

# The concept of Virtual Data Center: Scalable Online Environments for Data Science

*Stefano Chiappini & Stefano Cacciaguerra*



**ISTITUTO NAZIONALE  
DI GEOFISICA E VULCANOLOGIA**



# Modern Research Needs



## Complexity and Interdisciplinarity

Modern research merges **multiple disciplines**, utilizing different data and methods.

## Need for Collaboration

Increasingly **complex questions** require **geographical distribution** and **disciplinary collaborations** to harness various expertises.

## Technological Advances

Improvements in **data** and **computational management** facilitate handling large datasets and complex simulations, supporting effective collaboration despite distances

## Data Sharing

Essential for **validating results**, enabling **new discoveries** through open data availability and allow others to **build on existing work**.

## Reproducibility Challenges

Critical for **scientific trust**, yet complicated by the complexity of modern research.





INGV implemented the **NEw REsearch Infrastructure Datacenter for EMSO** in Sicily, Portopalo di Capopassero (SR), Italy.

An **ICT infrastructure** for:

- archiving, processing, and sharing **scientific data**
- developing **services** for data science
- promoting **multidisciplinary research**



# Infrastructure design from beginning to end...

Budget  
(1.5 M€)

Data Center  
Design

Market Research  
HW Available

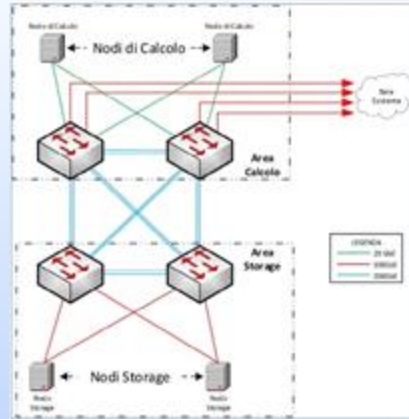
Technical  
Specification  
Document

Tender  
Process

HW Installation and  
firmware  
management

Network  
Configuration

Software  
Installation



## Computing:

48 nodes  
~2600 cpu cores  
36 TB ram

## Networking:

Internal VLT link  
4 x 400 Gbps

## Storage:

12 nodes  
240 HDD  
60 SSD  
3.6 PB raw disk



# The Architecture of NEREIDE



Open-source **Infrastructure-as-a-Service** platform. It enables seamless **scalability** and **flexibility** in **resource management**.

Workflow Automation with the integration of Metal-as-a-Service and Juju in the **deployment** and **maintenance** of physical and virtual resources.



Distributed **software-defined storage** engine, that offers **flexible** and **scalable storage** resources managing heavy data loads and ensuring both data security and integrity.



# The Concept of Virtual Data Centers



A **VDC** is a **managed environment** designed to facilitate **the control and operation of cloud resources** by DSs, including Virtual Machines (VMs), behind a **customizable gateway** with a dedicated **public IP address**.



VDC

## Customizable Gateway

Provides DSs with a unique, customizable **entry point**, enhancing security and personalization.

## Management levels

DSs possess **administrative privileges** over their **cloud segments**, enabling them to create, manage, and configure VMs, networks, **firewalls**, **VPNs**, and **routing** configurations autonomously.

The whole cloud infrastructure and physical data center are managed by **Infrastructure Administrators (IAs)**, maintaining a balance between autonomy and control.



