

阿里云

阿里云最佳实践 数据迁移

文档版本：20220428

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
 危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 危险 重置操作将丢失用户配置数据。
 警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告 重启操作将导致业务中断，恢复业务时间约十分钟。
 注意	用于警示信息、补充说明等，是用户必须了解的内容。	 注意 权重设置为0，该服务器不会再接受新请求。
 说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击设置> 网络> 设置网络类型。
粗体	表示按键、菜单、页面名称等UI元素。	在结果确认页面，单击确定。
Courier字体	命令或代码。	执行 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid</code> <code>Instance_ID</code>
[] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1.自建ELK迁移阿里云日志服务	05
2.Oracle RAC云上部署最佳实践	06
3.湖仓一体架构EMR元数据迁移DLF	07
4.开源Flink迁移实时计算Flink全托管版最佳实践	08
5.CDH迁移升级CDP最佳实践	09
6.自建Hadoop迁移MaxCompute	10
7.自建Hadoop数据迁移到阿里云EMR	11
8.影视数据分发汇集与传输加速	12
9.企业用户多账号合并之存储迁移集中	13
10.利用低成本链路完成业务数据迁移上云	14
11.低成本跨境文件传输	15
12.自建Hive数据仓库跨版本迁移到阿里云EMR	16
13.数据迁移上云最佳实践	17

1.自建ELK迁移阿里云日志服务

ELK Stack (Elastic/Logstash/Kibana) 通常被用于日志实时分析，在开源社区中有大量的内容和使用案例。本文介绍如何将自建的ELK系统迁移到阿里云SLS日志服务。

直达最佳实践

[点击查看最佳实践详情](#)

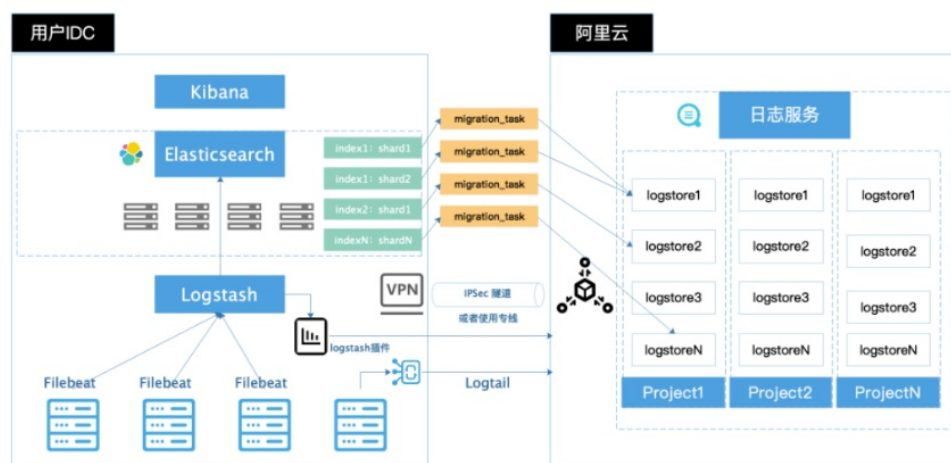
更多最佳实践

[点击查看更多阿里云最佳实践。](#)

