

Mis à jour le 05/03/2026

S'inscrire

Formation Qdrant Database Vector

2 jours (14 heures)

Présentation

Qdrant est une base de données vectorielle conçue pour la recherche sémantique, la recommandation et les cas d'usage RAG (chatbots, assistants, recherche documentaire). Elle permet d'indexer des embeddings et de retrouver rapidement les contenus les plus proches, avec filtres et scoring.

Cette formation vous guide pour déployer Qdrant, modéliser des collections, ingérer des vecteurs et interroger l'API pour construire un moteur de recherche sémantique fiable et performant. Vous verrez comment combiner similarité vectorielle et filtres métier (tags, dates, droits) pour des résultats exploitables en production.

L'approche de notre formation est résolument pratique : ateliers d'indexation, requêtes k-NN, upsert, payloads, et tuning (HNSW, quantization). Livrables : un projet de démonstration, scripts d'import, requêtes types et une checklist d'exploitation (sauvegarde, monitoring, sécurité).

Comme toutes nos formations, celle-ci vous présentera **la dernière version stable** de la technologie et ses nouveautés.

Objectifs

- Déployer Qdrant en local et en conteneur.
- Créer des collections et définir schémas de payload.
- Indexer des embeddings et gérer les mises à jour.
- Exécuter des recherches vectorielles avec filtres et reranking.
- Optimiser performances, persistance et observabilité.

Public visé

- Développeurs back-end et full-stack
- Data engineers / ML engineers
- Architectes logiciels
- Chefs de projet technique

Pré-requis

- Notions d'API REST et formats JSON
- Bases de Python ou d'un langage équivalent
- Compréhension des embeddings et de la similarité cosinus
- Familiarité avec Docker et la ligne de commande

Pré-requis techniques

- 8 Go de RAM minimum (16 Go recommandé)
- Linux, macOS ou Windows avec WSL2
- Docker et Docker Compose installés
- Un éditeur de code et un terminal
- Python 3.10+ (ou client HTTP équivalent) pour les ateliers

Programme de formation Qdrant Database Vector

[Jour 1 - Matin]

Fondamentaux des bases vectorielles et prise en main de Qdrant

- Comprendre les embeddings : dimensions, similarité (cosine/dot/Euclidean) et impacts sur la recherche
- Concepts Qdrant : collections, points, vectors, payload, indexes
- Installation et démarrage : Docker, configuration minimale, vérifications via API
- Premiers appels API : création de collection, insertion de points, recherche k-NN
- Atelier pratique : Démarrer Qdrant et réaliser une première recherche de similarité via REST.

[Jour 1 - Après-midi]

Modélisation des données : payload, filtres et qualité de recherche

- Structurer le payload (métadonnées) pour des cas d'usage : RAG, recommandation, recherche produit
- Requêtes hybrides : combinaison similarité vectorielle + filtres (must/should/must_not)
- Gestion des mises à jour : upsert, suppression, versionnement logique des documents
- Évaluer la pertinence : top-k, seuils, déduplication, gestion des doublons et normalisation
- Atelier pratique : Implémenter une recherche filtrée (par langue, date, catégorie) sur une collection existante.

[Jour 2 - Matin]

Performance et indexation : HNSW, quantization et tuning

- Index HNSW : principes, paramètres clés (m, ef_construct, ef_search) et compromis latence/recall
- Optimiser l'ingestion : batch upsert, parallélisation, dimensionnement et bonnes pratiques
- Stockage et mémoire : segments, compaction, impacts sur les temps de réponse
- Quantization (selon version) : objectifs, cas d'usage, validation de la perte de qualité
- Atelier pratique : Mesurer recall/latence et ajuster ef_search et m pour atteindre un objectif de performance.

[Jour 2 - Après-midi]

Industrialisation : sécurité, exploitation et intégration applicative

- Architecture de déploiement : standalone vs cluster, réplication, sharding et haute disponibilité
- Sécurité et gouvernance : clés/API, isolation réseau, gestion des données sensibles dans le payload
- Supervision : métriques, logs, alerting, diagnostics de requêtes lentes
- Intégration applicative : workflow RAG (chunking, embeddings, retrieval), pagination et reranking (concepts)
- Atelier pratique : Construire un mini-pipeline RAG (ingestion + retrieval) et valider le comportement en production (timeouts, retries, monitoring).

Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d'auto-évaluation nous permettant d'apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d'anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.