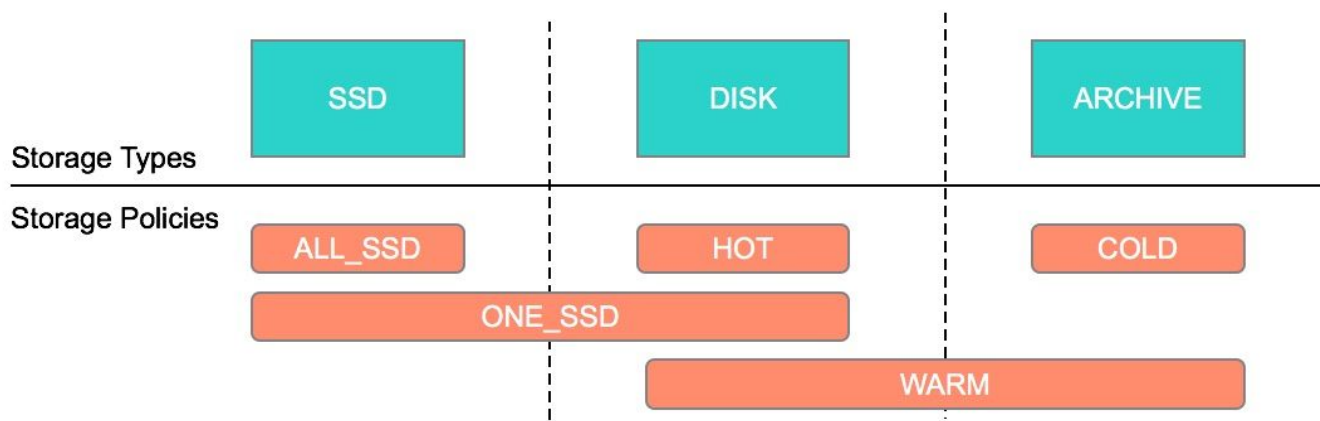


HDFS 归档存储编程指南

介绍

HDFS 归档存储 (Archival Storage) 是从 Hadoop 2.6.0 开始引入的 (参见 [HDFS-6584](#))。归档存储是一种将增长的存储容量与计算容量解耦的解决方案。我们可以在集群中部署一些具有更高密度、更便宜的存储且提供更低计算能力的节点，并且可以用作集群中的冷数据存储。根据我们的设置，可以将热数据移到冷存储介质中。通过添加更多的冷存储节点，可以为集群增加存储能力。

异构存储 (Heterogeneous Storage) 和归档存储 (Archival Storage) 提供的框架使得 HDFS 体系结构变得更加通用，使得 Hadoop 集群支持包括 SSD 和内存等存储介质。为了获得更好的性能，用户可以选择将数据存储在 SSD 或内存中。



如果想及时了解Spark、Hadoop或者HBase相关的文章，欢迎关注微信公众号：iteblog_hadoop

存储类型和存储策略

存储类型

异构存储的第一阶段([HDFS-2832](#))将 datanode 存储模型从单个存储(可能对应多个物理存储介质)更改为存储的集合，每个存储对应一个物理存储介质。它还添加了存储类型(磁盘和 SSD)的概念，其中磁盘是默认的存储类型。

在归档存储中，增加了一种新的存储类型：ARCHIVE，它具有高存储密度(pb级的存储)，但计算能力很小。

另外添加了一种新的存储类型：RAM_DISK，用于支持在内存中写入单个副本文件。

所以，到目前为止，Hadoop 支持四种存储类型：ARCHIVE, DISK, SSD 和 RAM_DISK。

存储策略

引入了存储策略的新概念，以便根据存储策略将文件存储在不同的存储类型中。

截止到 Hadoop 3.2.1，支持以下几种存储策略：

- Hot - 用于存储和计算，那些经常访问和存储的数据一般使用这种策略。当一个块是 hot 存储策略，则其他副本的存储策略为 DISK；
- Cold - 仅用于存储且提供有限的计算能力。那些不再使用的数据或者需要归档的数据会从 hot 存储介质移到 cold 存储介质。当一个块属于 cold 存储介质，则其他副本的存储策略为 ARCHIVE。
- Warm - 部分数据属于热数据部分属于冷数据。当一个块属于 Warm 存储策略，其一些副本存储在 DISK 中，剩下的存储在 ARCHIVE 介质中；
- All_SSD - 把所有副本存储在 SSD 介质中
- One_SSD - 用于将一个副本存储在 SSD 介质中，剩下的副本存储在 DISK 介质中；
- Lazy_Persist - 用于在内存中写入具有单个副本的块，需要持久化时，副本会先写到 RAM_DISK 中，最后延迟保存到 DISK 中；
- Provided - 用于将数据存储在 HDFS 之外。

