

Introduction à Neo4j

La base de données de graphes

Cédric.Fauvet@neotechnology.
com



Neo Technology



Société

- Neo Technology, Inc., fondée en 2007
- 100+ employés, QG dans la Silicon Valley; Bureaux à Londres, Munich et Malmö
- 160+ clients dont 60+ parmi les Global 2000

Communauté

- 27,000+ développeurs ont validé une formation Neo4j les 18 derniers mois.
- 25,000+ membres Meetup dans le monde dont 1100 à Paris.

Produit

- Neo4j – Leader mondial des bases de données de graphes
- Graphe natif et les meilleures performances
- 1M+ de téléchargement, +50k par mois

Clients & Partenaires

- Commerce: Walmart, Target, Adidas, eBay
- Telco: Telenor, Tre, CenturyLink, StarHub
- Finances: UBS, ICE ...
- 100+ partenaires technologiques et service, incluant Accenture, Ekino, Zenika, Sopra-Steria

Programme



- Quelle valeur pour les relations ?
- Les relations d'un point de vue SQL
- Les relations dans le monde NoSQL
- Les relations dans la base de données de graphes Neo4j
- Modélisation des données dans un graphe
- Cas d'usage
- Pourquoi choisir Neo4j ?



Programme



- **Quelle valeur pour les relations ?**
- Les relations d'un point de vue SQL
- Les relations dans le monde NoSQL
- Les relations dans la base de données de graphes Neo4j
- Modélisation des données dans un graphe
- Cas d'usage
- Pourquoi choisir Neo4j ?



Quelle valeur pour les relations ?



Le volume de données augmente...

- Nouveaux processus digitaux
- Augmentation des transactions en ligne
- Nouveaux réseaux sociaux
- Plus de périphériques

... et elles sont de plus en plus connectées

Clients, produits, processus, matériel interagissent et dépendent les uns des autres

Les relations ont une valeur à débloquent

- Recommandations temps-réel (TR)
- Détection de Fraude TR
- Gestion des données de référence
- Gestion des réseaux

Premiers utilisateurs

- Moteur de recherche basé sur les graphes

LinkedIn **Google** **facebook**



Programme



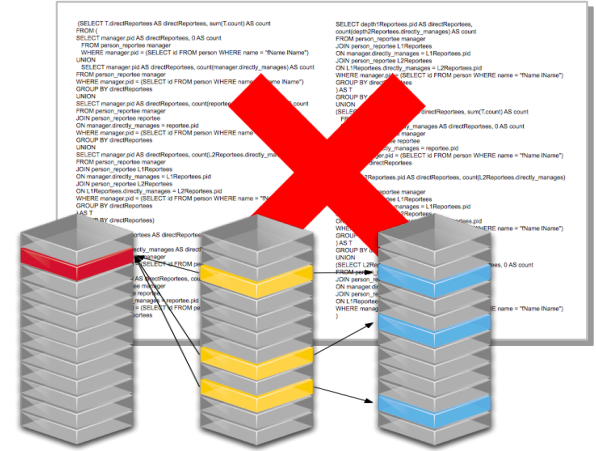
- Quelle valeur pour les relations ?
- **Les relations d'un point de vue SQL**
- Les relations dans le monde NoSQL
- Les relations dans la base de données de graphes Neo4j
- Modélisation des données dans un graphe
- Cas d'usage
- Pourquoi choisir Neo4j ?



Les bases de données relationnelles ne gèrent pas bien les relations



- *Impossible de modéliser les données ET les relations* sans complexité
 - *Les performance se dégradent* avec les relations nombreuses, profondes et la taille de la base
 - *La complexité des requêtes augmente* avec la nécessité des Jointures
 - *L'ajout de nouveaux types de données et relations* nécessite une redéfinition du modèle, augmentant les délais de mise sur le marché
- ... et rendent les bases de données traditionnelles **inappropriées** lorsque les relations sont exploitées **en temps-réel**



Développement lent
Faibles performances
Passage à l'échelle limité
Difficile à maintenir

Programme



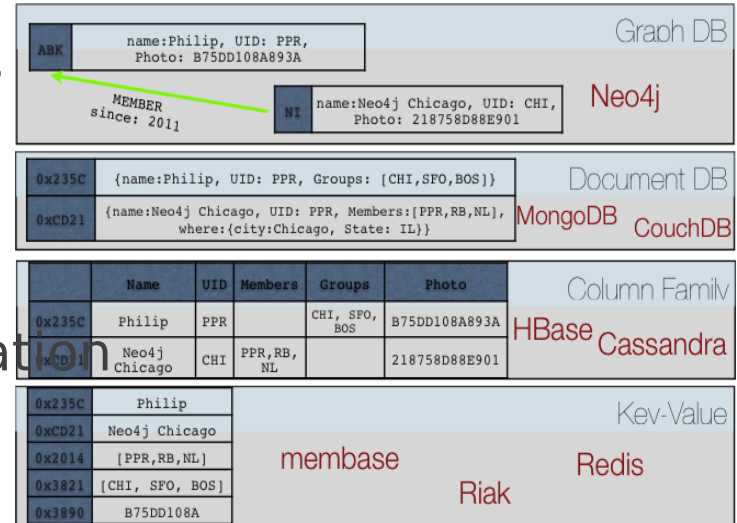
- Quelle valeur pour les relations ?
- Les relations d'un point de vue SQL
- **Le relations dans le monde NoSQL**
- Les relations dans la base de données de graphes Neo4j
- Modélisation des données dans un graphe
- Cas d'usage
- Pourquoi choisir Neo4j ?



Les bases de données NoSQL ne gèrent *pas* les relations



- **Pas de structures de données** pour modéliser ou sauvegarder les relations
- **Langage de requête inadapté** pour supporter les relations des données
- **Relier les données nécessite une "Logique de jointure"** dans l'application
- **Pas d'acidité** dans les transactions



... et rendent les bases de données NoSQL **inappropriées** lorsque les relations sont exploitées **en temps-réel**



Programme



- Quelle valeur pour les relations ?
- Les relations d'un point de vue SQL
- Les relations dans le monde NoSQL
- **Les relations dans la base de données de graphes Neo4j**
- Modélisation des données dans un graphe
- Cas d'usage
- Pourquoi choisir Neo4j ?

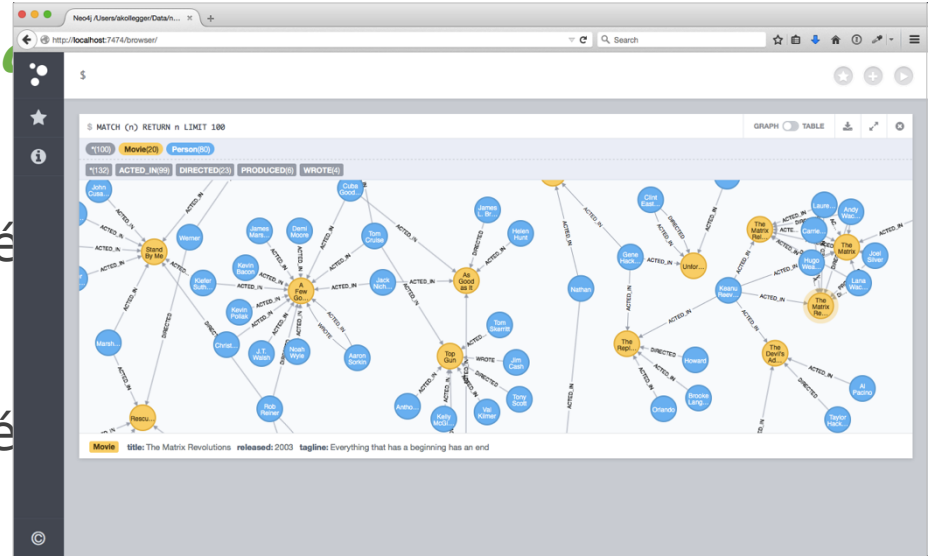


Neo4j - Re-Imagine vos données dans un graphe



Neo4j est *une base de données de graphes d'entreprise* qui vous permet de:

- **Modéliser et stocker** vos données dans un graphe
- **Requêter les données de relations** facilement en temps réel
- **Faites évoluer facilement vos applications** par l'ajout de nouvelles fonctionnalités, données et relations



Développement Agile
Haute performance
Echelonnement Vertical
et Horizontal



Neo4j: La base de données de graphes native



Stockage natif en Graphe

Garanti la persistance et la performance

Moteur de Graphes natif

Plusieurs Millions de 'jointures' par seconde

Modélisation 'sur tableau blanc'

Modélisation intuitive et naturelle

Cohérence des données

Langage de requêtes adapté aux 'Jointures'

Nécessite 10 à 100 fois moins de code que SQL

Mise à l'échelle et haute disponibilité

Echelonnement vertical et horizontal adapté aux graphes

ETL Intégré

Migration et import sans effort

Intégration

Prix & reconnaissance du marché et bibliothèques pour les langages



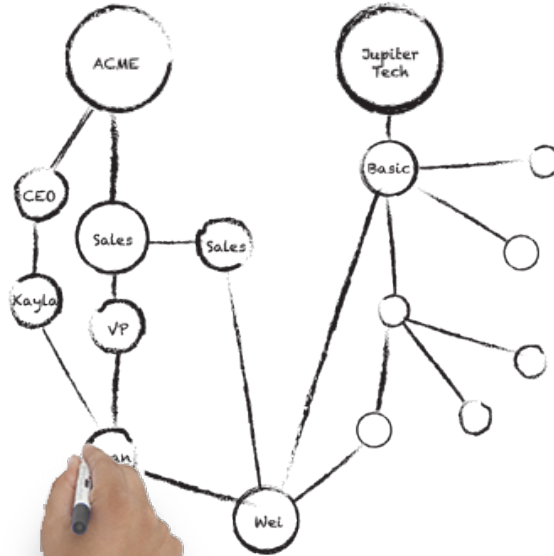
Programme



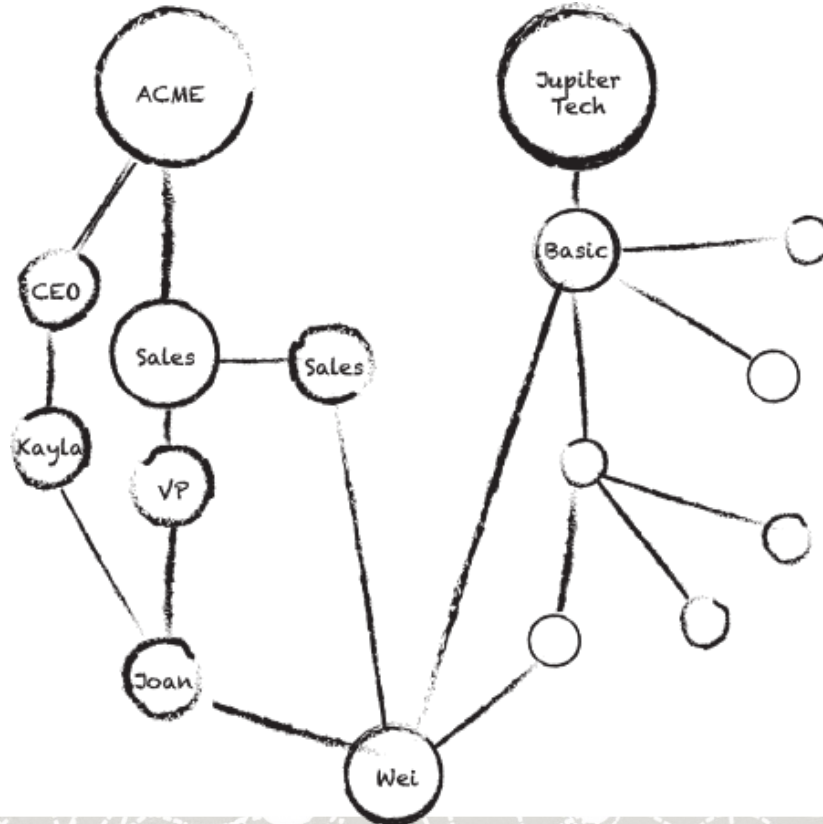
- Quelle valeur pour les relations ?
- Les relations d'un point de vue SQL
- Les relations dans le monde NoSQL
- Les relations dans la base de données de graphes Neo4j
- **Modélisation des données dans un graphe**
- Cas d'usage
- Pourquoi choisir Neo4j ?



