

DATA DRIVEN GLOBAL VISION CLOUD PLATFORM STRATEG
ON POWERFUL RELEVANT PERFORMANCE SOLUTION CLU
VIRTUAL BIG DATA SOLUTION ROI FLEXIBLE DATA DRIVEN

Changhong Data Platform

内容平台软件架构白皮书

持久可靠、管理简单且横向扩展的对象存储平台，提供最高级别数据保护能力、海量扩展能力、更高存储效率及全面的兼容性

目录

1. 序言.....	4
2. 内容平台软件产品概要.....	6
2.1. 架构概览.....	7
2.2. 常见适用场景.....	8
2.3. 主要概念和特点.....	10
2.4. 免备份功能详细描述.....	18
3. 内容平台软件架构详细介绍.....	23
3.1. 基于云计算的存储系统.....	23
3.2. 系统软件架构.....	25

1. 序言

数字业务被一致认为是企业未来发展的必然方向，而当今各行各业的企业都在面临数字化的搅局者 (Digital Disruptors)带来的挑战，比如金融行业的数字银行和互联网金融、媒体行业的互联网新媒体等。这是一种长期的趋势，企业要么拥抱数字化的产品、服务和信息并将其作为业务的核心所在，要么就是被数字化的搅局者所打败。IT 在企业的数字化转型战略中处于至关重要的位置，云、大数据、移动和社交等 IT 发展趋势也不仅仅是口号，而是迫在眉睫。如今 IT 的最大挑战之一便是非结构化数据不受控制的爆发性增长。电子邮件和文档、视频、Web 页面、演示文稿、医疗影像等数据的持续增长增加了复杂性和风险。这种效果特别是在分布式 IT 环境中十分明显，例如云服务供应商和拥有分支办事处或远程办事处的企业。生成的大量数据难以管理，并且恰当处理非结构化内容十分困难，为更多用户和应用提供支持也产生了复杂性，这些问题都对 IT 部门提出了多种挑战。企业最终会出现各种大量应用和工作负载滋生的存储数据孤岛，只有很少的资源可以进行管理、监管、保护和搜索数据。随着 IT 环境的不断演变，确保在正确的时间将正确的数据放在正确的位置成为严峻的挑战。新技术、第三方应用和移动设备意味着数据无处不在且持续变化。如何始终保持数据的安全性、可控性和可见性变成了一种挑战。

随着 IT 不断发展和数据的积累，企业发现越来越难管理和利用这些数据，特别是分散在不同应用系统的这些数据。而新的技术，互联网或者移动设备等更需要随时随地的访问这些数据。那么如何实现这些数据的集中管控、持久可靠的存储、更好的被保护、随时随地安全访问呢？

内容平台软件是一款多用途的分布式对象存储系统，可以提供另一种解决方案，通过单一的对象存储平台来应对这些挑战，该平台划分为多个虚拟存储系统，每个系统都针对指定的服务水平进行配置。这款解决方案的巨大规模和丰富功能可以帮助私有企业和云服务供应商的 IT 组织管理分布式 IT 环境。这款

解决方案可以帮助企业控制非结构化内容的大量存储需求，并且应对各种不同的工作负载。智能的解决方案使 IT 机构和云服务提供商能够使用单一存储平台存储、保护、保存、检索和分发非结构化内容。通过支持多级服务，并随着技术和规模的变化而迅速演变。借助大量数据管理、数据保护和内容保存技术，这个经济的系统可以显著降低资源需求，甚至可以消除自身基于磁带的备份或与平台连接的边缘设备备份。

内容平台软件构建的分布式对象存储系统，旨在支持大规模非结构化数据资源库。它可以帮助 IT 企业和云服务供应商采用单一存储平台存储、保护、保存和检索非结构化内容。它支持多种服务水平，随着技术和规模的变化稳定发展。利用一系列广泛的数据保护和内容保存技术，该系统可以大幅降低或者甚至消除磁带备份或者与平台相连接的边缘设备备份。内容平台软件消除了存储非结构化内容的孤岛式方法，为每一位租户提供统一的存储资源池，在此基础上支持范围广泛的应用。通过将物理系统划分为多个进行独特配置的租户，管理员可以创建“虚拟内容平台”，这些平台可以进一步划分为多个命名空间，对内容、策略和访问进一步组织。由于可以支持数千个租户、成千上万个命名空间以及一个系统内可达 PB 级容量，内容平台软件真正实现了云就绪。

内容平台软件为存储来自多个应用和数据源的数据提供了单一的基础平台。通过灵活的架构，它支持传统应用，例如多种归档应用（例如文件、电子邮件、录音、数据库、Microsoft SharePoint 和医疗影像）。它还可以作为新的 Web 2.0、支持亚马逊简单存储服务 (S3) 的云、大数据、移动同步和共享、远程和分支机构以及文件和开源应用数据的存储库 — 所有这些都通过单点管理实现。

2. 内容平台软件产品概要

内容平台软件（Changhong Data Platform）用于构建多用途的分布式对象存储系统，旨在支持大规模非结构化数据资源库。内容平台软件可以帮助 IT 企业和云服务供应商采用单一存储平台存储、保护、保存和检索非结构化内容。内容平台软件消除了存储非结构化内容的孤岛式方法，提供分布式的云存储服务，可以通过多租户为每一位租户提供的大规模、多个存储层、持久可靠、不停机的软件更新、多租户和可配置属性等，以实现单一物理实例支持范围广泛的组织和应用系统。通过将物理系统划分为多个进行独特配置的租户，管理员可以创建“虚拟云存储服务平台”，这些平台可以进一步划分为多个命名空间给到不同的应用系统使用，对数据存储空间、存储策略和访问进一步配置。由于可以支持数千个租户、成千上万个命名空间以及一个系统内可达上百 PB 级容量，内容平台软件真正实现了云就绪。

简单的说，相较于传统的文件存储来存放海量的非结构化数据，内容平台软件构建的对象存储系统具备以下特点：

大规模的横向扩展以支持海量数据

单个对象存储系统可以支持上百 PB 级的存储容量，即使单个桶即可支持 1000 亿个文件，无需构建分散的存储系统和创建众多文件系统，即可轻松满足您的海量数据存储需求

适用于您的数据的最持久且可靠的平台

提供全面的数据保护能力，无需要借助于耗时且耗资的传统备份，即可实现物理错误、逻辑错误(人为和病毒等)、比特错误和灾难等场景的数据保护和修复，可以实现 99.999999999999% 的数据持久性。

最全面的安全性与合规性功能

保护您的数据免受内部和外部威胁，并确保可以通过防篡改、保存周期等策略保证数据始终合规，同时还支持更多安全性标准和合规性认证。

功能强大且管理轻松

将数据存储内容平台软件构建的对象存储系统中，您将拥有云中数据存储、管理和使用的最完整体验。您可以轻松管理让各业务部门按需使用存储资源，通过服务策略轻松分配不同的 SLA 级别，简单实现数据生命周期管理，自动化的进行设备更替而无需耗时耗力的传统数据迁移，提供标准化的数据访问乃至检索接口。

随时查询的数据湖

提供 HTTP(S)/CIFS/NFS 等多种访问协议，可以随时随地访问和检索您的数据。提供接口和大数据结合，允许您直接在静态数据上运行复杂大数据分析的云存储平台，无需您提取数据以及将其加载到单独的分析系统。

2.1. 架构概览

内容平台软件构建的对象存储系统采用横向的分布式存储架构，其系统本身包含了数据管理、保护、索引及存储池管理等功能，分布在横向扩展的数据管理(访问)节点上，可以从 4 个节点扩展到 80 个节点。后端存储空间可以是分布式管理节点本地磁盘、外接高密度大容量对象存储容器、NFS 文件存储、异质磁带库或光盘塔甚至是公有云存储，其架构图如下：

