Availability, Latency and Cost: Withstanding Regional Outages

@aarontblohowiak
aaront@netflix.com
What to expect

- Why?
- Overview!
- Algebraic Models
  - Availability!
  - Latency!
  - Cost!
- Architecture!
Why?
You never let a serious crisis go to waste. And what I mean by that it's an opportunity to do things you think you could not do before.

- Rahm Emanuel
Good, not great.
Good, not great.

1. Instability
Good, not great.

1. Instability
2. Infrequency
Good, not great.

1. Instability
2. Infrequency
3. GOTO 1.
Source: https://martinfowler.com/bliki/FrequencyReducesDifficulty.html
One of my favorite soundbites is: if it hurts, do it more often.

- Martin Fowler
Operational Burden

1. Alerts
Operational Burden

1. Alerts
2. Canaries
Operational Burden

1. Alerts
2. Canaries
3. WoW Metrics
From **Burden** to **Advantage**
In general, freedom and rapid recovery is better than trying to prevent error. We are in a creative business, not a safety-critical business.

- jobs.netflix.com/culture
Overview
Problem Description

Number of Regions
Demand by Hour

100% Capacity

@aaronblohowiak
Problem Description

**Number of Regions**
N+1 Architecture
1+0 (no spare)
1+1
1 + 1 = 200%
2+1
2+1 = 150%
2+1 = 150% ?!?!?!?!
2+1 Overview
Demand by Hour

- Global demand
- Peak global demand
- 1/2 peak demand
- Availability ideal
Algebraic Models
All models are wrong but some are useful

- George Box
Availability
Distribution of Change

Number of Regions

Balance of Traffic
Distribution of Change

Number of Regions

Balance of Traffic
Distribution of Change

**Number of Regions**

Balance of Traffic
Fault Domain Size

Impact of losing a region

\[ \frac{1}{(n + 1)} \]

Outage Impact

Number of Regions (n+1)
<table>
<thead>
<tr>
<th>Distribution of Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Regions</td>
</tr>
</tbody>
</table>

**Balance of Traffic**
Distribution of Change
Number of Regions

Balance of Traffic

\[
\sum_{r=0}^{n+1} \sum_{t=0}^{24} \max(0, R_{rt} - \frac{G_t}{n+1})
\]
\[
\sum_{r=0}^{n+1} \sum_{t=0}^{24} \max\left(0, R_{rt} - \frac{G_t}{n + 1}\right)
\]
Distribution of Change
Number of Regions

Balance of Traffic
Empirical Risk
Latency
Which Latency?
Normal vs Failover
Latency  

Availability \[ \frac{1}{n+1} \]

Cost \[ \frac{1}{n} \]
If you’re successful, hourly demand maps to population by longitude.

- Blohowiak’s Third Law
Measuring Latency
Measuring Latency

\[
\sum_{p=0}^{P} \left( CDF_{a}(P_{p}) - CDF_{b}(P_{p}) \right)
\]
Measuring Latency

\[ CDF_a(0.95) - CDF_b(0.95) \]
Cost
2+1

50%

50%

50%
N+1 Minimal Total Cost

Assume pure throughput-based cost

Cost Compared to 1 Region

Number of Regions (n+1)
In N+1 Architecture, minimal failover overhead is $1/N$. 
In N+1 Architecture, minimal failover overhead is $1/N$.

Cost = 100% + $1/N$
In N+1 Architecture, minimal failover overhead is $\frac{1}{N}$.

Cost = 100% + $\frac{1}{N}$

If costs are pure throughput
Throughput Portion

Database Portion

100%
2+1
2+1
All data everywhere
2+1
All data everywhere
>150%
Data Base Portion
Region Replication Factor
In \textbf{RRF=All}

\begin{align*}
T \text{ is Throughput Cost} & \quad T = (1 - \text{DBP}) \times (1 + \frac{1}{N}) \\
D \text{ is DB Cost} & \quad D = \text{DBP} \times (N + 1) \\
\text{Total} & = T + D
\end{align*}
All Data Everywhere RRF=ALL

Total Cost as a function of DBP
In RRF=2

T is Throughput Cost

\[ T = (1 - \text{DBP}) \times (1 + 1/N) \]

D is DB Cost

\[ D = \text{DBP} \times 2 \]

Total = T + D
Active/Passive Shards or RRF=2

Total Cost as function of DBP

Cost Compared to 1 Region

Number of Regions (n+1)
All Data Everywhere RRF=ALL

Total Cost as a function of DBP

Cost Compared to 1 Region

Number of Regions (n+1)

@aarontblohowiak
Cost Summary

- 50% throughput overhead plus tripled database cost for 3-region RRF=all.
- 25% throughput overhead plus doubled database cost for 5-region RRF=2, plus a lot of complexity.
Architecture
Multi-Site Fault Isolation

- No cross-region Requests!
- Stateless or Async\(^*\) Replication!
  - Cache Replication!
- Change One Region at a Time!
To shard or not to shard?
That is the question.
To shard or not to shard?  
That is the question.

- Steering
To shard or not to shard? That is the question.

- Steering
- Rebalancing & Rehoming
To shard or not to shard?
That is the question.

- Steering
- Rebalancing & Rehoming
- Cost
To shard or not to shard?
That is the question.

- Steering
- Rebalancing & Rehoming
- Cost
- Satellites
To shard or not to shard?
That is the question.

- Steering
- Rebalancing & Rehoming
- Cost
- Satellites
- Graph vs Multi-tenant
How to RRF=2 with $1/N$ overhead?

- Central Savior
- Ring
- Custom Hashing
Central Savior
Central Savior
Ring Regions
Ring Regions
Ring Regions
One More Thing
What percentage of your outages come from regional failures?
Many of the availability benefits come from isolation, not regions.
What percentage of your outages come from *database* failures?
Maybe for you and your org having logical stacks makes the most sense.
Closing Thoughts
Questions?

@aaronblohowiak
Choose Your Own Adventure
What do you want more details on?

- Steering
- Scaling
- Demand Mapping