从数据到智能

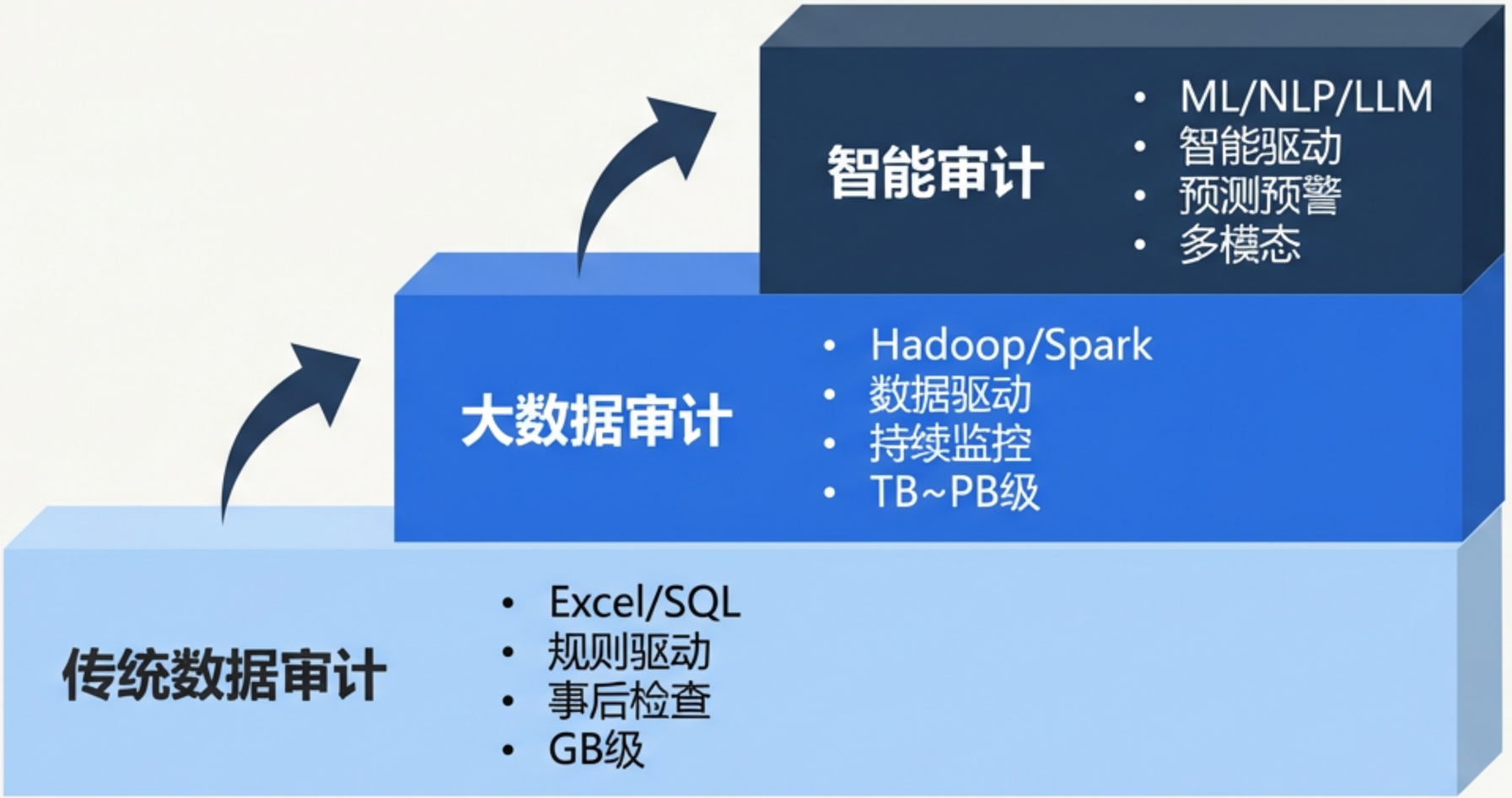
审计数字化转型全景 — 传统数据审计·大数据审计·智能审计



课程总览：两天四模块


Day1上午
传统数据审计方法论
— 方法·工具·流程
Day1下午
大数据审计基础与架构
— 全量·平台·治理
Day2上午
大数据审计分析实践
— 模型·场景·可视化
Day2下午
智能审计与未来趋势
— AI·LLM·转型

