

第3章Hadoop基础操作

(源自:<https://biglab.site>)

(版本:Ver2.1-20230926)

第3章Hadoop基础操作

任务前言

学习目标

任务背景

任务3.1认识Hadoop安全模式

3.1.1了解Hadoop安全模式

3.1.2查看、解除和开启Hadoop安装模式

任务3.2查看Hadoop集群基本信息

Hadoop集群的核心功能

操作终端机hosts修改

查询集群的存储系统信息HDFS

Web方式

命令行方式

查询 集群的计算资源信息YARN

Web方式

任务3.3上传文件到HDFS目录

任务描述

了解HDFS文件系统

掌握HDFS的基本操作

获取测试数据

方式1：通过wget获取云上数据

方式2：通过百度网盘下载：

创建hdfs目录(myname替换为本人)

上传文件

下载文件

查看文件

删除文件或目录

查看目录

HDFS基本操作练习

在Hdfs中创建私人目录

文件上传下载查看操作

要求

命令集

文件或目录删除操作

要求

命令集

了解HDFS的高级操作

删除非空目录

查目录的空间使用

合并hdfs文件(了解)

过滤内容(了解)

创建快照(了解)

快照恢复(了解)

任务3.4运行首个MapReduce任务

任务描述

了解Hadoop官方示例程序包

提交MapReduce任务给集群运行

任务3.5管理多个MapReduce任务

任务描述

执行MapReduce任务

执行估算PI值的任务

词频统计

计算单词平均长度

计算单词长度中位数

计算单词长度标准差

中断MapReduce任务

执行任务1：计算pi值，在slave1上执行

执行任务2：统计登录次数，在slave2上执行

通过本机浏览器，<http://master:8088> 单机MR任务1，进入详细信息页面

单击界面中的"Kill Application",中断任务1

观察任务1是否中断，任务2是否开始执行

版本历史

任务前言

学习目标

1. 掌握查看存储系统的基本信息。
2. 掌握查看Hadoop集群的计算资源。
3. 掌握HDFS文件系统的基本操作。
4. 了解Hadoop安全模式。
5. 掌握查看、解除与开启Hadoop安全模式。
6. 掌握以hadoop jar方式提交MapReduce任务。
7. 能够管理MapReduce多任务。

任务背景

- Hadoop框架的核心设计为HDFS和MapReduce，HDFS负责海量数据的存储，MapReduce则负责海量数据的计算。对海量数据进行计算前，数据的存储是必要的一步操作，因此，需要先掌握HDFS分布式文件系统的基本操作。
- 某社交网站有数百万注册用户，网站的服务器上保留了用户登录网站的日志记录，用户每登录一次网站，将在日志文件中记录一次用户的邮件地址，现有一份社交网站在2021年某一天的原始日志文件email_log.txt，共800万行记录
- 社交网站的运营管理者希望能定期获得用户在某天的登录次数信息，作为用户行为分析的基础数据，因此统计出日志文件中每个用户的登录次数，是一个迫切需要解决的实际任务

本章主要内容：

1. 搭建并配置好Hadoop集群后，本章将通过一个社交网站用户访问日志的实例任务，讲解Hadoop的基础操作。
2. 首先介绍Hadoop的安全模式及其基本操作，包括查看、解除与开启安全模式。
3. 其次将介绍查看Hadoop集群基本信息的方法，重点介绍HDFS文件系统及其基本操作。
4. 再以Hadoop官方的示例程序包为例，介绍提交MapReduce任务至Hadoop集群运行的方法，并实现社交网站的用户登录次数统计任务。
5. 最后介绍多个MapReduce任务运行时管理方法，包括查询与中断MapReduce任务。

任务3.1认识Hadoop安全模式

安全模式是保证一个系统保密性、完整性及可使用性的一种机制，一定程度上可以防止系统里的资源遭到破坏、更改和泄露，安全模式可以使整个系统持续可靠地正常运行。Hadoop集群也有安全模式，在安全模式下可保证Hadoop集群中数据块的安全性。本小节的任务是介绍Hadoop的安全模式，了解安全模式的作用，并学习查看、开启和解除安全模式的操作方法

3.1.1了解Hadoop安全模式

安全模式是Hadoop的保护机制，保障系统不受侵坏。当处于安全模式时，文件系统只接收读数据的请求，而不接受删除、修改等变更请求。若没有安全模式，则Hadoop将处在不受保护的状态，可能存在如下安全风险

1. Hadoop服务装不验证用户或其他服务
2. 攻击者可以伪装成Hadoop服务
3. DataNode不会对节点上数据块的访问实施任何访问控制。未授权的用户可以随时获取想要的数据块信息。

3.1.2查看、解除和开启Hadoop安装模式

1. 查看安全模式

不安全 | master:9870/dfshealth.html#tab-overview

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Overview 'master:8020' (active)

Started:	Wed Sep 13 12:11:32 +0800 2023
Version:	3.1.4, r1e877761e8dadd71effef30e592368f7fe66a61b
Compiled:	Tue Jul 21 16:05:00 +0800 2020 by gabota from branch-3.1.4
Cluster ID:	CID-dd303d94-0cac-4b85-9935-8d146f79c112
Block Pool ID:	BP-1706011580-192.168.128.130-1694360578231

Summary

Security is off.

Safemode is off.

7 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 7 total filesystem object(s).

Heap Memory used 26.44 MB of 35.77 MB Heap Memory. Max Heap Memory is 235.88 MB.

Non Heap Memory used 50.01 MB of 51.38 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	33.97 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	6.2 GB

任务3.2查看Hadoop集群基本信息

Hadoop集群的核心功能

- 分布式存储
 - HDFS
 - 查看方式
 - 命令方式
 - 浏览器方式
- 并行计算
 - 计算资源分布在集群的各个节点上
 - 通过ResourceManager与NodeManger协同调配

操作终端机hosts修改

- 说明:
在windows文件中找到System32-->drivers-->etc,进入到etc文件夹中就能看到hosts文件,添加如下行,其中IP和机器名要改为本小组集群的IP和名称

```
192.168.128.131 slave1
192.168.128.132 slave2
192.168.128.130 master
192.168.128.133 slave3
```

查询集群的存储系统信息HDFS

Web方式

- HDFS监控
 - <http://master:9870/>
- 统计信息
 - Configured Capacity: 89.94 GB
 - DFS Remaining: 81.29 GB (90.39%)
 - Non DFS Used: 8.65 GB
 - DFS Used: 16 KB (0%)
 - Live Nodes 2 (Decommissioned: 0, In Maintenance: 0)

如图:

Overview 'master:9864' (✔active)

Started:	Tue Sep 19 23:09:21 +0800 2023
Version:	3.3.1, ra3b9c37a397ad4188041dd80621bdeefc46885f2
Compiled:	Tue Jun 15 13:13:00 +0800 2021 by ubuntu from (HEAD detached at release-3.3.1-RC3)
Cluster ID:	CID-157e353f-8c91-4369-bb76-73bf7b25f76f
Block Pool ID:	BP-1277059469-192.168.137.100-1662365964733

