
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

STG222

Using data analytics to improve product quality and customer satisfaction

Seshu Adunuthula

Senior Director, Head of Data Platform
Coupang

Aldi Tjahjadi

Director, Data Intelligence
Coupang

Seth Markle

Senior Principal Engineer
Amazon S3

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Building at Amazon

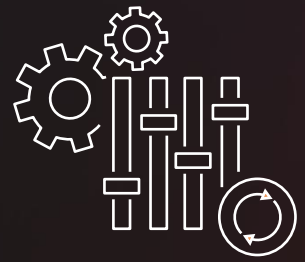


Huge scale

235+ microservices

Over 280 trillion objects

Averages 100 million requests per second



Request path

About a dozen services

Well-defined boundaries

Great model for ownership



Performance

Key business objective

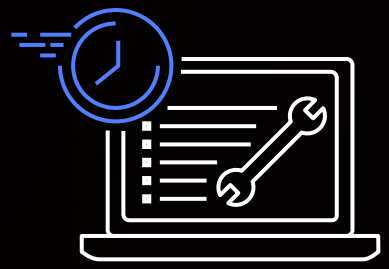
End-to-end view

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is performance?

Dimensions

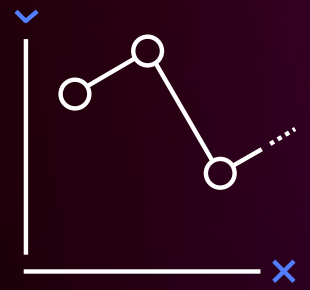


Latency

Throughput

Request rates

Measurement



Magnitude

Variance

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Tail latency matters

Data lake workloads are sensitive to latencies

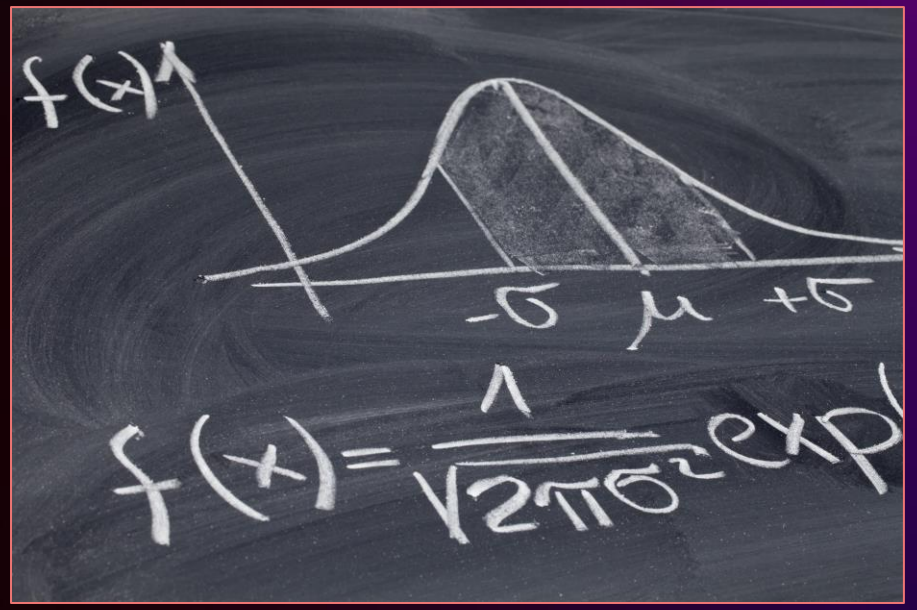
Outliers (99.9+ percentile) **matter**

To improve, we need to **root cause**

Also . . .

Each microservice writes its logs **separately**

We need to analyze **exabytes** of log data



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Where does latency come from?



Hops

Placement groups < intra AZ < inter AZ < inter Region



Storage media

HDDs: seek times, rotational latency

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Let’s talk about hard drives

