* **1.**The number of maps is usually driven by the total size of
  + ***A. Inputs***
  + B. Outputs
  + C. Tasks
  + D. None of the listed options
* **2.**You want to count the number of occurrences for each unique word in the supplied input data. You have decided to implement this by having your mapper tokenize each word and emit a literal value 1 and then have your reducer increment a counter for each literal 1 it receives. After successful implementation it occurs to you that you could optimise this by specifying a combiner. Will you be able to use your existing reducer as your combiner and why or why not
  + ***A. Yea because the sum operation is both associative and commutative and the input and output types of reduce method match***
  + B. No, because the sum operation in reducer is incompatible with operation of a reducer
  + C. No ,because combiner and reducers use different interfaces
  + D. No, because mapper and combiner must use the same input data types.
* **3.**\_\_\_\_ is the primary interface for a user to describe a MapReduce job to the Hadoop framework for execution
  + A. Map Parameters
  + ***B. JobConf***
  + C. MemoryConf
  + D. None of the listed options
* **4.**\_\_\_\_\_ is a generalization of the facility provided by the MapReduce framework to collect data output by the Mapper or the Reducer
  + A. Partitioner
  + ***B. OutputCollector***
  + C. Reporter
  + D. All of the listed options
* **5.**Which is the default Input Formats defined in Hadoop
  + A. SequenceFileInputFormat
  + B. ByteInputFormat
  + C. KeyValueInputFormat
  + ***D. TextInputFormat***
* **6.**Is Which daemon spawns child JVMs to perform MapReduce processing
  + A. JobTracker
  + ***B. NameNode***
  + C. DataNode
  + D. TaskTracker
  + E. Secondary NameNode
* **7.**Input to the \_\_\_\_\_\_\_ is the sorted output of the mappers
  + ***A. Reducer***
  + B. Mapper
  + C. Shuffle
  + D. All of the listed options
* **8.**Mapper class must have to extend with following class
  + ***A. Mapper***
  + B. Reducer
  + C. Partitioner
  + D. Combiner
* **9.**Map method will read how many records at a time?
  + ***A. 1***
  + B. 2
  + C. 3
  + D. 4
* **10.**Mapper class contains howmany arguments
  + A.2
  + B.3
  + ***C.4***
  + D.5
* **11.**\_\_\_\_ is a utility which allows users to create and run jobs with any executables as the mapper and/or the reducer
  + A. Hadoop Strdata
  + ***B. Hadoop Streaming***
  + C. Hadoop Stream
  + D. None of the listed options
* **12.**\_\_\_\_ is the default Partitioner for partitioning key space
  + A.HashPar
  + B.Partitioner
  + ***C.HashPartitioner***
  + D.None of the listed options
* **13.**A\_\_\_\_service acts in the Slave and is responsible for executing a Task assigned to it by the JobTracker
  + A.MapReduce
  + B.Mapper
  + ***C.TaskTracker***
  + D.JobTracker
* **14.**

Point out the correct statement

* + ***A.MapReduce tries to place the data and the compute as close as possible***
  + B.Map Task in MapReduce is performed using the Mapper() function
  + C.Reduce Task in MapReduce is performed using the Map() function
  + D.All of the listed options
* **15.**Mapper input howmany records
  + A.1
  + B.64
  + ***C.1 ton***
  + D.0 to n
* **16.**The output of the reduce task is typically written to the FileSystem via \_\_\_\_\_\_\_\_\_\_\_\_\_
  + ***A.OutputCollector.collect***
  + B.OutputCollector.get
  + C.OutputCollector.receive
  + D.OutputCollector.put
* **17.**How many instances of Job tracker can run on Hadoop cluster
  + ***A.1***
  + B.2
  + C.3
  + D.4
* **18.**Whats is the default input format
  + ***A.Testinputformat***
  + B.Testinputformat
  + C.Sequence file input format
  + D.Custom input format
* **19.**In the standard word count MapReduce algorithm, why might using a combiner reduce the overall Job Running time
  + A.Because combiners perform local aggregation of word counts, thereby allowing the mappers to process input data faster
  + B.Because combiners perform local aggregation of word counts, thereby reducing the number of mappers that need to run
  + C.Because combiners perform local aggregation of word counts, and the transfer that data to reducers without writing the intermediate data to disk
  + ***D.Because combiners perform local aggregation of word counts, thereby reducing the number of key-value pairs that needs to shuffled across the network to the reducers***
* **20.**What is the disadvantage of using multiple reduces with default HashPartioner and distributing your workload across your cluster
  + A.You will not be able to compress your intermediate data
  + B.You will no longer will be able to take the advantage of a Combiner
