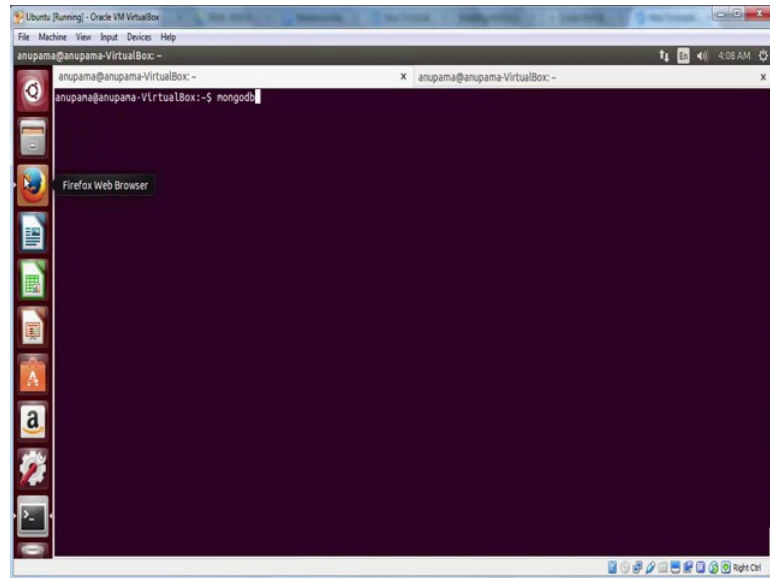


**Privacy and Security in Online Social Networks**  
**Department of Computer Science and Engineering**  
**Indian Institute of Technology, Madras**

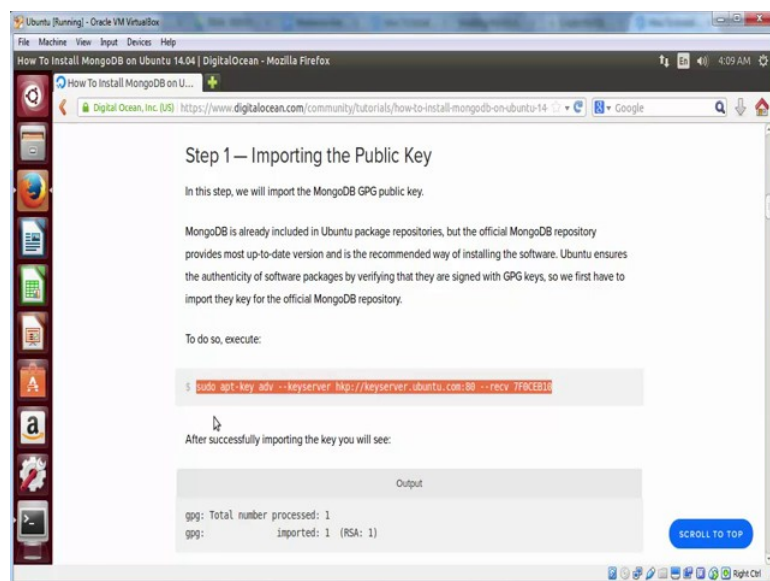
**Lecture – 14**  
**Tutorial 3, Part 3 MongoDB**

(Refer Slide Time: 00:13)



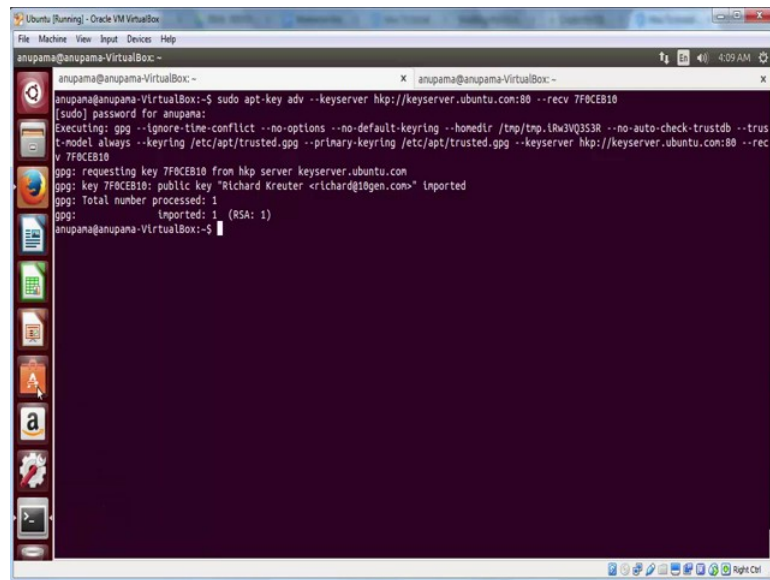
Now, we will look at another database for storing our collected data, MongoDB. MongoDB is an open source non relational document based data base.