解决问题

- 自建ELK如何迁移到阿里云日志服务。
- 如何使用Logtail采集日志。
- 如何使用阿里云日志服务对日志进行查询、分析。

部署架构



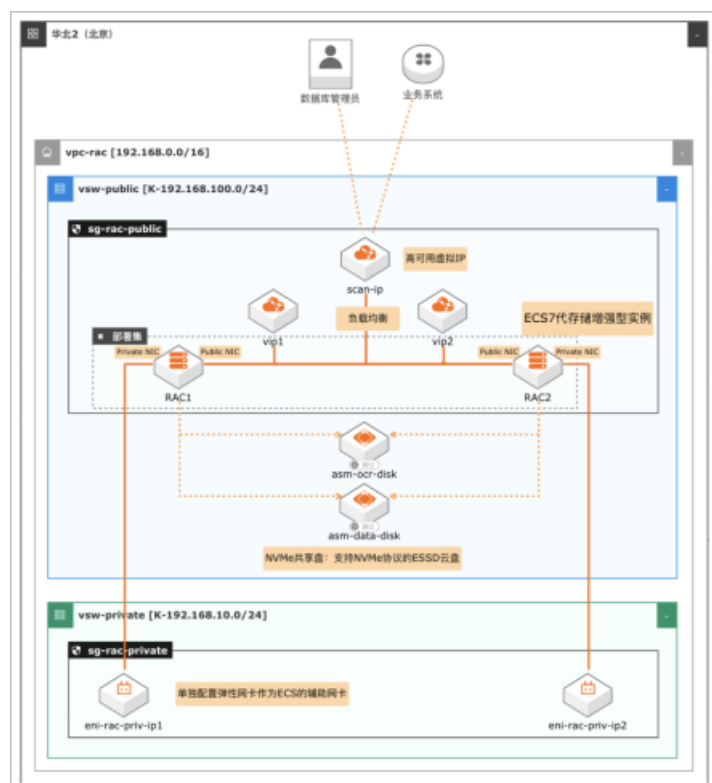
2.Oracle RAC云上部署最佳实践

Oracle RAC 架构迁移上云，提供高并发，高吞吐，高安全等特性，适用于金融，电力，电信，制造业等传统客户的核心交易系统。

方案优势

1. 基于 ESSD 存储 NVMe 协议提供的多实例共享能力的双机集群 Oracle RAC架构，为用户提供统一数据服务的同时提供故障切换与恢复能力（Fail Over 集群功能），避免单点故障，减少停机时间，确保系统全年 7*24 小时稳定运行。
2. 在集群环境运行中的所有实例通过共享的数据库运行事物，RAC 架构相对上层应用架构完全透明，整体数据负载流量可以在集群中尽可能平均合理地分摊处理，无资源闲置，保护用户原有资产的同时提升资源利用率。
3. RAC 集群架构，具有高伸缩能力和自动负载均衡等特性，当负载增加时可以通过增加节点来提高处理能力和吞吐量，水平扩充活动节点，实现性能平滑升级且近线性的扩展。

方案架构



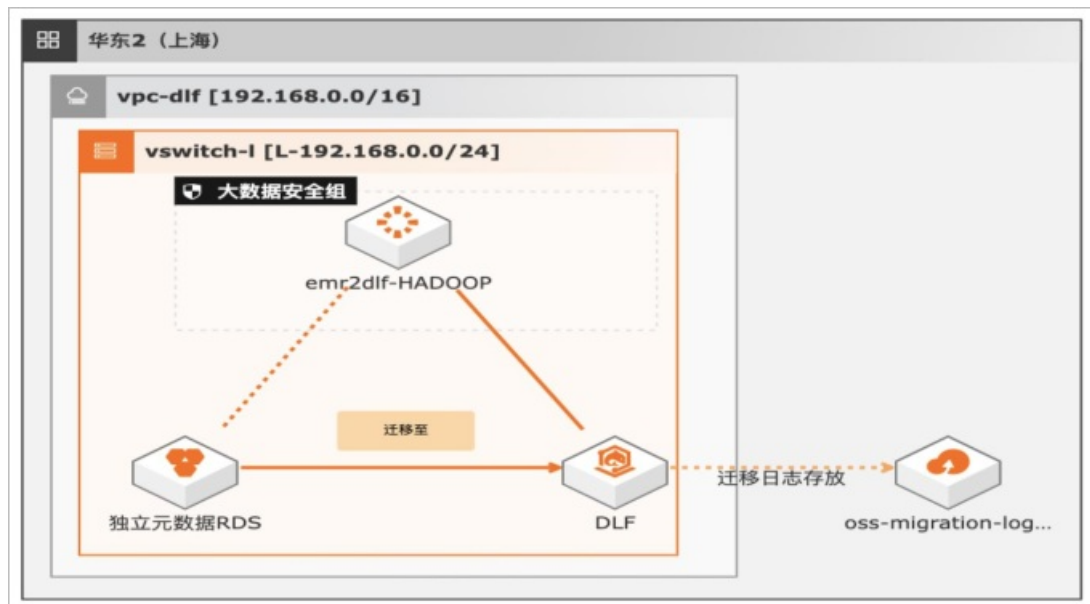
查看详情

[查看最佳实践](#)

3.湖仓一体架构EMR元数据迁移DLF

通过 EMR+DLF 数据湖方案，可以为企业提供数据湖内的统一的元数据管理，统一的权限管理，支持多源数据入湖以及一站式数据探索的能力。本方案支持已有EMR集群元数据库使用RDS或内置 MySQL数据库迁移 DLF，通过统一的元数据管理，多种数据源入湖，搭建高效的数据湖解决方案。

方案架构



解决问题

- EMR 元数据迁移至 DLF
- 元数据迁移验证
- 数据一致性校验

查看详情

[查看最佳实践](#)

