

SODA Data Management for Multi-Cloud Made Easy



ONE DATA FRAMEWORK, INFINITE POSSIBILITIES

an open source data framework enabling data mobility, data protection, data security, data lifecycle, and more

Archana Pathak – Click2Cloud
Roshan Wanaskar - Click2Cloud
Himanshu Varshney - Huawei Technologies India

SODA Data Framework: Introduction

- OpenSDS project an open source project under Linux foundation was started towards the end of 2016 by a group of companies joined to resolve key data management pain-points from storage vendors and end users.
- SODA Foundation is a sub foundation under Linux Foundation started in early 2019 that aims to establish an open, unified, and autonomous data management framework for data mobility from the edge, to core, to cloud.
- SODA is SODA Open Data Autonomy. It is an open source unified autonomous data framework for data mobility from edge to core to cloud.
- SODA Foundation offers a neutral forum for cross-projects collaboration and integration and provides end users quality end-to-end solutions.
- SODA brings together industry leaders to collaborate on building a common framework to promote standardization and best practices for data storage, data protection, data governance, data analytics, etc. to support IoT, big data, machine learning, and other applications.
- We are fostering collaboration and innovation across vendors, system integrators, cloud service providers, standards organizations, and consortiums across different industries, to provide quality end-to-end solutions to end users.



SODA Data Framework: Vision and Goals

- Currently, there are multiple challenges and pain points due to siloed data/storage services. This project's vision is to connect all the siloed data solutions to build a self-governed and intelligent data platform which is open, autonomous, ready and real.
- **SODA Holds Open Standard And Collaboration**
 - ✓ No vendor lock-in
 - ✓ Not under the control of a single or small group of companies
 - ✓ Accelerate development with shared components
- **SODA Is Autonomous Self-Governed And Intelligent Collaboration**
 - ✓ Siloed data solutions are connected.
 - ✓ Automated data service orchestration
 - ✓ Intelligence integrated

SODA Data Framework: Vision and Goals

- **SODA is READY having OpenSDS Certified Suppliers**

- ✓ Standard specification for products
- ✓ Compliance and Certification for Vendors
- ✓ Compliance Lab for seamless interoperability

- **SODA Is Real ,Designed For Real Use-Cases**

- ✓ Solve common end-user pain points
- ✓ Optimize storage operations and utilization
- ✓ Deploy in traditional or cloud-native environments

SODA Data Framework: Focus Areas

- SODA Foundation focuses to build unified frameworks, APIs and solutions in the areas of
 - ✓ Data Mobility
 - ✓ Data Protection
 - ✓ Data Lifecycle
 - ✓ Unified Storage Platform
 - ✓ Cloud Native Storage
 - ✓ Data Governance
 - ✓ Data Orchestration
 - ✓ Data Energy and more.
 - ✓ It envisions to provide data autonomy through its open source solutions and standards.
 - ✓ Solve common end-user pain points

SODA Data Framework: Project Governance

- The project is currently driven by TSC(Technical Steering Committee), EUAC(End User Advisory Committee) and open source community. It comprises of various organizations and experts from the industry
- In early 2019 , OPENSDDS got upgraded to SODA Foundation.

SODA Data Framework: Projects

The SODA Ecosystem has many projects under its umbrella, which work in unison to solve the various data and storage challenges. Some of the important ones are :

- **SODA Dashboard**

- ✓ SODA Dashboard provides a front end UI which integrates with the different APIs provided by SODA API. This dashboard can be used to test basic SODA functionality

- **SODA API**

- ✓ The key external interface to platforms, which can do a seamless integration with heterogeneous storage backends. Provides the standardization for Data / Storage Management APIs.

- **SODA Controller**

- ✓ In the API flow, controller plays a critical role for all the API flow management and tracking to handle all the state machine and metadata management requirements.

- **SODA Dock**

- ✓ It is a docking station for heterogeneous storage backends! This is where all the different storage vendors' drivers for various backends get attached.

SODA Data Framework: Projects

- **SODA Plugin**

- ✓ SODA North-Bound Plugin Project focuses to extend all the industry platforms and application solutions to interface with SODA API or be compliant with it.

- **SODA Multi-Cloud**

- ✓ SODA Multi-Cloud project provides a cloud vendor agnostic data management for hybrid cloud, cross-cloud or in-cloud. It can be hosted on prem or cloud native.

- **SODA Telemetry**

- ✓ Provides monitoring via standard interfaces to pull the statistical and health data from the infrastructure (including storage, host, ports) and a highly configurable alert management system

