

# Des datas & des graphes

Benoit Simard (@logisima)



# Presentation

# Benoit Simard

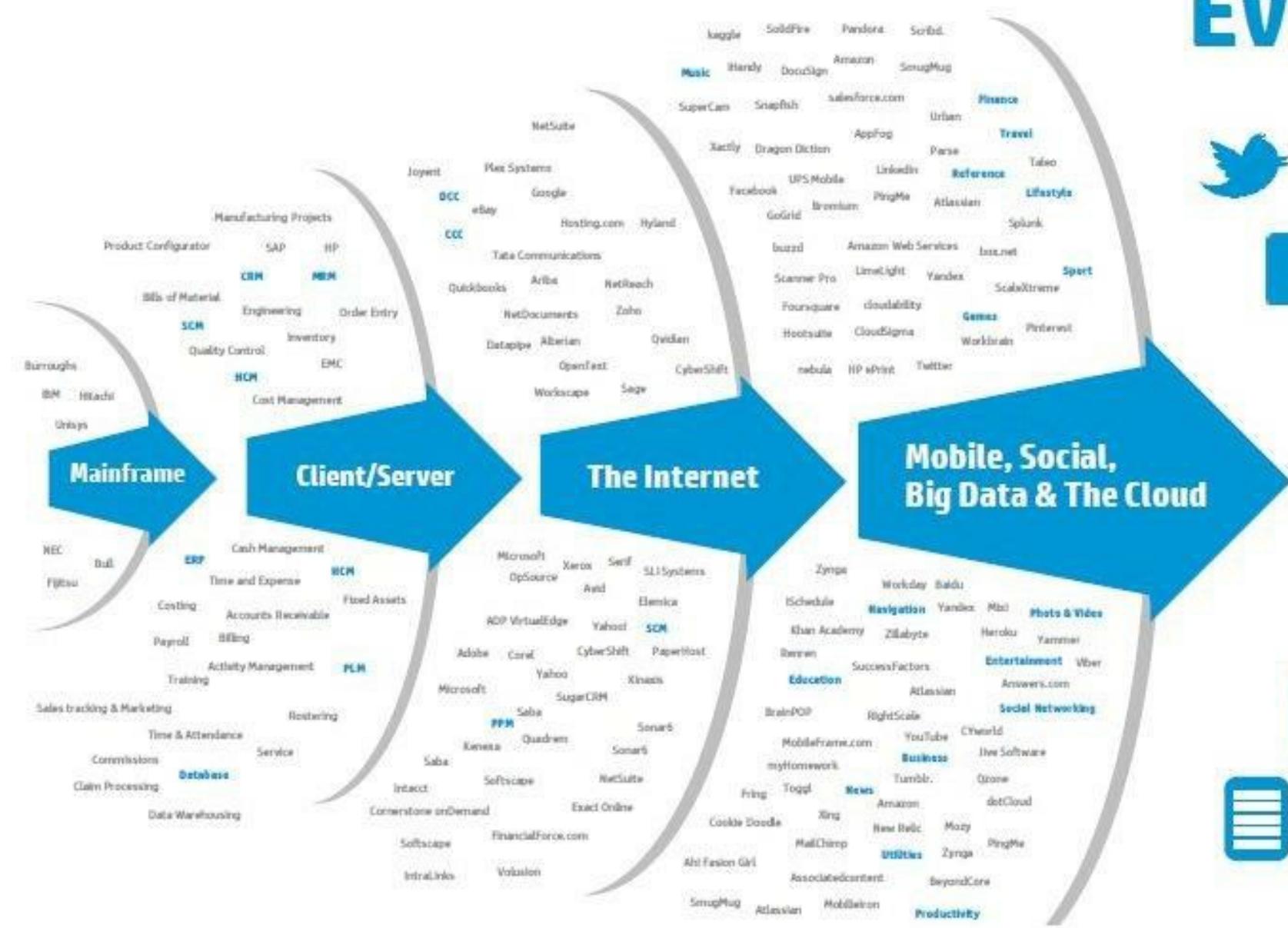
- Ancien de l'IMA
- **Graphe addicte**
- Libérateur de données
- [benoit.simard@neotechnology.com](mailto:benoit.simard@neotechnology.com)
- @logisima



