Prerequsites:

Rides.csv was generated using generate_data.ipynb

This repository was used as a reference for mongo cluster:

https://github.com/minhhungit/mongodb-cluster-docker-compose

Step 1:

Starting the cluster using docker-compose up -d

```
D:\All_my_Code\University\Master\HL-2022\Lab-3>docker-compose up -d
[+] Running 14/14
 - Network lab-3_default      Created                                    0.7s
 - Container shard-01-node-c  Started                                    4.3s
 - Container shard-02-node-a  Started                                   10.1s
 - Container mongo-config-01  Started                                    5.2s
 - Container router-01        Started                                    3.5s
 - Container shard-03-node-b  Started                                    9.7s
 - Container mongo-config-02  Started                                    8.0s
 - Container shard-02-node-b  Started                                    5.9s
 - Container shard-03-node-a  Started                                    6.2s
 - Container shard-02-node-c  Started                                    4.1s
 - Container shard-03-node-c  Started                                    9.0s
 - Container shard-01-node-b  Started                                    9.7s
 - Container mongo-config-03  Started                                    6.7s
 - Container shard-01-node-a  Started                                    4.5s
```

Step 2:

Initializing shards one by one.

docker-compose exec configsvr01 sh -c "mongosh < /scripts/init-configserver.js"

```
D:\All_my_Code\University\Master\HL-2022\Lab-3>docker-compose exec configsvr01 sh -c "mongosh < /scripts/init-configserver.js"
Current Mongosh Log ID: 63c9922cced348d4618e1ce6
Connecting to:          mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000&appName=mongosh+1.5.4
Using MongoDB:          6.0.1
Using Mongosh:          1.5.4

For mongosh info see: https://docs.mongodb.com/mongodb-shell/


To help improve our products, anonymous usage data is collected and sent to MongoDB periodically (https://www.mongodb.com/legal/privacy-policy).
You can opt-out by running the disableTelemetry() command.

------
   The server generated these startup warnings when booting
   2023-01-19T18:53:34.248+00:00: Using the XFS filesystem is strongly recommended with the WiredTiger storage engine. See http://dochub.mongodb.org/core/prodnotes-filesystem
   2023-01-19T18:53:36.316+00:00: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted
   2023-01-19T18:53:36.316+00:00: /sys/kernel/mm/transparent_hugepage/enabled is 'always'. We suggest setting it to 'never'
   2023-01-19T18:53:36.316+00:00: vm.max_map_count is too low
------

------
   Enable MongoDB's free cloud-based monitoring service, which will then receive and display
   metrics about your deployment (disk utilization, CPU, operation statistics, etc).

   The monitoring data will be available on a MongoDB website with a unique URL accessible to you
   and anyone you share the URL with. MongoDB may use this information to make product
   improvements and to suggest MongoDB products and deployment options to you.

   To enable free monitoring, run the following command: db.enableFreeMonitoring()
   To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
------

test> rs.initiate(
...   {
.....     _id: "rs-config-server",
.....     configsvr: true,
.....     version: 1,
.....     members: [
.....       { _id: 0, host : "configsvr01:27017" },
.....       { _id: 1, host : "configsvr02:27017" },
.....       { _id: 2, host : "configsvr03:27017" }
.....     ]
.....   }
... ){ ok: 1, lastCommittedOpTime: Timestamp({ t: 1674154541, i: 1 }) }
rs-config-server [direct: other] test> )
D:\All_my_Code\University\Master\HL-2022\Lab-3>
```

docker-compose exec shard01-a sh -c "mongosh < /scripts/init-shard01.js"

```
D:\All_my_Code\University\Master\HL-2022\Lab-3>docker-compose exec shard01-a sh -c "mongosh < /scripts/init-shard01.js"
Current Mongosh Log ID: 63c9925b81d252fbcd4759f6
Connecting to:          mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000&appName=mongosh+1.5.4
Using MongoDB:          6.0.1
Using Mongosh:          1.5.4

For mongosh info see: https://docs.mongodb.com/mongodb-shell/


To help improve our products, anonymous usage data is collected and sent to MongoDB periodically (https://www.mongodb.com/legal/privacy-policy).
You can opt-out by running the disableTelemetry() command.

------
   The server generated these startup warnings when booting
   2023-01-19T18:53:33.665+00:00: Using the XFS filesystem is strongly recommended with the WiredTiger storage engine. See http://dochub.mongodb.org/core/prodnotes-filesystem
   2023-01-19T18:53:35.659+00:00: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted
   2023-01-19T18:53:35.660+00:00: /sys/kernel/mm/transparent_hugepage/enabled is 'always'. We suggest setting it to 'never'
   2023-01-19T18:53:35.660+00:00: vm.max_map_count is too low
------

------
   Enable MongoDB's free cloud-based monitoring service, which will then receive and display
   metrics about your deployment (disk utilization, CPU, operation statistics, etc).

   The monitoring data will be available on a MongoDB website with a unique URL accessible to you
   and anyone you share the URL with. MongoDB may use this information to make product
   improvements and to suggest MongoDB products and deployment options to you.

   To enable free monitoring, run the following command: db.enableFreeMonitoring()
   To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
------

test> rs.initiate(
...     {
.....       _id: "rs-shard-01",
.....       version: 1,
.....       members: [
.....          { _id: 0, host : "shard01-a:27017" },
.....          { _id: 1, host : "shard01-b:27017" },
.....            { _id: 2, host : "shard01-c:27017" },
.....       ]
.....     }
... ){ ok: 1 }
rs-shard-01 [direct: other] test> )
```

