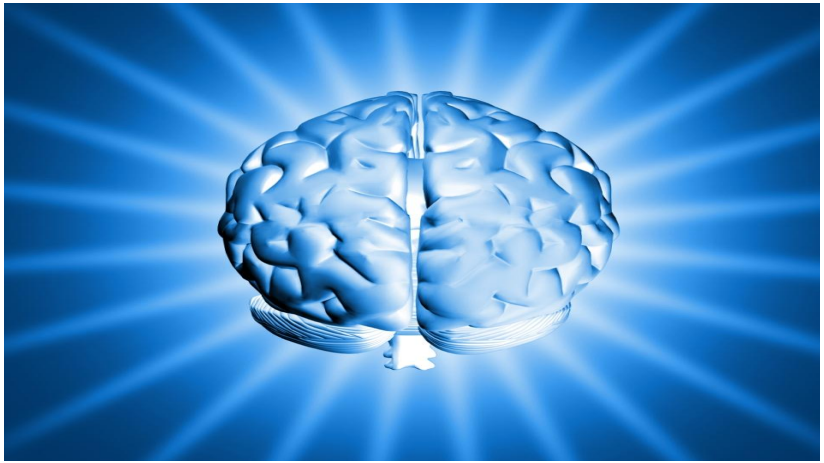
The Story of SODA Features...

Once upon a time...

soda foundation

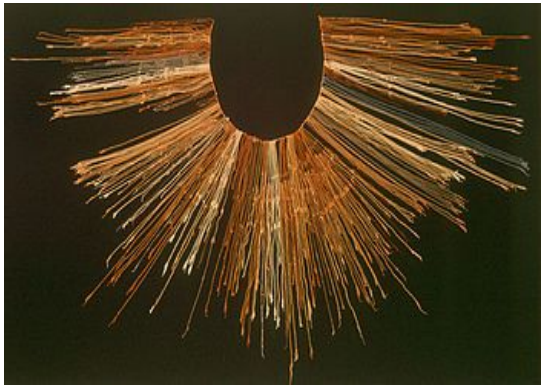
https://www.youtube.com/watch?v=B1hVx7BsvjU (3:30 - 4:25)

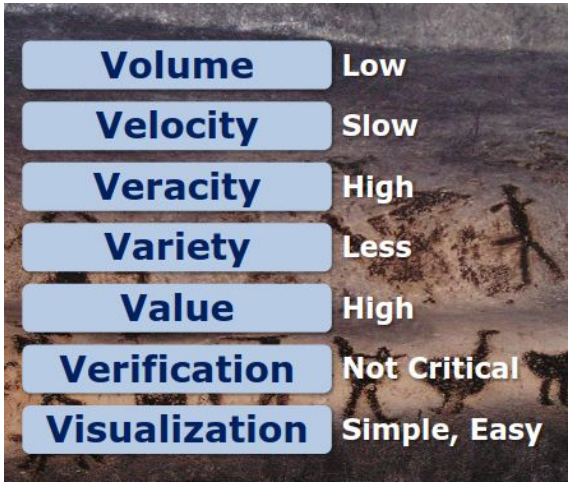
A Mankind is born with Big Data…

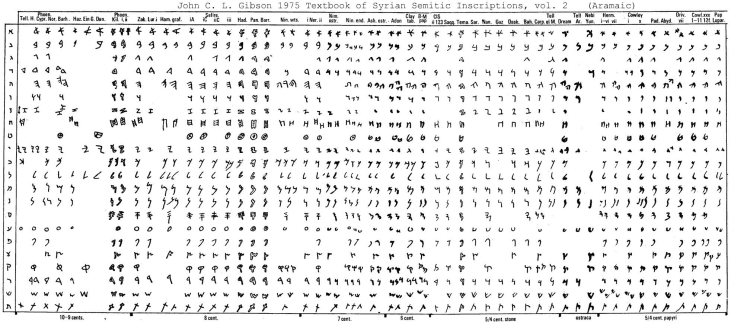


Stores 1 petabyte to 2.5 petabytes…

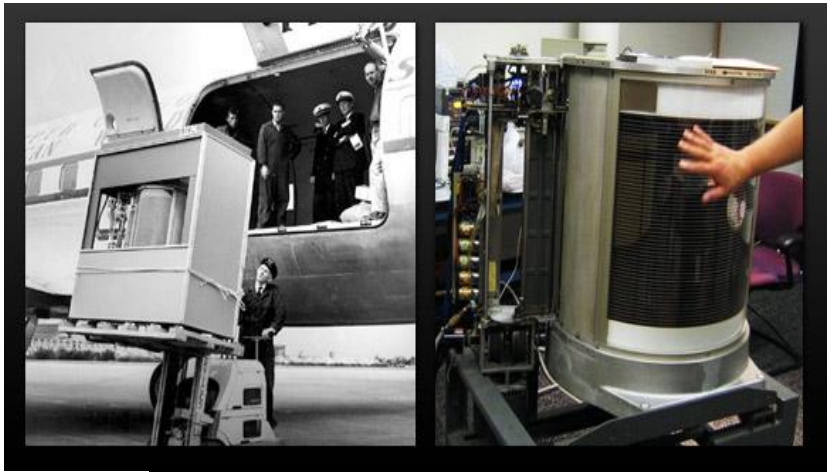
He started playing with Data…(unknowingly?)







But Data Started Growing...

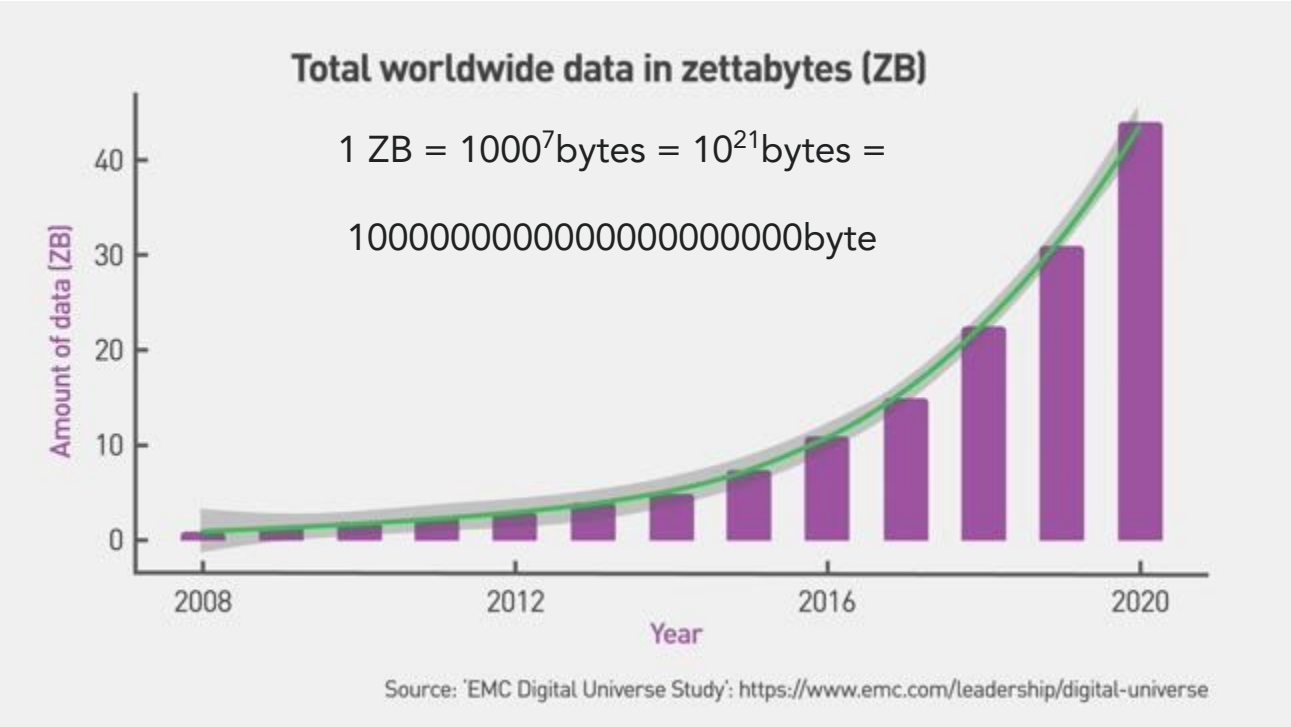


He invented better ways for Data...

..and kept generating tons of data

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

...Data Started Exploding...

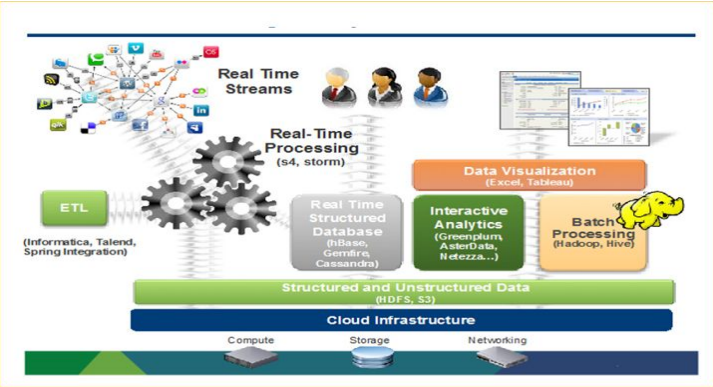


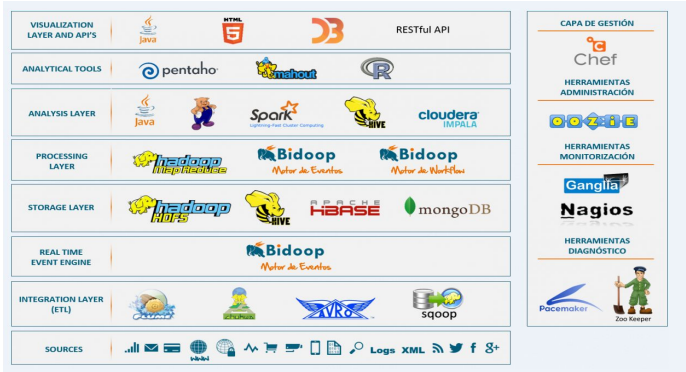
He kept innovating...

Different Storages, Different Platforms

Different Ways for Managing his Data

...Simple to Complex..

