FOUNDATION

- SODA Foundation is chartered under Linux Foundation

- Focuses on open source on data and storage management

- Launched on Jun 29, 2020

- Mission:

  - foster an ecosystem of open source data management and storage software for data autonomy

  - offer a neutral forum for cross-projects collaboration & integration,

 - provide end users quality end-to-end solutions

PREMIER MEMBERS

China unicom 中国联通   FUJITSU   HUAWEI   NTT Communications   TOYOTA

GENERAL MEMBERS

LIN·BIT   MayaData   SCALITY

wipro   YAHOO! JAPAN   SEAGATE

SUPPORTERS

建信金融科技 CCB Fintech   e·Hualu 易华录   GMOペパボ PEPABO

光华冠群 GUANGHUA CK   IBM   IIJ Internet Initiative Japan   kpn

ProphetStor   SAKURA internet   SCUTECH 鼎甲科技   vodafone

ASSOCIATE MEMBERS

CCICI   CESI 中国电子技术标准化研究院   JDSF Japan Data Storage Forum

Open Infrastructure FOUNDATION   Kristu Jayanti College

ALLIANCE PARTNER

SNIA

soda foundation

STORAGE DEVELOPER CONFERENCE

SDC 22

End Users

SODA is an end-user driven foundation. End users drive roadmap requirements, provide use cases, test and provide feedback, and guide opportunities for data and storage technologies.

The SODA End User Advisory Committee meets regularly and provides guidance to the Board and TOC. The organizations represented in the EUAC manages some of the biggest data in the world.

Yuji Yazawa Toyota Motor Corp	**Cosimo Rossetti** Vodafone	**Zhong Xin** China Unicom	**Kei Kusunoki** NTT Communications
Tomoko Kondo Softbank	**Zhan Shu** China Construction Bank Fintech	**Yusuke Sato** Yahoo! JAPAN	**Wim Jacobs** KPN
Michiharu Nakazawa Sakura Internet	**Mitchitaka Terada** Internet Initiative Japan	**Wei Rao** China Railway	**Shinya Tsunematsu** GMO Pepabo

SODA END USER ADVISORY COMMITTEE

STORAGE DEVELOPER CONFERENCE

SDC 22

BRIEF HISTORY OF SODA

Storage Management Problem

Management Complexity
Need to manage multiple storage clusters

Storage Silos
Hard to reallocate resources

Unknown Storage Support
Newer platforms have limited storage choices

kubernetes | Apache MESOS | docker | CLOUD FOUNDRY | openstack | vmware

controller | controller | controller | controller | controller | controller

Storage Cluster | Storage Cluster | Storage Cluster | Storage Cluster | Storage Cluster | Storage Cluster

reallocate

no support

soda foundation

STORAGE DEVELOPER CONFERENCE
SDC 22
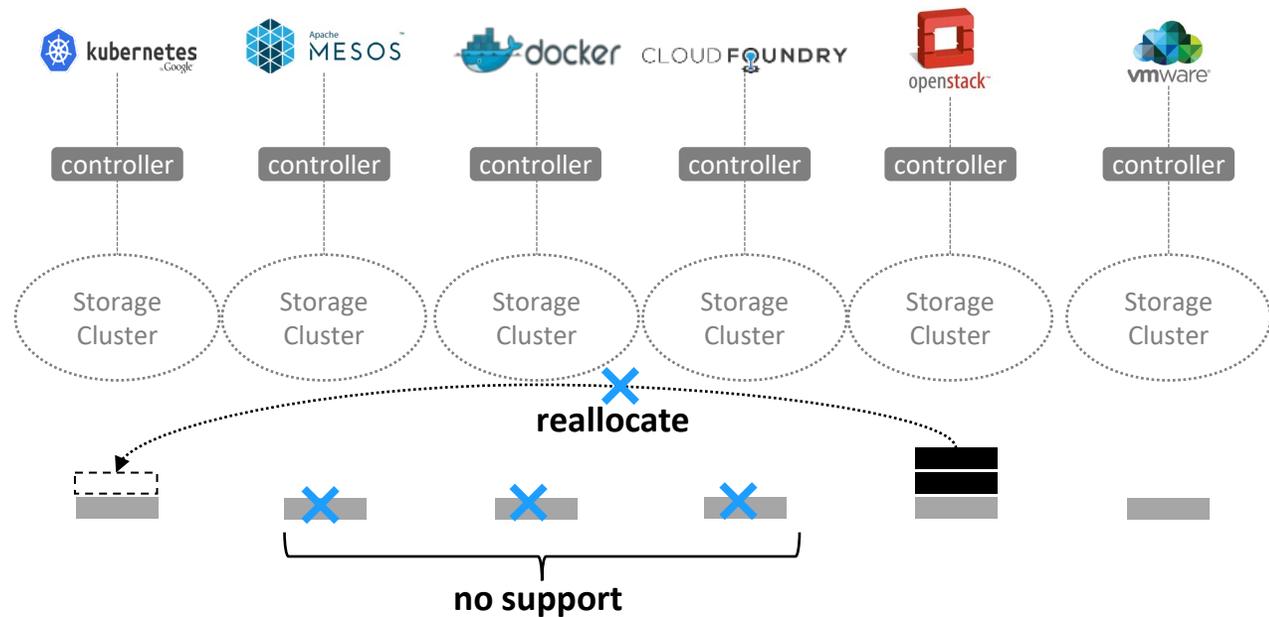
Solution: Make Storage Simple

Complex Admin Centric API's

```
1  cloud a: xxx create 10 --name test-001 --consisgroup-id consisgroup-01
2  --availability-zone xyz --allow-multiattach=true --volume-type vol-type-silver
3  --metadata type=bound --hint position=xyz --image-id 63fe1f3a-49b7-11e7-ab57-286
4  ed488d415
5
6  cloud b: xxx create --kind xyz --apiversion v1 --metadata-name test-001 --parame
   mter-backenddriver xyz --paramemter-volumetype vol-type-silver --paramemter-allo
7  w-multiattach true --paramemter-image-id 63fe1f3a-49b7-11e7-ab57-286ed488d415
8  --spec-accessmodes readwritemany --spec-resources-requests-storage 10Gi
9
```

Unified User Friendly API's

```
1  cloud a: osds create_volume 10 -n test-001 -p my-great-policy
2  cloud b: osds create_volume 10 -n test-001 -p my-great-policy
3
4
5
6
7
8
9
```