## Infrastructure Administrators



**Architecture design and  
VDC creation**

## Data Scientists



**Data Services  
development inside VDC**

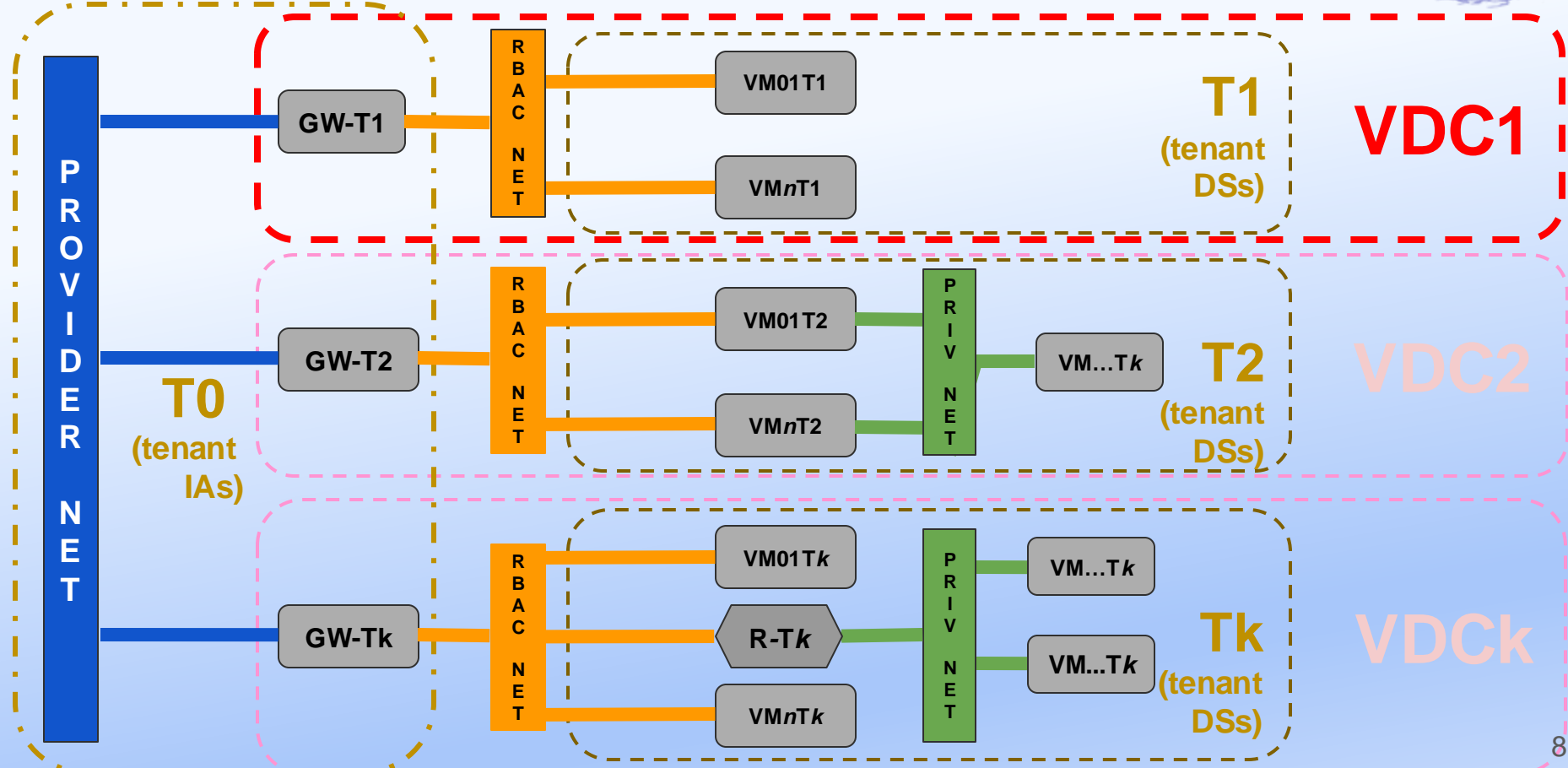
## End Users



**Use of  
Data Services**



# The Implementation of Virtual Data Centers

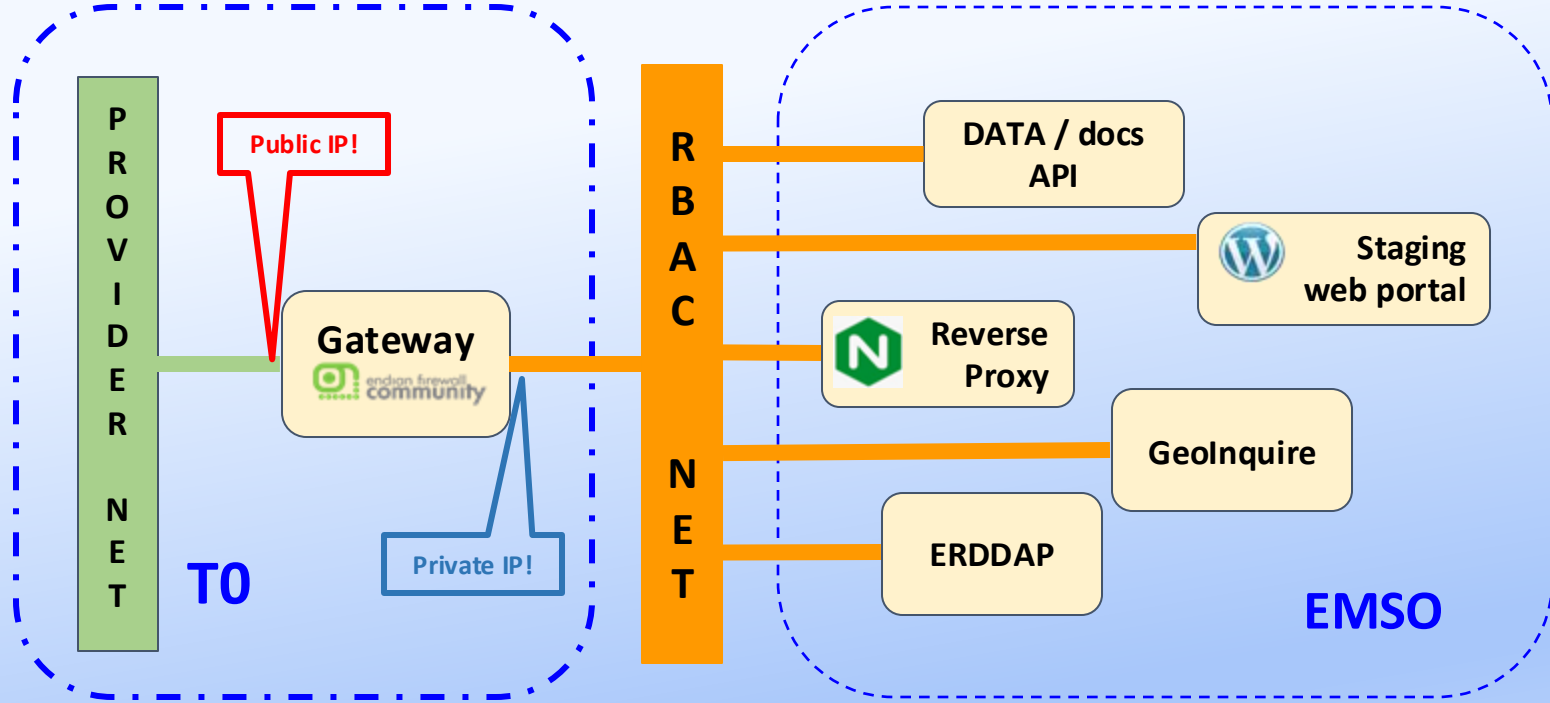




# VDC example: EMSO Services Migration



(project ANFITRITE)



# Endian Firewall as Gateway

It is an open-source **router**, **firewall** and **gateway** security Linux distribution

## Credentials of Endian Firewall:

- root of SO → Infrastructure Admins
