# Open IT Infrastructure to Drive Innovations at SKT

Kang-Won Lee, PhD

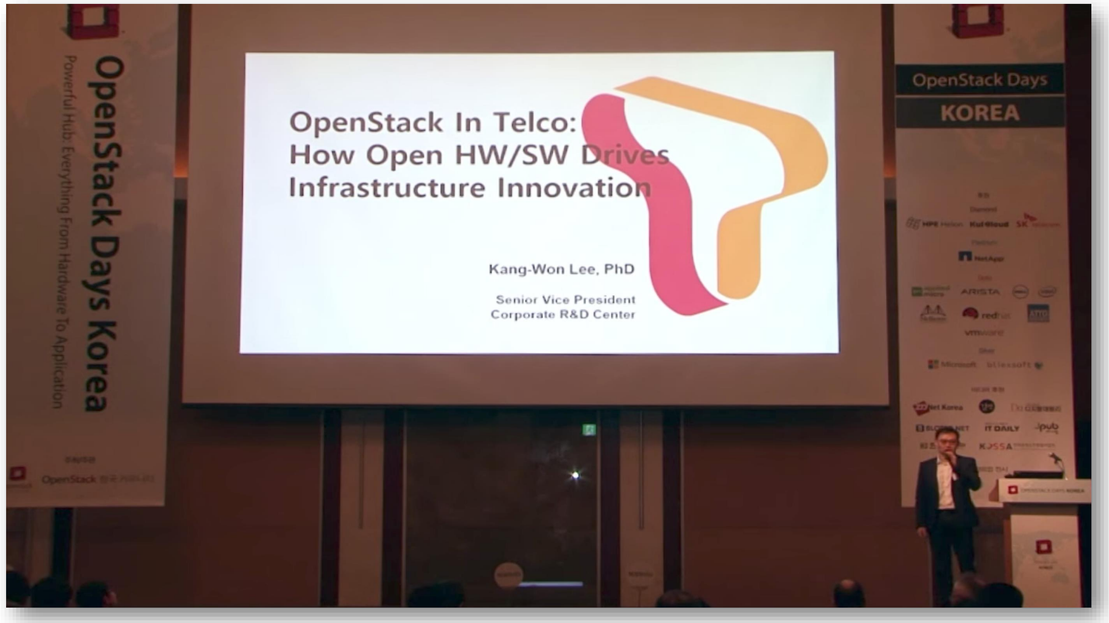
Senior Vice President
Corporate R&D Center

