

中国联通研究院的ceph应用实践

中国联通研究院

2015年10月

第一部分

联通沃云产品介绍

第二部分

Ceph集群在中国联通研究院的部署

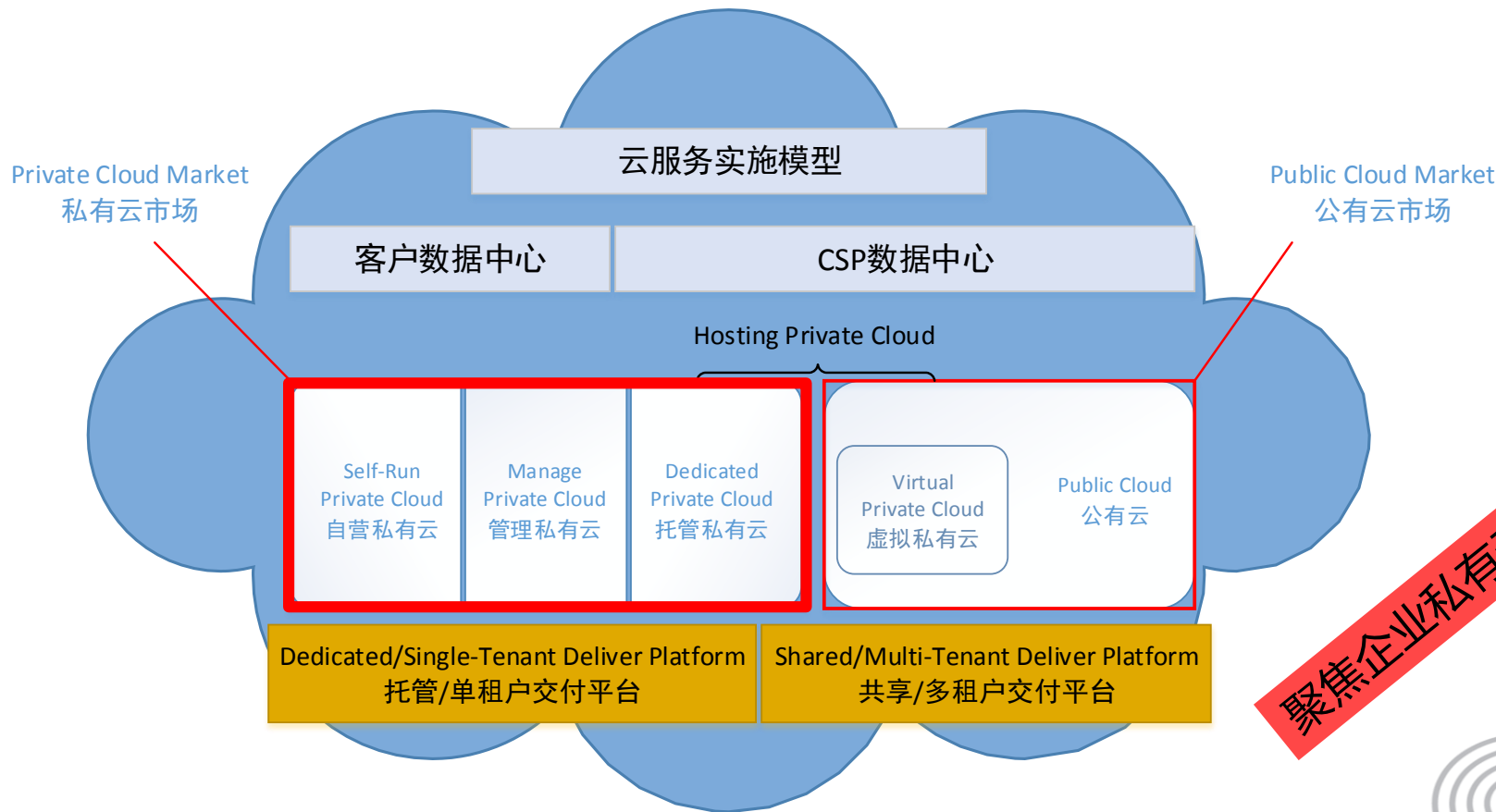
第三部分

中国联通研究院对ceph的测试优化

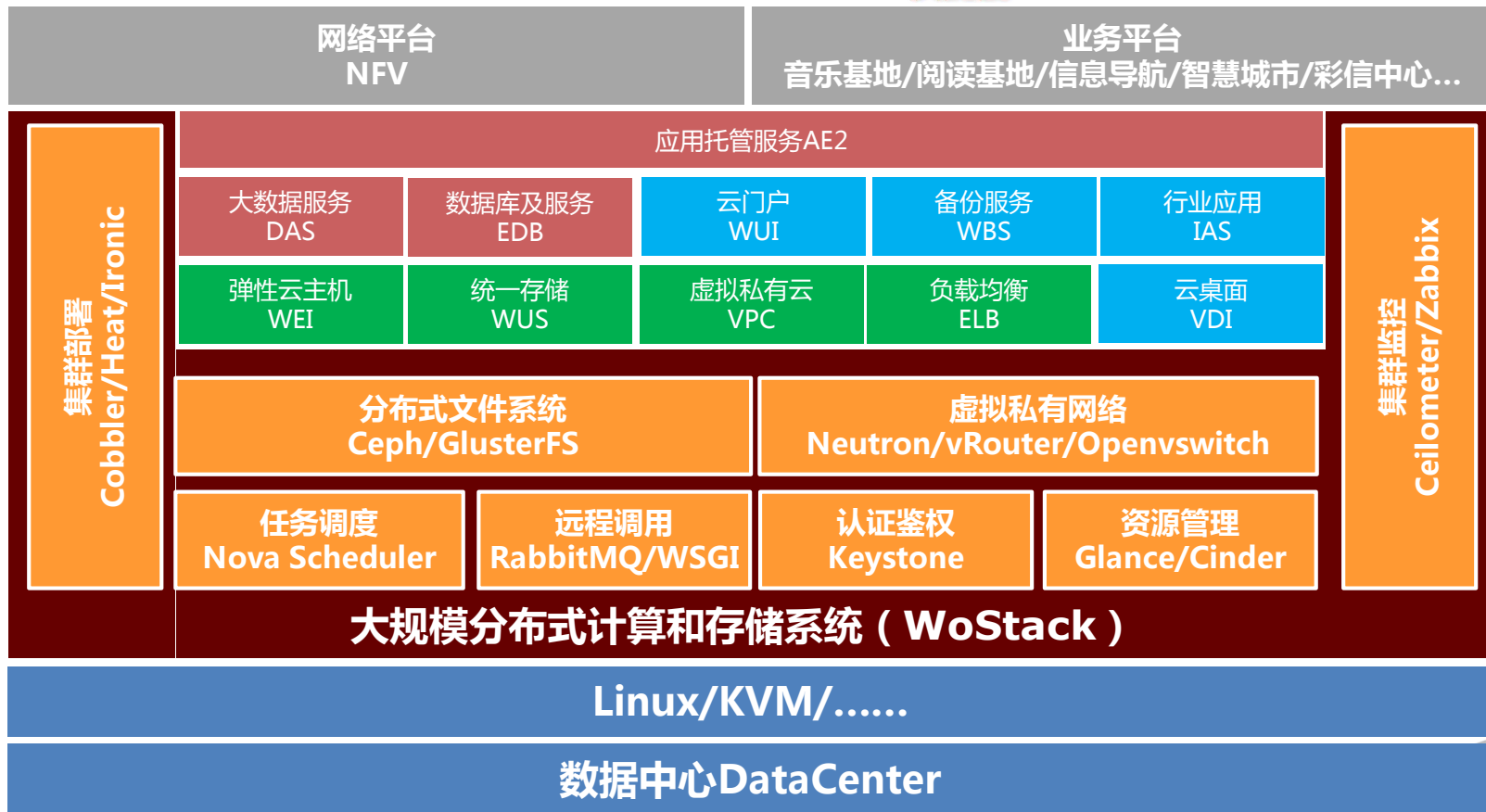
第四部分

联通沃云与ceph的对接简介

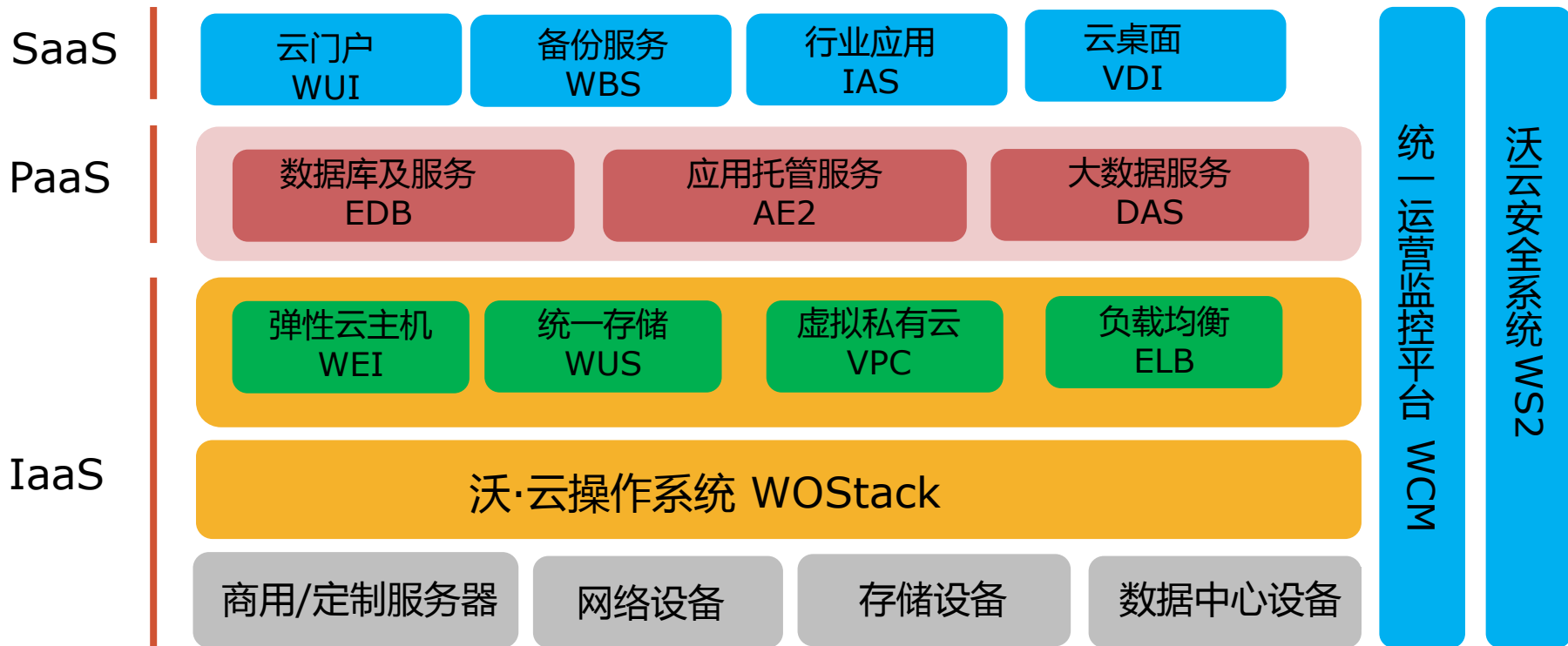
云计算服务模型的定位

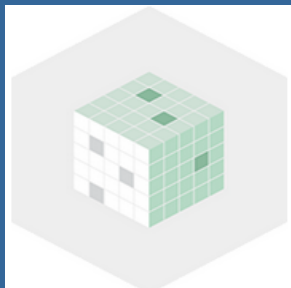


WoStack体系架构



沃云产品体系





弹性云主机

- 云服务器
- 按需分配
- 弹性伸缩
- 快速部署
- 在线迁移
- 磁盘快照



虚拟私有云

- 虚拟私有网络
- 隔离
- 网络解耦
- 灵活组网
- 自动配置
- VPN互联



负载均衡

- 四层和七层负载均衡
- 无单点故障
- 动态扩容
- 会话保持
- SLA
- 监控检查



WOSTack

- 一站式企业级私有云服
- 基于开源openstack定制化
- 云开源社区同步，迭代开发
- 卓越品质可靠运维
- 自主控制和独享资源
- 开源集成第三方设备与服务



弹性数据库
EDB



应用承载服务
AE2



数据分析云服务
DAS

- MySQL数据库即服务
- Oracle数据库即服务
- ▣ 特点：主流RDBMS支持，支持10000以上高并发请求，能力自动弹性伸缩
- 定制化承载环境（虚机模式）
- 自适应承载环境（容器模式）
- ▣ 特点：主流应用容器支持，部署-上线-管理-销毁全生命周期支持，能力自动弹性伸缩
- 大数据在线分析服务
- 弹性Hadoop集群服务
- ▣ 特点：支持10TB以上数据在线分析，支持用户自定义元数据，支持Hadoop集群构建与管理



