

对象存储服务

# 最佳实践

文档版本 01  
发布日期 2024-07-12



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

## 商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## 注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

# 华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

# 目录

<b>1 OBS 最佳实践汇总</b>	<b>1</b>
<b>2 OBS 安全配置建议</b>	<b>3</b>
<b>3 搬迁本地数据至 OBS</b>	<b>7</b>
3.1 概述	7
3.2 OBS 工具方式	8
3.3 CDM 方式	8
3.4 DES 磁盘方式	9
3.5 DES Teleport 方式	10
3.6 云专线方式	10
<b>4 迁移第三方云厂商数据至 OBS</b>	<b>12</b>
4.1 概述	12
4.2 OMS 方式	12
4.3 OBS 镜像回源方式	14
<b>5 OBS 之间数据迁移</b>	<b>15</b>
<b>6 使用备份软件实现本地数据备份至 OBS</b>	<b>17</b>
6.1 概述	17
6.2 使用 Commvault 备份本地 SAP HANA	17
<b>7 在 ECS 上通过内网访问 OBS</b>	<b>20</b>
7.1 方案概述	20
7.2 资源成本及规划	21
7.3 实施步骤	22
7.3.1 场景一：在 Windows ECS 上使用 OBS Browser+通过内网访问 OBS	22
7.3.2 场景二：在 Linux ECS 上使用 obsutil 通过内网访问 OBS	25
<b>8 使用自定义域名托管静态网站</b>	<b>29</b>
8.1 方案概述	29
8.2 资源和成本规划	30
8.3 操作流程	31
8.4 实施步骤	32
8.4.1 准备工作	32
8.4.2 上传静态网站文件	33

8.4.3 配置静态网站托管.....	34
8.4.4 配置自定义域名.....	35
8.4.5 创建和配置域名托管.....	36
8.4.6 验证.....	37
8.4.7 (后续操作)更新静态网站.....	38
<b>9 OBS 数据一致性校验.....</b>	<b>41</b>
9.1 方案概述.....	41
9.2 校验上传对象的一致性.....	44
9.3 校验下载对象的一致性.....	46
<b>10 OBS 性能优化建议.....</b>	<b>50</b>
<b>11 Web 端通过 PostObject 接口直传 OBS.....</b>	<b>51</b>
<b>12 移动应用直传.....</b>	<b>59</b>
12.1 方案概述.....	59
12.2 使用临时安全凭证直传 OBS.....	59
12.3 使用预签名 URL 直传 OBS.....	63
<b>13 小程序直传 OBS.....</b>	<b>68</b>
<b>14 通过 Nginx 反向代理访问 OBS.....</b>	<b>73</b>
<b>15 大数据场景下使用 OBS 实现存算分离.....</b>	<b>79</b>
15.1 方案概述.....	79
15.2 操作流程.....	82
15.3 对接大数据平台.....	82
15.3.1 支持的大数据平台简介.....	82
15.3.2 华为云 MRS 对接 OBS.....	83
15.3.3 Cloudera CDH 对接 OBS.....	83
15.3.4 Hortonworks HDP 对接 OBS.....	86
15.4 对接大数据组件.....	88
15.4.1 支持的大数据组件简介.....	88
15.4.2 Hadoop 对接 OBS.....	89
15.4.3 Hive 对接 OBS.....	94
15.4.4 Spark 对接 OBS.....	94
15.4.5 Presto 对接 OBS.....	95
15.4.6 Flume 对接 OBS.....	99
15.4.7 DataX 对接 OBS.....	100
15.4.8 Druid 对接 OBS.....	102
15.4.9 Flink 对接 OBS.....	104
15.4.10 Logstash 对接 OBS.....	105
15.5 迁移 HDFS 数据至 OBS.....	107
<b>16 通过 s3fs 挂载 OBS 对象桶.....</b>	<b>109</b>

# 1 OBS 最佳实践汇总

本文汇总了基于对象存储服务（OBS，Object Storage Service）常见应用场景的操作实践，为每个实践提供详细的方案描述和操作指导，帮助用户轻松构建基于OBS的存储业务。

表 1-1 OBS 最佳实践一览表

最佳实践	说明
<a href="#">OBS安全配置建议</a>	本章节提供了OBS使用过程中的安全最佳实践，旨在为提高整体安全能力提供可操作的规范性指导。
<a href="#">企业数据权限控制最佳实践</a>	本最佳实践提供了企业开通OBS后可以设置的四种常见权限控制。 <ul style="list-style-type: none"><li>• 为不同职能部门的员工设置不同的访问权限，以此达到不同部门人员访问公司数据的权限隔离。</li><li>• 设置权限允许其他部门/项目用户下载共享数据，禁止写删。</li><li>• 给各业务部门分配各自所需的IAM用户，通过桶策略给每个业务部门下的IAM用户授予独立的资源权限。</li><li>• 通过OBS Browser+添加外部桶的方式实现业务部门之间桶资源隔离。</li></ul>
<a href="#">搬迁本地数据至OBS</a>	本章节根据用户本地（个人电脑或自建存储服务器）数据大小，介绍了几种将本地数据搬迁至OBS的方式，并针对不同方式提供了对应操作流程及指导。
<a href="#">迁移第三方云厂商的数据至OBS</a>	本章节根据存储在第三方云厂商的数据量及迁移场景，介绍了几种迁移方式，并针对不同方式提供了对应操作流程及指导。
<a href="#">OBS之间数据迁移</a>	本章节介绍如何在华为云对象存储服务OBS之间进行跨账号、跨地域、以及同地域内的数据迁移。
<a href="#">使用备份软件实现本地数据备份至OBS</a>	本章节描述了备份本地数据至OBS的背景以及OBS支持的备份软件，并以Commvault备份软件为例，介绍了备份本地数据至OBS的基本流程。

最佳实践	说明
<a href="#">在ECS上通过内网访问OBS</a>	ECS支持通过公网和华为云内网两种方式访问OBS，为优化性能、节省开支，建议通过华为云内网访问OBS。本章节详细描述了在ECS上如何通过华为云内网访问OBS服务。
<a href="#">使用自定义域名托管静态网站</a>	本章节详细描述了在OBS上使用自定义域名托管静态网站的操作流程及步骤，无需搭建网站服务器，即可快速发布个人及企业静态网站。
<a href="#">OBS数据一致性校验</a>	对象数据在上传下载过程中，有可能会因为网络劫持、数据缓存等原因，存在数据不一致的问题。本章介绍如何利用OBS提供的通过计算MD5值的方式，对上传下载的数据进行一致性校验。
<a href="#">性能优化最佳实践</a>	本章节介绍如何通过给对象添加随机前缀名，对高速率访问请求进行水平扩展，以达到提升访问速率，降低访问时延的效果。
<a href="#">Web端通过PostObject接口直传OBS</a>	本章节介绍一种在Web端利用PostObject接口直传文件至OBS的方法，即使用表单上传方式上传文件至OBS。该方案省去了应用服务器这一步骤，提高了传输效率，不会对服务器产生压力，且服务端签名后直传可以保证传输的安全性。
<a href="#">移动应用直传</a>	本章节介绍了应用客户端访问OBS的两种方法，从而更好地保护应用数据，避免被攻击后数据泄露以及越权访问的风险。
<a href="#">小程序直传OBS</a>	本章节通过一个示例程序演示了如何通过微信小程序上传文件至OBS。
<a href="#">通过Nginx反向代理访问OBS</a>	本章节介绍如何通过ECS上配置Nginx反向代理，实现通过固定IP地址访问OBS。
<a href="#">通过s3fs挂载OBS对象桶</a>	s3fs是一款基于用户空间文件系统（Filesystem in Userspace，简称FUSE）的文件系统工具。本章节介绍在Linux系统中，将OBS对象桶挂载到本地文件系统，让您能够像操作本地文件一样操作对象，实现数据的共享。

# 2 OBS 安全配置建议

安全性是华为云与您的共同责任。华为云负责云服务自身的安全，提供安全的云；作为租户，您需要合理使用云服务提供的安全能力对数据进行保护，安全地使用云。详情请参见[责任共担](#)。

本文提供了OBS使用过程中的安全最佳实践，旨在为提高整体安全能力提供可操作的规范性指导。根据该指导文档您可以持续评估OBS资源的安全状态，更好的组合使用OBS提供的多种安全能力，提高对OBS资源的整体安全防御能力，保护存储在OBS桶内的数据不泄露、不被篡改，以及数据传输过程中不泄露、不被篡改。

本文从以下几个维度给出建议，您可以评估OBS使用情况，并根据业务需要在本指导的基础上进行安全配置。

- [建议妥善管理认证凭证，减小因凭证泄漏导致的数据泄露风险](#)
- [正确的使用OBS提供的访问控制能力保护数据不泄露、不被篡改](#)
- [加密存储数据](#)
- [构建数据的恢复、容灾能力避免数据被异常破坏](#)
- [确保您的数据在传输到OBS过程中不被窃取和篡改](#)
- [利用OBS提供的操作日志审计是否存在异常数据访问操作](#)
- [使用最新版本的SDK获得更好的操作体验和更强的安全能力](#)
- [使用其他云服务进一步增强对数据的安全防护](#)

## 建议妥善管理认证凭证，减小因凭证泄漏导致的数据泄露风险

### 1. 建议使用临时AK/SK进行业务处理，减小凭证泄漏导致您数据泄露的风险

当部署在ECS上的应用程序或者其他华为云服务需要访问OBS资源时，必须对访问OBS的请求进行签名，因此应用程序或服务需要持有一个可以访问OBS桶的凭证。建议您为应用程序或服务配置IAM委托或临时AK/SK，通过IAM委托可以获得一组临时AK/SK，临时AK/SK到期自动过期失效，可以有效降低凭证泄露造成的数据泄露风险。详情请参见[通过临时访问密钥访问OBS](#)和[通过委托获取临时AK/SK](#)。

### 2. 定期轮转永久AK/SK减小凭证泄漏导致您数据泄露的风险

如您必须使用永久AK/SK，建议对永久AK/SK进行定期凭证轮转，同时加密存储，避免凭证长期使用过程中预置的明文凭证泄露导致数据泄露。详情请参见[通过永久访问密钥访问OBS](#)。



## 正确的使用 OBS 提供的访问控制能力保护数据不泄露、不被篡改

正确的使用OBS提供的访问控制能力，可以有效预防您的数据被异常窃取或者破坏。

### 1. 建议对不同角色的IAM用户仅设置最小权限，避免权限过大导致数据泄露或被误操作

为了更好的进行权限隔离和管理，建议您配置独立的IAM管理员，授予IAM管理员IAM策略的管理权限。IAM管理员可以根据您业务的实际诉求创建不同的用户组，用户组对应不同的数据访问场景，通过将用户添加到用户组并将IAM策略绑定到对应用户组，IAM管理员可以为不同职能部门的员工按照最小权限原则授予不同的数据访问权限，详情请参见[部门公共数据权限管理](#)和[策略语法](#)。

### 2. 利用桶策略保护您的数据不被异常读取和操作

仅在资源上配置实际业务处理中所需权限，避免权限配置过大导致数据被错误分享给他人。详情请参见[OBS权限控制概述](#)。

OBS的桶策略支持多种Condition条件灵活设置，每一个Condition都是一个新的安全控制维度，建议您通过配置Condition进一步限定数据访问的上下文，如通过拒绝非指定源IP对桶的访问，限制数据只能被指定客户端访问，避免数据被窃取。详情请参见[限制指定IP地址对桶的访问权限](#)和[桶策略参数说明](#)。

### 3. 建议将需要公开访问的对象和私有的对象使用不同的桶进行存储，从而简化您的访问控制策略

推荐将需要公开访问的对象和私有的对象使用不同的桶进行存储。公开桶内请勿存放敏感数据，同时请避免私有桶误配置授权公开访问的桶策略导致私有对象泄露，建议存储私有对象的桶通过Condition进一步限制可以访问数据的源端确保数据不被外部攻击者窃取。详情请参见[限制指定IP地址对桶的访问权限](#)。

### 4. 建议使用OBS的数据临时分享功能来快速分享指定数据，无需进行复杂的桶策略编写

当需要将存放在OBS中对象（文件或文件夹）分享给其他用户时，建议使用OBS的数据临时分享功能，分享的URL可指定有效期，过期自动失效，避免数据长期暴露给其他用户导致泄露。详情请参见[通过临时URL访问OBS](#)。

### 5. 开启敏感操作多因子认证保护您的数据不被误删

OBS支持敏感操作保护，开启后执行删除桶等敏感操作时，系统会进行身份验证，进一步保证OBS配置和数据的安全性，对数据的高危操作进行控制。详情请参见[敏感操作](#)。

## 加密存储数据

OBS支持SSE-KMS服务端加密方案，为桶配置SSE-KMS服务端加密能力后，对于上传到桶中的每个对象，OBS都会访问KMS服务获取您指定的KMS密钥进行数据加密，避免数据明文存储。当下载对象时，OBS同样会访问KMS获取对应密钥进行数据解密，整个加解密过程中OBS都不会存储对应密钥。详情请参见[配置桶默认加密](#)。

OBS支持SSE-OBS服务端加密方案，为桶配置SSE-OBS服务端加密能力后，对于上传到桶中的每个对象，OBS都会进行数据加密，避免数据明文存储。当下载对象时，OBS会自动帮助用户进行数据解密。详情请参见[配置桶默认加密](#)。

除了SSE-KMS和SSE-OBS方案，OBS还支持SSE-C服务端加密方案，您可以在对象上传和下载的请求中携带密钥和解密算法，OBS使用您提供的密钥和算法对对象数据进行加解密，避免数据明文存储。OBS不存储您提供的加密密钥，如果您丢失加解密密钥，则会无法获取该对象明文数据。详情请参见[服务端加密SSE-C方式](#)。



## 构建数据的恢复、容灾能力避免数据被异常破坏

预先构建数据的容灾和恢复能力，可以有效避免异常数据处理场景下数据误删、破坏的问题。

### 1. 建议启用多版本获得异常场景数据快速恢复能力

利用多版本控制，您可以在一个桶中保留一个对象的多个版本，在意外操作或应用程序故障时通过历史版本对象快速恢复数据。详情请参见[多版本控制](#)。

### 2. 建议使用跨区域复制构建异地数据容灾能力

有些数据需要异地备份存储，跨区域复制为您提供跨区域数据容灾的能力，满足您将数据复制到异地进行备份的需求。详情请参见[跨区域复制](#)。

## 确保您的数据在传输到 OBS 过程中不被窃取和篡改

### 1. 建议使用HTTPS协议访问OBS，确保数据传输过程中不被窃取和破坏

HTTPS（超文本传输安全协议）是一种互联网通信协议，可保护客户端与服务端之间传输的数据的完整性和机密性。建议您使用HTTPS协议进行数据访问。

### 2. 建议使用桶策略限制对OBS桶的访问必须使用HTTPS协议

为避免客户端误使用HTTP协议进行OBS业务操作，建议通过桶策略中的SecureTransport条件进行限制，限制是否必须使用HTTPS协议发起请求对该桶进行操作。SecureTransport配置为True时，发起的请求必须使用SSL加密。如何配置桶策略中Condition以及SecureTransport条件，详情请参见[桶策略参数说明](#)。

## 利用 OBS 提供的操作日志审计是否存在异常数据访问操作

### 1. 开启云审计服务记录OBS的所有访问操作便于事后审查

云审计服务（Cloud Trace Service, CTS），是华为云安全解决方案中专业的日志审计服务，提供对各种云资源操作记录的收集、存储和查询功能，可用于支撑安全分析、合规审计、资源跟踪和问题定位等常见应用场景。

您开通云审计服务并创建和配置追踪器后，CTS可记录OBS的管理事件和数据事件用于审计。详情请参见[审计](#)。

### 2. 开启桶Logging记录桶的所有访问请求

通过访问日志记录，桶的拥有者可以深入分析访问该桶的用户请求性质、类型或趋势。当用户开启一个桶的日志记录功能后，OBS会自动对这个桶的访问请求记录日志，并生成日志文件写入用户指定的桶中。详情请参见[日志记录](#)。

### 3. 使用云监控服务对安全事件进行实时监控、告警

您在使用OBS的过程中也可能会遇到服务端返回的错误响应，为使您更好地掌握OBS桶的状态，华为云提供了云监控服务（Cloud Eye）。您可使用该服务监控自己的OBS桶，执行自动实时监控、告警和通知操作，帮助您实时掌握桶中所产生的请求、流量和错误响应等信息。

云监控服务不需要开通，会在用户创建资源（如OBS桶）后自动启动。

关于云监控服务的更多介绍，请参见[云监控服务产品介绍](#)。

## 使用最新版本的 SDK 获得更好的操作体验和更强的安全能力

建议客户升级SDK并使用最新版本，可以在您使用OBS的过程中对您的数据提供更好的保护。最新版本SDK在各语言对应界面下载，请参见[OBS SDK](#)。

## 使用其他云服务进一步增强对数据的安全防护

### 1. 启用WAF进行静态网站防护

如果您需要使用OBS静态网站托管功能，为进一步提升您网站的安全性，建议您使用华为云的WAF服务对您的域名进行保护，降低网站被网络攻击的风险。详情请参见[管理防护域名](#)。

### 2. 启用态势感知服务（SA）保障OBS资源安全

态势感知通过“安全上云合规检查1.0”、“等保2.0三级要求”、“护网检查”三种基线规则，检测OBS桶关键配置项，告警提示存在安全隐患的配置，并提供相应配置加固建议和帮助指导。您可以通过态势感知的资源管理功能，快速了解到OBS桶所属区域、安全状况等信息，帮助您定位安全风险问题。详情请参见[资源管理](#)。

### 3. 隐私保护

随着组织管理越来越多的数据，大规模地识别和保护它们的敏感数据会变得越来越复杂、昂贵和耗时。建议您通过DSC服务对OBS桶内数据进行敏感数据识别和管理，降低隐私数据识别和保护的成本以及复杂度。详情请参见[创建敏感数据识别任务](#)。

# 3 搬迁本地数据至 OBS

## 3.1 概述

### 背景

传统的自建存储服务器已不能满足大量的数据存储需求，主要原因可以归类为以下三点：

- 数据存储量受限于搭建服务器时使用的硬件设备，如果存储量不够，需要重新购买存储硬盘，进行人工扩容。
- 前期安装难、设备成本高、初始投资大、自建周期长、无法匹配快速变更的企业业务。
- 需承担网络信息安全、技术漏洞、误操作等各方面的数据安全风险。

OBS提供海量、稳定、安全的云存储能力，无需事先规划存储容量，存储资源可线性无限扩展，用户永远不必担心存储容量不够用。在OBS上可以存储任何类型和大小的非结构化数据，多级可靠性架构以及服务端加密、日志管理、权限控制等功能，保障存储在OBS上的数据高度稳定和安全。在成本方面，OBS即开即用，免去了自建存储服务器带来的资金、时间及人力成本的投入，后期的设备维护也全部交由OBS处理。

华为云提供[搬迁方案](#)，帮助用户将自建存储服务器上的数据短时间、低成本、安全、高效地搬迁至OBS。用户可根据数据量、耗时、费用等需求选择适合的方案进行数据搬迁。

### 搬迁方案

针对不同的搬迁场景及需求，华为云提供如[表3-1](#)所示的几种搬迁方案。

表 3-1 搬迁方案

搬迁方式	适用数据量	要求	耗时	费用
<a href="#">OBS工具方式</a> （在线）	不高于1TB的数据量	要求用户公网带宽空闲，需要人工操作客户端或脚本启动数据上传	家用100Mbps带宽，1TB耗时1天左右	数据传输不收取费用，仅OBS收取基本的存储费用
<a href="#">CDM方式</a> （在线）	单次小于8TB的数据量	需要用户单独购买CDM服务	1TB~8TB/天（取决于网络和数据读取源的读写性能）	根据购买CDM实例规格以及使用时长收费，具体参见 <a href="#">CDM价格详情</a>
<a href="#">DES磁盘方式</a> （离线）	单次小于30TB的数据量	需要用户自己提供磁盘	请参见 <a href="#">从创建DES服务单到数据导入华为云需要多长时间?</a>	根据磁盘数量以及使用时长收费，具体参见 <a href="#">DES价格详情</a>
<a href="#">DES Teleport方式</a> （离线）	单次30TB~500TB的数据量	由华为数据中心邮寄Teleport给用户使用	请参见 <a href="#">从创建DES服务单到数据导入华为云需要多长时间?</a>	根据磁盘数量以及使用时长收费，具体参见 <a href="#">DES价格详情</a>
<a href="#">云专线方式</a> （实时）	每月大于100TB的数据量，需要实时在线上传	需要部署专线	根据专线带宽决定	根据专线距离以及带宽收费，具体参见 <a href="#">云专线价格详情</a>

## 3.2 OBS 工具方式

OBS工具方式适用于百GB规模的数据搬迁。OBS提供OBS Browser+、obsutil等多种客户端工具，方便用户在本地直接将数据上传至OBS。由于上传需要占用用户公网带宽，为不影响用户在公网上的主营业务，建议利用公网带宽空闲的时间上传数据。

各工具使用场景及操作指导，请参见[OBS工具指南](#)。

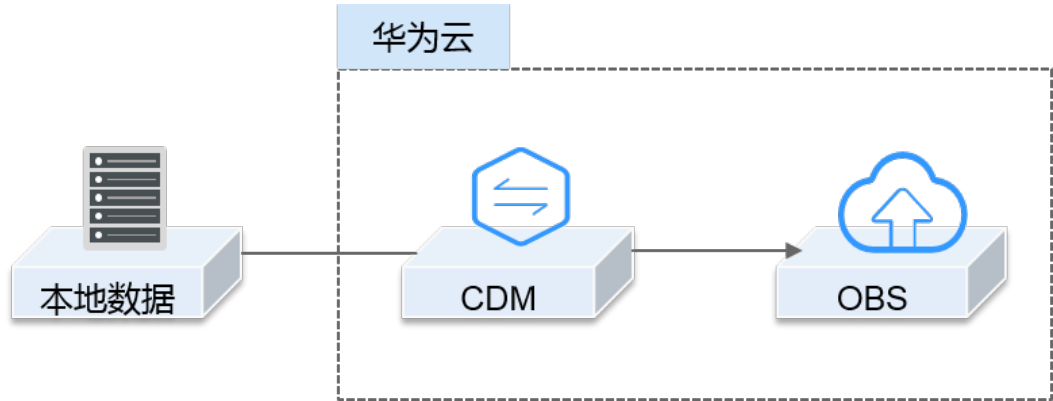
## 3.3 CDM 方式

### 简介

云数据迁移（Cloud Data Migration，CDM）提供同构/异构数据源之间批量数据迁移服务。CDM通过创建定时作业，将用户自建存储服务器上的文件系统、数据库、对象

存储等数据源与华为云OBS进行连接，从而实现定时、自动地将本地数据搬迁至OBS。

图 3-1 CDM 方式搬迁数据示意图

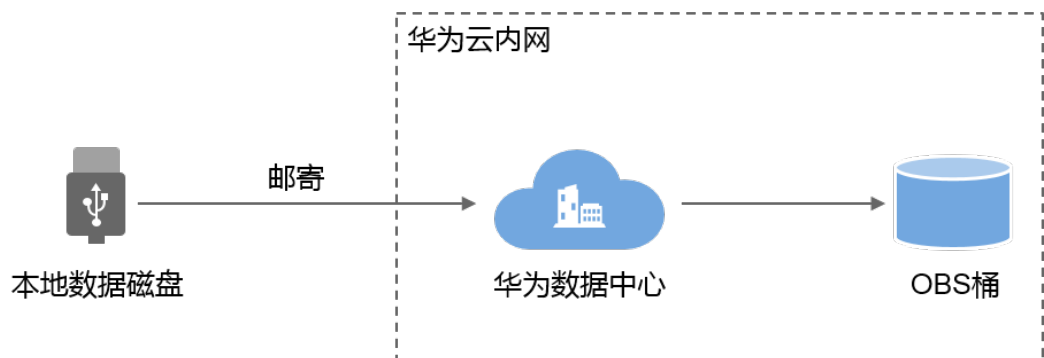


1. 创建OBS桶  
通过OBS控制台或OBS Browser+创建桶，用于存放原始数据。
2. 购买CDM  
购买CDM服务，即创建CDM集群，用于管理连接与作业。
3. 配置连接与作业  
在创建CDM集群中创建一个源连接和一个目的连接，分别与用户本地数据源和云端OBS连接。然后再创建CDM作业，执行从本地数据搬迁到云端OBS的任务。
4. 启动数据传输  
运行CDM作业，启动数据传输。用户可以通过作业管理界面查看作业进度。

### 3.4 DES 磁盘方式

DES磁盘方式同样采用离线的方式将用户数据磁盘（USB、eSATA接口的磁盘等）快递至华为云，实现数据高效传输。磁盘方式适用于30TB以下的数据量搬迁。

图 3-2 DES 磁盘方式搬迁数据示意图



1. 创建OBS桶  
通过OBS控制台或OBS Browser+创建桶，用于存放原始数据。

### 2. 创建DES磁盘方式服务单

登录DES控制台，创建一个磁盘方式的服务单。然后将DES提供的签名文件导入本地数据磁盘并邮寄给华为云数据中心。

### 3. 启动数据传输

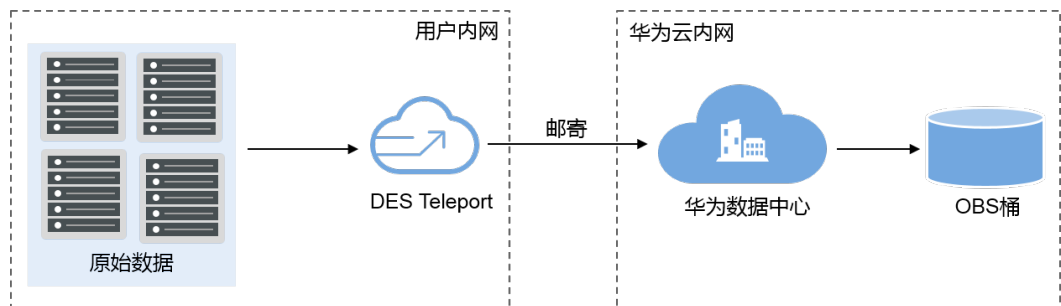
华为云数据中心收到用户的磁盘并将磁盘挂载到物理服务器后，会短信通知用户填写访问密钥（AK和SK）来启动数据上传。数据传输完成后，用户可以通过DES控制台和OBS控制台两种途径查看传输结果。同时华为云数据中心也会将用户磁盘回寄给用户。

具体操作步骤请参见[磁盘方式详细指导](#)。

## 3.5 DES Teleport 方式

Teleport是数据快递服务（Data Express Service, DES）专为30TB~500TB范围内数据搬迁至OBS而定制的存储设备，具有防尘防水、抗震抗压以及GPS锁定、传输加密等多重安全防护机制，配合离线传输的方式，能安全、高效的完成大规模数据搬迁。

图 3-3 DES Teleport 方式搬迁数据示意图



### 1. 创建OBS桶

通过OBS控制台或OBS Browser+创建桶，用于存放原始数据。

### 2. 创建DES Teleport服务单

DES提供Teleport和磁盘两种数据快递方式，在当前场景下选择Teleport方式。

### 3. 接收并导入数据至Teleport

成功创建DES服务单后，用户将接收到由华为数据中心邮寄的Teleport设备。接着进行简单配置操作使Teleport与用户客户端连接起来，然后执行数据拷贝并将设备回寄给华为云数据中心。

### 4. 启动数据传输

在华为云数据中心收到回寄的Teleport后，用户可以在DES控制台上输入访问密钥启动数据从Teleport到OBS指定桶的传输。数据传输完成后，用户可以通过DES控制台和OBS控制台两种途径查看传输结果。

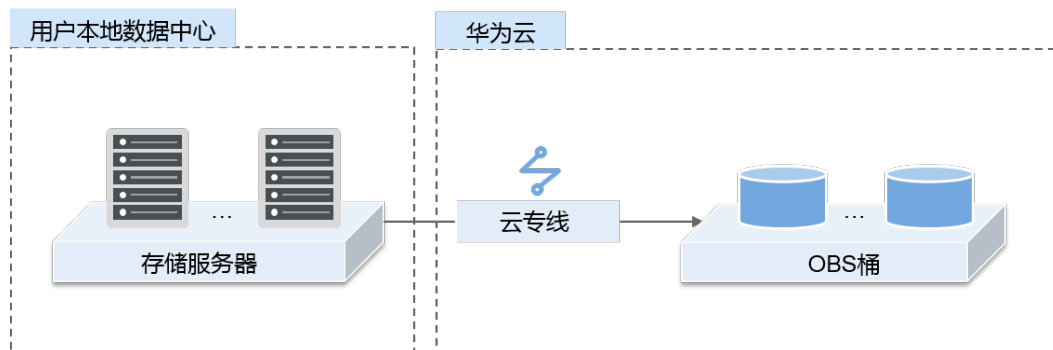
具体操作步骤请参见[Teleport方式详细指导](#)。

## 3.6 云专线方式

云专线方式由用户自己购买云专线服务（Direct Connect, DC），直接将用户本地的网络与华为云网络打通，实现专线直接访问OBS等服务。云专线方式适用于需要频繁

