CloudAEye
Intelligent Cloud Operations

Incident Detection and Root Cause Analysis

1. Context

Centralized Logging
Let's say we have centralized logging with ELK stack. Please see this page for examples from AWS: aesworkshops.com

One can get very detailed information of the logs. See here: aesworkshops.com/log-analytics/visual

See video: https://www.youtube.com/watch?v=n_rFNppLpQI

Centralized Monitoring
We also have centralized monitoring: www.cncf.io/blog/2018/12/18/cortex-a-multi-tenant-horizontally-scalable-prometheus-as-a-service

See video: https://www.youtube.com/watch?v=LW4RDJJI2tI&t=4s

Some example logs are given in log-data.csv.zip.

2. Problems

Now, let's say that I want to build an AI/ML model that will notify me when there is a “real failure”. We may define “real failure” as 'some functionality of an application that is not working due to a sub-system failure'. For example, database is down, networking outage, etc. So, user is not able to see product catalog.

How would you build this model? What type of information you would need to train such a model to detect 'real failures'?

Follow-up, the fact that database is down (let's assume that was the reason), that information is also in the logs. How would you do the correlation to build that knowledge/context so that you can offer to users the cause as a root-cause of the incident.

3. Examples

Here is an example videos to give you concrete idea of the problem I described based on logs:

- https://www.youtube.com/watch?v=jh676pNiBgM

4. Assignment

How would you design such a system? Please describe clearly the system architecture and type of AI/ML model(s) you would build. Please describe the type of data you need to train the models.