Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

terabit

INFN Cloud Object Storage Service: Gateway to a Multisite Infrastructure

Ahmad Alkhansa (ahmad.alkhansa@cnaf.infn.it)

Alessandro Costantini (alessandro.costantini@cnaf.infn.it)

Jacopo Gasparetto (jacopo.gasparetto@cnaf.infn.it)

Giada Malatesta (giada.malatesta@cnaf.infn.it)

Barbara Martelli (barbara.martelli@cnaf.infn.it)

Diego Michelotto (diego.michelotto@cnaf.infn.it)

Massimo Sgaravatto (massimo.sgaravatto@pd.infn.it)

Stefano Stalio (stefano.stalio@lngs.infn.it)

Palau, 21-05-2024

Workshop sul Calcolo nell'INFN

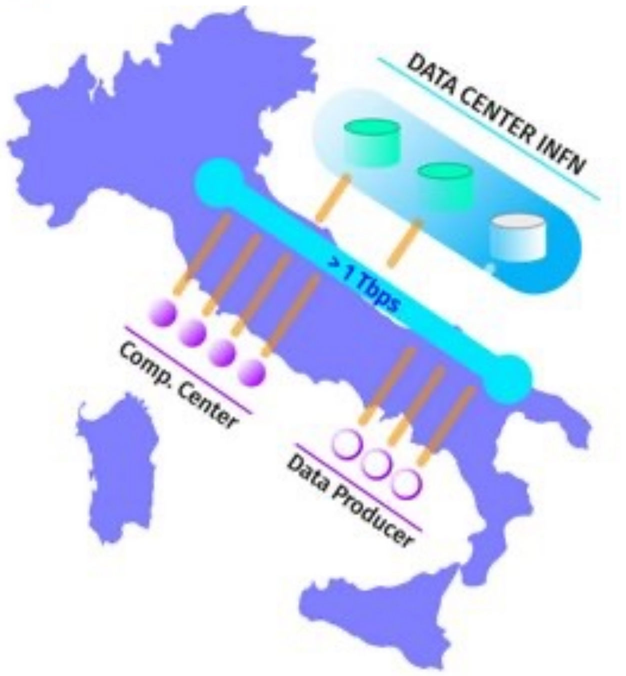
INFN Cloud

Infrastructure Architecture

- A **backbone** that consists of two main INFN computing sites.
- The sites are connected by high speed network.
- Multiple federated Cloud infrastructures.

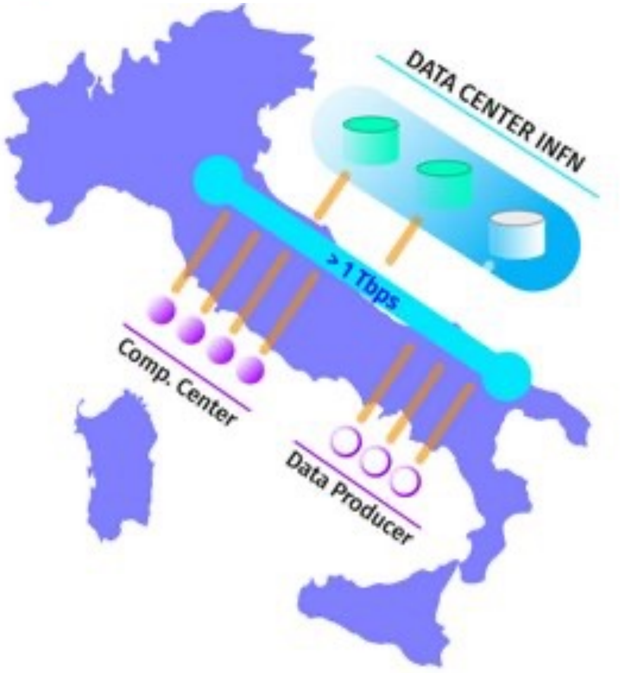
Services

- Federated Access to **IaaS, PaaS and SaaS**
- Virtualized computational resources.
- **Object Storage as a Service**.



Object Storage technology Requirements

- **Federated Access** allowing users to have **INDIGO IAM** as their identity provider**.**
- **Fine-grained Authorization** that permits different mechanisms of resource access.
- **Decentralized and Distributed storage solution** for high availability and disaster recovery.
- The ability to **Scale** and adapt to the increasing demand for storage.



Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

terabit

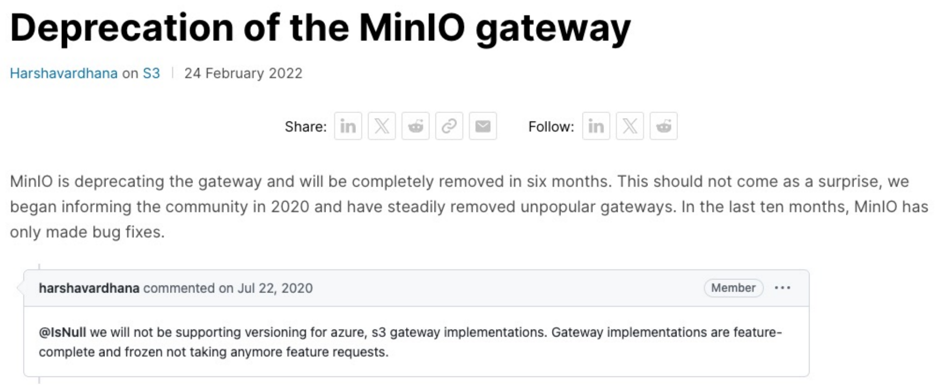
Object Storage technology Issue

Current State: MinIO Gateway

- **S3 compatible API.**
- **Uses Swift of OpenStack for storage.**
- **July 2020,** Not accepting feature requests.
- **February 2022,** Software deprecated.

