# Supporting Document for Data Modelling & Processes

## Introduction

What is web usage analysis? In Phase – 1 we have had done significant research and had studied various models to understand Web Usage Analysis. With this assumption that we have a strong background research we dived in Data Modelling and were extremely confused between the number of possible and potential Data Marts which could be created. There were numerous possibilities and after long discussions and readings we finally settled for previously planned Data Marts with many changes.

While making/creating these data marts we had to re-structure some tables in the OLTP and planned some other tables relating to comment section of an E-commerce website (or any other). We reached to this conclusion that we will make some changes and we will create possible other data marts and show their relevance in Web Usage Analysis and User Behavior On-Site.

Hence, we made some minor changes in OLTP, created two Data Marts Session Web Analysis and Cart Web Session, created some potential Snowflakes which can be implemented if required-be, planned a third Data Mart User Persona Through Comments Web Analysis

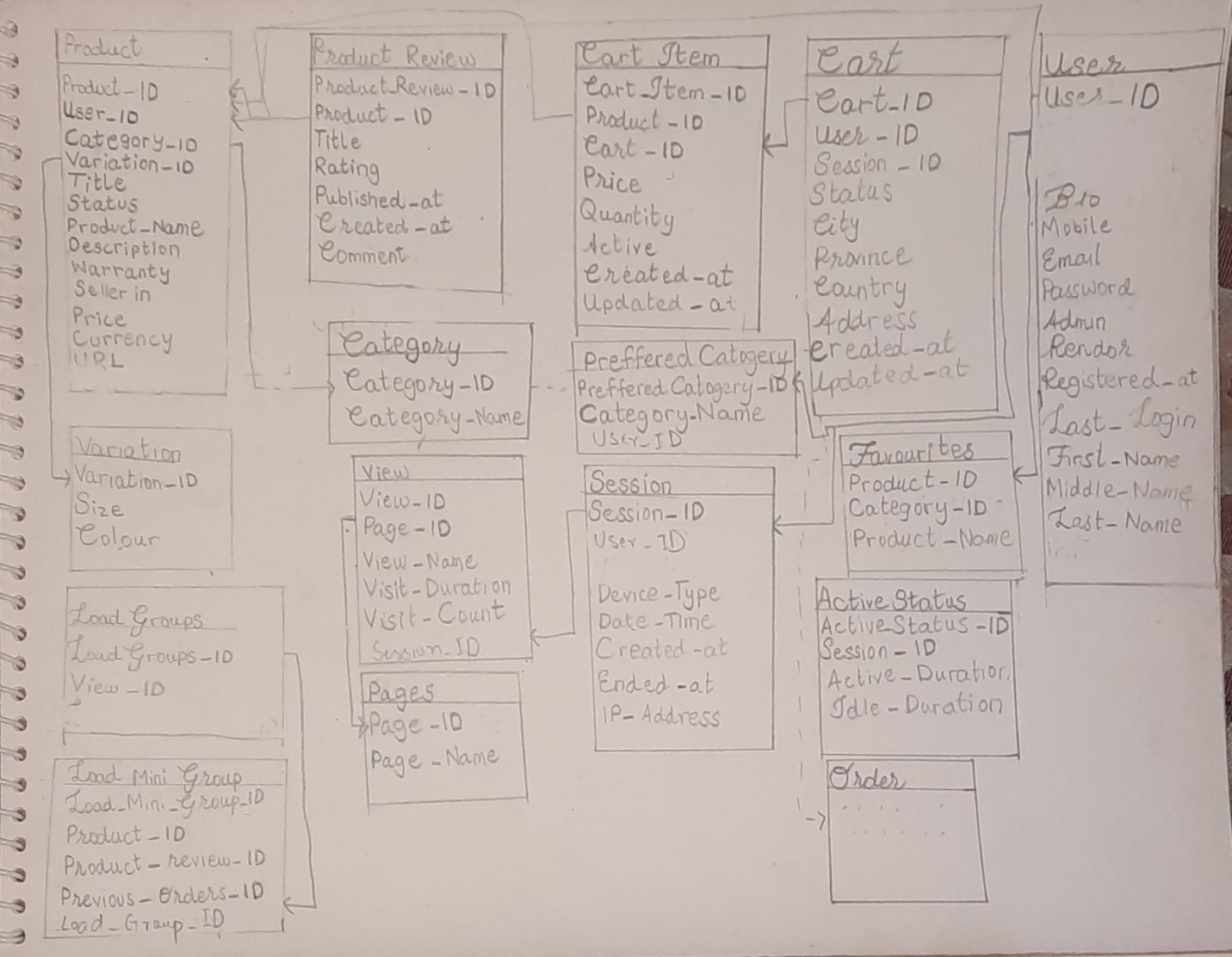
In the rest of the document you will all important topics that we covered, acted upon or thought about. Every mistake and every ambiguity will be accounted through reasoning in the later of this document.