Summary

Security is off.	
Safemode is off.	
54,431 files and directories, 21,730 blocks (21,730 replicated blocks, 0 erasure coded block groups) = 76,161 total filesystem object(s).	
Heap Memory used 37.58 MB of 57.84 MB Heap Memory. Max Heap Memory is 235.88 MB.	
Non Heap Memory used 63.01 MB of 64.5 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.	
Configured Capacity:	89.94 GB
Configured Remote Capacity:	0 B
DFS Used:	6.68 GB (7.43%)
Non DFS Used:	13.23 GB
DFS Remaining:	70.03 GB (77.86%)
Block Pool Used:	6.68 GB (7.43%)
DataNodes usages% (Min/Median/Max/stdDev):	7.40% / 7.46% / 7.46% / 0.03%

命令行方式

- `hdfs dfsadmin -report`
- `hdfs dfsadmin -report -live | dead | decommissioning`

如图：

```
[root@master hadoop]# hdfs dfsadmin -report -live
Configured Capacity: 96372809216 (89.94 GB)
Present Capacity: 82369077484 (76.71 GB)
DFS Remaining: 75191521280 (70.03 GB)
DFS Used: 717556204 (6.68 GB)
DFS Used%: 8.71%
Replicated Blocks:
  Under replicated blocks: 21730
  Blocks with corrupt replicas: 98
  Missing blocks: 0
  Missing blocks (with replication factor 1): 0
  Low redundancy blocks with highest priority to recover: 98
  Pending deletion blocks: 0
Erasure Coded Block Groups:
  Low redundancy block groups: 0
  Block groups with corrupt internal blocks: 0
  Missing block groups: 0
  Low redundancy blocks with highest priority to recover: 0
  Pending deletion blocks: 0
-----
Live datanodes (2):
Name: 192.168.137.101:9868 (slave1)
Hostname: slave1
Decommission Status : Normal
Configured Capacity: 48286404608 (44.97 GB)
DFS Used: 3603970166 (3.36 GB)
Non DFS Used: 5612832650 (5.23 GB)
DFS Remaining: 39069601792 (36.39 GB)
DFS Used%: 7.46%
DFS Remaining%: 80.91%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Tue Sep 19 13:18:55 EDT 2023
Last Block Report: Tue Sep 19 11:09:30 EDT 2023
Num of Blocks: 21730

Name: 192.168.137.102:9868 (slave2)
Hostname: slave2
Decommission Status : Normal
Configured Capacity: 48286404608 (44.97 GB)
DFS Used: 3573586038 (3.33 GB)
Non DFS Used: 850899082 (8.00 GB)
DFS Remaining: 36121919488 (33.64 GB)
DFS Used%: 7.40%
DFS Remaining%: 74.81%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Tue Sep 19 13:18:55 EDT 2023
Last Block Report: Tue Sep 19 11:09:30 EDT 2023
Num of Blocks: 21730
```

查询 集群的计算资源信息YARN

Web方式

- YARN监控
 - <http://master:8088/cluster/nodes>
- 统计信息
 - Active Nodes:3
 - Total Resources:<memory:3 GB, vCores:3>

如图:

hadoop

Nodes of the cluster

Cluster Metrics

Apps Submitted	0	Apps Pending	0	Apps Running	0	Apps Completed	0	Containers Running	0	Used Resources	<memory:0 B, vCores:0>	Total Resources	<memory:4 GB, vCores:4>	Reserved Resources	<memory:0 B, vCores:0>	Physical Mem Used	83 %	Physical Vcores Used	0 %
----------------	---	--------------	---	--------------	---	----------------	---	--------------------	---	----------------	------------------------	-----------------	-------------------------	--------------------	------------------------	-------------------	------	----------------------	-----

Cluster Nodes Metrics

Active Nodes	2	Decommissioning Nodes	0	Decommissioned Nodes	0	Lost Nodes	0	Unhealthy Nodes	0	Rebooted Nodes	0	Shutdown Nodes	0
--------------	---	-----------------------	---	----------------------	---	------------	---	-----------------	---	----------------	---	----------------	---

Scheduler Metrics

Scheduler Type	Capacity Scheduler	Scheduling Resource Type	[memory-mb (unit=Mi), vcores]	Minimum Allocation	<memory:128, vCores:1>	Maximum Allocation	<memory:2048, vCores:2>	Maximum Cluster Application Priority	0	Scheduler Busy %	0
----------------	--------------------	--------------------------	-------------------------------	--------------------	------------------------	--------------------	-------------------------	--------------------------------------	---	------------------	---

Tools

Showing 1 to 2 of 2 entries

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Allocation Tags	Mem Used	Mem Avail	Phys Mem Used %	VCores Used	VCores Avail	Phys Vcores Used %	Version
/default-rack		RUNNING	slave2-44414	slave2-8042	Sun Sep 24 22:44:03 -0400 2023		0		0 B	2 GB	84	0	2	0	3.3.1
/default-rack		RUNNING	slave1-43740	slave1-8042	Thu Sep 21 08:52:49 -0400 2023		0		0 B	2 GB	81	0	2	0	3.3.1

