



**QCon** 全球软件开发大会  
INTERNATIONAL SOFTWARE  
DEVELOPMENT CONFERENCE

BEIJING 2017

# 大数据驱动的实时信贷业务实践

张京磊

# 目录

01

业务背景

02

实时智能决策平台

03

数据的使用心得和案例

04

小微企业信贷业务实践

05

总结

# 宜信介绍

2006年 **宜信** 公司成立于北京，十年磨一剑，以强大科技基因引领金融产品创新，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** 和 **财富管理** 两大板块

- **10年普惠金融、财富管理经验**
- **年交易规模突破1,000亿**
- **香港、新加坡、美国、英国、澳大利亚等设立团队及办公室**
- **覆盖全国251个省市自治区93个农村地区**
- **合作伙伴机构数1,000+**
- **累计服务客户超过200万+**
- **在职员工4,5000+**



# 一、业务背景

商通贷是一个**专为小微企业解决融资难题**的综合性金融服务平台



在2016年全年，实现合作平台**40余家**、对接渠道**100余个**

- **简单**：仅需在线提交**个人有效身份信息**与一个或多个不同平台的**经营数据**，不需抵押，即可进行实时授信；
- **快速**：最快在**1分钟**内完成授信，在**8分钟**内放款；
- **费率低**：月费率低至**0.72**，最高授信**200万元**，最高期限24个月；
- **在线**：从注册申请、信用审核、到放款及还款的各个环节**全部线上完成**。



## 关于商通贷

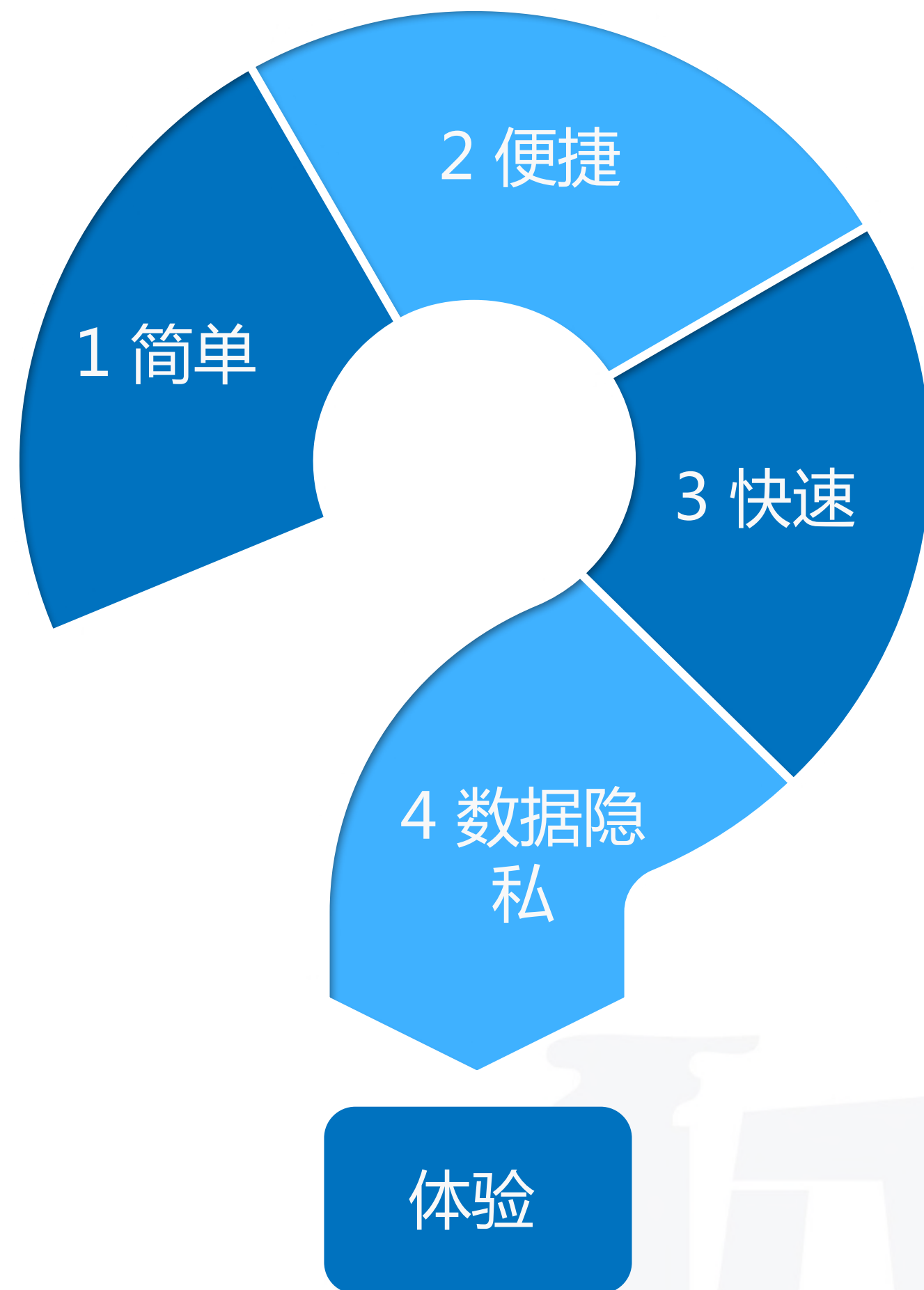
宜信商通贷，定位于专为小微企业解决融资难题的实时授信金融服务平台。基于Fintech价值理念、宜信大数据金融云平台，借款用户仅需在线提交个人有效身份信息与实际经营数据（网店经营、ERP管理、线上财务管理）即可进行实时授信，高效快捷的为用户解决资金周转问题。

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。

只要五步即可  
完成借款



# 互联网金融产品的体验要求



- 1 数据项少，资料在线提交
- 2 多端接入，便捷体验
- 3 高效率审批，快速放款
- 4 数据隐私，只收集授权必要项

## 二、实时智能决策平台



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

# 实时智能决策平台

进件申请

申请表 + 采集数据

数据收集

公网爬虫&授权爬虫

内部数据整合

三方数据网关

贷后数据

数据整合

风险搜索引擎

非结构化数据服务, 网查, 问答系统

知识图谱

结构化数据服务, 归户, 查重, 黑名

分析平台

风险分析, 建模, 政策评估, 离线训

智能决策

JSON Array

决策引擎

发布策略和模型

机器学习模型

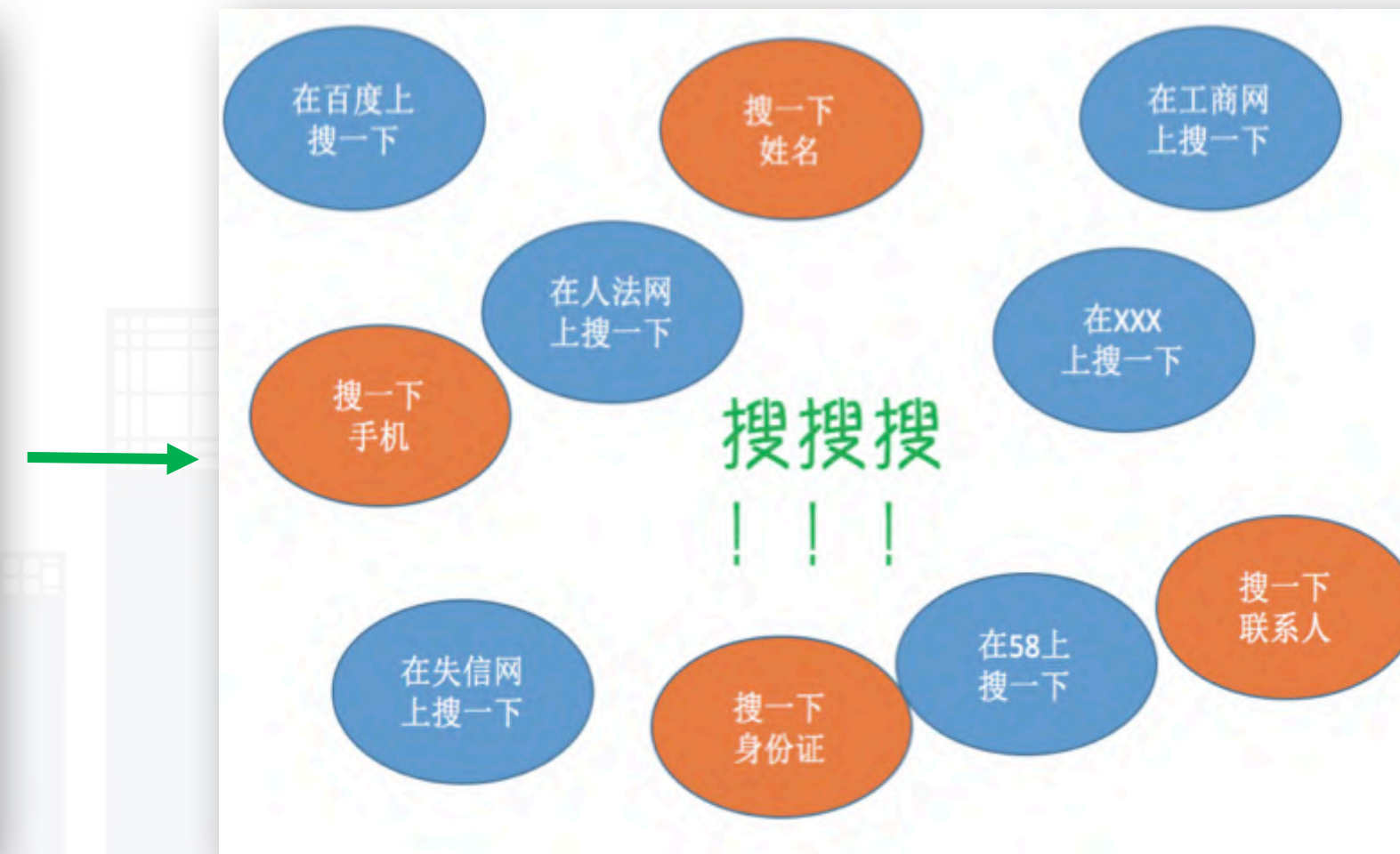
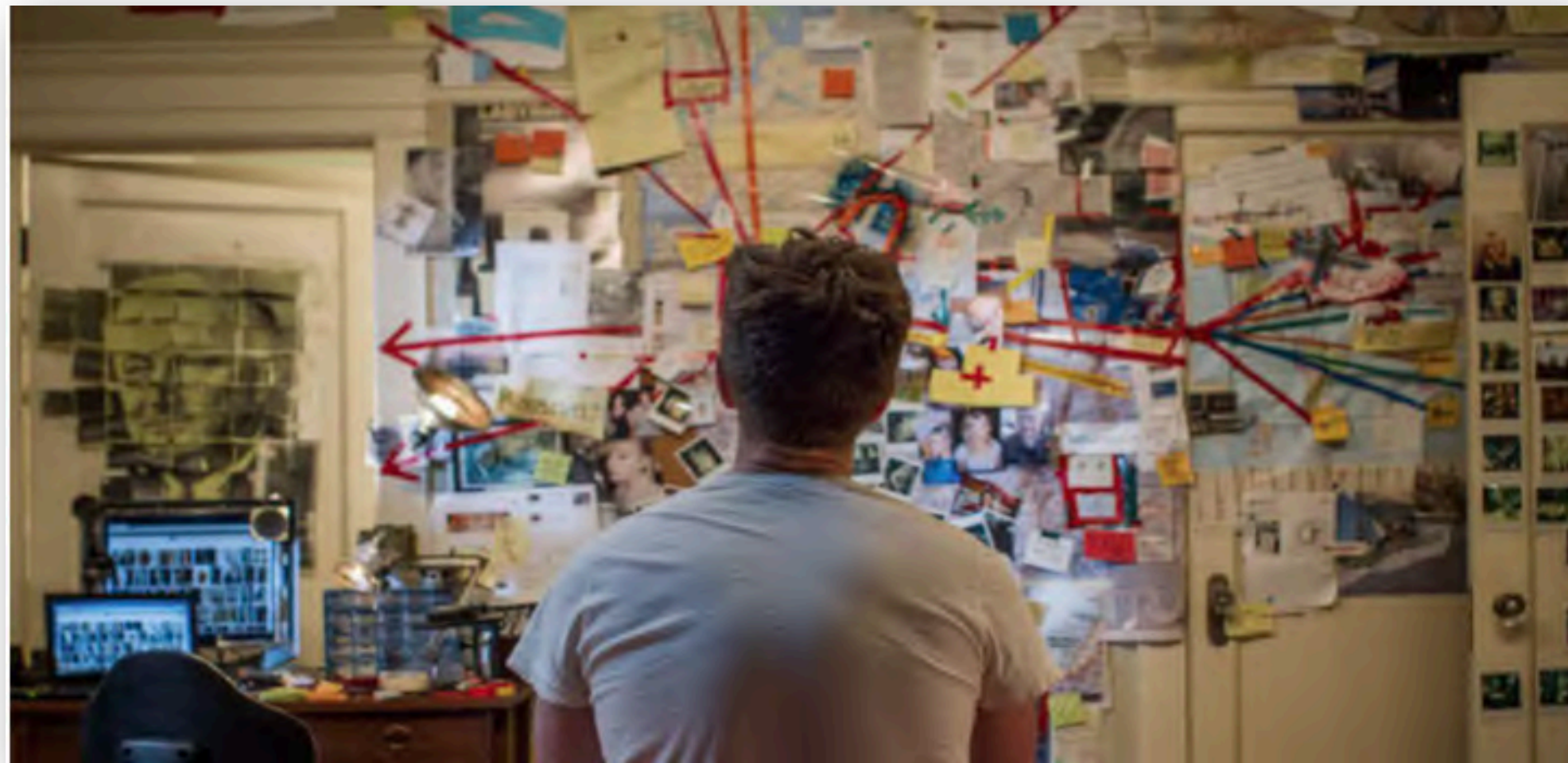
实时风险定价



