

collie 用户使用手册

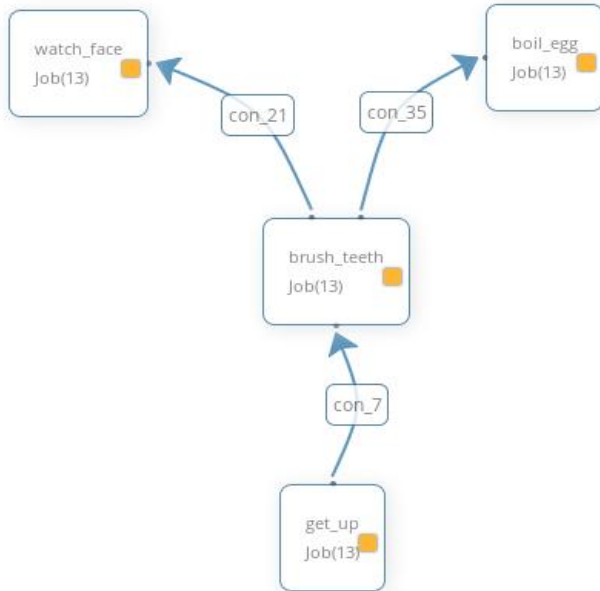
目录

collie 用户使用手册.....	1
目录	1
名词解释	2
1.权限管理	3
1.1创建或加入组	3
1.2Module 的权限	3
1.3Project 的权限	4
2.module	4
2.1创建 module	5
2.2删除和更新模块	7
3.projects	8
3.1创建一个 project	8
3.1.1创建空 project	8
3.1.2创建 new job.....	8
3.1.3建立 job 间的依赖	10
3.2删除/更新/修改 project	10
3.3运行 project	11
3.3.1 Flow View	12
3.3.2 Notification	12
3.3.3 Failure Options	13
3.3.4 Concurrent	13
3.3.5 Hadoop Options	14
3.3.6 Advanced Options.....	15
3.4运行详情查看	16
3.4.1 Graph	16
3.4.2 Job List.....	16
3.4.3 Flow log.....	17
4.Schedule.....	18
5.Executing.....	19
6.History	20

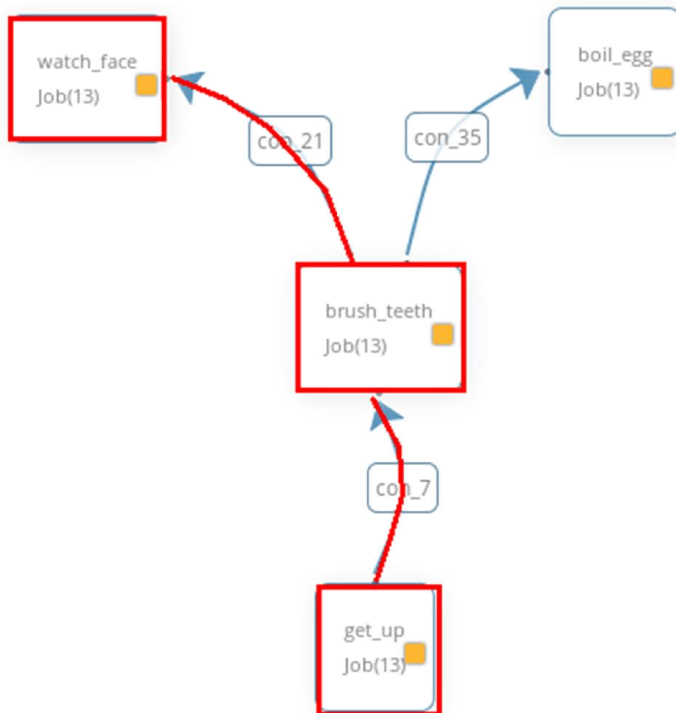
名词解释

collie 平台：提供任务分享和执行调度功能的平台。

Project: 项目，由若干任务组成的集合。其中任务之间有相应的依赖关系（此依赖关系构成了无环路的有向图，如下图）。箭头表示任务执行顺序，如图所示，brush_teeth 依赖 get_up，只有执行完 get_up 才能开始执行 brush_teeth。



Flow: 任务流，以项目中任何一个出度为零的节点为根节点，此节点所依赖的任务（直接依赖，以及间接依赖）。如图，红色标注即表示一个 Flow。



Job: 任务，表示一个独立的执行单元。如上图 get_up 即一个任务。

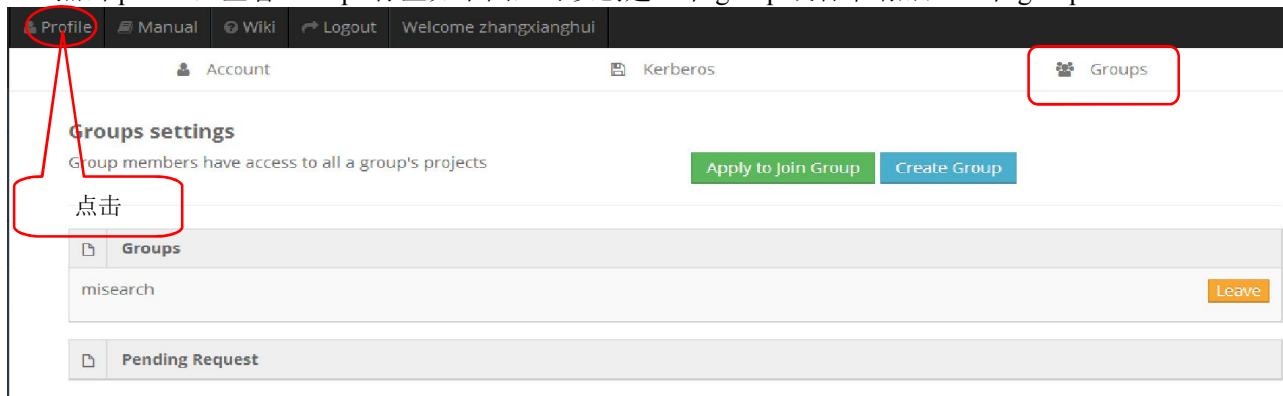
Module: 模块，是对 Job 的抽象，而 Job 是对 Module 参数的具体定制。

