KHAWARIZMI INTERNATIONAL COLLEGE

REPORT ON

DATA WAREHOUSE DESIGN

YOUR NAME

FIRST SEMESTER [OPTIONAL]

ROLL NO [ABC17]

## Database Design

**Identify the structured and less structured data sources needed in order to study the car market study of the 3 countries**

A structured data is comprised of clearly defined data types whose pattern makes them easily searchable; while unstructured data – “everything else” – is comprised of data that is usually not as easily searchable

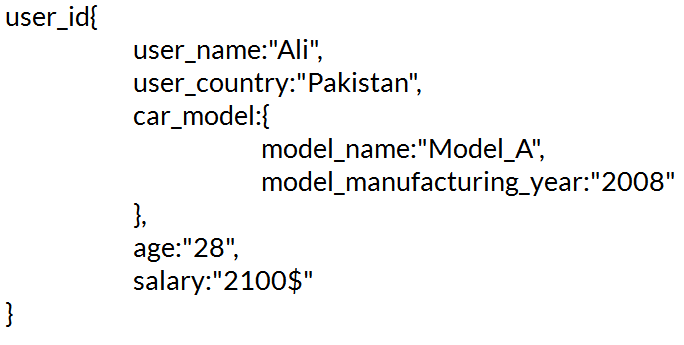
**Examples** of Structured data is Relation databases whereas unstructured data has no pre-defined format or organization, making it much more difficult to collect, process, and analyze examples include formats like audio, video, and social media postings.

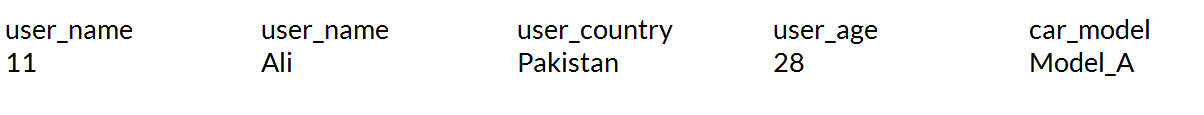
Further more **structured data** can be think of data that has specific data type that fits in the columns in relational databases and spreadsheets. The programming language used for managing structured data is called structured query language, also known as SQL

**Unstructured** **data** on the other hand is difficult to deconstruct because it has no pre-defined model, which mean it cannot be organized in relational databases. Instead, non-relational, or NoSQL databases, are best fit for managing unstructured data. Example may include mongoDB which is quite efficient way to handle data in a JSON format.

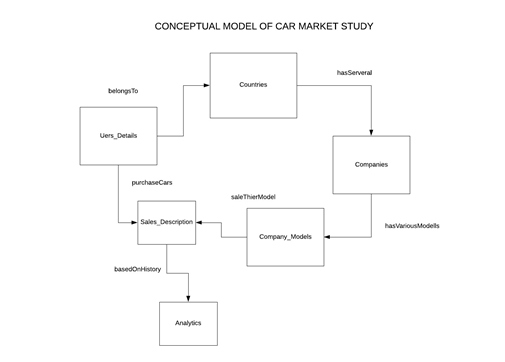
Addressing to the problem in hand i.e. “study ofcar market”data in structured format is more feasibly to handle, since it is more structured and there are powerful tools available to handle structured data, whereas unstructured sources need more processing and time since it’s tools are comparatively more complex than SQL, Let’s have a look at the “**car market study**”data in both possible formats

