

Bases de données graphes : comparaison de

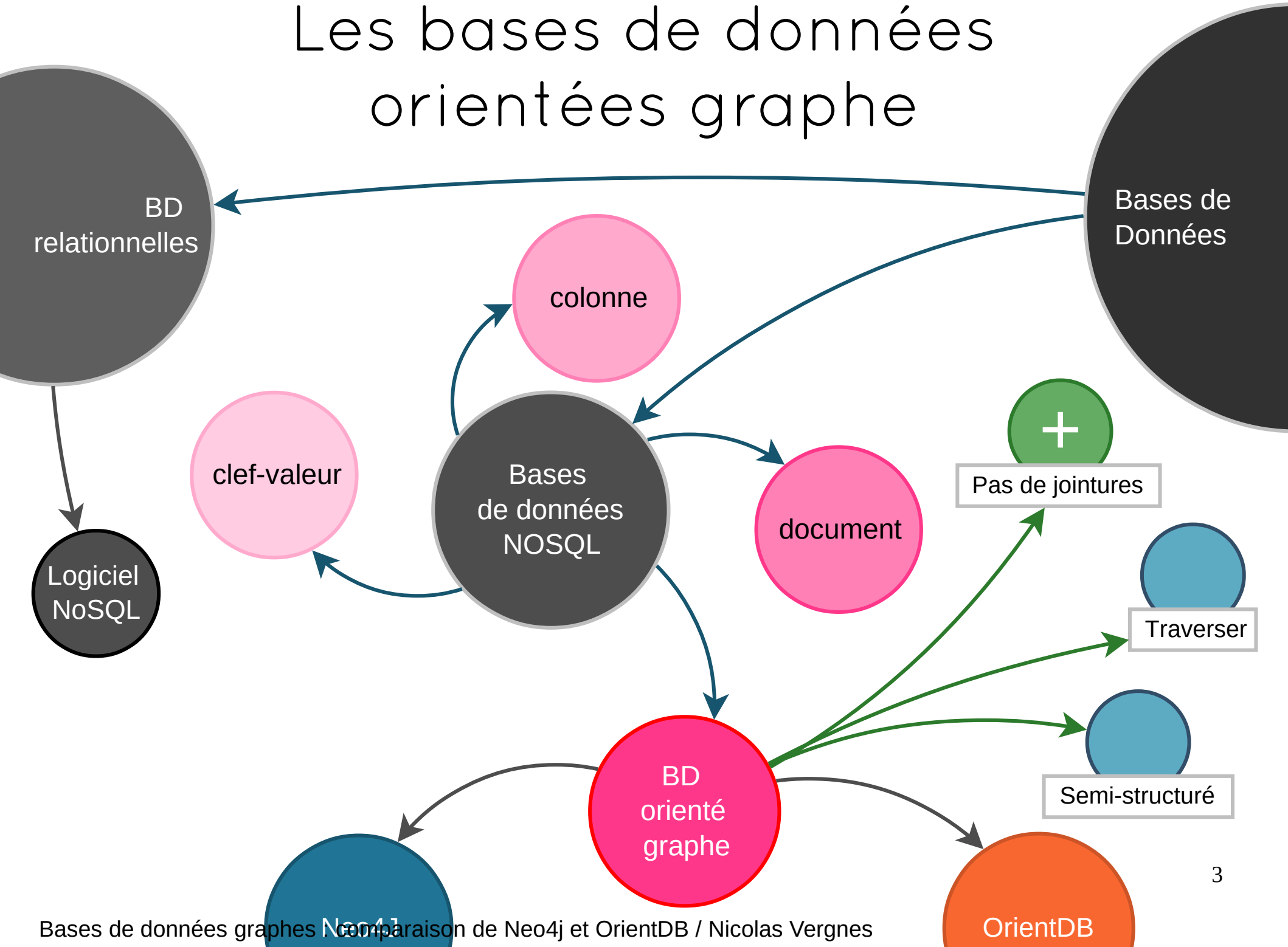


présenté par Nicolas Vergnes
le 13/05/2015

Sommaire

- Généralités
- Modèles de données (1..4)
- Fonctionnalités
- Utilisation en production
- Performances
 - Cas pratique (présentation, données, architecture, scénarios (1..4), synthèse)
- Conclusion