Complicated Configuration
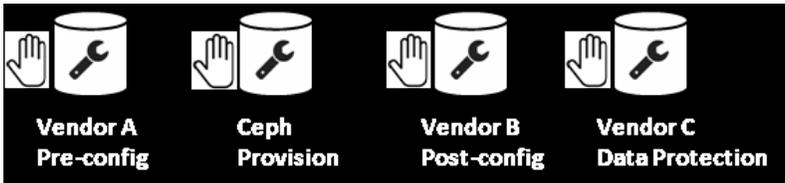
Policy-based Config and Ops
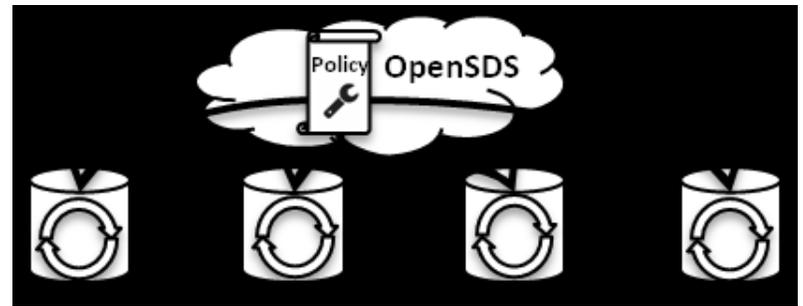
No Differentiation Between Storage
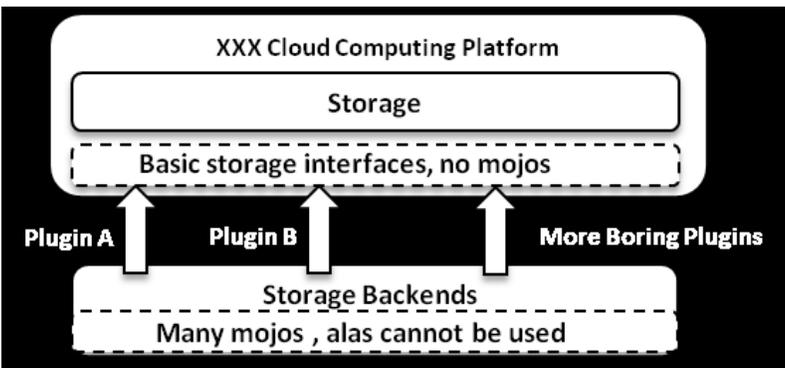
Storage With Unique Features
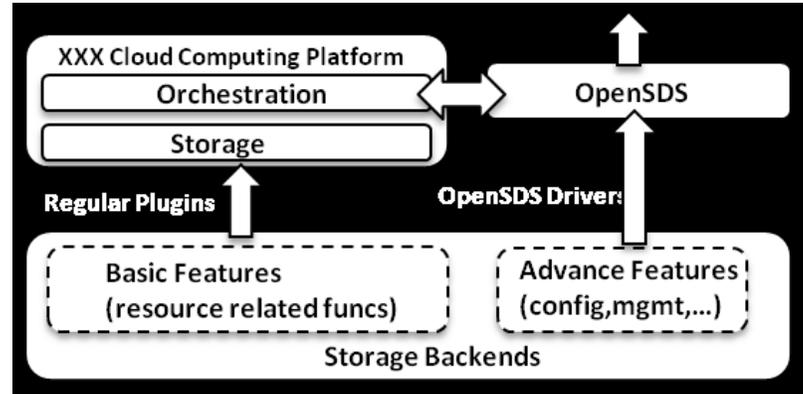
OpenSDS Architecture (TERRA)
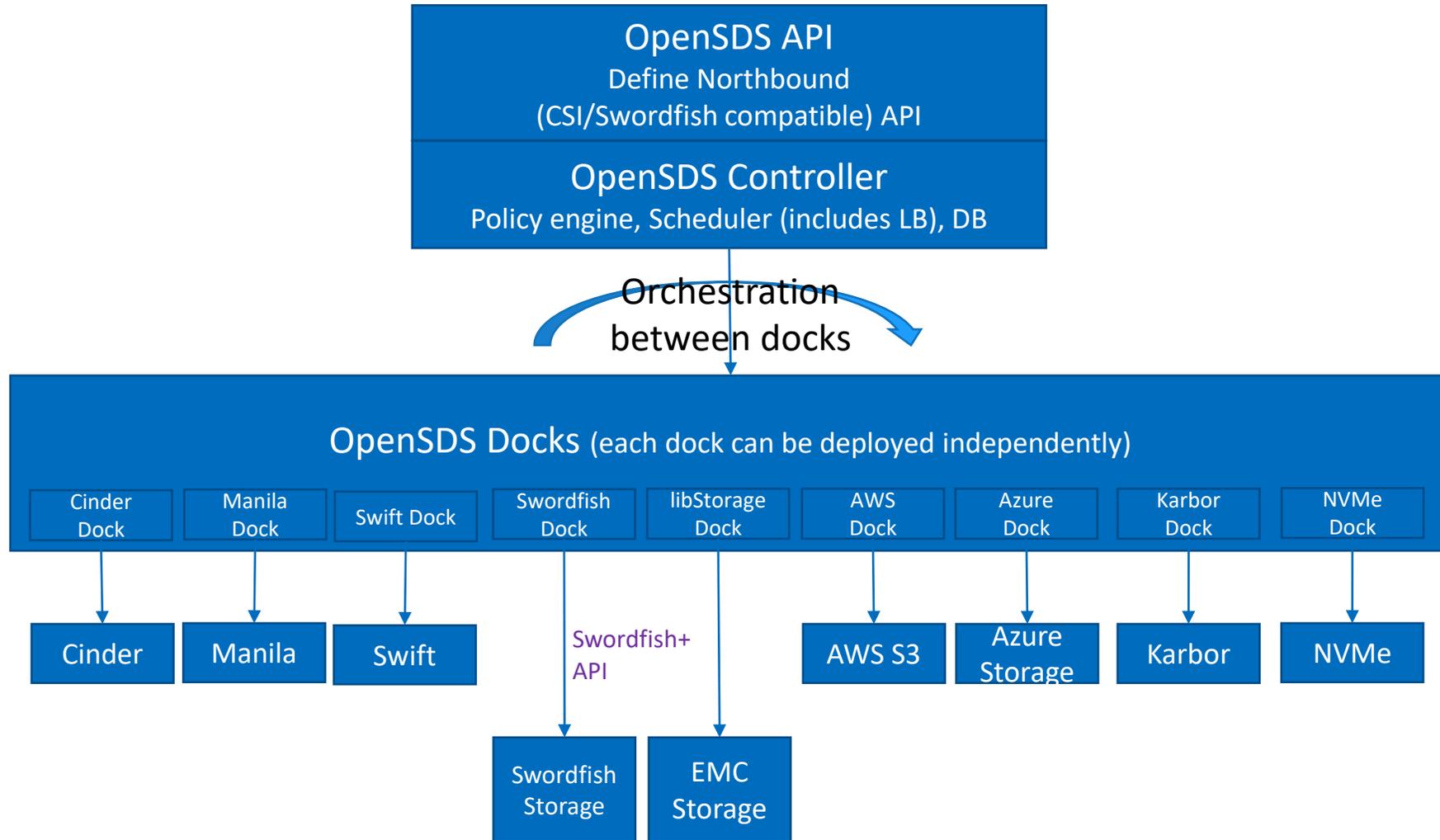
CURRENT SODA PROJECTS

Growing Project Ecosystem

⚙ ORCHESTRATION

SDS Controller	Container Data Management	Container Storage Management
TERRA	**KAHU**	**LINSTOR**

👁 OBSERVABILITY

Storage Monitoring	Storage Benchmark Kit
Delfin	**SBK** · Independent Open Source

📡 EDGE

Edge Container Management
KUBEEGE · CLOUD NATIVE COMPUTING FOUNDATION

☁ HYBRID/MULTICLOUD

Multicloud Data Management	Multicloud Data Lake	Multicloud Orchestration	Multicloud Data Controller
STRATO	**COMO**	**KARMADA** · CLOUD NATIVE COMPUTING FOUNDATION	**ZENKO** · SCALITY

🗄 STORAGE

Mass Object Storage	Mass Capacity Object Storage	NVM Object Storage	Container Attached Storage	Container Attached Storage
YIG · China unicom 中国联通	**CORTX** · SEAGATE	**DAOS** · intel.	**OPENEBS** · CLOUD NATIVE COMPUTING FOUNDATION	**CUBEFS** · CLOUD NATIVE COMPUTING FOUNDATION

SODA Framework Projects	New SODA Framework Projects (2022)	SODA ECO Projects

Framework Projects for SDS Controller, Multicloud, Container Data Management and Data Lake | 10 External Projects in Data & Storage joined as SODA ECO Projects to build collaborative solutions.

STORAGE DEVELOPER CONFERENCE
SDC 22

SODA Framework Projects :
Building Open Data Framework for Hybrid Data Management

STRATO – Multicloud Data Management

- Provides a cloud vendor agnostic data management for hybrid cloud, intercloud, or intracloud.
- Unified interface for file, block, and object services across multiple cloud vendors.
- S3 compatible APIs for object data management on cloud or on premise
- Deploy on-premise or in the cloud

DELFIN – Heterogeneous Storage Monitoring

- Unified, intelligent and scalable resource management
- Performance monitoring and alerting across heterogeneous storage
- Exporter framework to add custom clients from data processing and visualization. Prometheus supported out of the box
- Driver manager to add more storage vendors support seamlessly

TERRA – SDS Controller

- Standardized API, Controller for metadata and Dock for Drivers to provide seamless data management across various storage vendors
- Supports to connect different platforms like Kubernetes, Open Stack, VMware through plugins
- Supports custom vendors drivers and CSI plugins for heterogenous storages

https://github.com/sodafoundation/

Unified Data Management API

Global Controller (Metadata, Policy, ...)

| File | Block | Object |

Storage Backend Manager

| Edge | Core | Cloud |

Storage

open data
framework

STORAGE DEVELOPER CONFERENCE
SDC 22
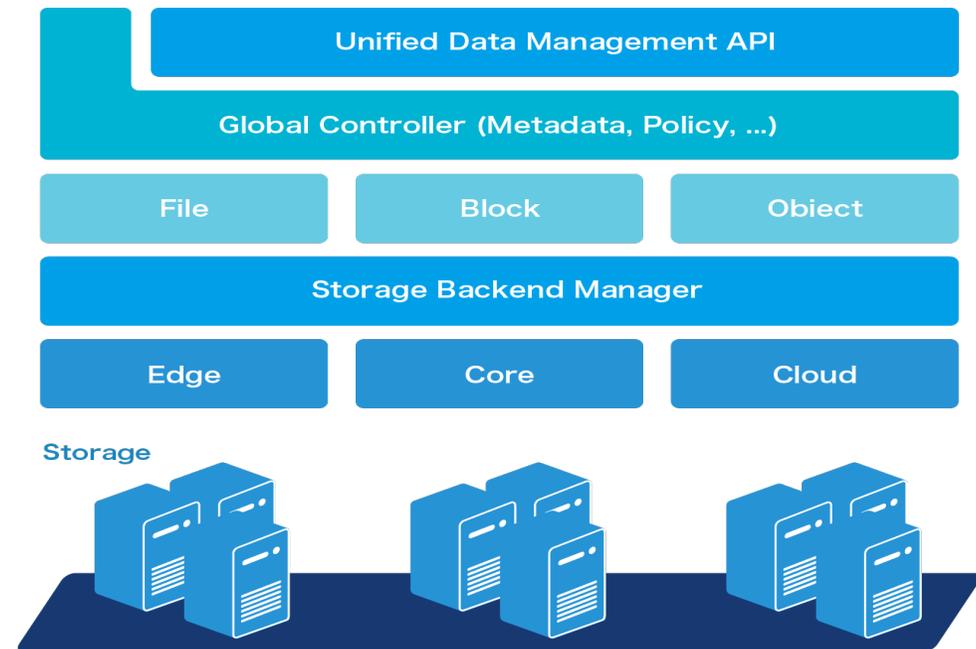
New SODA Framework Projects :
Moving towards Cloud Native Data Management & Data Lake
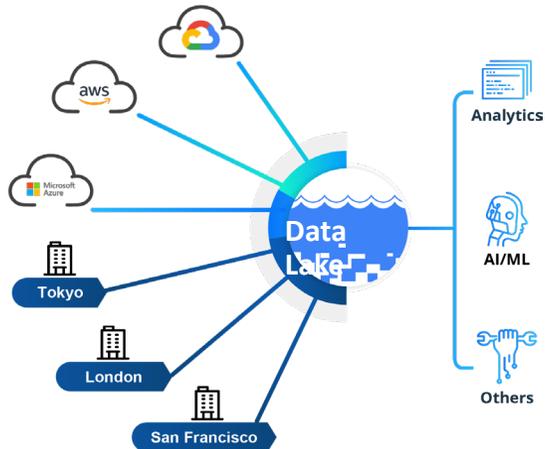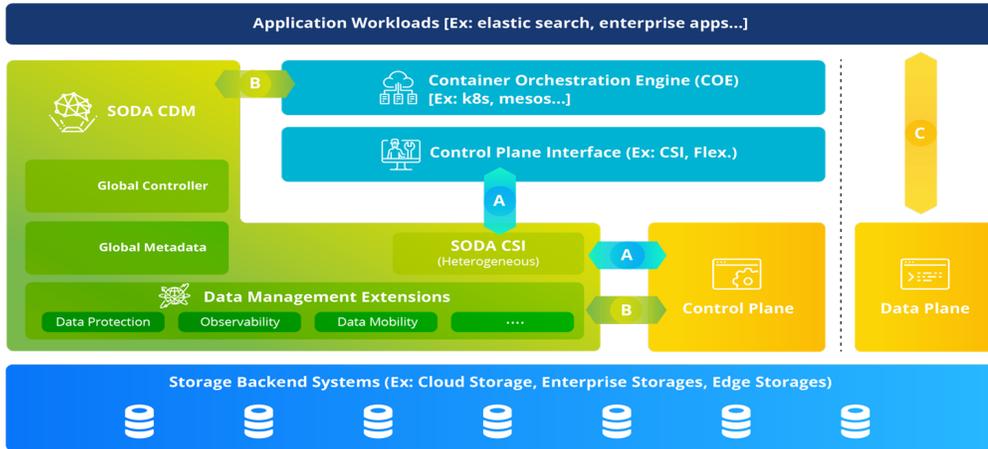
Kahu – Container Data Protection

