

DiPEAK

大模型趋势下的企业数据体系思考

何昌华 —— 数巔科技 CEO



何昌华



让大模型与企业数据充分协同

解锁数据潜力，释放企业数据价值

DiPEAK 2022 – 数巔科技CEO



2017 – 国家级人才计划



2017 – 蚂蚁集团
计算存储首席架构师

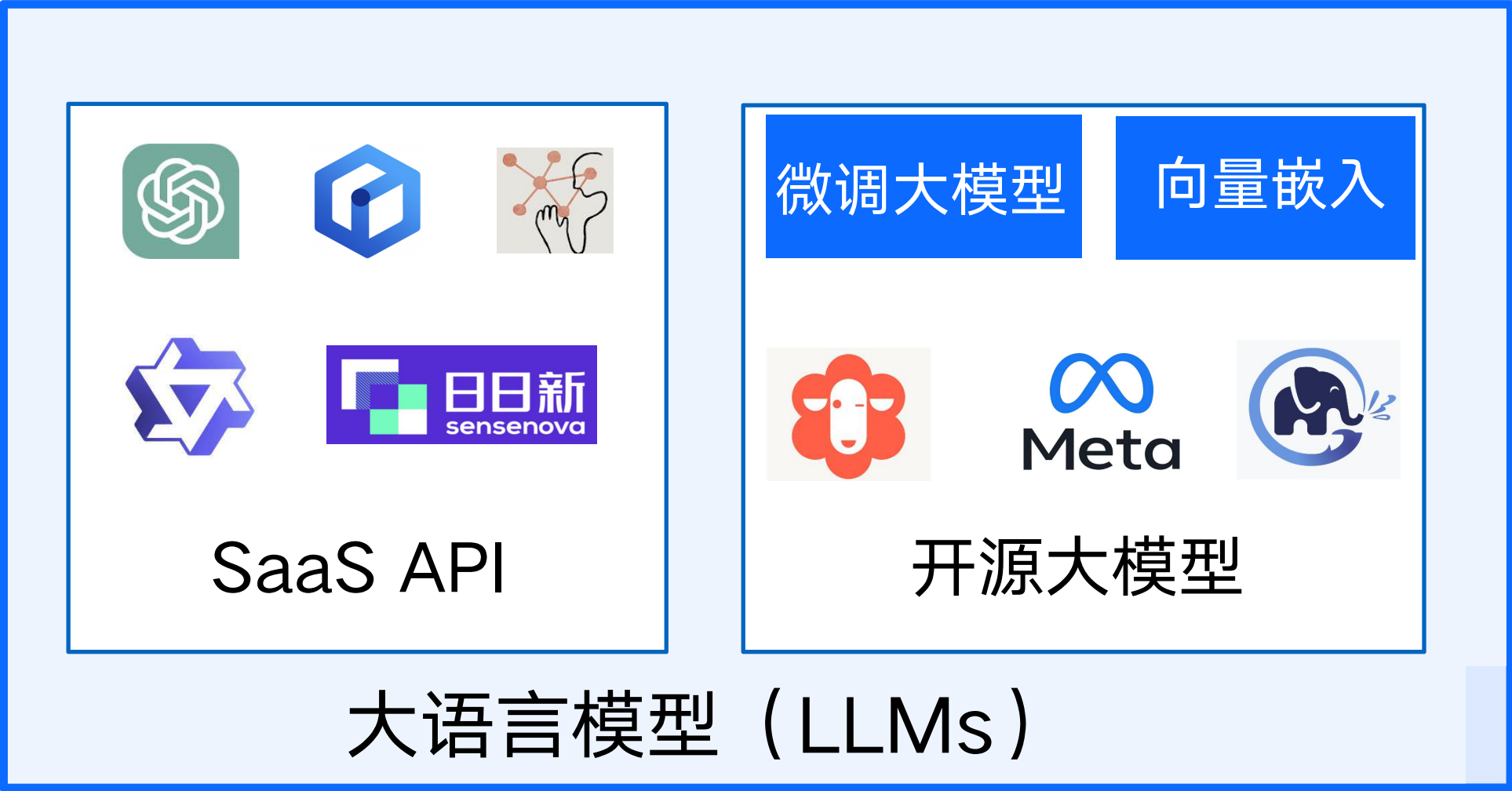
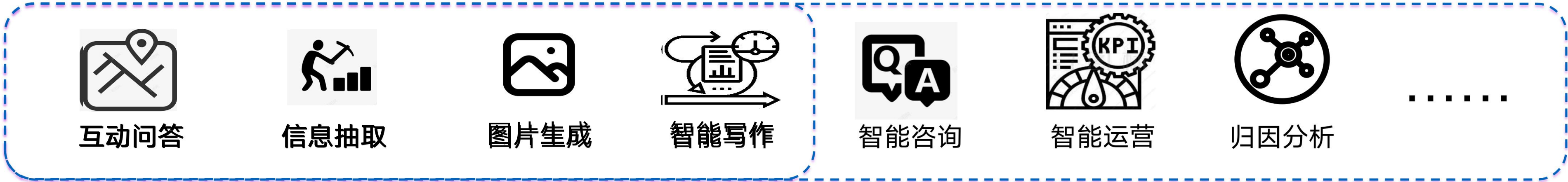


