

数据价值挖掘——群体信用违约预测模型

要解决的问题

识别股权、交易、担保网络的系统性风险
规避循环担保、过度授信
量化企业违约风险



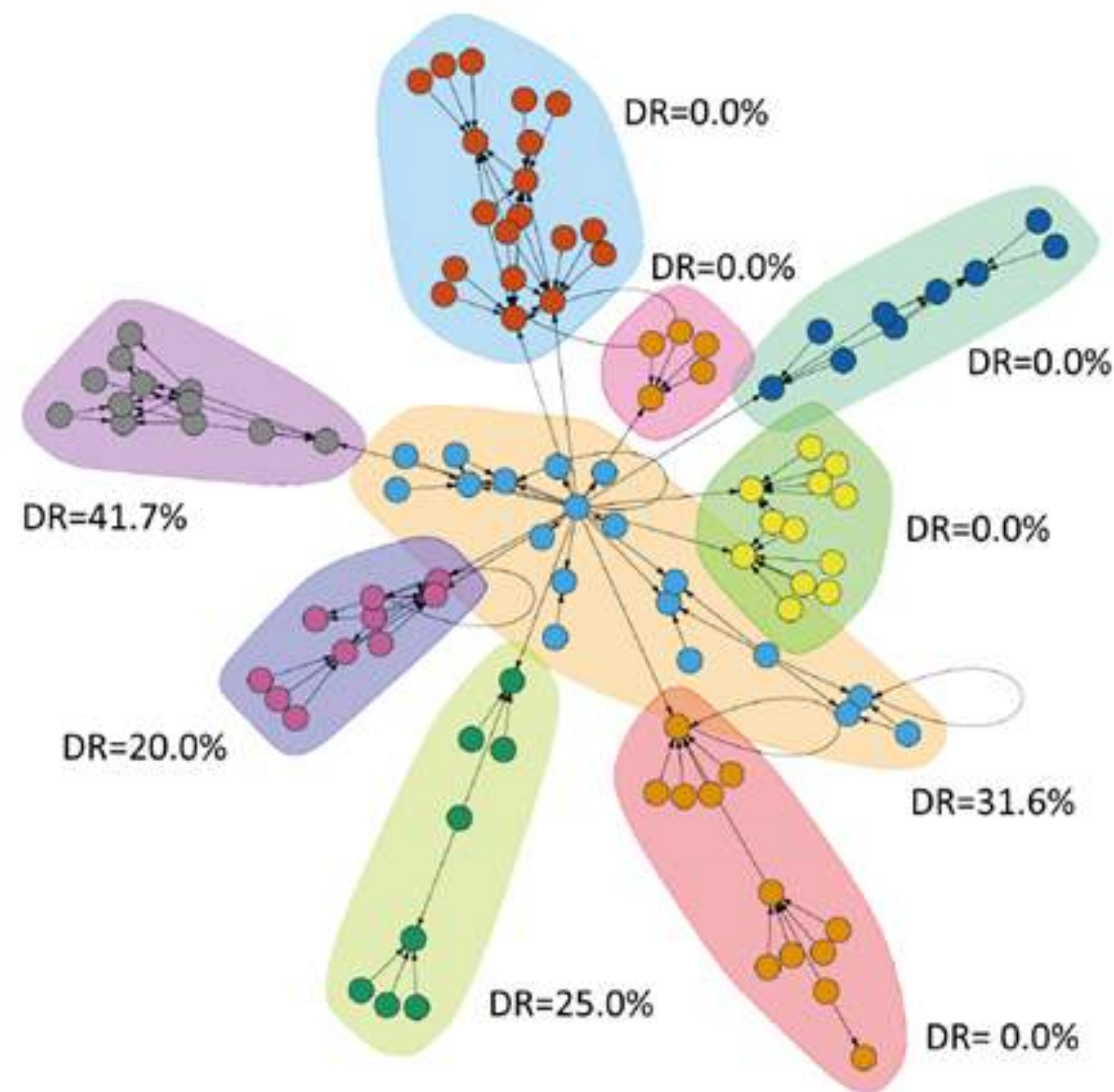
技术实现

客户属性、信用行为、关系图谱、社区特征
复杂网络技术挖掘违约风险影响分子
关系链图特征、客户行为特征建模



模型输出

关系网络可视化风险监控
高风险预警客户名单输出
