Storage Developer Conference
September 22-23, 2020

Autonomous Data Management at Edge

Challenges and Possibilities

Sanil Kumar D SODA Foundation/Huawei (skdsanil@gmail.com)
Vinod Eswar SODA Foundation/Wipro (vinod.eswar@wipro.com)

Edge Computing - A Recap

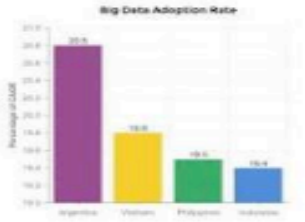
Google

data growth rate 2020

All | News | Images | Videos | Shopping | More | Settings | Tools

About 1,39,00,00,000 results (0.44 seconds)

The amount **of data** created each year is **growing** faster than ever before. By **2020**, every human on the planet will be creating 1.7 megabytes **of** information… each second! **In** only a year, the accumulated world **data** will **grow** to 44 zettabytes (that's 44 trillion gigabytes)!



Zettabytes = one sextillion (10^{21})
or,
2^{70} bytes.

How much data is generated daily 2020?

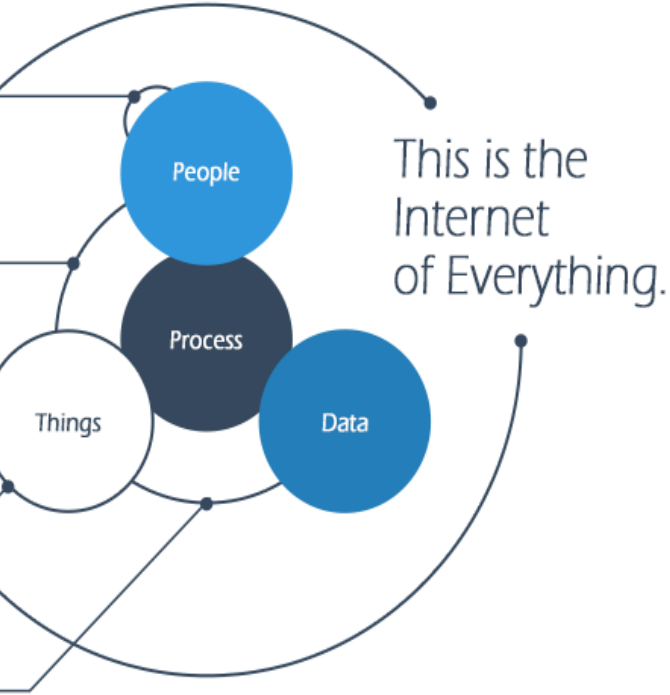
2.5 quintillion bytes of **data** are **produced** by humans **every day**. If you've wondered **how much data** the average person uses per month, you can start by looking at **how much data** is created **every day** in **2020 by the** average person. This currently stands at 2.5 quintillion bytes per person, per **day**. Aug 20, 2020

Who generates this much of data?

All these are valuable?

Where is it getting generated?

Where should we process it?

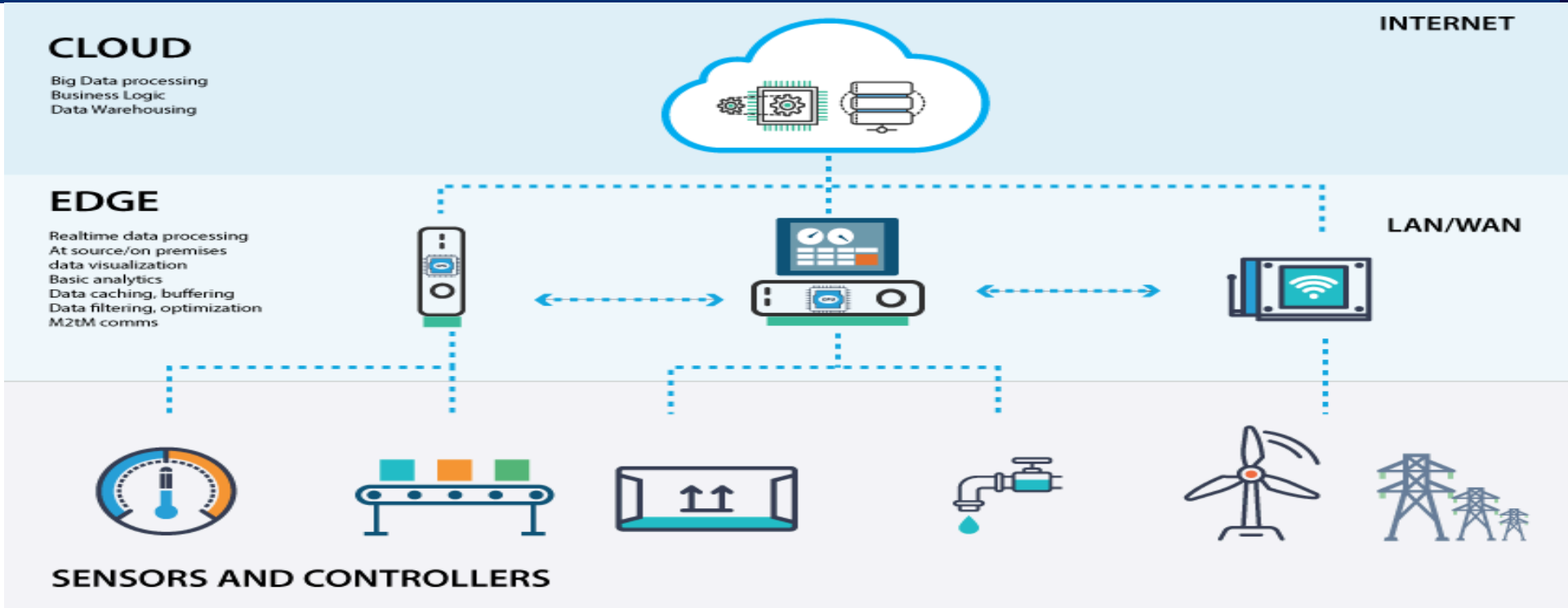


Moving to Edge!

Centralized to Decentralized to Distributed

Edge?

Edge computing is a method of **optimizing cloud** computing systems by **performing data processing at the edge of the network, near the source of the data.**



...and...why Edge?

Increasing costs of shipping the large volumes of data to the cloud for processing and storage.

Reduce the Cost

Trust & Security

Data governance and security – many organizations have sensitive data that they don't want to leave their premises under any circumstances.