Selected Substitute: Ceph Rados GateWay

- **Swift/S3 compatible API.**
- **Uses Ceph RADOS service for storage.**
- Part of **Ceph releases.**



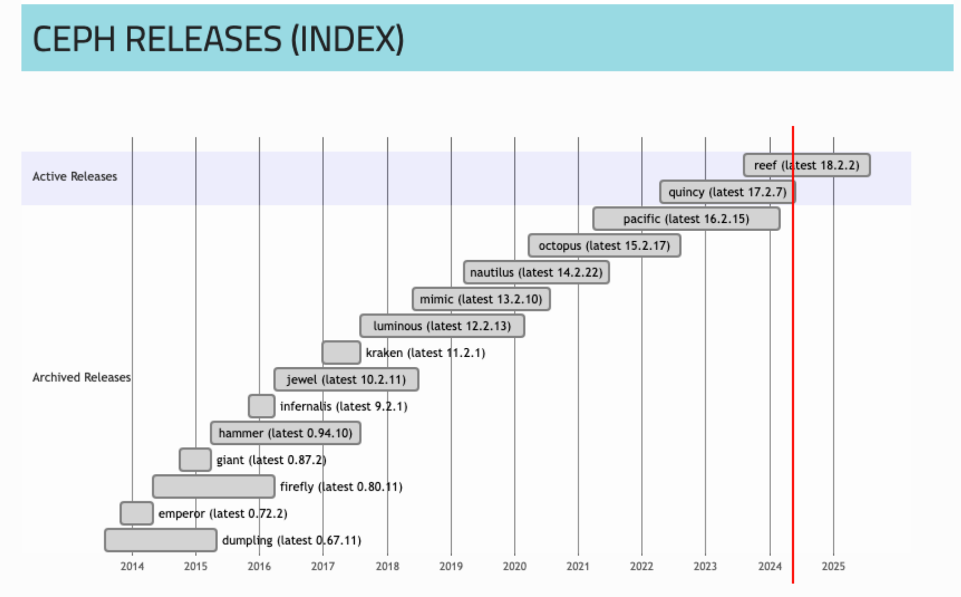
Deprecation of the MinIO gateway

Harshavardhana on S3 | 24 February 2022

MinIO is deprecating the gateway and will be completely removed in six months. This should not come as a surprise, we began informing the community in 2020 and have steadily removed unpopular gateways. In the last ten months, MinIO has only made bug fixes.

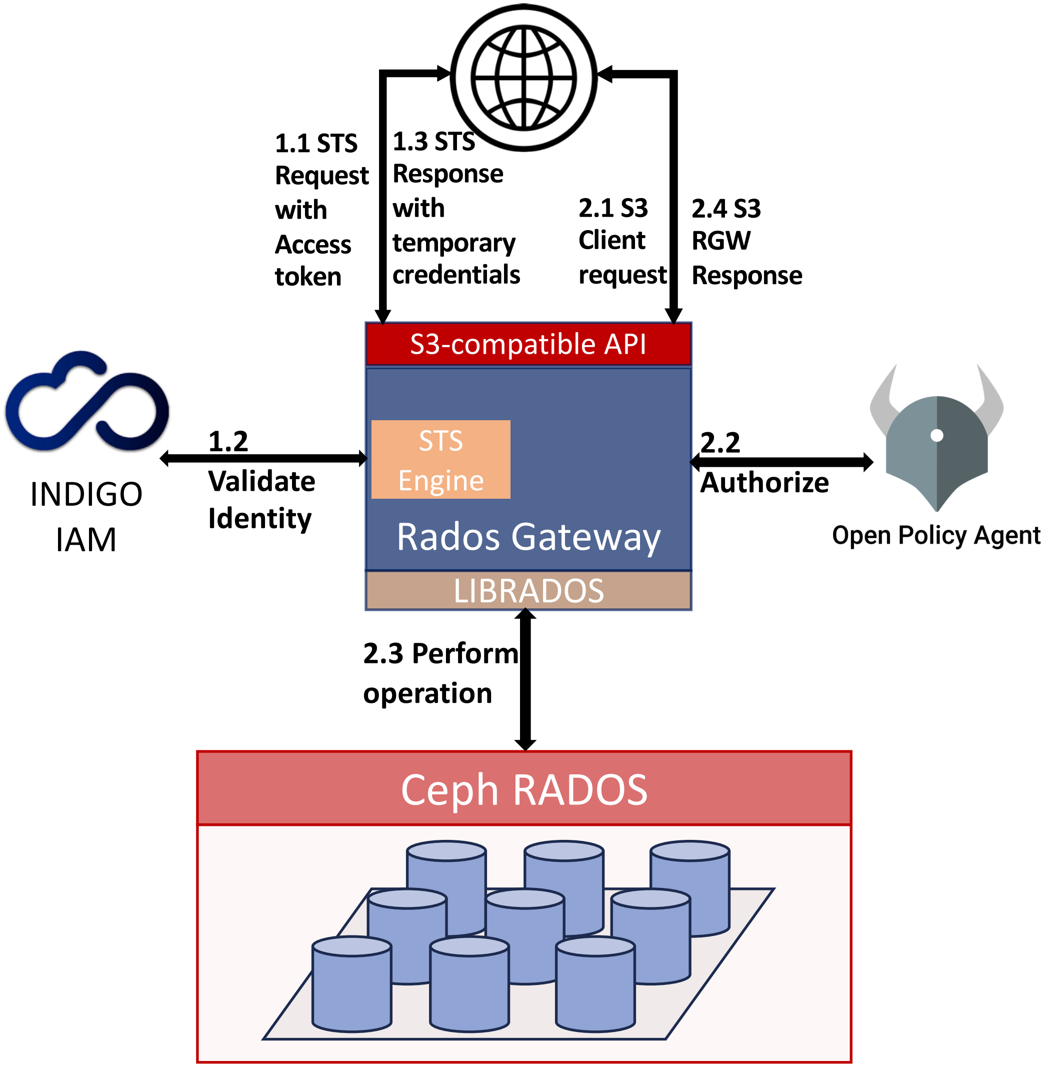
harshavardhana commented on Jul 22, 2020 Member

