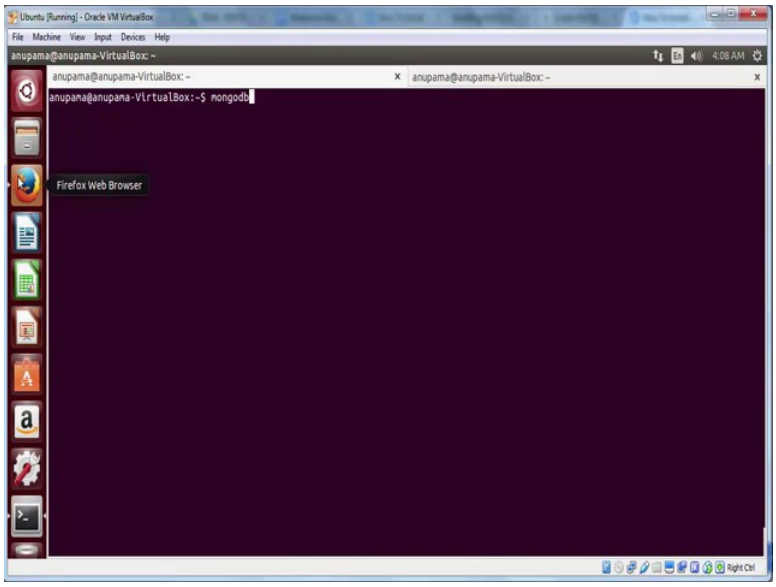
Privacy and Security in Online Social Networks
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

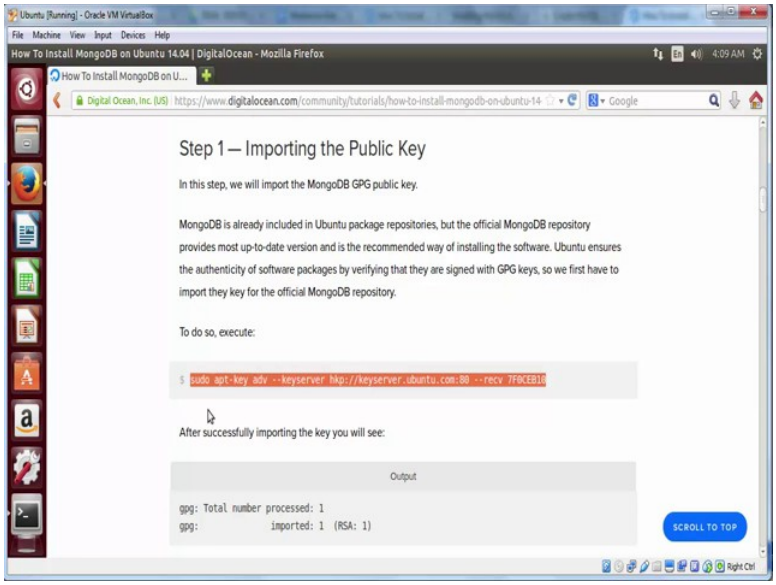
Lecture – 14
Tutorial 3, Part 3 MongoDB

(Refer Slide Time: 00:13)



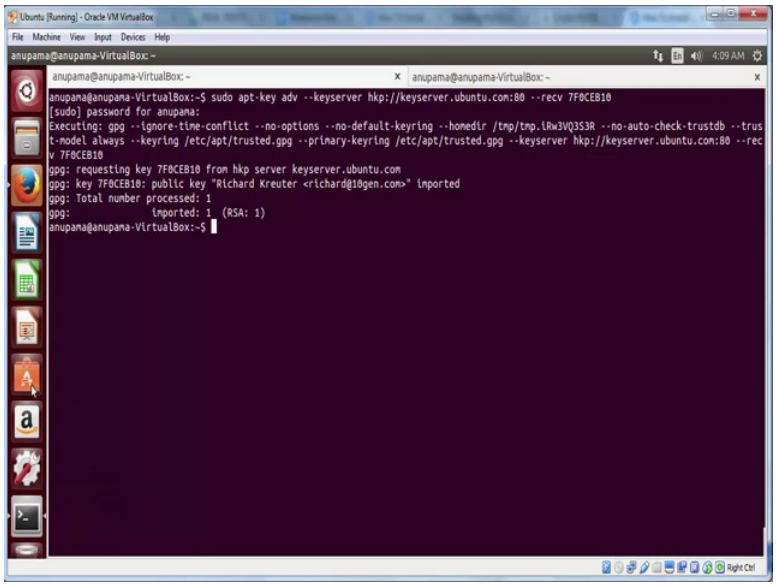
Now, we will look at another database for storing our collected data, MongoDB. MongoDB is an open source non relational document based data base.

