

Question Set O: The ApiJavaOne.java Program

1. *What is the overall function of this program? (If not sure from reading the script, explore the program and its output using the instructions below first)*

This Java program connects to a Hadoop Distributed File System (HDFS) instance located at "hdfs://sandbox-hdp.hortonworks.com:8020/" and lists all files and directories recursively under the "/user/aria_dev" directory. It utilizes the Hadoop FileSystem API to interact with HDFS. The `main` method sets up the Hadoop configuration, establishes a connection to the HDFS instance, iterates over the files and directories under the specified path, and prints out their details. Finally, it closes the FileSystem object to release any resources acquired during the process.

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

a. Describe, in natural language, what this handle does.

The `listFiles()` method is a functionality provided by the Hadoop FileSystem API that allows you to retrieve an iterator over the files and directories present in a given directory path within the HDFS. It recursively lists all files and directories under the specified path, providing access to their metadata such as file size, modification time, owner, and permissions.

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

Setting the argument to `true` for the `listFiles()` method indicates that the listing should be done recursively. This means that not only the files and directories directly under the specified path will be listed, but also those nested within subdirectories.

If the argument were set to `false`, the listing would only include files and directories present directly under the specified path. Subdirectories and their contents would not be included in the listing, and the listing would not be recursive.

3. *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).*

Considering how using a while loop or a for loop in the script provides benefits over using command-line tools like `hdfs dfs`.

1. - Using a loop in the script allows for customized processing logic for each file or directory listed. For example, you could filter files based on certain criteria, perform

specific operations on them, or format the output in a particular way. This level of customization is not easily achievable with command-line tools alone.

2. - The script, being in Java, can easily integrate with other Java applications or systems. This integration allows for seamless incorporation of HDFS file listing functionality into larger Java-based projects.
3. - With a loop in the script, you have more control over error handling and exception management. You can implement try-catch blocks to handle exceptions gracefully, providing better feedback to users or taking appropriate actions in case of errors.
4. - Using a loop in the script enables programmatic control over the listing process. You can dynamically adjust parameters or conditions within the loop based on runtime information or user input, providing more flexibility and control compared to static command-line invocations.
5. - Loops in the script allow for complex iteration patterns that may not be easily achievable using command-line tools alone. For example, you could iterate over files in a non-linear order, skip certain files based on conditions, or implement parallel processing strategies.

Overall, utilizing loops in the script provides finer control, flexibility, and integration capabilities compared to relying solely on command-line tools like `'hdfs dfs'`.

Question Set P: Java API program execution

1. *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)?*

```
[maria_dev@sandbox-hdp java]$ hadoop jar ajo.jar ApiJavaOne
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output/_SUCCESS; isDirectory=false; length=0;
replication=1; blockSize=134217728; modification_time=1708195324989; access_time=1708195324984; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymlink=false;
hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output/part-r-00000; isDirectory=false; length=
17; replication=1; blockSize=134217728; modification_time=1708195324828; access_time=1708195324624; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymlink=f
alse; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1/_SUCCESS; isDirectory=false; length=0;
replication=1; blockSize=134217728; modification_time=1708198182761; access_time=1708198182755; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymlink=false
; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1/part-r-00000; isDirectory=false; lengt
h=98177; replication=1; blockSize=134217728; modification_time=1708198182618; access_time=1708198182285; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymli
nk=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708197256524/_SUCCESS; isDirectory=fals
e; length=0; replication=1; blockSize=134217728; modification_time=1708196253103; access_time=1708196253095; owner=maria_dev; group=hdfs; permission=rw-r--r--; isS
ymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708197256524/part-r-00000; isDirectory=
false; length=159910; replication=1; blockSize=134217728; modification_time=1708196252900; access_time=1708196252607; owner=maria_dev; group=hdfs; permission=rw-r-
r--; isSymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708198645749/_SUCCESS; isDirectory=fals
e; length=0; replication=1; blockSize=134217728; modification_time=1708197304396; access_time=1708197304391; owner=maria_dev; group=hdfs; permission=rw-r--r--; isS
ymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708198645749/part-r-00000; isDirectory=
false; length=174; replication=1; blockSize=134217728; modification_time=1708197304243; access_time=1708197303958; owner=maria_dev; group=hdfs; permission=rw-r-
r--; isSymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708199137224/_SUCCESS; isDirectory=fals
e; length=0; replication=1; blockSize=134217728; modification_time=1708199119576; access_time=1708199119572; owner=maria_dev; group=hdfs; permission=rw-r--r--; isS
ymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708199137224/part-r-00000; isDirectory=
false; length=174; replication=1; blockSize=134217728; modification_time=1708199119438; access_time=1708199119226; owner=maria_dev; group=hdfs; permission=rw-r-
r--; isSymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708199539165/_SUCCESS; isDirectory=fals
e; length=0; replication=1; blockSize=134217728; modification_time=1708199171062; access_time=1708199171045; owner=maria_dev; group=hdfs; permission=rw-r--r--; isS
ymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708199539165/part-r-00000; isDirectory=
false; length=174; replication=1; blockSize=134217728; modification_time=1708199170874; access_time=1708199170707; owner=maria_dev; group=hdfs; permission=rw-r-
r--; isSymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/Input; isDirectory=false; length=55; replication=1; blockSize=134217728; modifika
tion_time=1708187205243; access_time=1708187204395; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymlink=false; hasAcl=false; isEncrypted=false; isErasureC
oded=false}
```

```
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/Output/_SUCCESS; isDirectory=false; length=0; replication=1; blockSize=134217728;
modification_time=1708199635450; access_time=1708199635446; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymlink=false; hasAcl=false; isEncrypted=false; i
sErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/Output/part-r-00000; isDirectory=false; length=163; replication=1; blockSize=1342
17728; modification_time=1708199635248; access_time=1708199635052; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymlink=false; hasAcl=false; isEncrypted=fals
e; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/girl.txt; isDirectory=false; length=553337; replication=1; blockSize=134217728; m
odification_time=1708195975345; access_time=1708195975152; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymlink=false; hasAcl=false; isEncrypted=false; isE
rasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/input.txt; isDirectory=false; length=12; replication=1; blockSize=134217728; modi
fication_time=1708194020957; access_time=1708194020741; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymlink=false; hasAcl=false; isEncrypted=false; isEras
ureCoded=false}
```

