STORAGE DEVELOPER CONFERENCE

=SDC 22 | Fremont, CA
September 12-15, 2022

BY Developers FOR Developers

soda
foundation

A **SNIA** Event

Container Data Backup: Going beyond CSI

Rakesh Jain
IBM
STSM and Researcher, SODA TOC Co-Chair

Sanil Kumar D
Huawei
Chief Architect, Head DM(India), SODA TOC Co-Chair

Container, Kubernetes, Cloud, Cloud Native, Data Lake...

Setting the context

*Container Application Deployment is growing, hence the container storage deployments!*

STORAGE DEVELOPER CONFERENCE

SDC 22

Connected.
Anywhere

soda
foundation

STORAGE DEVELOPER CONFERENCE

SDC 22

Data Management Across

| Edge | Core | Cloud |

Storage

| File |

| Platforms |

| Block |

| Applications |

| Object |

| Use cases |

We want
Unified
Global
Across Edge Core and Cloud
Across File, Block and Object
Across platforms, applications and use cases

soda
foundation

STORAGE DEVELOPER CONFERENCE
SDC 22
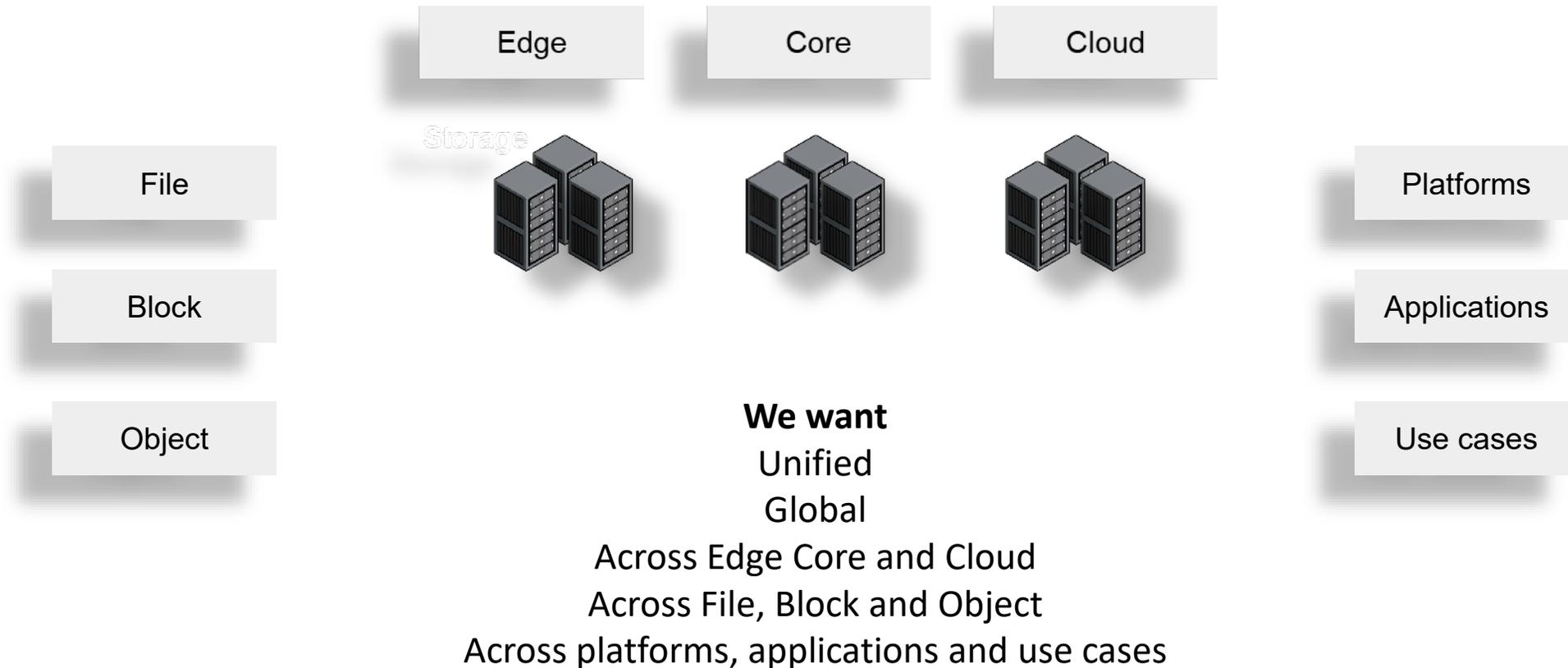
STaaS to CSTaaS

(STorage as a Service to Container STorage as a Service)

SODA Data & Storage Survey 2021[#0]

- Platform use pivoting to **container-based environments**

- Over the last 5 years there has been a seemingly **rapid transition from VMs to containers.**

- Kubernetes deployments came top (in fact **top 3 positions - K8S cloud, Hybrid K8S, K8S on prem**)

- Organizations are using **Container Storage for real use cases deployments**

Other Industry Reports / Surveys

- **Application Container** Market CAGR **26.5%** (2019-25)[#1]

- **CaaS** CAGR **35%** (2021-26)[#2]

- Installed base of **container instances** CAGR **62.1%**(2019-23)[#3]

- **84%** using Container in **Production**[#4]

Container centric storage solutions | Hybrid Cloud Data Management for DR | Momentum towards CSTaaS

STORAGE DEVELOPER CONFERENCE
SDC 22

Solutions Trend

- **Logical solutions to provide end to end data management**
- With Storage Boxes + Features
- Outside box features for Data Management, O&M, …
- Third party storage support
- Container Deployment Support
- Multi-Cloud support
- Application Aware Data Management
- Cross Vendor Product Ecosystems

A few examples:

- NetApp Astra : Application aware data management service built for Kubernetes
- VMware Tanzu & VMware Velero : enables you to build, run and manage modern apps on any cloud, back up and restore your Kubernetes cluster resources and persistent volumes
- Pure Service Orchestrator: Storage as a service for containers; gives developers the agility of public cloud with the reliability and security of on-premises infrastructure
- Product and open source ecosystem across Vmware Tanzu, Velero, NetApp / other storage vendor solutions for end to end container data management

Note: There are similar other products from other vendors and above are just examples for reference.

