
AWS
re:Invent

BLD209

Inversion of scale: Outnumbered but in control

Joe Magerramov

Senior Principal Engineer

AWS

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

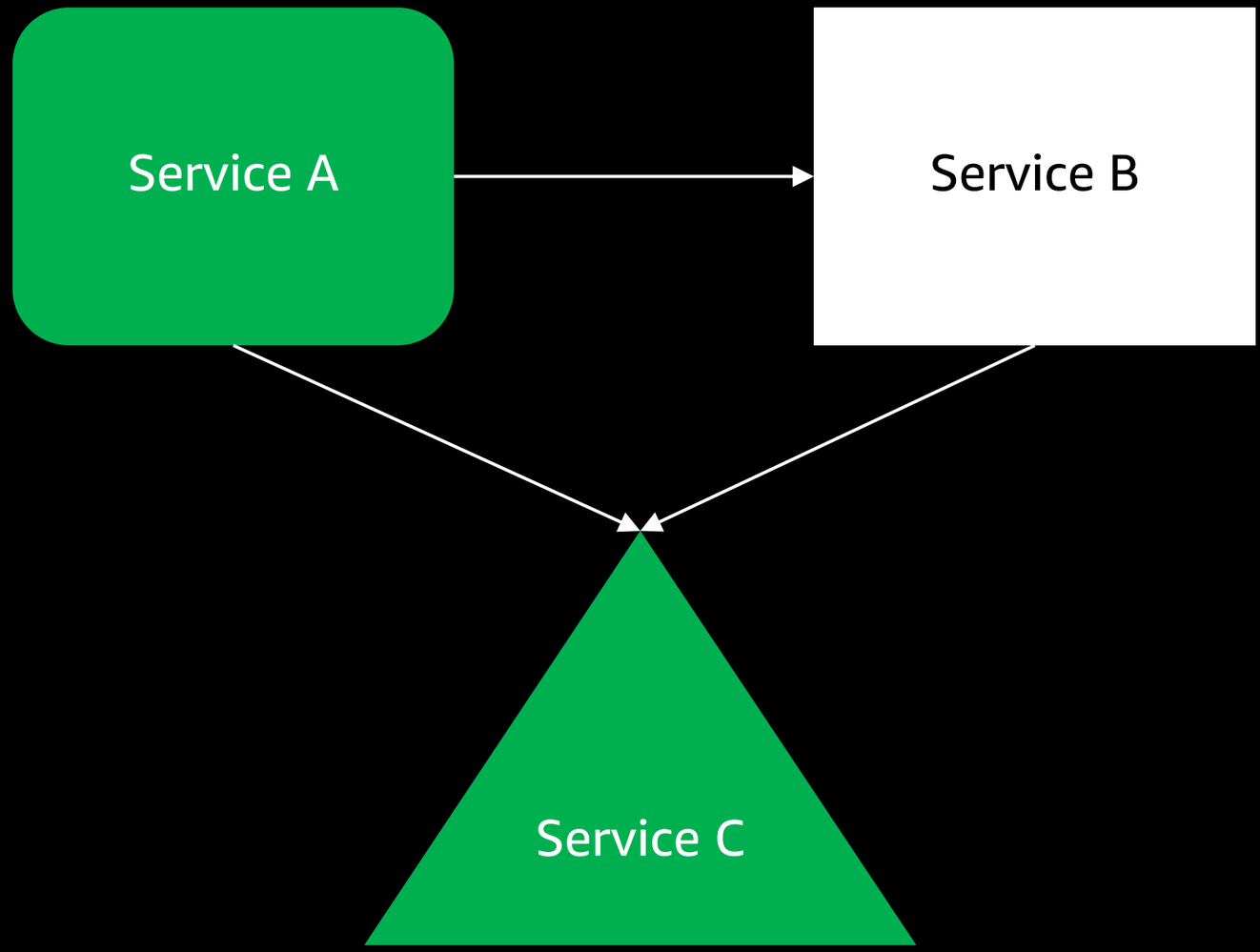
Why do we build service-oriented architectures?



