天翼云

云存储网关

产品白皮书

天翼云科技有限公司

修订记录

版本	发布日期	描述
2.0	2021 年 05 月 28 日	第一次发布
2.1	2021 年 08 月 27 日	增加对 ARM 服务器的支持。
2.3	2021 年 09 月 22 日	支持设置卷的写策略

目 录

产品概述

产品定义

云存储网关（HBlock）是天翼云自主研发的一款可在线下和云上部署的网关软件，方便用户将本地数据轻松上传到天翼云对象存储中，实现存储空间的弹性扩展。云存储网关作为本地与云端存储之间的桥梁，通过标准 iSCSI 协议提供块存储服务，帮助用户将全量数据自动同步到天翼云对象存储 OOS 中，本地仅保留热数据以节省本地存储空间，或者保留全量数据以保障本地 I/O 性能，实现混合云存储。同时，云存储网关也可以将通用服务器及其管理的存储资源转换成高性能的虚拟存储阵列，承载企业核心业务数据。

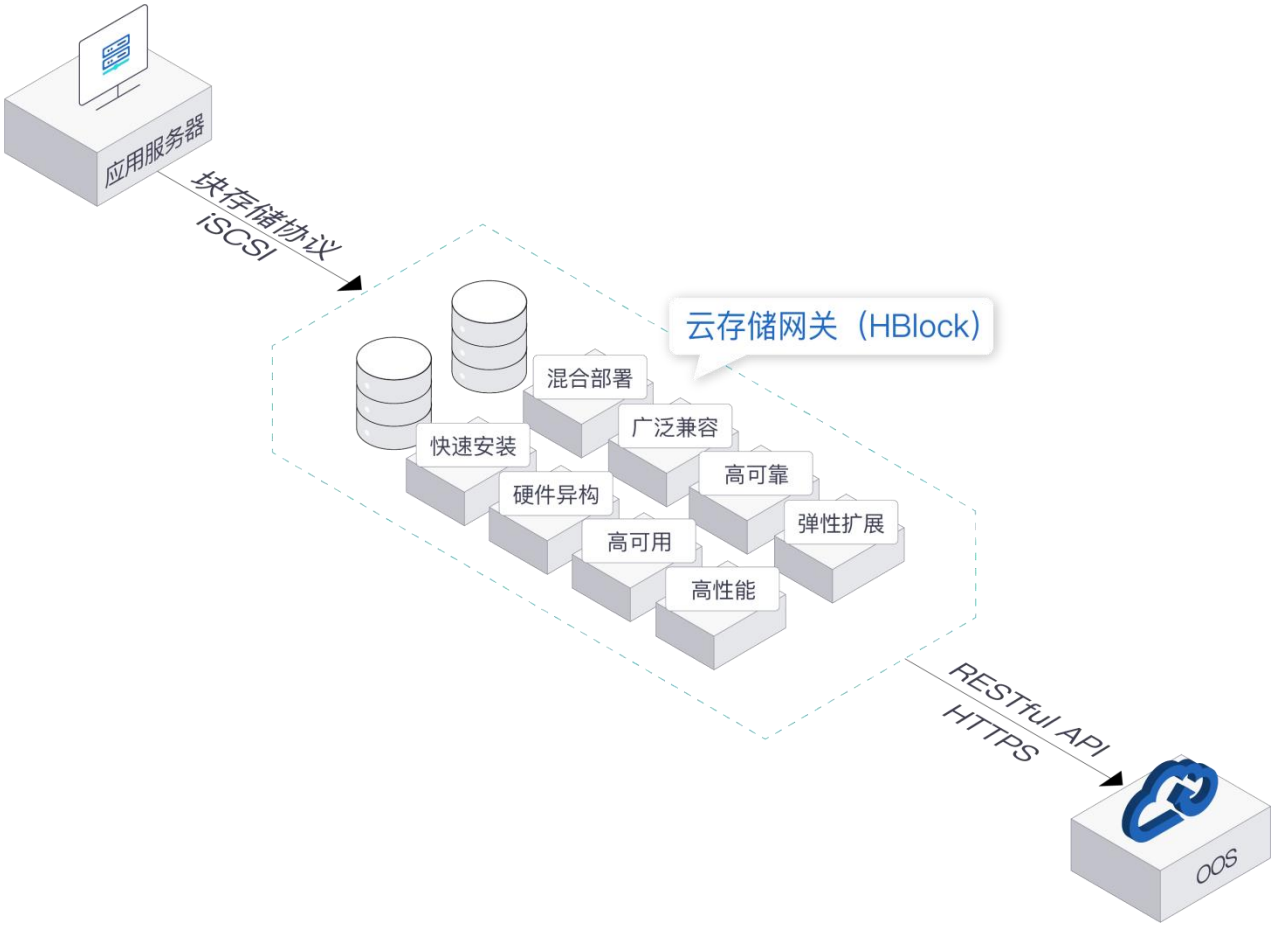

图1.　云存储网关（HBlock）架构图

云存储网关具有以下优势：

- **易安装**：云存储网关安装包是一个 zip 包，可以安装在通用 64 位 x86 服务器和 ARM 服务器的主流 Linux 操作系统上，支持物理服务器、裸金属服务器、虚拟机。云存储网关与硬件驱动程序完全解耦，用户可以自由使用市场上最新的硬件，减少供应商锁定。
- **绿色**：云存储网关作为一组用户态进程运行，不依赖任何特定版本的 Linux 内核或发行版，不依赖、不修改操作系统环境，不独占整个硬盘，不干扰其他进程的执行。因此，云存储网关可以与其他应用同时运行在同一 Linux 操作系统实例中。我们称此功能为“**绿色**”。一方面，它可以帮助用户提高现有硬件资源的利用率，另一方面，它也降低了用户使用云存储网关的门槛 — 甚至不需要虚拟机。