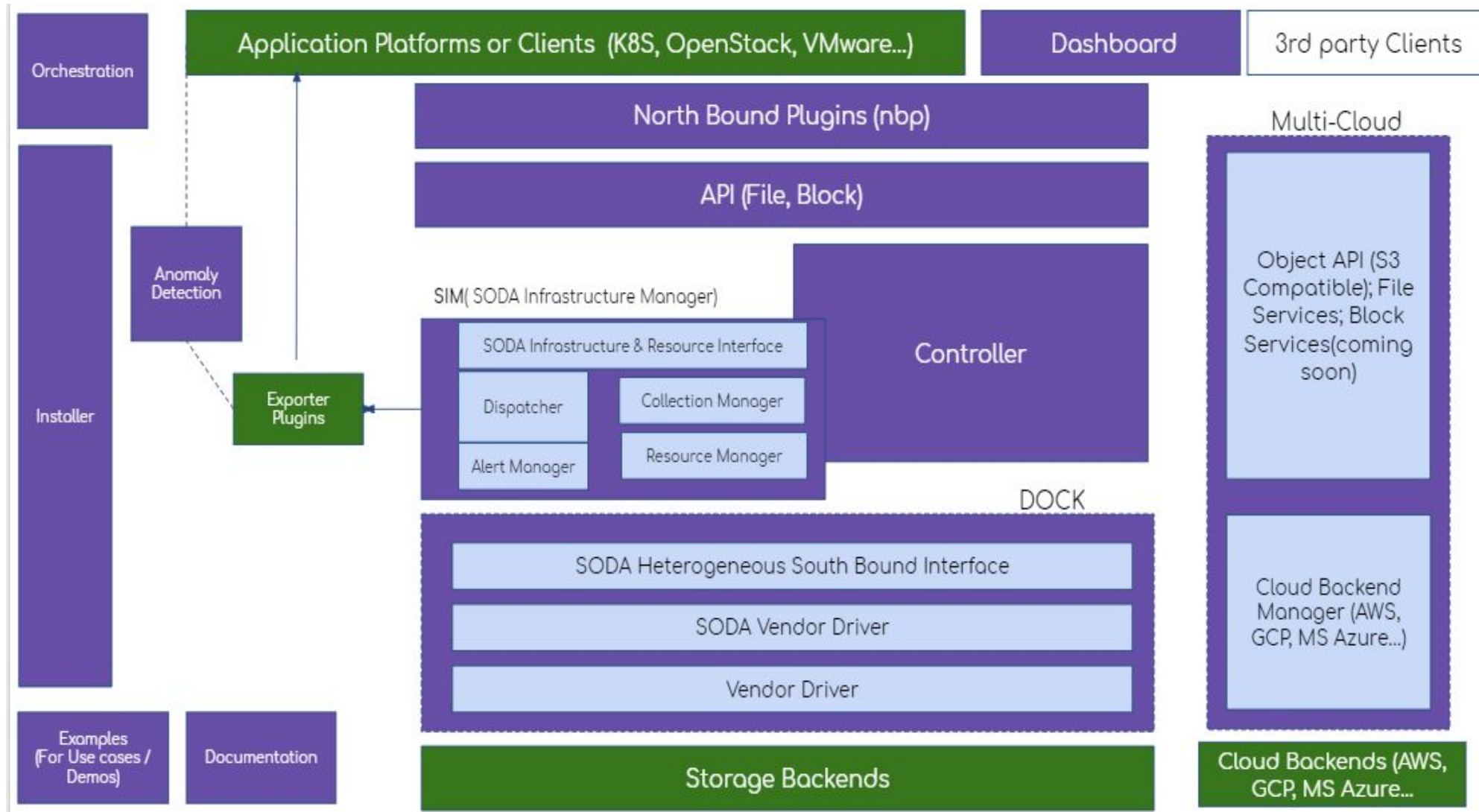
- **SODA Orchestration**

- ✓ The Orchestration framework provides flexibility to use existing workflows or define customized workflows to get the simplified execution of tasks.

- **SODA Intelligence**

- ✓ It is a reference implementation of intelligent monitoring which can utilize SODA Telemetry, exporter with Kafka (or similar) and a standard ML algorithm.

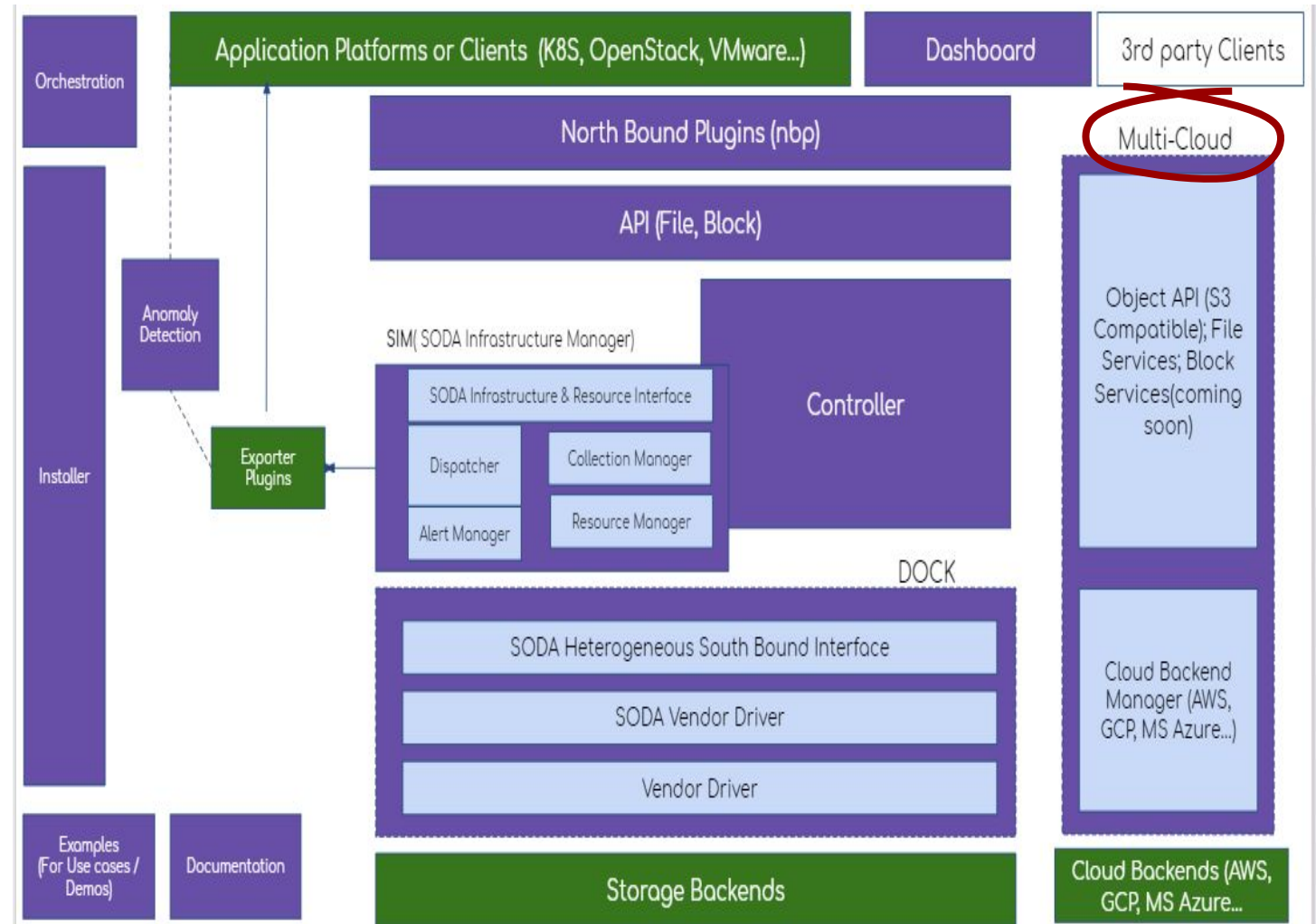
SODA Data Framework: Projects



(github.com/sodafoundation)

SODA Multi-Cloud

- Provides a cloud vendor agnostic data management for hybrid cloud, intercloud or intracloud.
- Supports S3 compatible backend manager which is to connect with any cloud vendors.
- Supports various cloud backend like MS Azure, GCP, AWS, Huawei, IBM and more.
- Integrated YIG-Ceph backend from China Unicom YIG project
- Provides policy-based data mobility across public and private clouds
- Provides Object, File and Block services from the cloud vendors.



(github.com/sodafoundation/multi-cloud)



SODA Multi-Cloud: Features

- **Supported Cloud Vendors/Storage**
 - ✓ **Object Storage:** AWS S3, Azure Blob, GCP, IBM, Huawei, Alibaba, Ceph and YIG
 - ✓ **Block Storage:** AWS EBS
 - ✓ **File Storage:** AWS EFS, Azure File Share and GCP Cloud Filestore
- **Object Operation and management**
 - ✓ Upload Object
 - ✓ Upload Multi-part Object
 - ✓ Download Object
 - ✓ Delete Object
 - ✓ Copy Object
- **Migration feature**
 - ✓ In-Cloud Migration
 - ✓ Cross-cloud Migration
 - ✓ Scheduled Migration

