北京三期数据迁移使用说明

目录

[北京三期数据迁移使用说明 1](#_Toc499305486)

[1 压缩包内容说明： 2](#_Toc499305487)

[2 修改配置文件 3](#_Toc499305488)

[2.1 修改配置文件说明 3](#_Toc499305489)

[2.2 Windows下修改BeiJingThirdPeriod.jar操作步骤 4](#_Toc499305490)

[2.3 修改配置文件 6](#_Toc499305491)

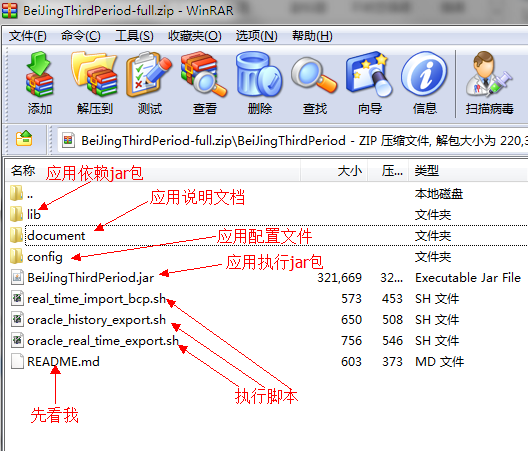
[3 部署说明 9](#_Toc499305492)

[3.1 部署应用 9](#_Toc499305493)

[3.2 导入数据 12](#_Toc499305494)

## 1 压缩包内容说明：

项目压缩包名：BeiJingThirdPeriod-full.zip



可执行脚本说明：

oracle\_real\_time\_export.sh Oracle数据实时导入脚本

real\_time\_import\_bcp.sh Bcp文件实时导入脚本

oracle\_history\_export.sh Oracle历史数据导入脚本

注意:

* 单机版用oracle\_real\_time\_export.sh脚本
* 集群版用脚本:
  + oracle\_history\_export.sh
  + real\_time\_import\_bcp.sh

## 修改配置文件

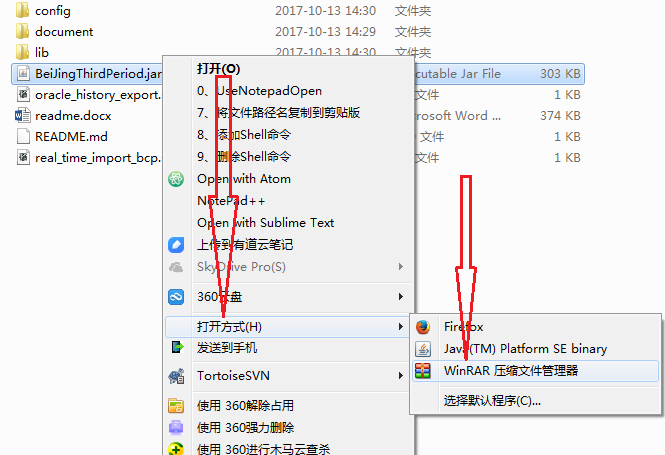
### 2.1 修改配置文件说明

主要修改BeiJingThirdPeriod.jar包三方面的配置:

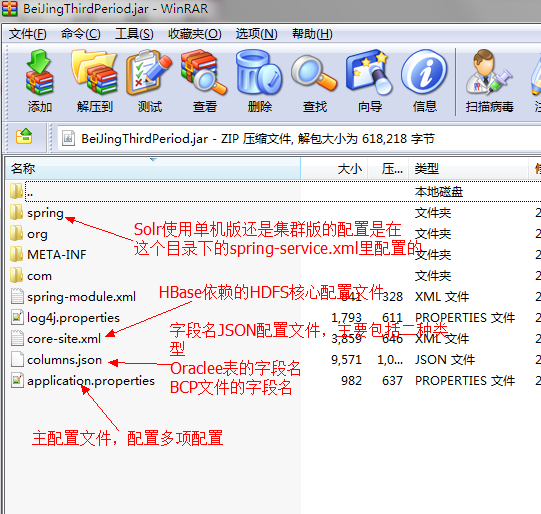
1. Solr的配置
   1. Solr是单机版的还是集群版的(在spring/spring-service.xml)
   2. Solr的连接地址(在application.properties配置文件中)
2. Oracle的配置
   1. Oracle的连接地址、用户名、密码(在application.properties配置文件中)
3. HBase的配置
   1. 是否需要导入到HBase(在application.properties配置文件中)
   2. HBase需要依赖的HDFS的核心配置文件core-site.xml
4. 字段顺序的配置（用默认的配置就行）
   1. Oracle或者Bcp文件字段一般为默认顺序，如果Oracle默认字段顺序及个数变了，或者Bcp文件添加了字段的话修改此项  
      (在columns.json配置文件)

