



This repository Search

Explore Features Enterprise Blog

Sign up

Sign in

neo4j / neo4j

Watch 238

Star 1,885

Fork 730

# Import tool has ability to skip duplicate nodes

Merged mneedham merged 1 commit into neo4j:2.3 from mneedham:tinwelint-2.3-bip-skip-duplicate-nodes Apr 2, 2015

Conversation 3 Commits 1 Files changed 32

+753 -428



mneedham commented Mar 30, 2015

Collaborator

rather nodes that have the same input ids in the same group. This ability is controlled with `--skip-duplicate-nodes<=true/false>`, whereas the skipping of bad relationships `--skip-bad-relationships<=true/false>`. All and any allowed bad entities (both relationships and duplicate nodes) are collected and controlled using `--bad` and `--bad-tolerance`.

Duplicate nodes are detected after node stage, when preparing IdMapper and deleted as a side-effect in node->relationship stage. For this the handling of not-in-use records is controlled using inUse record flag as opposed to null record due to a detail in the BatchingPageCache. It's generally better to not leave a record to chance, rather write it as inUse or not.

Labels

2.3

kernel

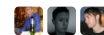
Milestone

No milestone

Assignee

No one assigned

3 participants



mneedham commented Mar 30, 2015

Collaborator

Rebased #4205 and added some tests around BadCollector to help me understand how it works



mneedham referenced this pull request Mar 30, 2015

Import tool has ability to skip duplicate nodes #4205

Closed



mneedham added kernel 2.3 labels Mar 30, 2015



chrisvest commented Apr 1, 2015

Collaborator

@mneedham You forgot to add a license header to the `BadCollectorTest.java` file. Also, this lot needs a rebase.



mneedham commented Apr 1, 2015

Collaborator

@chrisvest done & done



Import tool has ability to skip duplicate nodes ...

a56eb15



mneedham merged commit bafd1ab into neo4j:2.3 Apr 2, 2015  
1 check passed

View details

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



