

阿里云 混合云容灾服务

Hadoop集群容灾

文档版本：20190221

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
	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。	 禁止： 重置操作将丢失用户配置数据。
	该类警示信息可能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，恢复业务所需时间约10分钟。
	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。	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。	单击 确定 。
<code>courier</code> 字体	命令。	执行 <code>cd /d C:/windows</code> 命令，进入Windows系统文件夹。
<code>##</code>	表示参数、变量。	<code>bae log list --instanceid Instance_ID</code>
<code>[]或者[a b]</code>	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
<code>{ }或者{a b}</code>	表示必选项，至多选择一个。	<code>swich {stand slave}</code>

目录

法律声明.....	I
通用约定.....	I
1 概述.....	1

1 概述

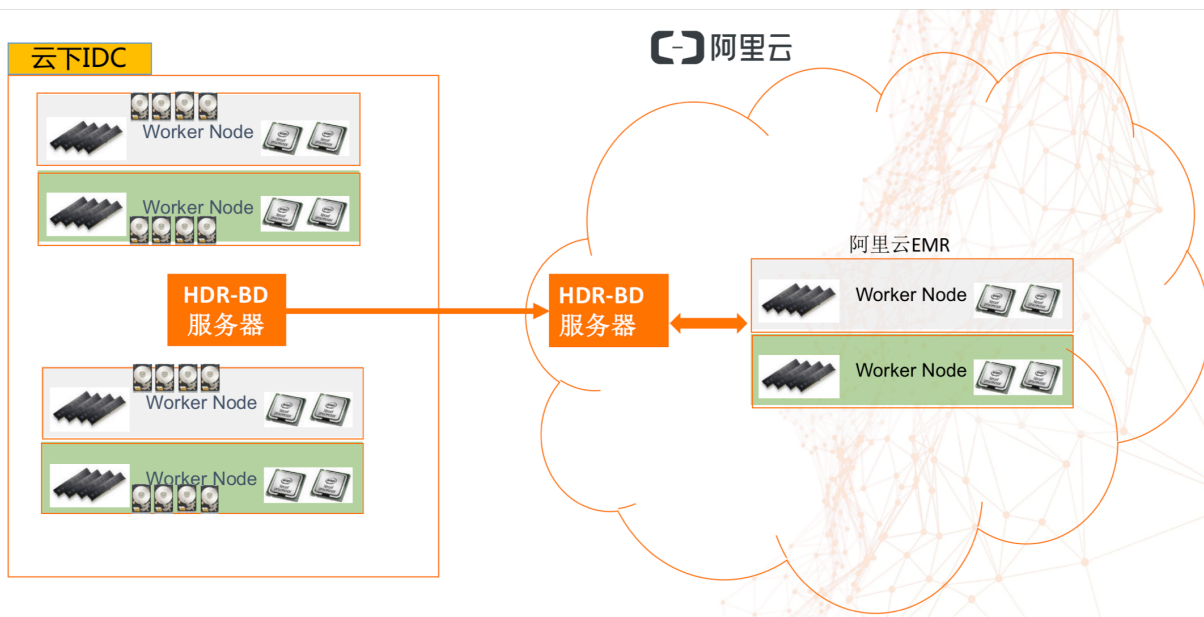
混合云大数据容灾（英文 HDR for Big Data，简称 HDR-BD）是针对 Hadoop 集群的灾备方案。