@IsNull we will not be supporting versioning for azure, s3 gateway implementations. Gateway implementations are feature-complete and frozen not taking anymore feature requests.



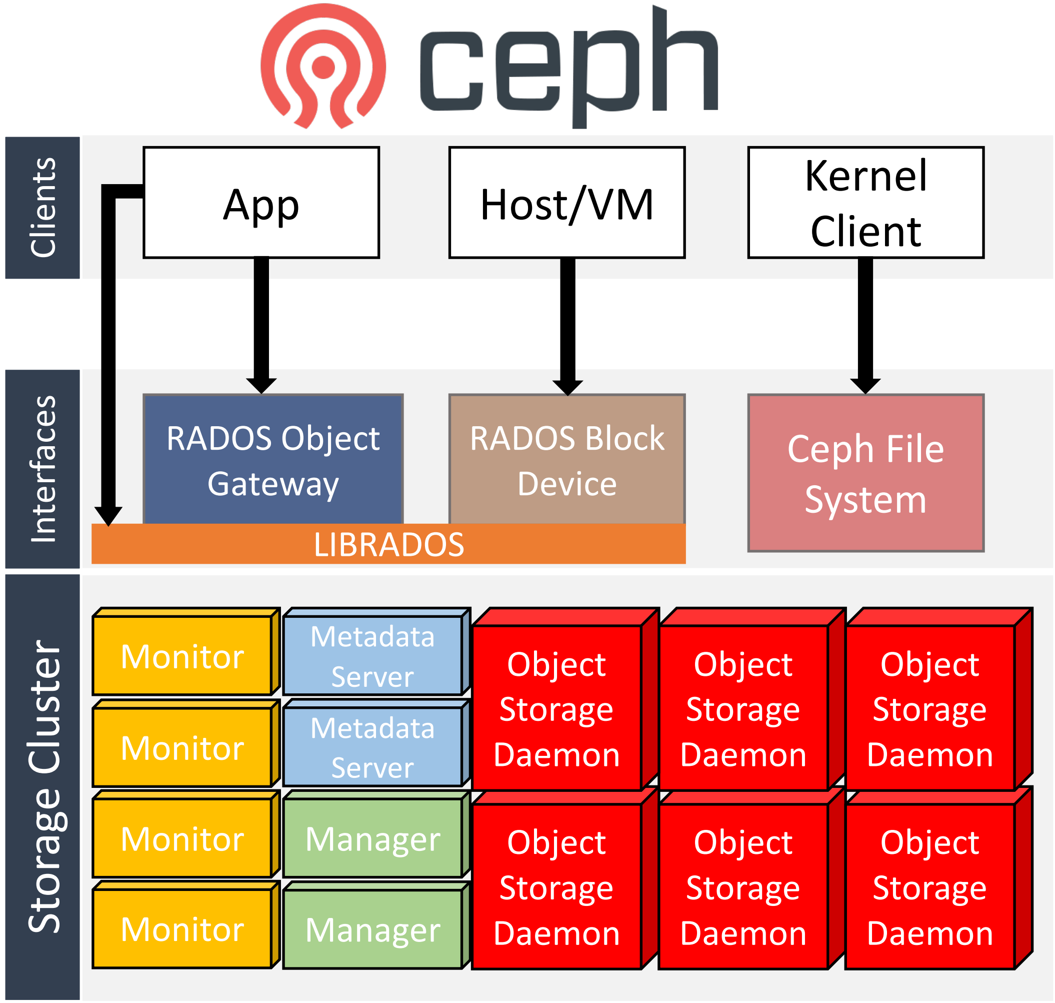
Rados GateWay

- Implements a large subset of **S3 RESTful API.**
- **Security Token Service (STS)** allowing access using **INDIGO IAM** as **OIDC provider.**
- **Integration** with **Open Policy Agent (OPA)** for **fine-grained Authorization.**
- **High availability** through **Mulitisite Configuration**
- **HTTP server** interacts with **Ceph Storage Cluster** using **LIBRADOS.**



