KubeCon | CloudNativeCon

Europe 2023

Archetypes for Reliable Systems

Steve McGhee	*Reliability Advocate*	*Google*
Ameer Abbas	*Product Manager*	*Google*

Who are we?

Steve McGhee

Reliability Advocate

Google

Ameer Abbas

Product Manager

Google

Agenda

01 Reliability Terms

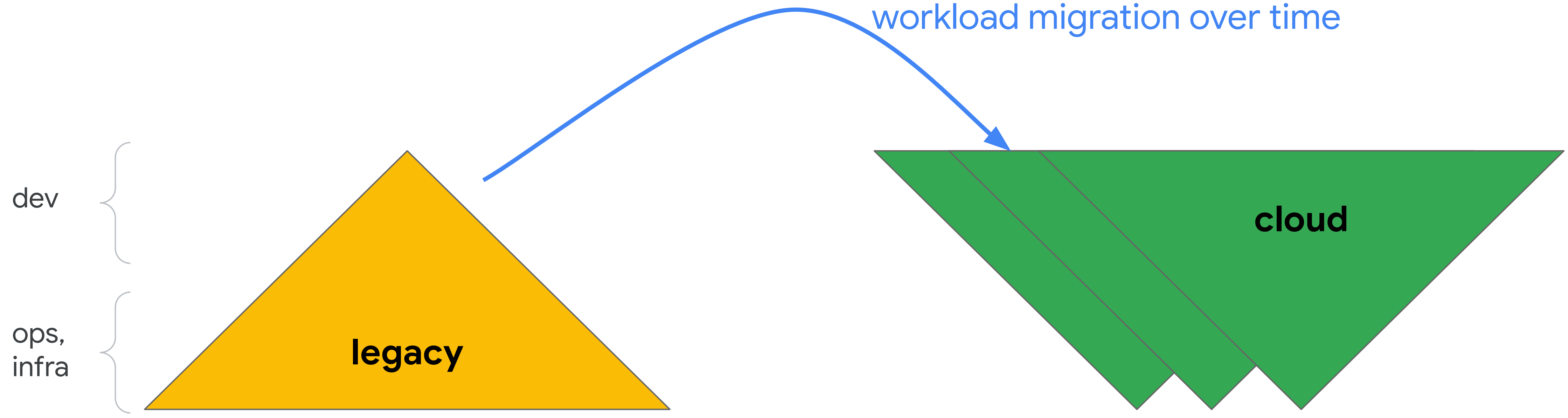
02 Application Archetypes

03 SLO Math

04 Reliable Architectures

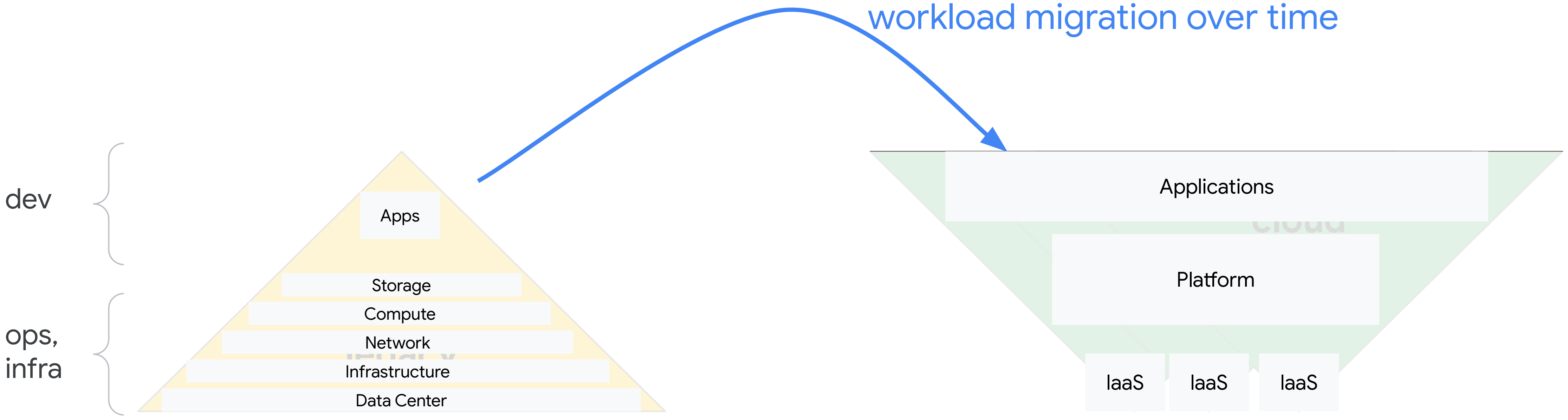
Can you build
99.99 services
on
99.9 infrastructure?

Pyramids of Reliability



software *inherits* reliability from base

software *improves* availability



workload migration over time

dev

ops, infra

Apps

Storage

Compute

Network

Infrastructure

Data Center

Applications

Platform

IaaS IaaS IaaS

software *inherits* reliability from base

software *improves* availability

Application Archetypes

Archetypes to Reliability

- **Archetype** - Abstract model

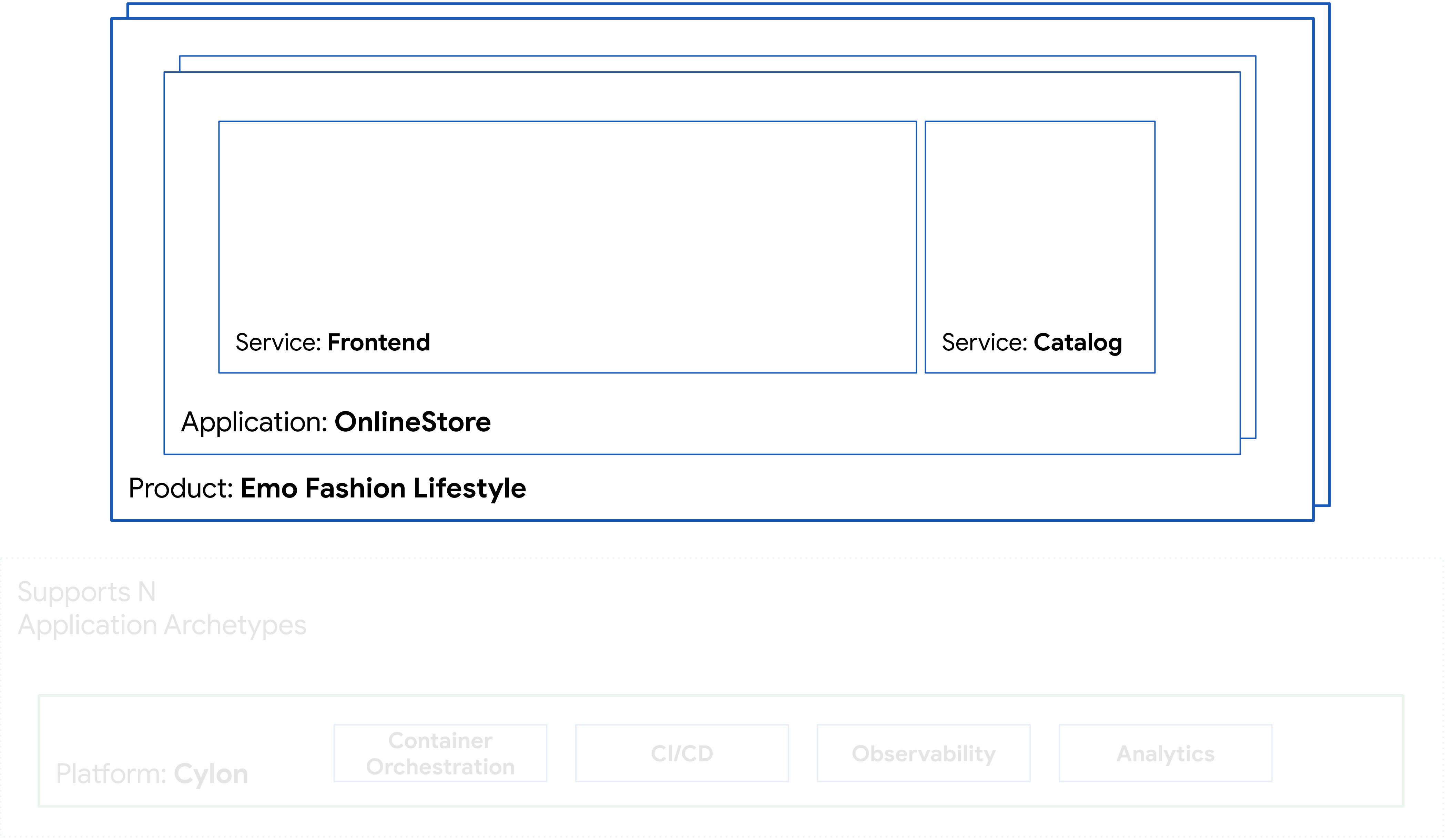
  Replication, redundancy, RTO/RPO, DR, cost

- **Architecture** - Products and Service design

  K8s, Mesh, CICD, DBs, Storage and backup

- **App/Service** & Footprint - *always changing*
- **SLOs** - expectations, guardrails