Real-time decision making –the latencies involved in shipping the data to the cloud for analytics are unacceptable.

Real time, Ultra Low Latency

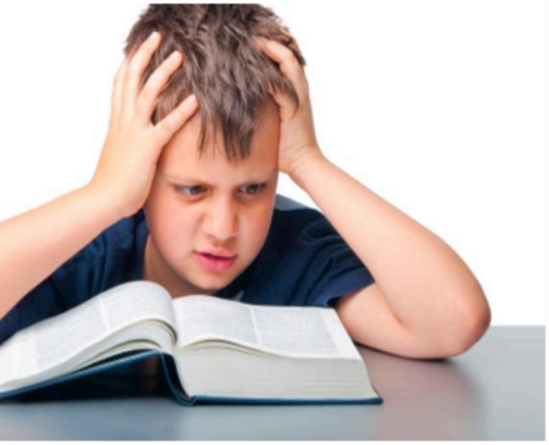
Offline, Independent

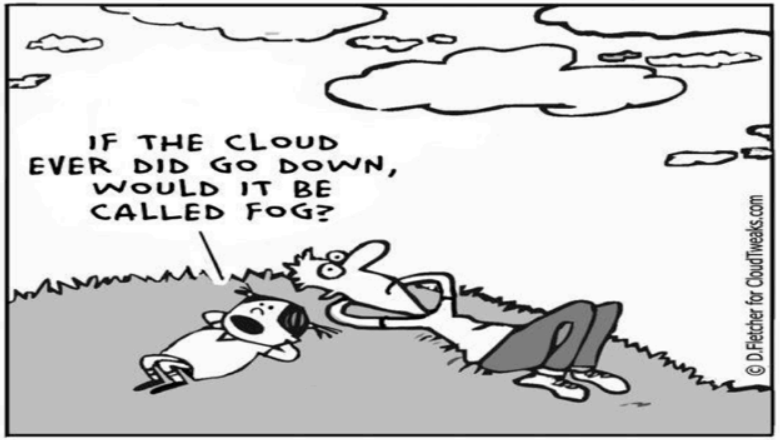
The possibility of intermittent cloud connectivity is a serious concern for mission-critical IoT applications such as a connected vehicle or other types of autonomous systems.

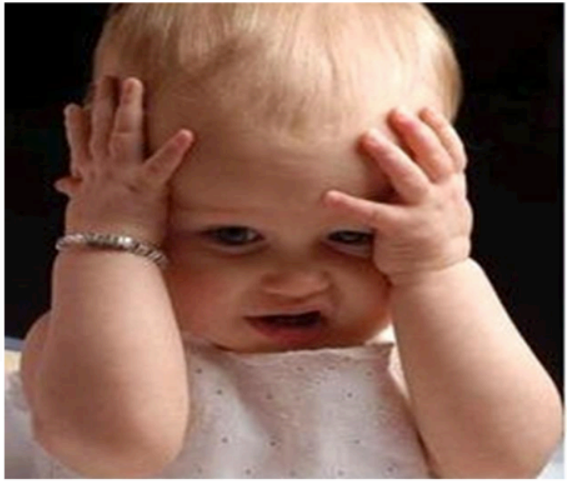
Roof Computing?

MEC ?

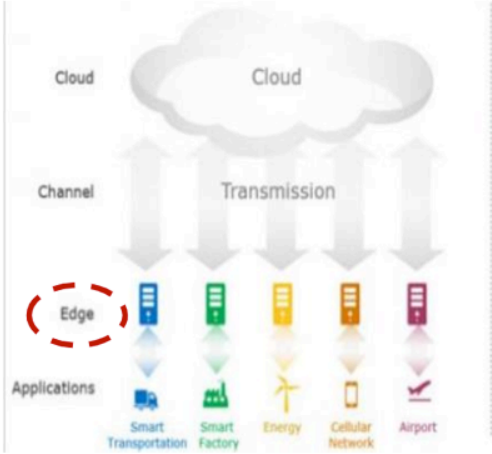
Fog Computing?