# SKT in Past OpenStack Days


**2015: All-IT Network Vision**

SDDC Vision & Architecture

SKT Commitment on OpenStack




**2016: Open HW/SW Development**

Open HW Development

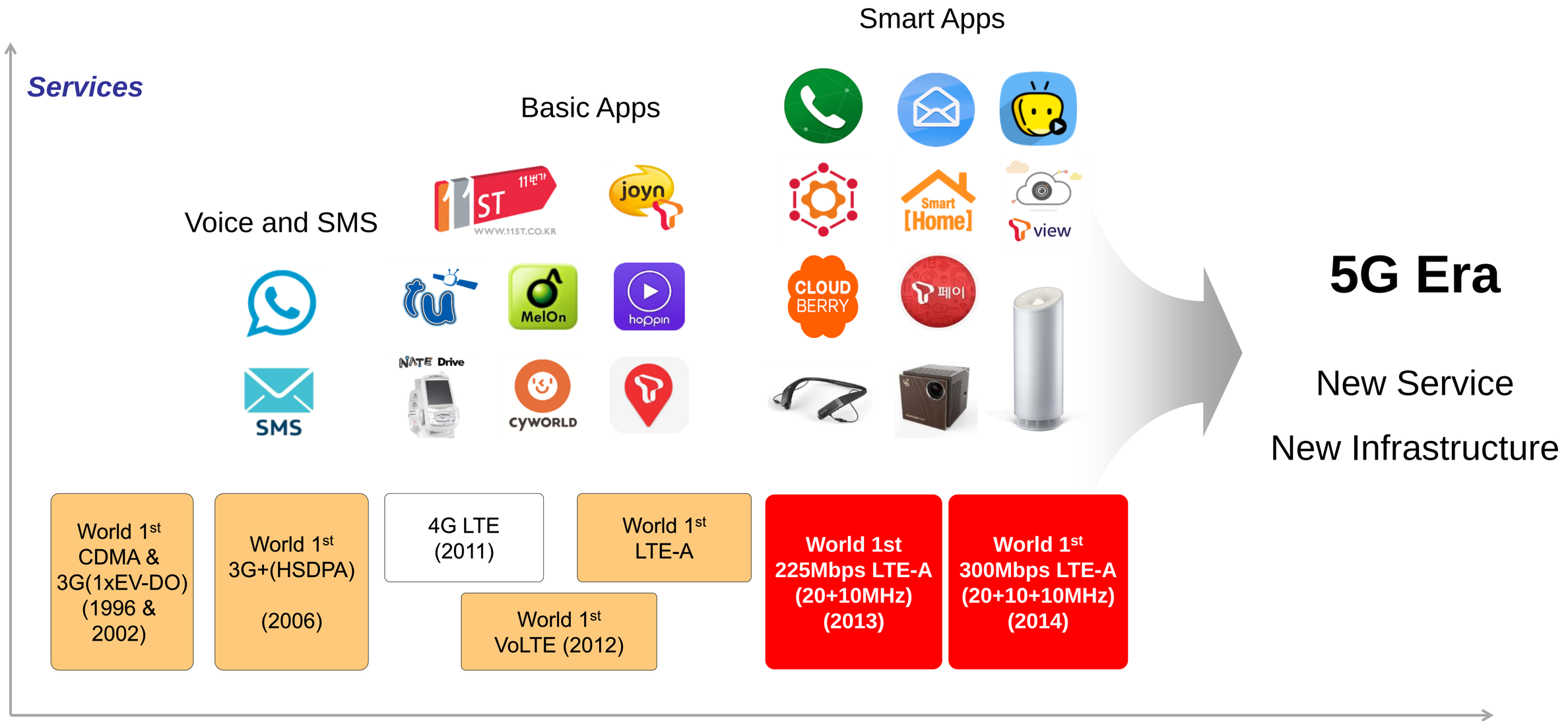
SDDC Operation/Analytics SW Development

**2017: OpenStack with Agility**

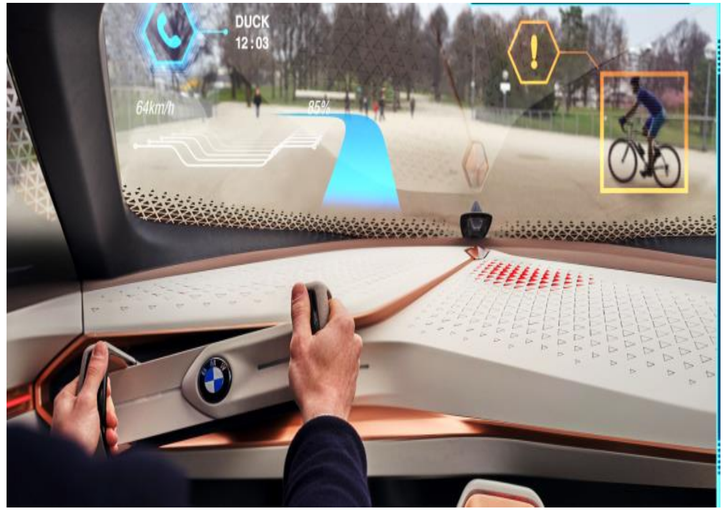
Containerized OpenStack Deploy and Manage

