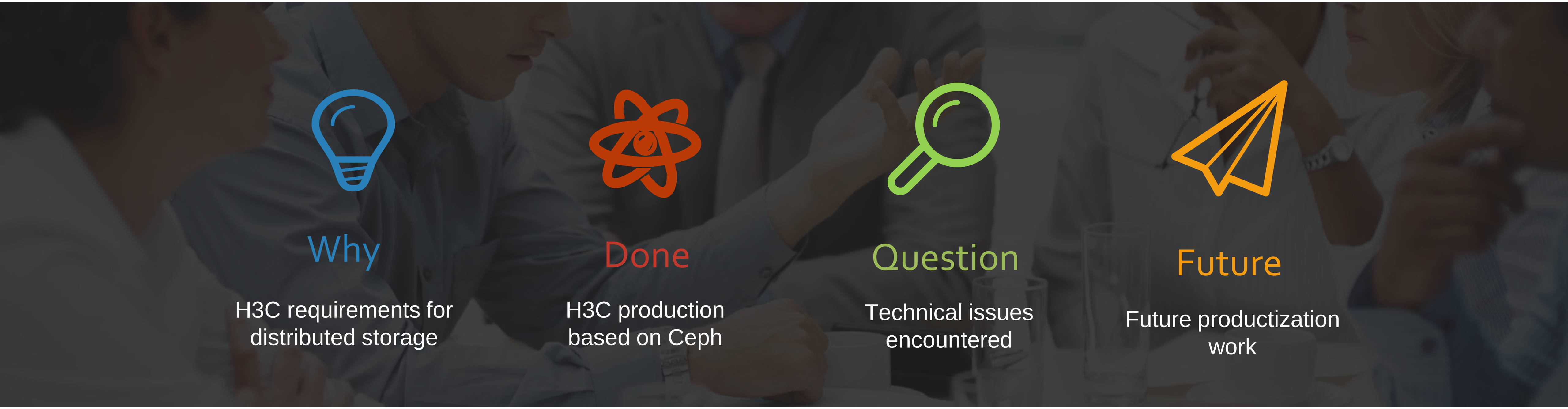
On the Productization Practice of Ceph

Hangzhou H3C Tech. Co. Ltd

Winter

wentao@h3c.com

Agenda



Why

H3C requirements for distributed storage

Done

H3C production based on Ceph

Question

Technical issues encountered

Future

Future productization work

WHY-Concerns of storage for Productization

Ease
Cost
Reliability
Maintainability