docker-compose exec shard02-a sh -c "mongosh < /scripts/init-shard02.js"

```
D:\All_my_Code\University\Master\HL-2022\Lab-3>docker-compose exec shard02-a sh -c "mongosh < /scripts/init-shard02.js"
Current Mongosh Log ID: 63c9928c84f5ad3849fe3a18
Connecting to:          mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000&appName=mongosh+1.5.4
Using MongoDB:          6.0.1
Using Mongosh:          1.5.4

For mongosh info see: https://docs.mongodb.com/mongodb-shell/


To help improve our products, anonymous usage data is collected and sent to MongoDB periodically (https://www.mongodb.com/legal/privacy-policy).
You can opt-out by running the disableTelemetry() command.

------
   The server generated these startup warnings when booting
   2023-01-19T18:53:39.186+00:00: Using the XFS filesystem is strongly recommended with the WiredTiger storage engine. See http://dochub.mongodb.org/core/prodnotes-filesystem
   2023-01-19T18:53:40.381+00:00: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted
   2023-01-19T18:53:40.381+00:00: /sys/kernel/mm/transparent_hugepage/enabled is 'always'. We suggest setting it to 'never'
   2023-01-19T18:53:40.381+00:00: vm.max_map_count is too low
------

------
   Enable MongoDB's free cloud-based monitoring service, which will then receive and display
   metrics about your deployment (disk utilization, CPU, operation statistics, etc).

   The monitoring data will be available on a MongoDB website with a unique URL accessible to you
   and anyone you share the URL with. MongoDB may use this information to make product
   improvements and to suggest MongoDB products and deployment options to you.

   To enable free monitoring, run the following command: db.enableFreeMonitoring()
   To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
------

test> rs.initiate(
...     {
.....       _id: "rs-shard-02",
.....       version: 1,
.....       members: [
.....          { _id: 0, host : "shard02-a:27017" },
.....          { _id: 1, host : "shard02-b:27017" },
.....            { _id: 2, host : "shard02-c:27017" },
.....       ]
.....     }
... ){ ok: 1 }
rs-shard-02 [direct: other] test> )
```

docker-compose exec shard03-a sh -c "mongosh < /scripts/init-shard03.js"

```
D:\All_my_Code\University\Master\HL-2022\Lab-3>docker-compose exec shard03-a sh -c "mongosh < /scripts/init-shard03.js"
Current Mongosh Log ID: 63c992a8300936b2cab514e9
Connecting to:          mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000&appName=mongosh+1.5.4
Using MongoDB:          6.0.1
Using Mongosh:          1.5.4

For mongosh info see: https://docs.mongodb.com/mongodb-shell/


To help improve our products, anonymous usage data is collected and sent to MongoDB periodically (https://www.mongodb.com/legal/privacy-policy).
You can opt-out by running the disableTelemetry() command.

------
   The server generated these startup warnings when booting
   2023-01-19T18:53:35.460+00:00: Using the XFS filesystem is strongly recommended with the WiredTiger storage engine. See http://dochub.mongodb.org/core/prodnotes-filesystem
   2023-01-19T18:53:37.282+00:00: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted
   2023-01-19T18:53:37.282+00:00: /sys/kernel/mm/transparent_hugepage/enabled is 'always'. We suggest setting it to 'never'
   2023-01-19T18:53:37.282+00:00: vm.max_map_count is too low
------

------
   Enable MongoDB's free cloud-based monitoring service, which will then receive and display
   metrics about your deployment (disk utilization, CPU, operation statistics, etc).

   The monitoring data will be available on a MongoDB website with a unique URL accessible to you
   and anyone you share the URL with. MongoDB may use this information to make product
   improvements and to suggest MongoDB products and deployment options to you.

   To enable free monitoring, run the following command: db.enableFreeMonitoring()
   To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
------

test> rs.initiate(
...     {
.....       _id: "rs-shard-03",
.....       version: 1,
.....       members: [
.....         { _id: 0, host : "shard03-a:27017" },
.....         { _id: 1, host : "shard03-b:27017" },
.....         { _id: 2, host : "shard03-c:27017" },
.....       ]
.....     }
... ){ ok: 1 }
rs-shard-03 [direct: other] test> )
```

