
KubeCon
CloudNativeCon
North America 2022
BUILDING FOR THE ROAD AHEAD
DETROIT 2022
TiKV
TiKV
HELM

BUILDING FOR THE ROAD AHEAD

DETROIT 2022

Cloud Native Storage: Storage TAG Intro, Projects, Landscape & Technology

Alex Chircop, Ondat
Xing Yang, VMware
Raffaele Spazzoli, Red Hat

Agenda

- Overview of the TAG
  - How to join and how to help
  - Overview of storage projects in the CNCF
- What is Cloud Native Storage and
  - Why it is important
  - What's New
- Overview of the CNCF Storage Landscape document
  - New: Template Use Cases!
- Overview of the Performance and Benchmarking document
- Overview of the Cloud Native DR document
- Community

Meetings

TAG Storage

CNCF SIGs were renamed TAGs (Technical Advisory Groups)

Meetings are on the 2nd and 4th Wednesday of every month at 8AM PT (USA Pacific)

- **Home:** https://github.com/cncf/tag-storage
- **Conf call:** http://bit.ly/cncf-storage-tag-call
- **Agenda:** http://bit.ly/cncf-storage-tag-minutes
- **Recordings:** http://bit.ly/cncf-storage-tag-recordings
- **Mail list:** https://lists.cncf.io/g/cncf-tag-storage

Our calls and membership are open!

Who are we

- We are a diverse set of users & developers of Cloud Native technologies with a storage focus
- We are leaders & early adopters

Co-Chairs	Tech Leads	TOC Liaisons
■Alex Chircop ■Quinton Hoole ■Xing Yang	■Raffaele Spazzoli ■Luis Pabón ■Sheng Yang ■Nick Connolly	■Erin Boyd

Questions? Reach out and feel free to connect on our mailing list, and CNCF Slack!

What we do

“Scale contributions by the CNCF technical and user community, while retaining integrity and increasing quality in support of the CNCF mission (to make cloud native computing ubiquitous).”

What we do

“Scale contributions by the CNCF technical and user community, while retaining integrity and increasing quality in support of the CNCF mission (to make cloud native computing ubiquitous).”

...this means we

What we do

“Scale contributions by the CNCF technical and user community, while retaining integrity and increasing quality in support of the CNCF mission (to make cloud native computing ubiquitous).”

...this means we

- **Educate**
- **Review Projects**
- **Engage with the user community**
- **Provide subject matter expertise**

CNCF Storage Projects

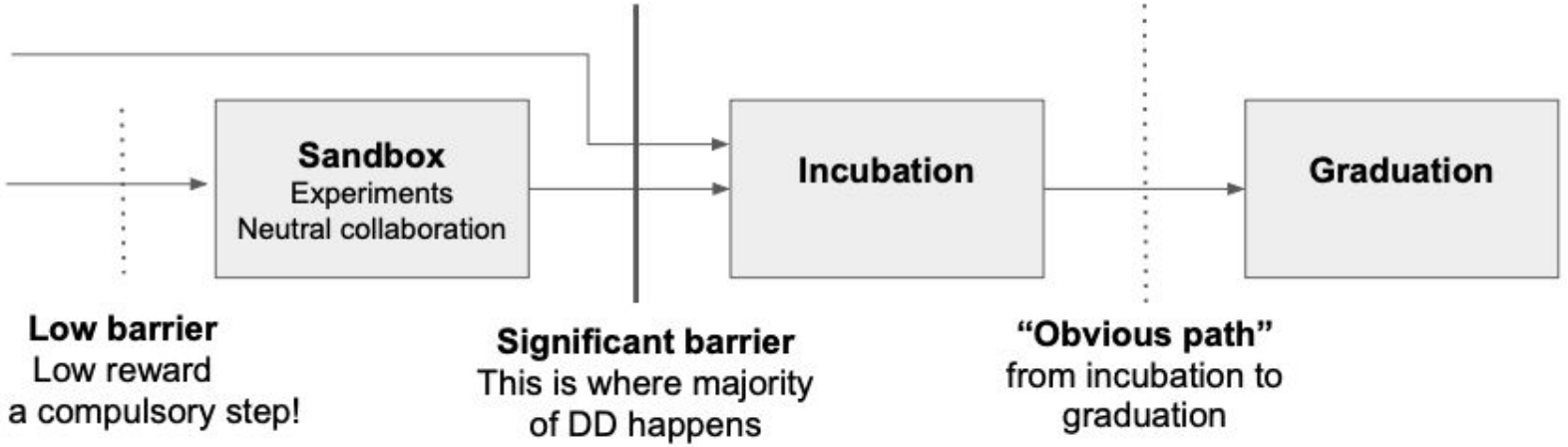
Graduated

Incubating

CNCF Projects: https://www.cncf.io/projects/
Sandbox Projects: https://www.cncf.io/sandbox-projects/

