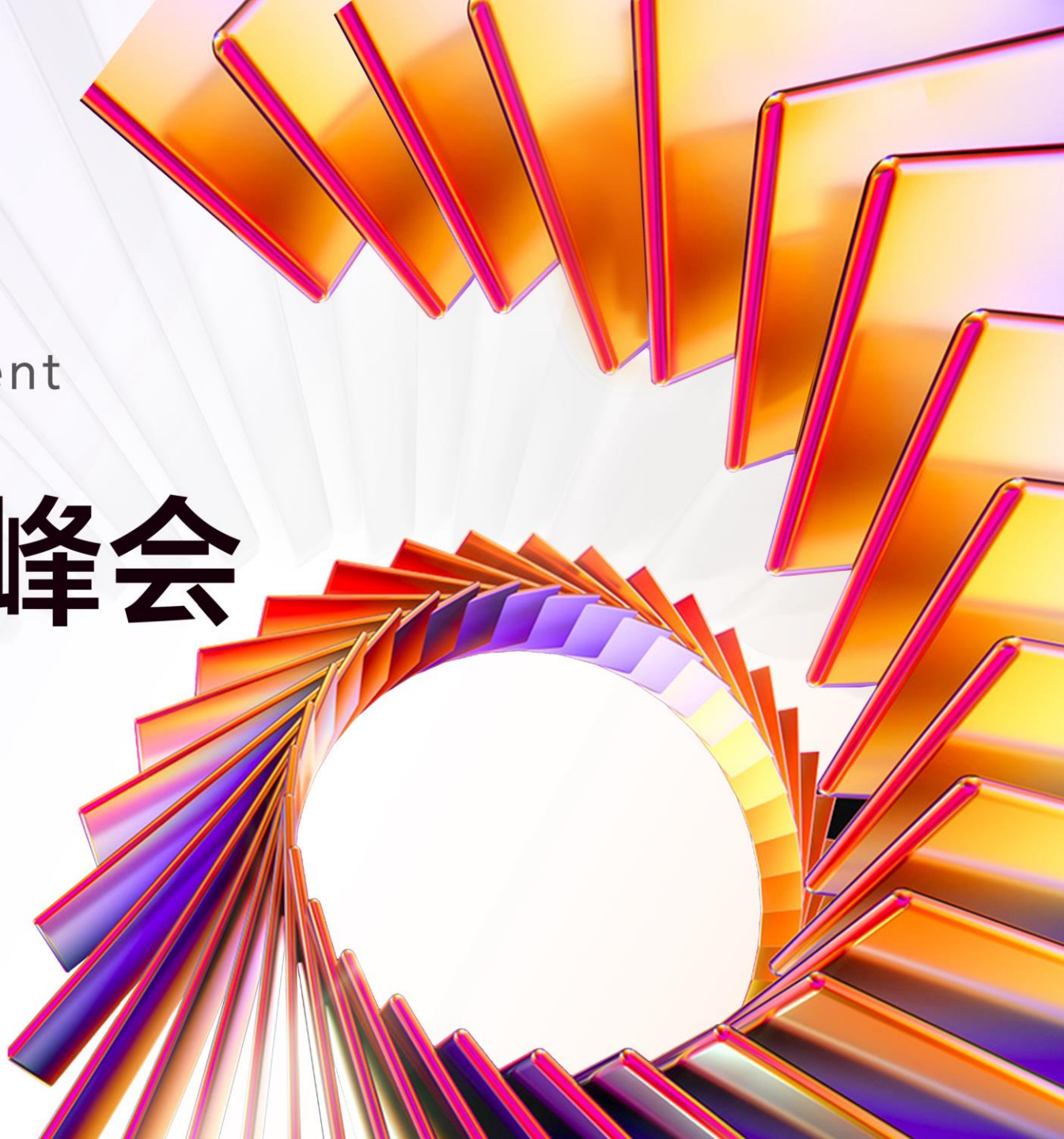




第7届 AI+ Development Digital Summit AI+研发数字峰会

拥抱AI 重塑研发

8月8-9日 | 北京站





第8届AI+研发数字峰会

拥抱AI 重塑研发 AI+ Development Digital Summit

下一站预告

11/14-15 | 深圳站

12/19-20 | 上海站



查看会议详情

深圳站论坛设置

智能装备与机器人

超越“编程 Copilot”

下一代知识工程

智能网联与汽车智能化

AI 测试工具开发与应用

AI 基础设施和运维

数据智能及其行业应用

可信 AI 安全工程

大模型和 AI 应用评测

多 Agent 协同框架

从智能测试到自主测试

大模型推理优化

多模态 LLM 训练与应用

智能化 DevOps 流水线

上下文工程

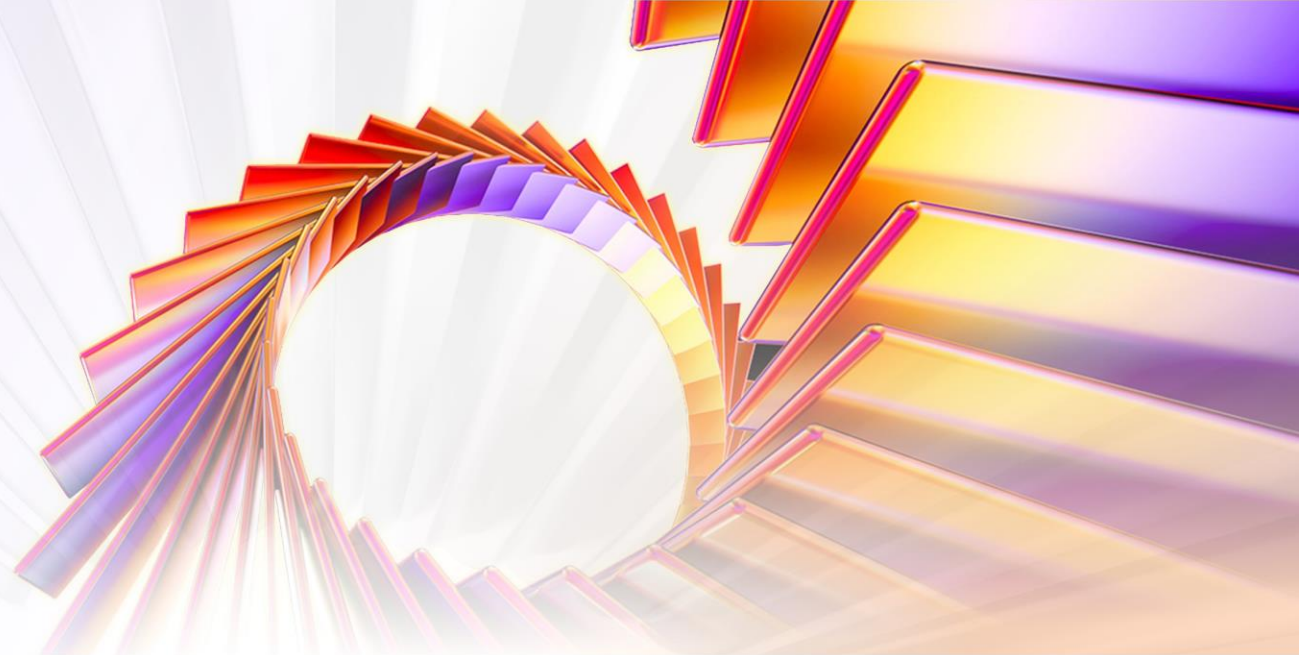


| 8月8-9日 | 北京站

第7届 AI+ Development
Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发



InfiniSynapse 基于RAG与 InfiniSQL的Agentic智能数据洞察

赵伟楠 | InfiniSynapse team



赵伟楠

InfiniSynapse Team

AI与大数据践行者，专注AI驱动大数据与分析领域，拥有近十年行业实战经验。主导多个国际零售与运动品牌项目架构实施，深度参与全球500强企业数字化转型战略落地。Apache开源社区贡献者，InfiniSynapse项目核心开发者，推动技术创新与知识共享实践。