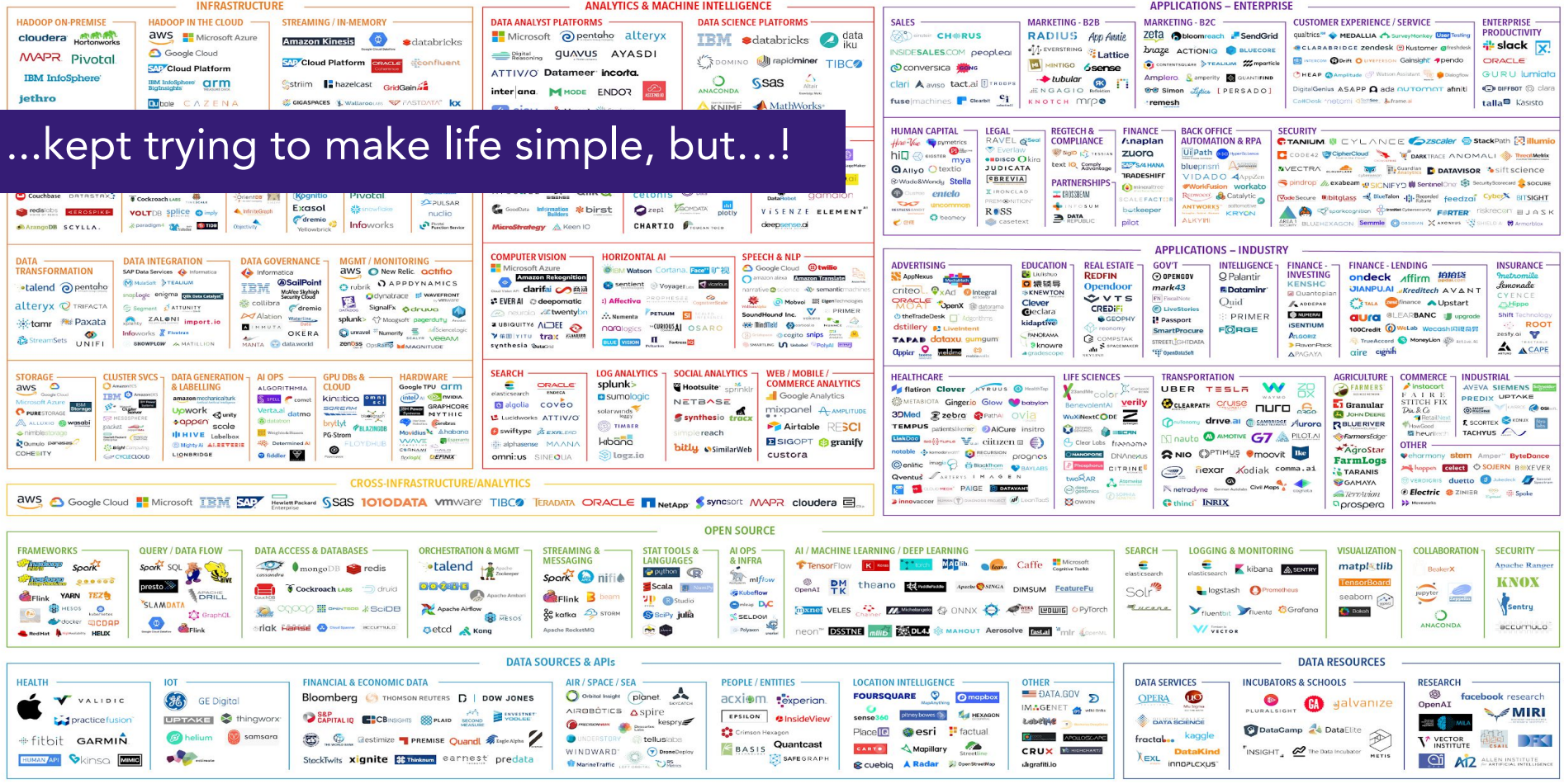




kubernetes

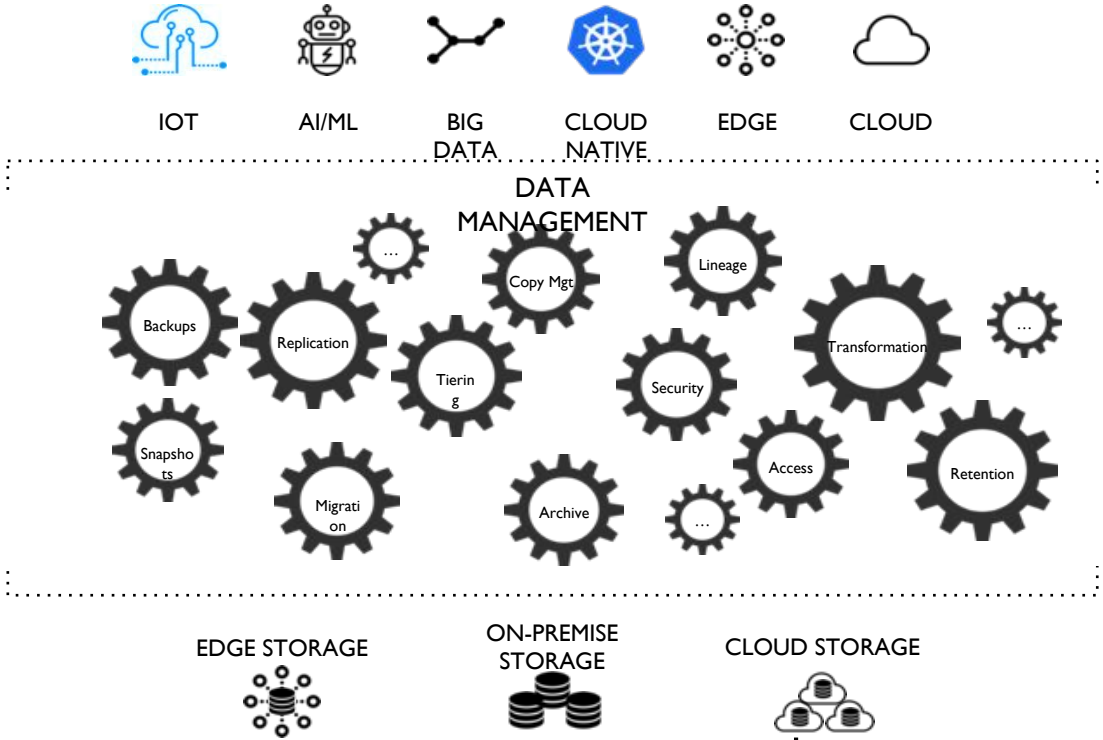
DATA & AI LANDSCAPE 2019

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



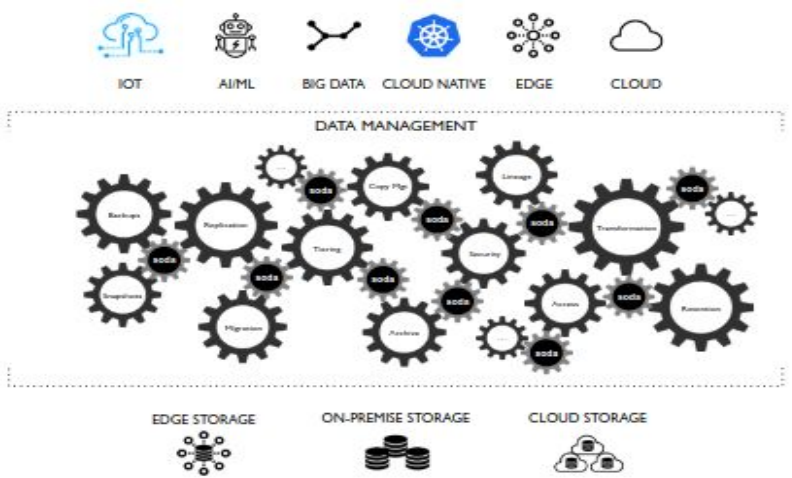
FIRSTMARK
EARLY STAGE VENTURE CAPITAL

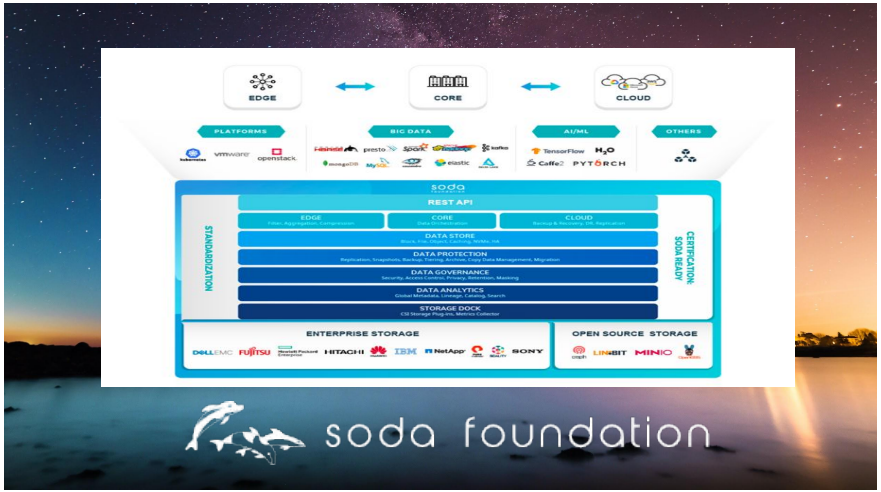
...Fragmented Data Management



...he wants it to be simpler, so..

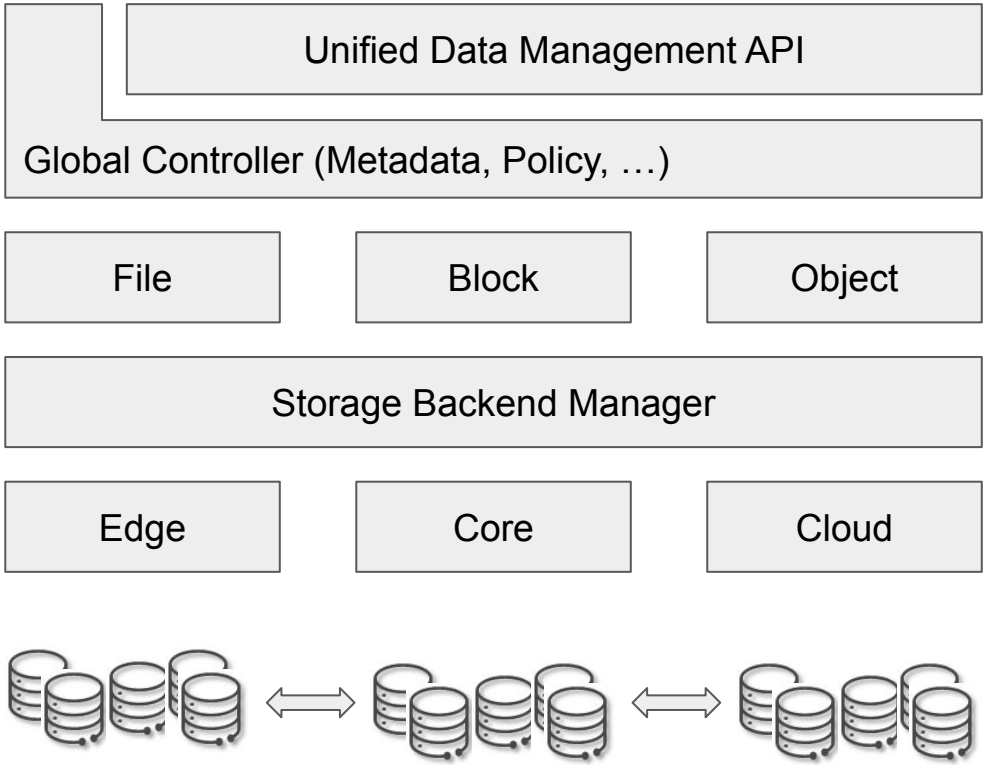
One Data Framework Infinite Possibilities





He started dreaming...

One Data Store



He started the journey...

