



THAIKI Cloud

AEGIS

Contents

1.	Market Background & Industry Trends	04	4.	User-Centric Cloud Experience	015
				Web-Based Desktop Environment	
				IAM	
				Compute	
				Container	
				Storage	
				Cloud Builder	
				Cloud AIOps	
2.	Aegis	06	5.	TCO & Cost Efficiency	023
3.	Core Architecture	08	6.	Deployment & Adoption Models	025
	Hyperscale-Ready Architecture				
	Infrastructure Layer				
	Virtualization & Cloud OS Layer				
	IaaS Management Layer				
	Observability & Operations				
	Security & Compliance		7.	Roadmap & Future Plans	027

1. Market Background & Industry trends

Enterprise infrastructure environments are becoming more complex, and decisions are becoming more cautious.

Enterprises need a Private Cloud that is safer, more efficient and guarantees governance, due to the rising costs, security requirements and evolving regulatory and licensing environments.



Increase in Licence Costs & Unpredictable Costs

- Rapid cost increase caused by global solution policy
- Transition from perpetual to subscription models, resulting in additional expenses.
- CPU/core-based pricing structure increases burden on large-scale customers



Scalability & Modern Cloud Trend Requirements

- Kubernetes and container-centric architecture becoming the standard
- Need for integration with DevOps, AI/ML pipelines, data lakes, etc.
- Existing on-premises environments have limitations in responding to modern development patterns



Data Security & Compliance Requirements Strengthening

- Financial, healthcare, and public sectors require Data Sovereignty
- Organizations are considering a shift from public to private cloud to achieve clearer data storage visibility.
- Risk management for certification, regulatory compliance, and security incidents becomes the responsibility of the enterprise.



Continuously Increasing Operating Costs

- Public Cloud require higher costs than predicted in the long-term
- Rapid cost increases when operating high-performance computing and AI workloads
- Difficult to customize optimization, limiting tuning and performance management

2. Aegis

Aegis

Thaki Cloud

In an ever-changing and uncertain market,
Thaki Cloud provides predictable costs, enterprise-grade stability, sovereign security, and cloud-native operability.

Scalable and Secure Infrastructure Architecture

High-performance architecture capable of stable operation even with large-scale workloads, providing scalability that doesn't limit organizational growth.

Optimized Resource Model for Long-term TCO Reduction

Predictable and efficient pricing is delivered through an objective cost structure, offering a clear advantage over public cloud providers.

Simplified Complex Cloud Management, Stronger Enterprise Operations

Easy Experience from Deployment to Operations Management

Reducing complex Cloud management burden through intuitive UI and automation-based operations.

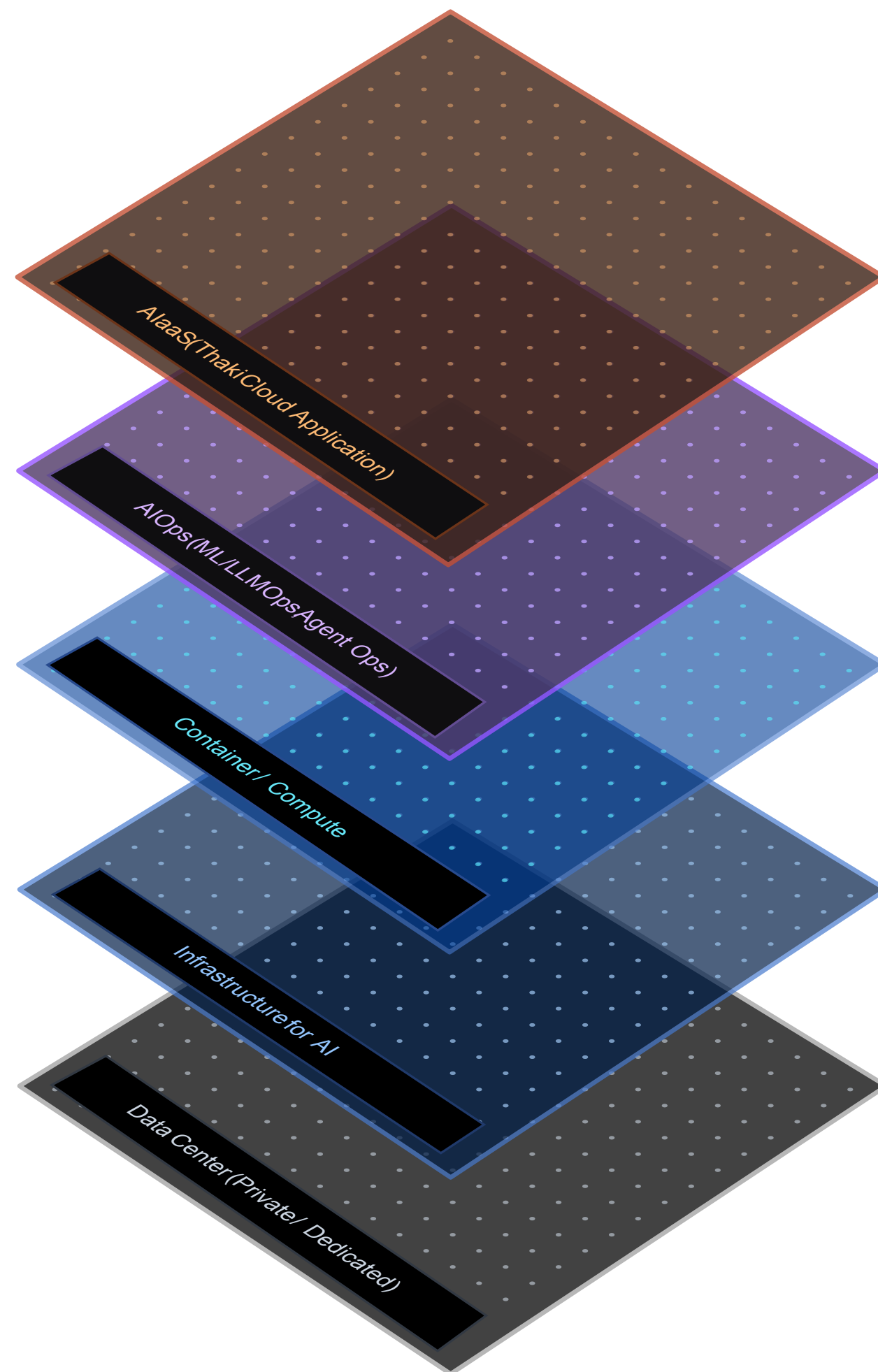
Enterprise Security Framework Ensuring Data Sovereignty

Full control over sensitive data and AI workloads ensures governance aligned with regulatory and security demands.

3. Core Architecture

End-to-End Cloud Spanning IaaS, PaaS, and SaaS

The entire cloud layer from Infrastructure to application operating on a single platform.



Agent Platform (Praxis)

- Provides a production-ready Praxis performing decision-making based on corporate internal data.