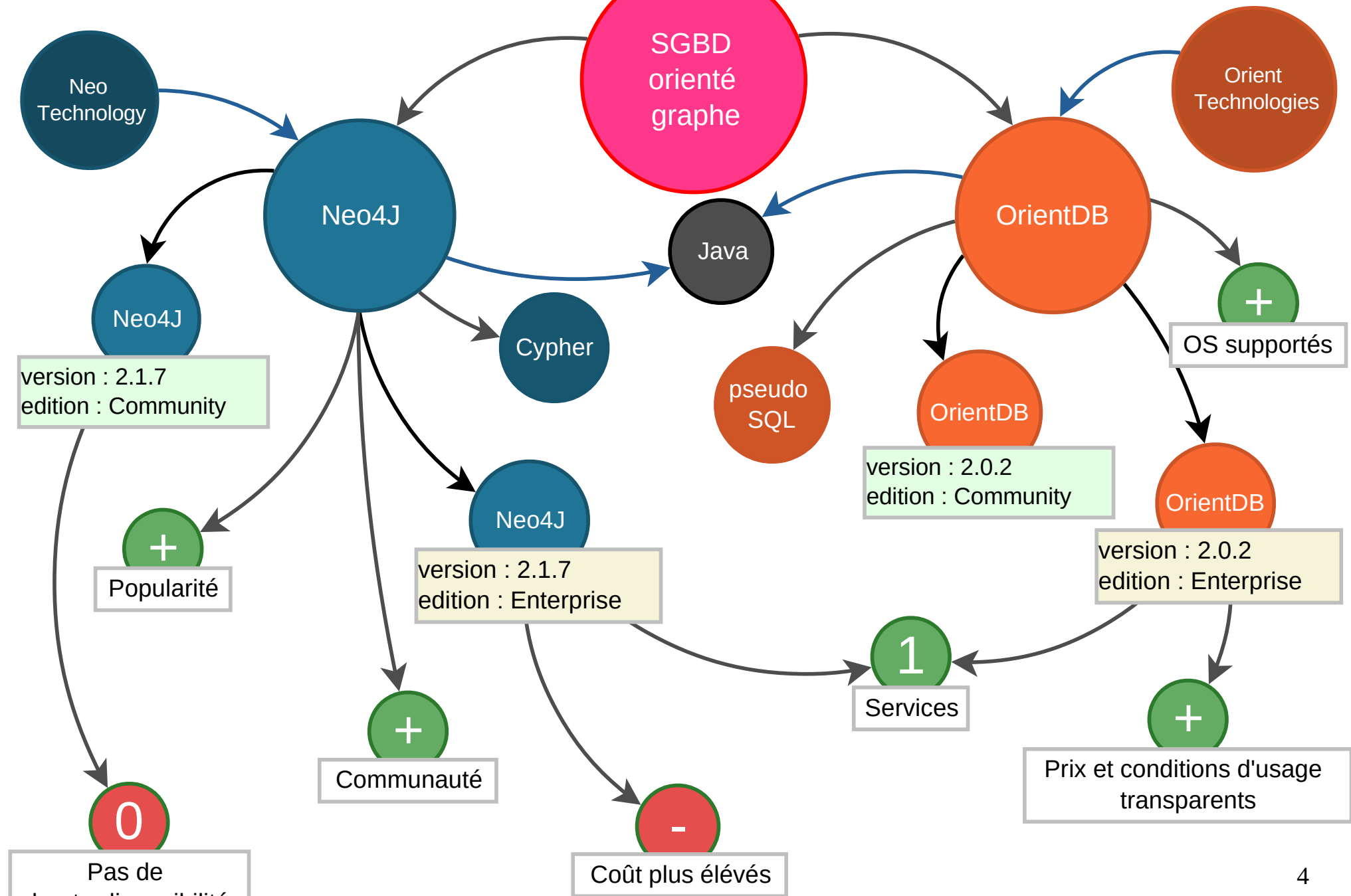
Les bases de données orientées graphe



Neo4J

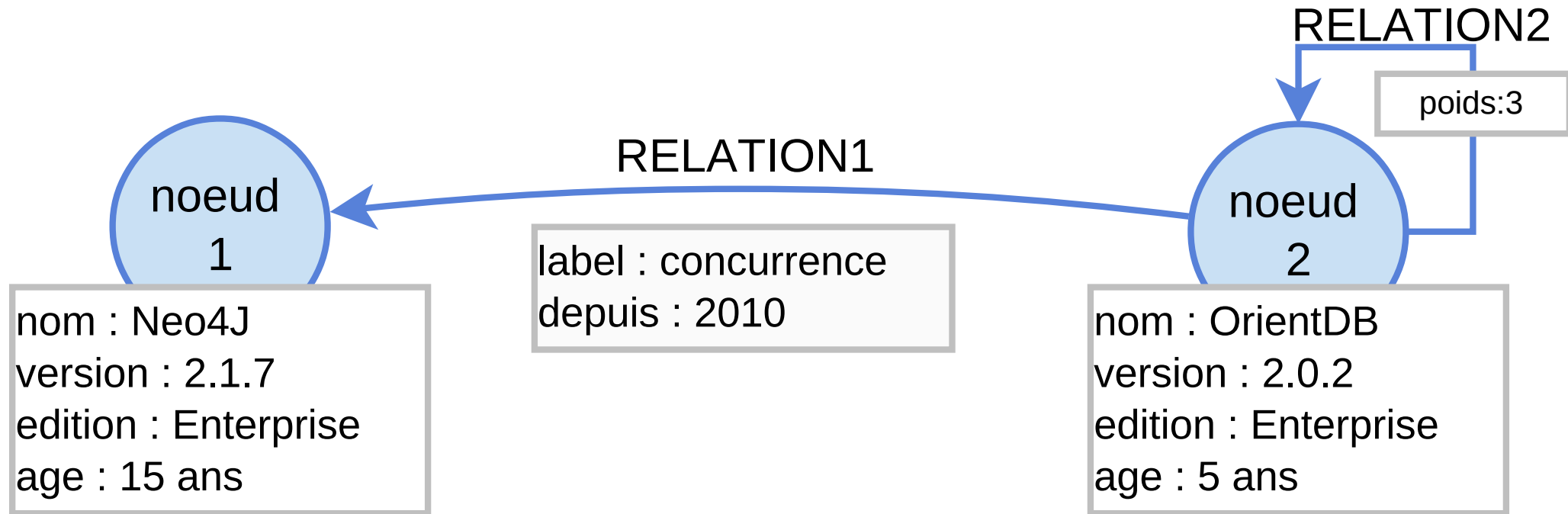
et

OrientDB



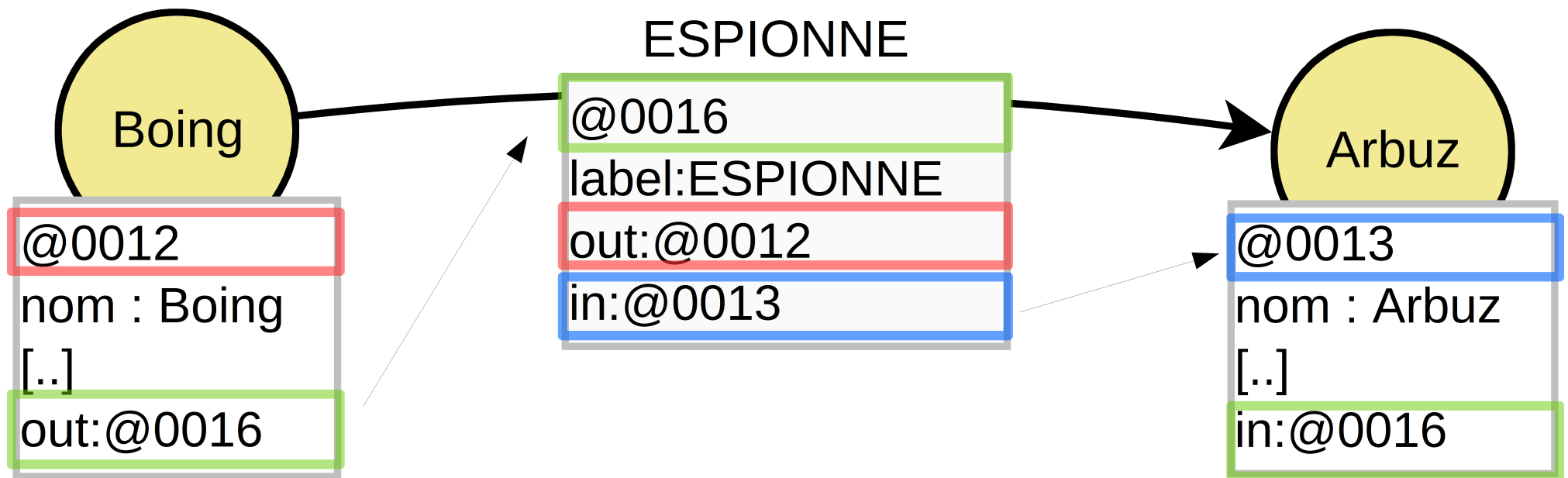
Modèles de données 1/4

- Les graphes de propriétés



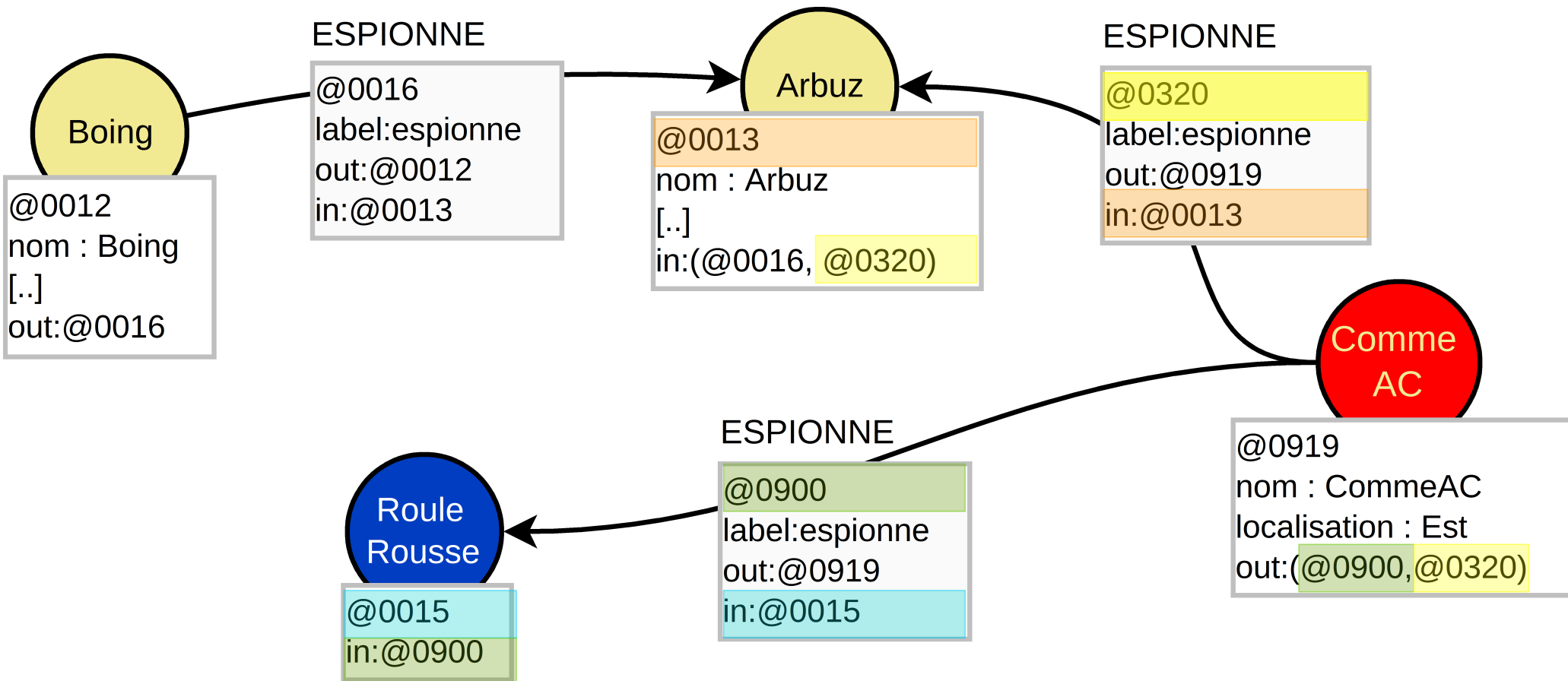
Modèles de données 2/4

- Modèle de données détaillé d'OrientDB
- Relation **1->1**



