查看Oracle字符集及如何修改字符集

文章分类:数据库

一、什么是oracle字符集

Oracle字符集是一个字节数据的解释的符号集合,有大小之分,有相互的包容关系。ORACLE 支持国家语言的体系结构允许你使用本地化语言来存储，处理，检索数据。它使数据库工具，错误消息，排序次序，日期，时间，货币，数字，和日历自动适应本地 化语言和平台。

影响oracle数据库字符集最重要的参数是NLS\_LANG参数。

它的格式如下: 　　NLS\_LANG = language\_territory.charset

它有三个组成部分(语言、地域和字符集)，每个成分控制了NLS子集的特性。

其中:

Language 指定服务器消息的语言，territory 指定服务器的日期和数字格式，charset 指定字符集。如:AMERICAN \_ AMERICA. ZHS16GBK

从NLS\_LANG的组成我们可以看出，真正影响数据库字符集的其实是第三部分。

所以两个数据库之间的字符集只要第三部分一样就可以相互导入导出数据，前面影响的只是提示信息是中文还是英文。

二.查看数据库字符集

这涉及三方面的字符集，

一是oracel server端的字符集;

二是oracle client端的字符集;

三是dmp文件的字符集。

在做数据导入的时候，需要这三个字符集都一致才能正确导入。

1、查询oracle server端的字符集

有很多种方法可以查出oracle server端的字符集，比较直观的查询方法是以下这种:

SQL>select userenv(‘language’) from dual;

结果类似如下:AMERICAN \_ AMERICA. ZHS16GBK

2、如何查询dmp文件的字符集

用oracle的exp工具导出的dmp文件也包含了字符集信息，dmp文件的第2和第3个字节记录了dmp文件的字符集。如果dmp文件不大，比 如只有几M或几十M，可以用UltraEdit打开(16进制方式)，看第2第3个字节的内容，如0354，然后用以下SQL查出它对应的字符集:

SQL> select nls\_charset\_name(to\_number('0354','xxxx')) from dual;

ZHS16GBK

如果dmp文件很大，比如有2G以上(这也是最常见的情况)，用文本编辑器打开很慢或者完全打不开，可以用以下命令(在unix主机上):

cat exp.dmp |od -x|head -1|awk '{print $2 $3}'|cut -c 3-6

然后用上述SQL也可以得到它对应的字符集。

3、查询oracle client端的字符集

这个比较简单。

在windows平台下，就是注册表里面相应OracleHome的NLS\_LANG。还可以在dos窗口里面自己设置，比如: 　　set nls\_lang=AMERICAN\_AMERICA.ZHS16GBK

这样就只影响这个窗口里面的环境变量。

在unix平台下，就是环境变量NLS\_LANG。

$echo $NLS\_LANG

AMERICAN\_AMERICA.ZHS16GBK

如果检查的结果发现server端与client端字符集不一致，请统一修改为同server端相同的字符集。

补充：

(1).数据库服务器字符集

select \* from nls\_database\_parameters

来源于props$，是表示数据库的字符集。

(2).客户端字符集环境

select \* from nls\_instance\_parameters

其来源于v$parameter，表示客户端的字符集的设置，可能是参数文件，环境变量或者是注册表

(3).会话字符集环境

select \* from nls\_session\_parameters

来源于v$nls\_parameters，表示会话自己的设置，可能是会话的环境变量或者是alter session完成，如果会话没有特殊的设置，将与nls\_instance\_parameters一致。

(4).客户端的字符集要求与服务器一致，才能正确显示数据库的非Ascii字符。如果多个设置存在的时候，alter session>环境变量>注册表>参数文件

字符集要求一致，但是语言设置却可以不同，语言设置建议用英文。如字符集是zhs16gbk，则nls\_lang可以是 American\_America.zhs16gbk。

三、修改oracle的字符集

上文说过，oracle的字符集有互相的包容关系。如us7ascii就是zhs16gbk的子集,从us7ascii到zhs16gbk不会有数 据解释上的问题,不会有数据丢失。在所有的字符集中utf8应该是最大,因为它基于unicode,双字节保存字符(也因此在存储空间上占用更多)。

