drant

Qdrant

Enterprise-Ready, Massive-Scale Vector Search Technology for the Next AI Generation

drant

Vector Search

An essential part of the AI Transformation

INDUSTRIES

HR-Tech | Ad-Tech | Online Dating | Gig Economy

E-Commerce | Law-Tech | Fashion | Med-Tech

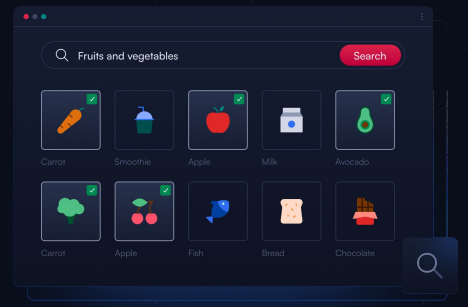
Ed-Tech | Media & News | Biometrics

Agriculture | Manufacturing | Streaming Services

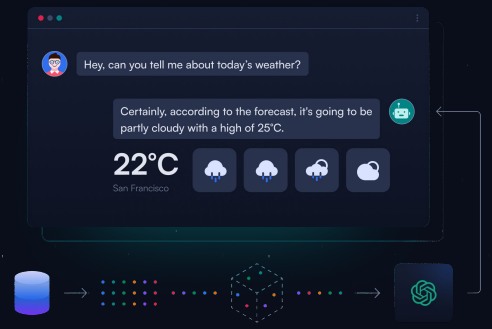
Marketplaces | Anti-fraud

NN Encoders + Vector Database

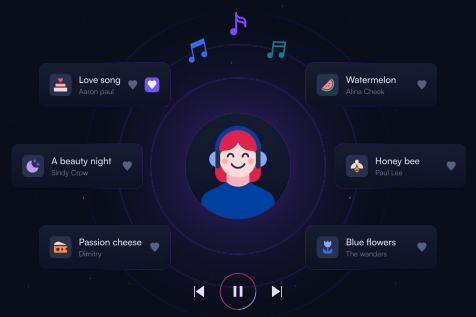
SEARCH SYSTEMS



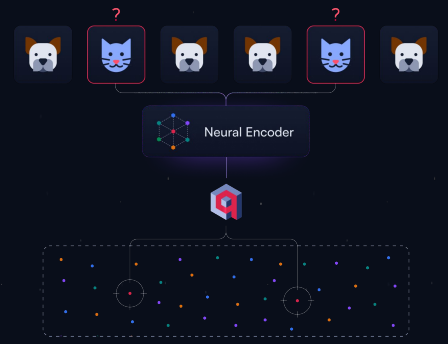
RAG / INFORMATION ASSISTANTS



RECOMMENDATIONS



ANOMALY DETECTION



Getting Started

What is Qdrant?

Qdrant is a vector similarity search engine that provides a production-ready service with a convenient API to store, search, and manage points (i.e. vectors) with an additional payload.

You can think of the payloads as additional pieces of information that can help you hone in on your search and also receive useful information that you can give to your users.

TRACTION

Most Loved

open-source vector search database

>10K
OS Adopters Worldwide

>10M
Downloads

>30K
Community

Need for Dedicated Vector Databases

Key Advantage

As AI applications become production-ready, dedicated vector databases are essential for high-performance, precise data retrieval, and scalable solutions at a reasonable cost, tailored to the demands of modern AI environments.

"Dedicated [vector] databases tend to be fully focused on specific use cases and hence can design their architecture for performance on the tasks needed, as well as user experience, compared to general-purpose databases, which need to fit it in the current design."

- Peter Zaitsev, Percona

Why vector databases are having a moment as the AI hype cycle peaks.

