zilliz

Unlocking the Power of Vector Search with Zilliz Cloud

Speaker

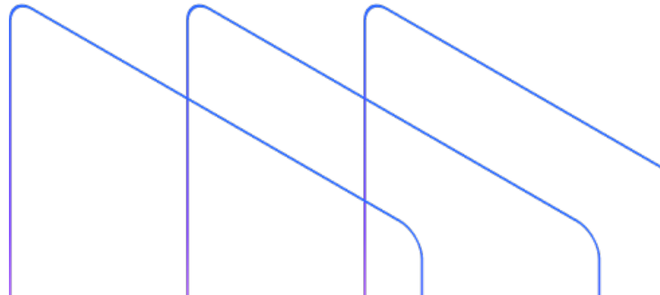
Frank Liu

Director of Operations
ML Architect

frank@zilliz.com
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



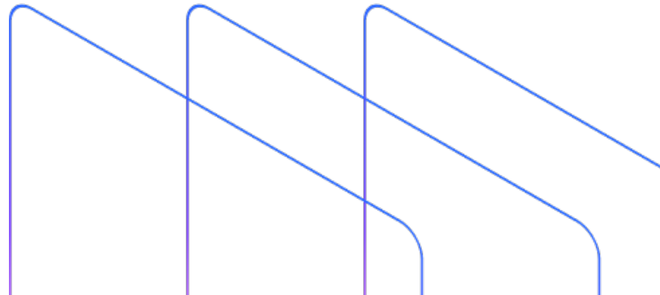
CONTENTS

01

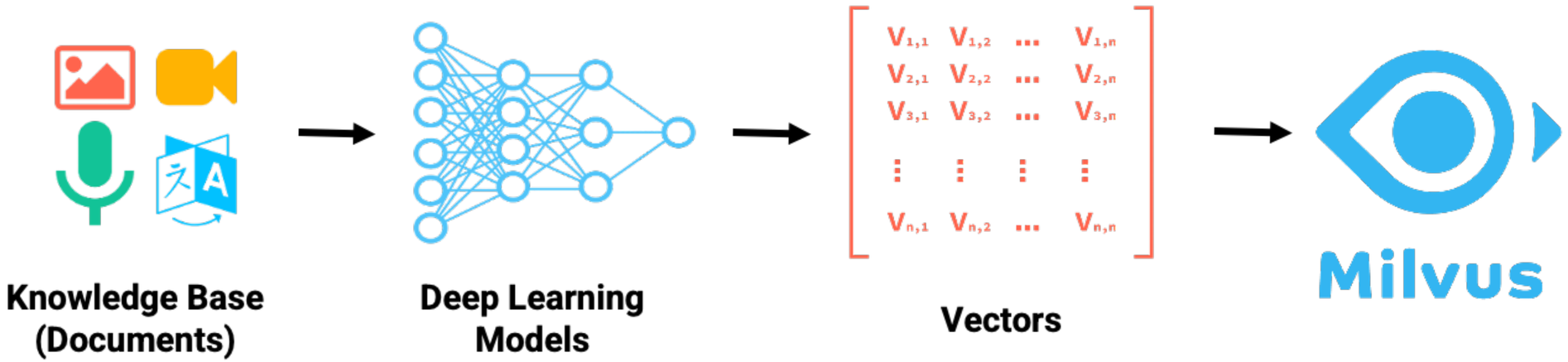
Introduction to Zilliz Cloud

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

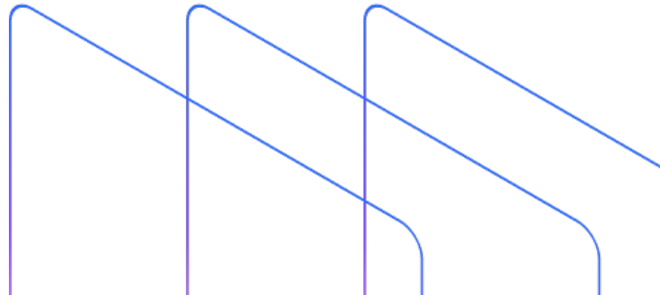
Applications - Using with LLMs

Claude

Bard Experiment

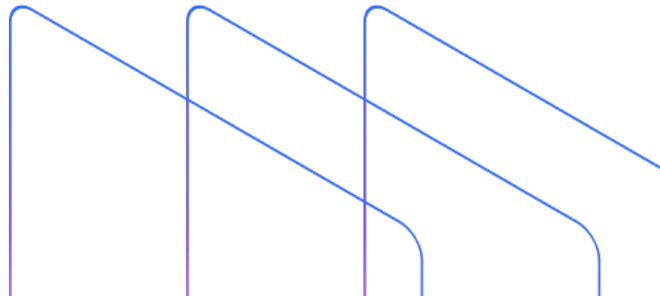
Why Choose Zilliz over Milvus?

	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



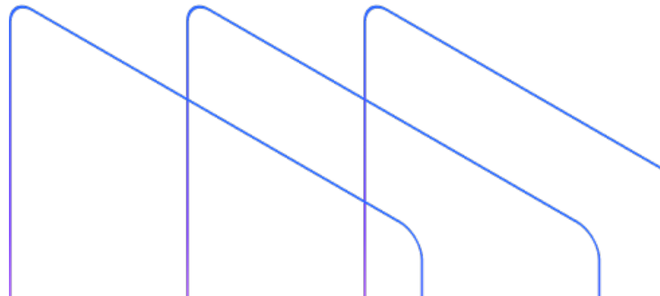
Why Choose Zilliz over Milvus?