Showing 1 to 2 of 2 entries

First Previous 1 Next Last

任务3.3上传文件到HDFS目录

任务描述

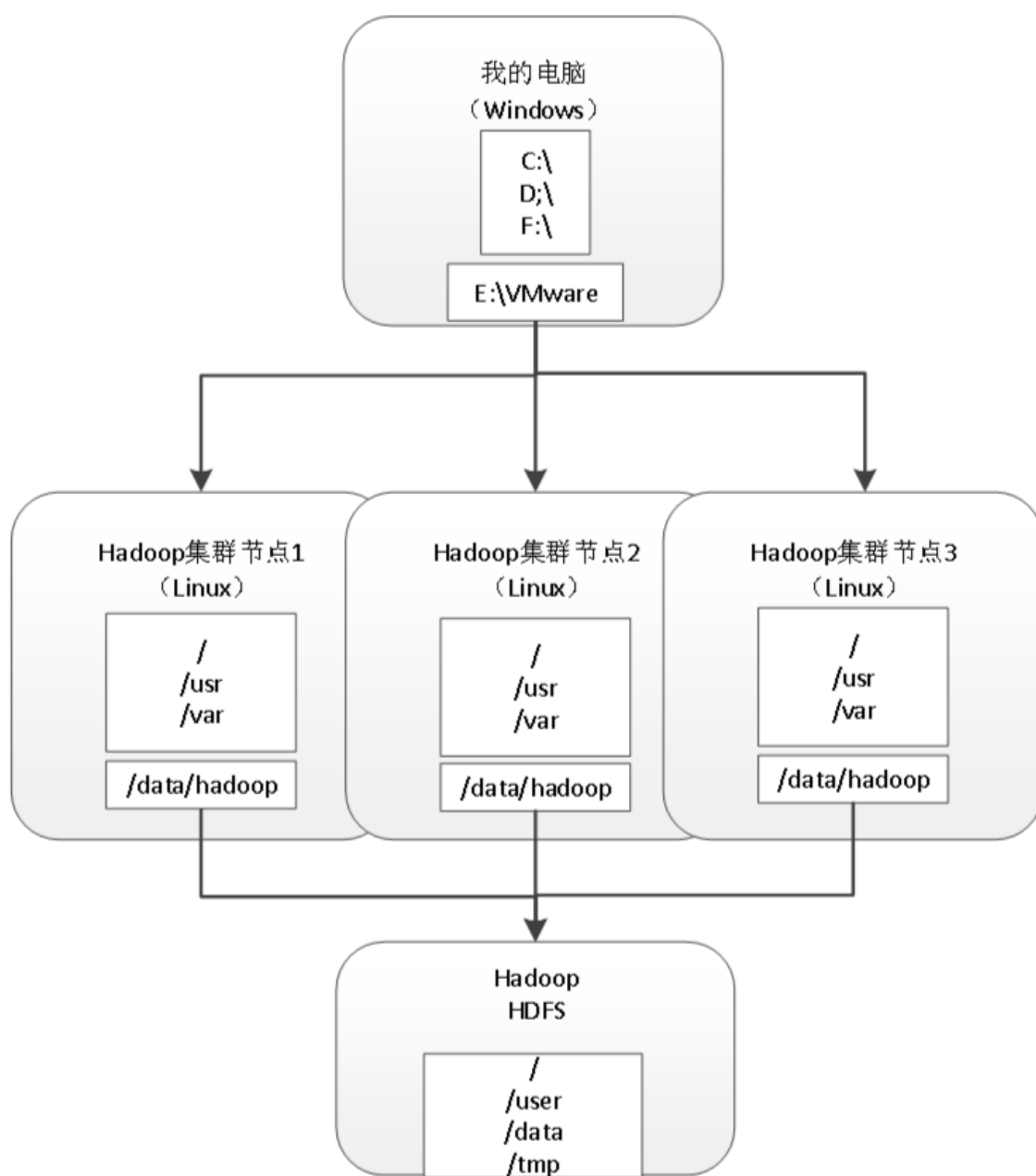
HDFS是Hadoop的核心组件之一，负责文件数据的存储。

本小节的任务如下。

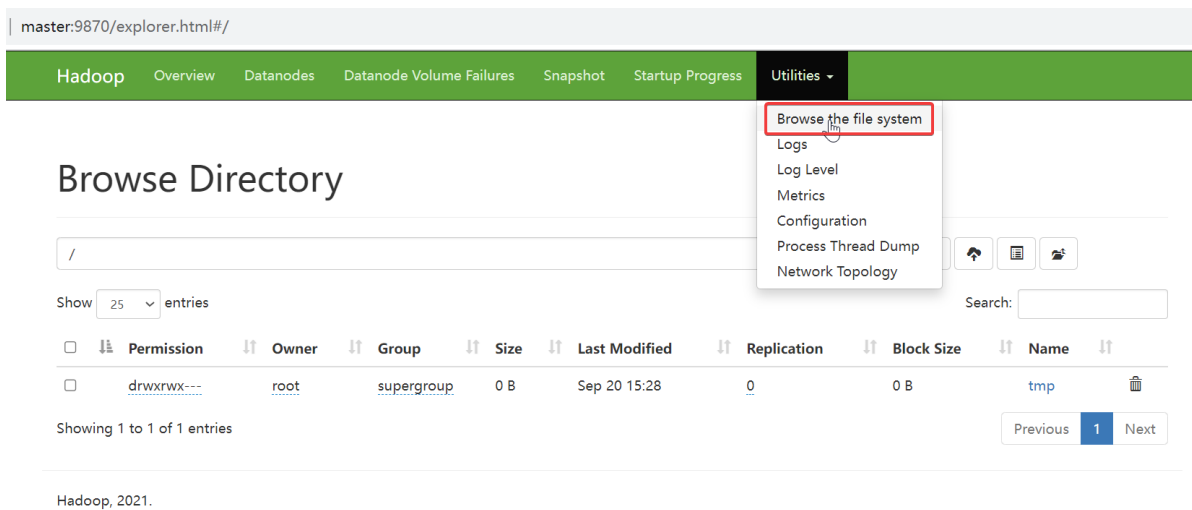
1. 了解HDFS文件系统，以及HDFS与其他文件系统之间的关系。
2. 其次介绍HDFS文件系统的基本操作。
3. 将存储在Linux本地中的日志文件email_log.txt上传至HDFS的/user/root/目录下

了解HDFS文件系统

以第2章安装及配置的Hadoop集群为例，介绍HDFS与本地计算机的文件系统、Linux本地的文件系统之间的关系：



查看HDFS文件：在本地浏览器的地址栏输入“<http://master:9870>”网址，远程访问HDFS的监控服务端口。当需要访问HDFS上的目录及文件时，可以单击网页的“Utilities”标签栏，在下拉选项中选择“Browse the file system”选项：



掌握HDFS的基本操作

说明:以下内容中myname都要替换为本人姓名全拼

获取测试数据

方式1: 通过wget获取云上数据

```
1 mkdir /root/hadoop
2 cd /root/hadoop
3 yum install -y wget
4 wget https://hddly.oss-cn-hangzhou.aliyuncs.com/down/file/email_log.tar.gz
5 tar -zxvf ./email_log.tar.gz
6 mkdir /root/hadoop/hadoop_data
7 cp ./email_log.txt /root/hadoop/hadoop_data/
```

方式2: 通过百度网盘下载:

下载地址: 链接: https://pan.baidu.com/s/1DHGRnjvi4yMY41Se7_vDA 提取码: vbgu

软件	大小	版本	安装包称
hadoop_data	99.7M	20230925	hadoop_data.tar.gz

下载hadoop_data.tar.gz后, 通过xftp工具上传到master主机的/root/hadoop目录下, 然后xshell进入master执行:

```
1 cd /root/hadoop
2 tar -zxvf ./hadoop_data.tar.gz
```

创建hdfs目录(myname替换为本人)

```
1 hdfs dfs -mkdir -p /user/myname
```

上传文件

```
1 | hdfs dfs -copyFromLocal /root/hadoop/hadoop_data/email_log.txt /user/myname/  
2 | hdfs dfs -moveFromLocal /root/hadoop/hadoop_data/email_log.txt /user/myname/  
3 | hdfs dfs -put /root/hadoop/hadoop_data/email_log.txt  
   | /user/myname/email_log.txt
```

下载文件

```
1 | hdfs dfs -copyToLocal /user/myname/email_log.txt ./  
2 | hdfs dfs -get /user/myname/email_log.txt ./
```

查看文件

```
1 | hdfs dfs -cat /user/myname/email_log.txt  
2 | hdfs dfs -tail /user/myname/email_log.txt
```

删除文件或目录

```
1 | hdfs dfs -rm 文件  
2 | hdfs dfs -rmdir 目录
```

查看目录

hdfs dfs -ls 目录

```
1 | hdfs dfs -ls /user/myname
```

hdfs dfs -ls -R 目录

```
1 | hdfs dfs -ls -R /user/myname
```

HDFS基本操作练习

说明:以下内容中myname都要替换为本人姓名全拼

在Hdfs中创建私人目录

要求:请通过命令行方式在本小组集群的HDFS文件系统中

创建目录: /user/myname

(myname都要替换为本人姓名全拼)

```
1 | hdfs dfs -mkdir -p /user/myname
```

文件上传下载查看操作

要求

通过命令行方式实现：

1. 下载文件email_log.txt (参考：获取测试数据)；
2. 上传文件email_log.txt到HDFS的本组员私人目录下；
3. 上传文件email_log.txt到HDFS的本组员私人目录下并重命名为1.txt；
4. 将HDFS上的私人目录下的1.txt下载到本地 /user/hadoop目录下
5. 将本地目录下的1.txt转移到HDFS的本组员私人目录下
6. 使用tail命令查看本组员私人目录下的2.txt文件
7. 使用cat命令查看本组员私人目录下的2.txt文件，按Ctrl+C中止查看

