The Story of Hybrid Data Management across Heterogeneous Storage for Cloud, Core and, Edge

Sanil Kumar D, TOC, Arch WG Lead,
SODA Foundation | Chief Architect,
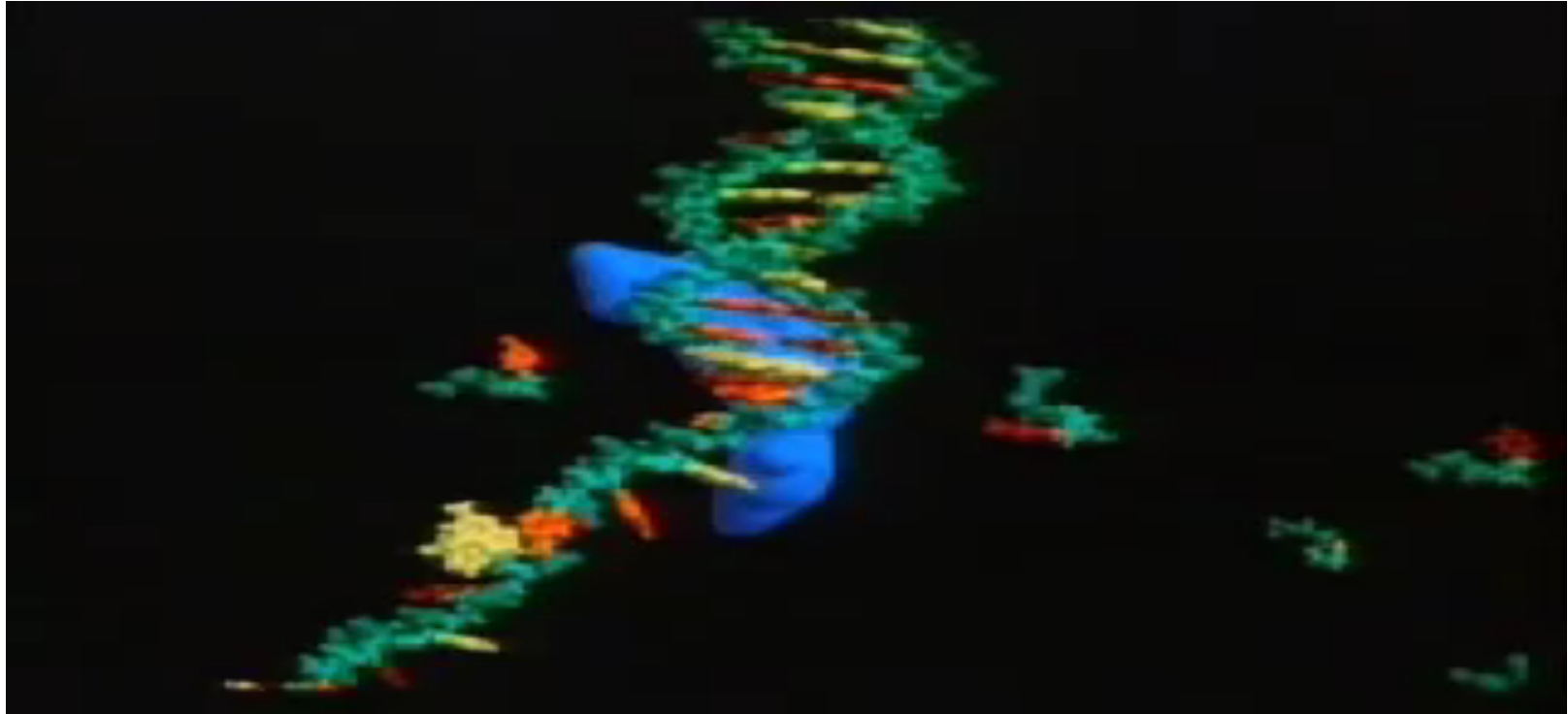Huawei Technologies, India

Ashit Kumar, India OC,
SODA Foundation | Lead Architect.
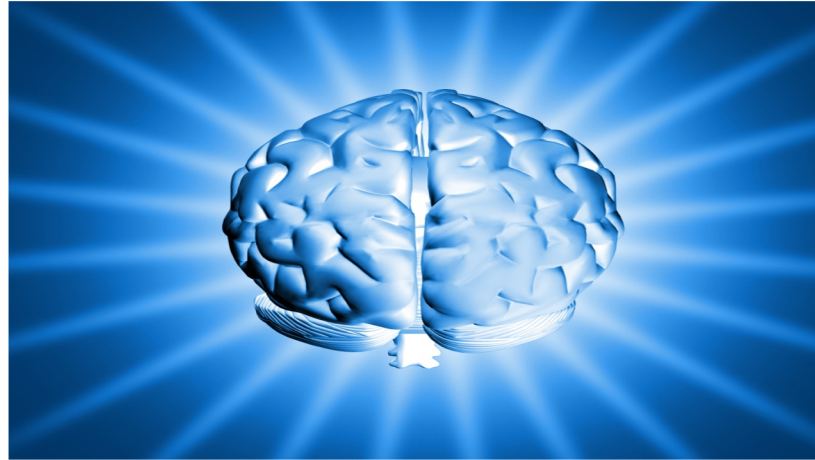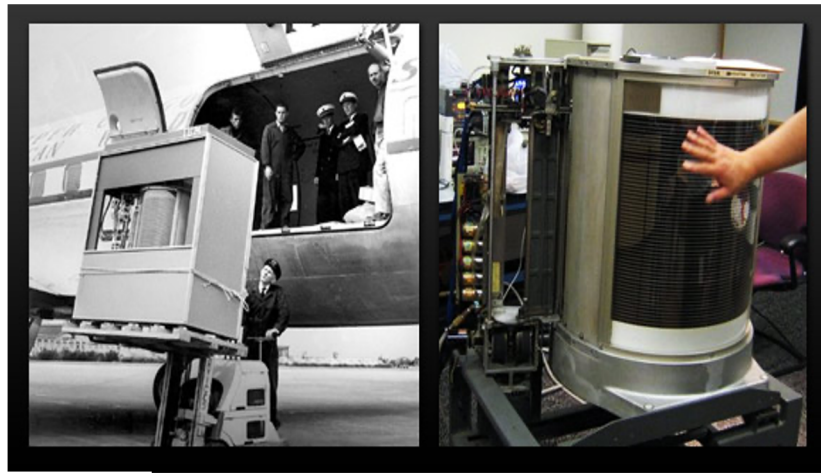Huawei Technologies, India

soda foundation

Once upon a time...

ttps://www.youtube.com/watch?v=B1hVx7BsvjU (3:30 - 4:25)

A Mankind is born with Big Data...

Stores 1 petabyte to 2.5 petabytes…

He started playing with Data…(unknowingly?)

Volume	Low
Velocity	Slow
Veracity	High
Variety	Less
Value	High
Verification	Not Critical
Visualization	Simple, Easy

John C. L. Gibson 1975 Textbook of Syrian Semitic Inscriptions, vol. 2   (Aramaic)

But Data Started Growing…

Punch Card in
Punch Card Machine

He invented better ways for Data...

soda foundation

..and kept generating tons of data

..every day..every second..and..

...Data Started Exploding...

