

Monitoring and Operations

Darren Moore

STFC RAL

GripPP41 – August 2018

Agenda

- New Production Manager
- Tier-1 On-Call
- Monitoring with InfluxDB/Grafana
- Other operational news...

Who am I? – New kid on the block

The new kid on the block

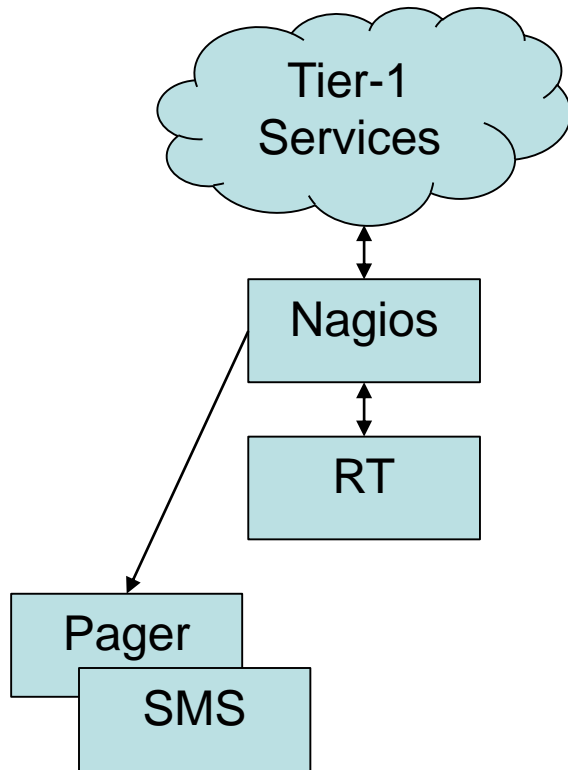
- The new Scientific Computing Services Manager (read as “Cat Herder”).
- Have been in post ~10 months.
- Taking over from Gareth Smith after 10 years.
 - Gareth is retiring March 2019.
- Coming from a commercial background with a career spanning software development thru project management...
- ...I even had a period on the ‘dark side’ - I was a SysAdmin!

Who am I? – New kid on the block

What the new kid is doing

- Responsible for (but not limited to):
 - All the stuff that Gareth do and does.
 - Improving the monitoring of Tier-1 services and providing a more customer (VO), focused service:
 - Be more proactive than reactive.
 - We want to know the problem before the VO does.
 - Question – how do you folks monitor stuff?
 - Covering the position of CMS liaison at RAL:
 - Making up for the number of ATLAS liaisons there seems to be at RAL.
 - Keeping the service running with improvements where required.
 - More of that later....
- If it's the Tier-1 that goes wrong, it's most likely me you are going to be calling!

