雲端 Cloud Computing Center for Mobile Applications

ITRI Cloud Operating System and OpenStack

Tzi-cker Chiueh 闕志克

Cloud Computing Research Center for Mobile Applications (CCMA)

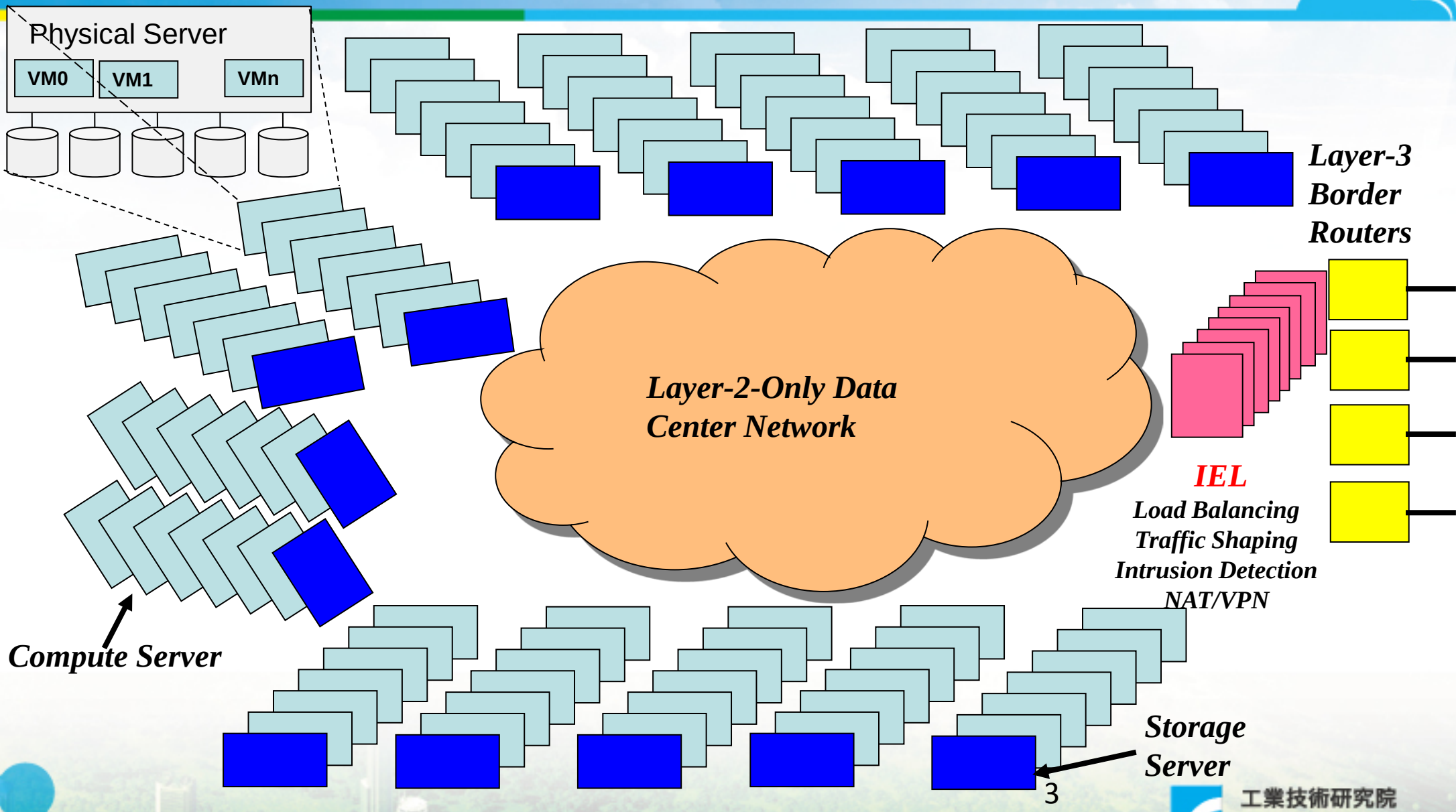
雲端運算行動應用研究中心

Cloud Data Center Solution

- Renting rather than buying IT infrastructure ➔ Build-up of cloud-scale data centers ➔ Need for inexpensive integrated cloud data center solutions
- The user pain point: integration