对象存储服务

- 互联网云存储服务
- PB级海量数据存储
- 低价格易使用易管理
- 兼容S3/Swift接口
- 多副本保证数据安全
- 提供异地备份



块存储服务

- TB级块存储服务
- 支持物理机/虚拟机挂载
- iSCSI/Qemu/RBD等多种接口
- 配置简单/动态扩容
- 支持快照功能



文件存储服务

- TB级共享文件存储服务
- 提供账号/权限管理能力
- 支持NFS/CIFS/SMB接口
- 支持配额管理
- 支持动态扩容



文件备份服务

- 支持本地备份和云备份方式
- 支持全量/增量/差异备份
- 提供备份策略管理
- 支持多历史版本管理
- 支持本地备份+异地容灾方案

智慧政务

电子公文传输

大舆情服务

通用办公系统

政务一号通

市民邮箱系统

政务微博

工程领域监管

短信平台系统

政务公开

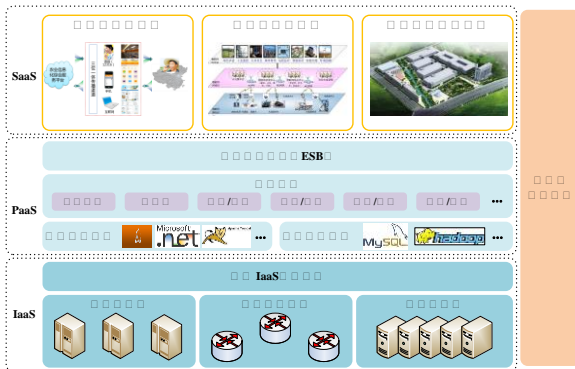
文字防护系统

自助建站

政务公开和服务

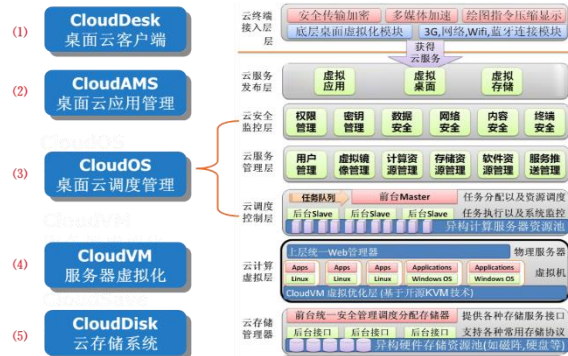
- ❑ 以使用部门的视角，设计满足电子政务公共平台需求的各项服务功能和服务产品。
- ❑ 将各部门应用效果较好、通用性较强的业务应用软件，经统一规划、设计和开发并部署在电子政务公共平台上，形成**12**个SaaS应用服务，供各部门直接使用。

智慧农业



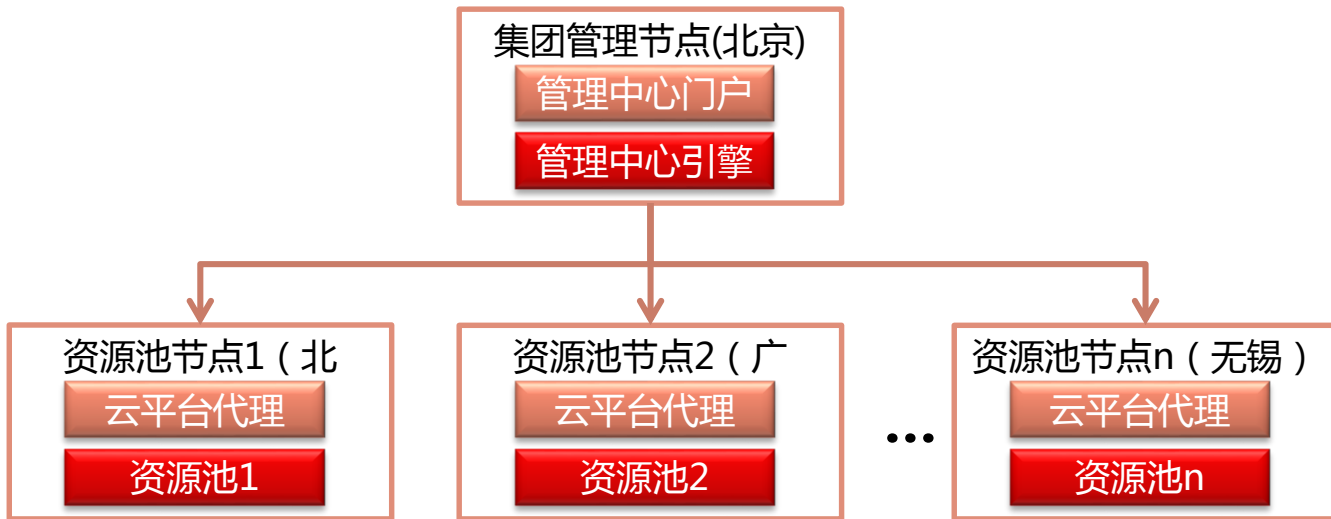
- ❑ **农机通**面向农机管理者和农机手，提供农机跨区作业调度服务。农民也可共享、共用农机服务信息。
- ❑ **沃农**作为农业部农技推广信息服务平台的延伸，不仅能为广大农技推广人员提供农业信息服务，同时可为广大农民直接提供各项农业信息服务。

桌面云



- ❑ TCO成本较传统PC模式低30%
- ❑ 自主研发，联通自有
- ❑ 自主掌控，信息安全可控，不受限于厂商
- ❑ 降低成本，无license成本
- ❑ 商用产品成本高，通过低成本开源产品，迫使商用产品降价

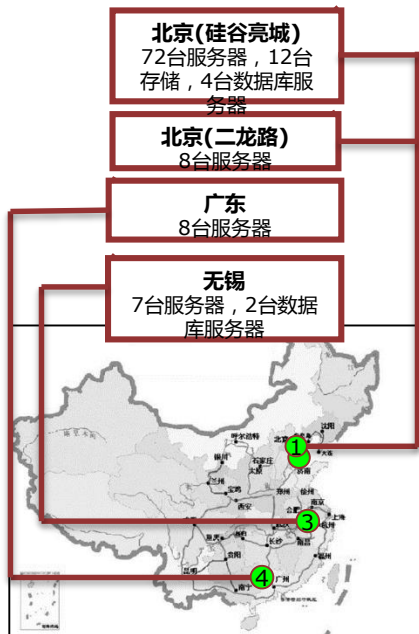
- ❑ **“一级架构，统一管理”**:云平台方案采用逻辑集中的一级架构，多地资源池独立运维，北京实现一级架构云平台管理模式，管理中心门户是统一对外展示云门户。
- ❑ **“集中调度，就近服务”**:管理中心引擎负责封装从资源池提供出来的接口，对全局资源进行统一调度，为客户请求就近选择最优化资源节点提供服务。