(Refer Slide Time: 00:26)

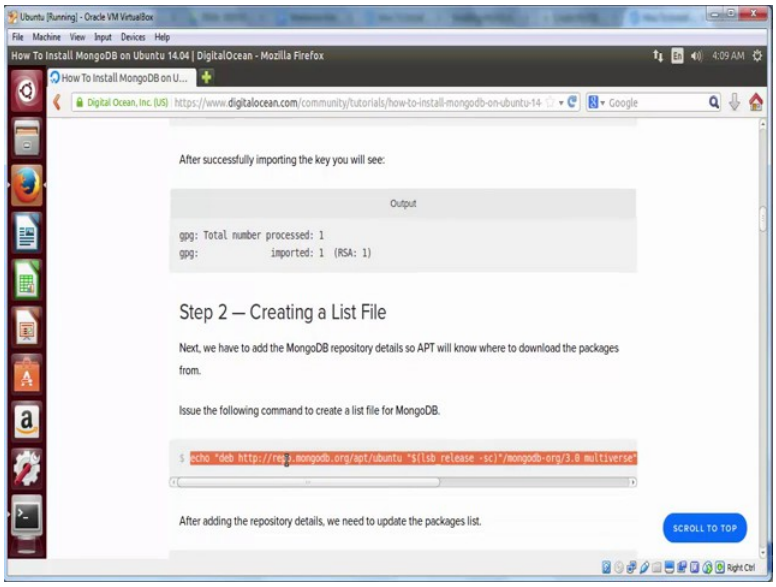


We are going to follow this tutorial to install MongoDB. We will begin by importing the MongoDB GPG public key. To do that let us copy paste this following command into a terminal.

(Refer Slide Time: 00:44)

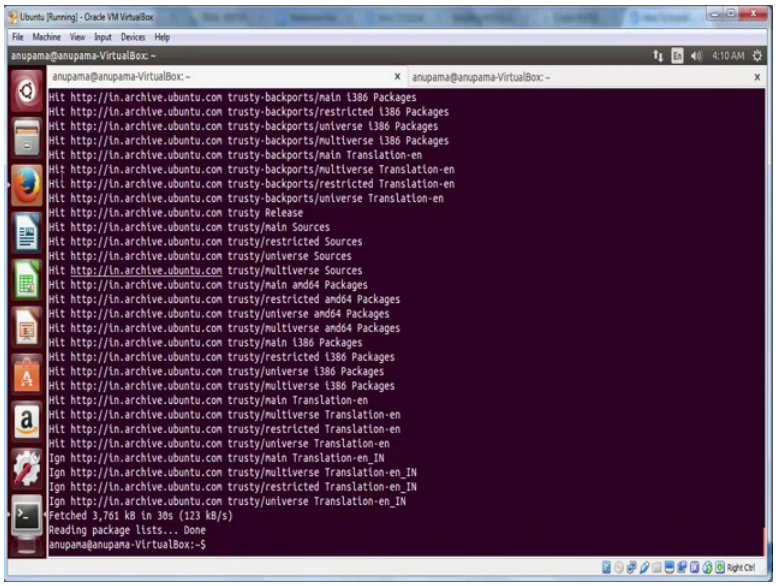


(Refer Slide Time: 01:00)



