

信贷业务持续创新中的 大数据风控架构

张军

宜信大数据创新中心



主办方 **Geekbang** 极客邦科技 **InfoQ**

AiCon

全球人工智能与机器学习技术大会

助力人工智能落地

2018.1.13 - 1.14 北京国际会议中心



扫描关注大会官网

SPEAKER INTRODUCE



张军

大数据创新中心 工程总监

负责宜信大数据风控平台

张军2005年硕士毕业于哈尔滨工业大学计算机科学与技术专业。之后加入富士通研发中心，先后从事自然语言处理、云计算等技术的研发；带领团队开发了舆情监控，企业服务融合，业务决策管理等等多项在实际业务上产生重大影响的系统。2014年张军加入百度，参与了百度核心的搜索排序系统由规则系统向深度学习系统转型的研发过程，上线了大幅提升长尾Query的搜索效果的RNN排序模型；之后，张军带领团队研发了百度内部的深度学习平台Lego，该平台应用在百度搜索体系内的搜索排序，信息流，点击预估，机器翻译等等多项核心业务当中。同时，张军还拥有三十余件发明专利。

TABLE OF

CONTENTS 大纲

- 持续的信贷业务创新对风控技术带来的挑战
- 存量信贷数据的数据整合和使用
- 风险政策与模型的快速迭代：决策引擎
- AI在风控当中的应用
- 案例分享：商通贷

关于宜信

2006年宜信公司成立于北京，今天的宜信已经成为中国的**金融科技**领军企业，业务涵盖**普惠金融、财富管理、资产管理**三大板块，年交易规模超过**1000亿**

