

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 x

Why would the same load csv in neo4j take exponential longer when run a second time (on a clean db)?

Add  projects to your  stackoverflowcareers profile.

1 I am new to using Neo4j so my first attempting at loading the database was a learning experiment. I realized when I attempted some of the queries I wanted to do, I didn't create the right model. I cleared the database, from the command line, using: `rm -rf data/*` and started over (after stopping the database and starting it after). The first 3 loads were almost exactly the same as the first time I imported data.

```
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS from 'file' AS line
WITH line
MERGE (p:Provider {pid:line.pid});

USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS from 'file' AS line
WITH line
MERGE (c:Credential {name:credential});

USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS from 'file' AS line
WITH line
MATCH (p:Provider {pid:line.pid});
MATCH (c:Credential {name:credential});
MATCH (p)-[:IS_A]->(c);
```

The only difference the second time I ran these load statements, was the first node was given two labels:

```
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS from 'file' AS line
WITH line
MERGE (p:Provider:Person {pid:line.pid});

USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS from 'file' AS line
WITH line
MERGE (c:Credential {name:credential});

USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS from 'file' AS line
WITH line
MATCH (p:Person {pid:line.pid});
MATCH (c:Credential {name:credential});
MATCH (p)-[:IS_A]->(c);
```

The first time I ran these 3 imports, it took maybe 20 minutes. However, the second time, it has been running for 3 days. The first two loads, were still pretty fast, maybe 5 minutes each. It is the relationship that has been running ever since. I can not figure out why it is taking so long.

csv neo4j

share improve this question

asked Apr 6 at 13:57




Mel18  
75 • 4

asked 4 months ago

viewed 14 times

active 4 months ago





#### Blog

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

#### Related

- 5 [How to save neo4j database?](#)
- 1 [Neo4j: is there a way to optimize input for CSV LOAD, like regex?](#)
- 1 [How to set relationship type and label in LOAD CSV?](#)
- 0 [Why import a 10X bigger file takes way more time than 10X time of the original fiel in Neo4j?](#)
- 1 [Create nodes and relations conditionally when loading nodes from csv in Neo4j](#)
- 0 [Neo4j creating relationships using csv](#)
- 1 [Neo4j LOAD CSV extremely slow when merging relations from a label of 3mil nodes to a lable of 20 nodes](#)
- 1 [Should I use Neo4j's Import Tool or Load Command to Insert Several Million Rows?](#)
- 0 [neo4j shell -creating edges from .csv](#)
- 0 [Problems to load .csv files generated in Social Network Benchmark to Neo4j](#)

#### Hot Network Questions

-  [Dullness vs. going overboard: Should I be calling people 'enfants terribles' in an academic paper?](#)
-  [Aerofoil that gives reasonably good lift for both flow directions: forward and backward](#)
-  [I added more water to my ciabatta dough by accident](#)
-  [Preventing a Print expression from generating a new cell each time it executed](#)

I copied my code over wrong for the last load. It should be: USING PERIODIC COMMIT 1000 LOAD CSV WITH HEADERS from 'file' AS line WITH line MATCH (p:Person {pid:line.pid}) MATCH (c:Credential {name:credential}) MERGE (p)-[:IS\_A]->(c); – [Mel18](#) Apr 6 at 14:40

[add a comment](#)

1 Answer

active oldest votes

Do you have indexes on the properties?

0

```
CREATE INDEX ON :Person(pid)
CREATE INDEX ON :Credential(name)
```



Alternatively, if you want uniqueness to be enforced on the columns you should create a constraint (this also creates an index):

```
CREATE CONSTRAINT ON (n:Person) ASSERT n.pid IS UNIQUE
CREATE CONSTRAINT ON (n:Credential) ASSERT n.name IS UNIQUE
```

Also, I think your last `MATCH` should be a `MERGE`, and you shouldn't have semi-colons:

```
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS from 'file' AS line
WITH line
MATCH
  (p:Person {pid:line.pid}),
  (c:Credential {name:credential})
MERGE (p)-[:IS_A]->(c);
```

[share](#) [improve this answer](#)

answered Apr 6 at 14:04



[Brian Underwood](#)

4,364 • 4 • 18

I have uniqueness constraints set up. For Person, I used just the `CREATE CONSTRAINT ON (n:Provider) ASSERT n.pid IS UNIQUE`, would I need to do that for the Person label as well, i.e. `CREATE CONSTRAINT ON (n:Person) ASSERT n.pid IS UNIQUE`? For the last `MATCH` you are right, its suppose to be a `MERGE`. That was a copy error. – [Mel18](#) Apr 6 at 14:35

I think adding the Person constraint or index would help a lot with performance – [Brian Underwood](#) Apr 6 at 15:22

Great thanks! I think thats it. I just needed an index/uniqueness constraint for every label. – [Mel18](#) Apr 6 at 16:07

[add a comment](#)

Your Answer

**B** *I* [link](#) [quote](#) [code](#) [image](#) [list](#) [table](#) [undo](#) [redo](#)

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](#)

[✈ Can a plane fly backward if its prop's in reverse?](#)

[📖 Justification that 'good exposition' is uniquely defined](#)

[💬 `cond` with less redundancy](#)

[🗺 Carrying a handgun in other countries with a US concealed carry permit](#)

[📁 Shut down list of servers given in a file](#)

[📖 Why does Scala allow nested data structures like List or Array](#)

[🔗 Using siunitx, \ohm results in an italic Omega](#)

[👤 Where did the nickname of 'Bones' for McCoy come from?](#)

[💬 Is it possible to automate tests for in progress sprint](#)

[👑 Is it modest to compliment a woman on her modesty?](#)

[👤 Why did Holly decide to simply wait out the radioactivity?](#)

[📈 Weird performance increase in simple benchmark](#)

[🔧 Should you always minimize cognitive load](#)

[👤 Is there a single-word antonym for "Opportunity"?](#)

[🚗 Using Emergency Fund to Sell Upside-down Car](#)

[💻 The Programming Language Quiz](#)

[?? A short, brutal riddle](#)

[🔧 How to show pagination "go to page" text box?](#)

[📖 What's the philosophy behind delaying writing data to disk](#)

[👤 Name for a device purposefully put together from faulty parts](#)

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

 [question feed](#)

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

#### TECHNOLOGY

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

#### 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