授信评审阻断策略因子



交易反欺诈系统

对全渠道的客户交易行为进行全方位监控，最大程度避免客户的经济损失。

覆盖全
电子渠
道

实时
侦测

灵活的
规则配
置

一站式
管理平
台

风险事
件持续
跟踪

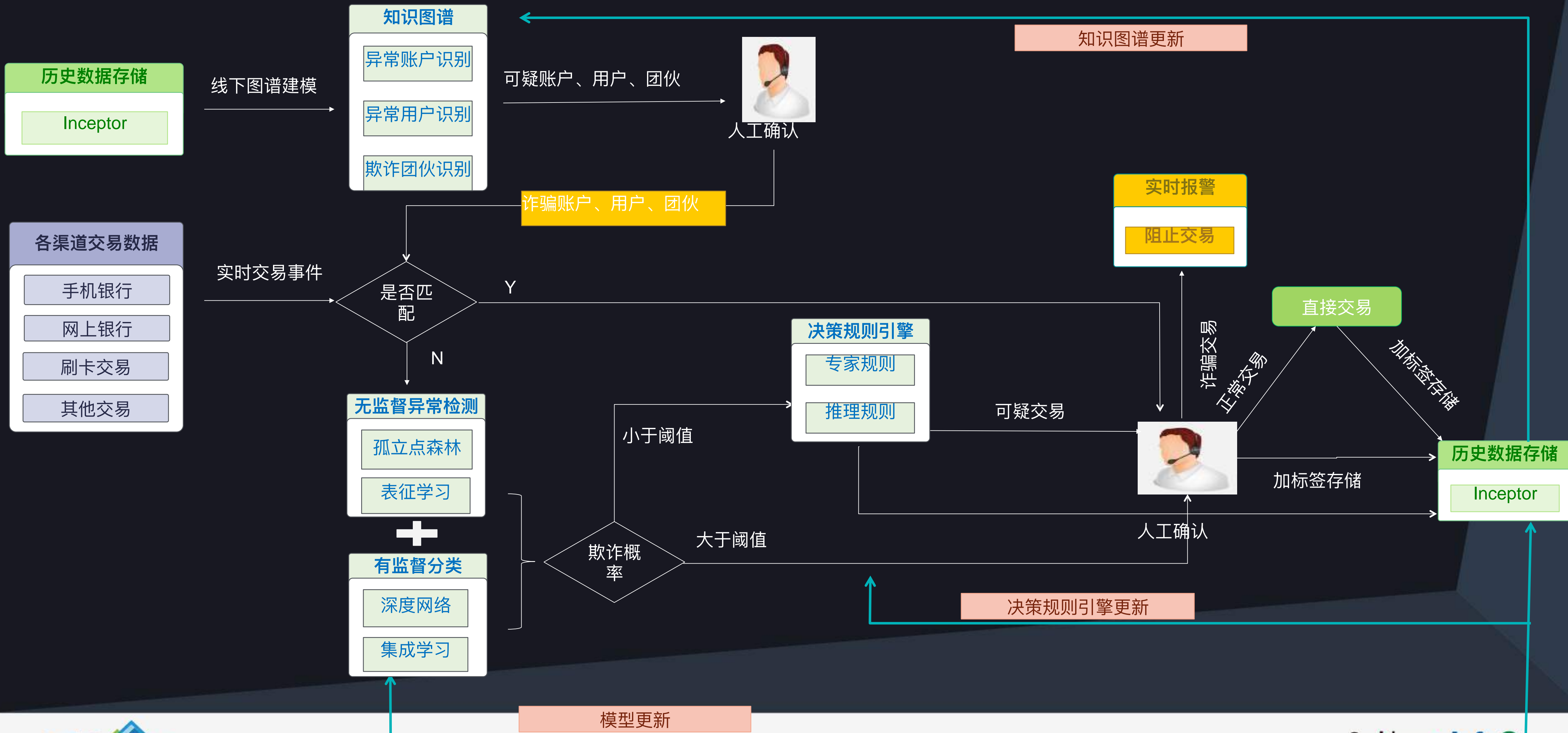
实时采集客户交易数据，支持实时风险侦测，对于欺诈交易实施事中阻断，对于可疑事件进行事后提醒。