创始人 唐宁

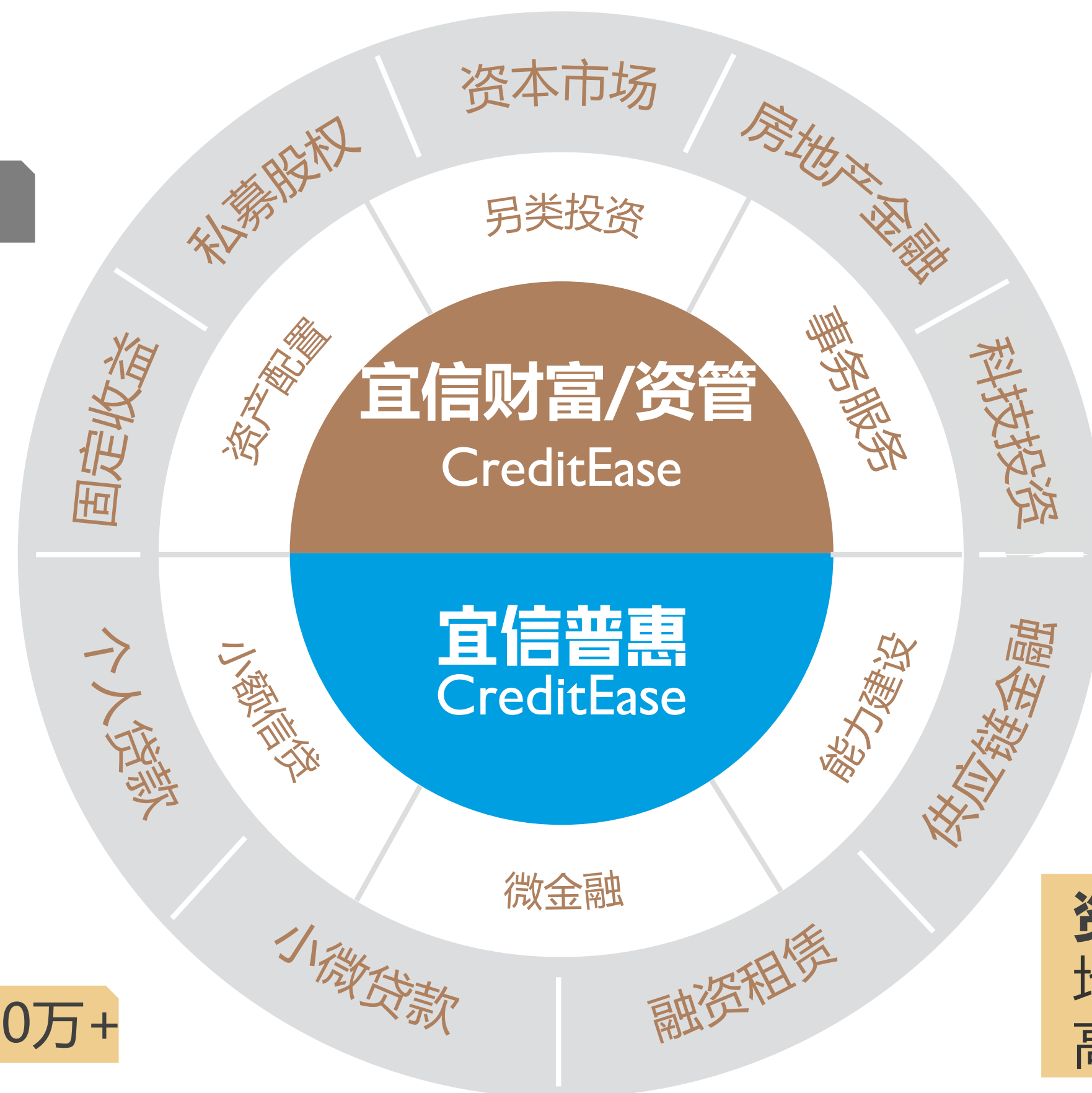
创立时间 2006

总部所在地 北京

经营行业 - 金融服务、普惠金融、财富管理、资产管理、金融科技

员工人数 45,000人

借款人客户数量 200万+



财富管理、资产管理
10万+高净值客户及机构客户

版图覆盖 全国 40余个城市
香港、新加坡、美国、英国、
澳大利等地设立团队及办公室

年交易规模 超过1000亿

资产配置 覆盖私募股权、资本市场、房
地产、保险保障、税务规划、移民置业、
高端游学等全方位产品和服务

持续创新的信贷业务

- 信用借款：基础信贷、车主贷、房主贷、商易贷、宜学贷
- 抵押借款：车抵贷、宜房贷
- 电商融资：商通贷
- 供应链金融：商超贷、翼启融
- 线上信贷（宜人贷）：极速模式、精英模式、寿险模式
- 其他孵化当中的创新的信贷产品

风险存在于信贷业务的各个环节

业务

城市信贷、宜人贷、商通贷、车贷、... ..

环节

营销

信用审核

反欺诈

贷后管理

合规检测

技术

数据准备

知识图谱

决策平台

机器学习

用户画像

语义分析

数据

业务数据

公开数据

用户提交数据

大数据风控的风险决策开发典型流程



S	数据产品	分析平台	决策引擎	业务流程
I	数据接口 决策场景	数据、算法、实验	新的模型 决策规则	决策流程的监测
P	<ul style="list-style-type: none"> 数据对接 数据价值评估 设计业务线实验 	<ul style="list-style-type: none"> 数据抽样 模型开发和决策设计 实验测试和实验分析 	<ul style="list-style-type: none"> 风险决策的开发 新决策流程的测试 生产环境发布 业务线确认 	<ul style="list-style-type: none"> 定期评估数据价值 数据退化、淘汰、备案 数据分析及决策迭代 数据安全及合规管理
O	评估报告	评估报告、应用原型	决策流程	监测报告、改进建议
C	分析平台	研发团队	业务线产品和风控	决策者（管理者）

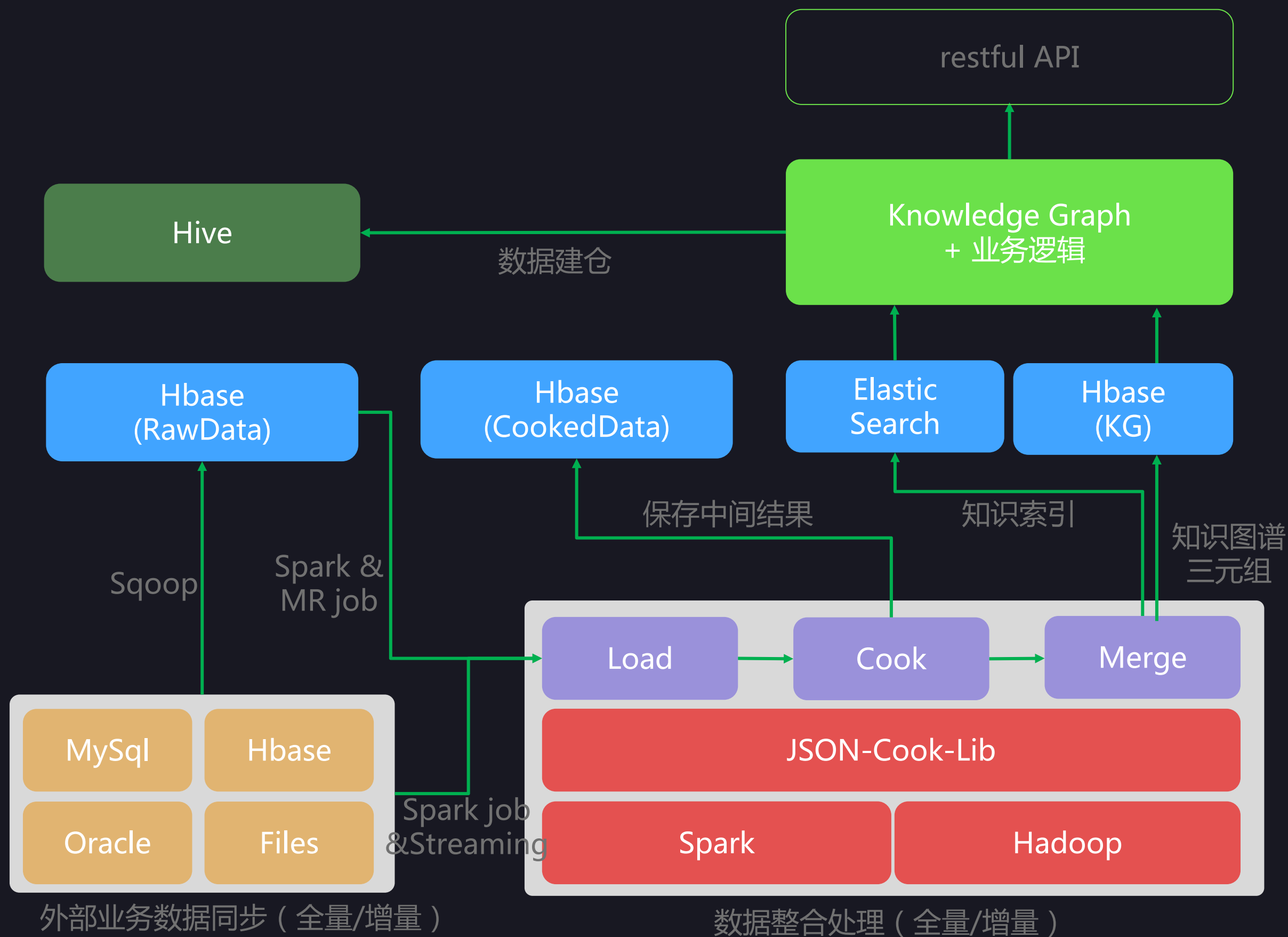
持续的信贷业务创新对风控技术带来的挑战

- 产品不同阶段对风控的需求
- 风控的类型：信用风险、欺诈风险、操作风险
- 存量数据的有效利用：实时数据服务和离线数据分析
- 风险政策和模型的全生命周期管理
- 用AI来赋能