SODA Multi-Cloud: Features

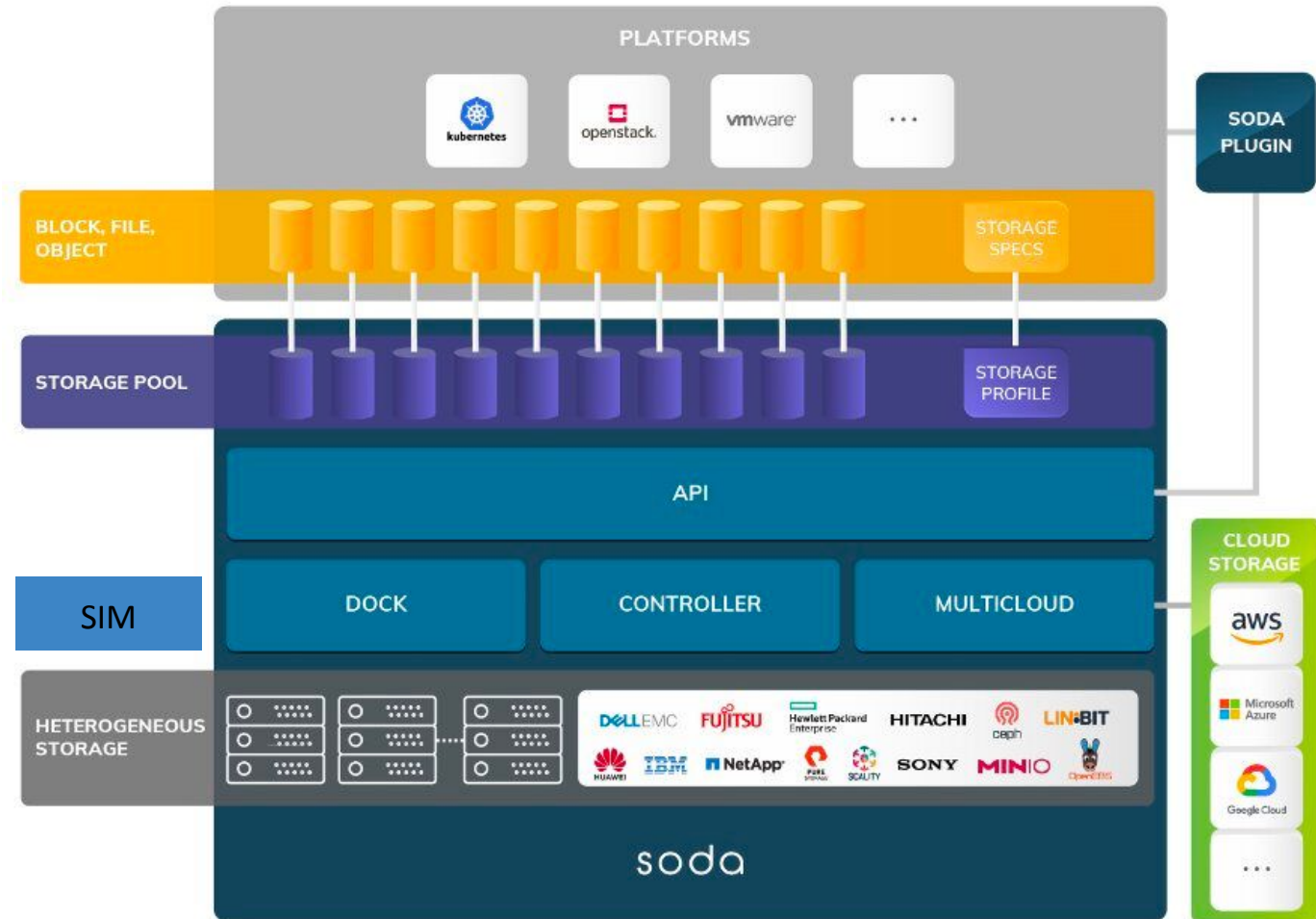
- **Object Lifecycle Management**
 - ✓ In-cloud Lifecycle transition
 - ✓ Cross-cloud Lifecycle transition
- **Object Encryption:** SSE encryption supported (AES algorithm)
- **Object Versioning:** Multiple versions of an object can be stored and managed
- **Object & Bucket ACL:** SODA access control lists(ACLs) enables user to manage access to buckets and objects.
- **Block Storage(Cloud)**
 - ✓ Manage Cloud Volumes(Create, List, Update and Delete Operation)
- **File Storage(Cloud)**
 - ✓ Creation/Deletion/Updation of File shares

SODA Data Framework: Architecture

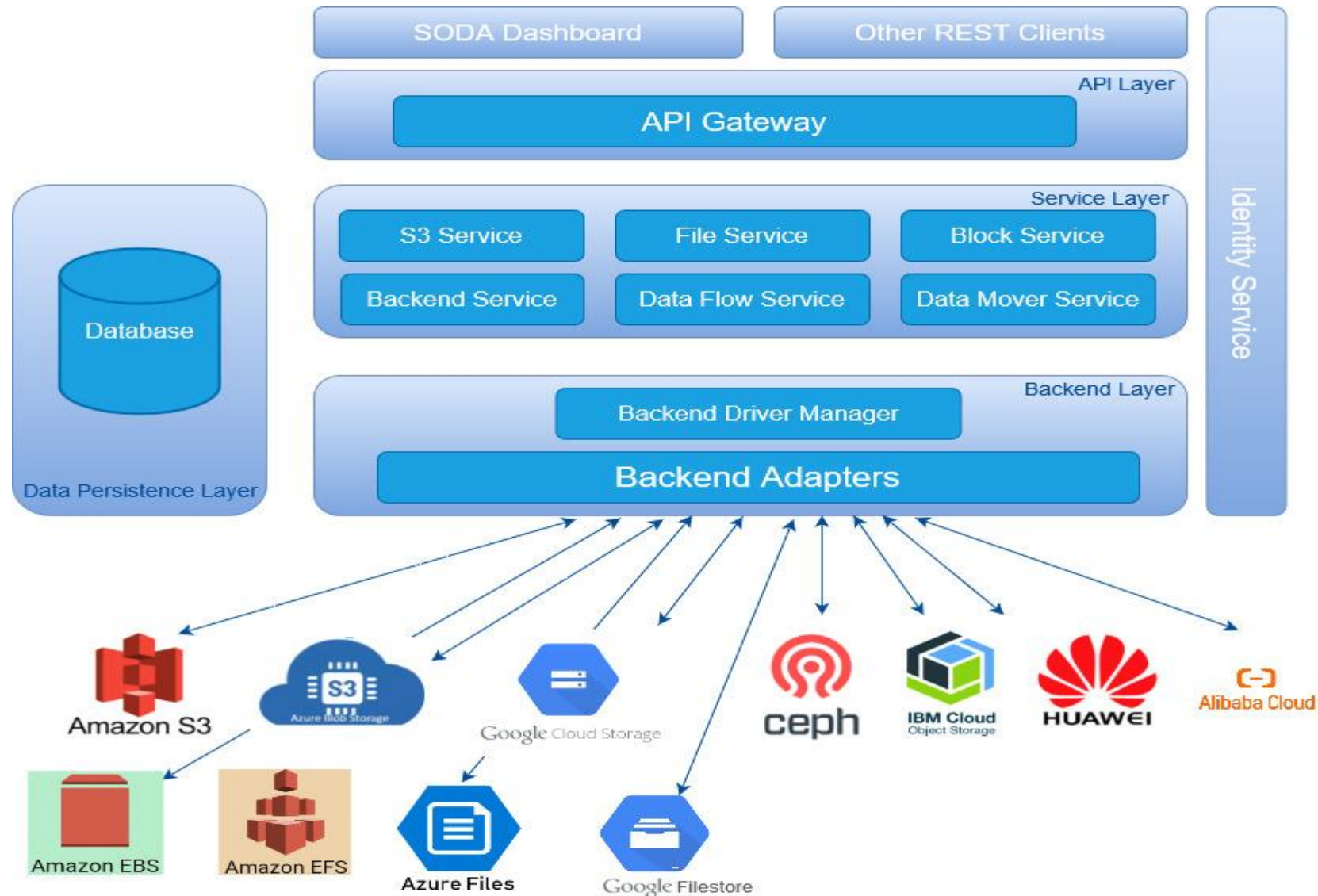
SODA Core Projects - API, Dock, Controller, Multi-cloud, Plugins, and SODA Infrastructure Manager, form the core of SODA framework architecture.