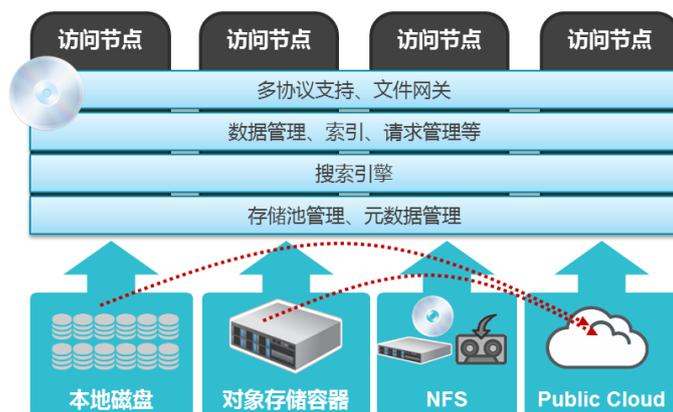


图 1：内容平台软件构建的分布式对象存储架构图

分布式横向扩展集群结构

采用集群架构，元数据和数据分散在集群内的所有节点上。一个本地的名称空间分配，被元数据控制和管理引擎资源监控，文件将被本地化的，在一个集群内部公平地分散。

数据存储负载均衡

当一个节点被加到一个集群中，集群的均衡策略被激活去维持每个数据和元数据的分布。物理地移动数据和元数据从已经存在的节点到一个新的节点，设法达到相同的利用率在一个集群内部的所有节点上。为了保持平台的均衡，每个节点要确保最佳性能和扩展性。

前端访问负载均衡

当应用程序作为前端访问对象存储，对象存储会根据负载情况分配分配处理阶段，使得性能分担到分布式架构上的所有节点，实现最好性能。

2.2. 常见适用场景

内容平台软件构建的对象存储系统是一个多用途的海量非结构化数据存储平台，适用于多种场景：

归档：

归档是对象存储的常见并且重要的用途。这一用途的主要原因之一是提供一种可以扩展的经济型存储和高级数据保护，从而消除对于备份的需求。多种不同应用已与对象存储平台直接集成，用于多种数据类型的归档，例如电子邮件、呼叫数据记录、文档管理数据、病历、医疗影像、媒体和娱乐文件，以及文件系统中的非活动文件。

合规与发现：

通过一系列先进的特性为归档增加了价值，这些特性包括数据保留管理、合法保留和自动数据处理，可帮助企业满足合规要求。通过结合内置的自定义元数据查询和应用的内容搜索能力，允许管理员识别诉讼数据，并自动对数据集执行合法保留，从而防止删除或篡改。

备份缩减和优化：

通常情况下，归档用于从备份流中删除不再变化的数据。企业纷纷将这一概念应用到新的用途：有些实际上将备份映像存储到内容平台软件构建的分布式对象存储中。在这种情况下，压缩、高效数据保护和更快的数据召回率使这一功能的价值不局限于备份存储，而是扩展到了磁带或昂贵的块级重删备份设备。

云应用存储：

大多数新应用正在开发中，这些应用利用轻量级、Web 和云友好的 REST API，并针对私有云、公有云或混合云存储。对象存储是最常见的存储类型，它们呈现这些 API 并优化简单的横向扩展存储，并且可以将丰富的元数据与每个文件捆绑在一起。

非结构化数据管理：

内容平台软件构建的对象存储系统在支持基于策略的数据管理能力方面保持领先。通过定义策略，数据可在对象存储系统的集群内的不同存储类别之间移动，甚至可以在集群外部移动，这些能力针对常见的公有云服务，例如阿里云、AWS S3、Microsoft Azure。这样，内容平台软件构建的对象存储系统可以作为这些云服务的代理，在服务之间移动数据，同时为 REST 应用提供一定程度的 API 抽象。内容平台软件构建的对象存储系统提供了控制数据地理位置的方法，以用于灾难恢复保护和数据分发目的。最后，通过为定制元数据和内置查询能力提供最先进的架构，对象存储系统能够充分利用这些信息，并引领数据管理的新时代。

云服务支持：

许多企业的 IT 机构正在改进其服务交付模式，目的是与不断增长的公有云服务模式保持一致。同样，第 2 层、区域或特定垂直行业服务提供商也在努力做同样的事情。对于这些群体，内容平台软件构建的对象存储系统是优秀的存储选择，提供了大量常见的 REST API 和服务差异化特性。

大数据存储：

从本质上讲，大数据涉及大量非结构化数据。企业希望通过分析从这些数据中释放战略价值。内容平台软件构建的对象存储系统安全、可靠、高效地存储这些数据。此外，先进的元数据架构可以为非结构化数据提供一种结构，允许分析应用查询特定的数据子集，加速分析过程并改善结果。

2.3. 主要概念和特点

对象存储

作为一款通用式对象存储系统，允许将非结构化数据文件作为对象存储起来。一个对象实际上就是包括文件数据及其相关的元数据（用于描述数据）。元数据包括系统元数据和用户自定义的元数据，系统元数据除了常规的文件名、创建时间等信息还可以设置文件的保存策略等，比如数据保护等级（DPL）、保留周期、数据的哈希指纹校验码等；用户自定义元数据可以为应用系统或者人为增加的描述信息，比如视频的节目名、码流、分辨率等，用于数据的检索和访问。

内容平台软件构建的对象存储系统将存储的数据与硬件层相隔离。在内部，接收的文件表示为对象，它包含所需要的数据和元数据，为各种应用提供支持。应用系统可以通过多种协议访问，包括 HTTP/HTTPS 方式的 S3/Swift/REST 等协议接口，也支持传统的 CIFS/NFS 方式访问标准目录结构。

对象结构

内容平台软件将对象存储在资源库，资源库对象由文件数据及其元数据组成，输入并存储在资源库中的数据与该数据的相关信息构成了一种永久性的关系，这些信息称为元数据。元数据则由系统元数据和自定义元数据（可选）以及访问控制列表(ACL)组成。每一个数据对象都包含对象数据和元数据，并被视为一个独立单元。对象的结构如图 2 所示：



图 2; 对象的结构

文件数据是在写入实际文件内容时对内容的准确拷贝。如果对象有合规和保留周期需求，在保留期限到期之前不可删除，除非采用特殊权限的操作。可以配置版本控制功能，可以保留一个文件的多个版本，如果发生人为误操作或病毒等逻辑错误，可以实现数据恢复。

元数据是系统或用户生成的数据，用于描述对象的描述数据，并且定义对象的属性。系统元数据即对象的系统管理属性。包括将对象添加到命名空间的数据和时间（输入时间）、最后更改对象的日期和时间（更改时间）、采用命名哈希算法的对象加密哈希值（用于生成数值）和输入对象的协议等。它还包括对象的策略设置，例如数据保护等级（DPL）、保留、销毁、索引和版本控制（仅适用于命名空间）。

自定义元数据是可选的、用户提供的数据对象描述性信息，通常以格式良好的 XML 形式提供。这通常用于描述对象更加详尽的信息。这种元数据还可以未来的用户和应用理解对象内容，并且重新利用。系统可以支持每一个对象的多种自定义元数据字段。

