BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI (RAJ)
I SEMESTER 2019-2020
ASSIGNMENT-1

Course No.: CS/SS G527 **Course Title:** Cloud Computing
Deadline: 22nd September **Maximum Marks:** 40 (10%)

Note:

- **Group of maximum four students**
- **Submit your assignment on http://nalanda.bits-pilani.ac.in as a single ZIP file**
- **Deliverables: (a) Source code (b) Design document containing rationale behind design choices**
- **Book your demo slot here (23rd to 26th Sep):** https://docs.google.com/spreadsheets/d/1FKDIb-PtYqnH-o6rSSFDyv2BvV2TqRjy5yZX2FJtBNg/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 one of the two templates 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 two standard SQL query templates are

(1) JOIN TEMPLATE:

```
SELECT * FROM <TABLE1> INNER JOIN <TABLE2> ON <CONDITION1>
 WHERE <CONDITION2>
```

(2) GROUP BY TEMPLATE:

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

Description of the dataset is given below:
Users file stores the demographic information about the users. The zipcodes file contains the city and state information for the zipcodes. Zipcode is referenced by users file. Movies file stores the information about movies. The last 19 fields in movies table are the genres, a 1 indicates the movie is of that genre, a 0 indicates it is not; movies can be in several genres. Rating file contains the ratings given by users for movies. Userid and movieid refer to the users and movies files respectively.

```
Users:
userid | age | gender | occupation | zipcode
Zipcodes:
 zipcode  | zipcodetype | city | state
Movies table:
movieid | title | releasedate | unknown | Action | Adventure | Animation |
              Children | Comedy | Crime | Documentary | Drama | Fantasy |
              Film_Noir | Horror | Musical | Mystery | Romance | Sci_Fi |
              Thriller | War | Western |
Rating:
userid | movieid | rating | timestamp
```

These files can be downloaded from the links given below.
https://drive.google.com/file/d/1LDngsYyPyilgkVdVMuenHzRv_PR5TPjF/view?usp=sharing

--&--