

PART 3 - MongoDB indexing

Most of the time, you'll want to declare your indexes before putting your application into production. This allows indexes to be built incrementally, as the data is inserted. But there are two cases where you might choose to build an index after the fact. The first case occurs when you need to import a lot of data before switching into production. For instance, you might be migrating an application to MongoDB and need to seed the database with user information from a data warehouse. You could create the indexes on your user data in advance, but doing so after you have imported the data will ensure an ideally balanced and compacted index from the start. This will also minimize the net time to build the index. Use the NYSE dataset to declare your indexes before putting your application into production.

1. Created new db => use a03;
2. Created new collection => db.createCollection("nyse")
3. Create index on date => db.nyse.createIndex({"date": 1})

```
..oop-2.7.3/bin

[a03> db.createCollection("nyse")
{ ok: 1 }
[a03> db.nyse.find()

a03> db.nyse.createIndex({"date": 1})
Uncaught:
SyntaxError: Unexpected character '''. (1:21)

> 1 | db.nyse.createIndex({"date": 1})
   | ^

a03> db.nyse.createIndex({"date": 1})
date_1
[a03> db.nyse.find()

[a03> db.nyse.getIndexes()
[
  { v: 2, key: { _id: 1 }, name: '_id_' },
  { v: 2, key: { date: 1 }, name: 'date_1' }
]
a03> ]
```

4. Imported nyse data using bat file

```

#!/bin/bash

for dataFileSuffix in {A..Z};
do

FILELOC="/Users/dikshabhatia/Desktop/info/a3/NYSE/NYSE_daily_prices_${dataFileSuffix}.csv"
echo $FILELOC
mongoimport --db=a03 --collection=nyse --type=csv
--file=$FILELOC --headerline