Availability

1 **Ease of use**
Easy-deployment
Easy-installation
Easy-configuration

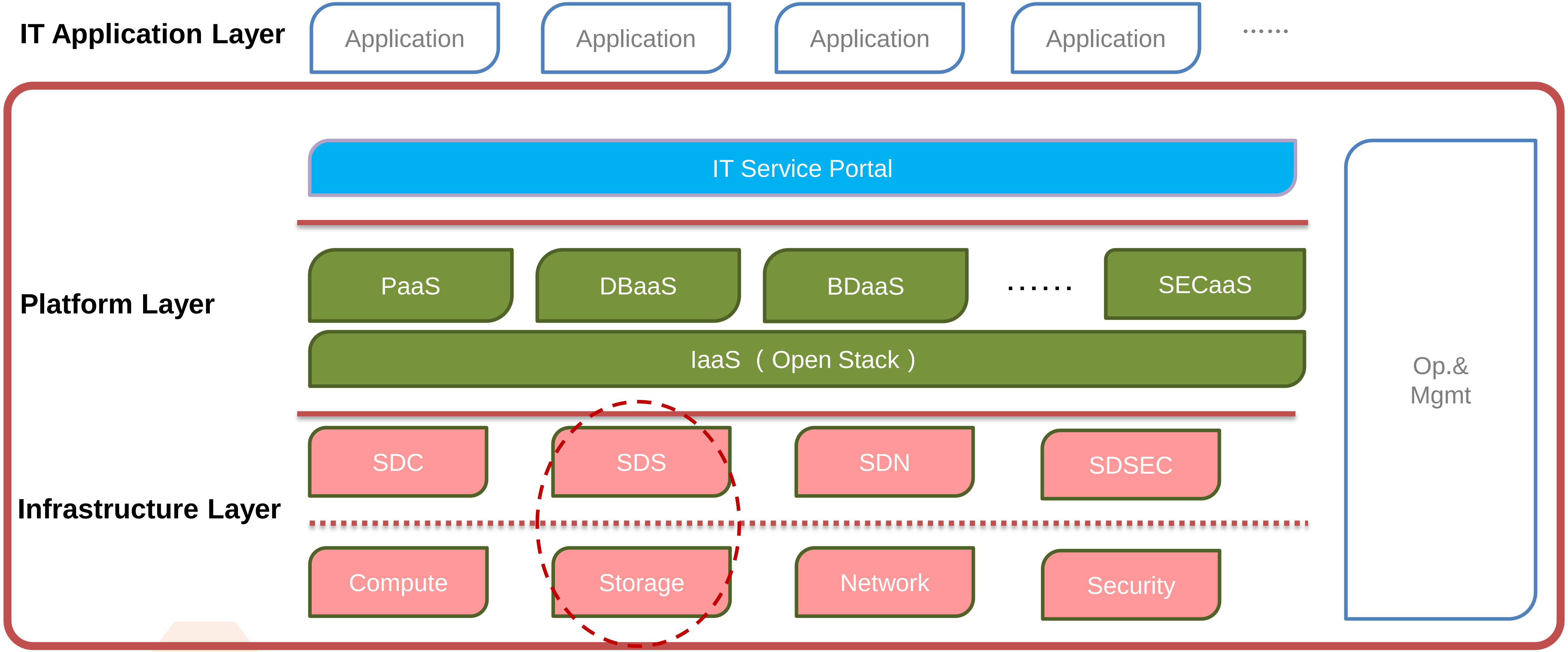
2 **Reliability**
No data loss

3 **Availability**
Provide normal service when cluster is shrunk or extended
Fast self-repair

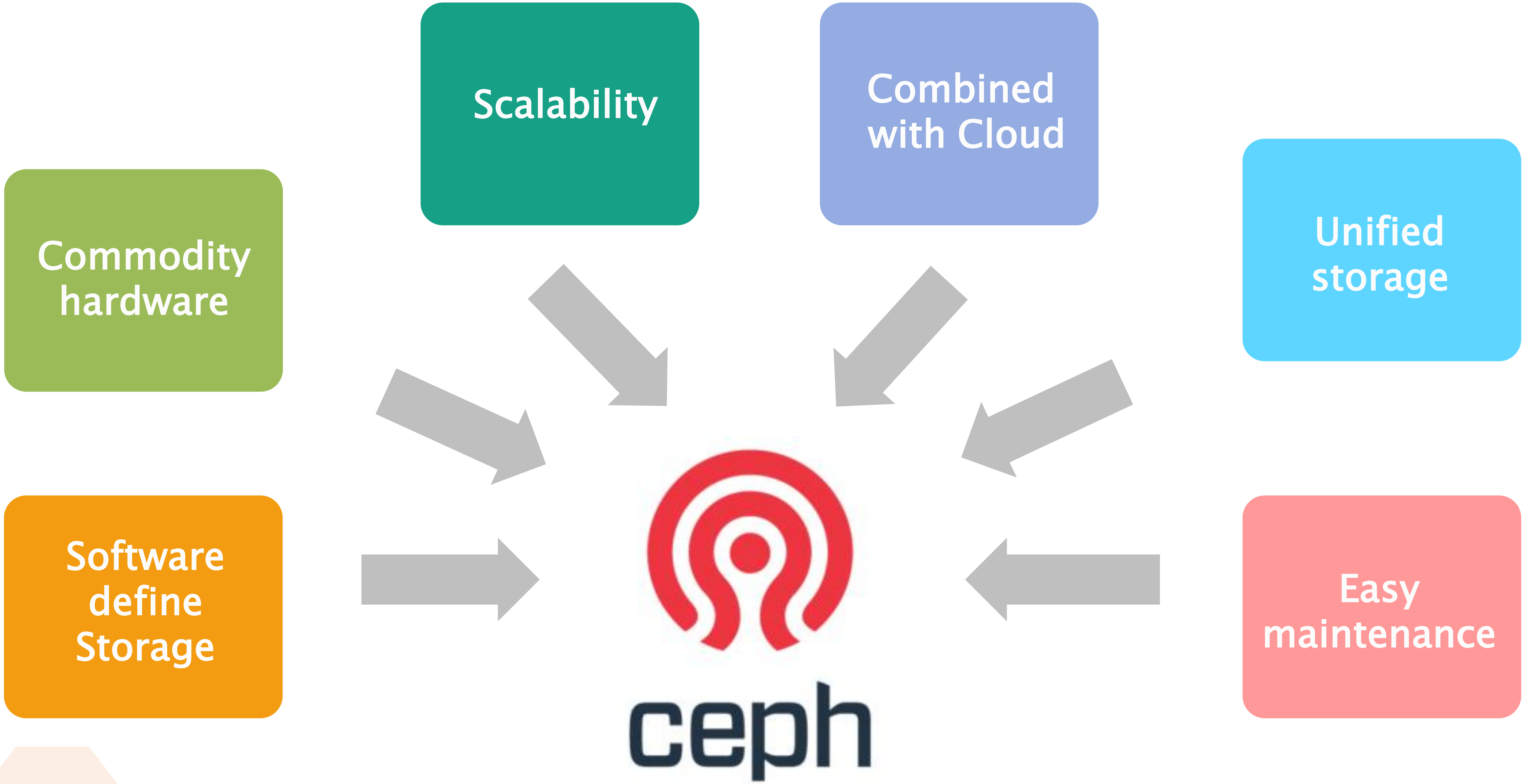
4 **Maintainability**
Variety of monitoring tools
Breakdown reported in time
Multi-levels log record

