

Top 5 des cas d'utilisations

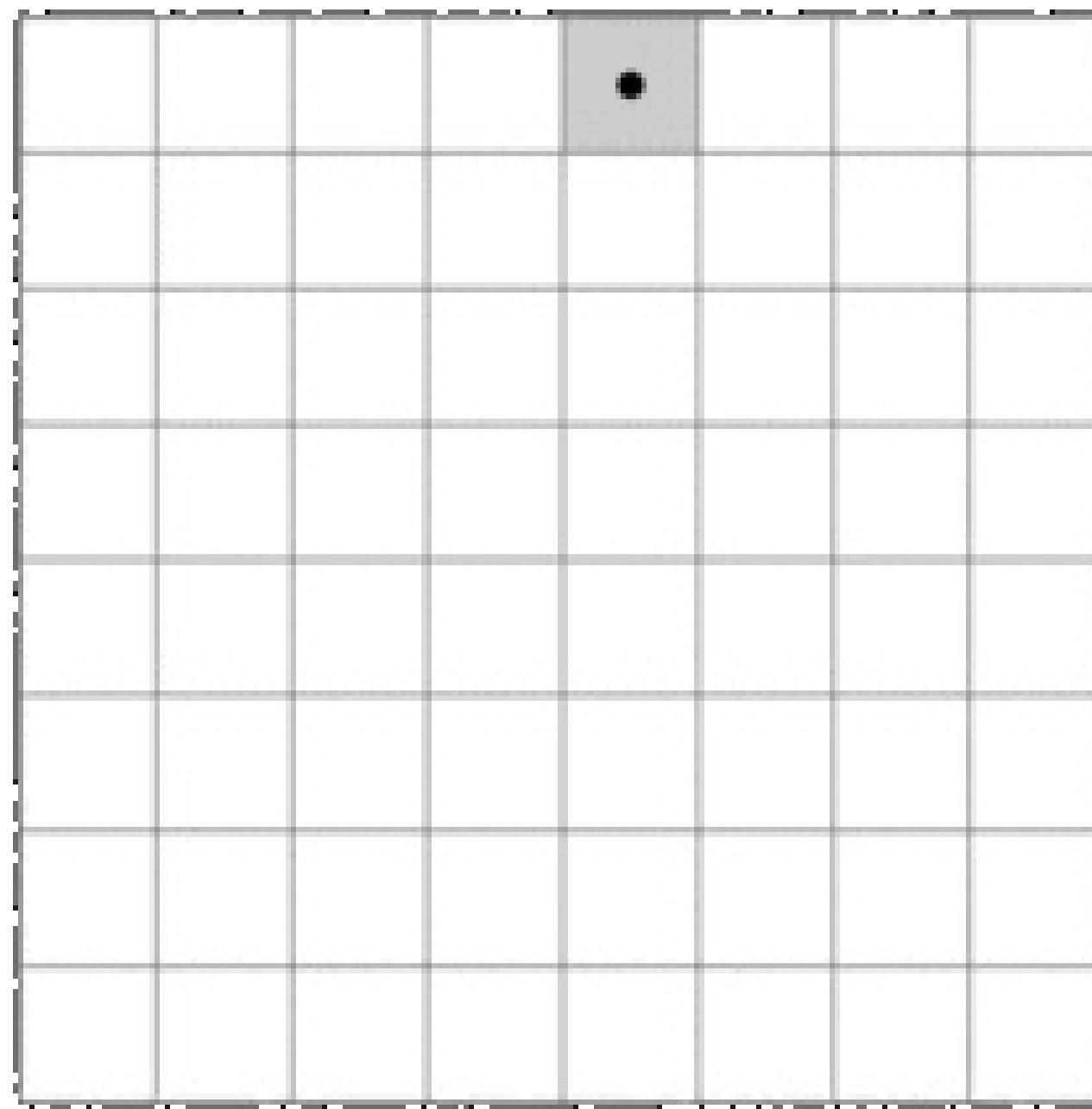
Benoit Simard (@logisima)



UN BREF HISTORIQUE

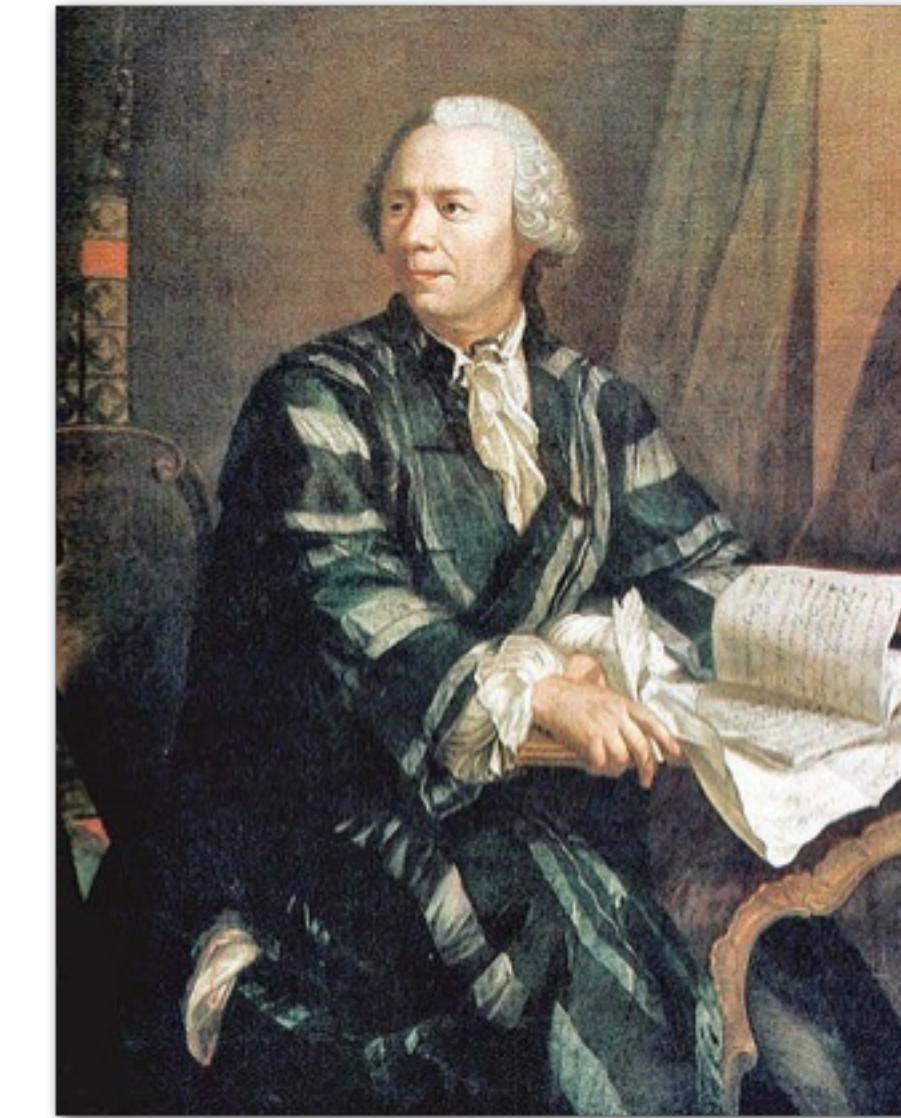
VERS 800 :LE PROBLÈME DU CAVALIER

Un cavalier posé sur une case d'un échiquier doit en visiter toutes les cases sans passer deux fois sur la même.



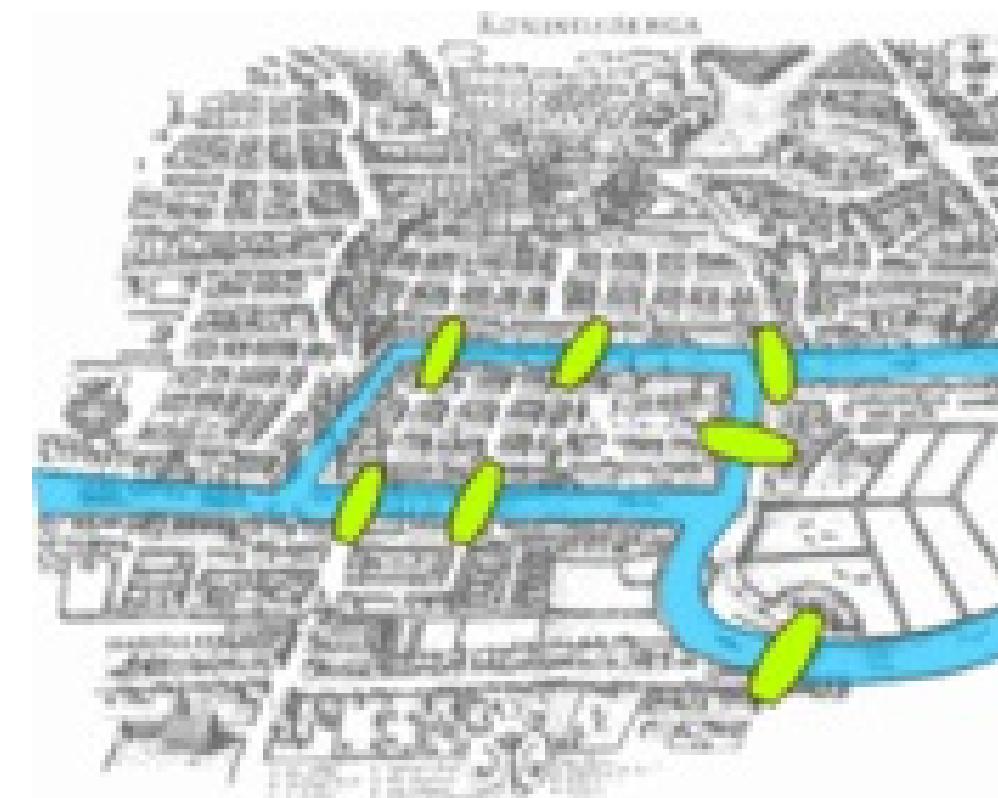
LEONHARD EULER 1707-1783

Mathématicien Suisse à l'origine de la théorie des graphes.

