



CEPHALOCON APAC 2018
THE FUTURE OF STORAGE
22-23 March 2018 | BEIJING

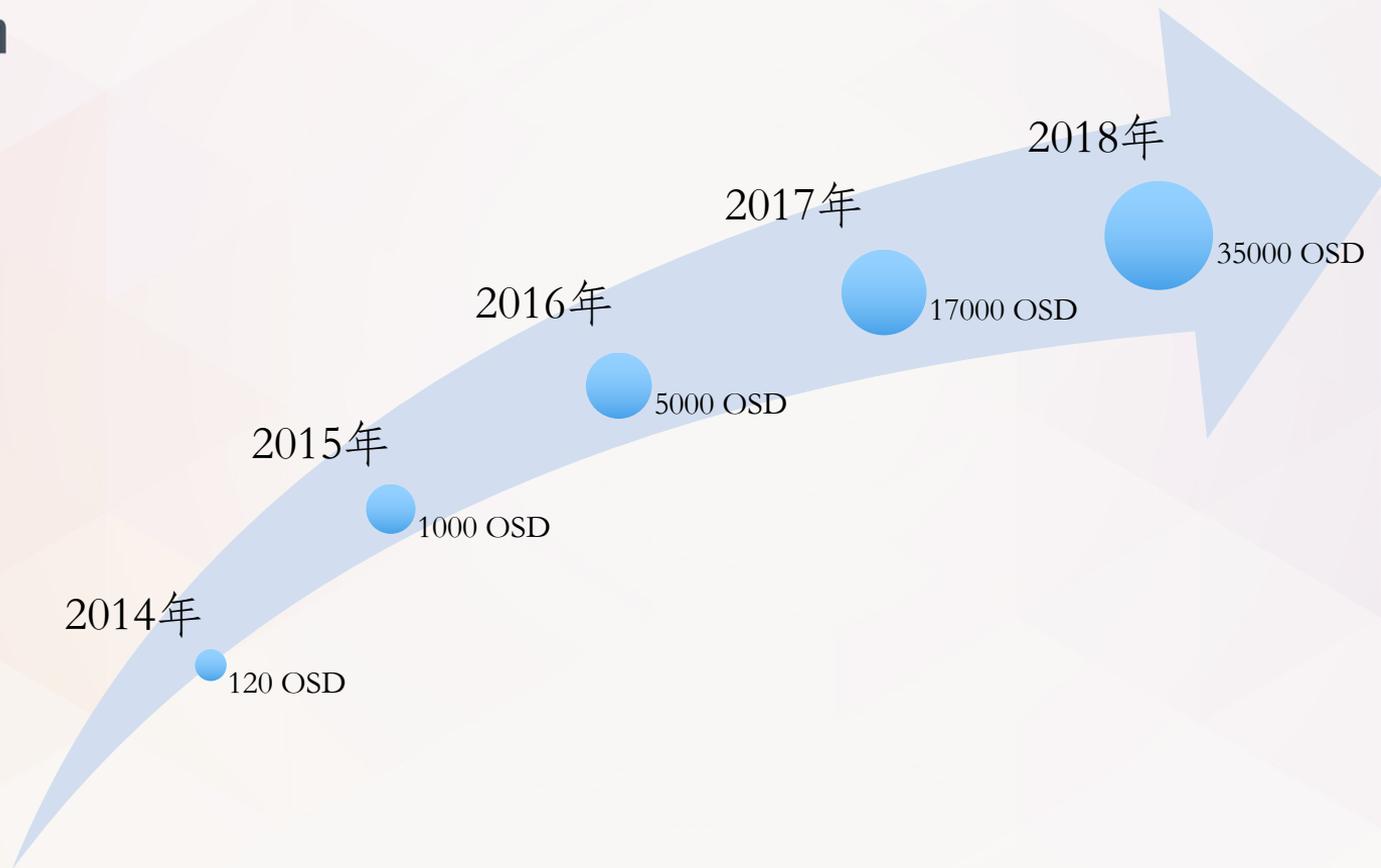
网易云大规模全闪存Ceph存储集群 的运维“智”路

赵文科
技术总监



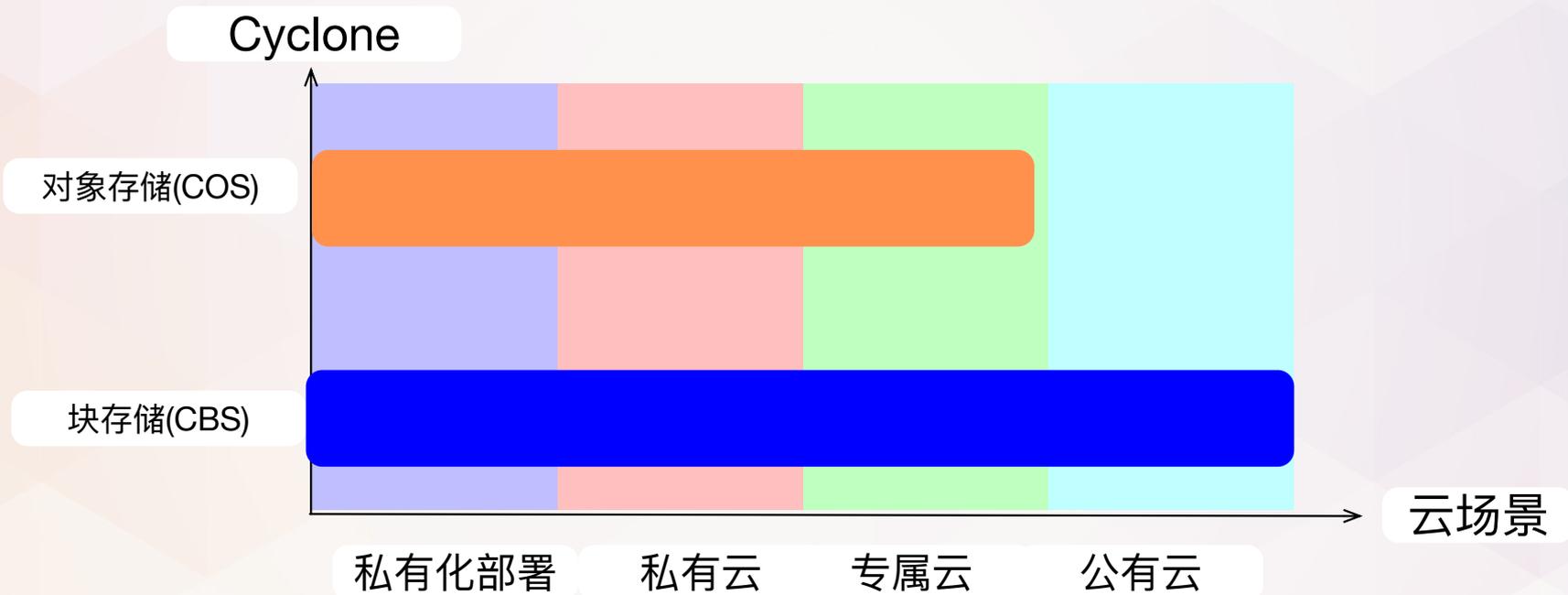


Ceph在网易云的发展历史





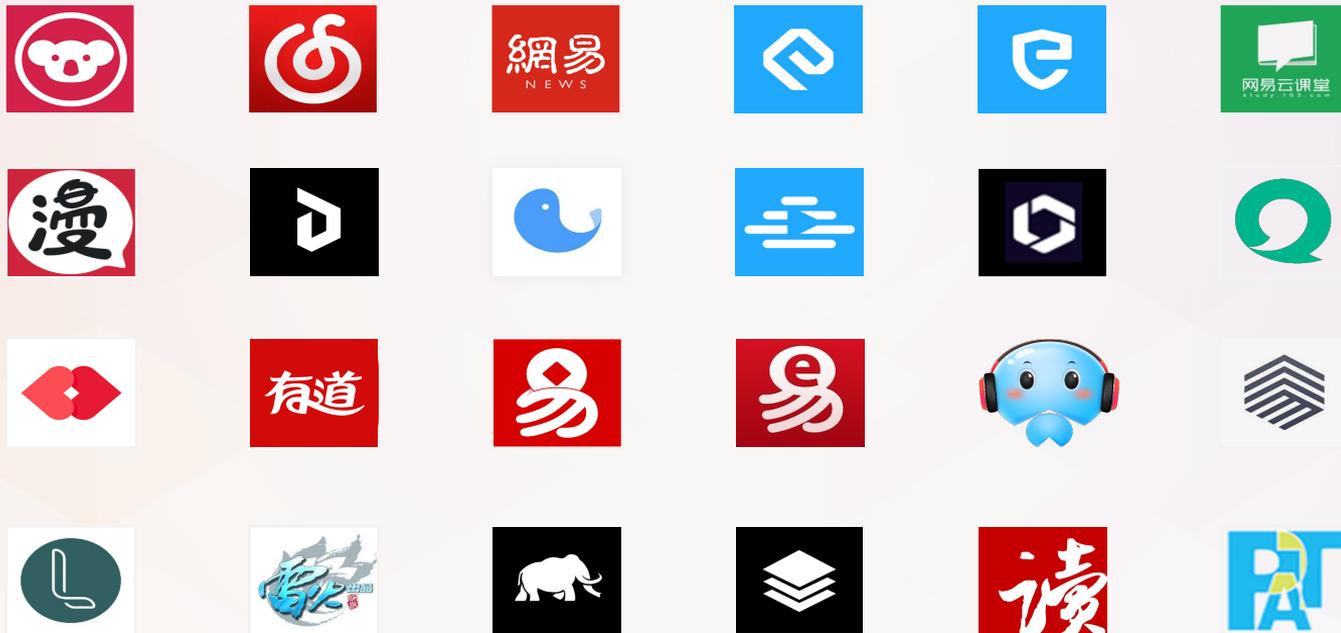