# La data

# Une question de taille ?

Tous les deux ans, on double la quantité de données produites



# Every 60 seconds

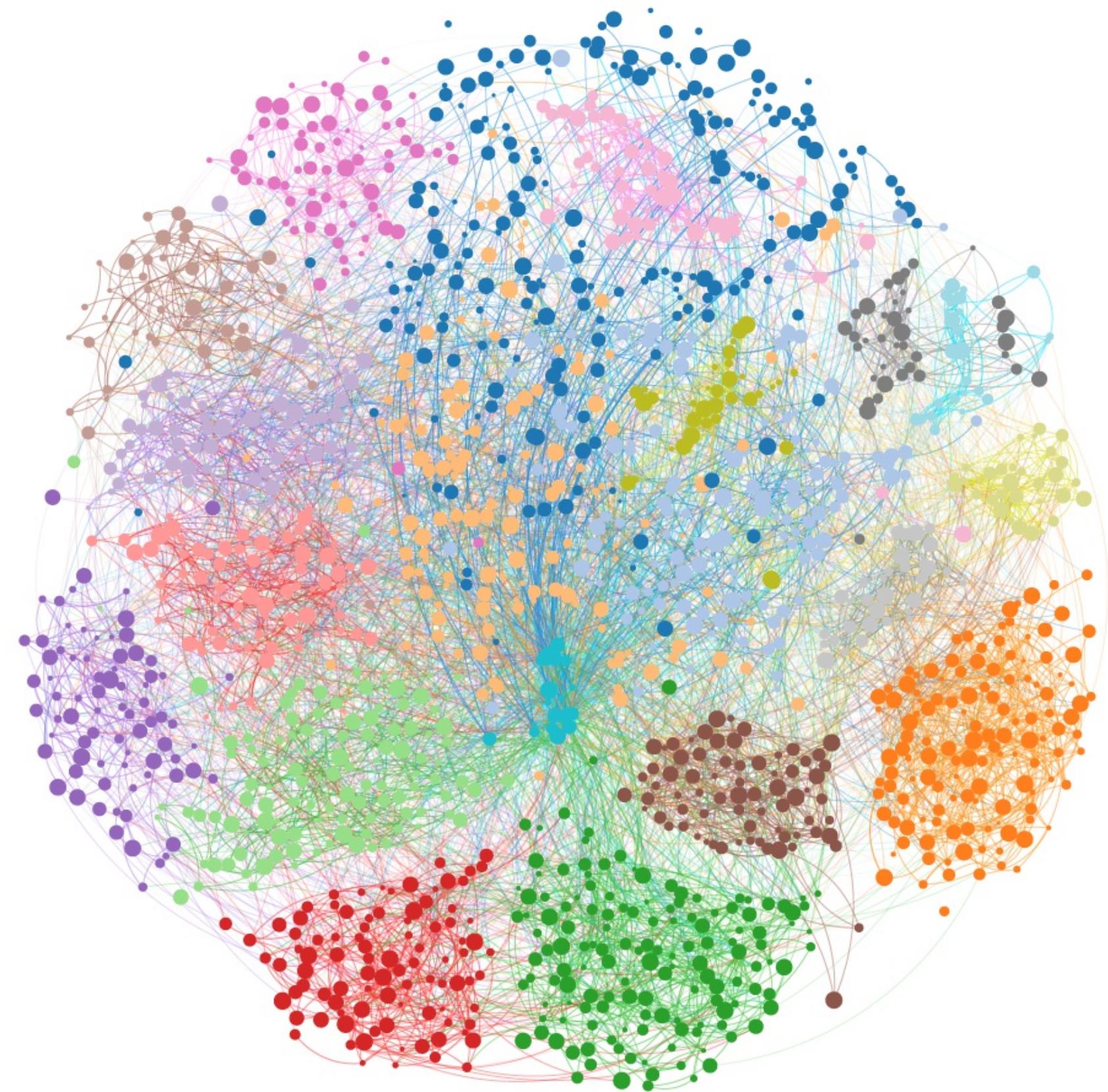
-  **98,000+** tweets
  -  **695,000** status updates
  -  **11million** instant messages
  -  **698,445** Google searches
  -  **168 million+** emails sent
  -  **1,820TB** of data created
  -  **217** new mobile web users

# Le Big Data



# Linked data

**La valeur n'est pas dans la donnée, mais dans ses relations**



# Le processing

complexity = f(size, **connectivité**)



*A quand les ordinateurs quantiques ?*

# Les bases de données

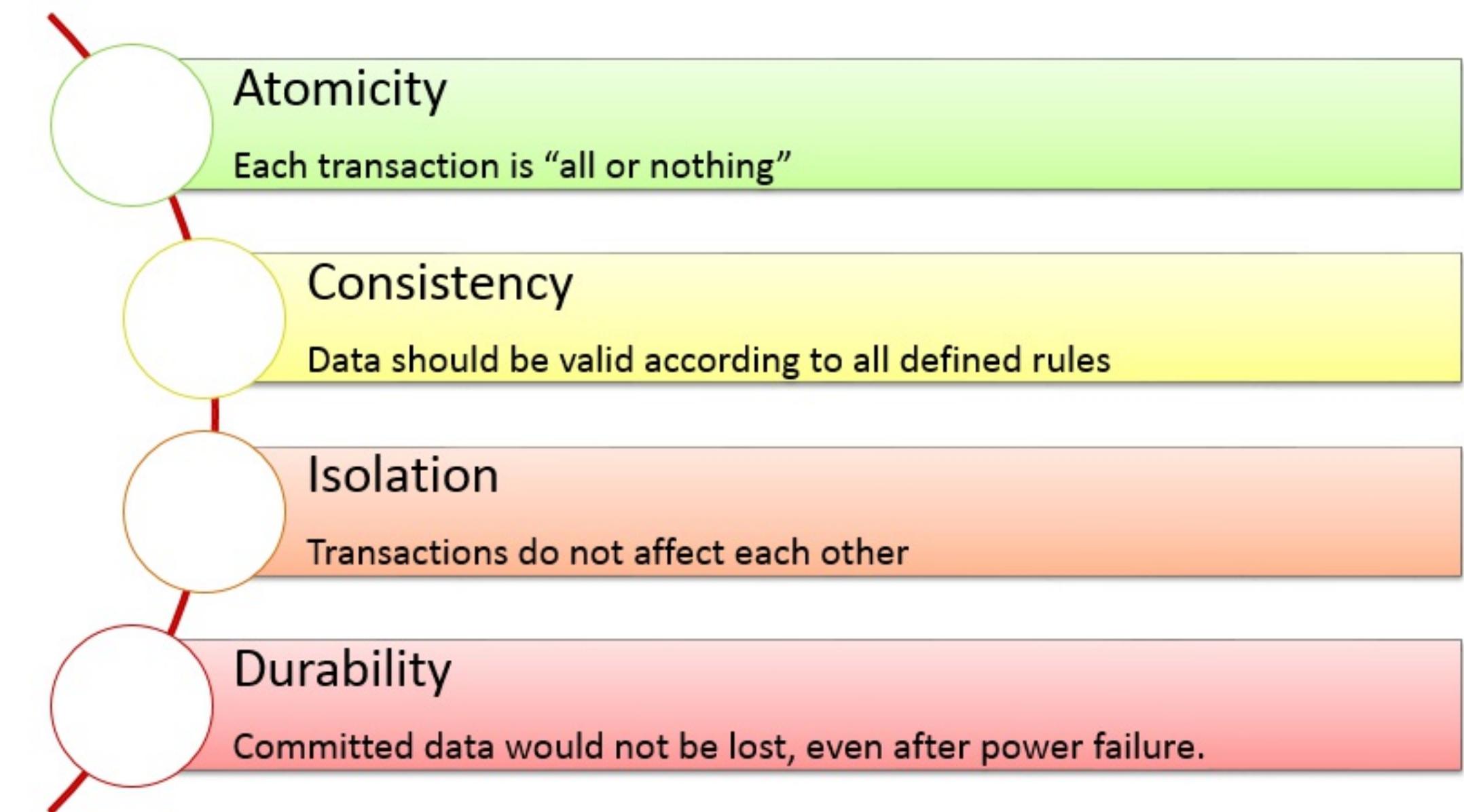
# Les bases relationnelles

- Existent depuis les années 80
- Base ACID
- Gèrent les relations
- Données très structurées
- Peux stocker de grosse volumétrie (plusieurs To)

