

# Cloud-native & Data-centric AI+X Infrastructure and Platform

**Keynote Talk @ IoTCloudServe@TEIN**  
**3rd Collaboration Community Meeting**

Dr. JongWon Kim

Networked Intelligence Laboratory  
AI Graduate School  
Gwangju Institute of Science & Technology (GIST)





# Smart + X

---

Providing Intelligence

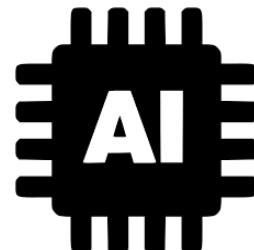
All services that are  
Flexible and Adaptable

Provides Flexible/Agile  
**SmartX User-defined**  
Services and Infrastructure

# AI + X

AI for X & X with AI

Data-  
Networking-  
AI  
for  
AI > ML > DL



(Tools)

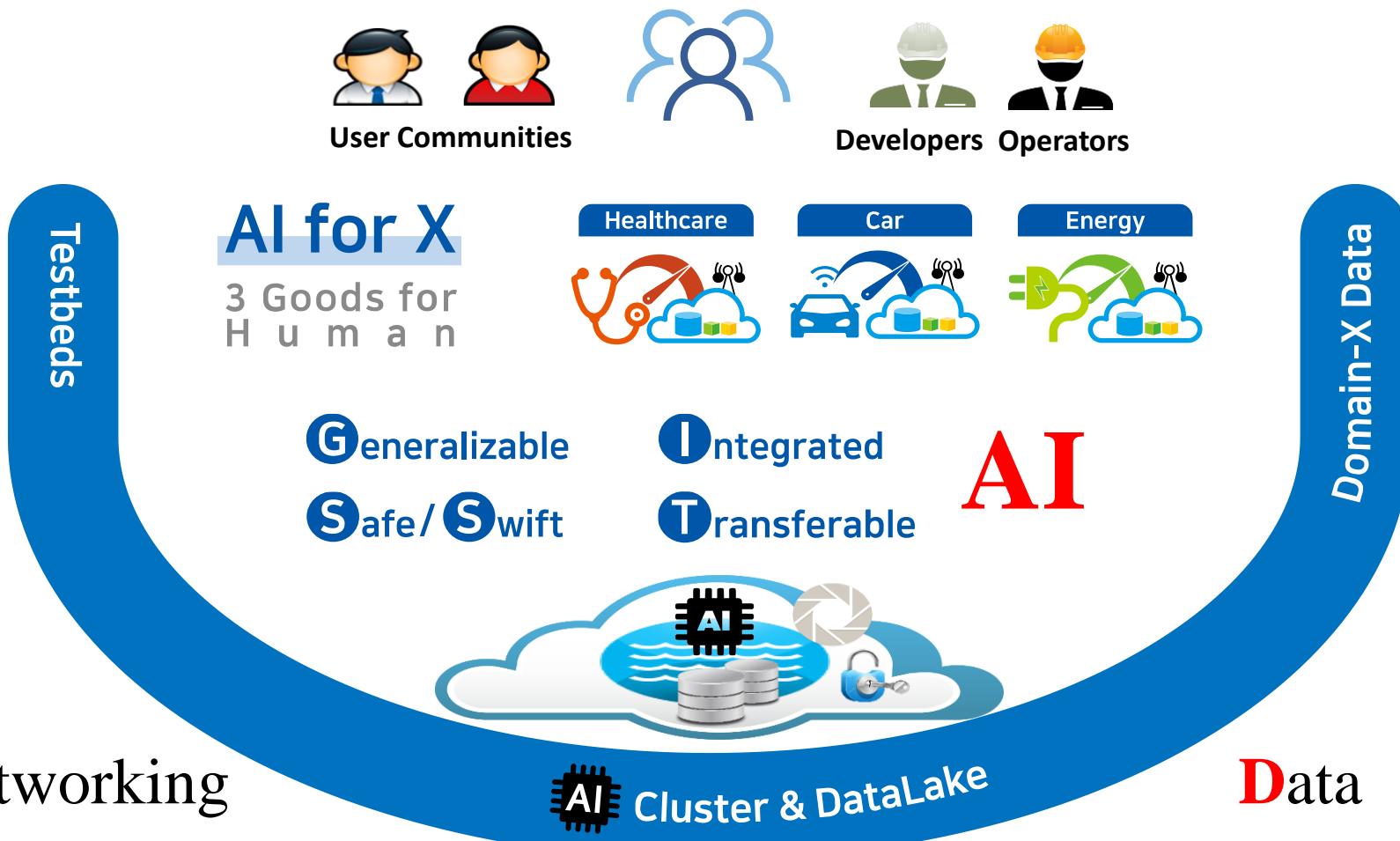
Data X

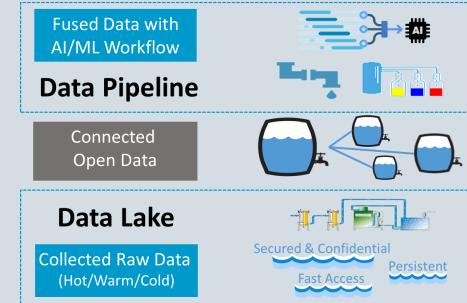
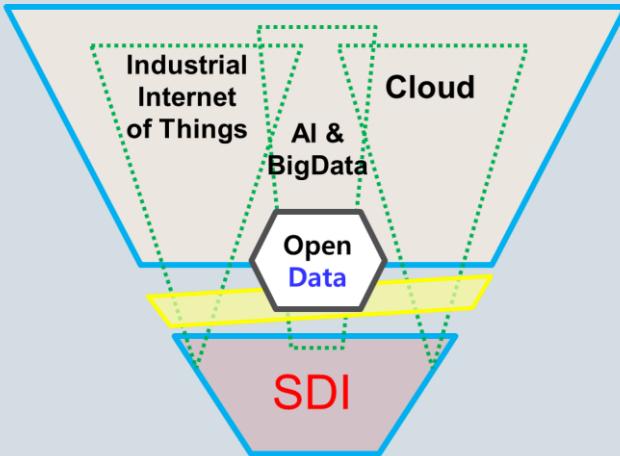
(Domains)

Core Fusion / Inter-disciplinary

# Shared Infrastructure & Common Platform for AI+X Ecosystem

## Private-Public Partnership (PPP)





# Cloud-native & Data-centric Infrastructure for AI+X Ecosystem

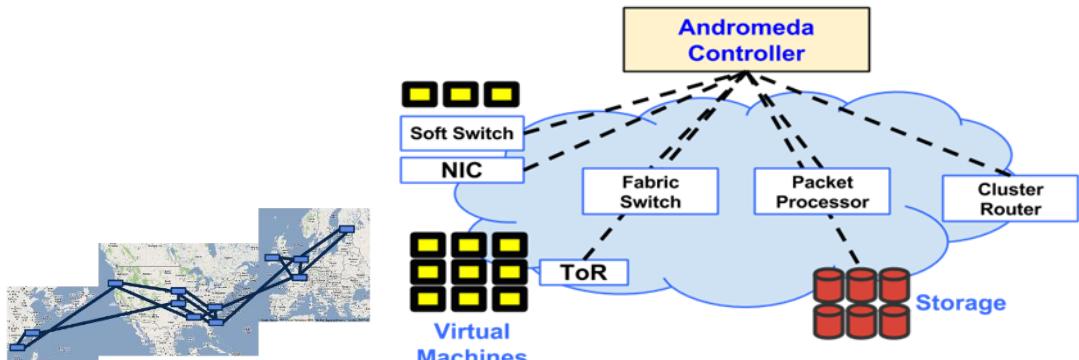
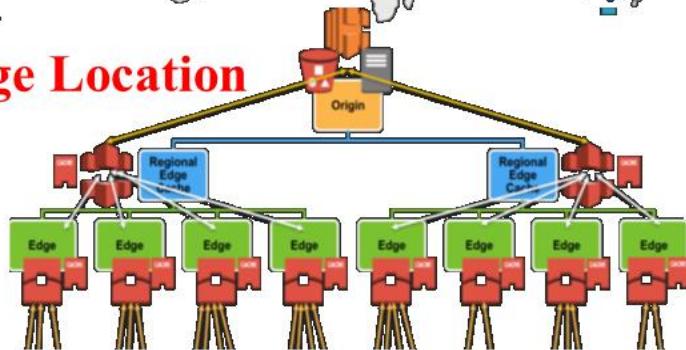
# Worldwide SDI @ Hyper-scale Cloud Players



## Regions

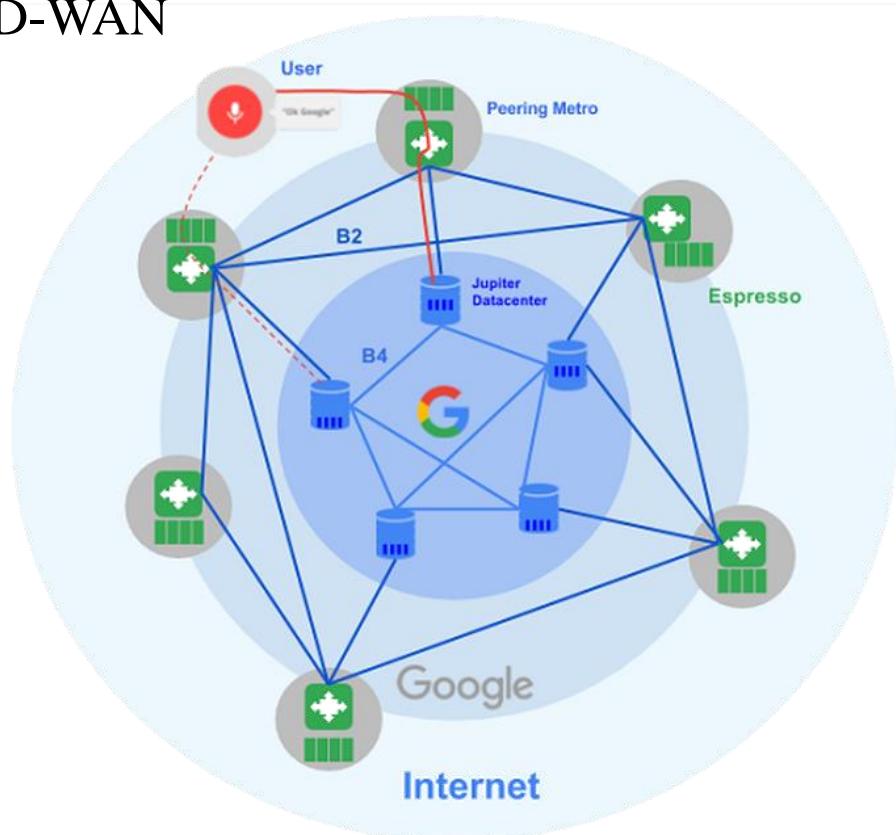


## Edge Location



**B4** Global SD-WAN

## Andromeda SDN/NFV



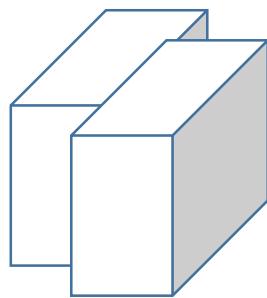
# Cloud-native: From Baremetal to Containers / Serverless

Monolithic Apps

N-tier Apps

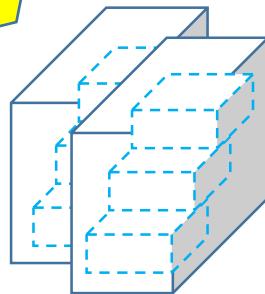
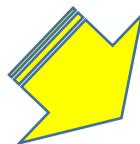
**Micro-services**

Nano-services



Cloud (VMs, BM, ...)