- **高利用率**：云存储网关支持异构硬件，集群中的每个 Linux 操作系统实例可以具有不同的硬件配置，例如不同数量的 CPU、不同的内存大小、不同容量的本地硬盘等。因此可以提高现有硬件资源的利用率。
- **高性能**：云存储网关采用分布式双控架构，提供像传统硬件存储阵列一样的低延迟和高可用性，以及像传统分布式存储一样的高扩展性和高吞吐量。 支持在不中断业务的情况下，从 3 台服务器扩展到数千台服务器，并从数千台服务器逐台缩小到 3 台服务器。
- **高质量**：当集群中同时发生的磁盘故障数不大于逻辑卷冗余模式允许的故障数（对于 3 副本模式，允许的故障数为 2；对于纠删码 N+M 模式，允许的故障数为 M），不影响云存储网关的数据持久性。在集群中发生单个服务器、链路或磁盘故障时，云存储网关保证服务可用。云存储网关是面向混沌（Chaos）环境设计的，可适用于弱网、弱电、弱盘等不确定环境，并在发布之前已经在复杂和大规模的环境中进行了充分的测试。
- **支持数据上云**：云存储网关可以与天翼云对象存储 OOS 结合，创建存储和缓存模式的卷。
  - 对于存储模式的卷：全量数据不仅存储在本地，还会被异步地复制到 OOS 中，实现本地高性能和异地数据灾备；

- 对于缓存模式的卷：最近读写的数据会缓存在本地以尽可能提高性能，全量数据将保存在 OOS 上以降低成本，使得很小的本地容量可以存储海量的数据，特别适合于数据备份、归档等对实时性要求不高的业务，以及文档卷宗、医疗影像、视频监控等写入多调阅少的业务。云存储网关可将本地应用与云端存储无缝连接，实现存储空间的按需使用，弹性扩展。

- **一致性：**云存储网关利用了 OOS 的原子操作，能够真正确保云上数据的一致性（即任何时候云上数据都是本地数据的一个快照），不会出现因云上数据的不一致而导致无法恢复整个业务的情况，从而保证数据安全。