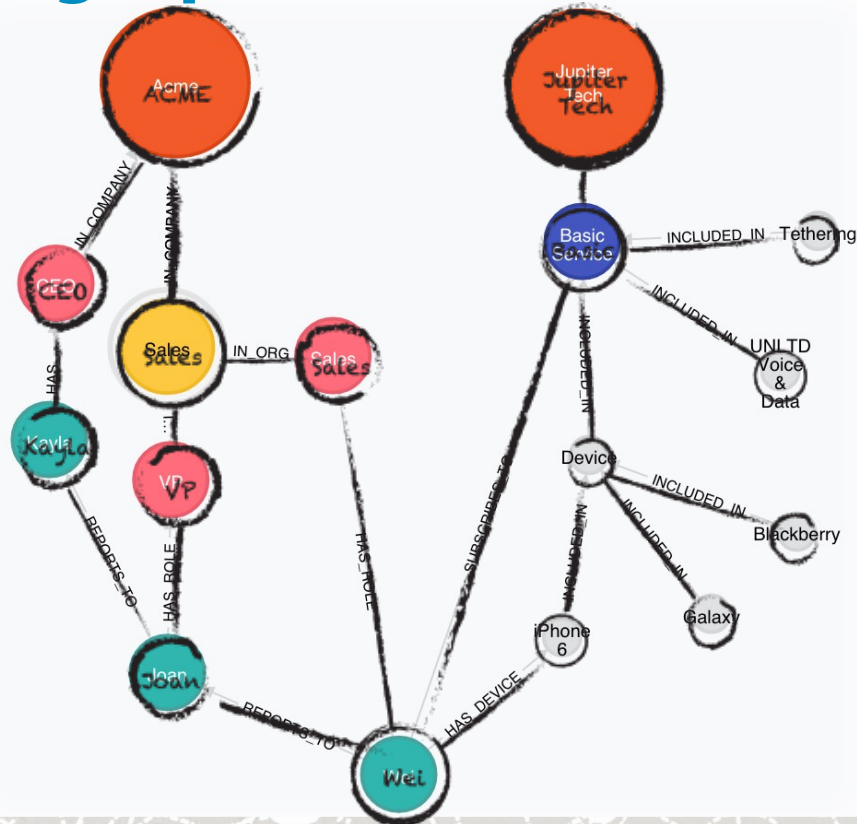
Modèle du tableau blanc



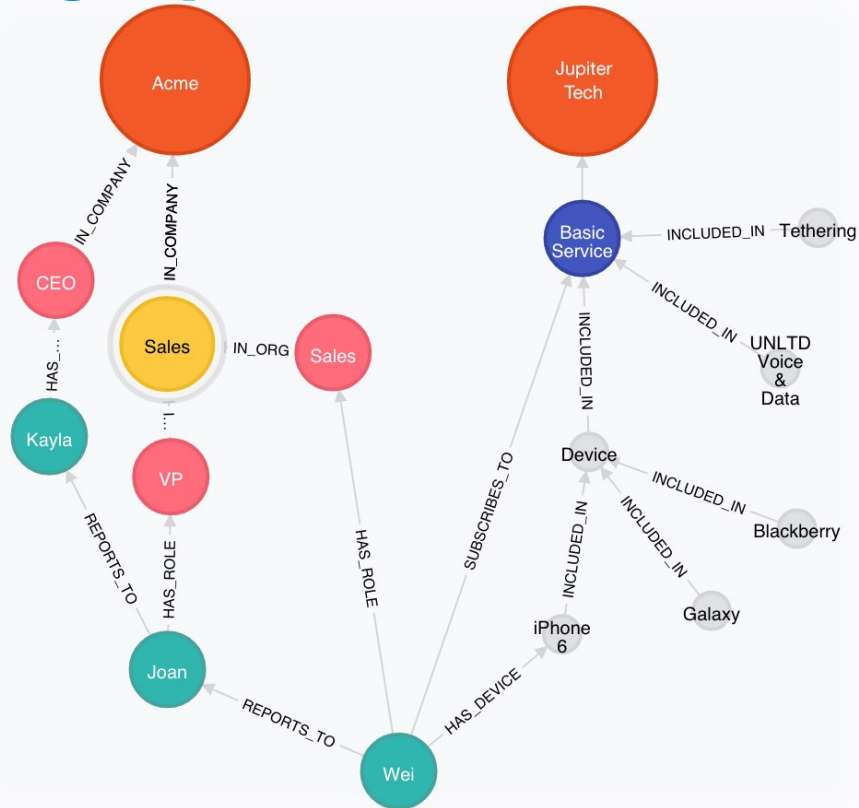
Modèle du tableau blanc



Modèle du tableau blanc = Modèle physique de graphes



Modèle du tableau blanc = Modèle physique de graphes



Programme



- Quelle valeur pour les relations ?
- Les relations d'un point de vue SQL
- Les relations dans le monde NoSQL
- Les relations dans la base de données de graphes Neo4j
- Modélisation des données dans un graphe
- **Cas d'usage**
- Pourquoi choisir Neo4j ?



Adoption de Neo4j par industrie



Finance



Télécom



Santé



RH



Média



Web Social



Industrie & Logistique



Jeux



Commerce



Business Services



Information Services



Usage: Réseau social

Moteur de recommandation

Contexte

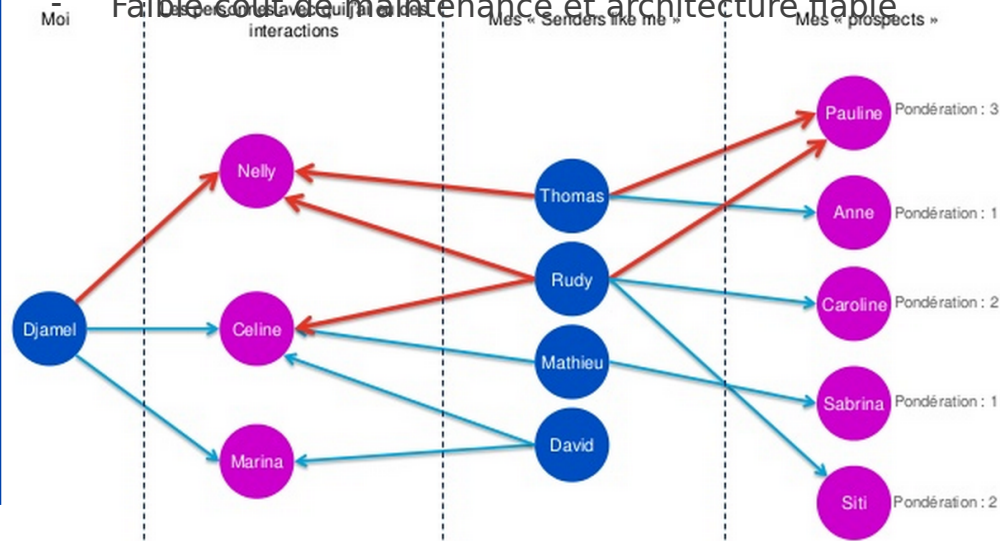
- Leader Européen des sites de rencontre.
- 126,1 million euro/an
- 1 million de membres

Problème métier

- A l'origine réalisé par un traitement batch écrit en PHP
- Isoler et traiter les cas des super-dragueurs et super-dragués.
- Recommandations précise afin de fidéliser les utilisateurs et attirer de nouveaux membres.

Solution & Bénéfices

- Expérience utilisateur personnalisée au maximum
- Solution scalable et simple à mettre en place
- Faible coût de maintenance et architecture fiable



Usage: eCommerce et Marketing

Moteur de recommandation



Contexte

- En position 36 des Fortune 500
- 1,916 magasins aux US et Canada
- Compétition directe avec Kmart and Walmart

Problème métier

- Manque des ventes à cause d'une recommandation non personnalisée
- Processus de création des promotions trop long (+1 jour) avec fort risque d'erreurs
- Remplacement de IBM WebSphere Commerce

Solution & Bénéfices

- Expérience client basée sur les promotions, parcours client, stock et autre informations du profil
- Neo4j permet d'établir les promotions en fonction de l'activité spécifique de chaque utilisateur
- Calcul automatique du prix en fonction des promotions

- Amélioration du **chiffre d'affaire** eCommerce de **+3 à +5%** !

Usage: Marketing

Gestion de référentiel (MDM) et Recommandation



Contexte

- Fabricant de chaussures et d'articles de sports Allemand, Multinationale.
- Fondée en 1949
- Emploie plus de de 53 000 personnes dans 160 pays
- Produit plus de 660 million de produits par an
- Chiffre d'affaire de 14.5 milliards en 2014

Problème métier

- Les applications internes communiquent peu et difficilement du fait de la complexité du référentiel existant
- Le contenu poussé aux visiteurs des plateformes web est standard par manque d'accès au référentiel de métadonnées

Solution & Bénéfices

- Un seul référentiel centralisé, une seule version de la vérité
- Recommandation personnalisée pour les visiteurs des différents sites internet de Adidas

Usage: Marketing

Gestion de référentiel (MDM)

Contexte

Adobe est le leader mondial des solutions multimédias et de marketing numérique.

- **Problème métier**
Volume de données important lié aux membres, groupes d'utilisateurs, contenu des membres - le tout massivement interconnecté
- Besoin d'en déduire les relations de collaboration basé sur les contenu utilisateurs.

Solution & Bénéfices

- Cluster de Neo4j Enterprise
- Composant d'une plus grande infrastructure
- Déploiement Multi-régions AWS
- Neo4j sélectionné sur une solution custom et Oracle
- Analyse des données hautement flexible
- Résultats temps réels malgré des données hautement interconnectées
- Expérience de l'utilisateur - Avantage compétitif
- Durée du projet : 12 mois



Usage: Logistique

Parcours de graphes, calcul de tournées

accenture



Contexte

- Opérateur historique de logistique d'un grand pays EU.
- +5000 colis par seconde dans les pics.
- Routage temps-réel : 8 Millions par jour !
- Suivi des colis pour les clients B2C & B2B.

Problème métier

- Système critique en ligne 24x7.
- Capacité du système historique dépassée d'ici peu.
- Nombreux logiciels en interaction complexe.
- Performance attendue : Linéaire et scalable.
- Changements quotidiens des routes possibles.

Solution & Bénéfices

Nouveau système de routage = Graphe = Neo4j.

- Plus simple : Cypher Vs SQL
- Plus Flexible : Reflet du monde réel.
- Plus rapide et hautement disponible :
Cluster Neo4j
- Une seule « source de vérité » pour tout le réseau.



Usage: Télécom

Gestion de droits d'accès (ACL)

Contexte

- Opérateur téléphonique leader dans les pays nordiques
- Besoin d'un contrôle d'accès fiable pour 5 millions de clients, souscriptions et accords.

Problème métier

- Dépendances complexes entre les groupes, sociétés, individus, comptes, produits, souscriptions, services et accords
- Graphs globaux et profonds (Client principaux avec 1000 fils, souscriptions et accords)



Solution & 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.

- Neo4j solution entreprise.
- Embarqué et haute disponibilité.
- En remplacement de bases de données Oracles
- vieilles de 10 ans, Berkeley DB et un environnement mainframe.

