

Session 3: HDFS Internals

Assignment 1

**A C A D G I L D Page 1**

*Assignment 1 – Try the given quiz questions and provide the answers in a word document.*

**Table of Contents**

1. Introduction .......................................................................................................................................... 3

2. Objective ............................................................................................................................................... 3

3. Prerequisites: ........................................................................................................................................ 3

4. Associated Data Files ............................................................................................................................ 3

5. Problem Statement: .............................................................................................................................. 3

6. Expected Output ................................................................................................................................... 5

7. Approximate Time to Complete Task ................................................................................................... 5



|  |  |  |  |
| --- | --- | --- | --- |
|  | **A C A D G I L D** | **Page 2** |  |
|  |  | | |

**1. Introduction**

In this assignment you need to select one right choice for the questions given on the topics discussed in the third session.

**2. Objective**

This assignment will help you to consolidate the concepts learnt in the session 3.

**3. Prerequisites:**

None

**4. Associated Data Files**

None

**5. Problem Statement:**

**1. HDFS is built around the idea that data is written but read many times.**

a) many

b) twice

c) data already exists d) once

**Solution: d**

**2. Hadoop divides input into fixed size pieces called what?**

a) output result b) input splits

c) input data d) input blogs

**Solution: b**

**3. All the blocks are replicated in other nodes for**

a) security b) big data c) pool

d) fault tolerance



**Solution: d**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **A C A D G I L D** | **Page 3** |  |
|  |  | | |

**4. Block size can be changed using the properties in**

a) core-site.xml

b) Hadoop-env.sh c) hdfs-site.xml

d) yarn-site.xml

**Solution: c**

**5. Hadoop uses the representation of the data stored in the file blocks known as Input splits.**

a) physical b) logical

c) mechanical d) none

**Solution: b**

**6. DFS calls NameNode to create file in file system’s**

a) dataspace

b) resourcespace c) namespace

d) nodespace

**Solution: c**

**7. Data packets are streamed to first DataNode in the \_**

a) handshake b) pipeline

c) hard disk

d) hdfs

**Solution: b**

**8. The client has finished writing data, it calls \_on the stream.**

a) close() b) read() c) open() d) check()



|  |  |  |  |
| --- | --- | --- | --- |
|  | **A C A D G I L D** | **Page 4** |  |
|  |  | | |

**Solution: a**

**9. Blocks are read in order, with the opening new**

**connections to datanodes as the client reads through the stream.**

a) DFSoutputstream b) DFSInputStream

c) DFStrackManager

d) DFSStringConcatination

**Solution: a**

**10. If I have 100 input splits, how many maps will run?**

a) 200 b) 50 c) 100 d) 1

**Solution: c**

**6. Expected Output**

None

**7. Approximate Time to Complete Task**

15 mins



|  |  |  |  |
| --- | --- | --- | --- |
|  | **A C A D G I L D** | **Page 5** |  |
|  |  | | |