



The night I accidentally killed a Kubernetes service and realized 3 days later

Lucila Stancato

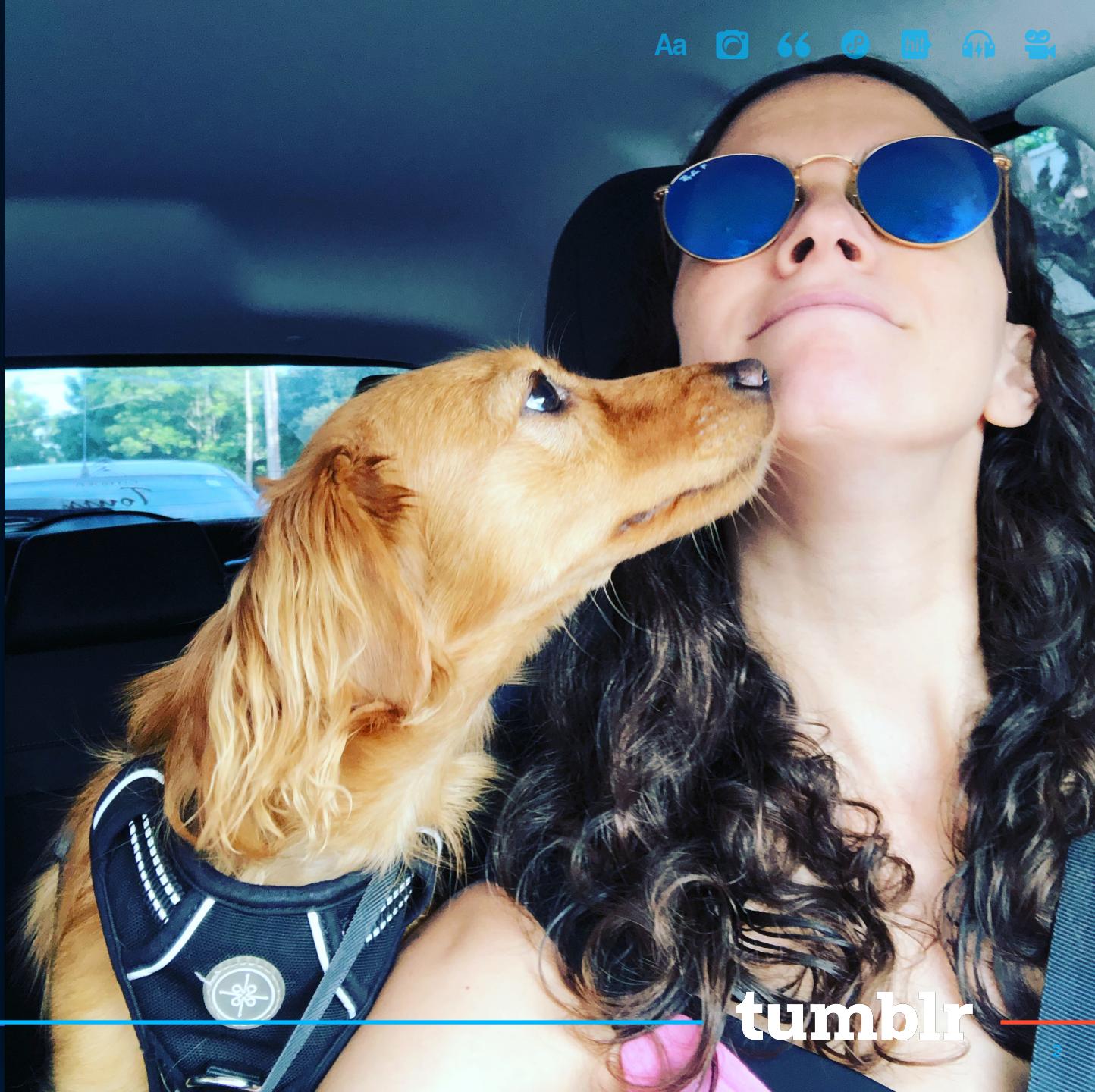
norm Conf



tumblr

Lucila Stancato

Senior Search Engineer at tumblr



tumblr