Not limited to Storage Box Features
Best utilize existing open source / own products to build the solution
Application Aware data management & Container Storage push this trend

soda foundation

STORAGE DEVELOPER CONFERENCE
SDC 22

Container Landscape is growing exponentially!
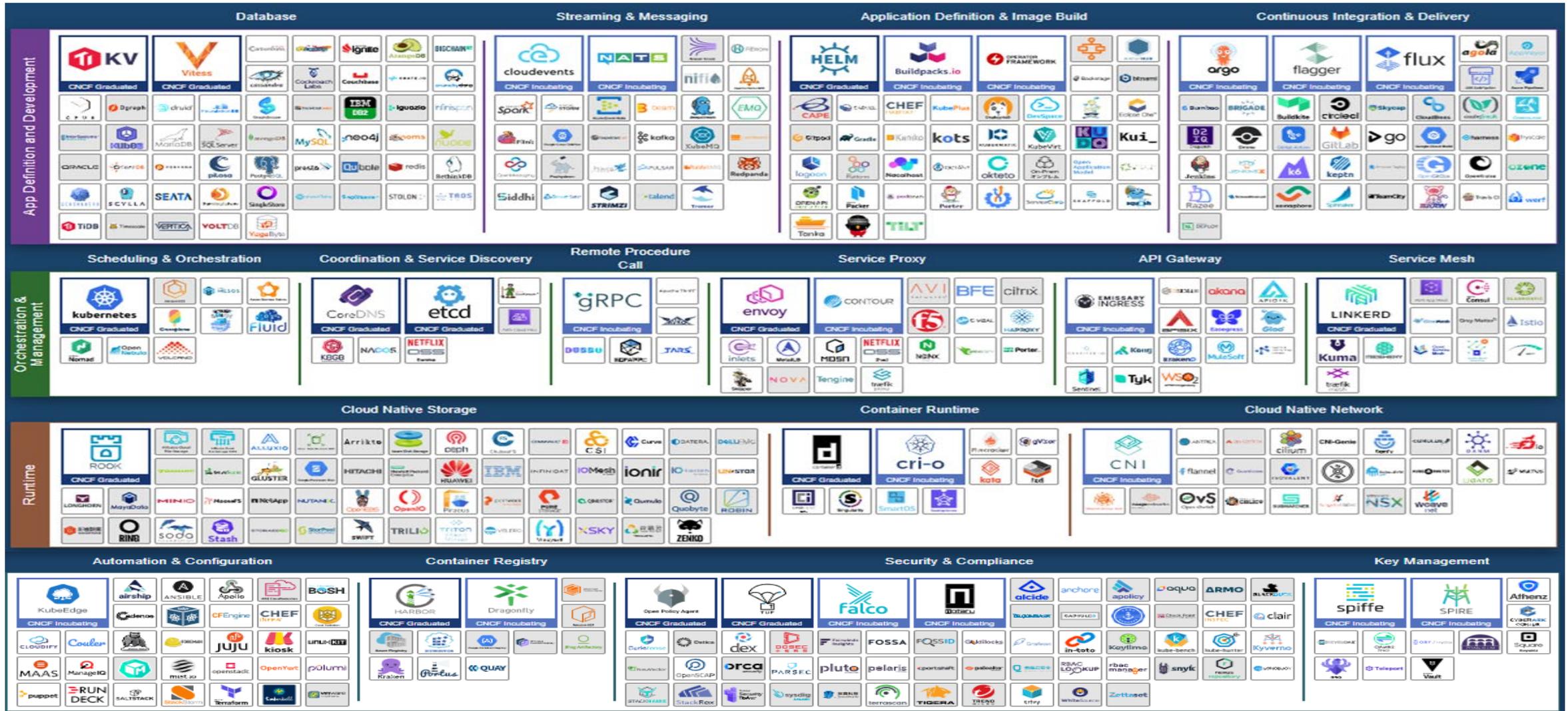
Hence, Container Data Management and Container Storage are critical!

Cloud Native Storage Landscape and Trends

What is happening and what it indicates…?

soda
foundation

STORAGE DEVELOPER CONFERENCE

SDC 22

Revisit the terms - Cloud Native Storage, Container Storage, CSI...

Container Storage

Cloud Native Storage(CNS)

Software Defined Storage - that is API driven and customers can auto-provision. A storage solution that is secure, performant and scalable to application demands.

Container centric storage solutions are Cloud Native Storage solutions and more.

Declarative API (GitOps), Auto-healing and 1000x in terms of Volume Churn. Scale up and down.

Developer productivity, cost optimization and truly hybrid.

Container Storage Interface

Container Storage Interface (CSI) is a specification on how Container Orchestrators interact with Storage Solutions to connect to Containers.

Container Storage is CNS, other way need not be true | CSI is an interface specification