Open Infrastructure for AI

# SKT over years

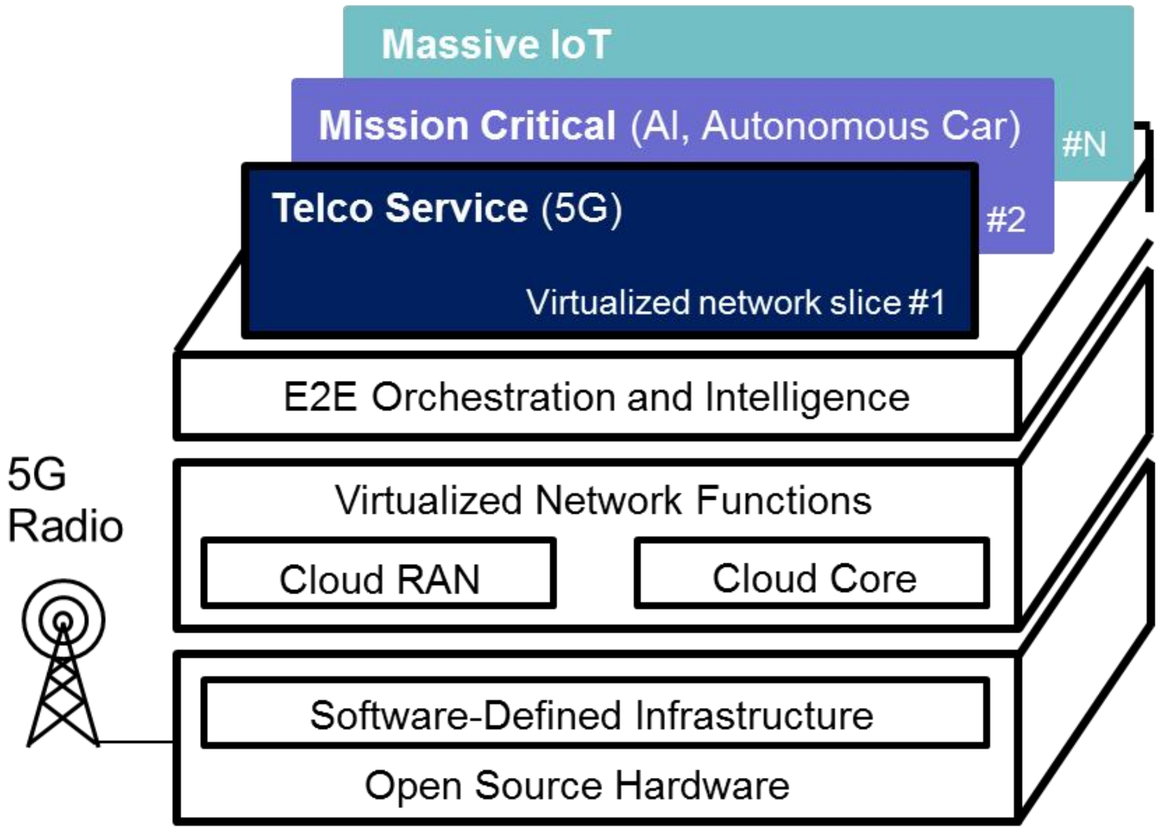


# Innovative Services of 5G



# 5G Evolution to Open Infrastructure




## SKT Services

- Ultra High Data Rate
- Mission Critical Service
- Massive Connectivity

## ATSCALE

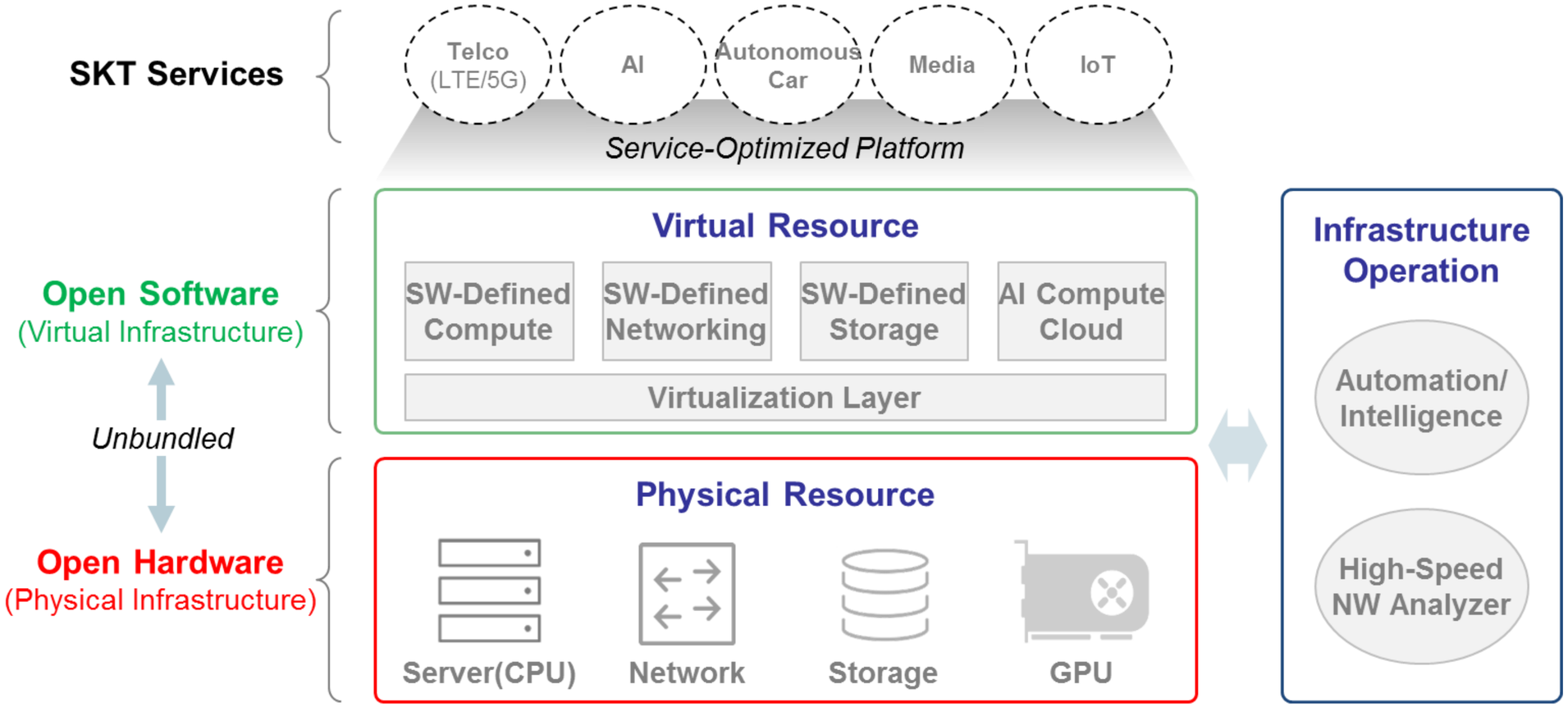
- Virtualized Network Functions
- Network & Service Slicing
- Next-Generation OSS (TANGO)

## COSMOS

- Software-Defined Infrastructure
- Open Hardware and Software
- Telco & Mission Critical Services

# COSMOS Vision

- **Composable, Open, Scalable with Open Software and Hardware**
- **Mission Critical Services (5G, AI, Autonomous Car, etc.)**

