中兴通讯
技术丛书

HZ BOOKS
华章IT

# Ceph

## 设计原理与实现

CEPH PRINCIPLE AND IMPLEMENTATION

谢型果◎等著

**Ceph创始人Sage Weil亲自作序**

中兴Clove团队核心成员撰写，Clove团队在Ceph项目的Commit数量，中国第一，世界第二，仅次于创始团队RedHat

从设计者和使用者角度系统剖析Ceph的核心设计理念与实战技巧

机械工业出版社
China Machine Press

## 内容介绍

本书是中兴Clove团队多年研究和实践经验的总结。Ceph创始人Sage Weil的高度评价并亲自作序。

Clove团队是Ceph项目的核心贡献者，从贡献的Commit数上看，连续多个版本的贡献在中国排名第一，世界排名第二。Clove团队对Ceph有非常深入的研究，在中兴通讯内部进行了大量的生产实践。

本书同时从设计者和使用者的角度系统剖析了Ceph 的整体架构、核心设计理念，以及各个组件的功能与原理；同时，结合大量在生产环境中积累的真实案例，展示了大量实战技巧。每一章都从基本原理切入，采用循序渐进的方式自然过渡至Ceph，并结合 Ceph 的核心设计理念指出需要进行哪些必要的改进和裁剪，使得读者不但能够知其然，而且能够知其所以然，真正做到了“源于 Ceph ，高于 Ceph”。此外，写作时尽量避免涉及过多非必要的专业术语，做到深入浅出并且每章相对独立，以最大程度减少阅读障碍。

**本书核心内容：**

- Ceph 核心算法 CRUSH 设计算法分析及拓展
- Ceph 新型高性能存储引擎BlueStore的特性及关键流程分析
- Ceph 高级特性EC Overwrites
- Ceph PG 状态机及数据修复、平衡机制
- Ceph RBD、RGW、CephFS三大主要组件的实现与拓展
- Ceph 生产环境实战技巧

中兴通讯
技术丛书

# Ceph

# 设计原理与实现

CEPH PRINCIPLE AND IMPLEMENTATION

谢型果 任焕文 严 军
罗润兵 韦巧苗 骆科学 ◎著

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

Ceph 设计原理与实现 / 谢型果等著 . --北京：机械工业出版社，2017.8
（中兴通讯技术丛书）

ISBN 978-7-111-57842-0

I. C… II. 谢… III. 分布式文件系统 IV. TP316

中国版本图书馆 CIP 数据核字（2017）第 206048 号

# Ceph 设计原理与实现

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）
责任编辑：何欣阳　　责任校对：李秋荣
印　　刷：三河市宏图印务有限公司　　版　　次：2017 年 9 月第 1 版第 1 次印刷
开　　本：186mm × 240mm　1/16　　印　　张：19.75
书　　号：ISBN 978-7-111-57842-0　　定　　价：69.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88379426　88361066　　投稿热线：（010）88379604
购书热线：（010）68326294　88379649　68995259　　读者信箱：hzit@hzbook.com

# *Foreword* 推荐序一

Ceph began more than ten years ago as an effort to build a better distributed file system for large supercomputers. In the course of designing for scalability and failure, we ended up creating something that was perfectly timed and well-suited for the explosion of cloud computing infrastructure deployments a few years later. Ceph was open source, scalable, avoiding single points of failure, and provided a block interface that was already well integrated with KVM, the preferred open source hypervisor. The rest is history.

Today, Ceph provides several storage interfaces: the RBD block interface, used widely for backing virtual machines; the RGW object interface, which provides an S3-compatible object storage interface; and CephFS, the scalable POSIX distributed file system we originally set out to build. It is scalable, fast, runs on commodity storage components, and, most importantly, it is 100% free and open source software. Ceph is deployed by a majority of OpenStack clouds and is used in a broad range of other domains, from genomics research to high energy physics to video streaming. As the world's demands for data storage continue to expand we expect to see many more.