  + ***C.The output files may not be in global sorted orDER***
  + D.There is no problem
* **21.**Users can control which keys (and hence records) go to which Reducer by implementing a custom
  + ***A.Partitioner***
  + B.OutputSplit
  + C.Reporter
  + D.All of the listed options
* **22.**Two files need to be joined over a common column. Which technique is faster and why
  + A.The reduce –side joining is faster as it receives the records sorted by keys
  + B.The reduce side joining is faster as it uses secondary sort
  + ***C.The map –side joining faster as it caches the data from one file in-memory***
  + D.The map –side joining faster as it writes the intermediate data on local file system
* **23.**You are developing a combiner that takes as input Text keys, IntWritable Values, and emits Text keys, IntWritable values. Which interface should your class implement
  + A.Combiner< Text, IntWritable, Text, IntWritable>
  + B.Reducer< Text, Text,IntWritable,IntWritable
  + C.Reducer
  + ***D.Reducer< Text, IntWritable, Text, IntWritable>***
* **24.**Which of the following phases occur simultaneously
  + ***A.Shuffle and Sort***
  + B.Reduce and Sort
  + C.Shuffle and Map
  + D.All of the listed options
* **25.**Combiners increase the efficiency of a MapReduce program because
  + A.They provide a mechanism for different mappers to communicate with each other, thereby reducing synchronization overhead
  + B.They provide an optimization and reduce the total number of computations that are needed to execute an algorithm by a factor of n; where n is the number of reducers
  + ***C.They aggregate map output locally in each individual machine and therefore reduce the amount of data that needs to shuffled across the network to the reducers***
  + D.They aggregate intermediate map output to a small number of nearby (i.e., rack local) machines and therefore reduce the amount of the amount data that needs to be shuffled across the reducers
* **26.**You write a MapReduce job to process 100 files in HDFS. Your MapReduce algorithm uses Text Input Format and the IdentityReducer. The mapper applies a regular expression over input values and emits key –value pairs with the key consisting of the matching text, and the value containing the filename and byte offset. Determine the difference between setting the number of reducer to zero.
  + A.There is no difference in output between the two settings
  + B.With Zero reducers, no reducer runs and the job throw an exception, with one reducer; instances of matching patterns are stored in a single file on HDFS
  + C.With zero reducer, all instances of matching patterns stored in multiple files on HDFS
  + ***D.With zero reducers, instances of matching pattern are stored in multiple files on HDFS.With one reducer; all instances of matching patterns are collected in one on HDFS***
* **27.**In a MapReduce job, you want each of your output files processed by a single map task. How do you configure a MapReduce job so that a single map task processes each input file regardless of how many blocks the input file occupies
  + A.Increase the parameter that controls minimum splits size in the job configuration
  + B.Write a custom MapRunner that iterates over all key-value pairs in the entire file
  + C.Set the number of mappers equal to the number of input files you want to process
  + **D.Write a custom FileInputFormat and override the method is splittable to always return false**
* **28.**In a MapReduce job, you want each of you input files processed by a single map task. How do youconfigure a MapReduce job so that a single map task processes each input file regardless of howmany blocks the input file occupies
  + A.Increase the parameter that controls minimum split size in the job configuration
  + B.Write a custom MapRunner that iterates over all key-value pairs in the entire file
  + C.Set the number of mappers equal to the number of input files you want to process
  + **D.Write a custom FileInputFormat and override the method isSplittable to always return false**
* **29.**Point out the correct statement
  + A.Applications can use the Reporter to report progress
  + B.The Hadoop MapReduce framework spawns one map task for each InputSplit generated by the InputFormat for the job
  + C.The intermediate, sorted outputs are always stored in a simple (key-len, key, value-len, value) format
  + ***D.All of the listed options***
* **30.**Which of the following is not the Dameon process that runs on a hadoopcluster
  + A.JobTracker
  + B.DataNode
  + ***C.TaskTracker***
  + D.TaskNode
* **31.**When combiner will work?