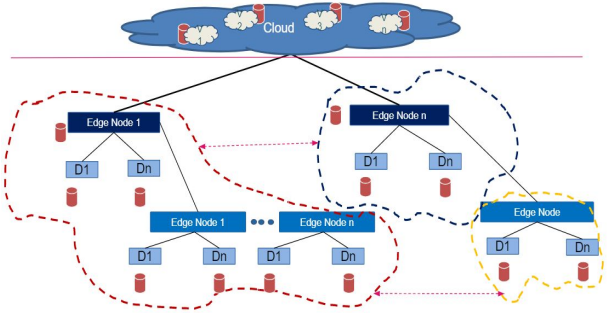
On-premise/Core

Heterogeneous
Storages and
Platforms

Cloud

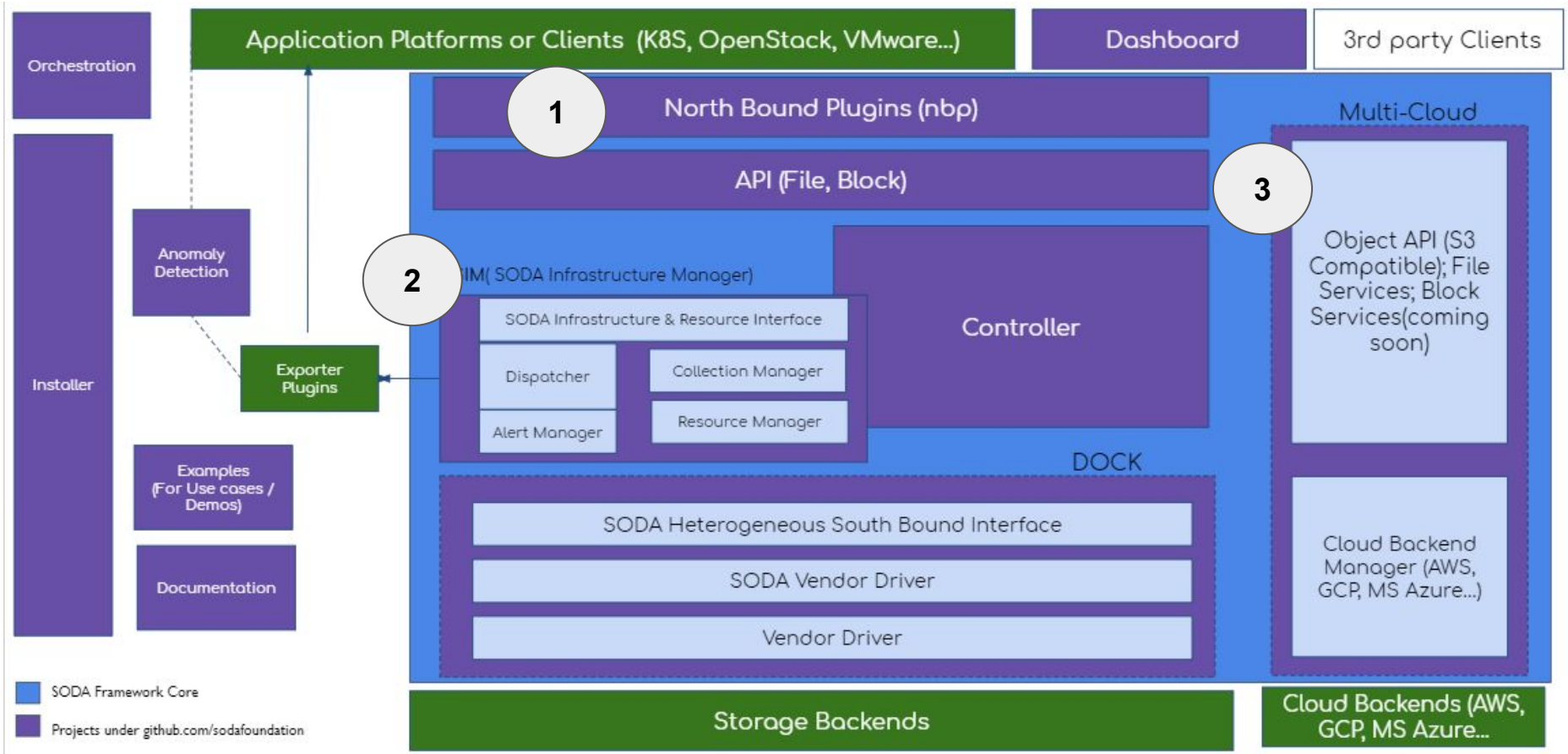
Multiple Cloud
Data Mobility

Edge



Distributed
Heterogeneous
Low Resource
Data Mobility

..with on-premise and cloud

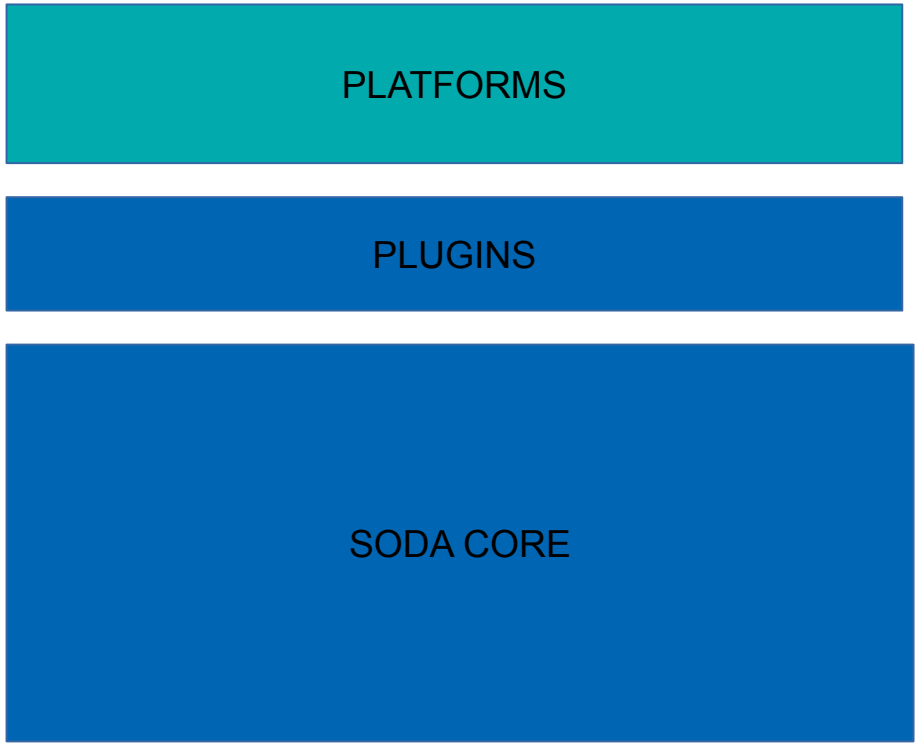


1

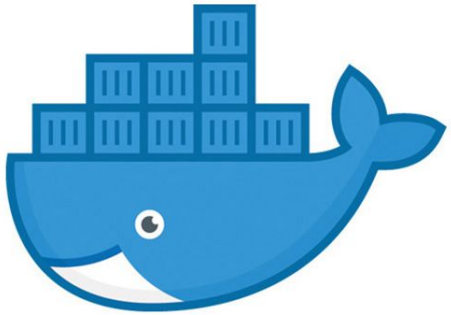
..tackling heterogeneous storages and platforms on-premise

Have different platforms??

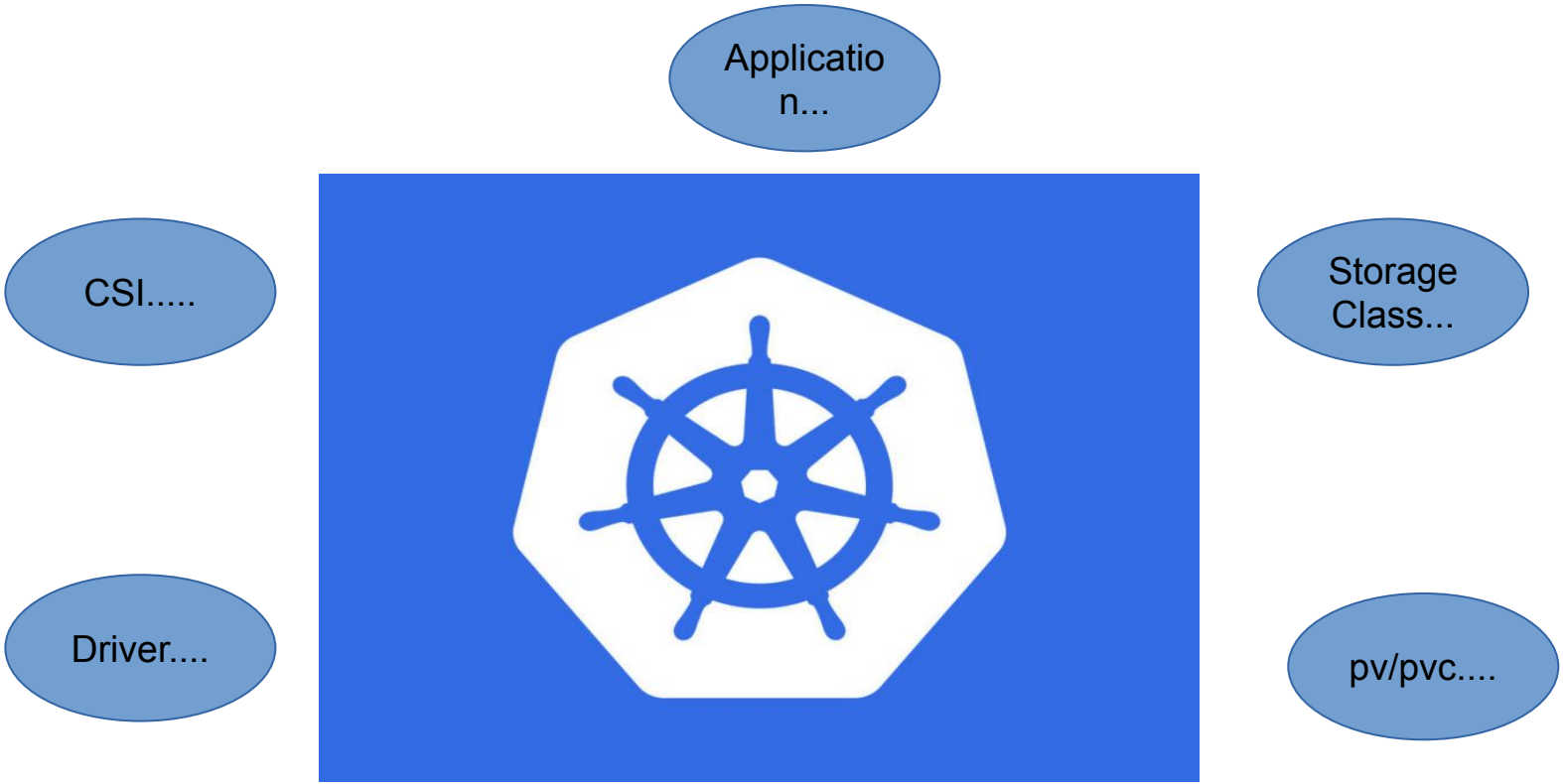
Then how !!



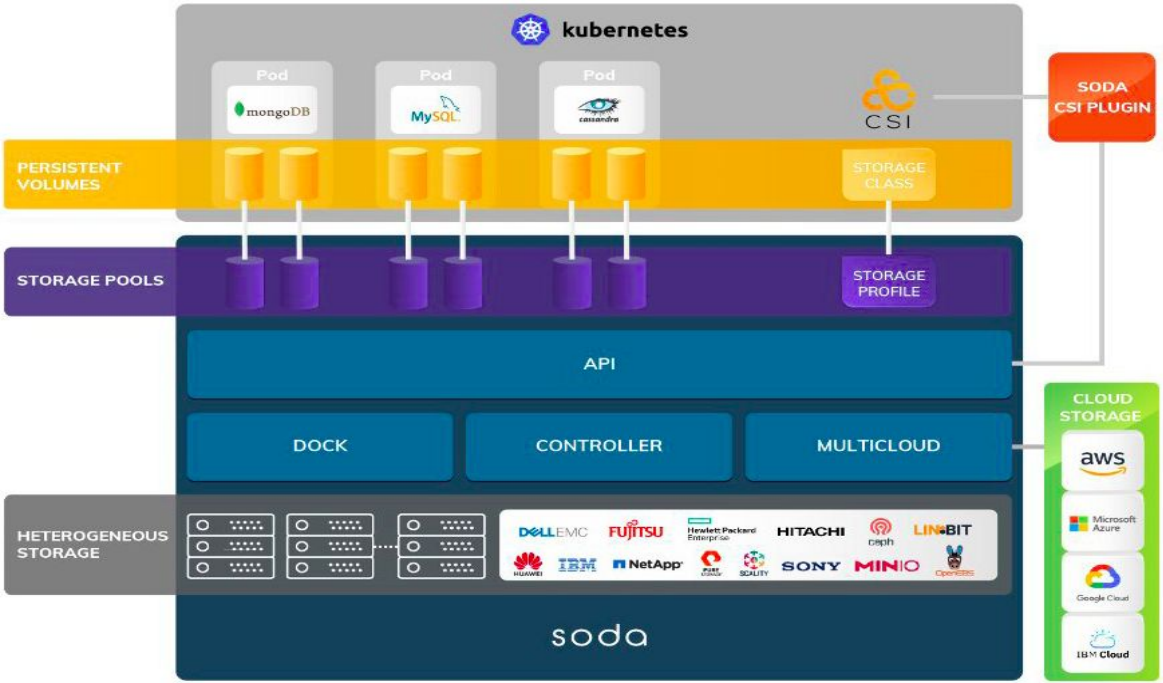
...moving to a containerized world



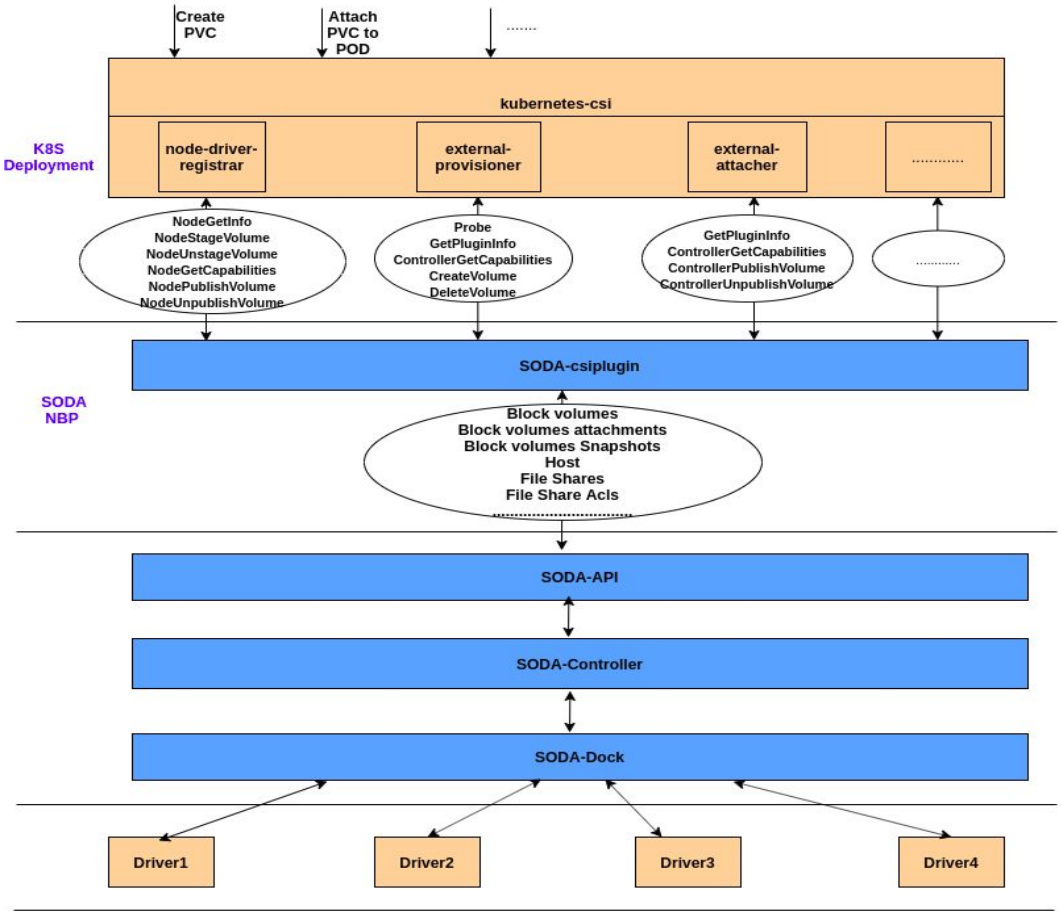
Using K8S??



Kubernetes to SODA, along with others!



Zoom In...


