IRIS/Ensemble as an ETL

Article
Guillaume Rongier · Apr 9, 2019 3m read

Open Exchange

IRIS/Ensemble as an ETL

IRIS and Ensemble are designed to act as an ESB/EAI. This mean they are build to process lots of small messages.

But some times, in real life we have to use them as ETL. The down side is not that they can't do so, but it can take a long time to process millions of row at once.

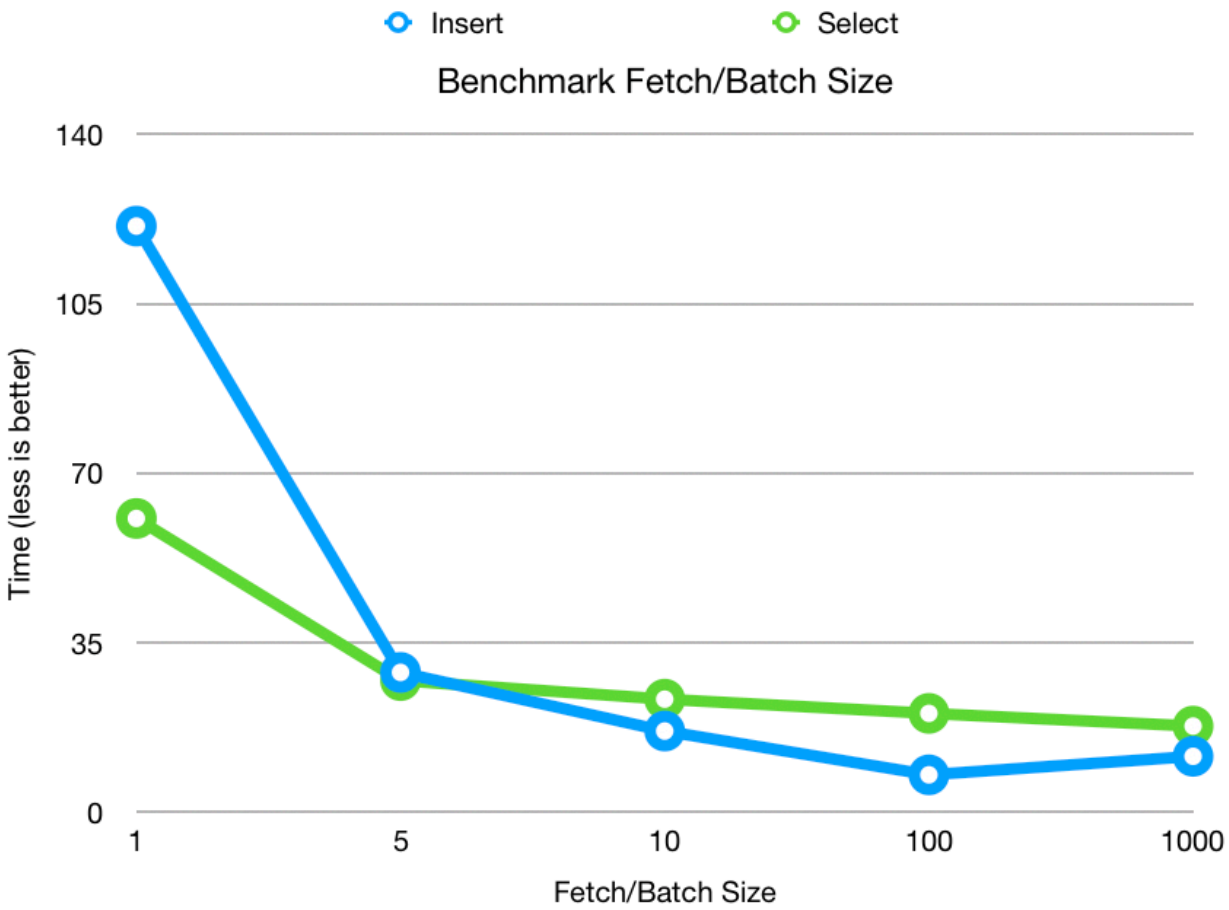
To improve performance, I have created a new SQLOutboundAdaptor who only works with JDBC.

BatchSqlOutboundAdapter

Extend EnsLib.SQL.OutboundAdapter to add batch **batch** and **fetch** support on JDBC connection.

Benchmark

Benchmarks released on Postgres 11.2 with 1 000 000 rows fetched and 100 000 rows inserted on 2 columns.



Prerequisites

Can be used on IRIS or Ensemble 2017.2+.

Installing

Clone this repository

```
git clone https://github.com/grongierisc/BatchSqlOutboundAdapter.git
```

Use Grongier.SQL.SqlOutboundAdapter adaptor.

New methods from the adaptor

- Method **ExecuteQueryBatchParmArray**(ByRef pRS As Grongier.SQL.GatewayResultSet, pQueryStatement As %String, pBatchSize As %Integer, ByRef pParms) As %Status
    - *pRS* is the ResultSet can be use as any EnsLib.SQL.GatewayResultSet
    - *pQueryStatement* is the SQL query you like to execute
    - *pBatchSize* is the fetch size JDBC parameter
- Method **ExecuteUpdateBatchParamArray**(Output pNumRowsAffected As %Integer, pUpdateStatement As %String, pParms...) As %Status

    - *pNumRowsAffected* is the number of row inserted
    - *pUpdateStatement* is teh update/insert SQL statement
    - *pParms* is Caché Multidimensional Array
        - pParms indicate the number of row in batch
        - pParms(integer) indicate the number of parameters in the row
        - pParms(integer,integerParam) indicate the value of the parameter whose position is integerParam.
        - pParms(integer,integerParam,"SqlType") indicate the SqlType of the parameter whose position is integerParam, by default it will be $$$SqlVarchar

Example

- **Grongier.Example.SqlSelectOperation** show an example of ExecuteQueryBatchParmArray
- **Grongier.Example.SqlSelectOperation** show an example of ExecuteUpdateBatchParamArray

Content of this project

This adaptor include :

- Grongier.SQL.Common
    - No modification, simple extend of EnsLib.SQL.Common
- Grongier.SQL.CommonJ
    - No modification, simple extend of EnsLib.SQL.CommonJ
- Grongier.SQL.GatewayResultSet
    - Extension of EnsLib.SQL.GatewayResultSet to gain the ablility to use fetch size.
- Grongier.SQL.JDBCGateway
    - Use to allow compilation and support on Ensemble 2017.1 and lower
- Grongier.SQL.OutboundAdapter
    - The new adaptor with :
    - ExecuteQueryBatchParmArray allow SQL query a distant database and specify the JDBC fetchSize
    - ExecuteUpdateBatchParamArray allow insertion in a distant database with JDBC addBatch and executeBatch
- Grongier.SQL.Snapshot
    - Extend of EnsLib.SQL.Snapshot to handle Grongier.SQL.GatewayResultSet and the fetch size property

#Business Operation #Databases #Interoperability #JDBC #Performance #Ensemble #InterSystems IRIS #InterSystems IRIS for Health
Check the related application on InterSystems Open Exchange

30 1 1 2 611

Log in or sign up to continue
Add reply

Source URL: https://community.intersystems.com/post/irisensemble-etl