  + A.Before mapper phase
  + ***B.After mapper phase***
  + C.After reducer phase
  + D.None of the listed options
* **32.**Which of the following is key values ?
  + A.Mapper input
  + B.Mapper output
  + ***C.Reducer input***
  + D.Reducer output
* **33.**Point out the wrong statement
  + A.A MapReduce job usually splits the input data-set into independent chunks which are processed by the map tasks in a completely parallel manner
  + B.The MapReduce framework operates exclusively on pairs
  + C.Applications typically implement the Mapper and Reducer interfaces to provide the map and reduce methods
  + ***D.None of the listed options***
* **34.**What is the command you will use to run a driver named “Sales Analysis “whose complied code is available in a jar file “Sales Analytics.jar” with input data in directory “/sales/data “ and output in a directory “Sales/analytics”
  + A.Hadoop fs-jar SalesAnalytics.jar Sales Analysis- input /sales /data- output /sales/analysis
  + B.Hadoop fs jar SalesAnalytics.jar –input /sales/data- output /sales/analysis
  + C.Hadoop –jar SalesAnalytics.jar Sales Analysis –input/sales/data –output/sales/analysis
  + ***D.Hadoop jar Sales Analytics. Jar Sales Analysis / sales/data / sales/ analysis***
* **35.**In a MapReduce job, the reducer receives all values associated with the same key. Which statement is most accurate about the ordering of these values
  + A.The values are in sorted order
  + B.The values are arbitrarily ordered , and the ordering may vary from run to run of the same MapReduce job
  + C.The values are arbitrarily ordered, but multiple runs of the same MapReduce job will always have the same ordering
  + ***D.Since the values come from mapper outputs, the reducers will receive contiguous sections of sorted values***
* **36.**

What is the maximum limit for key-value pair that a mapper can emit

* + A.Its equivalent to number of lines in input files
  + B.Its equivalent to number of times mapt) method is called in mapper task
  + ***C.There is no such restriction. It depends on the use case and logic***
  + D.10000
* **37.**Which of the following is not a phase of Reducer
  + ***A.Map***
  + B.Reduce
  + C.Shuffle
  + D.Sort
* **38.**One map-reduce programme takes a text were each line break is considered one complete record and the line offset as a key. The map method parses the record into words and for each word it creates multiple key value pair where keys are the words itself and values are the characters in the word. The reducer finds the characters used for each unique word.This programme may not be a perfect programme but it works correctly. The problem this program has is That, it creates more key value pairs in the intermediate output of mappers from single input (key-value). This leads to increasese of which of the following> (Select the correct answer)
  + ***A. Disk I/O and network traffic***
  + B. Memory foot-print of mappers and network traffic
  + C. Disk-io and memory foot print of mappers
  + D. Block size and disk-io
* **39.**In one job, Howmany combiner tasks will work?
  + A.Equal to block size
  + ***B.Equal to number of mapper tasks***
  + C.Equal to number of reducer tasks
  + D.Equal to number of replications
* **40.**During the standard sort and shuffle phase of MapReduce, keys and values are passed to reducers, which of the following is true
  + ***A. Keys are presented to a reducer in sorted order; values for a given key are not sorted***
  + B. Keys are presented to a reducer in sorted order; values for a given key are sorted in ascending order
  + C. Keys are presented to a reducer in random order; values for a given key are not sorted
  + D. Keys are presented to a reducer in random order ; values for a given key are sorted in ascending order
* **41.**What is a Sequence File
  + A.A Sequence File contains a binary encoding of an arbitrary number of homogeneous writable objects
  + B.A Sequence File contains a binary encoding of an arbitrary number of heterogeneous writable objects
  + C.A Sequence File contains a binary encoding of an arbitrary number of WritableComparable objects, in sorted order
  + ***D.A Sequence File contains a binary encoding of an arbitrary number key-value pairs. Each key must be the same type. Each value must be same type***
* **42.**One large data set has fewer keysets but each key has large number of occurrences in the data. A single reducer may not be able to process the whole data set. So, you decided to create one reducer task per key ranges. What is component you will use to make each key is processed by the appropriate reducer
  + A.Combiner
  + B.OOZIE
  + C.PIG
  + ***D.Total Order Partitioner***
* **43.**Which of the following statements best describes how a large (100 GB) is stored in HDFS
  + A. The file is divided into variable size blocks, which are stored on multiple data nodes .Each block is replicated three times by default
  + B.The file is replicated three times by default. Each copy of the file is stored on a separate data nodes
  + C.The master copy of the file is stored on a single data node. The replica copies are divided into fixed –size block, which are stored on multiple data nodes
  + D.The file is divided into fixed -size blocks, which are stored on multiple data nodes .Each block is replicated three times by default .Multiple blocks from the same file might reside on the same data node
  + ***E.The file is divided into fixed –size blocks which are stored on multiple datanodes. Each block is replicated three times by default .HDFS guarantees that different blocks from the same file are never on the same datanode***
* **44.**Which of the following is wrong?