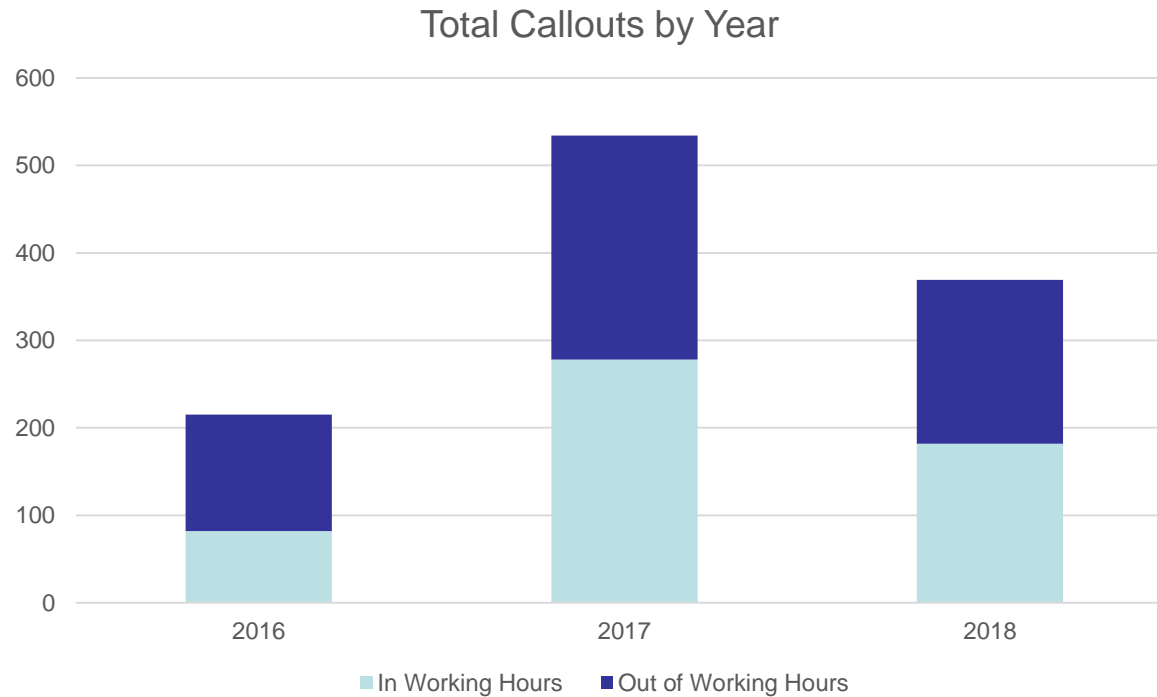
Tier-1 OnCall Services



On-Call – the why, what where an the how

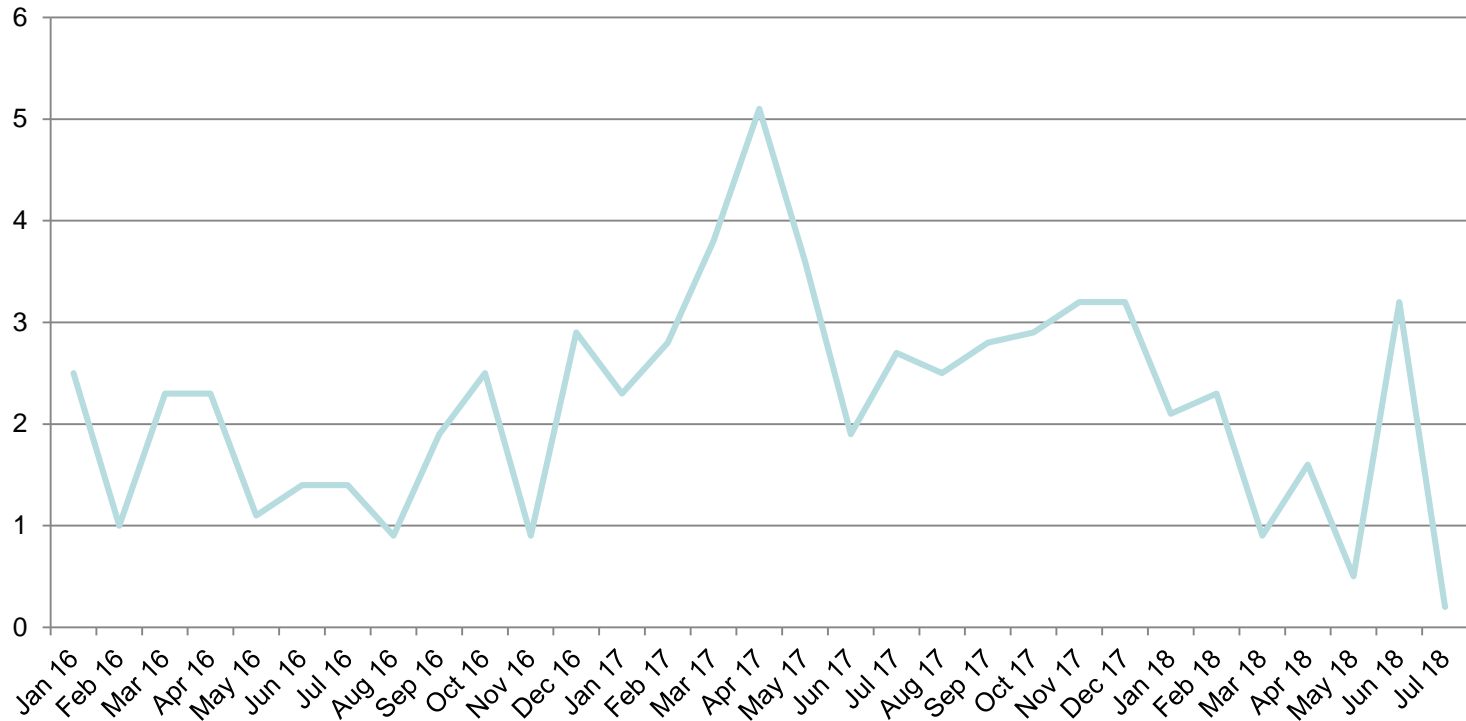
- Currently providing a 24x7x365 VO support for Tier-1 services
 - Continuing to improve the Tier-1 proactive approach to service.
 - Currently provides a good VO response < 2hr, normally <10mins.
- This is done with a two level support approach.
 - The Primary on Call (PoC), performs triage for any given support issue.
 - The PoC either resolves or passes on to the second level support Subject Matter Expert (SME).
 - Current SME's cover Grid, Echo, Castor, Databases and Fabric.
 - Nagios alerts achieved by a dual SMS and pager system.
 - Notified via Nagios
- Will be migrating from Nagios to Icinga.