1.权限管理

首先是登录界面，使用内网统一认证系统。

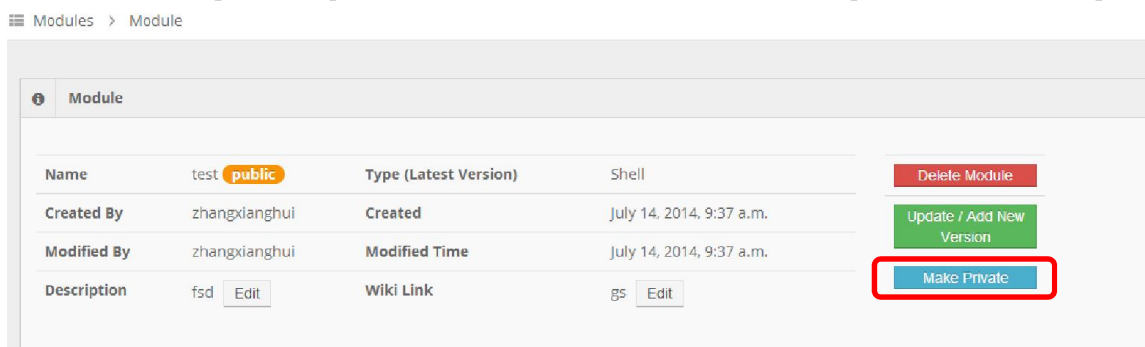
1.1创建或加入组

点击 profile，查看 Groups 标签如下图，可以创建一个 group 或者申请加入一个 group。



1.2Module 的权限

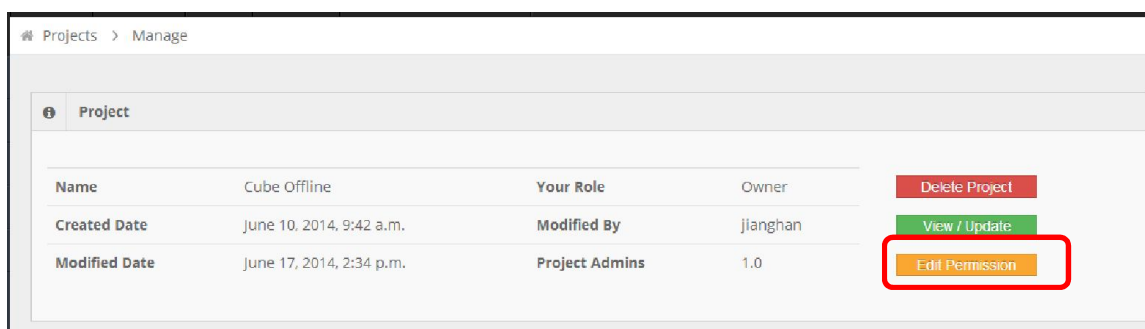
Module 现有 public 和 private 权限，如下图设置 module 的权限，public 所有人可见，private 自己可见。



1.3Project 的权限

可以把 project 授权给某个组，授予其 owner 或者 user 权限，如果是 owner 权限可以删除/更新 project，user 权限只能查看和执行 project。

点击 Edit Permission，进入授权页面，将该 project 的 owner/user 权限授予某个组。



Project Permissions

Following users & groups are allowed to operate on this project with role.

New Permission

Grant Items
<div> <div></div> <div>misearch</div> <div>Owner</div> <div>Delete</div> </div>

2.module

首先上传自己的功能模块，当然也可以不上传模块而使用别人上传的模块。各种类型的 module 创建及参数设置请参考：http://wiki.mioffice.cn/xmg/User_Guide#Module
进入 module 模块，点击 Create Module 创建模块。

The screenshot shows the 'Modules' section of the application. On the left is a sidebar with navigation links: Projects, Scheduled, Executing, History, and Module (which is selected). The main content area has a 'Modules' header with a 'Show 10 entries' dropdown and a search bar. Below this is a table with columns: Name, Id, Created By, Created Date, Modified By, and Modified Date. The table contains one entry: 'Cube Offline' with Id '1', Created By 'jianghan', Created Date 'June 10, 2014, 9:42 a.m.', Modified By 'jianghan', and Modified Date 'June 10, 2014, 9:42 a.m.'. At the bottom of the table are pagination controls: First, Previous, 1 (selected), Next, Last. A red circle highlights the 'Create Module' button in the top right corner, with a callout bubble containing the text '点击创建模块'.

2.1创建 module

目前支持 shell,jar,yarn 类型的模块上传，请将.sh 或.jar 打包成.zip 包后再上传。

Wiki Link:上传模块的功能/参数等说明所在的链接

Select Type:可以选择上传类型 shell，jar 或者 yarn

File upload:选择自己打包后 zip 包上传

shell Options: 首先填写的是主 shell 脚本名，然后填写 shell 脚步的参数，可以设置默认值。目前 shell 脚本只支持短参数。

Hadoop Support 选项，如果选择该选项，shell 脚本中可以使用如下目录下的命令

\$HADOOP_HOME/bin/Hadoop

\$HADOOP_HOME/bin/yarn

\$SPARK_HOME/bin/spark-submit

jar Options:首先是 java 运行入口类，包括包名

可以选填 classpath,使用的最大内存(Xmx)和最小内存(Xms)

然后是 java 运行的参数，可以填写默认值和参数描述。

创建 shell 类型模块:

New Module

Module Name

echo_module

required

Module Description

输出两个words

required

Wiki Link

http://wiki.mioffice.cn/xmg/%E9%A6%96%E9%A1%B5

required

Select Type

shell

File upload

echo.zip

Choose File

Name: echo.zip

Size: 0.31KB

Type: application/zip

Upload

success

Close

Save

New Module

Module Name

echo_module

required

Module Description

输出两个words

required

Wiki Link

http://wiki.mioffice.cn/xmg/User_Guide

optional

Select Type

shell

File upload

No file sel...

Choose File

Upload

Shell Options

☒ Hadoop Support

echo.sh

required (main script)

+ args +

创建 jar 类型模块:

New Module

Module Name

contain_module

required

Module Description

判断字符串中是否包含相应子串

required

Wiki Link

http://wiki.mioffice.cn/xmg/%E9%A6%96%E9%A1%B5

required

Select Type

jar

File upload

search.zip

Choose File

Name: search.zip

Size: 138.27KB

Type: application/zip

Upload

success

Close

Save

New Module

File upload

search.zip

Choose File

Name: search.zip

Size: 138.27KB

Type: application/zip

Upload

success

Jar Options

mi.collie.samples.SearchStrMain

required (java.class)

classpath

optional (classpath)

Xms

optional (Xms)

Xmx

optional (Xmx)

+ jvm.args +

+ main.args +

all

Default Value

Description

sub

Default Value

Description

Close

Save

2.2删除和更新模块

点击自己创建的模块，进入下面的页面，点击 Delete Module，可以删除模块

The screenshot shows the 'Module' page with a table containing the following data:

Name	test	Type (Latest Version)	Shell
Created By	zhangxianghui	Created	July 14, 2014, 9:37 a.m.
Modified By	zhangxianghui	Modified Time	July 14, 2014, 3:23 p.m.
Description	fsd	Wiki Link	gs

Below the table, there are buttons: 'Delete Module' (red), 'Update / Add New Version' (green), and 'Make Private' (blue). Red callouts point to these buttons with the text '点击删除模块' and '点击更新模块'.