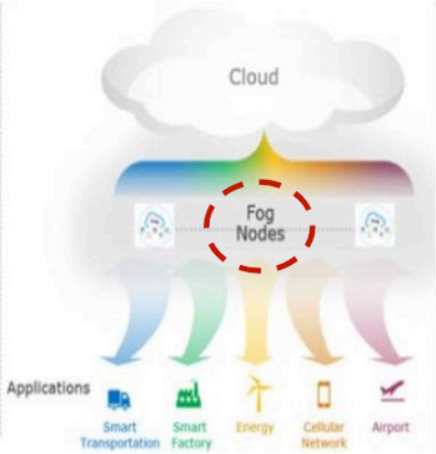


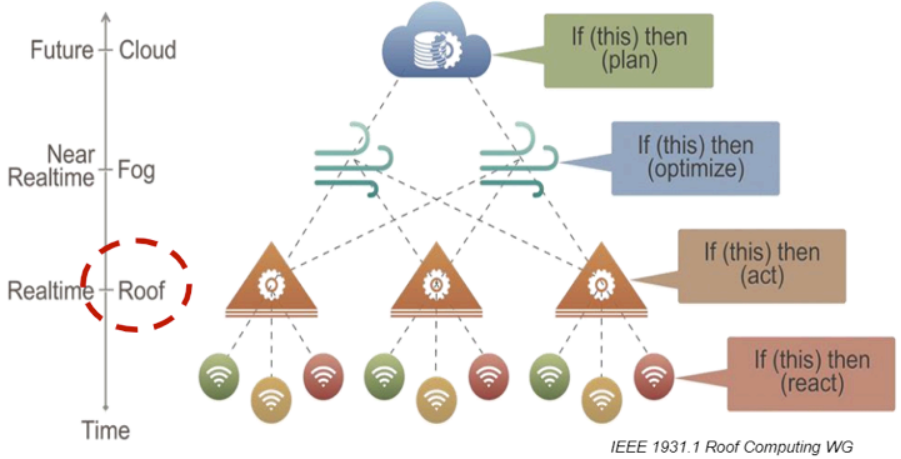




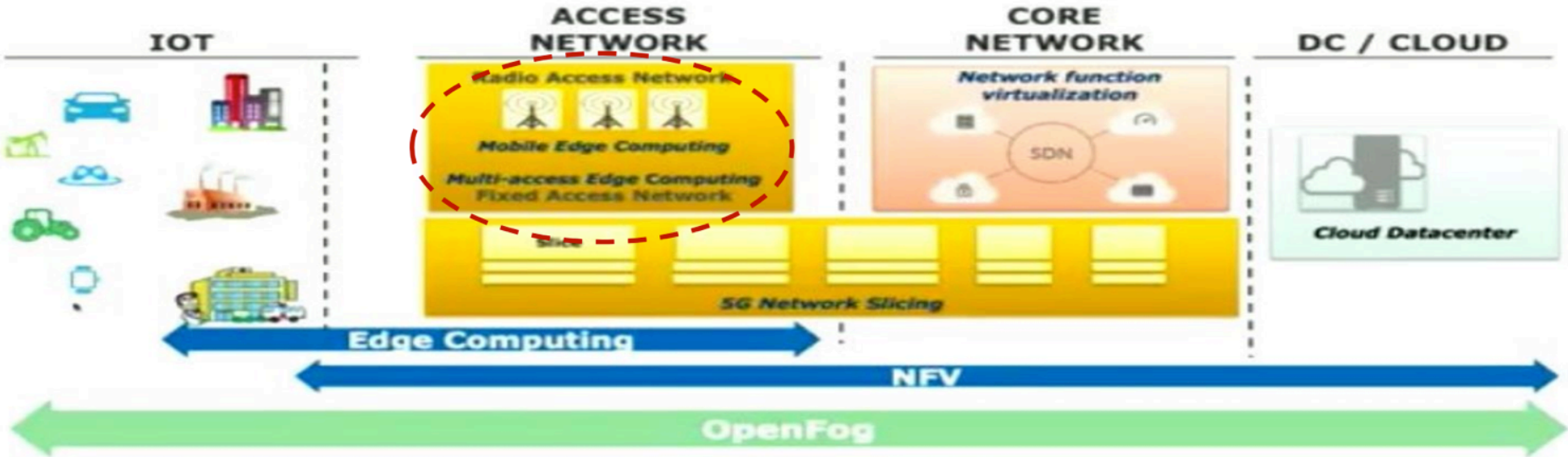
Edge Computing?







IEEE 1931.1 Roof Computing WG



https://www.openfogconsortium.org

Key Requirements

- **Maximize the computing at Edge**
  - Real time Scheduling
  - Distributed Application Execution
  - Resource Optimization and Compute Efficiency
  - Efficient Orchestration, Monitoring
  - Resource utilization across the cluster
  - Efficient Runtime Support (Container, LWC, Serverless)
  - Low Latency

- **Offline Scenarios and Communication**
  - Edge Node/Cluster Offline Working
  - Vendor Agnostic Cloud Interface
  - East – West Communication
  - Reverse Proxy, Address Resolution, Routing
  - Workload-Workload, Device to Workload Commn

- **Security & Privacy**
  - Workload to Workload Secure Communication
  - Device Identity and Authorization
  - Node level identity
  - Private Data Isolation

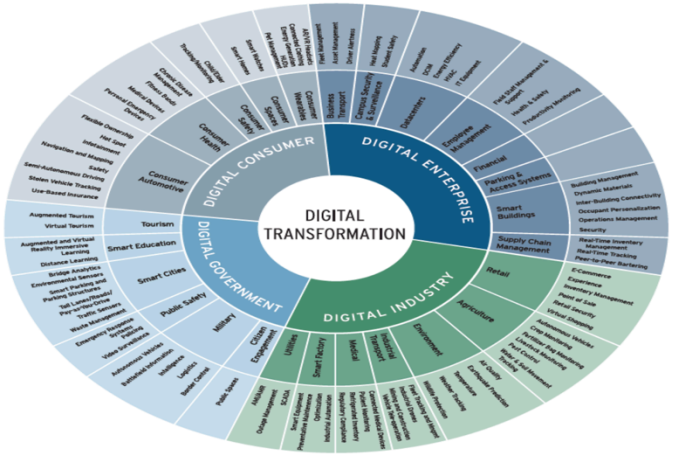
- **Scalability – Platform and Clusters**
  - Edge Cloud - Clusters
  - Microservice based core platform
  - Platform extensions and plugins

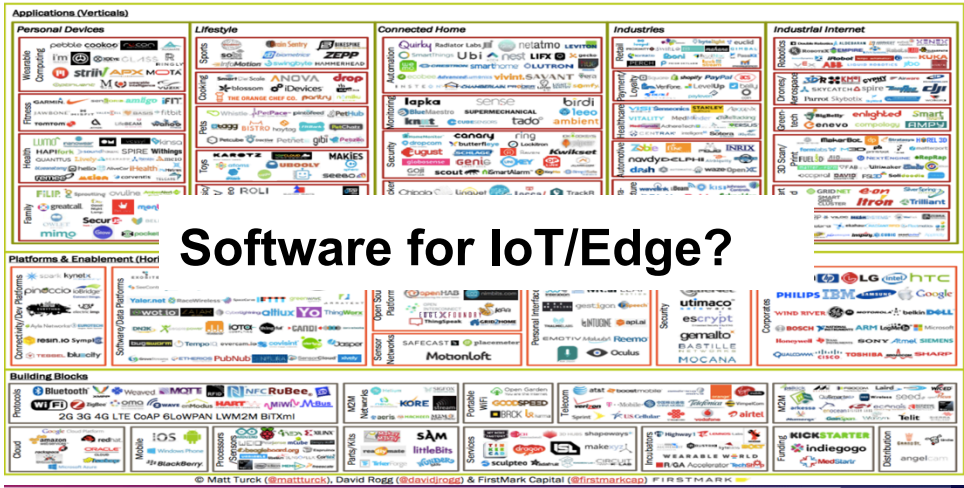
- **Device Life Cycle & Management**
  - Device, Node, Application Provisioning
  - Repositories and Registry (Device, Mapper, Node, Workload)
  - Discovery (Device, Node, Application/ Service)

- **Data & Data Analytics**
  - Data Storage, Sharing, Distributed
  - Distributed and customizable Data Analytics
  - AI/ML, Big Data, Streaming Data processing…

- **Remote Management & Visualization**
  - Consolidated and Efficient Dashboard (nodes, devices, workloads, resources so on)
  - Dashboard at North and South
  - Upgrade, Rollback, Reset, Enable/Disable