**Unstructured Data Format**

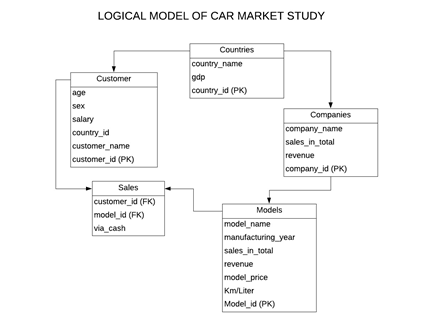


**Structured Data Format**

**Design the conceptual models for the databases**

****

**Design the logical model suitable for the DBMS**

****

**Select what DBMS you want to use for the implementation and justify your choice**

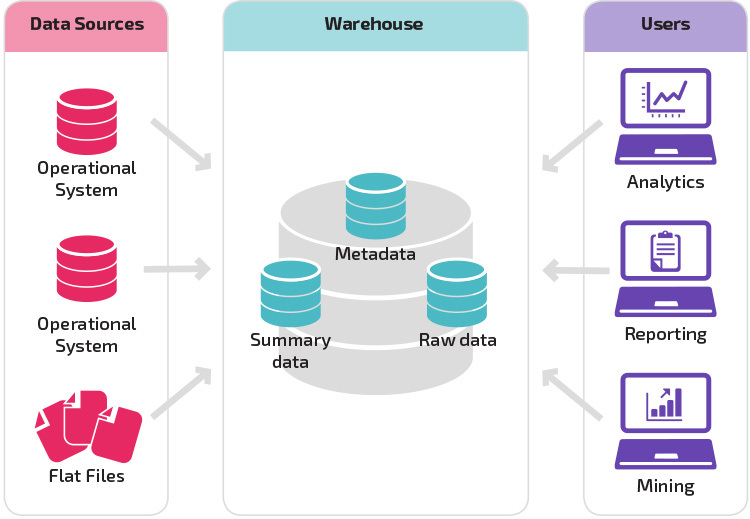
I used ORACLE and Microsoft Access DBMS since, these are the most popular databases all around the world. Most importantly these DBMS has plenty of documentation so, one can easily refer to its documentation, can post questions on developer blogs without wasting much time. Another reason for selecting RDBMS (Relational Database management system) is it requires structure query language (i.e. SQL) working on unstructured data involves much more complexity than to structure one.

## Data warehouse Design

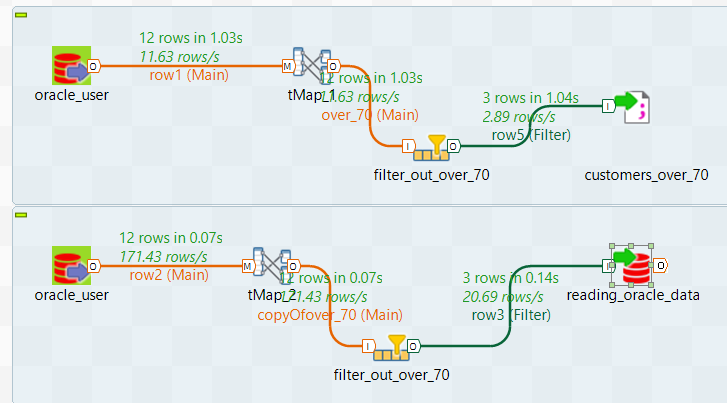
**Design the Data warehouse distributed architecture using a suitable modeling language and including the external data sources**

This is a very important step in the data warehousing project. Indeed, it is fair to say that the foundation of the data warehousing system is the data model. A good data model will allow the data warehousing system to grow easily, as well as allowing for good performance.

In data warehousing project, the logical data model is built based on, then it is translated into the physical data model. I already design logical and conceptual model as describe in **database design section.** Let’s have a look at data warehouse design.



As per study I understood data warehouse is like a central repository that keep track of data from several sources (including Database, flat files etc.). For implementing **“car study model”** data warehouse I used Talend Open Studio that server’s role for ETL (i.e. Extract Transfer and Loading files from several sources into the data warehouse) and databases namely ORACLE and MICROSOFT ACCESS, these database act as external data sources to my data warehouse.



In the above fig I loaded data into the data warehouse from ORACLE database (unfortunately due to some technical issues I am unable to load it from another source MICROSOFT ACCESS).