Baremetal Boxes

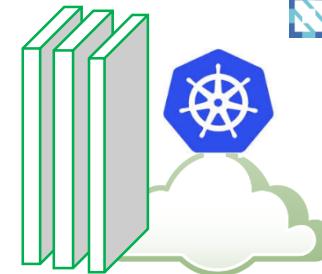


Virtual Machines



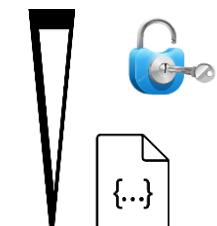
Serverless  
(functions as a service)

CLOUD NATIVE COMPUTING FOUNDATION



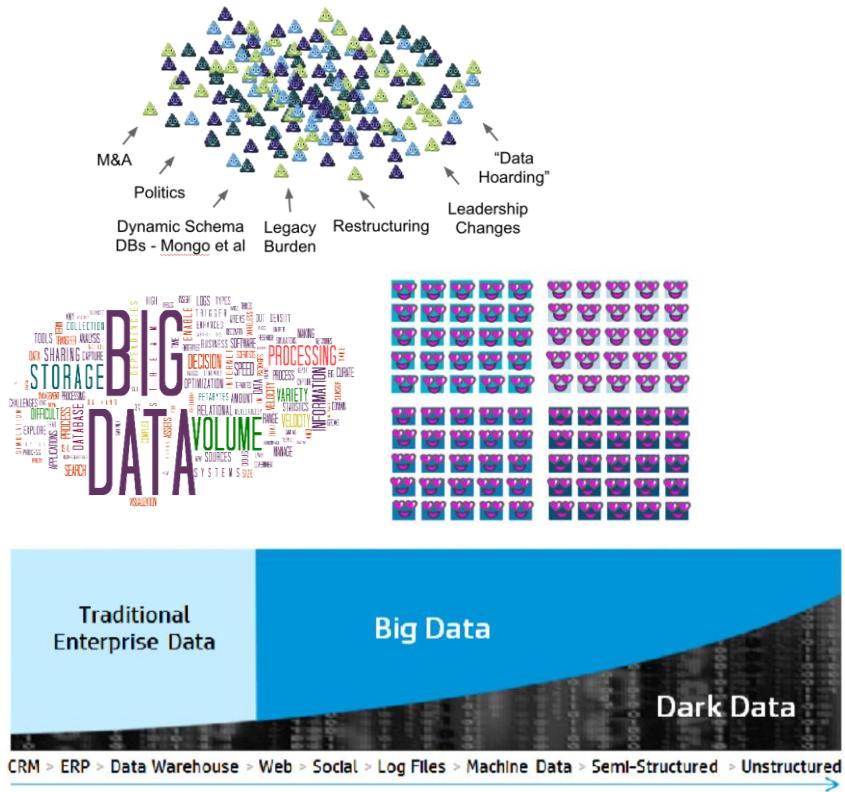
Containers

(Docker, Singularity, LXD, ...) <sub>6</sub>

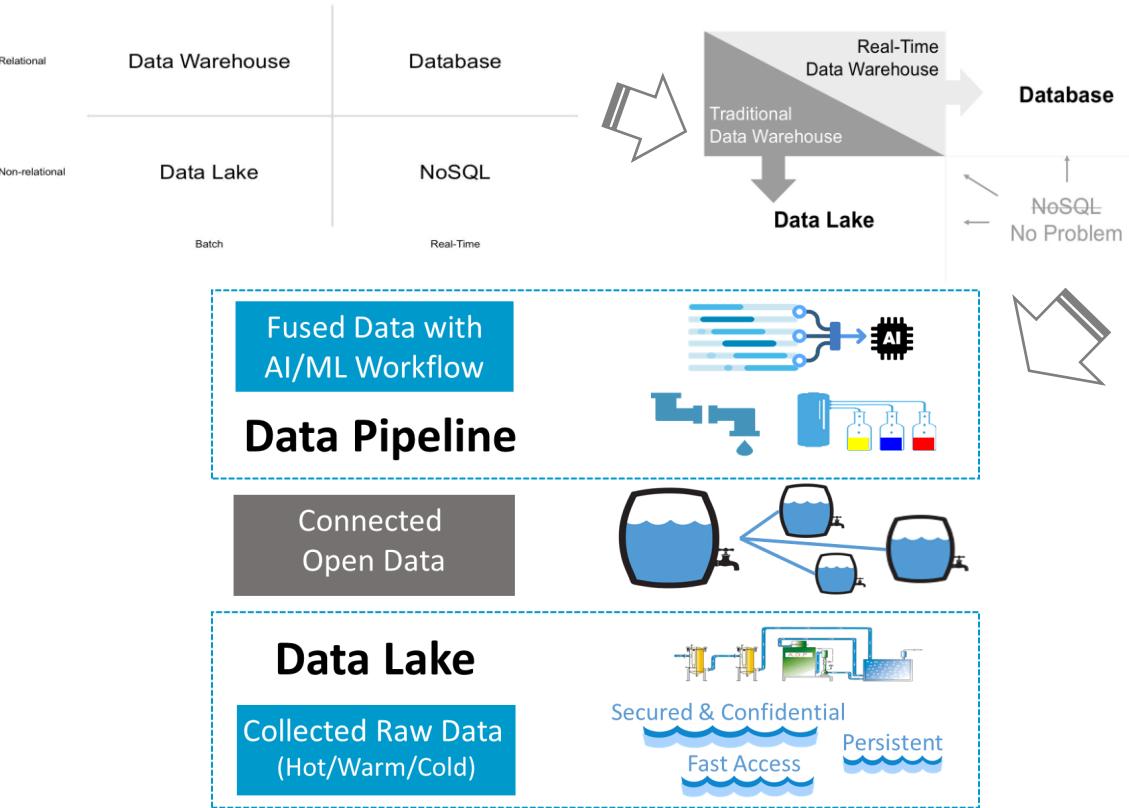


unikernel

# Open & Connected Data Landscape



## (Dark, Raw → Big) Data



# → Insight & Wisdom

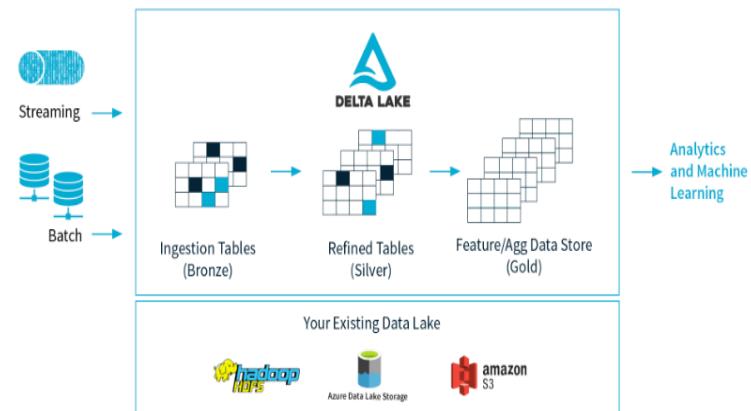
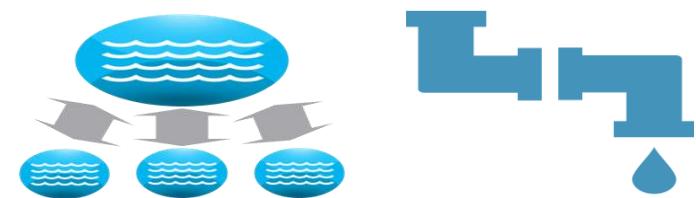
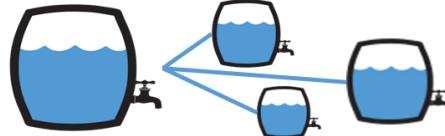
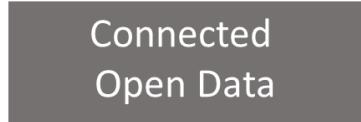
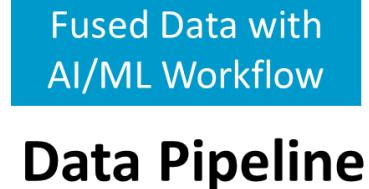
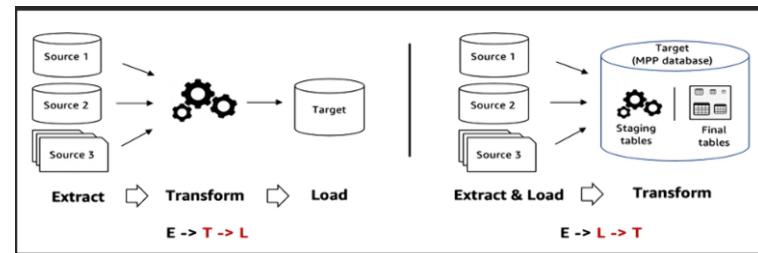
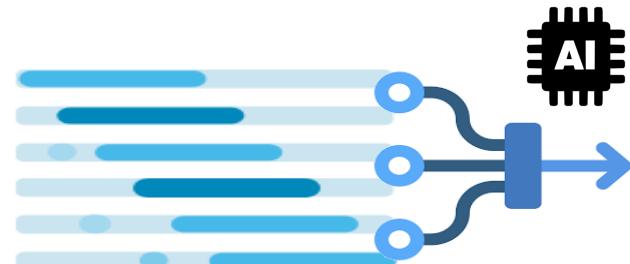
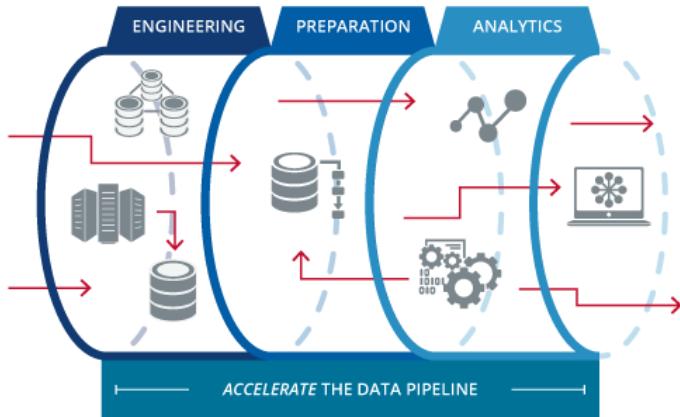
# → Intelligence (Learning & Inference)