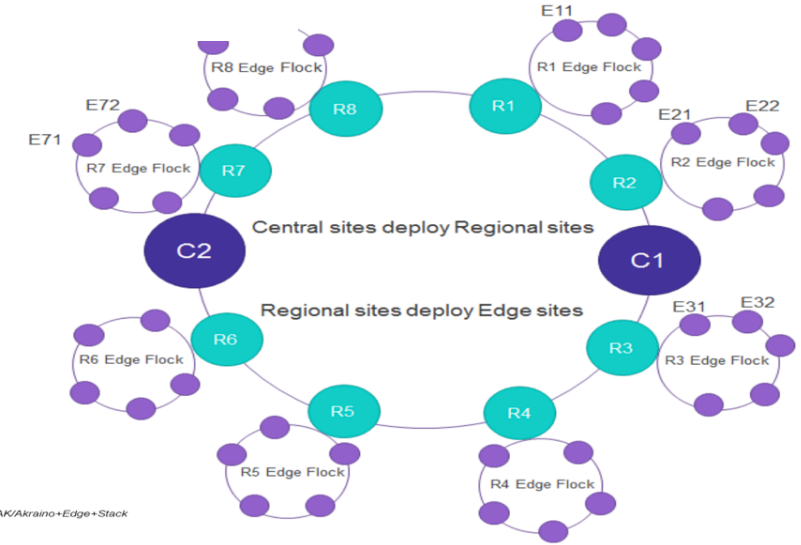
- **Efficient Energy Management**
  - Energy aware workload scheduler
  - Energy optimized Devices/Nodes
  - Energy Monitoring





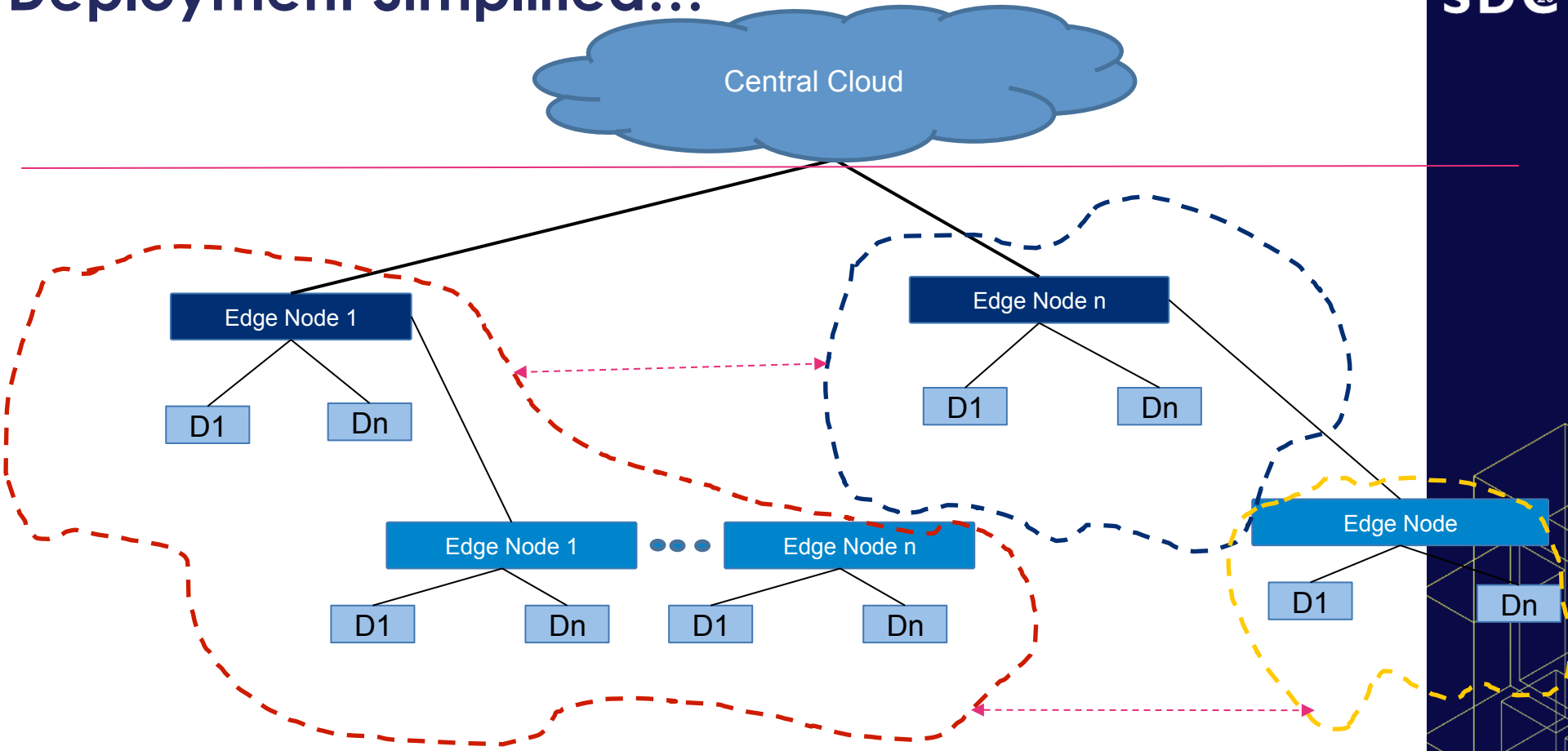
Software for IoT/Edge?

Edge Deployment Possibilities..?