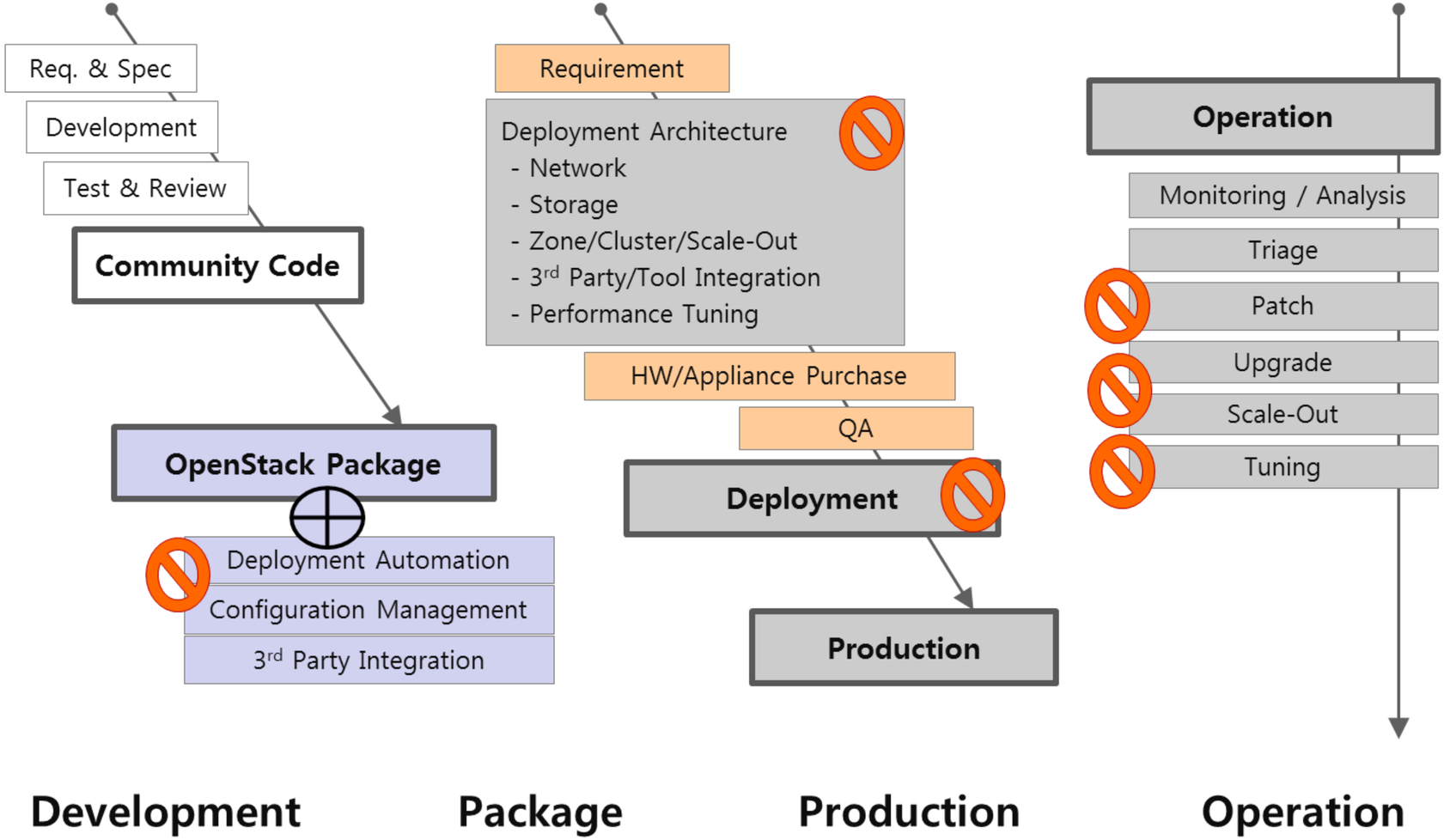


# Open Infrastructure for Virtual Resources

- **OpenStack for Virtual Infrastructure Management**
- **Telco (5G), Media, IoT Applications**

# Challenges

- OpenStack is still very **complicated system to deploy and manage**
- Current way of **automating** OpenStack still has **challenges**



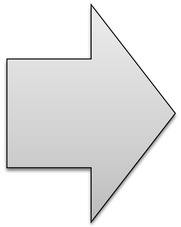
# TACO (SKT All Container OpenStack)

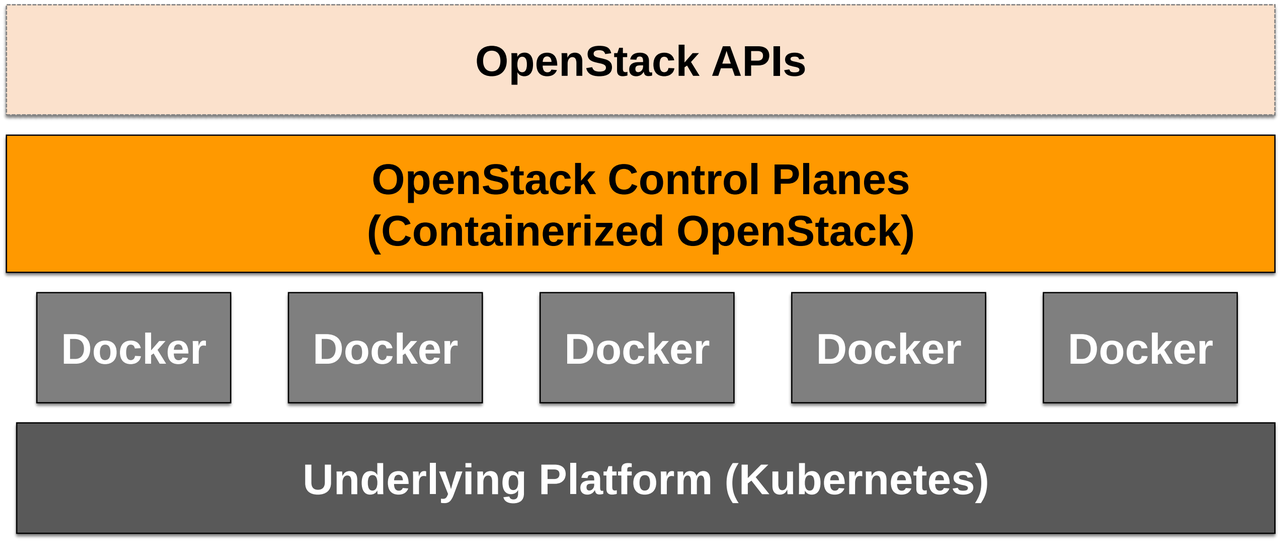
- Developed by **SK Telecom**, leveraging **Container and Kubernetes**
- **Community Version** with **Continuous Integration / Delivery** System
- Enhanced OpenStack Lifecycle Management: **Self-Healing, Upgrade w/o Service Interruption, Simple and Easy Deployment, Highly Flexible Customization**

