**<<TASK 0>>** Create instance (Amazon Linux 2)

1. Go to AWS EC2 and select create instance
2. Create keypair (automatically downloads - .ppk/.pem)
3. Defaults for rest
4. Connect with PuTTY – host (IP address); port (22); SSH->Auth->Key (.ppk file)
5. username: ec2-user
6. If it does not connect; make sure that port 22 is allowed in the inbound rules of that instance

**<<TASK 1>>** Install MongoDB

Guide

1. add repo file as given in guide
2. sudo yum install mongodb-org -y
3. ps --no-headers -o comm 1
4. for systemd:
    - start:- sudo systemctl start mongod
    - verify:- sudo systemctl status mongod
    - stop:- sudo systemctl stop mongod
    - restart:- sudo systemctl restart mongod
    - to use:- mongosh (mongo db shell)
      [to exit; type 'exit']

**<<TASK 2>>** move log data to instance

In PuTTY:

1. [in /home/ec2-user/] or where ever you want to add the logs
2. mkdir log_data
3. sudo chown -R ec2-user /home/ec2-user/log_data

In your system (cmd):

1. Transfer log files from local system to linux 2 instance
2. Syntax:
   pscp -i keyPath fileToBeUploadedPath username@ip:pathInRemoteSys
3. Example:
   pscp -i " /Downloads/myKeyPair.ppk" "/Downloads/log_data/2018/11/*"
   ec2-user@32.153.269.114:/home/ec2-user/log_data
4. All the .json files from the log_data should be successfully transferred to the instance

[To confirm that files have been transferred;
you can use 'ls' in the log_data folder in the instance – using PuTTY]

**<<SUBTASK>>** Create a virtual python env in the instance (PuTTY)

- Creating venv: [only once – activate/deactivate as needed]
  python3 -m venv venv
- Activate:
  source venv/bin/activate
- continue using as normally would [Potentially; go to TASK 3]

- Deactivate: [only do this when you wish to no longer use python env]
  deactivate

**<<TASK 3>>** insert logs into MongoDB collections

1. sudo su
2. pip3 install pymongo
3. create db and collection
4. Insert logs into collection
   You can insert one by one – each json object (insert_one) Or
   insert all json objects in a single file together (insert_many)
5. how to check?
   - Use mongoDB shell: mongosh
     - show dbs [view all dbs in mongo]
     - use dbName [go into the db that you created in step 2]
     - show collections [view all collections in that db]
     - View a single log file example:
       db.collectionName.findOne({userId: '101'})

**<<TASK 4>>** fetch top 10 artists and songs

- Use mongoDB aggregations to fetch top 10 artists and songs
  [you can do them both separately too for ease]
  [References: Documentation; TutorialsPoint; Simple example]
- Save the output however you like
  (you can save a dictionary, or a list, or create an html page which can be used directly in flask)

**<<TASK 5>>** setting up and using Flask

- pip install Flask
- mkdir templates
- add .html files to templates folder – whatever files you wish to have
  [topSongs.html; topArtists.html; topall.html; etc]

- in venv:
    - create an app.py file [vim app.py] as below

    [home.html – can be replaced with whatever html file you want flask to display]

```
from flask import Flask,render_template
app=Flask(__name__)
@app.route("/")
def home():
    return render_template("home.html")

if __name__ == "__main__":
    app.run(host='0.0.0.0',port=80)
```

    - to run:

      python3 app.py

- Add port 80 to inbound rule of instance
- use IP addr:80 to view webpage

insertToCollections.py:

```
import pymongo
import os
import json
import itertools

client = pymongo.MongoClient("mongodb://localhost:27017/")
mydb = client["db"]
col = mydb["songs"]

dire = r"/var/www/data/"
week = 1
day = 1
for filename in os.listdir(dire):
    with open(dire + filename) as f:
            stuList = []
            for jsonObj in f:
                dic = json.loads(jsonObj)
                dic['week'] = week
                stuList.append(dic)
            col.insert_many(stuList)
            day += 1
            if day == 8:
                week += 1
                day = 1
```