Create PVC
Attach PVC to POD
.....
kubernetes-csi
K8S Deployment
node-driver-registrar
external-provisioner
external-attacher
.....
NodeGetInfo
NodeStageVolume
NodeUnstageVolume
NodeGetCapabilities
NodePublishVolume
NodeUnpublishVolume
Probe
GetPluginInfo
ControllerGetCapabilities
CreateVolume
DeleteVolume
GetPluginInfo
ControllerGetCapabilities
ControllerPublishVolume
ControllerUnpublishVolume
.....
SODA-csiplugin
SODA NBP
Block volumes
Block volumes attachments
Block volumes Snapshots
Host
File Shares
File Share Acls
.....
SODA-API
SODA-Controller
SODA-Dock
Driver1
Driver2
Driver3
Driver4

Deployment view...

```
root@root123-VirtualBox:~/gopath/src/github.com/sodafoundation/installer/ansible# kubectl get pods | grep block
csi-attacher-opensdsplugin-block-0         3/3     Running   0          37m
csi-nodeplugin-opensdsplugin-block-mvp8g   2/2     Running   0          37m
csi-provisioner-opensdsplugin-block-0      2/2     Running   0          37m
csi-snapshotter-opensdsplugin-block-0      2/2     Running   0          37m
root@root123-VirtualBox:~/gopath/src/github.com/sodafoundation/installer/ansible#
root@root123-VirtualBox:~/gopath/src/github.com/sodafoundation/installer/ansible# kubectl get pods | grep file
csi-attacher-opensdsplugin-file-0          3/3     Running   0          37m
csi-nodeplugin-opensdsplugin-file-bj68v    2/2     Running   0          37m
csi-provisioner-opensdsplugin-file-0       2/2     Running   0          37m
csi-snapshotter-opensdsplugin-file-0       2/2     Running   0          37m
```

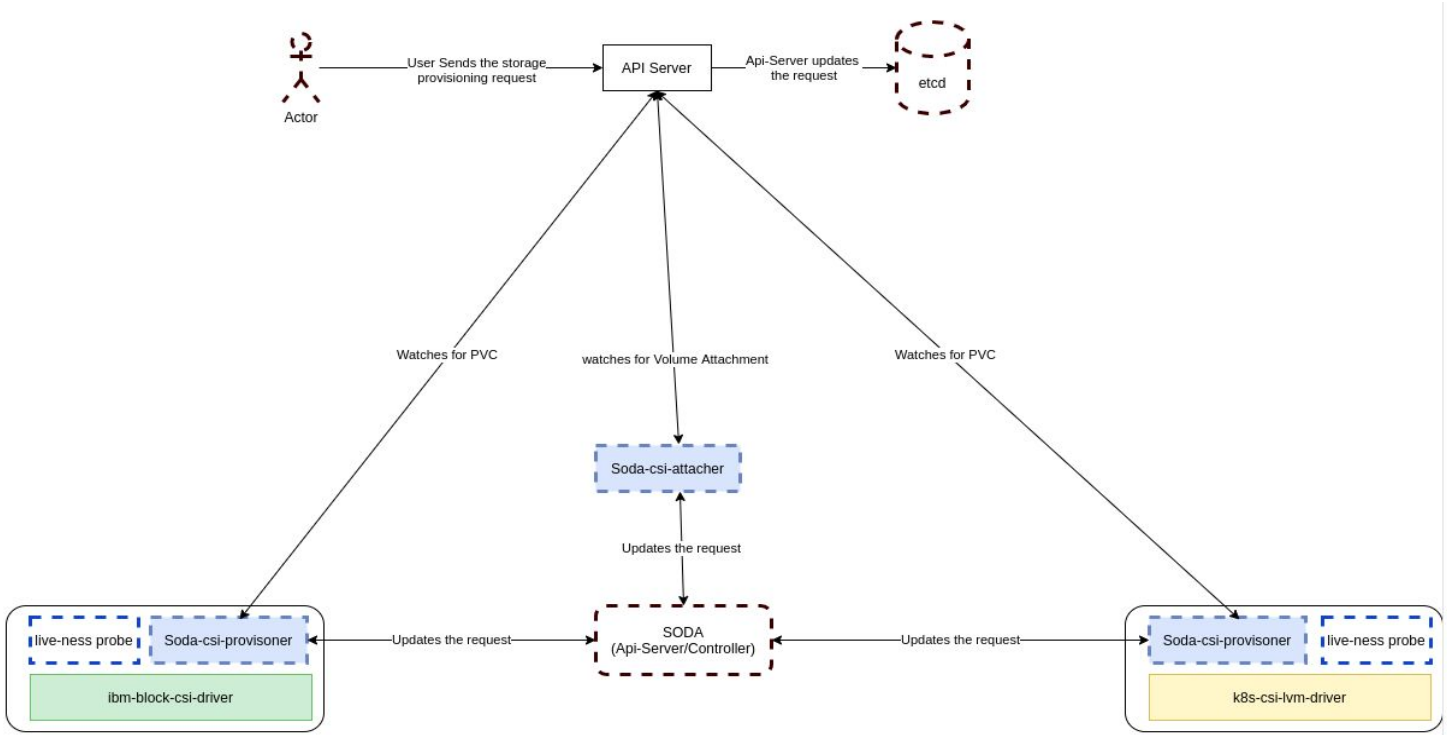
```
apiVersion: storage.k8s.io/v1
kind: StorageClass
metadata:
  name: csi-sc-opensdsplugin-block
provisioner: csi-opensdsplugin-block
parameters:
  attachMode: rw
  profile: 7e4d039f-6404-4d5b-956e-67be335fad8f
allowedTopologies:
- matchLabelExpressions:
  - key: topology.csi-opensdsplugin-block/zone
    values:
    - default
```

```
root@root123-VirtualBox:/opt/opensds-sushi-linux-amd64# kubectl get pods | grep nginx-block
nginx-block                                1/1     Running   0          163m
root@root123-VirtualBox:/opt/opensds-sushi-linux-amd64#
root@root123-VirtualBox:/opt/opensds-sushi-linux-amd64#
root@root123-VirtualBox:/opt/opensds-sushi-linux-amd64# kubectl get pvc
NAME                          STATUS   VOLUME                                     CAPACITY   ACCESS MODES   STORAGECLASS                AGE
csi-pvc-opensdsplugin-block   Bound    pvc-381f5c2c-0969-11eb-a3ac-080027310244   1Gi        RWX            csi-sc-opensdsplugin-block   163m
root@root123-VirtualBox:/opt/opensds-sushi-linux-amd64#
root@root123-VirtualBox:/opt/opensds-sushi-linux-amd64# kubectl get pv
NAME                                       CAPACITY   ACCESS MODES   RECLAIM POLICY   STATUS   CLAIM                                 STORAGECLASS                REASON   AGE
pvc-381f5c2c-0969-11eb-a3ac-080027310244   1Gi        RWX            Delete           Bound    default/csi-pvc-opensdsplugin-block   csi-sc-opensdsplugin-block           163m
root@root123-VirtualBox:/opt/opensds-sushi-linux-amd64#
root@root123-VirtualBox:/opt/opensds-sushi-linux-amd64#
root@root123-VirtualBox:/opt/opensds-sushi-linux-amd64# osdsctl volume list
WARNING: Not found Env OPENSDS_AUTH_STRATEGY, use default(noauth)
+--------------------------------------+--------------------------------------------+-------------+------+------------------+--------+--------------------------------------+
| Id                                   | Name                                       | Description | Size | AvailabilityZone | Status | ProfileId                            |
+--------------------------------------+--------------------------------------------+-------------+------+------------------+--------+--------------------------------------+
| 9ca997a4-1072-481d-adeb-da767b86c3dc | pvc-381f5c2c-0969-11eb-a3ac-080027310244 |             | 1    | default          | inUse  | 7e4d039f-6404-4d5b-956e-67be335fad8f |
+--------------------------------------+--------------------------------------------+-------------+------+------------------+--------+--------------------------------------+
root@root123-VirtualBox:/opt/opensds-sushi-linux-amd64#
root@root123-VirtualBox:/opt/opensds-sushi-linux-amd64#
```

So many storages...

so many csi drivers...

..can we just plug and play?!



..next on containerised world?

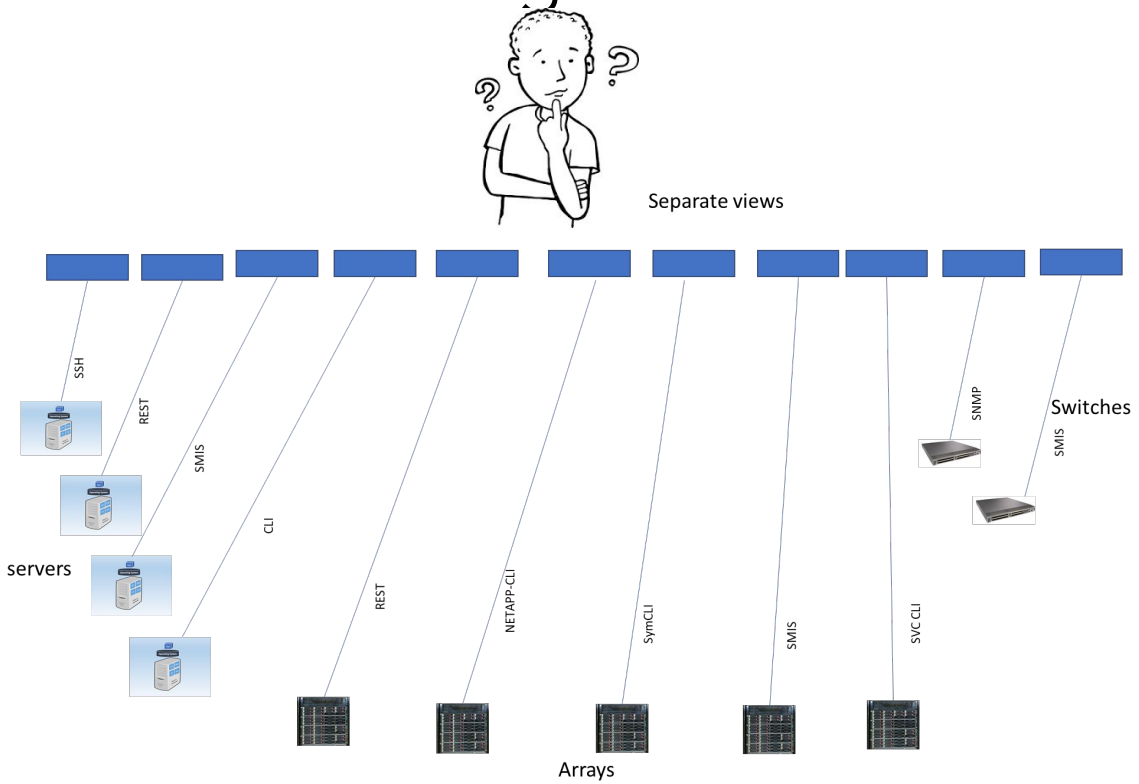
- Kubernetes and SODA with cloud backends
- PlunNPlay - kickstarted, moving ahead
- Performance benchmarking for reference deployment and usecases

2

..tackling the intelligent monitoring at scale

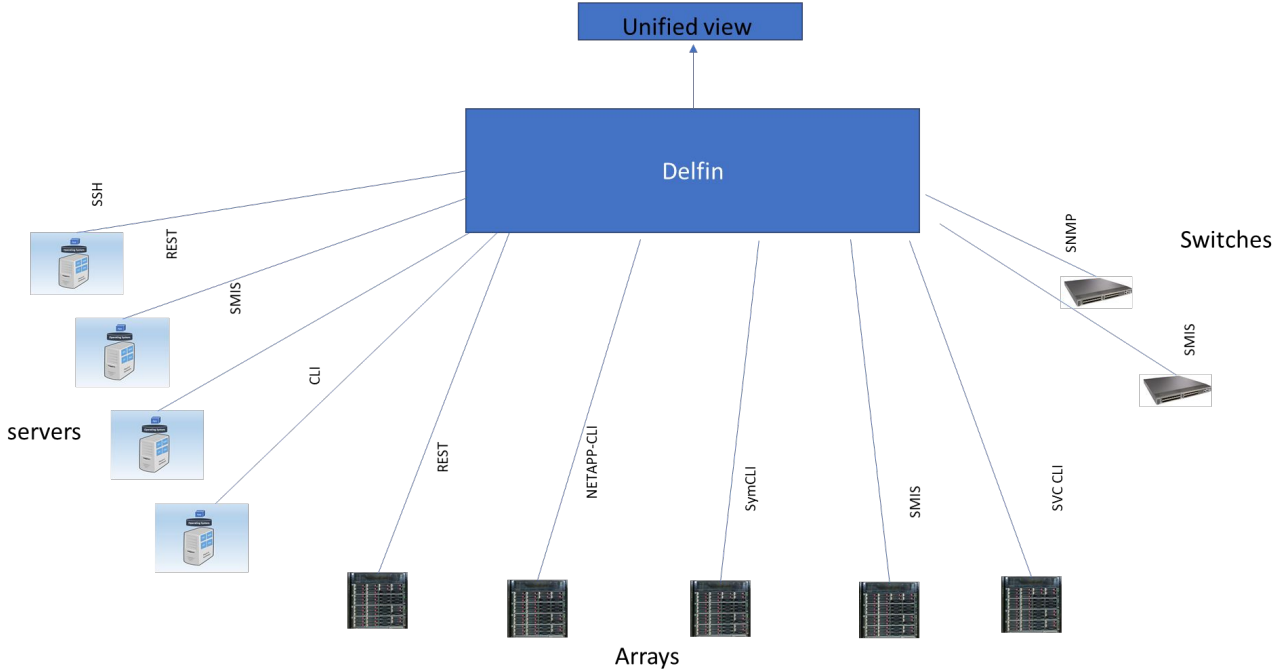
...world without unified monitoring...