STORAGE DEVELOPER CONFERENCE

SDC 22
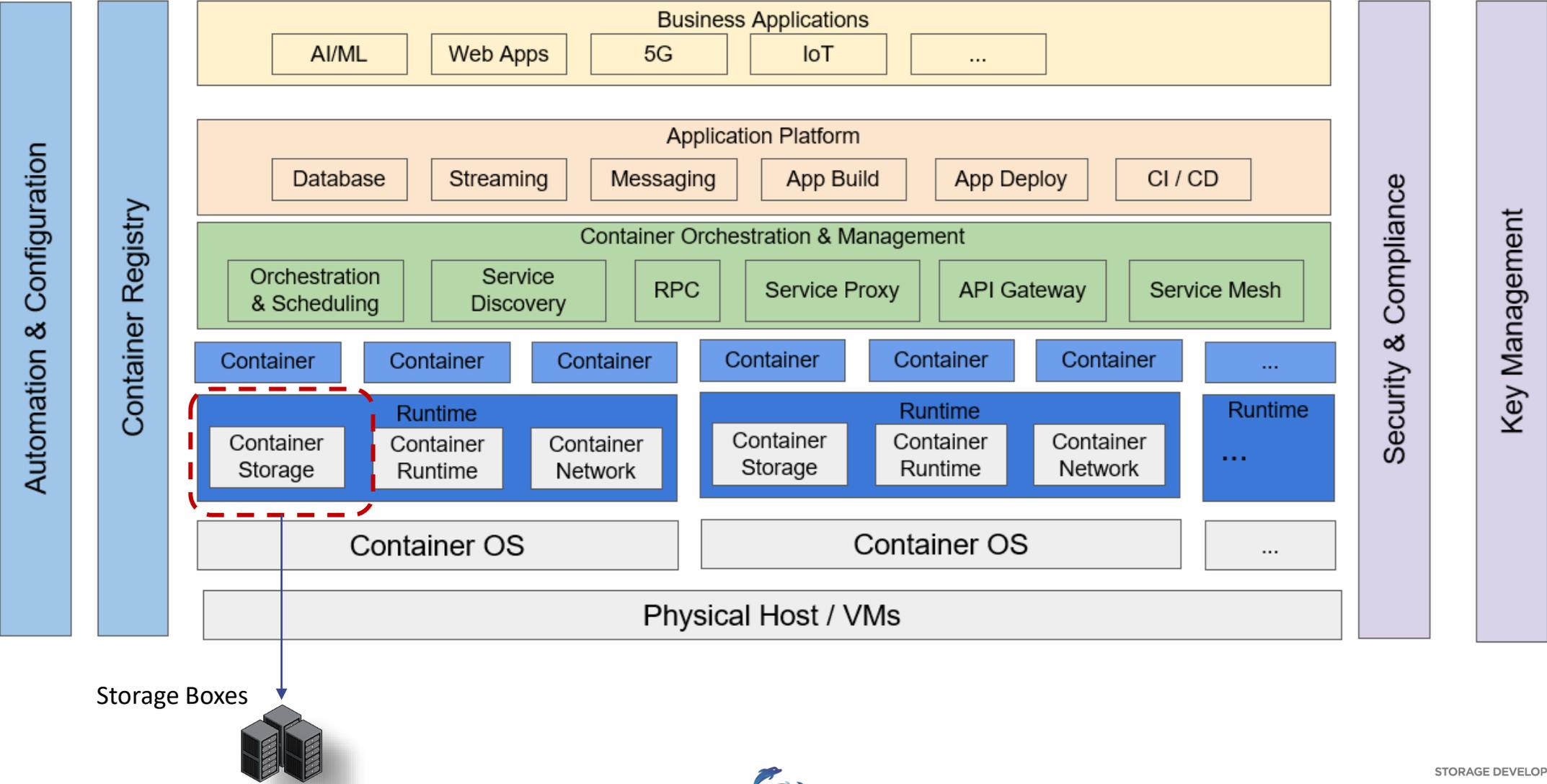
Cloud Native Stack

Cloud Native Storage Projects

SODA ECO Projects

STORAGE DEVELOPER CONFERENCE

# Industry Product Trends : Storage and Platform Vendors

Product Name	Company	Key Use Cases	Supported Infrastructure	Supported Data Types (File/Block/Object)	Deployment Mode	Data Management Technologies Covered	Open Source
Netapp Astra	NetApp	1. Backup/Restore across on prem & cloud (s3 backup supported) 2. Data Migration across clusters 3. Instant clones (working applications) 4. Cloud Bursting	a) within and across public clouds and on-premises b) Mainly NetApp on prem and cloud storages c) k8s multicluster deployments	File/Block/Object	SaaS	1. DR / Snapshots 2. Data Migration 3. Data Provisioning for k8s (NetApp Trident CSI) 4. Lifecycle Management 5. Cloud Bursting	No (Trident is Open Source)
PSO	PureStorage	1. Container STAAS (persistent storage service for containerized app)	a) Within and across public clouds and on-premises b) K8s multicluster deployments (private namespace) c) PureStorage FlashArray and FlashBlade	File and Block	SaaS	1. Data Provisioning (policy based for Containers) 2. DR 3. Data Migration	No (except CSI SP)
Ezmeral	HPE	1. Easy to deploy applications inside Contianers 2. Application Data provisioning and Management 3. Container app data Intelligence	On-prem and Cloud supported by HPE storages	File and Block	SaaS	1. Data Fabric 2. Container and Container storage Monitoring 3. Container appliation deployment and management	No
Karavi	Dell EMC	Observability, Data Mobility and Resiliency for stateful applications	On-prem EMC storage	Block	SaaS	1. Data Observability/Monitoring 2. Data Resiliency	Yes (Karavi is open source)
HSPC	Hitachi	1. Data Protection 2. Snapshot based backup 3. Dynamic storage provisioning	K8S clusters Hitachi VSP series storage volume	Block	SaaS	1. DR/ Snapshot 2. Data Provisioning for K8S (Hitachi CSI) 3. High availability/DR	No (except CSI SP)
Velero	VMWare	1. DR 2. Data Protection 3. Data Migration (hetrogenous/distributed env)	a) Within and across cloud and on-premises b) Any storage provider supporting CSI Plugin c) multi-cluster deployment d) FS based snapshot using restic	Block File Object	SaaS	1. DR/Snapshot 2. Data Migration across distributed env 3. Snapshots (scheduled/on-demand, File and Block)	Yes
Longhorn	Rancher	1. Snapshot based Data Protection 2. Block storage for K8S	On-premise and Cloud	Block	SaaS	1. DR/Snapshot	Yes
Restic	Open Source	1. FS based snapshot 2. uses Object Data Store	On-Premise and Cloud	File	native	1. DR/FS Snapshot	Yes
OpenEBS	MayaData	1. Container Attached Storage (CAS) 2. Persistent Storage 3. Stateful 4. SDS	on-prem and Cloud Plugins for different Storage CSI TP, Cloud Storage	Block File	SaaS	1. Storage (SDS) 2. Availability 3. Persistence	yes