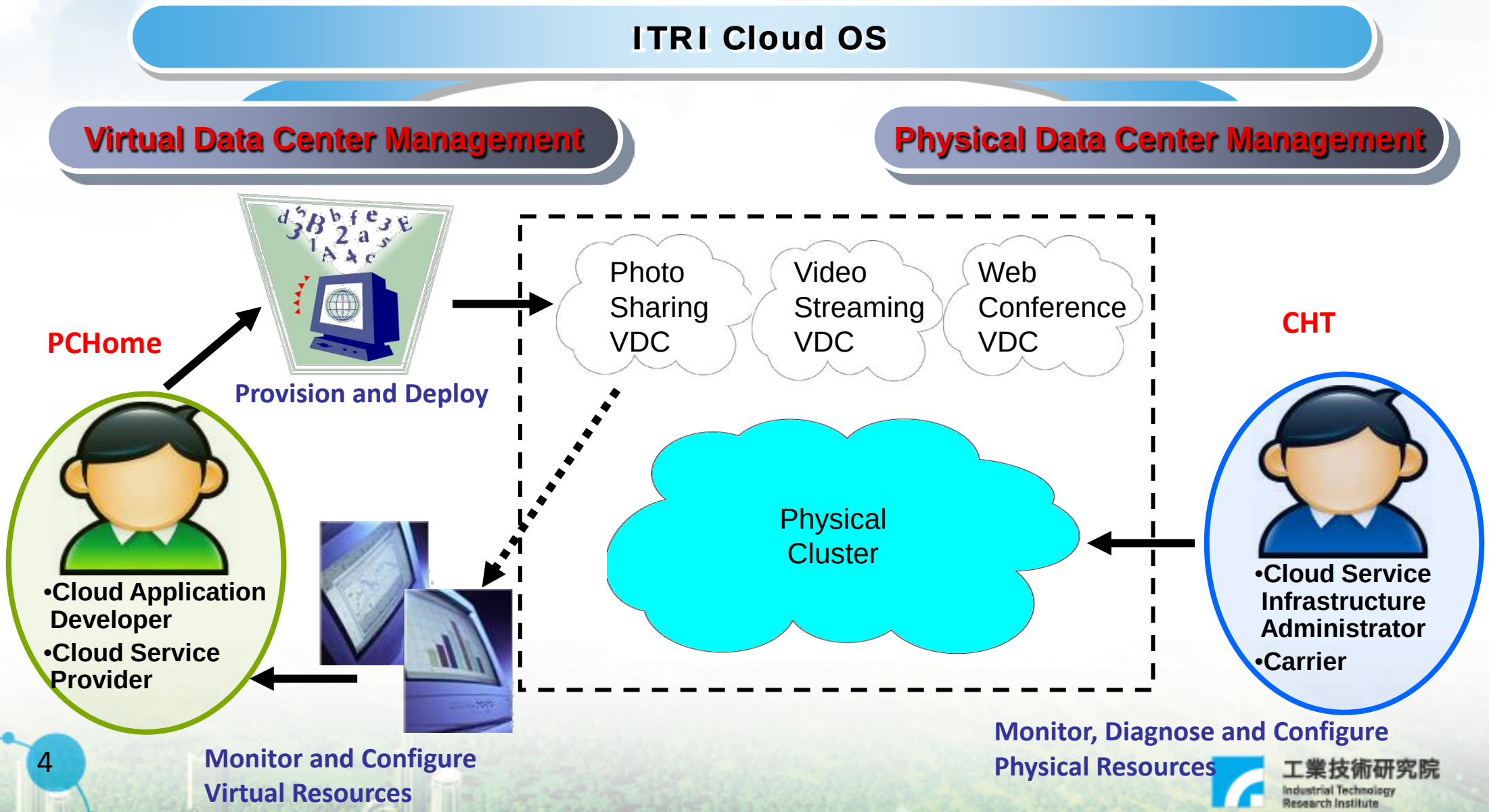
  Is it possible to build a cloud data center like “take a HW box, install OS on it, and have an AWS-like IaaS ready to go”?
- A total IaaS solution for both public and private clouds

Container Computer Architecture



工業技術研究院
Industrial Technology
Research Institute

What is Cloud OS?



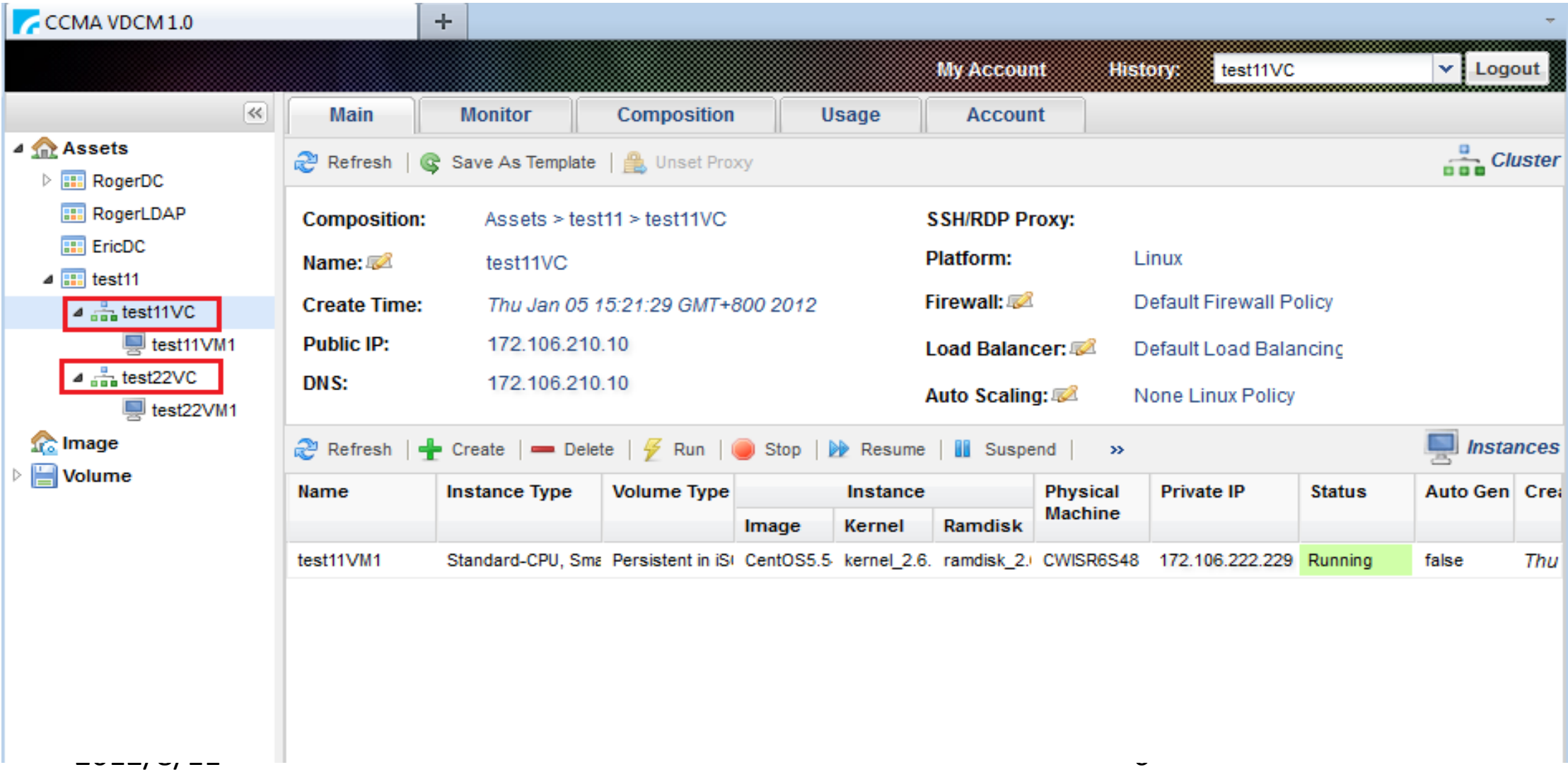
工業技術研究院
Industrial Technology Research Institute

