Data Test Directions

Part I

Background: You have been provided a small dataset of impressions and activities for an online advertising campaign. Each row contains the User ID, Event Type and the associated Timestamp. Event Type is either an impression or an activity. Impression indicates that the user was shown an ad for the campaign. Activity indicates that the user performed an activity on the advertiser's website. The timestamp is provided in unix format.

In addition, you have been given a dataset containing user information. In this simple example the user data contains a column for the user id and a region code.

Lastly you will be asked to create a table with the following region information:

1000 Northeast

2000 Northwest

3000 Southeast

4000 Southwest

5000 Middle of Nowhere

1. Please write sql statements to create table for these datasets. Postgresql is preferred, however if you'd like to use a different RDBMS please identify the database you are using, and do not use database specific functionality that is not present in Postgres. Two requirements for the tables are that a timestamp data type should be used for readability for the impression and activities, and that duplicate User ID/Event Type/Timestamp are not allowed.

2. Show the code you use to load each of the tables you've created. The data should be cleansed of invalid data and duplicates.

3. Write both a sql query and java code that would count the number of users that were shown an advertisement a specific number of times. (i.e. event_type=impression). Your output should look something like below:

Number of ads they saw	Number of users who saw that number of ads
1	nnn
2	nnn
3	nnn
etc.	etc.

Please provide your output as well as the code/query statement.

4. Use SQL to produce a table of the number of times a user was presented an impression, with the regions as column headings and the number of impressions shown as the row title. For example:

Number of ads they saw	Northeast	Northwest	Southeast	Southwest	Middle of Nowhere
1	nn	nn	nn	nn	nn
2	nn	nn	nn	nn	nn
3	nn	nn	nn	nn	nn
etc.	etc.	etc.	etc.	etc.	etc.

5. Activities are only counted for the campaign if a user performs an activity after seeing an impression. Write sql query and java code to count the number of attributed activities. An activity should only be counted once. For example if a user sees two impressions before performing a single activity, then there is only 1 attributed activity for the user.

   Please provide your output as well as the code/query statement.

6. Attributed impression is an impression that is shown any time prior to an activity. For example, if a user saw 5 impressions prior to performing an

activity, there would be 5 attributed impressions. If a user has multiple activities after seeing an impression, then the impression should only be counted once (i.e. I,I,I,A,A where I=impression and A=action would result in 3 attributed impressions and not 6). Write sql query and java code to count the number of attributed impressions in this dataset. Please provide your output as well as the code/query statement.

7. How would you go about tuning the statements you have provided for better performance? You can assume the data sets will grow much larger.

Part II

Background: A business has asked you to design a database for them that will provide phone numbers to a widget that the UI team is developing. This widget will be used to provide listings for things like an employee phone directory, a customer contact list, and the contact numbers for various departments in the corporate web site. The widget will always provide at a minimum the name (of the organization or person), purpose of phone number (fax, cell, work, home, day, evening, etc), and the phone number itself.

The purpose of phone number may be different based on the type of listings being requested, and new purposes may be added at any time. For example, a listing may ask for 'Office', 'Fax. and 'Cell' for an employee, while a personal listing may list 'Day' and 'Evening', and 'Cell'.

In addition to the basic fields of name, purpose, and number, the widget will also return specific descriptive information about the listing depending the type of listing requested. For organizations (e.g. the corporte customer service department) it will also list the geographic region (e.g. Northeast) and the employee who is the contact for that organization. For an employee listing it will also list their business title. For a personal listing it will also list an occupation.

Provide the DDL and supporting documentation as you see fit (e.g. ERD's or Physical Data Models) for your design.