(Refer Slide Time: 00:26)



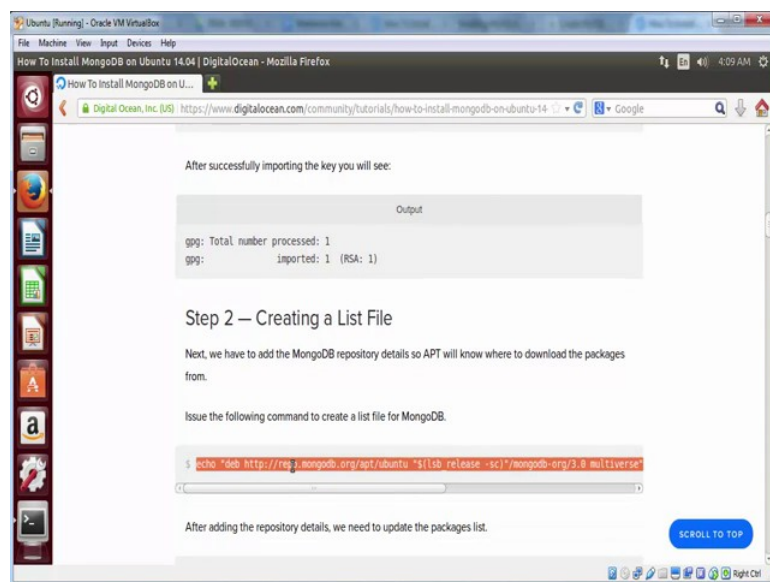
We are going to follow this tutorial to install MongoDB. We will begin by importing the MongoDB GPG public key. To do that let us copy paste this following command into a terminal.

(Refer Slide Time: 00:44)



```
anupama@anupama-VirtualBox:~$ sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv 7F0CEB10
[sudo] password for anupama:
Executing: gpg --ignore-time-conflict --no-options --no-default-keyring --homedir /tmp/.tmptmp.LW3VQ353R --no-auto-check-trustdb --trust-model always --keyring /etc/apt/trusted.gpg --primary-keyring /etc/apt/trusted.gpg --keyserver hkp://keyserver.ubuntu.com:80 --recv 7F0CEB10
gpg: requesting key 7F0CEB10 from hkp server keyserver.ubuntu.com
gpg: key 7F0CEB10: public key "Richard Kreuter <richard@10gen.com>" imported
gpg: Total number processed: 1
gpg:      imported: 1 (RSA: 1)
anupama@anupama-VirtualBox:~$
```

(Refer Slide Time: 01:00)



How To Install MongoDB on Ubuntu 14.04 | DigitalOcean - Mozilla Firefox

Digital Ocean, Inc. (US) <https://www.digitalocean.com/community/tutorials/how-to-install-mongodb-on-ubuntu-14-04>

After successfully importing the key you will see:

```
Output
gpg: Total number processed: 1
gpg:      imported: 1 (RSA: 1)
```

Step 2 — Creating a List File

Next, we have to add the MongoDB repository details so APT will know where to download the packages from.

Issue the following command to create a list file for MongoDB.

