

Alya - Task #17943

Implement batch processing: client library

15/02/2024 10:15 AM - Sachin Divekar

Status:	Closed	Start date:	
Priority:	Normal	Due date:	15/03/2024
Assignee:	Sachin Divekar	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	Sprint 29 Mar		
Description			
This includes the library functions to submit a batch, append more jobs to a batch, mark wait-off for a batch, abort a batch, the function to register an initialiser for an app and a processing function for a batch type, etc. It does not include the asynchronous framework which runs the processing functions.			
Related issues:			
Copied to Alya - Task #18088: Implement job manager for slow queries and batches		Closed	05/03/2024 15/03/2024

Associated revisions

Revision ddf0876 - 12/03/2024 07:59 AM - Sachin Divekar

Implement Submit, Done and Abort for batches #17943 #18088 #17944

- Implement Submit, Done and Abort for batches
- Implement Abort for slowquery
- Write programs to use and test them

Revision d4fa2389 - 14/03/2024 10:49 PM - Sachin Divekar

Implement batch Abort and Waitoff refs #17943 #18088 #17944

Implement batch Abort and Waitoff

Refactor code for better organization and simplicity

Revision 5bad17fa - 18/03/2024 11:34 PM - Sachin Divekar

Implement object storage integration #17943 #18088 #17944

Revision 13f184cb - 20/03/2024 11:33 PM - Sachin Divekar

Fix issue with concurrent nsuccess update with batchabort and summarize #17943

- Run all the queries in BatchAbort in the same transaction
- Calculate the counters inside BatchAbort itself after setting the status of the batchrows to Aborted
- Update the summary in batches table inside the BatchAbort itself

Revision 4f8c6fca - 10/04/2024 11:39 PM - Sachin Divekar

Update example programs to use minio #17943 #17944

Revision 87ca45da - 12/04/2024 01:06 AM - Sachin Divekar

Provide MigrateDatabase function to migrate alya batch related tables #17943 #17944

Revision eace4d49 - 12/04/2024 01:27 AM - Sachin Divekar

Convert op (operation) param value to lowercase before inserting into database #17943

Revision 17807588 - 12/04/2024 01:33 AM - Sachin Divekar

Convert op (operation) param value to lowercase before inserting into database #17943 #17944

Revision f1a9ff40 - 15/04/2024 01:21 PM - Sachin Divekar

Use sync.Map for Processor registration #17943 #17944

Replace regular map and mutex with sync.Map to manage the registration and retrieval of slow query processor functions.

Rationale:

- sync.Map is optimized for scenarios where there are many reads and infrequent writes, which aligns with our usage pattern of registering processors once and querying them frequently.
- The use of sync.Map reduces the complexity of manual locking and unlocking, thus lowering the risk of deadlocks and improving performance on concurrent access.

Revision e719dec5 - 15/04/2024 11:37 PM - Sachin Divekar

Write test for RegisterInitializer and add comments #17943 #17944

Revision c79f1f19 - 16/04/2024 12:17 AM - Sachin Divekar

Write a test blobrows handling in batch processing #17943

Revision 8eac6dbe - 16/04/2024 09:59 AM - Sachin Divekar

Add test case for successful batch summarization #17943

Write a test case for the summarizeBatch function in the batch package. The test focuses on the scenario where all rows in a batch are successfully processed and the batch is ready for summarization.

The mocked UpdateBatchSummary method is also set up to verify that the expected batch summary parameters are passed when the method is called during the summarization process. This includes checking the batch ID, status, and counts for successful, failed, and aborted rows.

The test case assumes that all rows in the batch are already processed and have a final status, so mocking the CountBatchRowsByBatchIDAndStatus method is not required in this specific scenario.

Revision 6c70fc00 - 22/04/2024 12:31 PM - Sachin Divekar

Ignore empty content when creating temp files refs #17943

When creating temporary files and appending blob rows to them, check if the content is not empty before proceeding.

Also, trim whitespace from the content before writing it to the file, and add a newline only if the content was successfully written.

Revision 31b49bba - 22/04/2024 12:31 PM - Sachin Divekar

Fix BatchDone method to return correct batchOutput and counts when status is retrieved from Redis #17943

Revision 4cd64afd - 22/04/2024 01:35 PM - Sachin Divekar

Fix issue in BatchDone where it was returning incorrect data #17943

Resolved a critical bug in the BatchDone method where the 'batch' variable was being reassigned, leading to incorrect output of BatchDone. This reassignment was overriding the initial batch data fetched, causing subsequent operations to work with stale or incorrect data.

Revision 7029b6c4 - 22/04/2024 04:37 PM - Sachin Divekar

Example to demonstrate selective output file generation in batch processing example #17943

This example highlights the ability to process transactions in batches without necessarily generating output for every input record. Specifically, the example skips generating output files for certain records based on the line number, showcasing a more flexible handling of batch outputs.

