Characterizing Private Clouds: A Large-Scale Empirical Analysis of Enterprise Clusters

Ignacio Cano, Srinivas Aiyar, Arvind Krishnamurthy

University of Washington – Nutanix Inc.

ACM Symposium on Cloud Computing

October 2016

Private Clouds

Private Clouds

- Cloud computing that **delivers service** to a **single organization**, as opposed to public clouds, which service many.
- Direct **control** of **infrastructure** and **data**.
- Carry **management** and **maintenance costs**.

Motivation

- **Increasing trend** in the use of **private clouds** within companies.
- Private clouds **deployments** require **careful consideration** of what will happen in the future:
  - Capacity
  - Failures
  - …

Motivation

- Research Questions:
  - What are the **most common failures**?
  - Wh…
  - Ho… **usage**?
  - How do additional **replicas** impact **data durability**?
  - What **causes** companies to **expand** their **clusters**?

Need Measurement Data!

Related Work

Setting \ Study	Hardware Failures	Storage	Compute
Desktops	• **HW Fai** [Nightin	• **Metadata in Windows PCs**	• **Disk/CPU Usage and Load** [Bolosky et al., SIGMETRICS'00]
Public Clouds	• **HW rel** [Vishwanath et al., SoCC'11]	**Access Patterns** [Liu et al., IEEE/ACM CCGrid'13]	• **Workloads characterization** [Mishra et al., SIGMETRICS'10] • **Scheduling on Heterogeneous Clusters** [Reiss et al., SoCC'12]

Limited prior work on Private Clouds!

In this talk

- Large-Scale Measurement Study of Private Clouds
  - **Lower hardware failure rates**
  - **Nodes overprovisioned**
  - **Stable storage and CPU usage**
- Modeling based on the Measurements
  - **Each extra replica provides substantial durability improvements**
  - **Storage needs drive growth more than compute**

Outline

- Large-Scale Measurement Study of Private Clouds
  - Context
  - Cluster Profiles
  - Failure Analysis
  - Workload Characteristics
- Modeling based on the Measurements
  - Durability
  - Cluster Growth