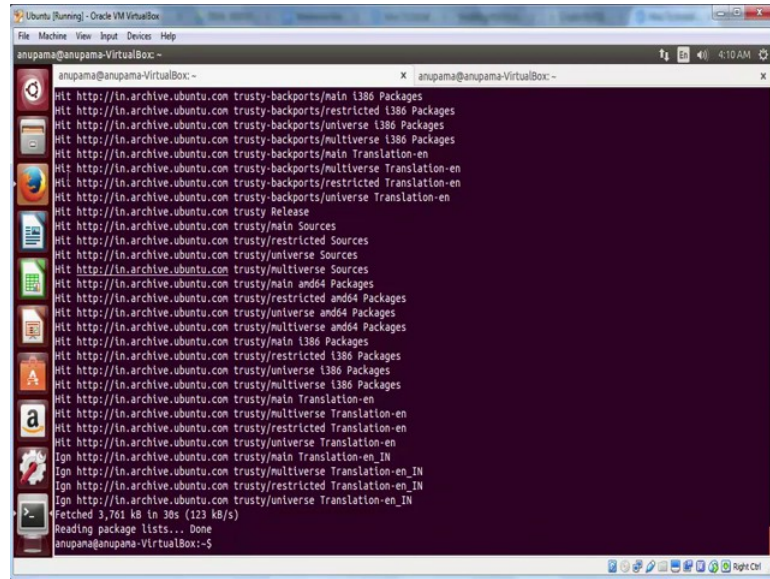
```
$ echo "deb http://rep.mongodb.org/apt/ubuntu \"$lsb.release -sc\"/mongodb-org/3.0 multiverse" > /etc/apt/sources.list.d/mongodb-org-3.0.list
```

After adding the repository details, we need to update the packages list.

SCROLL TO TOP

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.

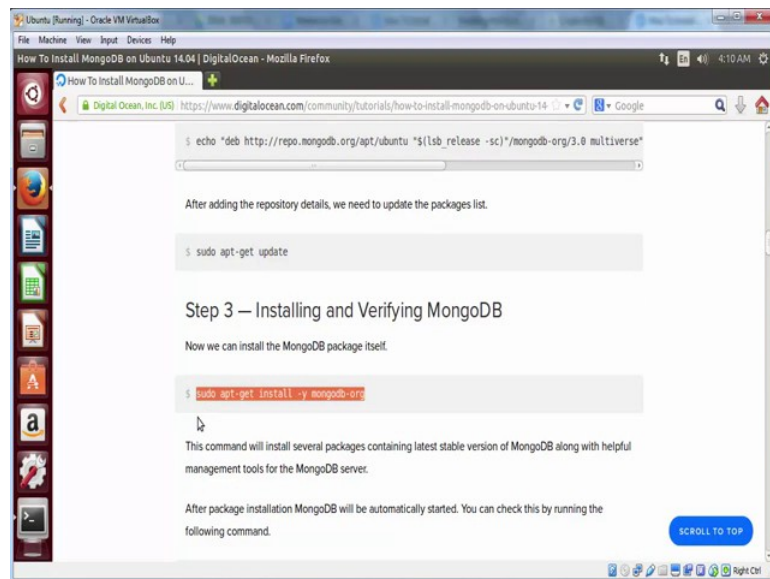
(Refer Slide Time: 01:16)



```
anupama@anupama-VirtualBox:~$ apt-get update
Hit http://ln.archive.ubuntu.com trusty-backports/main 1386 Packages
Hit http://ln.archive.ubuntu.com trusty-backports/restricted 1386 Packages
Hit http://ln.archive.ubuntu.com trusty-backports/universe 1386 Packages
Hit http://ln.archive.ubuntu.com trusty-backports/multiverse 1386 Packages
Hit http://ln.archive.ubuntu.com trusty-backports/main Translation-en
Hit http://ln.archive.ubuntu.com trusty-backports/multiverse Translation-en
Hit http://ln.archive.ubuntu.com trusty-backports/restricted Translation-en
Hit http://ln.archive.ubuntu.com trusty-backports/universe Translation-en
Hit http://ln.archive.ubuntu.com trusty Release
Hit http://ln.archive.ubuntu.com trusty/main Sources
Hit http://ln.archive.ubuntu.com trusty/restricted Sources
Hit http://ln.archive.ubuntu.com trusty/universe Sources
Hit http://ln.archive.ubuntu.com trusty/multiverse Sources
Hit http://ln.archive.ubuntu.com trusty/main amd64 Packages
Hit http://ln.archive.ubuntu.com trusty/restricted amd64 Packages
Hit http://ln.archive.ubuntu.com trusty/universe amd64 Packages
Hit http://ln.archive.ubuntu.com trusty/multiverse amd64 Packages
Hit http://ln.archive.ubuntu.com trusty/main i386 Packages
Hit http://ln.archive.ubuntu.com trusty/restricted i386 Packages
Hit http://ln.archive.ubuntu.com trusty/universe i386 Packages
Hit http://ln.archive.ubuntu.com trusty/multiverse i386 Packages
Hit http://ln.archive.ubuntu.com trusty/main Translation-en
Hit http://ln.archive.ubuntu.com trusty/restricted Translation-en
Hit http://ln.archive.ubuntu.com trusty/universe Translation-en
Ign http://ln.archive.ubuntu.com trusty/main Translation-en_IN
Ign http://ln.archive.ubuntu.com trusty/multiverse Translation-en_IN
Ign http://ln.archive.ubuntu.com trusty/restricted Translation-en_IN
Ign http://ln.archive.ubuntu.com trusty/universe Translation-en_IN
Fetched 3,761 kB in 30s (123 kB/s)
Reading package lists... Done
anupama@anupama-VirtualBox:~$
```