Usage: Commerce et CRM

Gestion de données de référence client



Contexte

- Leader mondial de l'infrastructure réseau
- Vaste organisation commerciale

Problème métier

- Règles complexes régissant la propriété des comptes.
- Règles complexes de calcul des commissions
- Requêtes complexes dues à la structure des RDBMS
- Performances de la base de données Oracle insuffisante pour une gestion temps réel des

Solution & Bénéfices

- Vue temps réel sur les comptes et leurs propriétaires
- Capacité de modéliser des règles complexes pour la gestion de la propriété des comptes
- Calcul des commissions automatique pour toute l'organisation commerciale
- Environ 12 mois de développement
- 2 clusters Neo4J hautement disponibles
- Un cluster en production un second en sauvegarde dans des Datacenter différents
- Au total : 6 bases de données Neo4J Enterprise embarquées

Usage: Réseau social

Détection de Fraude

Contexte

- Plus grand gestionnaire de jeux sociaux d'argent en ligne d'Angleterre et d'Europe
- Créateur du premier jeu d'argent sur Facebook
- Bingo, Machine à sous et Casino en ligne
- 5 millions de parieurs par an, 70 000€

Problème métier

- Augmenter la durée de moyenne de la visite
- SQL trop lent pour exécuter les millions de jointures nécessaires à la détection de Fraude
- Réduire les phases de tests, par des simulations dans le graphe



gamesys



Solution & Bénéfices

- Fidéliser les clients par l'ajout de couches sociales
- Réduire les fraude par l'analyse de relations de parrainage en temps réel et analytique
- Temps de développements plus courts que prévu
- Pas d'interruption de service les 18 derniers mois



Usage: Télécom

Analyse d'impact



Contexte

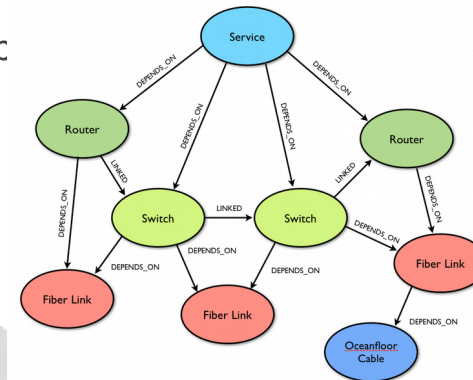
- Deuxième opérateur Télécom Français.
- 12 Milliard de CA en 2012.
- 165 000 Clients entreprise.
- 250 Clients opérateurs.

Problème métier

- 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.

Solution & Bénéfices

- 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 œuvre de changements métiers.



Programme



- Quelle valeur pour les relations ?
- Les relations d'un point de vue SQL
- Les relations dans le monde NoSQL
- Les relations dans la base de données de graphes Neo4j
- Modélisation des données dans un graphe
- Cas d'usage
- **Pourquoi choisir Neo4j ?**

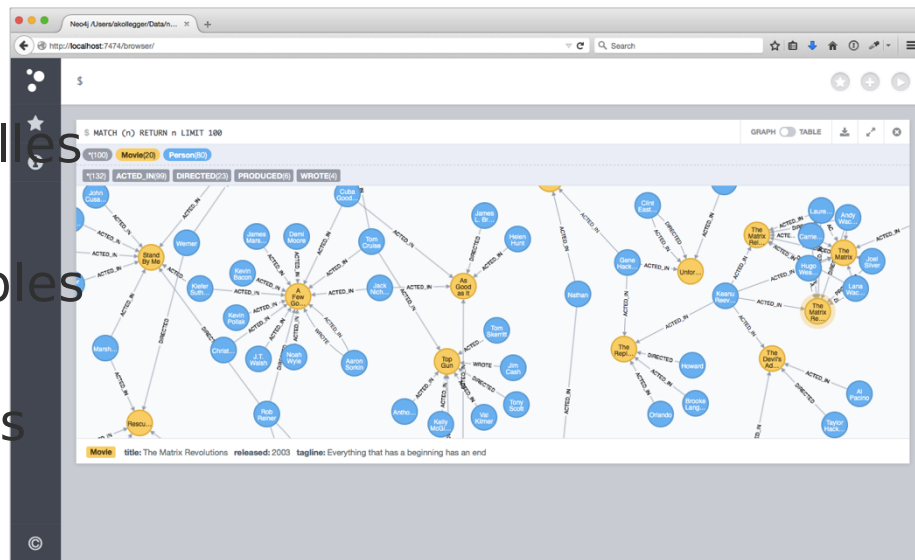


Pour conclure



Les trois grandes raisons de choisir Neo4j :

- Les performances exceptionnelles
- Des développements plus simples
- La découverte de nouveaux cas d'usage et la valorisation des relations



Merci

[cedric@neotechnolo](mailto:cedric@neotechnology.com)

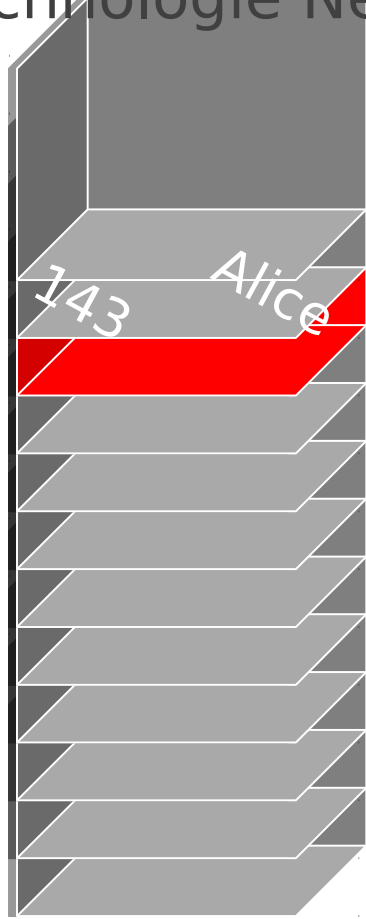
gy.com

Twitter:

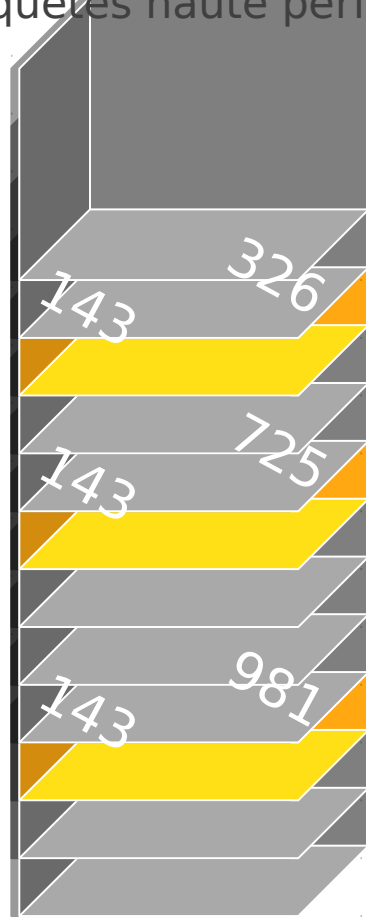
@CedricFauvet



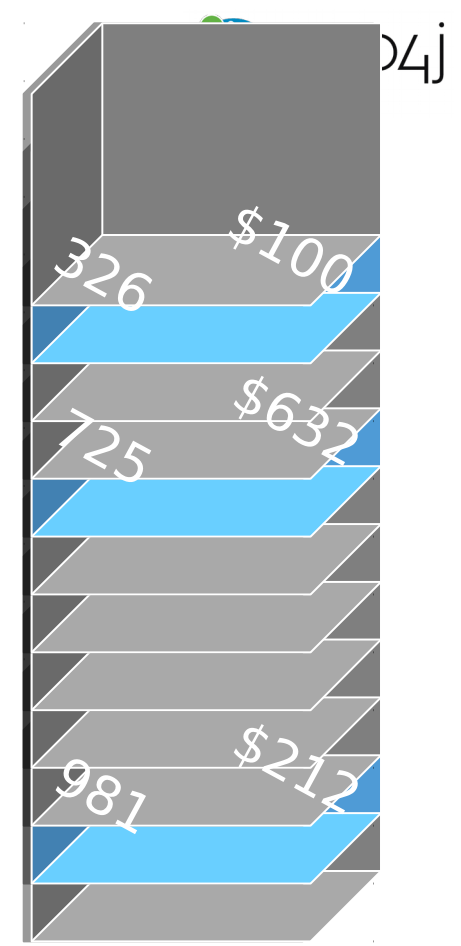
Technologie Neo4j, des requêtes haute performance



