# Data Warehousing Project Report

**PRN: 225100111**

**Name: Daivik Karbhari**

**1. Introduction:**

## Objective and Scope of the Project

The objective of this project is to design and implement a data warehouse that enables efficient storage, retrieval, and analysis of business data. The scope includes integrating various data sources, applying ETL (Extract, Transform, Load) processes, designing a star schema, implementing data cubes, and generating OLAP (Online Analytical Processing) reports for business intelligence.

**2. Business Requirement:**

## Application Specification of Data Warehousing for Users

The data warehouse is designed to meet the needs of business users, analysts, and decision-makers by providing a centralized repository of historical and real-time data. It supports:

* Sales analysis and forecasting
* Customer segmentation and behavior analysis
* Inventory management and stock tracking
* Store performance and revenue monitoring
* Efficient reporting and visualization of business insights

**3. Functional Specification:**

## Input and Output Specification of Data Warehousing

***Input Data Sources:***

* **Customer Data:** Customer ID, Name, City, Type, First Order Date
* **Store Data:** Store ID, City, Phone
* **Item Data:** Item ID, Description, Size, Weight, Unit Price
* **Order Data:** Order Number, Store ID, Item ID, Customer ID, Quantity, Ordered Price, Order Date
* **Headquarter Data:** City ID, Headquarter Address, State

***Output Reports:***

* Sales analysis reports by store and item
* Customer order trends
* Stock availability reports
* Revenue and profit tracking

**4. Data Warehousing Design:**

## Stepwise Methodology of Designing the Data Warehouse

1. **Requirement Analysis** - Understanding business objectives and identifying key metrics.
2. **Data Source Identification** - Collecting data from transactional databases.
3. **ETL Process** - Extracting data, transforming it into a suitable format, and loading it into the warehouse.
4. **Schema Design** - Implementing a **Star Schema**:
   1. **Fact Table:** FactSales (Order\_no, Store\_id, Item\_id, Customer\_id,

Quantity\_ordered, Ordered\_price, Order\_date\_id)

* 1. **Dimension Tables:** DimCustomer, DimStore, DimItem, DimHeadquarter, DimDate

1. **Data Cube Construction** - Creating OLAP cubes for multidimensional analysis.
2. **Data Validation and Testing** - Ensuring accuracy and consistency of the data.

**5. Data Cube Implementation:**

## Computer Automation of Implementing Data Warehouse and Loading Data into Data Cubes

* **Step 1:** Implement the schema in a database system (MySQL, PostgreSQL, etc.).
* **Step 2:** Load transformed data into the data warehouse.
* **Step 3:** Create OLAP cubes using tools like Microsoft Analysis Services, Apache Kylin, or Pentaho BI.
* **Step 4:** Perform slice, dice, drill-down, and roll-up operations on the cube.

**6. Observations:**

## (a) Online Analytical Processing Reports

* Invoking SQL queries or BI tools to generate interactive reports.
* Example SQL query for sales analysis: SELECT Store\_id, SUM(Ordered\_price) AS Total\_Revenue

|  |  |  |
| --- | --- | --- |
| FROM |  | FactSales |
| GROUP | BY | Store\_id; |

* Visualization using Power BI or Tableau.

## (b) Data Verification

* Compare OLAP report values with raw data from relational tables.
* Cross-check total revenue from FactSales with transactional sales data.
* Validate customer orders with DimCustomer and FactSales data.

**7. Conclusion:**

The data warehouse implementation successfully centralized business data, improved analytical capabilities, and enabled better decision-making through OLAP reporting. The integration of a star schema with fact and dimension tables ensured efficient querying and storage. Future enhancements may include real-time data processing, machine learning-based insights, and automated reporting features.