

Hortonworks HDPCD Exam

Volume: 108 Questions

Question No : 1

Which one of the following statements describes a Hive user-defined aggregate function?

- A. Operates on multiple input rows and creates a single row as output
- B. Operates on a single input row and produces a single row as output
- C. Operates on a single input row and produces a table as output
- D. Operates on multiple input rows and produces a table as output

Answer: A

Question No : 2

Identify the tool best suited to import a portion of a relational database every day as files into HDFS, and generate Java classes to interact with that imported data?

- A. Oozie
- B. Flume
- C. Pig
- D. Hue
- E. Hive
- F. Sqoop
- G. fuse-dfs

Answer: F

Question No : 3

Review the following data and Pig code.

M,38,95111 F,29,95060 F,45,95192 M,62,95102 F,56,95102

A = LOAD 'data' USING PigStorage(',') as (gender:Chararray, age:int,

Hortonworks HDPCD Exam

zlp:chararray);

B = FOREACH A GENERATE age;

Which one of the following commands would save the results of B to a folder in hdfs named myoutput?

- A. STORE A INTO 'myoutput' USING PigStorage('');
- B. DUMP B using PigStorage('myoutput');
- C. STORE B INTO 'myoutput';
- D. DUMP B INTO 'myoutput';

Answer: C

Question No : 4

You have user profile records in your OLPT database, that you want to join with web logs you have already ingested into the Hadoop file system. How will you obtain these user records?

- A. HDFS command
- B. Pig LOAD command
- C. Sqoop import
- D. Hive LOAD DATA command
- E. Ingest with Flume agents
- F. Ingest with Hadoop Streaming

Answer: C

Question No : 5

Consider the following two relations, A and B.

Hortonworks HDPCD Exam

```
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)
```

Which Pig statement combines A by its first field and B by its second field?

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

Answer: B

Question No : 6

Which one of the following statements is true regarding a MapReduce job?

- A. The job's Partitioner shuffles and sorts all (key.value) pairs and sends the output to all reducers
- B. The default Hash Partitioner sends key value pairs with the same key to the same Reducer
- C. The reduce method is invoked once for each unique value
- D. The Mapper must sort its output of (key.value) pairs in descending order based on value

Answer: A

Question No : 7

Review the following data and Pig code:

Hortonworks HDPCD Exam

```
M, 38, 95111
F, 29, 95060
F, 45, 95192
M, 62, 95102
F, 56, 95102
```

```
A = LOAD 'data' USING PigStorage(',')
AS (gender:chararray, age:int, zip:chararray);
```

What command to define B would produce the output (M,62,95102) when invoking the DUMP operator on B?

- A. B = FILTER A BY (zip == '95102' AND gender == 'M');
- B. B = FOREACH A BY (gender == 'M' AND zip == '95102');
- C. B = JOIN A BY (gender == 'M' AND zip == '95102');
- D. B = GROUP A BY (zip == '95102' AND gender == 'M');

Answer: A

Question No : 8

Which one of the following classes would a Pig command use to store data in a table defined in HCatalog?

- A. org.apache.hcatalog.pig.HCatOutputFormat
- B. org.apache.hcatalog.pig.HCatStorer
- C. No special class is needed for a Pig script to store data in an HCatalog table
- D. Pig scripts cannot use an HCatalog table

Answer: B

Question No : 9

Given the following Pig command:

Hortonworks HDPCD Exam

logevents = LOAD 'input/my.log' AS (date:chararray, level:string, 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 No : 10

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 No : 11

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)
```

What is the output of the following Pig commands?