数据的管理、治理、和服务

- 需求：离线的数据分析；实时性的风险政策
- 高实时性的数据流转方案
- 数据仓库体系
- 外部数据源的管理
- 数据的有效组织的知识图谱

数据服务：实时整合各环节数据，并提供消费



1. Data Cook工具，对异构数据的整理、清洗、合并
2. 全量字段mapping管理
3. schema统一视图
4. Azkaban任务调度
5. spark实时流式处理做到准实时数据同步
6. Restful的数据服务API供下游系统使用



DBus 实时数据总线

DBus (寓意数据管道)

专注于数据的收集及实时流计算，以无侵入方式对源端数据进行采集，采用高可用流式计算框架，对数据进行汇聚，转换处理。以统一数据格式 (UMS) 提供给不同数据使用方订阅和消费，充当数仓平台、大数据分析平台、实时报表和实时营销等业务的数据源

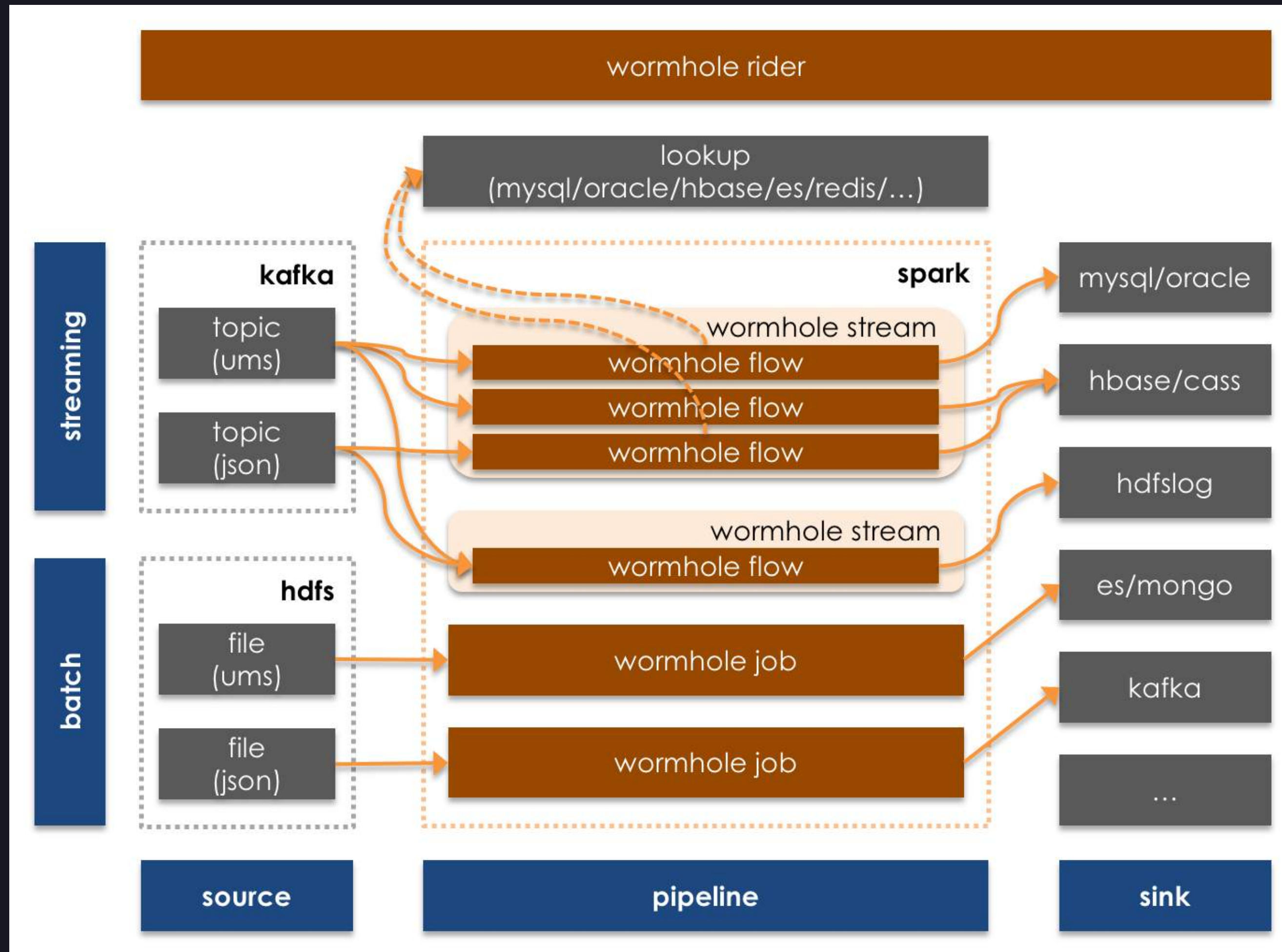
主要特性：

- 无侵入方式接入多种数据源
- 海量数据实时传输
- 感知源端schema变更
- 数据实时脱敏
- 高效的初始化加载和重新加载
- 统一标准化消息传输协议
- 可靠多路消息订阅分发
- 可视化后台管理系统
- 支持分表数据汇集