Rotational Latency

Seek Time

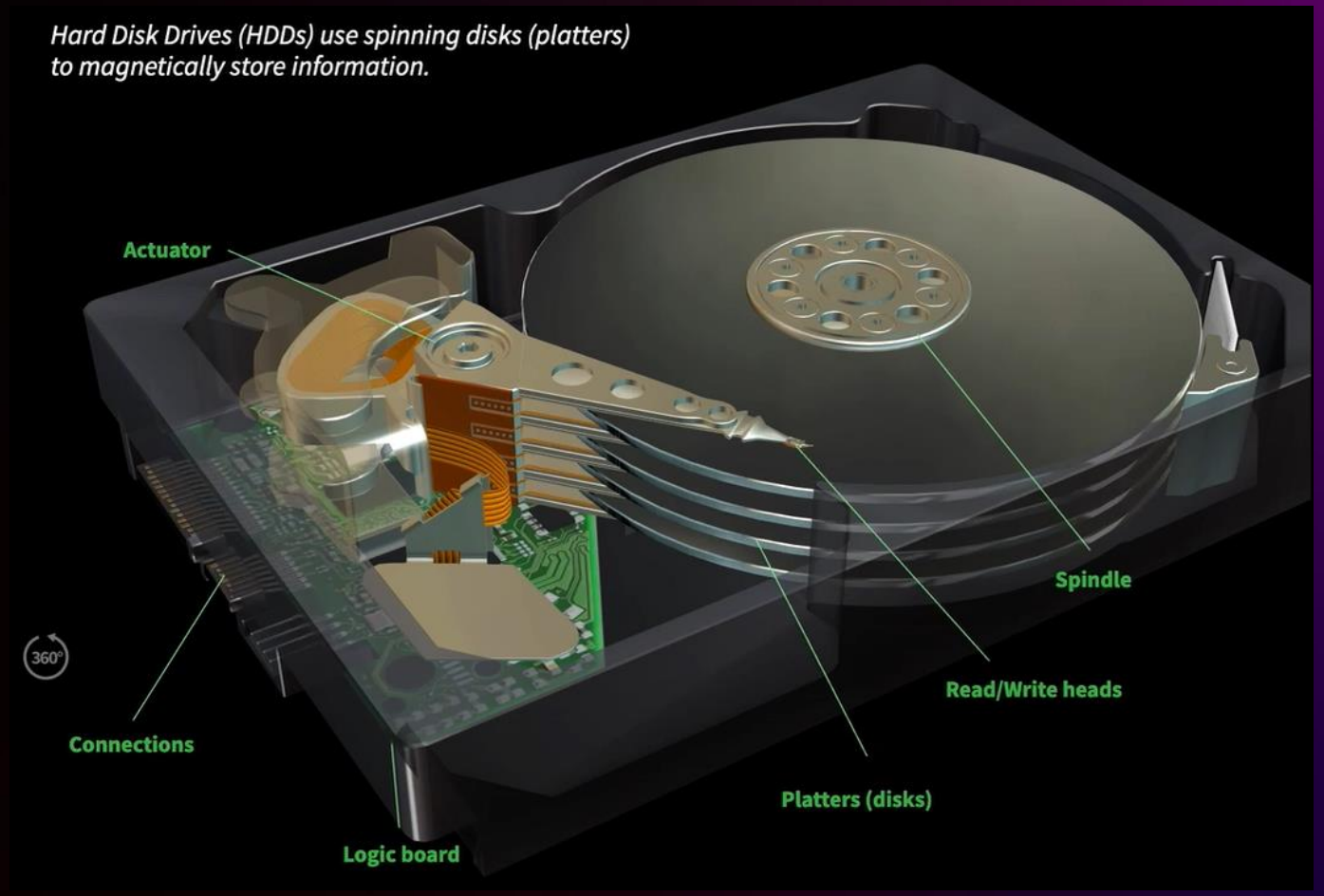


Image by Animagraffs
https://animagraffs.com/hard-disk-drive/

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Where does latency come from?



Hops

Placement groups < intra AZ < inter AZ < inter Region



Storage media

HDDs: seek times, rotational latency



Computational overhead

Crypto, signing, garbage collection, etc.



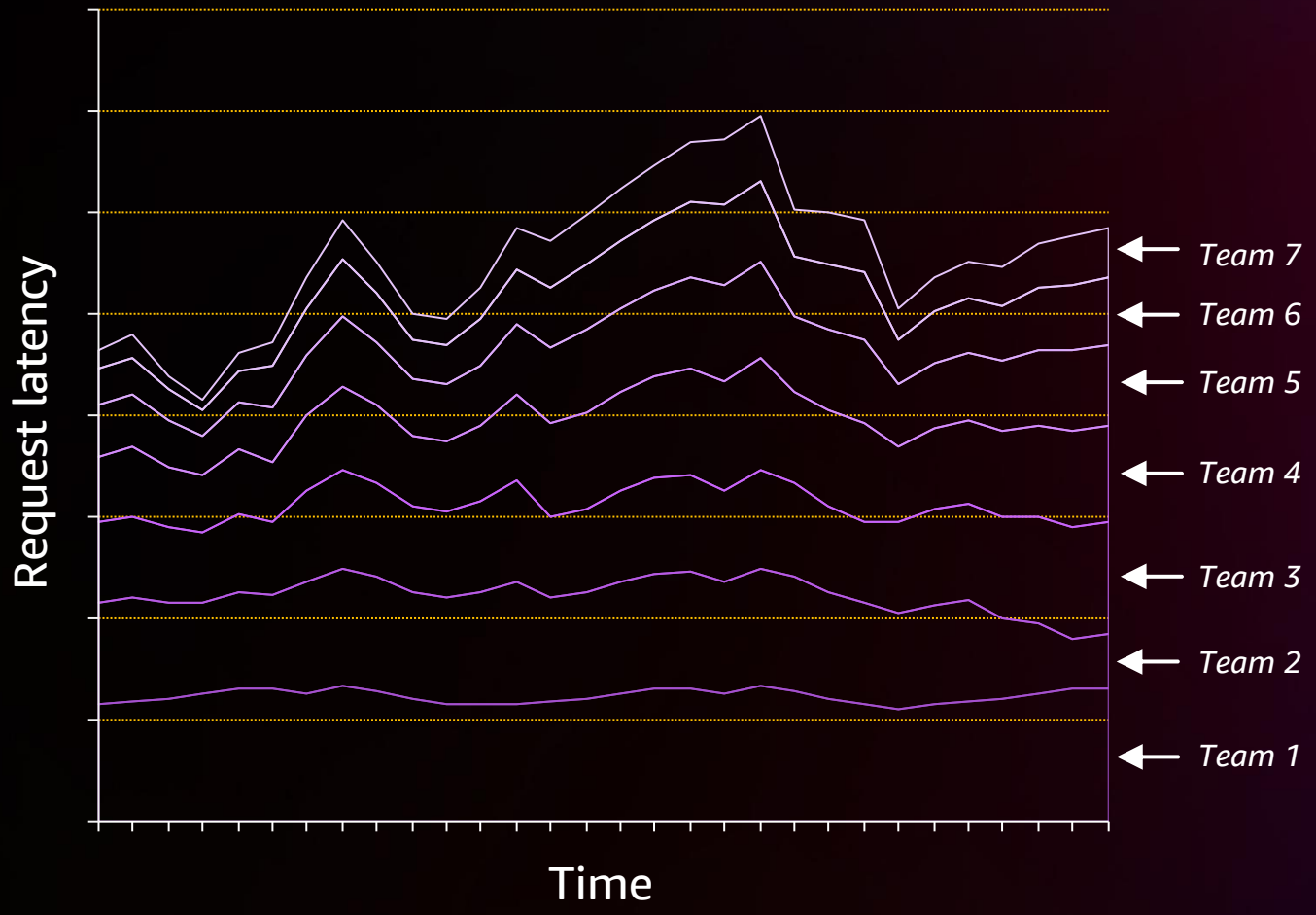
Queuing

Time spent waiting adds latency

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Providing insights to Amazon S3 dev teams



