CS 4287

Professor Gokhale
Team 10
Zhitao Shu, Duke Jiang

What we have done and challenges:

At first we started by setting the VMs on local machine and on the cloud. That part was straightforward and we were able to setup all three instances with some minor tweaking.

The challenge came when we were trying to setup the dependencies on our cloud machines. We first checked if python is installed on our machine and did a apt update

We did a “pip install python-kafka”. It was installed successfully. Then we tried to install kafka sever by using ‘pip install kafka” not expecting it to work and it didn’t. We then went online a found a couple tutorials. We were able to successfully install kafka server with this tutorial: https://tecadmin.net/install-apache-kafka-ubuntu/

We ran in some problems while trying to follow this tutorial.

1. Unable to extract the file: we found that the file extractor on one machine might be out-dated. We ran an update and fixed that problem
2. Transport-https: we installed the package using this tutorial: https://zoomadmin.com/HowToInstall/UbuntuPackage/apt-transport-https

With all that we were able to get kafka server up and running

We then modified the consumer.py code but when we tested it, a “no broker” error popped up. We were eventually able to identify the error and fixed it by 1.modifying the security group on the Chameleon cloud. 2.opening up required ports in the firewall on VM2 and VM3.

The next challenge is to parse the data stream into JSON format and save that into couchdb. We were able to setup couchdb using this tutorial: https://linuxize.com/post/how-to-install-couchdb-on-ubuntu-18-04/
We tested by running producer and consumer but quickly found out that we needed to create a “test” database. We used this tutorial to create that database and verified it exists:https://www.tutorialspoint.com/couchdb/couchdb_creating_a_database.htm

We ran the process again and we were able to monitor the data feeding in on our console for VM3

Conclusion:

In this assignment, we gained hands-on experience with cloud development environments and a deeper understanding of how different components of the network work with one another to send, process, store, and receive data. We also learned how to troubleshoot and look for the right resources when things are not going as expected. It is pretty exciting to start a project from scratch and little by little build up the whole architecture.