产品优势

易于安装

只需 3 个命令即可将云存储网关安装在 Linux 操作系统上，从安装包解压到集群初始化不超过 3 分钟，即可享受本地磁盘的读写体验与云上无限存储空间。

安装包小

安装包为 zip 类型，经过了高度优化，只有大约 140MB，安装部署非常方便。

硬件驱动程序解耦

云存储网关与硬件驱动程序完全解耦，可以安装在物理服务器、裸金属服务器、虚拟机的 Linux 操作系统上。因此，用户可以自由使用市场上最新的硬件，减少供应商锁定。

高利用率

混合部署

云存储网关是一个用户态进程级的软件，不依赖、不修改操作系统环境，不独占整个硬盘，也不干扰其他进程的执行。因此，它可以与其他应用同时运行在同一 Linux 操作系统实例中，帮助用户提高现有硬件资源的利用率，同时也降低了用户使用云存储网关的门槛。

异构硬件部署

集群中的每个 Linux 操作系统实例可以具有不同的硬件配置，例如不同数量的 CPU、不同的内存大小、不同容量的本地硬盘等。因此，可以提高硬件资源利用率。

自动精简配置

精简配置为应用程序提供了比实际物理存储设备上更多可用的虚拟存储空间。在数据写入逻辑卷之前，云存储网关即可以为上层应用提供存储设备，而不占用任何物理存储空间。云存储网关的卷默认自动支持精简配置，提高了存储空间的有效利用。

兼容性强

云存储网关与通用 64 位 x86 服务器和 ARM 服务器上的主流 Linux 操作系统兼容。支持部署在物理服务器、裸金属服务器、虚拟机的 Linux 操作系统上。

硬件兼容性

硬件兼容性包括:

- **CPU 架构**：通用 x86 服务器，ARM 服务器
- **存储介质**：NVMe SSD、SAS SSD、SATA SSD、SAS HDD、NL-SAS HDD、SATA HDD。

软件兼容性

软件兼容性包括：

- **操作系统**：云存储网关可以部署在 Linux 操作系统上，不依赖任何特定版本的 Linux 内核或发行版。客户端支持 Windows 和 Linux 操作系统。
- **虚拟化平台**：支持与 KVM 和 VMware 的虚拟化平台整合。
- **数据库**：支持多种数据库应用程序，如 Oracle、MySQL、SQL Server、PostgreSQL、MongoDB、DB2 等。
- **应用**：支持各种企业 IT 应用、行业应用和 web 应用。

高可用

秒级故障切换

在集群模式下，一个逻辑卷对应两个 Target：Active Target 和 Standby Target。当卷对应的 Active Target 所在服务器故障时，云存储网关将在几秒内自动切换到 Standby Target，而不会导致业务中断。

传统的虚拟 IP 模式采用“节点和虚拟 IP 对应”的设计方式，要求有权限变更节点的网络配置，存储系统与网络系统之间紧耦合，不利于系统扩展和更新。云存储网关采用先进的双控架构，只需要确保客户端能连接 Active Target 和 Standby Target 所在服务器的 IP，就可以通过标准的 MPIO 技术实现秒级故障切换， 不需要增加额外的 IP 地址，简化了云存储网关的部署，同时也能够和现有的应用并行运行。

无单点故障

云存储网关采用双控架构，并且集群中的服务器都采用冗余模式部署。在集群中，当单个服务器、单链路或单个磁盘发生故障时，云存储网关可确保高可用。任何人为引起的单点故障或者系统引起的单点故障，都不会影响服务的可用性。

智能调速器

云存储网关监控数据读/写过程中磁盘空间、内存和其他资源的使用情况。当资源不足时，速度调节器将自动调整数据写入速度，以确保磁盘始终可写、服务始终可用。而其他存储产品在资源不足时，还会持续写入数据，最终把磁盘写满，导致服务突然中断。

高可靠性

支持纠删码

纠删码（Erasure Code，EC）是一种数据冗余保护机制，广泛应用于分布式存储领域。数据写入云存储网关后，EC 模式会将源数据分割成 N 个片段，然后从 N 个片段中生成 M 个校

验数据，得到 N+M 个数据片段，并将这些数据存放到 N+M 个不同的服务器上。对于纠删码 N+M 冗余模式，最多允许 M 台服务器宕机，这 M 台的数据可以通过另外 N 台服务器的数据进行恢复。相对于三副本模式，EC 模式提供了相同或者更高的可靠性，显著提高了磁盘的利用率，节省了总成本。

