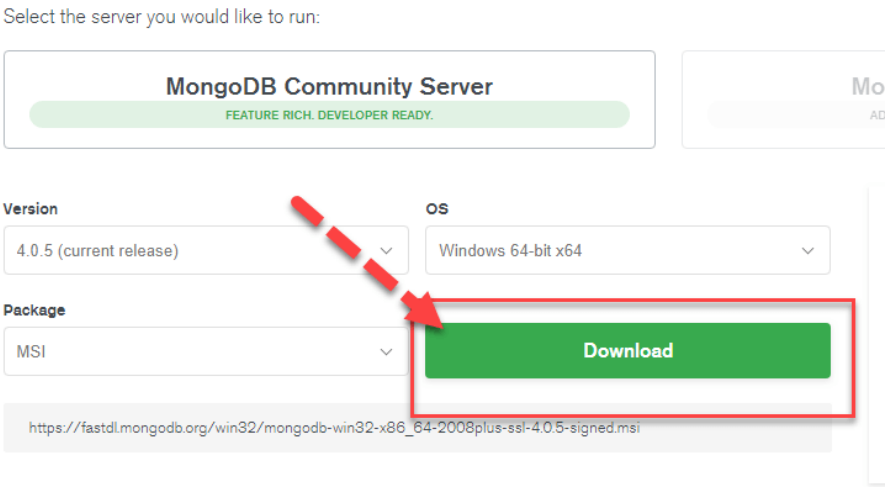
Aim : Implement an application that stores big data in Hbase / MongoDB and manipulate it using R / Python

Requirement

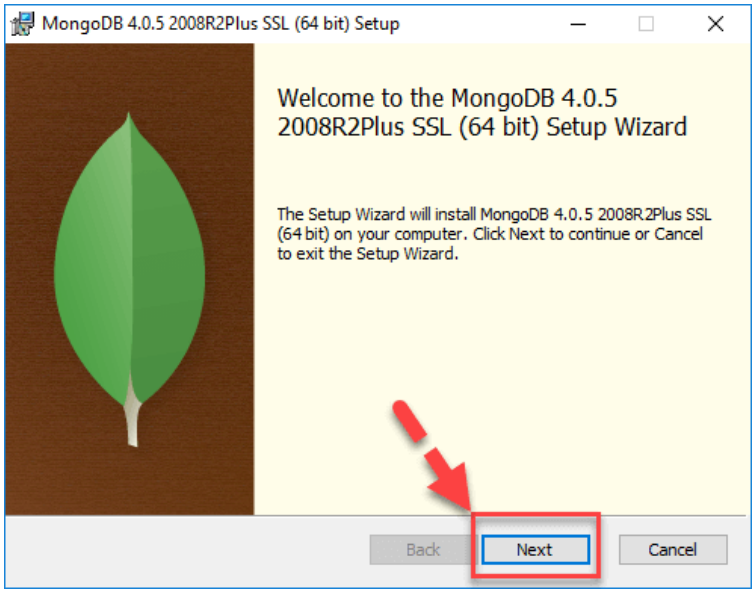
a. PyMongo
b. Mongo Database

Step A: Install Mongo database

Step 1) Go to (https://www.mongodb.com/download-center/community) and Download MongoDB Community Server. We will install the 64-bit version for Windows.

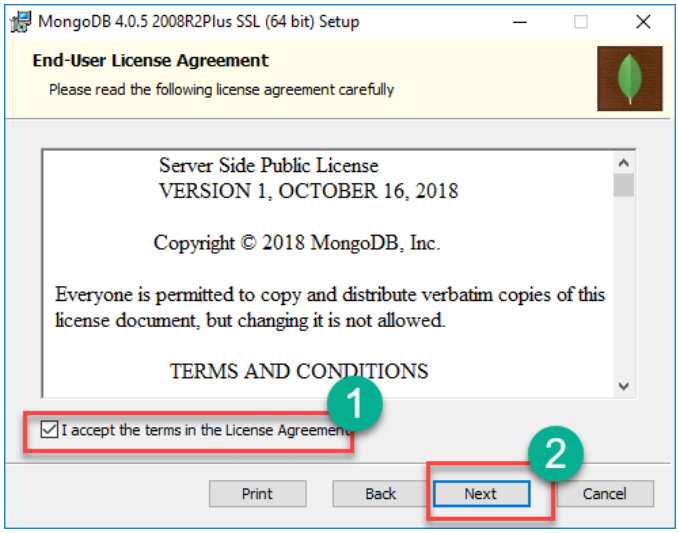


Step 2) Once download is complete open the msi file. Click Next in the start up screen