**Design the appropriate queries, user interfaces and forms using suitable query and modeling**

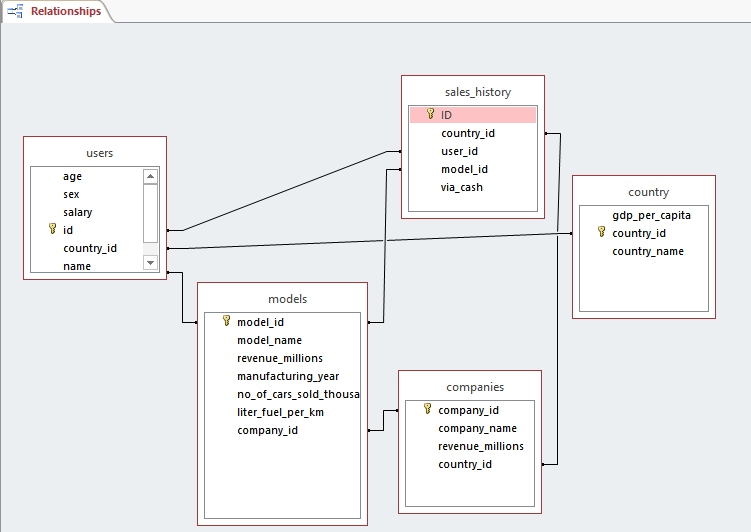
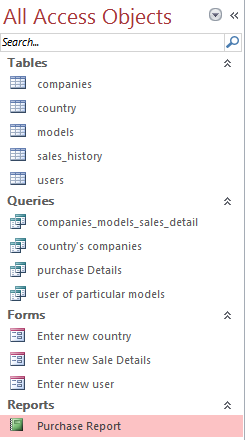
**Languages**

I used MICROSOFT ACCESS Database to implement queries, interfaces and forms.

## Data warehouse Implementation

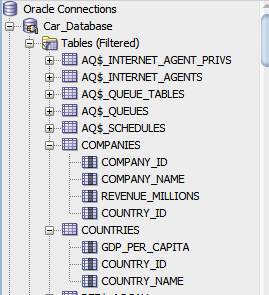
**Implement the different databases**

## Microsoft Access

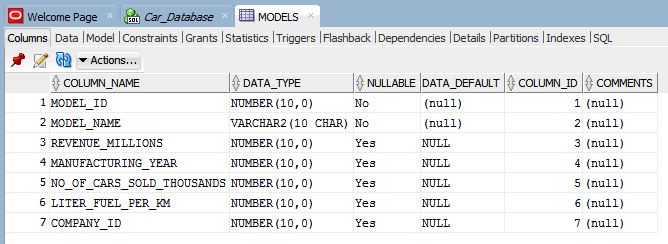
****

Above fig shows list of **Tables**, **Queries**, **Forms** and **Reports** created for the implementation of car study model along with the relationship between these tables.

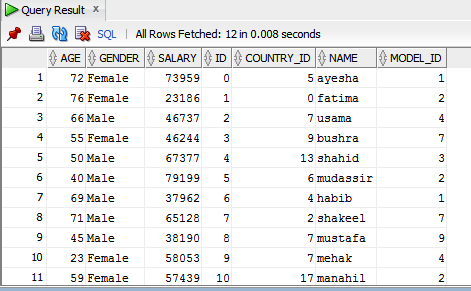
## ORACLE DATABASE



As mentioned in the conceptual design I need four tables to implement car study model so, here I created these tables in oracle as well



Above fig show the table structure (its shows car model table structure inside oracle database)

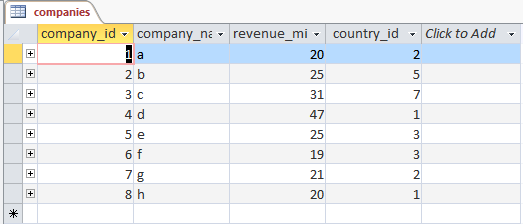


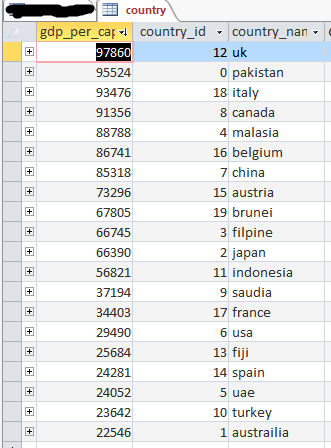
Above fig shows sample data filled inside User’s table in oracle (Query: SELECT \* FROM USERS)

**Create sample Data**

**Above fig also depicts Sample data used in oracle. For better understanding I imported these in .csv format as shown below in Microsoft Access Database**