说明：

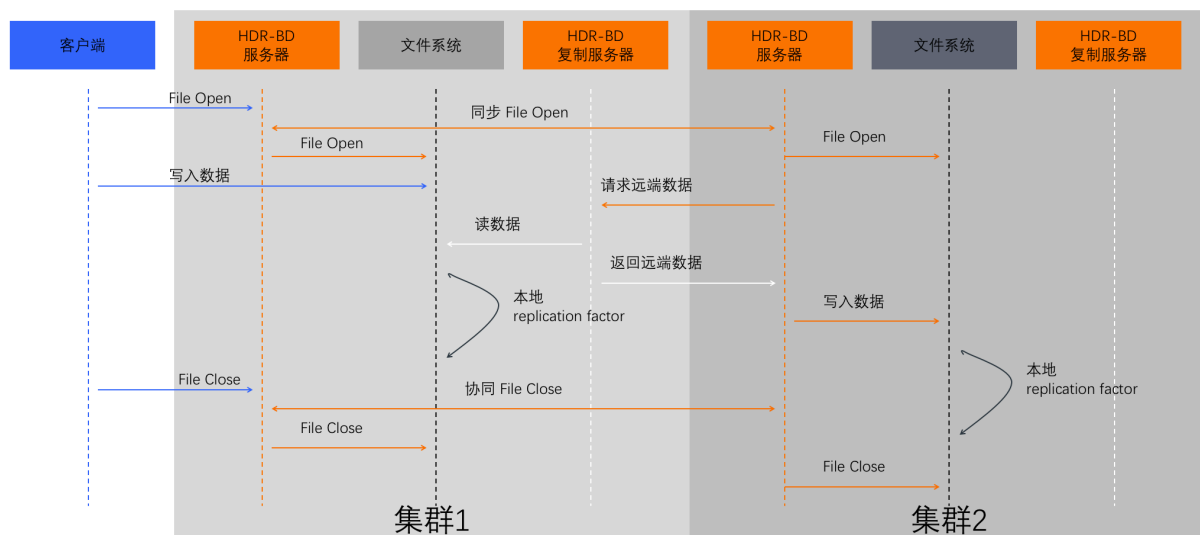
目前 HDR-BD 正在邀测，您可以到混合云容灾服务控制台中注册，我们会评估后尽早联系您。您也可以联系阿里云 CBM 获取邀测资格。HDR-BD 在邀测阶段不收取服务费用，但您可能需要承担测试产生的 EMR、存储、网络等资源费用。

基于领先的数据双向实时复制技术，HDR-BD 可以实现 Hadoop 集群双活和准 0 RPO 容灾。同时，通过与阿里云对象存储 OSS 的紧密集成，您也可以将集群数据实时备份至阿里云 OSS，实现非常简单高效的数据备份。



技术原理

HDR-BD 的核心原理是 Hadoop 集群间的数据实时双向复制。对于 HDFS 上的文件写操作，通过元数据同步复制和数据异步复制，实现了极低的 RPO，系统轻松实现数据保护和集群双活。一个文件写操作的具体流程如下图所示：



主要优势

HDR-BD 为大数据灾备设立了一个全新的标准，其主要优势包括运维复杂度低、近 0 RPO、高资源利用率等。

- 运维复杂度低

HDR-BD 实现了数据的全自动实时双向复制，无需特定时间窗口，也无需担心对业务的影响

- 近 0 RPO

数据的实时复制决定了每一个写操作都会实时复制到配对集群，RPO 几乎为0。

- 高资源利用率

数据双向复制确保运行在不同应用的两个集群共享一份数据，不要求一个集群必须处于只读或备用状态，计算资源得以充分利用。

核心场景

HDR-BD 的核心场景包括跨集群大数据湖、Hadoop 集群双活容灾、Hadoop 数据实时云备份。

- 跨集群大数据湖

在多个 Hadoop 集群之间建立实时双向复制，使数据在多个集群之间无缝流动，而且可以根据不同集群的角色，制订数据复制规则。每个集群在运行不同任务的同时，又能够聚合、分发、共享数据，形成大数据湖，充分发挥数据的效能。

- Hadoop 集群双活容灾

通过在自建 Hadoop 集群和云上 EMR 集群间，或者在云上 EMR 集群间安装配置 HDR-BD 服务，可以确保两个集群间数据实时双向复制，实现近 0 RPO 的大数据集群容灾。两个集群可以各自运行不同的业务，避免资源浪费。所有数据变化都会两个集群间同步，当一个集群发生重大故障时，您可以在另一个集群启动业务，期间几乎没有数据丢失。

- **Hadoop 数据实时云备份**

自建 Hadoop 集群可以通过 HDR-BD 服务与阿里云对象存储 OSS 建立实时单向复制。集群上的数据将会以极短的 RPO 复制到 OSS 上。无需配置备份策略、分配备份时间窗口，Hadoop 集群数据即可轻松得到保护。