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls
Found 6 items
drwx----- - maria_dev hdfs          0 2024-02-26 18:00 .Trash
drwx----- - maria_dev hdfs          0 2024-02-17 19:53 .staging
-rw-r--r--  1 maria_dev hdfs          55 2024-02-17 16:26 Input
drwxr-xr-x  - maria_dev hdfs          0 2024-02-17 19:53 Output
-rw-r--r--  1 maria_dev hdfs      553337 2024-02-17 18:52 girl.txt
-rw-r--r--  1 maria_dev hdfs         12 2024-02-17 18:20 input.txt
```

The program `ApiJavaOne.java` is designed to list the files and directories present in the HDFS directory `/user/maria_dev`. It does this by using the Hadoop FileSystem API to interact with HDFS and retrieve the file information.

On the other hand, the command `hdfs dfs -ls /insert/same/path/here` is a direct Hadoop command-line operation that lists the contents of the specified HDFS directory `/insert/same/path/here`.

The key differences between the two are:

- The Java program `ApiJavaOne.java` is executed as a Java program, which means it needs to be compiled and run as a Java application.

- The ``hdfs dfs -ls`` command is executed directly in the command line without needing to compile any code. It's a built-in Hadoop command for listing files in HDFS.
- The output of the Java program ``ApiJavaOne.java`` will be the list of files and directories present in the ``/user/maria_dev`` directory printed to the console.
- The output of the ``hdfs dfs -ls /insert/same/path/here`` command will be the list of files and directories present in the specified HDFS directory ``/insert/same/path/here``.
- The Java program allows for more flexibility and customization in handling the file listing operation. You can modify the program to include additional logic or filters as needed.
- The ``hdfs dfs -ls`` command is a predefined operation and provides limited flexibility compared to writing a custom Java program.

Question Set O: The ApiJavaTwo.java Program

1. *What is the overall function of this program? (If not sure from reading the script, explore the program and its output using the instructions below first)*

This program looks for a file named "test.txt" in a specific place on a big data system called Hadoop. If it finds "test.txt", it adds the words "Some text" to it.

2. *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?*

`file.getPath().getName().equalsIgnoreCase("test.txt")` is the most important line for the user to modify. This line specifies the name of the file ("test.txt") that the program checks for existence. If the user wants to extend the functionality to other files, they would need to modify this line to check for the existence of those files instead.

Question Set Q: Adapting the Program

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

