

PROJECTE Neo4J

**Comparativa sobre diferents esquemes de base de dades
al tema de padrons històrics de població**



Laia Alcalde, Cristina Huanca, Elena Gutiérrez i Laia Cámara

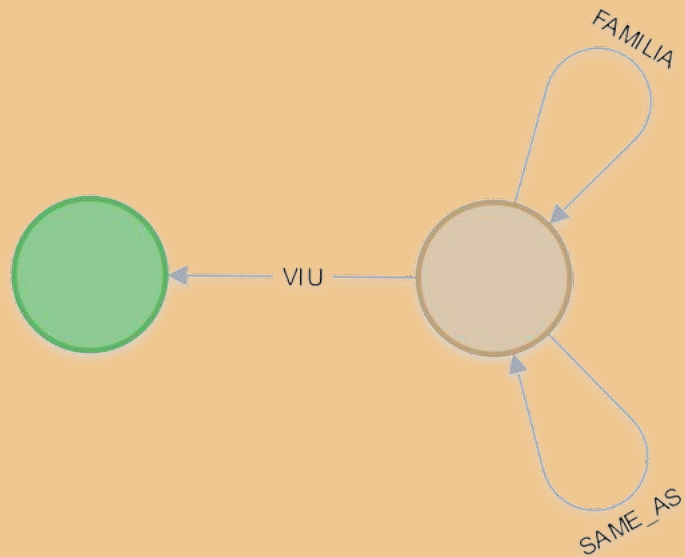
ÍNDEX

1. Proposta d'esquema
2. Importació de dades
3. Consultes Cypher
4. Avantatges i inconvenients
5. Conclusions
6. Acomiadament

ESQUEMA ORIGINAL

Individu

Habitatge

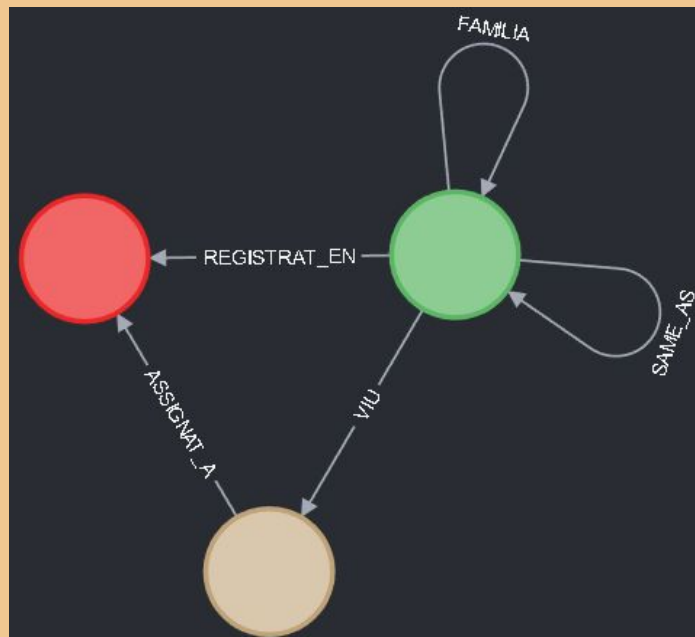


PROPOSTA D'ESQUEMA

Individu

Habitatge

Cens



PROPOSTA D'ESQUEMA

(:Individu)-[:**REGISTRAT_EN**]->(:Cens) → empadronament!

(:Habitatge)-[:**ASSIGNAT_A**]->(:Cens) → segons any i municipi!

(:Individu)-[:**VIU**]->(:Habitatge) → on viu actualment!

(:Individu)-[:**FAMILIA**]->(:Individu) → vincles!

(:Individu)-[:**SAME_AS**]->(:Individu) → instància per registre!

