Real Life is Uncertain. Consensus Should Be Too!

Reginald Frank, Octavio Lomeli, Neil Giridharan,

Soujanya Ponnapalli, Marcos K. Aguilera, Natacha Crooks

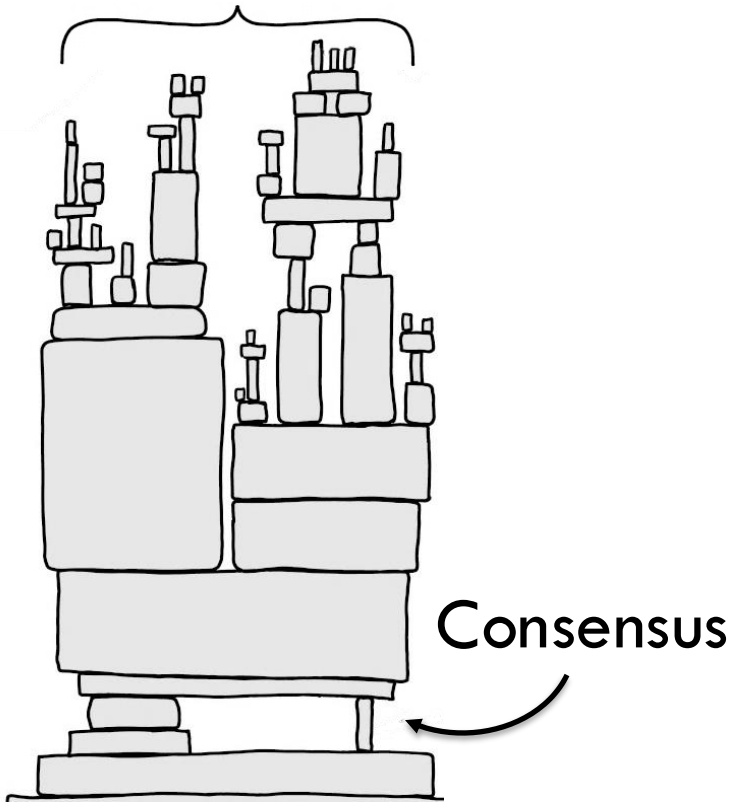
Real Life is Uncertain. Consensus Should Be Too!