Total worldwide data in zettabytes (ZB)

$$1 \text{ ZB} = 1000^7 \text{bytes} = 10^{21} \text{bytes} =$$

$$1000000000000000000000 \text{byte}$$

Amount of data [ZB]

40

30

20

10

0

2008   2012   2016   2020

Year

Source: 'EMC Digital Universe Study': https://www.emc.com/leadership/digital-universe
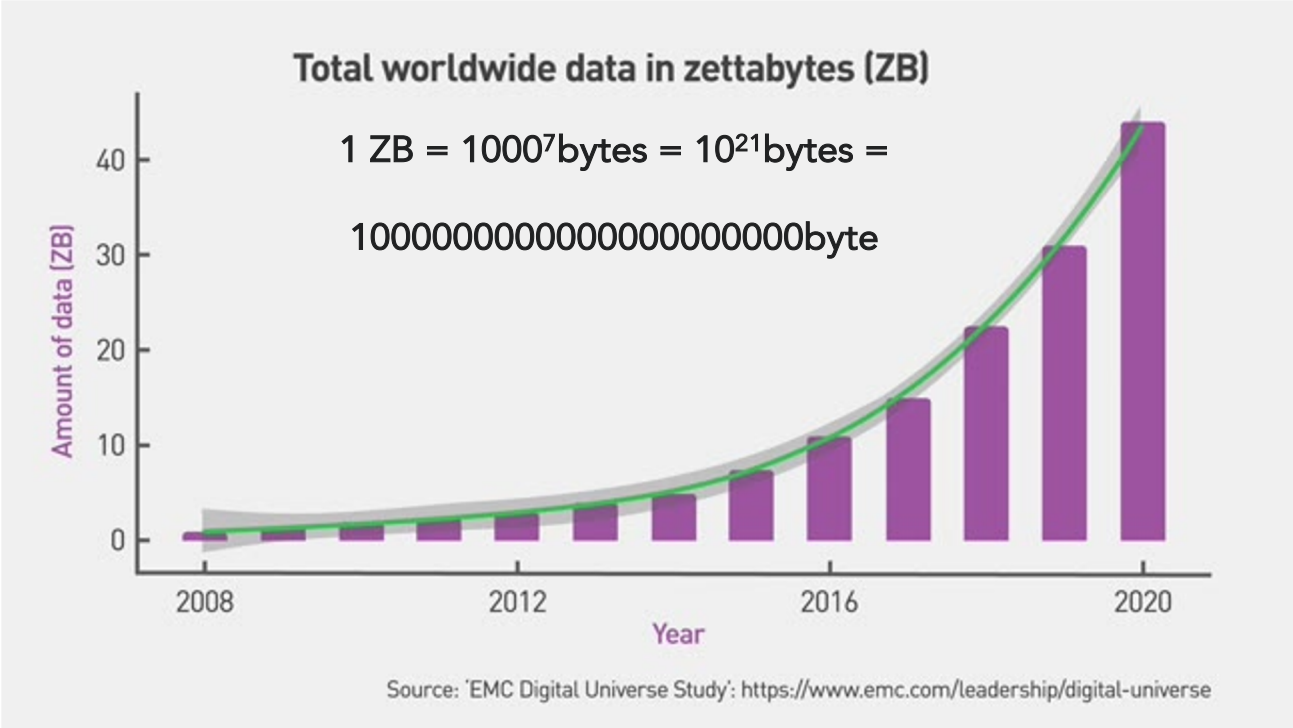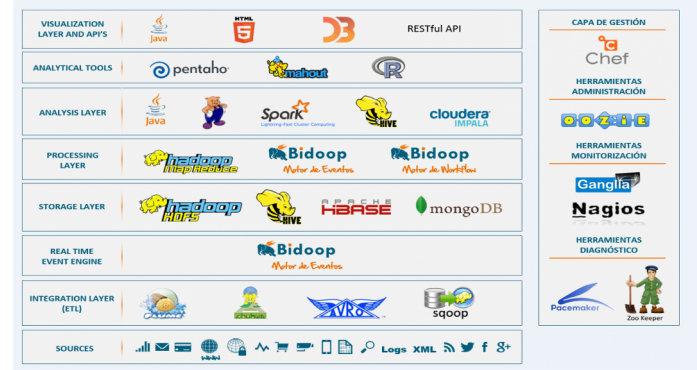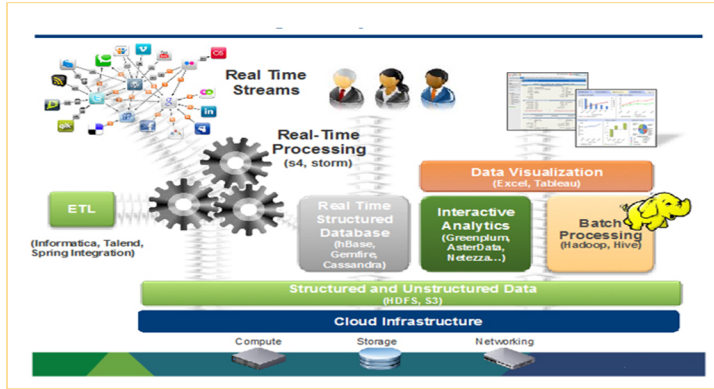
He kept innovating...

Different Storages, Different Platforms

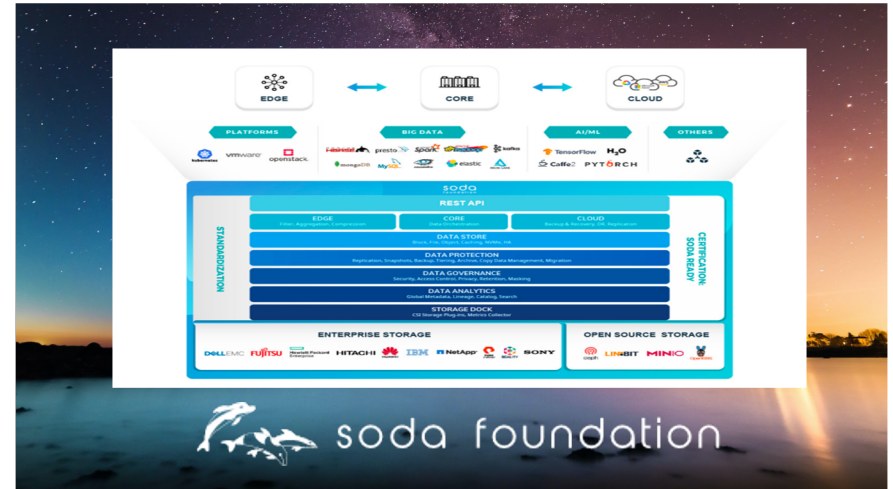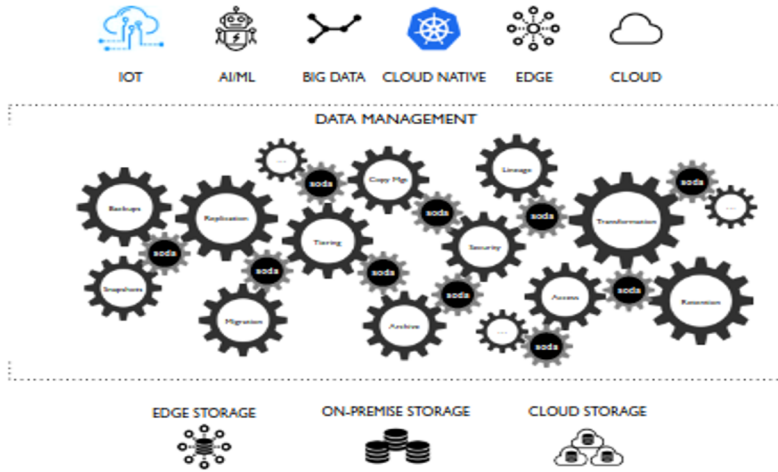## Different Ways for Managing his Data

...Simple to Complex..

...kept trying to make life simple, but…!

...Fragmented Data Management

IOT   AI/ML   BIG DATA   CLOUD NATIVE   EDGE   CLOUD

DATA MANAGEMENT

Backups   Replication   ...   Copy Mgt   Lineage   Transformation   ...

