

INTERNSHIP REPORT

Company: FogLogic Inc.

Position: Data Science Intern

Duration: 12 weeks

During my internship, I worked on a Data Visualization project.

The motive of the project was to visualize the changes happening in different system & performance metrics compared to the baseline date.

Use case: Suppose a company is undergoing some major migration. They wanted to analyze, what and why are the changes observed across different days when compared to the state of system on the day of migration.

STEPS:

1. Computed the kernel density estimate for baseline date.
2. For the subsequent days, compared the values from baseline date and assigned it a score from 0 to 10.
3. Computed the delta scores for each day compared to the baseline date.
4. Created an animated bubble chart with hours on x axis, metric names on y axis and slider to show the trend over days. We can play this visualization like a movie.
Size of circle denotes amount of change, plus and square shows that the change is like the previous day's change. Color denotes that change is positive/negative.

There were various filters for visualization.

KGS Visualization

Please select filters

Top N changing metrics:

5

×

▼

Select Colorscale

11

×

▼

Blue To Teal

×

▼

Select desired scores for end state

7

×

▼

Select desired scores for highlights

5

×

▼

Select System:

FG1

×

▼

Select Server:

ICBLCQSBQS0_FG1_00

×

▼

Select Metrics:

☒ System Utilization ☒ User Utilization ☒ 5minLoadAverage ☒ NumberOfLocks ☒ cancelledJobRate ☒ sysLogRate ☒ rteRate ☒ UsersLoggedIn ☒ ResponseTime ☒ DBRequestTime ☒ QueueTime ☒ RollTime ☒ Used Memory ☒ Used Data Space ☒ Used Log Space ☒ InboundQueueTotal ☒ OutboundQueueTotal ☒ Interrupts ☒ System Calls ☒ Context Switches