Reginald Frank, Octavio Lomeli, Neil Giridharan,

Soujanya Ponnapalli, Marcos K. Aguilera, Natacha Crooks

Consensus and Distributed Systems

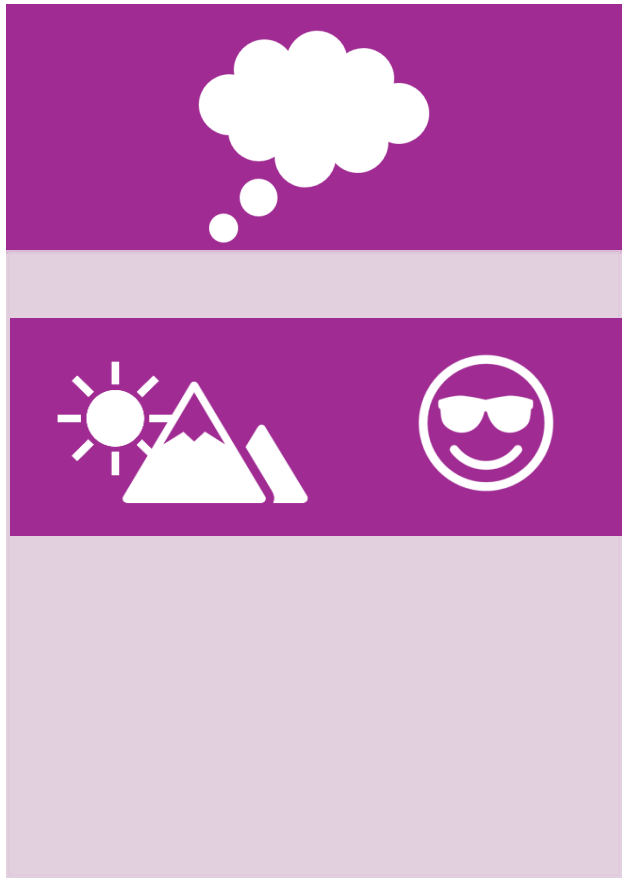
Distributed Systems

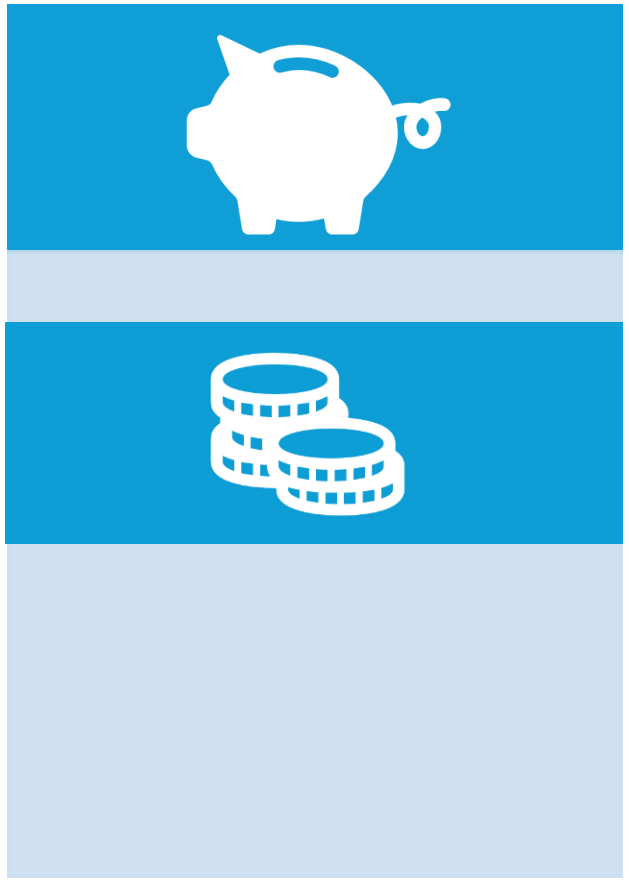


The state of distributed systems

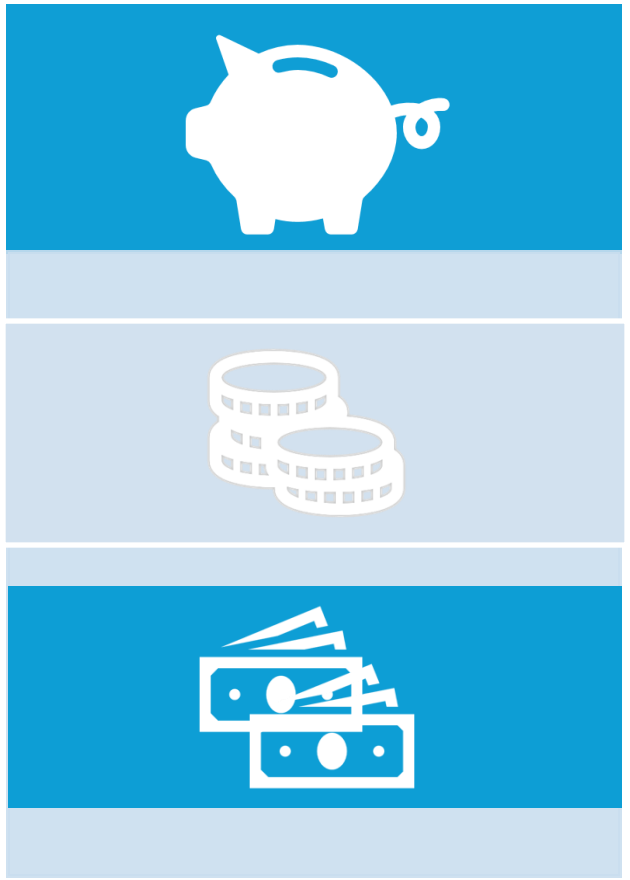
Car Insurance Policies

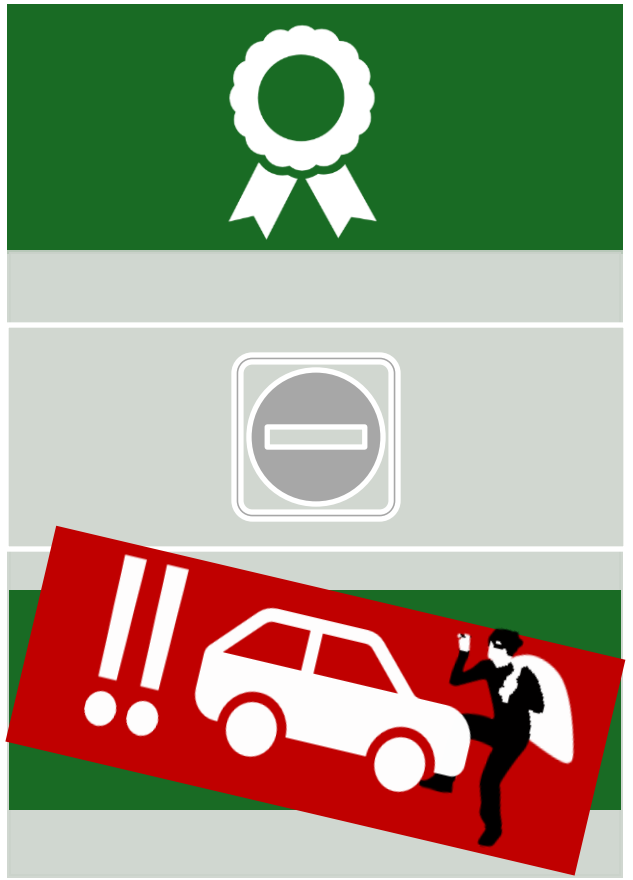
Car Insurance Policies



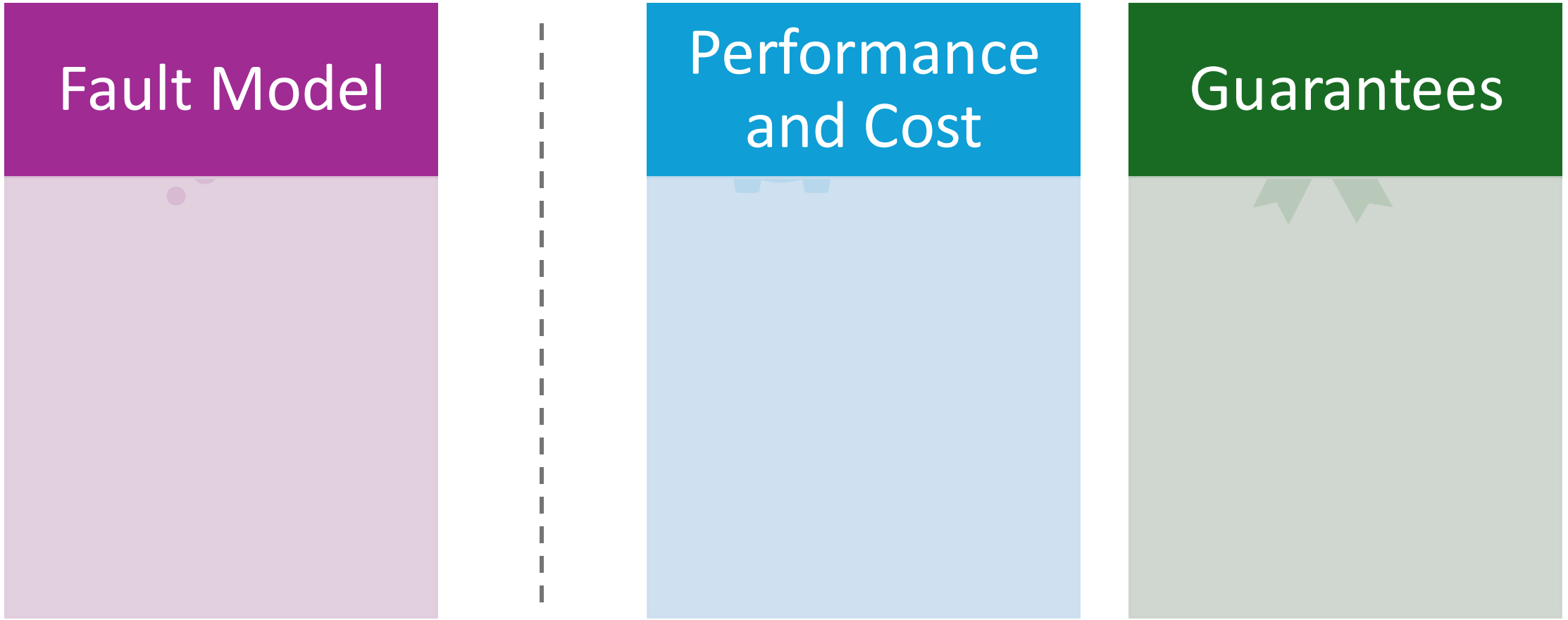