数据审计发展三阶段



三阶段是能力叠加，不是替代关系

传统数据审计：五大核心方法

	**账表核对法** — 科目余额表↔明细账↔凭证
	**趋势分析法** — 月度收入同比增长突变检测
	**比率分析法** — 毛利率偏离行业均值>2σ
	**分层抽样法** — 按金额分层，大额全查
	**交叉验证法** — 发货量×单价≈开票金额

核心逻辑：多源数据交叉比对，发现不一致

四大常见审计场景



收入确认审计 — 发货→开票→回款时间线匹配，跨期确认风险



费用报销审计 — 标准超限检测，重复报销识别，关联交易



往来款审计 — 账龄分析，长期挂账预警，坏账准备



资产审计 — 折旧计算正确性，账实一致性

传统审计工具矩阵

	数据分析	异常检测	项目管理	协作
Excel	●	◔ (限<100万行, 透视表/VLOOKUP)	○	○
SQL	●	● (多表关联/GROUP BY)	○	○
ACL/IDEA	●	● (规则引擎/连续审计)	○	○
TeamMate	○	○	●	●

传统审计五大瓶颈 — 转型是必答题

- **数据量瓶颈：**GB级数据，单机处理力不从心
- **类型单一：**仅结构化数据，无法处理合同/影像
- **时效不足：**事后审计，风险已成既定事实
- **深度有限：**规则驱动，难发现隐蔽模式
- **人力依赖：**经验驱动，质量因人而异



结论：转型不是选择题，是必答题

大数据4V的审计映射

V

Volume（量）



全量审计替代抽样审计

V

Velocity（速）



实时监控替代事后检查

V

Variety（多样）



合同+发票+日志+影像

V

Value（价值）



从发现问题到预测风险

全量审计 vs 抽样审计

- 基于统计学假设
- 存在抽样风险
- 可能遗漏关键异常
- ✔ 仍适用：物理盘点、实地访谈

VS

- **100%数据覆盖**
- 消除抽样风险
- 发现低频异常模式
- **✔ 适用：电子化交易数据**

不是取消抽样，而是在能全量的地方全量

大数据审计思维转变

找错 → 画像

从查具体违规到主体行为全景刻画

核对 → 关联

从两张表对数到跨系统关系发现

事后 → 事中

从年度审计到嵌入业务流监控

大数据审计技术架构

应用层 — 审计模型管理 | 可视化报告 | 预警工单

计算层 — 离线(Spark) | 实时(Flink) | 交互(Presto)

存储层 — 数据仓库(ClickHouse) | 数据湖(HDFS/S3) | 图数据库

采集层 — ETL(DataX) | CDC(实时) | OCR | API对接

数据源 — ERP | 财务 | OA | HR | 合同 | 外部数据

审计数据治理三支柱


可信审计数据
数据标准化
• 跨系统字段映射
• 主题域划分
• 数据字典
数据质量
• 完整性
• 一致性
• 准确性
• 时效性
数据安全
• 敏感性分级
• 脱敏策略
• 访问控制
• 操作留痕