AI Platform (Metis)

- Full-stack AI platform allowing automation throughout the cloud lifecycle
- Automated AI operations from data engineering to training and inference

Container / Compute (Aegis)

- Fully managed Kubernetes service connecting applications and infrastructure
- Multi-cluster based Container as a Service environment
- Virtual Machine service based on multi-tenancy architecture

Infrastructure for AI

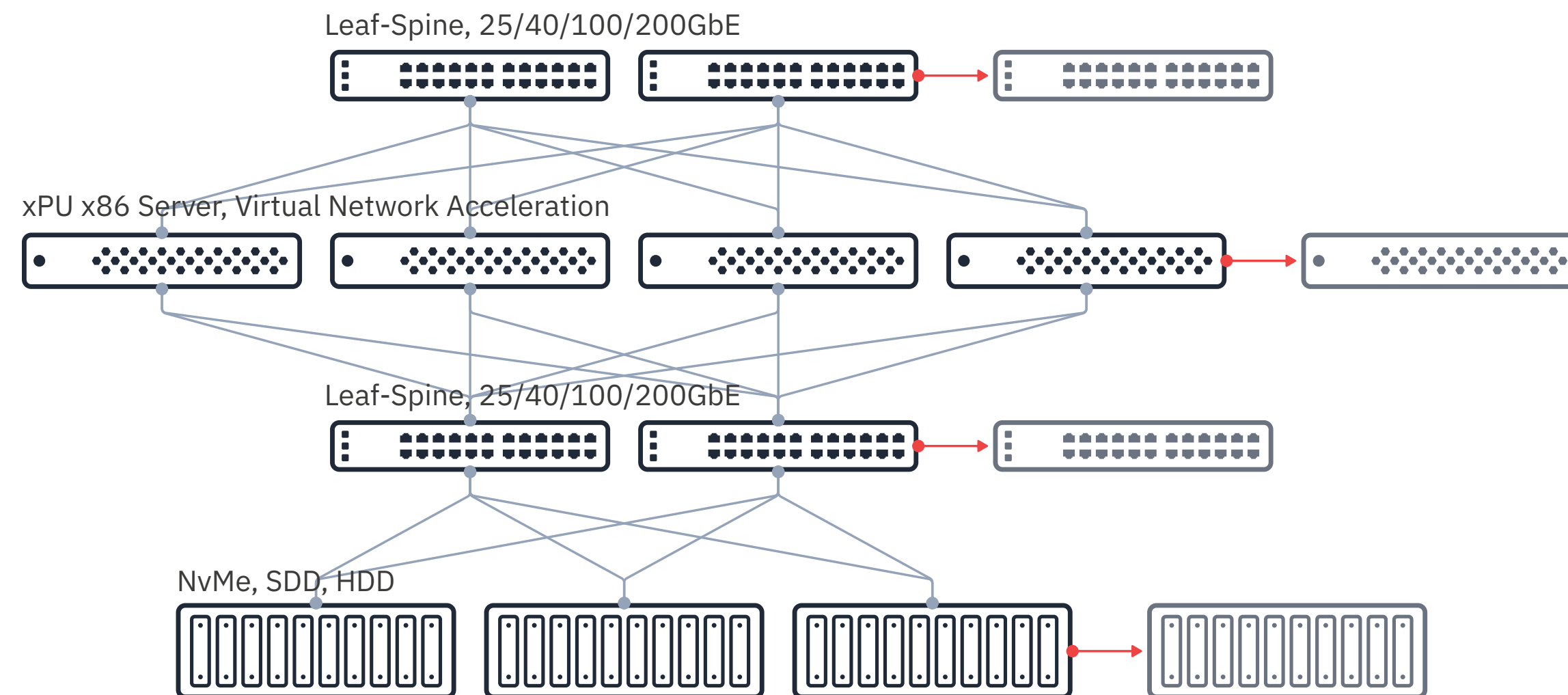
- ML-dedicated infrastructure for AI workloads
- Support for xPU multi-chip environments (CPU, GPU, NPU)

Data Center (Private / Dedicated)

- Private / hybrid AI cloud infrastructure deployable according to customer environment

Scale-Out Based High-Availability Infrastructure Structure

By combining a leaf-spine fabric with distributed storage, the system delivers DPU-based network acceleration and RoCE for ultra-low-latency, high-performance data paths. This ensures stable performance and scalable growth in large environments.



Leaf-Spine Based Scalable Fabric Structure

- Supporting consistent network performance and non-disruptive Scale-Out, allows for a stable infrastructure operation even for large-scale workloads

Distributed Storage Based High Availability

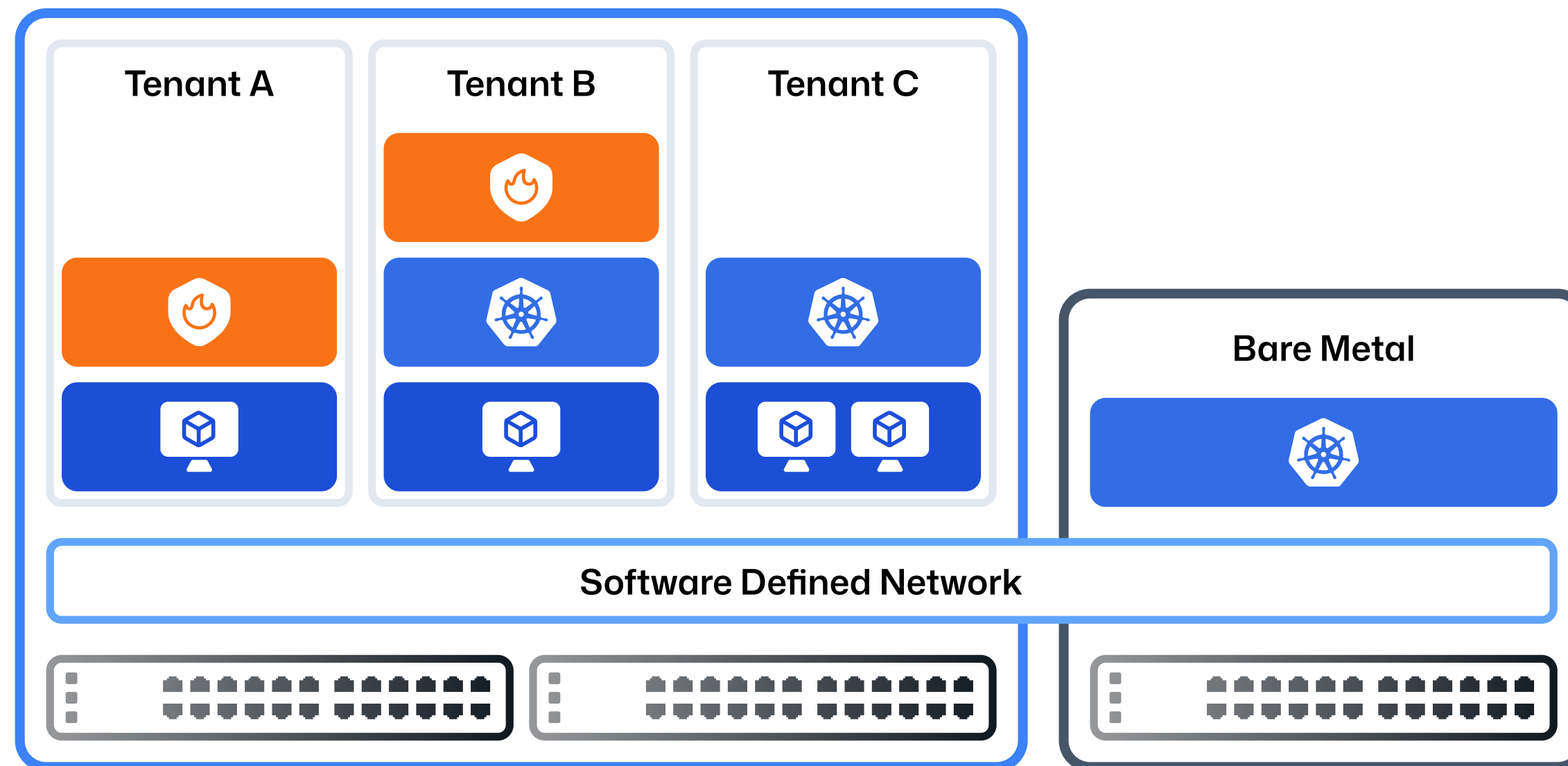
- Data path redundancy and fault isolation structure allows for a stable service continuity without storage performance degradation.

