HW2

PART 1. Reading Assignment [MongoDB in Action - eBook Links Provided]
Chapter 5. Constructing queries [https://www.safaribooksonline.com/library/view/mongodb-in-action/9781617291609/kindle_split_015.html]
Chapter 6. Aggregation [https://www.safaribooksonline.com/library/view/mongodb-in-action/9781617291609/kindle_split_016.html]
Chapter 7. Updates, atomic operations, and deletes
[https://www.safaribooksonline.com/library/view/mongodb-in-action/9781617291609/kindle_split_017.html]
Chapter 8. Indexing and query optimization [https://www.safaribooksonline.com/library/view/mongodb-in-action/9781617291609/kindle_split_019.html]

PART 2.1. Use the NYSE database to find the average price of stock_price_high values for each stock using MapReduce.
PART 2.2. Part 2.1 result will not be correct as AVERAGE is a commutative operation but nor associative. Use a FINALIZER to find the correct average.
Hint: pass sum and count from the reducer
   https://docs.mongodb.com/manual/reference/method/db.collection.mapReduce/index.html)
PART 3. Modify PART 2 by adding a finalizer to find out the average stock price of each price of all stocks in the finalizer.

PART 4 - 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.

PART 5 - 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.

PART 6 - Programming Assignment
MapReduce to find each of the followings:
   Task 1. MapReduce to find the top 25 rated movies in the movieLens dataset
   Task 2. MapReduce to find the number of males and females in the movielens dataset
   Task 3. MapReduce to find the number of movies rated by different users
PART 7 - Programming Assignment
Install Hadoop and run each of the following File System Shell commands, and place the screenshots in a word document after running each command.
https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/FileSystemShell.html