Car Insurance Policies





Consensus is an Insurance Policy

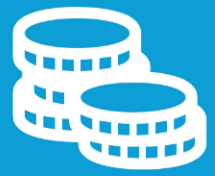


Consensus is an Insurance Policy

Fault Model

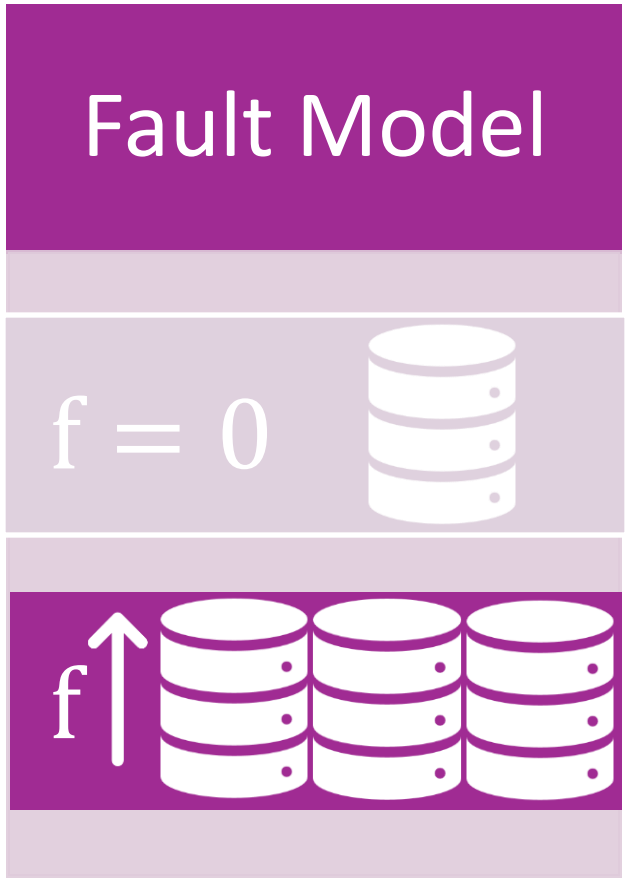
$f = 0$

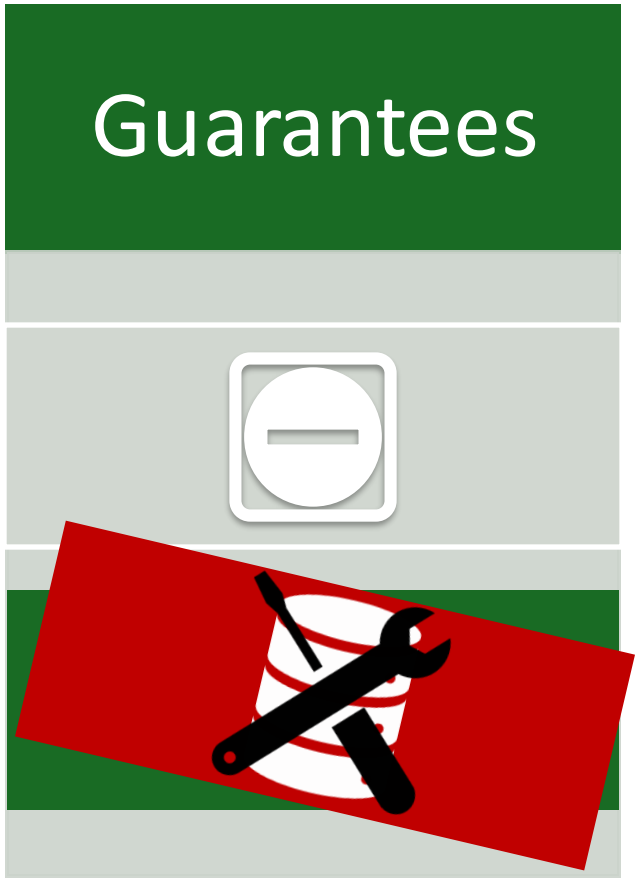
Performance and Cost



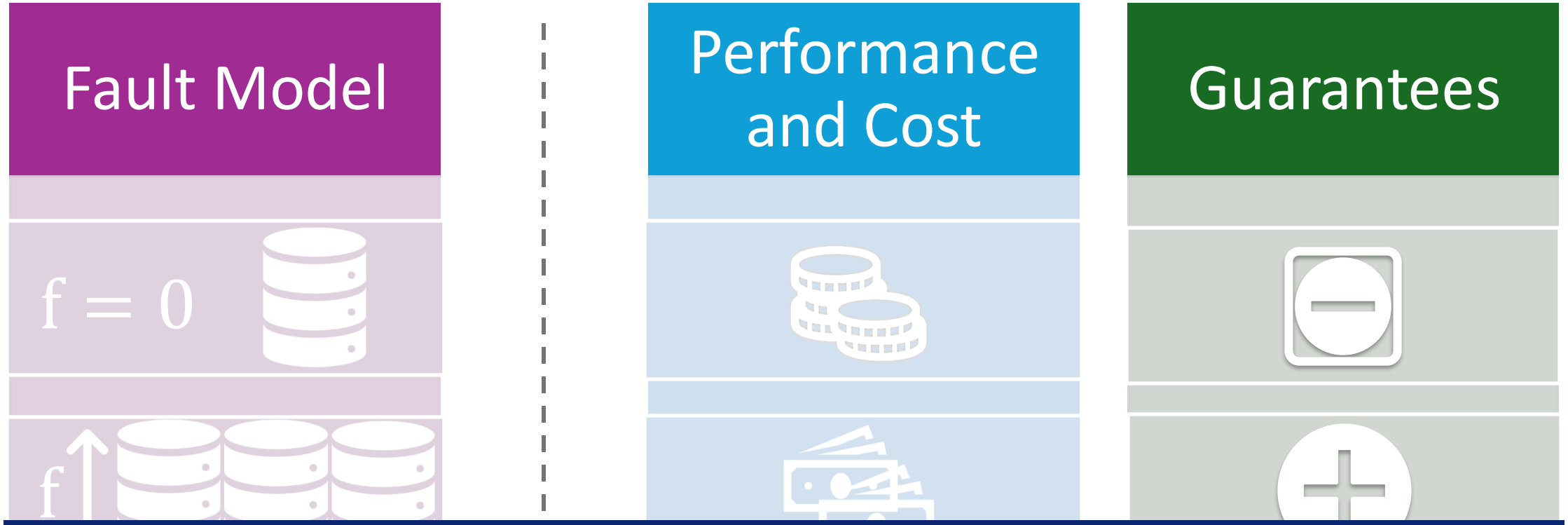
Guarantees

Consensus is an Insurance Policy





Consensus is an Insurance Policy



You buy protection for your world view!

