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](#_Toc497240297)

[1 压缩包内容说明： 2](#_Toc497240298)

[2 修改配置文件 3](#_Toc497240299)

[2.1 修改配置文件说明 3](#_Toc497240300)

[2.2 操作步骤 3](#_Toc497240301)

[2.3 修改配置文件 5](#_Toc497240302)

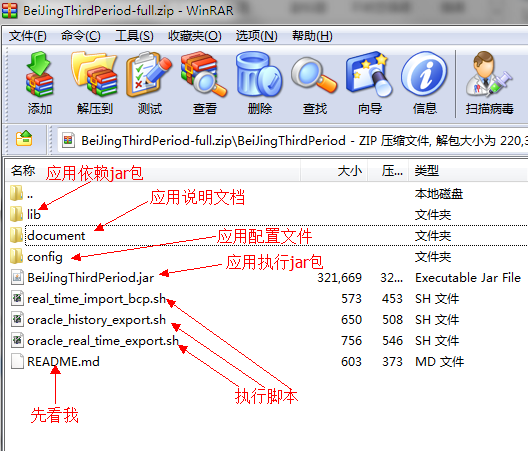
[3 部署说明 8](#_Toc497240303)

[3.1 部署应用 8](#_Toc497240304)

[3.2 导入数据 11](#_Toc497240305)

## 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: oracle\_history\_export.sh脚本一般不用，它是在这种情况使用的:实时数据是从Bcp文件导入的，历史数据从Bcp文件取不了，于是从Oracle取就用这个脚本

## 修改配置文件

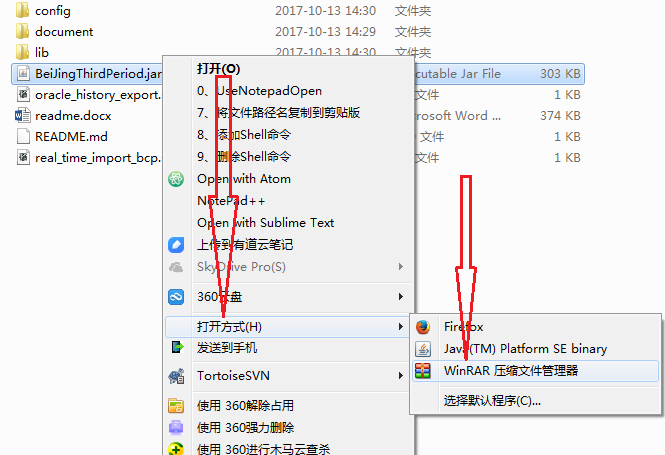
### 2.1 修改配置文件说明

主要修改BeiJingThirdPeriod.jar包三方面的配置:

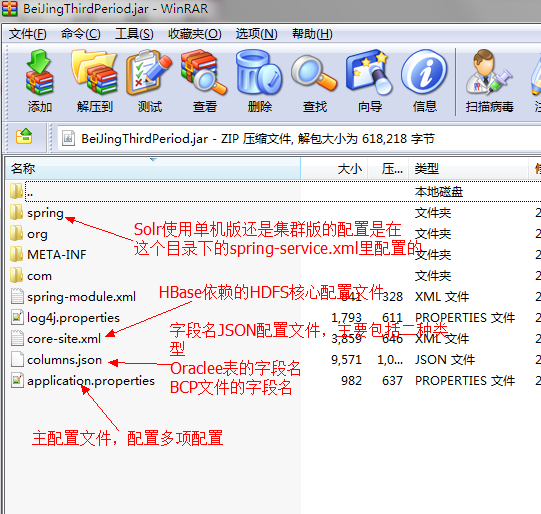
1. Solr的配置
   1. Solr是单机版的还是集群版的(在spring/spring-service.xml)
   2. Solr的连接地址(在application.properties配置文件中)
2. Oracle的配置
   1. Oracle的连接地址、用户名、密码(在application.properties配置文件中)
3. HBase的配置
   1. 是否需要到HBase(在application.properties配置文件中)
   2. HBase需要依赖的HDFS的核心配置文件core-site.xml
4. 字段顺序的配置
   1. Oracle或者Bcp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 操作步骤

