Object, Block, or File Storage??

Choosing the Right Storage to Integrate Into K8s

Mitch Becker
Sr. WW GTM Specialist Storage

Tom McDonald
Sr. WW Specialist SA Storage

Agenda

01 Who we are

02 Market Adoptions

03 Workloads

04 CSI

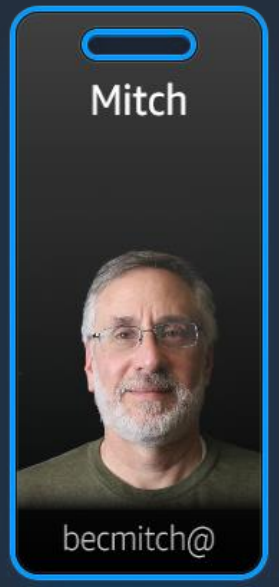
05 Storage Types

06 Best Practices and Blogs

07 CYOA

08 Conclusion

Meet the team – your AWS presenters



Mitch Becker

(he/him)
Sr. WW GTM Specialist Storage
Amazon Web Services

Tom McDonald

(he/him)
Sr. WW Specialist SA Storage
Amazon Web Services

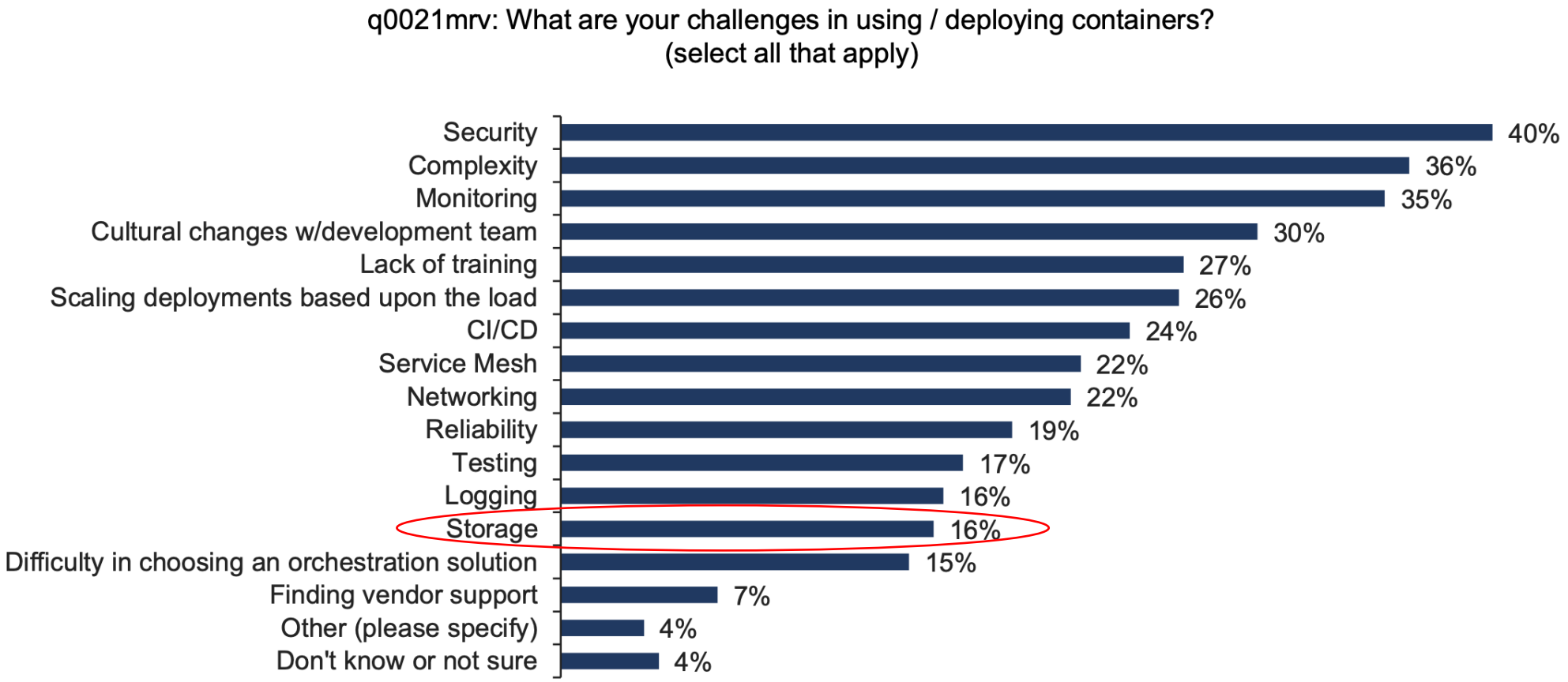
Why does storage matter ?

