

sodaCON 2020

DATA CONNECTED

DATA CONNECTED FOR AI, BIG DATA, CLOUD, EDGE AND 5G

Steven Tan

SODA Foundation Chair
VP & CTO Cloud Solution – Storage, Futurewei

 @stevenptan



#sodacon2020

TECHNOLOGIES IN 2020



EDGE

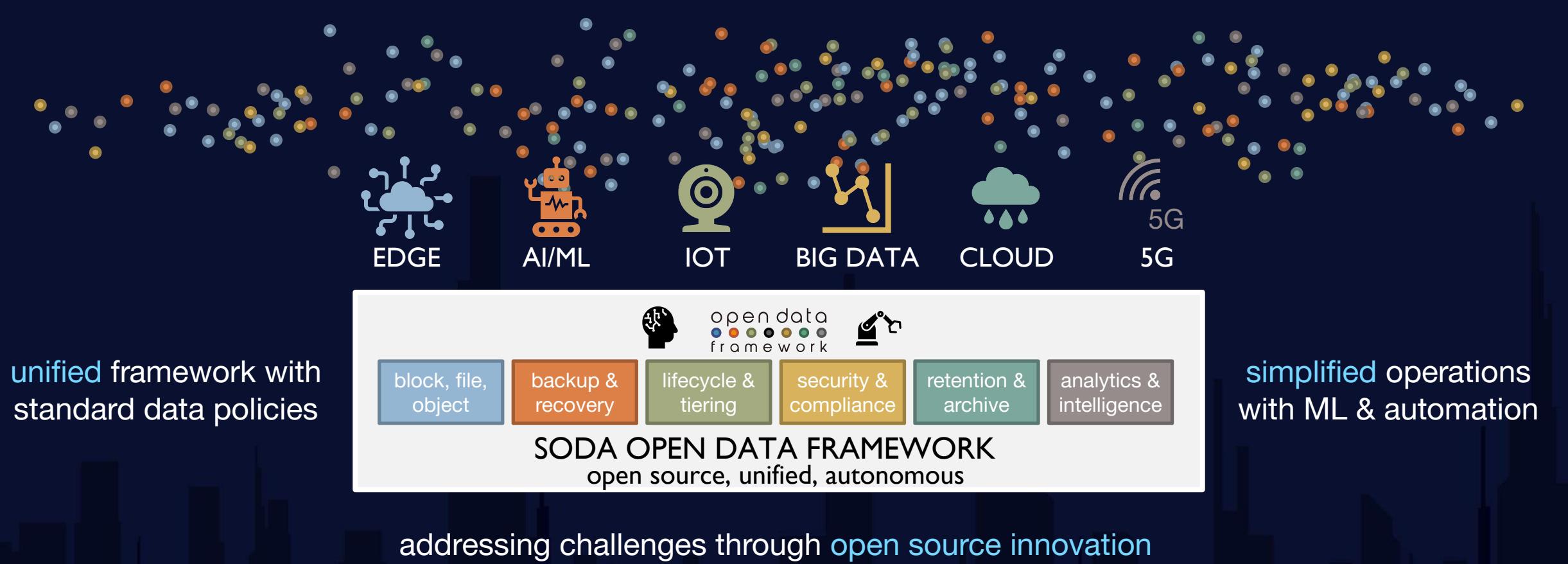
AI/ML

IOT

BIG DATA

CLOUD

5G





Services For Connected Car Platform

block, file, object
block, file, object storage for edge, DC
multicloud storage

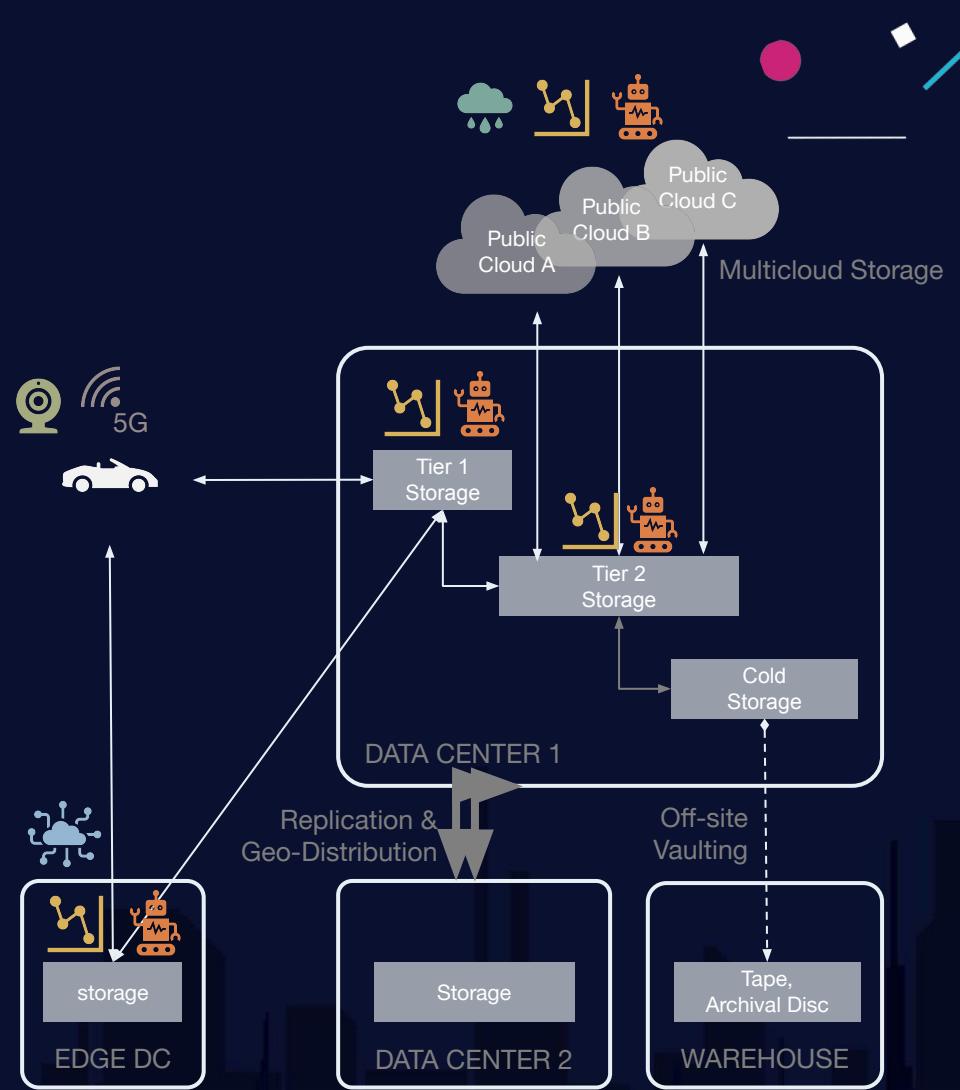
backup & recovery
backup to cloud or tape
snapshots from tier2 to cloud

lifecycle & tiering
edge to DC
tier 1 to tier 2, tier2 to cloud, tier2 to cold

security & compliance
edge, DC, & cloud security and compliance
offsite tape/archive

retention & archive
tier2 to cold storage (tape/archive disc)
tier2 to cloud for long term retention

analytics & intelligence
data integration for analytics and AI/ML applications



60PB of vehicle data goes to the DC each month
~20GB/month/vehicle x 3M vehicles. source: AECC



