# **Communication to DE Team**

**By Abhijna Kulal**

**Contents**

[**Communication to DE Team** 1](#_Toc175317219)

[**1.Context** 3](#_Toc175317220)

[**2.Limitation of existing data model** 4](#_Toc175317221)

[**3.Proposed new data model** 4](#_Toc175317222)

[**3.1 Assumptions made** 4](#_Toc175317223)

[**3.2 New data model** 5](#_Toc175317224)

[**3.3 Sales Conceptual data model.** 6](#_Toc175317225)

[**4.Conclusion** 7](#_Toc175317226)

# **1.Context**

Sales analytics project was developed for the sales team had many limitations with respect to the data and modelling of the data.

This document consists of the advantages of the proposed data model by specifying the significance of the same.

# **2.Limitation of existing data model**

* Existing data model don’t have any data constraints and hence future data loads will have issues. Data upload time must be added.
* Query load time will have significant impact due present data mode
* Addition of the few more datasets will have higher impact to present data architecture
* There is no referential integrity between the orders and shipping table
* There is no indexing in the data model which will have more impact.
* No handling of historical data.
* There are no detailed attributes for each table
* Frequency of the data loads and the source system must be identified. Data scalability will be a problem in the present data architecture.
* Data Security and GDPR mandate should be created for each table.

# **3.Proposed new data model**

Current data model had many limitations hence we have made few assumptions and tried to enhance the existing data model.

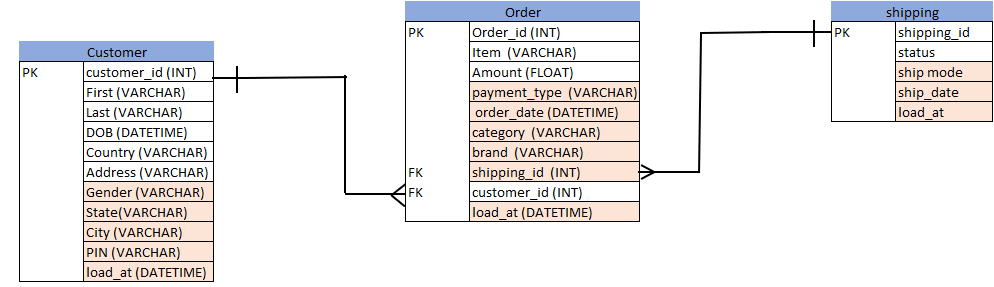
## **3.1 Assumptions made**

* Current data model doesn’t have any referential integrity between the order table and shipping. We have assumed each order will have a shipping\_id or group of orders might have the same shipping\_id. So, we can say that shipping\_id in order table will be a foreign key to join with shipping table.
* There were lot of inconsistencies in the tables that’s why we have added the loadt\_at column in each table, which would specify the data updating date and time for each table.
* There were many missing attributes in customer table. We have added DOB column instead of age and other customer details such as state, city, PIN, gender and address added.
* Orders table, we have added payment type, Order date, category, shipping\_id as well as the load\_at columns.
* Shipping table will have additional columns such as ship\_mode, ship\_date.

## **3.2 New data model**

Below diagram indicates the new data model. We have clearly mentioned the primary key constraints for three tables and their foreign key relationships.

Added few attributes which will help in understanding more about the customer demographics, sales trends, year over year growth, shipping delays which is directly proportional to customer satisfaction ratings.



The new data model can include indexing which will improve the faster query response.

Order table can be acted as fact table, which will have the measures.

SQL scripts also given for the data model.

## **3.3 Sales Conceptual data model.**

Previous proposed data model had only addition of few attributes, it’s a stop gap solution.

But in this conceptual data model, we tried identifying all the process and department involved in sales. Many datasets can be included for the whole process.

In the below conceptual model, we have added few more datasets related to payment, product and vendor.

Below example shows the flexibility of adding of datasets based on the processA diagram of a data flow

Description automatically generated with medium confidence

# **4.Conclusion**

We must maintain the data model in such a way that, it can be scalable, gives flexibility to add more tables and attributes based on the new process. Data should be secure, access given to specific the AD groups and user groups.