



# PostgreSQL 技术分享： pgvector 的应用和原理

唐颢

泽拓科技数据库技术专家

[www.klustron.com](http://www.klustron.com)

2023/08/31

# CONTENTS

## 目录

01 嵌入概念

02 使用范例

03 实现原理

04 Q&A

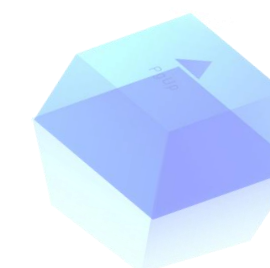


# 01

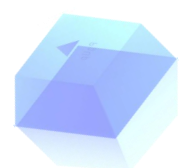
## PART ONE



**Klustron**

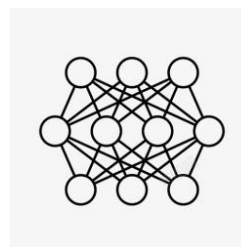


# 嵌入概念





文档、图片、音频等

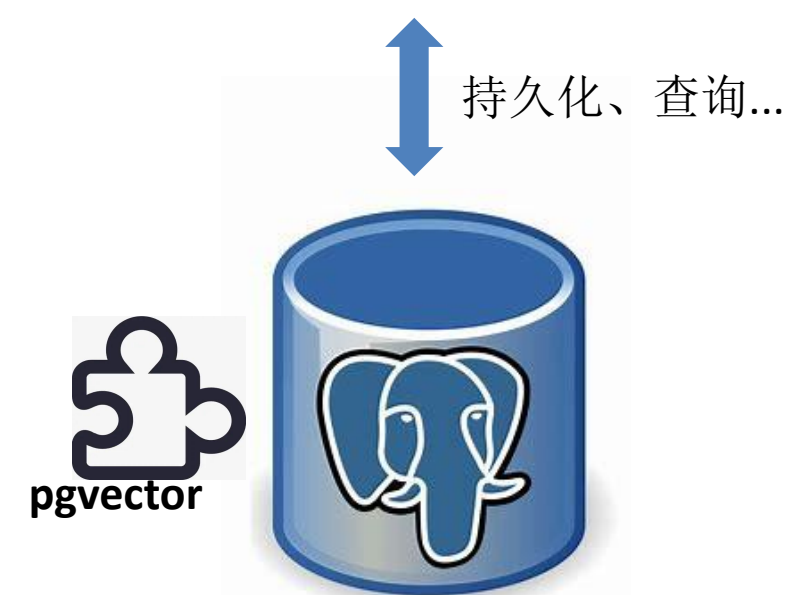


嵌入模型



【0.121, 0.31, ...】

向量