5 **Cost**
Commodity hardware
Low system resource required

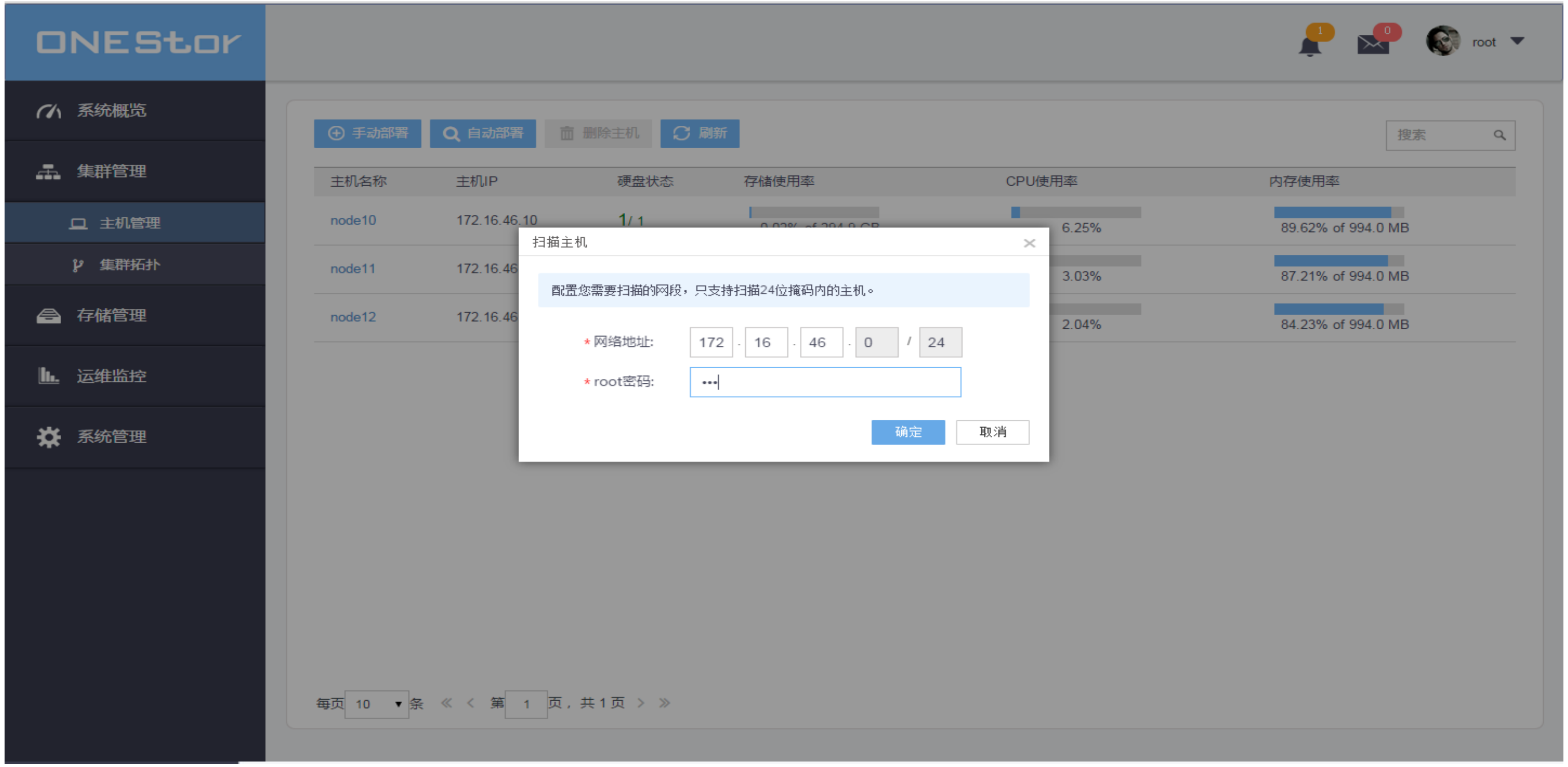
WHY-Enterprise market trend: integrated solution



WHY-Choice of Ceph



DONE- WEB UI and deployment



主机名称	主机IP	硬盘状态	存储使用率	CPU使用率	内存使用率
node10	172.16.46.10	1/ 1	[illegible]	6.25%	89.62% of 994.0 MB
node11	172.16.46			3.03%	87.21% of 994.0 MB
node12	172.16.46			2.04%	84.23% of 994.0 MB