```
[maria_dev@sandbox-hdp ~]$ hadoop jar ajo.jar ApiJavaTwo
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output/_SUCCESS; isDirectory=false; length=0;
replication=1; blockSize=134217728; modification_time=1708195324989; access_time=1708195324984; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymlink=false;
hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output/part-r-00000; isDirectory=false; length
=17; replication=1; blockSize=134217728; modification_time=1708195324828; access_time=1708195324624; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymlink=fals
e; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1/_SUCCESS; isDirectory=false; length=0;
replication=1; blockSize=134217728; modification_time=1708198182761; access_time=1708198182755; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymlink=false
; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1/part-r-00000; isDirectory=false; lengt
h=98177; replication=1; blockSize=134217728; modification_time=1708198182618; access_time=1708198182285; owner=maria_dev; group=hdfs; permission=rw-r--r--; isSymlin
k=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708197256524/_SUCCESS; isDirectory=fals
e; length=0; replication=1; blockSize=134217728; modification_time=1708196253103; access_time=1708196253095; owner=maria_dev; group=hdfs; permission=rw-r--r--; isS
ymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708197256524/part-r-00000; isDirectory=
false; length=159910; replication=1; blockSize=134217728; modification_time=1708196252900; access_time=1708196252607; owner=maria_dev; group=hdfs; permission=rw-r-
r--; isSymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708198645749/_SUCCESS; isDirectory=fals
e; length=0; replication=1; blockSize=134217728; modification_time=1708197304396; access_time=1708197304391; owner=maria_dev; group=hdfs; permission=rw-r--r--; isS
ymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708198645749/part-r-00000; isDirectory=
false; length=159910; replication=1; blockSize=134217728; modification_time=1708197304243; access_time=1708197303958; owner=maria_dev; group=hdfs; permission=rw-r-
r--; isSymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708199137224/_SUCCESS; isDirectory=fals
e; length=0; replication=1; blockSize=134217728; modification_time=1708199119576; access_time=1708199119572; owner=maria_dev; group=hdfs; permission=rw-r--r--; isS
ymlink=false; hasAcl=false; isEncrypted=false; isErasureCoded=false}
HdfsLocatedFileStatus{path=hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/240226180000/user/maria_dev/Output1708199137224/part-r-00000; isDirectory=
```

1. Implement error handling to catch and handle potential exceptions that may occur during filesystem operations (e.g., IOException). This ensures graceful handling of errors and prevents the program from crashing unexpectedly.
 2. Instead of just printing messages to the console, consider using a logging framework like Log4j to log informational, warning, and error messages to a log file. This provides better visibility into the application's behavior and facilitates debugging and monitoring.
 3. Depending on the size of the HDFS directory and the number of files it contains, iterating through all files in the directory might be inefficient. If possible, consider optimizing the search process, such as by using more specific file search criteria or parallelizing file processing.
 4. Instead of hardcoding the HDFS configuration parameters (e.g., default filesystem URL), consider externalizing these configurations to a properties file or using environment variables. This enhances flexibility and maintainability, allowing configurations to be easily modified without changing the code.
 5. Ensure proper resource management by closing filesystem handles (FileSystem) and output streams (FSDataOutputStream) in a finally block or using try-with-resources construct to prevent resource leaks.
 6. Validate input parameters such as HDFS path and file name to ensure they meet expected criteria and prevent potential security vulnerabilities or unexpected behavior.
2. *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
```



```
[maria_dev@sandbox-hdp ~]$ echo -e "linux-file-copy" | hdfs dfs -appendToFile - /linux-file-copy.txt
appendToFile: Permission denied: user=maria_dev, access=WRITE, inode="/":hdfs:hdfs:drwxr-xr-x
[maria_dev@sandbox-hdp ~]$ echo -e "linux-file-copy" | hdfs dfs -appendToFile - ./linux-file-copy.txt
[maria_dev@sandbox-hdp ~]$ hdfs dfs -cat ./linux-file-copy.txt
linux-file-copy
[maria_dev@sandbox-hdp ~]$ hadoop jar ajo.jar ApiJavaTwo2
[maria_dev@sandbox-hdp ~]$ █
```

The output of the provided Java program and the command line code you provided may differ in several aspects:

1. The error message "Permission denied" suggests that the user `maria_dev` does not have the necessary permissions to write to the specified HDFS directory or file. This could be due to restrictive HDFS permissions or ACLs (Access Control Lists) that prevent the user from writing to the target location. In contrast, the Java program might execute with different permissions depending on how it's run (e.g., as a different user or with elevated privileges).