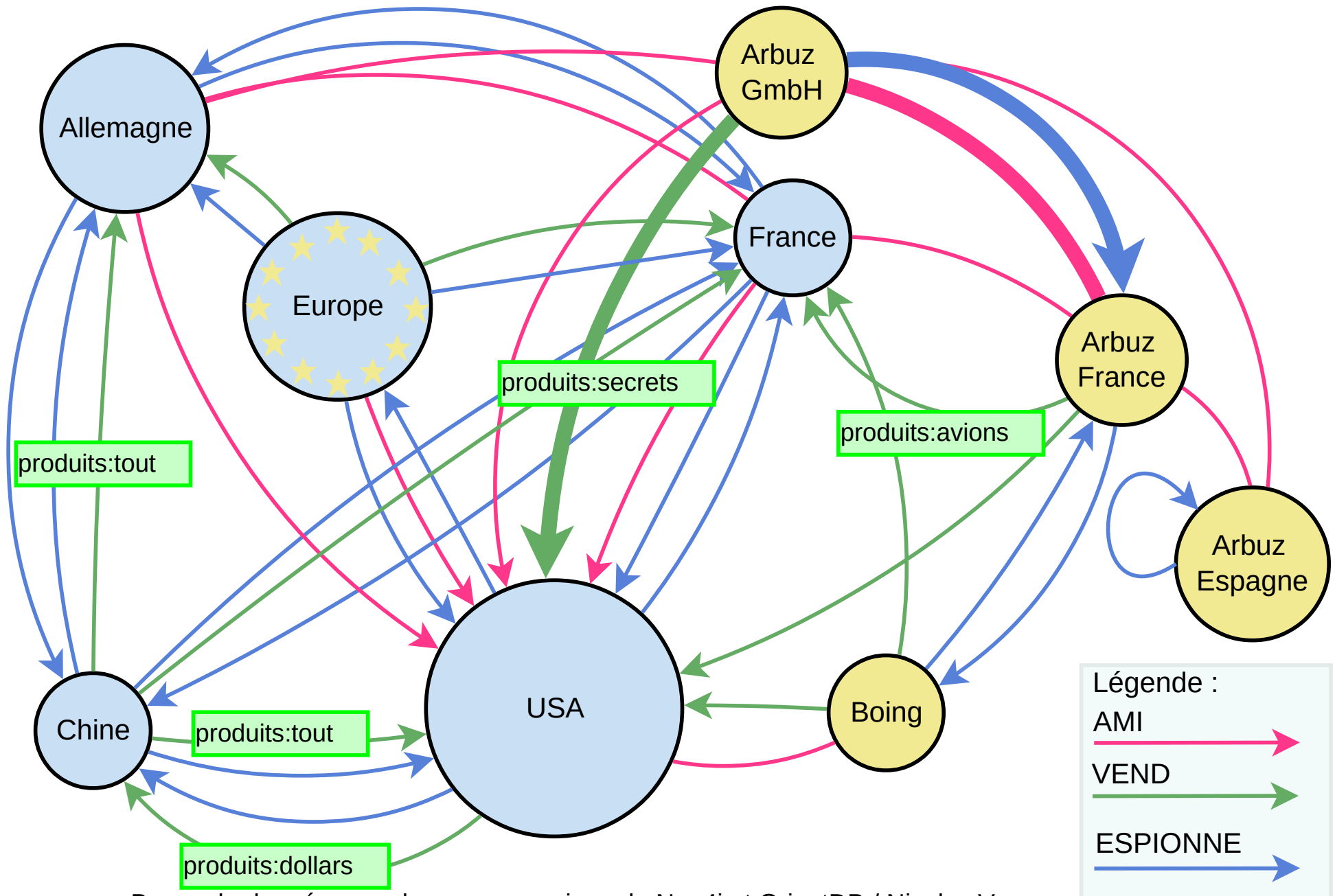
Modèles de données 3/4

- Modèle de données détaillé d'OrientDB
- Relations **1->N**

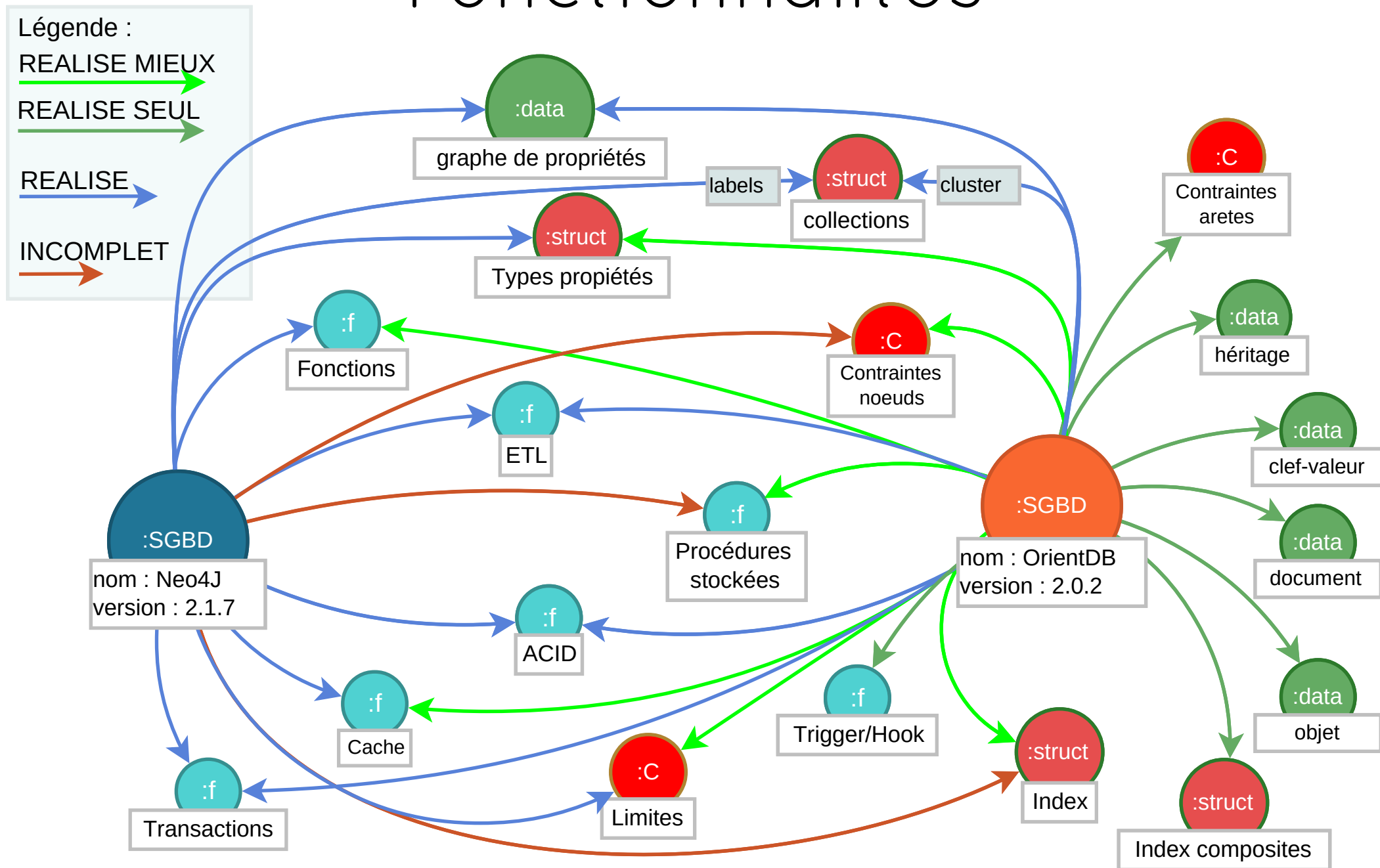


Modèles de données 4/4

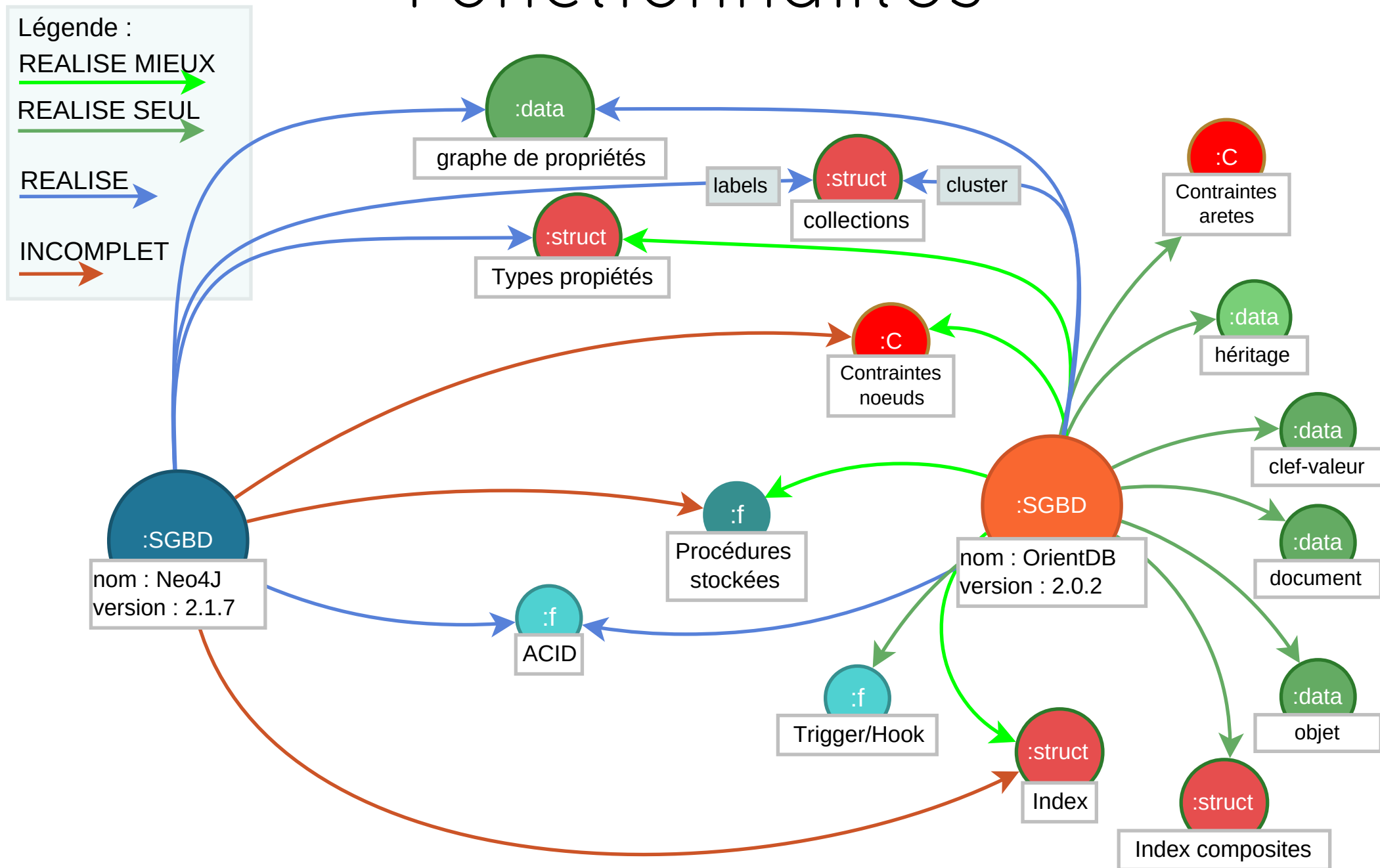
- Relations **N-→M** et traversées



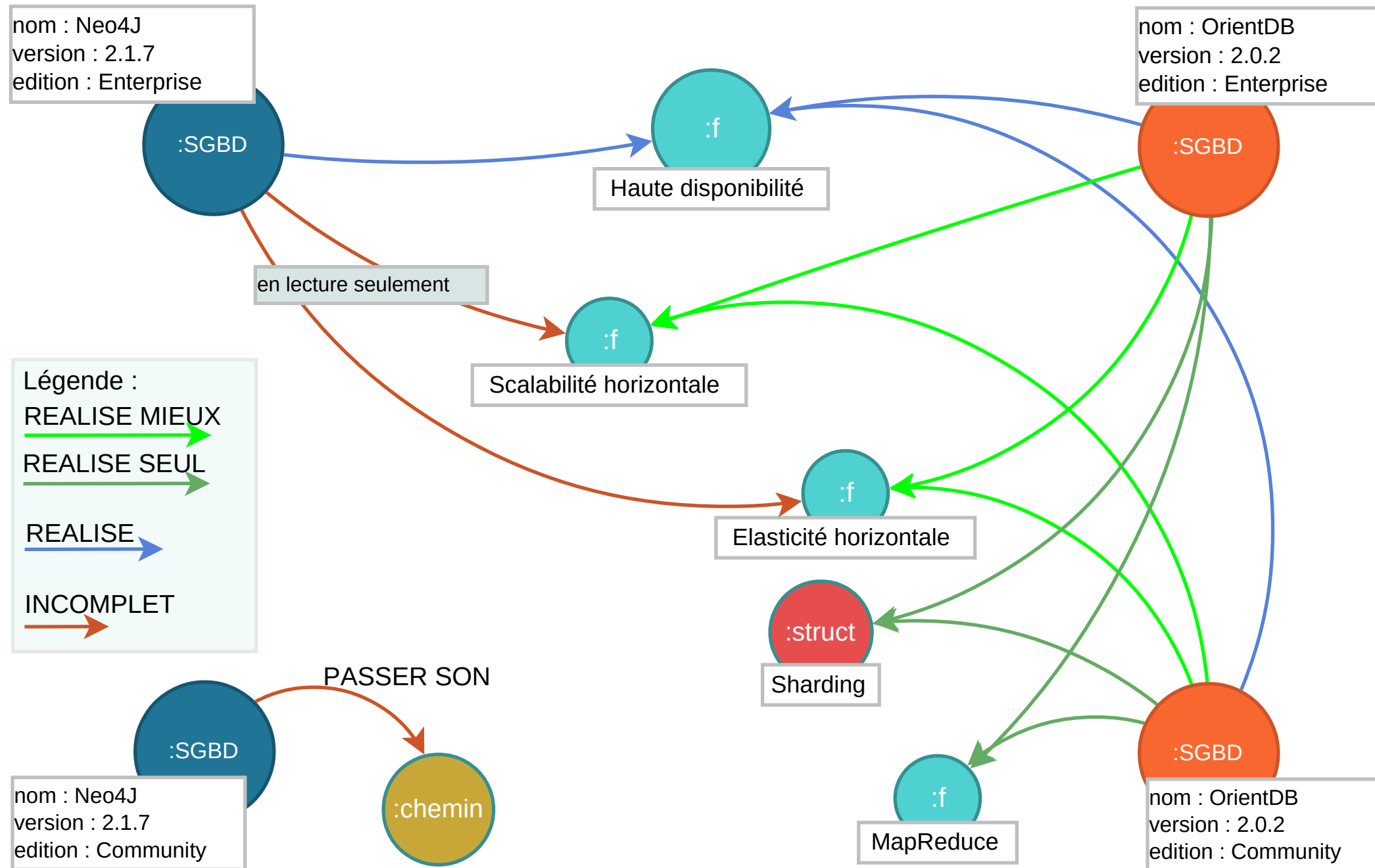
Fonctionnalités



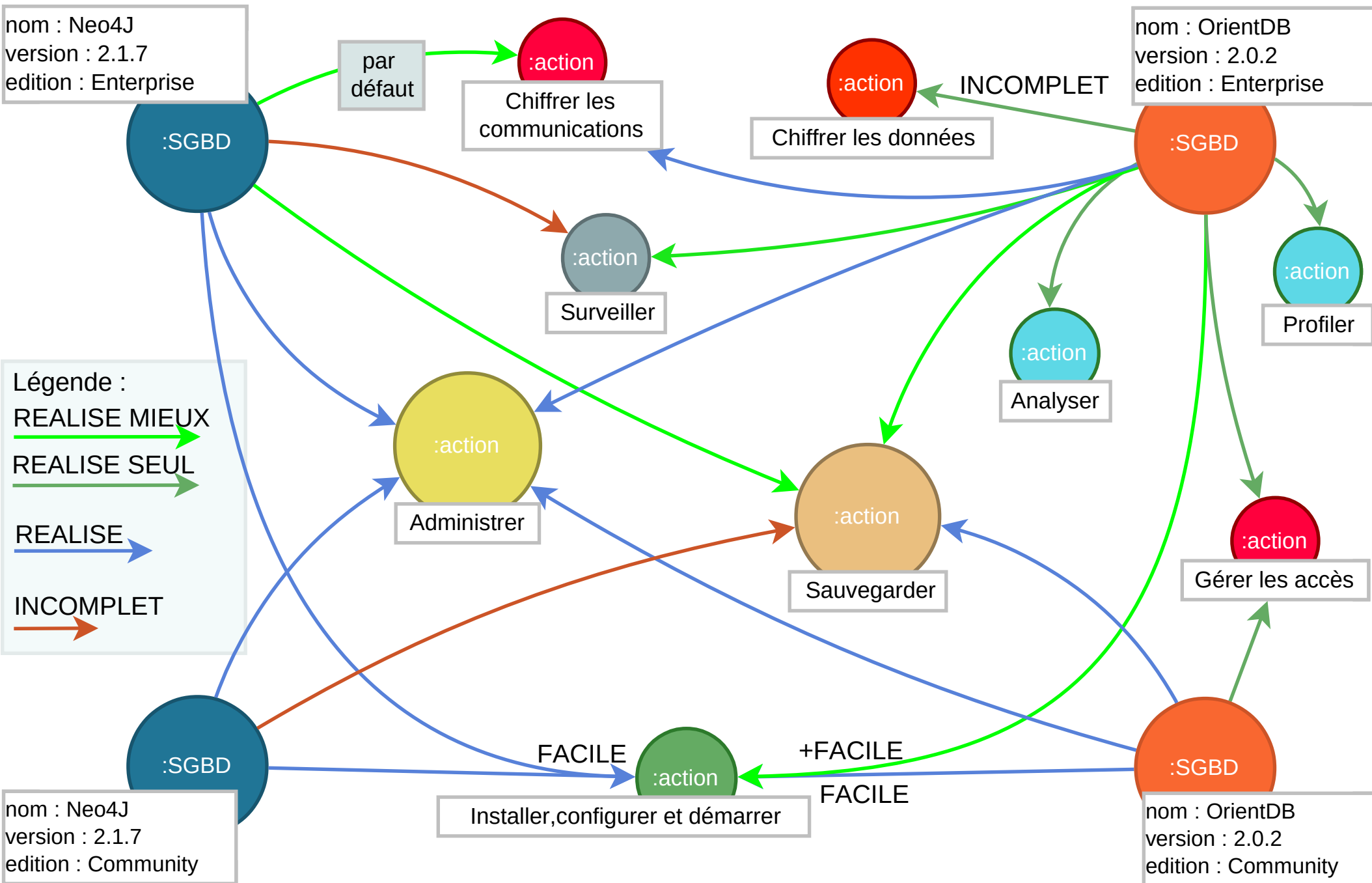
Fonctionnalités