• <https://github.com/BriData/DBus>



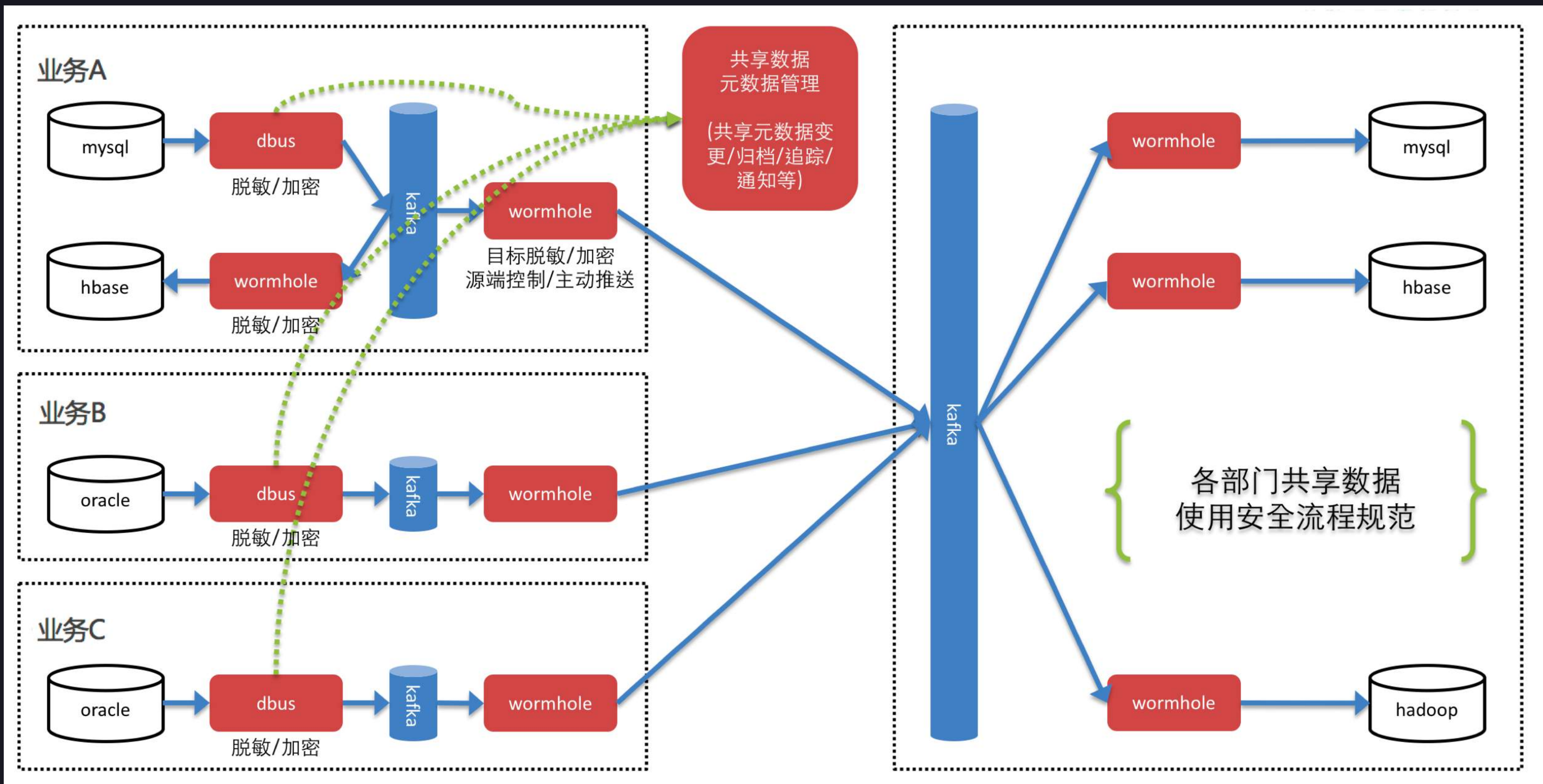
Wormhole 流式处理平台



- Wormhole 流式处理平台
- Wormhole是一个SPAAS (Stream Processing as a Service) 平台解决方案。Wormhole面向大数据项目开发和管理人员，致力于简化和统一开发管理流程。尤其针对流式实时和准实时数据处理场景，Wormhole提供了可视化的操作界面，极简的配置流程，基于SQL的业务开发方式，并屏蔽了大数据处理底层技术细节，极大的降低了流式处理项目开发管理运维门槛，使得大数据项目开发和管理变得更加轻量敏捷可控可靠。
- 开源地址
- <https://github.com/edp963/wormhole>



