UNIT – II

1. _____ is a platform for constructing data flows for extract, transform, and load (ETL) processing and analysis of large datasets.
a) Pig Latin
b) Oozie
c) Pig
d) Hive

2. 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 mentioned

3. _____ hides the limitations of Java behind a powerful and concise Clojure API for Cascading.
a) Scalding
b) HCatalog
c) Cascalog
d) All of the mentioned

4. Hive also support custom extensions written in :
a) C#
b) Java
c) C
d) C++

5. Point out the wrong statement :
a) Elastic MapReduce (EMR) is Facebook's packaged Hadoop offering
b) Amazon Web Service Elastic MapReduce (EMR) is Amazon's packaged Hadoop offering

c) Scalding is a Scala API on top of Cascading that removes most Java boilerplate

d) All of the mentioned

6. _____ is the most popular high-level Java API in Hadoop Ecosystem

a) Scalding

b) HCatalog

c) Cascalog

d) Cascading

7. _____ is general-purpose computing model and runtime system for distributed data analytics.

a) Mapreduce

b) Drill

c) Oozie

d) None of the mentioned

8. The Pig Latin scripting language is not only a higher-level data flow language but also has operators similar to :

a) SQL

b) JSON

c) XML

d) All of the mentioned

9. _____ jobs are optimized for scalability but not latency.

a) Mapreduce

b) Drill

c) Oozie

d) Hive

10. _____ is a framework for performing remote procedure calls and data serialization.

a) Drill

b) BigTop

c) Avro

d) Chukwa