DPU RoCE Based Ultra-Low Latency Data Path

- By using DPU and RoCE to directly exchange memory between servers at high speed, it provides ultra-low latency and high-performance network paths.

Unified Cloud OS Operating VM, Baremetal, and Kubernetes on a Single Platform

Providing KVM/QEMU-based hypervisor and NFV-based network virtualization with SDN technology, enabling integrated management of VM-Container-Baremetal on a single platform, supporting various Kubernetes deployment models for diverse workload operating environments.



KVM/QEMU Based VM and Baremetal Simultaneous Provision

- Providing stable VM operation and Baremetal instances at the same time.
- Delivering optimal computing environment tailored to workloads.

Kubernetes Integrated Operating Environment

- Supporting both Kubernetes on VMs and Baremetal.
- Supports flexible operation of various applications.

Software-Defined Network Based Network Virtualization

- Providing Distributed Virtual Router, LBaaS, FW functions through software.
- Connecting and protecting VM, Container, and Baremetal with consistent policies.

Managing Resource Provisioning and Operations on a Single Platform

Providing core IaaS services including Compute, Container, and Storage based on unified IAM, supporting essential cloud operation monitoring and log collection functions. Also enabling user-driven resource provisioning through Self-Service portal, and providing powerful resource isolation through Multi-Tenant architecture and Quota-based resource management.

Provision of a unified IaaS Service

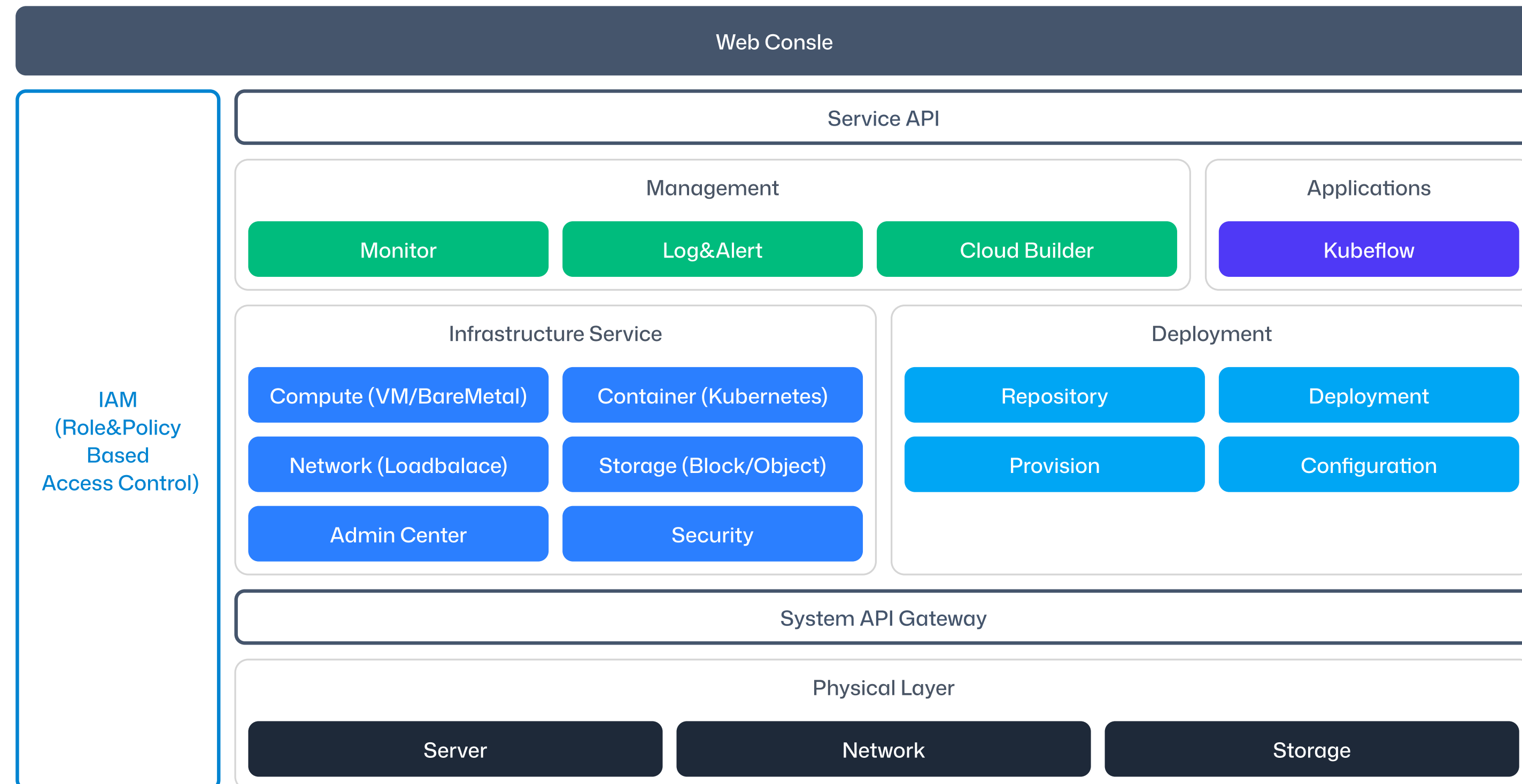
- Providing and managing core infrastructure services including VM, Kubernetes, Storage, and Network on a single platform

IAM-Based Security and Policy Control

- Safely controlling permissions and resources at user, project, and tenant levels through Role-Based/Policy-Based Access Control (RBAC/PBAC)

Cloud Lifecycle Management

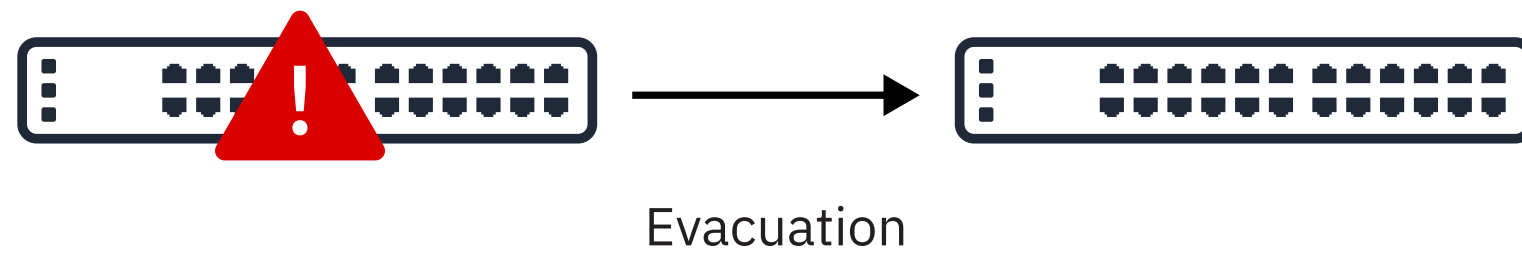
- Supporting automated update, upgrade and standardization of build for core infrastructure including VM, Kubernetes, and Storage through Cloud Builder



Self-Healing Based Fault Response and Real-Time Observability

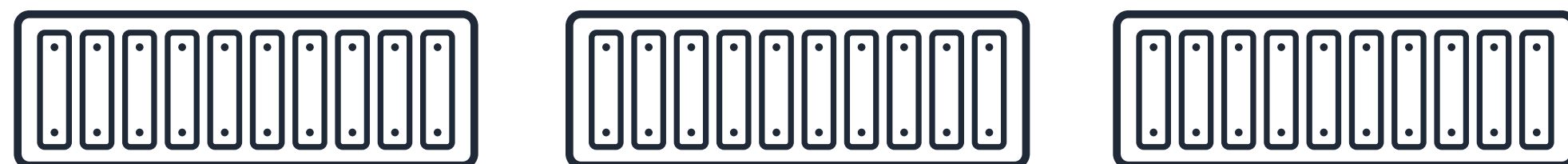
Detecting status across Compute and Storage in real-time, ensuring stable service continuity through automatic recovery (Evacuation/Replica) during failures, and dramatically improving operational transparency and fault response time by centrally managing all system logs on a single platform.