Tiering   Security

Snapshots   Migration   Archive   ...   Access   Retention
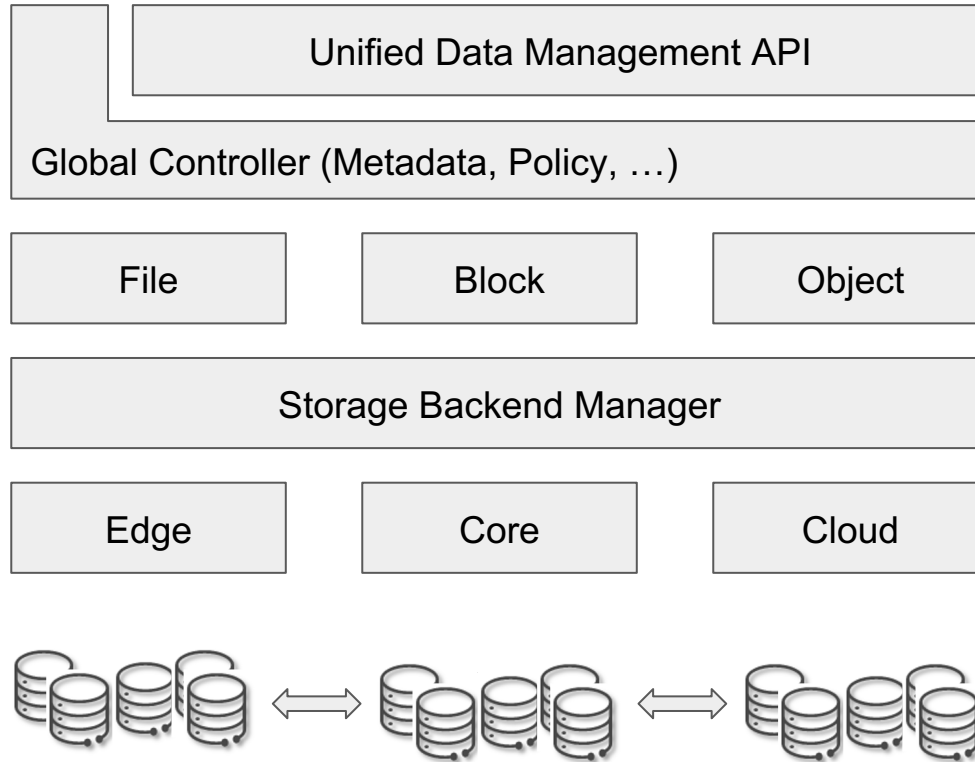
EDGE STORAGE   ON-PREMISE STORAGE   CLOUD STORAGE
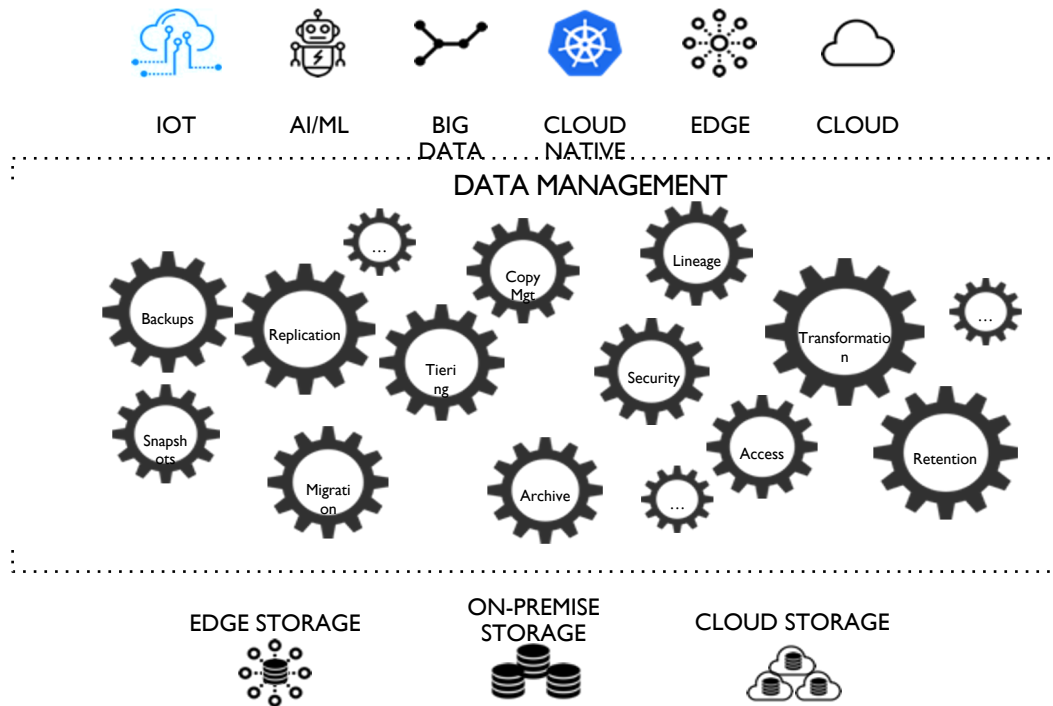
...he wants it to be simpler, so..

One Open Data Framework, Infinite Possibilities

He started dreaming...

One Open Data Store

Unified Data Management API

Global Controller (Metadata, Policy, …)

File	Block	Object

Storage Backend Manager

Edge	Core	Cloud

He started the journey...
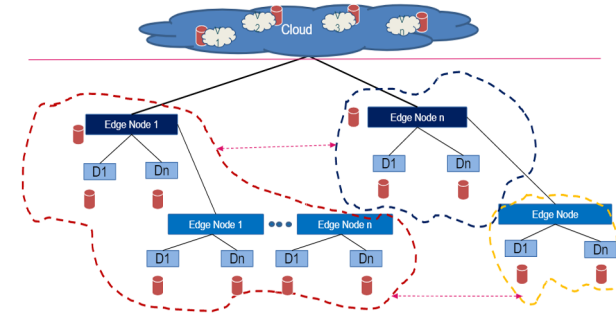
On-premise/Core	Cloud	Edge
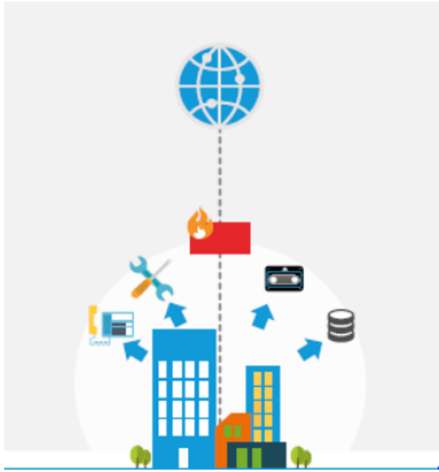
Heterogeneous
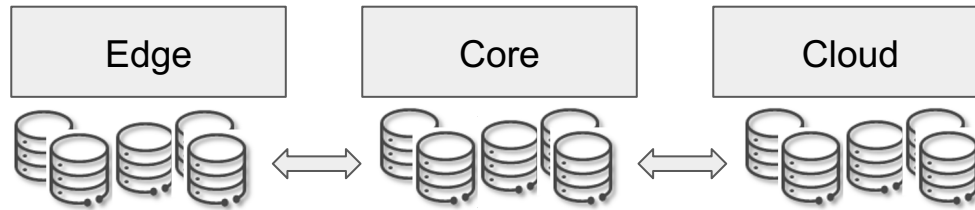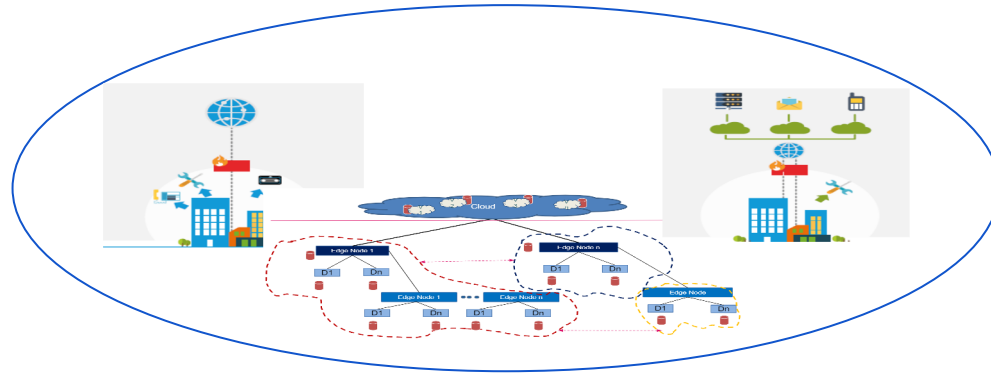Storages and
Platforms

Multiple Cloud
Data Mobility

Distributed
Heterogeneous
Low Resource
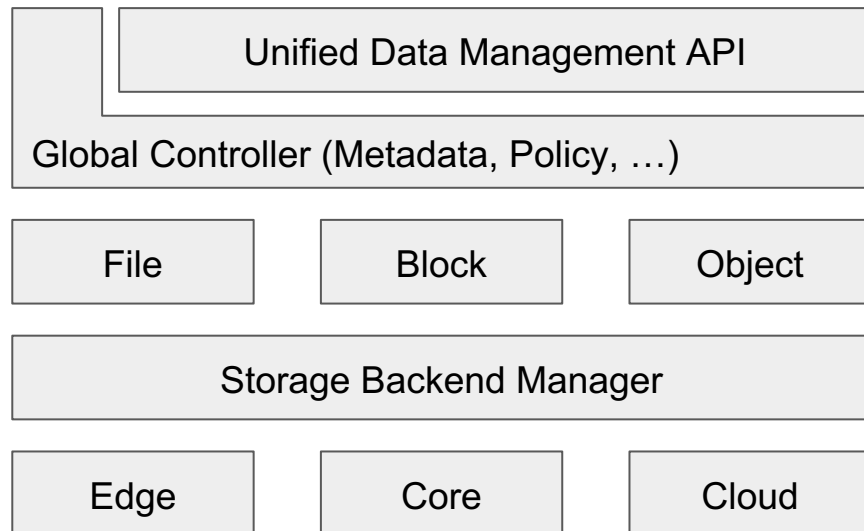Data Mobility

He landed where everything needs to co-work...