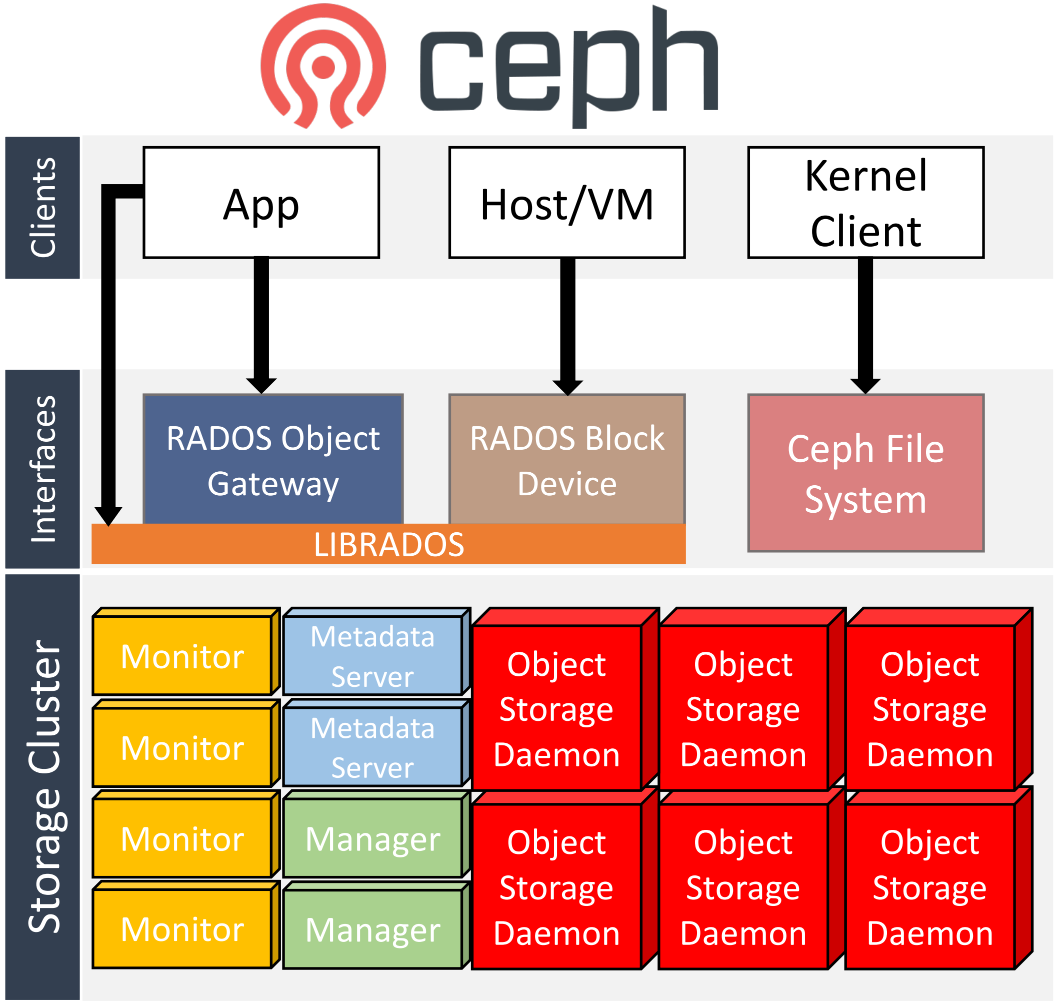
Ceph Storage Cluster

- **RADOS object store** that is able to reach **hyper-scale**.
- **Replicated RADOS** objects are stored and managed by **Ceph OSDs**.
- **High Availability** provided by **Ceph Monitors** that store a master copy of the cluster map.
- **Ceph OSDs and Clients** participate in the computation of the **objects localization** by using **CRUSH** algorithm, achieving **Decentralization**.