# Products providing Data Management across Cloud and On prem (1)

NetApp Astra
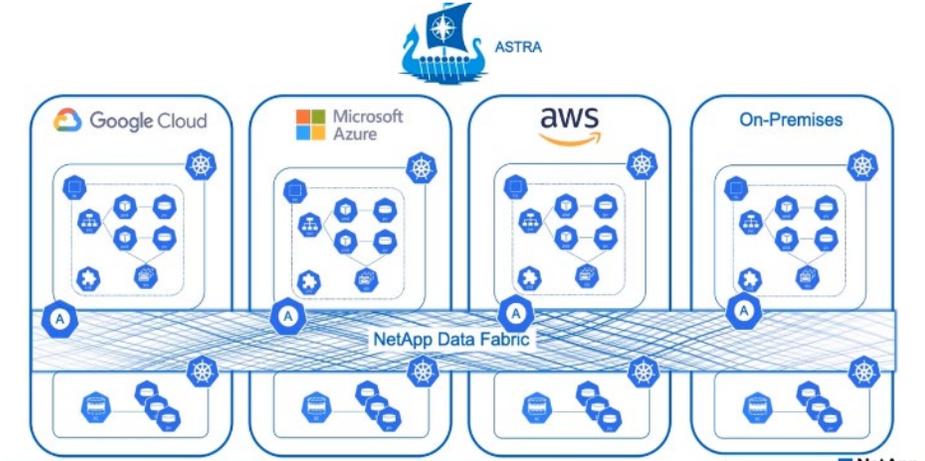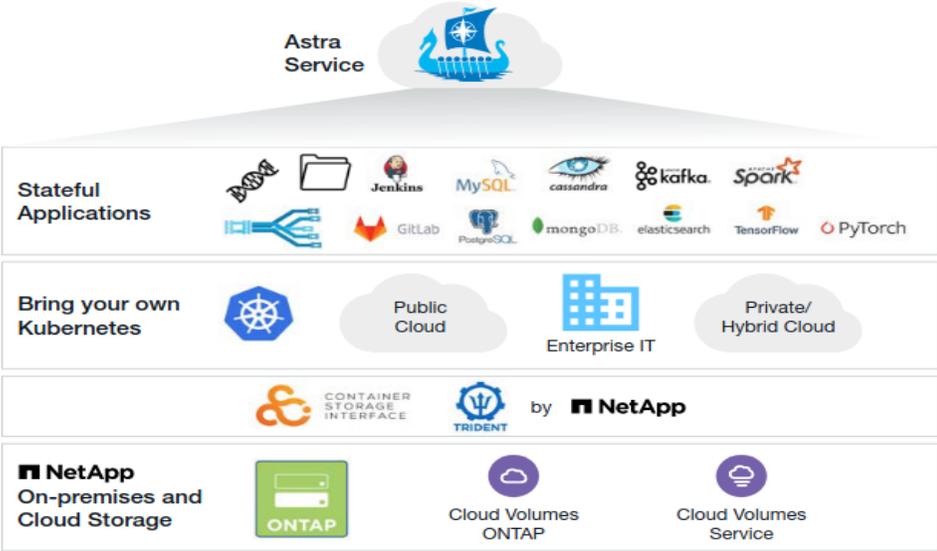
Figure 1: NetApp Astra SaaS for end-to-end application data lifecycle management.

Pure Service Orchestrator
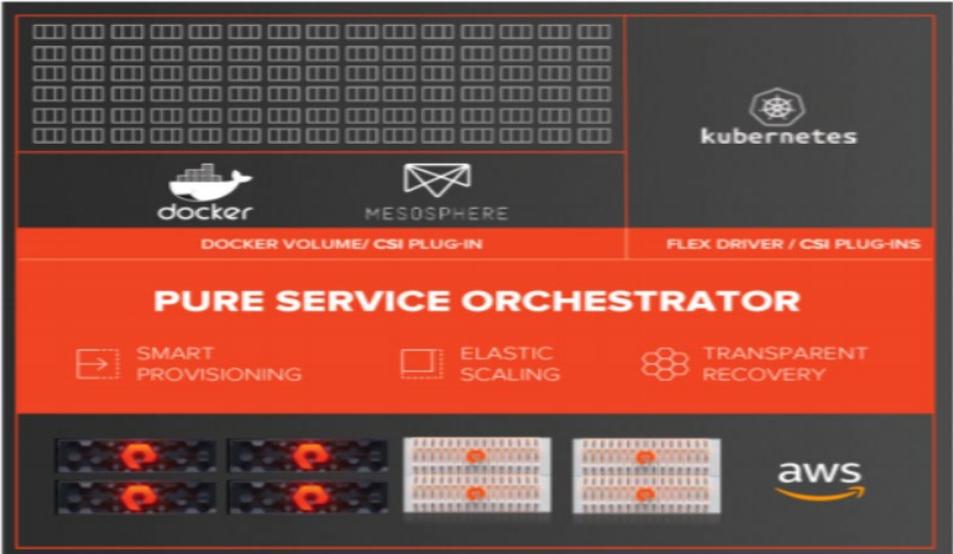
Products providing Data Management across Cloud and On prem (2)