Step 3:

Initializing router.

docker-compose exec router sh -c "mongosh < /scripts/init-router.js"

```
D:\All_my_Code\University\Master\HL-2022\Lab-3>docker-compose exec router sh -c "mongosh < /scripts/init-router.js"
Current Mongosh Log ID: 63c994a48c754f5119ca1fcc
Connecting to:          mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000&appName=mongosh+1.5.4
Using MongoDB:          6.0.1
Using Mongosh:          1.5.4

For mongosh info see: https://docs.mongodb.com/mongodb-shell/


To help improve our products, anonymous usage data is collected and sent to MongoDB periodically (https://www.mongodb.com/legal/privacy-policy).
You can opt-out by running the disableTelemetry() command.

------
   The server generated these startup warnings when booting
   2023-01-19T19:04:53.977+00:00: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted
------

[direct: mongos] test> sh.addShard("rs-shard-01/shard01-a:27017")


[direct: mongos] test> sh.addShard("rs-shard-01/shard01-b:27017")
{
  shardAdded: 'rs-shard-01',
  ok: 1,
  '$clusterTime': {
    clusterTime: Timestamp({ t: 1674155172, i: 8 }),
    signature: {
      hash: Binary(Buffer.from("0000000000000000000000000000000000000000", "hex"), 0),
      keyId: Long("0")
    }
  },
  operationTime: Timestamp({ t: 1674155172, i: 8 })
}
[direct: mongos] test> sh.addShard("rs-shard-01/shard01-c:27017")
{
  shardAdded: 'rs-shard-01',
  ok: 1,
  '$clusterTime': {
    clusterTime: Timestamp({ t: 1674155172, i: 8 }),
    signature: {
      hash: Binary(Buffer.from("0000000000000000000000000000000000000000", "hex"), 0),
      keyId: Long("0")
    }
  },
  operationTime: Timestamp({ t: 1674155172, i: 8 })
}
[direct: mongos] test>
```

```
[direct: mongos] test> sh.addShard("rs-shard-02/shard02-a:27017")
{
  shardAdded: 'rs-shard-01',
  ok: 1,
  '$clusterTime': {
    clusterTime: Timestamp({ t: 1674155172, i: 8 }),
    signature: {
      hash: Binary(Buffer.from("0000000000000000000000000000000000000000", "hex"), 0),
      keyId: Long("0")
    }
  },
  operationTime: Timestamp({ t: 1674155172, i: 8 })
}
[direct: mongos] test> sh.addShard("rs-shard-02/shard02-b:27017")
{
  shardAdded: 'rs-shard-02',
  ok: 1,
  '$clusterTime': {
    clusterTime: Timestamp({ t: 1674155173, i: 8 }),
    signature: {
      hash: Binary(Buffer.from("0000000000000000000000000000000000000000", "hex"), 0),
      keyId: Long("0")
    }
  },
  operationTime: Timestamp({ t: 1674155173, i: 8 })
}
[direct: mongos] test> sh.addShard("rs-shard-02/shard02-c:27017")
{
  shardAdded: 'rs-shard-02',
  ok: 1,
  '$clusterTime': {
    clusterTime: Timestamp({ t: 1674155173, i: 8 }),
    signature: {
      hash: Binary(Buffer.from("0000000000000000000000000000000000000000", "hex"), 0),
      keyId: Long("0")
    }
  },
  operationTime: Timestamp({ t: 1674155173, i: 8 })
}
[direct: mongos] test>

[direct: mongos] test> sh.addShard("rs-shard-03/shard03-a:27017")
{
  shardAdded: 'rs-shard-02',
  ok: 1,
  '$clusterTime': {
    clusterTime: Timestamp({ t: 1674155173, i: 8 }),
    signature: {
      hash: Binary(Buffer.from("0000000000000000000000000000000000000000", "hex"), 0),
      keyId: Long("0")
    }
  },
  operationTime: Timestamp({ t: 1674155173, i: 8 })
}
[direct: mongos] test> sh.addShard("rs-shard-03/shard03-b:27017")
{
  shardAdded: 'rs-shard-03',
  ok: 1,
  '$clusterTime': {
    clusterTime: Timestamp({ t: 1674155173, i: 15 }),
```

