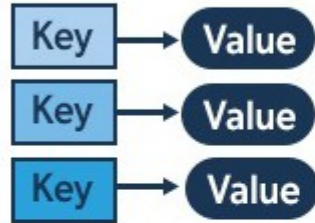
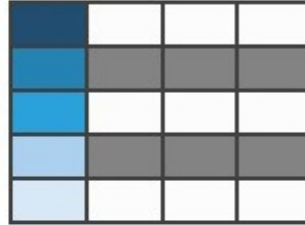


NoSQL

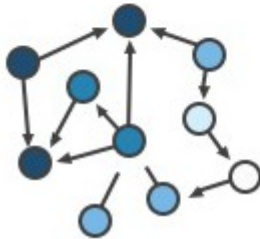
Key-Value



Column-Family



Graph



Document



Introduction au No SQL

Big Data

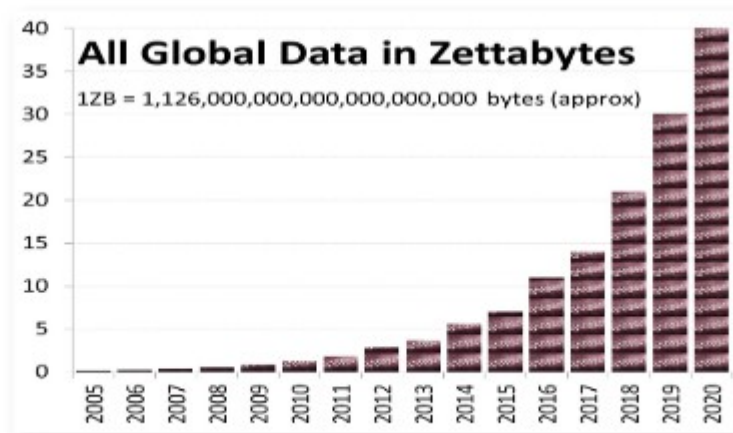
- Introduction à cette formation
 - Votre formateur ... Et Vous
 - Le matériel
 - Le support de cours
 - La distribution Anaconda
 - L'organisation – horaires
 - Formation de 3 jours
 - La forme : cours, exercices et TP

Big Data

- Des liens utiles
 - https://chewbii.com/wp-content/uploads/2015/11/NoSQL_intro_handout.pdf

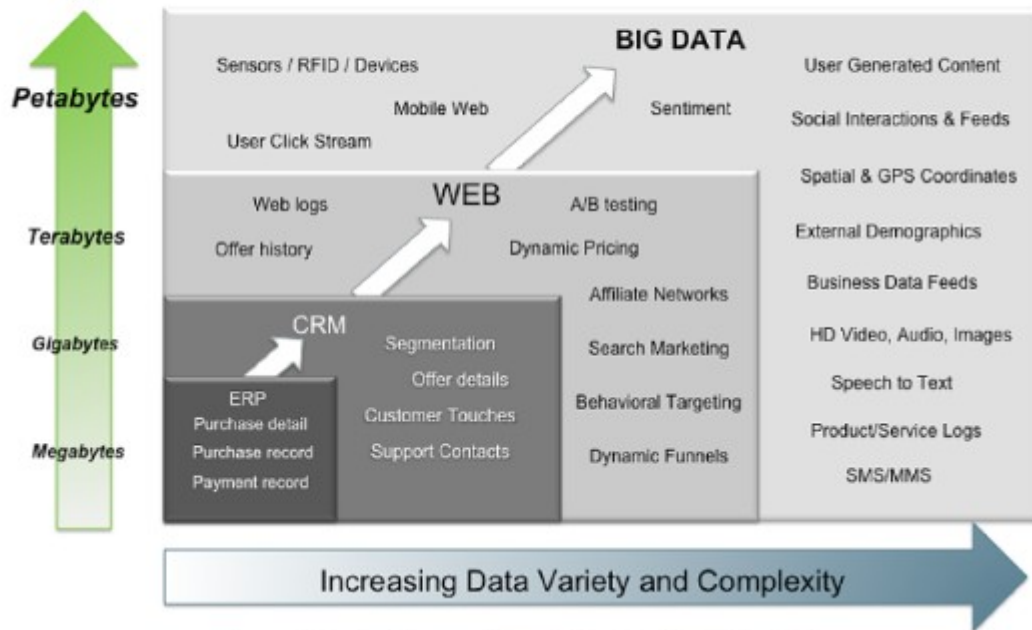
- Aperçu des bases NoSQL
- MongoDB
- MongoDB et Python