Cloud OS Service Model

- Virtual data center consists of one or multiple virtual clusters, each of which comprises one or multiple VMs
- Users provide a Virtual Cluster specification
  - No. of VM instances each with CPU performance and memory size requirement
  - Per-VM storage space requirement
  - External network bandwidth requirement
  - Security policy
  - Backup policy
  - Load balancing policy
  - Network configuration, e.g. public IP address and private IP address range
  - OS image and application image

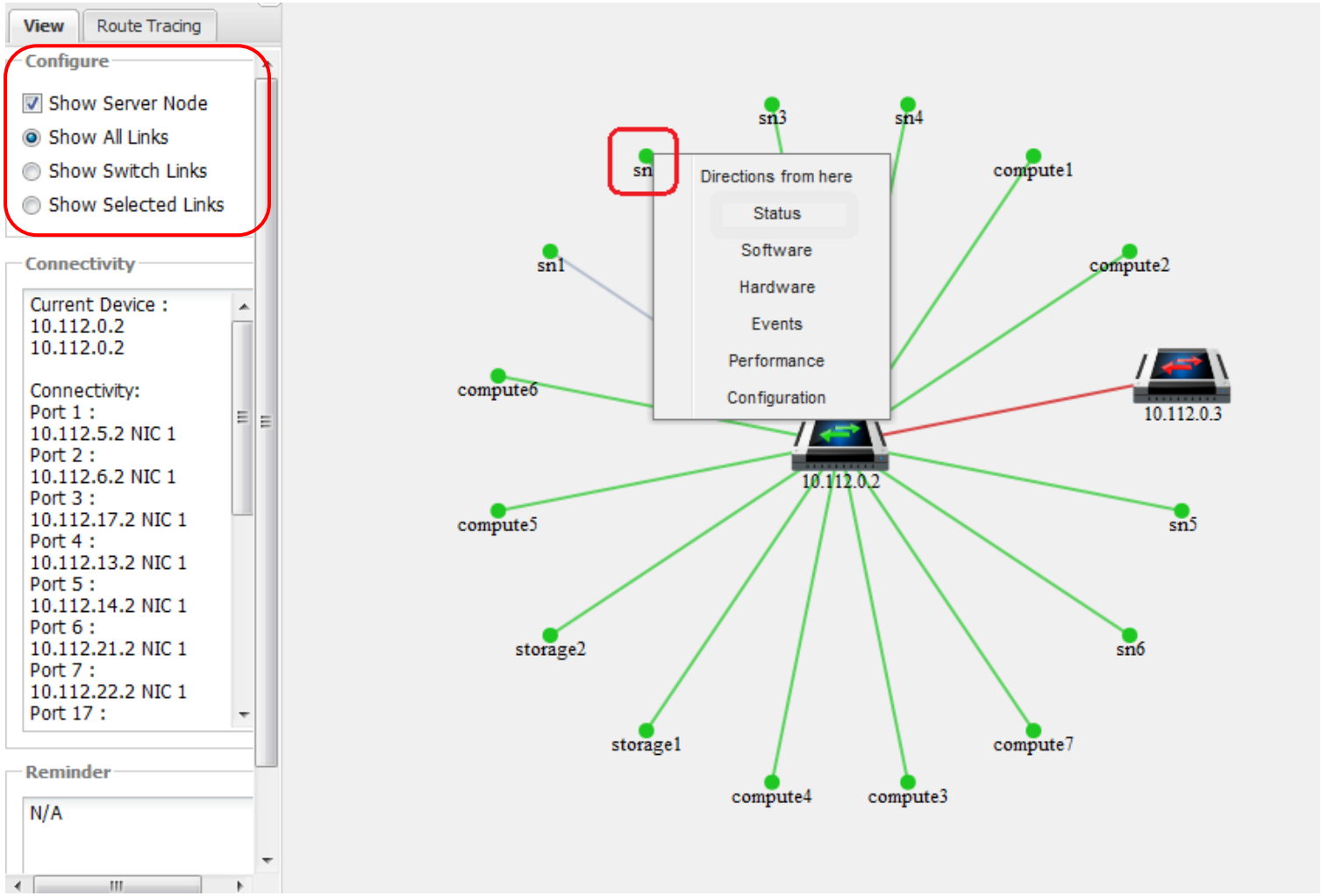
工業技術研究院
Industrial Technology Research Institute

VDCM – Assets (VDC, VC, VM)



