

前言

Flink On YARN 默认作业挂了之后打开的话，是一个如下这样的页面：

The screenshot shows the Hadoop JobManager web UI for application `application_1596708697506_1974`. The interface includes a sidebar with navigation links and a main content area with sections for 'Application Overview', 'Application Metrics', and a table of application attempts. The 'Application Overview' section displays various status and configuration details. The 'Application Metrics' section shows resource preemption statistics. The table at the bottom lists application attempts with columns for Attempt ID, Started, Node, Logs, Nodes blacklisted by the app, and Nodes blacklisted by the system.

对于这种我们页面我们只能查看 JobManager 的日志，不再可以查看作业挂掉之前的运行的 Web UI，很难清楚知道作业在挂的那一刻到底发生了啥？如果我们还没有 Metrics 监控的话，那么完全就只能通过日志去分析和定位问题了，所以如果能还原之前的 Web UI，我们可以通过 UI 发现和定位一些问题。

History Server

那么这里就需要利用 Flink 中的 History Server 来解决这个问题，如何开启这个呢？

你需要在 `flink-conf.yml` 中配置如下：

```
1  #=====
2  # HistoryServer
3  #=====
4
5  # The HistoryServer is started and stopped via bin/historyserver.sh
6  (start|stop)
7
8  # Directory to upload completed jobs to. Add this directory to the list of
9  # monitored directories of the HistoryServer as well (see below).
10 # flink job 运行完成后的日志存放目录
11 jobmanager.archive.fs.dir: hdfs:///flink/history-log
12
13 # The address under which the web-based HistoryServer listens.
14 # flink history进程所在的主机
```

```

14 #historyserver.web.address: 0.0.0.0
15
16 # The port under which the web-based HistoryServer listens.
17 # flink history进程的占用端口
18 #historyserver.web.port: 8082
19
20 # Comma separated list of directories to monitor for completed jobs.
21 # flink history进程的hdfs监控目录
22 historyserver.archive.fs.dir: hdfs:///flink/history-log
23
24 # Interval in milliseconds for refreshing the monitored directories.
25 # 刷新受监视目录的时间间隔（以毫秒为单位）
26 #historyserver.archive.fs.refresh-interval: 10000

```

注意：**jobmanager.archive.fs.dir** 要和 **historyserver.archive.fs.dir** 配置的路径要一样

执行命令：

```
1 | ./bin/historyserver.sh start
```

发现报错如下：

```

1 2020-10-13 21:21:01,310 main INFO org.apache.flink.core.fs.FileSystem
    - Hadoop is not in the classpath/dependencies. The
    extended set of supported File Systems via Hadoop is not available.
2 2020-10-13 21:21:01,336 main INFO
    org.apache.flink.runtime.security.modules.HadoopModuleFactory - Cannot
    create Hadoop Security Module because Hadoop cannot be found in the
    Classpath.
3 2020-10-13 21:21:01,352 main INFO
    org.apache.flink.runtime.security.modules.JaasModule - Jaas file
    will be created as /tmp/jaas-354359771751866787.conf.
4 2020-10-13 21:21:01,355 main INFO
    org.apache.flink.runtime.security.SecurityUtils - Cannot
    install HadoopSecurityContext because Hadoop cannot be found in the
    Classpath.
5 2020-10-13 21:21:01,363 main WARN
    org.apache.flink.runtime.webmonitor.history.HistoryServer - Failed to
    create Path or FileSystem for directory 'hdfs:///flink/history-log'.
    Directory will not be monitored.
6 org.apache.flink.core.fs.UnsupportedFileSystemSchemeException: Could not
    find a file system implementation for scheme 'hdfs'. The scheme is not
    directly supported by Flink and no Hadoop file system to support this
    scheme could be loaded.
7     at
    org.apache.flink.core.fs.FileSystem.getUnguardedFileSystem(FileSystem.java
    :450)
8         at org.apache.flink.core.fs.FileSystem.get(FileSystem.java:362)
9         at org.apache.flink.core.fs.Path.getFileSystem(Path.java:298)
10        at org.apache.flink.runtime.webmonitor.history.HistoryServer.
    <init>(HistoryServer.java:187)
11        at org.apache.flink.runtime.webmonitor.history.HistoryServer.
    <init>(HistoryServer.java:137)
12        at
    org.apache.flink.runtime.webmonitor.history.HistoryServer$1.call(HistorySe
    rver.java:122)