Tier-1 OnCall Services



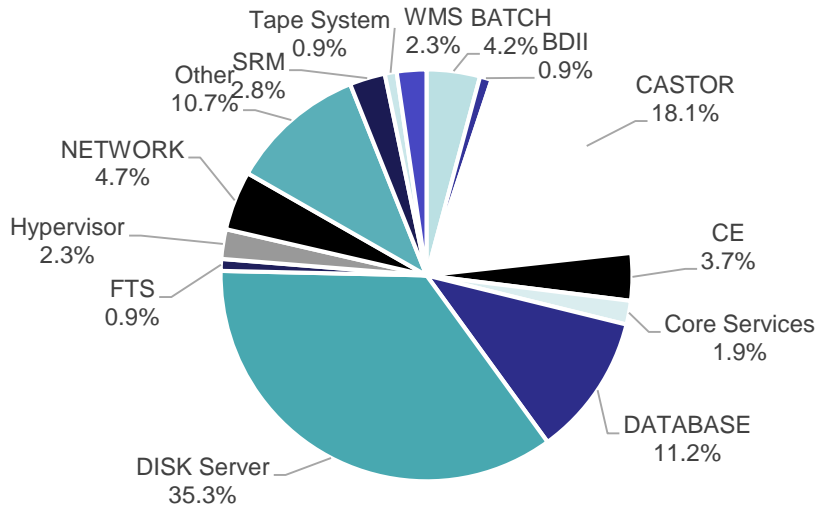
Tier-1 On-Call Services

Number of nights with a callout per week

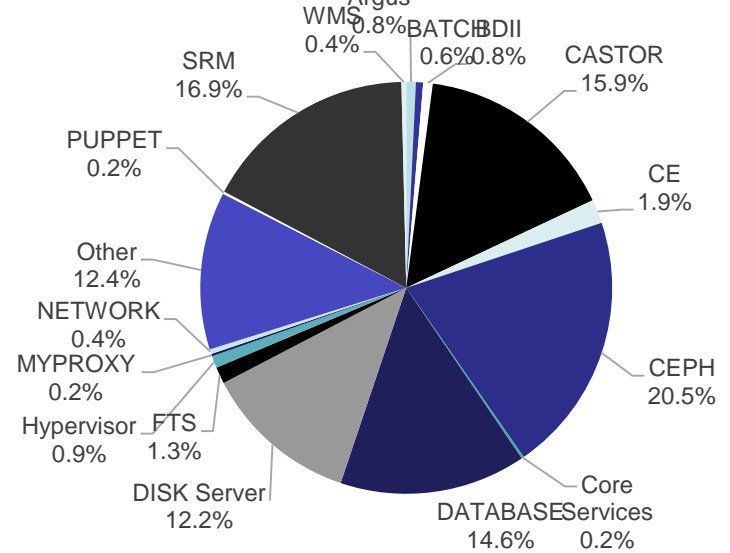


Tier-1 On-Call Services

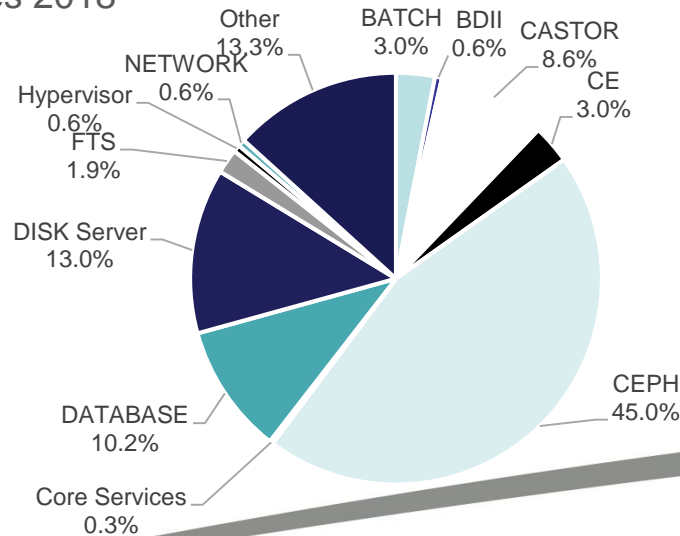
Callouts by Service 2016



Callouts by Services 2017



Callouts by Services 2018



Monitoring with InfluxDB/Grafana

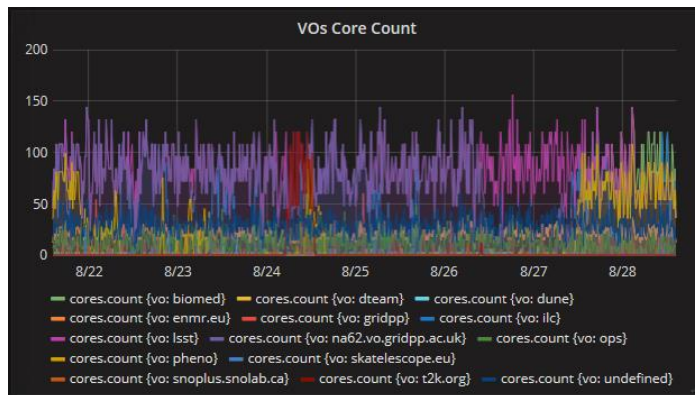


InfluxDB/Grafana to replace Ganglia

- Ganglia no longer a supported Tier-1 service:
 - To be retired by end of Q2 2019.
 - User base now far smaller in Tier-1
 - Not enough resources to support multiple monitoring systems.
- Grafana more flexible and integrates more easily with InfluxDB
- Supported integration includes Ceph, AWS and Docker.

