

Cours de Bases de données NOSQL (MongoDB)

madaniabdellah@gmail.com

Plan du cours

- Rappels :
 - JSON
 - Bases de Données Relationnelles
- **Introduction : Mouvement NoSQL**
- MongoDB
 - Caractéristiques de MongoDB
 - Requêtes avec MongoDB
 - MongoDB et Java
 - MongoDB et PHP
 - Réplication et reprise sur panne dans MongoDB
 - Partitionnement dans MongoDB
 - TextSearch

HOW TO WRITE A CV



Le mouvement NoSQL

madaniabdellah@gmail.com

Le mouvement NoSQL

Le NoSQL (Not Only SQL) c'est quoi ?

- SGBD non relationnel issue du monde Web
- Les données ne sont plus manipulées avec le SQL
- L'unité de stockage n'est plus la table
- La définition d'un schéma de données relationnel n'est plus nécessaire (schemaless)
- Renonciation aux fonctionnalités des SGBDR (ACID)
- Les principaux axes sont la haute disponibilité et la scalabilité des données
- Permet de gérer de très grosses volumétries de données (Big Data)

Le mouvement NoSQL

Que signifie le terme NoSQL ?

- Contrairement à ce qu'on pourrait croire au premier abord, une base de données NoSQL ne veut pas dire qu'on n'effectue plus de requêtes
- NoSQL signifie simplement *Not Only SQL*.
- Il ne s'agit pas d'un nouveau langage de requête permettant de dialoguer avec un SGBD
- Il s'agit d'une nouvelle approche du stockage de données

Le mouvement NoSQL

Que signifie le terme NoSQL ?

- Désigne l'ensemble des bases de données qui s'opposent à la notion relationnelle des SGBDR.
- NoSQL ne vient pas remplacer les BD relationnelles mais proposer une alternative ou compléter les fonctionnalités des SGBDR pour donner des solutions plus intéressantes dans certains contextes.

