BCP 工具的使用

The **b**ulk **c**opy **p**rogram utility (**bcp**) bulk copies data between an instance of Microsoft SQL Server and a data file in a user-specified format. The **bcp** utility can be used to import large numbers of new rows into SQL Server tables or to export data out of tables into data files. Except when used with the **queryout** option, the utility requires no knowledge of Transact-SQL. To import data into a table, you must either use a format file created for that table or understand the structure of the table and the types of data that are valid for its columns.

大容量复制程序实用工具 (bcp) 可以在 Microsoft SQL Server 实例和用户指定格式的数据文件间大容量复制数据。使用 **bcp** 实用工具可以将大量新行导入 SQL Server 表，或将表数据导出到数据文件。 除非与 **queryout** 选项一起使用，否则使用该实用工具不需要了解 Transact-SQL知识。 若要将数据导入表中，必须使用为该表创建的格式文件，或者必须了解表的结构以及对于该表中的列有效的数据类型。

如果使用 **bcp** 备份数据，请创建一个格式化文件来记录数据格式。 **bcp** 数据文件 **不包括** 任何架构或格式信息，因此如果已删除表或视图并且不具备格式化文件，则可能无法导入数据。

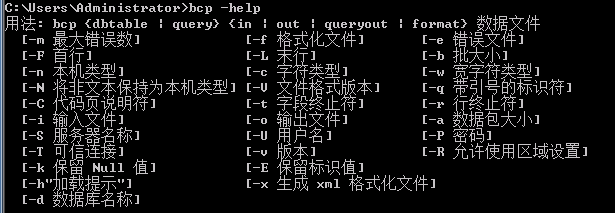
* **版本信息**

我们打开CMD命令 输入 CMD –v 就可以看到BCP是否安装以及其版本



我们在这里使用的是SQL server 2008 R2自带的版本，目前最新的是基于SQL server 2017 的14.0版本。最新版本支持对微软虚拟机上的数据库进行操作

|  |
| --- |
| * **基本语法** |



***数据文件***  
数据文件的完整路径。 将数据批量导入 SQL Server时，数据文件将包含要复制到指定的表或视图中的数据。 从SQL Server中批量导出数据时，数据文件将包含从表或视图中复制的数据。 路径可以有 1 到 255 个字符。 数据文件最多可包含 2^63 - 1 行。

***d******btable***  
指定的表或视图所在数据库的名称。 如果未指定，则使用用户的默认数据库。

你也可以使用 **d-** 显式指定数据库名称。

**in**  | **out**  | **queryout** | **format nul**  
指定大容量复制的方向，具体如下：

* **in** 从文件复制到数据库表或视图。
* **out** 从数据库表或视图复制到文件。 如果指定了现有文件，则该文件将被覆盖。 提取数据时，**请注意 bcp 实用工具将空字符串表示为 null，而将 null 字符串表示为空字符串。**
* **queryout** 从查询中复制，仅当从查询大容量复制数据时才必须指定此选项。
* **format** 根据指定的选项（**-n**、 **-c**、 **-w**或 **-N**）以及表或视图的分隔符创建格式化文件。 大容量复制数据时， **bcp** 命令可以引用一个格式化文件，从而避免以交互方式重复输入格式信息。 **format** 选项要求指定 **-f**选项；创建 XML 格式化文件时还需要指定 **-x** 选项。 有关详细信息，请参阅 [创建格式化文件 (SQL Server)](https://docs.microsoft.com/zh-cn/sql/relational-databases/import-export/create-a-format-file-sql-server?view=sql-server-2017)。必须将 **null** 指定为值 (**format nul**)。

***owner***  
表或视图的所有者的名称。 如果执行该操作的用户拥有指定的表或视图，则*owner* 是可选的。 如果未指定*owner* ，并且执行该操作的用户不是指定的表或视图的所有者，则 SQL Server 将返回错误消息，而且该操作将取消。