Our goal then and now is to ensure that the most compelling storage system available is a free one. Until recently, the market for systems that would store data at scale was dominated by expensive storage appliance vendors with few (if any) open solutions. This drove up costs, driving many cash-strapped users (like research and academic institutions) out of the market, but more importantly it meant that the engineering and development of storage systems was done in secret, in parallel, by many different organizations. Free

software allows organizations to effectively pool their engineering resources, taking advantage of their competitors ' investments as well as the community of users and casual contributors to improve the quality of the system.

I am delighted that Ceph has seen great interest and success in China, and excited to see this book published to help make open source storage more accessible to everyone. It has been a challenge to collaborate effectively with users and developers in China due to time zones and language differences, but I hope that this book (and others) will help bridge the divide.

——Sage Weil Ceph 创始人

# *Forword* 推荐序二

**张万春　中兴通讯股份有限公司副总裁**

阅读了谢型果、任焕文、严军、罗润兵、韦巧苗、骆科学 6 位同事创作的《 Ceph 设计原理与实现》，感到非常高兴，并由衷祝贺创作团队的杰出贡献！就这本书，我想谈三点看法：

**一、中兴通讯重视技术的发展**

云计算、大数据、人工智能三位一体，它们重新定义了 IT，重新定义了资产，重新定义了工具和效率，这些技术力量越来越快地驱动和改变了整个产业，成为支撑行业变革、选择技术伙伴、拓展创新业务和提供高效服务的技术平台。中兴通讯作为全球通信领域的重要厂商，非常重视技术的发展。在云计算领域，多年来致力于利用先进技术，研发更高速度、更大容量、更高安全、更具弹性、更低成本的云计算基础设施。其中 Ceph 就是开源分布式云储存领域中最具活力、最先进的基础技术社区之一，本书记录了我们在这个领域最新的探索实践。

**二、中兴通讯重视社区的力量**

中兴通讯非常认同和重视社区的力量，致力于建立开放合作的生态环境。我们参加了全球多个开源的社区项目，成为其中最关键的伙伴，包括我们和 Openstack、Ceph 社区的合作。中兴通讯在 Ceph 领域技术能力的发展，离不开与社区的合作。中兴通讯的 Ceph 团队是一个优秀的自组织、自管理、自激励的开放合作的敏捷组织，他们内通外联与社区合作，共同推动 Ceph 技术的发展演进。

### 三、中兴通讯乐于分享最佳实践

中兴通讯作为最负社会责任的高科技企业，非常愿意将我们的知识、经验和服务分享到社区、回馈到社会。本书的6位作者身处Ceph技术研发的最前沿，他们精心创作的《Ceph设计原理与实现》有着三个鲜明的特点：一是最系统，二是最前沿，三是最有深度。我们很乐意将这本优秀的著作分享给大家！

最后再次感谢创作团队的重大贡献，并欢迎各位读者开启精彩的《Ceph设计原理与实现》阅读之旅！

# *Foreword* 推荐序三

**陆平　中兴通讯股份有限公司副总裁**

近几年，随着IT信息技术的飞速发展，云计算、虚拟化及池化技术得到了广泛的应用。作为云计算最受追捧的开源项目，OpenStack让越来越多的人感受到了虚拟化的魅力，并在金融、政务、电力和制造业广泛被使用。在最新的OpenStack 2017用户调研中，Ceph RBD以绝对优势（65%）的环境占有率，证明了大家对Ceph充满信心，而作为OpenStack默认存储后端的Ceph，也不负众望，近两年发展得如火如荼，不仅吸引了越来越多的大厂商加入到Ceph生态圈，而且越来越多的行业也采用了Ceph作为其优选的存储解决方案。在大数据盛行的时代，数据量井喷式增长，动辄上PB、EB甚至是ZB的存储需求比比皆是，而且对性能、可靠性的要求也越来越高，Ceph以其优异的性能、可靠性及灵活的扩展性能受到各行各业的青睐，想想也是理所应当的事情。