2015 – Google搜索引擎
架构核心技术负责人

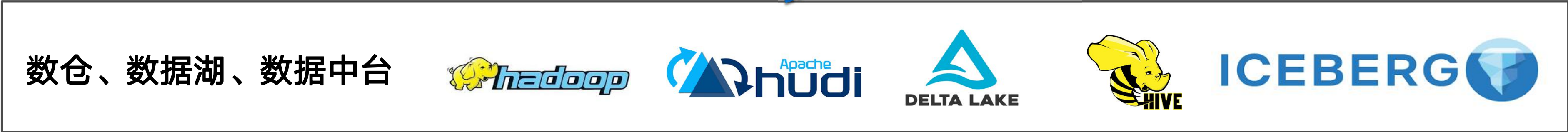


2005 – 斯坦福大学博士

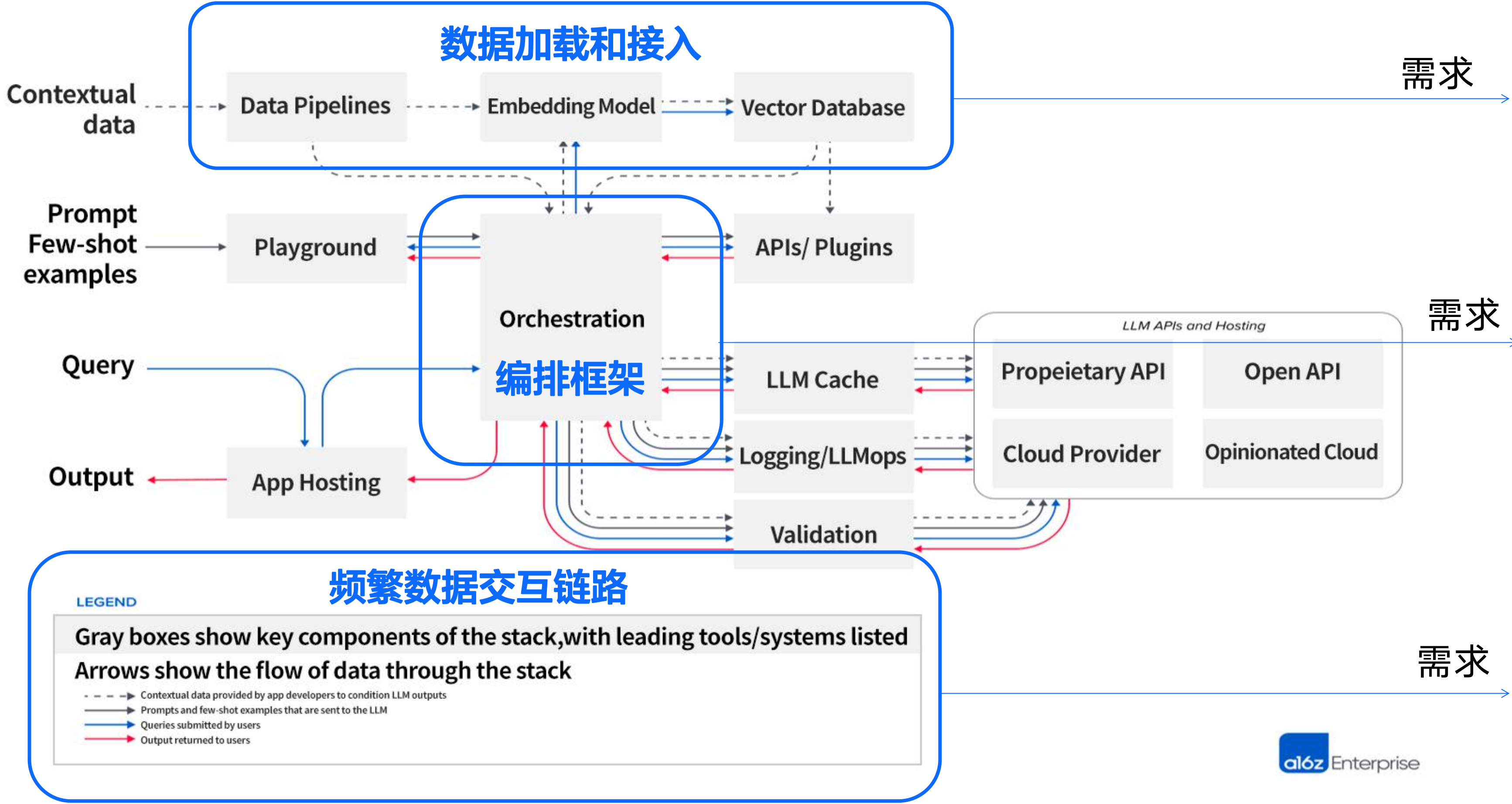
当 LLMs 遇到企业应用：需要与企业数据协同



实时数据	更新文件 视频监控	每日交易数据 行为日志
	存档文件 规章制度	历史日志 过往交易
历史数据	非结构化数据	结构化数据



大模型在企业落地对数据体系的关键需求



统一数据资产

- 多模态数据需要多种加载、转换方法