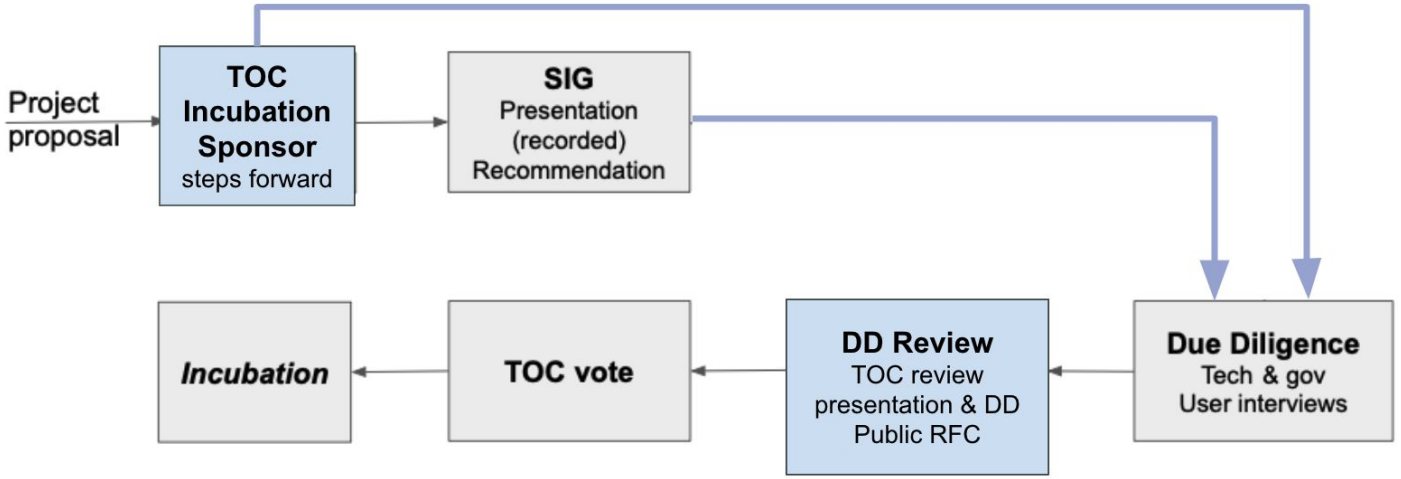
CNCF Project Stages

https://www.cncf.io/projects/



Sandbox	Incubation	Graduation
■Experiments ■IP Policy ■Build Community	■Used successfully in production ■Healthy number of committers ■Project metrics	■Mainstream production use ■Security audits ■Committers from multiple organisations

Project Review



- Understand and **document** a high level roadmap of projects within this space, including CNCF and non-CNCF projects.
- For projects that fall within the CNCF, perform **health checks**.
- Perform **discovery** of and **outreach** to candidate projects.
- **Help candidate projects** prepare for presentation to the TOC.
- Every CNCF project will be assigned to one suitable TAG by the TOC.

Why should you think about this?

Why should you think about this?

There is no such thing as a stateless application …

Why should you think about this?

There is no such thing as a stateless application …

… all applications store state somewhere!

Cloud Native Storage is Here!