目录

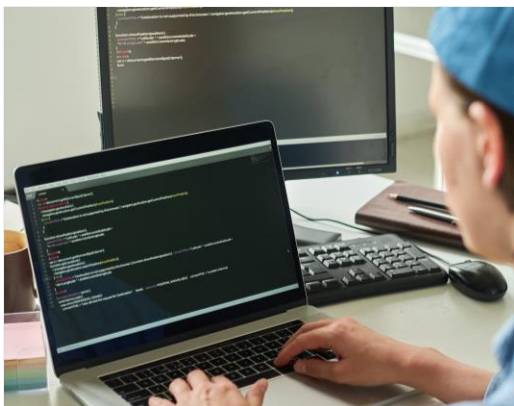
CONTENTS

- 01 AI分析的价值
- 02 传统Text2SQL的痛点
- 03 新一代AI数据分析底层架构设计
- 04 基于InfinSynapse的解决方案
- 05 未来的展望

PART 01

AI分析的价值

▶ 传统数据分析所面临的挑战



技术门槛高

对SQL或编程技能的依赖性强，导致业务人员必须等待技术团队的协助。



处理流程缓慢

从编码、调试到撰写文档，这一系列过程消耗大量的时间。



洞察力有限

高度依赖业务提供业务知识，同时人看一个问题也会有局限性。



多模态数据利用有限

不能很好的分析音频，视频，语音数据等。



AI如何赋能数据分析



打破依赖链

- 传统：业务需求→排队等分析师→获得滞后报告→二次沟通修正
- AI时代：自然语言提问→实时生成洞察→自主迭代分析`



核心能力

- 自助查询
 - ☆示例：“上季度华东区高客单用户复购率下降原因？”
 - ☆AI响应：自动关联促销活动/竞品动因/物流异常
- 深度洞察延伸
 - ☆自动检测：区域性渠道窜货 → 预测库存风险 → 推荐调货方案
- 助力决策
 - ☆从“发生了什么”升级到“该怎么做”

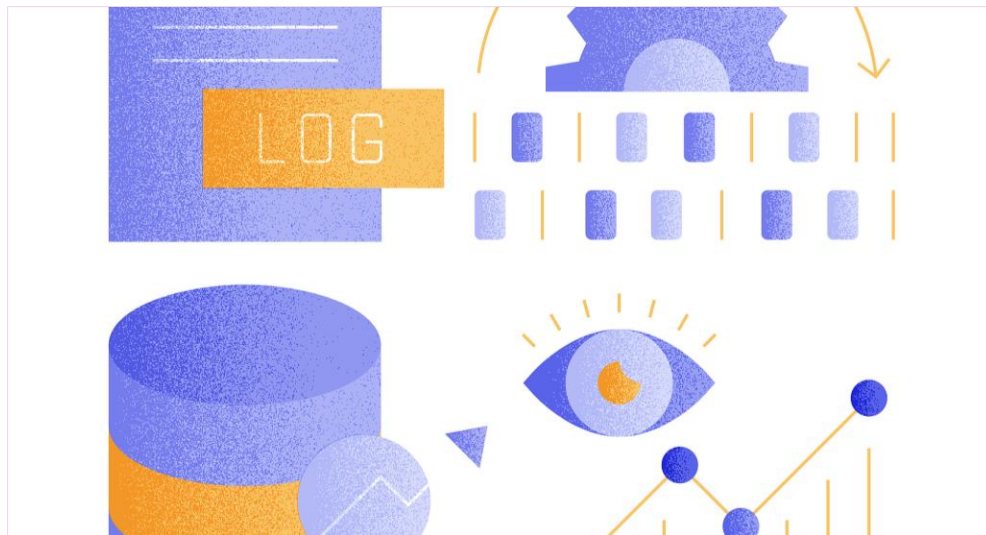


价值量化

- 需求响应时间：**7天→30分钟**



▶ AI赋能数据开发者——效率提升

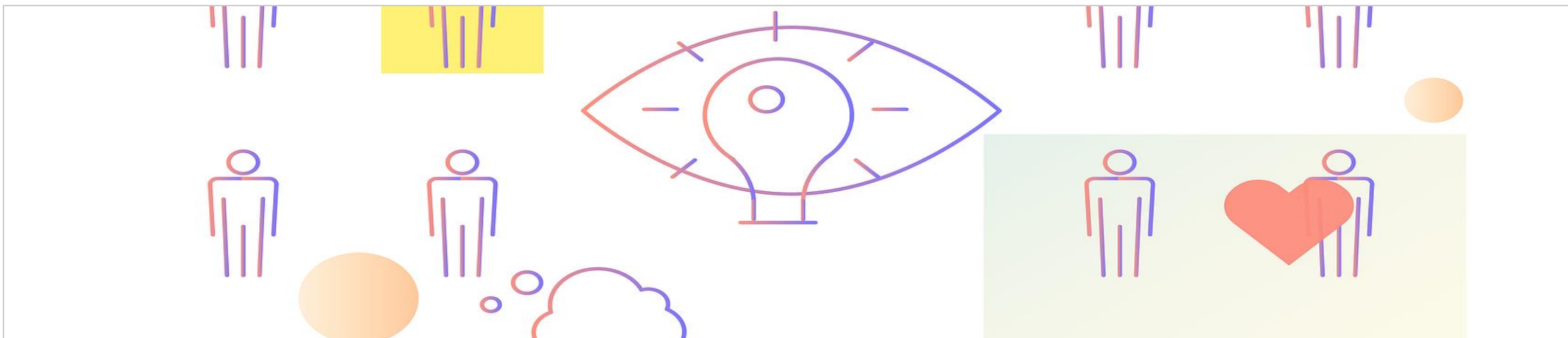


全流程提效

环节	传统方式	AI赋能	效率增益
数据查询	手动编写SQL	NL→SQL自动转换	节省50%+
文档管理	事后补文档	自动注释/同时生成数据字典/历史数据文档修复	100%覆盖
可视化开发	结合前端/BI才能可视化	AI一键生成图表	零配置生成