```

Done

5. Executed the script

The screenshot shows a terminal window with the command `/oop-2.7.3/bin/dikshabhatia:/Desktop/info/a3/NYSE/NYSE_daily_prices_A.csv` running. The output shows the progress of importing CSV files into MongoDB. A battery icon in the top right corner indicates a low battery level.

```

./oop-2.7.3/bin
...mebrew/bin:/opt/homebrew/sbin:/usr/local/bin:/bin:/sbin:/usr/sbin:/sbin
...oop-2.7.3/bin/dikshabhatia:/Desktop/info/a3/NYSE/NYSE_daily_prices_A.csv
2022-06-29T23:37:31.776-0700 [##,...] homework2.nsy Stock Data 4.84MB/39.1MB (10.3%)
2022-06-29T23:37:34.776-0700 [####,...] homework2.nsy Stock Data 8.39MB/39.1MB (21.5%)
2022-06-29T23:37:37.776-0700 [#####,...] homework2.nsy Stock Data 12.8MB/39.1MB (32.6%)
2022-06-29T23:37:40.776-0700 [#####,...] homework2.nsy Stock Data 17.2MB/39.1MB (44.1%)
2022-06-29T23:37:43.776-0700 [#####,...] homework2.nsy Stock Data 21.6MB/39.1MB (55.2%)
2022-06-29T23:37:46.776-0700 [#####,...] homework2.nsy Stock Data 26.1MB/39.1MB (66.3%)
2022-06-29T23:37:49.776-0700 [#####,...] homework2.nsy Stock Data 30.1MB/39.1MB (76.9%)
2022-06-29T23:37:52.776-0700 [#####,...] homework2.nsy Stock Data 34.3MB/39.1MB (87.6%)
2022-06-29T23:37:55.776-0700 [#####,...] homework2.nsy Stock Data 38.5MB/39.1MB (98.6%)
2022-06-29T23:37:56.177-0700 [#####,...] homework2.nsy Stock Data 39.1MB/39.1MB (100.0%)
2022-06-29T23:37:56.177-0700 735026 documents imported successfully. 0 document(s) failed to import.
/Users/dikshabhatia/Desktop/info/a3/NYSE/NYSE_daily_prices_B.csv
2022-06-29T23:37:59.282-0700 [##,...] homework2.nsy Stock Data 3.89MB/38.6MB (3.8%)
2022-06-29T23:38:02.281-0700 [####,...] homework2.nsy Stock Data 8.85MB/38.6MB (26.3%)
2022-06-29T23:38:08.280-0700 [#####,...] homework2.nsy Stock Data 12.4MB/38.6MB (40.5%)
2022-06-29T23:38:11.280-0700 [#####,...] homework2.nsy Stock Data 16.7MB/38.6MB (54.6%)
2022-06-29T23:38:11.280-0700 [#####,...] homework2.nsy Stock Data 20.8MB/38.6MB (68.2%)
2022-06-29T23:38:14.280-0700 [#####,...] homework2.nsy Stock Data 25.1MB/38.6MB (82.0%)
2022-06-29T23:38:17.280-0700 [#####,...] homework2.nsy Stock Data 29.4MB/38.6MB (96.3%)
2022-06-29T23:38:18.142-0700 [#####,...] homework2.nsy Stock Data 36.6MB/38.6MB (100.0%)
2022-06-29T23:38:18.142-0700 577983 documents imported successfully. 0 document(s) failed to import.
/Users/dikshabhatia/Desktop/info/a3/NYSE/NYSE_daily_prices_C.csv
2022-06-29T23:38:18.237-0700 connected to: mongodb://localhost/
2022-06-29T23:38:21.238-0700 [##,...] homework2.nsy Stock Data 4.28MB/43.7MB (9.8%)
2022-06-29T23:38:24.237-0700 [####,...] homework2.nsy Stock Data 8.61MB/43.7MB (19.7%)
2022-06-29T23:38:27.237-0700 [#####,...] homework2.nsy Stock Data 12.9MB/43.7MB (29.5%)
2022-06-29T23:38:30.237-0700 [#####,...] homework2.nsy Stock Data 17.2MB/43.7MB (39.3%)
2022-06-29T23:38:33.237-0700 [#####,...] homework2.nsy Stock Data 21.4MB/43.7MB (49.8%)
2022-06-29T23:38:36.236-0700 [#####,...] homework2.nsy Stock Data 25.6MB/43.7MB (58.6%)
2022-06-29T23:38:42.236-0700 [#####,...] homework2.nsy Stock Data 29.9MB/43.7MB (68.6%)
2022-06-29T23:38:42.236-0700 [#####,...] homework2.nsy Stock Data 34.4MB/43.7MB (78.8%)
2022-06-29T23:38:45.236-0700 [#####,...] homework2.nsy Stock Data 38.8MB/43.7MB (88.7%)
2022-06-29T23:38:48.236-0700 [#####,...] homework2.nsy Stock Data 43.0MB/43.7MB (98.5%)
2022-06-29T23:38:48.732-0700 [#####,...] homework2.nsy Stock Data 43.7MB/43.7MB (100.0%)
2022-06-29T23:38:48.732-0700 825935 documents imported successfully. 0 document(s) failed to import.
/Users/dikshabhatia/Desktop/info/a3/NYSE/NYSE_daily_prices_D.csv
2022-06-29T23:38:51.832-0700 connected to: mongodb://localhost/
2022-06-29T23:38:51.834-0700 [##,...] homework2.nsy Stock Data 4.43MB/18.3MB (24.1%)
2022-06-29T23:38:54.832-0700 [#####,...] homework2.nsy Stock Data 8.82MB/18.3MB (48.1%)
2022-06-29T23:38:57.832-0700 [#####,...] homework2.nsy Stock Data 13.2MB/18.3MB (71.1%)
2022-06-29T23:39:01.832-0700 [#####,...] homework2.nsy Stock Data 17.4MB/18.3MB (95.1%)
2022-06-29T23:39:01.832-0700 [#####,...] homework2.nsy Stock Data 18.3MB/18.3MB (100.0%)
2022-06-29T23:39:01.842-0700 345846 documents imported successfully. 0 document(s) failed to import.
/Users/dikshabhatia/Desktop/info/a3/NYSE/NYSE_daily_prices_E.csv
2022-06-29T23:39:01.855-0700 connected to: mongodb://localhost/
2022-06-29T23:39:08.566-0700 [##,...] homework2.nsy Stock Data 4.38MB/21.1MB (20.8%)
2022-06-29T23:39:08.569-0700 [####,...] homework2.nsy Stock Data 8.73MB/21.1MB (41.4%)
2022-06-29T23:39:08.569-0700 [#####,...] homework2.nsy Stock Data 12.9MB/21.1MB (61.3%)
2022-06-29T23:39:13.566-0700 [#####,...] homework2.nsy Stock Data 17.2MB/21.1MB (81.7%)
2022-06-29T23:39:16.272-0700 [#####,...] homework2.nsy Stock Data 21.1MB/21.1MB (100.0%)
2022-06-29T23:39:16.272-0700 396733 documents imported successfully. 0 document(s) failed to import.
/Users/dikshabhatia/Desktop/info/a3/NYSE/NYSE_daily_prices_F.csv
2022-06-29T23:39:16.368-0700 connected to: mongodb://localhost/
2022-06-29T23:39:19.369-0700 [##,...] homework2.nsy Stock Data 4.24MB/16.6MB (25.6%)
2022-06-29T23:39:22.368-0700 [#####,...] homework2.nsy Stock Data 8.48MB/16.6MB (51.1%)

```

6. Result of current indexes: db.nyse.getIndexes()

PART 4 - MongoDB Indexing

Insert the NYSE dataset into a new database. You may use the existing NYSE database created before. Now, create indexes on existing data sets.