Increase in persistent workloads (AI/ML, Spark, Kafka, Flink, etc.)
Organizations migrating from on-premises trying to understand options

Storage can be a significant portion of a deployment cost

Modern applications need durable shared scalable storage

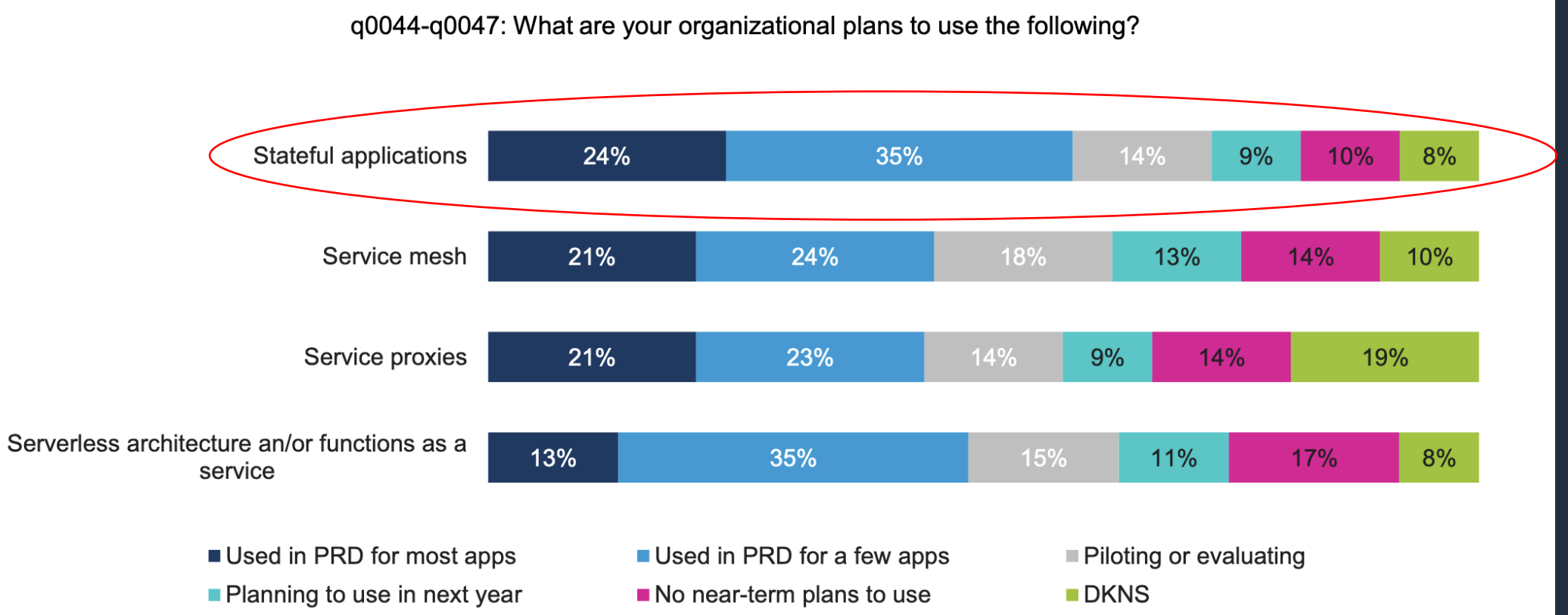
What is the market seeing?



2023 CNCF Annual Survey, q21, sample size = 477, valid cases = 477, total mentions = 1,720, organizations whose primary revenue is not from offering cloud native technologies or services

