**Phase 3: Development Part 1**

**In this part you will begin building your project.**

**Start building the data warehouse using IBM Cloud Db2 Warehouse.**

To begin building the data warehouse using IBM Cloud Db2 Warehouse, we'll first define the schema and structure of the data warehouse tables. Then, we'll identify data sources and design a strategy to integrate them into the data warehouse.

**1. Define Schema and Structure of Data Warehouse Tables:**

Let's create a simple example schema and structure for demonstration purposes. Assume we're building a data warehouse to store sales data for a retail company.

**Schema:**

* **Sales Data Warehouse:**
* **Dimension Tables:**

**1.Customer** (customer\_id, name, email, phone, address)

**2.Product** (product\_id, product\_name, category, price)

**3.Time** (date\_id, date, day\_of\_week, month, year)

* **Fact Table:**

**1.Sales** (sale\_id, customer\_id, product\_id, date\_id, quantity, total\_amount)

**2.Identify Data Sources and Design Integration Strategy:**

Data Sources:

**1.CSV Files:**

* CSV files containing customer data, product data, time data, and sales data.

**2.External Databases:**

* External databases with additional customer and product information.

Integration Strategy:

**1.CSV Data Integration:**

* Utilize Db2 Warehouse's data import capabilities to load CSV files into corresponding tables.
* Map CSV columns to appropriate columns in the data warehouse tables.

**2.External Database:**

* Establish a connection between Db2 Warehouse and the external databases containing additional customer and product information.
* Use SQL scripts and ETL (Extract, Transform, Load) processes to fetch and transform data from the external databases into the appropriate format and load it into the respective data warehouse tables.

**3.Regular Data Updates:**

* Implement scheduled ETL jobs to periodically update the data warehouse with fresh data from the CSV files and external databases.
* Use incremental loading techniques to only load new or modified records, minimizing the load time and optimizing efficiency.

**4.Data Quality and Transformation:**

Implement data quality checks and transformations during the ETL process to ensure consistency, accuracy, and completeness of the integrated data.

By following this schema definition and integration strategy, we can efficiently build and maintain the data warehouse using IBM Cloud Db2 Warehouse, integrating data from various sources to provide a unified and valuable analytical platform.



