



PROJECTE NEO4J : Padrons

Bases de Dades no Relacionals

Cèlia Martínez
Marta Matute
Goretti Pena
Judit Ugas

Introducció

Exercici 1: importació de dades

La importació de les dades per a crear el graf que farem servir per a les quèries es pot trobar al [GitHub](#), dins la carpeta "Code". Hem optat per escriure tota la informació en anglès (noms de les propietats, relacions, etc). En total, la importació de dades està formada per 8 comandes executades una darrera l'altra:

Una comanda per a assegurar-nos que la base de dades està buida:

```
match (n) detach delete n;
```

Dues comandes per a carregar tots els nodes:

```
// LOAD PEOPLE DATA
LOAD CSV WITH HEADERS FROM 'https://docs.google.com/spreadsheets/d/e/2PACX-
1vTfU6oJBZhmhzzkV_0-
avABPzHTdXy8851ySDbn2gq32WwaNmYxfiBtCGJG0ZsMgCWjz1EGX4Zh1wqe/pub?output=csv'
AS row
CREATE (p:PERSON {Id: toInteger(row.Id), Year: toInteger(row.Year), Name: row
.name, Last_Name: row.surname, Second_Last_Name: row.second_surname});

// LOAD HOUSE DATA
LOAD CSV WITH HEADERS FROM 'https://docs.google.com/spreadsheets/d/e/2PACX-
1vT0Zhr6BSO_M72JEmxXKs6GLuOwxm_Oy-
0UruLJeX8_R04KAcICuvrwn20ENQhtuvddU5RSJSclHRJf/pub?output=csv' AS row
CREATE (:HOUSE {Municipality: row.Municipi, Home_Id: toInteger(row.Id_Llar),
Registry_Year: toInteger(row.Any_Padro), Street: row.Carrer, Number: toIntege
r(row.Numero)});
```

Dues comandes per a crear índexs en algunes de les propietats (això ens servirà per accelerar el temps de creació de les relacions i les quèries), una per a esborrar qualsevol índex anterior, i l'altre per a crear els que ens calen:

```
// DELTE ANY EXISTING INDEXES
CALL apoc.schema.assert({}, {}, true) YIELD label, key
RETURN *;

// CREATE INDEXES (FOR FASTER DATA CREATION)
create index for (p:PERSON) on (p.Id);
create index for (p:PERSON) on (p.Year);
create index for (h:HOUSE) on (h.Home_Id);
create index for (h:HOUSE) on (h.Municipality);
create index for (h:HOUSE) on (h.Registry_Year);
```

I per últim tres comandes que creen les tres relacions entre nodes que necessitem:

```
// CREATION OF "LIVES" RELATIONSHIP BETWEEN A PERSON AND A HOUSE
LOAD CSV WITH HEADERS FROM 'https://docs.google.com/spreadsheets/d/e/2PACX-
1vRM4DPeQFmv7w6kLH5msNk6_Hdh1wuExRirgysZK0_Q70L21MKBkDISIyjdvm8shVix15Tcw_5zC
fdg/pub?output=csv' AS row
MATCH (p:PERSON {Id: toInteger(row.IND), Year: toInteger(row.Year)})
```

```

MATCH (h:HOUSE {Municipality: row.Location, Home_Id: toInteger(row.HOUSE_ID),
Registry_Year: toInteger(row.Year)})
MERGE (p)-[:LIVES {Year: toInteger(row.Year)}]->(h);

// CREATION OF "SAME_AS" BETWEEN TWO PEOPLE
LOAD CSV WITH HEADERS FROM 'https://docs.google.com/spreadsheets/d/e/2PACX-
1vTgC8TBmdXhjUOPKJxyiZSpetPYjaRC34gmXHj6H2AWvXTGbg7MLKVdJnwh5bIeer7WLUi00igI
6wc/pub?output=csv' AS row
MATCH (p:PERSON {Id: toInteger(row.Id_A)})
MATCH (q:PERSON {Id: toInteger(row.Id_B)})
MERGE (p)-[:SAME_AS]->(q);

// CREATION OF "FAMILY" RELATIONSHIP BETWEEN TWO PEOPLE
LOAD CSV WITH HEADERS FROM 'https://docs.google.com/spreadsheets/d/e/2PACX-
1vRV0oMAMoxHiGboTjCIHo2yT30CCWgVHgocGnVJxiCTgyurtmqCfAFahHajobVzwXFLwhqajz1fq
A8d/pub?output=csv' AS row
MATCH (p:PERSON {Id: toInteger(row.ID_1)})
MATCH (q:PERSON {Id: toInteger(row.ID_2)})
MERGE (p)-
[rel:FAMILY {Relation: row.Relacio, Harmonized_Relation: row.Relacio_Harmonit
zada}]->(q);