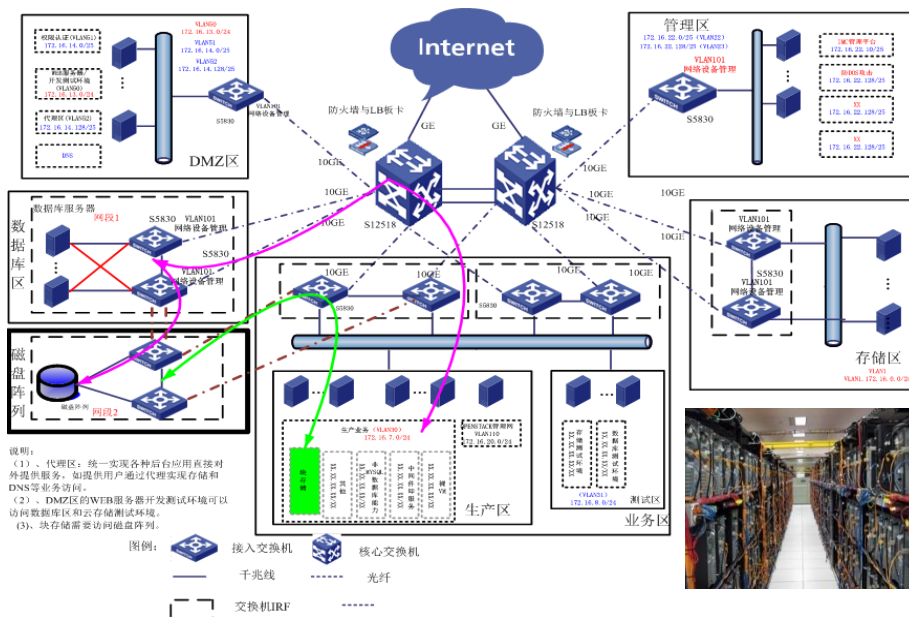
部署现状

- 95台x86服务器，12台存储服务器，6台数据库服务器，北京广州无锡三地四中心多活
- 支撑集团各部门、各基地、分子公司、研究院共35项业务，累计运行87.6万CPU小时

全国多中心部署



硅谷亮城主资源池拓扑



10个机架, 100M带宽

35项业务稳定运行3年以上，实践证明沃云系统性能与可靠性

产创部各基地创新业务

业务名称		调用沃云能力
悦云		计算/数据库/云存储
Wo+	通行证	云存储
沃商店	社区插件商用环境	计算/数据库/云存储
	社区插件开发测试环境	计算/数据库
	内部系统开发测试环境	计算/数据库
	大搜索的测试环境	计算/数据库
沃阅读	云终端环境	计算/数据库
		计算
宽带应用	沃游戏	计算/数据库/云存储
	视频	计算/数据库/云存储
联通时科新乐媒(超信)	个性化推送系统	计算
		计算/云存储

- 截止目前，研究院自身以及宽带公司、时科公司、沃商店、沃阅读等已有超过35项业务的部分子系统承载试点至云平台。
- 沃商店：
 - ◆ 插件系统：80万用户，部署到上线时间1天
 - ◆ 日志系统：日收集80万条日志数据，部署到上线时间1天
 - ◆ 沃游戏：累计470万用户，日活跃用户30万，部署到上线时间为3天
- 悦云：
 - 累计40W用户，数据库近5T的数据；部署到上线时间为1天
- 农业云：累计5万用户，数据库近25G的数据；

研究院自有业务

业务名称	调用沃云能力
市场中心-精细化营销分析	计算/数据库/存储
移动互联网-能力开放平台	计算
移动互联网-统计分析平台	计算/数据库/存储
DM	计算/数据库/存储
大容量智能卡	计算/数据库/存储
号码管理系统	计算/数据库/存储

第三方业务承载试点

业务名称	调用沃云能力
云联盟大学生创意大赛	计算
安徽农业云平台	计算/数据库
视频监控	计算
问卷调查系统	计算
木疙瘩	计算/云存储

第一部分

联通沃云产品介绍

第二部分

Ceph集群在中国联通研究院的部署

第三部分

中国联通研究院对ceph的测试优化

第四部分

联通沃云与ceph的对接简介

云存储部署情况

云存储管理平台

切换集群

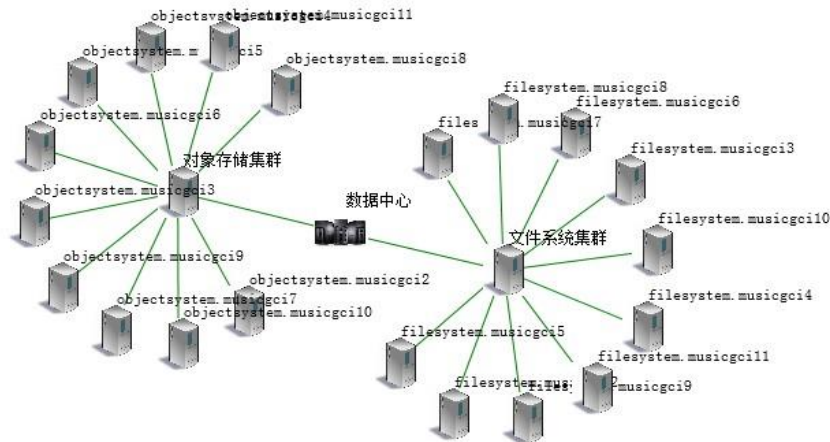


集群管理 对象存储 文件系统 系统管理

- 集群状态
- 主机列表
- 指标监控
- 数据中心

点击集群名称查看状态

	集群名称	集群状态	总容量	已用量	可用量
1	对象存储集群	HEALTH_OK	298TB	3086GB	295TB
2	文件系统集群	HEALTH_OK	298TB	3086GB	295TB



集群状态概述objectsystem1

[返回集群列表](#)

