OCEANBASE

LangChain Community

OceanBase SeekDB

为开发者打造的 AI 原生数据库

张海立

LangChain Ambassador

OCEANBASE

目录
Contents

OCEANBASE

01 AI 应用开发者需要怎样的数据库

OCEANBASE

时代的转折点：我们已进入 AI 原生时代

从数据驱动到 AI 原生，应用的核心正在被重构

Web2.0/ 业务在线化时代

一个可靠、精确的记录系统

移动互联网/数据智能化时代

一个强大的分析系统

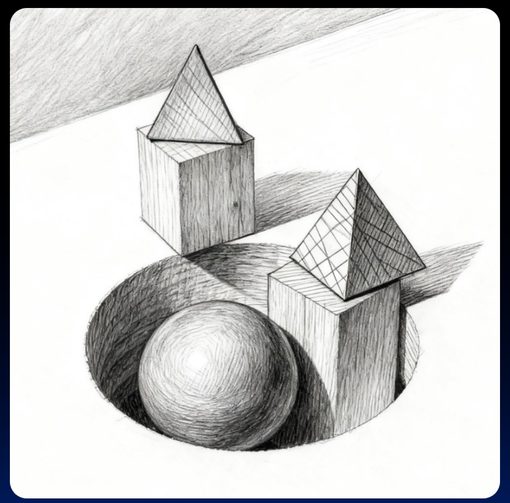
AI 原生时代

不仅用来查询和分析，更用来理解和推理

OCEANBASE

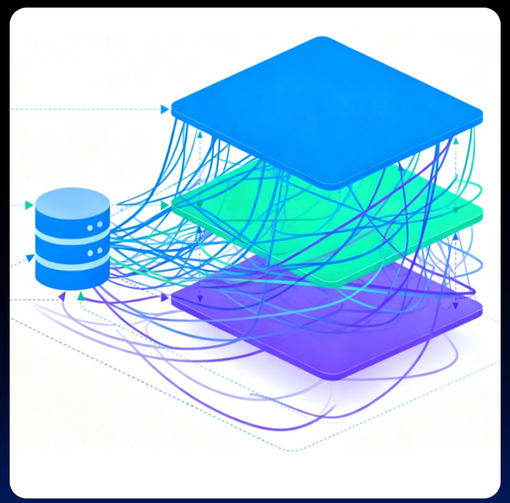
新挑战，新需求：AI 时代开发者的四大痛点

数据类型的“多维化”

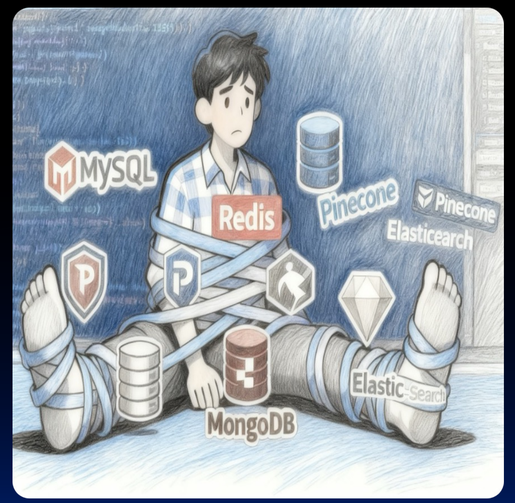


性能与规模的“极致化”

智能处理的“内生化”



开发流程的“敏捷化”



OCEANBASE

我们期待的 AI 时代数据库

多模态

统一的多模数据平台

高性能

为 AI 工作负载而
重塑的引擎

智能

内嵌 AI 运行时
的大脑

简单

以开发者为导向

强大、智能、一体化的数据与 AI 融合平台

02 | OceanBase SeekDB

OCEANBASE

一体化架构全新突破：OceanBase SeekDB

一款为 AI 时代开发者和智能应用设计的数据库

更轻量

- 推荐配置 1C2G
- Pip install 一键安装，秒级启动
- 本机部署 / 嵌入式部署

更开放

- Apache 2.0 协议
- LangChain/Dify 等 30+ 生态集成

AI 原生

- 混合搜索 /AI Function
- 全新 Doc in, data out API

平滑升级

- SeekDB API 是 OceanBase 子集
- 应用平滑迁移到分布式版本

访问地址：http://oceanbase.ai

OCEANBASE

OceanBase SeekDB: AI 原生，为开发者打造

```python
import pyseekdb

client = pyseekdb.Client()
collection = client.get_or_create_collection("products_database")

# Add product documents
collection.upsert(
    documents=[
        "Laptop Pro with 16GB RAM, 512GB SSD, and Intel i7 processor",
        "Gaming Laptop with 32GB RAM, 1TB SSD, and NVIDIA RTX 4080",
        "Business Ultrabook with 8GB RAM, 256GB SSD, and long battery life",
        "Tablet with 6GB RAM, 128GB storage, and 10-inch display"
    ],
    metadatas=[
        {"category": "laptop", "ram": 16, "storage": 512, "price": 1200, "type": "professional"},
        {"category": "laptop", "ram": 32, "storage": 1000, "price": 2500, "type": "gaming"},
        {"category": "laptop", "ram": 8, "storage": 256, "price": 900, "type": "business"},
        {"category": "tablet", "ram": 6, "storage": 128, "price": 600, "type": "consumer"}
    ],
    ids=["1", "2", "3", "4"]
)

print("📚 Product database built\n")

# Hybrid search for high-performance laptops
print("🔍 Hybrid Search: High-performance laptops for professional work")
results = collection.query(
    query_texts=["powerful computer for professional work"],  # Vector search
    where={                                                   # Scalar filters
        "category": "laptop",
        "ram": {"$gte": 16}
    },
    where_document={"$contains": "RAM"},                      # Full-text search
    n_results=2
)

print("Search Criteria:")
print("  • Vector: 'powerful computer for professional work'")
print("  • Filters: laptops with 16GB+ RAM")
print("  • Full-text: contains 'RAM'")
print("\nResults:")
for i, (doc, metadata) in enumerate(zip(results['documents'][0], results['metadatas'][0])):
    print(f"  {i+1}. {doc}")
```

```
📚Product database built

🔍Hybrid Search: High-performance laptops for professional work
Search Criteria:
  • Vector: 'powerful computer for professional work'
  • Filters: laptops with 16GB+ RAM
  • Full-text: contains 'RAM'

Results:
  1. Laptop Pro with 16GB RAM, 512GB SSD, and Intel i7 processor
  2. Gaming Laptop with 32GB RAM, 1TB SSD, and NVIDIA RTX 4080
```