# → Knowledge

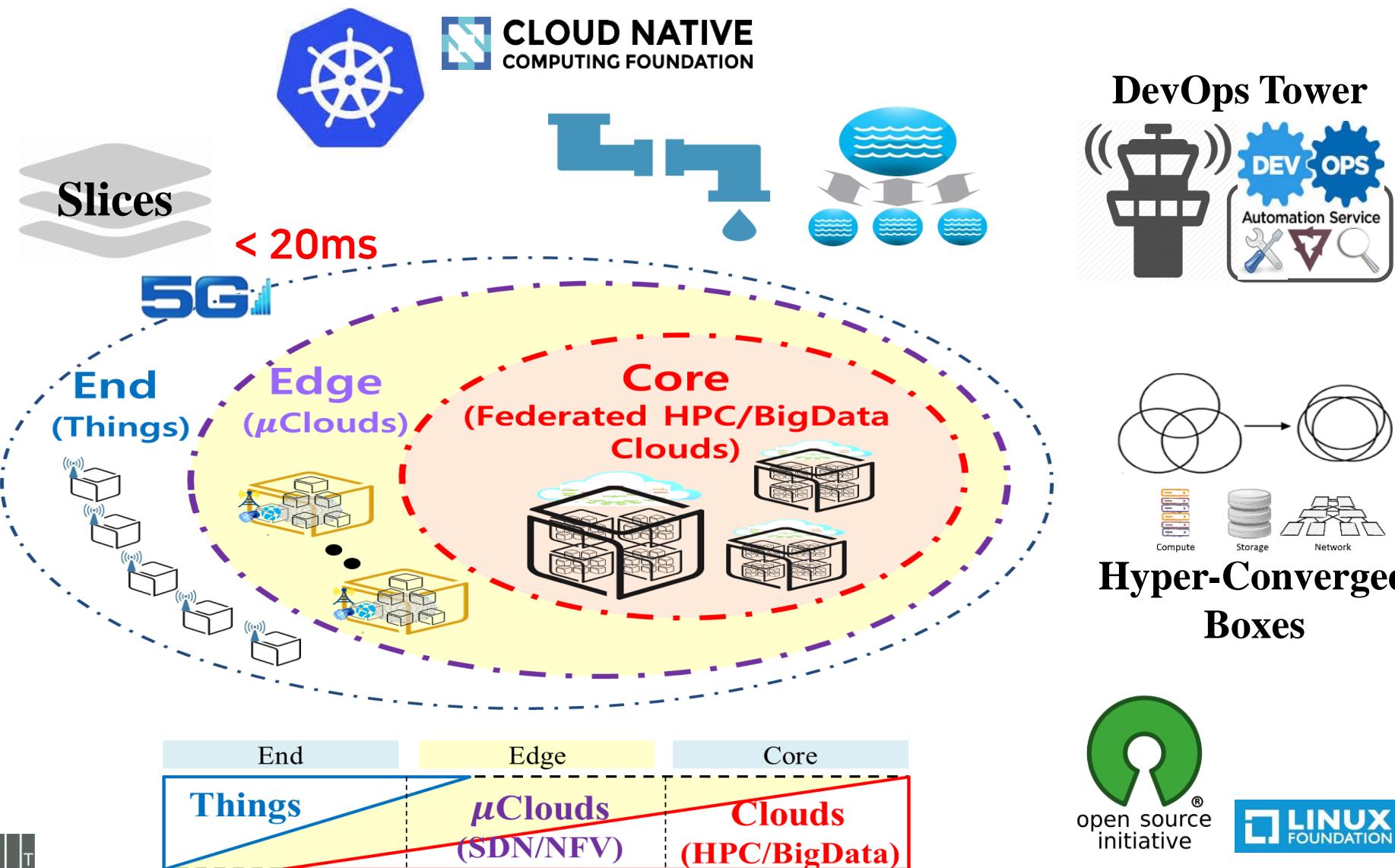
# → Information



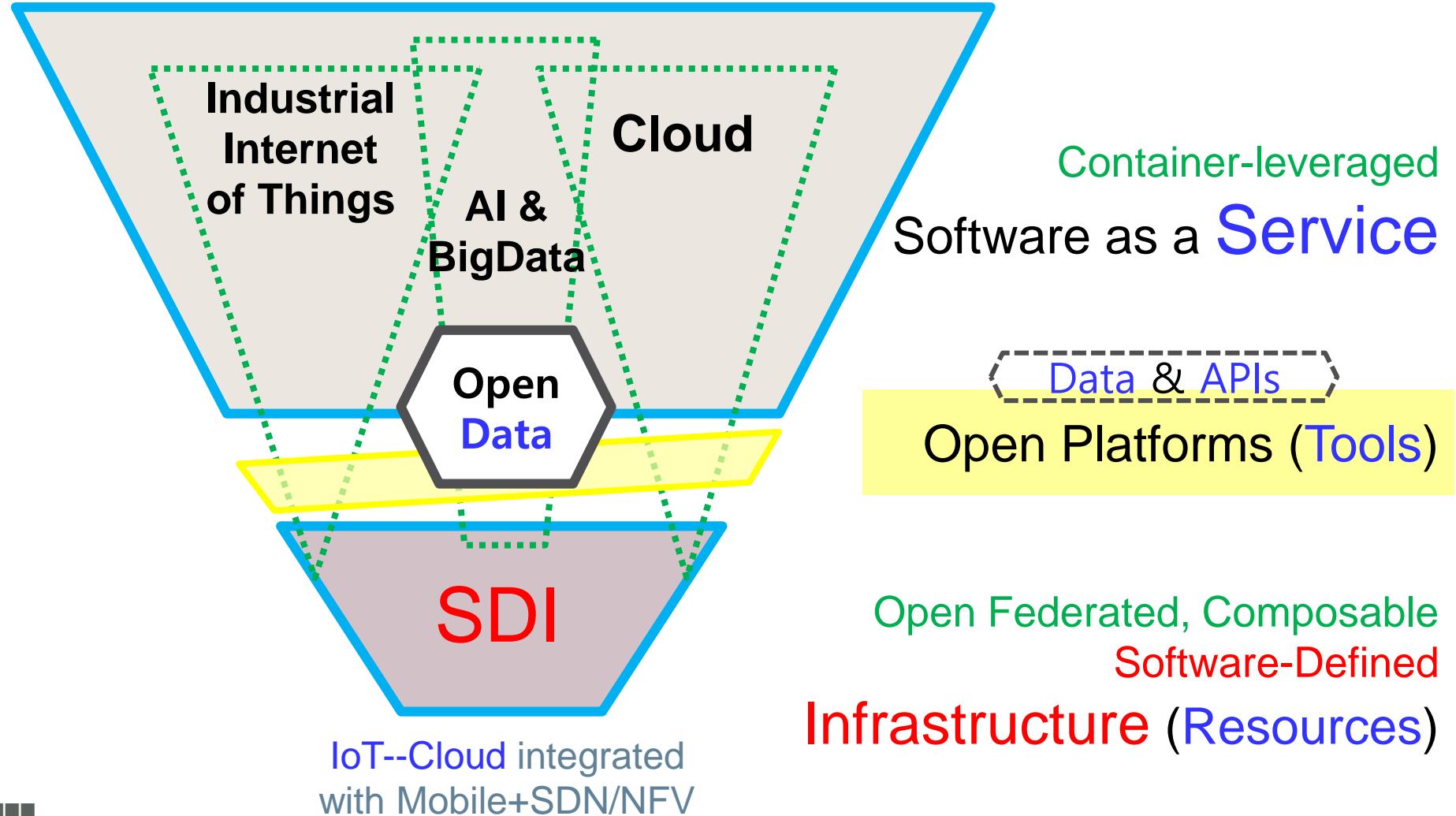
# Data-centric: Connected DataLake/Pipeline for AI



# Cloud-native & Data-centric 5G-/OpenSource-leveraged Converged SDI



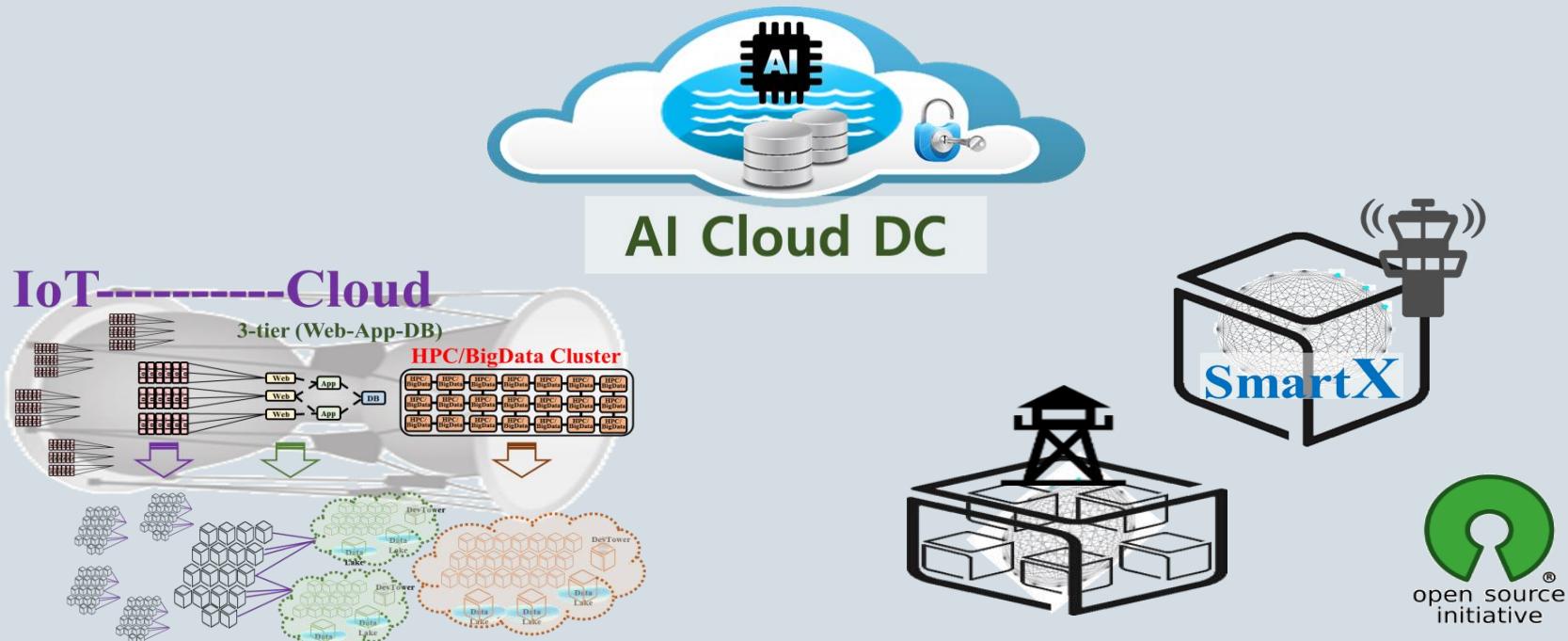
# Flipped Framework for Infrastructure – Platform – Services



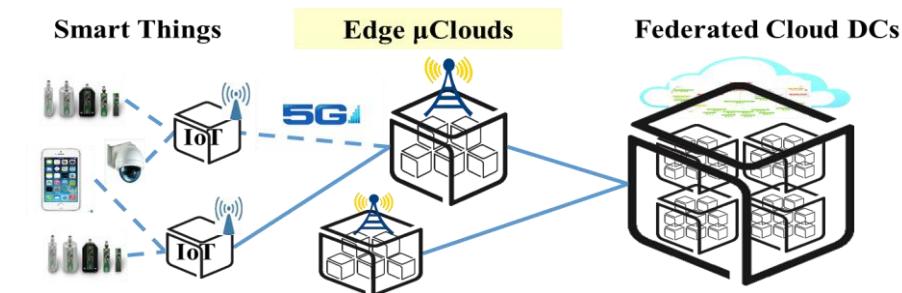
# SmartX (AI+X)