After this step, we will install the MongoDB package.

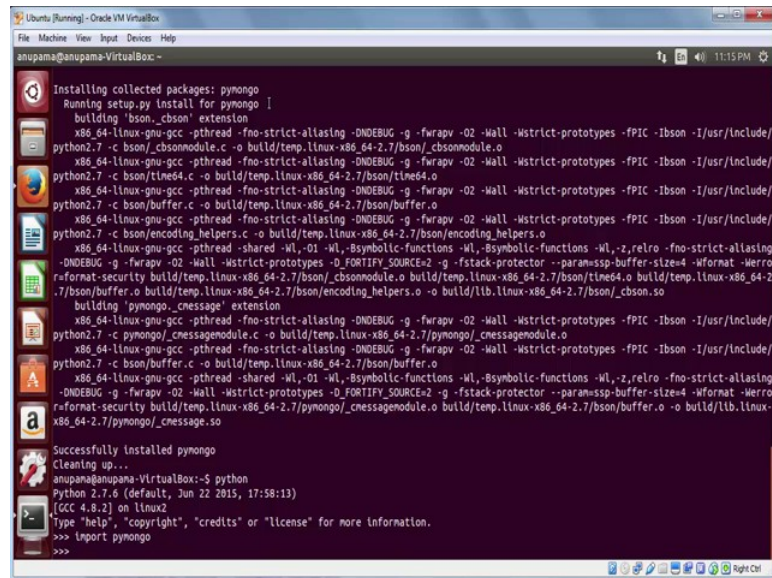
(Refer Slide Time: 01:38)



```
How To Install MongoDB on Ubuntu 14.04 | DigitalOcean - Mozilla Firefox
Digital Ocean, Inc. (US) https://www.digitalocean.com/community/tutorials/how-to-install-mongodb-on-ubuntu-14-04
$ echo "deb http://repo.mongodb.org/apt/ubuntu "$(lsb_release -sc)"/mongodb-org/3.0 multiverse"
After adding the repository details, we need to update the packages list.
$ sudo apt-get update
Step 3 — Installing and Verifying MongoDB
Now we can install the MongoDB package itself.
$ sudo apt-get install -y mongodb-org
This command will install several packages containing latest stable version of MongoDB along with helpful management tools for the MongoDB server.
After package installation MongoDB will be automatically started. You can check this by running the following command.
```