Ceph作为一个十多年前就已经诞生的开源项目，能够发展到今天，它的生命力是由每一个社区贡献者释放和延续的，我们很欣喜地发现，这种由参与的力量所带来的生命力，随着Ceph开源社区的不断发展及贡献者的日益增多，而变得越来越旺盛。让我们更兴奋的是在广大贡献者的不断努力下，Ceph依然在飞速发展，中兴通讯作为Ceph开源社区中持续活跃的贡献者，无疑给Ceph开源社区注入了更多的激情和活力。

本书是中兴通讯在Ceph开源社区中长期积累的创作成果，不仅从设计原理及思想上对Ceph进行了剖析，而且结合实践深入浅出地将Ceph的独特魅力展现给大家，对于想进阶参与Ceph开源社区的人来说，绝对是一本不可多得的好书。

Ceph是“存储的未来”，相信在大家的共同努力下，这个“未来”不会远了。

# 前　言 *Preface*

诞生于 2006 年的 Ceph，是开源社区的明星项目，也是私有云事实上的标准——OpenStack 的默认存储后端。作为当前最火爆的分布式存储系统，Ceph 拥有诸多引人注目的特性。

首先，Ceph 是一种软件定义存储，可以运行在几乎所有主流的 Linux 发行版（典型如 CentOS 和 Ubuntu）和其他类 UNIX 操作系统（典型如 FreeBSD）上。2016 年，社区进一步将 Ceph 从 x86 架构移植到 ARM 架构中，令 Ceph 应用场景进一步扩展至移动、低功耗等前沿领域，使得 Ceph 未来充满无限可能。

其次，Ceph 的分布式基因使其可以轻易管理成百上千个节点、PB 级及以上存储容量的大规模集群，同时基于计算的扁平寻址设计使得 Ceph 客户端可以直接和服务端的任意节点通信，从而避免因为存在访问热点而导致性能瓶颈。实际上，在没有网络传输限制的前提下，Ceph 可以呈现我们所梦寐以求的、性能与集群规模成线性扩展的优秀特性。

最后，Ceph 是一个统一存储系统，既支持传统的块、文件存储协议，例如 SAN 和 NAS；也支持新兴的对象存储协议，例如 S3 和 Swift，这使得 Ceph 理论上可以满足时下一切主流的存储应用需求。此外，良好的架构设计使得 Ceph 可以轻易拓展至需要存储的任何领域。

上述这一切使得理论上只要存在存储需求，Ceph 就能找到用武之地。因此，诚如 Ceph 社区所言：Ceph 是存储的未来！

## 为什么写这本书

在 Ceph 的设计理念中，高可扩展性、高可靠性和高性能都是其核心考虑要素。此

外，为了能够最大程度地拓展 Ceph 的“触角”（Ceph 本意就是章鱼），Ceph 当中所有组件都被设计成松耦合和高度可定制的。基于上述考虑，Ceph 采用面向对象的语言——C++ 进行开发，并且在具体实现上大量采用了 STL 和 Boost 库中的高级特性。一方面，C++ 被公认为最复杂的编程语言之一；另一方面，经过 10 年的发展，Ceph 已经成为一个代码行数超过百万的庞然大物，各种组件多如牛毛，组件之间关系错综复杂。更加令人望而生畏的是：随着 Ceph 应用场景日益广泛，大量新需求新特性持续涌入，Ceph 正加速向前发展！社区代码每天都在发生翻天覆地的变化——一方面很多模块从无到有，另一方面很多模块从有到无，即便是一些仍然存在的模块，短短几个开发周期之后就会变得面目全非。上述这一切都成为大量渴望接触 Ceph、玩转 Ceph 和深度参与 Ceph 的开发人士的梦魇，足以令他们手足无措，对 Ceph 望而却步。