Creating Index on stock volume: `db.nyse.createIndex({"stock_volume": -1})`

```
..oop-2.7.3/bin ..mebrew/bin:/opt/homebrew

a03> db.nyse.createIndex({"stock_volume": -1})
stock_volume_-1
a03>

a03>

[a03> db.nyse.getIndexes()
[
  { v: 2, key: { _id: 1 }, name: '_id_' },
  { v: 2, key: { date: 1 }, name: 'date_1' },
  { v: 2, key: { stock_volume: -1 }, name: 'stock_volume_-1' }
]
a03> ]
```

PART 5 - MongoDB Text search (refer to chapter 9).

Write and execute one query to perform each of the followings on any collection of your choice.

Created stores collection for performing this part

```
db.stores.insertMany(
[
  { _id: 1, name: "Java Hut", description: "Coffee and cakes" },
  { _id: 2, name: "Burger Buns", description: "Gourmet hamburgers" },
  { _id: 3, name: "Coffee Shop", description: "Just coffee" },
  { _id: 4, name: "Clothes Clothes Clothes", description: "Discount clothing" },
  { _id: 5, name: "Java Shopping", description: "Indonesian goods" }
]
```

1. Specify and word matches instead of or word matches.

```
db.stores.find({$text: {$search: "Clothes -Java"}}, { _id: 0, name: 1, description: 1 })
```

```
..oop-2.7.3/bin ...mebrew/bin:/opt/homebrew/s

[a03> db.stores.find({$text: {$search: "Coffee"}}, {_id: 0, name: 1, description: 1 })
MongoServerError: text index required for $text query
[a03> db.stores.createIndex({name: "text", description: "text"})
name_text_description_text
[a03> db.stores.find({$text: {$search: "Coffee"}}, {_id: 0, name: 1, description: 1 })
[ { name: 'Coffee Shop', description: 'Just coffee' } ]
a03> ]
```

2. Perform exact phrase matches.

```
db.stores.find({$text: {$search: "\"Coffee Shop\""} })
```

```
[a03> db.stores.find({$text: {$search: "\"Coffee Shop\""} })
[ { _id: 3, name: 'Coffee Shop', description: 'Just coffee' } ]
a03> ]
```

3. Exclude documents with certain words.

```
db.stores.find({$text: {$search: "Clothes -Java"}}, {_id: 0, name: 1, description: 1 })
```

```
a03> db.stores.find({$text: {$search: "Clothes -Java"}}, {_id: 0, name: 1, description: 1 })
[
  { name: 'Clothes Clothes Clothes', description: 'Discount clothing' }
]
[a03> db.stores.find({$text: {$search: " -Java"}}, {_id: 0, name: 1, description: 1 })
```

4. Exclude documents with certain phrases.

```
db.stores.find({$text: {$search: "Coffee -\"Java Shopping\""}}, {_id: 0, name: 1, description: 1 })
```

```
2 |  
a03> db.stores.find({$text: {$search: "Coffee -\"Java Shopping\""}}, {_id: 0, name: 1, description: 1 })  
[ { name: 'Coffee Shop', description: 'Just coffee' } ]  
a03> █
```

PART 6. Programming Assignment

All hadoop commands are invoked by the bin/hadoop script. Running the hadoop script without any arguments prints the description for all commands.

Usage: hadoop [--config confdir] [--loglevel loglevel] [COMMAND] [GENERIC_OPTIONS] [COMMAND_OPTIONS] Execute each hadoop command once, and place the screenshots into a word file. If a command cannot be executed for any reason (such as, a distributed environment is needed), you may write the definition of the command, and skip execution.

<http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/FileSystemShell.html>

=> Commands:

Start Hadoop

- cd /usr/local/bin/hadoop-2.7.3/sbin
- ./start-all.sh
- cd /usr/local/bin/hadoop-2.7.3/bin

- **appendToFile:** Append single src, or multiple srcs from local file system to the destination file system. Also reads input from stdin and appends to destination file system.
- **./hadoop fs -appendToFile /Users/dikshabhatia/Desktop/info/a3/example.docx /testdir/test.docx**

```

Tear 22/06/28 19:00:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $ ./hadoop fs -appendToFile /Users/dikshabhatia/Desktop/info/a3/example.docx /Users/dikshabhatia/Desktop/info/a3/test3.docx [23:01:04]
dikshabhatia:bin/ $ ls
container-executor      hadoop.cmd          hdfs.cmd          mapred.cmd        test-container-executor yarn.cmd
hadoop                  hdfs               mapred            rcc              yarn
dikshabhatia:bin/ $ ./hadoop fs -appendToFile /Users/dikshabhatia/Desktop/info/a3/example.docx /testdir/test3.docx [23:10:46]
dikshabhatia:bin/ $ ./hadoop fs -appendToFile /Users/dikshabhatia/Desktop/info/a3/example.docx /testdir/test3.docx [23:10:46]
22/06/28 23:06:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

```

localhost:50070/explorer.html#/testdir

Bookmarks Coding Akash Gre home Java iCloud Technologies JEST New Tab interview Direct quant News Other Bookmarks

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/testdir								Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
-rw-r--r--	dikshabhatia	supergroup	3.45 KB	6/28/2022, 11:14:03 PM	1	128 MB	test3.docx	

Hadoop, 2016.

