001. In Iterator, nextElement() method of Enumeration has been changed to -----   **A**
   A   next()   B   getNext()
   C   returnNext()   D   name remains same

002. TreeMap implements------------   **D**
   A   Dictionary   B   HashMap
   C   AbstractMap   D   NavigableMap

003. TreeSet internally uses which one to store elements?   **C**
   A   HashMap   B   LinkedHashMap
   C   TreeMap   D   Tree

004. HashSet internally uses   **A**
   A   HashMap   B   LinkedHashMap
   C   TreeMap   D   Tree

005. What guarantees type-safety in a collection?   **A**
   A   Generics   B   Abstract classes
   C   Interfaces   D   Collection

006. HashSet internally uses?   **C**
   A   Set   B   HashMap
   C   List   D   Collection

007. What implementation of Iterator can traverse a collection in both directions?   **B**
   A   Iterator   B   ListIterator
   C   SetIterator   D   MapIterator

008. Which does NOT implement the Collection interface?   **B**
   A   List   B   Map
   C   Set   D   Array

009. The Comparable interface contains method called------------   **C**
   A   to Compare   B   compare
   C   compareTo   D   compareWith

010. java.util.Collections is a: **A**
   A   class   B   interface
   C   package   D   object

011. What is invalid about serialization in java?   **B**
   A   Serialized object can be persisted into   B   Serialized object can &#39t be
       database.   persisted into file.
   C   Serialization is process of converting   D   Serialized object can be transferred
       object into byte stream   over network.

012. What method should we use for writing object in Serialization in java?   **C**
   A   objectWrite   B   objWrite
   C   writeObject   D   write

013. Can you customize serialization process when you have implemented Serializable   **D**
interface
   A   By defining objectWrite() and   B   It &#39s not possible
       objectRead() methods
   C   By defining write() and read()   D   By defining writeObject() and
       methods   readObject() methods

014. Which of these provides a get(in index) method?   **C**
   A   Map   B   Set
   C   List   D   Stack

015. Which of these does NOT have an index based structure?   **B**
   A   Map   B   Set
   C   List   D   Stack

016. An attempt to add null key to a TreeSet will result in: **D**
   A   Will compile   B   Compile time Exception
   C   Error   D   Runtime - NullPointerException

017. How can you sort given HashMap on basis of values ?   **A**

| A | Implement Comparator interface and override its compare method | B | It &#39s not possible |
| C | Implement Comparator interface and override its compareTo method | D | Implement Comparator and Comparable interface |

018. Which interface must be implemented for sorting on basis many criteria? **A**
| A | Comparator | B | Comparable |
| C | Serializable | D | Compare |

019. Which of the following are not a wrapper class? **B**
| A | Double | B | String |
| C | Integer | D | Boolean |

020. Which of these is a super class of wrapper classesLong & Integer? **A**
| A | Number | B | Int |
| C | Character | D | Long |

021. A -----------lets you insert and remove elements at only one end **B**
| A | queue | B | stack |
| C | list | D | set |

022. Which of these is a process of extracting/removing the state of an object from a stream? **D**
| A | Serialization | B | Externalization |
| C | File Filtering | D | Deserialization |

023. All the wrapper classes (Integer, Boolean, Float, Short, Long, DoubleandCharacter) in Java are _____ **D**
| A | immutable | B | private |
| C | serializable | D | final |

024. Which of these type parameters is used for a generic class to return and accept any type of object? **C**
| A | K | B | N |
| C | T | D | V |

025. What will happen if we try to serialize List, Set and Map as member of class in java? **B**
| A | NotSerializableException | B | They will get serialized |
| C | compilation error | D | SerializableNotPossibleException |

026. What will happen if we try to serialize primitive types in java? **A**
| A | They will get serialized | B | compilation error |
| C | SerializableNotPossibleException | D | NotSerializableException |