- admin of Web Dashboard → Infrastructure Admins
- admin user of Web Dashboard → Data Scientists

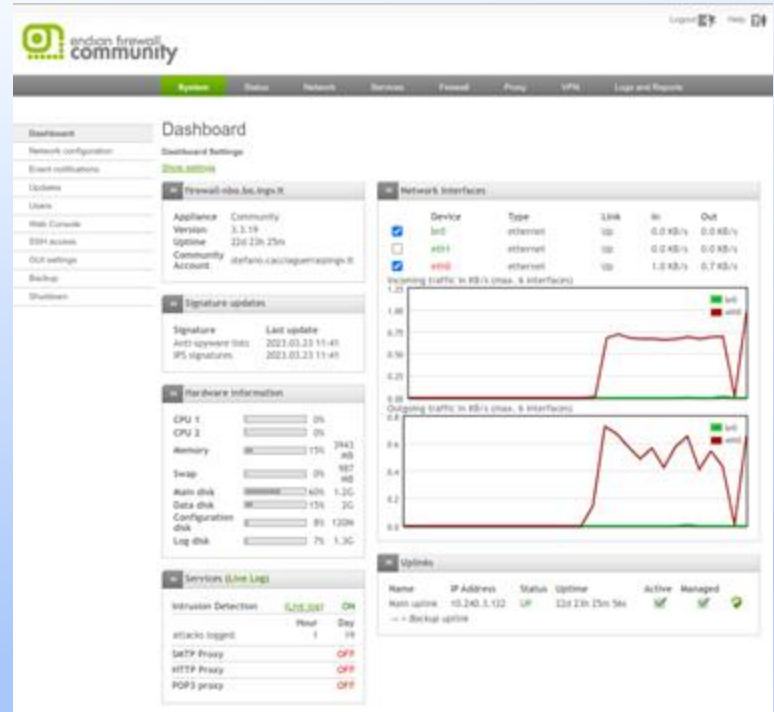
```
Release: Endian Firewall Community release 3.3.19
Product: Community (64 bit)
Hostname: firewall-nho

GREEN Zone
Management URL: https://172.16.0.1:10443
IPs: 172.16.0.1/24
Devices: eth1 (UP)

Uplink - wlan (ACTIVE)
IPs: 10.240.3.122/24 (STATIC)
Device: eth0 (UP)

0 Shell
1 Reboot
2 Change Root Password
3 Change Admin Password
4 Restore Factory Defaults
5 Network Configuration Wizard

















Choice: _
```



