

BDTC 2017 中国大数据技术大会
Big Data Technology Conference 2017

小微企业信贷的大数据风控模型 和架构实践

张京磊

宜信大数据创新中心 高级研发经理



个人介绍

1. 宜信大数据创新中心 高级研发经理
2. 负责多个数据驱动型实时信贷业务的架构设计和技术开发工作，曾负责大数据风控搜索引擎和决策引擎的技术工作；
3. 之前任职于腾讯移动互联网事业群，先后负责QQ音乐基础服务、腾讯文学推荐服务、高性能海量服务框架的架构和开发；
4. 曾2次获得腾讯S级员工，所在团队曾获得腾讯卓越研发银奖；拥有一项大数据方向的发明专利，2014年获评腾讯A级专利；
5. 张京磊毕业于中科院研究生院，获硕士学位。

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**

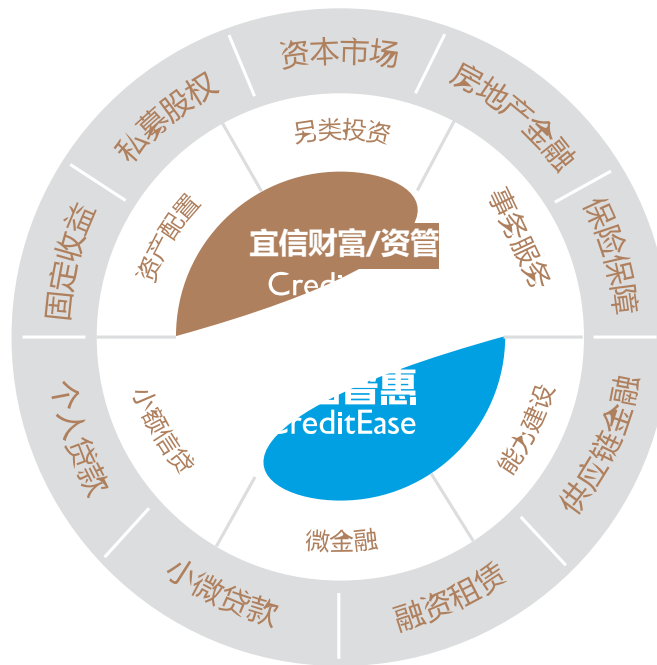
创始人：唐宁

创立时间：2006年

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

澳大利亚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、移民职业、高端游学等全方位产品和服务

目录

CONTENTS

- 01、小微企业现状和信贷挑战
- 02、基于大数据+AI的实时智能决策平台
- 03、风控数据和模型的应用
- 04、持续集成和发布的PaaS云平台
- 05、总结回顾

01 小微企业信贷现状和挑战

中国小微企业现状：“三高两难”

原材料成本高

人力成本高

运营成本高

找生意难

融资难

资源依赖程度
高，受外部环
境波动影响大

利润率低，税
负重，行业平
均亏损率高

老板个人
与公司
边界模糊

管理规范化程
度低

金融意识和数
据价值意识有
所提升

管理信息化、
移动互联网化
程度提高

寻求整合性、
体验好的信息、
交易与金融服
务

借助行业服务
平台获得信息
和资源

业务背景



支持主流内外贸电商平台授信

支持跨平台授信

已累积海量线上数据征信的能力和 data 储存，商通贷对小微企业有足够深刻的认知和理解

面向小微企业的全方位服务体系

—— 专为小微企业解决融资难题 ——

0.72% 超低综合月成本	1 分钟 预估额度时间	2 分钟 最快申请时间
8 分钟 最快放款时间	24 个月 期限畅享	200万 最高放款额

纯信用贷款 0 抵押

关于商通贷

宜信商通贷，定位于专为小微企业解决融资难题的实时授信金融服务平台。基于Fintech价值理念、宜信大数据金融云平台，借款用户仅需在线提交个人有效身份信息与实际经营数据（网店经营、ERP管理、线上财务管理）即可进行实时授信，高效快捷的为用户解决资金周转问题。

宜信商通贷已将小微企业融资业务，从注册申请、信用审核、实时放款、用户还款等各个环节全部网络化操作，并利用大数据建立小微企业信用体系，将零散数据结构化后高效转化为个体信用，多维度、多视角地对小微企业进行风控综合评定。