```

Exercici 2: Queries

Un cop s'han carregat les dades, l'exercici 2 consistia en resoldre 10 queries:

1. Dels padró de 1866 de Castellví de Rosanes (CR), retorna el número d'habitants i la llista de noms. Elimina duplicats i nan.

```
neo4j$ match (h:HOUSE)←[:LIVES]-(p:PERSON) where p.Last_Name <>
'nan' and h.Registry_Year = 1866 and h.Municipality = "CR"
return count(p) as `Num habitants`, collect(distinct
p.Last_Name) as Llistat
```

	Num habitants	Llistat
1	336	["olle", "galceran", "suñol", "rusell", "julibert", "bargallo", "anglada", "vila", "ros", "julia", "julivert", "gaset", "rurr"]

Started streaming 1 records in less than 1 ms and completed after 1 ms.

2. Dels padrons de Sant Feliu de Llobregat (SFLL) d'abans de l'any 1840 (no inclòs), retorna la població, l'any del padró i la llista d'identificadors dels habitatges de cada padró. Ordena els resultats per l'any de padró.

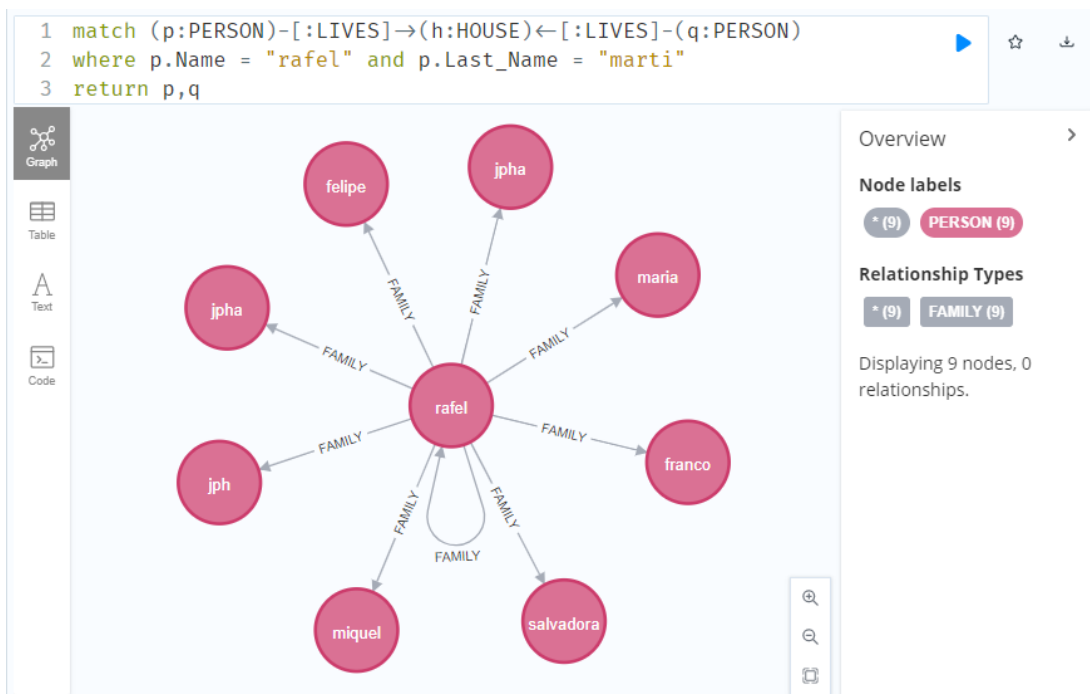
```
1 match (h:HOUSE)←[:LIVES]-(p:PERSON)
2 where h.Registry_Year < 1840 and h.Municipality = "SFLL"
3 with count(h) as poblacio,h
4 return distinct h.Registry_Year as `Any`, count(h) as Població,
   collect(DISTINCT h.Home_Id) as `Llista llars`
5 order by h.Registry_Year
```

	Any	Població	Llista llars
1	1833	304	[95, 99, 101, 103, 105, 107, 109, 111, 94, 96, 97, 98, 108, 100, 104, 102, 106, 110, 113, 115, 118, 119]
2	1838	55	[321, 324, 326, 328, 320, 322, 323, 325, 327, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339]
3	1839	377	[721, 722, 723, 724, 725, 731, 733, 734, 726, 735, 727, 729, 730, 728, 732, 629, 631, 633, 634, 637]

Started streaming 3 records in less than 1 ms and completed after 4 ms.

