Mayank Gangrade
INFO 7250: Fall 2019

Part 2.1

```
db.nysedata.mapReduce(
   function(){
       if(this.stock_price_high != null)
             emit( this.stock_symbol, this.stock_price_high);
   },
   function(key, values) {
       return Array.avg(values)
   },
   {
       "out": "part2"
   }
)
```

```
> db.part2.find()
{ "_id" : NaN, "value" : 14.348379501117657 }
{ "_id" : "AA", "value" : 64.26183302168995 }
{ "_id" : "AAI", "value" : 21.94922394259419 }
{ "_id" : "AAN", "value" : 8.94769631910608 }
{ "_id" : "AAP", "value" : 45.020595856255326 }
{ "_id" : "AAR", "value" : 22.850850444401562 }
{ "_id" : "AAV", "value" : 14.410471696271534 }
{ "_id" : "AB", "value" : 4.039495082136035 }
{ "_id" : "ABA", "value" : 25.357926773840354 }
{ "_id" : "ABB", "value" : 17.79397760193093 }
{ "_id" : "ABC", "value" : 23.223538539062925 }
{ "_id" : "ABD", "value" : 25.75448769531479 }
{ "_id" : "ABG", "value" : 17.71472997300015 }
{ "_id" : "ABK", "value" : 24.142518872829548 }
{ "_id" : "ABM", "value" : 22.647381781987022 }
{ "_id" : "ABR", "value" : 19.66816309383276 }
{ "_id" : "ABT", "value" : 45.674215680863774 }
{ "_id" : "ABV", "value" : 11.12086561737346 }
{ "_id" : "ABVT", "value" : 45.699488142200536 }
{ "_id" : "ABX", "value" : 4.829179083567142 }
Type "it" for more
```

Part 2.2

```
db.nysedata.mapReduce(
   function(){
      if(this.stock_price_high != null)
         emit( this.stock_symbol, { "cnt": 1, "price":
this.stock_price_high } );
   },
   function(key, values) {
      rv = { "cnt" : 0, "price" : 0 };
      values.forEach(function(value) {
         rv.cnt += value.cnt;
         rv.price += value.price;
      });
      return rv;
   },
   {
      "out": "part3" ,
      "finalize": function(key, reduceVal) {
         reduceVal.avg = reduceVal.price / reduceVal.cnt;
         return reduceVal;
      }
   }
)
```

Mayank Gangrade
INFO 7250: Fall 2019

Part 3

```
> db.part3.find()
{ "_id" : NaN, "value" : { "cnt" : 2687, "price" : 38560.6600000002, "avg" : 14.350822478600744 } }
{ "_id" : "AA", "value" : { "cnt" : 12109, "price" : 635234.290000002, "avg" : 52.459682054670246 } }
{ "_id" : "AAI", "value" : { "cnt" : 3933, "price" : 41369.05000000041, "avg" : 10.518446478515234 } }
{ "_id" : "AAN", "value" : { "cnt" : 4218, "price" : 83717.14999999902, "avg" : 19.847593646277623 } }
{ "_id" : "AAP", "value" : { "cnt" : 2058, "price" : 92036.46000000005, "avg" : 44.721311953352796 } }
{ "_id" : "AAR", "value" : { "cnt" : 2632, "price" : 50557.92000000018, "avg" : 19.208936170212834 } }
{ "_id" : "AAV", "value" : { "cnt" : 1435, "price" : 17935.320000000014, "avg" : 12.498480836236944 } }
{ "_id" : "AB", "value" : { "cnt" : 5495, "price" : 168401.26999999976, "avg" : 30.64627297543217 } }
{ "_id" : "ABA", "value" : { "cnt" : 906, "price" : 23550.98999999999, "avg" : 25.994470198675486 } }
{ "_id" : "ABB", "value" : { "cnt" : 2221, "price" : 27948.20000000001, "avg" : 12.583610986042329 } }
{ "_id" : "ABC", "value" : { "cnt" : 3733, "price" : 178398.48000000106, "avg" : 47.7895740691136 } }
{ "_id" : "ABD", "value" : { "cnt" : 1127, "price" : 17718.600000000013, "avg" : 15.721916592724059 } }
{ "_id" : "ABG", "value" : { "cnt" : 1984, "price" : 30611.2300000001, "avg" : 15.429047379032308 } }
{ "_id" : "ABK", "value" : { "cnt" : 4682, "price" : 240267.1700000026, "avg" : 51.31720845792452 } }
{ "_id" : "ABM", "value" : { "cnt" : 6446, "price" : 158110.459999997, "avg" : 24.528461061122712 } }
{ "_id" : "ABR", "value" : { "cnt" : 1450, "price" : 26724.670000000006, "avg" : 18.43080689655173 } }
{ "_id" : "ABT", "value" : { "cnt" : 6772, "price" : 326329.2000000044, "avg" : 48.188009450679914 } }
{ "_id" : "ABV", "value" : { "cnt" : 3284, "price" : 105043.44000000024, "avg" : 31.986431181486065 } }
{ "_id" : "ABVT", "value" : { "cnt" : 1520, "price" : 74779.0200000001, "avg" : 49.196723684210596 } }
{ "_id" : "ABX", "value" : { "cnt" : 6303, "price" : 142971.01000000082, "avg" : 22.683009677931274 } }
Type "it" for more
>
```