### 2.2 Windows下修改BeiJingThirdPeriod.jar操作步骤

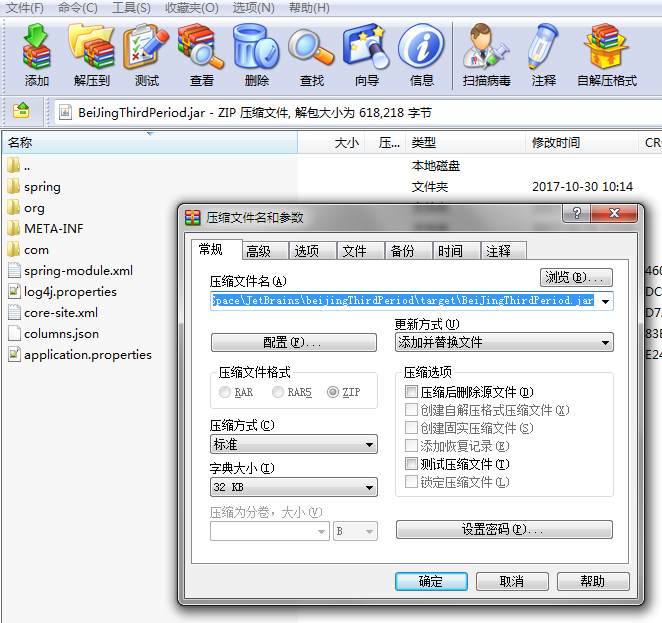
1. 用解压缩软件打开Jar包



1. 把需要修改的配置文件拖出来用文本编辑器打开进行修改

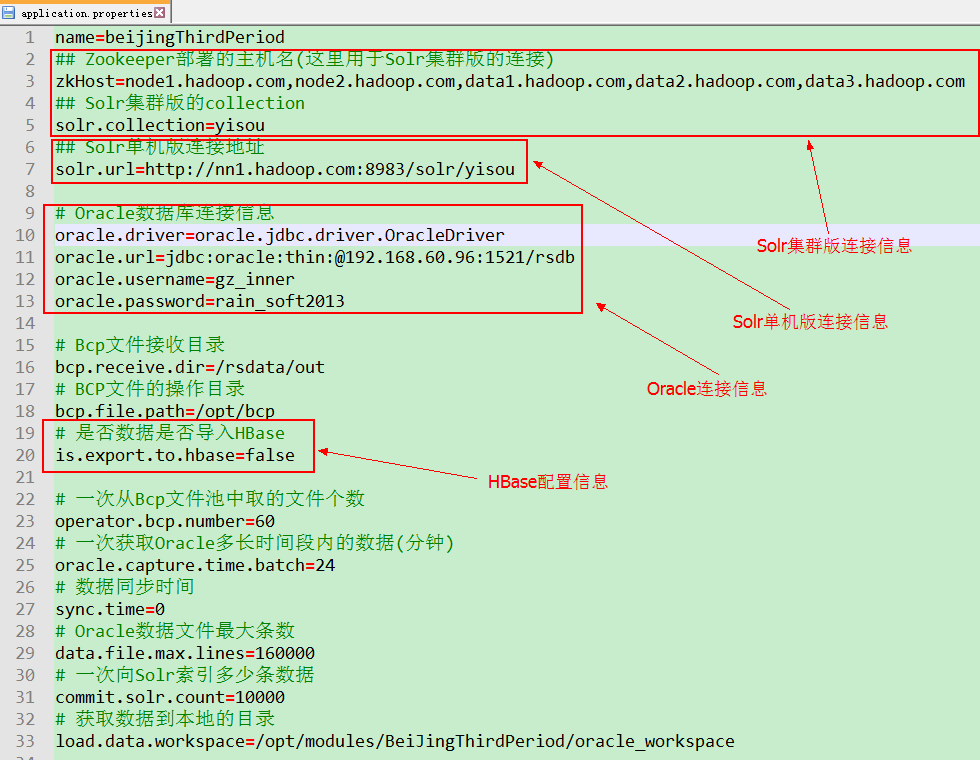


1. 保存修改的配置文件（具体哪个配置文件怎么修改下一章讲）
2. 把修改后的配置文件拖回BeiJingThirdPeriod.jar压缩包中覆盖原来的配置文件
3. 拖回压缩包后会提示是否覆盖，点击确定，提示框如下:



### 修改配置文件

application.properties配置文件截图



spring-service.xml配置文件截图



#### 2.3.1修改application.properties配置文件

**修改Solr的配置**

集群版需要修改zkHost和solr.collection

solr.collection属性，不用动，用默认的就行

zkHost就是所有安装Zookeeper服务的服务器的主机名

**修改Oracle的配置**

oracle.url:Oracle连接地址

oracle.username:Oracle用户名

oracle.password:Oracle密码

**修改HBase的配置**

集群版需要导入到HBase，这里配置为true

#### 修改****spring/spring-service.xml配置文件****

单机版与集群版的配置只能留一个

至此jar包配置文件修改完成

#### 替换core-site.xml配置文件

集群上此文件夹的位置一般为:

/opt/cloudera/parcels/CDH/lib/hadoop/etc/hadoop/core-site.xml

## 部署说明

### 3.1 部署应用

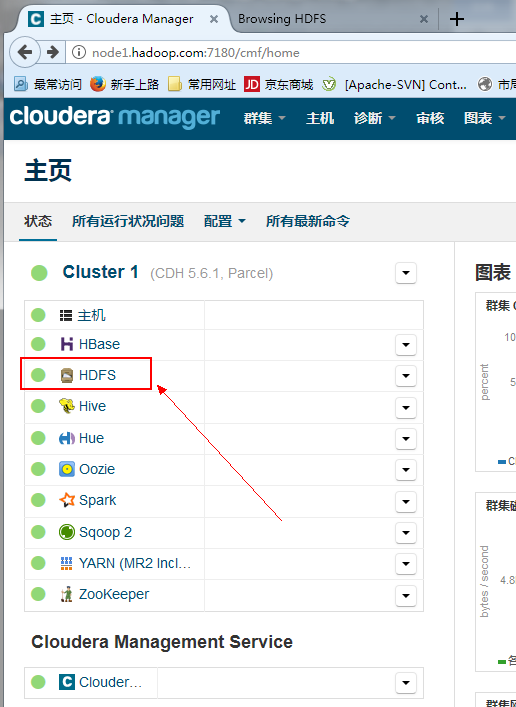
1. 将BeiJingThirdPeriod-full.zip上传到服务器(必须)
2. 创建目录(必须)

**mkdir** -p **/**opt**/**modules

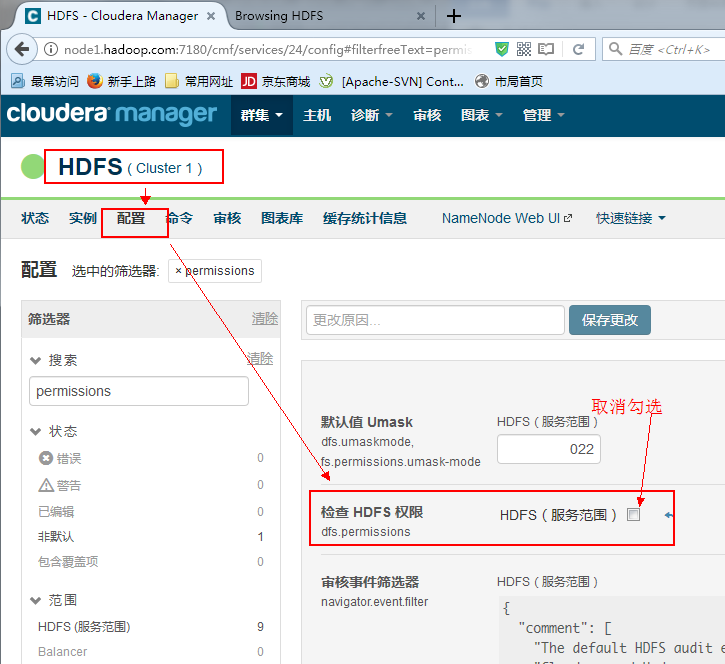
1. 解压压缩包(必须)

**unzip** BeiJingThirdPeriod-full.zip -d **/**opt**/**modules**/**

1. 将你在Windows平台下修改过配置文件的jar包替换刚刚解压的jar包(必须)
2. ClouderaManager集群关闭HDFS服务的权限检查(集群版)
3. 登陆到ClouderaManager的管理界面



1. 关闭HDFS的权限检查



1. 重启集群让配置生效



### 3.2 导入数据

根据情况看是从Bcp文件导入数据还是从Oracle导入数据

#### 3.2.1 从Oracle实时导入数据

执行Oracle实时任务脚本:

**nohup** sh oracle\_real\_time\_export.sh start **&**

结束导入命令

sh oracle\_real\_time\_export.sh stop

#### 3.2.2 从Bcp文件实时导入数据

**首先从Oracle导入历史数据**

执行导入的命令:

**nohup** sh oracle\_history\_export.sh start "${结束时间}" **&**

结束时间,结束时间是根据Bcp文件最早生成的时间来确定的，比如说最早的Bcp文件是从2017-11-01 00:00:00开始生成的，截止时间就应该是:” 2017-11-01 00:00:00”。注意一定要加上英文的双引号,不然会被认为是两个参数。

下面是一个完整的例子：

**nohup** sh oracle\_history\_export.sh start " 2017-12-01 00:00:00" **&**

结束导入的命令:

**sh** oracle\_history\_export.sh stop

**从Bcp导入实时数据**

**nohup** sh real\_time\_import\_bcp.sh start **&**

#### 3.2.3 查看导入日志

查看程序是否正常执行:可以查看执行日志:

tail -f logs**/**log.log