AWS re:Invent

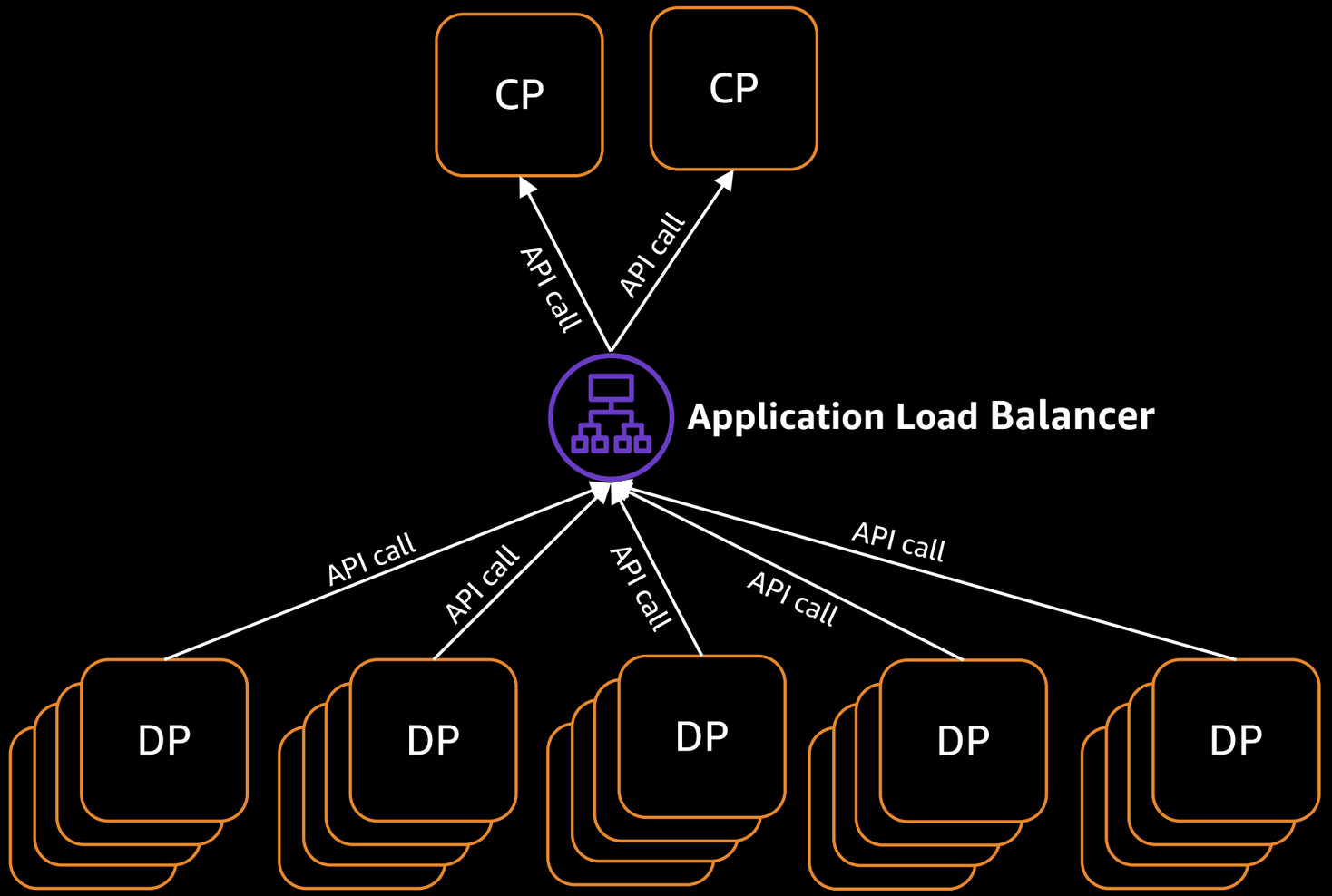
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Blast radius


Service A
Service B
Service C

At AWS

- Data plane – what you pay for
- Control plane – configures and manages the data plane

