

Zeus3的使用与二次开发

Andrew Cai

CONTENTS

1. Zeus3介绍
2. Zeus3的安装与部署
3. Zeus3的使用
4. Zeus3的改进之处
5. 以上工程代码下载与Q&A

Zeus3介绍

1. Zeus3介绍

宙斯是一个完整的Hadoop的作业平台从Hadoop任务的调试运行到生产任务的周期调度宙斯支持任务的整个生命周期

从功能上来说，支持：

- HadoopMapReduce任务的调试运行

- Hive任务的调试运行

- Shell任务的运行

- Hive元数据的可视化查询与数据预览

- Hadoop任务的自动调度

- 完整的文档管理

Andrew.cai在Zeus2的基础上做二次开发，历时2周，并完善使用文档贡献给大家。

Zeus3的安装与部署

[1]配置目录权限

```
mkdir -p /data/zeus/log  
sudo -u hdfs hadoop fs -mkdir /zeus  
sudo -u hdfs hadoop fs -mkdir /zeus/hdfs-upload-dir  
mkdir -p /data/zeus/users/zhoufang/zeus/run_job_dir  
mkdir -p /data/zeus/group/tbdataapplication/zhoufang  
mkdir -p /data/zeus/users/zhoufang/zeus/logs/
```

[2]配置文件修改

- 2.1 zeus-web.properties //新增加的配置文件
- 2.2 persistence.xml
- 2.3 log4j.xml
- 2.4 broadcast-applicationContext.xml

[3]zeus的table创建文件

zeus3.sql
先创建zeus3数据库

[4]启动

通过tomcat起来，tomcat的内存请开大，如：12G

[5]浏览器

Zeus只支持firefox,chrome不支持IE

Zeus3的安装与部署

[6]自动发邮件设置

zeus-web.properties

#mail

mail.host=smtp.163.com

mail.smtp.auth=false

mail.smtp.timeout=25000

mail.username=xxx@163.com

mail.password=xxx

Zeus3的使用

[1]调试Hive

我的文档

- 个人文档
 - Zack
 - andy.cai
 - leo
- 共享文档

运行 运行选中代码 同步任务 扩展功能 全局配置

[12] 文件 2.hive [1] CRM投诉明细报表.hive [16] 文件 3.hive [11] 文件 1.hive [15] 文件 1.sh

```
1 show databases;  
2  
3
```

17 个字符 已保存

公共文档

编辑 调试历史

调试文件，按姓名建立目录

调试Hive，默认是default库，查询其他库请用use或者xx.tabley

运行选定代码，注意：选定后要等运行按钮显示出来才能执行。否则显示NULL

Zeus3的使用

[1]调试Hive

调试结果：

调试的历史记录
很多场合这块显示记录更全面，如MR的过程log等...

显示的结果

```
1 show databases;  
2  
3  
4
```

CONSOLE# default
CONSOLE# insurance_calc
CONSOLE#
CONSOLE#
CONSOLE#
CONSOLE#
CONSOLE# ph
CONSOLE# st
CONSOLE# s
CONSOLE# t
CONSOLE# test2
CONSOLE# test_dawn
CONSOLE# vms_gps_db
CONSOLE# zeus_test
CONSOLE# Time taken: 3.045 seconds
ZEUS# 核心Job任务处理完毕
ZEUS# exitCode=0

ID	执行机器IP	状态	开始时间	结束时间	操作
112	10.0.137.51	success	2014-11-27 10:58	2014-11-27 10:58	查看日志