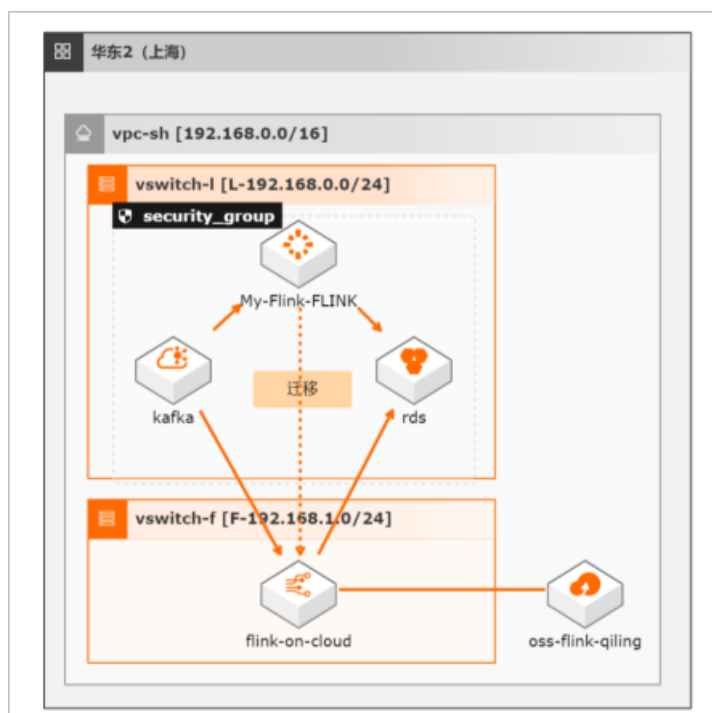
4. 开源Flink迁移实时计算Flink全托管版最佳实践

本方案介绍如何将自建开源Flink集群的流式任务（包含Datastream、Table/SQL、PyFlink任务）迁移至阿里云实时计算全托管版。

场景描述

Flink 全托管产品（Flink Serverless）是一款基于Apache Flink 构建的全托管产品，为您提供全托管一站式的实时计算服务，具有免运费、高增值、低成本等特性。本方案介绍如何将自建开源 Flink 集群的流式任务（包含 Datastream、Table/SQL、PyFlink 任务）迁移至阿里云实时计算全托管版。

方案架构



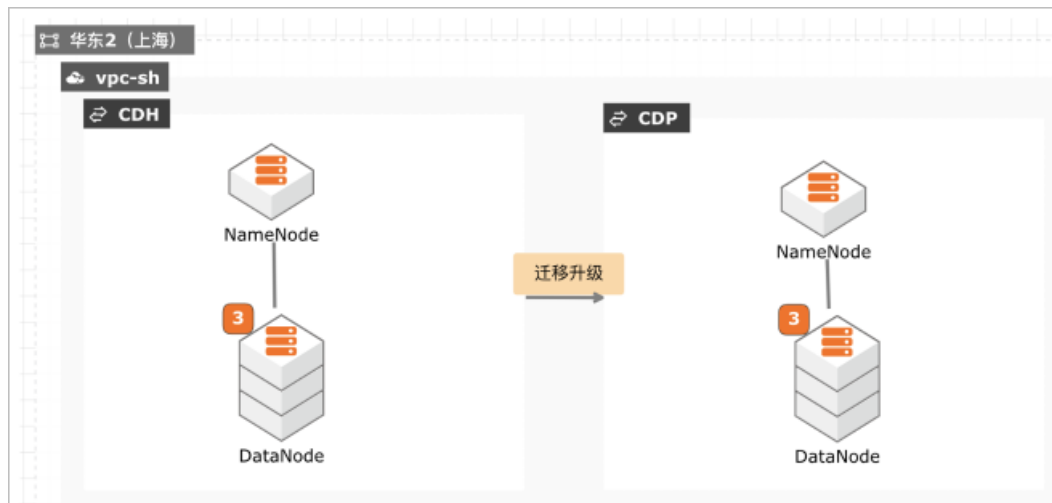
查看详情

[查看最佳实践](#)

5.CDH迁移升级CDP最佳实践

当前 CDH 免费版停止下载，终止服务，针对需要企业版服务能力并且 CDH 升级过程对业务影响较小的客户，通过安装新的 CDP 集群，将现有数据拷贝至新集群，然后将新集群切换为生产集群，升级过程没有数据丢失风险，停机时间较短，适合大部分互联网客户升级使用。

方案架构



解决问题

- CDH 升级至 CDP
- 要求升级过程无数据丢失风险
- 要求升级宕机时间及短
- 数据和组件迁移
- 数据一致性校验

查看详情

[查看最佳实践](#)

6. 自建Hadoop迁移MaxCompute

介绍如何将自建Hadoop集群及生态组件迁移到阿里云MaxCompute大数据服务。

直达最佳实践

[点击查看最佳实践详情](#)

更多最佳实践

[点击查看更多阿里云最佳实践。](#)