Fault Models Must Match Reality

Optimistic

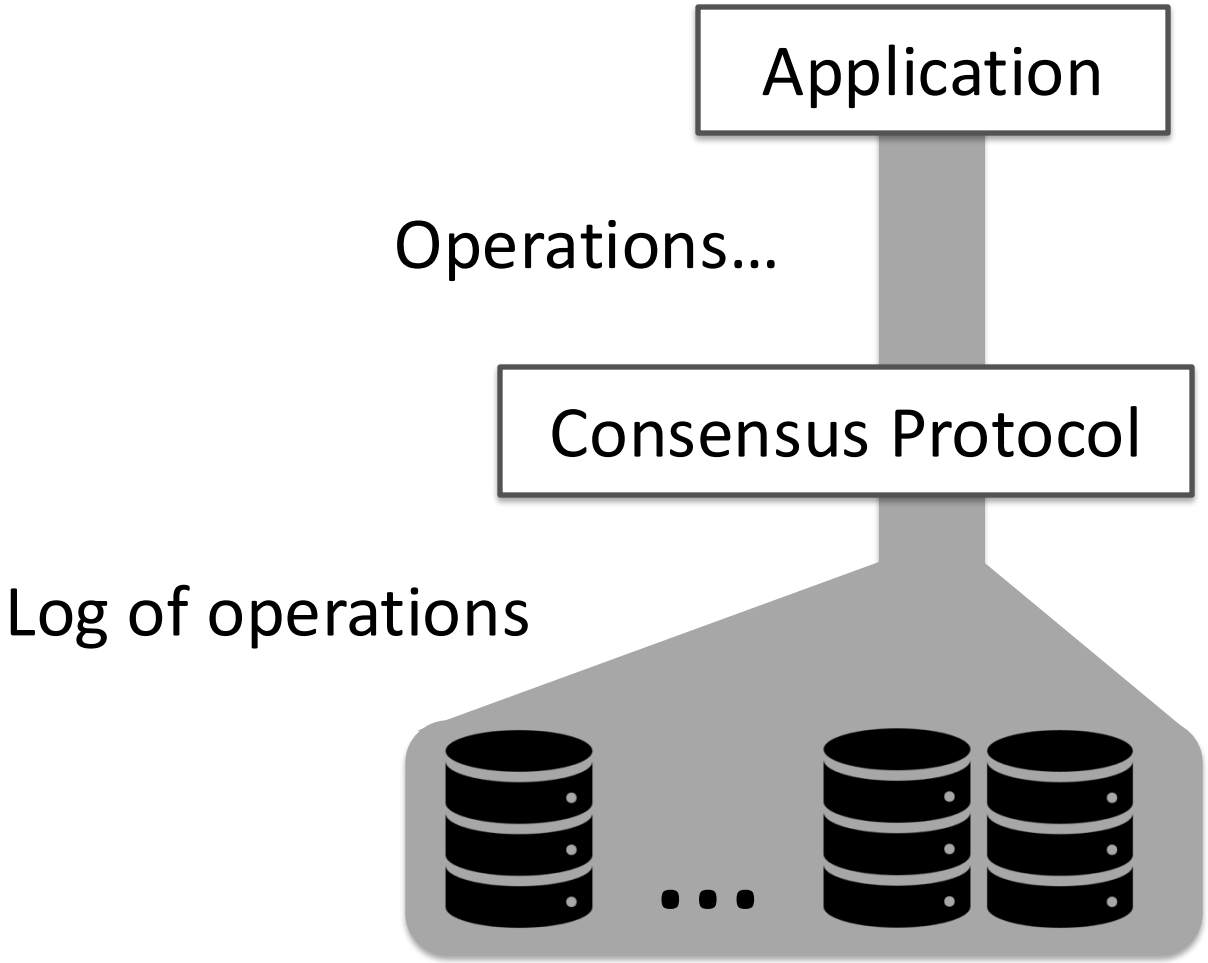
Reality

Pessimistic

Current fault models do not accurately capture reality

Protocols overpay for protection or do not fully protect against failures!

Consensus in the Real World

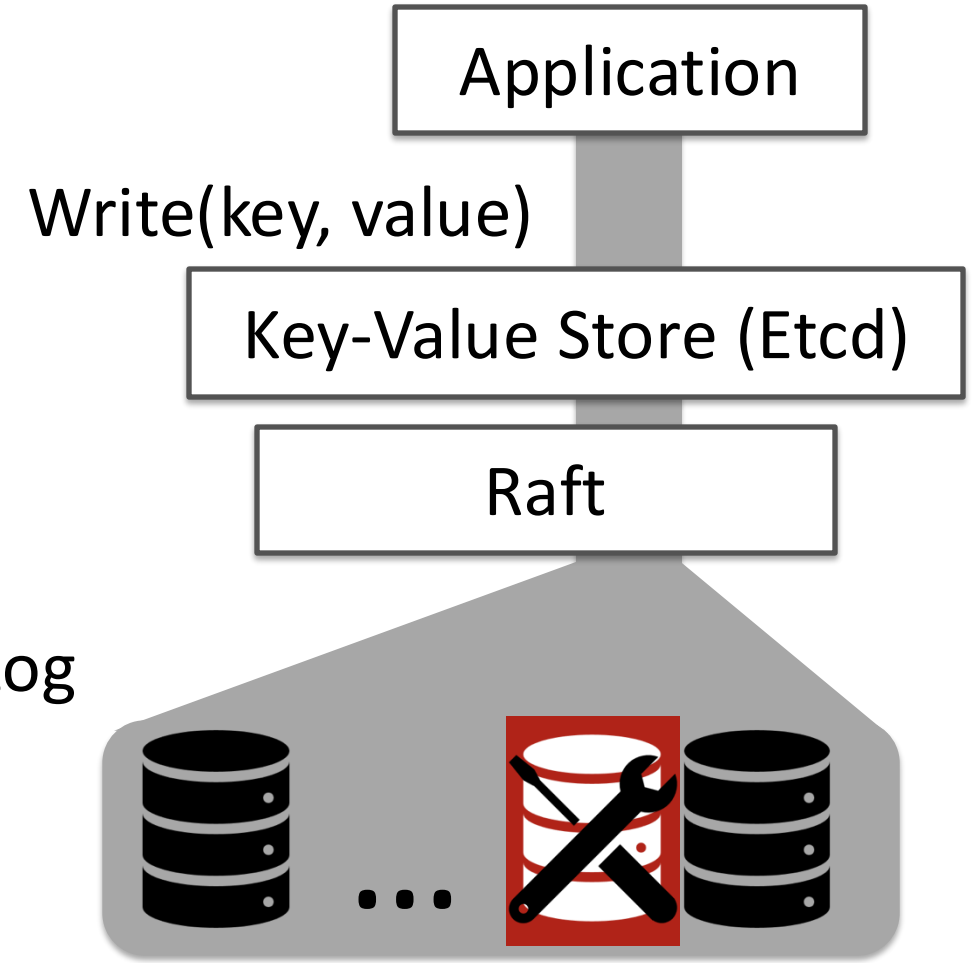


Guarantees

Safety: Operations commit in the same global order

Liveness: Every operation eventually commits

Consensus in the Real World



Fault Model

$f = 1$

Consensus in the Real World

f-Threshold Fault Model

Consensus Guarantees when #faults≤ f

Key Observations

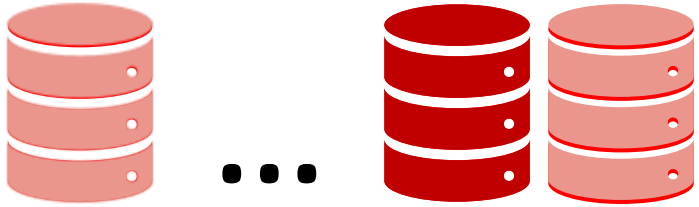
The f-threshold fault model fails to captures reality

In practice, machine failures are probabilistic

Protocols cannot provide better than probabilistic guarantees

f-Threshold Fault Model Fails to Capture Reality

Nodes are either correct or faulty	All nodes can fail
Faults are uniform	Some nodes are more likely to fail



(a) Fault rates are not uniform

Temperature	Failure Rate AFR	Increase
59°F	2.1%	1x
90.5°F	5.1 – 6.8%	3x

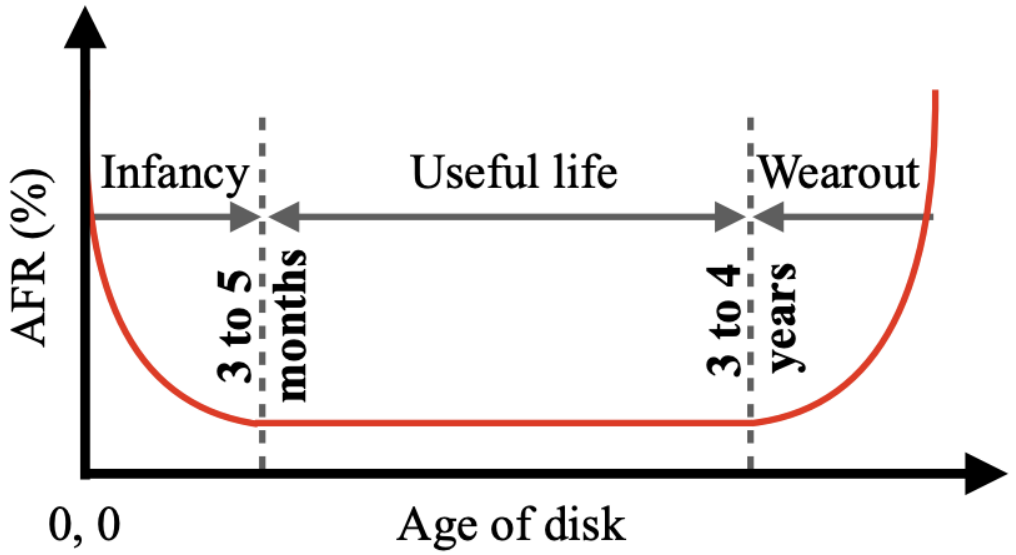
Manufacturer	Model	AFR Q4-2024	Increase
Toshiba	MG08ACA16TA	1.06%	1x
Seagate	ST14000NM0138	5.95%	6x
Seagate	ST14000NM000J	0.00%	0x