日志

```
ZEUS# 2014-11-27 10:58:43 进入任务队列  
ZEUS# 2014-11-27 10:58:45 开始运行  
ZEUS# 开始执行前置处理单元: DownloadJob  
ZEUS# 前置处理单元: DownloadJob 处理完毕  
ZEUS# 开始执行核心Job任务  
ZEUS# DEBUG Command:hive -f /data/zeus/users/zhoufang/zeus/run_job_dir/2014-11-27/debug-112/1417057125531.hive  
CONSOLE# Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties  
CONSOLE# Hive history file=/tmp/root/hive_job_log_dd25d50f-969c-45e3-0-a3f9ff75ec10_1405795299.txt  
CONSOLE# OK  
CONSOLE# cbsdb1  
CONSOLE# cbsdb4  
CONSOLE# ccmdw  
CONSOLE# crm_complain_dashboard  
CONSOLE# dawn  
CONSOLE# default  
CONSOLE# insurance_calc  
CONSOLE# kettledb  
CONSOLE# member  
CONSOLE# newccs
```

Page 1 of 2
Displaying 1 - 20 of 25

Zeus3的使用

[2]调试Shell

调试页面，大体同调试Hive的界面：

The image displays two screenshots of the Zeus3 IDE interface, specifically the Shell debugging window. The left screenshot shows the initial state with a script being edited. The right screenshot shows the execution log with a table of task details and a detailed log of commands and output.

Left Screenshot: Script Editor

```
1 #!/bin/bash
2 date1=`date +%Y_%m`
3 echo 'hello'
4 echo $date1
5
```

Right Screenshot: Execution Log

ID	执行机器IP	状态	开始时间	结束时间	操作
116	10.0.137.51	success	2014-11-27 11:10	2014-11-27 11:10	查看日志

日志

```
ZEUS# 2014-11-27 11:10:42 进入任务队列
ZEUS# 2014-11-27 11:10:45 开始运行
ZEUS# 开始执行前置处理单元: DownloadJob
ZEUS# 前置处理单元: DownloadJob 处理完毕
ZEUS# 开始执行核心Job任务
ZEUS# DEBUG Command:chmod u+x /data/zeus/us
116/1417057845523. sh
ZEUS# DEBUG Command:dos2unix /data/zeus/us
116/1417057845523. sh
CONSOLE# dos2unix: converting file /data/ze
116/1417057845523. sh to Unix format...
ZEUS# DEBUG Command:sh /data/zeus/users/zho
CONSOLE# hello
CONSOLE# 2014_11
ZEUS# 核心Job任务处理完毕
ZEUS# exitCode=0
```