审计模型设计方法论

合规类

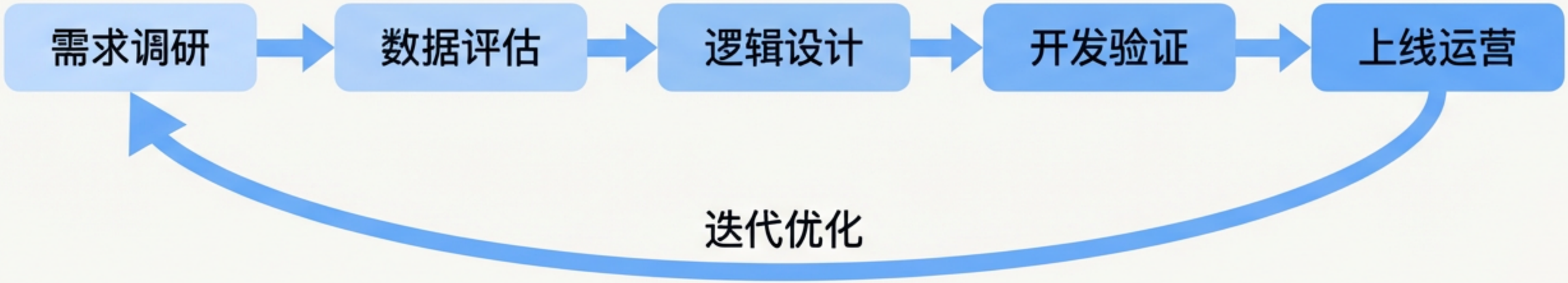
规则匹配/阈值判定（单笔超限、审批缺失）

分析类

趋势预测/异常检测（费用突增、指标偏离）

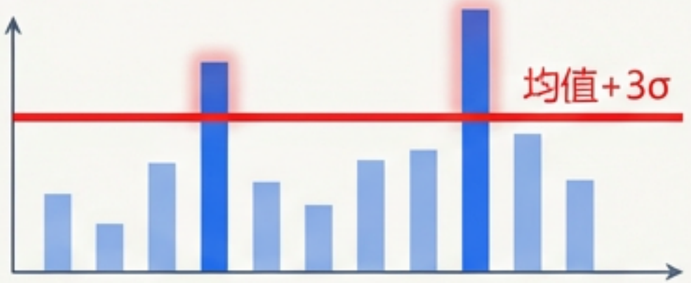
关联类

关系网络/资金链路（关联交易、循环转账）



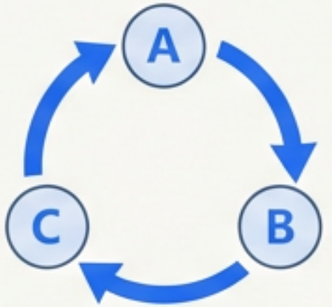
资金流分析四维度

大额交易



突发大额转账预警

资金闭环



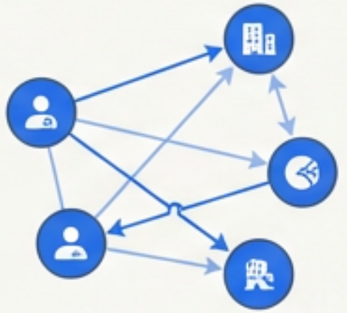
循环转账识别

异常时间



非常规时段集中交易

关联网络



股权穿透+交易图谱

收入审计 & 费用审计要点

收入审计

- 收入-成本配比异常
- 客户集中度分析
- 跨期确认风险检测

费用审计

- 费用-业务量弹性分析
- 供应商集中度分析