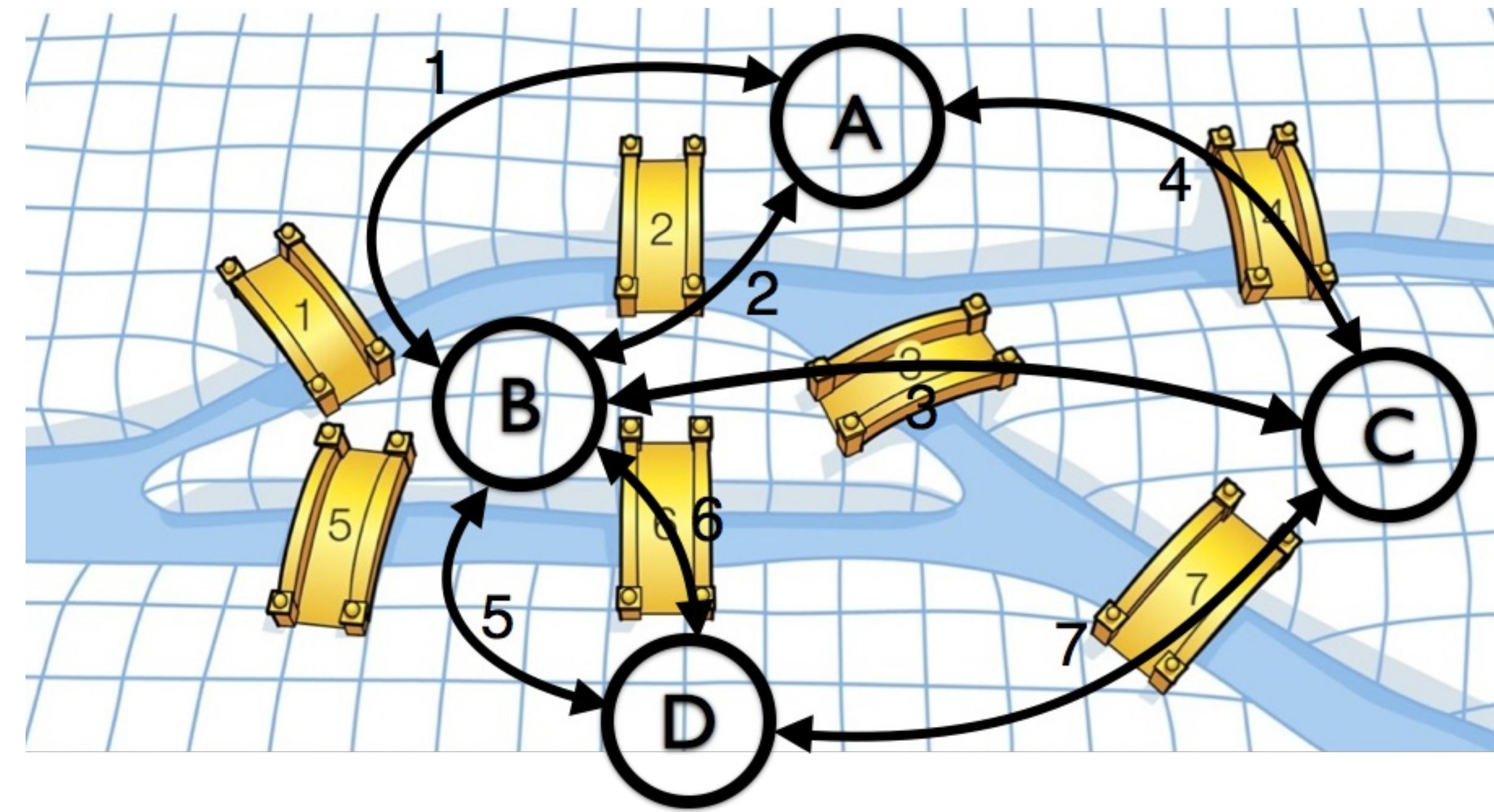


LES 7 PONTS DE KONIGSBERGS

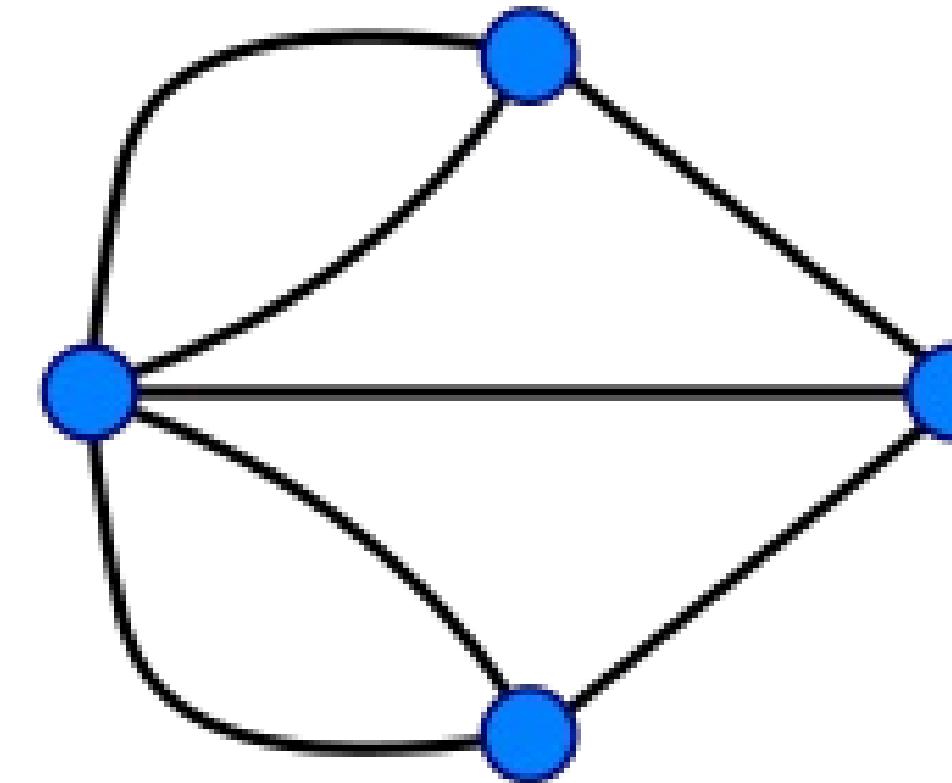
Est-ce qu'il existe une promenade permettant à partir d'un point de départ au choix, de passer une et une seule fois par chaque pont.



LES 7 PONTS DE KONIGSBERGS



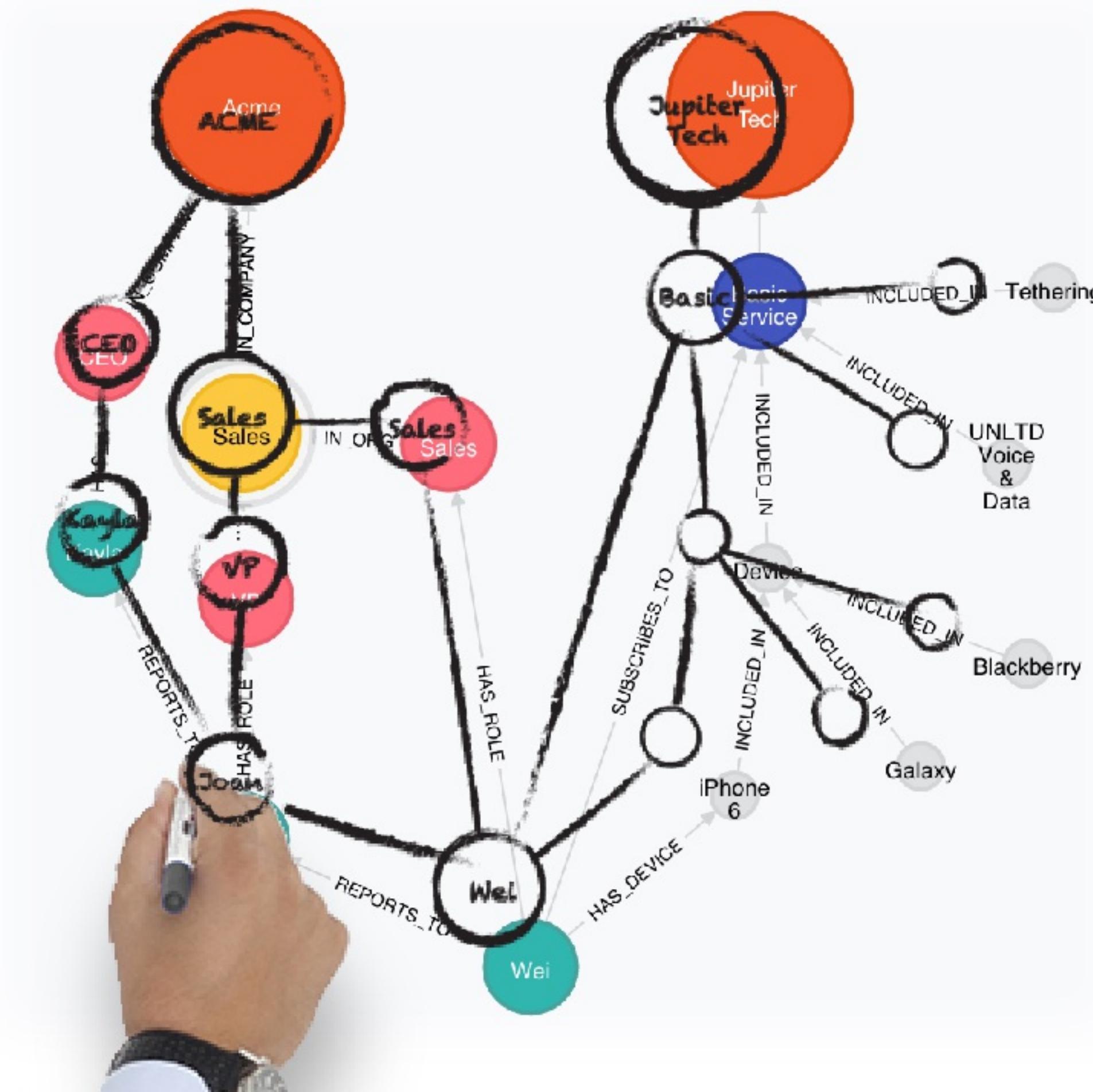
LES 7 PONTS DE KONIGSBERGS



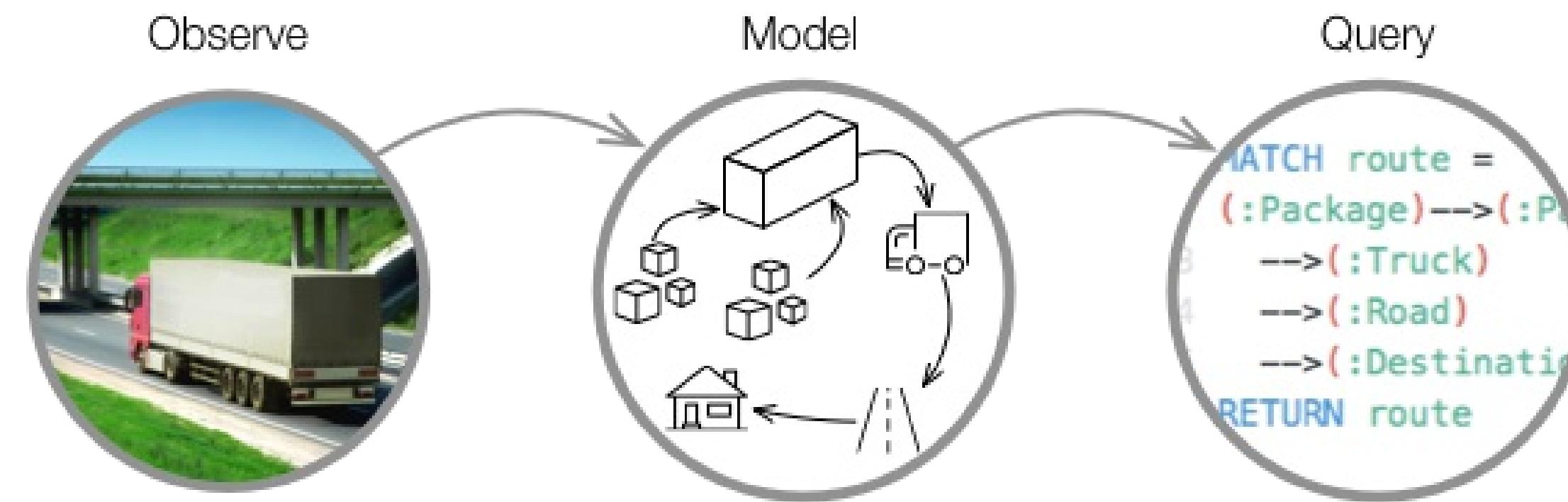
Théorie des graphes et des parcours Eulerien