*Source: Cloud Native Computing Foundation (CNCF) Customer Survey 2023

Workload adoption trends



2023 CNCF Annual Survey, q44-q47, sample size = 988

*Source: Cloud Native Computing Foundation (CNCF) Customer Survey 2023

Workload examples

Databases

Strict resilience & security

Frequent data refreshes

MSSQL, PostgreSQL

Developer tools

Keeping data while containers are scaled or moved around

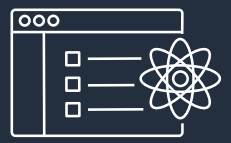
JIRA, Artifactory, Git

Content management

Single source of data for all content apps

WordPress, Drupal

Analytics



Shared notebooks for data scientists

Jupyter, Jupyterhub

HPC

High Performance workloads for research and enterprise

BLAST, OpenFoam

AI/ML

Shared training set for multiple services

MXNet, TensorFlow

Extensibility

Three simple storage **tenants**

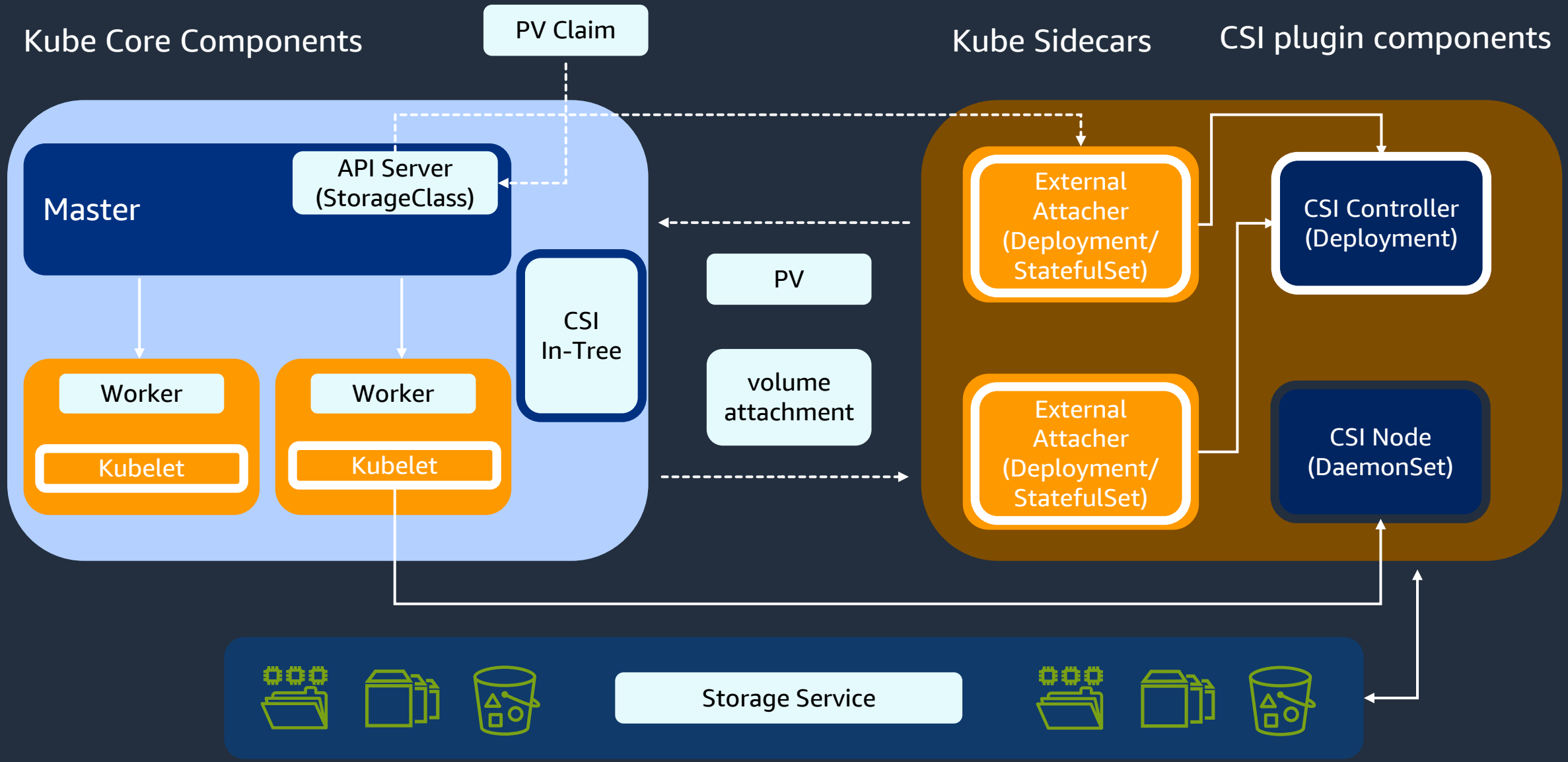
1. Storage **must not** depend on the pod lifecycle.

2. Storage is **available from all pods and nodes** in the Kubernetes cluster.

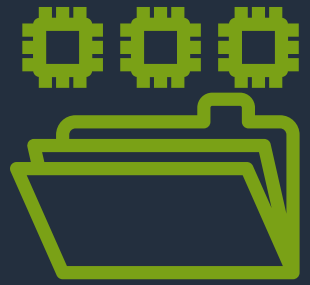
3. Storage must be **highly available** regardless of crashes or application failures.

