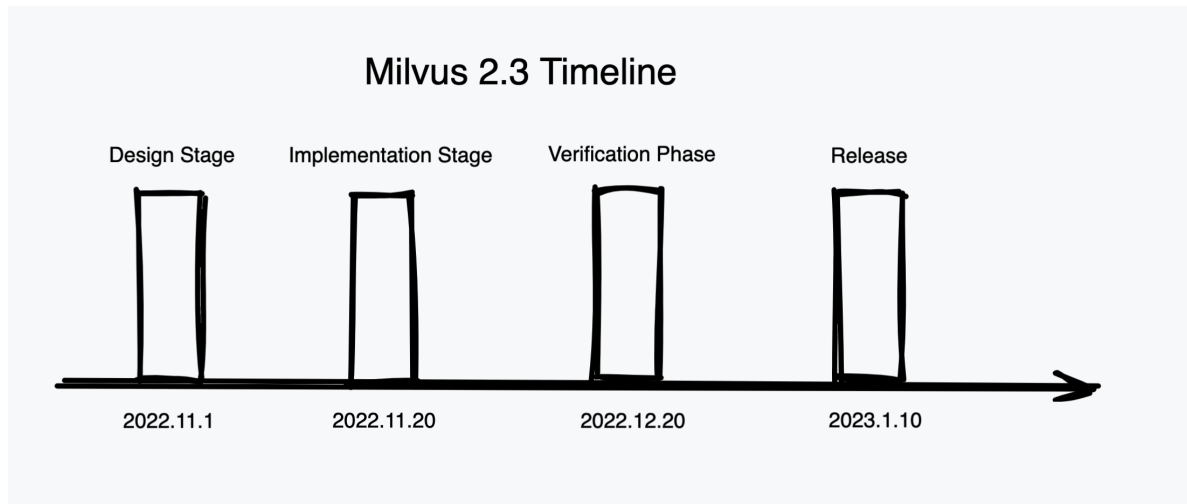


Feature plans for Milvus 2.3 and timeline



Stability improvement

Implement Mini Milvus Cluster and add more e2e test in the repo.

Clean all the chaos test issue and ensure all failure can be recovered in less than 3 Minutes.

Improve the availability with multiple replicas

Better memory usage control mechanism

Performance

2C8G 1M 128 dim Vector, serial search latency < 5ms

2C8G 1M 128 dim Vector, 50 thread concurrent search latency < 50ms(QPS > 1000)

Release vector search benchmark and tool

Support large cluster with tens of billion vectors and scale linearly

Improve the hybrid search/retrieve performance

Major functionalities

Partition refine

1. Support dynamically load/release one partition
2. Manage the memory when many partitions are created.
3. Improve partition search performance

Data backup

Resource Group

1. physical isolation between resource groups
2. resource group can be also combined with a namespace.

Range Search

Rolling Upgrade/Graceful shutdown

Change Data capture

GPU Index support (Optional)

SDK

Go SDK fully tested

Restful API Documentation

Refine python, java, go, node SDK and sync up all APIs

Enhancement

1. Refactoring Data storage format to support new scalar execution engine and better compression
2. Refactoring data coordinator and index coordinator, combine to single process and make it easy to understand and deploy
3. Implement QueryNode V2 to simplify search path and improve search performance

Ecosystem

1. Support Huggingface , PaddlePaddle and OpenAI as the inference engine
2. Support Data Migration from Milvus 1.0, Faiss, Hnswlib to Milvus 2.0
3. Support to deploy Milvus on GCP easily
4. Feder support - Vector data visualization
5. More trouble solving tools such as tracing, birdwatcher and dynamic config support

Documents

1. Update all user document
2. First version of milvus technical guide