**Container Orchestration**
(Kubernetes, OpenStack-Helm)

**OpenStack Containerization**
(Docker, Kolla)

**CI/CD & Testing**
(Jenkins, Rally/Tempest, Chaos Monkey)

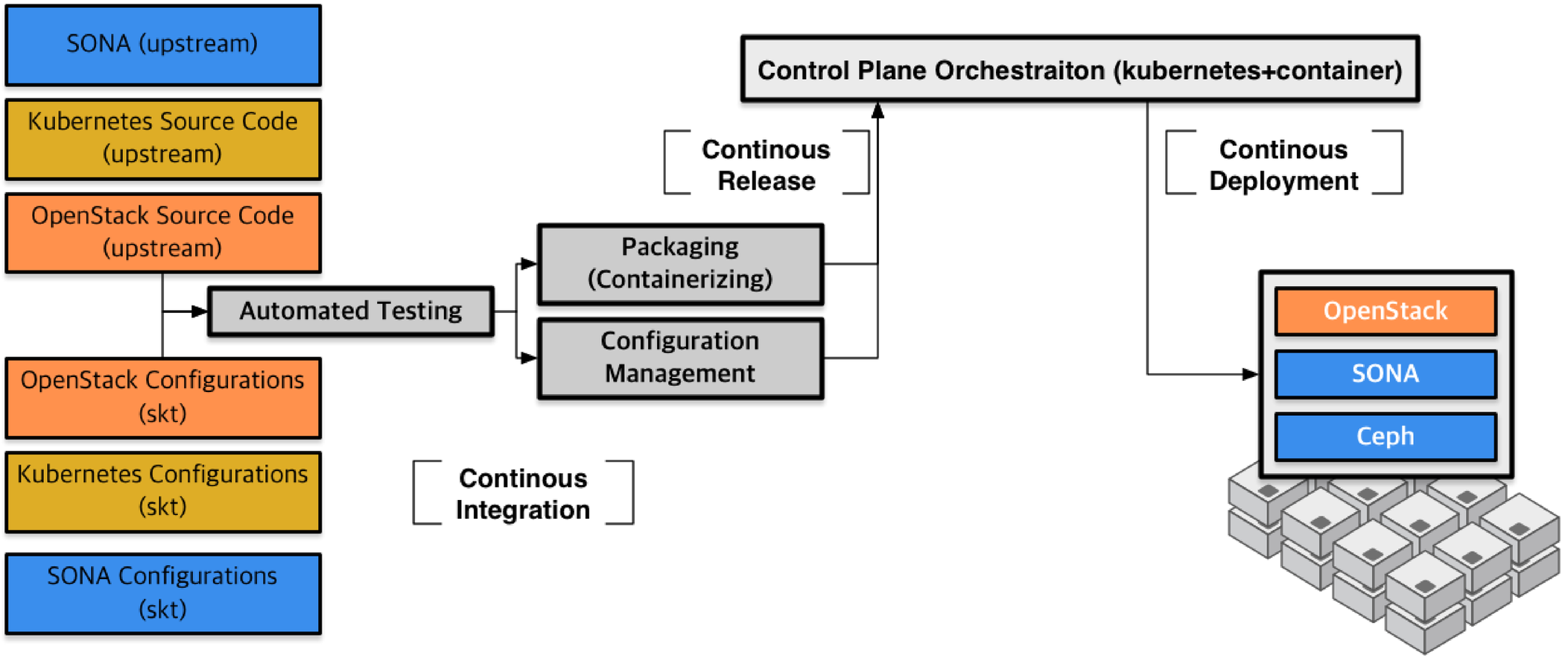





**Containerized OpenStack Lifecycle Management**

# TACO SW Delivery

- **Automated Continuous Integration** Pipeline w/ Various Tests (100% sync to Upstream Code)
- Standardized **Packaging, Versioning, Release** Process and Tool Sets