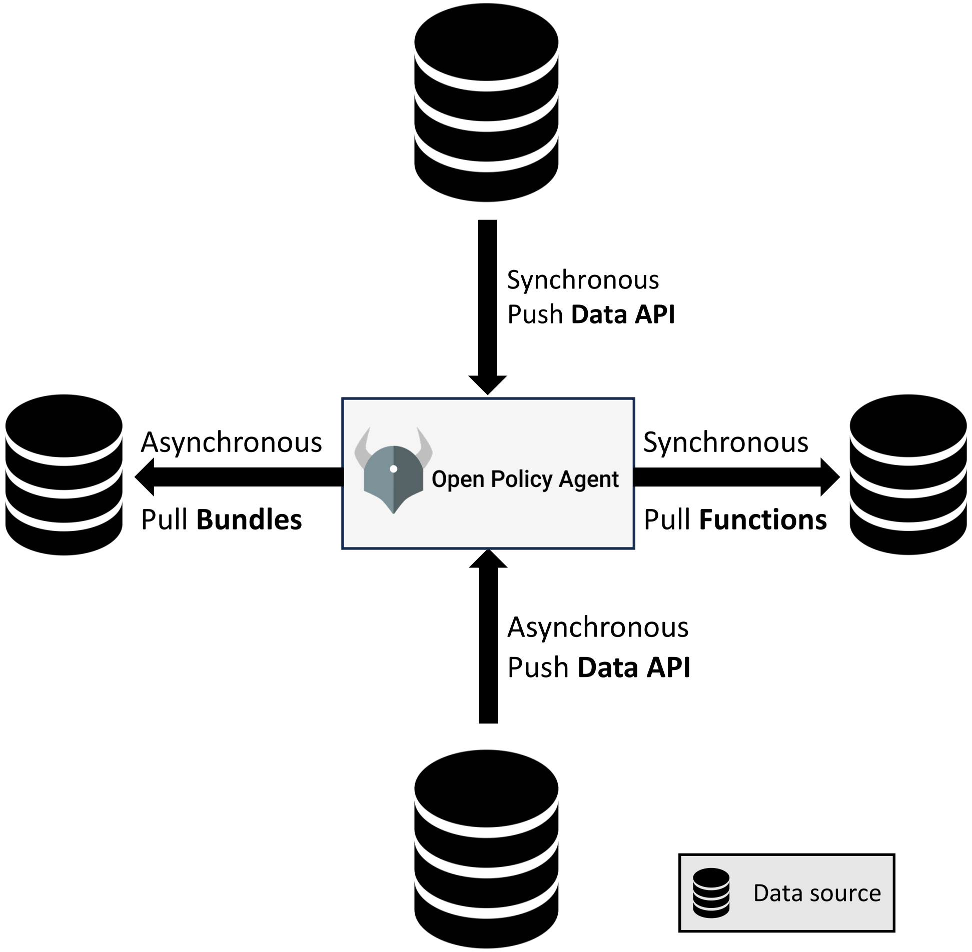
Ceph Interfaces

- **Posix-Compliant CephFS** distributed file system for **CephFS Kernel Clients or FUSE mount**.
- **Rados Block Device (RBD)** with virtualization in consideration for **RBD Kernel Clients** with **QEMU/KVM** drivers.
- **Rados GateWay** that provides **Swift/S3 compatible API for Object storage**.
- **LIBRADOS** C++ library that allows **direct access to RADOS**.



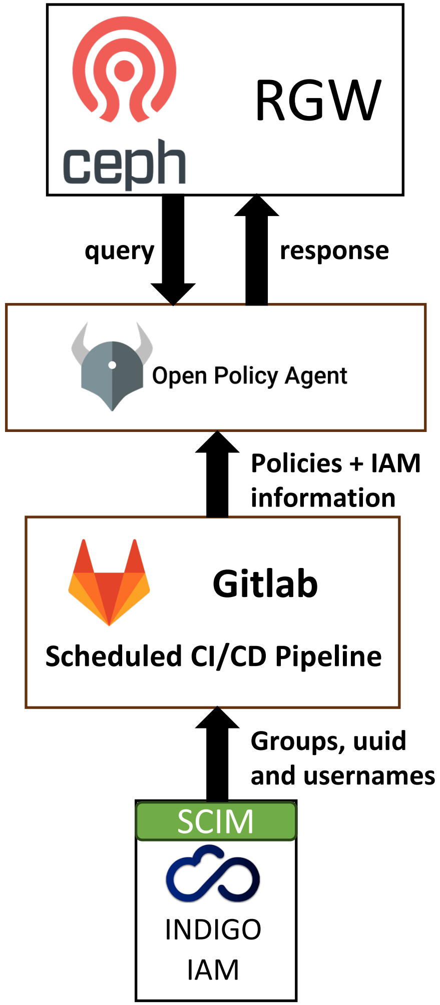
Open Policy Agent

- **Scalable** and **Lightweight** policy engine.
- Applies **Policy Decoupling** that offloads **decision making** from the software.
- Receives and Produces **documents** that contribute to the **authorization process**.
- Different models for **dynamic policy and data management**.
- Handles **authorization of S3 operations** requested to **RGW**.