- 发票批量验真

共同逻辑：多源数据交叉比对发现不一致

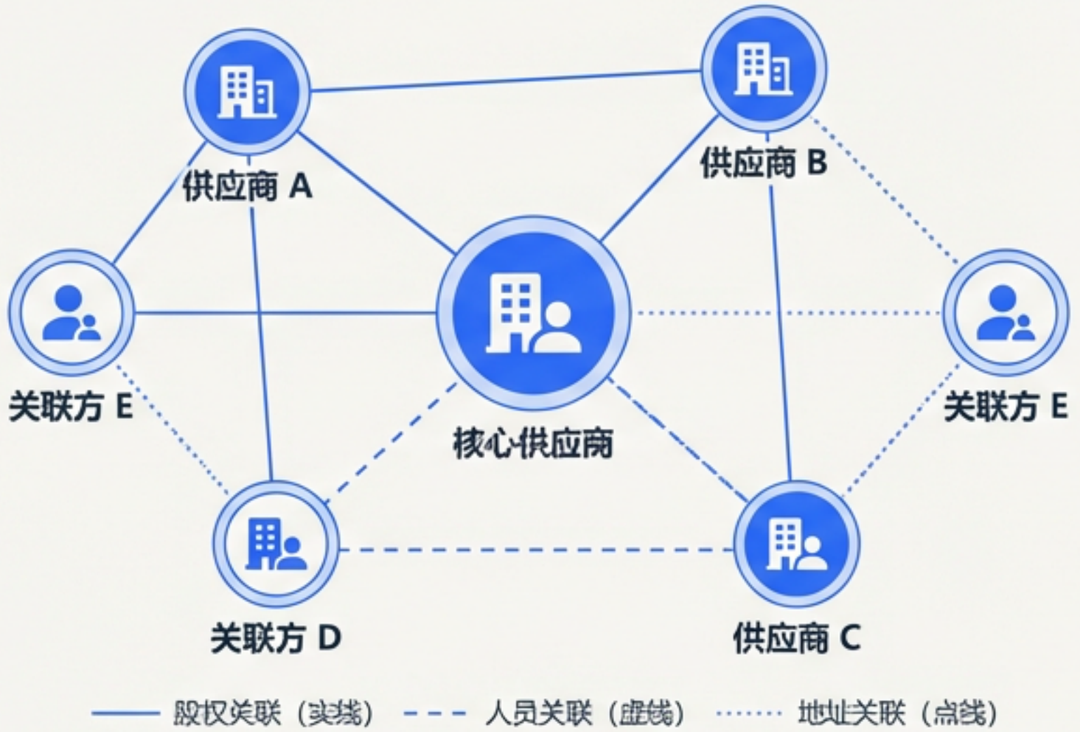
采购审计：围标串标与供应商网络

围标识别



报价高度接近，价差<1%

供应商关系网络



多维关联发现隐蔽关系

工程项目 & 人力资源审计



工程项目审计

- 工程量清单 vs 实际施工量
- 变更签证合理性分析
- 超进度付款预警



人力资源审计

- 同岗位薪酬离散度检测
- '吃空饷'多源比对：社保 ✓ 工资 ✓ 考勤 ✗ → 预警
- 绩效-薪酬一致性分析

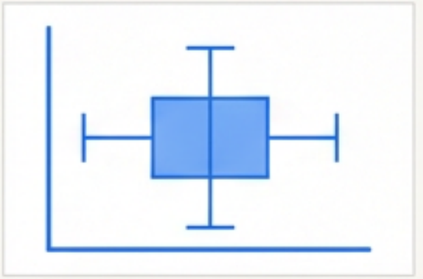
审计数据可视化与报告转型



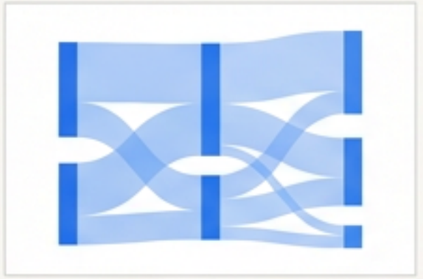
折线图(趋势)



网络图(关系)



箱线图(分布)



桑基图(流向)



树图(占比)



热力图(地理)



