

Unlocking the Power of Vector Search with Zilliz Cloud

Speaker



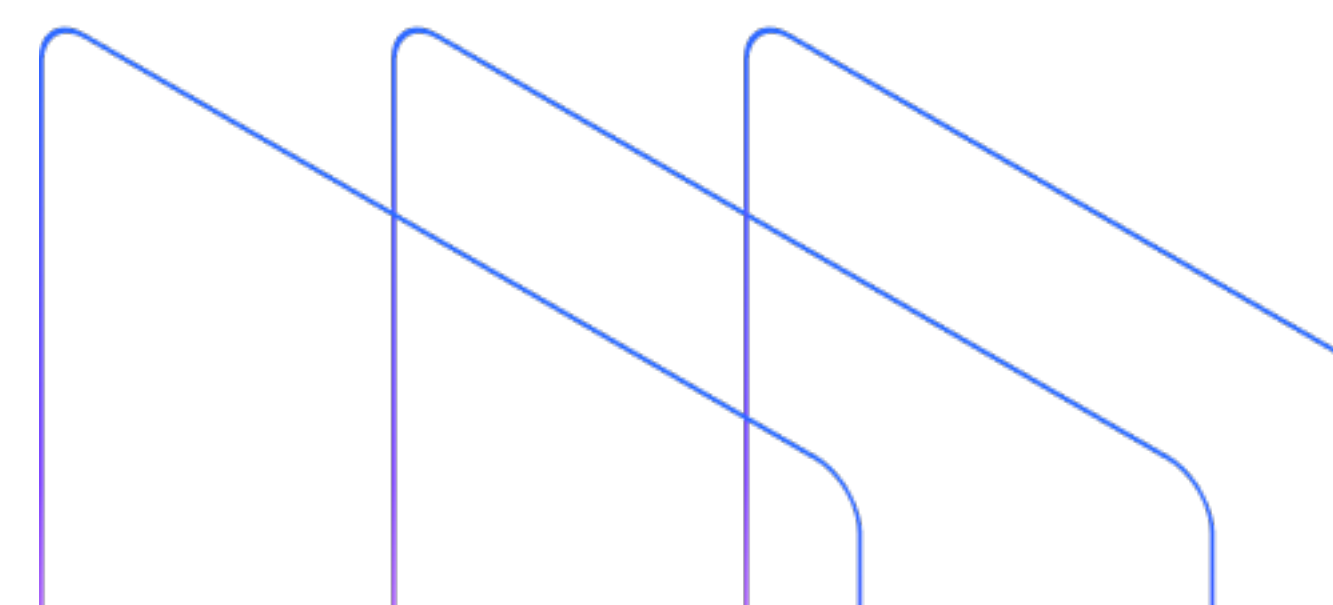
Frank Liu

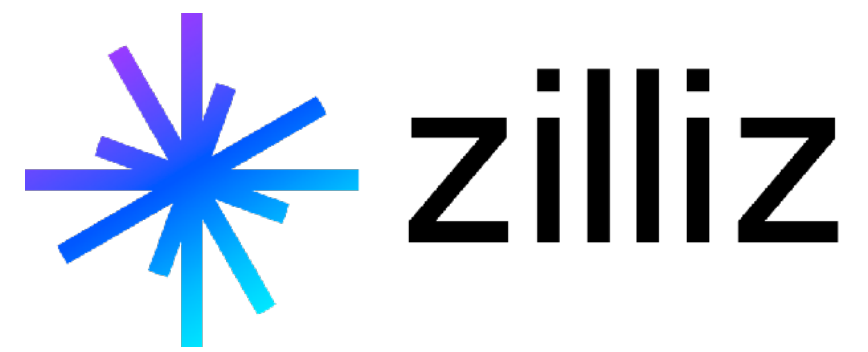
Director of Operations
ML Architect

frank@zilliz.com

[linkedin.com/in/fzliu](https://www.linkedin.com/in/fzliu)

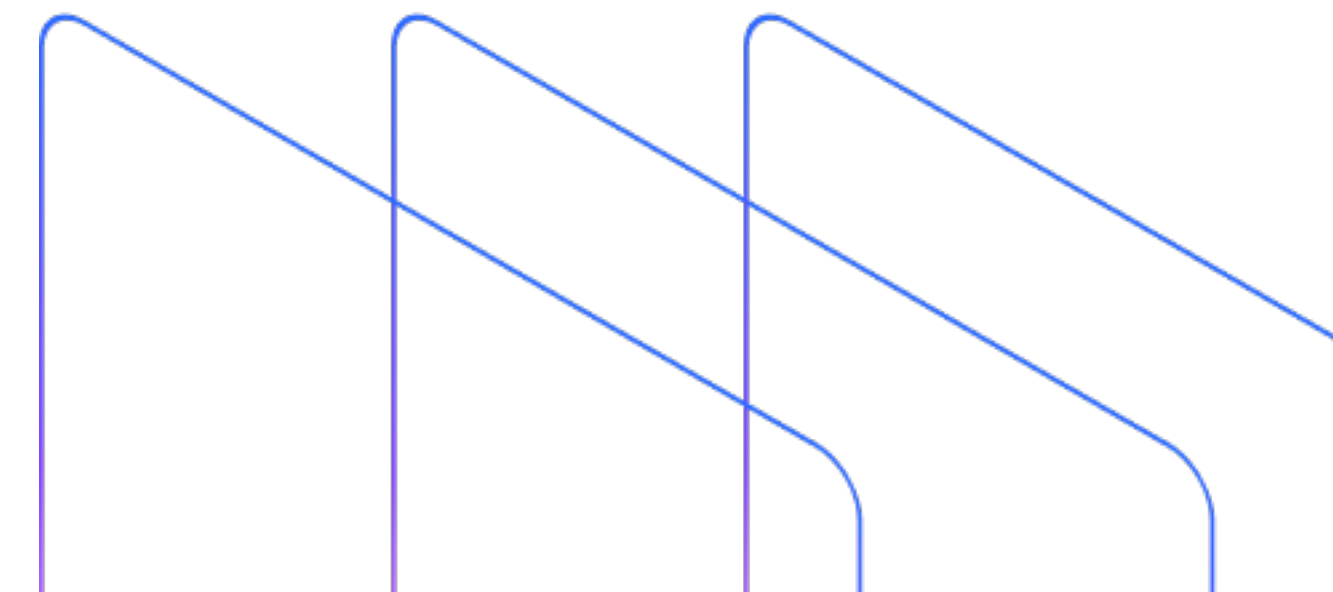
Twitter @frankzliu





Zilliz Cloud is a fully-managed Milvus vector database service, made by the creators of Milvus. It simplifies the process of deploying and scaling vector search applications by eliminating the need to create and maintain complex data infrastructure.