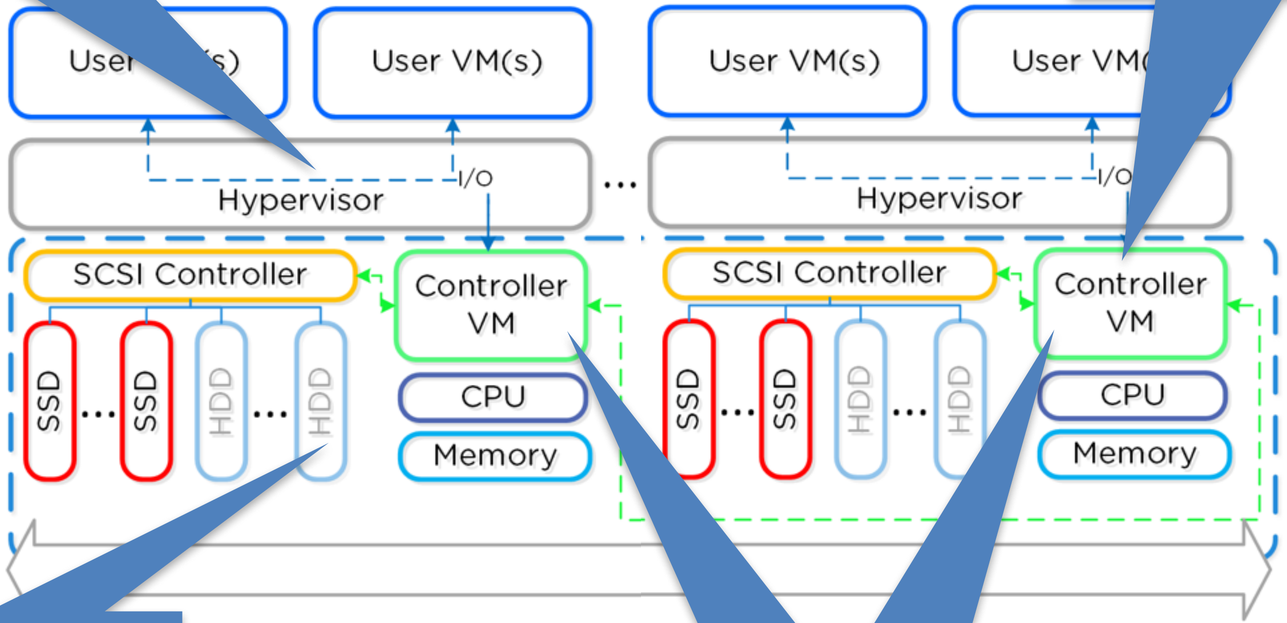
Outline

- **Large-Scale Measurement Study of Private Clouds**
  - **Context**
  - Cluster Profiles
  - Failure Analysis
  - Workload Characteristics
- Modeling based on the Measurements
  - Durability
  - Cluster Growth

Nutanix Clusters

Operations interposed at the hypervisor level and redirected to CVMs

Random replication
VMs migration
…



Integrated Compute-Storage

Global view of cluster state

Outline

- **Large-Scale Measurement Study of Private Clouds**
  - Context
  - **Cluster Profiles**
  - Failure Analysis
  - Workload Characteristics
- Modeling based on the Measurements
  - Durability
  - Cluster Growth

Clusters

Summary Statistics	Value
# of Clusters	2168

Clusters

Summary Statistics	Value
# of Clusters	2168
# of Nodes	13394

6.18 Nodes/Cluster

Clusters

Summary Statistics	Value
# of Clusters	2168
# of Nodes	13394
Cluster Sizes	3 - 40

Clusters

Summary Statistics	Value
# of Clusters	2168
# of Nodes	13394
Cluster Sizes	3 - 40
# of Disks	~ 70K

Node Configurations

Configuration	Storage		Compute		Memory (GB)
	SSD (TB)	HDD (TB)	Cores	Clock Rate (GHz)	
Config-1	1.6	8	24	2.5	384
Config-2	0.8	4	12	2.4	128
Config-3	0.8	30	16	2.4	256

Storage-heavy

Compute-heavy

Mostly homogeneous within a cluster

Workloads

Workload	Example Applications	Configuration
Virtual Desktop Infrastructure	Citrix XenDesktop VMware Horizon/View	Config-1

Workloads

Workload	Example Applications	Configuration
Virtual Desktop Infrastructure	Citrix XenDesktop VMware Horizon/View	Config-1
Server	SQL Server Exchange Mail Server	Config-2 Config-3