Part 4

```
> use nyse_before_index
switched to db nyse_before_index
> db.nysedata.createIndex({stock_symbol: 1, date: 1})
{
        "createdCollectionAutomatically" : true,
        "numIndexesBefore" : 1,
        "numIndexesAfter" : 2,
        "ok" : 1
}
>
```

```
> db.nysedata.getIndexes()
[
        {
                "v" : 2,
                "key" : {
                        "_id" : 1
                },
                "name" : "_id_",
                "ns" : "nyse_before_index.nysedata"
        },
        {
                "v" : 2,
                "key" : {
                        "stock_symbol" : 1,
                        "date" : 1
                },
                "name" : "stock_symbol_1_date_1",
                "ns" : "nyse_before_index.nysedata"
        }
]
```

Mayank Gangrade
INFO 7250: Fall 2019

Part 5

```
> use nyse_after_index
switched to db nyse_after_index
> db.nysedata.getIndexes()
[
        {
                "v" : 2,
                "key" : {
                        "_id" : 1
                },
                "name" : "_id_",
                "ns" : "nyse_after_index.nysedata"
        }
]
> db.nysedata.createIndex({stock_symbol: 1, date: 1})
{
        "createdCollectionAutomatically" : false,
        "numIndexesBefore" : 1,
        "numIndexesAfter" : 2,
        "ok" : 1
}
> db.nysedata.getIndexes()
[
        {
                "v" : 2,
                "key" : {
                        "_id" : 1
                },
                "name" : "_id_",
                "ns" : "nyse_after_index.nysedata"
        },
        {
                "v" : 2,
                "key" : {
                        "stock_symbol" : 1,
                        "date" : 1
                },
                "name" : "stock_symbol_1_date_1",
                "ns" : "nyse_after_index.nysedata"
        }
]
> |
```