1. **Company Table**

****

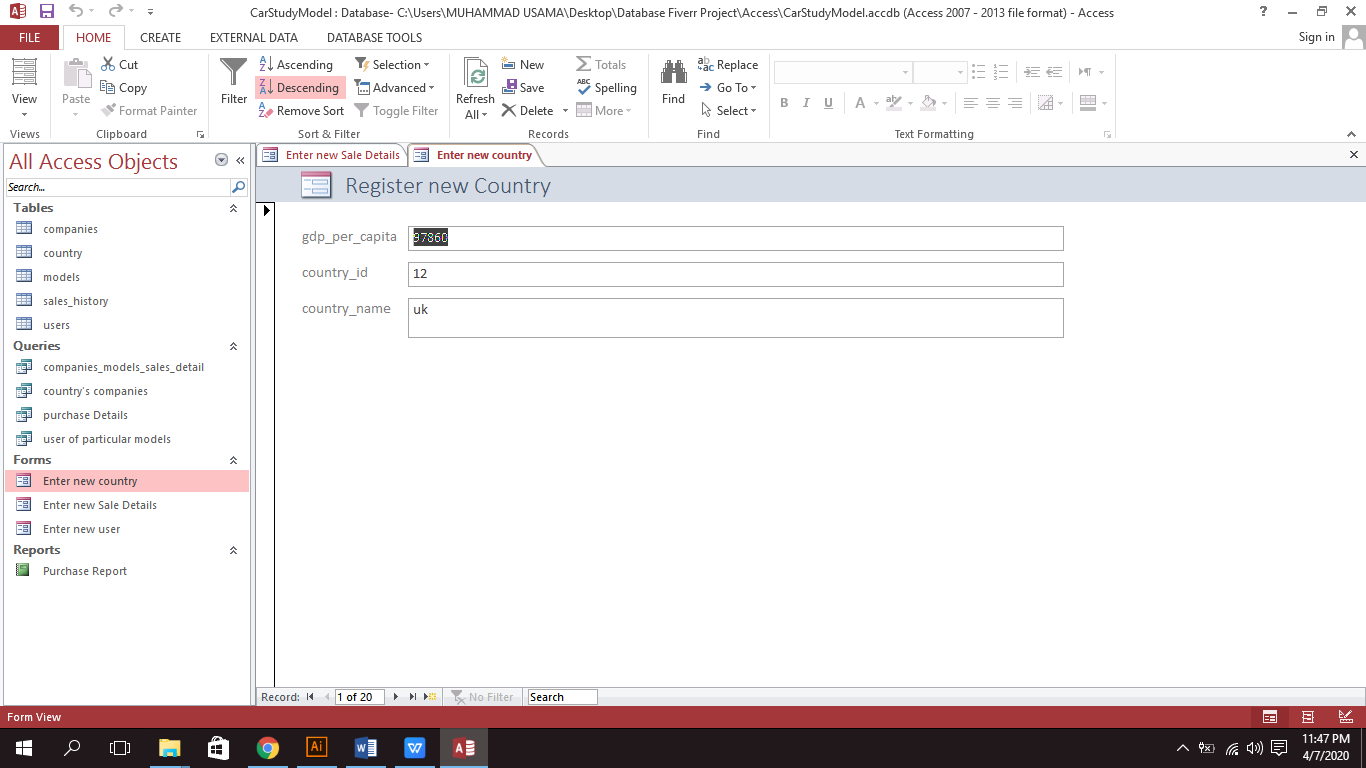