Move Stateful Workloads to K8s

- Automation
- Scale
- Performance
- Failover

➪ **Broad ecosystem and CSI support**

➪ **Operators for databases, message queues, *and many more!***

CNCF Storage Whitepaper

KubeCon CloudNativeCon North America 2022

Whitepaper: https://bit.ly/cncf-storage-whitepaperV2

1. Definition of the attributes of a storage system
2. Definition of the layers in a storage solution with a focus on terminology and how they impact the attributes
3. Definition of the data access interfaces in terms of volumes and application APIs
4. Definition of the management interfaces

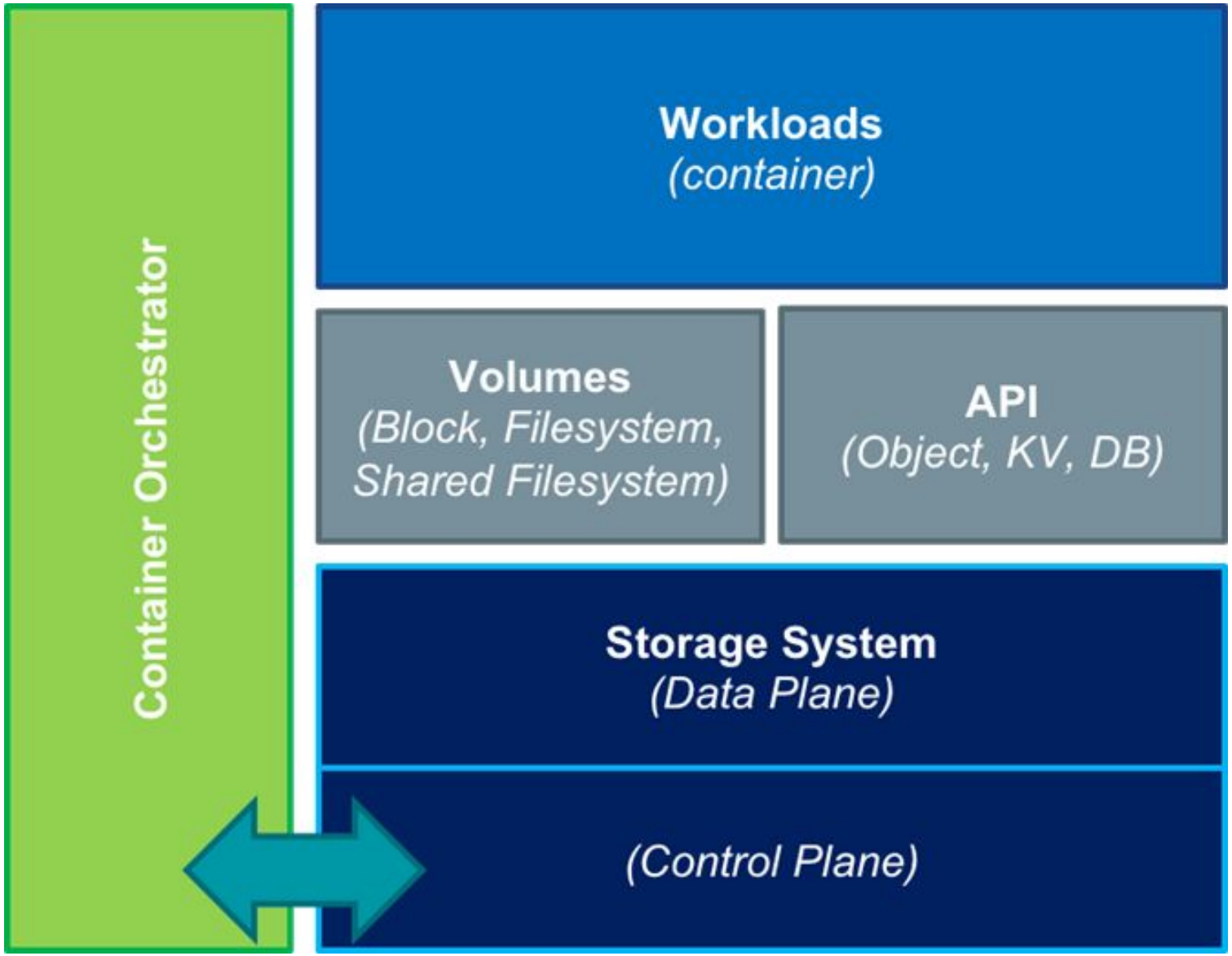
Storage Attributes

Availability	Scalability	Performance	Consistency	Durability
Failover	Clients	Latency	Delay to access correct data after a commit	Data protection
Moving access between nodes	Operations	Operations		Redundancy
Redundancy	Throughput	Throughput	Delay between commit and data being committed to non-volatile store	Bit-Rot
Data Protection	Components			