或实时地将本地数据搬迁至OBS的场景。专线提供的低时延、高带宽，可以满足用户随时上传数据至OBS的需求。

图 3-4 云专线方式搬迁数据示意图



1. 创建OBS桶  
登录OBS控制台，创建一个或多个用于存储用户数据的桶。
2. 开通云专线服务  
登录云专线控制台，根据业务需求填写专线申请并提交订单，待管理员审核通过后，用户支付订单，联系运营商安排工程师接通两端物理线路，华为工程师配合进行连接配置。具体操作步骤请参见[开通云专线](#)。
3. 启动数据传输  
专线搭建成功后，用户便可以通过控制台、工具、API、SDK等多种方式将本地数据上传至OBS。



# 4 迁移第三方云厂商数据至 OBS

## 4.1 概述

有大量数据在第三方云厂商对象存储上的用户，需要先将第三方云厂商上的对象数据下载到本地，再通过OBS控制台、客户端等工具上传到OBS，整个过程耗时又耗力，容易存在漏传、误传等问题。

针对迁移第三方云厂商的对象数据至OBS的场景，华为云提供OMS服务。通过迁移服务，用户只需在控制台配置简单的连接参数以及迁移任务，即可把数据从第三方云厂商轻松、平滑地迁移至OBS。如表4-1所示，用户可根据适用场景、迁移速率和费用选择适合的迁移方式。

表 4-1 迁移方案

迁移方式	适用场景	支持的迁移源端	迁移速率	费用
<b>OMS方式</b>	适用于大规模对象数据（500TB以下）全量或增量迁移场景。	请参见 <b>OMS支持迁移的第三方云厂商</b>	10~20TB/天	收费方法请参见 <b>OMS计费说明</b> 。

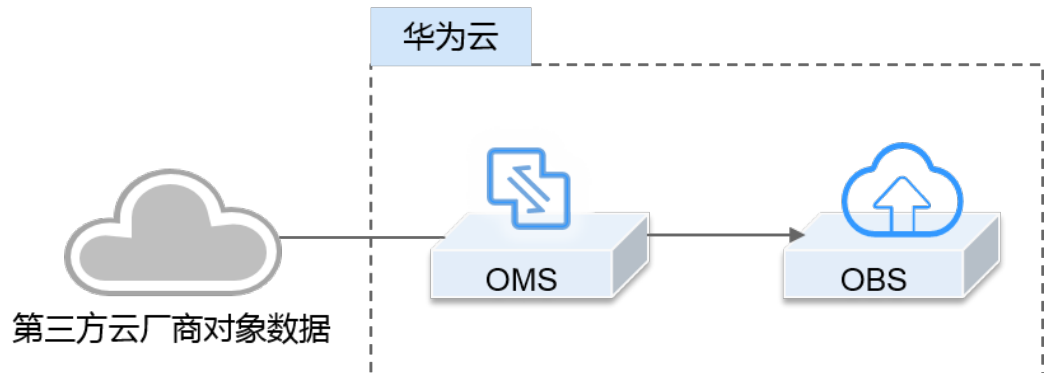
### 📖 说明

低成本、500TB以上的数据迁移请在华为云上[提交工单](#)或联系人工客服，根据您的业务模型，为您定制适合超大规模数据迁移的方案。

## 4.2 OMS 方式

OMS是一种云上的对象数据迁移服务，具有安全、高效等优势。数据迁移时，采用HTTPS数据加密通道，确保数据的传输安全。在重复迁移过程中，只迁移有变动或新增的对象，降低成本。

图 4-1 OMS 方式迁移数据示意图



1. 创建OBS桶  
登录OBS控制台，创建桶用于存放迁移数据。
2. 创建OMS迁移任务  
在OMS控制台创建迁移任务，通过配置访问密钥、桶名等参数将第三方云厂商（源端）与OBS（目的端）关联。  
具体操作步骤请参见[创建对象存储迁移任务](#)。
3. 开始数据迁移  
执行OMS迁移任务，开始数据迁移。数据迁移过程中，可以在OMS查看任务执行状态以及查看到最终结果。

## 迁移示例

1. **阿里云OSS迁移至华为云OBS**  
本方案介绍了如何将阿里云对象存储（Object Storage Service，简称OSS）上的数据迁移到华为云对象存储OBS。  
具体方案请参见[操作教程](#)。
2. **腾讯云COS迁移至华为云OBS**  
本方案介绍了如何将腾讯云对象存储（Cloud Object Storage，简称COS）上的数据迁移到华为云对象存储OBS。  
具体方案请参见[操作教程](#)。
3. **优刻得US3迁移至华为云OBS**  
本方案介绍了如何将优刻得对象存储（US3）上的数据迁移到华为云对象存储OBS。  
具体方案请参见[操作教程](#)。
4. **金山云KS3迁移至华为云OBS**  
本方案介绍了金山云对象存储（Kingsoft Standard Storage Service，简称KS3）上的数据迁移到华为云对象存储OBS。  
具体方案请参见[操作教程](#)。
5. **HTTP/HTTPS数据源迁移至华为云OBS**  
本方案介绍了如何将网络资源迁移至华为云对象存储OBS。  
具体方案请参见[操作教程](#)。

## 业务割接方案

如涉及业务割接，请参考以下方案。

1. 使用“源端/目的端双写方案”方案迁移增量对象数据
2. 使用“多次同步方案”迁移增量对象数据

## 4.3 OBS 镜像回源方式

一般情况下，当客户端访问OBS时，如果OBS中没有被访问的数据，将会返回404错误。OBS提供镜像回源功能，可以在被请求的数据不存在时，通过回源规则从源站获取对应数据。

用户为桶定义镜像回源规则后，如果客户端访问OBS桶中不存在的资源，且该资源符合镜像回源规则，OBS将以镜像回源的方式去数据源站获取资源，将该资源上传到OBS中并返回给客户端。整个过程不中断业务，适用于客户源站无缝迁移数据到OBS，用户可以在无感知的情况下，低成本地迁移业务到OBS上来。镜像回源流程如图4-2所示。

图 4-2 镜像回源流程



## 配置方法

配置方法请参见[创建镜像回源规则](#)。

# 5 OBS 之间数据迁移

本教程介绍如何通过对象存储迁移服务OMS实现对象存储服务OBS之间跨账号、跨区域、以及同区域内的数据迁移。

- 跨账号迁移：不同的华为云账号之间桶数据迁移。
- 跨区域迁移：不同区域之间的桶数据迁移。
- 同区域迁移：同区域内的桶数据迁移。

## 什么是对象存储迁移服务

对象存储迁移服务（Object Storage Migration Service，OMS）是一种线上数据迁移服务，可以帮助您将其他云服务商对象存储服务中的数据在线迁移至华为云的对象存储服务（Object Storage Service，OBS）中，也可以在对象存储服务OBS之间进行灵活的数据迁移。

使用对象存储迁移服务，您只需在控制台填写源端OBS数据信息和目的端OBS数据信息，并创建迁移任务或创建迁移任务组即可，迁移任务与迁移任务组的区别请参见[迁移任务与迁移任务组的适用场景](#)。

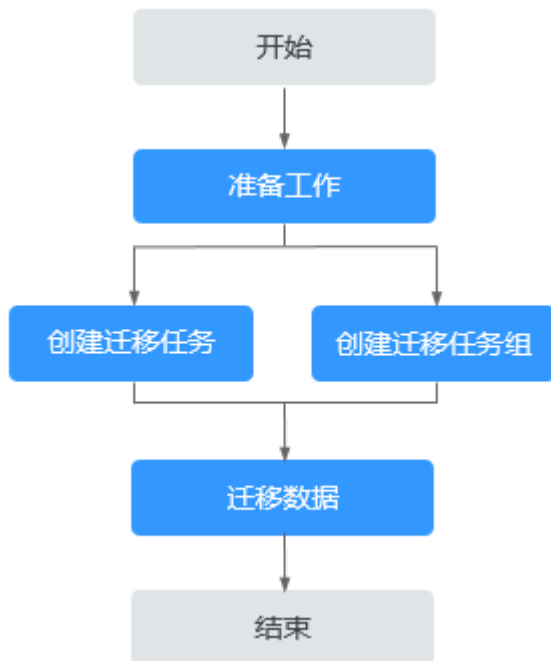
### 注意

- 对象存储迁移服务（OMS）处于免费期，免费期结束后服务将根据您累计使用的迁移流量进行收费，计费方式简单灵活、易于预测。迁移过程中会调用源端和目的端的对象存储服务的API进行数据的上传、下载，所以会产生一定的API请求及下载流量费用，具体费用说明请参见[计费说明](#)。
- 跨账号迁移过程除了会产生对象存储迁移服务费用（当前免费）外，还会产生请求费用和流量费用。
  - 请求费用：该项费用在调用OBS API时产生，按请求次数计费，包括PUT/POST/COPY/LIST/GET/HEAD等。
  - 流量费用：源端下载数据时，会产生流量费用，流量由数据的实际大小决定，费用由源端数据所在云服务商收取；数据上传至华为云OBS不收取流量费用。
- 对象存储迁移服务暂不支持迁移多版本的对象存储数据。
- OMS同样适用于OBS桶和并行文件系统桶之间的数据迁移。

## 迁移流程

迁移流程如图5-1所示，具体操作请参见[华为云OBS之间迁移操作指导](#)。

图 5-1 OBS 之间数据迁移流程



# 6 使用备份软件实现本地数据备份至 OBS

## 6.1 概述

传统的备份与恢复方案需要将备份数据写入磁带等存储设备，然后再运输至数据中心。在此过程中数据的安全及完整性依赖很多因素，比如硬件、人员等等。无论是从前期搭建数据中心还是后期的维护，都使得传统的备份与恢复方案面临着管理复杂、投入成本高的难题。

云存储定位于简单、安全、高效且低成本，使其成为磁带等传统存储设备的非常有吸引力的替代品。OBS即一种云存储服务，它提供海量、可弹性扩展的存储服务。OBS所有的业务、存储节点采用分布集群方式工作使得OBS的可扩展性更高。提供数据多份冗余、一致性检查等功能使得存储在OBS中的数据更加安全、可靠。OBS按照使用量付费，使得成本易于预测。

Commvault、爱数云备份服务（AnyBackup Cloud）等第三方备份软件，都支持对接OBS进行数据备份。通过这些备份软件，用户可以根据自身需求制定合适的备份策略，达到安全、高效的备份目的。

## 6.2 使用 Commvault 备份本地 SAP HANA

SAP HANA是基于内存计算技术的高性能实时数据计算平台，多应用于需要处理大量实时业务数据的企业。备份软件Commvault，与SAP HANA、OBS无缝集成，支持在线数据库、日志的备份。当SAP HANA系统出现故障或业务迁移时，Commvault能帮助用户从OBS快速、轻松地恢复数据，从而为SAP HANA提供企业级数据保护。

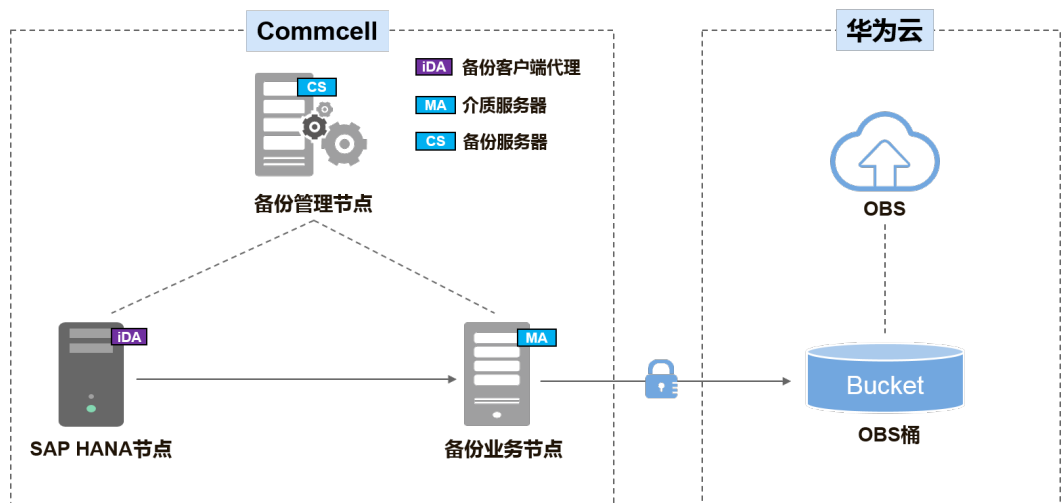
### 📖 说明

在本场景下推荐使用Commvault V11版本。

## 逻辑架构

此处以使用Commvault备份本地单节点部署的SAP HANA为例，其逻辑架构如[图6-1](#)所示。

图 6-1 逻辑架构



逻辑架构中各组件说明如表6-1所示：

表 6-1 组件说明

Name	说明
iDataAgent (iDA)	备份客户端代理，Commvault备份软件的组成部分，部署在 SAP HANA节点上，负责获取SAP HANA上需要备份的数据。
CommServe (CS)	备份服务器，Commvault备份软件的组成部分，部署在备份管理节点，负责全局备份策略的制定和备份业务的调度。
Media Agent (MA)	备份介质，Commvault备份软件的组成部分，部署在备份业务节点，负责直接将备份数据存储至OBS。
OBS	在备份场景下OBS负责存储备份数据，桶是OBS中存储数据的容器，最终数据都存储在OBS桶中。

### 说明

一个CommCell是一个备份管理域，是软件的逻辑组合，包含获取数据、传输数据、管理数据和信息的所有软件组件。

## 备份流程

### 1. 安装和预配置备份软件

在备份SAP HANA场景下，需要安装和配置备份服务器（CommServe）、备份介质（MediaAgent）及SAP HANA备份客户端代理（iDataAgent）三个组件。

### 2. 创建备份存储空间（OBS桶）

- a. 登录OBS控制台，创建一个桶，作为备份数据存储空间。详细创建桶操作请参见[创建桶](#)。
- b. 在CommCell Console上创建云存储库，输入OBS终端节点地址、访问密钥、桶名，用以将Commvault的备份介质（MediaAgent）与OBS关联。



### 说明

CommCell Console是用于管理CommCell环境、监视和控制活动作业以及查看与活动相关的事件的图形用户界面。

#### 3. 制定Commvault备份策略

在Commcell Console上创建备份策略，指定数据备份的周期、时间以及加密方式等。

#### 4. 检查备份执行情况

备份策略执行期间，用户可以通过Commcell Console查看备份执行情况。

#### 5. （可选）执行数据恢复

在SAP HANA源机上执行数据恢复。

### 说明

Commvault的具体操作请参见[Commvault官方文档](#)。

# 7 在 ECS 上通过内网访问 OBS

## 7.1 方案概述

### 应用场景

某企业基于弹性云服务器（Elastic Cloud Server，ECS）构建好基础的业务后，随着数据增长，硬盘已无法满足大量的图片、视频等数据存取需求。了解到华为云提供有海量、弹性的云存储服务OBS后，决定将OBS作为数据存储资源池，以减轻服务器负担。

在ECS上可以通过公网和华为云内网两种网络访问OBS。当有存取对象数据的需求时，公网方式响应速度会因为网络质量而受到影响，读取数据还将收取一定的流量费用。为最大化的优化性能、节省开支，企业管理者希望通过内网的方式访问OBS。

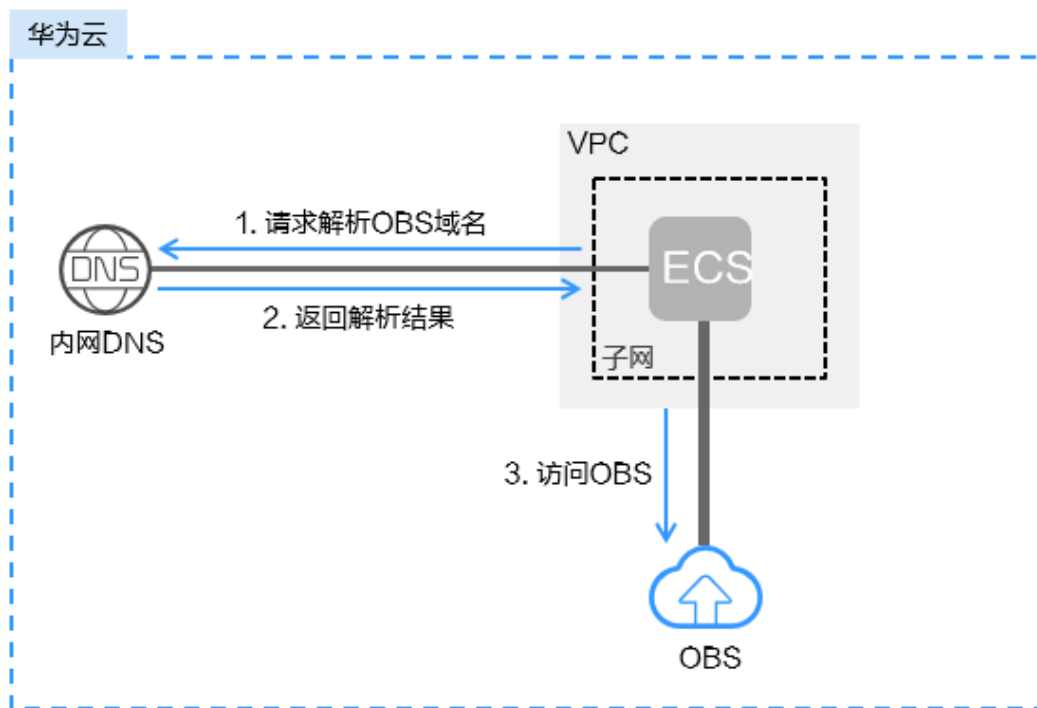
#### 📖 说明

当通过内网访问OBS时，需要确保待访问的OBS资源与ECS属于同一个区域，如果不属于同一个区域，将采用公网访问。

### 方案架构

在已搭建的ECS上通过配置内网DNS，由内网DNS解析OBS域名，即可实现在ECS上经由内网访问OBS。访问过程示意图如[图7-1](#)所示。

图 7-1 内网访问 OBS 示意图



- 对于Windows ECS，推荐使用OBS Browser+工具，实现内网访问OBS的目的，详细操作请参见：  
[在Windows ECS上使用OBS Browser+通过内网访问OBS](#)
- 对于Linux ECS，推荐使用obsutil工具，实现内网访问OBS的目的，详细操作请参见：  
[在Linux ECS上使用obsutil通过内网访问OBS](#)

当在ECS上通过内网访问OBS时，即可在内网进行数据读取、备份归档等业务，而不影响外网带宽。

## 7.2 资源成本及规划

最佳实践中涉及的资源如下：

表 7-1 资源说明

资源	资源说明
弹性云服务器 (ECS)	<ul style="list-style-type: none"> <li>• Windows系统：需安装OBS Browser+</li> <li>• Linux系统：需安装obsutil</li> </ul>
对象存储服务 (OBS)	<p>OBS作为数据存储资源池，以减轻服务器负担。</p> <p><b>须知</b> 确保待访问的OBS资源与ECS属于同一个区域。如果不属于同一个区域，将采用公网访问。</p>

资源	资源说明
虚拟私有云（VPC）	VPC主要负责为ECS构建隔离的、用户自主配置和管理的虚拟网络环境，提升用户云中资源的安全性，简化用户的网络部署。 子网是VPC中用来为ECS提供IP地址管理、DNS服务的一个网络，子网内ECS的IP地址都属于该子网。
云解析服务（DNS）	DNS提供内网DNS，专门用于处理华为云内网域名以及OBS域名的解析请求，简化域名解析流程，减少因访问公网产生的流量费用。

## 7.3 实施步骤

### 7.3.1 场景一：在 Windows ECS 上使用 OBS Browser+通过内网访问 OBS

OBSBrowser+是一款用于访问和管理对象存储服务的图形化工具，支持通过配置内网DNS服务器地址的方式，使在华为云上的Windows ECS通过内网直接访问OBS，下面将介绍具体操作流程和操作步骤。

---

#### 注意

OBSBrowser+需通过公网下载，或者从其他可以访问公网的云服务器下载后传到当前云服务器安装。

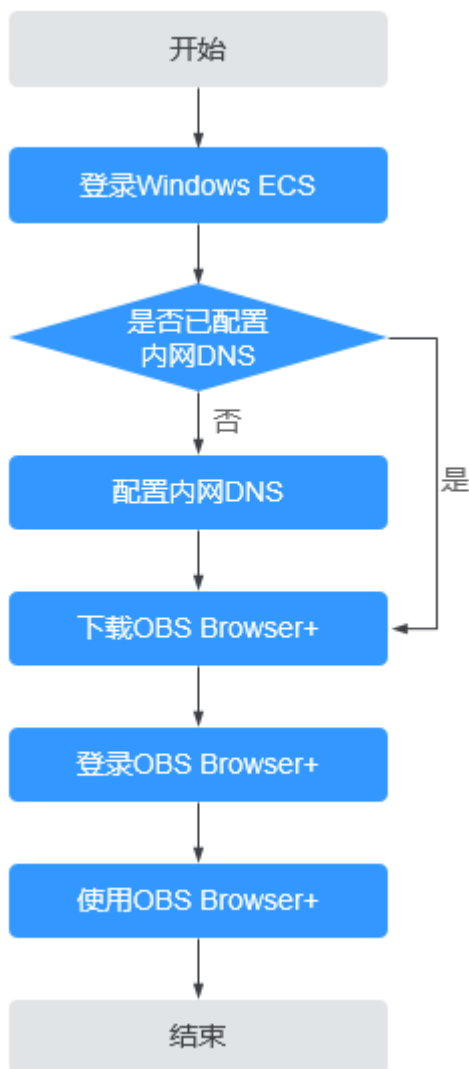
---

#### 前提条件

在登录ECS云服务器之前，请确保您已[购买云服务器](#)。

## 操作流程

图 7-2 在 Windows ECS 上使用 OBS Browser+通过内网访问 OBS 的流程



## 操作步骤

### 步骤1 登录Windows ECS

1. 登录[华为云](#)，在页面右上角单击“控制台”，进入“管理控制台”页面。
2. 在服务列表中，选择“计算 > 弹性云服务器 ECS”。
3. 选择待登录的云服务器，登录弹性云服务器。  
具体操作请参见[登录Windows弹性云服务器](#)。

### 步骤2 查看Windows ECS是否已配置内网DNS

在Windows ECS上，您可以通过图形界面和命令行两种方式查看当前的DNS配置。此处以通过命令行方式为例，介绍如何查看DNS配置。

1. 成功登录弹性云服务器后，打开cmd命令行。

2. 运行 `ipconfig /all` 命令，查看“DNS服务器”是否为当前ECS所在区域的内网DNS地址。

#### 说明

华为云针对各区域提供了不同的内网DNS服务器地址，具体请参见[华为云提供的内网DNS服务器地址](#)。

- 否，执行[步骤3](#)。
- 是，执行[步骤5](#)。

### 步骤3 配置内网DNS

修改ECS的DNS服务器地址为华为云提供的内网DNS，可以通过修改VPC子网DNS地址和修改本地DNS配置两种方式实现。

#### • 方式一：修改VPC子网DNS地址

确定ECS所在VPC，并修改VPC子网的DNS服务器地址为内网DNS地址后，可以使整个VPC内的ECS都通过内网DNS进行解析，从而访问在华为云内网的OBS服务。详细操作请参见[修改子网网络信息](#)。

#### • 方式二：修改本地DNS配置

采用此方式配置的内网DNS会在ECS每次重启后失效，在重启后需要重新配置内网DNS才可以通过内网访问OBS。此处以通过命令行配置为例，介绍如何在本地修改DNS配置。

1. 打开cmd命令行。
2. 运行以下命令，配置首选DNS服务器地址。

```
netsh interface ip set dns name="本地连接" source=static addr=内网DNS服务器地址 register=primary
```

#### 说明

- 本地连接：网卡名称，需要根据实际正在使用的网卡进行修改。
  - 内网DNS服务器地址：需要根据ECS所在区域选择内网DNS服务器地址，具体的地址信息请参见[华为云提供的内网DNS服务器地址](#)。
3. （可选）运行以下命令，配置备选DNS服务器地址。

```
netsh interface ip add dns name="本地连接" addr=备选DNS服务器地址 index=2
```

#### 说明

- 本地连接：网卡名称，需要根据实际正在使用的网卡进行修改。
- 备选DNS服务器地址：是在首选DNS服务器出现故障、不可用或无法解析请求的域名时使用的DNS服务器，因此您可以设置为华为云内网DNS服务器的地址。

### 步骤4 确认是否已经是内网访问OBS

具体方法请参见[如何判断是否内网访问OBS?](#)

### 步骤5 下载OBS Browser+

OBS Browser+下载地址及具体操作请参见[下载OBS Browser+](#)。

### 步骤6 登录OBS Browser+

由于OBS Browser+默认使用公网访问OBS，因此在登录OBS Browser+时，“服务提供商”和“服务器地址”需要按照以下要求填写：



- 服务提供商：选择“其他对象存储服务”。
- 服务器地址：根据ECS所在区域输入OBS在此区域的终端节点（Endpoint）和端口号（HTTPS协议端口号为“443”，HTTP协议端口号为“80”。系统默认服务器为HTTPS服务器）。

示例：obs.ap-southeast-1.myhuaweicloud.com:443

#### 📖 说明

OBS区域和终端节点信息请参见[地区和终端节点](#)。

### 步骤7 开始使用OBS Browser+

成功登录OBS Browser+后，便可以在Windows ECS上直接通过华为云内网访问OBS，进行基本的数据存取操作以及其他的高级设置操作。

详细使用指南请参见[对象存储服务工具指南（OBS Browser+）](#)。

----结束

## 7.3.2 场景二：在 Linux ECS 上使用 obsutil 通过内网访问 OBS

obsutil是适用于Windows、macOS和Linux操作系统的命令行工具，支持通过配置内网DNS服务器地址的方式，使在华为云上的Linux ECS通过内网直接访问OBS，下面将介绍其具体操作流程和操作步骤。

#### ⚠️ 注意

obsutil需通过公网下载，或者从其他可以访问公网的云服务器下载后传到当前云服务器安装。

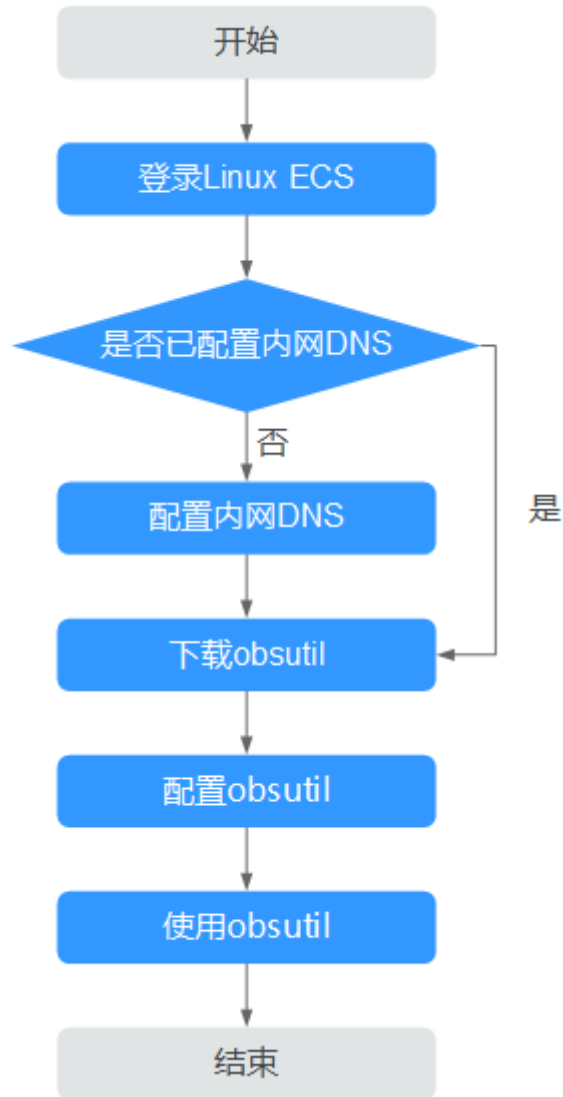


## 前提条件

在登录ECS云服务器之前，请确保您已[购买云服务器](#)。

## 操作流程

图 7-3 在 Linux ECS 上使用 obsutil 通过内网访问 OBS 的流程



## 操作步骤

### 步骤1 登录Linux ECS

1. 登录[华为云](#)，在页面右上角单击“控制台”，进入“管理控制台”页面。
2. 在打开的控制台首页，选择“计算 > 弹性云服务器”。
3. 选择待登录的云服务器，登录弹性云服务器。

