

JPO : Webinar d'information sur nos formations → RDV mardi prochain à 17h30. **Je m'inscris !** (<https://app.livestorm.co/datascientest/httpsapplivestormcodatascientestjpos>)



≡Menu

JPO : Webinar d'information sur nos formations → RDV mardi prochain à 17h30. **Je m'inscris !** (<https://app.livestorm.co/datascientest/httpsapplivestormcodatascientestjpos>)



≡Menu

Prendre RDV (/rdv)

- **Retour accueil blog**
(<https://datascientest.com/blog>)

MongoDB

**Recherchez
Dans Le
Blog :**

Recherchez un art

**Articles
Récents**

[La Data Science,
nouvel eldorado
des cabinets
d'expertise
comptable](#)
(<https://datascientest.com/la-data->

**MongoDB : tout savoir sur la
base de données NoSQL
orientée document**

**Margot P** ([https://datascientest.com/author](https://datascientest.com/author/margot)[/margot](#))

• 📅 3 février 2021



4 min

MongoDB est une base de données NoSQL orientée document. Elle se distingue des bases de données relationnelles par sa flexibilité et ses performances. Découvrez tout ce que vous devez savoir sur cet outil incontournable pour l'ingénierie des données.

Apparue au milieu des années 2000, [MongoDB](#) (<https://www.mongodb.com/fr>) est [une base de données](#) (<https://datascientest.com/base-de-donnees-definition>) NoSQL orientée document. Elle est utilisée pour le stockage de volumes massifs de données.

Contrairement à une base de données [relationnelle SQL](#) (<https://datascientest.com/sql-tout-savoir>) traditionnelle, MongoDB ne repose pas sur des tableaux et des colonnes. Les données sont stockées sous forme de collections et de documents.

Les **documents sont des paires de valeurs / clés** servant d'unité de données de base. Les collections quant à elles contiennent des ensembles de documents et de fonctions. Elles sont l'équivalent des tableaux dans les bases de données relationnelles classiques.

Les caractéristiques de MongoDB

[science-nouvel-eldorado-des-cabinets-dexpertise-comptable](#)

[Pandas : tout savoir sur la bibliothèque Python dédiée à la Data Science](#) (<https://datascientest.com/pandas-python-data-science>)

[Architecting on AWS : tout savoir sur la formation d'architecture cloud](#) (<https://datascientest.com/architecting-on-aws-tout-savoir-sur-la-formation-darchitecture-cloud>)

**Vidéo
D'alumnis**

Catégories

Chaque base de données MongoDB contient des collections, contenant elles-même des documents. Chaque document est différent et peut comporter un nombre de champs variable. La taille et le contenu de chaque document varient également.

La [structure d'un document](#) correspond à la façon dont les développeurs construisent leurs classes et objets dans le langage de programmation utilisé. En général, les classes ne sont pas de lignes et des colonnes mais ont une structure claire composée de paires de valeurs / clés.

Les [documents n'ont pas de schéma prédéfini](#) et des champs peuvent être ajoutés à volonté. Le modèle de données disponible au sein de MongoDB permet de représenter des relations hiérarchiques ou autre structure complexe plus facilement.

Une autre caractéristique majeure de MongoDB est [l'élasticité de ses environnements](#). De nombreuses entreprises ont des clusters de plus de 100 noeuds pour des bases de données contenant des millions de documents.

Commencer

une

formation

[Actualités](#)

<https://datascientest.com/blog-data-ia-actualites/actualites>

[Deep Learning](#)

<https://datascientest.com/blog-data-ia-actualites/deep-learning>

[Machine](#)

[Learning](#)

<https://datascientest.com/blog-data-ia-actualites/machine-learning>

[Programmation](#)

[Python](#)

<https://datascientest.com/blog-data-ia-actualites/programmation-python>

**S'inscrire À
La
Newsletter**

**Plus de 30
nouveaux
articles et
fichiers**

chaque mois !

Votre ac

Je m'a

MongoDB

(https://datas

cientest.com

/formation-

data-

engineer)

L'architecture MongoDB et ses composants

L'architecture MongoDB repose sur plusieurs composants principaux. Tout d'abord, " **_id** " est un champ requis pour chaque document. Il représente une valeur unique, et peut être considéré comme la clé principale du document permettant de l'identifier au sein de la collection.

Un **document** est l'équivalent d'un enregistrement dans

une base de données traditionnelle. Il est constitué de champs de nom et de valeurs. Chaque champ est une association entre un nom et une valeur, et s'apparente à une colonne dans une base de données relationnelle.

Une **collection est un groupe de documents** MongoDB, et correspond à un tableau créé avec n'importe quel autre RDMS comme Oracle ou MS SQL sur une base de données relationnelle. Elle n'a pas de structure prédéfinie.

Une base de données est **un conteneur de collections**, au même titre qu'un RDMS est un conteneur de tableaux pour les bases de données relationnelles. Chacune a son propre ensemble de fichiers sur le système fichier. Un serveur MongoDB peut stocker de multiples bases de données.