Le mouvement NoSQL

Caractéristiques du NoSQL

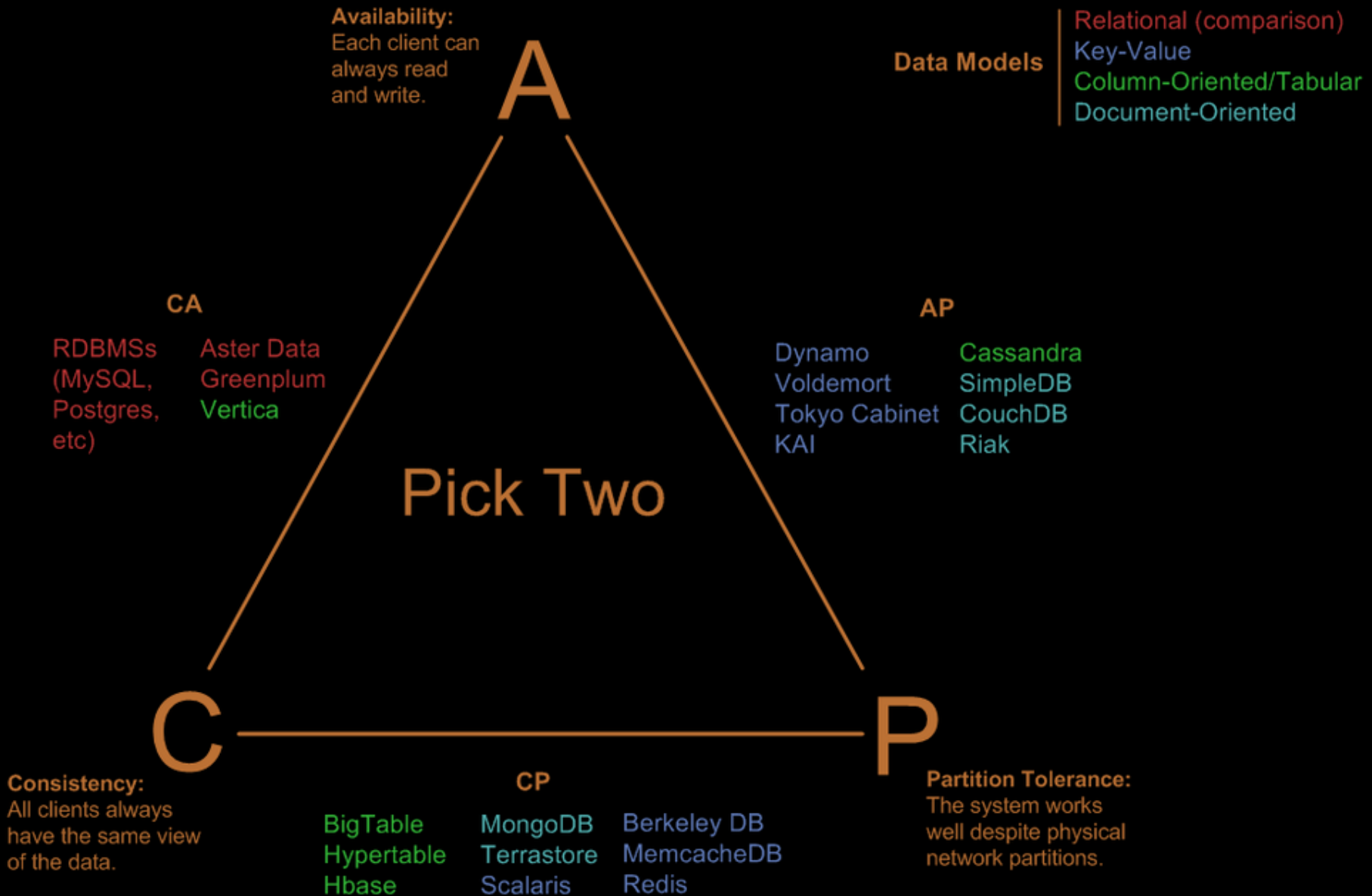
- Ne plus répondre au modèle relationnel.
- Les bases de données NoSQL répondent aussi au théorème du CAP d'Eric Brewer qui est plus adapté aux systèmes distribués.
- Ce théorème énonce que tout système distribué peut répondre à 2 contraintes parmi :
 - **Consistency** : Cohérence
 - **Availability** : Haute disponibilité
 - **Partition Tolerance** : Tolérance au Partitionnement

Le mouvement NoSQL

Le théorème CAP

- Cohérence (Consistency)
 - Tous les nœuds du système voient exactement les mêmes données au même moment
- Disponibilité (Availability)
 - En cas de panne, les données restent accessibles
 - Garantie que les requêtes reçoivent une réponse même si un ou plusieurs nœuds sont défectueux
- Résistance au morcellement (Partition Tolerance)
 - Le système peut être partitionné
 - Aucune panne autre qu'une coupure réseau totale ne doit empêcher le système de répondre. Idéalement, le système doit être en mesure de réconcilier les mises à jour une fois les nœuds à nouveau accessibles les uns des autres

Visual Guide to NoSQL Systems



Le mouvement NoSQL

Caractéristiques du NoSQL

- D'autres caractéristiques communes aux différentes bases de données NoSQL peuvent être citées tel que :
 - le partitionnement horizontal sur plusieurs nœuds
 - la réplication des données
 - les schémas dynamiques ou l'absence de schéma.

Le mouvement NoSQL

Avantages/Inconvénients

- Les avantages souvent mis en avant :
 - Plus souples
 - Plus faciles d'utilisation
 - Conçus pour être utilisés sur des architectures massivement parallèles, sur des très grandes quantités de données
- Inconvénients majeurs :
 - Moins de garanties de cohérence et de consistance des données.
 - Ces contraintes sont laissées à la charge du programmeur.