PDCM Event Monitor

PDCM Network Topology



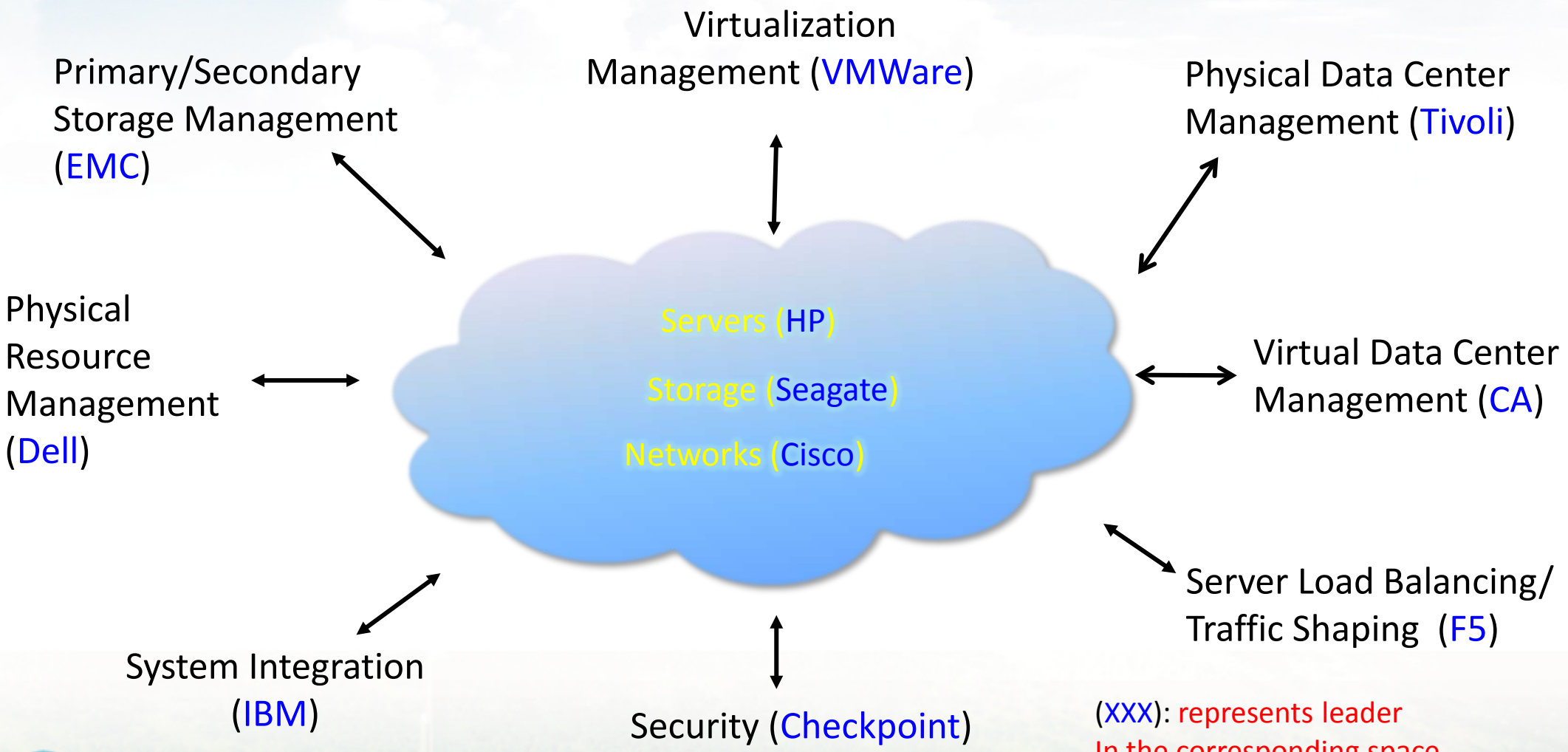
Key Cloud OS 1.0 Features – 1

- Physical resource management (PRM): BIOS
  - Centralized installation of all systems software
  - Start up, shut down, and recover a data center computer
- Data center storage management: file management
  - Main storage (DMS) : Forming a highly available global storage pool from: a set of commodity JBOD storage servers
  - Secondary storage (DSS): Offering streamlined disk-based snapshot/backup with configurable policy, and scalable de-duplication
- Virtualization management: process management
  - Resource provisioning management (RPM): allocate physical data center resources for a given virtual data center and auto-scaling
  - Dynamic virtual resource management (DVMM): use VM migration to support consolidation, load balancing and high availability

Key Cloud OS 1.0 Features – 2

- Physical data center management (PDCM): system administration
  - Comprehensive SNMP-based monitoring
  - Integrated virtual/physical resource mapping view
  - Unified event logging
  - Integrated trouble ticking support
- Virtual data center management (VDCM): system administration
  - VDC/VC/VM specification
  - Real-time resource usage and performance monitoring
- Security: security
  - Inter-VDC isolation
  - Centralized L3 and distributed L7 and web application firewalling
- Internet edge logic
  - Supporting inter-VM load balancing within a VC
  - DDoS attack mediation
  - Distributed traffic shaping

