

NETFLIX

How Netflix Enables CockroachDB- as-a-Service to Improve Customer Satisfaction Across Multi-Region Environments

Shengwei Wang, Senior Software Engineer, Netflix
Ram Srivatsa Kannan, Software Engineer, Netflix

Agenda

- **Intro**
- Fleet summary
- Why do you need multi-region
- Topology
- Future work

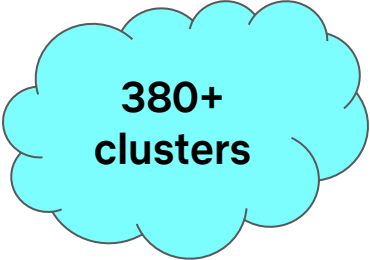
CockroachDB in Netflix

- Who are we?
 - Infrastructure Engineering -
Online Datastores (Cass/EVCache/ES/RDBMS)
 - CockroachDB as a Service
 - Customer self-provisioned
 - Platform automated maintenance
- Use Cases
 - Since 2020
 - Multi-region capabilities
 - HA database
 - Distributed Transactions

Agenda

- Intro
- **Fleet summary**
- Why do you need multi-region
- Topology
- Future work

Fleet Summary



380+
clusters

Fleet Summary

**Prod Clusters
(160+)**

**380+
clusters**

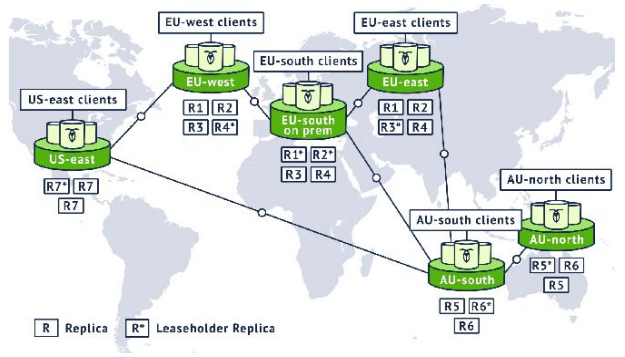
Fleet Summary

380+ clusters

Prod Clusters (160+)

