Group #: 10

Group Members

Sai Pratheek Banda

Prithvi Gude

Nikhil Kumar Kasham

Sriharshith Reddy Nimmala

**SQL Project – Google Store Visitor Data**

BUAN 6320.003

Contents

[Data Model 3](#_Toc528400248)

[Assumptions/Notes About Data Entities and Relationships 3](#_Toc528400249)

[Entity-Relationship Diagram 3](#_Toc528400250)

[Physical Database 4](#_Toc528400251)

[Assumptions/Notes About Data Set 4](#_Toc528400252)

[Screen shot of Physical Database objects 4](#_Toc528400253)

[Data in the Database 4](#_Toc528400254)

[SQL Queries 5](#_Toc528400255)

[Query 1 5](#_Toc528400256)

[Question 5](#_Toc528400257)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 5](#_Toc528400258)

[Translation 5](#_Toc528400259)

[Screen Shot of SQL Query and Results 5](#_Toc528400260)

[Query 2 6](#_Toc528400261)

[Question 6](#_Toc528400262)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 6](#_Toc528400263)

[Translation 6](#_Toc528400264)

[Screen Shot of SQL Query and Results 6](#_Toc528400265)

[Query 3 7](#_Toc528400266)

[Question 7](#_Toc528400267)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 7](#_Toc528400268)

[Translation 7](#_Toc528400269)

[Screen Shot of SQL Query and Results 7](#_Toc528400270)

[Query 4 8](#_Toc528400271)

[Question 8](#_Toc528400272)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 8](#_Toc528400273)

[Translation 8](#_Toc528400274)

[Screen Shot of SQL Query and Results 8](#_Toc528400275)

[Query 5 9](#_Toc528400276)

[Question 9](#_Toc528400277)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 9](#_Toc528400278)

[Translation 9](#_Toc528400279)

[Screen Shot of SQL Query and Results 9](#_Toc528400280)

[Query 6 10](#_Toc528400281)

[Question 10](#_Toc528400282)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 10](#_Toc528400283)

[Translation 10](#_Toc528400284)

[Screen Shot of SQL Query and Results 10](#_Toc528400285)

[Query 7 11](#_Toc528400286)

[Question 11](#_Toc528400287)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 11](#_Toc528400288)

[Translation 11](#_Toc528400289)