点击 Update/Add NewVersion 更新模块，更新模块除了模块名字其他都是可以修改的。

The 'Update Module' form contains the following fields:

- Module Name: test (required)
- Select Type: shell
- File upload: monkey.zip (Choose File) and Upload button
- Shell Options: ☐ Hadoop Support, main (required (main script)), and + args +

At the bottom right, there are 'Close' and 'Save' buttons.

3.projects

进入 Projects 页面，点击 Create Project 创建项目。

The screenshot shows the 'Projects' page with a table containing the following data:

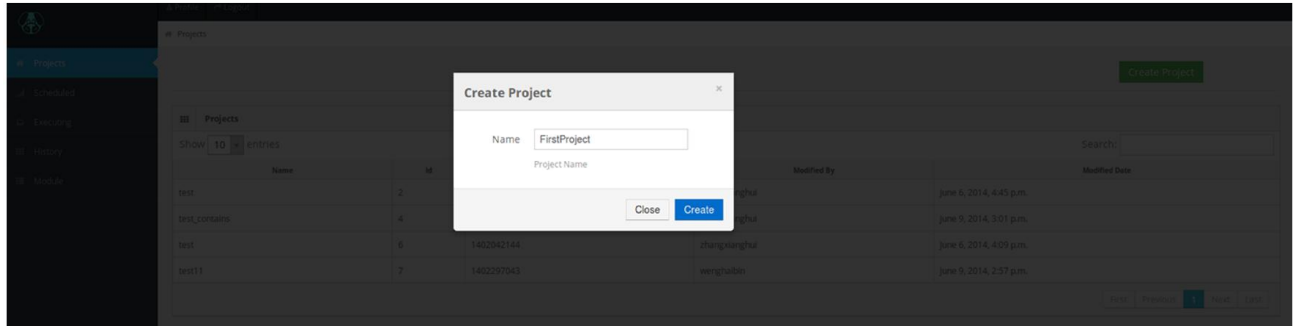
Name	Id	Latest Version	Modified By	Modified Date
test	2	1402044326	zhangxianghui	June 6, 2014, 4:45 p.m.
test_contains	4	1402297287	zhangxianghui	June 9, 2014, 3:01 p.m.
test	6	1402042144	zhangxianghui	June 6, 2014, 4:09 p.m.
test11	7	1402297043	wenghaibin	June 9, 2014, 2:57 p.m.
FirstProject	8	1402302288	zhangxianghui	June 9, 2014, 4:24 p.m.

A red callout points to the 'Create Project' button (green) at the top right of the page with the text '点击创建 project'.

3.1 创建一个 project

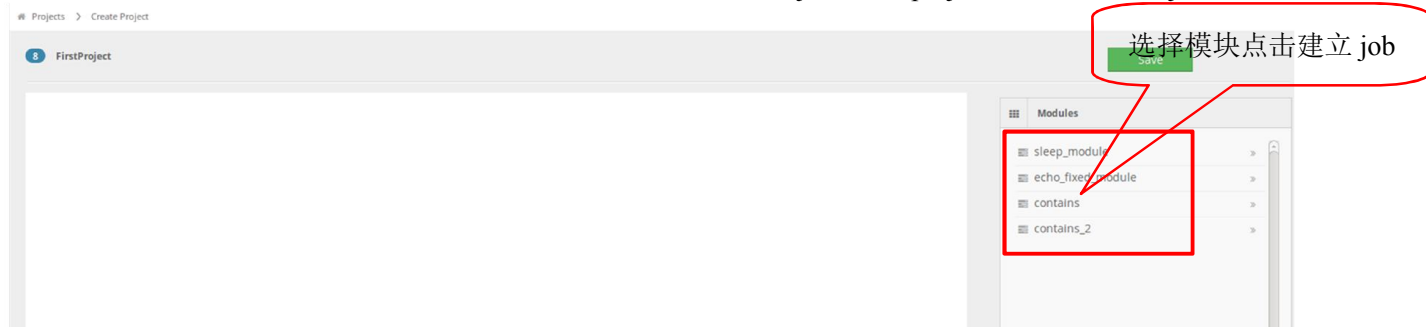
3.1.1 创建空 project

填写 Project Name 点击 Create 即可。



3.1.2 创建 new job

创建一个空的项目后，点击右边希望运行的模块，创建一个 new job，1个 project 可以包含多个 job



创建1个 echo job,写上 job 名字,描述信息以及该模块参数的值，点击 save

New Job

Job Name

echo

Description

echo

Module

echo_fixed_module

Version

3

options

words1

args

→ hello

words2

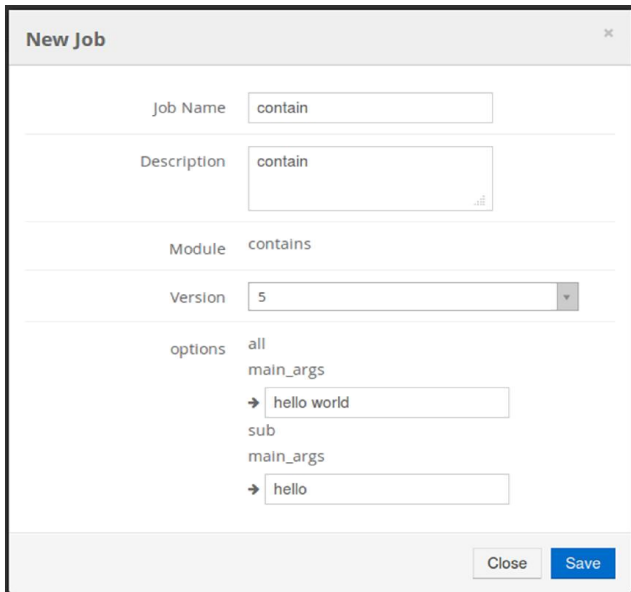
args

→ world

Close

Save

创建一个 cotain job



The 'New Job' form contains the following fields and options:

- Job Name: contain
- Description: contain
- Module: contains
- Version: 5
- options: all
 - main_args: hello world
 - sub: main_args
 - main_args: hello

Buttons: Close, Save

3.1.3建立 job 间的依赖

将新建立的2个 job 连接起来，这样创建 job 的依赖关系，echo job 运行完成后开始运行 contain job. 点击保存，一个 project 即建立完成。