- Powerful, flexible support for multiple Machine learning algorithms
- Lightning-fast queries on any size data set
- Cost-effective storage of vectors
- Zero ops overhead



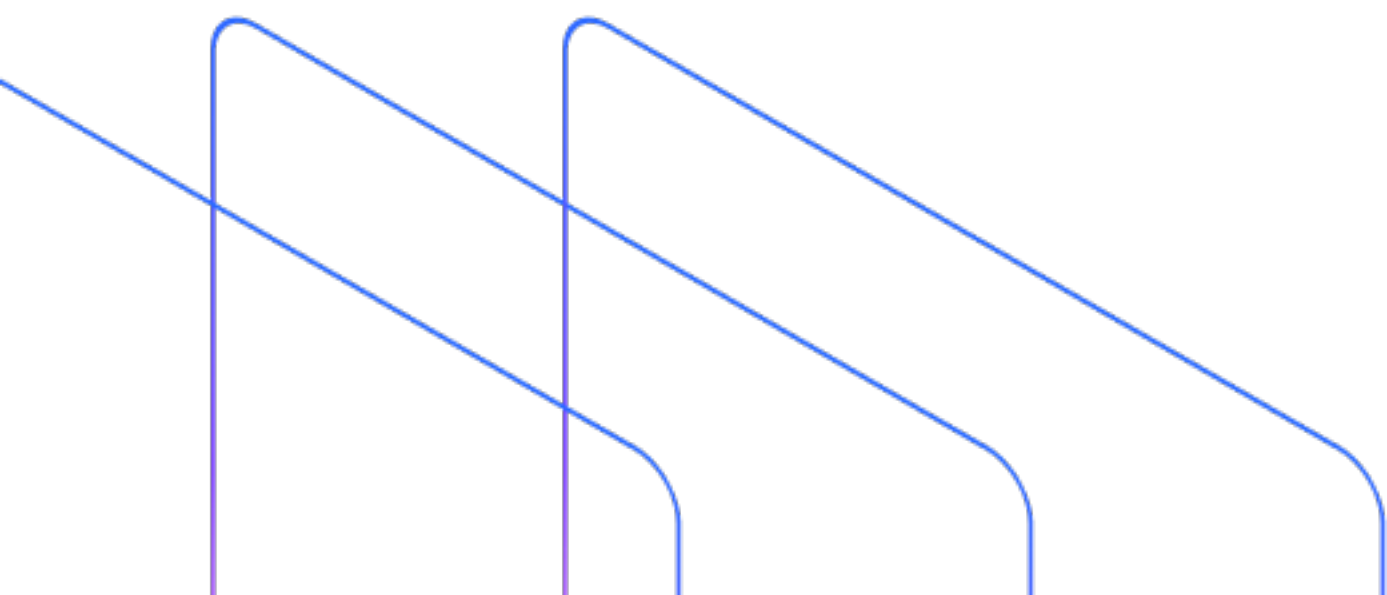
01 Introduction to Zilliz Cloud

02 Features and Functionality

03 Migrating from Milvus

04 Use cases

05 A Quick Demo



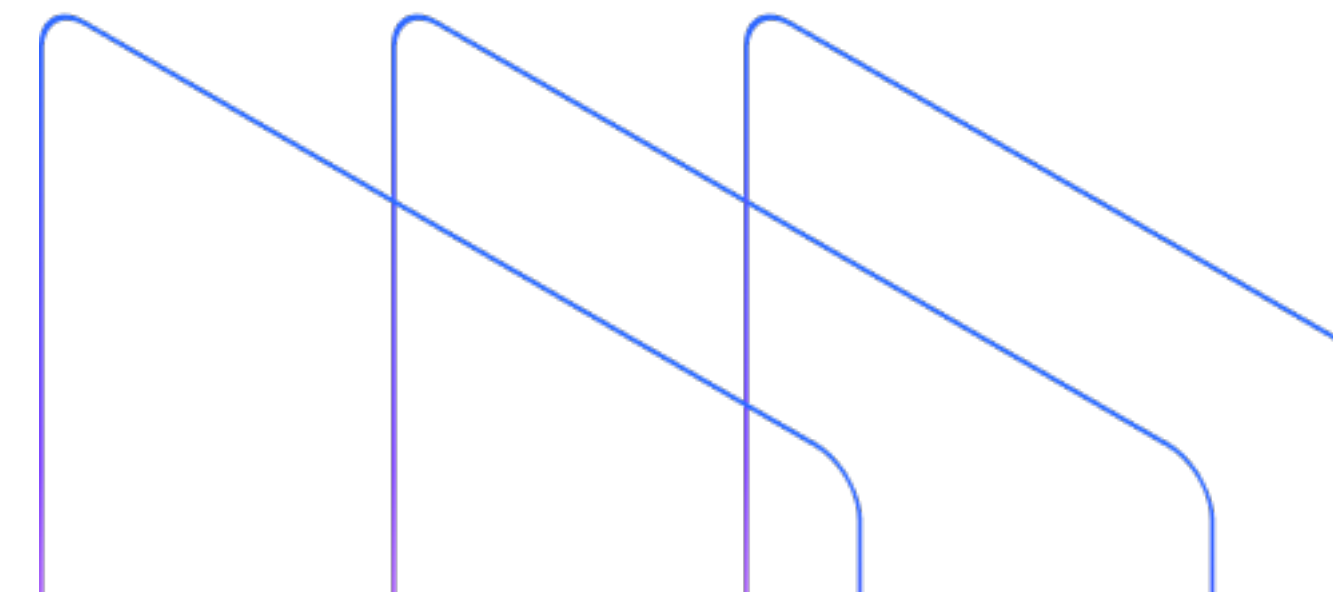
01

Introduction to Zilliz Cloud

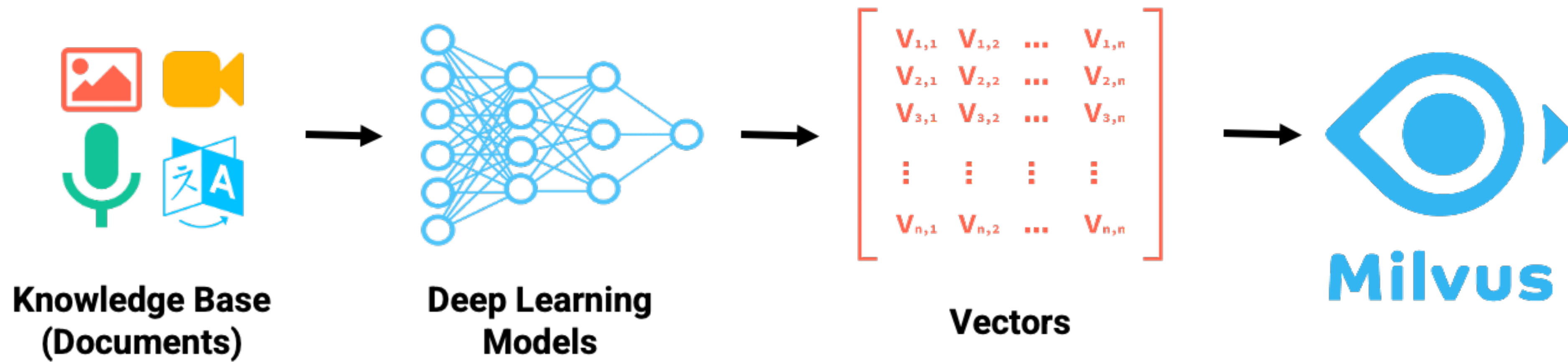


What is Zilliz Cloud?