ACL 是用户提供的可选元数据，包含一级由用户或用户组授权的访问许可，用于执行对象的各种操作。ACL 只能通过命名空间支持。

在命名空间中支持的完整元数据结构如图 2 中所示。它包括所有支持的对象元文件，为样本数据结构而生成（假设每一个对象都添加了自定义元数据和 ACL）。

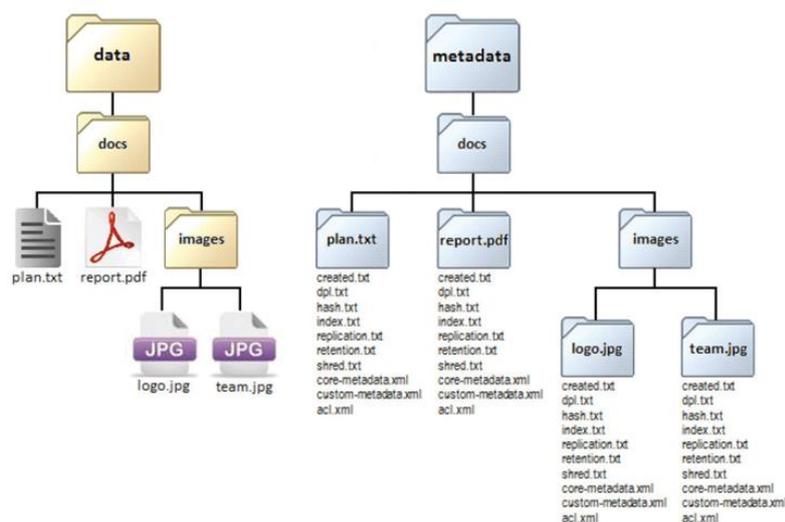


图 3. 完整的文件目录结构（左）和对应的元数据结构（右）

分布式设计

整个系统由硬件和软件组成，且包含众多互相联系的不同组件，为对象存储构成了一个强大、可扩展的架构。内容平台软件运行在一系列的服务器或者节点中，它们都通过网络联系在一起，构成了一个单独的物理实例。所有运行时操作和物理存储，包括数据和元数据在内，都分布式存储在各个节点中。资源库中的所有对象都分配在所有可用的存储空间，但仍然显示为采用标准目录结构的文件。以物理形式存储在特定节点的对象都可以通过所有其他节点上获取。

开放式架构

内容平台软件构建的对象存储系统拥有开放式架构，它可以将存储的数据与技术的变更以及由于产品升级导致的变更实现隔离。这种开放式架构确保用户在添加到资源库很久之后仍然可以访问数据。系统既可以作为一个存储客户数据的资源库，也可以作为一个在线门户网站，通过多种业界标准接口访问数据。HTTP 或者 HTTPS、S3, SWIFT, WebDAV、CIFS 和 NFS 协议可以支持各种不同的操作。这些操作包括：存储数据、创建并浏览目录、浏览并检索对象及其元数据、修改对象元数据和删除对象。采用任何协

议添加的对象都可以立即通过其他支持的协议进行访问。这些协议可以采用 Web 浏览器、客户端工具、第三方应用、Microsoft® Windows® Explorer 或者本地 Windows 或者 UNIX 工具访问数据。

系统允许通过 SMTP 协议实现对资源库的特殊用途访问，仅适用于存储电子邮件。对于数据备份和还原，还可以支持 NDMP 协议。

多租户

多租户支持单一物理实例的资源库划分为多个命名空间。一个命名空间就是一个逻辑分区，它包含对应一个或多个应用的一组特定对象。每一个命名空间都是由独立目录结构表示的专用对象存储库，都拥有一套经过独立配置的属性。命名空间提供数据隔离功能，而租户或者命名空间的分组则提供管理隔离功能。一个对象存储系统可以拥有多达 1,000 个租户。每一个租户及其命名空间都由虚拟对象存储系统构成，可以独立于用户和应用进行访问和管理。这种特点在企业、云和服务供应商环境中至关重要。

根据访问协议的类型和配置，数据访问命名空间可以经过验证，或者不经过验证。采用本地账户或者 Microsoft Active Directory® 群组可以执行验证。

对象版本控制

内容平台软件支持对象版本控制，这是一种命名空间功能，可以创建、存储和管理资源库中多个版本的对象。这种功能可以提供数据变更的历史记录。版本控制可以支持演变内容的存储和复制，因此为内容平台软件在内容仓储和工作流应用等市场开辟了全新的机遇。

版本控制适用于命名空间，可在命名空间等级进行配置。版本控制仅支持 HTTP 或者 REST。如果命名空间启用了版本控制，则无法支持其他协议。版本控制仅适用于对象，不适用于目录或者符号链接。当具有相同名称和位置的对象作为现有对象被添加到命名空间时，便会创建一个新版本的对象。一种特殊类

型的版本称为删除版本，当删除对象会创建删除版本。对象元数据的更新只会影响当前版本的对象，不会创建新的版本。

如果以前版本的对象比指定的时间更久，则会自动删除或者修改。不可能删除对象的特定历史版本；但是，拥有适当许可权限的用户或应用可以清除对象，删除所有版本，包括当前的版本在内。

存储分层

内容平台软件可以支持存储分层，以作为支持信息生命周期管理 (ILM) 和智能对象这一长期目标的初期发展步骤。根据客户指定的存储分层策略，可以把需要长期存留的数据进一步存放到公有云当中去。

搜索

内容平台软件构建的对象存储系统可以提供集成的元数据查询引擎。包括各种综合全面的搜索功能，支持用户搜索命名空间中的对象，分析命名空间内容，以及操作对象组。为了满足监管需求，支持面向审计和诉讼的电子取证。

自定义元数据可以提供快速、准确访问内容的工具，高效和智能地提供所需要的有用信息，可以支持构建大量的非结构化数据仓库，并且应用正确的对象策略，满足所有业务、监管达标和保护的需求。监管达标功能包括命名保留模式（监管达标和企业）、保留类别、保留控制、自动内容处置和特权删除与清除。内容平台软件的搜索功能包括支持诉讼或审计用途的电子取证。采用内容平台软件，开放式 API 允许第三方直接集成。

内容平台软件提供交互式界面的搜索工具。搜索控制台可以提供结构化环境，用于创建和执行查询（在搜索结果中每一个对象必须满足一系列标准）。用户可以应用各种选择标准，例如特定日期之前或者大于指定规格的存储对象。查询返回的对象元数据包括在搜索结果之中。这种元数据可以用于检索对象。通过

搜索控制台，用户可以打开对象，执行对象的批量操作（保持、释放、删除、清除、特权删除和清除、变更所有者、设置 ACL），并且以标准文件格式输出搜索结果，用作为其他应用的输入。

内容平台软件内嵌元数据查询引擎，它也可以通过元数据查询 API 直接使用，这是一种用于查询命名空间的编程接口。元数据查询引擎的索引驻留在存储节点的指定逻辑卷，根据系统类型和卷配置，与对象数据分享或不分享这些卷的空间。

租户级和命名空间级可以启用搜索。每个命名空间都可以启用索引。系统级和命名空间级的设置决定了除系统元数据和 ACL 以外，是否对自定义元数据进行索引。如果禁用自定义元数据索引，索引则不包括自定义元数据。如果命名空间完全没有编入索引，搜索则不会返回任何这个命名空间中对象的结果。

每一个对象都有一个索引设置，它对元数据查询引擎的索引内容具有不同的影响。如果启用了命名空间的索引，元数据查询引擎一直会索引系统元数据和 ACL，与对象的索引设置无关。如果索引设置为真，元数据查询引擎也会索引这个对象的自定义元数据。

复制

复制是内容平台软件的附加功能，是在 2 个或者多个内容平台软件构成的对象存储系统中，将选定的租户和命名空间与其他租户和命名空间保持同步的过程。复制服务可以将一个或多个租户或命名空间从一个内容平台软件系统拷贝到另一个内容平台软件系统中，支持对象创建、对象删除和元数据变更。内容平台软件还可以复制租户和命名空间配置、租户级用户账户、监管达标和租户日志消息与保留类别。

最初创建对象的内容平台软件系统被称为主系统。第 2 个系统被称为复制系统。通常而言，主系统和复制系统处于不同的独立地理位置，通过高速广域网络相联。内容平台软件可以支持不同的复制拓扑技术，包括多对一复制和链式复制等。

降低数据管理成本

内容平台软件构建的对象存储系统省去使企业数据系统时刻处于生产工作状态的复杂和昂贵的管理成本。对象存储根据数据生命周期内的价值和不同阶段形成合理的层级和保护级别，降低生产系统数据量，达到降低数据管理成本的目的。

