

阿里云 E-MapReduce

产品内核增强

文档版本：20190610

法律声明

阿里云提醒您在使用或阅读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</code> <code>Instance_ID</code> |
| <code>[]</code> 或者 <code>[a b]</code> | 表示可选项，至多选择一个。 | <code>ipconfig [-all -t]</code> |
| <code>{ }</code> 或者 <code>{a b}</code> | 表示必选项，至多选择一个。 | <code>swich {stand slave}</code> |

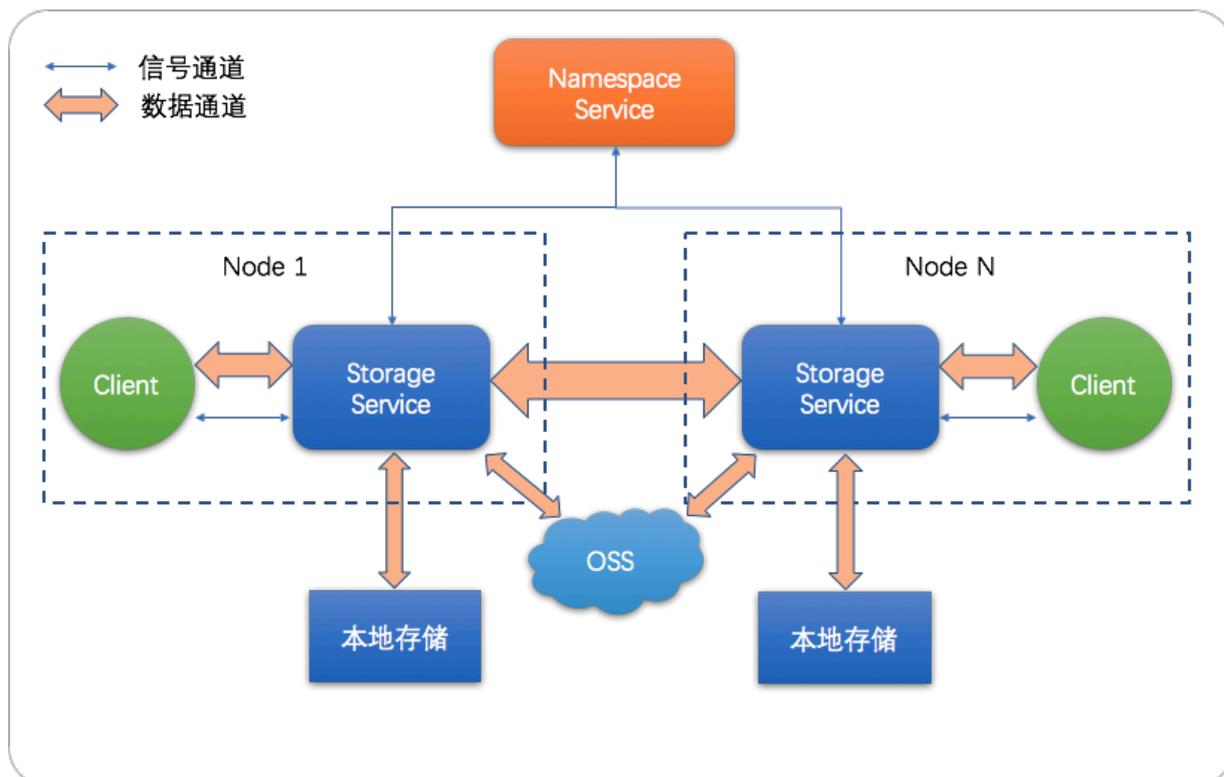
目录

| | |
|-----------------------|---|
| 法律声明..... | I |
| 通用约定..... | I |
| 1 JindoFS 参考使用说明..... | 1 |

1 JindoFS 参考使用说明

JindoFS 是一种云原生的文件系统，结合 OSS 和本地存储，成为 EMR 产品的新一代存储系统，为上层计算提供了高效可靠的存储。本文主要说明 JindoFS 的配置使用方式，以及介绍一些典型的应用场景。

概述



JindoFS 采用了本地存储和 OSS 的异构多备份机制，Storage Service 提供了数据存储能力，首先使用 OSS 作为存储后端，保证数据的高可靠性，同时利用本地存储实现冗余备份，利用本地的备份，可以加速数据读取；另外，JindoFS 的元数据通过本地服务 Namespace Service 管理，从而保证了元数据操作的性能（和 HDFS 元数据操作性能相当）。EMR-3.20.0 及以上版本支持 JindoFS，您可以在创建集群时勾选相关服务来使用 JindoFS。

环境准备

· 创建集群

选择 EMR-3.20.0 及以上版本，勾选可选服务中的 SmartData 和 Bigboot，具体操作步骤请参见[创建集群](#)。Bigboot 服务提供了 EMR 平台上的基础的分布式数据管理交互服务以及一些组

