Week12_Assignment

Umais Siddiqui

November 16, 2017

Rpubs: http://rpubs.com/umais/week12_assignment

Github: https://github.com/umais/DATA-607/tree/master/Week12_Assignment

Assignment Week 12

In this assignment I will be taking data from a relational database and migrating it to MongoDB database that is set up on MongoLab. The data set that I have chosen is from Project 3 . I am going to retrieve the Job Title, Job Location and Skill Name from the Normalized tables DataScienceJob and DataScienceSkills and store them in a mongo database collection called jobskills.

Reading the Data from MySQL database and migrating it to MongoDB No SQL Storage

In this step I will be using the RMySQL library to connect to MySQL data store and retrieve the Skills data. I will then connect to the MongoDB database called openletter on MongoLab and Insert the Skill data reteived from MySQL in the collection called jobskills.

```
my_collection = mongo(collection = "jobskills", url = "mongodb://ruser:Tw1stone23@ds031618.mlab.com:316
mydb = dbConnect(MySQL(), user='root', password='Welcome@1', dbname='project3', host='localhost')

results = dbSendQuery(mydb, "SELECT j.JobTitle,j.JobLocation,s.SkillName FROM DataScienceJobs j INNER J

jobSkills=fetch(results, n=-1)
my_collection$drop()
```

```
## [1] TRUE
```

```
my_collection$insert(jobSkills)
```

```
##
Processed 1000 rows...
Complete! Processed total of 1643 rows.
```

```
## $nInserted
## [1] 1643
##
## $nMatched
## [1] 0
##
## $nRemoved
## [1] 0
##
## $nUpserted
## [1] 0
##
## $writeErrors
## list()
```

```
head(jobSkills)
```

```
##           JobTitle         JobLocation           SkillName
## 1  Data Scientist       Sunnyvale, CA              Python
## 2  Data Scientist       Sunnyvale, CA               C/C++
## 3  Data Scientist       Sunnyvale, CA        Apache Spark
## 4  Data Scientist       Sunnyvale, CA               Kafka
## 5  Data Scientist       Sunnyvale, CA       ElasticSearch
## 6  Data Scientist   San Francisco, CA   Postgres/Redshift
```

Reading data from the MongoDB collection to ensure it got Saved correctly.

```
alldata <- my_collection$find('{}')
```

```
##
 Found 1000 records...
 Found 1643 records...
 Imported 1643 records. Simplifying into dataframe...
```

```
head(alldata)
```

```
##           JobTitle         JobLocation           SkillName
## 1  Data Scientist       Sunnyvale, CA              Python
## 2  Data Scientist       Sunnyvale, CA               C/C++
## 3  Data Scientist       Sunnyvale, CA        Apache Spark
## 4  Data Scientist       Sunnyvale, CA               Kafka
## 5  Data Scientist       Sunnyvale, CA       ElasticSearch
## 6  Data Scientist   San Francisco, CA   Postgres/Redshift
```

Pros and Cons of Storing data in Relational DB vs NoSQL

1) The advanatage of storing data in a relational database is that data can be stored in a relational manner that is good for data integrity and elimnates alot of anomolies.
2) The advantage of storing data in a NoSQL store is that it is easier to retrieve and the data access is faster as the data is stored in a manner of document.