Mayank Gangrade
INFO 7250: Fall 2019

Part 6.1

```
db.rating.mapReduce(
  function(){
          emit(this.movieId, 1);
  },
  function(key, values) {
      return Array.sum(values);
  },
  {
      "out": "part6_1",
        "sort": {"values": 1}
  }
)
```

```
> db.part6_1.find()
{ "_id" : "1", "value" : 26449 }
{ "_id" : "10", "value" : 16918 }
{ "_id" : "100", "value" : 2640 }
{ "_id" : "1000", "value" : 128 }
{ "_id" : "1001", "value" : 23 }
{ "_id" : "1002", "value" : 103 }
{ "_id" : "1003", "value" : 1027 }
{ "_id" : "1004", "value" : 977 }
{ "_id" : "1005", "value" : 1128 }
{ "_id" : "1006", "value" : 991 }
{ "_id" : "1007", "value" : 1103 }
{ "_id" : "1008", "value" : 491 }
{ "_id" : "1009", "value" : 1598 }
{ "_id" : "101", "value" : 2226 }
{ "_id" : "1010", "value" : 1540 }
{ "_id" : "1011", "value" : 973 }
{ "_id" : "1012", "value" : 2039 }
{ "_id" : "1013", "value" : 2365 }
{ "_id" : "1014", "value" : 1000 }
{ "_id" : "1015", "value" : 1870 }
Type "it" for more
```

```
db.part6_1.aggregate(
    [
      { $sort : { value : -1 } },
        { $limit : 25 }
    ]
)
```

```
> db.part6_1.aggregate(
...    [
...        { $sort : { value : -1 } },
...    { $limit : 25 }
...    ]
... )
{ "_id" : "296", "value" : 34864 }
{ "_id" : "356", "value" : 34457 }
{ "_id" : "593", "value" : 33668 }
{ "_id" : "480", "value" : 32631 }
{ "_id" : "318", "value" : 31126 }
{ "_id" : "110", "value" : 29154 }
{ "_id" : "457", "value" : 28951 }
{ "_id" : "589", "value" : 28948 }
{ "_id" : "260", "value" : 28566 }
{ "_id" : "150", "value" : 27035 }
{ "_id" : "592", "value" : 26996 }
{ "_id" : "1", "value" : 26449 }
{ "_id" : "780", "value" : 26042 }
{ "_id" : "590", "value" : 25912 }
{ "_id" : "527", "value" : 25777 }
{ "_id" : "380", "value" : 25381 }
{ "_id" : "1210", "value" : 25098 }
{ "_id" : "32", "value" : 24397 }
{ "_id" : "50", "value" : 24037 }
{ "_id" : "608", "value" : 23794 }
Type "it" for more
```

Mayank Gangrade
INFO 7250: Fall 2019

Part 6.2

```
db.users.mapReduce(
    function(){
            emit( this.gender, 1);
    },
    function(key, values) {
        return Array.sum(values)
    },
    {
        "out": "part6_2"
    }
)
```

```
> db.part6_2.find()
{ "_id" : "F", "value" : 1709 }
{ "_id" : "M", "value" : 4331 }
>
```

Mayank Gangrade
INFO 7250: Fall 2019

Part 6.3

```
db.rating.mapReduce(
    function(){
            emit( this.movieId, 1);
    },
    function(key, values) {
        return Array.sum(values)
    },
    {
        "out": "part6_3"
    }
)
```

```
> db.part6_3.find()
{ "_id" : "1", "value" : 26449 }
{ "_id" : "10", "value" : 16918 }
{ "_id" : "100", "value" : 2640 }
{ "_id" : "1000", "value" : 128 }
{ "_id" : "1001", "value" : 23 }
{ "_id" : "1002", "value" : 103 }
{ "_id" : "1003", "value" : 1027 }
{ "_id" : "1004", "value" : 977 }
{ "_id" : "1005", "value" : 1128 }
{ "_id" : "1006", "value" : 991 }
{ "_id" : "1007", "value" : 1103 }
{ "_id" : "1008", "value" : 491 }
{ "_id" : "1009", "value" : 1598 }
{ "_id" : "101", "value" : 2226 }
{ "_id" : "1010", "value" : 1540 }
{ "_id" : "1011", "value" : 973 }
{ "_id" : "1012", "value" : 2039 }
{ "_id" : "1013", "value" : 2365 }
{ "_id" : "1014", "value" : 1000 }
{ "_id" : "1015", "value" : 1870 }
Type "it" for more
```

Mayank Gangrade
INFO 7250: Fall 2019

Part 7

-appendToFile

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -appendToFile /home/mayank/myFile3.txt /testdir/myfile1.txt
2019-10-07 03:04:24,948 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$
```

-cat

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -cat /testdir/myfile1.txt
2019-10-07 03:13:30,480 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Hi this is the file 1
this is empty file
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$
```

