

SNU Chennai
Big Data Analytics – Model lab exam
B.Tech AI & Data Science

1.

- a) Create IAM users(alice & bob), put them under group(server admin) and give full access to EC2 services
- b) Create IAM users(cathy & david) put them under group(dns admin) and give full access to Route 53 services
- c) Create IAM user(eve) give him access to billing
- d) Create IAM user(your_name) and give full access to all the services
- e) Create an alias name for your account & check the login via alias url instead of account ID.
- f) Login as alice & create an EC2 instance (Name : Web Server) with elastic public ip with required port numbers open in the security group.
- g) check whether the port 80 is open for your instance from the below url
<https://portchecker.co/>

2.

- a) Create an EC2 instance (Name : Web Server) with elastic public ip
- b) Connect to the instance via 'EC2 Instance Connect' to get the cli of the instance via browser
- c) Install apache application(web service) and make vm as web server
- d) create a security group to allow http, ssh & ping to the web server
- e) Create an EC2 instance (Name : Web Client) with auto assign public ip
- f) Install 'links' application(web client) if required and access the web page of the Web Server

3.

Use the file u.data and do the following using PySpark:

- Read the data as a spark dataframe
- Create the schema using StructType – first column is movie id, second column is user id, third column is the rating and final column is the timestamp.
- Find the top 10 most popular movie – movie that has highest 5 ratings
- Find the 10 most popular users – users who have given most ratings

4. Use the book.txt file and do the following using Hadoop:

- Upload the book.txt to the HDFS.
- Display the content of book.txt from HDFS
- Write a mapreduce code to perform the wordcount
- Print the output file and display the result.