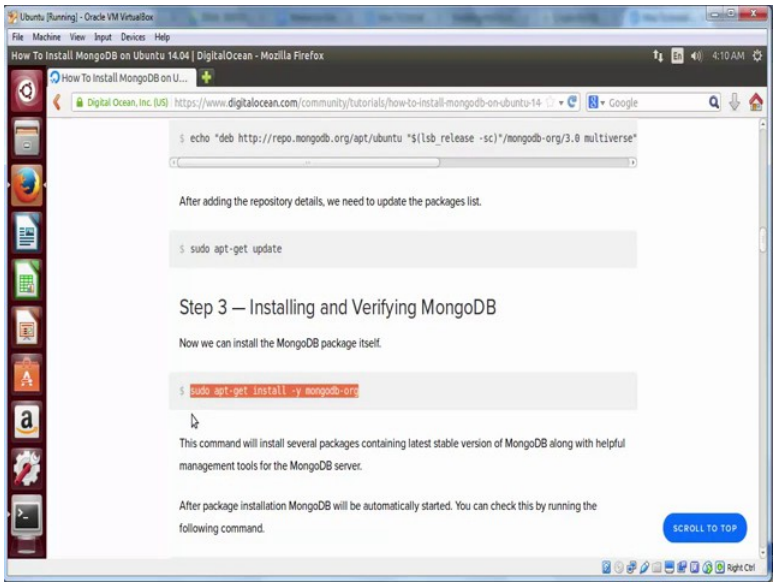
Next we are going to create a file list, that is, we will add the MongoDB repository details to the apt and then, update the package list. Now, to update the package list let us type the command sudo apt, get update and press enter.

(Refer Slide Time: 01:16)



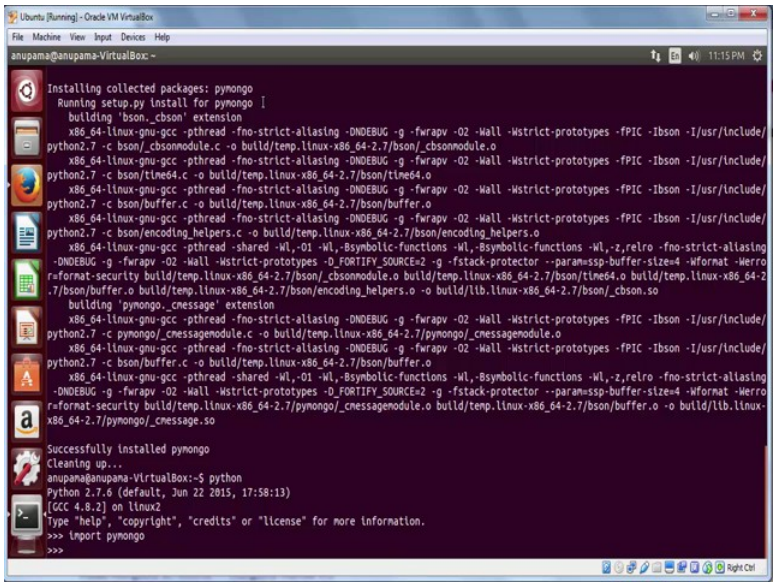
After this step, we will install the MongoDB package.

(Refer Slide Time: 01:38)



To do that use the command sudo apt-get install -y MongoDB-org and copy paste it in your terminal.

(Refer Slide Time: 01:49)

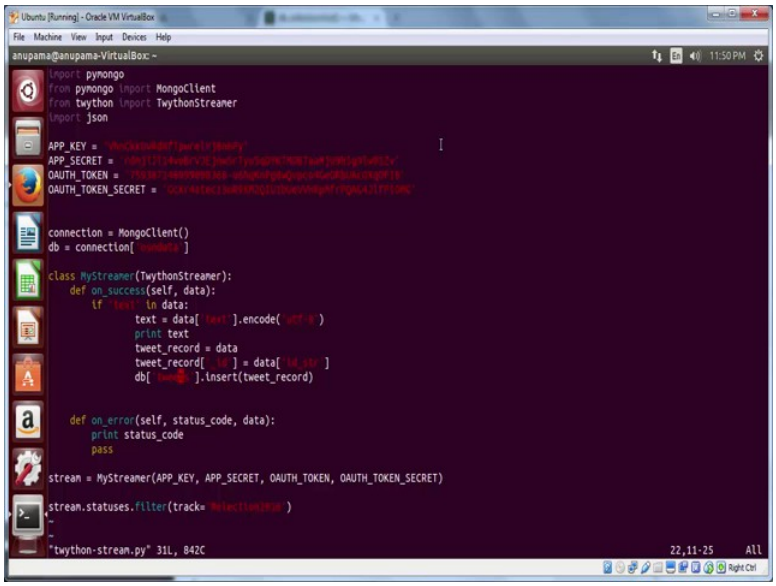


Now, let us check whether MongoDB has been successfully installed or not. Type mongo in your terminal and press enter. It should give you MongoDB console. In MongoDB, you can less the existing data bases by using the command show dbs. Now, let us create a database using the command use and the database name. We will use osndata as our database name. Press enter. It will give a message switched to db osndata.