```

```

13         at
org.apache.flink.runtime.webmonitor.history.HistoryServer$1.call (HistorySe
rver.java:119)
14         at
org.apache.flink.runtime.security.NoOpSecurityContext.runSecured (NoOpSecur
ityContext.java:30)
15         at
org.apache.flink.runtime.webmonitor.history.HistoryServer.main (HistoryServ
er.java:119)
16 Caused by: org.apache.flink.core.fs.UnsupportedFileSystemSchemeException:
Hadoop is not in the classpath/dependencies.
17         at
org.apache.flink.core.fs.UnsupportedSchemeFactory.create (UnsupportedScheme
Factory.java:58)
18         at
org.apache.flink.core.fs.FileSystem.getUnguardedFileSystem (FileSystem.java
:446)
19         ... 8 more
20 2020-10-13 21:21:01,367 main ERROR
org.apache.flink.runtime.webmonitor.history.HistoryServer      - Failed to
run HistoryServer.
21 org.apache.flink.util.FlinkException: Failed to validate any of the
configured directories to monitor.
22         at org.apache.flink.runtime.webmonitor.history.HistoryServer.
<init> (HistoryServer.java:196)
23         at org.apache.flink.runtime.webmonitor.history.HistoryServer.
<init> (HistoryServer.java:137)
24         at
org.apache.flink.runtime.webmonitor.history.HistoryServer$1.call (HistorySe
rver.java:122)
25         at
org.apache.flink.runtime.webmonitor.history.HistoryServer$1.call (HistorySe
rver.java:119)
26         at
org.apache.flink.runtime.security.NoOpSecurityContext.runSecured (NoOpSecur
ityContext.java:30)
27         at
org.apache.flink.runtime.webmonitor.history.HistoryServer.main (HistoryServ
er.java:119)

```

这个异常的原因是因为 Flink 集群的 CLASS_PATH 下缺少了 HDFS 相关的 jar，我们可以引入 HDFS 的依赖放到 lib 目录下面或者添加 Hadoop 的环境变量。

这里我们在 historyserver.sh 脚本中增加下面脚本，目的就是添加 Hadoop 的环境变量：

```

1 # export hadoop classpath
2 if [ `command -v hadoop` ];then
3     export HADOOP_CLASSPATH=`hadoop classpath`
4 else
5     echo "hadoop command not found in path!"
6 fi

```

效果

Apache Flink Dashboard

Overview

Jobs

Running Jobs

Completed Jobs

Task Managers

Job Manager

Version: 1.10.0

Commit: aa4eb8f @ 07.02.2020 @ 19:18:19 CET

Message:

Running Job List

Job Name	Start Time	Duration	End Time	Tasks	Status
No Data					

Completed Job List

Job Name	Start Time	Duration	End Time	Tasks	Status
State machine job	2020-10-13 22:08:44	4m 13s	2020-10-13 22:12:57	4/4	CANCELED
State machine job	2020-10-13 21:53:41	1m 37s	2020-10-13 21:55:18	4/4	CANCELED
full_indexer_main_indexer_2020_10_13_16_42_49	2020-10-13 19:29:09	1m 4s	2020-10-13 19:30:14	8/8	FINISHED
full_indexer_main_indexer_2020_10_13_16_42_49	2020-10-13 16:43:30	1m 5s	2020-10-13 16:44:36	8/8	FINISHED
full_indexer_main_indexer_2020_10_13_16_34_56	2020-10-13 14:39:47	24s	2020-10-13 14:40:12	8/8	FINISHED
Flink Java Job at Mon Oct 12 16:46:53 CST 2020	2020-10-12 16:47:11	11s	2020-10-12 16:47:22	2/2	FINISHED
Flink Java Job at Mon Oct 12 16:45:09 CST 2020	2020-10-12 16:45:24	11s	2020-10-12 16:45:35	2/2	FINISHED
Flink Java Job at Mon Oct 12 16:44:13 CST 2020	2020-10-12 16:44:31	13s	2020-10-12 16:44:45	2/2	FINISHED
Flink Java Job at Mon Oct 12 16:37:43 CST 2020	2020-10-12 16:38:01	13s	2020-10-12 16:38:14	2/2	FINISHED
Flink Java Job at Mon Oct 12 16:34:54 CST 2020	2020-10-12 16:35:11	12s	2020-10-12 16:35:23	2/2	FINISHED
Flink Java Job at Mon Oct 12 16:11:51 CST 2020	2020-10-12 16:12:09	10s	2020-10-12 16:12:19	2/2	FINISHED

→ → → 不安全 | 2024-10-09 14:02:27 #job=te24e02748df85afed9ee3da2abf23a/bview
Version: 1.10.0 Commit: aa4eb8f @ 07.02.2020 @ 19:18:19 CET Message:

Apache Flink Dashboard

- Overview
- Jobs
- Running Jobs
- Completed Jobs
- Task Managers
- Job Manager

Flink Streaming Job FAILED 1 5

ID: **1e24e02748df85afed9ee3da2abf23a** Start Time: **2020-10-09 10:20:17** End Time: **2020-10-09 14:40:21** Duration: **4h 20m 4s**

[Overview](#) Exceptions TimeLine Checkpoints Configuration

```

graph TD
    S[Source: Custom Source -> Filter] -- "filter" --> P1[Process: ...]
    S -- "filter" --> P2[Process: ...]
    S -- "filter" --> P3[Process: ...]
    S -- "filter" --> P4[Process: ...]
    S -- "filter" --> P5[Process: ...]
    
```

