**HW 4 - Hive**

**Due date is 28th Oct.**

Download amazon video game review data from the link https://nijianmo.github.io/amazon/index.html or https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon\_reviews.   
  
For this assignment,

* you must run different queries on Apache **HIVE**using this dataset.
* you must submit all your queries along with the screenshot of the results.
* every part is depended on the data insertion in 'part 1'.

**Point Distribution:**

**Part 1: Student 1**

* Download the dataset and insert into a table named 'GameTable'. Based on the data, design the fields of your table with appropriate data types.
* For each ‘marketplace’ & ‘product category’, find the total number of ‘review\_id’ & average ‘star\_rating’ where  ‘marketplace’ does not include ‘US’.

**Part 2: Student 2**

* Create a table 'GameTablePart' that supports partitioning on 'star\_rating' field. Fill up data for partitions 'star\_rating=1' and 'star\_rating=2' from the 'GameTable'.
* For each ‘star\_rating’ from 'GameTablePart' table, find sum of ‘helpful\_votes’ and sum of ‘total\_votes’ ordered by sum of 'total\_votes' descending.

**Part 3: Student 3**

* Create a table 'GameTableBuck' that supports bucketing on 'review\_date' field. Consider 3 buckets. For each bucket, find the minimum and maximum 'review\_date'.
* For each ‘product\_id’, find the average ‘helpful\_votes’ and average ‘total\_votes’ where average 'helpful\_votes' is greater than 1.