由于购买Linux ECS时设置的登录鉴权方式不同，登录方式因此也存在差异，不同方式登录的具体操作请参见[登录Linux弹性云服务器](#)。

### 步骤2 查看Linux ECS是否已配置内网DNS

1. 成功登录Linux ECS后，打开命令行终端。
2. 运行`cat /etc/resolv.conf`命令，查看“nameserver”后的IP地址是否为当前ECS所在区域的内网DNS地址。

#### 📖 说明

华为云针对各区域提供了不同的内网DNS服务器地址，具体请参见[华为云提供的内网DNS服务器地址](#)。

- 否，执行[步骤3](#)。
- 是，执行[步骤5](#)。

### 步骤3 配置内网DNS

修改ECS的DNS服务器地址为华为云提供的内网DNS，可以通过修改VPC子网DNS地址和修改本地DNS配置两种方式实现。

#### • 方式一：修改VPC子网DNS地址

确定ECS所在VPC，并修改VPC子网的DNS服务器地址为内网DNS地址后，可以使整个VPC内的ECS都通过内网DNS进行解析，从而访问在华为云内网的OBS服务。详细操作请参见[修改子网网络信息](#)。

#### • 方式二：修改本地DNS配置

此处以CentOS 6.x 64bit弹性云服务器为例，介绍如何修改本地DNS配置。

- a. 打开命令行终端。
- b. 运行以下命令，打开“/etc/resolv.conf”文件。

```
vi /etc/resolv.conf
```
- c. 按下i键进入编辑模式，在“/etc/resolv.conf”文件中按照以下格式，在原有的DNS服务器地址之前新增内网DNS服务器地址。

```
nameserver 内网DNS服务器地址
```

#### 📖 说明

- 内网DNS服务器地址：需要根据ECS所在区域选择内网DNS服务器地址，具体的地址信息请参见[华为云提供的内网DNS服务器地址](#)。
  - 新增的DNS服务器地址必须位于所有原有的DNS服务器地址之前。
  - DNS服务器按照nameserver顺序选择，且仅在前一个DNS服务器出现故障、不可用或无法解析请求的域名时，才选择下一个DNS服务器。因此，后续如果想切换成公网方式，需要将首行DNS地址改为公网的DNS，或者在已有DNS服务器地址前增加一条公网DNS服务器地址。
- d. 按下Esc键，并输入:wq!，保存并退出文件。



#### 📖 说明

修改后的DNS地址在保存“/etc/resolv.conf”文件的修改操作后立即生效。

### 步骤4 确认是否已经是内网访问OBS

具体方法请参见[如何判断是否内网访问OBS?](#)

### 步骤5 下载匹配云服务器架构的obsutil

1. 单击管理控制台左上角的，选择区域和项目。
2. 单击“”，选择“计算 > 弹性云服务器”。

系统进入弹性云服务器列表页，您可以在本页面查看您已购买的弹性云服务器，以及弹性云服务器的规格，如果规格有前缀“k”则为ARM架构，此时您应该下载Linux ARM 64位的obsutil安装包，如果没有前缀“k”则为x86架构，此时您应该下载Linux AMD 64位（Linux x86 64位）的obsutil安装包。更多关于云服务器的规格信息请参见[实例类型](#)。

3. obsutil最新版本和下载链接请参见[下载obsutil](#)。

## 步骤6 配置obsutil

使用obsutil之前，您需要配置obsutil与OBS的对接信息，包括OBS终端节点（Endpoint）和访问密钥（AK和SK）。具体操作请参见obsutil指南的[初始化配置](#)章节。

### 说明

其中OBS终端节点（Endpoint）需要根据ECS所在区域输入。OBS区域和终端节点信息请参见[地区和终端节点](#)。

## 步骤7 使用obsutil

obsutil配置成功后，便可以在Linux ECS上直接通过内网访问OBS，进行基本的数据存取操作以及其他的高级设置操作。

常见的数据存储操作请参见：

- [上传对象](#)
- [下载对象](#)

详细使用指南请参见[对象存储服务工具指南（obsutil）](#)。

----结束

# 8 使用自定义域名托管静态网站

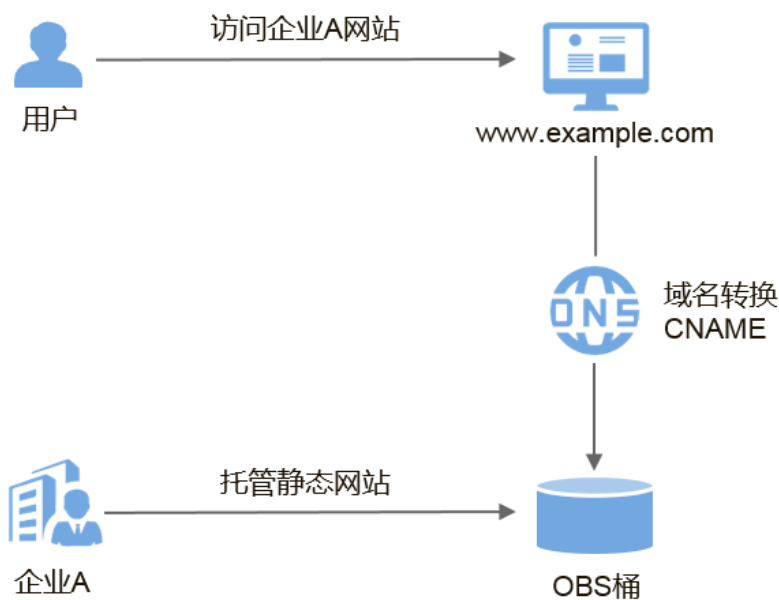
## 8.1 方案概述

### 应用场景

当企业有大量静态网站需要提供给用户访问，却不想搭建网站服务器时，可以将静态网站托管在OBS桶中，用户可以通过OBS桶绑定的自定义域名访问托管的静态网站。

### 方案架构

图 8-1 使用自定义域名访问静态网站示意图



1. 企业将静态网站托管至OBS桶中，并为OBS桶绑定对外的自定义域名。
2. 用户访问自定义域名时，即可直接访问到存储在OBS桶中的静态网站。

在此之前，您可能需要了解一些关于OBS静态网站托管的基本概念及操作，详情请参阅[静态网站托管](#)。

## 方案优势

- 快速构建基于静态内容的网站，简化建站流程，降低运营成本。
- 无需搭建服务器，静态网站轻松上线。

## 8.2 资源和成本规划

本节介绍最佳实践中资源规划情况，包含以下内容：

表 8-1 资源和成本规划说明

资源	资源说明	成本说明
OBS	需要创建一个OBS桶用于存放静态网站文件，同时在OBS桶上完成静态网站托管配置和自定义域名绑定。	OBS的使用涉及以下几项费用： <ul style="list-style-type: none"><li>• 存储费用：静态网站文件存储在OBS中产生的<b>存储费用</b>。</li><li>• 请求费用：用户访问OBS中存储的静态网站文件时产生的<b>请求费用</b>。</li><li>• 流量费用：用户使用自定义域名通过公网访问OBS时产生的<b>流量费用</b>。</li></ul> 实际产生的费用与存储的文件大小、用户访问所产生的请求次数和流量大小有关，请根据自己的业务进行预估。
静态网站文件	<ul style="list-style-type: none"><li>• 静态网站首页：访问静态网站时返回的索引页面，即首页。 示例：index.html</li><li>• 404错误页面：当访问错误的静态网站路径时，返回的404错误页面。 示例：error.html</li></ul>	免费
自定义域名	用户自己的域名地址，需要绑定在OBS桶上。 按照工信部要求，您绑定自定义域名的桶如果在以下区域，需要提前完成ICP备案。 包括：华北-北京一、华北-北京四、华东-上海一、华东-上海二、华南-广州 示例：www.example.com	准备自定义域名涉及域名注册费用，由域名注册商收取，具体费用以实际为准。

资源	资源说明	成本说明
DNS	OBS桶绑定的自定义域名需要在DNS上配置CNAME记录。	免费

本例中，静态网站文件的示例如下：

- index.html的内容为：

```
<html>
<head>
  <title>Hello OBS!</title>
  <meta charset="utf-8">
</head>
<body>
  <p>欢迎使用OBS静态网站托管功能</p>
  <p>这是首页</p>
</body>
</html>
```

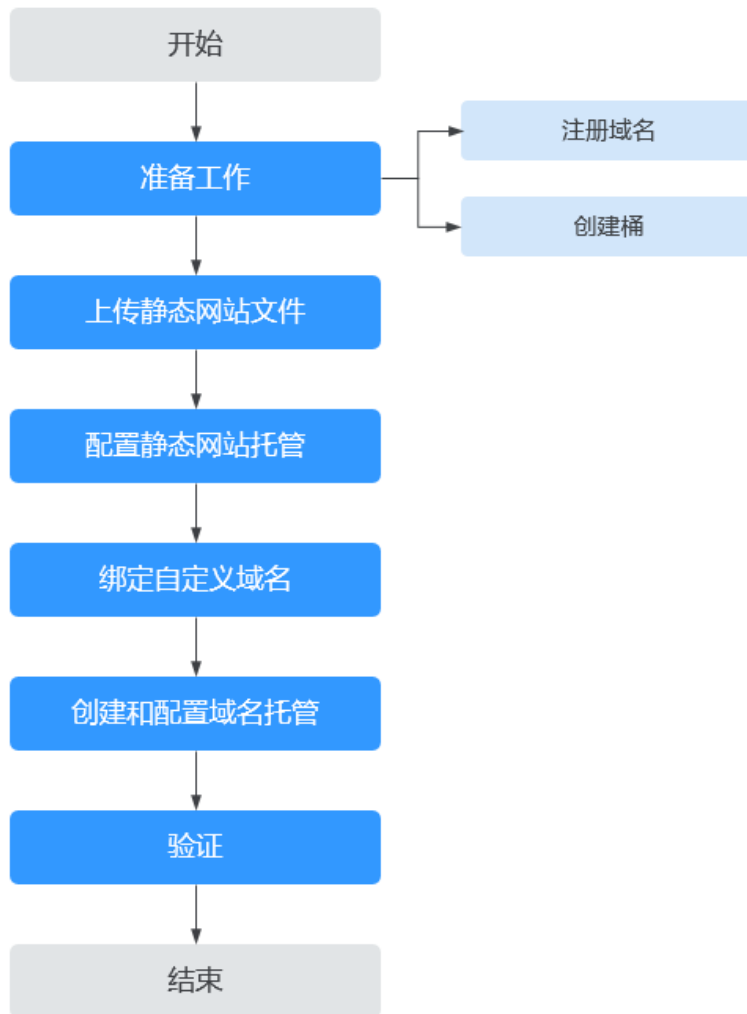
- error.html的内容为：

```
<html>
<head>
  <title>Hello OBS!</title>
  <meta charset="utf-8">
</head>
<body>
  <p>欢迎使用OBS静态网站托管功能</p>
  <p>这是404错误页面</p>
</body>
</html>
```

## 8.3 操作流程

您需要先在OBS管理控制台上创建一个桶，用于存放静态网站资源，并启用该桶的静态网站托管，然后通过OBS提供的绑定自定义域名功能，将自定义域名与新创建的桶绑定，再通过云解析服务（Domain Name Service，DNS）创建和配置域名托管，实现自定义域名访问托管在OBS上的静态网站。具体操作流程如下：

图 8-2 使用自定义域名托管静态网站流程图



## 8.4 实施步骤

### 8.4.1 准备工作

#### 注册域名

如果您拥有一个已注册的域名，可跳过本步骤。

如果您还没有，请选择一个合适的注册商注册一个属于自己企业的域名。在本场景下，以数据规划中的示例域名www.example.com进行注册，在实际操作中，您需要将此域名替换为您自己规划的域名。

#### 创建桶

桶名没有特殊要求，您只需要按照界面提示的命名规则创建一个桶用于存储静态网站文件。此处以创建一个桶名称为example的桶为例，其具体操作步骤如下：

**步骤1** 打开[OBS管理控制台](#)，根据页面提示进行登录。

**步骤2** 在页面上方单击“创建桶”。

**步骤3** 在弹出的对话框中配置以下参数。

- **区域**：根据就近原则选择离业务较近的区域。
- **默认存储类别**：推荐选择“标准存储”。

#### 说明

您也可以根据网站的访问频率以及对响应速度的要求，选择“低频访问存储”或“归档存储”。存储类别详细介绍请参见[桶存储类别简介](#)。

- **桶名称**：输入“example”。
- **桶策略**：选择“公共读”使桶内对象能够被任何用户访问。
- **服务端加密**：选择“不开启加密”。
- **企业项目**：请先[创建企业项目](#)，默认为default企业项目，然后在创建桶时选择对应企业项目。仅企业账号能够配置企业项目。

**步骤4** 单击“立即创建”，完成桶创建。

----结束

## 8.4.2 上传静态网站文件

整理好待上传的静态网站文件，在OBS控制台重复执行以下步骤，直至所有的静态网站文件都上传至[准备工作](#)创建的桶中。

#### 说明

在支持批量上传的区域，OBS控制台每次最多支持100个文件同时上传，总大小不超过5GB，如果网站文件较多，建议使用OBS Browser+上传，具体操作步骤请参见[使用OBS Browser+上传文件或文件夹](#)。

**步骤1** 单击待操作的桶名称，进入桶对象页面。

**步骤2** 单击“上传对象”，系统将弹出如下所示对话框。



图 8-3 上传对象

上传对象 [超过5GB如何上传?](#)

① 上传对象 ———— ② 高级配置 (可选)

1 上传操作将产生请求费用, 上传成功后将产生存储费用。

存储类别: **标准存储** | 低频访问存储 | 归档存储

适用于有大量热点文件或小文件, 且需要频繁访问 (平均一个月多次) 并快速获取数据的业务场景。  
对象默认与桶的存储类别相同, 也可以根据适用场景修改。 [了解更多](#)

上传对象

桶内如有同名文件/文件夹, 将被新上传的文件/文件夹覆盖, 如您需要保留多个版本的文件/文件夹, 请开启多版本控制。

拖拽本地文件或文件夹至此处, 或 [添加文件](#)  
(单次最多支持100个文件同时上传, 总大小不超过5GB)

服务端加密: SSE-KMS | SSE-OBS | **不开启加密**

开启服务端加密后, 上传到当前桶的对象会被加密。 [了解更多](#)

建议开启加密, 核心数据更安全, 如果您使用KMS加密模式, 超过免费配额会收取相应费用。 [价格详情](#)

下一步: [高级配置 \(可选\)](#) 取消 上传

### 步骤3 添加待上传的文件。

#### 说明

- 不可加密上传静态网站文件。
- 存储类别建议选择“标准”。如果静态网站文件的存储类别为“归档存储”，则需要先恢复才能被访问，具体恢复步骤请参见[恢复归档存储文件](#)。
- 网站首页文件（index.html）和404错误页面（error.html），需要存放在桶的根目录下。

### 步骤4 单击“上传”完成文件上传。

---结束

## 8.4.3 配置静态网站托管

上传完静态网站文件后，您需要执行以下步骤，将当前桶设置为静态网站托管模式。

#### 说明

您也可以将整个静态网站直接重定向至另一个桶或域名，配置操作请参见[重定向请求](#)。

**步骤1** 单击桶名称，进入桶对象页面后单击“基础配置 > 静态网站托管”。

**步骤2** 单击“配置静态网站托管”按钮。

**步骤3** 在弹出的对话框中，开启静态网站托管并选择“配置到当前桶”，将“默认首页”配置为数据规划中的index.html，将“默认404错误页面”配置为数据规划中的error.html，如下所示。

图 8-4 配置静态网站托管



#### 说明

您也可以根据业务需求配置重定向规则，实现网站内容重定向，具体操作请参见[配置静态网站托管](#)。

**步骤4** 单击“确定”。

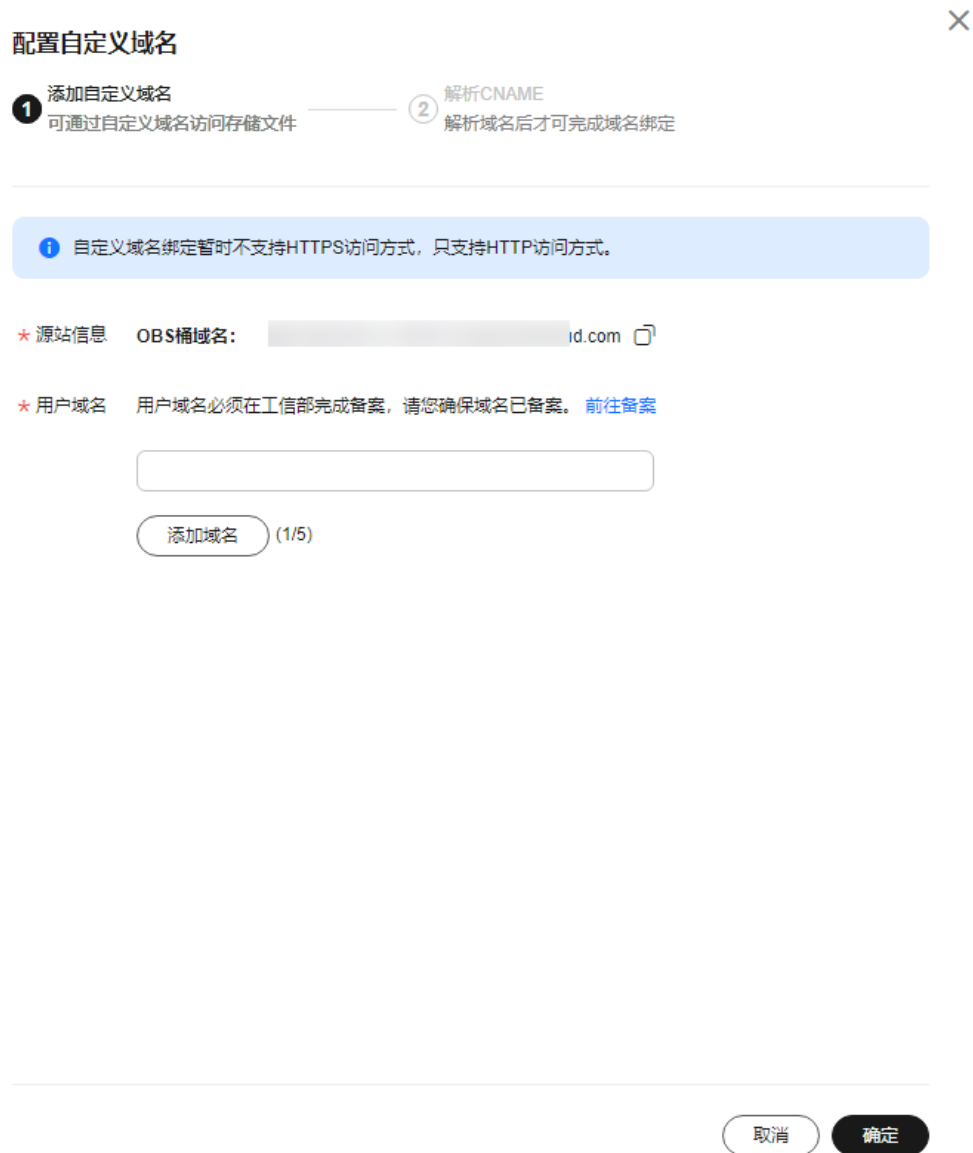
----结束

## 8.4.4 配置自定义域名

**步骤1** 单击桶名称进入“对象”页面，在左侧导航栏选择“域名管理”。

**步骤2** 单击页面上方的“配置自定义域名”，如果没有绑定自定义域名也可以在页面下方的配置自定义域名卡片中单击“配置自定义域名”，在“用户域名”输入“www.example.com”，如下所示。

图 8-5 配置自定义域名



**步骤3** 单击“确定”。

**步骤4** 根据页面提示，支持单击“一键解析”，或手动完成解析CNAME，单击右下角的“我已了解”。

#### 📖 说明

华为云域名支持一键解析添加CNAME记录；非华为云域名不支持一键解析，请用户自行配置解析规则。

---结束

## 8.4.5 创建和配置域名托管

为了方便对您的自定义域名和静态网站统一管理，实现业务全面云化，您可以直接在华为云提供的云解析服务（Domain Name Service，DNS）上托管您的自定义域名。托管完成后，后续域名解析的管理都可以在云解析服务上进行，包括：管理记录集、管理反向解析、设置域名泛解析等等。

### 📖 说明

您也可以直接在域名注册商域名解析中添加一条别名记录，指向桶的访问域名。例如：桶“www.example.com”所处区域“中国-香港”，则需要在域名注册商添加一条值为“www.example.com CNAME www.example.com.obs-website.ap-southeast-1.myhuaweicloud.com”的记录。

使用云解析服务创建和配置域名托管的操作步骤如下：

#### 步骤1 创建公网域名。

在云解析服务中创建公网域名，使用**准备工作**中注册的根域名“example.com”作为创建公网域名。详细的创建方法请参见**配置网站解析**章节中的“添加域名”部分内容。

#### 步骤2 添加别名记录。

在云解析服务中为托管域名子域名“www.example.com”添加记录集，配置该子域名别名指向OBS的静态网站托管域名。在添加别名记录时参数配置如下：

- **主机记录**：输入“www”。
- **类型**：选择“CNAME – 将域名指向另外一个域名”。
- **线路类型**：选择“全网默认”。
- **TTL(秒)**：保持默认。
- **值**：需指向的域名。如果绑定自定义域名时没有开启CDN加速，此处填写OBS的桶的静态网站托管域名；如果开启了CDN加速，此处填写CDN提供的加速域名（即CNAME）。

详细的创建方法请参见**增加CNAME类型记录集**。

#### 步骤3 在域名注册商处修改域名解析服务器地址。

在域名注册商处，将该根域名对应的NS记录中域名解析服务器地址修改为云解析服务（DNS）服务器的地址，具体地址为云解析服务中该公网域名记录集中NS记录的值字段内容信息。

详细的更改域名解析服务器地址的方法请参见**配置网站解析**章节中的“更改域名的DNS服务器”部分。

### 📖 说明

更改后的域名解析服务器地址将于48小时内生效，具体生效时间请以域名注册商处的说明为准。

----结束

## 8.4.6 验证

- 在浏览器中输入访问地址：www.example.com，验证能否访问到配置的默认首页，如**图8-6**所示。

图 8-6 默认首页



- 在浏览器中输入一个桶中不存在的静态文件访问地址，例如：`www.example.com/imgs`，验证能否访问到配置的404错误页面，如图8-7所示。

图 8-7 404 错误页面



#### 📖 说明

由于浏览器缓存等原因，您可能需要清除浏览器缓存后才能查看到预期效果。

## 8.4.7 （后续操作）更新静态网站

后续如果需要对网站某个静态文件（如：图片、音乐、html文件、css文件等）进行更新，您可以重新上传该静态文件。但需要注意的是，默认情况下，在OBS同一路径下新上传的文件会覆盖OBS上已存在的同名文件。为避免文件覆盖的情况，您可以选择启用OBS的多版本控制功能。利用多版本控制，可以保留静态文件的多个版本，使您更方便地检索和还原各个版本，在意外操作或应用程序故障时快速恢复数据。

### 启用多版本控制

- 步骤1** 登录OBS管理控制台。
- 步骤2** 在桶列表中单击待操作的桶，进入桶对象页面后在左侧导航栏单击“概览”。
- 步骤3** 在“基础配置”区域下，单击“多版本控制”卡片，系统弹出多版本控制对话框。

图 8-8 多版本控制

### 多版本控制

启用

暂停

暂停多版本控制，不会影响已经存在的历史版本对象。

取消

确定

**步骤4** 勾选“启用”后单击“确定”，启用目标桶中对象的多版本控制。

----结束

关于多版本控制的更多介绍以及操作指导，请参见[多版本控制](#)。

## 更新静态文件

**步骤1** 登录OBS管理控制台。

**步骤2** 在桶列表中单击待操作的桶，进入对象页面。

**步骤3** 单击“上传对象”，或选择待更新文件所在文件夹后单击“上传对象”。

图 8-9 上传对象

**上传对象** [超过5GB如何上传?](#)

① 上传对象 ———— ② 高级配置 (可选)

1 上传操作将产生请求费用, 上传成功后将产生存储费用。

存储类别: **标准存储** | 低频访问存储 | 归档存储

适用于有大量热点文件或小文件, 且需要频繁访问 (平均一个月多次) 并快速获取数据的业务场景。  
对象默认与桶的存储类别相同, 也可以根据适用场景修改。 [了解更多](#)

上传对象

桶内如有同名文件/文件夹, 将被新上传的文件/文件夹覆盖。如您想要保留多个版本的文件/文件夹, 请开启多版本控制。

拖拽本地文件或文件夹至此处, 或 [添加文件](#)  
(单次最多支持100个文件同时上传, 总大小不超过5GB)

服务端加密: SSE-KMS | SSE-OBS | **不开启加密**

开启服务端加密后, 上传到当前桶的对象会被加密。 [了解更多](#)

建议开启加密, 核心数据更安全, 如果您使用KMS加密模式, 超过免费配额会收取相应费用。 [价格详情](#)

下一步: 高级配置 (可选) 取消 上传

#### 步骤4 添加待上传文件。

##### 📖 说明

- 不可加密上传静态网站文件。
- 存储类别建议选择“标准”。如果静态网站文件的存储类别为“归档存储”，则需要先恢复才能被访问，具体恢复步骤请参见[恢复归档存储文件](#)。

#### 步骤5 单击“上传”完成文件上传。

在同一路径下新上传的同名文件会作为“最新版本”显示在对象列表，每次访问此文件时，都是访问的此文件的最新版本，以此达到更新静态网站文件的效果。

----结束

# 9 OBS 数据一致性校验

## 9.1 方案概述

### 应用场景

对象数据在上传下载过程中，有可能会因为网络劫持、数据缓存等原因，存在数据不一致的问题。

### 方案架构

OBS提供通过计算MD5值的方式对上传下载的数据进行一致性校验。默认情况下，OBS不会进行一致性校验，您可以通过以下方式在上传下载时主动启用校验。

#### 说明

- 各种方式的一致性校验结果互通，即无论您使用以下何种方式在上传时通过了一致性校验，都可以使用其他方式在下载时校验一致性。
- 下载对象时，只有当待下载对象具有MD5值时，MD5校验才会生效。
- 启用MD5进行数据一致性校验会影响上传下载性能。