	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.	✓	✓

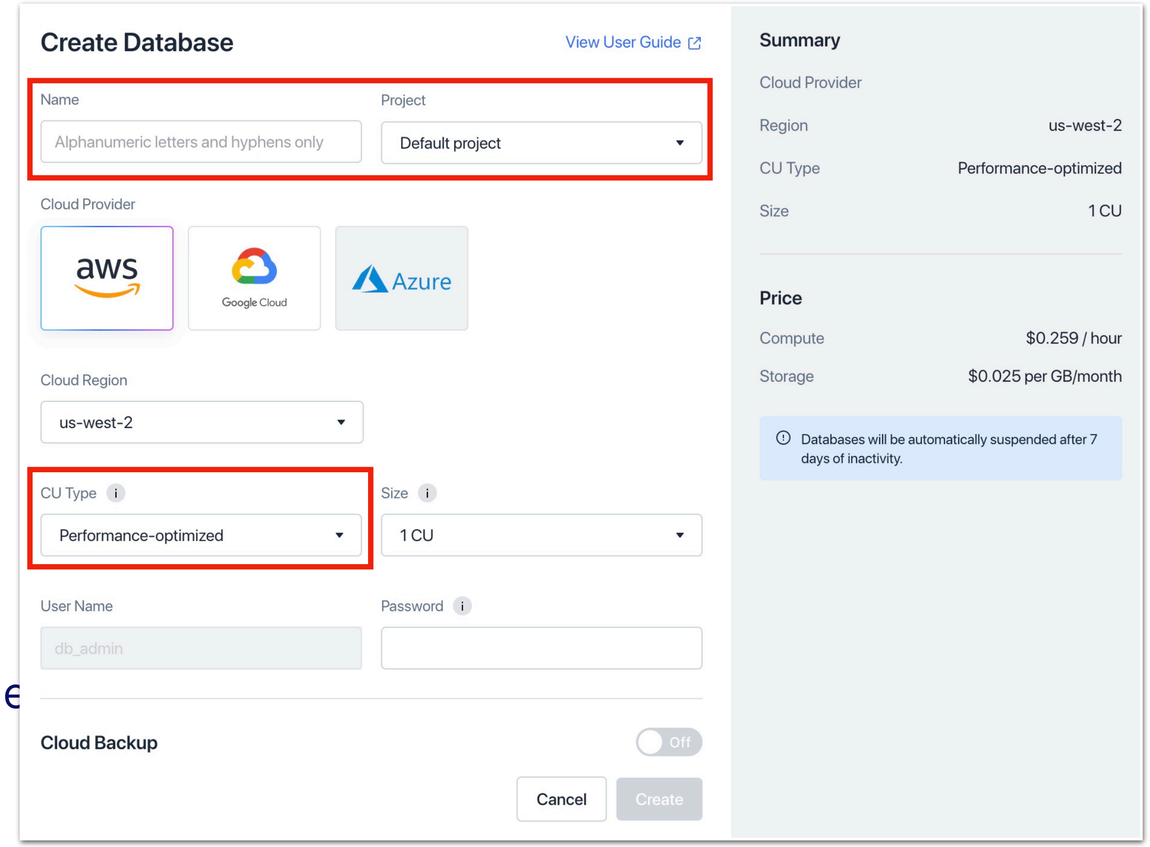


02

Features and Functionality

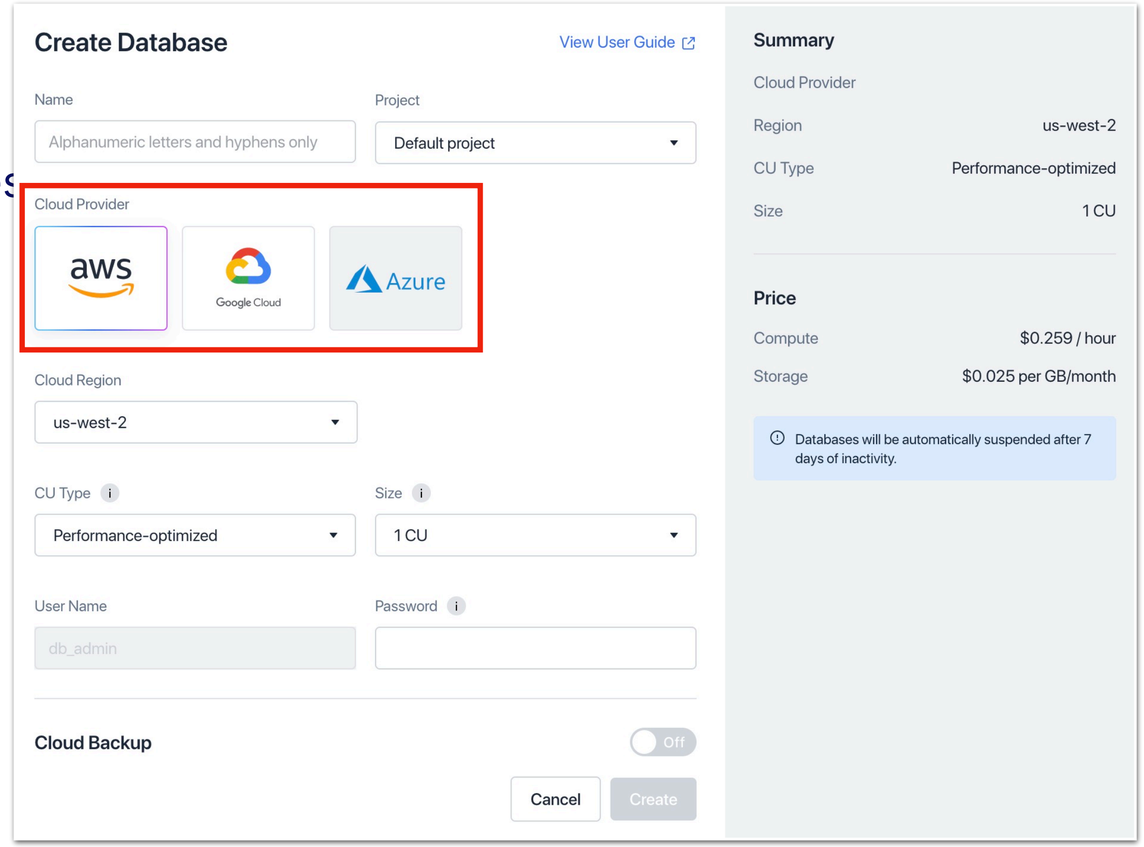
Zilliz Cloud Basics

- "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 mode
  - Specific vector dimensionality and schema
- Supported SDKs
  - Python, Java