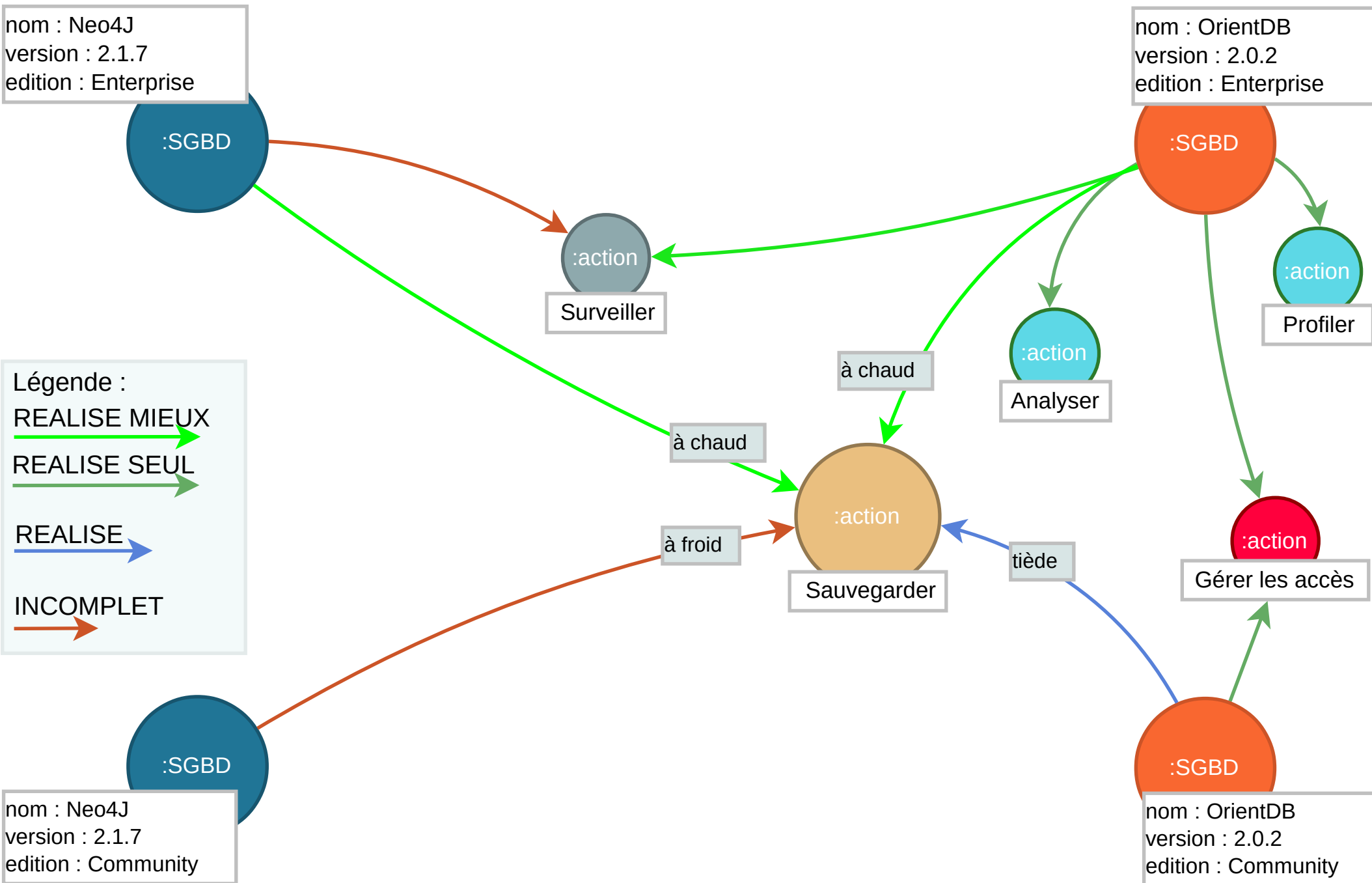
Architectures distribuées



Utilisation en production



Utilisation en production

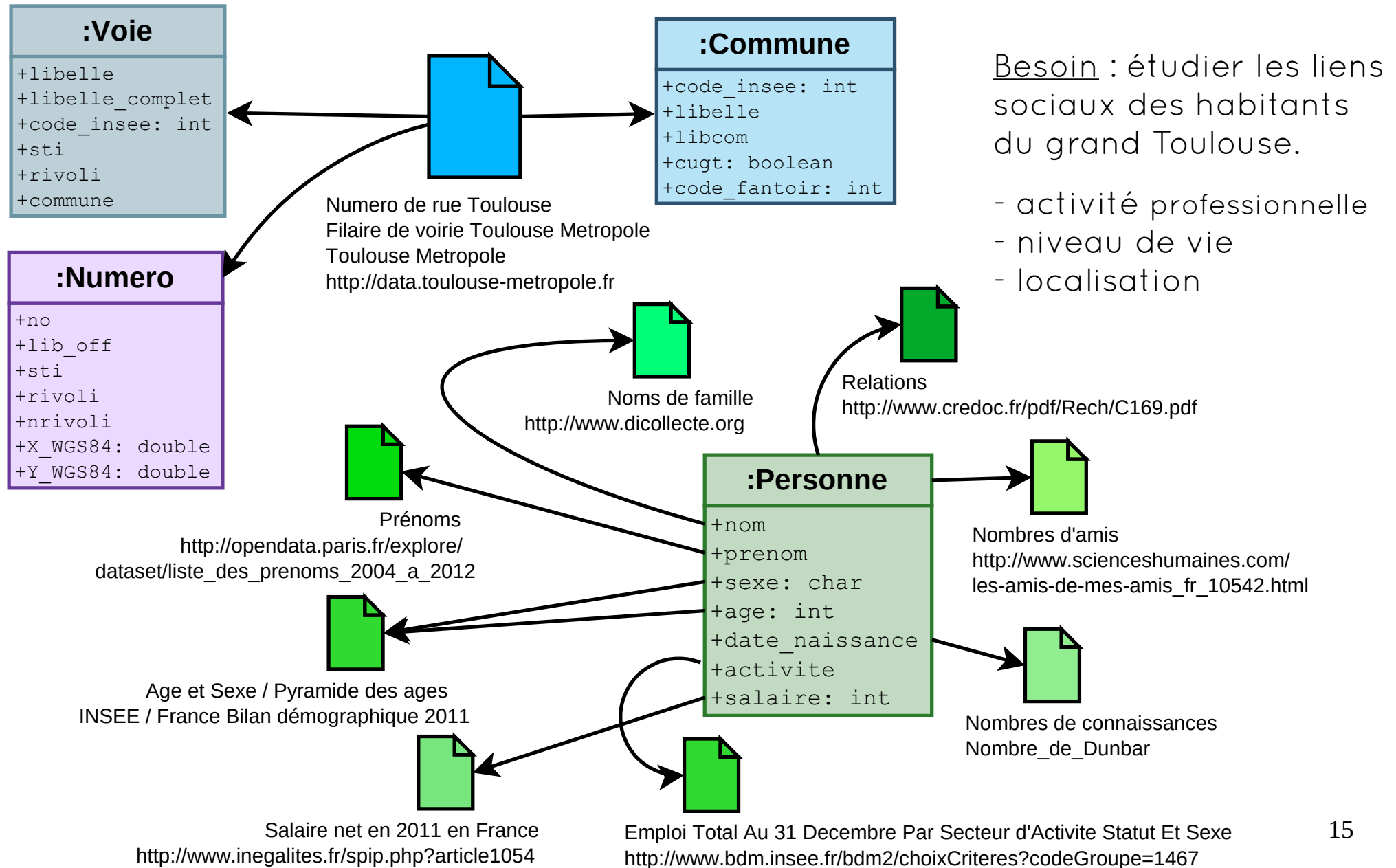


Performances

- Difficulté de trouver des tests utiles.
- L'enjeu : « De combien et de quelles ressources physiques et logicielles avons-nous besoin pour répondre au besoin métier présent et futur » .
- Le comportement d'un SGBD varie suivant sa version et son environnement.
