

**Attempt 1 of 1**

Written Feb 14, 2023 3:30 PM - Feb 14, 2023 5:20 PM

Your quiz has been submitted successfully.

Attempt Score 69 / 100 - D+

Overall Grade (Highest Attempt) 69 / 100 - D+

**1. Firebase, 25 points**

Consider storing air quality index (AQI) data on 7/21/2022 in Firebase database as shown below. Write a curl command for each of the following requests. The base url is: <https://dsci55x-default-rtdb.firebaseio.com>. You may assume all needed rule settings have been set properly.

Your codes are expected to run properly in the terminal. Please make sure your code will not generate any random key.

**Question 1****4 / 5 points**

Add a new entry for USA under AQI with status = "Unhealthy" and Value = 116. Please use the same layout shown in the screenshot.

```
curl -X PUT "https://dsci55x-default-rtdb.firebaseio.com/AQI/USA.json" -d "{\"status": \"Unhealthy\", \"Value\": 116}"
```

▼ Hide question 1 feedback

quotation for json String must be double -0.5

quotation for json object must be single -0.5

make sure you understand the meaning of quotation for json

**Question 2****5 / 5 points**

Update the value for Canada from 62 to 63. Your command should not remove/overwrite the Status of Canada.

```
curl -X PATCH "https://dsci55x-default-rtdb.firebaseio.com/AQI/Canada.json" -d "{\"Value\": 63}"
```

▼ Hide question 2 feedback

same problem as q1, only deduct points once

**Question 3****1 / 5 points**

Find all countries whose status is Good.

```
curl "https://dsci55x-default-rtdb.firebaseio.com/AQI.json?"
```

[▼ Hide question 3 feedback](#)

```
orderBy="Status" -2  
equalTo="Good" -2
```

#### Question 4

4.5 / 5 points

Find the first five countries ordered by the names of countries (ascending).

```
curl "https://dsci55x-default.firebaseio.com/AQI.json?orderBy='$key'&limitToFirst=5"
```

[▼ Hide question 4 feedback](#)

```
orderBy = "$key" -0.5
```

#### Question 5

5 / 5 points

Find the countries with the largest AQI value.

```
curl "https://dsci55x-default.firebaseio.com/AQI.json?orderBy='Value'&limitToLast=1"
```

[▼ Hide question 5 feedback](#)

same problem as q4, only deduct points once

### 2. Storage System - Hard Disk, 20 points

Consider a hard disk drive with the following characteristics:

- 10,000 RPM
- Maximum seek time: 15ms
- Sector size: 4KB (assume a block contains a single sector)
- Maximum bandwidth: 100MB/sec

You can assume 1000KB is roughly 1MB in this question.

#### Question 6

3 / 5 points

Compute the completion time for reading 4MB of data, assuming that the data are located on the same track on the drive.

For 10000 rpm  $\rightarrow$  60000ms/10000 = 6ms/rotation

avg rotation time = 3ms

Since the data is same track it will be read sequentially

completion\_time = t seek + t avg rotation + t transfer

= 15 + 3 + 4mb/100mb\*1000 = 15 + 3 + 40 = 58ms

[▼ Hide question 6 feedback](#)

Wrong answer and seeking time

#### Question 7

3 / 5 points

What is the actual bandwidth of the drive for the workload in question 6? Will the actual bandwidth increase, decrease or remain the same if we are to read 40MB of data on the same track instead? Explain your answer.

actual bandwidth = 4mb/completion time = 4mb/58ms = 68.96 mb/s

Actual bandwidth will increase as the completion time for reading 40mb data will increase which increases the actual bandwidth.

completion time for 40mb = 418ms

actual bandwidth = 40mb/418ms = 95.69 mb/s

[▼ Hide question 7 feedback](#)

Wrong actual bandwidth

#### Question 8

3 / 5 points

Compute the completion time for reading 4MB of data, assuming that the blocks of the data are randomly located on the drive.

$$\text{blocks that will be read} = 4\text{mb}/4\text{kb} = 1000 \text{ blocks}$$
$$\text{completion time} = 1000 * (15 + 3 + 4/100) = 18.04\text{s}$$