实时数据整合示意图



数据服务：实时构建知识图谱

• 主要应用

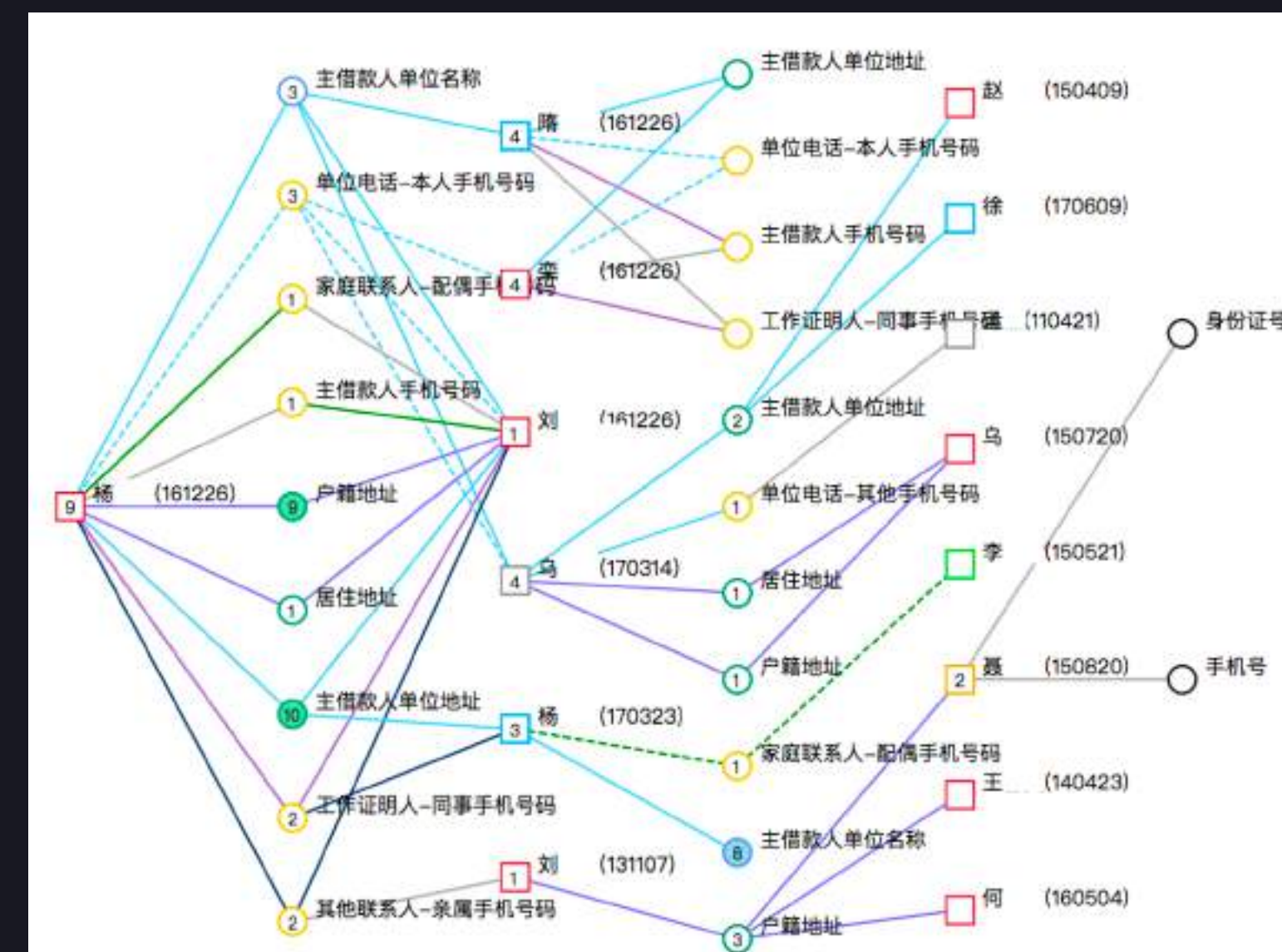
- 使用申请id、电话、设备、地址等等信息组成一个复杂的网络，不仅用在**授信建模**上，更重要的是能够**从关系的角度用数据挖掘方法识别出手段隐蔽的欺诈案件。**

• 知识图谱构建

- 身份id、电话号码类信息进行标准化处理
- 地址类信息使用深度学习方法进行匹配（例：光华路4号东方梅地亚中心C座20层、光华路东方梅地亚C座20楼是同一个地址）
- 大量非结构化的文本类数据，使用序列标注等算法进行关系挖掘