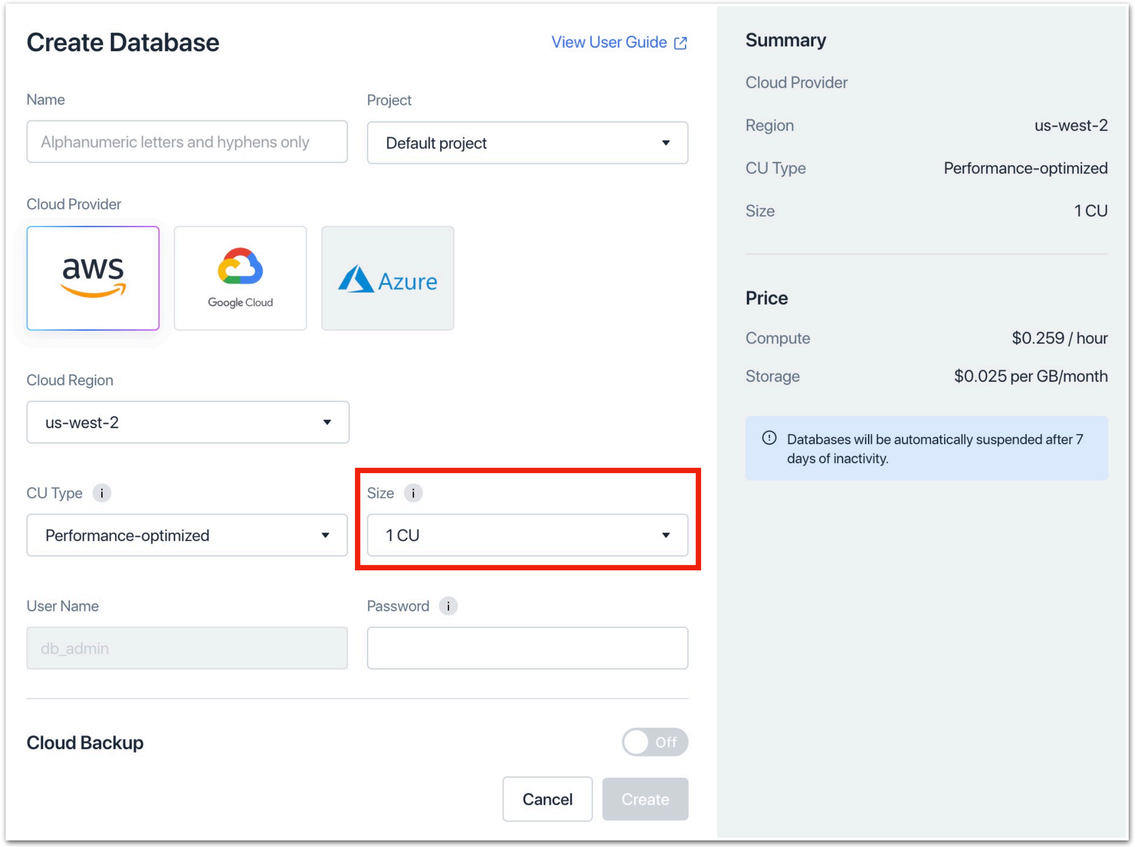
AWS and GCP Support

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

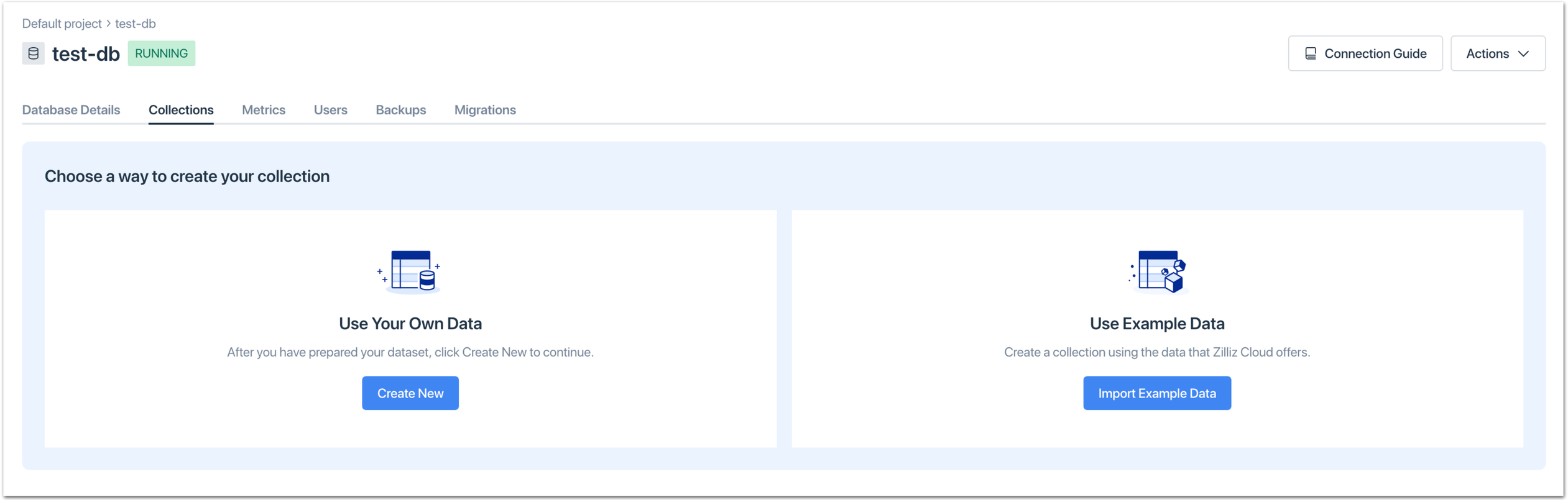


Billion-scale Collections

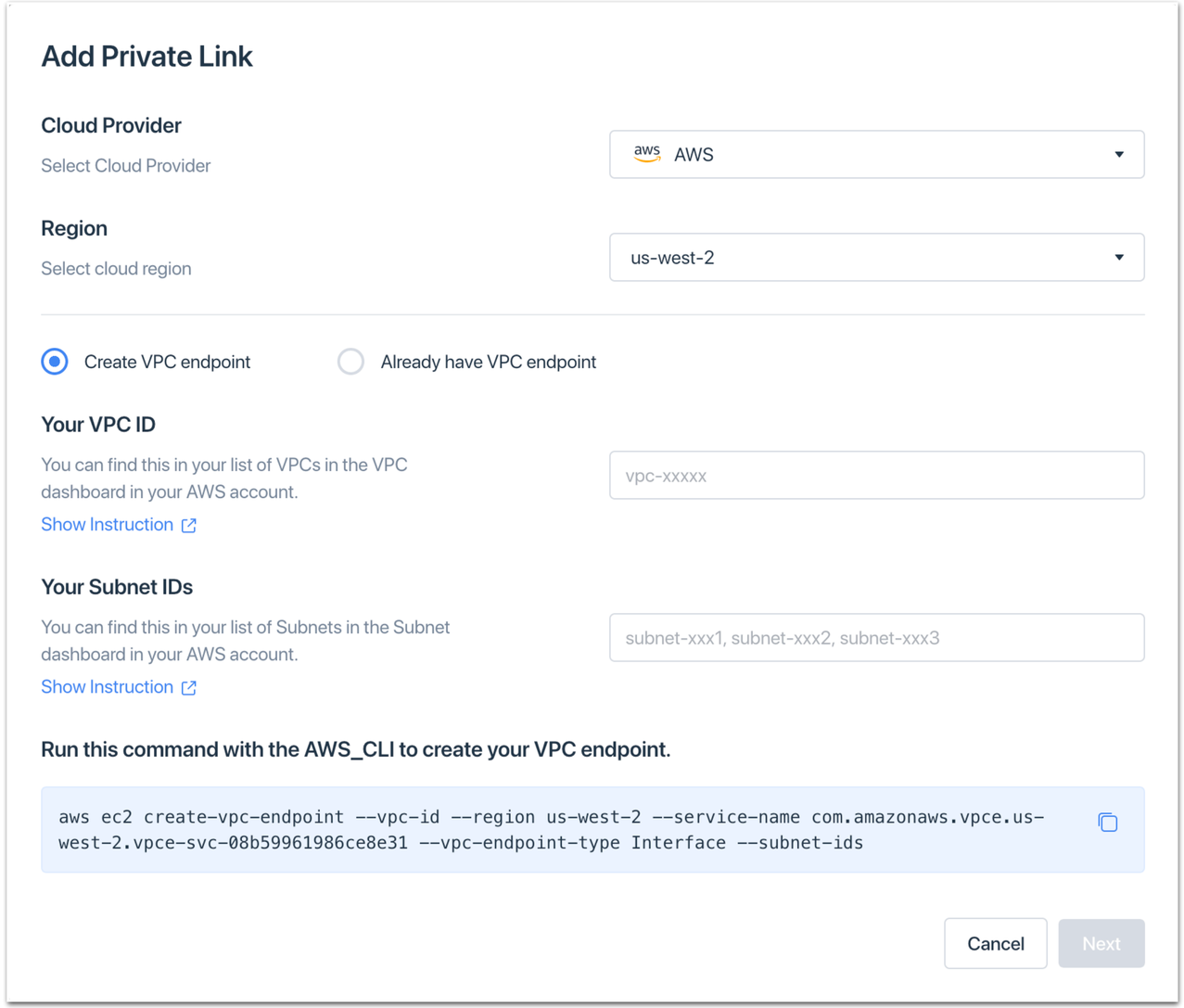
- Performance-optimized
  - Low-latency, high-throughput
  - Maximum 256 CU
  - Millisecond query response times
- Capacity-optimized
  - High storage capacity
  - Maximum128 CU
  - Up to 3.2 billion vectors