	Dedicated Vector Database	General Database with Vector Add-ons
Specialization	**Highly specialized on vector data type for optimal efficiency.**	Broad focus with secondary vector support.
Performance	**Optimized for high-speed vector operations.**	Performance may lag in vector-specific tasks.
Scalability	**Designed for complex, high-dimensional, large-scale vector data.**	Potential scalability issues with large-scale vector data.
Semantic Search	**Superior handling of unstructured data.**	Less effective with unstructured data.
Cost Effectiveness	**Cost-effective for extensive vector use. Dedicated compression features.**	Higher operational costs at scale.
Community Support	**Specialized community driving innovation focused on vector search.**	Broader community focus, less vector expertise.

qdrant

Key Features

Performance Centric

Scalability Oriented

Quick and easy to start

Resource optimization focussed

Fully Open Source project

All embeddings supported OOTB

qdrant

Performance Centric

Exceptional performance and handling of vectors

- Purpose built for handling high-dimensional and complexity of billion scale vectors.
- Support for dense and sparse vectors
- Optional GPU support for HNSW indexing

Fastest Vector Search Engine

Up to 10 times faster according to benchmarks

drant

Resource Optimization focused

Memory Maps and storage optimizations

Effective on-disk storage options and low level hardware optimization.

Quantization - Scalar, Product, and unique Binary Quantization

Up to 64x Memory usage reduction and x40 search speed up*.

qdrant

High performant search queries

Advanced one-stage Filtering

Store any kind of data including geolocation along with vectors and filter search results during the similarity search without pre- or post-filtering.

Advanced Scalability

Scaling

Scaling options to support data growth and performance.

- Horizontal
- Vertical

Custom sharding

Achieved through Payload-based sharding

Deployment flexibility

Zero downtime

Zero downtime update of vectors in case of model switch (blue-green deployment)

Read/Write

segregation and dynamic read scaling

qdrant

Dedicated Vector DB Solution

- **Pure vector-based Hybrid Search**
  Better performance and accuracy with less resources consumption.

- **Recommendations API**
  Get instant recommendation by just providing positive and negative examples.

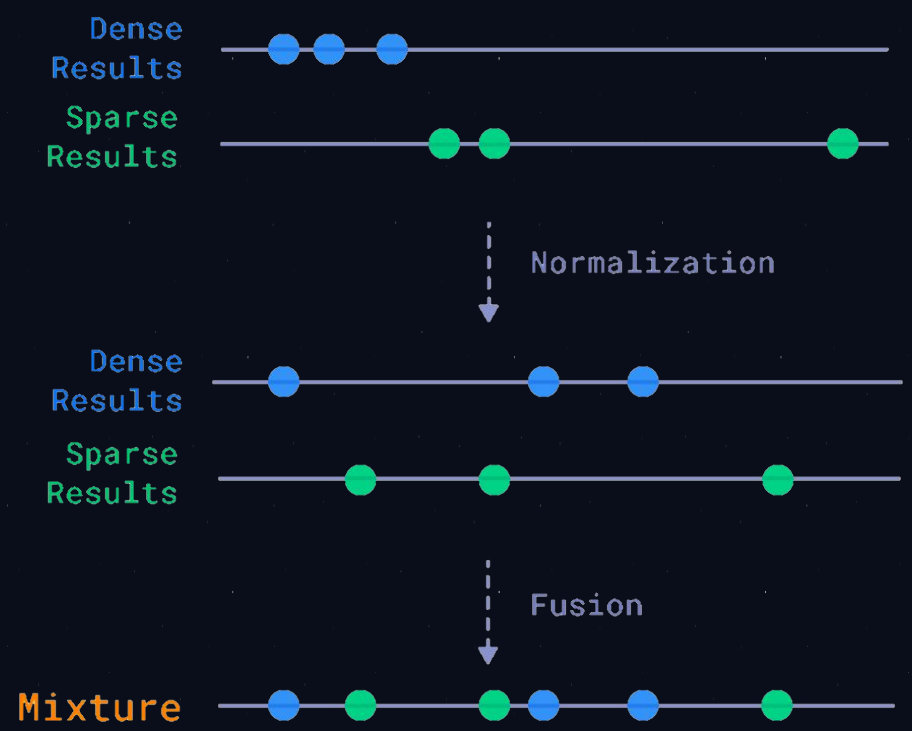
- **Semantic Search as you type**
  Real-time search result suggestions, faster than inference.