## Common Platform

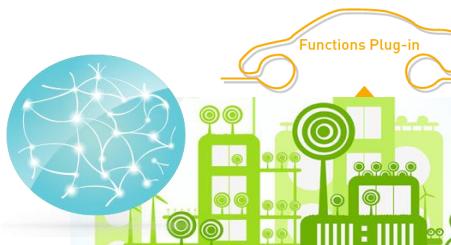
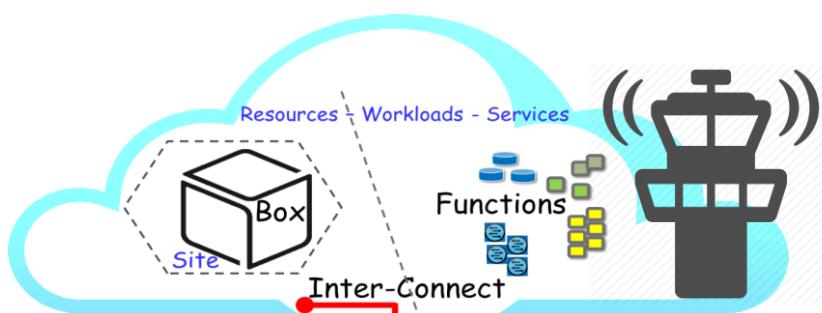
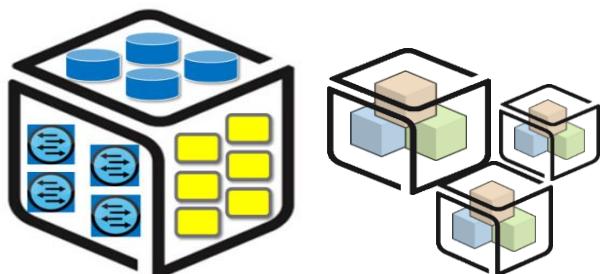
### for AI-inspired Services



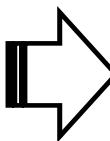
# Realize SmartX Services with Open APIs enabled by SmartX Boxes



**Software-Defined Infrastructure:**  
Build Programmable & Virtualized Resource Pools with Hyper-converged SmartX Boxes



**Open Platform:** Prepare **Open APIs** by Placing and Tuning Inter-connected Functions

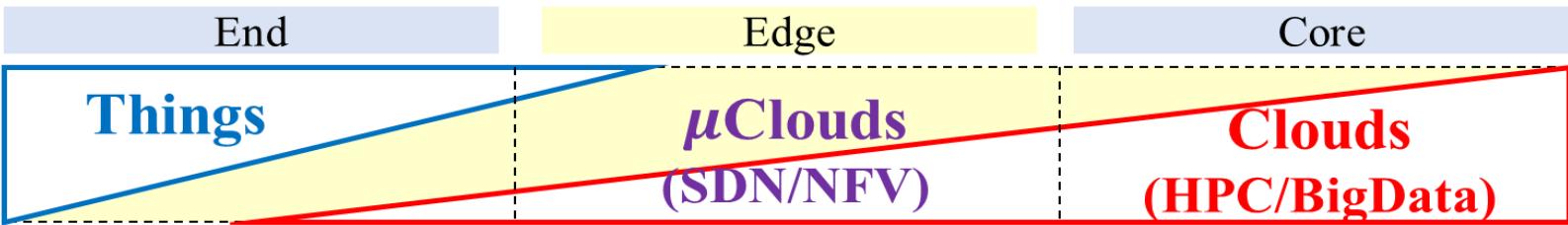
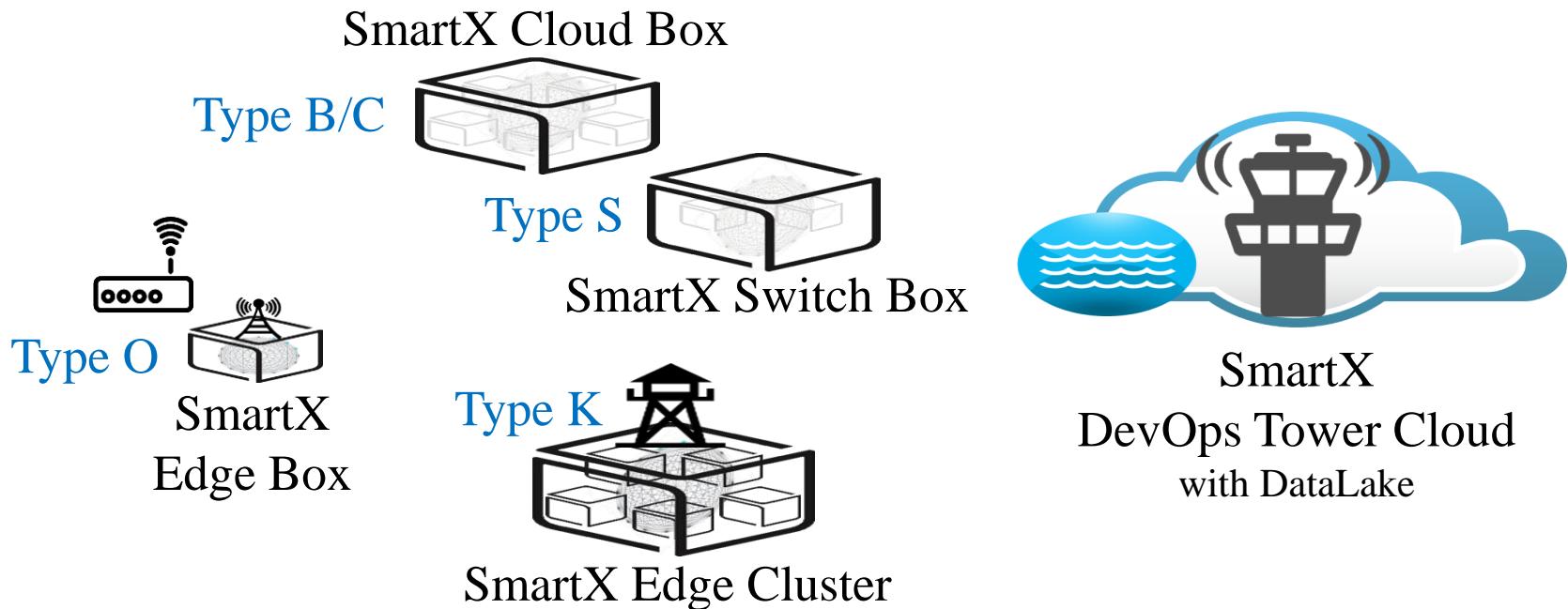
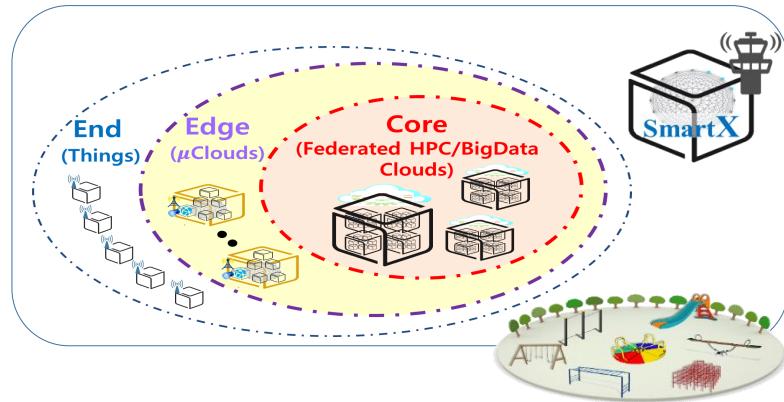


**User-Defined Services:** Create and Orchestrate Smart IoT-Cloud Services with Domain-specific **Open Data**



# SmartX Playgrounds

## (Infrastructure + Platform)



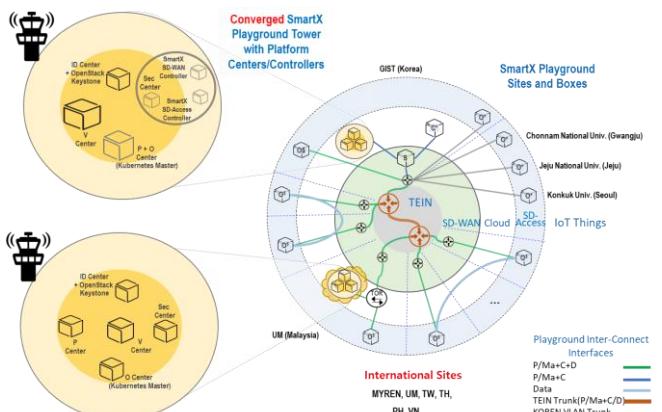
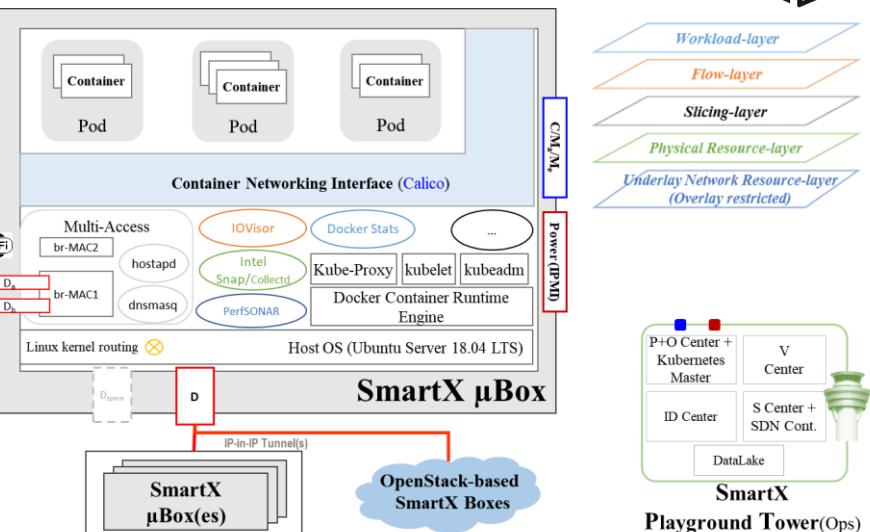
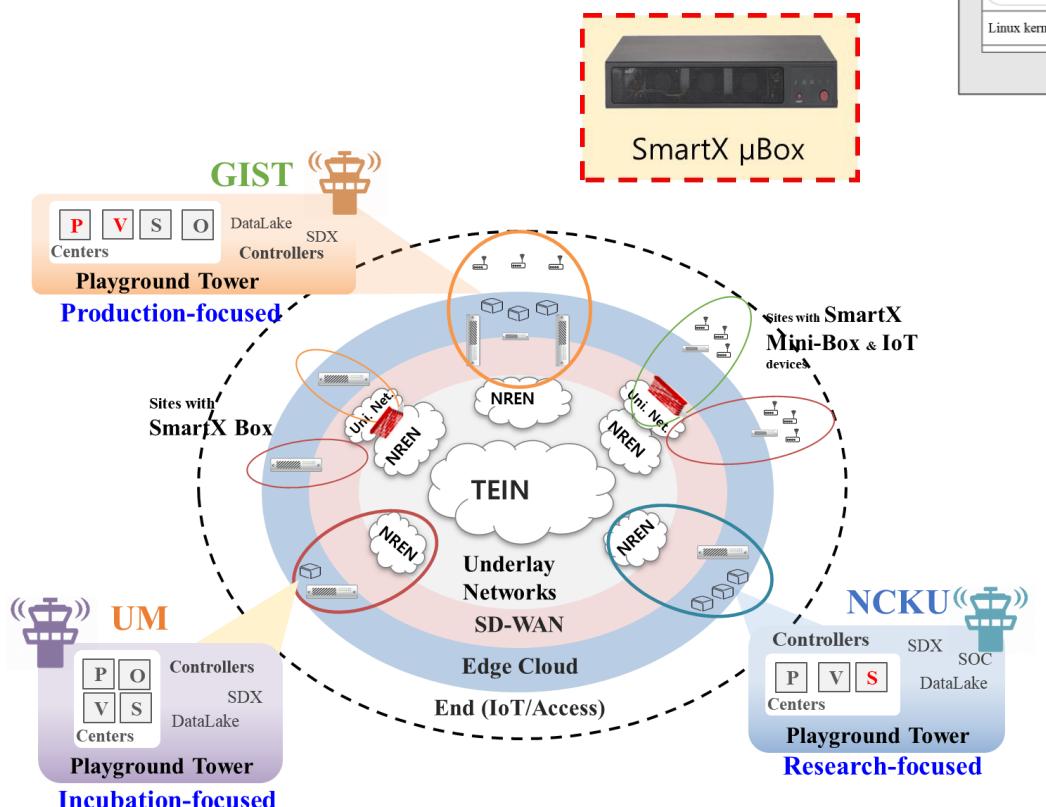
# Federated OF@TEIN+ Playground (2017~2019)