▶ AI赋能算法团队——攻克高阶战场



突破结构化边界

数据类型

传统方案局限

AI解决方案

价值案例

文本

分析文本需要大量算法的投入

LLM情感分析+主题聚类

从一些评论或者文字消息判断人的特征爱好等

图像

需定制CV模型

LLM快速理解视频

基于视频的AI Training

语音

转译准确率生硬

各种方言自适应分析

具有情感色彩的翻译，或者会议记录



PART 02

传统Text2SQL的痛点

方案原理

“喂” 数据库Schema + 业务文档 + 自然语言

大模型

直接输出可执行SQL



方言适配难

- 模型难以完美适配不同数据库（MySQL, PostgreSQL等），常出现函数/语法错误。



数据计算依赖

- 需将所有数据导入单一数据库，限制灵活性与效率。



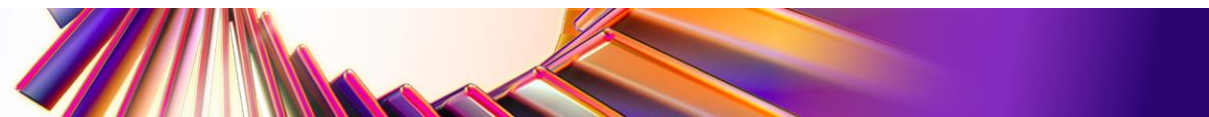
上下文控制难

- 严重依赖完备的文档，模型决策能力受限于信息质量。



复杂SQL生成困境

缺乏“阶段性思考”，生成复杂长SQL（含CTE等）成功率极低。



▶ 方案困境



黑箱操作，缺乏信任

- 生成SQL过程不透明，结果难以验证，用户信任度低。



自我修复能力缺失

- 需将所有数据导入单一数据库，限制灵活性与效率。



维护成本高、扩展性差

- 严重依赖完备的文档，模型决策能力受限于信息质量。



.....



PART 03

新一代AI数据分析底层架构

统一数据访问层

★目标：

数据库无关的通用SQL接口

★能力：

跨源计算 + 分布式执行





智能业务知识引擎

☆核心：强大的RAG

☆价值：企业级业务百科全书，精准理解业务背景

第7届 AI+ 研发数字峰会 | 拥抱 AI 重塑研发



自主思考AI智能体

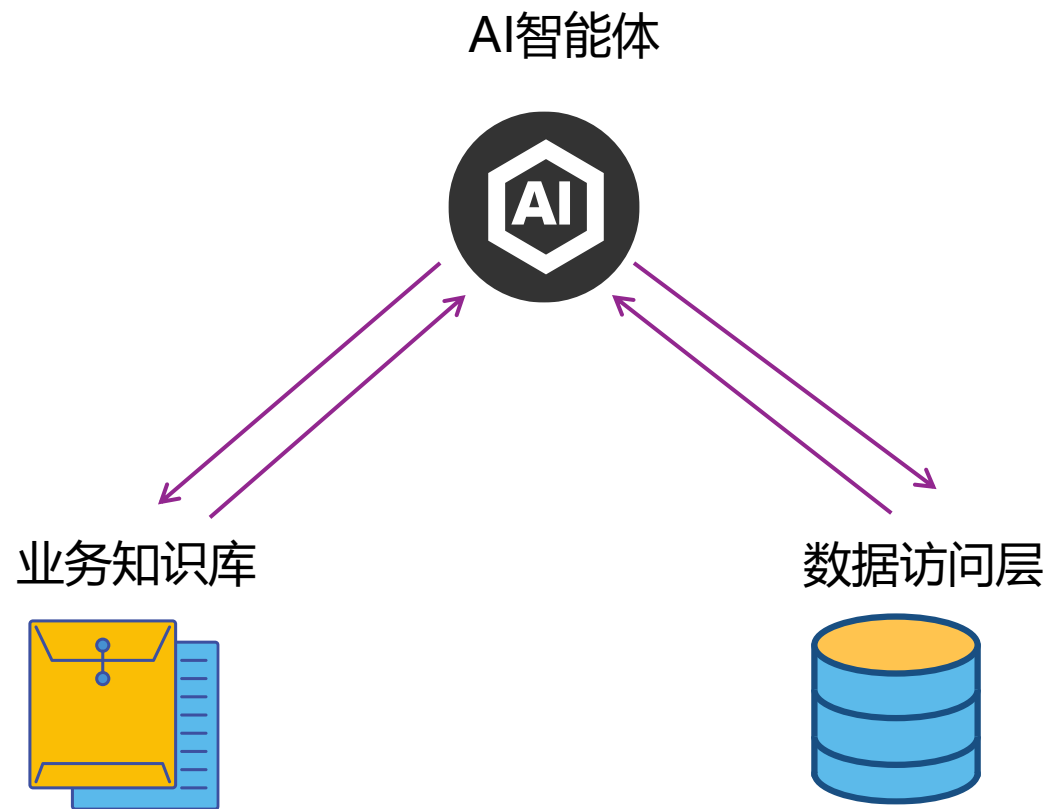


核心能力:

- 任务自主规划 & 解决方案思考
- 自我纠错机制
- 无缝协同SQL查询与RAG知识



▶▶ 数据分析架构图



PART 04

基于InfinSynapse的解决方案



核心定位:

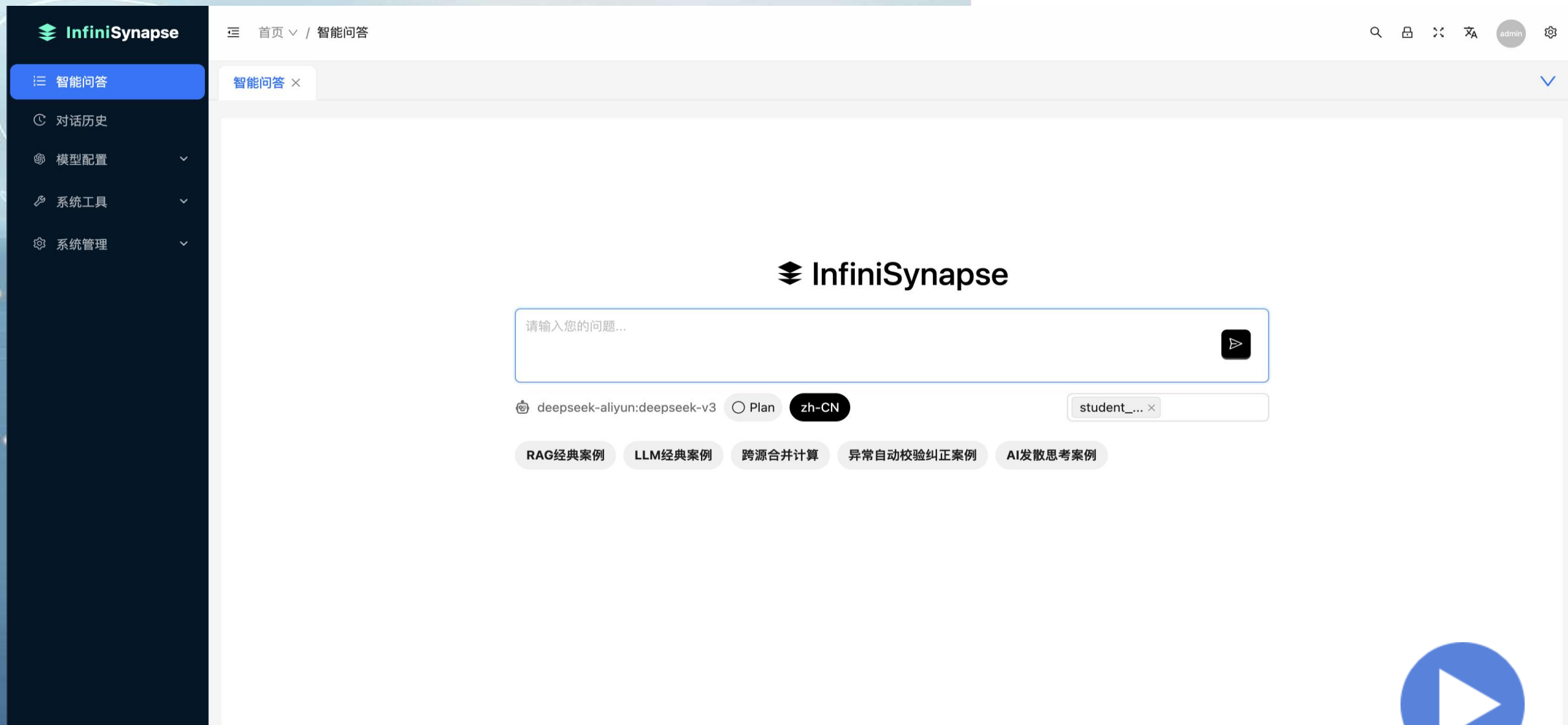
InfiniSynapse是一个具有数据分析师思路，基于Agentic范式落地的AI分析工具。