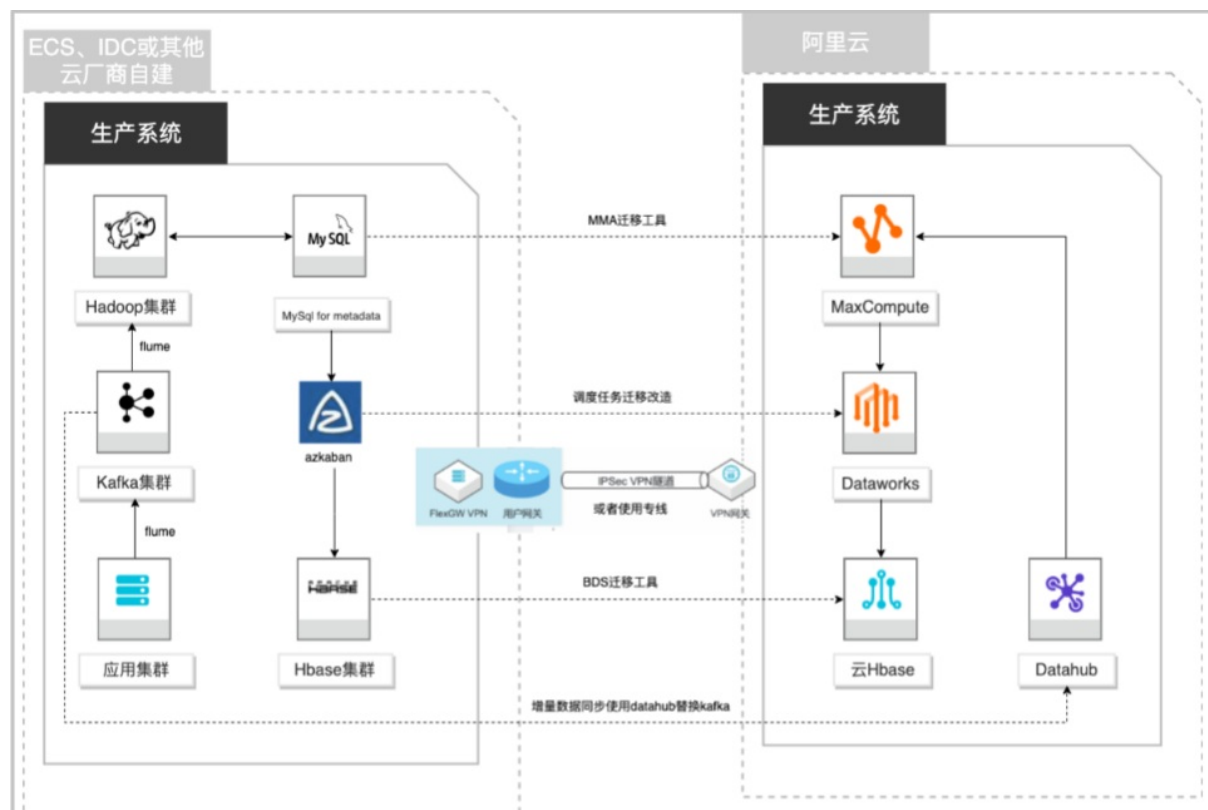
场景描述

客户基于ECS、IDC自建或在友商云平台自建了大数据集群，为了降低企业大数据计算平台的成本，提高大数据应用开发效率，更有效保障数据安全，把大数据集群的数据、作业、调度任务以及业务数据库整体迁移到MaxCompute和其他云产品。

解决的问题

- 自建Hadoop集群搬迁到MaxCompute
- 自建Hbase集群搬迁到云Hbase
- 自建Kafka或应用数据准实时同步到 MaxCompute
- 自建Azkaban任务迁移到Dataworks任务

部署架构



7. 自建Hadoop数据迁移到阿里云EMR

介绍如何将客户自建Hadoop/Hive数据仓库迁移到阿里云EMR的技术实现方案和实践步骤。

直达最佳实践

[点击查看最佳实践详情](#)

更多最佳实践

[点击查看更多阿里云最佳实践](#)

场景描述

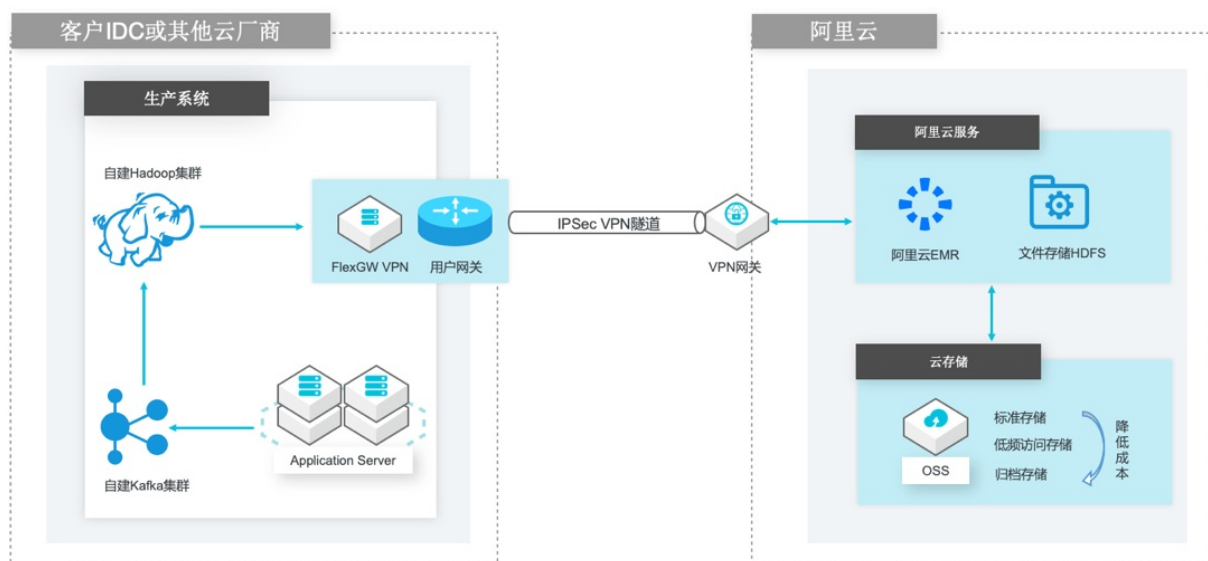
场景1：自建Hadoop集群数据（HDFS）迁移到阿里云EMR集群的HDFS文件系统。