一旦数据库创建后，数据库的字符集理论上讲是不能改变的。因此，在设计和安装之初考虑使用哪一种字符集十分重要。根据Oracle的官方说明，字符 集的转换是从子集到超集受支持,反之不行。如果两种字符集之间根本没有子集和超集的关系，那么字符集的转换是不受oracle支持的。对数据库 server而言，错误的修改字符集将会导致很多不可测的后果，可能会严重影响数据库的正常运行，所以在修改之前一定要确认两种字符集是否存在子集和超集 的关系。一般来说，除非万不得已，我们不建议修改oracle数据库server端的字符集。特别说明，我们最常用的两种字符集ZHS16GBK和 ZHS16CGB231280之间不存在子集和超集关系，因此理论上讲这两种字符集之间的相互转换不受支持。

1、修改server端字符集(不建议使用)

在oracle 8之前，可以用直接修改数据字典表props$来改变数据库的字符集。但oracle8之后，至少有三张系统表记录了数据库字符集的信息，只改 props$表并不完全，可能引起严重的后果。正确的修改方法如下:

$sqlplus /nolog

SQL>conn / as sysdba; 　　若此时数据库服务器已启动，则先执行SHUTDOWN IMMEDIATE命令关闭数据库服务器，然后执行以下命令:

SQL>STARTUP MOUNT;

SQL>ALTER SYSTEM ENABLE RESTRICTED SESSION;

SQL>ALTER SYSTEM SET JOB\_QUEUE\_PROCESSES=0;

SQL>ALTER SYSTEM SET AQ\_TM\_PROCESSES=0;

SQL>ALTER DATABASE OPEN;

SQL>ALTER DATABASE CHARACTER SET ZHS16GBK;

SQL>ALTER DATABASE national CHARACTER SET ZHS16GBK;

SQL>SHUTDOWN IMMEDIATE;

SQL>STARTUP

注意：如果没有大对象，在使用过程中进行语言转换没有什么影响，（切记设定的字符集必须是ORACLE支持，不然不能start） 按上面的做法就可以,但是可能会出现‘ORA-12717: Cannot ALTER DATABASE NATIONAL CHARACTER SET when NCLOB data exists’ 这样的提示信息

要解决这个问题有两种方法

一个是,利用INTERNAL\_USE 关键字修改区域设置,

还有一个是利用re-create,但是re-create有点复杂,所以请用internal\_use,

SQL>SHUTDOWN IMMEDIATE;

SQL>STARTUP MOUNT EXCLUSIVE;

SQL>ALTER SYSTEM ENABLE RESTRICTED SESSION;

SQL>ALTER SYSTEM SET JOB\_QUEUE\_PROCESSES=0;

SQL>ALTER SYSTEM SET AQ\_TM\_PROCESSES=0;

SQL>ALTER DATABASE OPEN;

SQL>ALTER DATABASE NATIONAL CHARACTER SET INTERNAL\_USE UTF8;

SQL>SHUTDOWN immediate;

SQL>startup;

如果按上面的做法做,National charset的区域设置就没有问题

2、修改dmp文件字符集

上文说过，dmp文件的第2第3字节记录了字符集信息，因此直接修改dmp文件的第2第3字节的内容就可以‘骗’过oracle的检查。这样做理论 上也仅是从子集到超集可以修改，但很多情况下在没有子集和超集关系的情况下也可以修改，我们常用的一些字符集，如 US7ASCII，WE8ISO8859P1，ZHS16CGB231280，ZHS16GBK基本都可以改。因为改的只是dmp文件，所以影响不大。

具体的修改方法比较多，最简单的就是直接用UltraEdit修改dmp文件的第2和第3个字节。

比如想将dmp文件的字符集改为ZHS16GBK，可以用以下SQL查出该种字符集对应的16进制代码: 　　SQL> select to\_char(nls\_charset\_id('ZHS16GBK'), 'xxxx') from dual;

0354

然后将dmp文件的2、3字节修改为0354即可。

如果dmp文件很大，用ue无法打开，就需要用程序的方法了

安装时采用什么字符集主要看你的需求，一般测试用的话，就用UTF8.

后面可以修改，

10g：UTF8 到ZHS16GBK

SHUTDOWN IMMEDIATE;

STARTUP MOUNT ;

ALTER SYSTEM ENABLE RESTRICTED SESSION;

ALTER SYSTEM SET JOB\_QUEUE\_PROCESSES=0;

ALTER SYSTEM SET AQ\_TM\_PROCESSES=0;

ALTER DATABASE OPEN;

----ALTER DATABASE NATIONAL CHARACTER SET AL16UTF16; 更改国际字符集

===========

----ALTER DATABASE CHARACTER SET ZHS16GBK; 更改本地字符集

ALTER DATABASE character set INTERNAL\_USE ZHS16GBK;

===========

SHUTDOWN IMMEDIATE;

STARTUP ;