- Augment Kubernetes (or COE) container data management in a cloud native way
- Container Data Protection
- Heterogeneous and hybrid ready

    https://github.com/soda-cdm

COMO – Multicloud Data Lake

- multicloud virtual data lake solution
- centralized repository with a single common interface
- public or private cloud. interface

   https://github.com/soda-lake

More projects and direction towards cloud native

- SODA Data & Storage Survey 2021 confirms the direction towards cloud native data management
- Research and project proposals for data protection, data observability and data mobility
- Solutions for data management across edge, cloud and code

SODA ECO Projects :
Building Collaborative Solutions for Data & Storage

massive object storage

container storage management

NVM object storage

edge computing management

container attached storage

multi-cloud data controller

mass capacity object storage

multicluster k8s controller

storage benchmarking

cloud-native file and object storage
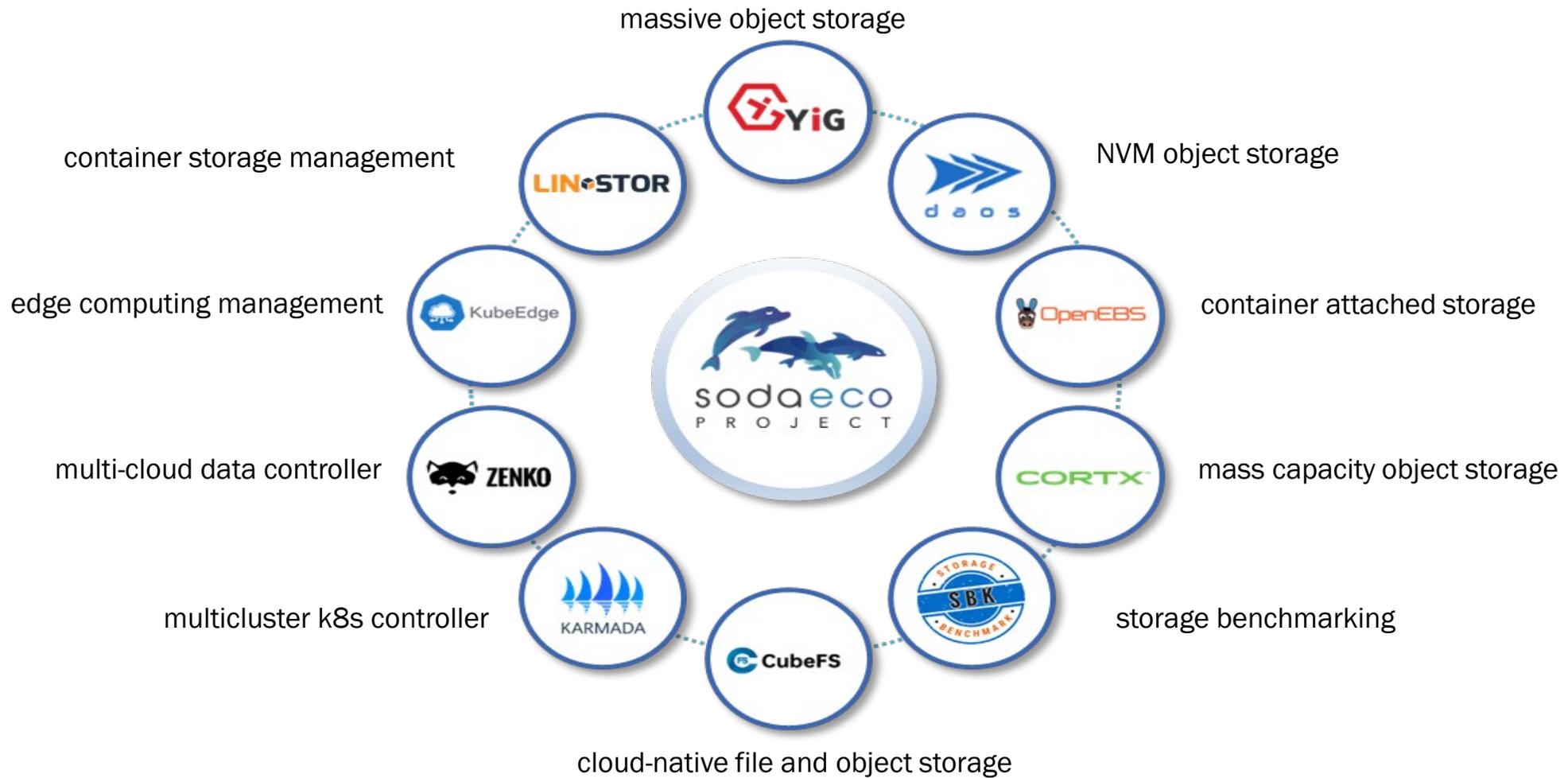
YIG

YIG is a massively scalable object storage project by China Unicom. YIG is deployed in production at China Unicom with about 3-4PB of data

YIG ties together multiple Ceph clusters at its backend to form a massive storage resource pool

Storage can be expanded easily by adding new Ceph clusters without performance degradation due to Ceph rebalancing and other operations
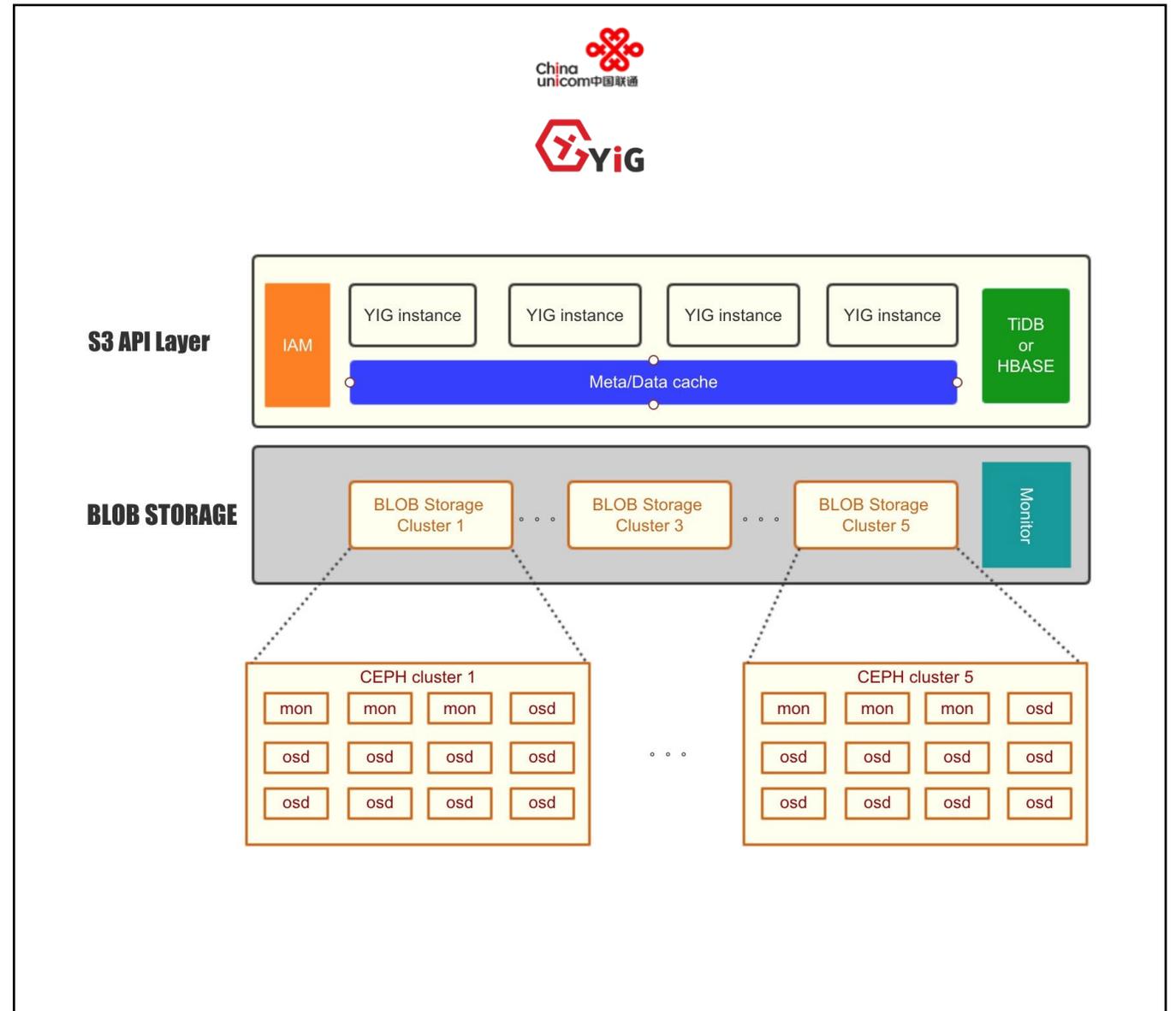
DAOS

DAOS stands for Distributed Asynchronous Object Storage, a project by Intel that was recently incubated into SODA.