场景2：自建Hadoop集群数据（HDFS）迁移到计算存储分离架构的阿里云EMR集群，以OSS和JindoFS作为EMR集群的后端存储。

解决的问题

- 客户自建Hadoop迁移到阿里云EMR集群的技术方案。
- 基于IPSec VPN隧道构建安全和低成本数据传输链路。

部署架构图



8. 影视数据分发汇集与传输加速

使用阿里云服务实现影视行业数据传输、汇集、分发。

直达最佳实践

[点击查看最佳实践详情](#)

更多最佳实践

[点击查看更多阿里云最佳实践。](#)

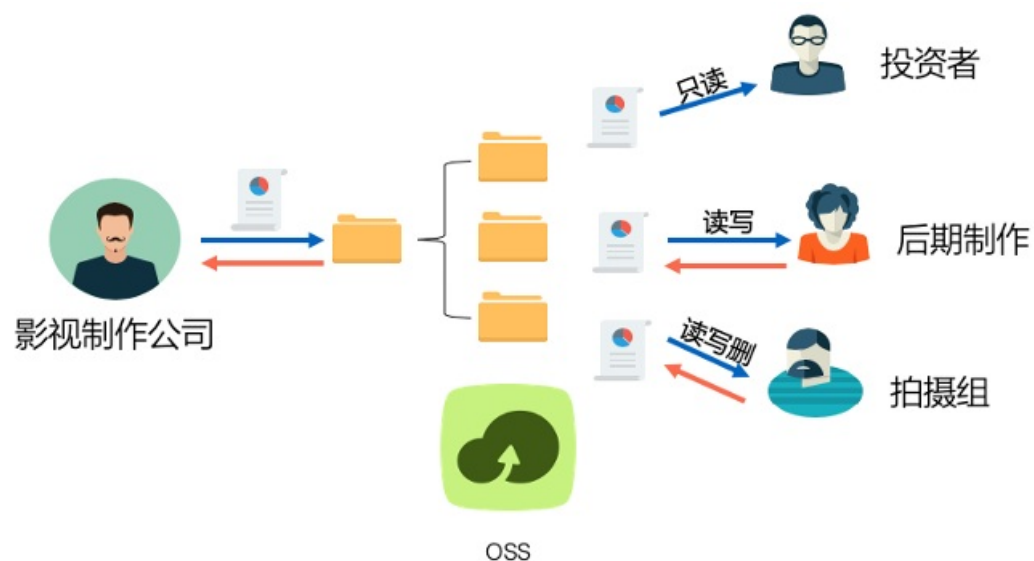
场景描述

使用阿里云对象存储服务OSS及OSSBrowser工具，实现影视数据多用户多地域分发、汇集，以及传输加速的最佳实践。

解决的问题

- 跨地域、多角色、分权限文件上传、下载。
- 基于OSS及其客户端实现Serverless文件分发服务（可替换传统FTP服务）。
- OSS文件传输全球加速。

部署架构图



9.企业用户多账号合并之存储迁移集中

企业多云账号场景下将对象存储服务中的数据归集至一个云账号下为例，介绍多账号场景下合并数据集中存储。

直达最佳实践

[点击查看最佳实践详情](#)

更多最佳实践

[点击查看更多阿里云最佳实践。](#)

