliverables:

Job or workflow definition files (e.g., Apache Airflow DAG, cron job script).

Configuration files specifying the scheduling frequency and pipeline execution details.

Task: Data Visualization

Outcome: Generate visualisations to showcase customer behaviour insights using the processed data.

Deliverables:

Visualization artefacts (reports, dashboards) created using Apache Superset, Tableau, or Power BI.

SQL queries or scripts are used to extract the necessary data for visualization.

Task: Documentation and Presentation

Outcome: Prepare detailed documentation and present the project implementation and outcomes.

Deliverables:

Project documentation describing the architecture, data flow, and implementation steps.

Presentation slides highlighting the project objectives, challenges, and achieved outcomes.