Customers

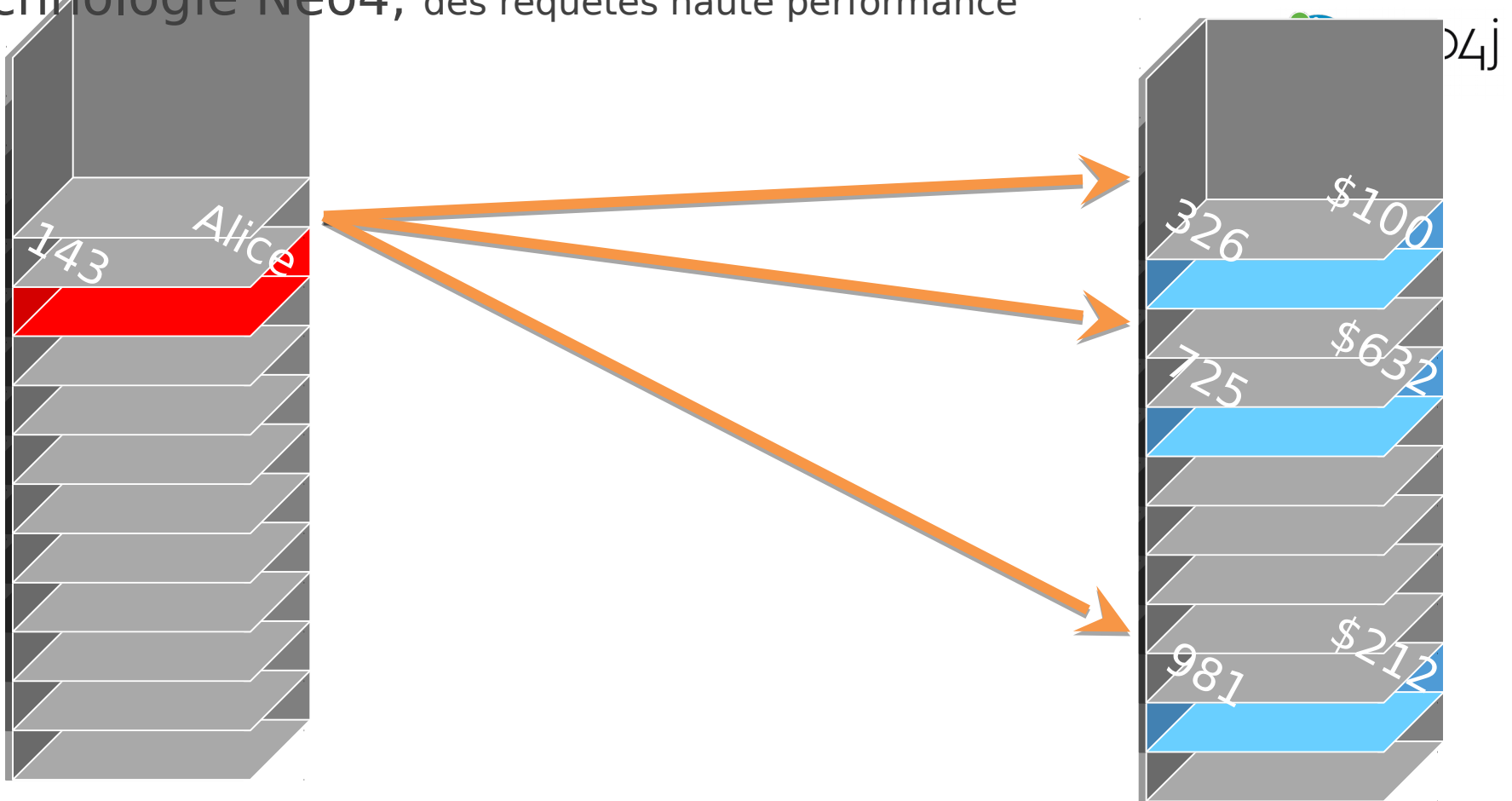


Customers_Accounts



Accounts

Technologie Neo4j, des requêtes haute performance

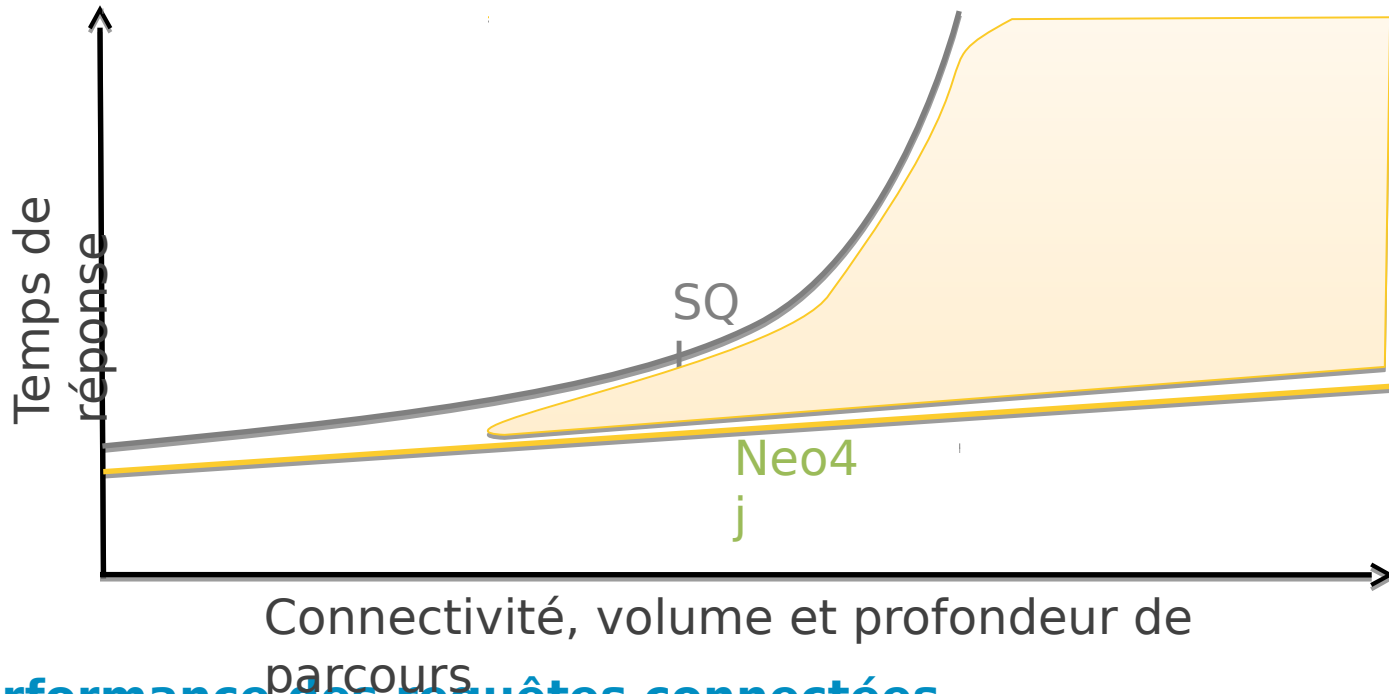


Customers

Accounts

Technologie Neo4j

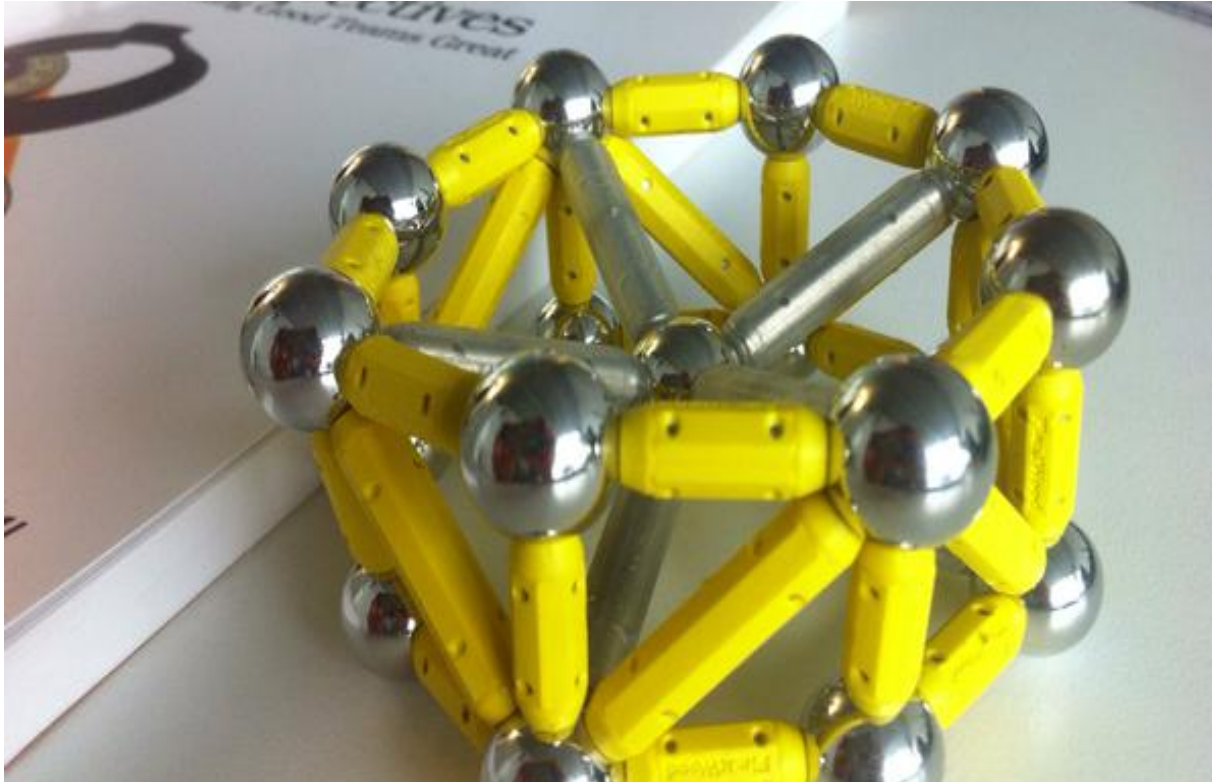