Platforms and Applications



Service: **Frontend**

Service: **Catalog**

Application: **OnlineStore**

Product: **Emo Fashion Lifestyle**

Supports N
Application Archetypes

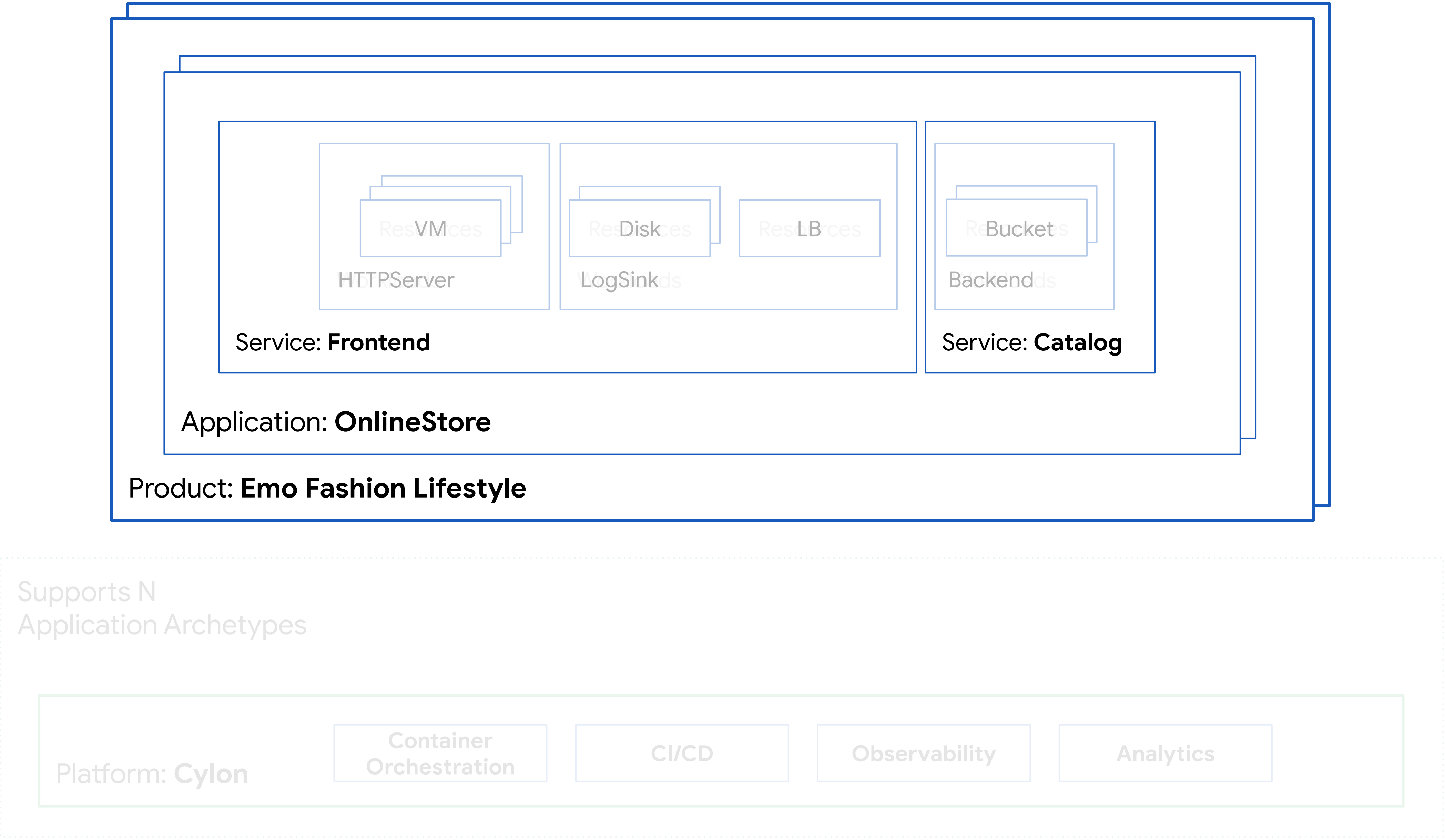
Platform: **Cylon**

Container Orchestration

CI/CD

Observability

Analytics



VM

HTTPServer

Disk

LB

LogSink

Service: **Frontend**

Bucket

Backend

Service: **Catalog**

Application: **OnlineStore**

Product: **Emo Fashion Lifestyle**

Supports N
Application Archetypes

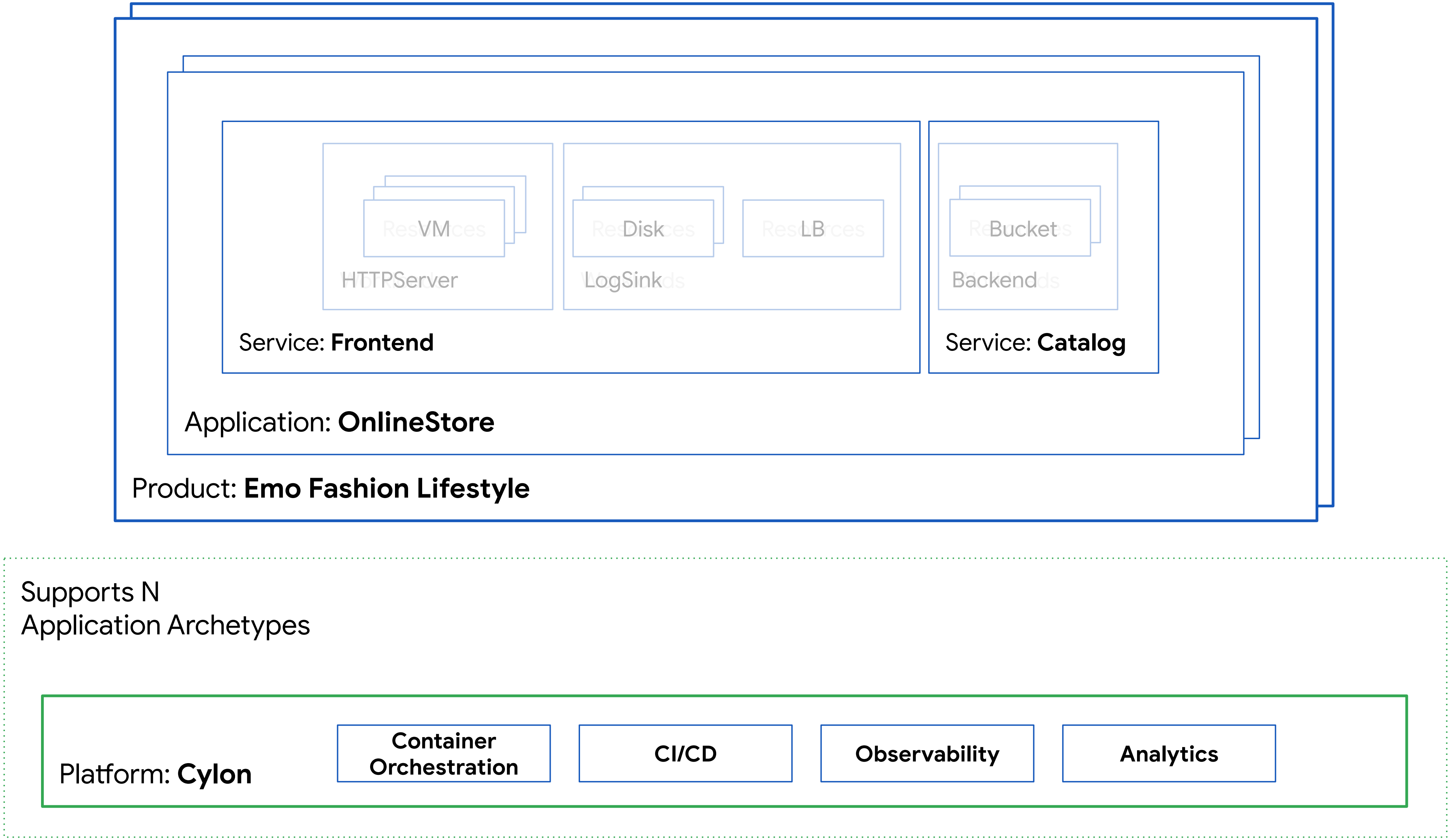
Platform: **Cylon**

Container Orchestration

CI/CD

Observability

Analytics



VM

HTTPServer

Disk

LB

LogSink

Service: **Frontend**

Bucket

Backend

Service: **Catalog**

Application: **OnlineStore**

Product: **Emo Fashion Lifestyle**

Supports N
Application Archetypes

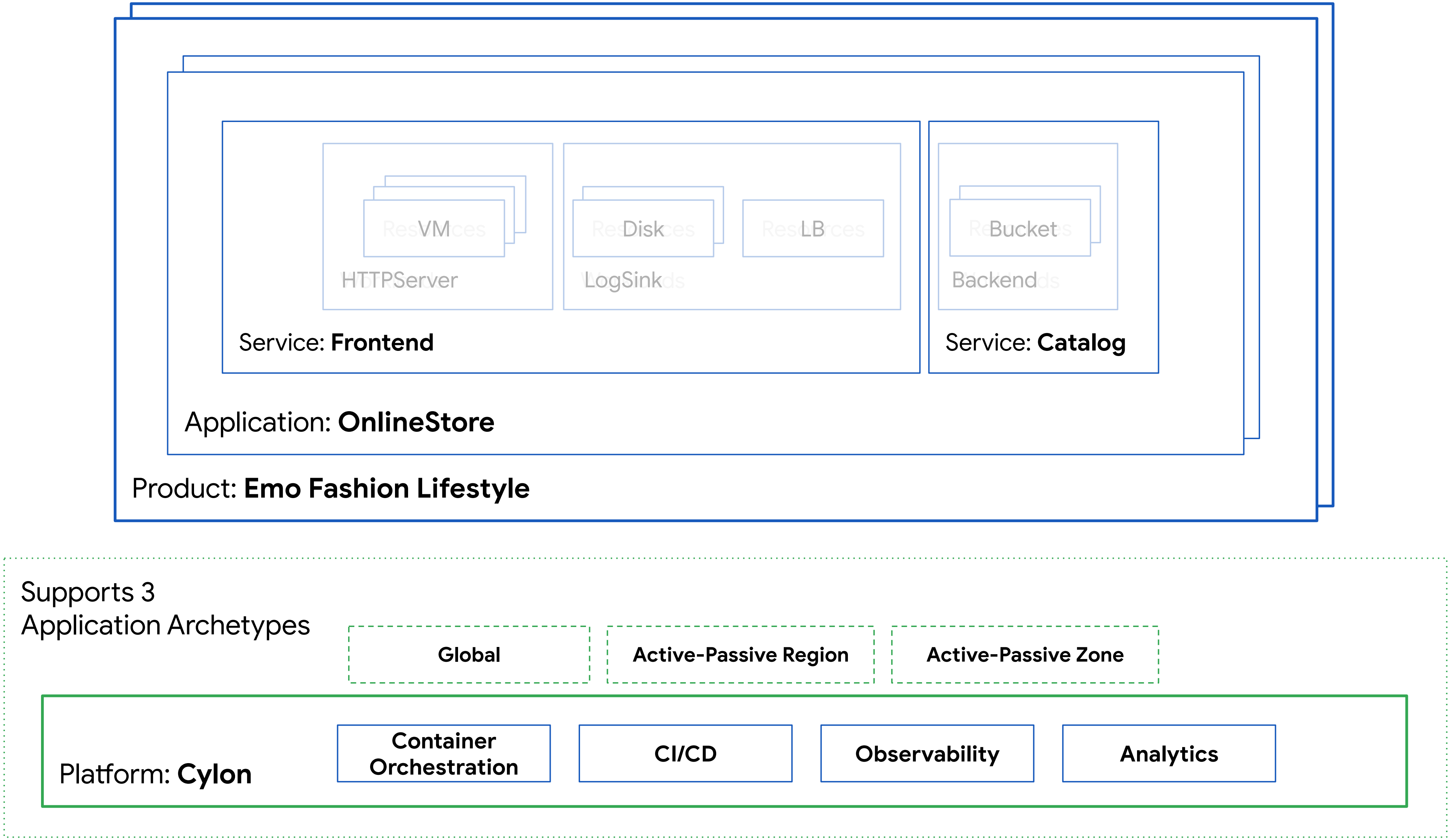
Platform: **Cylon**

Container Orchestration

CI/CD

Observability

Analytics



VM

HTTPServer

Disk

LB

LogSink

Service: **Frontend**

Bucket

Backend

Service: **Catalog**

Application: **OnlineStore**

Product: **Emo Fashion Lifestyle**

Supports 3
Application Archetypes

Global

Active-Passive Region

Active-Passive Zone

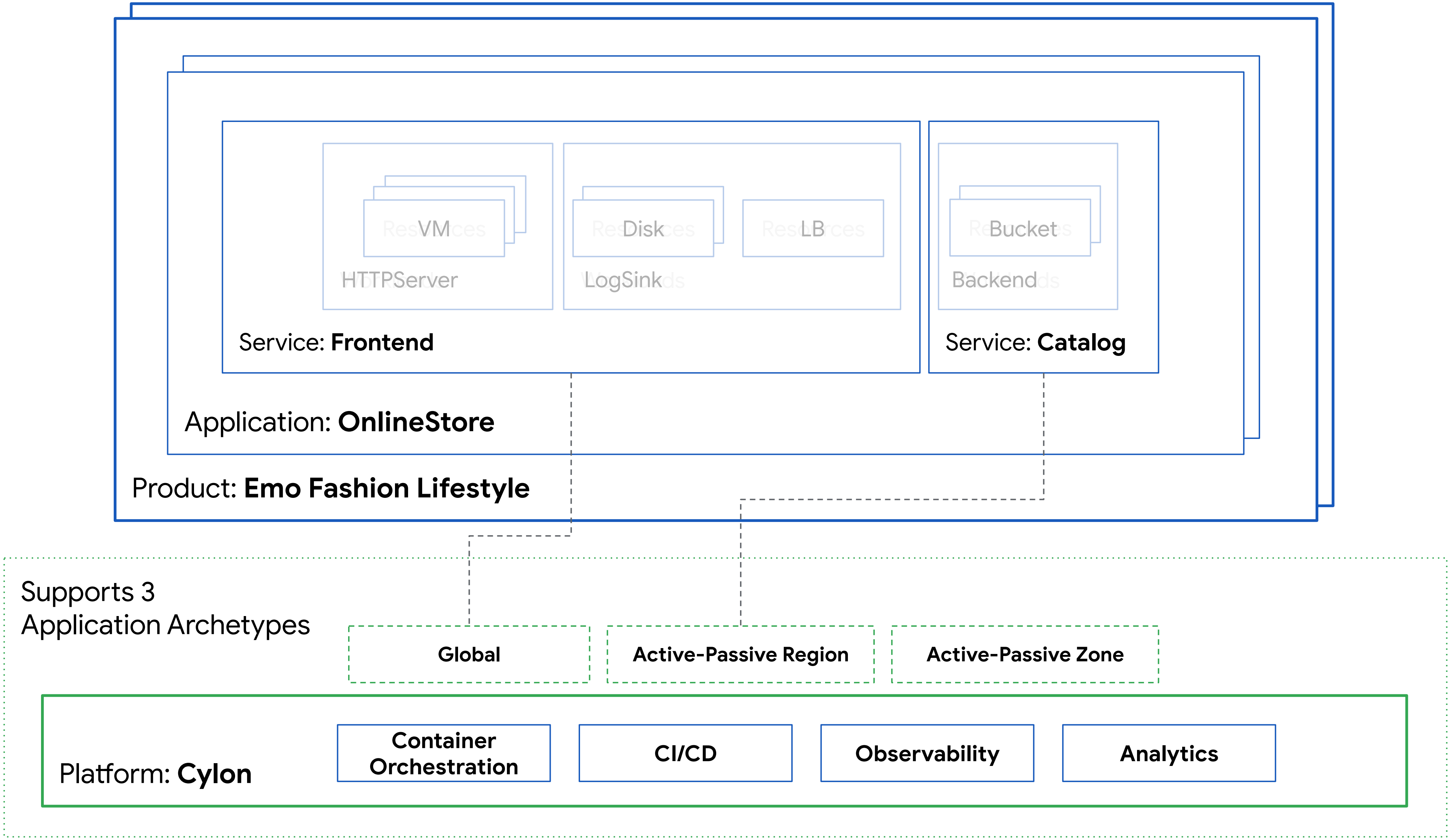
Platform: **Cylon**

Container Orchestration

CI/CD

Observability

Analytics



VM

HTTPServer

Disk

LB

LogSink

Bucket

Backend

Service: **Frontend**

Service: **Catalog**

Application: **OnlineStore**

Product: **Emo Fashion Lifestyle**

Supports 3
Application Archetypes

Global

Active-Passive Region

Active-Passive Zone

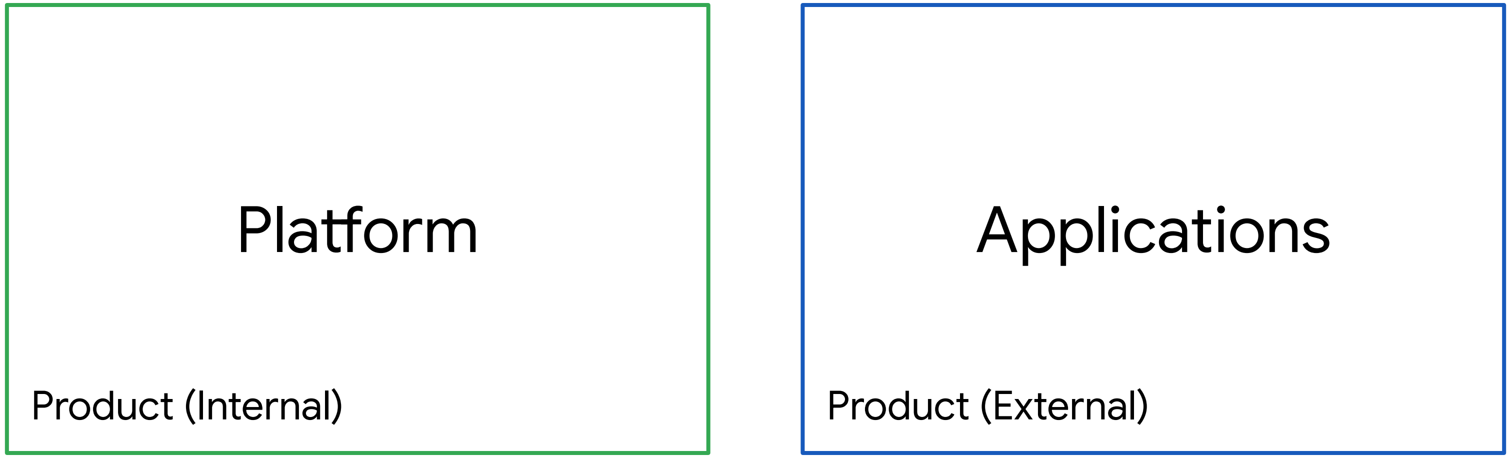
Platform: **Cylon**

Container Orchestration

CI/CD

Observability

Analytics

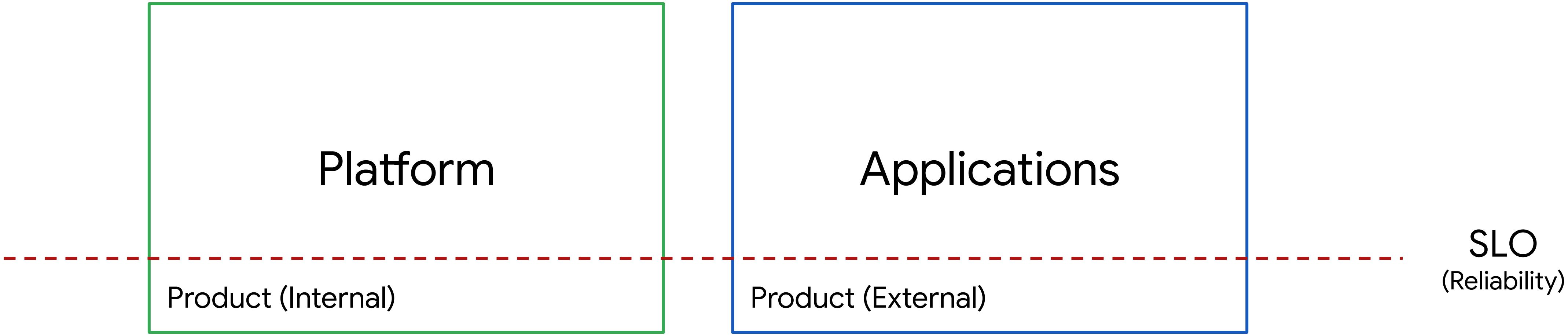


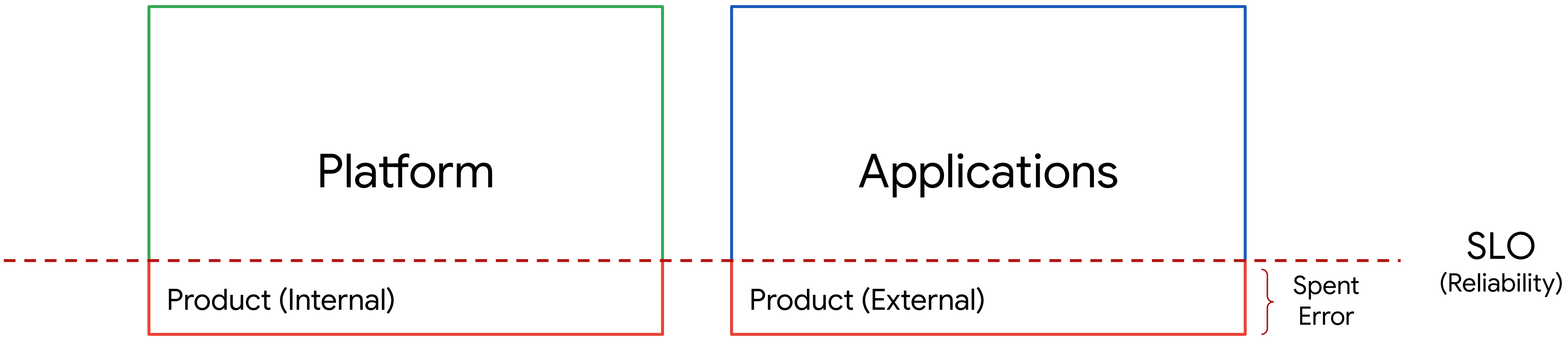
Platform

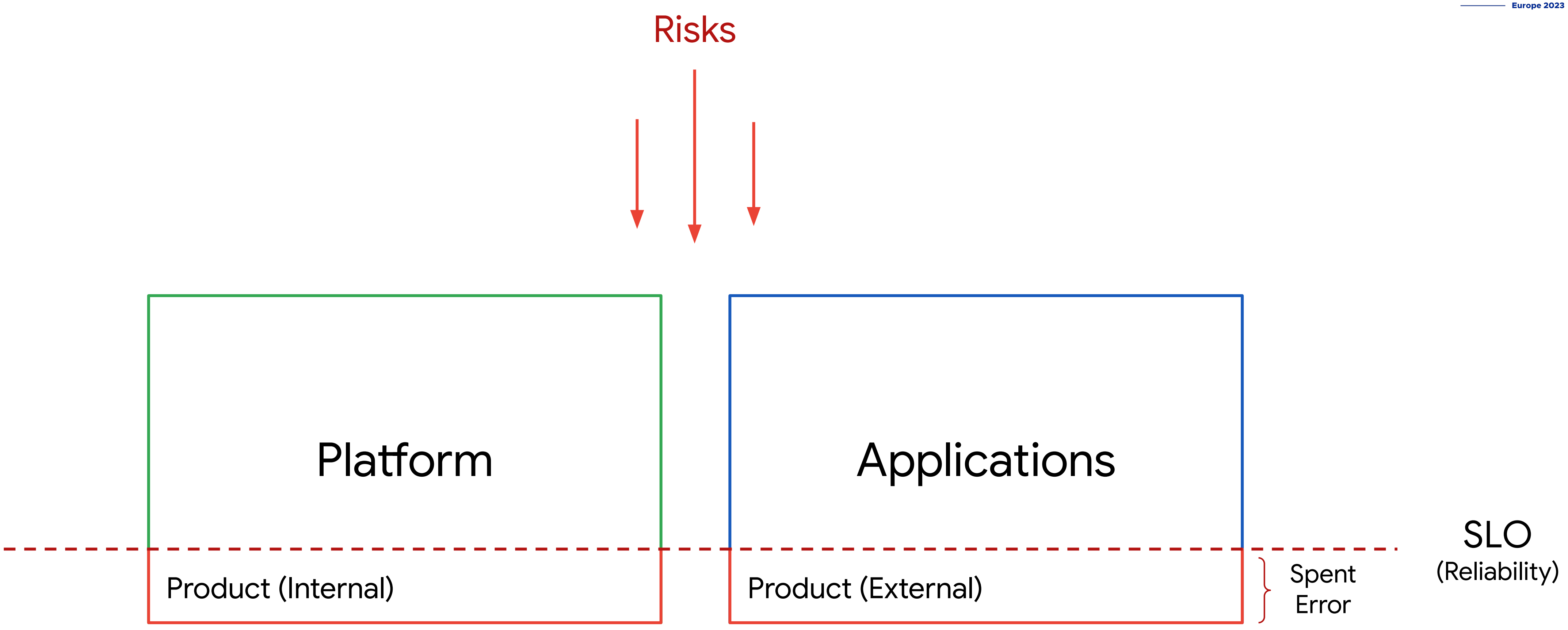
Product (Internal)

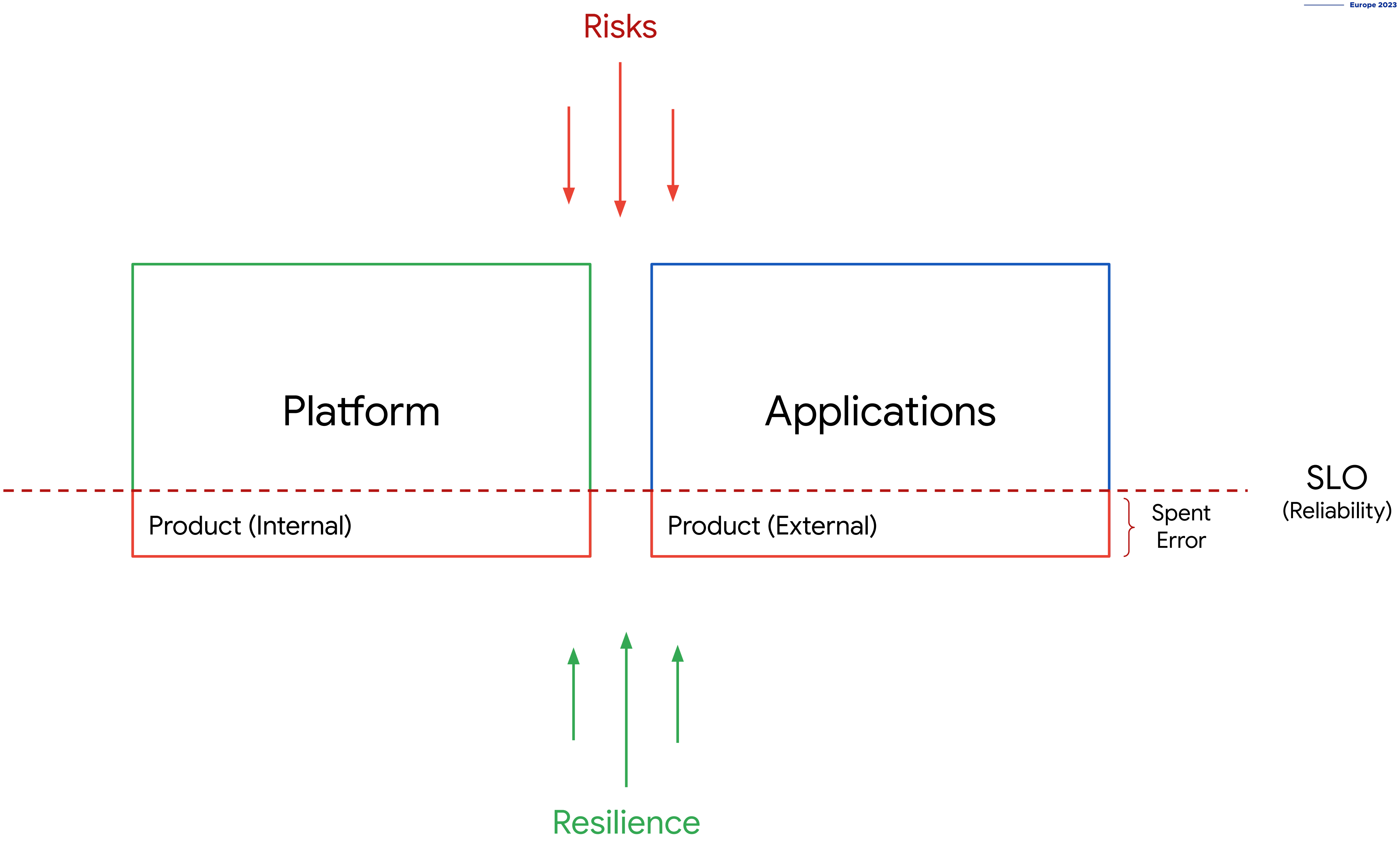
Applications

Product (External)