DAOS is a software-defined scale-out object store that provides high bandwidth, low latency, and high I/O operations per second (IOPS) storage containers to HPC applications. It enables next-generation data-centric workflows that combine simulation, data analytics and AI.
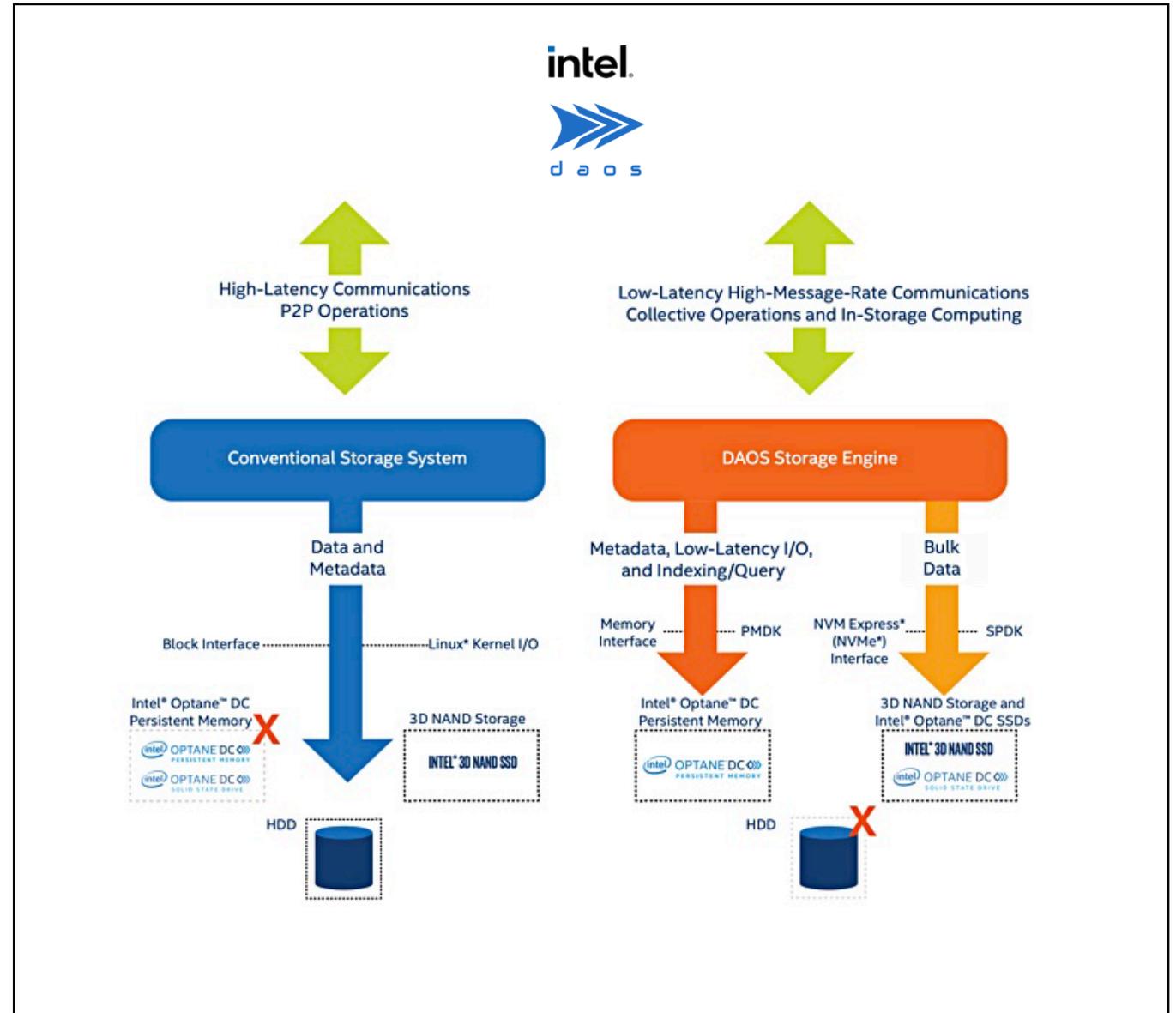
LINSTOR

LINSTOR is a **software defined storage** designed to manage block storage devices for Linux server clusters.

With its large set of features, LINSTOR can manage the stack for you from simple interfaces such as CLI, GUI or API & Drivers.

You can use LINSTOR with Kubernetes, Openshift, Opennebula, Proxmox, Xcp-ng, Openstack, Nomad, Docket etc.

The LINSTOR® system consists of multiple server and client components.
- A LINSTOR controller manages the configuration of the LINSTOR cluster and all its managed storage resources
- The LINSTOR satellite component manages the creation, modification, and deletion of storage resources on each node that provides or uses LINSTOR -managed storage resources
- All communication between LINSTOR components use LINSTOR's network protocol, based on TCP/IP network connections
- Manage the storage system by using a command-line utility to interact with the active LINSTOR controller.
- Alternatively, integrate the LINSTOR system into the storage architecture of other software systems, such as Kubernetes

SODA DELFIN

Heterogeneous Storage Monitoring

STORAGE DEVELOPER CONFERENCE

SODA Delfin – A Unified Platform
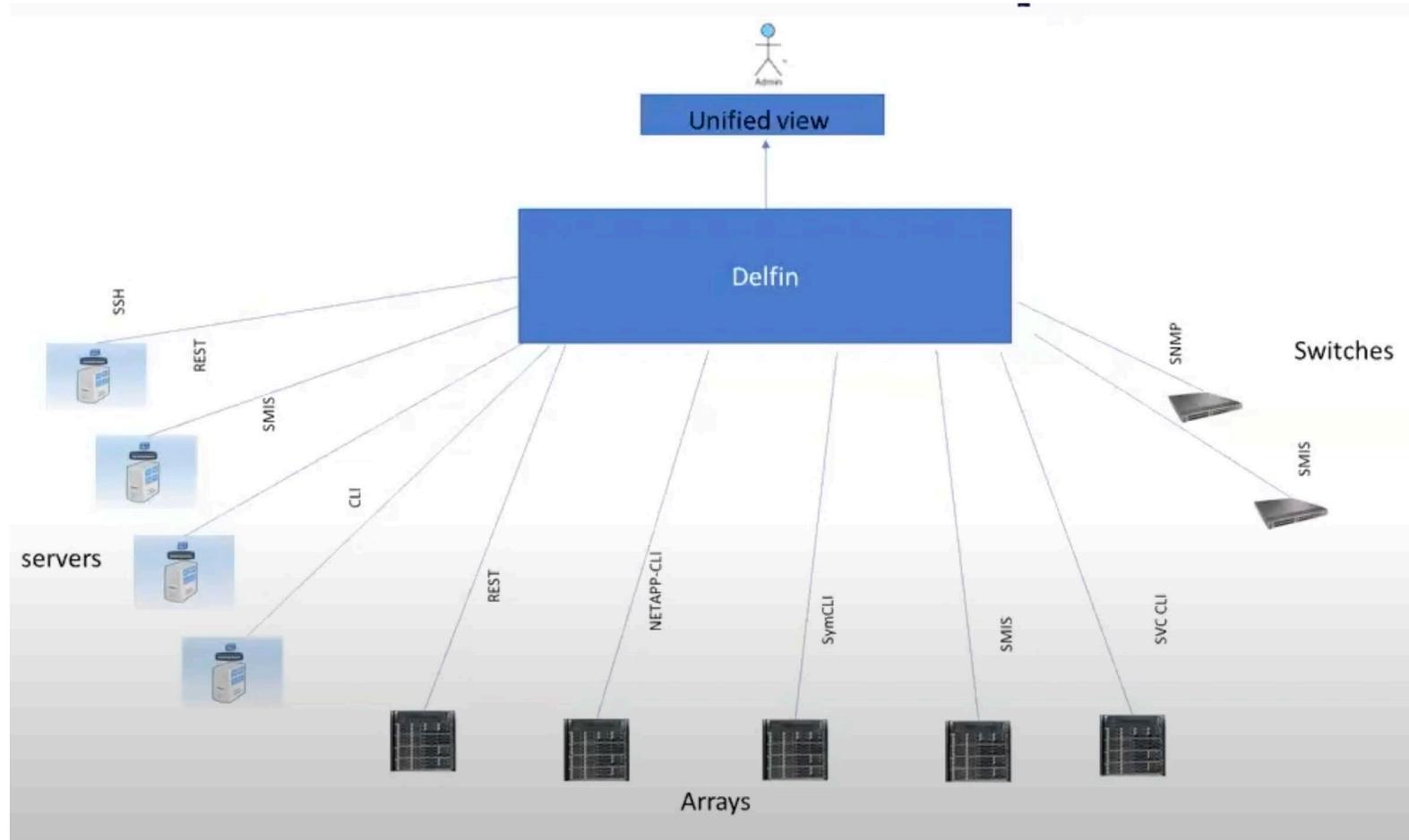
SODA Delfin : Unified Heterogeneous Storage Monitoring Framework

Support heterogeneous storage monitoring with all storage internal data needed.