扫描主机

配置您需要扫描的网段，只支持扫描24位掩码内的主机。

* 网络地址: 172 . 16 . 46 . 0 / 24

* root密码: •••

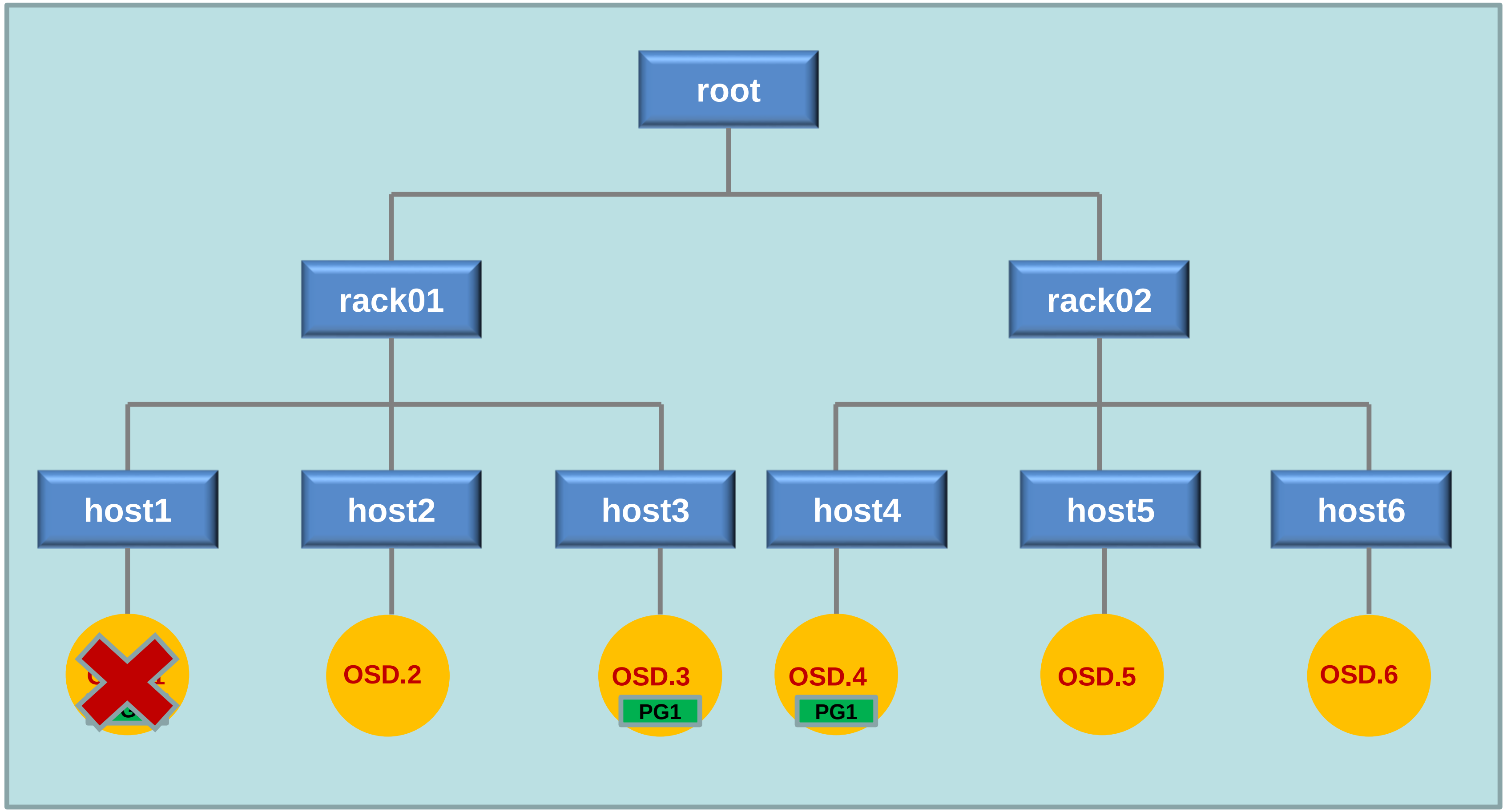
确定 取消

每页 10 条 « ‹ 第 1 页，共 1 页 › »

Design principles of UI

- Visualization：status of physical devices
- Classic scenario configuration：VDI, data backup, cloud drive, etc.
- Accessibility：one-click installation, automated deployment, log management, user management

QUESTION-CRUSH STRAW2 suboptimal solution



CRUSH Algorithm Suboptimal solution:

□ Preset configuration：

step take default

step chooseleaf firstn 0 type host

step emit

□ Operation：

Move one host out of crushmap

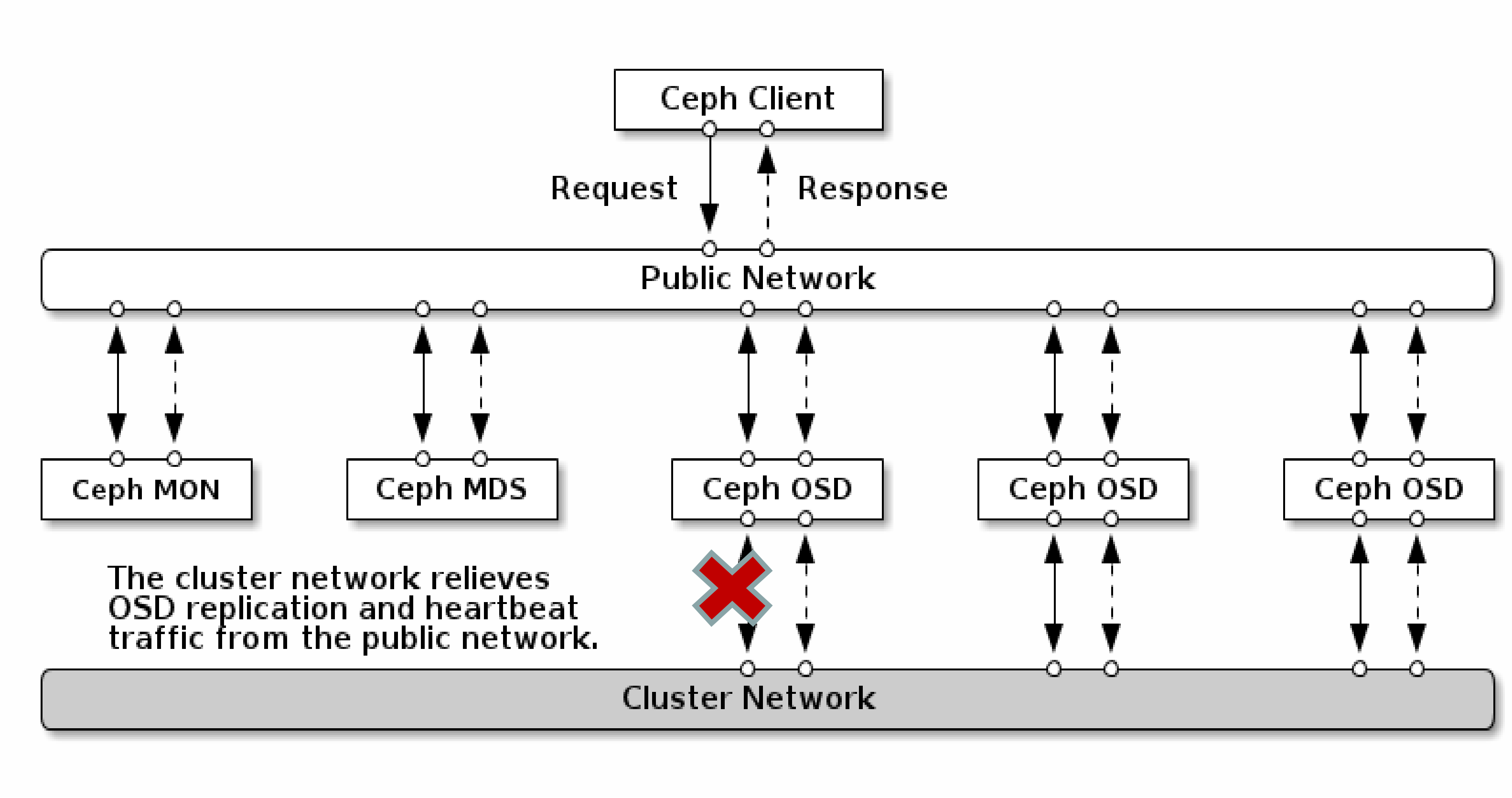
□ Expected result：

As the weight value adjusts, all the data and only the data that was originally located in the removed host is migrated to other hosts.

□ Actual result:

Extra but unnecessary migration occurs.

QUESTION-OSD flapping problem



CRUSH Algorithm Suboptimal solution:

□ Preset configuration：

Create Ceph Cluster with 3 nodes and each node has 10 OSDs.

Two networks: Public network and Cluster network.