027. Why are generics used? **C**
| A | generics make code more fast | B | generics make code more optimized and readable |
| C | generics add stability to the code by making bugs detectable at compile time | D | generics add stability to the code by making bugs detectable at run time |

028. ---------------- involves the design and implementation of data structures and algorithms that work for multiple types **C**
| A | top down programming | B | module programming |
| C | Generic programming | D | abstract programming |

029. A -------------is a data type that keeps associations between keys and values **B**
| A | set | B | map |
| C | tree | D | heap |

030. ---------dont have duplicates. Adding a duplicate of an element that is already present is silently ignored. **A**
| A | set | B | map |
| C | tree | D | heap |

031. Which of these is method of ObjectOutput interface used to write the object to input or output stream as required? **D**
| A | write() | B | Write() |

C	StreamWrite()	D	writeObject()

032. Which of these is a method of ObjectInput interface used to deserialize an object from a stream? **C**

A	int read()	B	void close()
C	Object readObject()	D	Object WriteObject()

033. A------------ is an unordered collection of distinct elements. Elements can be added, located, and removed **A**

A	set	B	map
C	tree	D	heap

034. The Stack class in the Java library uses --------to implement a stack **B**

A	array	B	array list
C	list	D	set

035. A list iterator object has a reference to the ------------visited node. **A**

A	last	B	first
C	recently	D	element before first

036. An abstract list is an ordered sequence of items that can be traversed sequentially and that allows for ------------insertion and removal of elements at any position. **B**

A	O(n)	B	O(1)
C	O(m*n)	D	O(log n)

037. Type variables of a generic class follow the class name and are enclosed in **B**

A	{ }	B	<>
C	[ ]	D	( )

038. An -----------------defines the fundamental operations on the data but does not specify an implementation. **B**

A	Iterator	B	abstract data type
C	data structure	D	data type

039. A --------------- is a data structure used for collecting a sequence of objects that allows efficient addition and removal of elements in the middle of the sequence. **A**

A	linked list	B	stack
C	queue	D	array

040. Which of these interface extends DataOutput interface? **C**

A	Serializable	B	Externalization
C	ObjectOutput	D	ObjectInput

041. ----------- defines the position of the elements present anywhere in the linked list. **A**

A	ListIterator	B	Iterator
C	Index	D	Subscript

042. The ----------class is used to serialize an Object. **B**

A	ObjectInputStream	B	ObjectOutputStream
C	ObjectoutputStream	D	ObjectinputStream

043. The automatic conversion of primitive into object is known as ------ **B**

A	unboxing	B	autoboxing
C	wrapping	D	streaming

044. When serializing an object to a file, the standard convention in Java is to give the file a --------extension. **A**

A	.ser	B	.deser
C	.serde	D	.sar

045. In TreeMap ------------- method adds an association. **B**

A	add()	B	put()
C	get()	D	set()

046. Which of these is a super class of wrappers Long, Character & Integer? **D**

A	Long	B	Digits
C	Float	D	Number

047. Which of these is a wrapper for data type int? **A**

A	Integer	B	Long

C   Byte   D   Double

048. Which of the following statement is not correct about HashMap class?   **D**
    A   HashMap class extends AbstractMap   B   HashMap class is not synchronized
       and implements Map interface
    C   HashMap uses a hashtable to store   D   HashMap maintain order of its
       the map   element

049. Which of the following method is used to return a Set that contains the entries in a map **C**
?
    A   keyset()   B   getSet()
    C   entrySet()   D   getAll()

050. Which of these is a wrapper for simple data type char?   **B**
    A   Float   B   Character
    C   String   D   Integer

051. Which of the following methods is a method of wrapper Integer for obtaining hash code **C**
for the invoking object?
    A   int hash()   B   int hashcode()
    C   int hashCode()   D   Integer hashcode()

052. Which of the following is method of wrapper Integer for converting the value of an   **B**
object into int?
    A   bytevalue()   B   int intValue()
    C   Bytevalue()   D   Byte Bytevalue()