Data Management Across Hybrid Clouds

Edge	Core	Cloud

Across Heterogeneous Storages

Architecture for Hybrid Cloud Data Management...

One Data Store

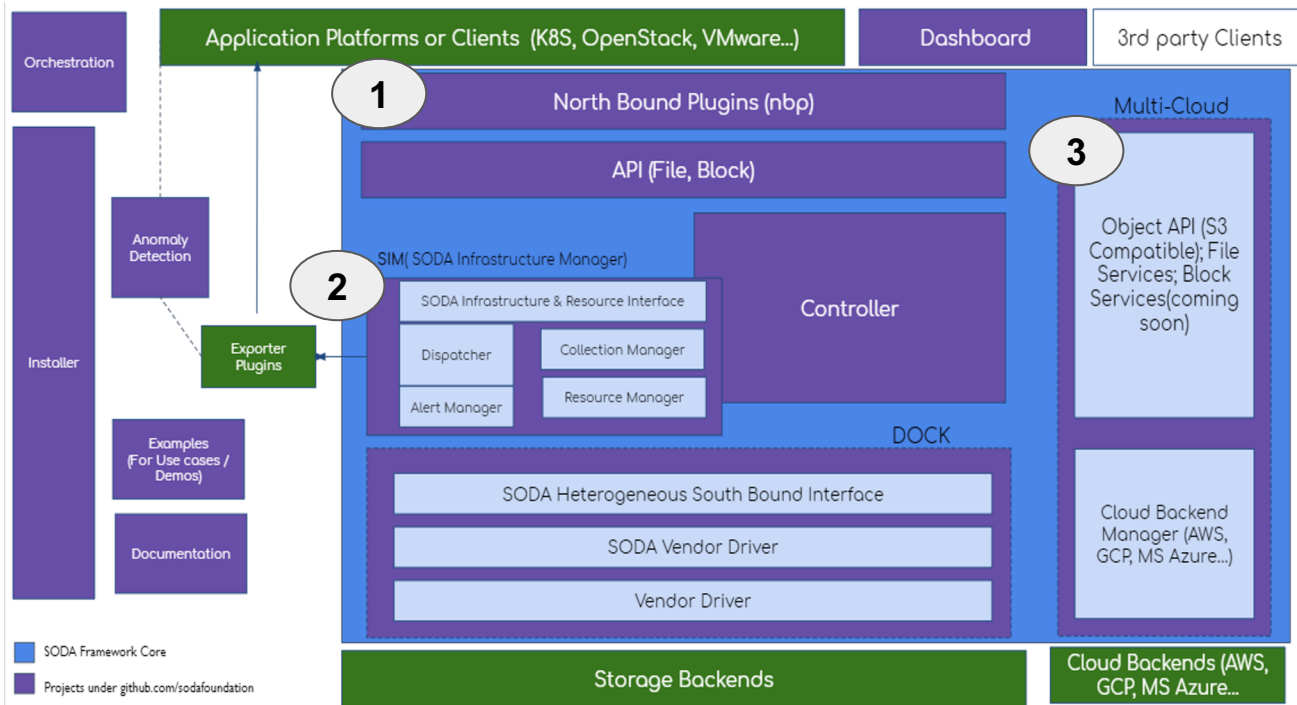| Unified Data Management API | Data Management APIs are unified, works across Edge, Core and Cloud + Heterogeneous Storage Vendors (Cloud Vendors / Storage Vendors/Types of Storages) |

Global Controller (Metadata, Policy, …)

| File | Block | Object | Support different Storage Classes |

| Storage Backend Manager | Dockable Storage Backend Manager |

| Edge | Core | Cloud | Microservice architecture to suit deployments across Edge, Core and Cloud |

Current Architecture Supporting Core and Multi Cloud

1

Core (On-prem) Data Management Across Heterogeneous Storages

2

Core (On-prem) Storage Monitoring Across Heterogeneous Storages
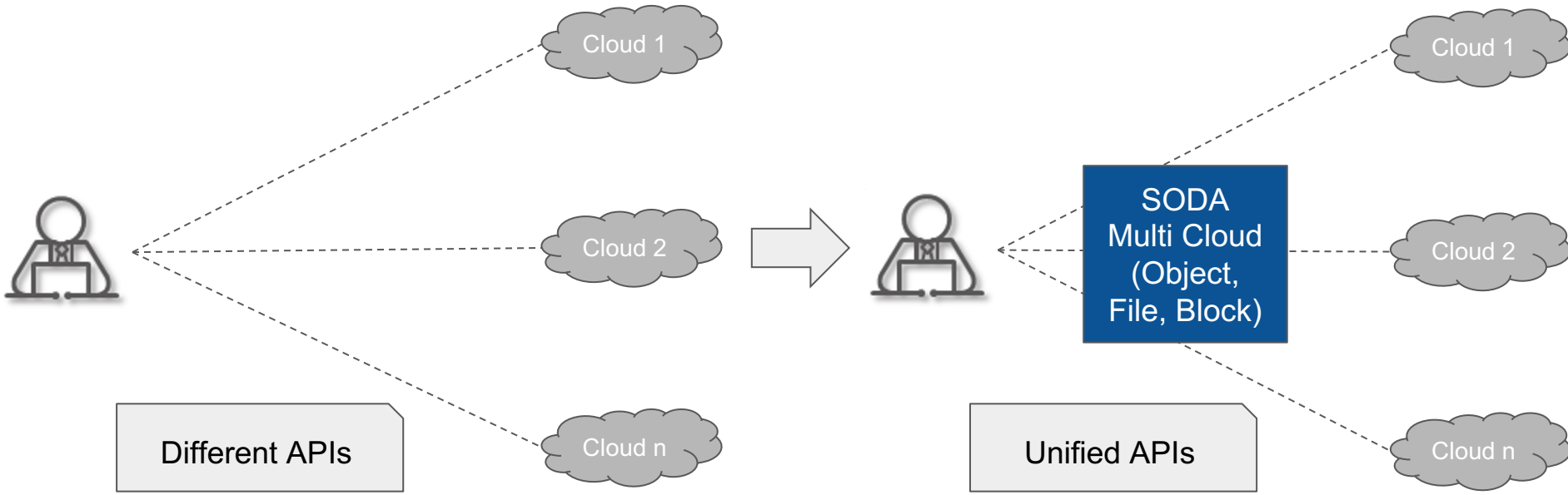
3

Multi Cloud Data Management Across Heterogeneous Cloud Vendors (and on prem)
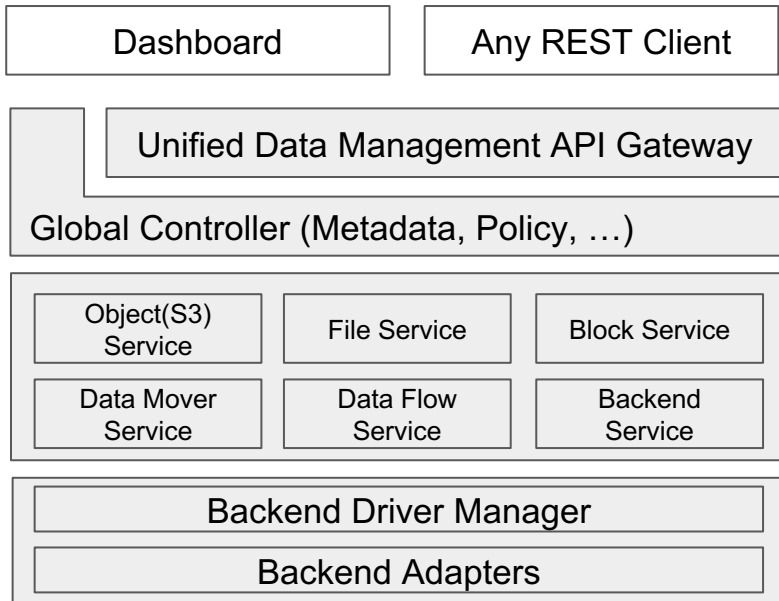
How does it work?