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

S3 performance profiling in 3 easy steps

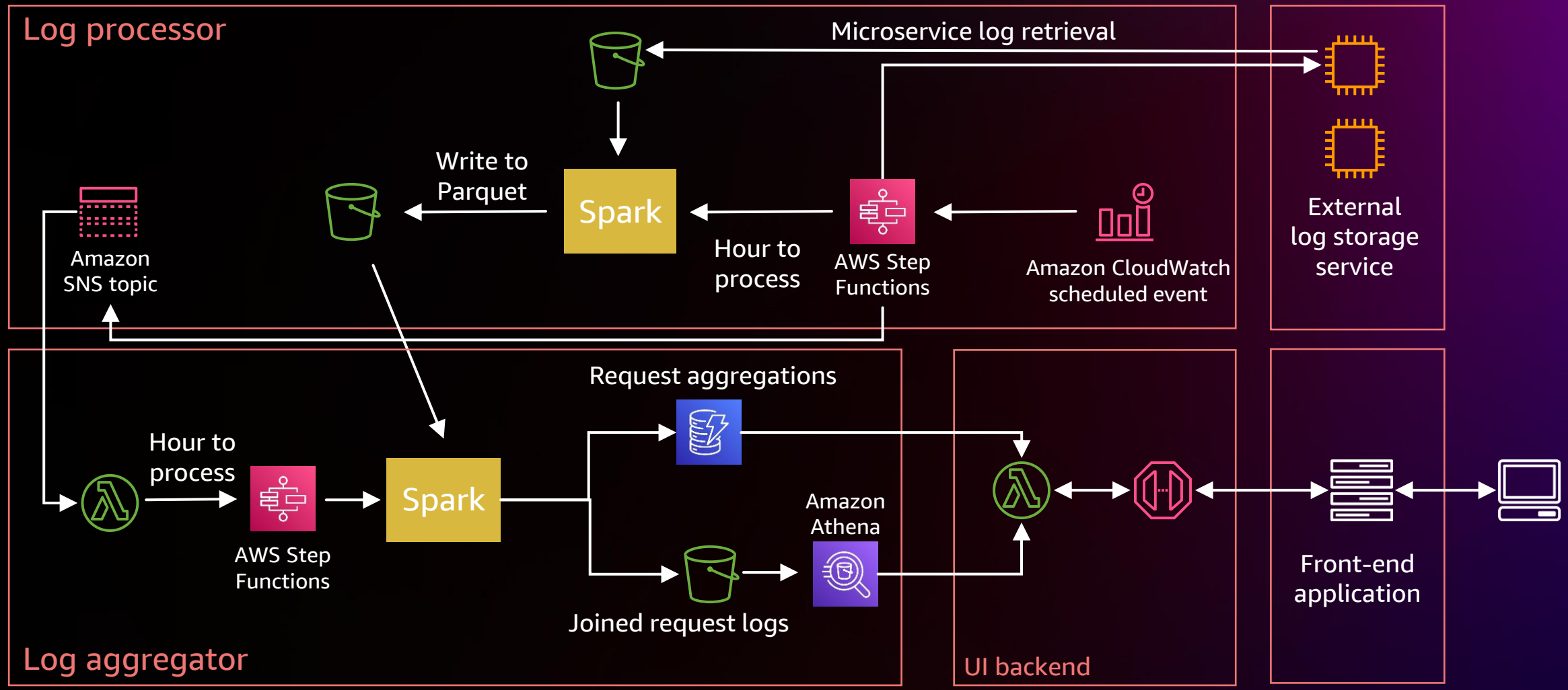
1. Pull and prepare **petabytes of log data** from thousands of hosts
2. Reassemble and measure the latency of **trillions of requests**, based on the logs of microservices
3. **Display results** in a user-friendly way, in a completely **self-service tool**

EVERY HOUR

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What we built: Our architecture



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Success stories

Timeout tuning

Hardware/Kernel combinations

Capacity scaling enhancements

Unintended Outcomes

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Let’s talk about hard drives, again

Hard drives are getting denser

Ability to access the data isn’t increasing meaningfully

IO scarcity

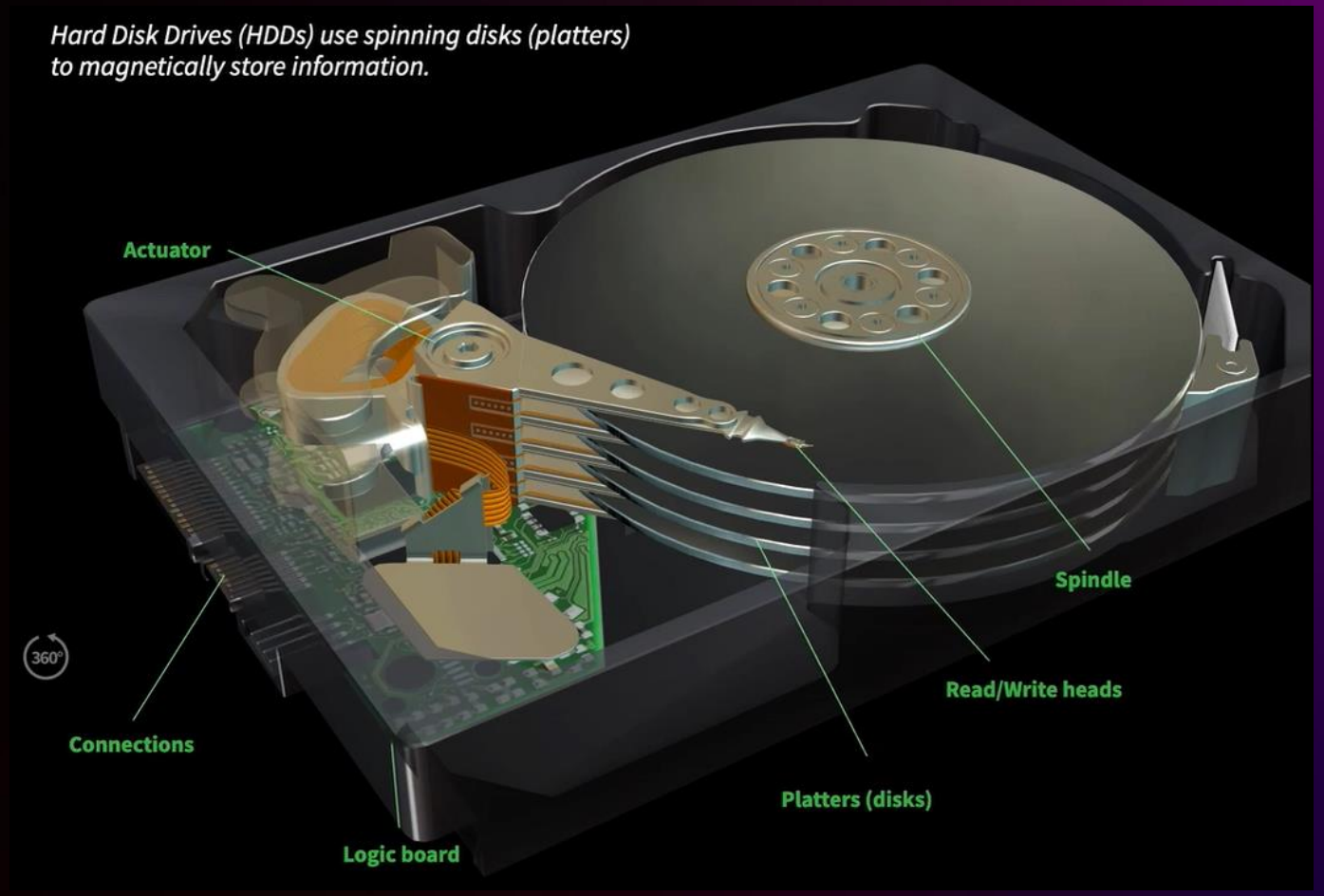


Image by Animagraffs
https://animagraffs.com/hard-disk-drive/

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Let’s talk about hard drives, again

Hard drives are getting denser

Ability to access the data isn’t increasing meaningfully

- IO scarcity

Access density

- 250 GB drive =~ 400 IOPS/TB
- 20 TB drive =~ 5 IOPS/TB (80x reduction!)

