UCLH PEACH and NHS Open Source: OpenEHR Architecture and Analytics

Team 38: Report 4

Sandipan Ganguly
Mengyang Wu
Desislava Koleva

Overview:

During the past two weeks, we have started installing the different services onto our DC/OS cluster such as Apache Kafka and Apache Spark. Furthermore we started extensive research on Apache Nifi to understand how we can create dataflows and connect it to Kafka.

We also started working on the PEACH Generator and read the previous documentation thoroughly. We imported the previous code from the Gitlab Repository and run the Generator source file. We tested the application and figured out which areas are working and which aren't.

Furthermore , we have made significant progress in our Project Website such as finishing up the Research and Resources page. We have also created an experiment log that keeps a record of tasks attempted and completed and uploaded that into our website.

Problems Faced:

- We reached the limit for the maximum number of available cores for our main subscription and the information about the subscription quota was not available in the Azure Portal Service.
- We had some trouble establishing the requirements for the PEACH Generator since we were not aware of how much work that was already done on it.

Successes:

- We successfully installed Confluent Kafka and Spark inside our DC/OS cluster.
- Successfully imported the previous Generator source code and run it in our local machine

Plan for next two weeks:

No.	Task
1	Create a successful dataflow between Apache Nifi and Kafka
2	Configure Kafka so that it can exchange basic messages
3	Create a secure server to store the Radiology data provided by Dr Ramachandran
4	Create our initial prototype which allows us to input a data file through Nifi and read it from Kafka
5	Document the entire dataflow diagram of our prototype
6	Draw sketches for a mockup of the GUI for Kafka
7	Finalize the Requirements for the Generator

Summary of meetings held:

Meeting Date	Who attended	What we did
25/11/2016	Sandipan Mengyang Desislava	● Downloaded and installed Apache NiFi ● Created Experiment Log
28/11/2016	Sandipan Mengyang Desislava	● Discussed progress made so far ● Finalised Experiment Log by making sure all experiments done so far are recorded in the table ● Set targets for the upcoming weeks
02/12/2016	Sandipan Mengyang Desislava	● Finalised our project website ● Discussed the next steps with our client, Dr Ramachandran.

Individual Contributions:

Sandipan Ganguly

During the last week, I installed Powershell and configured it with my Azure account so that I could find out the maximum number of cores as requested by Dr Ramachandran. This allowed him to submit a ticket to increase the cores. After the quota was increased, I deployed my own DC/OS cluster and installed Confluent Kafka in it. This will allow me to experiment with it and configure it for further use. Furthermore I installed Nifi and did extensive research on it and the previous file so that I can create and work with dataflows in the coming weeks.

Mengyang Wu

During the last two weeks, I tried to deploy the whole system including kafka, spark and nifi on our running DC/OS system. In that period, I was faced with several difficulties but successfully installed spark. Now I'm trying to figure out the problem in Kafka installation. Meanwhile, I received a new task to continue developing the generator from the previous project. I did test on the old version of code and find some errors. In the following week I will finalize the requirements for the generator with our client and started to work on it.

Desislava Koleva

Throughout the past two weeks, I have mostly been focused on Apache NiFi. I have downloaded and installed Apache NiFi following another online tutorial and have begun practising using it by creating and monitoring dataflows, adding, configuring and connecting processors, as well as transforming data. Furthermore, I have started conducting research on Apache Spark and its advantages and disadvantages.