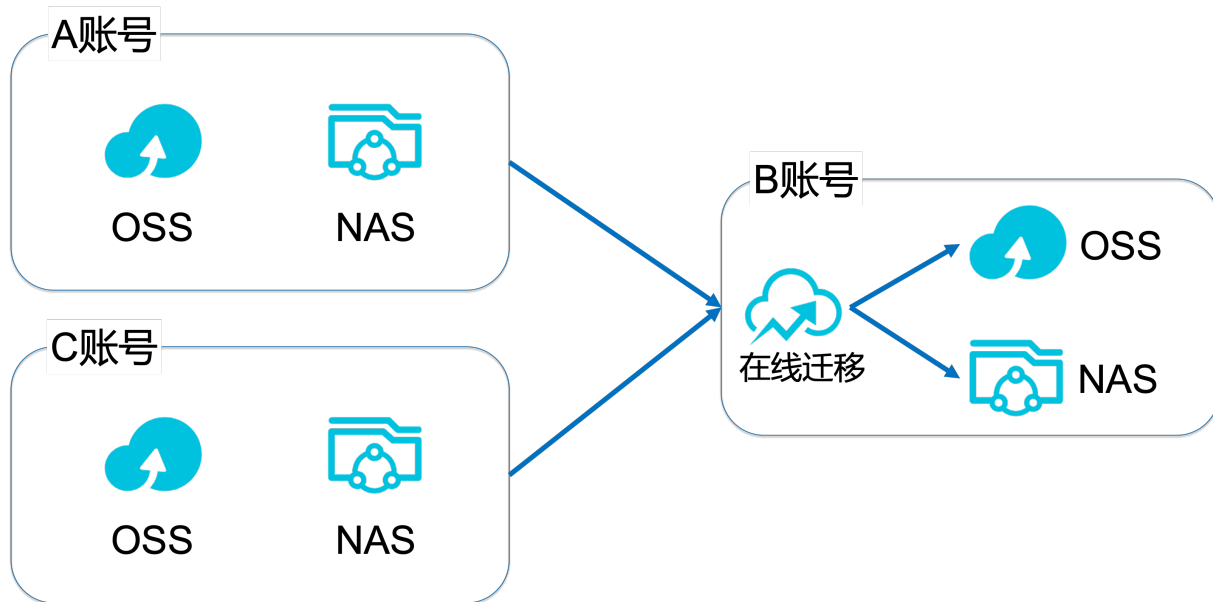
场景描述

本文介绍使用在线迁移服务，将分布在各个云账号中的对象存储、文件存储数据集中到一个账号的对象存储或文件存储下。

解决的问题

- 安全治理需求，统一的账户体系及身份、权限管理。
- 业务系统相互访问，数据统一需求。
- 系统架构及资源成本优化需求，多账号下的带宽、流量、存储包等资源整合优化。
- 企业或部门合并时云账号的合并。

部署架构图



10.利用低成本链路完成业务数据迁移上云

在上云数据迁移阶段，降低所使用的网络成本支出。

直达最佳实践

[点击查看最佳实践详情](#)

更多最佳实践

[点击查看更多阿里云最佳实践](#)

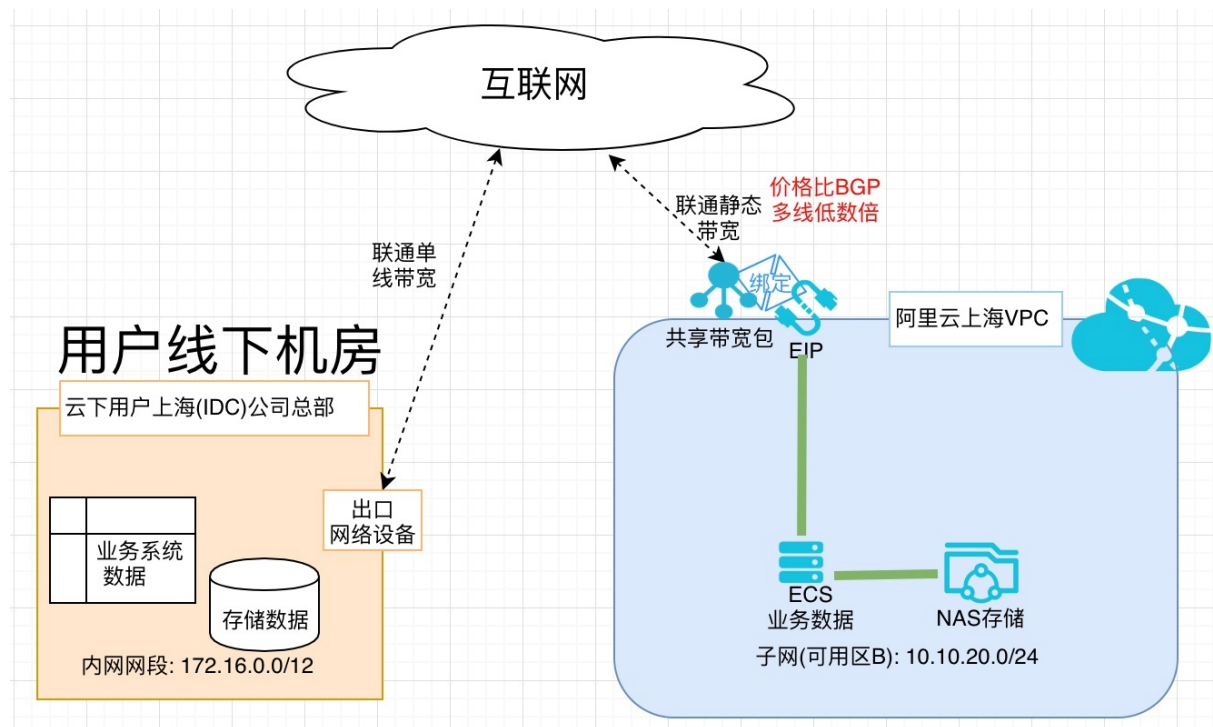
场景描述

随着云计算被越来越多的客户所接受，除业务系统上云外，很多客户已经把业务数据搬迁上云。业务数据量一般都比较大会，迁移上云需要大量的网络带宽，BGP费用比较高。阿里云对用户开放所需地域购买静态单线共享带宽包的权限（移动/联通/电信均可），可用为迁移数据有效降低成本。

解决的问题

业务数据上云网络成本高。

部署架构图



11.低成本跨境文件传输

通过OSS加速域名，配合函数计算和函数工作流，打造低成本、高效的跨境对象存储数据同步传输系统。

直达最佳实践

[点击查看最佳实践详情](#)

更多最佳实践

[点击查看更多阿里云最佳实践](#)