Name	Status	Bytes Received	Records Received	Bytes Sent	Records Sent	Parallelism	Start Time	Duration	Tasks
Source: Custom Source -> Flat Map -> Filter -> Process -> [TimeWindow]	CANCELED	0 B	0	5.01 MB	10,031	3	2020-10-09 14:40:10	9s	3
Window[TumblingProcessingTimeWindows(10000), ProcessingTimeTrigger(10000)]	CANCELED	252 B	0	0 B	0	3	2020-10-09 14:40:10	9s	3
Window[TumblingProcessingTimeWindows(10000), ProcessingTimeTrigger(10000)]	FAILED	5.13 MB	10,028	0 B	0	3	2020-10-09 14:40:10	11s	1 2
Window[TumblingProcessingTimeWindows(10000), ProcessingTimeTrigger(10000)]	CANCELED	252 B	0	0 B	0	3	2020-10-09 14:40:10	9s	3
Window[TumblingProcessingTimeWindows(10000), ProcessingTimeTrigger(10000)]	CANCELED	252 B	0	0 B	0	3	2020-10-09 14:40:10	9s	3
Window[TumblingProcessingTimeWindows(10000), ProcessingTimeTrigger(10000)]	CANCELED	252 B	0	0 B	0	3	2020-10-09 14:40:10	9s	3

```
1 | hdfs dfs -ls /flink/history-log/
```

```

-rw-r--r-- 3 flink hadoop 100093 2020-05-26 20:21 /flink/history-log/ca7b4fff8d307f1685351e69aa54c0a9
-rw-r--r-- 3 flink hadoop 69964 2020-01-06 16:22 /flink/history-log/cad2470f25cf459ba6a625956a9fa71b
-rw-r--r-- 3 flink hadoop 15173 2020-10-12 15:43 /flink/history-log/cb22640d59769697ae41381c0921234b
-rw-r--r-- 3 flink hadoop 9623 2020-05-29 11:50 /flink/history-log/cb5345b808b1226a483602746722ccae
-rw-r--r-- 3 flink hadoop 45794 2020-06-13 23:25 /flink/history-log/cb5410bde25fc52caa0341b06e4667c4
-rw-r--r-- 3 flink hadoop 138719 2019-11-26 10:35 /flink/history-log/cb81f692e69a24a9adcc2c3bfa006e87
-rw-r--r-- 3 flink hadoop 39995 2020-06-03 17:43 /flink/history-log/cb93230496ced3a55db44d8c94a50b6b
-rw-r--r-- 3 flink hadoop 35586 2020-08-19 11:46 /flink/history-log/cbe8631965b6fbc6cb01dbd23d72fb60
-rw-r--r-- 3 flink hadoop 72428 2019-11-24 21:33 /flink/history-log/cbfd919a1b49f6204b43b1c9b994c6ba
-rw-r--r-- 3 flink hadoop 30642 2020-07-08 18:53 /flink/history-log/cc16af1223366d67392e63e6e54b43a5
-rw-r--r-- 3 flink hadoop 134756 2020-09-02 11:30 /flink/history-log/cc40f4ee366deaf7f19b8728cc7da07a
-rw-r--r-- 3 flink hadoop 96848 2020-06-20 17:16 /flink/history-log/cc651178f420db1dfe01a9d340f866bf
-rw-r--r-- 3 flink hadoop 16534 2020-04-09 17:24 /flink/history-log/ccb434908ae66845e4f1b3570803673e
-rw-r--r-- 3 flink hadoop 363138 2020-08-20 17:18 /flink/history-log/cc9220d67b01135e6cf19fd97942188
-rw-r--r-- 3 flink hadoop 25493 2020-06-15 14:46 /flink/history-log/cccd8de87334d1445a0cd5351a52530b
-rw-r--r-- 3 flink hadoop 43152 2019-11-15 18:13 /flink/history-log/cd172210c2b872c33e793311eb92a797
-rw-r--r-- 3 flink hadoop 386580 2020-06-15 16:09 /flink/history-log/cd3b419db0c3d02a8a3e04610aac7192
-rw-r--r-- 3 flink hadoop 95277 2020-04-13 16:06 /flink/history-log/cda1b03b02b8ac44e2ba908648e23cfe
-rw-r--r-- 3 flink hadoop 97020 2020-04-03 15:52 /flink/history-log/cdb210e3b3a0fba201cb56f32caec65c
-rw-r--r-- 3 flink hadoop 22938 2020-05-29 10:24 /flink/history-log/cdbcf62ce40b2caf451ab44a75879c93
-rw-r--r-- 3 flink hadoop 104487 2020-06-12 15:38 /flink/history-log/cdee8e7b49a5717a5efbfc1841658ed85

```

其实 history server 会在本地存储已结束 Job 信息，你可以配置 `historyserver.web.tmpdir` 来决定存储在哪，默认的拼接规则为：

```

1 | System.getProperty("java.io.tmpdir") + File.separator + "flink-web-history-"
  | + UUID.randomUUID()

```

Linux 系统临时目录为 `/tmp`，你可以看到源码中 `HistoryServerOptions` 该类中的可选参数。

```

1 | /**
2 |  * The local directory used by the HistoryServer web-frontend.
3 |  */
4 | public static final ConfigOption<String> HISTORY_SERVER_WEB_DIR =
5 |     key("historyserver.web.tmpdir")
6 |         .noDefaultValue()
7 |         .withDescription("This configuration parameter allows defining the
8 |             Flink web directory to be used by the" +
9 |                 " history server web interface. The web interface will copy its
10 |                 static files into the directory.");