Illustration through SODA Multi Cloud Project

# ...Provides Unified APIs for Data Management Across Multiple Cloud Vendors

Supports Data Management for Object, File and Block

# ...supports hybrid cloud across On Premise and Multiple Cloud Vendors

- Provides a cloud vendor agnostic data management for hybrid cloud, intercloud or intracloud.
- REST based interface
- Global Configurations
- Supports Object, File and Block
- Policy Based Data Management:
  - Data Migration
  - Data Lifecycle Management
- Heterogeneous On Prem:
  - Ceph and YIG Ceph
  - SODA On Prem Connectivity for heterogeneous storages (in progress)
- S3 Compatible Object API
- Pluggable Adapter Model for Backends (Easy to add backends)
- Security (Enhancing..)

Object Service
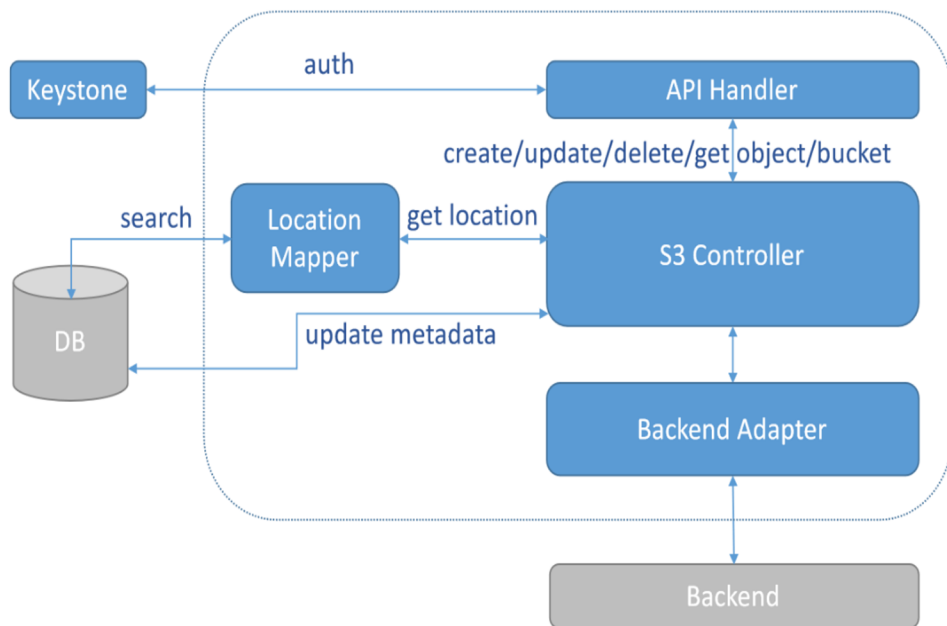
- Support Authentication with Keystone.
- Supports S3 Compatible APIs.
- S3 Controller serves as a bucket & object metadata placeholder.
- A bucket in SODA Multi-Cloud is a virtual bucket that is mapped to a physical bucket in the object store on premise or in the cloud.
- Location Mapper is responsible 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.

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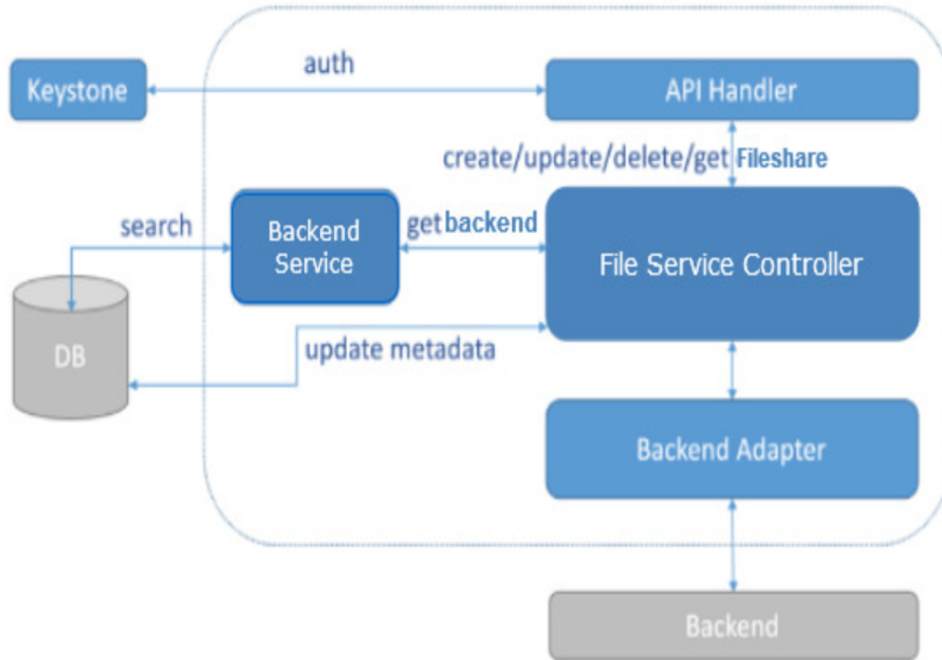
- Support Authentication with Keystone.
- Cloud File Share or Filestore Management support
- File Service Controller serves as a metadata placeholder for Cloud File/Filestore Resources.
- File Service support Create/ Update/Delete/Get a Fileshare.
- Support Cloud Vendors including AWS, Azure, GCP, Huawei.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

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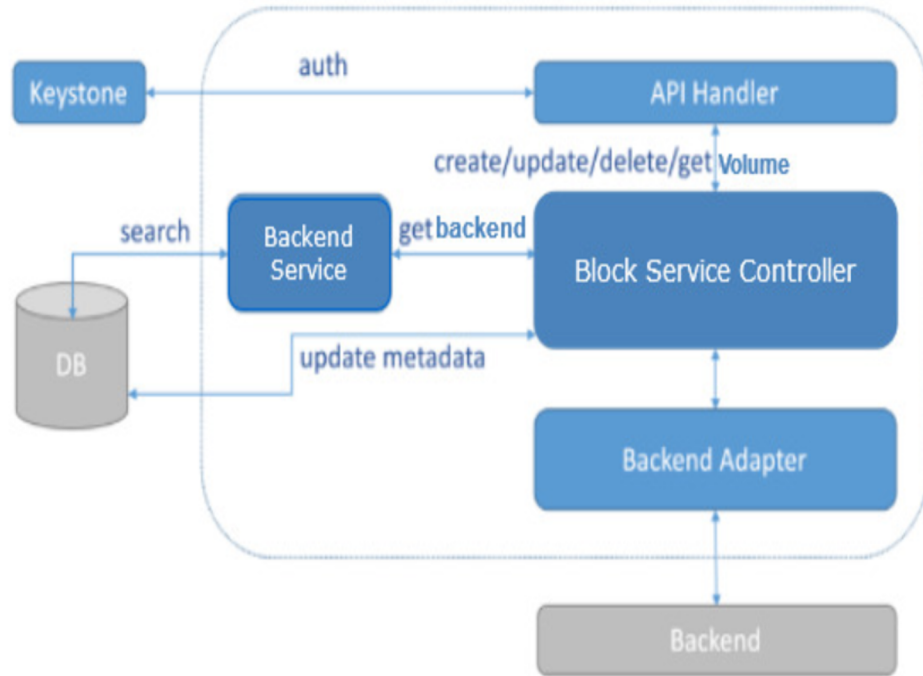
- Support Authentication with Keystone.
- Cloud Volumes/Block Device Management support
- Block Service Controller serves as a metadata placeholder for Cloud Volume Resources.
- Block Service support Create/Update/Delete/Get a Volume.
- Support Cloud Vendors like AWS, Huawei.(Adding More)