Resource monitoring
- Storage, Pool, Volume, Port, Network & Controllers
- Filesystem, Share, Quota, Qtree
- Mapping views, topology relations

Performance monitoring
- Bandwith, Throughput, IOPS, Respose time, Capacity

Alert Management
- Alert Notifications and Management

Driver Manager
- Easy support for new storage devices
- Support different vendor management platforms

Plugin Manager
- Easy support for new clients (export monitoring data to Prometheus, Grafana, Kafka or custom clients)
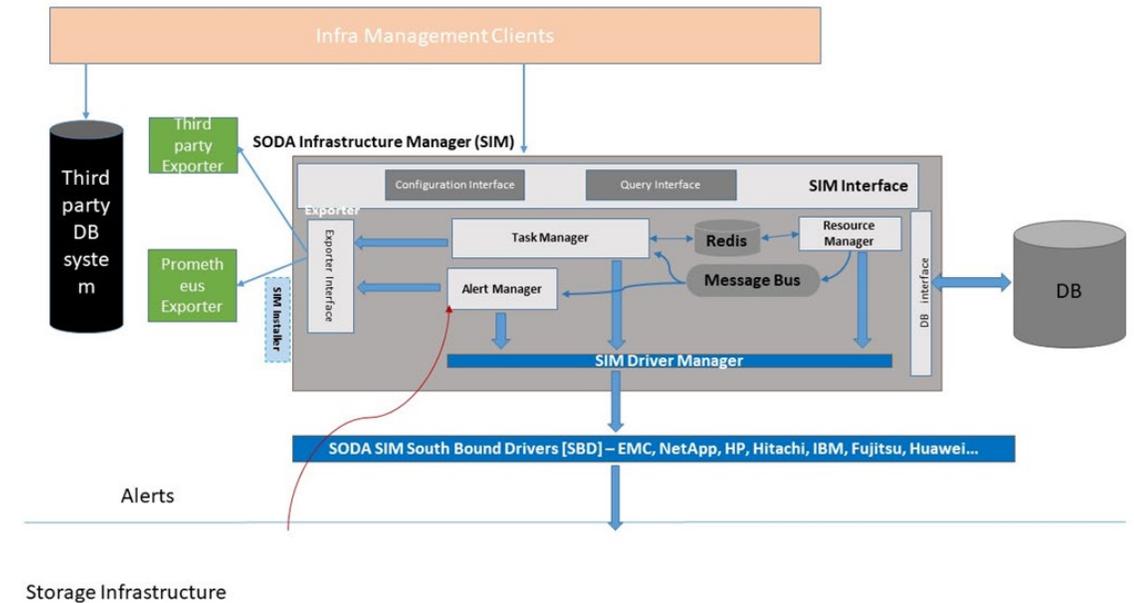
Deployable in container environment
(Trial deployment in Kubernetes Cluster is available in SODA examples repo)

https://github.com/sodafoundation/delfin

STORAGE DEVELOPER CONFERENCE

Delfin Architecture
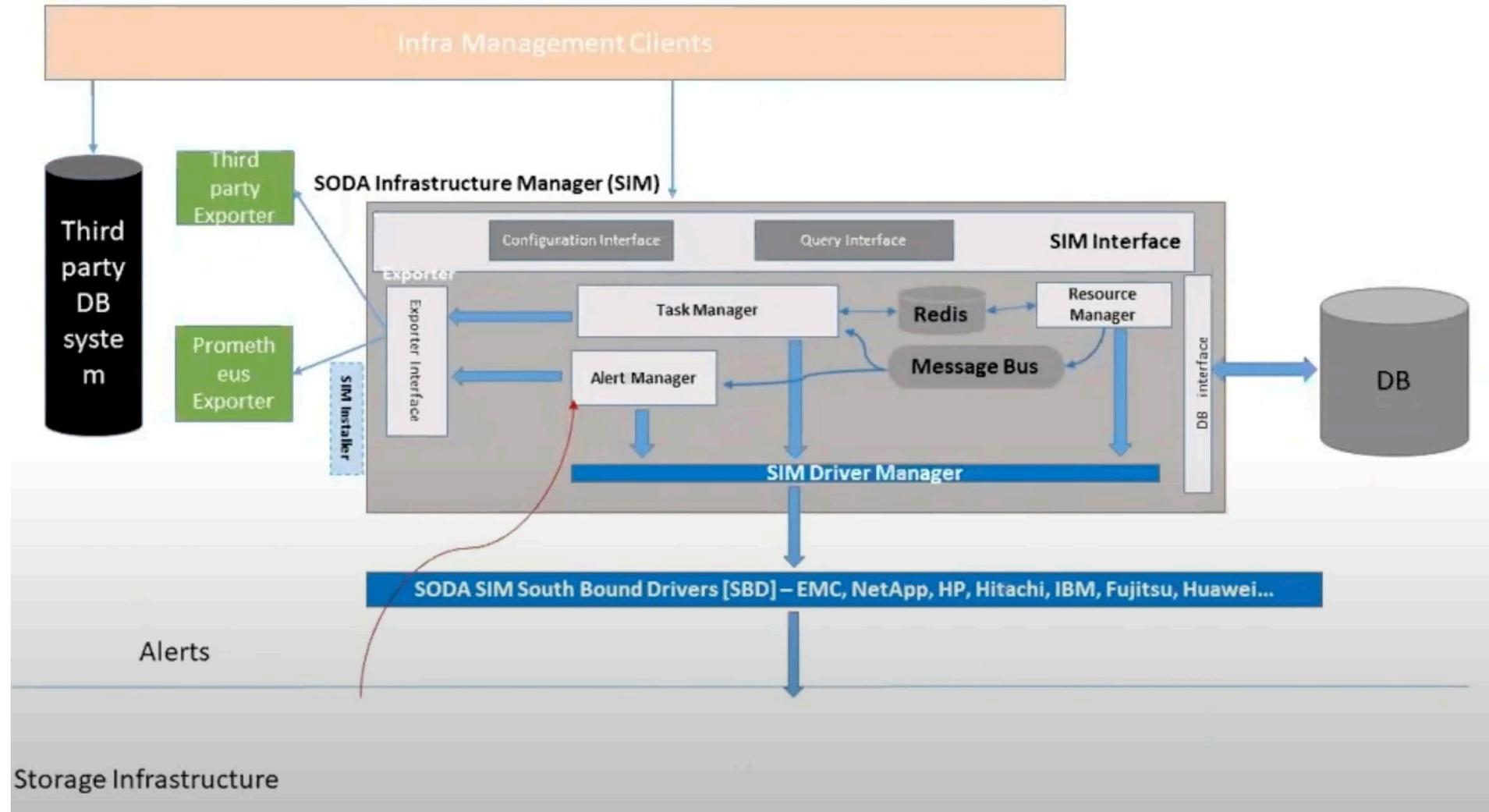
Delfin Architecture

Microservice based scalable architecture

Unified REST API
- Storage register, configure, sync & collect

Resource Manager Service
- Add, update, remove storages resource collection

Task Manager Service
- Periodic collection of performance metrics

Alert Manager Service
- Async SNMP trap handling, alert parsing, exporting

Driver Manager
- Easy support for new storage devices using pluggable driver
- Support for REST, SSH, CLI etc

Exporters Interface
- Export monitoring data to Prometheus, Grafana, Kafka or custom clients

STORAGE DEVELOPER CONFERENCE
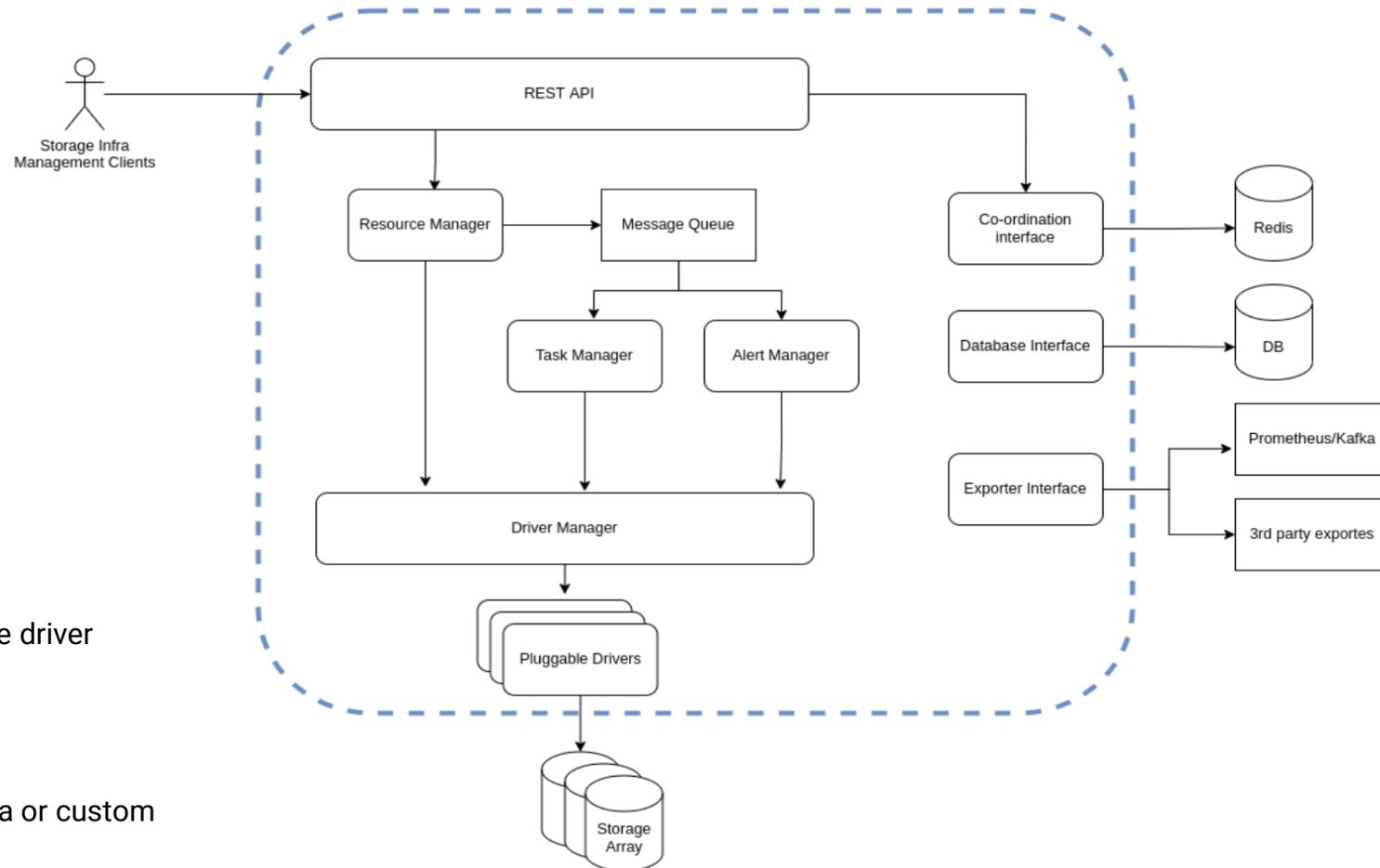
Delfin Summary