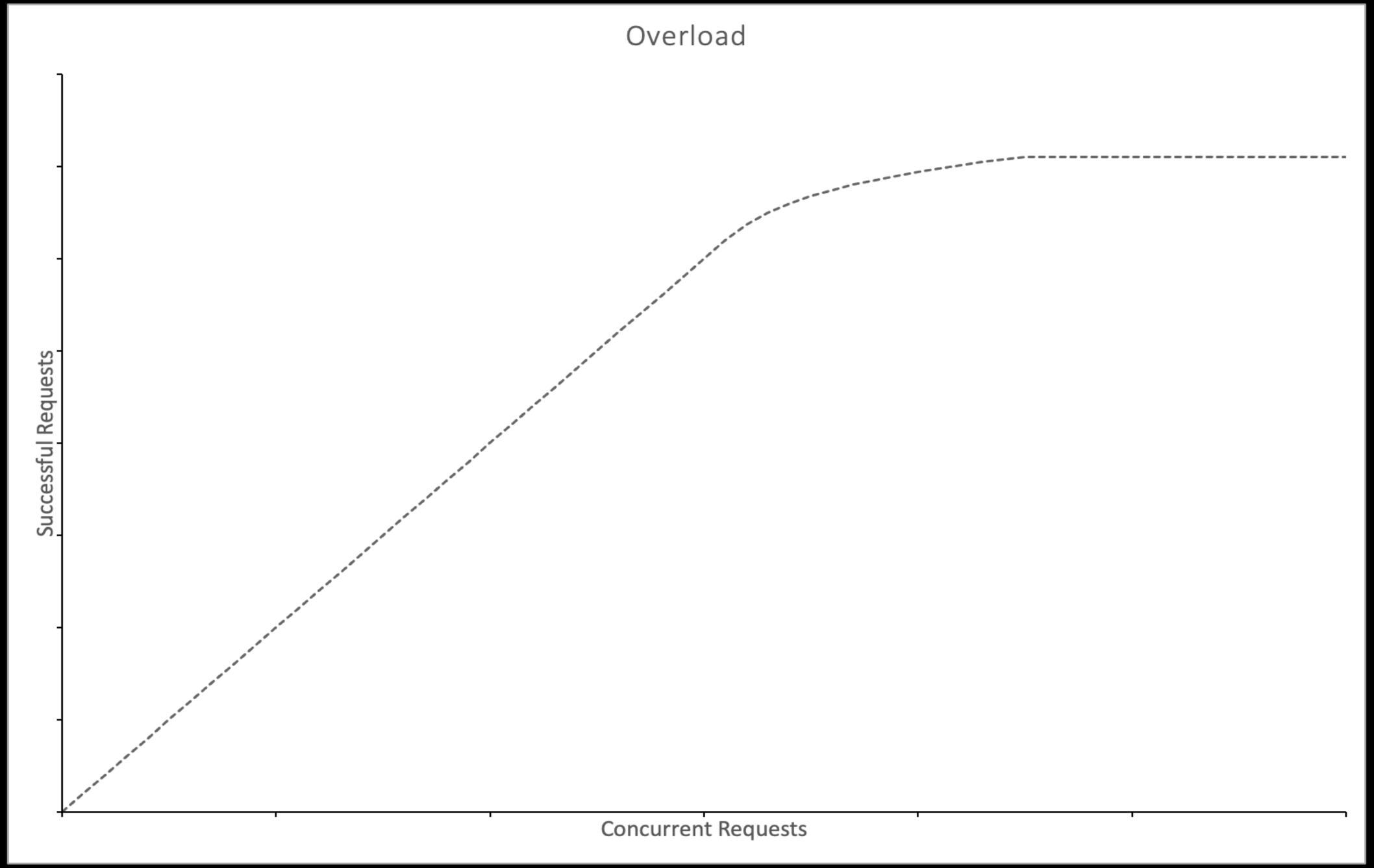

CP
CP
API call
API call
Application Load Balancer
API call
API call
API call
API call
API call
DP
DP
DP
DP
DP