- 需要管理及沉淀优质资产

自动化工具（包括常用模型）

- 自动编排执行任务需要各种插件工具和模型算法自动化驱动

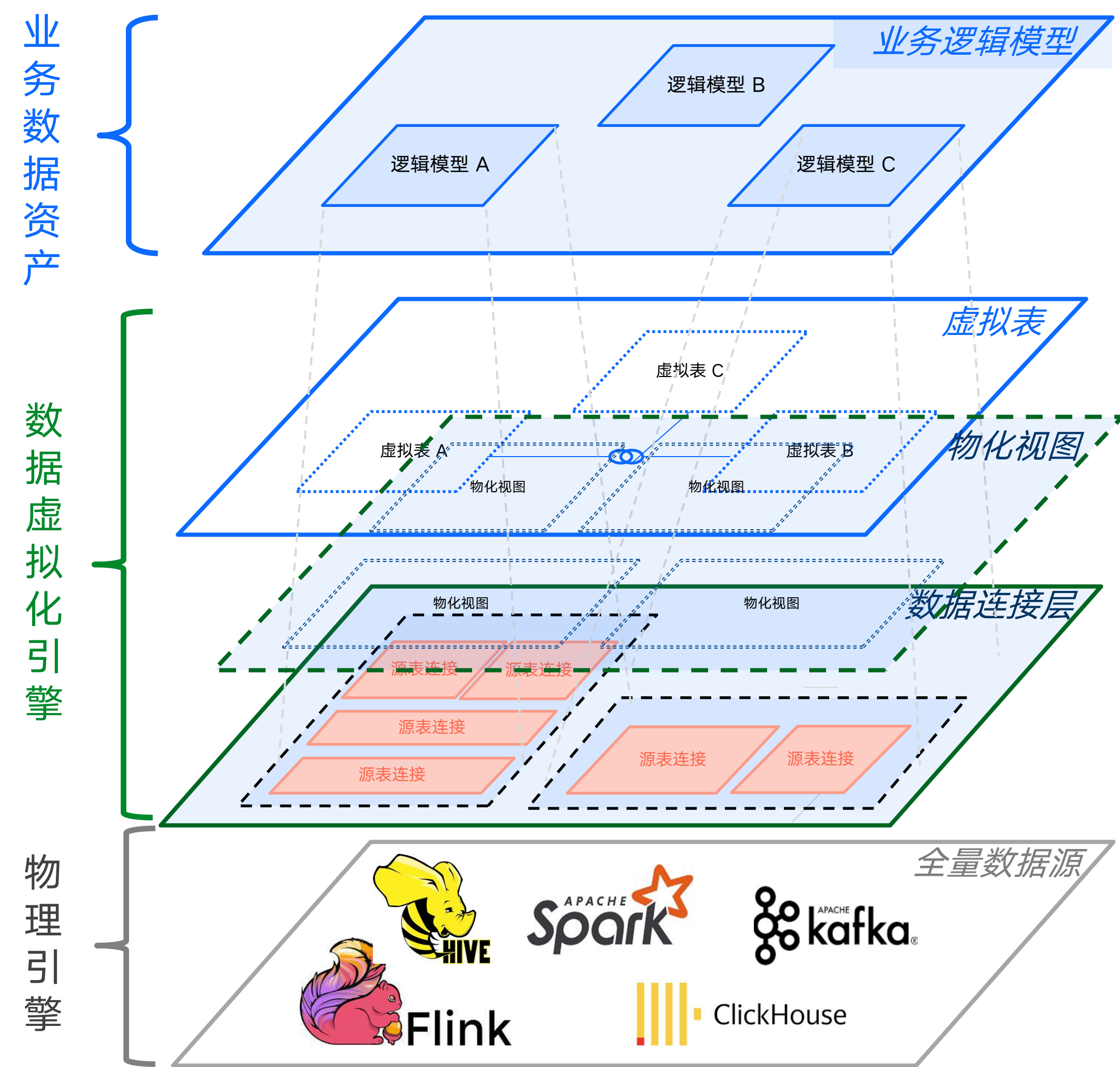
高效计算能力

自动链路中数据频繁交互带来计算量倍增

——世界顶流科技投行a16z 2023年6月发布最新大模型应用一般流程

如何解决：智能数据虚拟化技术

- 统一
将异构数据源虚拟化，沉淀统一语义的**数据资产**
- 自动
根据业务语义实现资产即刻**自动扩充**，驱动自动化工具
- 高效
智能加速实现数量级性能提升