053. What package is a part of the wrapper class which is imported by default into all Java **A**
programs?
    A   java.lang   B   java.awt
    C   java.io   D   java.util

054. Which of these is a process of writing the state of an object to a byte stream?   **A**
    A   Serialization   B   Externalization
    C   File Filtering   D   Deserialization

055. How will you implement dynamic arrays in Java?   **D**
    A   Set   B   Map
    C   HashMap   D   List

056. Which of the following is not a goal of HDFS?   **C**
    A   Fault detection and recovery   B   Handle huge dataset
    C   Prevent deletion of data   D   Provide high network bandwidth for
                                        data movement

057. In which mode each daemon runs on a single node but there is separate java process **D**
for each daemon
    A   local   B   standalone
    C   fully distributed   D   pseudo distributed

058. HDFS refers to   **D**
    A   Hadoop Distributed Flat file System B   Hadoop Distributive File System
    C   Hadoop Data File System   D   Hadoop Distributed File System

059. Which of the following interface guarantees that no duplicates elements are stored and **B**
all elements are accessed in natural order ?
    A   java.util.List   B   java.util.Set
    C   java.util.Deque   D   java.util.Map

060. What does get(int index) method define by List interface do ?   **A**
    A   returns an object stored at the   B   stores an object at the specified index
       specified index
    C   returns a list containing elements D   returns index
       between specified index and end in
       the collection

061. Which interface provides the capability to store objects using a key- value pair ? **A**
    A   java.util.Map   B   java.util.List

C java.util.Set	D java.util.Collection

062. Which interface is used to traverse a list in both forward and backward direction? **B**
- A   Iterator interface
- B   ListIterator interface
- C   Enumeration interface
- D   DoubleList Iterator

063. Which of the following statement is true about TreeSet class ? **C**
- A   TreeSet extends AbstractList class and implements the List interface
- B   Uses a LinkedList structure to store elements
- C   TreeSet stores elements sorted ascending order
- D   Access and retrieval times are quite slow

064. The logical division of data is known as -----while a physical division of data is known as ------- **D**
- A   record, block
- B   split, record
- C   HDFS block, split
- D   Split, HDFS Block

065. The process by which the system performs the sort and transfers the map outputs to the reducer as inputs is known as the **B**
- A   reduce
- B   shuffle
- C   sort
- D   combine

066. The HDFS command to create the copy of a file from a local system is which of the following? **A**
- A   copyFromLocal
- B   copyfromlocal
- C   CopyFromLocal
- D   copyLocal

067. A GFS cluster consists of a ----------master and -------chunk-servers and is accessed by multiple clients **B**
- A   one , one
- B   single, multiple
- C   multiple, multiple
- D   multiple, single

068. Rack awareness is the way in which the ------determines on how to place blocks based on the rack definitions. **D**
- A   job tracker
- B   secondary name node
- C   data node
- D   namenode

069. When using HDFS, what occurs when a file is deleted from the command line? **C**
- A   It is permanently deleted if trash is enabled.
- B   It is placed into a trash directory common to all users for that cluster.
- C   It is permanently deleted and the file attributes are recorded in a log file.
- D   It is moved into the trash directory of the user who deleted it if trash is enabled

070. The information mapping data blocks with their corresponding files is stored in **A**
- A   namenode
- B   datanode
- C   job tracker
- D   task tracker

071. GFS refers to **B**
- A   Google Fit file System
- B   Google File System
- C   Google Flat file System
- D   Google Fixed flle System

072. The fully distributed mode of installation (without virtualization) needs a minimum of ---------- physical machines **C**
- A   1
- B   2
- C   3
- D   4

073. When a jobTracker schedules a task is first looks for **A**
- A   A node with empty slot in the same rack as datanode
- B   Any node on the same rack as the datanode
- C   Any node on the rack adjacent to rack of the datanode
- D   Just any node in the cluster