▼ Hide question 8 feedback

Wrong answer and seeking time

#### Question 9

1 / 5 points

What is the actual bandwidth of the drive for the workload in question 8? Will the actual bandwidth increase, decrease or remain the same if we are to read 40MB of data whose blocks are randomly located on the drive? Explain your answer.

$$\text{actual bandwidth} = 40\text{mb}/18.004 = 2.22 \text{ mb/s}$$

Actual bandwidth decreases as the completion for reading the data is nearly the increases with number of blocks to read increases to 10000 blocks

$$\text{completion time} = 180.4\text{s}$$

$$\text{actual bandwidth} = 40\text{mb}/180.4\text{s} = 0.221\text{mb/s}$$

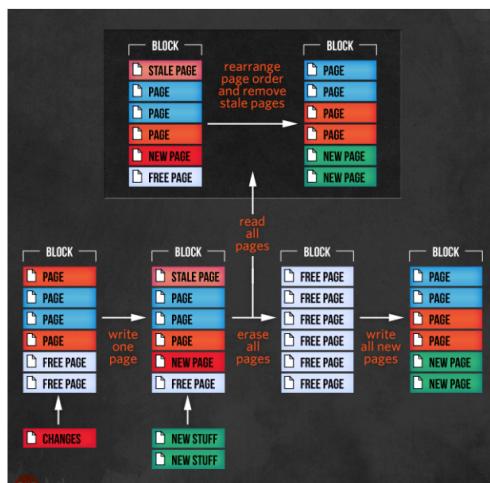
▼ Hide question 9 feedback

Wrong actual bandwidth, conclusion and explanation

### 3. Storage System - SSD, 15 points

Consider the SSD device you have seen in class. Recall that the device has only a single block, and the first request is to modify the first (red) page in the block and the second request is to add two (green) pages. Suppose SSD has the following characteristics:

- Read latency: 10 microseconds
- Write latency: 100 microseconds
- Erase latency: 1000 microseconds



Suppose that "read all pages" in the diagram means that the SSD reads all the six pages in the block into the device RAM buffer.

#### Question 10

5 / 5 points

Explain what the device controller will be doing to complete the requests.

The device controller reads the entire block into memory. Then it writes the new page to be written. All the pages are read and block is then erased. The read pages are rearranged and then written the erased

block.

#### Question 11

1 / 5 points

What is the completion time for the above requests? Be sure to show how you compute the completion time.

$$100+60+6000+200+600 = 6960\text{s}$$

[▼ Hide question 11 feedback](#)

Wrong answer, erase, write time

#### Question 12

0 / 5 points

Now consider the requests arriving in different order. That is, the first request will be to add two new pages and the second is to modify the red page. What is the completion time in this case? Show your derivation.

#### 4. File System, 15 points

Recall the file system image (XML format) that you have seen in class and is shown below.

```
<inode>
  <id>16390</id>
  <type>FILE</type>
  <name>core-site.xml</name>
  <replication>1</replication>
  <mtime>1675117007609</mtime>
  <atime>1675117007580</atime>
  <preferredBlockSize>134217728</preferredBlockSize>
  <permission>ubuntu:supergroup:0644</permission>
  <blocks>
    <block>
      <id>1073741826</id>
      <genstamp>1002</genstamp>
      <numBytes>884</numBytes>
    </block>
  </blocks>
  <storagePolicyId>0</storagePolicyId>
</inode>
```

#### Question 13

4.5 / 5 points

Explain the meaning of each value in the above fragment of XML file. No need to explain genstamp and storagePolicyId.

inode.id is the inumber

inode.type is file means it has metadata for file.

core-site.xml is the file name

1 replica of the file is present since there is only one datanode.

mtime is last modification time

atime is last access time

ubuntu:supergroup:0644 mentions the access for usergroups, owner and others

inode.blocks.block.id represent the block where metdata is stored

inode.blocks.block.numBytes = 884 which the size of core-site.xml

[▼ Hide question 13 feedback](#)

meaning of preferred block size is missing

#### Question 14

5 / 5 points

Explain what the 644 permission means.