Now, let us try to again list databases using the previous command. You will notice that the freshly created database does not appear in the list. That is because it does not have any data stored yet. Before we insert data into a MongoDB, let us first install pymongo which acts like a connector between MongoDB and python. Exit mongo shell by using Ctrl+C and then, type sudo pip install pymongo and press enter. This will install pymongo. Now, let us verify whether pymongo has been successfully installed or not. To do that, start python and enter import pymongo. If it successfully gets imported, pymongo has been installed.

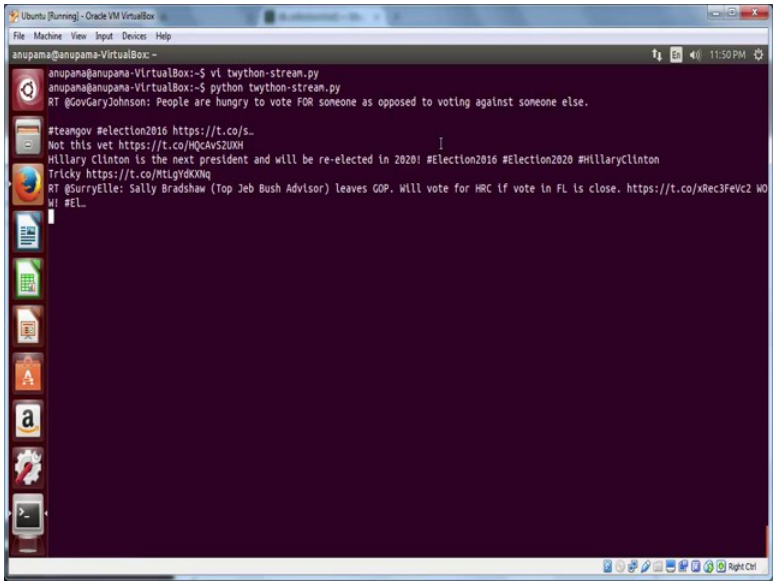
Now, let us go back to our previous example of tweet collection from Twitter Streaming API.

(Refer Slide Time: 04:12)



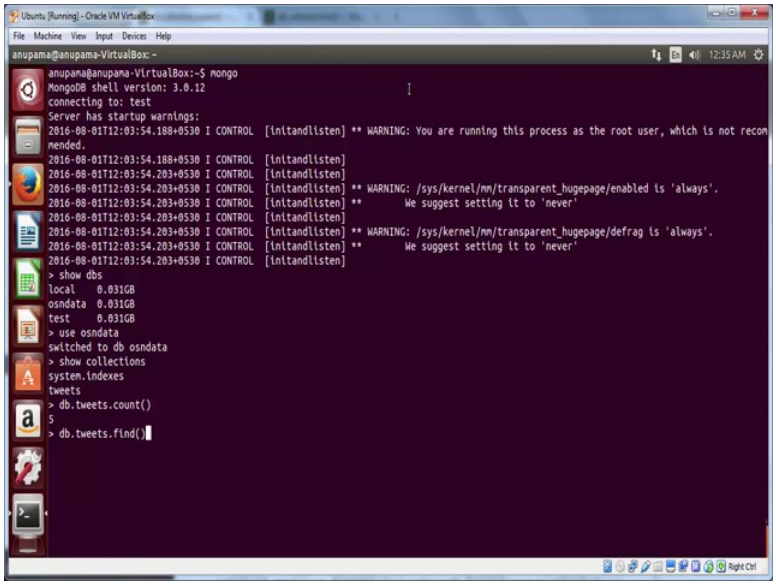
We have two import commands, pymongo, and MongoClient from pymongo. Connection to the database is defined by connection equal to MongoClient, and in the next line we define the database which we will be accessing in this connection. The rest of the program is same except that this time after fetching the tweet, we will store the entire tweet object in JSON format in MongoDB. Note that data variable is the tweet object which you want to insert. Additionally, set another key underscore id to the tweet id. Underscore id is used as a unique identifier for every record in MongoDB. Now, let us save this file and run it.