Storage Layers

Orchestrator, Host and Operating System
Storage Topology (centralized, distributed, sharded, hyperconverged)
Data Protection (RAID, Erasure coding, Replicas)
Data Services (Replication, Snapshots, Clones, etc.)
Physical, Non-Volatile Layer

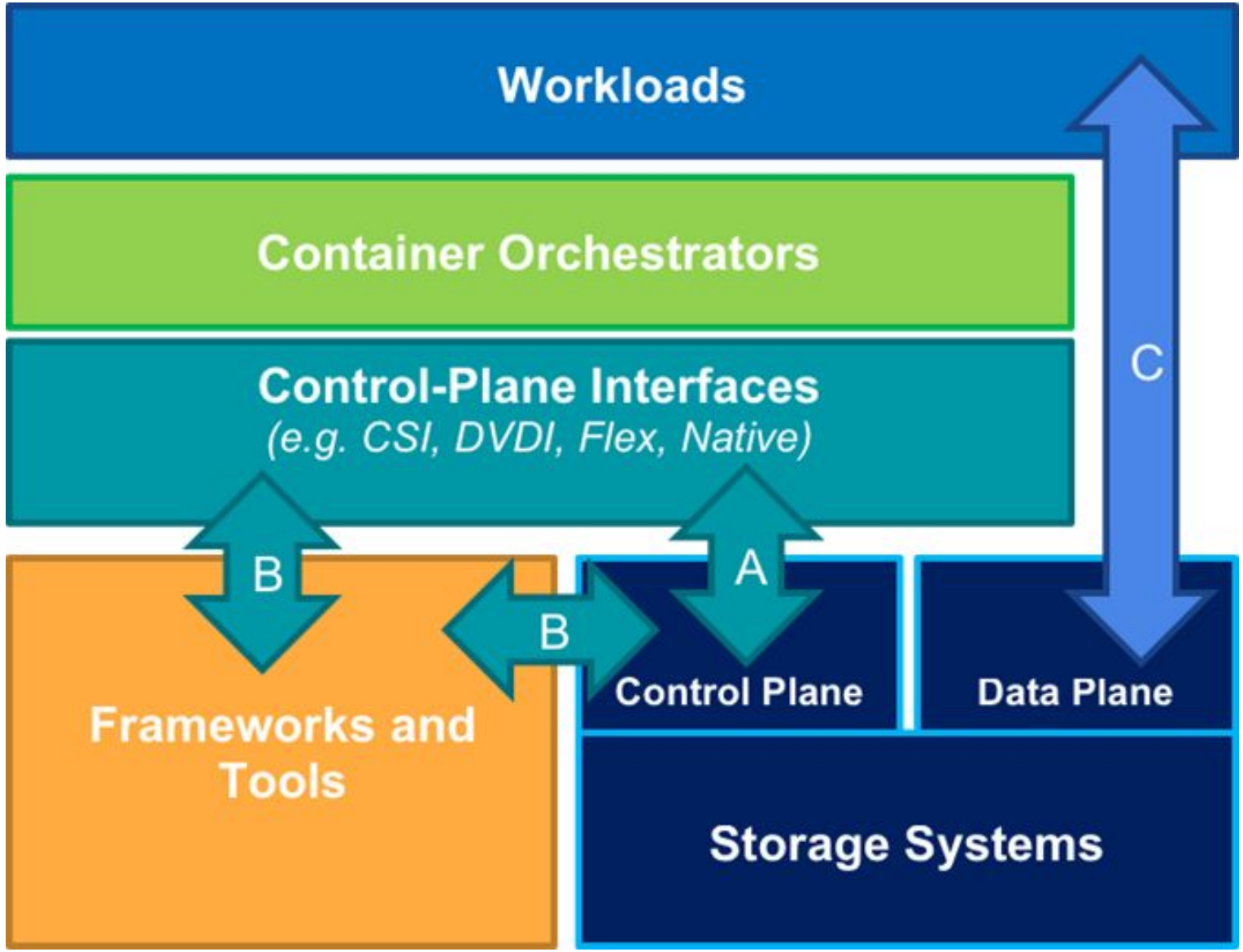
Data Access Interfaces



Storage can be accessed via **Data Access Interfaces**:

- **Volumes** – accessed through a more traditional file interface in a **block** or **filesystem** interface
- **API** – other ways to persist data such as **object stores**, **KV stores** or **databases**

Management Interfaces



Container Orchestration system (**CO**) uses an interface to interact with a storage system

The storage system can:
(A) support control-plane API directly
(B) interact via an API Framework layer or other Tools

Workloads consume **(C)** storage via a data access interface

Use Case Template

- **Based on CNCF Storage Landscape Whitepaper: https://bit.ly/cncf-storage-whitepaperV2**
- **Cloud native open source projects are welcome to fill out the template and give a presentation at a TAG Storage meeting**
- **Use cases will be published at CNCF TAG Storage community repo**

WIP Example Use Case Template

- Storage system: etcd
- etcd is a distributed key-value store used for Kubernetes.
- Strong consistency. Uses Raft consensus algorithm.
- etcd is a “CP” database: Can’t serve write requests if quorum isn’t satisfied.
- Provides stability, reliability, scalability and performance.
- Recording on 6/8/2022
- Presentation on etcd

Storage Topology	
Centralized	
Distribued	Y
Sharded	
Hyperconverged	

Data Access Interfaces	
Object store	
Key-value store	Y
Databases	

Data Protection	
RAID	
Erasure Coding	
Replicas	Y

Data Services	
Replication	Y
Snapshots	Y
Clones	

Performance Whitepaper

Whitepaper: https://bit.ly/cncf-tag-storage-performance-benchmarking

- Definition of common concepts for measuring performance and benchmarking for volumes and databases

- Definition of common pitfalls and considerations
  - Basics: operations vs throughput
  - Topology, Data Protection, Data Reduction, Encryption
  - Latency
  - Concurrency: queue depths, multiple clients and backends
  - Caching at multiple layers
  - Managing the environment, Cloud, and Client headroom

Performance Whitepaper

KubeCon CloudNativeCon North America 2022

Whitepaper:
https://bit.ly/cncf-tag-storage-performance-benchmarking

TL;DR - important takeaway:

published results are not useful for making comparisons - it is hard to compare published results without a deep understanding of the test conditions, so it is always important to run your own test, on your own environment with your own applications

Cloud Native Disaster Recovery

Whitepaper: http://bit.ly/cncf-cloud-native-DR

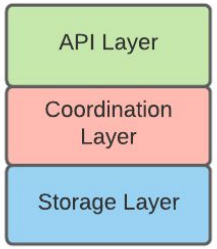
Concern	Traditional DR	Cloud Native DR
Type of deployment	active/passive, rarely active/active	Active / active
Disaster Detection and Recovery Trigger	Human	Autonomous
Disaster Recovery Procedure execution	Mix of manual and automated tasks	Automated
Recovery Time Objective (RTO)	From close to zero to hours	Close to zero
Recovery Point Objective (RPO)	From zero to hours	Exactly zero for strongly consistent deployments. Theoretically unbounded, practically close to zero for eventual consistent deployments.
DR Process Owner	Often the Storage Team	Application Team
Capabilities needed for DR	From storage (backup/restore, volume replication)	From networking (east-west communication, global load balancer)

CNDR Whitepaper content

- Cloud Native Disaster Recovery (CNDR) definition
- Definitions: Failure Domain, HA, DR
- CAP theorem: Consistency, Availability, Network Partitioning
- Anatomy of Distributed Stateful workloads: Replicas, Shards, API Layer, Storage Layer
- Consensus Protocols: Paxos, Raft, 2PC, 3PC
- CNDR Reference Architectures:
  - Strong Consistency
  - Eventual Consistency