074. Which of the following is/are true about combiners? **D**
- A   Combiners can be used for mapper only job
- B   Combiners can be used for any Map Reduce operation
- C   Mappers can be used as a combiner
- D   Combiners are primarily aimed to

	class		improve Map Reduce performance	

075. Which of the following is a common hadoop maintenance issue? **B**

    A   Lack of tools   B   Lack of configuration management

    C   Lack of web interface   D   lack of memory

076. Which of following statement(s) are correct? **C**

    A   Master and slaves files are optional in Hadoop 2.x   B   Master file has list of all name nodes

    C   Core-site has hdfs and MapReduce related common properties   D   hdfs-site file is now deprecated in Hadoop 2.x

077. What is the default HDFS replication factor? **C**

    A   4   B   1

    C   3   D   2

078. The current limiting factor to the size of a hadoop cluster is **C**

    A   Excess heat generated in data center   B   Upper limit of the network bandwidth

    C   Upper limit of the RAM in namenode   D   4000 data nodes

079. When a job tracker schedules a task , it first looks for **C**

    A   any node on the rack adjacent to rack of datanode   B   any node on the same rack as datanode

    C   a node with empty slot in the same rack as datanode   D   any node in the cluster

080. Master node in GFS, periodically communicates with each chunkserver in -------- messages **B**

    A   RPC   B   HeartBeat

    C   IPC   D   text

081. By default, the GFS makes ---- replicas per chunk **C**

    A   2   B   4

    C   3   D   1

082. Which of the following can be used to control the number of part files in a map reduce program output directory? **B**

    A   Number of Mappers   B   Number of Reducers

    C   Counter   D   Partitioner

083. Which of the following is NOT a type of metadata in NameNode? **C**

    A   List of files   B   Block locations of files

    C   No. of file records   D   File access control information

084. What does Velocity in Big Data mean? **D**

    A   Speed of input data generation   B   Speed of individual machine processors

    C   Speed of ONLY storing data   D   Speed of storing and processing data

085. _____ has the worlds largest Hadoop cluster. **C**

    A   Apple   B   Datamatics

    C   Facebook   D   Microsoft

086. Which of the following are true for Hadoop Pseudo Distributed Mode? **C**

    A   It runs on multiple machines   B   Runs on multiple machines without any daemons

    C   Runs on Single Machine with all daemons   D   Runs on Single Machine without all daemons

087. Which of the following command runs the HDFS secondary namenode? **B**

    A   secondary namenode   B   secondarynamenode

    C   secondary_namenode   D   secondary-namenode

088. A file in HDFS that is smaller than a single block size **C**

    A   Cannot be stored in HDFS.   B   Occupies the full block &#39s size.

    C   Occupies only the size it needs and not the full block   D   Can span over multiple blocks.

089. The command hadoop fs -test -z URI gives the result 0 if **D**

A	if the path is a directory	B	if the path is a file
C	if the path is not empty	D	if the file is zero length

090. The hdfs command to create the copy of a file from a local system is **D**
A	CopyFromLocal	B	copyfromlocal
C	CopyLocal	D	copyFromLocal

091. In order to read any file in HDFS, instance of _____ is required. **A**
A	filesystem	B	datastream
C	outstream	D	inputstream

092. Which one of the following stores data? **B**
A	Name node	B	Data node
C	Master node	D	task tracker

093. Which hdfs command is used to check for various inconsistencies? **B**
A	fsk	B	fsck
C	fetchdt	D	sck

094. HDFS block size is larger as compared to the size of the disk blocks so that **A**
A	A single file larger than the disk size can be stored across many disks in the cluster.	B	Transfer of a large files made of multiple disk blocks is not possible.
C	The seek time is maximum	D	Only HDFS files can be stored in the disk used

095. The split size is normally the size of an _____ block, which is appropriate for most applications. **D**
A	Generic	B	Task
C	Library	D	HDFS

096. For YARN, the _____ Manager UI provides host and port information. **C**
A	Data Node	B	NameNode
C	Resource	D	Replication