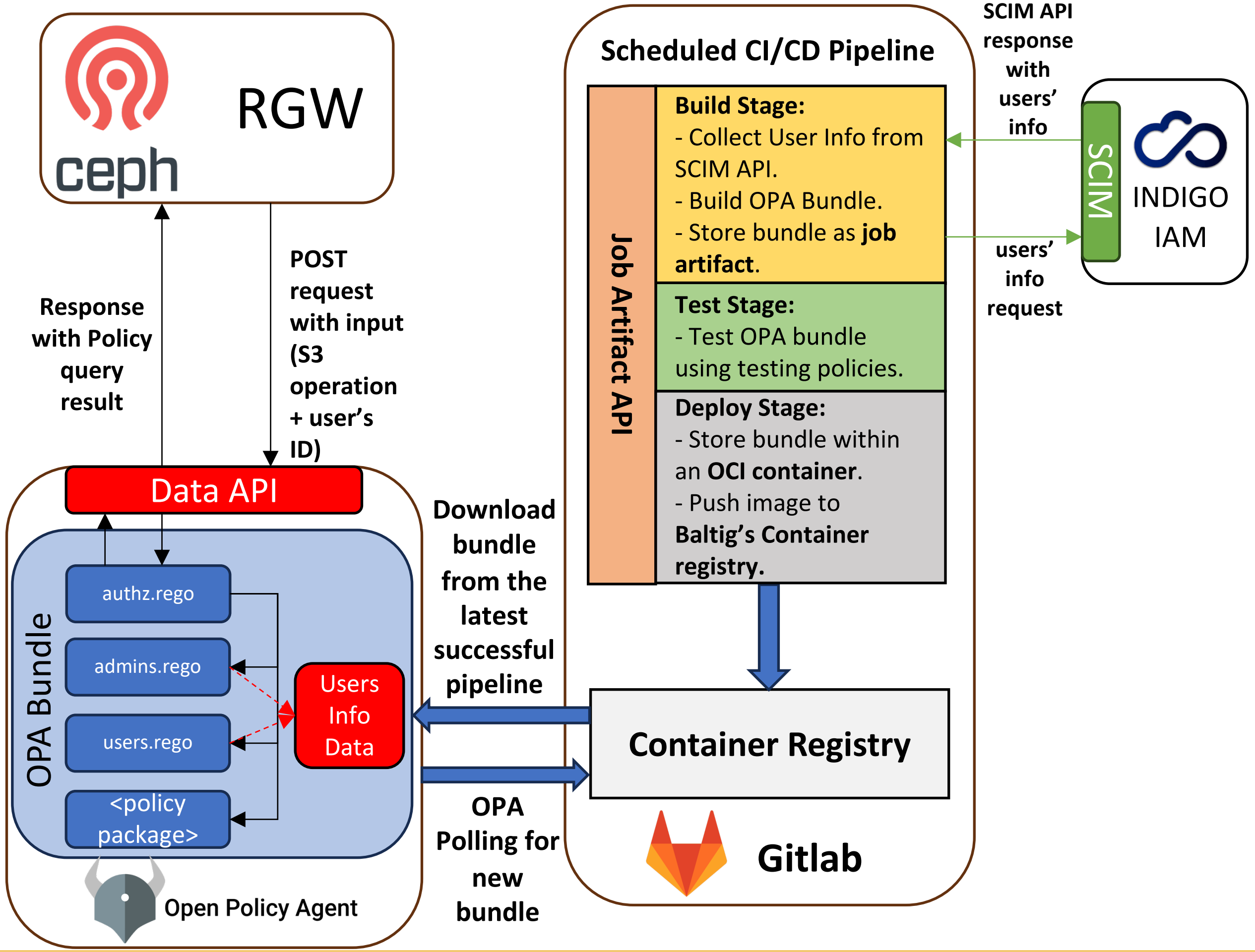
Open Policy Agent – Authorization Architecture

- Obtains information from **IAM** allowing **fine-grained authorization.**
- **Rados Gateway** authorization request contains **user's uuid and the S3 operation**.
- **Open Policy Agent** requires additional information such as **groups and username** in its decision making.
- **Gitlab** repository that contains **OPA policies** managed by administrators.
- **Gitlab CI/CD Pipline** periodically provide **policies and users' information**.



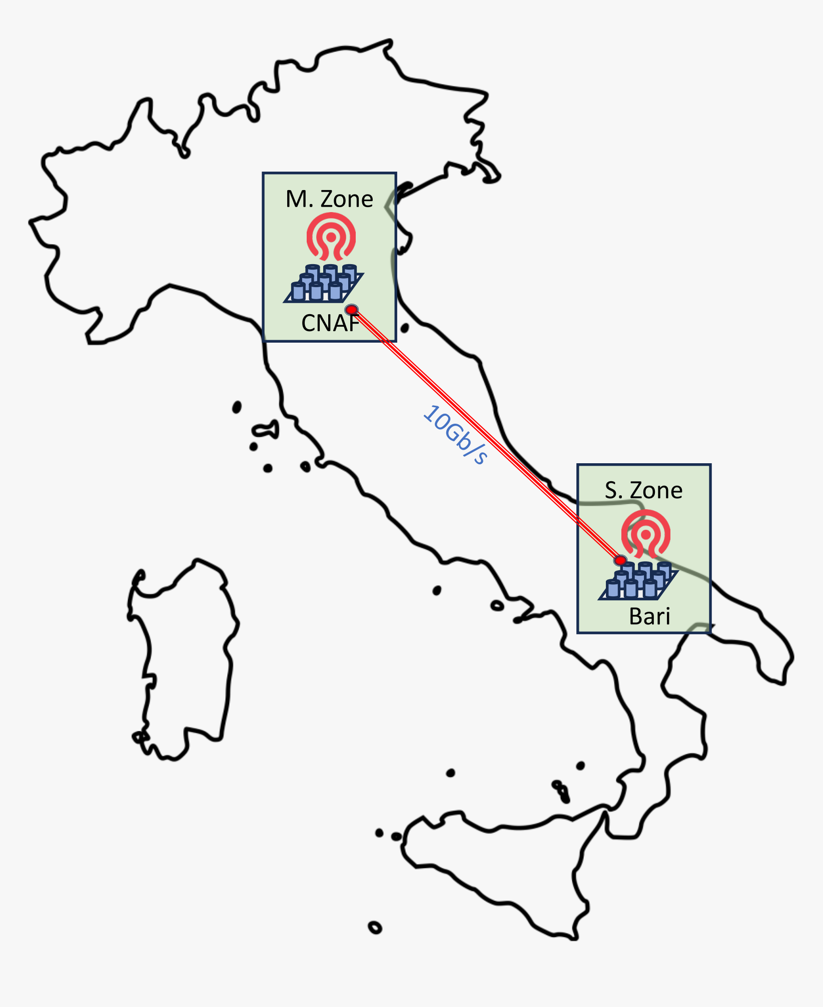
Open Policy Agent – Integration Mechanism