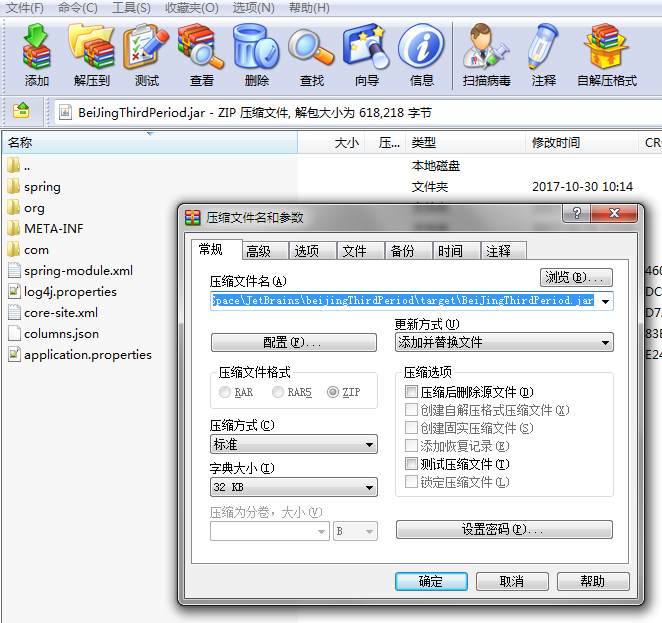
1. 用解压缩软件打开Jar包



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

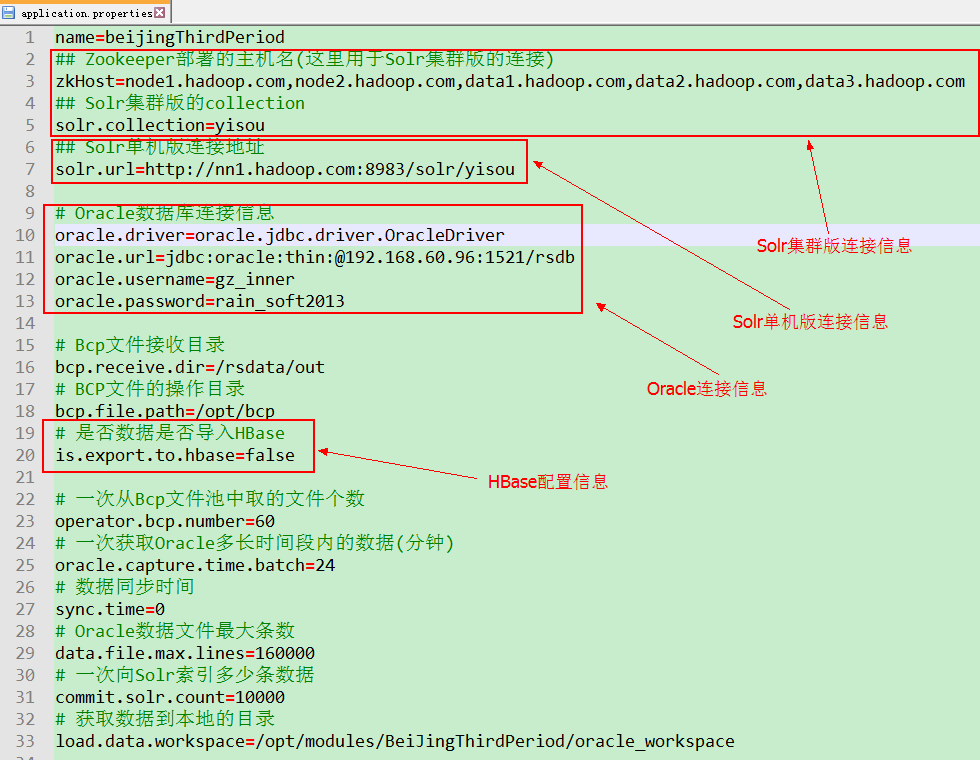


1. 保存修改的配置文件
2. 把修改后的配置文件拖回BeiJingThirdPeriod.jar压缩包
3. 拖回压缩包后会提示是否覆盖，点击确定，提示框如下:



### 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Oracle配置

配置文件名: application.properties

oracle.url:Oracle连接地址

oracle.username:Oracle用户名

oracle.password:Oracle密码

#### 修改Solr配置

1. **配置Solr单机版还是集群版(spring/spring-service.xml)**

截图见上

配置文件:

spring-service.xml

如果是单机版的就用把集群版的配置注释掉，反之亦然

1. **Solr连接地址配置**

截图见上

配置文件:

application.properties

Solr单机版配置信息:

solr.url: Solr单机版连接地址

Solr集群版配置信息:

zkHost:大数据集群Zookeeper服务所在的节点名

solr.collection:Solr当前使用的collection(一般不变)

如果是集群版配置集群版的连接信息，单机版的不用管，反之亦然。

#### 修改HBase配置

1. **配置是否需要导入到HBase**

配置文件:

application.properties

修改配置项:

is.export.to.hbase=false

如果需要导入到HBase就改成true，如果不需要就false,北京是需要导入到HBase的。

1. **替换core-site.xml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

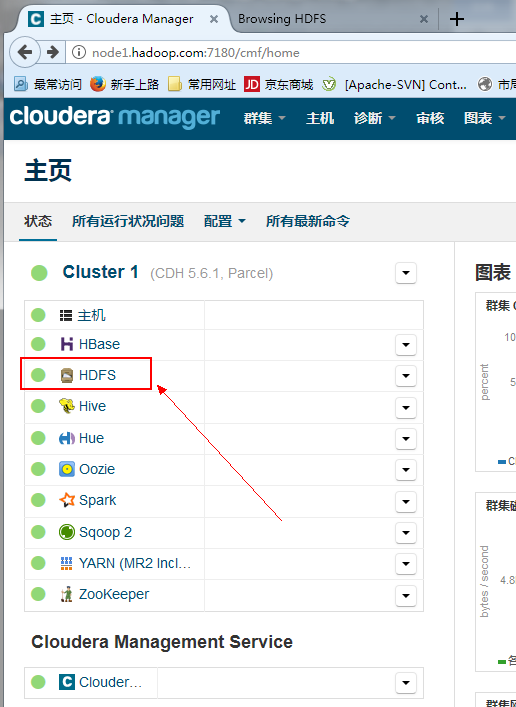
1. 将BeiJingThirdPeriod-full.zip上传到服务器(必须)
2. 创建目录(必须)