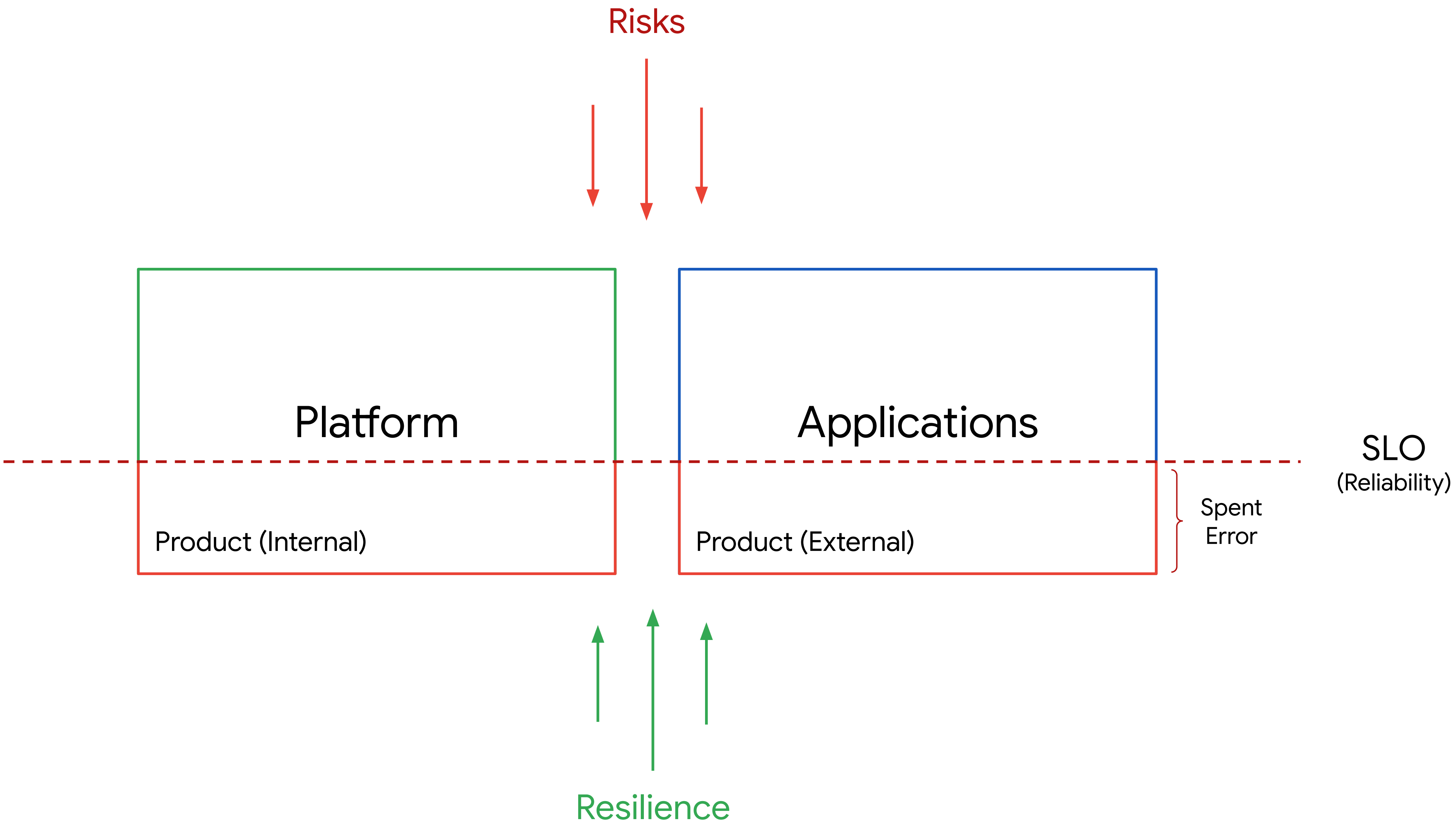

Platform
Product (Internal)
Applications
Product (External)
SLO
(Reliability)


Platform
Applications
SLO
(Reliability)
Product (Internal)
Product (External)
Spent
Error


Risks
Platform
Applications
Product (Internal)
Product (External)
Spent
Error
SLO
(Reliability)



Risks
Platform
Applications
SLO
(Reliability)
Product (Internal)
Product (External)
Spent
Error
Resilience



Risks

Platform

Applications

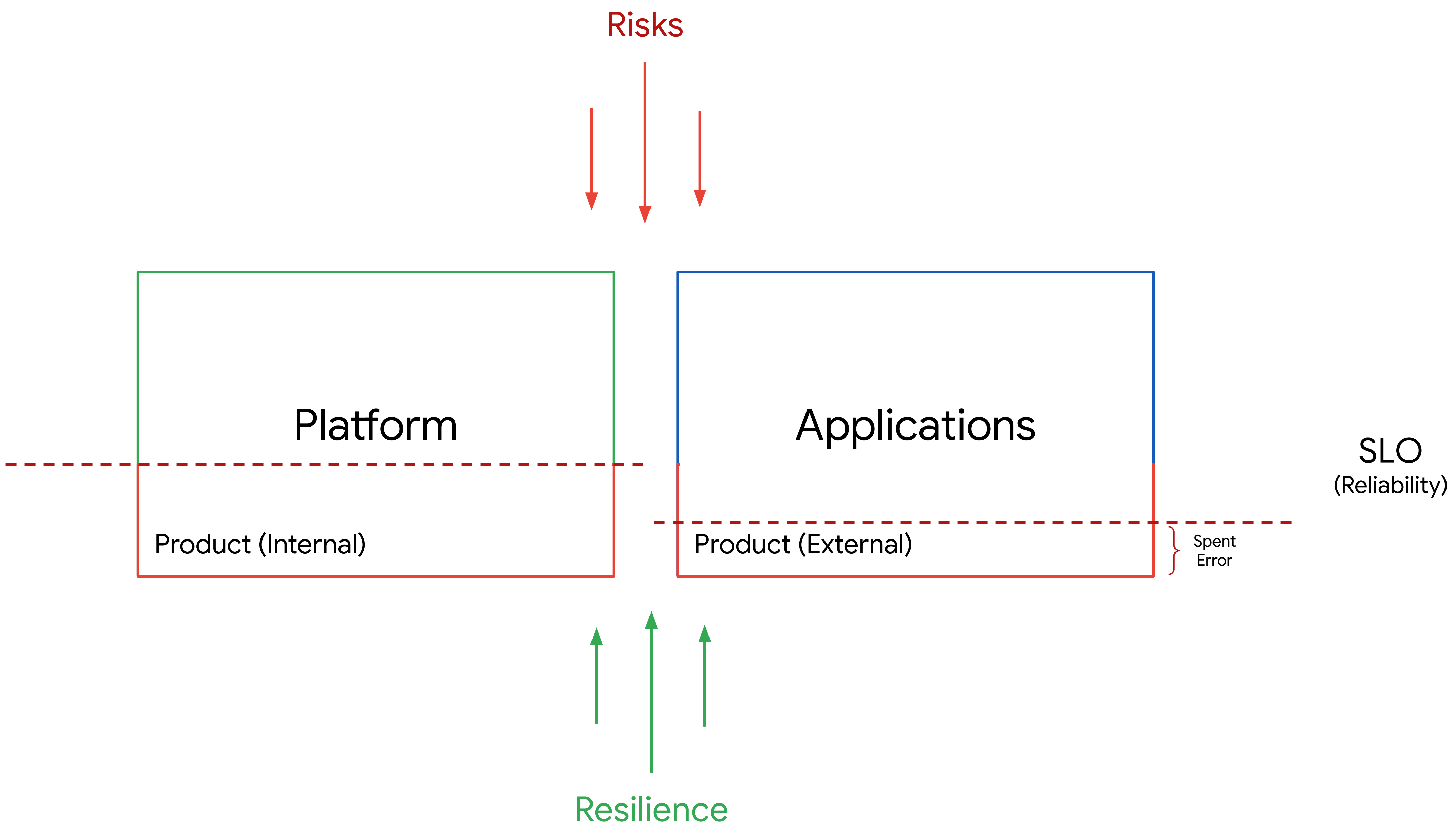
SLO
(Reliability)

Spent
Error

Product (Internal)

Product (External)

Resilience


Risks
Platform
Product (Internal)
Applications
Product (External)
SLO
(Reliability)
Spent
Error
Resilience

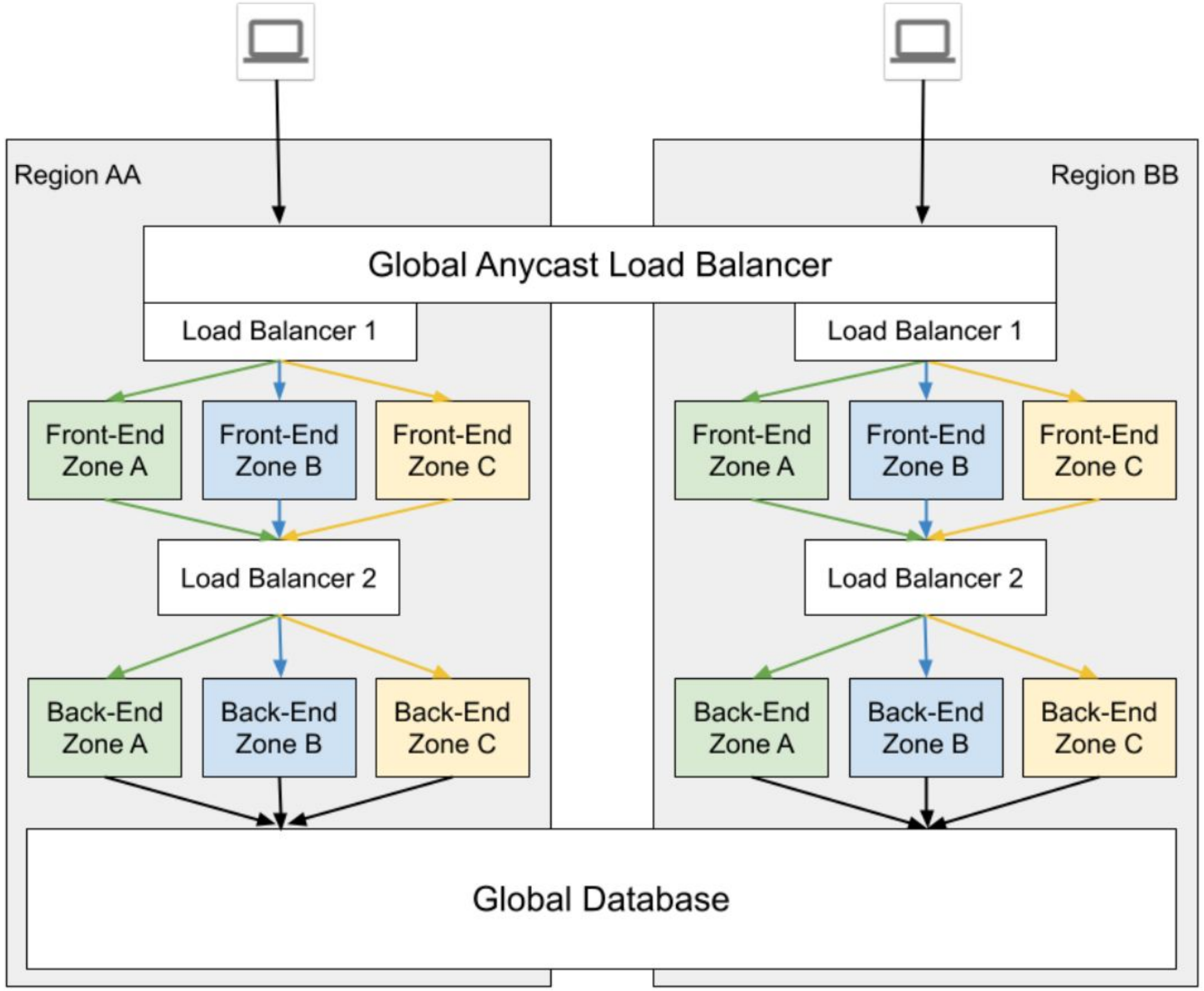
5 Application Archetypes

5 Application Archetypes

bit.ly/cloudarchetypes

Anna Berenberg, Brad Calder

ACM Computing Surveys, vol. 55 (2022), pp. 1-48

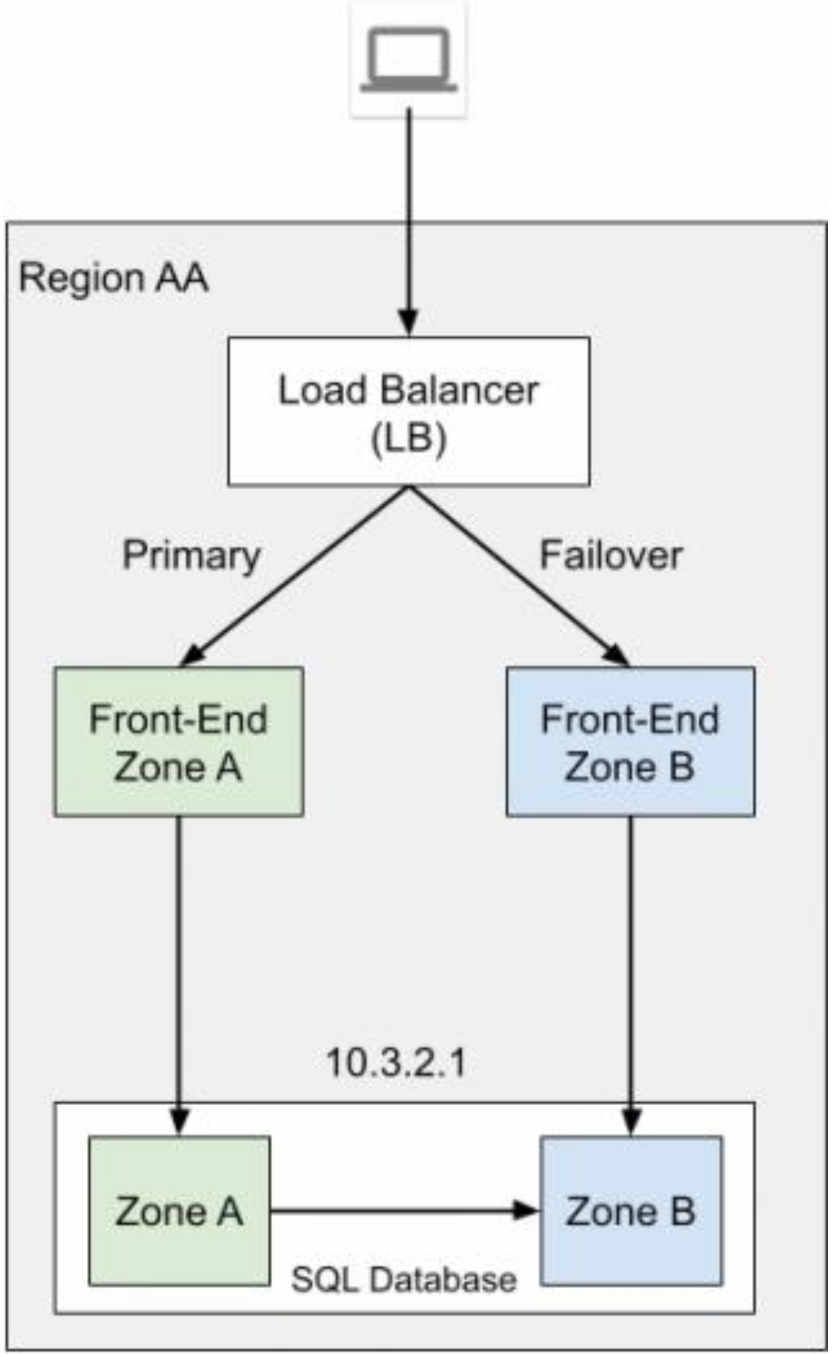


Archetype 5.2: Global Anycast with regional isolated stacks and global database deployment model