对可疑事件进行持续跟踪，建立风险等级评定体系。

建立风险特征模型库，支持灵活且快速部署新的业务规则，能及时有效地防范各类风险。

对接客服系统和短信平台，系统内实现一键触达客户；且支持报表统计。

数据价值挖掘——交易反欺诈应用流图



实施过程的困难与挑战

外部数据成本过高

缺乏高度整合的公开市场数据

政府、国营事业单位数据时效性差

业务团队风险理念转变

高层强力支持

风控流程和操作管理办法的适应调整

智能技术成熟应用需要时间

学术成果的转换成本、试错成本

数据科学人才缺乏、培养需要时间

行业知识图谱构建缺乏业务专家

外部信息源、产业标准化信息

实体逻辑关系、风控专家规则



目录

- 1 平台建设背景
- 2 数仓应用体系建设
- 3 风控领域创新应用
- 4 取得成果与未来展望

原数仓

大数据平台

VS

 数据量规模 6T

 数据量规模 400多T

 模型处理个数 1500

 模型处理个数 2300

 处理效率 13个小时

 处理效率 2个小时

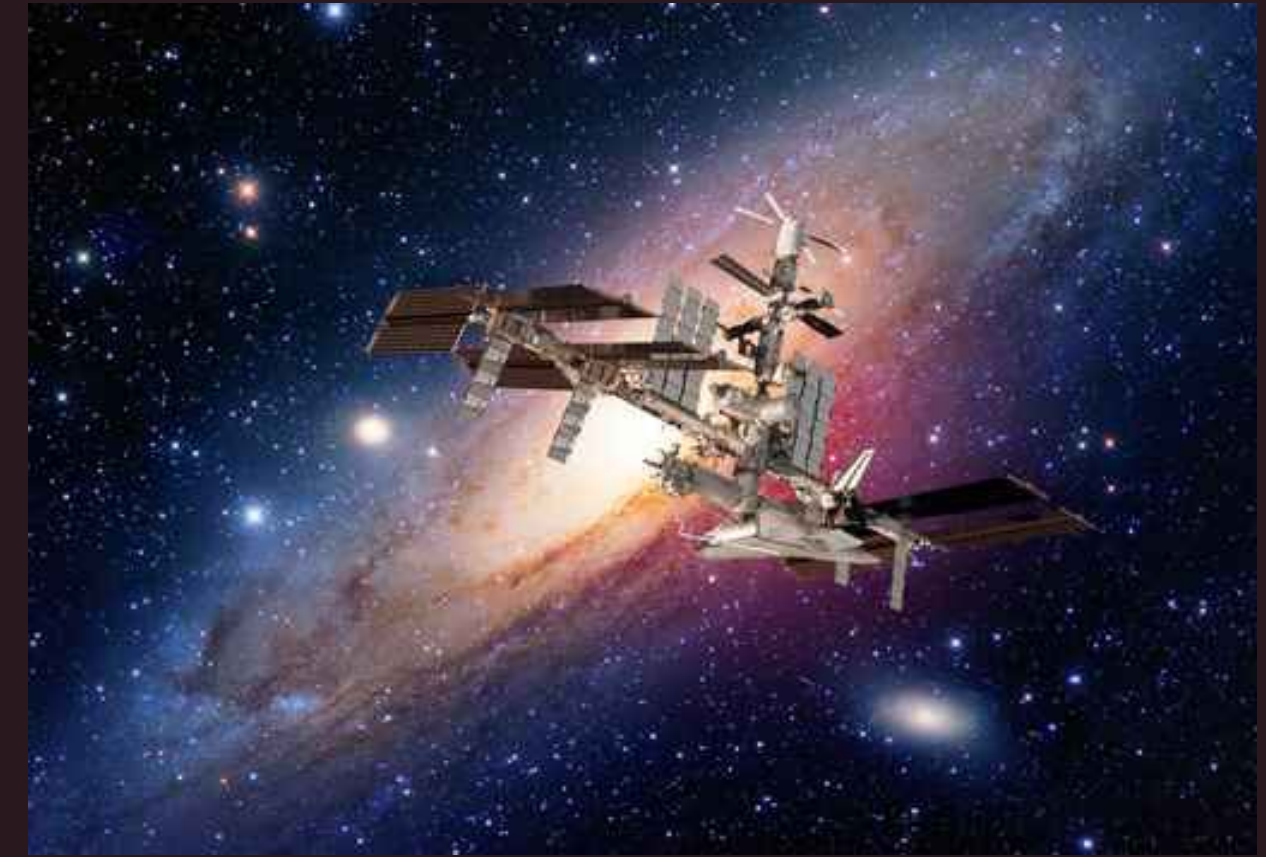
单个模型效率对比：

□ 处理效率是指，从接入核心系统数据算起到模型数据处理完成的日终时间。

以存款账户表为例：

- 数据处理逻辑相近，数据量一致
- 原数仓存款模型 --平均耗时：3小时30分钟
- 大数据存款平台模型 --平均耗时：39分钟

取得成果-提升全行数字化支撑能力



实时的大数据平台能力