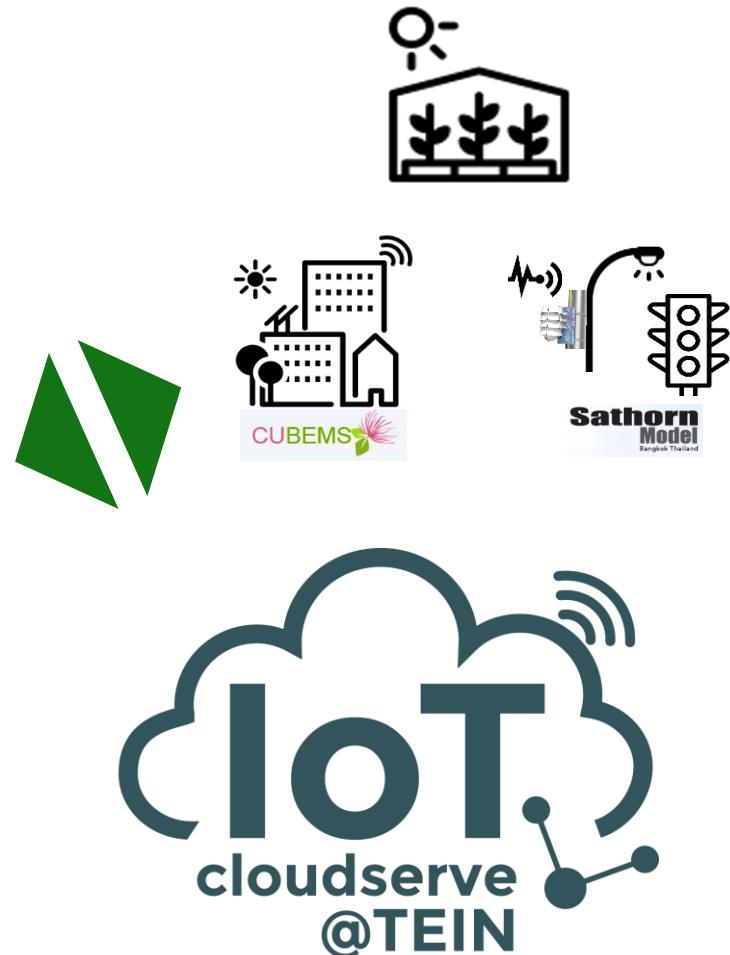
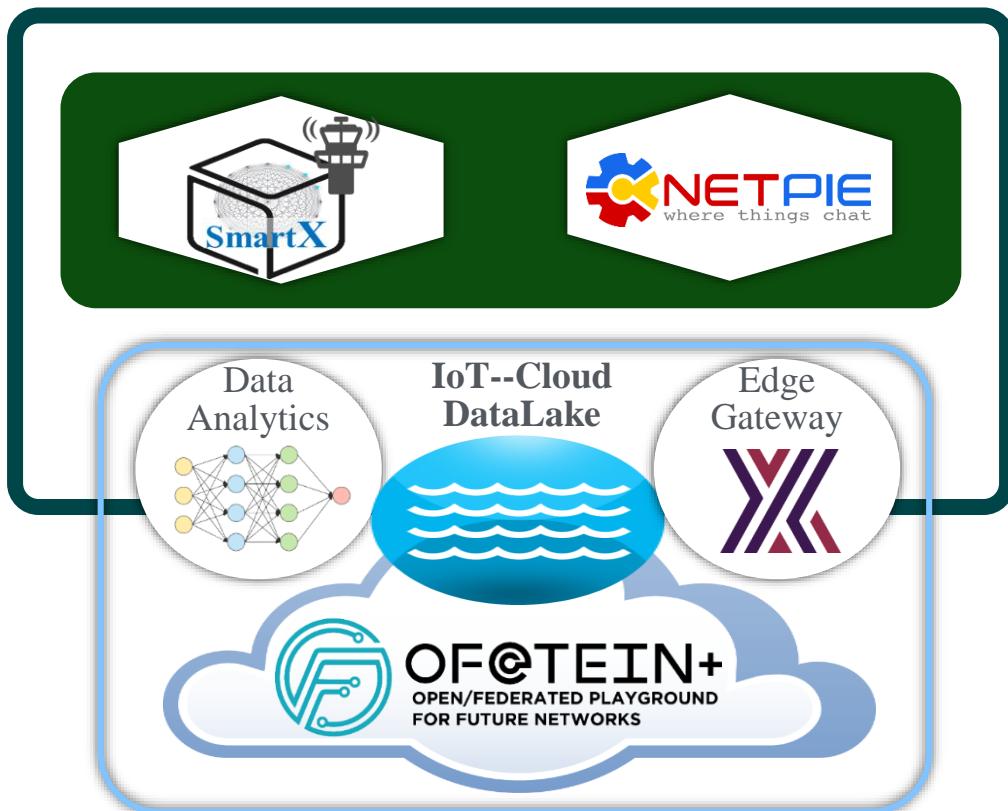
Extending OF@TEIN Playground (2012~)



**OF@TEIN+**  
OPEN/FEDERATED PLAYGROUND  
FOR FUTURE NETWORKS



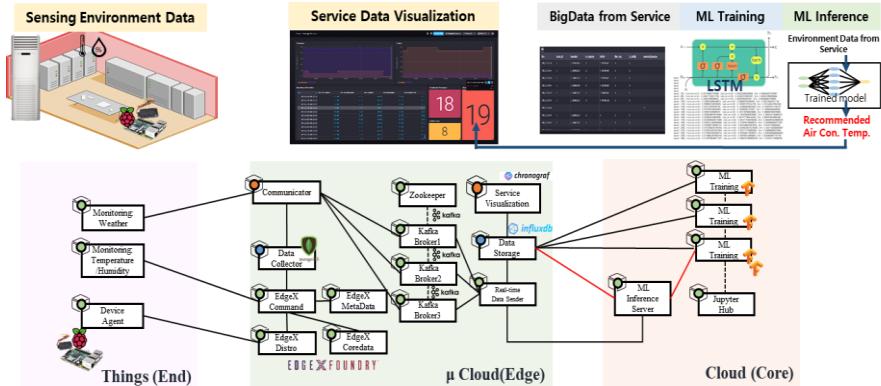
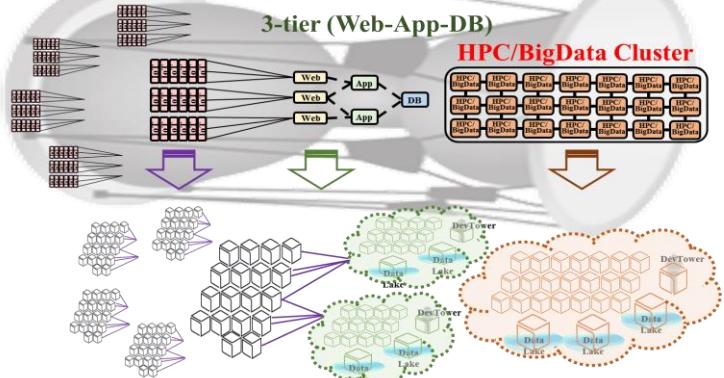
# IoTcloudServe@TEIN: Data-Centric IoT-Cloud Service Platform for Smart Communities (2018~2020)