## Additional Tables in OLTP & Reasonings

As we moved towards the creation of Data Marts and Data Modelling. We realized some issues as we molded our planned data marts in OLPT. The issue was not that we had not thought about such tables but we removed them from OLTP thinking that they might be ‘necessary’. But eventually we did add:

* Category
* Preferred Categories
* Favorites
* Order
* Pages

## Final OLTP



## Data Marts

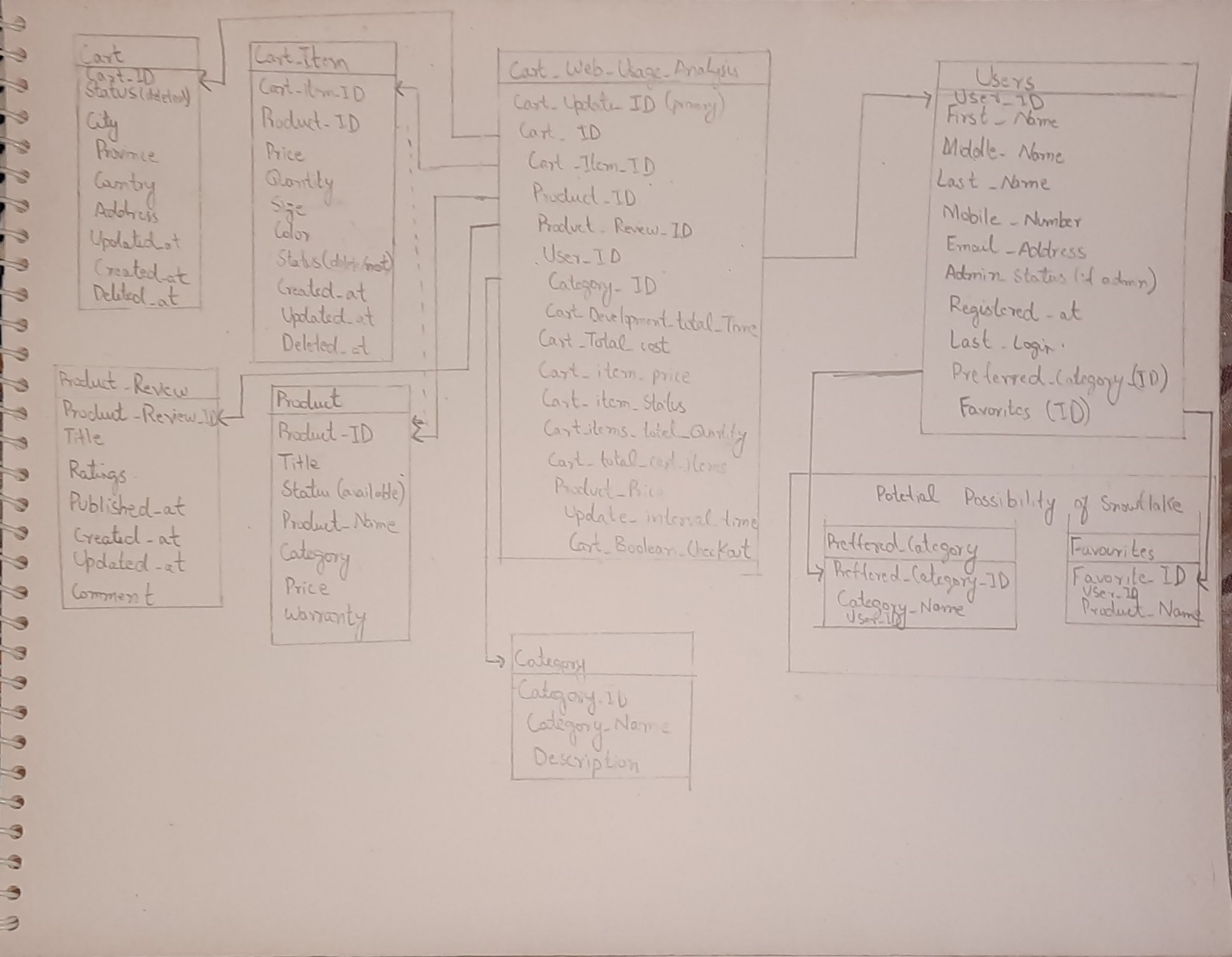
Moving on we created 2 Data Marts:

* Session Web Usage Analysis
* Cart Web Usage Analysis

You may see the excel file which is a more conventional way of creating Data Marts which we definitely followed but we couldn’t resist ourselves from making the Data Mart on a chart paper to have a visual representation of the data mart. Which in Ali’s view gives him the ability to design clearer.

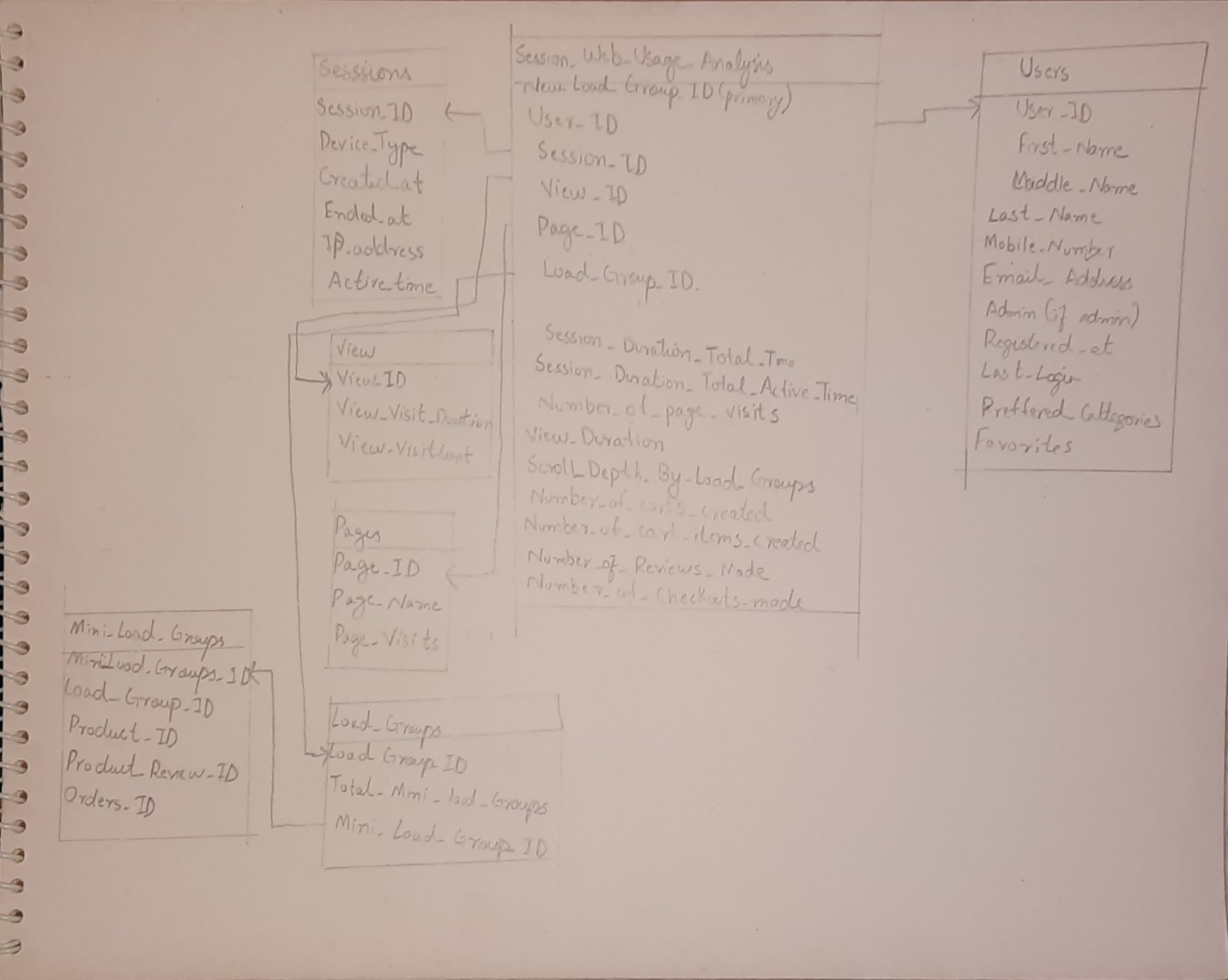
Hence you will see Data Charts Diagrams and design on Paper on the Next Page:

Cart Web Usage Analysis



If the pictures aren’t clear or zoomable, you may want to find these files in the files folder on the drive link and see the pictures clearly.

Session Web Usage Analysis



## Significance of Keys

In the Data Modelling, in the attributes and metrics tab, you may see that almost all tables have Surrogate Keys and Primary Keys. This is done because, depending on the type of the e-commerce website, these may vary considering what kind of most queries will be run. Hence to make generic data marts, we created keys in this manner to keep everything as much connected as possible.

## Hidden Processes

Some hidden processes need to be explained and some assumptions. Well, this project is, mostly based on assumptions as we have built out own OLTP. However, we think it is imminent if we explain some hidden processes that we have assumed.

As you may see, there are two tables named, Preferred Category and Favorites. We have assumed, that the user will only mark the product as favorite and the system can identify the top favorite products’ category and add them in to preferred category table.

Similarly, we had a huge debate that we haven’t added a separate table for Location in Data Mart but are using location from the cart table. It is because, the IP address can be used to identify location in Session Data Mart and Cart Table can be used to identify address in Carts Data Mart.

Parallelly, table for time also doesn’t exist. It is because, both data marts are run either on one update of cart or load\_groups. The time interval between update of a cart accounts to majority of the time issues and similar answer goes for session data mart with session total active duration.

## Possible Issues & Solutions

Possible issues are that in the initial OLTP, we didn’t create the tables of Preferred Categories and Favorites. We realized we need to because there will be multiple rows in both tables. Now in the Data Mart, these tables can either be ignored or we can take one row (most recent) and add that in users table or we can create a snowflake and use that.

In the Session Web Usage Analysis, there is a table Load\_Groups and Load\_Mini\_Groups. The Data Mart’s one row is one new row of Lood\_Group, now one row of Load\_Group creates multiple rows in Load\_Mini\_Group as a lot of data could be transferred at once. Now, this definitely should create a snowflake and we have created it. However, as so much data will be collected in the Snowflake, it is better to ignore this table which we can’t do because this table contains all the data which accessed at a given time. So what we can do is shift the dynamics of the Data Marts and make the grain as one row of the Load\_Mini\_Group is one row of Session which will solve the problem but again it depends on the company, its objectives and issues.

The active status table coming from the session table isn’t directly added as a table in the Session Data mart. Argument made by Rafay was that it will be near to impossible to note Active Session Duration if we omit this table which holds true, but Ali’s point was that this entity can be added in the Session table instead of creating a whole table for it and then note all the time intervals and then calculate the active duration of session, it is better to directly calculate and then save it in dimensional table.

## An Exempted Data Mart

We came through an interesting piece of research which claimed and explained how a user reads, views and creates ratings and comments for a product or anything can lead to how a customer behaves on-site. First we realized that we may divert from the original domain of our project but web analysis definition is to study users behavior by how he/she acts on the web and this means studying comments, studying chat with vendors, studying call complaints etc all can fall in Web Usage Analysis. Even more, some people also go to the extent to analyze how much they are paying for computing power of a website and how efficient that computing/processing power is?

We created a rough data mart for comments but that required multiple changes in OLTP, we planned that if we will have remaining time, we will make it and write its detailing in supporting document and may present it while presenting the project. We have made it now we can show it in the presentation if we have the required time. It is a very interesting data mart which will require a lot of data mining and analysis of language used in the comments by users.