提供更好的数据保护，更高的数据可用性

内容平台软件构建的对象存储系统支持被复制成多个副本以保证数据可靠性。如果其中一份数据出现问题，数据恢复后台自动运行，而且数据可用性不会中断，性能也不会明显退化。

提供高可扩展性

内容平台软件构建的对象存储系统中，没有目录层次结构(树)，对象的存储通过 ID 值管理数据，而且不受文件(对象)数量、文件大小和文件系统容量的限制。文件系统无法实现的元数据利用，但是对象存储系统可以不需要文件名、日期和其他文件属性就可以查找文件，通过内置的搜索引擎实现数据快速定位。

自动负载均衡

对象存储集群几乎是完全对称的。每个节点都是独立的，提供了集群的切入点，并运行相同的代码。这使得工作量可以平均分配到集群中的所有节点上，避免 NAS 和集群文件系统中常见的热节点问题的出现。自动负载均衡可以让 I/O 自动选择合理的节点，保证系统性能最大化。

升级及数据迁移简单

在对象存储架构中，可以免却传统硬件移植或者大规模硬件升级的麻烦。对象存储结构只需要采用常规移植就可以实现。整个过程中都可以随时添加新的设备并会自动加载到集群中，而旧的设备单元也可以解除。

设备更换简单

根据存档和法规要求，存储的数据需要保持数年。技术更新的成本和复杂性是一个需要考虑的重要因素，特别是连接到昂贵的专有硬件平台系统，对象存储提供平滑设备更新工具。

软件升级无需停机

内容平台软件采用大量技术将软件升级期间的正常系统功能中断降到最小程度，或者将其完全排除。不停机的软件升级 (NDSU) 属于这些技术之一，它包括大幅增强的在线升级支持、不停机的补丁管理和在线升级性能改善等功能。内容平台软件支持无介质的远程升级、HTTP 或 REST 排出模式和并行操作系统 (OS) 安装。它还可以支持自动在线升级委托、离线升级持续时间预估、增强监控和电子邮件警报和其他功能。

无缝应用故障切换

这种功能通过多种复制拓扑技术来实现。该功能包括无缝故障切换路由选择功能，它允许采用复制拓扑技术的系统维护 HTTP 请求，可以支持与所有客户所有的负载均衡器进行直接集成。无缝域名系统 (DNS) 故障切换是一种内容平台软件内置的多站点负载均衡和高可用性技术，十分适用于经济高效、全力为客户服务的环境。

更高的磁盘利用率

相比块存储，对象存储可以提供更高的磁盘利用率。

无需备份

对象存储系统并不需要备份。如果需要的话，多个副本可以确保数据始终保持可用状态，而且异地灾难恢复备份也可以被自动创建。一旦主集群数据不可用，可以使用备份数据。因为集群中所有内容的副本中所存储的是一致的。而这些在文件系统中是几乎不可能发生的，它们需要克服繁琐的备份窗口和既漫长又艰难的备份还原操作。

2.4. 免备份功能详细描述

需要数据长期保存并满足行业合规需求，这需要建设一个专业的、更加安全可靠的数据长期保存平台。

对象存储可以实现对数据安全的存储和保护，可以更好的集中保护和长期保留企业数据。那么智能的数据保护是如何实现的呢？

数据保护/备份往往是为了在发生数据丢失的时候可以去恢复数据，那么那些情况下数据会丢失呢？根据第三方统计的结果，数据丢失的场景包括系统故障 44%、人为错误 32%、软件损坏 14%、计算机病毒 7%和自然灾害 3%。我们也可以简单的将其分类为物理故障、逻辑错误和灾难，而在这些情况下，对象存储系统都会通过其智能的数据保护方案去保障您的数据。

对象存储系统并不需要备份。如果需要的话，多个副本可以确保数据始终保持可用状态，而且异地灾难恢复备份也可以被自动创建。一旦主集群数据不可用，可以使用备份数据。因为集群中所有内容的与副本中所存储的是一致的。除了物理上的故障，增对逻辑错误，具备版本管理、WORM 等有效保护手段，可以设置数据的保存周期，以实现保存周期内的强合规需求。

内容平台软件具有免备份的特点。内容平台软件系统无需备份便可以全方位的保护其中的内容。它利用复杂的数据保存技术，例如可配置数据和元数据保护级别 (MDPL)、对象版本控制和变更跟踪、无缝应用故障切换的多站点复制以及其他技术等。内容平台软件通过多种功能组合，可以保证数据的完整性和一致性，并且确保存储数据的可用性和安全性。以下归纳概括了内容平台软件关键的数据保护功能：

多副本

虽然单个存储节点整体发生故障的概率很小，但并不意味着不重要。内容平台软件构建的对象存储系统后端可以挂接多个存储池，而独有的数据多副本技术可以将数据存放到不同的存储池，可根据业务需要设置副本保护级别(DPL)，比如设置 DPL=2 则保存 2 个数据副本，通过多副本即使极端情况单个存储池故障也不会丢失数据且对应用透明。内容平台软件可以利用软件镜像将每一个对象的数据存储在不同节点的

多个位置。内容平台软件将存储节点分为各个保护集合，每个集合都带有相同数量的节点，并且试图在单一保护集合中存储一个对象的所有数据拷贝，在这个集合中每一个拷贝都被存储在不同的节点。通过检查和修复保护集合，保护服务会强制执行所需的数据冗余度等级。如果违反规定，它会创建额外的拷贝，或者删除对象的多余拷贝。如果启用了复制，如果主系统的拷贝不可使用，保护服务可以采用复制系统的对象拷贝。

除了 DPL 指定数据冗余保护之外，内容平台软件还会在不同节点创建多份元数据的拷贝。元数据保护等级或 MDPL 是一种指定元数据拷贝数量的系统范围设置，内容平台软件系统必须维护这种设置（通常为 2 个拷贝、MDPL2）。MDPL 冗余度的管理独立于数据的冗余拷贝管理 DPL。

清理服务可以确保资源库中的所有对象都具有有效的元数据。当元数据丢失或损坏时，该服务将会利用元数据副本（与对象数据每一个拷贝一起存储的元数据拷贝）尝试重建元数据。

WORM

逻辑错误往往是指人为误操作或者蓄意的篡改/删除、病毒的恶意破坏等，那么怎样保护呢？最简单的答案从底层存储直接不允许修改和删除，WORM 一次写多次读技术正是这样的技术，系统可以根据应用系统或者基于单个文件设置其保存周期，通过 WORM 防止在保存周期内的数据误操作或恶意篡改，保证数据完整性。内容平台软件经过优化，适用于固定内容的数据归档。固定内容数据是不变更但需要长期保存的信息，以供日后需要时可以轻松访问。固定内容存储系统是一个不可修改数据的系统。内容平台软件采用“一次写入、多次读取”（WORM）的存储技术和多种策略与服务（例如保留、内容验证和保护等），以确保资源库中的数据完整性。WORM 存储代表一旦数据进入到资源库，便不可更新或修改，保障数据从最初存储时便一直保持不变。如果内容平台软件系统启用了版本控制功能，则会存储并恢复不同版本的数据，在这种情况下每一个版本均为 WORM 状态。

版本控制

内容平台软件构建的对象存储系统支持对象版本控制 (Versioning)，这是一种命名空间功能，可以创建、存储和管理资源库中多个版本的对象。这种功能可以提供数据变更的历史记录。版本控制可以支持演变内容的存储和复制，因此为内容仓储和工作流应用等市场开辟了全新的机遇。

版本控制适用于命名空间，可在命名空间等级进行配置。版本控制仅支持 HTTP 或者 REST。如果命名空间启用了版本控制，则无法支持其他协议。版本控制仅适用于对象，不适用于目录或者符号链接。当具有相同名称和位置的对象作为现有对象被添加到命名空间时，便会创建一个新版本的对象。一种特殊类型的版本称为删除版本，当删除对象会创建删除版本。对象元数据的更新只会影响当前版本的对象，不会创建新的版本。