- **Built on Milvus and optimized for performance**
- **Elastic and Scalable**
- **Pay-as-you-go**
- **Multi-Cloud (AWS, GCP)**
- **Cloud Native Resiliency** with 99.99% uptime SLA, and zero data corruption
- **Enterprise Security & Governance**
- **Integrated with key LLMs** (OpenAI, Hugging Face, LangChain, Cohere, PyTorch)
- **Supported SDKs** (Python, Java)



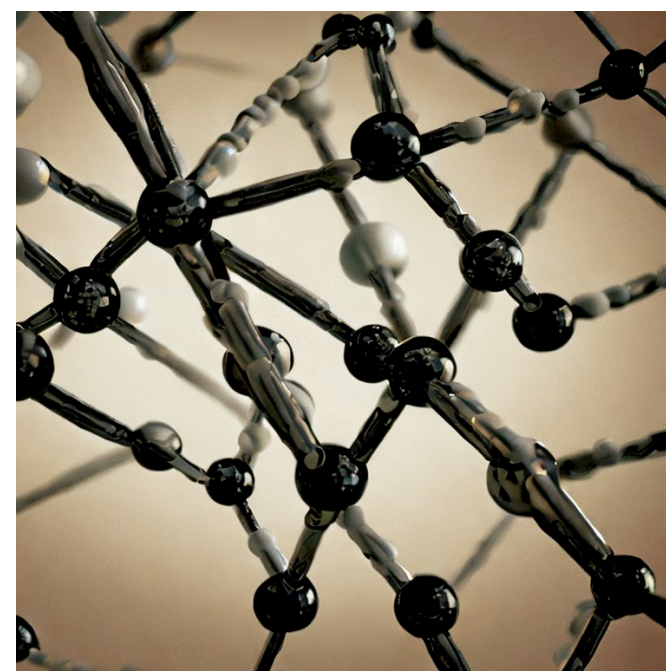
Milvus Refresher



Applications



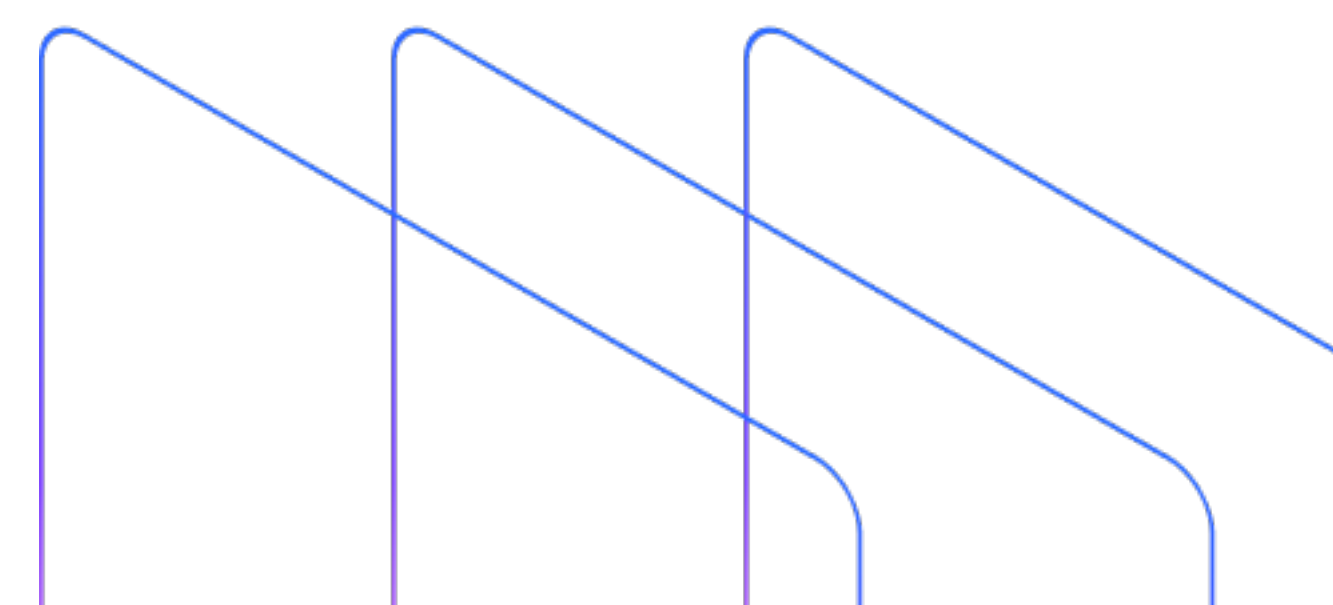
Semantic Text Search



Molecular Similarity Search



Image Search



Applications - Using with LLMs



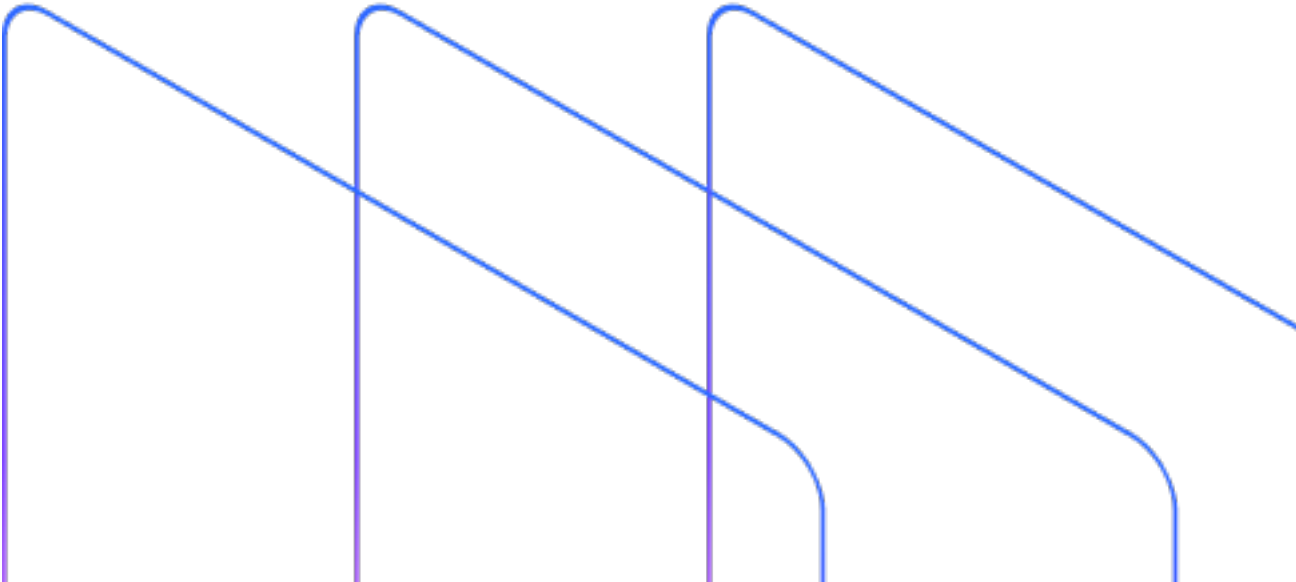
Claude

Bard

Experiment

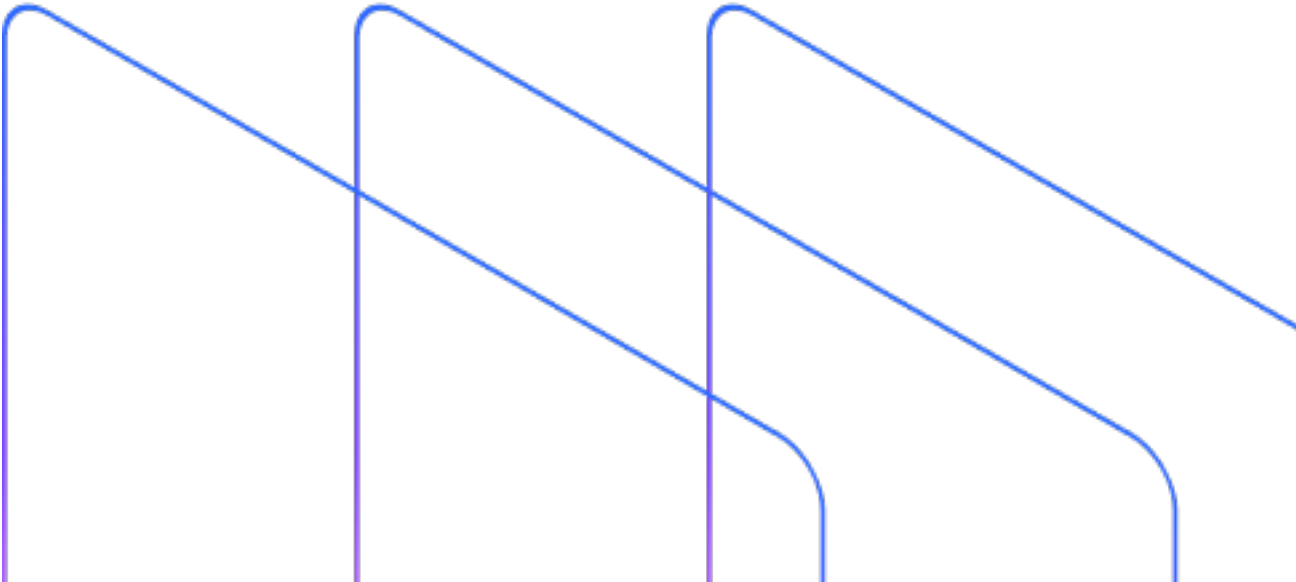
Why Choose Zilliz over Milvus?

	Zilliz	Milvus
Vector Similarity Search Efficiently store and search vectors for similarity. Fine-tuned parameters and software/hardware adaption for better performance.	✓	✓
Hybrid Search Combine vector similarity search with advanced filtering to match your search condition requirements.	✓	✓
Rich Schema Support Easily work with a variety of data types, ie. float/binary vector, integer, floating point, boolean, varchar, etc., for flexible schema modeling.	✓	✓
