## Milvus使用手册

~~NC5主机，代码位置：~/pwork/milvus~~

DS主机，代码位置：~/pwork/milvus

《向量数据库：大模型驱动的智能检索与应用(2025)》

1. 嵌入模型

下载bge嵌入模型，模型大小：1.5G

apt install git-lfs

git lfs install

git clone https://huggingface.co/BAAI/bge-large-zh-v1.5

使用bge嵌入模型，代码见：~/pwork/gpt/bge.py

from langchain\_community.embeddings import HuggingFaceBgeEmbeddings

bge\_embeddings = HuggingFaceBgeEmbeddings(model\_name="/root/model/bge-large-zh-v1.5", model\_kwargs={'device': 'cuda:0'})

计算嵌入向量

v=bge\_embeddings.embed\_query("青蛙是食草动物")

1. 轻量Milvus
2. 容器Milvus
3. 云端Milvus