核心能力:

1. 纯自然语言驱动分析，零学习成本
2. 支持分布式跨数据源的查询
3. 支持结构化/文档/语音/视频等全模态数据做联合分析

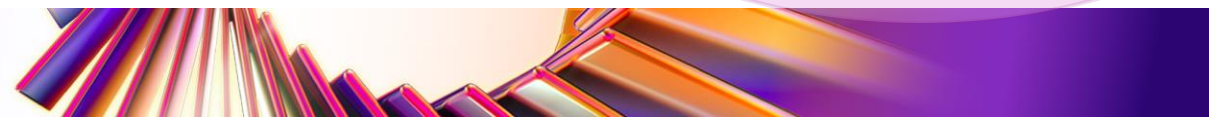
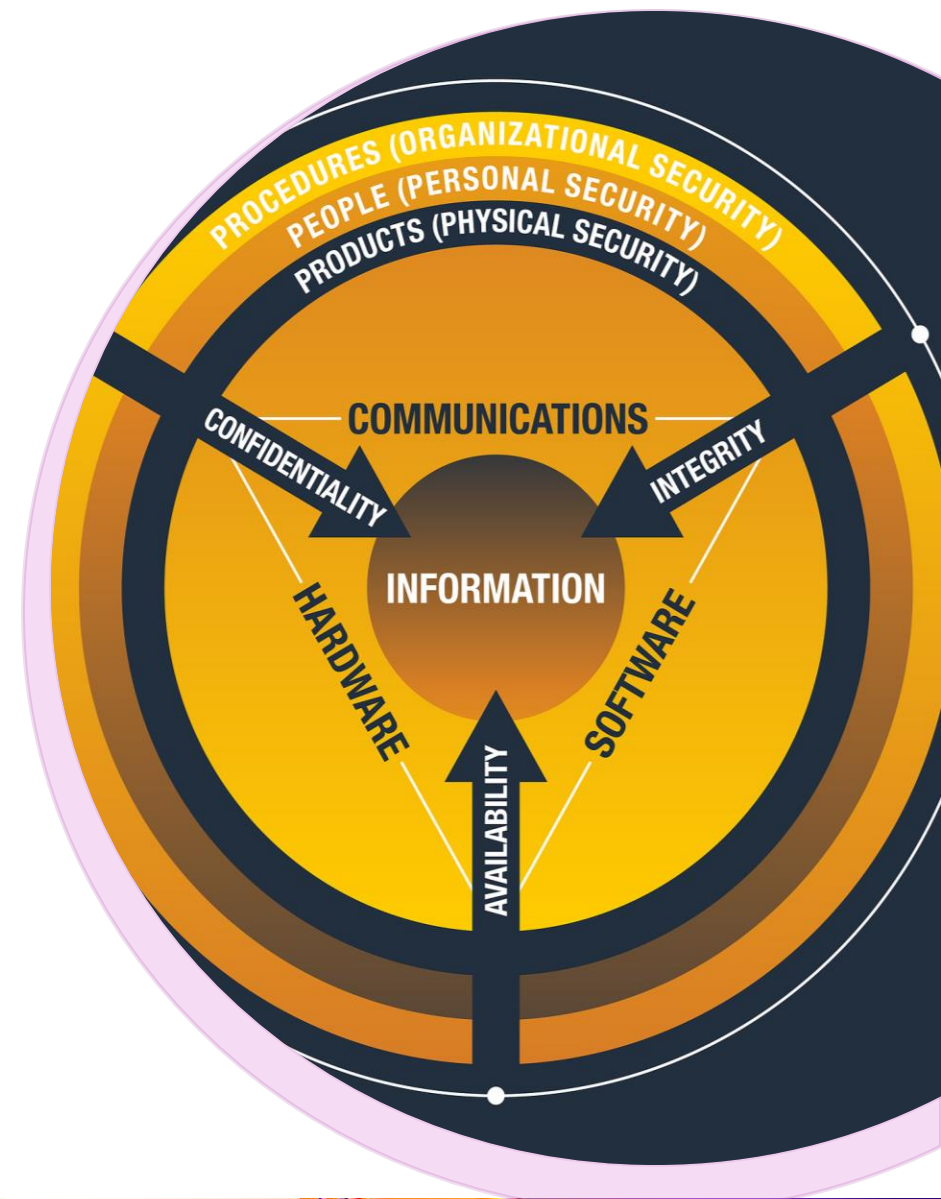