在使用 HDD 和小块读写数据的场景中，云存储网关也支持纠删码模式，并且可以实现低延迟。

数据零丢失

云存储网关支持多副本和纠删码数据冗余保护机制。数据被分割存储在不同的服务器上，当单个服务器、单条链路或者单个磁盘发生故障时，云存储网关使用存储在其他服务器上的数据片，在后台开始重建/重新平衡数据，以便重新分配数据。所有数据不会出现丢失或者暂时不可用的情况。

云存储网关运行过程中可能遇到各种亚健康状态。例如，服务器中其他应用引起的高负载，CPU/内存使用率高，网络异常（如数据包丢失或高延迟）以及其他异常情况。这些情况统称为亚健康状态。在亚健康状态下，云存储网关仍可以确保数据不丢失，服务不停止。

另外，云存储网关有其独特的内部时钟检查机制，可以确定每台服务器的运行时间。云存储网关中的每台服务器不需要配置 NTP 服务，服务器时钟可以任意设置。云存储网关对不同服务器的时钟偏差没有要求，不会因为时钟不同步而导致数据丢失或者服务不可用。但是对于传统分布式存储，服务器必须配置 NTP，否则会引起数据丢失。

数据一致性

为了确保数据一致性，云存储网关支持多种数据校验，包括客户端-服务器端校验、集群内部全流程数据校验、多台服务器间数据校验等。

- 多台服务器间的数据一致性：集群中多台服务器上的数据以版本号为比较标准，最新数据是带有最新版本号的数据。这确保了数据的严格一致性。当发现异常副本后，云存储网关将自动修复异常副本。

- 内存数据和持久数据的一致性：云存储网关定期扫描内存和磁盘数据。当磁盘数据不可访问或者校验失败时，云存储网关会自动启动数据恢复进程来重建数据。
- 云上数据与本地数据的一致性：云存储网关利用了天翼云对象存储 OOS 的原子操作，能够真正确保云上数据的一致性（即任何时候云上数据都是本地数据的一个快照），不会出现因云上数据的不一致而导致无法恢复整个业务的情况，从而保证数据安全。其他存储产品无法利用云端对象存储的原子操作接口，因此无法确保云上数据的一致性。并且，有的存储产品把元数据仅存储在本地系统中，很容易造成元数据丢失。云存储网关将元数据和数据都存储到云端，并保证一致性。

高性能

低延迟

云存储网关可以自动实现本地缓存和云端存储的冷热数据分离，提升读写性能，达到极低的读写延迟。

聚合吞吐

云存储网关中的 iSCSI Target 可以被创建在集群中的任何服务器上。创建 LUN 时，为了使系统负载均衡，云存储网关会选择集群中两个负载比较低的服务器作为 Target 服务器。因此，云存储网关没有单点的吞吐量瓶颈。但是，对于传统硬件存储阵列，控制器和磁盘框之间有限带宽会成为吞吐量的瓶颈。

避免性能瓶颈

云存储网关支持多副本和纠删码数据冗余保护，数据片段存储在不同的服务器上。当一个磁盘或者服务器出现故障、集群中添加新服务器或移出服务器时，云存储网关将在后台自动启动数据重建/重新平衡，来重新分配数据。由于数据片段分布在多个不同的服务器上，因此将在多个服务器上进行数据重建/重新平衡，从而有效避免了因单个服务器上大量数据重建/重新平衡造成的性能瓶颈，对业务的影响降到最低。

弹性扩展

云存储网关架构不仅支持纵向扩展（通过增加单服务器的处理器、内存、网络和磁盘进行扩展），还支持横向扩展（通过添加服务器进行扩展）。这使得云存储网关可以基于 IOPS、存储空间和带宽进行独立扩展。