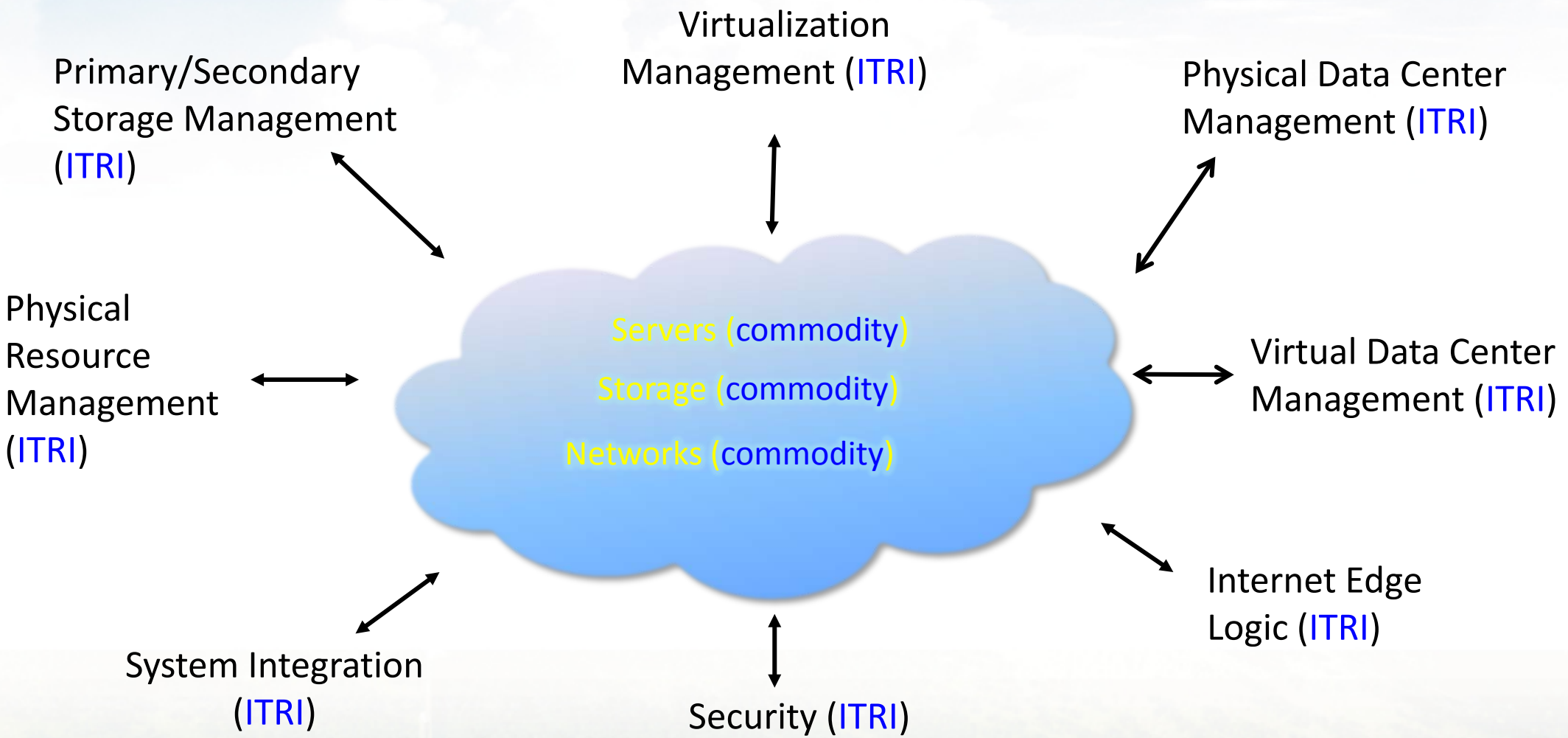
Building Cloud Data Center



(XXX): represents leader
In the corresponding space

工業技術研究院
Industrial Technology Research Institute

ITRI Cloud OS's Way



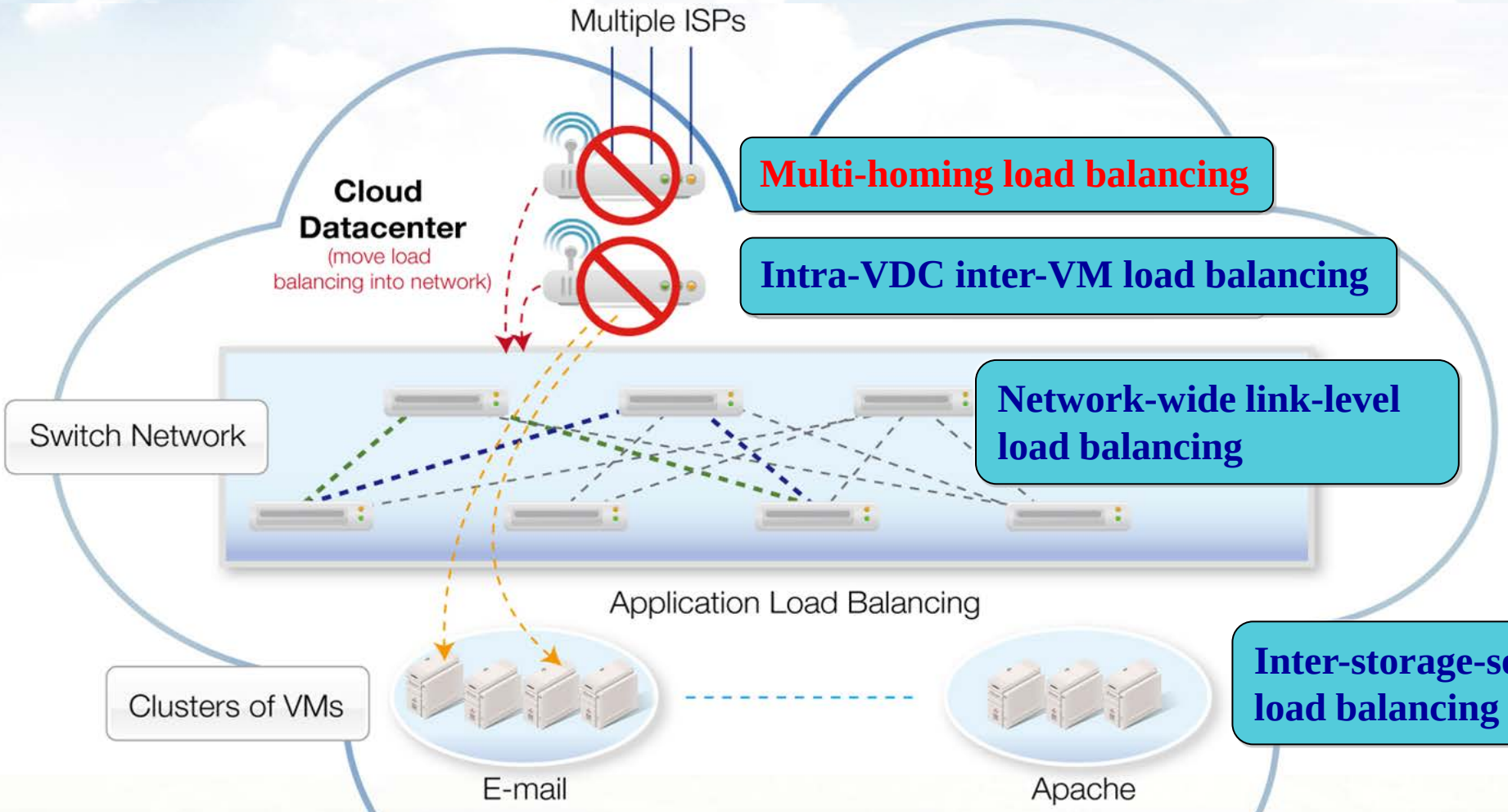
Strong Data Protection

- N-way data replication vs. RAID
  - End to end data availability: disk, server, and network failures
- Periodic snapshots for local data backup with de-duplication
- Wide-area data backup
  - Snapshot frequency: a couple of hours to days
- Wide-area data replication (Cloud OS 2.0)
  - Snapshot frequency: a couple of seconds to minutes

High Availability

- High availability support for Cloud OS subsystems
  - Active-passive: Linux HA + DRBD + edit logging/recovery
  - Active-active: MySQL and server load balancer
- Disk state-preserving fail-over for applications running inside VDCs
