What does ETL stand for?

ETL stands for Extracting data, or retrieving information from one or more sources, Transforming it according the operational needs, and Loading into the end target, like a database.

Star Schema:

**Time Bracket**

View id

Week

Quarter

**Views**

View\_id

Data\_hour

Name\_type

Size\_type

Bid\_price

Page\_id

**Ad\_type**

Name

Size

View\_id

**Page**

View\_id

Page\_id

Page\_name

**Operations**

Slice & Dice: Considering that Slice is to select a particular value in a dimension, and Dice is to consider a subset of the values in one dimension, and example of slice & dice in the schema could be pick the views only size\_type and then look only at the views of 478x60.

**Classifier**

1. Prepare training data
2. Set aside hold-out set
3. Teach the classifier
4. Input unlabeled new data into model
5. Output guesses on new data
6. Save the model parameters
7. Verification with testing data

**Information gain**

It is the comparison of how mixed the results are before and after splitting. Another way to understand Information Gain is as the entropy resulting of “mixed” condition of a dataset.

**Fixing the gradient descent**

Changing the slope direction seems to be the solution to fix the problem.

**Terminating conditions on neural networks:**

Changes in weights too small

Accuracy in training models is high (Or errors are minimum)

Maximum number or times for learning

**Dendogram**

Considering that L2 is Euclidean distance, the order of clustering is

1. AB
2. ABD4
3. CE
4. ABDC’E