- **Command Cat:** Copies source paths to stdout.

- Creating a text file: `./hadoop fs -touchz /testdir/sample.txt`
- Concatenating local file into file on hadoop: `./hadoop fs -cat file:///Users/dikshabhatia/Desktop/info/a3/example.docx /testdir/sample.txt`

```

bin - dikshabhatia@Dikshas-MacBook-Pro ..oop-2.7.3/bin - zsh - 161x24
-cat: Wrong FS: file:///Users/dikshabhatia/Desktop/info/a3/example.docx, expected: file:///
Usage: hadoop fs [-generic options] -cat [-ignoreCrc] <src> ...
dikshabhatia:bin/ $ ./hadoop fs -cat file:///Users/dikshabhatia/Desktop/info/a3/example.docx /testdir/sample.txt [23:35:5
22/06/28 23:37:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
P??T?K?"[Content_Types].xml"?N?E?|??m?8e?J??tQ>??'???'!?[g??_P.(?n"9s=??????"*?
??k+k+?y??$???wP# _?????V?K)?+1U?Lt?Rme
?(?.?KY2x]>A#W)b??=5?G?'??
7J&?????(_??s`g.H??,a??
8??h2?h`k??>j??Yr?cf??Mc??CZ?^"?????b?q??H4q??_c.!I?> m0??q_+]:@L?OD??~01_?a??_P??T??(_M
_rels/.rels??J1
??>E?}'?+?v?"??Dmf?
??Vpt`]?i??K?vw?z\\
?M
??oQ?s??U??9p??j??3I?)?KEUH(&t?X?6]??3??I??e?hi?_?jo5?_?_
?G?9i?"_R?Y\F??
x?hs??Ng!Kbh?>?????e?({?W??W?)P??T??x??word/_rels/document.xml.rels??MK1
?G?9i?"_R?Y\F??
W?!"[?K?@Ru0??;?g?T?DN-?D?6?&x??l0U)3?W?jW??e??_?^8??XU?K?Z??D;ao"+x
??B??mP?_.?t?????k?;?P??T??Q?aword/document.xml?S?N?0
??U?#?4??"$@\????Rw?
?[?6-??.?1??u?O??<)?%[^?,+_?VvW?????
?Jvb??/j?{d?R?d?+b'gt?L6??w??
?l?mQ,?? ?o?2?4J'?????Y?=??V?o?=?;??
#[?I?{???(?((?)yd\?g|?o?8??o?_?*??o?qy?Lo{ _e????B?T?dO?]
```

Browse Directory

/testdir								Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
-rw-r--r--	dikshabhatia	supergroup	0 B	6/28/2022, 11:34:16 PM	1	128 MB	sample.txt	
-rw-r--r--	dikshabhatia	supergroup	3.45 KB	6/28/2022, 11:14:03 PM	1	128 MB	test3.docx	

Hadoop, 2016.

- Checksum: Returns the checksum information of a file.

Command: **./hadoop fs -checksum /testdir/sample.txt**

```
docProps/meta.xml?K
dikshabhatia:bin/ $ ./hadoop fs -checksum /testdir/sample.txt [23:37:19]
zsh: no such file or directory: /hadoop
dikshabhatia:bin/ $ ./hadoop fs -checksum /testdir/sample.txt [23:43:42]
22/06/28 23:43:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
/testdir/sample.txt      MD5-of-0MD5-of-0CRC32  000000000000000000000000000000007bc8f4b72a86921468bf8e8441dce51
dikshabhatia:bin/ $ [23:43:50]
```

- chgrp

Change group association of files. The user must be the owner of files, or else a super-user.

Listing files: **./hadoop fs -ls /testdir/**

Changing owner: **./hadoop fs -chgrp newusr /testdir/sample.txt**

```
cd: Unknown command
dikshabhatia:bin/ $ ./hadoop fs ls /testdir/
ls: Unknown command
Did you mean -ls? This command begins with a dash.
dikshabhatia:bin/ $ ./hadoop fs -ls /testdir/ [23:48:53]
22/06/28 23:49:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 dikshabhatia supergroup 0 2022-06-28 23:34 /testdir/sample.txt
-rw-r--r-- 1 dikshabhatia supergroup 3530 2022-06-28 23:14 /testdir/test3.docx
dikshabhatia:bin/ $ ./hadoop fs -chgrp newusr /testdir/sample.txt [23:49:09]
22/06/28 23:50:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $ ./hadoop fs -ls /testdir/ [23:49:17]
22/06/28 23:50:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 dikshabhatia newusr 0 2022-06-28 23:34 /testdir/sample.txt
-rw-r--r-- 1 dikshabhatia supergroup 3530 2022-06-28 23:14 /testdir/test3.docx
dikshabhatia:bin/ $ [23:50:16]
```

- chmod

Change the permissions of files.