- **Beyond similarity search**
  Dissimilarity Search, Diversity search, Discovery.

Think outside of the search box

Hybrid Search

- Combines semantic and keyword search
- Improved retrieval quality
- Based on sparse vectors
- Combination of results with fusion or re-ranking



⚡ FastEmbed

- Fast, Accurate, Lightweight Python library to make State of the Art Embeddings for Qdrant
- Support for a lot of popular models out-of-the box
- Simplify the implementation of Hybrid Search
- Designed for speed
- Great for serverless runtimes (e.g. AWS Lambda)
- Supports multilingual models
- Increased developer productivity
- Fully open-source

https://qdrant.github.io/fastembed/

Enterprise ready

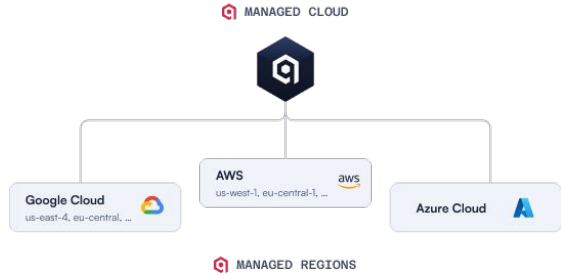
Qdrant Cloud

- Central Cluster Management
- Horizontal/Vertical up and down Scaling
- Automatic shard distribution and re-sharding
- High Availability, Auto-Healing
- Zero-Downtime Upgrades
- Monitoring, Log Management and Alerting
- Backup & Disaster Recovery
- Role based access control
- Enterprise SSO (*)

* Only with Premium subscription

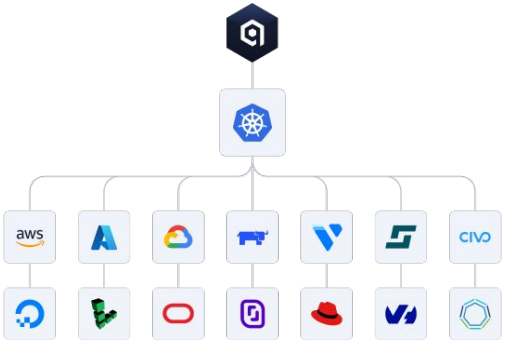
Managed Cloud

Optimal flexibility and a suite of features focused on efficient, scalable vector search - fully managed. Available on AWS, Google Cloud, and Azure.



Hybrid Cloud

Bring your own Kubernetes clusters from any cloud provider, on-premise infrastructure, or edge locations - fully managed.



Private Cloud

Experience maximum control and security by deploying Qdrant in your own infrastructure or edge locations.

qdrant

Customer Success Stories

DAILYMOTION

Recommendation System

Dailymotion leveraged Qdrant to optimize its **video recommendation engine**, managing over 420 million videos and processing 13 million recommendations daily. This enhancement **reduced content processing times from hours to minutes** and **increased user interactions and click-through rates by more than 3x**.
Case Study

DUST

RAG

Dust leveraged Qdrant to significantly enhance its AI platform, utilizing retrieval augmented generation (RAG) to enable **rapid data retrieval times in milliseconds**, **reduce operational costs by 50%**, and improve system scalability and performance
Case Study

VISUA

Anomaly Detection

Visua, a leader in high-volume computer vision data analysis, leveraged Qdrant to significantly improve their quality control process with anomaly detection. With Qdrant they achieved **40x faster query processing** and were able to **handle 10x more data** in the quality assurance and learning processes, **enhancing the accuracy** of its algorithms. Case Study

Qdrant powers thousands of top AI solutions

BOSCH
Johnson & Johnson
thoughtworks
moz://a
perplexity
Voic
Tripadvisor
BAYER
Deloitte.
Discord
DAILYMOTION
Kau
CBINSIGHTS
Hewlett Packard Enterprise
Flipkart