• 知识图谱的在线应用

- 依赖ElasticSearch、HBase及图数据库来架构出高可用、高并发的在线应用



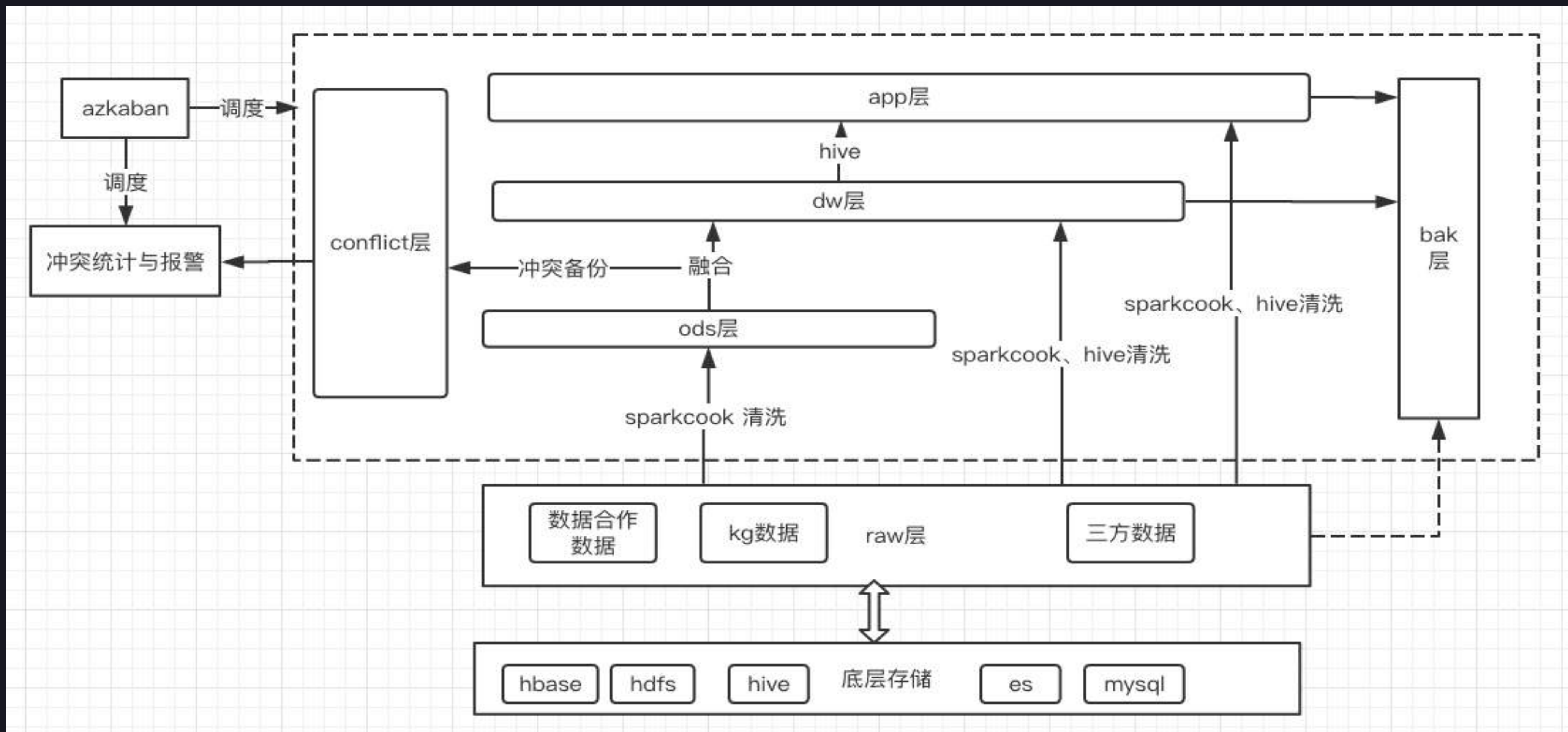
知识图谱目前的积累：

7亿实体

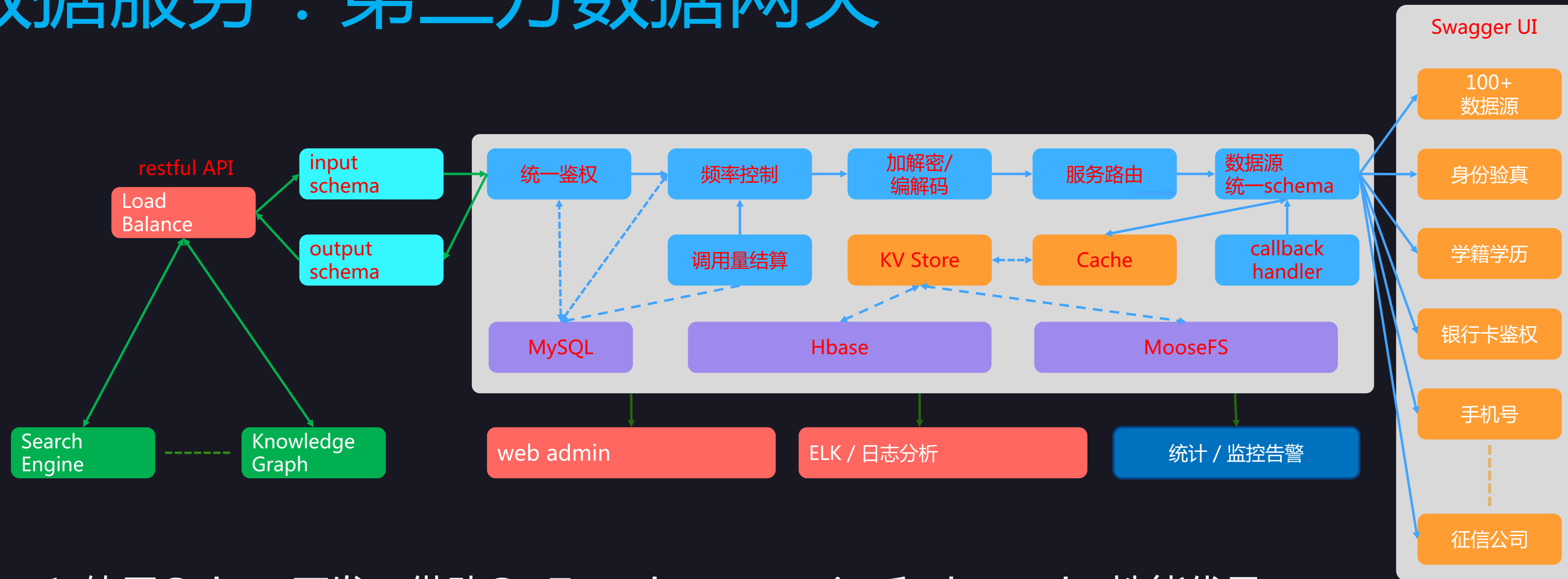
45亿事实

0.4亿风险名单

数据仓库：建立体系以支持离线分析



数据服务：第三方数据网关



1. 使用Golang开发，借助Go Template、routine和channel，性能优异；
2. 极大简化异构 / 多维数据源的接入；
3. 统一鉴权、频率控制及调用量结算；
4. 全数据源维度的数据分析，实时监控告警。