命令集

```
1 | cd /root/hadoop
2 | hdfs dfs -copyFromLocal email_log.txt /user/myname/
3 | hdfs dfs -put email_log.txt /user/myname/1.txt
4 | hdfs dfs -get /user/myname/1.txt ./
5 | hdfs dfs -moveFromLocal ./1.txt /user/myname/2.txt
6 | hdfs dfs -tail /user/myname/2.txt
7 | hdfs dfs -cat /user/myname/2.txt
```

Ctrl+c 退出cat内容查看

文件或目录删除操作

要求

通过命令行方式实现：

- 1) 在HDFS文件系统中本组员的私人目录下创建tmp目录；
- 2) 将本地/user/hadoop目录下的email_log.txt上传到HDFS文件系统中本组员的私人目录下的tmp目录下，并重命名为2.txt；
- 3) 删除DFS文件系统中本组员的私人目录下的tmp目录下的2.txt；
- 4) 删除DFS文件系统中本组员的私人目录下的tmp目录

命令集

```
1 | hdfs dfs -mkdir -p /user/myname/tmp
2 | hdfs dfs -put email_log.txt /user/myname/tmp/2.txt
3 | hdfs dfs -rm /user/myname/tmp/2.txt
4 | hdfs dfs -rmdir /user/myname/tmp
```

了解HDFS的高级操作

删除非空目录

hdfs dfs -rm -r 目录

```
1 | hdfs dfs -rm -r /user/myname/tmp
```

查目录的空间使用

hdfs dfs -du 目录

```
1 | hdfs dfs -du /user/myname
```

合并hdfs文件(了解)

-getmerge

将源目录和目标文件作为输入，并将src中的文件连接到目标本地文件（把两个文件的内容合并起来）

Usage: hdfs dfs -getmerge <src> <localdst> [addnl]

注：合并后的文件位于当前目录，不在hdfs中，是本地文件

过滤内容(了解)

-grep

从hdfs上过滤包含某个字符的行内容

Usage: hdfs dfs -cat <srcpath> | grep 过滤字段

hdfs dfs -cat /user/yuxm/output_music_data2/part-r-00000 | grep 俞老师

创建快照(了解)

设置允许快照:

```
1 | hdfs dfsadmin -allowSnapshot /user/myname
```

创建快照

```
1 | hdfs dfs -createSnapshot /user/myname myname_bak1
```

快照恢复(了解)

先删除测试文件

```
1 | hdfs dfs -rm /user/myname/2.txt
```

再快照恢复

```
1 | hdfs dfs -cp /user/myname/.snapshot/myname_bak1/2.txt /user/myname/
```

任务3.4运行首个MapReduce任务

任务描述

MapReduce是Hadoop的数据处理引擎，是运算程序的核心。

Hadoop官方提供了一个hadoop-mapreduce-examples-3.1.4.jar示例程序包给使用者初步运行MapReduce任务。

在该示例程序包中，存在一个词频统计模块，调用该模块即可完成用户登录次数统计任务。

本小节的任务是提交运行首个MapReduce任务，并对HDFS的/user/root/目录下的日志文件email_log.txt进行计算，调用官方示例程序包中的词频统计模块统计出每个用户的登录次数

了解Hadoop官方示例程序包

hadoop-mapreduce-examples位于：\$HADOOP_HOME/share/hadoop/mapreduce/

```
1 | cd $HADOOP_HOME/share/hadoop/mapreduce/
2 | ls hadoop-mapreduce-examples-3.1.4.jar
```

模块名称	内 容
multifilewc	统计多个文件中单词的数量
pi	应用拟蒙特卡罗（Quasi-Monte Carlo）算法估算圆周率π的值
randomtextwriter	在每个数据节点随机生成一个10GB的文本文件
wordcount	对输入文件中的单词进行频数统计
wordmean	计算输入文件中单词的平均长度
wordmedian	计算输入文件中单词长度的中位数
wordstandarddeviation	计算输入文件中单词长度的标准差

提交MapReduce任务给集群运行

提交MapReduce任务，通常使用hadoop jar命令。

hadoop jar命令的基本语法格式如下。

hadoop jar [mainClass] args

在master或slave上执行：

```
1 | cd $HADOOP_HOME/share/hadoop/mapreduce/
2 | hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar wordcount
   /user/myname/email_log.txt /user/myname/output_email_log
```

运行结果如：

```
1 | [root@slave1 ~]# cd $HADOOP_HOME/share/hadoop/mapreduce/
2 | [root@slave1 mapreduce]# ls hadoop-mapreduce-examples-3.1.4.jar
3 | hadoop-mapreduce-examples-3.1.4.jar
4 | [root@slave1 mapreduce]# hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar
   wordcount /user/myname/email_log.txt /user/myname/output_email_log
5 | 2023-09-26 21:39:46,540 INFO client.RMProxy: Connecting to ResourceManager
   at master/192.168.128.130:8032
6 | 2023-09-26 21:39:47,646 INFO mapreduce.JobResourceUploader: Disabling
   Erasure Coding for path: /tmp/hadoop-
   yarn/staging/root/.staging/job_1695733172104_0001
```

```
7 2023-09-26 21:39:48,112 INFO input.FileInputFormat: Total input files to
process : 1
8 2023-09-26 21:39:48,276 INFO mapreduce.JobSubmitter: number of splits:2
9 2023-09-26 21:39:48,571 INFO mapreduce.JobSubmitter: Submitting tokens for
job: job_1695733172104_0001
10 2023-09-26 21:39:48,573 INFO mapreduce.JobSubmitter: Executing with tokens:
[]
11 2023-09-26 21:39:48,983 INFO conf.Configuration: resource-types.xml not
found
12 2023-09-26 21:39:48,983 INFO resource.ResourceUtils: Unable to find
'resource-types.xml'.
13 2023-09-26 21:39:49,642 INFO impl.YarnClientImpl: Submitted application
application_1695733172104_0001
14 2023-09-26 21:39:49,720 INFO mapreduce.Job: The url to track the job:
http://master:8088/proxy/application_1695733172104_0001/
15 2023-09-26 21:39:49,721 INFO mapreduce.Job: Running job:
job_1695733172104_0001
16 2023-09-26 21:40:03,003 INFO mapreduce.Job: Job job_1695733172104_0001
running in uber mode : false
17 2023-09-26 21:40:03,005 INFO mapreduce.Job: map 0% reduce 0%
18 2023-09-26 21:40:23,499 INFO mapreduce.Job: map 24% reduce 0%
19 2023-09-26 21:40:29,563 INFO mapreduce.Job: map 28% reduce 0%
20 2023-09-26 21:40:35,673 INFO mapreduce.Job: map 33% reduce 0%
21 2023-09-26 21:40:40,779 INFO mapreduce.Job: map 50% reduce 0%
22 2023-09-26 21:40:58,239 INFO mapreduce.Job: map 83% reduce 0%
23 2023-09-26 21:41:04,354 INFO mapreduce.Job: map 84% reduce 0%
24 2023-09-26 21:41:07,398 INFO mapreduce.Job: map 100% reduce 0%
25 2023-09-26 21:41:23,744 INFO mapreduce.Job: map 100% reduce 100%
26 2023-09-26 21:41:24,763 INFO mapreduce.Job: Job job_1695733172104_0001
completed successfully
27 2023-09-26 21:41:24,903 INFO mapreduce.Job: Counters: 54
28     File System Counters
29         FILE: Number of bytes read=416431057
30         FILE: Number of bytes written=585286648
31         FILE: Number of read operations=0
32         FILE: Number of large read operations=0
33         FILE: Number of write operations=0
34         HDFS: Number of bytes read=226383989
35         HDFS: Number of bytes written=114167885
36         HDFS: Number of read operations=11
37         HDFS: Number of large read operations=0
38         HDFS: Number of write operations=2
39     Job Counters
40         Killed map tasks=1
41         Launched map tasks=2
42         Launched reduce tasks=1
43         Data-local map tasks=2
44         Total time spent by all maps in occupied slots (ms)=239212
45         Total time spent by all reduces in occupied slots (ms)=64196
46         Total time spent by all map tasks (ms)=59803
47         Total time spent by all reduce tasks (ms)=16049
48         Total vcore-milliseconds taken by all map tasks=59803
49         Total vcore-milliseconds taken by all reduce tasks=16049
50         Total megabyte-milliseconds taken by all map tasks=122476544
```

```

51      Total megabyte-milliseconds taken by all reduce
tasks=32868352
52      Map-Reduce Framework
53          Map input records=8000000
54          Map output records=8000000
55          Map output bytes=250379675
56          Map output materialized bytes=168189616
57          Input split bytes=218
58          Combine input records=12301355
59          Combine output records=9352725
60          Reduce input groups=3896706
61          Reduce shuffle bytes=168189616
62          Reduce input records=5051370
63          Reduce output records=3896706
64          Spilled Records=17558337
65          Shuffled Maps =2
66          Failed Shuffles=0
67          Merged Map outputs=2
68          GC time elapsed (ms)=1546
69          CPU time spent (ms)=73920
70          Physical memory (bytes) snapshot=707313664
71          Virtual memory (bytes) snapshot=10799833088
72          Total committed heap usage (bytes)=522625024
73          Peak Map Physical memory (bytes)=213454848
74          Peak Map Virtual memory (bytes)=3597885440
75          Peak Reduce Physical memory (bytes)=310030336
76          Peak Reduce Virtual memory (bytes)=3604062208
77      Shuffle Errors
78          BAD_ID=0
79          CONNECTION=0
80          IO_ERROR=0
81          WRONG_LENGTH=0
82          WRONG_MAP=0
83          WRONG_REDUCE=0
84      File Input Format Counters
85          Bytes Read=226383771
86      File Output Format Counters
87          Bytes Written=114167885
88      [root@slave1 mapreduce]#