Timeline	SpotVM Evictions	Increase
4pm – 4am	<5%	1x
4am – 12pm	60%	30x

Algorand: Scaling Byzantine Agreements for Cryptocurrencies

[1] https://blocksandfiles.com/2024/05/02/disk-failure-rates-in-data-centres-are-falling-says-backblaze/
[2] https://www.backblaze.com/blog/backblaze-drive-stats-for-2024/
[3] Memory-Harvesting VMs in Cloud Platforms, ASPLOS'22
[4] Snape: Reliable and Low-Cost Computing with Mixture of Spot and On-Demand VMs, ASPLOS'23

(b) Fault rates evolve over time



VMs	Evictions 10 mins	Evictions 1 day
SpotVMs	10%	55.3%

[1] Cluster storage systems gotta have HeART: improving storage efficiency by exploiting disk-reliability heterogeneity
[2] https://www.backblaze.com/blog/drive-failure-over-time-the-bathtub-curve-is-leaking/
[3] Snape: Reliable and Low-Cost Computing with Mixture of Spot and On-Demand VMs, ASPLOS'23

(c) Faults may be correlated

https://blogs.microsoft.com/blog/2024/07/20/helping-our-customers-through-the-crowdstrike-outage/
https://www.techtarget.com/whatis/feature/Explaining-the-largest-IT-outage-in-history-and-whats-next

(d) Faults are complex

Cores that don't count

Peter H. Hochschild, Paul Turner, Jeffrey C. Mogul — Google, Sunnyvale, CA, US

Rama Govindaraju Parthasarathy Ranganathan — Google, Sunnyvale, CA, US

David E. Culler, Amin Vahdat — Google, Sunnyvale, CA, US

Silent Data Corruptions at Scale

Harish Dattatraya Dixit — Facebook, Inc. — hdd@fb.com

Sneha Pendharkar — Facebook, Inc. — spendharkar@fb.com

Matt Beadon — Facebook, Inc. — mbeadon@fb.com

Chris Mason — Facebook, Inc. — clm@fb.com

Tejasvi Chakravarthy — Facebook, Inc. — teju@fb.com

Bharath Muthiah — Facebook, Inc. — bharathm@fb.com

Sriram Sankar — Facebook Inc. — sriramsankar@fb.com

[1] Cluster storage systems gotta have HeART: improving storage efficiency by exploiting disk-reliability heterogeneity

[2] https://www.backblaze.com/blog/drive-failure-over-time-the-bathtub-curve-is-leaking/

f-Threshold Fault Model Fails to Capture Reality

(a) Fault rates are not uniform

(b) Fault rates evolve over time

(c) Faults may be correlated

(d) Faults are complex

In practice, machine faults are probabilistic

Consensus Guarantees are Mismatched

All-or-nothing guarantees

Safe and Live
when failures $\leq$ f

Undefined guarantees
if failures $>$ f

Consensus Guarantees are Mismatched

All-or-nothing guarantees	100% - ε guarantees
Safe and Live when failures $\leq f$ Undefined guarantees if failures $> f$	Safe: Every node has a non-zero failure probability and eventually all nodes will fail

No protocol can do better than probabilistic guarantees

Consensus Guarantees are Mismatched

All-or-nothing guarantees	100% - ε guarantees
Safe and Live when failures $\leq f$ Undefined guarantees if failures $> f$	Raft is only 99.97% [safe&live] in three-node deployments if nodes suffer a 1% fault rate!

Consensus Guarantees are Mismatched

AWS	Standard S3	S3 Intelligent Tiering	S3 Express One Zone
Designed for availability	99.99%	99.9%	99.95%
Availability SLA	99.9%	99%	99.9%

S3 is designed to exceed 11 nines of data durability

Google Spanner is designed to support 5 nines of availability

[1] https://aws.amazon.com/s3/storage-classes/
[2] https://www.frictionlesspost.com/p/google-spanner-serves-trillions-of-rows-with-99-999-reliability

Our Vision: Probabilistic Consensus for the Real World!