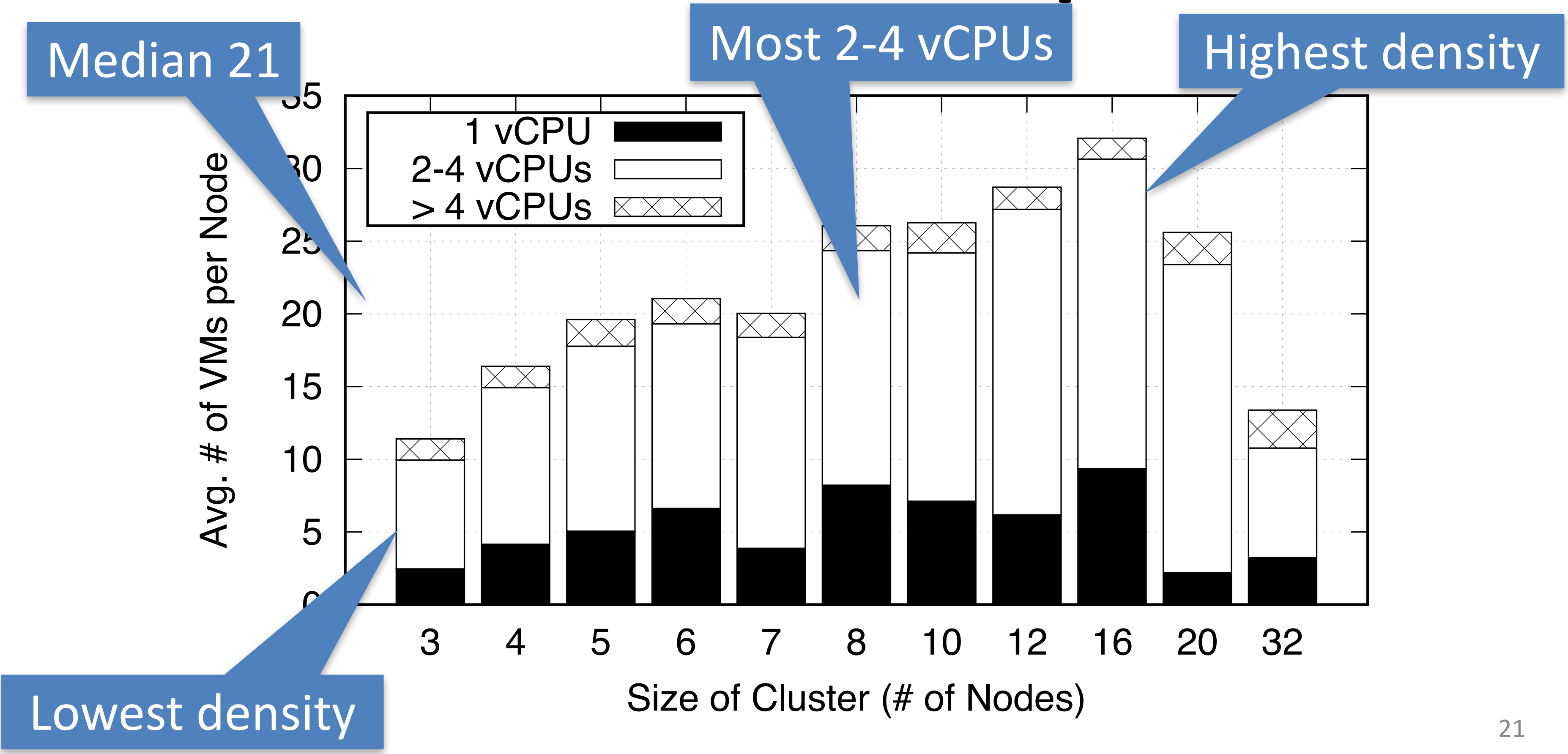
Workloads

Workload	Example Applications	Configuration
Virtual Desktop Infrastructure	Citrix XenDesktop VMware Horizon/View	Config-1
Server	SQL Server Exchange Mail Server	Config-2 Config-3
Big Data	Splunk Hadoop	Config-3

Workloads

Workload	Example Applications	Configuration
Virtual Desktop Infrastructure	Citrix XenDesktop VMware Horizon/View	Config-1
Server	SQL Server Exchange Mail Server	Config-2 Config-3
Big Data	Splunk Hadoop	Config-3
Others	IT Infrastructure Custom applications	Mix