- Se rapprocher le plus possible de ce que serait l'environnement de production.
- Faire attention en extrapolant les résultats des tests : comportement non-linéaire.

Cas pratique *présentation

- Urbanisation, population et liens sociaux

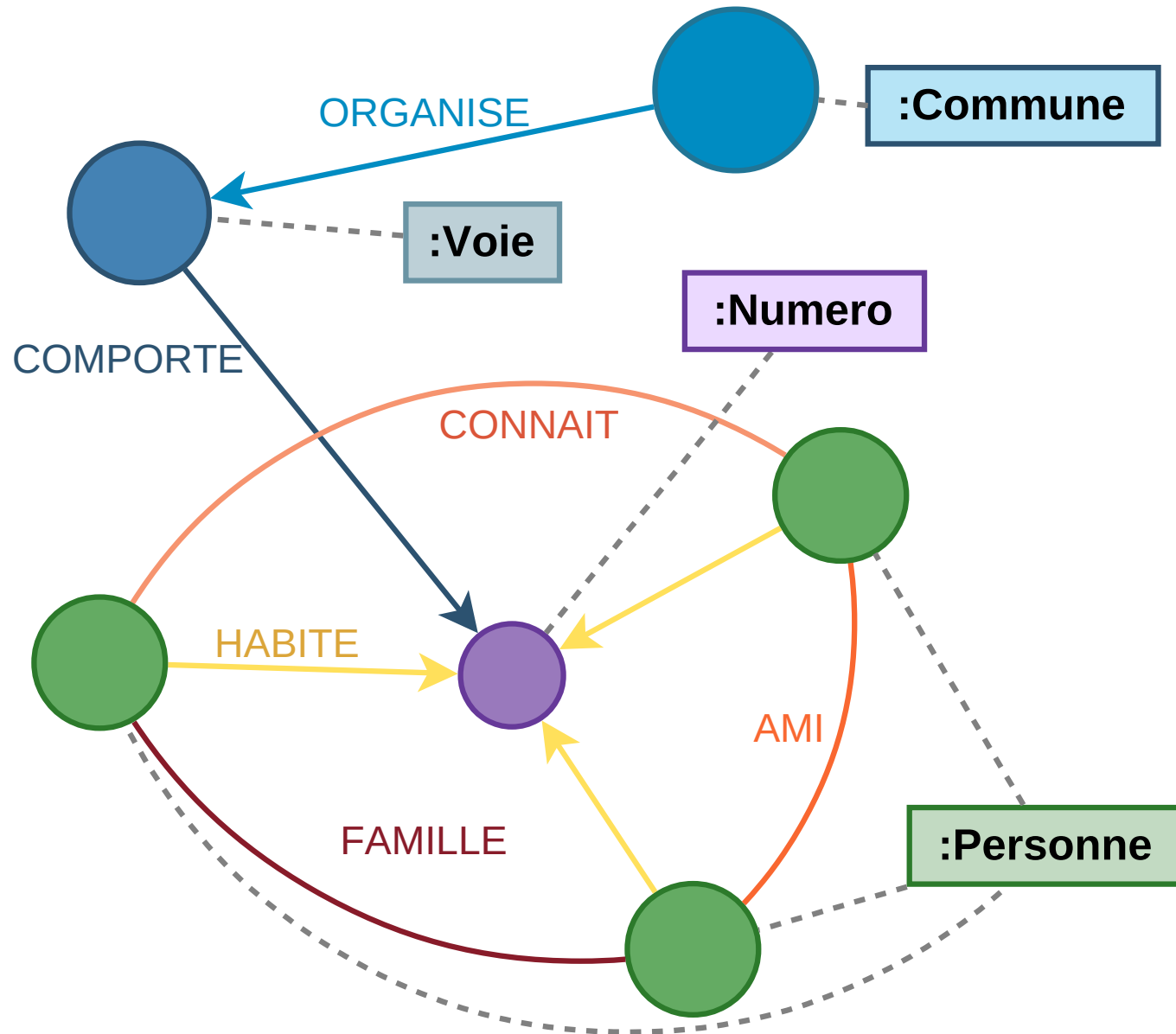


Besoin : étudier les liens sociaux des habitants du grand Toulouse.

