MySQL configuration for batch inserts

Revision 20200615

Goal

This document explains the behavior of the default configuration in MySQL databases for batch inserts and how to configure MySQL properly to take advantage of batch inserts when used as cache or as a data movement enabled data source in Virtual DataPort.

Content

When the cache for a Virtual DataPort view is being populated, the rows obtained from the source are stored in the cache database using batch insertions. The size of the batches is determined by the 'batch insert size' parameter in the cache configuration.
In the same way, when a data source is the target of a Data Movement operation, the Virtual DataPort server inserts data obtained from another data source into the target data source. To speed up the data movement, the INSERT statements are also executed in batches.

When using a MySQL database as a cache database or as a data movement target source, it is important to notice that, by default, MySQL does not allow Batched statements. This implies that Virtual DataPort's cache batch inserts are not going to be effective and all the rows are going to be inserted one at a time:

```
INSERT INTO jdbc (`name`) VALUES ('value_name_1');
INSERT INTO jdbc (`name`) VALUES ('value_name_2');
[...]
```

To take advantage of batch inserts when using MySQL is necessary to change how the MySQL jdbc driver performs the connection. In order to do this, it is just needed to add the property

        'rewriteBatchedStatements=true'

to the connection URL in the configuration of the data source or cache. Once this property is set to true, the rows will effectively be inserted in batches:

```
INSERT INTO jdbc (`name`) VALUES ('value_name_1'),('value_name_2')[...];
```

There is also another important thing to take into account when dealing with batch inserts on MySQL. The property 'max_allowed_packet' defines the maximum size of a single network packet, so this parameter will limit the number of inserts included in a batch. This parameter is defined on the MySQL server side.
It is important to notice that this limit is more restrictive than the Batch size configured in Virtual DataPort for the data source. This way, the number of inserts per batch will be the minimum between the batch insert size defined in VDP and the value of the property 'max_allowed_packet' defined in MySQL.

Configuring MySQL data sources to enable batch insertions is recommended as it will improve the performance when acting as a VDP cache or as a data movement target source.

References

Driver/Datasource Class Names, URL Syntax and Configuration Properties for Connector/J
MySQL and jdbc batch inserts