Des requêtes haute performance



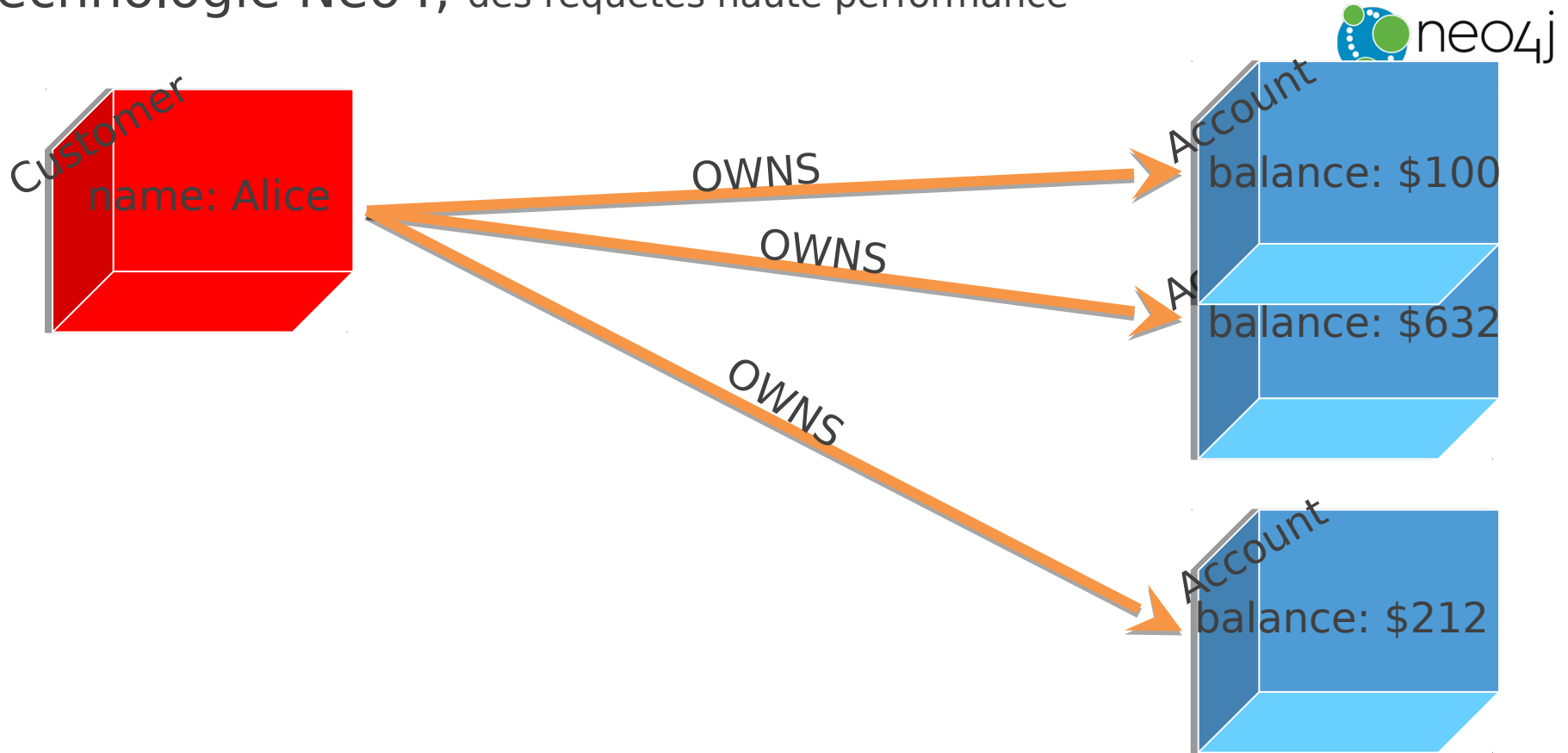
Performance des requêtes connectées



Les relations en jaune sont persistées dans Neo4j

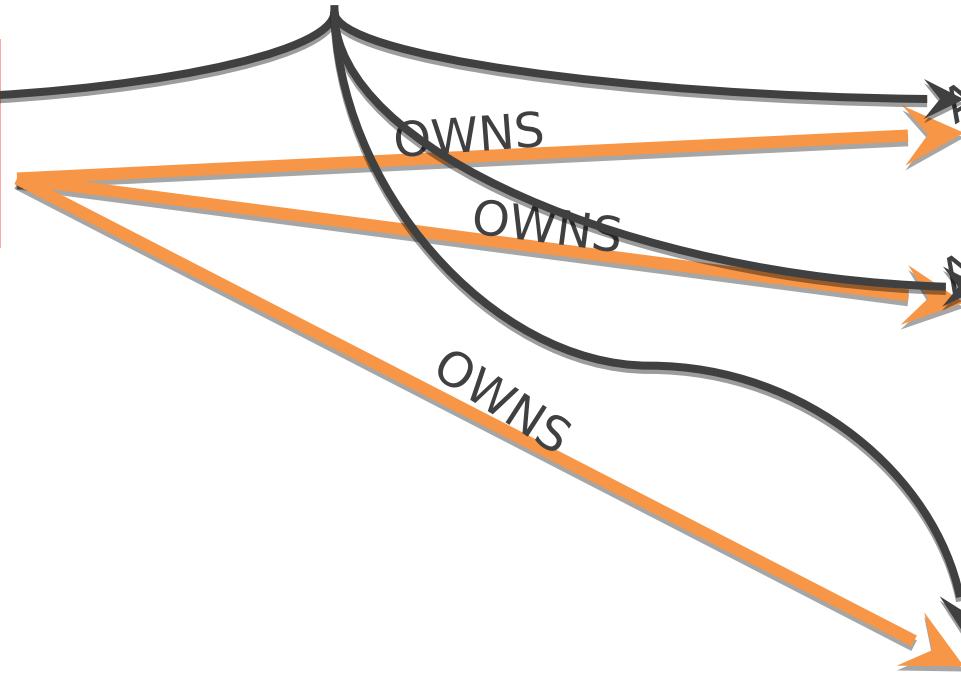
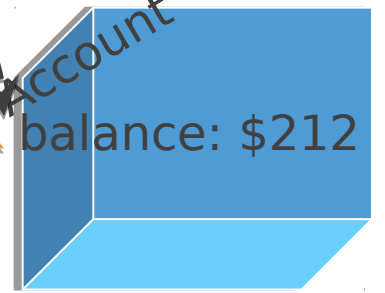