sodacon 2020

#sodacon2020

sodaCON 2020

DATA CONNECTED

CCB'S CLOUD DATABASE PLATFORM PRACTICE
BASED ON KUBERNETES & SODA

Xion Lai

Cloud Platform R&D Department, General Manager

China Construction Bank Fintech



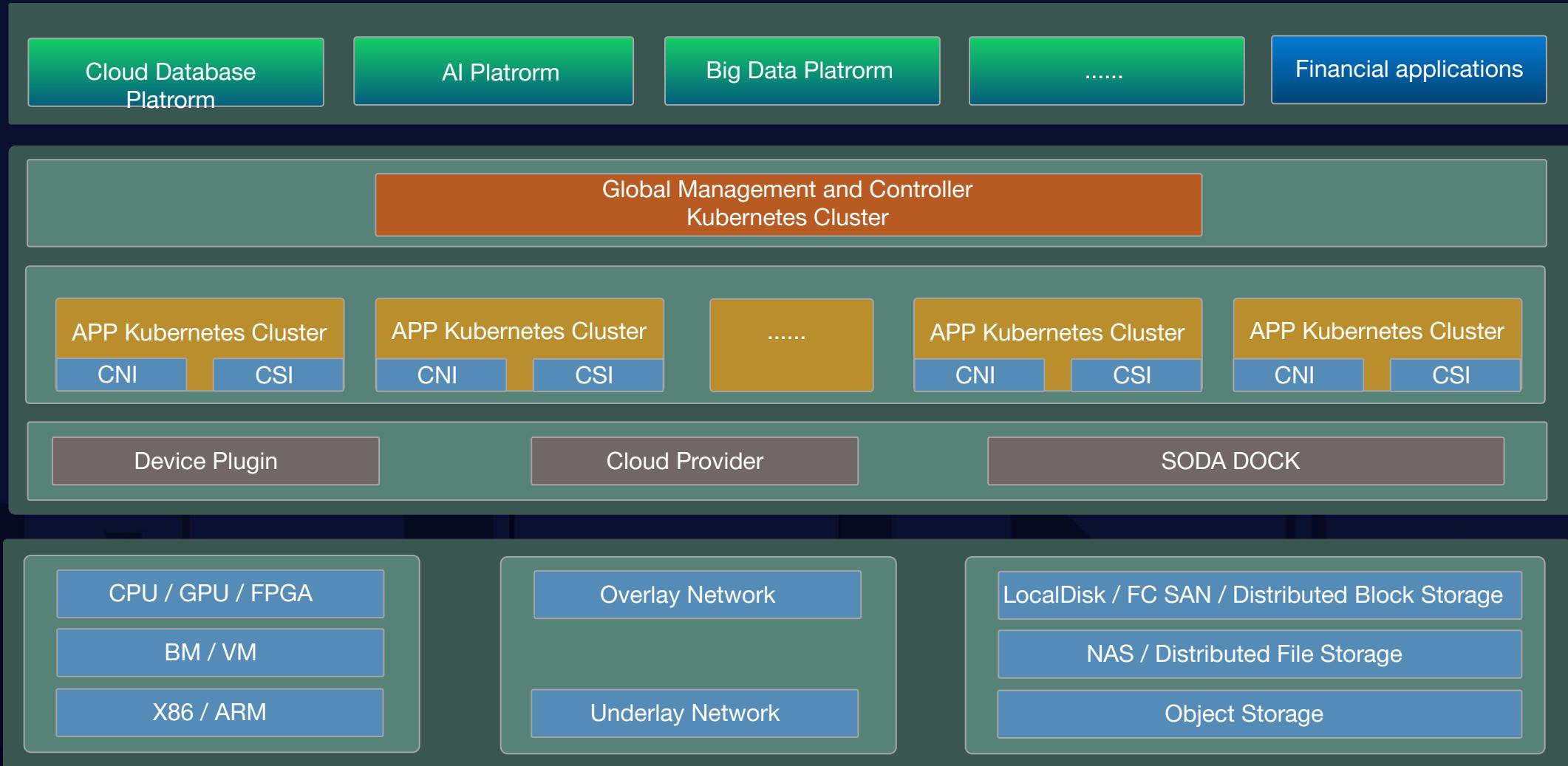
#sodacon2020

COMPANY OVERVIEW

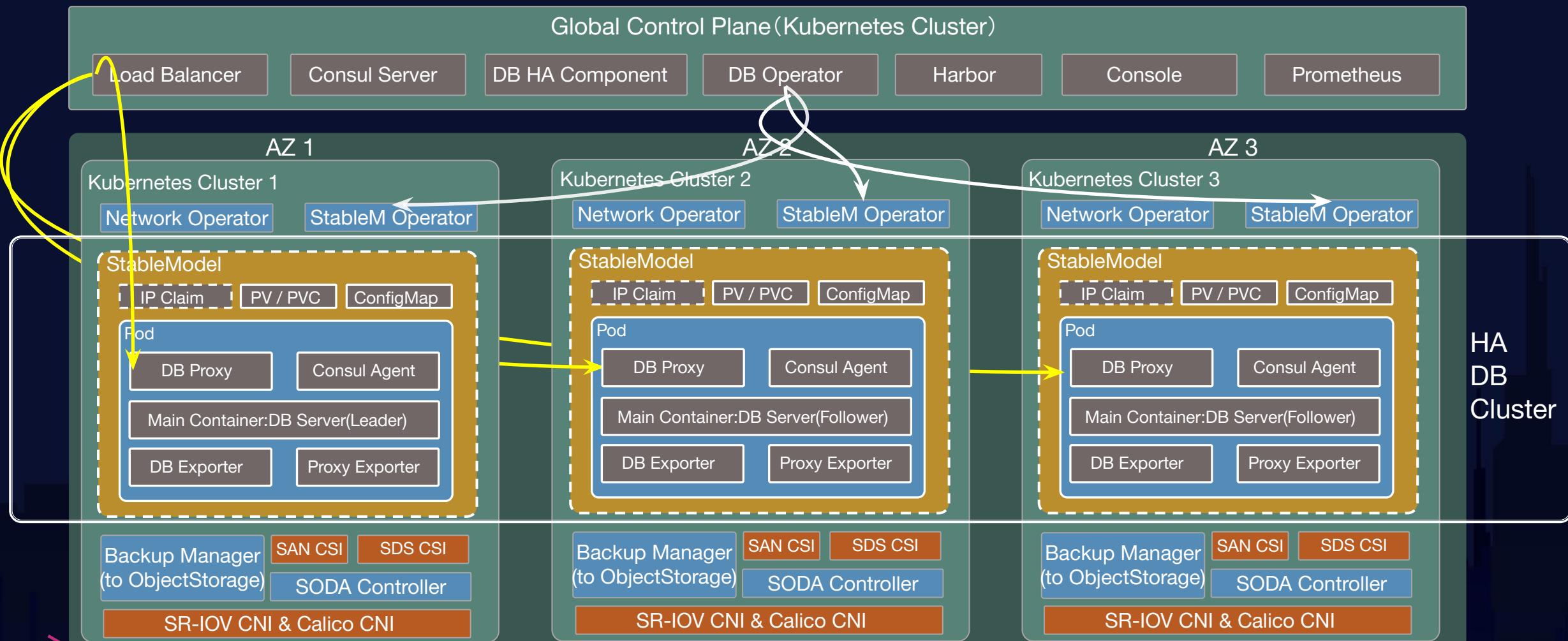


- Ranked 2nd in 'TOP 1000 WORLD BANKS 2020'
--<The Banker Magazine>
- 30th in 'GLOBAL FORTUNE 500, 2020'
--<Fortune Magazine>
- First fintech company held by China nationalized major banks
- Development centers in 7 major cities of China, 6300+ employees