**Note：**

**"** ***query*** **"** 是一个返回结果集的 Transact-SQL 查询。 如果该查询返回多个结果集，则只将第一个结果集复制到数据文件，而忽略其余的结果集。 将查询用双引号括起来，将查询中嵌入的任何内容用单引号括起来。 从查询大容量复制数据时，也必须指定。**queryout** 。

只要在执行 bcp 语句之前存储过程内引用的所有表均存在，查询就可以引用该存储过程。 例如，如果存储过程生成一个临时表，则 **bcp** 语句便会失败，因为该临时表只在运行时可用，而在语句执行时不可用。 在这种情况下，应考虑将存储过程的结果插入表中，然后使用 **bcp** 将数据从表复制到数据文件中。

* **数据导出**
* **导出特定表**

从数据库中导出一张表的所有内容：

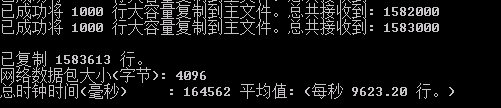
我们输入命令：

bcp wggc.dbo.afp\_reportinfo out C:\report.txt -S localhost\GCPACSWS -U sa -P sa20021224$



BCP会要求我们输入到处字段的数据类型，使用这种方式依次输出，会耗费我们大量的时间，实际上我们可以使用代码页，取消类型的输入操作。命令如下：

bcp wggc.dbo.afp\_reportinfo out –c –T C:\report.txt -S localhost\GCPACSWS -U sa -P sa20021224$



我们看到工具耗时 164秒导出了 1583613条记录，每秒可以处理9623条记录。速度还是比较快速的。导出的txt文件大小为549M， 数据文件较大。

需要注意的是， **bcp 实用工具将空字符串表示为 null，而将 null 字符串表示为空字符串。**

* **导出查询语句**

我们还可以使用查询语句读取我们需要的语句，并导出。

bcp "select top 10 \* from wggc.dbo.AFP\_ReportInfo where CreatedTime >'2016-08-18 15:51:48.497'" queryout C:\test.txt -c -E -k -T -S localhost\GCPACSWS -U sa -P sa20021224$

****

* **导出存储过程**

我们还可以使用执行存储过程，然后获取我们所需的数据导出，需要注意的是，如果存储过程中存在临时表，那么导出操作将不能正常运行。因为命令是去获取表中的信息，无法获取内存空间中的临时数据。

我们新建一个存储过程：

USE [WGGC]

GO

/\*\*\*\*\*\* Object: StoredProcedure [dbo].[AFP\_SP\_AddFilmPrintTask] Script Date: 11/06/2018 10:11:42 \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

-- =============================================

Create PROCEDURE [dbo].BCPTest

AS

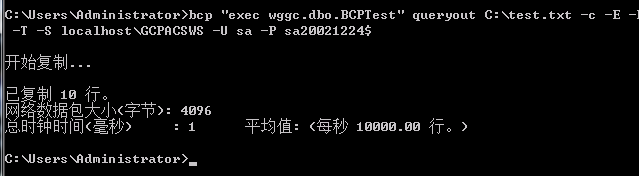
BEGIN

select top 10 \* from wggc.dbo.AFP\_ReportInfo where CreatedTime >'2016-08-18 15:51:48.497'

END

使用以下命令可以将数据导出

bcp "exec wggc.dbo.BCPTest" queryout C:\test.txt -c -E -k -T -S localhost\GCPACSWS -U sa -P sa20021224$



* **数据导入操作：**
* **使用BCP导入**

数据导入操作基本语法和导出相类似，需要须改out to in，例如：

bcp wggc.dbo.afp\_reportinfo in "C:\ test.txt" -c -E -k –T -S localhost\GCPACSWS -U sa -P sa20021224$