How is this done?

Container Storage Interface (CSI)



Storage Types

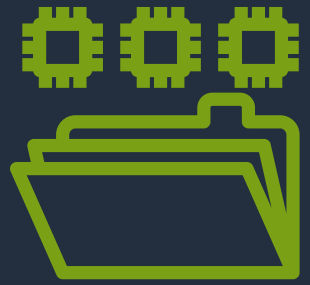


File Storage is a data storage architecture that manages data as files in a traditional hierarchical directory structure.

Block Storage is a data storage architecture that manages data as either physical or virtual storage devices, as a sequence of fixed-size blocks rather than a hierarchical file system.

Object Storage is a data storage architecture that manages data as objects, rather than as files in a traditional hierarchical directory structure or as blocks of data on a storage device.

CSI Drivers by Storage Type



File Storage CSI Drivers
give the ability to access via a protocol level (ex: NFS, SMB)

Typically used in 1:M environments.

Or at a product or service level such as the Amazon FSx for Lustre CSI Driver.

Block Storage CSI Drivers
designed to prioritize low latency and high throughput performance.

iSCSI, NVMe/TCP, Fibre Channel, Local...

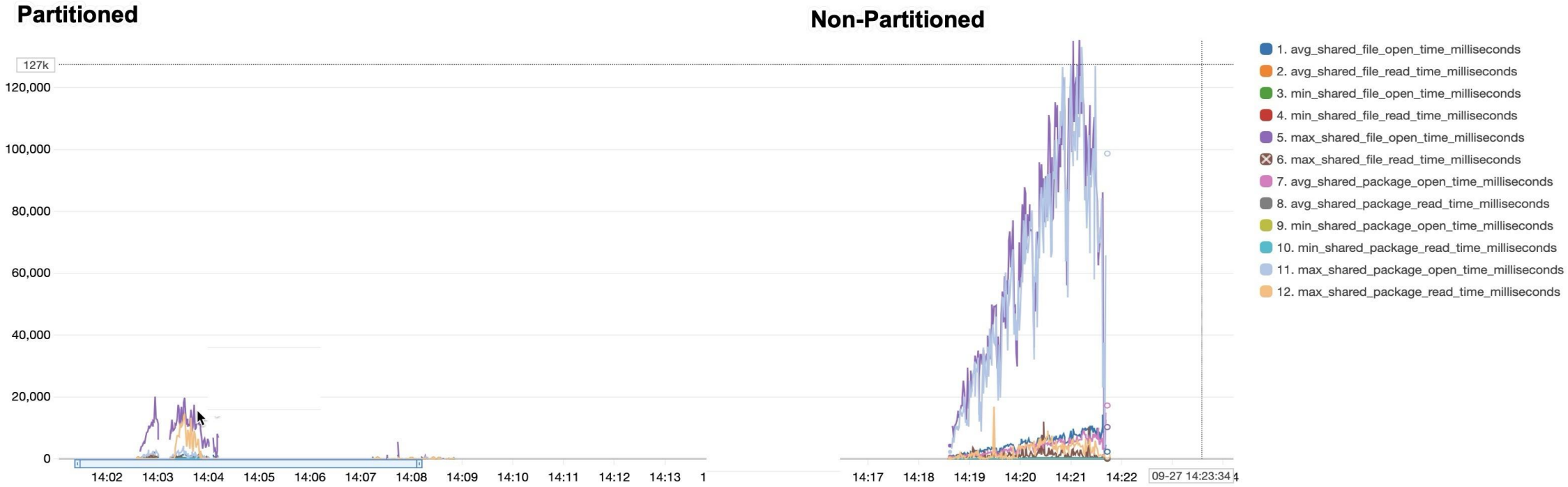
Also by Storage product/service, like Amazon Elastic Block Storage (EBS) Driver.

Object Storage CSI Drivers
FUSE based drivers present objects as a file system.

Understanding the functionality crucial for successful use.

Most Object driven applications use APIs to access and use objects either in memory or on local filesystems.

