

第2章 HDFS文件系统

2.1 【实验】HDFS的基本操作

本节实验操作在“Hadoop_模板2”中进行，需要为学生分配1台虚拟机，虚拟机的配置为4096MB内存，500MHZ CPU。

2.1.1 实验目的

Hadoop配置好后，可以通过命令行工具快速的进行HDFS文件系统的访问。本节实验主要进行HDFS的一些基本文件操作，例如读文件、创建文件存储路径、转移文件、删除文件、列出文件列表、并行复制、文件归档等操作。

2.1.2 实验环境

操作系统：CentOS6.5操作系统

计算机资源：CPU 1核 0.5GHz 内存 4GB 硬盘 10.00GB

实验环境：易优云大数据实验实训平台 Hadoop_模板2

2.1.3 实验类型及实验课时

实验类型：验证性实验

实验课时：2课时

2.1.4 实验原理

Hadoop配置完毕后，可以通过命令行接口来和HDFS进行交互。当然，命令行接口只是HDFS的访问接口之一，它的特点是更加简单直观，便于使用，可以进行一些基本操作。同时我们也可以直接通过<http://NameNode IP地址:50070>访问HDFS的Web管理界面，HDFS的Web管理界面提供了基本的文件系统信息，其中包括集群启动时间、版本号、编译时间及是否可升级等信息。

HDFS的Web界面还提供了文件系统的基本功能，例如可以通过HDFS的Web管理界面将HDFS的文件结构通过目录的形式展现出来，增加了对文件系统的可读性；可以直接通过Web界面访问文件内容；HDFS的Web管理界面还可以将该文件所在的节点位置展现出来。此外HDFS的Web管理界面还提供了NameNode的日志列表、运行中的节点列表以及宕机的节点列表等信息。

在对HDFS进行命令行操作时会存在hadoop fs命令和hadoop dfs命令两种模式，fs是一个比较抽象的层面，在分布环境中，fs就是dfs，但在本地部署环境中，例如单机环境中，fs就是local file system，这个时候dfs就不能用。在本实验中，我们使用伪分布式环境进行HDFS的基本操作，因此可以直接使用hadoop fs命令来代替hadoop dfs。

在HDFS中每个文件采用块方式进行存储，在系统运行时，文件块的元数据信息会被存在NameNode的内存中，因此对HDFS来说，大规模存储小文件显然是低效的，很多小文件会耗尽NameNode的大部分内存。Hadoop归档文件和HAR文件可以将文件高效地放入HDFS块中的文件存档设备，在减少NameNode内存使用的同时，仍然允许对文件进行透明访问。Hadoop归档文件是通过archive命令工具根据文件集合创建的，因为这个工具需要运行一个MapReduce来并行处理输入文件，所以需要有一个运行MapReduce的集群。Hadoop归档文件可以作为Mapreduce的输入，这里需要注意的是，小文件并不会占用太多的磁盘空间，

例如设定一个128M的文件块来存储1MB的文件，实际上存储这个文件只需要1MB磁盘空间，而不是128MB。

HDFS提供了一个非常实用的复制命令`distcp`。常规的Java API等多种接口对HDFS数据访问模型往往集中于单线程的数据存取，如果希望高效的对一个文件集合进行操作，就需要编写一个程序进行并行操作，用来在Hadoop文件系统中并行地复制大量文件。通过`distcp`命令可以快速的在两个HDFS集群间并行的进行数据传送。如果两个集群都运行在同一个Hadoop版本上，那么可以使用HDFS模式的`distcp`。`distcp`操作有很多选项可以设置，比如忽略失败、限制文件或者复制的数据量等。直接输入指令或者不附加选项可以查看此操作的使用说明。具体实现时，`distcp`操作会被解析为一个MapReduce操作来执行，当没有Reducer操作时，复制被作为Map操作并行地在集群节点中运行，因此每个文件都可以被当成一个Map操作来执行复制。`distcp`会通过执行多个文件聚集捆绑操作，尽可能地保证每个Map操作执行相同数量的数据。由于系统需要保证每个Map操作执行的数据量是合理的，来最大化地减少Map执行的开销，而按规定，每个Map最少要执行256MB的数据量（此值不是固定的，复制的全部数据量小于256MB除外）。例如要复制1GB的数据，那么系统会分配4个Map任务，当数据量非常大时，就需要限制执行的Map任务数，以限制网络带宽和集群的使用率。默认情况下，每个集群的一个节点最多执行20个Map任务。例如要复制1000GB数据到100节点的集群中，那么系统就会分配2000个Map任务（每个节点20个），也就是说每个节点会平均复制512MB，可以通过调整`distcp`的`-m`参数减少Map任务量，比如`-m 1000`就意味着分配1000个Map任务，此时每个节点将分配1GB数据量。