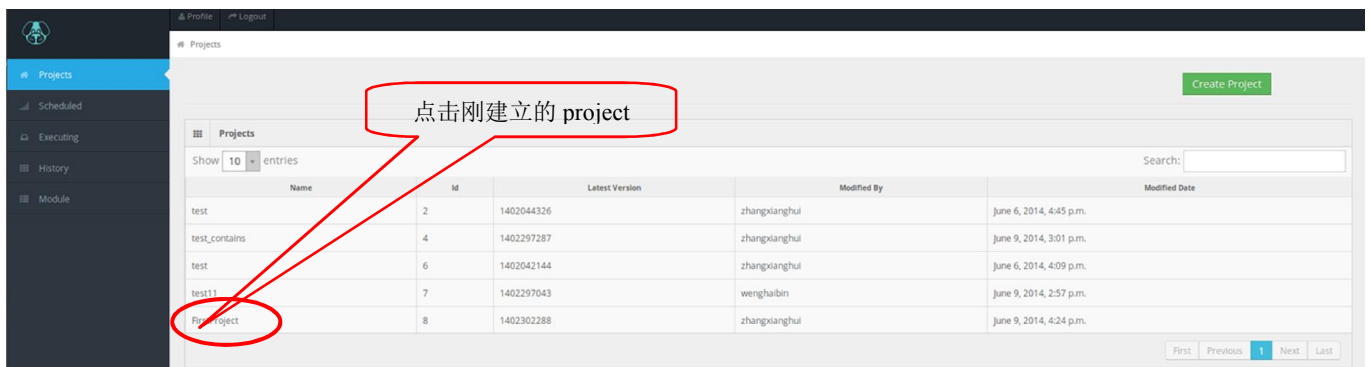
The interface shows a dependency graph where 'echo' (echo_fixed_module3) points to 'contain' (contains2). A red circle highlights the 'Save' button with the text '点击保存，project 建立完成'.

Modules list:

- sleep_module
- echo_fixed_module
- contains
- contains_2

3.2删除/更新/修改 project

回到 projects 页面，找到你刚才建立的 project, 点击进入。

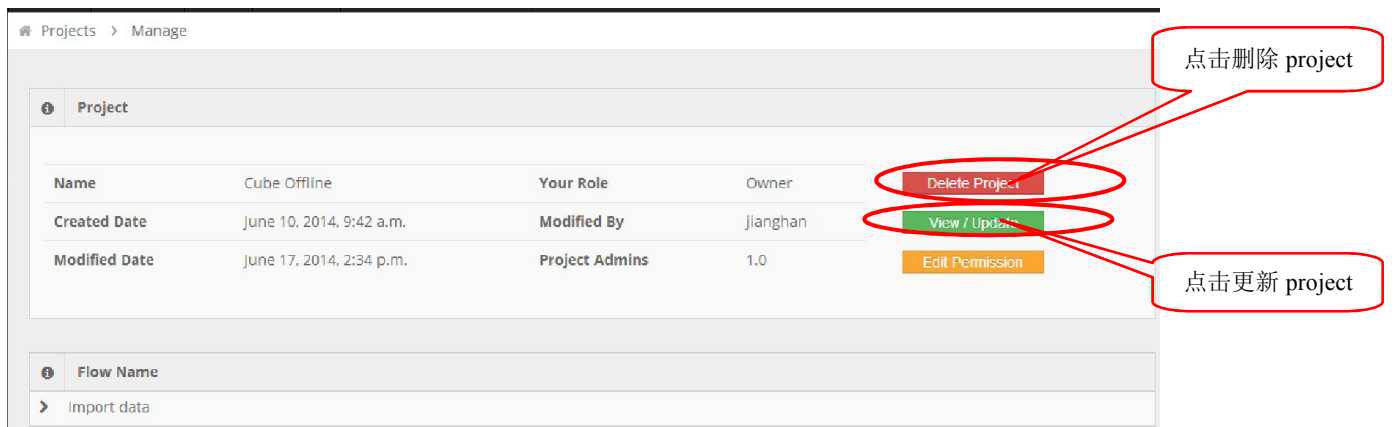


The 'Projects' table lists the following data:

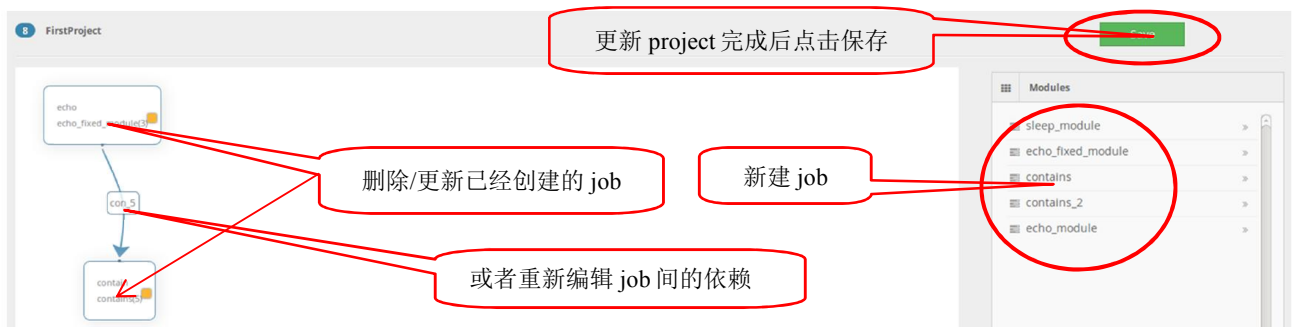
Name	Id	Latest Version	Modified By	Modified Date
test	2	1402044326	zhangxianghui	June 6, 2014, 4:45 p.m.
test_contains	4	1402297287	zhangxianghui	June 9, 2014, 3:01 p.m.
test	6	1402042144	zhangxianghui	June 6, 2014, 4:09 p.m.
test1	7	1402297043	wenghaibin	June 9, 2014, 2:57 p.m.
FirstProject	8	1402302288	zhangxianghui	June 9, 2014, 4:24 p.m.

A red circle highlights the 'FirstProject' row with the text '点击刚建立的 project'.