[http://www.postgresqlfr.org/temoignages:le\\_bon\\_coin](http://www.postgresqlfr.org/temoignages:le_bon_coin)

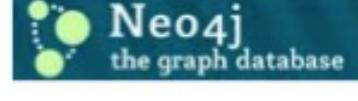
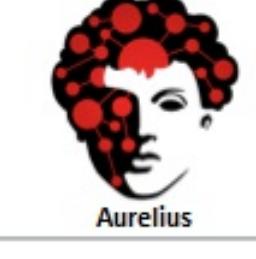
(3To de données - 1To RAM - 160 cores)

# ACID



*Opposé à eventuellement consistant, mais distribué (BASE & CAP théorème)*

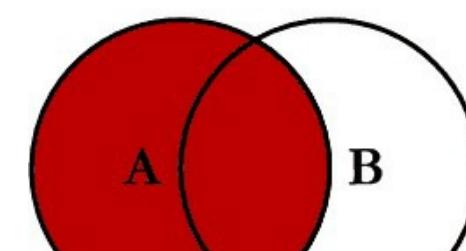
# La mouvance NoSQL

Type	Examples
Key-Value Store	 
Wide Column Store	 
Document Store	 
Graph Store	 

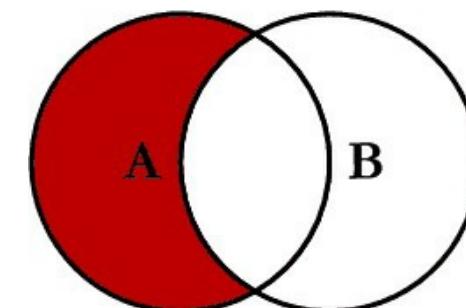
Increasing Data Complexity  
↓

# Le monde ensembliste

## SQL JOINS



```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
```



```
SELECT <select_list>
FROM TableA A
INNER JOIN TableB B
ON A.Key = B.Key
```



```
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
```



```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
WHERE B.Key IS NULL
```



```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
```



```
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
```

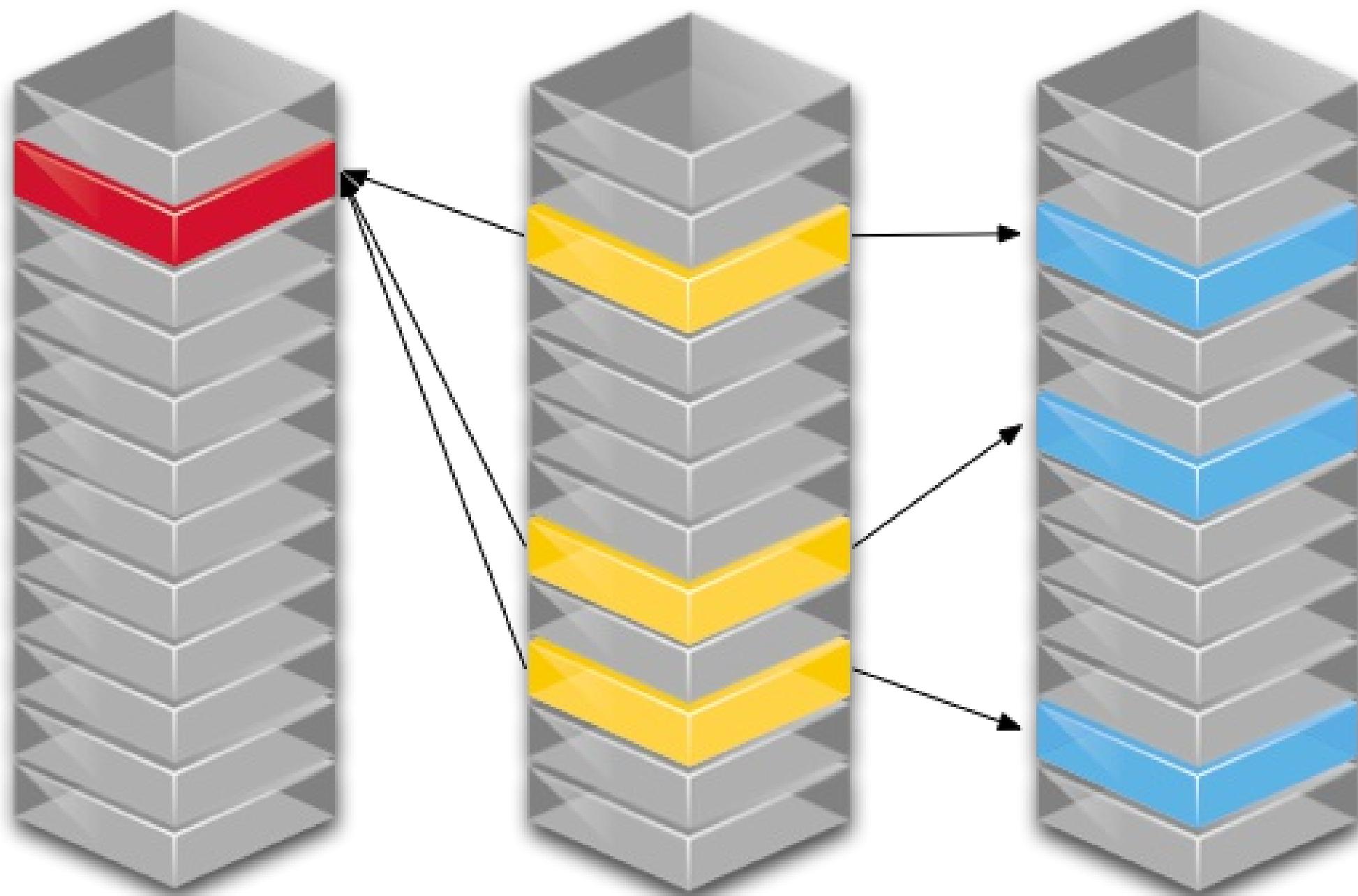


```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
OR B.Key IS NULL
```

