schema index not being used

7 posts by 3 authors

Clark Richey   Apr 15

Hello,
When I run EXPLAIN on this query I can see that the index isn't being used. See below. However, when I execute the schema command I can see that that index is on-line. (ON :Place(_geocoded) ONLINE)

Why isn't the query performing a nodeIndexSeek?

```
neo4j-sh (?)$ explain match (p:Place)where p._geocoded = "true" return count(p);
+-------------------------------------------+
| No data returned, and nothing was changed. |
+-------------------------------------------+
10 ms

Compiler CYPHER 2.2

Planner COST

EagerAggregation
  |
  +Filter
    |
    +NodeByLabelScan


+-----------------+--------------+-------------+-----------------------------+
|        Operator | EstimatedRows | Identifiers |                       Other |
+-----------------+--------------+-------------+-----------------------------+
| EagerAggregation |         1592 |    count(p) |                             |
|          Filter |      2535411 |         p | p._geocoded == {  AUTOSTRING0} |
|   NodeByLabelScan |      7606234 |         p |                      :Place |
+-----------------+--------------+-------------+-----------------------------+
```

Clark D. Richey, Jr
CHIEF TECHNOLOGY OFFICER

240.252.7507
CLARK@FACTGEM.COM
WWW.FACTGEM.COM

Clark Richey
clark....@gmail.com

Michael Hunger   Apr 15   ◀◀

Good question, perhaps it decided that the index selectivity was too low as it is a boolean index?

You can force it and see how it changes:

    explain

    match (p:Place)

using index p:Place(_geocoded)

    where p._geocoded = "true"

    return count(p);

- show quoted text -
- show quoted text -

<fg-logo.png>

Clark D. Richey, Jr
CHIEF TECHNOLOGY OFFICER
240.252.7507
CLARK@FACTGEM.COM
WWW.FACTGEM.COM

Clark Richey
clark....@gmail.com

Clark Richey   Apr 15   ◀◀

It's actually supposed to be a String index.See node value below

Node[425]{uniqueID:"Place:a1447d21-1fb2-426f-bba0-e91ddb98ee7b",_type:"Place",confidenceScore:1, dateModified:1428639760183,source:"Ohio Secretary of State",state:"OH",street1:"P.O. Box 494",city:"West Union",zip:"45693",lat:38.7945166,lon:-83.5451934,formattedAddress:"West Union, OH 45693, USA",_geocoded:"true"}

There are some values that are 'true' and 'false' (as Strings) but there are other values as well. How does neo determine what type of index to create?

Results forcing the index are below with changing count to returning nodes shows the index being used.

neo4j-sh (?)$ explain match (p:Place) using index p:Place(_geocoded) where p._geocoded = "true" return p;;

```
+-------------------------------------------+
| No data returned, and nothing was changed. |
+-------------------------------------------+
```
65 ms

Compiler CYPHER 2.2

Planner COST

NodeIndexSeek

```
+---------------+---------------+-------------+------------------+
|     Operator | EstimatedRows | Identifiers |         Other |
+---------------+---------------+-------------+------------------+
| NodeIndexSeek |     2535411 |         p | :Place(_geocoded) |
+---------------+---------------+-------------+_____+
```

However, if I do a count it still doesn't use the index:

neo4j-sh (?)$ explain match (p:Place) using index p:Place(_geocoded) where p._geocoded = "true" return count(p);

```
+-------------------------------------------+
| No data returned, and nothing was changed. |
+-------------------------------------------+
```
1 ms

Compiler CYPHER 2.2

Planner COST

EagerAggregation
  |
  +NodeIndexSeek

```
+------------------+---------------+-------------+------------------+
|       Operator | EstimatedRows | Identifiers |         Other |
+------------------+---------------+-------------+------------------+
| EagerAggregation |         1592 |   count(p) |                |
|   NodeIndexSeek |     2535411 |         p | :Place(_geocoded) |
+------------------+---------------+-------------+------------------+
```

Clark Richey
clark....@gmail.com

- show quoted text -

Michael Hunger   Apr 15   ⬱

Am 15.04.2015 um 22:47 schrieb Clark Richey <clark....@gmail.com>:

It's actually supposed to be a String index.See node value below

Node[425]{uniqueID:"Place:a1447d21-1fb2-426f-bba0-e91ddb98ee7b",_type:"Place",
confidenceScore:1,dateModified:1428639760183,source:"Ohio Secretary of
State",state:"OH",street1:"P.O. Box 494",city:"West Union",zip:"45693",lat:38.
7945166,lon:-83.5451934,formattedAddress:"West Union, OH 45693, USA",_geocoded:"true"}

There are some values that are 'true' and 'false' (as Strings) but there are other values as well.
How does neo determine what type of index to create?

It creates the string index, but afaik it gets information from the index how selective it is, e.g. if it rather stores
4 different values or 4M.

Results forcing the index are below with changing count to returning nodes shows the index
being used.

-> still a lot of rows which would explain that it chose the label-scan

```
neo4j-sh (?)$ explain match (p:Place) using index p:Place(_geocoded) where p._geocoded =
"true" return p;;
+------------------------------------------+
| No data returned, and nothing was changed. |
+------------------------------------------+
65 ms

Compiler CYPHER 2.2

Planner COST

NodeIndexSeek

+--------------+--------------+------------+------------------+
|    Operator | EstimatedRows | Identifiers |         Other |
+--------------+--------------+------------+------------------+
| NodeIndexSeek |    2535411 |        p | :Place(_geocoded) |
+--------------+--------------+------------+——————————————+
```

However, if I do a count it still doesn't use the index:

-> what do you mean? It clearly says NodeIndexSeek ?
- show quoted text -

Clark Richey

Apr 15

Thanks

Yes it did the directIndex. Reading is hard sometimes.

Sent from my iPhone
- show quoted text -

Andres Taylor   Apr 15

Hi there!

Index seeks are slow, compared to label scans. So when a lookup is expected to return a large portion of the nodes with a label, doing a label scan is cheaper. You can see in estimatedRows that Neo4j estimated that 2.5M nodes will be returned by an index seek, which sounds like a lot. How many nodes have the

Andrés
- show quoted text -

Andres Taylor   Apr 15

...hit send way too soon there...

I was about to say - how many nodes have the :Place label on them? And, have you run the query using PROFILE to see how many rows are actually returned?

Andrés