Ceph在网易云的应用场景





Ceph在网易云的业务发展

支持网易100+业务，总存储容量超过30PB (块存储)



块存储

对象存储



Ceph优化1 - Ceph Peering

➤ 背景

当生产环境下集群规模大到一定规模(比如>100节点), Mon和OSD的压力都会指数级的增长; 坏盘、坏存储节点或者扩容都会导致长时间的PG Peering流程, 此过程中, 所有的用户IO都被阻塞, 不能被响应; 从用户来看, 云盘就会长时间卡顿(我们观察到的的卡顿长达2分钟多), 极大地影响用户体验和业务的服

➤ 目标

解决坏盘、宕机、扩容时用户IO卡顿的问题

➤ 设计思路

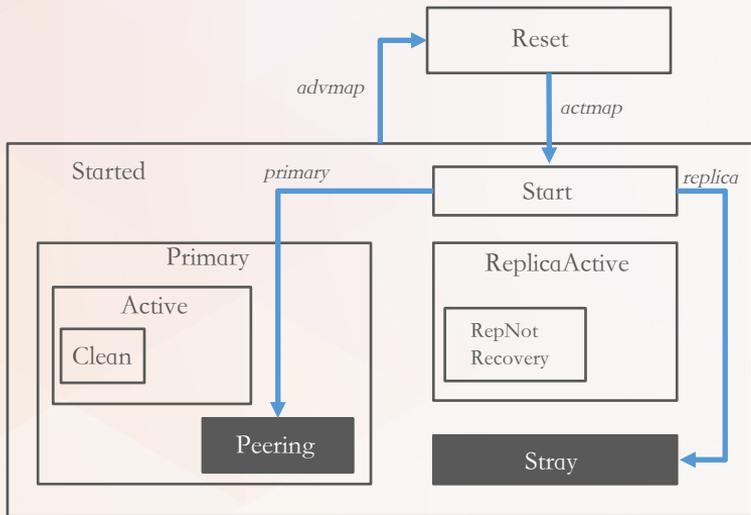
- ✓ PG Peering时会阻塞用户IO, 一个OSD包含100个左右PG。OSD上线或离线时100个PG会同时进行Peering, 但实际上只有1/3的PG需要立即进行Peering, 剩余2/3的PG可以延迟进行。通过减少同时进行Peering的PG数量来缓解IO卡顿问题
- ✓ 主动检测坏盘或者有问题的磁盘控制器的发生, 缩短器件损坏的周期, 主动上报Mon