Key Features:

- Policy-based storage provisioning with profiles
- Dynamic Kubernetes persistent storage with SODA CSI driver
- Integration with VMware and OpenStack
- S3 multi-cloud data control for AWS, Azure, Google Cloud and local Ceph S3 gateway
- Heterogeneous vendor storage support including OpenStack Cinder/Manila, and CSI drivers support in future
- Telemetry, automation, anomaly detection, and more



SODA Multi-Cloud: Architecture



SODA Multi-Cloud: Architecture

SODA Multi-Cloud Data Management platform includes the following components:

Identity Service: Provides Authentication and Authorization using Keystone.

Backend Service: Manages storage backends. A backend in SODA multi-cloud is corresponding to a Cloud Backend profile metadata.

S3 Service: Includes S3 controller S3 API requests/responses handling, communicates with database and object storage backends on premise or in the cloud, and returns results to the caller.

Dataflow Service: In-charge of data migration.

Data Mover: Responsible for moving the data from one location to another.

Database: Data Persistence for SODA multi-cloud resources' metadata.

Backend Adapters: Communicates with actual cloud backend vendors such as AWS, Azure, GCP etc. to support backend management.

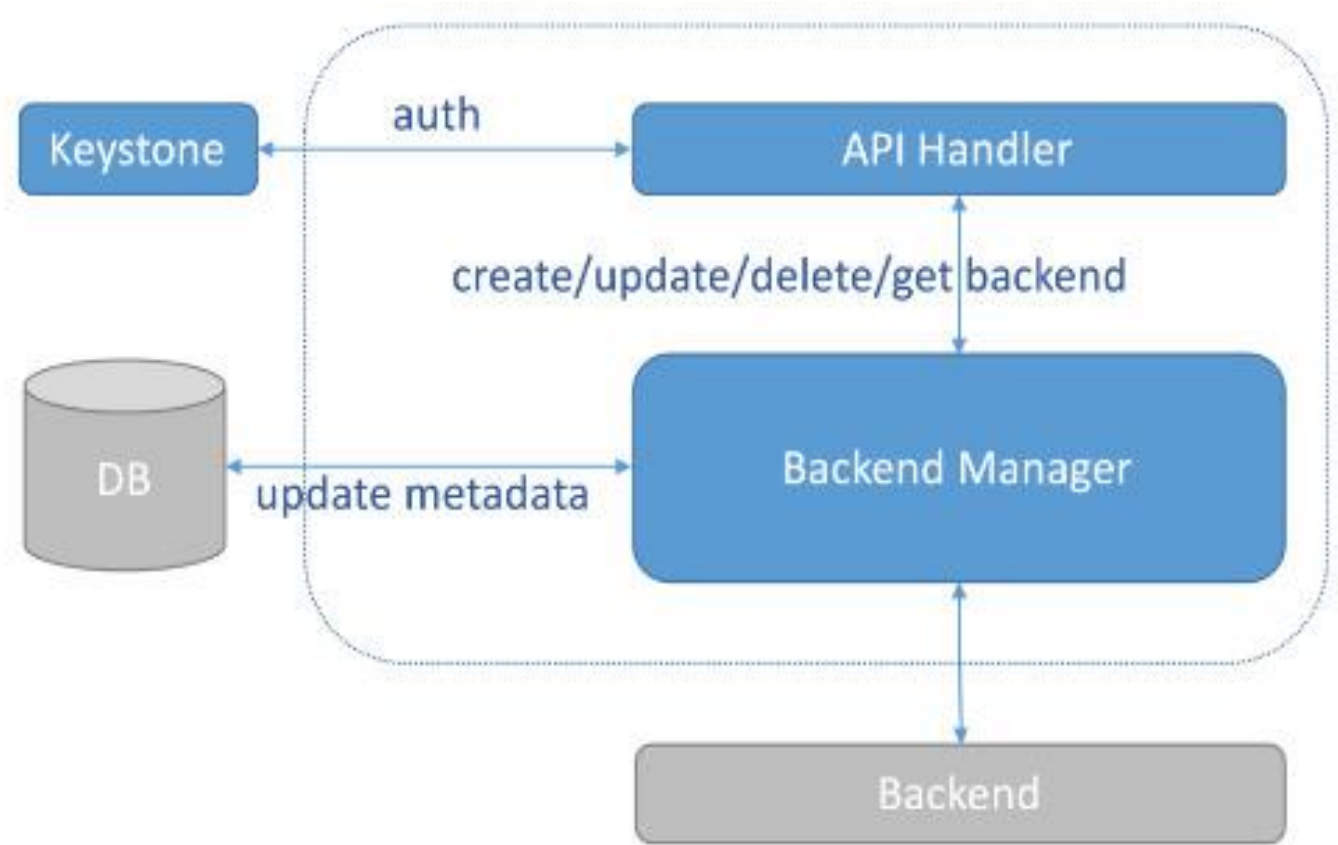
File Service: Includes File controller Fileshare API requests/responses handling, communicates with database and file storage backends in the cloud, and returns results to the caller.

Block Service: Includes Block controller Volume API requests/responses handling, communicates with database and block storage backends in the cloud, and returns results to the caller.



