BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI (RAJ)
I SEMESTER 2020-2021
ASSIGNMENT-1

Course No.:	CS/SS G527	**Course Title:**	Cloud Computing
Deadline:	30th September	**Maximum Marks:**	40 (10%)

Note:
- **Group of maximum two students**
- **Deliverables: (a) Source code   (b) Design document containing rationale behind design choices**
- **Book your demo slot here (Oct 1-3):**
  https://docs.google.com/spreadsheets/d/17rR7kNN7PJoi2xNrj1qD666TclIc98BuVjFmxDoApEo/edit?usp=sharing

In this assignment, you are required to develop a REST-based service for the following dataset. This service will accept a query in the form of the template provided below and translate into MapReduce jobs and also into Spark job. Service should run these two jobs separately and return the following in a JSON object:

- Time taken for Hadoop MapReduce execution

- Time taken for Spark execution

- Input and output of map and reduce tasks in a chain as they are applied on the data

- Spark transformations and actions, in the order they are applied

- Result of the query

The standard SQL query templates is

```
GROUP BY TEMPLATE:
    SELECT <COLUMNS>, FUNC(COLUMN1)
    FROM <TABLE>
    WHERE <COLUMN1> = Y
    GROUP BY <COLUMNS>
    HAVING FUNC(COLUMN1)>X
    --Here FUNC can be COUNT, MAX, MIN, SUM
```

Description of a product dataset is given below. This dataset can be seen as a collection of tables namely products, similar, categories and reviews.

Data format:
- **Id:** Product id (number 0, ..., 548551)
- **ASIN:** Amazon Standard Identification Number
- **title:** Name/title of the product

- **group:** Product group (Book, DVD, Video or Music)
- **salesrank:** Amazon Salesrank
- **similar:** ASINs of co-purchased products (people who buy X also buy Y)
- **categories:** Location in product category hierarchy to which the product belongs (separated by |, category id in [])
- **reviews:** Product review information: time, user id, rating, total number of votes on the review, total number of helpfulness votes (how many people found the review to be helpful)

These file can be downloaded from the link given below.

--&--