Unify Data and Storage Management with SODA ODF

An open source project for data & storage management

Steven Tan, VP & CTO Cloud Solution, Futurewei
Anjaneya ”Reddy” Chagam, Intel
Part 1: Introduction
The Foundation

- SODA Foundation focuses on open source on data and storage management
- Launched Jun 29, 2020 under Linux Foundation
- 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
SODA End Users

SODA end users represent some of the largest and most innovative companies around the world.

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.

END USER ADVISORY COMMITTEE
SODA Data & Storage Trends
2021 Survey
SODA Foundation &
Linux Foundation Research

From April 15 to May 24, 2021, SODA and The Linux Foundation shared the survey to individuals via social media, The Linux Foundation and Linux.com websites, the Linux Foundation Newsletter, and with the support of the following partners:

- Cloud Native Computing Foundation (CNCF)
- Storage Networking Industry Association (SNIA)
- Open Infrastructure Foundation (OIF)
- Japan Data Storage Forum (JDSF)
- China Open Source Cloud League (COSCL)
- Mulan Open Source Community
- Storage Performance Council (SPC)
Open Invitation for Projects & Developers
Email: hackathon@sodafoundation.io
Part 2: The Open Data Framework
Technology Stacks Create Environments That Are Hard To Monitor and Control
Multi-DC, Cloud, and Edge Add To Monitor and Control Challenges

key challenges: capacity - performance - data protection - data compliance - ...
Unify Data And Storage Management With A Single Open Framework Across The Core, Cloud And Edge
ODF For Connected Car Platform

Services For Connected Car Platform

- **block, file, object** block, file, object storage for edge, DC
- **multicloud storage**
- **backup & recovery** backup to cloud or tape
- **snapshots from tier2 to cloud**
- **lifecycle & tiering** edge to DC
- **tier 1 to tier 2, tier2 to cloud, tier2 to cold**
- **security & compliance** edge, DC, & cloud security and compliance
- **offsite tape/archive**
- **retention & archive** tier2 to cold storage (tape/archive disc)
- **tier2 to cloud for long term retention**
- **analytics & intelligence** data integration for analytics and AI/ML applications

60PB of vehicle data goes to the DC each month
~20GB/month/vehicle x 3M vehicles. *source: AECC*
OPEN ARCHITECTURE

Features:

- API integration for platforms and applications
- Seamless plug-in integration with K8S, OpenStack, VMware
- Block, file, object storage services
- Policy-based storage provisioning and data management for protection and lifecycle & tiering
- Storage performance monitoring & visualization
- Container protection with application consistent snapshot to cloud
- Hybrid/multicloud to AWS, GCP, Azure, ...
- Prometheus & Kafka integration
- Plug & play CSI storage
ODF API

ODF API is based on SNIA Swordfish standard, functions are easily extensible.
ODF STORAGE MANAGEMENT

ODF supports CSI, OpenStack Cinder & Manila, and Swordfish based storage
Releases

TOWARDS OPEN DATA AUTONOMY

2021 Focus: Cloud Native Data Protection and Data Optimization

<table>
<thead>
<tr>
<th>Year</th>
<th>Release</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017-2019</td>
<td>OPENSDS (PRE-SODA)</td>
<td>2017H2 Zealand • 2018H1 Aruba • 2018H2 Bali • 2019H1 Capri • 2019H2 Daito • 2020Q1 Elba</td>
</tr>
<tr>
<td>JUL 2020</td>
<td>FAROE V1.0</td>
<td>• Heterogeneous Storage Management • Block/File Multi cloud • CSI Plug &amp; Play experiment</td>
</tr>
<tr>
<td>OCT 2020</td>
<td>GREENLAND V1.1</td>
<td>• Prometheus &amp; Kafka integration • Storage Performance Monitoring (SPM) • Multicloud object and file - AWS, Azure, GCP • CSI Plug &amp; Play • Edge data management • NetApp ONTAP &amp; more</td>
</tr>
<tr>
<td>JAN 2021</td>
<td>HAWAII V1.2</td>
<td>• Performance anomaly detection • Performance visualization with Grafana • Enhanced cloud file shares for AWS, GCP, Azure, Huawei • Enhanced block AWS • More storage support - IBM SVC, HDS VSP, EMC VNX</td>
</tr>
<tr>
<td>APR 2021</td>
<td>ISABLELA V1.3</td>
<td>• Improved storage monitoring • Monitor NAS performance • HA support with multi-cloud • Cold storage • CSI plug-n-play with more drivers • More on-prem and cloud backends</td>
</tr>
<tr>
<td>JUL 2021</td>
<td>JERBA V1.4</td>
<td>• Plug-in any CSI driver • Multiple CSI drivers in K8S • Container data protection (Restic) • Application consistent snapshot to cloud • Multi-cloud storage tiering • Storage performance monitoring with more metrics metrics • Bucket management for all cloud backends</td>
</tr>
</tbody>
</table>