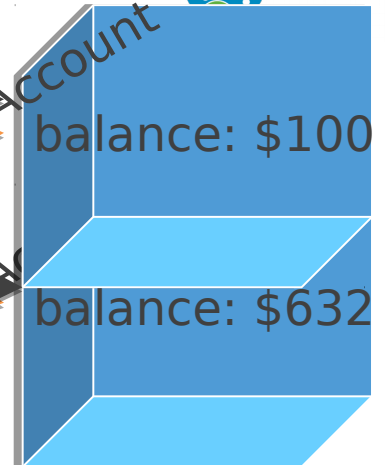
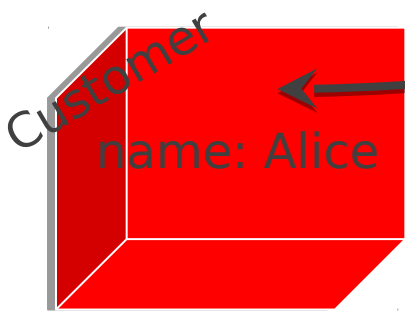


Technologie Neo4, des requêtes haute performance

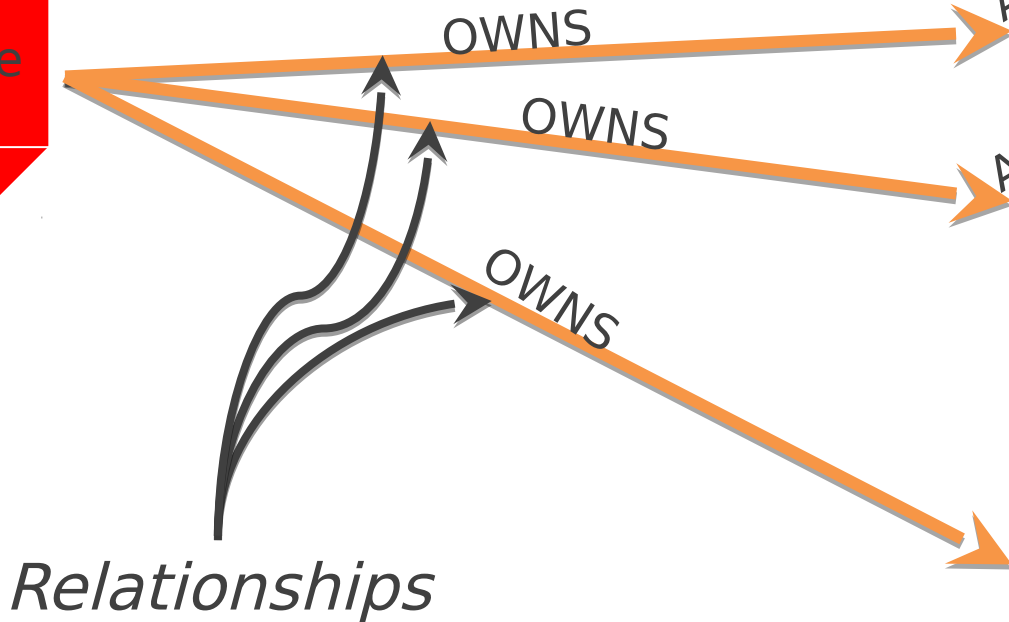
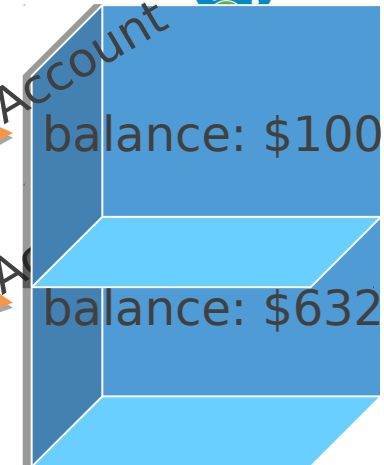
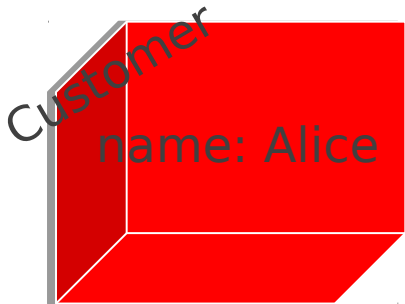