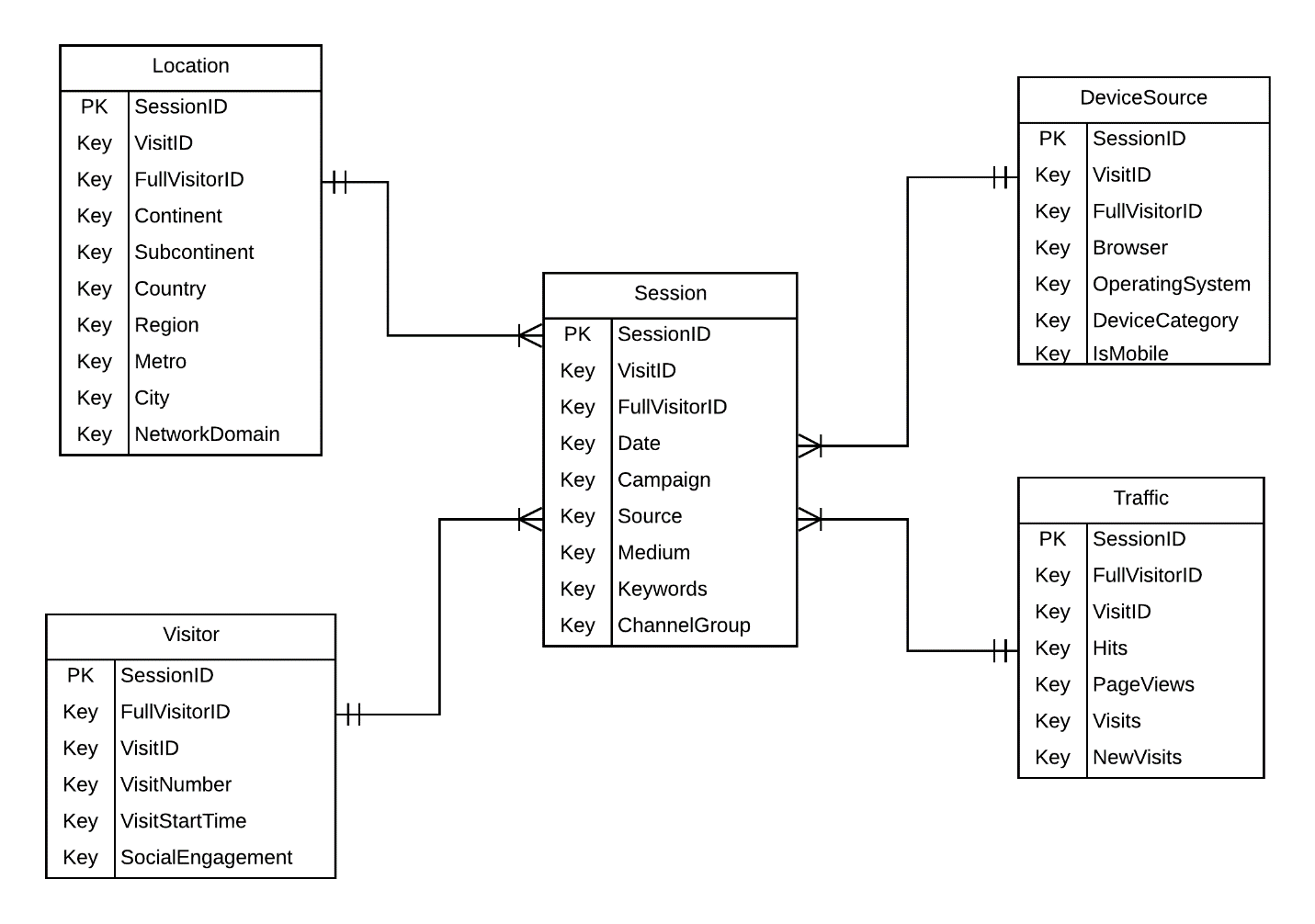
[Screen Shot of SQL Query and Results 11](#_Toc528400290)

# Data Model

## Assumptions/Notes About Data Entities and Relationships

* In the given dataset, the data is not in 1NF in the first place because, the columns “device”, “geoNetwork”, “totals” & “trafficSource” have multiple data entries, which stops us from going forward with the data model.
* As soon as those clustered columns are parsed and the data set is expanded, the data is transformed into 1 NF as every column in the data set is dependent on the sessionID
* The data set is in 2NF, because no composite primary key in our dataset. Instead all the columns are dependent solely on the primary key i.e., sessionID
* The data set is in 3NF, because every column in the dataset is dependent solely on the primary key, and there is no dependency between any other columns except for the primary key