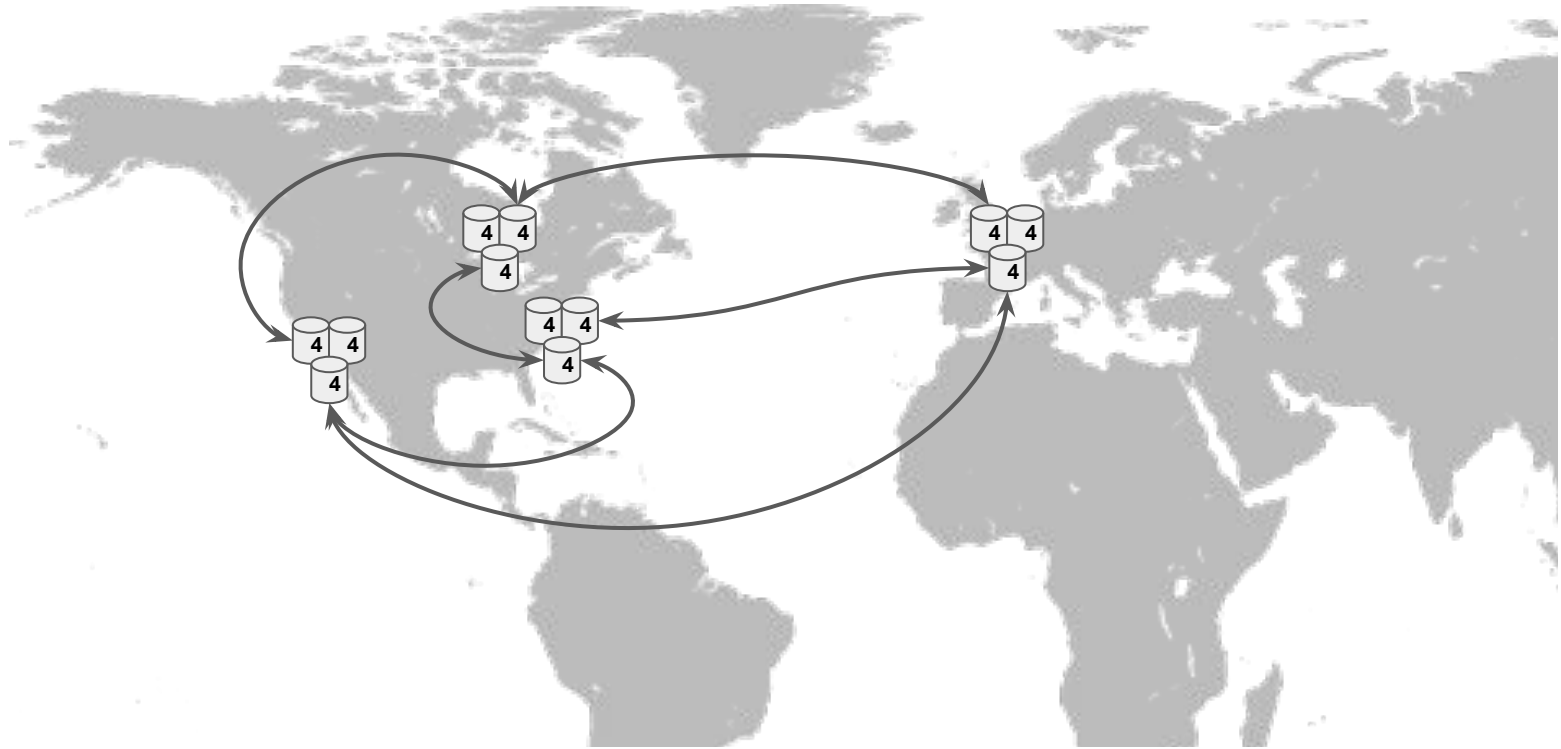


60 clusters

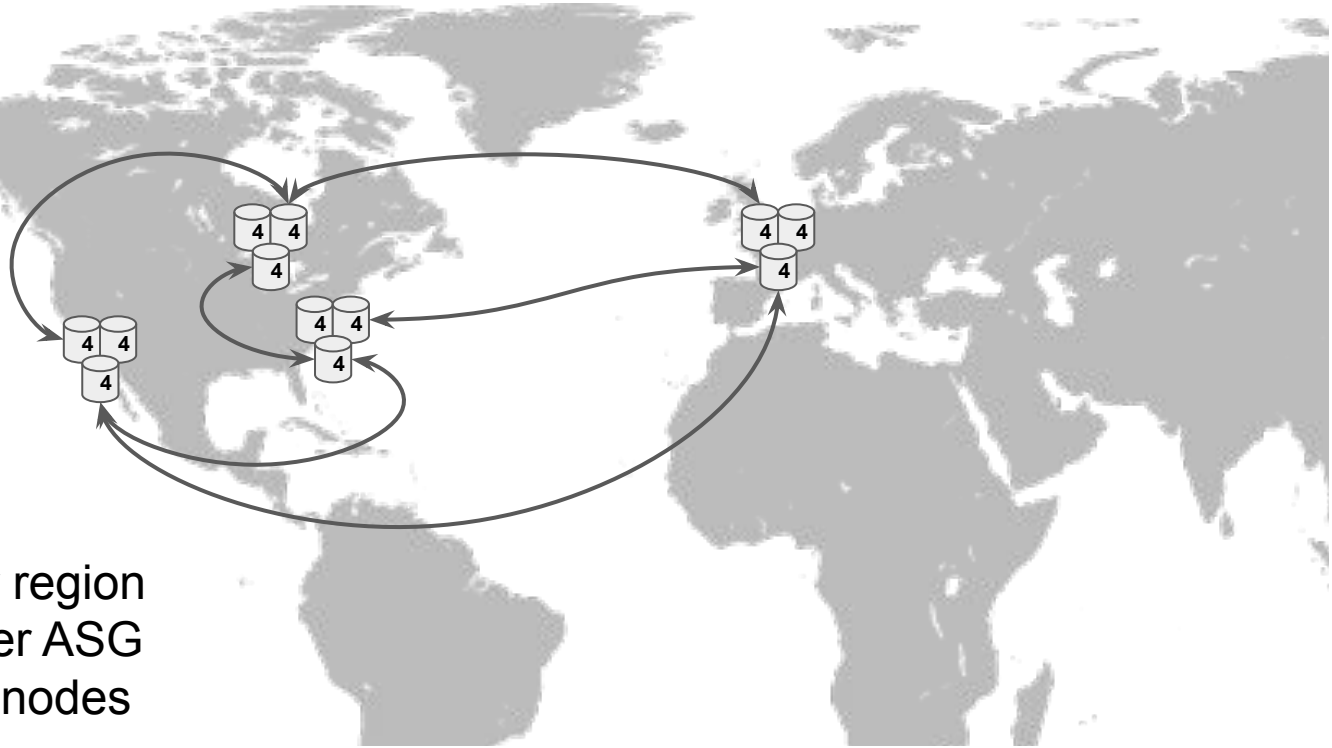


	2 Regions	3 Regions	4 Regions
PROD	12	28	20

Fleet Summary – Eg. crdb_ngpenv



Fleet Summary – Eg. crdb_ngpenv



- 4 Regions
- 3 ASG per region
- 4 nodes per ASG
- Total = 48 nodes

Agenda

- Intro
- Fleet summary
- **Why do you need multi-region**
- Topology
- Future work

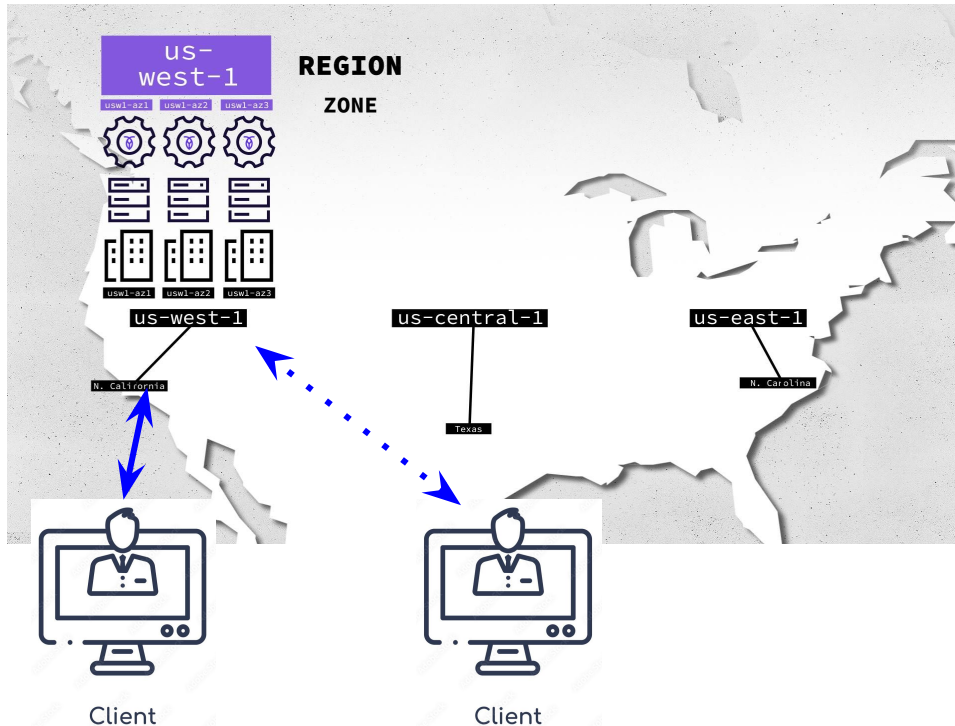
Why do you need multi-region?

Why do you need multi-region?