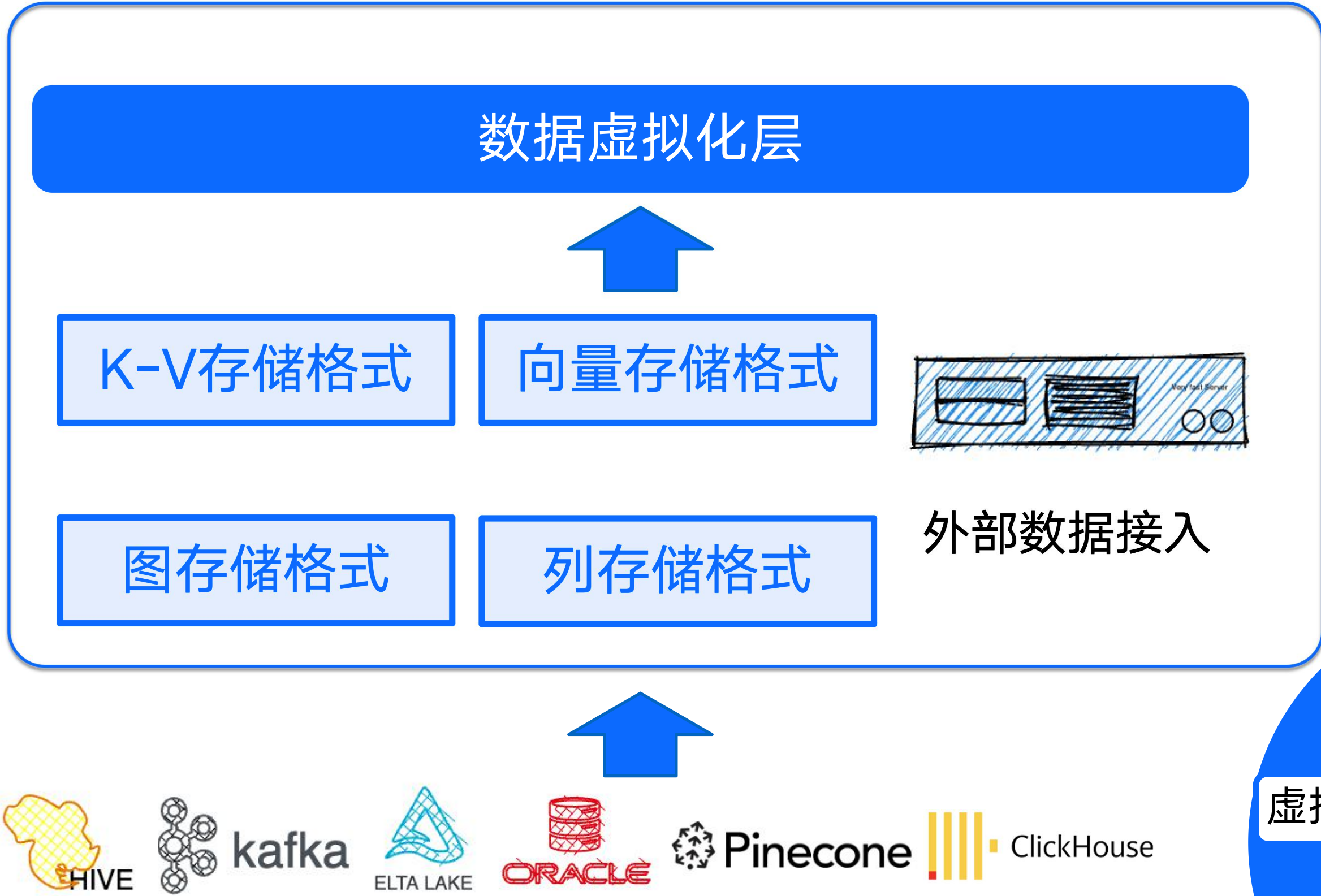
数据虚拟化引擎全面实现三大需求



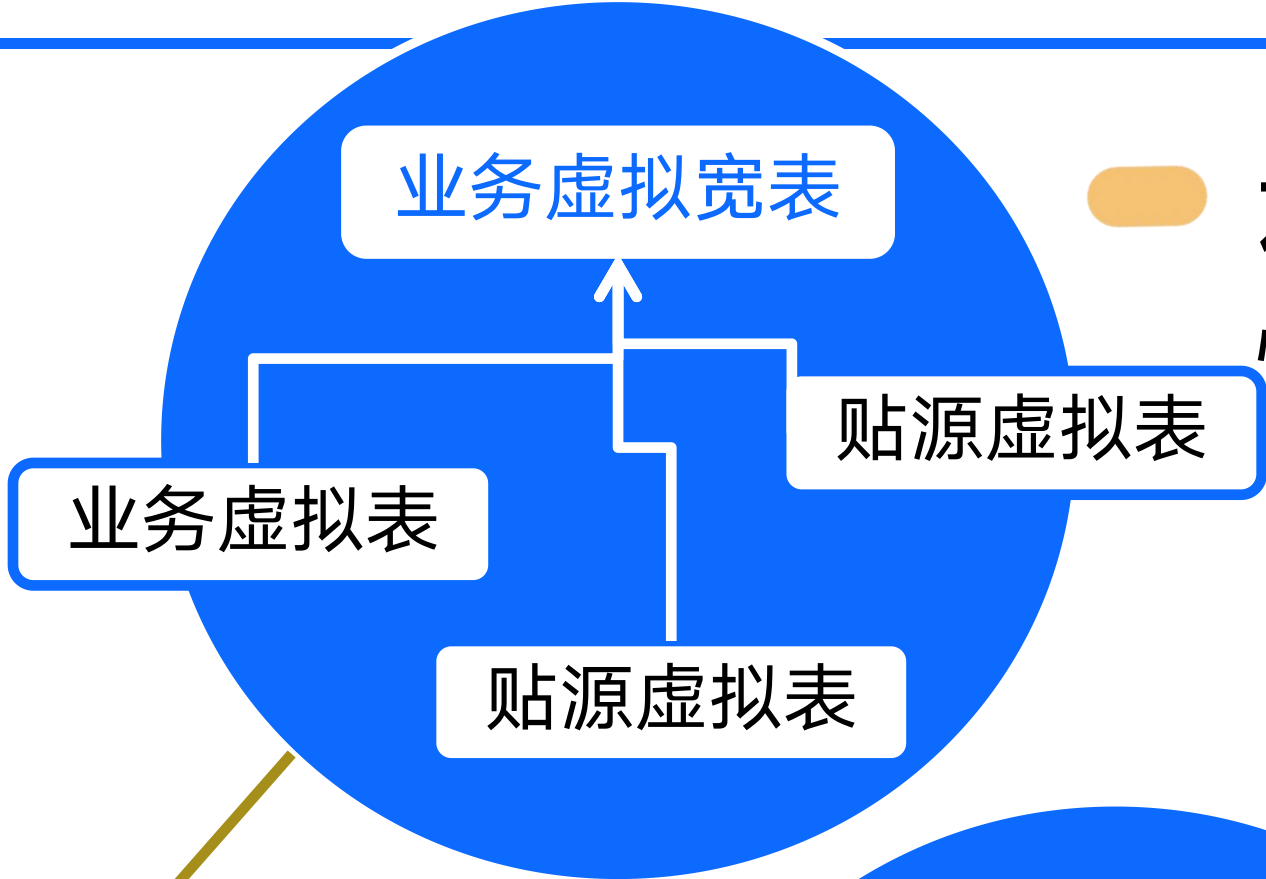
实现需求一：统一数据资产

更多数据治理方案请关注7月22日【智能化数据治理】分论坛

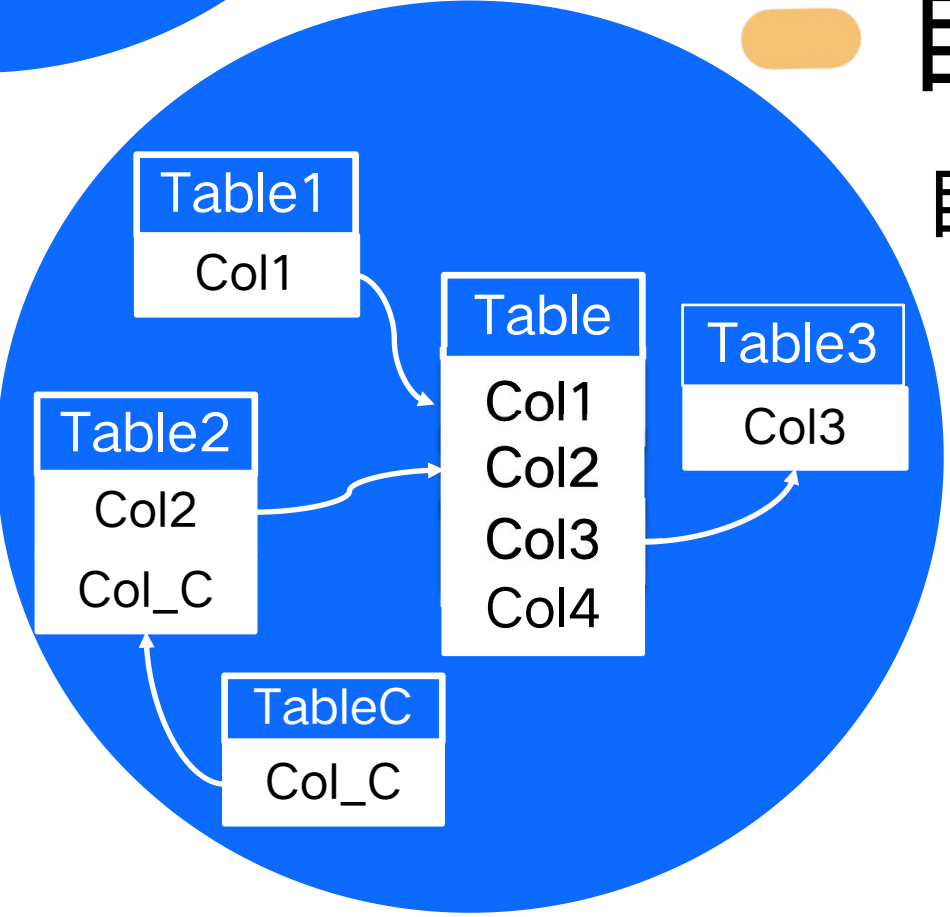