Storage Node



Evacuation

Storage Node



Replication

Replication

Automatic Fault Recovery

- Evacuation function supported to automatically relocate workloads to other nodes during Compute Node failures.
- Storage Node provides data stability and high availability with basic 3-Replica structure.

Real-Time Status Monitoring

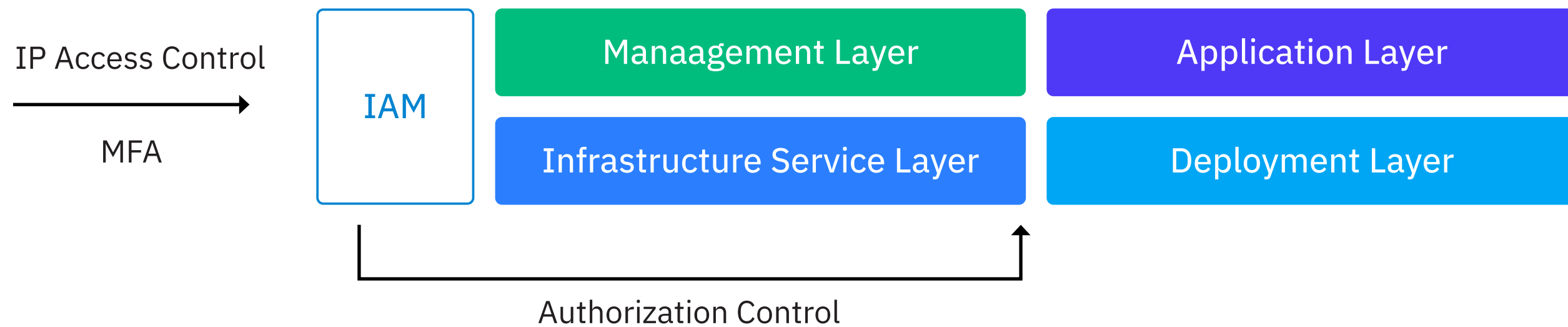
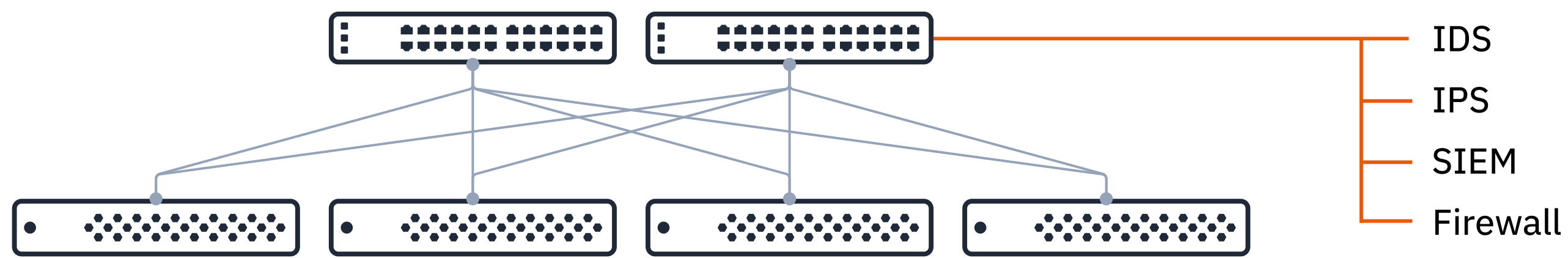
- Real-time monitoring of infrastructure status within the cloud including Storage Node, Disk, and Hypervisor

Unified Log-Based Operational Transparency

- Strengthening fault tracking and operational transparency by centrally collecting and analyzing events and logs across the platform

Enterprise Security and Control Meeting Organizational Security Standards and Regulatory Requirements

Providing powerful security and consistent control across all areas including VM, Container, and Network through on-premises based security model, RBAC/PBAC IAM, MFA, and network/tenant isolation.



On-Premises Based Customized Security Model

- Designing and applying security policies at network, workload, and tenant levels according to organizational regulatory and security requirements

RBAC/PBAC Based IAM; Authentication Framework

- Safely controlling permissions at user, project, and tenant levels through role/policy-based access control with MFA and IP access control support

Network and Tenant Isolation Based Access Control

- Providing consistent security and isolation across all layers including VM, Container, and Network through Tenant/namespace Isolation, Security Group, Network Policy, FWaaS