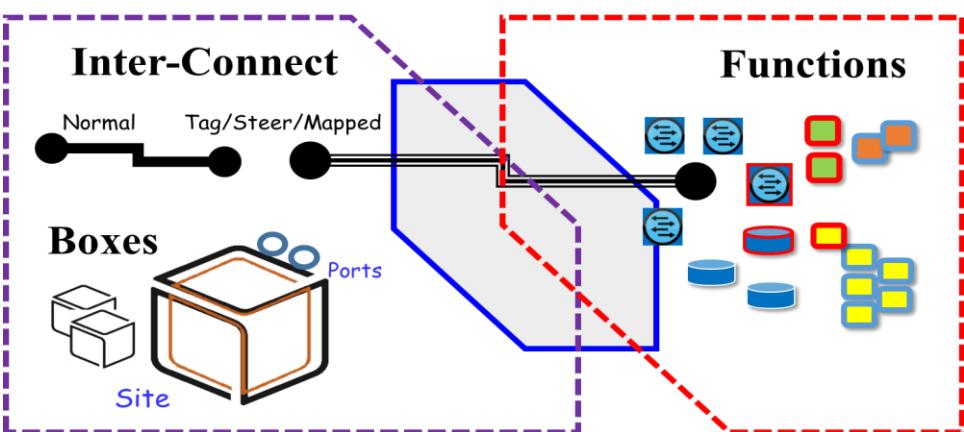


# SmartX IoT—Cloud Services with MSA (MicroServices Architecture) Applications (2016~2018)

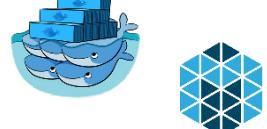
## IoT - Cloud



**p(Baremetal) + c(Container) + v(VM)**



Process containers  
N SHIFTEK



application containers  
S



Machine containers

(Monster containers)  
KVM

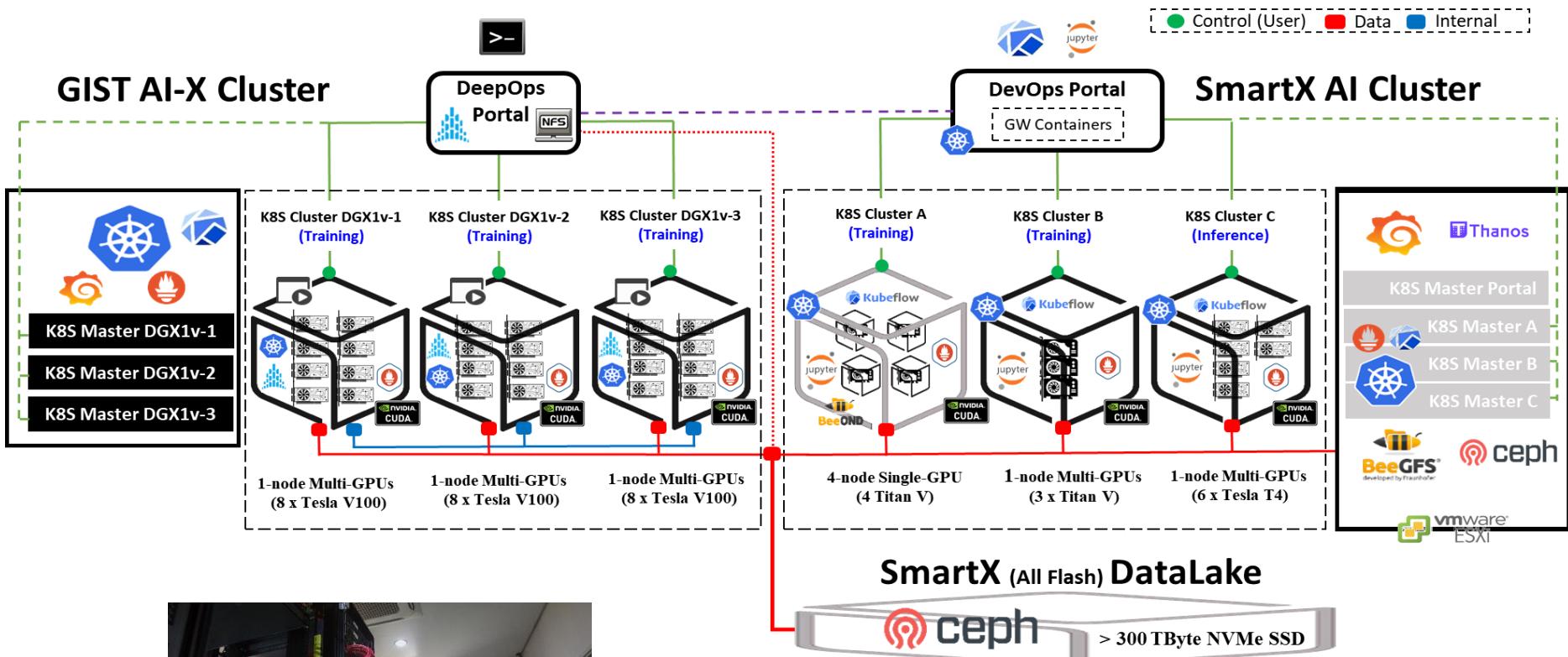


VM hypervisors



# Cloud-native AI-X & SmartX AI Clusters (2018~2020)

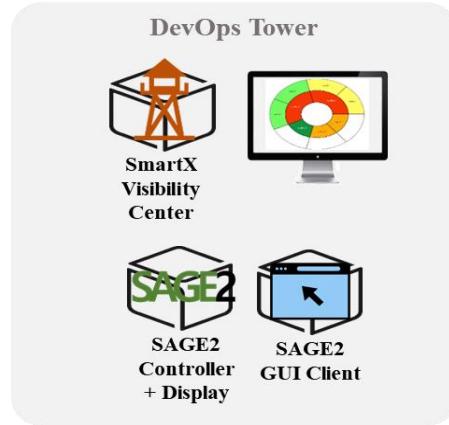
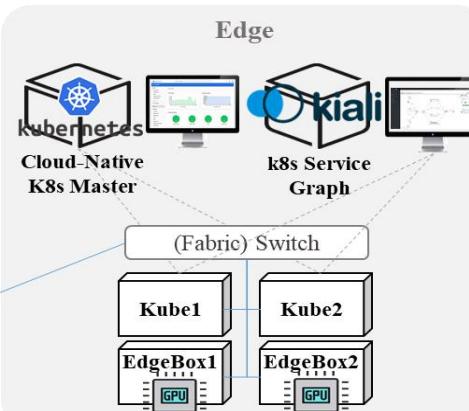
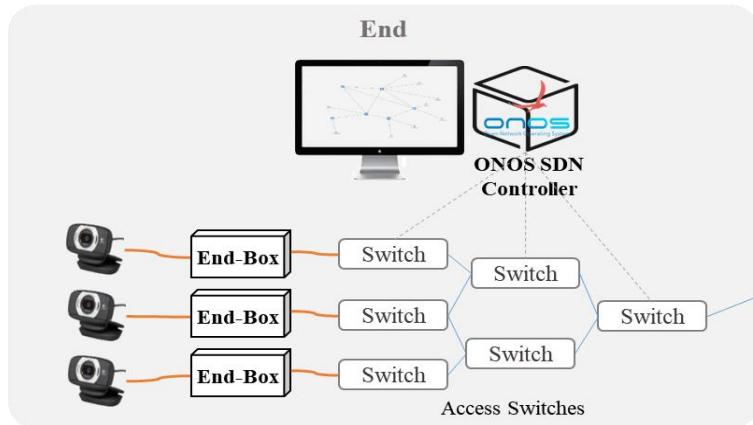
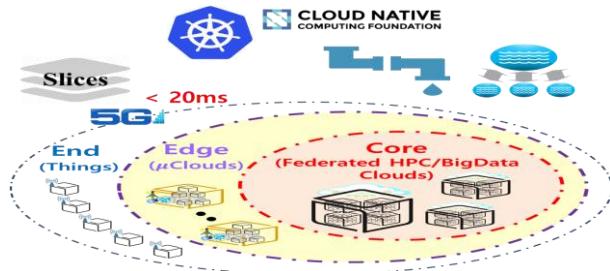
Domain-specific AI-inspired HPC/HPDA Workloads for Multiple Tenants



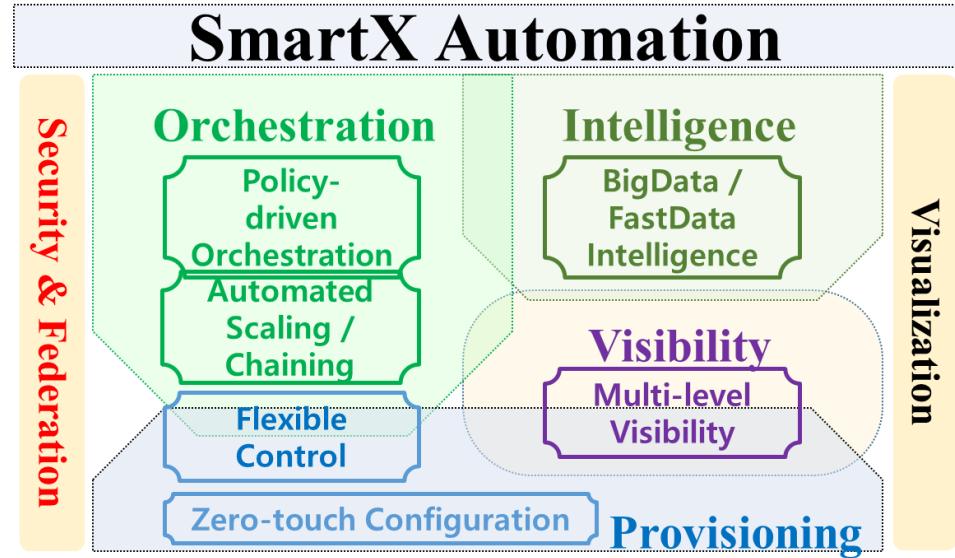
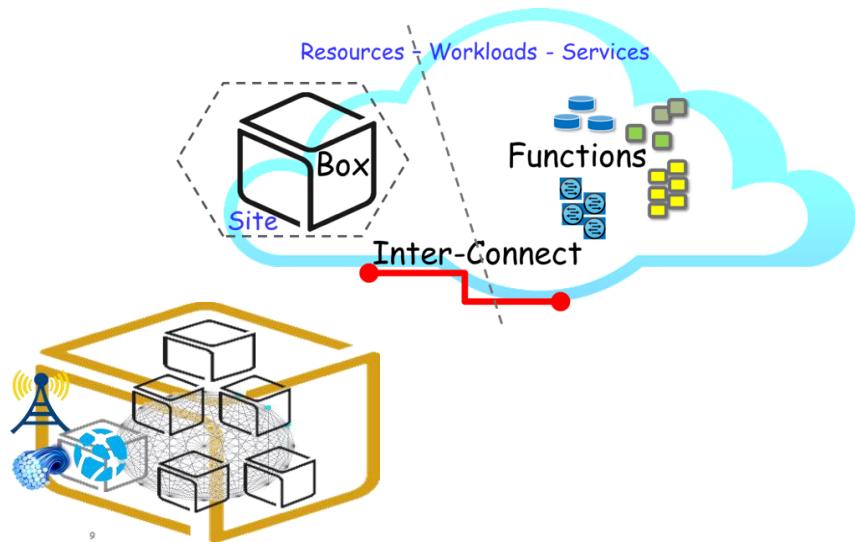
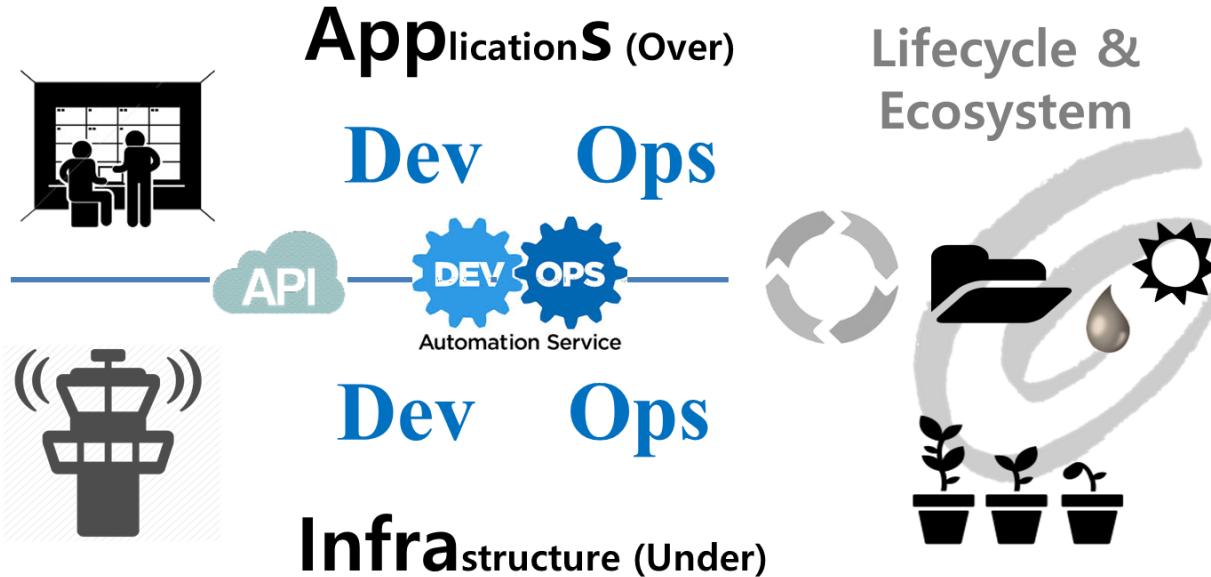
GIST | AI 대학원