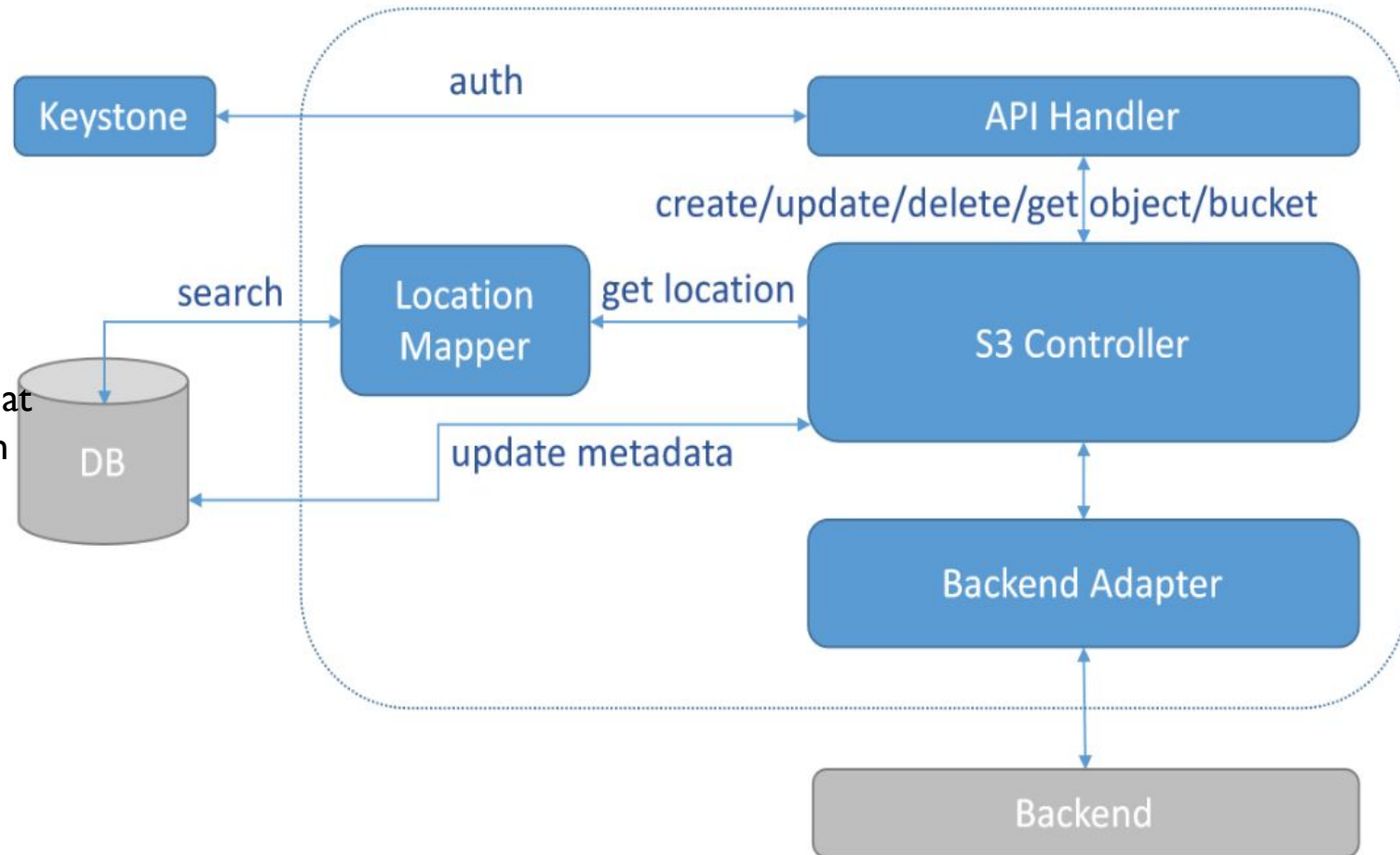
SODA Multi-Cloud: Backend Service

- Support Authentication with Keystone.
- Backend Manager serves as a metadata placeholder for Cloud Backend profiles.
- Backend Service Manager support Create/Update/Delete/Get a backend.
- Supports object/blob, block, file storage types
- Support Cloud Backend Profile Registration for Cloud Vendors including AWS, Azure, GCP and many more.



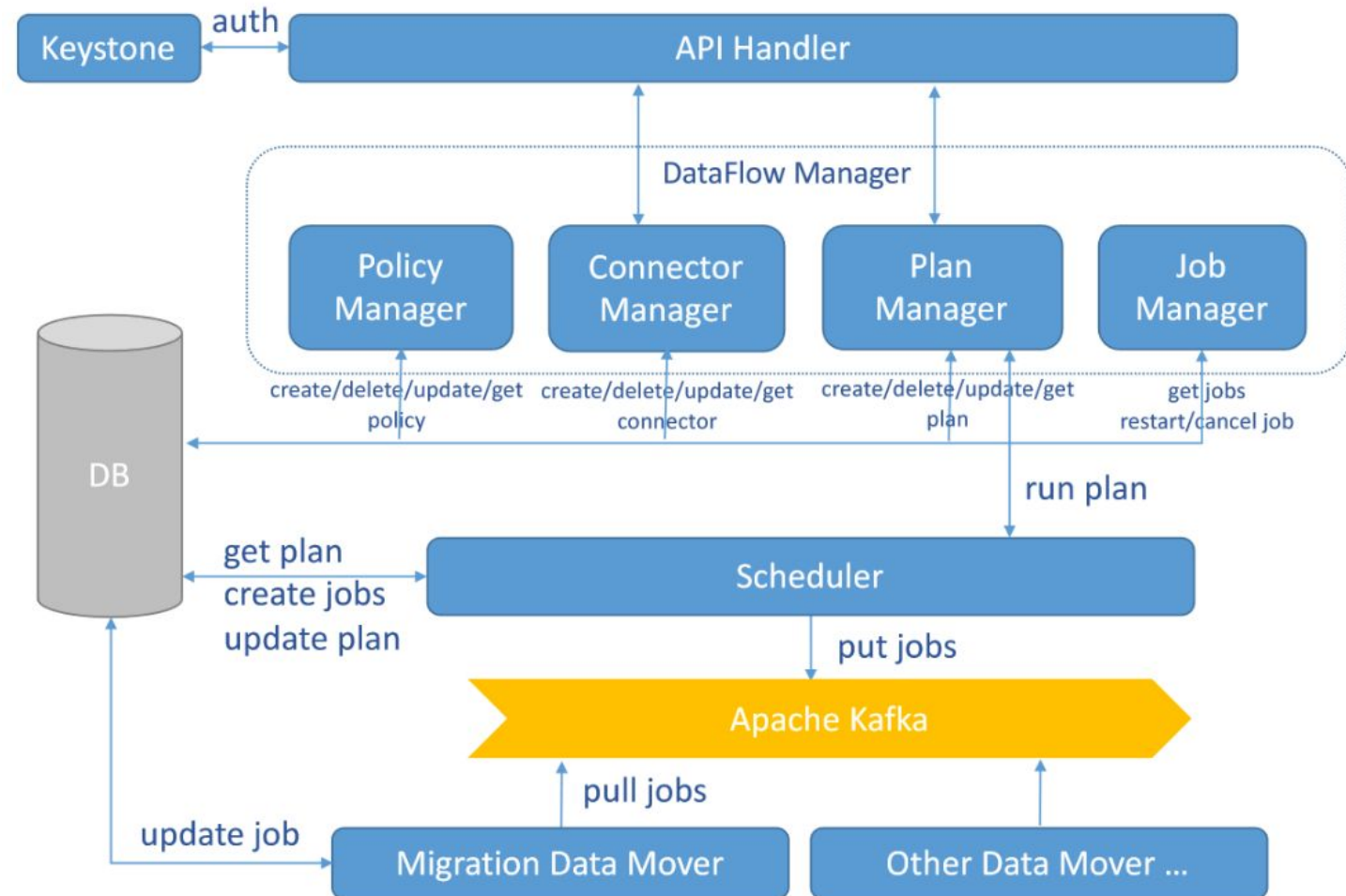
SODA Multi-cloud: S3 Service

- Support Authentication with Keystone.
- Supports S3 Compatible APIs.
- S3 Controller serves as a bucket & object .metadata placeholder.
- A bucket in SODA Multi-Cloud is a virtual bucket that is mapped to a physical bucket in the object store on premise or in the cloud.
- Location Mapper is responsible for searching the database to find the physical location of an object.
- S3 Service support Create/ Update/Delete/Get a virtual buckets.
- S3 Service support Upload/Download Objects.