TOWARDS OPEN DATA AUTONOMY

2021 Focus: Cloud Native Data Protection and Data Optimization

<table>
<thead>
<tr>
<th>Open Data Framework</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPENSDS (PRE-SODA)</td>
<td>2017H2 Zealand • 2018H1 Aruba • 2018H2 Bali • 2019H1 Capri • 2019H2 Daito • 2020Q1 Elba</td>
</tr>
<tr>
<td>FAROE V1.0</td>
<td>• Heterogeneous Storage Management • Block/File Multi cloud • CSI Plug &amp; Play experiment</td>
</tr>
<tr>
<td>GREENLAND V1.1</td>
<td>• Prometheus &amp; Kafka integration • Storage Performance Monitoring (SPM) • Multicloud object and file - AWS, Azure, GCP • CSI Plug &amp; Play • Edge data management • NetApp ONTAP &amp; more</td>
</tr>
<tr>
<td>HAWAII V1.2</td>
<td>• Performance anomaly detection • Performance visualization with Grafana • Enhanced cloud file shares for AWS, GCP, Azure, Huawei • Enhanced block AWS • More storage support - IBM SVC, HDS VSP, EMC VNX</td>
</tr>
<tr>
<td>ISABLELA V1.3</td>
<td>• Improved storage monitoring • Monitor NAS performance • HA support with multi-cloud • Cold storage • CSI plug-n-play with more drivers • More on-prem and cloud backends</td>
</tr>
<tr>
<td>JERBA V1.4</td>
<td>• Plug-in any CSI driver • Multiple CSI drivers in K8S • Container data protection (Restic) • Application consistent snapshot to cloud • Multi-cloud storage tiering • Storage performance monitoring with more metrics metrics • Bucket management for all cloud backends</td>
</tr>
</tbody>
</table>
The Open Data Framework

Unify Data And Storage Management With A Single Open Framework Across The Core, Cloud And Edge
Built on Open Source, Open Standard, Open Ecosystem and Open Collaboration
Part 3: ODF Uses

Container and Edge Data Management
SODA ODF for Container Data Management

Augment Kubernetes (or COE) capabilities for heterogeneous and hybrid container data management.

- Unified CSI
- Heterogeneous Ready
- Designed for Container Data Management: Data Protection, Data Observability, Data Mobility and more
- Hybrid Data Management Ready

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.
SODA CSI

VANILLA CSI V/S SODA CSI

- 3 Different PVC requests
- There is no information on pods (other than pod name passed down to storage)
- There is no zone or tenant information passed down.
- Distributed applications using distributed storage. Even with single vendor - multiple storage classes.

- Unified CSI for All
- Any vendor CSI plug and play
- Future ready to enhance for container data management services like Data Protection, Observability and Global Metadata Management
SODA ODF Features for Container Data Management

CURRENT

- CSI Plug and Play: Plug-in support for any CSI driver
- Support multiple concurrent CSI drivers in Kubernetes deployment
- Container data protection framework based on Restic
- Policy-based application-consistent snapshot to cloud

NEXT

- CSI Enhancements
- Data Protection (Snapshot, Backup, Recovery) Enhance
- Observability: Intelligent monitoring

Jerba Release: https://github.com/sodafoundation/soda/releases/tag/v1.4.0
SODA EDGE

Building seamless ODF Data Management capabilities at Edge

- Edge Data Autonomy: autonomous deployment, orchestration and management
- Container Data Management at Edge with Edge compute platforms
- Native to COEs (Kubernetes Focus)
- Low Resource
- Heterogeneous Storage support at Edge
- Enable seamless data management across Edge and Cloud/On Premise

Note:
a) Initial trials with KubeEdge done
Part 4: SODA Projects

Incubated + Eco Projects
The Distributed Asynchronous Object Storage (DAOS) is an open-source object store designed from the ground up for massively distributed Non-Volatile Memory (NVM).