4. User-Centric Cloud Experience

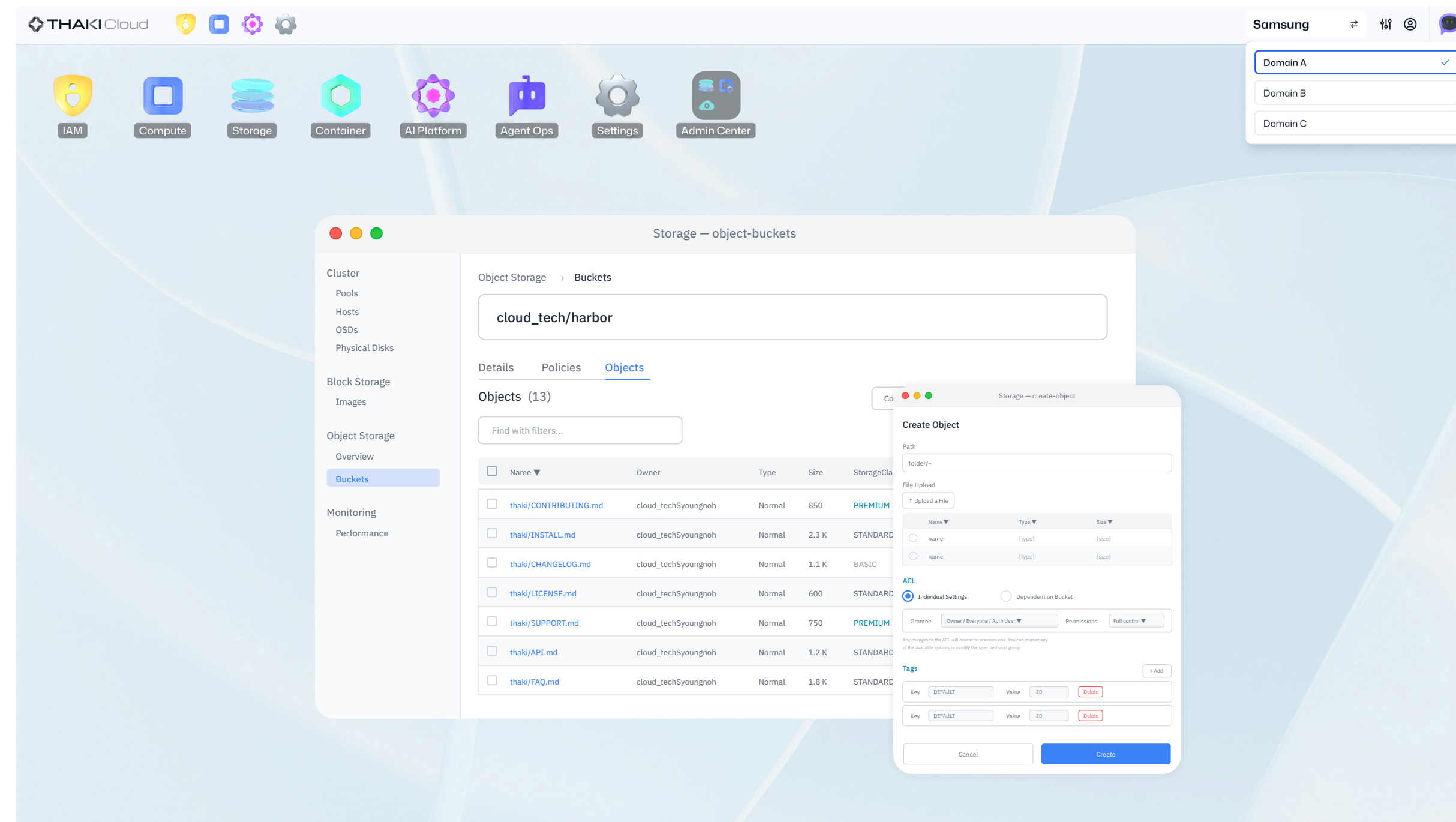
User-Centric Cloud Experience

Web-based Desktop Environment

Run and manage all cloud services as apps from a web-based desktop UI. Easily control versions and permission of each app, with support for multi-tasking by opening multiple windows at the same time.

One-Click Expansion from Store

Install and upgrade apps from the App Store to quickly expand cloud functionality



App-Based Service Management

Each function is set up as an independent app, allowing selective installation, execution, and permission management

Window-Based Multi-Tasking

Open multiple windows simultaneously for desktop-level multi-tasking experience

IAM (Identity and Access Management)

Role-based permission management (RBAC) provides fine-grained access control per user, group, and project. MFA and IP-based access control strengthen cloud access security.

Unified User, Group, and Policy Management

- Easily manage account creation, policy deployment, and permission changes
- Stable management even as cloud operations scale grows

RBAC-Based Permission Management

- Role-based permission management supporting direct organization/project-level management to detailed service/resource-level management

Full Audit Log Tracking

- Real-time audit logs confirming all user activities, API calls, and resource change records
- Analyzing security incidents and threats to improve operational stability

MFA and IP Policy-Based Security Enforcement

- Applying additional authentication steps for detailed operations
- Configuring strong security policies allowing only permitted networks/IPs to connect

The image displays four screenshots of the IAM console interface, illustrating various management and security features:

- iam — user-groups:** A table listing user groups with columns for Name, Type, Roles, Users, Created, and Action. The table shows five entries, all of type 'Built-in' with 'Viewer (+3)' roles and 130 users, created on 2025-09-12.
- IAM — create-role:** A 'Create Role' page with 'Step 1: Select trusted entity'. The 'Cloud Service' option is selected, indicating services that need to perform actions.
- IAM — mfa-settings:** 'MFA Policies Settings' page for configuring multi-factor authentication. The 'Voluntary' option is selected, allowing users to optionally enable MFA.
- IAM — audit-logs:** 'Audit Logs' page showing a table of events. The table includes columns for Timestamp, Event Type, Actor, and Policy Action. Recent events include 'MFA Enabled', 'User Deleted', and 'Group Deleted'.

User-Centric Cloud Experience

Compute

Template structure allows easy saving and repeated deployment of frequently used compute environments. Multi-tenancy and Quota settings enable efficient resource management for team and project.

The screenshot displays the 'Compute — dashboard' interface for 'Proj-1'. The main content area includes:

- Project Overview:** 'proj-1' with ID: 7284d91f4e81431e93060a9bbcf2cdfd, created on 2025-07-25. My Role: Member. Description: Dev environment.
- Instance Summary:**

Total	Active	Error	Shutoff	Others
13	10	0	0	3
- Core Compute Quota:** Visual progress indicators for vCPU, RAM, Disk, and GPU.
- Infrastructure Quota:**

Volumes	2/10
Networks	10/100
Routers	5/10
Ports	49/500
Floating IPs	2/50
Security Groups	16/100

Overlaid on the dashboard are two other windows:

- Compute — templates:** A table listing various templates for selection.

Name	Image	vCPU	RAM	Disk	Network	Floating IP	SG	Access	Edit
th.tiny	Ubuntu 22.04 LTS	2	16 GiB	40 GiB	net-default	None	default	Personal	Edit
th.tiny	Ubuntu 22.04 LTS	3	32 GiB	80 GiB	net-default	None	default	Personal	Edit
th.tiny	Ubuntu 22.04 LTS	3	32 GiB	80 GiB	net-default	None	default	Personal	Edit
th.tiny	Ubuntu 22.04 LTS	3	32 GiB	80 GiB	net-default	None	default	Personal	Edit
th.tiny	Ubuntu 22.04 LTS	2	16 GiB	40 GiB	net-default	None	default	Personal	Edit
- compute — select-project:** A list of projects for selection, including Project-01, Project-02, Project-03, and Project-04.

Flexible Scaling Options

- Freely scale instance specs including CPU, memory, and storage according to workload
- Supporting diverse environments from AI training to general web services

Multi-Tenancy and Quota-Based Resource Management

- Providing independent tenancy environments for each organization or team
- Setting Quotas per project to control resource usage
- Preventing excessive usage and ensuring stable operation

Template-Based Repeated Deployment Automation

- Install and upgrade apps from the App Store to quickly expand cloud functionality

User-Centric Cloud Experience

Container

Supports Multi-Cluster to enable operation of various Kubernetes environments in one place. Web UI and YAML-based IaC are provided together to enhance container management convenience.

The screenshot displays a web interface for managing Kubernetes nodes. The main window, titled "container — node-details", shows the following information:

- Namespace:** default
- Cluster:** Dashboard, Nodes (selected), Namespaces
- Workloads:** Pods (111)
- Storage:** Info, Images, Taints, Conditions
- Policy:**

The node details section shows:

- Node: nodeName**
- Status:** Active
- K8s Version:** v1.34
- OS:** Ubuntu 24.04
- Runtime:** containerd
- Labels:** 9
- Created:** 2025-07-25

Health checks are shown as "Ready":

- PID Pressure: Ready
- Disk Pressure: Ready
- Memory: Ready
- kubelet: Ready

CPU usage is shown as 27%.

An overlay window titled "deployment.yaml" shows the following YAML configuration:

```

1  apiVersion : apps/v1
2  kind : Deployment
3  metadata :
4  name : sample-app-deployment
5  namespace : production
6
7  annotations :
8  app.kubernetes.io/description : "Primary app"
9  app.kubernetes.io/owner : "platform-team"
10 app.kubernetes.io/last-reviewed : "2025-01-15"
11
12 labels :
13 app : sample-app
14 environment : production
15 tier : backend
16 version : "v2"
17 managed-by : "kubernetes"
18
19 spec :
20 replicas : 3
21 selector :
22 matchLabels :
```

Unified Multi-Cluster Operation

- Unified management of Kubernetes clusters across multiple regions, teams, and environments in one UI
- Reducing operational complexity and increasing scalability

Simple Cluster Creation & Operation

- Straightforward experience of cluster creation, node expansion, and upgrades with just a few clicks through UI