- Discrete and Unconnected!!
- No Unification!
- No Common APIs!



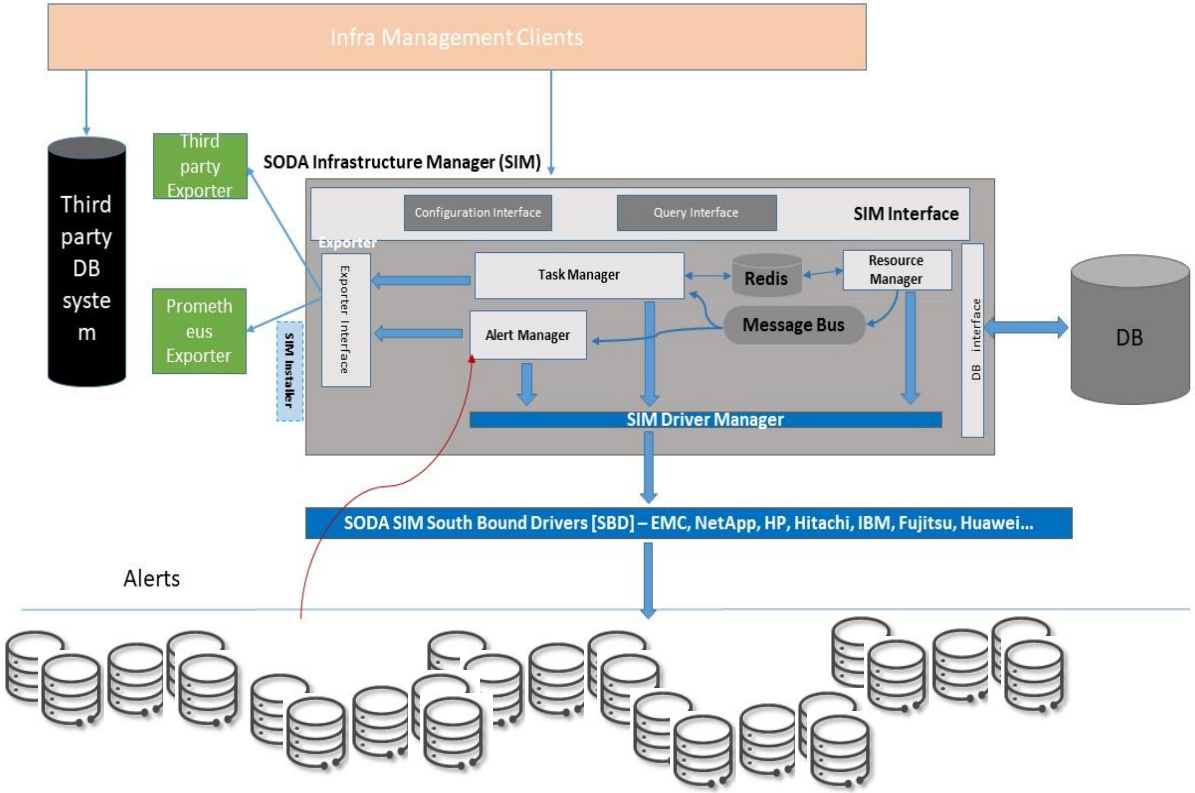
...but we have beautiful Delfin 😊

- Unified
- Intelligent
- Automated
- Extensible



Delfin...?

- SODA foundation project for unified heterogeneous resource and performance monitoring and alerting
- Drivers to collect data
- Exporters to push data
- A scalable architecture to add drivers and exporters.



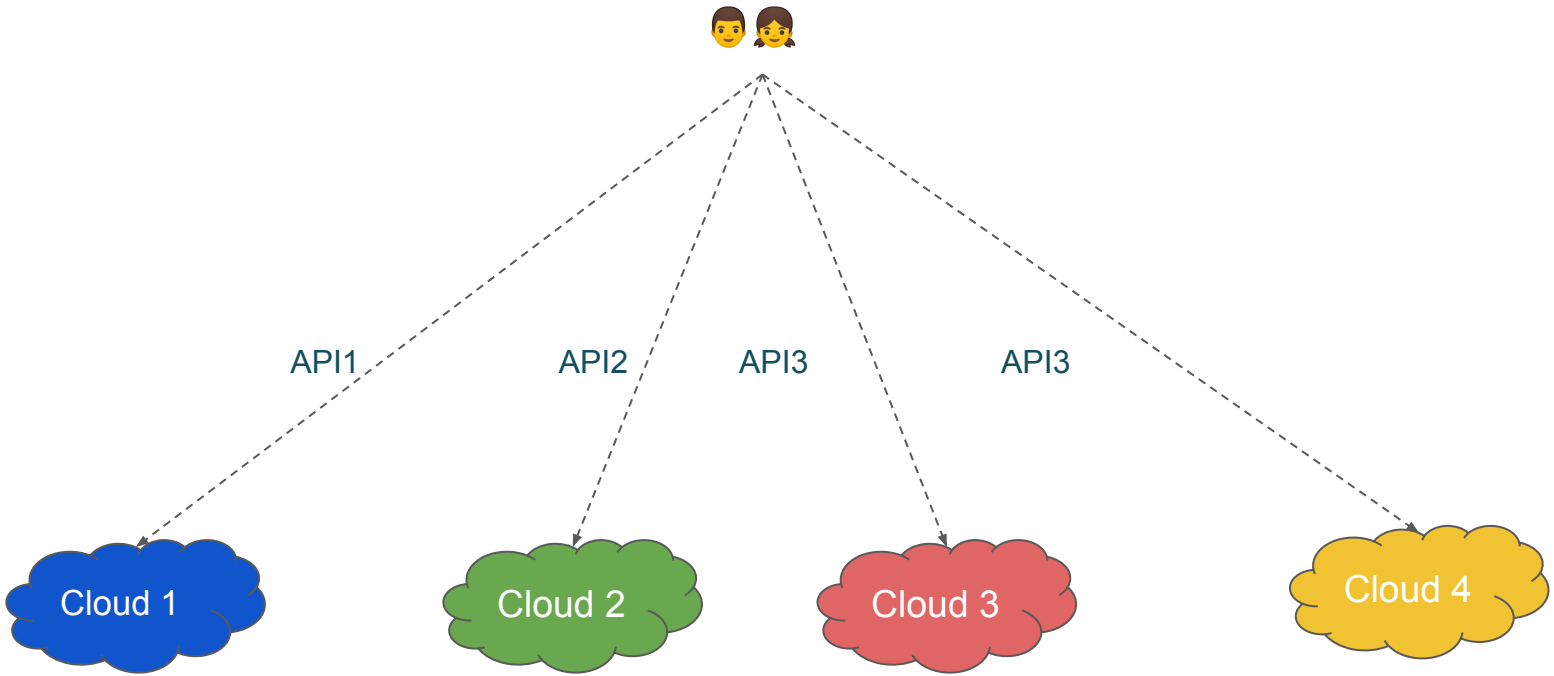
...one monitoring, what next?

- More Storage backend support...
- More resources (Network , Compute,Applications ….)
- Intelligent Monitoring….
- Plug and Play Integrations….
- Connecting SODA Provisioning, Analysis, Orchestration and ...

3

..tackling the multiple cloud data management

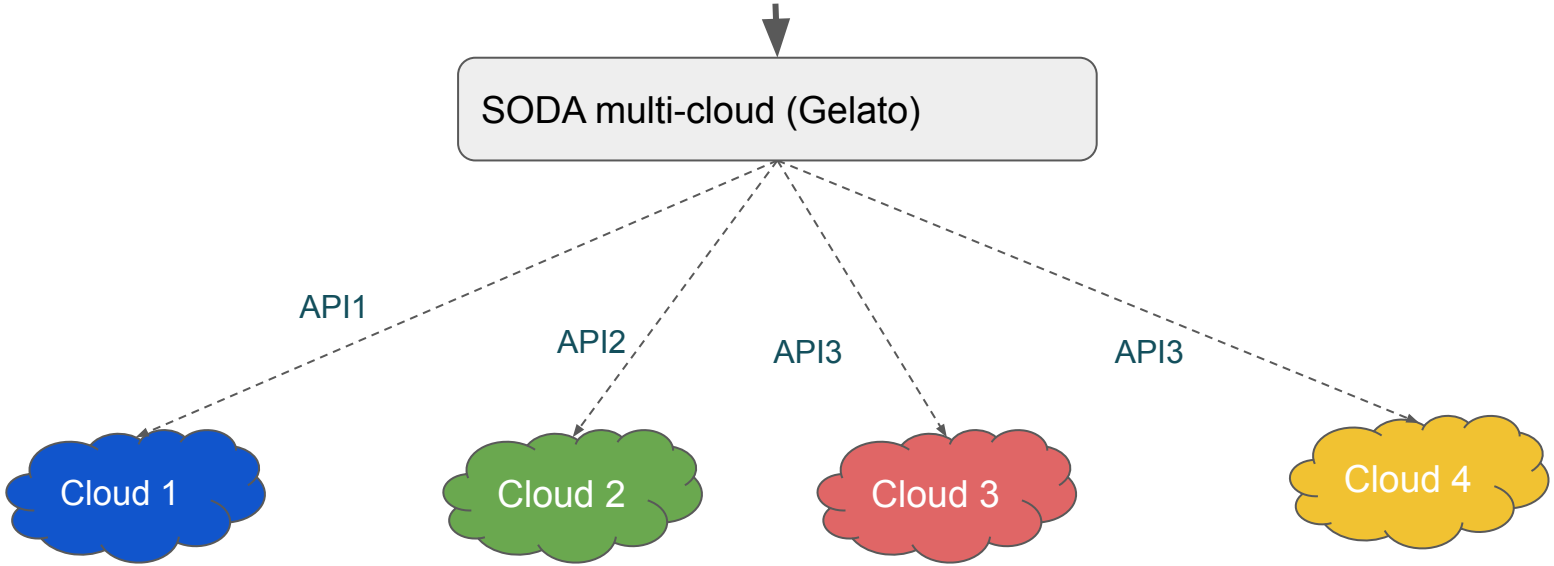
Before…..



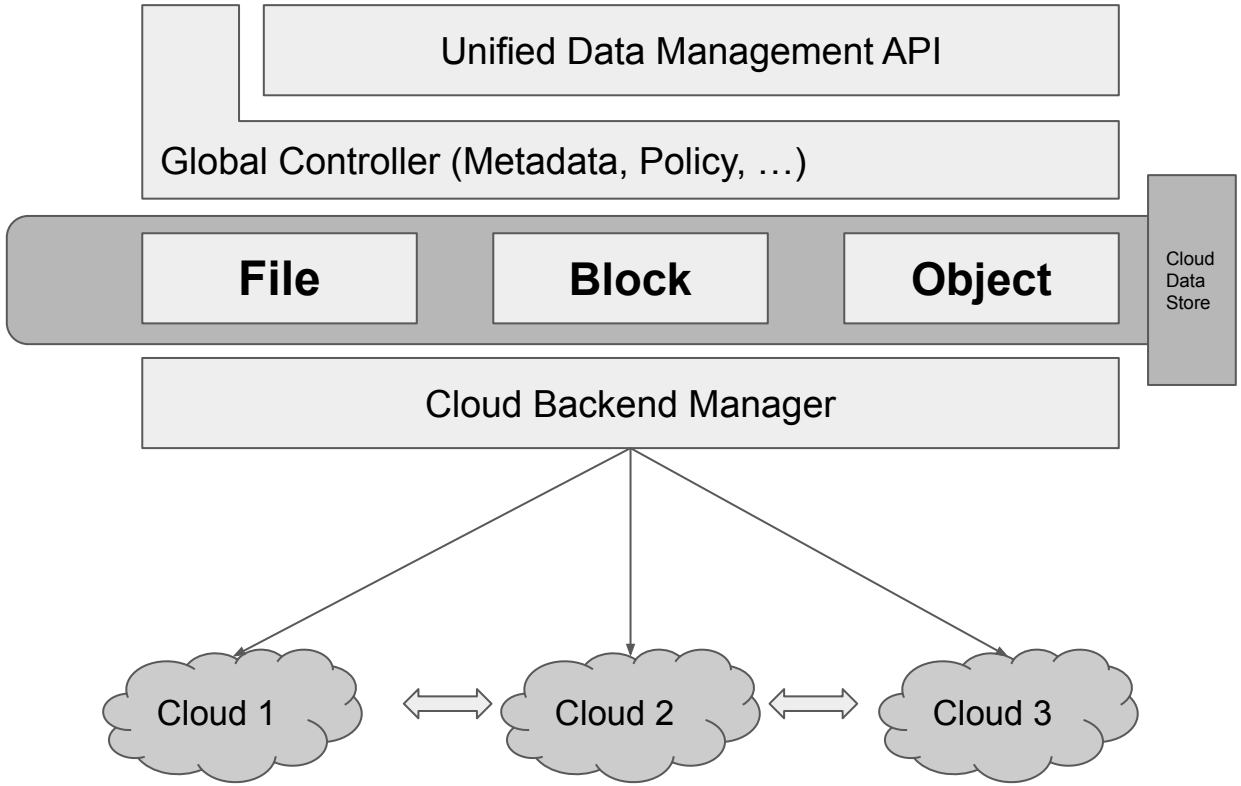
Let’s make Life Sweeter....