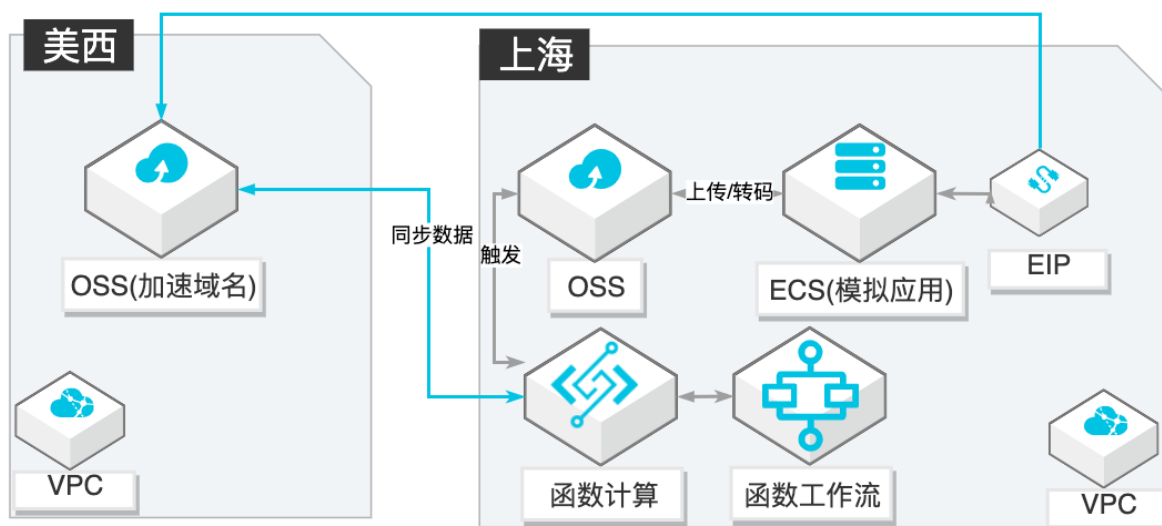
场景描述

- 跨境对象存储数据传输需求。
- 对跨境传输成本控制严苛。
- 能接受一定程度网络抖动带来的文件同步延迟。
- 系统需要有足够的弹性和扩展性应对大规模文件的写入。

解决问题

- 运维低成本：开发人员关注代码逻辑即可。
- 网络成本低：相比CEN高速通道的方式网络成本降低。
- 同步服务部署成本低：文件发生变化触发函数计算任务，按量触发，不用准备ECS。
- 弹性高效：一个文件同步触发一个函数工作流任务，充分利用资源高效同步。

部署架构图



12. 自建Hive数据仓库跨版本迁移到阿里云EMR

本最佳实践介绍如何将客户自建Hadoop/Hive数据仓库迁移到阿里云EMR的技术实现方案和实践步骤。

直达最佳实践

[点击查看最佳实践详情](#)

更多最佳实践

[点击查看更多阿里云最佳实践](#)

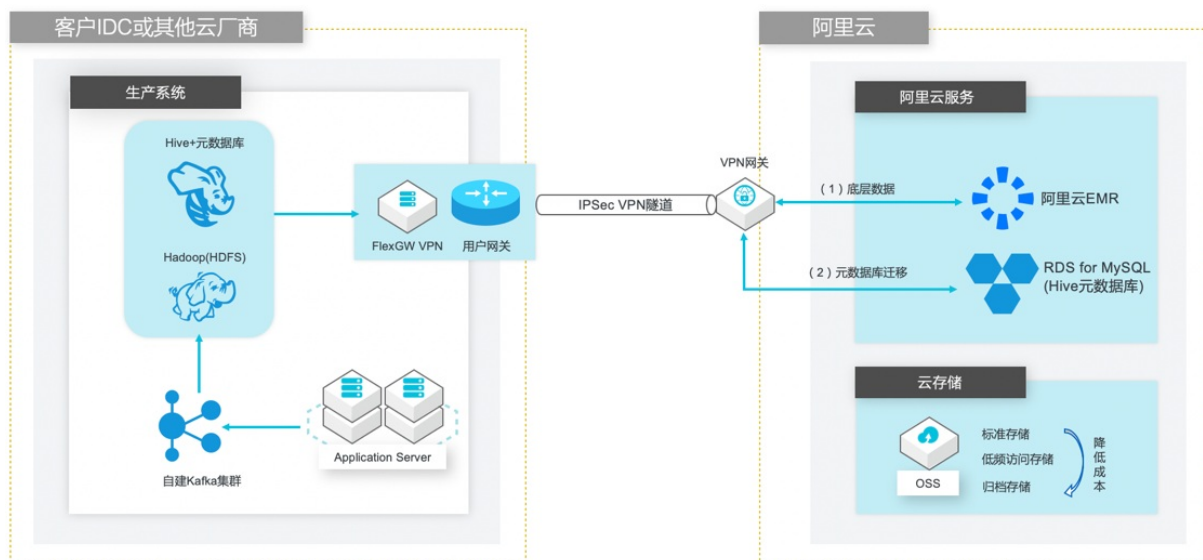
场景描述

客户在IDC或者公有云环境自建Hadoop集群构建数据仓库和分析系统，购买阿里云EMR集群之后，涉及到将数据仓库和Hive元数据的数据库迁移上云。目前主流Hive数据仓库迁移场景为1.x版本迁移到阿里云EMR(Hive 2.x版本)，涉及到数据订正更新步骤。