2. In the Java program, the file path is constructed programmatically using the HDFS path and file name variables (`hdfsPath` and `fileName`). This ensures flexibility and allows dynamic generation of file paths based on runtime conditions. In the command line code, the file path is explicitly specified as `/linux-file-copy.txt` or `./linux-file-copy.txt`, which may not match the path used in the Java program (`/user/maria_dev/linux-file-copy`).

3. The content appended to the file also differs. In the command line code, the content appended is "linux-file-copy", while in the Java program, the content appended is "Some text\n". This is because the Java program explicitly writes "Some text\n" to the file, whereas the command line code echoes "linux-file-copy".

4. The Java program iterates through files in the specified directory and appends content to a file if it exists and matches the specified name. In contrast, the command line code directly appends content to the specified file without checking if it exists or matching its name. Therefore, if the file does not exist, the command line code may create a new file with the specified name and append content to it.

We can say that, while both the Java program and the command line code aim to append content to a file on HDFS, they differ in terms of permissions, file path handling, content appended, and handling of existing files. The differences in behavior may lead to variations in output and outcomes.

Question Set Q: Running WebHDFS

1. This part of the lab will be largely explorative.

```
[maria_dev@sandbox-hdp ~]$ dbox-hdp.hortonworks.com:50070/webhdfs/v1/tmp?op=GETFILESTATUS"
HTTP/1.1 200 OK
Date: Mon, 26 Feb 2024 22:28:15 GMT
Cache-Control: no-cache
Expires: Mon, 26 Feb 2024 22:28:15 GMT
Date: Mon, 26 Feb 2024 22:28:15 GMT
Pragma: no-cache
X-FRAME-OPTIONS: SAMEORIGIN
Content-Type: application/json
Transfer-Encoding: chunked

{"FileStatus":{"accessTime":0,"blockSize":0,"childrenNum":5,"fileId":16386,"group":"hdfs","length":0,"modificationTime":1708176089386,"owner":"hdfs","pathSuffix":"",
"permission":"777","replication":0,"storagePolicy":0,"type":"DIRECTORY"}}[maria_dev@sandbox-hdp ~]$
```

2. Show that you are able to execute at least 3 WebHDFS commands from the "webhdfs_api_manual.pdf" document in the lab folder.
3. Then include screenshots of minimum 3 successful execution of WebHDFS (with generally HTTP 200 indicating success) in the lab report.

List the contents of the /tmp directory:

```
[maria_dev@sandbox-hdp ~]$ curl -i -L "http://sandbox-hdp.hortonworks.com:50070/webhdfs/v1/tmp?op=LISTSTATUS"
HTTP/1.1 200 OK
Date: Tue, 27 Feb 2024 18:42:54 GMT
Cache-Control: no-cache
Expires: Tue, 27 Feb 2024 18:42:54 GMT
Date: Tue, 27 Feb 2024 18:42:54 GMT
Pragma: no-cache
X-FRAME-OPTIONS: SAMEORIGIN
Content-Type: application/json
Transfer-Encoding: chunked

{"FileStatuses":{"FileStatus":[{"accessTime":0,"blockSize":0,"childrenNum":0,"fileId":18752,"group":"hadoop","length":0,"modificationTime":1543518115118,"owner":"druid","pathSuffix":"druid-indexing",
"permission":"775","replication":0,"storagePolicy":0,"type":"DIRECTORY"}, {"accessTime":0,"blockSize":0,"childrenNum":1,"fileId":16393,"group":"hdfs","length":0,"modificationTime":1543512359901,"owner":"hdfs","pathSuffix":"entity-file-history",
"permission":"755","replication":0,"storagePolicy":0,"type":"DIRECTORY"}, {"accessTime":0,"blockSize":0,"childrenNum":3,"fileId":17488,"group":"hdfs","length":0,"modificationTime":1543514294701,"owner":"hive","pathSuffix":"hive",
"permission":"733","replication":0,"storagePolicy":0,"type":"DIRECTORY"}, {"accessTime":0,"blockSize":0,"childrenNum":2,"fileId":22964,"group":"hdfs","length":0,"modificationTime":1708176965323,"owner":"maria_dev",
"pathSuffix":"maria_dev-newfolder",
"permission":"755","replication":0,"storagePolicy":0,"type":"DIRECTORY"}, {"accessTime":0,"blockSize":0,"childrenNum":0,"fileId":22958,"group":"hdfs","length":0,"modificationTime":1707909090293,"owner":"root",
"pathSuffix":"root-newfolder",
"permission":"755","replication":0,"storagePolicy":0,"type":"DIRECTORY"}]}}
[maria_dev@sandbox-hdp ~]$
```