脚本

```
#!/bin/bash
```

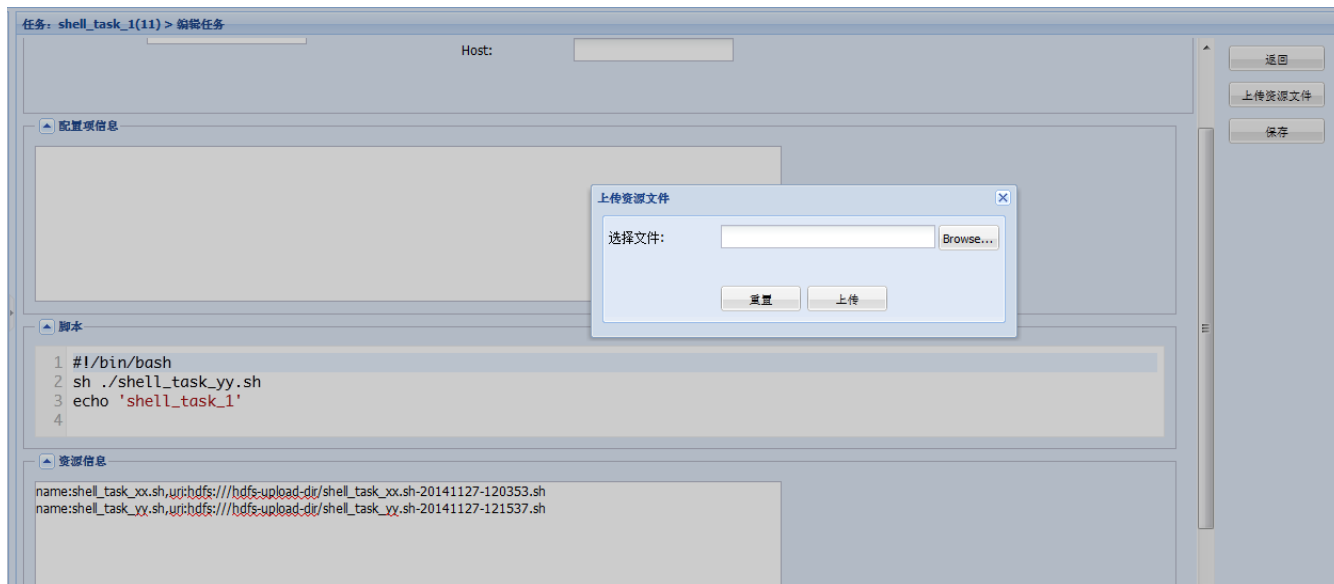
Page 1 of 1 | Displaying 1 - 7 of 7

Zeus3的使用

[3]调度Shell

场景1：shell_task_1.sh 内部调用shell_task_yy.sh

Step1:上传shell_task_yy.sh作为资源



Step2:编辑调用

```
#!/bin/bash
```

```
sh ./shell_task_yy.sh
```

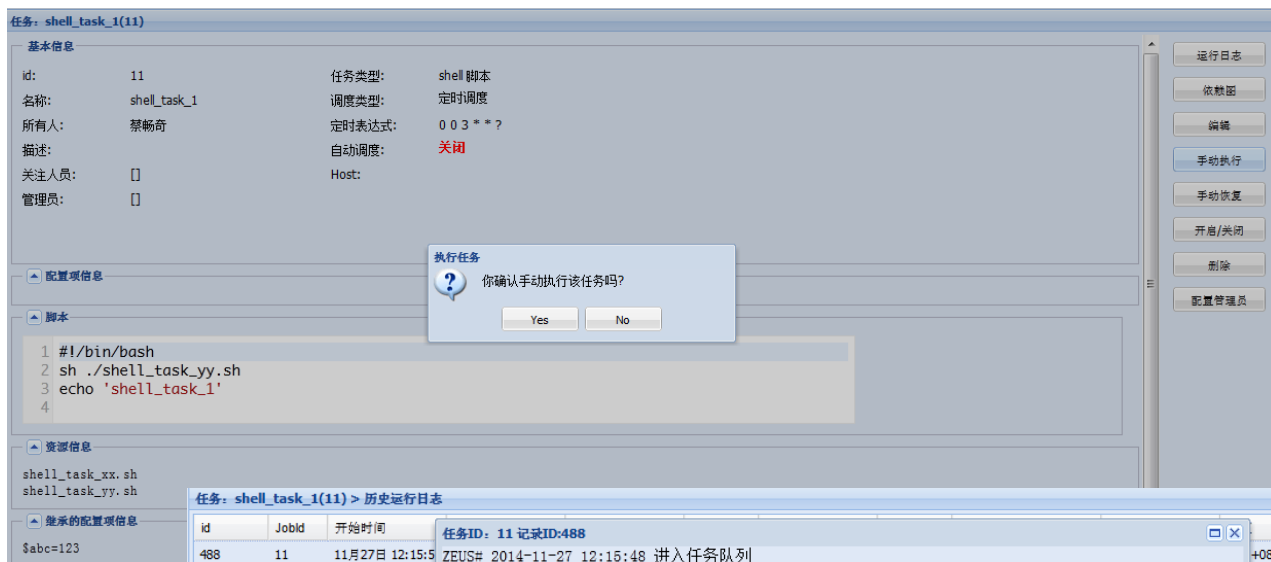
```
echo 'shell_task_1'
```

Zeus3的使用

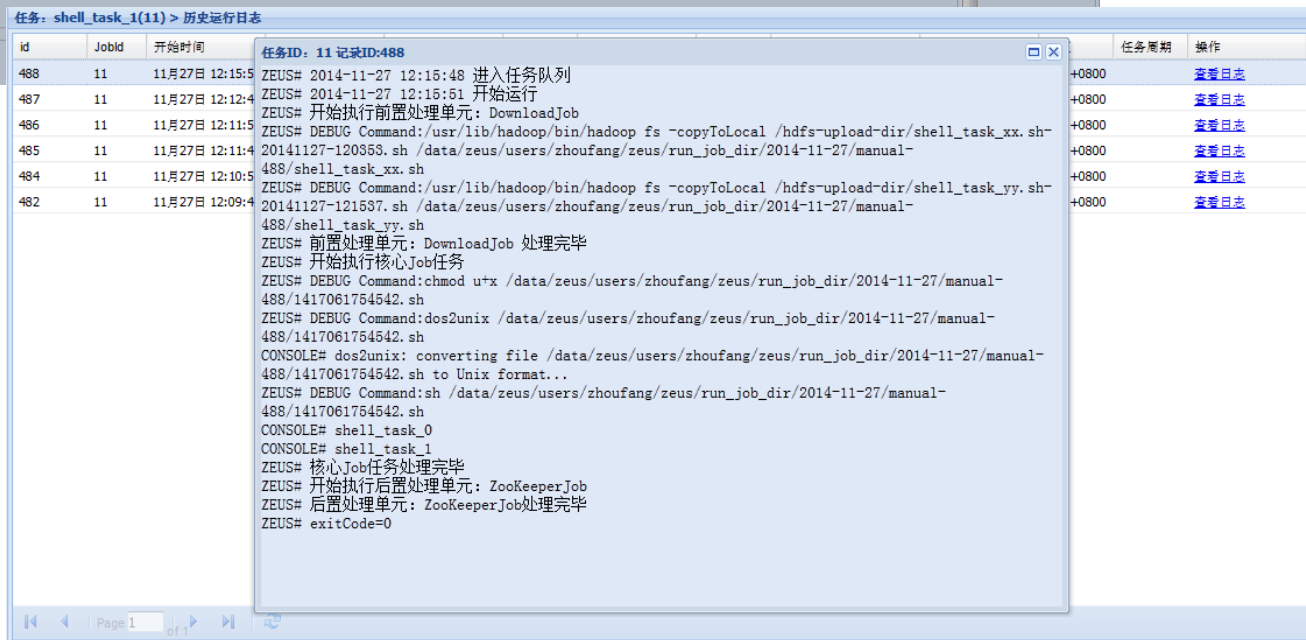
[3]调度Shell

场景1：shell_task_1.sh 内部调用shell_task_yy.sh

Step3:执行



Step4:执行结果



Zeus3的使用

[3]调度Shell

场景2：定时调度任务，每隔1分钟调度一次，不断循环

0 0/1 * * * ?

任务: shell_task_1(11) > 编辑任务

基本信息

名称: shell_task_1

描述:

定时表达式: 0 0/1 * * * ?

Host:

配置项信息

配置调度

任务: shell_task_1(11)

基本信息

id: 11
名称: shell_task_1
所有人: 蔡畅奇
描述:
关注人员:
管理员:

任务类型: shell 脚本
调度类型: 定时调度
定时表达式: 0 0/1 * * * ?
自动调度: 开启
Host:

配置项信息

脚本

成功
开启/关闭 自动调度 成功 运行日志

依赖图
编辑
手动执行
手动恢复
开启/关闭
删除
配置管理员

开启调度

调度日志

任务: shell_task_1(11) > 历史运行日志

id	JobId	开始时间	结束时间	执行服务器	执行状态	执行人	触发类型	说明	统计时间	时区	任务周期	操作
493	11	11月27日 12:46:00	11月27日 12:46:03	10.0.137.51	success	null	自动调度	null		GMT+0800	day	查看日志
492	11	11月27日 12:45:00	11月27日 12:45:03	10.0.137.51	success	null	自动调度	null		GMT+0800	day	查看日志

Zeus3的使用

[3]调度Shell

场景3：依赖调度，先调度A再调度B

0 0/1 * * * ?