```

任务3.5管理多个MapReduce任务

任务描述

Hadoop是一个多任务系统，可以同时为多个用户、多个作业处理多个数据集。

用户可以在本地计算机的浏览器中输入“<http://master:8088/>”网址，打开Hadoop集群的资源管理器的监控服务主界面，对提交到Hadoop集群上的多个MapReduce任务进行管理。

本小节的任务是查询和中断MapReduce任务，掌握MapReduce任务的管理方法。

执行MapReduce任务

执行估算PI值的任务

```
1 | cd $HADOOP_HOME/share/hadoop/mapreduce/  
2 | hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar pi 10 100
```

运行结果如：

```
1 | [root@slave1 mapreduce]# cd $HADOOP_HOME/share/hadoop/mapreduce/  
2 | [root@slave1 mapreduce]# hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar  
   pi 10 100  
3 | Number of Maps = 10  
4 | Samples per Map = 100  
5 | Wrote input for Map #0  
6 | Wrote input for Map #1  
7 | Wrote input for Map #2  
8 | Wrote input for Map #3  
9 | Wrote input for Map #4  
10 | Wrote input for Map #5  
11 | Wrote input for Map #6  
12 | Wrote input for Map #7  
13 | Wrote input for Map #8  
14 | Wrote input for Map #9  
15 | Starting Job  
16 | 2023-09-26 21:52:51,786 INFO client.RMProxy: Connecting to ResourceManager  
   at master/192.168.128.130:8032  
17 | 2023-09-26 21:52:52,605 INFO mapreduce.JobResourceUploader: Disabling  
   Erasure Coding for path: /tmp/hadoop-  
   yarn/staging/root/.staging/job_1695733172104_0002  
18 | 2023-09-26 21:52:52,843 INFO input.FileInputFormat: Total input files to  
   process : 10  
19 | 2023-09-26 21:52:52,961 INFO mapreduce.JobSubmitter: number of splits:10  
20 | 2023-09-26 21:52:53,335 INFO mapreduce.JobSubmitter: Submitting tokens for  
   job: job_1695733172104_0002  
21 | 2023-09-26 21:52:53,337 INFO mapreduce.JobSubmitter: Executing with tokens:  
   []  
22 | 2023-09-26 21:52:53,665 INFO conf.Configuration: resource-types.xml not  
   found  
23 | 2023-09-26 21:52:53,665 INFO resource.ResourceUtils: Unable to find  
   'resource-types.xml'.  
24 | 2023-09-26 21:52:53,803 INFO impl.YarnClientImpl: Submitted application  
   application_1695733172104_0002  
25 | 2023-09-26 21:52:53,869 INFO mapreduce.Job: The url to track the job:  
   http://master:8088/proxy/application_1695733172104_0002/  
26 | 2023-09-26 21:52:53,870 INFO mapreduce.Job: Running job:  
   job_1695733172104_0002  
27 | 2023-09-26 21:53:03,175 INFO mapreduce.Job: Job job_1695733172104_0002  
   running in uber mode : false  
28 | 2023-09-26 21:53:03,177 INFO mapreduce.Job: map 0% reduce 0%  
29 | 2023-09-26 21:53:10,406 INFO mapreduce.Job: map 10% reduce 0%  
30 | 2023-09-26 21:53:14,505 INFO mapreduce.Job: map 20% reduce 0%  
31 | 2023-09-26 21:53:19,677 INFO mapreduce.Job: map 30% reduce 0%  
32 | 2023-09-26 21:53:24,817 INFO mapreduce.Job: map 40% reduce 0%
```

```

33 2023-09-26 21:53:28,921 INFO mapreduce.Job: map 50% reduce 0%
34 2023-09-26 21:53:35,087 INFO mapreduce.Job: map 60% reduce 0%
35 2023-09-26 21:53:39,143 INFO mapreduce.Job: map 70% reduce 0%
36 2023-09-26 21:53:44,233 INFO mapreduce.Job: map 80% reduce 0%
37 2023-09-26 21:53:49,338 INFO mapreduce.Job: map 90% reduce 0%
38 2023-09-26 21:53:54,480 INFO mapreduce.Job: map 100% reduce 0%
39 2023-09-26 21:53:59,584 INFO mapreduce.Job: map 100% reduce 100%
40 2023-09-26 21:54:00,615 INFO mapreduce.Job: Job job_1695733172104_0002
    completed successfully
41 2023-09-26 21:54:00,755 INFO mapreduce.Job: Counters: 53
42     File System Counters
43         FILE: Number of bytes read=226
44         FILE: Number of bytes written=2445938
45         FILE: Number of read operations=0
46         FILE: Number of large read operations=0
47         FILE: Number of write operations=0
48         HDFS: Number of bytes read=2600
49         HDFS: Number of bytes written=215
50         HDFS: Number of read operations=45
51         HDFS: Number of large read operations=0
52         HDFS: Number of write operations=3
53     Job Counters
54         Launched map tasks=10
55         Launched reduce tasks=1
56         Data-local map tasks=10
57         Total time spent by all maps in occupied slots (ms)=149080
58         Total time spent by all reduces in occupied slots
59         (ms)=16116
60         Total time spent by all map tasks (ms)=37270
61         Total time spent by all reduce tasks (ms)=4029
62         Total vcore-milliseconds taken by all map tasks=37270
63         Total vcore-milliseconds taken by all reduce tasks=4029
64         Total megabyte-milliseconds taken by all map tasks=76328960
65         Total megabyte-milliseconds taken by all reduce
66         tasks=8251392
67     Map-Reduce Framework
68         Map input records=10
69         Map output records=20
70         Map output bytes=180
71         Map output materialized bytes=280
72         Input split bytes=1420
73         Combine input records=0
74         Combine output records=0
75         Reduce input groups=2
76         Reduce shuffle bytes=280
77         Reduce input records=20
78         Reduce output records=0
79         Spilled Records=40
80         Shuffled Maps =10
81         Failed Shuffles=0
82         Merged Map outputs=10
83         GC time elapsed (ms)=1381
84         CPU time spent (ms)=8110
85         Physical memory (bytes) snapshot=2159034368
86         Virtual memory (bytes) snapshot=39551234048