Le mouvement NoSQL

NoSQL vs SGBDR

SGBDR

- Les points forts :
 - Beaucoup de fonctionnalités et de règles pour garantir des bases cohérentes et complètes : mécanisme de verrous pour la gestion des accès concurrentiels, respect des propriétés ACID (Atomicité, Cohérence, Isolation, Durabilité).
 - Utilisation des outils spécialisés comme le modèle entité-relation qui est le Modèle Conceptuel de Données (MCD) utilisé pour modéliser la structure de la base.
- Les points faibles :
 - Performances qui déclinent avec l'augmentation du volume de données
 - Problématique de distribution des données d'une base.

Le mouvement NoSQL

NoSQL / SGBDR

- NoSQL
 - Les points forts :
 - Bons temps de réponse malgré de très gros volumes de données
 - Facilement distribuable
 - Plus flexible en cas de panne (disponibilité partielle voire totale)

Le mouvement NoSQL

NoSQL / SGBDR

NoSQL

- Les points faibles :
 - Moins de propriétés garantissant un état cohérent de la base.
Conformément au théorème de Brewer (aussi appelé théorème du CAP), seules deux des trois propriétés suivantes peuvent être respectées par un SGBD NoSQL : Cohérence, Disponibilité et Résistance au morcellement.
 - Pas de mécanismes de jointures, le côté client doit pallier à ce problème.
 - Il est rare qu'un SGBD de type NoSQL implémente un mécanisme de verrous pour garantir la cohérence pendant des accès concurrents.

Le mouvement NoSQL

Types de bases de données NoSQL

- Les solutions NoSQL existantes peuvent être regroupées en 4 grandes familles.
- Les bases clefs/valeurs
 - Redis, Voldemort
- Les bases orientées documents
 - **MongoDB**, CouchDB, Riak
- Les bases orientées colonnes
 - Cassandra, BigTable, HBase
- Les bases orientées graphes
 - Neo4j: en Java
 - FlockDB: utilisé par Twitter

Le mouvement NoSQL

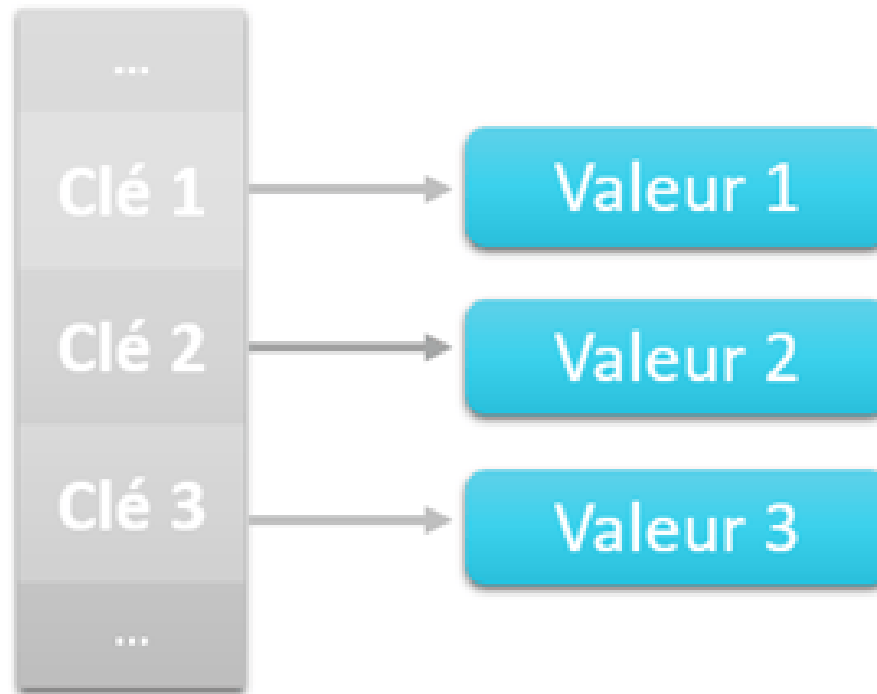
Types de bases de données NoSQL

Bases Orientées Clé/Valeur

- Ce modèle peut être assimilé à une hashmap distribuée.
- Les données sont représentées par un couple clé/valeur.
- La valeur peut être une simple chaîne de caractères, un objet, etc.
- Les solutions les plus connues sont Redis, Riak et Voldemort créé par LinkedIn.

