Mukesh Patel School of Technology Management and Engineering

M2 Test – Big Data Technology

Time: 1.5 hours

1. Setting up the data
   1. Create DFS directory called /data/logs
   2. Copy all the log files into /data/logs/.



* 1. Create table with
     1. log\_date STRING
     2. log\_time STRING
     3. c\_ip STRING
     4. cs\_username STRING
     5. s\_ip STRING
     6. s\_port STRING
     7. cs\_method STRING
     8. cs\_uri\_stem STRING
     9. cs\_uri\_query STRING
     10. sc\_status STRING
     11. sc\_bytes INT
     12. cs\_bytes INT
     13. time\_taken INT
     14. cs\_user\_agent STRING
     15. cs\_referrer STRING
  2. Load the above data
  3. how many records are there?
  4. Clean the data (hints: SUBSTR(log\_date, 1, 1) <> '#‘)
  5. Create view with group the data with log\_date
  6. Order by log\_date
  7. Create partition tables with log\_year and log\_month

1. Where you should be able to run query like

SELECT log\_day, count(\*) AS page\_hits

FROM partitionedlog

WHERE log\_year=2008 AND log\_month=6

GROUP BY log\_day;