The screenshot displays the Zeus3 configuration interface. On the left, two task configuration panels are visible. The top panel is for '任务: A(12)' and the bottom for '任务: B(13)'. Both tasks are of type 'shell 脚本' (shell script) and have a '定时表达式' (cron expression) of '0 0/1 * * * ?'. Task A's script is '#!/bin/bash' followed by 'echo 'A''. Task B's script is '#!/bin/bash' followed by 'echo 'B''. Yellow starburst callouts labeled '创建A' and '创建B' are placed over the task configuration panels. In the center, a dialog box titled '选择依赖任务(可以多选)' (Select dependent tasks (can select multiple)) is open. It shows a search bar with '12' and a list of tasks. Task A(12) is selected with a checkmark. A yellow starburst callout labeled 'A依赖B' is placed over the dialog. On the right, another yellow starburst callout labeled 'A依赖B' is visible over a partially obscured task configuration panel. At the bottom right, there are buttons for '删除' (Delete) and '配置管理员' (Configure Administrator).

任务: A(12)

基本信息

id: 12 任务类型: shell 脚本

名称: A 调度类型: 定时调度

所有人: 蔡畅奇 定时表达式: 0 0/1 * * * ?

描述: 自动调度: 关闭

关注人员: [] Host:

管理员: []

配置项信息

脚本

```
1 #!/bin/bash
2 echo 'A'
3
```

资源信息

任务: B(13)

基本信息

id: 13 任务类型: shell 脚本

名称: B 调度类型: 定时调度

所有人: 蔡畅奇 定时表达式: 0 0/1 * * * ?

描述: 自动调度: 关闭

关注人员: [] Host:

管理员: []

配置项信息

脚本

```
1 #!/bin/bash
2 echo 'B'
3
```

选择依赖任务(可以多选)

'12'

☐ SF大数据(1)

☐ 调度中心(2)

☐ 测试用(3)

☒ A(12)

☐ B(13)

☐ crm_unified_view_run_time.sh(7)

☐ hive-shell-test01(2)

☐ oracle_to_hive_refresh_daily-2(5)

☐ oracle_to_hive_refresh_daily_2014_11_24(3)

☐ parameter_test(4)

☐ parameter_test2(6)

☐ qswqw(1)

☐ shell_task_1(11)

☐ 调度(4)

删除

配置管理员

Zeus3的使用

[3]调度Shell

场景3：依赖调度，先调度A再调度B

具体配置细节：

- (1)A任务调度配置自动一分钟一次：0 0/1 * * * ?
- (2)B任务调度要配置依赖A任务，然后选择启动
- (3)启动A任务
- (4)查看workflow的调度结果

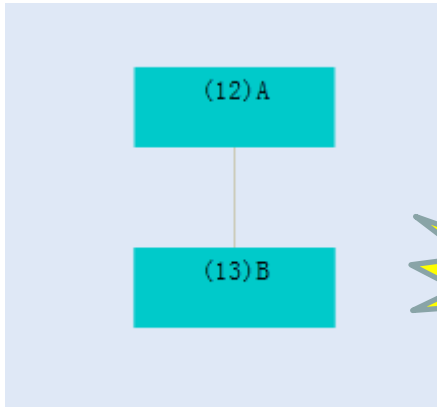
整个workflow由A触发

基本信息

id:	13	任务类型:	shell 脚本
名称:	B	调度类型:	依赖调度
所有人:	蔡畅奇	自动调度:	开启
描述:		依赖任务:	[12]
关注人员:	<input type="checkbox"/>	依赖周期:	没有限制
管理员:	<input type="checkbox"/>	Host:	

基本信息

id:	12	任务类型:	shell 脚本
名称:	A	调度类型:	定时调度
所有人:	蔡畅奇	定时表达式:	0 0/1 * * * ?
描述:		自动调度:	开启
关注人员:	<input type="checkbox"/>	Host:	
管理员:	<input type="checkbox"/>		