097. HDFS supports the _____ command to fetch Delegation Token and store it in a file on the local system. **B**
A	fetdt	B	fetchdt
C	fsk	D	rec

098. In _____ mode, the NameNode will interactively prompt you at the command line about possible courses of action you can take to recover your data. **C**
A	full	B	partial
C	recovery	D	commit

099. HDFS is implemented in _____ programming language. **B**
A	C++	B	Java
C	Scala	D	SQL

100. When the namenode finds that some blocks are over replicated, it **C**
A	Stops the replication job in the entire hdfs file system.	B	It slows down the replication process for those blocks
C	It deletes the extra blocks.	D	It leaves the extra blocks as it is.

101. The following is not permitted on HDFS files **D**
A	Deleting	B	Renaming
C	Moving	D	Executing

102. The client reading the data from HDFS filesystem in Hadoop does which of the following? **A**
A	Gets only the block locations from the namenode	B	Gets the data from the namenode
C	Gets both the data and block location from the namenode	D	Gets the block location from the datanode

103. Which configuration file is used to control the HDFS replication factor? (S) **C**
A	mapred-site.xml	B	core-site.xml
C	hdfs-site.xml	D	yarn-site.xml

104. A _____ serves as the master and there is only one NameNode per cluster.   **B**
 A   Data Node   B   NameNode
 C   Data block   D   Replication

105. A _____ node acts as the Slave and is responsible for executing a Task assigned   **C**
 to it by the JobTracker.
 A   MapReduce   B   Mapper
 C   TaskTracker   D   JobTracker

106. HDFS provides a command line interface called _____ used to interact with   **B**
 HDFS.
 A   HDFS Shell   B   FS Shell
 C   DFS Shell   D   vi shell

107. Which of the following is used for the MapReduce job Tracker node ?   **C**
 A   mradmin   B   tasktracker
 C   jobtracker   D   job-tracker

108. _____ is the slave/worker node and holds the user data in the form of Data Blocks. **A**
 A   DataNode   B   NameNode
 C   Data block   D   Replication

109. Which command is used to show all the Hadoop daemons that are running on the   **B**
 machine
 A   distcp   B   jps
 C   dir   D   fsck

110. All the files in a directory in HDFS can be merged together using which of the   **A**
 following?
 A   Put merge   B   Get merge
 C   Remerge   D   Merge all

111. What does commodity Hardware in Hadoop world mean?   **D**
 A   Very cheap hardware   B   Industry standard hardware
 C   Discarded hardware   D   Low specifications Industry grade
                                             hardware

112. Which of the following is not a valid Hadoop config file?   **C**
 A   core-default.xml   B   hdfs-default.xml
 C   hadoop-default.xml   D   mapred-default.xml

113. Point out the wrong statement :   **D**
 A   Replication Factor can be configured B   Block Report from each DataNode
     at a cluster level (Default is set to 3)   contains a list of all the blocks that
     and also at a file level   are stored on that DataNode
 C   User data is stored on the local file D   DataNode is aware of the files to
     system of DataNodes   which the blocks stored on it belong

114. Which of the following platforms does Apache Hadoop run on ?   **C**
 A   Dos like   B   Unix like
 C   cross platform   D   debian

115. In which mode all daemons execute in separate nodes   **C**
 A   local mode   B   pseudo distributed mode
 C   fully distributed mode   D   stand alone mode

116. All of the following accurately describe Hadoop, EXCEPT   **B**
 A   Open source   B   Real time
 C   Java-based   D   Distributed computing approach

117. Apache Hadoop achieves reliability by replicating the data across multiple hosts, and   **B**
 hence does not require _____ storage on hosts.
 A   Standard RAID levels   B   RAID
 C   ZFS   D   Operating system

118. When You are developing a combiner that takes as input Text keys, IntWritable values, **B**
 and emits Text keys, IntWritable values. Which interface should your class implement?
 A   Combiner   B   Reducer