Archetype 2.2

Active Passive Zones

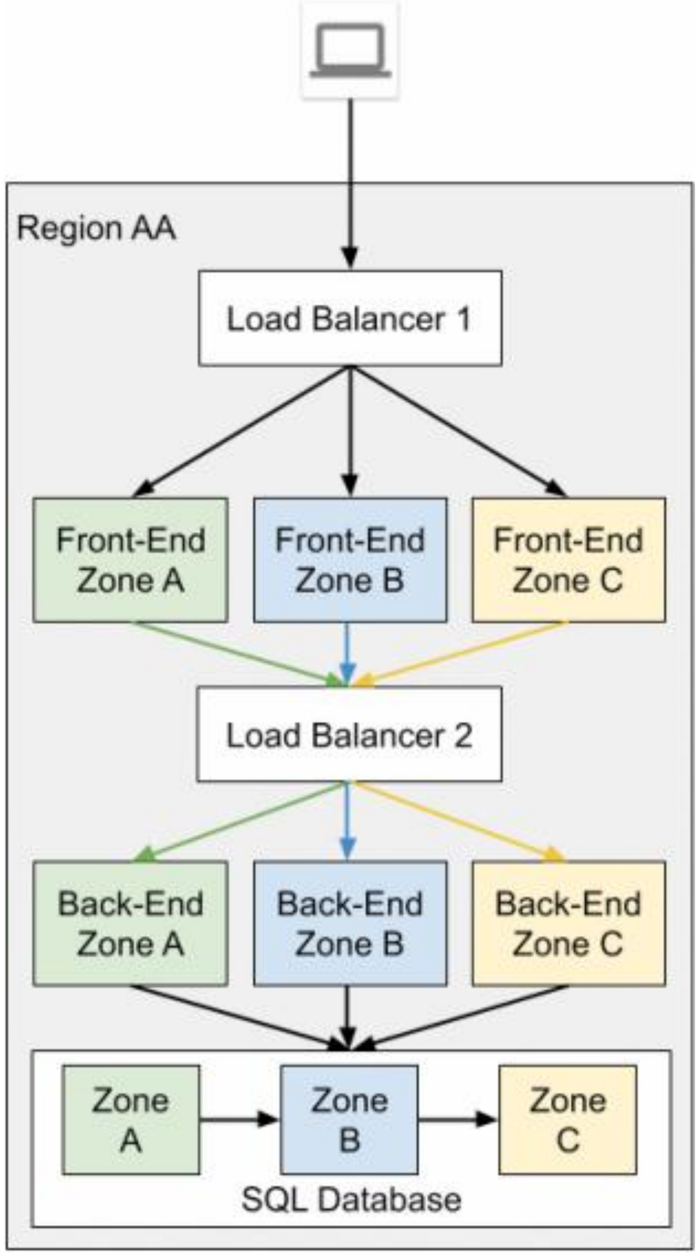
- **Deploy** all services of app to two zones in one region
- **Data** in SQL with a read replica
- **L4 LB** with one backend group
- **Survives zone** failure. Does not survive region failure.
- **Fail-Ops**: Change LB backend, promote read replica
- **Cost**: 2x serving + 2x data (1 replica)
- **Complexity**: Low
- **App Refactoring**: None (lift and shift)
- **Type**: COTS, licensing



Archetype 3.1

Multi Zonal

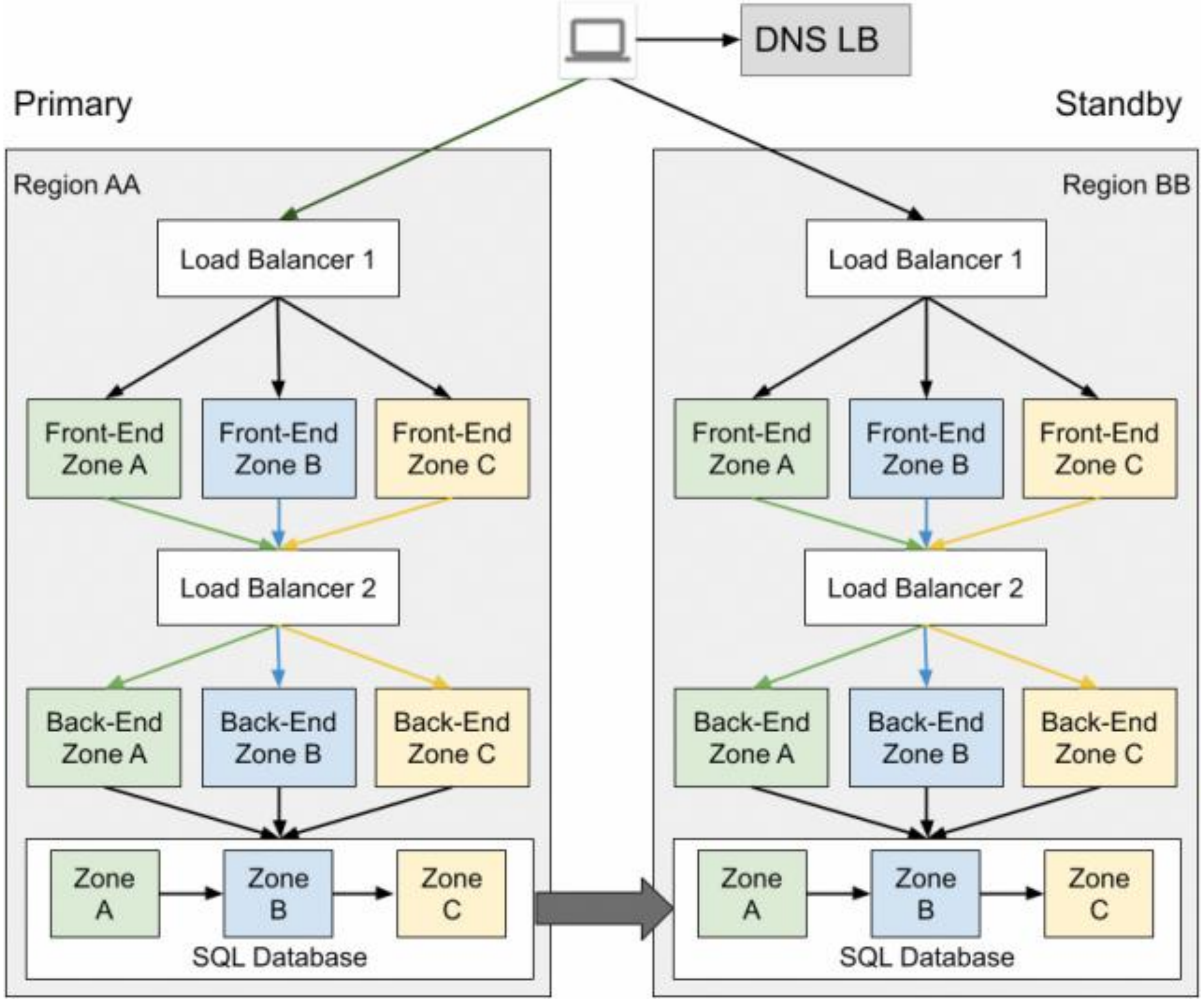
- **Deploy** all services of app to all three zones in one region
- **Data:** Use HA SQL
- Use **Global LB or Reg LB** with 3 backend groups
- **Survives zone** failure. Does not survive region failure.
- **Fail-Ops**: Initiate DB failover
- **Cost**: 1.5x serving + 2x data (HA SQL)
- **Complexity**: Medium
- **App Refactoring**: Low (multi instance)
- **Type**: Web services



Archetype 3.2

Active Passive Region

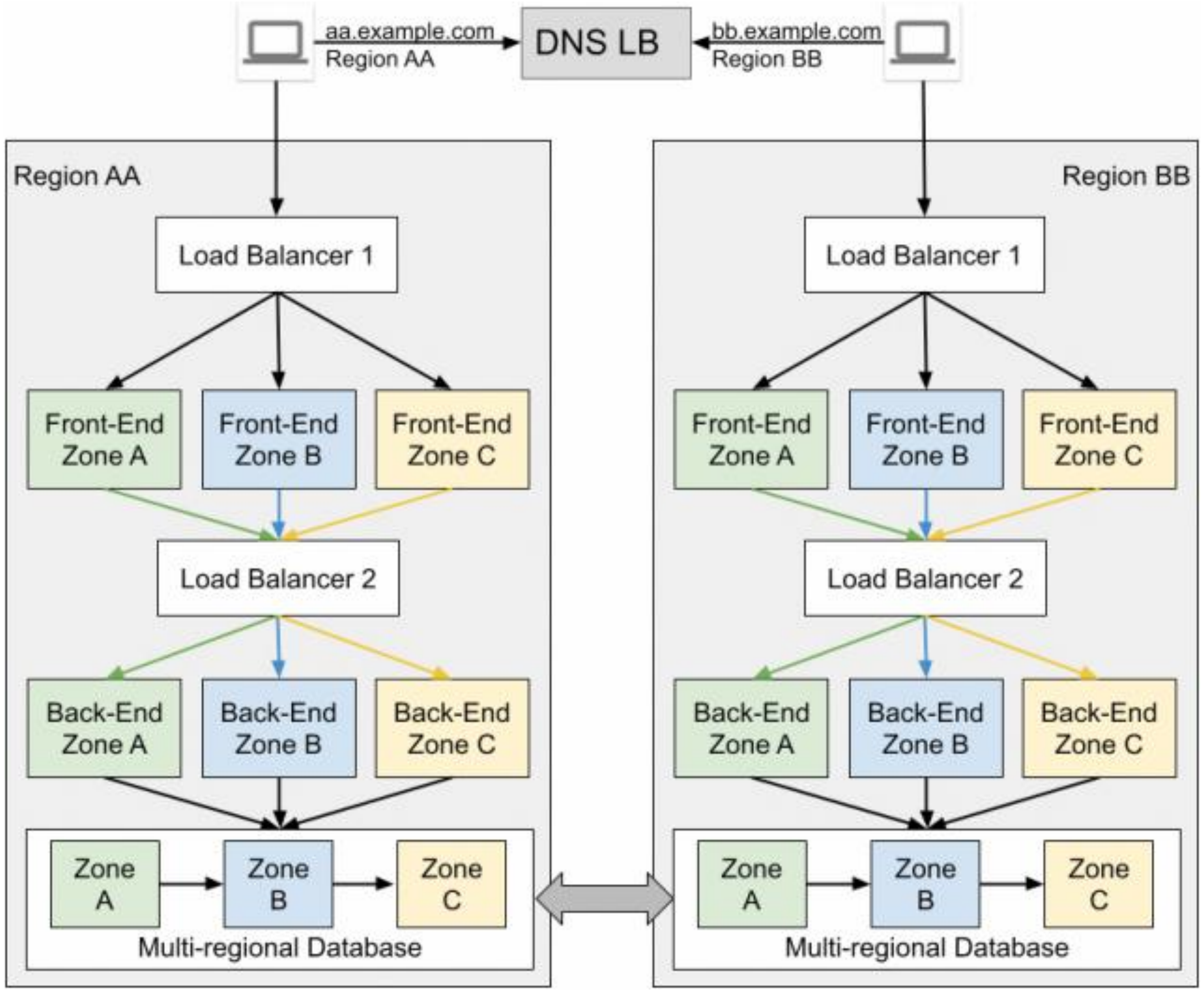
- **Deploy** all services of app to all three zones in each of two regions
- **Data:** SQL with cross-region replication
- **DNS** points at one LB (until disaster)
- **Survives zone and region** failures
- **Fail-Ops:** No action for zone failure.
  - Update DNS to point at standby LB
  - Cross region DR failover process for DB
- **Cost**: 3x serving + 2x data (HA SQL)
- **Complexity**: Medium
- **App Refactoring**: Medium (multi instance, multi regional data)
- **Type**: HA web services



Archetype 4.3

Isolated Regions

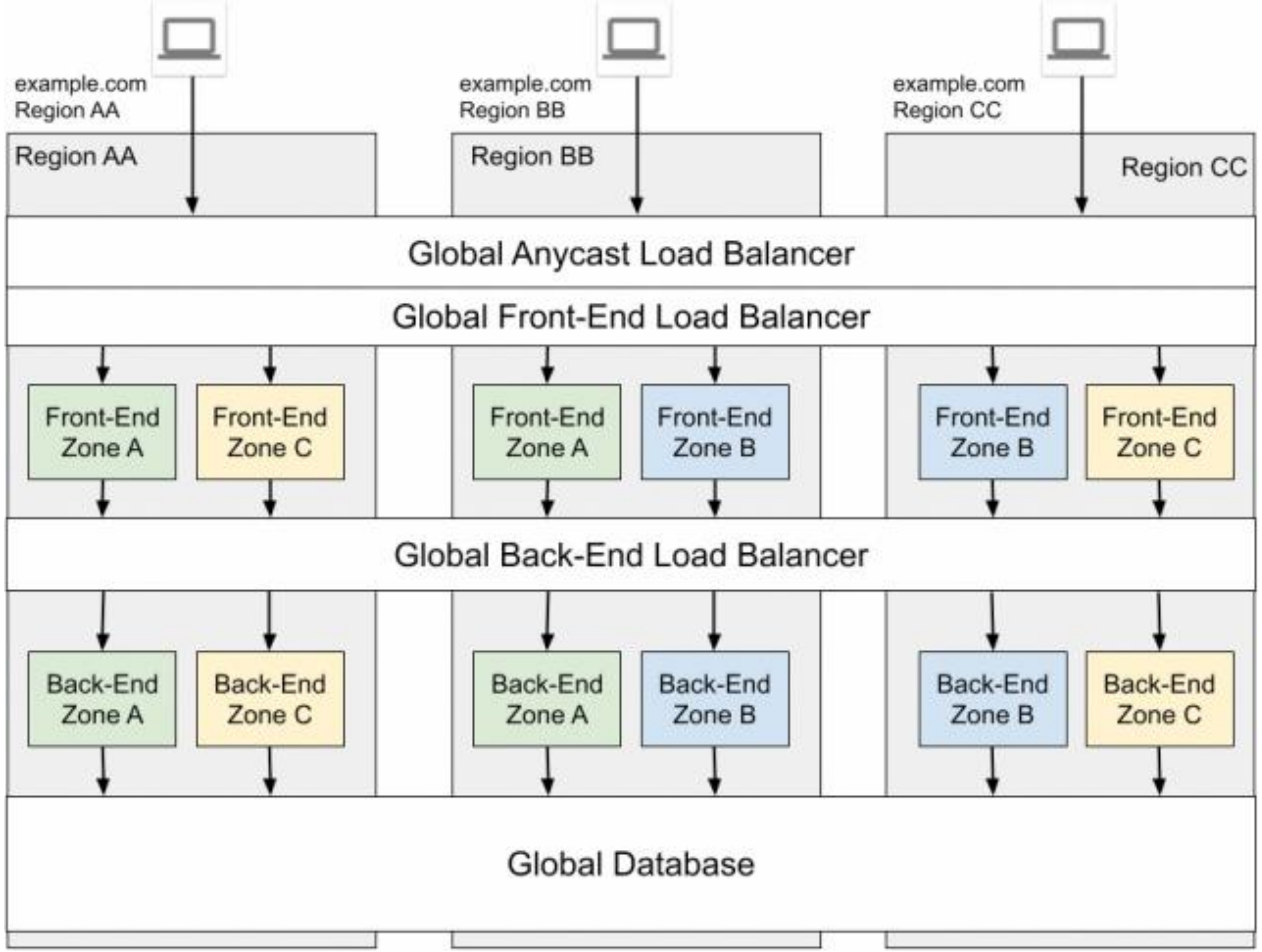
- **Deploy** all services of app to all three zones in each of two regions
- **Data:** Spanner or CockroachDB
- **DNS** points at two Regional LBs
- **Survives zone and region** failures. No impact for ½ consumers. Possible manual failover
- **Fail-Ops**: No action for zone failure. Optional regional failover like Arch 3.2
- **Cost**: 1.5 cost per region for zone failure
- **Complexity**: Medium/High
- **App Refactoring**: Medium (multi instance, multi regional data)
- **Type**: Regulated HA services



Archetype 5.2

Global

- **Deploy** all services of app to all three zones in each of two or more regions
- **Data:** Spanner or CockroachDB
- **Global LB** points at regional backend groups
- **Survives zone and region** failures
- **Fail-Ops**: None
- **Cost**: N+m cost modelling. Global DBs are more expensive
- **Complexity**: High
- **App Refactoring**: High (multi instance, global DBs)
- **Type**: Global consumer services



How to use Archetypes?

01

Services can be deployed to a **single archetype**

How to use Archetypes?

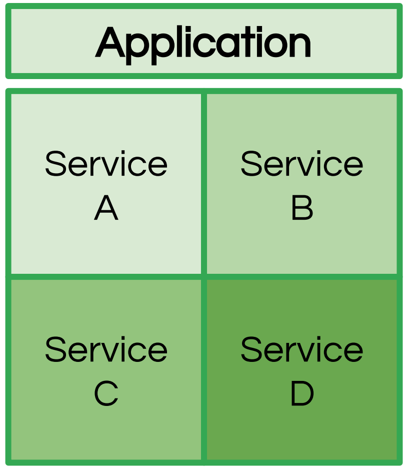
01

Services can be deployed to a **single archetype**

Service A

02

Application can use services across **multiple archetypes**



Application

Service A	Service B
Service C	Service D

How to use Archetypes?