CCB'S CLOUDNATIVE PLATFORM OVERVIEW



CLOUD DATABASE PLATFORM BASED ON KUBERNETES&SODA

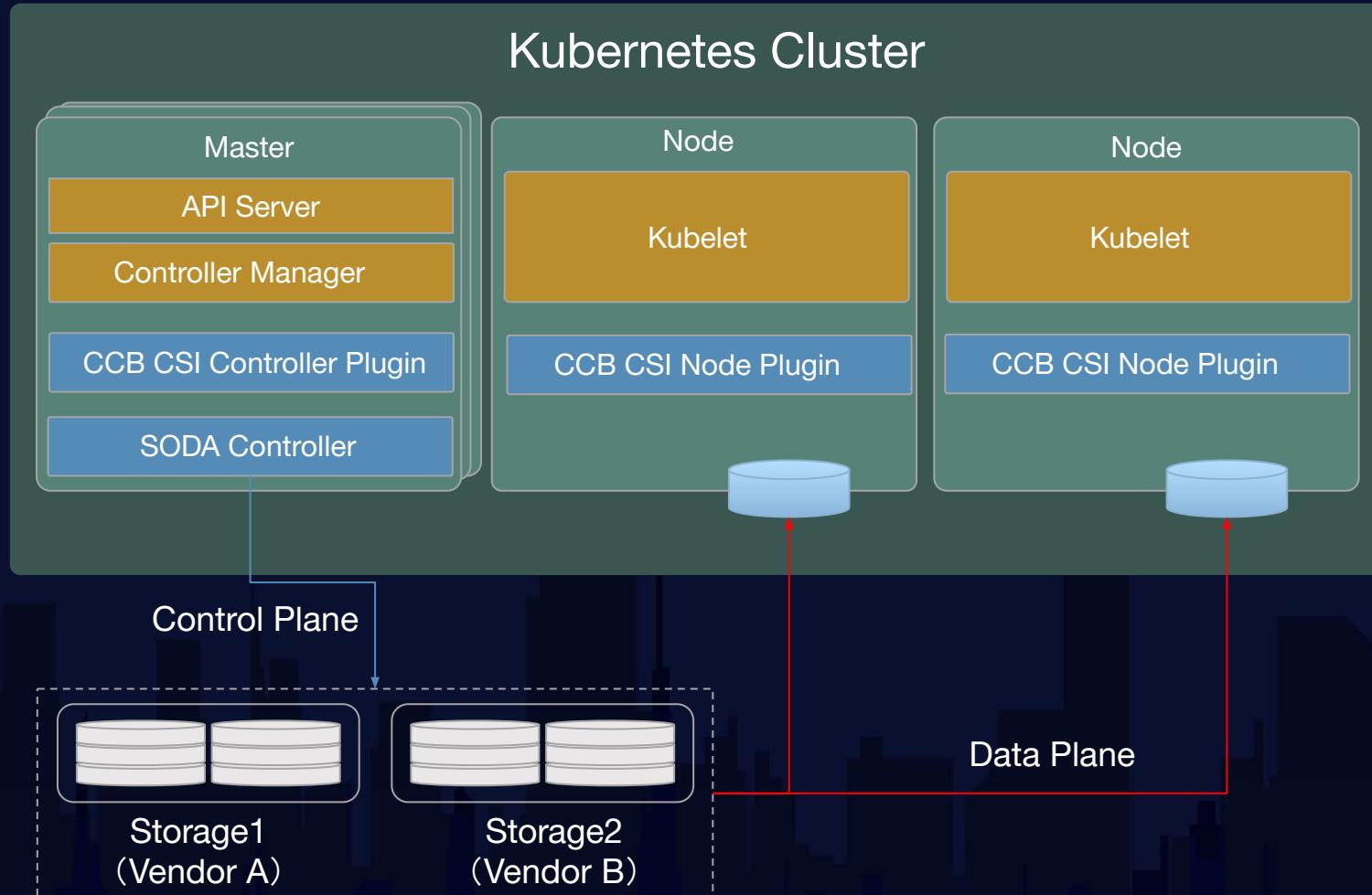


ACHIEVEMENTS OF THE PLATFORM

The screenshot shows the MySQL (k8s版) management interface. On the left, there's a sidebar with navigation links: 建行云, 总览, 云产品, MySQL(k8s版), MySQL服务, Redis服务, 告警管理, 操作日志, and 项目信息. The main content area has tabs: 详细信息, 监控, 数据库管理, 参数设置, 用户管理, 任务日志, 备份管理, and 慢日志. The 详细信息 tab is selected. It displays basic attributes like 实例名称: mysql-tqwiuryr, GTID 信息: 已开启, 访问地址: ..., 端口: 3306, APPID, VPCID, SubnetID, 所属用户ID, 标签, and 创建时间: 2019-09-30 17:40:52. Below that is a 配置信息 section with 架构: 高可用版, 实例规格: 4核 8G 200GB, 数据库版本: 5.7.28, 部署方式: 多可用区, 主库所在可用区: az1, 从库所在可用区: az2, az3, and 可维护时间段: 23:00-05:00. To the right is a 拓扑图 (Topology Map) showing a master node (mysql-tqwiuryr-1) connected to three slave nodes (mysql-tqwiuryr-2, mysql-tqwiuryr-3) in different availability zones (az1, az2, az3). A replication lag of 0M is indicated. There's also a 备份中心 (Backup Center) icon. At the bottom is a 运行信息 (Runtime Information) section with 复制信息: 半同步, 复制状态: 运行中, 任务: 无运行中的任务.

- **Scale:** 600+ database clusters delivered in one year
- **Efficiency:** HA database cluster deployed within 1 minute
- **Security:** Multiple data protection mechanisms, ensuring data security
- **Automation:** Automated maintenance, seconds level event perception, error self-healing, minimizing db maintenance cost

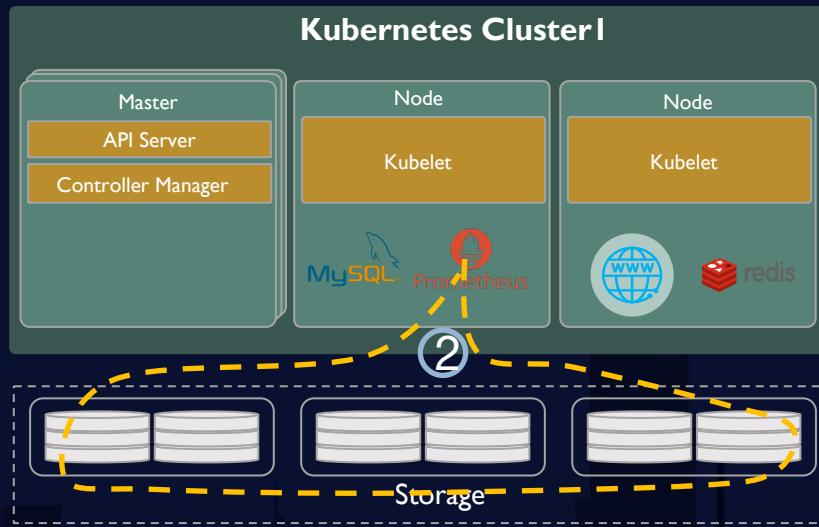
HOW CAN SODA HELP



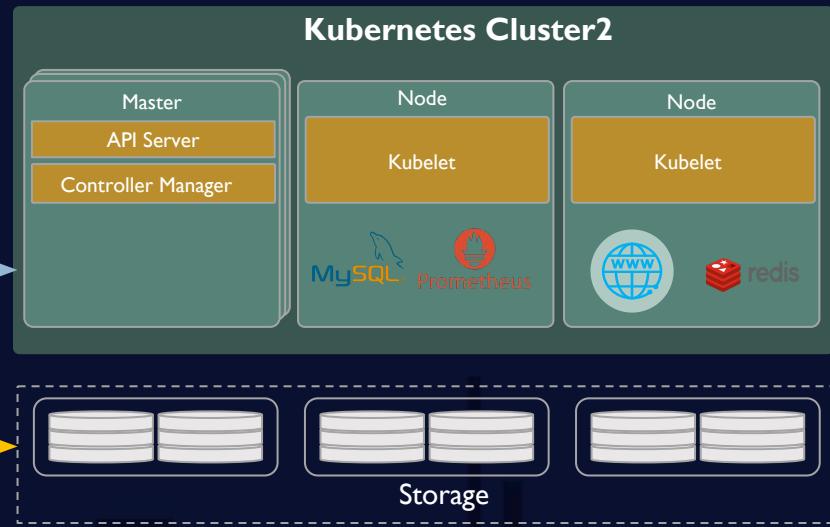
- Single plugin, minimizing problems of deployment and maintenance caused by multiple CSI plugins(CCB SAN CSI, CCB NAS CSI, CCB SDS CSI..etc)
- SODA Dock as interface for different storage backends to shield protocol differences
- Provide a scalable monitoring framework where backends and client exporters can be managed dynamically

CHALLENGES IN THE FUTURE

Data Center1



Data Center2



- Data migration between different Kubernetes clusters and storage;
- Unified monitoring of storage infrastructure through Prometheus.