挑战1：小微企业经营和数据特点



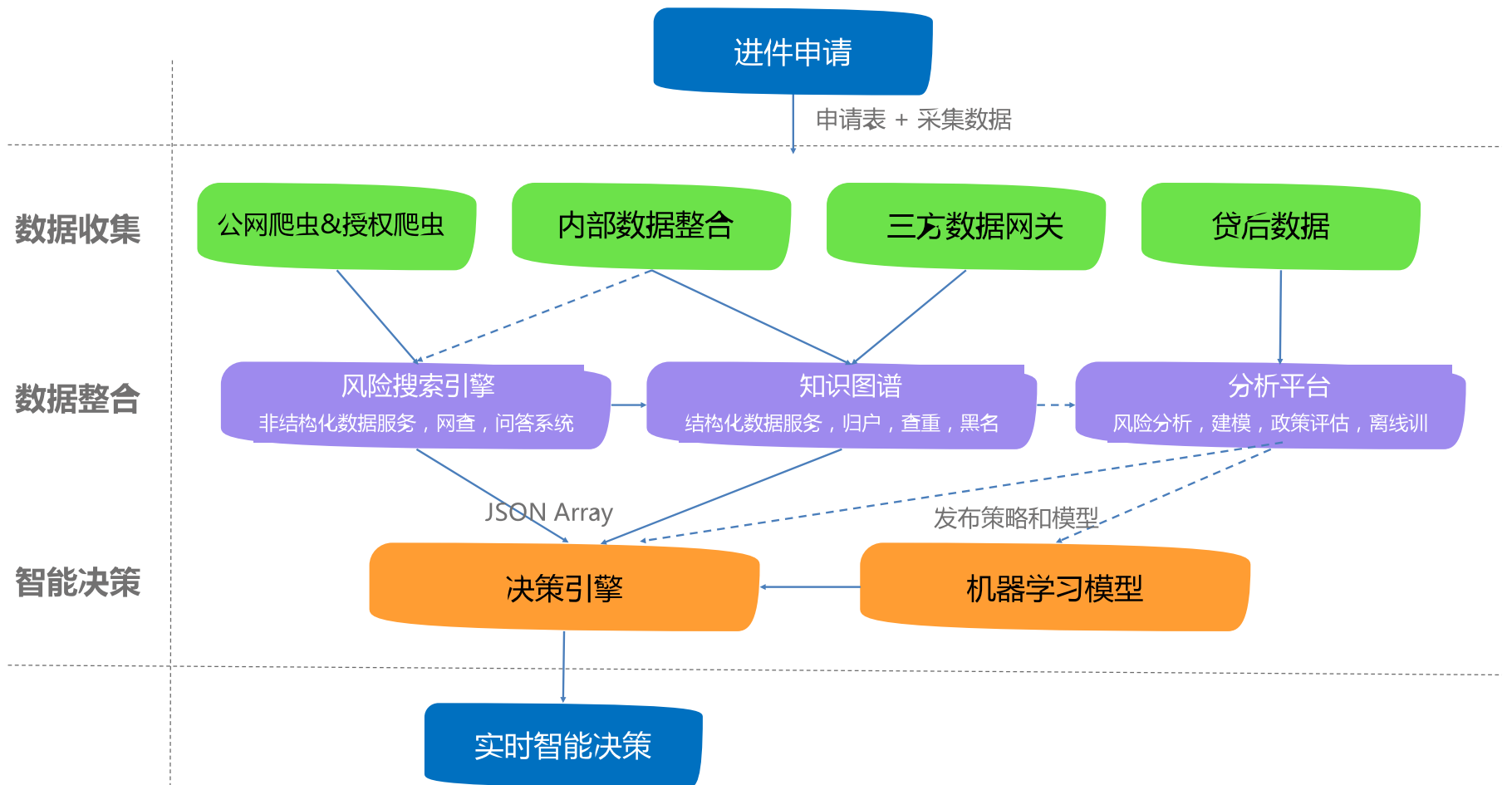
挑战2：持续业务创新带来的压力

- 产品不同发展阶段，对技术的需求
- 风险的类型：信用风险、欺诈风险、操作风险
- 存量数据的有效利用
- 用AI来赋能
- 数据和技术的平台化（数据服务、技术服务）