-checksum

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -checksum /testdir/myfile1.txt
/testdir/myfile1.txt    MD5-of-0MD5-of-512CRC32C        000002000000000000000000099f03d1fa705eb3d0cfb2a45f1fdca26
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$
```

-copyFromLocal

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -copyFromLocal /home/mayank/myFile4.txt /testdir
2019-10-07 03:17:31,148 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$
```

-copyToLocal

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -copyToLocal /testdir/myFile4.txt /home/mayank/
2019-10-07 03:19:04,280 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$
```

-count

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -count /testdir
           1            4            101 /testdir
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$
```

-cp

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -cp /testdir/myFile4.txt /testdir1
2019-10-07 03:21:58,544 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2019-10-07 03:21:58,736 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$
```

-df

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -df /testdir
Filesystem                  Size     Used    Available   Use%
hdfs://localhost:9000   69332221952   65626   43284004864     0%
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$
```

Mayank Gangrade
INFO 7250: Fall 2019

-du

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -du /testdir
19   19   /testdir/myFile3.txt
19   19   /testdir/myFile4.txt
41   41   /testdir/myfile1.txt
22   22   /testdir/myfile2.txt
19   19   /testdir/myfile5.txt
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$
```

-dus (-du -s)

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -du -s /testdir
120   120   /testdir
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$
```

-find

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -find /testdir
/testdir
/testdir/myFile3.txt
/testdir/myFile4.txt
/testdir/myfile1.txt
/testdir/myfile2.txt
/testdir/myfile5.txt
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$
```

-get

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -get /testdir/myFile4.txt /home/mayank/myfile5.txt
2019-10-07 03:30:11,163 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$
```

-getfacl

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -getfacl /testdir
# file: /testdir
# owner: mayank
# group: supergroup
user::rwx
group::r-x
other::r-x
```

-head
-help

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -head /testdir/myfile1.txt
2019-10-07 03:32:19,839 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Hi this is the file 1
this is empty file
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -help
Usage: hadoop fs [generic options]
        [-appendToFile <localsrc> ... <dst>]
        [-cat [-ignoreCrc] <src> ...]
        [-checksum <src> ...]
        [-chgrp [-R] GROUP PATH...]
        [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
        [-chown [-R] [OWNER][:[GROUP]] PATH...]
```

Mayank Gangrade
INFO 7250: Fall 2019

-ls
-mkdir

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -ls /testdir
Found 5 items
-rw-r--r--   1 mayank supergroup         19 2019-10-07 03:03 /testdir/myFile3.txt
-rw-r--r--   1 mayank supergroup         19 2019-10-07 03:17 /testdir/myFile4.txt
-rw-r--r--   1 mayank supergroup         41 2019-10-07 03:04 /testdir/myfile1.txt
-rw-r--r--   1 mayank supergroup         22 2019-10-07 03:01 /testdir/myfile2.txt
-rw-r--r--   1 mayank supergroup         19 2019-10-07 03:24 /testdir/myfile5.txt
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -mkdir /testdir/inside_dir
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ |
```

-moveFromLocal
-mv

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -moveFromLocal /home/mayank/myfile1.txt /testdir/inside_dir
2019-10-07 03:35:55,249 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -mv /testdir/myfile2.txt /testdir/inside_dir
```

-rm
-rmdir

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -rm /testdir/inside_dir/myfile1.txt
Deleted /testdir/inside_dir/myfile1.txt
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -rmdir /testdir/inside_dir
rmdir: `/testdir/inside_dir': Directory is not empty
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -rmdir -ignore-fail-on-empty /testdir/inside_dir
```

-tail
-test

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -tail /testdir/myfile5.txt
2019-10-07 03:41:27,886 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
this is empty file
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -test -e /testdir/myfile5.txt
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -test -f /testdir/myfile5.txt
```

-touchz
-truncate

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -touchz /testdir/myfile10.txt
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -truncate /testdir/myfile10.txt
```

-usage

```
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -usage ls
Usage: hadoop fs [generic options] -ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [-e] [<path> ...]
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -usage mkdir
Usage: hadoop fs [generic options] -mkdir [-p] <path> ...
mayank@mayank:/usr/local/bin/hadoop-3.2.1/bin$ |
```