表 9-1 校验数据一致性的方式

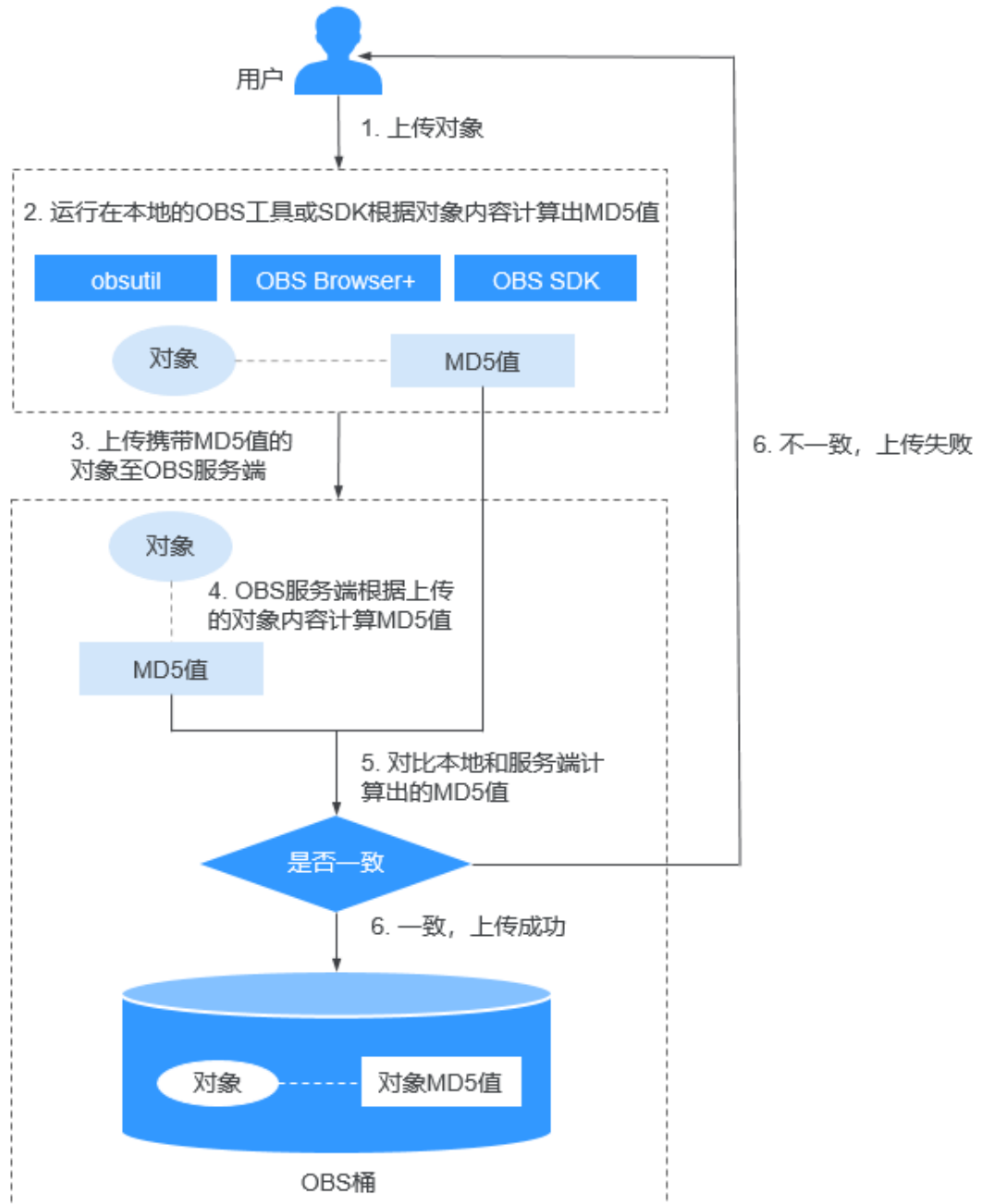
方式	说明	操作指导
obsutil	命令行工具，可以通过简单的一行命令实现上传下载，并且在命令中选择是否采用MD5校验。	使用obsutil校验上传对象的一致性
		使用obsutil校验下载对象的一致性
OBS Browser +	图形化界面工具，可以一键开启或关闭MD5校验，同时提供任务管理，方便查看校验状态。	使用OBS Browser+校验上传对象的一致性
		使用OBS Browser+校验下载对象的一致性



方式	说明	操作指导
OBS SDK	开发者可以通过OBS SDK进行二次开发，自行判断MD5校验结果，并根据实际业务进行结果处理。	使用OBS SDK校验上传对象的一致性
		使用OBS SDK校验下载对象的一致性

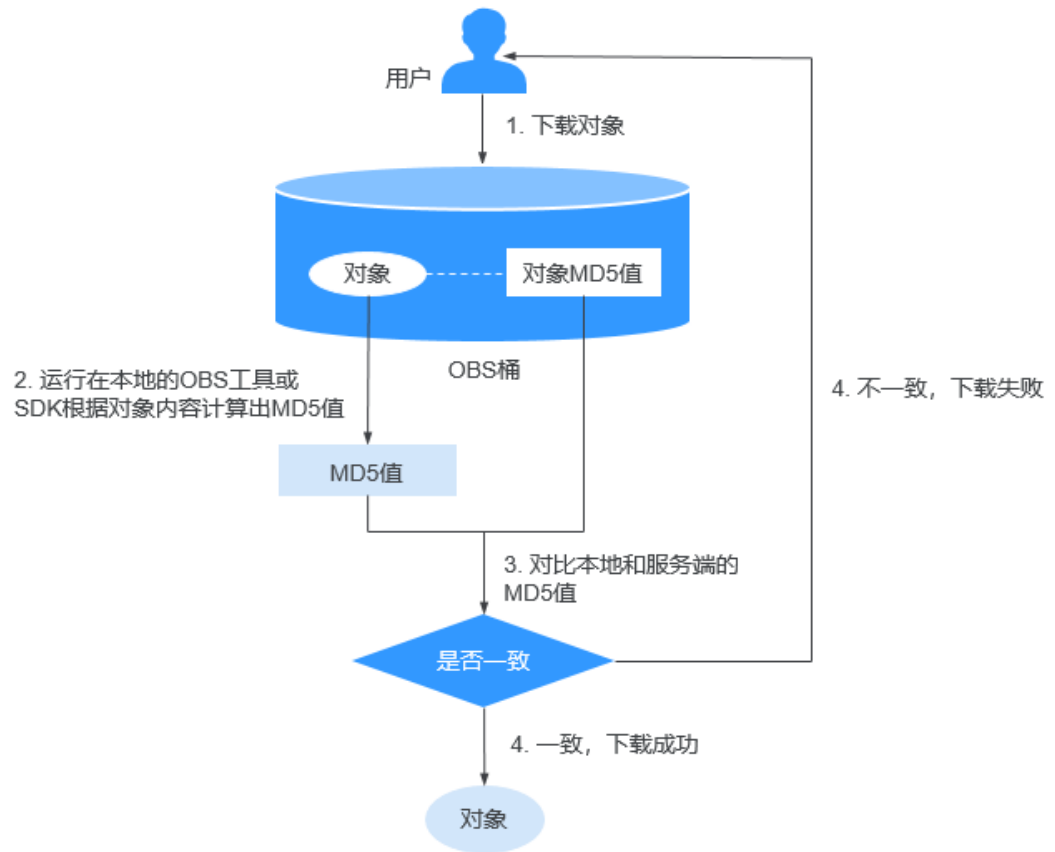
上传对象时，OBS会先在客户端计算出对象的MD5值然后携带上传至OBS，OBS服务端再根据上传的对象内容计算出MD5值，最终与携带上传的MD5值进行对比，如果对比结果一致，对象上传成功，否则上传失败。使用MD5值对上传数据进行一致性校验的示意图如[图9-1](#)所示。

图 9-1 校验上传对象的一致性



下载对象时，OBS会将对象已有的MD5值与根据下载的对象内容计算出来的MD5值进行对比，如果对比结果一致，对象下载成功，否则下载失败。使用MD5值对下载数据进行一致性校验的示意图如图9-2所示。

图 9-2 校验下载对象的一致性



## 9.2 校验上传对象的一致性

obsutil、OBS Browser+以及OBS SDK都支持在上传对象时进行一致性校验，您可以根据自己的业务选择任意一种方式进行校验。本文分别介绍了几种方式在上传对象时进行一致性校验的操作指导。

### 方式一：使用 obsutil 校验上传对象的一致性

obsutil支持在上传对象时通过附加参数（vmd5）来校验数据的一致性。

以在Windows操作系统上传本地一个位于D盘的test.txt文件至mytestbucket桶为例，开启一致性校验的命令示例如下：

```
obsutil cp D:\test.txt obs://mytestbucket/test.txt -vmd5
```

校验通过后，对象上传成功，系统显示Upload successfully的回显信息。

```
D:\obsutil>obsutil cp D:\test.txt obs://mytestbucket/text.txt -vmd5
Start at 2019-06-11 01:43:07.5075864 +0000 UTC

Parallel:      5                Jobs:          5
Threshold:    50.00MB          PartSize:     5.00MB
VerifyLength: false          VerifyMd5:    true
Exclude:
Include:
CheckpointDir: C:\Users\...\.obsutil_checkpoint

[-----] 100.00% ?/s 291B/291B 50ms
Upload successfully, 291B, 2b0ec156a68e0035ea18d85c1b3ce619, D:\test.txt --> obs
://mytestbucket/text.txt, cost [108], status [200], request id [0000016B44342DC6
860CCDE792CFFC25]
```

## 方式二：使用 OBS Browser+校验上传对象的一致性

OBS Browser+默认关闭MD5校验，在OBS Browser+上启用MD5校验一致性并上传对象的步骤如下：


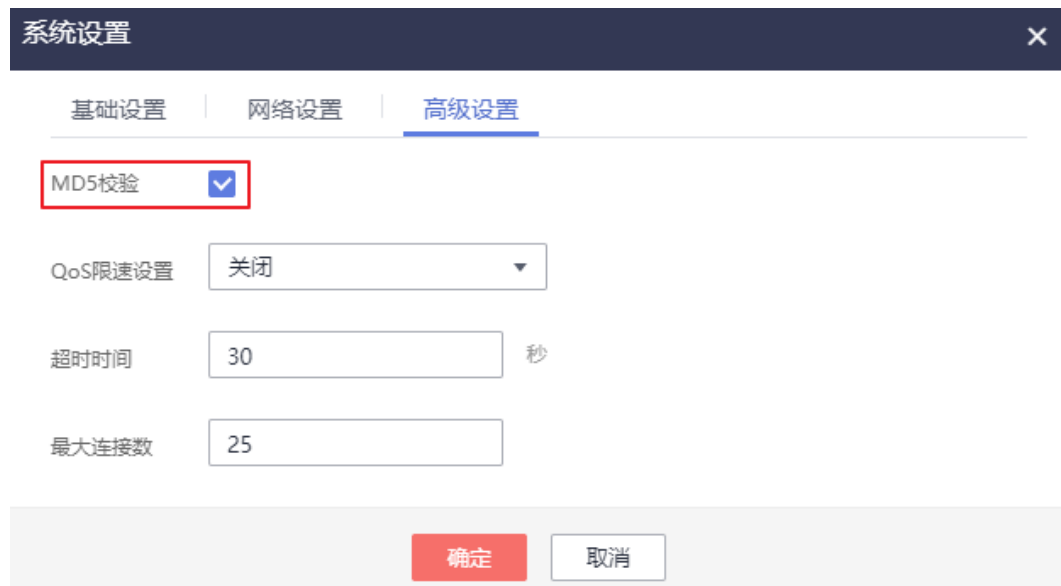
- 步骤1 登录OBS Browser+。
- 步骤2 单击客户端右上方的  设置，并选择“高级设置”。
- 步骤3 勾选“MD5校验”，如图9-3所示。

图 9-3 配置 MD5 校验



- 步骤4 单击“确定”。
- 步骤5 选择待上传文件的桶，上传文件。
  - 如果MD5校验成功，则文件上传成功。
  - 如果MD5校验失败，则文件上传失败，且在任务管理中提示失败原因：校验文件MD5失败。

----结束

## 方式三：使用 OBS SDK 校验上传对象的一致性

OBS提供Java、Python等多种语言的SDK，各SDK通过在上传对象时设置对象的Content-MD5值以开启一致性校验。如何计算并设置对象MD5值请前往[OBS SDK](#)参见各自开发指南的setObjectMetadata接口。

此处以使用OBS Java SDK上传Windows本地D盘一个名为text.txt的文本文件至mytestbucket为例，上传过程使用MD5值校验数据一致性的示例代码如下：

```
String endPoint = "https://your-endpoint";
// 认证用的ak和sk硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密文存放，使用时解密，确保安全；本示例以ak和sk保存在环境变量中为例，运行本示例前请先在本地环境中设置环境变量ACCESS_KEY_ID和SECRET_ACCESS_KEY_ID。
// 您可以登录访问管理控制台获取访问密钥AK/SK，获取方式请参见https://support.huaweicloud.com/intl/zh-cn/usermanual-ca/ca_01_0003.html
String ak = System.getenv("ACCESS_KEY_ID");
String sk = System.getenv("SECRET_ACCESS_KEY_ID");
// 创建ObsClient实例
ObsClient obsClient = new ObsClient(ak, sk, endPoint);
// 计算并设置MD5值
ObjectMetadata metadata = new ObjectMetadata();
File file = new File("D:\\text.txt");
FileInputStream fis = new FileInputStream(file);
InputStream is = (InputStream) fis;
String contentMd5 = obsClient.base64Md5(is);
metadata.setContentMd5(contentMd5);
// 带MD5值上传文件
obsClient.putObject("mytestbucket", "text.txt", file, metadata);
```

### 说明

- 对象数据的MD5值必须经过Base64编码。
- OBS服务端会将该MD5值与对象数据计算出的MD5值进行对比，如果不匹配则上传失败，返回HTTP 400错误。如果匹配，对象上传成功，返回HTTP 200状态码。

## 9.3 校验下载对象的一致性

OBS Browser+、obsutil以及OBS SDK都支持在下载对象时进行一致性校验，您可以根据自己的业务选择任意一种方式进行校验，本文就几种方式如何使用一致性校验进行了详细说明。

### 前提条件

待下载对象已有MD5值，如果没有MD5值，将不会进行一致性校验。对象的MD5值需要在上传的时候计算并设置，详细操作请参见[校验上传对象的一致性](#)。

### 方式一：使用 obsutil 校验下载对象的一致性

obsutil支持在下载对象时通过附加参数（vmd5）来校验下载数据的一致性。

以在Windows操作系统下载mytestbucket桶中的test.txt文件至本地为例，开启数据一致性校验的步骤如下：

**步骤1** 执行以下命令，检查待下载对象是否具有MD5信息。

```
obsutil stat obs://test-bucket/test.txt
```

- 返回的对象基本信息中，包含MD5信息，如下图所示，执行**步骤2**。

```
D:\obsutil_windows_amd64>obsutil stat obs://mytestbucket/text.txt
Start at 2019-06-10 09:07:00.9978182 +0000 UTC

Key:
  obs://mytestbucket/text.txt
LastModified:
  2019-06-10T09:04:26Z
Size:
  291
StorageClass:
  standard
MD5:
  2b0ec156a68e0035ea18d85c1b3ce619
Etag:
  2b0ec156a68e0035ea18d85c1b3ce619
ContentType:
  text/plain
```

- 不包含MD5信息，下载对象时无法进行一致性校验。

### 步骤2 执行以下命令，下载对象。

```
obsutil cp obs://mytestbucket/test.txt D:\test.txt -vmd5
```

- 对象下载成功且通过一致性校验，回显信息如下：

```
Download successfully, 317B, a6d2a254f93af83c6efe59232bdbb4e0, obs://mytestbucket/test.txt --> D:\test.txt, cost [50], status [200], request id [0000016B4466E8C3860BFF29740B5669]
```

- 如果桶中对象没有MD5值，对象能够下载成功，但不会校验一致性，回显信息如下：

```
Download successfully, 317B, n/a, obs://mytestbucket/test.txt --> D:\test.txt, cost [100], status [200], request id [0000016B445FA2CB860DCF05B537DF8E]
Warn: Cannot get the valid md5 value of key [text.txt] in bucket [mytestbucket] to check
```

----结束

## 方式二：使用 OBS Browser+校验下载对象的一致性

OBS Browser+默认关闭MD5校验，在OBS Browser+上启用MD5校验一致性并下载对象的步骤如下：

步骤1 登录OBS Browser+。

步骤2 单击客户端右上方的  设置，并选择“高级设置”。

步骤3 勾选“MD5校验”，如图9-4所示。

图 9-4 配置 MD5 校验



**步骤4** 单击“确定”。

**步骤5** 选择待下载文件的桶，下载文件。

- 如果MD5校验成功，则文件下载成功。
- 如果MD5校验失败，则文件下载失败，且在任务管理中提示失败原因：校验文件MD5失败。

----结束

### 方式三：使用 OBS SDK 校验下载对象的一致性

OBS SDK对待下载对象的自定义元数据中的MD5值和下载到本地的对象的MD5值进行对比，通过对比结果判断下载对象的一致性。

此处以使用OBS Java SDK下载mytestbucket桶中一个名为test.txt的文本文件为例，下载过程使用MD5值校验数据一致性的示例代码如下：

```
String endPoint = "https://your-endpoint";
// 认证用的ak和sk硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密文存放，使用时解密，确保安全；本示例以ak和sk保存在环境变量中为例，运行本示例前请先在本地环境中设置环境变量ACCESS_KEY_ID和SECRET_ACCESS_KEY_ID。
// 您可以登录访问管理控制台获取访问密钥AK/SK，获取方式请参见https://support.huaweicloud.com/intl/zh-cn/usermanual-ca/ca_01_0003.html
String ak = System.getenv("ACCESS_KEY_ID");
String sk = System.getenv("SECRET_ACCESS_KEY_ID");
// 创建ObsClient实例
final ObsClient obsClient = new ObsClient(ak, sk, endPoint);
// 获取对象的MD5值
ObjectMetadata metadata = obsClient.getObjectMetadata("mytestbucket", "test.txt");
String md5Origin = metadata.getUserMetadata("contentMd5");
// 计算下载后对象的MD5值
ObsObject obsObject = obsClient.getObject("mytestbucket", "test.txt");
String md5Download = obsClient.base64Md5(obsObject.getObjectContent());
// 对比MD5值
if(md5Origin.contentEquals(md5Download))
    System.out.println("Object MD5 validation passes!\n");
else
    System.out.println("Object MD5 validation failed!\n");
```

### 说明

在以上示例代码中，获取对象MD5值时的contentMd5是在[上传](#)时设置的自定义元数据，实际开发中需要根据自定义的元数据名称修改。



# 10 OBS 性能优化建议

OBS按照对象名的UTF-8编码范围来进行分区管理，对系统进行水平扩展与动态负载均衡。如果用户在对象命名规则上使用了顺序前缀（如时间戳或字母顺序），可能导致大量对象的请求访问集中于某个特定分区，造成访问热点。热点分区上的请求速率受限，访问时延上升。

**推荐使用随机前缀对象名，这样请求就会均匀分布在多个分区，达到水平扩展的效果。**

示例：

比如典型的日志归档场景，可能上传的对象名都是如下形式：

```
yourbucket/obslog/20190610-01.log.tar.gz
yourbucket/obslog/20190610-02.log.tar.gz
yourbucket/obslog/20190610-03.log.tar.gz
yourbucket/obslog/20190610-04.log.tar.gz
...
yourbucket/obslog/20190611-01.log.tar.gz
yourbucket/obslog/20190611-02.log.tar.gz
yourbucket/obslog/20190611-03.log.tar.gz
yourbucket/obslog/20190611-04.log.tar.gz
```

建议为对象名添加3位以上16进制哈希前缀：

```
yourbucket/6ac-obslog/20140610-01.log.tar.gz
yourbucket/b42-obslog/20140610-02.log.tar.gz
yourbucket/17f-obslog/20140610-03.log.tar.gz
yourbucket/ac9-obslog/20140610-04.log.tar.gz
...
yourbucket/95d-obslog/20140611-01.log.tar.gz
yourbucket/4a5-obslog/20140611-02.log.tar.gz
yourbucket/ea2-obslog/20140611-03.log.tar.gz
yourbucket/ba3-obslog/20140611-04.log.tar.gz
```

# 11 Web 端通过 PostObject 接口直传 OBS

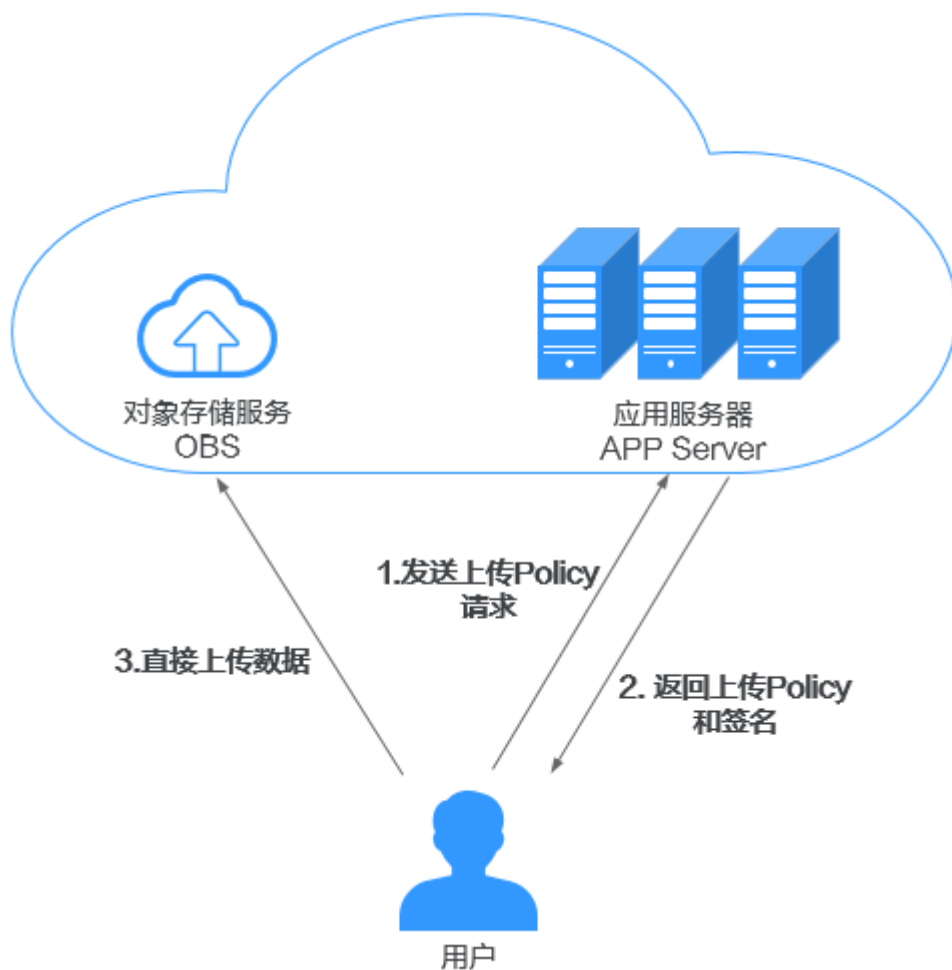
---

## 背景信息

常见的Web端上传方法是用户通过浏览器上传文件至应用服务器，再由应用服务器上传至OBS，数据需要在应用服务器中转，传输效率较低，且多任务同时上传时应用服务器压力大。

本文介绍一种在Web端利用PostObject接口直传文件至OBS的方法，即使用表单上传方式上传文件至OBS。如图11-1所示，该方案省去了应用服务器这一步骤，提高了传输效率，不会对服务器产生压力，且服务端签名后直传可以保证传输的安全性。

图 11-1 Web 端 PostObject 直传流程图



## 前提条件

已创建桶。具体操作请参见[创建桶](#)。

## 操作步骤

配置分为两大步：配置跨域资源共享和使用表单上传。

### 第一步：配置跨域资源共享

在通常的网页请求中，由于同源安全策略SOP的存在，不同域之间的网站脚本和内容是无法进行交互的。

跨域资源共享CORS是一种浏览器的规范机制，定义了一个域中加载的客户端Web应用程序与另一个域中的资源交互的方式。OBS支持CORS规范，允许跨域请求访问OBS中的资源。

**步骤1** 在OBS管理控制台左侧导航栏选择“对象存储”。

**步骤2** 在桶列表单击待操作的桶，进入对象页面。

**步骤3** 在左侧导航栏，单击“访问权限控制 > CORS规则”。

**步骤4** 单击“创建”，系统弹出“创建CORS规则”对话框，如[图11-2](#)所示。

### 说明

一个桶最多可设置100条CORS规则。

图 11-2 创建 CORS 规则

创建CORS规则 如何配置?

\* 允许的来源  0/1,024

\* 允许的方法  Get  Post  Put  Delete  Head

允许的头域  0/1,024

补充头域  0/1,024

缓存时间(秒)

**步骤5** 在“CORS规则”中配置“允许的来源”、“允许的方法”、“允许的头域”、“补充头域”和“缓存时间”。

### 说明

如果该OBS桶同时开启了CDN加速，CDN需配置HTTP header，详见[HTTP header配置](#)。

表 11-1 CORS 规则

参数	说明	配置建议
允许的来源	必选参数，指定允许的跨域请求的来源，即允许来自该域名下的请求访问该桶。 允许多条匹配规则，以回车换行为间隔。每个匹配规则允许使用最多一个“*”通配符。例如： http://rds.example.com https://*.vbs.example.com	*
允许的方法	必选参数，指定允许的跨域请求方法，即桶和对象的几种操作类型。包括：Get、Post、Put、Delete、Head。	全选

参数	说明	配置建议
允许的头域	<p>可选参数，指定允许的跨域请求的头域。只有匹配上允许的头域中的配置，才被视为是合法的CORS请求。</p> <p>允许的头域可设置多个，多个头域之间换行隔开，每行最多可填写一个*符号，不支持&amp;、:、&lt;、空格以及中文字符。</p>	*
补充头域	<p>可选参数，指CORS响应中带的补充头域，给客户端提供额外的信息。</p> <p>补充头域可设置多个，多个头域之间换行隔开，不支持*、&amp;、:、&lt;、空格以及中文字符。</p>	<ul style="list-style-type: none"> <li>● ETag</li> <li>● x-obs-request-id</li> <li>● x-obs-api</li> <li>● Content-Type</li> <li>● Content-Length</li> <li>● Cache-Control</li> <li>● Content-Disposition</li> <li>● Content-Encoding</li> <li>● Content-Language</li> <li>● Expires</li> <li>● x-obs-id-2</li> <li>● x-reserved-indicator</li> <li>● x-obs-version-id</li> <li>● x-obs-copy-source-version-id</li> <li>● x-obs-storage-class</li> <li>● x-obs-delete-marker</li> <li>● x-obs-expiration</li> <li>● x-obs-website-redirect-location</li> <li>● x-obs-restore</li> <li>● x-obs-version</li> <li>● x-obs-object-type</li> <li>● x-obs-next-append-position</li> </ul>
缓存时间	<p>必选参数，请求来源的客户端可以缓存的CORS响应时间，以秒为单位，默认为100秒。</p>	<p>根据实际业务设置。</p>

**步骤6** 单击“确定”。

“CORS规则”页签显示“创建CORS规则成功”提示创建桶的CORS配置成功。CORS配置会在两分钟内生效。

CORS配置成功后，便仅允许跨域请求来源的地址通过允许的方法访问OBS的桶。例如：为桶“testbucket”允许的来源配置为“https://www.example.com”，允许的方法配置为“GET”，允许的头域配置为“\*”，补充头域配置为“ETag”，缓存时间设置为“100”，则OBS仅允许来源为“https://www.example.com”的“GET”请求访问桶“testbucket”，且不限该请求的头域，允许响应中返回ETag值，请求来源的客户端可缓存的该CORS请求的响应时间为100秒。

----结束

## 第二步：使用表单上传

以BrowserJS为例，演示如何直接使用SDK计算签名。

基于表单上传是使用HTML表单形式上传对象到指定桶中，对象最大不能超过5GB。

您可以通过ObsClient.createPostSignatureSync生成基于表单上传的请求参数。使用BrowserJS代码模拟表单上传的完整代码示例，可单击此处下载：[post-object-sample](#)。您也可以通过如下步骤进行表单上传：

### 步骤1 使用ObsClient.createPostSignatureSync生成用于鉴权的请求参数。

使用SDK生成用于鉴权的请求参数包括两个：

- Policy：对应表单中policy字段。
- Signature：对应表单中的signature字段。

代码示例如下：

```
// 创建ObsClient实例
var obsClient = new ObsClient({
  // 认证用的ak和sk硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密文存放，使用时解密，确保安全；本示例以ak和sk保存在环境变量中为例，运行本示例前请先在本地环境中设置环境变量AccessKeyId和SecretAccessKey。
  // 您可以登录访问管理控制台获取访问密钥AK/SK，获取方式请参见https://support.huaweicloud.com/intl/zh-cn/usermanual-ca/ca_01_0003.html
  access_key_id: process.env.AccessKeyId,
  secret_access_key: process.env.SecretAccessKey,
  server: 'https://your-endpoint',
  signature: 'obs'
});

// 设置表单参数
var formParams = {
  // 设置对象访问权限为公共读
  'x-obs-acl': obsClient.enums.AclPublicRead,
  // 设置对象MIME类型
  'content-type': 'text/plain'
};

// 设置表单上传请求有效期，单位：秒
var expires = 3600;

var res = obsClient.createPostSignatureSync({Expires:expires, FormParams: formParams});

// 获取表单上传请求参数
console.log('\t' + res.Policy);
console.log('\t' + res.Signature);
```

### 步骤2 准备表单HTML页面。

表单HTML代码示例如下：

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
</head>
```

```
<body>
<form action="http://bucketname.your-endpoint/" method="post" enctype="multipart/form-data">
Object key
<!-- 对象名 -->
<input type="text" name="key" value="objectname" />
<p>
ACL
<!-- 对象ACL权限 -->
<input type="text" name="x-obs-acl" value="public-read" />
<p>
Content-Type
<!-- 对象MIME类型 -->
<input type="text" name="content-type" value="text/plain" />
<p>
<!-- policy的base64编码值 -->
<input type="hidden" name="policy" value="**** Provide your policy ****" />
<!-- AK -->
<input type="hidden" name="AccessKeyId" value="**** Provide your access key ****"/>
<!-- 签名串信息 -->
<input type="hidden" name="signature" value="**** Provide your signature ****"/>

