**ABSTRACT**

Logistics is the support function of an organization and it means having the right object, at the right place, in the right time. Logistics deals with various kinds of methods to control the flow of resources from one place to another. One of the major and the most important factors that is costing is being dealt with utmost attention. The project is being designed keeping in attention the details of the various requirements of logistics such as keeping records of the goods i.e. their details and the kind of content that is stored in the shipment which is to be delivered.

A Relational Database Management System (RDBMS) is similar to DBMS. The difference is that in RDBMS, the entities and values in tables are related to one another. Also the tables are related to each other. Thus, it is called “Relational”.

### **PROBLEM DESCRIPTION**

The logistics company (GLlogistics) provides services in both the international and domestic sectors. The logistics management takes into consideration every facility that has an impact on cost. It plays an important role in making the product conform to customer requirements. Also it involves efficient integration of suppliers, manufacturers, Import & export and other activities at many levels, from the strategic level through the tactical to the operational level.

Create a database schema and table relationships that can be used in any technology. Perform data pre-processing and Exploratory data analysis.

**Attention:**

Below can be solved, if needed. (Not mandatory, it’s up to the interest of the students)

Customers can send different types of shipping contents. Payment is to be paid at the same time the product is delivered to the client. Delivery person and centre head can update the status of the shipment.

### 

### **SCOPE**

It is of critical importance to the organization how it delivers products & services to the customer, whether the product is tangible or intangible. Effective and efficient physical movement of the tangible product will speak of intangible services associated with the product and the organization which is delivering it.

In case of intangible products, the delivery of tangibles at the right place & right time will speak about its quality. On the macro level infrastructure such as various modes of transport, transportation equipment, storage facilities, connectivity and information processing are contributing to a large extent in the physical movement of goods produced in manufacturing, mining and agriculture Sectors.

This speed and reliability in distribution of products and services contribute to a great extent in the growth of a country’s domestic and international trade.

### **TABLE DEFINITIONS**

1. Employee Table:

This table contains the information of the employees.

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| E\_ID | INT (5) | Employee ID (Primary Key) |
| E\_NAME | VARCHAR (30) | Name of the employee |
| E\_BRANCH | VARCHAR (15) | Branch name |
| E\_DESIGNATION | VARCHAR (40) | Designation of the employee |
| E\_ADDR | VARCHAR (100) | Address of the employee |
| E\_CONT\_NO | VARCHAR (10) | Contact Number of the employee |

1. Membership Table:

This table contains the membership details of the customer or client.

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| M\_ID | INT | Membership ID associated with the client (Primary Key) |
| START\_DATE | TEXT | Start date of the membership |
| END\_DATE | TEXT | End date of the membership |

1. Client Table:

This table contains the information of the customers or clients.

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| C\_ID | INT (4) | Client ID (Primary Key) |
| C\_NAME | VARCHAR (30) | Name of the client |
| C\_EMAIL\_ID | VARCHAR (50) | Email of the client |
| C\_CONT\_NO | VARCHAR (10) | Contact Number of the client |
| C\_ADDR | VARCHAR (100) | Address of the client |
| C\_TYPE | VARCHAR (30) | Type of client (Wholesale, Retail, Internal Goods) |
| Membership\_M\_ID | INT | Membership ID (Foreign Key) |

1. Payment Table:

This table contains the payment details.

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| PAYMENT\_ID | VARCHAR (40) | Payment Unique ID (Primary Key) |
| AMOUNT | INT | Price to be paid by the client |
| PAYMENT\_STATUS | VARCHAR (10) | Payment status (Paid / Not Paid) |
| PAYMENT\_DATE | TEXT | Date when payment is made by the client |
| PAYMENT\_MODE | VARCHAR (25) | Mode of payment (COD / Card Payment) |
| Shipment\_SH\_ID | VARCHAR (6) | Shipment ID (Foreign Key) |
| Shipment\_Client\_C\_ID | INT (4) | Client ID (Foreign Key) |

1. Shipment Table:

This table contains the shipment details.

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| SH\_ID | VARCHAR (6) | Shipment ID (Primary Key) |
| SH\_CONTENT | VARCHAR (40) | Type of shipping content |
| SH\_DOMAIN | VARCHAR (15) | Shipment Domain (International / Domestic) |
| SER\_TYPE | VARCHAR (15) | Service Type (Express / Regular) |
| SH\_WEIGHT | VARCHAR (10) | Shipment Weight |
| SH\_CHARGES | INT (10) | Shipment Charges |
| SR\_ADDR | VARCHAR (100) | Source Address |
| DS\_ADDR | VARCHAR (100) | Destination Address |
| Client\_C\_ID | INT (4) | Client ID (Foreign Key) |

1. Status table:

This table contains the details about the delivery status.