Le mouvement NoSQL (Bases Orientées Clé/Valeur)

Les bases NoSQL Clé Valeur



Le mouvement NoSQL

Types de bases de données NoSQL

Bases Orientées Colonnes

- Ce modèle ressemble à première vue à une table dans un SGBDR
- La différence avec une BD NoSQL orientée colonne, le nombre de colonnes est dynamique. En effet :
 - Dans une table relationnelle, le nombre de colonnes est fixé dès la création du schéma de la table pour tous les enregistrements.
 - Par contre, avec ce modèle, le nombre de colonnes peut varier d'un enregistrement à un autre ce qui évite de retrouver des colonnes ayant des valeurs NULL.
- Comme solutions, on retrouve principalement :
 - HBase (implémentation Open Source du modèle BigTable publié par Google)
 - Cassandra (projet Apache qui respecte l'architecture distribuée de Dynamo d'Amazon et le modèle BigTable de Google).

Le mouvement NoSQL (Bases orientées colonnes)

Le stockage dans Bases Relationnelles

Clé	Date début validité	Date fin Validité	Nom	Prénom	Csp
A	01/01/2010	02/02/2015	Dupont	Null	Cadre
A	03/02/2015	Null	Cap	Null	Cadre
B	01/01/2010	Null	Null	Joséphine	Null
C	01/01/2010	03/03/2016	Black	George	Cadre
C	04/04/2016	Null	Black	George	Retraité

VS. Dans les bases orientées colonnes

A	Nom : Dupont DATE : 01/01/2010	Csp: Cadre DATE: 01/01/2010
	Nom: Cap DATE : 03/02/2015	
B	Prénom : Joséphine DATE: 01/01/2010	
C	Nom : Black DATE: 01/01/2010	Prénom : George DATE: 01/01/2010
		Csp : Cadre DATE: 01/01/2010 Csp: Retraité DATE: 04/04/2016