Enfin, **le JSON** (JavaScript Object Notation) est un format en texte clair pour exprimer les données structurées. Il est pris en charge par de nombreux langages de programmation.

Pourquoi utiliser MongoDB ? Quels sont les avantages ?

MongoDB présente plusieurs avantages majeurs. Tout d'abord, cette base de données NoSQL orientée document se révèle **très flexible** et adaptée aux cas d'usage concrets d'une entreprise.

Les **requêtes " Ad hoc "** permettent par ailleurs de retrouver des champs spécifiques au sein des documents. Notons aussi la possibilité de créer des index pour améliorer la performance des recherches. N'importe quel champ peut être indexé.

Un autre avantage est la possibilité de créer **des " replica sets "** constitués de deux instances MongoDB ou plus. Chaque membre peut jouer le rôle de réplique secondaire ou primaire à tout moment.

La **réplique primaire** est le serveur principal, qui interagit avec le client et effectue toutes les opérations de lecture et d'écriture. Les répliques secondaires quant à elles conservent une copie des données. Ainsi, en cas de panne de la réplique primaire, le basculement vers la secondaire se fait automatiquement. Ce système garantit une haute disponibilité.

Enfin, **le concept du sharding** permet le scaling horizontal en répartissant les données entre de multiples instances MongoDB. La base de données peut être exécutée sur de multiples serveurs, et ceci permet d'équilibrer la charge ou de dupliquer les données pour garder le système fonctionnel en cas de panne matérielle.

De par ces nombreux avantages, MongoDB est aujourd'hui un outil très utilisé dans le domaine de l'ingénierie des données. Il s'agit d'une solution à maîtriser impérativement [pour les Data Engineers \(https://datascientest.com/data-engineer-tout-savoir\)](https://datascientest.com/data-engineer-tout-savoir).

MongoDB vs RDBMS : quelles sont les différences ?

Il existe plusieurs différences majeures entre MongoDB et les RDBMS (système de gestion de bases de données relationnelles). Comme évoqué auparavant, les données ne sont pas stockées dans des tableaux mais dans des collections de documents. Ces **documents remplacent les lignes des RDBMS**. Ils contiennent des champs de paires de valeurs / clés, remplaçant eux-mêmes les colonnes.

En outre, l'intégrité des données n'est pas une contrainte sur MongoDB. Les **données n'ont pas non plus besoin d'être "normalisées"** avant usage comme sur un RDBMS. C'est un véritable avantage, car la contrainte de normalisation peut atténuer les performances à mesure que la base de données prend de l'ampleur.

La modélisation de données sur MongoDB

Contrairement aux bases de données SQL, MongoDB n'implique **aucune contrainte en termes de structure** des documents. Les données n'ont pas de schéma préconçu, et c'est cette flexibilité qui rend MongoDB si puissant et performant.

La **modélisation des données** et la structure des documents doivent uniquement répondre aux besoins de l'utilisateur. Il convient de prendre en compte les besoins de l'application, et donc quels seront les données et les types de données nécessaires.

Si de nombreuses requêtes sont à prévoir, il est pertinent d'utiliser **des index** dans le modèle de données pour améliorer l'efficacité des requêtes. Enfin, en cas d'ajouts, de mises à jour et de suppressions de données fréquentes, on utilisera là encore des index et le système de sharding pour améliorer l'efficacité générale de l'environnement.

TOUT SAVOIR SUR SQL

Ca tombe
bien, on en
a fait une
fiche
mémo !

DÉCOUVRI

Pourquoi et comment apprendre à
utiliser MongoDB ?

(https://datascientest.com/wp-content/formation/2020/11/Data/infography_Software_Engineer)

(<https://datascientest.com/formation->

data- engineer)

MongoDB fait partie des outils indispensables de l'ingénierie des données. Pour apprendre à l'utiliser, vous pouvez vous tourner vers [les formations DataScientest](#).

La [formation Data Engineer \(https://datascientest.com/formation-data-engineer\)](https://datascientest.com/formation-data-engineer) vous apprendra le métier d'ingénieur des données, et notamment à construire des pipelines d'acquisition et de traitement automatique des données. Au sein du module " base de données ", vous apprendrez notamment à utiliser MongoDB, mais aussi Cassandra, Elastic Search, Neo4J et le langage SQL.

Si vous êtes déjà Data Scientist et que vous souhaitez apprendre à mettre en production des modèles de Machine Learning, vous pouvez vous tourner vers [notre formation Machine Learning Engineer \(https://datascientest.com/formation-machine-learning-engineer\)](https://datascientest.com/formation-machine-learning-engineer). MongoDB fait partie des outils que vous apprendrez à utiliser.

Débuter la formation

Machine

Learning

Engineer

(https://datas

cientest.com

/formation-

machine-

learning-

engineer)

Vous savez tout sur MongoDB. Découvrez d'autres outils de Data Engineering, comme [la plateforme de conteneurisation Docker](https://datascientest.com/formation-docker) (<https://datascientest.com/formation-docker>) ou [le Cloud Data Warehouse de Snowflake](https://datascientest.com/snowflake) (<https://datascientest.com/snowflake>).