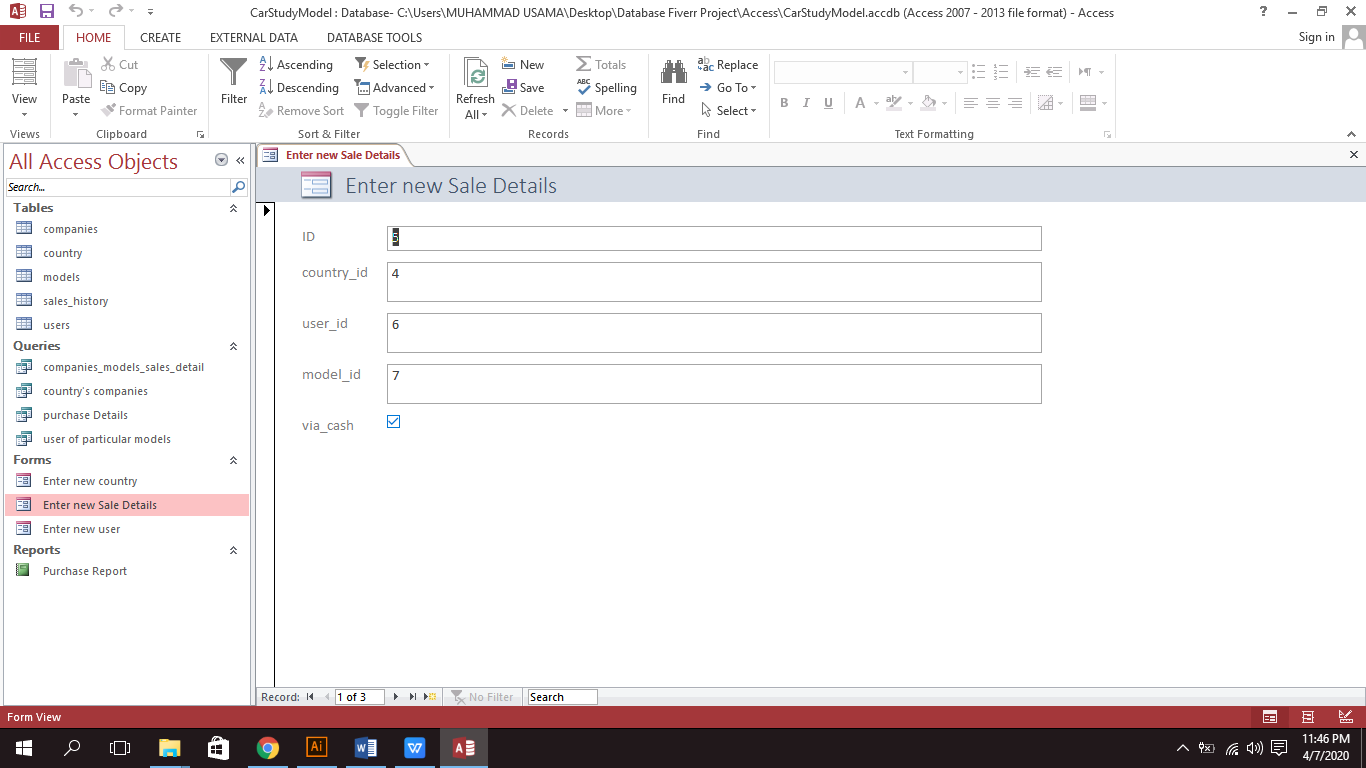
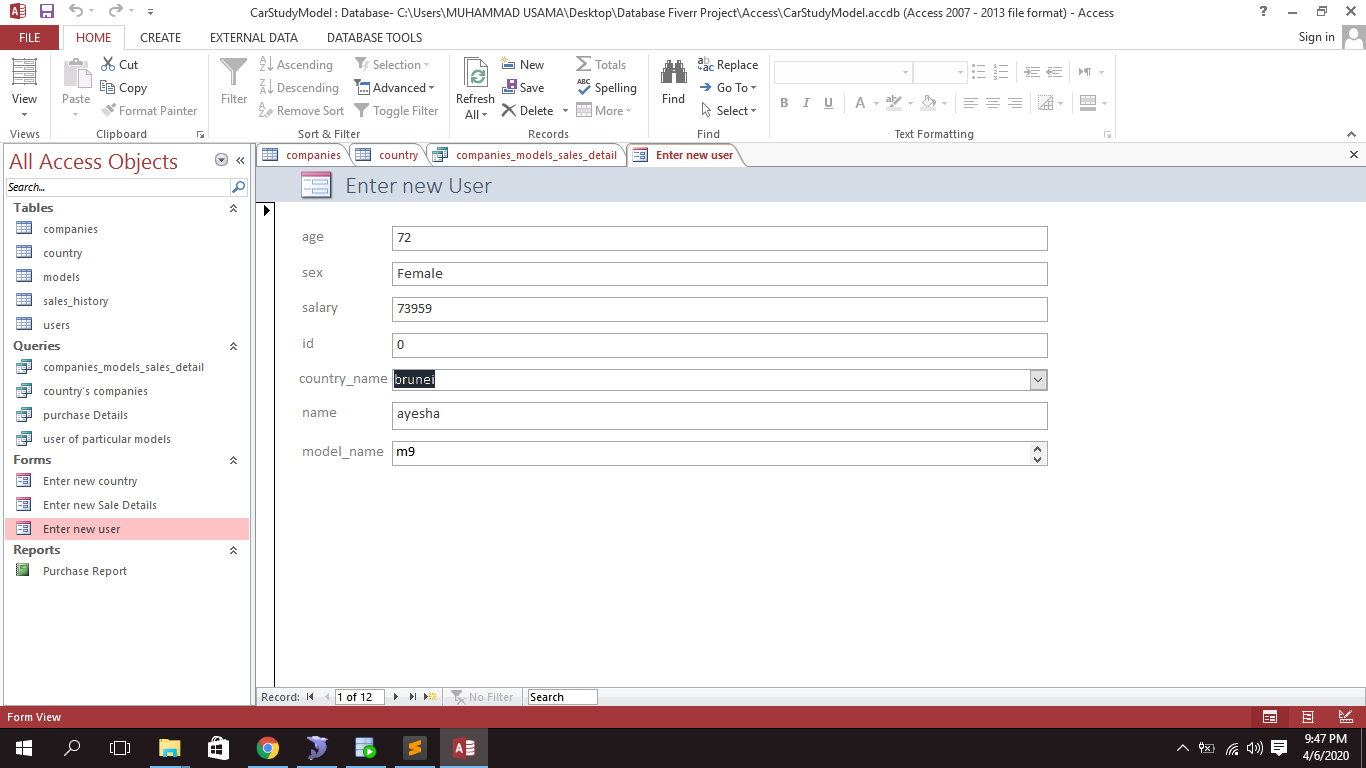
1. **Country Table**