- 01.13个部门集市、18个分行集市、26个应用集市管理380TB数据，日实时处理200万交易数据
- 02.对外发布110个服务接口，月均调用近200万次
- 03.对接26个外部数据源，月采集企业舆情80万条每天聚合8大行业资讯、23类市场指数、200多份投研报告

创新应用助力业务发展

- 01.32个创新应用，发布1100个业务功能
- 02.全行2500个用户，月均使用6万余次
- 03.335张业务可视化报表，每次支撑6000次统计分析
- 04.月推送实时业务提醒60万条，累积推荐潜在客户5万户

全天候的风险监测体系

- 01.实时跟踪30万行业标杆客户
- 02.300多个信用风险预警规则，年触发风险预警信号8000次
- 03.各类平台贷风险服务接口23个，月均调用5万余次
- 04.196个运营风险监测模型，月均生成工单4500笔

工作展望

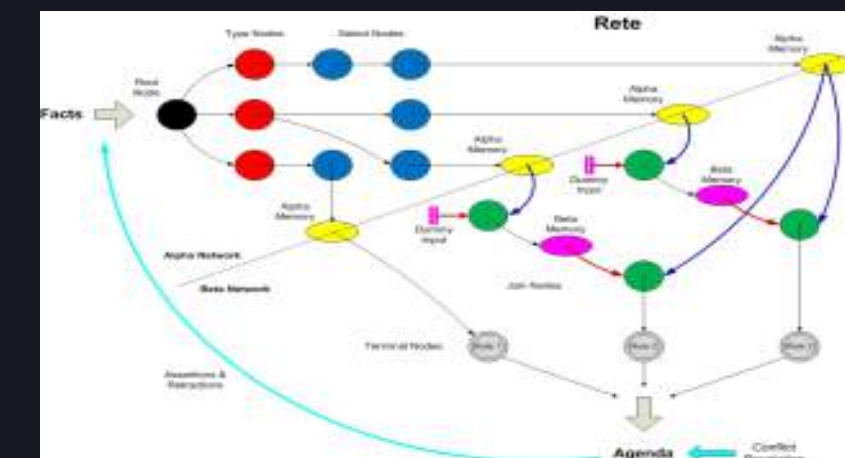
1

基于Docker 容器技术，建设面向多租户的大数据平台，实现更细粒度的资源管控与调度



2

研发面向业务团队、可定制的实时智能业务决策引擎，满足实时营销、实时风险管理、实时交易反欺诈等多种场景需求。



3

融合行内外多样化数据，深度提炼数据价值，优化业务领域模型，为业务一线和客户提供更多场景的智能化数据服务



4

建立行业知识图谱技术平台，满足客户价值评估，行业风险传导路径，重大事件影响分析等业务需求



THANK YOU

如有需求，欢迎至 [讲师交流会议室] 与我们的讲师进一步交流