I am Loving it……

After.. The Story is sweeter…..



And HE envisioned it… One Data Store

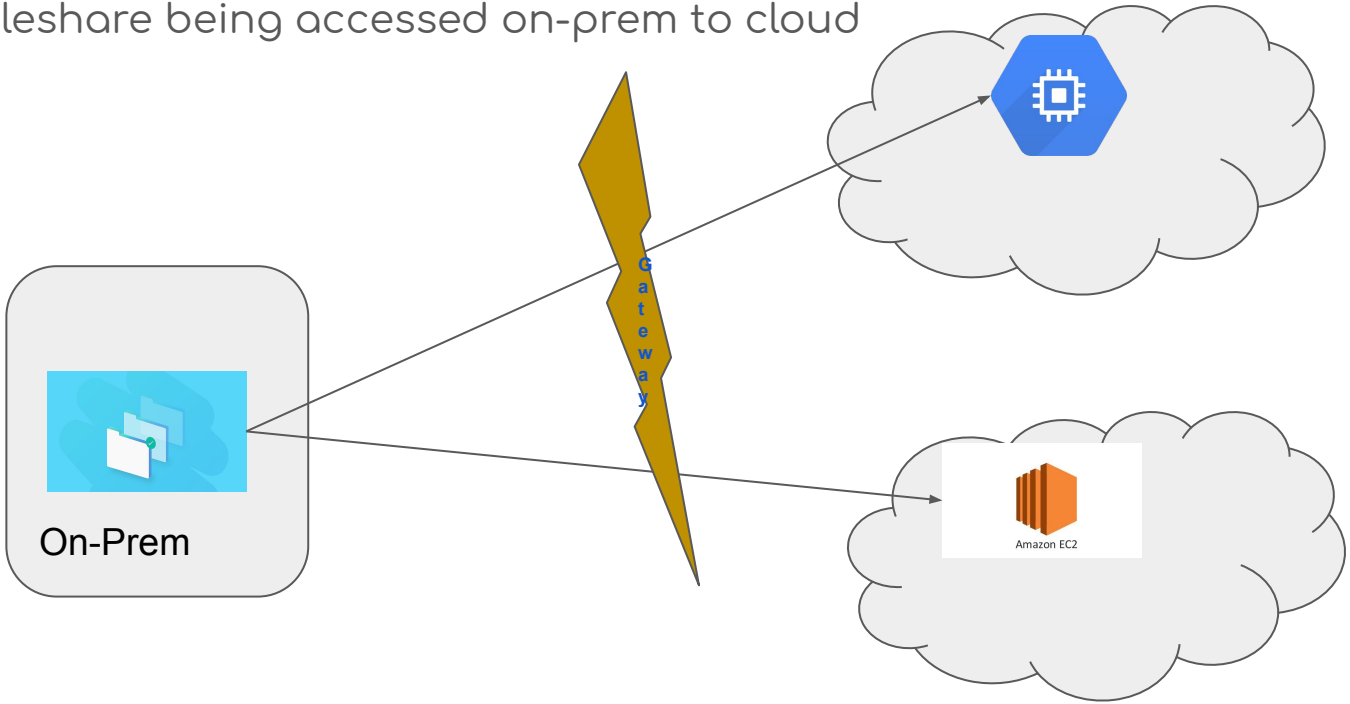


And HIS Vision had the fact behind….

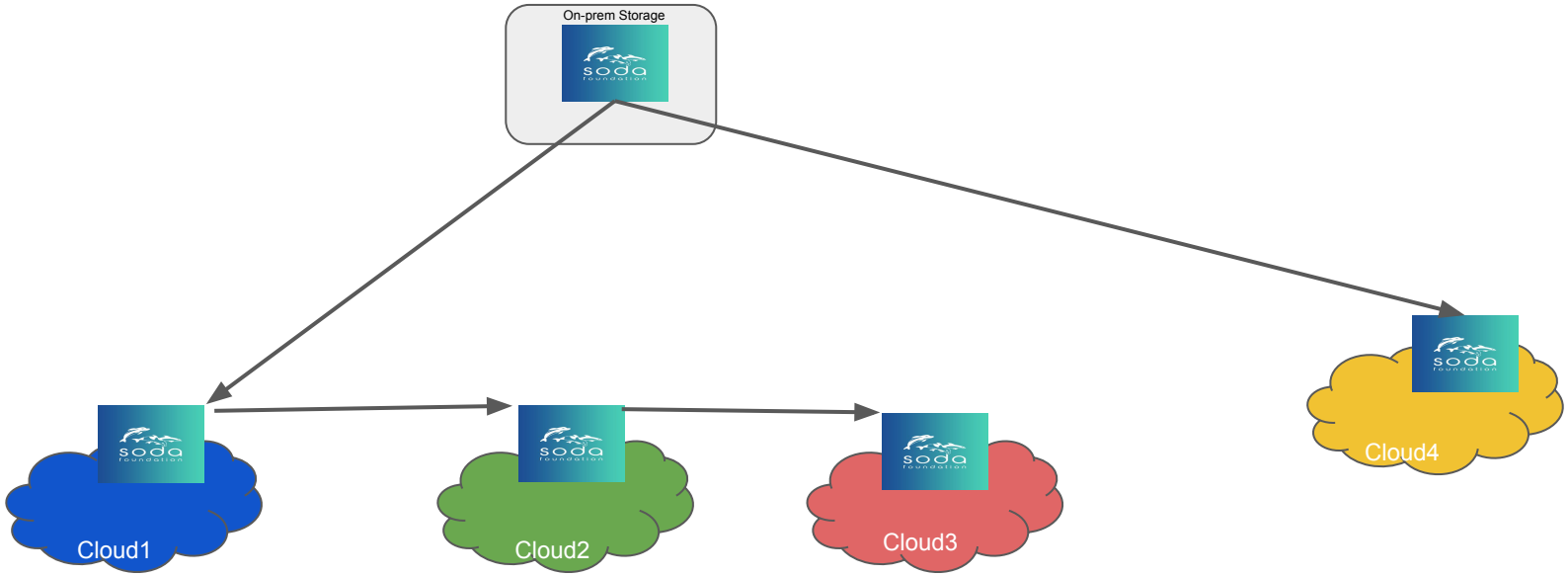
Let’s do it……

Fileshare: Find me anywhere...

NFS Fileshare being accessed on-prem to cloud

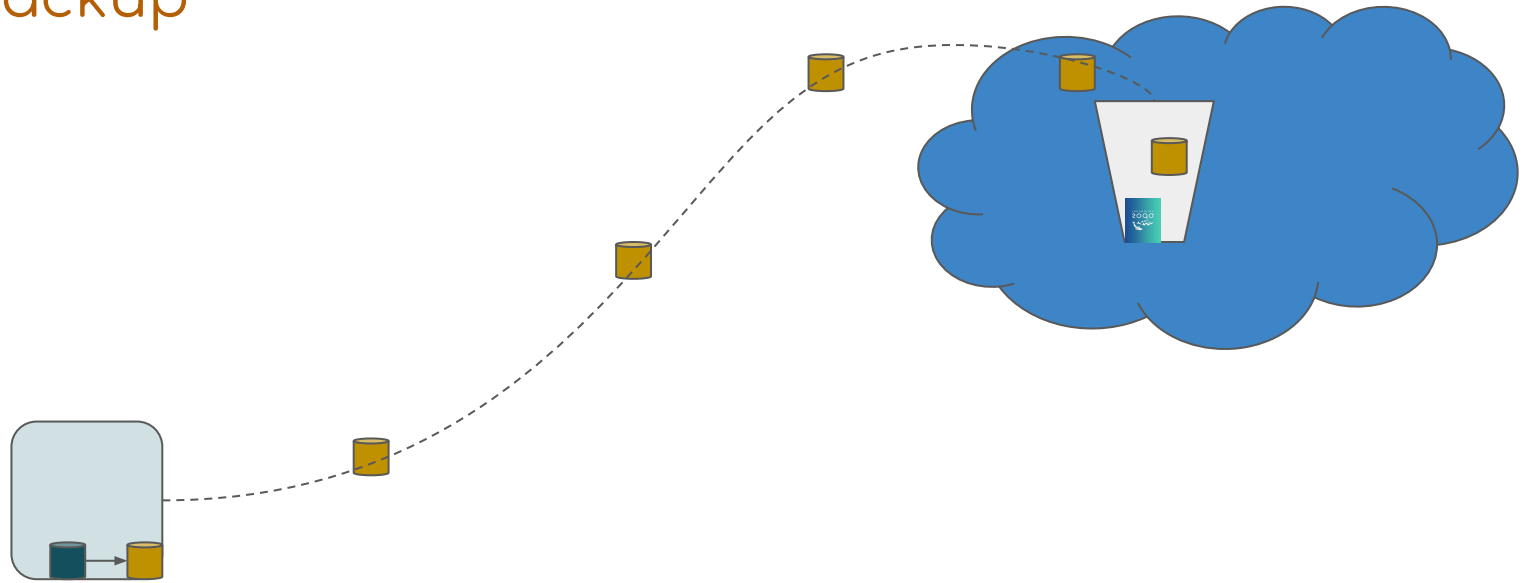


Objects Flying and Roaming around!!

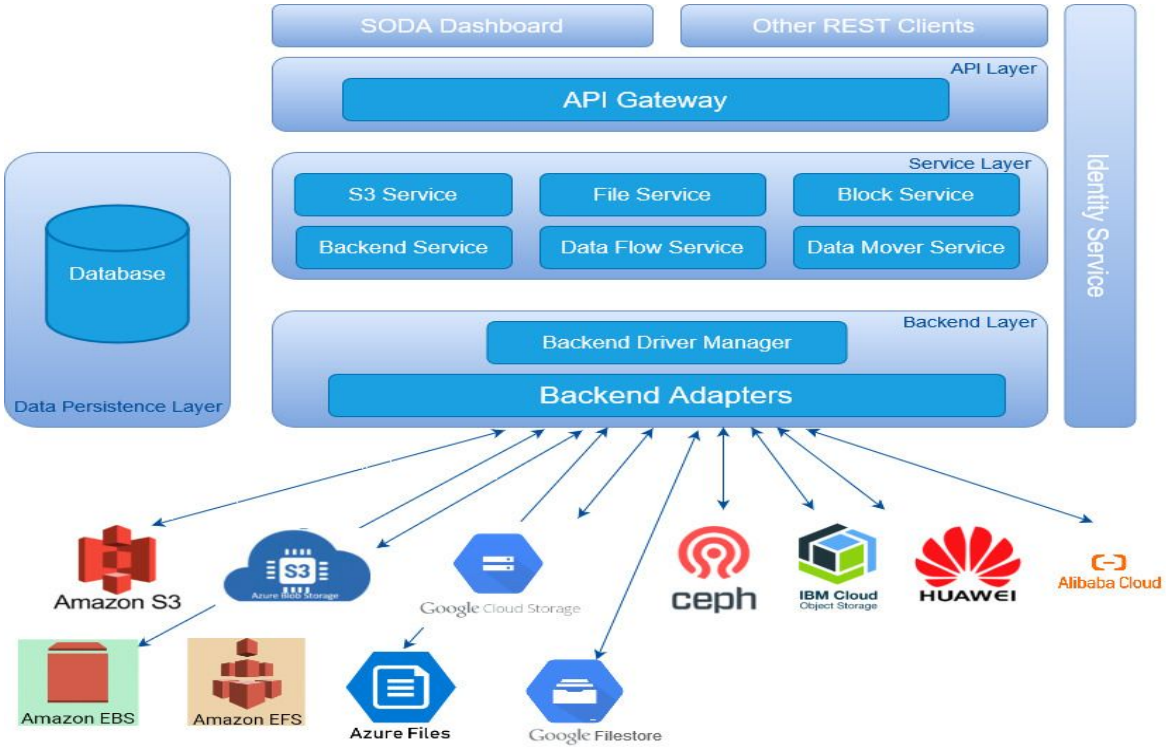


Block data: Save it for difficult time...