Distribution of VMs per Node

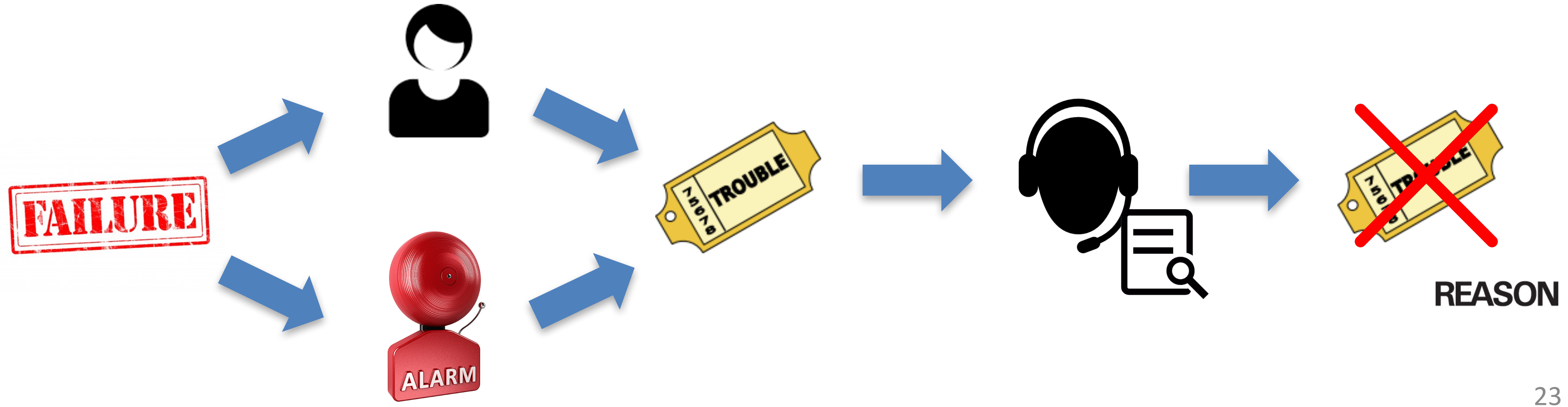


Outline

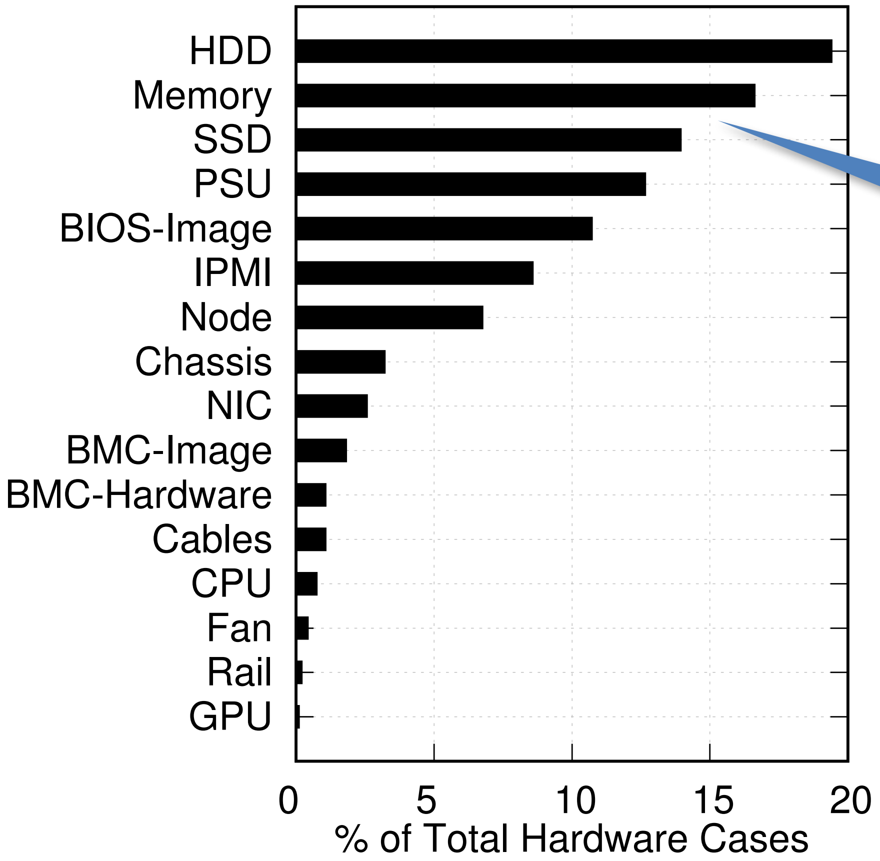
- **Large-Scale Measurement Study of Private Clouds**
  - Context
  - Cluster Profiles
  - **Failure Analysis**
  - Workload Characteristics
- Modeling based on the Measurements
  - Durability
  - Cluster Growth

Failures

- We only **consider failures** that require **manual intervention, i.e., human** operators **annotate** the cause of the problem.