THANK YOU

xionlai.cloud@gmail.com

sodacon 2020
DATA CONNECTED

sodaCON 2020

DATA CONNECTED

#sodacon2020

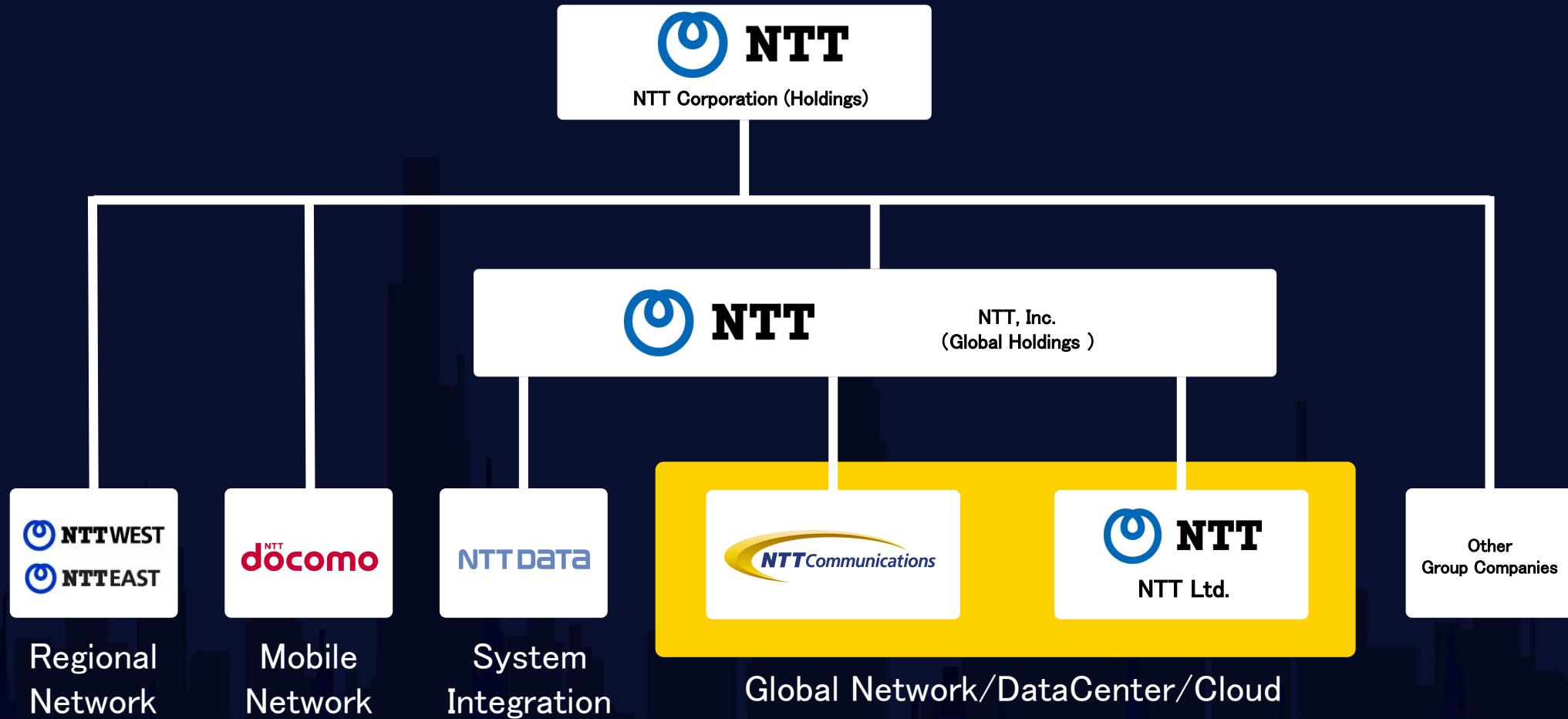
SERVICE PROVIDER USE-CASE AND PROSPECTS

Kei Kusunoki

TOC Co-Chair, SODA Foundation
Storage Architect at NTT Communications, Japan



JAPANESE TELECOM CAREER GROUP



INFRASTRUCTURE BUSSINESS DOMAIN

- Global Network
 - Inter Cloud/DC Network
 - Closed Network for on-premises
- Data Center
- Private/Public Cloud Service
 - VMware based IaaS
 - OpenStack based IaaS