# Cloud-native SmartX Playground: AI+X Playground (2020~)



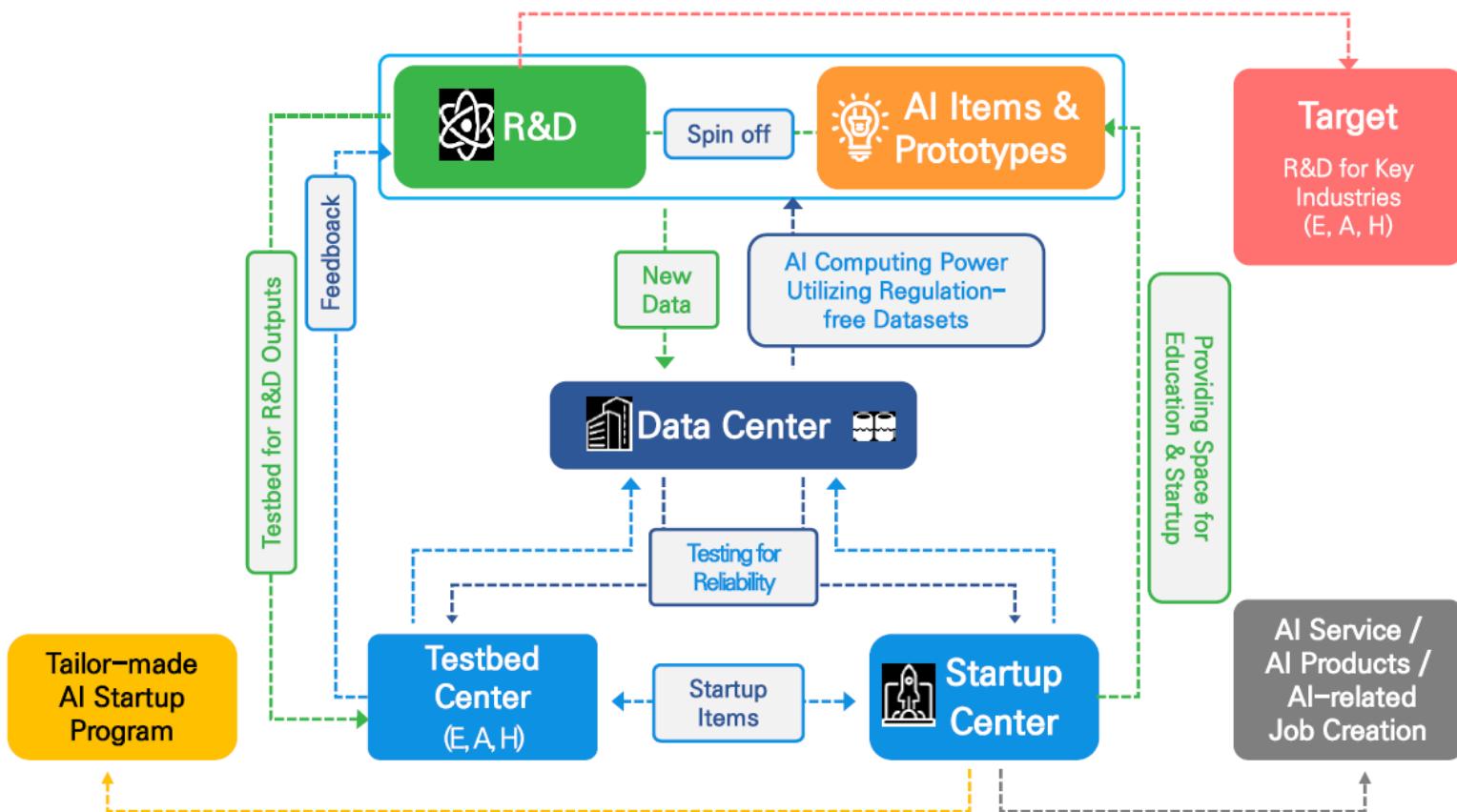
# SmartX Automation Framework



# Gwangju AI Complex for National AI Ecosystem

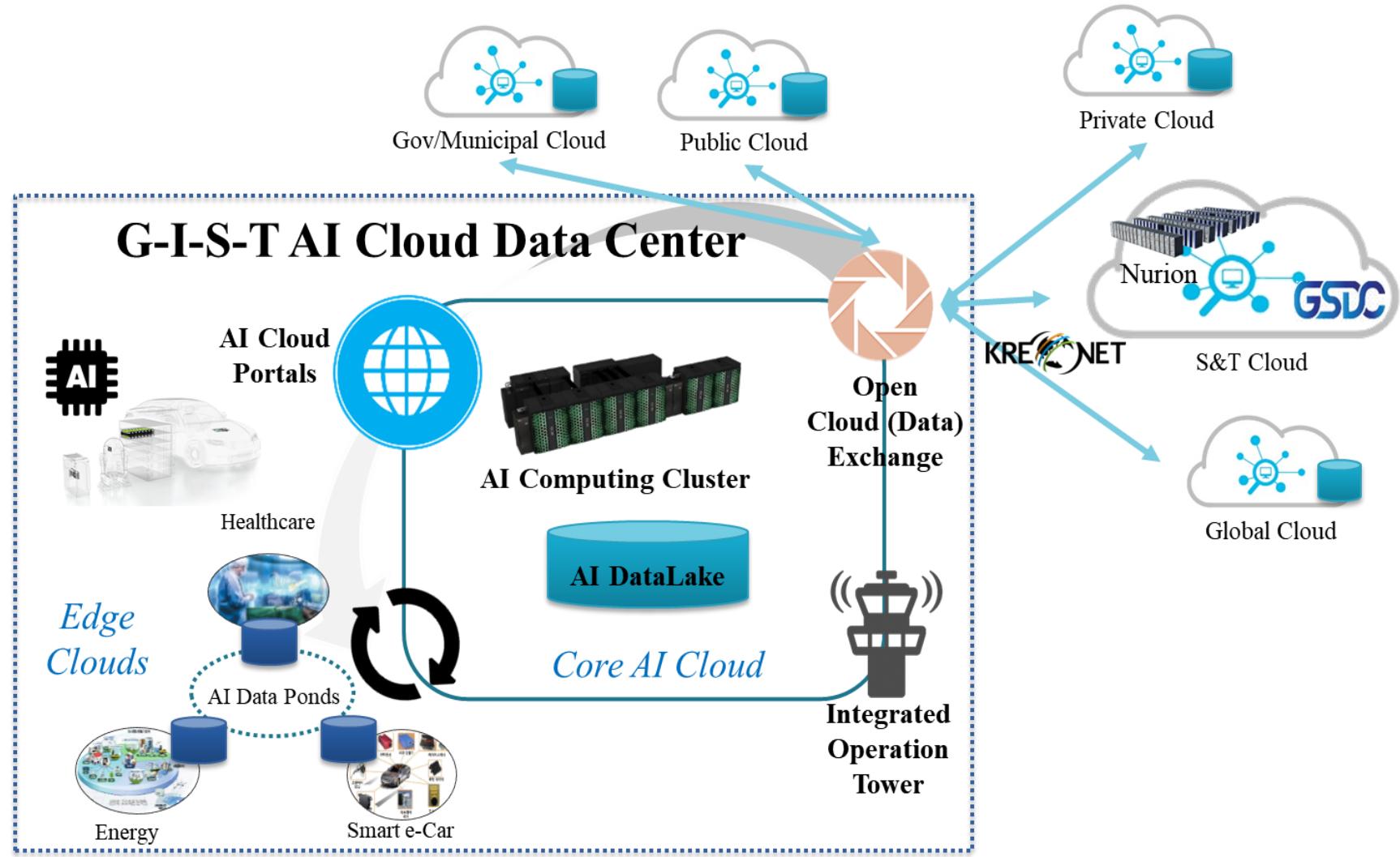


# Gwangju AI Complex for AI+X Specialization ('20~'24)

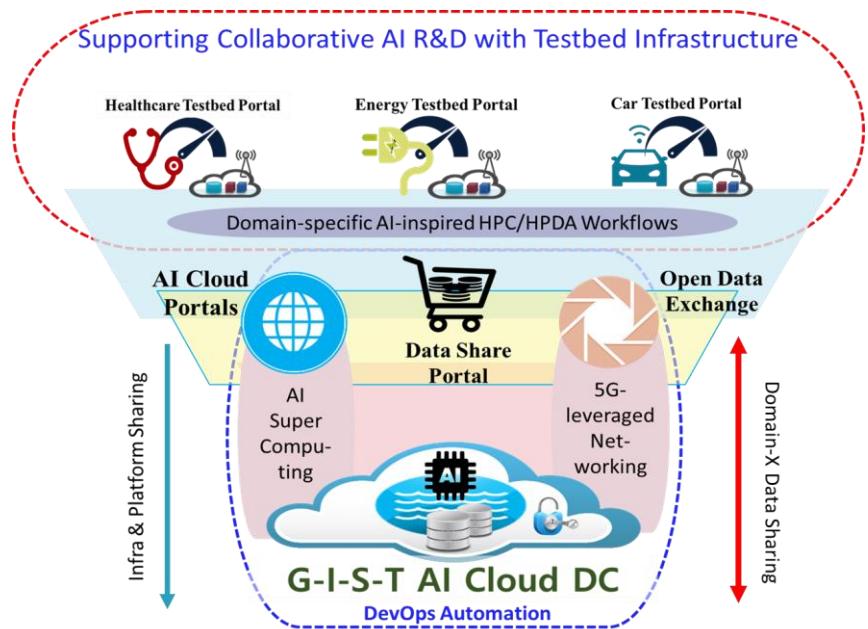
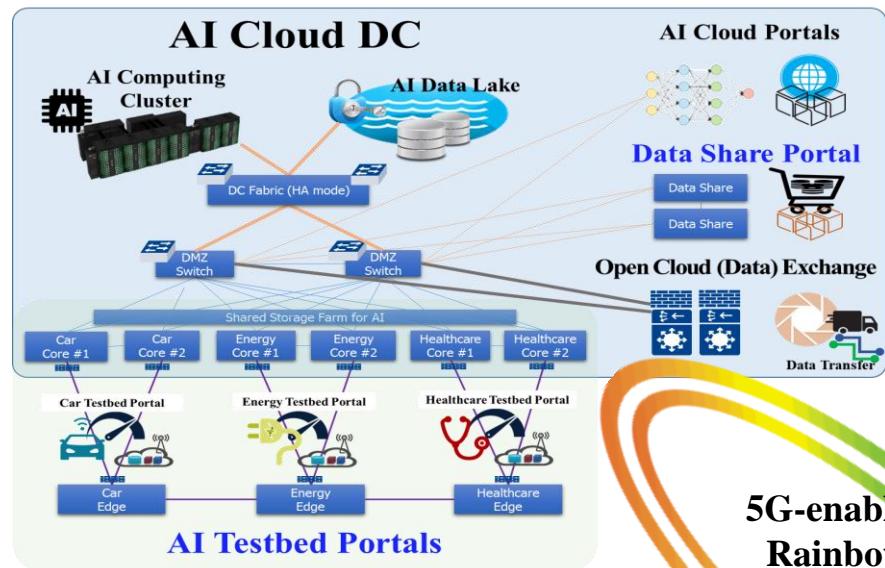


Cloud-leveraged/Data-centric

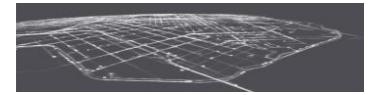
# G-I-S-T (Green / Intelligent / Sustained / Technical) AI Cloud Data Center



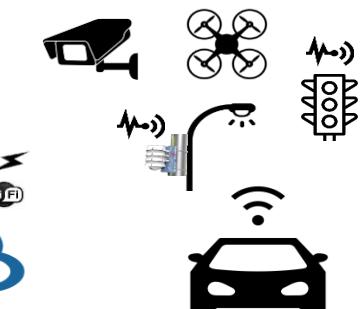
# Live! AI+X R&D over Gwangju AI Complex



