

# 阿里云 文件存储HDFS

## 产品简介

文档版本：20200326

## 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. 阿里云文档中所有内容，包括但不限于图片、架构设计、页面布局、文字描述，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。	 <b>禁止：</b> 重置操作将丢失用户配置数据。
	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。	 <b>警告：</b> 重启操作将导致业务中断，恢复业务时间约十分钟。
	用于警示信息、补充说明等，是用户必须了解的内容。	 <b>注意：</b> 权重设置为0，该服务器不会再接受新请求。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 <b>说明：</b> 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。
<b>粗体</b>	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。	执行cd /d C:/window命令，进入Windows系统文件夹。
##	表示参数、变量。	bae log list --instanceid Instance_ID
[ ]或者[a b]	表示可选项，至多选择一个。	ipconfig [-all -t]
{ }或者{a b}	表示必选项，至多选择一个。	switch {active stand}

# 目录

---

法律声明.....	I
通用约定.....	I
1 什么是文件存储HDFS.....	1
2 功能特性.....	2
3 使用限制.....	3
4 应用场景.....	5
5 基本概念.....	6

# 1 什么是文件存储HDFS

---

阿里云文件存储HDFS（Apsara File Storage for HDFS）是面向阿里云ECS实例及容器服务等计算资源的文件存储服务。

文件存储HDFS允许您就像在Hadoop分布式文件系统 (Hadoop Distributed File System) 中管理和访问数据。您无需对现有大数据分析应用做任何修改，即可使用具备无限容量及性能扩展、单一命名空间、多共享、高可靠和高可用等特性的分布式文件系统。

文件存储HDFS适用于互联网行业、金融行业等有大数据计算与存储分析需求的行业客户，进行海量数据存储和离线计算的业务场景，充分满足以Hadoop为代表的分布式计算业务类型对分布式存储性能、容量和可靠性的多方面要求。

创建文件存储HDFS实例后，即可在ECS及容器服务等计算资源内通过标准的HDFS协议接口访问文件系统。此外，多个计算节点可以同时访问同一个文件存储HDFS，共享文件和目录。

## 2 功能特性

---

文件存储HDFS具有无缝集成、共享访问、安全控制、线性扩展等多种特性。

### 无缝集成

- 文件存储HDFS允许您就像在Hadoop分布式文件系统（HDFS）中管理和访问数据。
- 文件存储HDFS SDK可在所有Apache Hadoop 2.x环境中使用，包括阿里云EMR和阿里云Flink。
- 文件存储HDFS支持的原生文件系统语义被开发人员和用户熟知，您无需在迁移到云时学习新的存储范例。
- 从传统HDFS迁移到文件存储HDFS，不会丢失任何文件元信息，包括目录结构、文件创建时间等。

### 共享访问

一个文件存储HDFS实例可以被多个计算节点同时访问，非常适合跨多个ECS或容器服务实例部署的应用程序访问相同数据来源的应用场景。

### 安全控制

文件存储HDFS具有文件系统标准权限控制、权限组访问控制和RAM主子账号授权等多种安全机制，从而保证文件系统数据安全万无一失。

### 线性扩展

文件存储HDFS能够为应用负载提供高吞吐量、高IOPS及低时延的存储性能。同时，其性能与容量成线性关系，可满足业务增长时对更高容量与存储性能的需求。

## 3 使用限制

---

本文档介绍文件存储HDFS及命令行工具存在的限制。

如要了解更多有关开源Hadoop的信息，请参考[Hadoop 文档](#)。

Hadoop FileSystem/AbstractFileSystem

- 不支持目录修改时间（mtime）、访问时间（atime），不支持通过setTimes设置文件的mtime和atime。
- 不支持符号链接（Symbolic Links）。
- 不支持文件截断（truncate）。
- 不支持文件拼接（concat）。
- 不支持扩展属性（XAttrs）相关操作。
- 不支持快照（Snapshot）相关操作。
- 不支持代理令牌（Delegation Token）相关操作。
- 不支持checksum相关操作（setWriteChecksum、setVerifyChecksum）。
- 不支持ACL相关操作。
- 不支持文件块位置信息（FileBlockLocations）。

Hadoop fs 命令行工具

- 不支持快照（snapshot）相关命令（createSnapshot、deleteSnapshot、renameSnapshot）。
- 不支持ACL相关命令（setfacl、getfacl）。
- 不支持XAttr相关命令（setfattr、getfattr）。
- 不支持文件截断相关命令（truncate）。

WebHDFS REST API

不支持

文件存储HDFS文件/目录命名规范

文件存储HDFS中文件/目录的命名规则有如下限制：

- 路径须是以unicode编码的字符或者字符串。

- 文件名和目录名不能为如下字符串：
  - 空字符串
  - 单个点 (.)
  - 两个点 (..)
- 文件名和目录名中不能包含如下字符：
  - 冒号 (:)
  - 斜杠 (/)
  - 波浪号 (~)
- 路径的元素不能包含0-31的ASCII码。



## 4 应用场景

---

文件存储HDFS适用于对吞吐要求较高的大数据分析与机器学习的业务需求场景。

### 大数据分析与机器学习

在大数据分析与机器学习场景中，应用对数据访问的吞吐性能和延迟有较高要求。而文件存储HDFS能够提供高吞吐量和低延迟的访问能力，无需将数据迁移到计算资源本地，因此在该场景下推荐使用文件存储HDFS存储数据。

将数据存入文件存储HDFS，ECS实例或其他计算资源即可直接访问这些数据。将 Hadoop 或其他机器学习应用部署在多个计算资源上，这样应用可以直接通过Hadoopfs接口访问数据进行离线或在线计算，也可以直接将计算结果输出到文件存储HDFS做永久保存。

## 5 基本概念

下表列出了在文档存储HDFS中使用的一些术语。

术语	描述
文件系统实例	一个拥有全局命名空间，可用来存储数据的文件系统。您可以通过挂载点访问文件系统实例中的数据。
挂载点	挂载点是文件系统实例在专有网络或经典网络内的一个访问目标地址。每个挂载点都对应一个域名，您需要修改 <code>core-site.xml</code> 中的配置来访问文件存储HDFS中的文件。
权限组	<p>权限组是若干访问文件系统的权限规则的集合。包括授权IP地址、读写权限、用户权限等。</p> <div> 说明： 每个挂载点都必须与一个权限组绑定。</div>
权限规则	指通过设置授权地址、访问类型以及优先级来约束用户访问文件系统的权限。
授权地址	<p>授权地址是权限组规则的一个属性，代表一条权限组规则被应用的目标。</p> <p>在专有网络内，授权地址允许单个IP或网段。</p>