Probabilistic fault models
that accurately capture reality

Simple Abstractions: Key Challenge

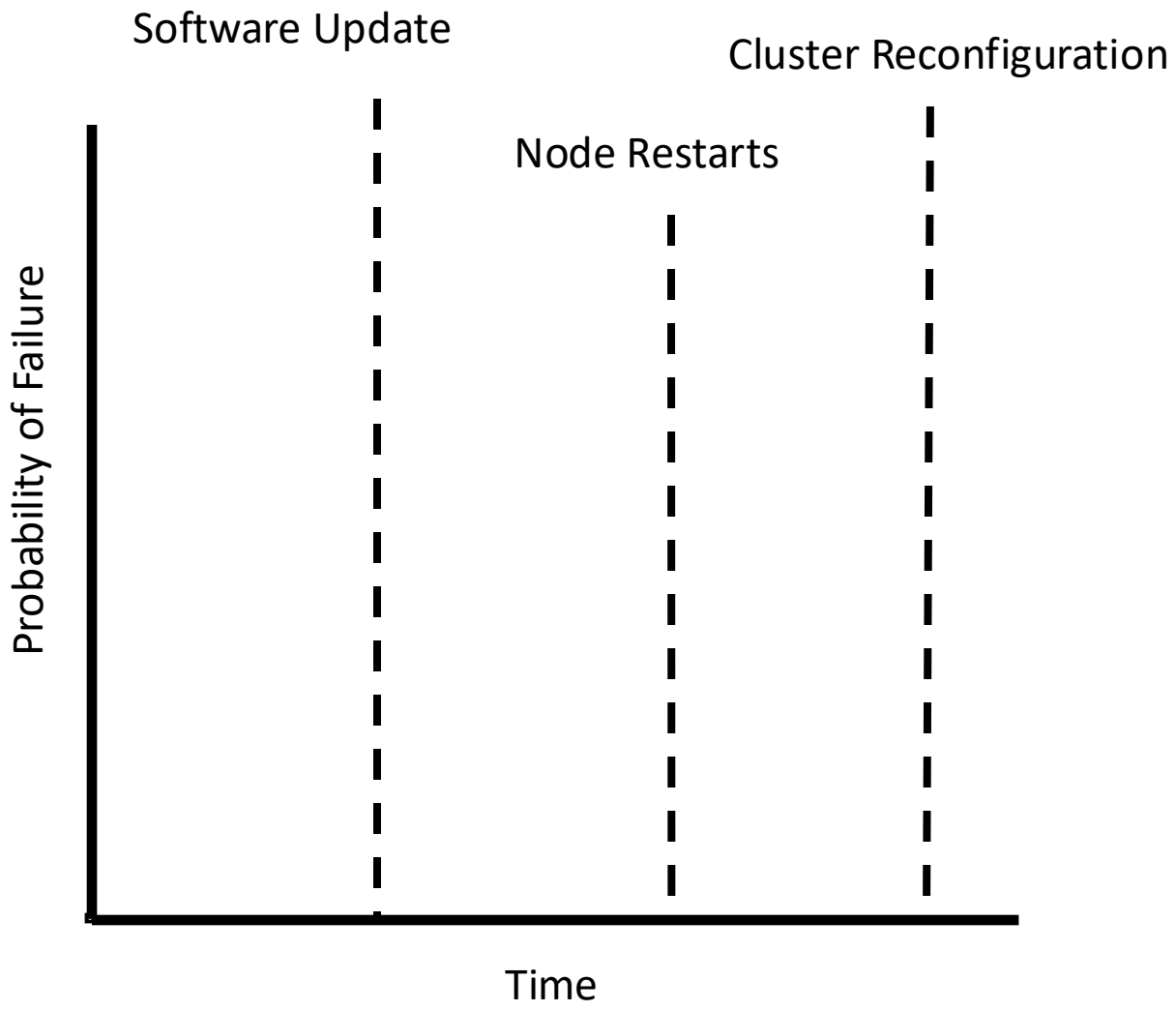
Traditional Consensus

Single parameter f to decide number of replicas

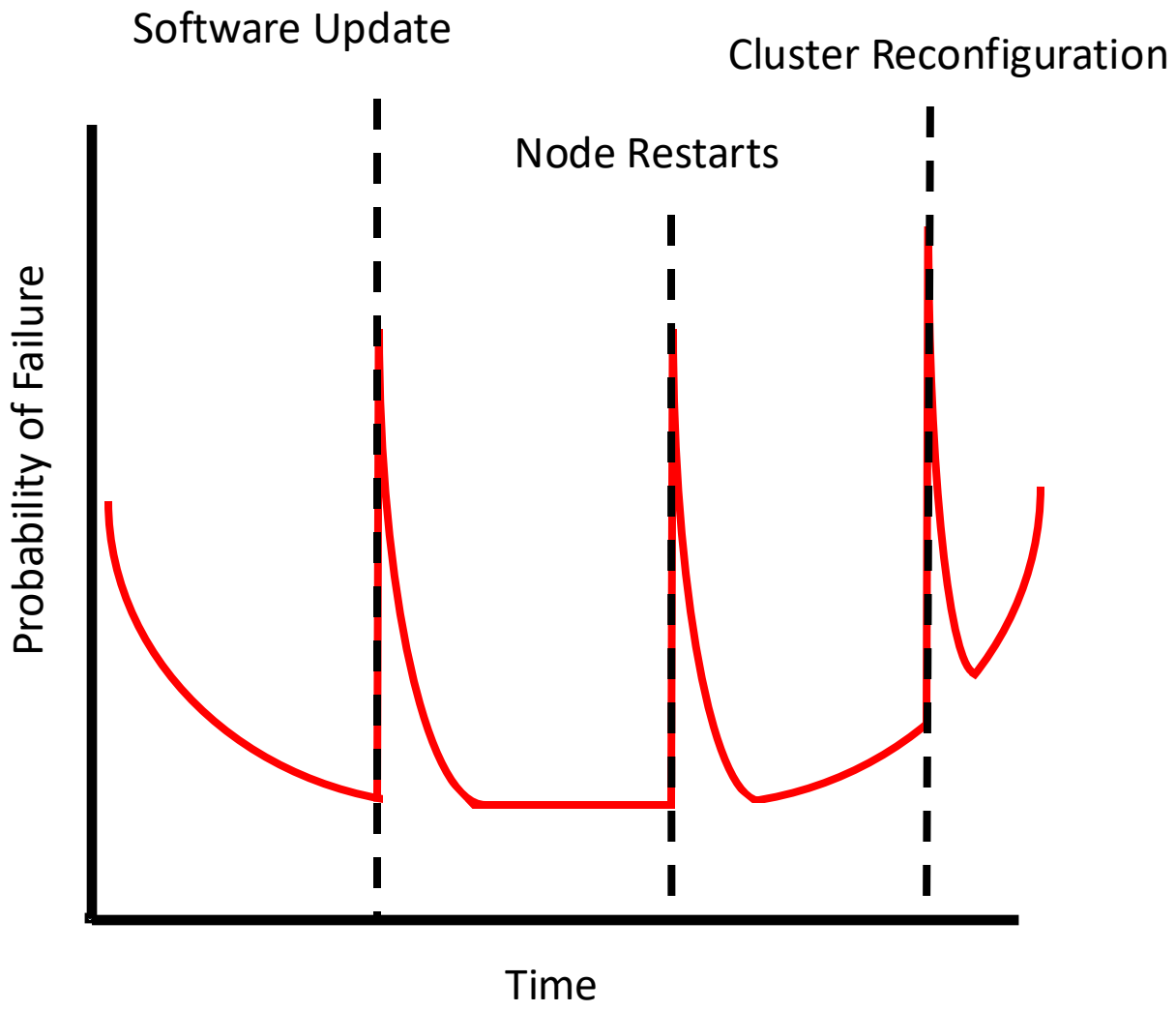
100% safety + liveness (*)

* f nodes fail, and others are correct

Accurate Per-Node Fault Curves



Accurate Per-Node Fault Curves



Potential Opportunities

Protocols can better utilize reliable nodes

Larger clusters of less reliable nodes can help

Explore constant instead of linear size quorums

Exploit the tradeoff between safety and liveness

Potential Opportunities

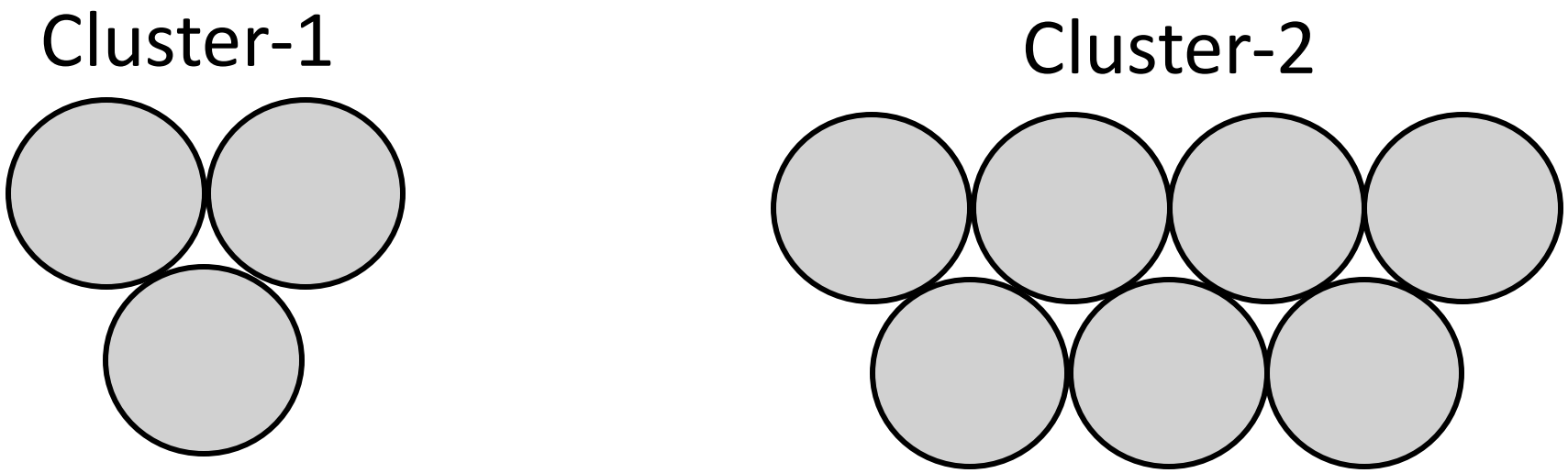
Protocols can better utilize reliable nodes

