REVISION 6

Answers:

1. In 2005 Roger Mougalas.
2. Sd
3. Big Data Real Time Processing
4. **HBase** is a column-oriented non-relational database management system that runs on top of Hadoop Distributed File System (HDFS)
5. Apache **Hive** is a data warehouse software project built on top of Apache Hadoop for providing data query and analysis.
6. It is best for live streaming of data
7. MapReduce, HDFS, YARN
8. 3
9. **open source** software
10. Ruby, Perl, Python, R, Awk, etc.
11. abstraction over MapReduce
12. velocity,**distributed storage**
13. **Hadoop Distributed File System**
14. HDFS.
15. MapReduce.
16. **HDFS** has a master/slave architecture. An **HDFS cluster** consists of a single NameNode.
17. **Data Processing**
18. MasterNode
19. SlaveNode, JobTracker, Task Tracker,job.
20. JobTracker.
21. JobTracker and TaskTracker are 2 essential process involved in MapReduce execution in MRv1
22. Terminology. PayLoad
23. Job
24. Hadoop
25. File concepts.
26. **-mv <src><dest>**
27. **-cp <src> <dest>**