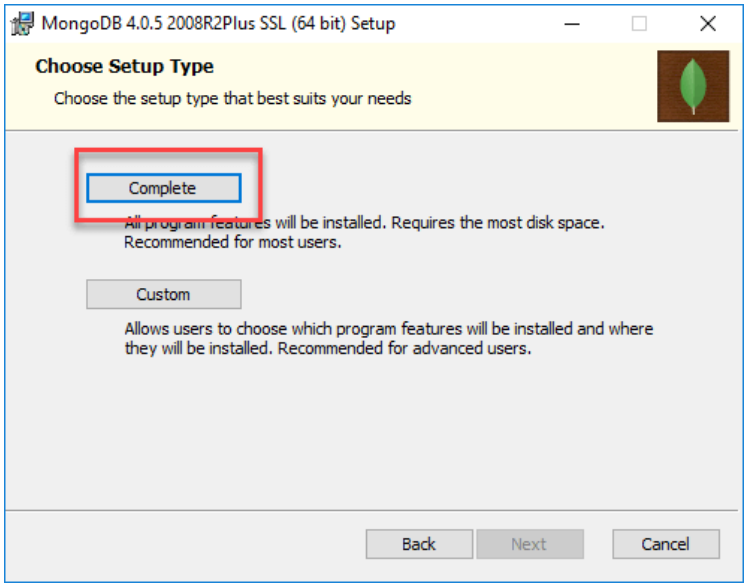


Step 3)

1. Accept the End-User License Agreement
2. Click Next

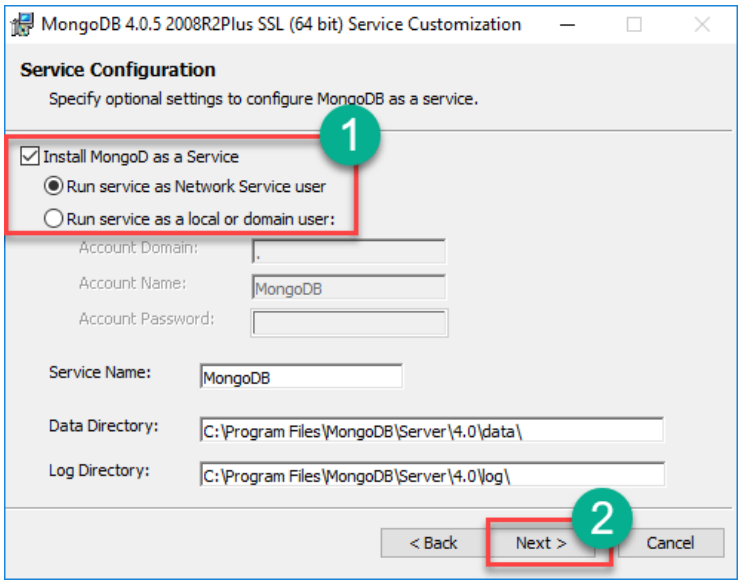


Step 4) Click on the "complete" button to install all of the components. The custom option can be used to install selective components or if you want to change the location of the installation.