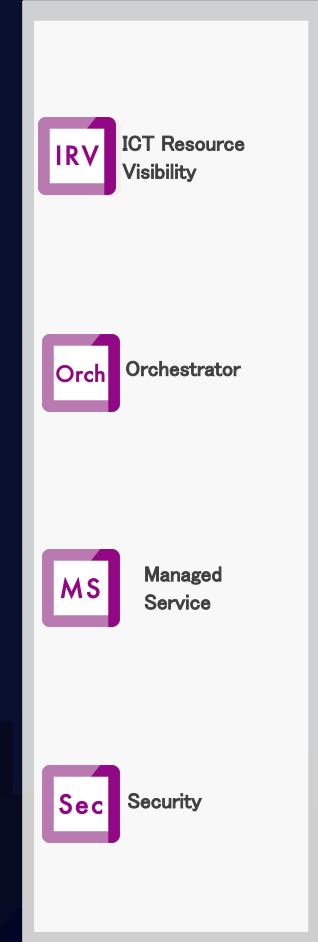


NTT COMMUNICATION'S PLATFORM

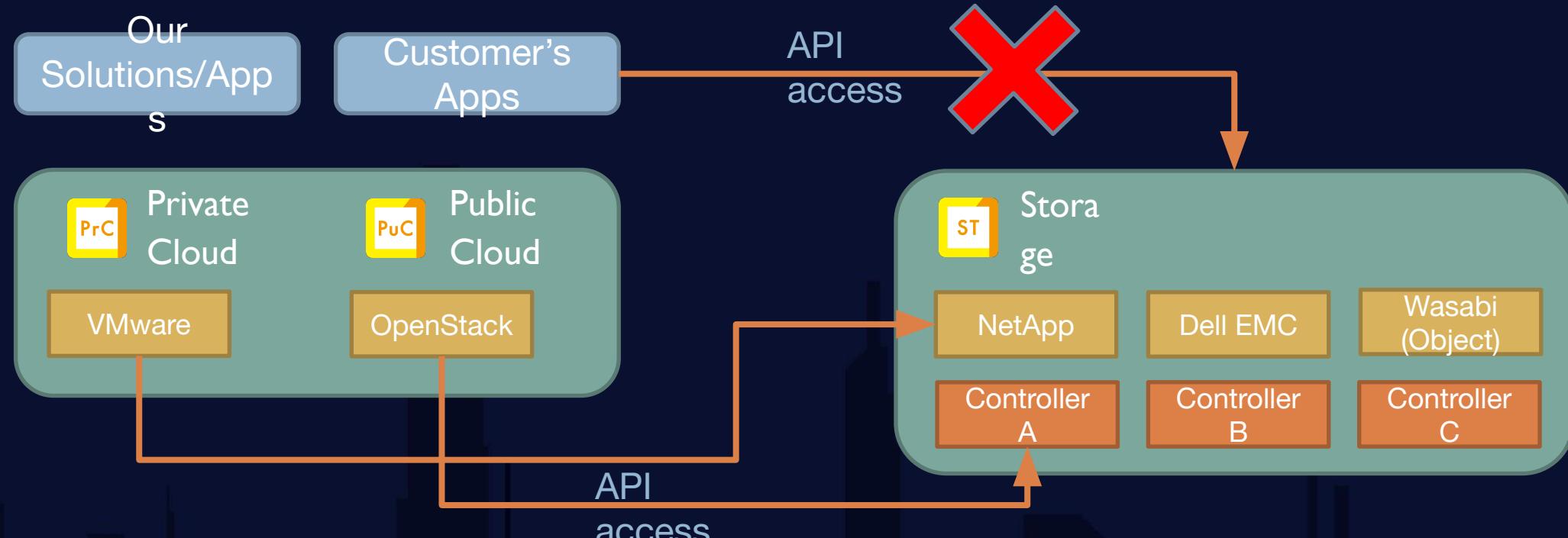
Apps on Smart Data Platform



Smart Data Platform

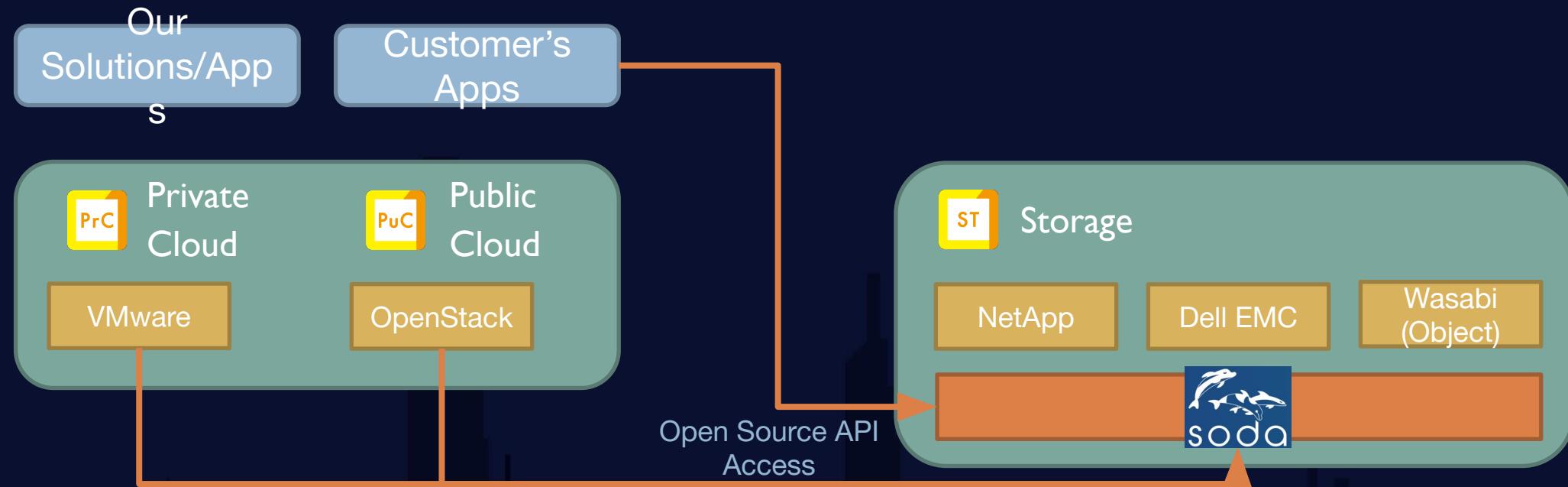


PAIN POINT ON CURRENT STORAGE ORCHESTRATION



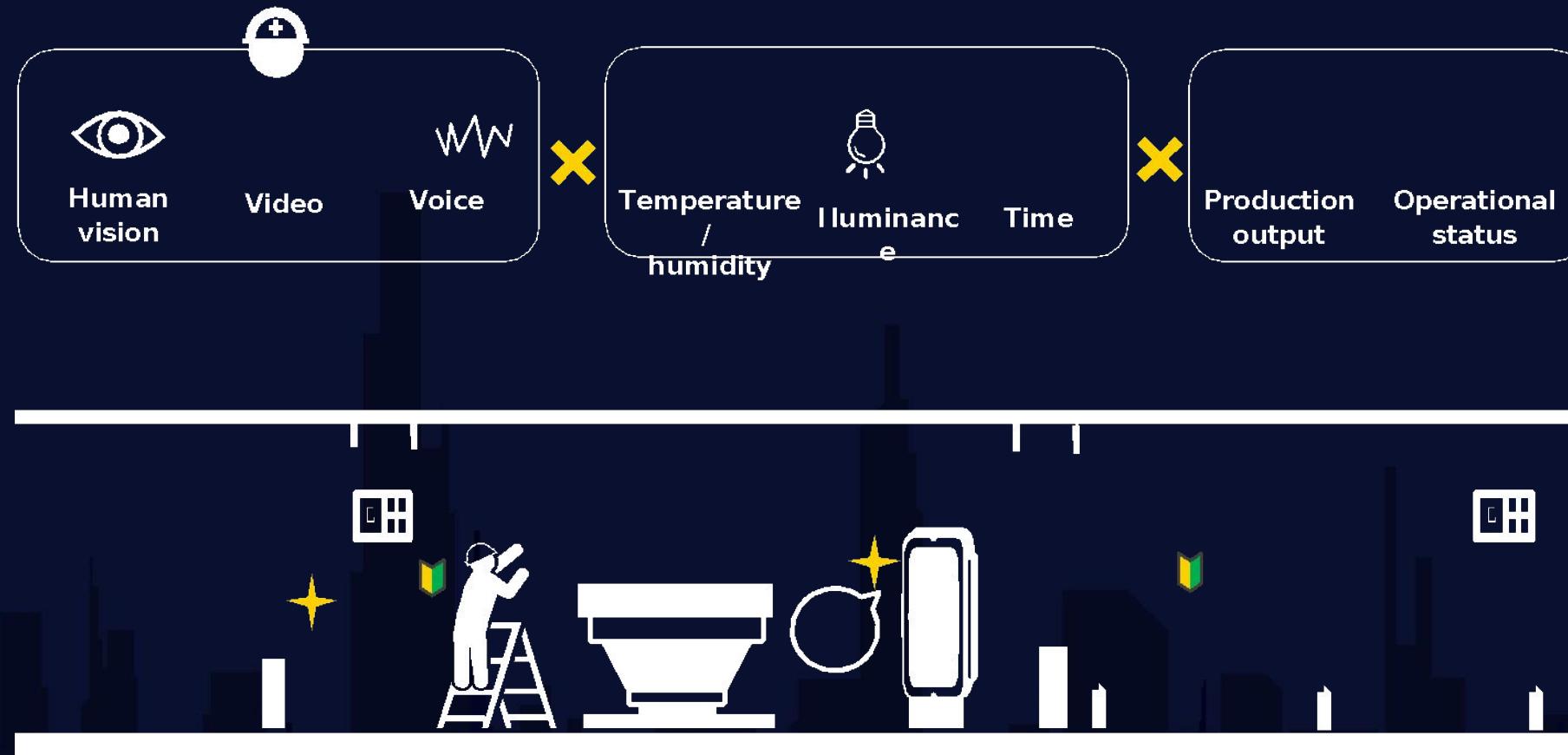
- High Development Cost for original controllers
- Fragmented controllers have original API
 - Customer cannot utilize Storage directly from their apps

SODA UNIFIED STORAGE ORCHESTRATION



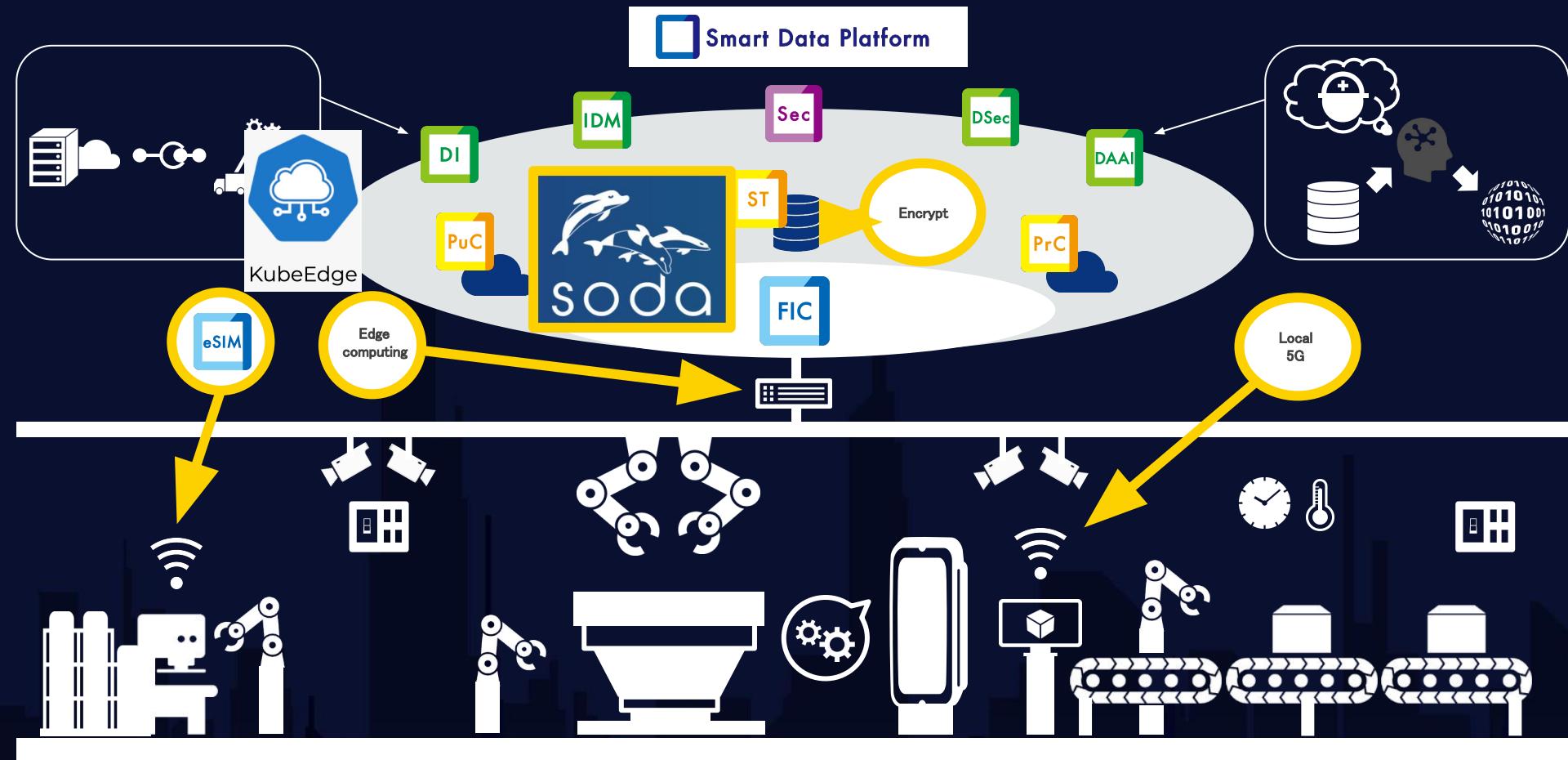
- With SODA ready Storage
 - No need for controller development
 - Ease of Data Management
- Open Source API can be used by Customer's Apps

OTHER SODA USECASE EXAMPLE: DIVERSE DATA IN FACTORY



Solution Example on Smart Data Platform

SODA INTEGRATION IMAGE

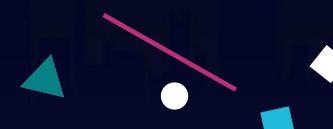


SODA will enable IoT devices to utilize Data



OUR ACTIVITY OF THE YEAR

- PoC about VMware and NetApp
 - VMware vCenter Integration
 - Basic CRUD operation through NetApp Driver
 - Result: Some fundamental features don't work
- Cold Storage and Object Storage Usecase Discussion in community



PROSPECTS FOR SODA

- Further Storage Driver Support
 - More Vendor
 - More development to current Storage Drivers
 - FeedBack and Next PoC
- Object Storage Usecase/Architecture Discussion
 - Service Provider Requirements such as Reliability/Scalability
 - New usecase (Ex. Datalake)



