|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **DSCP507** | **MAP REDUCE PROGRAMMING WITH HADOOP LAB** | **L** | **T** | **P** | **C** |
| **0** | **0** | **3** | **1.5** |

# COURSE OBJECTIVES:

To learn how to setup standalone Hadoopv2 on a local machine, Hadoop YARN and Hadoop ecosystem in a distributed cluster environment and HDFS.

To gather knowledge to executeHadoopMapReducev2 computations on standalone Hadoopv2 on a local machine and distributed cluster environment.

To understand how to runHadoopMapReducev2 computations using Amazon Elastic Map Reduce cloud environment.

To perform simple analytics, accomplish mass text data processing and develop applications such as classifications, recommendations and finding relationships.

# LIST OF EXERCISES

1. Study on setting up standalone Hadoopv2 on a local machine and Hadoop YARN in a distributed cluster environment.
2. Write a Map Reduce application to count the number of occurrences of words in a dataset and run it using the Hadoop local mode.
3. Write a Map Reduce application to count the number of occurrences of words in a dataset and run it in the Hadoop distributed cluster environment.
4. Execute Word Count Map Reduce application (count the number of occurrences of words in a dataset) on Amazon Elastic Map Reduce (EMR).
5. Write a Map Reduce application to calculate simple aggregate metrics about the weblog dataset.
6. Write a Map Reduce application to group web server log data and calculate histogram and other analytics.
7. Write a Map Reduce application to calculate frequency distributions; the number of hits received by each URL.
8. Write a Map Reduce application to calculate the correlation between two datasets using scatter plots.
9. Write a Map Reduce application to parse the Tomcat e-mail list dataset that has complex data format using Hadoop by writing an input formatter.
10. Write a Map Reduce application to join two MBOX-formatted e-mail datasets.
11. Write a Map Reduce application to perform content-based recommendations for the Amazon product co-purchasing network metadata dataset.
12. Write a Map Reduce application to assign advertisements to keywords using the AdWords balance algorithm for the Amazon product co-purchasing network metadata dataset.
13. Write a Map Reduce application to clean and extract data from the 20news dataset and store the data as a tab-separated file and remove duplicate mail records using Python.

# COURSE OUTCOMES:

At the end of this course, the students will be able to

1. Install standalone Hadoop v2 on a local machine, Hadoop YARN in a distributed cluster environment and Execute Map Reduce applications on Amazon Elastic Map Reduce.
2. Formulate new solutions for programming problems or improve existing code using learned Map Reduce techniques.
3. Demonstrate an ability to listen and answer the viva questions related to programming skills needed for solving real-world problems in Computer Science and Engineering.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mapping of Course Outcomes with Programme Outcomes** | | | | | | | | | | | | |
|  | **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** |
| **CO1** | 2 | 2 | 3 | - | 2 | - | - | - | - | - | - | - |
| **CO2** | 2 | 2 | 2 | 2 | 2 | - | - | - | - | - | - | 2 |
| **CO3** | 2 | 2 | - | - | - | - | - | - | - | 2 | - | 2 |