PASS

*killexams.com*

**Hitachi**

# HCAHD

*Apache Hadoop Developer*

ORDER FULL VERSION

## Question: 24

Assuming the following Hive query executes successfully:

```
from inputdata select context_ngrams(sentences(lines),
array("you", "are", null), 80);
```

Which one of the following statements describes the result set?

A. A bigram of the top 80 sentences that contain the substring "you are" in the lines column of the input data A1 table.
B. An 80-value ngram of sentences that contain the words "you" or "are" in the lines column of the inputdata table.
C. A trigram of the top 80 sentences that contain "you are" followed by a null space in the lines column of the inputdata table.
D. A frequency distribution of the top 80 words that follow the subsequence "you are" in the lines column of the inputdata table.

**Answer: D**

## Question: 25

Given the following Pig commands:

```
logevents = LOAD 'input/my.log';
severe = FILTER logevents BY ($1 =='severe' AND $2 >= 500);
grouped = GROUP severe BY $2;
DUMP grouped;
```

Which one of the following statements is true?

A. The $1 variable represents the first column of data in 'my.log'
B. The $1 variable represents the second column of data in 'my.log'
C. The severe relation is not valid
D. The grouped relation is not valid

**Answer: B**

## Question: 26

What does Pig provide to the overall Hadoop solution?

A. Legacy language Integration with MapReduce framework
B. Simple scripting language for writing MapReduce programs
C. Database table and storage management services
D. C++ interface to MapReduce and data warehouse infrastructure

**Answer: B**

## Question: 27

What types of algorithms are difficult to express in MapReduce v1 (MRv1)?

A. Algorithms that require applying the same mathematical function to large numbers of individual binary records.
B. Relational operations on large amounts of structured and semi-structured data.
C. Algorithms that require global, sharing states.
D. Large-scale graph algorithms that require one-step link traversal.
E. Text analysis algorithms on large collections of unstructured text (e.g, Web crawls).

## Answer: A,C,E

Explanation:

See 3) below.

Limitations of Mapreduce C where not to use Mapreduce

While very powerful and applicable to a wide variety of problems, MapReduce is not the answer to every problem. Here are some problems I found where MapReudce is not suited and some papers that address the limitations of MapReuce.

## Question: 28

You need to create a job that does frequency analysis on input data. You will do this by writing a Mapper that uses TextInputFormat and splits each value (a line of text from an input file) into individual characters. For each one of these characters, you will emit the character as a key and an InputWritable as the value.

As this will produce proportionally more intermediate data than input data, which two resources should you expect to be bottlenecks?

A. Processor and network I/O
B. Disk I/O and network I/O
C. Processor and RAM
D. Processor and disk I/O

## Answer: B

## Question: 29

Which one of the following statements regarding the components of YARN is FALSE?

A. A Container executes a specific task as assigned by the ApplicationMaster
B. The ResourceManager is responsible for scheduling and allocating resources
C. A client application submits a YARW job to the ResourceManager
D. The ResourceManager monitors and restarts any failed Containers

## Answer: D

## Question: 30

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. Mapper <Text, IntWritable, Text, IntWritable>
C. Reducer <Text, Text, IntWritable, IntWritable>
D. Reducer <Text, IntWritable, Text, IntWritable>
E. Combiner <Text, Text, IntWritable, IntWritable>

## Answer: D

## Question: 31

Which one of the following Hive commands uses an HCatalog table named x?

A. SELECT * FROM x;
B. SELECT x.-FROM org.apache.hcatalog.hive.HCatLoader('x');
C. SELECT * FROM org.apache.hcatalog.hive.HCatLoader('x');
D. Hive commands cannot reference an HCatalog table

## Answer: C

## Question: 32

Given the following Pig command:

logevents = LOAD &apos;input/my.log&apos; AS (date:chararray, levehstring, code:int, message:string);

Which one of the following statements is true?