  - Shared persistent state + VM restart + take-over
- Memory state-preserving fail-over for applications running inside VDCs (Cloud OS 2.0)
  - Shared memory/persistent state + VM resume + take-over

Multi-Dimensional Load Balancing



Multi-homing load balancing

Intra-VDC inter-VM load balancing

Network-wide link-level load balancing

Inter-storage-server load balancing

Inter-physical-server load balancing

Thermal capacity load balancing

工業技術研究院
Industrial Technology Research Institute

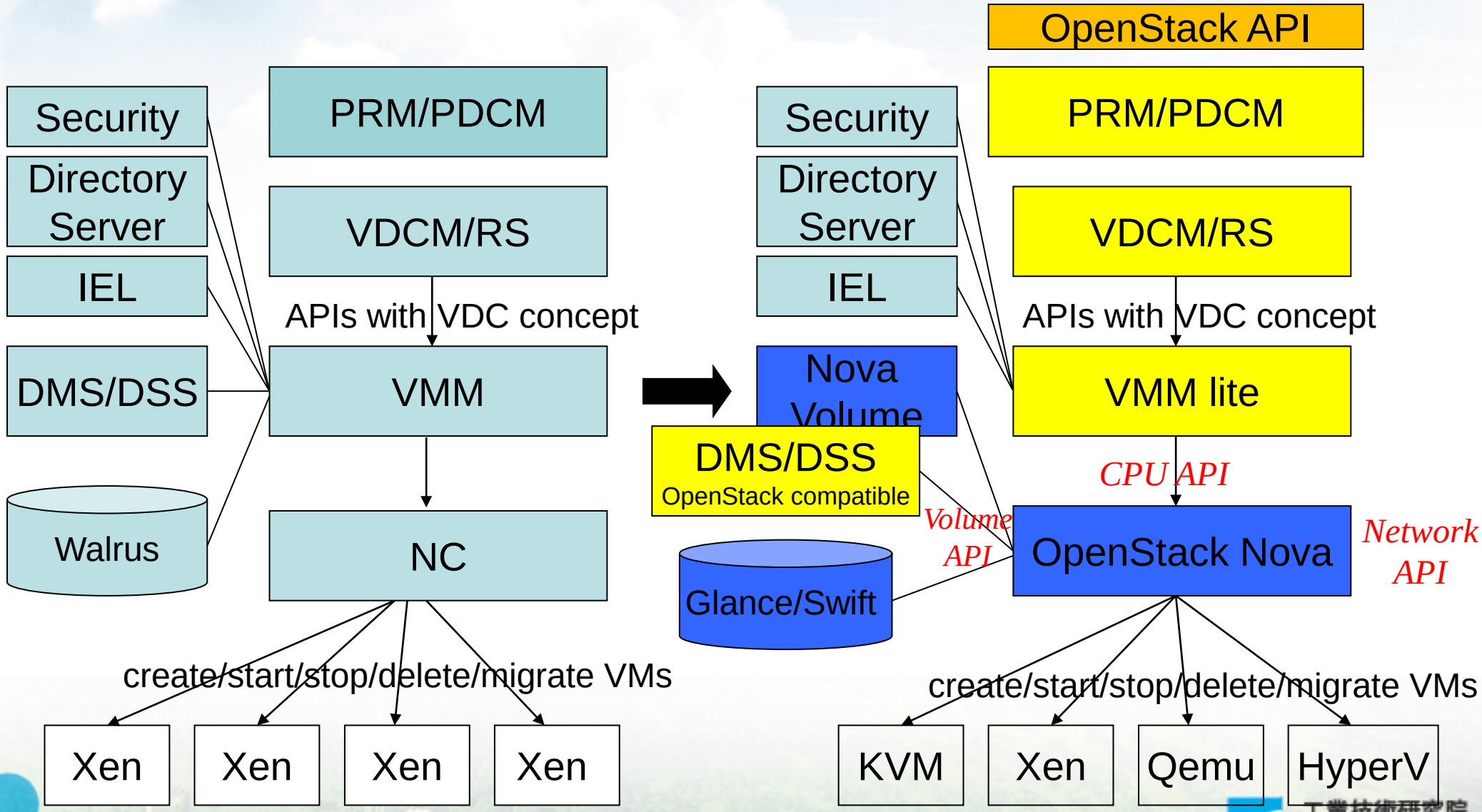
Cloud Security

- Any security breaches that are possible for a physical data center are equally likely for a virtual data center
  - L4/L7 and Web Application Firewall
- New security concerns
  - Interference between tenants on the same physical machines
    - Inter-VDC isolation vs. VLAN isolation