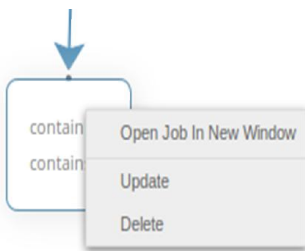
进入建立的 project 后，可以点击 Delete Project 删除项目，点击 View/Update 更新修改 project。



点击 View/Update,进入 project 的编辑页面，可以点击右边的模块新建模块，或者点击已经建立的模块删除模块/或者点击 update 更新模块，然后点击 save 更新完毕。



右键点击建立的 job，可以更新/删除 job，新窗口打开 job 功能还没有实现。



点击 Upate 进入的 job 更新页面

Update Job

Job Name

contain

Description

contain

Module

contains

Version

5

options

all

main_args

→ hello world

sub

main_args

→ hello

Close

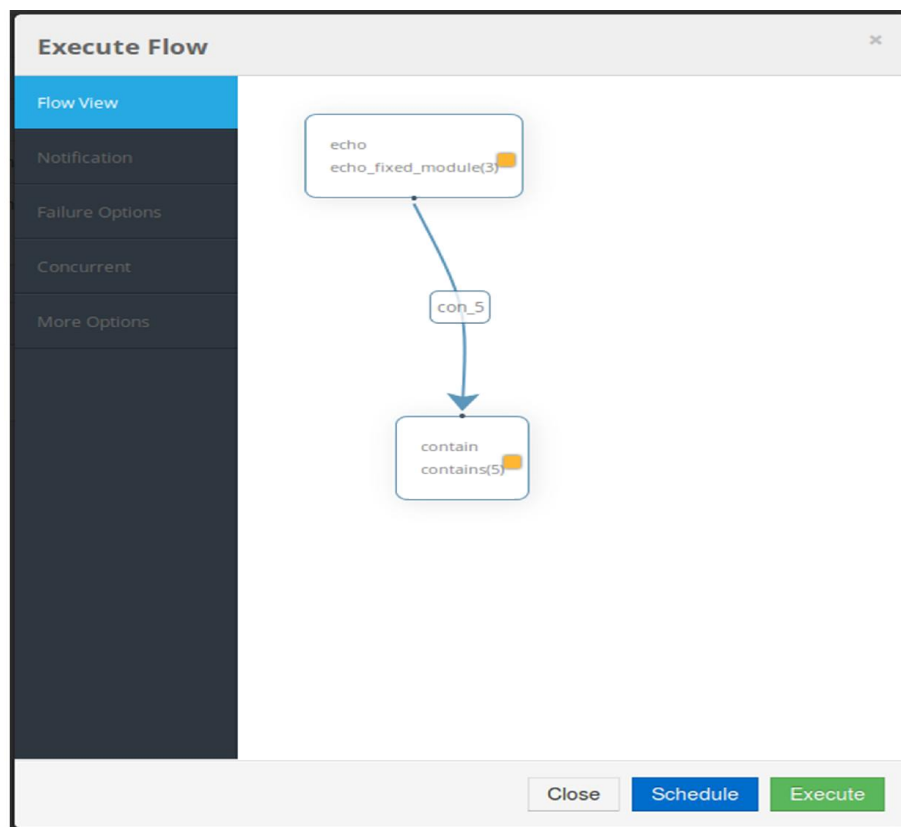
Save

3.3运行 project

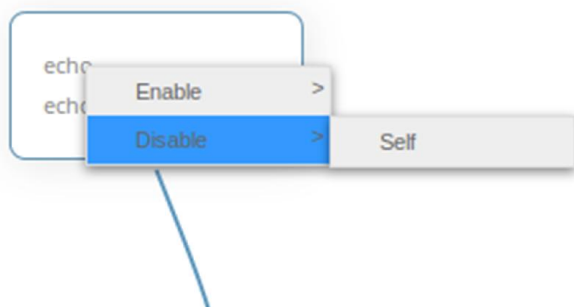
进入创建的 project,点击 job 右边的 Run,进入运行页面。



3.3.1 Flow View



右键点击 job, 可以使 job Disable, 运行时就不再执行, 点击 Enable 激活 job.



3.3.2 Notification

First Failure:第一次检测到失败后立即发送邮件

Flow Finished:如果有 job 失败，整个项目所有 job 执行完后再发送邮件

下面填写失败/成功邮件发送到的邮箱，可以写多个，用逗号/空格/分号分割。

The screenshot shows the 'Execute Flow' dialog box with the 'Notification' tab selected. The left sidebar contains options: Flow View, Notification (selected), Failure Options, Concurrent, and More Options. The main area is divided into two sections: 'Notify On Failure' and 'Success Emails'. Under 'Notify On Failure', there are checkboxes for 'First Failure' and 'Flow Finished', both of which are checked. Below this is the 'Failure Emails' section, which has a checkbox for 'Override flow email settings' (checked) and a text input field containing 'zhangxianghui@xiaomi.com'. The 'Success Emails' section also has a checkbox for 'Override flow email settings' (checked) and a text input field containing 'zhangxianghui@xiaomi.com'. At the bottom right, there are three buttons: 'Close', 'Schedule', and 'Execute'.

3.3.3 Failure Options

Finish Current Running:如果有 job 失败，运行完当前正在运行的所有 job 后停止。

Cancel All: 如果有 job 失败，立即停止该项目所有 job 的运行。

Finish All Possible: 如果有 job 失败，依然继续运行该项目，直到不能再运行为止。

The screenshot shows the 'Execute Flow' dialog box with the 'Failure Options' tab selected. The left sidebar contains options: Flow View, Notification, Failure Options (selected), Concurrent, and More Options. The main area is titled 'Failure Options' and features a dropdown menu currently set to 'Finish All Possible'. Below the dropdown, there is explanatory text: 'When a failure first occurs in the flow, select the execution behavior.' followed by three radio button options: 'Finish Current Running' (selected), 'Cancel All', and 'Finish All Possible'. Each option has a brief description of its behavior. At the bottom right, there are three buttons: 'Close', 'Schedule', and 'Execute'.

3.3.4 Concurrent

如果该项目已经在运行了，如果再次点击运行有下面的设置选择。

Skip Execution: 如果已经在运行了，即不再运行

Run Concurrently : 如果已经在运行了，还可以并行运行，互相不影响。

Pipeline: 串行运行

Level 1: 本进程中的 job A 必须等到并行进程中的 job A 运行完成后才能运行。

Level 2: 本进程中的 job A 等到并行进程中的 job A 的子 job 运行完成后即可开始运行。

The screenshot shows a dialog box titled "Execute Flow" with a sidebar on the left containing "Flow View", "Notification", "Failure Options", "Concurrent" (highlighted), and "More Options". The main area is titled "Concurrent Execution Options" and contains the following text: "If the flow is currently running, these are the options that can be set." Below this are three radio buttons: "Skip Execution" (unselected), "Run Concurrently" (selected), and "Pipeline" (unselected). Under "Run Concurrently" is a text box with "Level 1" and a dropdown arrow. Under "Pipeline" is a text box with "Level 1" and a dropdown arrow. Below these are two checkboxes: "Level 1: block job A until the previous flow job A has completed." (checked) and "Level 2: block job A until the previous flow job A's children have completed." (checked). At the bottom of the dialog are three buttons: "Close", "Schedule", and "Execute".

3.3.5 Hadoop Options

Hadoop Options 仅在 shell (hadoop support 为 true)、yarn、spark 类型下才出现，其他情形不会出现。

Kerberos Key Options: 选择 kerberos key，可以点击上方的 Profile,然后点击 Kerberos 创建。

Mapred Queue Options: 选定 Queue

The screenshot shows a web interface with a top navigation bar containing "Profile" and "Logout". Below this is a header with "Account", "Kerberos", and "Groups". The main content area is titled "My Kerberos keys" and contains the text "Kerberos keys allow you to submit job on hadoop". On the right side of this text is a green button labeled "New Key". Below this is a table with one row containing a trash icon and the text "Kerberos Keys".