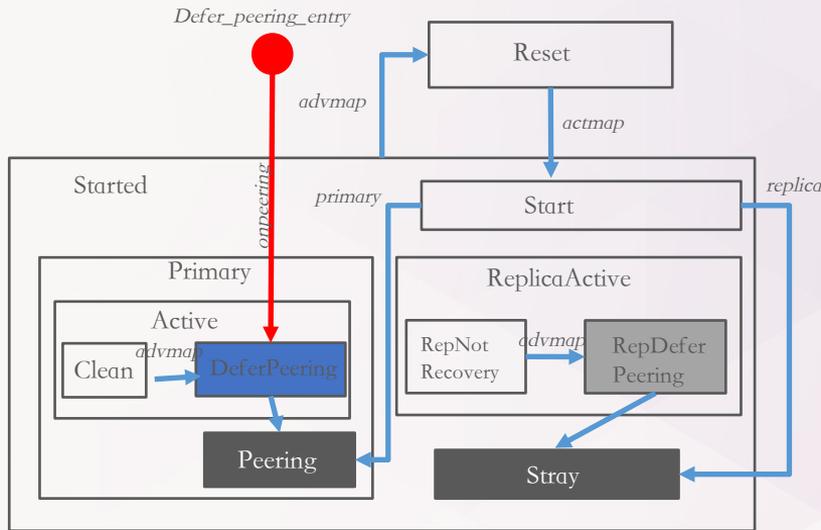
Ceph优化1 - Ceph Peering cont.

- 推迟非主节点的PG同步
- 发现坏盘，主动上报Mon

- Peering阶段控制流量
- 坏盘或扩容，静默处理



原PG中osdmap更新流程



新PG中osdmap更新流程



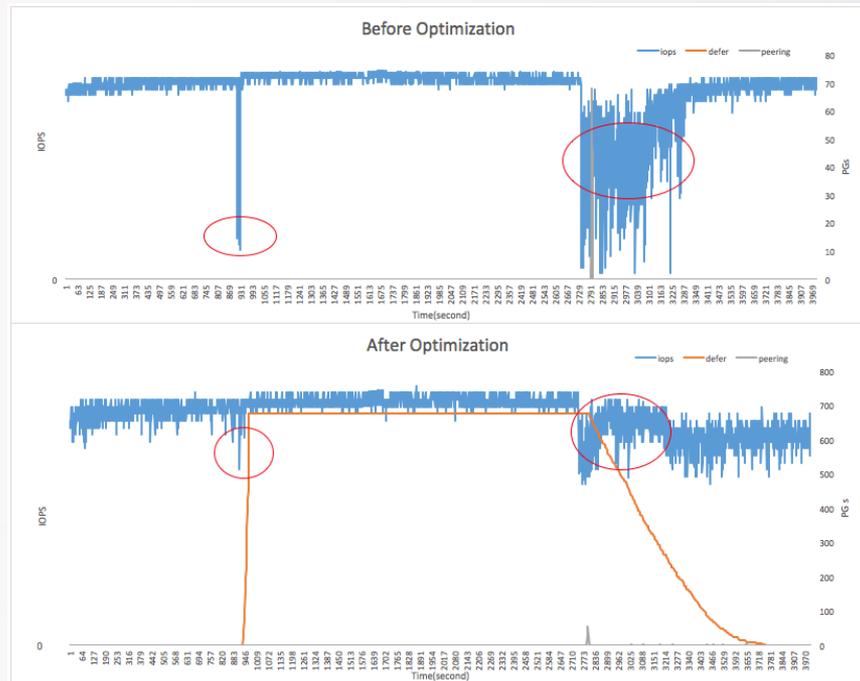
Ceph优化1 – Ceph Peering Cont.