Same is the case for User and Models tables.

**Implement the forms and interfaces**

I create three forms

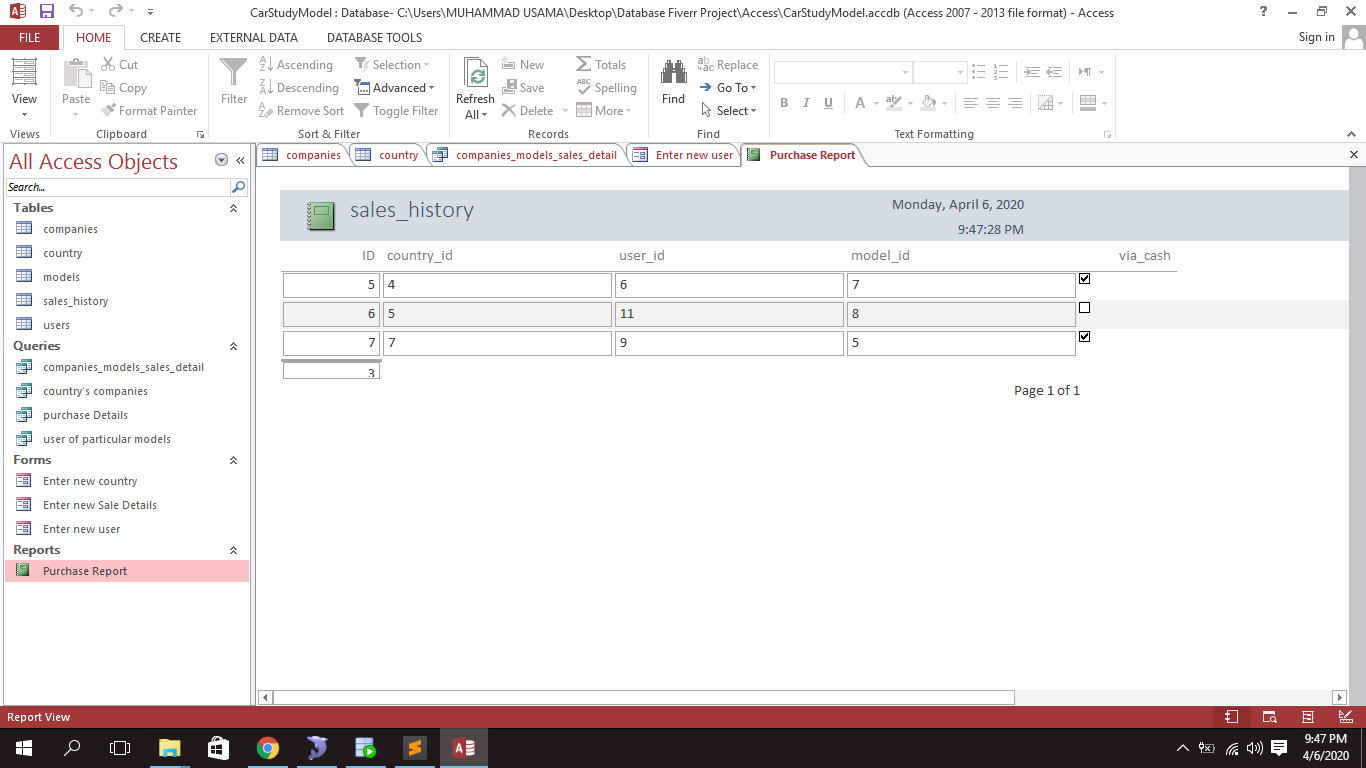
1. Entering a new user
2. Entering Sales Detail
3. Entering a new Country



