HDFS experiment

1. Download, install and setup Hadoop- completed as part of previous assignment.
2. Format datanode and namenode, start dfs and yarn.
3. Verify everything is running properly for this experiment:

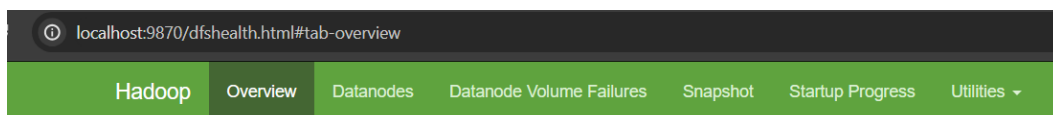- Start hdfs and yarn using command as shown below:

```
C:\hadoop\sbin>start-dfs.cmd
```

```
C:\hadoop\sbin>start-yarn
starting yarn daemons
```

- Verify using jps command:

```
C:\hadoop\sbin>jps
16064 ResourceManager
27584 SpringToolSuite4
32096 BootLanguageServerBootApp
11140 NodeManager
37188 DataNode
41316 NameNode
49764 Jps
```

- Verify in browser by visiting http://localhost:9870/

localhost:9870/dfshealth.html#tab-overview

| Hadoop | Overview | Datanodes | Datanode Volume Failures | Snapshot | Startup Progress | Utilities ▾ |

Overview 'localhost:9000' (✔active)

Started:	Sun Mar 24 15:25:41 -0400 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012bf9c
Compiled:	Sun Jun 18 04:22:00 -0400 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-aa42c928-1993-410d-ad66-1b94e81f3d58
Block Pool ID:	BP-333416306-192.168.1.74-1711308221175

- Browse the file system under utilities:

4. Create a new directory /user/your_user_name/books on HDFS:

- Navigate to bin directory and run command to create a home directory first then the subsequent directories as shown:

```
C:\hadoop\bin>hdfs dfs -mkdir /user

C:\hadoop\bin>hdfs dfs -mkdir /user/spatidar

C:\hadoop\bin>hdfs dfs -mkdir /user/spatidar/books
```

- Verify the directories from browser:

		Permission		Owner		Group		Size		Last Modified		Replication		Block Size		Name		
☐		drwxr-xr-x		charu		supergroup		0 B		Mar 24 16:22		0		0 B		user		🗑

		Permission		Owner		Group		Size		Last Modified		Replication		Block Size		Name		
☐		drwxr-xr-x		charu		supergroup		0 B		Mar 24 16:23		0		0 B		spatidar		🗑

		Permission		Owner		Group		Size		Last Modified		Replication		Block Size		Name		
☐		drwxr-xr-x		charu		supergroup		0 B		Mar 24 16:23		0		0 B		books		🗑

5. Create a file, name it harry_potter.txt, on your local filesystem, write something to it, and then upload it to HDFS under /user/your_user_name/books

   - Create a file as above and copy the file to HDFS as shown:

   ```
   C:\hadoop\bin>hdfs dfs -copyFromLocal C:\hadoop\temp\harry_potter.txt /user/spatidar/books/
   ```

   - Verify from browser:

/user/spatidar/books								Go!

Show 25 entries   Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
☐	-rw-r--r--	charu	supergroup	13 B	Mar 24 16:43	1	128 MB	harry_potter.txt	🗑

6. Print the content of /user/your_user_name/books/harry_potter.txt: use the -cat to print content of file as shown

   ```
   C:\hadoop\bin>hdfs dfs -cat /user/spatidar/books/harry_potter.txt
   Expelliarmus!
   C:\hadoop\bin>
   ```

7. Print the size of /user/your_user_name/books/harry_potter.txt: use -du to print the size of a file. Below file size is 13 Bytes

   ```
   C:\hadoop\bin>hdfs dfs -du /user/spatidar/books/harry_potter.txt
   13   13  /user/spatidar/books/harry_potter.txt
   ```

8. Print the last modified time of file: use -stat to print the last modified time of file

   ```
   C:\hadoop\bin>hdfs dfs -stat /user/spatidar/books/harry_potter.txt
   2024-03-24 20:43:08
   ```

9. **Copy within HDFS:** Copy within HDFS using -cp and verify in browser as shown

```
C:\hadoop\bin>hdfs dfs -cp /user/spatidar/books/harry_potter.txt /user/spatidar/books/harry_potter.txt_copied
```

/user/spatidar/books   Go!

Show 25 entries   Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
☐	-rw-r--r--	charu	supergroup	13 B	Mar 24 16:43	1	128 MB	harry_potter.txt	🗑
☐	-rw-r--r--	charu	supergroup	13 B	Mar 24 17:06	1	128 MB	harry_potter.txt_copied	🗑

10. **Download from HDFS to your local file system:** download file using -get or -copyToLocal:

```
C:\hadoop\bin>hdfs dfs -copyToLocal /user/spatidar/books/harry_potter.txt_copied  C:\hadoop\temp\
```

This PC > Windows (C:) > hadoop > temp

Sort   View   •••

Name	Date modified	Type	Size
harry_potter	24-03-2024 16:41	Text Document	1 KB
harry_potter.txt_copied	24-03-2024 17:16	TXT_COPIED File	1 KB