



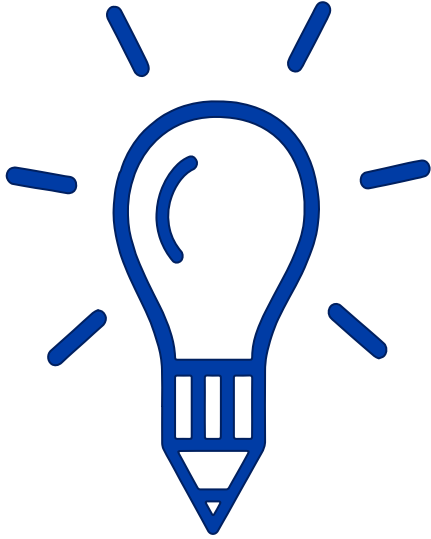
# Formation NoSQL

*Partie 3 : Principaux acteurs*



# **Principaux acteurs**

# Introduction



**01**

Les solutions NoSQL et leurs choix techniques :  
CouchDB, MongoDB, Cassandra, Hbase  
(Hadoop), ElasticSearch, ...

**02**

Démonstrations avec Cassandra et couchDB

**03**

Critères de choix

**04**

Mise en œuvre



# Solutions NoSQL

Avec le grand nombre des bases de données NoSQL  
open source et commerciales sur le marché,  
comment choisir le bon produit ou service cloud?

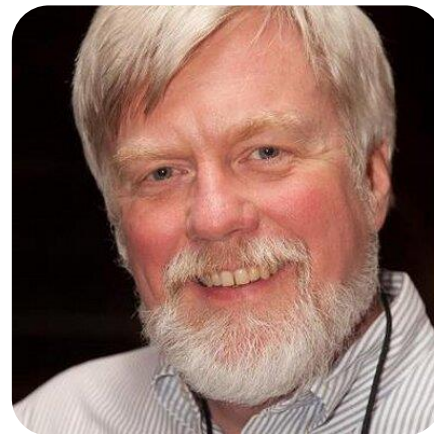




# Solutions NoSQL

L'un des facteurs essentiels est de savoir à quelles fins vous voulez utiliser les données.

*Carl Olofson, vice-président chargé de la recherche sur les technologies des bases de données, IDC (**International Data Corporation**) Carl Olofson effectue des recherches et des analyses pour IDC depuis 1997, et gère le programme de logiciels de gestion de bases de données d'IDC.*



Carl Olofson



# Solutions NoSQL

Les bases de données NoSQL varient en termes d'architecture et de fonction. Vous devez donc choisir le type qui convient le mieux à la tâche souhaitée :

**Stockage Clé-Valeur sont les meilleurs pour le partage persistant de données par plusieurs processus ou microservices dans une application**

**Si vous prévoyez d'effectuer une analyse profonde des relations pour le calcul de proximité, la détection des fraudes ou l'évaluation de la structure associative, une base de données de graphes pourrait être le meilleur choix**





# Solutions NoSQL

Les bases de données NoSQL varient en termes d'architecture et de fonction. Vous devez donc choisir le type qui convient le mieux à la tâche souhaitée :

*Si vous avez besoin de collecter des données très rapidement et à des volumes élevés pour des analyses, tournez-vous vers un stockage à colonnes larges.*

*Ces bases de données NoSQL ont tendance à offrir également un support pour les documents et les graphes.*





# Solutions NoSQL

## Critères de choix des bases de données NoSQL

### Le coût

- Le coût de la solution est un facteur qui peut avoir son importance. Il est bon de consulter l'offre et de voir **si une version professionnelle est disponible (payante)**, les services associés (**facilité d'administration/exploitation/déploiement**).
- Le coût d'infrastructure est similaire quelle que soit la solution car dépendant de vos besoins et de l'implémentation des algorithmes que vous aurez réalisés. Toutefois, les offres « Cloud » ne sont pas forcément toutes adaptées à votre choix. **Par exemple : DynamoDB n'est disponible que chez Amazon.**