IMPORTACIÓ DE DADES

Creació de constraints i índexs

Creació de nodes

Creació de relacions

Creació de constraints i índexs

```
CREATE CONSTRAINT individu_id_unique IF NOT EXISTS  
FOR (i:Individu) REQUIRE i.id IS UNIQUE;  
  
CREATE CONSTRAINT habitatge_id_unique IF NOT EXISTS  
FOR (h:Habitatge) REQUIRE h.id IS UNIQUE;  
  
CREATE CONSTRAINT cens_unique IF NOT EXISTS  
FOR (c:Cens) REQUIRE (c.any, c.municipi) IS NODE KEY;
```



Creació de nodes

2.1 Nodes Cens

```
LOAD CSV WITH HEADERS FROM 'file:///HABITATGES.csv' AS row
WITH DISTINCT row.Any_Padro AS any, row.Municipi AS municipi
WHERE any IS NOT NULL AND municipi IS NOT NULL AND any <> "" AND
municipi <> "" AND any <> "null" AND municipi <> "null"
MERGE (:Cens {any: toInteger(any), municipi: municipi});
```

2.2 Nodes Habitatge

```
LOAD CSV WITH HEADERS FROM 'file:///HABITATGES.csv' AS row
WITH row
WHERE row.Id_Llar IS NOT NULL AND row.Id_Llar <> "" AND row.Id_Llar
<> "null"
MERGE (h:Habitatge {id: toInteger(row.Id_Llar)})
SET h.carrer = row.Carrer,
    h.numero = CASE
        WHEN row.Numero IS NULL OR row.Numero = "" OR
row.Numero = "null"
        THEN null
        ELSE row.Numero
    END;
```

2.3 Nodes Individu

```
LOAD CSV WITH HEADERS FROM 'file:///INDIVIDUAL.csv' AS row
WITH row
WHERE row.Id IS NOT NULL AND row.Id <> "" AND row.Id <> "null"
MERGE (i:Individu {id: toInteger(row.Id)})
SET i.nom = row.name,
    i.cognom = row.surname,
    i.segonCognom = row.second_surname,
    i.any = toInteger(row.Year);
```


Creació de relacions

3.1 (:Habitatge)-[:ASSIGNAT_A]->(:Cens)

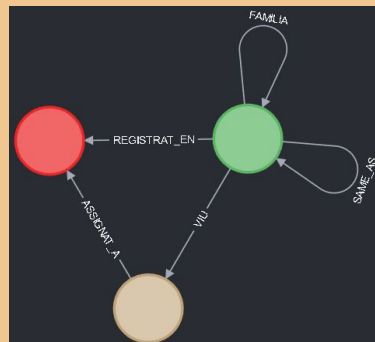
```
LOAD CSV WITH HEADERS FROM 'file:///HABITATGES.csv' AS row
WITH row
WHERE row.Id_Llar IS NOT NULL AND row.Id_Llar <> "" AND
row.Id_Llar <> "null"
MATCH (h:Habitatge {id: toInteger(row.Id_Llar)})
MATCH (c:Cens {any: toInteger(row.Any_Padro), municipi:
row.Municipi})
MERGE (h)-[:ASSIGNAT_A]->(c);
```

3.2 (:Individu)-[:VIU]->(:Habitatge)

```
LOAD CSV WITH HEADERS FROM 'file:///VIU.csv' AS row
WITH row
WHERE row.IND IS NOT NULL AND row.IND <> "null" AND
row.HOUSE_ID IS NOT NULL AND row.HOUSE_ID <> "null"
MATCH (i:Individu {id: toInteger(row.IND)})
MATCH (h:Habitatge {id: toInteger(row.HOUSE_ID)})
MERGE (i)-[:VIU]->(h);
```

3.3 (:Individu)-[:REGISTRAT_EN]->(:Cens)

```
LOAD CSV WITH HEADERS FROM 'file:///VIU.csv' AS row
WITH row
WHERE row.IND IS NOT NULL AND row.IND <> "null" AND row.Year
IS NOT NULL AND row.Location IS NOT NULL
MATCH (i:Individu {id: toInteger(row.IND)})
MATCH (c:Cens {any: toInteger(row.Year), municipi:
row.Location})
MERGE (i)-[:REGISTRAT_EN]->(c);
```