Command:

Listing files: **./hadoop fs -ls /testdir/**

Changing owner: **./hadoop fs -chmod 777 /testdir/sample.txt**

```

22/06/28 23:50:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $ ./hadoop fs -ls /testdir/
22/06/28 23:50:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 dikshabhatia newusr          0 2022-06-28 23:34 /testdir/sample.txt
-rw-r--r-- 1 dikshabhatia supergroup     3530 2022-06-28 23:14 /testdir/test3.docx
dikshabhatia:bin/ $ ls -l /testdir/sample.txt
22/06/28 23:54:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $ ./hadoop fs -ls /testdir/
22/06/28 23:54:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rwxrwxrwx 1 dikshabhatia newusr          0 2022-06-28 23:34 /testdir/sample.txt
-rw-r--r-- 1 dikshabhatia supergroup     3530 2022-06-28 23:14 /testdir/test3.docx
dikshabhatia:bin/ $ 

```

- **copyFromLocal**
- Identical to the **-put** command

Command: **\$./hadoop fs -copyFromLocal
file:///Users/dikshabhatia/Desktop/info/a3/example6.pages /testdir/**

Hadoop Overview Datanodes Snapshot Startup Progress Utilities ▾

Browse Directory

/testdir							
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	dikshabhatia	supergroup	612.26 KB	6/29/2022, 12:00:51 AM	1	128 MB	example6.pages
-rwxrwxrwx	dikshabhatia	newusr	0 B	6/28/2022, 11:34:16 PM	1	128 MB	sample.txt
-rw-r--r--	dikshabhatia	supergroup	3.45 KB	6/28/2022, 11:14:03 PM	1	128 MB	test3.docx

Hadoop, 2016.

- **copyToLocal**
- Identical to the **-get** command

Command:

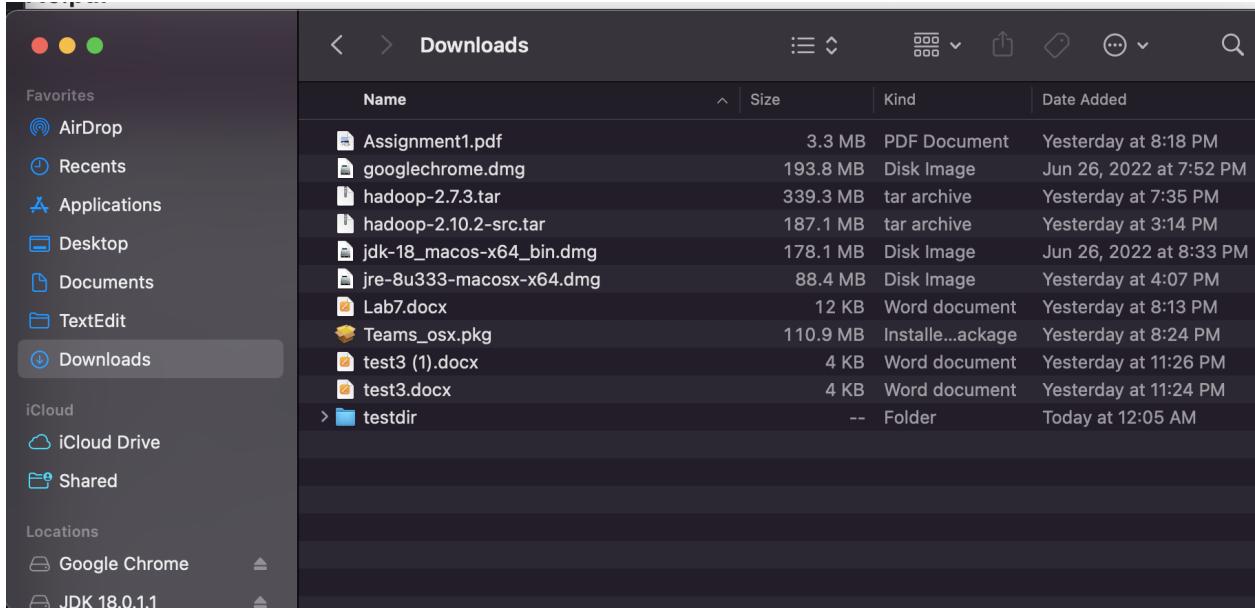
./hadoop fs -copyToLocal /testdir/ file:///Users/dikshabhatia/Downloads/

```

dikshabhatia:bin/ $ ./hadoop fs -copyToLocal /testdir/ file:///Users/dikshabhatia/Downloads/
22/06/29 00:05:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $ 

```

Screenshot of testdir in the downloads folder



- Count
- Count the number of directories, files and bytes under the paths that match the specified file pattern.

Command: **./hadoop fs -count /testdir**

```
22/06/29 00:08:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $ ./hadoop fs -count /testdir
22/06/29 00:08:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
      1          3        638486 /testdir
dikshabhatia:bin/ $ [0:05:15] [0:08:56]
```

- Cp
- Copy files from source to destination.