© C.L. Moffatt, 2008

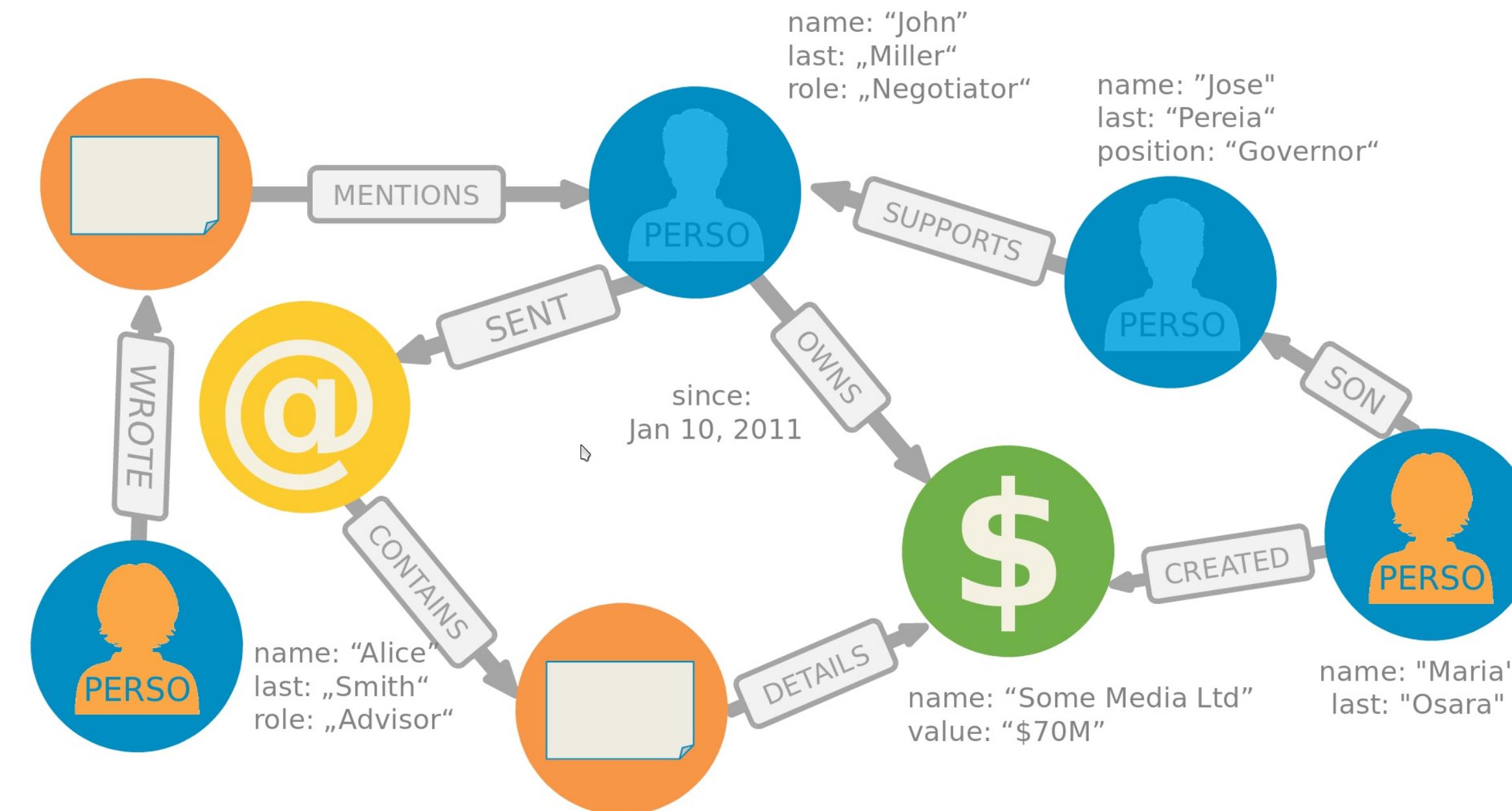
# RDBMS

**Les relations sont calculés ⇒ dans le pire des cas cela revient à faire un produit cartésian des tables**



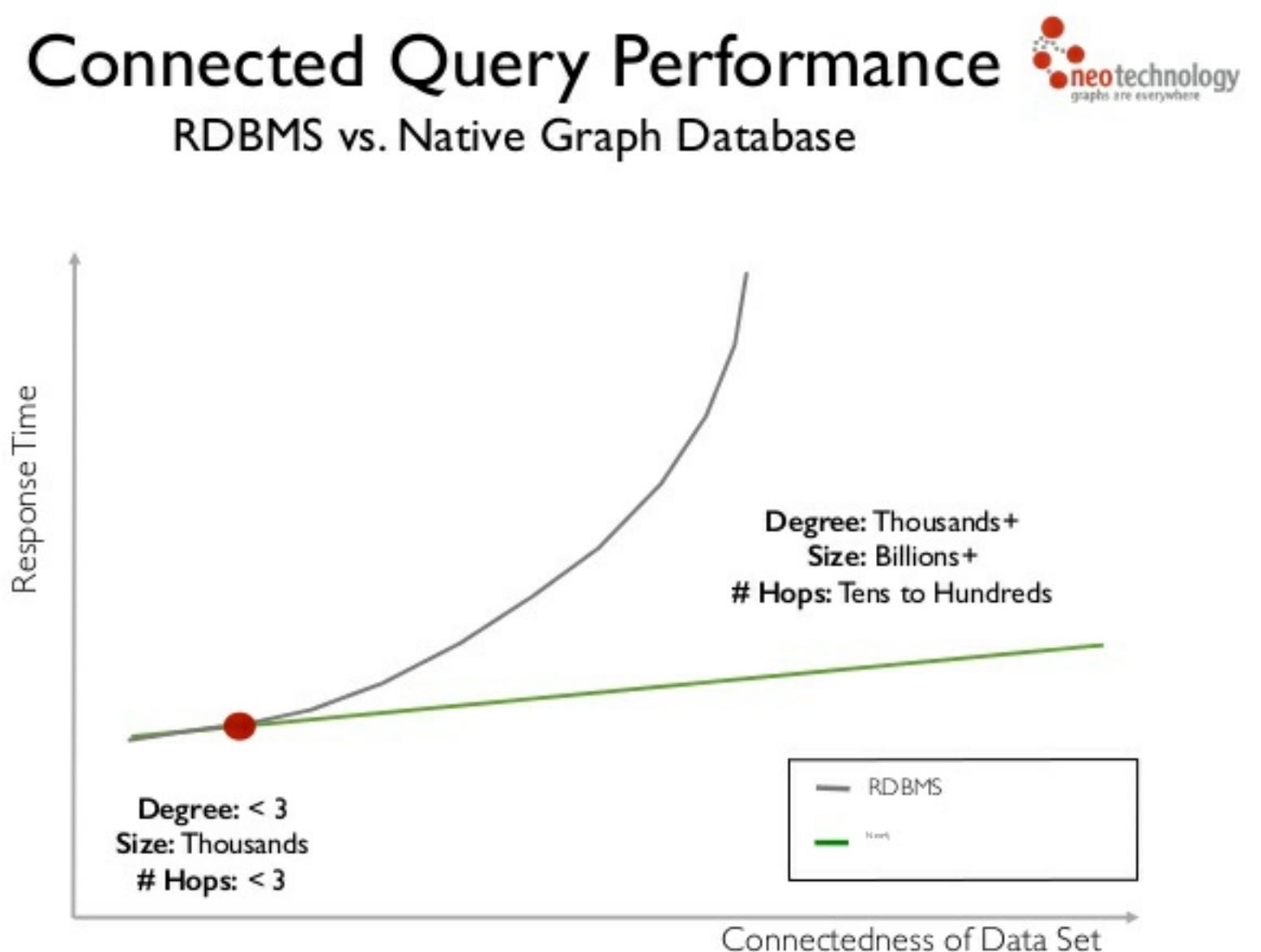
# Les bases graphes

Les relations sont persistés sur le disque, le travail est local



# Performance

- les données/requêtes connectées
- le temps-réel



# **Qui s'en sert ?**

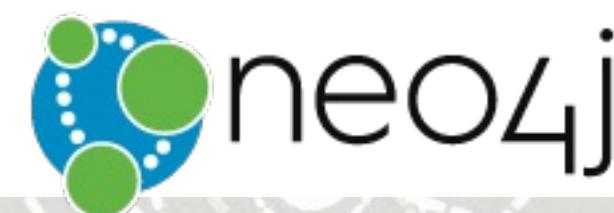


Google est devenu le premier moteur de recherche grâce à son algorithme basé sur les graphes

:

**le page rank**

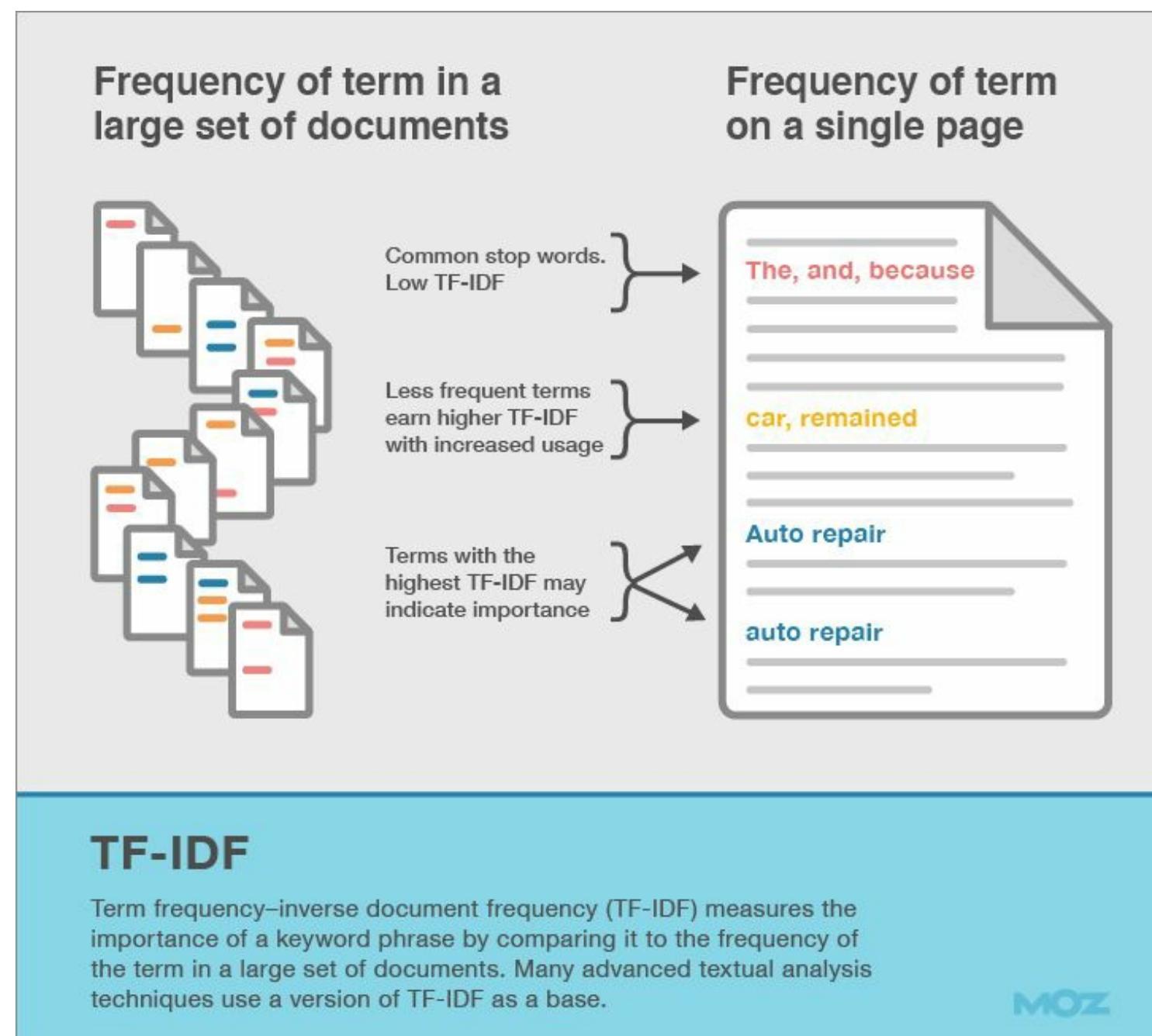
Google™  
YAHOO!®  
LYCOS



# TF / IDF

**Terme frequency** : Nombre de fois que le mot apparaît dans un document

**Inverse Document Frequency** : Nombre de fois que le mot apparaît dans un ensemble de documents

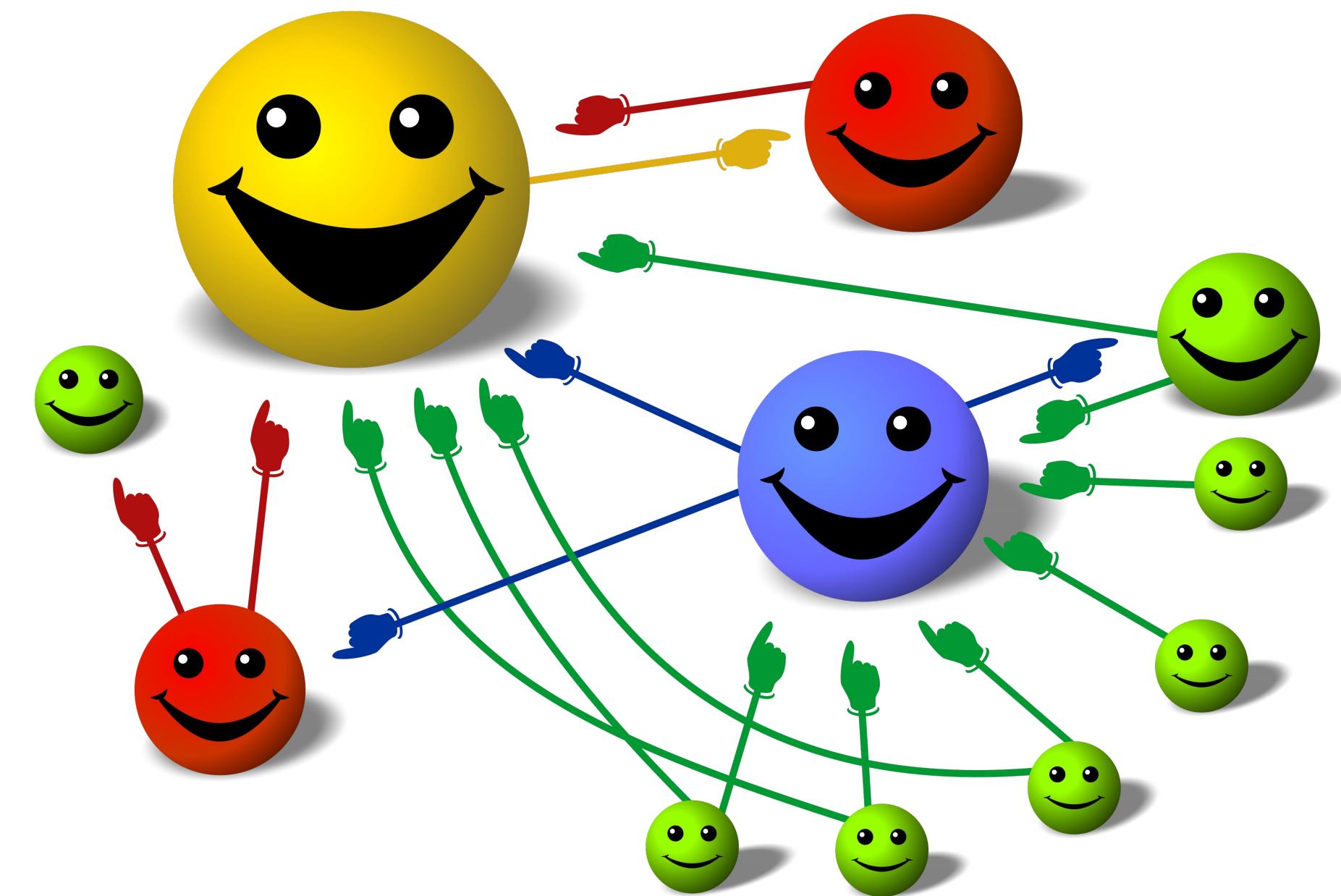


# Invisible touch



# Page Rank

Cartographie thématique du web



# Recommendation

# Moteur de recommandation en temps réel

Des réseaux sociaux, à la vente au détail, de service ou dans le secteur des médias, la recommandation est partout !

**Objectif :** offrir une recommandation en temps réel, hautement ciblés et contextuelles afin quelle soit pertinentes pour l'utilisateur

LES CLIENTS ONT ÉGALEMENT ACHETÉ...