Creació de relacions

3.4 (:Individu)-[:FAMILIA]->(:Individu)

```
LOAD CSV WITH HEADERS FROM 'file:///FAMILIA.csv' AS row
WITH row
WHERE row.ID_1 IS NOT NULL AND row.ID_1 <> "null"
  AND row.ID_2 IS NOT NULL AND row.ID_2 <> "null"
MATCH (i1:Individu {id: toInteger(row.ID_1)})
MATCH (i2:Individu {id: toInteger(row.ID_2)})
WITH i1, i2, row,
  row.Relacio AS rel,
  row.Relacio_Harmonitzada AS rel_h
FOREACH (_ IN CASE WHEN rel IS NOT NULL AND rel <> "null" AND
rel_h IS NOT NULL AND rel_h <> "null" THEN [1] ELSE [] END |
  MERGE (i1)-[r:FAMILIA]->(i2)
  SET r.relacio = rel, r.relacio_harmonitzada = rel_h
)
FOREACH (_ IN CASE WHEN rel IS NULL OR rel = "null" OR rel_h IS
NULL OR rel_h = "null" THEN [1] ELSE [] END |
  MERGE (i1)-[:FAMILIA]->(i2)
);
```

3.5 (:Individu)-[:SAME_AS]-(:Individu)

```
LOAD CSV WITH HEADERS FROM 'file:///SAME_AS.csv' AS row
WITH row
WHERE row.Id_A IS NOT NULL AND row.Id_B IS NOT NULL AND
row.Id_A <> "null" AND row.Id_B <> "null"
MATCH (i1:Individu {id: toInteger(row.Id_A)})
MATCH (i2:Individu {id: toInteger(row.Id_B)})
MERGE (i1)-[:SAME_AS]->(i2);
```

CONSULTES CYPHER

CONSULTA 1 - Totes les persones registrades al cens de l'any 1866 i municipi "CR"

—NOU ESQUEMA—

```
MATCH (i:Individu)-[:REGISTRAT_EN]->(c:Cens)
WHERE c.any = 1866 AND c.municipi = 'CR'
RETURN i.nom AS Nom, i.id AS ID, c.any AS Any, c.municipi
AS Municipi
LIMIT 10;
```

| | Nom | ID | Any | Municipi |
|----|-----------|----------|------|----------|
| 1 | "rosa" | 18660002 | 1866 | "CR" |
| 2 | "antonio" | 18660003 | 1866 | "CR" |
| 3 | "rosa" | 18660004 | 1866 | "CR" |
| 4 | "elisa" | 18660005 | 1866 | "CR" |
| 5 | "manuel" | 18660006 | 1866 | "CR" |
| 6 | "miguel" | 18660007 | 1866 | "CR" |
| 7 | "emilia" | 18660008 | 1866 | "CR" |
| 8 | "isidro" | 18660009 | 1866 | "CR" |
| 9 | "paula" | 18660010 | 1866 | "CR" |
| 10 | "jaine" | 18660011 | 1866 | "CR" |

CONSULTES CYPHER

CONSULTA 1 - Totes les persones registrades al cens de l'any 1866 i municipi "CR"

—ESQUEMA ORIGINAL—

```
MATCH (i:Individu)-[:VIU_EN]->(h:Habitatge)
WHERE h.any = 1866 AND h.municipi = "CR"
RETURN i.nom AS Nom, i.id AS ID, h.any AS Any,
h.municipi AS Municipi
LIMIT 10;
```

| | Nom | ID | Any | Municipi |
|----|-----------|----------|------|----------|
| 1 | "rosa" | 18660002 | 1866 | "CR" |
| 2 | "antonio" | 18660003 | 1866 | "CR" |
| 3 | "rosa" | 18660004 | 1866 | "CR" |
| 4 | "elisa" | 18660005 | 1866 | "CR" |
| 5 | "manuel" | 18660006 | 1866 | "CR" |
| 6 | "miguel" | 18660007 | 1866 | "CR" |
| 7 | "emilia" | 18660008 | 1866 | "CR" |
| 8 | "isidro" | 18660009 | 1866 | "CR" |
| 9 | "paula" | 18660010 | 1866 | "CR" |
| 10 | "jaime" | 18660011 | 1866 | "CR" |

CONSULTES CYPHER

CONSULTA 2 - Totes les persones que viuen en habitatges assignats al cens 1881 de “SFLl”