T All Container OpenStack

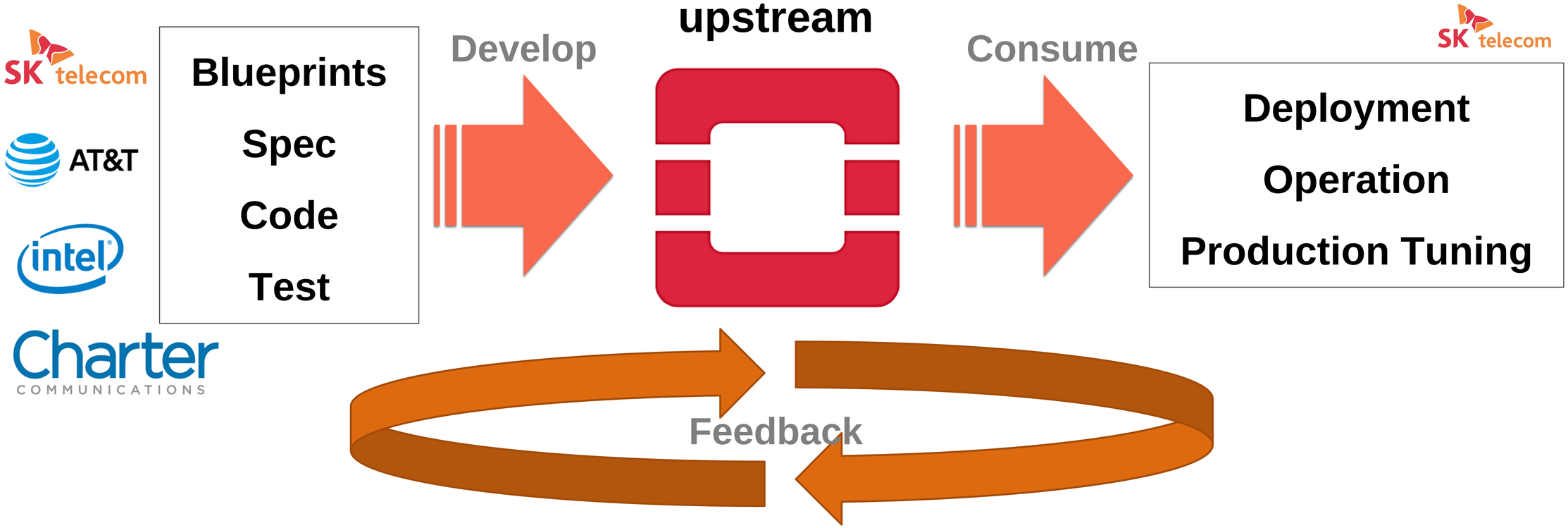
SK telecom

# OpenStack on Kubernetes

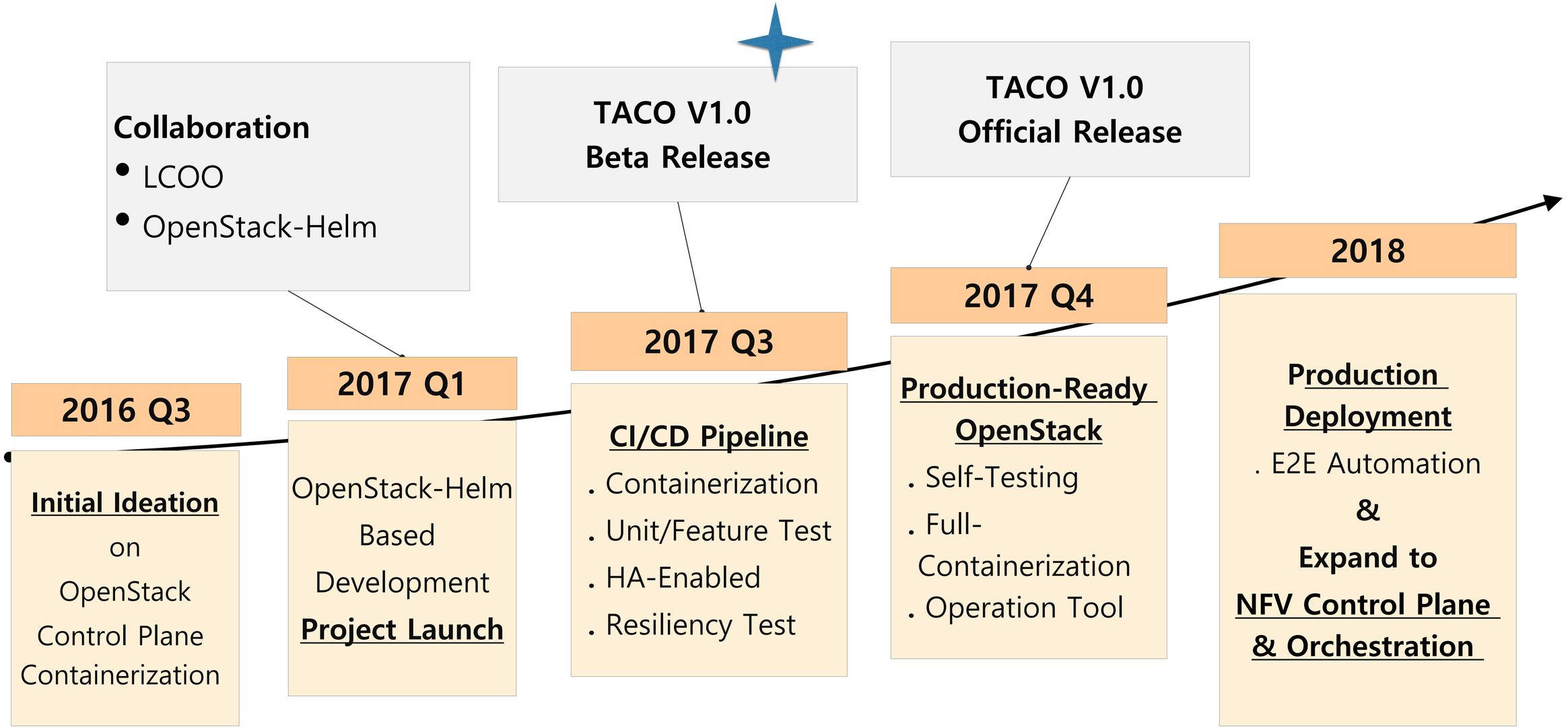
SK Telecom
Open System Lab

# Upstream First Philosophy

- **Upstream First**: Develop on upstream; Consume directly from the upstream
- **Benefits:** Zero Silo Code, Strong ecosystem (Your code are used everywhere), Efficient Development Effort (Loosely Coupled Co-Development with Various Community Partners)