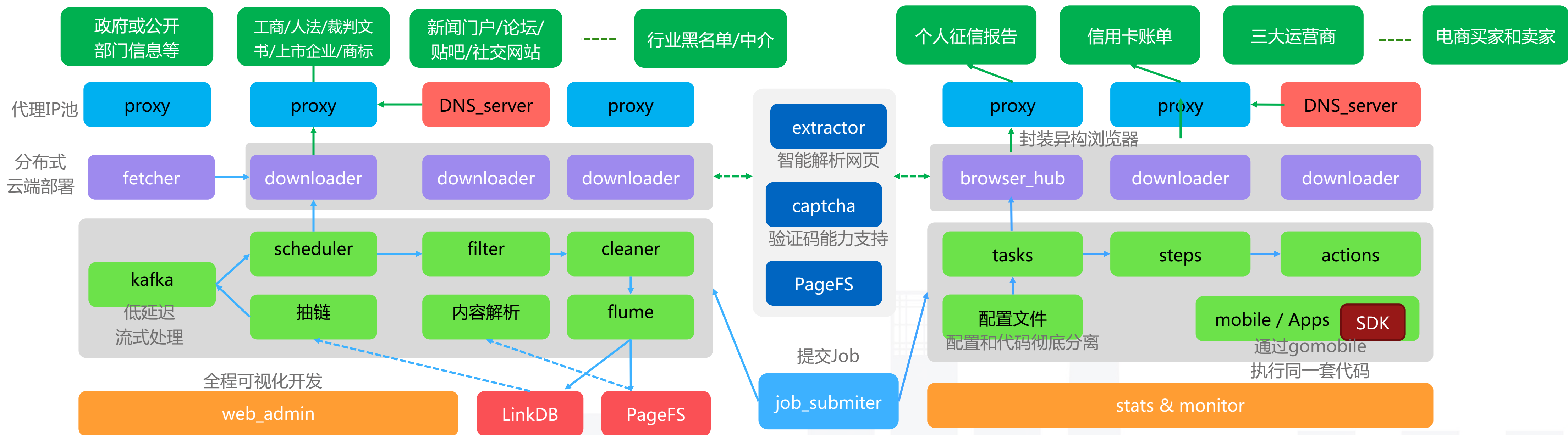
# 1. 数据收集：公开爬虫&授权爬虫

## Motivations：

- 数据分散在各种异构网站
- 线下提交资料，繁琐缓慢，容易造假
- 人工搜索效率低，精度差
- 对数据变化难以实时感知反应



# 爬虫技术架构 ( 公开爬虫 & 授权爬虫 )



# 配置文件简洁灵活

1. 类似Python的string Template和Java的Velocity模板引擎；
2. Go template的独特优势：
  - 支持.Funcs方法自定义函数；支持管道，类似shell中的“|”，将前一个结果作为后一个函数的参数；
  - 由此可以自定义函数，并且随意组合，简洁有效；
3. 抽象语法：配置文件有变量和step、goto、condition/if/else、for循环、context等概念。

```
func add(a, b int) int{
    return a + b
}
func mul(a, b int) int{
    return a * b
}
func sub(a, b int) int{
    return a - b
}
```

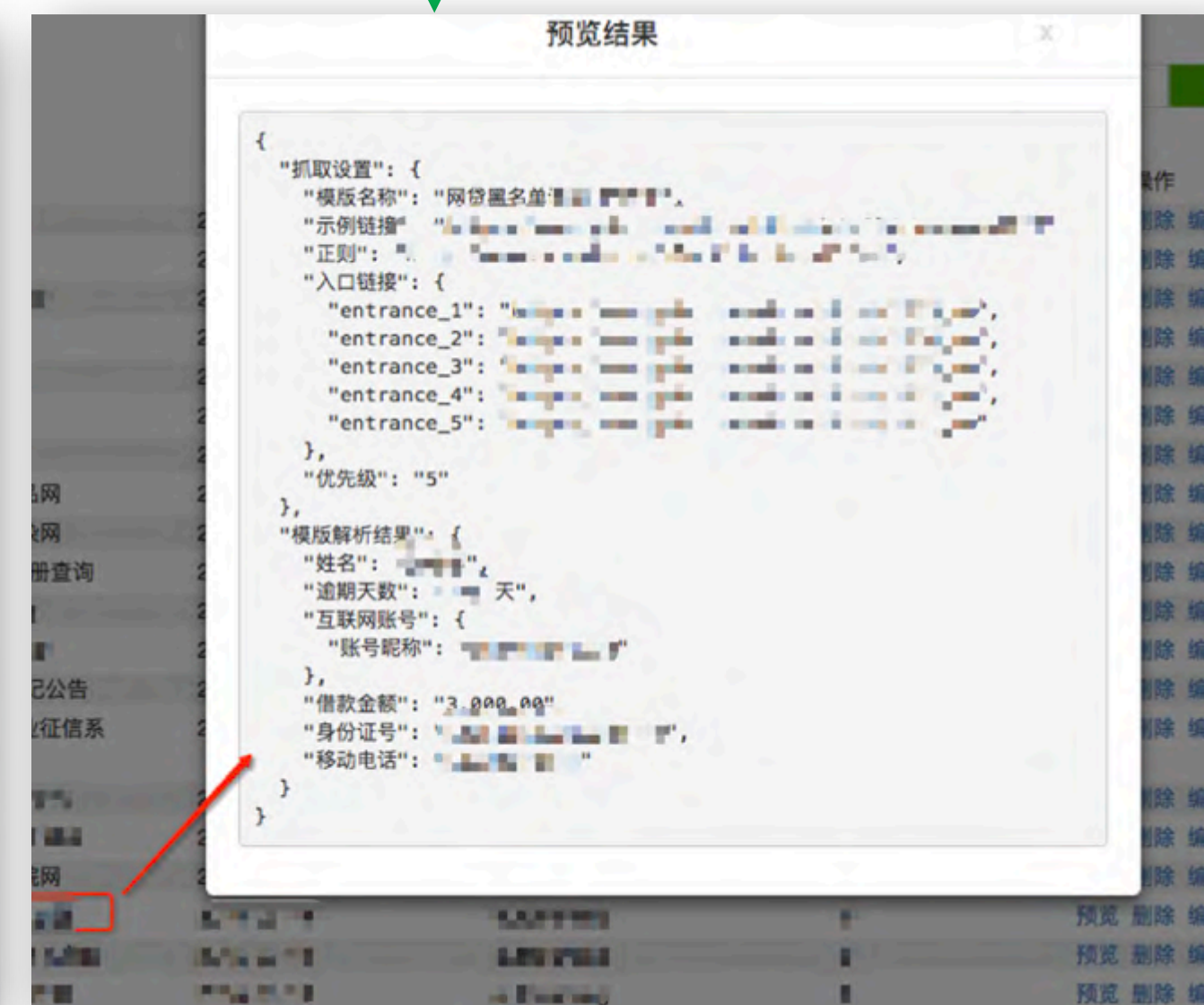
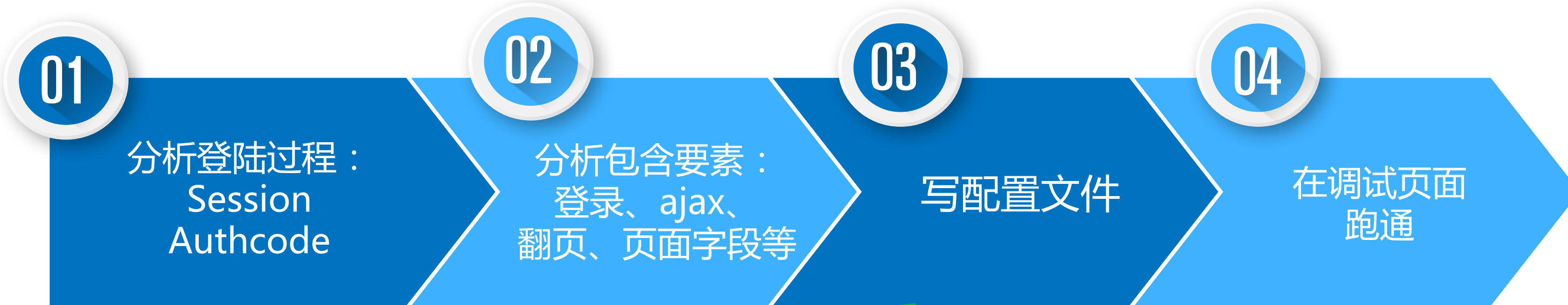
```
func newTemplate() *template.Template{
    return template.New("").Funcs(template.FuncMap{
        "add":add,
        "mul":mul,
        "sub":sub,
    })
}
```

```
/** 15-4*(1+2)=3 **
fmt.Println(parse("{{add 1 2 | mul 4 | sub 15}}"))
/** 2*(6+(5-3))=16 **
fmt.Println(parse("{{sub 5 3 | add 6 | mul 2}}"))
```

# 通过gomobile执行跨平台代码

- 1** Golang的移动开发支持两种模式，一种是使用Golang开发整个应用，另外一种便是使用Golang开发common library；
- 2 打包到Android**：gomobile bind-target=android <http://git.cc/{groups}/repositories> -o common\_server.aar，在Android的项目中直接引用这个aar文件；
- 3 打包到iOS**：gomobile bind-target=ios <http://git.cc/{groups}/repositories>，把生成的CommonServer.framework 加入iOS 项目中即可；
- 4** 在go 1.5提供，在go 1.6之后走向成熟，有了显著提升和优化。