```

那么我们找到本地该临时目录，可以观察到里面保存着很多 JS 文件，其实就是我们刚才看到的页面

```

22:51:09deploy@: /app/flink-1.10.0
$ ll /tmp/flink-web-history-02036d66-be85-4193-907b-1529cad44118
total 3352
-rw-rw-r-- 1 deploy deploy 14470 Oct 13 21:52 10.c9ca91a95a4d63748ee1.js
-rw-rw-r-- 1 deploy deploy 38097 Oct 13 21:52 11.56c9972eef60dc16521e.js
-rw-rw-r-- 1 deploy deploy 153830 Oct 13 21:52 12.8cc05a4b060334cb0c35.js
-rw-rw-r-- 1 deploy deploy 146748 Oct 13 21:52 13.d5131dc0a5369f20f02d.js
-rw-rw-r-- 1 deploy deploy 226348 Oct 13 21:52 2.f6fb98ae7d83e12a2a6d.js
-rw-rw-r-- 1 deploy deploy 570037 Oct 13 21:52 3.ec614c80ee284892305c.js
-rw-rw-r-- 1 deploy deploy 16304 Oct 13 21:52 8.84125283c79153121580.js
-rw-rw-r-- 1 deploy deploy 70302 Oct 13 21:52 9.86fa9c404a7042039c2d.js
drwxrwxr-x 4 deploy deploy 4096 Oct 13 21:52 assets
-rw-rw-r-- 1 deploy deploy 16664 Oct 13 21:52 common.a709d9aae860ac87013a.js
-rw-rw-r-- 1 deploy deploy 194 Oct 13 21:52 config.json
-rw-rw-r-- 1 deploy deploy 2137 Oct 13 21:52 index.html
drwxrwxr-x 2093 deploy deploy 278528 Oct 13 22:13 jobs
drwxrwxr-x 3 deploy deploy 4096 Oct 13 21:54 libs
-rw-rw-r-- 1 deploy deploy 1219946 Oct 13 21:52 main.7b22367c6265ef1a481d.js
drwxrwxr-x 2 deploy deploy 147456 Oct 13 22:13 overviews
-rw-rw-r-- 1 deploy deploy 41994 Oct 13 21:52 polyfills.b37850e8279bc3caafc9.js
-rw-rw-r-- 1 deploy deploy 2410 Oct 13 21:52 runtime.6deb22162a17754db1bc.js
-rw-rw-r-- 1 deploy deploy 435784 Oct 13 21:52 styles.30d0912c1ece284d8d9a.css

```


历史服务存储文件中，存储了用于页面展示的模板配置。历史任务信息存储在 Jobs 路径下，其中包含了已经完成的 Job，每次启动都会从 historyserver.archive.fs.dir 拉取所有的任务元数据信息。

```
-rw-rw-r-- 1 deploy deploy 4106 Oct 13 21:55 fe9f19661875c2603010e11bd5629f2c.json
drwxrwxr-x 4 deploy deploy 4096 Oct 13 21:55 fededf204b6546ee582c999f2ef1fd80
-rw-rw-r-- 1 deploy deploy 3517 Oct 13 21:55 fededf204b6546ee582c999f2ef1fd80.json
drwxrwxr-x 4 deploy deploy 4096 Oct 13 21:55 ff065df585c692f8bc9f4c791fdd31ac
-rw-rw-r-- 1 deploy deploy 4521 Oct 13 21:55 ff065df585c692f8bc9f4c791fdd31ac.json
drwxrwxr-x 4 deploy deploy 4096 Oct 13 21:55 ff53218afce4a45b4b46b749eed0fbae
-rw-rw-r-- 1 deploy deploy 3423 Oct 13 21:55 ff53218afce4a45b4b46b749eed0fbae.json
drwxrwxr-x 4 deploy deploy 4096 Oct 13 21:55 ff6326d9ef2a6c5ccc20a7c94704ebf7
-rw-rw-r-- 1 deploy deploy 7043 Oct 13 21:55 ff6326d9ef2a6c5ccc20a7c94704ebf7.json
drwxrwxr-x 4 deploy deploy 4096 Oct 13 21:55 ff68ce2c6171e4d858c61b5d29d3548a
-rw-rw-r-- 1 deploy deploy 1464 Oct 13 21:55 ff68ce2c6171e4d858c61b5d29d3548a.json
drwxrwxr-x 4 deploy deploy 4096 Oct 13 21:55 ff8ba46de3b15bdc0dceb24462b22b6e
-rw-rw-r-- 1 deploy deploy 2577 Oct 13 21:55 ff8ba46de3b15bdc0dceb24462b22b6e.json
drwxrwxr-x 4 deploy deploy 4096 Oct 13 21:55 fffb75b942bea3669338b2e87b3ee0cb2
-rw-rw-r-- 1 deploy deploy 13699 Oct 13 21:55 ffb75b942bea3669338b2e87b3ee0cb2.json
drwxrwxr-x 4 deploy deploy 4096 Oct 13 21:55 ffbb07e5e0039d318718d99304eef33a
-rw-rw-r-- 1 deploy deploy 1411 Oct 13 21:55 ffbb07e5e0039d318718d99304eef33a.json
drwxrwxr-x 4 deploy deploy 4096 Oct 13 21:55 fff16030ed22c35e2f3436e027cde2a7
-rw-rw-r-- 1 deploy deploy 1411 Oct 13 21:55 fff16030ed22c35e2f3436e027cde2a7.json
drwxrwxr-x 3 deploy deploy 4096 Oct 13 21:52 :jobid
-rw-rw-r-- 1 deploy deploy 766303 Oct 13 22:13 overview.json
```