数据接入 虚拟表批量映射，内置多模态存储加速



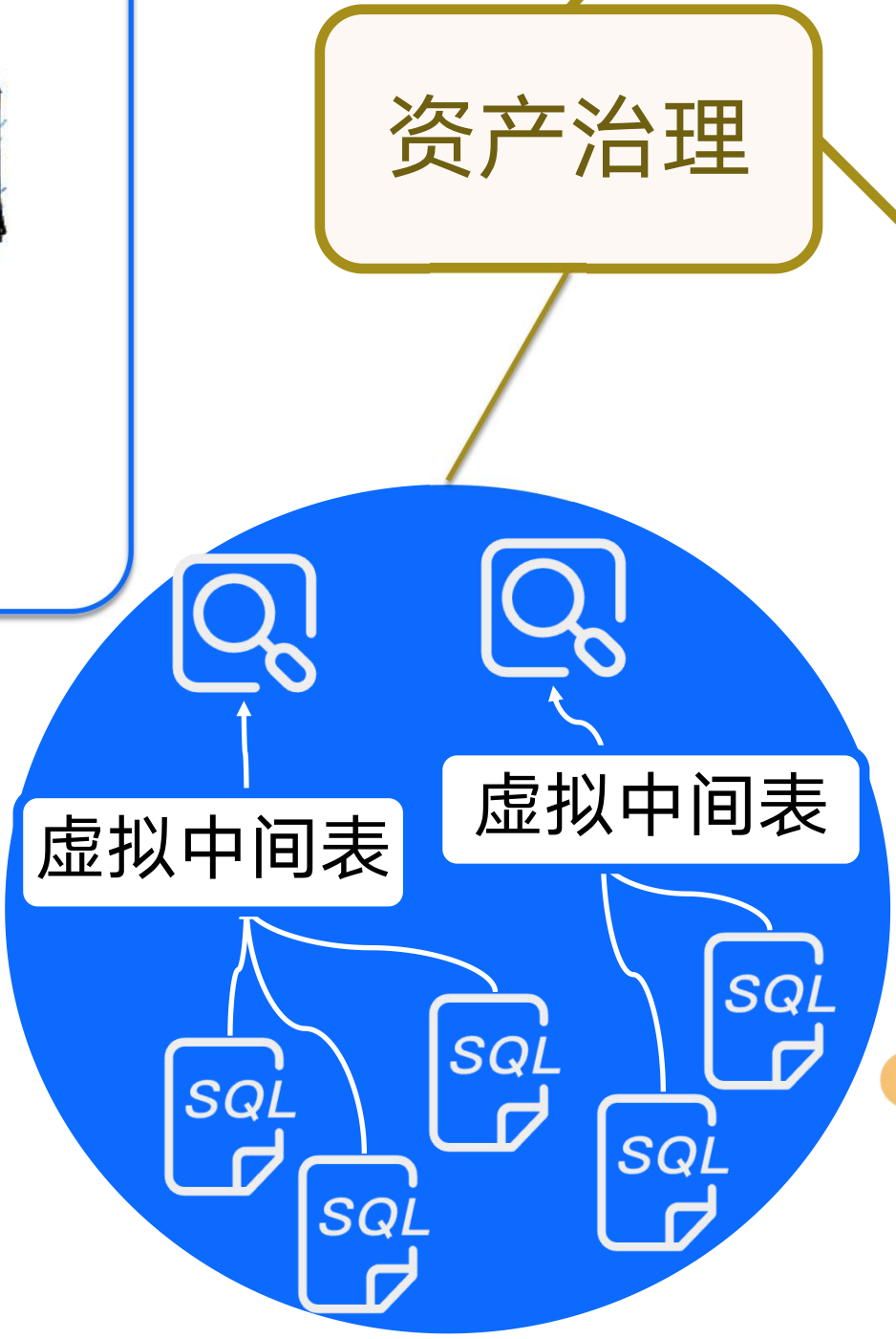
逻辑建模 快速灵活业务建模



自动血缘 自动构建字段级血缘

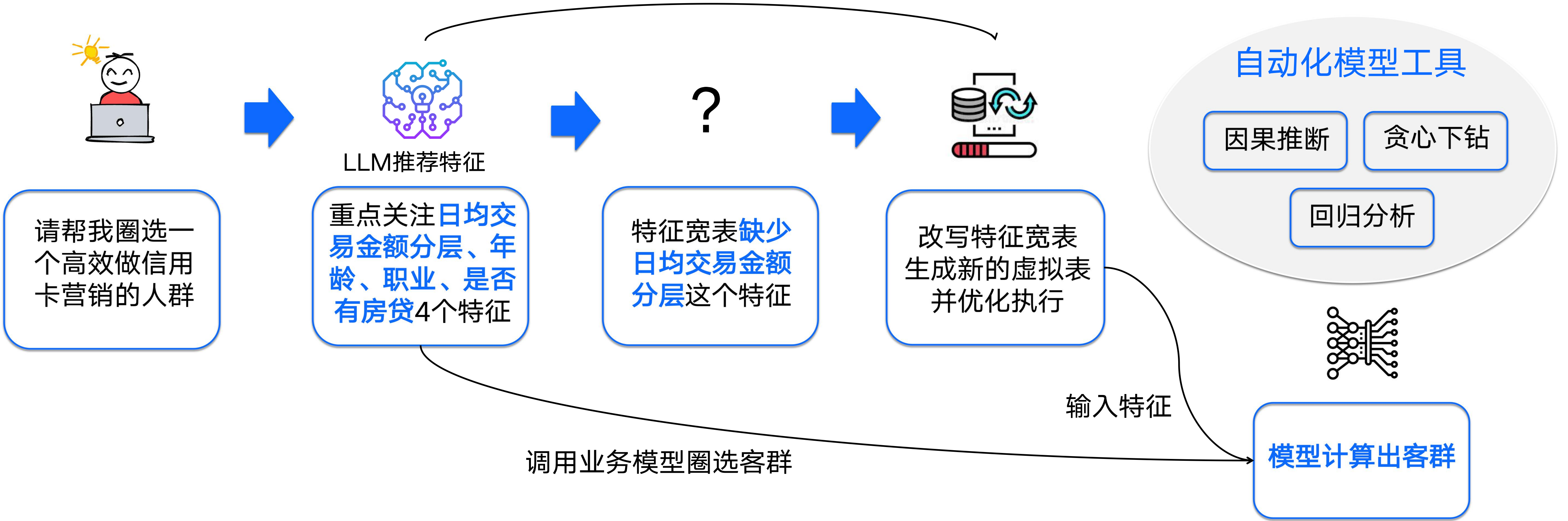


自动合并等价SQL 避免资产二义性