如果以前版本的对象比指定的时间更久，则会自动删除或者修改。不可能删除对象的特定历史版本；但是，拥有适当许可权限的用户或应用可以清除对象，删除所有版本，包括当前的版本在内。

数据的自校验和自修复

对于合规性要求很严苛的数据，数字指纹可以作为数据原始性和真实性的凭证。当文件存入内容平台软件构建的对象存储系统时，系统针对每个文件会产生唯一的 Hash 值（即数字指纹），同时系统还会不断在后台运行内容验证服务，确保每一个文件的所有副本数据都与这个唯一的哈希值相匹配，从而保证写入和读出文件的一致性、原始性和真实性。

另外，通过多副本机制产生的多份副本数据，如果发生数据损坏，系统也可以实现自动修复。内容平台软件构建的对象存储系统首先通过内容验证服务执行数据的自校验，以判断是否有文件的某个副本发生损坏，如果副本受损（例如发生了比特错误），那么可以自动通过完好副本覆盖受损副本实现数据的自修复，无需人为干预从而节省恢复的时间和人力。

站点复制

如果启用了这项功能，它可以提供大量可以确保数据可用性的机制。复制系统可以作为灾难恢复的来源，也可以提供良好的对象拷贝用于保护和内容验证服务，以便维护数据可用性。如果无法通过主系统读取对象，假如启用了复制读取功能，内容平台软件则会尝试通过复制系统读取对象。

复制是内容平台软件构建的对象存储系统的附加功能，是在 2 个或者多个对象存储系统中将选定的租户和命名空间与其他租户和命名空间保持同步的流程。复制服务可以将一个或多个租户或命名空间从一个对象存储系统拷贝到另一个对象存储系统中，支持对象创建、对象删除和元数据变更。系统还可以复制租户和命名空间配置、租户级用户账户、监管达标和租户日志消息与保留类别。

最初创建对象的系统被称为主系统，第 2 个系统被称为复制系统。通常而言，主系统和复制系统处于不同的独立地理位置，通过高速广域网络相联。系统可以支持不同的复制拓扑技术，包括多对一和链配置，也可以设置为双活和多活架构，应用可以同时访问主系统和复制系统实现跨站点全局访问和读写负载均衡。

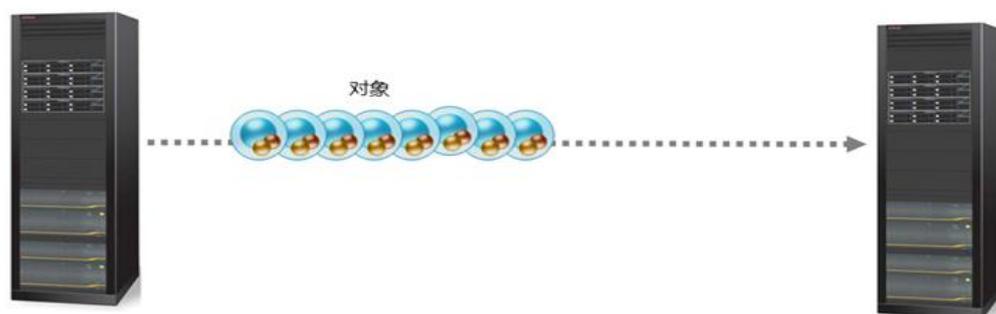


图 4：对象存储系统主/从复制架构

当有 3 个或 3 个以上站点时，采用多站点复制显然存储效率不高，会带来很多的空间开销。系统提供多站点纠删码技术，当数据写入时可以根据纠删码技术做出 N+1 的保护单元，应用可以从任意站点全局访问数据，而当其中一个站点故障时不影响数据的可用性。

综上，我们可以看到通过全面的数据保护技术，系统可以保证在发生物理故障、逻辑错误情况下数据均不丢失，而且通过这些智能的数据保护技术，解决了传统非结构化数据备份时间、恢复时间、成本等巨

大挑战。同时，在数据长期保存的情况下，系统可以提供保存周期的灵活配置，甚至可以根据文件设置及保存周期，以满足合规及长期保存的需求。

3. 内容平台软件架构详细介绍

3.1. 基于云计算的存储系统

基于内容平台软件的基础架构解决方案具备充分的灵活性，可以容纳所有云部署模型（公共云、私有云或混合云），并且简化服务供应商和订阅用户的云迁移过程。内容平台软件可以提供边缘到核心、安全多租户和强大的管理功能，以及优化云存储操作的众多功能。

内容平台软件作为在线数据资源库，真正为云市场需求做好了充分准备。尽管在这份白皮书中之前已经探讨了大量的内容平台软件功能，本部分内容的目的在于归纳这些内容平台软件的云功能。它们包括：

大规模多租户

管理隔离：内容平台软件支持多达 1,000 个租户，每一个租户都可以经过唯一配置，提供给单独的云服务订阅用户使用。

数据隔离：内容平台软件支持多达 10,000 个命名空间，每一个命名空间可以经过唯一配置，用于特定的应用或工作负载。

大规模扩展能力

内容平台软件构建的单个对象存储系统可以支持百 PB 级的存储容量，即使单个桶即也可支持 1000 亿个文件，每个目录 1500 个用户，全部都位于单一的物理系统之内，无需构建分散的存储系统和创建众多文件系统，即可轻松满足您的海量数据存储需求。

每个节点支持 12.5 亿个对象。由于节点的数量更少，内容平台软件降低了所需的电力、冷却和占用空间。

可以按需添加节点，扩展系统的存储和吞吐量，不会产生中断。多个存储节点受单一内容平台软件系统的管理。

轻松的租户与存储预配置。

地理分散性和全球可访问能力

WAN 友好 REST 界面，便于命名空间数据访问和复制。

采用先进、灵活的复制拓扑技术实现多个站点的内容复制。

WAN 优化的高吞吐量数据传输。

多种 REST 界面

这些界面包括用于命名空间数据访问的 REST API、管理 API 和元数据查询 API。REST API 是一种为云支持供应商和消费者提供的技术，它之所以普及，原因包括高效率、低开销、客户端和服务器的缓存以及 API 一致性。除此之外，这种技术可以提供无状态特性，可以接受互联网接入的延迟和潜在的复杂防火墙配置。

安全、细粒度地访问租户、命名空间和对象

这对所有云环境都至关重要。这种访问权由内容平台软件多层次、灵活的许可机制所支持，其中包括对象级 ACL 在内。

使用量计量

内容平台软件具有内置的内部计费 (chargeback) 功能，这对于云的使用不可或缺，可以促进供应商和订阅用户的交易。内容平台软件还可以为第三方供应商和客户提供工具，编写 API，从而进行计费和报告。

自动监控、自动管理和自愈功能

内容平台软件包含高级监控、审计和报告功能。如果出现问题，内容平台软件服务可以自动修复。

支持多种服务级别

通过内容平台软件策略、服务规划和配额提供这一支持，为每一个租户进行配置，有助于强制执行服务水平协议 (SLA)。它让该平台可以采用单一系统支持范围广泛的用户和商业模型。

3.2. 系统软件架构

软件概述

系统软件由操作系统（装置操作系统 AOS）和核心软件（Core Application）组成。核心软件包括以下部分任务：

- 通过业界标准的 HTTP 或者 HTTPS、S3、SWIFT、WebDAV、CIFS、NFS、SMTP 和 NDMP 协议，支持对象资源库的访问。

- 输入固定内容数据，将其转化为对象，并且根据时间进度管理对象数据和元数据。

- 强制执行资源库策略并且执行系统服务，维持存储数据的完整性、稳定性、可用性和安全性。

- 通过可读界面，支持系统的配置、监控和管理。

- 通过交互式 Web 界面（搜索控制台）和编程界面（元数据查询 API），支持资源库搜索。

AOS 和 Core Application 均内置了以下核心软件包：

FUSE(Filesystem in Usespace)：将其用于 RISFS 中。

Device-Mapper-multipath

Samba

Jetty:HTTP 服务组件和 servlet 容器,用于提供HTTP/WebDAV 网关以及 Web 管理控制台服务,另外 MAPI 和搜索控制台服务也由它提供

