PRAGMA Data Storage Monitoring and Visualization

[1]Hajeong Cho, [1] Minsun Lee, [2] Nadya Williams

[1] Chungnam National University, [2] University of California-San Diego

Background and Motivations

It's difficult to process and visualize a huge data, especially, a real-time data between server and client. This data is inserted into Influx DB and visualized inside Grafana dashboard. When an administrator tries to monitor the data in the DB with the host's IP address or port number, it takes a long time to process and query the DB directly. The purpose of this project is to enhance data monitoring efficiency for the PRAGMA Cloud.

System Architecture
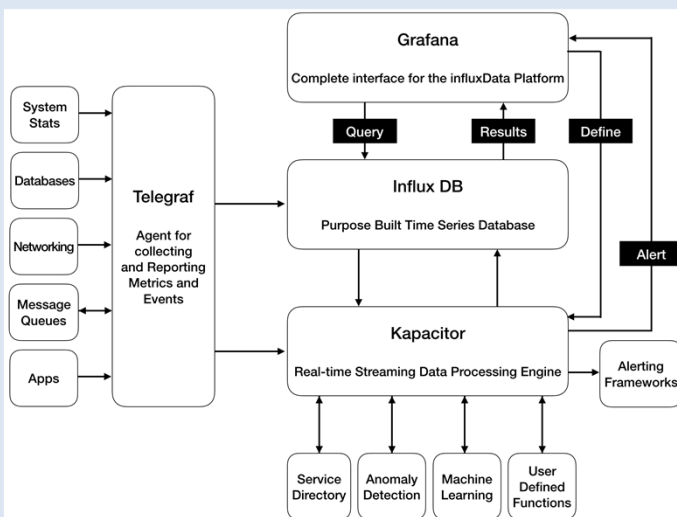
- ➤ The data is inserted into the Influx DB using this Python program. Grafana calls the Influx DB to query data and visualize a time-series data.

- ➤ Telegraf stores data synchronously in the database, so it is excluded from this project that requires asynchronous updates.

- ➤ Kapacitor, which is an open source sw, sends an alert whenever configured condition is satisfied.

Results and Visualization

- ➤ The Python program exports data from log files without converting types, each of which was collected for three months and has 785,085 lines.

- ➤ In earlier versions of the program, the processing time was too long to measure, but the modified program took 5 minutes.

- ➤ The processed data is inserted into the Influx DB using the built-in HTTP API, which helps you insert data at multiple points.

- ➤ The data is exported from influx DB, and Grafana creates a dashboard which shows the network status information such as average RTT, retransmission, and payload etc.

- ➤ The following figure shows the various network status of servers.
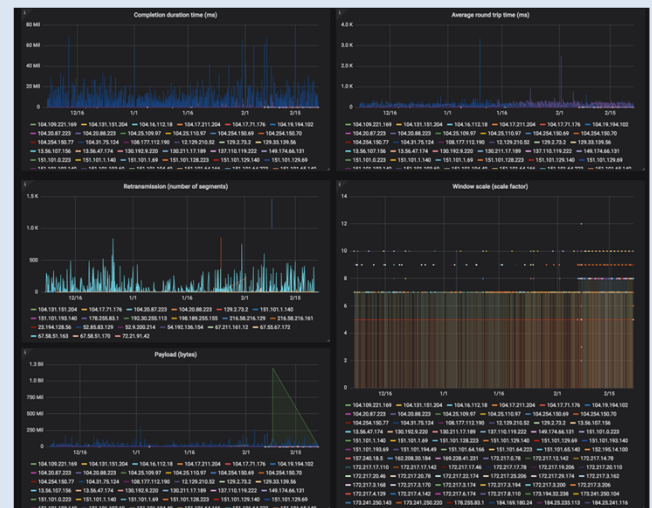
Conclusions

- ➤ Visualization dashboard with Grafana which queries events through Influx DB was created.

- ➤ The modified Python program exports data without converting the types.

- ➤ This program also inserts data into the Influx DB at multiple points.

- ➤ As a result, the processing time has been reduced compared to previous version of the program.

Future Works

- ➤ The time to process and insert data into DB should be reduced less than 2 minutes.

- ➤ Future work includes monitoring the most and the least connected host address(server or client)

References

- ➤ Grafana Documentation,
  http://docs.grafana.org/guides/getting_started/

- ➤ Influx DB Documentation,
  https://docs.influxdata.com/