OpenStack

- Open Stack core:
  - Nova: VM provisioning
  - Glance: VM image upload and delivery
  - Swift: Object data storage
- RPM vs. Nova
  - Boot from remote cloned volume
  - Dynamic load balancing
  - Power consolidation
  - Dedicated physical machine pool
  - Auto-scaling

OpenStack-Compatible Cloud OS



工業技術研究院
Industrial Technology Research Institute

Cloud OS 2.0

- OpenStack Compatible:
  - Nova's compute, volume and network API
  - OpenStack web service API
  - Target date: 10/1/2012
- Data center federation: Support for multi-site data centers
- Network virtualization: Support for hybrid cloud
- Wide-area data replication
- Memory de-duplication

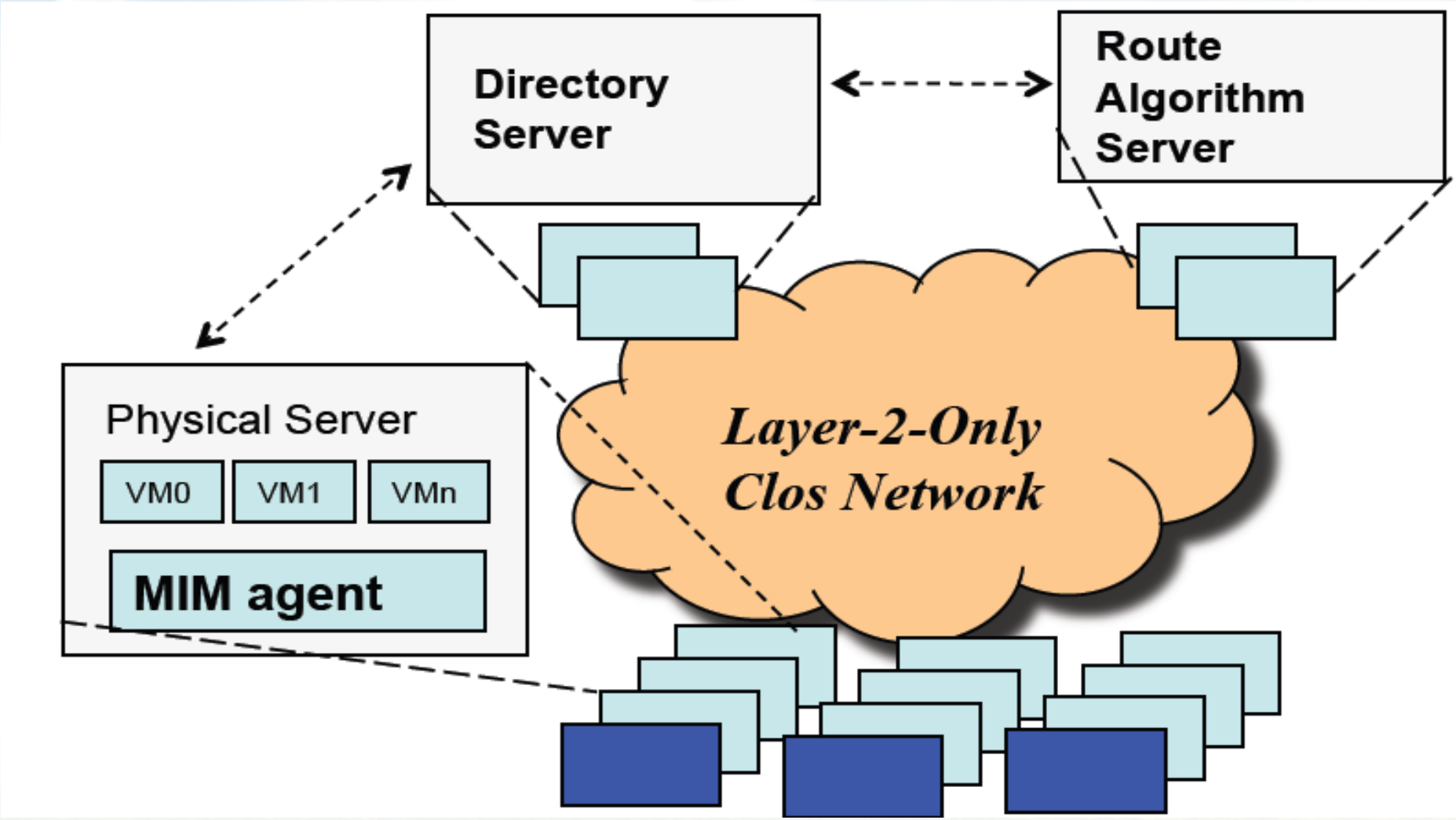
Cloud Data Center Network

- Cloud data centers are Big and Shared
- Scalable and available data center fabrics
  - Not all links are used
  - No load-sensitive routing
  - Fail-over latency is high (> 5 seconds)
- Network virtualization: Each virtual data center (VDC) gets to define its own network
  - All VMs in a VDC belong to one flat subnet
  - Each VDC has its own private IP address space
  - Each VDC has a set of public IP addresses
  - Each VDC has a set of external VPN connections
  - Per-VDC Internet traffic shaping policy, intra-VDC and inter-VDC firewalling policy, and server load balancing policy