## Entity-Relationship Diagram



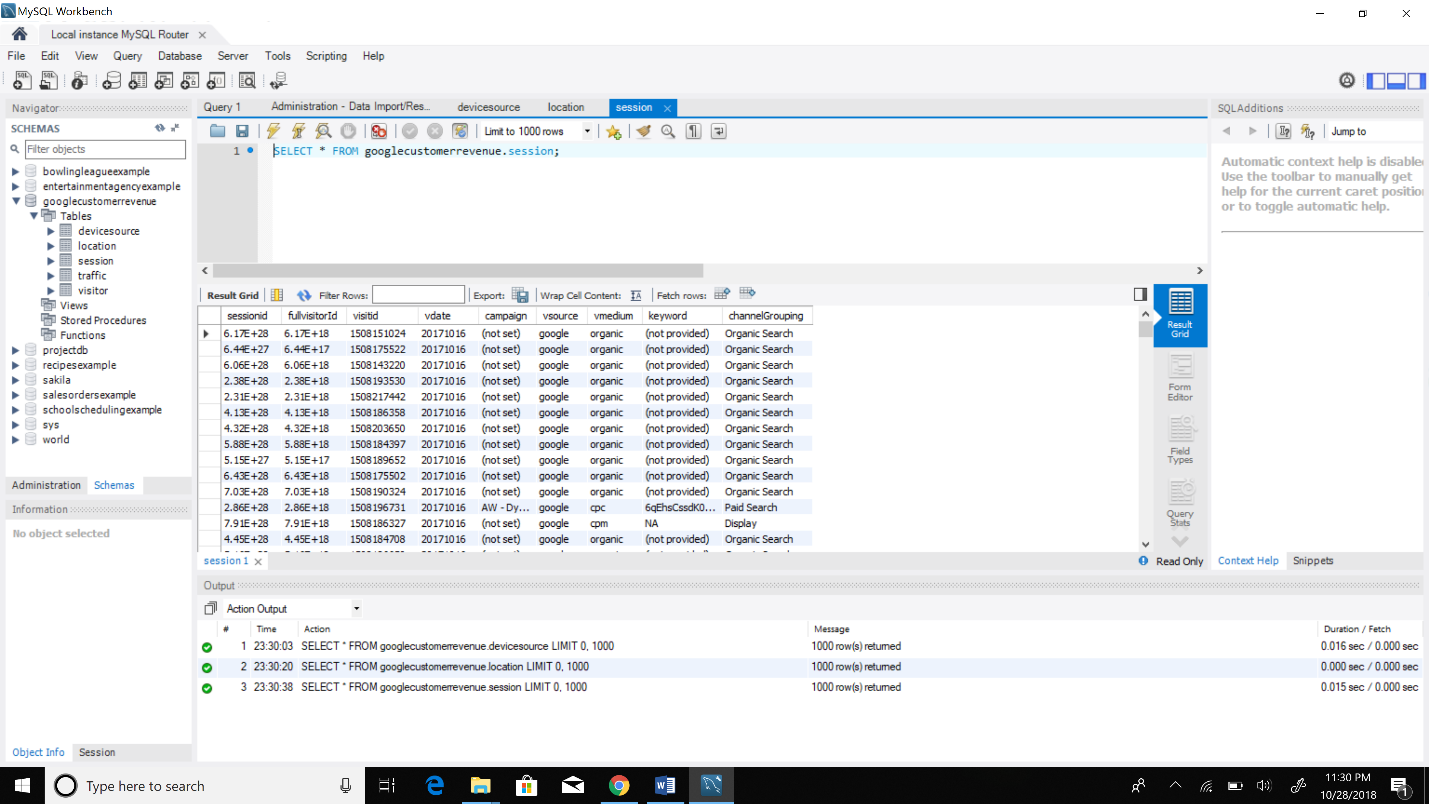
# Physical Database

## Assumptions/Notes About