Suggestions · Actualiser · Tout afficher

linkurious @Linkurious

Philip Rathle @prathle

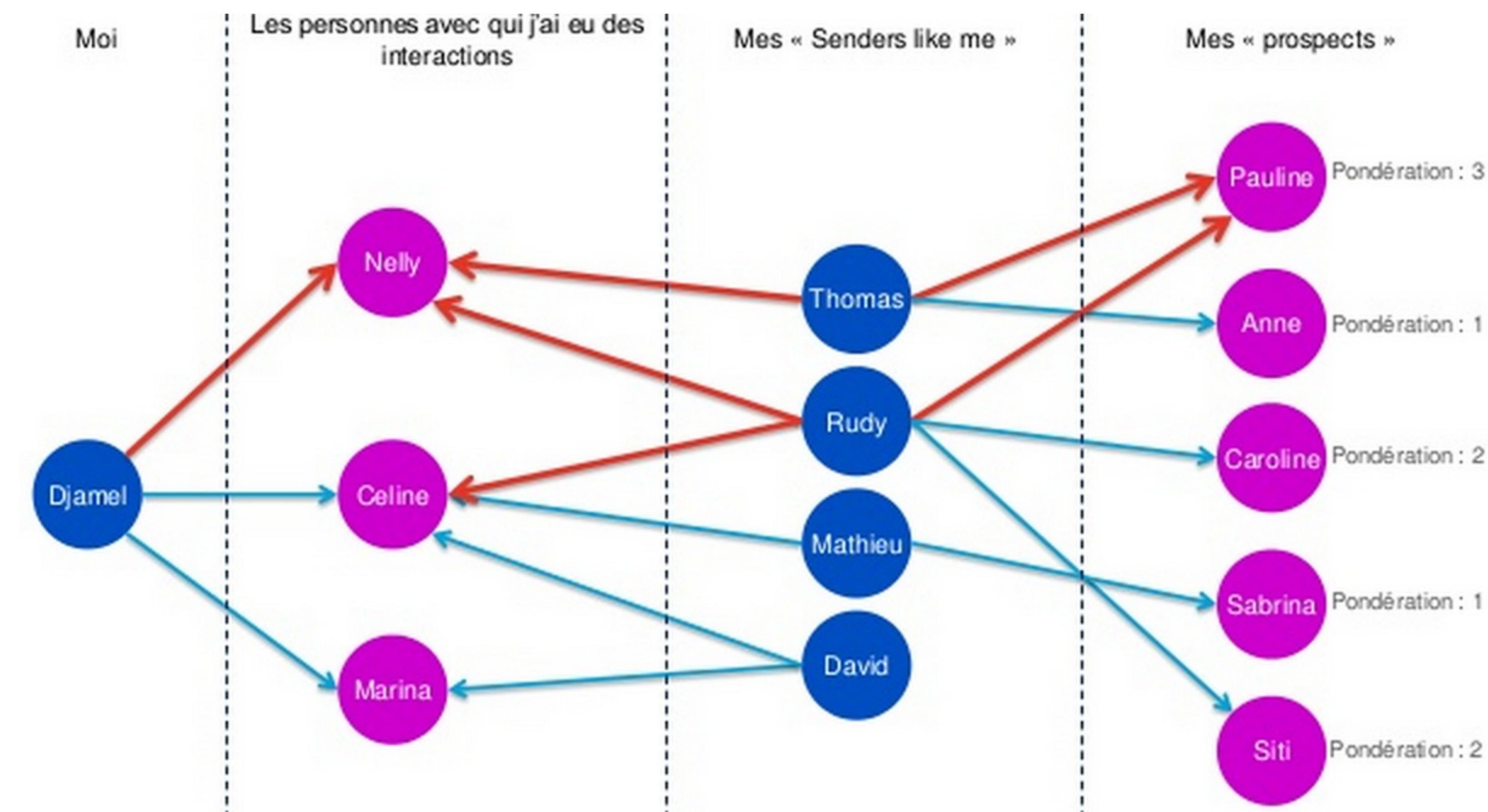
Trouver des amis

# Filtrage collaboratif

Le système est simple : trouver des personnes qui me ressemblent, et voir ce qu'elles ont fait et que je n'ai pas fait

**Les amis de mes amis avec lesquels je ne suis pas ami**

# Meetic



# La détection de fraude

# Les panama papers

C'est la plus large fuite de données de l'histoire, et de loin !

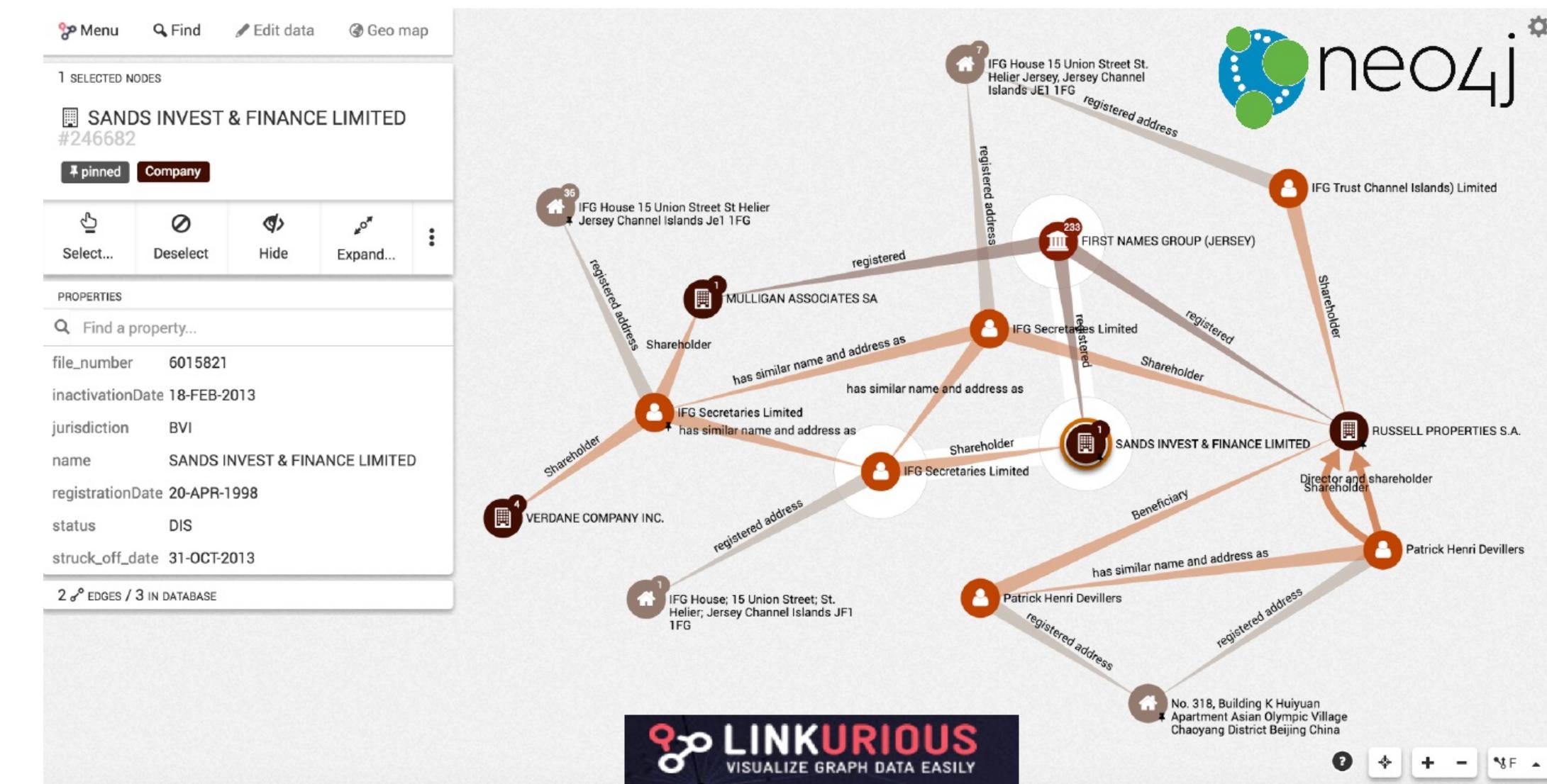


# L'opacité

Ajouter un maximum d'intermédiaire pour rendre le plus opaque possible (holding + offshore).