- Delfin is Python based heterogeneous storage monitoring framework which supports performance, resource and alerts

- Delfin provides agentless & less intrusive

- Control Plane

- Fully open source, vendor neutral, secure deployments

- Delfin provides pluggable drivers for dynamic storage additions with REST, CLI, SSH, SMI-S support

- Metrics Export and Visualization (Grafana, Custom Dashboard..)

- Exporter Framework (out of the box support for Prometheus) for Kafka, AI/ML integration, visualization or custom management.

- Delfin provides performance scaling with dynamic process creation, depending on the number of storages managed by Delfin.

Delfin Next: Towards Container Storage Monitoring
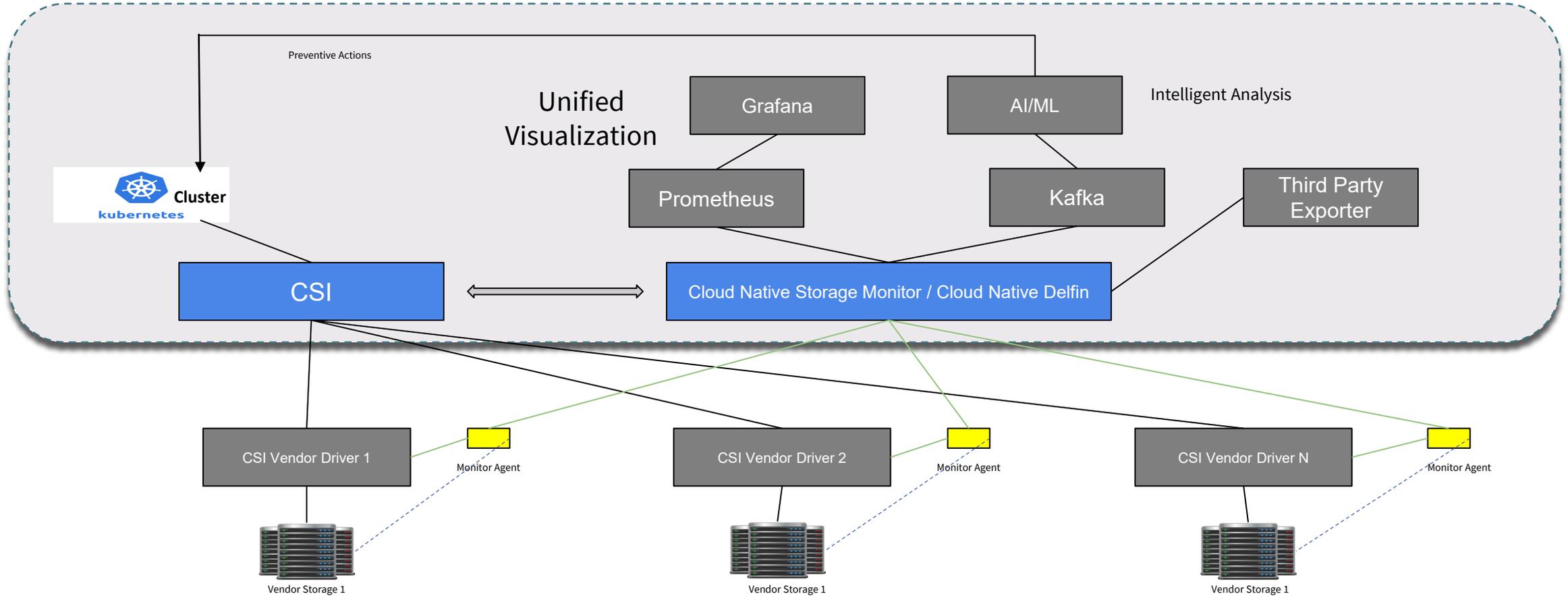
Kubernetes Native Observability

Use Cases:

- Heterogeneous Container Storage Observability
- Unified Visualization and Reporting for CN Storages
- Unified Storage analysis and apply AI/ML
- Predictive analysis and Preventive Actions (live debug and heal)
- Hybrid Observability across cloud, edge and on-premise
- e2e AIOps

Key Highlights:

- Cloud Native Ready
- Microservice Architecture
- Unified across heterogeneous storage vendors
- Multiple CSI Driver Instances
- Exporters for Prometheus, Kafka and third party management platforms
- Agent model for capability extensions

Future potential research areas…

- Federation Support
- Real Time Performance
- Cloud Observability
- Edge Observability
- Distributed Observability
- Tools Market Place

STORAGE DEVELOPER CONFERENCE

SODA COMO

Multicloud Data Lake

Data Virtualization

- Enterprise landscapes are filled with disparate data sources
- In many forms and is stored in many locations
- Multiple data warehouses, data marts, data lakes

How do you get an overview of your far-flung data and manage it in all of its disparate forms?

- <u>Data virtualization</u> integrates data from disparate sources without copying or moving the data, thus giving users a single virtual layer that spans multiple applications, formats, and physical locations. This means faster, easier access to data.

soda
foundation

STORAGE DEVELOPER CONFERENCE
SDC 22

SODA COMO : First SODA Lake Project for Multicloud Data Management

COMO is a multi-cloud virtual data lake providing a centralized repository with a single common interface for data stored in any public or private cloud.

Benefits:

- connect with a single interface

- unified view of data from multiple sources

- minimal data transfers

- enhanced security and governance

- faster integration and deployment

- better performance

- highly versatile and scalable

2

STORAGE DEVELOPER CONFERENCE

SDC 22

COMO Phase 1 : 2022

Customer DC   Service Provider DC   Cloud Object Storage
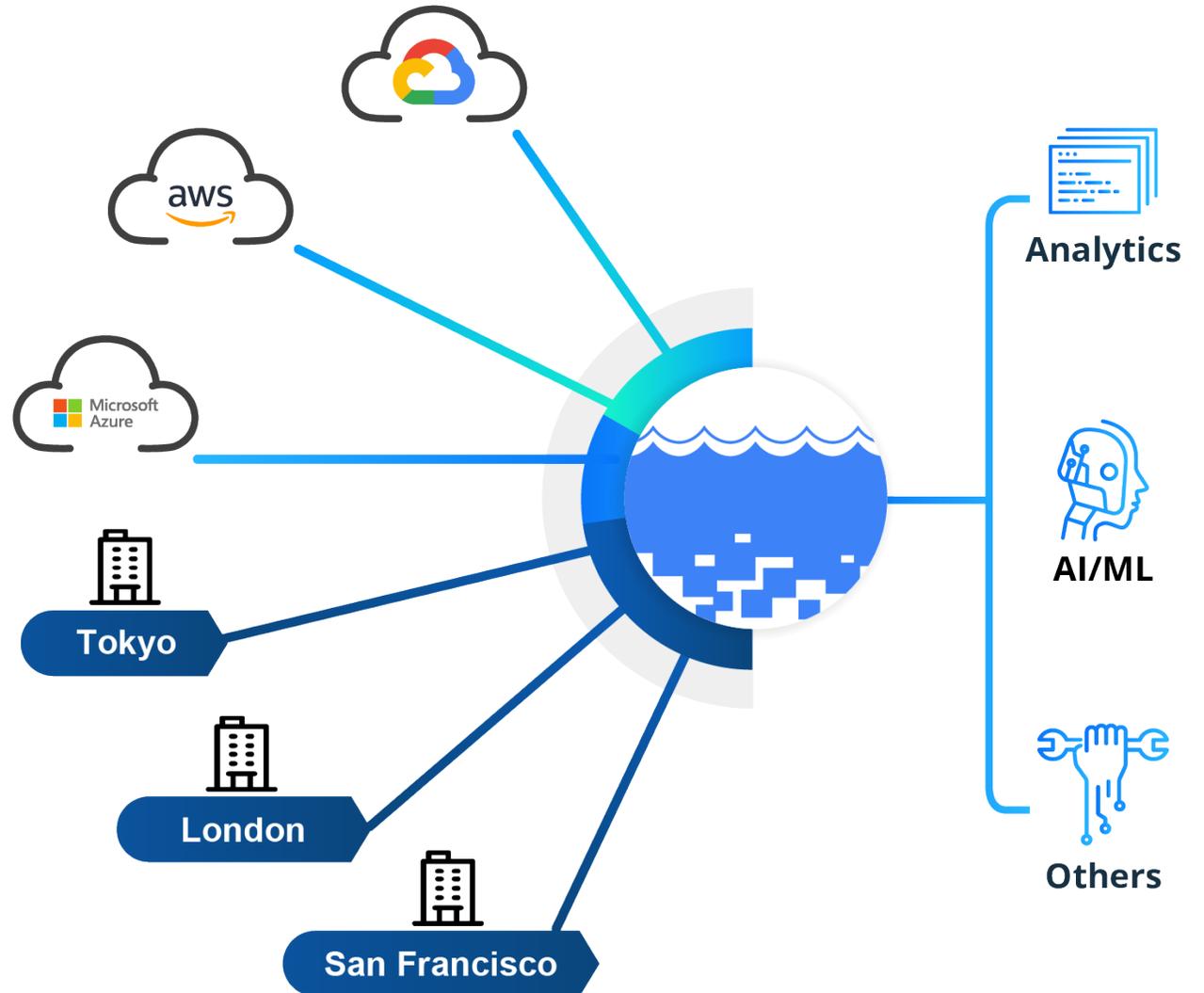
1. **Multicloud Object Data Management**
   - Provisioning
   - Data Access
   - Data Movement

2. **Service Provider Driven Access Management**
   - Storage Service Groups
   - RBAC

Storage Service Groups

AWS

Azure

Ex: Softbank DC

GCP

....