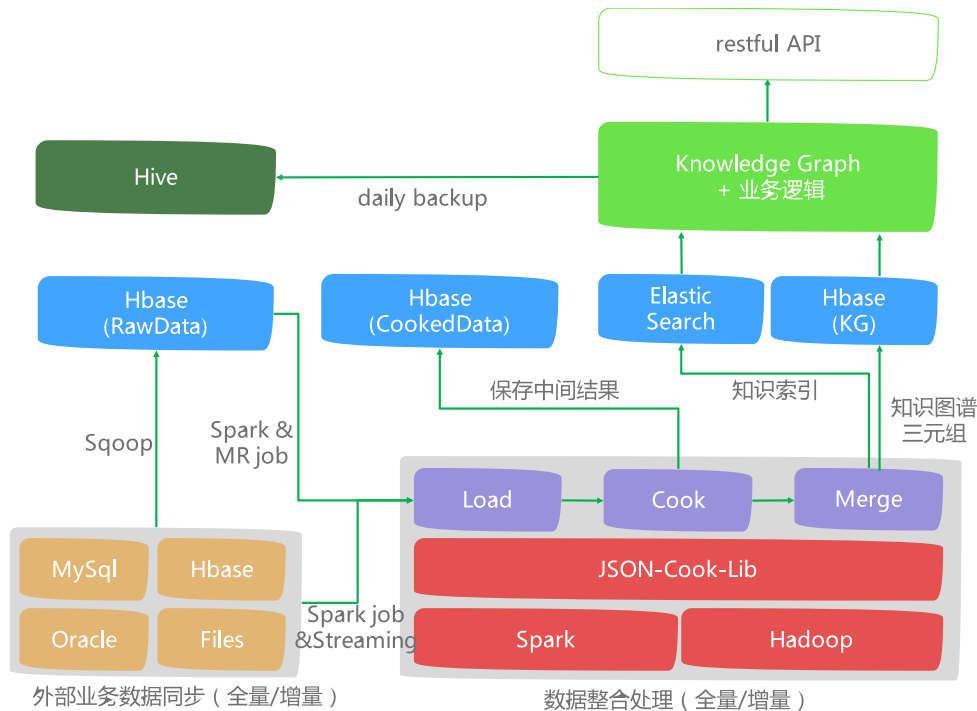
02 基于大数据+AI的实时决策平台



1. 数据是基础，领域知识、专家系统和智能决策是核心竞争力
2. 将数据加工，通过智能决策引擎实现：
 - 分析申请人的信用状况
 - 检测欺诈风险，还款意愿
 - 预估还款能力，实时给出授信额度
 - 其它大量模型的应用适配
3. 最终形成面向互联网用户的，可灵活定义的金融产品



数据服务：实时整合各环节数据，并提供消费



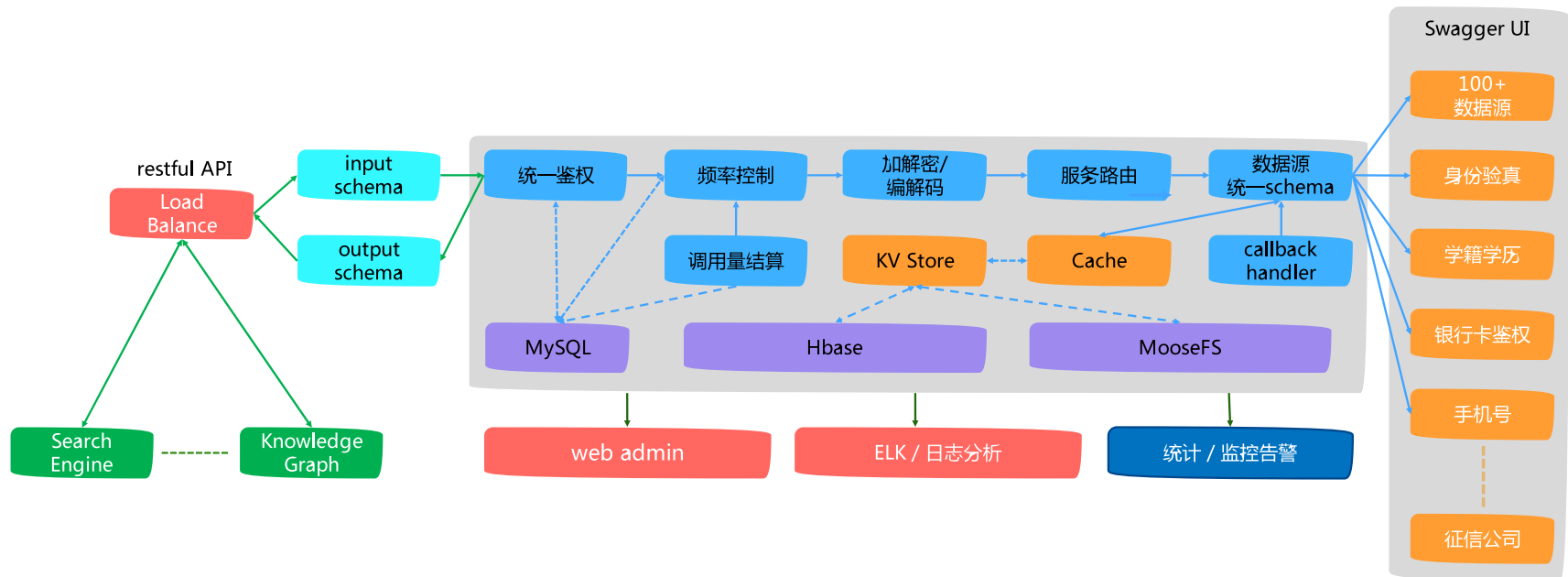
- 1、Data Cook工具，对异构数据的整理、清洗、合并
- 2、全量字段mapping管理
- 3、schema统一视图
- 4、Azkaban任务调度
- 5、spark实时流式处理做到准实时数据同步
- 6、Restful的数据服务API供下游系统使用

