CERN openlab

EOS Winston

Expert Systems for Automated Diagnosis and Remediation

Ishank Arora, IT-ST-FDO
Supervisor: Luca Mascetti
EOS

A disk-based, low-latency storage service

Scale:
- Over 280 PB of raw disk space
- 4.88 billion files
- 50,000 disks
- 25+ namespaces
Issues and Anomalies

*It works! It works! Wait a second...*

- The **very-critical-mount** on EOSPublic is 95% full (of logs dating a year back).
- User 1729 is performing read operations at 7201.21 Hz.
- Boot failure on 20 nodes in EOSAlice.
- Node `heart-attack.cern.ch` has a heartbeat of 83018 *(actual figures).*
Project Requirements

*Check. Check. Wait a second...*

- Reduce MTTR and pager fatigue
- Remove false positives
- Enable the team to focus on the business logic rather than the infrastructure
- Provide a common platform for alerts and collaboration
EOS Winston

I'm Winston Wolf. I solve problems.

https://giphy.com/gifs/quentin-tarantino-pulp-fiction-harvey-keitel-DeOa0SqsDH5sc
EOS Winston

An event-driven alerting and mitigation automation platform. It employs expert rules and streaming anomaly detection algorithms to categorise events and execute corresponding runbooks.
The architecture

EOS Instances → Diagnostic Information → EOS Winston → Deploys runbook → Assisted diagnostics
False positives
Anomaly detected/ no runbook defined
StackStorm\textsuperscript{1} for automation

*(Thank God for open-source)*

- Pluggable rule engine comprising of triggers and actions
- MongoDB for data storage and RabbitMQ for message broking
- Supports a wide variety of interfaces, available as packs
- Deployed by firms such as Netflix\textsuperscript{2} and NASA

[1]: https://stackstorm.com/
Anomalies in the Average MGM Load

*Nupic*: *Real-time Streaming Anomaly Detection*

- Try to identify if there’s a sudden unexplained increase in the average load across all nodes of an EOS instance, or a given user’s activity.
- *Nupic*, a real-time streaming algorithm, employs Hierarchical Temporal Memory (HTM) networks

*Nupic*: Real-time Streaming Anomaly Detection

Why not LSTMs?
An EOS Winston instance
Implemented Runbooks

*Winston in action*
Implemented Runbooks

Winston in action

- Drain expired:

  Instance: eosalice
  Hostname: eosalice-srv-m1
  Draining issues. Please review:
  ✔ Drained 11 file systems:
  5363, 6521, 7275, 7489, 8532, 10258, 10944, 11052, 14589, 15735, 21127

- /var usage > 80%:

  Instance: eoshome-i00
  Disk usage issues. Please review:
<table>
<thead>
<tr>
<th>Mount</th>
<th>Disk Usage</th>
<th>Usage %</th>
</tr>
</thead>
<tbody>
<tr>
<td>/var</td>
<td>1.39 TB/1.80 TB</td>
<td>82%</td>
</tr>
</tbody>
</table>

  ✔ Cleaned files on /var/tmp/
Implemented Runbooks

*Winston in action*

- User activity through the roof:

ChatOps Integration

Conversation-driven development

- Execute workflows as commands through Mattermost
- Unifying the communication about what work should get done with the actual history of the work being done
- New commands can be implemented by just adding the business logic
More of Winston

Latest EOS reports as of 09 Aug 2019 15:07

Instance: eosalice
Draining issues. Please review:
- Drained 2 file systems: 19845, 21891

Instance: eosatlases
Node issues. Please review:
- Node: ldfsref2308.cern.ch:1095 has a heartbeat of 79419.

Instance: eoscms
Node issues. Please review:
- Node: ldfsref5603.cern.ch:1095 is offline.

Instance: eoshome-03
Disk usage issues. Please review:

<table>
<thead>
<tr>
<th>Mount</th>
<th>Disk Usage</th>
<th>Usage %</th>
</tr>
</thead>
<tbody>
<tr>
<td>/var</td>
<td>1.42 TB/1.80 TB</td>
<td>84%</td>
</tr>
</tbody>
</table>

Instance: eospublic
User processes crossing specified thresholds
- User 1005596 is performing Stat operations at 707.81 Hz.
- User 25567 is performing Stat operations at 306.21 Hz.
More of Winston (cuz why not)
Impact of the Project

Mr. Wolf, I just wanna tell you it was a real pleasure watching you work.

https://www.netflix.com/watch/880640
Thank you!
Questions?

ishank.arora@cern.ch