QM 7093: Enterprise Data Systems: NoSQL with MongoDB

Noah L. Schrick

13 December, 2022

Contents

1 MongoDB

Due to the flexible nature of the project assignment and the ability to take alternate database approaches, MongoDB was chosen as the database implementation. While both MongoDB and CouchDB are document-based NoSQL databases, they each have differing advantages. My primary research focuses and interests revolve around the High-Performance Computing (HPC) space, and MongoDB sees greater usage in this area. MongoDB has greater scalability and better performance than CouchDB, though it does lack the design priorities of availability that CouchDB offers. MongoDB is also one of the most-widely used databases across all models, ranking at position 5 on `https://db-engines.com/en/ranking?utm_source=xp&utm_medium=blog&utm_campaign=content`.

2 Insertions and Queries

2.1 Inserting Data

Insert all records from the provided datasheet with the following properties:

- Create all the records with columns StockCode, Description, Quantity, Price, Customer ID and Country (that means you should NOT include invoice and invoice Date in your columns).
- The code for records without customer ID should NOT have a customer ID column.
- Create another column "HighDemand" but ONLY for records with Quantity more than 12 (12 included). In the column put "Yes".

To reduce the amount of manual insertions and minimize the risk of human insertion error, the xlsx datasheet was converted to a csv. Each text cell ("Description", for example) was encapsulated in quotes before the conversion. The delimiter used was a comma (","). Saving the data in a csv format allows for an easy insertion by MongoDB using mongoimport.

```
mongoimport --db QM_7093_Final --headerline --file Project_Data.csv
--type csv
```

Removing Invoice and InvoiceDate from the Project_Data collection can be performed with:

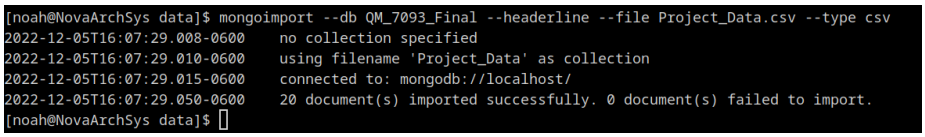

Figure 1: Part 1.a: Importing from CSV

```
db.Project_Data.updateMany({}, {$unset: { "Invoice": "", "InvoiceDate":
""}} )
```

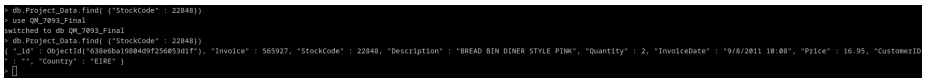

Figure 2: Collection Sample Prior to Removing Invoice and Invoice Date

Figure 3: Collection Sample After Removing Invoice and Invoice Date

Removing the CutomerID field when empty can be performed with:

```
db.Project_Data.updateMany({"CustomerID" : ""}, { $unset : {"CustomerID"
: 1 } } )
```

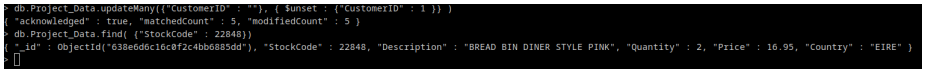

Figure 4: Collection Sample After Removing Empty CustomerID fields

2.2 Queries

Question 1: How many records have the column "HighDemand"? (Must have a code to answer this, one way to answer this is to have a code that displays all the records except those with the column HighDemand and then subtract the number from total number of records)

Question 2: Display the records with price more than 4 (4 excluded)

3 Metadata