  + A.Number of mapper tasks is equal to input splits
  + B.Number of mapper tasks is equal to number of combiner tasks
  + ***C.Number of mapper tasks is equal to number of reducer tasks***
  + D.Number of input split depended on block size.
* **45.**Your client application submits a MapReduce to your Hadoop cluster. Identify the Hadoop cluster . Identify the Hadoop daemon on which the Hadoop framework will look for an available slot to schedule a MapReduce operation
  + A.TaskTracker
  + B.NameNode
  + ***C.DataNode***
  + D.JobTracker
* **46.**The number of reduces for the job is set by the user via
  + A.JobConf.setNumTasks(int)
  + ***B.JobConf.setNumReduceTasks(int)***
  + C.JobConf.setNumMapTasks(int)
  + D.All of the listed options
* **47.**Which of the following best describes the workings of TextInputFormat
  + A.Input file splits may cross line boundary .A line that crosses tile splits is ignored
  + B.The input file is split exactly at the breaks, so each Record Reader will read a series of blocks
  + C.Input file splits may cross line boundary. A line that crosses file splits is read RecordReaders of both splits containing the broken line
  + ***D.Input file splits may cross line***
  + E.Input file splits may cross line boundary. A line that crosses file splits is read by the RecordReader of split that contains the beginning of the broken line
* **48.**Which of the following is a valid flow in Hadoop
  + A.Input -> Reducer -> Mapper -> Combiner -> -> Output
  + B.Input -> Mapper -> Reducer -> Combiner -> Output
  + ***C.Input -> Mapper -> Combiner -> Reducer -> Output***
  + D.Input -> Reducer -> Combiner -> Mapper -> Output
* **49.**MapReduce was devised by
  + A.Apple
  + ***B.Google***
  + C.Microsoft
  + D.Samsung
* **50.**Which of the following should be used when possible to improve performance
  + ***A.Combiner***
  + B.Partitioner
  + C.Comparator
  + D.Reducer
  + E.All of the listed options
* **51.**What is a sequence file
  + A.A SequenceFile contains a binary encoding of an arbitrary number of homogeneous writable objects
  + B.A SequenceFile contains a binary encoding of an arbitrary number of heterogeneous writable objects
  + C.A SequenceFile contains a binary encoding of an arbitrary number of WritableComparableobjects, in sorted order
  + ***D.A SequenceFile contains a binary encoding of an arbitrary number key-value pairs. Each keymust be the same type. Each value must be same type***
* **52.**If we set setnumreducetasks(0), what will be the default output file name?
  + A.Part-r-00000
  + ***B.Part-m-00000***
  + C.Part-0-00000
  + D.Part-a-00000
* **53.**Which of the following is not an input format in Hadoop
  + A.TextInputFormat
  + ***B.ByteInputFormat***
  + C.SequenceFileInputFormat
  + D.KeyValueInputFormat
* **54.**What is data localization?
  + ***A.Executing application on the machine where data is residing***
  + B.Moving data to the machine where application is running
  + C.Distributing the data across several machine
  + D.Localizing the entire data on single machine
  + E.None of the listed options
* **55.**What is mean by batch data?