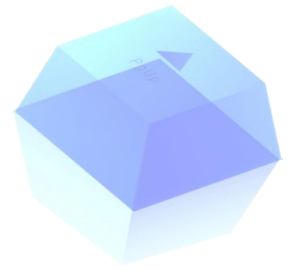


# 02

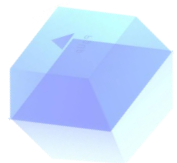
## PART TWO



**Klustron**



# 使用范例



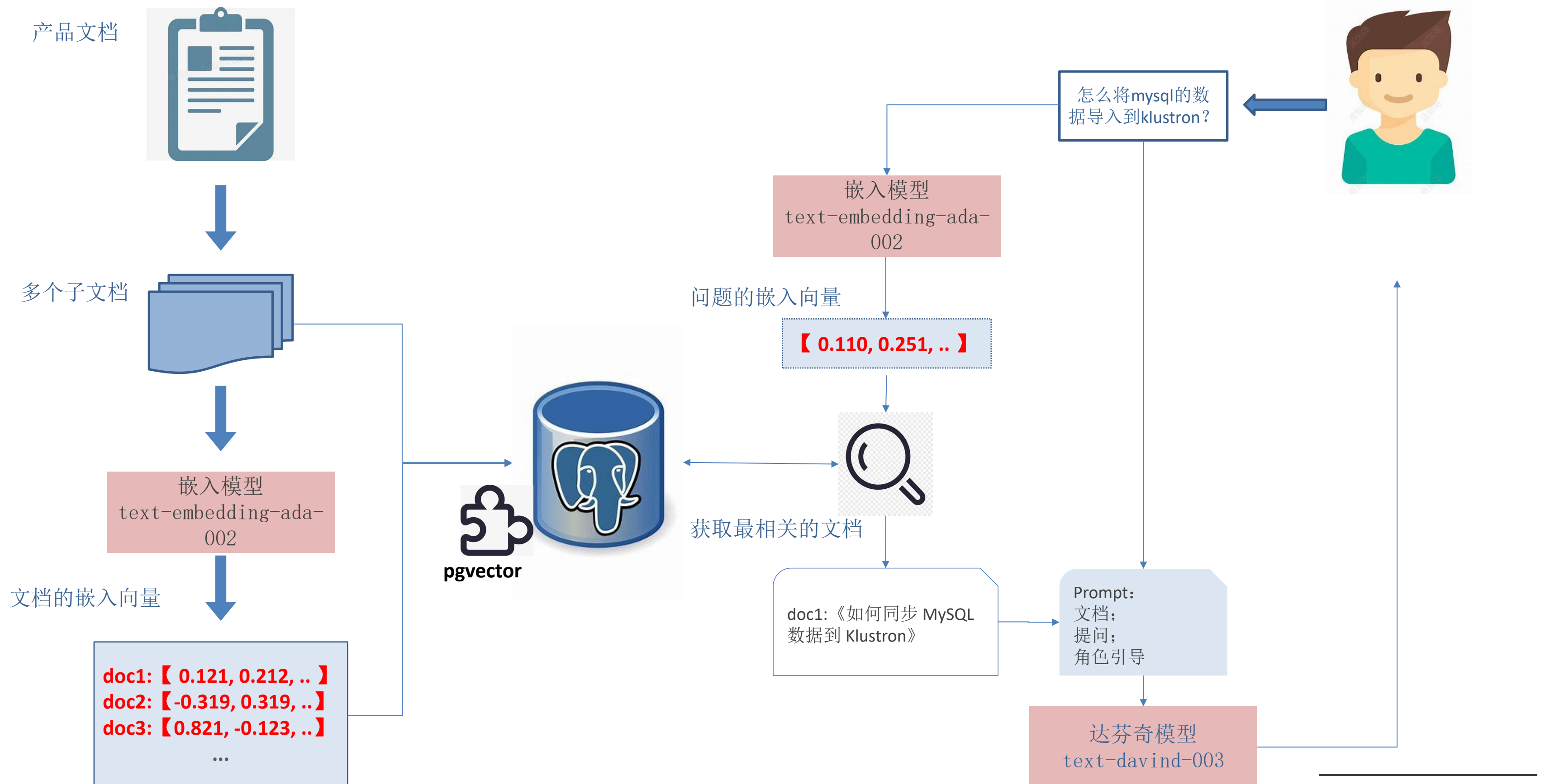


# 场景1. 和ChatGPT的联用

PART 02



Klustron



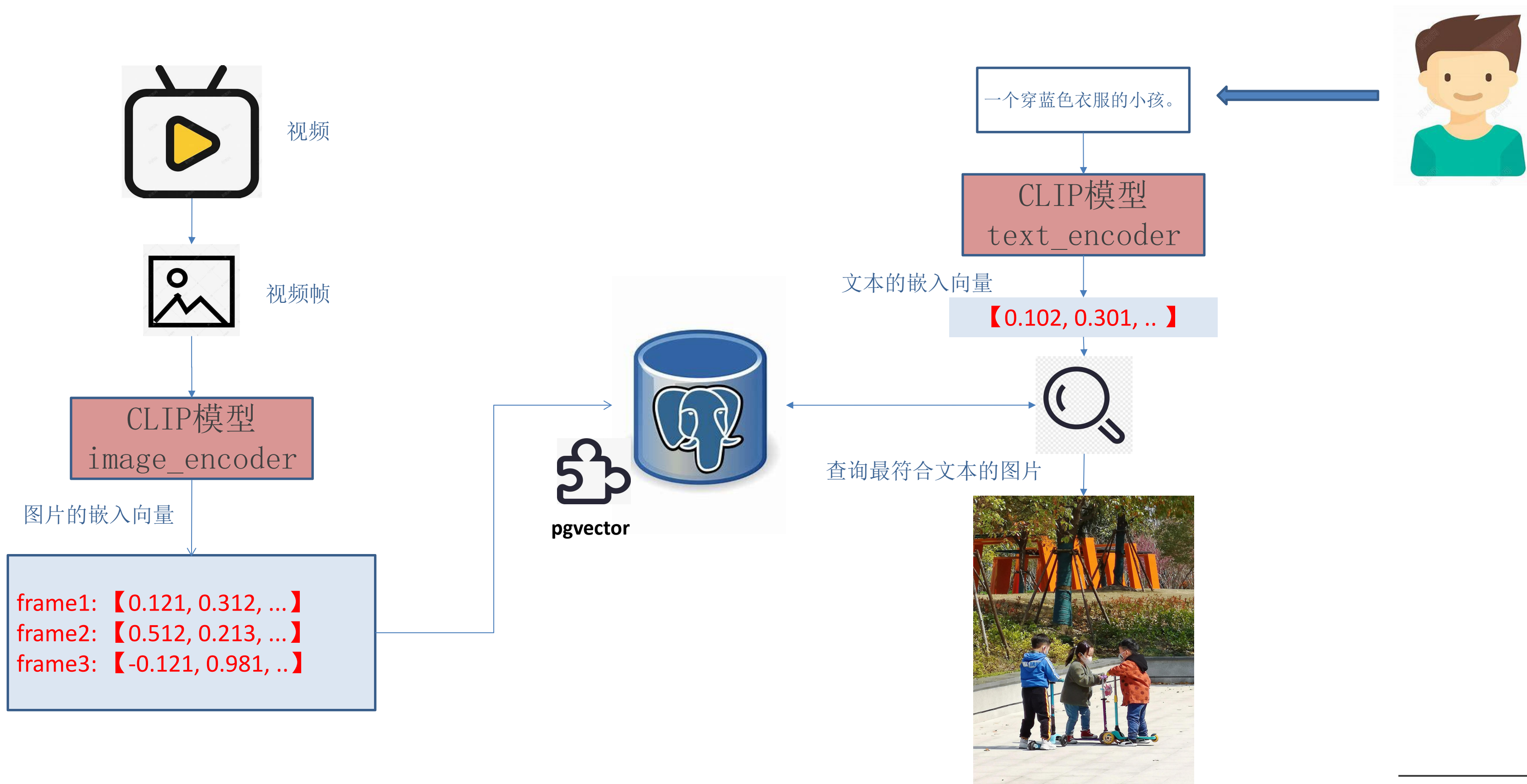


# 场景2. 图片搜索

PART 02



Klustron

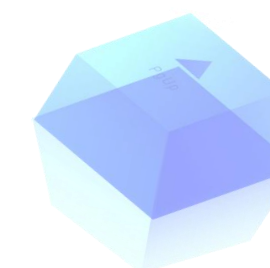


# 03

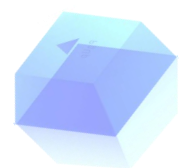
## PART THREE



**Klustron**



# 实现原理







```
typedef struct Vector
{
    int32    vl_len_;      /* varlena header (do not touch directly!) */
    int16    dim;          /* number of dimensions */
    int16    unused;
    float    x[FLEXIBLE_ARRAY_MEMBER];
}          Vector;
```

-- 定义类型vector

```
CREATE TYPE vector (
    INPUT      = vector_in,
    OUTPUT     = vector_out,
    TYPMOD_IN  = vector_typmod_in,
    RECEIVE    = vector_recv,
    SEND       = vector_send,
    STORAGE    = extended
);
```

