

Data Warehousing with IBM Cloud Db2 Warehouse

921821104051: U. Yasvindhini

PHASE -3 Development part 1

Consider two data sources: a CSV file containing sales data and a relational database with customer information.

1. Sales Data CSV:

- Table Name: **SalesData**
- Columns:
 - **TransactionID** (Primary Key)
 - **ProductID**
 - **Date**
 - **QuantitySold**
 - **TotalAmount**

Sample code:

Transaction ID,Product ID,Date,Quantity Sold,Total Amount
1,P001,2023-10-18,10,250.00
2,P002,2023-10-19,8,180.00
3,P003,2023-10-20,15,375.00

2. Customer Database:

- Table Name: **CustomerInfo**
- Columns:
 - **CustomerID** (Primary Key)
 - **FirstName**
 - **LastName**
 - **Email**
 - **Phone**

Sample code:

Customer ID,First Name,Last Name,Email,Phone
101,John,Doe,john.doe@email.com,123-456-7890
102,Jane,Smith,jane.smith@email.com,987-654-3210
103,Robert,Johnson,robert.johnson@email.com,456-789-0123

3. Integration Strategy:

- Establish a common key between the tables (e.g., **CustomerID** in both).
- Create a new table, say **IntegratedData**, linking the two:
 - Columns:

- **TransactionID** (Primary Key)
- **ProductID**
- **Date**
- **QuantitySold**
- **TotalAmount**
- **CustomerID** (Foreign Key)

4. Loading Data:

- Use Db2's LOAD utility to import CSV data into the **SalesData** table.
- Execute SQL statements to insert data from the Customer Database into the **CustomerInfo** table.

5. Integration Process:

- Write SQL queries to join tables based on the common key (**CustomerID**).
- Populate the **IntegratedData** table with relevant information from both sources.

6. Testing:

- Run sample queries to ensure accurate integration.
- Validate relationships and check for any data inconsistencies.