- La quantité de données digitales produites double **tous les 2 ans**.
- En d'autres termes, on a produit autant de données digitales ces 2 dernières années que tout ce qui a été produit auparavant.

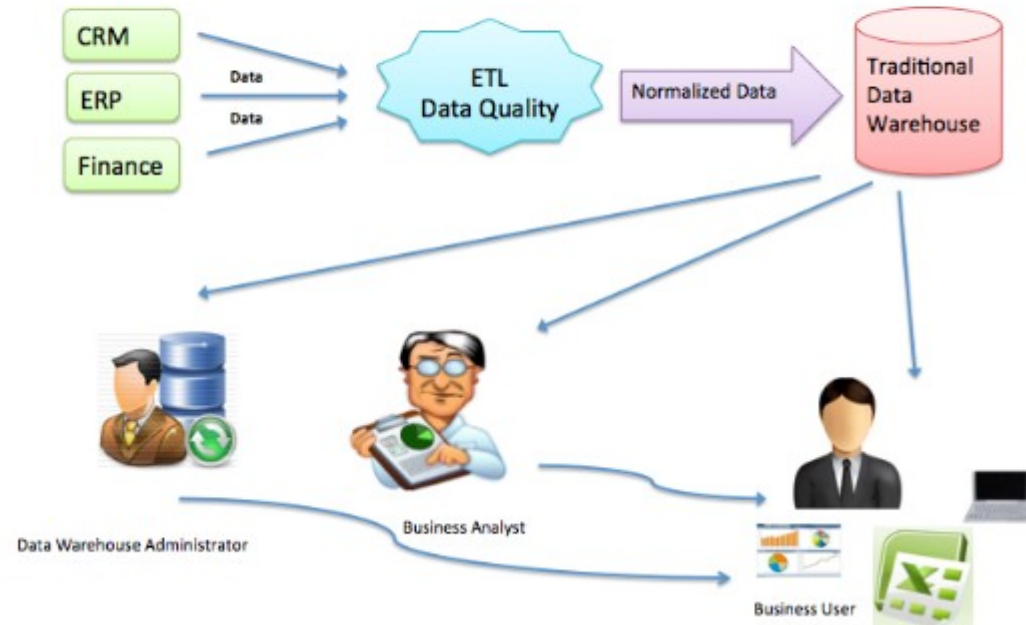


Volume et Variété

Big Data = Transactions + Interactions + Observations



Décisionnel : ancienne méthode



Décisionnel vs 3V

- L'approche classique incompatible avec les 3V du BigData :
- Le **Volume**: les entrepôts sont conçus pour gérer des Go ou To de données alors que la croissance exponentielle des données nous conduit aux Po ou Eo
- Le **type (Variety)**: le nombre de types, incluant les données textuelles semi ou non structurées, augmente
- La **vitesse (Velocity)**: les données sont créées de plus en plus vite et nécessitent des traitements en temps-réel

SGBDR vs Distribution

- **Fonctionnalités**
 - Jointures entre les tables
 - Langage d'interrogation riche
 - Contraintes d'intégrité solides
- **Limites dans le contexte distribué :**
 - **Comment distribuer/partitionner les données**
 - Liens entre entités -> Même serveur
 - Mais plus on a de liens, plus le placement des données est complexe