INTRODUCTION TO CHINA MOBILE CLOUD STORAGE AND ITS REQUIREMENTS IN MULTI-CLOUD MANAGEMENT

Deng Jin

Storage Service Architect

China Mobile

COMPANY OVERVIEW



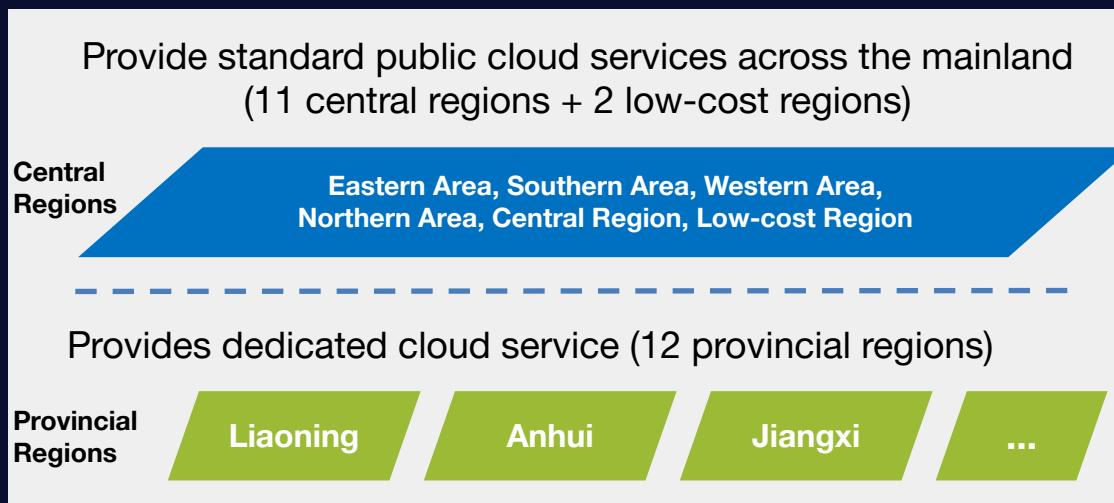
- The world's largest telecommunications operator
- No. 65 on the Fortune Global 500 in 2020



- CMCC's public cloud services (big cloud) were provided in 2015
- Number of online customers is up to 710,000 by Dec 2020

BIGCLOUD INFRASTRUCTURE OVERVIEW

With the "**Centralized Management, Unified Division, and Hierarchical Construction**" principle, China Mobile offers 25 regions across the mainland, and **the provincial coverage rate exceeds 80%** (As of December 2020, excluding Hong Kong, Macao and Taiwan).



Certifications

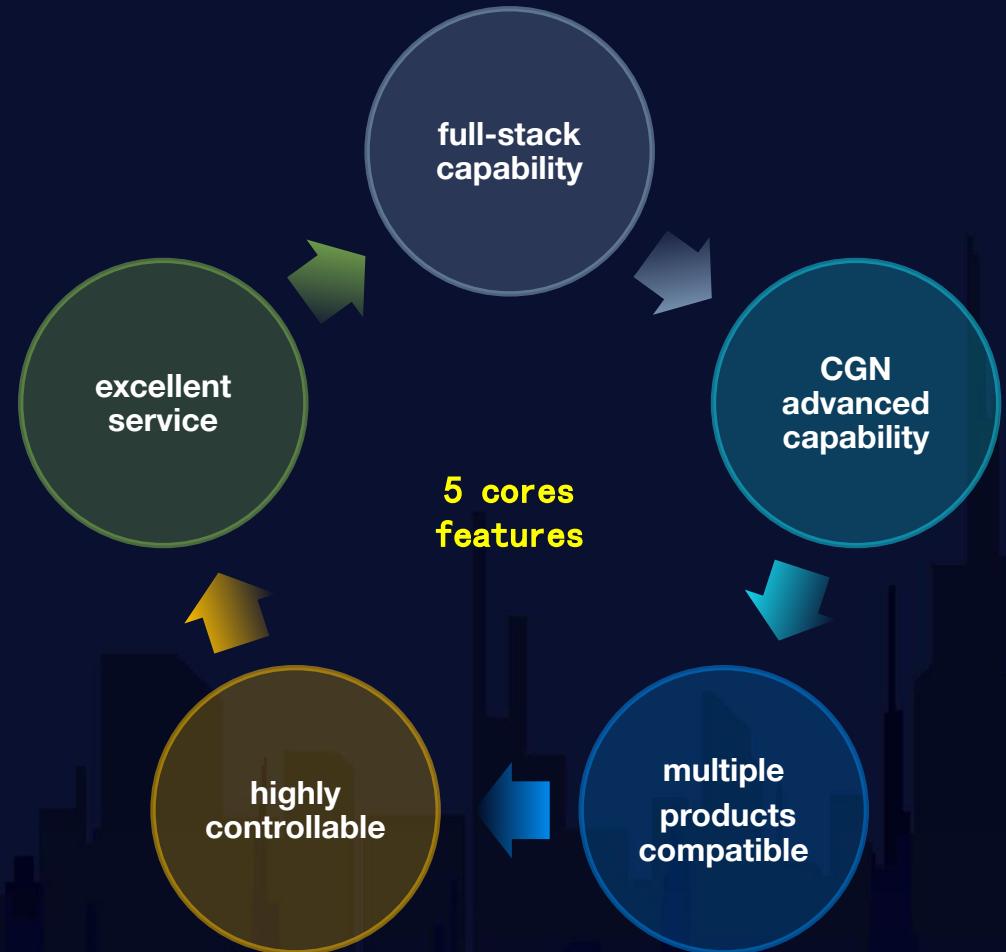


BIGCLOUD PLATFORM OVERVIEW



provides

sodacon 2020



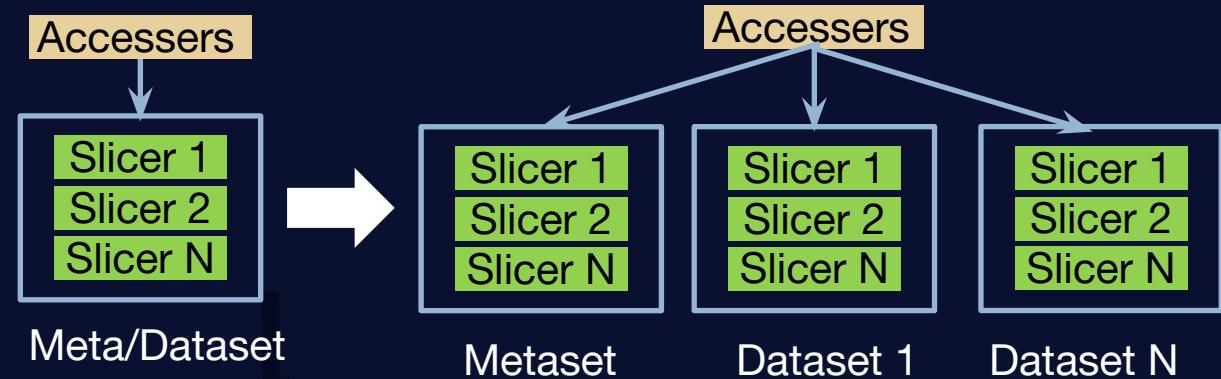
#sodacon2020

OBJECT STORAGE PROBLEMS & SOLUTIONS

Challenge:

- Capacity is limited by cluster size
 - A single bucket can hold only several petabytes of data
 - A single bucket can hold no more than one billion objects
- Expansion will slow down user requests
 - Data rebalance has a great impact on service

Solutions:



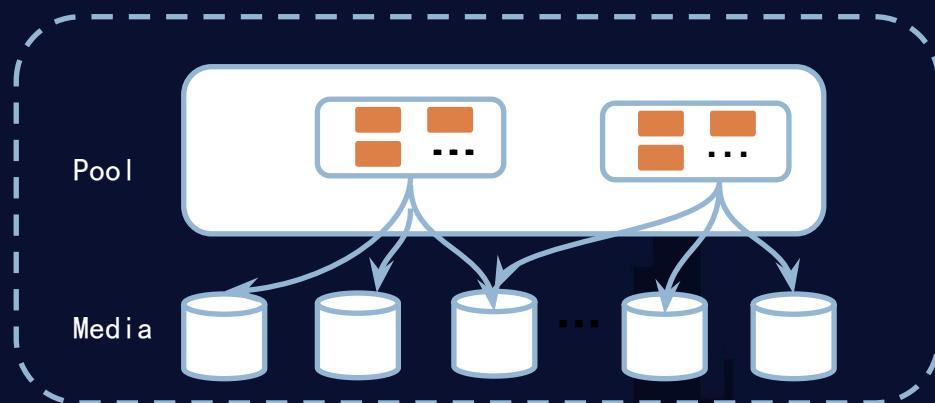
- ✓ Single Bucket can hold over 100 petabytes of data
- ✓ Single Bucket can hold over 10 billions of objects
- ✓ No data rebalance in expansion

Coming soon:

- Erasure code across multiple Azs/regions
- Multi-region active-active architecture