  + A.Write many read many
  + ***B.Wire once read many***
  + C.Write many read once
  + D.Write once read once
* **56.**Does Hadoop efficiently solve every kind of problem
  + A.Yes, it is like any framework and is capable of solving any problem efficiently
  + ***B.Hadoop can solve those problems very efficiently where the data is independent of each other***
  + C.Hadoop can solve only data insensitive problem efficiently
  + D.Hadoop can solve only computational insensitive problems efficiently
* **57.**How does Hadoop process large volumes of data
  + ***A.Hadoop uses a lot of machines in parallel. This optimises data processing***
  + B.Hadoop was specifically designed to process large amount of data by taking advantage of MPP hardware
  + C. Hadoop ships the code to the data instead of sending the data to the code
  + D. Hadoop uses sophisticated caching teaching on name node to speed processing of data
* **58.**Which technology is used to store data in Hadoop?
  + ***A.Flume***
  + B.HDFS
  + C.Avro
  + D.Sqoop
  + E.Zookeeper
* **59.**Hadoop \_\_\_\_\_\_\_\_\_ is a SWIG- compatible C++ API to implement MapReduce applications
  + A.Streaming
  + ***B.Pipes***
  + C.Orchestration
  + D.All of the listed options
* **60.**As compared to RDBMS, Hadoop
  + A.Has higher data Integrity.
  + B.Does ACID transactions
  + C.IS suitable for read and write many times
  + ***D.Works better on unstructured and semi-structured data.***
* **61.**What is true about Local Job Runner
  + A.It can configure as many reduces as it need
  + B.You can use “Practitioners”
  + C.It can use local file system as well as HDFS
  + ***D.It can only use local file system***
* **62.**What is the disadvantage of traditional application/systems
  + A.Data is local to the application
  + B.Data is moved to the application over low latency network
  + C.Finite network bandwidth is used
  + ***D.Reading is slow***
  + E.None of the listed options
* **63.**In Hadoop 1.x the following service not available.
  + A.Namenode
  + B.Job tracker
  + C.Task tracker
  + ***D.Node manager***
* **64.**Point out the correct statement
  + ***A.DataNode is the slave/worker node and holds the user data in the form of Data Blocks***
  + B. Each incoming file is broken into 32 MB by default
  + C. Data blocks are replicated across different nodes in the cluster to ensure a low degree of fault tolerance
  + D. None of the listed options
* **65.**Which demon is responsible for replication of data in Hadoop
  + A.HDFS
  + B.Task Tracker
  + C.Job Tracker
  + ***D.Name Node***
  + E.Data Node.
* **66.**How fault tolerance with respect to data is achieved in Hadoop
  + A.By breaking the data into smaller blocks and distributing these smaller blocks into several machines
  + B.By adding extra nodes
  + ***C.By breaking the data into smaller blocks and copying each block several times and distributing these replicas across several machines***
  + D.None of the listed options
* **67.**\_\_\_ is the slave node and holds the user data in the form of Data Blocks
  + ***A.DataNode***
  + B.NameNode
  + C.Data block
  + D.Replication
* **68.**A client wants to read a file from HDFS. How does the data get from the DataNodes to the client
  + ***A.TheNameNode reads the blocks from the DataNodes, and caches them. Then, the applicationreads the blocks from the NameNode***
  + B.The application readsthe blocks directly from the DataNodes
  + C. The blocks are sent to a single DataNode, then the application reads the blocks from that DataNode
* **69.**How does the Namenode detect that a DataNode has failed
  + A.The NameNode does not need to know that DataNode has failed
  + ***B.When the NameNode fails to receive periodic heartbeats from the DataNode, it considers the DataNode as failed***
  + C.The NameNode pings the DataNode .If the DataNode does not respond, the NameNode consider the DataNode failed
  + D.When HDFS starts up, the NameNode tries to communicate with the DataNode and consider the DataNodes failed if it does not respond
