Session 4: Hadoop Configuration and Eco System

Assignment

ACADGILD

Assignment 2–Quiz

Table of Contents

1. Introduction

In this assignment you need perform quiz on topic covered in the class and additional topics related to the class.

2. Objective

This assignment will help you to understand concepts of HDFS and it's components

3. Associated Data Files

No files

4. Problem Statement

1. The _____is the heart of an HDFS file system. It keeps the metadata such as directory tree of all files in the file system and tracks where across the cluster the file data is kept. The actual data is stored on _____as HDFS blocks

a) datanode,datanode
b) resourcemanager,namenode
c) namenode,datanode
d) tasktraker,jobtraker

2. When the end of the block is reached, _____closes the connection to the datanode, then finds the best datanode for the next block.

    a) DFSOutputStream
    b) client
    c) DFSInputStream
    d) Resourcemanager

3. When the client finishes reading, it calls ____method on the close the stream

    a) read()
    b) write()
    c) shuffle
    d) close()

4. which configuration file contains Environmental variable settings used by Hadoop
    a) Core-site.xml
    b) mapred-site.xml
    c) yarn-site.xml

d) Hadoop-env.sh

5. Which MapReduce daemon instantiates user code, and executes map and reduce tasks on a cluster running MapReducevl (MRvl)?

   A. NameNode

   B. DataNode

   C. JobTracker

   D. TaskTracker

6. Identify the function performed by the Secondary NameNode daemon on a cluster configured to run with a single NameNode.

   A. In this configuration, the Secondary NameNode performs a checkpoint operation on the files by the NameNode.

   B. In this configuration, the Secondary NameNode is standby NameNode, ready to failover and provide high availability.

   C. In this configuration, the Secondary NameNode performs deal-time backups of the NameNode.

   D. In this configuration, the Secondary NameNode servers as alternate data channel for clients to reach HDFS, should the NameNode become too busy.

7. Hadoop administrators write a script called Topology script to determine the rack location of nodes. It triggers to know the distance of the nodes to replicate the data and Configure's this script in _____.

a) yarn-site.xml
b) Hadoop-env.sh
c) core-site.xml
d) mapred-site.env

8. _____is the master that arbitrates all the available cluster resources and thus helps manage the distributed applications running on the YARN system

a) Node manager
b) data manager
c) ResourceManager (RM)
d) Taskmanager

9. _____take instructions from the ResourceManager and manage resources available on a single node.

a) NodeManagers
b) data manager
c) ResourceManager (RM)
d) Taskmanager

10. How does HDFS Federation help HDFS Scale horizontally?

A. HDFS Federation improves the resiliency of HDFS in the face of network issues by removing the NameNode as a single-point-of-failure.

B. HDFS Federation allows the Standby NameNode to automatically resume the services of an activeNameNode.

C. HDFS Federation provides cross-data center (non-local) support for HDFS, allowing a cluster administrator to split the Block Storage outside the local cluster.

D. HDFS Federation reduces the load on any single NameNode by using the multiple, independent NameNode to manage individual pars of the filesystem namespace.

5. Approximate Time to Complete Task

20 min