Revision 06f4f3b5 - 22/04/2024 10:55 PM - Sachin Divekar

Add an example to process a large batch #17943

Revision 103dcd27 - 23/04/2024 12:38 AM - Sachin Divekar

Fix a bug related to batch and row status updates in job manager #17943

- Update batch row status to "inprog" before processing each row
- Check and update corresponding batch status to "inprog" if "queued"

Revision ba27321d - 24/04/2024 12:53 AM - Sachin Divekar

Update batch status update in Redis atomically in a transaction #17943

Update the batch status in Redis using a transaction to ensure atomic updates when multiple jobmanager instances are running concurrently. The transaction monitors the key for changes, checks

the current status, and only updates if the status has changed.

This prevents incorrect updates due to concurrent writes.

Revision 94168bf3 - 24/04/2024 09:27 AM - Sachin Divekar

README file for Alya batch package #17943

README file for Alya batch package, describing high-level steps followed by specific explanations and code snippets.

Revision 2389dd5a - 27/04/2024 01:23 PM - Sachin Divekar

Test batch run with multiple concurrent instances of jobmanager and other refactorings for logging, etc. #17943

- Add logharbour logger to JobManager for structured logging
- Introduce JSONstr type for safe JSON handling and validation
- Log batch and row status changes for visibility and auditing
- Refactor code to use JSONstr for context and input parsing

Revision 59867901 - 27/04/2024 02:52 PM - Sachin Divekar

Pass logger to JobManager constructor in all the examples and tests #17943

Revision b9623439 - 27/04/2024 06:45 PM - Sachin Divekar

Add an example of returning wscutils.ErrorMessage from doBatchJob #17943

Revision 22aefe6b - 27/04/2024 10:10 PM - Sachin Divekar

Update logharbour and a small fix #17943

History

#1 - 16/02/2024 06:48 PM - Shuvam Misra

- Target version set to Sprint 15 Mar
- Assignee set to Sachin Divekar
- Due date set to 15/03/2024

#2 - 25/02/2024 09:27 PM - Sachin Divekar

- Status changed from New to In Progress

#3 - 05/03/2024 04:08 PM - Shuvam Misra

- Copied to Task #18088: Implement job manager for slow queries and batches added

#4 - 07/03/2024 05:42 PM - Shuvam Misra

- Description updated

#5 - 07/03/2024 05:45 PM - Shuvam Misra

- Description updated
- Subject changed from Implement batch processing to Implement batch processing: client library

#6 - 08/03/2024 05:53 PM - Shuvam Misra

- % Done changed from 0 to 10

#7 - 12/03/2024 08:02 AM - Sachin Divekar

- % Done changed from 10 to 60

Implemented Submit, Done and Abort for batches and wrote a program to use them and verify their working.

#8 - 13/03/2024 10:59 PM - Sachin Divekar

- % Done changed from 60 to 80

Implement batch.Append()

#9 - 14/03/2024 10:51 PM - Sachin Divekar

Implemented batch Abort and Waitoff.

#10 - 18/03/2024 11:37 PM - Sachin Divekar

- % Done changed from 80 to 90

Wrote examples for Submit, Abort and Append for testing.

#11 - 20/03/2024 03:23 PM - Shuvam Misra

- Target version changed from Sprint 15 Mar to Sprint 29 Mar

#12 - 20/03/2024 11:34 PM - Sachin Divekar

Fix issue with concurrent nsuccess update with batchabort and summarize.

- Run all the queries in BatchAbort in the same transaction
- Calculate the counters inside BatchAbort itself after setting the status of the batchrows to Aborted
- Update the summary in batches table inside the BatchAbort itself

#13 - 26/03/2024 07:19 PM - Shuvam Misra

- Target version changed from Sprint 29 Mar to Sprint 12 Apr

#14 - 26/03/2024 07:21 PM - Shuvam Misra

- Target version changed from Sprint 12 Apr to Sprint 29 Mar

#15 - 26/03/2024 07:22 PM - Shuvam Misra

- Status changed from In Progress to Testing

#16 - 26/03/2024 07:27 PM - Shuvam Misra

- % Done changed from 90 to 100

#17 - 12/04/2024 01:35 AM - Sachin Divekar

Provided MigrateDatabase() function to migrate Alya batch related tables.

Fixed a problem where non-lowercase values for "op" parameter cause issues because we have lowercase check. The updated code converts the value of "op" passed to batch, slowquery and register processor functions to lowercase before using them.

#18 - 16/04/2024 10:01 AM - Sachin Divekar

Wrote automated unit tests for initializers and processors registrations, batch summarization, temporary files creation and handling of blobrows.

#19 - 23/04/2024 01:11 AM - Sachin Divekar

Worked on testing the package. Fixed bugs. Wrote unit tests.

- Fixed issues in the output file where some lines were empty
- Fixed an issue where BatchDone was returning incorrect values when the batch status is present in Redis
- Fixed an issue in BatchDone where it was returning
- Fixed a bug where batch status was not getting updated to inprog
- Wrote an example to process a large batch.

#20 - 29/04/2024 09:04 AM - Sachin Divekar

- Status changed from Testing to Closed