每个任务文件夹中包含我们需要获取的一些信息，通过 REST API 获取时指标时，就是返回这些内容（Checkpoint/Exception 信息等）。

```
$ ll /tmp/flink-web-history-02036d66-be85-4193-907b-1529cad44118/jobs/fff16030ed22c35e2f3436e027cde2a7
total 28
-rw-rw-r-- 1 deploy deploy 90 Oct 13 21:55 accumulators.json
drwxrwxr-x 2 deploy deploy 4096 Oct 13 21:55 checkpoints
-rw-rw-r-- 1 deploy deploy 310 Oct 13 21:55 checkpoints.json
-rw-rw-r-- 1 deploy deploy 351 Oct 13 21:55 config.json
-rw-rw-r-- 1 deploy deploy 78 Oct 13 21:55 exceptions.json
-rw-rw-r-- 1 deploy deploy 315 Oct 13 21:55 plan.json
drwxrwxr-x 3 deploy deploy 4096 Oct 13 21:55 vertices
```

REST API

以下是可用且带有示例 JSON 响应的请求列表。所有请求格式样例均为

http://hostname:8082/jobs，下面我们仅列出了 URLs 的 *path* 部分。尖括号中的值为变量，例如作业 7684be6004e4e955c2a558a9bc463f65 的

http://hostname:port/jobs/<jobid>/exceptions 请求须写为

http://hostname:port/jobs/7684be6004e4e955c2a558a9bc463f65/exceptions。

- /config
- /jobs/overview
- /jobs/<jobid>
- /jobs/<jobid>/vertices
- /jobs/<jobid>/config

- /jobs/<jobid>/exceptions
- /jobs/<jobid>/accumulators
- /jobs/<jobid>/vertices/<vertexid>
- /jobs/<jobid>/vertices/<vertexid>/subtasktimes
- /jobs/<jobid>/vertices/<vertexid>/taskmanagers
- /jobs/<jobid>/vertices/<vertexid>/accumulators
- /jobs/<jobid>/vertices/<vertexid>/subtasks/accumulators
- /jobs/<jobid>/vertices/<vertexid>/subtasks/<subtasknum>
- /jobs/<jobid>/vertices/<vertexid>/subtasks/<subtasknum>/attempts
/<attempt>
- /jobs/<jobid>/vertices/<vertexid>/subtasks/<subtasknum>/attempts
/<attempt>/accumulators
- /jobs/<jobid>/plan

总结

这样我们就可以开心的去查看作业挂之前的 Web UI 信息了，妈妈在也不用担心我的作业挂了! 😊

参考文章

- [History Server](#)
- [flink历史服务](#)