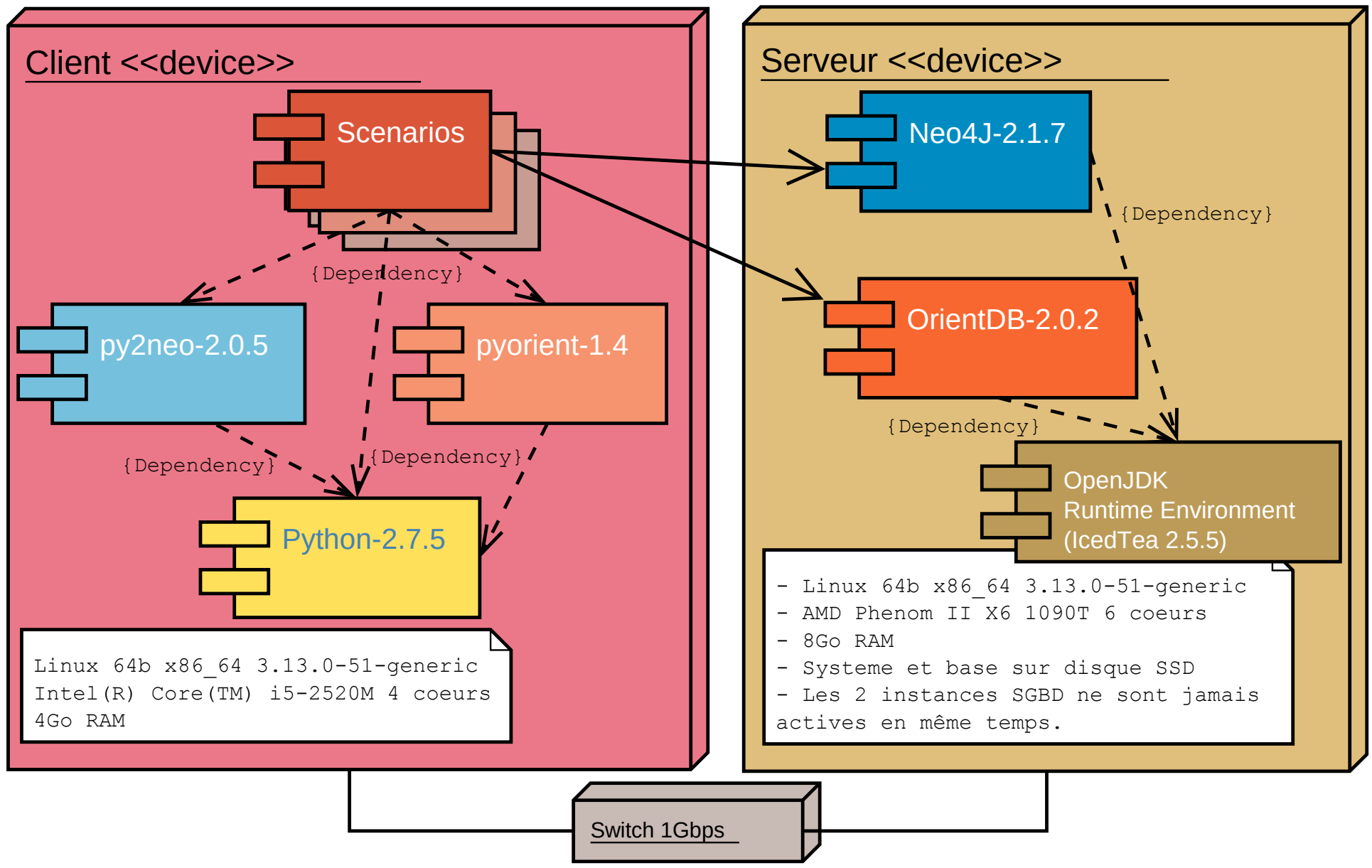
- activité professionnelle
- niveau de vie
- localisation