—NOU ESQUEMA—

```
MATCH(i:Individu)-[:VIU]->(h:Habitatge)-[:ASSIGNAT_A]->(c:Cens)
WHERE c.any = 1881 AND c.municipi = 'SFLl'
RETURN i.nom AS Nom, i.id AS ID, h.id AS HabitatgeID,
c.any AS Any, c.municipi AS Municipi
LIMIT 10;
```

| | Nom | ID | HabitatgeID | Any | Municipi |
|----|-----------|----------|-------------|------|----------|
| 1 | "rosa" | 18660002 | 1 | 1881 | "SFLl" |
| 2 | "antonio" | 18660003 | 1 | 1881 | "SFLl" |
| 3 | "rosa" | 18660004 | 1 | 1881 | "SFLl" |
| 4 | "elisa" | 18660005 | 1 | 1881 | "SFLl" |
| 5 | "manuel" | 18660006 | 1 | 1881 | "SFLl" |
| 6 | "miguel" | 18660007 | 1 | 1881 | "SFLl" |
| 7 | "emilia" | 18660008 | 1 | 1881 | "SFLl" |
| 8 | "jose" | 18660001 | 1 | 1881 | "SFLl" |
| 9 | "martin" | 18780001 | 1 | 1881 | "SFLl" |
| 10 | "manuela" | 18780002 | 1 | 1881 | "SFLl" |

CONSULTES CYPHER

CONSULTA 2 - Totes les persones que viuen en habitatges assignats al cens 1881 de “SFLl”

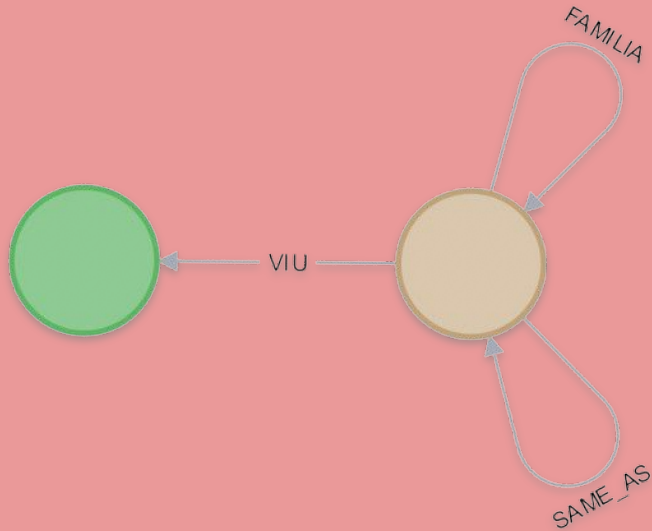
—ESQUEMA ORIGINAL—

```
MATCH (i:Individu)-[:VIU_EN]->(h:Habitatge)
WHERE h.any = 1881 AND h.municipi = "SFLl"
RETURN i.nom AS Nom, i.id AS ID, h.any AS Any, h.municipi
AS Municipi
LIMIT 10;
```