* **70.**In Hadoop 2.x release HDFS federation means
  + A.Allowing namenodes to communicate with each other
  + B. Allow a cluster to scale by adding more datanodes under one namenod
  + ***C. Allow a cluster to scale by adding more namenodes***
  + D. Adding more physical memory to both namenode and datanode
* **71.**Which of the file is used for copying file from one cluster to another cluster
  + A.Shell script
  + B. Java program
  + ***C. DistCp***
  + D. None of the listed options
  + E. All of the listed options
* **72.**The \_\_\_\_\_\_\_\_\_\_\_\_ and the EditLog are central data structures of HDFS
  + A. DsImage
  + ***B. FsImage***
  + C. DFsImages
  + D. All of the listed options
* **73.**How Hadoop achieves scaling in terms of storage
  + A.By increasing the hard disk capacity of the machine
  + B. By increasing the RAM capacity of the machine
  + C. By increasing both hard disk and RAM capacity of the machine
  + ***D. By increasing the hard disk capacity of the machine and by adding more machine***
* **74.**\_\_\_\_\_ service is used when the Primary NameNode goes down
  + A. Rack
  + B. Datanode
  + ***C. Secondaryname***
  + D. None of the listed options
* **75.**How does Hadoop does the reading faster
  + A. Hadoop uses high end machines which has lower disk latency
  + B. Hadoop minimizes the seek rate by reading the full block of data at once
  + ***C. By adding more machines to the cluster, so that it can read the data faster***
  + D. By increasing the hard disk size of the machine where data is stored
* **76.**The NameNode uses RAM for the following purpose:
  + A.To store the contents in HDFS
  + ***B. To store the filenames, list of blocks and other meta information***
  + C. To store log that keeps track of changes in HDFS
  + D. To manage distributed read and write locks on files in HDFS
* **77.**The Hadoop tool used for uniformly spreading the data across the data nodes is named
  + A.Scheduler
  + ***B. Balancer***
  + C. Spreader
  + D. Reporter
* **78.**What is the scalability limit of Hadoop
  + A. NameNode’s RAM
  + B. NameNode’s hard disk
  + C. Both Hard disk and RAM of the NameNode
  + ***D. Hadoop can scale up any limit***
* **79.**In what all parameters Hadoop scales up
  + A. Storage only
  + B. Performance only
  + ***C. Storage and performance both***
  + D. Storage , performance and 10 bandwidth
* **80.**The Hadoop framework provides a mechanism for coping with machine issues such as faulty configuration or impending hardware failure. MapReduce that one or a number of machines are performing poorly and starts more copies of a map or reduce tasks. All these tasks runs simultaneously and which one finishes first is used. This is called
  + A.Combine
  + B.IdentityMapper
  + C. IdentityReducer
  + D. Default Partitioner
  + ***E. Speculative execution***
* **81.**The archive file created in Hadoop always has the extension of
  + A. .hrc
  + ***B. .har***
  + C. .hrh
  + D. .hrar
* **82.**I have 100 files, each file size is 100 mb. Howmany input splits will be there in hadoop1.x?
  + A. 2
  + B. 100
  + ***C. 200***
  + D. Infinite
* **83.**When you put the files on HDFS, what does it do
  + ***A. Break the file into blocks, each block is replicated and replicas are distributed over the machines and Name Node updates its meta data***
  + B. File is replicate and is distributed across several machines and Name Node update its metadata
  + C. File is broken into blocks , each block is replicated and distributed across machines and Data Node’s update its meta data
  + D. File is kept as it is on the machine , along with the replica
* **84.**HDFS can be accessed over HTTP using
  + A. Viewfs URI scheme
  + ***B. Webhdfs URI scheme***
  + C. Wasb URI scheme
  + D. HDFS ftp
* **85.**Underreplication in HDFS means
  + A. No replication is happening in the data nodes.
  + B. Replication process is very slow in the data nodes.
  + C. The frequency of replication in data nodes is very low.