LES BASES GRAPHE

LA MODÉLISATION EN GRAPHE



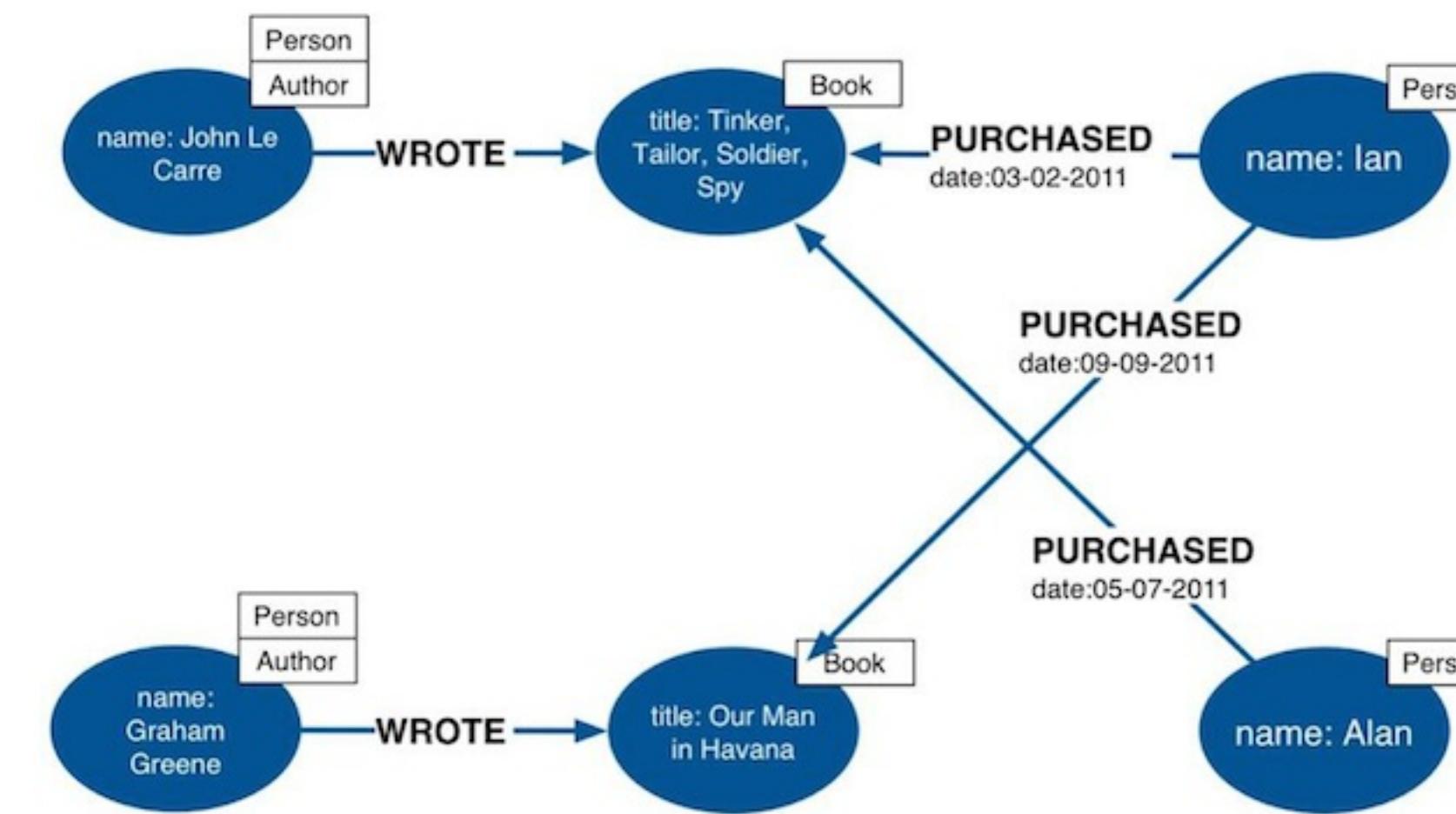
UNE MODÉLISATION NATURELLE



La modélisation du monde réel et celle de la base sont identiques

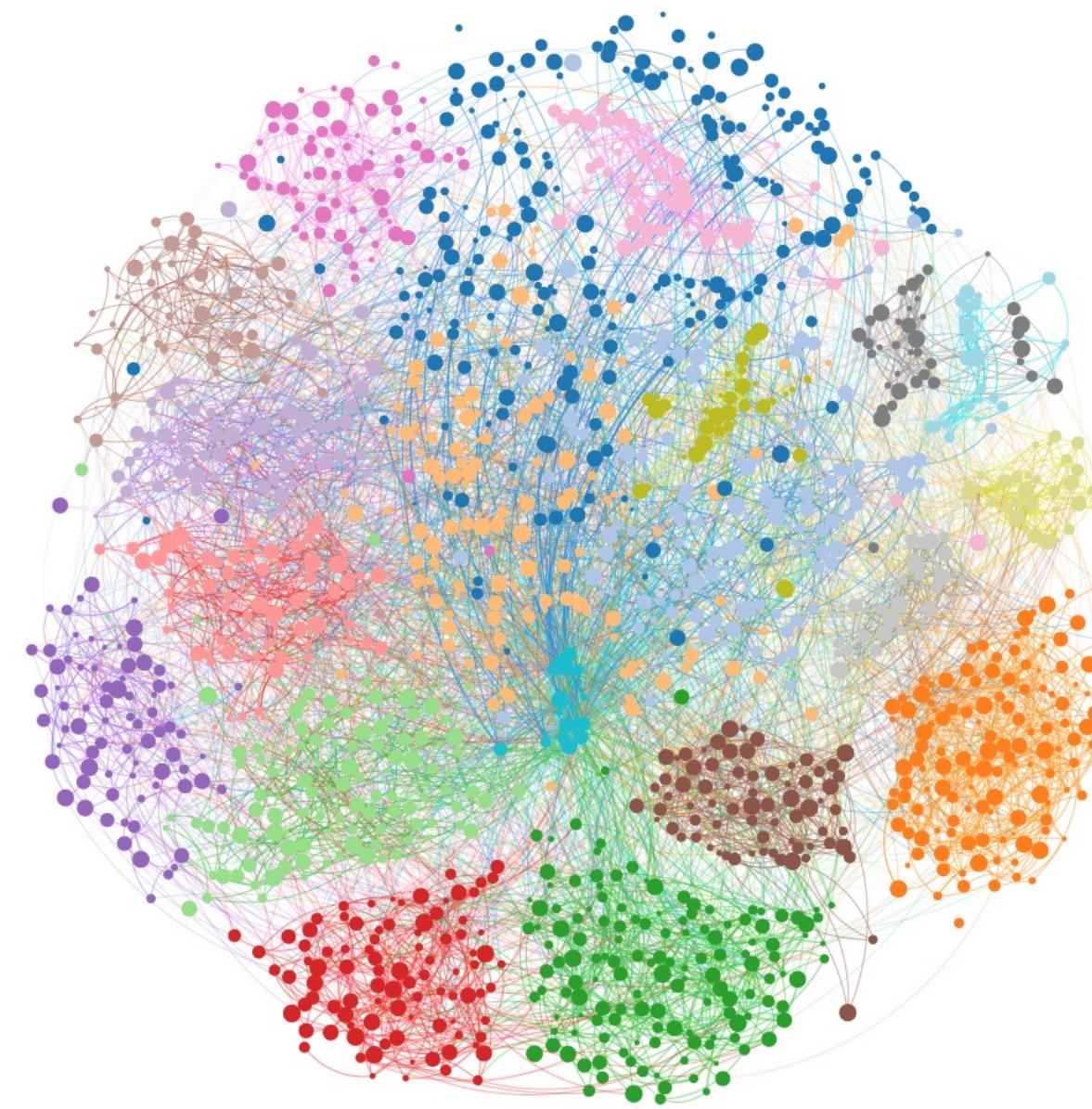
UN EXAMPLE DE GRAPHE

- Des noeuds avec des propriétés et des labels
- Des relations typés et orientés avec des propriétés



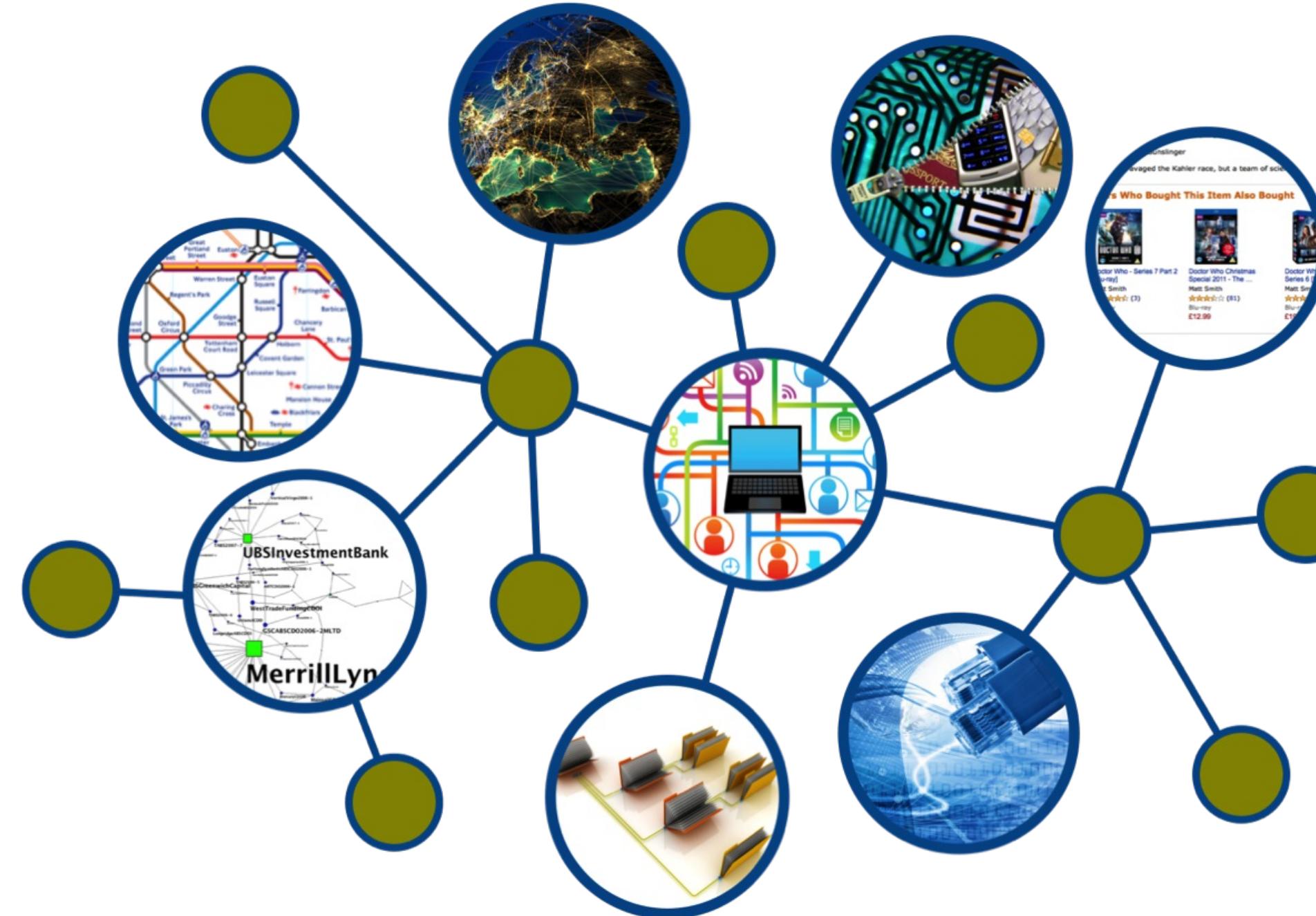
LINKED DATA

Les relations existant entre les données sont presques plus importantes que les données elles-mêmes



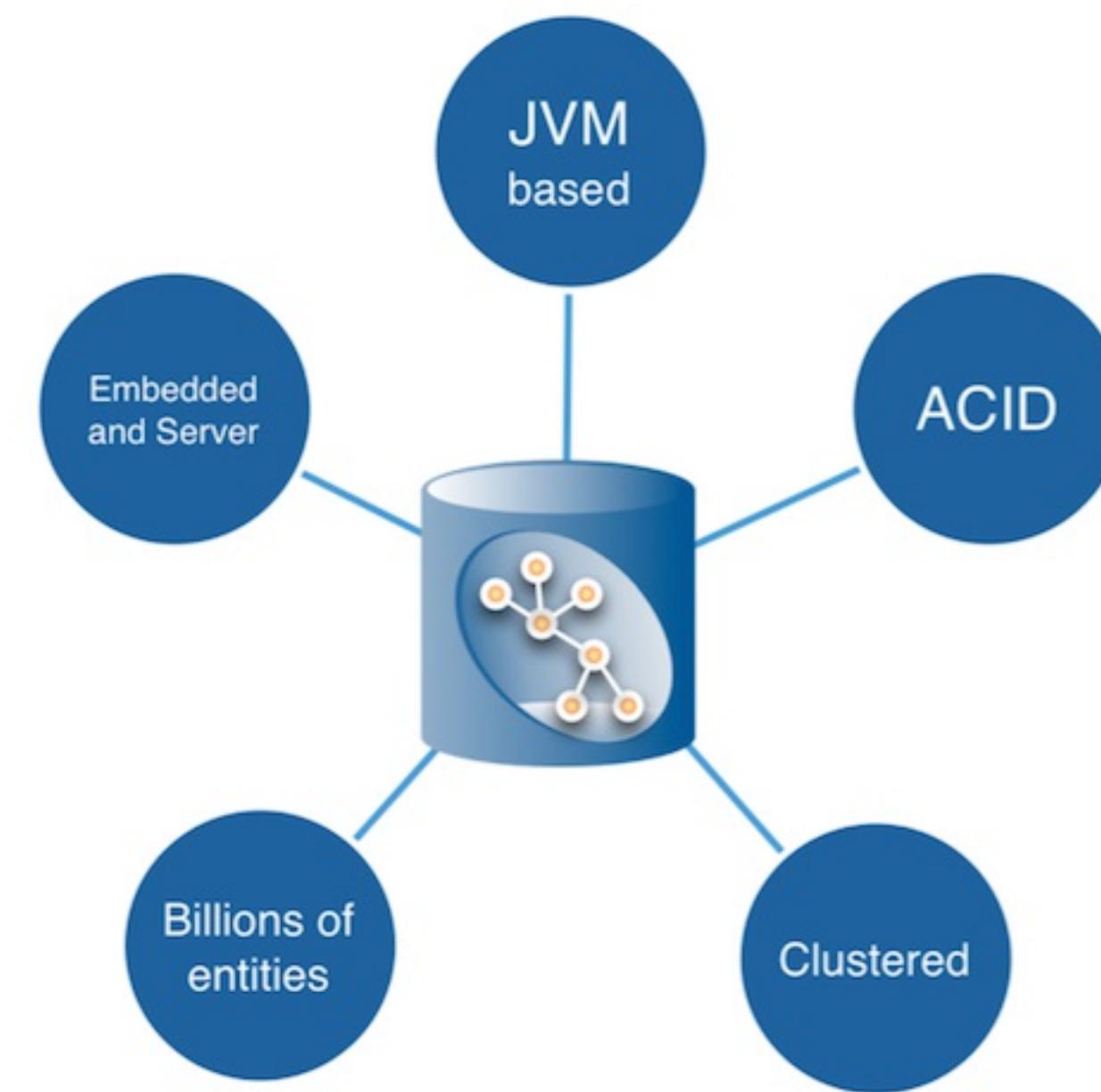
COMPLÉXITÉ DES DONNÉES

complexité = f(volumétrie, semi-structure, connectivité)



NEO4J

Premières lignes de code datent de **2000** et la **version 1.0** est sortie en **2010**.