- **Gitlab CI/CD pipeline:**
  - **Collects Users' information** from INDIGO IAM **SCIM API**.
  - Generates **OPA bundle** from the collected data and the **stored policy files** within the **repository**.
  - Create an **OCI container** from the bundle and push it the **registry**.
- **OPA** service **Polls** Gitlab's **Container registry** for new **bundle.**


ceph
RGW
Response with Policy query result
POST request with input (S3 operation + user's ID)
Data API
OPA Bundle
authz.rego
admins.rego
users.rego
<policy package>
Users Info Data
Open Policy Agent
Download bundle from the latest successful pipeline
OPA Polling for new bundle
Scheduled CI/CD Pipeline
Job Artifact API
Build Stage:
- Collect User Info from SCIM API.
- Build OPA Bundle.
- Store bundle as job artifact.
Test Stage:
- Test OPA bundle using testing policies.
Deploy Stage:
- Store bundle within an OCI container.
- Push image to Baltig's Container registry.
Container Registry
Gitlab
SCIM API response with users' info
users' info request
SCIM
INDIGO IAM

Backbone Multisite Configuration

- **Single Realm** with a **unique Zonegroup**.
- **2 Zones** with independent **Ceph Storage Clusters**.
- **3 instances** of RADOS Gateway running in **high availability** within each zone.
- **Every Zone** connects to a dedicated **OPA deployment** through a private network.
- The zones have **Active-Passive** Configuration.
- 600 KMs apart with 9 ms latency and 10Gb/s of bandwidth.



Rados Gateway Clients

Posix-like Mount of S3 Buckets

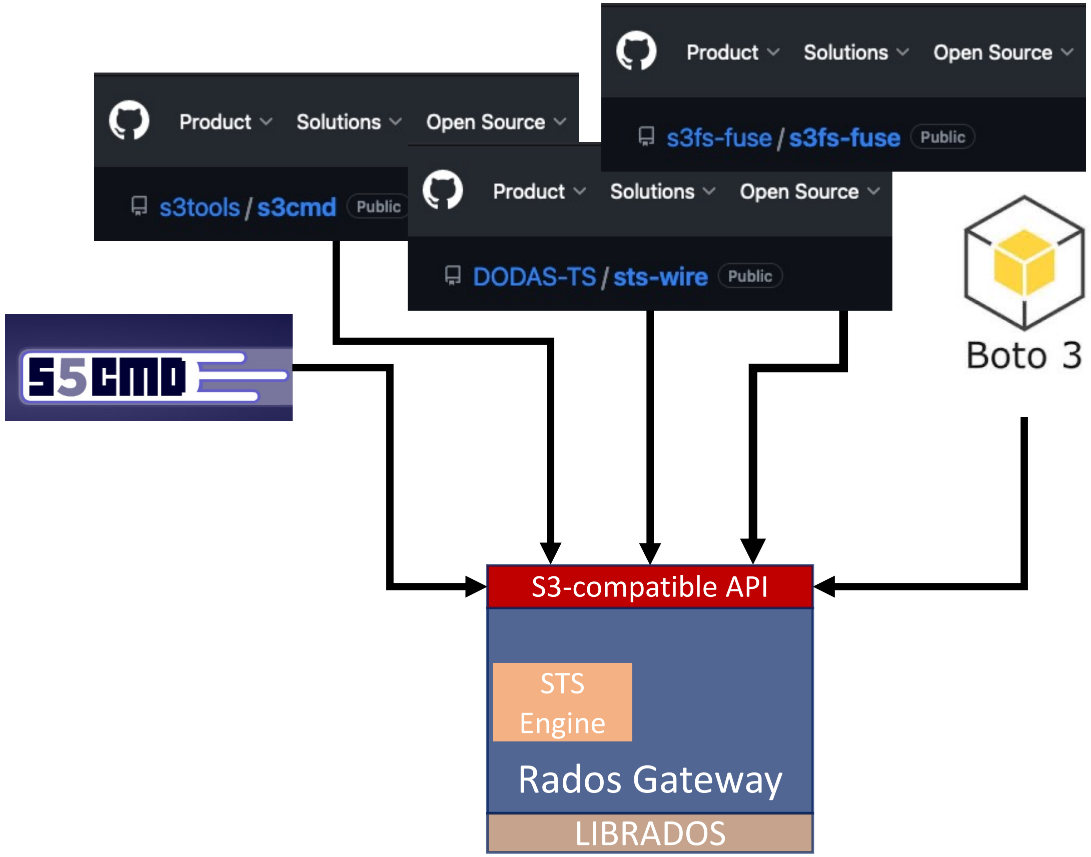
- **sts-wire** is a wrapper around **rclone.** It allows continuous refresh of **IAM Access tokens.**
- **S3fs-fuse** uses **FUSE (Filesystem in Userspace)** to mount **S3 Buckets** with large subset of **POSIX**.