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

IO: A mental model

IO improvement levers:

- Increase the supply
- Decrease the demand
- Spread better

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Access patterns drive IO

- Workload observations
  - Read frequency
  - Object lifetimes

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

coupang

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Business areas

Coupang

Coupang Eats

Coupang Play

Coupang Pay

Jet Delivery

No. of employees



60K persons

Source: National Pension Service (as of June 2022)

Annual sales



$18.4 billion

(increased by 54% YoY)

Source: Coupang Quarterly Report (as of 2021 Q4)

Quarterly sales



$5.0 billion

(increased by 12% YoY)

Source: Coupang Quarterly Report (as of 2022 Q2)

Logistics infrastructure scale



3.9km^2

nationwide

Souce: Coupang Annual Report (as of 2021)

Regional investment



Establishing 1.4km^2 FCs

in year 2021

Source: Coupang (as of 2021 Q4)

Customers

17.9M persons

Coupang Active Customers*

*Active Customers: A customer who has made a purchase at least once in the last three months

Source: Coupang Quarterly Report (as of 2022 Q2)

9000K persons

Wow members

Source: Coupang Quarterly Report (as of 2021 Q4)

Rocket Delivery



Millions of products

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Coupang data platform overview

31.5 petabytes of data

1,300+ users

120K jobs per day

500K queries per day

7.1M logs per minute ingested

55K EMR nodes per day

150K data sets

5,500 certified data sets

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

A/B testing

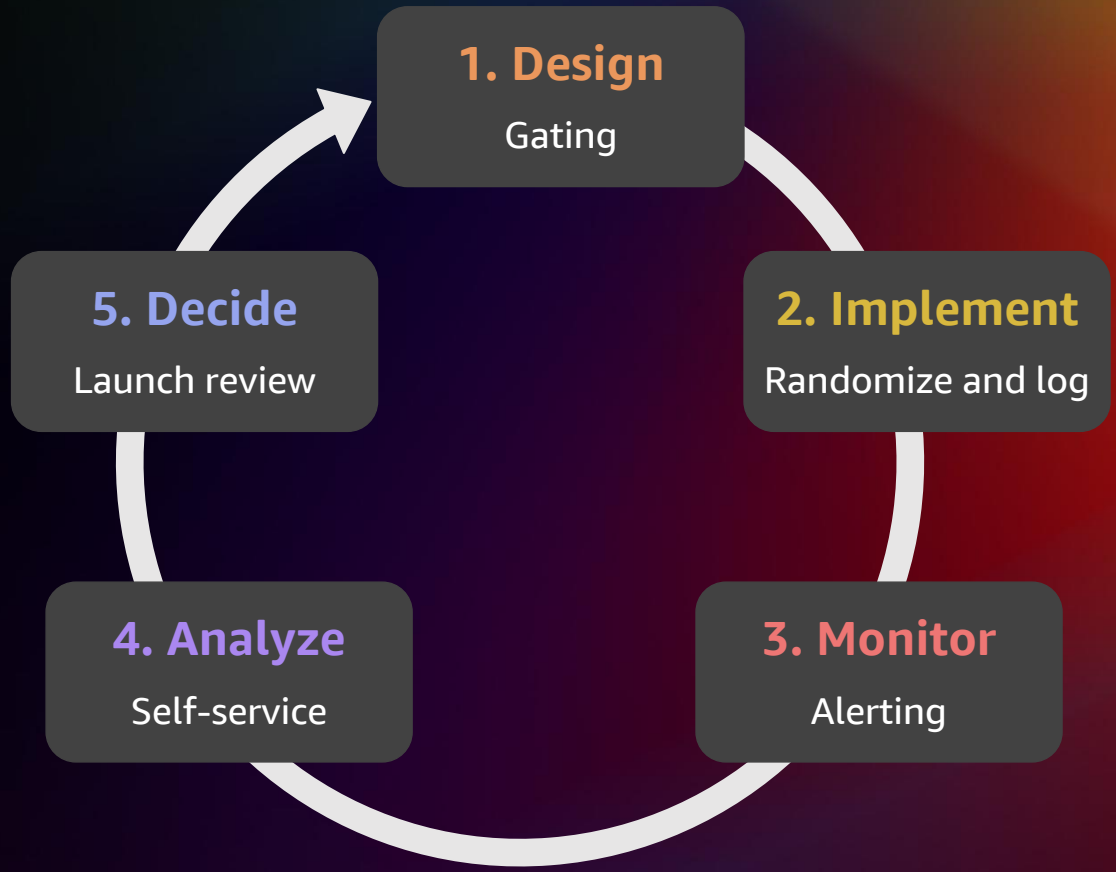
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

“How did I ever live without Coupang?”