Auto ANNS Index No more suffering on matching your workload to the right index type, and the endless optimizations. With auto-index, you can just focus on the business, and Zilliz Cloud will handle the rest.	Auto-index type	Manual index type selection
Compatible SDKs	Python and Java SDK Fully supported	Community version only



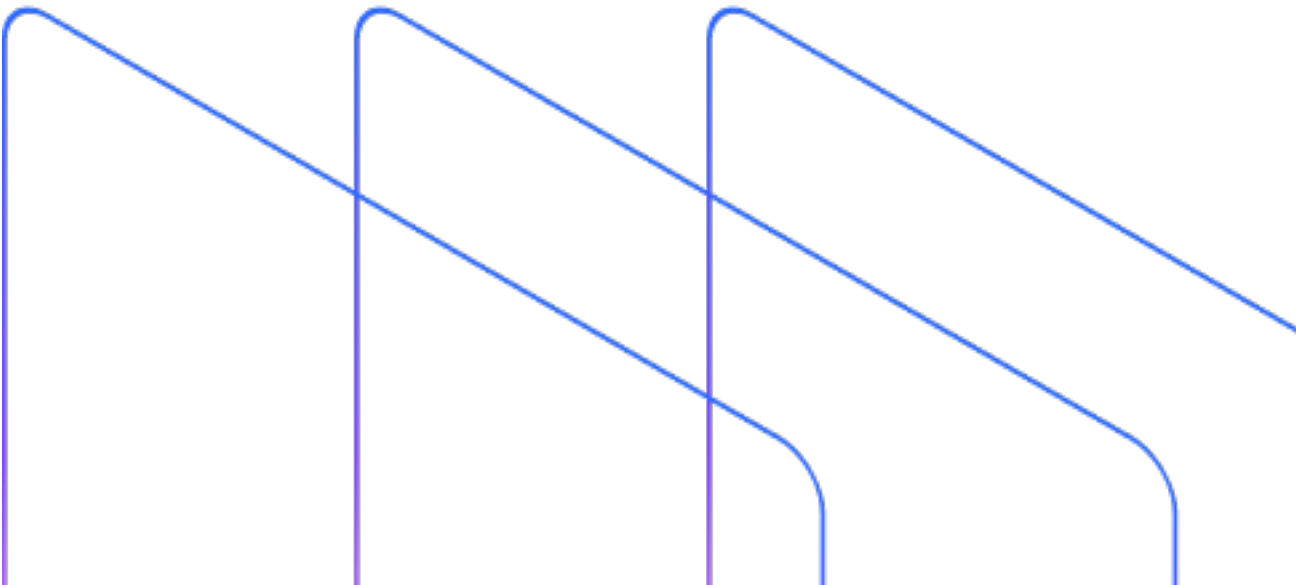
Why Choose Zilliz over Milvus?

	Zilliz	Milvus
Fully Managed Milvus clusters that are fully managed and automated with zero operations required.	✓	—
High Availability 99.99% uptime SLA with built-in failover to ensure your Milvus clusters are always available.	✓	—
Elastic Scaling Easily scale up to a billion-scale or down without the need to over-provision infrastructure.	✓	—
Different Machine types for best price performance provides performance-optimized and capacity-optimized compute unit types to fit different use cases.	✓	—
Infinite Storage Cost-effectively store data at any scale without the need to increase compute resources.	✓	—
Cloud UI User-friendly GUI to easily manage and monitor your Milvus clusters at any scale in the cloud.	✓	—
Resource monitors Get automatic notifications to avoid service overloading.	✓	—



Why Choose Zilliz over Milvus?