宜信的相关开源项目：

<https://github.com/BriData/DBus>

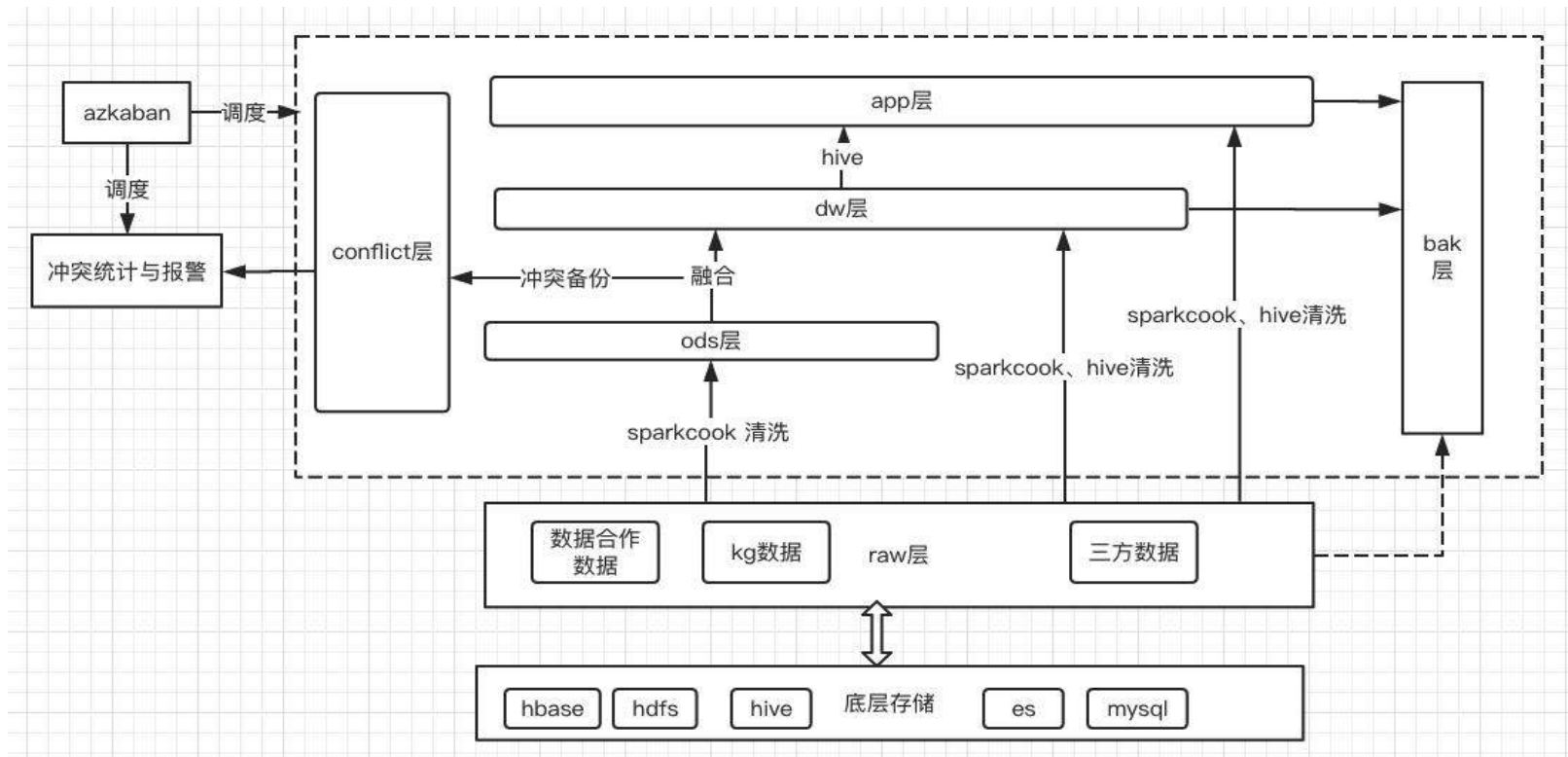
<https://github.com/edp963/wormhole>

数据服务：第三方数据网关

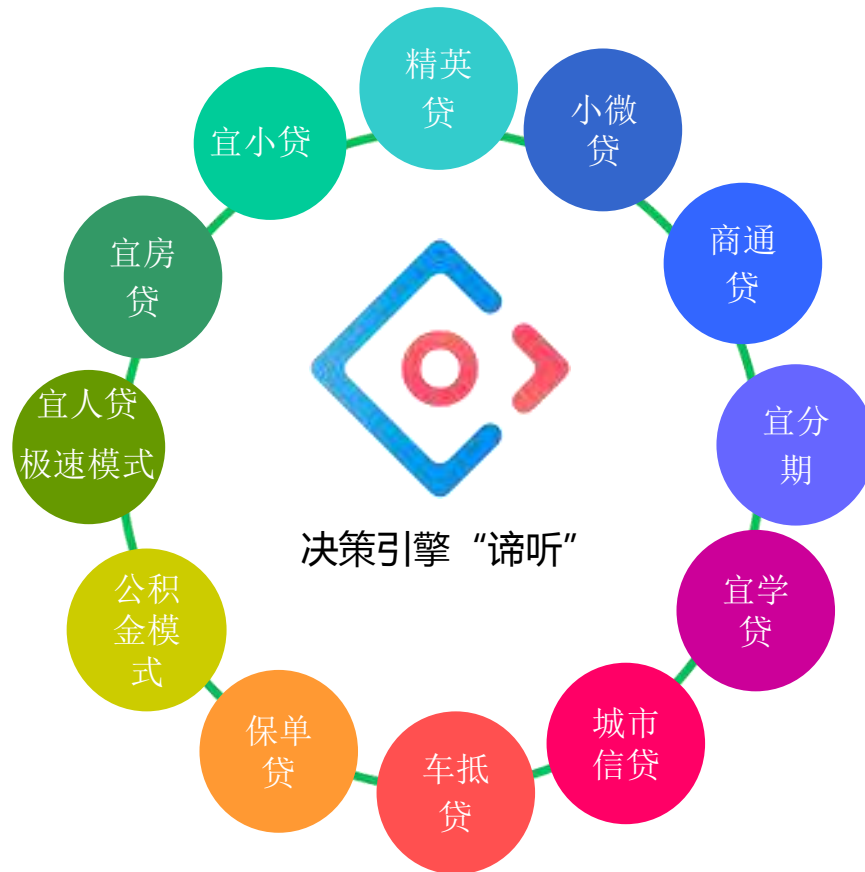


1. 使用Golang开发，借助Go Template、routine和channel，性能优异；
2. 极大简化异构 / 多维数据源的接入；
3. 统一鉴权、频率控制及调用量结算；
4. 全数据源维度的数据分析，实时监控告警。