Customer experience feedback

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Experiment life cycle

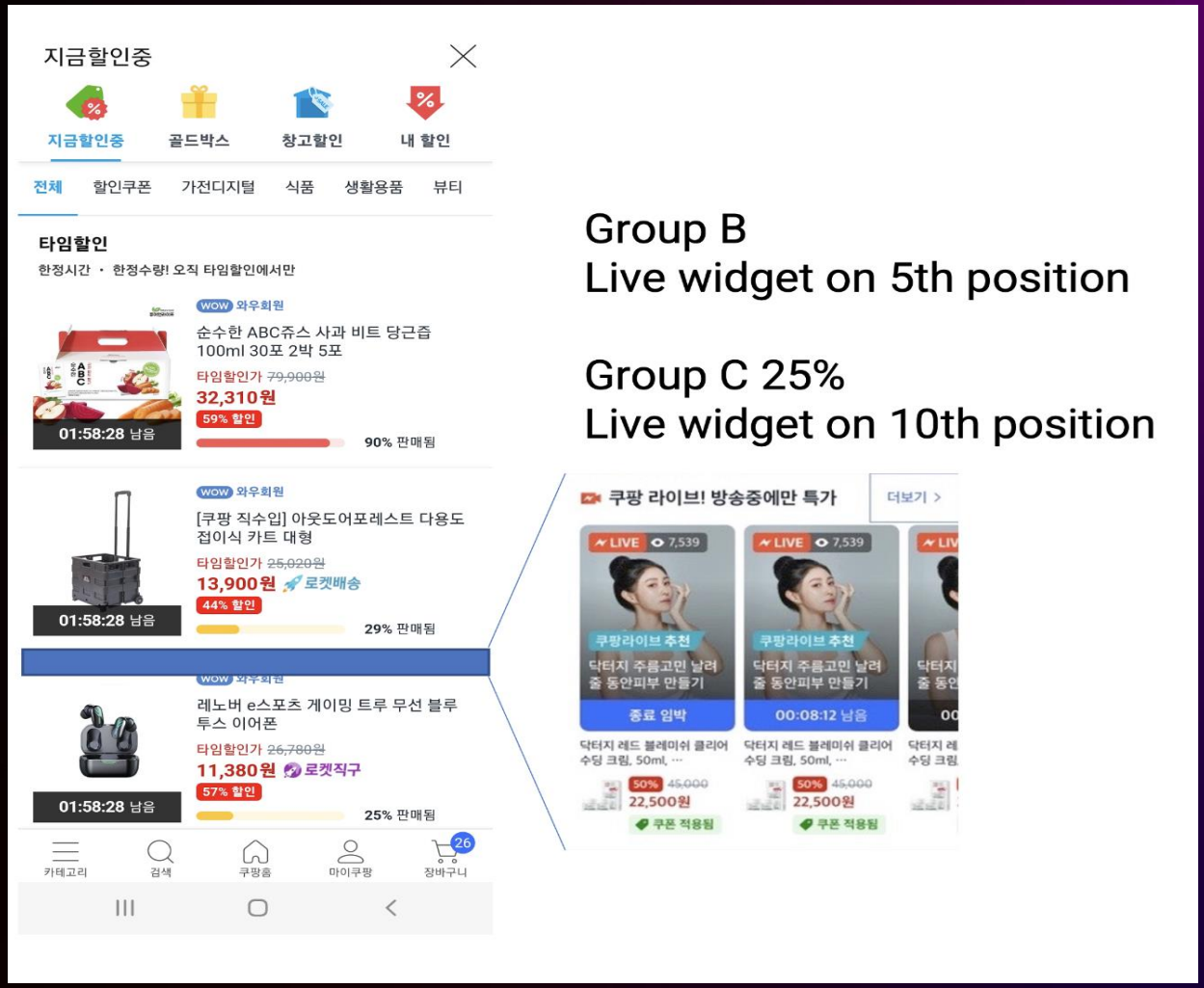


aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Experimentations on live widget

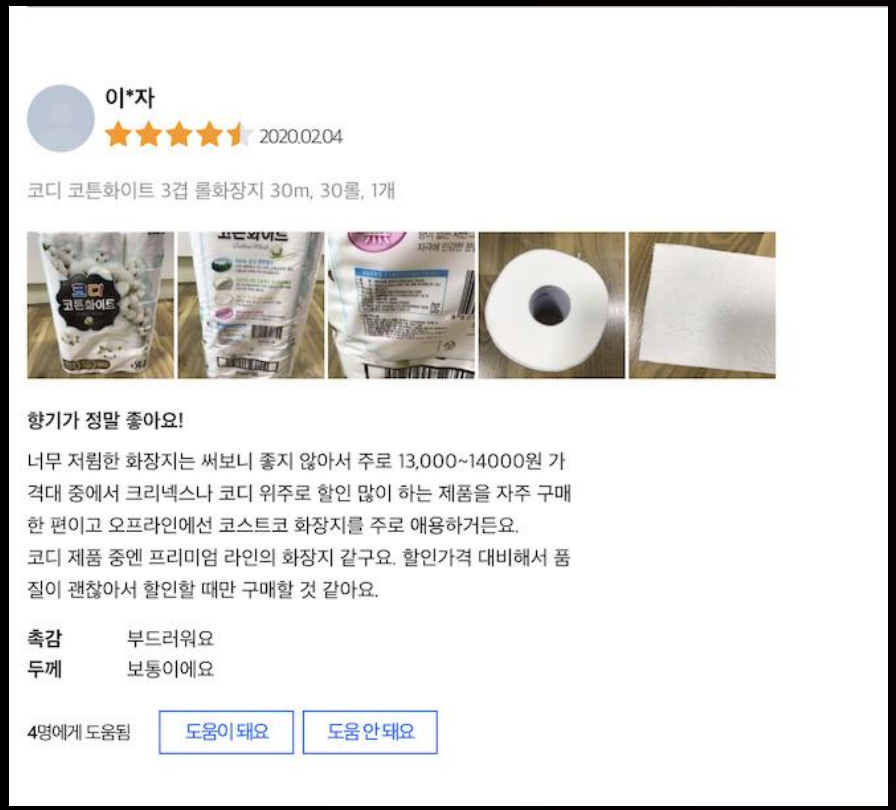
Control: without live widget

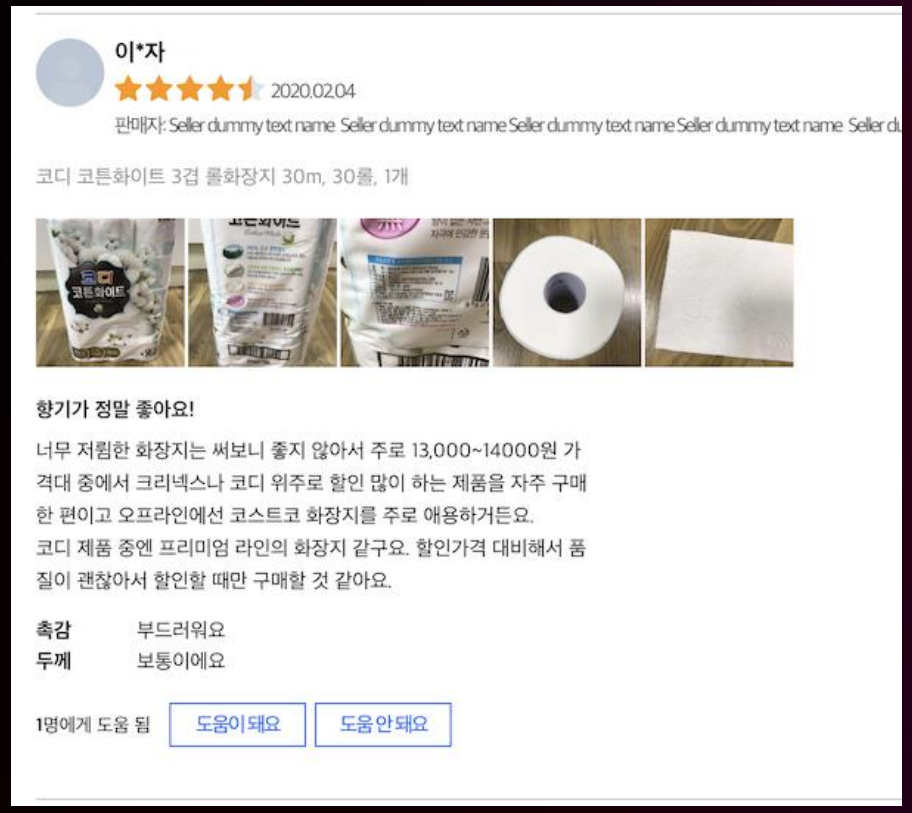
Increase buyer conversion and overall spend per customer