大家要注意参数 **–E –k**的使用

**-E：**表示保留标识符，使用该参数的话，导入的时候会保留原有的标识符，而不会自增长。

**-k：**表示保留null值，使用该参数，在导入的时候不会将null导入成空字符。

建议使用**-o + 文件路径将**导入的过程，以日志的形式将记录保留下来

使用BCP导入的优势在于快速，效率高，并且可以避开触发器的相关问题。

* **使用BulkInsert SQL 语句导入**

我们还可以使用SQL语句来导入我们已经备份好的数据文件：

如：

BULK INSERT wggc.dbo.afp\_reportinfo

FROM "C:\ test.txt"

WITH

(

KEEPIDENTITY

)

Bulk insert的语法如下：

BULK INSERT [ [ 'database\_name'.][ 'owner' ].]{ 'table\_name' FROM 'data\_file' }

WITH (

[ BATCHSIZE [ = batch\_size ] ],

[ CHECK\_CONSTRAINTS ],

[ CODEPAGE [ = 'ACP' | 'OEM' | 'RAW' | 'code\_page' ] ],

[ DATAFILETYPE [ = 'char' | 'native'| 'widechar' | 'widenative' ] ],

[ FIELDTERMINATOR [ = 'field\_terminator' ] ],

[ FIRSTROW [ = first\_row ] ],

[ FIRE\_TRIGGERS ],

[ FORMATFILE = 'format\_file\_path' ],

[ KEEPIDENTITY ],

[ KEEPNULLS ],

[ KILOBYTES\_PER\_BATCH [ = kilobytes\_per\_batch ] ],

[ LASTROW [ = last\_row ] ],

[ MAXERRORS [ = max\_errors ] ],

[ ORDER ( { column [ ASC | DESC ] } [ ,...n ] ) ],

[ ROWS\_PER\_BATCH [ = rows\_per\_batch ] ],

[ ROWTERMINATOR [ = 'row\_terminator' ] ],

[ TABLOCK ],

)

详细参数说明：

****

* **Sample1 大批量数据导出导入**

最后我们来看看使用BCP大批量导入数据的性能表现：

导出数据：

bcp wggc.dbo.afp\_reportinfo out C:\report.txt -c -T -S localhost\GCPACSWS -U sa -P sa20021224$

删除数据：

truncate table wggc.dbo.afp\_reportinfo

导入数据：

bcp wggc.dbo.afp\_reportinfo in "C:\report.txt" -c -E -k -T -S localhost\GCPACSWS -U sa -P sa20021224$



我们可以看到数据导入总共耗时549秒，平均每秒可以导入2884.91条数据，性能还是比较好的

* **Sample2 大批量数据导出导入工具**

我个人使用autoit写了一个工具，用于数据的导入导出操作，可以从\\10.184.129.235\puma\Team\Ralf\Tool\BCP\_20180822\_1805中获取。

文件名分别为 Export and import, 该文件夹里面有一些基于1805的历史数据，大家也可以直接使用。

我会逐步优化该工具，以满足项目组的使用。

* **Sample3 BCP在项目中的使用**

之前services team 反映升级过程过于缓慢，其中有部分时间是从film 和 report表中获取信息，并插入到exam表中。我们可以使用BCP或者bulk insert的方式生成新的数据。我使用下列语句：

使用CTE建立子查询， 收集film 和report 表中既有film和report的数据，并导入到新生成的exam表中。

1. 导出数据：bcp "exec wggc.dbo.BCPTest1" queryout C:\exzaminfo.txt -c -E –k -T -S localhost\GCPACSWS -U sa -P sa20021224$

其中film表数据190707条

其中report表数据1583612 条

总共导出数据 81655条，耗时677秒。大部分的时间耗费中数据查询中：



1. 使用语句bcp wggc.dbo.AFP\_Examinfo IN C:\exzaminfo.txt -c -E -k -T -S localhost\GCPACSWS -U sa -P sa20021224$



我们可以看到导入81655行花费8.8秒，速度非常快。