云存储网关支持灵活的扩展方法：通过添加新磁盘扩展现有服务器容量，或者通过添加新服务器来扩展容量。扩容后，无需重新定位大量数据，系统便可自动实现负载均衡。

使用云存储网关，用户不需要进行大量的前期投入。可以在使用过程中，随时按需添加服务器或磁盘，这些硬件可以是价格低廉易用的通用硬件，添加过程中不会中断业务。

安全认证

云存储网关支持挑战握手认证协议（Challenge-Handshake Authentication Protocol，CHAP）。CHAP 是一种对等身份认证协议，允许 iSCSI 客户端和 Target 端基于密码进行安全身份认证。CHAP 包括单向认证和双向认证。对于单向 CHAP，Target 在连接时对客户端 initiator 进行身份认证。对于双向 CHAP，客户端和 Target 端基于各自的密码进行认证。云存储网关支持单向 CHAP，后续版本会支持双向 CHAP。

易操作和维护

支持故障告警

云存储网关监控系统中所有的资源，实时了解资源使用情况和紧急情况。当系统中的组件或资源出现异常时，云存储网关会自动发送邮件通知用户。

以慢速磁盘检测为例：磁盘长时间工作后，可能会出现组件老化等问题，导致 I/O 响应时间变慢，最终导致服务不可用。云存储网关会定期执行磁盘检测、监控、分析和诊断磁盘的读写请求，以评估磁盘是否为慢盘，在发现慢盘后及时通知用户。

支持 NAT 访问

通常，iSCSI initiator 通过内网访问云存储网关中的 Target。如果内网路由器上配置了网络地址转换（Network Address Translation，NAT），iSCSI initiator 可以通过 NAT 的外网 IP 连接到 Target 所在服务器，从而通过云存储网关将 iSCSI 作为云服务进行远程提供。而其他存储产品不支持用户设置 Target 的 portal IP，因此无法支持 NAT 访问。

滚动升级

云存储网关支持自动滚动升级，即在不中断业务的情况下，允许用户一键升级至最新版本。云存储网关自动逐个升级集群中的服务器，在同一时间只能重启一台服务器。被重启的服务器上的 Target 将进行主备切换，但切换期间不会影响客户端业务。

业务场景适配

云存储网关支持 3 种写策略，用户可以根据自己的业务场景特点，设置卷级别的数据写入方式。3 种方式如下：

- 回写：数据写入到内存后，立刻返回给客户端写成功，之后再异步写入磁盘。适用于对性能要求较高，稳定性要求不高的场景。
- 透写：数据同时写入内存和磁盘，并在两处都写成功后，再返回客户端写成功。适用于稳定性要求较高，写性能要求不高，且最近写入的数据会较快被读取的场景。
- 绕写：数据写入磁盘后即释放相应内存，写入磁盘成功后，立刻返回客户端写成功。适用于稳定性要求较高，性能要求不高，且写多读少的场景。

应用场景

承载关键业务

云存储网关采用分布式双控架构，实现了秒级故障切换、低时延、和高吞吐量，满足企业级核心业务高负载和高性能的要求。云存储网关支持纠删码数据冗余保护，数据片段分布在不同的服务器上，增强了企业数据的持久性，同时提升磁盘空间利用率。同时云存储网关支持弹性扩展，可以在不中断企业业务的情况下进行扩容。因此，云存储网关可以承载企业的核心业务数据，包括：数据库、虚拟化平台、应用程序等。

数据上云

对于需要存储海量数据的客户，如备份数据，视频监控，归档数据等，可以通过云存储网关将数据上传到天翼云对象存储中，存储空间按需使用，可弹性伸缩至 PB 级别，为大规模数据、高带宽型应用提供有力支持。同时，利用云端对象存储，可以降低存储总成本。

无缝接入

利旧原有业务系统，可通过云存储网关使用行业标准 iSCSI 协议连接，无需重写本地应用程序，无需修改现有使用架构，即可将本地应用与云端存储无缝连接，享受低成本、易扩展，无上限的云端存储。此外，本地可以保留需要经常访问的热数据，实现低延迟访问。