- Load balancing during traffic evacuation (App side region failover)

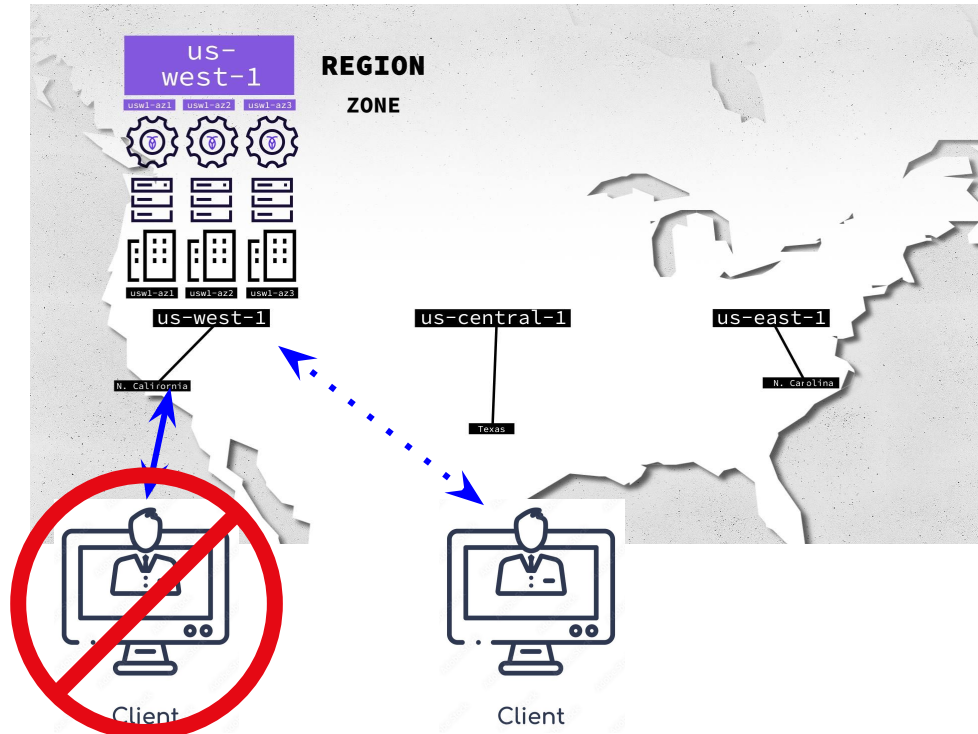
Why do you need multi-region?

- Load balancing during traffic evacuation (App side region failover)



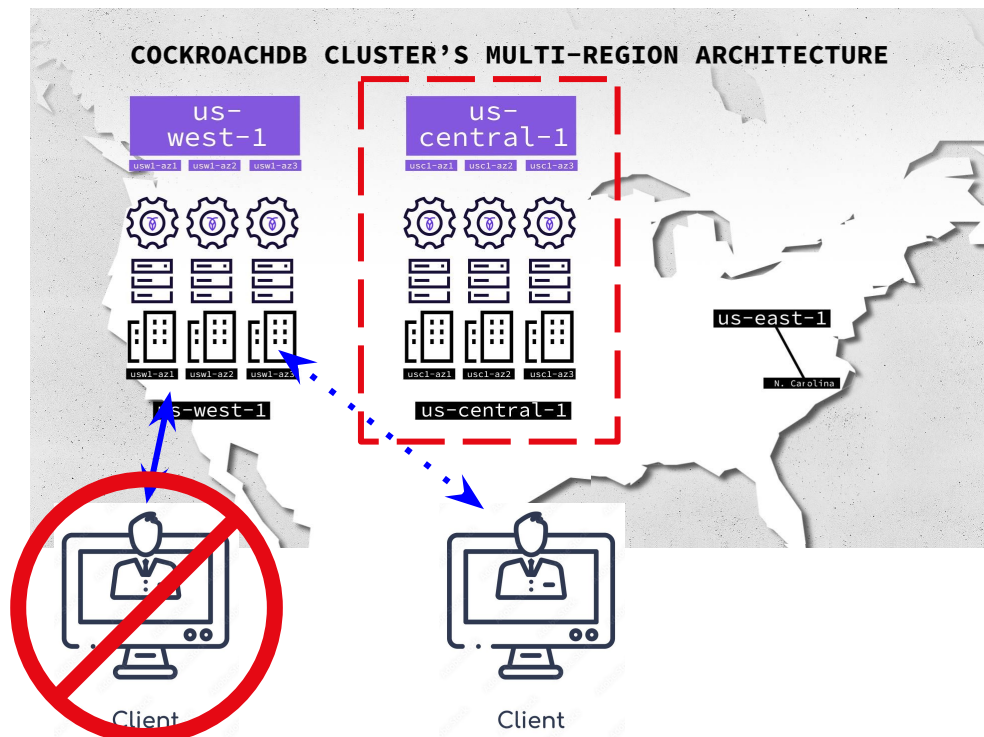
Why do you need multi-region?

- Load balancing during traffic evacuation (App side region failover)