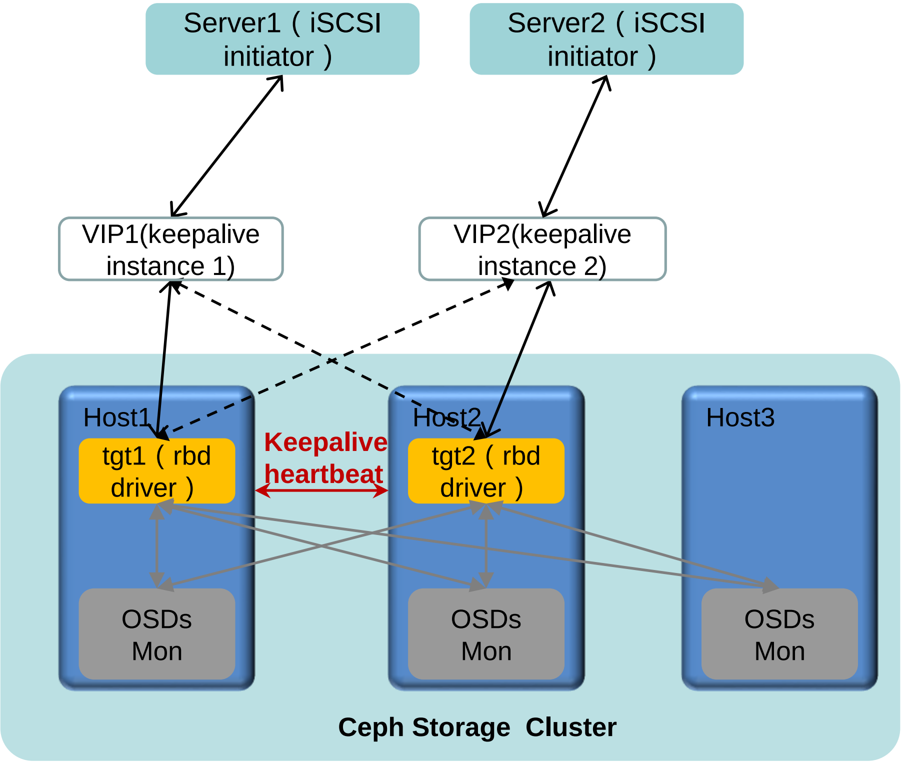
□ Operation：

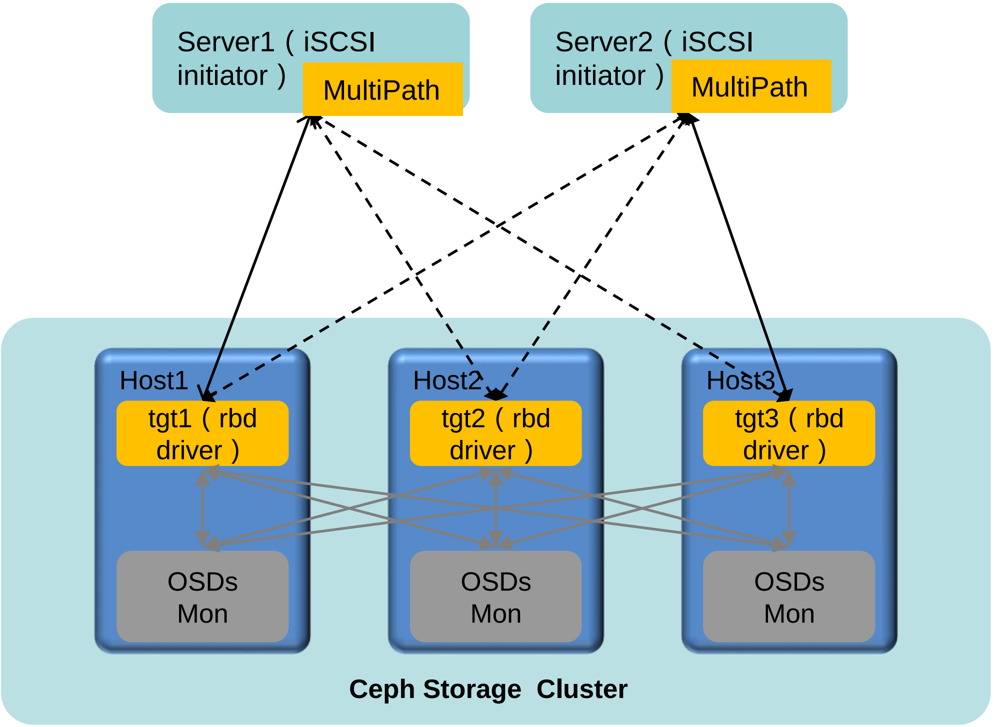
Unplug the Cluster network cable of one node

□ Result：

OSD status become flapping and will not stabilize eventually.

QUESTION-HA solution of iSCSI protocol in Ceph





Requirement：

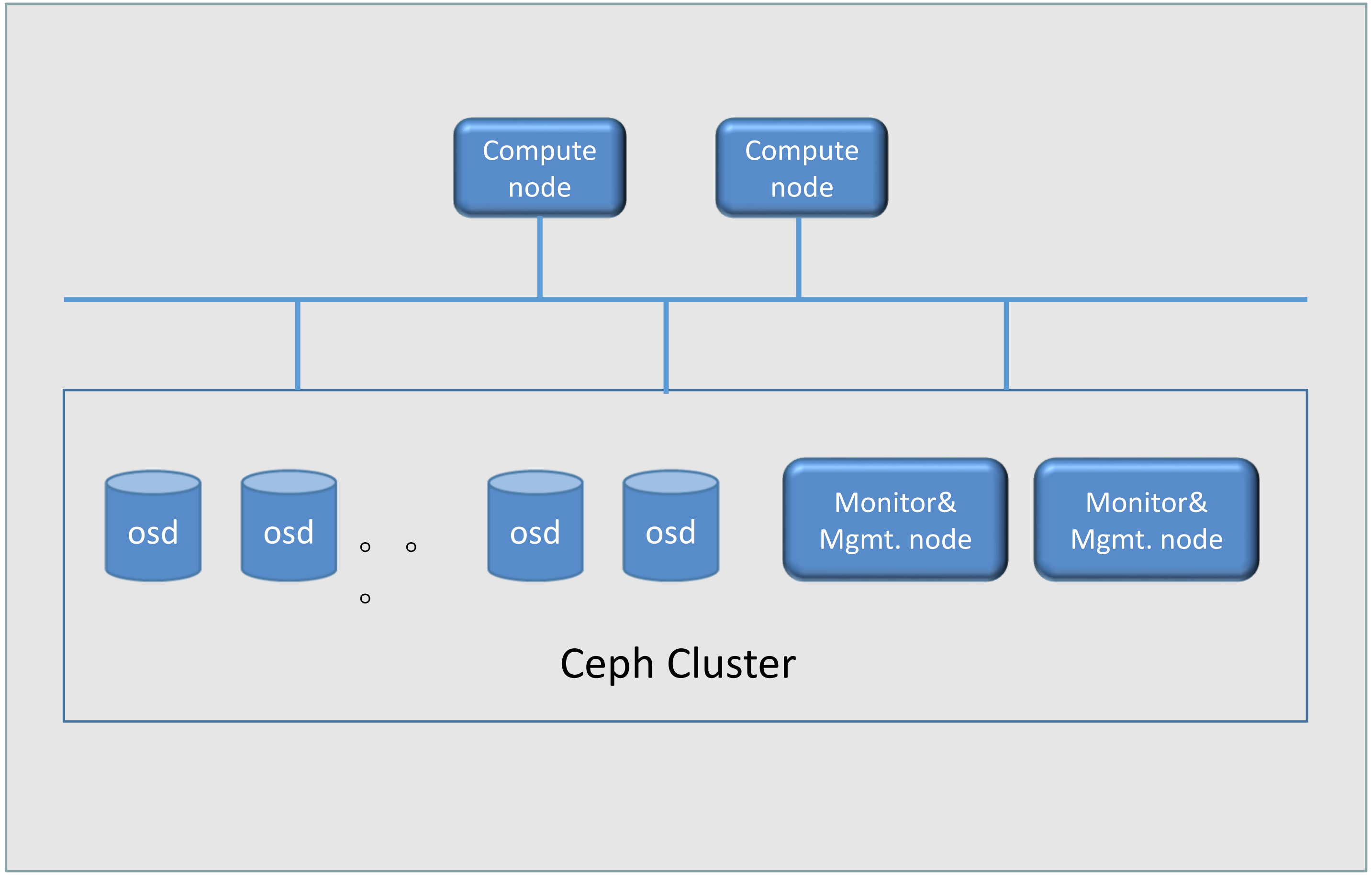
When used for iSCSI in Ceph, target module (such as the TGT) should support HA features.