	C	Combinator		D	Partitioner	

119. The total number of partitioners is equal to -------------------   **A**

A	Number of reducers		B	number of nodes
C	number of combiners		D	number of mappers

120. Hadoop is a framework that allows the distributed processing of: **C**

A	Small Data Sets		B	Semi-Large Data Sets
C	Large Data Sets		D	Large and Small Data sets

121. Which of the following is a valid flow in Hadoop ?   **C**

A	Input -> Reducer -> Mapper -> Combiner -> -> Output	B	Input -> Mapper -> Reducer -> Combiner -> Output
C	Input -> Mapper -> Combiner -> Reducer -> Output	D	Input -> Reducer -> Combiner -> Mapper -> Output

122. Which of the following method used to set the output directory?   **C**

A	FileOutputFormat.setOutputgetpath()	B	OutputFormat.setOutputpath()
C	FileOutputFormat.setOutputpath()	D	OutputFormat.setOutputgetpath()

123. Which of the following is/are true about combiners?   **D**

A	Combiners can be used for mapper only job	B	Combiners can be used for any Map Reduce operation
C	Mappers can be used as a combiner class	D	Combiners are primarily aimed to improve Map Reduce performance

124. Hadoop Framework is written in   **A**

A	Java		B	Python
C	C++		D	Scala

125. Which of the following Hadoop config files is used to define the heap size?   **C**

A	hdfs-site.xml		B	core-site.xml
C	hadoop-env.sh		D	mapred-site.xml

126. Which of the following property gets configured on mapred-site.xml ?   **A**

A	replication factor	B	java environment variables
C	directory names to store hdfs files	D	host and port where MapReduce job runs

127. What is the default HDFS replication factor?   **C**

A	4		B	1
C	3		D	2

128. _____ can best be described as a programming model used to develop Hadoop-based applications that can process massive amounts of data.   **A**

A	MapReduce		B	Mahout
C	Oozie		D	Unix

129. _____ maps input key/value pairs to a set of intermediate key/value pairs.   **D**

A	shuffle		B	Reducer
C	Both Mapper and Reducer		D	Mapper

130. What was Hadoop named after?   **C**

A	Creator Doug Cuttings favorite circus act	B	Cuttings high school rock band
C	The toy elephant of Cuttings son	D	A sound Cuttings laptop made during Hadoop development

131. MapReduce was devised by ...   **B**

A	Apple		B	Google
C	Microsoft		D	Samsung

132. Number of mappers is decided by the   **D**

A	Mappers specified by the programmer	B	Available Mapper slots
C	Available heap memory	D	Input Splits

133. What is the default HDFS block size?   **D**

A	32 MB		B	64 KB
C	128 KB		D	64 MB

134. What decides number of Mappers for a MapReduce job? **C**
A   File Location   B   mapred.map.tasks parameter
C   Input file size   D   Input Splits

135. Above the file systems comes the _____ engine, which consists of one Job **A**
Tracker, to which client applications submit MapReduce jobs.
A   MapReduce   B   Google
C   Functional programming   D   Facebook

136. _____ part of the MapReduce is responsible for processing one or more **A**
chunks of data and producing the output results.
A   Maptask   B   Mapper
C   Task execution   D   Shuffle

137. _____ refers to incremental costs with no major impact on solution design, **C**
performance and complexity.
A   Scale-out   B   Scale-down
C   Scale-up   D   Scale-in

138. _____ is a generalization of the facility provided by the MapReduce framework to **B**
collect data output by the Mapper or the Reducer
A   Partitioner   B   OutputCollector
C   Reporter   D   Collector

139. Output of the mapper is first written on the local disk for sorting and _____ **A**
process.
A   shuffling   B   secondary sorting
C   forking   D   reducing

140. Mapper and Reducer implementations can use the _____ to report progress or just **C**
indicate that they are alive.
A   Partitioner   B   OutputCollector
C   Reporter   D   Combiner