3. Retorna el nom de les persones que vivien al mateix habitatge que "rafel marti" (no té segon cognom) segons el padró de 1838 de Sant Feliu de Llobregat (SFLL). Retorna la informació en mode graf i mode llista.

En mode de graf:



I en mode de llista:

```
1 match (p:PERSON)-[:LIVES]-(h:HOUSE)<-[:LIVES]-(q:PERSON)
2 where p.Name = "rafel" and p.Last_Name = "martí"
3 return p.Name as nom, collect(q.Name) as convivents
```

	nom	convivents
1	"rafel"	["salvadora", "franco", "maria", "jpha", "felipe", "jpha", "jph", "miquel"]

Started streaming 1 records after 4 ms and completed after 15 ms.

4. Retorna totes les aparicions de "Miguel ballester". Fes servir la relació SAME_AS per poder retornar totes les instàncies, independentment de si hi ha variacions lèxiques (ex. diferents formes d'escriure el seu nom/cognoms). Mostra la informació en forma de subgraf.

```
1 match (p:PERSON)-[:SAME_AS]-(q:PERSON)
2 where p.Name = "miguel" and p.Last_Name = "ballester"
3 return p,q;
```

Overview

Node labels

- * (4) PERSON (4)

Relationship Types

- * (6) SAME_AS (6)

Displaying 4 nodes, 0 relationships.

5. Mostra totes les persones relacionades amb "antonio farran". Mostra la informació en forma de taula: el nom, cognom1, cognom2, i tipus de relació.

```
1 match (p:PERSON)-[rel]-(q:PERSON)
2 where p.Name = "antonio" and p.Last_Name = "farran"
3 return q.Name as Nom, q.Last_Name as Cognom1, q.Second_Last_Name
   as Cognom2, type(rel) as Tipus;
```

	Nom	Cognom1	Cognom2	Tipus
1	"isidro"	"farran"	"colet"	"FAMILY"
2	"catalina"	"farran"	"colet"	"FAMILY"
3	"francisco"	"farran"	"colet"	"FAMILY"
4	"esperanza"	"farran"	"colet"	"FAMILY"
5	"esperanza"	"colet"	"gavarro"	"FAMILY"
6	"antonio"	"farran"	"sole"	"FAMILY"
7	"antonio"	"ferran"	"sele"	"SAME_AS"
8	"antonio"	"ferran"	"sole"	"SAME_AS"

Started streaming 8 records after 4 ms and completed after 11 ms.

6. Llisteu totes les relacions familiars que hi ha.

```
1 match (p:PERSON)-[r:FAMILY]-(q:PERSON)
2 where r.Harmonized_Relation <> 'null'
3 return distinct r.Harmonized_Relation as `Relacions Familiars`;
```

	Relacions Familiars
1	"esposa"
2	"fill"
3	"filla"
4	"jefe"
5	"net"
6	"gendre"
7	

Started streaming 47 records after 3 ms and completed after 35 ms.

7. Identifiqueu els nodes que representen el mateix habitatge (carrer i numero) al llarg dels anys de Sant Feliu del Llobregat (SFL). Mostreu el resultat dels habitatges que tingueu totes dues informacions (carrer i numero), el nombre total d'habitatges, el llistat d'anys dels padrons i el llistat de les Ids de les llars. Ordeneu de més a menys segons el total d'habitatges i mostreu-ne els 10 primers.

```
1 match (h1:HOUSE) match (h2:HOUSE)
2 where h1.Street = h2.Street and h1.Number = h2.Number and
  h1.Home_Id <> h2.Home_Id and h1.Municipality = "SFL" and
  h2.Municipality = "SFL" and h1.Street <> "null" and h1.Number
  <> "nan"
3 return h1.Street as Carrer, h1.Number as Numero, count(distinct
  h2) as Total, collect(distinct h2.Registry_Year) as any,
  collect(DISTINCT h2.Home_Id) as Ids
4 order by Total desc limit 10
```