User-friendly UI

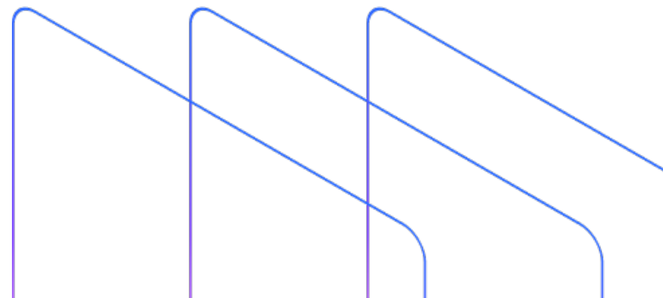


Privatelink Support



Rolling Upgrade

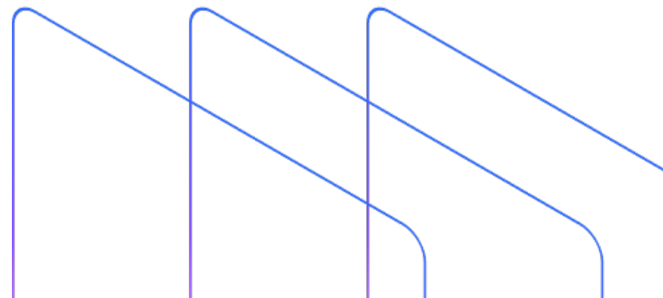
- Zilliz Cloud supports rolling upgrades
  - No scheduled maintenance requires
  - Rolling upgrade downtime <1 min/month
- Latest and greatest vector search experience
  - Focus on your application and business needs instead!



