HiBI壳工具部署指南

(仅供内部使用）

For internal use only

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 拟制:  Prepared by | 罗向龙 201279 |  | 日期：  Date | 2016-5-12 |
| 审核:  Reviewed by |  |  | 日期：  Date | yyyy-mm-dd |
| 批准:  Granted by |  |  | 日期：  Date | yyyy-mm-dd |



华为技术有限公司

Huawei Technologies Co., Ltd.

版权所有 侵权必究

All rights reserved

修订记录Revision record·

|  |  |  |  |
| --- | --- | --- | --- |
| 日期  Date | 修订版本Revision version | 修改描述  change Description | 作者  Author |
| 2016-05-12 | V1.0 | 初稿 | 罗向龙 201279 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

[1 概述 4](#_Toc340741479)

[2 部署 4](#_Toc340741480)

[2.1 前置条件 4](#_Toc340741481)

[2.2 部署步骤 4](#_Toc340741482)

[3 配置 4](#_Toc340741483)

[3.1 环境配置 4](#_Toc340741484)

[3.2 业务配置 7](#_Toc340741485)

[3.2.1 数据库类推送任务 7](#_Toc340741486)

[3.2.2 文件类推送任务 9](#_Toc340741487)

[3.2.3 目标文件格式配置 14](#_Toc340741488)

[3.2.4 传输选项配置 15](#_Toc340741489)

[3.3 日志配置 16](#_Toc340741490)

[4 运行 16](#_Toc340741491)

[4.1 自动调度 16](#_Toc340741492)

[4.2 手动执行 17](#_Toc340741493)

[4.3 日常维护 18](#_Toc340741494)

[5 附录 18](#_Toc340741495)

[5.1 SSH密钥认证 20](#_Toc340741496)

[5.2 FAQ 21](#_Toc340741497)

# 概述

壳工具主要用于替换执行网关中的后台写好的hive sql脚本中的变量。主要特点：

* 获取脚本的输入输出
* 替换脚本中的变量
* 替换脚本中的表达式
* 运行hive sql

# 系统架构

# 部署

## 前置条件

1. Java

壳工具采用java语言写就，故要确保java运行环境正常。安装合适的JDK，版本要求在1.6及以上。

1. Hive客户端

壳工具主要是用来执行hive sql的，所以需要安装hive客户端，也可以安装FI客户端。

## 部署步骤

1. 创建用户

可以为壳工具创建1个单独的操作系统用户，比如hibi；也可以与其他业务合用1个用户。

1. 软件包上传

将软件包HIBI-ExecuteShell.zip上传至*$HOME/hib*i目录下，并解压。

# 配置

## 环境配置

环境配置文件sysconfig.properties用于配置Hive安装路径，Hadoop安装路径等运行环境信息。

sysconfig.properties配置项：

|  |  |  |
| --- | --- | --- |
| 配置项名称 | 含义 | 可能取值 |
| HIVE\_PATH | hive安装路径 |  |
| HADOOP\_PATH | hadoop安装路径 |  |
| HIBI\_PATH | HIBI路径，主要用于脚本内变量替换 |  |
| hadoopuser | hdfs路径，主要用于脚本内变量替换 |  |

典型配置示例如下（按实际情况填写）：

HIVE\_PATH=/MFS/Share/BICOmmon/task/HiBI/FIClient/BIClient/Hive/bin/

HADOOP\_PATH=/MFS/Share/BICOmmon/task/HiBI/FIClient/HDFS/hadoop/bin/

HIBI\_PATH= /$USER/task/HiBI/HIBI

hadoopuser=/AppData/HomeCloudProdGrp/HomeCloud

## 业务配置

业务配置文件outinlist.xml用于配置每个脚本文件的输入输出。

outinlist.xml配置项说明：

|  |  |  |
| --- | --- | --- |
| 配置项名称 | 含义 | 可能取值 |
| name | hive脚本名称 |  |
| in | 脚本的输入 |  |
| out | 脚本的输出 |  |

典型配置实例如下：

<list>

<name>dw\_hicloud\_user\_behavior\_stat\_dm.sql</name>

<in>dw\_hicloud\_user\_behavior\_hm</in>

<out>dw\_hicloud\_user\_behavior\_stat\_dm</out>

</list>

## 日志配置

推送工具采用log4j作为日志组件，可以设置日志级别、输出目的，对日志输出进行控制。

通常，调试阶段，日志级别可设置为DEBUG级别，添加stdout输出。

**在线运行时需将日志级别设置为INFO，去掉stdout输出！**

#日志级别

log4j.rootLogger=debug, stdout, file

# 运行

## 自动调度

推送工具需要传入2个调度参数：

* 脚本名称 –表示要运行的hive脚本
* 起始周期 – 格式：yyyymmdd-HHmm，支持按天、按小时执行脚本。

壳工具不提供调度功能，业务可采用cron或TCC调度壳工具以完成数据挖掘任务。

## 手动执行

举例：exec-hive.sh test.sql 20160512-0000

# 安全特性

# 安全配置/加固、维护

## 操作系统安全性

需要基础运维保证壳工具部署服务器的操作系统均满足安全红线要求，特别注意：sshd的安全性。

分配的账号需要满足最低复杂度要求，保证定期更新账号密码。

部署壳工具的主机必须要有操作记录。方便后期对壳工具的操作进行安全审计。

## JDK安全性

推送工具需要用到jdk，需保证jdk不存在重大安全问题，有安全补丁需要及时更新。请业务运维提供最新安全版本。

## 文件访问权限

为了最大限度保证安全性，壳工具要求：所有目录、文件必需本用户可操作，其他用户无任何访问权限。