件管理监控和支持性服务，SmartData 服务基于 Bigboot 之上对应用层提供了 JindoFS 文件
系统。



· 配置集群

SmartData 提供的 JindoFS 文件系统使用 OSS 作为存储后端，因此在使用 JindoFS 之前需配置一些 OSS 相关参数。下面提供两种配置方式，第一种是先创建好集群，修改 Bigboot 相关参数，需重启 SmartData 服务生效；第二种是创建集群过程中添加自定义配置，这样集群创建好后相关服务就能按照自定义参数启动：

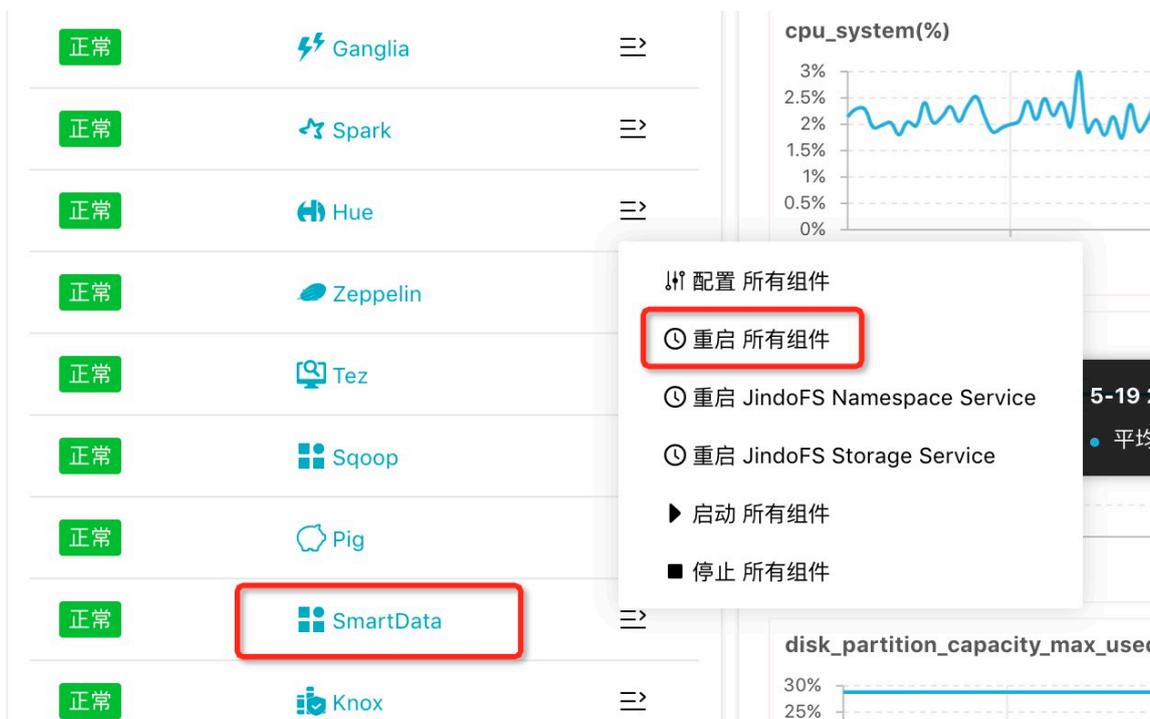
- 集群创建好后参数初始化

所有 JindoFS 相关配置都在 Bigboot 组件中，配置如下图所示，红框中为必填的配置项，oss.access.bucket 为 OSS bucket 名字，oss.data-dir 为 JindoFS 在 OSS bucket 中所使用的目录（注：该目录为 JindoFS 后端存储目录，生成的数据不能人为破坏，并且保证该目录仅用于 JindoFS 后端存储，JindoFS 在写入数据时会自动创建用户所配置的目录，无需在 OSS 上事先创建）。oss.access.endpoint、oss.access.key、oss.access.secret 含义可参考页面注

解，一般考虑到性能和稳定性，推荐使用同 region 下的 OSS bucket 作为存储后端，这种情况下，EMR 集群能够免密访问 OSS，无需配置这几项。



配置完成后保存并部署，然后在 SmartData 服务中重启所有组件，即开始使用 JindoFS。



- 创建集群时添加自定义配置

EMR 集群在创建集群时支持添加自定义配置，以同 region 下免密访问 OSS 为例，如下图勾选软件自定义配置，添加如下配置，配置 oss.data-dir 和 oss.access.bucket。

```
[
  {
    "ServiceName": "BIGBOOT",
    "FileName": "bigboot",
    "ConfigKey": "oss.data-dir",
    "ConfigValue": "jindoFS-1"
  },
  {
    "ServiceName": "BIGBOOT",
    "FileName": "bigboot",
```

```

"ConfigKey": "oss.access.bucket",
"ConfigValue": "oss-bucket-name"
}
]