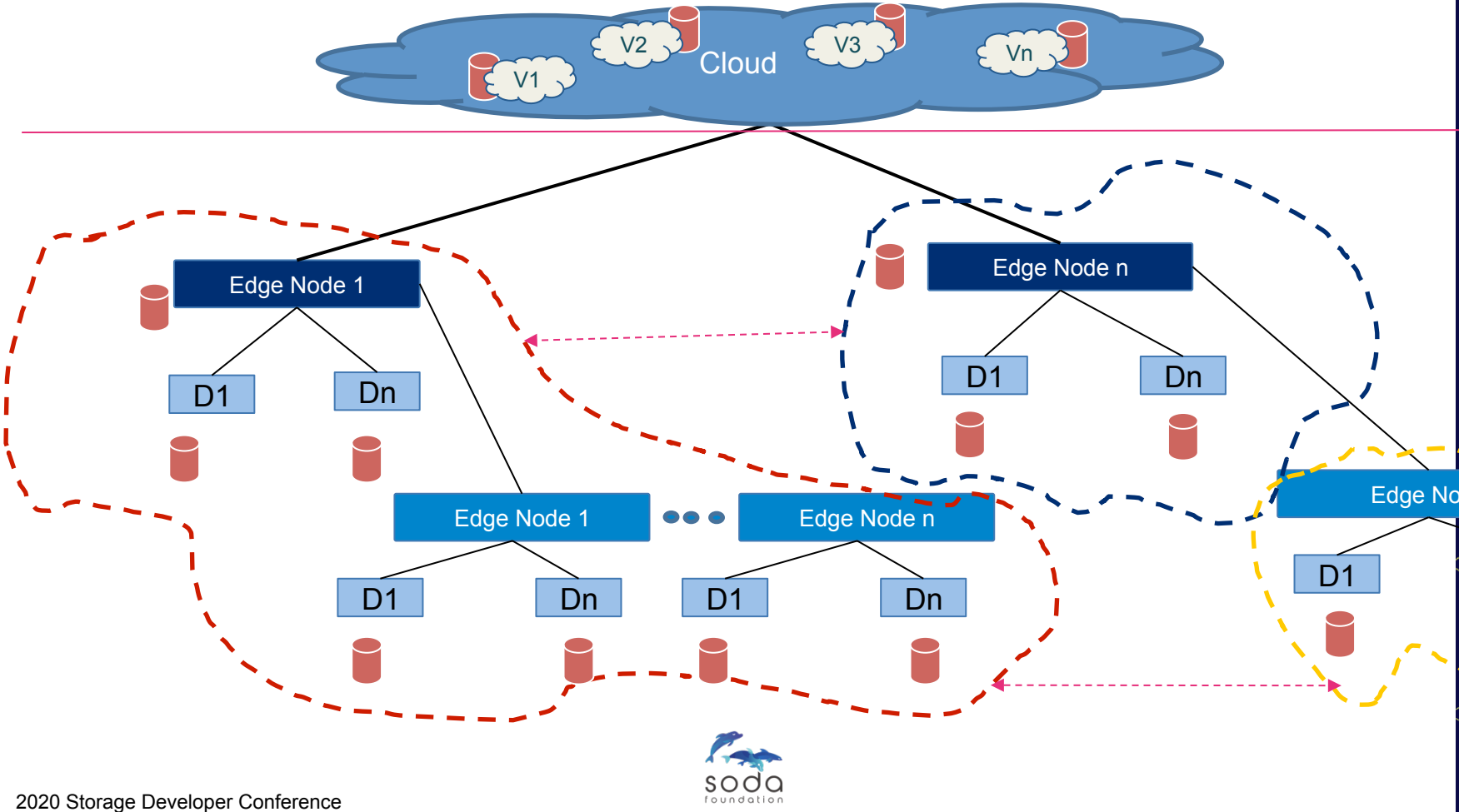
https://wiki.akraino.org/display/AK/Akraino+Edge+Stack

Deployment Simplified…



Data @ Edge
Deployment and Challenges

Distributed Data @ Edge



Data @ Edge

- Need for processing closer to the source of the data for:
  - Real time response
  - Data Efficiency
  - Data Energy
  - Data Privacy
  - Data Security
- High Demand of Use cases

But....

- Edge Platforms are not ready
- Storage research underway
- Current - mostly core / cloud data storage

Multi-Source Data Generation

Data from Each Device at the Edge
Data from Each Node at the Edge
Data from each cloud vendor

Data Operated from multiple sources from Edge to Core to Cloud!

Heterogeneous Data Storage

Store at Edge
Store at Core
Store at Cloud
Different Kinds of Storages (Object, File, Block)
Different Types (SSD, Flash..)
Different Vendors

Data Stored in heterogeneous storages across Edge,Core and Cloud!

Distributed...Really!

Real P2P Network
Huge Number of Devices/Nodes

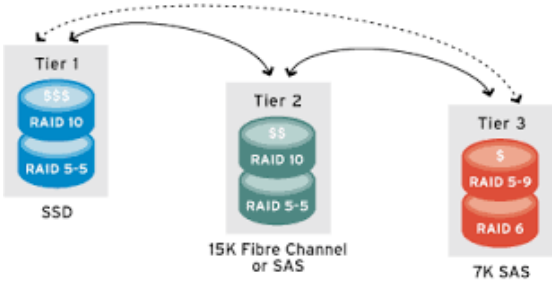
Data Ownership and Consistency

Multi-point Data Access and Ownership
Making Data Consistency
Real time consistency?

Low Latency

Less Bandwidth, yet low latency demand for real time response

- Made for Real-Time Response
- Low-latency edge computing
- low-latency distributed data stores



- The heterogeneity of the storage nodes at the Edge is far more diverse than the Cloud
- Diverse Data Store Design
- What should be the caching strategy?

A uniform storage tiering architecture – software defined across heterogeneous storage architectures

Offline Scenarios

A Key feature of Edge Architecture is Offline Data storage and processing capacity

- Network failure
- Congestion
- Planned upgrade

Ability to aggregate data from Devices/Sensors in offline environment

- How is the data synced with the cloud backend?
- A data synchronization framework that works across storage systems
- Storage that supports Offline-first application architecture

Cloud Offline
DC or Cluster Offline

Data Mobility

Device, Node... Entry Exit
Node to Node, Cluster to Cluster Data Mobility

- There is an implicit storage hierarchy
- Workload/Use case demands may vary

- Data mobility is the movement of Data from – Device, Node, Core to Cloud.
- The protocol, lifecycle management, and data protection are key to data mobility

Modern Data requirements drive mobility:
- supported for application development and test
- backup, disaster recovery
- analytics,
- Archive
- And more….

Storage services should focused on managing the data, not the infra.

So
Many
Challenges…!

Many Possibilities; Research ; Solutions

What we need?

Distributed & Heterogeneous Data Management Platform@Edge

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

SODA Framework Introduction or Recap!

SODA FOUNDATION

SODA Foundation is an open source project that aims to foster an ecosystem of open source data and storage software for data autonomy. SODA Foundation is organized as a Directed Fund Project under the Linux Foundation with funding from Premier Members and General Members.

SODA Foundation is transformed from the OpenSDS project which focused on SDS management.