InfiniSynapse首页界面

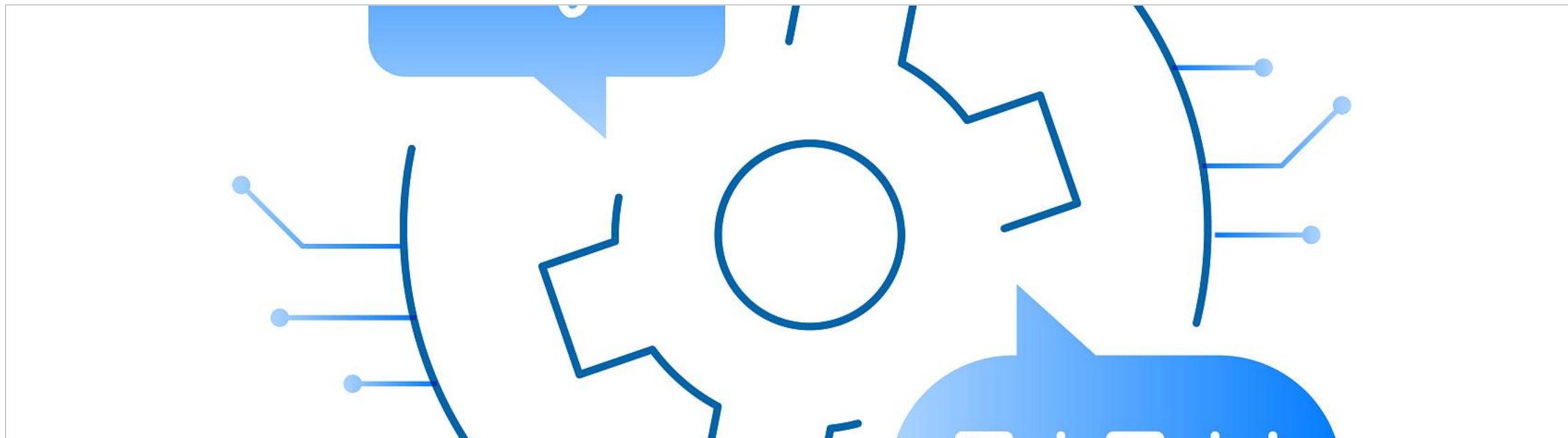


▶ InfiniSynapse组件介绍

1. InfiniSynapse(Agentic) - 智能中枢
2. InfiniSynapse(RAG) - 知识库大脑
3. InfiniSynapse(InfiniSQL) - 计算引擎



► 功能介绍——智能中枢

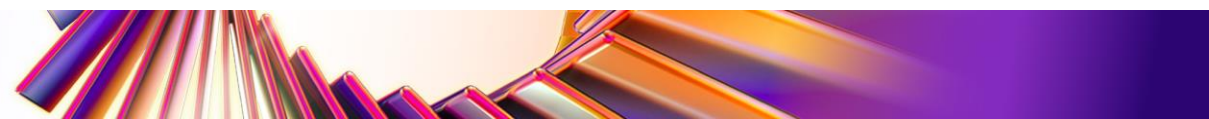


1. 负责与用户进行自然语言交互，深度理解用户意图。

2. 将复杂需求智能拆解为可执行的子任务。

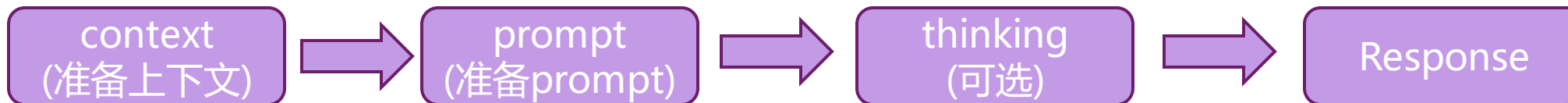
3. 制定最优执行计划，并调度协调 InfiniRag 和 InfiniSQL 协同工作。

4. 监控执行过程，自我诊断与修复问题，确保任务闭环完成。

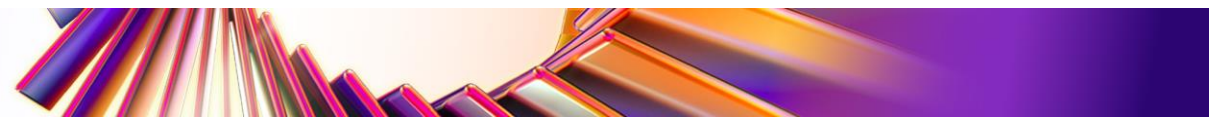
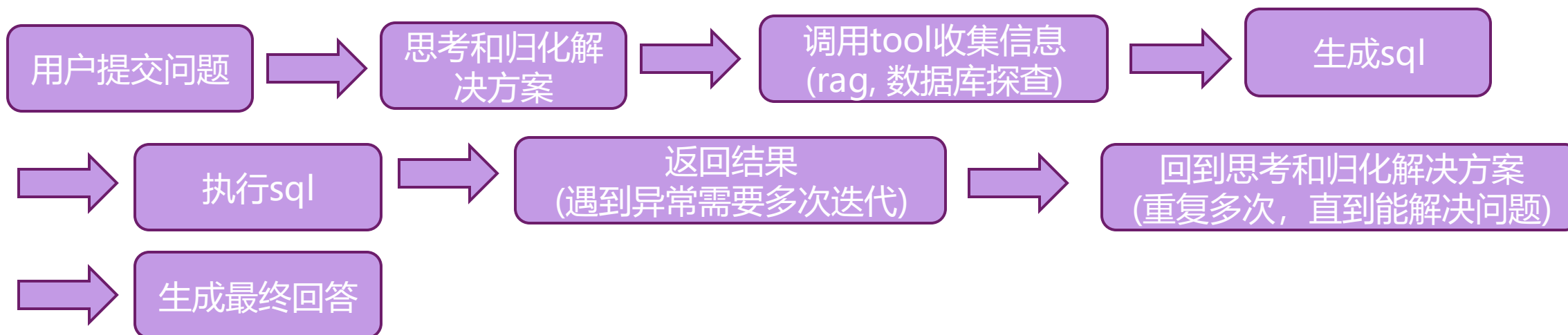


► 什么是Agentic?

1.传统的Agent方案



2.Agentic实现



▶ Agentic的自我纠错

> Task: 帮我注册jd里面的user表



📄 InfiniSynapse wants to register jd: user

⊗ Error We encountered some minor issues and are trying to fix them automatically...



Error register_database_table:

Infinity SQL Error: Error: Failed to execute script: ERROR: relation "public.user" does not exist

位置: 15

✓ API Request



注册user表失败, 错误显示"public.user"表不存在

需要先确认jd数据库中实际存在的表名

使用list_all_tables工具列出jd数据库中的所有表

可能需要使用正则表达式过滤包含"user"关键字的表



▶ InfiniSynapse(RAG)-介绍

知识库大脑：

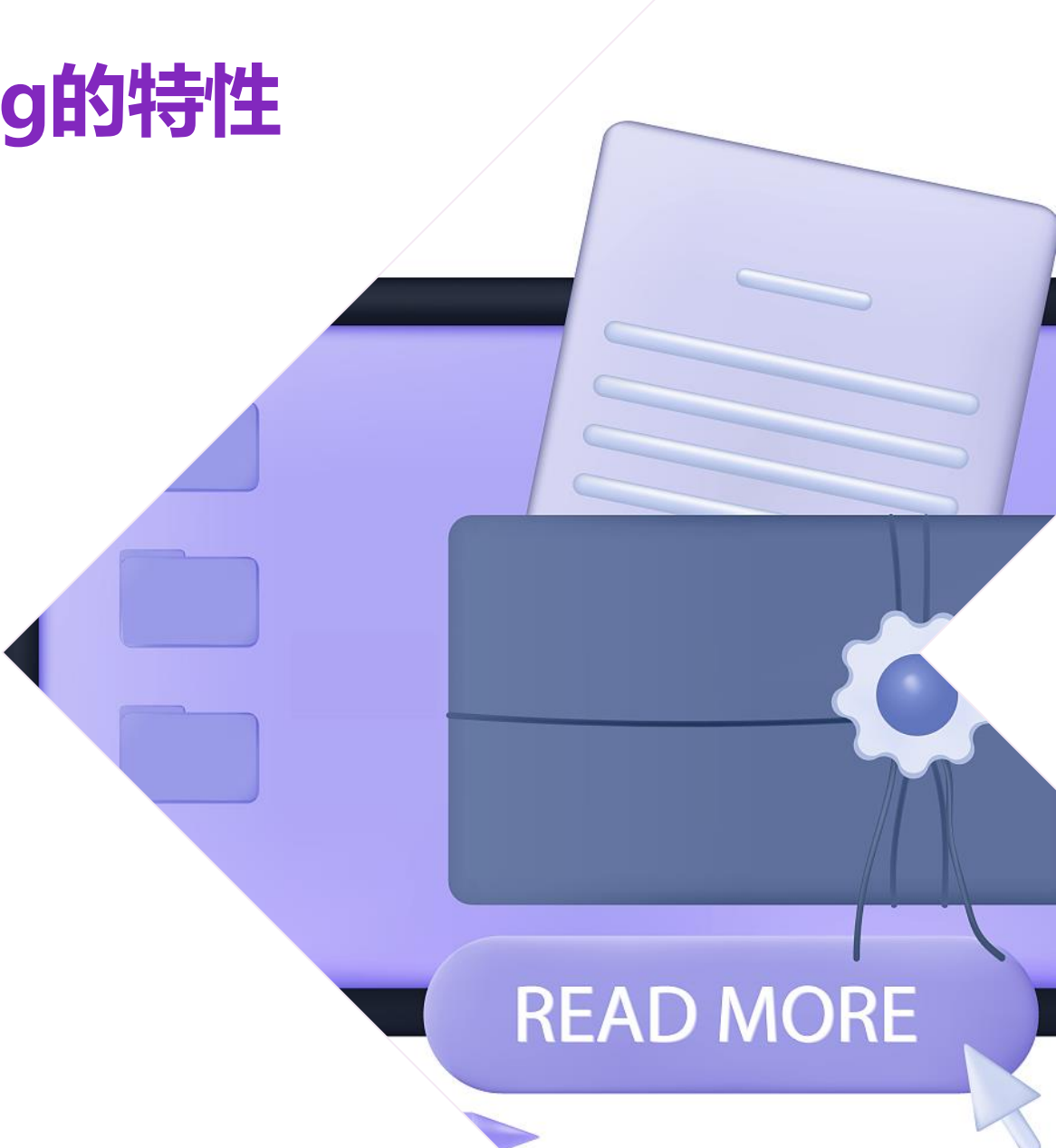
- ★接入各类业务文档：上传文档，InfiniSynapse即刻理解。
- ★沉淀历史任务经验：自动总结过往任务，支持快速复用。
-
- ★解答业务核心问题：实时提供精准的业务概念、流程、规则知识。