Backend Service
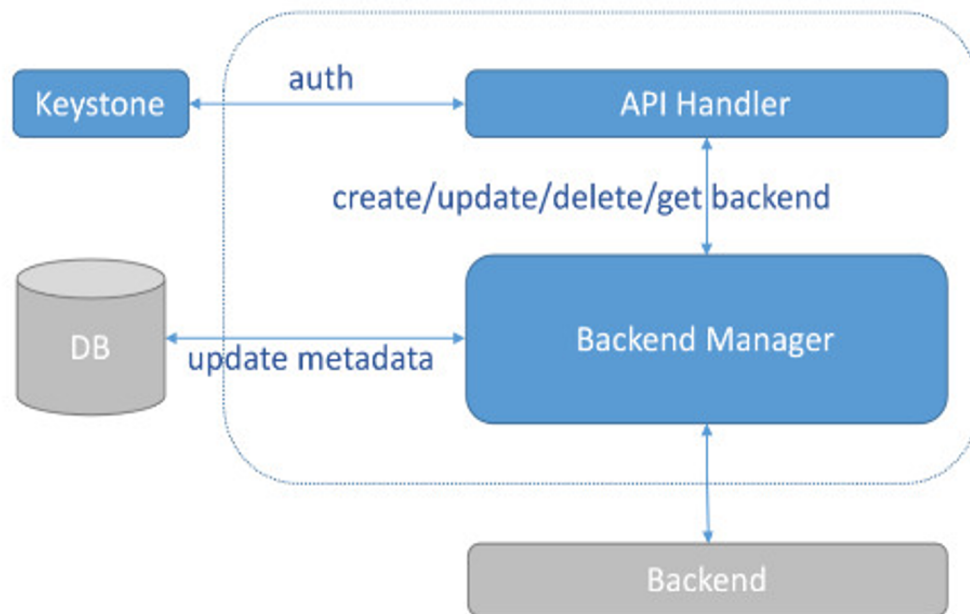
- Support Authentication with Keystone.
- Backend Manager serves as a metadata placeholder for Cloud Backend profiles.
- Backend Service Manager support Create/ Update/Delete/Get a backend.
- Supports object/blob, block, file storage types
- Support Cloud Backend Profile Registration
- Adapter Model (easy to add backends)

Data Mover / Flow Service
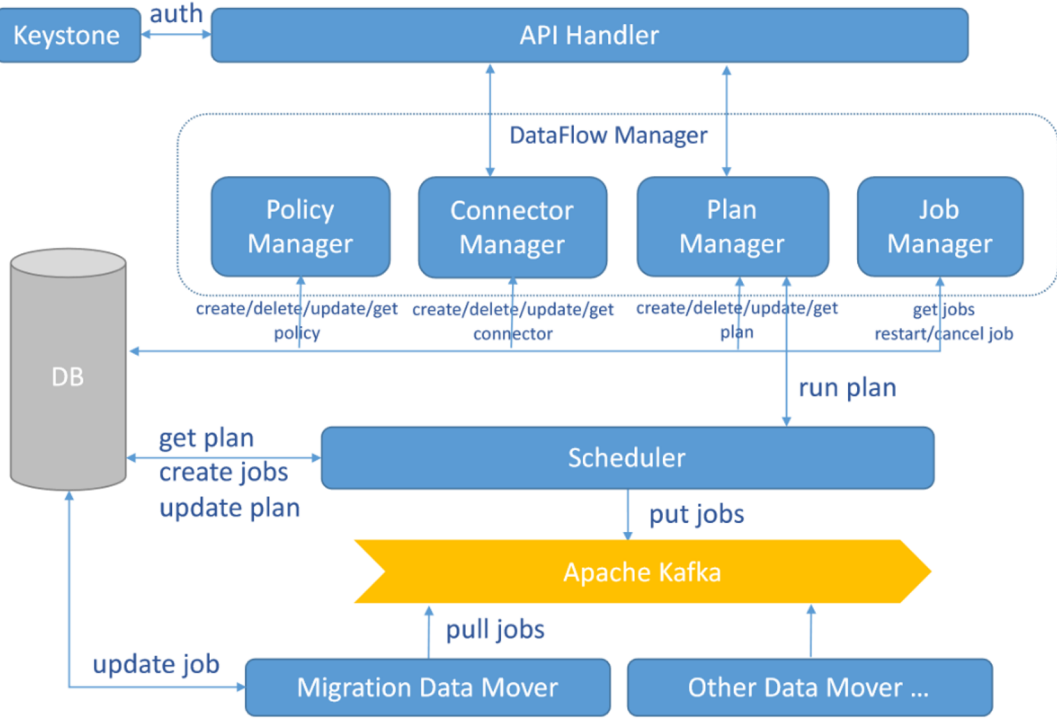
- Support Authentication with Keystone.
- Dataflow Manager serves as a Migration Plan, Policy, Job & Connectors metadata placeholder.
- Policy Manager support Create/Update/Delete/Get a Migration Policy.
- Plan Manager support Create/Update/Delete/Get a Migration Plan.
- Job Manager support Create/Update/Delete/Get 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. (Pluggable)
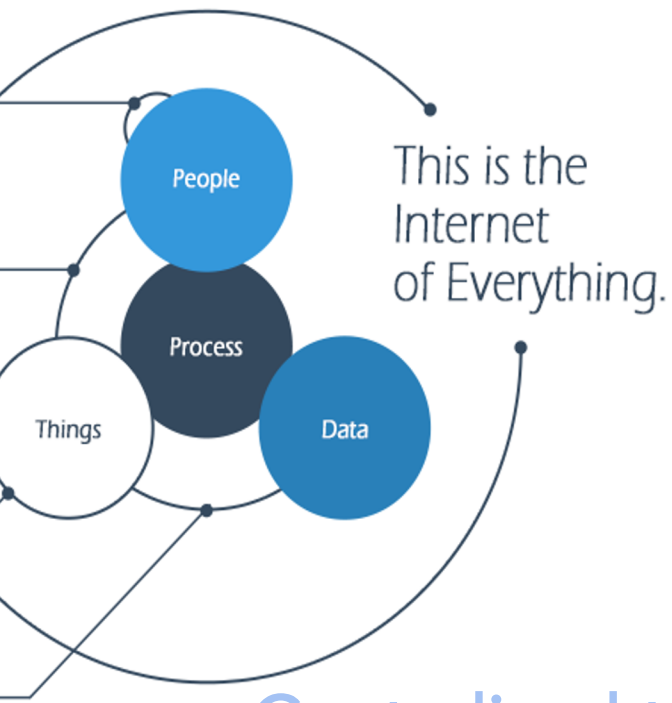
Let us see a demo...!

Multi Cloud Data Migration Managed from SODA Dashboard

https://www.youtube.com/watch?v=XCIuIIm11zw

..moving to edge...data@edge!

This is the
Internet
of Everything.