SGBDR vs Distribution

ACID vs BASE

- Propriétés **ACID** pour les transactions

- Atomicité** : une transaction s'effectue entièrement ou pas du tout
- Cohérence** : le contenu d'une base doit être cohérent au début et à la fin d'une transaction
- Isolation** : les modifications d'une transaction ne sont visibles/modifiables que quand celle-ci a validé
- Durabilité** : une fois la transaction validée, l'état de la base est permanent (non affecté par les pannes ou autre)

- Systèmes distribués : modèle **BASE**

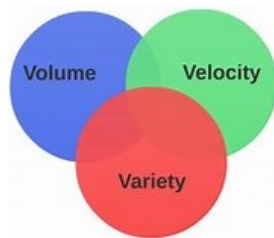
- Basically Available** : garantie minimale pour taux de disponibilité face grande quantité de requêtes
- Soft-state** : l'état du système peut changer au cours du temps même sans nouveaux inputs (cela est du au modèle de consistance).
- Eventually consistent** : tous les réplicas atteignent le même état, et le système devient à un moment consistant, si on stoppe les inputs



NoSQL : une solution

- **NoSQL : Not Only SQL**
 - Nouvelle approche de stockage et de gestion de données
 - Permet le passage à l'échelle via un contexte hautement distribué
 - Gestion de données complexes et hétérogènes
 - Pas de schéma pour les objets
- **Ne remplace pas les SGBDR !!**
 - Quantité de données énorme (PétaBytes)
 - Besoin de temps de réponse
 - Cohérence de données faible

-
- Not Only SQL : il n'y a pas que le SQL
- Face aux 3V – Volume, Velocity, Variety – le modèle relationnel n'est pas adapté



variety signifie ici la diversité des données vs rigidité d'un modèle SQL

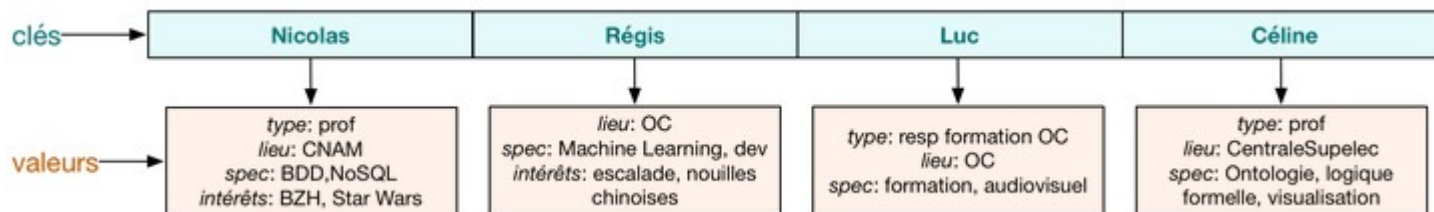
- Big Data

Les types de BD NoSQL

Les types de BD NoSQL :

- Les clés-valeurs : simple et efficace, sorte de grande table de hachage distribuée sur un réseau, entre N serveurs.

Aucun schéma, aucune structure, pas de langage SQL



SimpleDB (Amazon)
Azure Cosmos DB (Microsoft)
...

- Big Data

Les types de BD NoSQL

- Stockage orienté colonnes

Stockage orienté lignes				
id	type	lieu	spec	intérêts
Nicolas	prof	CNAM	BDD, NoSQL	BZH, Star Wars
Régis		OC	Machine Learning, Dev	escalade, nouilles chinoises
Luc	resp formation OC	OC	formation, audiovisuel	
Céline	prof	CentraleSupelec	Ontologie, logique formelle, visualisation	

Stockage orienté colonnes							
id	type	id	lieu	id	spec	id	intérêts
Nicolas	prof	Céline	Centrale Supelec	Nicolas	BDD	Nicolas	BZH
Céline	prof	Nicolas	CNAM	Nicolas	NoSQL	Nicolas	Star Wars
Luc	resp formation OC	Régis	OC	Régis	Machine Learning	Régis	escalade
		Luc	OC	Régis	Dev	Régis	nouilles chinoises
				Luc	formation		
				Luc	audiovisuel		
				Céline	Ontologie		
				Céline	logique formelle		
				Céline	visualisation		