部署方式

云存储网关支持两种部署方式：

- 部署在云端

云存储网关可以部署在云端服务器上，用户的本地应用服务器和云主机通过专线相连，将数据通过云存储网关上传到天翼云对象存储中。

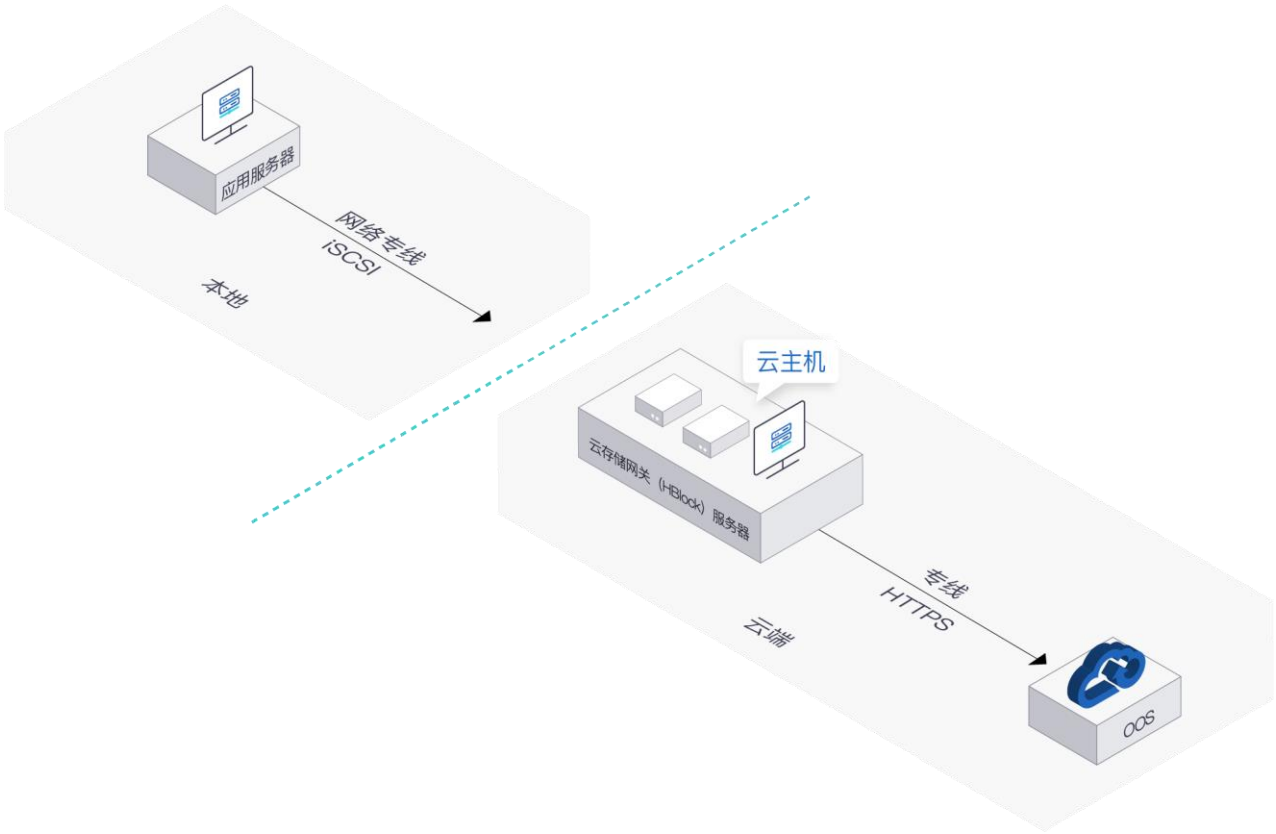

图2.　云存储网关（HBlock）部署在云端

- 部署在用户本地

云存储网关也可以部署用户本地服务器上，可以是单独的服务器，也可以利旧用户已经有的服务器。云存储网关是用户态进程级软件，可以和用户的已有应用混合部署。用户本地的应用服务器与云存储网关通过内部局域网连接，通过 iSCSI 协议进行数据读写。如果用户要将本地数据上传到云端，云存储网关可通过互联网或者专线与天翼云对象存储相连，并通过 HTTP RESTful API 协议将本地数据上传到云端。

