Hadoop配置日志聚集

应用场景

为了让应用运行完成以后，将程序运行日志信息上传到HDFS系统上，有了日志之后就可以查看程序中的报错信息，从而调试程序

配置步骤

1.配置yarn-site.xml

```
<!-- 日志聚集功能使能 -->
<property>
    <name>yarn.log-aggregation-enable</name>
    <value>true</value>
</property>


<!-- 日志保留时间设置7天 -->
<property>
    <name>yarn.log-aggregation.retain-seconds</name>
    <value>604800</value>
</property>
```

```
</property>

<!-- 指定YARN的ResourceManager的地址 -->
<property>
            <name>yarn.resourcemanager.hostname</name>
            <value>hadoop103</value>
</property>

<!-- 日志聚集功能使能 -->
<property>
<name>yarn.log-aggregation-enable</name>
<value>true</value>
</property>

<!-- 日志保留时间设置7天 -->
<property>
<name>yarn.log-aggregation.retain-seconds</name>
<value>604800</value>
</property>
-- 插入 --                                    40,12        91%
```

2.重启集群

```
sbin/stop-yarn.sh
sbin/stop-dfs.sh
sbin/mr-jobhistory-daemon.sh stop historyserver

sbin/start-dfs.sh
sbin/start-yarn.sh
sbin/mr-jobhistory-daemon.sh start historyserver
```

3.运行程序

上传数据到HDFS

```
[zhutian@hadoop102 hadoop-2.7.2]$ hadoop fs -mkdir /input
[zhutian@hadoop102 hadoop-2.7.2]$ hadoop fs -copyFromLocal ./input/word_data.txt /input
[zhutian@hadoop102 hadoop-2.7.2]$ hadoop fs -ls -R /
drwxr-xr-x   - zhutian supergroup          0 2020-05-13 15:12 /input
-rw-r--r--   3 zhutian supergroup         38 2020-05-13 15:12 /input/word_data.txt
drwxrwx---   - zhutian supergroup          0 2020-05-13 14:50 /tmp
drwxrwx---   - zhutian supergroup          0 2020-05-13 14:50 /tmp/hadoop-yarn
drwxrwx---   - zhutian supergroup          0 2020-05-13 14:50 /tmp/hadoop-yarn/staging
drwxrwx---   - zhutian supergroup          0 2020-05-13 14:50 /tmp/hadoop-yarn/staging/history
drwxrwx---   - zhutian supergroup          0 2020-05-13 14:50 /tmp/hadoop-yarn/staging/history/d
one
drwxrwxrwt   - zhutian supergroup          0 2020-05-13 14:50 /tmp/hadoop-yarn/staging/history/d
one_intermediate
```

赋给文件权限

```
[zhutian@hadoop102 hadoop-2.7.2]$ hadoop fs -chmod 777 /input/word_data.txt
```

运行程序

```
[zhutian@hadoop102 hadoop-2.7.2]$ hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar wordcoun
t /input /output
20/05/13 15:17:31 INFO client.RMProxy: Connecting to ResourceManager at hadoop103/192.168.159.103:8032
20/05/13 15:17:31 INFO input.FileInputFormat: Total input paths to process : 1
20/05/13 15:17:31 INFO mapreduce.JobSubmitter: number of splits:1
20/05/13 15:17:31 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1589354225366_0001
20/05/13 15:17:32 INFO impl.YarnClientImpl: Submitted application application_1589354225366_0001
20/05/13 15:17:32 INFO mapreduce.Job: The url to track the job: http://hadoop103:8088/proxy/application_15893542
25366_0001/
20/05/13 15:17:32 INFO mapreduce.Job: Running job: job_1589354225366_0001
20/05/13 15:17:39 INFO mapreduce.Job: Job job_1589354225366_0001 running in uber mode : false
20/05/13 15:17:39 INFO mapreduce.Job:  map 0% reduce 0%
20/05/13 15:17:44 INFO mapreduce.Job:  map 100% reduce 0%
```
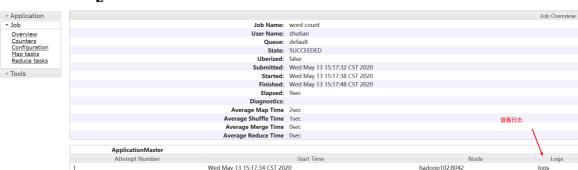
查看运行结果

```
[zhutian@hadoop102 hadoop-2.7.2]$ hadoop fs -cat /output/part-r-00000
flink   3
hadoop  2
spark   1
[zhutian@hadoop102 hadoop-2.7.2]$
```

4.查看日志

MapReduce Job job_1589354225366_0001

- Application
- ▾ Job
  - Overview
  - Counters
  - Configuration
  - Map tasks
  - Reduce tasks
- Tools

	Job Overview
Job Name:	word count
User Name:	zhutian
Queue:	default
State:	SUCCEEDED
Uberized:	false
Submitted:	Wed May 13 15:17:32 CST 2020
Started:	Wed May 13 15:17:38 CST 2020
Finished:	Wed May 13 15:17:48 CST 2020
Elapsed:	9sec
Diagnostics:	
Average Map Time	2sec
Average Shuffle Time	1sec
Average Merge Time	0sec
Average Reduce Time	0sec

查看日志

ApplicationMaster			
Attempt Number	Start Time	Node	Logs
1	Wed May 13 15:17:34 CST 2020	hadoop102:8042	logs

- ▾ Application
  - About
  - Jobs
- Tools

Log Type: stderr
Log Upload Time: 星期三 五月 13 15:17:54 +0800 2020
Log Length: 1766

程序日志

五月 13, 2020 3:17:37 下午 com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
信息: Registering org.apache.hadoop.mapreduce.v2.app.webapp.JAXBContextResolver as a provider class
五月 13, 2020 3:17:37 下午 com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
信息: Registering org.apache.hadoop.yarn.webapp.GenericExceptionHandler as a provider class
五月 13, 2020 3:17:37 下午 com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory register
信息: Registering org.apache.hadoop.mapreduce.v2.app.webapp.AMWebServices as a root resource class
五月 13, 2020 3:17:37 下午 com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
信息: Initiating Jersey application, version 'Jersey: 1.9 09/02/2011 11:17 AM'
五月 13, 2020 3:17:37 下午 com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
信息: Binding org.apache.hadoop.mapreduce.v2.app.webapp.JAXBContextResolver to GuiceManagedComponentProvider with the scope "Singleton"
五月 13, 2020 3:17:37 下午 com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
信息: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Singleton"
五月 13, 2020 3:17:37 下午 com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
信息: Binding org.apache.hadoop.mapreduce.v2.app.webapp.AMWebServices to GuiceManagedComponentProvider with the scope "PerRequest"
log4j:WARN No appenders could be found for logger (org.apache.hadoop.ipc.Server).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.