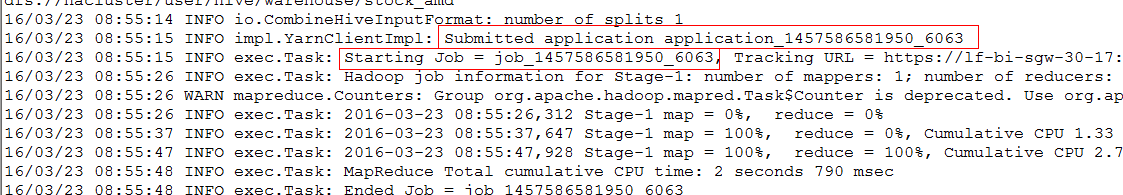
# TCC失效特性列表

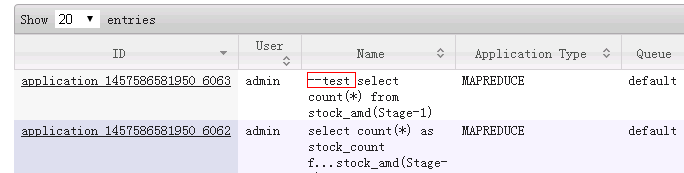
|  |  |  |
| --- | --- | --- |
| **序号** | **失效特性** | **失效原因** |
| 1 | 壳日志没有任务执行进度信息 | TCC壳日志是通过记录ssh客户端的标准输出作为壳执行日志的。由于FI将hive的任务进度及jobid信息屏蔽了，导致ssh客户端的标准输出也不包含任务进度和jobid信息。 |
| 2 | 在壳日志界面以及步骤运行状态界面中没有任务对应的jobid列表，无法快速查看对应job的详细信息 | TCC通过解析ssh客户端的标准输出，得到任务相应的所有jobid。由于FI将hive的任务进度及jobid信息屏蔽了，导致ssh客户端的标准输出也不包含任务进度和jobid信息。所以TCC也无法将这些jobid列出来。 |
| 3 | 任务重做或停止任务时，无法停止相应的已经在运行中的hadoop job | 同上，JOBID无法获取，导致无法停止hadoop job |
| 4 | 任务依赖图中的任务资源消耗表失效 | hadoop日志存储方式不一样 |

# 失效特性1的解决方案

**方案1**：让FI提供补丁，将任务进度及jobid信息输出到客户端。



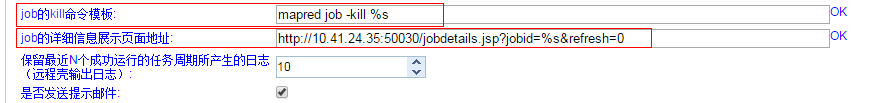
**方案2**：修改所有任务的脚本，将每一个任务的脚本前面加上与任务相对应的注释（比如任务ID），TCC从YARN上根据名称获取相应的jobid。如下图所示



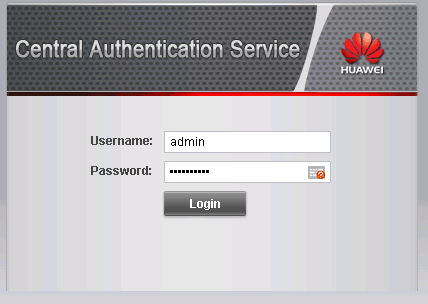
**建议方案**：建议使用方案1，只需要简单的修改，能够很快上线。方案1的缺点是依赖FI的补丁。