illustradata.com

Le mouvement NoSQL

Types de bases de données NoSQL

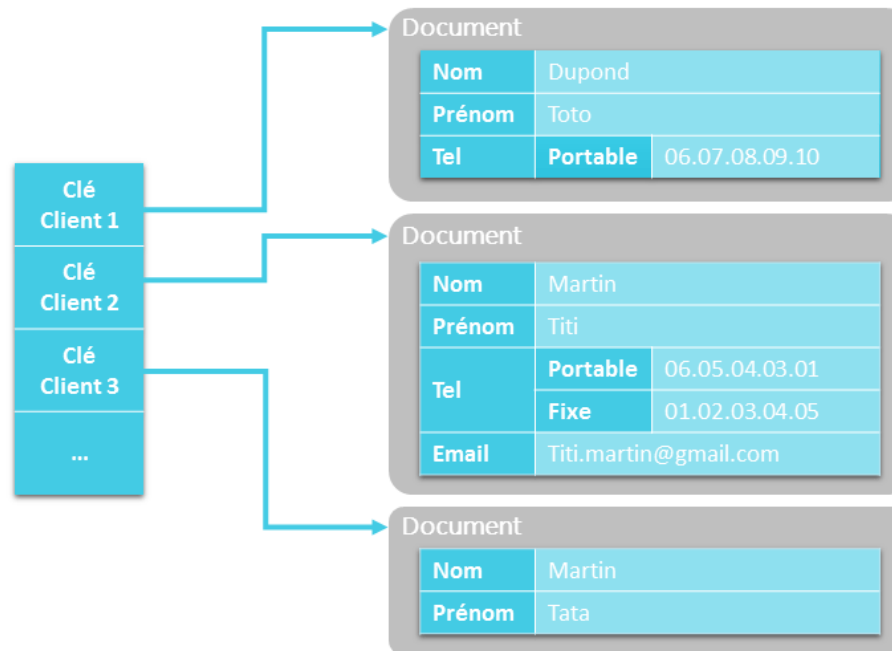
Bases Orientées Document

- Ce modèle se base sur le paradigme clé valeur.
- La valeur, dans ce cas, est un document de type JSON ou XML.
- L'avantage est de pouvoir récupérer, via une seule clé, un ensemble d'informations structurées de manière hiérarchique.
- La même opération dans le monde relationnel impliquerait plusieurs jointures.
- Pour ce modèle, les implémentations les plus populaires sont :
 - CouchDB d'Apache,
 - RavenDB (destiné aux plateformes .NET/Windows)
 - MongoDB.

Le mouvement NoSQL (Bases Orientées Document)

Un exemple de stockage de « fiches clients » dans une base NoSQL orientée document