Data Set

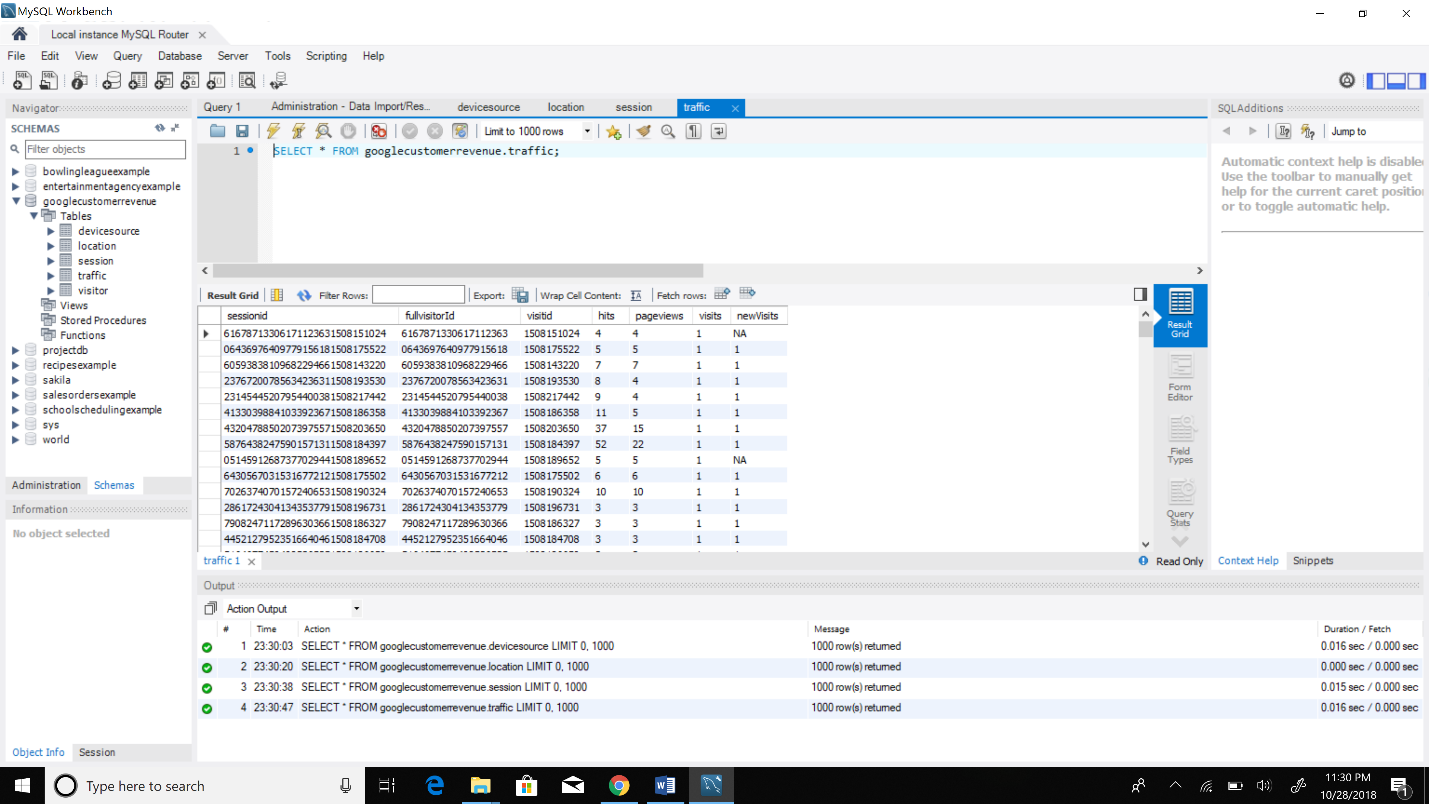
The tables created are Session, DeviceSource, Location, Traffic, Visitor.