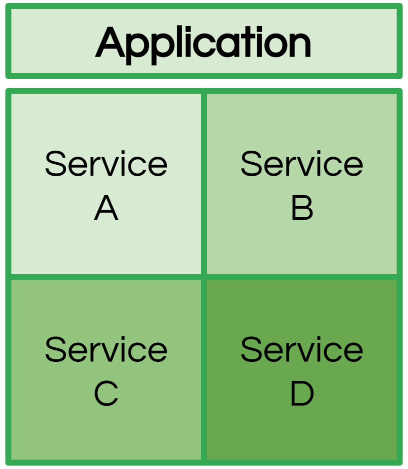
01

Services can be deployed to a **single archetype**

Service A

02

Application can use services across **multiple archetypes**



Application

Service A

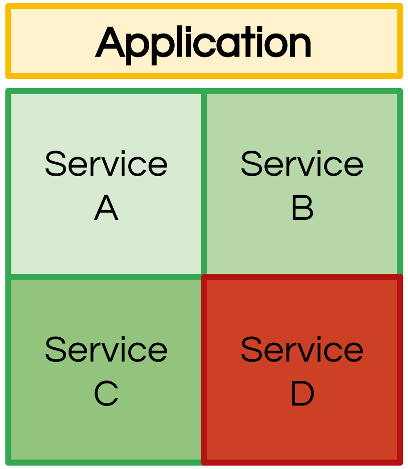
Service B

Service C

Service D

03

Applications should be designed for **graceful degradation**



Application

Service A

Service B

Service C

Service D

SLOs and SLIs

SLI

Quantitative measure of some aspect of the level of service

aka

latency, throughput, availability

from Google SRE Handbook

SLI

Quantitative measure of some aspect of the level of service

aka

latency, throughput, availability

SLO

a target value or range of values for a service level that is measured by an SLI

aka

99% of Get RPC calls will complete in less than 100 ms

from Google SRE Handbook

SLOs in one slide

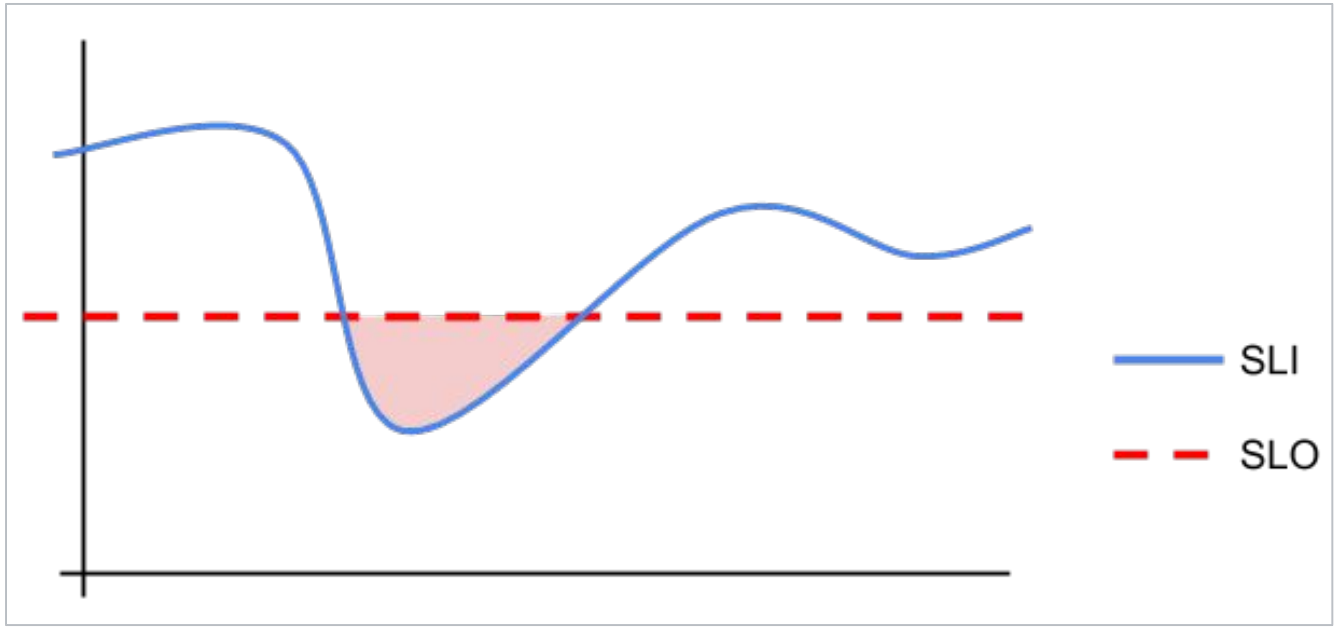
A **ratio-rate** of good/total, measured over a time duration.

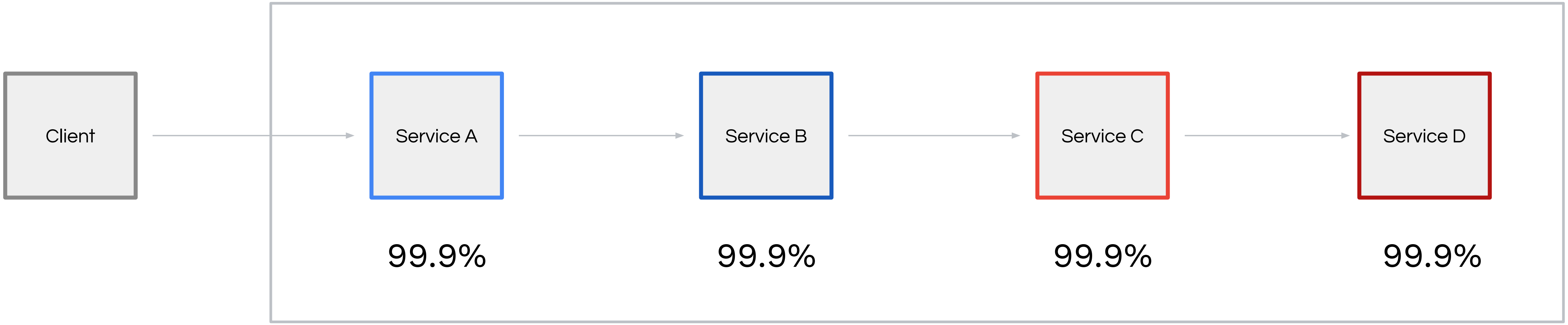
If too much non-good, for too long, tell a human.

SLI is the squiggly line

SLO is the straight one

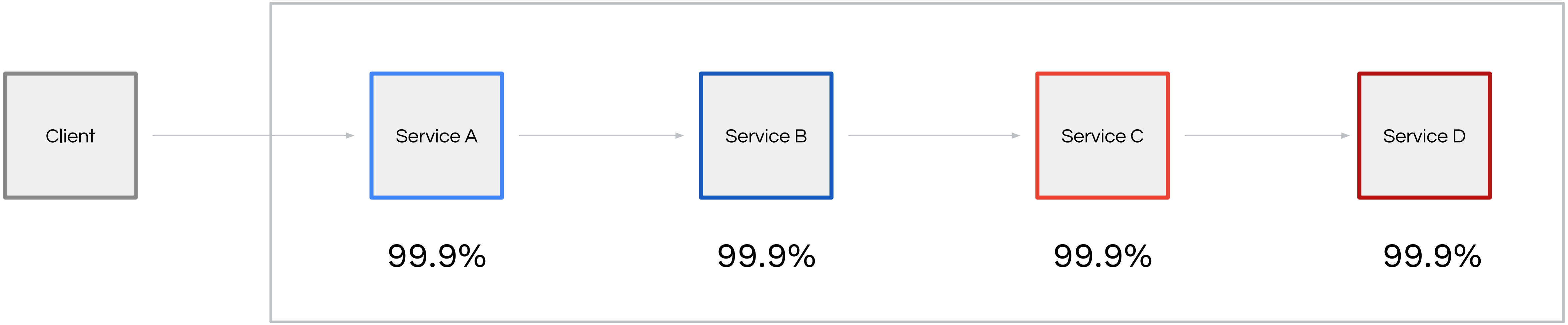
Area is time **exceeding SLO**





[from Steve McGhee's SLO Conf 2021 Talk]()

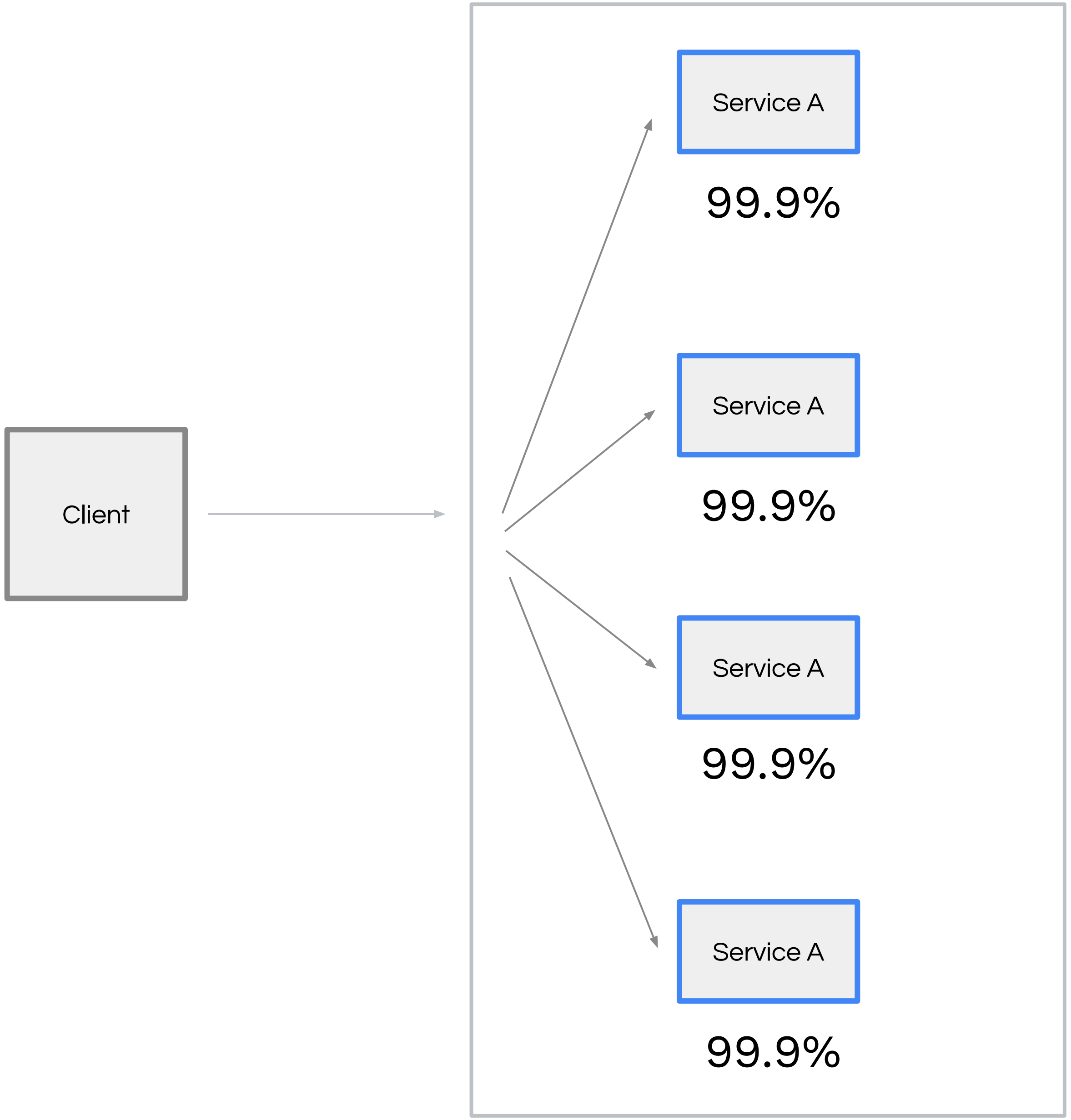
Intersection (or serial)



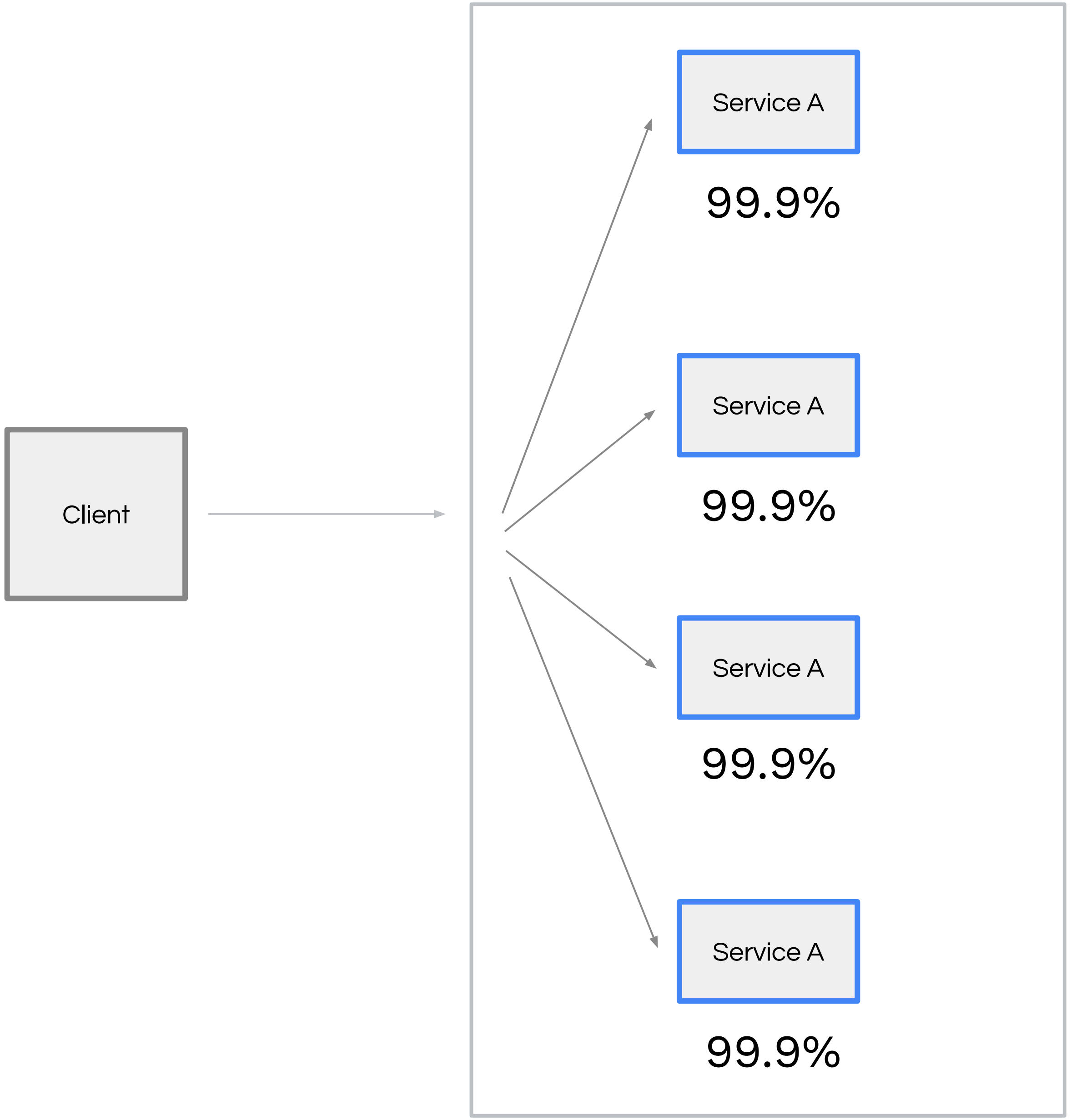
0.999 x 0.999 x 0.999 x 0.999

99.6% SLO

[from Steve McGhee's SLO Conf 2021 Talk]



from Steve McGhee's SLO Conf 2021 Talk



Union (aka parallel)

$1 - (0.001)^4$

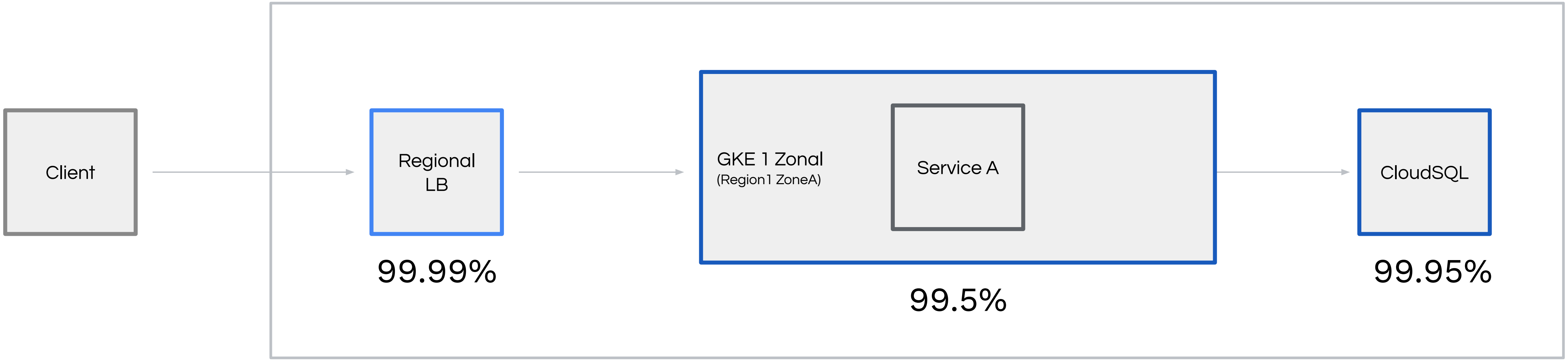
99.999999999% SLO

or 11 nines

This is strictly mathematical and does not include any dependent variables like network, LBs, capacity planning, connectivity, and other dependent services

Building Reliable Platforms on Kubernetes

Archetype 2.1 Single zonal GKE cluster with Cloud SQL



Client

Regional LB

99.99%

GKE 1 Zonal
(Region1 ZoneA)