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



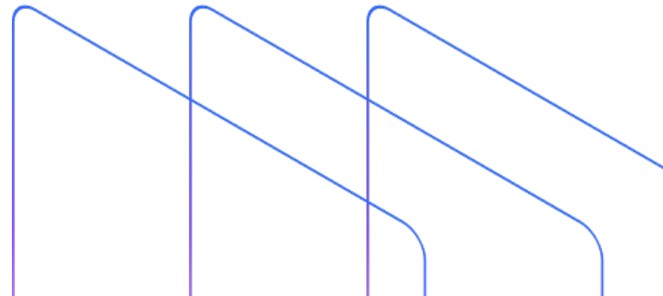
03

Migrating from Milvus

Step 1: Preparation

Download **milvus-backup**

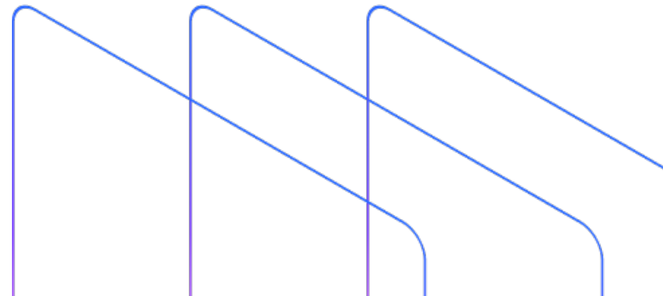
```
wget https://github.com/zilliztech/milvus-backup/releases/download/v0.2.2/milvus-backup_Darwin_x86_64.tar.gz
```



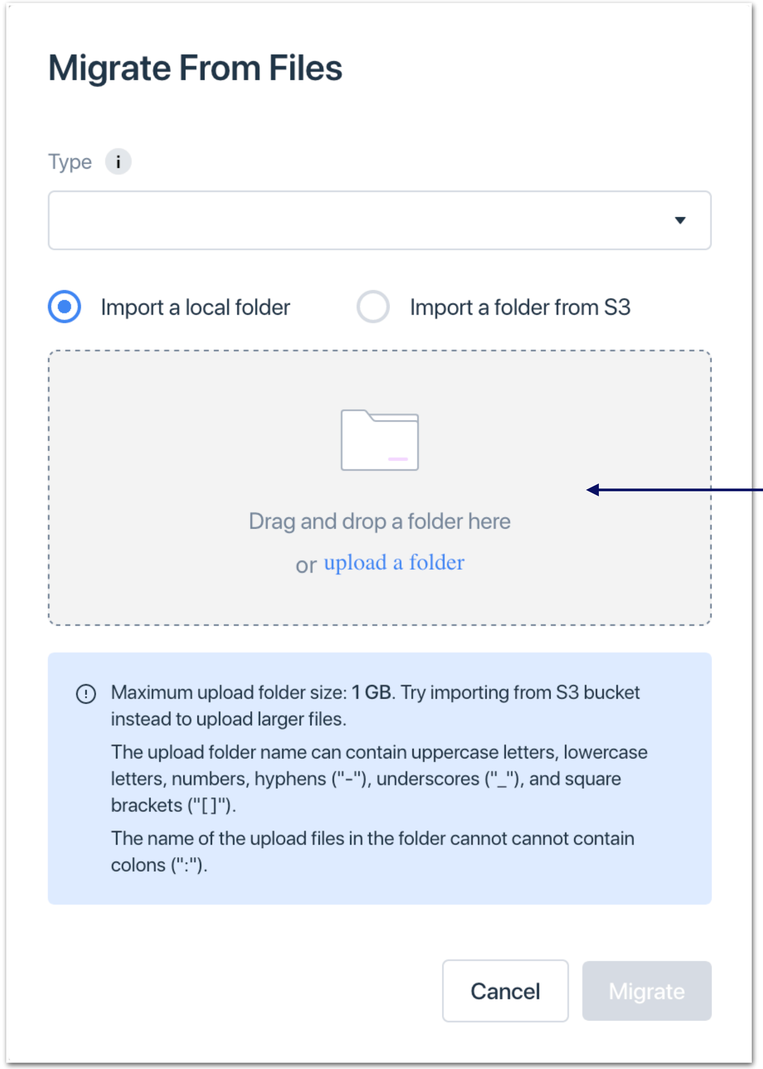
Step 2: Backup

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



Drag and drop backup folder here

05

A Quick Demo

Try Zilliz Cloud free today at zilliz.com/cloud

@Zilliz_Universe

linkedin.com/in/zilliz

milvusio.slack.com

github.com/zilliztech