Hardware Failures



Top 3 account for around 50% of HW failures

Annual Return Rate

Component	ARR (%)
HDD	0.76

2-9 % prior studies

Annual Return Rate

Component	ARR (%)
HDD	0.76
SSD	0.72

Lower return rates

Enterprise-grade commodity HW

% prior
s (4 years)

Outline

- **Large-Scale Measurement Study of Private Clouds**
  - Context
  - Cluster Profiles
  - Failure Analysis
  - **Workload Characteristics**
- Modeling based on the Measurements
  - Durability
  - Cluster Growth

Workload Characteristics

- **Usage over time** seems to be **stable/predictable:** 80% of the clusters use
  - **Storage:** mean <= 50%, std <= 8%
  - **CPU:** mean <= 20%, std <= 5%

- **SSDs** can generally **maintain** the **working set**
  - 80% of nodes use <= 500 GB for the working set

Outline

- Large-Scale Measurement Study of Private Clouds
  - Context
  - Cluster Profiles
  - Failure Analysis
  - Workload Characteristics
- **Modeling based on the Measurements**
  - **Durability**
  - Cluster Growth

Durability Model

- Estimate the **probability of data loss.**
- Assumptions:
  - **replication factor** of **2**
  - **random replication** (replicate to a random node)
- The **time** required **to create a new replica** when a node goes down:

$$\Delta t = \frac{d}{(n-1)v}$$

Data to be replicated

Remaining live nodes

Data transfer rate

Durability Model

- $p(\Delta t)$ = **probability of node failure** in Δt time.
- We **decompose** the overall period over which we want to provide the durability guarantee into a **sequence of intervals**, each of length Δt.
- *Q* **= data loss event** where two failures occur within Δt time, i.e. data could not be replicated.

Durability Model

- Then the **probability** that there is **no data loss** in an **interval** Δt:

$$P(\neg Q, \Delta t) \leq \underbrace{(1 - p(\Delta t))^n}_{\text{No failures}} + \underbrace{np(\Delta t)(1 - p(\Delta t))^{n-1}}_{\text{Exactly one node fails}} \underbrace{(1 - p(\Delta t))^{n-1}}_{\text{The remaining n-1 nodes do not fail within }\Delta t\text{ time}}$$