HPE

Dell EMC Karavi

Open Source
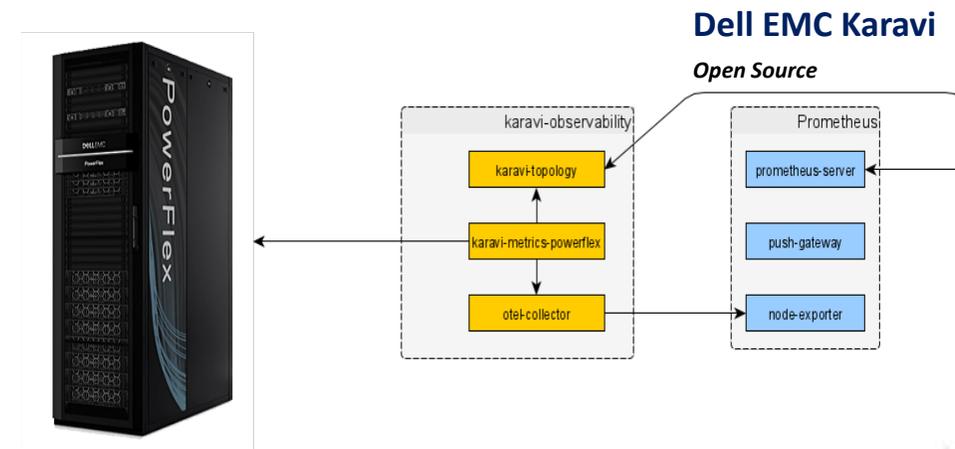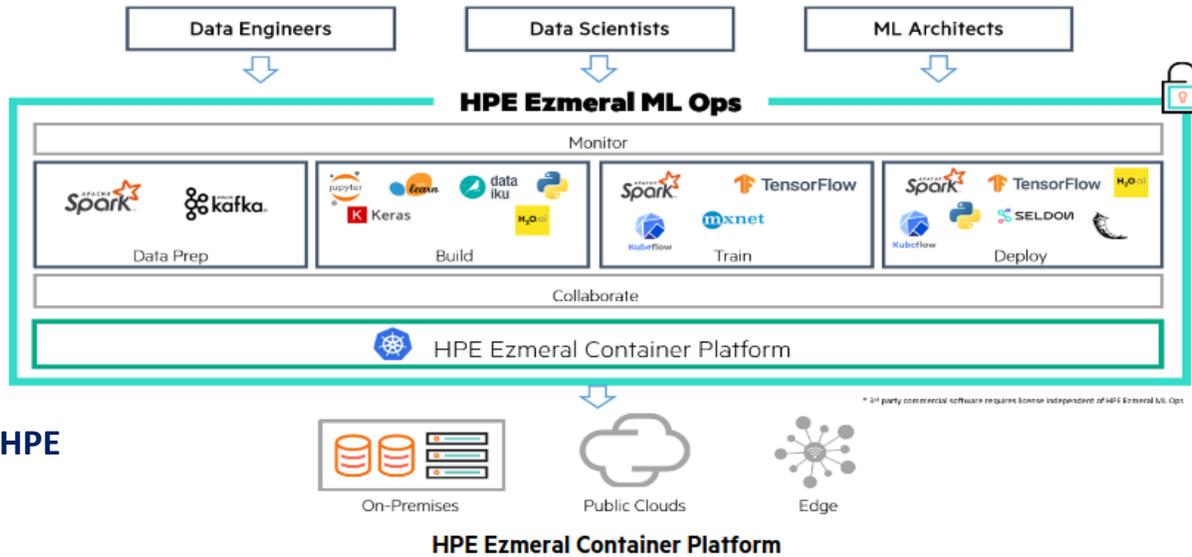
Hitachi

VMware Velero

Open Source
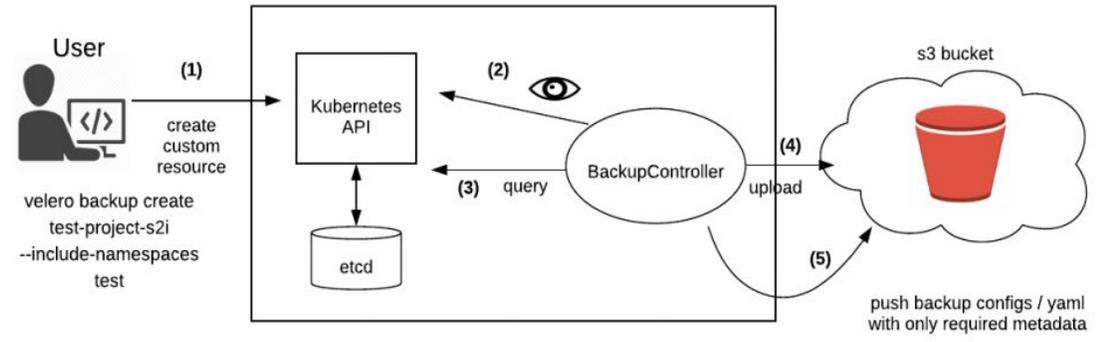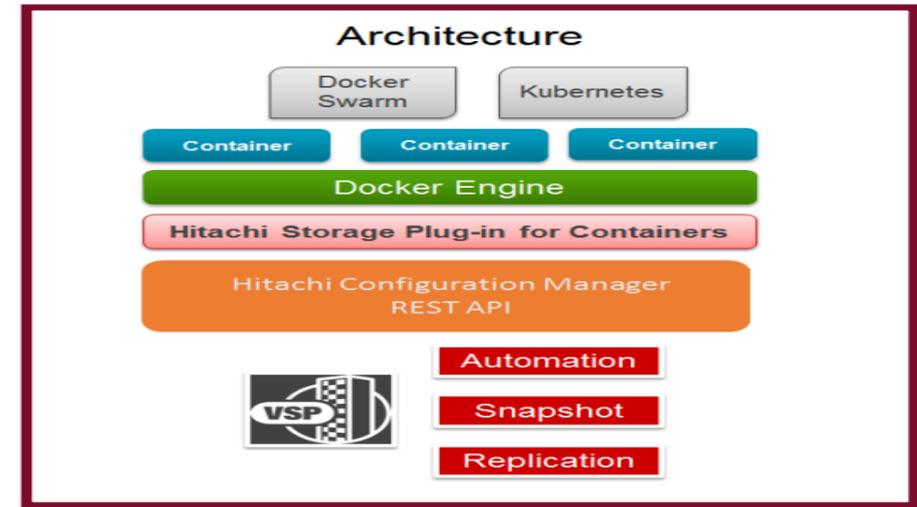
Product Direction is inclined towards hybrid cloud data management, more on cloud native storage support for data protection and mobility

<u>Backup-Restore</u> is kind of primary use case under container data protection

STORAGE DEVELOPER CONFERENCE

SDC 22

Data Protection in Kubernetes

Data Protection is a focus area

Data Protection Working Group

A Working Group dedicated to promoting data protection support in Kubernetes, identifying missing functionality and working together to design features to enable data protection support. Involves collaboration with multiple SIGs such as Apps and Storage. This work-in-progress doc tracks missing building blocks we have identified and what we are working on to fill the gaps.