3 行代码构建 AI 应用

内置 Embedding

混合搜索 = 向量检索 + 标量过滤 + 全文搜索

Schemaless API + SQL 接口

OCEANBASE

OceanBase SeekDB：融合数据与智能

Unified API	SQL-based Unified Query Language with Stored Procedure				Schemaless API	
Multi-model Computing Layer	Vectorized Engine	Parallel Execution	Adaptive TP/AP Execution	**Advanced Optimizer**	**Hybrid Search**	**AI Functions**
Multi-model Data Types	Table	**Vector**	Wide Column	**JSON**	**Text**	**GIS/Array/ Bitmap/...**	UDT
Multi-model Index Layer	Secondary Index	LOB	**Vector Index**	Multi-value Index	**Full text Index**	GIS Index
Storage Layer with ACID	Row Store	Column Store	LSM-Tree Storage Engine		Vector Store	Multi-model Store
Flexible Architecture	Embedded Mode		Single Node		Primary/Standby Cluster	

统一存储

融合计算

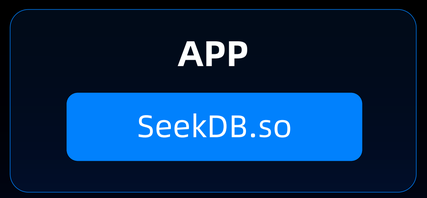
原生智能

灵活接口

OceanBase SeekDB：从原型开发到海量用户

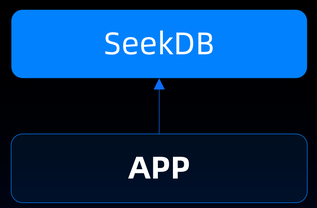
OceanBase SeekDB

SeekDB as an Embedded Library



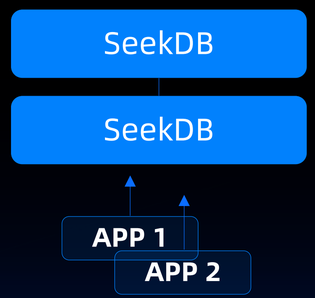
- Students
- Prototyping

Single Node SeekDB



- Individual Developers
- Test & Prod Environment

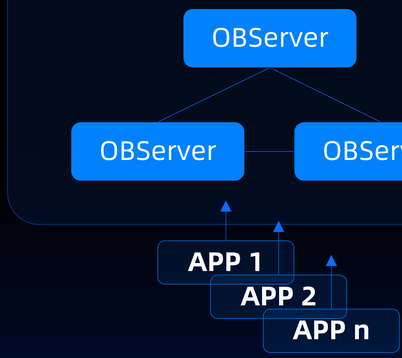
Single Node With Standby



- Small to medium scale workloads
- Disaster Recovery

OceanBase Distributed

Distributed Cluster



- Large scale workloads
- Mission-critical workloads

SeekDB takes developer experience seriously

03 LangChain x Oceanbase 工作坊

实验任务：基于 SeekDB 快速构建 Hybrid RAG

1. Knowledge Base

使用给定的 Nike 10-K 2023 财报文件，基于 OceanBsase SeekDB 构建一个基础的知识库

2. Agentic RAG

基于 langchain-oceanbase 构建一个 Agentic RAG 的智能体，实现对于 Nike 财报知识库的检索

3. Hybrid RAG

请参照前两个实验的代码和相关文档，基于 LangChain V1 和 langchain-oceanbase，实现可实现混合检索的 Agentic RAG

实验环境

uv + python

docs.astral.sh/uv 下载

seekdb

oceanbase.io 下载

juypter

随 uv sync 安装

安装依赖

项目目录下执行 uv sync

Docker 环境

可下载桌面版进行安装

Linux 环境

可使用 yum install seekdb 安装

IDE 插件

请在您的 IDE 中安装 Juypter Notebook 插件

实验资源包下载地址：？？？

01-kb.ipynb

演示: 知识库构建

02-agentic.ipynb

演示：Agentic RAG 构建

03-hybrid.ipynb

实验：实现 Hybrid RAG

实验要求和评审标准

20 分钟

实验时间
（不含演示时间）

填充 ipynb

填充 03-hybrid.ipynb 中
的空缺部分，使其可运行

运行成功

03-hybrid.ipynb 应可以
完全运行并显示成功

截图上传

把实验结果截图发送到工
作坊的微信群中

实验速通攻略

仔细阅读
官方文档

使用 AI 辅
助编程

提供样例
指导 AI

使用MCP
context7.co
m

1.X

4.X

OCEANBASE

谢 谢