# TACO Roadmap



# Open Infrastructure for AI

- **Cloud Infrastructure of AI Training**
- **Large GPU Cluster for Diverse AI Applications**

# AI Infra R&D Overview

## Deep Learning Infrastructure with Increased Performance and Utilization for SKT's AI Services

**SKT AI Services**

NUGU

T-map

T-view

Autonomous Driving

⋮

**1 AI Compute Cloud**

- Cloudification of Large-Scale GPU Cluster
- GPU Virtualization by Container
- Dynamic Job Scheduler for Better Utilization

**2 AI Compute Appliance**

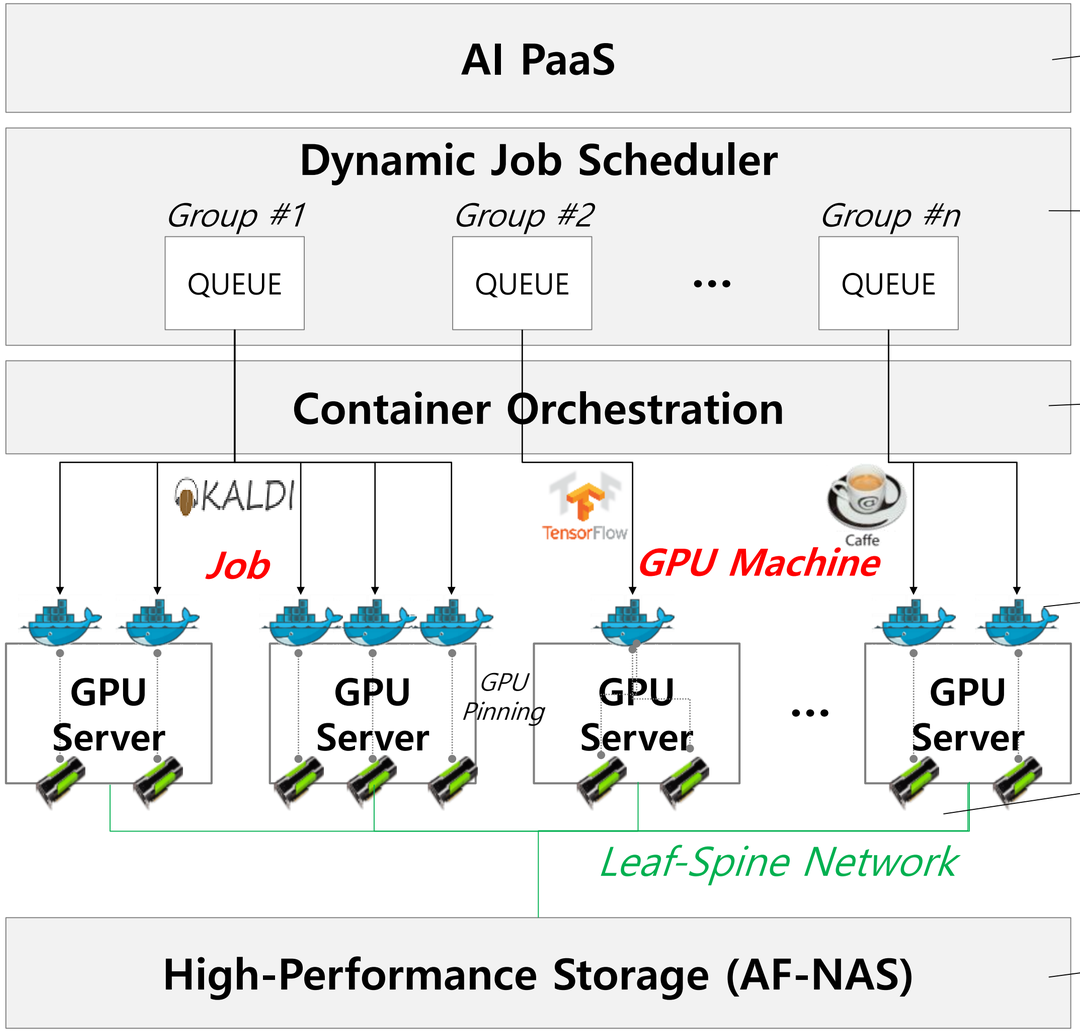
- High-Density GPU-based HPC Appliance
- Deep Learning SW Framework Optimization
- Rack-Scale Design by Dynamic Configuration of CPU-GPU-Storage

**3 AI Inference Accelerator**

- FPGA Optimized for Diverse ML Framework
- Lower Cost, Higher Performance, Lower Power Consumption than CPU/GPU

# AI Infra R&D: AI Compute Cloud

## Highly Efficient Multi-User Multi-Node GPU Cloud Using Container & Job Scheduler




- Job Creation for Data Parallelism
- Dynamic GPU Resource Allocation
- Container Life Cycle Management
- GPU Containerization for Diverse ML Frameworks
- SONA Fabric Network
- All-Flash NAS Storage for Large Model and Data

