elastic | The Search Analytics Company

Capacity Constraints Unveiled: Navigating Cloud Scaling Realities

Kevin Sonney

Senior SRE, Capacity

@ksonney.bsky.social

@ksonney@redwombat.social

@ksonney

elastic

Marc-Andre Dufresne

Principal SRE, Capacity

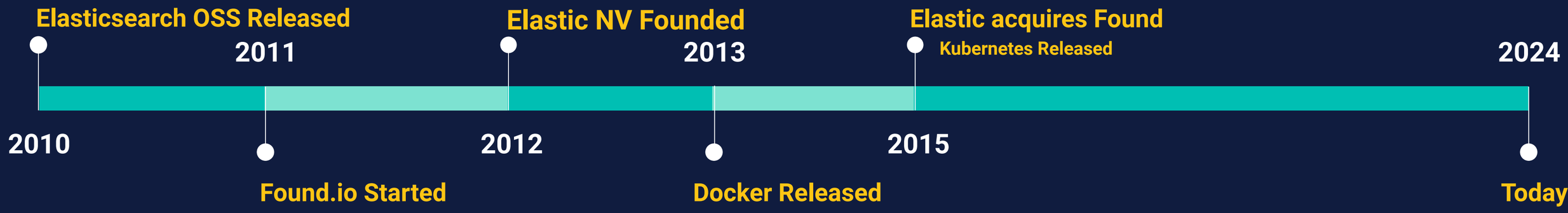
elastic

Who is Elastic?
What is Elastic Cloud?

elastic

A Brief History of ~~Time~~ Elastic Cloud

(Not to scale)



elastic

In the Dark Time, there were machines

In the dark time before “the cloud” there were servers, data centers, and physical constraints.

The cloud changed all that. An endless supply of computing that someone else maintained, at a reasonable price.

Or at least that’s what we were told...

A brief history of the Dark Time

In the old days, there was a long-term plan

Request a budget, get approvals, and place an order

Wait on build, shipment, and delivery

Rack ‘em, stack ‘em, image ‘em, and run

Rinse and repeat every quarter

KEEP
CALM
because
EVERYTHING IS
AWESOME!

But in today's world, it's better!

- No need to plan for major purchases
- Click the button and deploy in seconds
- Need more? Just click!
- Automated growth: GitOps, Infrastructure as Code, ChatOps
- You don't even need to click anymore!

And we all lived happily ever after.

