Midterm Review

1) no skew

5) ETL stands for extract transform and load

6)

```
                                                    Page Table
                                                    ID   Page
                  Time Table                        0     name
                  ID    Time                        1    search
                  0     hour                         2    login
                  1     week                         .    ...
                  2    quarter                        .   ...

                              Fact Table          ID table
                              -type_id             ID    Value
                              -size_id             0     user_id
                              -page_id             1     query
                              -id_other             2    ...
                              -time_id              .    ...
                              -bid_price            .    ...

      Size Table
      ID    Size                                  Type table
      0     728x90                                ID    Type
      .     ....                                   0    banner
      .     ...                                    1    text
      5     478x60                                 2    search
      .     ....                                   .    ...
                                                   .    ...
```
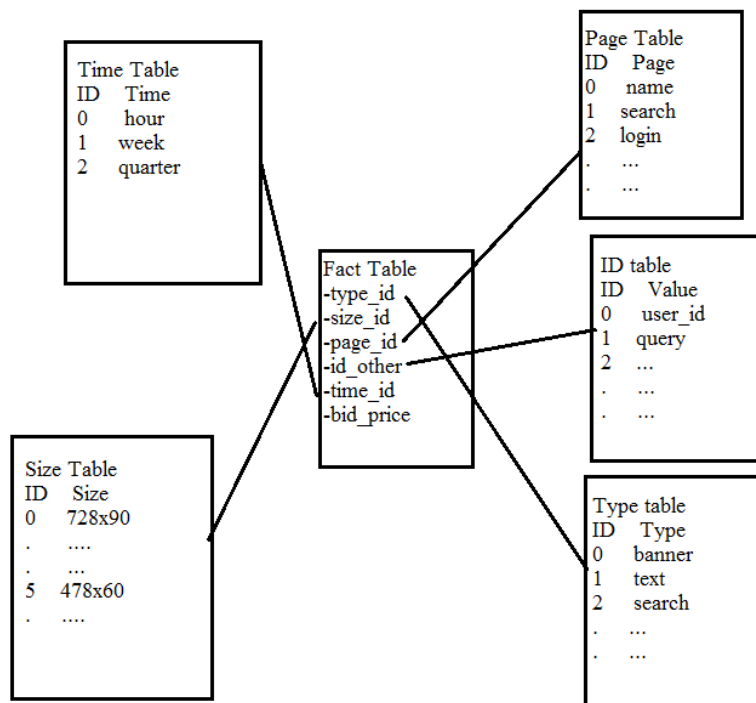
7.1) You can use rollup to summarize the data by any of the id fields and their respective values. For example if you wanted to rollup the type table you can see the resulting ads by banner, text, search, etc each separted into the different categories for type of ad.

7.4) You can use pivot to change the visualization of a data cube.

8) 1, 5, 2, 6, 7, 4, 3

10) Information gain in decision trees is an attribute selection measure. We split on some node which the highest information gain to reduce impurity in the following partitions, resulting in fewer tests to classify an object. This is important because without it you decision tree may never converge or even give incorrect results. We hope to reduce the entropy or uncertainty in our decisions.

14)