BLOCK/FILE STORAGE PROBLEMS & SOLUTIONS

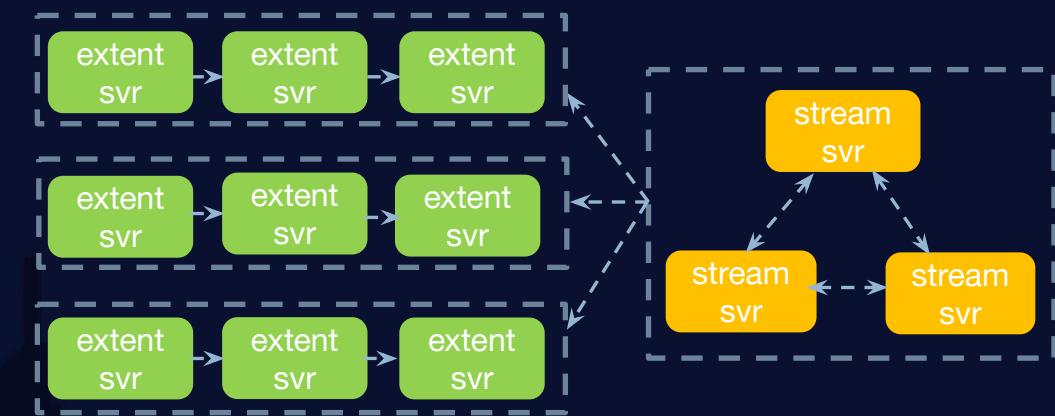
Current implementation:



Challenge:

- Too many queues and threads switch in an IO request loop
- Too many locks for synchronization between threads
- Mixed processing of synchronous and asynchronous requests
- Limited cluster scale

New Plan:



Advantage:

- Light thread, lockless programming
- Network optimization, RDMA support
- Full path user space IO
- Large-scale expansion

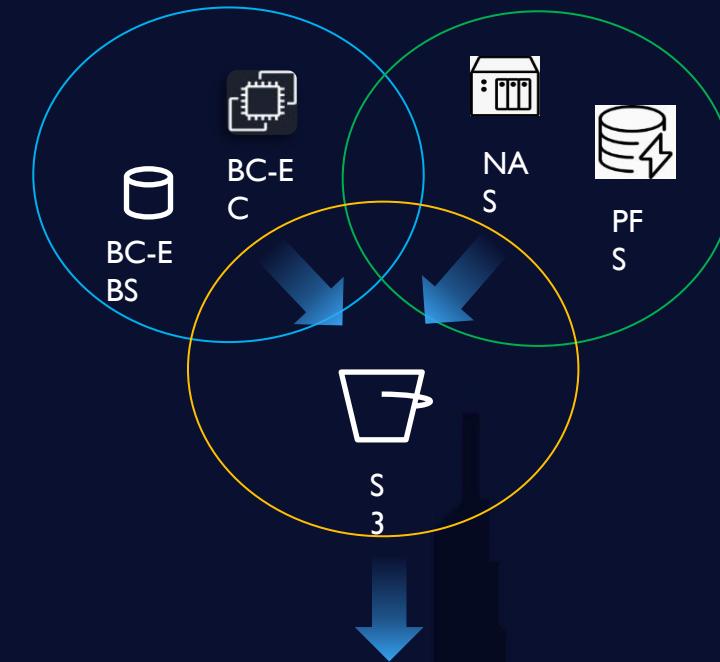
HOW CAN SODA HELP

BACKUP

Backup block/file storage to the object-based storage by S3 interface

- Requirement:

Provides more powerful disaster resilience ability (for example, backup block/file storage to multi-cloud s3 storage)



OBJECT STORAGE

Standard Storage
Low-frequency Storage
Archive Storage

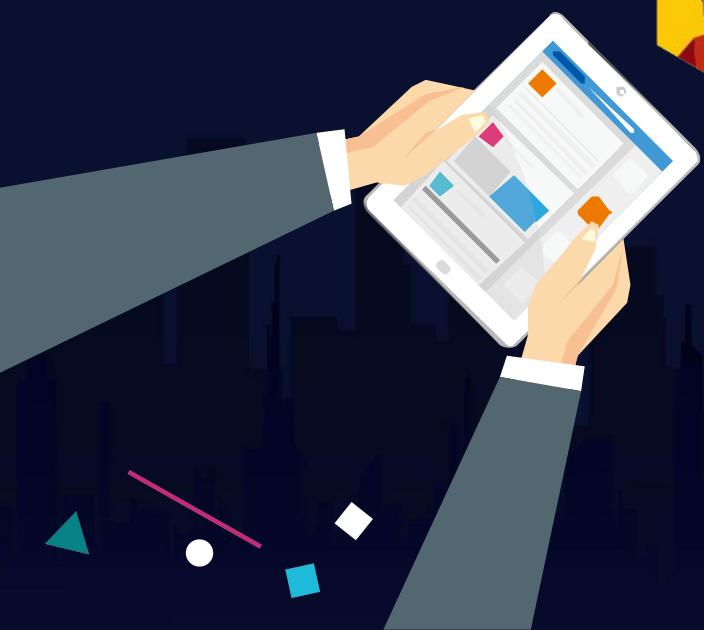
- Requirement:

Integrate China Mobile big cloud S3 storage into SODA multi-cloud object lifecycle management



SUMMARY

**Embracing open source
Giving back to the community**





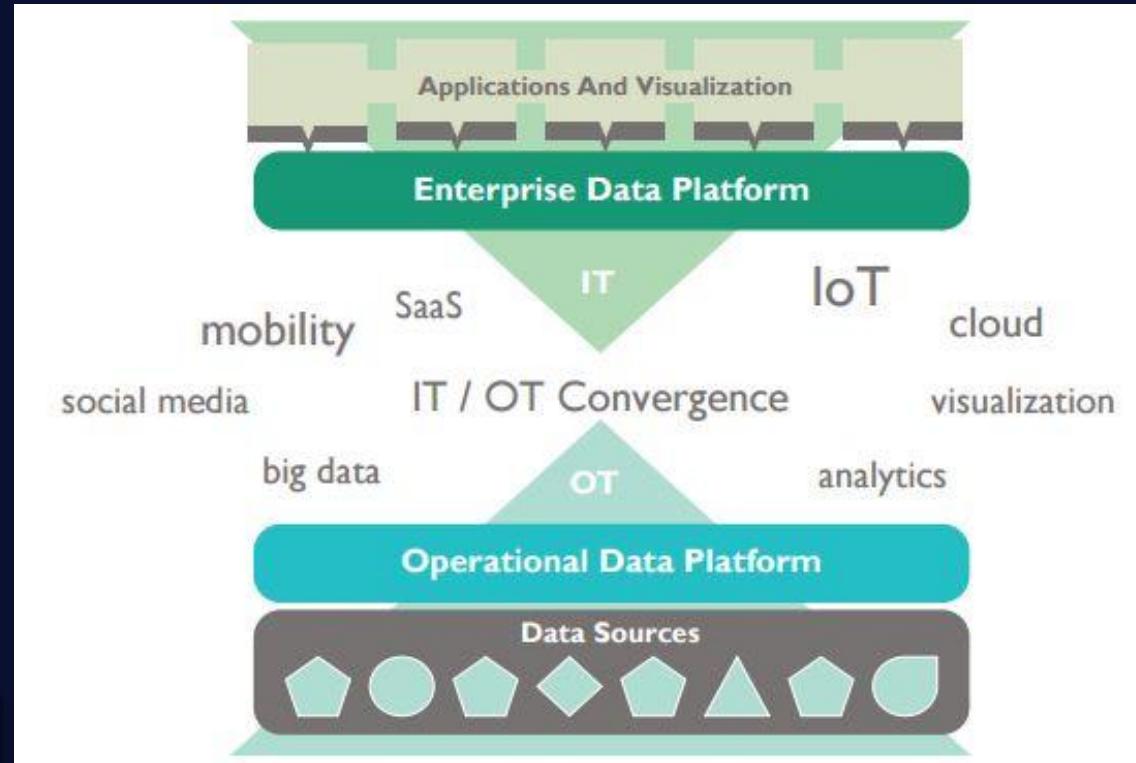
CHALLENGES AND COORDINATION ON DATA MANAGEMENT OF INDUSTRIAL INTERNET ERA

Wang Yousheng

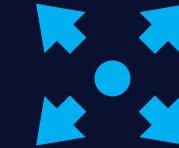
Director of Open Source Developer Department

