

Ten. Million. Questions. Let's celebrate [all we've done together](#).

Stack Overflow is a question and answer site for professional and enthusiast programmers. It's 100% free.

Take the 2-minute tour ×

Garbage collection tuning/performance degradation for neo4j bulk .csv import



I am running a bulk import of data into a neo4j instance (I have run against 2.2.0 community and enterprise editions as well as 2.1.7 community) running in server mode. My application creates a bunch of nodes in memory, and will periodically stop to write a series .csv files and send cypher to the neo4j instance to upload the files. (this was done due to performance issues with running the application using the plain old REST API).

Overall, I'm looking to upload something like 150-5000 million nodes, so this is, in principle, the type of thing that neo4j claims to be able to handle relatively well.

Well, anyway, what I'm noticing when I run this against production data is that the application runs in two states -- one where the csv upload processes between 2k-8k of nodes per second, and one where it processes between 80-200 nodes per second. The two states are interwoven when you look at the upload as a time series, and as time goes on, it spends increasingly long amounts of time in the slow state.

Nodes are created through a series of

```
MERGE (:{NODE_TYPE} {csvLine.key = n.primaryKey}) on create set [PROPERTY LIST];
```

statements, and I have indexes on everything that I'm doing merges against. This doesn't feel like a degradation in the insert statements, because the slowdown is not linear, but rather bimodal, this feels like there are garbage collection in the neo4j instance. What is the best way to tune the neo4j JVM garbage collector for frequent bulk inserts?

neo4j.properties:

```
neostore.nodestore.db.mapped_memory=50M
neostore.relationshipstore.db.mapped_memory=500M
#neostore.relationshipgroupstore.db.mapped_memory=10M
neostore.propertystore.db.mapped_memory=100M
#neostore.propertystore.db.strings.mapped_memory=130M
neostore.propertystore.db.arrays.mapped_memory=130M
```

neo4j-wrapper.conf:

```
wrapper.java.additional=-XX:+UseConcMarkSweepGC
wrapper.java.additional=-XX:+CMSClassUnloadingEnabled
wrapper.java.additional=-XX:-OmitStackTraceInFastThrow
wrapper.java.additional=-XX:hashCode=5

wrapper.java.initmemory=8194
wrapper.java.maxmemory=8194
```

This felt like the sweet spot for both the overall heap memory and the neostore stuff. Increasing the overall heap degraded performance. That said, the neo4j garbage collection logs frequently have that GC (Allocation Failure) message.

EDIT: in response to Michael Hunger:

the machine has 64 GB of RAM, and nothing seems to be maxed out. It also seems like only a small number of cores are being used at any time. Garbage collector profiling shows that the garbage collector seems to be running quite frequently.

The exact cypher statements are, for example:

asked 4 months ago

viewed 84 times

active 4 months ago









Blog

 [Why Stack Overflow is a Good Workplace for Women](#)

Related

- 208 [What is JavaScript garbage collection?](#)
- 11 [Tuning Neo4j for Performance](#)
- 1 [Neo4j write performance degradation](#)
- 1 [How to import data by csv in bulk from files in one directory using Cypher in Neo4j?](#)
- 2 [Importing 64k tuples into neo4j](#)
- 2 [Neo4j: Slow input and what garbage collector to use?](#)
- 0 [Neo4j Cypher Load CSV Failure on Unique Constraint](#)
- 1 [How to create relationships \(a->b->c\) among those nodes by importing csv data](#)
- 0 [reduce pause in garbage collection](#)
- 0 [How to import relationships between nodes of the same label from a csv file in neo4j?](#)

Hot Network Questions

-  [How to show pagination "go to page" text box?](#)
-  [Where did the nickname of 'Bones' for McCoy come from?](#)
-  [Lunch with the CEO](#)
-  [How do I know when my sauce is reduced enough?](#)
-  [What is aliasing and what causes it?](#)
-  [Password strength of random words and brainflayer](#)
-  [Why moving fan seems transparent?](#)
-  [Using Emergency Fund to Sell Upside-down Car](#)

```
USING PERIODIC COMMIT 110000 LOAD CSV WITH HEADERS FROM 'file:///home/jschirmer/Even
USING PERIODIC COMMIT 110000 LOAD CSV WITH HEADERS FROM 'file:///home/jschirmer/Even
```

Though there are several such statements being made I have tried a single concurrent transaction as well as running several (~3) such statements in parallel (which gives a roughly 2x improvement). I've tried tuning the periodic commit frequency, and the size of the file. It seems that this maximizes performance when the csv file is roughly 100k lines, which means that really, the periodic commit can be off.