存储策略由以下字段组成：

1. Policy ID
2. Policy name
3. 块放置的存储类型列表
4. 用于文件创建的后备存储类型列表
5. 用于存储副本的后备存储类型列表n

当有足够的空间时，将根据上面第三条指定的存储类型列表存储块副本。当上面第三条中的某些存储类型空间不足时，在上面第四条和第五条中指定的后备存储类型列表将分别用于替换空间不足的存储类型，以用于文件创建和复制

以下是典型的存储策略表

Policy ID	Policy Name	Block Placement (n replicas)	Fallback storages for creation	Fallback storages for replication
15	Lazy_Persist	RAM_DISK: 1, DISK: n-1	DISK	DISK
12	All_SSD	SSD: n	DISK	DISK
10	One_SSD	SSD: 1, DISK: n-1	SSD, DISK	SSD, DISK
7	Hot (default)	DISK: n	<none>	ARCHIVE
5	Warm	DISK: 1, ARCHIVE: n-1	ARCHIVE, DISK	ARCHIVE, DISK
2	Cold	ARCHIVE: n	<none>	<none>
1	Provided	PROVIDED: 1, DISK: n-1	PROVIDED, DISK	PROVIDED, DISK

注意一：Lazy_Persist 策略仅对单个复制块有用。对于具有多个副本的块，所有副本都将被写到

DISK，因为只将一个副本写到 RAM_DISK 并不能提高总体性能。

注意二：对于带条带布局的纠删码文件，适用的存储策略是 All_SSD、Hot、Cold。因此，如果用户为带条带的纠删码文件配置除上述策略之外的存储策略，则在创建或移动块时不会遵循该策略。

存储策略解析

当创建文件或目录时，其存储策略是未知的。我们可以使用 `storagepolicies -setStoragePolicy` 命令指定存储策略。文件或目录的有效存储策略由以下规则解析。

1. 如果文件或者目录指定了存储策略，那直接返回这个策略；
2. 对于没有指定存储策略的文件或者目录，如果其父根目录，则返回默认存储策略；否则返回其父目录的存储策略。

我们可以通过 `storagepolicies -getStoragePolicy` 命令获取有效的存储策略。

配置

- `dfs.storage.policy.enabled` - 启用或者关闭存储策略这个功能，默认值是 `true`；
- `dfs.datanode.data.dir` - 每个 `DataNode` 上配置，以逗号分隔的存储位置，并在每个路径都需要加上存储类型标记。这样可以存储策略选择器根据标记的存储类型来存储相应的块。比如
 1. `datanode` 的存储位置 `/grid/dn/disk0` 在 DISK 上，那么我们需要这么配置
`[DISK]file:///grid/dn/disk0`
 2. `datanode` 的存储位置 `/grid/dn/ssd0` 在 SSD 上，那么我们需要这么配置
`[SSD]file:///grid/dn/ssd0`
 3. `datanode` 的存储位置 `/grid/dn/archive0` 在 ARCHIVE，那么我们需要这么配置
`[ARCHIVE]file:///grid/dn/archive0`
 4. `datanode` 的存储位置 `/grid/dn/ram0` 在 RAM_DISK，那么我们需要这么配置
`[RAM_DISK]file:///grid/dn/ram0`

`DataNode` 的存储目录默认的存储类型为 DISK，所以如果我们没有显示指定存储类型，那就是 DISK 存储类型。

存储策略命令

展示存储策略

可以使用下面命令显示所有的存储策略：

```
hdfs storagepolicies -listPolicies
```

设置存储策略

我们可以使用下面命令给文件或者目录设置存储策略：

```
hdfs storagepolicies -setStoragePolicy -path <path> -policy <policy>
```

- -path ：需要设置存储策略的目录或文件路径；
- -policy ：存储策略的名称。

删除存储策略

可以使用下面命令删除文件或者目录的存储策略。当文件或者目录的存储策略被删除，那么其存储策略使用最近父目录的存储策略，如果其所有父目录都没有设置存储策略，那么将会使用默认的存储策略：

```
hdfs storagepolicies -unsetStoragePolicy -path <path>
```

获取存储策略

可以使用下面命令获取文件或者目录的存储策略。

```
hdfs storagepolicies -getStoragePolicy -path <path>
```

应用存储策略

使用下面命令把某个块移动到相应的存储策略存储设备上：

```
hdfs storagepolicies -satisfyStoragePolicy -path <path>
```

更多关于 HDFS 的归档存储可以参见 [Archival Storage, SSD & Memory](#)。

本博客文章除特别声明，全部都是原创！
原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。
本文链接: **【】**（**）**