Cas pratique *présentation

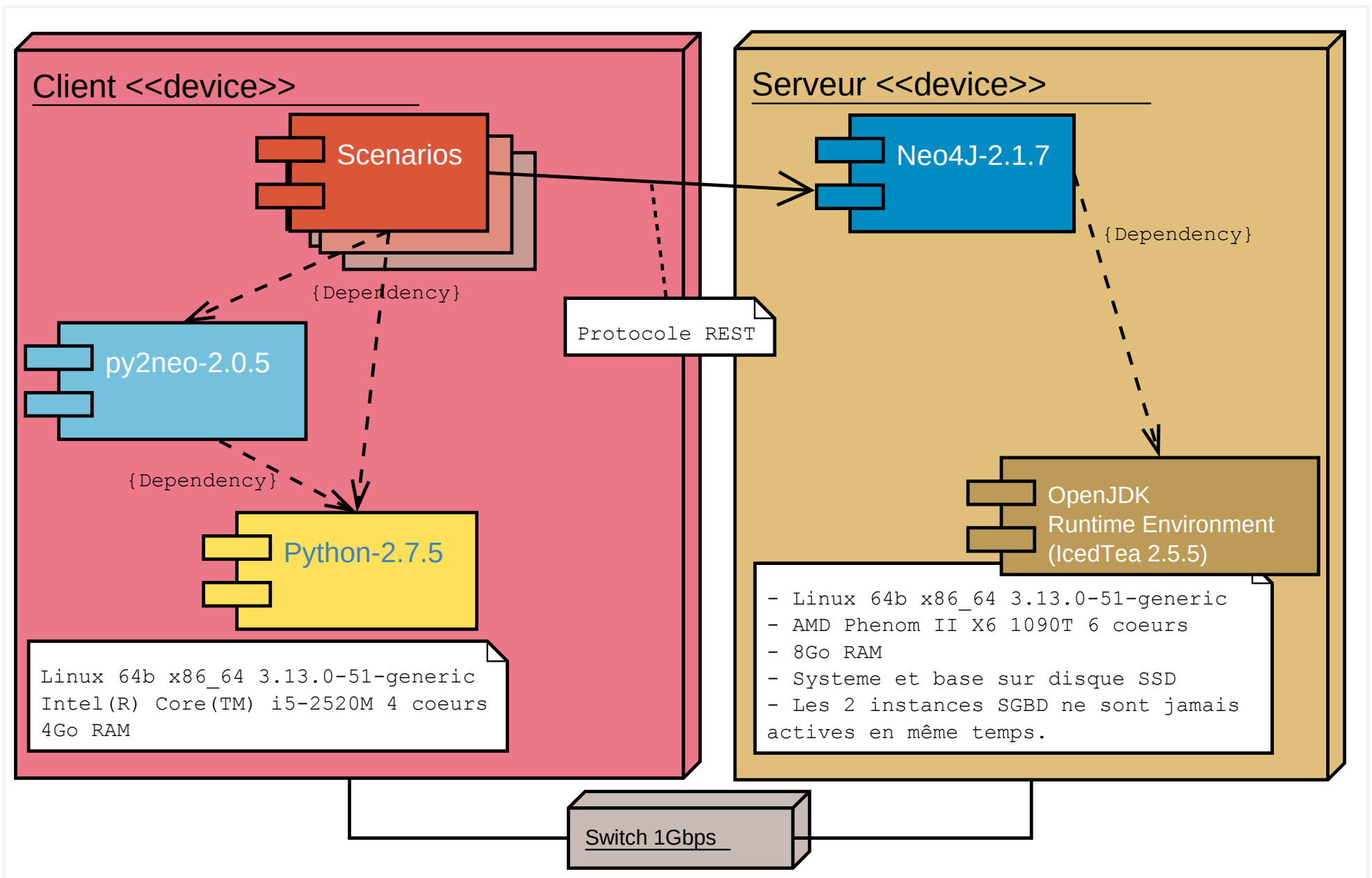
- Modèle de données



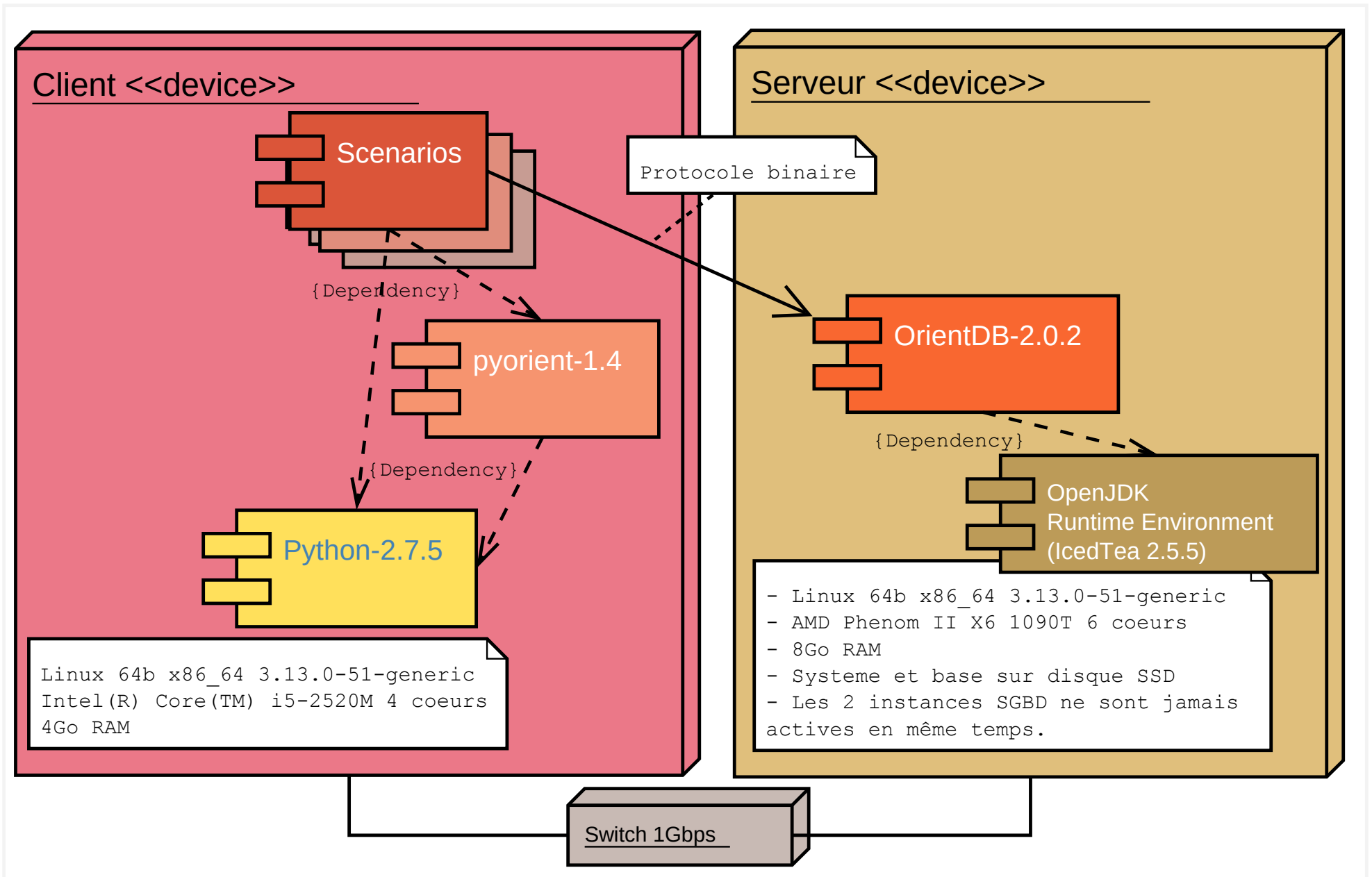
Cas pratique *architecture



Cas pratique *architecture



Cas pratique *architecture

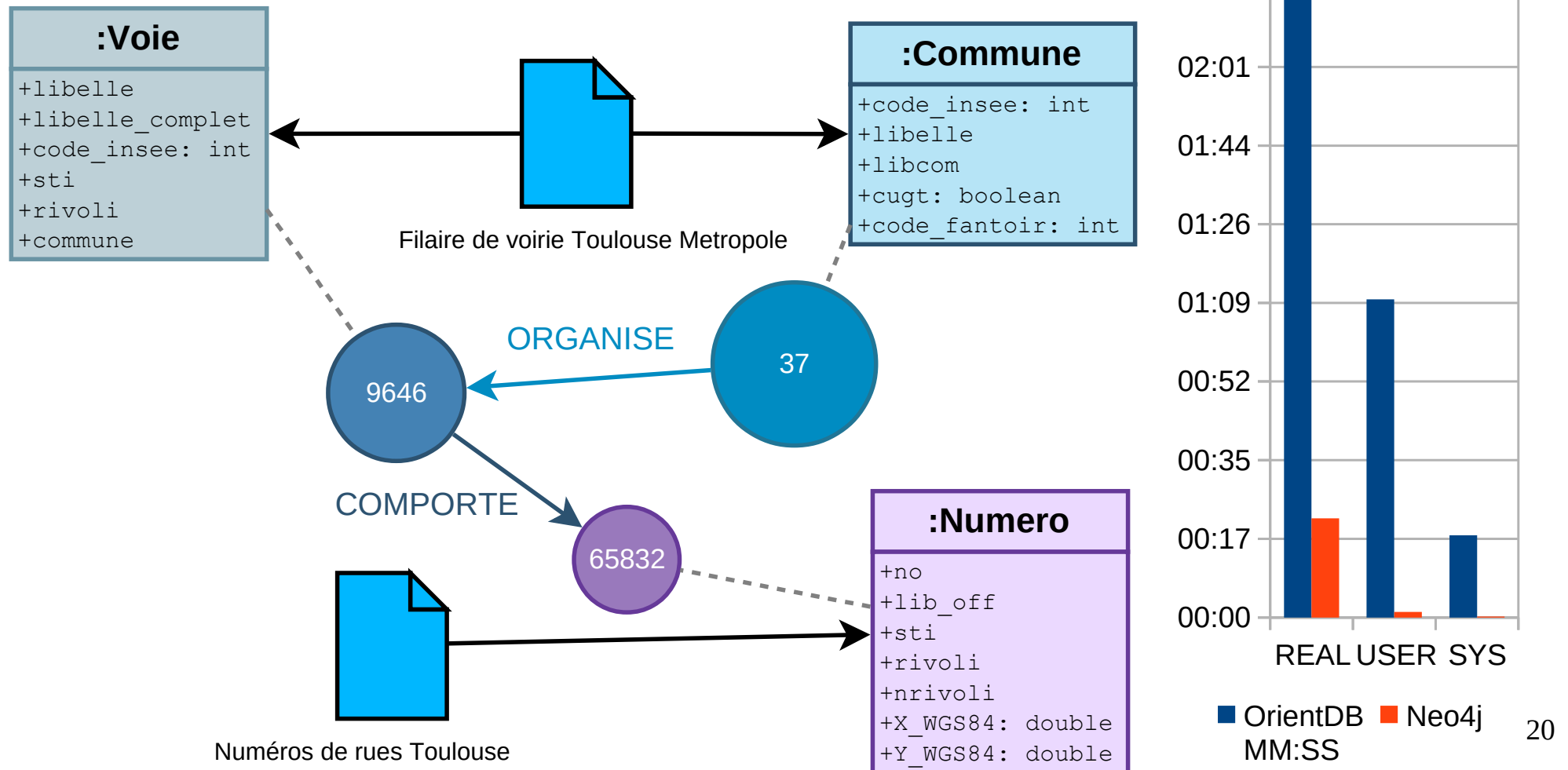


Cas pratique *scénario1

Chargement ETL des 3 fichiers CSV en local.

► Création de 75 515 nœuds et
75 478 relations.

(lecture-écriture 0-100)

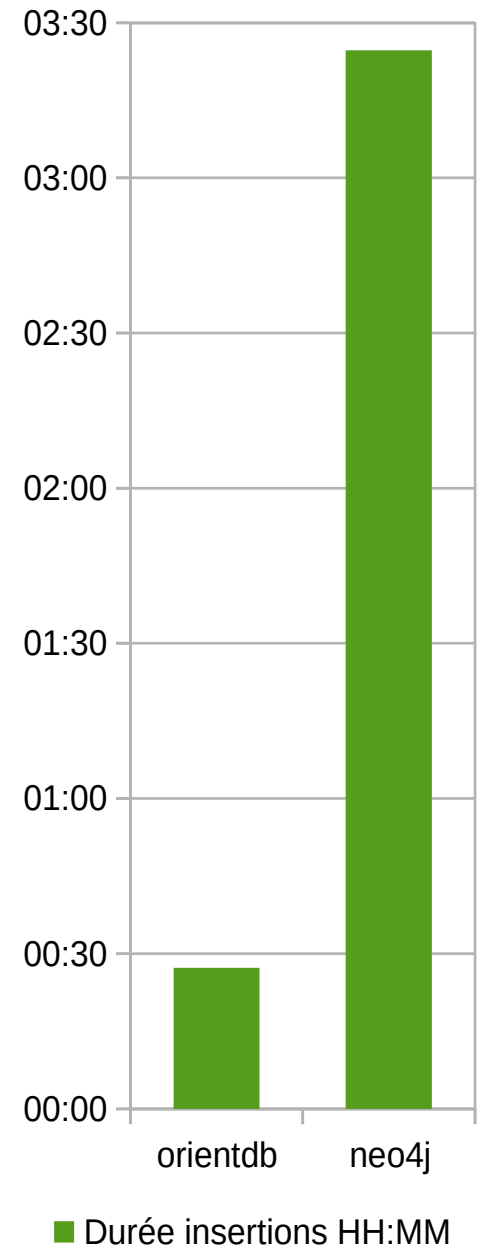
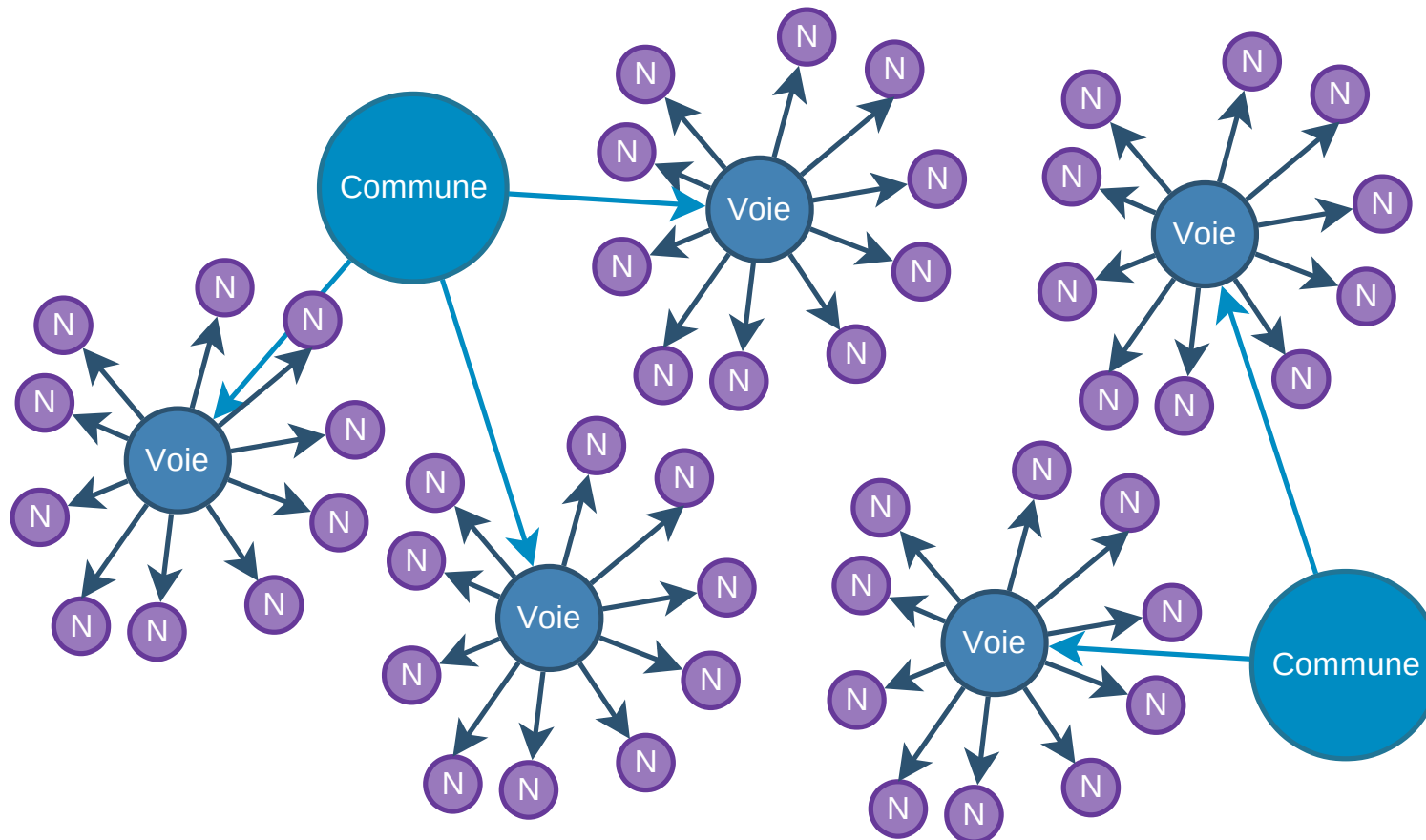


Cas pratique *scénario2

Ajout des 100 numéros par rues.

- Création de 610 000 nœuds et 610 000 relations.