► InfiniSynapse(RAG)-Infini Rag的特性

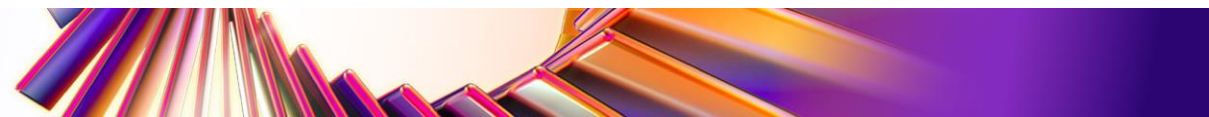
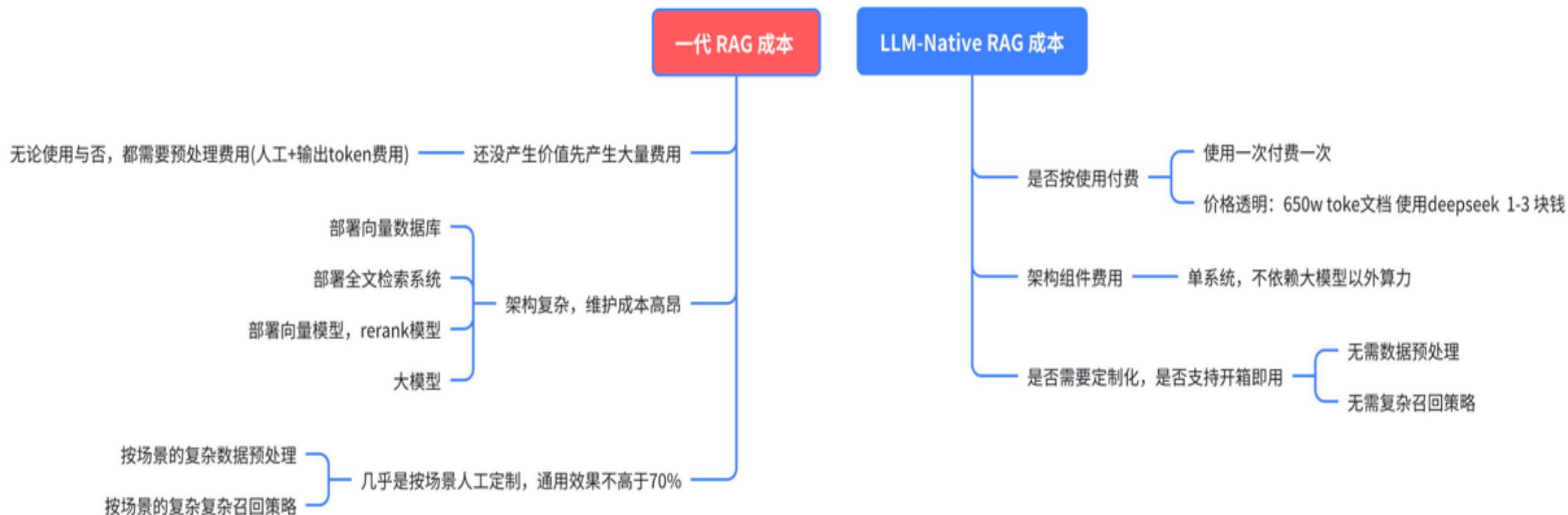


1. 基于LLM-native 完全有LLM驱动, 抛弃了传统emb的方案。
2. 无需数据预处理, PDF ,Word, Excel 等各类文档丢进去后, 启动即可使用。
3. 兼容 OpenAI 接口,任何大模型聊天工具皆可使用。
4. 效果非常优秀, 至少达到 google notebookllm 水平。



READ MORE

► InfiniSynapse(RAG)-成本对比



▶ InfiniSynapse(RAG)-效果对比

	问题等级	auto-coder RAG (0.1.165)		RAGLAB	备注
20240911	简单	100%(16/16)	87.5%(14/16)		
	中等	100%(15/15)	80%(12/15)		
	困难	88.89%(8/9)	44.44%(4/9)		
20240913	困难	80%(4/5)	20%(1/5)		



> Task: MT银行的一级分行有哪几家?

☆ ×

markdown 复制代码

1 MT银行的一级分行信息存储在数据库的哪个表中? 请提供表名和关键字段名称。

Reasoning: ...h deepseek_v3, first round cost seconds, second round cost 0 seconds, input tokens: 0, output tokens: 0, first round full docs: 0, second round extracted docs: 0 Send to model deepseek_v3 with 11618 tokens qa_model_thinking >

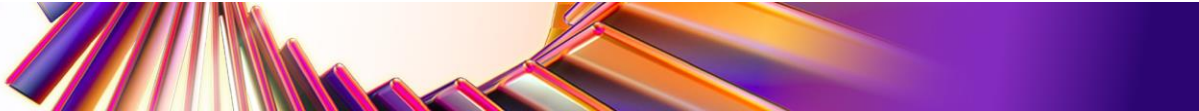
根据提供的文档，MT银行的一级分行信息存储在以下数据表中：

表名

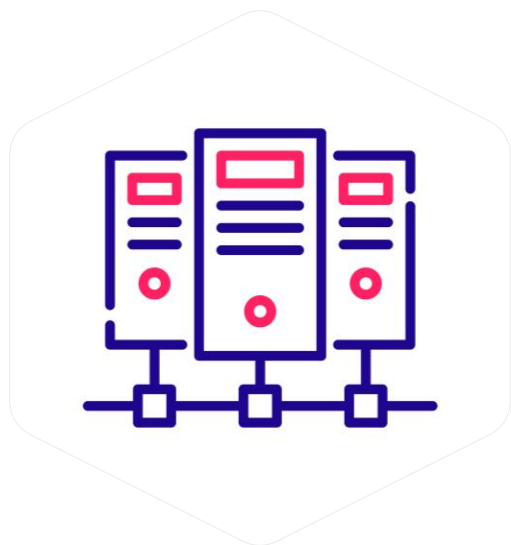
`institutions` (机构维度表)