静态报告

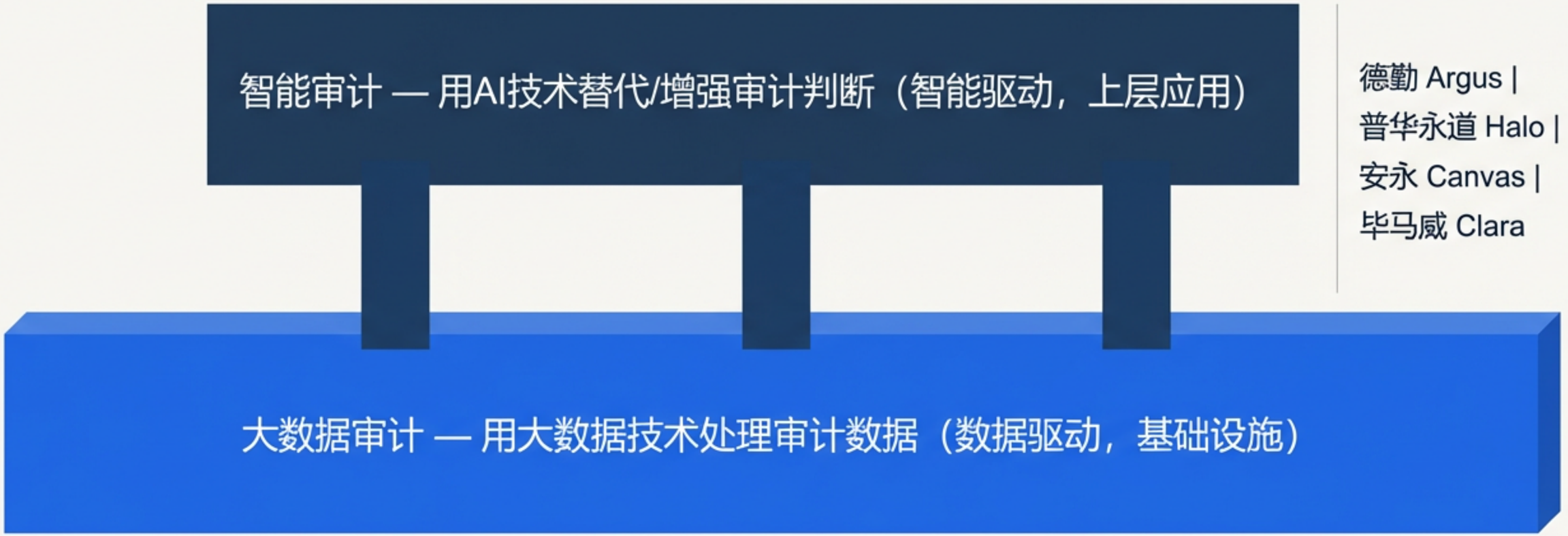




交互式看板

审计叙事逻辑：问题→证据→影响→建议

智能审计：从数据处理到判断增强



德勤 Argus |
普华永道 Halo |
安永 Canvas |
毕马威 Clara

递进关系：大数据是基础设施，AI是上层应用

机器学习×审计

监督学习

标注数据 → 训练模型 → 预测分类

挑战：审计标注样本稀缺

无监督学习

原始数据 → 聚类分析 → 异常发现

优势：适合探索性分析

算法应用应用场景

算法应用		应用场景
Isolation Forest	→	交易异常检测
K-Means聚类	→	供应商/客户行为分群
随机森林	→	财务舞弊预警评分