The charter defines the scope and governance of the Data Protection Working Group.

Stakeholder SIGs

- SIG Apps
- SIG Storage

Meetings

Joining the mailing list for the group will typically add invites for the following meetings to your calendar.

https://github.com/kubernetes/community/blob/master/wg-data-protection/README.md

WIP Document : https://docs.google.com/document/d/1yHbW0hxHehQzdaL7AWSl81OW4f2OcBoskXTbezx92-U/edit#

STORAGE DEVELOPER CONFERENCE

SDC 22

What is available & missing in Kubernetes

- ## Available:
  - Application : Workload APIs: StatefulSet, Deployment, DaemonSet, etc. | Application CRD
  - Storage : Basic Building Block: Volume Snapshots

- ## Missing:
  - Volume Backups
  - Backup Repositories
  - Volume Group, Group Snapshot
  - CBT
  - …

CSI API spec and implementation are considered for enhancements for above missing features under the Data Protection WG in Kubernetes

Multiple projects/products to plug this gap, currently

- Storage vendor specific backup/restore solutions!
- Third Party Commercial products
- Open Source Projects
    - Velero
    - Kahu (SODA Foundation) – We will focus on this today!

SODA CDM : Towards Easy Container Data Management

Augment Kubernetes (or COE) capabilities for heterogeneous and hybrid container data management.
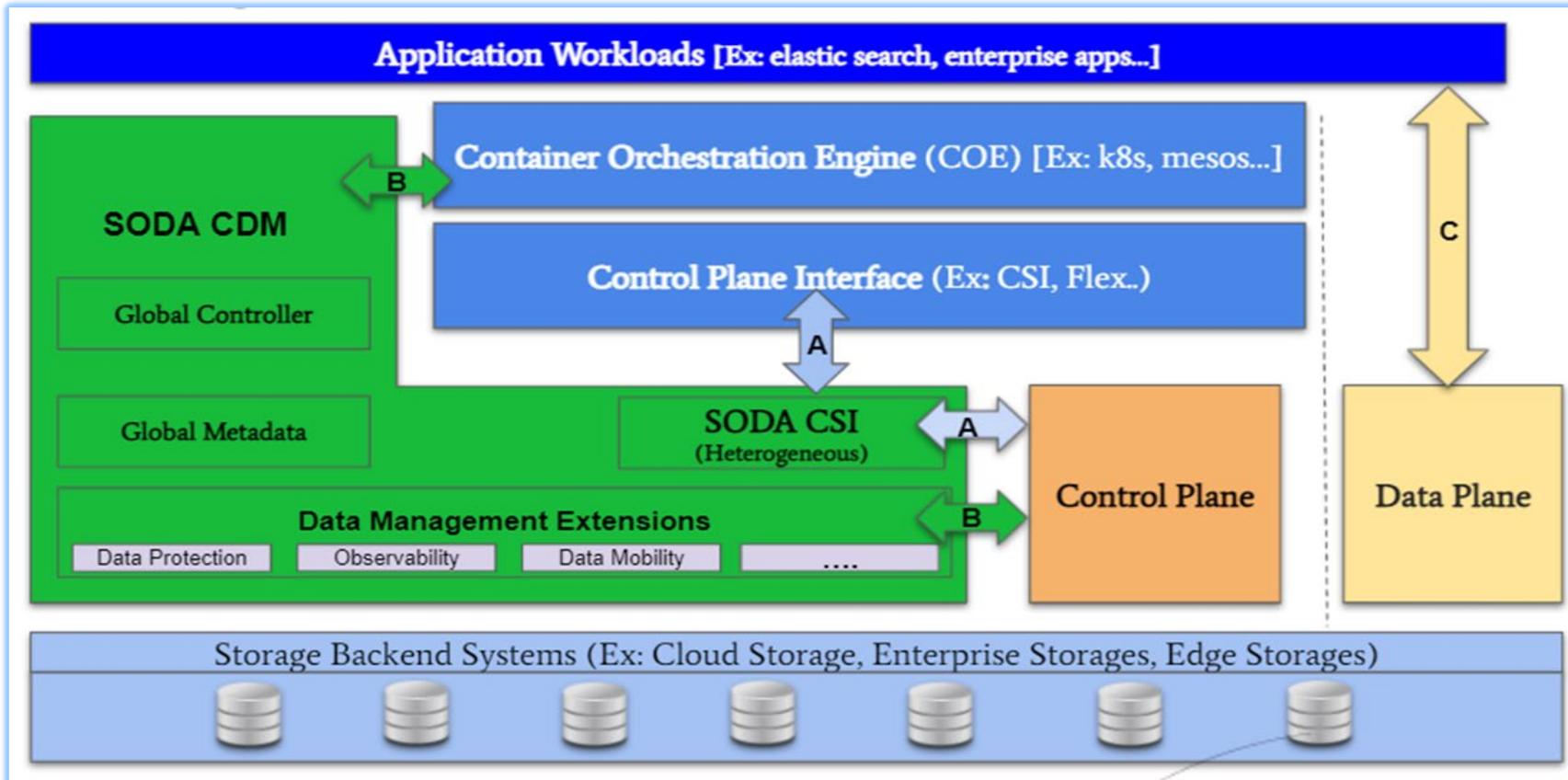
- Designed for Container Data Management: Data Protection, Data Observability, Data Mobility and more
- Aligned with CSI
- Unified CSI
- Heterogeneous Ready
- Multi-cluster and Federation Support
- Hybrid Container Data Management (Kubernetes across on prem, cloud and edge)

A: *Support control plane interface API directly and interact with respective COE*
B: *Interact with Orchestrator through Data Management Framework(SODA)*
C: *Workloads consume storage through data access interfaces.*

STORAGE DEVELOPER CONFERENCE
SDC 22

Kahu : First project under SODA CDM for Seamless Data Backup & Restore!