<input name="file" type="file" />
<input name="submit" value="Upload" type="submit" />
</form>
</body>
</html>
```

### 📖 说明

- HTML表单中的policy, signature的值均是从ObsClient.createPostSignatureSync的返回结果中获取。
- 表单HTML示例可单击此处下载：[PostDemo](#)。

**步骤3** 将生成的请求参数填入HTML页面。

**步骤4** 选择本地文件，进行表单上传。

----结束

## 知识扩展

采用BrowserJS SDK直接计算签名时，AK/SK可能会展现在前端界面，有一定风险。

您还可以采用客户端-服务端模型，服务端可以采用Java、Python等SDK计算POST上传签名，客户端采用JavaScript向服务端获取签名信息后利用签名信息访问OBS。

其中，计算POST上传签名信息请参考各SDK语言：

- [Java](#)
- [Python](#)
- [PHP](#)
- [BrowserJS](#)
- [Node.js](#)

除POST上传外，在其他场景中，为了避免前端代码直接使用AK/SK访问OBS造成敏感信息泄露，可以通过后台计算临时URL，前端使用临时URL授权访问OBS。

利用GO SDK计算临时URL，前端JS使用临时URL列举OBS桶内对象。示例如下：

1. GO SDK后台计算列举桶临时URL。  
// 引入依赖包  
import (

```
"fmt"  
"obs"  
"strings"  
)  
  
//推荐通过环境变量获取AKSK，这里也可以使用其他外部引入方式传入，如果使用硬编码可能会存在泄露  
//风险。  
//您可以登录访问管理控制台获取访问密钥AK/SK，获取方式请参见https://support.huaweicloud.com/intl/zh-cn/usermanual-ca/ca\_01\_0003.html。  
var ak = os.Getenv("AccessKeyID")  
var sk = os.Getenv("SecretAccessKey")  
var endpoint = "https://your-endpoint"  
  
// 创建ObsClient结构体  
var obsClient, _ = obs.New(ak, sk, endpoint)  
  
func main() {  
    input := &obs.CreateSignedUrlInput{}  
    input.Expires = 3600  
  
    // 生成列举对象临时URL  
    // 指定为GET请求，传入桶名  
    input.Method = obs.HttpMethodGet  
    input.Bucket = "bucketname"  
    output, _ := obsClient.CreateSignedUrl(input)  
    // 获取生成的临时URL及请求头域信息  
    fmt.Printf("SignedUrl:%s\n", output.SignedUrl)  
    fmt.Printf("ActualSignedRequestHeaders:%v\n", output.ActualSignedRequestHeaders)  
}
```

## 2. 前台获取到签名URL SignedUrl及请求头域信息ActualSignedRequestHeaders后，访问OBS进行列举桶操作。

```
// 使用GET请求获取对象列表  
var bucketName = 'bucketname';  
var method = 'GET';  
  
// SignedUrl为上一步骤中后端服务计算得到的临时URL，  
// ActualSignedRequestHeaders为上一步骤中后端服务计算临时URL时使用的请求头域，前台实际请求应  
// 保持一致；  
var reopt = {  
    method : method,  
    url : SignedUrl,  
    withCredentials: false,  
    headers : ActualSignedRequestHeaders || {},  
    validateStatus: function(status){  
        return status >= 200;  
    },  
    maxRedirects : 0,  
    responseType : 'text',  
};  
  
axios.request(reopt).then(function (response) {  
    if(response.status < 300){  
        console.log('Listing object using temporary signature succeed.');    }else{  
        console.log('Listing object using temporary signature failed!');  
        console.log('status:' + response.status);  
        console.log('\n');  
    }  
    console.log(response.data);  
    console.log('\n');  
}).catch(function (err) {  
    console.log('Listing object using temporary signature failed!');  
    console.log(err);  
    console.log('\n');  
});
```

其中，生成临时授权访问URL请参考各SDK语言：



- [Java](#)
- [Python](#)
- [Go](#)
- [Node.js](#)

# 12 移动应用直传

## 12.1 方案概述

在互联网中，使用OBS作为存储在移动APP（手机Android、iOS应用）中获得了越来越广泛的应用。Android和iOS应用使用OBS服务时，不能直接存储**访问密钥（AK/SK）**，这样可能会导致访问密钥（AK/SK）被黑客软件破解获取，进而可能导致存放在云存储中的文件数据被窃取，甚至被篡改。

为了更好地保护应用数据，避免被攻击后数据泄露以及越权访问的风险，为您推荐以下两种方法。

- **方法一：使用临时安全凭证直传OBS**
- **方法二：使用预签名URL访问OBS**

方法一使用临时的AK/SK，可以避免AK/SK泄露的风险。推荐您优先使用临时安全凭证直传OBS。

## 12.2 使用临时安全凭证直传 OBS

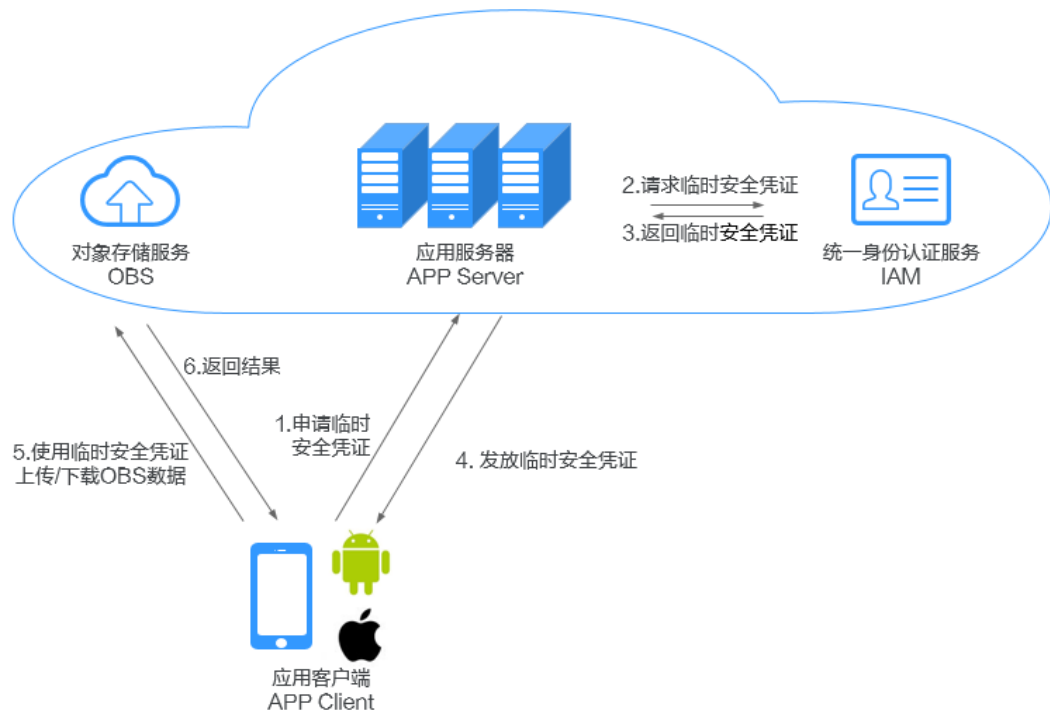
### 方案架构

使用的方式，可以将应用客户端的数据直传至OBS，或者将存储在OBS里的数据进行下载。具体流程如**图12-1**。

OBS支持使用临时安全凭证（临时AK/SK和securitytoken）进行授权访问，同时支持为临时安全凭证配置权限策略来指定使用该临时安全凭证时允许执行的操作。**什么是临时安全凭证？**

移动应用客户端可以使用指定了权限策略的临时安全凭证来访问OBS，实现数据直传，整个过程不会暴露用户的永久AK/SK，降低了账号泄露带来的安全风险。

图 12-1 使用临时安全凭证直传 OBS 流程



### 角色分析如下：

- 应用客户端：即最终用户手机上的APP，负责向应用服务器发出申请临时安全凭证的请求，以及访问OBS完成数据上传或下载。
- 应用服务器：即提供该Android/iOS应用的开发者开发的APP后台服务，用于用户管理和授权管理等。
- 对象存储服务：即华为云对象存储服务，负责处理移动应用的数据请求。
- 统一身份认证服务：即华为云统一身份认证服务，负责生成临时安全凭证。

### 实现流程如下：

1. 应用客户端向应用服务器申请一个临时操作凭证。
2. 应用服务器向统一身份认证服务请求临时安全凭证。
3. 统一身份认证服务向应用服务器返回临时安全凭证。
4. 应用服务器将临时安全凭证发放给应用客户端。
5. 应用客户端使用安全凭证完成OBS数据上传下载。

## 前提条件

已创建桶，并将桶权限设置为私有读写或者公共读私有写。

详细操作步骤请参见[创建桶](#)和[配置桶策略](#)。

## 资源和成本规划

最佳实践中涉及的资源如下：

表 12-1 资源说明

资源	资源说明
应用客户端（APP Client）	最终用户手机上的APP，负责向应用服务器发出申请临时安全凭证的请求，以及访问OBS完成数据上传或下载。
应用服务器（APP Server）	提供该Android/iOS应用的开发者开发的APP后台服务，用于用户管理和授权管理等。
对象存储服务（OBS）	华为云对象存储服务，负责处理移动应用的数据请求。
统一身份认证服务（IAM）	华为云统一身份认证服务，负责生成临时安全凭证。

## 实施步骤

**步骤1** 获取OBS SDK开发包和IAM SDK开发包。

OBS SDK请在[SDK开发指南](#)中获取。

IAM SDK开发包请在[IAM开发工具包](#)获取。

**步骤2** 模拟应用服务器向IAM请求临时安全凭证和返回安全凭证。

过程如下：

1. 获取用户的IAM用户Token。

API请参见[获取IAM用户Token（使用密码）](#)，SDK请参见[开发工具包（SDK）](#)。

2. 使用Token获取临时安全凭证（临时AK/SK和securitytoken），获取时需要通过Policy字段指定该安全凭证允许执行的操作权限。

API请参见[通过token获取临时访问密钥和securitytoken](#)，SDK请参见[开发工具包（SDK）](#)。

示例：获取一个有效期为900秒的临时安全凭证，该凭证只允许上传数据到桶hi-company的APPClient/APP-1/目录下。

```
{
  "auth":{
    "identity":{
      "policy":{
        "Version":"1.1",
        "Statement":[
          {
            "Action":[
              "obs:object:PutObject"
            ],
            "Resource":[
              "obs:*:*:object:hi-company/APPClient/APP-1/*"
            ],
            "Effect":"Allow"
          }
        ]
      }
    },
    "token":{
      "duration-seconds":900,
      "id":"MIIDkgYJKoZIhvcNAQcCollDgzCCA38CAQExDTALMEXXXXX..."
    },
    "methods":[
```

```
    "token"  
  ]  
}  
}
```

### 步骤3 初始化应用客户端中OBS client。

初始化示例：

- Android

```
String endPoint = "https://your-endpoint";  
// 认证用的ak和sk硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密  
文存放，使用时解密，确保安全；本示例以ak和sk保存在环境变量中为例，运行本示例前请先在本地环境  
中设置环境变量ACCESS_KEY_ID和SECRET_ACCESS_KEY_ID。  
// 您可以登录访问管理控制台获取访问密钥AK/SK，获取方式请参见https://support.huaweicloud.com/  
intl/zh-cn/usermanual-ca/ca_01_0003.html  
String ak = System.getenv("ACCESS_KEY_ID");  
String sk = System.getenv("SECRET_ACCESS_KEY_ID");  
String token = System.getenv("Security-Token");  
  
// 创建ObsConfiguration配置类实例  
ObsConfiguration config = new ObsConfiguration();  
config.setEndPoint(endPoint);  
config.setTimeout(30000);  
config.setConnectionTimeout(10000);  
  
// 创建ObsClient实例  
ObsClient obsClient = new ObsClient(ak, sk, token, config);  
  
// 使用访问OBS  
  
// 关闭obsClient  
obsClient.close();
```

#### 📖 说明

- endPoint即终端节点，可通过[地区和终端节点查询](#)。
- ak和sk即临时AK/SK，token即securitytoken，获取方式请参见[访问密钥（AK/SK）](#)。

- iOS

```
NSString *endPoint = @"your-endpoint";  
// 认证用的ak和sk硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密  
文存放，使用时解密，确保安全；本示例以ak和sk保存在环境变量中为例，运行本示例前请先在本地环境  
中设置环境变量AccessKeyID和SecretAccessKey。  
// 您可以登录访问管理控制台获取访问密钥AK/SK，获取方式请参见https://support.huaweicloud.com/  
intl/zh-cn/usermanual-ca/ca_01_0003.html  
NSString *SK = getenv("AccessKeyID");  
NSString *AK = getenv("SecretAccessKey");  
// 初始化身份验证  
OBSStaticCredentialProvider *credentialProvider = [[OBSStaticCredentialProvider alloc]  
initWithAccessKey:AK secretKey:SK];  
securityTokenCredentialProvider.securityToken = @"*** Provide your Security Token ***";  
// 初始化服务配置  
OBSServiceConfiguration *conf = [[OBSServiceConfiguration alloc] initWithURLString:endPoint  
credentialProvider:credentialProvider];  
// 初始化  
clientObsClient *client = [[ObsClient alloc] initWithConfiguration:conf];
```

#### 📖 说明

- endPoint即终端节点，可通过[地区和终端节点查询](#)。
- ak和sk即临时AK/SK，token即securitytoken，获取方式请参见[访问密钥（AK/SK）](#)。

- web js

```
// 未引入AMD，直接通过构造函数创建ObsClient实例  
var obsClient = new ObsClient({  
  // 认证用的ak和sk硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量
```

```
中密文存放，使用时解密，确保安全；本示例以ak和sk保存在环境变量中为例，运行本示例前请先在本地环境中设置环境变量AccessKeyId和SecretAccessKey。  
// 您可以登录访问管理控制台获取访问密钥AK/SK，获取方式请参见https://support.huaweicloud.com/intl/zh-cn/usermanual-ca/ca\_01\_0003.html  
access_key_id: process.env.AccessKeyId,  
secret_access_key: process.env.SecretAccessKey,  
security_token: process.env.SecurityToken,  
server : 'https://your-endpoint'  
});  
// 使用访问OBS  
  
// 引入AMD，通过依赖注入的构造函数创建ObsClient实例  
var obsClient;  
define(['ObsClient'], function(ObsClient){  
  obsClient = new ObsClient({  
    // 认证用的ak和sk硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密文存放，使用时解密，确保安全；本示例以ak和sk保存在环境变量中为例，运行本示例前请先在本地环境中设置环境变量AccessKeyId和SecretAccessKey。  
    // 您可以登录访问管理控制台获取访问密钥AK/SK，获取方式请参见https://support.huaweicloud.com/intl/zh-cn/usermanual-ca/ca\_01\_0003.html  
    access_key_id: process.env.AccessKeyId,  
    secret_access_key: process.env.SecretAccessKey,  
    security_token: process.env.SecurityToken,  
    server : 'https://your-endpoint'  
  });  
  // 使用访问OBS  
});
```

#### 📖 说明

- endPoint即终端节点，可通过[地区和终端节点](#)查询。
- ak和sk即临时AK/SK，token即securitytoken，获取方式请参见[访问密钥（AK/SK）](#)。

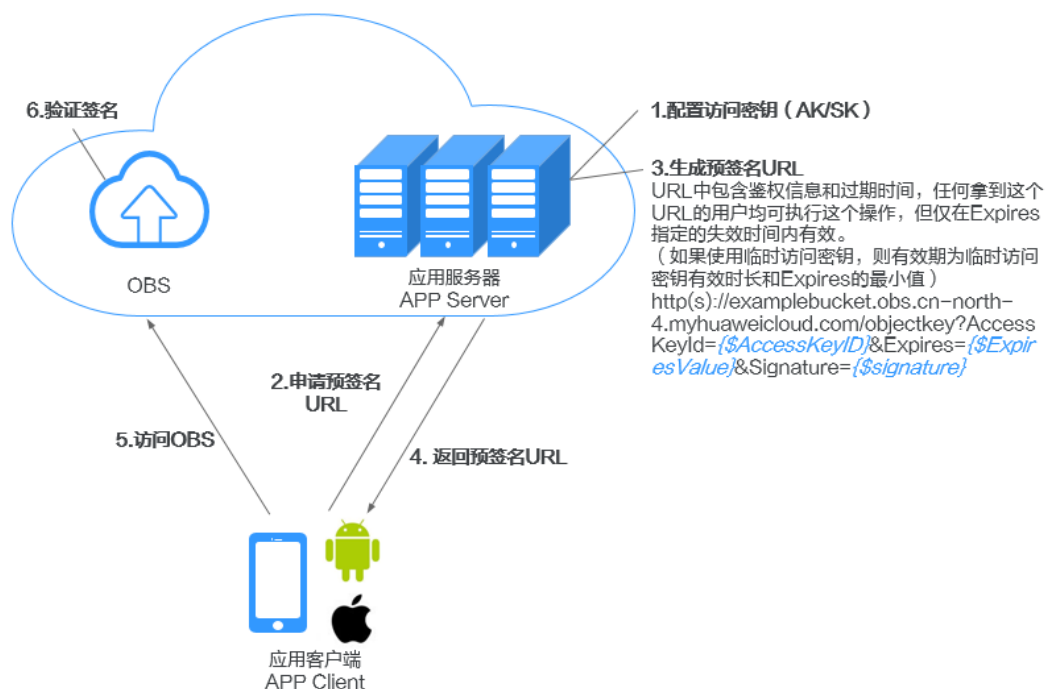
----结束

## 12.3 使用预签名 URL 直传 OBS

### 方案架构

应用客户端每个请求都将向应用服务器申请预签名URL，该预签名URL有效期由应用服务器管理。具体流程如[图12-2](#)。

图 12-2 移动应用访问 OBS 数据流程



#### 角色分析如下：

- 应用客户端：即最终用户手机上的APP，负责向应用服务器申请包含预签名的URL，以及访问OBS完成数据上传或下载。
- 应用服务器：即提供该Android/iOS应用的开发者开发的APP后台服务，用于管理凭证信息以及发放预签名URL。
- OBS：即华为云对象存储，负责处理移动应用的数据请求。

#### 实现流程如下：

1. 移动应用客户端向应用服务器申请一个预签名的URL。

Android和iOS应用使用OBS服务时，不需要存储访问密钥（AK/SK）。应用在上传前必须向用户的应用服务器申请访问OBS的URL，并携带必须信息，包括请求类型、资源路径和资源名称。比如上传操作需要标识该URL为上传请求，需要包含上传的路径以及上传对象的名称；下载操作需要标识该URL为下载请求，需要包含所下载对象的名称。

2. 应用服务器作为可信设备，在应用服务器上存储访问密钥（AK/SK）。应用服务器在验证客户端身份合法之后，使用应用服务器保存的访问密钥（AK/SK）以及客户端访问的资源、操作类型生成预签名URL。举例：

```
https://examplebucket.obs.ap-southeast-1.myhuaweicloud.com/objectkey?  
AccessKeyID=AccessKey/D&Expires=1532779451&Signature=0Akylf43Bm3mD1bh2rM3dmVp1Bo%3D
```

3. Android/iOS移动应用获取此URL，直接使用该URL操作数据，比如上传或者下载操作。

URL中会包含用户的AK、签名、有效期、资源等信息，任何拿到这个URL的人均可执行这个操作。OBS服务收到这个请求并验证签名后，认为该请求就是签发URL的用户自己在执行操作。例如构造一个携带签名信息的下载对象URL，拿到相应URL的人能下载这个对象，但该URL只在Expires指定的失效时间内有效（如果使用临时访问密钥，有效期为临时访问密钥有效时长和Expires的最小值）。URL中携带签名主要用于在不提供给其他人SK的情况下，让其他人能用预签发的URL来进行身份认证，并执行预定义的操作。

## 前提条件

- 创建桶。  
在OBS控制台上创建桶。需要将桶权限设置为私有读写或者公共读私有写。  
详细操作步骤请参见[创建桶](#)和[配置桶策略](#)。
- 获取访问密钥（AK/SK）。  
预签名URL需要通过访问密钥生成，请参考[访问密钥（AK/SK）](#)获取。其中访问密钥（AK/SK）对应的用户需设置所需的最小权限，具体权限设置方法参考[向IAM用户授予OBS资源权限](#)。

## 资源和成本规划

最佳实践中涉及的资源如下：

表 12-2 资源说明

资源	资源说明
应用客户端（APP Client）	最终用户手机上的APP，负责向应用服务器申请包含预签名的URL，以及访问OBS完成数据上传或下载。
应用服务器（APP Server）	提供该Android/iOS应用的开发者开发的APP后台服务，用于管理凭证信息以及发放预签名URL。
对象存储服务（OBS）	华为云对象存储服务，负责处理移动应用的数据请求。

## 实施步骤

### 步骤1 配置应用服务器。

1. 获取SDK开发包。  
请在Java [SDK开发指南](#)中获取。
2. 生成预签名URL的代码。  
预签名URL的计算方法请参考[URL中携带签名](#)。  
下述示例以在应用服务器中使用Java语言开发进行举例。

#### 说明

应用服务器需要根据APP操作类型，识别公共请求消息头与自定义请求消息头，并将其加入到预签名URL生成签名计算中。

- 公共请求消息头，请参考[构造请求](#)。
- 自定义请求消息头，请参考对应操作的API文档。例如PUT上传，参考[PUT上传API](#)。

```
// 本次请求的桶的endpoint  
String endPoint = "http://your-endpoint";
```

```
// 认证用的ak和sk硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密文存放，使用时解密，确保安全；本示例以ak和sk保存在环境变量中为例，运行本示例前请先在本地环境中设置环境变量ACCESS_KEY_ID和SECRET_ACCESS_KEY_ID。
```

```
// 您可以登录访问管理控制台获取访问密钥AK/SK，获取方式请参见https://support.huaweicloud.com/intl/zh-cn/usermanual-ca/ca_01_0003.html  
String ak = System.getenv("ACCESS_KEY_ID");  
String sk = System.getenv("SECRET_ACCESS_KEY_ID");
```



```
// 创建ObsClient实例
ObsClient obsClient = new ObsClient(ak, sk, endPoint);
// 替换您的过期时间, 单位是秒
long expireSeconds = 3600L;

// 替换成您对应的操作
TemporarySignatureRequest request = new TemporarySignatureRequest(HttpMethodEnum.PUT,
expireSeconds);

// 替换为请求本次操作访问的桶名和对象名
request.setBucketName("bucketname");
request.setObjectKey("objectname");

TemporarySignatureResponse response = obsClient.createTemporarySignature(request);

// 成功返回预签名URL, 如下打印URL信息
System.out.println(response.getSignedUrl());
```

更多相关介绍和示例代码, 请参见[使用URL进行授权访问](#)。

## 步骤2 移动应用客户端使用获取到的预签名URL发送OBS请求。

```
public class Demo extends Activity
{
    private static String bucketName = "my-obs-bucket-demo";
    private static String objectKey = "my-obs-object-key-demo";
    private static OkHttpClient httpClient;
    private static StringBuffer sb;

    @Override
    protected void onCreate(Bundle savedInstanceState)
    {
        super.onCreate(savedInstanceState);
        setContentView(R.layout.activity_main);
        sb = new StringBuffer();
        /*
        * Constructs a client instance with your account for accessing OBS
        */
        httpClient = new OkHttpClient.Builder().followRedirects(false).retryOnConnectionFailure(false)
            .cache(null).build();
        final TextView tv = (TextView)findViewById(R.id.tv);
        tv.setText("Click to start test");
        tv.setOnClickListener(new View.OnClickListener()
        {
            @Override
            public void onClick(View v)
            {
                tv.setClickable(false);
                AsyncTask<Void, Void, String> task = new DownloadTask();
                task.execute();
            }
        });
    }

    class DownloadTask extends AsyncTask<Void, Void, String>
    {
        @Override
        protected String doInBackground(Void... params)
        {
            try
            {
                /*
                * 这里需要您自己构造上传对象请求到应用服务器来生成到OBS请求的预签名URL
                * 假如响应结果存放在: response, 通过方法获取getSignedUrl()
                */
                sb.append("Uploading a new object to OBS from a file\n\n");
                Request.Builder builder = new Request.Builder();
                // 使用PUT请求上传对象
                Request httpRequest =
                builder.url(response.getSignedUrl()).put(RequestBody.create(MediaType.parse(contentType), "Hello
```

```
OBS".getBytes("UTF-8"))).build();
    Call c = httpClient.newCall(httpRequest);
    Response res = c.execute();
    sb.append("\tStatus:" + res.code());
    if (res.body() != null) {
        sb.append("\tContent:" + res.body().string() + "\n");
    }
    res.close();

    /*
     * 这里需要您自己构造下载对象请求到应用服务器来生成到OBS请求的预签名URL
     * 假如响应结果存放在: response, 通过方法获取getSignedUrl()
     */
    sb.append("Downloading an object\n\n");
    Request.Builder builder = new Request.Builder();
    // 使用GET请求下载对象
    Request httpRequest = builder.url(response.getSignedUrl()).get().build();
    OkHttpClient httpClient = new
OkHttpClient.Builder().followRedirects(false).retryOnConnectionFailure(false).cache(null).build();
    Call c = httpClient.newCall(httpRequest);
    Response res = c.execute();
    System.out.println("\tStatus:" + res.code());
    if (res.body() != null) {
        sb.append("\tContent:" + res.body().string() + "\n");
    }
    res.close();

    return sb.toString();
}
catch (Exception e)
{
    sb.append("\n\n");
    sb.append(e.getMessage());
    return sb.toString();
}
finally
{
    if (httpClient != null)
    {
        try
        {
            /*
             * Close obs client
             */
            httpClient.close();
        }
        catch (IOException e)
        {
        }
    }
}
}

@Override
protected void onPostExecute(String result)
{
    TextView tv = (TextView)findViewById(R.id.tv);
    tv.setText(result);
    tv.setOnClickListener(null);
    tv.setMovementMethod(ScrollingMovementMethod.getInstance());
}
}
```

----结束

# 13 小程序直传 OBS

## 背景信息

微信小程序作为当下流行的移动应用，具有广泛的应用场景。如何通过微信小程序上传文件至对象存储服务OBS成为了一个热点问题，本文将通过一个示例程序进行演示。

## 注意事项

- 客户端计算签名时依赖引用“crypto-js”及“js-base64”两个开源组件，因此需要在微信小程序项目中设置使用NPM模块。
- 在微信小程序中进行编译时，如果在引入“crypto-js”包时出现“Maximum call stack size exceed”报错，请升级微信小程序开发客户端至最新版本。
- 上传过程中返回405时，请检查指定的endpoint是否为对应上传桶的桶域名。

## 操作步骤

### 步骤1 设置桶的跨域访问权限。

微信小程序基于BrowerJS进行开发，受同源安全策略的要求，不同域间的网站脚本和内容如需交互，需要配置跨域资源共享（CORS）规范。华为云对象存储服务OBS支持CORS规范，允许跨域访问OBS中的资源，具体配置步骤请参见[配置跨域资源共享](#)。

CORS规则配置项建议：

表 13-1 CORS 规则

参数	说明	配置建议
允许的来源	必选参数，指定允许的跨域请求的来源，即允许来自该域名下的请求访问该桶。 允许多条匹配规则，以回车换行为间隔。每个匹配规则允许使用最多一个“*”通配符。例如： http://rds.example.com https://*.vbs.example.com	*

参数	说明	配置建议
允许的方法	必选参数，指定允许的跨域请求方法，即桶和对象的几种操作类型。包括：Get、Post、Put、Delete、Head。	全选
允许的头域	可选参数，指定允许的跨域请求的头域。只有匹配上允许的头域中的配置，才被视为是合法的CORS请求。 允许的头域可设置多个，多个头域之间换行隔开，每行最多可填写一个*符号，不支持&、:、<、空格以及中文字符。	*
补充头域	可选参数，指CORS响应中带的补充头域，给客户端提供额外的信息。 补充头域可设置多个，多个头域之间换行隔开，不支持*、&、:、<、空格以及中文字符。	<ul style="list-style-type: none"> <li>● ETag</li> <li>● x-obs-request-id</li> <li>● x-obs-api</li> <li>● Content-Type</li> <li>● Content-Length</li> <li>● Cache-Control</li> <li>● Content-Disposition</li> <li>● Content-Encoding</li> <li>● Content-Language</li> <li>● Expires</li> <li>● x-obs-id-2</li> <li>● x-reserved-indicator</li> <li>● x-obs-version-id</li> <li>● x-obs-copy-source-version-id</li> <li>● x-obs-storage-class</li> <li>● x-obs-delete-marker</li> <li>● x-obs-expiration</li> <li>● x-obs-website-redirect-location</li> <li>● x-obs-restore</li> <li>● x-obs-version</li> <li>● x-obs-object-type</li> <li>● x-obs-next-append-position</li> </ul>
缓存时间	必选参数，请求来源的客户端可以缓存的CORS响应时间，以秒为单位，默认为100秒。	根据实际业务设置。

**步骤2 配置小程序上传域名白名单。**

微信小程序利用白名单机制管理跨域访问，想要实现数据上传，需要在微信小程序平台域名白名单中配置桶的访问域名。

#### 1. 获取桶的访问域名。

在桶列表单击待操作的桶，进入对象页面后单击“概览”。在“域名信息”下查看桶的访问域名。

图 13-1 桶的访问域名

域名信息			
类型	域名	协议	操作
Endpoint(终端节点) <sup>①</sup>	...	HTTPS/HTTP	--
访问域名 <sup>②</sup>	...	HTTPS/HTTP	🔗 绑定用户域名
静态网页托管域名	--	HTTPS/HTTP	🔗 配置

#### 2. 在微信小程序服务器域名配置中指定桶域名为合法域名。

### 步骤3 计算POST上传签名。

POST上传前需要根据上传时自定义使用的policy字段计算相关签名信息，签名计算规则请参见[基于浏览器上传的表单中携带签名](#)，计算签名相关源代码如下：

对policy进行base64编码（GetPolicy.js）：