实现需求二：数据驱动的自动化工具

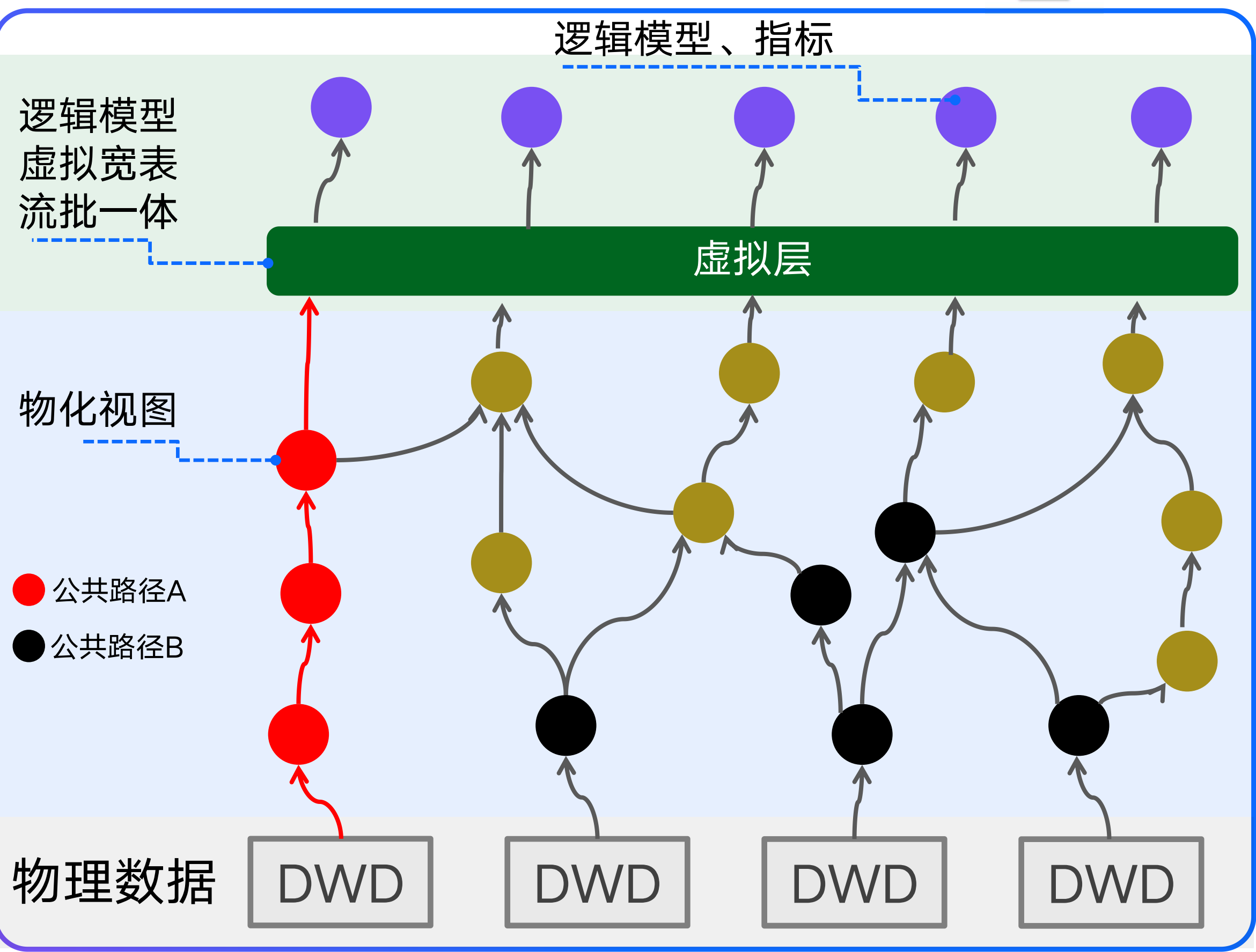
大模型依据已有【日均交易金额】字段推荐特征加工逻辑



实现需求三： 高效计算能力

⚡ 全技术栈智能优化，性能整体可提升10倍以上

智能加速
智能物化+SQL优化



先进计算存储底盘