(Refer Slide Time: 05:03)



You will soon see the data starting to come in for #election2016. Now, let us see whether this data is getting stored in the database or not.

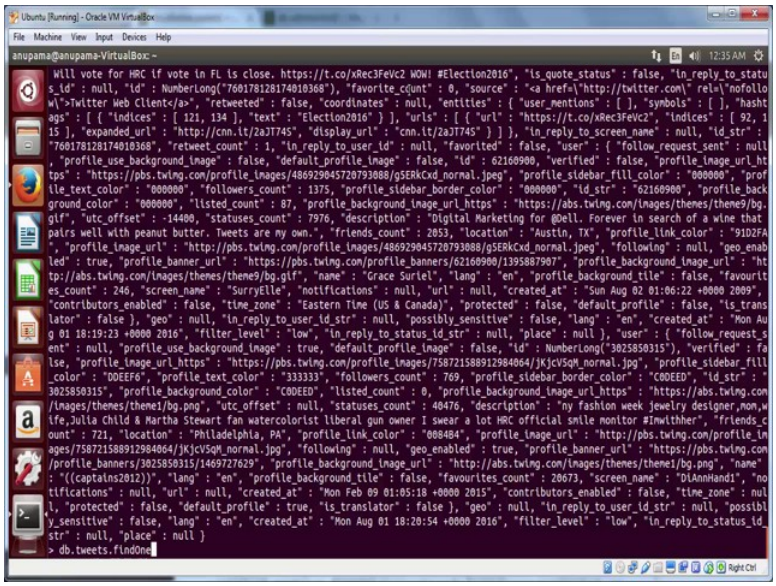
(Refer Slide Time: 05:30)



Let us start the mongo shell by typing mongo. Let us again list the existing databases by entering show dbs. This time you will note osndata listed. Recall that before data insertion, it was not showing up which means that it now holds some data. Let us verify this further. Switch to osn database by using osndata. Then, enter show collections. Now, you will be able to see two collections, system.indexes and twitter.

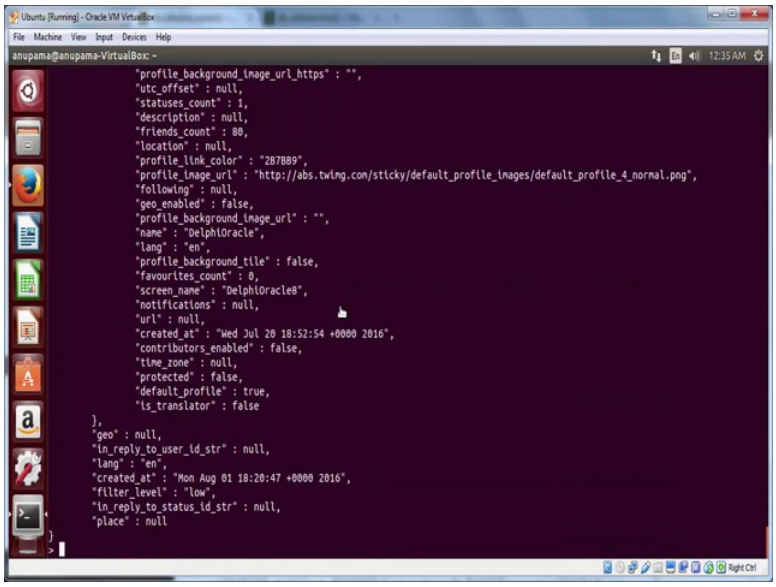
system.indexes is a default collection by mongo and tweets collection is the one which we have just created and it has data stored. Let us look at few basic commands to explore this data in tweet collection. Let us count the number of tweets stored by entering db.tweets.count(). Here tweet is the name of the collection and the output in our case is 5. Let us also look at the data stored by using the command db.tweets.find().

(Refer Slide Time: 07:03)



Notice that the data stored is in JSON format. However, it looks all jumbled up. Therefore, to print a well formatted singular record, we can use the command db.tweets.findOne().

(Refer Slide Time: 07:25)



Now, you can clearly see a well formatted tweet object. So, now we know how to collect OSN data via Facebook API, Twitter API and store the data either into a relational database that is MySQL or a non-relational database that is MongoDB.