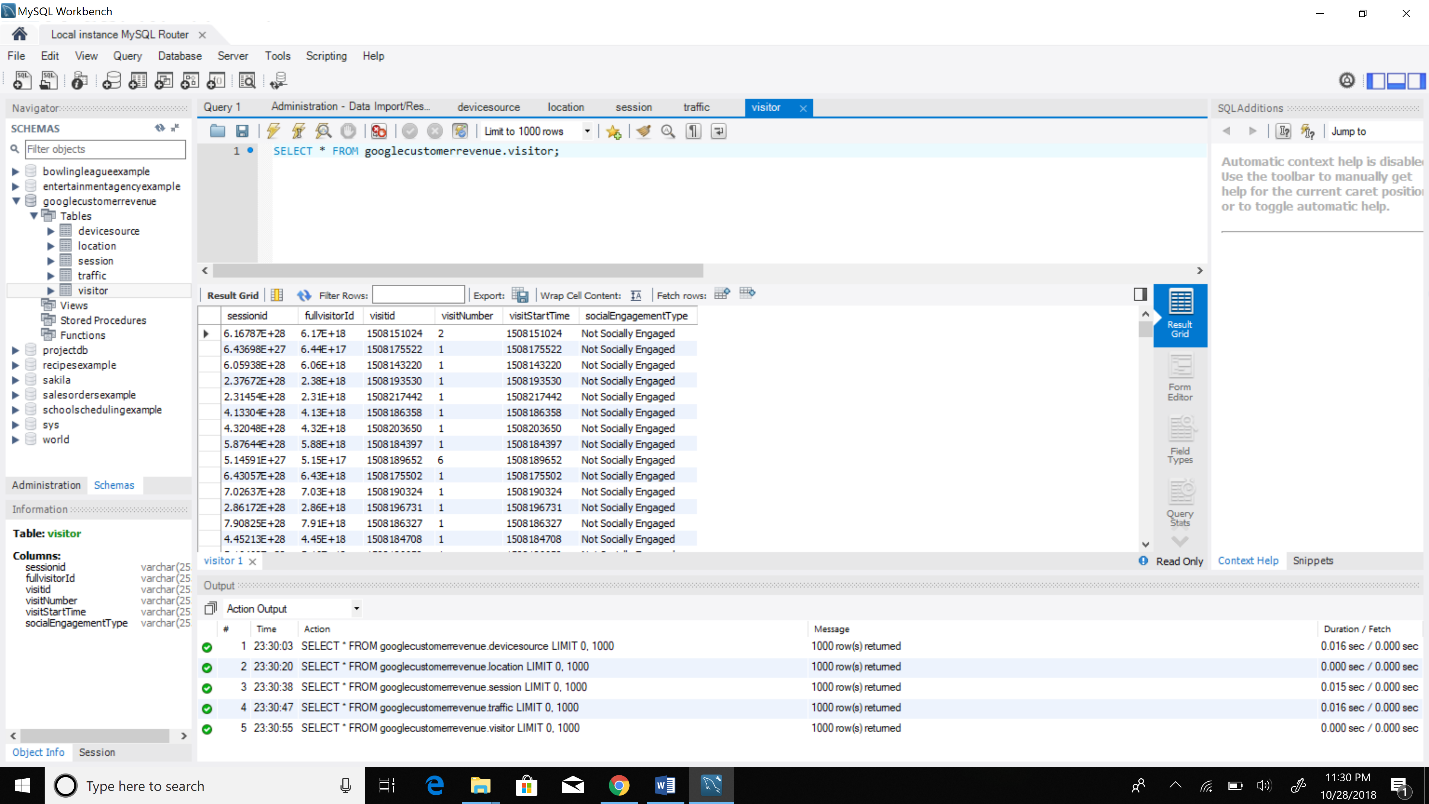
* The table “Location” contains SessionID, VisitID, FullVisitorID, Continent, Subcontinent, Country, Region, Metro, City, NetworkDomain where SessionID is the primary key. These columns have been put together in this table because all the columns give out the location details of the user and their session details.
* The table “Session” contains SessionID, VisitID, FullVisitorID, Date, Campaign, Source, Medium, Keywords and Channelgroup where SessionID is the primary key. These columns have been put together in this table because all the columns give out the session details of the user and their medium of entry, channel through which they made use of keywords and made a search, and through which campaign they got to know about the product and made the purchase
* The table “Device Source” contains SessionID, VisitID, FullVisitorID, Browser, OperatingSystem, DeviceCategory and isMobile where SessionID is the primary key. These columns have been put together in this table because all the columns give out the Device details of the user and whether they have used a mobile or desktop to make a session and the browser and OS used to run the session
* The table “Visitor” contains SessionID, VisitID, FullVisitorID, VisitNumber, VisitStarttime and SocialEngagement where SessionID is the primary key. These columns have been put together in this table because all the columns give out the respective visit details of the user and their social engagement status, also the time their session starts