Neo4j is a NoSQL Graph Database

- Open source
- Welcoming UI
- Easy data modeling
- Readable queries
- Active community
- High performance
- optional schema

MOTEUR DE RECOMMANDATION EN TEMPS RÉEL

LA RECOMMANDATION

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

Suivre

Philip Rathle @prathle

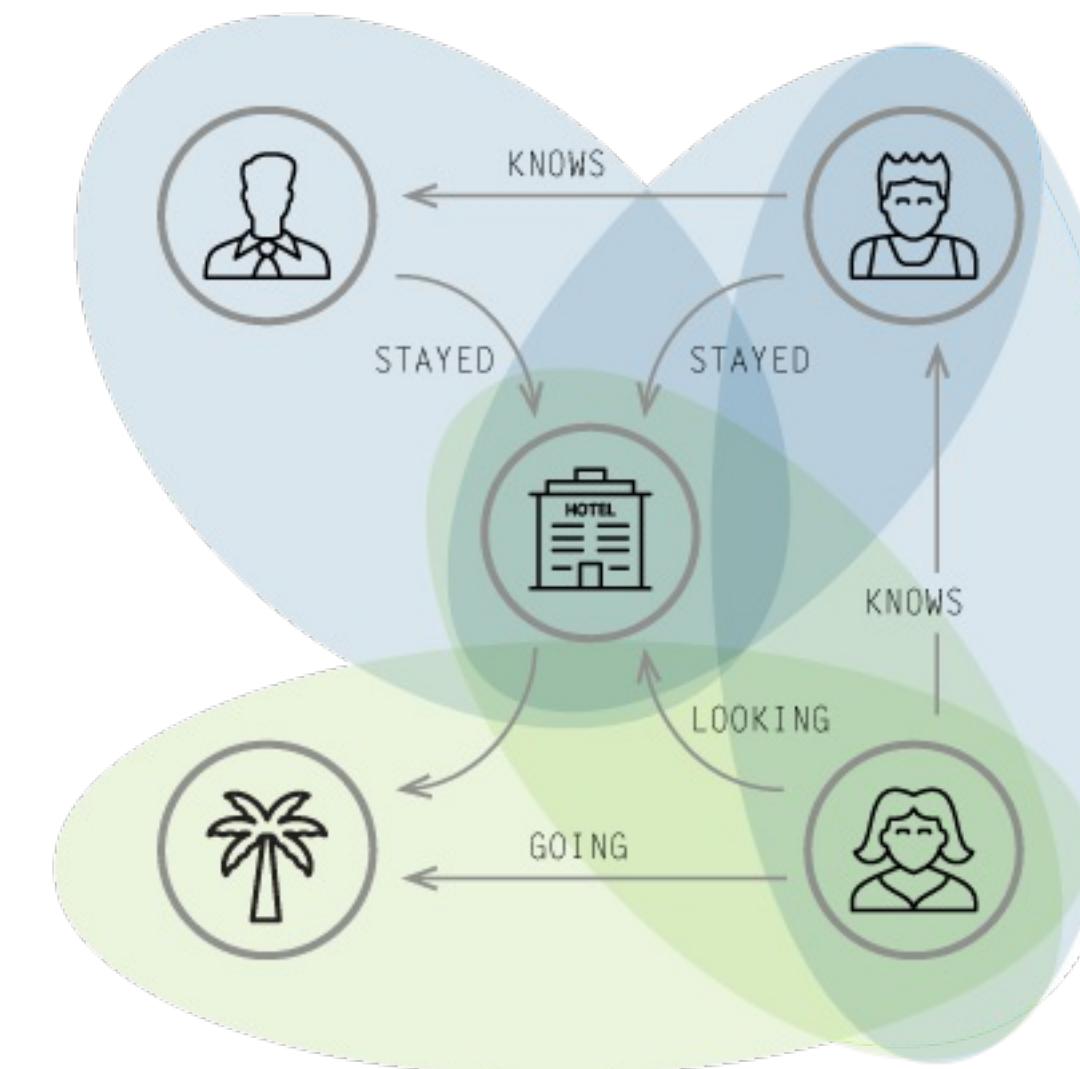
Suivre

Trouver des amis

POURQUOI DU GRAPHE ?

Faire des recommandations efficaces repose sur une base qui comprend les relations existants entre les entités, ainsi que la qualité et la solidité de ces connexions.

C'est un graphe !



LES DÉFIS LIÉS À LA RECOMMANDATION

Traiter de gros volumes de données et de relations pour comprendre le contexte

Les algorithmes standard de recommandations s'appuient sur des données hautement connectées en expansion rapide.

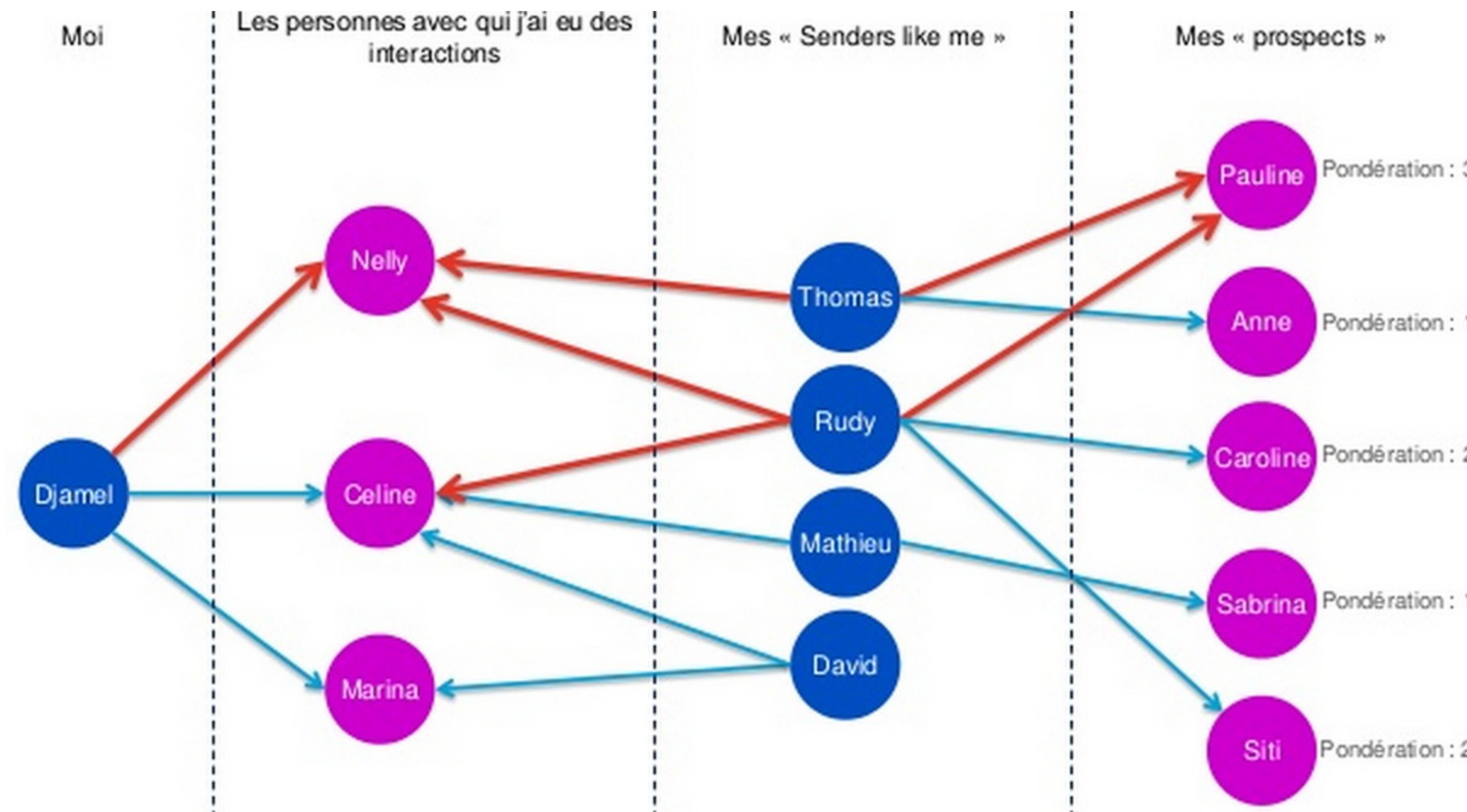
Offrir des recommandations pertinentes en temps réel, basé sur des faits immédiats

La puissance d'un système de suggestion repose sur sa capacité à faire des recommandations en temps réel à l'aide de faits immédiats.

Prendre continuellement en compte des données et relations nouvelles

Du fait de la croissance rapide dans la taille et le volume des données, le système doit prendre en compte à la fois les besoins actuels et futurs.

ETUDE DE CAS : MEETIC



ETUDE DE CAS : MEETIC

Problème

- Recommandations temps-réel afin de fidéliser les utilisateurs et attirer de nouveaux membres.
- Isoler et traiter les cas des super-dragueurs et super-dragués
- A l'origine réalisé par un traitement batch écrit avec MS SQL Server
- La volumétrie : 10 millions de noeuds (ie. de membres) | 100 millions de relations (ie. d'interactions)

Solution

- Expérience temps réel pour les utilisateurs de Meetic
- Solution scalable.
- Faible coût de maintenance et architecture fiable

AUTRES CLIENTS



Fournit des recommandations en temps réel sur leur service de vente en ligne, en utilisant les informations relatives aux préférences des clients.



Le géant du commerce électronique eBay a également utilisé avec succès un moteur de suggestion fonctionnant sur un graphe, en l'occurrence pour une solution sophistiquée d'acheminement de paquets/courrier en temps réel (Ebay now).

LA DÉTECTION DE FRAUDES

LA DÉTECTION DE FRAUDES

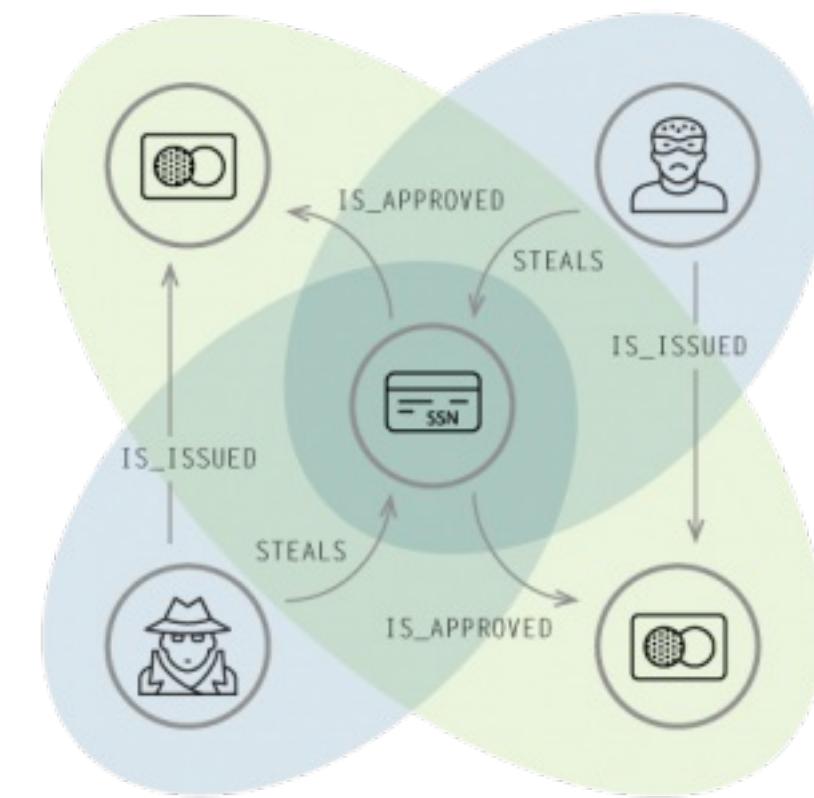
- La fraude coûte des milliards chaque années aux banques et assurances
- Les techniques de fraudes sont de plus en plus sophistiquées
- Il faut constater la fraude en temps réel (ie. avant qu'elle est lieu)



POURQUOI DU GRAPHE ?