Command: **./hadoop fs -cp /testdir /testdir1**

Created 1 more folder testdir1 in the files directory

Browse Directory

Browse Directory							
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	dikshabhatia	supergroup	0 B	6/28/2022, 11:00:51 PM	0	0 B	Users
drwxr-xr-x	dikshabhatia	supergroup	0 B	6/29/2022, 12:00:51 AM	0	0 B	testdir
drwxr-xr-x	dikshabhatia	supergroup	0 B	6/29/2022, 12:10:47 AM	0	0 B	testdir1
drwxr-xr-x	dikshabhatia	supergroup	0 B	6/28/2022, 10:57:23 PM	0	0 B	user

Hadoop, 2016.

- Df
- Displays free space

Command: **./hadoop fs -df /testdir/**

```
Error: No command named '-df' was found. Perhaps you meant 'hadoop df'
[dikshabhatia:bin] $ ./hadoop fs -df /testdir/
[0:15:46]

22/06/29 08:15:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Filesystem          Size   Used  Available Usage%
hdfs://localhost:9000  245107195904  1323008  202117662656   0%
[dikshabhatia:bin] $ [0:15:55]
```

- **Du, dus**
- Displays sizes of files and directories contained in the given directory or the length of a file in case its just a file.
- Displays a summary of file lengths.

Command:

```
./hadoop fs -du /testdir/
./hadoop fs -dus /testdir/
```

```
dikshabhatia:bin] $ ./hadoop fs -du /testdir/
[0:15:55]

22/06/29 08:18:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
626956  /testdir/example6.pages
0       /testdir/sample.txt
3538   /testdir/test3.docx
[dikshabhatia:bin] $ ./hadoop fs -dus /testdir/
[0:19:00]

du: DEPRECATED: Please use 'du -s' instead.
22/06/29 08:19:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
630486  /testdir
[dikshabhatia:bin] $ [0:19:06]
```

- **Find**
- Command:
- **./hadoop fs -find /testdir/**

```
-----[REDACTED]-----
1      26    511085627 /nyse_local
[dikshabhatia:bin] $ ./hadoop fs -find /testdir/
[1:01:14]

22/06/29 15:28:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
/testdir
/testdir/example6.pages
/testdir/sample.txt
/testdir/test3.docx
[dikshabhatia:bin] $ [15:28:05]
```

- **Get**
- Copy files to the local file system.
- Command: **./hadoop fs -get /testdir/**

```
[testdir/test3.docx]
[dikshabhatia:bin] $ ./hadoop fs -get /testdir/
[15:20:05]

22/06/29 15:22:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin] $ ./hadoop fs -get /testdir/
[15:22:18]

22/06/29 15:23:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
get: '/testdir/test3.docx': File does not exist
get: 'testdir/sample.txt': File exists
get: 'testdir/test3.docx': File exists
[dikshabhatia:bin] $ [15:23:06]
```

- **getfacl**
- Command: **./hadoop fs -getfacl -R /**

```
-getfacl: <path> is missing
Usage: hadoop fs [generic options] -getfacl [-R] <path>
[dikshabhatia:bin] $ ./hadoop fs -getfacl -R /
[15:25:29]

22/06/30 15:25:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
# file: /
# owner: dikshabhatia
# group: supergroup
getfacl: The ACL operation has been rejected. Support for ACLs has been disabled by setting dfs.namenode.acls.enabled to false.
[dikshabhatia:bin] $ [15:25:58]
```

- **getattr**
- Displays the extended attribute names and values (if any) for a file or directory.

- Command: **./hadoop fs -getfattr -d /testdir**

```
dikshabhatia:bin/ $ ./hadoop fs -getfattr -d /testdir
getfattr: The ACL operation has been rejected. Support for ACLs has been disabled by setting dfs.namenode.acls.enabled to false.
22/06/30 15:27:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
# file: /testdir
dikshabhatia:bin/ $ [15:27:56]
```

- **getmerge**

- Takes a source directory and a destination file as input and concatenates files in src into the destination local file.
- Command: **./hadoop fs -getmerge -d /homework3/sample.txt /homework3/samplecopy.txt /homework3/merged.txt**

```
..oop-2.7.3/bin ...mebrew/bin:/opt/homebrew/sbin:/usr/local/bin:/usr/bin:/usr/sbin:/sbin ..sktop
dikshabhatia:bin/ $ ./hadoop fs -mkdir /homework3
22/06/30 15:34:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $ ./hadoop fs -copyFromLocal file:///Users/dikshabhatia/Desktop/info/a3/sample.txt /homework3/
22/06/30 15:35:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $ ./hadoop fs -copyFromLocal file:///Users/dikshabhatia/Desktop/info/a3/samplecopy.txt /homework3/
22/06/30 15:35:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $ ./hadoop fs -getmerge -d /homework3/sample.txt /homework3/samplecopy.txt /homework3/merged.txt
-getmerge: Illegal option -d
Usage: hadoop fs [generic options] -getmerge [-nl] <src> <localdst>
dikshabhatia:bin/ $ ./hadoop fs -getmerge /homework3/sample.txt /homework3/samplecopy.txt /homework3/merged.txt
22/06/30 15:38:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
getmerge: Mkdir failed to create file:/homework3 (exists=false, cwd=file:/usr/local/bin/hadoop-2.7.3/bin)
dikshabhatia:bin/ $ ./hadoop fs -getmerge /homework3/sample.txt /homework3/samplecopy.txt /homework3/merged.txt
22/06/30 15:39:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
getmerge: Mkdir failed to create file:/homework3/merged (exists=false, cwd=file:/usr/local/bin/hadoop-2.7.3/bin)
dikshabhatia:bin/ $ ./hadoop fs -getmerge /homework3/sample.txt /homework3/samplecopy.txt merged.txt
22/06/30 15:39:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $ [15:39:24]
```