Control Flow   Data Flow

soda foundation

STORAGE DEVELOPER CONFERENCE
SDC 22
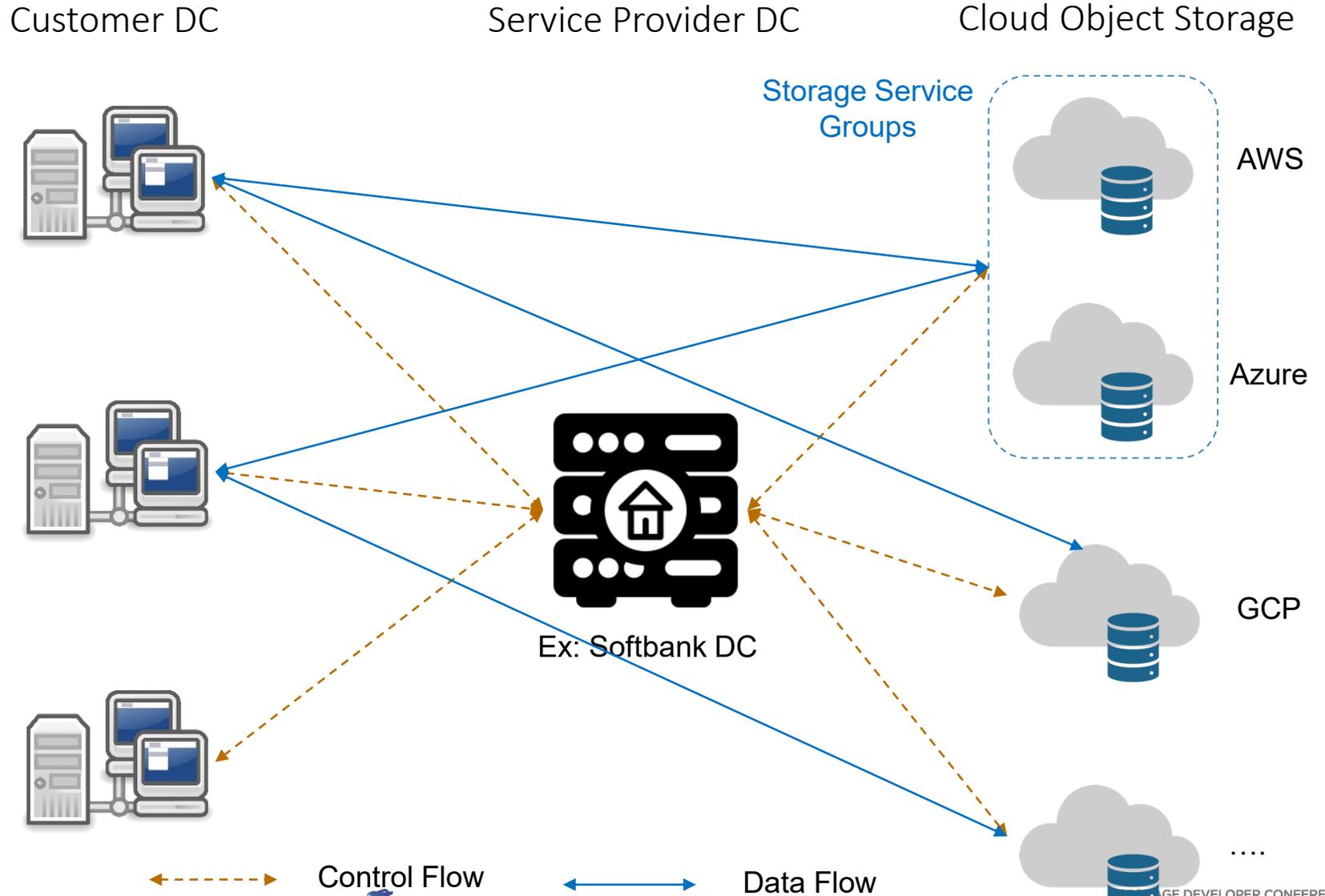
High Level Stack and Architecture Tenets

- Microservice Based
- Distributed
- Scalable
- Plugin and Provider Based extensions
- REST, Open API
- S3 and other backend API compatibility, utilize backend APIs transparently as much as possible

API Gateway ←- - - - - - - - - - - → API Auth

API Layer

Control Services	Core Services	Job Manager
Data Services	Manage Services	Connection Services
Data Mover Services	Data Process / ETL Services	Cache Services

Metadata store

Service Layer

Backend Dock

Backend Providers (AWS, GCP, Ceph…)

Backend Layer

soda foundation

STORAGE DEVELOPER CONFERENCE
SDC 22
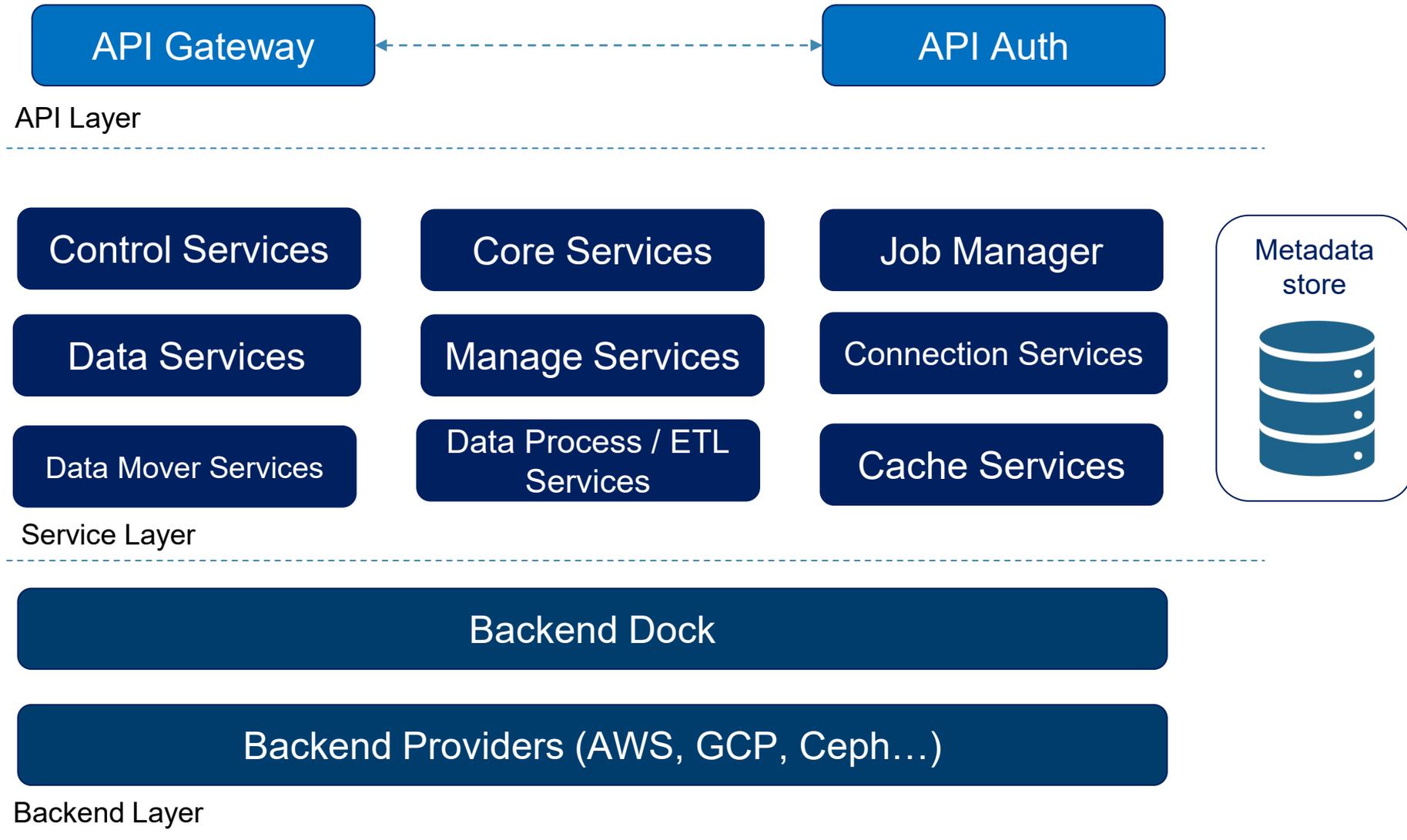
SODA Ecosystem

Growing the Community

ECOSYSTEM

- End users drive roadmap requirements, test & adopt, provide feedback and validation

- Community developers build SODA projects and explore cross-project integration

- Vendors build interoperability solutions around SODA projects

- Industry organizations collaboration for standardization, certification, co-marketing and events