关键字段

字段名	数据类型	描述	备注
<code>branch_id</code>	VARCHAR(10)	机构唯一ID	主键，例如 <code>SH001</code>
<code>branch_name</code>	VARCHAR(100)	机构官方全称	例如 "MT银行市南分行"
<code>branch_level</code>	INT	机构层级	<code>1</code> : 总行/中心, <code>2</code> : 一级分行, <code>3</code> : 三级支行/网点
<code>parent_id</code>	VARCHAR(10)	上级机构ID	关联主表的 <code>branch_id</code>



► InfiniSynapse(InfiniSQL)-介绍



计算引擎



统一接入异构数据源

无缝适配 PostgreSQL、MySQL 等各类数据库。



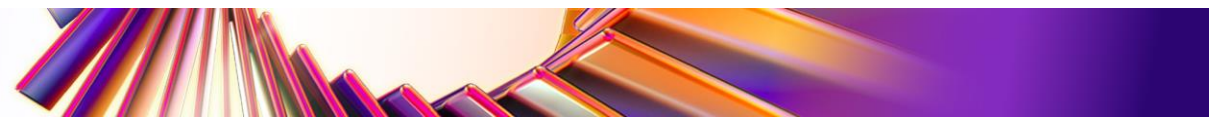
智能执行优化指令

精准运行 InfiniSynapse 生成的查询 (InfiniSQL或数据库SQL) 。



高效分布式计算

分布式计算，易于扩展。



▶ InfiniSynapse(InfiniSQL)-设计理念

1.万物皆表，统一抽象

将结构化/半结构化/非结构化数据乃至模型等实体统一抽象为“表”。

2.语法简洁，学习高效

融合声明式与命令式优点，基础语法接近标准SQL，同时也易于AI理解。

3.统一语言，融合 Data & AI

无需切换语言/工具，使用SQL即可完成从数据接入、ETL、特征工程到模型训练、部署、调用的全流程。

4.强大引擎，生态复用

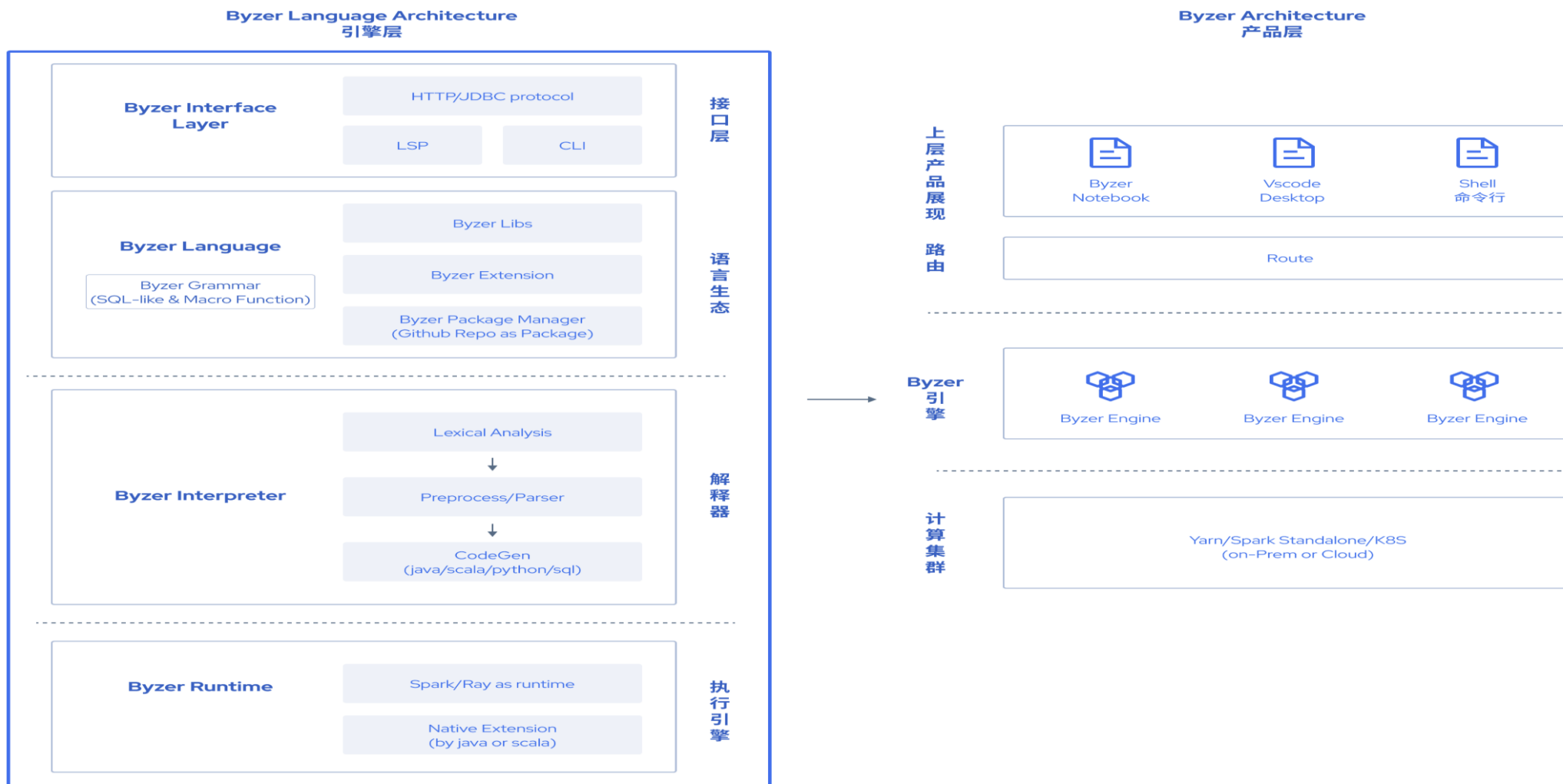
底层采用 Spark (大数据处理) + Ray (分布式计算/AI) 混合执行引擎。