SODA Multi-Cloud: Dataflow Service

- Support Authentication with Keystone.
- Dataflow Manager serves as a Migration Plan, Policy, Job & Connectors metadata placeholder.
- Policy Manager support Create/ Update/Delete/Get a Migration Policy.
- Plan Manager support Create/ Update/Delete/Get a Migration Plan.
- Job Manager support Create/ Update/Delete/Get a Migration Job.
- Scheduler supports Scheduling of Migration Jobs based on the policies.
- Migration Job execution to Data Mover through the Apache Kafka, a distributed streaming platform.
- Data Mover has different plugins to handle different migration tasks.



SODA Multi-Cloud: Dataflow Service

- Dataflow Service includes:
 - Dataflow Manager
 - Scheduler
- The Dataflow Manager includes:
 - Policy Manager
 - Connector Manager
 - Plan Manager
 - Job Manager

Dataflow Manager: Responsible to make plans for data movement such as Migration.

Scheduler: Executes data movement tasks based on plans made by the Dataflow Manager.

Data Mover: Responsible for moving the data across the physical buckets.

Policy Manager: Responsible for policy definitions which includes a schedule that describes when and how often a migration will happen.

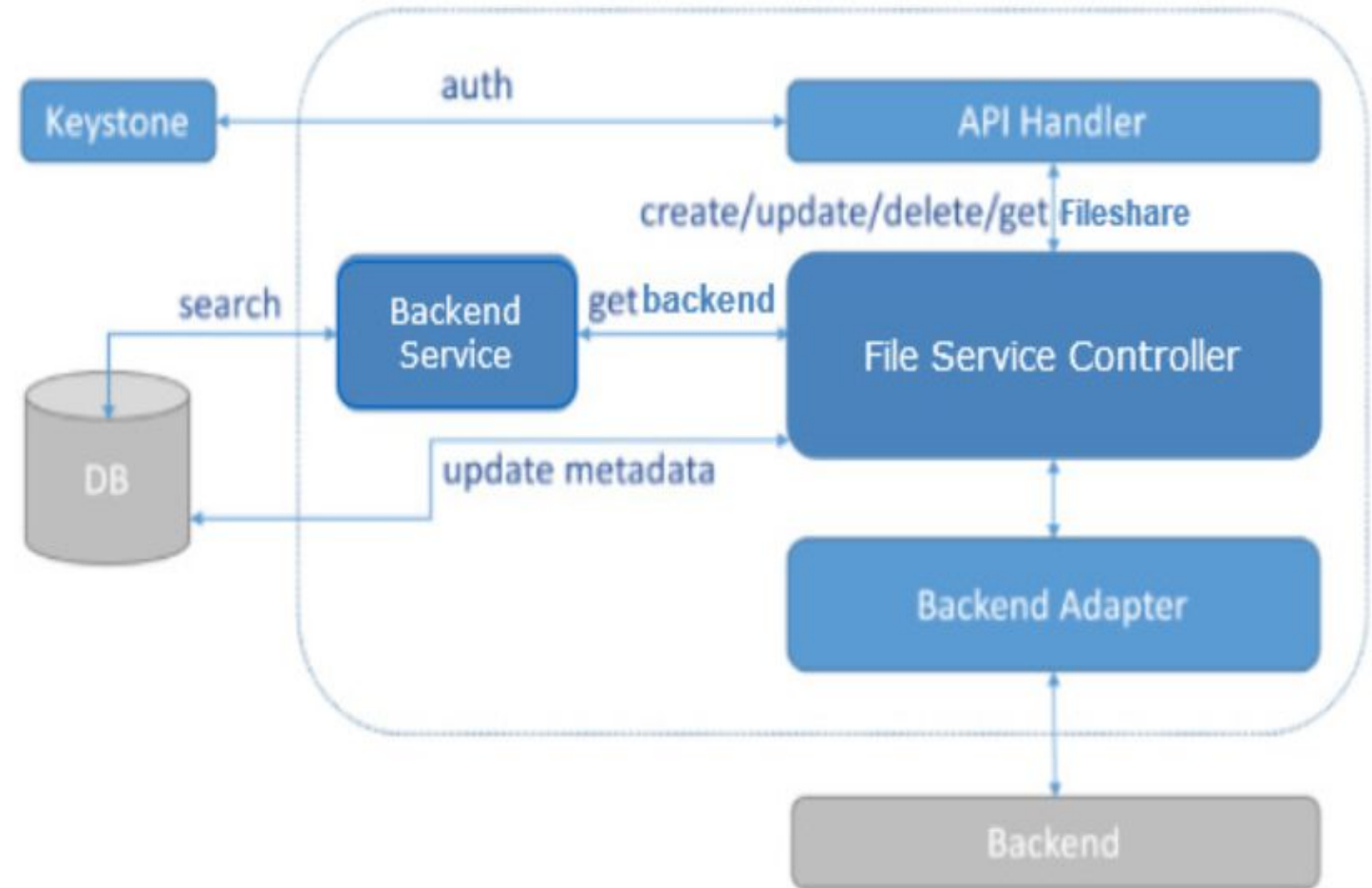
Connector Manager: Contains information to communicate with a bucket in the backend. It includes source and destination connectors.

Plan Manager: Controls plan executions for automatic or manual run.

Job Manager: Responsible to keep tracks the status of a migration job created after a plan starts execution based on the schedule.