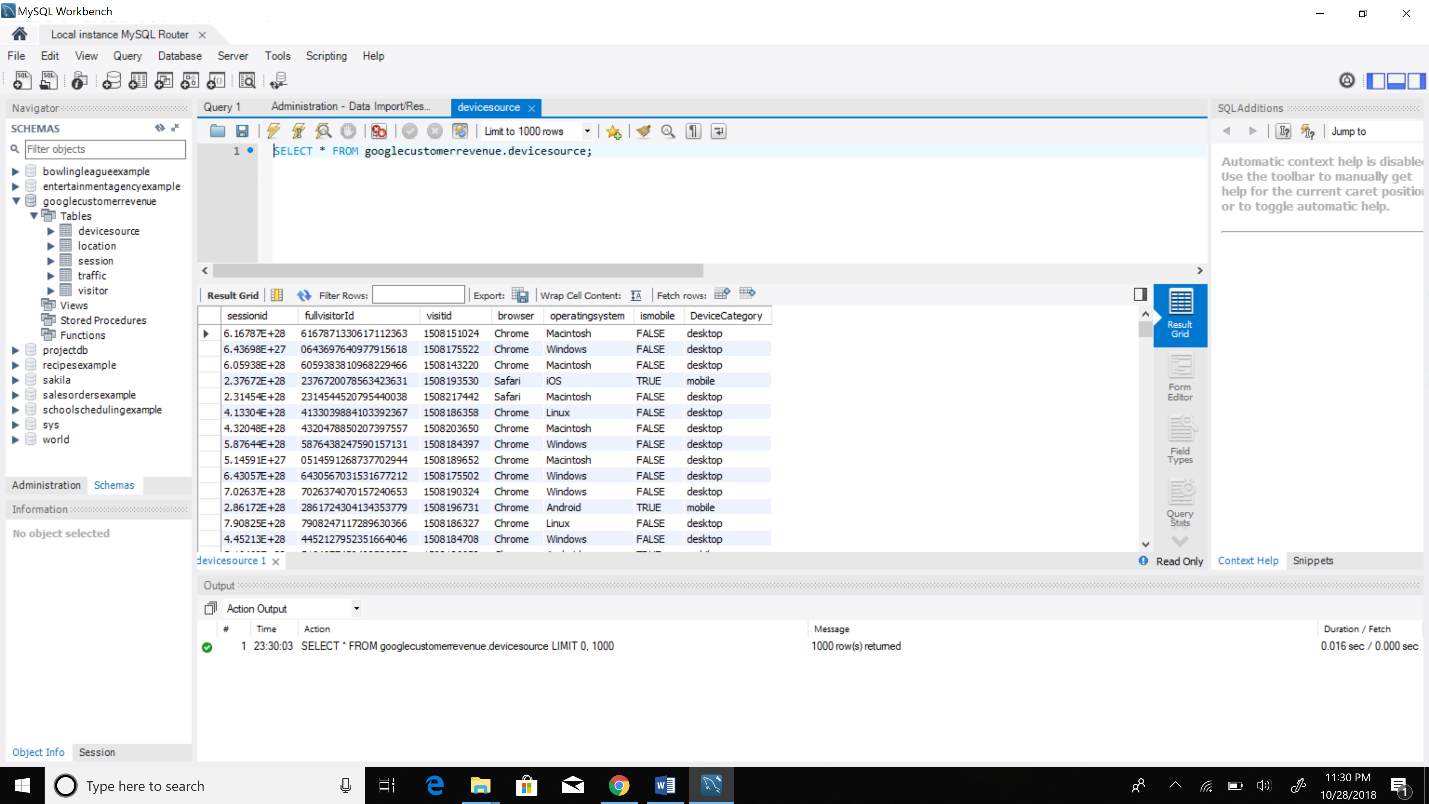
## Screen shot of Physical Database objects

Session

Traffic



Visitor

DeviceSource

Location

## Data in the Database

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Primary Key** | **Foreign Key** | **# of Rows in Table** |
| Session | SessionID |  | 804684 |
| Traffic | SessionID |  | 804684 |
| Visits | SessionID |  | 804684 |
| Location | SessionID |  | 804684 |
| DeviceSource | SessionID |  | 804684 |