The End

If that was really the end, you wouldn’t be here

elastic

Elastic Cloud Today

64
regions

4
CSPs

40K
hosts

50K
Elastic Deployments

ALL OF THIS IS GROWING DAILY

elastic

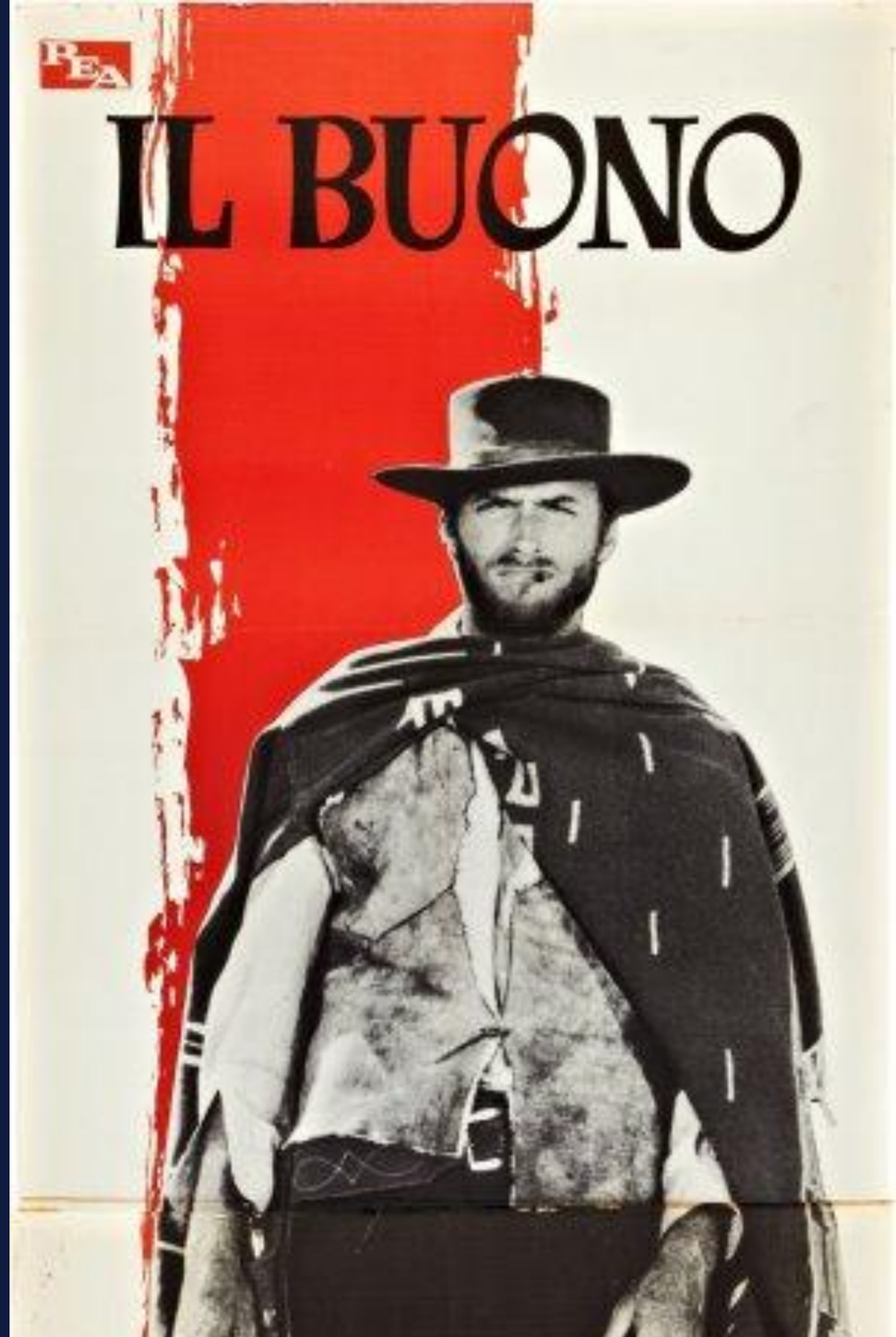
“We have a 💩-TON of containers”

(Author’s Note: both metric and imperial 💩-tons)