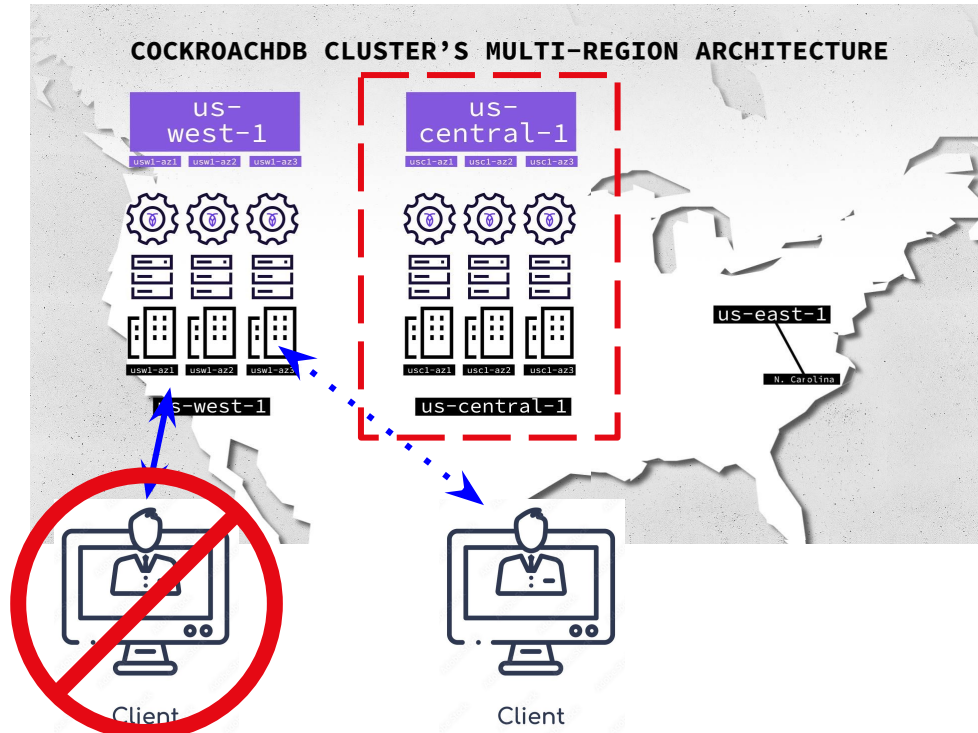
Why do you need multi-region?

- Load balancing during traffic evacuation (App side region failover)



Why do you need multi-region?

- Load balancing during traffic evacuation (App side region failover)



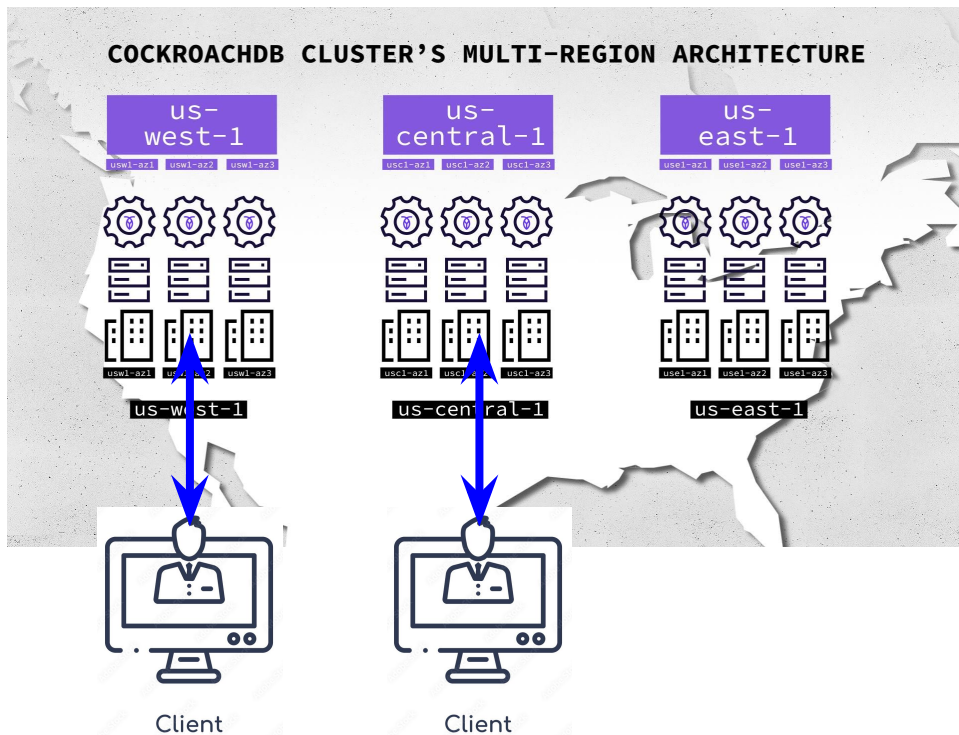
Not very helpful!

Why do you need multi-region?

- Load balancing during traffic evacuation (App side region failover)
- Region Survivability (Database side region failover)

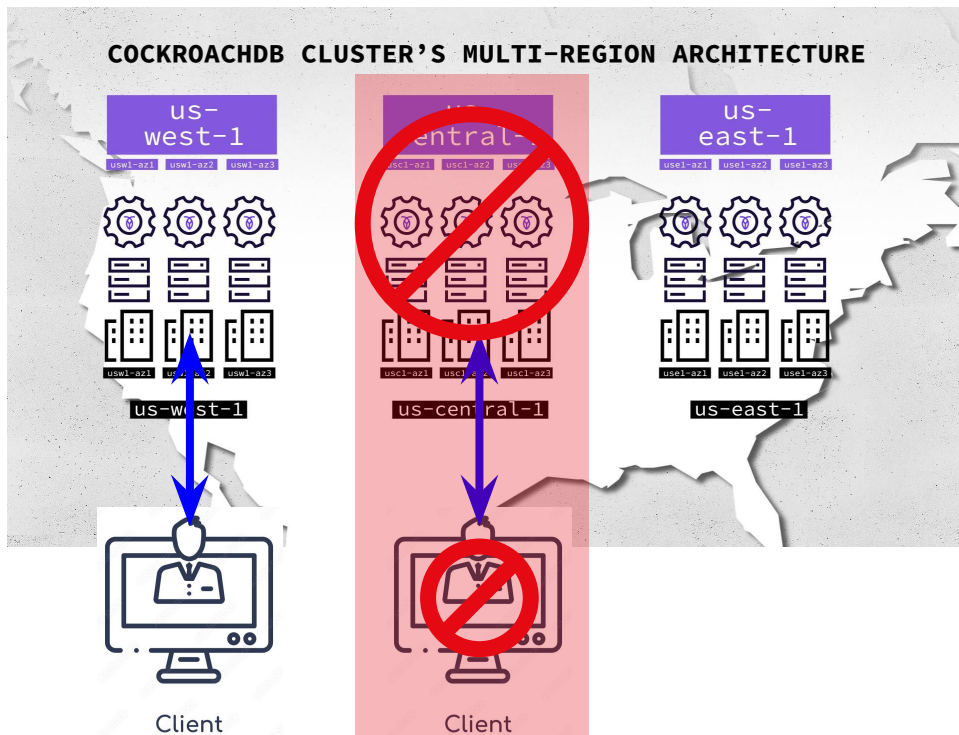
Why do you need multi-region?