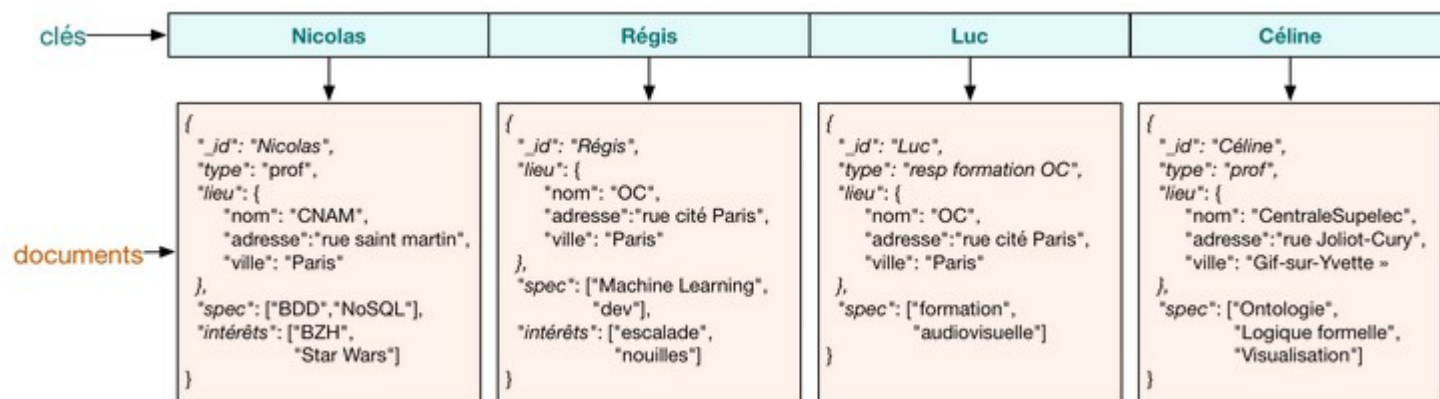
- Adapté pour des traitements sur les colonnes (somme, moyenne,...)
- BigTable (Google), Hbase (Apache, Hadoop), Elasticsearch (elastic)

- Big Data

Les types de BD NoSQL

- Bases orientées documents

- Repose sur le type clé/valeur avec un document pour valeur

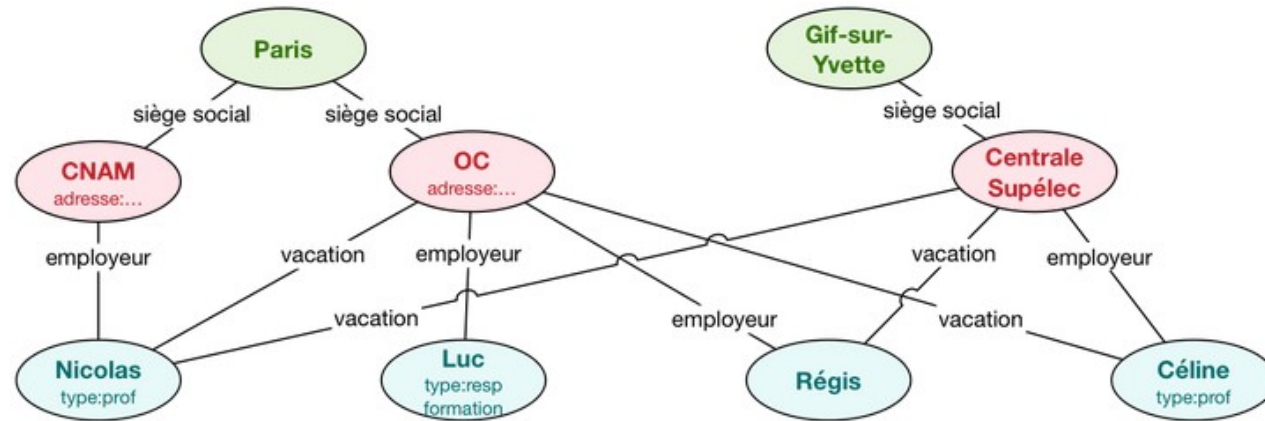


- La valeur est structurée. Des langages d'interrogation peuvent exister
- MongoDB, CouchBase, DynamoDB, Cassandra