# Endian Firewall Services

## Port Forwarding / DNAT

to make **accessible** services from Internet

| # | Incoming IP                | Service  | Policy  | Translate to       | Remark | Actions   |
|---|----------------------------|----------|---|--------------------|--------|---|
| 1 | 10.240.3.122 (Uplink main) | TCP/8888 |  | 172.16.0.20 : 8888 |        |      |
|   | ALLOW with IPS from:       |          |   | <ANY>              |        |     |
| 2 | 10.240.3.122 (Uplink main) | TCP/8000 |  | 172.16.0.20 : 80   |        |      |
|   | ALLOW with IPS from:       |          |   | <ANY>              |        |     |

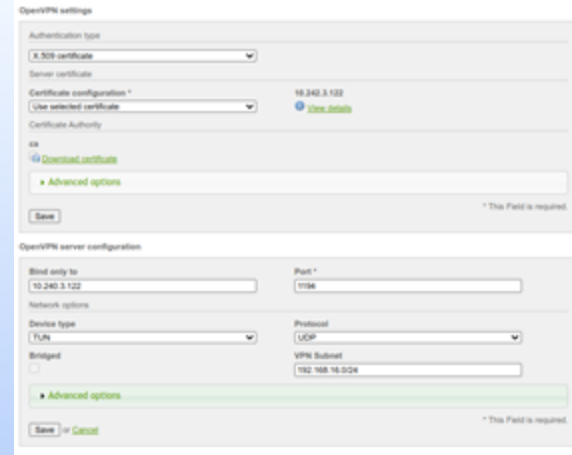
## Masquerading / SNAT

to allow VMs to **access Internet**

| # | Source | Destination | Service | NAT to | Remark               | Actions |
|---|--------|-------------|---------|--------|----------------------|---------|
| 1 | <ANY>  | Uplink ANY  | <ANY>   | Auto   | standard uplink SNAT |         |

## openVPN

to allow users to **access VMs**



The screenshot shows two configuration panels. The top panel, 'OpenVPN settings', includes fields for 'Authentication type' (set to 'x 309 certificate'), 'Server certificate' (set to '10.240.3.122'), and 'Certificate Authority'. The bottom panel, 'OpenVPN server configuration', includes fields for 'Bind only to' (10.240.3.122), 'Port' (1194), 'Device type' (TUN), 'Protocol' (UDP), and 'VPN Subnet' (192.168.16.0/24).

## Virtual Private Networking



The screenshot shows the 'Virtual Private Networking' settings. It includes a section for 'IPsec L2TP' with a toggle for 'Enable IPsec' (checked). Below that, the 'IPsec settings' section includes a field for 'Roadwarriors virtual IP (inner IP) pool'.

## IPsec

to allow VDC to **access other VDC**

# Conclusions



NEREIDE is a fully operational ICT infrastructure hosting EMSO services on a VDC (ANFITRITE project).

VDCs represent the **evolution** in the **concept of virtualization**, moving beyond individual virtual machines to include comprehensive data center infrastructures, **ready-to-use, fully equipped** with advanced tools.

NEREIDE allows easier and more **effective resource management** and **data sharing**, accelerating the achievement of **Open Science goals**, simplifying the complexities of modern research.