➤ 硬件的问题

- ✓ RAID卡或者HBA卡的4S固定时间的Discovery流程
- ✓ 放大Ceph IO的影响效果(10S)

➤ 优化的结果

- ✓ 通过Ceph Peering的优化，坏盘、坏节点、扩容等行为对用户IO的影响降低到1S以内
- ✓ 整个集群的稳定性大大提升，和运维时间缩短非常明显

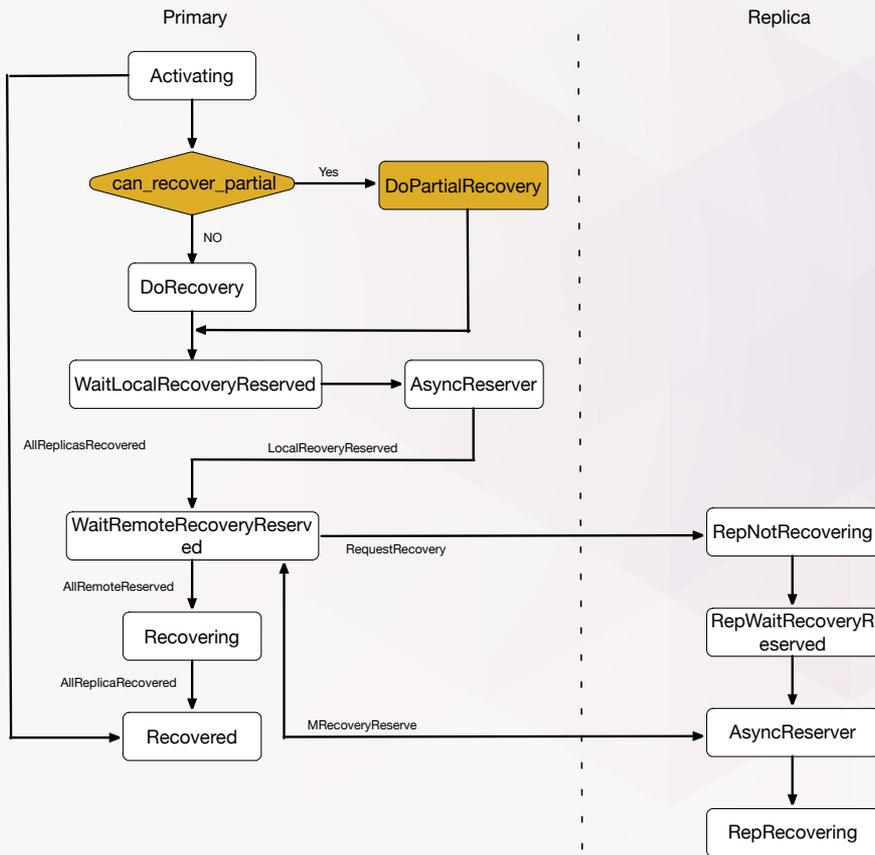




Ceph优化2 – Partial Recovery

➤ 基本思路

- ✓ 利用PGLog中dirty extent数据的offset/len信息，在恢复过程中进行增量数据恢复
- ✓ 增加恢复的流控控制参数 (OSD_RESTORE_THROUGHPUT)，控制恢复的进度和影响，将对磁盘和网络的带宽压力降低到最小程度





Ceph优化2 – Partial Recovery Cont.