People
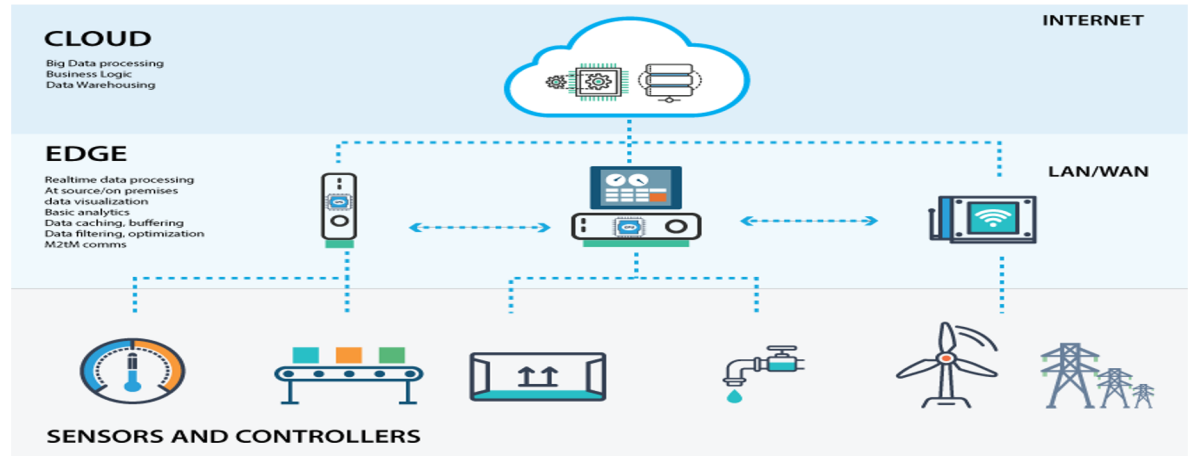
Process

Things

Data

Moving to Edge!

Centralized to Decentralized to Distributed

Edge

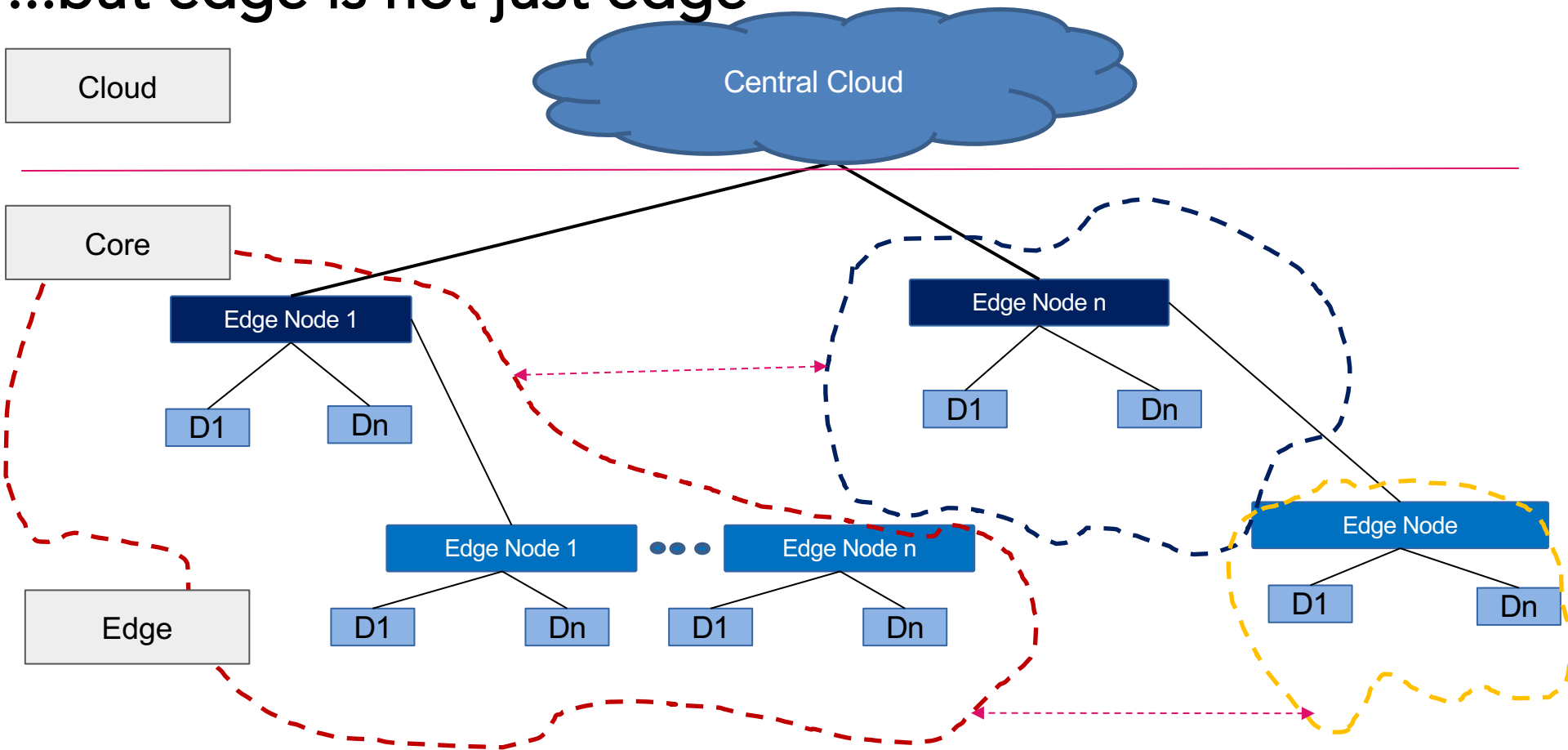
..Alexa, show me the features of SODA Framework…!

..want the response now!
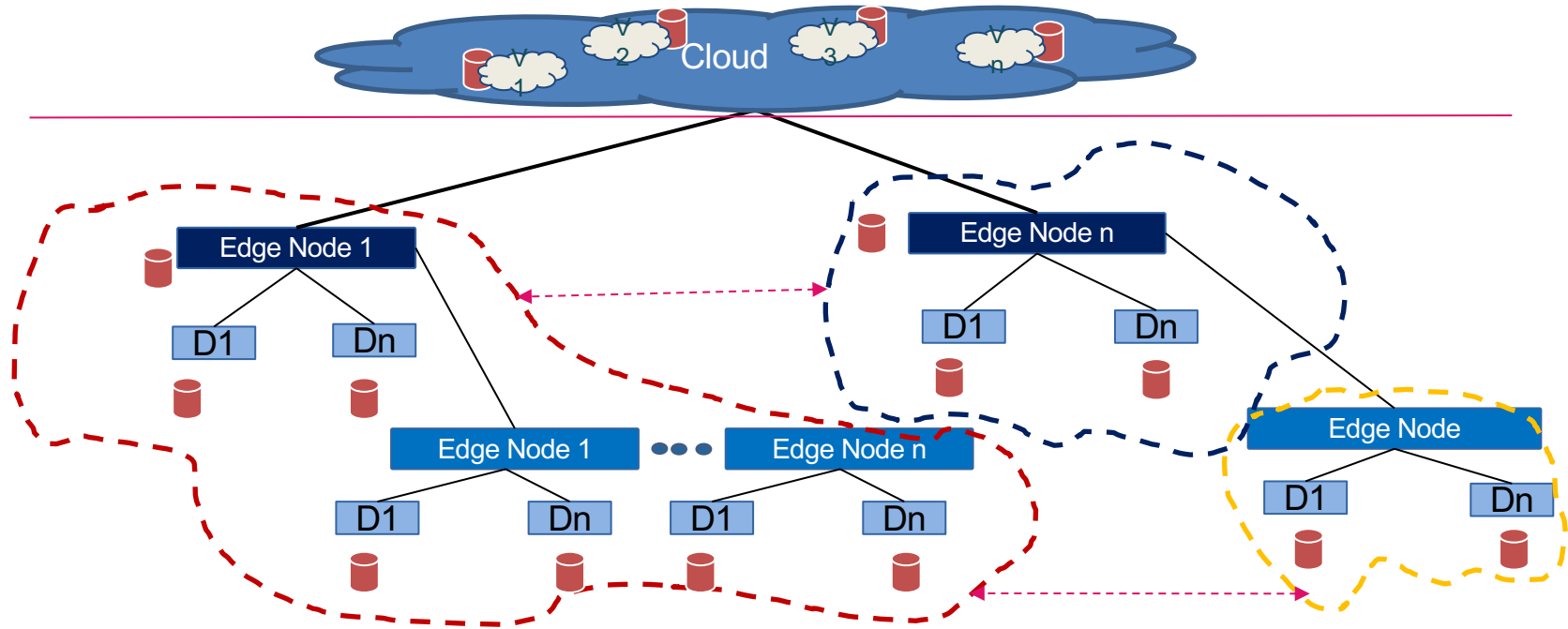
..so, let us process the data closer to the source of the data (edge)

…but edge is not just edge

Cloud

Central Cloud

Core

Edge Node 1

D1   Dn

Edge Node n

D1   Dn

Edge Node 1

D1   Dn

Edge Node n

D1   Dn

Edge Node

D1   Dn

Edge

...then data everywhere...!