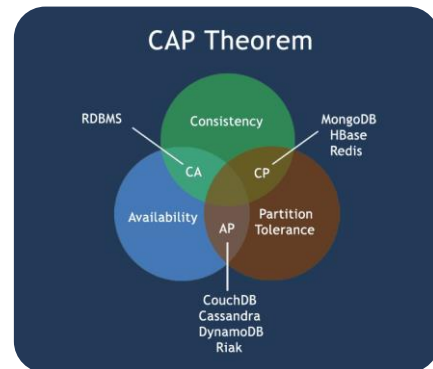


# Solutions NoSQL

## Critères de choix des bases de données NoSQL

### La cohérence des données

La cohérence des données est un critère fondamental pour des applications reposant sur des données « fraîches » ou des applications reposant sur des décisions critiques.



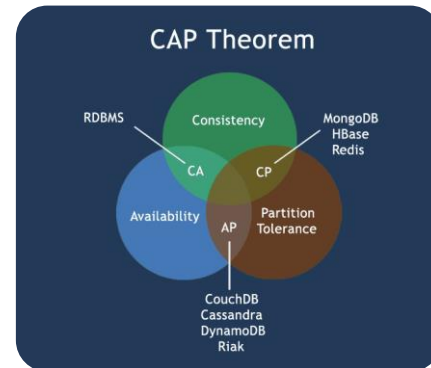


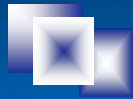
# Solutions NoSQL

## Critères de choix des bases de données NoSQL

### La disponibilité du système

À opposer à la **disponibilité** grâce au théorème de CAP, nous pouvons déduire que ces deux critères ne sont pas compatibles. Attention, les temps de latence restent faibles (quelques centaines de millisecondes), mais si cela reste crucial, il ne faut pas hésiter. Toutefois, cela ne concerne que les mises à jour de données déjà existantes, pas les insertions.





# Solutions NoSQL

## Critères de choix des bases de données NoSQL

### Le langage d'interrogation

La richesse du **langage** permet également de manipuler simplement la base NoSQL. Il facilite la conduite du changement des développeurs dans le processus d'intégration de la solution. Plus le langage est riche, plus vous serez à même d'exprimer vos besoins. Toutefois, une formation est souvent nécessaire pour comprendre les subtilités du langage.





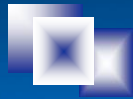
# Solutions NoSQL

## Critères de choix des bases de données NoSQL

### Les fonctionnalités

La **fonctionnalité** recherchée est souvent un critère clé dans le choix. Il est fréquemment à l'origine du changement relationnel/NoSQL. Le but étant de répondre à une demande précise : data mining, recherche textuelle, temps-réel, graphe, outil compatible.





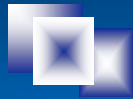
# Solutions NoSQL

## Spécifier le besoin

### Matrices multicritères

- Maintenant que nous avons défini nos critères de choix, il faut prendre une décision. Pour cela, prenez le panel de solutions NoSQL candidates mettez-le dans une **matrice multicritères**. Cette matrice donne en abscisse les critères de choix, et en ordonnée les solutions proposées.
- Il faut tout d'abord donner un poids à chacun de ces critères. Si on choisit cinq critères , il suffit de répartir pour chaque critère une valeur entre 0 et 5. La seule contrainte est que le somme de ces poids soit égale à 5. **Un poids de 0 signifie que ce critère n'a aucune importance**, et **une note supérieure à 2,5 qu'il est primordiale** (car il représente à lui seul la moitié des critères).





