Question Set O: The ApiJavaOne.java Program

1. *Q: What is the overall function of this program?*

   The functionality of ApiJavaOne is to read the files that resides in the HDFS and list them with detailed information of what the files contains.

2. *Q: Study the handle of "fs" called listFiles().*

   a. *Q: Describe, in natural language, what this handle does.*

      A: The handle *listFiles* is responsible for listing statuses and block locations of the files that are in each path.

   b. *Q: The value 'true' as an argument for listFiles() refers to 'recursive'. What does this argument do, and what would have happened if it were set to false?*

      A: Q: If we set the recursive to false it will return files in the directory, but if the recursive shows true it will return files that are sub-rooted. If the path is a file, it will return the statue of the file and block locations.

   Q: The script contains a while loop, which could obviously also have been repurposed as a for loop. Give some examples of how these control structures could benefit the user as opposed to using the command line (hdfs dfs).

   A: By using loops we save a lot of time, instead of manually using hdfs commands. When we create a script and utilize the different kinds of loop, we can repeat a action set number of times until desired process is complete. Otherwise had to do manually over. In this case we loop through the Filesystem which save us time.

Question Set P: Java API Program Execution

1. *Q: How does the program output differ from the output of simply running the command line code given below (where the path is changed to the same as in ApiJavaOne.java script)?*

   A: The Java API deliver information that would not be available through -ls command. Therefore, using java code is a tool is much more powerful, and allow for changes in the content of the files by running the jar file, this would not be possible with an -ls command.

Question Set Q: The ApiJavaTwo.java Program

1. *Q: What is the overall function of this program?*

   A: The function of the ApiJavaTwo program is almost identical to ApiJavaOne, the new program has an additional functionality. The new functionality allows us to check if the files is file formatted and will append a line or text to the file.

   Q: Besides the hdfsPath, what line(-s) (argument of handle) is most important for the user to change in this script to make it useful and extend to other files than the example?

   A: The script could be modified to become more dynamic. Currently it calls for a specific file, by making it more dynamic it could accept different files. We could also change it in a way so it appends the content of the code once more instead of a decided string.

Question Set R: Adapting the Program

1. *Q: Discuss the functionality of the program and given this functionality, what else you might change in your script to improve it.*

   A: The main functionality of the program is to find files in the specified hdfs path we have inputted in our java script and alter the files. If i could change something I would change the wat the files append data to files and which is more dynamic than hard coded.

   Q: How does the program output differ from the output of simply running the command line code given below (where the path and .txt file name is changed to the same as in your script)

   ```
   echo -e "some text" | hdfs dfs -appendToFile -
   insert/same/path/here/test.txt
   ```

   A: The code above does not provide an output; in the java program it shows that the length of the file is longer. We save us time of having to perform commands like -cat to ensure a change has happened.

Question Set S: Running WebHDFS

1. *Q: This part of the lab will be largely explorative.*

2. *Q: Show that you are able to execute at least 3 WebHDFS commands from the "webhdfs_api_manual.pdf" document in the lab folder.*

   A: I will provide the screenshots when executing the commands.

3. *Q: Then include screenshots of minimum 3 successful execution of WebHDFS (with generally HTTP 200 indicating success) in the lab report.*

A: Screenshots are on the next page.

Screenshot 1: first one is GetFileStatus .

```
[maria_dev@sandbox-hdp ~]$ curl -i "http://sandbox-hdp.hortonworks.com:50070/webhdfs/v1/user/maria_dev?op=GETFILESTATUS"
HTTP/1.1 200 OK
Date: Fri, 11 Mar 2022 15:41:07 GMT
Cache-Control: no-cache
Expires: Fri, 11 Mar 2022 15:41:07 GMT
Date: Fri, 11 Mar 2022 15:41:07 GMT
Pragma: no-cache
X-FRAME-OPTIONS: SAMEORIGIN
Content-Type: application/json
Transfer-Encoding: chunked

{"FileStatus":{"accessTime":0,"blockSize":0,"childrenNum":2,"fileId":17269,"group":"hdfs","length":0,"modificationTime":1646999524599,"owner":"maria_dev","pathSuffi
x":"","permission":"755","replication":0,"storagePolicy":0,"type":"DIRECTORY"}}[maria_dev@sandbox-hdp ~]$
```

Screenshot second one is ListFileStatus.

```
{"FileStatus":{"accessTime":0,"blockSize":0,"childrenNum":2,"fileId":17269,"group":"hdfs","length":0,"modificationTime":1646999524599,"owner":"maria_dev","pathSuffi
x":"","permission":"755","replication":0,"storagePolicy":0,"type":"DIRECTORY"}}[maria_dev@sandbox-hdp ~]$ curl -i "http://sandbox-hdp.hortonworks.com:50070/webhdfs/
v1/user/maria_dev?op=LISTSTATUS"
HTTP/1.1 200 OK
Date: Fri, 11 Mar 2022 15:43:35 GMT
Cache-Control: no-cache
Expires: Fri, 11 Mar 2022 15:43:35 GMT
Date: Fri, 11 Mar 2022 15:43:35 GMT
Pragma: no-cache
X-FRAME-OPTIONS: SAMEORIGIN
Content-Type: application/json
Transfer-Encoding: chunked

{"FileStatuses":{"FileStatus":[
{"accessTime":0,"blockSize":0,"childrenNum":1,"fileId":46048,"group":"hdfs","length":0,"modificationTime":1647000000078,"owner":"maria_dev","pathSuffix":".Trash","p
ermission":"700","replication":0,"storagePolicy":0,"type":"DIRECTORY"},
{"accessTime":0,"blockSize":0,"childrenNum":1,"fileId":51428,"group":"hdfs","length":0,"modificationTime":1646999710175,"owner":"maria_dev","pathSuffix":"texts","pe
rmission":"755","replication":0,"storagePolicy":0,"type":"DIRECTORY"}
]}}
```

Screenshot third is GetContentSummary .

```
[root@sandbox-hdp ~]# curl -i "http://sandbox-hdp.hortonworks.com:50070/webhdfs/v1/user/root?op=GETCONTENTSUMMARY"
HTTP/1.1 200 OK
Date: Fri, 11 Mar 2022 16:02:37 GMT
Cache-Control: no-cache
Expires: Fri, 11 Mar 2022 16:02:37 GMT
Date: Fri, 11 Mar 2022 16:02:37 GMT
Pragma: no-cache
X-FRAME-OPTIONS: SAMEORIGIN
Content-Type: application/json
Transfer-Encoding: chunked

{"ContentSummary":{"directoryCount":1,"fileCount":0,"length":0,"quota":-1,"spaceConsumed":0,"spaceQuota":-1,"typeQuota":{}}}[root@sandbox-hdp ~]#
```