NLP×审计：解锁非结构化文本



合同审查 — 关键条款提取，风险标注，效率提升10x+

条款自动识别，风险等级分类，高亮显示



底稿自动化 — 工作记录结构化，自动生成

工作流整合，格式化输出，减少人工录入





舆情监控 — 被审计单位负面信息实时预警

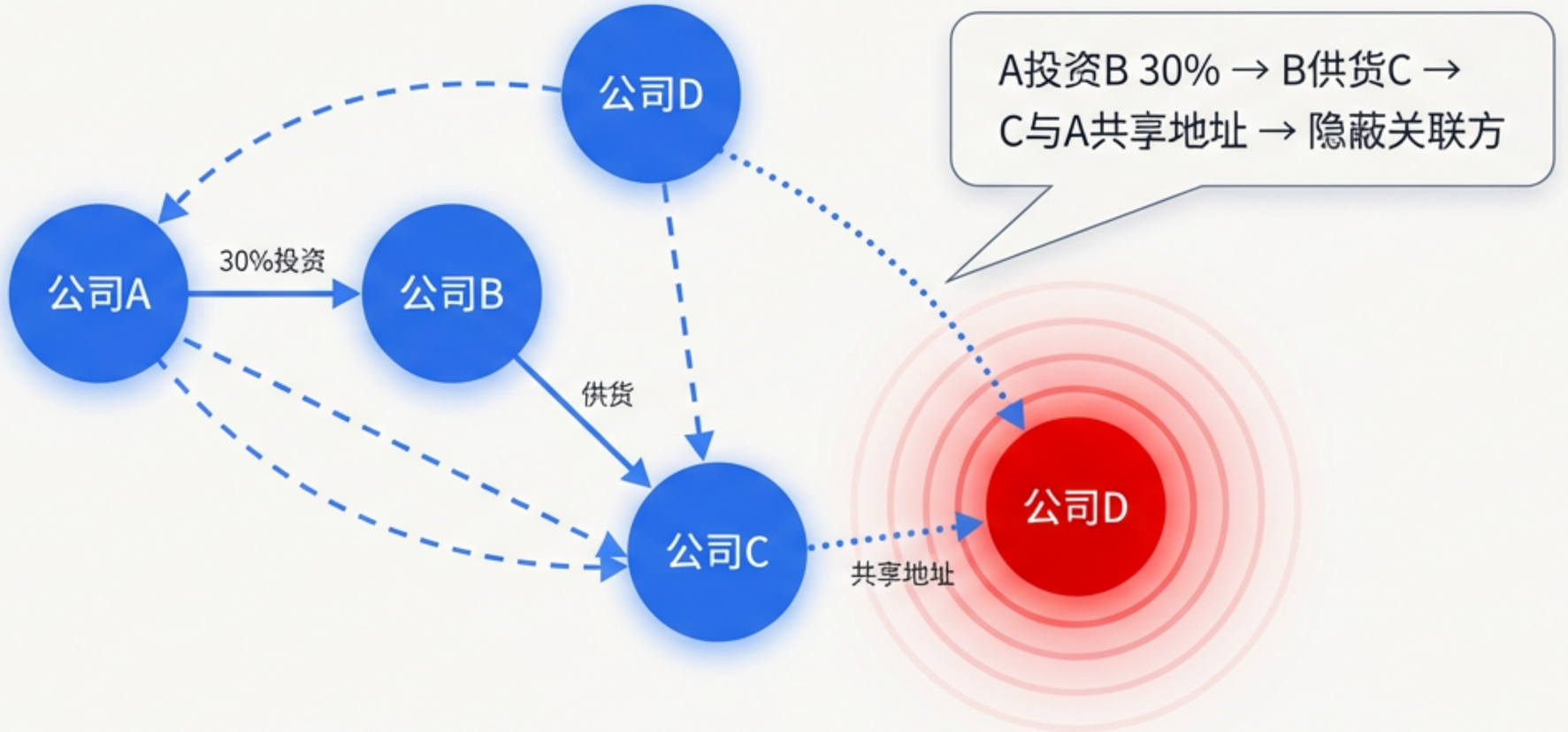
全网数据采集，关键词监控，风险等级预报



决策还原 — 会议纪要/邮件分析，审批链路追溯

因果关系识别，时间轴可视化，流程还原

知识图谱：隐蔽关联的发现引擎

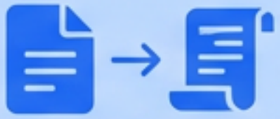


大语言模型与审计：能力与边界

能力



审计问询
