**风控系统批处理开发环境搭建指引**

版本信息

|  |  |  |  |
| --- | --- | --- | --- |
| **版本号** | **修订日期** | **修订内容** | **编辑人** |
| V1.0 | 2019.11.21 | 新建 | 李德和 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**前言**

本文档旨在指导用户完成批处理任务的开发和部署。本文假设大数据环境（Hadoop、Spark、Hive、MySQL）已经完成搭建，并且有zip解压工具。

本文在上述假设的条件下，对批量开发环境的搭建作出一些指导。

**步骤**

1. 从大数据平台的一个节点，拷贝 hadoop安装目录、hive安装目录、spark安装目 录到要部署批处理开发环境的机器（以下简称“目标机器”），并配置好三者的环 境变量（仅配置当前用户即可）
2. Java环境配置
3. 文件上传：以下文件均上传 到目标机器上用户目录下

3.1 pdi-ce-8.3.0.0-371.zip

3.2 run.sh

3.3 install.sh

1. 在目标机器上执行 sh install.sh
2. 在Mysql库里新建kettle数据库，建表batch：

CREATE TABLE `batch` (

`job` varchar(255) NOT NULL,

`datadate` date DEFAULT NULL

) ENGINE=InnoDB DEFAULT CHARSET=utf8

1. 将目标机器的以下信息填写到管理台的配置文件中：

编辑application-prod.xml（如果是开发环境，编辑application-dev.xml）

编辑内容：  
 **ssh\_host**: 目标机器ip  
 **ssh\_port**: 22  
 **ssh\_user**: 目标机器授权用户名  
 **ssh\_password**: 目标机器授权用户密码

至此，批处理开发环境搭建完成，具体开发流程请参考《批处理开发说明.docx》