此外，虽然 Ceph 诞生至今已经超过 10 年的时间，但是在国内兴起却是近几年的事情（感谢 OpenStack），因此相关书籍异常匮乏。市面上仅有的几本，或者单纯从实践角度针对如何使用 Ceph 进行介绍，因为缺乏理论作为指导，加之 Ceph 的命令集一直处于进化之中并且越来越庞大，普通读者可能无法留下深刻印象；或者单纯从源码角度对 Ceph 进行分解和剖析，一方面牵涉到大量实现细节，另一方面源码日新月异，因此非资深开发者可能不易上手。再将视野转向国外——Ceph 官方社区虽然早有专门的文档库对 Ceph 进行系统性的介绍，但是一方面文档库过于庞大并且涉及大量专业术语，另一方面作者和国内读者语言、文化背景存在巨大差异，导致直接阅读这类文档困难重重、举步维艰。

来自 ZTE 的 Clove 团队，自 2014 年开始接触 Ceph，是国内最早从事 Ceph 研究和开发的团队之一。团队从传统存储领域转型，大部分成员此前都有从事 SAN 或者 NAS 开发的背景，因此转战 Ceph 可谓如鱼得水。自成立之日起，Clove 团队就一直和 Ceph 社区保持着良好的互动，我们在使用 Ceph、享受 Ceph 带给我们种种好处的同时，一方面通过反馈故障、修复故障、推送特性等方式持续回馈社区，另一方面通过参与和举办线下沙龙等方式不遗余力地宣传和推广 Ceph 。时至今日，团队中不少人都已经成长为国内在 Ceph 社区中独当一面的活跃开发者。

因为我们在多年的摸索过程中深切体会到学习资料匮乏对 Ceph 初学者所造成的巨大困扰；加之，普及 Ceph、推广 Ceph，与社区共筑良好的 Ceph 生态圈并最终实现社区广大开发者和用户双赢也是我们和社区的共识，我们自 2016 年年中开始动笔编写本书。之所以选择这个时间点，一是因为我们团队已经在传统存储领域以及 Ceph 社区耕耘多年，自身积淀已经逐步殷实；二是 Ceph 这两年在国内发展如火如荼，受众日益广泛，时机

逐渐成熟。书中大部分内容基于社区最新（2017 年 1 月）发布的 Kraken 稳定版，侧重于 BlueStore、EC overwrites、QoS 等一众新增组件和新增特性的介绍，写作时每章务必追求从基本原理切入，采用循序渐进的方式自然过渡和推广至 Ceph，并结合 Ceph 的核心设计理念指出需要进行哪些必要的改进和裁剪，使得读者不但能够知其然，而且能够知其所以然；同时，写作时尽量避免涉及过多、非必要的专业术语，做到深入浅出；并且每章相对独立，最大程度地减少阅读障碍。此外，为了进一步加深读者印象，每个章节都穿插了不少实用案例，最后一章的素材更是全部源于我们日常积累的、从客户处收集的生产案例，极具代表性和通用性，如果读者能够在阅读、学习的同时进行实战演练，理论结合实践，相信必定能够取得更大收益。