110 100 100

-rw-r--r--

The permission states that the owner of the file can read and write to the file while the users in the usergroup and rest can only read the file.

Since the file is not an executable none has an executable permission.

#### Question 15

2 / 2 points

What is the default block size in HDFS? Which value in the above XML fragment is most likely to be the default block size of HDFS?

Default block size of HDFS is 128mb.  
inode.preferredBlockSize

#### Question 16

2 / 3 points

The file system image file also has a INodeDirectorySection shown below. Explain the purpose of data shown in the INodeDirectorySection section. Is it possible that the inode shown in the above question (with id = 16390) appears as the value of some parent element in the INodeDirectorySection? Explain your answer.

The INodeDirectorySection gives the mapping for directory and file tree corresponding to the root folder in hdfs.

No it is not possible to be representing a parent folder

[▼ Hide question 16 feedback](#)

Explanation missing - 16390 is not a parent because it is a file

#### 5. HDFS, 15 points

Suppose an HDFS client will be executing the following HDFS commands. For each command, explain the meaning of the command and how HDFS will execute the command. Be sure to specify which servers the client will need to contact, what the server should do to fulfill the request, and what may be returned from the server. Also explain the possible scenarios where the server may need to report an error.

#### Question 17

2 / 3 points

hdfs dfs -ls /user

This command lists all the files and directories under the /user folder.

First the client will request the search through the namenode for metadata corresponding to the user folder -> It will then lookup of the data in the nearest datanode to fetch data to list all the files and folder in that directory and will once again go through the namenode to get metadata for those files and folders.

[▼ Hide question 17 feedback](#)

Case of error missing (-1)

#### Question 18

3 / 4 points

hdfs dfs -mkdir /user/dsci551

Creates a new directory named dsci551 under user folder. Client request the namenode to generate metadata for the new directory to be created. Next the dir is replicated in all the datanodes sequentially with intermediate acknowledgements between namenode and datanodes.

[▼ Hide question 18 feedback](#)

Case of error missing (-1)

#### Question 19

4 / 5 points

hdfs -dfs -put hello.txt /user/dsci551

This command copies file hello.txt from local file system to hdfs dir /user/dsci551.

The file from local fs is read and stored into memory. If the dir in hdfs exists the file is copied by the datanode. Before the acknowledgement made to namenode where metadata for the file is created and then the file is copied all the datanodes and finally the request is closed.

[▼ Hide question 19 feedback](#)

Case of error missing (-1)

**Question 20****2.5 / 3 points**

```
hdfs -dfs -cat /user/dsci551/hello.txt
```

Print the contents of /user/dsci551/hello.txt to standard output.

Clients request the namenode for metadata corresponding the file to read. It then send the metadata to the nearest datanode having the replica of the file. Then the datanode reads the content of file from the block and inode information received from namenode finally printing it to standard output.

▼ [Hide question 20 feedback](#)

Case of error missing (-0.5)

**6. Miscellaneous, 10 points****Question 21****3 / 3 points**

What are 3V's of big data? Explain your answer.

Velocity - The pace at which we are receiving the data from data sources. Streaming and Batch data.

Variety - receiving data in different formats - structured(CSV, TSV, etc), unstructured(images, videos, etc), semi-structured(JSON, etc)

Volume - The volume of the incoming data and data to be processed.

**Question 22****2 / 2 points**

Give a Linux command that only changes the metadata of a file. Explain your answer.

chmod - change mode

This command changes the access permissions related to a file. It controls how user, usergroups and others can work with the file.

**Question 23****3.5 / 5 points**

Write a Python script that doubles the even values in a list of integers. Note that you are required to use map and/or filter function of Python. Suppose the list is contained in a variable called lst. For example, lst may be [3, 2, 5, 4, 6]. In this case, the script will return [3, 4, 5, 8, 12].

```
lst = [3, 2, 5, 8, 12]
lst = map(lst, lambda: x*2 if x%2==0 else x)
```

▼ [Hide question 23 feedback](#)

map function syntax wrong(-1)

.....lambda x: x\*2.....: x is missing after lambda(-0.5)

**Done**