```

```

85      Total committed heap usage (bytes)=1245011968
86      Peak Map Physical memory (bytes)=207187968
87      Peak Map Virtual memory (bytes)=3597455360
88      Peak Reduce Physical memory (bytes)=105111552
89      Peak Reduce Virtual memory (bytes)=3601231872
90      Shuffle Errors
91          BAD_ID=0
92          CONNECTION=0
93          IO_ERROR=0
94          WRONG_LENGTH=0
95          WRONG_MAP=0
96          WRONG_REDUCE=0
97      File Input Format Counters
98          Bytes Read=1180
99      File Output Format Counters
100         Bytes Written=97
101 Job Finished in 69.132 seconds
102 Estimated value of Pi is 3.14800000000000000000
103 [root@slave1 mapreduce]#

```

词频统计

```

1 cd $HADOOP_HOME/share/hadoop/mapreduce/
2 hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar wordcount
  /user/myname/email_log.txt /user/root/output4

```

计算单词平均长度

```

1 cd $HADOOP_HOME/share/hadoop/mapreduce/
2 hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar wordmean
  /user/myname/email_log.txt /user/root/output5

```

运行结果如：

```

1 [root@slave1 mapreduce]# cd $HADOOP_HOME/share/hadoop/mapreduce/
2 [root@slave1 mapreduce]# hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar
  wordmean /user/myname/email_log.txt /user/root/output5
3 2023-09-26 22:02:33,556 INFO client.RMPProxy: Connecting to ResourceManager
  at master/192.168.128.130:8032
4 2023-09-26 22:02:34,525 INFO mapreduce.JobResourceUploader: Disabling
  Erasure Coding for path: /tmp/hadoop-
  yarn/staging/root/.staging/job_1695733172104_0004
5 2023-09-26 22:02:34,924 INFO input.FileInputFormat: Total input files to
  process : 1
6 2023-09-26 22:02:35,054 INFO mapreduce.JobSubmitter: number of splits:2
7 2023-09-26 22:02:35,359 INFO mapreduce.JobSubmitter: Submitting tokens for
  job: job_1695733172104_0004
8 2023-09-26 22:02:35,361 INFO mapreduce.JobSubmitter: Executing with tokens:
  []
9 2023-09-26 22:02:35,710 INFO conf.Configuration: resource-types.xml not
  found

```

```
10 2023-09-26 22:02:35,710 INFO resource.ResourceUtils: Unable to find
    'resource-types.xml'.
11 2023-09-26 22:02:35,821 INFO impl.YarnClientImpl: Submitted application
    application_1695733172104_0004
12 2023-09-26 22:02:35,896 INFO mapreduce.Job: The url to track the job:
    http://master:8088/proxy/application_1695733172104_0004/
13 2023-09-26 22:02:35,897 INFO mapreduce.Job: Running job:
    job_1695733172104_0004
14 2023-09-26 22:02:45,099 INFO mapreduce.Job: Job job_1695733172104_0004
    running in uber mode : false
15 2023-09-26 22:02:45,100 INFO mapreduce.Job: map 0% reduce 0%
16 2023-09-26 22:03:01,539 INFO mapreduce.Job: map 50% reduce 0%
17 2023-09-26 22:03:12,854 INFO mapreduce.Job: map 100% reduce 0%
18 2023-09-26 22:03:18,015 INFO mapreduce.Job: map 100% reduce 100%
19 2023-09-26 22:03:18,031 INFO mapreduce.Job: Job job_1695733172104_0004
    completed successfully
20 2023-09-26 22:03:18,186 INFO mapreduce.Job: Counters: 54
21     File System Counters
22         FILE: Number of bytes read=345
23         FILE: Number of bytes written=666365
24         FILE: Number of read operations=0
25         FILE: Number of large read operations=0
26         FILE: Number of write operations=0
27         HDFS: Number of bytes read=226383989
28         HDFS: Number of bytes written=31
29         HDFS: Number of read operations=11
30         HDFS: Number of large read operations=0
31         HDFS: Number of write operations=2
32     Job Counters
33         Killed map tasks=1
34         Launched map tasks=2
35         Launched reduce tasks=1
36         Data-local map tasks=2
37         Total time spent by all maps in occupied slots (ms)=100308
38         Total time spent by all reduces in occupied slots (ms)=15768
39         Total time spent by all map tasks (ms)=25077
40         Total time spent by all reduce tasks (ms)=3942
41         Total vcore-milliseconds taken by all map tasks=25077
42         Total vcore-milliseconds taken by all reduce tasks=3942
43         Total megabyte-milliseconds taken by all map tasks=51357696
44         Total megabyte-milliseconds taken by all reduce
    tasks=8073216
45     Map-Reduce Framework
46         Map input records=8000000
47         Map output records=16000000
48         Map output bytes=232000000
49         Map output materialized bytes=78
50         Input split bytes=218
51         Combine input records=16000014
52         Combine output records=18
53         Reduce input groups=2
54         Reduce shuffle bytes=78
55         Reduce input records=4
56         Reduce output records=2
57         Spilled Records=22
```

```

58         Shuffled Maps =2
59         Failed Shuffles=0
60         Merged Map outputs=2
61         GC time elapsed (ms)=1102
62         CPU time spent (ms)=28310
63         Physical memory (bytes) snapshot=521674752
64         Virtual memory (bytes) snapshot=10796236800
65         Total committed heap usage (bytes)=262107136
66         Peak Map Physical memory (bytes)=207740928
67         Peak Map Virtual memory (bytes)=3597885440
68         Peak Reduce Physical memory (bytes)=106577920
69         Peak Reduce Virtual memory (bytes)=3601518592
70     Shuffle Errors
71         BAD_ID=0
72         CONNECTION=0
73         IO_ERROR=0
74         WRONG_LENGTH=0
75         WRONG_MAP=0
76         WRONG_REDUCE=0
77     File Input Format Counters
78         Bytes Read=226383771
79     File Output Format Counters
80         Bytes Written=31
81 The mean is: 26.297459375
82 [root@slave1 mapreduce]#