2.1.5 【学生版】实验步骤

2.1.5.1 启动Hadoop集群

在模板中，我们已经配置好了**Hadoop**伪分布式环境，同学们不需要再次配置，可以直接启动使用。

步骤1. 启动Hadoop

打开一个终端模拟器，通过命令启动Hadoop。

步骤2. 验证Hadoop是否启动成功

通过命令，查看相应的JVM进程确定Hadoop是否启动成功。

2.1.5.2 HDFS的命令行操作

步骤1. 查看HDFS命令参数

使用命令查看HDFS文件系统操作的一些基本参数。

步骤2. 列出HDFS文件

使用命令列出了HDFS文件系统根目录（/）下的所有文件。

步骤3. 列出HDFS子目录下的文件

使用命令浏览HDFS下/user子目录的文件。

步骤4. 在HDFS中创建子目录

1. 使用命令在HDFS根目录下（/）创建一个名为test的子目录。
2. 使用命令列出HDFS文件系统根目录（/）下的所有文件。

步骤5. 上传文件到HDFS

通过`hadoop fs -put`命令可以向HDFS中上传文件，其中需要指定两个参数，第一个参数代表需要上传到HDFS中的本地文件的绝对路径，第二个参数代表HDFS中对应的指定目录。

1. 使用命令创建需要上传的数据文件。
2. 使用命令查看数据文件是否创建成功。
3. 使用命令实现将本地的/home/input.txt文件上传到HDFS文件系统的/test子目录下。
4. 使用命令浏览HDFS下/test子目录的文件。

步骤6. 将HDFS中文件复制到本地系统中

通过`hadoop fs -get`命令可以获取HDFS中的文件，其中需要指定两个参数，第一个参数代表需要获取的文件在HDFS中的绝对路径，第二个参数代表将文件存储到本地的指定接收目录。

1. 使用命令实现将HDFS中的/test/input.txt文件复制到本地文件系统的/usr目录下。
2. 使用命令查看文件是否复制成功。

步骤7. 查看HDFS下某个文件

使用命令查看HDFS文件系统中/test/input.txt文件中的内容。

步骤8. 删除HDFS下的文档

1. 使用命令删除HDFS中的/test子目录。
2. 使用命令列出HDFS文件系统根目录 (/) 下的所有文件。

2.1.5.3 使用Hadoop归档文件

在HDFS中每个文件采用块方式进行存储，在系统运行时，文件块的元数据信息会被存放在NameNode的内存中，因此对HDFS来说，大规模存储小文件显然是低效的，很多小文件会耗尽NameNode的大部分内存。Hadoop归档文件可以将文件高效地放入HDFS块中的存储，在减少NameNode内存使用的同时，仍然允许对文件进行透明访问。

通过`hadoop archive`命令可以创建归档文件，我们通常需要传入3个参数：通过`-archiveName`指定第一个参数设置归档文件的名称；通过`-p`指定第二个参数设置要归档的文件源；最后一个参数即指定归档文件的输出目录。

步骤1. 创建工作目录

1. 使用命令在HDFS中创建/archive工作目录。
2. 使用命令在/archive目录下创建子目录file1。
3. 使用命令在/archive目录下创建子目录file2。