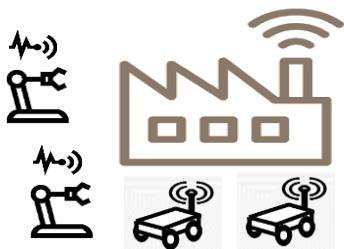
## AI Integrated Operation Tower



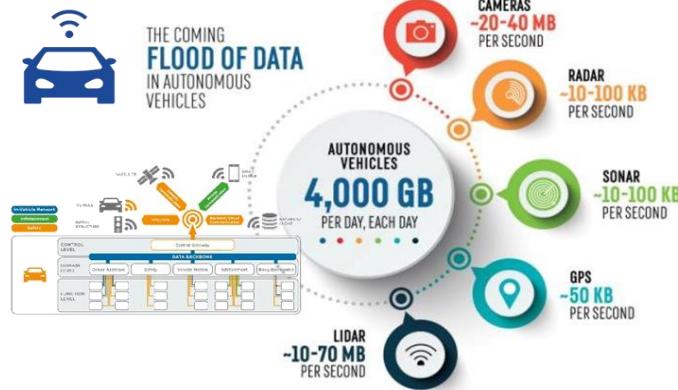
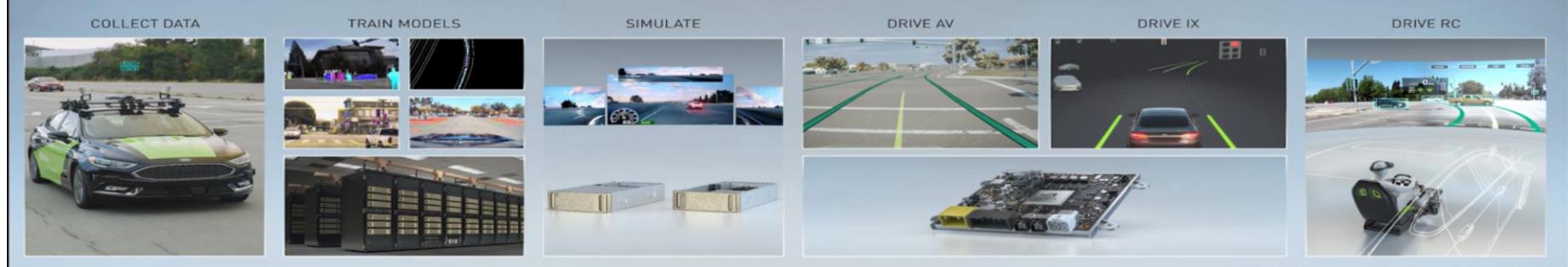
## Smart Mobility & Safety Demo Services



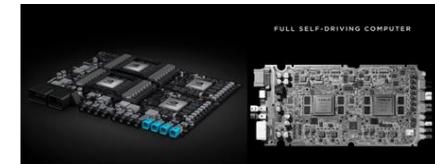
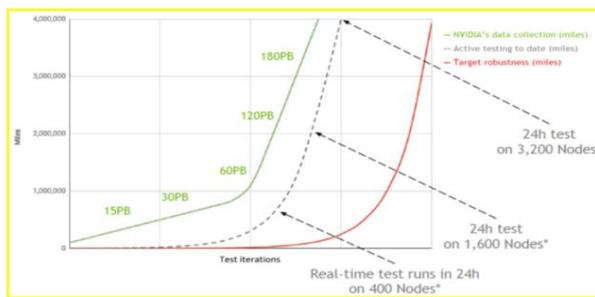
## Smart Factory Demo Services



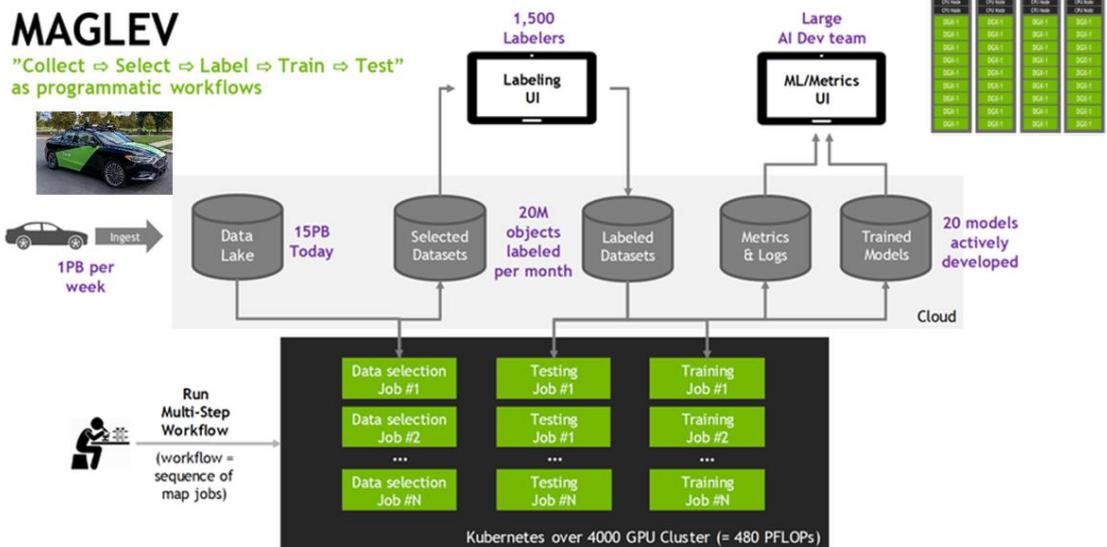
# On-Off-Mix Verification for Self-Driving Car



12-camera+Radar+Lidar RIG mounted on 30 cars	1,500 labelers	4,000 GPUs in cluster = 500 PFLOPs
> 1PB collected/month	20M objects labeled/mo	100 DRIVE Pegasus in cluster (Constellations)
15PB active training+test dataset	20 unique models 50 labeling tasks	1PB of in-rack object cache per 72 GPUs, 30PB provisioned



NVIDIA Drive AGX Pegasus vs Tesla HW3



# G-I-S-T Cloud-native AI Infrastructure & Platform

AI-inspired Domain Services Realization with DevSecOps-automated Workflows

Super  
Special  
Long-tail



Domain-based Edge Clouds with Testbed Service Portals



External Clouds



Cloud-leveraged OpenAPI Portals  
(ID Federated, Customizable, ...)

AI Service Software Frameworks for Data / Workload  
with Programming languages, models, environments, libraries, tools

Cloud-native K8S Orchestration  
with AI/HPC/HPDA Container/Job Scheduling

Protected Data Pipeline Tools with Connected DataLake

SD-Networking with Fabric



SD-Compute

Accelerators

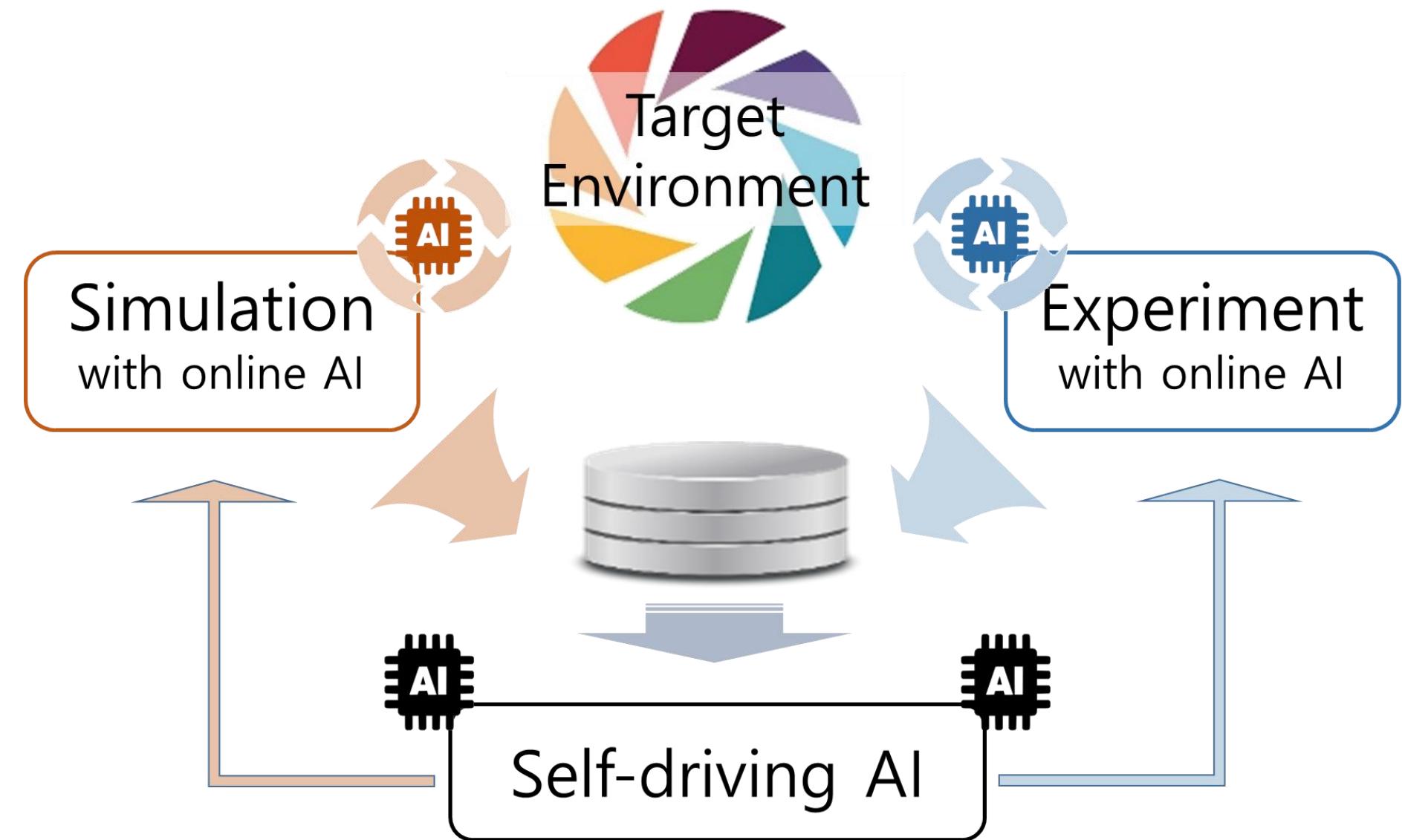
SD-Storage



Automated Infrastructure Operation  
(Composability, Visibility/Security, Accountability, ...)

AI Core Cloud

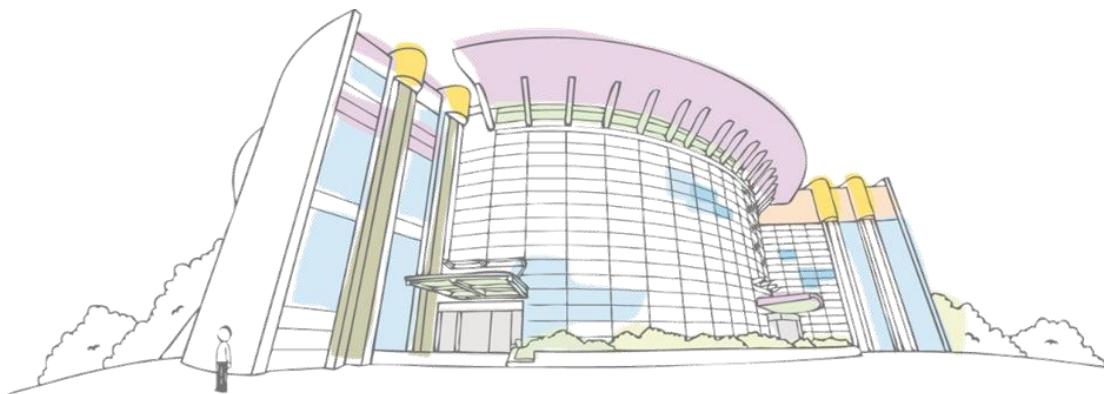
# On-Off-Mix AI+X R&D for Sci. & Tech.



# GIST | AI Graduate School



*Thank You!*



*AI@gist.ac.kr*