	Zilliz	Milvus
24×7×365 Expert Support Access to the world’s foremost Milvus experts to offer you 24×7×365 support.	✓	—
SOC2 Type II Our platform is compliant with SOC 2, Type 2 so that you can be confident in the integrity of your data.	✓	—
Data Recovery Backup and restore, time travel and recycler bin to help restore data in the event of accidental loss.	✓	✓
Data Encryption in Transit Ensure the security of your data while it’s being transferred.	✓	✓
Role-Based Access Control (RBAC) Permission management to protect private and sensitive information and control user access.	✓	✓



02

Features and Functionality



Zilliz Cloud Basics

- "Capacity Unit" (CU)
 - A single unit of compute
 - Higher CUs = more storage and compute
 - Capacity-optimized and compute-optimized
- "Project"
 - Each DB is assigned to a project
- "Collection"
 - Database "table" in Zilliz Cloud
 - Container for vectors generated by the same model
 - Specific vector dimensionality and schema
- Supported SDKs
 - Python, Java

Create Database [View User Guide](#)

Name: Project:

Cloud Provider:

Cloud Region:

CU Type: Size:

User Name: Password:

Cloud Backup: ☐ Off

[Cancel](#) [Create](#)

Summary

Cloud Provider:
Region: us-west-2
CU Type: Performance-optimized
Size: 1 CU

Price

Compute: \$0.259 / hour
Storage: \$0.025 per GB/month

ⓘ Databases will be automatically suspended after 7 days of inactivity.

AWS and GCP Support

- Select a preferred cloud provider
 - Different providers have different pricing structures
- Also available on AWS marketplace

Create Database

[View User Guide](#)


Name


Alphanumeric letters and hyphens only


Project

Default project

Cloud Provider

aws

Google Cloud

Azure

Cloud Region

us-west-2

CU Type

Performance-optimized

Size

1 CU

User Name

db_admin

Password

Cloud Backup

☐ Off

Cancel

Create

Summary

Cloud Provider	
Region	us-west-2
CU Type	Performance-optimized
Size	1 CU

Price

Compute	\$0.259 / hour
Storage	\$0.025 per GB/month

Databases will be automatically suspended after 7 days of inactivity.

Billion-scale Collections

- Performance-optimized
 - Low-latency, high-throughput
 - Maximum 256 CU
 - Millisecond query response times
- Capacity-optimized
 - High storage capacity
 - Maximum 128 CU
 - Up to 3.2 billion vectors

Create Database

[View User Guide](#)

Name

Alphanumeric letters and hyphens only

Project

Default project

Cloud Provider

aws

Google Cloud

Azure

Cloud Region

us-west-2

CU Type

Performance-optimized

Size

1 CU

User Name

db_admin

Password

Cloud Backup

Off

Cancel

Create

Summary

Cloud Provider	
Region	us-west-2
CU Type	Performance-optimized
Size	1 CU

Price

Compute	\$0.259 / hour
Storage	\$0.025 per GB/month

Databases will be automatically suspended after 7 days of inactivity.

User-friendly UI

Default project > test-db

test-db

RUNNING

Connection Guide

Actions

Database Details

Collections

Metrics

Users

Backups

Migrations

Choose a way to create your collection

Use Your Own Data

After you have prepared your dataset, click Create New to continue.

Create New

Use Example Data

Create a collection using the data that Zilliz Cloud offers.

Import Example Data

Privatelink Support

Add Private Link

Cloud Provider

Select Cloud Provider

aws

AWS

Region

Select cloud region

us-west-2

☒

Create VPC endpoint

☐

Already have VPC endpoint

Your VPC ID

You can find this in your list of VPCs in the VPC dashboard in your AWS account.

[Show Instruction](#)

Your Subnet IDs

You can find this in your list of Subnets in the Subnet dashboard in your AWS account.

[Show Instruction](#)

Run this command with the AWS_CLI to create your VPC endpoint.

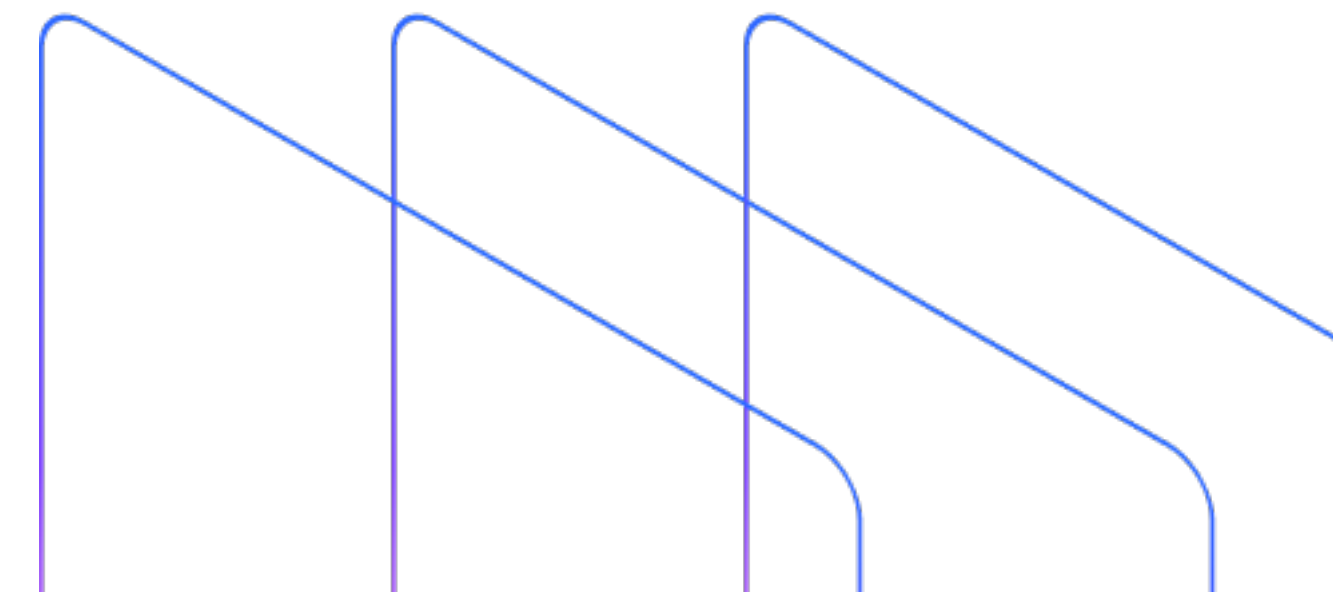
aws ec2 create-vpc-endpoint --vpc-id --region us-west-2 --service-name com.amazonaws.vpce.us-west-2.vpce-svc-08b59961986ce8e31 --vpc-endpoint-type Interface --subnet-ids