3) > 历史运行日志

id	JobId	开始时间	结束时间	执行服务器	执行	执行人	触发类型	说明	统计...
5...	12	11月27日 14:17:00	11月27日 14:17:00	10.0.137.51	s...	null	自动调度	依赖任务全部到位，开始执行	
5...	12	11月27日 14:16:00	11月27日 14:16:00	10.0.137.51	s...	null	自动调度	依赖任务全部到位，开始执行	
5...	12	11月27日 14:15:00	11月27日 14:15:00	10.0.137.51	s...	null	自动调度	依赖任务全部到位，开始执行	

任务: A(12) > 历史运行日志

id	JobId	开始时间	结束时间	执行服务器	执行	执行人	触发类型	说明
5...	12	11月27日 14:17:00	11月27日 14:17:00	10.0.137.51	s...	null	自动调度	null
5...	12	11月27日 14:16:00	11月27日 14:16:00	10.0.137.51	s...	null	自动调度	null
5...	12	11月27日 14:15:00	11月27日 14:15:00	10.0.137.51	s...	null	自动调度	null

Zeus3的使用

[3]调度Hive

场景1-3：直接调用Hive-SQL，与SHELL类似

任务: Hive-01(14)

基本信息

id: 14

名称: Hive-01

所有人: 蔡畅奇

描述:

关注人员: []

管理员: []

任务类型: hive 脚本

调度类型: 定时调度

定时表达式: 0 0 3 * * ?

自动调度: 关闭

Host:

配置项信息

脚本

任务: Hive-01(14)

基本信息

id: 14

名称: Hive-01

所有人: 蔡畅奇

描述:

关注人员: []

管理员: []

任务类型: hive 脚本

调度类型: 定时调度

定时表达式: 0 0 3 * * ?

自动调度: 关闭

Host:

配置项信息

脚本

1 show databases;

2

任务: Hive-01(14) > 历史运行日志

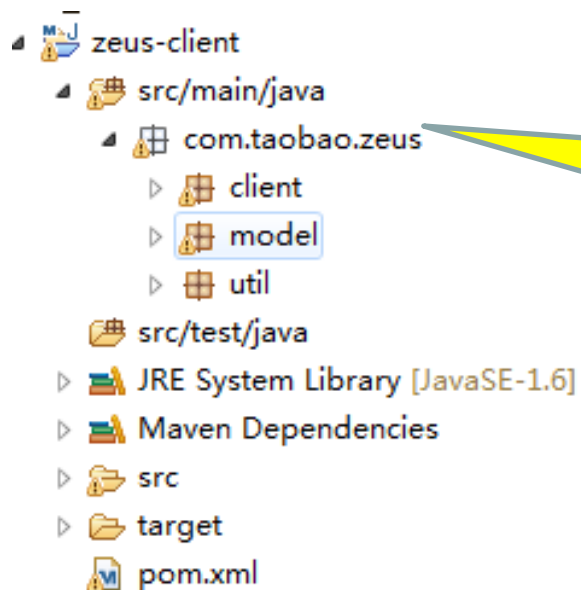
id	JobId	开始时间	结束时间	执行服务器	执行状态	执...	触发类型	说明	统...	时区	任务	操作
5...	14	11月27日 14:22:45	11月27日 14:22:53	10.0.137...	success	and...	手动调度	触发人: andy.cai		G...		查看...
5...	任务ID: 14 记录ID:572											
5...	CONSOLE# Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties											
5...	CONSOLE# Hive history file=/tmp/root/hive_job_log_b985ea90-263b-4b5d-9edb-8173497a48ba_1547262289.txt											
5...	CONSOLE# OK											
5...	CONSOLE# cbsdb1											
5...	CONSOLE# cbsdb4											
5...	CONSOLE# ccmdw											
5...	CONSOLE# crm_complain_dashboard											
5...	CONSOLE# dawn											

