

Neo4j ›

## 2.2.0 - LOAD CSV performance dramatically degrades after 25 millions relationships

3 posts by 2 authors  



Andrii Stesin

Apr 8



Hi neo4j gurus,

let me describe the problem we hit while evaluating 2.2.0 enterprise ed. We are trying to import some sample data which is basically just pretty chaotic graph with about 50 millions nodes and about 450 millions relationships between them (unevenly distributed where 90% relationships hit 10% nodes but this does not matter as of now). Data has chaotically distributed duplicates of some nodes and some relationships.

Each node has its string property `key_id` which is correctly indexed **with schema index w/ unique constraint** and index is Ok, up and working.

First of all, we created All Our Nodes with massive LOAD CSV and it went fine they got into database correctly (uniquely) indexed by their `key_id`

Than we split our relationships into 100k-lined CSV files as simple as possible - just source node ID, destination node ID, and some 3 properties for each relationships.

Wrapped query:

```
USING PERIODIC COMMIT 10000
LOAD CSV FROM "file:///tmp/rels.csv" AS line
MATCH (p1:PP {key_id: line[0]})
WITH p1, line
MATCH (p2:PP {key_id: line[1]})
WITH p1, p2, line
MERGE (p1)-[l1:LINK]->(p2)
ON CREATE SET l1.x = line[2], l1.y = line[3], l1.z =line[4];
```

EXPLAIN-ing query confirms, that both lookups for p1 and p2 nodes are correct index lookups.

Server is given `-Xms=10240 -Xmx=10240` options. No other configuration options were tweaked, cache is HPC as by default.

First 100k relationships CSV file took about 4 minutes to load - not too brilliant, but acceptable speed, The problem is that after about 25 millions of relationships the 100k file load time just jumped up to 40 minutes.

That was not a slow degradation of performance time from file to file - times were about 4:00-4:30 all the way and at some point just hit some barrier where dramatically increased.

What may be the cause of this? Any suggestions?

Thanks in advance!

WBR,  
Andrii



**Andrii Stesin**

Apr 23



Seems that no one have any ideas? :(

We were starting from a clean fresh empty database, btw.



**cui r**

Apr 28



We run into the dense node issue, i.e., a node with a lot of relationships. If this is your case, try to set the commit size to 100 or so.  
That helped us.

On Thursday, April 23, 2015 at 8:12:26 AM UTC-4, Andrii Stesin wrote:

| Seems that no one have any ideas? :(