Listing the contents of a specific directory within /tmp, by providing the full path to that directory (/tmp/maria_dev-newfolder)

```
[maria_dev@sandbox-hdp ~]$ curl -i -L "http://sandbox-hdp.hortonworks.com:50070/webhdfs/v1/tmp/maria_dev-newfolder?op=LISTSTATUS"
HTTP/1.1 200 OK
Date: Tue, 27 Feb 2024 18:44:59 GMT
Cache-Control: no-cache
Expires: Tue, 27 Feb 2024 18:44:59 GMT
Date: Tue, 27 Feb 2024 18:44:59 GMT
Pragma: no-cache
X-FRAME-OPTIONS: SAMEORIGIN
Content-Type: application/json
Transfer-Encoding: chunked

{"FileStatuses":{"FileStatus":[{"accessTime":1708176965026,"blockSize":134217728,"childrenNum":0,"fileId":22966,"group":"hdfs","length":32,"modificationTime":1708176965312,"owner":"maria_dev",
"pathSuffix":"linux-file-copy.txt",
"permission":"644","replication":1,"storagePolicy":0,"type":"FILE"}, {"accessTime":1708176343950,"blockSize":134217728,"childrenNum":0,"fileId":22965,"group":"hdfs","length":32,"modificationTime":1708176344693,"owner":"maria_dev",
"pathSuffix":"linux-file.txt",
"permission":"644","replication":1,"storagePolicy":0,"type":"FILE"}]}}
[maria_dev@sandbox-hdp ~]$
```

A command to open a file named linux-file.txt within the maria_dev-newfolder directory:

```
[maria_dev@sandbox-hdp ~]$ curl -i -L "http://sandbox-hdp.hortonworks.com:50070/webhdfs/v1/tmp/maria_dev-newfolder/linux-file.txt?op=OPEN"
HTTP/1.1 307 Temporary Redirect
Date: Tue, 27 Feb 2024 19:02:58 GMT
Cache-Control: no-cache
Expires: Tue, 27 Feb 2024 19:02:58 GMT
Date: Tue, 27 Feb 2024 19:02:58 GMT
Pragma: no-cache
X-FRAME-OPTIONS: SAMEORIGIN
Location: http://sandbox-hdp.hortonworks.com:50075/webhdfs/v1/tmp/maria_dev-newfolder/linux-file.txt?op=OPEN&namenoderpcaddress=sandbox-hdp.hortonworks.com:8020&offset=0
Content-Type: application/octet-stream
Content-Length: 0

HTTP/1.1 200 OK
Access-Control-Allow-Methods: GET
Access-Control-Allow-Origin: *
Content-Type: application/octet-stream
Connection: close
Content-Length: 32

This file was created in Linux.
[maria_dev@sandbox-hdp ~]$ █
```

A command to read a file named linux-file.txt within the maria_dev-newfolder directory:

```
[maria_dev@sandbox-hdp ~]$ curl -v -s -X GET "http://sandbox-hdp.hortonworks.com:50075/webhdfs/v1/tmp/maria_dev-newfolder/linux-file.txt?op=OPEN&namenoderpcaddress=sandbox-hdp.hortonworks.com:8020&offset=0" > output.txt
* About to connect() to sandbox-hdp.hortonworks.com port 50075 (#0)
*   Trying 172.18.0.2...
* Connected to sandbox-hdp.hortonworks.com (172.18.0.2) port 50075 (#0)
> GET /webhdfs/v1/tmp/maria_dev-newfolder/linux-file.txt?op=OPEN&namenoderpcaddress=sandbox-hdp.hortonworks.com:8020&offset=0 HTTP/1.1
> User-Agent: curl/7.29.0
> Host: sandbox-hdp.hortonworks.com:50075
> Accept: */*
>
< HTTP/1.1 200 OK
< Access-Control-Allow-Methods: GET
< Access-Control-Allow-Origin: *
< Content-Type: application/octet-stream
< Connection: close
< Content-Length: 32
<
[data not shown]
* Closing connection 0
[maria_dev@sandbox-hdp ~]$ curl -X GET "http://sandbox-hdp.hortonworks.com:50075/webhdfs/v1/tmp/maria_dev-newfolder/linux-file.txt?op=OPEN&namenoderpcaddress=sandbox-hdp.hortonworks.com:8020&offset=0"
This file was created in Linux.
[maria_dev@sandbox-hdp ~]$ █
```