Follow Best Practices



#	Metric	Standard	Pre-Partitioned	Improvements
1	avg_shared_file_open_time_milliseconds	14,000	1,500	933%
5	max_shared_file_open_time_milliseconds	130,000	20,000	650%
6	max_shared_file_read_time_milliseconds	11,800	2,400	492%
7	avg_shared_package_open_time_milliseconds	17,000	300	5667%
11	max_shared_package_open_time_milliseconds	13,000	4000	325%

And follow blogs for similar Workloads

Some file systems require init containers to modify host side settings.

```
apiVersion: v1
kind: Pod
metadata:
  name: readahead-cmd-shell
spec:
  initContainers:
  - name: set-lustre-params
    image: amazonlinux:2
    command: ["/bin/sh","-c"]
    args: ["amazon-linux-extras install lustre -y && /sbin/lctl set_param llite.*.max_read_ahead_mb=1024"]
    securityContext:
      privileged: true
  containers:
  - name: app
    image: $ECR_REPO_URI:latest
    command: ["/bin/sh", "-c"]
    args: ["while true; do sleep 30; done"]
    volumeMounts:
    - name: persistent-storage
      mountPath: /fsx/lustre
    - name: dshm
      mountPath: /dev/shm
  volumes:
  - name: persistent-storage
    persistentVolumeClaim:
      claimName: fsx-claim
  - name: dshm
    emptyDir:
      medium: Memory
  restartPolicy: Never
```

For this mmap() workload, adjusting the read-ahead parameter on FSx Lustre saw a 27% increase in workload performance without any additional costs.

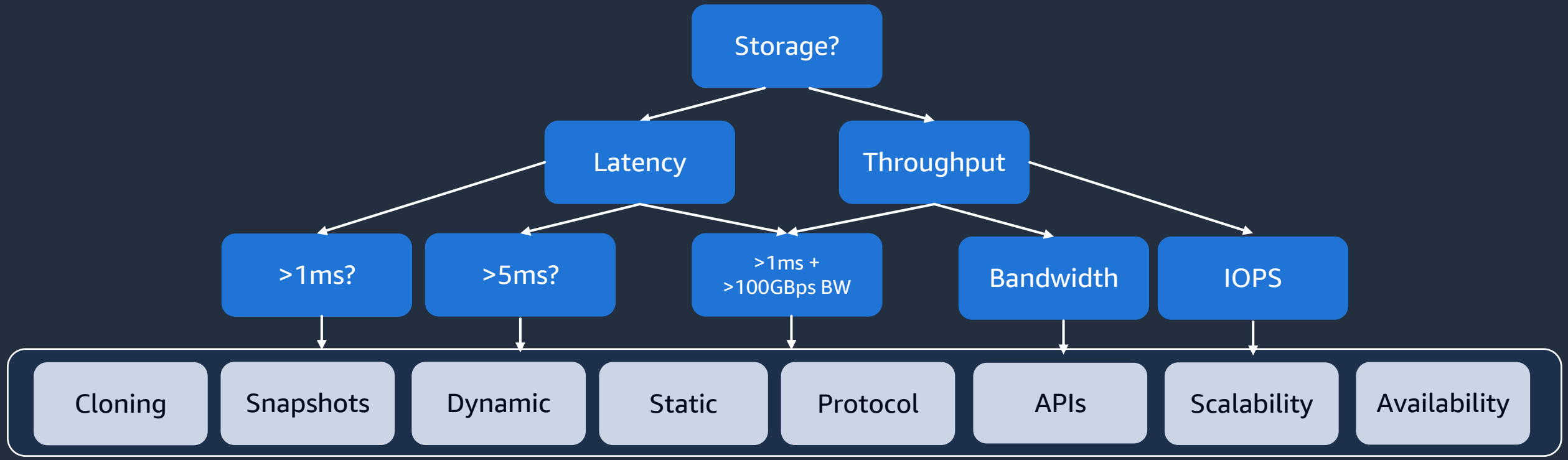
Job client settings	Time in Seconds
T_1 = Job with max_read_ahead_mb defaults changed to 1024 MiB	877
T_2 = Job with max_read_ahead_mb	1201