141. _____ is the primary interface for a user to describe a MapReduce job to the **B**
Hadoop framework for execution.
A   Map Parameters   B   JobConf
C   MemoryConf   D   Conf

142. Which of the following phases occur simultaneously? **D**
A   Map and Reduce   B   Reduce and Sort
C   Shuffle and Map   D   Shuffle and Sort

143. Mapper implementations are passed the JobConf for the job via the _____ method **B**
A   JobConfigure.configure   B   JobConfigurable.configure
C   JobConfigurable.configureable   D   Job.conf

144. _____ is general-purpose computing model and runtime system for distributed **A**
data analytics.
A   Mapreduce   B   Drill
C   Oozie   D   SQL

145. _____ controls the partitioning of the keys of the intermediate map-outputs. **B**
A   Collector   B   Partitioner
C   InputFormat   D   Combiner

146. The JobTracker pushes work out to available _____ nodes in the cluster, striving to **B**
keep the work as close to the data as possible.
A   DataNodes   B   TaskTracker
C   ActionNodes   D   reduce-tracker

147. The daemons associated with the MapReduce phase are _____ and task-trackers. **A**
A   job-tracker   B   map-tracker
C   reduce-tracker   D   reduce-tracker

148. A Map reduce job can be written in: **D**
A   Java   B   Ruby
C   Python   D   Any Language which can read from

input stream

149. _____ is a programming model designed for processing large volumes of data in parallel by dividing the work into a set of independent tasks. **B**

A   Hive   B   MapReduce
C   Pig   D   Lucene

150. The output of the reduce task is typically written to the File System via --------- **A**

A   OutputCollector   B   InputCollector
C   OutputCollect   D   InputCollect

151. _____ function is responsible for consolidating the results produced by each of the Map() functions/tasks. **A**

A   Reduce   B   Map
C   Reducer   D   sum

152. Which of the following parameter is to collect keys and combined values? **D**

A   key   B   values
C   reporter   D   output

153. Input to the _____ is the sorted output of the mappers. **A**

A   Reducer   B   Mapper
C   Shuffle   D   conf

154. Which one of the following statements is false regarding a MapReduce job? **C**

A   You can specify the number of Reducers using the setNumReduceTasks method of the JobConf object.   B   During the shuffling, the records are sorted by key and the values are combined into a collection.
C   You can specify the number of Mappers using the setNumMapTasks method of the JobConf object.   D   All records that share the same key are sent to the same Reducer.

155. Which MapReduce stage serves as a barrier, where all previous stages must be completed before it may proceed? **A**

A   Combine   B   shuffle
C   Reduce   D   Write

156. Which MapReduce phase is theoretically able to utilize features of the underlying file system in order to optimize parallel execution? **A**

A   Split   B   Combine
C   Map   D   sort

157. The new Hadoop API is in the ----------package **B**

A   org.apache.hadoop.mapred   B   org.apache.hadoop.mapreduce
C   org.apache.hadoop.maprer   D   org.apache.hadoop.map

158. In GFS, Chunkservers store the chunks on local disks as ------ **B**

A   Dos files   B   Linux files
C   Text file   D   Binary file

159. What are the steps in the lifecycle of mapper? **B**

A   JobTracker calls the map method of the mapper   B   Jobtracker calls the task tracker and task tracker calls the map method
C   Task tracker calls the job tracker,which inturn calls the map method   D   job tracker and map are parallelly called

160. Which of the following class is responsible for converting inputs to key-value Pairs of Map Reduce? **C**

A   FileInputFormat   B   InputSplit
C   RecordReader   D   Mapper

161. When 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? **B**

A   Combiner <Text, IntWritable,Text, IntWritable>   B   Reducer <Text, IntWritable,Text, IntWritable>

C   Combiner <Text,Text, IntWritable,   D   Combiner <Text, Text, IntWritable,
     IntWritable>   IntWritable>

162. How are keys and values presented and passed to the reducers during a standard sort **A**
and shuffle phase of MapReduce?