A. The logevents relation represents the data from the my.log file, using a comma as the parsing delimiter
B. The logevents relation represents the data from the my.log file, using a tab as the parsing delimiter
C. The first field of logevents must be a properly-formatted date string or table return an error
D. The statement is not a valid Pig command

## Answer: B

## Question: 33

Consider the following two relations, A and B.

```
A = LOAD 'data1' AS (a1:int,a2:chararray);
DUMP A;
(1,apple)
(3,orange)
(4,peach)
(2,cherry)

B = LOAD 'data2' AS (b1:chararray,b2:int);
DUMP B;
(Jim,2)
(Brian,4)
(Kim,0)
(Terry,3)
(Chris,2)
```

A. C = DOIN B BY a1, A by b2;
B. C = JOIN A by al, B by b2;
C. C = JOIN A a1, B b2;
D. C = JOIN A SO, B $1;

**Answer: B**

## Question: 34

Given the following Hive commands:

```
CREATE TABLE mytable (name chararray, age int) ROW FORMAT
DELIMITED FIELDS TERMINATED BY ',' STORED AS
TEXTFILE;
LOAD DATA INPATH '/home/user/mydata.txt' INTO TABLE mytable;
```

Which one of the following statements Is true?

A. The file mydata.txt is copied to a subfolder of /apps/hive/warehouse
B. The file mydata.txt is moved to a subfolder of /apps/hive/warehouse
C. The file mydata.txt is copied into Hive's underlying relational database 0.
D. The file mydata.txt does not move from Its current location in HDFS

**Answer: A**

## Question: 35

In a MapReduce job, the reducer receives all values associated with same key.

Which 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 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.

**Answer: A,B**

Explanation:

Note:

* Input to the Reducer is the sorted output of the mappers.

* The framework calls the application's Reduce function once for each unique key in the sorted order.

* Example:

For the given sample input the first map emits:

< Hello, 1>

< World, 1>

< Bye, 1>

< World, 1>

The second map emits:

< Hello, 1>

< Hadoop, 1>

< Goodbye, 1>

< Hadoop, 1>

## Question: 36

Which describes how a client reads a file from HDFS?

A. The client queries the NameNode for the block location(s). The NameNode returns the block location(s) to the client. The client reads the data directory off the DataNode(s).
B. The client queries all DataNodes in parallel. The DataNode that contains the requested data responds directly to the client. The client reads the data directly off the DataNode.
C. The client contacts the NameNode for the block location(s). The NameNode then queries the DataNodes for block locations. The DataNodes respond to the NameNode, and the NameNode redirects the client to the DataNode that holds the requested data block(s). The client then reads the data directly off the DataNode.
D. The client contacts the NameNode for the block location(s). The NameNode contacts the DataNode that holds the requested data block. Data is transferred from the DataNode to the NameNode, and then from the NameNode to the client.

## Answer: A,C,D

Explanation:

Reference: 24 Interview Questions & Answers for Hadoop MapReduce developers, How the Client communicates with HDFS?

## Question: 37

For each input key-value pair, mappers can emit:

A. As many intermediate key-value pairs as designed. There are no restrictions on the types of those key-value pairs (i.e., they can be heterogeneous).
B. As many intermediate key-value pairs as designed, but they cannot be of the same type as the input key-value pair.
C. One intermediate key-value pair, of a different type.
D. One intermediate key-value pair, but of the same type.
E. As many intermediate key-value pairs as designed, as long as all the keys have the same types and all the values have the same type.

**Answer: A,E**

Explanation:

Mapper maps input key/value pairs to a set of intermediate key/value pairs.

Maps are the individual tasks that transform input records into intermediate records. The transformed intermediate records do not need to be of the same type as the input records. A given input pair may map to zero or many output pairs.

Reference: Hadoop Map-Reduce Tutorial

## Question: 38

You write MapReduce job to process 100 files in HDFS. Your MapReduce algorithm uses TextInputFormat: the mapper applies a regular expression over input values and emits key-values 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 reduces to one and settings the number of reducers to zero.

A. There is no difference in output between the two settings.
B. With zero reducers, no reducer runs and the job throws an exception. With one reducer, instances of matching patterns are stored in a single file on HDF
C. With zero reducers, all instances of matching patterns are gathered together in one file on HDF
D. With one reducer, instances of matching patterns are stored in multiple files on HDF
E. With zero reducers, instances of matching patterns are stored in multiple files on HDF
F. With one reducer, all instances of matching patterns are gathered together in one file on HDF

**Answer: A,C,D**

Explanation:

* It is legal to set the number of reduce-tasks to zero if no reduction is desired.