Formula	$1-(T_1-T_2)$
Increase in performance	27%

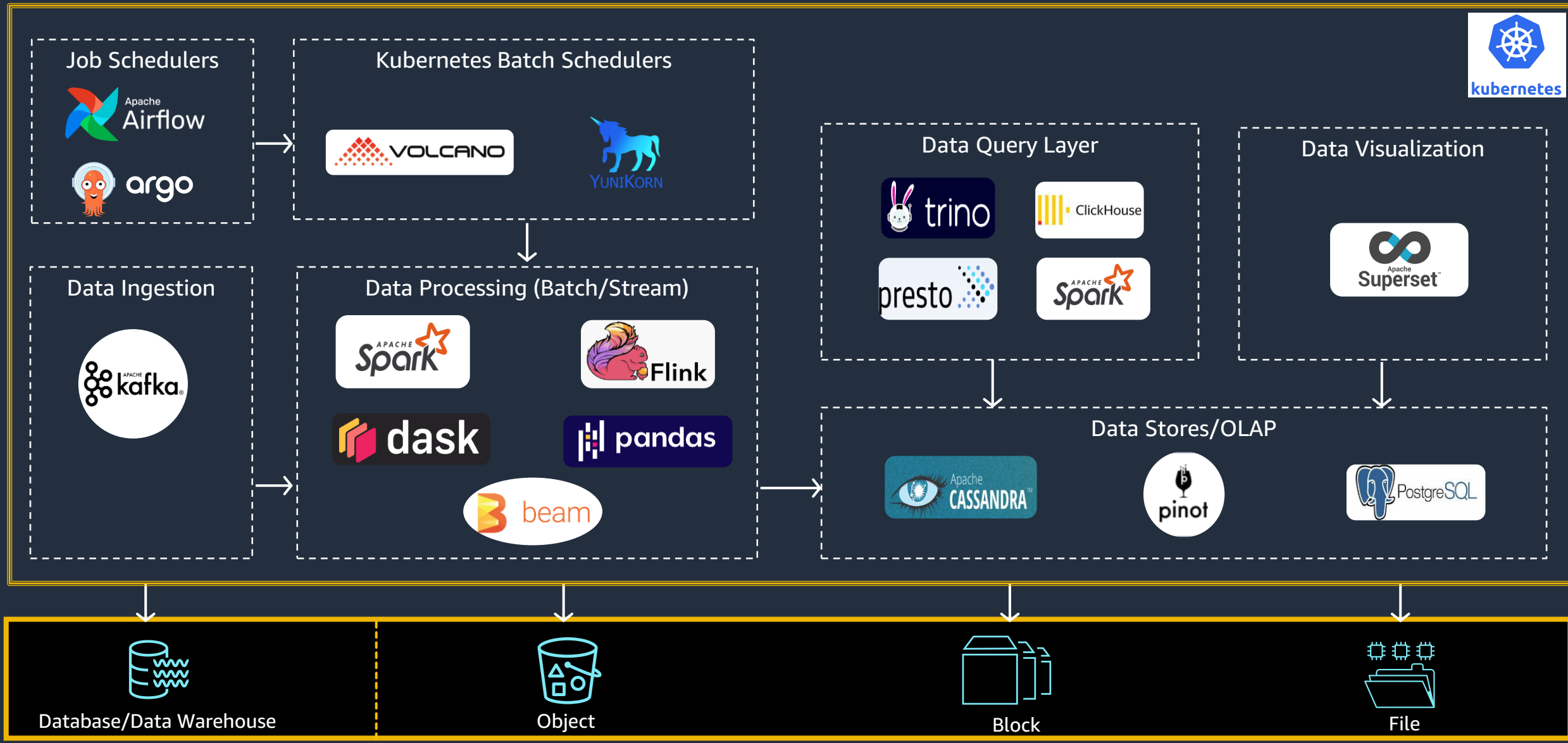
Choose your own Adventure or Endless Quest?

Factors to Consider when Choosing Cloud Storage for Kubernetes

- Data access patterns (read-heavy, write-heavy, mixed)
- Data durability and availability requirements.
- Scalability (1:1, 1:N)
- Performance needs (throughput, latency, etc.)
- Integration and automation with Kubernetes.
- What API access does the application need?
- There is no “F” in Cloud: Cost implications!
- Etc…



Open Source Software (OSS) - Storage Extensibility



Summary

What we're seeing: Increasing persistent workloads, migration to cloud, significant deployment cost.

Workload adoption trends: Stateful applications in AI/ML and analytics are gaining stream quickly.

Extensibility thru the Container Storage Interface (CSI): Enables cross-platform storage drivers, defines APIs for provisioning, attaching, mounting volumes.

Factors to consider: Integration, access patterns, API needs, durability, cost, scalability, performance
- is this File, Block or an Object need?

No single storage solution fits all: Choose based on specific workload requirements and trade-offs of the application requirements.

Work with your peers, community and partners: If the Workload has unknown requirements – someone else may have already seen and solved the issue.

Questions???

Thank you!

Mitch Becker
becmitch@amazon.com
https://www.linkedin.com/in/mitch-becker-577600a/

Tom McDonald
tjm@amazon.com
linkedin.com/in/tomcdonald