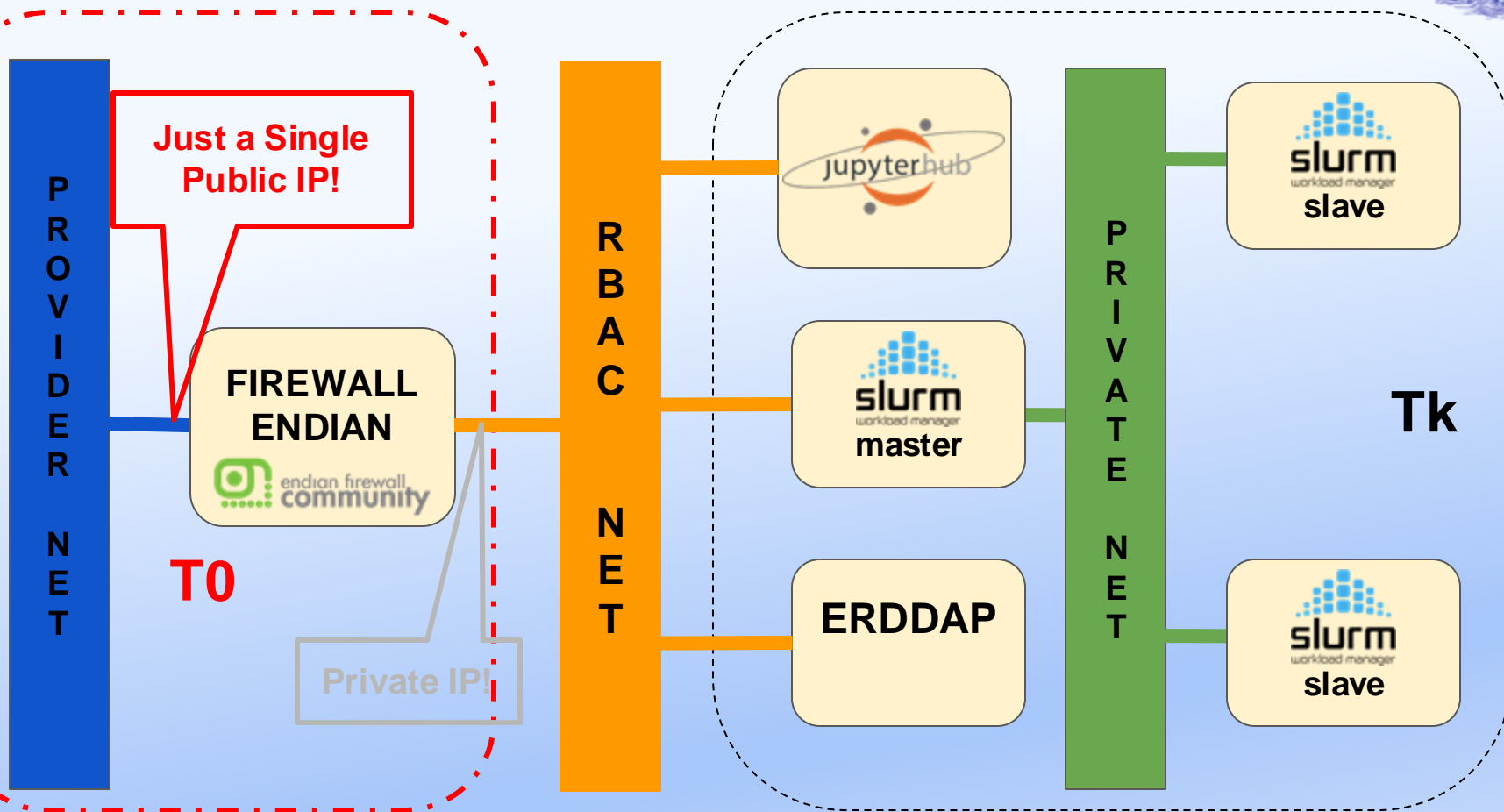


# Questions ?



**Stefano Chiappini & Stefano  
Cacciaguerra**

# Single Tenant on Openstack Networking



# Cloud-init



A tool for initializing and automatically customizing VMs:

- ❖ **Package installation, configuration management,** and **script execution**
- ❖ **Configurations** can be **reused** for deploying **multiple VMs**, significantly reducing the time and effort required for their setup.

IT works to possible into two different phases:

- ❖ **Local Boot:** Before the network is active
- ❖ **Late Boot:** After applying the network configuration



It is **native** in Openstack and responsible for **initializing a VM** during their **installation**,



# Bash, CLI and Cloud-init



Using a combination of Bash, CLI and Cloud-init, it is possible to obtain a **rudimentary cloud orchestration**

1. write a **Cloud-init YAML file**
2. with a **CLI command** assign a Cloud-Init script
3. use **bash script** to run more CLI commands

```
#cloud-config
package_update: true
package_upgrade: true
packages:
  - nginx
runcmd:
  - [ systemctl, enable, nginx ]
  - [ systemctl, start, nginx ]
```