集群状态	HEALTH_OK
存储节点	6个
监控节点	3个
总容量	98238MB

集群存储使用情况

集群存储容量使用情况



集群主机物理状态

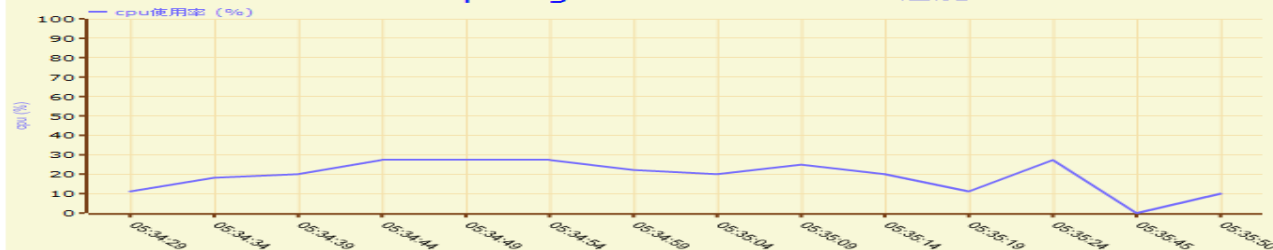
选择主机

s3ceph2.gcis3test.com
s3ceph2.gcis3test.com
s3ceph3.gcis3test.com
s3ceph1.gcis3test.com

网络状态

磁盘状态

s3ceph2.gcis3test.com CPU性能



集群状态

主机列表

配置信息

对象存储集群主机列表

新增主机

序号	主机名	已安装服务	操作
1	s3ceph2.gcis3test.com	osd存储服务X2 mon监控服务X1	增加硬盘 移除硬盘 从集群删除
2	s3ceph3.gcis3test.com	osd存储服务X2 s3网关服务X2 mon监控服务X1	增加硬盘 移除硬盘 从集群删除
3	s3ceph1.gcis3test.com	osd存储服务X2 mon监控服务X1	增加硬盘 移除硬盘 从集群删除

文件系统集群主机列表

新增主机

序号	主机名	已安装服务	操作
1	fsceph1.gcis3test.com	osd存储服务X2 mon监控服务X1 samba文件服务X2	增加硬盘 移除硬盘 从集群删除
2	fsceph2.gcis3test.com	osd存储服务X2 mon监控服务X1 samba文件服务X2	增加硬盘 移除硬盘 从集群删除

账户管理

账户标识: my

账户状态: 已启用

查询

重置

账户列表

新增 删除 修改 启用 停用 获取访问信息

	账户标识	账户名称	邮箱地址	账户状态	
<input type="checkbox"/>	mytestuser	abc		已启用	
<input type="checkbox"/>	myuser	myuser		已启用	
<input type="checkbox"/>	myuser2	myuser2		已启用	

对象列表

账户列表

对象列表

对象列表

10 每页记录数

Search:

序号	账户标识	对象名称	对象类型	对象大小	最后修改时间
1	42284	20131120225259.png	文件	230635bytes	2013-11-22T06:56:04.000Z
2	42284	DSC00268.JPG	文件	3593034bytes	2013-11-22T06:56:13.000Z
3	mytestuser	20110705162641934.doc	文件	34304bytes	2013-11-21T10:04:42.000Z
4	mytestuser	CreativeCloudSet-Up.exe	文件	3867000bytes	2013-11-21T10:05:34.000Z
5	mytestuser	ali_020.gif	文件	9596bytes	2013-11-21T10:03:50.000Z
6	mytestuser	blazeds-turnkey-4.0.0.14931.zip	文件	45842358bytes	2013-11-21T10:05:29.000Z
7	mytestuser	boto-latest.txt	文件	1290bytes	2013-11-20T10:16:51.000Z
8	mytestuser	hello.txt	文件	12bytes	2013-11-20T08:45:26.000Z
9	mytestuser	下载说明.txt	文件	2571bytes	2013-11-20T10:33:57.000Z
10	42341	VMware-workstation-full-8.0.2-591240.exe	文件	491232944bytes	2013-11-21T03:12:35.000Z

文件存储系统用户

文件系统用户功能菜单

用户名:

wu

查询

新增用户

文件系统用户列表

序号	用户名	空间大小	已用空间	文件数量	用户状态	访问方式	操作
1	wu1	1024M	4K	1	在用	\\10.122.129.193\wu1	修改配额 停用用户 删除
2	wu2	120M	1K	1	在用	\\10.122.129.193\wu2	修改配额 停用用户 删除
3	wu3	12288K	1K	1	在用	\\10.122.129.193\wu3	修改配额 停用用户 删除
4	wu4	12288K	1K	1	在用	\\10.122.129.191\wu4	修改配额 停用用户 删除

复制该访问方式粘贴到windows系统的文件资源管理器地址栏，再输入用户名和密码访问即可进行文件的存取操作

序号	用户名	空间大小	已用空间	文件数量	用户状态	访问方式	操作
21	test	123M	8K	2	在用	\\10.1.9.129\test	修改配额 停用用户
22	test123	256M	1K	1	在用	\\10.1.9.129\test123	修改配额 停用用户
23	testJohnny1	120M	1K	1	在用	\\10.1.9.129\testJohnny1	修改配额 停用用户

用户配额修改

修改用户配额界面

温馨提示：【8bit(位)=1Byte(字节)】【1024Byte(字节)=1KB】【1024KB=1MB】【1024MB=1GB】【1024GB=1TB】【1024TB=1PB】
【“修改之后的用户空间大小”必须大于“原空间大小”】

用户名称：test123 原空间大小：256 MB

新的用户空间大小

256

（单位：MB）

修改用户

启用/停用用户

文件系统用户列表

序号	用户名	空间大小	已用空间	文件数量	用户状态	访问方式	操作
21	test	123M	8K	2	已停用	\\10.1.9.129\test	修改配额 启用用户
22	test123	256M	1K	1	在用	\\10.1.9.129\test123	修改配额 停用用户
23	testJohnny1	120M	1K	1	在用	\\10.1.9.129\testJohnny1	修改配额 停用用户

用户管理

- 用户管理
- 日志信息
- 告警信息
- 接口调用记录

用户ID: 用户名: 密码: 用户权限:

用户列表

用户ID	用户名	用户权限	最近登录时间	最近登录ip地址
1	admin	管理员	2013-11-27 16:30:40	0:0:0:0:0:0:1
2	lichhao	普通用户		
4	james	管理员		
6	test1	普通用户		
7	chlk	管理员	2013-11-25 18:03:26	0:0:0:0:0:0:1

日志信息

- 用户管理
- 日志信息
- 告警信息
- 接口调用记录

日志内容: 日志类型: 操作用户登陆IP:

日志列表

日志表ID	日志内容	日志类型	用户登陆IP	用户ID	用户名称	创建时间
1	用户【admin】退出系统	退出	127.0.0.1	1	admin	2013-11-04 16:09:11
2	用户【admin】登陆系统	登陆	127.0.0.1	1	admin	2013-11-04 16:09:13
3	用户【admin】登陆系统	登陆	127.0.0.1	1	admin	2013-11-04 16:16:55
4	新增文件系统用户操作，新增用户信息: UserId: 5	新增	127.0.0.1	1	admin	2013-11-04 16:19:53
5	停用文件系统用户: 用户名: test	停用用户	127.0.0.1	1	admin	2013-11-04 16:20:15
6	启用文件系统用户: 用户名: test	启用用户	127.0.0.1	1	admin	2013-11-04 16:20:19
7	删除文件系统用户操作: 删除的实体对应的ID【5】	删除	127.0.0.1	1	admin	2013-11-04 16:20:32
8	线程【Thread-82-36719】执行结束，详情: 线程	删除	127.0.0.1	1	admin	2013-11-04 16:20:33
9	用户【admin】登陆系统	登陆	127.0.0.1	1	admin	2013-11-04 16:37:12
10	用户【admin】登陆系统	登陆	10.122.129.71	1	admin	2013-11-04 17:22:33

第一部分

联通沃云产品介绍

第二部分

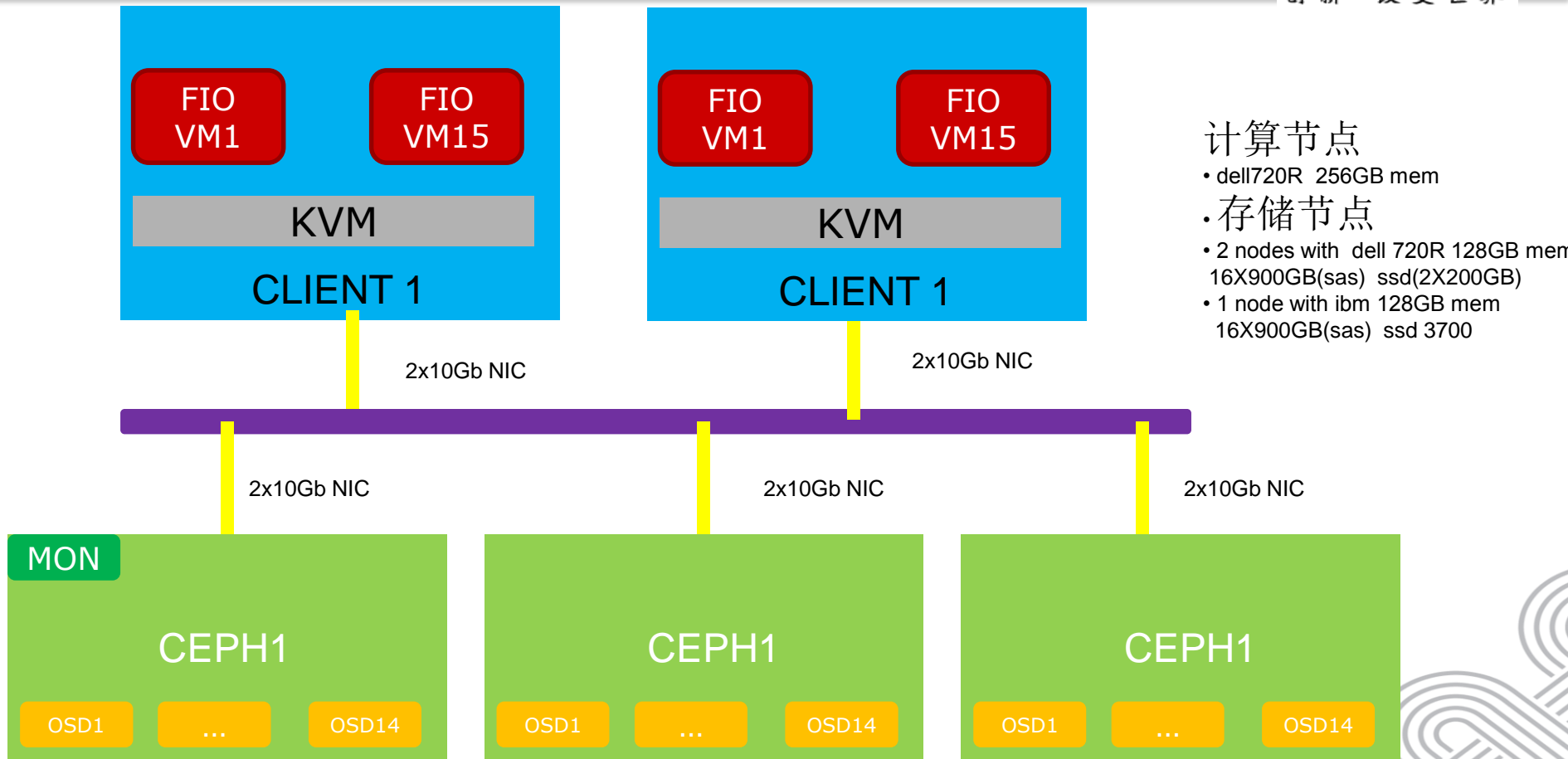
Ceph集群在中国联通研究院的部署

第三部分

中国联通研究院对ceph的测试优化

第四部分

联通沃云与ceph的对接简介



1 万兆+SSD 20G (journal)

❑配置参数默认值

- 不限速4k、64k顺序随机读写
- 限速4k、64k、顺序随机读写
- Run_queue_depth调优
- 大文件测试

❑配置参数优化

- 不限速4k、64k顺序随机读写
- 限速4k、64k、顺序随机读写
- Run_queue_depth调优
- 大文件测试

2 千兆+SSD 20G (journal)

❑配置参数默认值

- 不限速4k、64k顺序随机读写
- 限速4k、64k、顺序随机读写
- Run_queue_depth调优
- 大文件测试

❑配置参数优化

- 不限速4k、64k顺序随机读写
- 限速4k、64k、顺序随机读写
- Run_queue_depth调优
- 大文件测试

3 万兆+20G (journal)

❑配置参数默认值

- 不限速4k、64k顺序随机读写
- 限速4k、64k、顺序随机读写
- Run_queue_depth调优
- 大文件测试

❑配置参数优化

- 不限速4k、64k顺序随机读写
- 限速4k、64k、顺序随机读写
- Run_queue_depth调优
- 大文件测试

4 千兆+20G (journal)

❑配置参数默认值

- 不限速4k、64k顺序随机读写
- 限速4k、64k、顺序随机读写
- Run_queue_depth调优
- 大文件测试

❑配置参数优化

- 不限速4k、64k顺序随机读写
- 限速4k、64k、顺序随机读写
- Run_queue_depth调优
- 大文件测试

1 万兆+SSD 20G (journal)