- Merges the output in a txt file shown below

```
nano ...mebrew/bin:/opt/homebrew/sbin:/usr/local/bin:/usr/bin:/usr/sbin:/sbin ..sktop/info/a3
Uw PICO 5.09
Test 1test 1
Sample copy
File: merged.txt
```

- **Lsr, ls**

- Lists all directories
- Command: **dikshabhatia:bin/ \$./hadoop fs -lsr /homework3**
- Command: **./hadoop fs -ls**

```
dikshabhatia:bin/ $ ./hadoop fs -ls
22/06/30 15:43:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items:
drwxr-xr-x - dikshabhatia supergroup 0 2022-06-28 22:57 Data
drwxr-xr-x - dikshabhatia supergroup 0 2022-06-29 00:47 nsye
dikshabhatia:bin/ $ ./hadoop fs -lsr /homework3
lsr: DEPRECATED: Please use 'ls -R' instead.
22/06/30 15:43:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
drwxr-xr-x - dikshabhatia supergroup 0 2022-06-30 15:43 /homework3/homework3a
-rw-r--r-- 1 dikshabhatia supergroup 6 2022-06-30 15:35 /homework3/sample.txt
-rw-r--r-- 1 dikshabhatia supergroup 20 2022-06-30 15:35 /homework3/samplecopy.txt
dikshabhatia:bin/ $ [15:43:25]
```

- **mv**

- Command: **./hadoop fs -mv /testdir /testdir1**

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	dikshabhatia	supergroup	612.26 KB	6/29/2022, 12:10:46 AM	1	128 MB	example6.pages
-rw-r--r--	dikshabhatia	supergroup	0 B	6/29/2022, 12:10:46 AM	1	128 MB	sample.txt
-rw-r--r--	dikshabhatia	supergroup	3.45 KB	6/29/2022, 12:10:47 AM	1	128 MB	test3.docx
drwxr-xr-x	dikshabhatia	supergroup	0 B	6/29/2022, 12:00:51 AM	0	0 B	testdir

Hadoop, 2016.

- Rm: removes file
- Command: ./hadoop fs -rm /homework3/sample.txt

```
rm: '/user/hadoop/emptydir': No such file or directory
dikshabhatia@bin: ~ $ ./hadoop fs -rm /homework3/sample.txt [15:50:46]
22/06/30 15:51:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
22/06/30 15:51:32 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /homework3/sample.txt [15:51:32]
```

- Rmdir, rmr
- Command: ./hadoop fs -rmdir /empty
- Command: ./hadoop fs -rmr /homework3

```
rmdir: '/empty': No such file or directory
dikshabhatia@bin: ~ $ ./hadoop fs -mkdir /empty [15:54:56]
22/06/30 15:55:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia@bin: ~ $ ./hadoop fs -rmdir /empty [15:55:18]
22/06/30 15:55:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia@bin: ~ $ ./hadoop fs -rmr /nyse_local [15:55:27]
rmr: DEPRECATED: Please use 'rm -r' instead.
22/06/30 15:55:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
22/06/30 15:55:48 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /nyse_local [15:55:48]
```

- setattr
- Sets an extended attribute name and value for a file or directory.
- Command: ./hadoop fs -setattr -n **diksha.myName** -v myValue /testdir1
- setrep
- ./hadoop fs -setrep -w 5 /testdir1/sample1.txt

```
dikshabhatia:bin/ $ ./hadoop fs -setrep -w 5 /testdir1/sample1.txt [16:02:26]
22/06/30 16:04:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
setrep: /testdir1/sample1.txt: No such file or directory
dikshabhatia:bin/ $ ./hadoop fs -setrep -w 5 /testdir1/sample.txt [16:04:57]
22/06/30 16:05:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Replication 5 set: /testdir1/sample.txt
Waiting for /testdir1/sample.txt ... done
dikshabhatia:bin/ $ [16:05:15]
```