Usage

- Allowing the Tier-1 to monitor large AND small VO's
 - Facilitating better engagement with the smaller VO's
 - Small VO's tend to run lower job counts
 - These counts are monitored weekly
 - Notable changes in usage can be acted upon



Monitoring with InfluxDB/Grafana



Telegraf

Collects time-series data
from a variety of sources

Telegraf

- Is part of the TICK Stack and is a plugin-driven server agent for collecting and reporting metrics.



Science & Technology
Facilities Council

UK Research
and Innovation

Monitoring with InfluxDB/Grafana



InfluxDB

- Used as a data store for any use case involving large amounts of time-stamped data, including DevOps monitoring, log data, application metrics, IoT sensor data, and real-time analytics

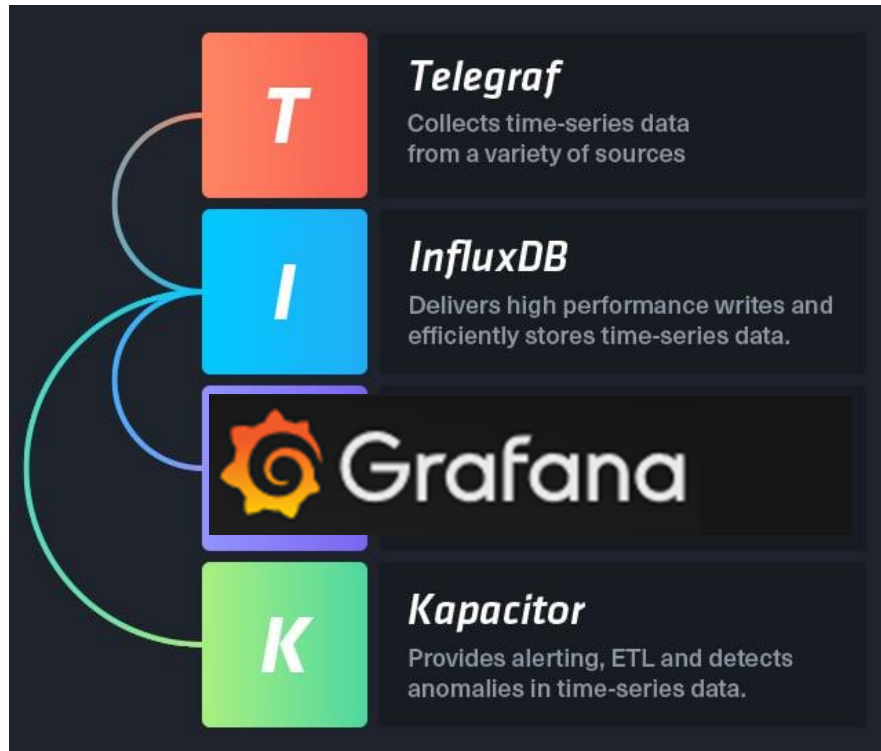
Monitoring with InfluxDB/Grafana



Chronograf

- Chronograf is the user interface component of InfluxData's TICK Stack. It allows you to quickly see the data that you have stored in InfluxDB so you can build robust queries and alerts.
- We replace this with Grafana. This allows us to query, visualize, alert on our metrics.

Monitoring with InfluxDB/Grafana



Kapacitor

- Kapacitor is a native data processing engine in the TICK Stack. It can process both stream and batch data from InfluxDB.

Other Operational news...



Security patching

- Like Sisyphus [2], this is a never ending task.
- Patching taken seriously in Tier-1
 - Patching now considered a security incident.
 - From most machines to ALL machines patched.
 - Using a Pakiti and nagging/moaning combination.
- All SL5 machine to be retired or upgraded by 1st November 2018.

[2] In Greek mythology Sisyphus was the king of Ephyra (now known as Corinth). His punishment for his self-aggrandizing craftiness and deceitfulness was being forced to roll an immense boulder up a hill only for it to roll down when it neared the top, repeating this action for eternity.

In conclusion

- Proactive rather than reactive.
 - We want to know before you do!
- Improving the monitoring and response.
 - Prevention is better than cure
- A more user oriented service.
 - A better VO experience.

Any Questions?