# Solutions NoSQL

## Spécifier le besoin

### Exemple

	Poids	MongoDB	Cassandra	ElasticSearch
Coût	0,5	4	4	4
Cohérence	1	4	2	2
Disponibilité	0,5	2	4	3
Langage	1	4	2	3
Fonctionnalité : Symfony	2	4	2	4
Total	<b>5</b>	<b>19</b>	<b>12</b>	<b>16,5</b>





# Solutions NoSQL

## Spécifier le besoin

### Exemple

	Poids	MongoDB	Cassandra	ElasticSearch
Coût	0,5	4	4	4
Cohérence	1	4	2	2
Disponibilité	0,5	2	4	3
Langage	1	4	2	3
Fonctionnalité : Symfony	2	4	2	4
Total	<b>5</b>	<b>19</b>	<b>12</b>	<b>16,5</b>





# Solutions NoSQL

## Spécifier le besoin

### Exemple

	Poids	MongoDB	Cassandra	ElasticSearch
Coût	0,5	4	4	4
Cohérence	1	4	2	2
Disponibilité	0,5	2	4	3
Langage	1	4	2	3
Fonctionnalité : Symfony	2	4	2	4
Total	5	19	12	16,5







# Solutions NoSQL

## Spécifier le besoin

### Exemple

	Poids	MongoDB	Cassandra	ElasticSearch
Coût	0,5	4	4	4
Cohérence	1	4	2	2
Disponibilité	0,5	2	4	3
Langage	1	4	2	3
Fonctionnalité : Symfony	2	4	2	4
Total	5	19	12	16,5





# Solutions NoSQL

## Spécifier le besoin

### Exemple

La pondération d'un projet utilisant le framework Symfony, avec gestion de la cohérence et un langage riche oriente le choix vers MongoDB. On peut être surpris par le résultat de Cassandra, mais le fait qu'il ne soit pas intégré nativement à Symfony et son langage trop réduit par rapport à SQL en fait une solution non envisageable. De son côté, Elasticsearch est une bonne solution, mais est rejetée pour sa gestion de la cohérence et son langage trop complexe.





# Solutions NoSQL

## BDEngines

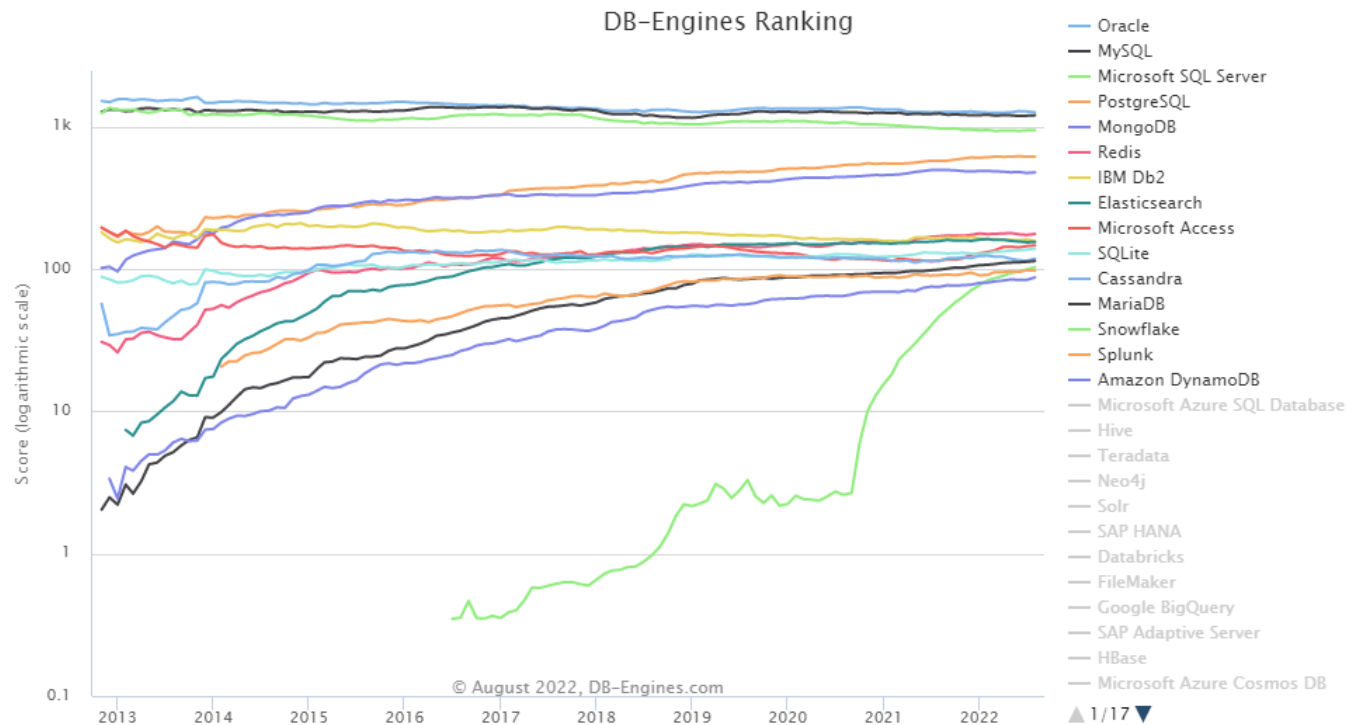
Le site web DB-Engines.org donne de nombreuses caractéristiques sur toutes les bases de données référencées sur le Web. Non seulement il donne les caractéristiques pour chaque BDD, mais il évalue également son impact sur la toile

