Migration Plugin - Task #7778

Ensure a migration can be interrupted within reasonable time at any stage

11/03/2020 04:25 PM - ttereshc

Status:	CLOSED - COMPLETE	**Start date:**	
Priority:	High	**Due date:**	
Assignee:	dalley	**% Done:**	0%
Category:		**Estimated time:**	0:00 hour
Sprint/Milestone:	0.9.0		
Platform Release:		**Tags:**	Katello
Groomed:	No	**Sprint:**	Sprint 92
Sprint Candidate:	No	**Quarter:**	

Description

Motivation

Users who have pulp 2 and pulp 3 on the same machine need to have a way to control when migration is running. The migration plugin inevitably provides certain load on the system and uses resources. Users might want to choose windows when they run a migration. E.g. Run every Saturday for 4 hours at most.

If the system is large, it can take hours to migrate it, users need a way to cancel the task and free all the related resource within a reasonable timeframe.

Potential issues

When there is a very slow db query and the task which triggered that query is cancelled, how quickly the db query will be aborted as well?

History

#1 - 11/13/2020 05:29 PM - ttereshc

- Priority changed from Normal to High

- Tags Katello added

#2 - 01/19/2021 07:47 PM - dalley

- Sprint set to Sprint 88

#3 - 01/19/2021 08:16 PM - ttereshc

- Sprint/Milestone set to 0.9.0

#4 - 01/22/2021 09:32 PM - rchan

- Sprint changed from Sprint 88 to Sprint 89

#5 - 02/04/2021 11:09 PM - rchan

- Sprint changed from Sprint 89 to Sprint 90

#7 - 02/19/2021 11:08 PM - rchan

- Sprint changed from Sprint 90 to Sprint 91

#8 - 03/05/2021 09:09 PM - rchan

- Sprint changed from Sprint 91 to Sprint 92

#9 - 03/05/2021 10:31 PM - ttereshc

- Status changed from NEW to ASSIGNED

- Assignee set to dalley

#10 - 03/07/2021 05:15 PM - dalley

- Status changed from ASSIGNED to CLOSED - COMPLETE

https://dba.stackexchange.com/questions/81408/is-a-postgres-long-running-query-aborted-if-the-connection-is-lost-broken/81424#81424

```
 <knbk> dalley: I guess it largely depends on when the DB notices the connection is lost. If it's streaming the response for a server-side cursor it should get a write error reasonably quickly and probably kill the query. If there's a big upfront cost where it's not interacting with the connection, that's more likely to continue for a while
 <knbk> and of course it depends on the exact implementation details of the db
```

We use .iterator() in most places, and the places where we don't tend to be very small queries. I'm still investigating mongoengine but I expect it's roughly the same story.

#11 - 03/09/2021 02:09 AM - dalley

For MongoDB:

```
<d4rkp1r4t3> dalley, if the client dies mid-query, similar to a socket timeout, the cursor would remain open in mongodb until it trys to return the resultset back to the client. you can verify this by checking db.currentOp() when this happens. for time sensitive queries use the maxTimeMS setting in the client instead of the socket timeout setting to ensure the cursor is killed after a specific time has passed
```

The general guidelines here seem to be - just make sure that we're using reasonable batch sizes that don't take too long to calculate. For both PostgreSQL and MongoDB.