步骤2. 上传数据文件

1. 使用命令创建需要上传的数据文件test_file1.txt。
2. 使用命令创建需要上传的数据文件test_file2.txt。
3. 使用命令将本地的/home/test_file1.txt文件上传到HDFS文件系统的/archive/file1子目录下。
4. 使用命令测试文件是否上传成功。
5. 使用命令将本地的/home/test_file2.txt文件上传到HDFS文件系统的/archive/file2子目录下。

6. 使用命令测试文件是否上传成功。

步骤3. 创建归档文件

使用命令将HDFS文件系统/archive目录下的所有文件进行归档，命名为files.har并存储到HDFS文件系统的根目录（/）下。

步骤4. 归档文件测试

1. 使用命令查看HDFS根目录确定归档文件是否成功生成。
2. 使用命令进行归档文件子内容的查看。

2.1.5.4 通过distcp进行并行复制

HDFS提供了一个非常实用的复制命令distcp，通过distcp命令可以快速的在两个HDFS集群间并行的进行数据传送。具体实现时，distcp操作会被解析为一个MapReduce操作来执行，当没有Reducer操作时，复制被作为Map操作并行地在集群节点中运行，因此每个文件都可以被当成一个Map操作来执行复制。

使用distcp命令通常需要传入2个参数：第一个参数设置输入源文件的完整访问路径hdfs://输入源集群NameNode主机名:端口号/输入源目录；第二个参数设置目标位置的完整路径hdfs://目标集群NameNode主机名:端口号/输出目录。

其中端口号需要与core-site.xml配置文件中设置的端口号相一致，如下所示：

```
<property>
<name>fs.defaultFS</name>
<value>hdfs://master:9000</value>
</property>
```

为了操作方便，我们在同一个集群中演示distcp命令的使用，因此我们的输入源为hdfs://master:9000/，对应的输出源也为hdfs://master:9000/。

步骤1. 创建目标路径

1. 使用命令在HDFS中创建目标路径。
2. 使用命令查看HDFS根目录确定目录是否成功生成。

步骤2. 文件复制

使用命令将HDFS中/archive下的文件复制到/distcpdir目录下。

步骤3. 复制结果查看

使用命令进行目录内容的查看。

步骤4. 使用HFTP进行数据复制

当使用distcp进行HDFS集群间的复制时，任务需要在目标集群中执行，所以HDFS的RPC服务版本需要匹配，当HDFS运行在不同的Hadoop版本之上，复制将会因为RPC版本的不匹配而失败，此时可以使用基于HTTP的HFTP进行数据操作。需要注意的是，输入源的URI中要定义NameNode的网络接口，这个接口会通过dfs.hftp.address的属性值设定，默认值为50070。

1. 使用命令实现将HDFS中/archive下的文件以HFTP模式复制到/distcpdir1目录下。
2. 使用命令进行目录内容的查看。

2.1.6 常见问题

2.1.6.1 创建目录失败

当在HDFS中创建子目录时，首先需要保证父目录已经被创建，然后才能创建子目录，否则是无法完成子目录创建的。常见的报错信息如下所示：

```
[root@master ~]# hadoop fs -mkdir /parent/child
18/11/09 10:25:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
mkdir: `/parent/child': No such file or directory
[root@master ~]#
```

2.1.6.2 删除目录失败

当使用hadoop fs -rm命令删除HDFS中的目录时，如果被删除的目录中有相关的子目录文件，则会删除失败，提示信息如下所示：

```
[root@master ~]# hadoop fs -rm /archive
18/11/09 10:28:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
rm: `/archive': Is a directory
```

此时可以将被删除的目录中相关的子目录文件先删除，然后再删除目录。当然也可以直接使用hadoop fs -rmr命令以递归的方式删除目录。

2.1.6.3 使用distcp进行并行复制失败

当使用distcp命令进行数据并行复制的时候，可能会出现“拒绝连接”的相关错误，如下所示：

```
[root@master ~]# hadoop distcp hdfs://master:8020/archive hdfs://master:8020/distcpdir
18/11/09 10:34:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/11/09 10:34:40 ERROR tools.DistCp: Invalid arguments:
java.net.ConnectException: Call From master/127.0.0.1 to master:8020 failed on connection exception: java.net.ConnectException: 拒绝连接; For more details see: http://wiki.apache.org/hadoop/ConnectionRefused
    at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native Method)
    at sun.reflect.NativeConstructorAccessorImpl.newInstance(NativeConstructorAccessorImpl.java:62)
    at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
    (-----省略-----)
```