In this case the outputs of the map-tasks go directly to the FileSystem, into the output path set by setOutputPath(Path). The framework does not sort the map-outputs before writing them out to the FileSystem.

* Often, you may want to process input data using a map function only. To do this, simply set mapreduce.job.reduces to zero. The MapReduce framework will not create any reducer tasks. Rather, the outputs of the mapper tasks will be the final output of the job.

Note:

Reduce

In this phase the reduce(WritableComparable, Iterator, OutputCollector, Reporter) method is called for each <key, (list of values)> pair in the grouped inputs.

The output of the reduce task is typically written to the FileSystem via OutputCollector.collect(WritableComparable, Writable).

Applications can use the Reporter to report progress, set application-level status messages and update Counters, or just indicate that they are alive.

The output of the Reducer is not sorted.


## Question: 39

In Hadoop 2.0, which one of the following statements is true about a standby NameNode?

The Standby NameNode:

A. Communicates directly with the active NameNode to maintain the state of the active NameNode.
B. Receives the same block reports as the active NameNode.
C. Runs on the same machine and shares the memory of the active NameNode.
D. Processes all client requests and block reports from the appropriate DataNodes.

## Answer: B

## Question: 40

In the reducer, the MapReduce API provides you with an iterator over Writable values.

What does calling the next () method return?

A. It returns a reference to a different Writable object time.
B. It returns a reference to a Writable object from an object pool.
C. It returns a reference to the same Writable object each time, but populated with different data.
D. It returns a reference to a Writable object. The API leaves unspecified whether this is a reused object or a new object.
E. It returns a reference to the same Writable object if the next value is the same as the previous value, or a new Writable object otherwise.

## Answer: A,C,E

Explanation:

Calling Iterator.next() will always return the SAME EXACT instance of IntWritable, with the contents of that instance replaced with the next value.

Reference: manupulating iterator in mapreduce

# SAMPLE QUESTIONS

*These questions are for demo purpose only.* **Full version** *is up to date and contains actual questions and answers.*

Killexams.com is an online platform that offers a wide range of services related to certification exam preparation. The platform provides actual questions, exam dumps, and practice tests to help individuals prepare for various certification exams with confidence. Here are some key features and services offered by Killexams.com:

**ORDER FULL VERSION**
Special Discount Coupon

**Actual Exam Questions**: Killexams.com provides actual exam questions that are experienced in test centers. These questions are updated regularly to ensure they are up-to-date and relevant to the latest exam syllabus. By studying these actual questions, candidates can familiarize themselves with the content and format of the real exam.

**Exam Dumps**: Killexams.com offers exam dumps in PDF format. These dumps contain a comprehensive collection of questions and answers that cover the exam topics. By using these dumps, candidates can enhance their knowledge and improve their chances of success in the certification exam.

**Practice Tests**: Killexams.com provides practice tests through their desktop VCE exam simulator and online test engine. These practice tests simulate the real exam environment and help candidates assess their readiness for the actual exam. The practice tests cover a wide range of questions and enable candidates to identify their strengths and weaknesses.

**Guaranteed Success**: Killexams.com offers a success guarantee with their exam dumps. They claim that by using their materials, candidates will pass their exams on the first attempt or they will refund the purchase price. This guarantee provides assurance and confidence to individuals preparing for certification exams.

**Updated Content:** Killexams.com regularly updates its question bank and exam dumps to ensure that they are current and reflect the latest changes in the exam syllabus. This helps candidates stay up-to-date with the exam content and increases their chances of success.

**Technical Support**: Killexams.com provides free 24x7 technical support to assist candidates with any queries or issues they may encounter while using their services. Their certified experts are available to provide guidance and help candidates throughout their exam preparation journey.

For More exams visit https://killexams.com/vendors-exam-list
*Kill your exam at First Attempt....Guaranteed!*