PostgreSQL: 每个节点的软件部件部分相同,从软件功能可分为访问层(Access Layer)、事务和管理层(Transaction & Administration Layer)、核心组件层(Core Component Layer)、系统接口层(System Interface Layer)、操作系统层(OS)。每层的功能如下:

- 访问层 (Application Layer) 提供用户、用户和对象资源的交互能力,各种协议网关在该层实现,包括HTTP 网关、Admin UI、文件系统访问接口等
- 事务和管理层:接受来自访问层的处理请求。包括Request Manager(RM, 为JVM)和 Administrative Engine(AE, 为JVM)两个组件。RM 确保所有的读写请求在一个已知的确定状态。AE 主要提供系统层面的监控,如监控系统状态,资源使用率以及协调系统间的行为等。
- 核心组件层: 包括Metadata Manager (MM) 、 Storage Manager(SM)、 DNS Manager、 Protection Manager、 Search Manager、 Medata Query Engine (MQE) 、 Solr Manager、 Service Engine 等组件
- 系统接口层: 主要为C 组件,为访问节点和系统提供配置、状态和健康维护服务,包括 Volume Manager(volman)、SLAB (System Level Archive Backplane, 在节点间维护系统状态信息)、 Sysmon (负责监控OS 状态,并在节点间传递和共享这些信息)、 Timeload (负责同步节点时钟)、 Guard (确保组件工作正常,如果失败,会尝试重启服务进程)、 IPMI 等组件。

专用文件系统 CHDP-FS

长虹内容平台软件的文件系统为专用的文件系统 CHDP-FS,是一个标准的 POSIX 文件系统,它模拟成一个 Native 文件系统,提供到对象空间的完全访问,主要作用是允许用户和应用像常规文件、目录和符号链接一样浏览对象,并按以往熟悉的操作方式处理这些对象,它将对象和 metadata 以传统的目录结构和文件方式呈现。文件系统访问接口由 CIFS/NFS 网关提供。CIFS 使用 smbld 守护进程,NFS 使用 nfsd 守护进程。

服务引擎

内容平台软件构建的分布式对象存储系统包括 13 个后台服务引擎,每个引擎实现不同的作用,在后台运行:HCP 软件实施了 13 种后台服务,如表 1 所列。这些服务用于改善 HCP 系统的整体健康状况,优化效率,并维护存储的对象数据的完整性和可用性。服务可以连续、定期(按规定时间表)运行,也可以响应某些事件。系统级管理员可以启用、禁用、启动或停止任何服务,并控制每项服务的优先级或时间表。这些控制功能包括较长时间运行服务,单独运行或为其指定更高的优先级。通过指定简单的高、中和低等级,可通过限制服务生成的线程数而控制运行时系统的加载。

所有预定的服务同时但自主地运行,因此,每个服务可以同时元数据数据库的不同区域工作。每个服务迭代所存储的内容,并最终检查每个存储对象的元数据。在新的 HCP 系统上,每个服务都安排在某些天的特定时间段运行。如果特定服务在指定的时间段内完成完整扫描,则服务将停止。如果未完成,则服务将在下一个预定时间段停止。完成计划的扫描间隔后,服务会在 HCP 系统事件日志中发布摘要消息。

- **保护:** 执行数据保护级别 (DPL) 策略合规性检查,确保系统中每个对象的副本数正确。根据定义的 DPL 设置检查和维护数据冗余,每天运行一次。
- **内容验证:** 通过确保文件与其数字哈希签名匹配而保证存储库对象的数据完整性。如果哈希值不匹配,会自动修复对象。这一操作还检测并修复元数据差异。在每个存储节点

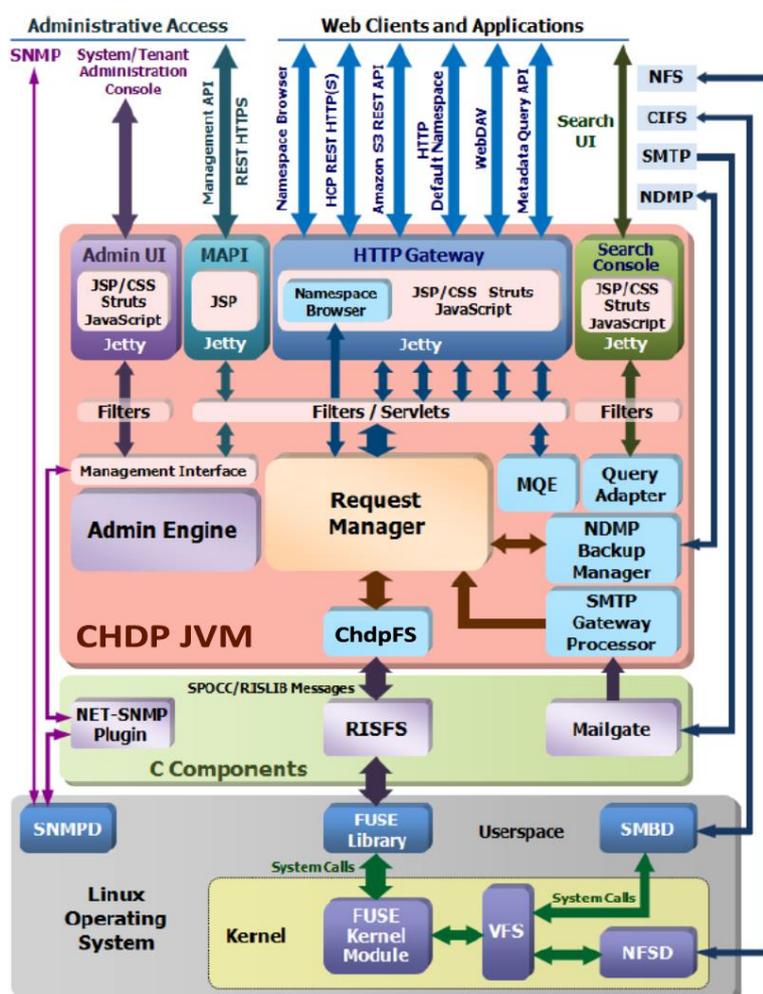
缺省每周运行一次。

- **清除：**确存储库中的所有对象都包含有效的元数据，并在元数据丢失或损坏的情况下重建元数据。每周运行一次。
- **垃圾采集/回收站回收：**通过清除标记为已删除的对象的隐藏数据和元数据而回收存储空间，或者由于未完成的事务（未关闭的NFS或CIFS文件）而遗留存储空间。比如在删除租户之前，需要把该租户下的所有命名空间中的对象清除，清除后再运行此服务可彻底清除垃圾文件。
- **重复数据删除：**识别并消除存储库中的冗余对象，并将重复数据合并到可用空间。
- **粉碎：**文件粉碎功能。出于安全原因，覆盖已删除对象的副本存储位置，使其数据或元数据都不能被重建。这也称为安全删除。默认的粉碎算法使用三次传递来覆盖对象。
- **处置：**自动清理过期对象。命名空间配置策略授权系统在保留期到期后自动删除对象，每天运行。
- **压缩：**压缩对象数据，更有效地利用物理存储空间。
- **容量平衡：**尝试在系统中的所有存储节点上保持可用存储容量平衡。如果节点的存储利用率相差很大，则服务会移动对象，以使节点更接近平衡状态
- **存储分层：**确定哪个存储分层策略适用于对象，根据应用的服务计划中的规则评估对象副本应驻留的位置。
- **迁移：**执行数据迁移策略，从所选节点中迁移数据，将这些数据清除。
- **复制：**将一个或多个命名空间从一个对象存储系统复制到另一个对象存储系统，确保数