（智能问答）



报告生成
（数据→叙述）



方案编制
（历史推荐）



法规解读
（合规建议）

RAG — 企业审计知识库+
精准条款引用，减少幻觉

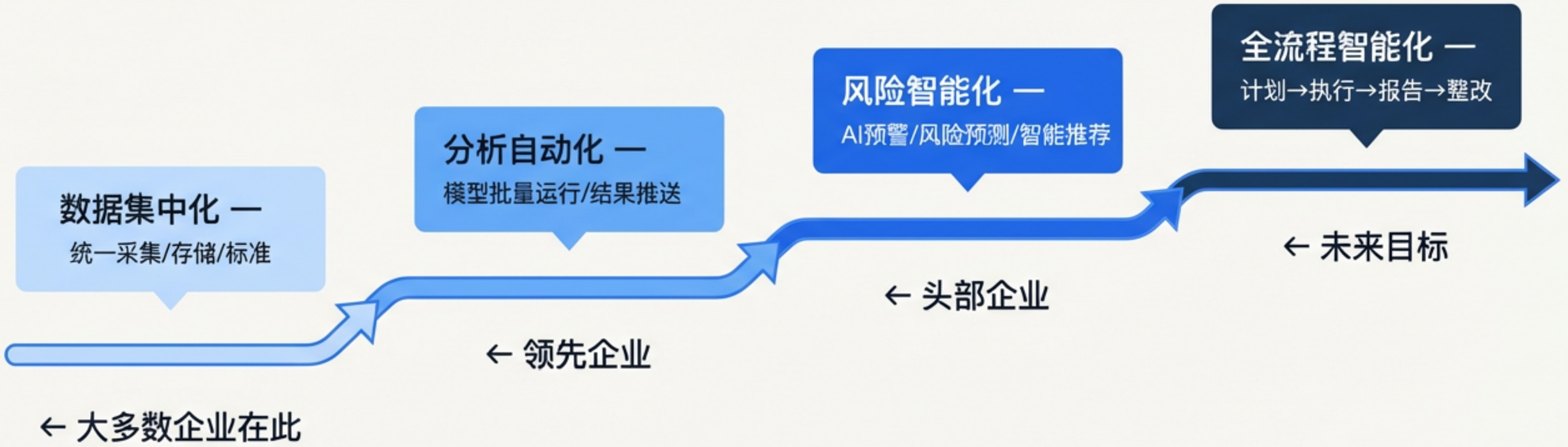
边界

四条红线

1. ⚠ **数据安全：** 审计数据不可外泄 → 私有化部署
2. ⚠ **判断不可委托：** AI辅助 ≠ AI替代
3. ⚠ **证据效力：** AI生成内容需人工复核
4. ⚠ **可解释性：** 黑箱用于线索发现，不用于定性

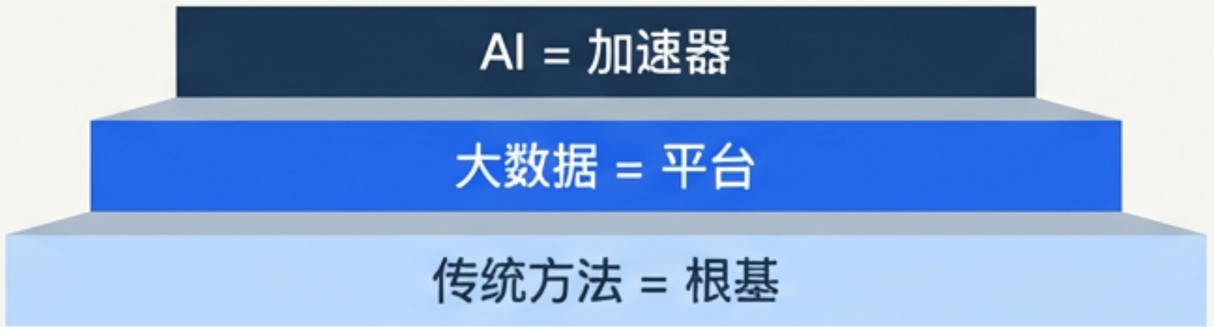


智能审计平台建设四阶段



从「审计工匠」到「审计架构师」





审计数字化转型，从认知开始