Current solutions：

1,rebuilding target module:such as the use of existing open source software or solutions（Keepalive/LVS etc.). The disadvantage is: fault detection/switching time is long (at least of the order of several seconds), and the stability is not high.

2,rebuilding initiator modules： such as configuring multiple target IP address. The disadvantage is: client-side modification is not a universal solution.

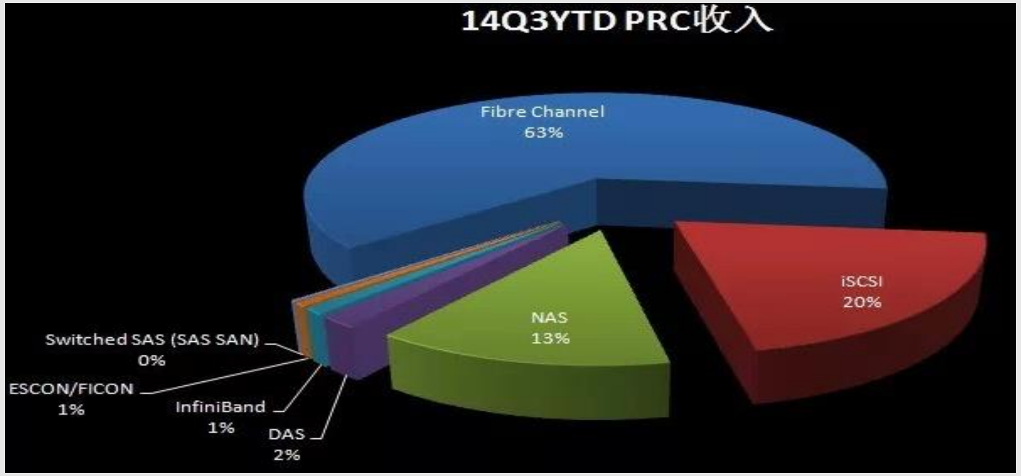
QUESTION-Two Monitor nodes issue



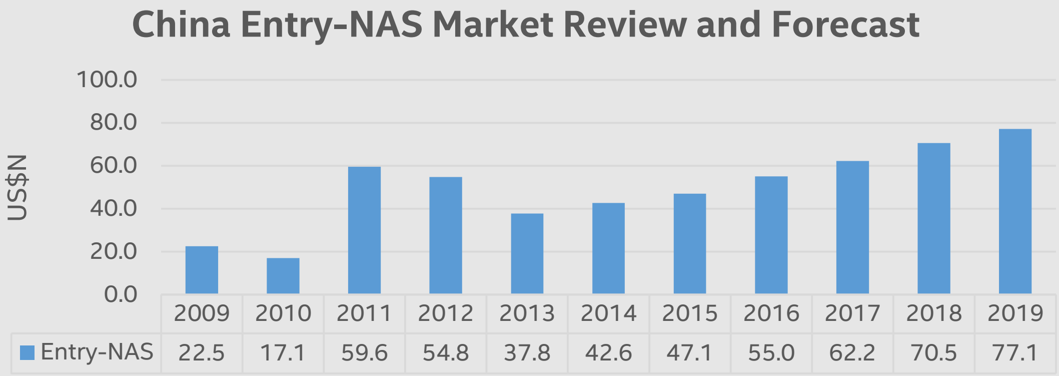
Description of issue:

- □ In certain scenario, some users use 2 dedicated physical hosts to deploy monitor and other management node together.
- □ Due to the quorum mechanism, when a monitor fails, the other one cannot work either.