Peregrine

- A unified Layer-2-only network for LAN and SA
- Centralized control plane and distributed data plane
- Use only Commodity Ethernet switches
  - Army of commodity switches vs. few high-port-density switches
  - Requirements on switches: run fast and has programmable routing table
- Centralized load-balancing routing using real-time traffic matrix
  - Support for incremental and QoS-aware routing
- Fast fail-over using pre-computed primary/back routes
- Native support for network virtualization
  - Private IP address space reuse
  - Multi-tenancy VPN, NAT and traffic shaping
  - Intra-VDC or inter-VDC firewall

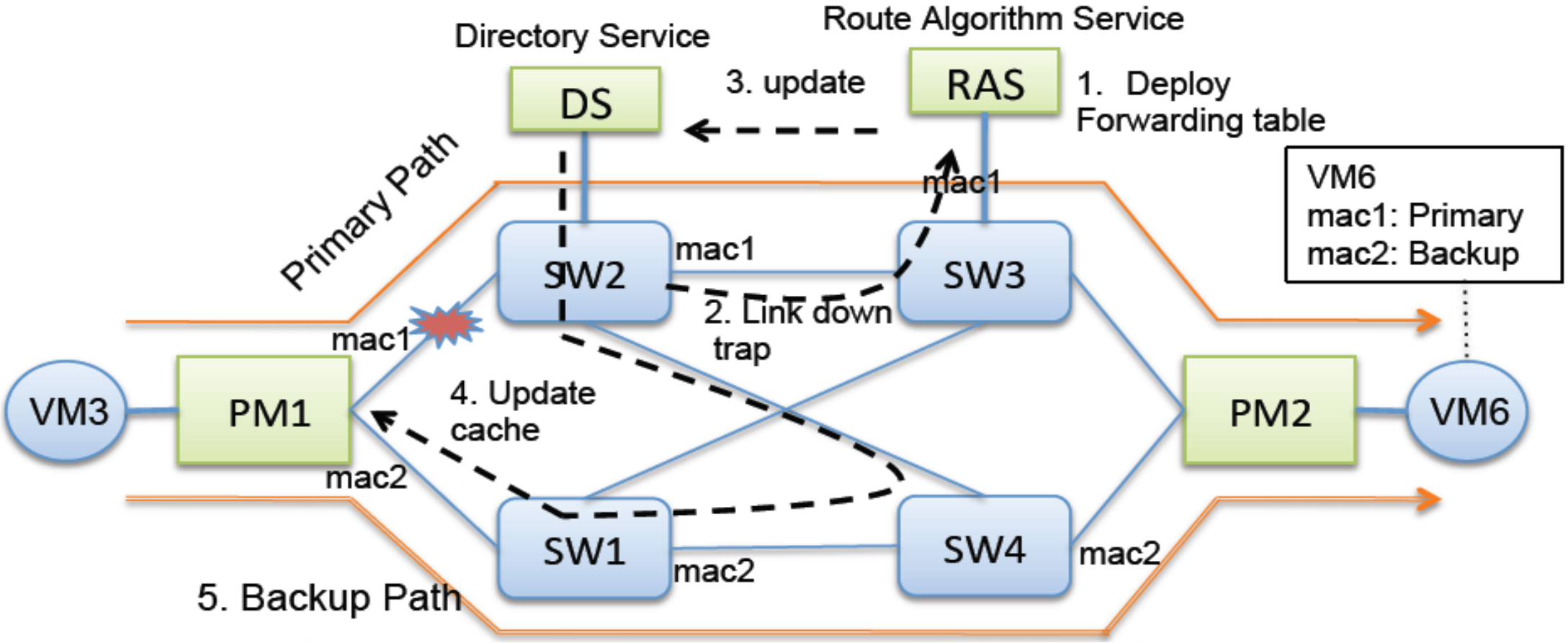
Software Architecture



Load Balancing Routing

- Collection of real-time traffic matrix
  - Traffic volume between each pair of VMs
  - Traffic volume between each pair of PMs
- Load balancing routing algorithm
  - Loads on the physical links
  - Number of hops
  - Forwarding table entries
  - Prioritization
- Computed routes are installed on switches

When a Network Link Fails



工業技術研究院
Industrial Technology Research Institute

Private IP Address Space Reuse

- Requirement: Every VDC has a VDC ID and its own full 24-bit private IP address space (10.x.x.x), even though multiple VDCs run on top of the same data center network
- Two approaches:
  - Ethernet over TCP/UDP:
    - Every Ethernet packet is encapsulated inside an TCP/UDP packet or TCP/UDP connection as an Ethernet link
    - Needs to implement in software such Ethernet switch functions as source learning, flooding, VLAN, etc.
    - Can work with arbitrary IP networks
  - Multi-tenancy-aware IP-MAC mapping: our approach
    - Runs directly on L2 networks, no need for Ethernet switch emulation
    - Inter-virtual-data-center isolation

Peregrine Summary

- Peregrine is a network system technology, not a network device technology, and consists of
  - A hypervisor module running on every compute node
  - A route server and an ARP server
  - A VDC-aware VPN
- Runs directly on commodity Ethernet switches and NICs: fully leverages the benefit of I/O virtualization, which encourages direct NIC access from VM
- Under development: Refactor Peregrine as a Quantum plug-in

Conclusion

- Cloud computing is all about consolidation of IT infrastructures and usage-based resource allocation
  - Data center as a computer paradigm
- Cloud-scale data center industry is emerging
  - Integration is a real user pain point
  - An integrated solution with lesser components is much more desirable than an un-integrated set of more capable components
- ITRI's integrated data center solution, Container Computer 1.0 + Cloud OS 1.0, is expected to provide 70% of the functionalities at 1/3 cost of leading solutions from US
  - Virtual data center service abstraction

Thank You!

Questions and Comments?

tcc@itri.org.tw