Backup and restore of Kubernetes Resource, Config and PV data, across multiple clusters and federation environment

- Backup and Restore for Kubernetes
- Federation / Multi Cluster
- Storage Provider Support
- Data Backup Plugins / Extensions
- Low RPO(Recovery Point Objective) / RTO(Recovery Time Objective)

Cloud Backup Support (Object)

Active

Restore

Metadata

Apps

Apps

Data Backup / restore

STORAGE DEVELOPER CONFERENCE
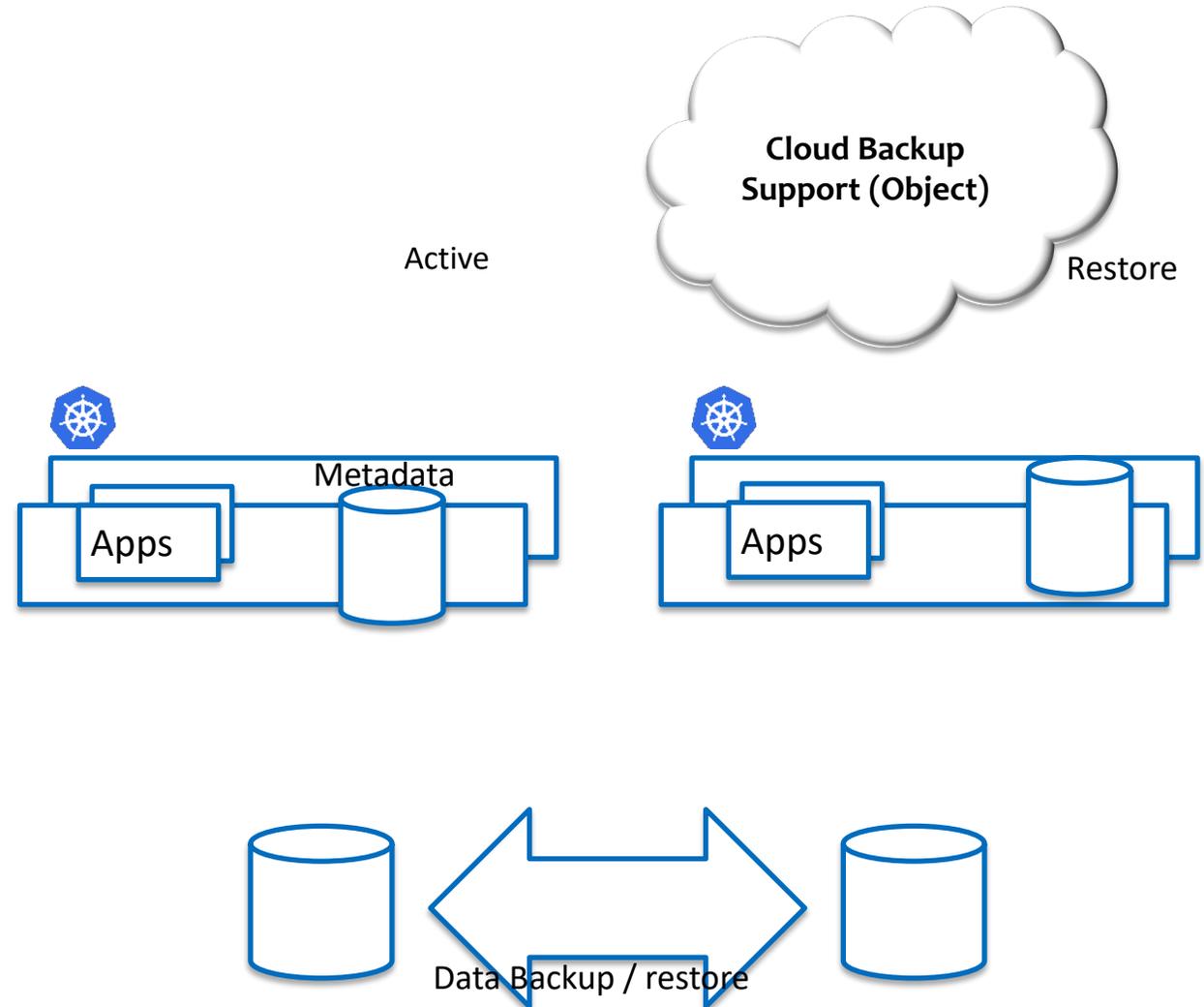
Kahu Phase 1 : 2022

Backup and Restore features
- Metadata backup/Restore of K8S resources at different scope such as cluster, namespace, label selector
- Data backup / Restore of applications in full snapshot way
- Incremental and differential backup
- Custom prehook/post hook
- Backup across storage providers
- CSI based and Non CSI based volume backups.

Storage provider framework support
- Dynamic integration of any storage providers for metadata or volume
- Support coexistence of multiple providers during runtime

Automation and Orchestration
- Scheduled backup
- Backup based on event driven mechanisms
- Backup based on policies
- Backup service plan – Provider independent service plan