QUESTION-Requirement for CephFS



14Q3YTD PRC收入



China Entry-NAS Market Review and Forecast

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
■ Entry-NAS	22.5	17.1	59.6	54.8	37.8	42.6	47.1	55.0	62.2	70.5	77.1

Requirement:

Extensive use of NAS devices in the enterprise network

Current solution：

RBD + NAS Controller

Scalability and efficiency are not good enough

Ideal solution：

Native CephFS

FUTURE-H3C future plan

OPEN

- Do as open-sourced
- Follow the open-sourced edition
- Take part in the community growing

FEEDBACK

- Contribute back to the community
- Focus on reliability, availability and maintainability

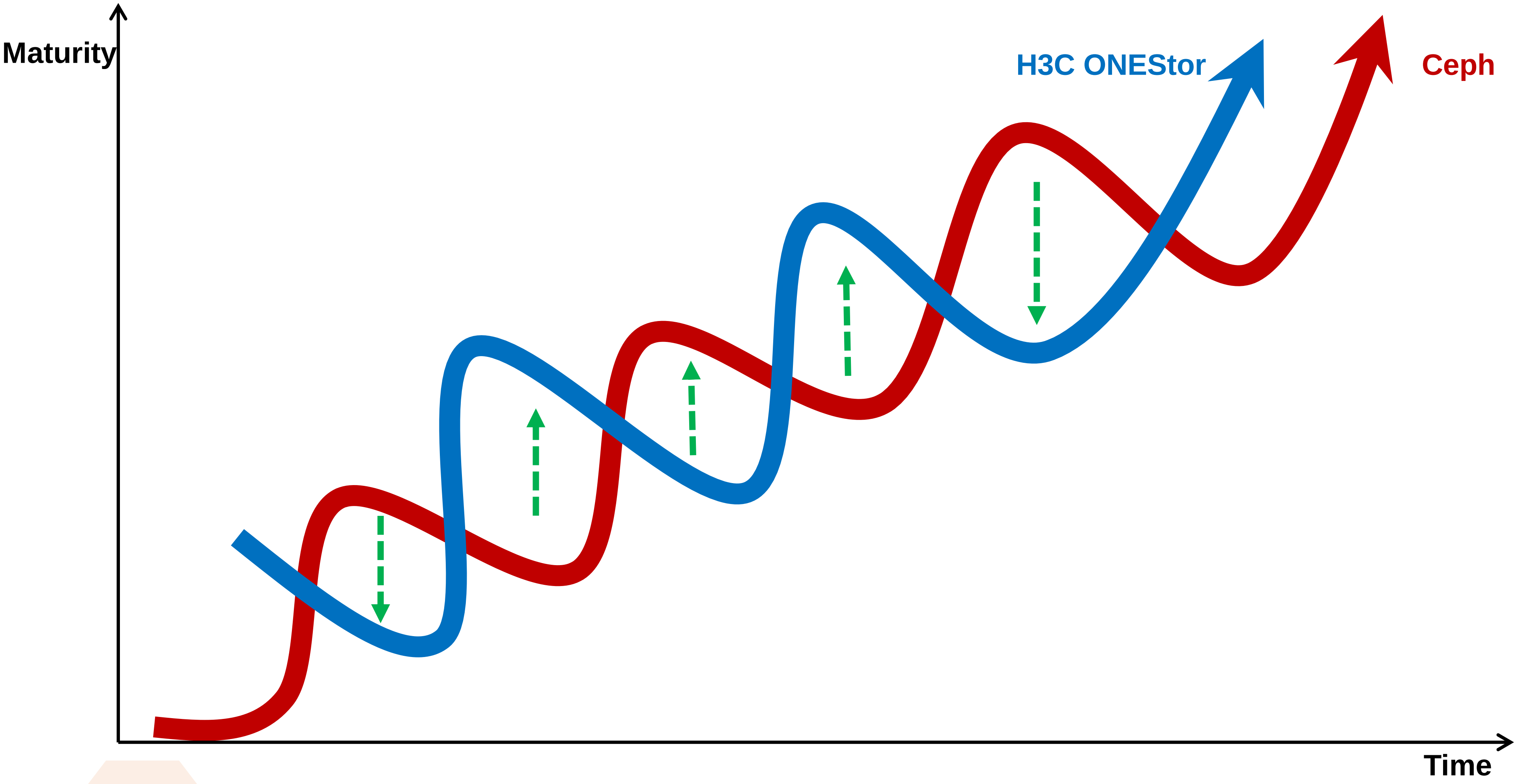
COOPERATE

- Cooperate with the community
- Cooperate with friend manufacturers

PRACTICE

- Collect issues and requirements from real enterprise customers to perfect Ceph
- Tests about hardware compatibility and system compatibility

FUTURE-Take From Ceph and Contribute to Ceph



H3C

IToIP解决方案专家

杭州华三通信技术有限公司

www.h3c.com.cn