此时可能存在的问题是所访问的HDFS文件系统的端口号与core-site.xml中配置的端口号不一致所致：

```
<property>
<name>fs.defaultFS</name>
<value>hdfs://master:9000</value>
</property>
```

当然也可能存在一种情况，即distcp所连接的两个HDFS文件系统对应的版本号不一致，因此导致复制失败，此时应该使用HFTP模式进行数据复制。

2.1.6.4 HFTP模式数据复制失败

当使用distcp命令以HFTP模式进行HDFS集群间的复制时，可能会出现下列错误信息：

```
[root@master ~]# hadoop distcp hftp://master:50070/archive hftp://master:50070/distcpdir
1
```

```

18/11/09 10:40:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/11/09 10:40:38 INFO tools.DistCp: Input Options: DistCpOptions{atomicCommit=false, syncFolder=false, deleteMissing=false, ignoreFailures=false, maxMaps=20, sslConfigurationFile='null', copyStrategy='uniformsize', sourceFileListing=null, sourcePaths=[hftp://master:50070/archive], targetPath=hftp://master:50070/distcpdir1, targetPathExists=true, preserveRawXattrs=false}
18/11/09 10:40:38 INFO client.RMPProxy: Connecting to ResourceManager at master/127.0.0.1:8032
18/11/09 10:40:39 INFO Configuration.deprecation: io.sort.mb is deprecated. Instead, use mapreduce.task.io.sort.mb
18/11/09 10:40:39 INFO Configuration.deprecation: io.sort.factor is deprecated. Instead, use mapreduce.task.io.sort.factor
18/11/09 10:40:40 INFO client.RMPProxy: Connecting to ResourceManager at master/127.0.0.1:8032
18/11/09 10:40:40 INFO mapreduce.JobSubmitter: number of splits:4
18/11/09 10:40:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1541659355931_0011
18/11/09 10:40:40 INFO impl.YarnClientImpl: Submitted application application_1541659355931_0011
18/11/09 10:40:41 INFO mapreduce.Job: The url to track the job: http://master:8088/proxy/application_1541659355931_0011/
18/11/09 10:40:41 INFO tools.DistCp: DistCp job-id: job_1541659355931_0011
18/11/09 10:40:41 INFO mapreduce.Job: Running job: job_1541659355931_0011
18/11/09 10:40:50 INFO mapreduce.Job: Job job_1541659355931_0011 running in uber mode : false
18/11/09 10:40:50 INFO mapreduce.Job: map 0% reduce 0%
18/11/09 10:41:14 INFO mapreduce.Job: map 36% reduce 0%
18/11/09 10:41:15 INFO mapreduce.Job: Task Id : attempt_1541659355931_0011_m_000000_0, Status : FAILED
Error: java.io.IOException: mkdir failed for hftp://master:50070/distcpdir1/archive
    at org.apache.hadoop.tools.mapred.CopyMapper.createTargetDirsWithRetry(CopyMapper.java:298)
    at org.apache.hadoop.tools.mapred.CopyMapper.map(CopyMapper.java:242)
    at org.apache.hadoop.tools.mapred.CopyMapper.map(CopyMapper.java:50)
    at org.apache.hadoop.mapreduce.Mapper.run(Mapper.java:146)
    at org.apache.hadoop.mapred.MapTask.runNewMapper(MapTask.java:787)
    at org.apache.hadoop.mapred.MapTask.run(MapTask.java:341)
    at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:164)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    (-----省略-----)

```

由于使用distcp命令进行HDFS集群间的复制时，任务需要在目标集群中执行，因此在目标位置依然需要使用hdfs://master:9000/distcpdir1形式，而不是HFTP形式hftp://master:50070/distcpdir1。

2.1.7 课后思考

1. 请详细说明为什么HDFS中不适合存储小文件。
2. 请总结使用归档文件的优势所在。