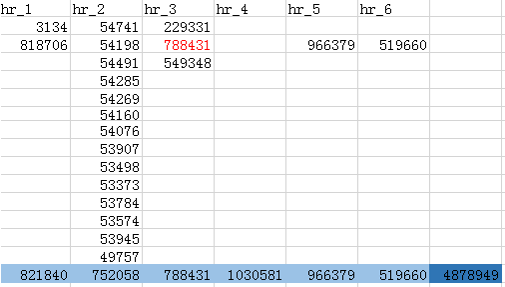
本地简历库MDF数据库数据的整理

一、打开MDF数据库文件

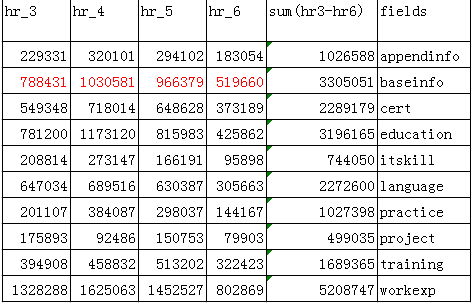
1.1 使用SqlServer2008 打开MDF数据库文件，在192.168.6.59sql数据库中新建了6个数据库，用于存放6个MDF文件多对应的.sql文件



1.2 查看数据库表结构，分析数据，select \*from dbo.table（where id=）；统计据条数为：



hr3--hr6数据库中各有10张表，相对应的表的schema相同，各个表中的数据统计如下：

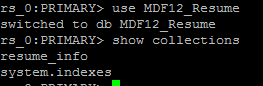


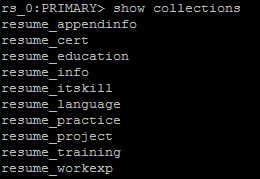
二、SqlServer数据迁移到mongodb上

对以上6个数据库内的数据进行分析，根据表结构及数据进行整合1、2数据库整合一起，包含一张总表，3-6数据库整合一起，包含10张分表

数据库数据迁移到192.168.6.62 mongo数据库，如下：







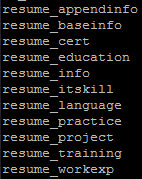
代码实现过程、调试及完成数据迁移

以及代码路径：.\MDF2Mongodb\之下

三、mongodb数据库collections之间的数据传递

**3.1 数据库MDF3\_6\_Resume子表数据整合到主表中**

1）实现功能：对MDF3\_6\_Resume数据库中的十张表，通过“id”的关联，将其他表中的数据整合到resume\_info表格中相应的字段中。



比如：resume\_info 与 resume\_cert两张表，表格之间数据关联通过‘id’属性，要实现的功能是：将resume\_cert表格中的证书信息（cert\_time+cert）整合一起存入到resume\_info表中的person\_certification字段中;一个id下的证书信息可能会有多条，将所有信息累加后存入person\_certification字段中。

代码实现过程、调试及完成数据整合

以及代码路径：.\conn\_joindata\join\_data.py

2）整理所需属性字段及对应关系及结果

Resume\_info表格与其余9个表格 resume\_info中数据为3305051

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 子表名称 | Resume\_info对应的字段 | 子表数据整合所包含的字段 | 更新条数 | 子表中不包括的id条数 |
| resume\_cert | Person\_certification | str1 = fc['cert\_time'] + '\t' + fc['cert'] | 2362763 | 942288 |
| resume\_education | c\_edu\_exp | str1=fc['edu\_start\_time']+'\t'+ fc['edu\_end\_time'] + '\t' +fc['school\_name']+'\t' +fc['edu\_major']+'\t' +fc['education']+'\t' +fc['major\_desc'] | 2794406 | 510645 |
| resume\_workexp | C\_work\_exp | str1=fc['work\_begin\_time']+'\t'+fc['work\_end\_time']+\t'+fc['work\_in\_company']+'\t'+fc['work\_department']+'\t'+fc['work\_position']+'\t' fc['work\_abstract']+'\t'fc['work\_grade']+'\t' fc['work\_xiashu\_num']+'\t'+fc['company\_industry']+'\t' +fc['report\_person']+'\t'+fc['leave\_reason']+'\t' fc['reference'] | 1878537 | 1426514 |
| resume\_project | c\_pro\_exp | str1 =fc['start\_date'] + '\t' + fc['end\_date']+ '\t' +fc['perject\_name']+ '\t' +fc['project\_desc']+ '\t' +fc['project\_grade']+ '\t' +fc['soft\_env']+ '\t' +fc['hard\_env'] | 137880 | 3167171 |
| resume\_appendinfo | person\_note | str1 = fc['title'] + '\t' + fc['remark'] | 692773 | 2623004 |
| resume\_practice | campus\_practice | str1 = fc['start\_time'] + '\t' + fc['end\_time'] + '\t' +fc['practice\_duty']+ '\t' +fc['practice\_type']+ '\t' +fc['practice\_desc'] | 518127 | 2862134 |
| resume\_language | person\_language\_skill | str1 = fc['language'] + '\t' + fc['level\_id'] | 1281052 | 2112243 |
| resume\_training | c\_train\_exp | str1 = fc['start\_time'] + '\t' + fc['end\_time']+ '\t' +fc['school']+ '\t' +fc['location']+ '\t' +fc['course']+ '\t' +fc['cert\_name']+ '\t' +fc['training\_desc'] | 1004369 | 2301839 |
| resume\_itskill | person\_profession\_skill | str1 = fc['skill\_type'] + '\t' + fc['use\_time'] + '\t' +fc['level\_id'] | 304714 | 3001057 |

**3.2 主表与子表建立resume\_id的关联**

实现功能：对MDF3\_6\_Resume数据库中的主表（resume\_info）以及三张分表（resume\_education、resume\_workexp、resume\_project），通过“id”的关联，将其他表中的resume\_id与resume\_info表格中的resume\_id一样。

 代码实现过程、调试及完成数据转换

 代码路径：.\conn\_joindata\connect\_join.py

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Resume\_education | | Resume\_workexp | Resume\_project |
| No\_count | | 510645 | | 1426514 | 3167171 |
| Update\_count | | 2794406 | | 1878536 | 137879 |
| Sum\_zibiao | | 3157225 | | 3629200 | 417936 |
| 更新条数 |  | |

No\_count:主表id在子表中没有数据

Update\_count：更新条数

Sum\_zibiao: 子表中总共更新的条数