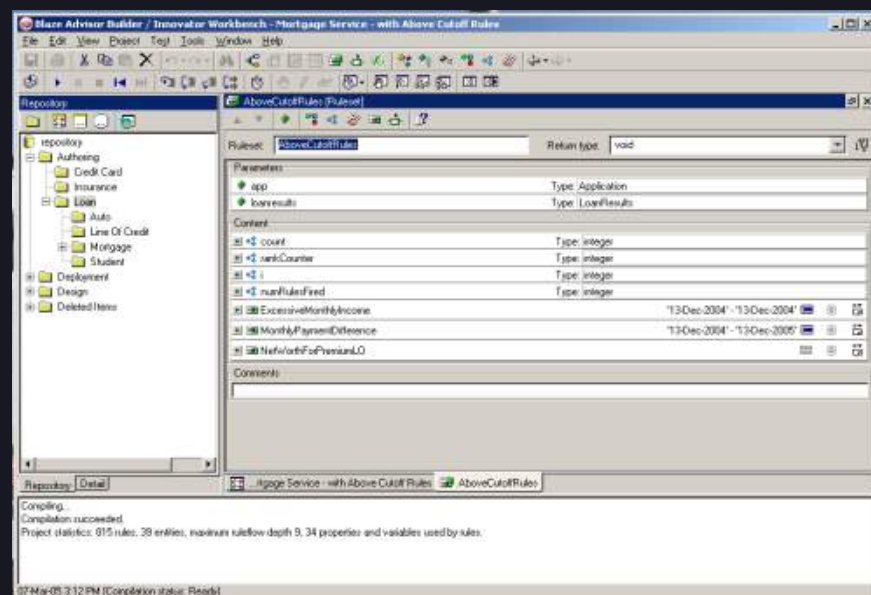
风险决策平台——谛听研发的意义和背景

- 信贷产品中的典型角色
 - 技术开发：造车修车
 - 风险管控：控制车速
 - 产品运营：把车开快
- 决策的依据：模型+规则



决策平台在宜信的演进

FICO Blaze



商业

Drools



开源

决策平台-谛听



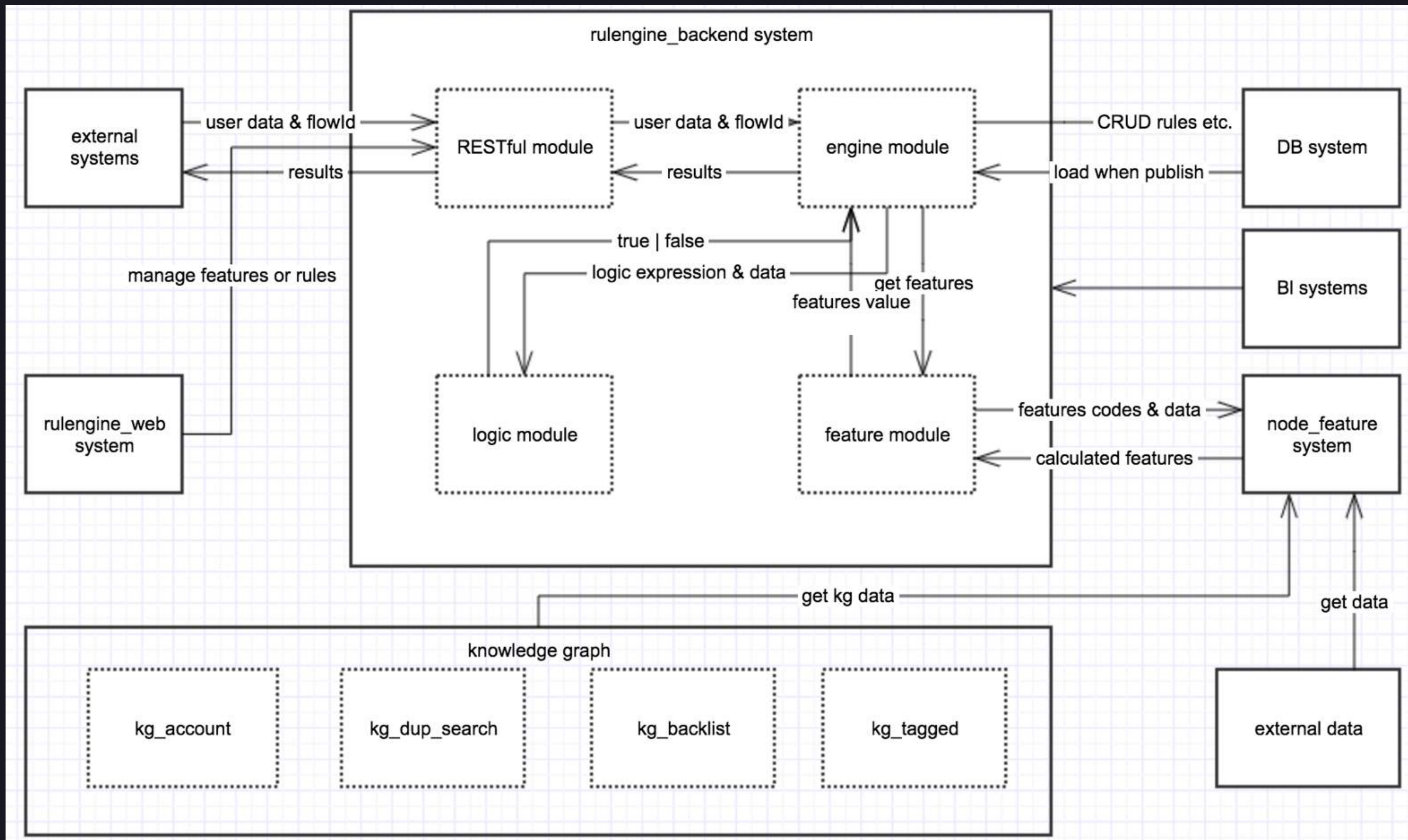
自研

风险决策平台——谛听

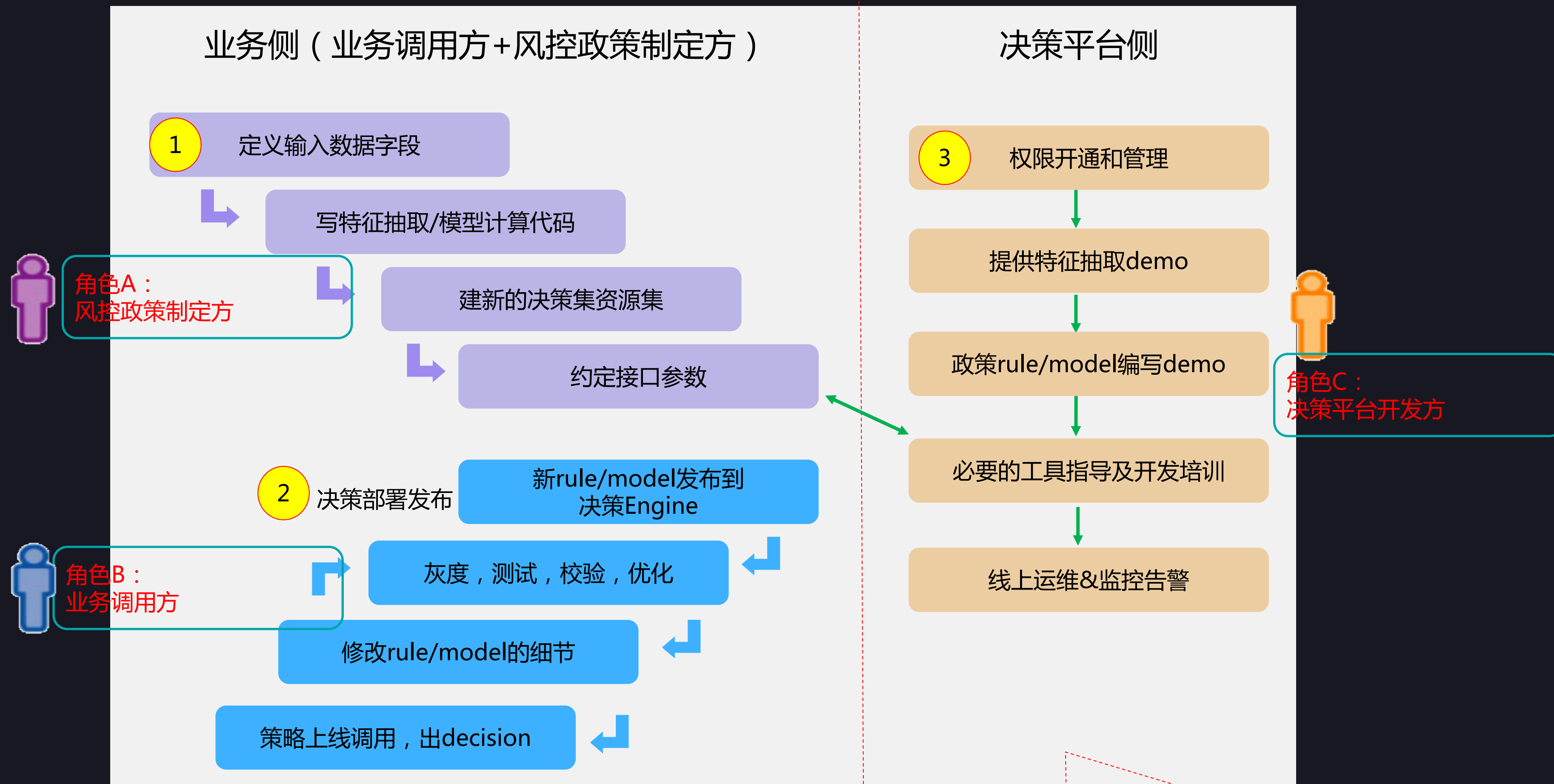
风险政策+风险模型的全生命周期管理



风险决策平台架构



风险决策实施步骤



原则上, 决策平台侧只需关注此类事情

风险决策平台同时作为风险政策的管理平台



- 制定风险策略是**金融业务的核心竞争力**，直接决定了风险表现；
- 决策平台同时把风险模型和规则作为**企业的核心资产**集中管理。

风险决策平台解决了的问题

- 模型的快速部署和应对
 - 风险政策：直接在UI上操作即可上线
 - 复杂模型：支持用js或者python进行计算，并快速上线
 - 对于业务系统透明，无须关注后面的逻辑实现和调整
- 风险政策的集中式管理
 - 风险政策是核心资产：权限管理+使用审计；
- 模型生命周期的管理
 - 数据分析&模型训练→测试验证→灰度上线→部署上线→持续监控