Project initiated : https://github.com/soda-cdm/kahu
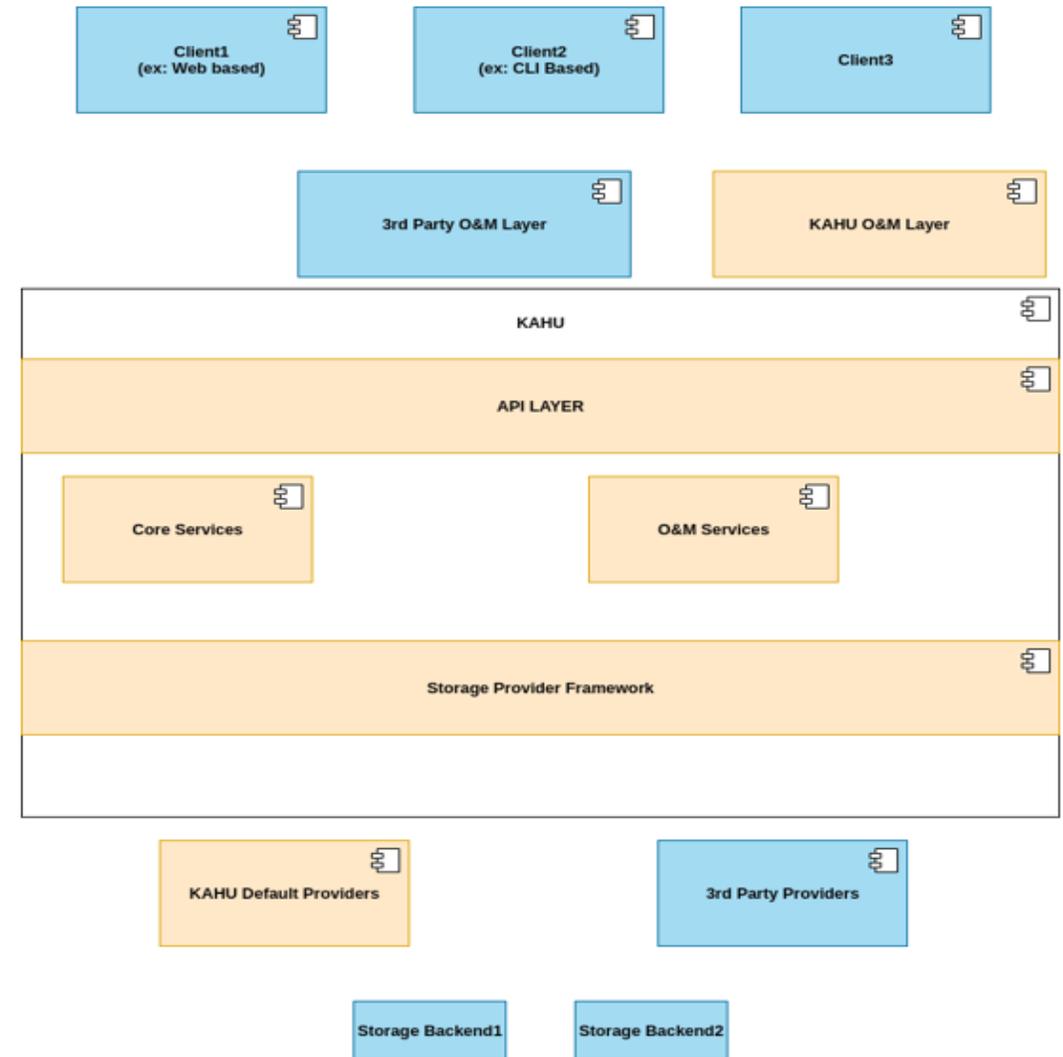
First POC of Kahu in SODA Madagascar v1.7.0 Release

Metadata backup PoC is getting tested!

```
root@ubuntu:~# k get pods
NAME                                          READY   STATUS    RESTARTS   AGE
kahu-backup-controller-58bd685ddf-wcnq6       1/1     Running   0          9m44s
kahu-nfs-provider-79754df9f5-22cqz            2/2     Running   0          30m
root@ubuntu:~# k get svc
NAME                  TYPE        CLUSTER-IP       EXTERNAL-IP   PORT(S)     AGE
kahu-nfs-provider     ClusterIP   10.96.167.143    <none>        8181/TCP    30m
kubernetes            ClusterIP   10.96.0.1        <none>        443/TCP     275d
```

1. Kahu Deployment

```
apiVersion: v1
kind: Pod
metadata:
  name: pod1
spec:
  containers:
  - name: webserver
    image: nginx:latest
    ports:
    - containerPort: 80
```

```
root@ubuntu:~# k get pods
NAME                                          READY   STATUS    RESTARTS   AGE
kahu-backup-controller-58bd685ddf-wcnq6       1/1     Running   0          12m
kahu-nfs-provider-79754df9f5-22cqz            2/2     Running   0          32m
pod1                                          1/1     Running   0          8s
```

2. Application pod deployed

```
apiVersion: kahu.io/v1beta1
kind: Backup
metadata:
  name: example-pod-backup
spec:
  includedResources: ["pod"]
```

3. Backup CRD

```
root@ubuntu:~# ls -lart /tmp/nfs
total 16
drwxrwxrwx 2 nobody nogroup   4096 May 25 22:06
drwxrwxrwt 9 root   root     12288 May 25 22:17
root@ubuntu:~# k apply -f backup-cr.yaml
backup.kahu.io/example-pod-backup created
root@ubuntu:~# k get backups.kahu.io
NAME                 AGE
example-pod-backup   19s
root@ubuntu:~# ls -lart /tmp/nfs
total 20
-rw-r--r-- 1 nobody nogroup   1132 May 25 22:18 example-pod-backup
drwxrwxrwx 2 nobody nogroup   4096 May 25 22:18
drwxrwxrwt 9 root   root     12288 May 25 22:18
```

4. Backup Created

SODA Madagascar Release : https://github.com/sodafoundation/soda/tree/master/madagascar

STORAGE DEVELOPER CONFERENCE

SDC 22

Join Us

Right time to collaborate! (Now is the right time! ☺)

We welcome developers of all levels of skills and experience to join to build these future technology solutions!

slack

https://sodafoundation.io/slack

Channels: #soda-cdm | #soda-lake

SODA Data and Storage Trends Survey is announced!

https://www.sodafoundation.io/data-storage-trends-survey-2022/

STORAGE DEVELOPER CONFERENCE

THANK YOU!

https://www.sodafoundation.io/

SODA Source Code:
https://github.com/sodafoundation
https://github.com/soda-cdm
SODA Docs: https://docs.sodafoundation.io/

Join SODA Slack:
https://sodafoundation.io/slack/
Follow SODA Twitter:
https://twitter.com/sodafoundation

Please take a moment to rate this session.

Your feedback is important to us.

STORAGE DEVELOPER CONFERENCE

SDC 22