Tuesday, August 16th 2022

8.28 PM: Incident started

9.54 PM: Incident resolved



“

**Do you remember killing the
Flink job during that incident?**

- my team lead
August 19th, 2022

tumblr



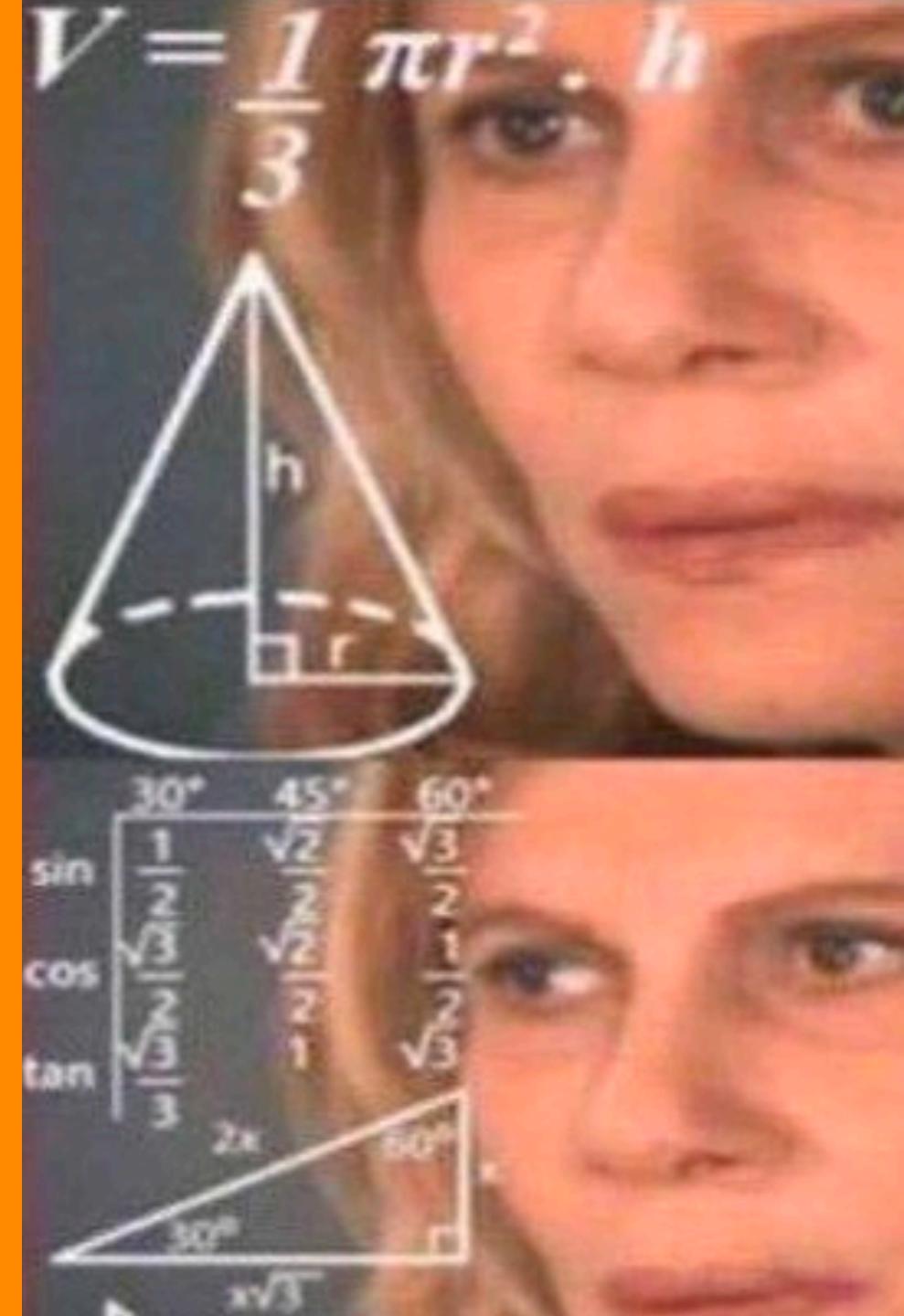
“

Well...