illustradata.com



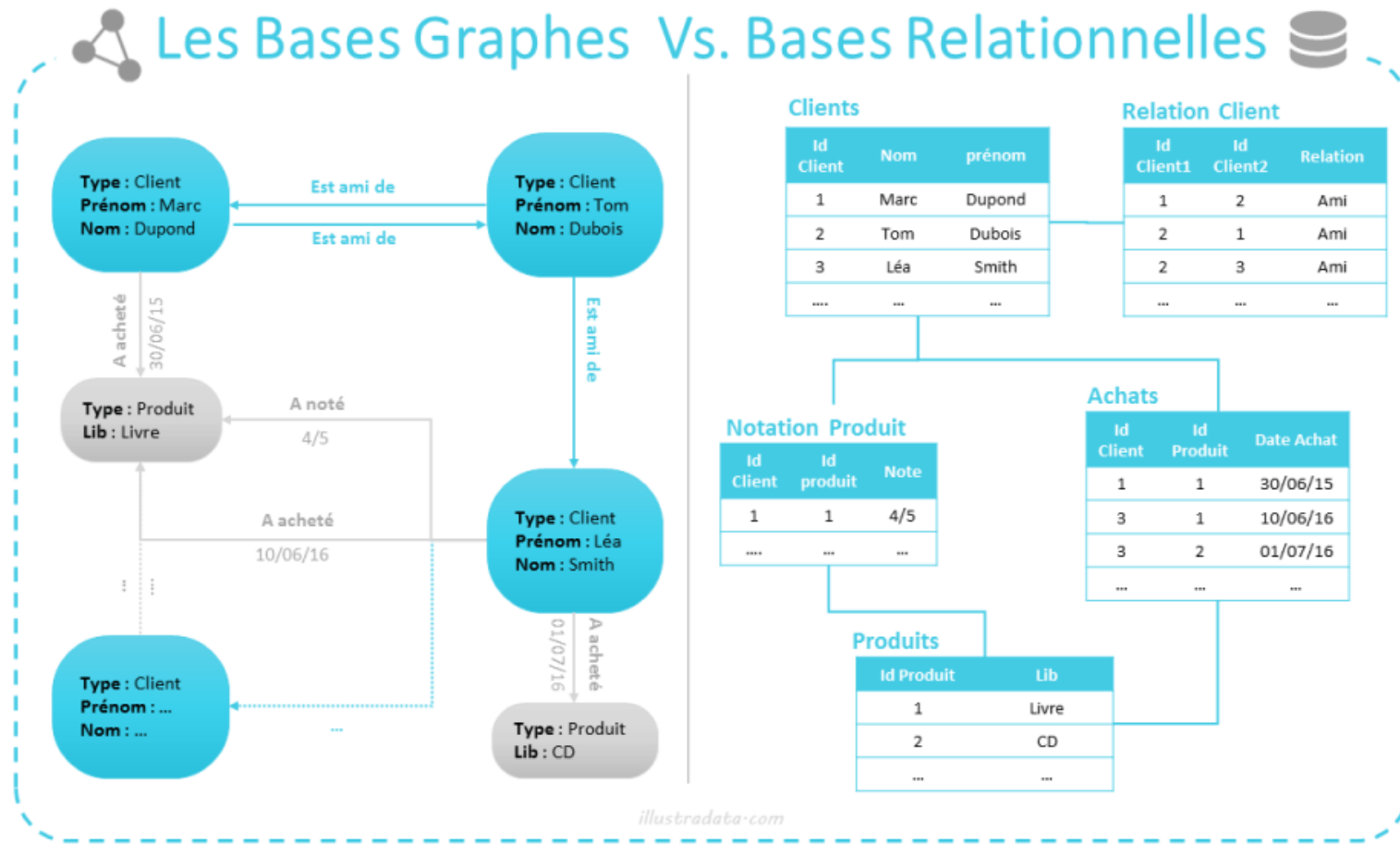
Le mouvement NoSQL

Types de bases de données NoSQL

Bases orienté graphe






- Ce modèle de représentation des données se base sur la théorie des graphes.
- Il s'appuie sur la notion de nœuds, de relations et de propriétés qui leur sont rattachées.
- Ce modèle facilite la représentation du monde réel, ce qui le rend adapté au traitement des données des réseaux sociaux.
- La principale solution est Neo4J.

Le mouvement NoSQL (Bases orientées graphe)



Classement des SGBD

(<http://db-engines.com/en/>)

Rank			DBMS	Database Model	Score		
Aug 2015	Jul 2015	Aug 2014			Aug 2015	Jul 2015	Aug 2014
1.	1.	1.	Oracle	Relational DBMS	1453.02	-3.70	-17.83
2.	2.	2.	MySQL	Relational DBMS	1292.03	+8.69	+10.81
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1108.66	+5.60	-133.84
4.	4.	↑ 5.	MongoDB 	Document store	294.65	+7.26	+57.30
5.	5.	↓ 4.	PostgreSQL	Relational DBMS	281.86	+9.04	+32.01
6.	6.	6.	DB2	Relational DBMS	201.23	+3.12	-5.19
7.	7.	7.	Microsoft Access	Relational DBMS	144.20	-0.10	+4.58
8.	8.	↑ 10.	Cassandra 	Wide column store	113.99	+1.28	+32.09
9.	9.	↓ 8.	SQLite	Relational DBMS	105.82	-0.05	+16.95
10.	10.	↑ 11.	Redis 	Key-value store	98.81	+3.73	+28.01
11.	11.	↓ 9.	SAP Adaptive Server	Relational DBMS	85.11	-2.10	-1.06
12.	12.	12.	Solr	Search engine	81.90	+2.61	+12.87
13.	13.	13.	Teradata	Relational DBMS	73.59	+1.28	+8.21
14.	14.	↑ 16.	Elasticsearch	Search engine	69.64	-0.52	+30.84
15.	15.	15.	HBase	Wide column store	59.95	-0.97	+18.03
16.	↑ 17.	↑ 19.	Hive	Relational DBMS	53.87	+5.67	+23.06
17.	↓ 16.	↓ 14.	FileMaker	Relational DBMS	51.87	+1.05	-0.20
18.	18.	↑ 20.	Splunk	Search engine	42.20	+0.55	+14.59
19.	19.	↑ 23.	SAP HANA	Relational DBMS	38.25	+1.63	+15.92
20.	20.	↓ 17.	Informix	Relational DBMS	36.80	+0.91	+3.64
21.	21.	↓ 18.	Memcached	Key-value store	33.38	+0.25	+2.39
22.	22.	22.	Neo4j 	Graph DBMS	33.16	+1.81	+10.25
23.	23.	↓ 21.	CouchDB	Document store	27.28	+0.35	+3.15
24.	24.	24.	Couchbase	Document store	26.16	-0.11	+8.62
25.	↑ 26.	↑ 27.	MariaDB 	Relational DBMS	24.12	+0.85	+8.76

Classement des SGBD

(<http://db-engines.com/en/>)

Popularity changes per category, August 2015