	Carrer	Numero	Total	any	Ids
1	"falguera"	5	9	[1833, 1878, 1881, 1889]	[247, 359, 360, 361, 416, 398, 414, 415, 417]
2	"falguera"	3	8	[1833, 1878, 1881, 1889]	[245, 357, 358, 668, 411, 397, 412, 413]
3	"san antonio"	1	7	[1889]	[562, 563, 564, 565, 566, 567, 568]
4	"falguera"	22	7	[1833, 1878, 1881, 1889]	[269, 376, 377, 378, 429, 430]
5	"carretera"	28	6	[1838, 1878, 1881, 1889]	[346, 347, 34, 35, 26, 128]
6	"iglesia"	3	6	[1833, 1839, 1878, 1881, 1889]	[97, 723, 724, 516, 511, 484]

Started streaming 10 records after 1 ms and completed after 36 ms.

8. Mostreu les famílies de Castellví de Rosanes amb més de 3 fills. Mostreu el nom i cognoms del cap de família i el nombre de fills. Ordeneu-les pel nombre de fills fins a un límit de 20, de més a menys.

```
1 match (h:HOUSE)-[:LIVES]-(p:PERSON)-[r:FAMILY]→(q:PERSON)
2 where r.Harmonized_Relation =~ "f.*" and r.Harmonized_Relation
   < "familiar" and h.Municipality = "CR"
3 with count(distinct q) as total, p.Name as nom, p.Last_Name as
   `1er cognom`, p.Second_Last_Name as `2n Cognom`
4 where total>3
5 return nom, `1er cognom`, `2n Cognom`, total
6 order by total desc
7 limit 20
```

	nom	1er cognom	2n Cognom	total
1	"pablo"	"astruch"	"julia"	7
2	"jose"	"olle"	"domenech"	6
3	"benito"	"julivert"	"parera"	6
4	"jose"	"canals"	"olle"	6
5	"pedro"	"bargallo"	"ilegible"	6
6	"jose"	"canals"	"mila"	6
7				

Started streaming 20 records in less than 1 ms and completed after 2 ms.

9. Mitja de fills a Sant Feliu del Llobregat l'any 1881 per família. Mostreu el total de fills, el nombre d'habitatges i la mitja.

```
1 call{
2   match (l:HOUSE)
3   where l.Municipality='SFL' and l.Registry_Year=1881
4   return count(l.Home_Id) as NumLlars
5 }
6 match (h:HOUSE)←[:LIVES]-(p:PERSON)-[r:FAMILY]→(q:PERSON)
7 where (r.Harmonized_Relation =~ "f.*" and r.Harmonized_Relation
8   <> "familiar") and p.Year = 1881 and h.Municipality = "SFL"
9 with p as persones, count(distinct q) as nombre_fills, NumLlars
10 return sum(nombre_fills) as total_fills, NumLlars as num_llars,
11   (sum(nombre_fills)*1.0)/NumLlars as mitja;
```

	total_fills	num_llars	mitja
1	1292	596	2.1677852348993287

Started streaming 1 records after 54 ms and completed after 68 ms.

10. Per cada any que hi ha a la base de dades, quin és el carrer amb menys habitants de Sant Feliu de Llobregat?

```
1 MATCH (p:PERSON)-[r:LIVES]→(h:HOUSE {Municipality: 'SFL'})
2 WITH COLLECT(r.Year) AS year, p, h
3 UNWIND h.Street as st
4 WITH count(p) as per, year, st
5 ORDER BY per
6 RETURN year, COLLECT(st)[0], min(per)
7 ORDER BY year
```

	year	COLLECT(st)[0]	min(per)
1	[1833]	"cartera de la part de molins de rey"	5
2	[1838]	"carretera de barna"	30
3	[1839]	"casas del 3onmany"	3
4	[1878]	"carrretera"	2
5	[1881]	"Carretera"	5
6	[1889]	"s n antonio"	1

Started streaming 6 records in less than 1 ms and completed after 32 ms.

Exercici 3: Anàlisi del graf

- Explicació raonada de cada exercici
- Treball en equip
-