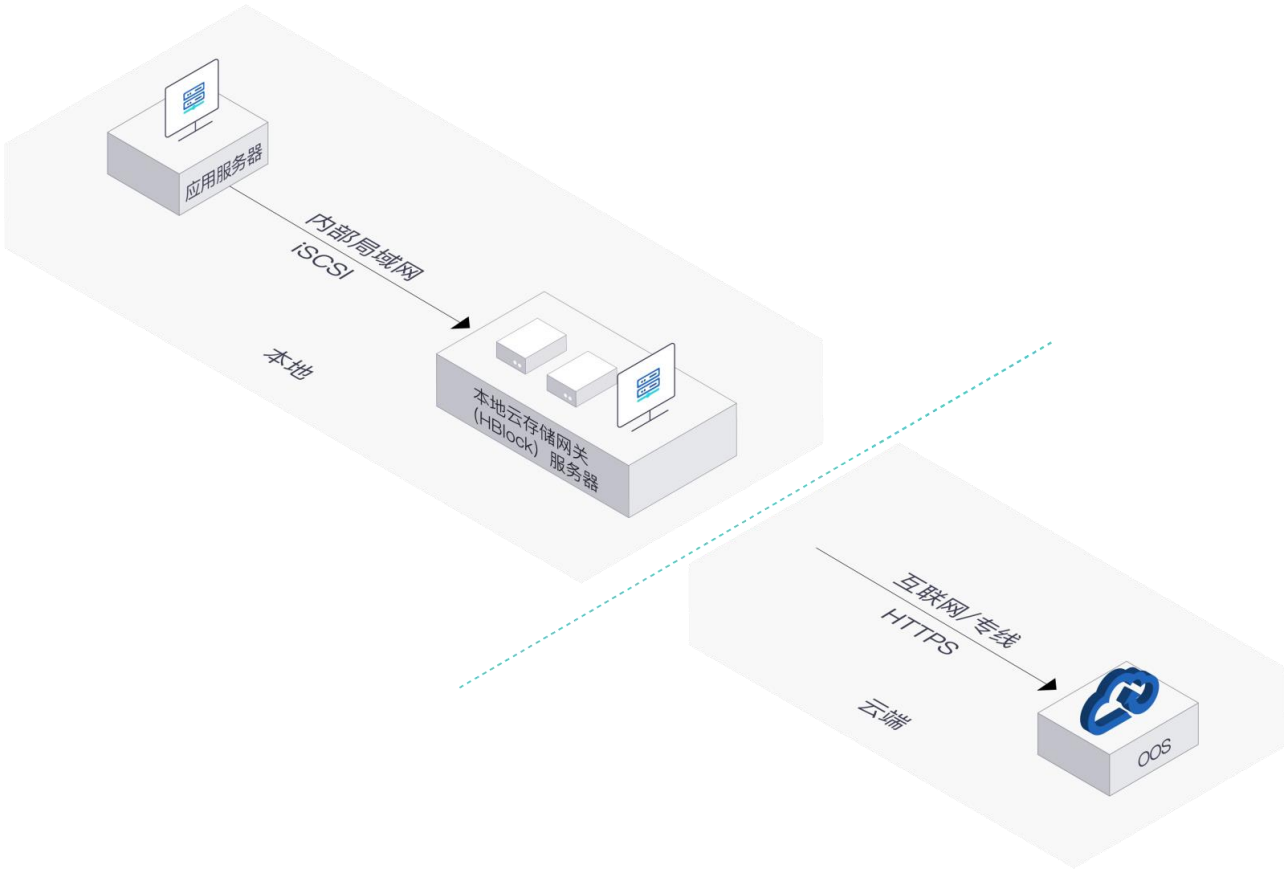

图3.　　云存储网关部署在本地

规格

架构	分布式双控架构
支持协议	标准 iSCSI 协议
服务器数	单机版，集群版：3-数千节点
LUN 的个数	不受限
CPU 体系结构	x86，ARM
操作系统	CentOS 7.x，64 位
混合部署	支持 可以与其他应用同时运行在同一 Linux 操作系统实例中
异构硬件部署	支持 允许集群中每个 Linux 操作系统实例有不同的硬件配置
高可用	支持，支持 MPIO
高性能	低时延，高吞吐量
数据冗余保护	多副本：3 副本 纠删码：N+M 冗余模式（M: +1, +2）
存储介质	NVMe SSD、SAS SSD、SATA SSD、SAS HDD、NL-SAS HDD、SATA HDD
网络介质	TCP/IP