Bien qu'aucune mesure de détection de fraudes ne soit parfaite, il est possible d'introduire des améliorations considérables si l'on se penche sur les connexions qui relient les points de données individuels.

Il faut étudier l'interconnexion des données



LES DÉFIS LIÉS À LA DÉTECTION DE FRAUDES

Effectuer une analyse complexe des liens pour découvrir les schémas de fraude

Découvrir les réseaux de fraude nécessite d'exécuter une analyse d'une complexité élevée sur les relations existant entre les données ; une difficulté exacerbée par la croissance des réseaux de fraude.

Déetecter et empêcher les fraudes en temps réel

Pour faire barrage à un réseau de fraude, vous devez effectuer une analyse des liens en temps réel au niveau d'un jeu de données interconnectées, à partir de la création d'un faux compte jusqu'à l'exécution d'une transaction frauduleuse.

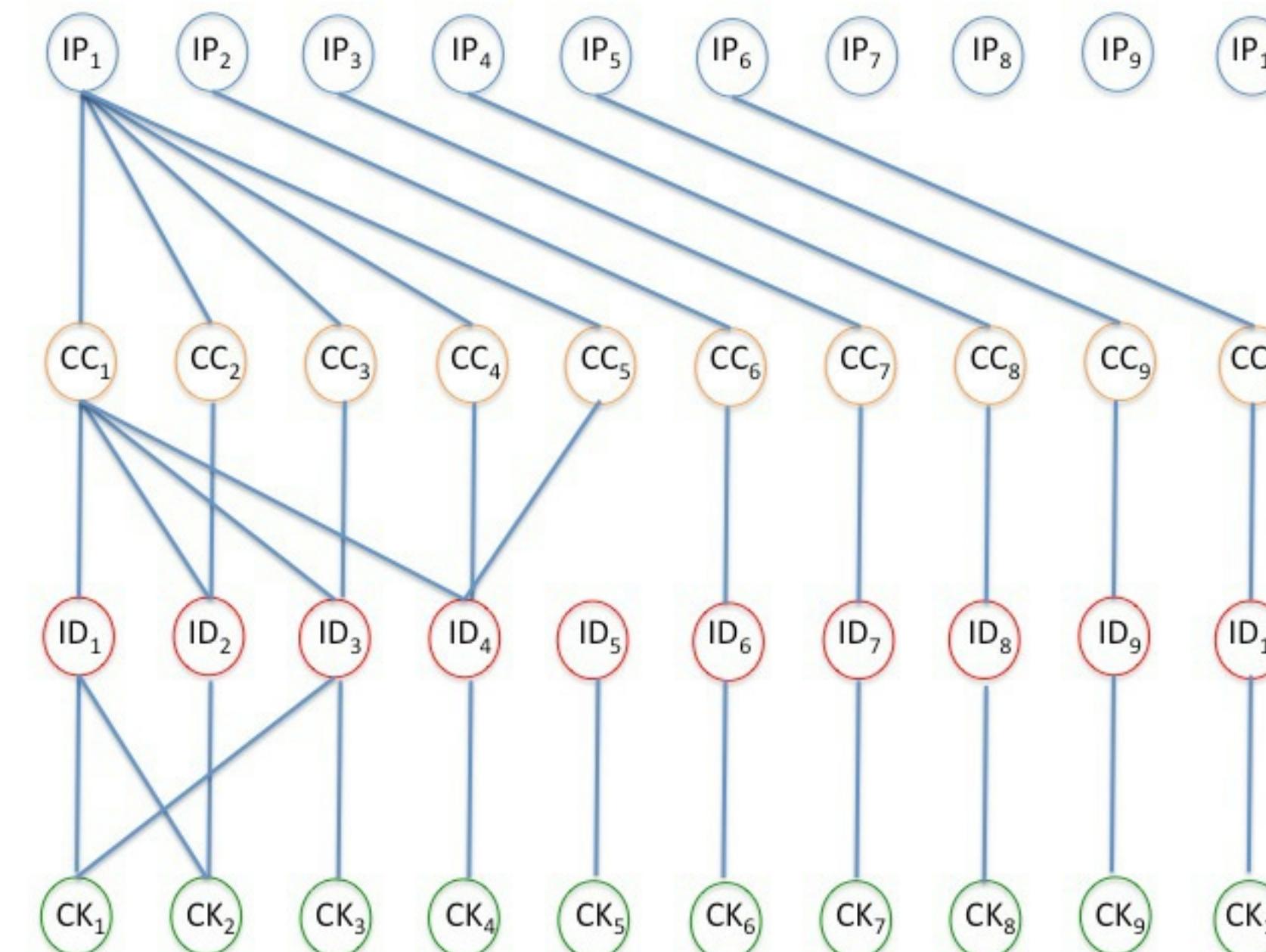
Réseaux de fraude en évolution et dynamiques

Les réseaux de fraude se développent continuellement en forme et en taille et votre application doit détecter ces schémas de fraude dans cet environnement hautement dynamique.

ETUDE DE CAS : FRAUDE DANS LE E-COMMERCE

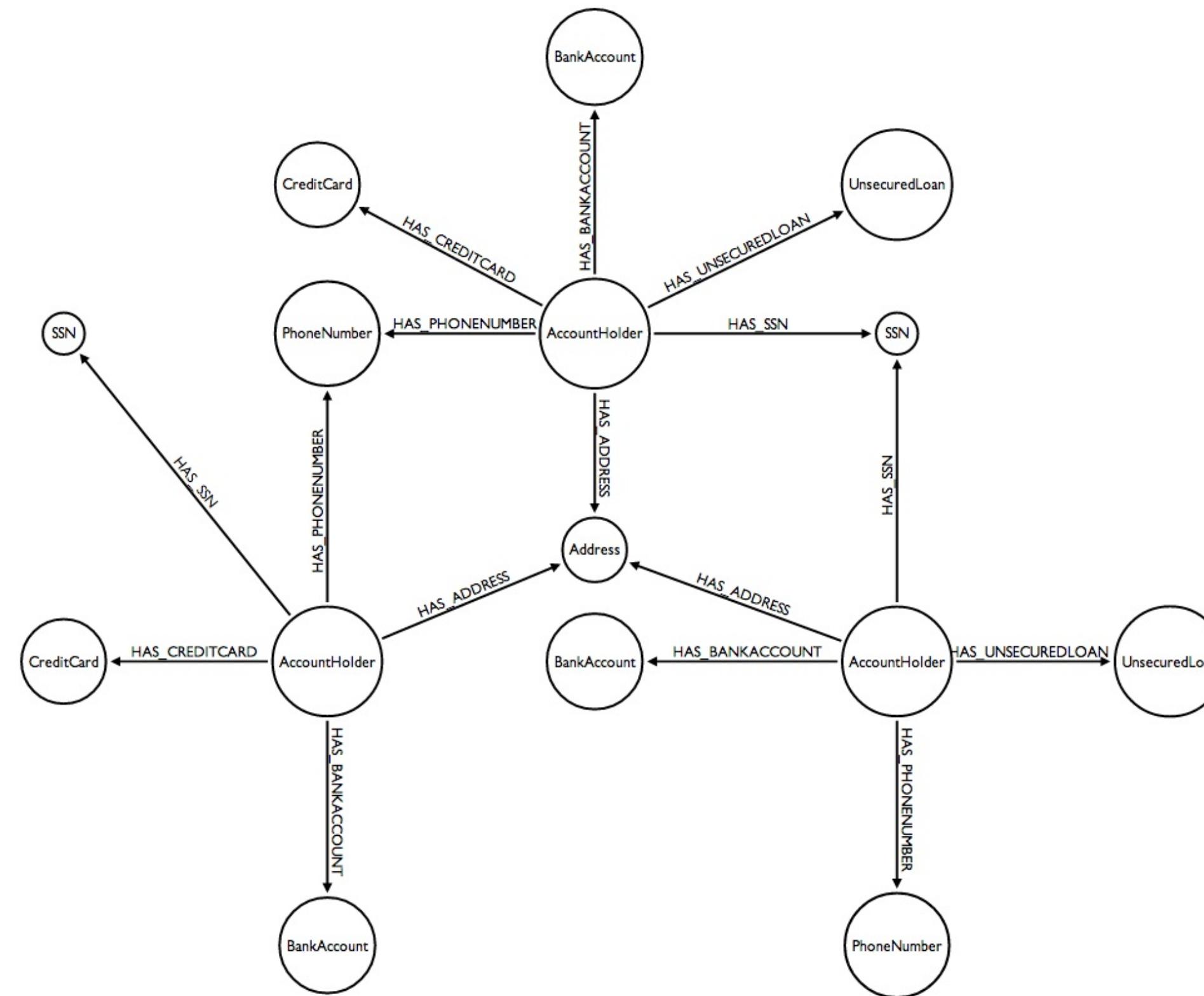
Graphe d'une série de transactions effectuée avec fraude probable se produisant à l'adresse IP1.

(IP : Adresse IP, CC : carte de crédit, ID : Identifiant de l'utilisateur, CK : Adresse de livraison)



ETUDE DE CAS : LA DÉTECTION D'ANNEAU

Constituer un ensemble de fausses identités à partir d'identités volées, dans l'objectif de commettre un acte frauduleux



LA GESTION DE DONNÉES DE RÉFÉRENCE

LA GESTION DE DONNÉES DE RÉFÉRENCE

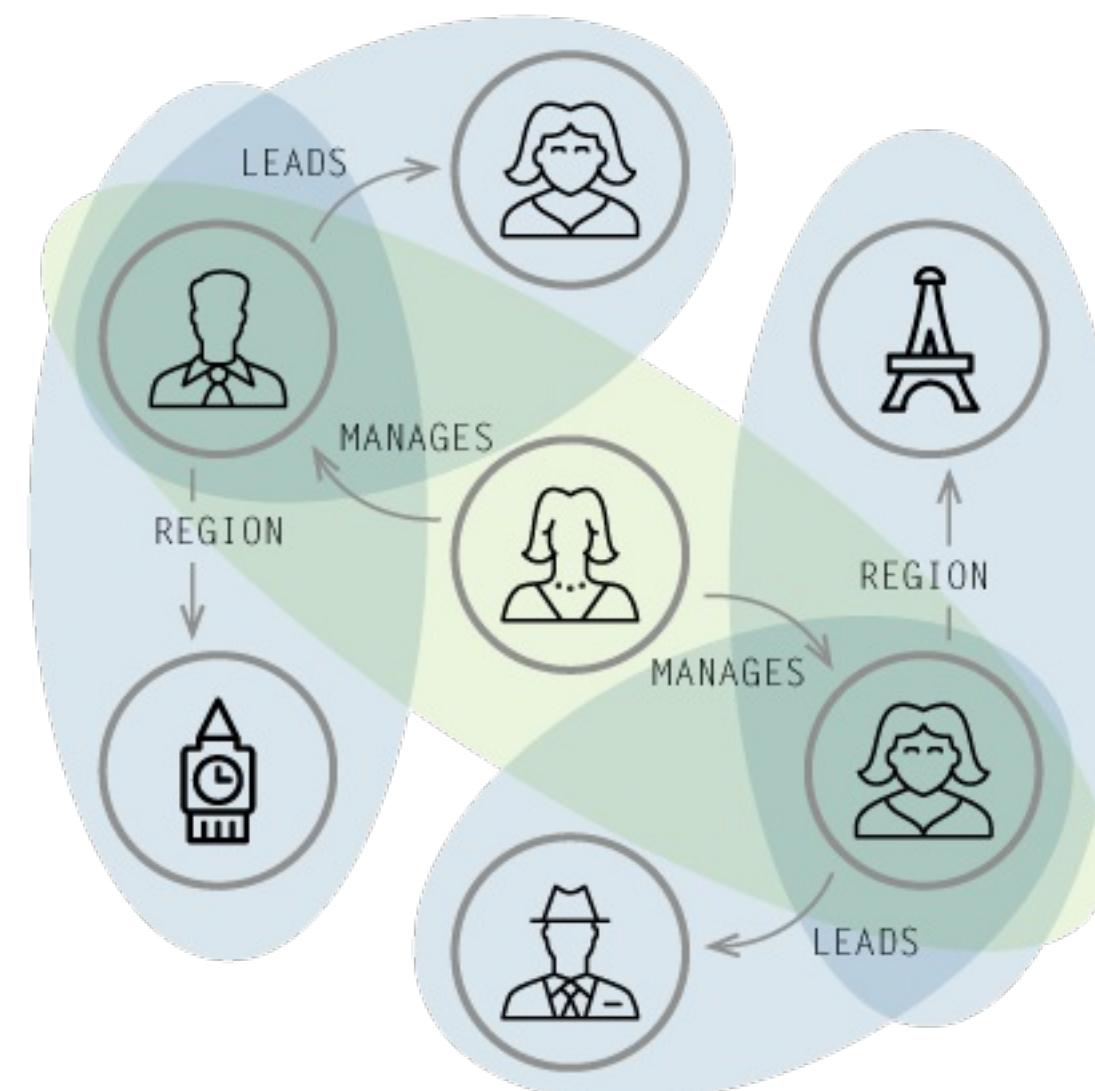
Les données de référence constituent le pilier de votre entreprise et incluent des données telles que les utilisateurs, les clients, les produits, les comptes, les partenaires, les sites, les unités commerciales, ...