- Academia partnerships for labs, internships and research

End Users
- requirements
- POC's
- testimonials

Developers
- projects
- features
- tests
- documentation

Vendors
- interoperability
- solutions
- partnerships
- vendor customers

Organizations
- standards
- certifications
- co-marketing
- events

Academia
- lab
- internships
- research

STORAGE DEVELOPER CONFERENCE

SDC 22
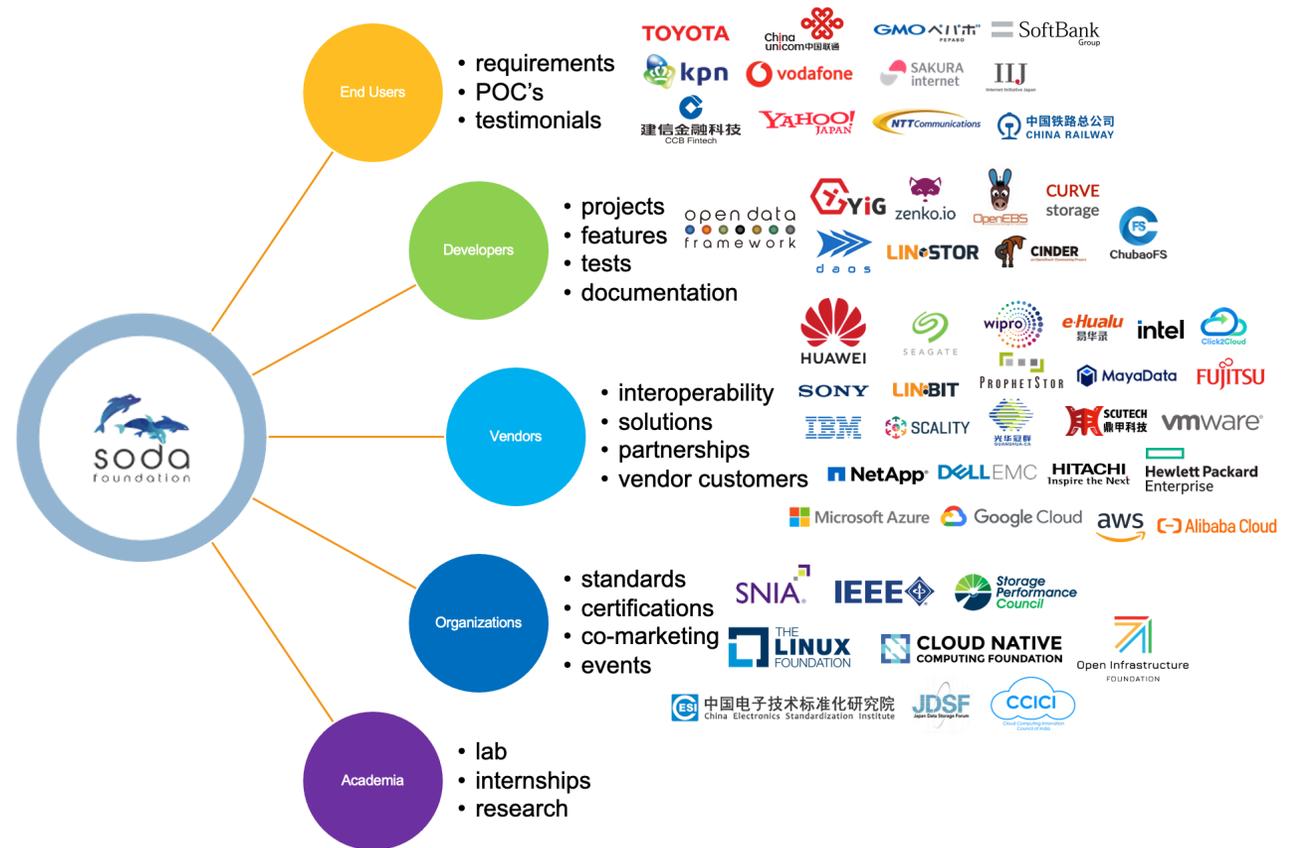
2021 Data and Storage Trends

October 2021

In Partnership With

74% of end-users **run database workloads all the time.**

61% of end-users **rate DPA as the most important capability** over the next three years.

64% of end-users identified **storage capacity as their #1 storage infrastructure challenge.**

55%: The share of total mentions **identifying public cloud services as the foundation** for application development and deployment.

10-100X: The increased amount of **annual data growth** by the top 9% of end-users compared to the majority of mainstream enterprises.

60%: The number of end-users that identify **Software Defined Storage as a key technology** used in their storage infrastructure.

Cloud application storage is the #1 use case for cloud storage.

Backup and recovery is the #2 cloud storage use case reported by **58%** of end-users.

Performance is the #1 cloud native storage pain point as identified by **49%** of end-users.

Hilary Carter

VP Research
The Linux Foundation

We trust that the actionable insights from this effort will help the community to make better decisions, improve products, and establish new technical directions.

STORAGE DEVELOPER CONFERENCE

2021

DATA &

STORAGE

TRENDS

HIGHLIGHTS

COLLABORATING WITH OPEN SOURCE ECOSYSTEMS IS A VENDOR PRIORITY

Figure 12: The Importance of Ecosystems and IT Industry Organizations to Vendors
Rate the importance of the following in your product strategy.

Sample Size = 98
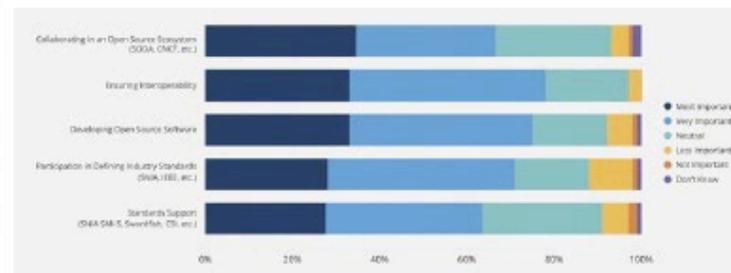
HOW SODA BRINGS VALUE TO YOUR ORGANIZATION

Figure 13: How SODA Brings Value to Your Organization
Rate how SODA can bring value to you and your organization.

Sample Size = 203
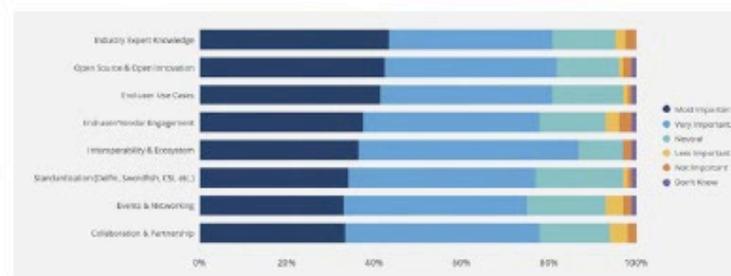
WHERE SODA CAN BRING VALUE TO YOUR ORGANIZATION

Figure 14: Rate the Importance of Potential SODA Development Work
Rate which SODA development work would benefit you.

Sample Size = 203
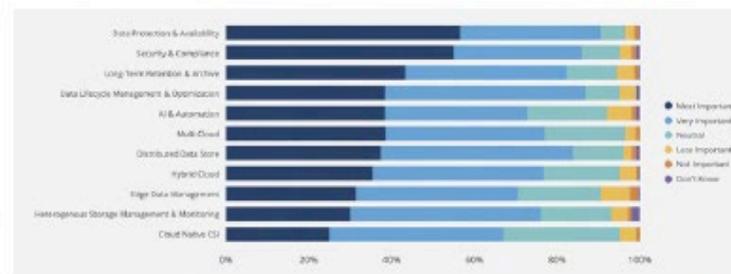
SODA Data and Storage Trends Survey is announced!

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

STORAGE DEVELOPER CONFERENCE

Participating Projects

PAST EVENTS

Open Source Summit, North America 2021

Open Source Summit, Seattle 2021

Kubernetes Forum, Delhi 2020

SODA China Community Launch, Beijing 2019

SODA Forum, Tokyo 2019

Open Source India, Bengaluru 2019

Open Source Summit, Lyon 2019

1st Japan Meetup, Tokyo 2018

OpenSDS Mini Summit, Prague 2017

SODACON
— Global 2021 — July 13-14
Data and Storage Matters
Virtual #sodacon2021

THIS YEAR, WE WELCOMED A RECORD-BREAKING NUMBER OF

SODACON 2021 Event Summary
DATA AND STORAGE MATTERS

1081 registrants
700+ attendees

from 25 countries

We hosted 5 tracks

N. America
Europe
India
China
Japan

9 Keynotes
4 panel discussions

with 54 sessions presented by 85 industry experts from 53 world class organizations

SPONSORS AND PARTNERS

PLATINUM SPONSORS

FUJIFILM   IBM   intel.   TOYOTA   HUAWEI

PREMIER PARTNERS

CLOUD NATIVE COMPUTING FOUNDATION   SNIA.

GOLD SPONSOR

SCALITY

SILVER SPONSORS

博雅云计算 BYCCTEC.COM   XSKY.

EVENT PARTNERS

CCICI Cloud Computing Innovation Council of India

ESI 中国电子技术标准化研究院 China Electronics Standardization Institute

JDSF Japan Data Storage Forum

Kristu Jayanti College AUTONOMOUS Bengaluru

木兰开源社区 MULAN OPEN SOURCE

Open Infrastructure FOUNDATION

Storage Performance Council

SNIA. CMSI | COMPUTE, MEMORY, AND STORAGE

SNIA. SMI | STORAGE MANAGEMENT

MEDIA PARTNERS

electronics FOR YOU EFYGROUP YOURS SINCE 1969

KonfHub

シンクイット ThinkIT thinkit.co.jp

soda foundation

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

STORAGE DEVELOPER CONFERENCE
SDC 22

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

STORAGE DEVELOPER CONFERENCE
SDC 22

Please take a moment to rate this session.

Your feedback is important to us.

soda
foundation

STORAGE DEVELOPER CONFERENCE
SDC 22