



This repository Search

Explore Features Enterprise Blog

Sign up

Sign in



neo4j / neo4j

Watch 238

Star 1,885

Fork 730

Parallel relationship import in batch importer

Merged

mneedham merged 1 commit into neo4j:2.2 from tinwelint:2.2-bip-parallel-rel-encoder Apr 17, 2015

Conversation 0

Commits 1

Files changed 30

+819 -47



tinwelint commented Apr 13, 2015

Collaborator

Batches of relationships to be imported are individually analyzed whether or not they can be processed in parallel with its surrounding batches. The deciding factor is whether or not there's any start/end node that is the same as the previous batch. There can be stretches of batches that can be run in parallel, batches that have completely disjoint sets of start/end nodes. Then when there's a batch that has some same start/end node as any in the current stretch of concurrently processing batches then it will await all previous batches to complete before starting. Consecutive batches will continue as described above.

There's a theoretical worst-case scenario where all batches have at least one start/end node id in common, effectively removing the ability to parallelize any batch. This scenario is not very likely and will still have the performance of the relationship import as it was before this commit.

Detection is done by comparing two sorted long[] in a new single-threaded step. The actual sorting is done in an already existing preparation step before this detection step and since preparation is parallelizable it will not become the bottleneck (given enough CPUs available of course).

Relationship import has been one of the most time consuming stages in the batch importer, partly because it usually contains the most data of an import, but also partly because there has been no parallelism in that stage. This commit should (again, given enough available CPUs) ensure that the bottleneck will be raw disk I/O.

Labels

2.2

kernel

Milestone

No milestone

Assignee

mneedham

2 participants



Parallel relationship import in batch importer ... 21149c8

tinwelint added kernel 2.2 labels Apr 13, 2015

mneedham was assigned by chrisvest Apr 14, 2015

mneedham merged commit 42fdec9 into neo4j:2.2 Apr 17, 2015
1 check passed

View details

Sign up for free

to join this conversation on GitHub. Already have an account? Sign in to comment