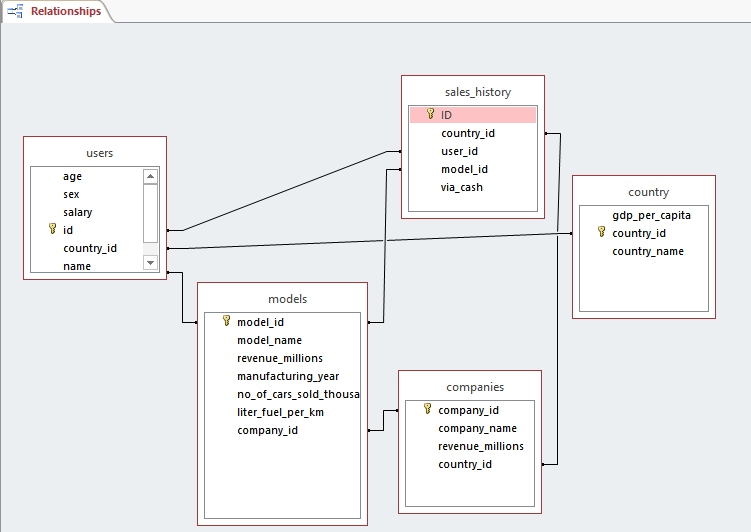
**Generate at least 3 reports**

Regardless of the strength of the OLAP engine and the integrity of the data, if the users cannot visualize the reports, the data warehouse brings zero value to them. Hence front end development is an important part of a data warehousing initiative.

1. Microsoft Access Reporting

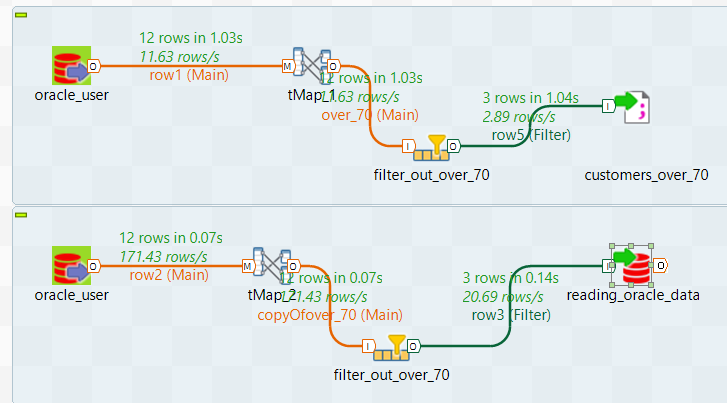


As you can see first basic report using Microsoft access, Report describe the sales history i.e. user identity, It’s country, car model he/she purchase and finally did he/she purchase via cash or loan. Relationship between the tables can be depicted from the fig below

****

1. Data warehouse Reporting (Most important)

Obviously our optimal goal is to implement data warehouse. The objective of data warehouse is to figure out most important statistics or facts from the data that we gathered from various sources, here in “car study model” I gather data from oracle database (tried to extract from Microsoft access as well but due to certain technical issue (like drivers) unable to load data into the data warehouse) and filter out those customers who bought car model and have age more than seventy (70), whether company is interested to give them special discounts or something else. How I did, I use Talend open studio for fetching data from oracle as an input, and filter out customers based on condition (i.e. age>70) and store the results back to oracle database for further analysis or OLAP purpose i.e. Online Analytical Processing. This is all to show you how data warehouse work in real scenarios



Above fig show I also store results in a flat file namely “customers\_over\_70”.

1. Oracle Report

Results of step “Data warehouse reporting (Most important)” is store in oracle table as shown below.