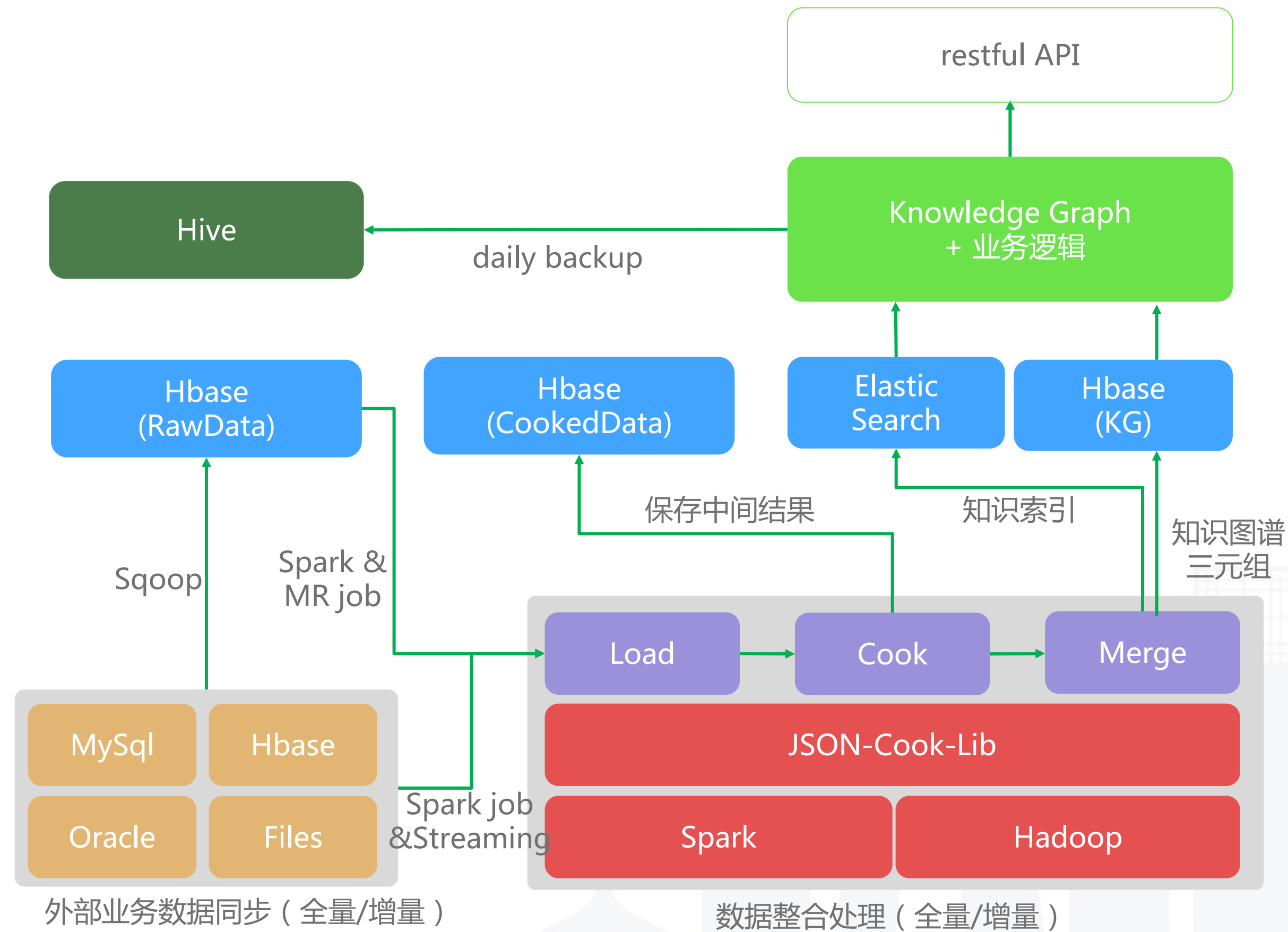
# 如何快速开发一个登陆爬虫



27 台爬虫 2400万 网页/天  
24 台大数据处理 5TB 数据/天  
12 台搜索 12TB 数据  
3亿 实体 20亿 关系 1000万 黑名单疑似中介

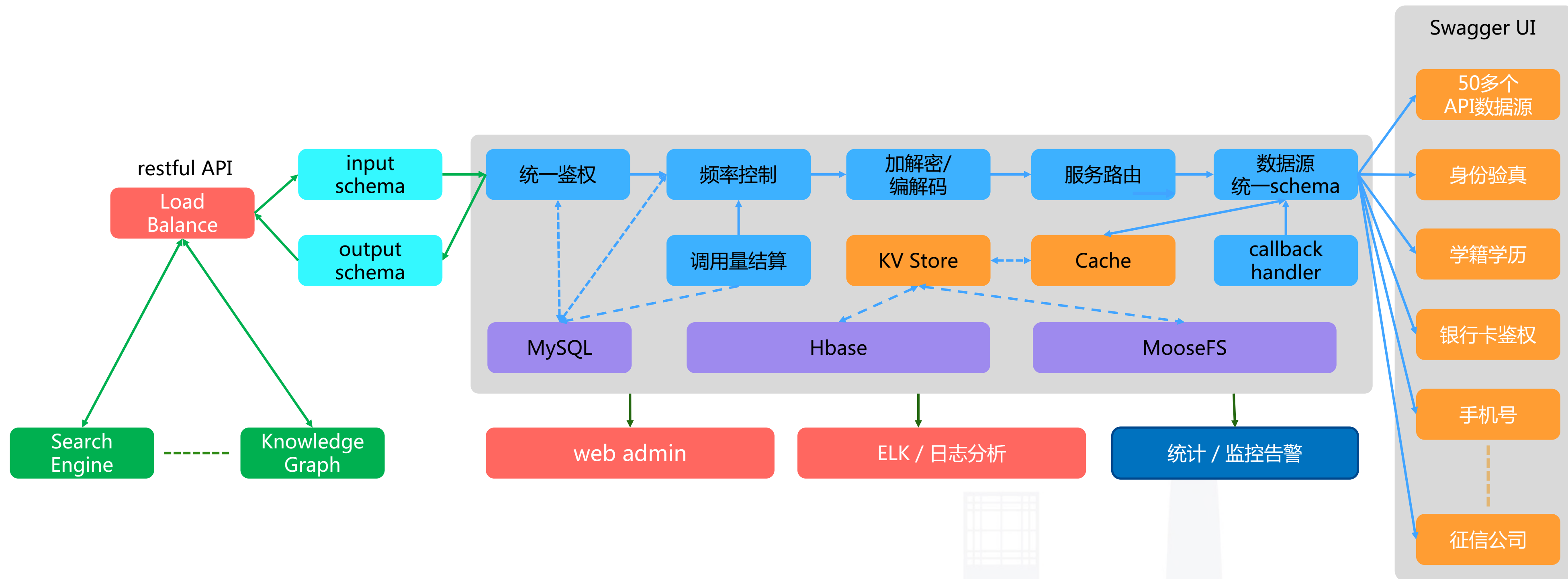
- 秒级抓取，实时解析，离线聚合查询分析；
- 抽象智能解析引擎，通过Web即可配置和管理模版；
- 整体识别成功率超过97%；
- 基于机器学习的风险数据拦截，能够智能识别欺诈伪装数据。

# 1. 数据收集：内部数据整合



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. airflow调度
5. spark实时流式处理

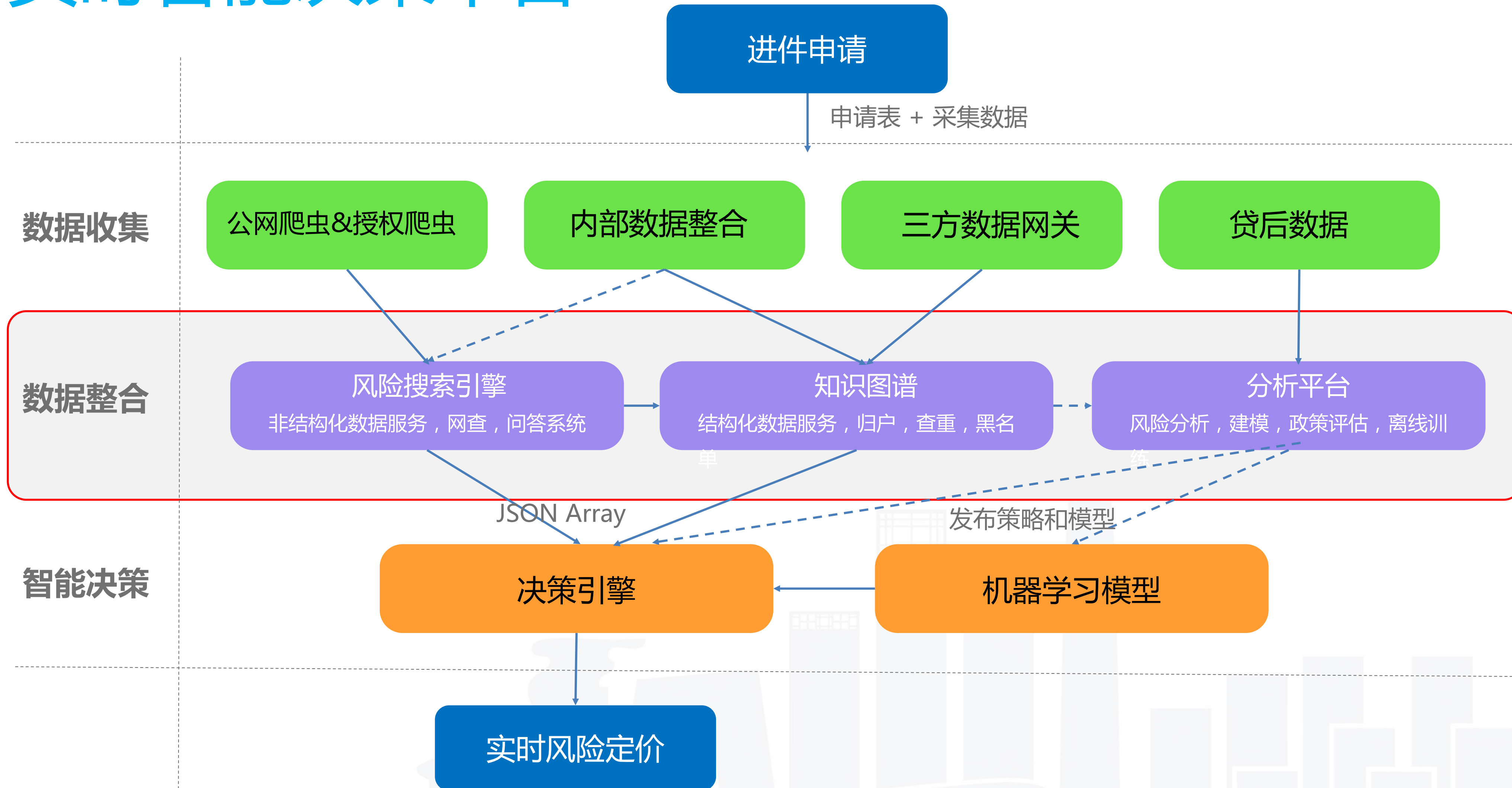
# 1. 数据收集：三方数据网关



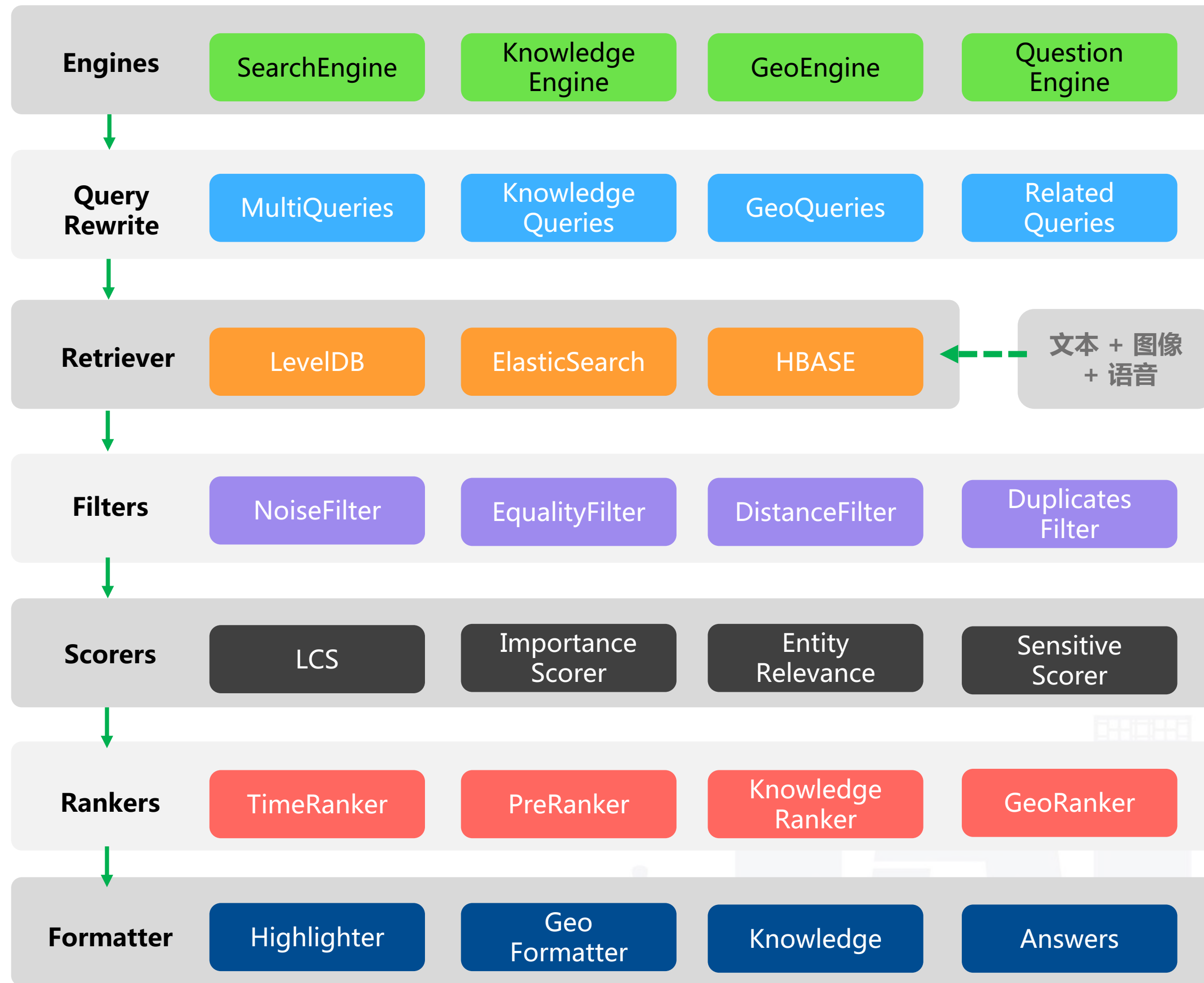
1. 使用Golang开发，借助Go Template、routine和channel，性能优异；
2. 极大简化异构 / 多维数据源的接入；
3. 统一鉴权、频率控制及调用量结算；
4. 全数据源维度的数据分析，实时监控告警。



