Best way to create millions of relationships for existing nodes

asked 4 months ago
viewed 36 times
active 4 months ago

1

I have an existing graph with about 1.5mln nodes with label Foo. Each Foo node has a property prop1, which is from a significantly smaller subset (a few thousand values). We now need to be able to traverse the graph using this property (it will be connected to other labels too), and we'd like to convert it into nodes (rather than property).

So basically, if I have

```
Foo1{prop1:1}
Foo2{prop1:2}
Foo3{prop1:1}
```

then I'd like to create 2 new nodes, Prop1{id:1} and Prop1{id:2}, and link them with a new relationship PROP1.

```
Foo1 -[:PROP1]-> Prop1{id:1}
Foo2 -[:PROP1]-> Prop1{id:2}
Foo3 -[:PROP1]-> Prop1{id:1}
```

I hope it makes sense.

The fundamental problem is that since it's 1.5mln nodes, trying to do it in one go simply runs OOM. I managed to come up with a cypher query that generally does small parts of it (I might have messed up something with the syntax as I'm writing from memory, but that was roughly how it worked):

```
MATCH (n:Foo) WHERE NOT(n-[:PROP1]->())
WITH n LIMIT 10000
MERGE (p:Prop1 {id = n.prop1})
MERGE (n) -[:PROP1] -> (p)
```

Unfortunately running it manually over and over again is a rather boring task ;) so I'm looking for 1) either way to loop it until it's finished (but commits in between parts) OR 2) some other way to execute this in smaller chunks.

Any hints anybody?

neo4j

share improve this question

asked Mar 30 at 19:41
Lili
124 ● 1 ● 8

How are you trying to do this in one query? That's probably the right approach. If you're running out of memory the way you're formulating that query may be problematic. You might want to post it. – FrobberOfBits Mar 30 at 19:57

If I run it in a single query it's just without the limit, MATCH (n:Foo) WHERE NOT(n-[:PROP1]->()) MERGE (p:Prop1 {id = n.prop1}) MERGE (n) -[:PROP1] -> (p) - it keeps running and running and running... and in the logs I just see GC overhead messages. I also tried on an untouched backup, without the WHERE clause (as I knew no relationships existed) and it had same problem. When I do that for a small subset it works fine, so either I need to have much, much more patience (it was running for quite a few minutes before I killed it), or I need to split it into chunks... – Lili Mar 30 at 20:08

I'm sorry I can't copy the actual exact query - it's for a client and unfortunately absolutely everything there is top secret and can't be copy-pasted (yes, that REALLY helps solving issues ;)) :/ – Lili Mar 30 at 20:12

add a comment

1 Answer

active oldest votes

Blog
Why Stack Overflow is a Good Workplace for Women

Related

0 Not able to create the relationships with existing nodes in batch processing
2 Create node with unique relationships in Neo4j
1 Create relationship between nodes having same property value in common, using one Cypher query
0 In Neo4J create relationship between two nodes, but if one of nodes don't exists create it in the same call
0 Create new Node with Relationship to existing Node
0 creating relationship between nodes with foreach
0 batch insertion of millions of records on the pre-existing nodes
1 How to create multiple nodes with cypher in neo4j
0 Creating multiple relationships to existing nodes with Cypher
1 Best way to delete all nodes and relationships in Cypher

Hot Network Questions

0

What happens if you run a slightly different but simple bulk query:

```
MERGE (prop1:Prop1 { id: 1 })
WITH prop1
MATCH (f:Foo { prop1: 1 })
MERGE (f)-[:PROP1]->(prop1)
RETURN count(f);
```

Note that if you `MERGE` the relationship between two nodes, you shouldn't really need the `OPTIONAL MATCH` bit you're doing.

share improve this answer

answered Mar 30 at 20:14

FrobberOfBits

8,397 ● 7 ● 29

Well, that only solves the problem partially - first, there are thousands of values for prop1 (so I need to execute it with 1, 2, 20, 5000... - except they are not actually sequential numbers but simply some sort of identifier). I wanted to look them up by checking what is actually referenced from Foos. When I was at work I run the match (f:Foo) merge (p:Prop1 {id : f.prop1}) merge (f) -[:PROP1]-> (p) and it completed after about 20 mins - but only created 600k relationships (the rest was already there, as I was running it batched with LIMITs the day before and these relationships stayed). – Lili Mar 31 at 20:07

Then I tried doing the same on a similar another property - and it never completed (after about 2 hours I gave up). So... it seems the query itself is okay, but it is simply too much for it to take in one go - so I'm looking what is the best way to chop it into chunks. Is there some kind of equivalent of periodic commit as exists for CSV import? – Lili Mar 31 at 20:08

Are you indexing your nodes? – FrobberOfBits Mar 31 at 20:16

Foos are indexed on their id, but not on that property - since I expect to run through pretty much all of the nodes I didn't think that was useful - do you think it would help? I can try adding an index on prop1 (well, prop2 now :)) tomorrow and see if that helps. – Lili Mar 31 at 20:51

Or did you mean adding indexing on Prop1 label ? – Lili Mar 31 at 20:52

show **1** more comment

Your Answer

Sign up or log in

Sign up using Google

Sign up using Facebook

Sign up using Stack Exchange

Post as a guest

Name

Email

required, but never shown

Post Your Answer

By posting your answer, you agree to the privacy policy and terms of service.

Not the answer you're looking for? Browse other questions tagged neo4j or ask your own question.

Select polygons one by one and export them using arcpy

Legacy application unreachable code

How does java.util.EnumSet<E> work?

Antonym of Overlap?

I added more water to my ciabatta dough by accident

How do I calculate approximate equity liquidity?

Can a Glaive be stowed in a Ruby Scabbard?

Using Emergency Fund to Sell Upside-down Car

Proving that all integers are even or odd

Dullness vs. going overboard: Should I be calling people 'enfants terribles' in an academic paper?

Approximation of Borel sets by a countable collection of majorants

Could an autotrophic civilisation develop, or will evolving life forms always eat each other?

Did a Kamikaze ever impact the hull of a ship, as opposed to the deck?

Seven overlapping circles

An "outside the box" sequence

If static methods can't be overriden, how its working here (For Java)?

How can I grant a non-Super User group permission to force check-in of articles?

What is aliasing and what causes it?

Weird performance increase in simple benchmark

Would allowing shorter passwords sometimes be more secure?

Funding government by only printing more money

Why moving fan seems transparent?

Was クリスマスケーキ used metaphorically about women?

Words of the Buddha

question feed