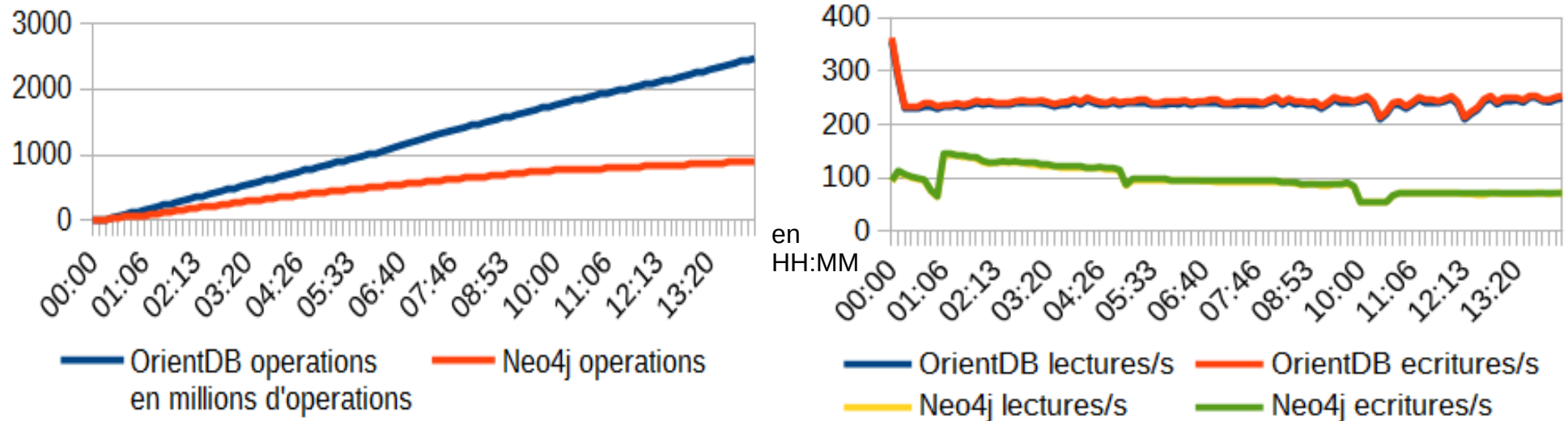
(lecture-écriture 0-100)



Cas pratique *scénario3

Création des habitants, de leurs liens sociaux et affectation de familles à des logements.
(lecture-écriture 50-50)

- ▶ Basé sur données statistiques d'OpenData et documents de sciences humaines.
- ▶ OrientDB : création de ~ 250K personnes et 16M relations
- ▶ Neo4j: création de ~ 88K personnes et 5M relations (avec baisse des statistiques des liens sociaux)



Cas pratique *scénario4

Séquences non tirées de 7 recherches diverses
(lecture-écriture 100-0)

Ses amis et les amis de ses amis jusqu'au 4e niveau de
sexe différent et du même âge ou plus jeunes :

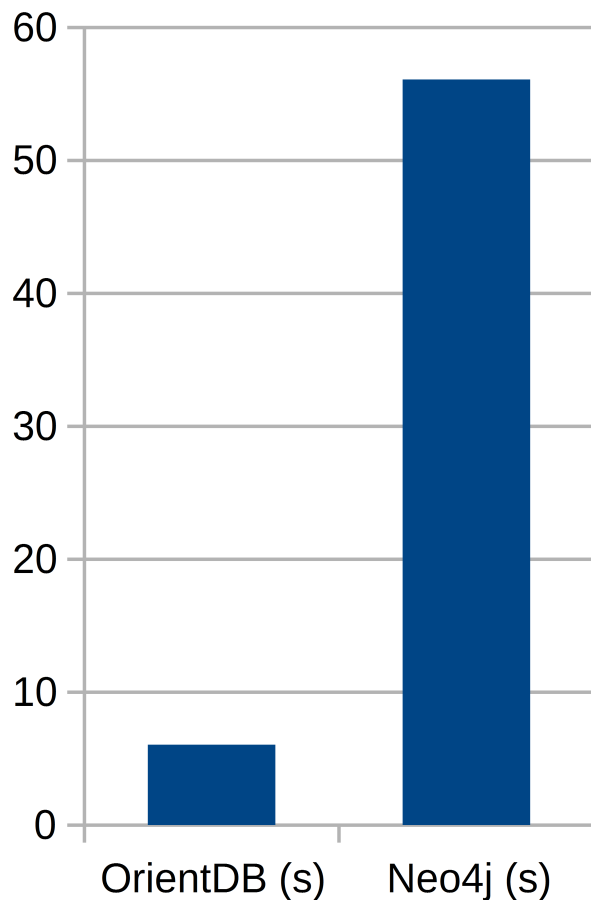
```
match (n:Personne)-[:AMI*..4]-(m:Personne)
where id(n) = %d and n.age >= m.age
and n.sexe <> m.sexe
return n,m
```

```
select from
(SELECT expand(
  set(both("AMI").both("AMI").both("AMI").both("AMI"))
) FROM %s )
where sexe <> "%s" and %s >= age
```

Cas pratique *scénario4

Séquences non tirées de 7 recherches diverses
(lecture-écriture 100-0)

Ses amis et les amis de ses amis jusqu'au 4e niveau de
du même âge ou plus jeunes :



ne)-[:AMI*..4]-(m:Personne)
%d and n.age >= m.age
n.sexe

").both("AMI").both("AMI").both("AMI"))

%s" and %s >= age

■ AMISAMISAMISAMIS -
PLUSJEUNESEXEDIFF

Conclusion

Question : Neo4j, OrientDB, lequel est le meilleur ?

Conclusion

Question : Neo4j, OrientDB, lequel est le meilleur ?

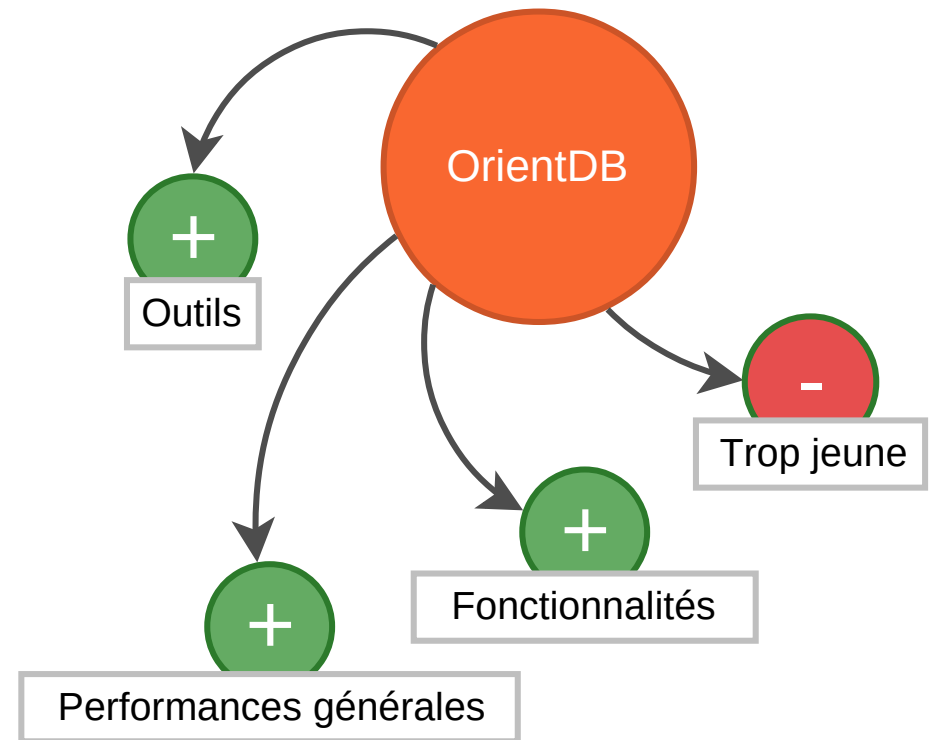
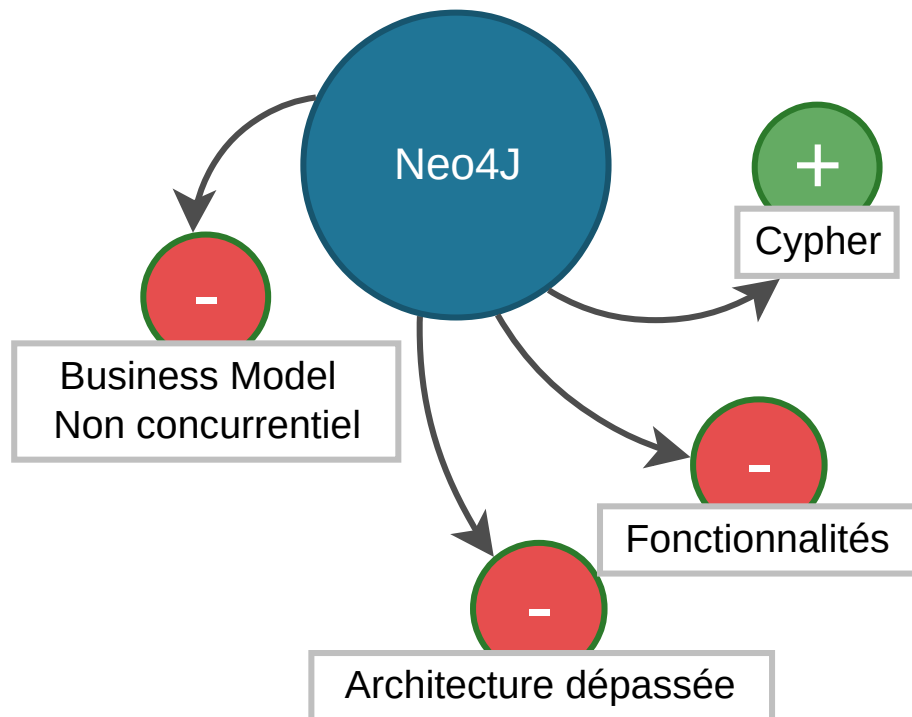
Réponse : ça dépend...

Conclusion

Question : Neo4j, OrientDB, lequel est le meilleur ?

Réponse : ça dépend...

Toutefois,



Fin

- Merci de votre attention
- Questions

{ Framework et les données
seront peut-être disponibles
sur <https://github.com/wh6b> }