Larger clusters of less reliable nodes can help

Explore constant instead of linear size quorums

Exploit the tradeoff between safety and liveness

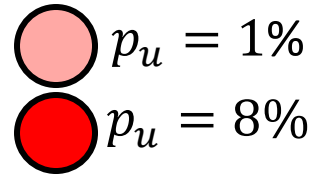
1. Protocols can better utilize reliable nodes



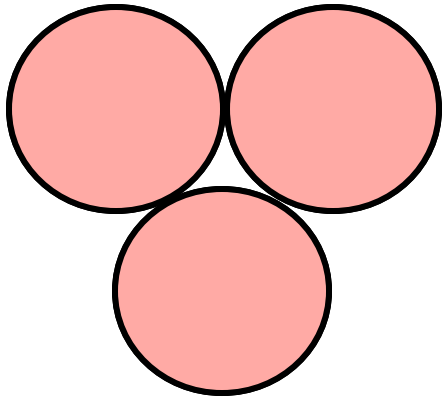
Both clusters guarantee safety and liveness

Cluster-2 tolerates $f = 3$ faults

1. Protocols can better utilize reliable nodes

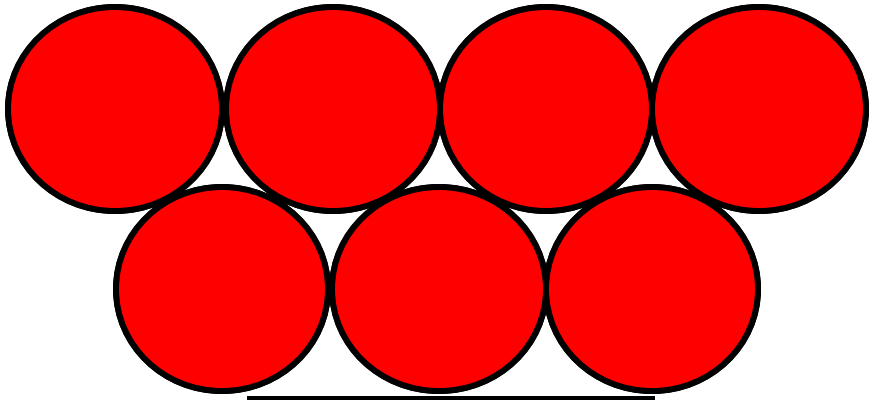


Cluster-1



Three 9s

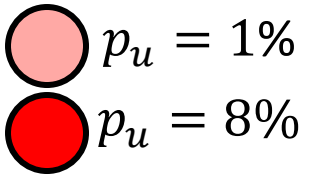
Cluster-2



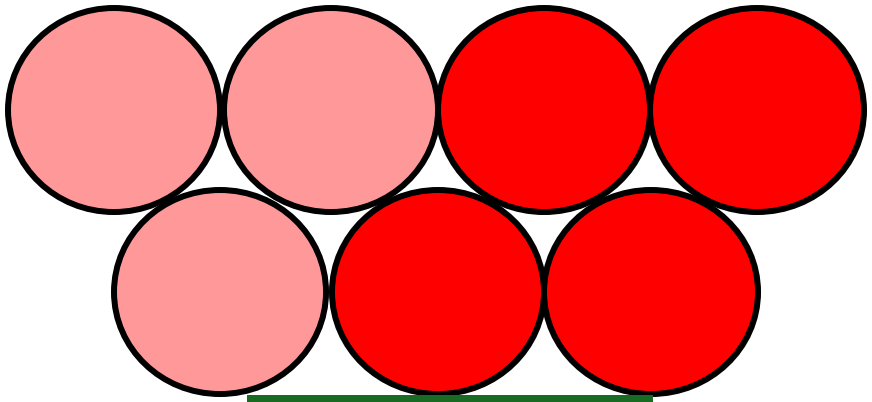
Two 9s

With fault rates, they achieve 9s of safety and liveness

1. Protocols can better utilize reliable nodes



Cluster-3



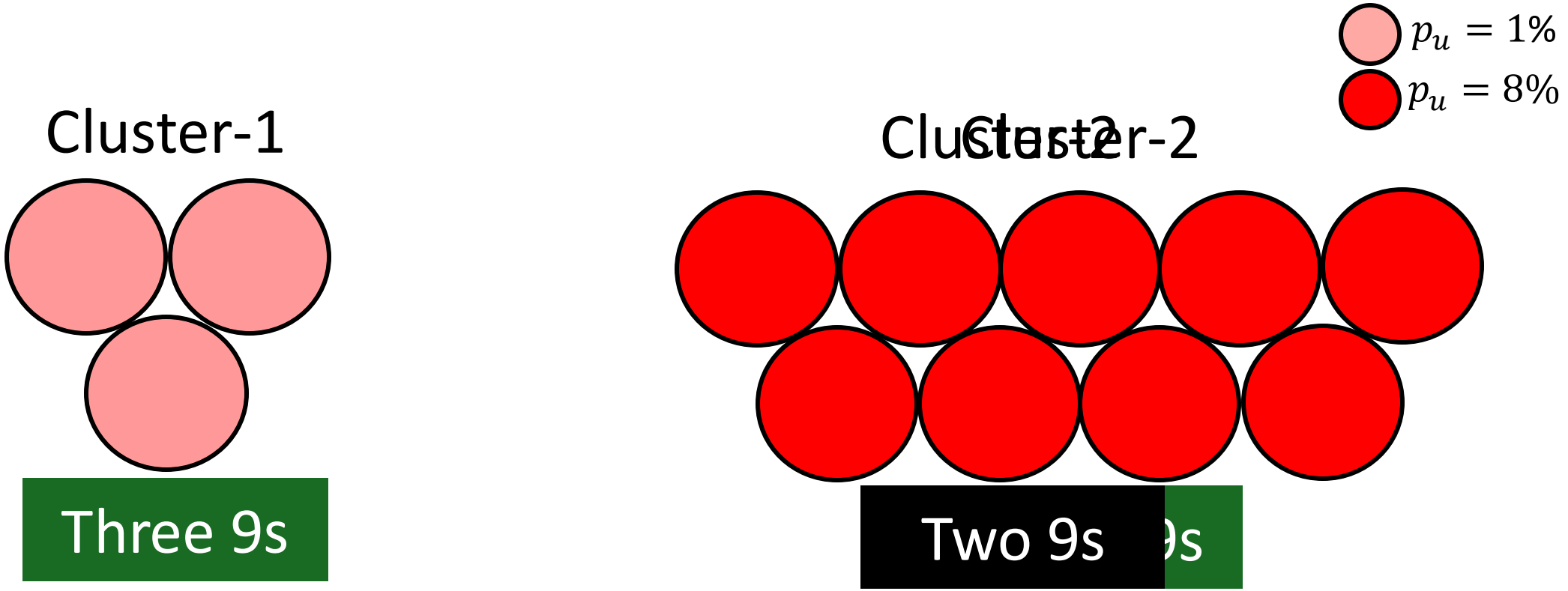
Three 9s

1. Leader selection from more reliable nodes