- Big Data

Les types de BD NoSQL

- Les graphes

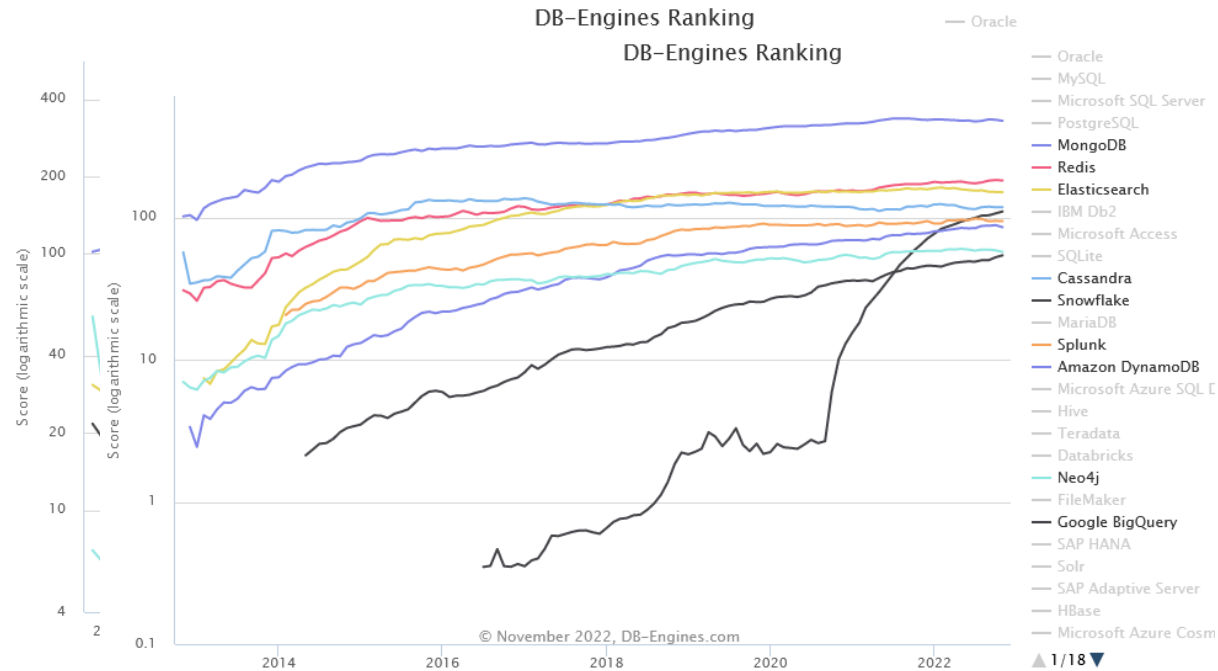


- Les données sont constituées par les nœuds et les liens.
- Requêtes d'interrogation basées sur la gestion des chemins
- Utilisés pour les réseaux sociaux, réseaux routier, électrique
- Neo4j, OrientDB, FlockDB

• Big Data

Les types de BD NoSQL

- Le choix d'une BD SQL se fait en fonction de son type et facilité de mise en œuvre, au regard du cas à traiter
 - Site https://db-engines.com/en/ranking_trend



- MongoDB est mis à disposition sous deux formes :
 - En local sur des machines
 - Sur le cloud : MongoDB Atlas
- Nous proposons une initiation à MongoDB en local:
 - Sa mise en place
 - Utilisation en mode console
 - Utilisation de MongoDB Compass
 - Interface avec Python

=> Utiliser le fichier 2025Mongodb.pdf