-- 创建表

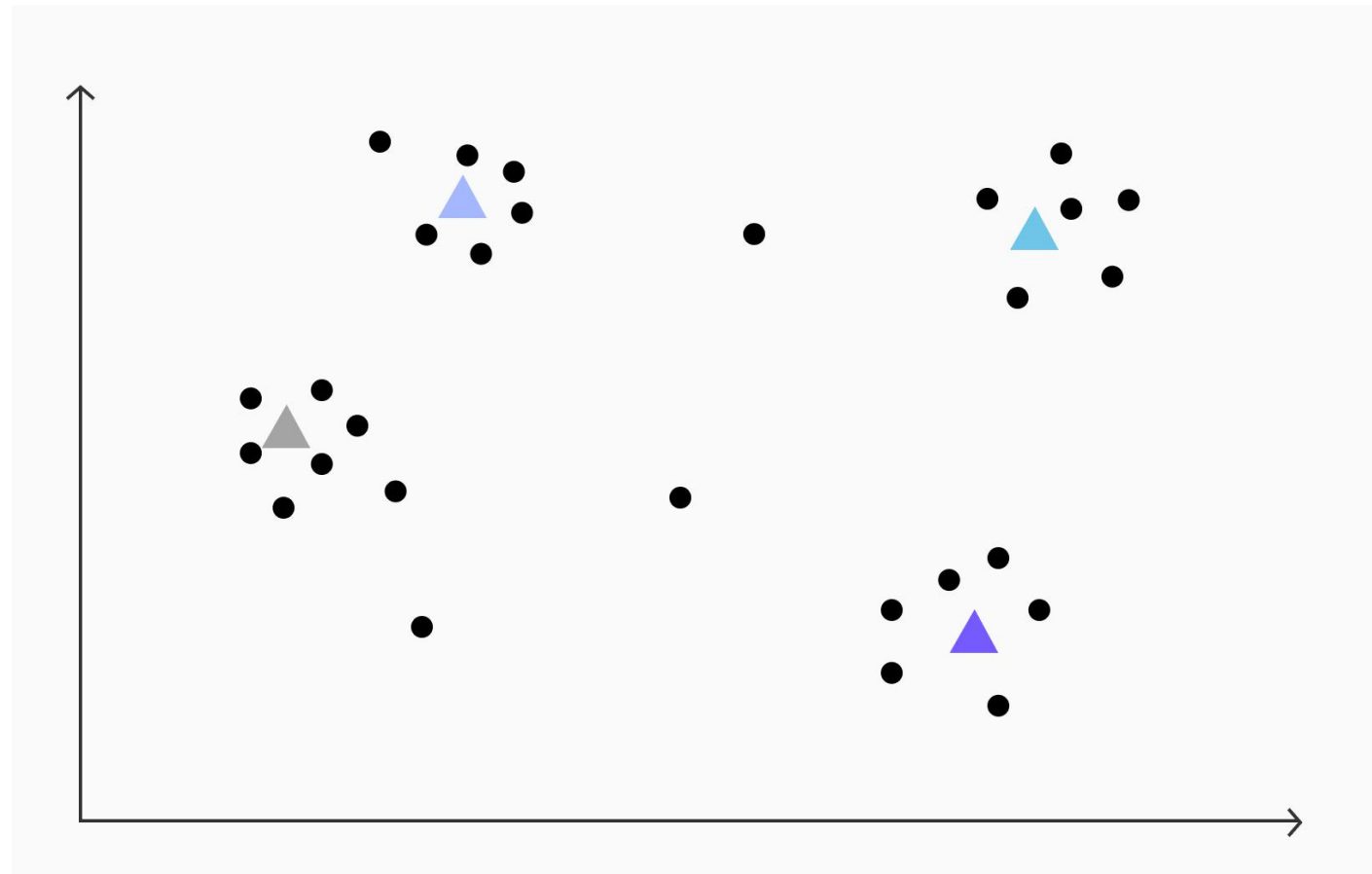
```
CREATE TABLE t(a vector(4));
```

-- 插入数据

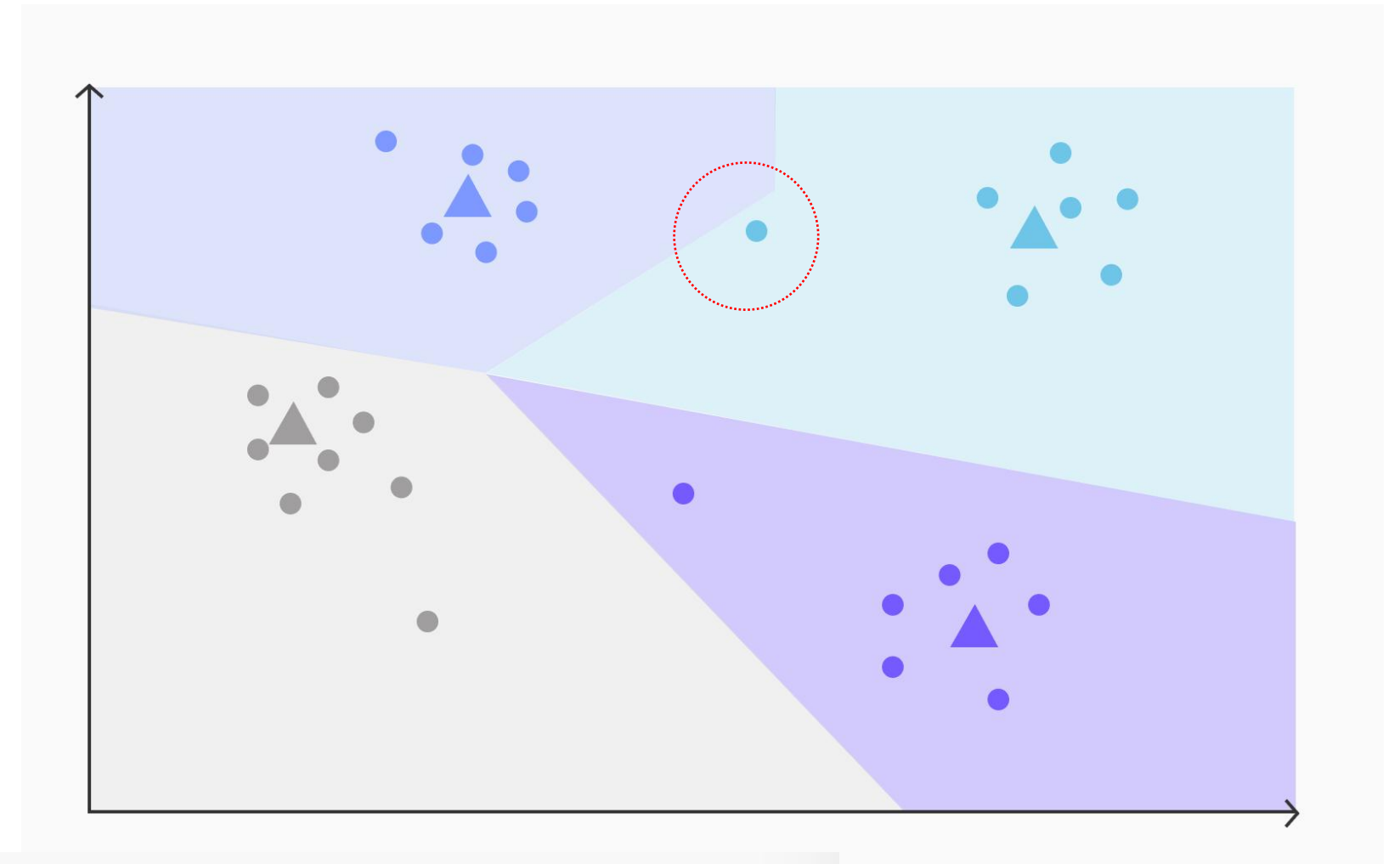
```
INSERT INTO t VALUES('[1, 2, 3, 4]')
```

距离类型	操作符	索引方法
L2 / Euclidean	<->	vector_l2_ops
Negative Inner product	<#>	vector_ip_ops
Cosine	<=>	vector_cosine_ops

```
SELECT image FROM vedio_features
      ORDER BY feature <=> %s::vector limit 5;
```