elastic

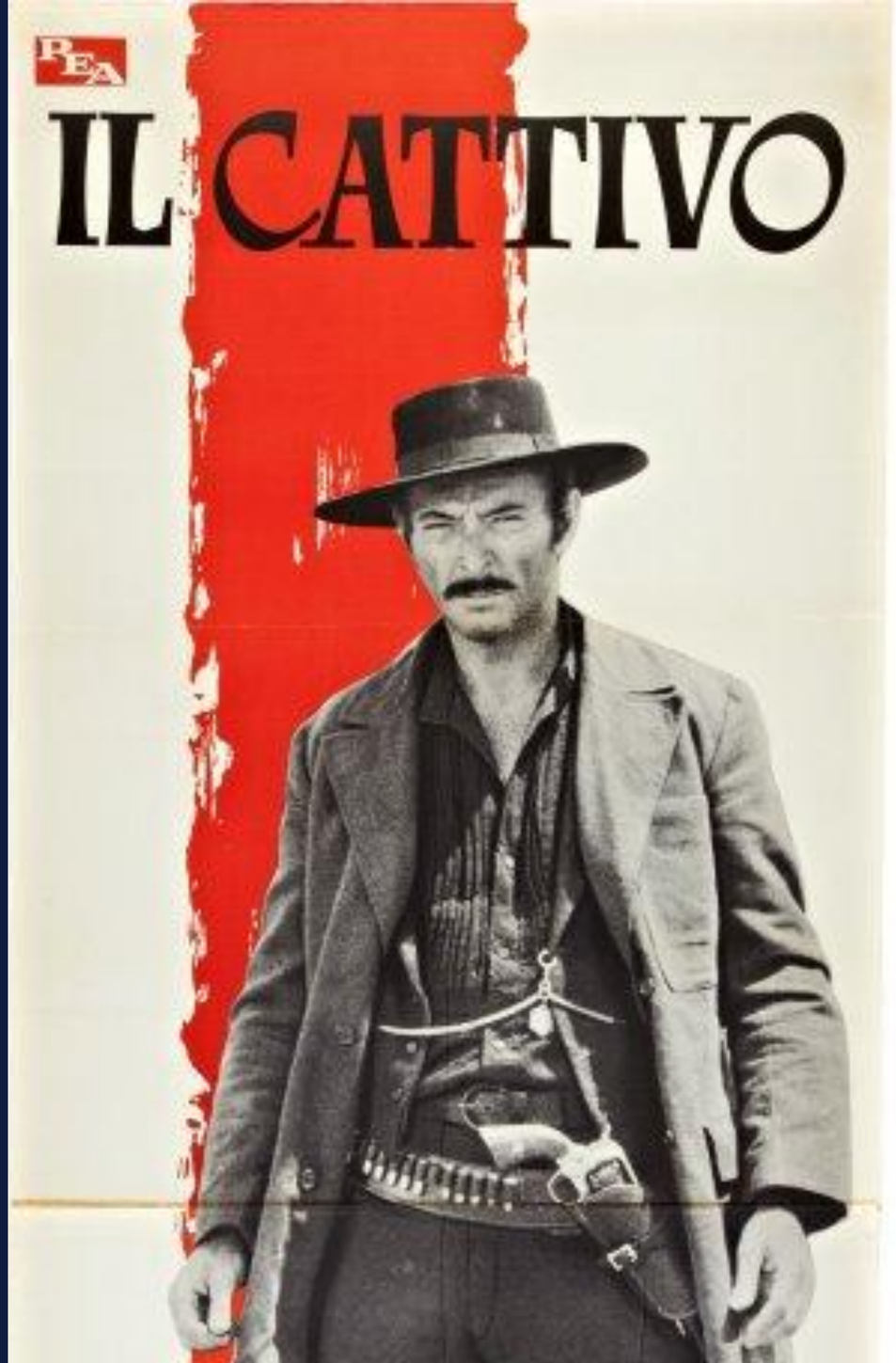
Elastic Cloud: The Good

- Containerized
- Orchestrated
- Transparent Infrastructure
- Customers love it!



Elastic Cloud:
The Bad

- Custom tooling to detect and update needed capacity
- Stateful
- Expensive to run



Elastic Cloud:
The Ugly

- Can’t use CSP provided auto-scaling
- Capacity availability is limited in many regions
- On-call still has to address capacity issues

Wait, out of Capacity?
But the Cloud is Infinite Computers, Right?

elastic

Things we have seen in the wild

elastic

Things we have seen in the wild

```
Status Reason: We currently do not have sufficient [type] capacity in the
Availability Zone you requested ([zone]). Our system will be working on
provisioning additional capacity. You can  currently get [type] capacity by
not specifying an Availability Zone in your request or choosing [zones].
```

elastic

Things we have seen in the wild

```
The zone 'projects/[project]/zones/[zone]' does not have enough resources
available to fulfill the request.
```

elastic

How did THAT happen?

