

# ***Veille Technologique : Cassandra***

## ***Introduction***

Apache Cassandra est un système de gestion de base de données distribuée conçu pour gérer de grandes quantités de données sur un grand nombre de serveurs.

### ***1. Cassandra 4.0***

La version 4 de Cassandra a été publiée le 27/07/2021. Elle apporte des améliorations significatives en termes de performances, de scalabilité et de stabilité. De nouvelles fonctionnalités comme l'indexation globale, l'amélioration du support JSON et l'amélioration de la gestion des ressources sont à noter.

### ***2. Gestion de la Conformité et de la Sécurité***

Avec les préoccupations croissantes en matière de sécurité des données, il y a une attention particulière portée sur la gestion de la conformité et la sécurité dans Cassandra. Des mises à jour récentes ont inclus des fonctionnalités de sécurité avancées telles que [l'authentification Kerberos](#), le chiffrement des données au repos et des améliorations dans le contrôle d'accès.

### ***3. Cassandra en Conteneurs et Docker***

L'utilisation de Cassandra avec Docker présente plusieurs avantages significatifs. Docker assure l'isolation complète de l'environnement, permettant d'éviter les conflits et de garantir une exécution cohérente de Cassandra. Il simplifie le déploiement sur divers environnements et offre une portabilité entre systèmes compatibles Docker. La gestion des versions est facilitée, et il est possible de configurer Cassandra via des fichiers Docker Compose. Les mises à jour sont aisées, il suffit de créer une nouvelle image Docker. Les données peuvent être stockées de manière persistante grâce aux volumes Docker. La mise en place et l'orchestration de clusters locaux sont facilitées avec Docker Compose.

### ***4. Time-Series Data et Cassandra***

Cassandra est bien adaptée pour gérer des données en série temporelle ([time-series data](#)). Cette tendance s'est renforcée avec l'augmentation des cas d'utilisation liés à l'Internet des Objets (IoT) et aux systèmes de surveillance en temps réel. Des modèles de données spécifiques et des stratégies de modélisation ont émergé pour tirer le meilleur parti de Cassandra dans ce contexte.

### ***5. Intégration avec d'Autres Technologies***

Cassandra est souvent utilisée en conjonction avec d'autres technologies pour répondre à des besoins spécifiques. Des intégrations avec des outils d'analyse en temps réel comme [Kafka](#) et [Spark](#) sont fréquentes, ainsi que des solutions de recherche comme [Elasticsearch](#).

### ***6. Communauté et Ressources***

La communauté Cassandra est active et propose une multitude de ressources, y compris des forums de discussion, des blogs, des webinaires et des conférences. Rejoindre la communauté peut être un excellent moyen de rester à jour avec les derniers développements.

# Veille Technologique : Cassandra

## Mise en place de Cassandra avec Docker :

- 1- Téléchargez et installez Docker Desktop pour Windows et installez-le en suivant les instructions du programme d'installation.
- 2- Lancez Docker Desktop.
- 3- Copier l'image de Cassandra : `docker pull cassandra`
- 4- Créez et lancez : `docker run --name my-cassandra-container -d cassandra`
- 5- Vérifiez le statut de Cassandra : `docker ps`
- 6- Vous pouvez accéder à la ligne de commande CQL : `docker exec -it my-cassandra-container cqlsh`
- 7- Vous pouvez stopper le conteneur Cassandra : `docker stop my-cassandra-container`
- 8- Vous pouvez supprimer : `docker rm my-cassandra-container`

## Comparaison avec les concurrents :

Cassandra vs Apache HBase :

- HBase est plus adapté aux cas d'usage nécessitant des transactions ACID strictes, tandis que Cassandra est conçue pour une haute disponibilité et une grande scalabilité.

## Cas d'usage :

Gestion des Big Data en temps réel : Cassandra est idéale pour les applications qui nécessitent une grande disponibilité et une haute performance lors du traitement de grandes quantités de données distribuées sur plusieurs nœuds.

## Forces :

- Haute disponibilité et tolérance aux pannes.
- Évolutivité linéaire.
- Performances élevées pour les opérations de lecture.
- Modèle de données flexible.

## Faiblesses :

- Modélisation de données complexe pour les opérations de requête complexes.
- Les opérations d'écriture peuvent être moins performantes que les bases de données orientées écriture.
- Pas adapté aux requêtes nécessitant des jointures complexes.

## Conclusion

Cassandra reste une technologie importante dans le domaine des bases de données distribuées, et son évolution continue apporte des améliorations significatives. Suivre les dernières tendances et les mises à jour de la communauté peut aider à maximiser les avantages de cette puissante base de données.