[◀Précédent](https://datascientest.com/mongodb) ([https://](https://datascientest.com/mongodb)) [▶Suivant](https://datascientest.com/data-engineering) ([https://data](https://datascientest.com/data-engineering))



[Jupyter Notebook : tout savoir sur l'outil de partage de code indispensable en Data Science \(https://datascientest.com/jupyter-notebook\)](https://datascientest.com/jupyter-notebook)

Jupyter Notebook est une application web permettant de partager du code informatique. Découvrez tout ce que vous devez savoir sur cet outil très utilisé par

[LIRE PLUS » \(HTTPS://DATASCIENTEST.COM/JUPYTER-NOTEBOOK\)](https://datascientest.com/jupyter-notebook)

Laura P | décembre 20, 2021

BUSINESS ET DATA SCIENCE

What's new in the cloud? #2 – Semaine du 21-12-2021 – (<https://datascientest.com/whats-new-in-the-cloud-2-semaine-du-21-12-2021>)

Chaque semaine, nous vous aidons à y voir plus clair dans le cloud... Actualités, innovations, mises à jour, événements à ne pas rater, etc. Vous

[LIRE PLUS » \(HTTPS://DATASCIENTEST.COM/WHATS-NEW-IN-THE-CLOUD-2-SEMAINE-DU-21-12-2021\)](https://datascientest.com/whats-new-in-the-cloud-2-semaine-du-21-12-2021)

Chouaieb N | décembre 22, 2021

[Le test de Turing est-il en mesure d'estimer l'intelligence d'une machine ?](https://datascientest.com/le-test-de-turing-est-il-en-mesure-destimer-lintelligence-dune-machine) [\(https://datascientest.com/le-test-de-turing-est-il-en-mesure-destimer-lintelligence-dune-machine\)](https://datascientest.com/le-test-de-turing-est-il-en-mesure-destimer-lintelligence-dune-machine)

Depuis l'émergence du Machine Learning et du Deep Learning, une des questions les plus discutables et controversées dans le domaine de l'intelligence artificielle est en

[LIRE PLUS » \(HTTPS://DATASCIENTEST.COM/LE-TEST-DE-TURING-EST-IL-EN-MESURE-DESTIMER-LINTELLIGENCE-DUNE-MACHINE\)](https://datascientest.com/le-test-de-turing-est-il-en-mesure-destimer-lintelligence-dune-machine)

Team redac | décembre 22, 2021

Liens populaires

[Se connecter](#)

[Inscrivez-vous à notre newsletter !](#)

[Carrières : Rejoignez-nous !](#)

[Lexique Data Science](#)

Infos pratiques

[Processus Inscription](#)

[Financement](#)

[Formation Data CPF](#)

[Accueil Handicap](#)

[Prix](#)

À propos de nous

[Aide au retour à l'emploi](#)

[FAQ Data Science](#)

[Avis Datascientest](#)

[Nos réseaux sociaux](#)

[Livret apprenant](#)

Contact



[1 Terrasse Bellini 92800 Puteaux \(tel:+33980807949\)](#)



[+33 9 80 80 79 49 \(tel:+33980807949\)](#)



[contact@datascientest.com \(mailto:contact@datascientest.com\)](mailto:contact@datascientest.com)



[Prenez RDV en ligne \(https://datascientest.com/rdv\)](https://datascientest.com/rdv)



Toutes nos formations sont certifiées Qualiopi.

CPF

DataScientest [est éligible au CPF \(/formation-data-cpf\)](/formation-data-cpf). Financez votre formation en data science grâce à votre Compte de Formation.



DataScientest est un organisme de formation qui propose des cours interactifs. Apprenez d'une équipe d'experts dans le domaine de la Data.

PPC

[_\(/https://www.institutdesactuaire.com/se-perfectionner](https://www.institutdesactuaire.com/se-perfectionner)



[/devenir-actuaire-qualifie-certifie-126\)](#)

DataScientest est aujourd'hui éligible aux points PPC. (<https://datascientest.com/actuaire-demarrez-une-formation-en-data-science-eligibile-aux-points-ppc>) Faites évoluer l'actuariat grâce aux Data Sciences !

Retrouvez-nous sur:

Facebook  (<https://www.facebook.com/datascientest>) Twitter  (<https://twitter.com>

[/DataScientest__](#))  (<https://www.instagram.com/datascientest>) LinkedIn 

(<https://www.linkedin.com/school/datascientest/>) Youtube  (https://www.youtube.com/channel/UCTVTIRE_gyvZvEFJOxRZIrA) Github  (<https://github.com/DataScientest-Studio>)

©2022 DataScientest - [Nos mentions légales \(https://datascientest.com/mentions-legales\)](https://datascientest.com/mentions-legales)