Anatomy of a Stateful Application

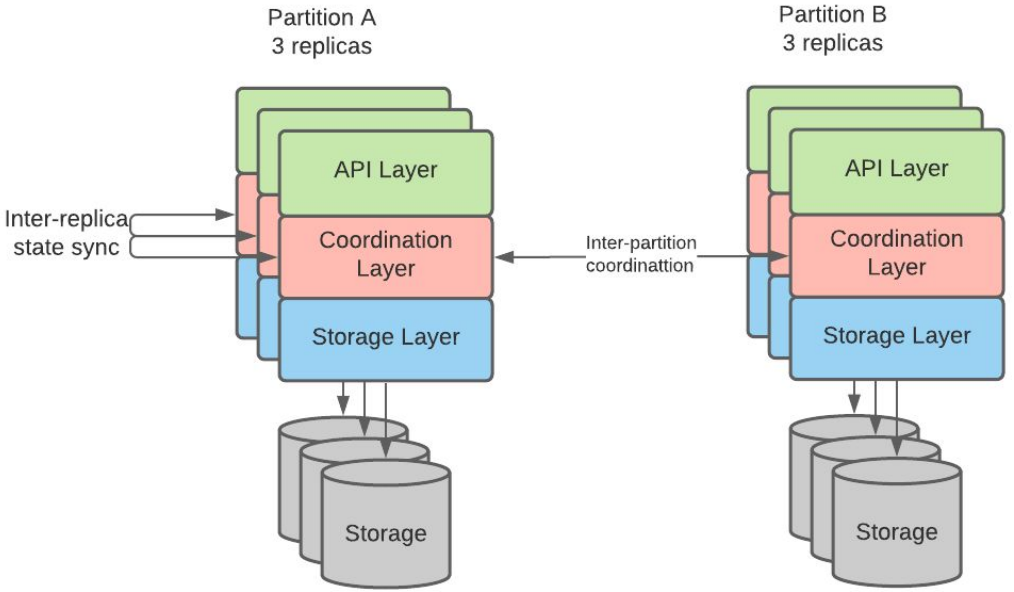
Stateful Workload Logical Tiers



Replicas

Replicas are a way to increase availability of a stateful workload.