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| CURRENT\_ST | VARCHAR (15) | Current status of the shipment |
| SENT\_DATE | TEXT | Date when shipment was sent |
| DELIVERY\_DATE | TEXT | Date when the product was/will be delivered |
| SH\_ID | VARCHAR (6) | Shipment ID (Primary Key) |

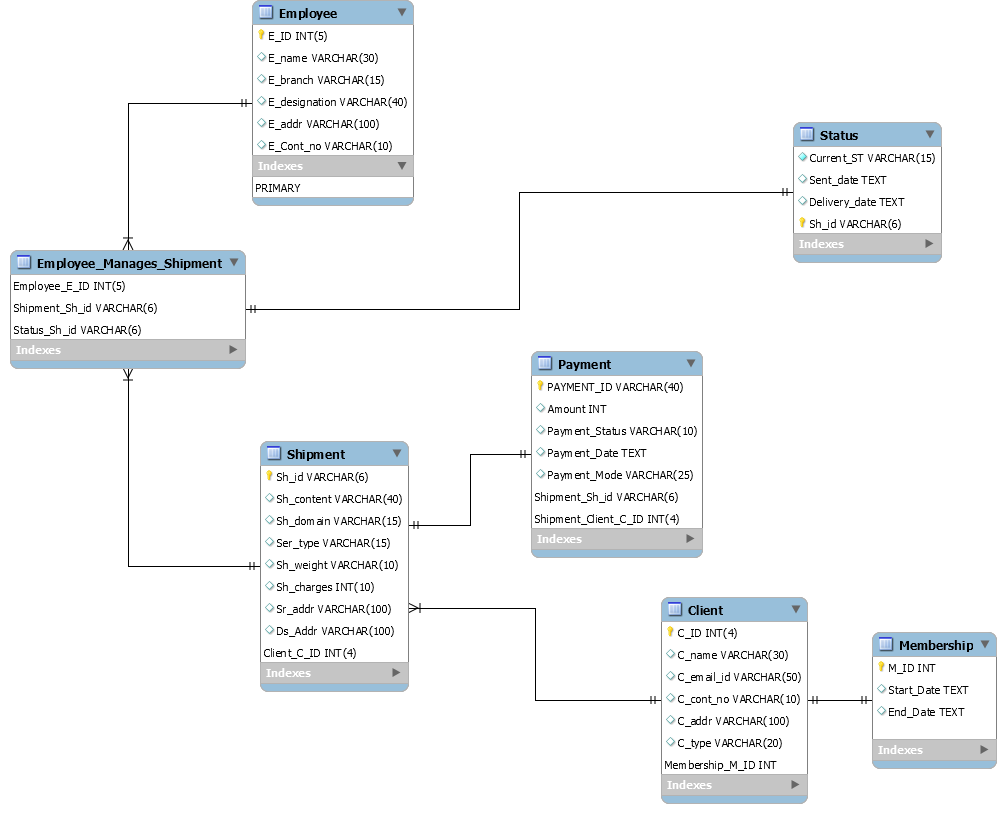
1. Employee Manages Shipment Table:

This is a relationship table between the employee and the shipment table.

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| Employee\_E\_ID | INT (5) | Employee ID (Foreign Key) |
| Shipment\_SH\_ID | VARCHAR (6) | Shipment ID (Foreign Key) |
| Status\_SH\_ID | VARCHAR (6) | Shipment\_ID from status table (Foreign Key) |

### 

### **CLASS DIAGRAM**



**DATA PREPROCESSING (Working with dates)**

It is not necessary that the dates obtained from csv files will always be in the same format. They can be separated by ‘/’ (slash) or by ‘-’. Also dates can be in any format like ‘dd-mm-yy’, ‘dd-mm-yyyy’, ‘yyyy-mm-dd’, etc and many more.

In tables PAYMENT and MEMBERSHIP, the date is in the format “%Y-%m-%d”. In the STATUS table the date is of the format “%m/%d/%Y”.

There can be some dates that are entered erroneously like 02/31/1999. There are only 28 or 29 days in the month of February. But the date ‘31’ can be wrongly entered.

Steps to perform while dealing with the dates after importing the data from csv files:

1. Look for erroneous dates

There can be dates where the month is greater than 12.

For example: Find the erroneous date from the column 'DELIVERY\_DATE' in the 'STATUS' table where the month is greater than 12.

Search for the records where the month is February but the date is erroneously entered as 30 and 31.

1. Convert the string in the date format
2. Change the data type of the column to DATE

### **CREATING A SINGLE SOURCE OF TRUTH (SSOT)**

SSOT means creating a new table by joining all the available tables. Here we will create a new table ‘LOGISTICS’ by joining the tables EMPLOYEE, CLIENT, MEMBERSHIP, SHIPMENT, PAYMENT, Employee\_Manages\_Shipment and STATUS.

### **EXPLORATORY DATA ANALYSIS (EDA)**

1. Find all the employees whose name starts with A and ends with A.
2. Find all the common names from employee names and client names.
3. Create a view of those clients who have not paid the amount.
4. What is the total percentage contribution of each Payment Mode?
5. Create a new column 'Total\_Payable\_Charges' using shipping cost and price of the product. (Use logistics)
6. What is the highest payable amount ? (Use logistics)
7. Extract the client id and the name of the clients who were or will be the member of the branch for more than 10 years (Use logistics)
8. Fetch the records which got the product delivered on the next day the product was sent? (Use logistics)
9. Which shipping content had the highest total amount (Top 5). (Use logistics)
10. Which product categories from shipment content are transferred more (Top 5)?
11. Create a new view ‘TXLogistics’ where the employee branch is Texas.
12. The Texas(TX) branch is giving a 5% discount on total payable amount. Create a new column ‘New\_Price’ for a new payable price obtained after applying a discount.
13. Drop the view TXLogistics
14. The employee branch in New York (NY) is shut down temporarily. Thus, the branch needs to be replaced with New Jersey (NJ).
15. Find the unique designations of the employees.
16. Find the frequency of each customer type (in percentage)?
17. Rename the column SER\_TYPE to SERVICE\_TYPE.
18. Which service type is preferred more?
19. Find the shipment id and shipment content where the weight is greater than the average weight.