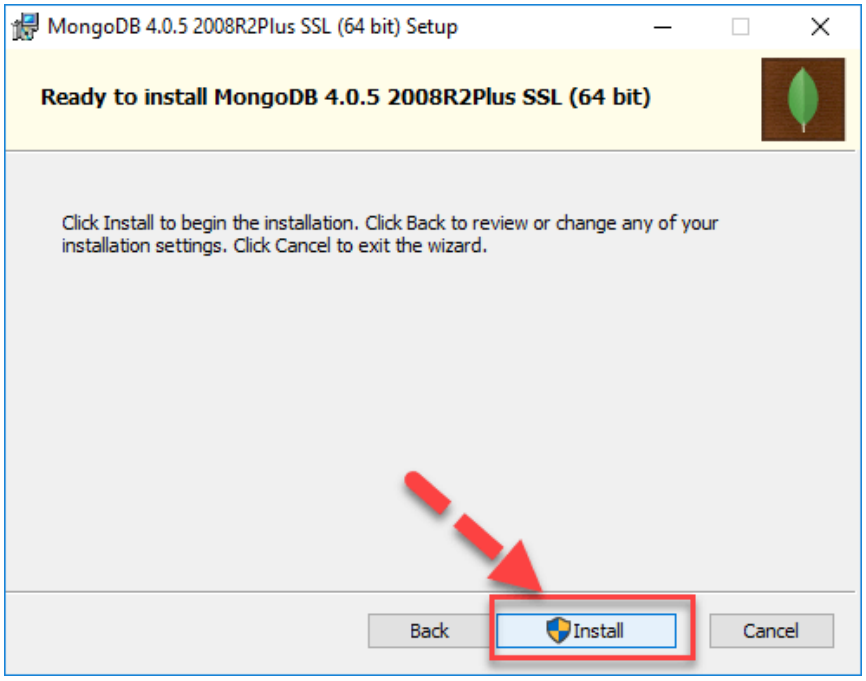


Step 5)

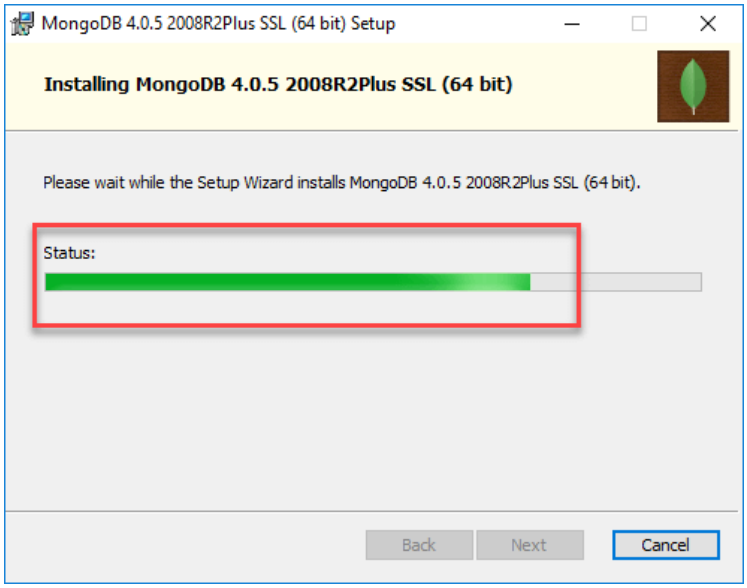
1. Select “Run service as Network Service user”. make a note of the data directory, we’ll need this later.
2. Click Next



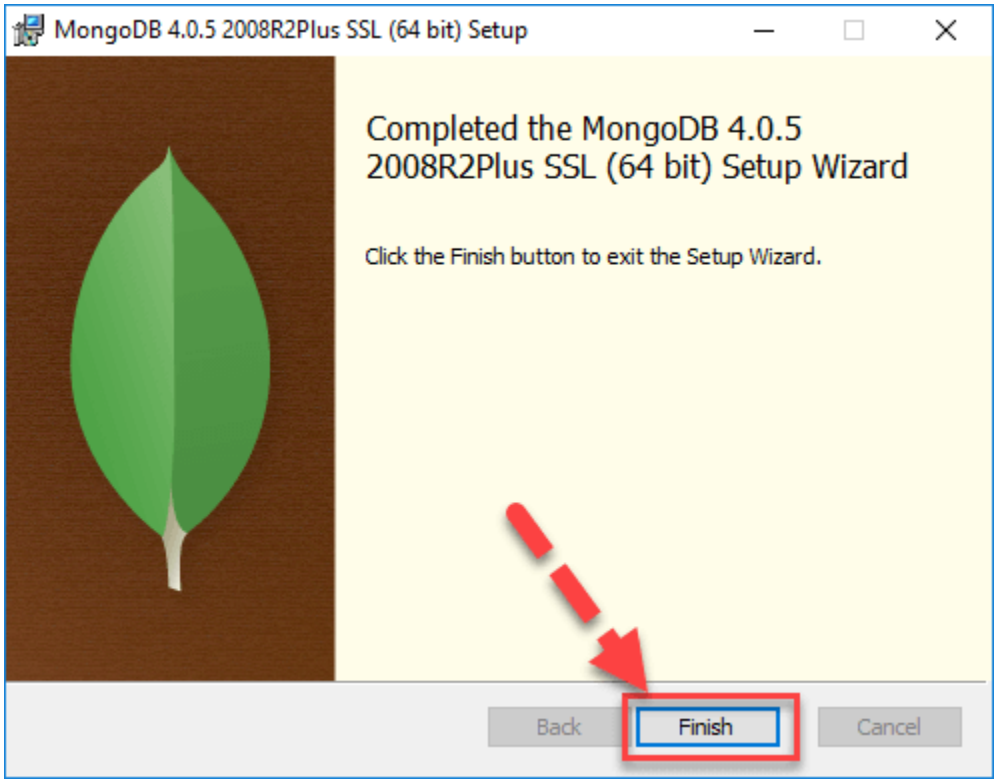
Step 6) Click on the Install button to start the installation.