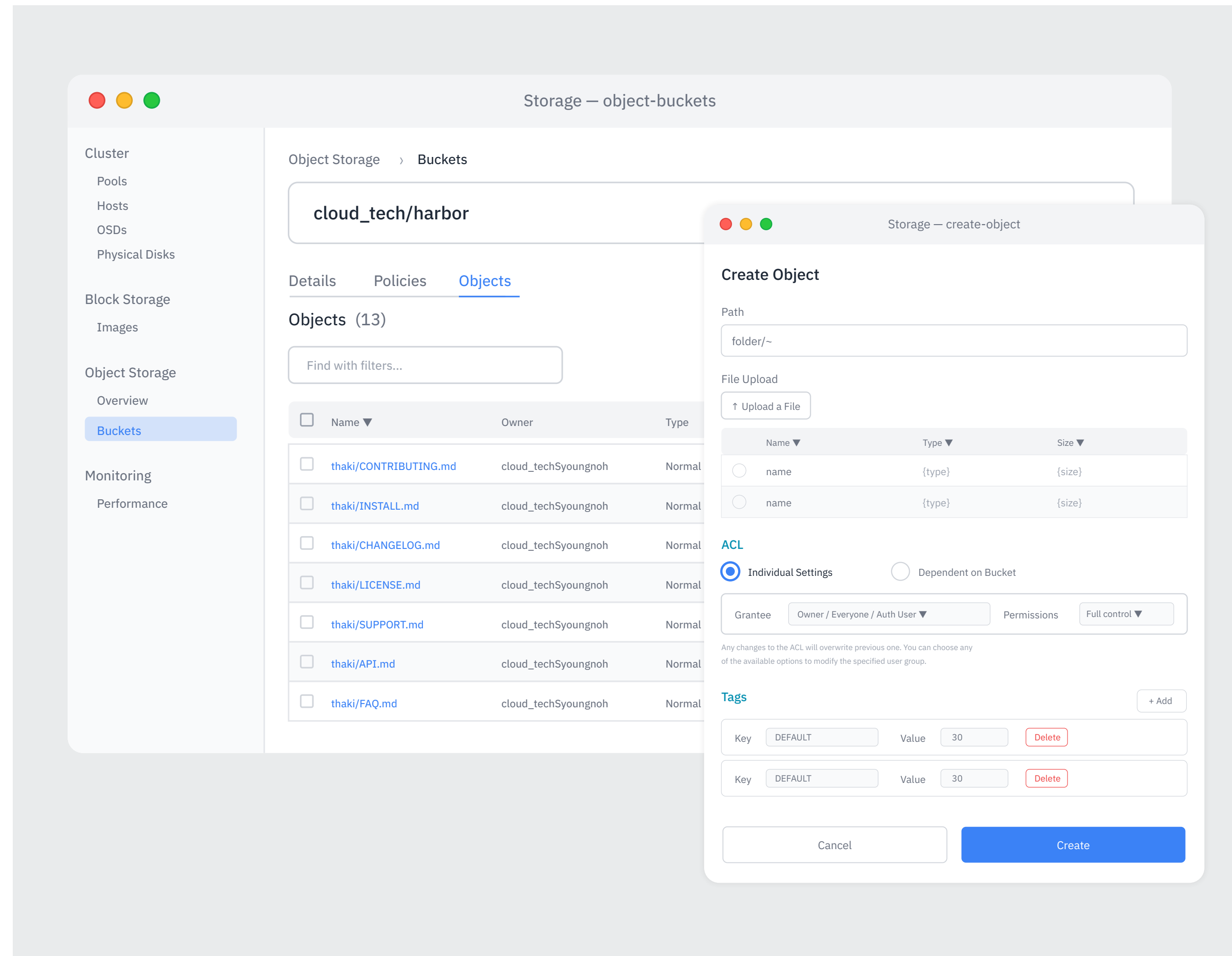
Web UI + YAML-Based IaC Support

- Supporting YAML-based infrastructure automation alongside UI
- Flexible selection according to development and operations methods

Storage

Block and Object storage both available for flexibly support of various workloads.

Disaster recovery features ensure data availability even during node and disk failure situations.



Block & Object Storage

- Providing both Block Storage for VM disks and Object Storage for large-scale data storage
- Processing various storage requirements on a single platform

High Availability and Disaster Recovery Structure

- Minimizing service interruption through data replication and automatic recovery during node or disk failures
- Providing enterprise-level stability and durability

Flexible Scalability

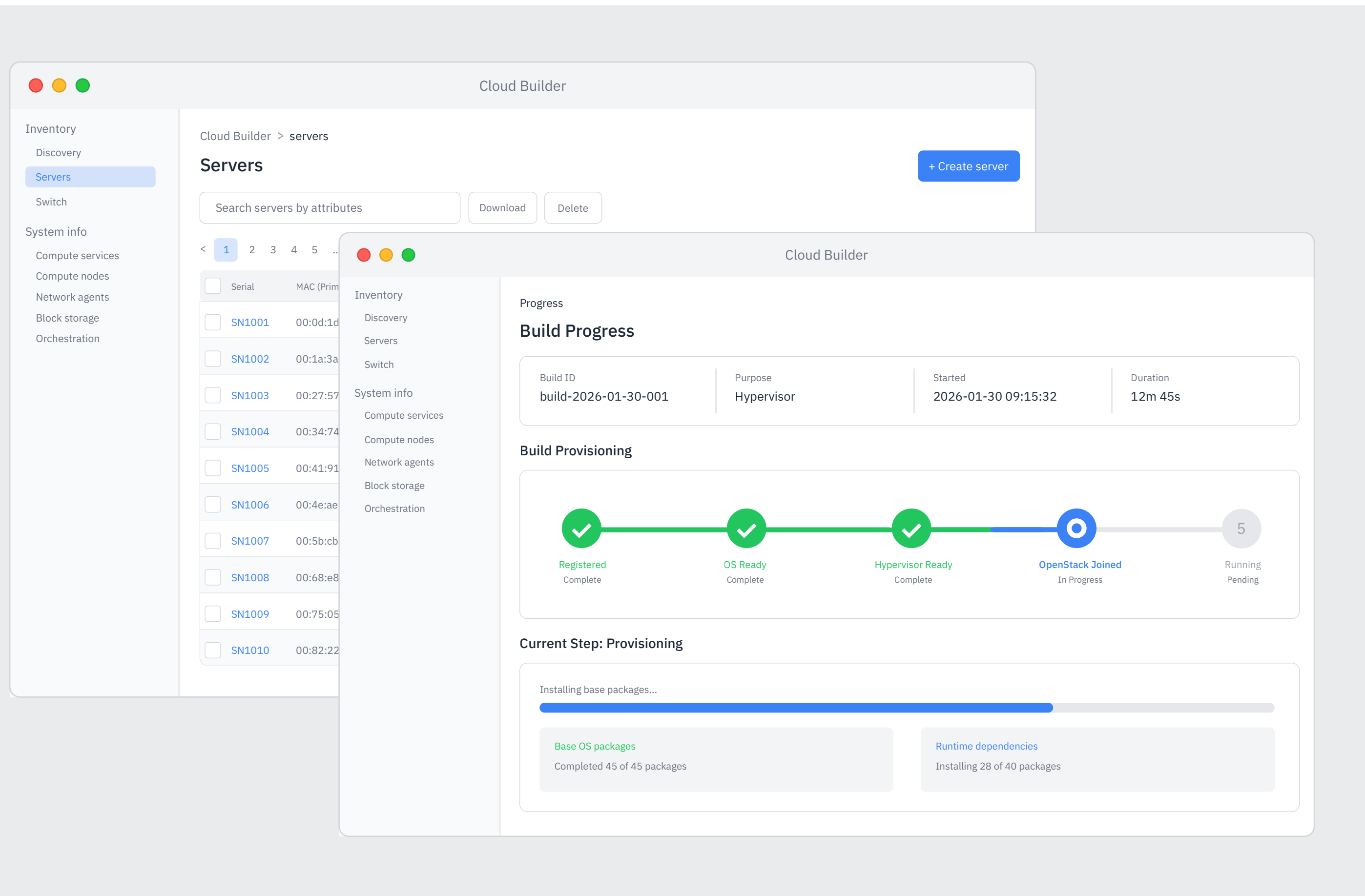
- Expanding storage capacity without interruption to flexibly respond to data growth

User-Centric Cloud Experience

Cloud Builder

Build cloud environments quickly and easily.

