

Project

Order Data ETL Pipeline with Hive and Sqoop

Overview:

The project focuses on building an **ETL pipeline** for order data stored in a MySQL database.

The goal is to process and analyze the data using Hive, followed by exporting the processed data back to MySQL for reporting and insights.

Steps :

- Import Data from MySQL to HDFS
- import the orders table from the MySQL database to HDFS.
- Create a Hive Database
- Create a Hive Managed Table
- Load Data into Hive Table
- Data Transformation with Hive - Partition the Data by order_status
- Perform Analytics with Hive Queries
- Find total orders by status
- filter orders with the status as PENDING
- Incremental Data Import with Sqoop
- Insert new records into the MySQL orders table - Import only new records into HDFS
- Create a table in MySQL to store processed data
- Export processed data from HDFS(hive) to MySQL