En répertoriant chaque ticket sur **stackoverflow (bug, solution, code, conseil...)**, **les blogs et tutoriels, leur utilisation par les entreprises (spécifié dans les rapports d'architecture)**, DB-Engines produit un score pour chaque solution. Le but est de **connaître les tendances**, aussi bien pour les développeurs que les entreprises. Toutefois, cela n'est qu'un prisme sur le réel, basé uniquement sur les informations présentes sur internet. Mais cela donne déjà un avis et permet de se poser des questions.



# Solutions NoSQL

## BDEngines





# Solutions NoSQL



[MongoDB](#) est la base de données NoSQL la plus populaire. Base de données orientée documents, gratuite et open source, multiplateforme, MongoDB utilise des documents de type JSON avec des schémas.

MongoDB a été conçu pour supporter de gros volumes de données. Il met en œuvre une logique de scalabilité\* horizontale avec sharding\*. Pour aller plus loin, MongoDB permet également de mettre en œuvre un système MapReduce\*.

[MongoDB Atlas](#) (Version cloud)

Pas d'API





# Solutions NoSQL



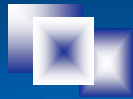
## DynamoDB



**Amazon DynamoDB** est une autre base de données NoSQL populaire basée sur le cloud. Amazon DynamoDB est une plateforme NoSQL entièrement gérée qui utilise un lecteur à état solide (SSD) pour stocker, traiter et accéder aux données afin de prendre en charge des applications hautes performances et orientées vers l'échelle.

Elle répartit automatiquement les données entre les serveurs en fonction du débit et des besoins de stockage de la charge de travail, et gère les cas d'utilisation de haute performance plus importants.





# Solutions NoSQL



## DynamoDB



Les utilisateurs peuvent faire évoluer, surveiller et gérer leurs tables à la fois via les interfaces de programmation d'applications (API) et la console de gestion d'Amazon Web Services. DynamoDB est étroitement intégré à Amazon EMR (un cadre géré pour Apache Hadoop, Apache Spark et HBase) qui offre la possibilité d'exécuter des requêtes couvrant plusieurs sources de données. La plateforme prend en charge les modèles clé-valeur et document.





# Solutions NoSQL



## CouchDB

Couchbase est une plateforme de base de données supportant les documents JSON.

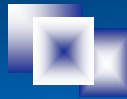
Couchbase Server, une base de données NoSQL open source de type clé-valeur et de type document.

Les entreprises utilisent Couchbase pour prendre en charge des applications sociales et mobiles, de stockage de contenu et de métadonnées, des transactions de commerce électronique et des applications de jeux en ligne.

Couchbase offre une prise en charge complète des documents, un modèle de données flexible, l'indexation, la recherche plein texte et MapReduce pour les analyses en temps réel.







# Solutions NoSQL



Parrainée par [Redis Labs](#), la plate-forme open source Redis Enterprise est l'une des bases de données NoSQL clé-valeur les plus courantes, selon M. Olofson d'IDC.

[Redis](#) est une base de données en mémoire très performante, qui prend en charge la cohérence forte et faible, un modèle flexible sans schéma, la haute disponibilité et la facilité de déploiement.

