硬件

Cpu：Inter® Xeon® E5-2670 v2 @ 2.5GHz 40个

内存：256G

硬盘: 10T

Oracle 数据库

1. Oracle版本：12.1.0
2. PIMDATA表空间：50个数据文件，空间30G\*50
3. Temp表空间：10个数据文件，自动扩展，上限至300G
4. 最大连接数 processes ：98304
5. 以noarchivelog方式启动数据库实例，写入的表无主键/nologging，读取的表 设置parallel，结果表加主键和parallel
6. 每cpu并行度 parallel\_threads\_per\_cpu ：2

操作系统

1. 版本: redhat 6.5 2.6.32-431.el6.x86\_64
2. kernel.sem = **5010 641280 5010 128**

java 虚拟机：

-Xms10240M –Xmx153600M

运行方法：

1. 根据机器内存调整java启动参数；
2. 确定源表存在，修改配置文件中的表名（有三处）；

配置数据库连接，修改druid.properties中url、username、password

1. 确保源表有分区（分区数不限，和配置文件config.properties 里的参数项check\_certno\_partition\_count设置保持一致）；
2. 根据需要调整配置文件，针对特定类别的证件则调整check\_certno\_cert\_type和merge\_cert\_type，全部类型则设为all；
3. 创建存储过程，同用户下执行pro\_zxzx.sql；
4. 导入最新的行政区划，执行DIC\_REGION\_DATA.sql 和 DIC\_REGION.sql（已导入可省略）；
5. 根据平台选择run\_linux.sh或run\_win.bat。

软件参数

1. config.properties

#计算数据总量并行数

count\_parallel=8

#证件号码校验，提取无效证件号使用的参数

#如名称包含valid关键字，则认为是迭代调用，会删除源表中变有效的结果

check\_certno\_table\_name=qita

#根据源表分区数设置，根据源表分区调整

check\_certno\_partition\_count=128

#可以为0~9及A,B,C,X，ALL表示全部

check\_certno\_cert\_type=ALL

#验证号码同时运行的线程数

check\_certno\_threadpool\_max\_size=64

#验证号码并行度

check\_certno\_parallel=8

#一号多名合并使用的参数

#如名称包含\_mergeInvalid关键字，则认为是迭代调用，会删除源文件中变有效的结果

merge\_table\_name=qita

#可以为0~9及A,B,C,X，ALL表示全部

merge\_cert\_type=ALL

#合并同时运行的线程数

merge\_threadpool\_max\_size=64

#合并并行度

merge\_parallel=4

#提取有效分组的参数

extract\_result\_table\_name=qita

#提取类型:0 全部 1 只做相似度提取 2 只做2名比较 3 只做3名比较 4 只做4名比较

extract\_result\_type = 0

#8进制,类型阈值，用类型位数表示

#取值可为:6只由汉字组成 5由汉字/空格组成 4只由汉字/空格/括号对组成

# 3由汉字/空格/括号对/英文字母、数字组成 2由汉字/空格/英文字符/括号对/其他字符组成 1其他

extract\_result\_type\_threshold = -1

#8进制,首位阈值

#取值可为:1音字都不等 4音同字不同 6音同字同 7原始字同或音同

extract\_result\_bit\_threshold = 2

#汉字的edit\_distance\_similarity

sim\_2\_dis\_threshold = 30

sim\_3\_1\_2\_dis\_threshold = 30

sim\_3\_1\_3\_dis\_threshold = 30

sim\_3\_2\_3\_dis\_threshold = 30

sim\_4\_1\_2\_dis\_threshold = 30

sim\_4\_1\_3\_dis\_threshold = 30

sim\_4\_1\_4\_dis\_threshold = 30

sim\_4\_2\_3\_dis\_threshold = 30

sim\_4\_2\_4\_dis\_threshold = 30

sim\_4\_3\_4\_dis\_threshold = 30

#拼音的JARO\_WINKLER\_SIMILARITY

sim\_2\_jw\_threshold = 30

sim\_3\_1\_2\_jw\_threshold = 90

sim\_3\_1\_3\_jw\_threshold = 90

sim\_3\_2\_3\_jw\_threshold = 90

sim\_4\_1\_2\_jw\_threshold = 90

sim\_4\_1\_3\_jw\_threshold = 90

sim\_4\_1\_4\_jw\_threshold = 90

sim\_4\_2\_3\_jw\_threshold = 90

sim\_4\_2\_4\_jw\_threshold = 90

sim\_4\_3\_4\_jw\_threshold = 90

#匹配时过滤的用户名

extract\_result\_except = 无名氏,无户名,佚名,暂缺,银联测试，测试

#在线验证行政区划代码使用的参数

check\_dic\_code\_table\_name=REGION\_CHECK

check\_dic\_code\_thread\_process\_count=5000

check\_dic\_code\_threadpool\_max\_size=10

1. druid.properties

driverClassName=oracle.jdbc.driver.OracleDriver

url=jdbc:oracle:thin:@127.0.0.1:1521:orcl

username=bpi

password=zxzx

filters=stat

#初始化连接数量

initialSize=1

#最大并发连接数

maxActive=51200

#配置获取连接等待超时的时间

maxWait=6000000

#配置间隔多久才进行一次检测，检测需要关闭的空闲连接，单位是毫秒

timeBetweenEvictionRunsMillis=20000

#配置一个连接在池中最小生存的时间，单位是毫秒

minEvictableIdleTimeMillis=300000

validationQuery=SELECT 1 FROM DUAL

testWhileIdle=true

#申请连接时执行validationQuery检测连接是否有效，配置为true会降低性能

testOnBorrow=false

#归还连接时执行validationQuery检测连接是否有效，配置为true会降低性能

testOnReturn=false

#打开PSCache，并且指定每个连接上PSCache的大小

poolPreparedStatements=true

maxPoolPreparedStatementPerConnectionSize=2000