- **text**

- `./hadoop fs -text /testdir1/sample.txt`

-

- **touchz**

- `./hadoop fs -touchz /testdir1/emptytext.txt`

```
..oop-2.7.3/bin ...mebrew/bin:/opt/homebrew/sbin:/usr/local/bin:/usr/bin:/bin:/usr/sbin:/sbin
dikshabhatia:bin/ $ ./hadoop fs -touchz /testdir1/emptytext.txt
22/06/30 16:06:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $
```

-

- **truncate**

- `./hadoop fs -truncate 1 /testdir1/sample.txt`

```
Usage: hadoop fs [generic options] -truncate [-w] <length> <path> ...
dikshabhatia:bin/ $ ./hadoop fs -truncate 1 /testdir1/sample.txt
22/06/30 16:09:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
-truncate: Cannot truncate to a larger file size. Current size: 0, truncate size: 1.
Usage: hadoop fs [generic options] -truncate [-w] <length> <path> ...
dikshabhatia:bin/ $
```

-

- **usage**

- `./hadoop fs -usage truncate`

```
..oop-2.7.3/bin ...mebrew/bin:/opt/homebrew/sbin:/usr/local/bin:/usr/bin:/bin:/usr/sbin:/sbin
dikshabhatia:bin/ $ ./hadoop fs -touchz /testdir1/emptytext.txt
22/06/30 16:06:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $ ./hadoop fs -usage truncate
Usage: hadoop fs [generic options] -truncate [-w] <length> <path> ...
dikshabhatia:bin/ $
```

-

PART 7. Programming Assignment

HDFS Command Assignment - Submit screenshots

Copy all the files (A to Z) from NYSE dataset into HDFS.

Step 1: Created local directory

`./hadoop fs -mkdir /nyse_local`

Step 2: Bash script to copy files from local to hdfs

```
#!/bin/bash
```

```
for dataFileSuffix in {A..Z};  
do
```

`FILELOC="file:///Users/dikshabhatia/Desktop/info/a3/NYSE/NYSE_daily_prices_$dataFileSuffix.csv"`

```
echo $FILELOC
```

```
./hadoop fs -copyFromLocal $FILELOC /nyse_local/
```

Done

Step 3: Run from bin

```
bash /Users/dikshabhatia/Desktop/info/a3/importnsye.sh
```

Step 5: Verify count

Command: `dikshabhatia:bin/ $./hadoop fs -count /nyse_local/`

```

22/06/29 00:54:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
dikshabhatia:bin/ $ ./hadoop fs -count /nyse_local/
22/06/29 00:54:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count: '/nyse_local/': No such file or directory
dikshabhatia:bin/ $ ./hadoop fs -count /nyse_local/
22/06/29 01:01:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
      1          26      511885627 /nyse_local
dikshabhatia:bin/ $ [0:54:57]
[1:00:56] [1:01:14]

```

Part 7:

PART 7. HDFS Command Assignment - Submit screenshots

Copy all the files (A to Z) from NYSE dataset into HDFS

```

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.*;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import java.io.IOException;
public class PutMergeAttempt {
    public static void main(String args[]) throws Exception{
        Configuration config = new Configuration();
        FileSystem hdfs = FileSystem.get(config);
        FileSystem local = FileSystem.getLocal(config);
        Path op = new Path(args[1]);
        Path ip = new Path(args[0]);
        try{
            FileStatus[] ipFile = localFile.listStatus(ip);
            FSDataOutputStream outputStream = hdfs.create(op);
            for(int i = 0; i<inputFiles.length; i++){
                FSDataInputStream input = localFile.open(ipFile[i].getPath()); byte buffer[] = new byte[256];
                int bytesRead = 0;

```

```

while( (bytesRead = input.read(buffer)) > 0){

output.write(buffer, 0 ,bytesRead);

}

input.close(); }

output.close();

}catch (IOException e){

e.printStackTrace(); }

}

```

Command:./hadoop jar **Users/dikshabhatia/Desktop/info/a3/assign3.jar** PutMergeAttempt /NYSE

The screenshot shows a web browser window with multiple tabs open. The active tab is titled 'localhost:50070/explorer.html#/testdir1'. The browser's address bar also displays this URL. The page content is a 'Browse Directory' interface for HDFS, showing a list of files in the '/testdir1' directory. The files listed are 'example6.pages', 'sample.txt', 'test3.docx', and 'testdir'. The interface includes a search bar at the top and a navigation menu below it.

Browse Directory

/testdir1								Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
-rw-r--r--	dikshabhatia	supergroup	612.26 KB	6/29/2022, 12:10:46 AM	1	128 MB	example6.pages	
-rw-r--r--	dikshabhatia	supergroup	0 B	6/29/2022, 12:10:46 AM	1	128 MB	sample.txt	
-rw-r--r--	dikshabhatia	supergroup	3.45 KB	6/29/2022, 12:10:47 AM	1	128 MB	test3.docx	
drwxr-xr-x	dikshabhatia	supergroup	0 B	6/29/2022, 12:00:51 AM	0	0 B	testdir	

Hadoop, 2016.