➤ 优化的结果

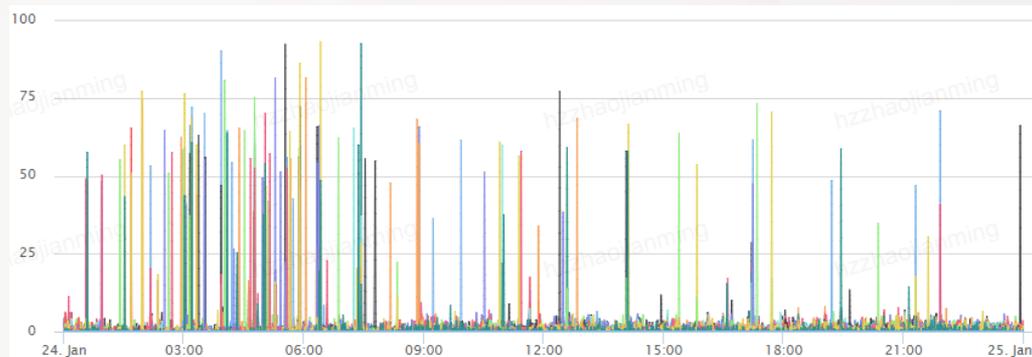
- ✓ 通过增量恢复和恢复带宽流控的优化，大量数据的恢复对用户IO影响的改善比较显著



Ceph优化3 - Scrub的流控

➤ Scrub的现状

- ✓ Scrub分为fast scrub 和 deep scrub
- ✓ Fast scrub只对元数据部分进行checksum计算并校验
- ✓ Deep scrub对整个数据块进行checksum计算并校验
- ✓ Scrub的配置是两种scrub都enable, 且一天多次进行scrub
- ✓ 读数据操作占用过多的磁盘带宽, 导致多个磁盘的diskUtil过高, 不同程度的影响用户IO和用户使用体验





Ceph优化3 – Scrub的流控 cont.

➤ 优化的结果

- ✓ 经过此次优化的效果，在Scrub操作过程中，通过对读磁盘带宽的细粒度流控管理，达到了进行Scrub的时候读磁盘行为对用户IO无明显影响的目标



Ceph优化4 - LibRBD客户端线程数膨胀



➤ 问题的背景

- ✓ 在大规模Ceph块存储的集群中，当一个Ceph数据池的OSD数量超过一定数值(比如600 ~ 840个)，客户端librbd为请求分配的线程数会发生膨胀。
- ✓ 比如三个机柜满配840个OSD，那么每个客户端需要维护(12个卷 * 840 * 2)个线程总数，那么需要宿主机为此预留50GB内存，严重地影响虚机交付数量和宿主机的稳定性；同时，线程的增加也会导致线程上下文切换的开销

➤ 优化的工作

- ✓ 基于高性能NRPC库的功能，对SimpleMessenger的模型重构成基于事件驱动的异步消息实现
- ✓ 由虚拟机产生的IO或者请求由librbd层固定数量的线程来服务，异步化实现下发IO或者请求到后端Ceph集群的模型





Ceph优化5 – OSD假死的问题

➤ 问题的背景

- ✓ 在大规模Ceph块存储的集群中，Mon节点的压力非常大，网络抖动或者存储节点的问题，会导致OSD与Mon之间、OSD之间的心跳会超时，某些情况下，个别OSD会被Mon节点错误的标识为Down，从而影响系统的稳定性

➤ 优化的工作

- ✓ 在秒级监控的系统帮助下，通过逐步的调整集群内Mon节点的心跳间隔来调控心跳本身的压力，尽量减少OSD被误报假死的情况



未来的工作

- Cyclone Store针对读写Cache的优化
- 裸机容器的成熟解决方案的优化
- Rados层面4K IO性能优化
- RGW小文件系统的优化
- Librbd客户端的优化
- 端到端的协程的实现与管理
- . . .



THANKS 感谢聆听