**部署手册**

**修订记录**

|  |  |  |  |
| --- | --- | --- | --- |
| **版本** | **修改人** | **修改时间** | **修改说明** |
| 1.0 | 龙昌茂 | 2015-10-7 | 初始版本 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**目录**

[1 数据导入程序部署 4](#_Toc391240057)

[1.1 解压 4](#_Toc391240058)

[1.2 修改配置 4](#_Toc391240059)

[1.2.1 数据库配置 4](#_Toc391240060)

[1.2.2 日志文件配置 4](#_Toc391240061)

[2 数据爬取程序部署 5](#_Toc391240062)

[2.1 解压 5](#_Toc391240063)

[2.2 修改配置 5](#_Toc391240064)

[2.2.1 数据文件配置 5](#_Toc391240065)

[2.2.2 日志文件配置 6](#_Toc391240066)

[3 程序运行 6](#_Toc391240067)

[3.1 运行数据导入程序 6](#_Toc391240068)

[3.2 运行数据爬取程序 6](#_Toc391240069)

# 设置环境变量

需要设置以下两个环境变量：

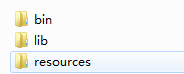
1. **file2db\_home** 其值设为数据导入程序所在目录，如D:\File2DB-1.0-SNAPSHOT
2. **crawl\_data\_path**其值设为存放爬取数据文件的目录，如D:\datas

# 数据导入程序部署

## 解压

将程序压缩包“File2DB-xxx-SNAPSHOT.zip”（xxx是版本号）解压到机器环境变量**file2db\_home**指定的目录下。

解压后的目录结构如下所示：



其中：

|-bin 存放可执行脚本文件

|-building2db.bat 楼盘户栋信息文件数据入库脚本

|-buildings2db.bat 楼盘文件数据入库脚本

|-land2db.bat 土地文件数据入库脚本

|-setEnv.cmd

|-batch2db.cmd 批量入库，供batch2db-xxx-xxx.bat类文件调用，用户手工维护

|-batch2db-4301-building2db.bat 长沙户栋详情批量入库，用户手工维护

|-batch2db-4301-buildings2db.bat长沙楼盘信息批量入库，用户手工维护

|-batch2db-4302-buildings2db.bat株洲楼盘信息批量入库，用户手工维护

|-batch2db-4303-buildings2db.bat湘潭楼盘信息批量入库，用户手工维护

|-batch2db-43-land2db-hn.bat 湖南土地信息批量入库，用户手工维护

|-lib 存放依赖的jar包文件

|-resources 存放资源文件，包括配置文件

|-applicationContext.xml 业务功能配置（不能随意修改）

|-config.properties 数据库等配置

|-log4j.properties 日志文件配置

## 修改配置

### 数据库配置

位于config.properties文件中，详细如下：

jdbc.driver jdbc驱动类名，一般不需要修改

jdbc.username 数据库用户名。

jdbc.password 数据库用户密码。

jdbc.url 数据库访问的url。

### 日志文件配置

位于log4j.properties文件中，重点只关注日志文件保存目录：

log4j.appender.file.File 普通级别日志存放路径

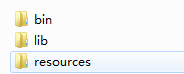
log4j.appender.err.File 告警级别日志存放路径

# 数据爬取程序部署

## 解压

将程序压缩包“TargetCrawler-xxx-SNAPSHOT.zip”（xxx是版本号）解压到机器的某个目录下，将该目录定为${TARGET\_CRAWLER\_HOME}。

解压后的目录结构如下所示：



其中：

|-bin 存放可执行脚本文件

|-run.bat 程序启动脚本

|-setEnv.cmd

|-lib 存放依赖的jar包文件

|-resources 存放资源文件，包括配置文件

|-applicationContext.xml 业务功能配置（不能随意修改）

|-config.properties 数据库等配置

|-log4j.properties 日志文件配置

|-0731fdc-floor-cs.xml 长沙楼盘信息爬取配置（不能随意修改）

|-0731fdc-floor-cs-detail.xml 长沙楼盘户栋信息爬取配置（不能随意修改）

|-0731fdc-floor-xt.xml 湘潭楼盘信息爬取配置（不能随意修改）

|-0731fdc-floor-zz.xml 株洲楼盘信息爬取配置（不能随意修改）

|-landchina-gonggaojieguo-hunan-common.xml 湖南土地信息爬取公共配置（不能随意修改）

|-landchina-gonggaojieguo-hunan-xycr.xml 湖南协议出让土地信息爬取配置（不能随意修改）

|-landchina-gonggaojieguo-hunan-zpgcr.xml 湖南招拍挂出让土地信息爬取配置（不能随意修改）

## 修改配置

### 数据文件配置

位于config.properties文件中，详细如下：

file.result.dir 存放爬取数据文件的目录（取环境变量crawl\_data\_path的值）

cmd\_completed\_handler\_prefix 任务执行完之后的shell命令前缀，包含数据导入程序命令脚本所在目录。

### 日志文件配置

位于log4j.properties文件中，重点只关注日志文件保存目录：

log4j.appender.file.File 普通级别日志存放路径

log4j.appender.err.File 告警级别日志存放路径

# 数据初始化

## 数据库初始化

将database.zip解压，然后在mysql客户端运行DDL目录下的sql文件内容以完成建表和表字段修改（tb\_land\_info）。

## 数据初始化

第一步，解压datas.zip到“存放爬取数据文件的目录”（参见“[数据文件配置](#_数据文件配置)”中的file.result.dir配置）；

第二步，运行“数据导入程序”bin目录下的batch2db-xxx-xxx.bat类文件以完成数据的初始化工作（可以直接运行batch2db-all.bat），运行情况参考对应的日志文件building2db.log、buildings2db.log、land2db.log

# 程序运行

## 运行数据导入程序

一般不需要手工运行，由数据爬取程序调用。

## 运行数据爬取程序

在数据爬取程序解压目录${TARGET\_CRAWLER\_HOME}的bin子目录下运行：run

也可以使用run脚本的完整路径进行运行；

建议配置操作系统的定时调度，每天运行一次！具体时间间隔可以根据需要调整。