**Hadoop Questions and Answers – Introduction to Pig**

**This set of Questions and Answers focuses on “Pig”.**

1. Pig operates in mainly how many nodes ?  
**a) Two**  
b) Three  
c) Four  
d) Five  
View Answer

**Answer: a  
Explanation: You can run Pig (execute Pig Latin statements and Pig commands) using various mode: Interactive and Batch Mode.**

2. Point out the correct statement :  
**a) You can run Pig in either mode using the “pig” command**  
b) You can run Pig in batch mode using the Grunt shell  
c) You can run Pig in interactive mode using the FS shell  
d) None of the mentioned  
  
  
 run Pig in either mode using the “pig” command (the bin/pig Perl script) or the “java” command (java -cp pig.jar …).

3. You can run Pig in batch mode using \_\_\_\_\_\_\_\_\_\_ .  
a) Pig shell command  
**b) Pig scripts**  
c) Pig options  
d) All of the mentioned  
**View Answer**

**Answer: b  
Explanation: Pig script contains Pig Latin statements.**

4. 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 mentioned  
View Answer**

**Answer: d  
Explanation: A DUMP or STORE statement is required to generate output.**

5. Point out the wrong statement :  
a) To run Pig in local mode, you need access to a single machine  
**b) The DISPLAY operator will display the results to your terminal screen**c) To run Pig in mapreduce mode, you need access to a Hadoop cluster and HDFS installation  
d) All of the mentioned  
View Answer

Answer: b  
Explanation: The DUMP operator will display the results to your terminal screen.

6. Which of the following function is used to read data in PIG ?  
a) WRITE  
b) READ  
**c) LOAD**  
d) None of the mentioned  
View Answer

**Answer: c  
Explanation: PigStorage is the default load function.**

7. You can run Pig in interactive mode using the \_\_\_\_\_\_ shell.  
**a) Grunt**  
b) FS  
c) HDFS  
d) None of the mentioned  
View Answer

8. Which of the following is the default mode ?  
**a) Mapreduce**  
b) Tez  
c) Local  
d) All of the mentioned  
View Answer

Answer: a  
Explanation: To run Pig in mapreduce mode, you need access to a Hadoop cluster and HDFS installation.

9. Which of the following will run pig in local mode ?  
**a) $ pig -x local …**  
b) $ pig -x tez\_local …  
c) $ pig …  
d) None of the mentioned  
View Answer

Answer: a  
Explanation: Specify local mode using the -x flag (pig -x local).

10.$ pig -x tez\_local … will enable \_\_\_\_\_\_\_\_ mode in Pig.  
a) Mapreduce  
b) Tez  
c) Local  
**d) None of the mentioned**  
View Answer

Answer: d  
Explanation: Tez Local Mode is similar to local mode, except internally Pig will invoke tez runtime engine.

**Hadoop Questions and Answers – Pig Latin**

**This set of Hadoop Multiple Choice Questions & Answers (MCQs) focuses on “Pig Latin”.**

1.\_\_\_\_\_\_\_\_\_ operator is used to review the schema of a relation.  
a) DUMP  
**b) DESCRIBE**  
c) STORE  
d) EXPLAIN  
**View Answer**

**Answer: b  
Explanation: DESCRIBE returns the schema of a relation.**

2. Point out the correct statement :  
a) During the testing phase of your implementation, you can use LOAD to display results to your terminal screen  
**b) You can view outer relations as well as relations defined in a nested FOREACH statement**  
c) Hadoop properties are interpreted by Pig  
d) None of the mentioned  
View Answer

Answer: b  
Explanation: Viewing outer relations is possible using DESCRIBE operator.

3. Which of the following operator is used to view the map reduce execution plans ?  
a) DUMP  
b) DESCRIBE  
c) STORE  
**d) EXPLAIN**  
View Answer

Answer: d  
Explanation: EXPLAIN displays execution plans.

4. \_\_\_\_\_\_\_\_\_\_\_ operator is used to view the step-by-step execution of a series of statements.  
**a) ILLUSTRATE**  
b) DESCRIBE  
c) STORE  
d) EXPLAIN  
View Answer

Answer: a  
Explanation: ILLUSTRATE allows you to test your programs on small datasets and get faster turnaround times.

5. Point out the wrong statement :  
a) ILLUSTRATE operator is used to review how data is transformed through a sequence of Pig Latin statements  
b) ILLUSTRATE is based on an example generator  
**c) Several new private classes make it harder for external tools such as Oozie to integrate with Pig statistics**  
d) None of the mentioned  
View Answer

Answer: c  
Explanation: Several new public classes make it easier for external tools such as Oozie to integrate with Pig statistics.

6. \_\_\_\_\_\_\_\_\_\_ is a framework for collecting and storing script-level statistics for Pig Latin.  
a) Pig Stats  
b) PStatistics  
**c) Pig Statistics**  
d) None of the mentioned  
View Answer

Answer: c  
Explanation: The new Pig statistics and the existing Hadoop statistics can also be accessed via the Hadoop job history file.

7. The \_\_\_\_\_\_\_\_ class mimics the behavior of the Main class but gives users a statistics object back.  
a) PigRun  
**b) PigRunner**  
c) RunnerPig  
d) None of the mentioned  
**View Answer**

**Answer: b  
Explanation: Optionally, you can call the API with an implementation of progress listener which will be invoked by Pig runtime during the execution.**

8. \_\_\_\_\_\_\_\_\_\_\_ is a simple xUnit framework that enables you to easily test your Pig scripts.  
a) PigUnit  
**b) PigXUnit**  
c) PigUnitX  
d) All of the mentioned  
View Answer

**Answer: b  
Explanation: With PigUnit you can perform unit testing, regression testing, and rapid prototyping. No cluster set up is required if you run Pig in local mode.**

9. Which of the following will compile the Pigunit ?  
**a) $pig\_trunk ant pigunit-jar**  
b) $pig\_tr ant pigunit-jar  
c) $pig\_ ant pigunit-jar  
d) None of the mentioned  
View Answer

Answer: a  
Explanation: The compile will create the pigunit.jar file.

10. PigUnit runs in Pig’s \_\_\_\_\_\_\_ mode by default.  
**a) local**  
b) tez  
c) mapreduce  
d) none of the mentioned  
**View Answer**

**Answer: a  
Explanation: Local mode does not require a real cluster but a new local one is created each time.**