I have not run profile on the statements. I will do that, but I thought that the eager merget problem was avoided by using MERGE ... on create statements.

add a comment
neo4j garbage-collection csv-import

1 Answer share improve this question

edited Apr 8 at 0:43

asked Apr 6 at 18:58

votes

 Jerry Schirmer
80 • 8

- ▲ IN general your config looks ok, what RAM does your machine have?
- 0 For the things you merge against I'd recommend constraint instead of an index.
- ▼ What's your tx size? And how many concurrent tx do you run?

Instead of your generic merge statement (which wouldn't compile) can you share the concrete statements?

Did you profile the statements? Perhaps you run into the eager pipe problem:

<http://www.markhneedham.com/blog/2014/10/23/neo4j-cypher-avoiding-the-eager/>

Do you use periodic commit?

How large are you CSV files?

See: <http://neo4j.com/developer/guide-import-csv/>

share improve this answer

answered Apr 7 at 21:38

 Michael Hunger
25.2k • 2 • 21 • 41

updated the OP. – Jerry Schirmer Apr 8 at 0:44


add a comment


Your Answer


B *I*         

Sign up or [log in](#)

Post as a guest

 Sign up using Google

 Sign up using Facebook

















 Sign up using Stack Exchange

Name

Email

required, but never shown

Post Your Answer

-  Does touching the tent outer wall from the inside make it leak?
-  Make subscript size smaller (always)
-  When does the designer pay for a reprint
-  Shut down list of servers given in a file
-  How do I remove Candy Crush Saga from Windows 10?
-  How to deal with people looking at their mobile phone during my presentation?
-  Name for a device purposefully put together from faulty parts
-  Seven overlapping circles
-  Select polygons one by one and export them using arcpy
-  I added more water to my ciabatta dough by accident
-  What are some techniques to avoid touching a wet tent wall?
-  Is it modest to compliment a woman on her modesty?
-  Did Arnold Rimmer kill the Red Dwarf crew?
-  Could an autotrophic civilisation develop, or will evolving life forms always eat each other?
-  Can a plane fly backward if its prop's in reverse?
-  The Hiding Rules Contradict Themselves. What Can I Make Of This?

Not the answer you're looking for? Browse other questions tagged [neo4j](#) [garbage-collection](#) [csv-import](#) or [ask your own question](#).



[tour](#) [help](#) [blog](#) [chat](#) [data](#) [legal](#) [privacy policy](#) [work here](#) [advertising info](#) [mobile](#) [contact us](#) [feedback](#)

TECHNOLOGY

- Stack Overflow
- Server Fault
- Super User
- Web Applications
- Ask Ubuntu
- Webmasters
- Game Development
- TeX - LaTeX
- Programmers
- Unix & Linux
- Ask Different (Apple)
- WordPress Development
- Geographic Information Systems
- Electrical Engineering
- Android Enthusiasts
- Information Security
- Database Administrators
- Drupal Answers
- SharePoint
- User Experience
- Mathematica
- Salesforce
- ExpressionEngine® Answers
- more (13)

LIFE / ARTS

- Photography
- Science Fiction & Fantasy
- Graphic Design
- Movies & TV
- Seasoned Advice (cooking)
- Home Improvement
- Personal Finance & Money
- Academia
- more (9)

CULTURE / RECREATION

- English Language & Usage
- Skeptics
- Mi Yodeya (Judaism)
- Travel
- Christianity
- Arqade (gaming)
- Bicycles
- Role-playing Games
- more (21)

SCIENCE

- Mathematics
- Cross Validated (stats)
- Theoretical Computer Science
- Physics
- MathOverflow
- Chemistry
- Biology
- more (5)

OTHER

- Stack Apps
- Meta Stack Exchange
- Area 51
- Stack Overflow Careers