Service A

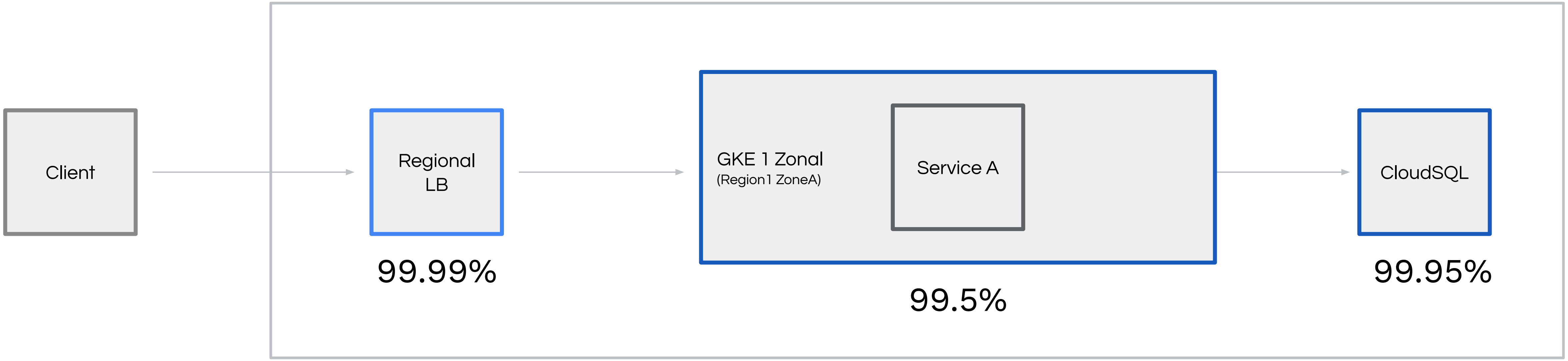
99.5%

CloudSQL

99.95%

GCP SLAs

Archetype 2.1 Single zonal GKE cluster with Cloud SQL



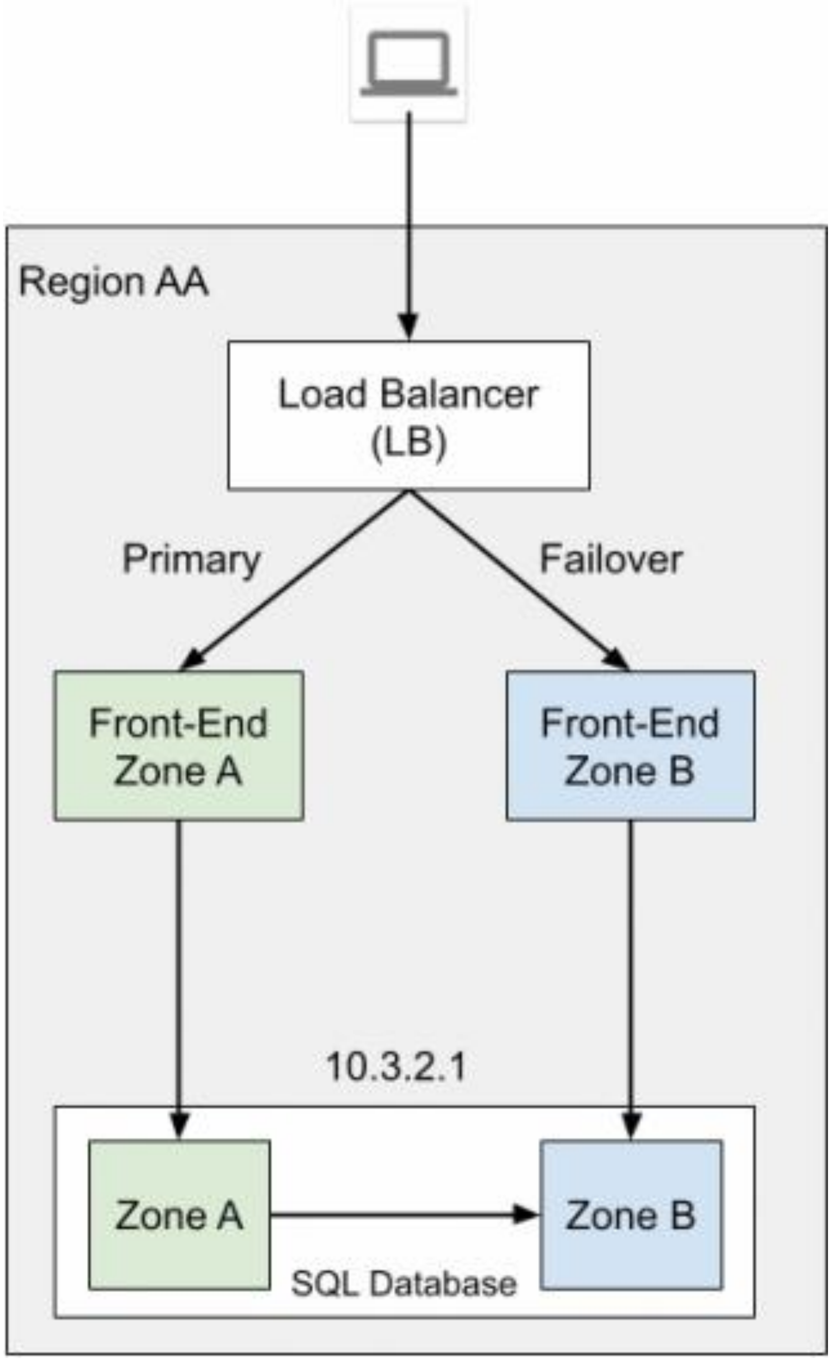
0.9999 x 0.995 x 0.9995

99.44% SLO

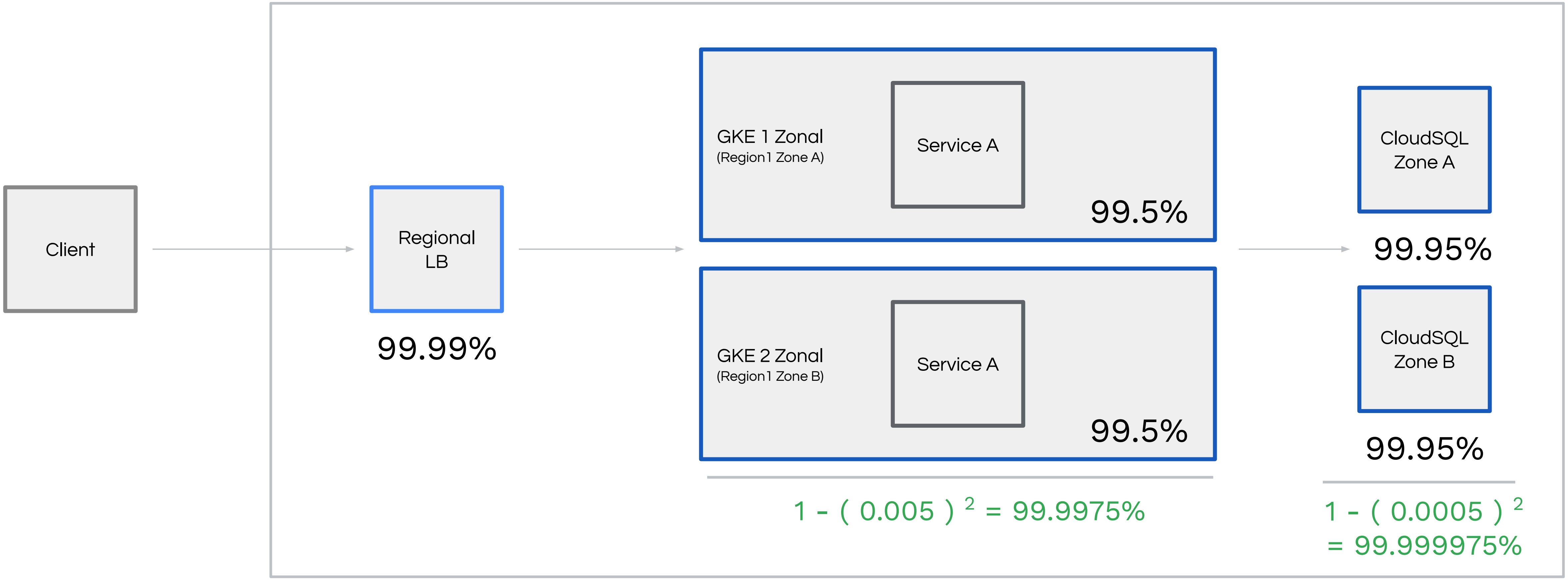
GCP SLAs

Archetype 2.2
Active Passive Zones

- **Deploy** all services of app to two zones in one region
- **Data** in Cloud SQL with a read replica
- **L4 LB** with one backend group
- **Survives zone** failure. Does not survive region failure.
- **Fail-Ops:** Change LB backend, promote read replica
- **Cost:** 2x serving + 2x data (1 replica)
- **Complexity:** Low
- **App Refactoring:** None (lift and shift)



Archetype 2.2 Active Passive Zone with Cloud SQL HA



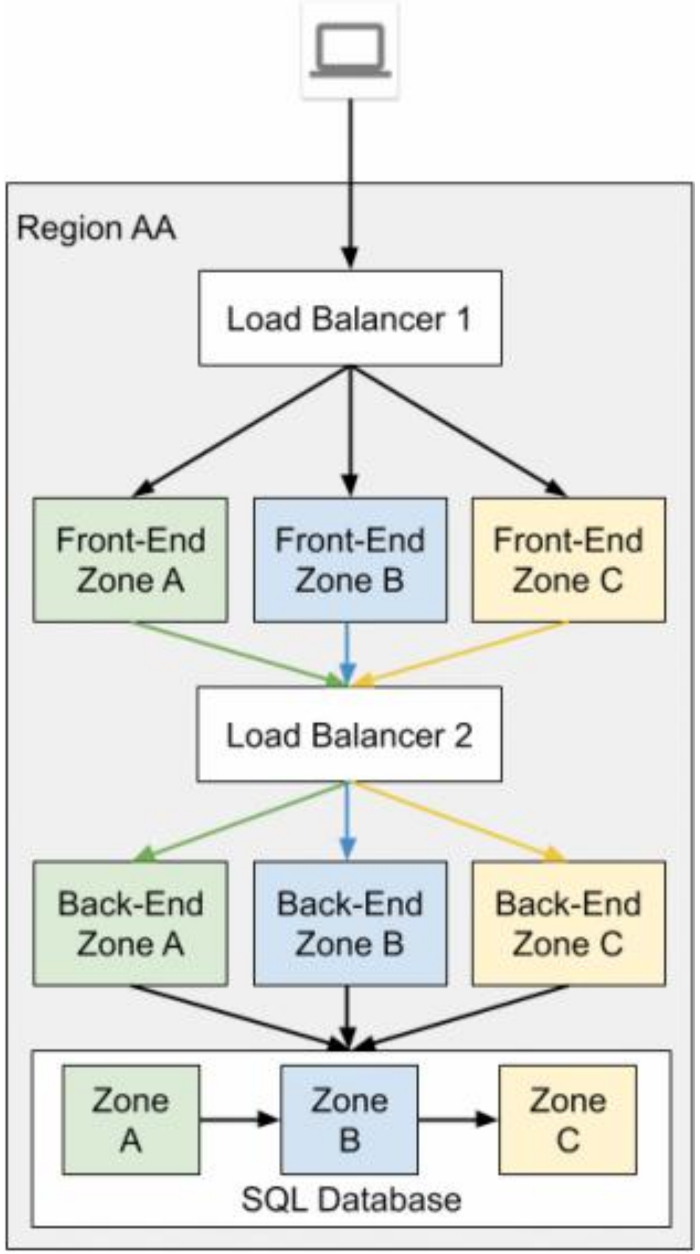
0.9999 x 0.999975 x 0.99999975

99.98% SLO

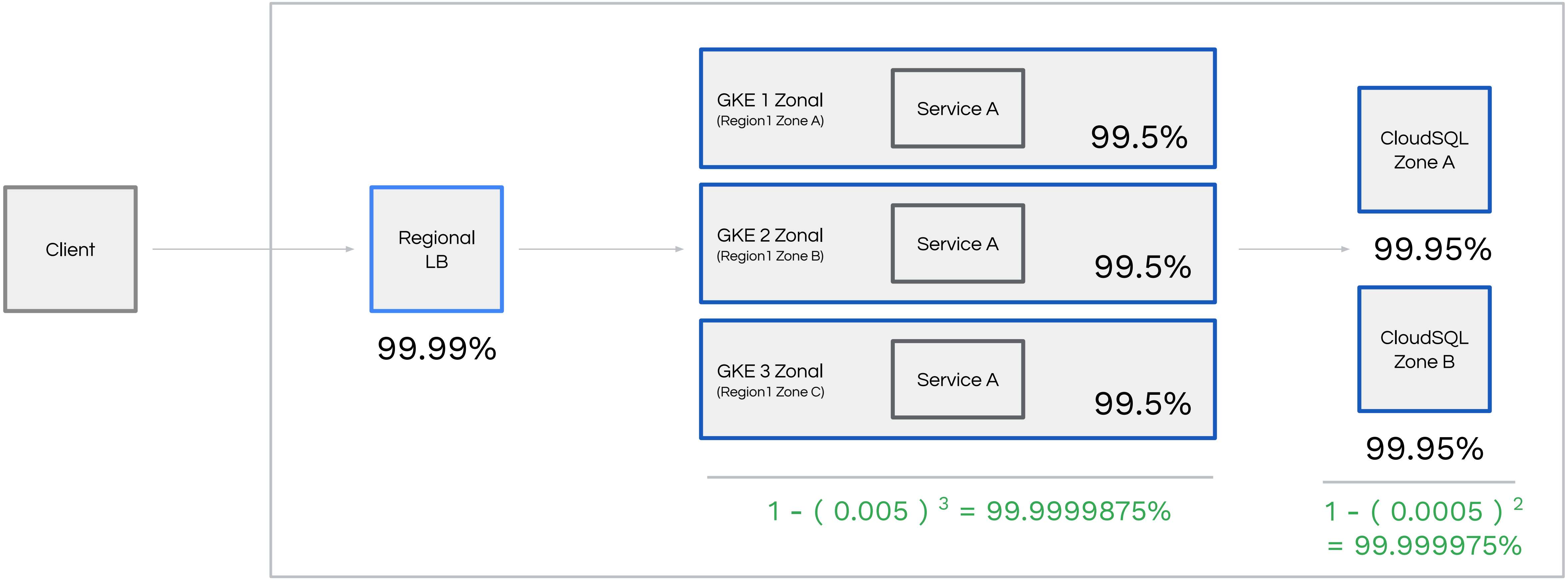
GCP SLAs

Archetype 3.1
Multi Zonal

- **Deploy** all services of app to all three zones in one region
- **Data:** Use HA Cloud SQL
- Use **GLB or RLB** with 3 backend groups
- **Survives zone** failure. Does not survive region failure.
- **Fail-Ops:** Initiate DB failover (testable)
- **Cost:** 1.5x serving + 2x data (SQL HA)
- **Complexity:** Medium
- **App Refactoring:** Low (multi instance)



Archetype 3.1 Multi Zonal with Cloud SQL HA



$1 - (0.005)^3 = 99.9999875\%$

$1 - (0.0005)^2 = 99.999975\%$

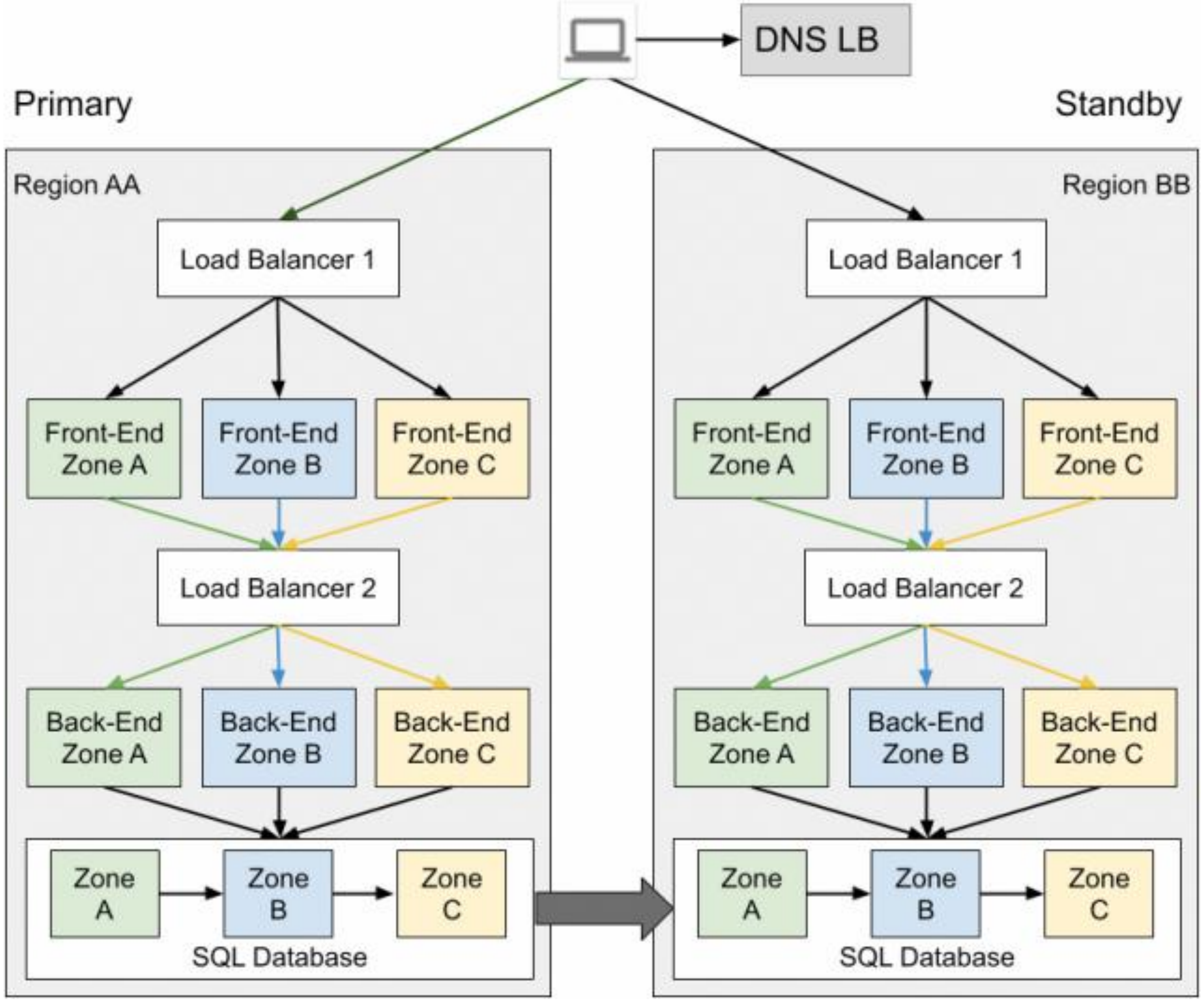
0.9999 x 0.999999875 x 0.99999975

99.99% SLO

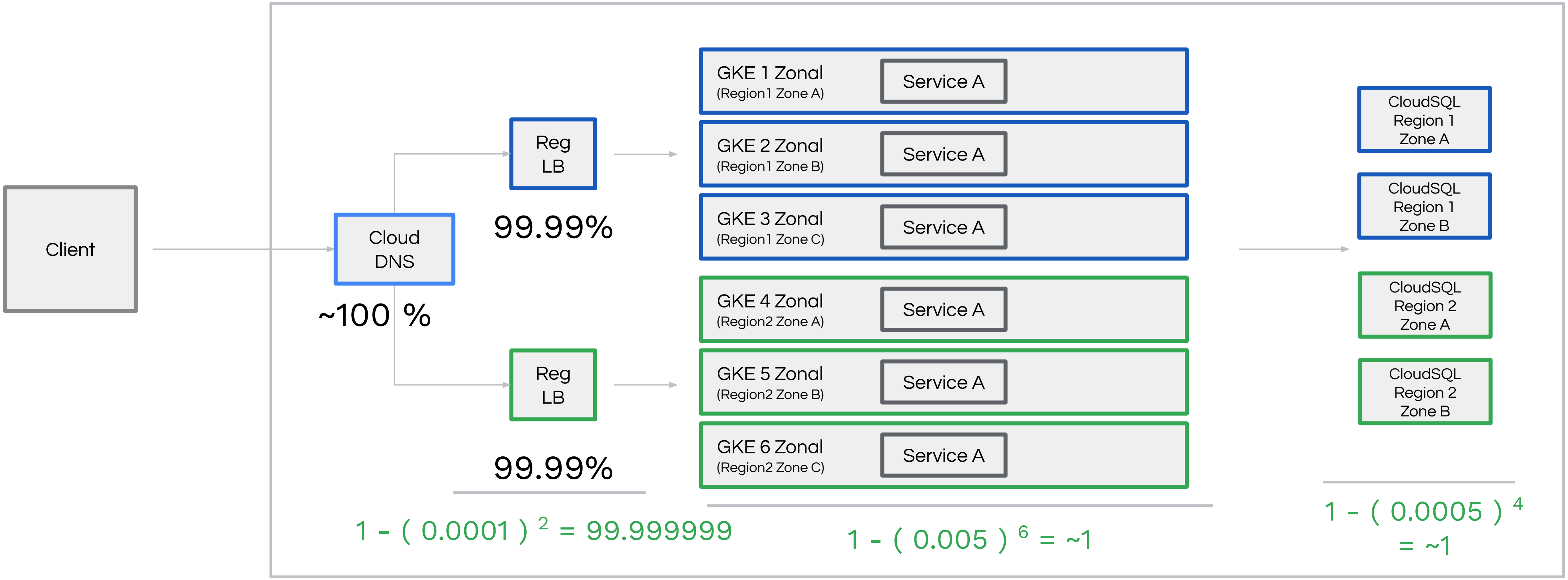
GCP SLAs

Archetype 3.2
Active Passive Region

- **Deploy** all services of app to all three zones in each of two regions
- **Data:** Cloud SQL with cross-region replication
- **Cloud DNS** points at one LB (until disaster)
- **Survives zone and region** failures
- **Fail-Ops**: No action for zone failure.
    - Update DNS to point at standby LB
    - Cross region DR failover process for DB
- **Cost:** 2x serving + 2x data (SQL HA)
- **Complexity**: Medium
- **App Refactoring**: Medium (multi instance, multi regional data)