Backup



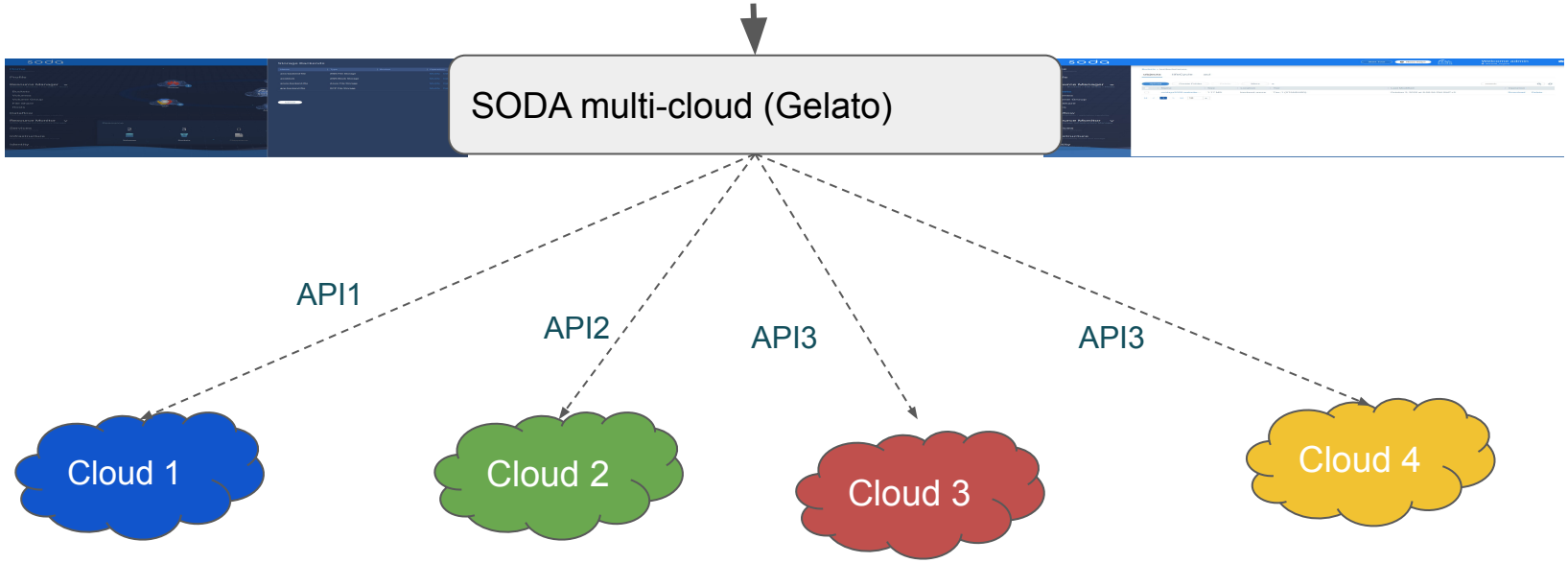
SODA multi-cloud architecture



And it's good be Greedy….

- S3 compatible APIs for Object Storage
- Data Migration
- Policy based Lifecycle migration
- Security
- On-premise/Hybrid/Public
- Multiple Vendors: GCP/AWS/Azure/IBM/HOS/Alibaba/Ceph/YIG
- Fileshare support for AWS/Azure/GCP
- Block Support for AWS

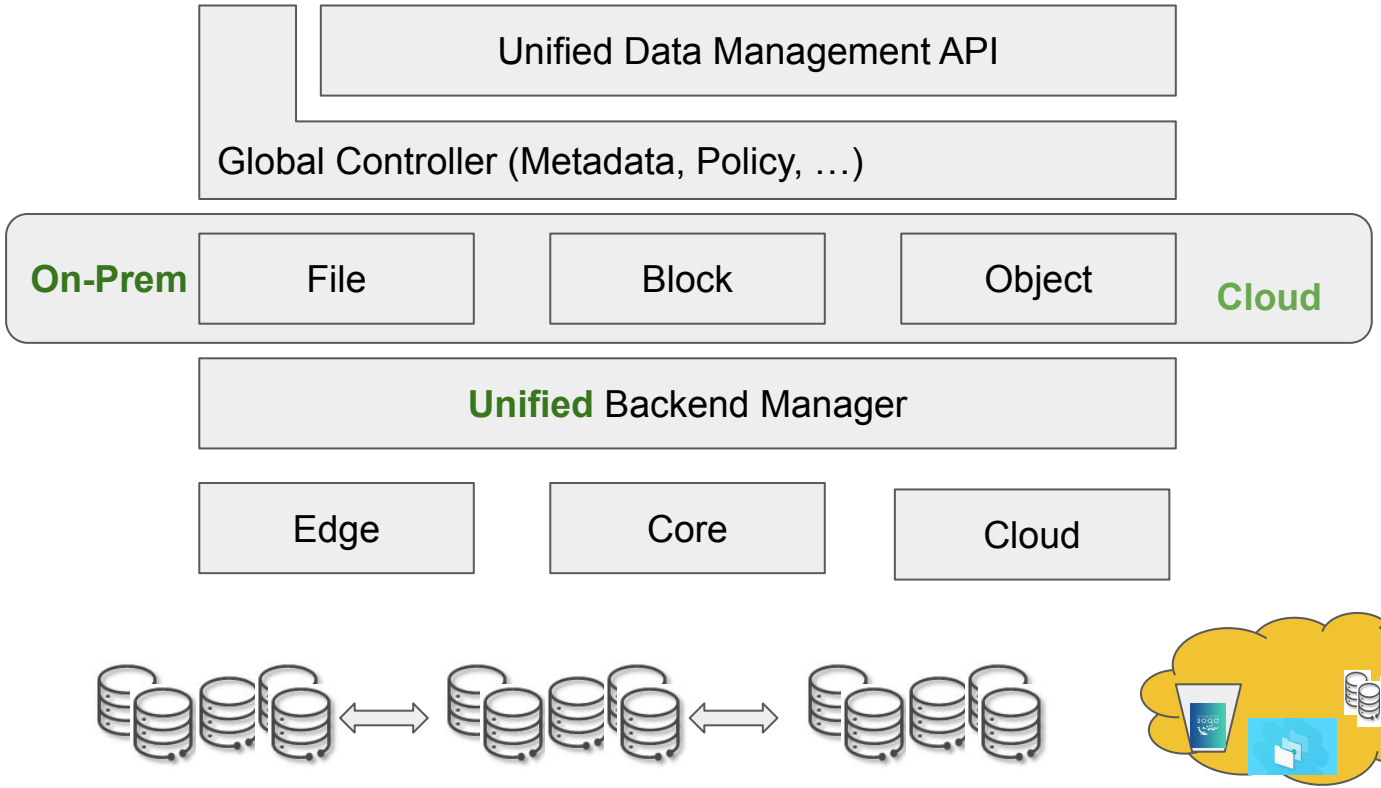
Unified View



Dream!!: A world where data doesn't have boundaries……

More Cloud Providers

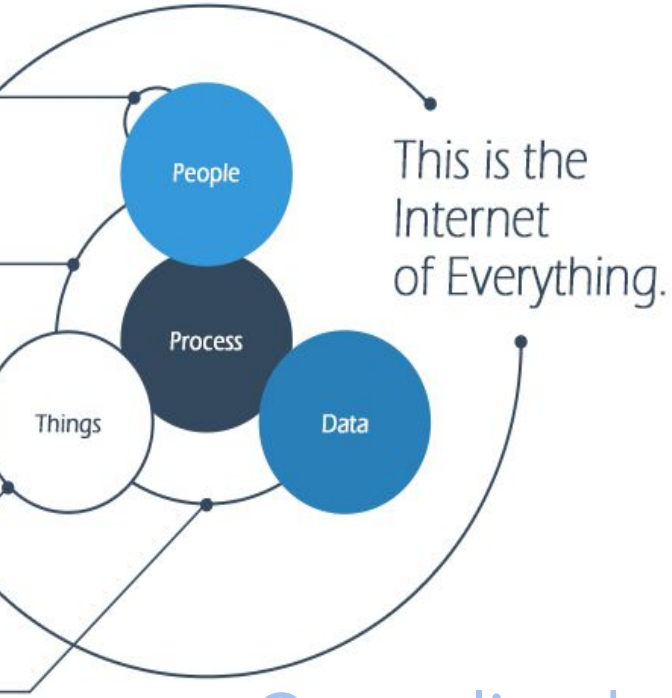
TRUE Hybrid-Cloud



...for an intelligent one cloud...

4

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



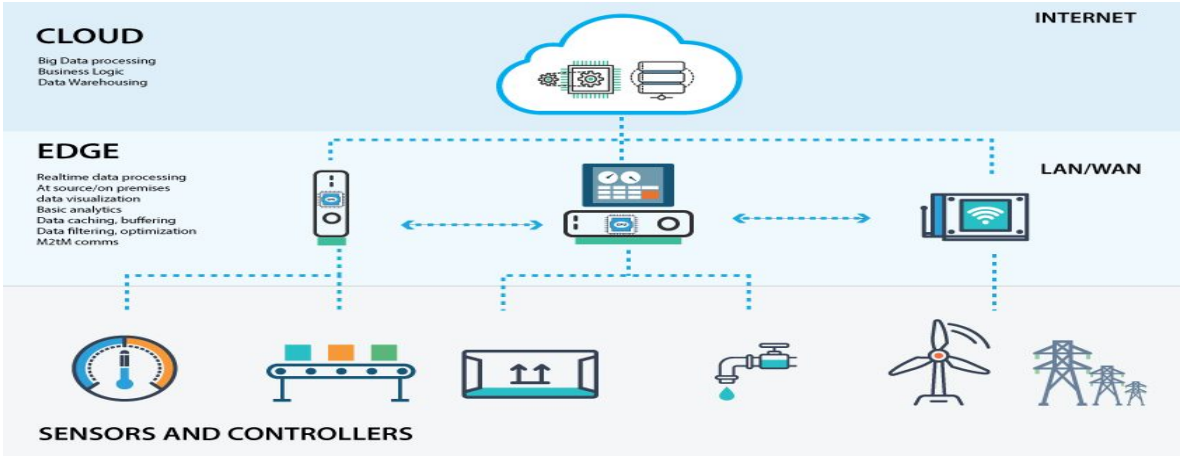
Moving to Edge!

Centralized to Decentralized to Distributed

Edge

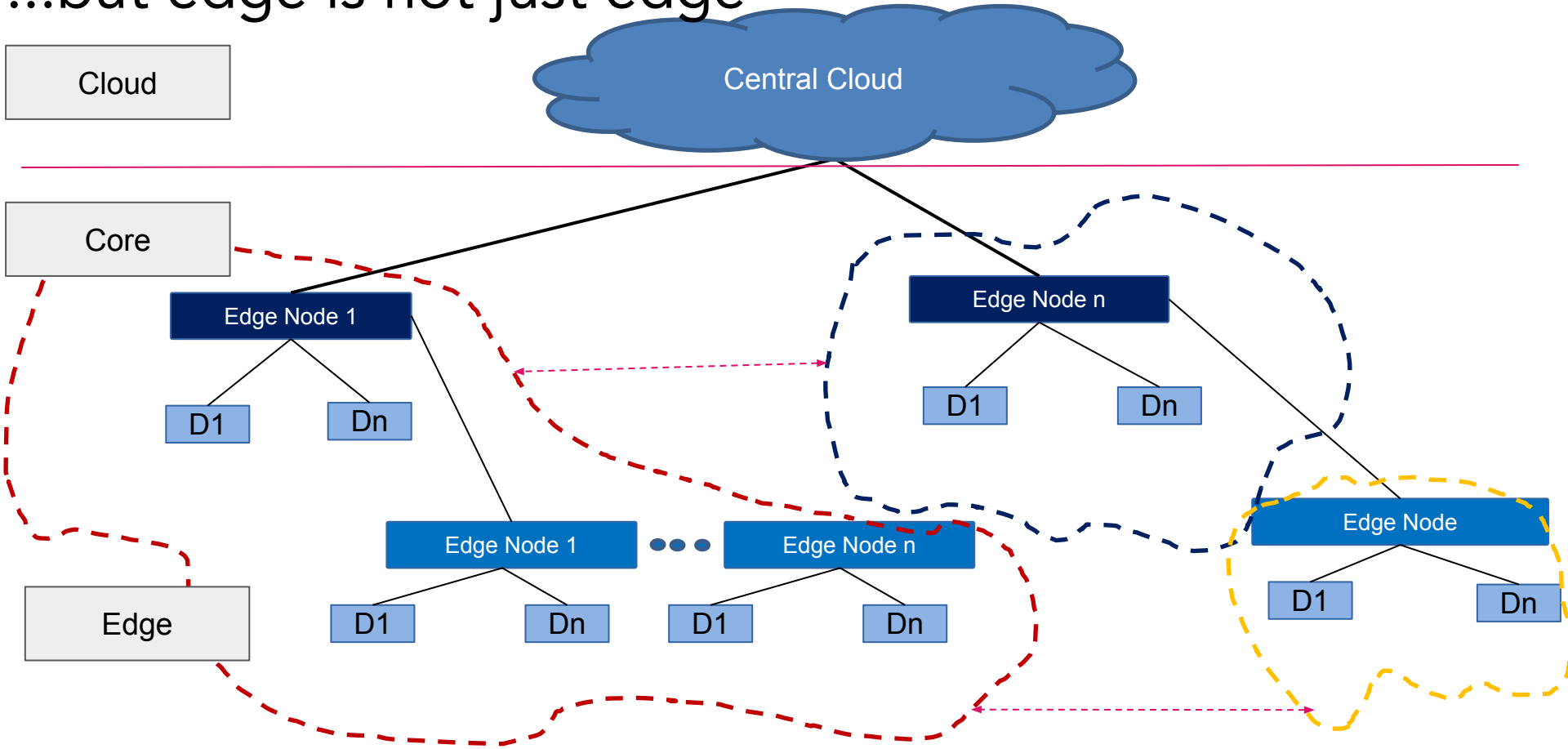
..Alexa, show me the features of SODA Framework…!

..want the response now!

