当获取一个dmp文件，如何确定其使用的抽取工具是什么呢（exp还是expdp）？dmp文件的数据库字符集是什么呢？dmp中源库表空间等架构是什么样的呢？

》得到了一个导出dmp文件。(不确定其 是exp 还是 expdb)

通过命令查看dmp文件头信息即可确认

windows：使用type命令

linux：使用strings命令

D:\>type hr.dmp |more

躈?

g€?

<eg?绪-郤ɡI

T?

?

"SYS"."SYS\_EXPORT\_TABLE\_01"

[oracle@db dpdump]$ strings hr.dmp |more

"SYS"."SYS\_EXPORT\_TABLE\_01"

x86\_64/Linux 2.4.xx

orcl

ZHS16GBK

11.02.00.00.00

001:001:000001:000001

HDR>T<?

8ISL\*

cdgwb

ap    A

/Kf q

以上含有生成系统表"SYS"."SYS\_EXPORT\_TABLE\_01"的为数据泵expdp工具生成的dmp文件

D:\>type hr2.dmp |more

EXPORT:V11.02.00

DSYS

RTABLES

8192

0

72

0

T?

[oracle@db dpdump]$ strings hr2.dmp |more

EXPORT:V11.02.00

DSYS

RTABLES

8192

                                        Thu Mar 22 16:35:38 2018hr2.dmp

#G#G

#G#G

+00:00

以上含有提示EXPORT工具版本EXPORT:V11.02.00 的为传统EXPORT工具生成的dmp文件

》dmp的字符集关系着数据能否正确导入

传统工具EXPORT生成的dmp文件字符集比较复杂

数据泵datapump中保存的字符集比较好确认

1 确认传统EXPORT字符集

了解传统EXPORT生成的dmp文件中保存的字符集时，需要先了解以下知识点

\*传统EXPORT工具是客户端工具，其字符集由NLS\_LANG参数定义

\*Oracle有全球化支持，当客户端字符集和数据库字符集不一致时，会做隐式自动字符集转换

即dmp字符集和数据库字符集不同，也可以导入到数据库中

\*数据库字符集应是导入会话字符集的超集 (或相同), 否则将无法正确转换特殊字符，导致数据丢失或乱码

   -------------------                       for all data pre-oracle9 and data-

   | db in character |     export            dictionary data in oracle9 and up:

   | set A           |---------------------> exp session is in character

   -------------------                       set B as defined by NLS\_LANG.

        source                               Therefore the dump file is in

                                             character set B. Character set

                                             conversion may occur.

                                                       |

                                                       |

                                                       | move file over to

                                                       | another machine

                                                       |

                                                       V

        destination

   -------------------

   | db in character |     import            imp session is in character

   | set D           | <-------------------  set C as defined by NLS\_LANG.

   -------------------                       The dump file is still in

                                             character set B. Character set

                                             conversion may occur.

   During the import process

   character set conversion

   may occur between character

   set C and the db's character

   set D if they differ.

为了最少字符集转换和数据丢失可能性，oracle建议做如下的配置

+export时，设置NLS\_LANG会话字符集和源库数据库字符集相同  --这意味着不需要做字符集转换，dmp文件中保存的字符和源库中的字符一致

+import时，设置NLS\_LANG会话字符集和export时一致  --这意味着在import session阶段不需要做转换

+如果字符集和目标库一致，则不需要转换，如果不一致，则需要进行一次转换，如果目标库字符集不是超集，可能出现数据丢失和乱码

基于以上

传统EXPORT生成的dmp文件中保存的字符集不是数据库字符集，是运行EXPORT工具的客户端设置的NLS\_LANG环境变量指定的字符集

虽然dmp文件可以自动转换字符集导入到数据库中，但如果想确认dmp文件中的字符集可用如下方法

以16进制查看dmp文件头

$ cat hr2.dmp |od -x |head

0000000 0003 4501 5058 524f 3a54 3156 2e31 3230

0000020 302e 0a30 5344 5359 520a 4154 4c42 5345

0000040 380a 3931 0a32 0a30 3237 300a 000a 0301

0000060 0754 00d0 0001 0000 0000 0000 0000 0007

0000100 2020 2020 2020 2020 2020 2020 2020 2020

unix平台大都是 big-endian，文件的开头都是03xx

linux平台大都是little-endian，文件开头都是xx03

第2、3个字节（byte）定义个EXPORT工具使用的字符集

2个16进制表示1个byte，则第2、3个字节为03（0x03），45（0x45）

使用内置函数nls\_charset\_name将charset id转换成charset name

select nls\_charset\_name(to\_number('0345','xxxx')) from dual;

NLS\_CHARSET\_NAME(TO\_NUMBER('0345','XXXX'

----------------------------------------

JA16EUCTILDE

以上显示传统EXPORT工具使用的字符集为JA16EUCTILDE，这是由于导出时未手动设置NLS\_LANG参数，其使用了默认值

注意：如果因字符集不同，发生了转换，在export或者import日志中会有如下类似记录

    import done in US7ASCII character set and AL16UTF16 NCHAR character set

      -> this is the current NLS\_LANG value set in the environment

      and the NCHAR characterset of the target database

    import server uses WE8MSWIN1252 character set (possible charset conversion)

      -> this is only shown if the NLS\_LANG during this import session is different

      from the target database characterset, so if you see 3 lines you might have problems :-)

    export client uses UTF8 character set (possible charset conversion)

      -> this is the characterset used during the export session and the

      characterset used in the dmp file.

2 确认数据泵datapump字符集

数据泵datapump是服务端字符集，不受NLS\_LANG影响，其字符集来源于数据库中NLS\_CHARACTERSET，即dmp文件中的字符集就是数据库中的字符集

其dmp文件的字符集，可以使用前面的命令type或strings

查看dmp文件头信息即可看到。

》是否有办法将此dmp文件导入到一个新的数据库, 而不知道源架构？？

将dmp文件转储成sql文件查看即可

传统的exp工具使用show=y参数

imp \'/ as sysdba\' file=hr2.dmp log=hr.log full=y show=y >hr.sql

数据泵datapump 使用sqlfile参数

impdp \'/ as sysdba\' directory=DATA\_PUMP\_DIR dumpfile=hr.dmp sqlfile=hr.sql full=y exclude=table\_statistics,index\_statistics

知道源架构后，对需要的表空间或者用户，新建还是remap都可以

reference

mos：How to tell whether a .dmp is an EXP oder datapump?

**NLS considerations in Import/Export - Frequently Asked Questions (Doc ID 227332.1)**