- Region Survivability (Database side region failover)



Why do you need multi-region?

- Region Survivability (Database side region failover)



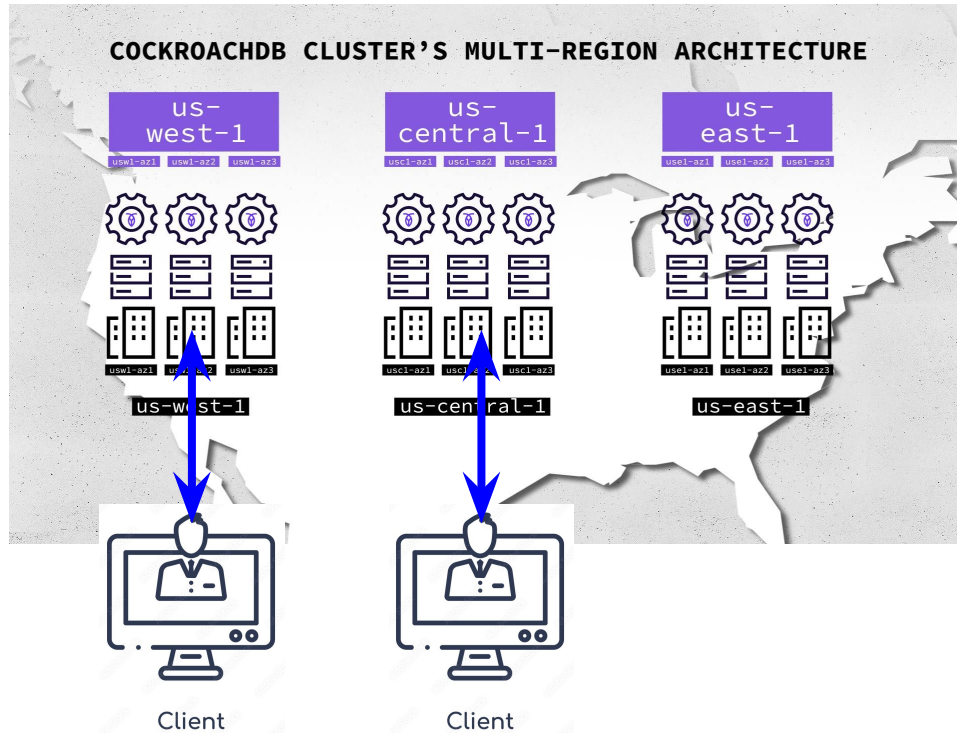
Why do you need multi-region?

- Load balancing during traffic evacuation (App side region failover)
- Region Survivability (Database side region failover)
- Performance

Performance considerations during multi-region?

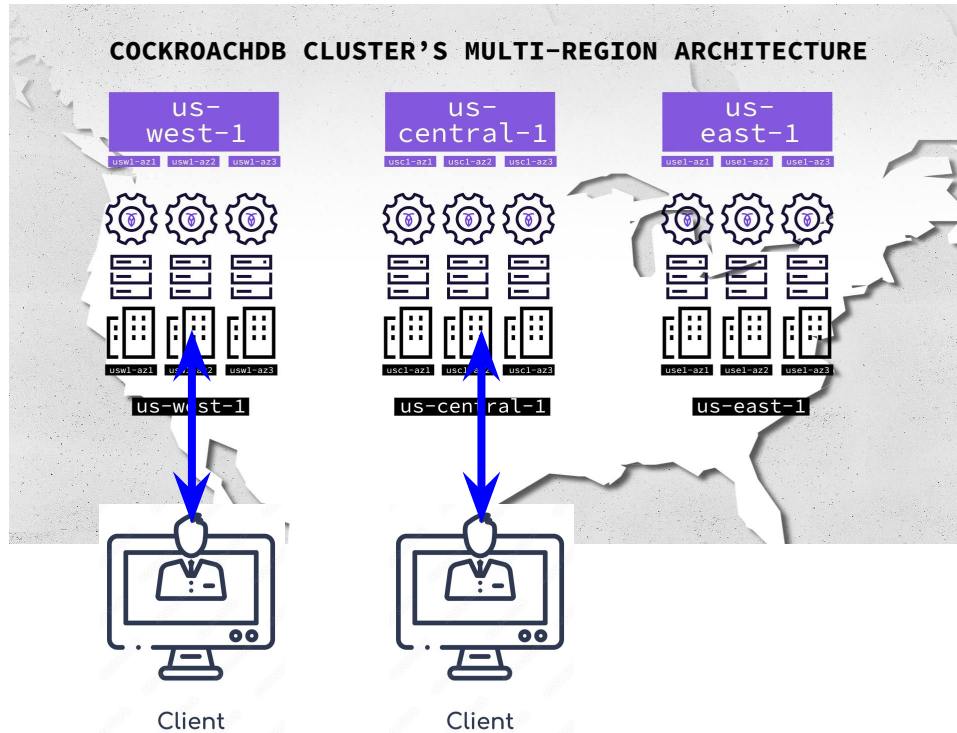
Performance considerations during multi-region?