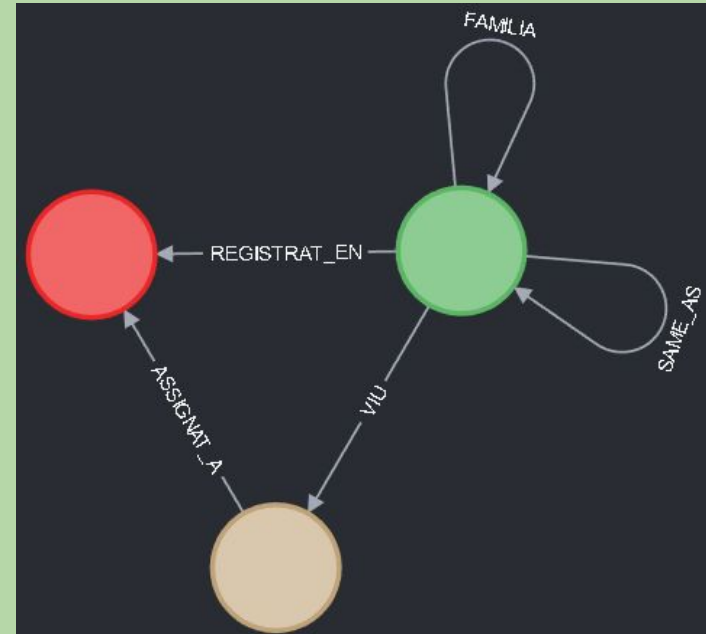
| | Nom | ID | HabitatgeID | Any | Municipi |
|----|-----------|----------|-------------|------|----------|
| 1 | "rosa" | 18660002 | 1 | 1881 | "SFLl" |
| 2 | "antonio" | 18660003 | 1 | 1881 | "SFLl" |
| 3 | "rosa" | 18660004 | 1 | 1881 | "SFLl" |
| 4 | "elisa" | 18660005 | 1 | 1881 | "SFLl" |
| 5 | "manuel" | 18660006 | 1 | 1881 | "SFLl" |
| 6 | "miguel" | 18660007 | 1 | 1881 | "SFLl" |
| 7 | "emilia" | 18660008 | 1 | 1881 | "SFLl" |
| 8 | "jose" | 18660001 | 1 | 1881 | "SFLl" |
| 9 | "martin" | 18780001 | 1 | 1881 | "SFLl" |
| 10 | "manuela" | 18780002 | 1 | 1881 | "SFLl" |

AVANTATGES I INCONVENIENTS

Esquema Original



Nou Esquema Proposat



AVANTATGES I INCONVENIENTS

Inconvenients

Major complexitat, un node de més.

Possible redundància, sense informació de cens.

Duplicació d'informació si es registra en diferents censos.

Avantatges

Millor traçabilitat amb node cens, facilitar anàlisi temporal.

Relació clara entre habitatge i cens.

Més ampliabilitat, més dades.

Més fidel a dades oficials.



CONCLUSIONS

Més complet i flexible, informació més detallada.

Complexitat més gran, node de més.

Millor per guardar dades reals, més informació.



Esquema de la Figura 3 (original)

Relacions:

(:Individu)-[:VIU]->(:Habitatge)

(:Individu)-[:FAMILIA]->(:Individu)

(:Individu)-[:SAME_AS]->(:Individu)

Esquema Proposat

Nodes:

:Individu

:Habitatge

:Cens (afegits)

Relacions:

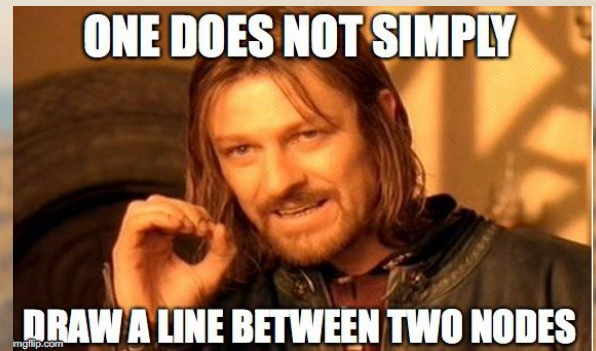
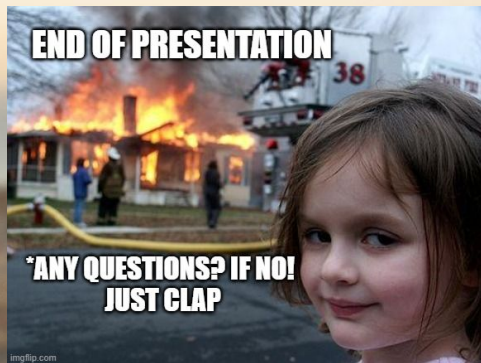
(:Individu)-[:REGISTRAT_EN]->(:Cens)

(:Habitatge)-[:ASSIGNAT_A]->(:Cens)

(:Individu)-[:VIU]->(:Habitatge)

(:Individu)-[:FAMILIA]->(:Individu)

(:Individu)-[:SAME_AS]->(:Individu)



**GRÀCIES PER
L'ATENCIÓ!**



Laia Alcalde, Cristina Huanca, Elena Gutiérrez i Laia Cámara