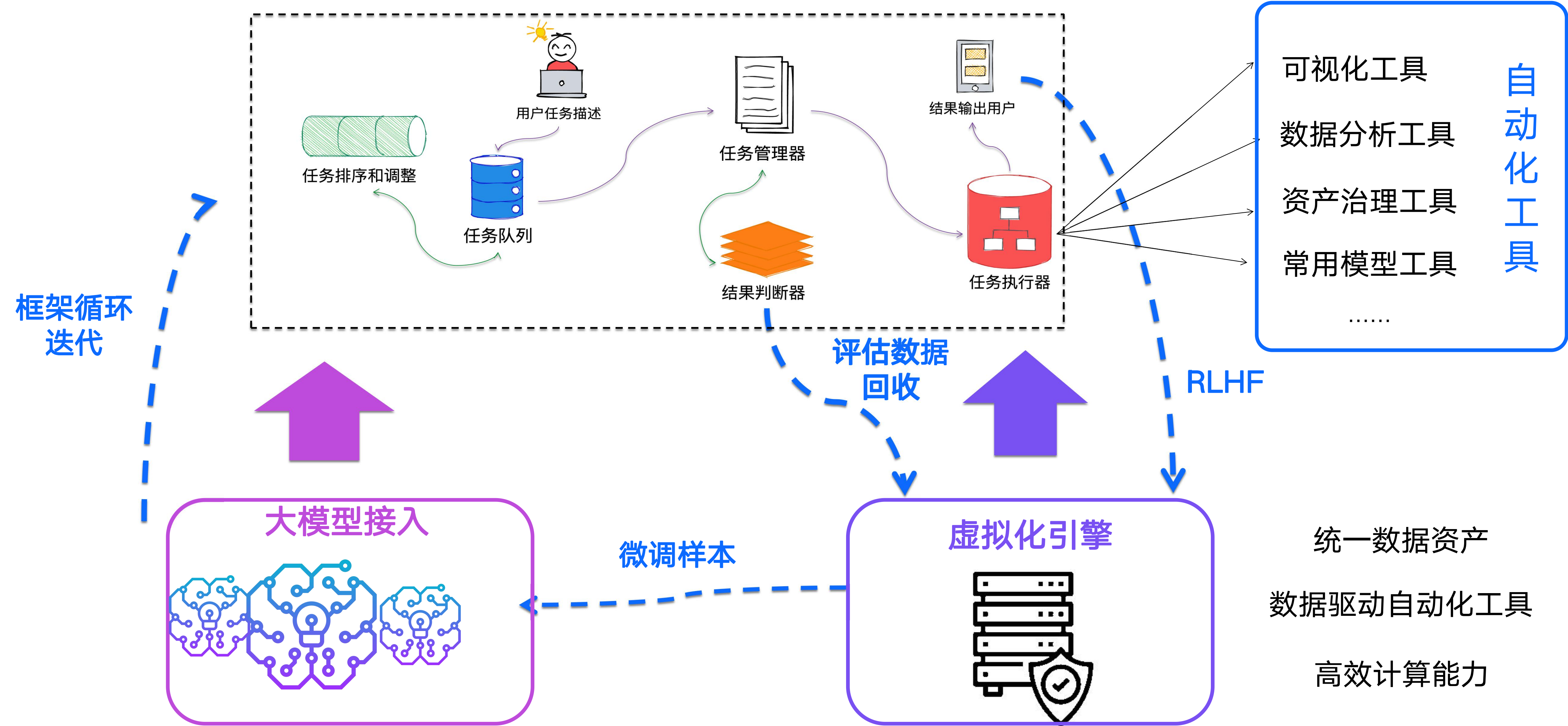
内置多格式高速存储

1 KV、Table、TS，Graph。性能10倍于主流KV存储。

计算性能优化

- 1 Join性能做到极致，2倍于主流引擎。
- 2 兼容BSP、MPP、Streaming等业界通用引擎。

自我迭代的企业大模型应用框架



数巔科技，让大模型和企业数据协同起来，实现广泛业务场景下的智能决策

You need both LLMs and data systems; Harness their synergy



感谢聆听

公众号请扫码：



官方网站

<https://www.dipeak.com>

邮箱

biz@dipeak.com

微信公众号

数巅科技