Update and upgrade support maintains stable cloud operations at the latest state.



Batch Deployment and Installation of OS and Cloud Components

- Template-based consistent deployment of Base OS, network configuration, storage, and OpenStack configuration
- Standardized installation quality

Automated Physical Server Registration and Provisioning

- Automating pre-cloud deployment stages including Baremetal server detection, BMC information auto-input, and HW status inquiry
- Greatly reducing initial deployment complexity

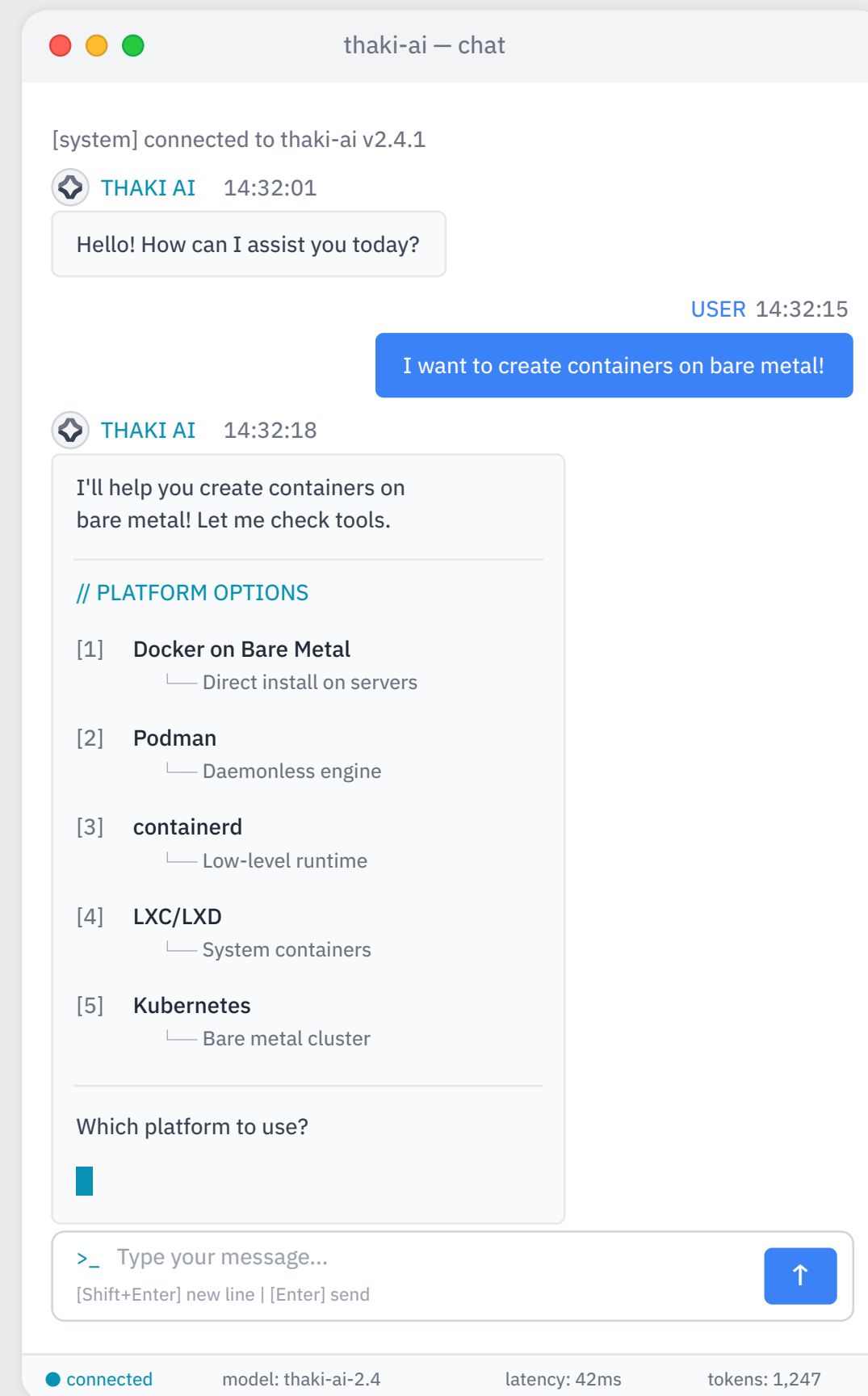
Stable Updates and Upgrades Available

- Applying updates and upgrades without service interruption
- Maintaining stable platform state

Thaki Cloud Assistant

Chatbot-based interface for infrastructure creation, deletion, and modification tasks in natural language.

Supporting Claude LLM and on-premises LLM integration to build AI operations environments that meet security requirements.



Natural Language Chatbot-Based Infrastructure Management

- Performing infrastructure creation, deletion, and modification through natural language-based chatbot
- Massive improvement in operational convenience and speed

Claude LLM Integration Support

- Claude integration allows intelligent analysis and operation assistance

On-Premises LLM Option

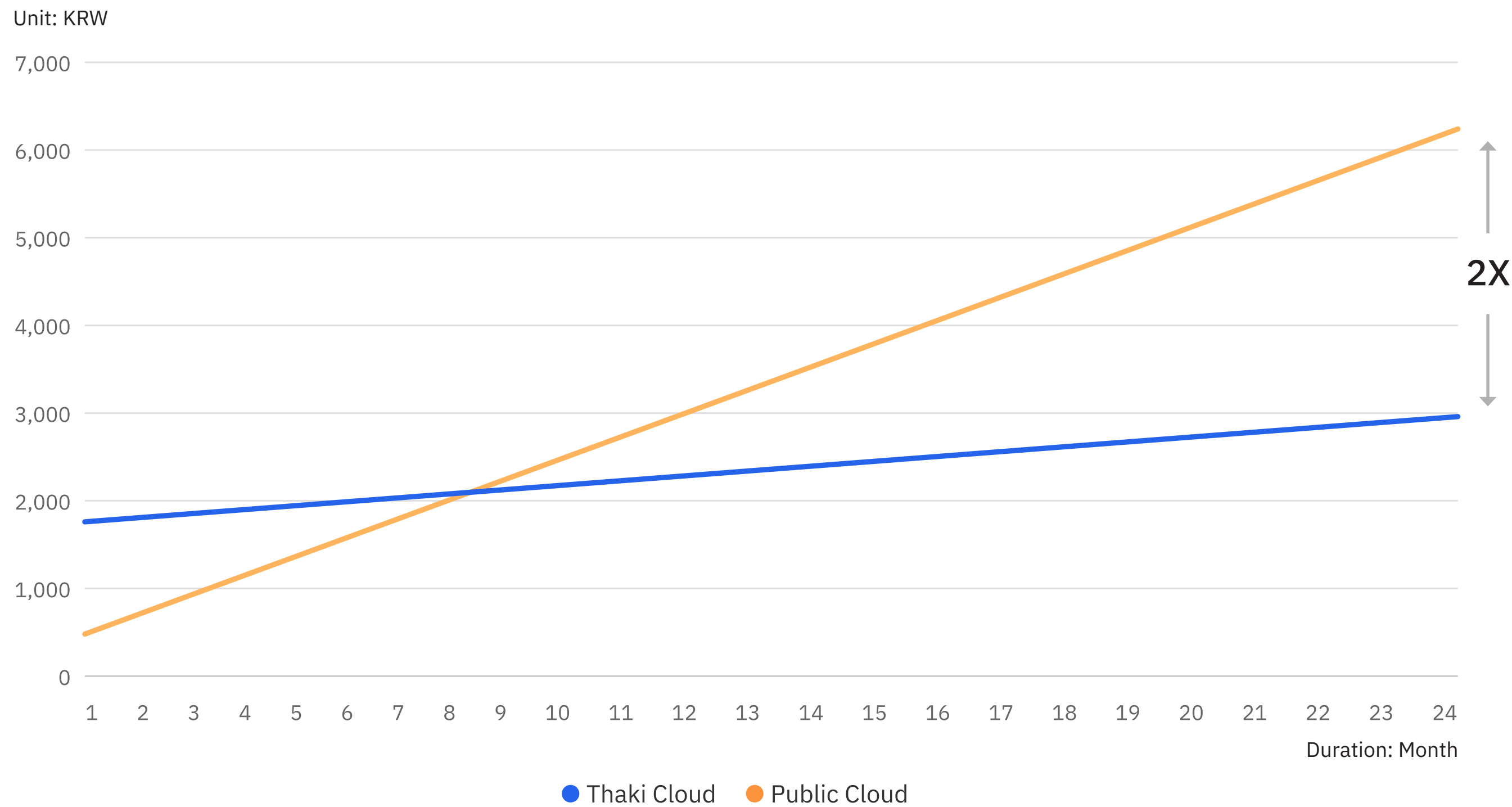
- Selecting LLM running in internal environment to implement secure AI operations where data doesn't leave externally
- Safe to use even in environments with high regulations and security requirements