I don't *remember* killing the Flink job. But it is possible that I *did*.

- me
3 minutes later

tumblr



“

I checked my command line history
and I confess:

I killed that Flink job.

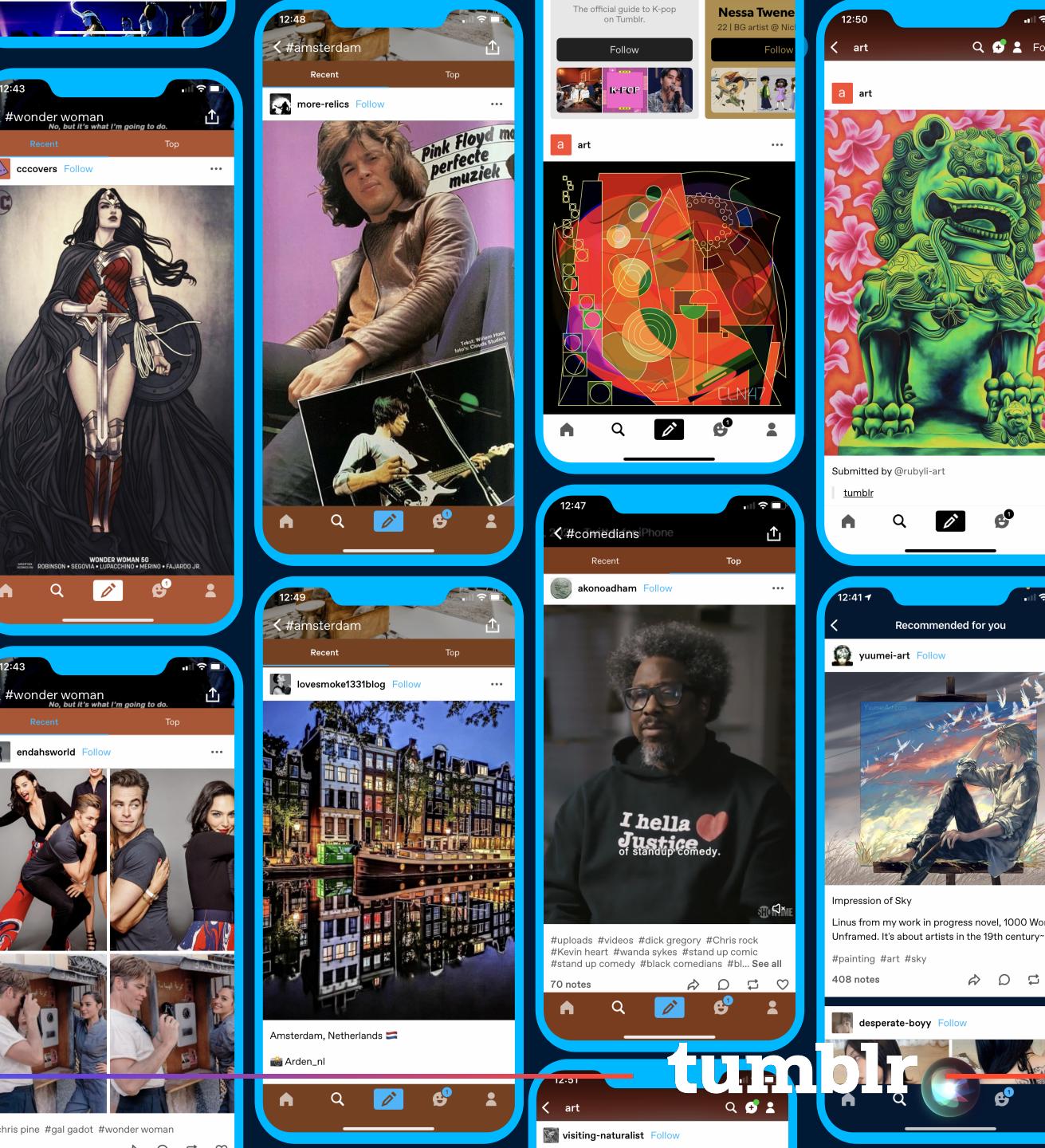
- also me
Later that day

tumblr



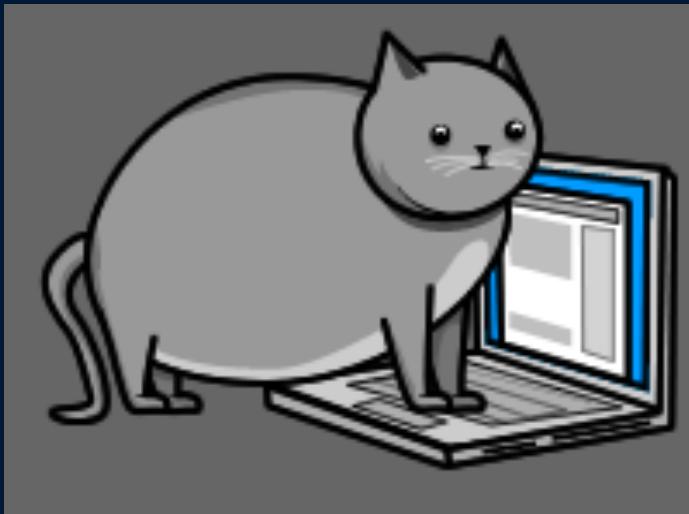
That Flink job

- Stores post impressions on Redis
- We avoid showing posts you have already seen



tumblr

Why we didn't realize the service was down for 3 days



- We monitor **read** error rate
- The service stores data on Redis
- The data was still on Redis but was not getting updated.
- We were not monitoring **write** error rate



What happened during the incident?

A horror story



- A critical service goes down. Making other services fail too.



- I'm on call, so I hear a lovely tune on my phone



- Because this incident is affecting many services, I'm hearing my phone over and over, playing the same lovely tune.



