Distributed Algorithm Design Using MapReduce

1. Write a map only algorithm which will read the original dataset as input and filter out all the records which have event_epoch_time, user_id, device_id, user_agent as NULL.

   Pseudo Code

2. An algorithm to read the user agent and extract OS Version and Platform from it.

   Pseudo Code

3. Assume there is a predefined method named getCounter(String name) which takes a name as the parameter and creates a global counter variable of the same name if already not created. This global counter variable is accessible to all the map tasks. To increment the value of a counter the method to be used is incrementBy(integer num). Here "num" is the number by which we want to increment the global variable. So the syntax to increment the value of a counter is:

   getCounter("Orders").incrementBy(1)

   Using the above info write algorithms to perform below mentioned tasks:

   a) Find out the number of veg and non-veg pizzas sold

      Pseudo Code

   b) Find out the size wise distribution of pizzas sold

      Pseudo Code

c) Find out how many cheese burst pizzas were sold

Pseudo Code

d) Find out how many small cheese burst pizzas were sold. Ideally, the count should be 0 because cheese burst is available for medium and large

Pseudo Code

e) Find out the number of cheese burst pizzas whose cost is below Rs 500

Pseudo Code

4. Assume that the predefined method getCounter does not exist. Write the updated algorithms for the tasks in point-3.

a) Find out the number of veg and non-veg pizzas sold

Pseudo Code

b) Find out the size wise distribution of pizzas sold

Pseudo Code

c) Find out how many cheese burst pizzas were sold

Pseudo Code

d) Find out how many small cheese burst pizzas were sold. Ideally, the count should be 0 because cheese burst is available for medium and large

Pseudo Code

e) Find out the number of cheese burst pizzas whose cost is below Rs 500

Pseudo Code