428 Quiz 2: Hadoop Cluster Setup and Management

1. **Which of the following is NOT a core component of Hadoop?**

    o   a. HDFS

    o   b. YARN

    o   c. MapReduce

    o   d. Hive

2. **What is the primary purpose of the core-site.xml file in Hadoop configuration?**

    o   a. Define HDFS namenode address

    o   b. Configure YARN resource manager

    o   c. Specify MapReduce job settings

    o   d. Set general Hadoop cluster properties

3. **Which Hadoop component is responsible for managing cluster resources and scheduling jobs?**

    o   a. HDFS

    o   b. YARN

    o   c. MapReduce

    o   d. ZooKeeper

4. **What is the main advantage of setting up a High Availability (HA) cluster in Hadoop?**

    o   a. Improved data locality

    o   b. Increased storage capacity

    o   c. Fault tolerance and continuous operation

    o   d. Faster job execution

5. **Which of the following tools can be used for comprehensive Hadoop cluster monitoring and performance analysis?**

    o   a. Ganglia

    o   b. Nagios

    o   c. Both a and b

    o   d. Neither a nor b

6. **What is the role of Kerberos in Hadoop security?**

    o   a. Data encryption

    o   b. Access control

    o   c. Authentication

- o d. Auditing

7. **Which configuration file is used to set the HDFS namenode address?**

   - o a. core-site.xml

   - o b. hdfs-site.xml

   - o c. yarn-site.xml

   - o d. mapred-site.xml

8. **What is the purpose of access control lists (ACLs) in Hadoop?**

   - o a. Encrypt data at rest

   - o b. Restrict access to sensitive data

   - o c. Manage cluster resources

   - o d. Schedule jobs

9. **Which of the following is an example of configuring Hadoop cluster architecture?**

   - o a. Setting up SSH key-based authentication

   - o b. Configuring a federated cluster with multiple namenodes

   - o c. Enabling Kerberos authentication

   - o d. Using Apache Ambari for cluster management

10. **What is the benefit of using tools like Elasticsearch, Logstash, and Kibana (ELK stack) in Hadoop cluster management?**

    - o a. Centralized cluster configuration

    - o b. Log aggregation and analysis for troubleshooting

    - o c. Data encryption

    - o d. Role-based access control

11. **Which Hadoop component stores the actual data files in the cluster?**

    - o a. Namenode

    - o b. Datanode

    - o c. Resource Manager

    - o d. Node Manager

12. **What is the purpose of configuring a cluster with multiple racks?**

    - o a. Increase storage capacity

    - o b. Improve fault tolerance

    - o c. Optimize data locality and network bandwidth

- o d. Enable Kerberos authentication

13. **Which of the following is NOT a typical task in Hadoop cluster management and monitoring?**

    - o a. Viewing cluster metrics and node status
    - o b. Setting up alerts and notifications
    - o c. Configuring data encryption
    - o d. Using tools like Ganglia or Nagios for performance analysis

14. **What is the main goal of implementing transparent data encryption in HDFS?**

    - o a. Control access to sensitive data
    - o b. Protect data at rest without application changes
    - o c. Ensure secure communication between cluster components
    - o d. Grant fine-grained permissions to users

15. **Which environment variable is commonly set in the hadoop-env.sh file?**

    - o a. HADOOP_HOME
    - o b. JAVA_HOME
    - o c. PATH
    - o d. CLASSPATH

16. **What is the advantage of setting up a heterogeneous cluster in Hadoop?**

    - o a. Simplified cluster management
    - o b. Optimized resource utilization with different hardware
    - o c. Increased storage capacity
    - o d. Improved fault tolerance

17. **Which tool can be used for centralized Hadoop cluster management, monitoring, and configuration?**

    - o a. Ganglia
    - o b. Nagios
    - o c. Apache Ambari
    - o d. Elasticsearch

18. **What is the purpose of role-based access control (RBAC) in Hadoop security?**

    - o a. Encrypt data in transit
    - o b. Grant fine-grained permissions to users and groups
    - o c. Restrict access to sensitive data

- o d. Authenticate users

19. **Which configuration file is used to set YARN resource manager settings?**

    - o a. core-site.xml

    - o b. hdfs-site.xml

    - o c. yarn-site.xml

    - o d. mapred-site.xml

20. **What is the primary focus of the "Hadoop Security" submodule?**

    - o a. Cluster performance optimization

    - o b. Data locality improvement

    - o c. Protecting the cluster and its data from unauthorized access

    - o d. Centralized cluster management

Answers:

1. d
2. d
3. b
4. c
5. c
6. c
7. b
8. b
9. b
10. b
11. b
12. c
13. c
14. b
15. b
16. b
17. c
18. b
19. c
20. c