# AI Infra R&D: AI Compute Appliance

## HPC Appliance with Optimized Deep Learning S/W on High-Density GPU Server

| | | | |
|---|---|---|---|
| **BI Tool (GUI)** | Jupyter | etc. | |
| **Job Scheduler** | AI Compute Cloud | | |
| **Deep Learning Framework** | Tensorflow | Kaldi | Caffe |
| **S/W Accel. Module** | Comp. Graph Optimization (JIT Compile) | | |
| **Library** | CuDNN | NCCL | NV Docker |
| **Accelerator Direct I/O Layer** | NVMe-GPU-XeonPhi-FPGA Direct | | |
| **H/W Accelerator** | GPU, XeonPhi | FPGA, NVMe SSD | |
| **Compute H/W** |  GPU Server |  |  JBOG |

**BI Tool for High Usability**

**Highly Efficient PaaS Platform**

- Cluster, Dynamic Scheduling
- Support for Diverse ML Framework

**Computation & Memory Optimization**

**H/W Acceleration Methods**

| Accelerator | Target |
|---|---|
| XeonPhi | Seq. Processing, CODEC |
| FPGA | Reduce Operation |
| NVMe SSD | Near Data Processing, ISP |

**Direct Communication b/w Accelerators**

- NVMe-GPU Direct I/O

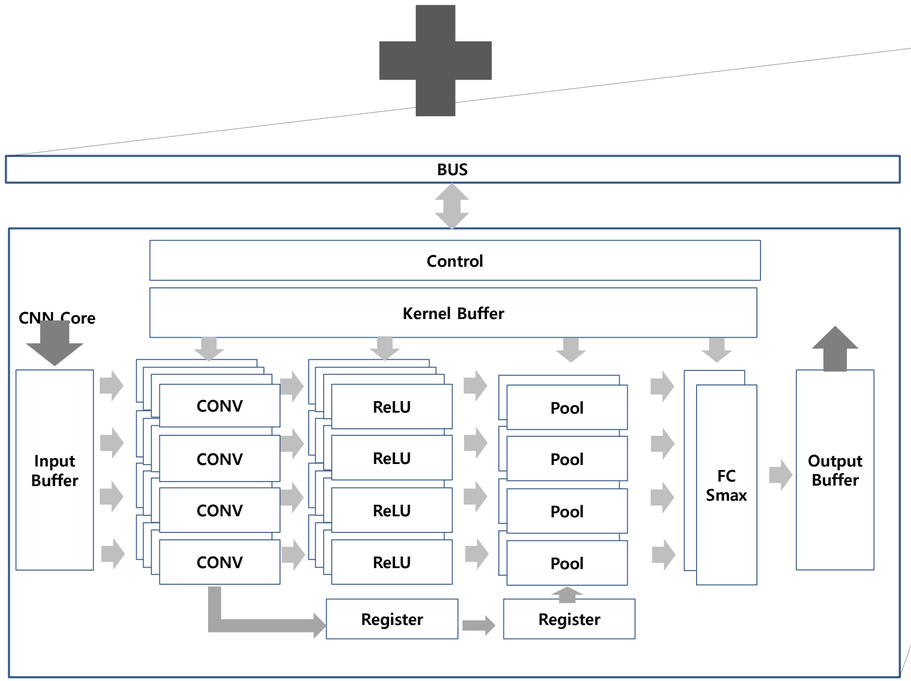
**ISP(In-Storage Processing)**

- Deep Learning Acceleration in NVMe SSD Core

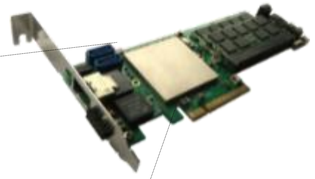
# AI Infra R&D: AI Inference Accelerator

## Inference Acceleration to Improve Performance and Power Efficiency

**Deep Learning Framework (S/W)**




**Deep Learning Inference Core (H/W)**



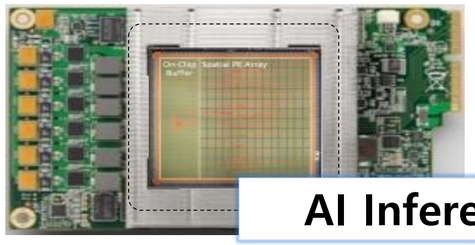
**FPGA Accelerator**

**PCIe Slot**

PCIe Card Based Server
in AI Service Cloud

**Xeon+FPGA**

Xeon+FPGA Server
for AI Service Cloud



**AI Inference ASIC**

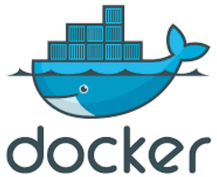
# Let’s Journey Together

SK telecom

## Community Collaboration


- **OpenStack-Helm Project**
- **Large Contributing OpenStack Operator WG (LCOO)**


- **Kubernetes**
- **Helm**
- **Prometheus**





- **Docker**
- **Linuxkit**

# Summary

## TACO (SKT All Container OpenStack)

- **Kubernetes/Container based Life Cycle Management**
- **Container-centric SW Delivery Pipeline (CI)**
- **Community Collaboration (Call for Action)**

## Open Infrastructure for AI

- **DL Training & Application Development**

# To Learn More…

**OpenStack의 컨테이너화 및 Kubernetes를 통한 Lifecycle 관리 기술**

Track I
13:00 ~ 13:30

**Kolla를 이용한 Production-Ready OpenStack Container 생성 및 CI파이프라인**

Track I
17:20 ~ 17:50


**Kubernetes/OpenStack-Helm 튜토리얼**

7월 14일 A-1 10:50~14:50

**Advanced Features for Ceph: Deduplication and QoS**

7월 14일 Ceph Day 14:00~14:50

# Thank You!