# 实时智能决策平台



## 2. 数据整合：风险搜索引擎



### 风险搜索VS人工信审

39%

效率提升

31%

效果提升

从文档添加，到被搜索到，小于5秒

## 2. 数据整合：知识图谱



### 知识图谱schema

- Ontology
- 类别，层次，属性，值，关系
- 自顶向下，自底向上

```
m:people.person rdf:type owl:Class ;  
  rdfs:label "人"^^xsd:string ;  
  rdfs:subClassOf m:common.topic ;  
  
m:people.person.name rdf:type  
owl:DatatypeProperty ;  
  rdfs:domain m:people.person ;  
  rdfs:label "姓名"^^xsd:string ;  
  rdfs:range xsd:string ;  
  
m:people.person.work_at rdf:type  
owl:ObjectProperty ;  
  rdfs:domain m:people.person ;  
  rdfs:label "工作单位"^^xsd:string ;  
  rdfs:range m:organization.organization .  
  
m:organization.organization rdf:type owl:Class ;  
  rdfs:label "组织"^^xsd:string ;  
  rdfs:subClassOf m:common.topic ;
```

定义结构化数据 ( RDF Schema )

```
e:borrower_1, rdf:type, m:finance.borrower  
e:borrower_1, m:people.person.name, "张三"  
e:borrower_1, m:people.person.phone, "13800138000"  
e:borrower_1, m:people.person.home_addr, e:location_1  
e:borrower_1, m:people.person.work_at, e:company_1  
e:location_1, rdf:type, m:common.location  
e:location_1, m:common.location.full_addr, "北京市朝阳区东方梅地亚中心"  
e:company_1, rdf:type, m:finance.company  
e:company_1, m:finance.company.name, "宜信普惠"  
e:company_1, m:finance.company.telephone, "400-6099-800"
```

知识图谱的数据转化

# 知识图谱的检索与应用

## 1 实体查询

- 黑名单查询：直接匹配 & 间接联系
- 客户查重

## 2 实体扩展

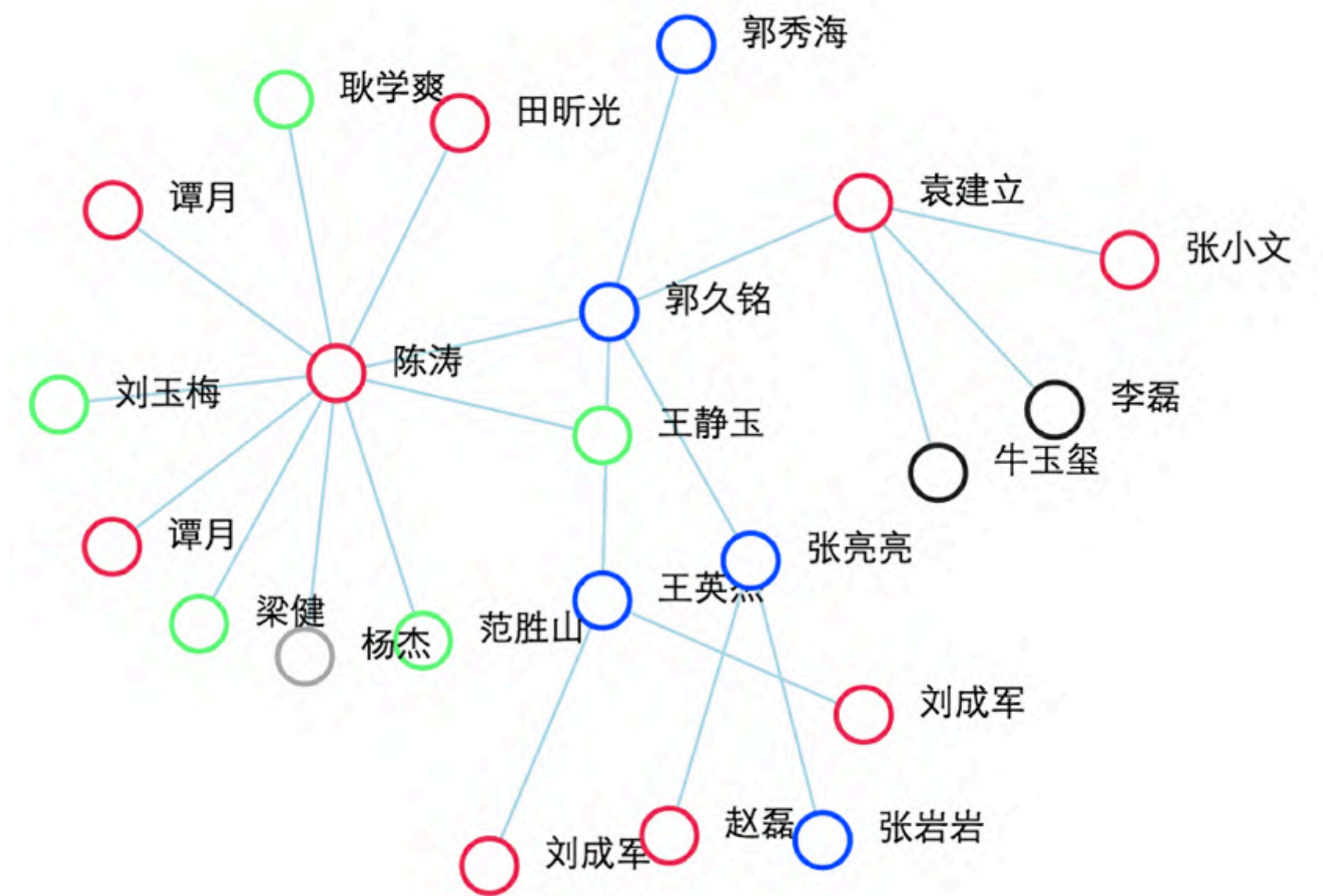
- 通过属性查询实体 => 获取实体的其他属性
- 获取相似实体

## 3 知识查询

- 例：XX公司的法人是否有失信记录
- 例：XX写字楼中有多少家机构入驻

## 4 个性化问答

- 由客户信息生成个性化问题
- 根据客户回答，检测虚假身份等欺诈风险



# 金融智能化之路

信息化

从线下到线上



自动化

从繁琐到简单



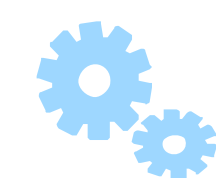
大数据化

从割裂到融合



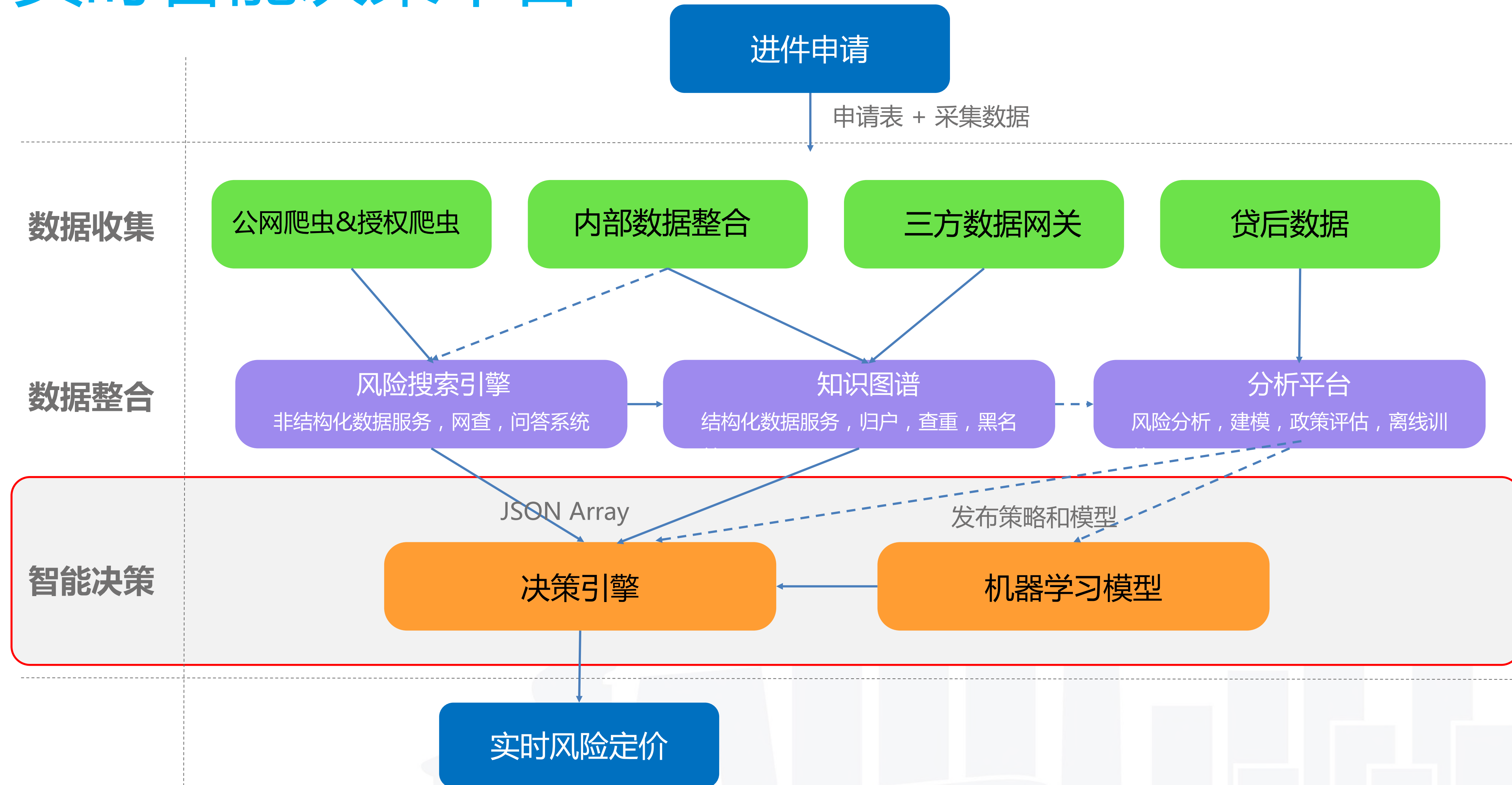
智能化

从经验到智能



智能决策引擎

# 实时智能决策平台



# 3. 智能决策：决策引擎

## 自定义核查信息

姓名、身份证号、固定电话、手机号码、现居住地、户籍地址、工资卡号、就职公司全称、公司电话、联系人电话.....