5.无缝集成大模型

"原生支持大模型的部署与调用(ByzerLLM)。



▶ InfiniSynapse(InfiniSQL)- Infini SQL架构图



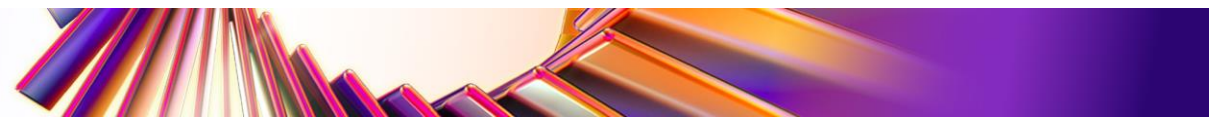
▶ InfiniSynapse(InfiniSQL)- 解决哪些问题

1.提供了AI友好的、更简单的创建临时视图语法，解决了CTE（公用表表达式）冗长的问题。

2.允许同时进行数据库计算和本地计算，便于AI执行优化。

3.具有强大的兼容性，支持多种数据源和计算引擎。

4.原生支持LLM SQL，允许在SQL语句中直接调用模型。



▶ InfiniSynapse(InfiniSQL)- 结合LLM的分析能力

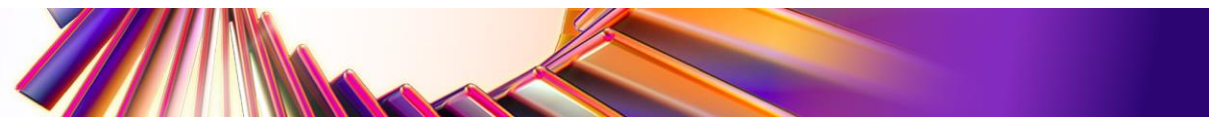
> Task: comment中有哪些是好评?



InfiniSynapse wants to execute SQL queries

sql 复制代码

```
1 load jdbc.`test_comment.test_comment` as comments;
2
3 select /*+ REPARTITION(500) */
4 `id`,
5 `comment`,
6 llm_result(deepseek_v3(llm_param(map(
7   'instruction', llm_prompt('判断以下评论的情感倾向是正面还是负面, 只回答"正面"或"负面":
8
9   评论内容: {0}', array(`comment`)
10  )))) as sentiment
11 from comments as analyzed_comments;
12
13 !cache analyzed_comments session;
14
15 select `id`, `comment`
16 from analyzed_comments
17 where `sentiment` like '%正面%' as positive_comments;
```



适用于数据分析人员的Plan模式

▶ 适用于数据分析人员的Plan模式



针对复杂数据分析问题，AI生成的执行计划可能存在局限性。此模式倡导数据分析师深度介入，引导AI执行过程，融合人类专业洞察与AI计算效率，共同制定并优化解决方案路径。

▶ InfiniSynapse-Plan Mode

> **Task:** 我想基于学生数据，分析比较火的app所具有的特征？



分析维度：您更关注哪些方面的特征？例如：

功能类别（社交、学习、游戏等）

使用时间段

社交属性

用户评价

建议的分析步骤：

首先获取数据集结构

根据定义的热门标准筛选出热门app

从多个维度分析这些app的共同特征

通过统计分析和可视化展示结果

您对这些步骤有什么建议或需要调整的地方吗？

基于使用频率定义热门app



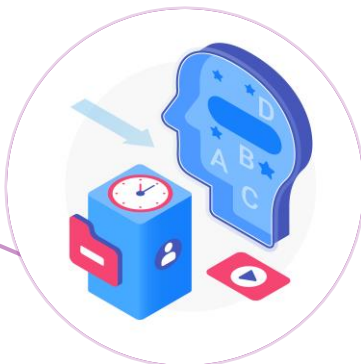
基于使用时长定义热门app

先查看数据集结构



Workflow的总结和自我归纳

知识“蒸馏”与总结加速任务完成



知识“蒸馏”与总结

为了提升任务执行效率，我们在完成复杂任务后，会进行关键知识的“蒸馏”与总结。这些提炼后的精华将通过 RAG (检索增强生成) 技术进行存储和索引。这样，在后续遇到类似任务时，就能快速检索并复用这些知识，显著加速完成过程。



选择RAG

Select RAG:

mt_rag

总结

知识蒸馏：学生数据集中最火的APP分析

1. 任务目标

核心目标：从学生数据集中找出最受欢迎的社交平台，并生成排名图表。

2. 数据库和表总结

数据文件： `student_social_data`

关键列：

`Most_Used_Platform` (记录学生最常使用的社交平台)

Cancel

Confirm

PART 05

未来的展望

功能层面

实现与BI系统及低代码报表平台对接：将AI分析结果无缝集成至主流BI工具或低代码报表平台，实现分析成果的可视化展示与深度交互。

构建行业分析报告自动生成能力：基于多源数据与子任务协同，驱动AI自动生成结构化的、具有深度的行业洞察与分析报告。

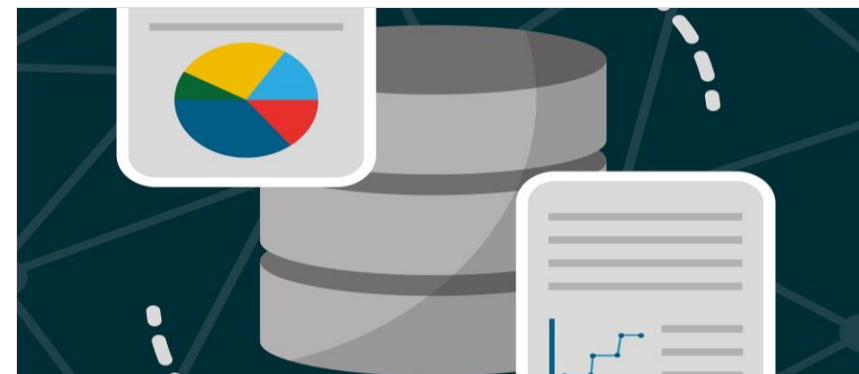
优化用户体验与系统交互：持续改进用户界面与交互设计，提升系统的易用性、直观性和操作效率。

技术层面



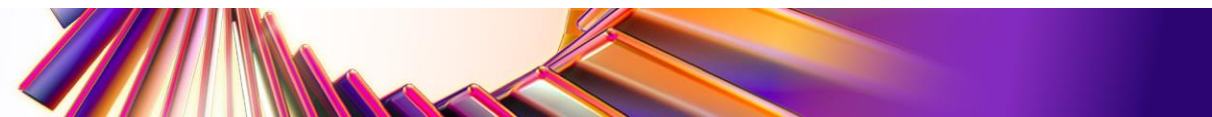
深化AI对SQL执行优化

运用AI算法持续精炼SQL语句，增强查询效率，缩短响应时间，实现系统性能飞跃。



强化LLM与SQL的协同分析能力

深化大型语言模型（LLM）与SQL引擎的融合，提升复杂数据查询、理解与分析的能力。





第8届AI+研发数字峰会

拥抱AI 重塑研发 AI+ Development
Digital Summit

下一站预告

11/14-15 | 深圳站

12/19-20 | 上海站



查看会议详情

深圳站论坛设置

智能装备与机器人

超越“编程 Copilot”

下一代知识工程

智能网联与汽车智能化

AI 测试工具开发与应用

AI 基础设施和运维

数据智能及其行业应用

可信 AI 安全工程

大模型和 AI 应用评测

多 Agent 协同框架

从智能测试到自主测试

大模型推理优化

多模态 LLM 训练与应用

智能化 DevOps 流水线

上下文工程

AiDD

「深行 · 浅智」

Walk Deep, Think Light.

2025.11.16

AiDD首届麦理浩径徒步





科技生态圈峰会 + 深度研习

——1000+ 技术团队的选择



AiDD峰会详情





第7届AI+研发数字峰会
AI+ Development Digital Summit

感谢聆听!

扫码领取会议PPT资料