数据仓库：建立体系以支持离线分析



风险决策引擎“谛听”：研发的意义和背景



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

决策引擎在宜信的演进

FICO Blaze



商业



Drools



开源



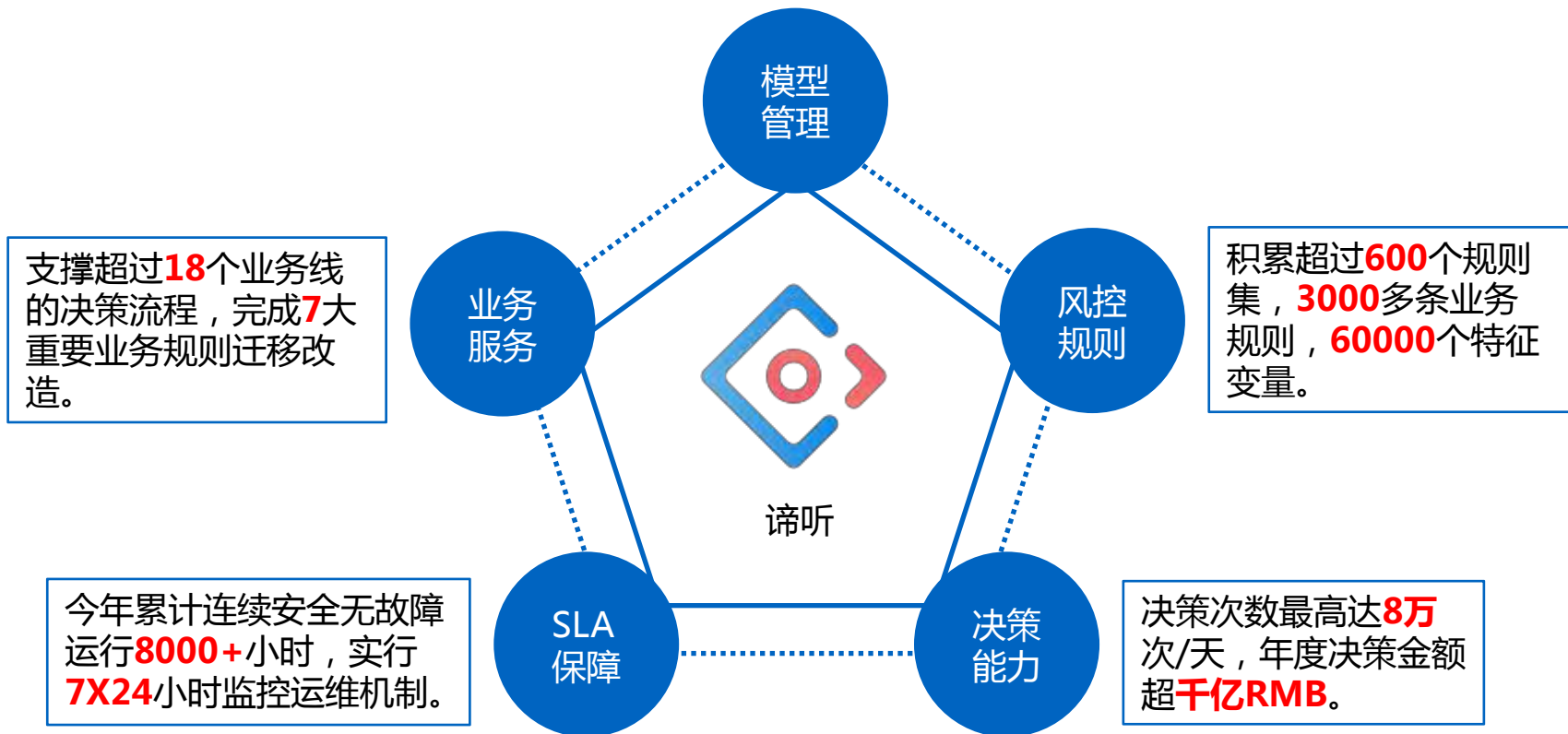
决策平台-谛听



