

Manipulating data.: About CHUNKs and BULKs

With the introduction of SQLyog version 5.1 fully configurable **BULK** and **CHUNK** settings were introduced. BULKs have been available for a while (but only little configurable) and CHUNKs is something quite new. The purpose of both is to make it possible to backup and restore huge amounts of data in a fast and reliable way even if server and network resources are limited.

BULKs and CHUNKs have effect with the 'Export as SQL' tool of **all** SQLyog versions and the (Scheduled) Backup 'powertool' of SQLyog Enterprise and SQLyog Ultimate.

The settings for the 'export' tool are available from 'preferences' and for the 'backup' 'powertool' the option is provided by the backup wizard. Now let's see what they are:

CHUNKs: CHUNKs have effect on the (source) server from where data are exported. The CHUNK size is specified in number of rows. If for instance a chunk size of 1000 rows is specified then data will not simply be fetched using one 'SELECT ...' but more SELECTs like 'SELECT ..LIMIT 1,1000', 'SELECT ... LIMIT 1001,1000' will be used until the end of data is reached. This assures that it is possible to specify a CHUNK size that does not exceed various types of resources (like the memory available) to user ... what would result in very slow operation or even maybe a 'hang' or 'deadlock'. Also specifying a not too big CHUNK setting will ensure that no timeout will occur. Such timeout may happen due to server 'net_write_timeout' setting or network settings not specifically related to MySQL. Also note that with HTTP-connection for each SELECT statement the HTTP-connection will have to be re-established. Each re-connect can sometimes take several seconds. So don't specify more CHUNKs/lower CHUNK setting than what you need to in order to back up successfully.

Note that implementation was changed with SQLyog 8.02 release. Before that CHUNK setting had only effect for HTTP-tunnelled connections. From 8.02 onwards it has effect with all types of connections. From version 11.31 CHUNK setting also has effect with 'copy database/table ..'. Before that it only had with backup/export to SQL.

BULKs: BULK on the opposite have effect on the client where the exported data are saved as a file. The BULK size is specified in number of KB (Kilo Bytes). There is a 'standard' setting available that will set a BULK size equal to the max-allowed-packet setting in MySQL configuration. The BULK setting specifies how big each INSERT statement in the saved file will be allowed to be. When that size is exceeded, a new INSERT statement will be generated from the next row of data. Building BULKs has normally practically no influence on speed/performance of the backup process. However in the restore process many and small BULKs can slow down things. However too large BULKs (larger than server max-allowed-packet setting) will cause the server to drop the connection (and SQLyog - or whatever client you use for restoring this dump - will return the MySQL client error 'MySQL server has gone away')

As you see BULKs and CHUNKs operate completely independently of one-another.

The settings that should be taken into consideration when specifying BULKs and CHUNKs are many, but primarily:

Manipulating data.: About CHUNKs and BULKs

- The max-allowed-packet setting in MySQL configuration
- MySQL 'net_write_timeout' setting as well as Storage Engine and User -specific settings in MySQL configuration
- The MAX_POST setting in PHP configuration (if HTTP-tunneling is used)
- Various settings on relays servers on the remote network, proxies etc.

It is not possible to tell what will be the most effective setting. It depends. But note that max-allowed-packet setting in MySQL configuration has a standard value of 1 MB with the most recent MySQL versions, and BULK size cannot exceed that if restore shall be possible. MAX_POST setting in PHP configuration often is 8 MB and that also sets an upper limit on the BULK size (if data shall be restored over a HTTP-connection).

Basically you will have to experiment a little by yourself to find the settings that are optimal for you - and they of course may be different with different hosting providers if you have more. A practical experience with most cheap hosting is that CHUNKs setting of 2000-10000 (rows - depending on how many and what type of columns you have) and BULK setting of 100-500 (KB) normally work fine.

Unique solution ID: #1084

Author: Peter Laursen

Last update: 2014-01-09 11:53