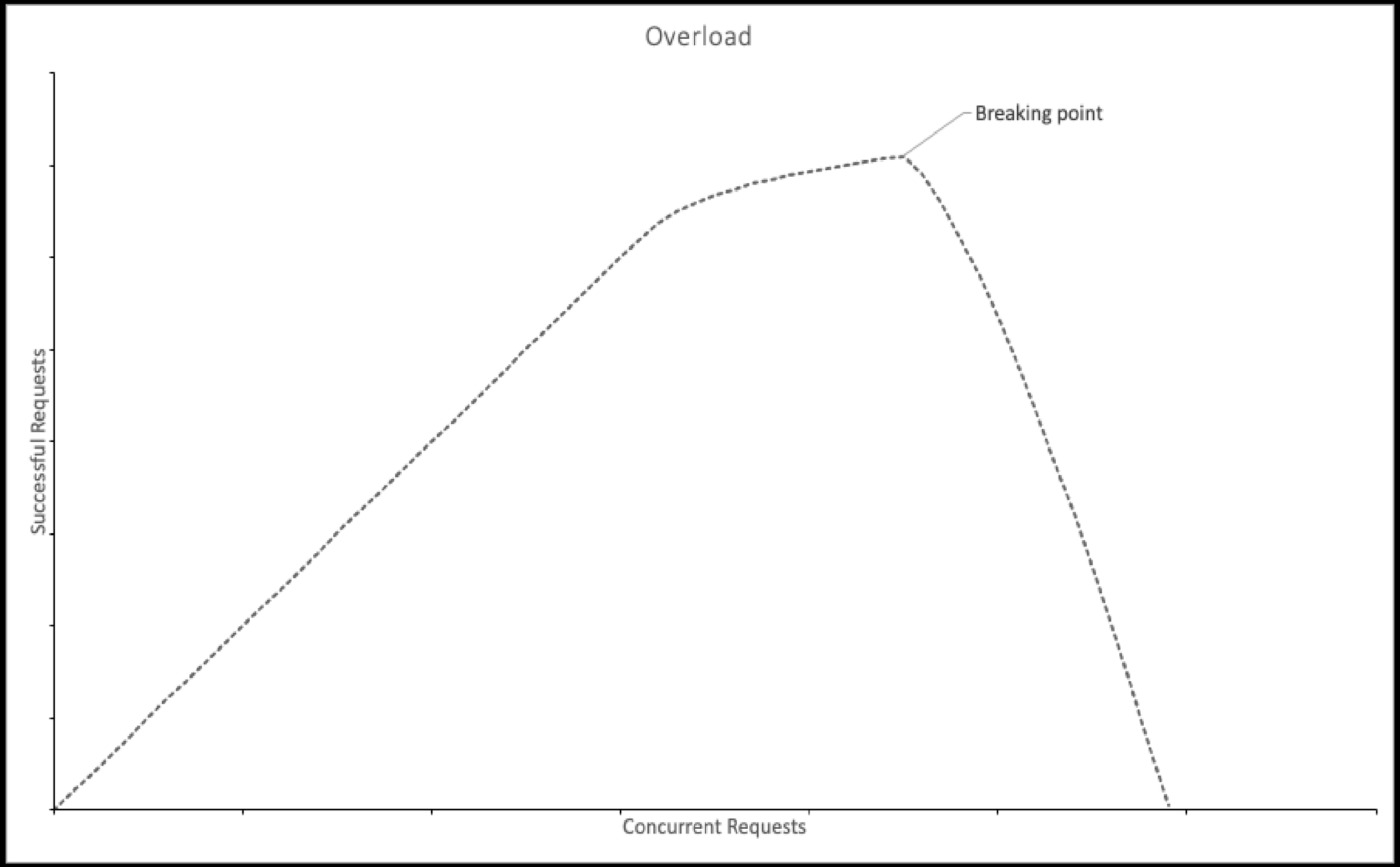





12
1
2
3
4
5
6
7
8
9
10
11


Overload
Successful Requests
Concurrent Requests

Not what really happens


Overload
Breaking point
Successful Requests
Concurrent Requests

What can we do?

- Load shedding
- Cooperating clients

Bellevue School District
@TheBSD405

UPDATE - all Bellevue schools are closed today. There is no transportation and all after school activities are cancelled. For more information visit: bsd405.org/closure

