UCLH PEACH and NHS Open Source: OpenEHR Architecture and Analytics

Team 38: Report 7

Sandipan Ganguly
Mengyang Wu
Desislava Koleva

Overview of progress :
During the last couple of weeks, we made a lot of progress with a live successful workflow in our local systems. This workflow mainly dealt with the Apache NiFi and Apache Kafka and the flow of data between them. To start with, we configured our own processors in NiFi and to understand the ingestion of data better and then moved onto working with the previous team's NiFi Template. We managed to fix the errors and configure it for our local machines. Furthermore, we configured a slack server and used its API to receive the error messages from NiFi. We also started looking into Spark and ways of implementing it in our local machines.

We also received the encrypted radiology data from our client, Dr Navin Ramachandran, which we will use to recreate the dataflow made by the previous team. In addition to that, we looked into Javafx so that we can start working on the new UI for the Generator as soon as possible.

Problems Faced:
Due to the lack of documentation from the previous team about their NiFi template, we had trouble figuring out what each Label contained and what each processor did. It was the same situation when we tried to understand the previous Kafka-Spark connector code.

Successes:
We were successfully able to use the output from the Generator (.csv files) to feed into Apache Nifi which then converts the data to JSON files and sends it to Kafka.

Meanwhile, we recovered the previous Kafka-Spark connector and performed tests on data transfer between Kafka and Spark.

Plan for next two weeks:

No.	Task
1	Implement Spark data analytics in our workflow
2	Deploy the workflow in the Azure DC/OS server
3	Research and attempt to implement the Kafka Streams in our local machines
4	Work on the new Generator UI

Summary of meetings held:

Meeting Date	Who attended	What we did
30/01/17	Sandipan Mengyang Desislava	● Discussed the plans for the upcoming weeks and divided the tasks between ourselves. (Each person were to install NiFi on their local machines and experiment with different processors so that we would be know what to use in our final workflow)
03/02/17	Sandipan Mengyang Desislava	● Having installed NiFi and Kafka in our machines, we used the Generator output which we converted to JSON files using our workflow ● Received the secure radiology data from Dr Ramachandran and were told on how to access it
06/02/17	Sandipan Mengyang Desislava	● Configured a Slack server from the previous team's NiFi template so that we could receive the error messages while the dataflow is active. ● Looked into the previous team's Spark deployment and started some research on our own so that we could test the dataflow between Kafka and Spark
10/02/17	Sandipan Mengyang Desislava	● Demonstrated our current Random Generator as well as our analytics platform to Dr Yun Fu ● Started looking into Kafka Streams. ● Attempted to implement the previous team's Spark installation

Individual Contributions:

Sandipan Ganguly

During the first week, I installed Apache NiFi and Kafka in my local system to get started with the test dataflow. In NiFi, I worked around with the GetFile and Fetch File processors which dealt with the data ingestion from a local disk. Then, I connected them to PublishKafka which copied the input files to a test Kafka topic that was running concurrently on my machine. In between, I used few other processors for example SplitText which split the the rows in the .csv files into single lines and ConvertCSVtoAvro , ConvertAvroToJSON to convert the input to JSON files. This helped me understand a lot more about the previous team's NiFi template and how we could implement that in our final version.

In addition to that, I started looking into Kafka Streams and Apache Spark which we will use in the next stage of our dataflow.

Mengyang Wu

My main contribution during the past two weeks is implementing the workflow with Sandipan. In the first week, I was focus on the Nifi preprocessor and its connection to Kafka server. I managed to understand the detailed structure of the original design and fixed bugs unsolved from previous team. In the second half, I contributed more on Spark-Kafka connector. Although the documentation of the previous code is weak, I successfully found a way to submit the "consumer task" to local Spark service. Next step for me is to run Nifi, Kafka and Spark simultaneously, and make data go through all the three components.

Desislava Koleva

Throughout the past two weeks, I have continued to practise using Apache NiFi and learn more about its uses as well as ways to integrate it into our system. Furthermore, I took up the task of writing the first draft of an article to be posted on Medium.com briefly describing our PEACH project and its challenges, as requested by our client Dr Ramachandran. Next, we will share the draft with our client so as to receive feedback and be able to apply any necessary changes.