```
const Base64 = require('js-base64');

function getPolicyEncode(policy) {
  // 传入表单上传的policy字段，对policy进行Base64编码
  const encodedPolicy = Base64.encode(JSON.stringify(policy));
  return encodedPolicy;
}

module.exports = getPolicyEncode;
```

计算签名的源代码（GetSignature.js）：

```
const Crypto = require('crypto-js');
const Base64 = require('js-base64');

function getSignature(policyEncoded, SecretKey){
  // 利用SK对Base64编码后的policy结果进行HMAC-SHA1签名计算
  const bytes = Crypto.HmacSHA1(policyEncoded, SecretKey);
  // 对计算结果进行Base64编码，得到最终的签名信息
  const signature = Crypto.enc.Base64.stringify(bytes);
  return signature;
}

module.exports = getSignature;
```

### 步骤4 使用小程序直传数据至对象存储桶中。

基于[步骤3](#)中得到的编码后的policy字段及signature字段，可以调用小程序中的上传接口，选择本地文件并上传。具体代码示例如下：

配置AK、SK、访问域名等信息的配置文件（Configuration.js）：

- 使用永久访问密钥（AK/SK）  
// 指定OBS服务相关信息：AK，SK，EndPoint  
var Configuration = {  
 // 认证用的ak和sk硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密文存放，使用时解密，确保安全；本示例以ak和sk保存在环境变量中为例，运行本示例前请先在本地环境中设置环境变量AccessKeyId和SecretKey。  
}

```
// 前端本身没有process对象，可以使用webpack类打包工具定义环境变量，就可以在代码中运行了。  
// 您可以登录访问管理控制台获取访问密钥AK/SK，获取方式请参见https://support.huaweicloud.com/intl/zh-cn/usermanual-ca/ca_01_0003.html  
AccessKeyId: process.env.AccessKeyId,  
SecretKey: process.env.SecretAccessKey,  
EndPoint: 'https://your-test-bucket.obs.myhuaweicloud.com', //完整的桶访问域名  
};
```

```
module.exports = Configuration;
```

- 使用临时访问密钥（AK/SK/securitytoken）

获取临时AK/SK和securitytoken的方法，请参见[获取临时AK/SK和securitytoken](#)。

// 指定OBS服务相关信息：AK，SK，SecurityToken，EndPoint

```
var Configuration = {
```

// 认证用的ak和sk硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密文存放，使用时解密，确保安全；本示例以ak和sk保存在环境变量中为例，运行本示例前请先在本地环境中设置环境变量AccessKeyId和SecretKey。

// 前端本身没有process对象，可以使用webpack类打包工具定义环境变量，就可以在代码中运行了。

// 您可以登录访问管理控制台获取访问密钥AK/SK，获取方式请参见https://support.huaweicloud.com/intl/zh-cn/usermanual-ca/ca\_01\_0003.html

```
AccessKeyId: process.env.AccessKeyId,  
SecretKey: process.env.SecretAccessKey,  
SecurityToken: process.env.SecurityToken, //securityToken  
EndPoint: 'https://your-test-bucket.obs.myhuaweicloud.com', //完整的桶访问域名  
};
```

```
module.exports = Configuration;
```

## 📖 说明

配置文件中传入的EndPoint应该为完整的桶访问域名，例如：<https://bucketName.obs.myhuaweicloud.com>，其中bucketName即小程序上传的目标桶名。

```
// 引入配置文件  
const config = require('./Configuration.js');  
// 引入policy编码计算方法  
const getPolicyEncode = require('./getPolicy.js');  
// 引入签名计算方法  
const getSignature = require('./GetSignature.js');  
  
const OBSUpload = function (filePath){  
  if(!filePath){  
    wx.showToast({  
      title: 'Invalid filePath',  
      icon: 'Please re-select path',  
    });  
  }  
  else{  
    const fileName = 'testMiniprogram.jpg'; // 指定上传到OBS桶中的对象名  
  
    const OBSPolicy = { // 设定policy内容，policy规则定义可参考步骤3中的超链接签名计算规则文档  
      "expiration": "2021-12-31T12:00:00.000Z",  
      "conditions": [  
        { "bucket": "your-test-bucket" }, // 桶名要和配置文件中endpoint中的桶名保持一致  
        // { "x-obs-security-token": config.SecurityToken } // 如果是临时访问密钥鉴权，必须设置该值  
        { 'key': fileName }  
      ]  
    }  
  
    const policyEncoded = getPolicyEncode(OBSPolicy); // 计算base64编码后的policy  
    const signature = getSignature(policyEncoded, config.SecretKey); // 计算signature  
  
    wx.uploadFile({  
      url: config.EndPoint,  
      filePath: filePath,
```

```
name: 'file',
header: {
  'content-type': 'multipart/form-data; boundary=-9431149156168',
},
formData: {
  // 从配置文件中获取到的AK信息、计算得到的编码后policy及signature信息
  'AccessKeyId': config.AccessKeyId,
  'policy': policyEncoded,
  'signature': signature,
  'key': fileName,
  // "x-obs-security-token": config.SecurityToken, // 如果是临时访问秘钥鉴权，必须设置该值
},

success: function(res){
  console.log(res.statusCode); //打印响应状态码
  if(res.statusCode=='204'){
    console.log('Uploaded successfully', res)
    wx.showToast({
      title: 'Uploaded successfully',
      icon: 'Success'
    });
  }
  else{
    console.log('Uploaded failed', res)
    wx.showToast({
      title: 'Uploaded failed',
      icon: 'Fail'
    });
  }
},
fail: function(e){
  console.log(e);
}
})

}
}

module.exports = OBSupload;
```

----结束

## 相关操作

上传完成后，要获取对应对象的访问URL，请参见[如何获取对象访问路径](#)。

# 14 通过 Nginx 反向代理访问 OBS

## 应用场景

一般情况下，用户会通过OBS提供的桶访问域名（例如`https://bucketname.obs.ap-southeast-1.myhuaweicloud.com`）或者绑定的自定义域名来访问OBS。

但在某些场景下，用户需要通过固定的IP地址访问OBS，例如：某些企业出于安全考虑，对于可访问的外部地址需要设置黑白名单，而这个时候对于OBS的访问则需要一个固定的IP地址。同样出于安全考虑，华为云OBS桶访问域名通过DNS解析的IP地址是会发生变化的，所以用户无法获取某个桶长期有效的固定IP地址。

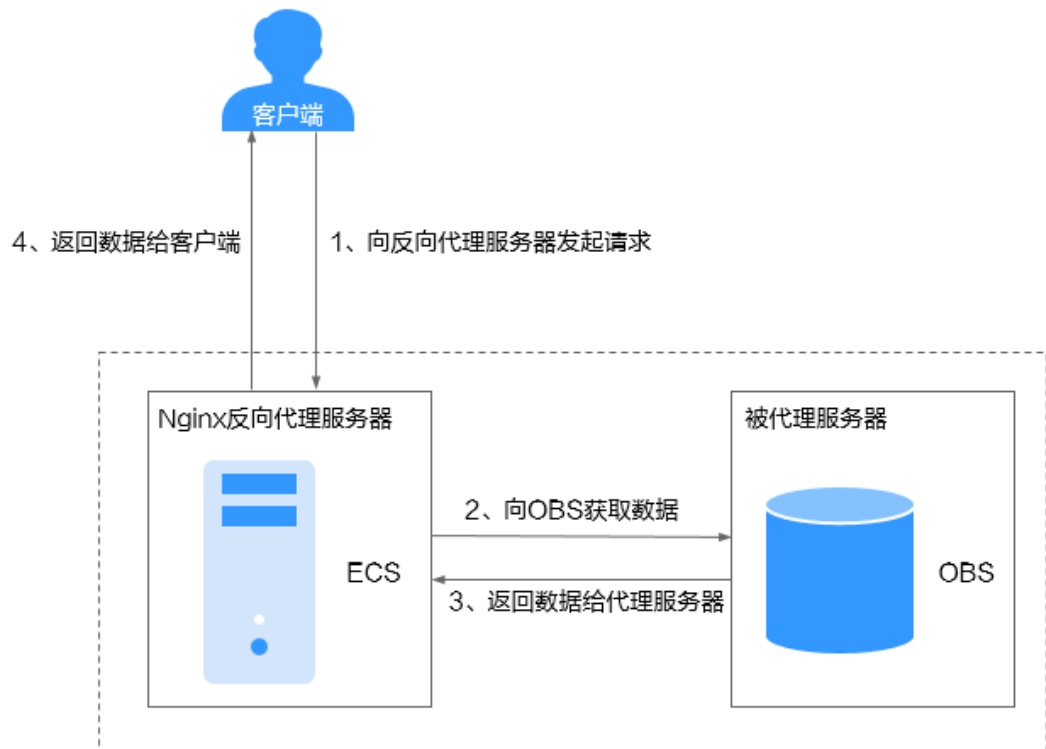
此时，可以通过在ECS上搭建Nginx反向代理服务器，来实现通过固定IP地址访问OBS。

## 方案架构

本实践将Nginx部署在ECS上，搭建Nginx反向代理服务器。用户对代理无感知，只需要将请求发送到反向代理服务器，然后由反向代理服务器向OBS获取数据，再返回给用户。反向代理服务器和OBS对外看做一个整体，仅暴露代理服务器的IP地址，隐藏了OBS真实的域名或IP地址。



图 14-1 通过 Nginx 反向代理访问 OBS 原理



## 约束与限制

- 已明确OBS桶所在区域和桶的访问域名，如中国-香港区域的桶：[nginx-obs.obs.ap-southeast-1.myhuaweicloud.com](#)。[查看方法](#)
- 已在同区域购买Linux操作系统的ECS，本文以CentOS系统为例。[购买ECS方法](#)
- ECS已绑定EIP，EIP用于从公网下载必要的Nginx安装包。

## 实施步骤

### 步骤1 在ECS上安装Nginx

此处以CentOS 7.6版本的操作系统为例。

1. 登录用于搭建Nginx反向代理服务器的ECS。
2. 使用wget命令，下载对应当前操作系统版本的Nginx安装包。  

```
wget http://nginx.org/packages/centos/7/noarch/RPMS/nginx-release-centos-7-0.el7ngx.noarch.rpm
```
3. 执行以下命令，建立Nginx的yum仓库。  

```
rpm -ivh nginx-release-centos-7-0.el7ngx.noarch.rpm
```
4. 执行以下命令，安装Nginx。  

```
yum -y install nginx
```
5. 执行以下命令，启动Nginx并设置开机启动。  

```
systemctl start nginx  
systemctl enable nginx
```
6. 在任意终端使用浏览器访问“[http://ECS弹性公网IP地址](#)”，显示如下图所示，说明Nginx安装成功。

图 14-2 Nginx 安装成功

# Welcome to nginx!

If you see this page, the nginx web server is successfully installed and working. Further configuration is required.

For online documentation and support please refer to [nginx.org](http://nginx.org).  
Commercial support is available at [nginx.com](http://nginx.com).

*Thank you for using nginx.*

## 步骤2 修改Nginx配置文件，反向代理OBS桶

1. 执行以下命令，打开“default.conf”配置文件。

```
vim /etc/nginx/conf.d/default.conf
```

2. 按“i”键进入编辑模式，修改“default.conf”配置文件。

```
server {  
    listen    80;  
    server_name  *.*.*.*, #此处填写ECS弹性公网IP地址  
    proxy_buffering off; #关闭代理缓冲区（内存）  
    location / {  
        proxy_pass https://nginx-obs.obs.ap-southeast-1.myhuaweicloud.com; #此处填写OBS桶访问域  
        #以http://或https://开头  
        index index.html index.htm; #指定网站初始页。如果包括多个文件，Nginx会根据文件的枚举顺  
        #序来检查。  
    }  
}
```

表 14-1 配置文件参数说明

参数	说明
server_name	提供反向代理服务的IP地址，即需要暴露给终端用户访问的固定IP地址。 此处填写搭建Nginx反向代理服务的ECS弹性公网IP地址，即当前登录的ECS弹性公网IP地址。
proxy_pass	被代理服务器的地址。 此处填写前提条件获取的OBS桶的访问域名，注意需要以http://或https://开头，例如： https://nginx-obs.obs.ap-southeast-1.myhuaweicloud.com 注意： 当使用API、SDK、obsutil调用时，此处填写区域域名，例如： obs.ap-southeast-1.myhuaweicloud.com

参数	说明
proxy_buffering	<p>设置是否开启代理缓冲区，取值为on（开启）或者off（关闭）。</p> <p>如果取值为on，Nginx会把后端返回的内容先放到缓冲区，然后再返回给客户端。</p> <p>如果取值为off，Nginx会立即把从后端收到的响应内容传送给客户端。</p> <p>默认值：on</p> <p>示例：proxy_buffering off</p>

- 按“Esc”，输入“:wq”保存并退出。
- 执行以下命令，测试Nginx配置文件状态。  
nginx -t
- 执行以下命令，重启Nginx服务使配置生效。  
systemctl stop nginx  
systemctl start nginx

### 步骤3 (可选)配置OBS桶策略，允许Nginx代理服务器的IP地址访问OBS

如果您的OBS桶为公共读，或者访问私有桶内对象时在[URL中携带签名](#)，则可跳过此步骤。

如果您的OBS桶为私有桶，且不希望使用携带签名的URL访问桶内资源，则建议配置以下桶策略：仅允许Nginx代理服务器的IP地址访问OBS桶。

- 在OBS管理控制台左侧导航栏选择“对象存储”。
- 在桶列表单击待操作的桶，进入对象页面。
- 在左侧导航栏，单击“访问权限控制 > 桶策略”。
- 单击“创建”。
- 根据使用习惯，策略配置方式以可视化视图为例。单击“可视化视图”。
- 配置如下参数。

表 14-2 桶策略参数配置

参数	说明						
策略名称	输入自定义的桶策略名称						
策略内容	<table border="1"> <tr> <td>效力</td> <td>允许</td> </tr> <tr> <td>被授权用户</td> <td>- 被授权用户：所有账号</td> </tr> <tr> <td>授权资源</td> <td>                     - 方式一：                     <ul style="list-style-type: none"> <li>资源范围：整个桶（包括桶内对象）</li> </ul>                     - 方式二：                     <ul style="list-style-type: none"> <li>资源范围：当前桶、指定对象</li> <li>指定对象 - 资源路径：*</li> </ul> </td> </tr> </table>	效力	允许	被授权用户	- 被授权用户：所有账号	授权资源	- 方式一： <ul style="list-style-type: none"> <li>资源范围：整个桶（包括桶内对象）</li> </ul> - 方式二： <ul style="list-style-type: none"> <li>资源范围：当前桶、指定对象</li> <li>指定对象 - 资源路径：*</li> </ul>
效力	允许						
被授权用户	- 被授权用户：所有账号						
授权资源	- 方式一： <ul style="list-style-type: none"> <li>资源范围：整个桶（包括桶内对象）</li> </ul> - 方式二： <ul style="list-style-type: none"> <li>资源范围：当前桶、指定对象</li> <li>指定对象 - 资源路径：*</li> </ul>						

参数		说明
	授权操作	<ul style="list-style-type: none"> <li>- 动作范围：自定义配置</li> <li>- 选择动作：Get*和List*</li> </ul>
	授权条件 (可选)	<ul style="list-style-type: none"> <li>- 键：SourceIp</li> <li>- 条件运算符：IpAddress</li> <li>- 值：                             <ul style="list-style-type: none"> <li>▪ 如果ECS使用公网DNS，取值为： <b><i>ECS的弹性公网IP地址</i></b></li> <li>▪ 如果ECS使用华为云内网DNS，取值为： <b><i>100.64.0.0/10,214.0.0.0/7,ECS的私有IP地址</i></b></li> </ul> </li> </ul> <p><b>说明</b> 取值需要同时配置三个IP地址（IP地址段），请单击“增加”按钮。 其中，100网段和214网段为ECS内网访问OBS的网段。</p>

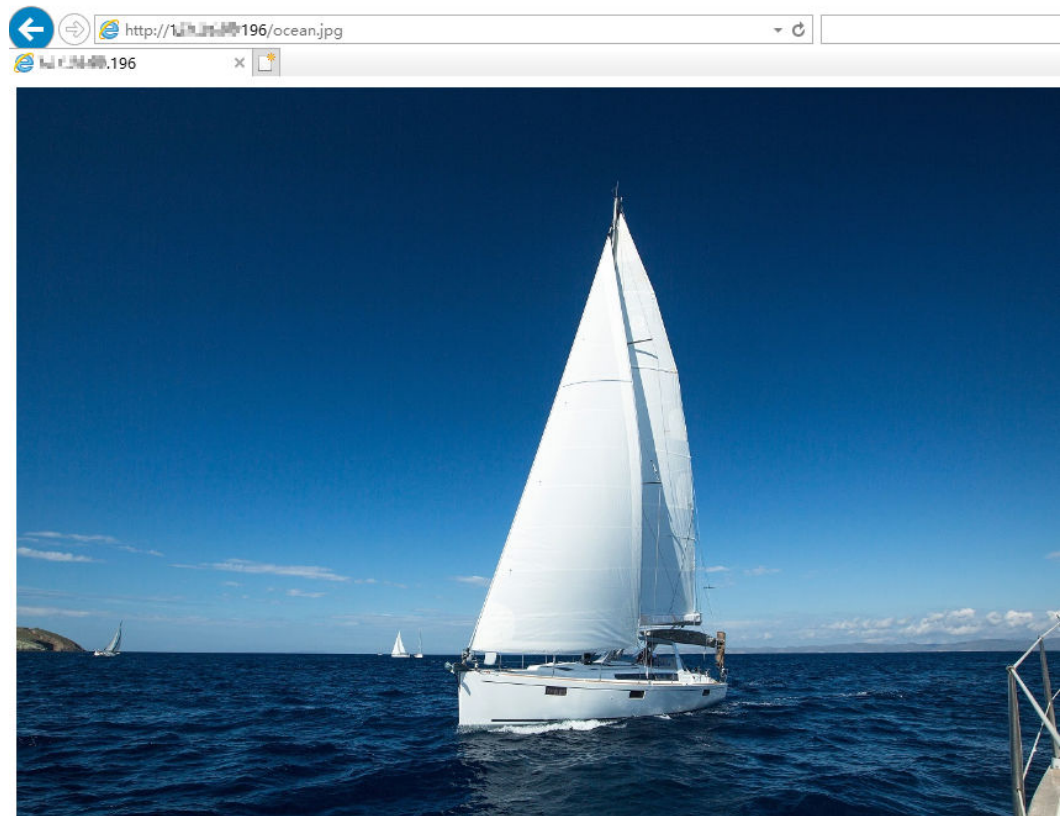
7. 单击右下角的“创建”，完成桶策略创建。

#### 步骤4 验证反向代理配置

在任意终端使用ECS弹性公网IP地址+对象名访问OBS资源，如果能正常访问，则说明配置成功。

例如访问<http://ECS弹性公网IP地址/ocean.jpg>

图 14-3 使用固定 IP 地址访问 OBS 资源



----结束

# 15 大数据场景下使用 OBS 实现存算分离

## 15.1 方案概述

### 应用场景

随着大数据技术的飞速发展，对数据价值的认识逐渐加深，大数据已经融入到了各行各业。根据相关调查报告数据显示，超过39.6%的企业正在应用大数据并从中获益；超过89.6%的企业已经成立或计划成立相关的大数据分析部门；超过六成的企业在扩大大数据的投入力度。对各行业来讲，大数据的使用能力成为未来取得竞争优势的关键能力之一。

在大数据场景下，数据已成为新资产，智能已成为新生产力。企业迫切需要完成数字化转型，提高生产力，使数据资产发挥最大价值。而传统企业在业务未上云之前，业务部署和数据存储往往都在本地IDC机房的多个集群，且一台服务器同时提供计算和存储能力，这种方式导致的如表15-1所示的几个关键问题，已成为企业数字化转型的阻碍。

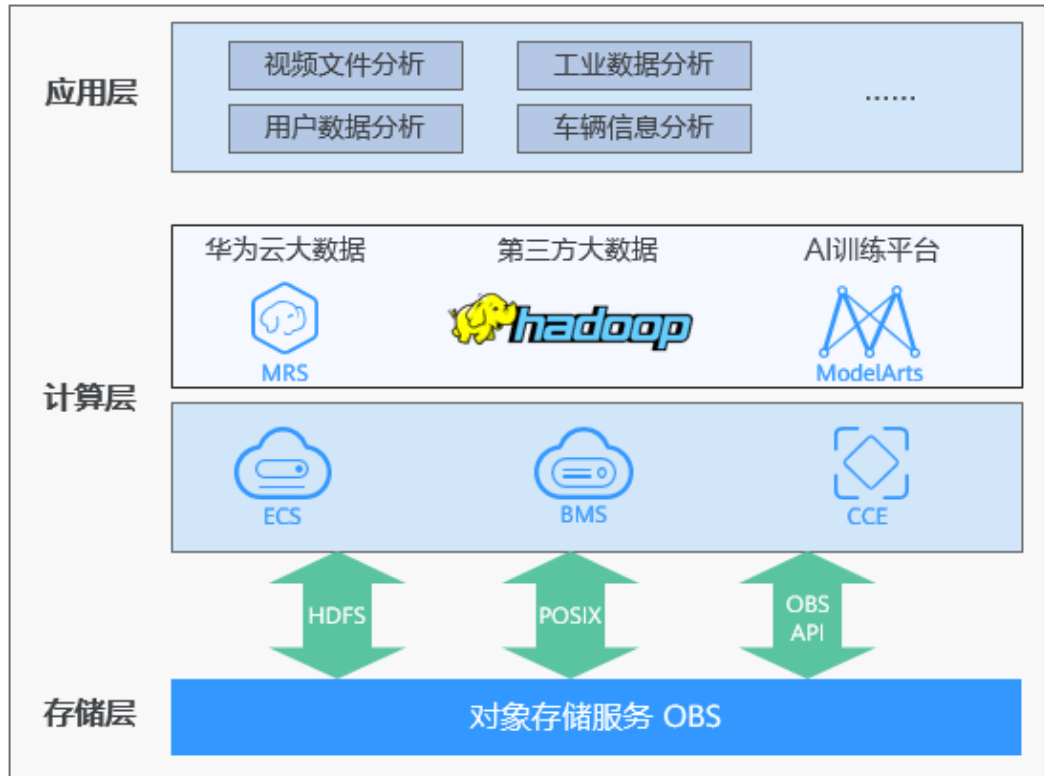
表 15-1 传统企业在大数据场景面临的关键问题

序号	关键问题	详细描述
1	多集群数据共享难	企业数据往往分别存储在IDC多个集群，存在如下问题： <ul style="list-style-type: none"><li>• 无全局视图，数据只能在集群内部使用。</li><li>• 拷贝是跨集群数据共享的唯一途径，数据拷贝耗时长。</li><li>• 公共数据集多份存储，数据冗余。</li></ul>
2	计算存储资源绑定，导致资源浪费	计算和存储资源无法均衡，当计算和存储需求不一致时，只能等比扩容，势必造成一种资源的浪费。
3	数据三副本存储，利用率低，成本高	Hadoop分布式文件系统（HDFS）使用三副本保存数据，磁盘空间利用率仅33%，单盘利用率低于70%。

## 方案架构

针对传统企业在大数据场景面临的问题，华为云提供了基于对象存储服务OBS作为统一数据湖存储的大数据存算分离方案。

图 15-1 基于 OBS 的华为云大数据存算分离方案



华为云大数据存算分离方案基于对象存储服务OBS的大容量高带宽能力，以及多协议共享访问技术（HDFS/POSIX/OBS API），实现Hadoop生态多计算引擎（Hive、Spark等）兼容对接。

## 方案优势

相比传统企业在本地上IDC机房部署大数据业务，华为云数据存算分离方案的主要优势如表15-2。

表 15-2 华为云大数据存算分离相比传统大数据方案的优势

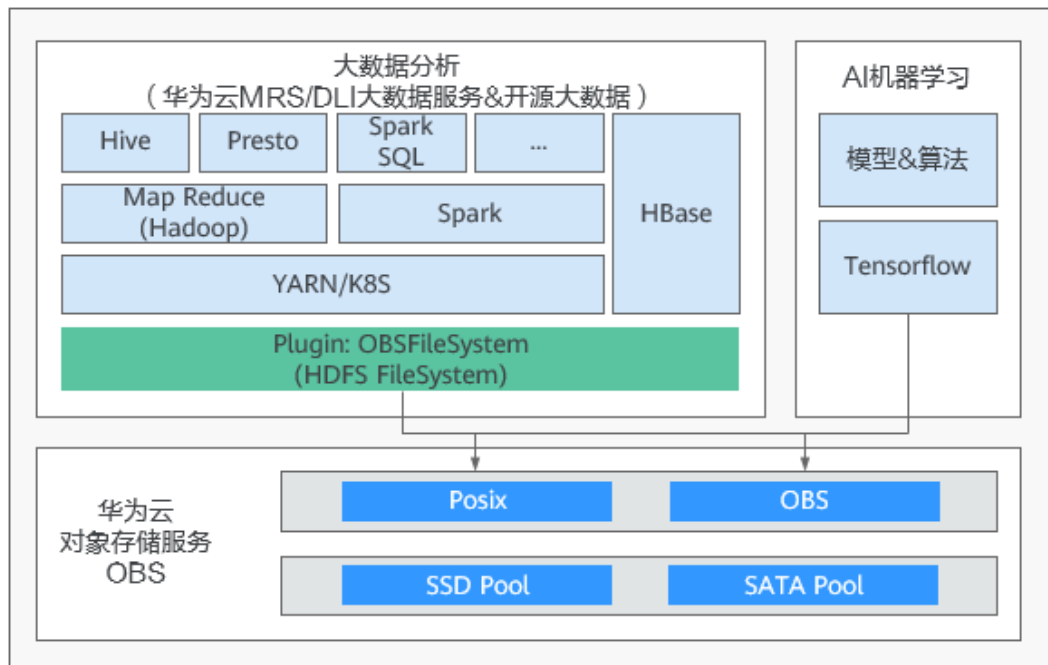
序号	主要优势	详细描述
1	融合高效，协同分析	<ul style="list-style-type: none"> <li>通过统一的权限控制，实现多集群间的数据共享。</li> <li>数据“0”拷贝。</li> <li>大数据和AI一体化，减少作业耗时。</li> </ul>
2	存算分离，资源利用率高	计算存储解耦，支持独立扩容或缩容，计算资源可弹性伸缩，资源利用率提升。

序号	主要优势	详细描述
3	数据EC冗余存储，利用率高，成本低	对象存储服务OBS支持利用率最高的分布式数据容错技术Erasure code，磁盘利用率大幅提升，数据存储空间需求远低于三副本。

此外，对象存储服务OBS提供了OBSFileSystem插件（OBSA-HDFS），可与上层大数据平台无缝对接，实现业务零改造。

OBSFileSystem的主要作用：提供HDFS文件系统的相关接口实现，让大数据计算引擎（Hive、Spark等）可以将OBS作为HDFS协议的底层存储。

图 15-2 存算分离方案中的 OBSFileSystem



### 说明

OBS服务支持对象存储桶（对象语义）和并行文件系统（POSIX文件语义），在大数据场景下建议选择并行文件系统。并行文件系统支持POSIX文件语义，通过OBSFileSystem封装，相较对象语义增加Rename、Append、hflush/hsync接口，实现完善的HDFS语义，为大数据计算提供了更好的性能。

基于上述优势，华为云存算分离大数据方案相比传统大数据方案，在同样的业务规模下所使用的计算资源、存储资源以及服务器数量都会有明显下降，同时资源利用率也能得到显著提升，可帮助企业降低业务综合成本。

## 文档使用范围

本最佳实践主要提供华为云大数据存算分离方案中不同大数据平台和大数据组件与对象存储服务OBS的对接指导，以及HDFS数据迁移至对象存储服务OBS的方案。



## 15.2 操作流程

大数据场景下使用OBS实现存算分离的操作流程如[图15-3](#)所示。

图 15-3 操作流程



1. 配置的核心是完成大数据平台与OBS对接，实现OBS作为大数据的统一数据湖存储。本文档提供三种主流大数据平台的对接指导，详情请参见[支持的大数据平台简介](#)。
2. （可选）OBS除了可以与主流大数据平台对接外，还可以直接与开源的大数据组件对接。当您使用开源的大数据组件时，可参考[支持的大数据组件简介](#)完成与OBS对接。
3. （可选）如果您的数据仍存储在本地HDFS，需要先将数据迁移到华为云OBS中。详情请参见[迁移HDFS数据至OBS](#)。

## 15.3 对接大数据平台

### 15.3.1 支持的大数据平台简介

华为云大数据存算分离方案中，OBS支持与多种大数据平台对接，包括华为云MapReduce服务（MRS）、Cloudera CDH和Hortonworks HDP，满足用户业务的灵活诉求。

#### 华为云 MapReduce 服务（MRS）

华为云MapReduce服务（MRS）是华为云提供的大数据服务，可以在华为云上部署和管理Hadoop系统，一键即可部署Hadoop集群。

MRS提供用户完全可控的一站式企业级大数据集群云服务，完全兼容开源接口，结合华为云计算、存储优势及大数据行业经验，为客户提供高性能、低成本、灵活易用的全栈大数据平台，轻松运行Hadoop、Spark、HBase、Kafka、Storm等大数据组件，并具备在后续根据业务需要进行定制开发的能力，帮助企业快速构建海量数据信息处

理系统，并通过对海量信息数据实时与非实时的分析挖掘，发现全新价值点和企业商机。

MRS与OBS对接的具体操作，请参见[华为云MRS对接OBS](#)。

## Cloudera CDH

CDH是Cloudera基于Apache Hadoop生态系统构建的大数据分析管理平台发行版。

Cloudera CDH与OBS对接的具体操作，请参见[Cloudera CDH对接OBS](#)。

## Hortonworks HDP

HDP是Hortonworks基于Apache Hadoop生态系统开源组件构建的大数据分析管理平台。

Hortonworks HDP与OBS对接的具体操作，请参见[Hortonworks HDP对接OBS](#)

## 15.3.2 华为云 MRS 对接 OBS

### 对接步骤

**步骤1** 配置存算分离集群。

详细操作，请参见[使用委托方式配置存算分离集群](#)。

**步骤2** 使用存算分离集群。

详细操作，请参见[使用存算分离集群](#)。

----结束

## 15.3.3 Cloudera CDH 对接 OBS

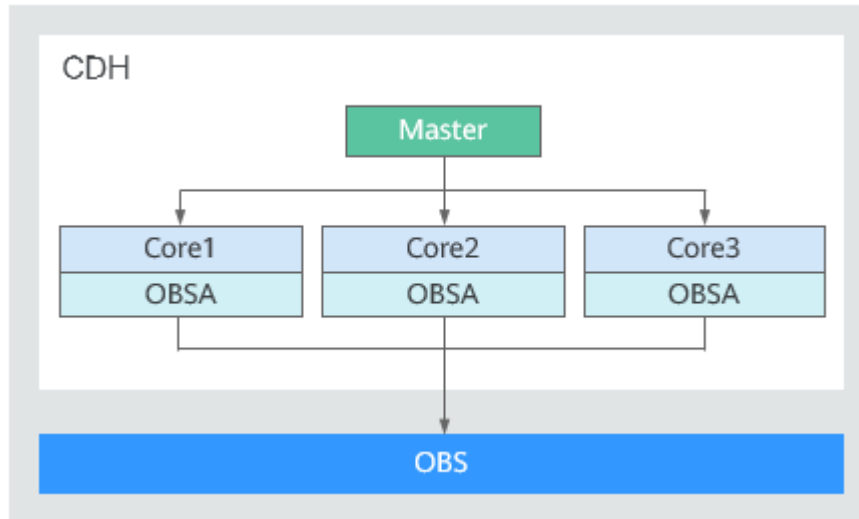
### 部署视图

#### 安装版本

硬件：1Master+3Core（配置：8U32G，操作系统：CentOS 7.5）

软件：CDH 6.0.1

#### 部署视图



## 更新 OBSA-HDFS 工具

**步骤1** 下载与hadoop版本配套的OBSA-HDFS工具：[下载地址](#)。

并将OBSA-HDFS工具jar包（如hadoop-huaweicloud-3.1.1-hw-53.8.jar）上传到CDH各节点/opt/obsa-hdfs目录中。

### 📖 说明

- hadoop-huaweicloud-x.x.x-hw-y.jar包含义：前三位x.x.x为配套hadoop版本号；最后一位y为OBSA版本号，y值最大为最新版本。如：hadoop-huaweicloud-3.1.1-hw-53.8.jar，3.1.1是配套hadoop版本号，53.8是OBSA的版本号。
- 如hadoop版本为3.1.x，则选择hadoop-huaweicloud-3.1.1-hw-53.8.jar。

**步骤2** 增加hadoop-huaweicloud的jar包。

在CDH集群各节点执行以下命令，命令请根据hadoop-huaweicloud的jar包名字及实际CDH版本进行适配使用。

1. 执行如下命令，将OBSA-HDFS工具的jar包放到/opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/目录中。