Technologie Neo4, des requêtes haute performance

Nodes

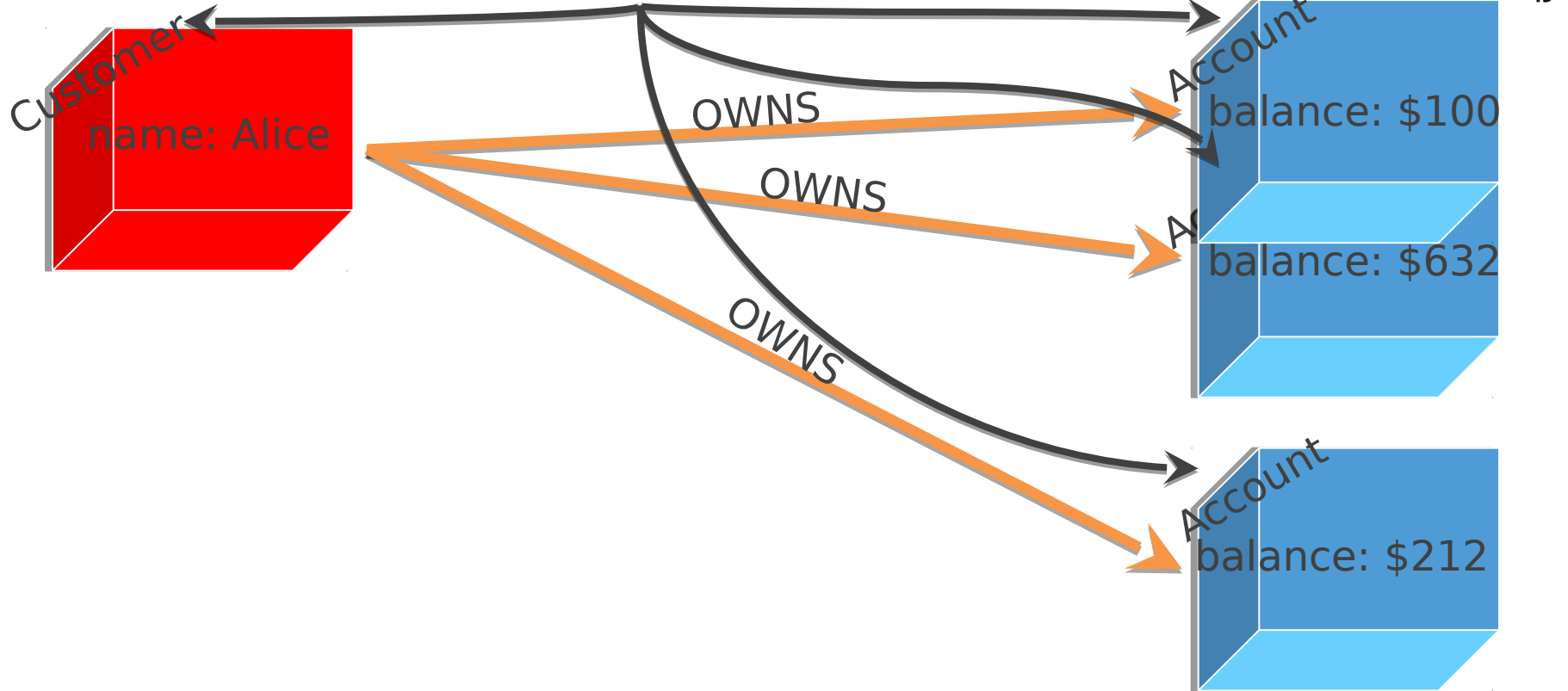


Technologie Neo4, des requêtes haute performance



Technologie Neo4, des requêtes haute performance

Labels



Technologie Neo4j

Des projets plus simples avec Cypher

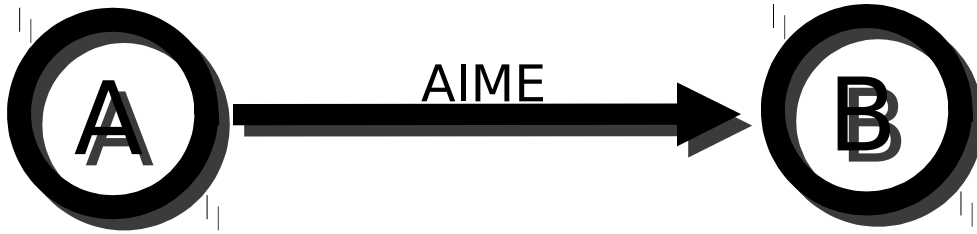


(A) --> (B)



Technologie Neo4j

Des projets plus simples avec Cypher

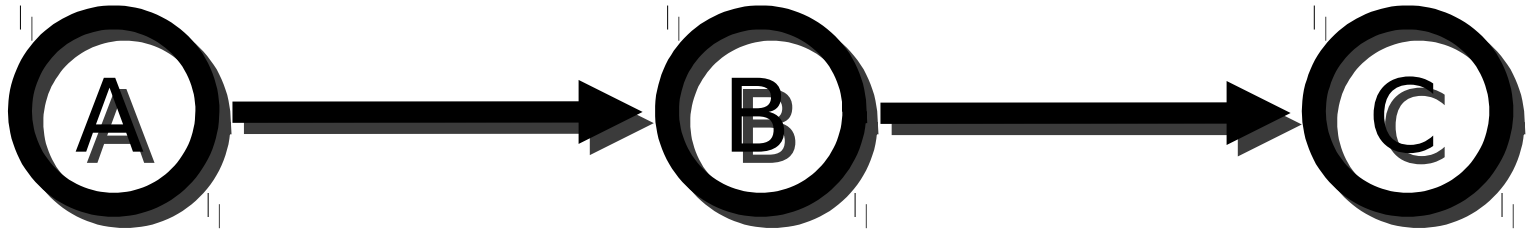


A -[:AIME]-> B



Technologie Neo4j

Des projets plus simples avec Cypher

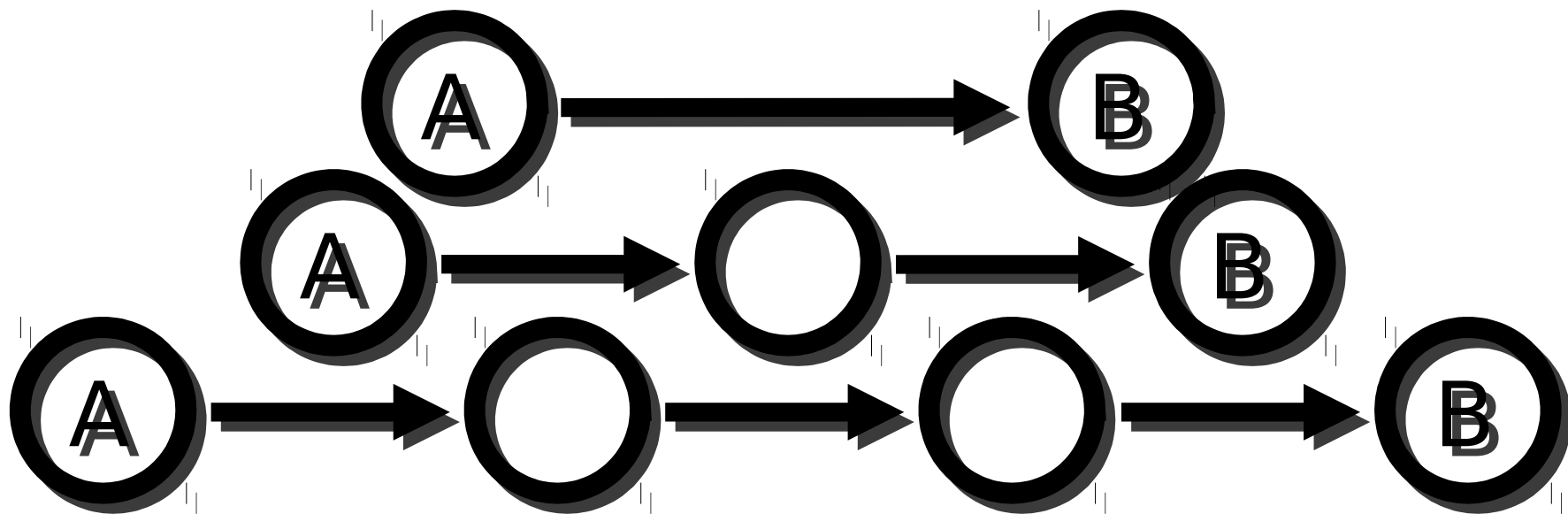


A --> B --> C



Technologie Neo4j

Des projets plus simples avec Cypher

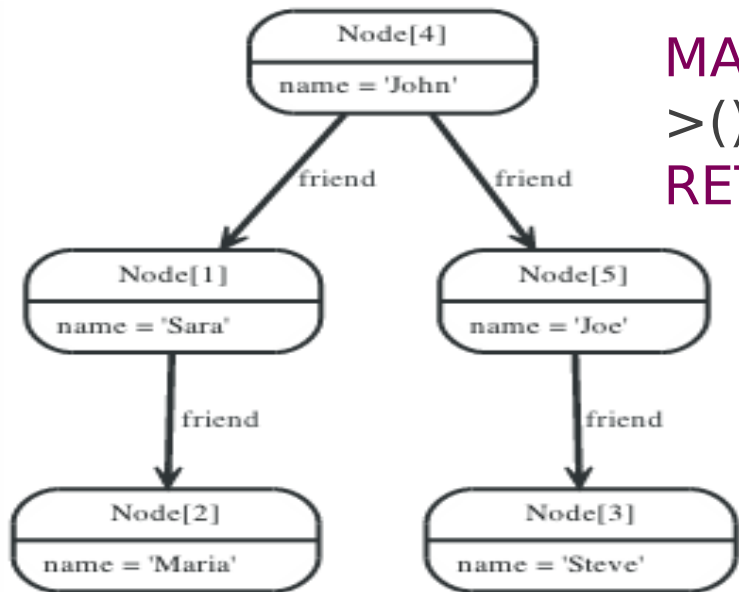


A -[*]-> B



Technologie Neo4j

Des projets plus simples avec Cypher



```
MATCH (VARjohn {name: 'John'})-[:friend]->()-[:friend]->(VARfof)  
RETURN VARjohn, VARfof
```

john	fof
Node[4]{name:"John"}	Node[2]{name:"Maria"}
Node[4]{name:"John"}	Node[3]{name:"Steve"}
2 rows	
2 ms	