```

计算单词长度中位数

一般地，n个数据按大小顺序排列，处于最中间位置的一个数据（或最中间位置的两个数据的平均数）叫这组数据的中位数。

```

1 cd $HADOOP_HOME/share/hadoop/mapreduce/
2 hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar wordmedian
  /user/myname/email_log.txt /user/root/output6

```

运行结果如：

```

1 [root@slave1 mapreduce]# cd $HADOOP_HOME/share/hadoop/mapreduce/
2 [root@slave1 mapreduce]# hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar
  wordmedian /user/myname/email_log.txt /user/root/output6
3 2023-09-26 22:05:54,484 INFO client.RMProxy: Connecting to ResourceManager
  at master/192.168.128.130:8032
4 2023-09-26 22:05:55,403 INFO mapreduce.JobResourceUploader: Disabling
  Erasure Coding for path: /tmp/hadoop-
  yarn/staging/root/.staging/job_1695733172104_0006
5 2023-09-26 22:05:55,788 INFO input.FileInputFormat: Total input files to
  process : 1
6 2023-09-26 22:05:55,931 INFO mapreduce.JobSubmitter: number of splits:2
7 2023-09-26 22:05:56,170 INFO mapreduce.JobSubmitter: Submitting tokens for
  job: job_1695733172104_0006
8 2023-09-26 22:05:56,173 INFO mapreduce.JobSubmitter: Executing with tokens:
  []
9 2023-09-26 22:05:56,517 INFO conf.Configuration: resource-types.xml not
  found

```

```
10 2023-09-26 22:05:56,517 INFO resource.ResourceUtils: Unable to find
    'resource-types.xml'.
11 2023-09-26 22:05:56,627 INFO impl.YarnClientImpl: Submitted application
    application_1695733172104_0006
12 2023-09-26 22:05:56,711 INFO mapreduce.Job: The url to track the job:
    http://master:8088/proxy/application_1695733172104_0006/
13 2023-09-26 22:05:56,712 INFO mapreduce.Job: Running job:
    job_1695733172104_0006
14 2023-09-26 22:06:06,205 INFO mapreduce.Job: Job job_1695733172104_0006
    running in uber mode : false
15 2023-09-26 22:06:06,207 INFO mapreduce.Job: map 0% reduce 0%
16 2023-09-26 22:06:21,381 INFO mapreduce.Job: map 50% reduce 0%
17 2023-09-26 22:06:32,464 INFO mapreduce.Job: map 100% reduce 0%
18 2023-09-26 22:06:37,512 INFO mapreduce.Job: map 100% reduce 100%
19 2023-09-26 22:06:38,532 INFO mapreduce.Job: Job job_1695733172104_0006
    completed successfully
20 2023-09-26 22:06:38,675 INFO mapreduce.Job: Counters: 54
21     File System Counters
22         FILE: Number of bytes read=2608
23         FILE: Number of bytes written=670172
24         FILE: Number of read operations=0
25         FILE: Number of large read operations=0
26         FILE: Number of write operations=0
27         HDFS: Number of bytes read=226383989
28         HDFS: Number of bytes written=447
29         HDFS: Number of read operations=11
30         HDFS: Number of large read operations=0
31         HDFS: Number of write operations=2
32     Job Counters
33         Killed map tasks=1
34         Launched map tasks=2
35         Launched reduce tasks=1
36         Data-local map tasks=2
37         Total time spent by all maps in occupied slots (ms)=91216
38         Total time spent by all reduces in occupied slots (ms)=16496
39         Total time spent by all map tasks (ms)=22804
40         Total time spent by all reduce tasks (ms)=4124
41         Total vcore-milliseconds taken by all map tasks=22804
42         Total vcore-milliseconds taken by all reduce tasks=4124
43         Total megabyte-milliseconds taken by all map tasks=46702592
44         Total megabyte-milliseconds taken by all reduce
tasks=8445952
45     Map-Reduce Framework
46         Map input records=8000000
47         Map output records=8000000
48         Map output bytes=64000000
49         Map output materialized bytes=1562
50         Input split bytes=218
51         Combine input records=8000000
52         Combine output records=155
53         Reduce input groups=52
54         Reduce shuffle bytes=1562
55         Reduce input records=155
56         Reduce output records=52
57         Spilled Records=414
```

```

58         Shuffled Maps =2
59         Failed Shuffles=0
60         Merged Map outputs=2
61         GC time elapsed (ms)=1150
62         CPU time spent (ms)=22520
63         Physical memory (bytes) snapshot=527749120
64         Virtual memory (bytes) snapshot=10792017920
65         Total committed heap usage (bytes)=262389760
66         Peak Map Physical memory (bytes)=208019456
67         Peak Map Virtual memory (bytes)=3595780096
68         Peak Reduce Physical memory (bytes)=112193536
69         Peak Reduce Virtual memory (bytes)=3601510400
70     Shuffle Errors
71         BAD_ID=0
72         CONNECTION=0
73         IO_ERROR=0
74         WRONG_LENGTH=0
75         WRONG_MAP=0
76         WRONG_REDUCE=0
77     File Input Format Counters
78         Bytes Read=226383771
79     File Output Format Counters
80         Bytes Written=447
81 The median is: 26
82 [root@slave1 mapreduce]#

```

计算单词长度标准差

```

1 cd $HADOOP_HOME/share/hadoop/mapreduce/
2 hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar wordstandarddeviation
  /user/myname/email_log.txt /user/root/output7

```

运行结果如：

```

1 [root@slave1 mapreduce]# cd $HADOOP_HOME/share/hadoop/mapreduce/
2 [root@slave1 mapreduce]# hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar
  wordstandarddeviation /user/myname/email_log.txt /user/root/output7
3 2023-09-26 22:07:51,141 INFO client.RMProxy: Connecting to ResourceManager
  at master/192.168.128.130:8032
4 2023-09-26 22:07:52,075 INFO mapreduce.JobResourceUploader: Disabling
  Erasure Coding for path: /tmp/hadoop-
  yarn/staging/root/.staging/job_1695733172104_0007
5 2023-09-26 22:07:52,486 INFO input.FileInputFormat: Total input files to
  process : 1
6 2023-09-26 22:07:52,649 INFO mapreduce.JobSubmitter: number of splits:2
7 2023-09-26 22:07:52,920 INFO mapreduce.JobSubmitter: Submitting tokens for
  job: job_1695733172104_0007
8 2023-09-26 22:07:52,921 INFO mapreduce.JobSubmitter: Executing with tokens:
  []
9 2023-09-26 22:07:53,259 INFO conf.Configuration: resource-types.xml not
  found