```
cp /opt/obsa-hdfs/hadoop-huaweicloud-3.1.1-hw-53.8.jar /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/
```

2. 执行如下命令，建立各目录的软连接，将hadoop-huaweicloud的jar包放入如下目录。

```
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud-3.1.1-hw-53.8.jar /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar
```

```
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar /opt/cloudera/cm/cloudera-navigator-server/libs/cdh6/hadoop-huaweicloud.jar
```

```
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar /opt/cloudera/cm/common_jars/hadoop-huaweicloud.jar
```

```
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar /opt/cloudera/cm/lib/cdh6/hadoop-huaweicloud.jar
```

```
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar /opt/cloudera/cm/cloudera-scm-telepub/libs/cdh6/hadoop-huaweicloud.jar
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/lib/hadoop/hadoop-huaweicloud.jar
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/lib/hadoop/client/hadoop-huaweicloud.jar
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/lib/spark/jars/hadoop-huaweicloud.jar
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/lib/impala/lib/hadoop-huaweicloud.jar
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/lib/hadoop-mapreduce/hadoop-huaweicloud.jar
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar /opt/cloudera/cm/lib/cdh5/hadoop-huaweicloud.jar
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar /opt/cloudera/cm/cloudera-scm-telepub/libs/cdh5/hadoop-huaweicloud.jar
ln -s /opt/cloudera/parcels/CDH-6.0.1-1.cdh6.0.1.p0.590678/jars/hadoop-huaweicloud.jar /opt/cloudera/cm/cloudera-navigator-server/libs/cdh5/hadoop-huaweicloud.jar
```

----结束

## HDFS 和 Yarn 集群对接 OBS 配置项

**步骤1** 在HDFS集群配置中选择“高级”，在core-site.xml的群集范围高级配置代码段（安全阀）增加OBS的ak、sk、endpoint和impl配置，对应名称为fs.obs.access.key、fs.obs.secret.key、fs.obs.endpoint、fs.obs.impl。

### 📖 说明

1. 访问密钥AK/SK和终端节点Endpoint请根据实际填写，AK/SK获取方式请参见[访问密钥（AK/SK）](#)，Endpoint获取方式请参见[终端节点（Endpoint）和访问域名](#)。
2. fs.obs.impl配置为org.apache.hadoop.fs.obs.OBSFileSystem。

**步骤2** 修改后“重启”或“滚动重启”HDFS集群，再重启“部署客户端配置”。

**步骤3** 进入YARN集群，重启“部署客户端配置”。

**步骤4** 查看节点中/etc/hadoop/conf/core-site.xml中是否已增加OBS的ak、sk、endpoint和impl配置。

```
<property>
<name>fs.obs.access.key</name>
<value>****</value>
</property>
<property>
```

```
<name>fs.obs.secret.key</name>
<value>*****</value>
</property>
<property>
<name>fs.obs.endpoint</name>
<value>{Target Endpoint}</value>
</property>
<property>
<name>fs.obs.impl</name>
<value>org.apache.hadoop.fs.obs.OBSFileSystem</value>
</property>
```

----结束

## Spark 集群对接 OBS 配置项

**步骤1** Spark应用对接OBS，需要在YARN集群中进行core-site.xml配置，包括：ak、sk、endpoint、impl等。

**步骤2** core-site.xml配置完成后“重启”YARN集群，再重启Spark集群的“部署客户端配置”。

----结束

## Hive 集群对接 OBS 配置项

**步骤1** Hive应用对接OBS，需要在Hive集群中进行core-site.xml配置，包括：ak、sk、endpoint、impl等。

**步骤2** core-site.xml配置完成后“重启”Hive集群，再重启Hive集群的“部署客户端配置”。

----结束

## 15.3.4 Hortonworks HDP 对接 OBS

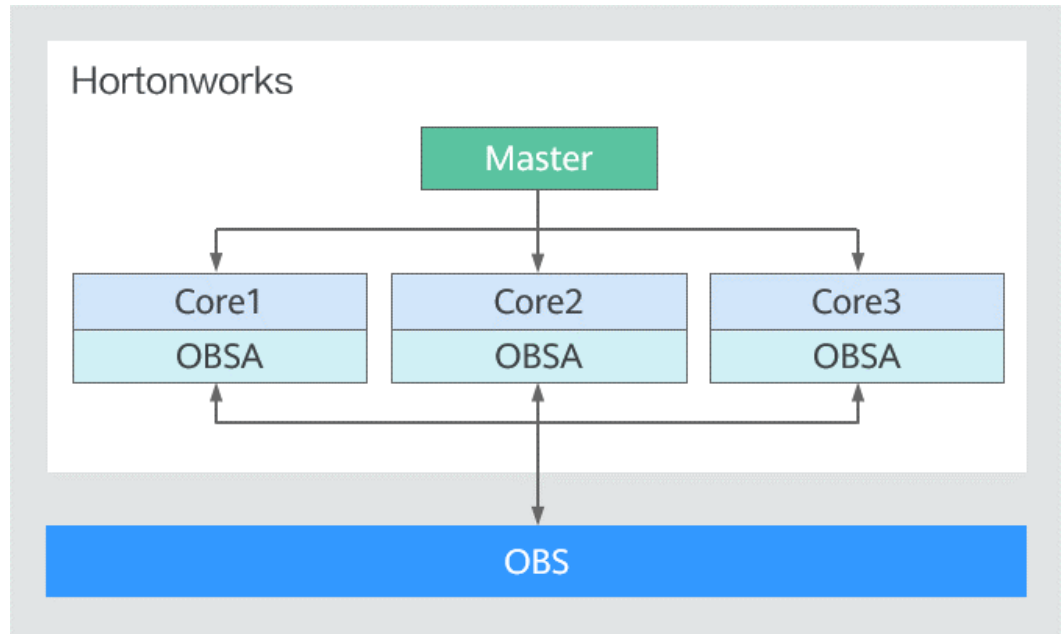
### 部署视图

#### 安装版本

硬件：1master+3core（配置：8U32G，操作系统：CentOS 7.5）

软件：Ambari: 2.7.1.0, HDP: 3.0.1.0

#### 部署视图



## 更新 OBSA-HDFS 工具

**步骤1** 下载与hadoop版本配套的OBSA-HDFS工具：[下载地址](#)。

下载OBSA-HDFS工具的jar包（如hadoop-huaweicloud-3.1.1-hw-53.8.jar）到/mnt/obsjar目录。

### 📖 说明

- hadoop-huaweicloud-x.x.x-hw-y.jar包含义：前三位x.x.x为配套hadoop版本号；最后一位y为OBSA版本号，y值最大为最新版本。如：hadoop-huaweicloud-3.1.1-hw-53.8.jar，3.1.1是配套hadoop版本号，53.8是OBSA的版本号。
- 如hadoop版本为3.1.x，则选择hadoop-huaweicloud-3.1.1-hw-53.8.jar。

**步骤2** 执行以下命令，将OBSA-HDFS工具jar包（如hadoop-huaweicloud-3.1.1-hw-53.8.jar）拷贝到如下目录中。

```
cp /mnt/obsjar/hadoop-huaweicloud-3.1.1-hw-53.8.jar /usr/hdp/share/hst/activity-explorer/lib/
```

```
cp /mnt/obsjar/hadoop-huaweicloud-3.1.1-hw-53.8.jar /usr/hdp/3.0.1.0-187/hadoop-mapreduce/
```

```
cp /mnt/obsjar/hadoop-huaweicloud-3.1.1-hw-53.8.jar /usr/hdp/3.0.1.0-187/spark2/jars/
```

```
cp /mnt/obsjar/hadoop-huaweicloud-3.1.1-hw-53.8.jar /usr/hdp/3.0.1.0-187/tez/lib/
```

```
cp /mnt/obsjar/hadoop-huaweicloud-3.1.1-hw-53.8.jar /var/lib/ambari-server/resources/views/work/CAPACITY-SCHEDULER{1.0.0}/WEB-INF/lib/
```

```
cp /mnt/obsjar/hadoop-huaweicloud-3.1.1-hw-53.8.jar /var/lib/ambari-server/resources/views/work/FILES{1.0.0}/WEB-INF/lib/
```

```
cp /mnt/obsjar/hadoop-huaweicloud-3.1.1-hw-53.8.jar /var/lib/ambari-server/resources/views/work/WORKFLOW_MANAGER{1.0.0}/WEB-INF/lib/
```

```
ln -s /usr/hdp/3.0.1.0-187/hadoop-mapreduce/hadoop-huaweicloud-3.1.1-  
hw-53.8.jar /usr/hdp/3.0.1.0-187/hadoop-mapreduce/hadoop-huaweicloud.jar  
----结束
```

## 在 HDFS 集群中增加配置项

**步骤1** 在HDFS集群CONFIGS的ADVANCED配置项中增加Custom core-site.xml文件中的配置项，包括：fs.obs.access.key，fs.obs.secret.key，fs.obs.endpoint和fs.obs.impl。

### 📖 说明

1. fs.obs.access.key、fs.obs.secret.key、fs.obs.endpoint分别为用户的ak、sk和endpoint。访问密钥AK/SK和终端节点Endpoint请根据实际填写，AK/SK获取方式请参见[访问密钥 \(AK/SK\)](#)，Endpoint获取方式请参见[终端节点 \(Endpoint\)](#) 和[访问域名](#)。
2. fs.obs.impl配置为org.apache.hadoop.fs.obs.OBSFileSystem。

**步骤2** 重启HDFS集群。

----结束

## 在 MapReduce2 集群中增加配置项

**步骤1** 在MapReduce2集群CONFIGS的ADVANCED配置项中修改mapred-site.xml文件中的mapreduce.application.classpath配置项，添加路径为/usr/hdp/3.0.1.0-187/hadoop-mapreduce/\*。

**步骤2** 重启MapReduce2集群。

----结束

## 增加 Hive 对接 OBS 的 jar 包

**步骤1** 执行以下命令，在Hive Server节点创建auxlib文件夹。

```
mkdir /usr/hdp/3.0.1.0-187/hive/auxlib
```

**步骤2** 执行以下命令，将OBSA-HDFS工具的jar包放到auxlib文件夹。

```
cp /mnt/obsjar/hadoop-huaweicloud-3.1.1-hw-53.8.jar /usr/hdp/3.0.1.0-187/  
hive/auxlib
```

**步骤3** 重启Hive集群。

----结束

# 15.4 对接大数据组件

## 15.4.1 支持的大数据组件简介

在华为云大数据存算分离方案中，OBS除了可以与大数据平台对接外，还可以直接与开源的大数据组件对接。

当前支持的大数据组件如下：

- [Hadoop](#)
- [Hive](#)
- [Spark](#)
- [Flume](#)
- [DataX](#)
- [Druid](#)
- [Flink](#)
- [logstash](#)

## 15.4.2 Hadoop 对接 OBS

### 概述

Hadoop系统提供了分布式存储、计算和资源调度引擎，用于大规模数据处理和分析。OBS服务实现了Hadoop的HDFS协议，在大数据场景中可以替代Hadoop系统中的HDFS服务，实现Spark、MapReduce、Hive等大数据生态与OBS服务的对接，为大数据计算提供“数据湖”存储。

#### 说明

HDFS协议：Hadoop中定义了HDFS协议（通过FileSystem抽象类），其他各类存储系统均可以实现HDFS协议，例如Hadoop中内置的HDFS服务，华为云的对象存储服务OBS。

### 约束与限制

不支持以下HDFS语义：

- Lease
- Symbolic link operations
- Proxy users
- File concat
- File checksum
- File replication factor
- Extended Attributes(XAttrs) operations
- Snapshot operations
- Storage policy
- Quota
- POSIX ACL
- Delegation token operations

### 注意事项

为了减少日志输出，在/opt/hadoop-3.1.1/etc/hadoop/log4j.properties文件中增加配置：

```
log4j.logger.com.obs=ERROR
```



## 对接步骤

以Hadoop 3.1.1为例（建议使用最新的版本，不推荐使用低于2.8.3版本的hadoop与hadoop-huaweicloud配套使用）。

**步骤1** 下载hadoop-3.1.1.tar.gz，并解压到/opt/hadoop-3.1.1目录。

**步骤2** 在/etc/profile文件中增加配置内容：

```
export HADOOP_HOME=/opt/hadoop-3.1.1
export PATH=$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$PATH
```

**步骤3** 安装hadoop-huaweicloud。

1. 在官方Github下载hadoop-huaweicloud：[下载地址](#)。

### 📖 说明

如果没有匹配版本的jar包，可自行修改hadoop-huaweicloud目录下pom文件中的hadoop版本重新编译生成。

2. 将hadoop-huaweicloud-x.x.x-hw-y.jar拷贝到/opt/hadoop-3.1.1/share/hadoop/tools/lib和/opt/hadoop-3.1.1/share/hadoop/common/lib目录下。

### 📖 说明

hadoop-huaweicloud-x.x.x-hw-y.jar包含义：前三位x.x.x为配套hadoop版本号；最后一位y为OBSA版本号，y值最大为最新版本。如：hadoop-huaweicloud-3.1.1-hw-40.jar，3.1.1是配套hadoop版本号，40是OBSA的版本号。

**步骤4** 配置hadoop。

修改 /opt/hadoop-3.1.1/etc/hadoop/core-site.xml，增加OBS相关配置信息：

```
<property>
<name>fs.obs.impl</name>
<value>org.apache.hadoop.fs.obs.OBSFileSystem</value>
</property>
<property>
<name>fs.AbstractFileSystem.obs.impl</name>
<value>org.apache.hadoop.fs.obs.OBS</value>
</property>
<property>
<name>fs.obs.access.key</name>
<value>xxx</value>
<description>HuaweiCloud Access Key Id</description>
</property>
<property>
<name>fs.obs.secret.key</name>
<value>xxx</value>
<description>HuaweiCloud Secret Access Key</description>
</property>
<property>
<name>fs.obs.endpoint</name>
<value>xxx</value>
<description>HuaweiCloud Endpoint</description>
</property>
```

**步骤5** 验证是否对接成功。

您可以通过命令行和MR程序两种方式进行验证。示例如下：

- 命令行  
**hadoop fs -ls obs://obs-bucket/**  
回显如下信息：

```
-rw-rw-rw- 1 root root 1087 2018-06-11 07:49 obs://obs-bucket/test1
-rw-rw-rw- 1 root root 1087 2018-06-11 07:49 obs://obs-bucket/test2
```

- MR程序

```
hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-
examples-3.1.1.jar wordcount obs://example-bucket/input/test.txt obs://
obs-bucket/output
```

----结束

## 【附】hadoop-huaweicloud 相关配置

配置项	默认值	是否必填	说明
fs.obs.impl	org.apache.hadoop.fs.obs.OBSFileSystem	是	-
fs.AbstractFileSystem.obs.impl	org.apache.hadoop.fs.obs.OBS	是	-
fs.obs.endpoint	无	是	华为云OBS的终端节点（Endpoint）。
fs.obs.access.key	无	是	华为云的AK（Access Key Id），需要具备访问OBS对应桶的权限。
fs.obs.secret.key	无	是	华为云的SK（Secret Access Key），需要具备访问OBS对应桶的权限。
fs.obs.session.token	无	否	华为云的securitytoken，需要具备访问OBS对应桶的权限。当使用临时AK/SK时需要。
fs.obs.security.provider	无	否	实现com.obs.services.IObsCredentialsProvider接口的类，用于获取访问OBS的凭证。
fs.obs.connection.ssl.enabled	FALSE	否	是否通过HTTPS访问OBS。
fs.obs.threads.keepalivetime	60	否	控制读写线程池参数keepAliveTime。
fs.obs.threads.max	20	否	控制读写线程池参数corePoolSize和maximumPoolSize
fs.obs.max.total.tasks	20	否	控制读写线程池参数BlockingQueue的容量，其等于fs.obs.threads.max+fs.obs.max.total.tasks

配置项	默认值	是否必填	说明
fs.obs.multipart.size	104857600	否	写相关配置，多段上传大小。
fs.obs.fast.upload.buffer	disk	否	写相关配置，所有数据在写入OBS前都会先缓存然后再上传到OBS，此参数用于设置缓存方式，取值范围： <ul style="list-style-type: none"><li>• disk：缓存在磁盘</li><li>• array：缓存在JVM堆内内存</li><li>• bytearray：缓存在JVM堆外内存</li></ul>
fs.obs.buffer.dir	\${hadoop.tmp.dir}	否	写相关配置，当fs.obs.fast.upload.buffer为disk时的缓存目录，支持多目录并以逗号分隔。
fs.obs.bufferdir.verify.enable	FALSE	否	写相关配置，当fs.obs.fast.upload.buffer为disk时是否验证缓存目录是否存在以及是否具备写权限。
fs.obs.fast.upload.active.blocks	4	否	写相关配置，每个流操作最大可以使用的缓存个数（通过多段上传线程池最多可以提交的线程任务个数），从而限制每个流操作最大可以使用的缓存空间 fs.obs.fast.upload.active.blocks*fs.obs.multipart.size。
fs.obs.fast.upload.array.first.buffer	1048576	否	写相关配置，当fs.obs.fast.upload.buffer为array时，此参数控制JVM堆内缓存初始化大小
fs.obs.readahead.range	1048576	否	写相关配置，预读片段大小。
fs.obs.multiobject.delete.enable	TRUE	否	删除相关配置，删除目录时是否启动批量删除。
fs.obs.delete.threads.max	20	否	删除相关配置，控制线程池参数maximumPoolSize和corePoolSize
fs.obs.multiobject.delete.maximum	1000	否	删除相关配置，批量删除时单次OBS批量删除请求中支持的最多可删除对象的个数，最大值为1000。
fs.obs.multiobject.delete.threshold	3	否	删除相关配置，批量删除时当对象个数小于此参数值时将不启动批量删除。

配置项	默认值	是否必填	说明
fs.obs.list.threads.core	30	否	List相关配置，控制线程池参数corePoolSize
fs.obs.list.threads.max	60	否	List相关配置，控制线程池参数maximumPoolSize
fs.obs.list.workqueue.capacity	1024	否	List相关配置，控制线程池参数BlockingQueue的容量
fs.obs.list.parallel.factor	30	否	List相关配置，控制并发因子参数。
fs.obs.paging.maximum	1000	否	List相关配置，单次OBS List请求最多返回的对象个数，最大值为1000。
fs.obs.copy.threads.max	40	否	对象桶rename相关配置，对象桶rename目录时copy线程池配置参数maximumPoolSize，corePoolSize的值为此参数的一半，BlockingQueue的容量为1024。
fs.obs.copypart.size	104857600	否	对象桶rename相关配置，单个对象copy时当对象的大小超过了此参数值则进行多段copy，且段大小为此参数值；否则进行简单copy。
fs.obs.copypart.threads.max	5368709120	否	对象桶rename相关配置，单个对象copy时如果进行了多段copy，多段copy线程池配置参数maximumPoolSize，corePoolSize的值为此参数的一半，BlockingQueue的容量为1024。
fs.obs.getcanonicalservicename.enable	FALSE	否	控制getCanonicalServiceName()接口的返回值。 <ul style="list-style-type: none"> <li>TRUE: obs://bucketname</li> <li>FALSE: null</li> </ul>
fs.obs.multipart.purge	FALSE	否	初始化OBSFilesystem时是否清理桶内的多段上传任务。
fs.obs.multipart.purge.age	86400	否	初始化OBSFilesystem时清理桶内多久之前的多段上传任务。
fs.obs.trash.enable	FALSE	否	是否开启垃圾回收功能。
fs.obs.trash.dir	无	否	垃圾回收目录。
fs.obs.block.size	134217728	否	块大小。

## 15.4.3 Hive 对接 OBS

### 概述

Hive是一个数据仓库工具，可以对存储在分布式存储中的大规模数据进行数据提取、转化和加载，它提供了丰富的SQL查询方式来进行数据分析。

### 前提条件

已安装Hadoop，具体请参见[Hadoop对接OBS](#)。

### 对接步骤

以Hive 2.3.3为例。

**步骤1** 下载apache-hive-2.3.3-bin.tar.gz，并解压到/opt/hive-2.3.3。

**步骤2** 在/etc/profile文件中增加配置内容：

```
export HIVE_HOME=/opt/hive-2.3.3
export PATH=$HIVE_HOME/bin:$PATH
```

**步骤3** 配置Hive。

1. 重命名/opt/hive-2.3.3/conf/hive-env.sh.template为hive-env.sh。
2. 重命名opt/hive-2.3.3/conf/hive-log4j2.properties.template为hive-log4j2.properties。
3. 创建hive-site.xml文件，添加配置：

```
<property>
<name>hive.metastore.warehouse.dir</name>
<value>obs://obs-bucket/warehouse/hive</value>
</property>
```

#### 说明

可选配置，当添加了上述配置后在创建Hive表时将不用再显示指定location，创建的Hive表将自动落在OBS上。

4. 执行以下命令，初始化元数据。

```
/opt/hive-2.3.3/bin/schematool -dbType derby -initSchema
```

**步骤4** 验证是否对接成功。

示例如下，示例中的location为obs://obs-bucket/warehouse/hive/student。

```
hive>
create table student(id int comment "学生id",name string comment "学生姓名",age int comment "学生年龄")
comment "学生信息表"
row format delimited fields terminated by ",";

insert into table student select 6,"yangdong",29;
```

----结束

## 15.4.4 Spark 对接 OBS

### 概述

Apache Spark是专为大规模数据处理而设计的快速通用的计算引擎。

## 前提条件

已安装Hadoop，具体请参见[Hadoop对接OBS](#)。

## 注意事项

为了减少日志输出，在/opt/spark-2.3.3/conf/log4j.properties文件中增加配置：

```
log4j.logger.com.obs= ERROR
```

## 对接步骤

以Spark2.3.3为例。

**步骤1** 下载spark-2.3.3-bin-without-hadoop.tgz，并解压到/opt/spark-2.3.3。

**步骤2** 在/etc/profile文件中增加配置内容：

```
export SPARK_HOME=/opt/spark-2.3.3
export PATH=$SPARK_HOME/bin:$SPARK_HOME/sbin:$PATH
```

**步骤3** 配置spark。

1. 重命名/opt/spark-2.3.3/conf/spark-env.sh.template为spark-env.sh并增加配置：  
export SPARK\_DIST\_CLASSPATH=\$(hadoop classpath)  
更多配置内容请参见[Apache Hadoop](#)。
2. 重命名/opt/spark-2.3.3/conf/log4j.properties.template为log4j.properties。

**步骤4** 执行以下命令，验证是否对接成功。

```
$SPARK_HOME/bin/run-example org.apache.spark.examples.JavaWordCount
obs://obs-bucket/input/test.txt
```

----结束

## 15.4.5 Presto 对接 OBS

### 概述

Presto分为prestoSql（现更名为Trino）和PrestoDB两大分支或是发行版。

Presto on OBS仅支持prestoSql/Trino发行版，下述对接步骤以prestoSql-333版本为例（从prestoSql-332版本开始Presto服务端必须使用JDK 11）。

#### 说明

本章节中的Presto指prestoSql/Trino发行版。

### 前提条件

- 已安装Hadoop，具体请参见[Hadoop对接OBS](#)。
- 已安装Hive，具体请参见[Hive对接OBS](#)。

### 安装 presto server

版本：prestoSQL-333

**步骤1** 下载Presto客户端和服务端。

[下载客户端](#)

[下载服务端](#)

**步骤2** 下载hadoop-huaweicloud插件：[下载地址](#)。

**步骤3** 执行以下命令，解压Presto服务端。