Command Line Tools

- **S3cmd** is a client written in **Python** that allows **S3 bucket management** and **data synchronization** with the **local directory**.
- **S5cmd** is a **GO** tool that focuses on **performance**.

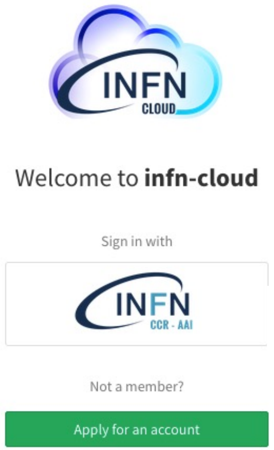
Software Development Kit (SDK)

- **Boto3** is developed by **Amazon** for **Python** that provides **object-oriented API**.



S3 Web User Interface

- Based on **React and FastAPI**.
- **OIDC** protocol with **IAM** to generate **Json Web Token (JWT)**.
- Uses **IAM Access Token** to perform **STS with RGW**.
- **S3 operations using AWS SDK library**.

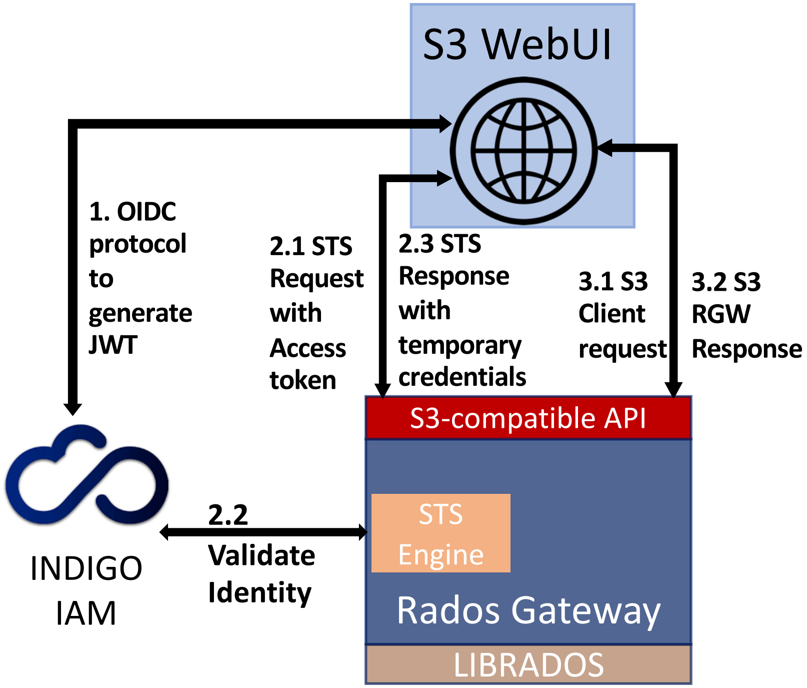


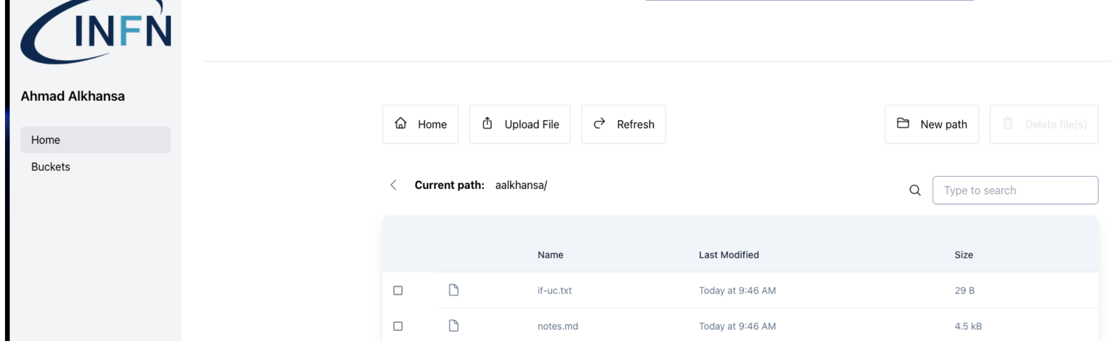
Welcome to **infn-cloud**

Sign in with

Not a member?

Apply for an account





Ahmad Alkhansa

Home

Buckets

Home | Upload File | Refresh | New path | Delete file(s)

Current path: aalkhansa/

Type to search

		Name	Last Modified	Size
□		if-uc.txt	Today at 9:46 AM	29 B
□		notes.md	Today at 9:46 AM	4.5 kB

Conclusion

- The migration of object storage service is expected soon.

- Rados Gateway
  - Provides S3 compatible API.
  - Integrated with OIDC providers using the Security Token Service Engine.
  - Allows fine grained authorization with Open Policy Agent.
  - Ensures High availability through a multi-site configuration.

- DevOps tools are advantageous for fast authorization model updates.

- Several open-source clients exist with different approach of data management.

- S3 Web App is a Graphical User Interface for the Object Storage Service.

INFN CLOUD

ceph

INDIGO IAM

Open Policy Agent

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA

terabit

Thank you

ahmad.alkhansa@cnaf.infn.it