5. TCO & Cost Efficiency

TCO & Cost Efficiency

Private Cloud Providing Long-Term TCO Efficiency Compared to Public Cloud

Thaki Cloud-based private cloud reaches TCO equality with public cloud at approximately 8.2 months, and after 24 months, private cloud shows approximately 2x lower cost structure compared to equivalent scale.



Cost efficiency catching up in 8.2 Months



2X TCO difference at 24-month point

Thaki Cloud

CAPEX: Servers, networks, firewalls, etc.

24/7 operations, staff of 6 people

30-50 Node

Public Cloud (A company)

Based on AMD Epyc CPU 4Core 8G Mem VM

1,200 units baseline

Exchange rate ₩1,440 applied

6. Deployment & Adoption Models

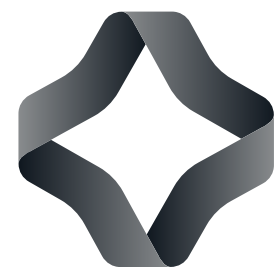
Two Private Cloud Delivery Models Selectable by Operating Method

Two models available:

Managed Private Cloud with public cloud-like experience, and On-Premise Private Cloud tailored to your infrastructure.

Managed Private Cloud

- Providing complete deployment, operation, and maintenance for datacenter, servers, networks, and storage
- Delivering the same operational experience as public cloud



License-based Aegis deployment on your existing infrastructure for a tailored Private Cloud solution.

On-Premise Private Cloud

- Providing Aegis in license form on customer-owned infrastructure environments
- Supporting optimized independent Private Cloud deployment for customer environments



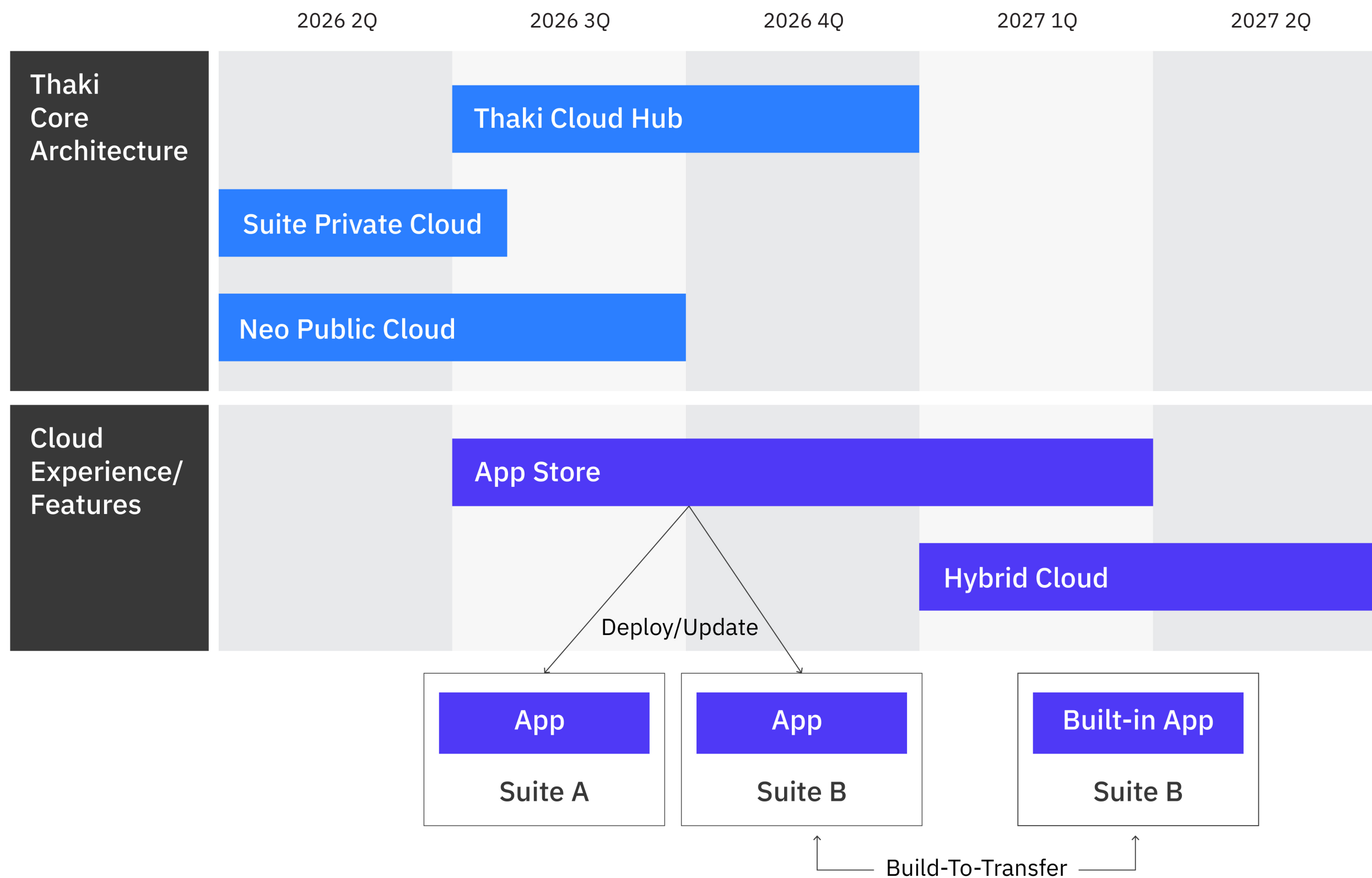
7.

Roadmap & Future Plans

Hub & Neo: From Private to Multi-Hybrid Cloud

Thaki Neo Cloud complements Private Cloud limitations through public cloud resources and ensures complete compatibility with Aegis.

Thaki Cloud Hub provides multi-private cloud management and Store-based application deployment/updates, delivering a complete hybrid cloud environment between public and private.



Thaki Neo Cloud: 2026 2Q

- Public Cloud service delivery
- Complementing Private Cloud resource limitations through Public
- Ensuring compatibility between Aegis and Neo Cloud

Thaki Cloud Hub: 2026 3Q

- Multi Private Cloud Management supported
- Service (Management, UI, Application) deployment/update available through App Store
- Hybrid Cloud implementation through Thaki Neo Cloud integration