```
tar -zxvf presto-server-333.tar.gz
```

在presto根目录/plugin/hive-hadoop2下放入如下两个jar包。

- [hadoop-huaweicloud-\\${hadoop.version}-hw-\\${version}.jar](#)
- Apache commons-lang-xxx.jar  
可从maven中央仓库下载或从hadoop目录中拷贝。

----结束

## 配置 presto

在安装目录里创建etc目录。这目录会有以下配置（自己创建）：

- 节点配置文件：每个节点的环境配置
- JVM配置文件：Java虚拟机的命令行选项
- Server配置文件（Config Properties）：Presto server的配置
- Catalog配置文件：配置presto的各种Connector（数据源）
- 日志配置文件：配置presto日志

## 节点配置文件

节点属性文件etc/node.properties，包含每个节点的配置。一个节点是一个Presto实例。这文件一般是在Presto第一次安装时创建的。以下是最小配置：

```
node.environment=production
node.id=ffffffff-ffff-ffff-ffff-ffffffffffff
node.data-dir=/var/presto/data
```

解释：

node.environment：环境名字，Presto集群中的节点的环境名字都必须是一样的。

node.id：唯一标识，每个节点的标识都必须是唯一的。就算重启或升级Presto都必须还保持原来的标识。

node.data-dir：数据目录，Presto用它来保存log和其他数据

示例：

```
node.environment=presto_cluster
```

```
node.id=bigdata00
```

```
node.data-dir=/home/modules/presto-server-0.215/data #data需要自己手动创建
```

## JVM 配置文件

JVM配置文件etc/jvm.config, 包含启动Java虚拟机时的命令行选项。格式是每一行是一个命令行选项。此文件数据是由shell解析, 所以选项中包含空格或特殊字符会被忽略。

以下是参考配置:

```
-server
-Xmx16G
-XX:-UseBiasedLocking
-XX:+UseG1GC
-XX:G1HeapRegionSize=32M
-XX:+ExplicitGCInvokesConcurrent
-XX:+ExitOnOutOfMemoryError
-XX:+UseGCOverheadLimit
-XX:+HeapDumpOnOutOfMemoryError
-XX:ReservedCodeCacheSize=512M
-Djdk.attach.allowAttachSelf=true
-Djdk.nio.maxCachedBufferSize=2000000
```

备注: 以上参数都是官网参数, 实际环境需要调整

## Server 配置文件

配置属性文件etc/config.properties, 包含Presto server的配置。Presto server可以同时为coordinator和worker, 但一个大集群里最好就是只指定一台机器为coordinator。

### 1. coordinator节点的配置文件

```
coordinator=true
node-scheduler.include-coordinator=true
http-server.http.port=5050
discovery-server.enabled=true
discovery.uri=http://192.168.XX.XX:5050
query.max-memory=20GB
query.max-memory-per-node=1GB
query.max-total-memory-per-node=2GB
```

### 2. worker节点的配置文件

```
coordinator=false
http-server.http.port=5050
discovery.uri=http://192.168.XX.XX:5050
query.max-memory=20GB
query.max-memory-per-node=1GB
query.max-total-memory-per-node=2GB
```

解释:

coordinator: 是否运行该实例为coordinator (接受client的查询和管理查询执行)。

node-scheduler.include-coordinator:coordinator是否也作为work。对于大型集群来说, 在coordinator里做worker的工作会影响查询性能。

http-server.http.port:指定HTTP端口。Presto使用HTTP来与外部和内部进行交流。

query.max-memory: 查询能用到的最大总内存。

query.max-memory-per-node: 查询能用到的最大单节点内存。

discovery-server.enabled: Presto使用Discovery服务去找到集群中的所有节点。每个Presto实例在启动时都会在Discovery服务里注册。这样可以简化部署, 不需要额外的服务, Presto的coordinator内置一个Discovery服务。



discovery.uri: Discovery服务的URI。将example.net:8080替换为coordinator的host和端口。这个URI不能以斜杠结尾，这个错误需特别注意，否则会报404错误。

另外还有以下属性:

jmx.rmiregistry.port: 指定JMX RMI的注册。JMX client可以连接此端口

jmx.rmiserver.port: 指定JMX RMI的服务器。可通过JMX监听。

## Catalog 配置文件 (重点)

hive connector配置如下:

1. 在etc目录下创建catalog目录
2. 创建一个hive connector的配置文件: hive.properties

```
# hive.properties
#连接名
connector.name=hive-hadoop2
#配置hive metastore连接
hive.metastore.uri=thrift://192.168.XX.XX:9083
#指定hadoop的配置文件，注意core-site.xml需要按照https://github.com/huaweicloud/obsa-hdfs/tree/master/
release/doc配置
hive.config.resources=/home/modules/hadoop-2.8.3/etc/hadoop/core-site.xml,/home/modules/
hadoop-2.8.3/etc/hadoop/hdfs-site.xml,/home/modules/hadoop-2.8.3/etc/hadoop/mapred-site.xml
#给删表权限
hive.allow-drop-table=true
```

## 日志配置文件

创建文件log.properties

写入内容: com.facebook.presto=INFO

备注: 日志级别有四种: DEBUG、INFO、WARN和ERROR。

## 启动 presto

步骤如下:

**步骤1** 启动hive metastore: hive --service metastore &

**步骤2** 启动presto server: bin/launcher start (如何关闭presto服务: bin/launcher stop)

**步骤3** 启动presto client:

1. 重命名presto-cli-333-executable.jar为presto,放在bin目录下,然后赋予执行权限: chmod +x presto
2. 启动client: ./presto --server XX.XX.XX.XX:5050 --catalog hive --schema default

----结束

## Presto 查询 OBS

### 创建hive表

```
hive>
CREATE TABLE sample01(id int,name string,address string)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
```

```
LOCATION 'obs://obs-east-bkt001/sample01';
insert into sample01 values(1,'xiaoming','cd');
insert into sample01 values(2,'daming','sh');
```

### presto查询hive表

```
./presto --server XX.XX.XX.XX:5050 --catalog hive --schema default
```

```
presto:default>
select * from sample01;
```

## 15.4.6 Flume 对接 OBS

### 概述

Flume是一个分布式的、可靠的和高可用的服务，用于收集、聚合以及移动大量日志数据，具体请参见[Apache Flume](#)。OBS在大数据场景中可以替代Hadoop系统中的HDFS服务。

### 注意事项

- 多sink写同一文件  
OBS和HDFS在一致性保证上是有差别的：HDFS租约机制可以保证并发写同一个文件时不会产生一致性问题，但是OBS实现的HDFS协议不支持租约Lease机制（并发写同一个文件时将产生不可确定的状态），所以在flume场景下可以通过文件命名规则进行解决。  
如sink文件的命名规则：hostname-sinkname作为文件的前缀，如果一个主机上部署了多个flume agent，不同的agent要有不同的sinkname。
- flume日志配置  
为了减少日志输出，在/opt/apache-flume-1.9.0-bin/conf/log4j.properties文件中增加配置：  
log4j.logger.com.obs=ERROR
- obsa写入时临时文件的目录配置  
Flume写OBS时会先写入本地磁盘缓冲区，然后上传到OBS，如果对写入OBS有极致性能要求请选择高性能磁盘作为缓冲区，在core-site.xml文件中增加配置：  

```
<property>
<name>fs.obs.buffer.dir</name>
<value>xxx</value>
</property>
```

### 对接步骤

以flume 1.9版本为例。

**步骤1** 下载apache-flume-1.9.0-bin.tar.gz。

**步骤2** 安装flume。

解压apache-flume-1.9.0-bin.tar.gz到/opt/apache-flume-1.9.0-bin目录。

- 已部署Hadoop的环境：无需额外操作，部署Hadoop请参见[Hadoop对接OBS](#)。
- 未部署Hadoop的环境：
  - a. 将hadoop中的相关jar包复制到/opt/apache-flume-1.9.0-bin/lib目录下，包含hadoop-huaweicloud-xxx.jar。

- b. 将添加了OBS相关配置的core-site.xml文件复制到/opt/apache-flume-1.9.0-bin/conf目录下。

### 步骤3 验证是否对接成功。

示例：以flume内置的StressSource为source，以file为channel，以obs为sink。

1. 创建flume配置文件：sink2obs.properties。

```
agent.sources = r1
agent.channels = c1
agent.sinks = k1

agent.sources.r1.type = org.apache.flume.source.StressSource
agent.sources.r1.channels = c1
agent.sources.r1.size = 1024
agent.sources.r1.maxTotalEvents = 100000
agent.sources.r1.maxEventsPerSecond = 10000
agent.sources.r1.batchSize=1000

agent.sources.r1.interceptors = i1
agent.sources.r1.interceptors.i1.type = host
agent.sources.r1.interceptors.i1.useIP = false

agent.channels.c1.type = file
agent.channels.c1.dataDirs = /data/agent/flume-data
agent.channels.c1.checkpointDir = /data/agent/flume-checkpoint
agent.channels.c1.capacity = 500000
agent.channels.c1.transactionCapacity = 50000

agent.sinks.k1.channel = c1
agent.sinks.k1.type = hdfs
agent.sinks.k1.hdfs.useLocalTimeStamp = true
agent.sinks.k1.hdfs.filePrefix = ${host}_k1
agent.sinks.k1.hdfs.path = obs://obs-bucket/flume/create_time=%Y-%m-%d-%H-%M
agent.sinks.k1.hdfs.fileType = DataStream
agent.sinks.k1.hdfs.writeFormat = Text
agent.sinks.k1.hdfs.rollSize = 0
agent.sinks.k1.hdfs.rollCount = 1000
agent.sinks.k1.hdfs.rollInterval = 0
agent.sinks.k1.hdfs.batchSize = 1000
agent.sinks.k1.hdfs.round = true
agent.sinks.k1.hdfs.roundValue = 10
agent.sinks.k1.hdfs.roundUnit = minute
```

2. 执行以下命令，启动flume agent。

```
./bin/flume-ng agent -n agent -c conf/ -f conf/sink2obs.properties
```

----结束

## 15.4.7 DataX 对接 OBS

### 概述

DataX是一个数据同步框架，实现了包括MySQL、SQL Server、Oracle、PostgreSQL、HDFS、Hive、HBase、OTS、ODPS等各种异构数据源之间高效的数据同步功能。OBS在大数据场景中可以替代Hadoop系统中的HDFS服务，本文介绍DataX如何对接OBS。

### 对接步骤

**步骤1** 下载datax源码，以发布版本datax\_v202308为例：[下载地址](#)。

**步骤2** 修改编译datax。

1. 升级hdfsreader和hdfswriter模块依赖的hadoop版本，以升级到2.8.3版本为例。  
修改datax\hdfswriter\pom.xml和datax\hdfsreader\pom.xml文件配置：

```
<properties>
<!--由2.7.1升级到2.8.3-->
<hadoop.version>2.8.3</hadoop.version>
</properties>
```

2. 编译datax。
3. 执行以下命令，在datax源码根目录/target目录下生成datax.tar.gz压缩文件。  
**mvn -U clean package assembly:assembly -Dmaven.test.skip=true**

### 步骤3 安装datax。

1. 解压datax.tar.gz到/opt/datax目录。
2. 在Github下载hadoop-huaweicloud：[下载地址](#)。（建议使用hadoop 2.8.3版本下最新版本的hadoop-huaweicloud版本，例如hadoop-huaweicloud-2.8.3-hw-53.8，以最新版本为准）
3. 将上述下载的jar包放入/opt/datax/plugin/writer/hdfswriter/libs和/opt/datax/plugin/reader/hdfsreader/libs。

### 步骤4 验证是否对接成功。

示例：以txtfilereader为源端，以OBS为目的端。

1. 创建作业配置文件file2obs.json。

```
{
  "setting":{
  },
  "job":{
    "setting":{
      "speed":{
        "channel":2
      }
    },
    "content":[
      {
        "reader":{
          "name":"txtfilereader",
          "parameter":{
            "path":[
              "/opt/test.txt"
            ],
            "encoding":"UTF-8",
            "column":[
              {
                "index":0,
                "type":"STRING"
              },
              {
                "index":1,
                "type":"STRING"
              }
            ]
          },
          "fieldDelimiter":"\t"
        },
        "writer":{
          "name":"hdfswriter",
          "parameter":{
            "defaultFS":"obs://obs-bucket",##obs桶
            "fileType":"text",
            "path":"/test",##obs桶中的路径
            "fileName":"test",
```

```
"column":[
  {
    "name":"col1",
    "type":"STRING"
  },
  {
    "name":"col2",
    "type":"STRING"
  }
],
"writeMode":"append",
"fieldDelimiter":"\t",
"hadoopConfig":{"##此部分hadoop配置必须添加
  "fs.obs.impl":"org.apache.hadoop.fs.obs.OBSFileSystem",
  "fs.obs.access.key":"可访问OBS的ak",
  "fs.obs.secret.key":"可访问OBS的sk",
  "fs.obs.endpoint":"OBS桶所在region"
}
}
```

2. 启动datax。

```
python /opt/datax/bin/datax.py file2obs.json
```

----结束

## 15.4.8 Druid 对接 OBS

### 概述

Druid专为需要快速数据查询与摄入的工作流程而设计，在即时数据可见性、即席查询、运营分析以及高并发等方面表现非常出色。

通过HDFS接口对接OBS，使用OBS提供的OBSA-HDFS工具，无需重新编译druid，将OBS配置为deep storage。

### 对接步骤

#### 步骤1 配置Druid。

1. 修改配置：

```
conf/druid/single-server/micro-quickstart/_common/common.runtime.properties
```

将druid-hdfs-storage加入druid.extensions.loadList。

```
# If you specify 'druid.extensions.loadList=[]', Druid won't load any extension from file system.
# If you don't specify 'druid.extensions.loadList', Druid will load all the extensions under root extension directory.
# More info: https://druid.apache.org/docs/latest/operations/including-extensions.html
druid.extensions.loadList=["druid-hdfs-storage", "druid-kafka-indexing-service", "druid-datasketches"]
```

2. 配置Deep storage在OBS中的存储路径。

```
#  
# Deep storage  
#  
# For local disk (only viable in a cluster if this is a network mount):  
#druid.storage.type=local  
#druid.storage.storageDirectory=var/druid/segments  
  
# For HDFS:  
druid.storage.type=hdfs  
druid.storage.storageDirectory=obs://wxg-sg-oms-test/druidhdfs/segments  
  
#  
# Indexing service logs  
#  
# For local disk (only viable in a cluster if this is a network mount):  
#druid.indexer.logs.type=file  
#druid.indexer.logs.directory=var/druid/indexing-logs  
  
# For HDFS:  
druid.indexer.logs.type=hdfs  
druid.indexer.logs.directory=obs://wxg-sg-oms-test/druidhdfs/indexing-logs
```

**步骤2** 配置OBSA-HDFS插件。

1. 在官方Github下载OBSA-HDFS插件：[下载地址](#)，然后拷贝到extensions/druid-hdfs-storage/ 目录。
2. 在配置目录conf/druid/single-server/micro-quickstart/\_common/下增加hdfs-site.xml，配置如下（其中endpoint按照桶所在的实际endpoint填写）：

```
<configuration>
  <property>
    <name>fs.obs.access.key</name>
    <value>[REDACTED]</value>
  </property>
  <property>
    <name>fs.obs.secret.key</name>
    <value>[REDACTED]</value>
  </property>
  <property>
    <name>fs.obs.endpoint</name>
    <value>obs.ap-southeast-3.myhuaweicloud.com</value>
  </property>
  <property>
    <name>fs.obs.buffer.dir</name>
    <value>/home/modules/data/buf</value>
  </property>
  <property>
    <name>fs.obs.impl</name>
    <value>org.apache.hadoop.fs.obs.OBSFileSystem</value>
  </property>
</configuration>
```

步骤3 启动Druid服务。

----结束

## 15.4.9 Flink 对接 OBS

### 概述

Flink是一个分布式的数据处理引擎，用于处理有界和无界流式数据。Flink定义了文件系统抽象，OBS服务实现了Flink的文件系统抽象，使得OBS可以作为flink StateBackend和数据读写的载体。

### 注意事项

- flink-obs-fs-hadoop目前仅支持OBS并行文件系统。
- 为了减少日志输出，在/opt/flink-1.12.1/conf/log4j.properties文件中增加配置：  
logger.obs.name=com.obs  
logger.obs.level=ERROR
- flink-obs-fs-hadoop的实现基于flink的plugin加载机制（flink从1.9开始引入），flink-obs-fs-hadoop必须通过flink的plugin机制进行加载，即将flink-obs-fs-hadoop放入/opt/flink-1.12.1/plugins/obs-fs-hadoop目录下。

### 对接步骤

以flink-1.12.1为例。

**步骤1** 下载flink-1.12.1-bin-scala\_2.11.tgz，并解压到/opt/flink-1.12.1目录。

**步骤2** 在/etc/profile文件中增加配置：

```
export FLINK_HOME=/opt/flink-1.12.1
export PATH=$FLINK_HOME/bin:$PATH
```

**步骤3** 安装flink-obs-fs-hadoop。

1. 在Github下载flink-obs-fs-hadoop：[下载地址](#)。

#### 📖 说明

- flink-obs-fs-hadoop-\${flinkversion}-hw-\${version}.jar版本规则：flinkversion为对应的flink版本号，version为flink-obs-fs-hadoop版本号。
  - 如果没有匹配版本的jar包，可自行修改flink-obs-fs-hadoop目录下pom文件中的flink版本重新编译生成。
  - 自行编译flink-obs-fs-hadoop时，推荐编译依赖的hadoop.huaweicloud版本（hadoop.huaweicloud.version）不低于53.8版本。
2. 在/opt/flink-1.12.1/plugins目录下创建obs-fs-hadoop目录，并将上述jar放入此目录。

**步骤4** 配置flink。

在/opt/flink-1.12.1/conf/flink-conf.yaml文件中或在代码中设置如下参数：

```
fs.obs.impl: org.apache.hadoop.fs.obs.OBSFileSystem
fs.obs.access.key: xxx
fs.obs.secret.key: xxx
fs.obs.endpoint: xxx
fs.obs.buffer.dir: /data/buf #写数据到OBS时需要的本地临时目录，flink程序需具备此目录读写权限
```

**步骤5** 编写flink应用程序。

1. StateBackend设置为OBS中的路径。

示例：

```
env.setStateBackend(new FsStateBackend("obs://obs-bucket/test/checkpoint"));
```

2. StreamingFileSink设置为OBS中的路径。

示例：

```
final StreamingFileSink<String> sink = StreamingFileSink
.forRowFormat(new Path("obs://obs-bucket/test/data"),
new SimpleStringEncoder<String>("UTF-8"))
.withBucketAssigner(new BasePathBucketAssigner())
.withRollingPolicy(rollingPolicy)
.withBucketCheckInterval(1000L)
.build();
```

----结束

## 15.4.10 Logstash 对接 OBS

### 概述

Logstash能够从多个来源采集数据、转换数据并将数据发送到存储系统中。本文用于描述Logstash如何对接使用OBS。

### 注意事项

请使用较新版本的logstash，例如≥7.10.2的版本，避免使用较老版本的logstash。



## 对接步骤

以logstash-7.10.2为例。

**步骤1** 下载logstash-7.10.2-linux-x86\_64.tar.gz，并解压到/opt/logstash-7.10.2-linux-x86\_64目录。

**步骤2** 验证是否对接成功。

示例：以file为源端，以OBS为目的端。

1. 创建配置文件file2obs.conf。参数说明见表15-3，更多详情请参见[这里](#)。

```
input {
  file {
    path => "/opt/nginx/logs/access.log"
    start_position => "beginning"
  }
}

output {
  s3 {
    endpoint => "obs endpoint" # The endpoint should be an HTTP or HTTPS URL
    access_key_id => "ak"
    secret_access_key => "sk"
    bucket => "obs bucket name"
    size_file => 1048576
    time_file => 1
    prefix => "logstash/"
    enable_metric => true
  }
}
```

表 15-3 参数说明

参数	说明
endpoint	OBS的endpoint，例如 - https://obs.ap-southeast-1.myhuaweicloud.com - http://obs.ap-southeast-1.myhuaweicloud.com
access_key_id	具备访问OBS权限的ak。
secret_access_key	具备访问OBS权限的sk。
bucket	OBS的桶名称。
size_file	指定文件滚动大小（字节）。当文件大小达到设定的值时，会生成一个新的文件。
time_file	设置文件滚动周期（分钟）。当数据写入达到设定周期时，会生成一个新的文件。
prefix	指定文件存储的目录，例如“logstash/”，此时文件会写入到桶的logstash/目录下（注意路径不要以/开头）。

2. 执行以下命令，运行logstash。

```
bin/logstash -f ../conf/file2obs.conf
```

----结束

## 15.5 迁移 HDFS 数据至 OBS

### 操作场景

在华为云大数据存算分离方案中，对象存储服务OBS作为统一数据湖存储数据。如果用户数据仍存储在本地HDFS中，则需要先将HDFS的数据迁移至OBS。

用户可以使用以下迁移方案中的任意一种完成数据迁移，包括：[Distcp方式迁移](#)、[CDM方式迁移](#)和[OMS方式迁移](#)。

### Distcp 方式迁移

Hadoop Distcp ( Distributed copy ) 主要是用于Hadoop文件系统内部或之间进行大规模数据复制的工具，它使用Map/Reduce实现文件分发，错误处理和恢复，以及报告生成。它把文件和目录的列表作为map任务的输入，每个任务会完成源列表中部分文件的拷贝。

#### 配置指南

参考[Hadoop对接OBS](#)中hadoop-huaweicloud的安装和配置方法，完成OBS相关配置。

#### 使用示例

**步骤1** 以迁移HDFS上的“/data/sample”目录为例，执行以下命令查看HDFS文件系统上此目录下的文件与目录。

```
hadoop fs -ls hdfs:///data/sample
```

**步骤2** 执行以下命令，将HDFS文件系统上“/data/sample”目录下所有文件与目录迁移到OBS桶“obs-bigdata-posix-bucket”的“data/sample”目录下。

```
hadoop distcp hdfs:///data/sample obs://obs-bigdata-posix-bucket/data/sample
```

**步骤3** 执行以下命令，查看拷贝的文件。

```
hadoop fs -ls obs://obs-bigdata-posix-bucket/data/sample
```

----结束

### CDM 方式迁移

云数据迁移 ( Cloud Data Migration, CDM ) 提供同构/异构数据源之间批量数据迁移服务，帮助您实现数据自由流动。支持关系数据库，数据仓库，NoSQL，大数据云服务数据源。

详细内容请参见[云数据迁移](#)。

### OMS 方式迁移

对象存储迁移服务 ( Object Storage Migration Service, OMS ) 是一种线上数据迁移服务，帮助您将其他云服务商对象存储服务中的数据在线迁移至华为云的对象存储服务 ( Object Storage Service, OBS ) 中。

详细内容请参见[对象存储迁移服务](#)。

# 16 通过 s3fs 挂载 OBS 对象桶

## 应用场景

对于原本使用本地文件系统存储数据的用户，希望把数据存储到OBS上但不改变数据的访问方式，推荐使用s3fs。

s3fs是一款基于用户空间文件系统（Filesystem in Userspace，简称FUSE）的文件系统工具，支持您在Linux系统中，将OBS对象桶挂载到本地文件系统，让您能够像操作本地文件一样操作对象，实现数据的共享。关于s3fs的更多信息，请参见[GitHub](#)。如果您在s3fs使用过程中遇到问题，请参见[常见问题](#)进行排查。

## 方案优势

- 支持上传和下载文件、目录、软链接、设置用户权限等POSIX文件协议标准的大部分功能。
- 支持多段上传。
- 可以使用本地磁盘作为缓存提高IO性能。

## 约束与限制

- 随机写或追加写文件场景需要重写整个文件，会有带宽浪费。
- 由于网络延迟，元数据操作（如列举目录）性能较差。
- 不支持文件或目录的原子性重命名。
- 支持挂载到多台云服务器，但数据一致性需由您自行维护，避免出现多台云服务器同时对一个文件并发写入的情况。
- 不支持硬链接。
- s3fs与存储服务端交互采用HTTP/HTTPS协议，客户端CPU开销会较大。
- 因为客户端存在操作系统元数据缓存，可能导致该缓存失效前，客户端元数据与存储服务端元数据呈现不一致现象。
- 因使用FUSE框架存在内核态与用户态切换，不适合高并发场景。

## 实施步骤

通过s3fs挂载OBS对象桶的步骤如下：

**步骤1** 下载相关依赖：

```
yum install fuse
sudo yum install automake fuse fuse-devel gcc-c++ git libcurl-devel libxml2-devel make openssl-devel
```

**⚠ 注意**

需要保证云服务器上fuse版本 $\geq 2.8.4$ ，部分过老操作系统可能不满足这个约束，需要自行适配。

**步骤2** 支持通过以下两种方式下载s3fs。

- 方式一：镜像源安装。

```
# Ubuntu
sudo apt install s3fs

# CentOS
sudo yum install epel-release
sudo yum install s3fs-fuse
```

- 方式二：GitHub获取指定版本安装。

推荐使用[s3fs 1.91 版本](#)，下载以后自行编译安装。

**步骤3** 编译。使用[方式一](#)安装s3fs请跳过该步骤。

进入s3fs-fuse解压目录，按照以下命令执行编译操作：

```
./autogen.sh
./configure
make
sudo make install
```

**步骤4** 执行以下命令检查安装情况。

```
s3fs --version
```

当命令回显出现s3fs版本信息时，表示s3fs能够成功运行。

**步骤5** 配置AK:SK。

按照以下命令配置AK:SK：

```
echo "AK:SK" >>/root/.passwd-s3fs
chmod 600 /root/.passwd-s3fs
```

**📖 说明**

获取AK:SK方法请参见[获取访问密钥（AK和SK）](#)。

**步骤6** 挂载。

按照以下格式执行挂载命令：

```
s3fs 对象桶名 本地挂载目录 -o passwd_file=密钥文件路径 -o url=区域终端节点地址 -o nonempty -o
big_writes -o max_write=131072 其他挂载参数
```

例如，通过s3fs挂载中国-香港区域的OBS对象桶“test-bucket”，命令示例如下：

```
s3fs test-bucket /s3fs-test -o passwd_file=/root/.passwd-s3fs -o url=https://obs.ap-
southeast-1.myhuaweicloud.com -o nonempty -o big_writes -o max_write=131072
```

**步骤7** 验证。

执行以下命令，检查挂载结果。

```
df -h
```

当界面显示如下类似回显信息时，表示对象桶挂载成功。

```
Filesystem      Size  Used Avail Use% Mounted on
s3fs            16E   0  16E   0% /s3fs-test
```

如果没有显示以上类似回显信息，表示OBS桶挂载失败。此时可在执行的挂载命令后面添加以下参数，用以在回显中输出挂载过程以及详细调试日志：

```
-d -d -f -o f2 -o curldb
```

----结束

## 常见参数说明

表 16-1 常见参数说明

参数	说明
umask	<p><b>参数解释：</b> 配置文件权限的掩码，使用此参数会覆盖并行文件系统内所有文件的权限。</p> <p><b>默认取值：</b> 0000</p>
nonempty	<p><b>参数解释：</b> 添加该参数后允许在非空目录上挂载。</p> <p><b>默认取值：</b> 无</p>
multipart_size	<p><b>参数解释：</b> 以多段上传模式上传文件时，段的大小，单位为MB。段的大小会影响能够上传的文件大小，具体参考<a href="#">多段上传</a>。</p> <p><b>取值范围：</b> 5~5120，单位：MB。</p> <p><b>默认取值：</b> 10</p>
no_check_certificate	<p><b>参数解释：</b> 对服务端的证书不做校验，仅在使用HTTPS协议时有效，默认开启证书校验。</p> <p><b>默认取值：</b> 无</p>
use_cache	<p><b>参数解释：</b> 用于本地文件缓存的地址，使用该参数后，IO性能会提升，但是磁盘占用会升高，可以和del_cache配合使用。</p> <p><b>默认取值：</b> ""（表示不使用缓存）</p>

参数	说明
del_cache	<b>参数解释:</b> s3fs进程在启动和退出时删除本地缓存文件。 <b>默认取值:</b> 无

更多参数详见[s3fs-fuse](#)。