...can we manage most of it at edge..?

...but..?!

Multi-Source	**Heterogeneous Data**
Distributed...Really!	**Data Consistency**
Data Ownership	**Low Latency**
Data Mobility	**Offline Scenarios**

So
Many
Challenges…!

…finding a way..

Distributed & Heterogeneous Data Management Platform@Edge
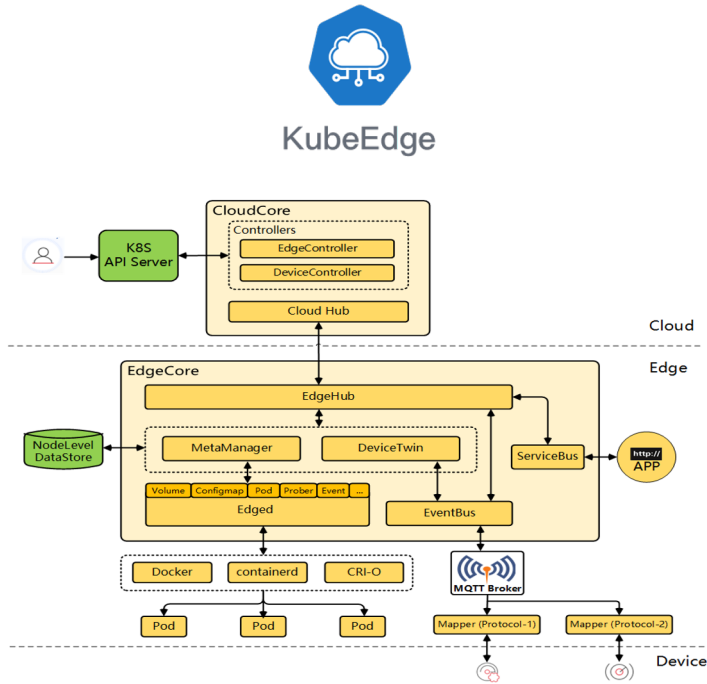
Unified Data Framework
Open
Vendor Agnostic
Platform Agnostic
Distributed
Low Resource
Extensible or Shrinkable
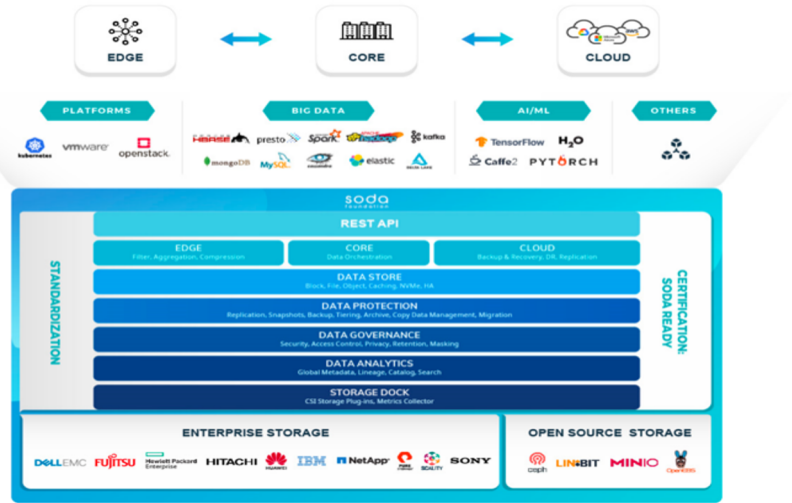Standardized

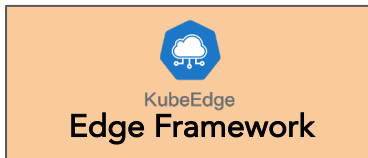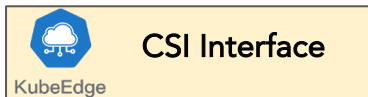**Edge Computing Platform** ➕ **Edge Data Management Platform**

..let us start somewhere...

KubeEdge

https://github.com/kubeedge

soda foundation

https://github.com/sodafoundation

...found compatible needed interface...CSI !

Edge Framework (KubeEdge)

Provides lightweight edge computing platform with compute, network and storage(csi) interfaces

CSI Interface (KubeEdge)

Container Storage Interface

CSI Plugin (soda foundation)

Single SODA CSI plugin which can support all the devices /drivers supported in SODA

Data Framework (soda foundation)

SODA Unified Heterogeneous Data Interface

Storage Backends (Diff types/vendors)

Heterogeneous Storages : different vendors, models, types, cloud storages...

https://github.com/sodafoundation/soda/releases/tag/v1.1.0
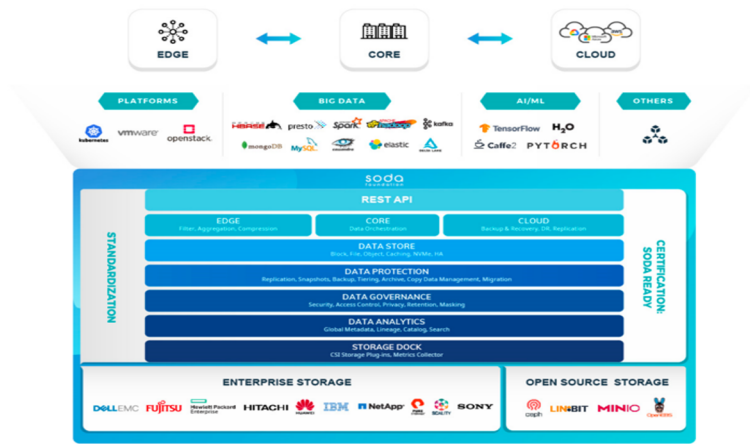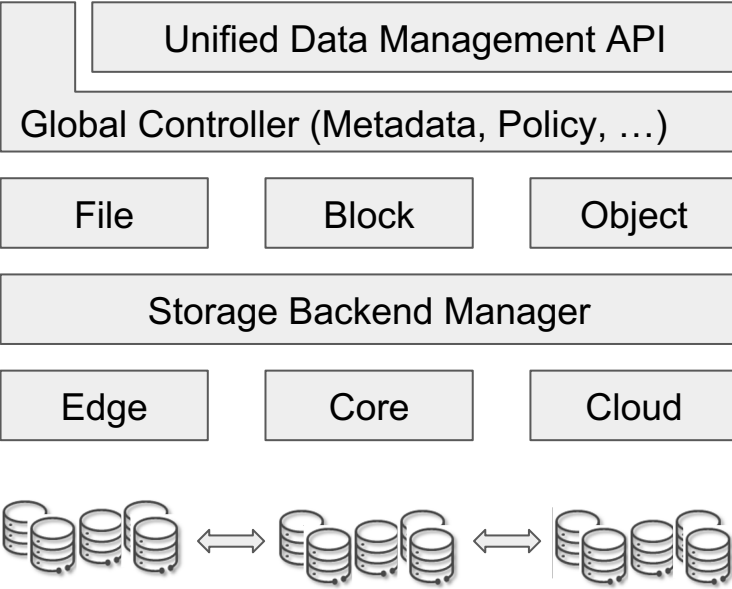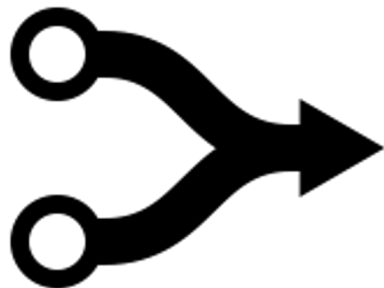
...way to go...

...SODA@Edge SIG

The SODA Story Continues...

Exploring One Open Data Framework

to realize infinite possibilities of
seamless hybrid cloud data management use cases...

Wanna Join?

SODA Github:
https://github.com/sodafoundation

Join SODA Slack:
https://sodafoundation.io/slack/

soda foundation

Thank You

 soda foundation

https://sodafoundation.io/

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

Join SODA Slack: https://sodafoundation.io/slack/
Follow SODA Twitter: https://twitter.com/sodafoundation
Join Us: https://sodafoundation.io/the-foundation/join/