

Milvus 설계 문서

1. 개요

Milvus는 대규모 벡터 데이터를 효율적으로 저장, 관리, 검색할 수 있는 벡터 데이터베이스입니다. 이 설계 문서는 세 개의 주요 컬렉션 `manual`, `genesis`, `insurance`에 대한 구조와 설정을 설명합니다.

2. 컬렉션 설계

2.1 `manual` 컬렉션

이 컬렉션은 차량 사용자 매뉴얼 PDF 텍스트 데이터를 벡터화하여 저장합니다.

주요 정보:

- 일관성: Strong Consistency
- 스키마: Dynamic Schema 지원

필드명	데이터 타입	설명	인덱스 유형	메트릭 타입
<code>pk</code>	VarChar(100)	기본 키	없음	없음
<code>car_id</code>	Int32	차량 모델 ID	없음	없음
<code>text</code>	VarChar(5000)	관련 텍스트	없음	없음
<code>sparse_vector</code>	SparseFloatVector	희소 벡터	SPARSE_INVERTED_INDEX	IP
<code>dense_vector</code>	FloatVector(1024)	고밀도 벡터	AUTOINDEX	IP
<code>dynamic_field</code>	Dynamic	동적으로 추가된 필드	없음	없음

2.2 `genesis` 컬렉션

이 컬렉션은 차량 추천에 사용할 사용자 후기 관련 데이터를 벡터화하여 저장합니다.

주요 정보:

- 엔티티 수: 5
- 일관성: Bounded Consistency
- 스키마: Dynamic Schema 지원

필드명	데이터 타입	설명	인덱스 유형	메트릭 타입
<code>id</code>	Int64	기본 키	없음	없음
<code>car_id</code>	Int32	차량 모델 ID	STL_SORT	없음
<code>vector</code>	FloatVector(768)	차량 관련 임베딩 벡터	AUTOINDEX	L2
<code>dynamic_field</code>	Dynamic	동적으로 추가된 필드	없음	없음

2.3 `insurance` 컬렉션

이 컬렉션은 보험 약관 PDF 텍스트 데이터를 벡터화하여 저장합니다.

주요 정보:

- **일관성:** Strong Consistency
- **스키마:** Dynamic Schema 지원

필드명	데이터 타입	설명	인덱스 유형	메트릭 타입
<code>pk</code>	VarChar(100)	기본 키	없음	없음
<code>text</code>	VarChar(5000)	관련 텍스트	없음	없음
<code>sparse_vector</code>	SparseFloatVector	희소 벡터	SPARSE_INVERTED_INDEX	IP
<code>dense_vector</code>	FloatVector(1024)	고밀도 벡터	AUTOINDEX	IP
<code>dynamic_field</code>	Dynamic	동적으로 추가된 필드	없음	없음

3. 데이터 흐름

1. 데이터 입력

- `manual` 및 `insurance` 컬렉션에 텍스트와 관련된 벡터를 삽입.
- `genesis` 컬렉션에 차량 데이터 임베딩 벡터를 삽입.

2. 벡터 검색

- `manual` 및 `insurance` 컬렉션에서는 `dense_vector` 및 `sparse_vector` 를 기반으로 IP 메트릭을 사용해 유사도 검색.
- `genesis` 컬렉션에서는 `vector` 를 기반으로 L2 메트릭을 사용해 유사도 검색.

3. 결과 반환

- 검색된 벡터와 매칭되는 데이터를 반환.