Using Python MongoDB Package - pymongo

lesson_3_2_1

Information for installing and remote access gathered form:

Information on installation is taken from MongoDB docs.

Information on remote access found at How to connect to your remote MongoDB server

Installing MongoDB on the Playground Server

Information on installation is taken from MongoDB docs.
Information on remote access found at How to connect to your remote MongoDB server

Import the GPG key from MongoDB

- `wget -qO - https://www.mongodb.org/static/pgp/server-4.2.asc | sudo apt-key add -`

Create a list file for MongoDB

- `echo "deb [ arch=amd64,arm64 ] https://repo.mongodb.org/apt/ubuntu bionic/mongodb-org/4.2 multiverse" | sudo tee /etc/apt/sources.list.d/mongodb-org-4.2.list`

Reload package database:

- `sudo apt update`

Install MongoDB

`-sudo apt install -y mongodb-org`

Start the MongoDB Service

- `sudo systemctl start mongod`

This starts the mongodb service as needed. To have the mongodb service start everytime you start the server:

- `sudo systemctl enable mongod`

I will only start the service when I need to, but you have the option.

Create User, Database and Grant Access

Create User

- At the terminal:

```
mongo
```

Now at the mongo terminal:

```
use cloud_user

db.createUser({
    user: 'cloud_user',
    pwd: 'cloud_user',
    roles: [{ role: 'readWrite', db:'cloud_user'}]
})
```

- leave the mongo shell with `CTRL-c`

Create Collection From File

A special thanks to MongoDB for the zips dataset

Import collection:

`-mongoimport -v --db=cloud_user --file=/home/cloud_user/python_data_course/data/zips.json`

Install MongoDB Driver to Your Virtual Environment

- `conda activate python_data_course
- `conda install pymongo`

Start and connect to the Jupyter Notebook server as usual.

Using Python to Connect to Server

Imports and DB Connection

```
from pymongo import MongoClient
import pandas as pd

client =
MongoClient("mongodb://cloud_user:cloud_user@localhost:27017/cloud_user")
```

```
df = pd.DataFrame.from_records(client.cloud_user.zips.find())
```

```
df.head()
```

```
oz_data = {"_id": 99990, "city": "EMERALD", "loc": [-510.9, -600.89], "pop":
564372, "state": "Munchkin Land"}
```

```
client.cloud_user.zips.insert_one(oz_data)
```

```
df1 = pd.DataFrame.from_records(client.cloud_user.zips.find())
```

```
df1.tail()
```

```
df.count()
```

```
df1.count()
```

```
client.close()
```