top10Songs.py [returns an html of top 10 songs per week]:

```
import pymongo
import pprint

myclient = pymongo.MongoClient("mongodb://localhost:27017/")
mydb = myclient['db']
mycol = mydb['songs']

pipeline = [
  {
    "$group": {
      "_id": {
        "week": "$week",
        "song": "$song"
      },
      "songCount": {
        "$sum": 1
      }
    }
  },
  {
    "$sort": {
      "songCount": -1
    }
  },
  {
    "$group": {
      "_id": "$_id.week",
      "songs": {
        "$push": {
```

```
            "songName": "$_id.song",
            "count": "$songCount"
          }
        },

      }
    },
    {
      "$sort": {
        "count": -1
      }
    }
]

finalSongs = dict()
for x in mydb.songs.aggregate(pipeline):
    x = dict(x)
    finalSongs[x['_id']] = []
    for y in x['songs'][1:11]:
        finalSongs[x['_id']].append(dict(y)['songName'])

html = "<html><div class='box'><h1>Top 10 Songs per week</h1>"
for key in sorted(finalSongs):
    html += '<h1>Week' + str(key) + '</h1><ol>'
    for y in finalSongs[key]:
        html += '<li>' + y + '</li>'
    html += '</ol>'

html += '</div></html>'
print(html)
```

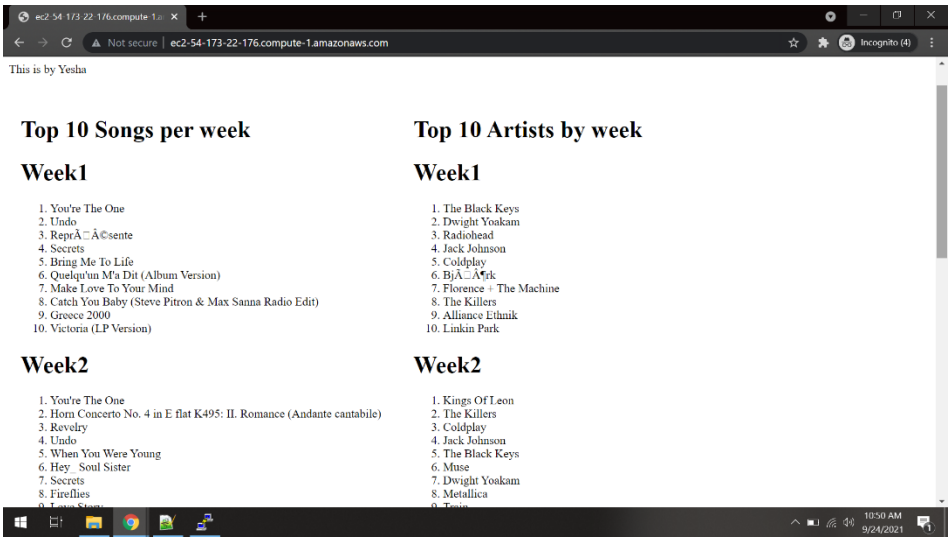
* you can use python3 top10Songs.py > templates/top10Songs.html to push results to an html file *

* modify above file to implement top10Artists.py and create top10Artists.html *

Home.html:

```
<html>
<head>
<style>
.box {
    float : left;
    width : 500px;
    margin : 1em;
}
</style>
</head>
<body>
    <h1>Welcome to flask application</h1>
    <p> This is by Yesha </p>
    {% include 'topSongs.html' %}
    {% include 'topArtists.html' %}
</body>
</html>
```

Output:


Not secure | ec2-54-173-22-176.compute-1.amazonaws.com
Incognito (4)
This is by Yesha
Top 10 Songs per week
Week1
1. You're The One
2. Undo
3. ReprÃ□Â©sente
4. Secrets
5. Bring Me To Life
6. Quelqu'un M'a Dit (Album Version)
7. Make Love To Your Mind
8. Catch You Baby (Steve Pitron & Max Sanna Radio Edit)
9. Greece 2000
10. Victoria (LP Version)
Week2
1. You're The One
2. Horn Concerto No. 4 in E flat K495: II. Romance (Andante cantabile)
3. Revelry
4. Undo
5. When You Were Young
6. Hey_ Soul Sister
7. Secrets
8. Fireflies
Top 10 Artists by week
Week1
1. The Black Keys
2. Dwight Yoakam
3. Radiohead
4. Jack Johnson
5. Coldplay
6. BjÃ□Â¶rk
7. Florence + The Machine
8. The Killers
9. Alliance Ethnik
10. Linkin Park
Week2
1. Kings Of Leon
2. The Killers
3. Coldplay
4. Jack Johnson
5. The Black Keys
6. Muse
7. Dwight Yoakam
8. Metallica
10:50 AM
9/24/2021