Date March, 8 2015:

1. Unix / Hadoop fs command to run latest file in a directory
2. Unix / Hadoop fs command to delete all directory / files except specified directory or files.
3. The industrial standard file split size of HDFS
4. What is CPU time spent (ms) in Hadoop MapReduce program execution summary report.
5. Like Hadoop Streaming in Ruby, Unix Pipe, Python and Hadoop Pipe do we have VBScript Interface with Hadoop MapReduce Programme?
6. What is Transfer rate and Seek time in disk ?
7. Transfer Time = SizeOf Block / TransferRate

Hence to read data from disk faster rate we have to keep (seek time / transfer time) ratio less as much as possible i.e.Seek time should be 1% of transfer time.

1. How to get Data Transfer rate of HDFS file system Disk and also get the Seek time of disk.
2. About hadoop fsck –files –blocks or how to check the file blocks allocation information from name node? /
3. How to navigate to namespace image and edit log of a name node in hadoop system.
4. Have u come across the scenario that if Primary namenode down and using secondary name node automatically . if yes how?
5. Check the hdfs namenode’s host and port i.e. fs.default.name property in configuration file with satish machine.
6. In which configuration file do we set dfs.replication property and its value.
7. How to add users to Supergroup in unix and how to list the users under Supergroup
8. Under stand file mode entries like drwxr-….
9. What is “Thrift”
10. Go through copyFromLocal, copyToLocal and FileCopyWithProgress commands
11. Try to using Hadoop Archives
12. What is “Checksum” and hot it helps in Data Integrity
13. assertThat() method
14. How to get the “core-site.xml” configuration file path
15. Check on configauration files of daemon’s site…
16. To list out all HDFS Directory and files comes under as core-site.xml configuration

i.e. hadoop fs -conf conf/core-site.xml –ls

24) what is java mock objects framework

25) Exploring Hadoop Output format: http://www.infoq.com/articles/HadoopOutputFormat

26) Difference between Sequencial and Random data access and what is seek ?

Sequential Access pattern is when you read your data in sequence (often from start to finish). Consider a book example. When reading a novel, you use sequential order: you start with page 1, then move to page 2 and so on. The other common pattern is called Random Access. This is when you jump from one place to another, and possibly even backwards when reading data. For a book example, consider a dictionary. You don't read it like you read a novel. Instead, you search for your word in the middle somewhere. And when you're done looking up that word, you may perhaps go look for another word that is located hundreds of pages away from where you have your book open to at the moment. That searching of where you should start reading from is called a "seek".

When you access sequentially, you only need to seek once and then read until you're done with that data. When doing random access, you need to seek every time you want to switch to a different place in your file. This can be quite a performance hit on hard drives, because seeking is really expensive on magnetic drives.

27) why HDFS block size is large i.e. 64 MB?

28) how to configure Namenode’s persistent state into local file system or remote NFS mount?

29) How to set up default filesystem for hadoop as a “HDFS” by coding.

30) Datastructures and algorithms:

<http://www.javaworld.com/article/2073390/core-java/datastructures-and-algorithms-part-1.html>

<http://www.tutorialspoint.com/java/java_data_structures.htm>

31) About Sorting algorithms

<http://www.sorting-algorithms.com/>

<http://www.java2novice.com/java-sorting-algorithms/>

32) About Lamda expressions?