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)



```
anupama@anupama-VirtualBox:~$ pip install pymongo
Installing collected packages: pymongo
  Running setup.py install for pymongo
    building 'bson' extension
      x86_64-linux-gnu-gcc -pthread -fno-strict-aliasing -DDEBUG -g -fwrapv -O2 -Wall -Wstrict-prototypes -fPIC -I/usr/include/python2.7 -c bson/_bsonmodule.c -o build/temp.linux-x86_64-2.7/bson/_bsonmodule.o
      x86_64-linux-gnu-gcc -pthread -fno-strict-aliasing -DDEBUG -g -fwrapv -O2 -Wall -Wstrict-prototypes -fPIC -I/usr/include/python2.7 -c bson/timedelta.c -o build/temp.linux-x86_64-2.7/bson/timedelta.o
      x86_64-linux-gnu-gcc -pthread -fno-strict-aliasing -DDEBUG -g -fwrapv -O2 -Wall -Wstrict-prototypes -fPIC -I/usr/include/python2.7 -c bson/buffer.c -o build/temp.linux-x86_64-2.7/bson/buffer.o
      x86_64-linux-gnu-gcc -pthread -fno-strict-aliasing -DDEBUG -g -fwrapv -O2 -Wall -Wstrict-prototypes -fPIC -I/usr/include/python2.7 -c bson/encoding_helpers.c -o build/temp.linux-x86_64-2.7/bson/encoding_helpers.o
      x86_64-linux-gnu-gcc -pthread -shared -Wl,-O1 -Wl,-Bsymbolic-functions -Wl,-Bsymbolic-functions -Wl,-z,relro -fno-strict-aliasing -DDEBUG -g -fwrapv -O2 -Wall -Wstrict-prototypes -D_FORTIFY_SOURCE=2 -g -fstack-protector --param=ssp-buffer-size=4 -Wformat -Werror=format-security build/temp.linux-x86_64-2.7/bson/_bsonmodule.o build/temp.linux-x86_64-2.7/bson/timedelta.o build/temp.linux-x86_64-2.7/bson/buffer.o build/temp.linux-x86_64-2.7/bson/encoding_helpers.o -o build/lib.linux-x86_64-2.7/bson/_bson.so
    building 'pymongo' extension
      x86_64-linux-gnu-gcc -pthread -fno-strict-aliasing -DDEBUG -g -fwrapv -O2 -Wall -Wstrict-prototypes -fPIC -I/usr/include/python2.7 -c pymongo/_cmessagemodule.c -o build/temp.linux-x86_64-2.7/pymongo/_cmessagemodule.o
      x86_64-linux-gnu-gcc -pthread -fno-strict-aliasing -DDEBUG -g -fwrapv -O2 -Wall -Wstrict-prototypes -fPIC -I/usr/include/python2.7 -c bson/buffer.c -o build/temp.linux-x86_64-2.7/bson/buffer.o
      x86_64-linux-gnu-gcc -pthread -shared -Wl,-O1 -Wl,-Bsymbolic-functions -Wl,-Bsymbolic-functions -Wl,-z,relro -fno-strict-aliasing -DDEBUG -g -fwrapv -O2 -Wall -Wstrict-prototypes -D_FORTIFY_SOURCE=2 -g -fstack-protector --param=ssp-buffer-size=4 -Wformat -Werror=format-security build/temp.linux-x86_64-2.7/pymongo/_cmessagemodule.o build/temp.linux-x86_64-2.7/bson/buffer.o -o build/lib.linux-x86_64-2.7/pymongo/_cmessage.so
Successfully installed pymongo
Cleaning up...
anupama@anupama-VirtualBox:~$ python
Python 2.7.6 (default, Jun 22 2015, 17:58:13)
[GCC 4.8.2] on linux2
Type "help", "copyright", "credits" or "license()" for more information.
>>> import pymongo
>>>
```

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 list 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)

```

import pymongo
from pymongo import MongoClient
from twython import TwythonStreamer
import json

APP_KEY = "XXXXXXXXXXXXXXX"
APP_SECRET = "XXXXXXXXXXXXXXX"
OAUTH_TOKEN = "XXXXXXXXXXXXXXX"
OAUTH_TOKEN_SECRET = "XXXXXXXXXXXXXXX"

connection = MongoClient()
db = connection['tweeter']

class MyStreamer(TwythonStreamer):
    def on_success(self, data):
        if 'text' in data:
            text = data['text'].encode('utf-8')
            print text
            tweet_record = data
            tweet_record['_id'] = data['_id_str']
            db[stream].insert(tweet_record)

    def on_error(self, status_code, data):
        print status_code
        pass

stream = MyStreamer(APP_KEY, APP_SECRET, OAUTH_TOKEN, OAUTH_TOKEN_SECRET)

stream.statuses.filter(track='election2016')

twython-stream.py 31L, 842C

```

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)