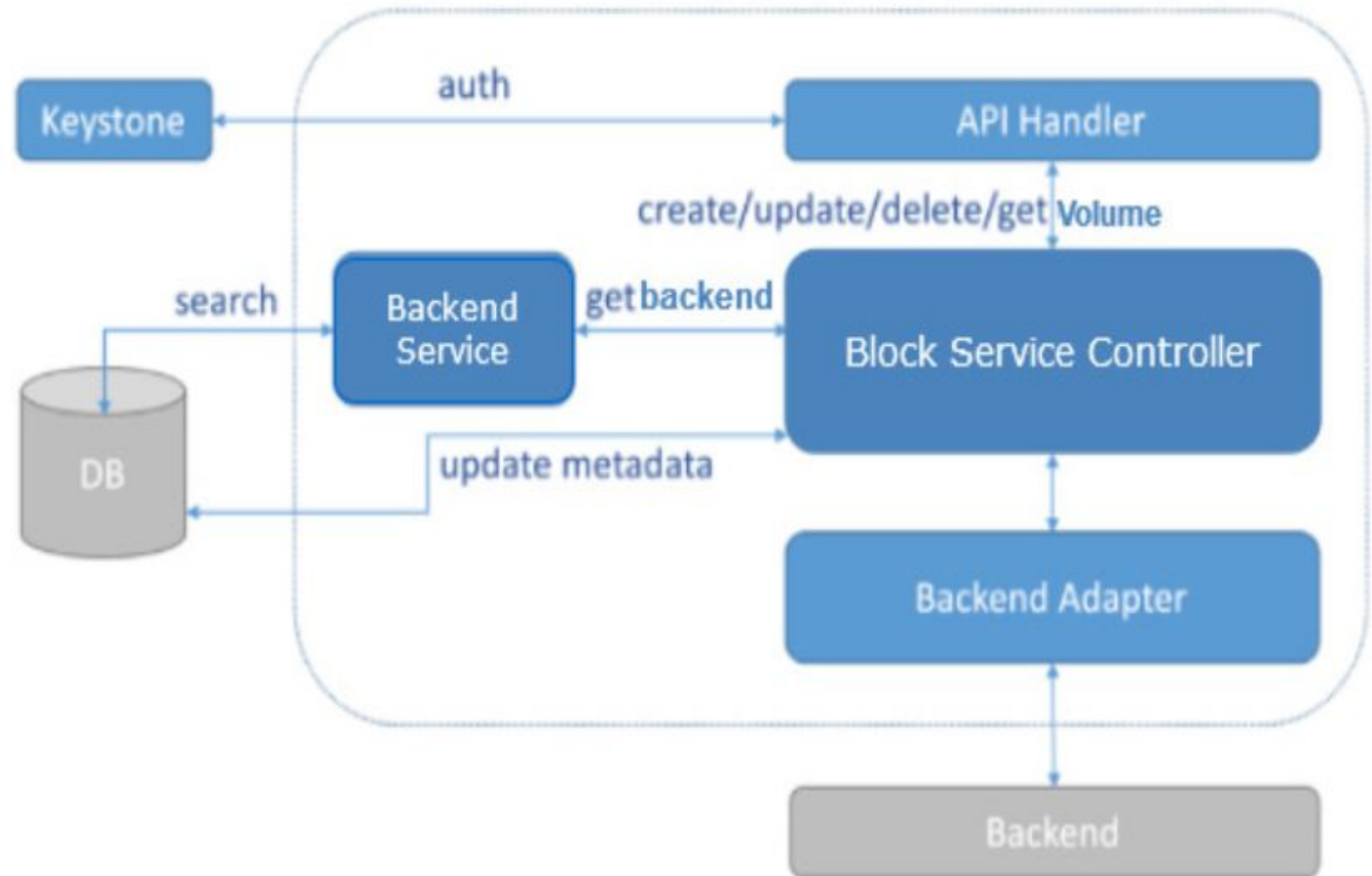
SODA Multi-Cloud: File Service

- Support Authentication with Keystone.
- Cloud File Share or Filestore Management support
- File Service Controller serves as a metadata placeholder for Cloud File/Filestore Resources.
- File Service support Create/ Update/Delete/Get a Fileshare.
- Support Cloud Vendors including AWS, Azure, GCP.



SODA Multi-Cloud: Block Service

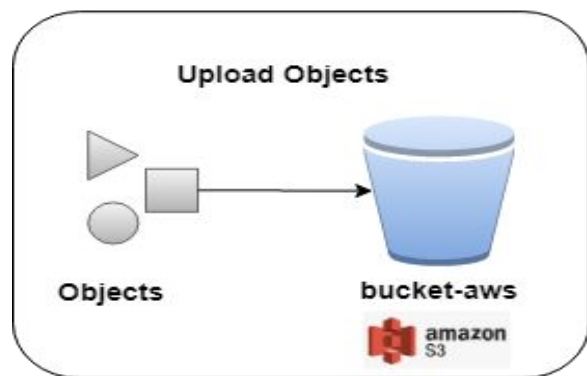
- Support Authentication with Keystone.
- Cloud Volumes/Block Device Management support
- Block Service Controller serves as a metadata placeholder for Cloud Volume Resources.
- Block Service support Create/ Update/Delete/Get a Volume.
- Support Cloud Vendors like AWS.



SODA Multi-Cloud: Services

```
fb921a8b2fab      sodafoundation/multi-cloud-dataflow:v1.1.0      "/dataflow"      29 hours ago
    Up 29 hours      opensds-gelato-linux-amd64
_dataflow_1_5fa2a0261a57
322a9a48a721      sodafoundation/multi-cloud-datamover:v1.1.0      "/datamover"      29 hours ago
    Up 29 hours      opensds-gelato-linux-amd64
_datamover_1_798177e2bb64
888d4c14e88a      sodafoundation/multi-cloud-s3:v1.1.0      "/initdb.sh"      29 hours ago
    Up 29 hours      opensds-gelato-linux-amd64
_s3_1_dd9d8c1c3bac
bc279fc4ff05      sodafoundation/multi-cloud-file:v1.1.0      "/file"      29 hours ago
    Up 29 hours      opensds-gelato-linux-amd64
_file_1_23ae3f28a442
92ff965485ee      sodafoundation/multi-cloud-api:v1.1.0      "/api"      29 hours ago
    Up 29 hours      0.0.0.0:8090->8090/tcp      opensds-gelato-linux-amd64
_s3api_1_c4e69c8aa738
68588212f081      sodafoundation/multi-cloud-block:v1.1.0      "/block"      29 hours ago
    Up 29 hours      opensds-gelato-linux-amd64
_block_1_7d123db2e506
b09bbff5347b      sodafoundation/multi-cloud-api:v1.1.0      "/api"      29 hours ago
    Up 29 hours      0.0.0.0:8089->8089/tcp      opensds-gelato-linux-amd64
_api_1_1217b92393ec
a7416329fbac      sodafoundation/multi-cloud-backend:v1.1.0      "/backend"      29 hours ago
    Up 29 hours      opensds-gelato-linux-amd64
_backend_1_b0a116be7f3b
```

SODA Multi-Cloud: Bucket/Object Migration

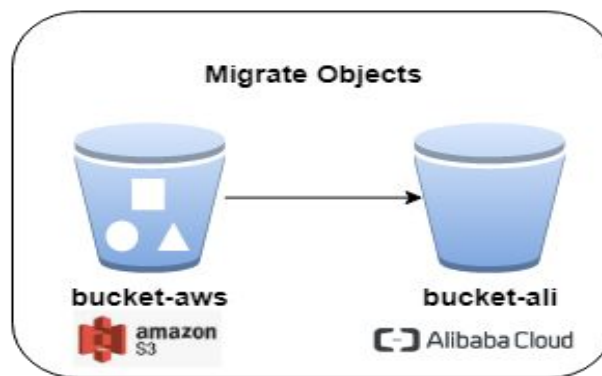


Pre-requisite:

- ✓ Register Backend (required AK, SK, region, endpoint, cloud-bucket name)
- ✓ Create a bucket on registered backend (eg. **bucket-aws**)
- ✓ Register another backend and create bucket (eg. **bucket-ali**)