Start Date

3-1-2019

End Date

3-20-2019

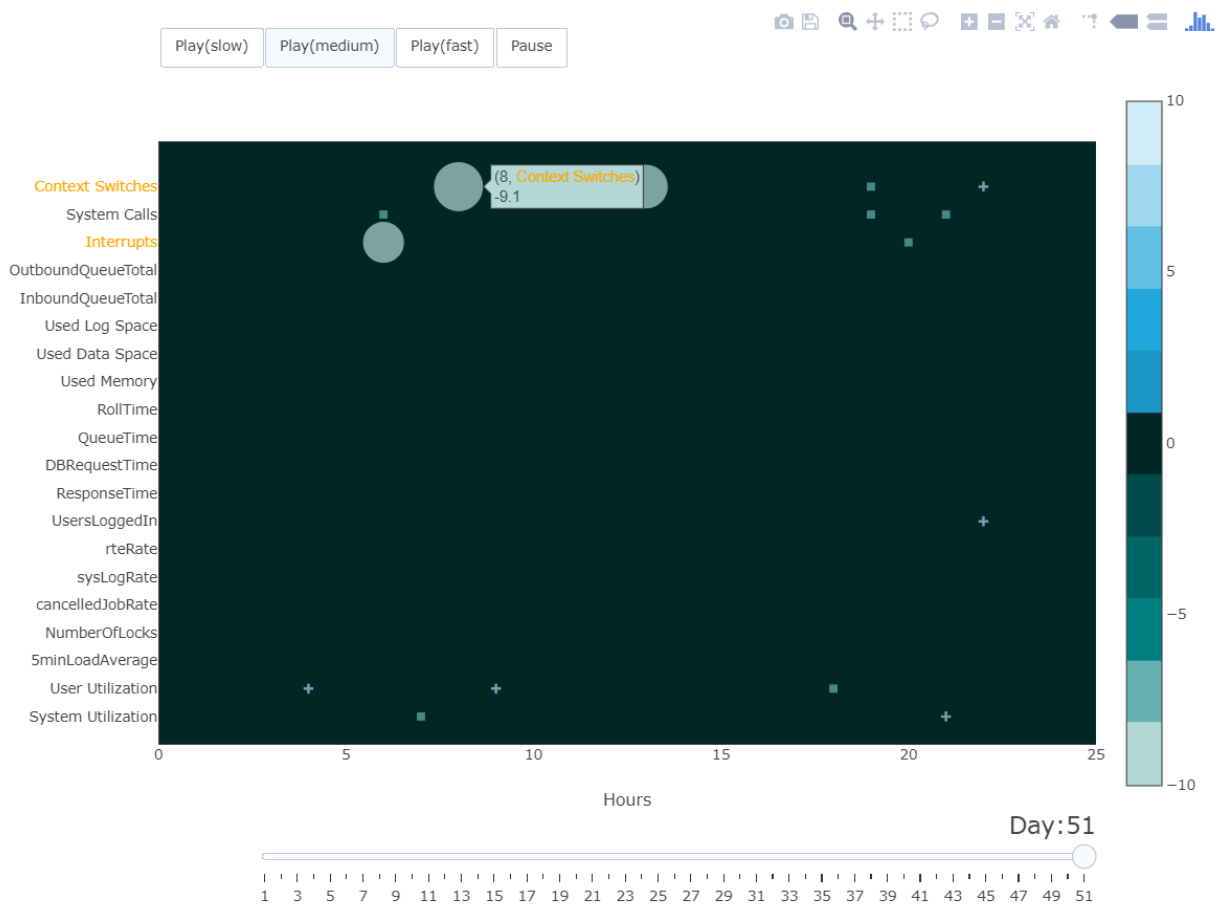
Submit

Play(slow)

Play(medium)

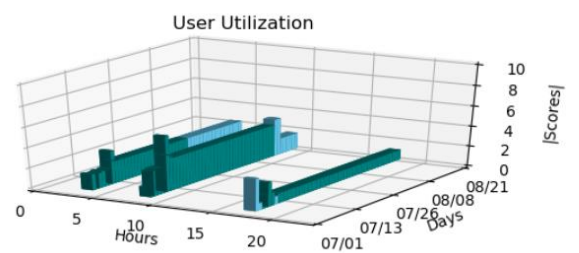
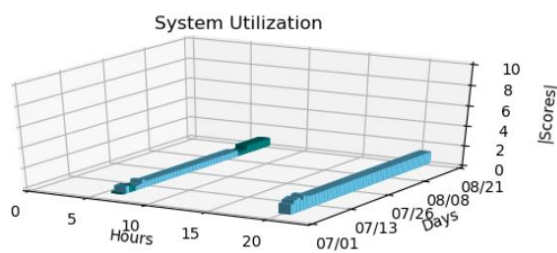
Play(fast)

Pause

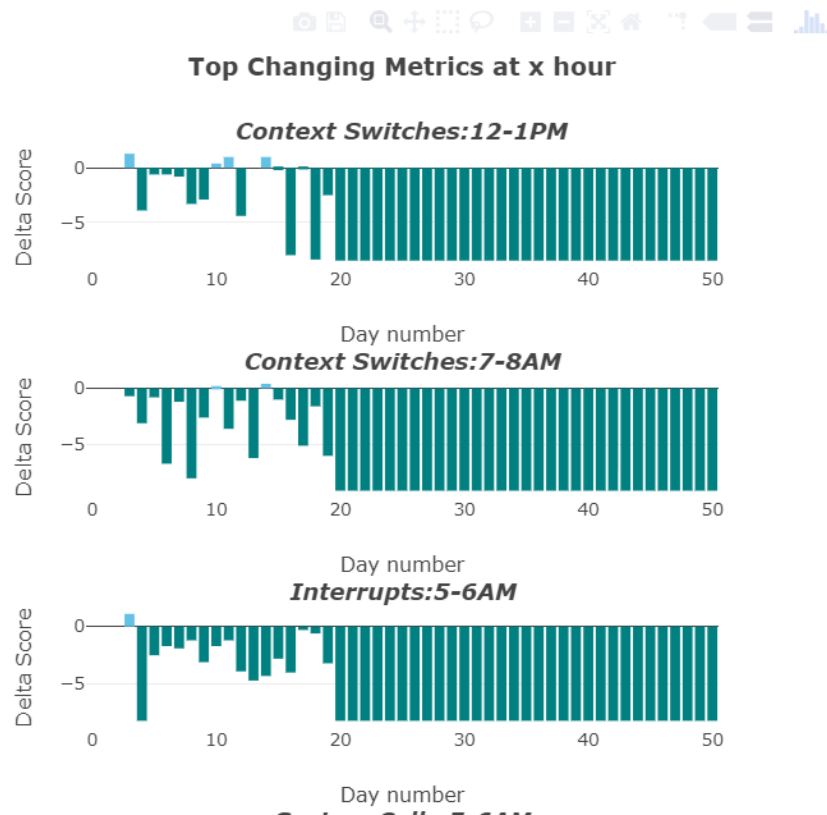


Positive change means that the metric is more likely to exhibit that value. Negative means that it is less likely.

- I also created 3 d animations to visualize change from a different angle. We can see that how is the change happening over days.



6. I also worked on finding the top N changing metrics during any time period entered.



This shows that context switched from 12 – 1 pm has most variability than all other metrics at any hour. The visualizations were created in Dash framework.

Second task which I worked on was anomaly detection and Incident generation. The task involved generating an incident if cluster of anomalies occur together. The code was written in python.

CONCLUSION

It was a great learning opportunity for me. It was my first experience of working in a startup environment. I also got the opportunity to work with amazing people. I had incisive mentors to guide me throughout my tenure.