OceanBase静态数据生成工具使用文档

# PreRequire

1. 原始数据：多个字段以某个特定分割符为分割。比如：1|2|3|4|5 该原始数据有五个字段，以|为分割符。分割符可以为不可见字符，比如’\0’，但不可以为多个字符。
2. OceanBase 的schema文件。具体参见 OceanBase的schema文件格式。
3. 原始数据和静态数据关系描述文件。该文件格式如下：

[1001] # table id,要与在schema文件中写的一致

#row\_key\_column\_info='the column index in raw data','column type','len'

#row\_key\_column\_info指出由原始数据中的哪几个字段构成OceanBase静态数据中的#rowkey。

#column type:

#0:INT8

#1:INT16

#2:INT32

#3:INT64

#4:VARCHAR

#row key will generate in this order.

row\_key\_column\_info=0,3,8

row\_key\_column\_info=2,0,1

row\_key\_column\_info=1,3,8

#column\_info指出静态数据的列和原始数据的关系。

#column\_info='the column id in schmea','the column index in raw data'

#if some column in the schema but not in raw\_data,just set column\_index to -1

#'the column id in schmea' must be ascending order

column\_info=0,0 #column id为0表示row key

column\_info=0,1

column\_info=0,2

column\_info=2,3

column\_info=3,4

column\_info=4,5

column\_info=5,-1 #如果这一列在原始数据中并不存在，将后一个值设为-1

1. 一般说来，原始数据的文件数量会比较多，可能分布在不同的机器上，如果生成完静态数据还要做多备份，需要将这些机器相互打通。

# 生成数据

准备好以上三个文件以后就可以开始生成数据，生成静态数据的工具为tools目录下的databuilder，使用格式：

./databuilder -s schema\_file #OceanBase的scheam文件

-t table\_id #指定要生成哪个table的静态数据

-f data\_file #原始数据文件

-d dest\_path #静态数据存放地点

-i disk\_no (default is 1) #静态数据存放点的磁盘号。

-b file\_no #原始数据编号。

-x syntax file #原始数据和静态数据关系描述文件

# 元数据

数据生成完毕后会生成一个元数据，命名方式为:

原始数据编号.idx\_磁盘号

原始数据编号即在生成数据的时候-b的参数，磁盘号亦然。

一般说来，原始数据的数量会大于一个，而每个原始数据都会生成这样一个元数据，这就需要将这些元数据合并起来，处理成一个OceanBase要求的(MIN,MAX)的区间。tools下的merge\_meta\_new可以完成这个事情，使用格式：

./merge\_meta\_new -f file\_list -n app\_name -d data\_dir -[rem]

-f 元数据列表

-n 应用名

-d data dir

-r 处理成(MIN,MAX)

-e 删除不存在的文件(处理成多备份时使用)

-m 只做简单的合并(即将多个meta合成一个)。