Ces données sont :

- stockées à différents emplacements
- avec des chevauchements et des redondances
- dans divers formats
- avec des qualités variables

POURQUOI DU GRAPHE ?

- Les données sont hautement connectées et partagées
- Les systèmes de gestion ne sont pas flexibles
- L'interconnexion des données de références permettent d'obtenir un avantage concurrentiel



LES DÉFIS LIÉS AU MDM

Jeux de données complexes et hiérarchiques

Gérer les hiérarchies descendantes des données de référence avec une base de données relationnelle se traduit par un code complexe lent à exécuter, cher à concevoir et long à maintenir.

Performances de stockage et de requête en temps réel

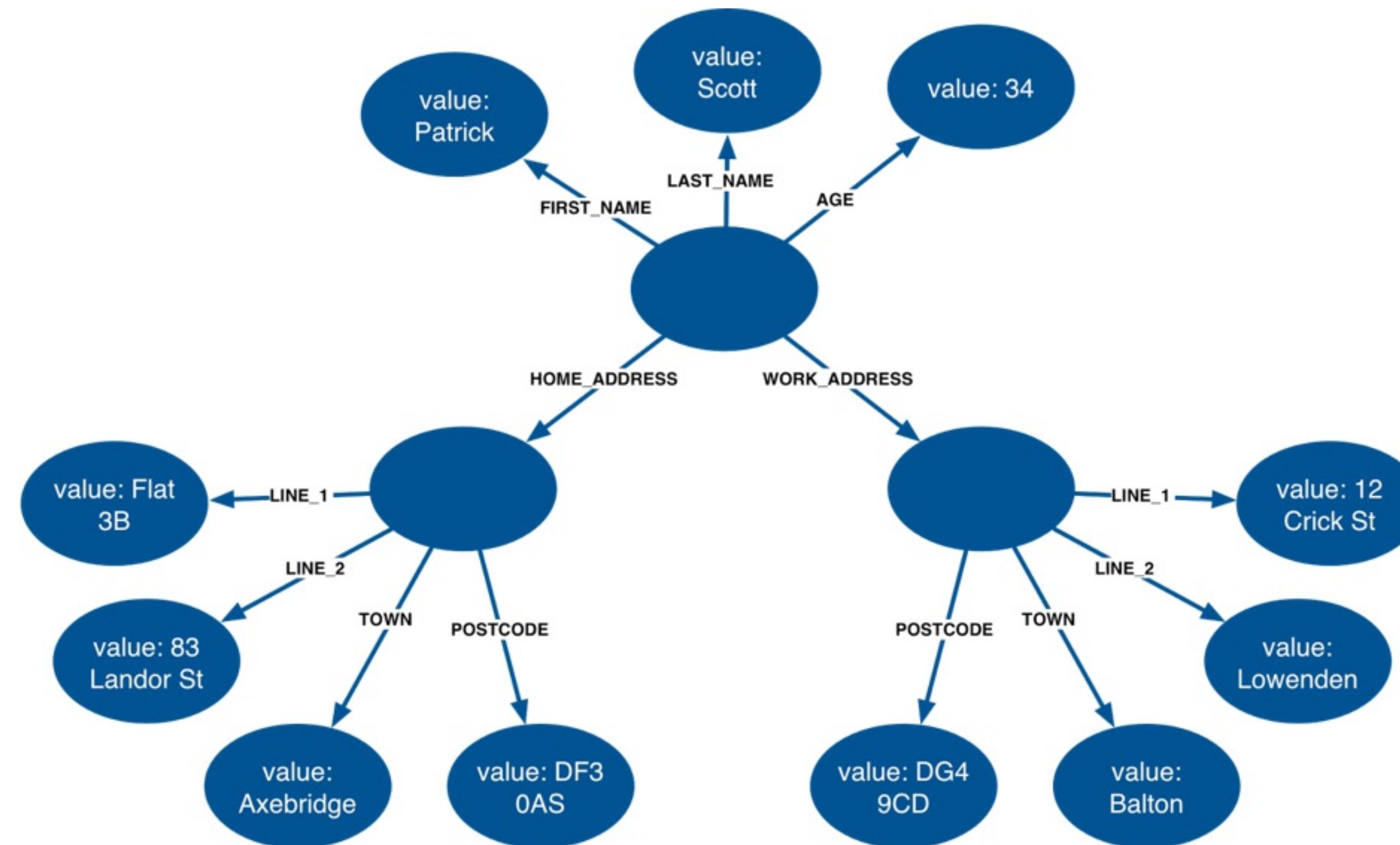
Le magasin de données de référence doit s'intégrer à une multitude d'applications au sein de l'entreprise et fournir des données à ces applications. Fournir des informations en temps réel sur des jeux de données complexes et hautement interconnectés est un véritable challenge.

Structure dynamique

Les données de référence étant dynamiques par nature, il est plus difficile pour les développeurs de concevoir des systèmes qui prennent en compte leur évolution.

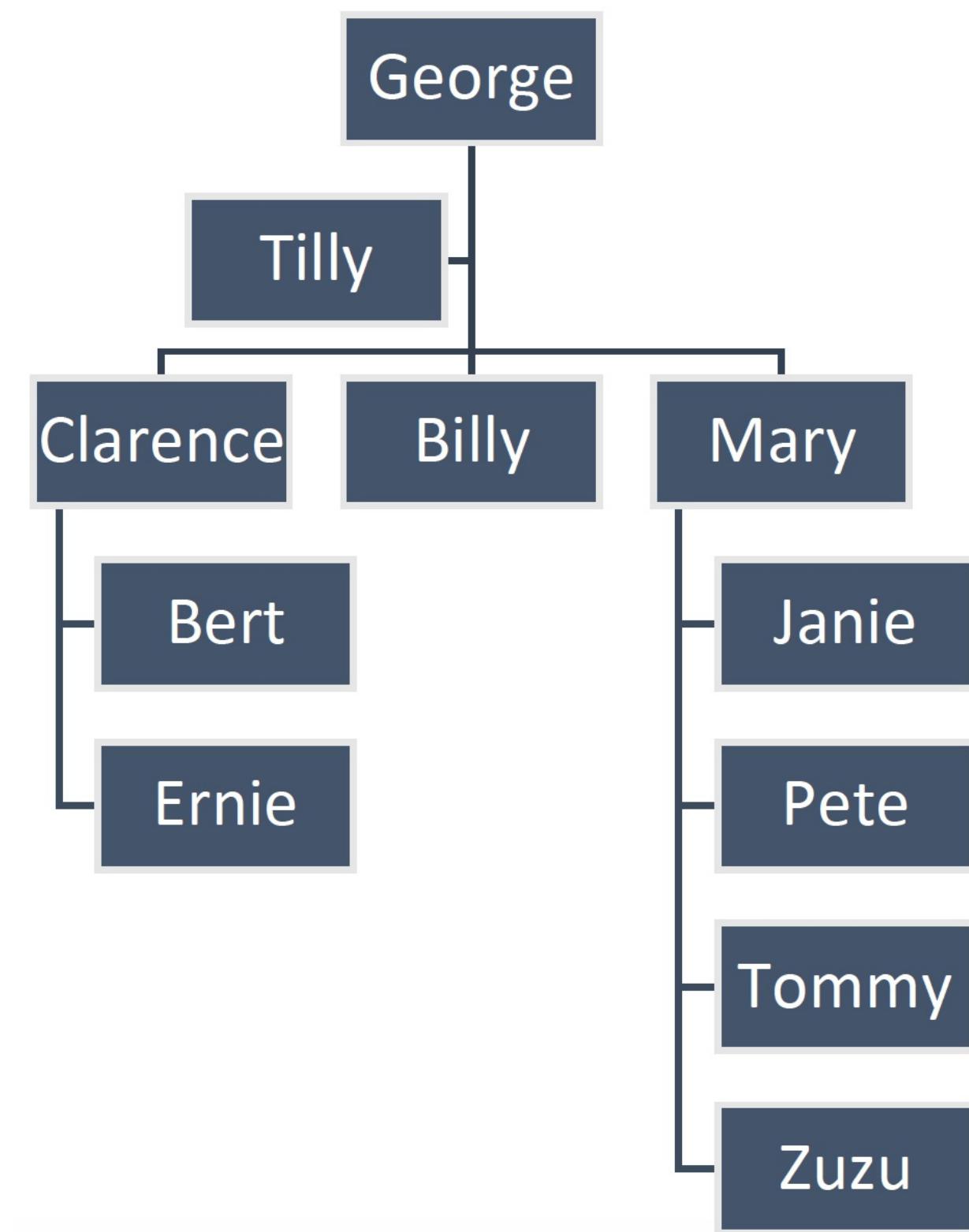
ETUDE DE CAS : RECONCILIATION DES DONNÉES