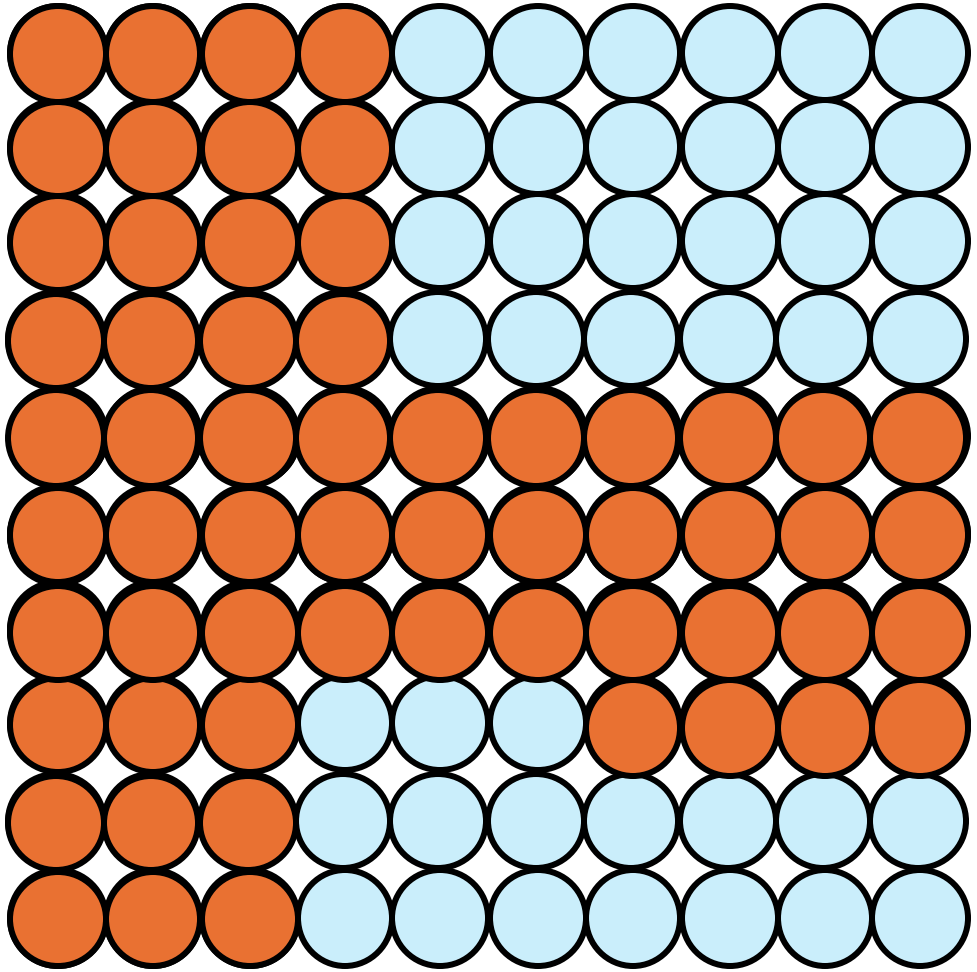
2. Skewing quorums to include reliable nodes

Protocols can leverage accurate fault rates for better guarantees

2. Larger clusters of less reliable nodes can help

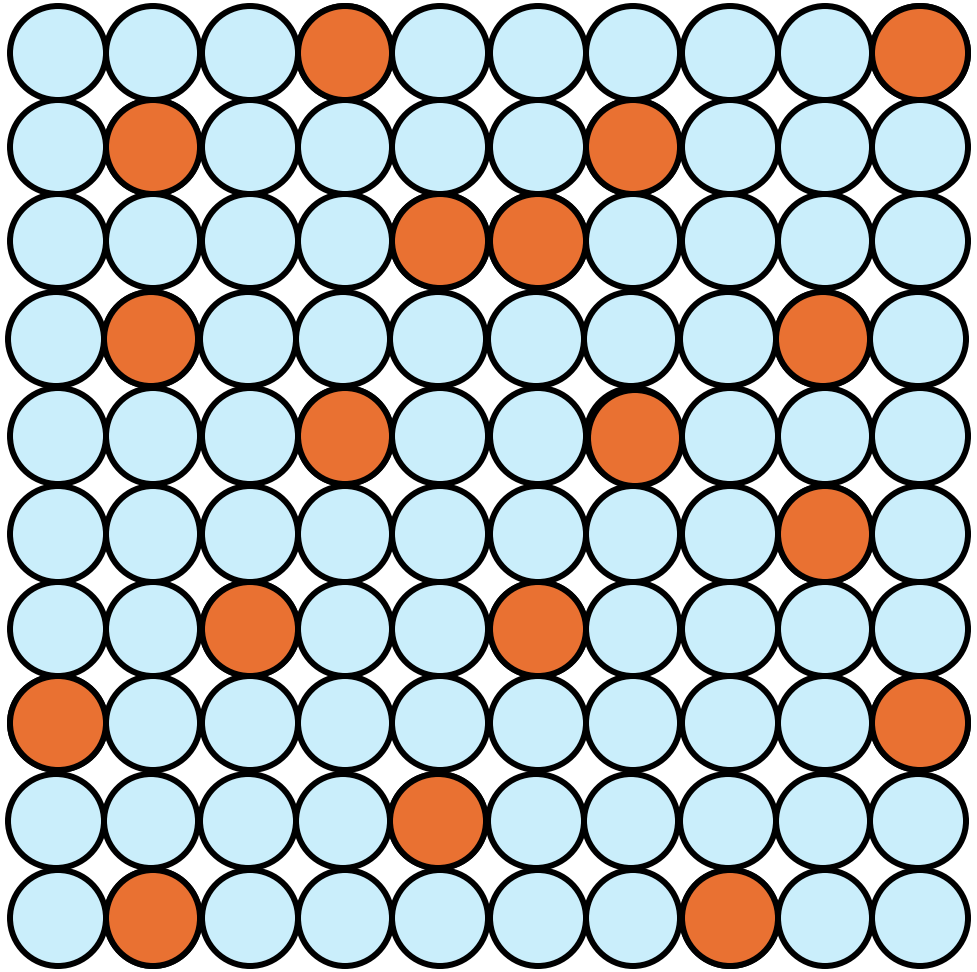


3. Exploring constant size quorums



$$N = 100$$
$$f = 33$$
$$Q = 34$$

3. Exploring constant size quorums



$$N = 100$$
$$f = 33$$

$$Q = 34$$

$$P[\checkmark] = 97.9\ldots\%$$

Probabilistic Consensus for Real World!

Current fault models fail to accurately capture reality

Today, consensus is probabilistic – like it or not!

Accurate fault curves for better fault modeling

Probability-native consensus protocols

- More efficient, cost-effective, and sustainable, and reliable

soujanya@berkeley.edu

reginaldfrank77@berkeley.edu