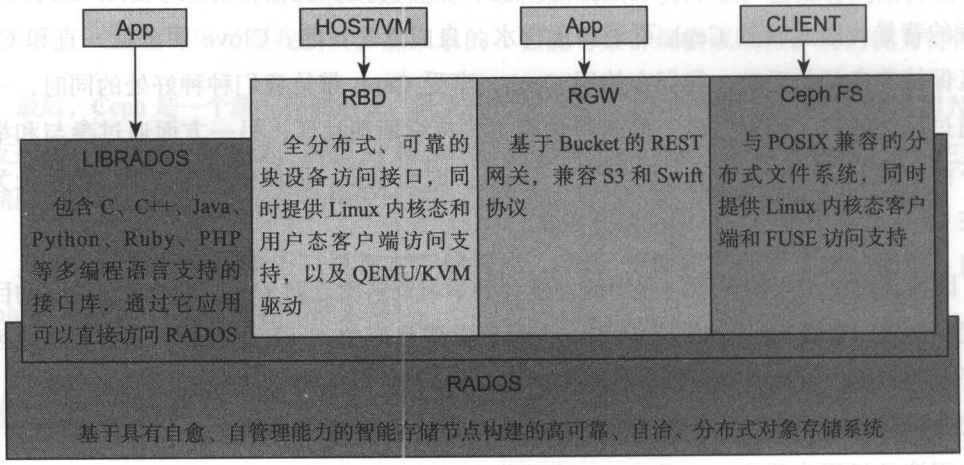
## 本书的读者对象

本书适合于对 Ceph 有一定了解，想更进一步参与到 Ceph 开源项目中来，并致力于为 Ceph 项目添砖加瓦的开发者阅读。

此外，高级运维人员通过阅读本书也能够了解和掌握 Ceph 的核心设计理念及高级应用技巧，从而在日常运维工作中更加得心应手。

## 本书的主要内容

Ceph 整体架构如下：




Ceph 整体架构

其中，RADOS 是 Ceph 的支撑组件，除了 Ceph 当前的三大核心应用组件 RBD、RGW 和 CephFS 之外（它们分别提供块、对象和文件访问接口），原则上，基于 RADOS 及其派生的 librados 标准库也可以开发任意类型的其他应用组件。本书侧重介绍 RADOS 及三大核心应用组件——RBD、RGW 和 CephFS，详细章节及简介如下：

**第 1 章　计算为王**

CRUSH 是 Ceph 两大核心设计之一。CRUSH 良好的设计理念使其具有计算寻址、高并发和动态数据均衡、可定制的副本策略等基本特性，进而能够非常方便地实现诸如去中心化、有效抵御物理结构变化并保证性能随集群规模呈线性扩展、高可靠等高级特性，因而非常适合应用于 Ceph 这类对可扩展性、性能和可靠性都具有严苛要求的大型分布式存储系统。

**第 2 章　性能之巅**

自 Jewel 版本开始，社区引入了一种新型的高性能对象存储引擎——BlueStore，用于取代服役已经超过 10 年的 FileStore。BlueStore 的引入毫无疑问是这两年来 Ceph 最引人注目的特性之一。

**第 3 章　时空博弈**

Ceph 传统的三副本数据备份方式能够在取得高可靠性的前提下最小化前端请求的响应时延，因而特别适合对可靠性和性能都有一定要求的上层应用。这种目前使用最广泛的备份方式缺点在于会大量占用额外的存储空间，因而导致集群的实际空间利用率不高。与之相反，纠删码以条带为单位，通过数学变换，将采用任意 k + m 备份策略所消耗的额外存储空间都成功控制在 1 倍以内，代价是计算资源消耗变大和前端请求响应时延变长，因而适合对时延不敏感的“冷数据”（例如备份数据）应用。在 Kraken 版本中，社区通过解决纠删码中最复杂的覆盖写难题，使得纠删码类型的存储池第一次见到了迈向生产环境的曙光。

**第 4 章　迁移之美**

PG 是 Ceph 最核心和最复杂的概念之一，这也使得学习和了解 PG 成为 Ceph 最富挑战性的工作之一。在 PG 为数众多的优秀特性中，也许最重要也最引人注目的是它可以在 OSD 之间（根据 CRUSH 的实时计算结果）自由进行迁移，这是 Ceph 赖以实现自动数据恢复、自动数据平衡等高级特性的基础。

**第 5 章　控制先行**

在虚拟化技术大行其道的今天，如何针对有限的资源进行集中管理并按需分配以最大化收益，一直是焦点议题之一。Ceph 通过积极引入 QoS 功能，有望对集群的 IOPS、带宽等 I/O 资源进行合理统筹，实现按需、定量分配，从而对外提供更加精细化的存储服务。