Step 5:

Importing data

docker-compose exec router sh -c "mongosh < /scripts/enable-sharding.js"

```
D:\All_my_Code\University\Master\HL-2022\Lab-3>docker-compose exec router sh -c "mongosh < /scripts/eanble-sharding.js"
Current Mongosh Log ID: 63c9952f199489fd3734a1f8
Connecting to:          mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000&appName=mongosh+1.5.4
Using MongoDB:          6.0.1
Using Mongosh:          1.5.4

For mongosh info see: https://docs.mongodb.com/mongodb-shell/

------
   The server generated these startup warnings when booting
   2023-01-19T19:04:53.977+00:00: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted
------

[direct: mongos] test> sh.enableSharding('taxi')


[direct: mongos] test> sh.shardCollection('taxi.rides', {start_latitude: "hashed"}){
  ok: 1,
  '$clusterTime': {
    clusterTime: Timestamp({ t: 1674155312, i: 5 }),
    signature: {
      hash: Binary(Buffer.from("0000000000000000000000000000000000000000", "hex"), 0),
      keyId: Long("0")
    }
  },
  operationTime: Timestamp({ t: 1674155312, i: 3 })
}
[direct: mongos] test> {h.shardCollection('taxi.rides', {start_latitude: "hashed"})
  collectionsharded: 'taxi.rides',
  ok: 1,
  '$clusterTime': {
    clusterTime: Timestamp({ t: 1674155314, i: 43 }),
    signature: {
      hash: Binary(Buffer.from("0000000000000000000000000000000000000000", "hex"), 0),
      keyId: Long("0")
    }
  },
  operationTime: Timestamp({ t: 1674155314, i: 39 })
}
[direct: mongos] test> sh.shardCollection('taxi.rides', {start_latitude: "hashed"})
```

docker-compose exec router sh -c "mongoimport --port 27017 -d taxi -c rides --type csv --file rides.csv --headerline"

```
D:\All_my_Code\University\Master\HL-2022\Lab-3>docker-compose exec router sh -c "mongoimport --port 27017 -d taxi -c rides --type csv --file scripts/rides.csv --headerline"
2023-01-19T19:12:15.238+0000    connected to: mongodb://localhost:27017/
2023-01-19T19:12:18.239+0000    [........................] taxi.rides   5.06MB/745MB (0.7%)
2023-01-19T19:12:21.238+0000    [........................] taxi.rides   11.2MB/745MB (1.5%)
2023-01-19T19:12:24.238+0000    [........................] taxi.rides   17.4MB/745MB (2.3%)
2023-01-19T19:12:27.238+0000    [........................] taxi.rides   24.0MB/745MB (3.2%)
2023-01-19T19:12:30.239+0000    [........................] taxi.rides   30.3MB/745MB (4.1%)
2023-01-19T19:12:33.238+0000    [#.......................] taxi.rides   37.0MB/745MB (5.0%)
2023-01-19T19:12:36.238+0000    [#.......................] taxi.rides   43.9MB/745MB (5.9%)
2023-01-19T19:12:39.238+0000    [#.......................] taxi.rides   47.6MB/745MB (6.4%)
2023-01-19T19:12:42.238+0000    [#.......................] taxi.rides   51.9MB/745MB (7.0%)
2023-01-19T19:12:45.239+0000    [#.......................] taxi.rides   58.0MB/745MB (7.8%)
2023-01-19T19:12:48.239+0000    [##......................] taxi.rides   64.8MB/745MB (8.7%)
2023-01-19T19:12:51.239+0000    [##......................] taxi.rides   71.2MB/745MB (9.6%)
2023-01-19T19:19:33.239+0000    [#######################.] taxi.rides   717MB/745MB (96.3%)
2023-01-19T19:19:36.239+0000    [#######################.] taxi.rides   722MB/745MB (97.0%)
2023-01-19T19:19:39.242+0000    [#######################.] taxi.rides   727MB/745MB (97.7%)
2023-01-19T19:19:42.239+0000    [#######################.] taxi.rides   732MB/745MB (98.3%)
2023-01-19T19:19:45.239+0000    [#######################.] taxi.rides   736MB/745MB (98.9%)
2023-01-19T19:19:48.229+0000    [#######################.] taxi.rides   740MB/745MB (99.3%)
2023-01-19T19:19:50.639+0000    [########################] taxi.rides   745MB/745MB (100.0%)
2023-01-19T19:19:50.642+0000    5000000 document(s) imported successfully. 0 document(s) failed to import.
```

Step 6:

Querying data

docker-compose exec router sh -c "mongosh < /scripts/query-data.js"