Step 7) Installation begins. Click Next once completed



Step 8) Click on the Finish button to complete the installation



Test Mongodb

Step 1) Go to " C:\Program Files\MongoDB\Server\4.0\bin" and double click on mongo.exe. Alternatively, you can also click on the MongoDB desktop item

- **Create the directory where MongoDB will store it's files.** From the command prompt run `md \data\db`. This is the default location. However, other locations can be specified using the `--dbpath` parameter. See the Mongo docs for more information.
    - C:\>md data
    - C:\md data\db
    - C:\Program Files\MongoDB\Server\4.05\bin>mongod.exe --dbpath "C:\data"
- **Start the mongodb daemon** by running `C:\mongodb\bin\mongod.exe` in the Command Prompt. Or by running, `C:\path\to\mongodb\bin\mongod.exe`
- **Connect to MongoDB using the Mongo shell** While the MongoDB daemon is running, from a different Command prompt window run `C:\mongodb\bin\mongo.exe`
    - C:\Program Files\MongoDB\Server\4.05\bin>mongod.exe --dbpath "C:\data"
    - C:\Program Files\MongoDB\Server\4.05\bin>mongo.exe

Step B: Install PyMongo

C:\Users*Your Name*\AppData\Local\Programs\Python\Python36-32\Scripts>python -m pip install pymongo

Now you have downloaded and installed a mongoDB driver.

Test PyMongo

demo_mongodb_test.py:

import pymongo

Program 1: Creating a Database

```
import pymongo
myclient = pymongo.MongoClient("mongodb://localhost:27017/")
mydb = myclient["mybigdata"]
print(myclient.list_database_names())
```

Progam 2: Creating a Collection

```
import pymongo
myclient = pymongo.MongoClient("mongodb://localhost:27017/")
mydb = myclient["mybigdata"]
mycol=mydb[“student’]
print(mydb.list_collection_names())
```

Progam 3: Insert into Collection

```
import pymongo
myclient = pymongo.MongoClient("mongodb://localhost:27017/")
mydb = myclient["mybigdata"]
mycol=mydb[“student’]
mydict={“name”:”Kaushal”, “address”:”Mumbai”}
x=mycol.insert_one(mydict) # insert_one(containing the name(s) and value(s) of each field
```

Program 4: Insert Multiple data into Collection

```
import pymongo
myclient = pymongo.MongoClient("mongodb://localhost:27017/")
mydb = myclient["mybigdata"]
mycol=mydb[“student’]
mylist=[{“name”:”Kaushal”, “address”:”Mumbai”}, {“name”:”A”, “address”:”Mumbai”}, {“name”:”B”, “address”:”Pune”}, {“name”:”C”, “address”:”Pune”},]
x=mycol.insert_many(mylist)
```

Test in Mongodb to check database and data inserted in collection

a. If you want to check your database list, use the command **show dbs** in mongo command prompt

b. If you want to use a database with name mybigdata, then use database statement would be as follow: **use mybigdata**

c. If you want to check collection in mongodb use the command **show collections**

d. If you want to display all the data from collection: **db.collection_name.find()** or **db.collection_name.find().pretty()**