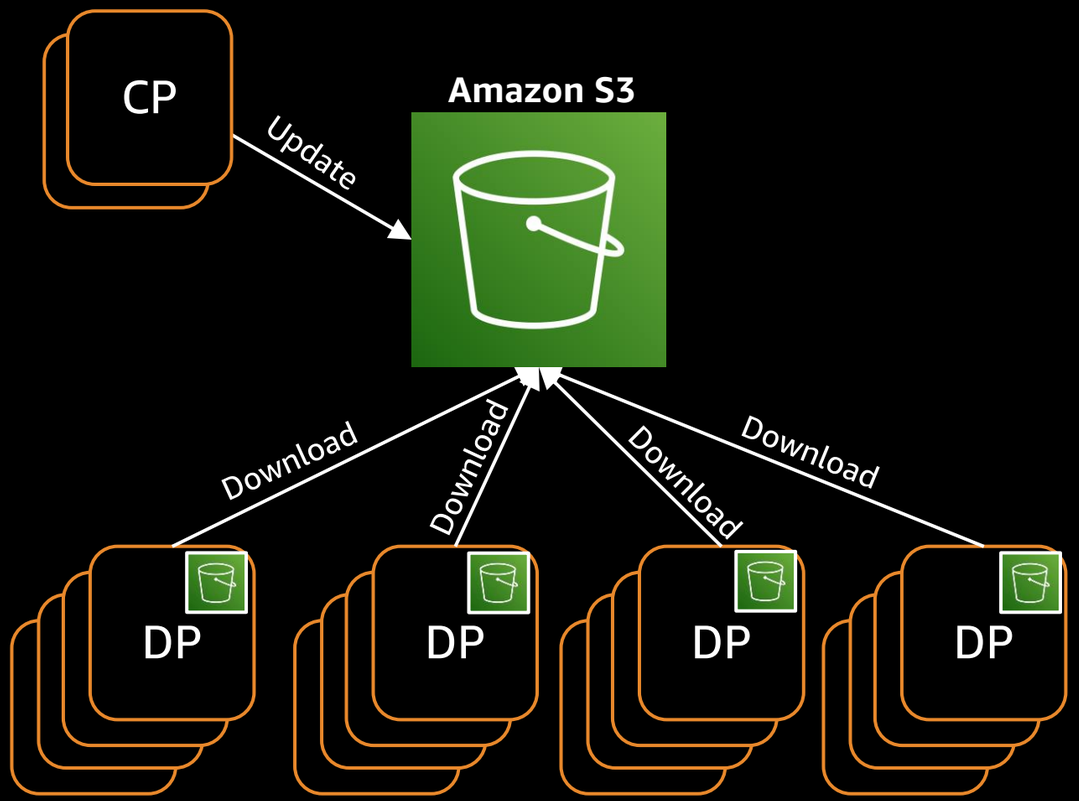
Jan. 13, 2020 at 6:55 a.m.



6:55 AM · Jan 13, 2020

23 18 people are Tweeting about this

- Control plane periodically updates an Amazon S3 file with the latest configuration
- Data plane servers download the configuration file and apply it locally
- Works in the opposite direction too

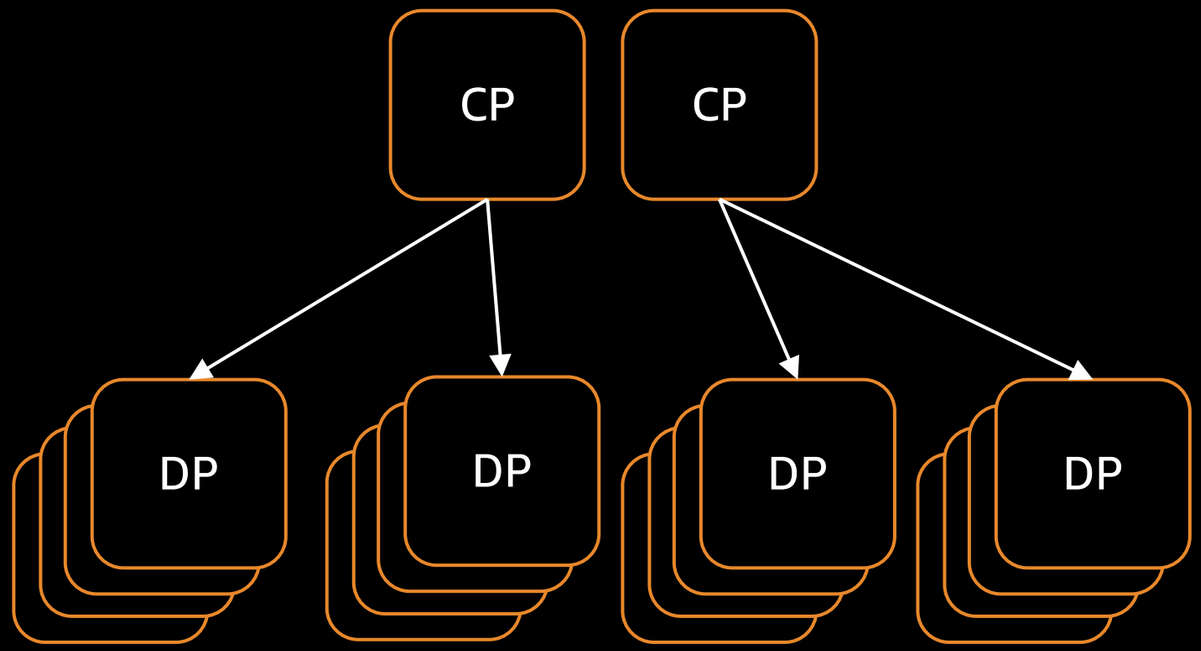


Not always practical

- Cannot precompute configuration ahead of time
- Changes need to propagate in seconds or faster