据可用性并实现灾难恢复。

数据处理流程

当内容平台系统接收到用户的请求时，它会路由到一系列软件组件进行处理，例如对于一个 NFS 请求，内容平台的处理流程如下：NFS (kernel) <=> VFS (kernel) <=> FUSE (kernel) <=> FUSE Library (userspace) <=> RISFS (userspace) <=> ChdpFS (JVM)<=> Request Manager (JVM) <=> Storage Manager (JVM) ，参考如下示意图（简化）：



系统架构

内容平台软件构建的对象存储系统是一个完全对称的分布式应用，可以存储并管理对象。对象包含原始的固定内容数据，由客户端应用写入，还包含相关的系统元数据和自定义元数据，元数据分布在所有访

问节点，参见下图。内容平台软件构建的对象存储系统的每一个节点都是一台 Linux 服务器，运行完整的实例。系统可以承受多个同步节点故障，并且自动操作，确保所有对象和命名空间策略的有效性。

外部系统通信由 DNS 管理器进行管理，这是一种分布式网络组件，用于平衡所有节点内的客户端请求，以便确保获得最大的系统吞吐量和可用性。DNS 管理器与企业 DNS 服务器结合使用，允许客户端作为一个单独实体访问系统，即使系统由多个独立节点构成。

内容平台软件系统配置为现有企业域的子域。客户端采用预定义的特定协议或特定命名空间名称访问系统。

尽管并非必需要求，采用 DNS 非常重要，这可以确保客户端均衡且无障碍地访问内容平台软件系统，特别是对于 HTTP 或 REST 客户端而言。

命名空间和租户

资源库被划分为命名空间。一个命名空间可以视为应用的一个逻辑资源库。每一个命名空间都由带有自身目录结构的不同逻辑对象组构成，这样位于一个命名空间中的对象在其他命名空间中是不可见的。访问一个命名空间不会授权用户访问其他命名空间。对于命名空间的用户而言，命名空间便是资源库。命名空间与预先分配的存储没有关系；它们共享相同的底层物理存储。命名空间可以提供一种机制，将存储在不同应用、业务部门或客户的数据实现隔离。例如，可以建立一个“应收帐款”的命名空间，再建立一个“应付帐款”的命名空间。尽管一个单独的命名空间可以托管一个或多个应用，通常它只托管一个应用。命名空间也可以针对选定的资源库对象子集进行操作。例如，搜索可以定位于“应收帐款”和“应付帐款”，而不选择“员工”命名空间。

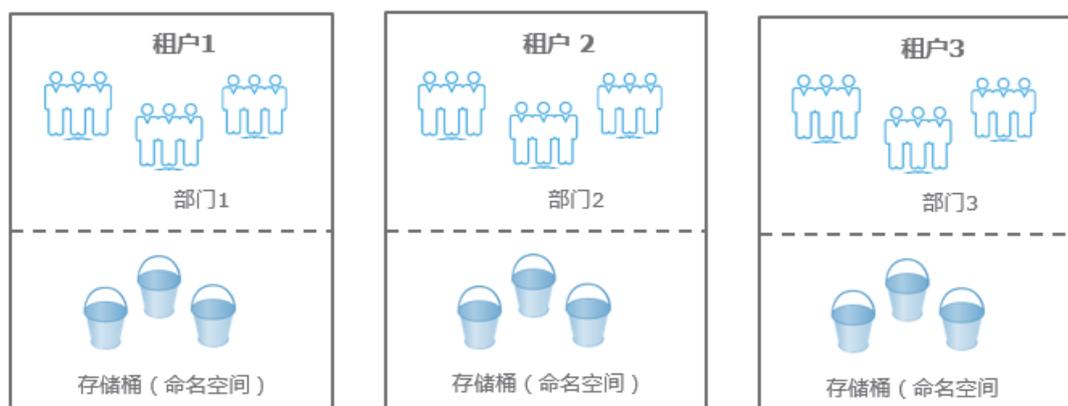


图 5：命名空间和租户

命名空间由租户所有并且管理。租户是提供管理隔离的管理实体，而命名空间则提供数据隔离。一个租户通常代表了一个实际的企业，例如使用资源库部分内容的一家公司或者一家公司的一个部门。命名空间管理在自有的租户级别上完成。

客户端可以通过 HTTP 或者 HTTPS、WebDAV、CIFS、NFS 和 SMTP 访问命名空间。这些协议可以支持已验证和/或匿名类型的访问（访问类型及其组合在本文后半部分会详细讨论）。命名空间由租户所有。一个对象存储系统可以拥有多个租户，每一个租户都拥有多个命名空间。每一个租户拥有的命名空间数量都会受到管理员的限制。

用户与组账户

用户与组账户可以控制不同界面的访问权限，为用户提供许可执行管理任务并且访问命名空间内容。

用户账户拥有一组证书、用户名和密码，存储在本地系统中。内容平台软件系统使用这些证书为用户进行验证，执行本地验证。

一个组账户代表一个 Active Directory (AD) 组。要创建组账户，内容平台软件必须进行配置，支持 Active Directory。组账户可以支持 AD 组中的 AD 用户，访问一个或多个界面。

如同用户账户一样，组账户可以分别以系统级和租户级进行定义。不同的租户拥有不同的用户和组账户。这些账户不可在租户间共享。组成员拥有不同的系统级和租户级。

管理角色与系统级和租户级用户，以及组账户相关。数据访问许可仅与租户级用户和组账户相关。因此，系统级本地和 AD 用户只是管理用户；而租户级本地和 AD 用户则既是管理用户，也拥有数据访问许可权限。租户级用户只是管理角色，没有命名空间数据许可权限；或者仅有命名空间数据许可权限，不是管理角色；或者为管理角色和命名空间数据许可权限的任意组合。

系统和租户管理

系统拥有系统级和租户级管理员：

系统级管理账户用于配置系统功能、监控系统软件以及整体资源库使用率，管理系统级用户。系统管理员用户界面、系统管理控制台可以提供系统维护人员所需的功能。例如，它允许维护人员关闭系统，查看节点的信息，管理策略和服务，并创建租户。系统管理员拥有系统整体视图，该视图包括了所有运行系统的软件组件，可以执行系统范围操作范围内的所有管理任务。

租户级管理账户用于创建命名空间。它们可以配置单独的租户和命名空间，监控租户级和命名空间级的命名空间使用率，管理租户级用户，并且控制命名空间的访问权限。这种必要的功能由租户管理员用户界面、租户管理控制台提供。该界面由虚拟系统的维护人员（拥有一组命名空间的单独租户）。租户级管理功能可以支持管理隔离，这对云环境十分重要。

租户可以选择为系统级用户授权访问自身的管理权限。在这种情况下，拥有监控、管理员、安全或监管角色的系统级用户可以登录进入租户管理控制台，或者使用该租户的管理 API。拥有监控或者管理员角色的系统级用户也可以通过系统管理控制台直接访问租户管理控制台。这可以有效支持系统管理员执行租户管理员的职责。系统级用户可以执行与其系统级角色相对应的租户级角色所允许的所有操作。AD 用户

属于 AD 组，相对应的组账户同时存在于系统级和租户级。该用户的角色与适用的系统级组账户和适用的租户级账户都有关系。

策略

一个命名空间中的对象具备各种各样的属性，例如保留设置或索引设置。这些属性都是由对象系统元数据为每一个对象定义而成。对象也可以被一些命名空间属性所影响，例如由存储在命名空间中的新对象继承的默认元数据设置或者版本控制设置等。属于对象元数据组成部分的命名空间级设置和属性可以作为系统的事务处理和服务参数，并且决定资源库生命周期期间的性质。这些设置被称为策略。