SODA : SODA Open Data Autonomy

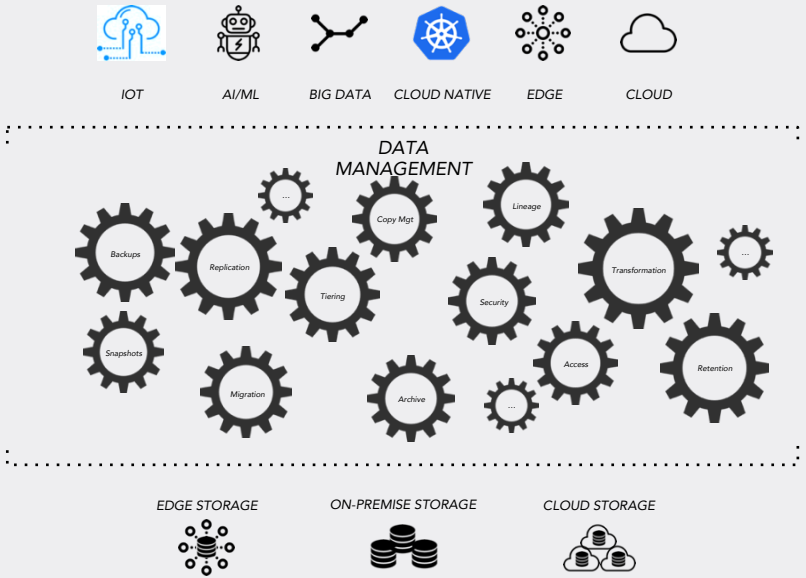
MISSION

To foster an ecosystem of open source data management and storage software for data autonomy

To offer a neutral forum for cross-projects collaboration and integration

To provide end users quality end-to-end solutions

soda foundation



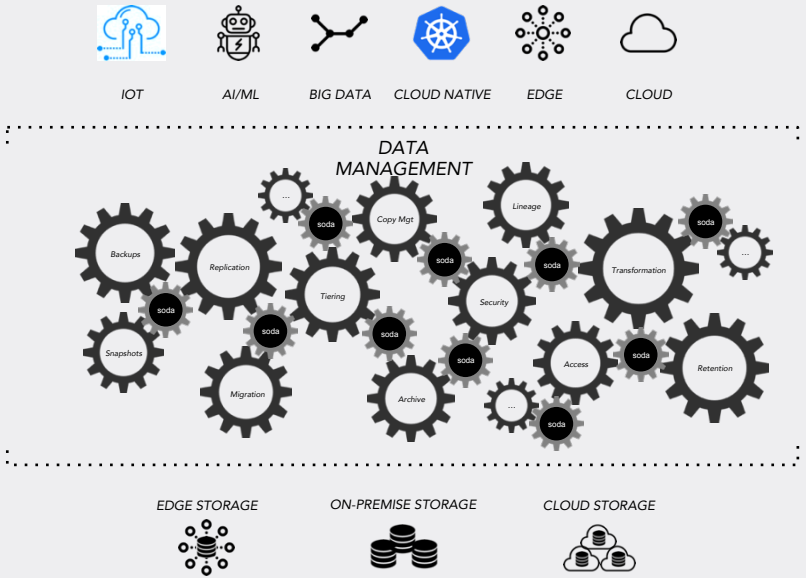
NOW

FRAGMENTATION

There is no single vendor solution that can undertake the humongous task of handling all kinds of data operations from the point of data creation, to data storage, to data archive and data disposal. Challenges include virtualization, cloud native, IoT, big data, AI and machine learning, with the added complexity of data residing everywhere beyond the data center in any cloud or at the edges.

The result is that multiple vendor solutions have to be put together causing these problems:

- *data silos*
- *non-standard interfaces*
- *spot solutions*
- *excessive data transfers*
- *mostly manual operations*
- *limited scalability*



SODA

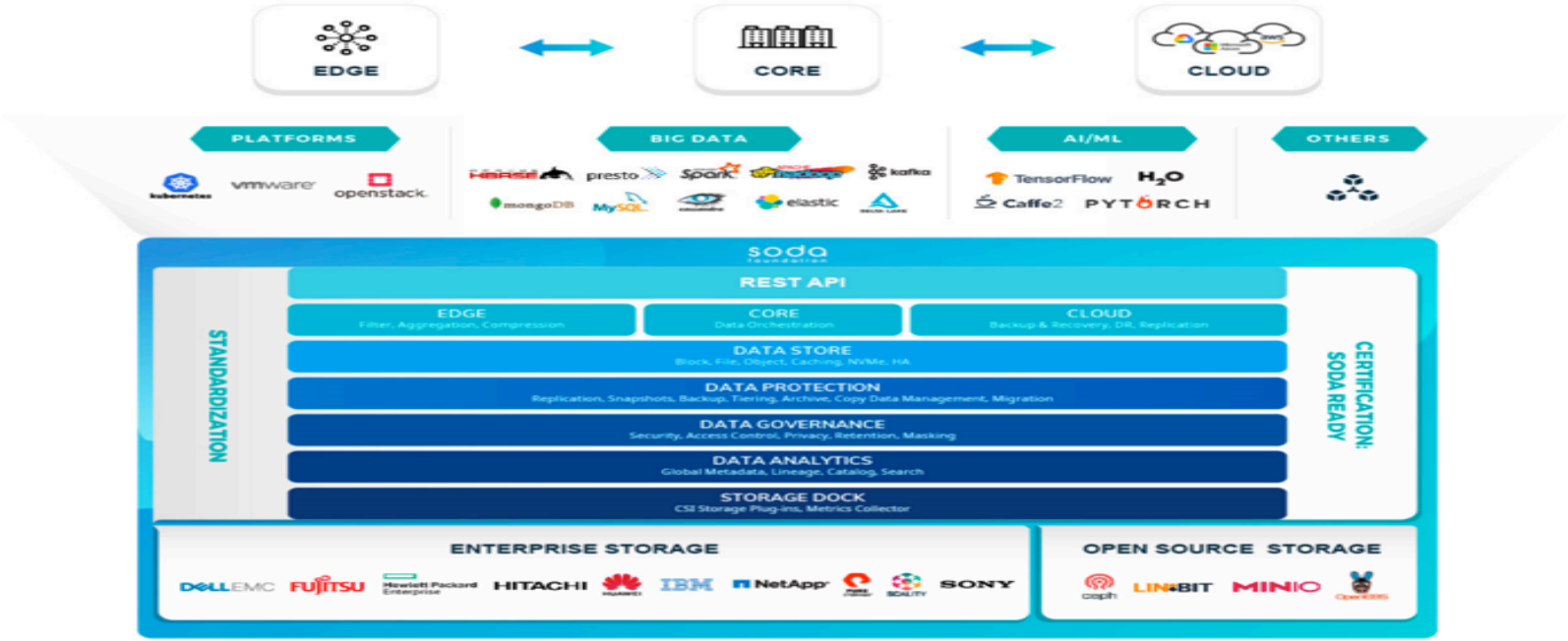
ONE DATA FRAMEWORK

SODA is a single data framework connecting disparate solutions into seamless end to end solutions. This framework is open source, allowing any developer, vendor, or end user to build and extend upon it.