Ubuntu [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

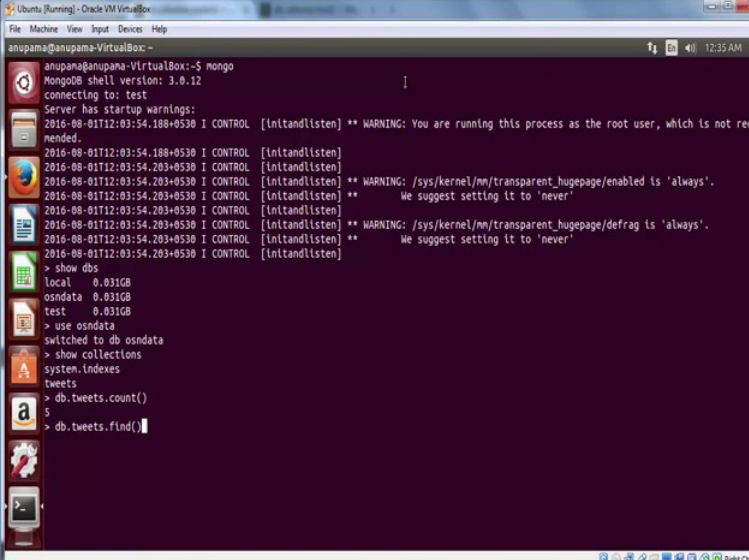
anupama@anupama-VirtualBox:~\$

```
anupama@anupama-VirtualBox:~$ vi twython-stream.py
anupama@anupama-VirtualBox:~$ python twython-stream.py
RT @GovGryJohnson: People are hungry to vote FOR someone as opposed to voting against someone else.

#teangov #election2016 https://t.co/s...
Not this vet https://t.co/HQcAVS2UOH
Hillary Clinton is the next president and will be re-elected in 2020! #Election2016 #Election2020 #HillaryClinton
Tricky https://t.co/MTLgYdKXNG
RT @SurreyElle: Sally Bradshaw (top Jeb Bush Advisor) leaves GOP. Will vote for HRC if vote in FL is close. https://t.co/xRc3Fv2c W
W! #EL.
```

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)



The screenshot shows a Windows desktop environment. A virtual machine window titled 'Ubuntu (Running) - Oracle VM VirtualBox' is open, displaying a terminal window. The terminal shows the user 'anupama' at the 'anupama@anupama-VirtualBox' prompt. The user has entered the command 'mongo', which has started the MongoDB shell. The shell version is 3.0.12. The user has entered 'connecting to: test' and 'wended'. The terminal shows several warnings from MongoDB, including 'WARNING: You are running this process as the root user, which is not recommended.' and 'WARNING: /sys/kernel/mn/transparent\_hugepage/enabled is 'always''. The user has entered 'show dbs' and 'show collections'. The terminal shows the output of these commands: 'local 0.031GB', 'osdata 0.031GB', 'test 0.031GB', and 'switched to db osdata'. The user has entered 'show collections' and 'system.indexes'. The terminal shows the output of these commands: 'tweets' and 'db.tweets.count()'. The user has entered '5' and 'db.tweets.find()'. The terminal shows the output of these commands: '5' and 'db.tweets.find()'. The desktop background is a blue Ubuntu logo, and various application icons are visible on the left sidebar.

```

anupama@anupama-VirtualBox:~$ mongo
MongoDB shell version: 3.0.12
connecting to: test
wended
2016-08-01T12:03:54.188+0530 I CONTROL [InlAndListen] ** WARNING: You are running this process as the root user, which is not recommended.
2016-08-01T12:03:54.203+0530 I CONTROL [InlAndListen]
2016-08-01T12:03:54.203+0530 I CONTROL [InlAndListen]
2016-08-01T12:03:54.203+0530 I CONTROL [InlAndListen] ** WARNING: /sys/kernel/mn/transparent_hugepage/enabled is 'always'.
2016-08-01T12:03:54.203+0530 I CONTROL [InlAndListen] We suggest setting it to 'never'
2016-08-01T12:03:54.203+0530 I CONTROL [InlAndListen]
2016-08-01T12:03:54.203+0530 I CONTROL [InlAndListen] ** WARNING: /sys/kernel/mn/transparent_hugepage/defrag is 'always'.
2016-08-01T12:03:54.203+0530 I CONTROL [InlAndListen] We suggest setting it to 'never'
2016-08-01T12:03:54.203+0530 I CONTROL [InlAndListen]
> show dbs
local 0.031GB
osdata 0.031GB
test 0.031GB
> use osdata
switched to db osdata
> show collections
system.indexes
tweets
> db.tweets.count()
5
> db.tweets.find()