[Redis Labs](#) a développé des fonctionnalités et une technologie supplémentaires qui encapsulent le logiciel open source et fournissent une architecture de déploiement améliorée pour Redis, tout en supportant l'API open source.





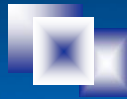
# Solutions NoSQL



ENTERPRISE

Le modèle de données prend en charge les clés et les valeurs, une variété de structures de données telles que les listes, les ensembles.etc ainsi qu'une gamme de modèles grâce à des modules enfichables tels que la recherche, les graphes, JSON et XML.

Redis prend en charge de nombreux cas d'utilisation, notamment l'analyse en temps réel, les transactions, l'ingestion de données, les médias sociaux, la gestion des tâches, la mise en file d'attente des messages et la mise en cache.

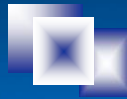


# Solutions NoSQL



[MarkLogic NoSQL Database](#) est une base de données d'entreprise opérationnelle et transactionnelle conçue pour la vitesse et l'échelle NoSQL. Grâce à une approche multimodèle, la base de données intègre et stocke les données critiques, puis vous permet de visualiser ces données sous forme de documents, de graphiques ou de données relationnelles, que ce soit sur site, en mode virtuel ou dans le cloud.

Il offre une haute disponibilité et des fonctions de sécurité au niveau des données, notamment la conformité ACID, la sécurité au niveau des éléments, l'anonymisation, la rédaction et le cryptage avancé. Pour ces raisons, il convient aux entreprises qui cherchent à partager des quantités massives d'informations sensibles. MarkLogic est également la seule base de données NoSQL à bénéficier d'une certification Critères Communs.



# Solutions NoSQL



D'autres fonctionnalités clés visent à améliorer l'expérience des utilisateurs en créant une vue unique et unifiée des données, consultable et pouvant être validée à tout moment grâce aux métadonnées. Ces fonctionnalités comprennent la sémantique bitemporale, la capacité d'ingérer des données structurées et non structurées (stockage natif pour JSON, XML, RDF, géospatial et grands binaires) et l'index universel "ask anything".

Un concentrateur de données opérationnelles qui contribue à la gouvernance et à la conformité de l'entreprise rend MarkLogic utile pour les grandes entreprises avec des silos de données, ainsi que pour celles qui sont confrontées à des réglementations et à des menaces accrues de cybersécurité.



# Solutions NoSQL

Elasticsearch est un moteur de recherche orienté document programmé en Java . Créé en 2012, il devient de plus en plus populaire avec une communauté grandissante. Il est gratuit et open source. Il dispose de plugins et d'outils payants.

Elasticsearch est très efficace pour effectuer des recherches complexes sur de gros volumes de données.



elasticsearch



# Solutions NoSQL



## ***cassandra***

**Cassandra est un système NoSQL**, créé par Facebook et sorti en 2008. Il est orienté colonnes et open source. Il est largement utilisé par de grandes entreprises : CERN, eBay (avec 250TB de données), Netflix (avec 420TB de données et un trillion de requêtes par jour), Github, Spotify, Instagram... Le projet est toujours en cours et des entreprises comme Facebook participent à l'avancement du produit. Une version commerciale existe également éditée par la société Datastax.

Cassandra a été créé pour permettre une forte évolutivité et garantir une haute disponibilité. Lorsque vous ajoutez un nœud Cassandra au sein d'un cluster, la puissance est améliorée proportionnellement aux nœuds ajoutés. En d'autres termes, les performances restent les mêmes, il n'y a donc aucun problème pour ajouter un nœud comme cela peut être le cas avec d'autres SGBD.



# Solutions NoSQL

D'autres offres de bases de données NoSQL open source et commerciales incluent :

- Blazegraph, from Systap
- Google Bigtable, from Google
- Helium, from Levyx
- Microsoft Azure Cosmos DB, from Microsoft
- Neo4j, from Neo4j
- Oracle NoSQL Database, from Oracle
- ThingSpan, from Objectivity