Execute Flow

Flow View

Notification

Failure Options

Concurrent

Hadoop Options

Advanced Options

Kerberos Key Options

Mapred Queue Options

queue name

required

Close

Schedule

Execute

3.3.6 Advanced Options

Running Node Options:可以选择在哪台机器上运行，该选项默认是隐藏的。

Execute Flow

Flow View

Notification

Failure Options

Concurrent

Advanced Options

Running Node Options

Random

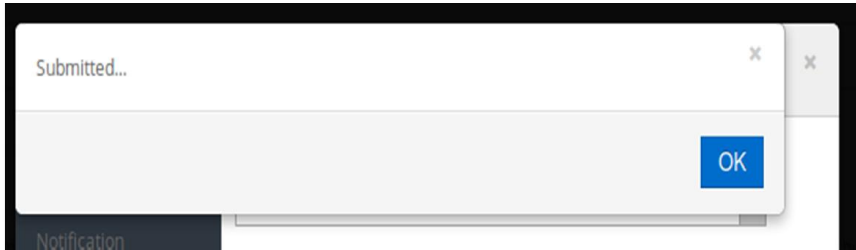
Hide Advanced Options

Close

Schedule

Execute

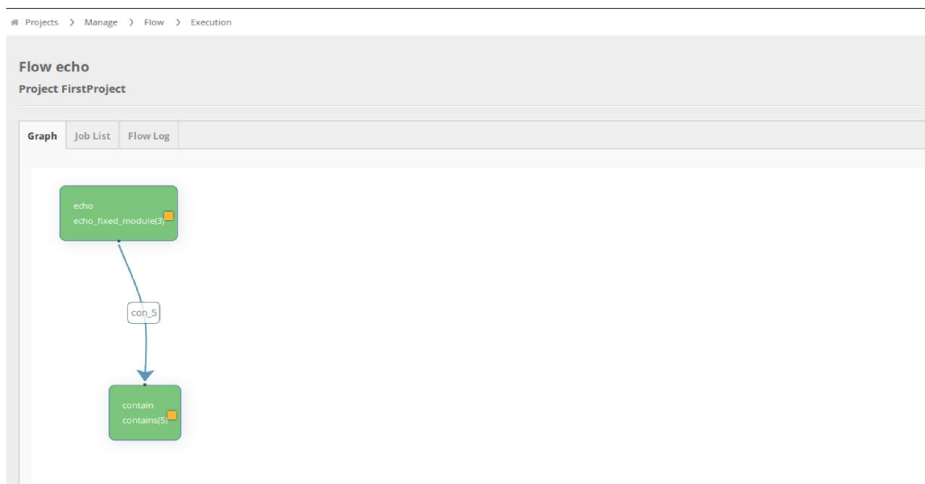
点击 Execute,点击 ok,进入运行页面。



3.4运行详情查看

3.4.1 Graph

进入运行页面如下, job 运行完成后该 job 节点变成绿色。



3.4.2 Job List

按照执行顺序显示 job,并显示开始时间/结束时间/状态等, 点击后面的 log,可以查看 job 的执行日志。

Projects > Manage > Flow > Execution

Flow echo
Project FirstProject

Graph Job List Flow Log

Show 10 entries

Name	Timeline	Start Time	End Time	Elapsed	Status	Logs
contain		June 9, 2014, 5:37 p.m.	June 9, 2014, 5:37 p.m.	0:00:00	SUCCEEDED	Log
echo		June 9, 2014, 5:37 p.m.	June 9, 2014, 5:37 p.m.	0:00:00	SUCCEEDED	Log

First Previous 1 Next Last

点击查看 job 日志

点击 contain 后面的 log, 进行日志查看页面。

ProfileLogout

Projects > Manage > Flow > Execution > Job

Project FirstProject > echo > contain

Refresh

2014/06/09 17:37:34 [INFO] Starting to run JobRunner.
2014/06/09 17:37:34 [INFO] Prepare to create job contain
2014/06/09 17:37:34 [INFO] Job contain's type is java
2014/06/09 17:37:34 [INFO] Successfully created java job for contain
2014/06/09 17:37:34 [INFO] Start to run job contain.
2014/06/09 17:37:34 [INFO] Prepare to build process:
command:[java, -Xms64M, -Xmx256M, -cp, SearchStr-1.0-SNAPSHOT-jar-with-dependencies.jar, com.xiaomi.colliie.samples.SearchStrMain, -sub, hello, -all, hello world]
directory:/home/shixin/workspace/collixe/release/executions/20
environment parameters:{}
2014/06/09 17:37:34 [INFO] Spawned thread with process id 13052
2014/06/09 17:37:34 [INFO] Start to run process 13052
2014/06/09 17:37:34 [INFO] all string is:hello world
2014/06/09 17:37:34 [INFO] sub string is:hello
2014/06/09 17:37:34 [INFO] true
2014/06/09 17:37:34 [INFO] Finish process 13052 with exit code 0
2014/06/09 17:37:34 [INFO] Job contain completed successfully in 0