```

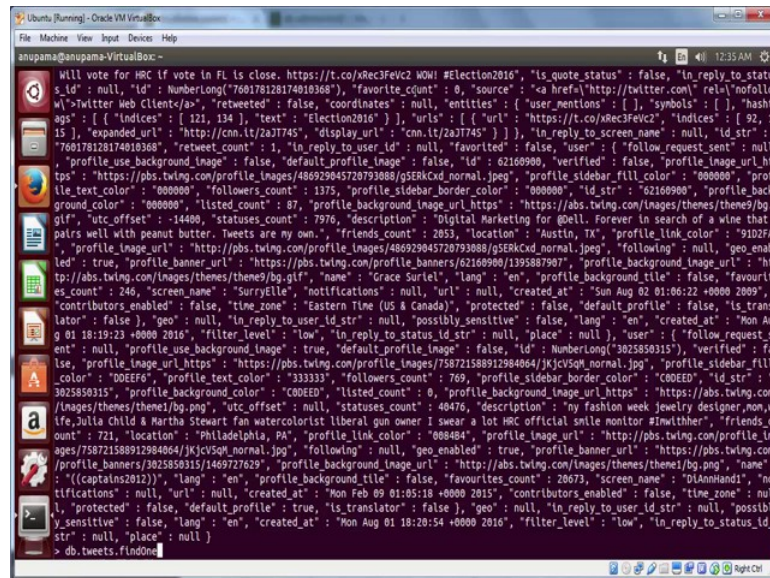
Let us start the mongo shell by typing mongo. Let us again list the existing databases by entering show dbs. This time you will note osndata listed. Recall that before data insertion, it was not showing up which means that it now holds some data. Let us verify this further. Switch to osn database by 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()`.

```
db.tweets.find()
```