A   Keys are presented to reducer in   B   Keys are presented to reducer in
     sorted order; values for a given key   sorted order; values for a given key
     are not sorted   are sorted in ascending order

C   Keys are presented to a reducer in   D   Keys are presented to a reducer in
     random order; values for a given key   random order; values for a given key
     are not sorted   are sorted in ascending order

163. How many instances of Job tracker can run on Hadoop cluster ?   **A**

A   1   B   2
C   3   D   4

164. ---------- is also known as semi reducer.   **B**

A   mapper   B   combiner
C   shuffler   D   sorter

165. The number of partitioners is equal to the number of ---------   **D**

A   nodes   B   mapper
C   combiner   D   reducers

166. Which among the following controls the partitioning of the keys of the intermediate **C**
map-outputs?

A   Reader   B   Combiner
C   Partitioner   D   Writer

167. Job control is performed through the ----------class in the new API,   **B**

A   JobConf   B   Job
C   JobClient   D   Config

168. Which of the following is not the Dameon process that runs on a hadoop cluster ?   **C**

A   JobTracker   B   DataNode
C   TaskTracker   D   TaskNode

169. In the old API both map and reduce outputs are named as   **A**

A   part-nnnnn   B   part-nnnn
C   part-nnnnnn   D   part-nnn

170. Calling the combiner function zero, one, or many times should produce --------output **B**
from the reducer

A   zero   B   the same
C   null   D   key

171. MapReduce programming model is _____   **B**

A   Platform Dependent but not   B   Neither platform- nor language-
     language-specific   specific

C   Platform independent but language-   D   Platform Dependent and language-
     specific   specific

172. In a MapReduce job, the reducer receives all values associated with same key. Which **B**
statement best describes 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 arbitrary ordered, but D   Since the values come from mapper
     multiple runs of the same MapReduce   outputs, the reducers will receive
     job will always have the same   contiguous sections of sorted values
     ordering.

173. How can you disable the reduce step?   **C**

A   The Hadoop administrator has to set   B   It is imposible to disable the reduce
     the number of the reducer slot to zero   step since it is critical part of the Mep-
     on all slave nodes. This will disable   Reduce abstraction.

the reduce step.

C   A developer can always set the number of the reducers to zero. That will completely disable the reduce step.

D   While you cannot completely disable reducers you can set output to one. There needs to be at least one reduce step in Map-Reduce abstraction

174. What are map files and why are they important?   **D**

A   Map files are stored on the namenode and capture the metadata for all blocks on a particular rack. This is how Hadoop is "rack aware"

B   Map files are the files that show how the data is distributed in the Hadoop cluster.

C   Map files are generated by Map-Reduce after the reduce step. They show the task distribution during job execution

D   Map files are sorted sequence files that also have an index. The index allows fast data look up.

175. What are supported programming languages for Map Reduce?   **A**

A   The most common programming language is Java, but scripting languages are also supported via Hadoop streaming.

B   Any programming language that can comply with Map Reduce concept can be supported.

C   Only Java supported since Hadoop was written in Java.

D   Currently Map Reduce supports Java, C, C++ and COBOL.

176. The ------command is used to check if the Hadoop daemons are running properly or not **B**

A   job
B   jps
C   dir
D   ls

177. Your client application submits a MapReduce job to your Hadoop cluster. Identify the Hadoop daemon on which the Hadoop framework will look for an available slot schedule a MapReduce operation.   **D**

A   TaskTracker
B   NameNode
C   DataNode
D   JobTracker

178. Which of the following is the default partitioner for partitioning key space?   **B**

A   Part
B   Hash partitioner
C   Partitioner
D   Hpartitioner

179. The number of maps is usually driven by the total size of   **A**

A   inputs
B   tasks
C   jobs
D   outputs

180. What happens if number of reducers are set to 0?   **D**

A   nothing happens
B   reducer output will be final output
C   Reduce only job takes place
D   Map only job takes place