Réconciliation des données par degré de similitude



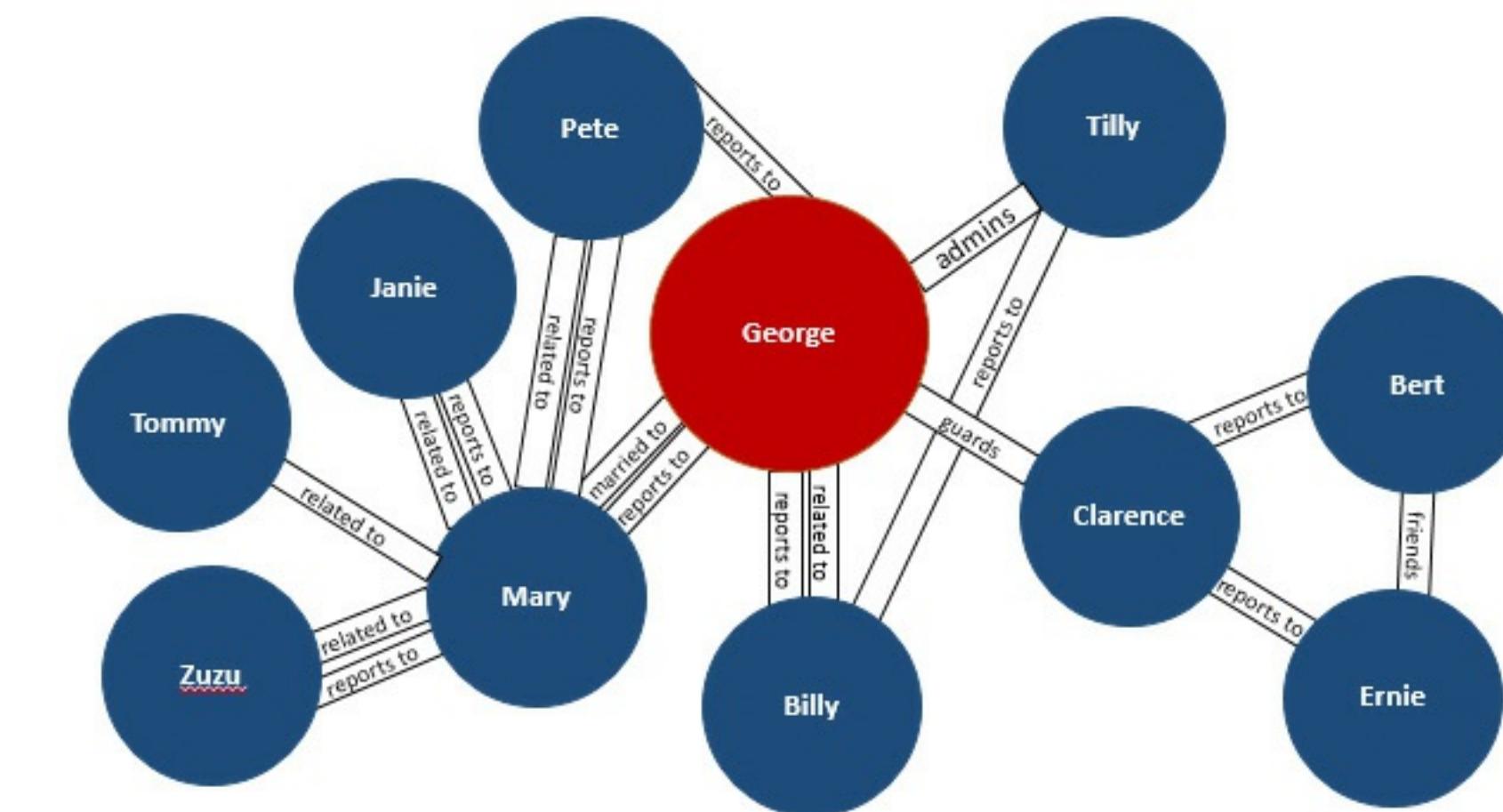
ETUDE DE CAS : LA MODÉLISATION RH D'UNE ENTREPRISE

Hiérarchie de données de référence illustrant des structures de rapport et de supervision d'employé.



ETUDE DE CAS : LA MODÉLISATION RH D'UNE ENTREPRISE

Réseau de données de référence détaillant les relations de rapport et de supervision d'employé, ici avec une plus grande complexité du **monde réel**.



CLIENT



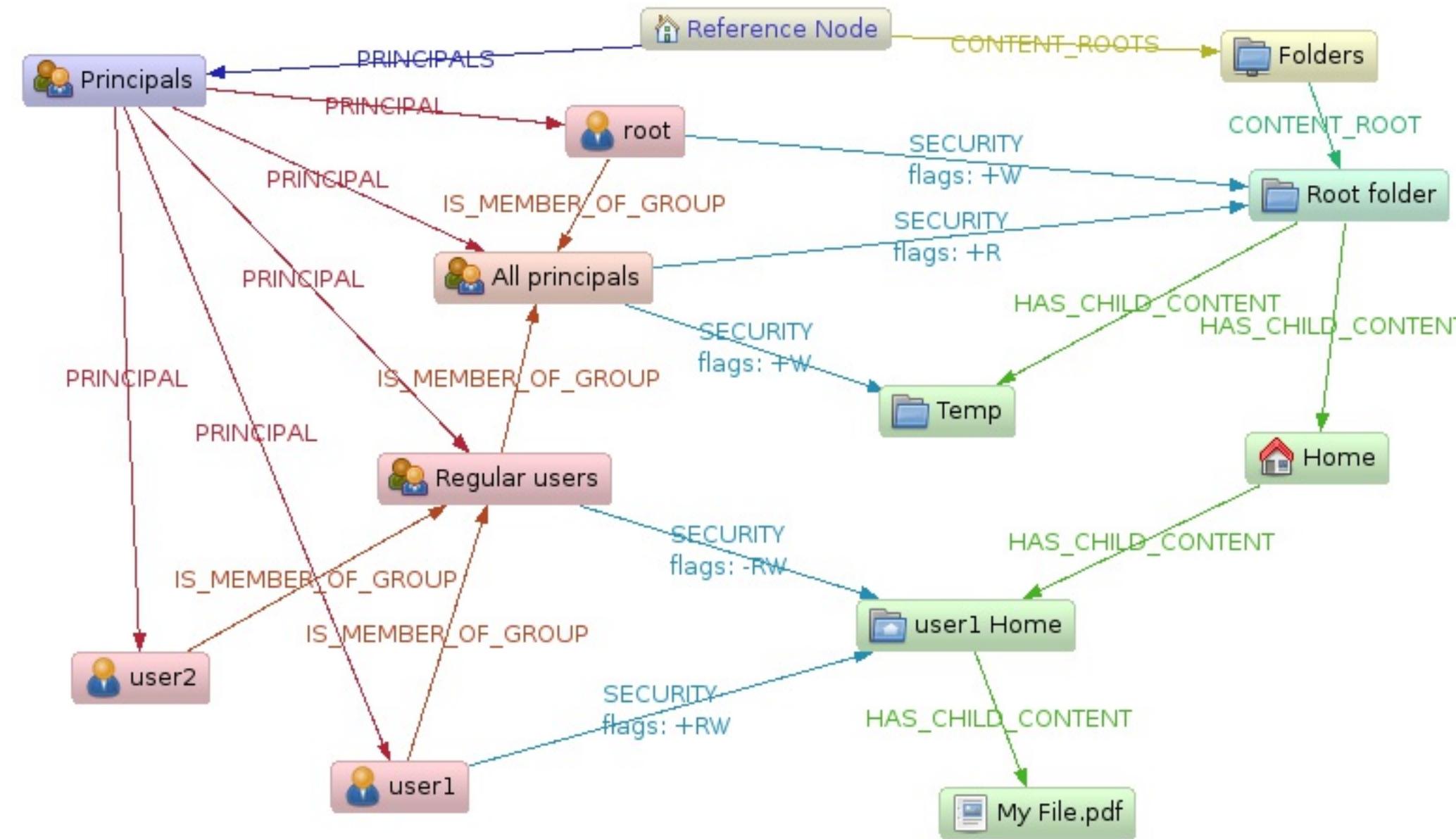
Adidas utilise Neo4j pour créer leur référentiel interne, incluant les données sur les produits, les marchés, les médias sociaux, les actifs numériques, ...



Wine Data System utilise Neo4j pour réconcilier les bases de données des négociants afin de construire un référenciel unique.

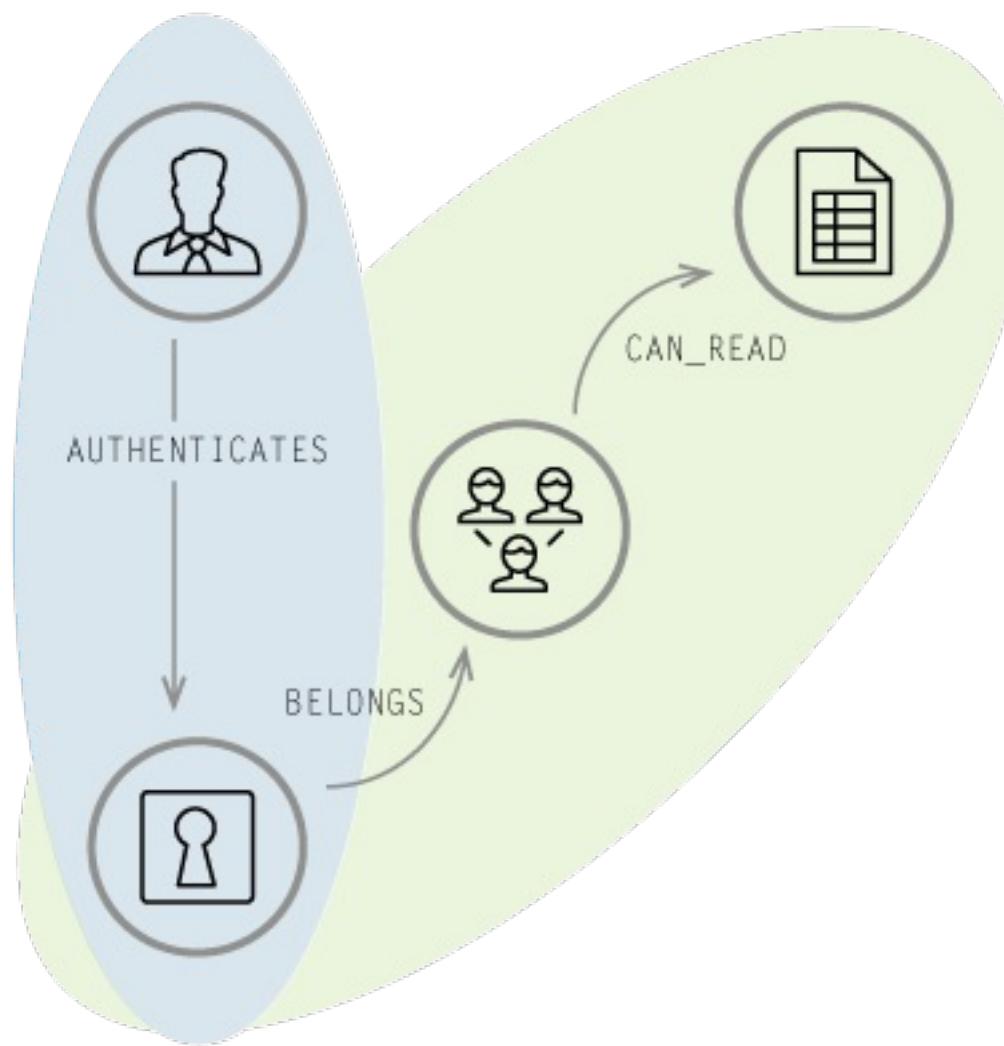
LA GESTION DES IDENTITÉS ET DES ACCÈS

LA GESTION DES IDENTITÉS ET DES ACCÈS



POURQUOI DU GRAPHE ?

C'est un graphe !



LES DÉFIS LIÉS À LA GESTION DES IDENTITÉS / ACCÈS

Autorisations d'identité et d'accès hautement interconnectées

Pour vérifier une identité précise et ses autorisations d'accès, le système doit traverser un jeu de données hautement interconnectées dont le volume et la complexité augmentent en permanence.

Productivité et satisfaction du client

À mesure que le volume d'utilisateurs, de produits et d'autorisations augmente, les systèmes traditionnels n'assurent plus de réponses réactives, ce qui se traduit par une frustration chez les utilisateurs.

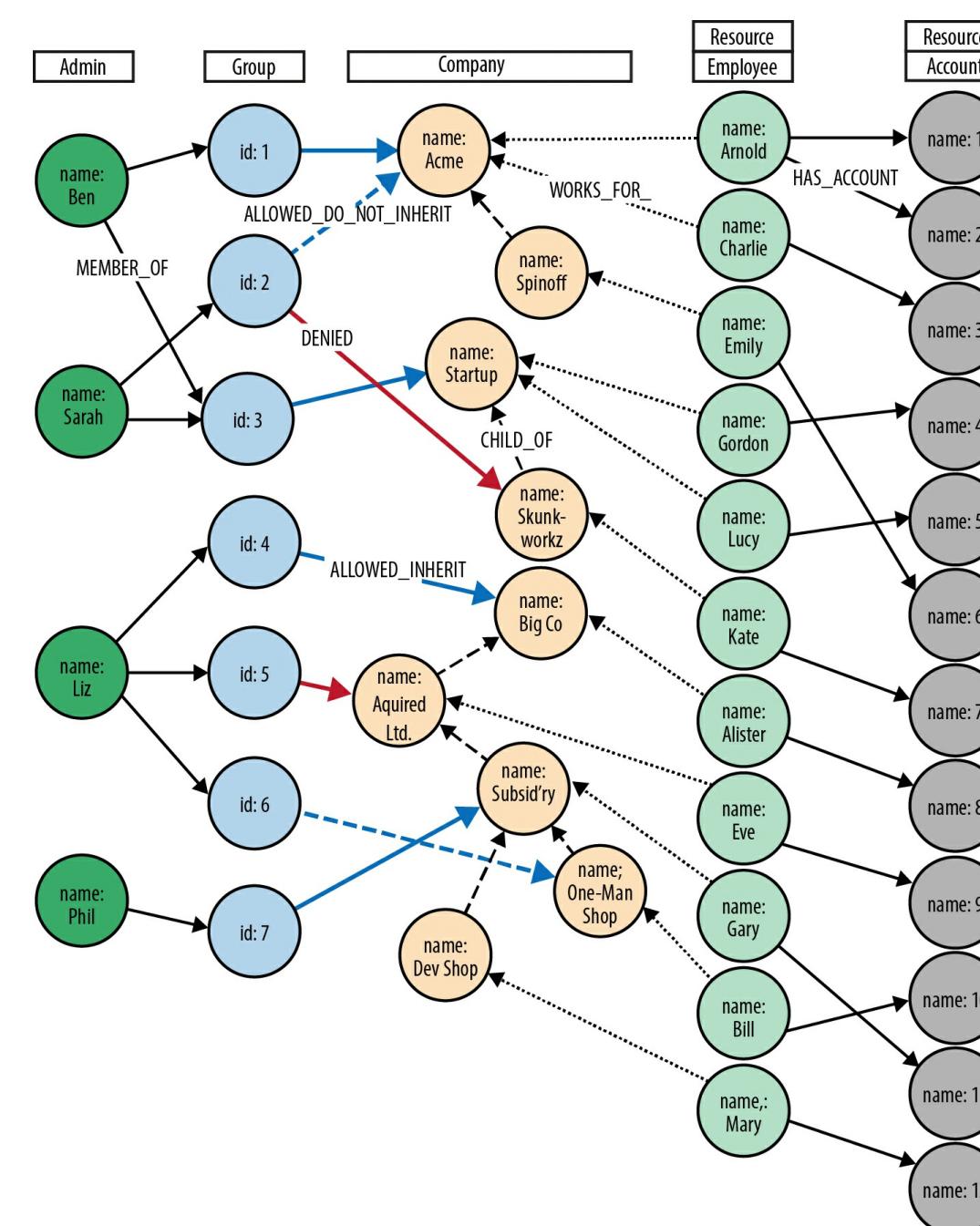
Structure et environnement dynamiques

Avec l'augmentation rapide de la taille des utilisateurs et de leurs métadonnées associées, votre application doit prendre en compte les besoins de gestion des identités à la fois actuels et futurs.