elastic

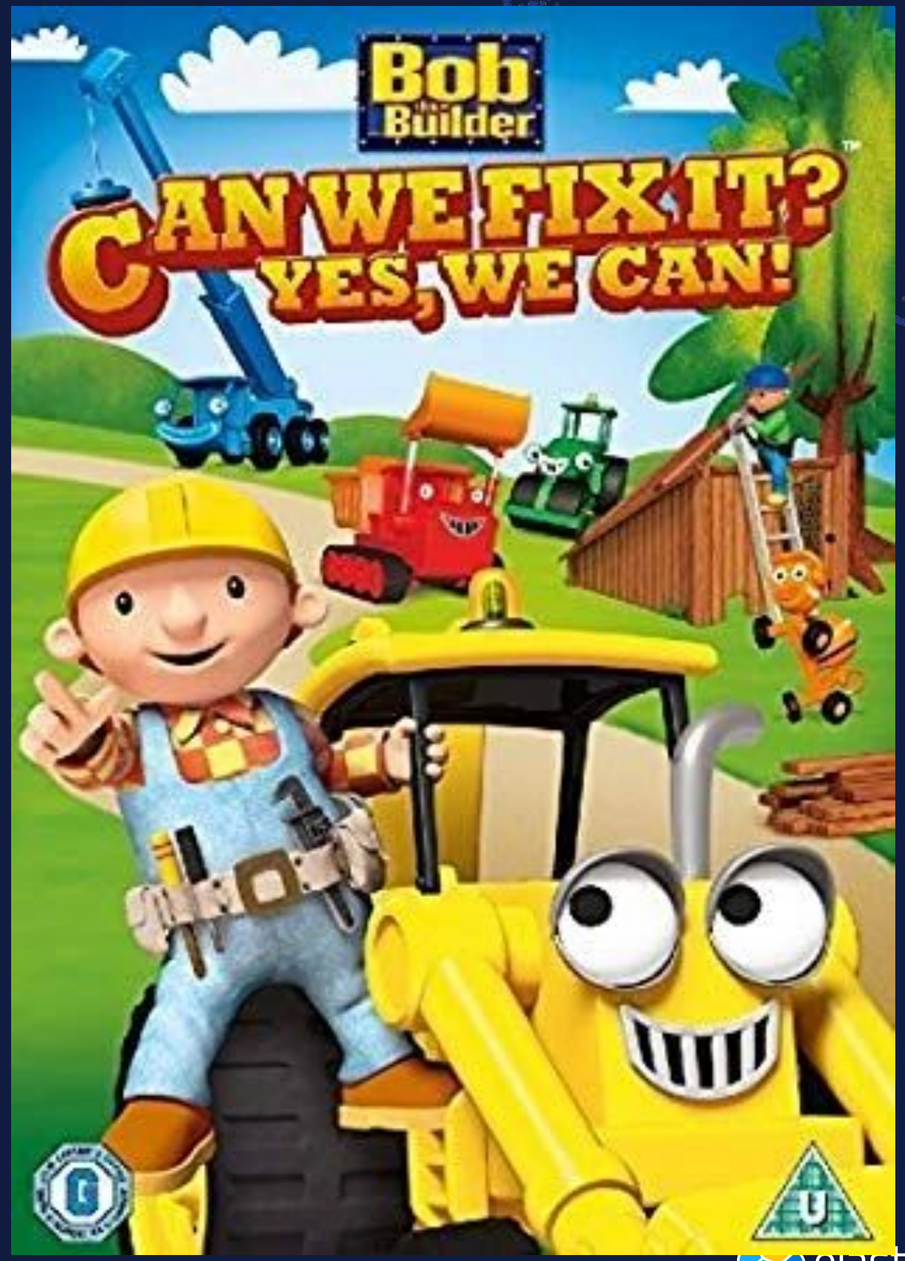
Live and Learn: Mistakes we've made

- Instance types tied to Elastic deployment types
- Not planning for high demand events
- Not building CSP relationships to plan capacity EARLY
- Not planning for CSP maintenance events
- Sunsetting older instance types
- Implementing newer instance types