- High throughput and IOPS
- Fine-grained I/O operations with true zero-copy I/O to SCM
- Support for massively distributed NVM storage
- Non-blocking data and metadata operations
- Advanced data placement considering fault domains
- Software-managed redundancy supporting both replication and erasure code with an online rebuild
- End-to-end data integrity
- Dataset snapshot
- And more…
YIG is a massively scalable object developed to support EB level deployments using Ceph clusters on the backend.

- Uses POSIX API
- Easy to use, no SDK integration
- Support broad applications, such as Spark, etc.
- Have high availability
- Have high capacity
With native integration to Kubernetes, LINSTOR® makes building, running, and controlling block storage simple.

- Multi-tier storage: Data can be stored on either HDD, SSD, NVME or PMEM. Live migration is possible between each other.
- Data Dedupe: Data deduplication is one such technology that enables better utilization of both storage devices and network bandwidth.
- Geo Clustering: Possibility to have multiple clusters in different geographical locations
- Ultra Fast Performance: World IOPS record with DRBD
- Wide Platform Support: OpenShift, OpenNebula, OpenStack, Kubernetes, Docker, HyperV, Vmware, Proxmox
- And more...
OpenEBS builds on Kubernetes to enable Stateful applications to easily access Dynamic Local PVs or Replicated PVs.

- Kubernetes native - ease of use and operations. Integrates into the standard cloud native tooling
- Lower footprint. Flexible deployment options. Fastest NVMe Replicated Storage.
- Controlled and predictable blast radius. Easy to visualize the location of the data of an application or volume
- Horizontally scalable. Scale up and/or down
- Highly composable. Choice of data engines matching the node capabilities and storage requirements
- Open Source and Avoid vendor lock-in
- And more...
Zenko is open-source infrastructure software for DevOps, storage and data managers to view and control data in multi-cloud IT environments.

- Single API (Amazon S3) data access to any storage location or cloud
- Global multi-cloud namespace
- Data remains in format of each storage system or cloud (open, readable, non-proprietary)
- Multi-cloud data management through lifecycle & replication policies
- Extensible metadata and search across clouds
- Zenko includes open-source Cloudserver (S3 endpoint service) and Backbeat workflow service (asynchronous processing engine) projects
- And more...
CORTX is an opensource distributed object storage system designed for great efficiency, massive capacity, and high HDD-utilization.

Object storage uniquely optimized for mass capacity storage devices
Works with any processor.
Highly flexible, works with HDD, SSD, and NVM
Massively Scalable. Scales up to a billion billion billion billion exabytes ($2^{206}$) and 1.3 billion billion billion billion objects ($2^{120}$) with unlimited object sizes.
Rapidly Responsive. Quickly retrieves data regardless of the scale using a novel Key-Value System that ensures low search latency across massive data sets.
And more.
Part 5: Summary
Key Takeaways

- SODA Open Data Framework unifies data & storage management for cloud native, the edge and more
- SODA Foundation helps data and storage projects to integrate and grow
WHY ORGANIZATIONS JOIN SODA

Vendors, end users, and other organizations join SODA for these key benefits:

- **Open Innovation**: accelerate development and bring value to organizations through open innovation in the SODA ecosystem.

- **Feature Request**: request features to be on the roadmap through the TOC or EUAC and the community developers will work on them.

- **POC Solution**: opportunities to participate in SODA proof-of-concept (POC) solution testing where vendors and end users work together closely.

- **Community Engagement**: engage with developers, vendors and end users in our meetings, meetups, and conferences.

- **Brand Recognition**: bring awareness to your organization, project, and things that matter to the SODA community and our partners’ communities.

- **Thought Leadership**: participate in our committees (TOC, AC, OC, EUAC) and workgroups to drive SODA technical direction and other activities.

- **Speaking Opportunities**: opportunities to speak at our meetups and SODACON’s virtual and worldwide.

- **Press Release**: press release announcement when joining SODA and possibilities of mentions or quotes in other SODA press releases.
Thank You

https://sodafoundation.io

https://github.com/sodafoundation