策略是一种或多种设置，它们可以对处理事务和内部流程（服务）对命名空间中的对象施加影响。策略可以确保对象以预期的方式运行。每一种策略都由一个或多个设置组成，设置具有不同的应用范围和配置方法。在对象级和命名空间级定义策略设置。注意根据命名空间，可以在不同的级别设定相同的策略设置。默认的保留、粉碎和索引设置在命名空间级进行设定。

内容平台软件策略的具体描述请见表 1：

表 1. 内容平台软件策略

策略名称	策略描述与组成部分	受影响的事务处理和服务
DPL	系统 DPL 设置、命名空间 DPL 设置。	对象创建、保护服务。
保留	默认保留设置、对象保留设置、保持设置、处于保留状态下对象的系统元数据和自定义元数据选项。	对象创建、对象删除、系统和自定义元数据处置、处理、垃圾采集服务。
粉碎	默认粉碎设置、对象粉碎设置。	对象删除、粉碎服务。
索引	默认索引设置、对象索引设置。	MQE。
版本控制	版本控制设置、修改设置。	对象创建与删除、垃圾采集服务。

自定义元数据验证	XML 语法验证。	添加/替换自定义元数据操作。
----------	-----------	----------------

表 2 列出了根据各自范围和配置方法所分类的所有策略设置。

表 2.内容平台软件策略设置：范围和配置

策略	策略设置	内容平台软件命名空间	
		范围/等级	配置方式
数据保护等级	系统 DPL: 1-4	系统	系统 UI
	命名空间 DPL: 1-4, 动态	命名空间	租户 UI, MAPI
保留	默认保留设置: 指定日期、偏移、特殊值、保留类别	命名空间	租户 UI, MAPI
	保留设置: 指定日期、偏移、特殊值、保留类别	对象	REST API, retention.txt
	保持设置: 真或假	对象	REST API
	保留期间的所有权和 POSIX 许可变更: 真或假	命名空间	租户 UI, MAPI
	保留期间允许的自定义元数据操作	命名空间	租户 UI, MAPI
索引	索引设置: 真或假 (1/0)	对象	REST API, index.txt
	默认索引设置: 真或假	命名空间	租户 UI, MAPI
粉碎	粉碎设置: 真或假 (1/0)	对象	REST API, shred.txt
	默认粉碎设置: 真或假	命名空间	租户 UI
自定义元数据验证	XML 验证: 真或假	命名空间	租户 UI

版本控制	版本控制设置：真或假	命名空间	租户 UI
	修改设置：真/假和主系统或复制系统的天数	命名空间	租户 UI

内容管理服务

后台服务就是旨在保护并提高对象存储系统整体性能正常而执行特定功能的一个后台流程。特别是，这些服务负责优化系统资源的使用，并且维护对象存储资源库储存数据的完整性和可用性。

系统可以执行 12 种服务：保护、内容验证、清理、垃圾采集、重复数据删除、粉碎、处置、压缩、容量平衡、存储分层、迁移和复制。

内容平台软件各项服务在表 3 中有简要介绍。

表 3. 内容平台软件服务

策略	说明
保护	确保系统中每一个对象的适当拷贝数量，强制执行 DPL 策略监管达标要求，并恢复损坏或丢失的对象。任何违规操作都会引发修复流程。提供定期和事件驱动的服务。事件触发全体服务运行，即使服务处于禁用状态，配置时间：节点关闭 90 分钟之后；逻辑卷故障 1 分钟后；节点移除 10 分钟之后。
内容验证	确保文件内容与其数字签名相互匹配，保证资源库对象的数据完整性。如果哈希值不匹配，修复对象。检测并修复主系统元数据与次系统元数据之间的不符之处。默认采用 SHA-256 哈希算法。计算外部和内部文件的校验和。运算密集型和耗时型服务。根据主动服务计划表运行。
清理	确保资源库的所有对象都拥有有效的元数据，并在元数据丢失或损坏时可以重建元数据，但数据文件仍然存在。该服务可以验证每个数据对象的主元数据以及与对象数据存储在一

	起的元数据拷贝是否完整、有效，并且相互同步。运算密集型和耗时型服务。定期服务。
垃圾采集	清除删除标记或者不完全事务遗留的隐藏数据和元数据，回收存储空间。它还可以删除旧版本的对象，以便进行修改。在适用条件下，删除会触发粉碎服务。定期服务，非事件驱动。
重复数据删除	发现并删除资源库中的冗余对象，合并重复数据，释放空间。外部文件表达式的哈希签名用于选择对象作为该服务的输入。之后这些对象会经过字节检查，以确保数据内容确实相同。定期服务。
粉碎	出于安全原因，被删除对象的拷贝存储的位置无法重建数据或元数据，因此需要重写存储位置。这也被称为安全删除。默认内容平台软件粉碎算法采用 3 种途径重写对象，并且符合 DoD 5220.22-M 标准要求。在安装时间选择算法。只是事件驱动的服务，非定期服务。由于删除对象的粉碎标记所触发。
处置	自动清除到期对象。所有命名空间经过配置，在对象保留期限到期之后都会自动删除对象。可以在系统级和命名空间级启用或禁用该服务；如果在系统级禁用该服务，则对命名空间的处置服务没有影响。处置服务只删除当前版本的对象。定期服务。
压缩	压缩对象数据，更加高效地利用系统存储空间。通过压缩回收的空间可以用于额外的存储。通过系统管理控制台可以提供大量的可配置参数。定期服务。
容量平衡	力图保持系统所有存储节点内存储容量的平衡（大致相等）。如果节点的存储利用率出现大范围差异，该服务会将对象移到更近的节点，以达到平衡状态。只能在人工启动时运行。对象的添加和删除不会触发该服务。通常情况下，授权服务提供商在将新的存储节点添加到系统之后会启动这项服务。除此之外，尽管并不属于该服务的一部分，在系统正常运行期间，新对象都会相当平均的比例自然分布在所有存储节点之中。原因在于：存储管理器选择算法的性质和管理引擎的资源监控。
存储分层	确定哪一个存储分层策略适用于对象，评估对象拷贝基于适用的服务计划应当驻留在何处，并且按照需要移动对象。定期服务。

结论

内容平台软件构建的对象存储系统解决方案可以采用更大数量和更加高效的方式智能存储内容，避免了传统文件系统的局限性。这些解决方案可以满足非结构化数据产生的新需求，对企业、合作伙伴、客户、政府及其股东的重要性

内容平台软件构建的对象存储系统解决方案将文件数据、文件元数据和自定义元数据视为一个单独的对象，在各种存储分层间进行跟踪和存储。凭借每个逻辑分区的安全多租户和可配置属性，对象存储可以划分为大量的小型虚拟对象存储库，其具有的可配置属性可以支持不同的服务水平。这可以让对象存储库支持范围广泛的工作负载，例如内容保存、数据保护、内容分配，甚至单一物理基础架构的云环境。对于每一种应用或一系列用户而言，与孤立的技术孤岛相比，单一的基础架构更加便于管理。通过将众多关键技术集成到单一的存储平台之内，内容平台软件构建的对象存储系统对象存储解决方案可以提供短期投资回报途径，以及明显的长期效率提升。它们帮助 IT 不断发展，从而应对全新的挑战，长期保持灵敏性，并且满足未来的变化和增长需求。

CHANGHONG

北京

北京市丰台区南四环西路188号18区26号楼长虹科技大厦

邮编：100738

电话：010-58292000

传真：010-58292000

上海

上海市静安区北京西路1701号静安中华大厦602单元

邮编：200040

电话：021-62889117

传真：021-62889115

广州

广州市天河北路898号信源大厦3408室

邮编：510898

电话：020-38182838

传真：020-38182835

深圳

深圳市福田区华强北路群星广场B座2408室

邮编：518000

电话：0755-25327693

传真：0755-83534550