3.4.3 Flow log

点击 Flow Log，查看流程的运行日志。

Projects > Manage > Flow > Execution

Flow echo

Project FirstProject

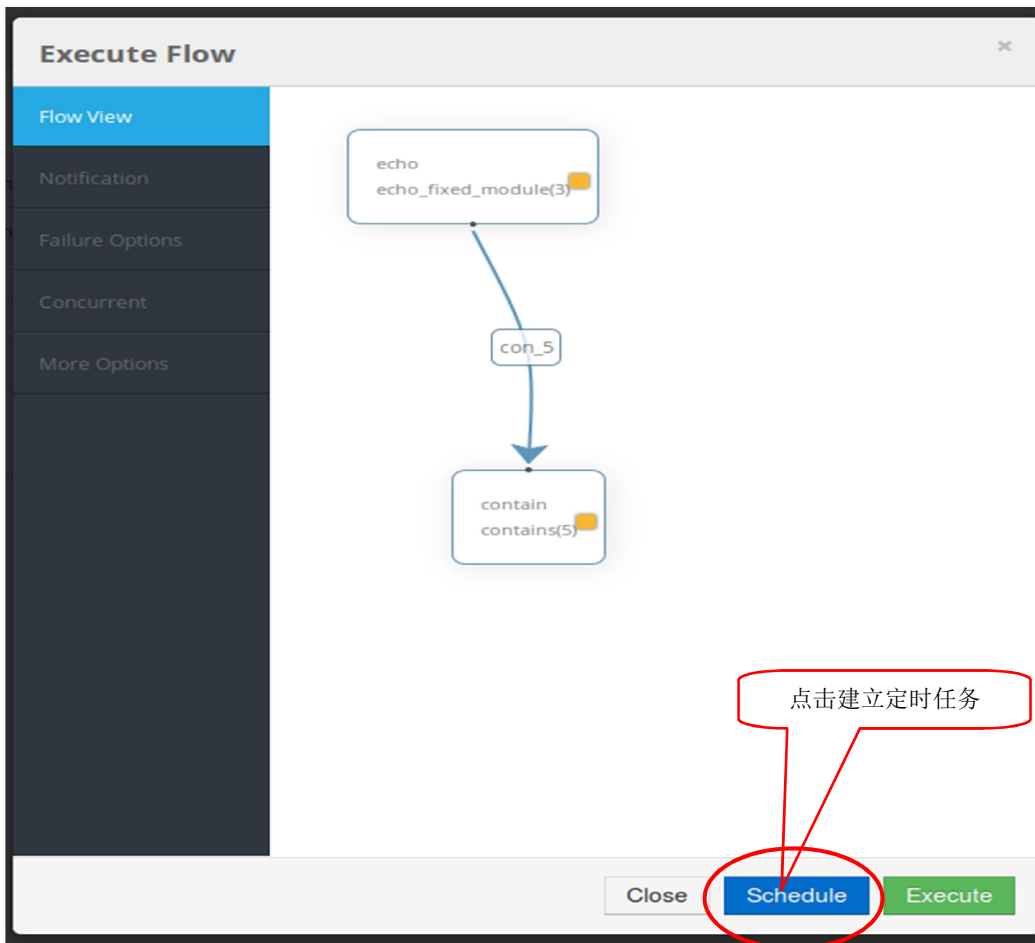
GraphJob ListFlow Log

Refresh

2014/06/09 17:37:34 [INFO] Starting flow runner : 20
2014/06/09 17:37:34 [INFO] Found job contain ready to run!
2014/06/09 17:37:34 [INFO] Try to create job runner for contain
2014/06/09 17:37:34 [INFO] Constructed job runner properties:
2014/06/09 17:37:34 [INFO] Key: path.yarn, value: /opt/soft/hadoop/bin
2014/06/09 17:37:34 [INFO] Key: type, value: java
2014/06/09 17:37:34 [INFO] Key: main.args.all, value: hello world
2014/06/09 17:37:34 [INFO] Key: path.hadoop, value: /opt/soft/hadoop/bin
2014/06/09 17:37:34 [INFO] Key: main.args.sub, value: hello
2014/06/09 17:37:34 [INFO] Key: java.class, value: com.xiaomi.colliie.samples.SearchStrMain
2014/06/09 17:37:34 [INFO] Submitting job contain to run!
2014/06/09 17:37:34 [INFO] Flow will wait for 300000 milliseconds.
2014/06/09 17:37:34 [INFO] Starting JobRunner contain with job status: QUEUED
2014/06/09 17:37:34 [INFO] Handle JobRunner started event, jobld : contain
2014/06/09 17:37:34 [INFO] JobEvent: job contain run succeeded!
2014/06/09 17:37:34 [INFO] Handle JobRunner finished event: contain, SUCCEEDED
2014/06/09 17:37:34 [INFO] Remove job contain from activeJobRunners.
2014/06/09 17:37:34 [INFO] Found job echo ready to run!
2014/06/09 17:37:34 [INFO] Try to create job runner for echo
2014/06/09 17:37:34 [INFO] Constructed job runner properties:
2014/06/09 17:37:34 [INFO] Key: path.yarn, value: /opt/soft/hadoop/bin
2014/06/09 17:37:34 [INFO] Key: dependencies, value: contain
2014/06/09 17:37:34 [INFO] Key: command, value: sh echo.sh -words2 "world" -words1 "hello"
2014/06/09 17:37:34 [INFO] Key: type, value: shell
2014/06/09 17:37:34 [INFO] Key: path.hadoop, value: /opt/soft/hadoop/bin
2014/06/09 17:37:34 [INFO] Submitting job echo to run!
2014/06/09 17:37:34 [INFO] Flow will wait for 300000 milliseconds.
2014/06/09 17:37:34 [INFO] Starting JobRunner echo with job status: QUEUED
2014/06/09 17:37:34 [INFO] Handle JobRunner started event, jobld : echo
2014/06/09 17:37:34 [INFO] Finishing job contain
2014/06/09 17:37:34 [INFO] JobEvent: job echo run succeeded!
2014/06/09 17:37:34 [INFO] Handle JobRunner finished event: echo, SUCCEEDED
2014/06/09 17:37:34 [INFO] Remove job echo from activeJobRunners.
2014/06/09 17:37:34 [INFO] Flow was finished!
2014/06/09 17:37:34 [INFO] Flow status is set to SUCCEEDED
2014/06/09 17:37:34 [INFO] Finishing job echo

4.Schedule

进入 Execute Flow 页面，点击下方的 Schedule.



进入定时任务设置页面

Schedule Time: 定时任务的执行时间

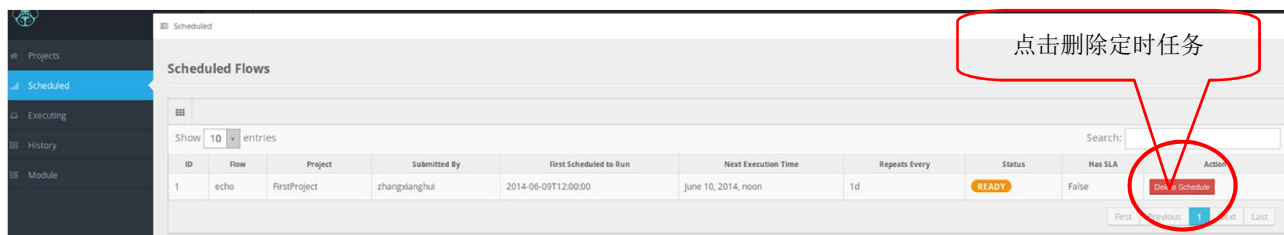
Schedule Date: 设置定时任务的时间

Recurrence: 多久重复执行一次。

The screenshot shows the 'Schedule Flow Options' form. It has three main sections: 'Schedule Time' with fields for '12', '00', 'am', and 'CST'; 'Schedule Date' with a field for '06/09/2014' and a calendar icon; and 'Recurrence' with a checked checkbox, the text 'repeat every', a field for '1', and a dropdown for 'Days'. At the bottom right are 'Close' and 'Schedule' buttons.