KNN聚类

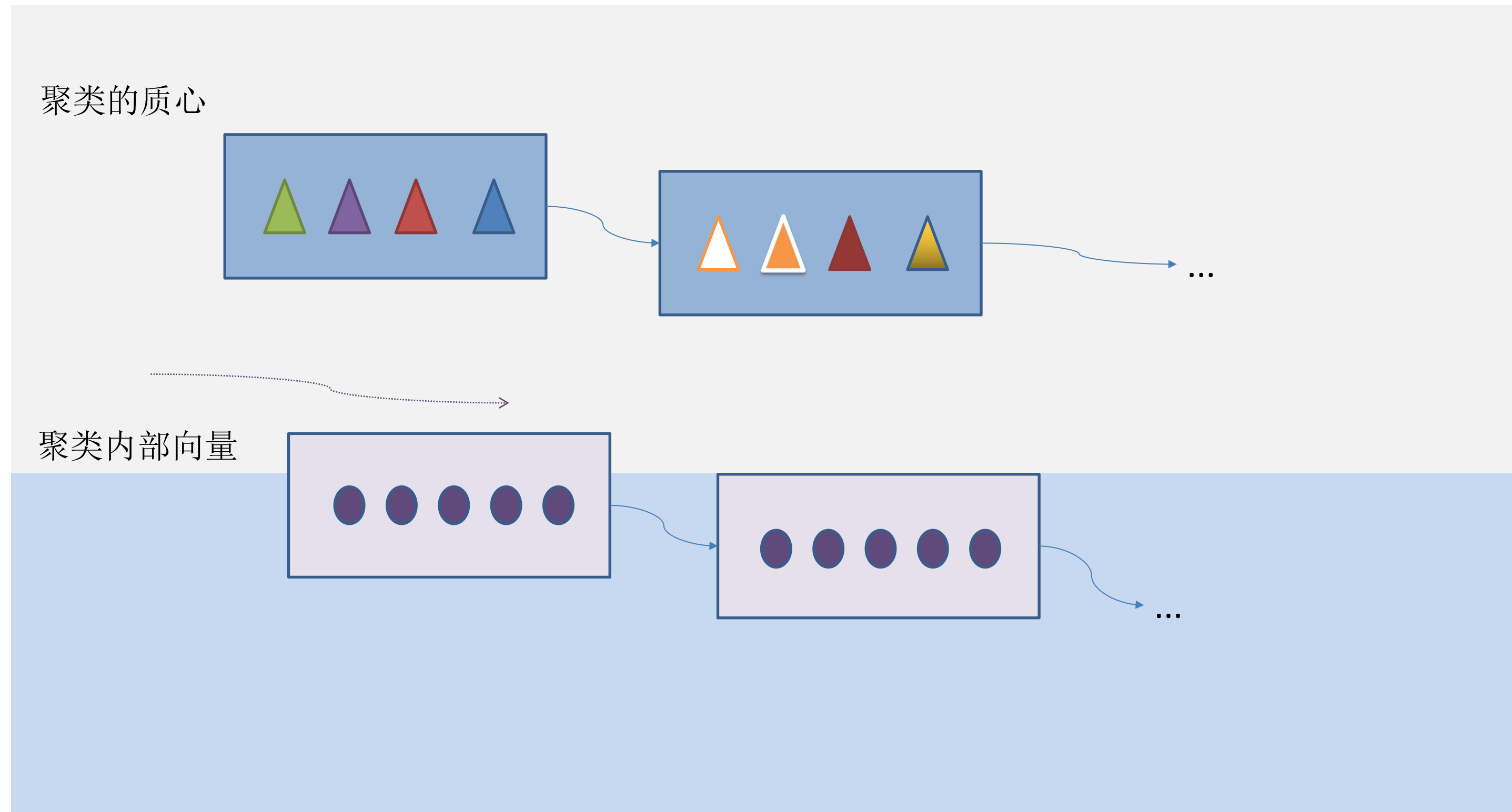


-- 索引参数lists

```
CREATE INDEX ON <table name> USING ivfflat  
  (<column name> <index method>) WITH (lists = <lists parameter>);
```

-- 运行时参数

```
set ivfflat.probes = 3 ;
```



- 空表建立索引，质心石随机生成的；
- 质心一旦确定，后续将保持不变；
- 只能用于优化 排序，并且返回元组数局限于扫描区域的中元组的数据。



# Q & A

T H A N K   Y O U

演示代码地址:

<https://github.com/zettadb/techtalk/blob/main/pgvector%E7%9A%84%E5%BA%94%E7%94%A8%E5%92%8C%E5%8E%9F%E7%90%86/clip.ipynb>

[www.klustron.com](http://www.klustron.com)

2023/08/31



Klustron公众号



Klustron小助手



交流群