- Performance (Leaseholder in another region)



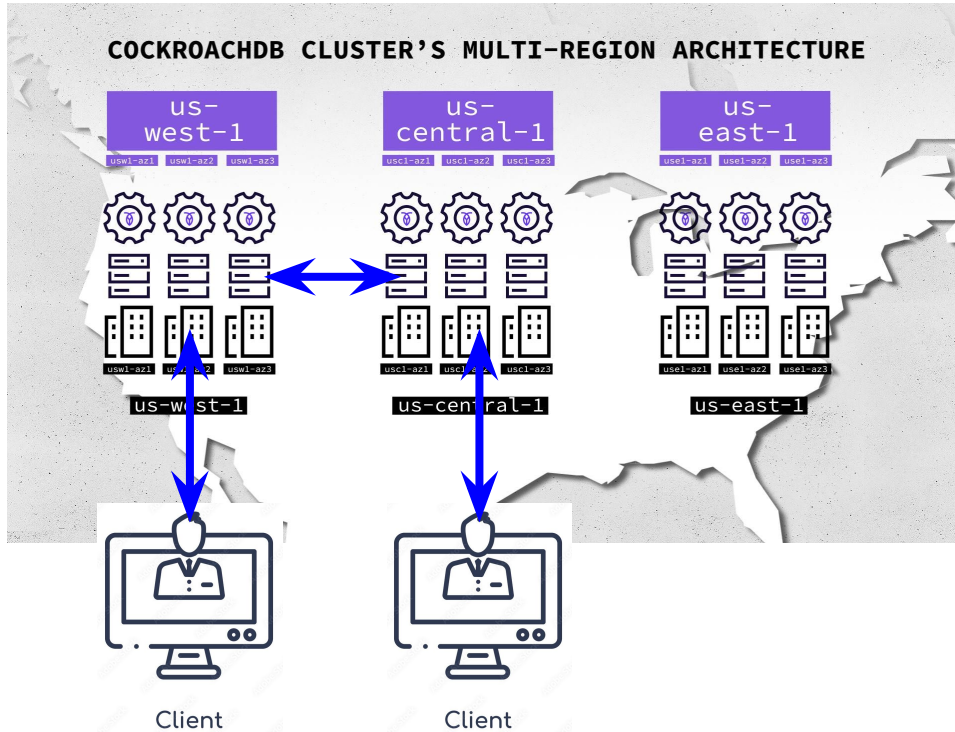
Performance considerations during multi-region?

- Performance (Leaseholder in another region)



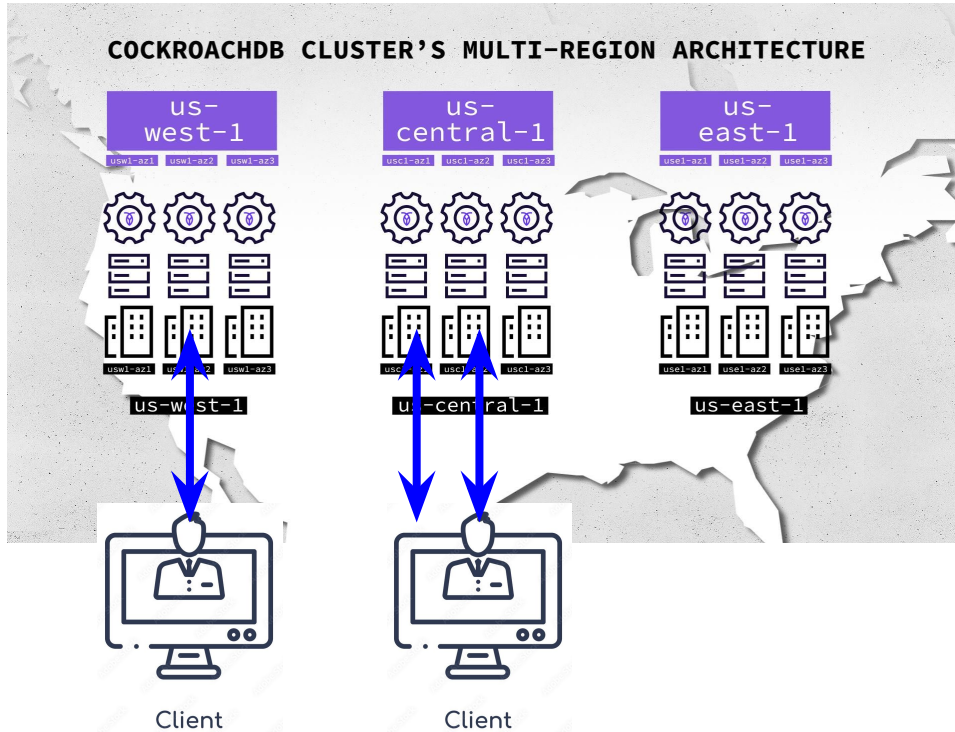
Optimize multi-region setup

- Performance (Leaseholder in another region)