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Experiment: Seller name on review section





By providing seller information on each product review:

Increased customer trust

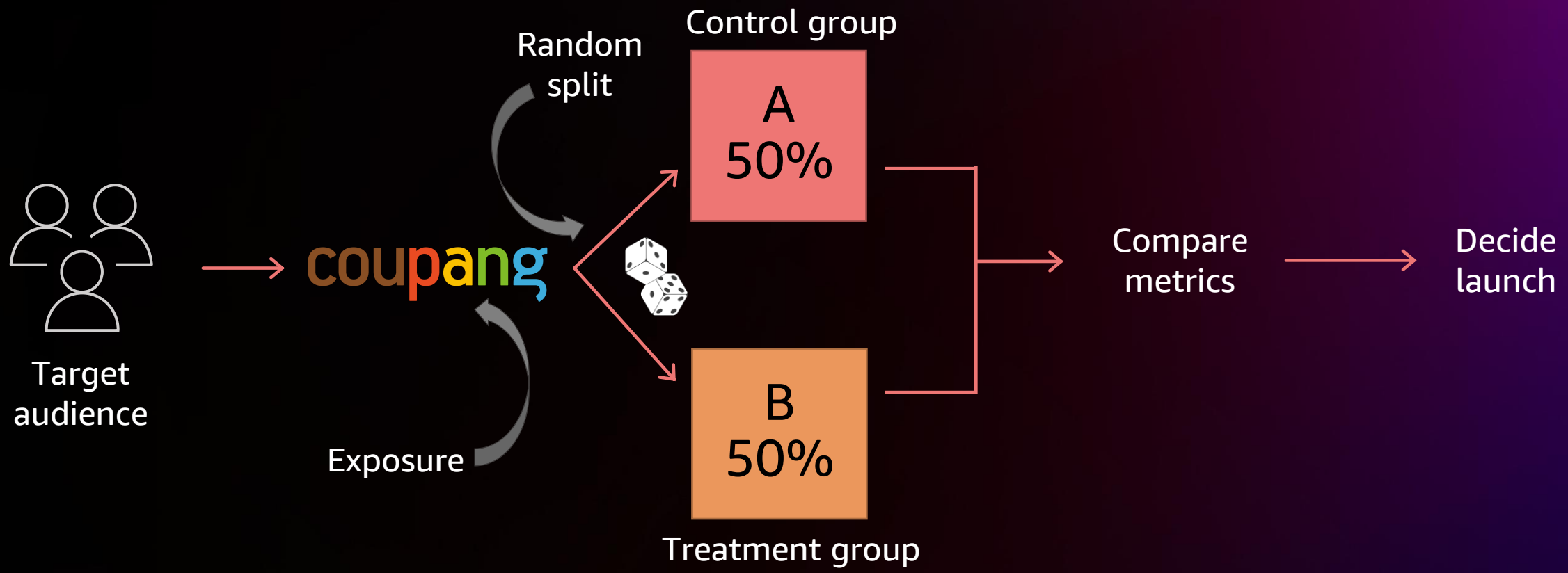
Increased percentage of customer revisits

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

A/B testing architecture

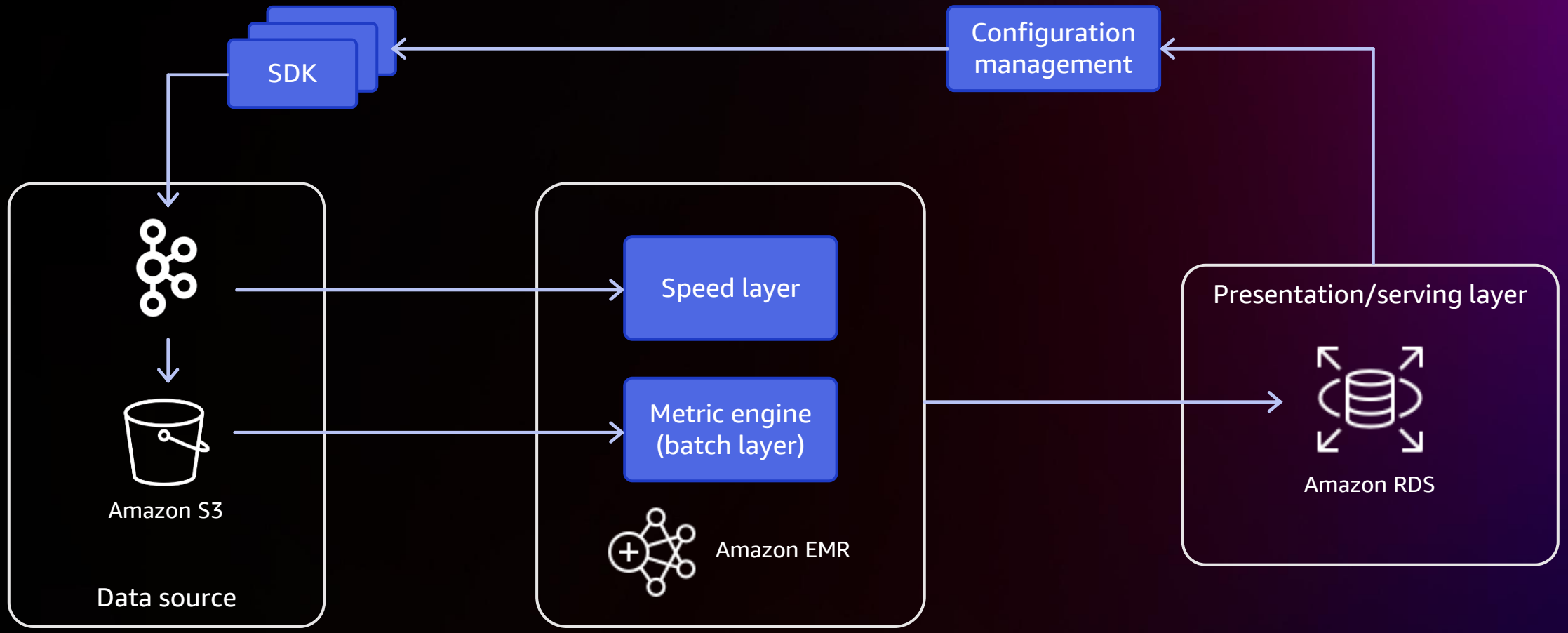
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is A/B testing?



