

SL	Question	Answer
1	would you please publish the Q&A's (previous and future) in the website? Some of them are really helpful and we have no way of using them once the video session is over.	The video recording is including the Q&A Session.
2	Sir, Can you kindly give us the instruction to setup our environment of this course? which version should we use CentOS 7 or 8? It would be better is you provide us a tutorial for that. I use ubuntu on Vbox. I think, CentOS is totally new for most of us.	Yes please go for Cent OS 7 according to provided guideline within "Download_Package_1".
3	What does it mean by Commodity Machine? Is there anything like COMMUNITY Machine?	Commodity hardware is a non-expensive system which is not of high quality or high-availability. Hadoop can be installed on any commodity hardware. We don't need super computers or high-end hardware to work on Hadoop
5	could you please briefly explain how the google use HDFS in the GGC ( Google global Cache server ) ?	Not sure whether GGC is work on HDFS or not & it is completely related to Google and related ISP(s) wanted to configure GGC.
6	Could you please describe what is noSQL and bigData and what is the difference between hadoop and mongodb?	NoSQL can handle all types of data starting from structured, semi and non-structure data and it can work even within a small data size environmnet e.g. MonogoDB we can install for a small web application. On the other hand, HBASE is also a No SQL DB which works over HDFS meaning Big data set.
7	Is there any documentation , how to configure CentOS operating systems hosted in Virtual Box for Hadoop?	See in "Download_Package_1"
8	In general what is the ratio of Name node and Data nodes for the efficient/optimum management	If you are referring for replication factor then it is 3. Otherwise name node and data node ratio depends on your HDFS size.
9	what is heart beat in the architecture	There is heart beat implemented between Name Node and Data Node to ensure reliable status of the cluster.
10	can u explain again read operation	Explained within the class again on request.
11	sir which commands should we learn to use CentOS linux distribution?	See in "Download_Package_1"
12	Is there any solution to node redundancy for Hadoop Nodes ? for the later versions of Hadoop ?	Yes secondary namenode is there by default but it will not take traffic live in case of primary name node failure. To achive that you can configure name node cluster. Please check hadoop.apache.org site.
13	sir. are user write in log files?	Log file is generally written by deamon processes.
14	what is the deference between data block and data node which one is distributed ?	Data block is created from user files and all data blocks are distributed over all data nodes as per replication factor.
15	sir..why user need to write in logfile?	Log file is generally written by deamon processes.
16	Please tell something about the process of DFS output System	Explained within the class again on request.
17	Name node functions are like master in the nodes ?	Yes

SL	Question	Answer
18	How is the replication factor identified? Is it based on the size of the file?	It is recommended to set 3.
19	what is affected if we configure the name node and data node in the same machine if we have adequate resource?	Yes it can be done, will see in our lab. Thinking Name name clustering or redundancy, it is recommended to keep name node separately. Also data node will consume more memory.
20	How the node selection is done by HDFS, i.e; when to write in which data node?	It is round robin or based on available free space in data node. However you can always do balancer in a periodic way when you feel it is required. But better not to run it frequently as it will shuffle your data blocks.
21	if median file size > (node storage size/2), will it make the node size useless, since next file won't fit in the node? so sizing is based on type of data - correct? more like file size	File will be broken down multiple blocks based on HDFS unit block size e.g. 128 MB. So as long as user file can be kept within overall HDFS size, it is fine.
22	What is the ideal % need to be free for HDFS sizing 55%?	It depends on your host OS partitioning plan and whether you are running Map Reduce or not.
25	Is there any process to assign the replicas' of a data file to select specific data nodes?	Not sure as it is managed by HDFS framework internally.
26	How to edit files, what are the commands, you did not mention here	Follow the next class material.
27	Are we taken the lowest number to calculated??	Answered within class if the question is understandable.
28	What will be if one or two of the Slave node is/are down? Is the slave node mandatory or it will work by master node?	According to replication factor it will still try to serve data to user.
29	Sir, is it possible to demonstrate the practical classes in container environment like docker or podman?	It is not planned within our course content.
30	Map reduce technique is good	Hmm
31	sir need instructions about Cent OS in details	See in "Download_Package_1"
32	if is data size more then 1 gb , how many storage location created?	You should be able to calculate now hopefully
33	sir..plz tell me why user need to write on logfiles?	Log file is generally written by daemon processes.
34	could you please give one real world example of write-once-read-many?	Given within the Q&A session of indexing web sites in Google.
35	is there any limitations in data size for hadoop to process ?	Not known practically
36	Can you please describe some more about the relationship between map reduce and HDFS? I mean about their interaction ?	Explained within the class again on request.
37	CentOS 7 having many options to install which one should we install, i.e gnome, kde, server etc	See in "Download_Package_1" I recommend Compute Node
38	what is the importance of using rack in data nodes in case of big data?	Rack plays an important role as it hosts all servers and network bandwidth matters.
39	I have installed centOS 7, it worked fine... But then some updates came... I installed the updates, and now centOS doesn't work... What shall I do?	Uninstall update
40	Can you please repeat the least data node memory part? How did the calculation end up with 28 GB?	Explained within the class again on request.

SL	Question	Answer
41	what happens in case of communications error?	Explained within the class again on request.
42	Have any reference book to understand about Big data & Hadoop?	See my slide reference
43	As HDFS has its own data redundancy system then do we need RAID.	NO
44	Is it okay to install CentOS on Windows using VMWare? Would it cause any problems later in the course? Some suggestions please.	Yes it is ok. See in "Download_Package_1"
45	sir can you give us reference book or link	See my slide reference
46	Can run on ubuntu server?	You can run Cent OS VM over Ubuntu server.