```

```

10 2023-09-26 22:07:53,259 INFO resource.ResourceUtils: Unable to find
    'resource-types.xml'.
11 2023-09-26 22:07:53,379 INFO impl.YarnClientImpl: Submitted application
    application_1695733172104_0007
12 2023-09-26 22:07:53,431 INFO mapreduce.Job: The url to track the job:
    http://master:8088/proxy/application_1695733172104_0007/
13 2023-09-26 22:07:53,432 INFO mapreduce.Job: Running job:
    job_1695733172104_0007
14 2023-09-26 22:08:01,724 INFO mapreduce.Job: Job job_1695733172104_0007
    running in uber mode : false
15 2023-09-26 22:08:01,726 INFO mapreduce.Job: map 0% reduce 0%
16 2023-09-26 22:08:19,908 INFO mapreduce.Job: map 31% reduce 0%
17 2023-09-26 22:08:21,951 INFO mapreduce.Job: map 50% reduce 0%
18 2023-09-26 22:08:37,069 INFO mapreduce.Job: map 100% reduce 0%
19 2023-09-26 22:08:42,101 INFO mapreduce.Job: map 100% reduce 100%
20 2023-09-26 22:08:42,114 INFO mapreduce.Job: Job job_1695733172104_0007
    completed successfully
21 2023-09-26 22:08:42,263 INFO mapreduce.Job: Counters: 54
22     File System Counters
23         FILE: Number of bytes read=666
24         FILE: Number of bytes written=666960
25         FILE: Number of read operations=0
26         FILE: Number of large read operations=0
27         FILE: Number of write operations=0
28         HDFS: Number of bytes read=226383989
29         HDFS: Number of bytes written=49
30         HDFS: Number of read operations=11
31         HDFS: Number of large read operations=0
32         HDFS: Number of write operations=2
33     Job Counters
34         Killed map tasks=1
35         Launched map tasks=2
36         Launched reduce tasks=1
37         Data-local map tasks=2
38         Total time spent by all maps in occupied slots (ms)=124832
39         Total time spent by all reduces in occupied slots (ms)=15720
40         Total time spent by all map tasks (ms)=31208
41         Total time spent by all reduce tasks (ms)=3930
42         Total vcore-milliseconds taken by all map tasks=31208
43         Total vcore-milliseconds taken by all reduce tasks=3930
44         Total megabyte-milliseconds taken by all map tasks=63913984
45         Total megabyte-milliseconds taken by all reduce
    tasks=8048640
46     Map-Reduce Framework
47         Map input records=8000000
48         Map output records=24000000
49         Map output bytes=352000000
50         Map output materialized bytes=112
51         Input split bytes=218
52         Combine input records=24000030
53         Combine output records=36
54         Reduce input groups=3
55         Reduce shuffle bytes=112
56         Reduce input records=6
57         Reduce output records=3

```

```

58         Spilled Records=42
59         Shuffled Maps =2
60         Failed Shuffles=0
61         Merged Map outputs=2
62         GC time elapsed (ms)=1164
63         CPU time spent (ms)=35940
64         Physical memory (bytes) snapshot=525041664
65         Virtual memory (bytes) snapshot=10797289472
66         Total committed heap usage (bytes)=262836224
67         Peak Map Physical memory (bytes)=208216064
68         Peak Map Virtual memory (bytes)=3597885440
69         Peak Reduce Physical memory (bytes)=109023232
70         Peak Reduce Virtual memory (bytes)=3601518592
71     Shuffle Errors
72         BAD_ID=0
73         CONNECTION=0
74         IO_ERROR=0
75         WRONG_LENGTH=0
76         WRONG_MAP=0
77         WRONG_REDUCE=0
78     File Input Format Counters
79         Bytes Read=226383771
80     File Output Format Counters
81         Bytes Written=49
82     The standard deviation is: NaN
83     [root@slave1 mapreduce]#

```

中断MapReduce任务

对于已提交的MapReduce任务，在某些特殊情况下需要中断它，如发现程序有异常、某个任务执行时间过长、占用大量计算资源等。

依次提交两个MapReduce任务，第1个任务占用大部分集群资源，且执行时间长，第2个任务一直处于等待状态。

人为中断第1个任务，观察第2个任务是否从等待状态切换为执行状态。

执行任务1：计算pi值，在slave1上执行

```

1 | cd $HADOOP_HOME/share/hadoop/mapreduce/
2 | hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar pi 30 99999

```

执行任务2：统计登录次数，在slave2上执行

```

1 | cd $HADOOP_HOME/share/hadoop/mapreduce/
2 | hadoop jar ./hadoop-mapreduce-examples-3.1.4.jar wordcount
   /user/myname/email_log.txt /user/root/output9

```

通过本机浏览器， <http://master:8088> 单机MR任务1， 进入详细信息页面

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW_SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserv
9	1	1	7	2	3.50 GB	4 GB	0 B

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
2	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory:512, vCores:1>	<memory:2048, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reser CP VCo
application_1695733172104_0011	root	word count	MAPREDUCE	default	0	Tue Sep 26 22:22:01 +0800 2023	N/A	N/A	ACCEPTED	UNDEFINED	0	0	0	0
application_1695733172104_0010	root	QuasiMonteCarlo	MAPREDUCE	default	0	Tue Sep 26 22:21:05 +0800 2023	Tue Sep 26 22:21:18 +0800 2023	N/A	RUNNING	UNDEFINED	2	2	3584	0

单击界面上的"Kill Application",中断任务1

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW_SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Application Overview

Kill Application

User: root

Name: QuasiMonteCarlo

Application Type: MAPREDUCE

Application Tags:

Application Priority: 0 (Higher Integer value indicates higher priority)

YarnApplicationState: RUNNING: AM has registered with RM and started running.

Queue: default

FinalStatus Reported by AM: Application has not completed yet.

Started: Tue Sep 26 22:21:05 +0800 2023

Launched: Tue Sep 26 22:21:18 +0800 2023

Finished: N/A

Elapsed: 2mins, 11sec

Tracking URL: ApplicationMaster

Log Aggregation Status: NOT_START

Application Timeout (Remaining Time): Unlimited

Diagnostics:

Unmanaged Application: false

Application Node Label expression: <Not set>

AM container Node Label expression: <DEFAULT_PARTITION>

Application Metrics

Total Resource Preempted:	<memory:0, vCores:0>
Total Number of Non-AM Containers Preempted:	0
Total Number of AM Containers Preempted:	0
Resource Preempted from Current Attempt:	<memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt:	0
Aggregate Resource Allocation:	405698 MB-seconds, 215 vcore-seconds
Aggregate Preempted Resource Allocation:	0 MB-seconds, 0 vcore-seconds

观察任务1是否中断， 任务2是否开始执行

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW_SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserv
9	0	1	8	1	1.50 GB	4 GB	0 B

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
2	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory:512, vCores:1>	<memory:2048, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reser CP VCo
application_1695733172104_0011	root	word count	MAPREDUCE	default	0	Tue Sep 26 22:22:01 +0800 2023	Tue Sep 26 22:23:57 +0800 2023	N/A	RUNNING	UNDEFINED	1	1	1536	0
application_1695733172104_0010	root	QuasiMonteCarlo	MAPREDUCE	default	0	Tue Sep 26 22:21:05 +0800 2023	Tue Sep 26 22:21:18 +0800 2023	Tue Sep 26 22:23:57 +0800 2023	KILLED	KILLED	N/A	N/A	N/A	N/A
application_1695733172104_0009	root	word count	MAPREDUCE	default	0	Tue Sep 26 22:21:05 +0800 2023	Tue Sep 26 22:21:18 +0800 2023	Tue Sep 26 22:23:57 +0800 2023	KILLED	KILLED	N/A	N/A	N/A	N/A

版本历史

- Ver1.1-20220121
 - 初始版本
- Ver1.2-20230828
 - 增加markdown版本
- Ver2.0-20230911
 - 修改hadoop版本为3.1.4重新撰写
- Ver2.1-20230926
 - 修改内容，更新hadoop版本和测试结果

[上一章节](#) [下一章节](#)