# C'est du graph !



*Il faut suivre les petits cailloux*

# Le routing

# Ebay Now !

La livraison en moins de deux heures !



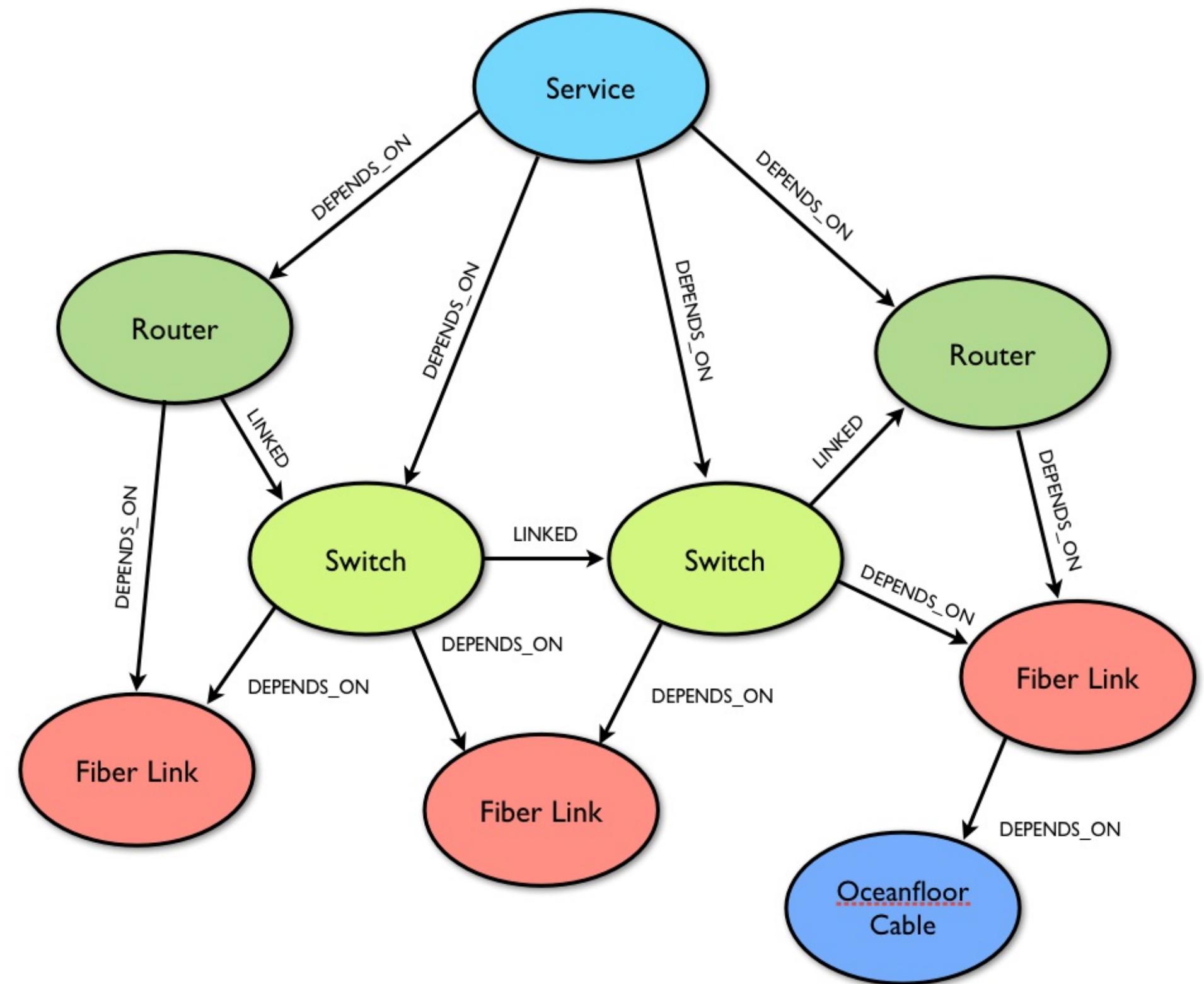
<https://vimeo.com/110675062>

*Même principe qu'Uber pool*

# L'analyse des réseaux

## Etudes d'impacts

- Pouvoir planifier une maintenance rapidement
- Identifier les clients impactés par une maintenance ou une panne.
- Identifier les faiblesses du réseau



# **Graphs are everywhere !**

# Les cas d'utilisation

## How Customers Use Neo4j



Social	Recom-mendations	Search & Discovery	Network & Data Center	Master Data Management	Identity & Access	GEO	
eHarmony meetic Hinge DOWN maail ZEPHYR HEALTH INC mallowstreet Care.com LIFECHURCH.TV doximity	careerbuilder classmates viaded NATIONAL GEOGRAPHIC CloudB zonelinestay gamesys ice bwin.party Uvestation	Walmart careerbuilder InfoJobs kooboodle LIQUID COMMON acuraspan™ NOMURA research now! Compete Cobrain shutl knedesk	IEEE First Data janssen e-Spirit codex TRIBAL technologies Lufthansa Didacti scribeaster springcm noble group	hp SFR NetApp Zenoss gen Lufthansa virtual instruments EarthLink	cisco Pitney Bowes ADVENT die Bayerische TechCrunch veda ZEPHYR HEALTH INC Juice PLUS+ EQUILAR	UBS telenor LifeWay i TechCrunch veda ZEPHYR HEALTH INC Juice PLUS+ EQUILAR	TOMTOM ebay now classmates- KiwiRail Dinglicom Perigee™ DOWN CHRONOTRACK YELAGO



# un est tout, tout est un

Le monde est connecté



# Merci



# Des questions ?