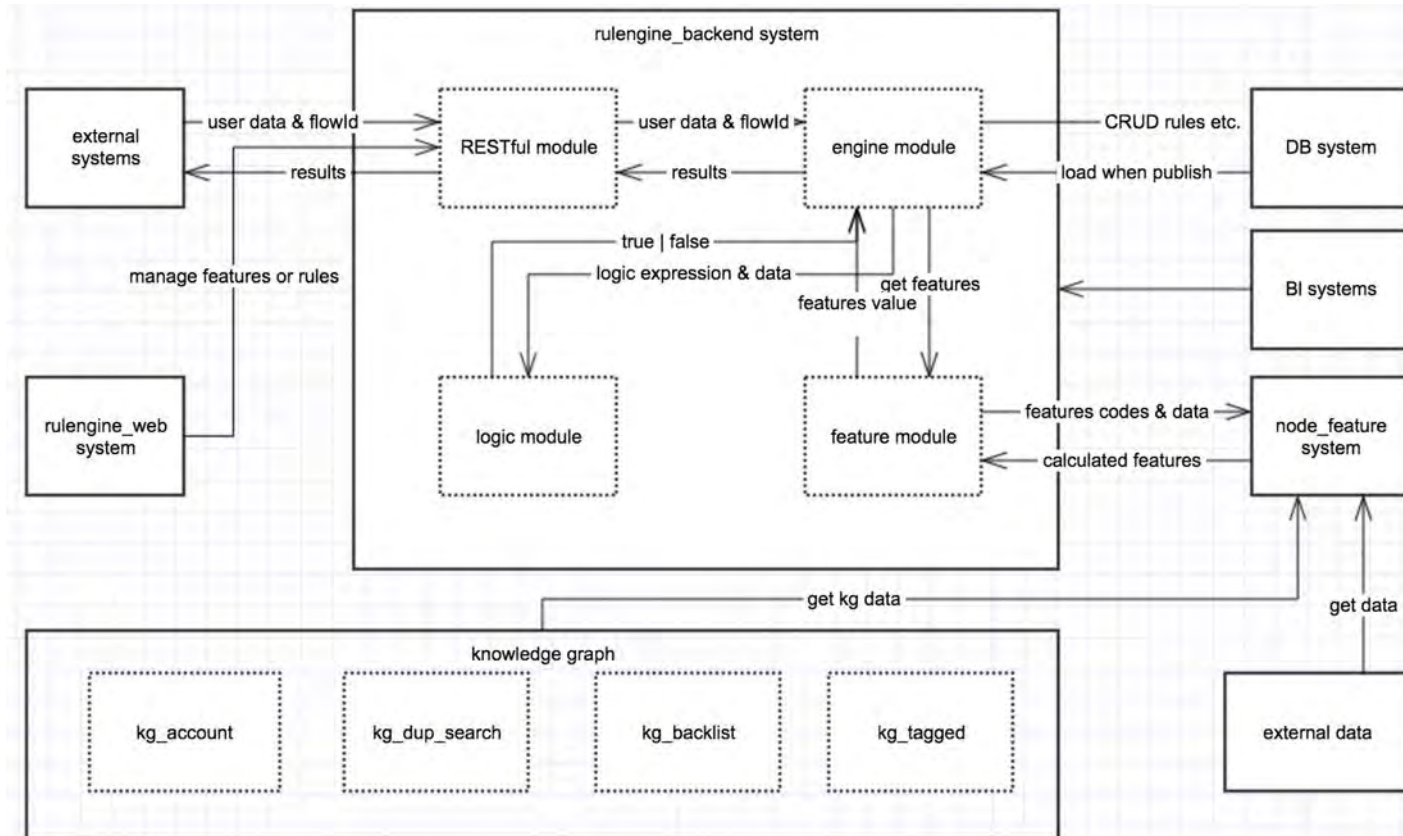
自研

决策引擎“谛听”的现状

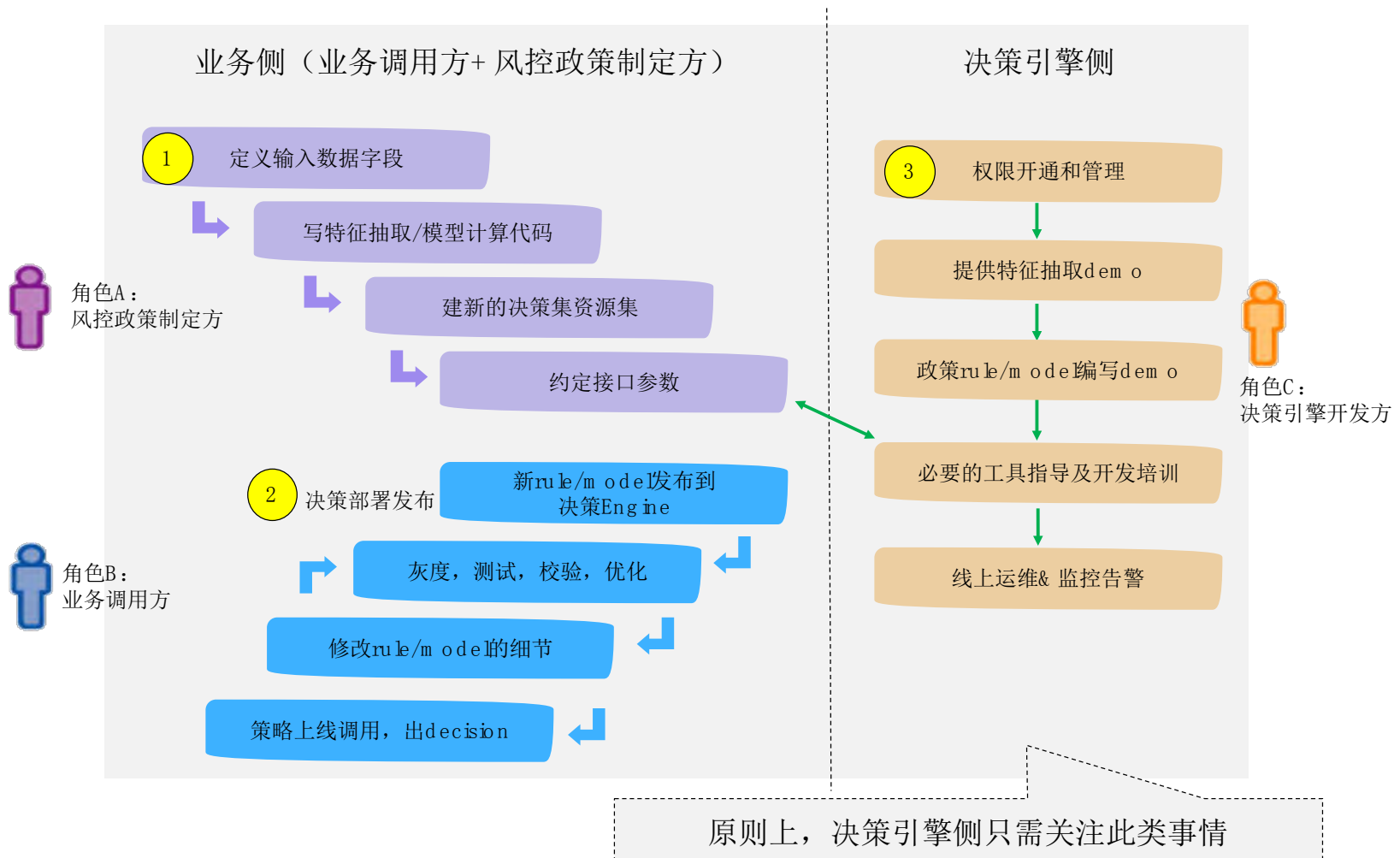
风险政策+风险模型的全生命周期管理



技术架构



风险决策实施步骤



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；
- BRMS将业务策略作为**企业的核心资产**集中管理。