**第 6 章　无心插柳**

自 2007 年 Sage A. Weil 正式发布 Ceph 以来，Ceph 实际上已经存在并发展了 10 余年时间。Ceph 在设计之初被定位为一个纯粹的分布式文件系统（CephFS），但随着虚拟化逐渐成为信息时代的主旋律和以 OpenStack 为代表的云计算技术闪电崛起，社区果断调整重心，开始着力发展新型分布式块存储服务组件——RBD，并使其逐渐成长为 OpenStack 等 IaaS 云计算环境中虚机、镜像、云盘等服务不可或缺的默认块设备存储后端。可以说，Ceph 能够在为数众多的同类软件竞争中脱颖而出，并逐渐成长为最炙手可热的分布式统一存储系统，很大程度上得益于收获了 OpenStack 的青睐，而 RBD 取代 CephFS 伴随 OpenStack 先一步进入公众视野则是意料之外、情理之中。

**第 7 章　应云而生**

在《浪潮之巅》一书的前言中，吴军博士开宗明义地提出："近一百多年来，总有一些公司很幸运地、有意识或无意识地站在技术革命的浪尖之上。在这十几年间，它们代表着科技的浪潮，直到下一波浪潮的来临。"

当前，方兴未艾的云计算无疑代表了科技发展下一波浪潮的到来，而率先基于 AWS 推出公有云服务并成为公有云事实标准的亚马逊公司无疑一只脚已经踏上了这波浪潮的浪潮之巅。事实上，自 2006 年面世以来，AWS 当前存储的对象规模已经高达千亿级别，并已经累积为亚马逊创造了超过百亿美元的利润，由此可见云计算所蕴含的巨大商机。AWS 要求存储系统能够提供与传统块、文件存储都不相同的第三类接口——对象存储接口，并采用自定义的 S3 协议通过互联网（HTTP）进行传输。在此背景下，为了赶上云计算为代表的这波科技浪潮，Ceph 兼容以 S3 为代表的对象存储协议簇的对象存储网关——RGW 应云而生。

**第 8 章　经典重现**

文件系统伴随操作系统一同诞生，是计算机科学中最基本和最经典的概念之一。Ceph 自诞生之日起就被定位为一个分布式文件系统。时至今日，在 Ceph 的三大典型应

用场景中，RBD 和 RGW 先后乘着云计算的东风后来居上获得了日益广泛的应用，但是起步最早的 CephFS 却一直迟迟未能有所建树。究其原因，一是文件系统采用树状结构管理数据（文件和目录）、基于查表进行寻址的设计理念，与 Ceph 采用扁平方式管理数据、基于计算进行寻址的设计理念格格不入；二是支持文件系统必然要求 Ceph 引入集中的元数据管理服务器（作为树状结构的统一入口用于寻址），这又与 Ceph 去中心化、追求近乎无限横向扩展能力的设计思想激烈冲突。

尽管颇具戏剧性，然而一个不可否认的事实是：RBD 和 RGW 的蓬勃发展反过来又促使 Ceph 在云计算以外的领域也迅速普及并逐渐变得广为人知。随着传统块、文件存储设备日薄西山，业界期待 Ceph 作为一个真正意义上的大一统存储系统接管传统存储的呼声越来越高。因此，尽管道阻且长，但是作为替代传统文件存储的重要一环，重启 CephFS 研究并使之早日进入生产环境已是势在必行。

**第 9 章　运用之妙**

运用之妙，存乎一心。经过漫长的 Ceph 基本原理学习之旅，相信大部分读者已经按捺不住、想要通过亲自动手实践来体验 Ceph 的种种神奇魅力。在本书的最后，我们精心准备了以 Ceph 应用于生产环境的各种案例为原材料烹制的饕餮盛宴，以飨读者。

## 勘误与支持

赠人玫瑰，手有余香。我们真诚地希望每位读者都能从阅读本书中找到乐趣并获得收益。当然，由于水平有限，书中难免存在错误和疏漏，我们将每位读者都当成是志同道合（关注 Ceph、爱好 Ceph）的朋友，朋友们的指正自然永远是欢迎的。