Cancel

Next

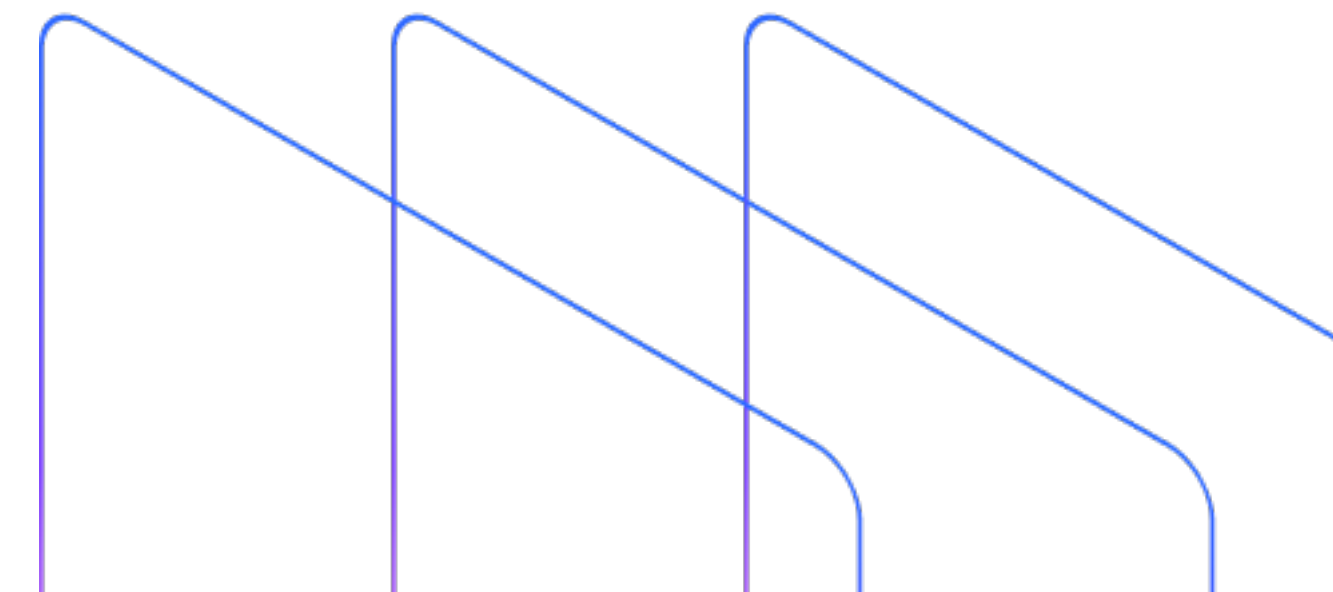
Rolling Upgrade

- Zilliz Cloud supports rolling upgrades
 - No scheduled maintenance requires
 - Rolling upgrade downtime <1 min/month
- Latest and greatest vector search experience
 - Focus on your application and business needs instead!



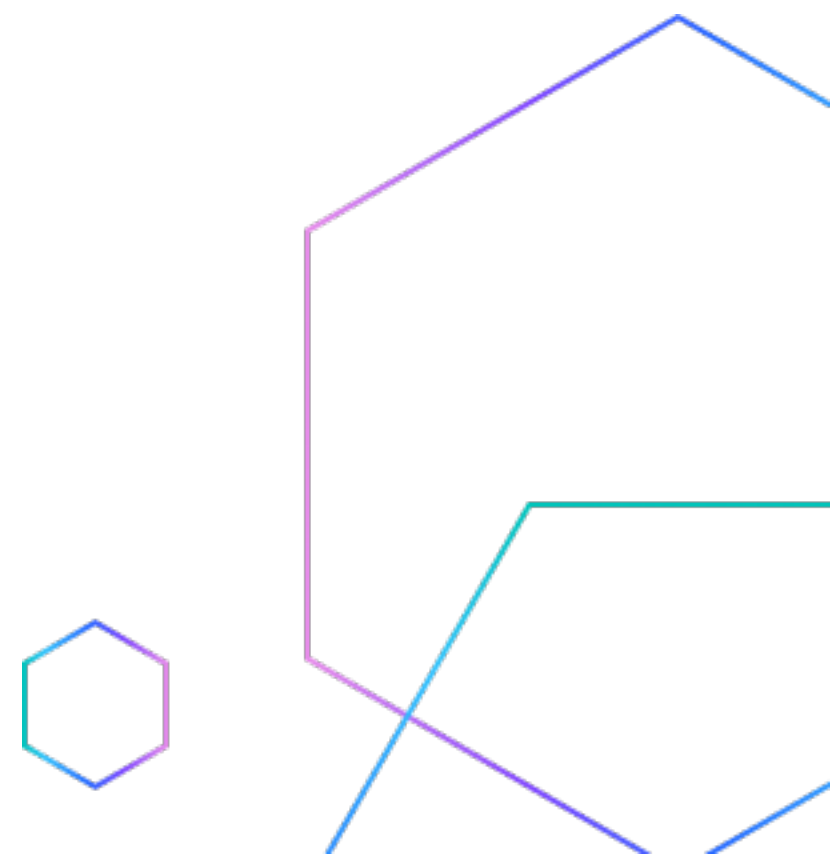
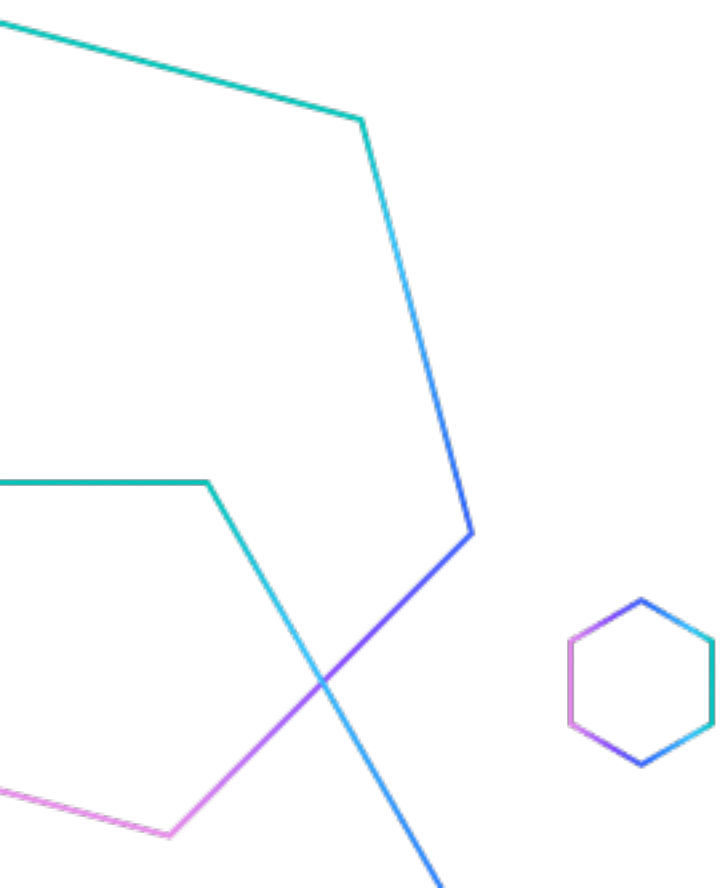
Backup and Restore

- Data loss can occur for a variety of reasons
 - Consequences can be devastating
 - Reliable backup and recovery solution is a necessity
- Zilliz cloud supports backup and restore
 - One-click solution to database recovery