风险决策引擎带来的收益

- 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

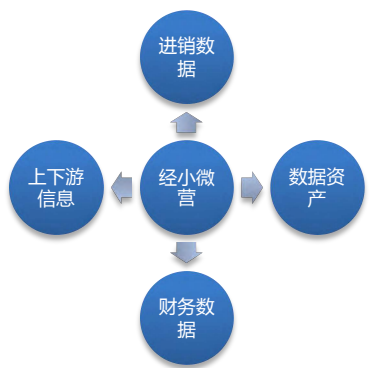
- 风险政策是核心资产：权限管理+使用审计；

- 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

03 风控数据和模型的应用

机器和专家经验的融合，数据的金融价值最大化



更加全面的经营数据



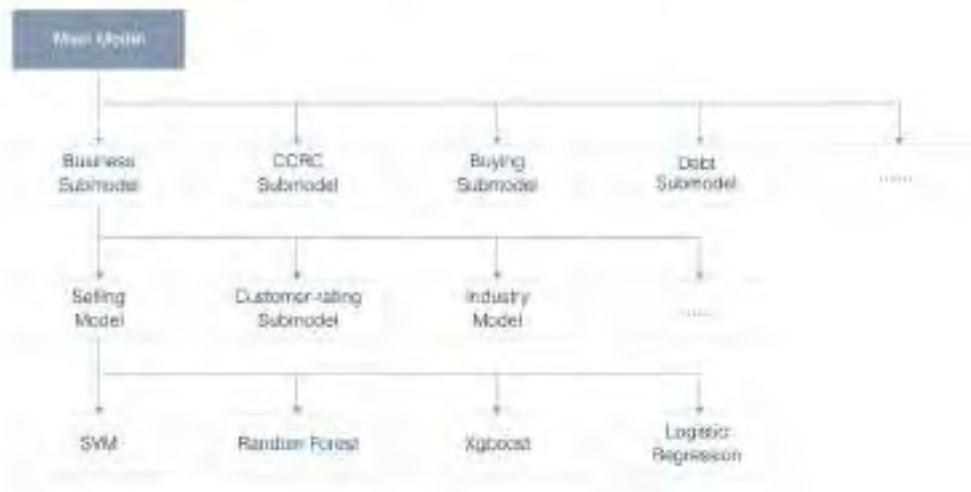
充分的校验模块



多数据源授信支持



树形模型融合



模型池：

充分挖掘电商数据

- 客户自身数据（经营能力、客户评价等）
- 行业趋势数据

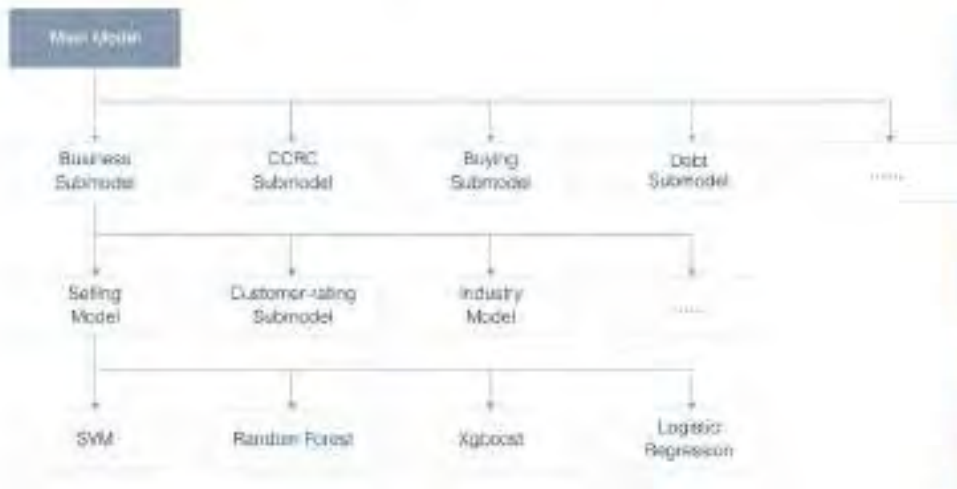
电商数据特征维度上千维

授权三方数据：对电商经营数据的有效补充，在提升模型区分度、反欺诈方面提供了有力支撑

- 征信数据
- 购物数据
- 外部负债数据等

上百个子模型池，利用特征共计上万维，强有力的模型可监控性，