  + ***D. The number of replicated copies is less than as specified by the replication factor.***
* **86.**What is HDFS
  + A. HDFS is regular file system like any other file system, and you can perform any operations on HOFS
  + B. HDFS is a layered file system on top of your native file system and you can do all the operations you want
  + C. HDFS is layered file system which modifies the local system in such a way that you can perform any operations
  + ***D. HDFS is layered file system on top of your local file system which does not modify local file system and there are some restrictions with respect to the operations which you perform***
* **87.**When you put the files on HDFS, where does the HDFS stores its blocks
  + ***A. On HDFS***
  + B. On Name Node’s local file system
  + C. On Data Node’s local file system
  + D. Blocks are placed both on Name Node’s and Data Node’s local file system so that if Data Node goes down , Name Node should be able to replicate the data from its own local file system
* **88.**If a file is broken into blocks and distributed across machines then how you read back file
  + A.You will search each of the data nodes and ask the data nodes list of blocks .Then you check each of the blocks and read the appropriate bloc
  + ***B.You will ask the Name Node, and since Name Node has the meta information, it will read the data from the data node s give back the file to you***
  + C. You will ask the Name Node and since the Name Node has the meta information, it will give you the list of data nodes which are hosting the blocks & then you go each of the data nodes & read the block
  + D. You will directly read the files from HDFS
* **89.**What is true about HDFS? (Select one)
  + A. It is suitable for storing large number of small files
  + B. It is suitable storing small number of small files
  + C. It is suitable for storing large number of large files
  + ***D. It is suitable for storing small number of large files***
* **90.**Your cluster has 10 Datanodes, each with a single 1TB hard drive. You utilize all your disk capacity for HDFS, reserving none for MapReduce. You implement default replication settings. What is the data size of your Hadoop Cluster? (Assuming no compression)
  + ***A. About 3TB***
  + B. About 5TB
  + C. About 10 TB
  + D. About 11 TB
* **91.**Hadoop fs –expunge
  + A. Gives the list of datanodes
  + B. Used to delete a file
  + ***C. A.    Used to exchange a file between two datanodes.***
  + D. Empties the trash
* **92.**What is the default block size in Hadoop1.x
  + ***A.64***
  + B.128
  + C. 32
  + D. All of the listed options
* **93.**For copying data from linux to Hadoop , which command will not help
  + A. Put
  + ***B. Get***
  + C. CopyFromLocal
  + D. MoveFromLocal
* **94.**You use the Hadoop fs-put command to write a 300 MB file using an HDFS block size of 64MB. Just after this command has finished writing 200MB of this file, what would another user see when trying to access this file
  + A. They would see no content until the whole file is written and closed
  + B. They would see the content of the file through the complete block.
  + C. They would see the current state of the file, up to the last bit written by the command
  + ***D. They would see Hadoop throw a concurrent File Access Exception when they try to access this file***
* **95.**HCatalog is installed with Hive, starting with Hive release
  + A. 0.10.0
  + B. 0.9.0
  + ***C. 0.11.0***
  + D. 0.12.0
* **96.**Which techonlogy hive uses to store logs?
  + A.Logj4
  + B.Log4l
  + C.Log4i
  + ***D.Log4j***
* **97.**Point out the correct statement
  + ***A. Hive is not a relational database, but a query engine that supports the parts of SQL specific to querying data***
  + B. Hive is a relational database with SQL support
  + C. Pig is a relational database with SQL support
  + D. All of the listed options
* **98.**Variable Substitution is disabled by using ------------
  + ***A.Set hive.variable.substitute=false;***
  + B.Set hive.variable.substitutevalues=false
  + C.Set hive.variable.substitute=true;
  + D.All of the listed options
* **99.**Hive also support custom extensions written in
  + A.C#
  + ***B.Java***
  + C.C
  + D.C++
* **100.**Which of the following command sets the value of a particular configuration variable (key)
  + A.Set -v
  + ***B. Set =***
  + C. Set
  + D. Reset
* **101.**HiveServer2 introduced in Hive 0.11 has a new CLI called \_\_\_\_\_\_\_\_\_\_
  + ***A.BeeLine***
  + B.SqlLine
  + C. HiveLine
  + D. CLilLine
* **102.**Which of the following is a command line option
  + ***A. -d,–define***
  + B. -e,–define
  + C. -f,–define
  + D. None of the listed options
* **103.**Which is the additional command line option is available in Hive 0.10.0
  + ***A. –database***
  + B. –db
  + C. –dbase
  + D. All of the listed options
* **104.**Which of the following will remove the resource(s) from the distributed cache
  + A.Delete FILE[S] \*
  + B.Delete JAR[S] \*
  + C. Delete ARCHIVE[S] \*
  + ***D. All of the listed options***
* **105.**Hiveconf variables are set as normal by using the following statement
  + A.Set -v x=myvalue
  + B. Set x=myvalue
  + C. Reset x=myvalue
  + ***D. None of the listed options***
* **106.**Pig Latin statements are generally organized in one of the following ways
  + A.A LOAD statement to read data from the file system
  + B.A series of “transformation” statements to process the data
  + C.A DUMP statement to view results or a STORE statement to save the results
  + ***D.All of the listed options***