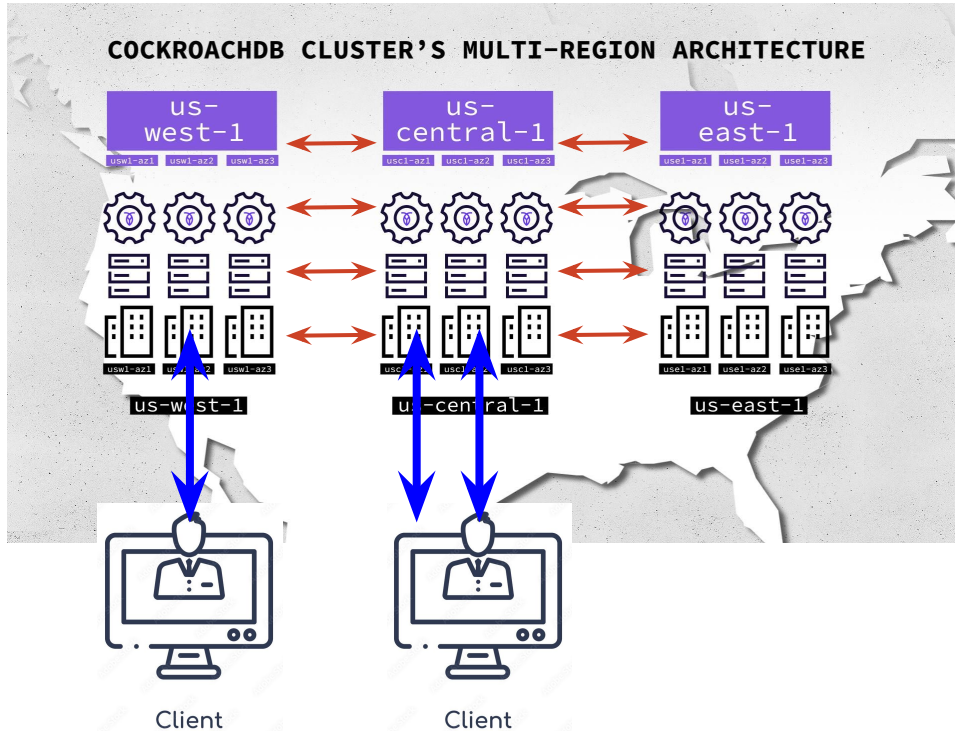
Optimize multi-region setup

- Optimize local reads – Follower reads (stale reads)