Archetype 3.2 Active Passive Regions with Cloud SQL HA



$1 - (0.0001)^{2} = 99.999999$

$1 - (0.005)^{6} = \sim 1$

$1 - (0.0005)^{4} = \sim 1$

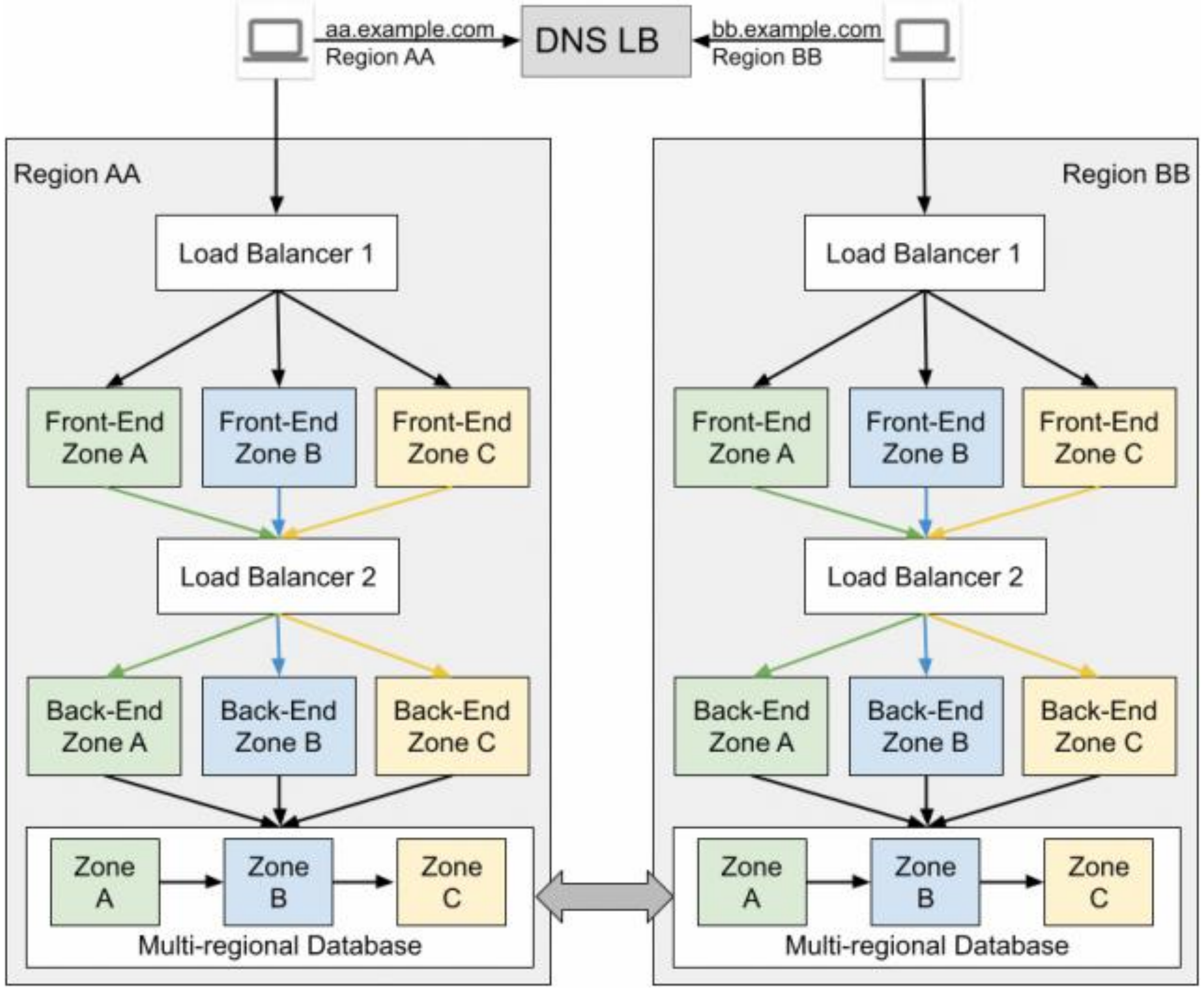
~1 x 0.99999999 x ~1 x ~1

99.999999% SLO

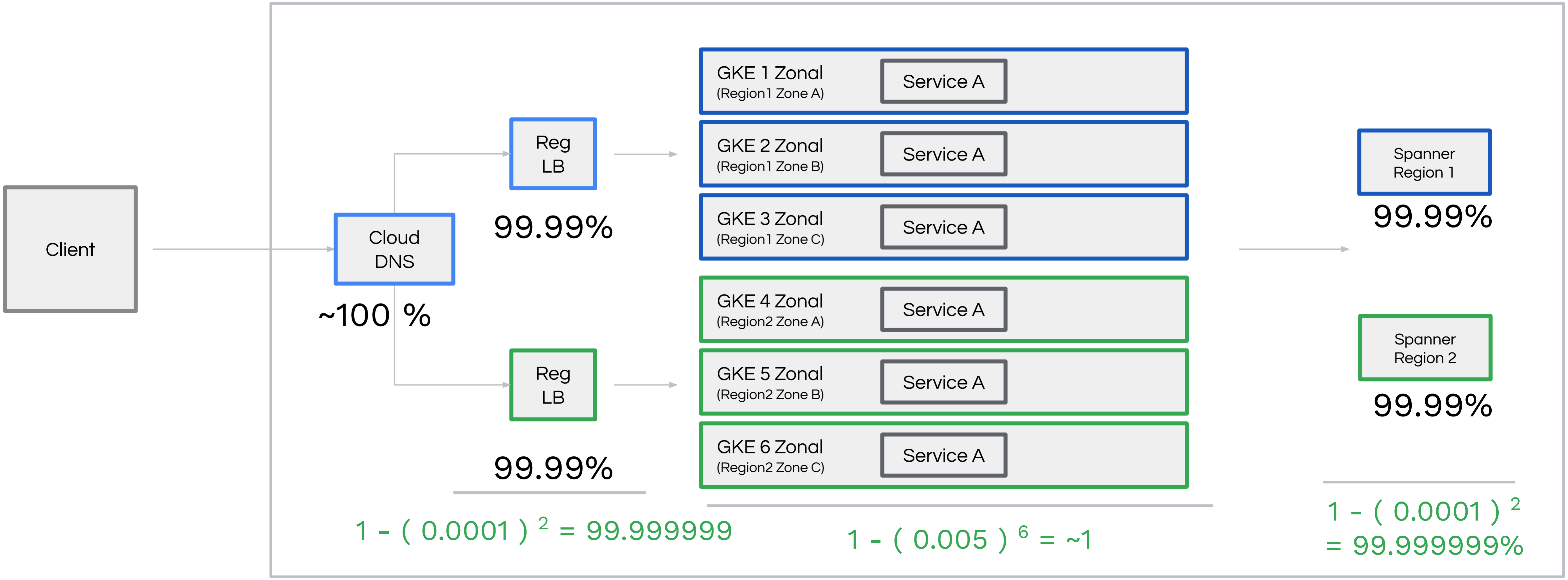
GCP SLAs

Archetype 4.3
Isolated Regions

- **Deploy** all services of app to all three zones in each of two regions
- **Data:** Spanner or CockroachDB
- **Cloud DNS** points at two Regional LBs
- **Survives zone and region** failures. No impact for ½ consumers. Possible manual failover
- **Fail-Ops:** No action for zone failure. Optional regional failover like Arch 3.2
- **Cost:** 1+⅓ cost for zone failure. Scale-up required if optional regional failover expected
- **Complexity:** Medium/High
- **App Refactoring:** Medium (multi instance, multi regional data)



Archetype 4.3 Isolated Regions with Cloud Spanner



Client

Cloud DNS

~100 %

Reg LB

99.99%

Reg LB

99.99%

GKE 1 Zonal (Region1 Zone A)	Service A
GKE 2 Zonal (Region1 Zone B)	Service A
GKE 3 Zonal (Region1 Zone C)	Service A
GKE 4 Zonal (Region2 Zone A)	Service A
GKE 5 Zonal (Region2 Zone B)	Service A
GKE 6 Zonal (Region2 Zone C)	Service A

Spanner Region 1

99.99%

Spanner Region 2

99.99%

$1 - (0.0001)^2 = 99.999999$

$1 - (0.005)^6 = \sim 1$

$1 - (0.0001)^2$
$= 99.999999\%$

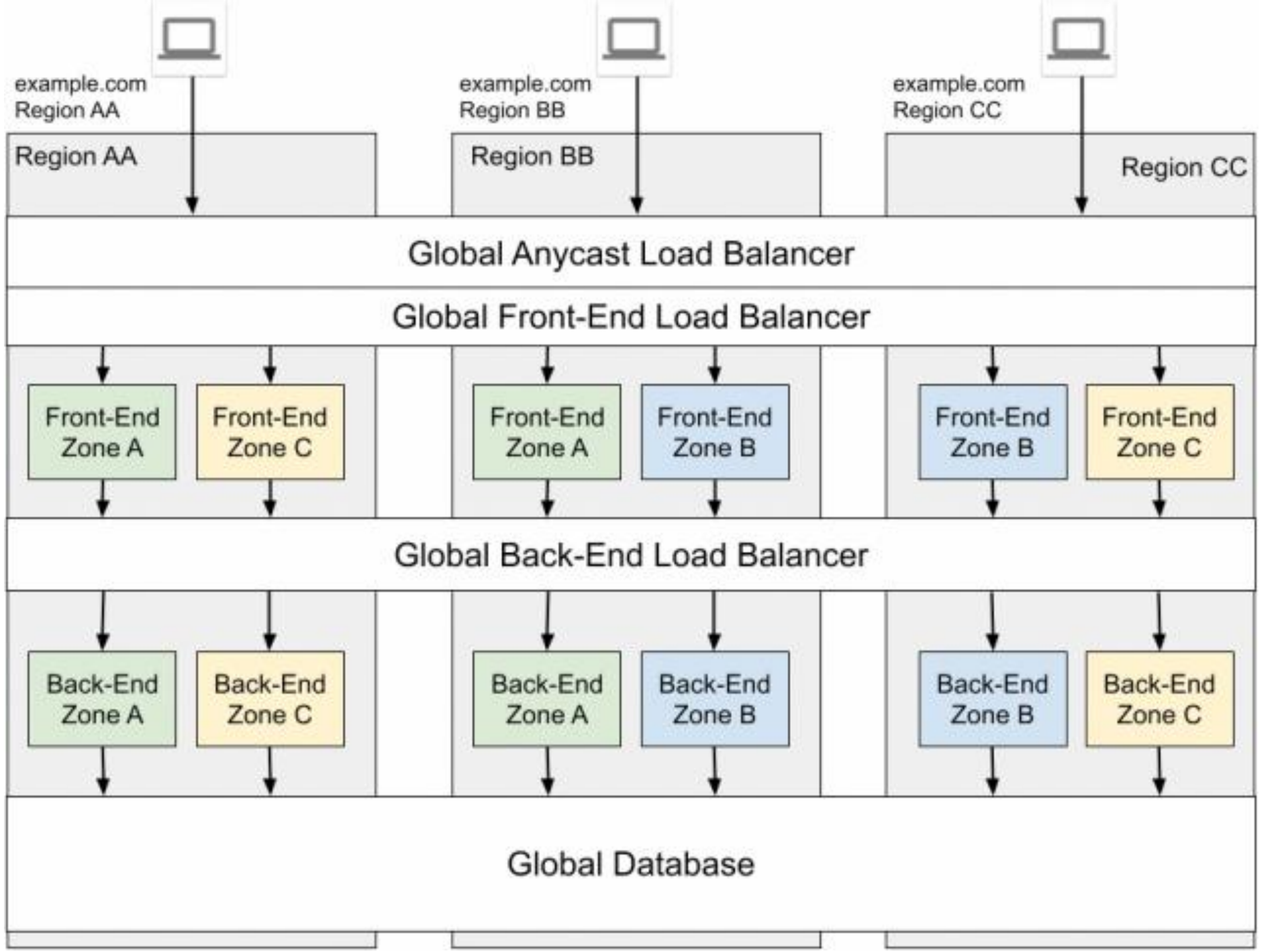
~1 x 0.99999999 x ~1 x 0.99999999

99.999998% SLO

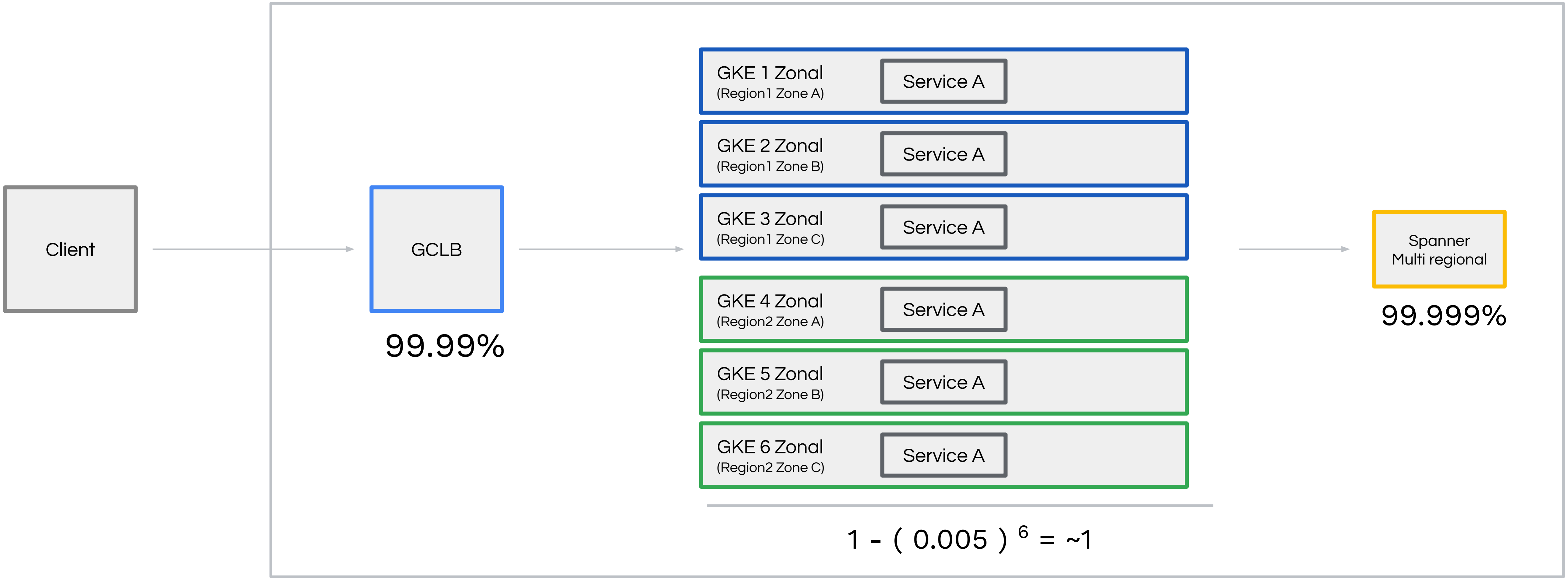
GCP SLAs

Archetype 5.2
Global

- **Deploy** all services of app to all three zones in each of two or more regions
- **Data:** Spanner or CockroachDB
- **Global LB** points at regional backend groups
- **Survives zone and region failures**
- **Fail-Ops:** None
- **Cost:** N+m cost modelling. Global DBs are more expensive
- **Complexity:** High
- **App Refactoring:** High (multi instance, global DBs)



Archetype 5.2 Global with Cloud Spanner (Multi regional)