## 动态策略调优

可视化BI报表  
命中分析  
策略建议

## 智能引擎

规则决策、决策树、随机森林、贝叶斯网络、深度学习



## 灵活接入

Web形式编辑测试  
API方式对接上线  
全可视化结果展现

## 可视化策略定义



特征集



样例规则

真实性核查

欺诈风险因素匹配

信息综合交叉验证

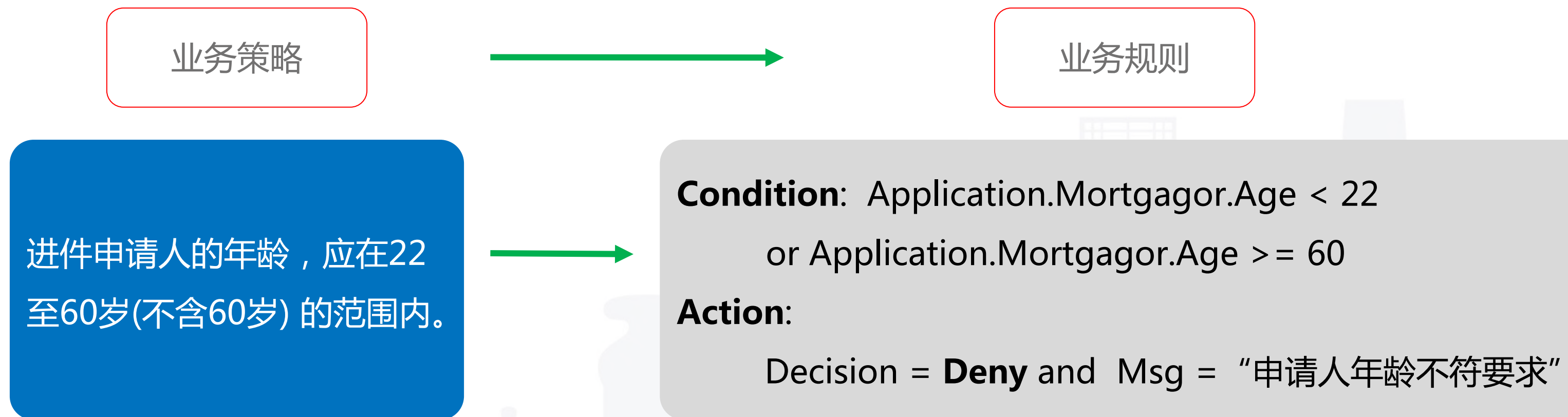
自定义评分策略

线下反欺诈支持

# 决策引擎技术组件

## 1 业务决策对象 ( 如Mortgagor , Decision , etc )

**决策 ( 规则 )** 实质上也可以理解为一组条件和在此条件下的操作，是一组准确凝练的语句，用于描述、约束及控制业务的结构、运作和战略。





# 决策引擎技术组件

## 2 决策Schema定义 ( 用JSON描述的业务规则和Model )

```
{  
  "probability": "1.0",  
  "desc": "年龄小于22岁或者大于等于60岁",  
  "priority": 25,  
  "status": "online",  
  "logic": {  
    "or": [  
      {  
        "op": "=",  
        "dst": "NULL",  
        "src": "$age_years"  
      },  
      {  
        "op": "<",  
        "dst": 22,  
        "src": "$age_years"  
      },  
      {  
        "op": ">=",  
        "dst": 60,  
        "src": "$age_years"  
      }  
    ]  
  },  
  "action": [  
    {  
      "action": "DENY",  
      "deny_month": 3  
    }  
  ]  
}
```

执行概率

判断条件

特征

执行动作

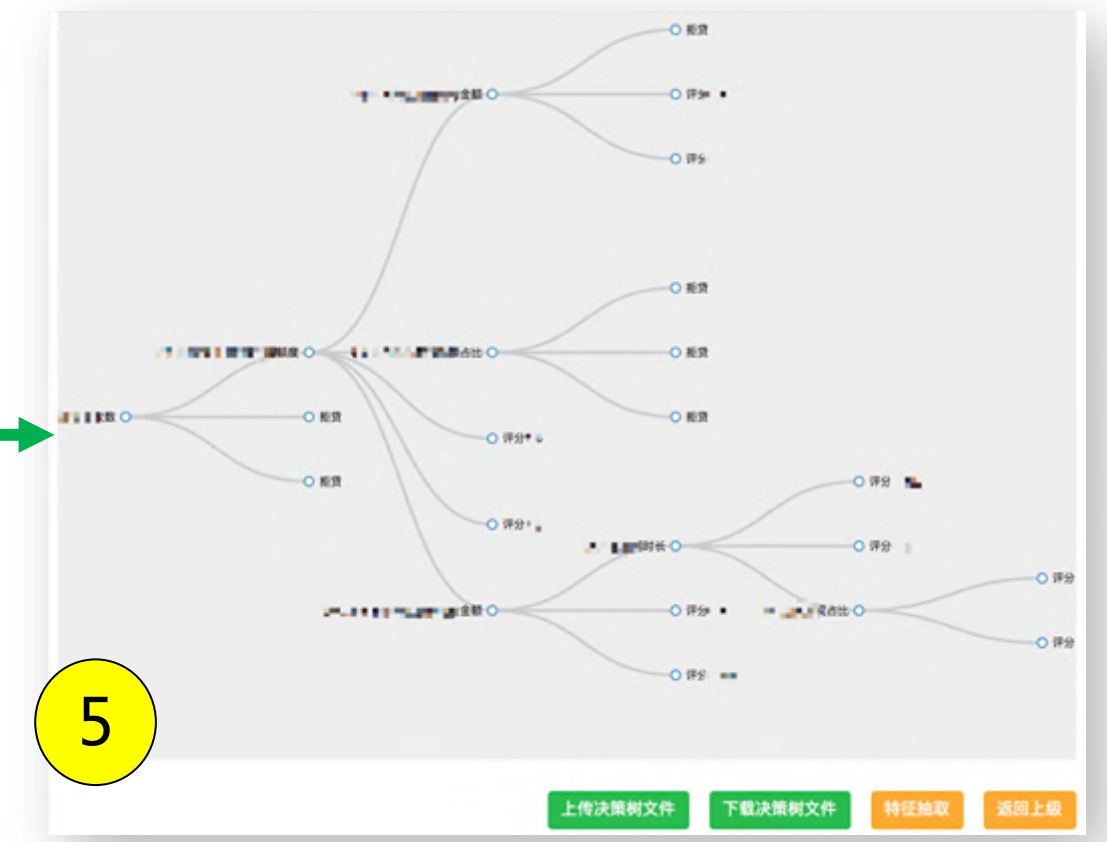
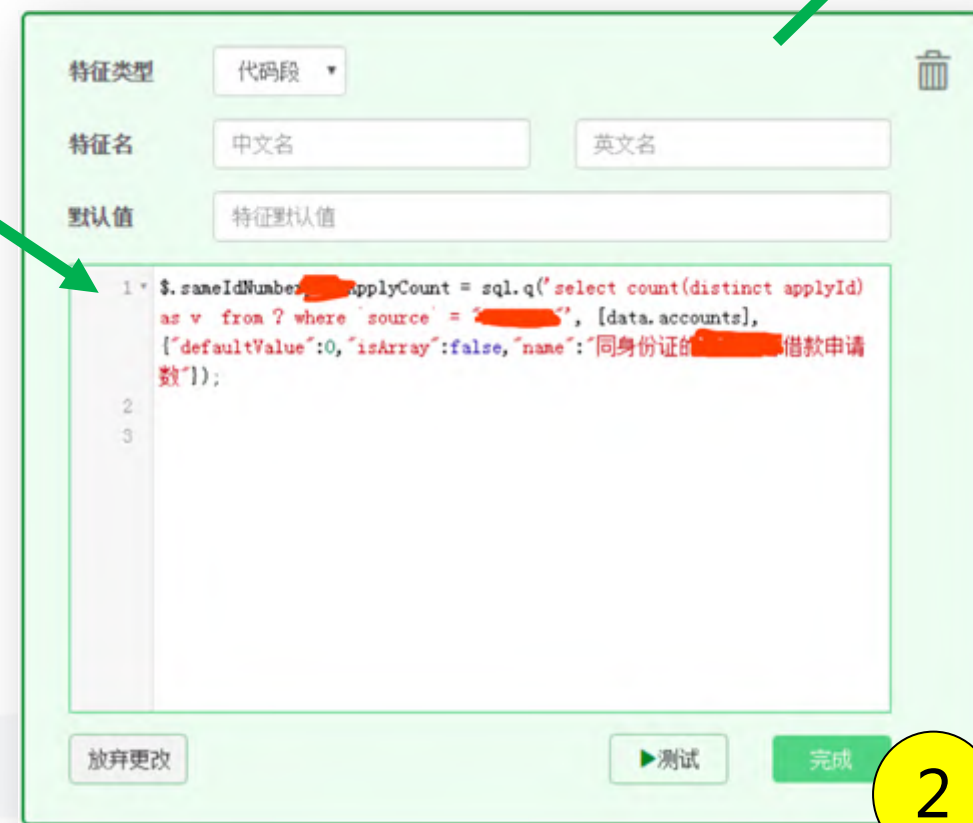
- 以JSON的格式，用业务和风控政策人员熟悉的业务语言描述；
- 支持Number（各种格式）、String、JSON、Array等多种数据类型；
- 支持常用操作符：四则运算、取余、取整、比较，支持Array操作，以及集合操作：in，contains one，并集，交集，求和，相减等；
- 支持自定义任意函数，如时间戳、随机数、编解码等。

# 决策引擎技术组件

3

## 决策规则和模型组件：

- 单策略：规则，特征，模型，评分，etc
- 复合策略：决策树，评分卡，决策表，etc
- 全自动化特征抽取，只需写简单的类SQL语句
- 极友好的可视化编辑，一键生成策略，简单直观



# 决策引擎技术组件

## 4 决策集和决策流

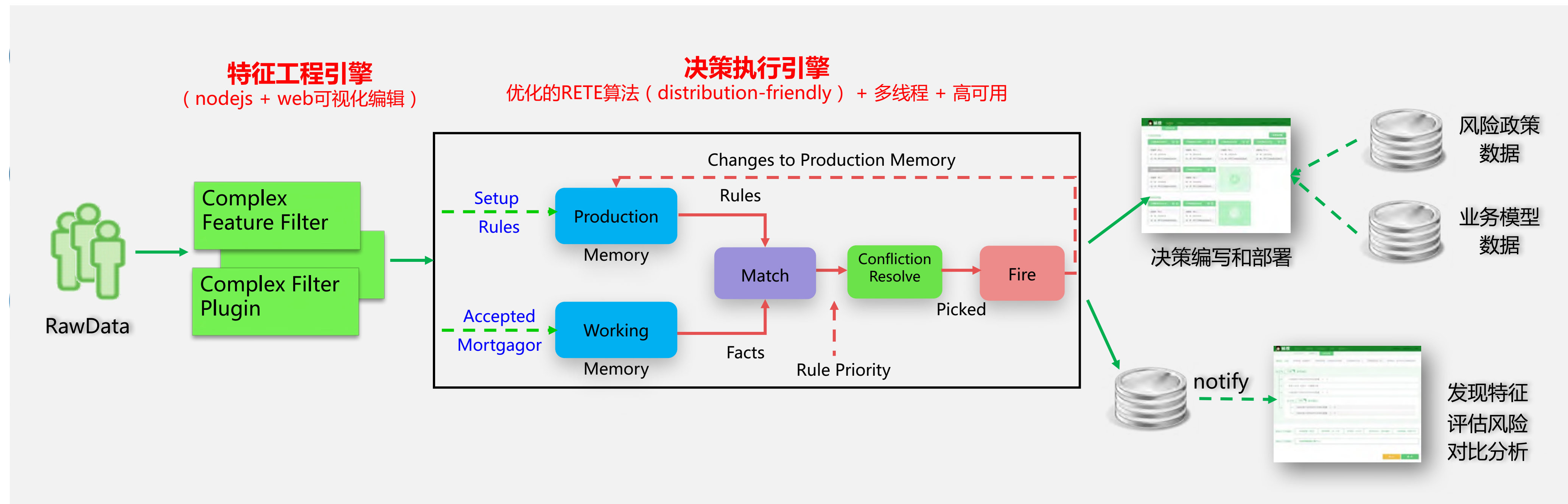
The image shows a two-step process in a decision engine interface:

- 1 进入列表页** (Enter List Page): A dashboard view titled "我的规则集" (My Rule Sets) with a "新建资源集" (New Resource Set) button. A red box highlights a green "+" button, with an arrow pointing to the next step.
- 2 定义决策流** (Define Decision Flow): A table view titled "风险定价申请" (Risk Pricing Application) with a search bar and a "特征抽取" (Feature Extraction) button. The table contains the following data:

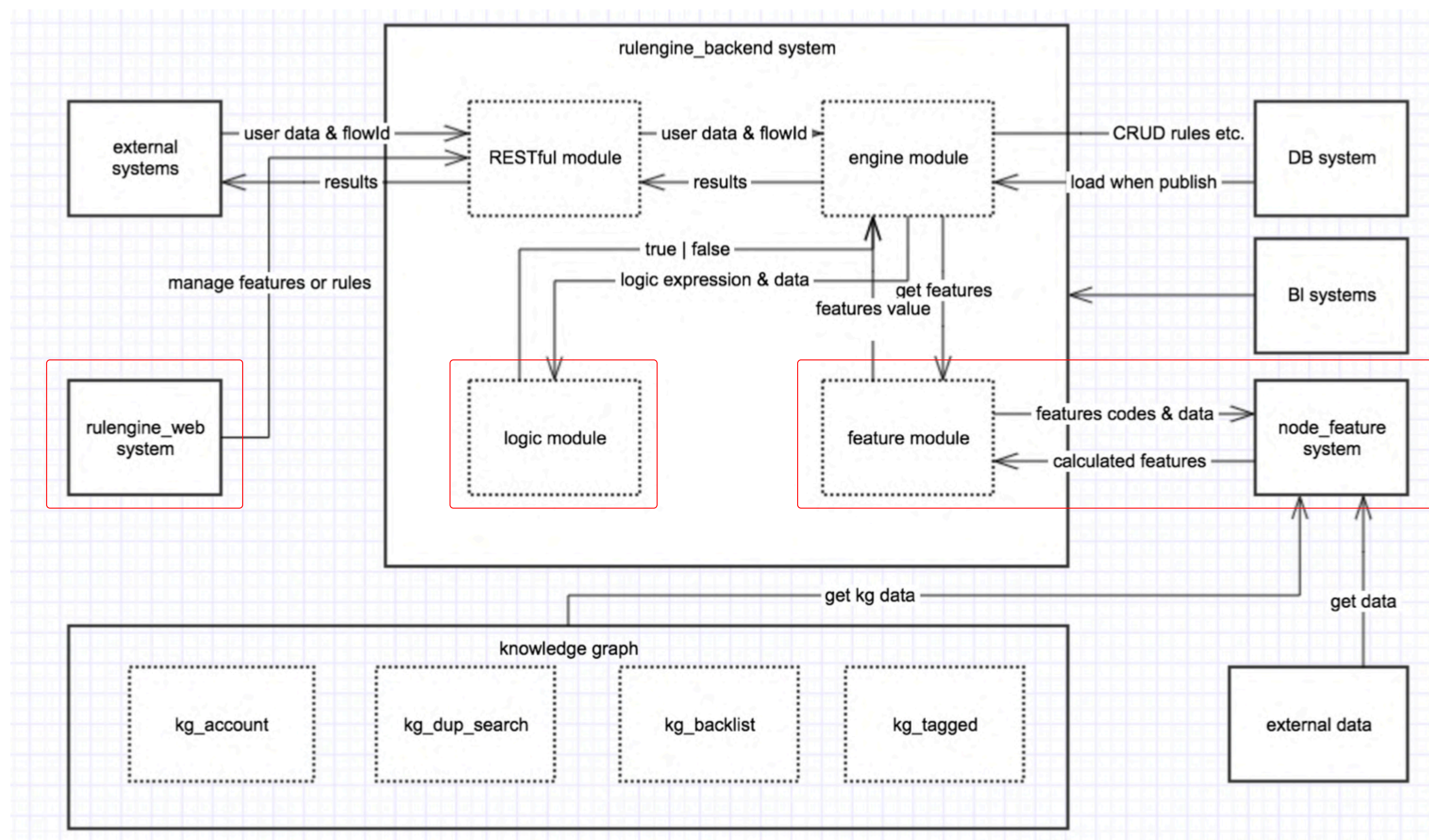
描述	审批结果	最近更新人	修改时间	规则优先级	状态	备注
风险定价申请...	审批通过	admin	2017-03-01 10:00:00	100	启用-在线	规则生效
风险定价申请...	审批通过	admin	2017-03-01 10:00:00	100	启用-在线	规则生效
风险定价申请...	审批通过	admin	2017-03-01 10:00:00	100	启用-在线	规则生效
风险定价申请...	审批通过	admin	2017-03-01 10:00:00	100	启用-在线	规则生效
风险定价申请...	审批通过	admin	2017-03-01 10:00:00	100	启用-在线	规则生效
风险定价申请...	审批通过	admin	2017-03-01 10:00:00	100	启用-在线	规则生效
风险定价申请...	审批通过	admin	2017-03-01 10:00:00	100	启用-在线	规则生效
风险定价申请...	审批通过	admin	2017-03-01 10:00:00	100	启用-在线	规则生效

# 决策引擎技术组件

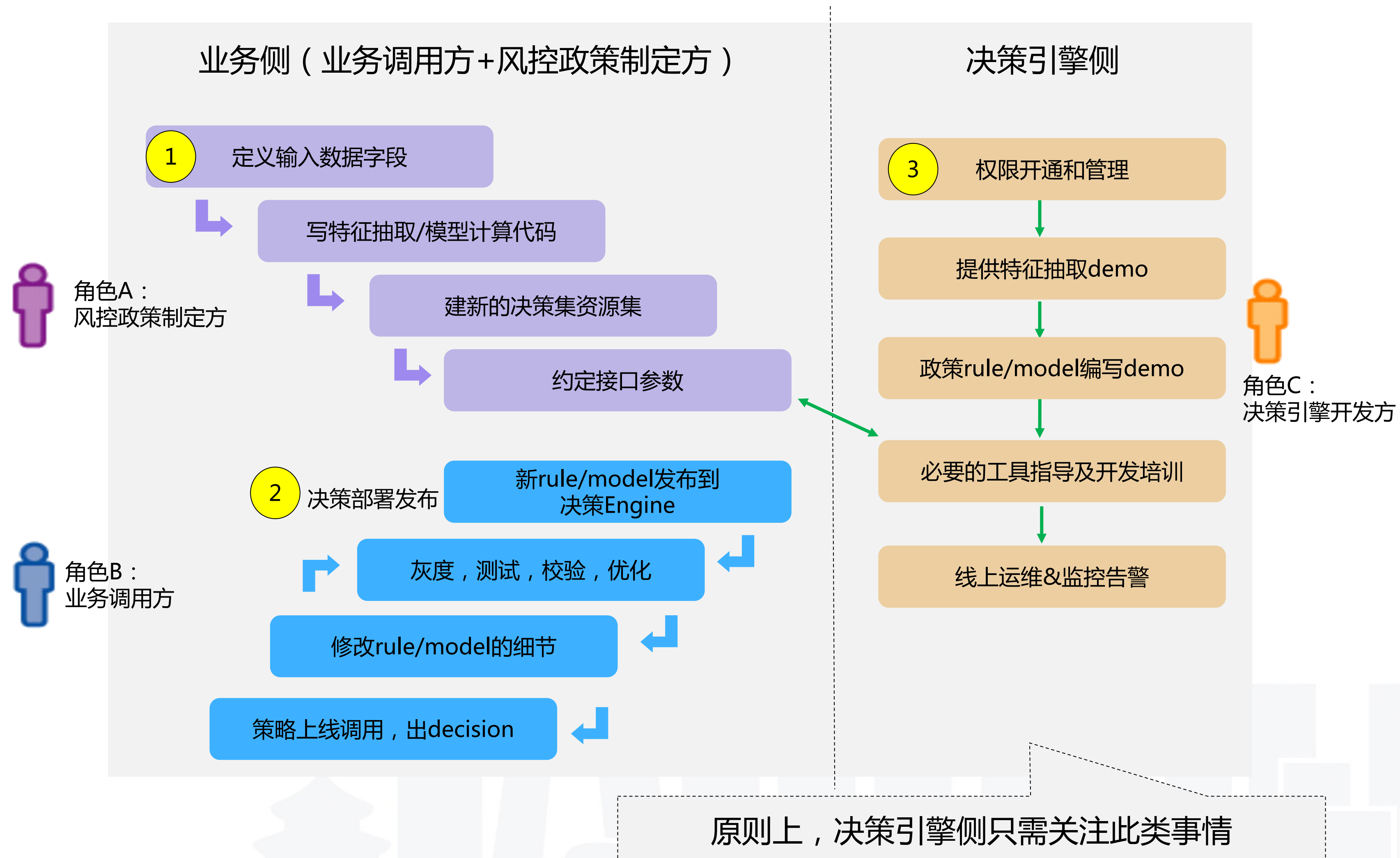
## 5 特征工程引擎 + 决策执行引擎



# 决策引擎技术架构



# 业务决策实施步骤



# 业务规则管理平台 ( BRMS )



- 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。



4000亿 累计决策金额 100% 可用性 (最近半年)