□ 64k顺序读写

虚拟机数	4	8	16	24
顺序写 (M/S)	714	943	944	1675
顺序读 (M/S)	266	518	872	565

□ 4k随机

虚拟机数	4	8	16	24
随机写 (IOPS)	2586	2817	2672	11614
随机读 (IOPS)	3123	4740	7662	15421

□ 限速60

虚拟机数	4	8	16	24
顺序写延时 (ms)	2.418	9.368	13.941	18.7491
顺序读延时 (ms)	7.474	9.548	10.107	10.689

2 千兆+SSD 20G (journal)

64k顺序读写

虚拟机数	4	8	16	24
顺序写 (M/S)	82.63	103.54	147.56	138.08
顺序读 (M/S)	137.54	198.35	218.32	233.33

4k随机

虚拟机数	4	8	16	24
随机写 (IOPS)	3397	3342	3352	3349
随机读 (IOPS)	2711	4503	6636	7824

限速60

虚拟机数	4	8	16	24
顺序写延时 (ms)	25.13	40.05	66.967	73.713
顺序读延时 (ms)	14.76	20.975	44.706	52.635

3 万兆+20G (journal)

64k顺序读写

虚拟机数	4	8	16	24
顺序写 (M/S)	581.34	714.70	738.75	780.68
顺序读 (M/S)	262.90	474.98	784.61	1056.81

4k随机

虚拟机数	4	8	16	24
随机写 (IOPS)	2602	2413	2393	2500
随机读 (IOPS)	2710	4242	5830	6241

限速60

虚拟机数	4	8	16	24
顺序写延时 (ms)	1.002	1.177	10.569	15.772
顺序读延时 (ms)	7.778	8.615	10.328	11.375

4 千兆+20G (journal)

64k顺序读写

虚拟机数	4	8	16	24
顺序写 (M/S)	79.178	104.49	125.18	137.11
顺序读 (M/S)	110.36	202.62	211.39	233.94

4k随机

虚拟机数	4	8	16	24
随机写 (IOPS)	4322	3709	3401	3276
随机读 (IOPS)	2862	4645	6682	7799

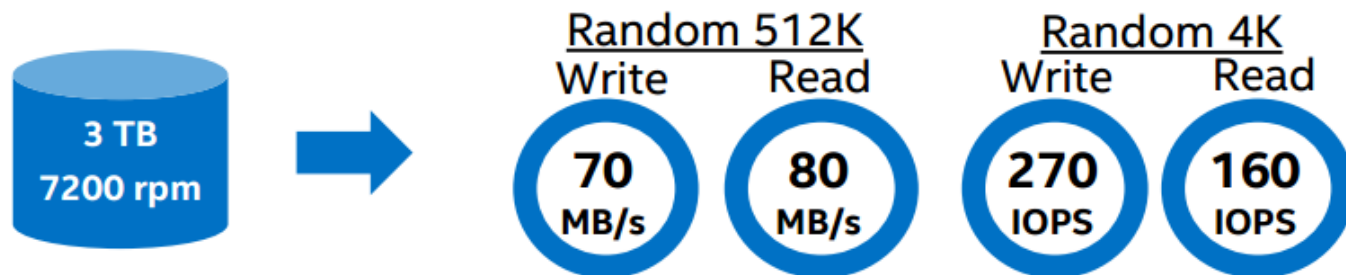
限速60

虚拟机数	4	8	16	24
顺序写延时 (ms)	25.739	41.54	68.11	90.71
顺序读延时 (ms)	18.29	20.213	44.82	52.62

测试结果汇总

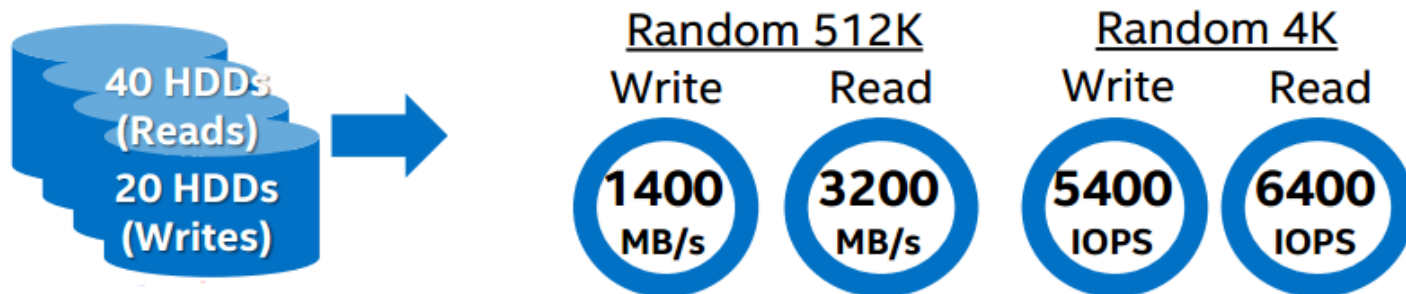
Work Load	Metrics	10Gb	10Gb/SSD	10Gb/SSD/Tuning Config
64k-seqwrite	IOPS	12185	26230	33956
	BW	780687	1679579	735695
	Latency	15.683	7.835	14.745
64k-seqread	IOPS	16822	19215	17437
	BW	1077433	1230930	1116898
	Latency	11.375	10.107	11.083
4k-randwrite	IOPS	2486	2817	3749
	BW	9952.600	11279.890	15006.800
	Latency	25.721	27.282	17.075
4k-randread	IOPS	6241	7662	6937
	BW	25023.940	30674.710	27757.700
	Latency	30.710	18.479	18.439