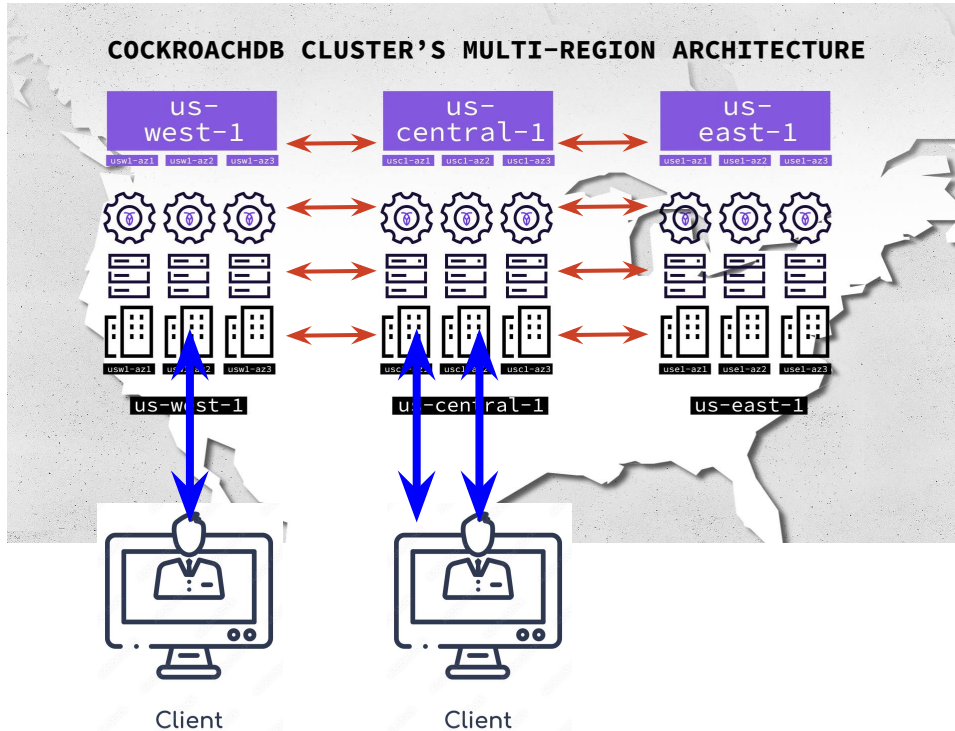
Optimize multi-region setup

- Optimize local reads – Global tables



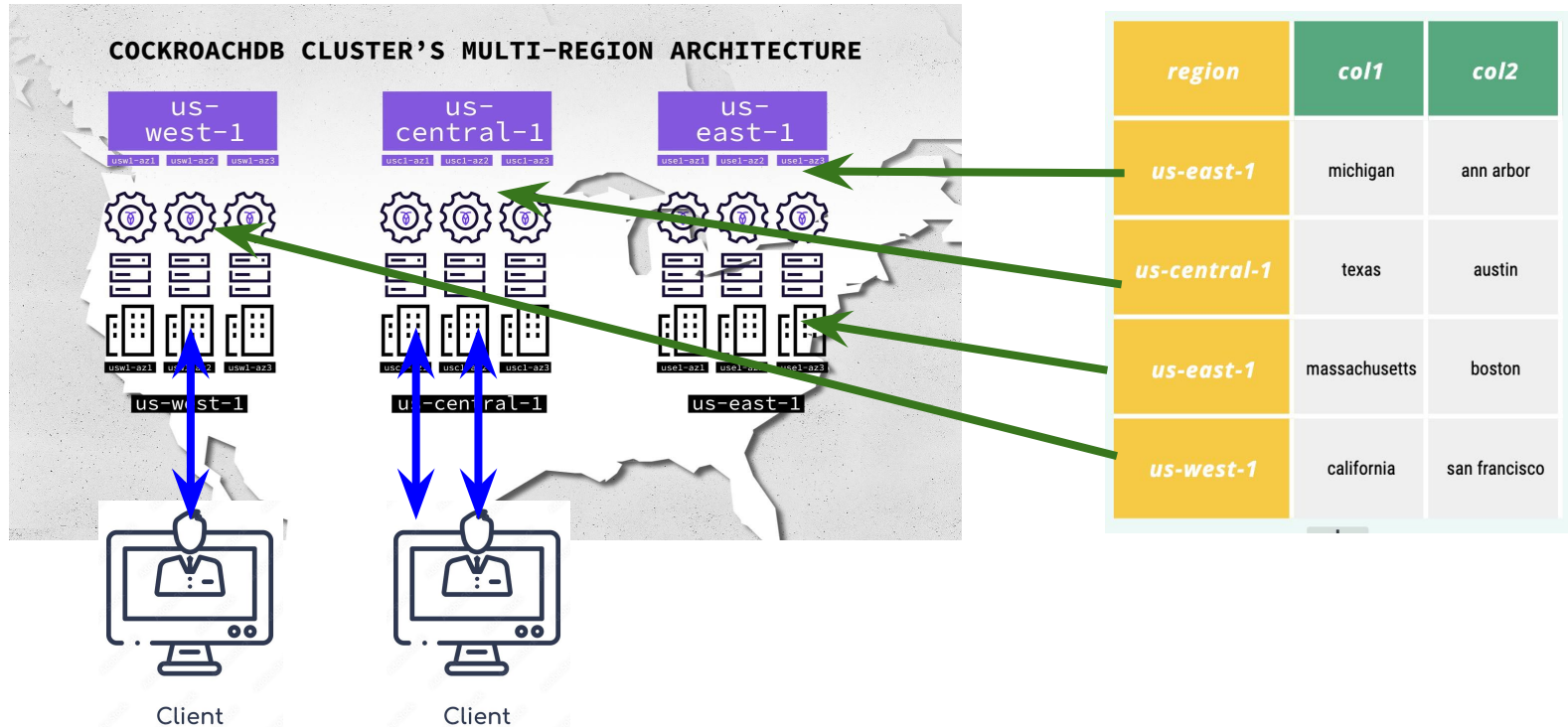
Optimize multi-region setup

- Optimize local reads – Global tables (slower writes by design)



Optimize multi-region setup

- Optimize local reads – Geo partitioned DB (Local reads + efficient writes)



Agenda

- Intro
- Fleet summary
- Why do you need multi-region
- **Topology**
- Future work

Customer interaction

- Ideal:
 - “Here is my use case”
 - “Here is my traffic pattern across all regions”
 - “I want to survive from app failure/ db failure”
 - “Here is the my latency expectation, and here is level of inconsistency I can trade-off”
- Reality
 - “I want to have a multi-region RDBMS system.”
 - “ah, this is not how cassandra worked”
 - “Oh yeah, if by that definition, I don’t really need regional survival”
- Rule of Thumb: Go with default setting, do optimization later

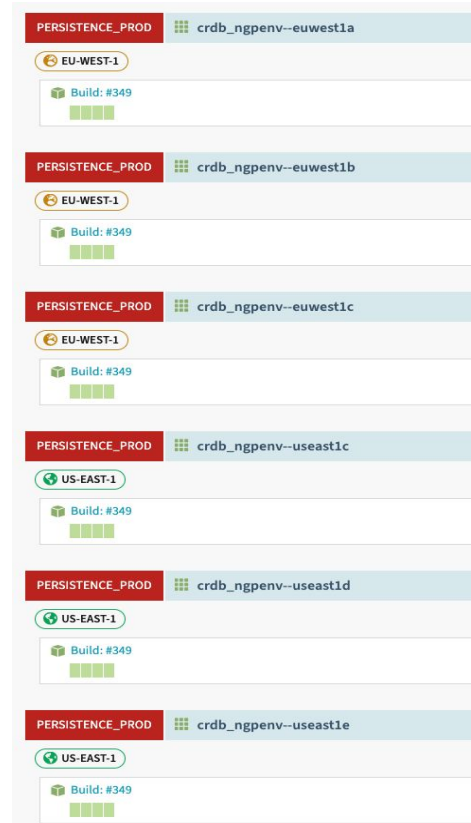
Topology Configuration

- Abstraction vs **Zone Configurations**
 - **Human Readable for discussion**
 - **Machine Readable**
 - **Harder for Customer Tuning**

```
RANGE default      | ALTER RANGE default CONFIGURE ZONE USING
                   |   range_min_bytes = 16777216,
                   |   range_max_bytes = 67108864,
                   |   gc.ttlseconds = 90000,
                   |   num_replicas = 9,
                   |   constraints = '{+region=eu-west-1: 3, +region=us-east-1: 3, +region=us-west-2: 3}',
                   |   lease_preferences = '[]'
```

Topology Configuration

- Default setup when initiate with multi-region
 - 3 AZs * X regions, each region will have 3 replicas
 - Share control plane with other data stores: Cass
 - Easier to understand from customer perspective
 - Minimal Cost implication



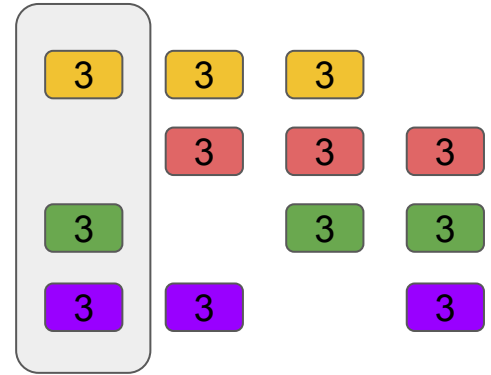
Topology Configuration

- Expand to new regions
 - **Add gateway nodes only**
 - Adding replica
 - Moving replica
 - Minimize app side impact

nodes	Node Count	Uptime	Replicas
▶ us-west-2a	2		197
▶ us-west-2c	3		310
▶ us-west-2b	2		198
▶ us-east-1d	2		200
▶ us-east-1c	2		201
▶ eu-west-1a	2		200
▶ eu-west-1b	2		200
▶ eu-west-1c	2		201
▶ us-east-1e	2		200
▶ us-east-2a	2		46

Topology Configuration - Example

- Application deployed in 4 regions
- Data are geo-shardable
- Regional Survival
- Cross region latency is tolerable, but may be limit to 1



```
ALTER PARTITION us_east_2 OF INDEX foo CONFIGURE ZONE USING
```

```
|   num_replicas = 9,
```

```
|   constraints = '{+region=us-east-1: 3, +region=us-east-2: 3, +region=us-west-2: 3}',
```

```
|   lease_preferences = '[[+region=us-east-2]]'
```

- Further optimization?

Agenda

- Intro
- Fleet summary
- Why do you need multi-region
- Topology
- **Future work**

Future work

1. Abstraction level customer coaching
2. Decide whether to use crlabs abstraction or Netflix customized abstraction
3. Control Plane improvement for Optimization