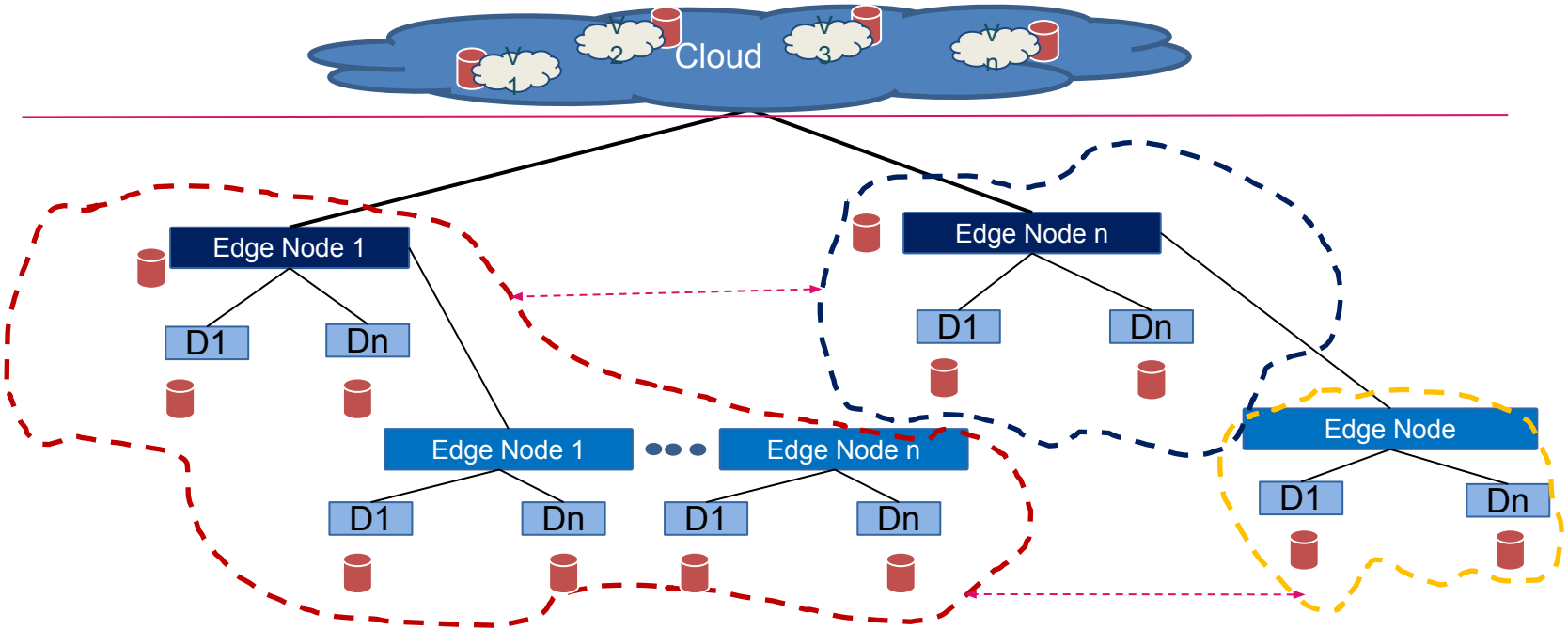


..so, let us process the data closer to the source of the data (edge)

...but edge is not just edge



...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

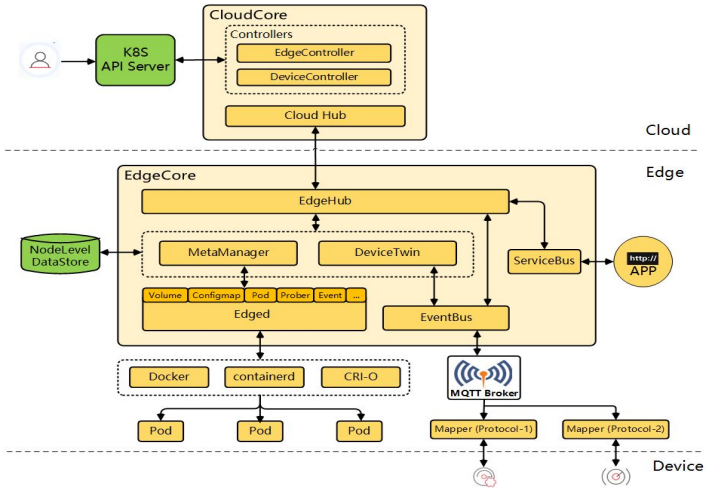
Unified Data Framework
Open
Vendor Agnostic
Platform Agnostic
Distributed
Low Resource
Extensible or Shrinkable
Standardized

Edge Computing Platform

Edge Computing Platform

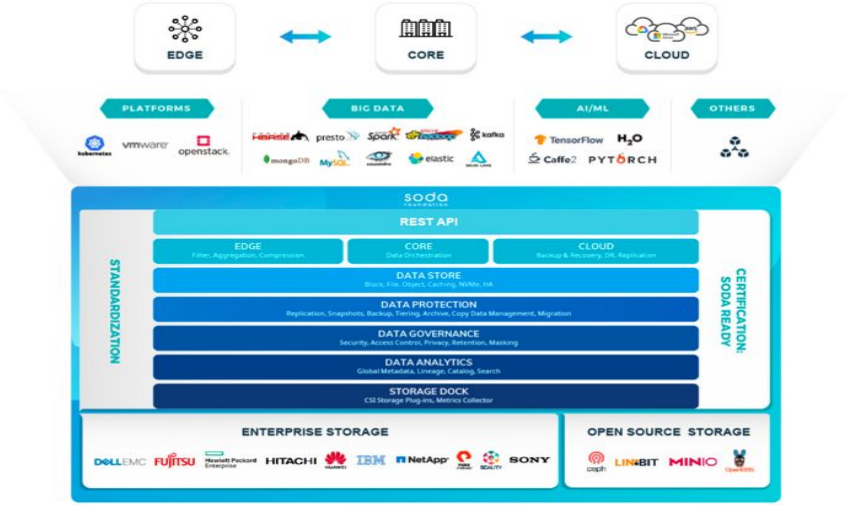
..let us start somewhere...

KubeEdge



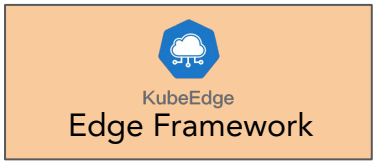
https://github.com/kubeedge

soda foundation



https://github.com/sodafoundation

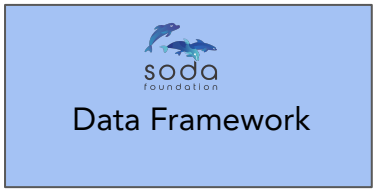
...found compatible needed interface...CSI !



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

Container Storage Interface

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



SODA Unified Heterogeneous Data Interface

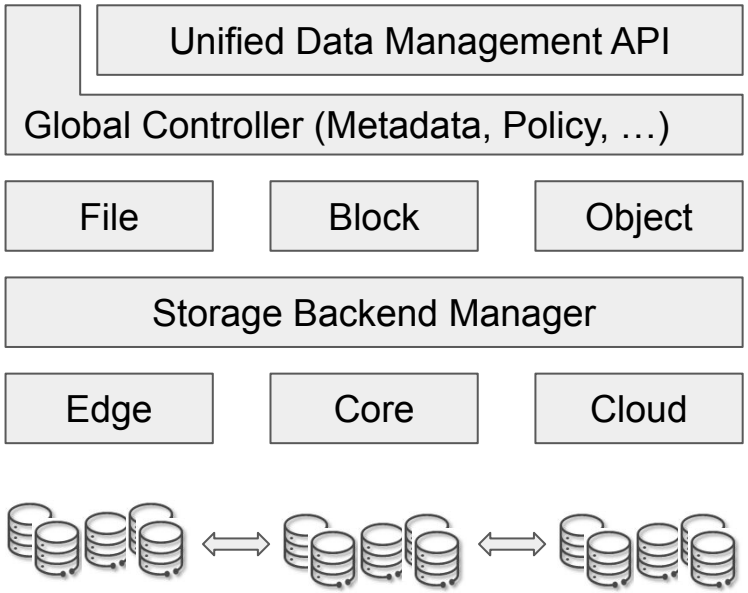
Heterogeneous Storages : different vendors, models, types, cloud storages...

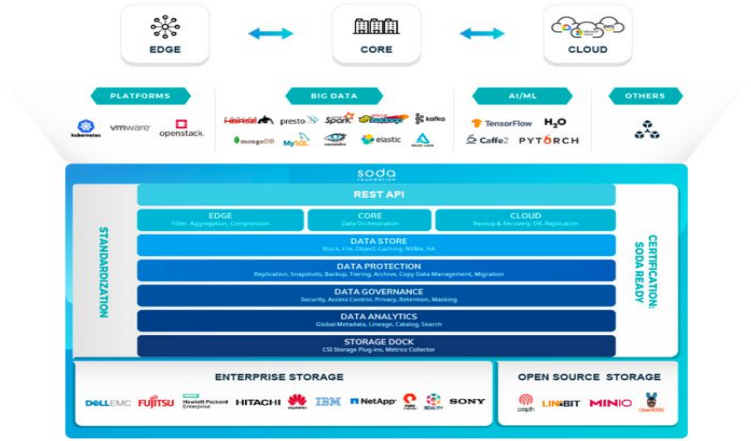
...way to go...

...SODA@Edge SIG

The SODA Story Continues...

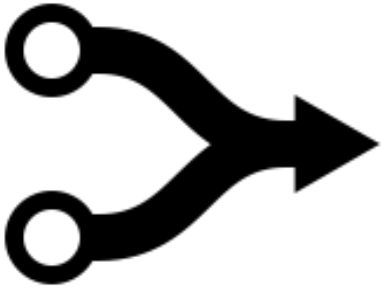
for One Data Framework





to realize infinite possibilities...

Wanna Join?



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

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

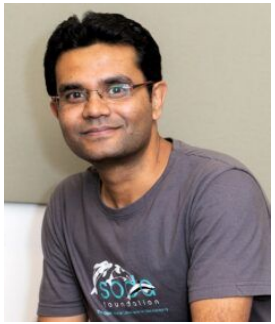
Sushanth Kumar

(SODA Project Maintainer)

Najmudheen CT

(SODA Project Maintainer)

Reach Us at SODA Slack
sodafoundation.io/slack



Ashit Kumar

(SODA Project Maintainer, India OC)

Sanil Kumar D

(SODA TOC member, AWG Lead)

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/

Background

- Why multi-cloud/Gelato project was required
  - Need of being Vendor agnostic
  - Take advantage of the best of the breed

Image: Forrester/Virtustream

May 7, 2019 | Contributor: Laurence Goasduff

Most organizations choose to work with multiple cloud providers, for a host of different reasons.

For an enterprise using cloud services across multiple geographies, finding just one public cloud infrastructure provider to meet its needs is a struggle. In organizations like this, the decision to use a multicloud strategy is clear.

"The 10 biggest public cloud providers will command, at a minimum, half of the total public cloud market until at least 2023"

In fact, most enterprise adopters of public cloud services use multiple providers. This is known as multicloud computing, a subset of the broader term hybrid cloud computing. In a recent Gartner survey of public cloud users, 81% of respondents said they are working with two or more providers.

The dominance of megavendors in the public cloud services market is behind the main reason that enterprise buyers choose multiple cloud providers, says Michael Warrilow, VP Analyst, Gartner.

Data into the Cloud

- Due to data deluge and the requirement of Fast data analytics, it's availability, accessibility and TCO, Cloud data storage is inevitable
- Data @Edge being generated by Edge applications say automated car

https://aws.amazon.com/what-is-cloud-storage/

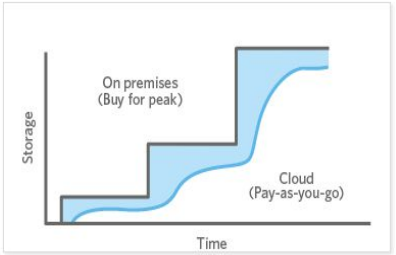
Benefits of Cloud Storage

Storing data in the cloud lets IT departments transform three areas:

1. **Total Cost of Ownership.** With cloud storage, there is no hardware to purchase, storage to provision, or capital being used for "someday" scenarios. You can add or remove capacity on demand, quickly change performance and retention characteristics, and only pay for storage that you actually use. Less frequently accessed data can even be automatically moved to lower cost tiers in accordance with auditable rules, driving economies of scale.

2. **Time to Deployment.** When development teams are ready to execute, infrastructure should never slow them down. Cloud storage allows IT to quickly deliver the exact amount of storage needed, right when it's needed. This allows IT to focus on solving complex application problems instead of having to manage storage systems.

3. **Information Management.** Centralizing storage in the cloud creates a tremendous leverage point for new use cases. By using cloud storage lifecycle management policies, you can perform powerful information management tasks including automated tiering or locking down data in support of compliance requirements.



Cherry on the Cake “Orchestration”

- To Orchestrate and Automate complex tasks in Cloud
- Pre-defined Workflow definitions

And HIS Vision had the fact behind….

Type of Cloud Storage

- **Object** (object storage vast scalability and metadata characteristics.)
- **File** (like large content repositories, development environments, media stores, or user home directories.)
- **Block** (Low latency for HPC)

Use cases

- Use case of Object storage
- Use case of File storage
  - Cloud Bursting/Analytics/On-prem <->Cloud
- Use case of Block storage
  - Replication
  - Cloud Native