```

软件配置

集群类型: **Hadoop** | Druid | Kafka | ZooKeeper | Data science

开源大数据离线、实时、Ad-hoc查询场景

E-MapReduce Hadoop是完全使用开源Hadoop生态，采用YARN管理集群资源，提供Hive、Spark高线大规模分布式数据存储和计算，SparkStreaming、Flink、Storm流式数据计算，Presto、Impala交互式查询，Oozie、Pig等Hadoop生态圈的组件，支持OSS存储，支持Kerberos的数据认证与加密。

产品版本: EMR-3.20.0

资源管理类型: **半托管** | 全托管

必选服务: HDFS (2.8.5) | YARN (2.8.5) | Hive (3.1.1) | Spark (2.4.2) | Knox (1.1.0) | Zeppelin (0.8.1) | Tez (0.9.1) | ApacheDS (2.0.0) | Ganglia (3.7.2) | Pig (0.14.0) | Sqoop (1.4.7) | Hue (4.1.0)

可选服务: HBase (1.4.9) | ZooKeeper (3.4.13) | Presto (0.213) | Impala (2.12.2) | Flume (1.8.0) | Livy (0.6.0) | Superset (0.28.1) | Ranger (1.2.0) | Flink (1.7.2) | Storm (1.2.2) | Phoenix (4.14.1) | **SmartData (1.0.0)** | **Bigboot (1.0.0)** | Oozie (5.1.0)

请点击选择

高级设置

Kerberos集群模式: 高安全集群中的各组件会通过Kerberos进行认证，详细信息参考[Kerberos简介](#)

软件自定义配置: 新建集群创建前，可以通过json文件定义集群组件的参数配置，详细信息参考 [软件配置](#)

```

[{"ServiceName": "BIGBOOT", "FileName": "bigboot", "ConfigKey": "oss.data-dir", "ConfigValue": "jindoFS"},
{"ServiceName": "BIGBOOT", "FileName": "bigboot", "ConfigKey": "oss.access.bucket", "ConfigValue": "oss-bucket-name"}]

```

使用 JindoFS

JindoFS 使用上与 HDFS 类似，提供 jfs 前缀，将 jfs 替代 hdfs 即可使用。简单示例：

```

hadoop fs -ls jfs:/// hadoop fs -mkdir jfs:///test-dir
hadoop fs -put test.log jfs:///test-dir/

```

目前，JindoFS 能够支持 EMR 集群上的 Hadoop、Hive、Spark 的作业进行访问，其余组件尚未完全支持。

应用场景

EMR 目前提供了三种大数据存储系统，EMR OssFileSystem、EMR HDFS 和 EMR JindoFS，OssFileSystem 和 JindoFS 都是云上存储的解决方案，下面比较了这三种存储系统各自的特点，同时也将开源 OSS 放入比较。

| 特点 | 开源 OSS | EMR OssFileSystem | EMR HDFS | EMR JindoFS |
|-------|--------|-------------------|----------------|-------------|
| 存储空间 | 海量 | 海量 | 取决于集群规模 | 海量 |
| 可靠性 | 高 | 高 | 高 | 高 |
| 吞吐率因素 | 服务端 | 集群内磁盘缓存 | 集群内磁盘 | 集群内磁盘 |
| 元数据效率 | 慢 | 中 | 快 | 快 |
| 扩容操作 | 容易 | 容易 | 容易 | 容易 |
| 缩容操作 | 容易 | 容易 | 需 Decommission | 容易 |
| 数据本地化 | 无 | 弱 | 强 | 较强 |

JindoFS 具有以下几个特点：

- 海量弹性的存储空间，基于 OSS 作为存储后端，存储不受限于本地集群，而且本地集群能够自由弹性伸缩。
- 能够利用本地集群的存储资源加速数据读取，适合具有一定本地存储能力的集群，能够利用有限的本地存储提升吞吐率，特别对于一写多读的场景效果显著。
- 元数据操作效率高，能够与 HDFS 相当，能够有效规避 OSS 文件系统元数据操作耗时以及高频访问下可能引发不稳定的痛点。
- 能够最大限度保证执行作业时的数据本地化，减少网络传输的压力，进一步提升读取性能。

磁盘空间水位控制

JindoFS 后端基于 OSS，可以提供海量的存储，但是本地盘的容量是有限的，因此 JindoFS 会自动淘汰本地较冷的数据备份。我们提供了 `node.data-dirs.watermark.high.ratio` 和 `node.data-dirs.watermark.low.ratio` 这两个参数用来调节本地存储的使用容量，值均为 0~1 的小数表示使用比例，JindoFS 默认使用所有数据盘，每块盘的使用容量默即为数据盘大小。前者表示使用量上水位比例，每块数据盘的 JindoFS 占用的空间到达上水位即会开始清理淘汰；后者表示使用量下水位比例，触发清理后会将 JindoFS 的占用空间清理到下水位。用户可以通过设置上水位比例调节期望分给 JindoFS 的磁盘空间，下水位必须小于上水位，设置合理的值即可。