树形模型融合



高性能

区分度强是衡量一个模型的根本

可解释性

可对模型的关键步骤进行解释，以进行实时监控

可插拔性

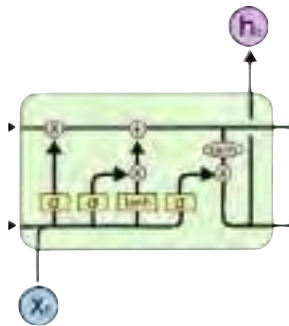
高通贷数据源较多，数据源的添加和删除较为频繁。新增数据源时保持原有模型的稳定性，在数据源异常时可及时进行停用

可复用性

在跨平台场景下，能够实现子模型良好的可复用性，降低跨维护成本

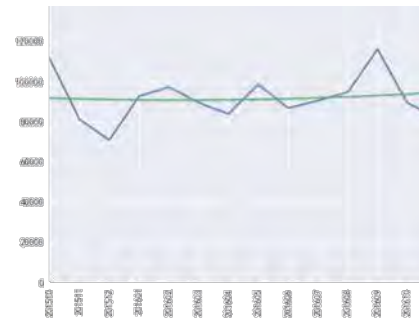
深度神经网络

基于时序序列的销售预测

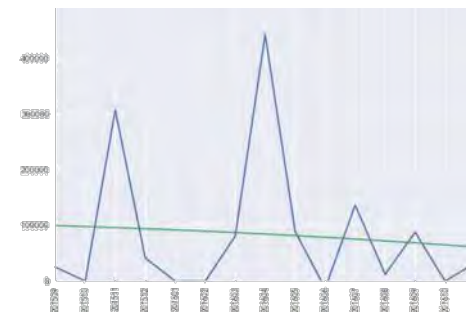


利用聚类技术对客户进行分类
利用循环神经网络预测未来销售额
针对不同经营模式企业，给予相应的额度及费率

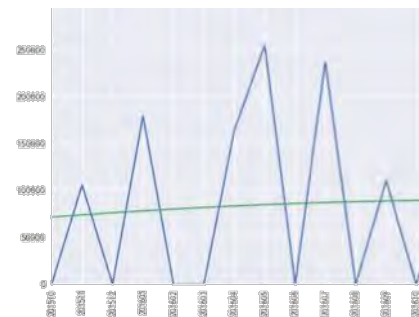
平稳型 XXXX教育科技有限公司（教育培训）



大单型 XX展览展示服务(上海)有限公司（婚庆服务）



波动型 XXXX网络科技有限公司（其它服务业）



季节性 XXXX农副产品有限公司（农副产品批发零售）