解决的问题

- Hive数据仓库的数据迁移方案。
- Hive元数据库的迁移方案。
- Hive跨版本迁移后的数据订正。

部署架构图



13.数据迁移上云最佳实践

企业上云过程中涉及的数据（FTP文件、第三方存储文件、NAS）迁移上云的最佳实践，主要含以下几种迁移方案：ossimport、在线迁移服务、rsync等。

直达最佳实践

[点击查看最佳实践详情](#)

更多最佳实践

[点击查看更多阿里云最佳实践](#)

场景描述

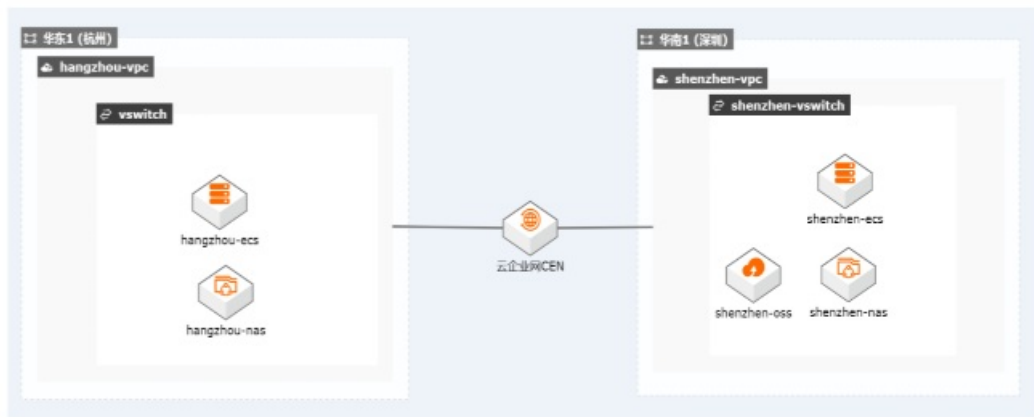
本最佳实践构建以下场景：

- 以杭州地域模拟线下IDC，使用ECS+云盘模拟本地存储数据，杭州地域模拟阿里云。通过ossimport、在线迁移服务等方式将数据从第三方存储迁移到OSS，并确保在迁移时候不影响现有业务。
- 云上用户通过云企业网+NAS在线迁移服务实现跨区域、跨账号迁移NAS上的数据，并确保能实现增量迁移（减少业务停机时长）。

应用场景

- 线下IDC数据迁移至阿里云OSS
- 线下NAS数据迁移至阿里云NAS
- 线上杭州区域NAS数据通过CEN迁移至深圳区域NAS

部署架构图



相关产品

- 专有网络VPC

专有网络VPC（Virtual Private Cloud）是用户基于阿里云创建的自定义私有网络，不同的专有网络之间二层逻辑隔离，用户可以在自己创建的专有网络内创建和管理云产品实例，比如ECS、负载均衡、RDS等。

更多关于专有网络VPC的介绍，参见[专有网络VPC产品详情页](#)。

- 云服务器ECS

云服务器ECS（Elastic Compute Service）是一种简单高效、处理能力可弹性伸缩的计算服务。帮助您构建更稳定、安全的应用，提升运维效率，降低IT成本，使您更专注于核心业务创新。

更多关于云服务器ECS的介绍，参见[云服务器ECS产品详情页](#)。

- **访问控制RAM**

访问控制RAM（Resource Access Management）是阿里云为客户提供的用户身份管理与访问控制服务。使用RAM，您可以创建、管理用户账号（比如员工、系统或应用程序），并可以控制这些用户账号对您名下资源具有的操作权限。当您的企业存在多用户协同操作资源时，RAM可以按需为用户分配最小权限，从而让您避免与其他用户共享云账号密钥，降低您企业的信息安全风险。

更多关于访问控制RAM的介绍，参见[访问控制RAM产品详情页](#)。

- **云企业网CEN**

云企业网CEN（Cloud Enterprise Network）是承载在阿里云提供的高性能、低延迟的私有全球网络上的一张高可用网络。云企业网可帮助您在不同地域VPC间，VPC与本地数据中心间搭建私网通信通道，通过自动路由分发及学习，提高网络的快速收敛和跨网络通信的质量和安全性，实现全网资源的互通，帮助您打造一张具有企业级规模和通信能力的互联网络。

更多关于云企业网CEN的介绍，参见[云企业网CEN产品详情页](#)。

- **文件存储NAS**

阿里云文件存储NAS是一个可共享访问，弹性扩展，高可靠，高性能的分布式文件系统。广泛应用于容器存储、大数据分析、Web服务和内容管理、应用程序开发和测试、媒体和娱乐工作流程、数据库备份。支持冷热数据分级存储。

更多关于文件存储NAS的介绍，参见[文件存储NAS产品详情页](#)。