No failures

Exactly one node fails

The remaining n-1 nodes do not fail within Δt time

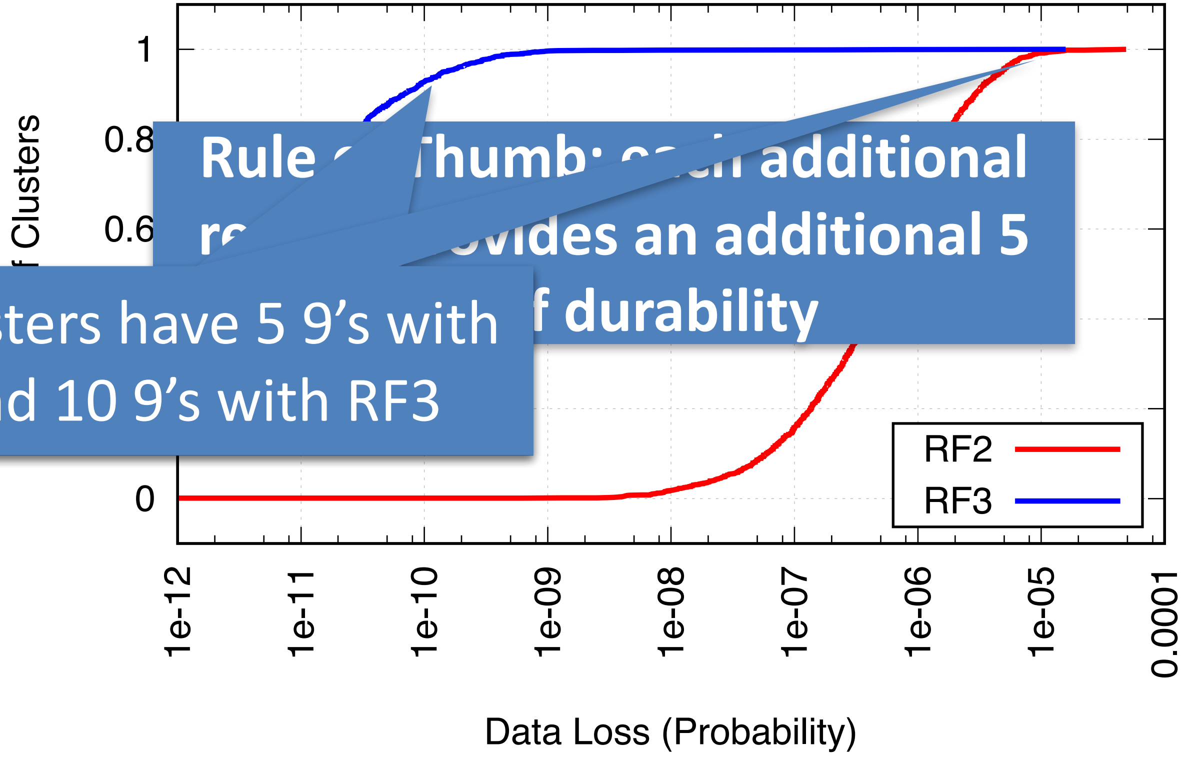
Durability Model

- On a yearly-basis, we consider all Δt intervals in a year.
- Probability of **no data loss** within a **year** is:

$$P_{durability} = P(\neg Q, \Delta t)^{N(\Delta t)}$$

# of intervals of Δt time in a year

Durability in Private Clouds



Rule o… Thumb: e… additional re… …ovides an additional 5 …f durability

Most clusters have 5 9's with RF2, and 10 9's with RF3

Outline

- Large-Scale Measurement Study of Private Clouds
  - Context
  - Cluster Profiles
  - Failure Analysis
  - Workload Characteristics
- **Modeling based on the Measurements**
  - Durability
  - **Cluster Growth**

Cluster Growth Analysis

- **Customers** periodically **add nodes** to their existing clusters.
- What **drives** such **growth**?
- We resort to **machine learning**
  - **Binary classification** problem
  - **Logistic Regression** with L1 regularization

Cluster Growth Analysis

- Use **200 clusters** than grew at least once in a period of 8 months.
- **15K examples** (70% train, 10% val, 20% test).
- Train with **different combination of features** to understand which are important.

Features

Cluster Features F^c	Description
n(nodes)	discretized # of nodes
n(vms)	# of vms per node

Storage Features F^s	Description
r(ssd)	ssd usage to ssd capacity ratio
r(hdd)	hdd usage to hdd capacity ratio
r(store)	storage usage to total capacity ratio

Performance Features F^p	Description
n(vcpus)	# of virtual cpus
n(iops)	# of iops per node

What drives cluster growth?

1. Cluster Size
2. Storage Needs
3. Compute Needs

Upgrades from 3-4 node clusters

HDD usage

Number of VMs

Storage more than compute!

Outline

- Large-Scale Measurement Study of Private Clouds
  - Context
  - Cluster Profiles
  - Failure Analysis
  - Workload Characteristics
- Modeling based on the Measurements
  - Durability
  - Cluster Growth

Conclusions

- Large-Scale Measurement Study of Private Clouds
  - **Lower hardware failure rates**
  - **Nodes overprovisioned**
  - **Stable storage and CPU usage**
- Modeling based on the Measurements
  - **Each extra replica provides substantial durability improvements**
  - **Storage needs drive growth more than compute**

Thanks!