1 Run FIO on one HDD, collect disk IO performance

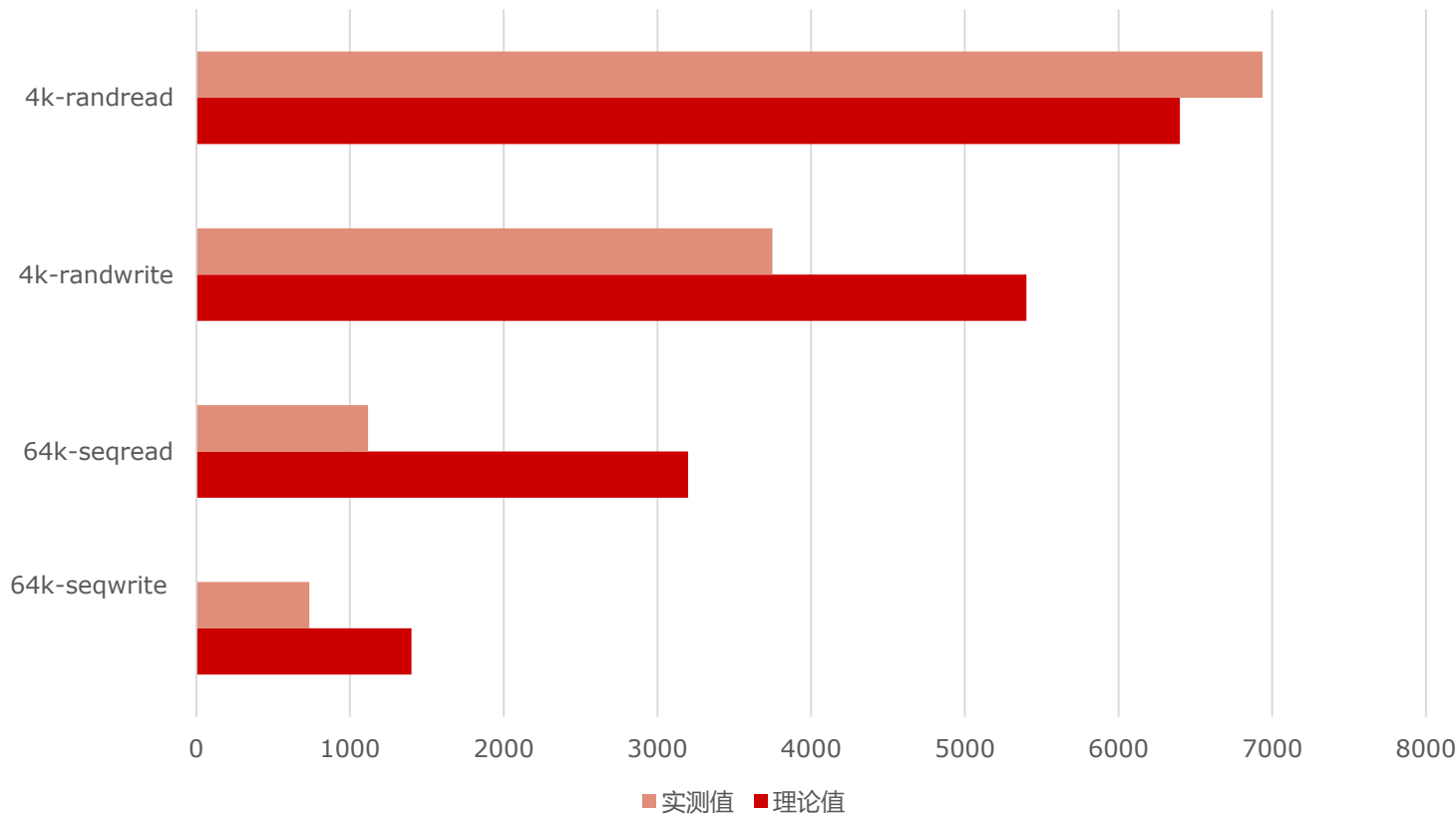


Note: Sequential 64K (Client) = Random 512K (Ceph OSD)

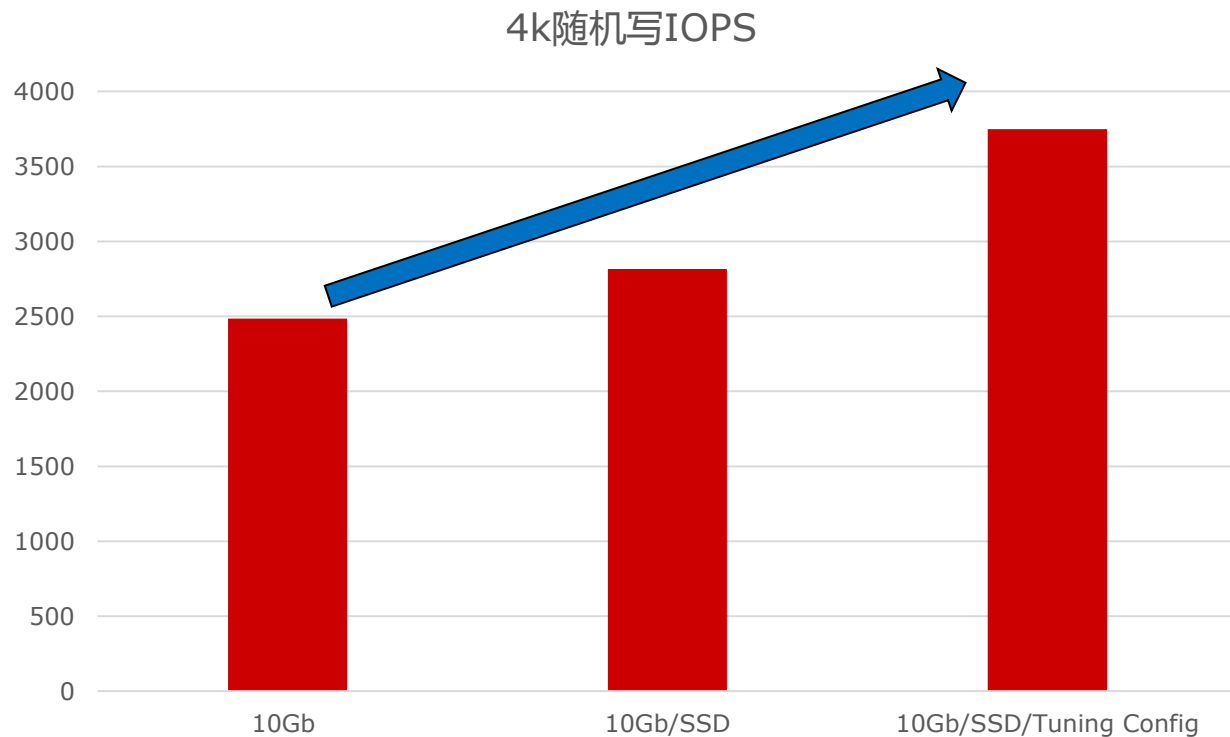
2 Estimate cluster performance (include replication overhead for writes – 2x in this test)



测试值与理论值对比



基于SSD与参数优化建议对随机写性能的提升



第一部分

沃云产品介绍

第二部分

Ceph集群在中国联通研究院的部署

第三部分

中国联通研究院对ceph的测试优化

第四部分

联通沃云与ceph对接简介

THANK YOU!