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

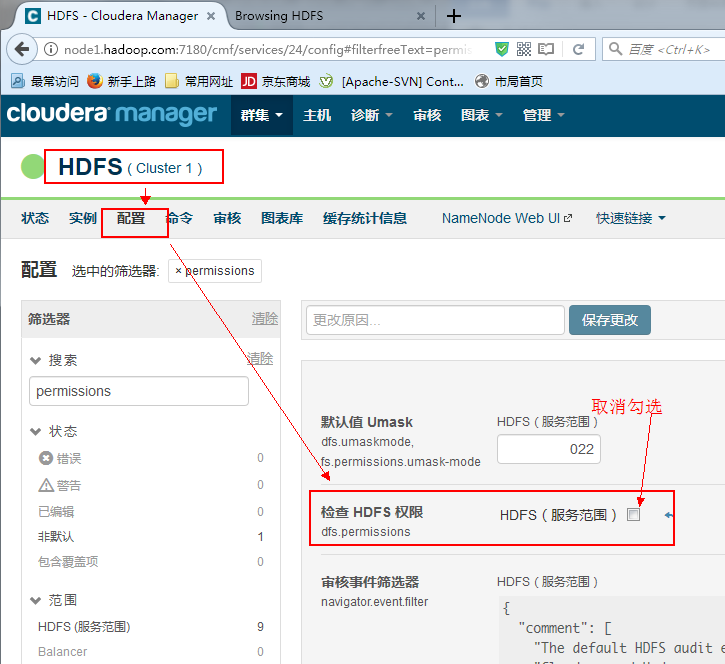
1. 解压压缩包(必须)

**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 **/**opt**/**modules**/**BeiJingThirdPeriod**/**oracle\_real\_time\_export.sh **&**

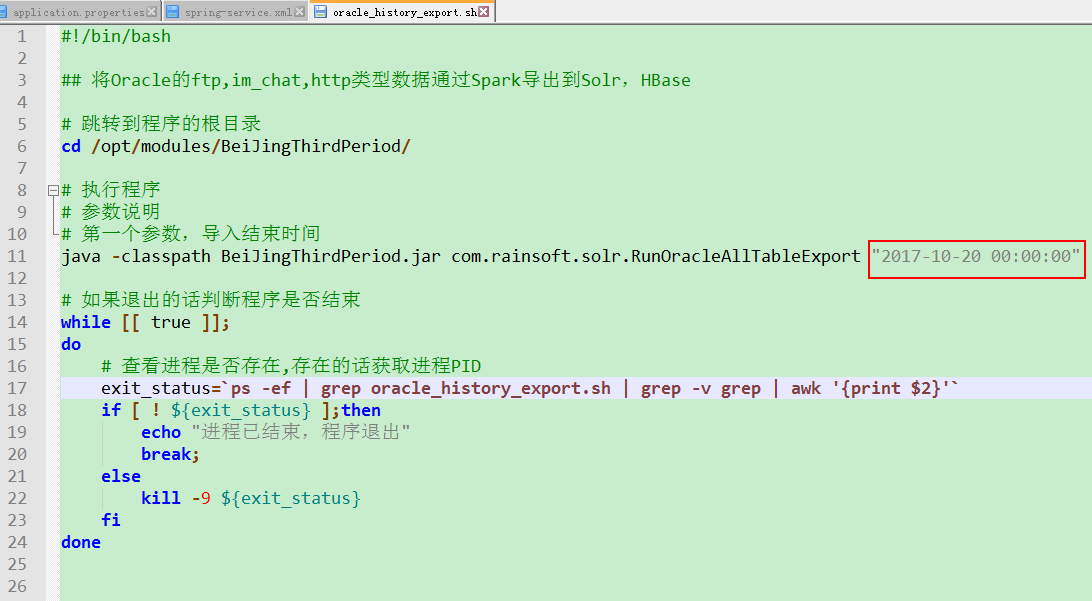
#### 3.2.2 从Bcp文件实时导入数据

首先需要从Oracle导入历史数据

修改历史任务脚本:

**vim** oracle\_history\_export.sh

修改oracle\_history\_export.sh历史任务的结束时间



参数为导入截止时间,截止时间是根据Bcp文件最早的时候来确定的，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。

执行历史任务脚本

**nohup** sh **/**opt**/**modules**/**BeiJingThirdPeriod**/**oracle\_history\_export.sh **&**

导入数据跑Bcp实时任务(根据情况确定执行)

**nohup** sh **/**opt**/**modules**/**BeiJingThirdPeriod**/**real\_time\_import\_bcp.sh **&**

#### 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:

**vim** **/**opt**/**modules**/**BeiJingThirdPeriod**/**logs**/**log.log