The value propositions of this framework are:

- *data mobility*
- *standardized interfaces*
- *interoperable solutions*
- *optimized transfers*
- *autonomous operations*
- *high scalability*

SODA is an open source unified autonomous data framework for data mobility from edge to core to cloud.



soda foundation

Focus Areas

DATA MOBILITY

SODA enables seamless data tiering, replication, migration, and archiving to the cloud using a common S3 interface with multi-cloud support for AWS, Azure, Google Cloud, IBM Cloud, and other local S3 object storage such as Ceph

DATA PROTECTION

SODA uses snapshots to provide efficient data protection and instant recovery. Using data protection policies, snapshots can be programmed to be taken at intervals, and can be replicated to different storage, or to the cloud for disaster recovery

DATA LIFECYCLE

SODA combines different storage resources on-premise, and across multiple clouds to build tiers. 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

UNIFIED STORAGE PLATFORM

SODA abstracts multi-vendor storage and offers block, file and object storage services, monitoring, and more that work natively with Kubernetes, VMWare, and OpenStack. SODA can be easily extended to support other application frameworks.

CLOUD NATIVE STORAGE

SODA provides persistent storage to Kubernetes stateful workloads with the SODA CSI driver. SODA consolidates multi-vendor storage backends into storage pools for dynamic provisioning. We are working on allowing cloud native storage to be plugged into SODA with their CSI drivers so , reducing complex multi-driver management

DATA GOVERNANCE

SODA is working to offer a common data governance framework from edge to core to cloud to meet the needs for data regulations. This framework will provide, access control, data security, data retention, key management, etc. for data at rest and data in flight. We are looking for data governance and security experts to join us in this effort.

DATA ORCHESTRATION

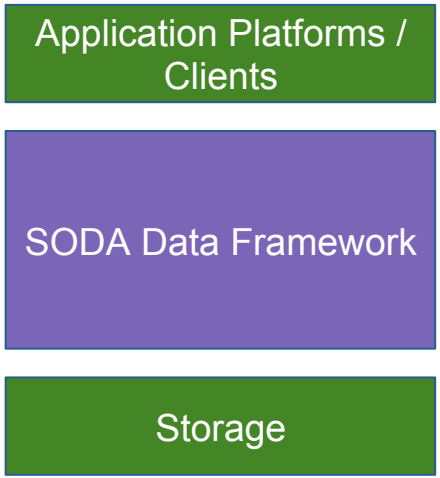
SODA is working to support data orchestration for IoT, big data, machine learning and other popular application frameworks, across different backend storage, and multiple clouds with a distributed data store. We welcome developers who are interested In working on this project.

DATA ENERGY

Data energy is a new area of focus which tries to profile, analyze and optimize the energy consumption for data management. We envision data energy to be a key ROI and a competitive parameter for data solutions. SODA community has identified this for standardization to bring about energy efficient data solutions.