1

```
openstack server create --image <image-id> --flavor <flavor-id> --key-name <keypair-name>
<...> --user-data /path/to/your/cloud-config.yml <instance-name>
```

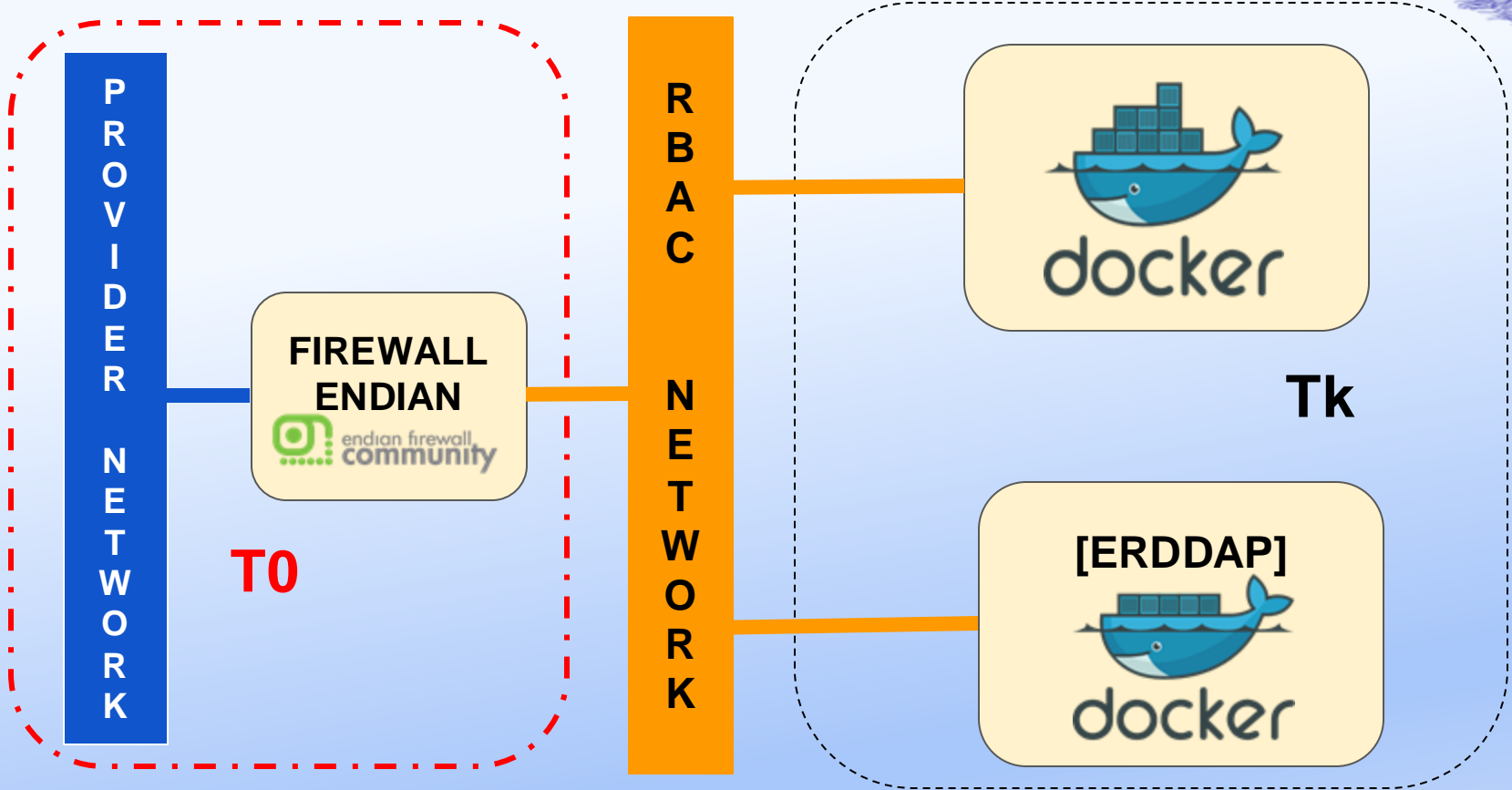
2

```
for (( i=1; i<=3; i++ ))
do
    openstack server create <...> --user-data ./cloud-init-slurm-slaves.yaml
    <...> v4-fixed-ip=172.16.0.2$i myslurmslave$i
done
```

3



# VDC - Docker



# VDC - SLURM