- I acknowledge each page. And start working on these failures one by one.



- Each service has a **runbook** for this type of failures.

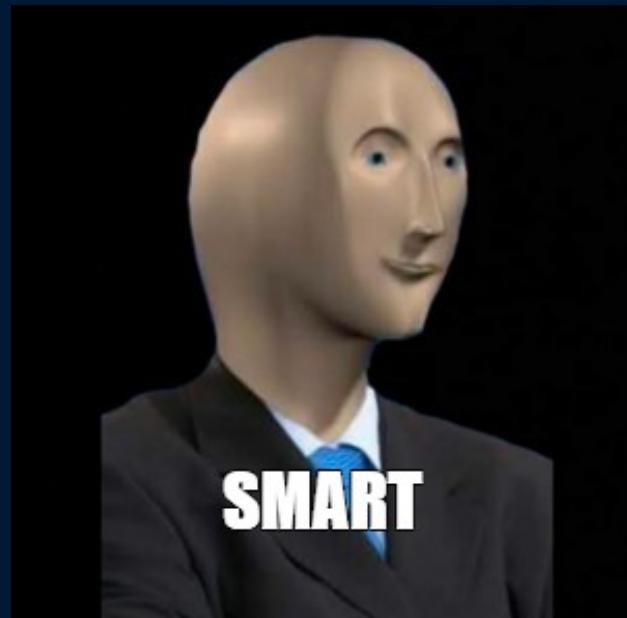


The Runbook

Back to the basics

- A link to the graphs monitoring the service
- **Delete the problematic pod and let k8s restart a new one.**
- If after restarting the service, the insert rate is still not back to normal levels, contact search backend team.

Delete the problematic node



- Google
- Stack Overflow
- k8s documentation
- Ctrl + r “kubectl”



Mystery Solved

This is how killed the Flink job



A good runbook

Tips

- Must be specific.
- Include step by step guide.
- Include command lines.
- Include references to monitoring graphs and what to expect.
- Avoid “everything should look good”.
- Explain what “looks good” means. Give examples, or numbers, acceptable levels.



Thank for your watching!



THE NORMCORE TECH CONFERENCE

tumblr