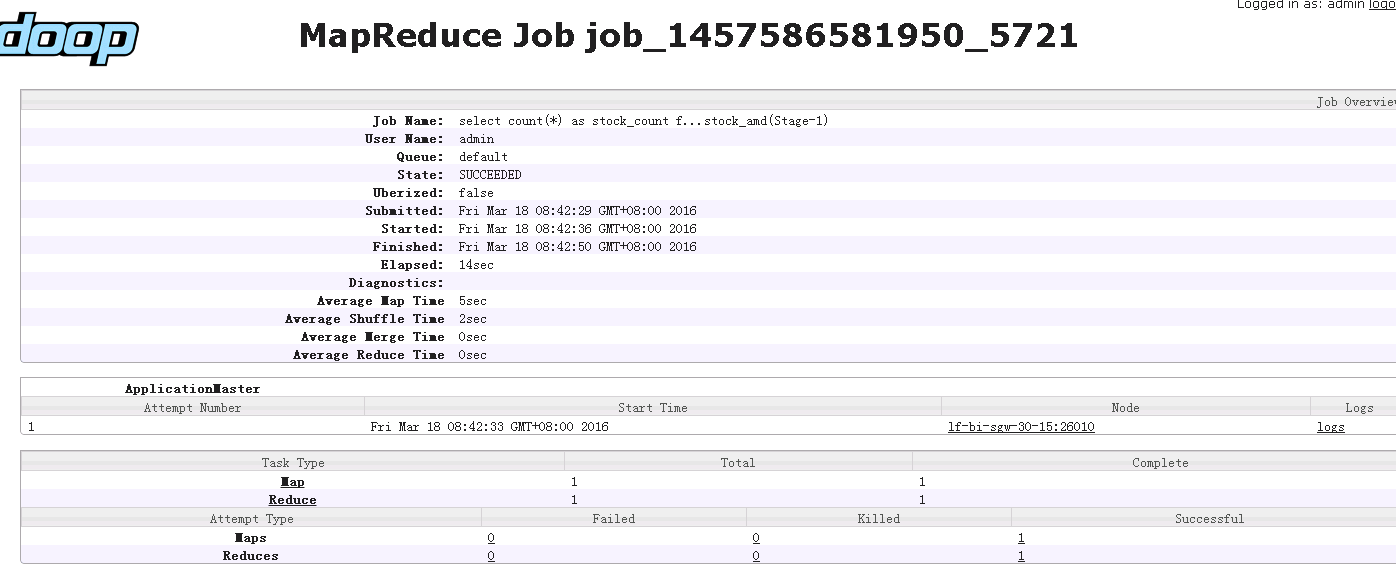
# 失效特性2/特性3的解决方案

当失效特性1解决后，特性2以及特性3的问题基本解决了。剩下的只需要修改一下TCC的链接配置。



当然，YARN的JOB详细信息需要登录后才能看的。如果已经登录，直接进入任务信息界面；没有登录时，会弹出登录界面，输入用户名密码后，会直接跳转到信息信息界面。



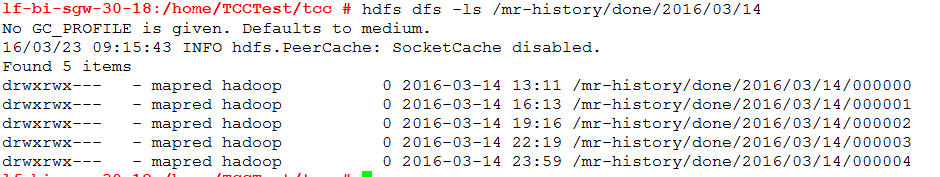


注：要求能够访问TCC的机器，同时也能访问YARN，否则无法正常链接过去。

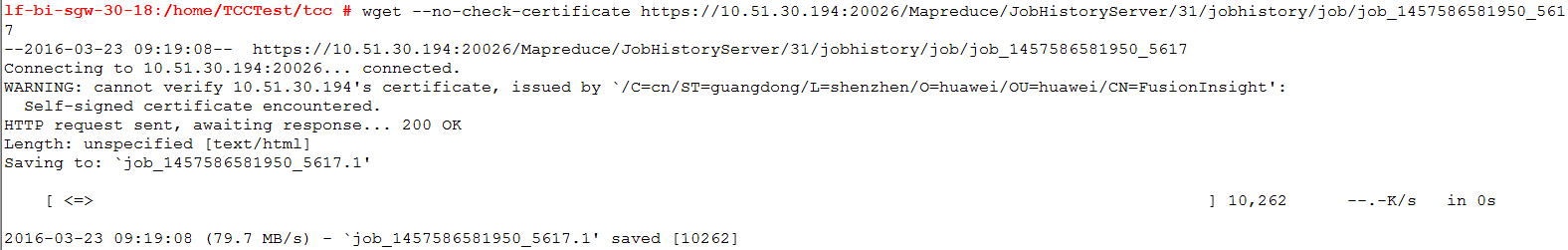
# 特性4的解决方案

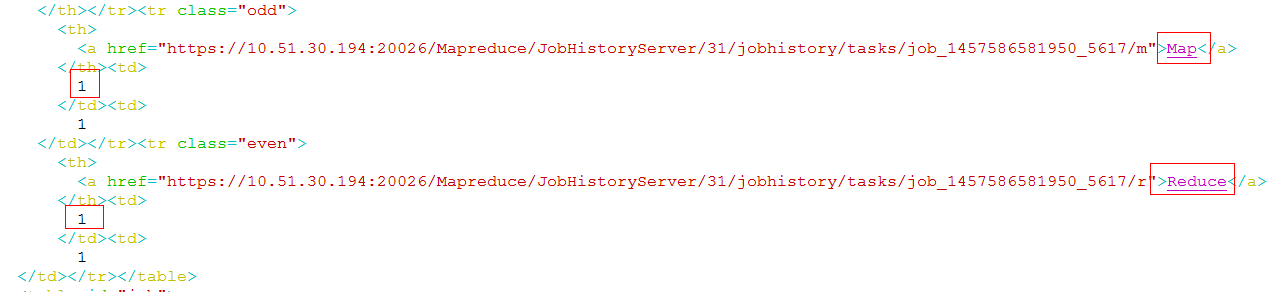
老的TCC方案是单独部署一个程序，去jobtracker服务器解析hadoop的日志信息。并将每个任务所有的job消耗的map/reduce汇总记录到数据库中。然后计算最近一个月的平均消耗数做为一个任务的资源消耗信息。

**方案1**：在tcc上安装FI客户端，通过hdfs命令获取mr-history的日志文件，解析当天的所有日志，然后再根据任务与jobid的对应关系，把每个任务的map/reduce消耗数汇总记录起来。

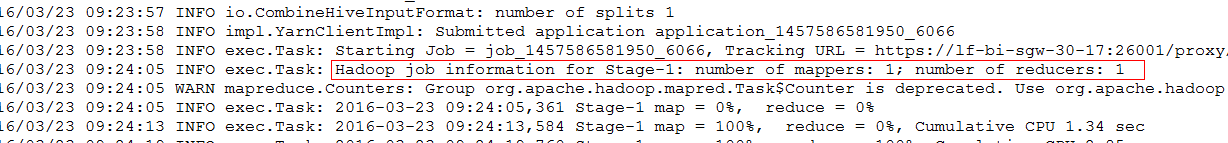


**方案2**：在tcc上安装FI客户端，通过解析YARN任务信息界面，获取jobid对应的map/reduce消耗数。





**方案3**：在任务运行过程中，解析壳日志中输出的map/reduce资源数。



**建议方案**：建议使用方案2。方案1中取分布式的文件太慢了。方案3比较简单，但是会加重TCC的负载，而且依赖于FI的补丁。方案2的工作量主要体现在解析html上，但是代码量不是很大。