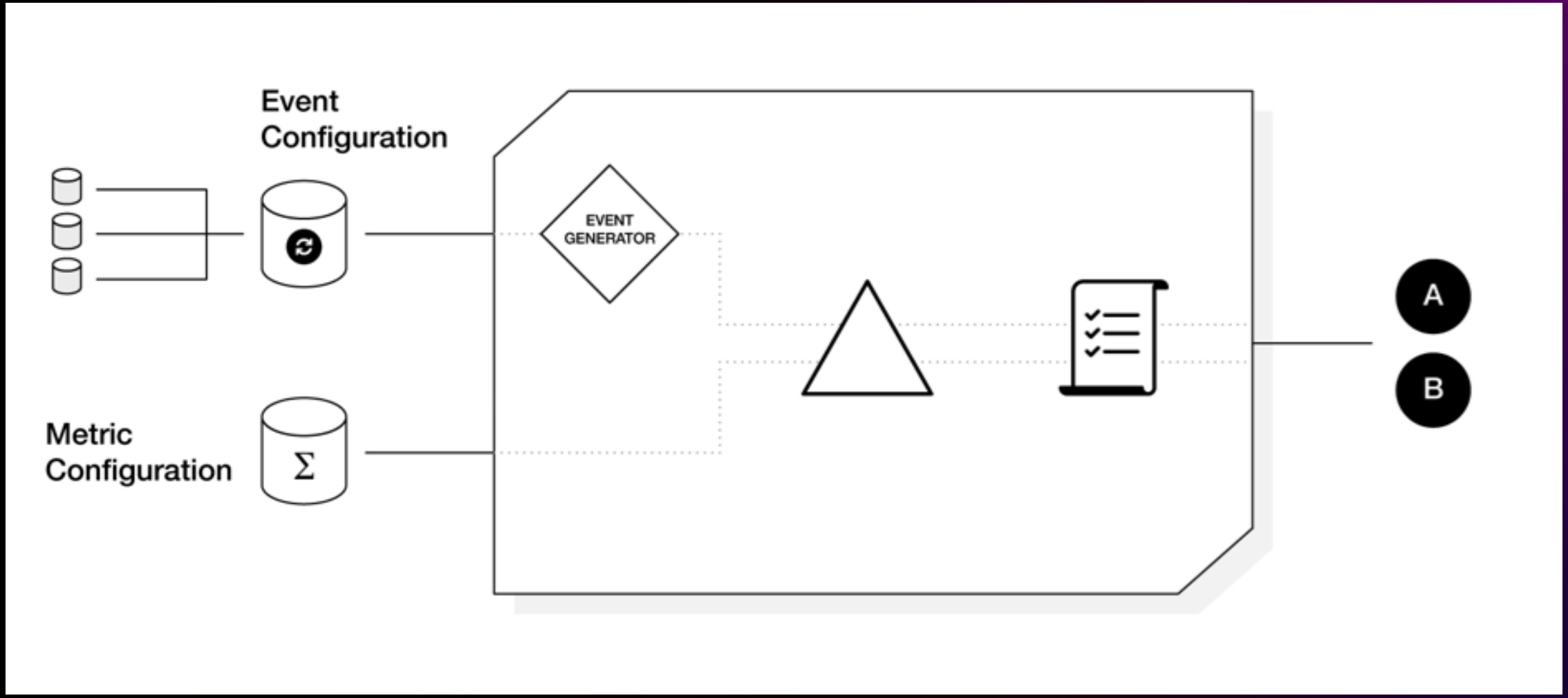
aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

A/B test infrastructure



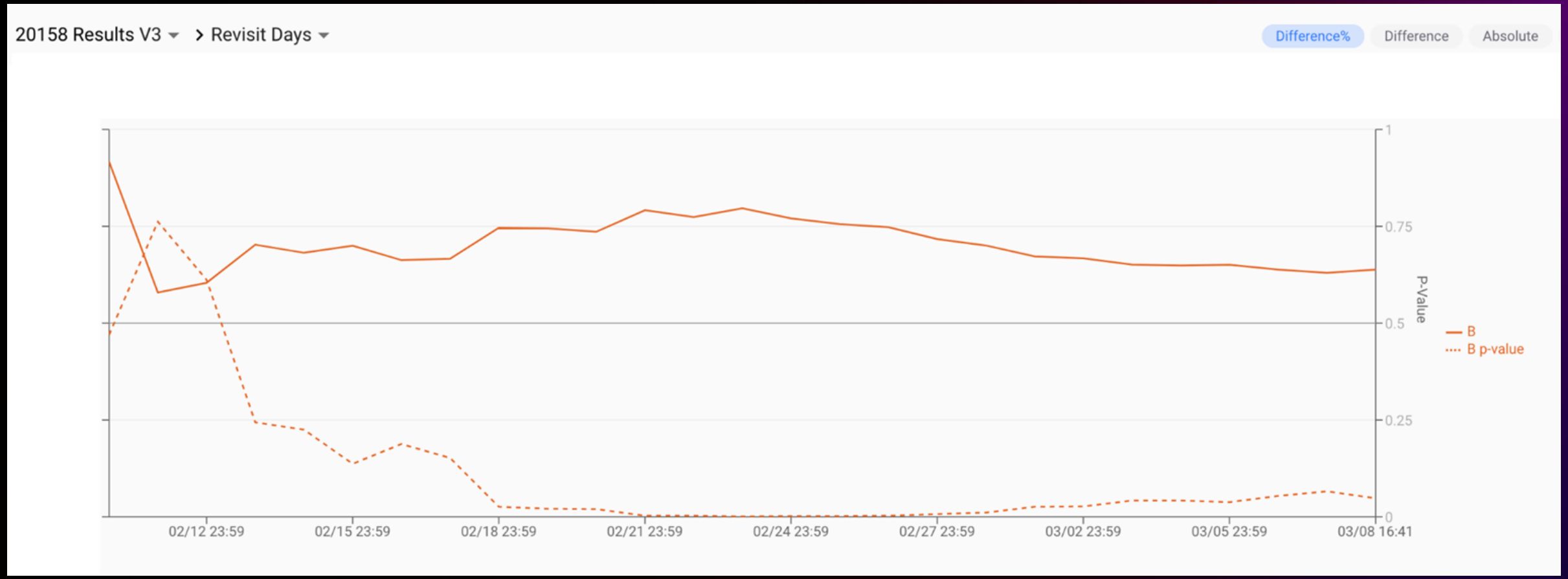
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Batch layer – metrics engine



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Revisit days – adding seller info on review