soda foundation

SODA Data Framework : A Simple View

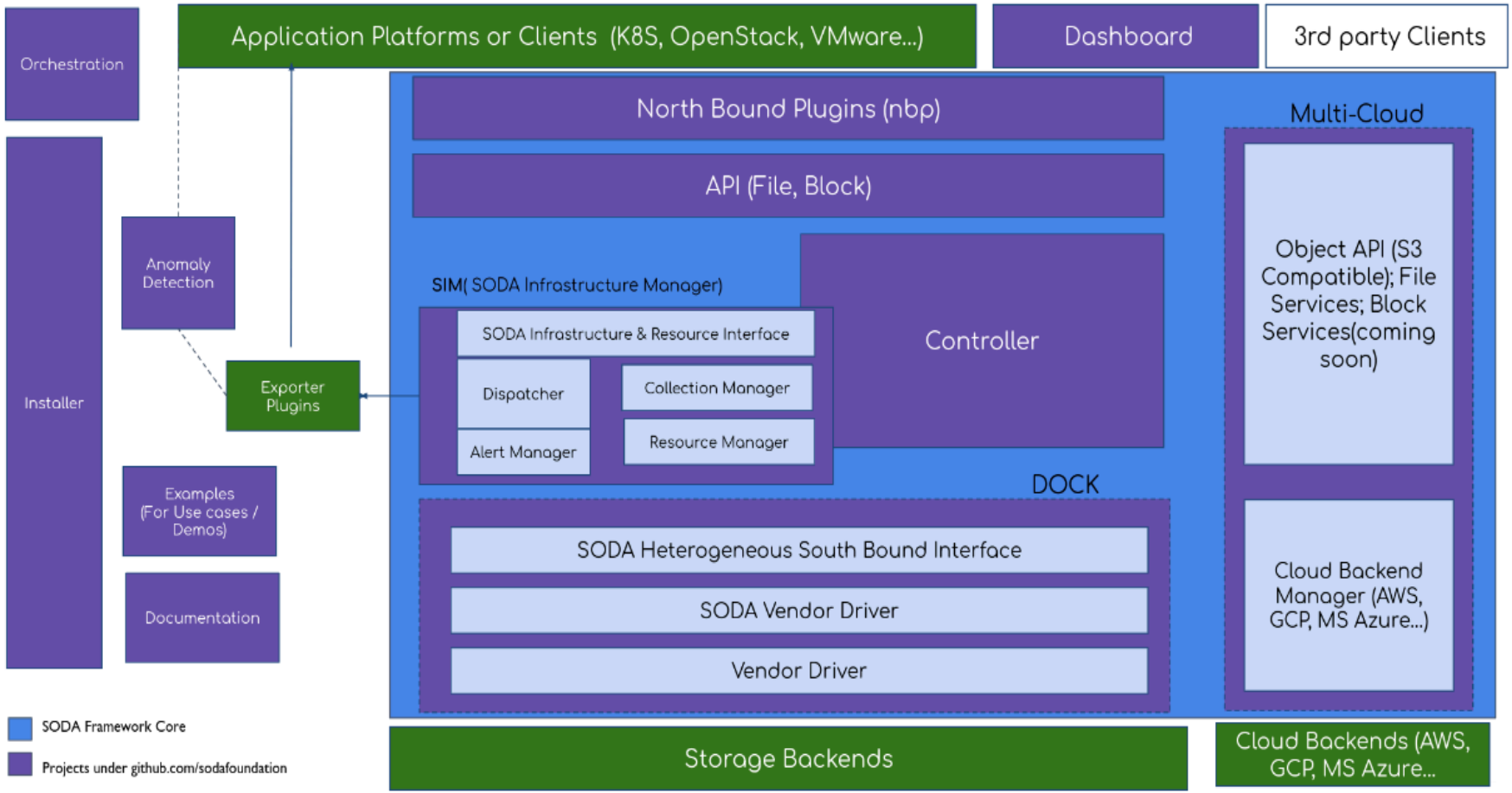


- Connects **Any Application Platforms or Clients** (Kubernetes, OpenStack, VMware...or any clients like Dashboard…) to **Any Storage** Backends(on-prem / cloud)
- **Unified Data Framework** (**Control** Plane and **Data** Plane) supporting all Data Services
- Support **Edge, Core and Cloud**

- Multiple Projects for different features and areas form the complete SODA Framework

github.com/sodafoundation

SODA Projects Architecture



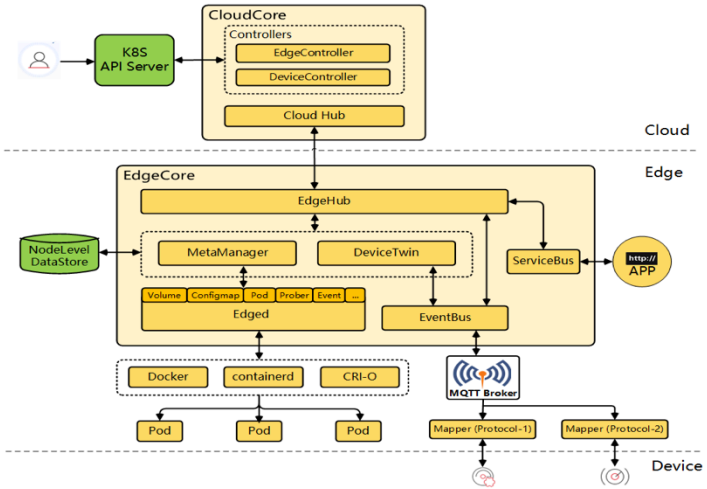
Data@Edge : What are we trying?

Where do we start?

Edge Computing Platform + Distributed & Heterogeneous Data Management Platform

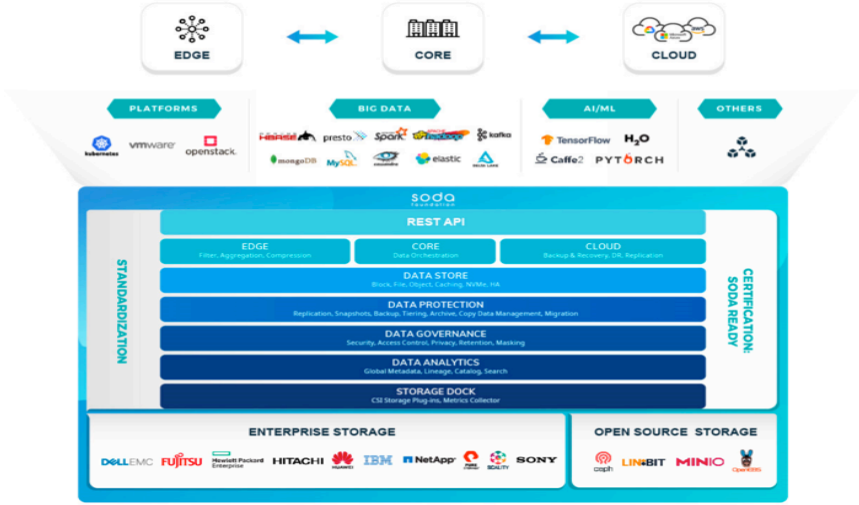
We have just started....

KubeEdge



https://github.com/kubeedge

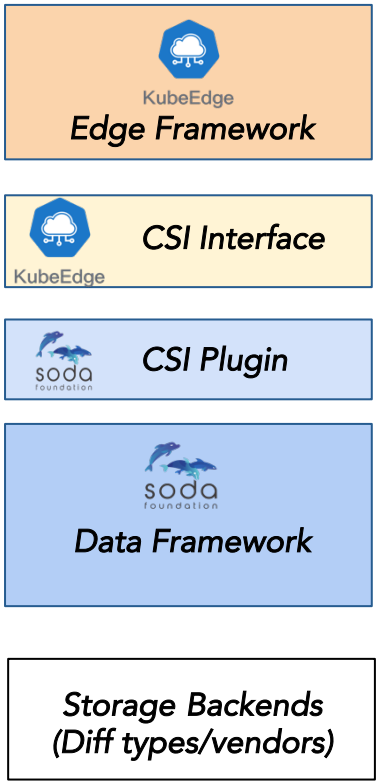
soda foundation



https://github.com/sodafoundation

soda foundation

Integration of KubeEdge and SODA



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...

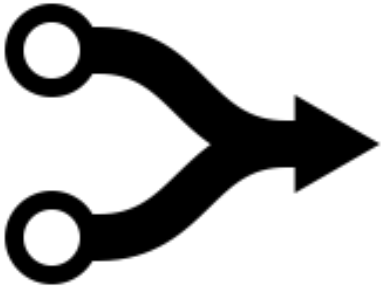
soda foundation

Data@Edge : Next

Next Plans

- *Basic PoC with KubeEdge and SODA (on a single node) - On going.*
- *Test with different storage types or vendors (on prem)*
- *Lightweight SODA*
- *Formation of SODA SIG for Data@Edge:*
  - *Better industry collaboration (Other edge/data platforms, requirements)*
  - *Research, Analysis*
  - *Recommendations to SODA Requirements*
  - *Standardization*
- *Development and Releases of the solutions with SODA Releases.*

Wanna Join?



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

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

Possibilities...

- Big potential for research (edge specific):
  - Distributed Data Management
  - Distributed Data Analytics
  - Data Energy
  - Data Compression
  - Data Privacy, Security
  - ….
- Increasing demands of edge use cases
- Data explosion
- High Business Potential : Specific Business Solutions and use cases
- ….

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/