ETUDE DE CAS : TELENO

Telenor permet à ses clients entreprises d'utiliser leurs comptes pour ajouter ou supprimer des services à leurs employés.

Ainsi il faut garantir que les utilisateurs et les administrateurs ne voient et ne modifient que les parties de l'entreprise et les services qu'ils ont le droit de gérer.



ETUDE DE CAS : TELENO

Définition du problème

- Besoin d'un contrôle d'accès fiable pour 5 millions de clients, souscriptions et accords.
- Dépendances complexes entre les groupes, sociétés, individus, comptes, produits, souscriptions, services et accords

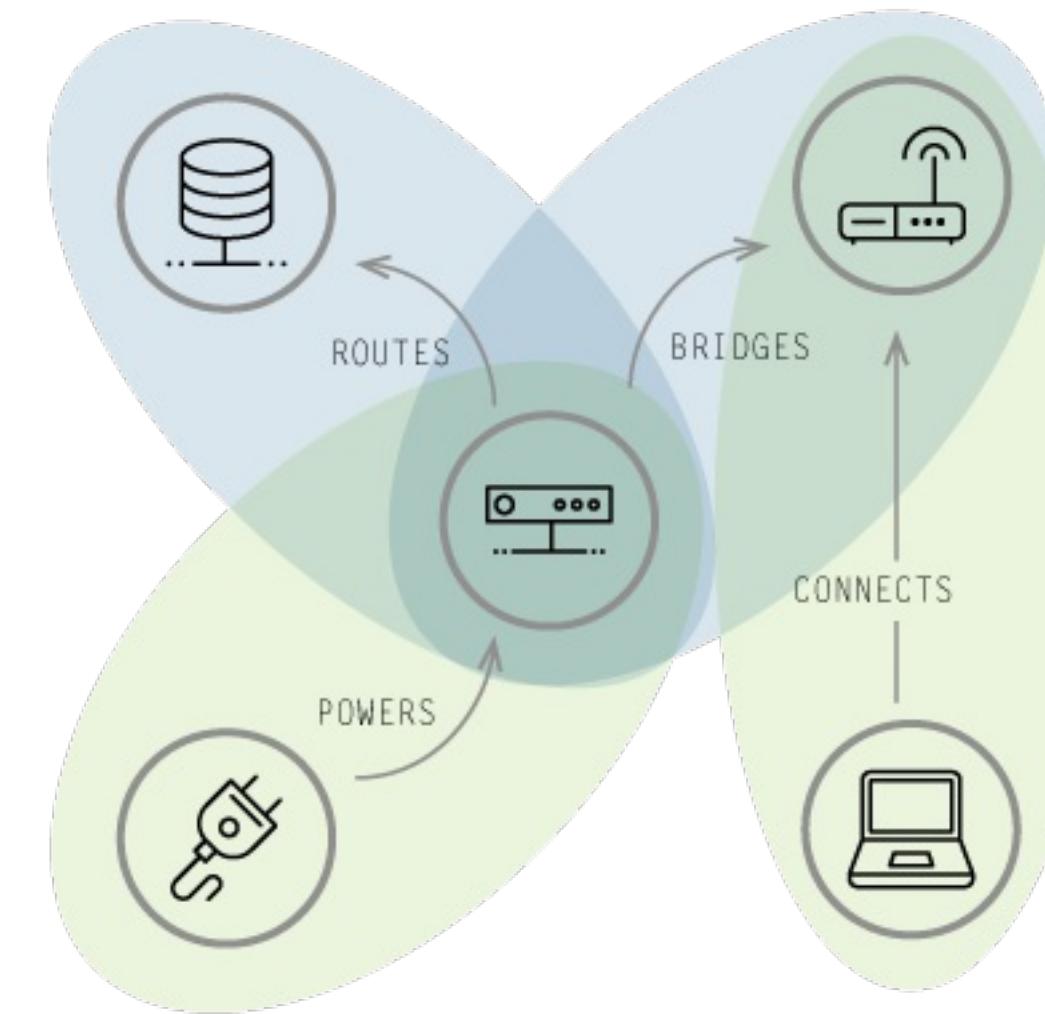
Bénéfices

- Architecture flexible et dynamique.
- Performance exceptionnelle.
- Bas coût comparé aux autres alternatives.
- Modèle de données extensible supportant l'ajout de nouvelles applications et fonctionnalités.

LES RÉSEAUX INFORMATIQUES

LES RÉSEAUX INFORMATIQUES

Un réseau est un graphe !



POURQUOI DU GRAPHE ?

- Modélisation/stockage naturelle
- Meilleur visualisation et apprehension des données
- Permet de réaliser des analyses d'impacts en cas de panne
- A l'inverse, voir les redondances du réseau

LES DÉFIS LIÉS AUX RÉSEAUX

Résolution des problèmes sur un réseau

Les interdépendances physiques et humaines sont extrêmement complexes dans un réseau ou un environnement informatique, ce qui rend difficile la résolution des problèmes.

Analyse d'impact

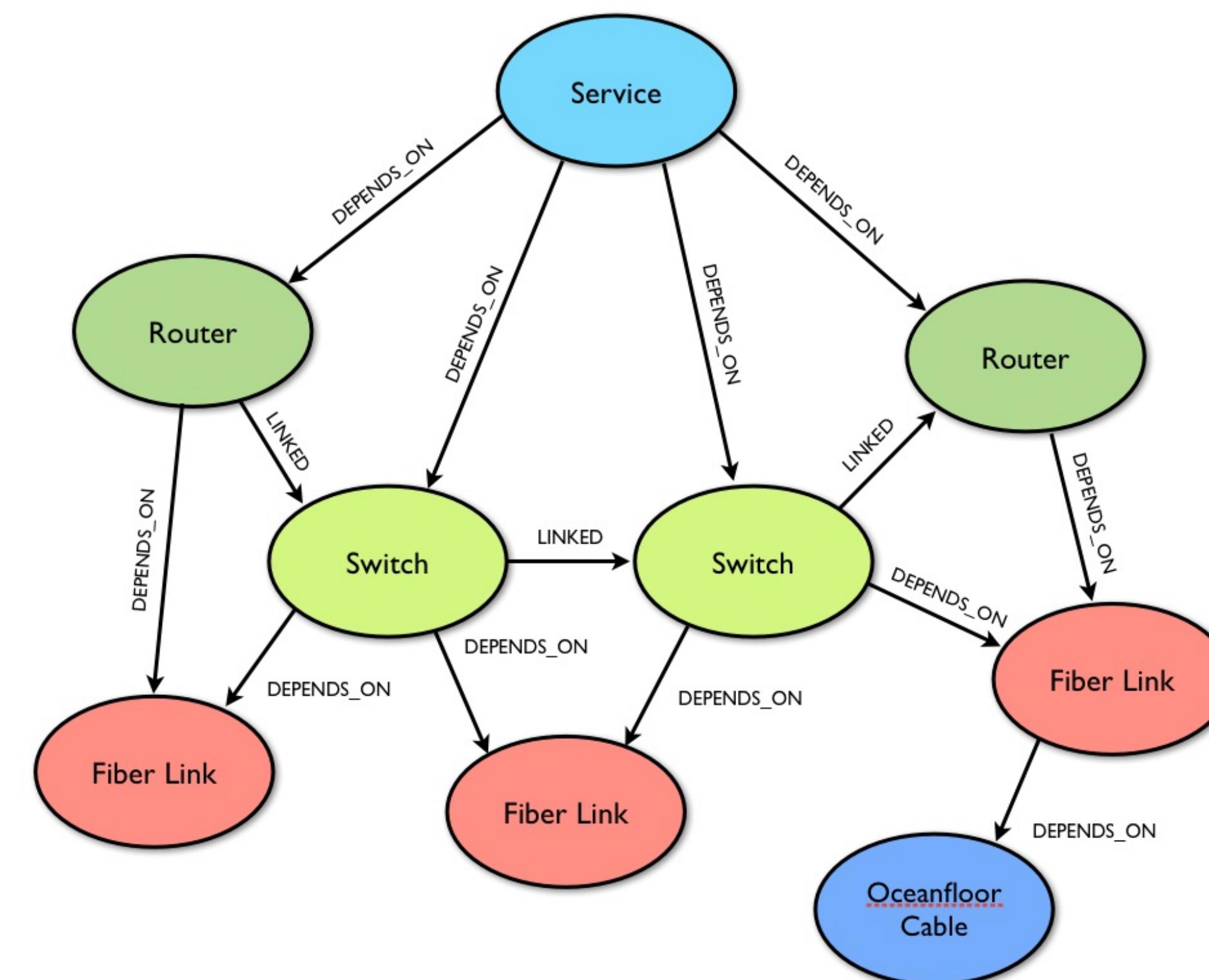
Les relations existant parmi les noeuds d'un réseau ne sont ni purement linéaires ni hiérarchiques et il est donc difficile de déterminer les interdépendances des éléments du réseau les uns envers les autres.

Nombre croissant de noeuds physiques et virtuels

Du fait de la croissance rapide dans la taille et le volume des données, le système doit prendre en compte à la fois les besoins actuels et futurs.

ETUDE DE CAS : ANALYSE D'IMPACT CHEZ SFR

Modélisation de leur réseau physique dans Neo4j.



ETUDE DE CAS : ANALYSE D'IMPACT CHEZ SFR

Problème

- Une semaine entière pour planifier une maintenance.
- Analyste d'impact sur de multiples systèmes (>30).
- Changements quotidiens dans l'infrastructure.
- Identifier les clients impactés par une maintenance.
- Identifier les faiblesses du réseau (Deuxième projet).
- Une semaine entière pour connaître le noeud de raccordement d'un nouveau client (Troisième projet).

ETUDE DE CAS : ANALYSE D'IMPACT CHEZ SFR

Bénéfice

- Identification temps réel des clients liées à une panne.
- Gestion d'inventaire réseaux très flexible.
- Une seule source de vérité (Neo4j) représentant le réseau entier.
- Système dynamique de chargement (ETL) des données issues de +30 sources variées.
- Mise en application rapide des changements métiers.

GRAPHE ARE EVERYWHERE !

GRAPHCONNECT EUROPE

20% de réduction avec le code NEWS20 !!!

Mardi 26 Avril, 2016, Queen Elizabeth II Centre, Londres

- 3 Thématiques + Lightning Talks
- GraphClinics: Rencontrez nos experts
- Networking

Formation, Lundi 25 Avril 2016, 9h30-17h30 : Fundamentals, Modelling, Neo4j in Production, Creating a Recommendation Engine



DES QUESTIONS ?

- **Twitter:** Suivez les comptes @neojFr & @neo4j
- **Google group :** Avec les groupes Neo4jFr & Neo4j
- **Stackoverflow :** avec les tags neo4j & cypher