03

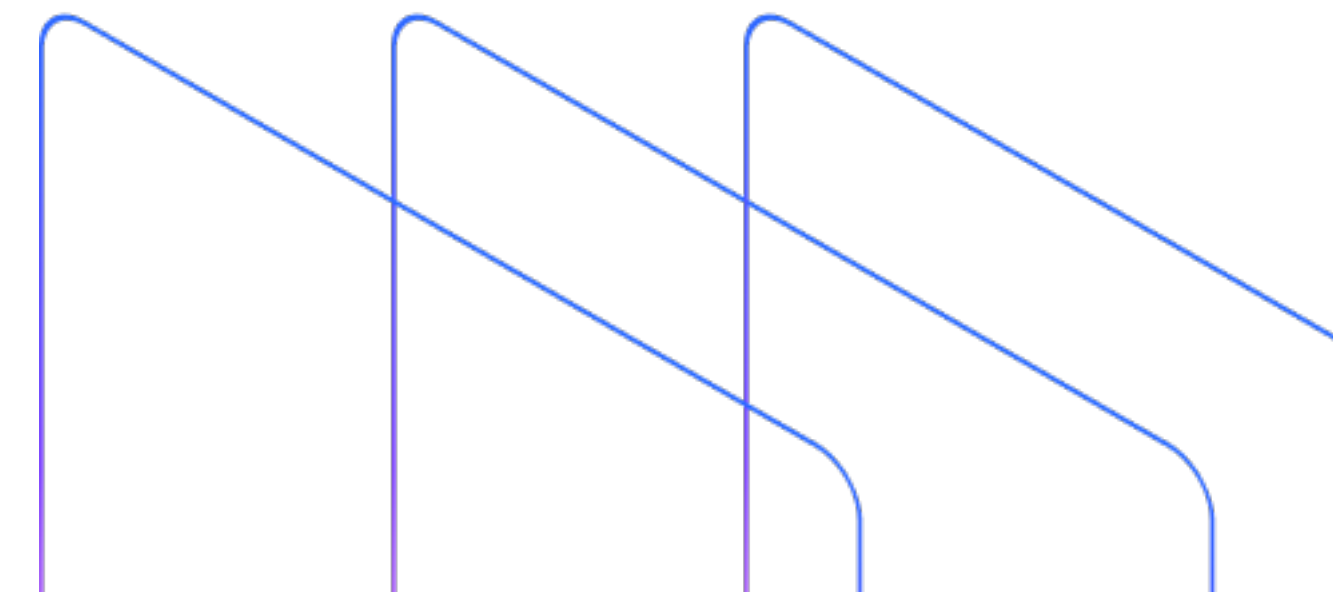
Migrating from Milvus



Step 1: Preparation

Download [milvus-backup](https://github.com/zilliztech/milvus-backup/releases/download/v0.2.2/milvus-backup_Darwin_x86_64.tar.gz)

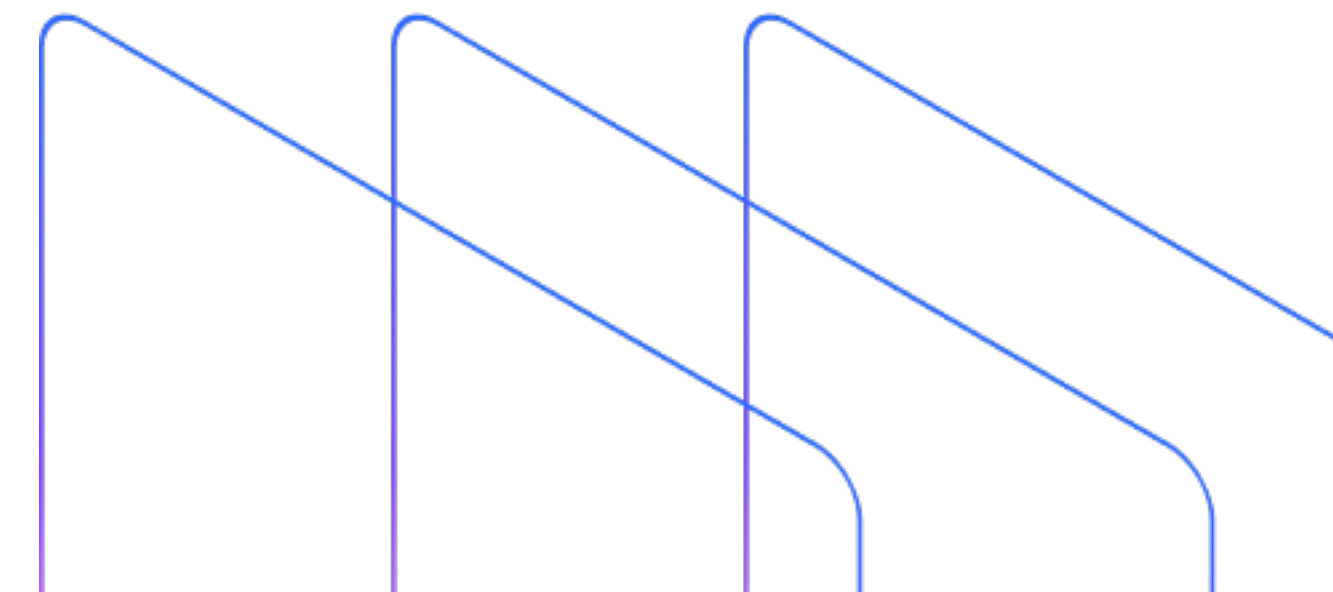
```
wget https://github.com/zilliztech/milvus-backup/releases/download/v0.2.2/milvus-backup_Darwin_x86_64.tar.gz
```



Step 2: Backup

Create a backup of your Milvus installation

```
./milvus-backup -config backup.yaml create -n my_backup  
./milvus-backup -config backup.yaml get -n my_backup
```




Step 3: Import

Migrate From Files

Type ⓘ

▼

☒ Import a local folder☐ Import a folder from S3



Drag and drop a folder here
or [upload a folder](#)

ⓘ

Maximum upload folder size: 1 GB. Try importing from S3 bucket instead to upload larger files.

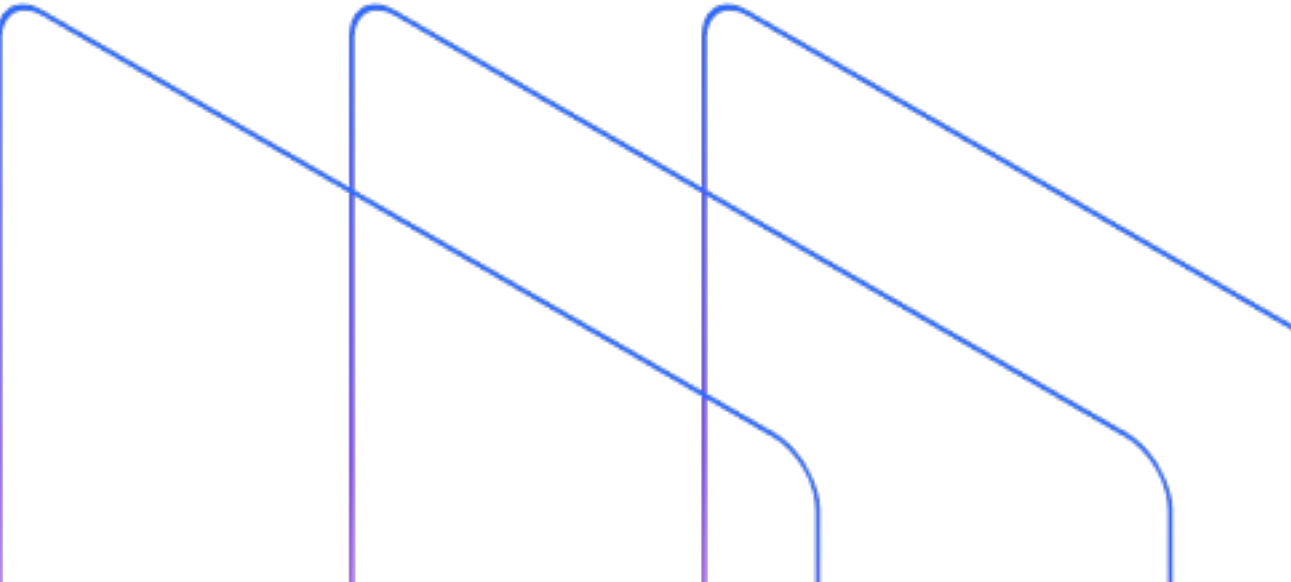
The upload folder name can contain uppercase letters, lowercase letters, numbers, hyphens ("-"), underscores ("_"), and square brackets ("[]").