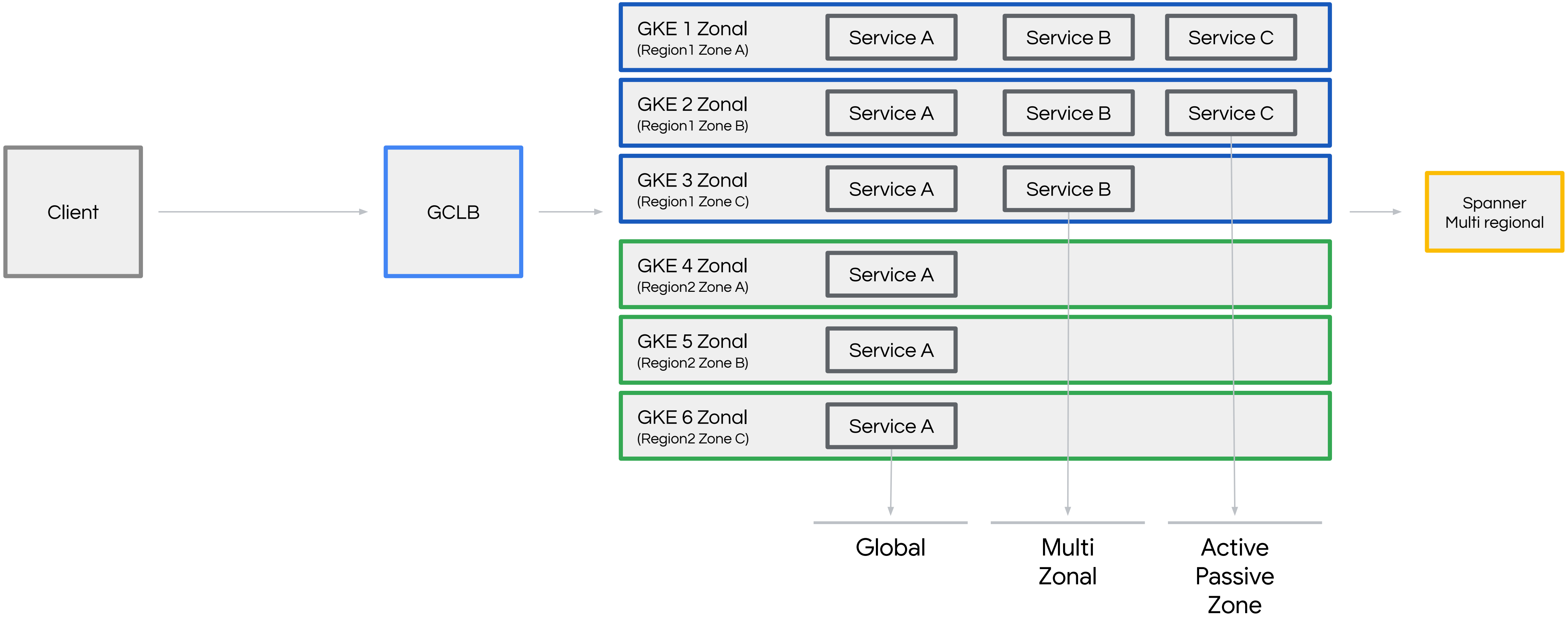
$1 - (0.005)^6 = \sim 1$

0.9999 x 1 x 0.99999

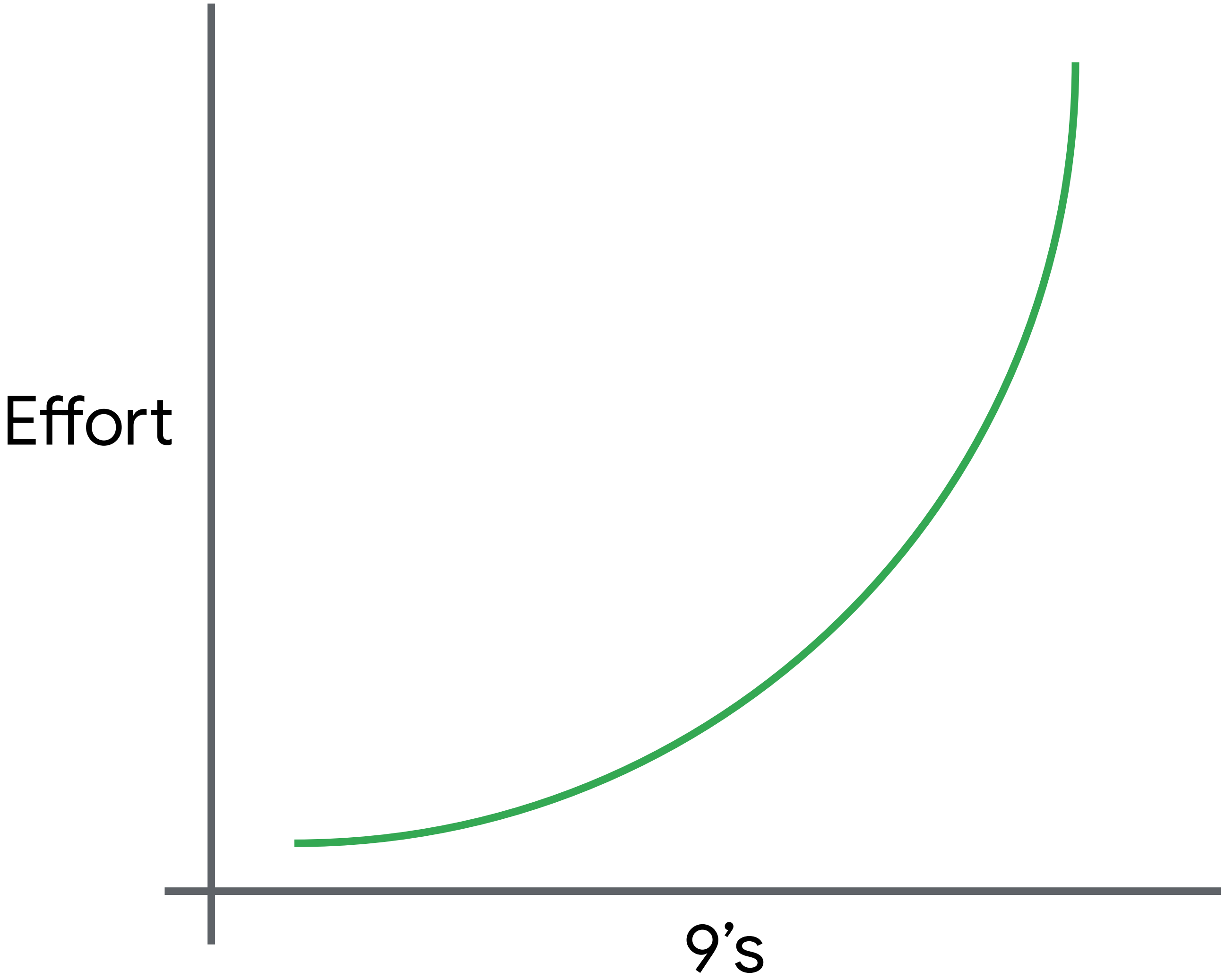
99.99% SLO

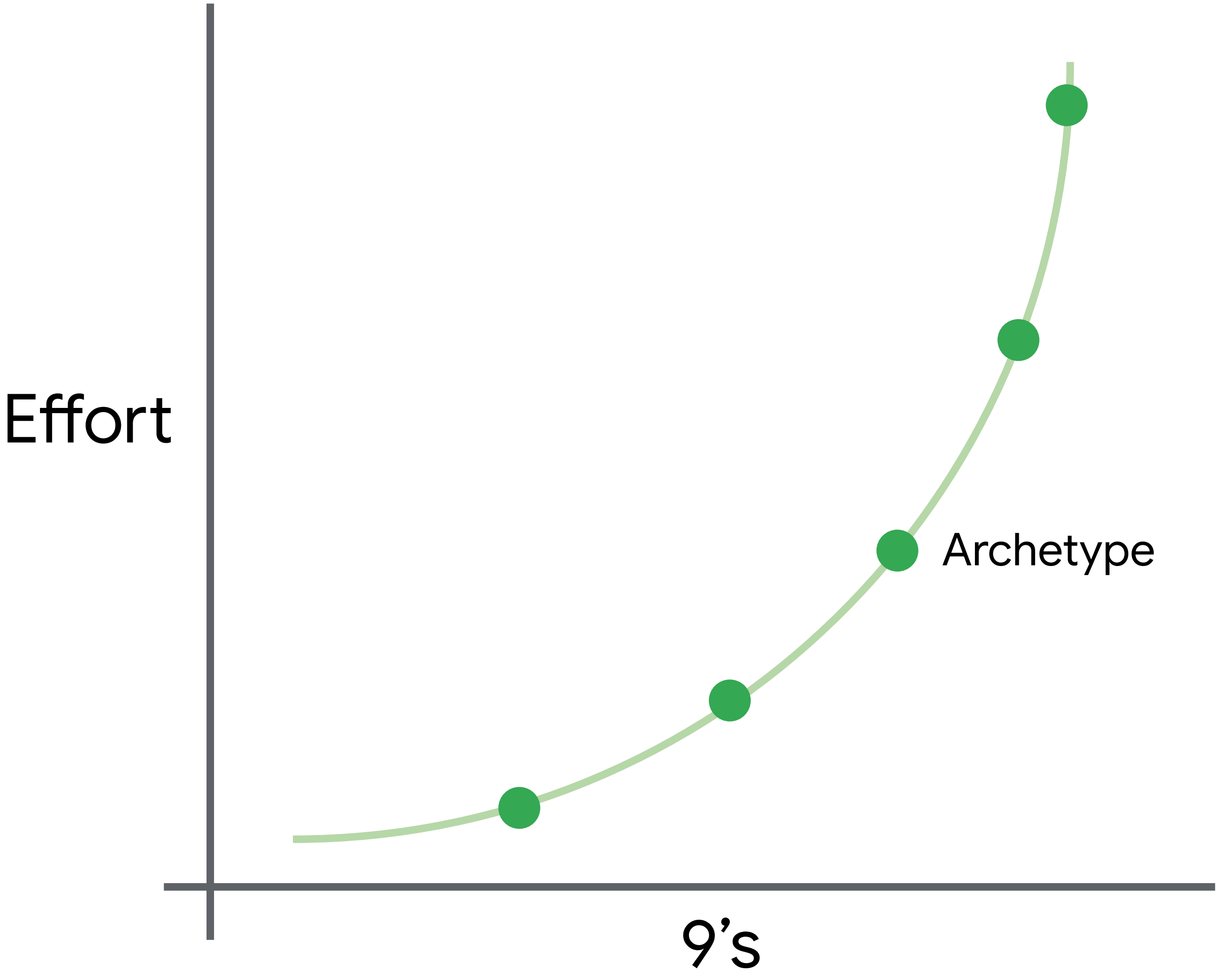
GCP SLAs

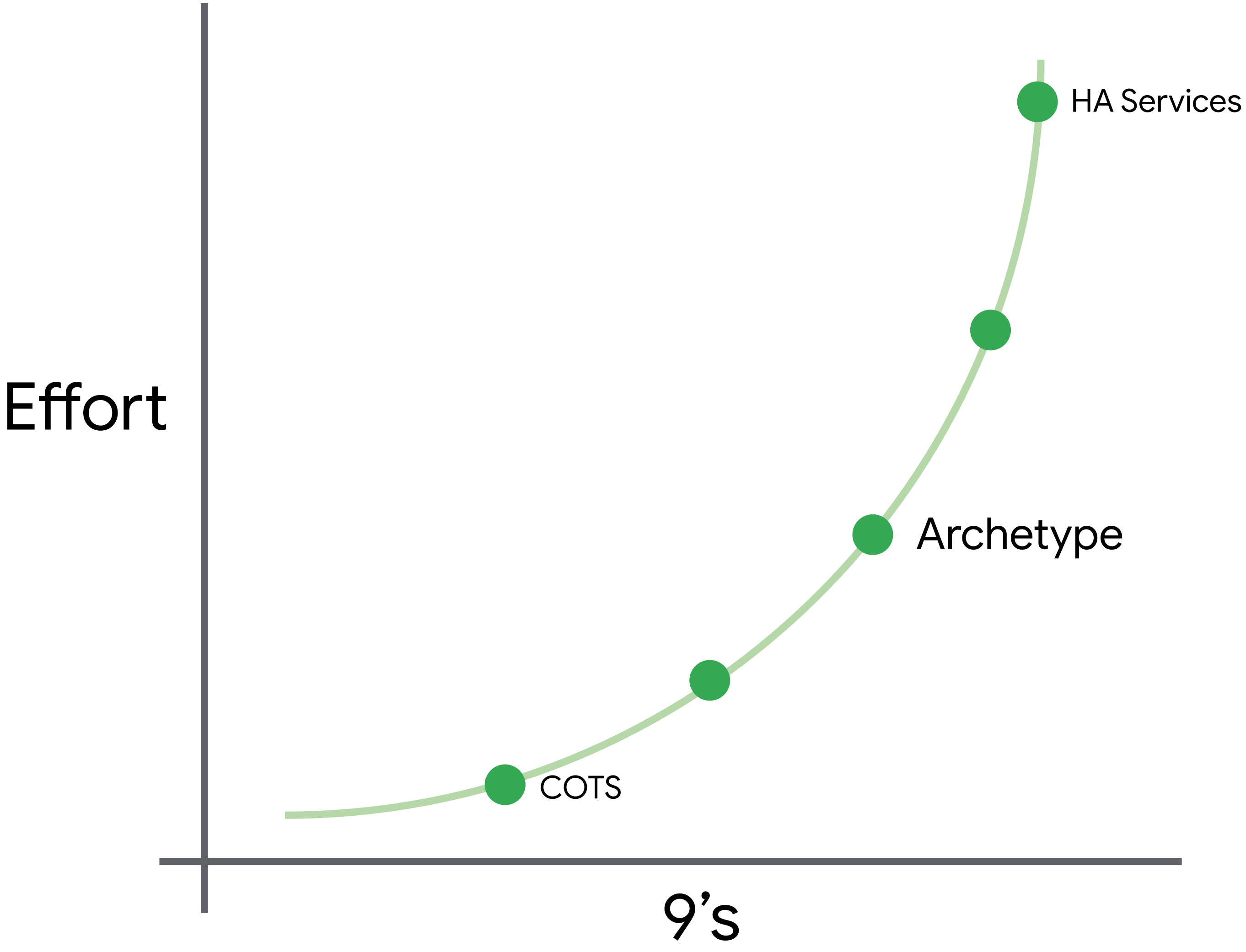
Application with Services using multiple Archetypes

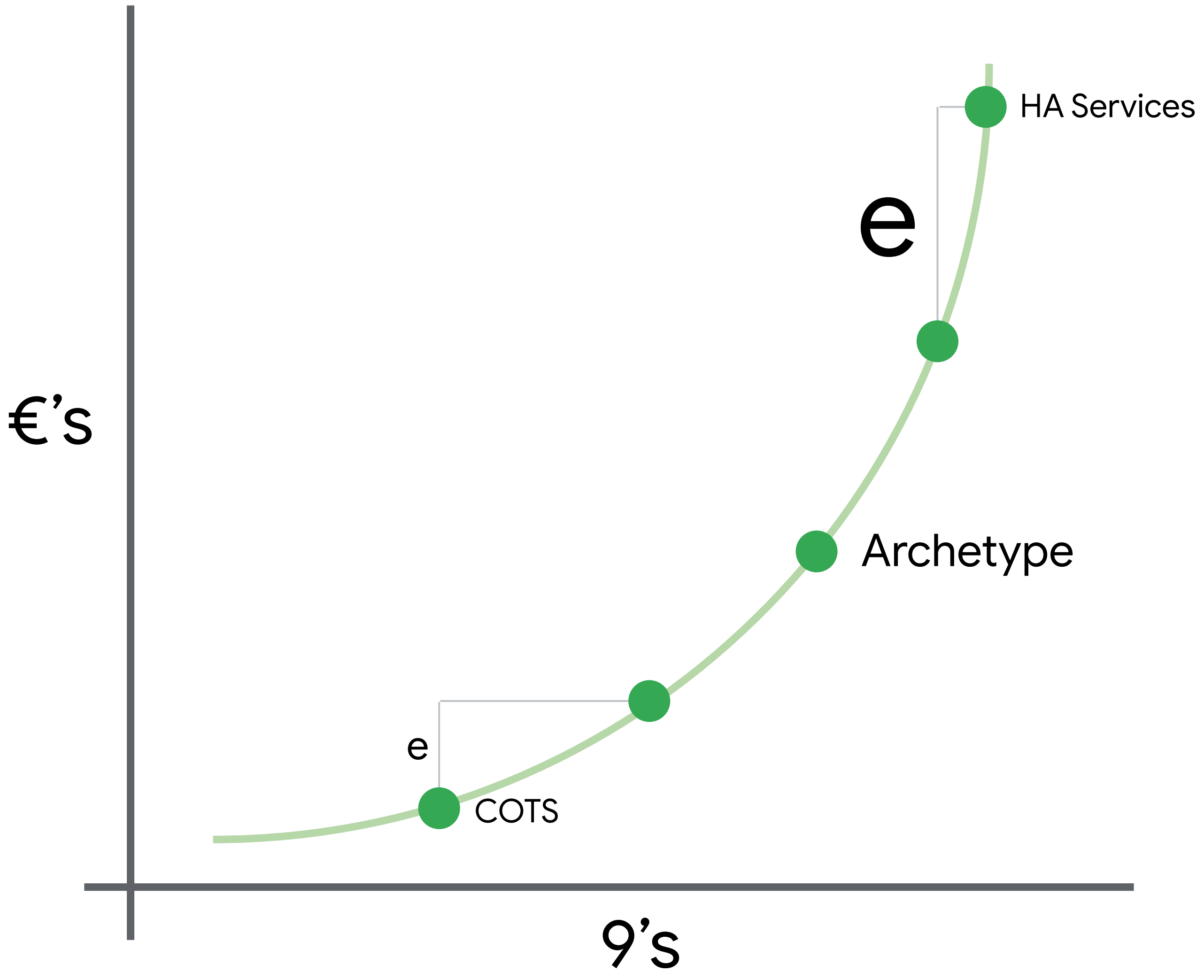


GCP SLAs


Effort
9’s


Effort
Archetype
9's


Effort
HA Services
Archetype
COTS
9's


€'s
9's
HA Services
e
Archetype
e
COTS

Conclusions

- Start with **Archetypes**
- Compose **Services into Applications** that can **degrade gracefully**
- Develop **resilient teams**
  robust platforms
  reliable products

Please scan the QR Code above
to leave feedback on this session

Thank you