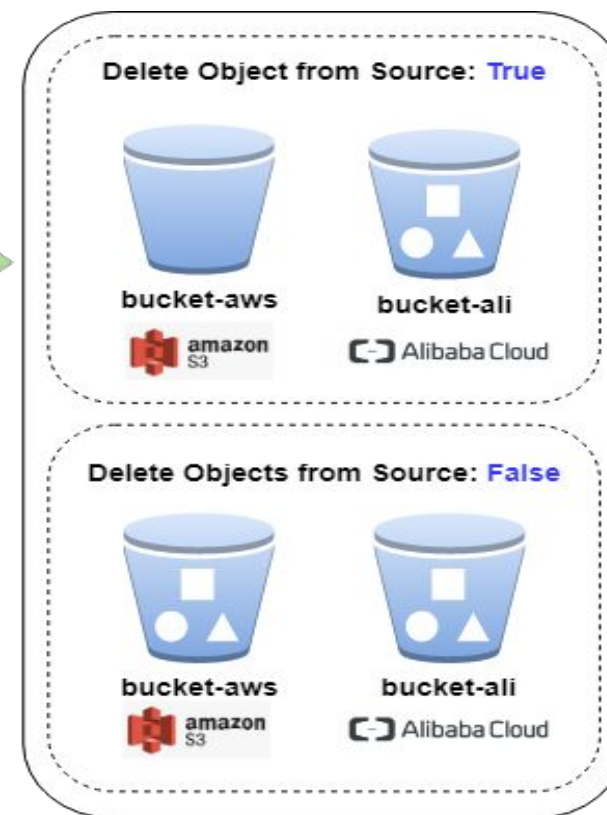
Upload Objects

- ✓ Upload multiple objects to bucket-aws
- ✓ User can upload object in multi-part also



Migrate Objects

- ✓ User can migrate data across the cloud.
- ✓ Supported storages: AWS-S3, Azure-Blob, Huawei, Alibaba OSS, Ceph, IBM, GCP, YIG
- ✓ User can migrate immediately as well as can schedule the migration to execute later.
- ✓ User have option to delete objects from source after successful migration to destination cloud.



After successful migration

- ✓ User can perform object operation and utilize objects of destination bucket for further use.

SODA Multi-Cloud: Bucket/Object Migration Demo

Drive Location for the Demo: [Click here](#)

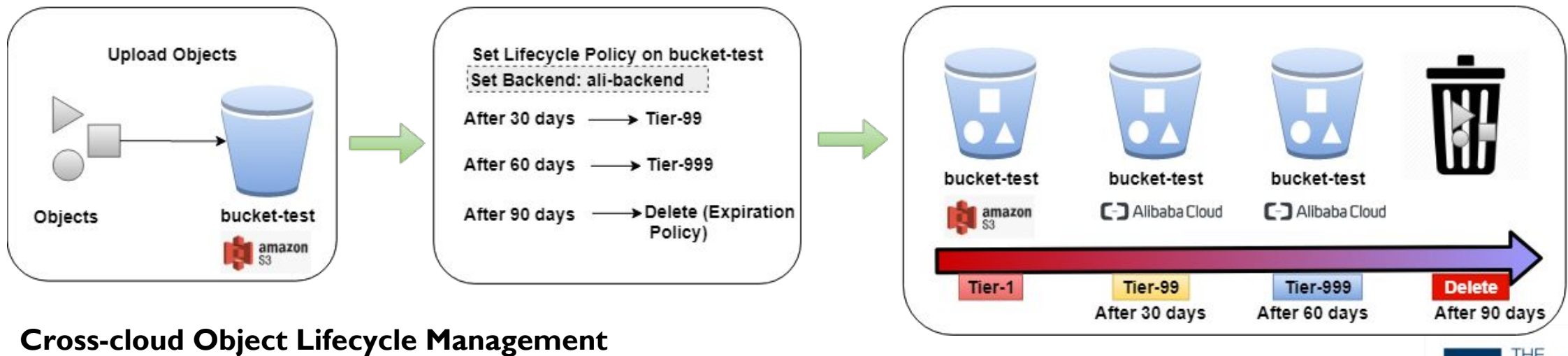
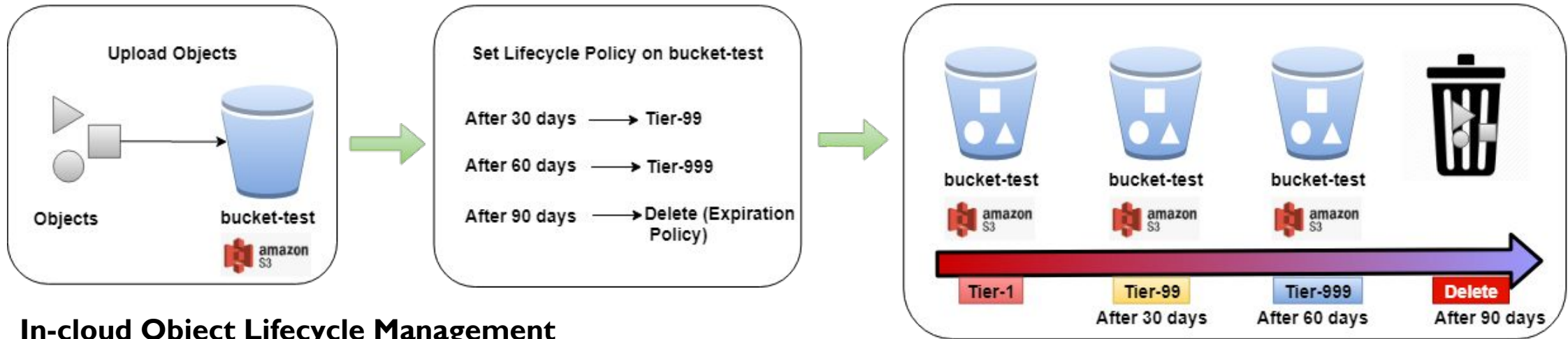


SODA Multi-Cloud: Object Life Cycle Management

- SODA combines different storage resources on-premise, and across multiple clouds to build tiers.
- In SODA, we have generalized storage class
 - If frequently used: **Tier-I**
 - If used infrequently: **Tier-99**
 - If not in use but still due to data policy, we store it for future purpose : **Tier-999**
- Using data lifecycle policies, data that meets the lifecycle conditions is moved to the next tier, allowing data to be stored efficiently throughout their lifecycle.

| SODA Storage Tier | AWS | Azure | Alibaba | Huawei |
|-------------------|-------------|---------|----------|----------|
| Tier-I | STANDARD | HOT | Standard | STANDARD |
| Tier-99 | STANDARD_IA | COOL | IA | WARM |
| Tier-999 | GLACIER | ARCHIVE | Archive | COLD |

SODA Multi-Cloud: Object Life Cycle Management



Thank You



<https://sodafoundation.io>



<https://github.com/sodafoundation>



<https://sodafoundation.io/slack>



[@sodafoundation](#)



<http://lists.sodafoundation.io/>