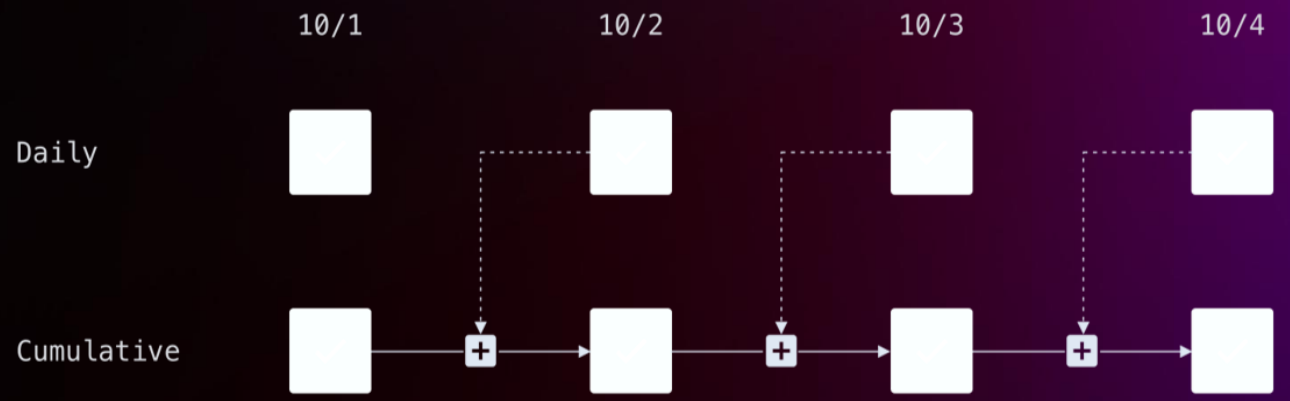


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lesson learned: Cumulative calculation

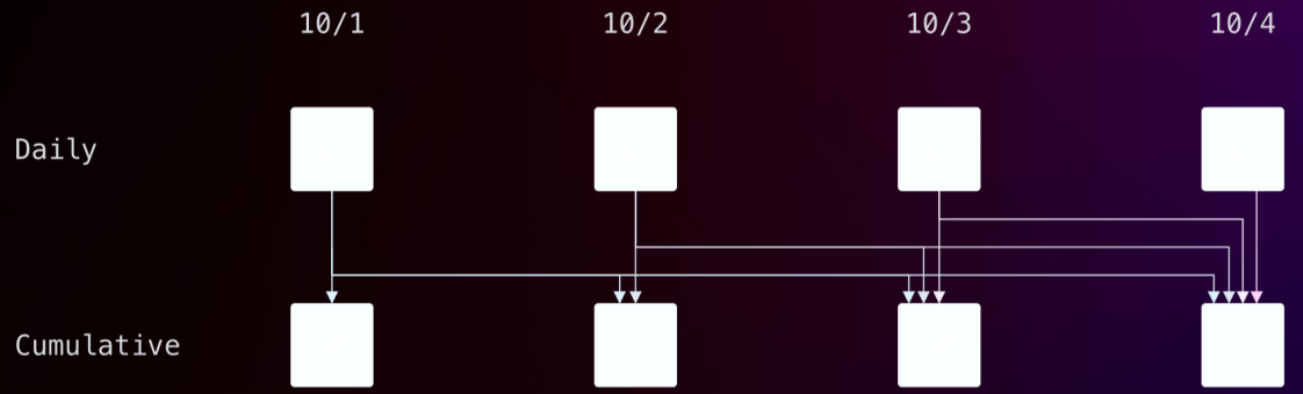
Incremental calculation

- Faster processing



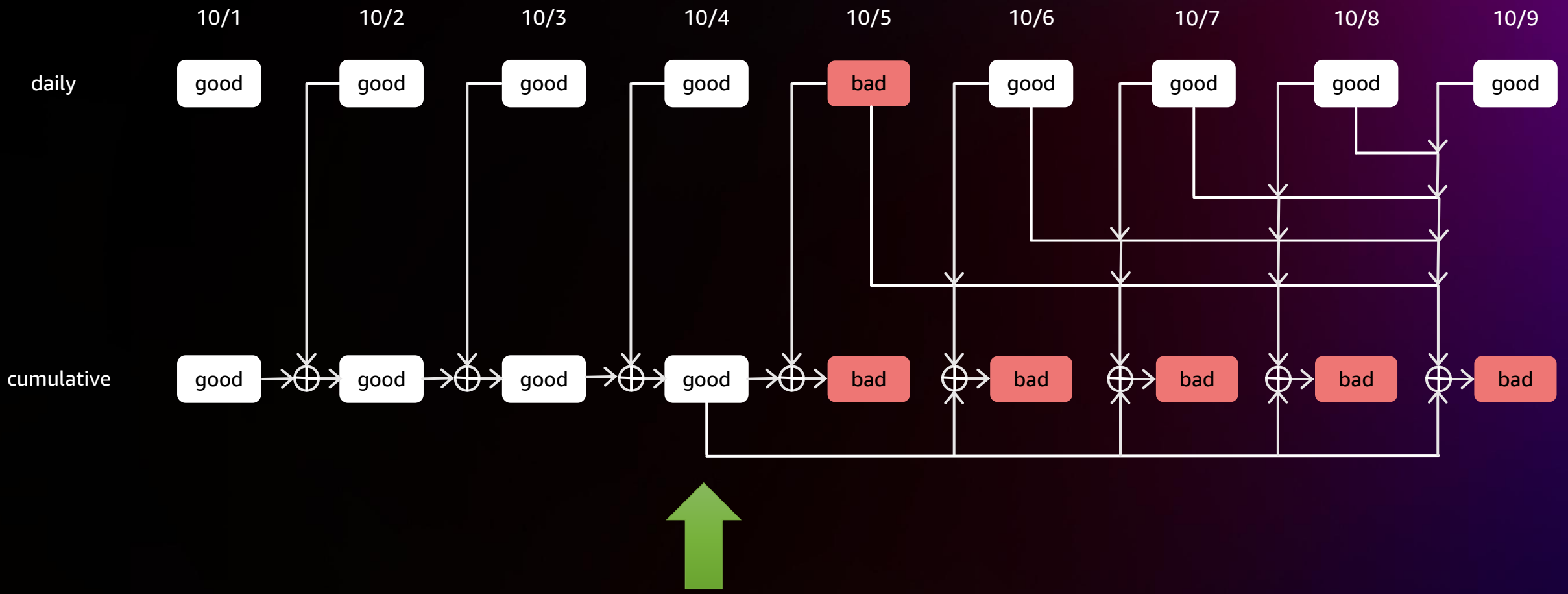
Non-incremental calculation

- Fault tolerance
- Scale calculation on Amazon EMR



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Incremental calculation with state



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Common Themes



- Plan for resumability
- Go serverless, if appropriate
- Sometimes you find what you’re not looking for

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Seshu Adunuthula
seshu@coupang.com

Aldi Tjahjadi
altjahja@coupang.com

Seth Markle
smarkle@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.