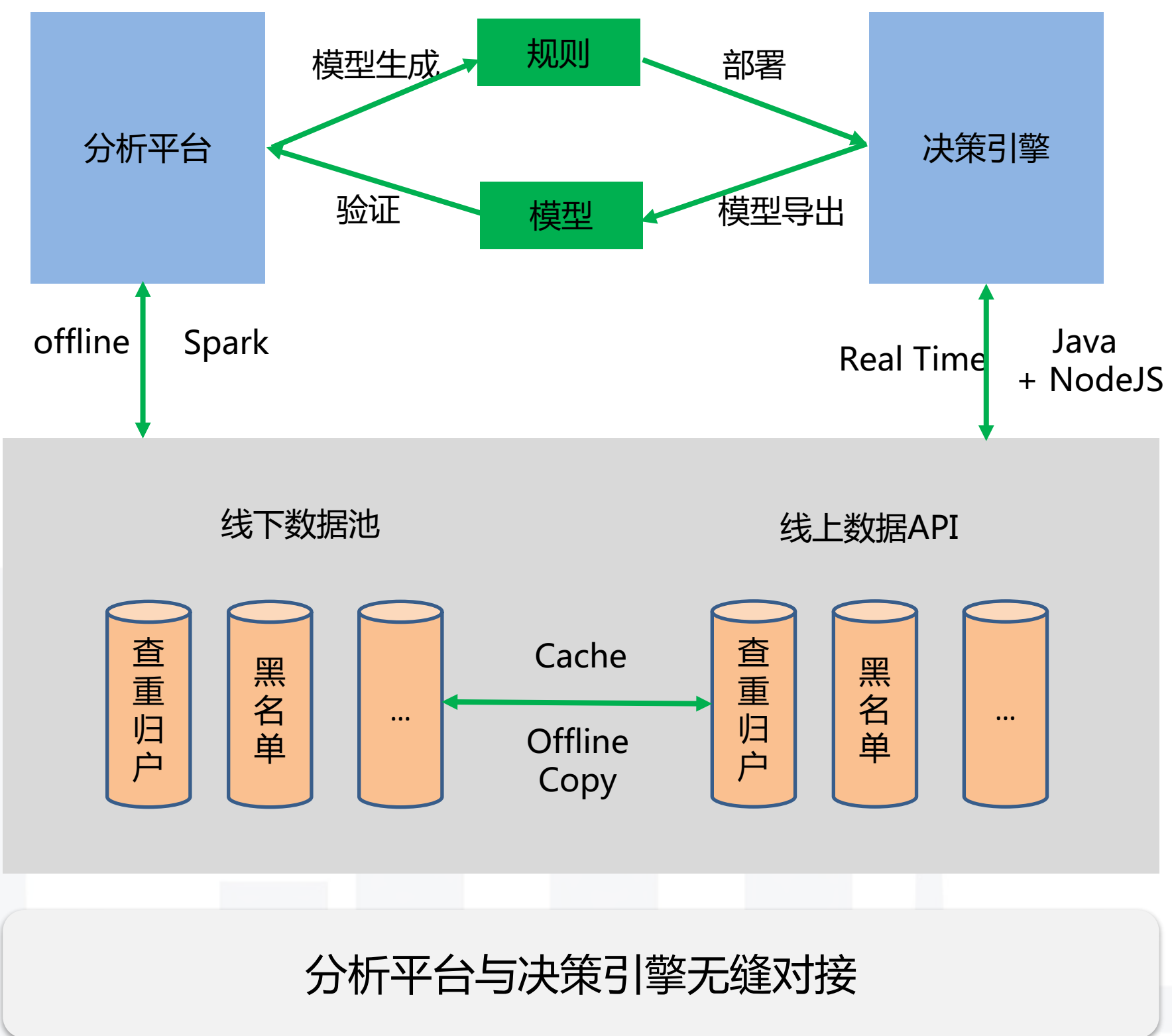
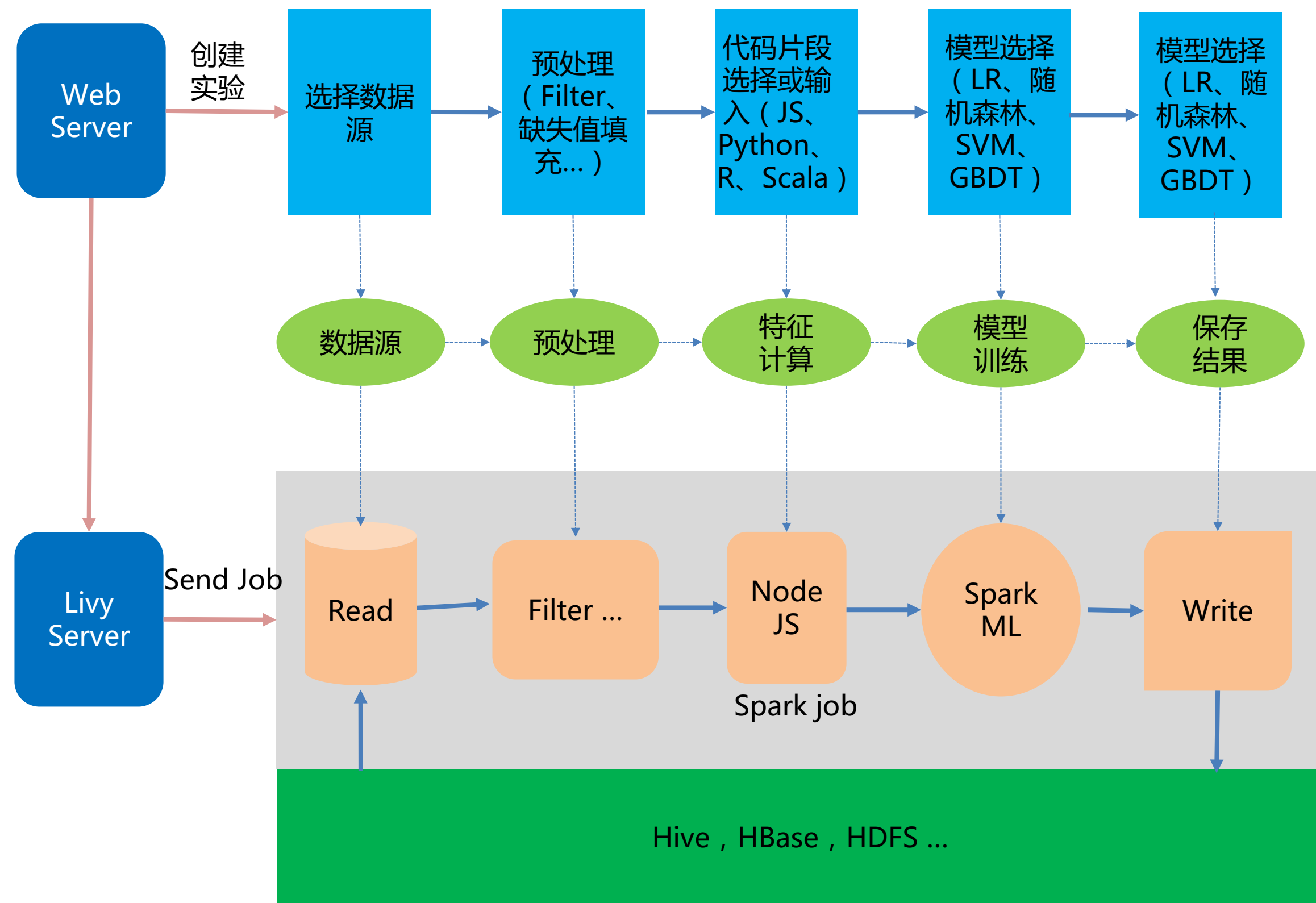
- 接入宜信16个业务线（宜人贷，商通贷，城市信贷等等），广泛提升了自动化风控决策能力；
- 大幅提高业务决策的实施速度，从过去的数天到现在的数分钟内完成；
- 集中管理风险政策，统一审批尺度，严格执行决策；防止**纰漏**、**舞弊**、甚至**内部欺诈**；
- **解放生产力**：一站式的政策模型编辑、测试、发布服务，稳定、高效、灵活、智能，使风险政策人员能够专注政策研究。



# 3. 智能决策：分析平台

实验定义阶段

实验运行阶段



# 三、数据的使用心得和案例

## 1 传统授信模型

- 仅作为人工的辅助手段
- 强特征
- 可解释

## 2 我们的反欺诈和授信模型：

- 一切数据皆为信用数据
- 维度无限制，强弱无限制
- 一切模型均可使用
- 没有边界

## 3 仅有数据和模型其实远远不够

- 寻找典型案例，寻找因果关系，依赖统计规律
- 方法论：变量构造，数据分析，发现规律和洞察

## 4 系统工程：

- 数据质量高，避免“garbage in, garbage out”（无用输入,无用输出）
- 重视工程实现的可靠性，可维护性

# 举个例子：数据源问题

## Motivations :

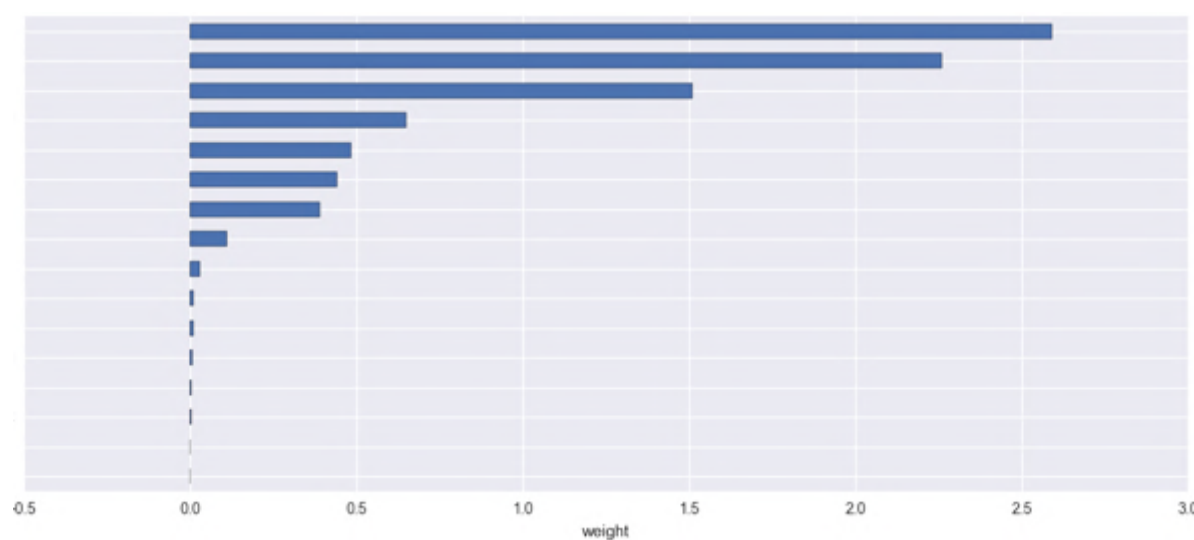
- **数据源的有限覆盖率**
  - 使用各个数据源的客户覆盖率在4% ~ 100%之间
  - 缺失值的种类: **unknown**, undefined, negative response
  - 缺失值的填充：不恰当填充会扭曲数据的真实分布
- **数据源的鲁棒性**
  - 数据源的故障很常见
  - 风控模型对数据源异常应该是可控、可预测的
  - 不会因为一个或几个数据源异常而失控
- **可扩展性 & 可解释性**
  - 异构特征和模型的融合导致解释性差
  - 能够并行开发不同数据源的特征和模型
  - 衡量不同数据源对风险预测的价值，并优化数据获取成本