elastic

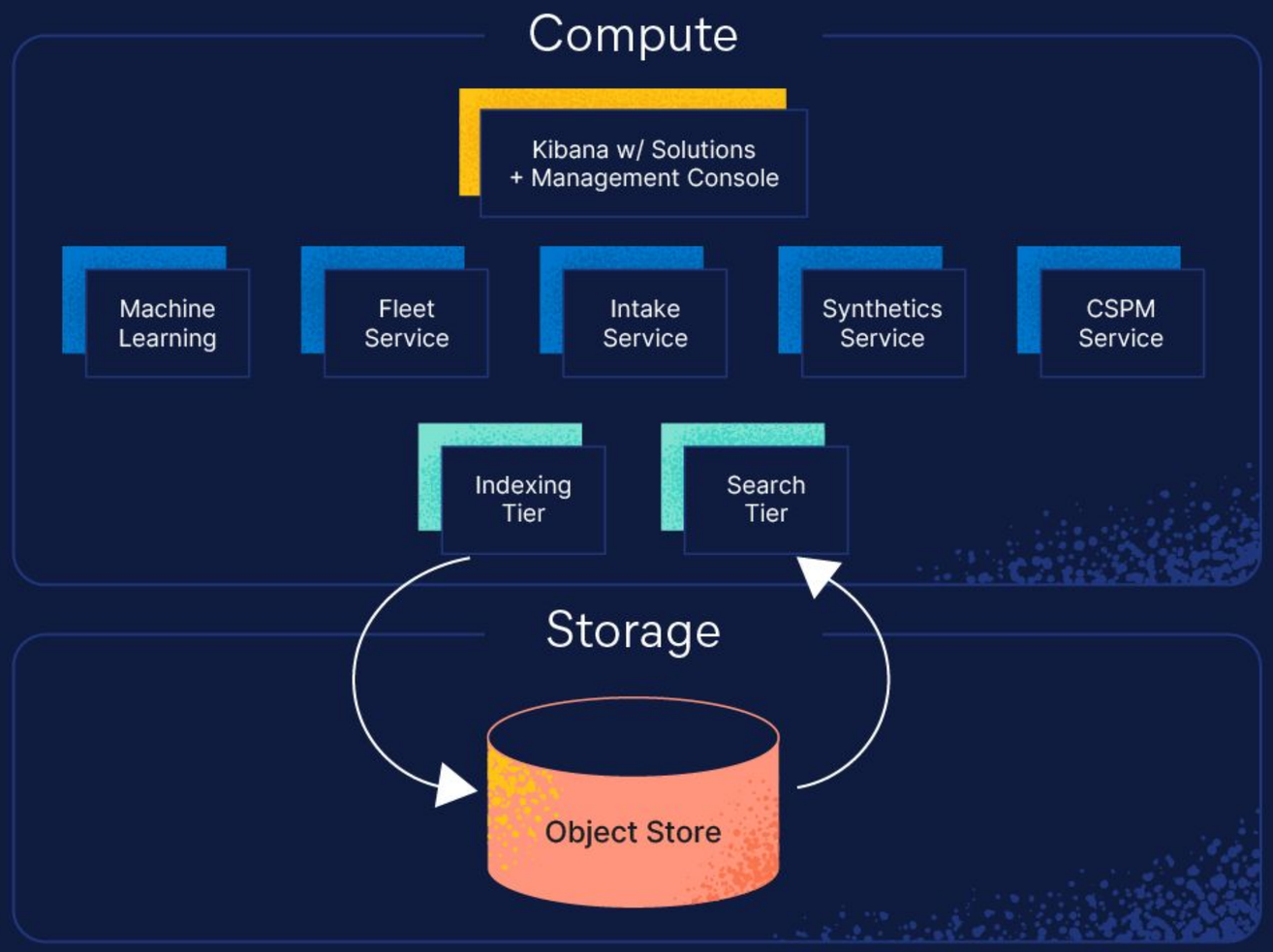
How are we fixing all that?

Can we fix it? YES WE CAN!



elastic

Elasticsearch Serverless



elastic

Get Capacity Now!

- Reserve Instances
- Reserve Instances in Advance!

elastic

Plan and Communicate

- Planning for high demand events with Customers and CSPs
- Roadmaps for new instance types
- Automating proactive actions for CSP maintenance events

elastic

Our Approach to Capacity Management

Taken From Lessons Learned

elastic

Concrete Take Home Actions

- Probe Capacity Reservations
- Start using Future Capacity Reservations
- Plan Availability Zones support
- Project/Plan/Reserve
- Design for flexibility

elastic

AMA Time!

Brunhilde has questions, do you?

elastic

Thank you!