Huawei

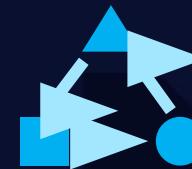
INDUSTRIAL INTERNET ERA NEEDS A DATA MANAGEMENT REDESIGN



Data Analysis/AI



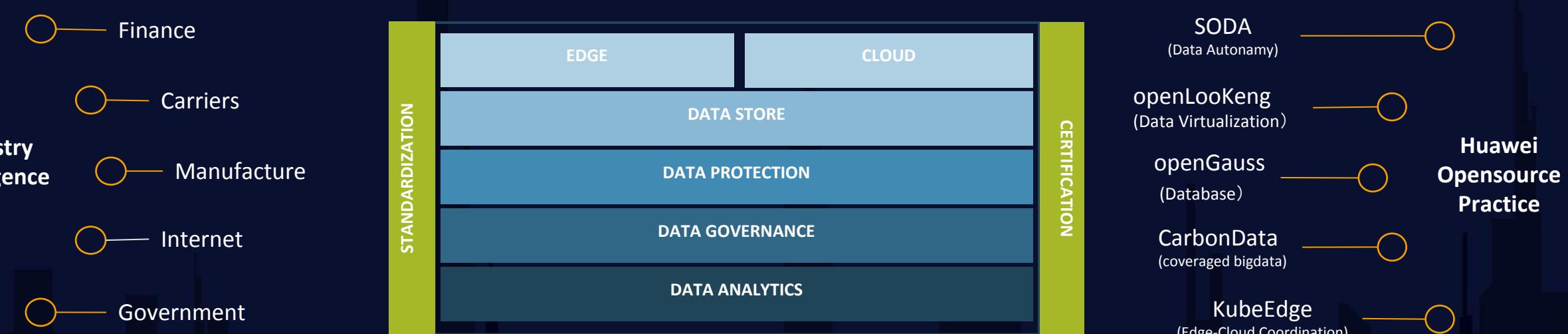
Data Distributed & Scalable



Data Interoperable

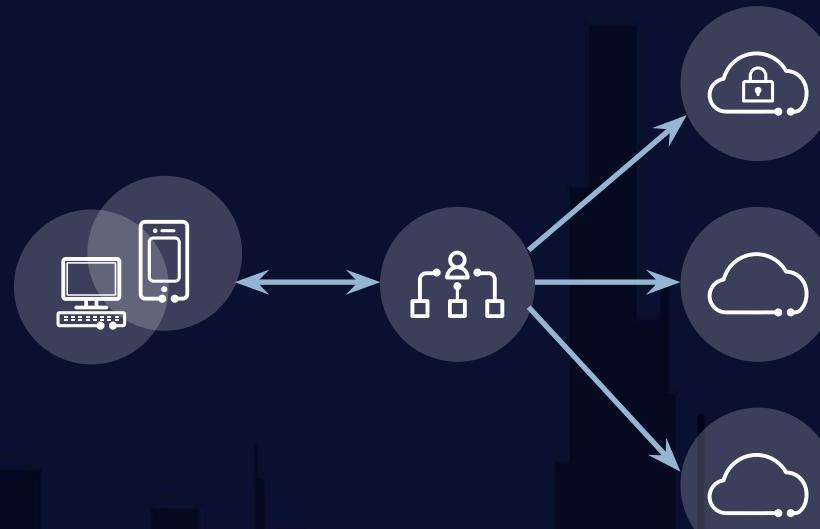
Redefine Open Architecture & Innovative Technologies for Whole Ecosystem Become a Must

EMBRACING OPEN SOURCE TO PROMOTE OPEN INNOVATION



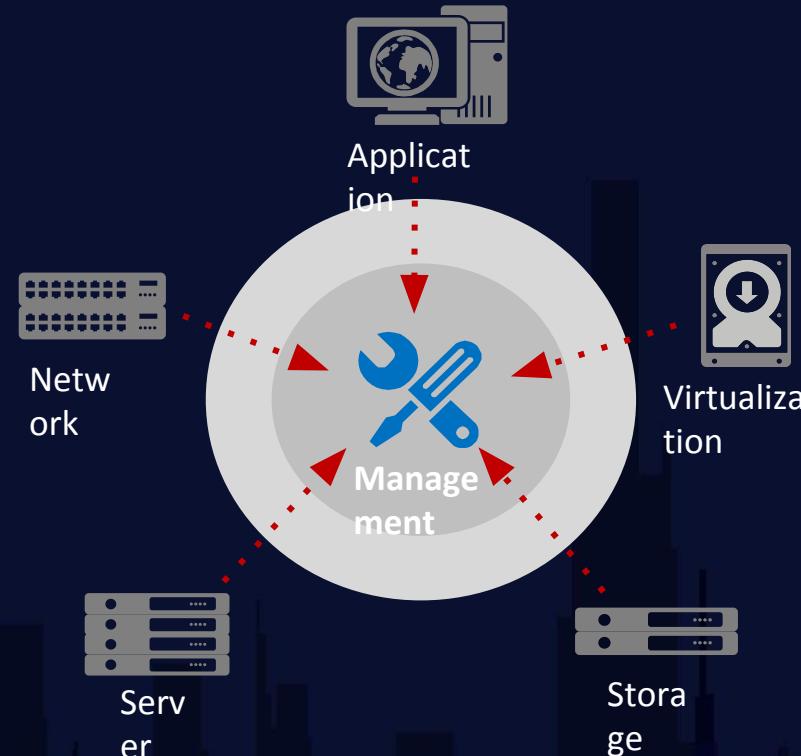
CHALLENGES IN HYBRID CLOUD

Hybrid Data across Edge, Core and Cloud



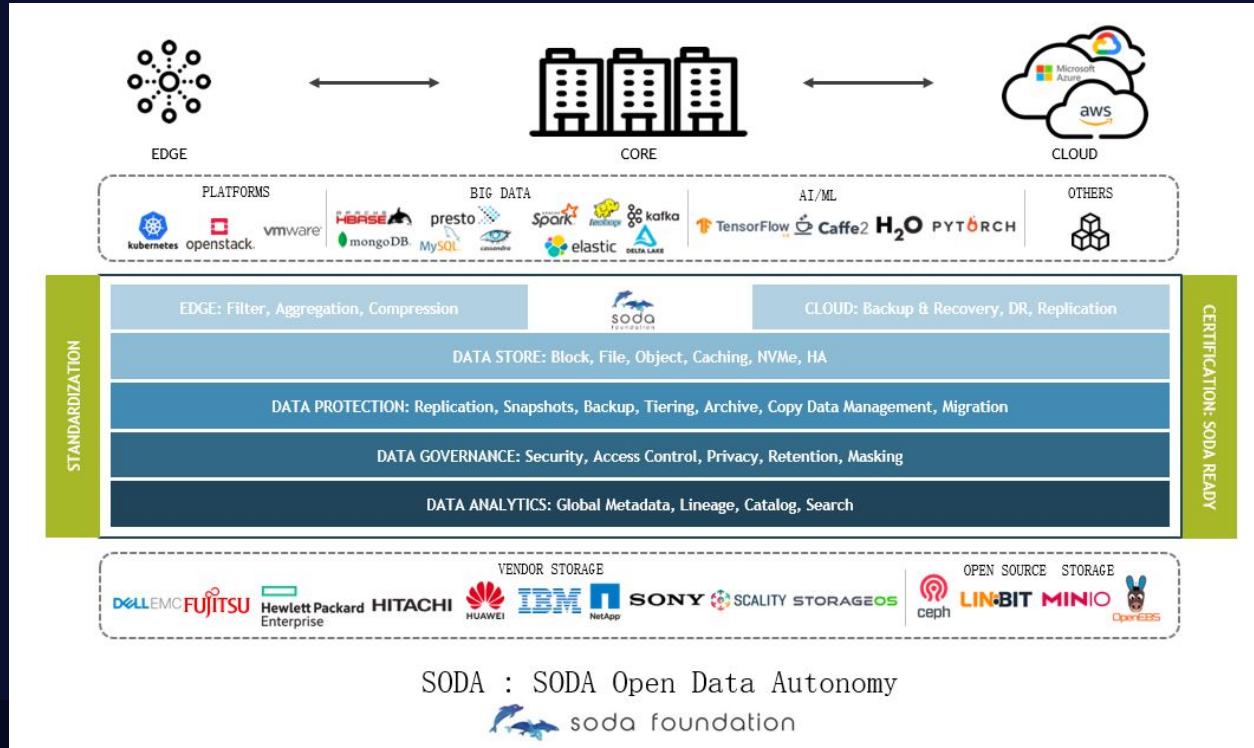
- Different requirements for edge, cloud and core
- Different Standards and technologies based on vendors or platforms
- Data from Multi-Source increases data cost and difficulty to keep data consistency
- Seamless Data mobility on demands
- Data governance and compliance (like GDPR)
- Data Security and Protection across different interfaces

CHALLENGES IN DATA CENTER



- Difficulty to manage heterogeneous storages and resources
- Low storage utilization and huge resource waste, without automatic control
- Difficulty to get consolidated prediction and analytics of capacity, fault, and performance trends
- Lack ability of Self Governance and Self healing

HOW CAN SODA HELP



- Unified API for data and storage management to adapt for multi-vendors
- Unified platform to build data solutions for different use cases and technologies
- Unified reference framework in hybrid cloud to manage global metadata with complex characteristics and multi-vendors

Collaboration, Innovation, and Success!

THANK YOU

sodacon 2020
DATA CONNECTED