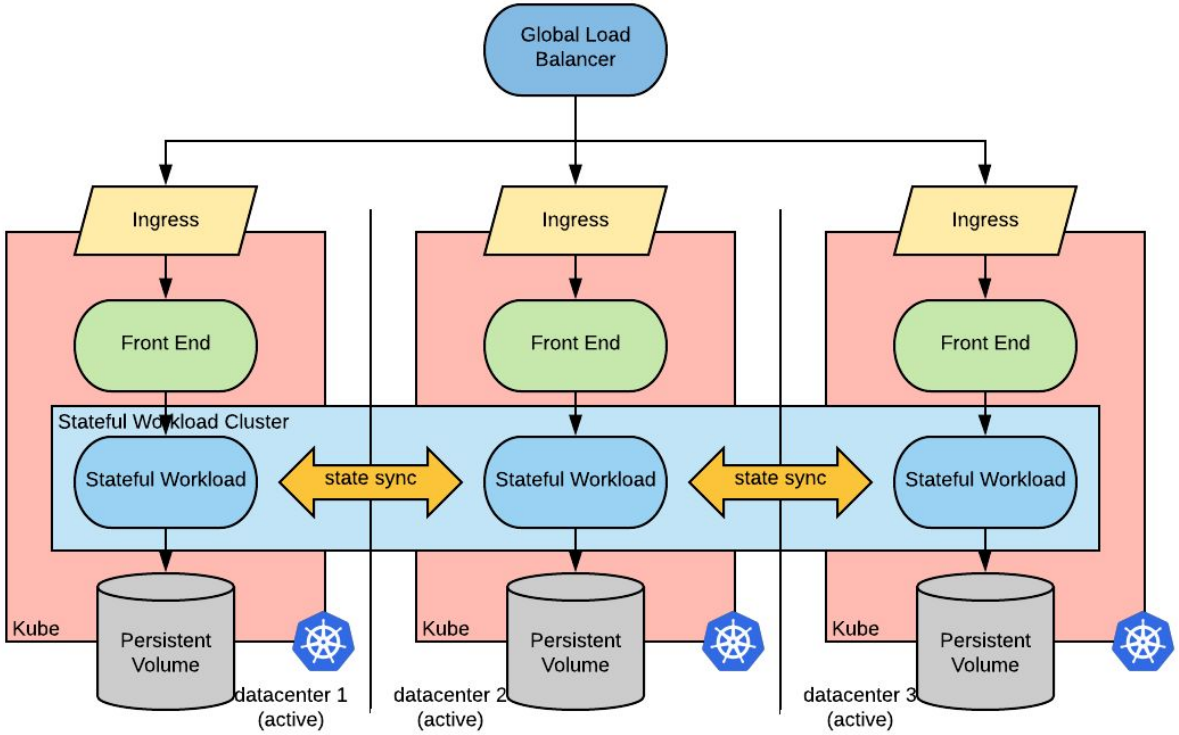
Partitions

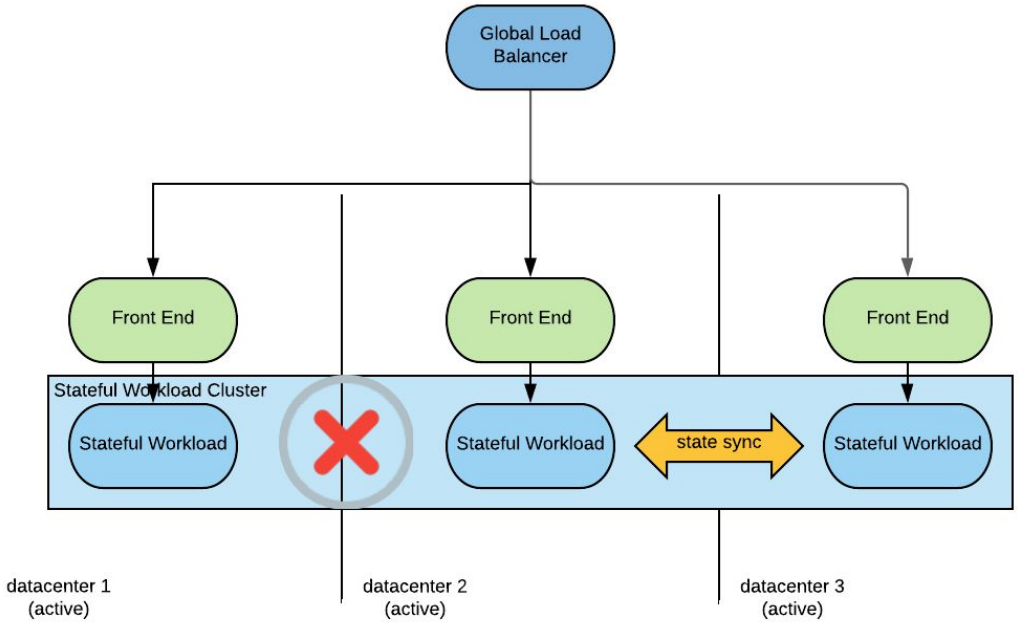


Examples of Consensus Protocol choices

Product	Replica consensus protocol	Shard consensus protocol
Etcd	Raft	N/A (no support for shards)
Consul	Raft	N/A (no support for shards)
Zookeeper	Atomic Broadcast (a derivative of Paxos)	N/A (no support for shards)
ElasticSearch	Paxos	N/A (No support for transactions)
Cassandra	Paxos	Supported, but details are not available.
MongoDB	Paxos	Homegrown protocol.
CockroachDB	Raft	2PC
YugabyteDB	Raft	2PC
TiKV	Raft	Percolator
Spanner	Raft	2PC+high-precision time service
Kafka	A custom derivative of PacificA	Custom Implementation of 2PC

CNDR -- Strong Consistency - Kubernetes Reference Architecture





Community

- How you can get involved?
  - Join our meeting
    - 2nd & 4th Wednesday each month
  - Submit and help review projects for consideration
    - https://github.com/cncf/toc/tree/master/process

- We value community presentations of projects in the cloud native storage space including, but not limited to: *management frameworks, block stores, filesystems, object stores, key-value stores and databases*

- Several projects have presented to the TAG such as: CSI, Rook, REX-Ray, TiKV, Dotmesh, Yugabyte, OpenEBS, Open Services Broker, Vitess, Minio, OpenSDS, Redfish/Swordfish, CubeFS (previously ChubaoFS), Longhorn, Dragonfly, Harbor, Pravega, Piraeus, Dataset Lifecycle Framework, Linstor, Vineyard, CurveStorage

KubeCon | CloudNativeCon
North America 2022

BUILDING FOR THE ROAD AHEAD

DETROIT 2022

Please scan the QR Code above to
leave feedback on this session