

Cloud Computing Final Report

Norbert Prokopiuk

January 2021

1 Introduction

The main purpose of this project was getting to know how to use cloud computing services. In this case it was AWS service.

2 How to configure computing instance

Project was started by creating AWS account. Then it was necessary to configure these components:

- IAM-module where you can configure security and access group
- S3 - module where you can configure storage
- EC - module where you can configure computing instances

Now I will describe how I created and configured this modules. I will describe it in the form of tutorial.

2.1 IAM configuration/creation

In this section I will describe how to configure security grup. You have to find section "Security Groups" on EC2 dashboard and enter "Create Security Group". When you click "Create Security Group" you will be able to name the security group as you like. After this you have click "Add Rule" and select all traffic in type section. Now click create security group.

2.2 S3 configuration/creation

To configure S3 module choose S3 on AWS services dashboard. It will be in storage section. When you enter this section enter "Create Bucket". Now you can choose name and region for your bucket. Other option might be in default state. When you set name and region, click create bucket. Now go to IAM section on AWS services dashboard and select users section. Now choose add user. Choose name for new user and select programmatic acces. In section Permission select Attach existing policies directly and choose AmazonS3FullAccess. Skip

other options and click create user. Now you can see access ID and secret access ID. You should save it in safety place, because they are needed to get access to your bucket.

2.3 EC2 configuration/creation

To create EC2 instances choose EC2 on AWS services dashboard. Now choose Launch instance. In next window select Ubuntu Server. After that choose t2.micro as a instance type. Now go to Configure Security Group section, choose "Select an existing security group" and select the group created in subsection 2.1. In the last window choose create new access key, download it and save for later. You have to create 2 EC2 instances. Instances created as a second can use the same access key as the first.

3 Linux configuration

In my case I have to create Spark cluster consisting of my two instances. I did it based on tutorial created by Oscar Bienko: <https://github.com/OskarBienko/Spark-cluster/>
I also use tutorial on this website: <https://medium.com/@josemarcialportilla/getting-spark-python-and-jupyter-notebook-running-on-amazon-ec2-dec599e1c297>
This article also was really helpful. Generally Linux configuration comes down to install Anaconda, Java, Scala and Apache Spark. I also prepare Jupyter notebook to be able to write my code more comfortably.

4 Linear Regression

To prepare linear regression script I used Karlijn Willems article on data-camp.com website (<https://www.datacamp.com/community/tutorials/apache-spark-tutorial-machine-learning>). But in my case I use other dataset. This dataset includes:

- age,
- sex,
- bmi,
- children,
- is smoker?,
- region,
- charges.

You can find it on my github (https://github.com/norbertprokopiuk/Cloud_Computing). Unfortunately I can't present my results grafically because for unknown reason adding matplotlib library ended with failure. Numerical results are as follow:

```
results
coefficients:[0.06702845515810069,0.0,0.00032264265856510907]
intercept: 33.099450658730554
RMSE: 13.269802665073149
R2: 0.09528124355001555
```

Figure 1: Results

5 Conclusions

AWS is a very powerful computation service, can be used to solve much faster engineering problem. However if we want to use all possibilities, we have to pay. Thats a huge minus but free version let you learn how to use this platform.