The name of the upload files in the folder cannot contain colons (":").

Cancel

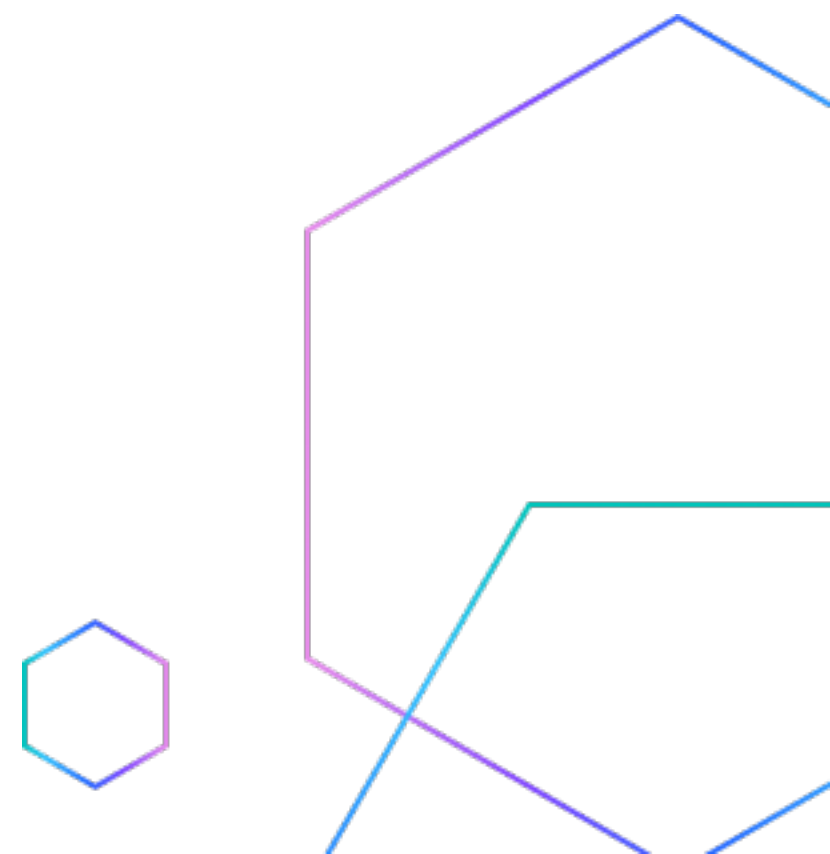
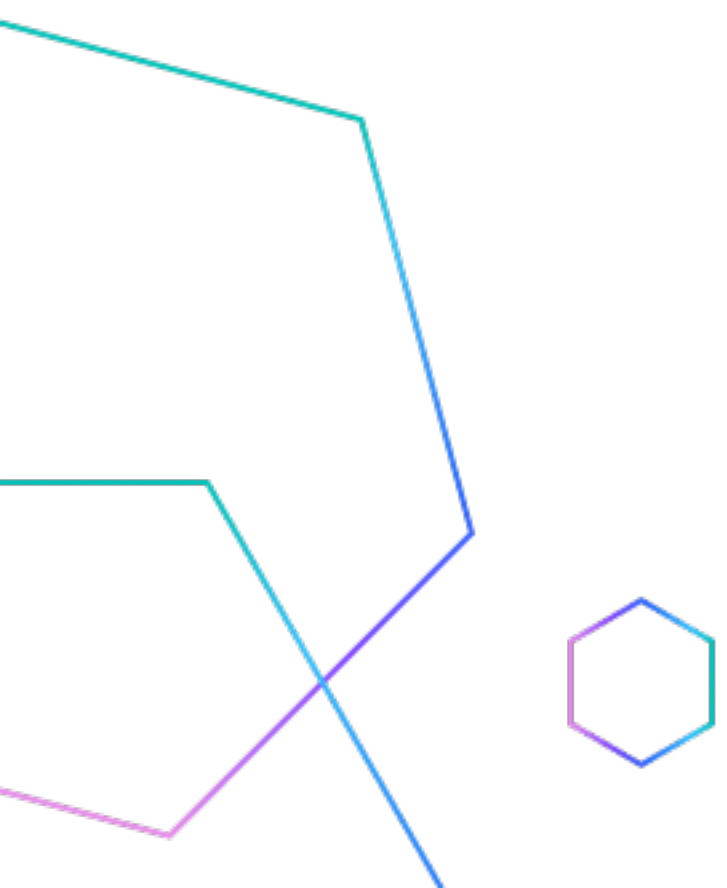
Migrate

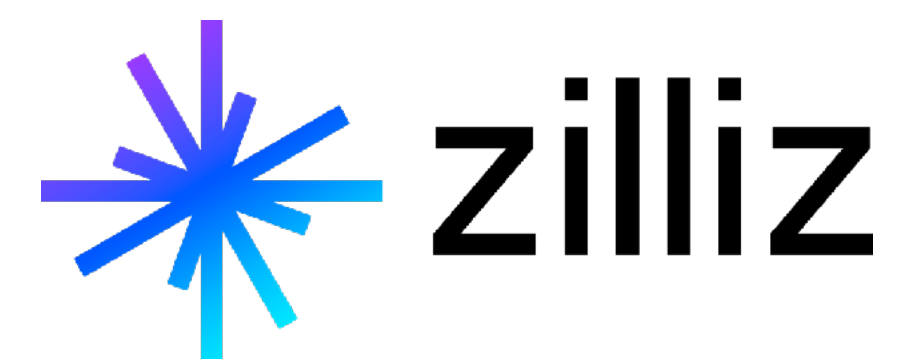
Drag and drop backup folder here



05

A Quick Demo





Try Zilliz Cloud free today at zilliz.com/cloud

 [@Zilliz_Universe](https://twitter.com/Zilliz_Universe)

 linkedin.com/in/zilliz

 milvusio.slack.com

 github.com/zilliztech