调度结果

返回

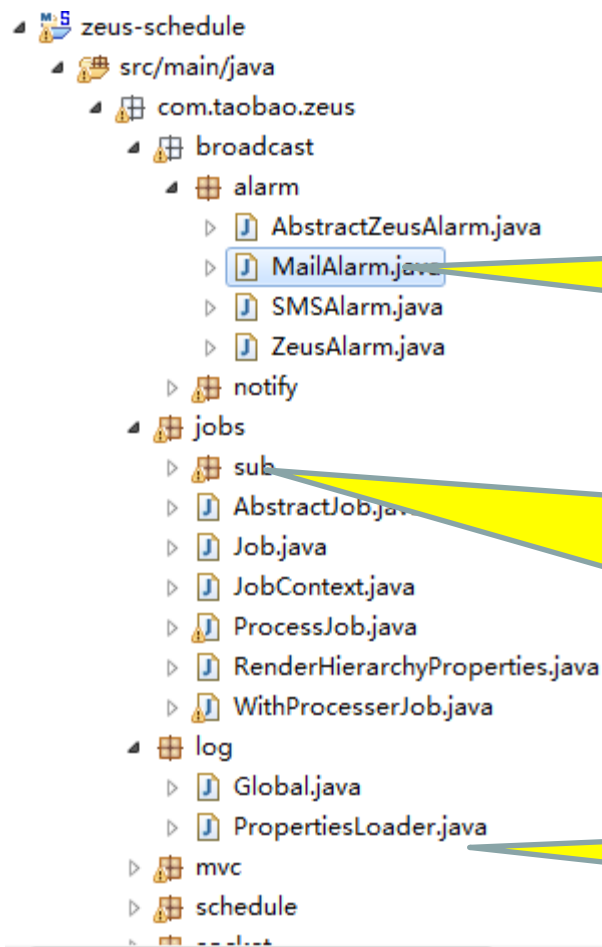
刷新

Zeus3二次开发介绍



Zeus-client工程
基本没有做改造

Zeus3二次开发介绍

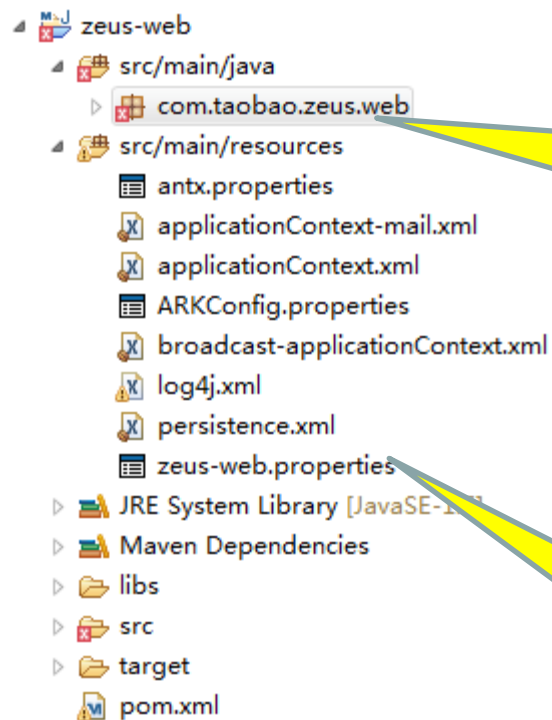


Zeus-schedule工程
1) 增加邮件报警功能

Zeus-schedule工程
2) 修改shell与hive调度功能，以前有问题

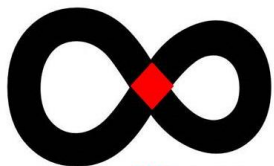
Zeus-schedule工程
3) 增加动态配置，将所有常量配置改为动态配置

Zeus3二次开发介绍



修改web部分配置

增加各种动态配置



We Achieve Limitless

与你共创无限

Q&A

Andrew Cai

Quora:<http://www.quora.com/Changqi-Cai>

QQ: 287888203

EMail: 287888203@qq.com

QQ Group(大数据QQ群): 221692953

Github:<https://github.com/caichangqi>

微信: caichangqiccq