4:50
Bellevue School Distr
Bellevue School District
Remind Me
Message
Decline
Accept

- Put the smaller control plane in charge of the larger data plane
- The control plane dictates the pace of all interactions
- Requires coordination between control plane servers



Watch out for zombies and ghosts

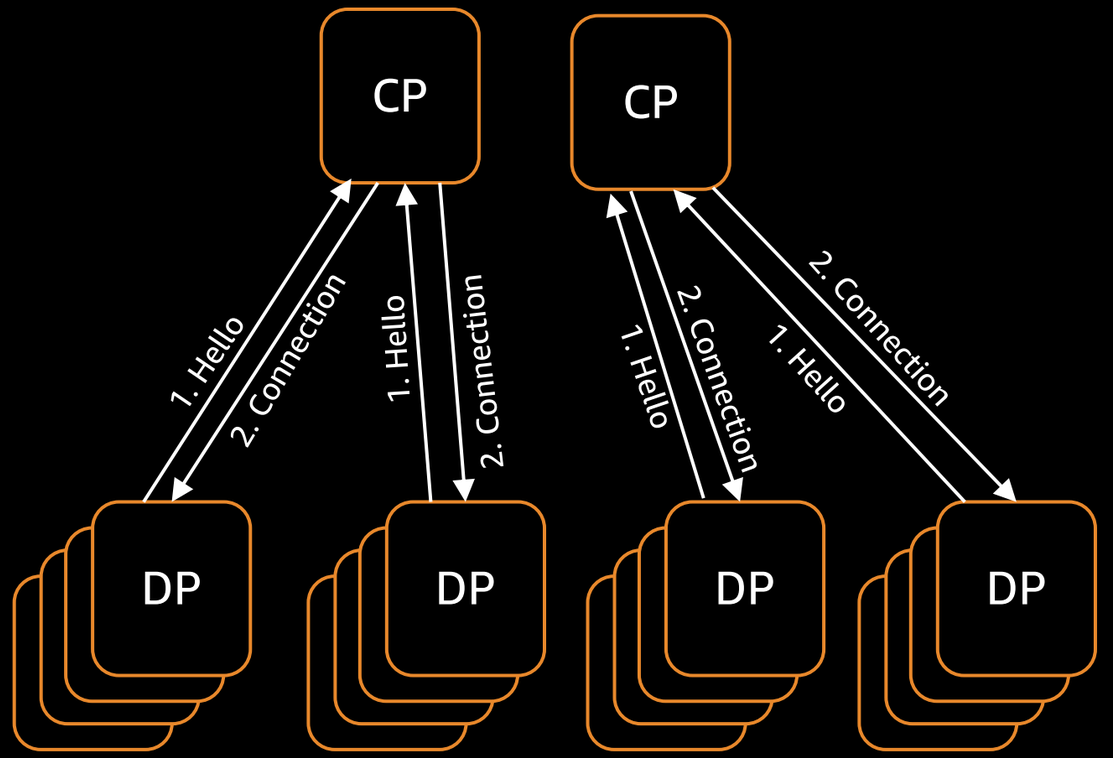
Enter your number and click Call Me Now. (You'll need an open phone line.)
We'll call you and connect you to a customer service associate.

Country/Region United States

Your number (206) 555 - 0148

Call me now

- Servers in the data plane fleet send out a small UDP “hello” message
- Control plane fleet responds by opening a long-lived connection
- Control plane still dictates the pace of interactions



Takeaways

- Pay attention to scale mismatch
- Simple architectures have inherent availability benefits
- When vastly outnumbered, put the smaller fleet in control

For more information, visit
https://aws.amazon.com/builders-library/

Thank you!

Joe Magerramov

Senior Principal Engineer
Twitter: @_joemag_

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.