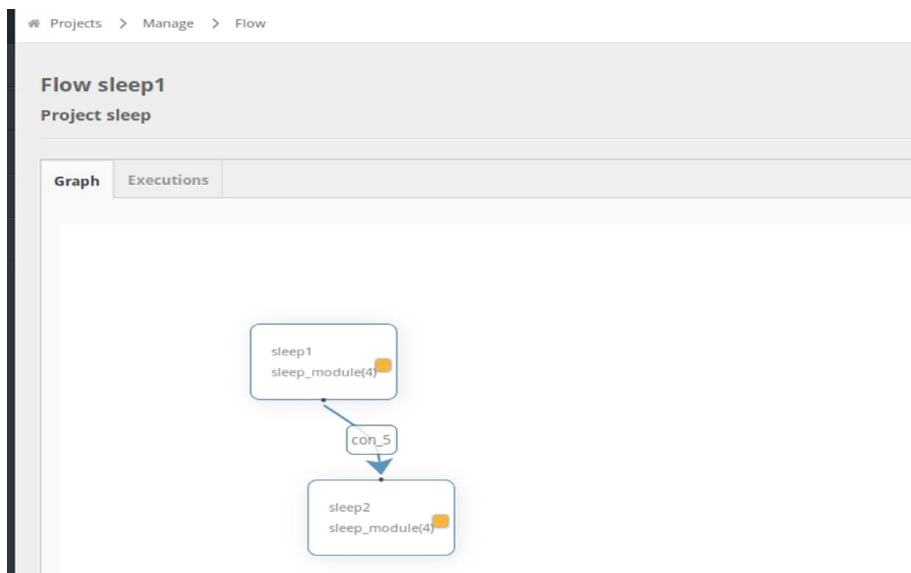
点击 Schedule，进入到 Schedule 页面。

Scheduled Flows 下面可以看到刚才建立的定时任务。



点击 Delete Schedule,删除定时任务。

点击 Flow 进入下面的页面



再次点击 Executions,进入下面的页面

Execution Id	User	Start Time	End Time	Elapsed	Status	Action
23	zhangxianghui	June 10, 2014, 4:18 a.m.	None	None	RUNNING	
22	zhangxianghui	June 9, 2014, 8:14 p.m.	June 9, 2014, 8:14 p.m.	0:00:36	SUCCEEDED	

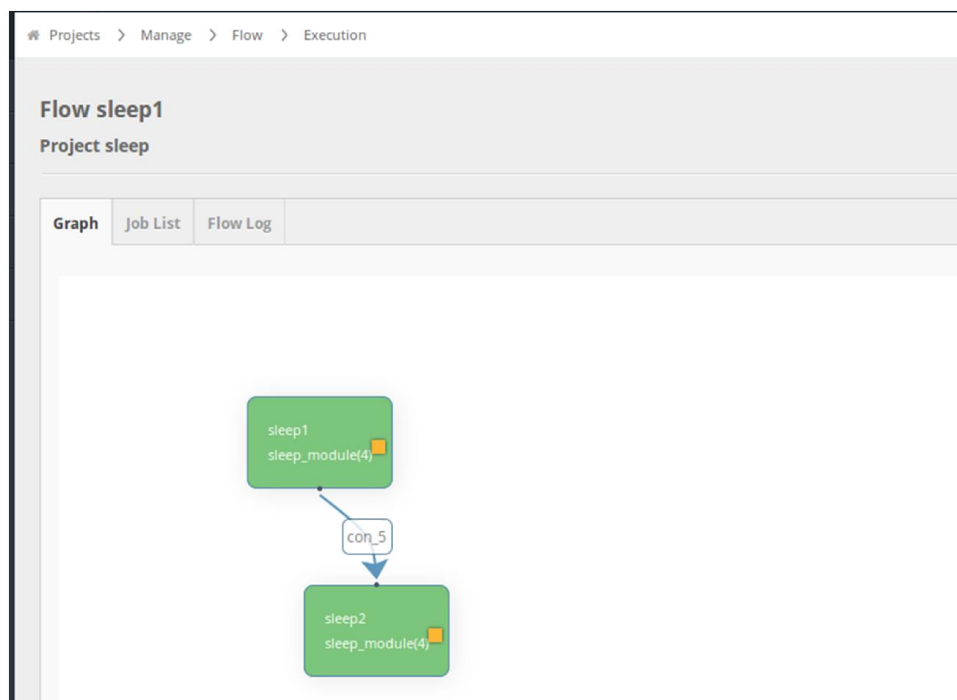
点击 Execution Id, 可以进入日志查看页面。

5.Executing

Executing 中显示当前正在运行的任务

Execution Id	Flow	Project	User	Proxy User	Start Time	End Time	Elapsed	Status	Action
22	sleep1	sleep	zhangxianghui	zhangxianghui	June 9, 2014, 8:14 p.m.	-	-	Running	

点击 Execution Id, 能查看 job 的日志和整个项目的日志



6. History

查看运行历史记录，其中 ID，Flow 和 Project 可以点击，可以查看日志/ workflows图/project 等信息。

History

Show 10 entries

Search:

ID	Flow	Project	User	Start Time	End Time	Elapsed	Status
24	sleep1	sleep	zhangxianghui	June 9, 2014, 8:21 p.m.	June 9, 2014, 8:22 p.m.	0:00:36	SUCCEEDED
23	sleep1	sleep	zhangxianghui	June 9, 2014, 8:18 p.m.	June 9, 2014, 8:19 p.m.	0:00:37	SUCCEEDED
22	sleep1	sleep	zhangxianghui	June 9, 2014, 8:14 p.m.	June 9, 2014, 8:14 p.m.	0:00:36	SUCCEEDED
21	echo	FirstProject	zhangxianghui	June 9, 2014, 5:48 p.m.	June 9, 2014, 5:48 p.m.	0:00:01	SUCCEEDED
20	echo	FirstProject	zhangxianghui	June 9, 2014, 5:37 p.m.	June 9, 2014, 5:37 p.m.	0:00:00	SUCCEEDED
19	echo	FirstProject	zhangxianghui	June 9, 2014, 5 p.m.	June 9, 2014, 5 p.m.	0:00:00	SUCCEEDED
18	echo	FirstProject	zhangxianghui	June 9, 2014, 4:55 p.m.	June 9, 2014, 4:55 p.m.	0:00:00	SUCCEEDED
17	echo	FirstProject	zhangxianghui	June 9, 2014, 4:11 p.m.	June 9, 2014, 4:11 p.m.	0:00:01	SUCCEEDED
16	echo	test11	zhangxianghui	June 9, 2014, 3:03 p.m.	June 9, 2014, 3:03 p.m.	0:00:00	SUCCEEDED
15	contains_job	test_contains	zhangxianghui	June 9, 2014, 2:59 p.m.	June 9, 2014, 2:59 p.m.	0:00:00	SUCCEEDED

First Previous 1 2 3 Next Last