Data source	Coverage
U*	100%
S*	30%
A*	24%
T*	4%

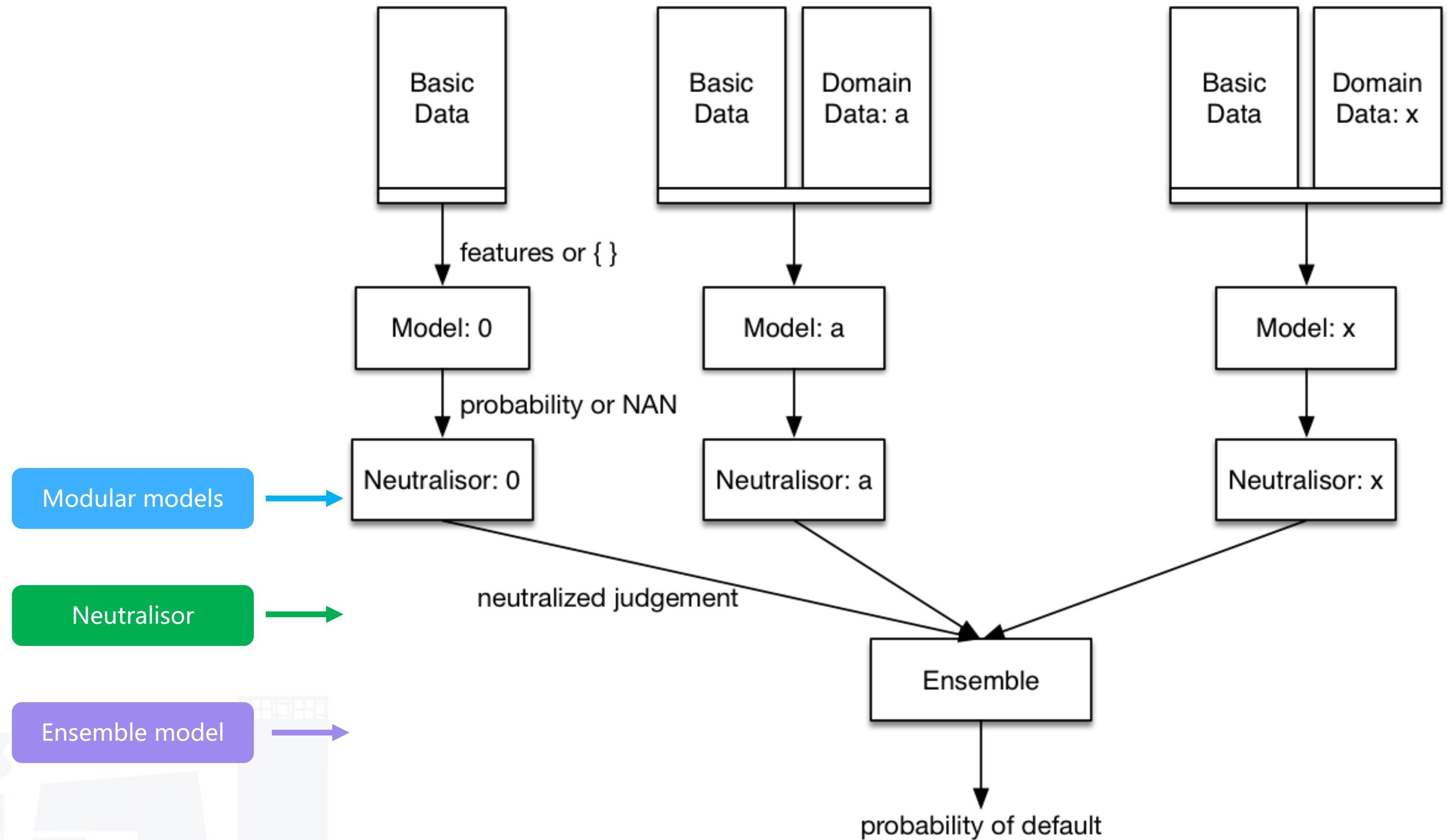
# 解决方案：模型融合

## Solutions :

- 数据源的覆盖率
- 数据源的鲁棒性
- 可扩展性 & 可解释性



对数据源的评估



# 四、小微企业信贷业务实践

## 商通贷

—— 专为小微企业解决融资难题 ——

**0.72%**

超低综合月成本

**1 minute**  
**1分钟**

预估额度时间

**2 minutes**  
**2分钟**

最快申请时间

**8 minutes**  
**8分钟**

最快放款时间

**24 months**  
**24个月**

期限畅享

**200万**

最高放款额

纯信用贷款 **0** 抵押

领先  
优势

- 自主研发，独立知识产权的宜信金融云
- 大数据风控体系
- 跨平台综合授信
- 互联网金融服务中成长最快平台之一

### 产品优势

#### 真正的信用贷

商通贷根据您的经营数据评估授信额度，无需任何抵押物，是一款真正的信用贷款。

#### 资金使用灵活

一旦在商通贷获得授信，授信额度在一个月内保持有效，不取用不计息，帮助小微企业提前轻松准备资金，灵活支取。

#### 支持跨平台授信

商通贷完全支持主流电商平台网店及实体经营数据（网店经营数据，ERP、物流、财务等管理数据），平台账号添加越多，获得授信额度越高。

#### 全程线上操作

从申请到放款，全程线上操作，点点鼠标，坐等资金到账。

### 跨平台综合授信

#### 网店销售数据

淘宝网  
Taobao.com

天猫  
TMALL.COM

JD.京东  
JD.COM

ebay

amazon

LAZADA  
来赞达

wish

All xpress  
全球速卖通

#### 实体经营数据

用友旗下企业

畅捷通  
Chanjet

起心金融  
管家婆品牌

# 商通贷整体技术架构

## 申请进件



PC官网



H5

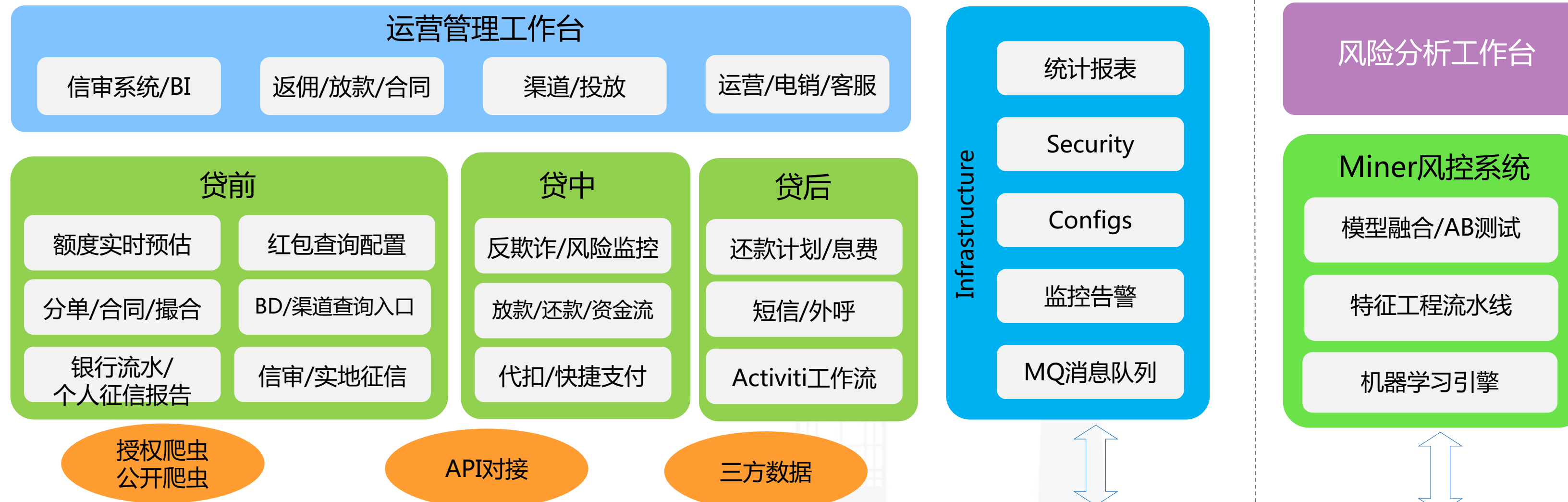


APP

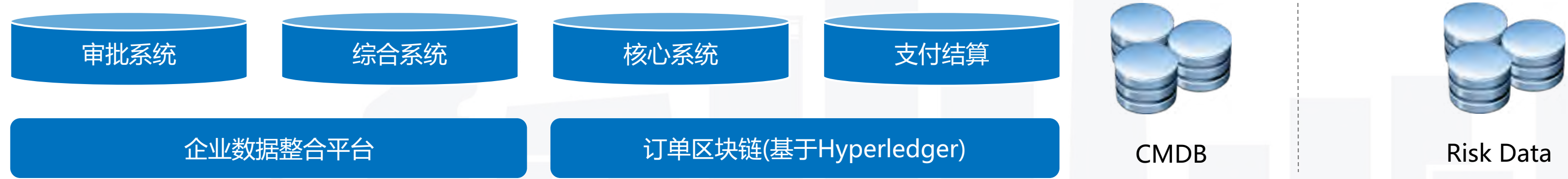


几百个合作渠道  
(通过API或iFrame灵活对接)

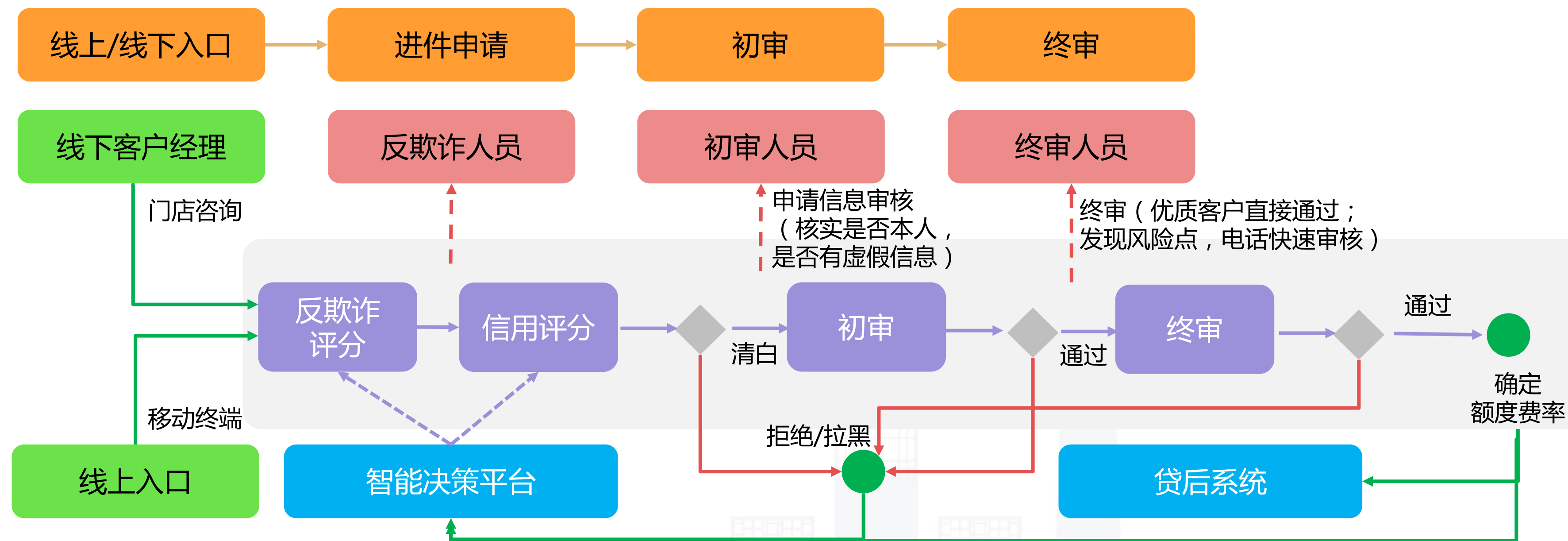
## 商通贷 支撑平台



## 基础服务



# 1. 进件审批系统



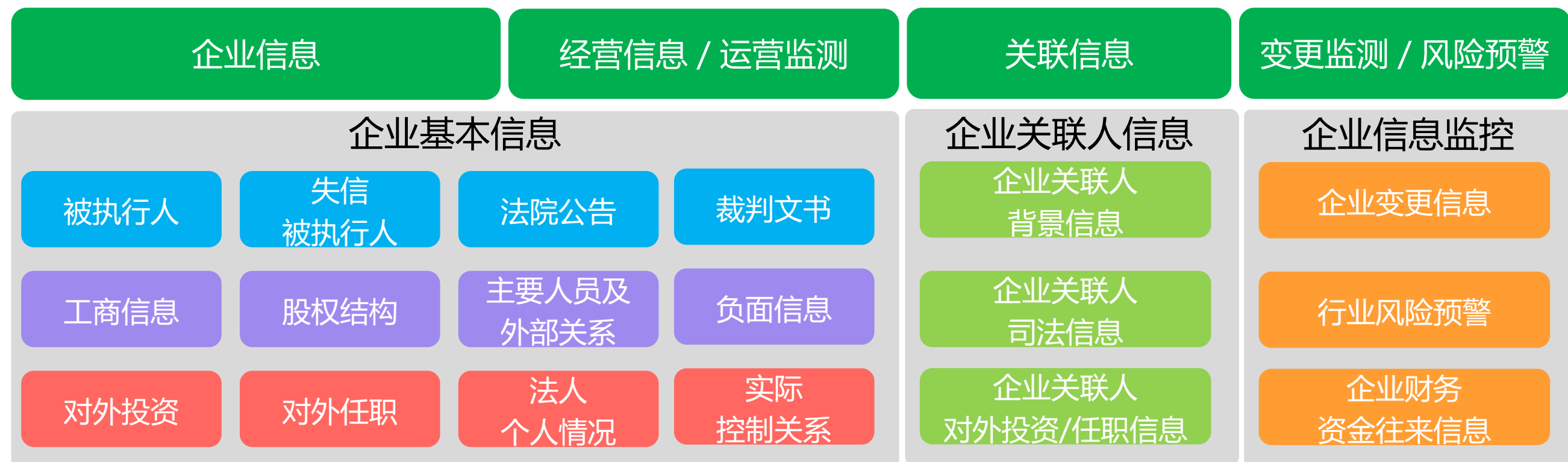
## 2. 授信、放款、撮合系统及贷后管理

- 1** 信用额度在一个月之内有效
  - 用户在有效期内，随时按需提款
  - 支持补充授信 / 循环授信
- 2** P2P撮合系统实时对接资金端
  - 分钟级完成资金撮合
  - 全部在银行托管，安全放心
- 3** 还款方式灵活多样
  - 先息后本/等额本息/随借随还

- 4** 定期信用重评估：提额或降额
  - 重点关注经营表现，根据一系列行为评分决定



# 3. 企业数据整合平台



- 利用**互联网数据**来替代和补充对传统数据的依赖；
- 提高信审效率，保证数据的真实性；
- 提升对决策数据使用的规模性和扩展性。

# 4. 订单区块链

- 1** 解决去中心化的信用问题
  - 让主要参与方成为节点
  - 每个环节形成不可篡改记录，并完整回溯
- 2** 节约成本，缩短流程，效率提升
  - 通过数据的记录、传递、核实及分析，形成信用
  - 降低交易结算风险
- 3** 基于Hyperledger
  - 以go语言开发，并提供多种语言的API
  - 比较充分发挥了go的性能，对Docker容器友好

# 此外，商通贷的大数据应用，还有很多：

- 获取数据：内部，第三方，客户授权，网上公开信息
- 了解客户：客户画像
- 精准获客：网络抓取，自动定制化营销配合线下转化
- 反欺诈：身份，行为，社交
- 风险评估：大数据信用评分模型，收益覆盖风险
- 授信模型：客户需求，还款能力
- 贷后管理：行为评分，失联查询
- ...

# 五、总结

- 金融业是世界上门槛最低的行业，不需要任何技能，只需要有钱就行；有的在短期内跑得快，有的在长期内活得久，管理好风险就行；
- 互联网信贷是一门数据的生意；而且，对于个人和小微企业的业务，数据关注维度非常不一样；
- 互联网信贷，既不是量越大越好，也不是风险越低越好；
- **把握 业务规模 与 金融风险 的平衡。**



关注QCon微信公众号，  
获得更多干货！

# Thanks!



主办方 **Geekbang** > **InfoQ**  
极客邦科技