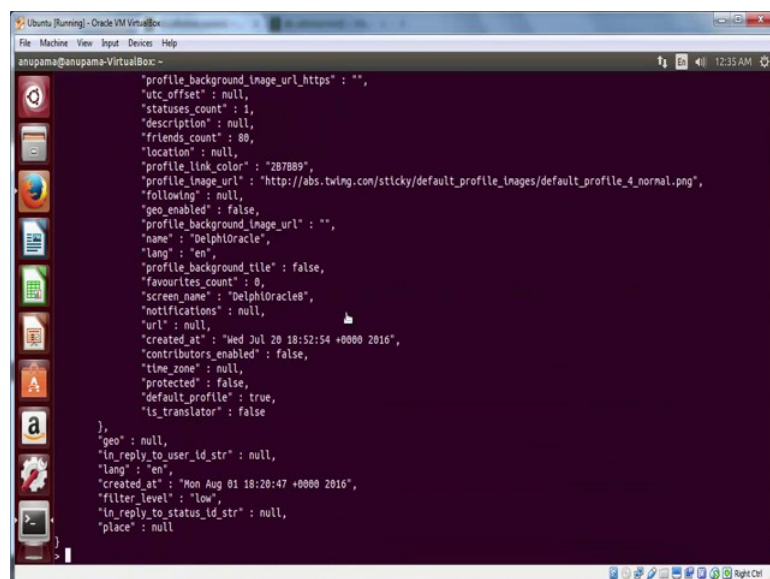
(Refer Slide Time: 07:03)



```
anupama@anupama-VirtualBox:~$ cat /tmp/tweet.json
{"s_id": null, "id": "NumberLong(760178128174010368)", "favorite_count": 0, "source": "a href='http://twitter.com' rel='nofollow'>Twitter Web Client/as", "retweeted": false, "coordinates": null, "entities": { "user_mentions": [ ], "symbols": [ ], "hashtags": [ ], "indices": [ 121, 134 ], "text": "Election2016" }, "urls": [ { "url": "https://t.co/xRec3FevC2", "indices": [ 92, 115 ], "expanded_url": "http://cnn.lt/2aJ7745", "display_url": "cnn.lt/2aJ7745" } ], "in_reply_to_screen_name": null, "id_str": "760178128174010368", "retweet_count": 1, "in_reply_to_user_id": null, "favorited": false, "user": { "follow_request_sent": null, "profile_use_background_image": false, "default_profile_image": false, "id": "62160990", "verified": false, "profile_image_url_https": "https://pbs.twimg.com/profile_images/486929045720793088/gSEKcxd_normal.jpeg", "profile_sidebar_fill_color": "000000", "profile_text_color": "000000", "followers_count": 1375, "profile_sidebar_border_color": "000000", "id_str": "62160990", "profile_background_color": "000000", "listed_count": 87, "profile_background_image_url_https": "https://abs.twimg.com/images/themes/themes/bg.gif", "utc_offset": -14400, "statuses_count": 7976, "description": "Digital Marketing for @Dell. Forever in search of a wine that pairs well with peanut butter. Tweets are my own.", "friends_count": 2053, "location": "Austin, TX", "profile_link_color": "91D2FA", "profile_image_url": "http://pbs.twimg.com/profile_images/486929045720793088/gSEKcxd_normal.jpeg", "following": null, "geo_enabled": true, "profile_banner_url": "https://pbs.twimg.com/profile_banners/62160990/1395887907", "profile_background_image_url": "https://abs.twimg.com/images/themes/themes/bg.gif", "name": "Grace Surler", "lang": "en", "profile_background_tile": false, "favourites_count": 246, "screen_name": "Surryelle", "notifications": null, "url": null, "created_at": "Sun Aug 02 01:06:22 +0000 2009", "contributors_enabled": false, "time_zone": "Eastern Time (US & Canada)", "protected": false, "default_profile": false, "is_translator": false }, "geo": null, "in_reply_to_user_id_str": null, "possibly_sensitive": false, "lang": "en", "created_at": "Mon Aug 01 18:19:23 +0000 2016", "filter_level": "low", "in_reply_to_status_id_str": null, "place": null, "user": { "follow_request_sent": null, "profile_use_background_image": true, "default_profile_image": false, "id": "NumberLong(3025850315)", "verified": false, "profile_image_url_https": "https://pbs.twimg.com/profile_images/758721588912944064/JKjvSgk_normal.jpg", "profile_sidebar_fill_color": "DDEEFF", "profile_text_color": "333333", "followers_count": 769, "profile_sidebar_border_color": "C0DEED", "id_str": "3025850315", "profile_background_color": "C0DEED", "listed_count": 0, "profile_background_image_url_https": "https://abs.twimg.com/images/themes/themes/bg.png", "utc_offset": null, "statuses_count": 40476, "description": "my fashion week jewelry designer, mom, wife, Julia Child & Martha Stewart fan watercolorist liberal gun owner I swear a lot HRC official smile monitor #mwithther", "friends_count": 721, "location": "Philadelphia, PA", "profile_link_color": "0084B4", "profile_image_url": "http://pbs.twimg.com/profile_images/758721588912944064/JKjvSgk_normal.jpg", "following": null, "geo_enabled": true, "profile_banner_url": "https://pbs.twimg.com/profile_banners/3025850315/1469727629", "profile_background_image_url": "http://abs.twimg.com/images/themes/themes/bg.png", "name": "(captains2012)", "lang": "en", "profile_background_tile": false, "favourites_count": 20673, "screen_name": "DlAnHandi", "notifications": null, "url": null, "created_at": "Mon Feb 09 01:05:18 +0000 2015", "contributors_enabled": false, "time_zone": null, "protected": false, "default_profile": true, "is_translator": false }, "geo": null, "in_reply_to_user_id_str": null, "possibly_sensitive": false, "lang": "en", "created_at": "Mon Aug 01 18:20:54 +0000 2016", "filter_level": "low", "in_reply_to_status_id_str": null, "place": null }
> db.tweets.findOne()
```

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)



```
anupama@anupama-VirtualBox:~$ cat /tmp/tweet.json
{"profile_background_image_url_https": "",
  "utc_offset": null,
  "statuses_count": 1,
  "description": null,
  "friends_count": 80,
  "location": null,
  "profile_link_color": "2B7889",
  "profile_image_url": "http://abs.twimg.com/sticky/default_profile_images/default_profile_4_normal.png",
  "following": null,
  "geo_enabled": false,
  "profile_background_image_url": "",
  "name": "Delphioracle",
  "lang": "en",
  "profile_background_tile": false,
  "favourites_count": 0,
  "screen_name": "Delphioracle",
  "notifications": null,
  "url": null,
  "created_at": "Wed Jul 20 18:52:54 +0000 2016",
  "contributors_enabled": false,
  "time_zone": null,
  "protected": false,
  "default_profile": true,
  "is_translator": false
},
  "geo": null,
  "in_reply_to_user_id_str": null,
  "lang": "en",
  "created_at": "Mon Aug 01 18:20:47 +0000 2016",
  "filter_level": "low",
  "in_reply_to_status_id_str": null,
  "place": null
}
```

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.