如果您在阅读本书过程中碰到任何问题，可以通过以下电子邮箱联系我们：

luo.kexue@zte.com.cn

xie.xingguo@zte.com.cn

## 致谢

Ceph 官方社区[⊖]的源代码[⊜]是创作本书的原始素材，因此我们首先要感谢 Ceph 官

⊖ http://ceph.com/

⊜ https://github.com/ceph/ceph

方社区，特别是社区领袖和 Ceph 创始人 Sage A. Weil 先生。Sage 学识渊博、为人和善，乐于接纳新人和帮助新人成长。在他的带领下，Ceph 欣欣向荣，十年间从一个默默无闻的学院派作品逐渐成长为开源社区万众瞩目的明星项目。作为 Ceph 官方社区的一分子，Clove 团队与有荣焉。

其次，我们要感谢所在部门的主管领导——谭芳部长，是他给予了 Clove 团队无微不至的关怀和无与伦比的信任，让我们有勇气去不断突破自身瓶颈，全力以赴追求心中的梦想。

再次，我们也非常感谢那些阅读过本书草稿并提出宝贵意见的人：宋维斌（针对本书的大部分章节，他都阅读了两遍以上，他是我们所见过的最细心的人）、朱尚忠（他指出了本书一些晦涩难懂之处，使得读者能够获得更加轻松愉快的阅读体验）和罗慕尧等。

最后，我们要特别感谢 IT 技术学院的闫林老师，如果没有他的鼓励和帮助，相信本书将不会有机会和广大读者朋友们见面。

开放正在成为这个时代的主旋律，开源正在成为软件开发的新信条。与大师同行，和开源社区共成长，让每个深度参与到开源社区中的开发者们都受益匪浅。而以 Linus、Sage 等为首的开源社区领袖，则完美阐释了约翰·邓普顿的名言，“It is nice to be important, but it's more important to be nice”，他们永远是后来者学习和追赶的榜样。

我们期待并将继续为之努力。

# Contents 目　录


推荐序一
推荐序二
推荐序三
前　言

**第1章　计算为王——基于可扩展哈希的受控副本分布策略CRUSH** ····· 1

1.1　straw 及 straw2 算法简介 ····· 2
1.2　CRUSH 算法详解 ····· 6
1.2.1　集群的层级化描述——Cluster Map ····· 7
1.2.2　数据分布策略——Placement Rule ····· 9
1.3　调制 CRUSH ····· 14
1.3.1　编辑 CRUSH Map ····· 15
1.3.2　定制 CRUSH 规则 ····· 19
1.3.3　数据重平衡 ····· 21
1.4　总结与展望 ····· 23

**第2章　性能之巅——新型对象存储引擎BlueStore** ····· 25

2.1　设计理念与指导原则 ····· 26
2.2　磁盘数据结构 ····· 30
2.2.1　PG ····· 30
2.2.2　对象 ····· 38
2.3　缓存管理 ····· 46
2.3.1　常见的缓存淘汰算法 ····· 46
2.3.2　BlueStore 中的缓存管理 ····· 49
2.4　磁盘空间管理 ····· 53
2.4.1　常见磁盘空间管理模式 ····· 53
2.4.2　BitmapFreelistManager ····· 56
2.4.3　BitmapAllocator ····· 57
2.5　BlueFS ····· 59
2.5.1　RocksDB 与 BlueFS ····· 59
2.5.2　磁盘数据结构 ····· 62
2.5.3　块设备 ····· 65
2.6　实现原理 ····· 66
2.6.1　mkfs ····· 66
2.6.2　mount ····· 67
2.6.3　read ····· 69
2.6.4　write ····· 72
2.7　使用指南 ····· 77
2.7.1　部署 BlueStore ····· 77
2.7.2　配置参数 ····· 80
2.8　总结与展望 ····· 83