

# 姨搜

信贷场景下全流程数据风控体系

侯松

2016年10月28日

# 大纲

1. 我们是谁
2. 理解风控
3. 理解用户
4. 风控系统的架构设计



# 1. 我们是谁

- 宜信：超十年的稳健经营历史，世界上规模最大的P2P金融公司之一。
- 姨搜：为整个宜信公司提供风控数据服务和模型服务，包括风控搜索引擎，大规模知识图谱，决策引擎，离线分析建模平台等。
- 本人：负责姨搜大数据风控的各种数据产品和基础设施，协助宜信各产品线和业务方提升风控水平，降低成本。



# 大纲

1. 我们是谁
2. 理解风控
3. 理解用户
4. 风控系统的架构设计

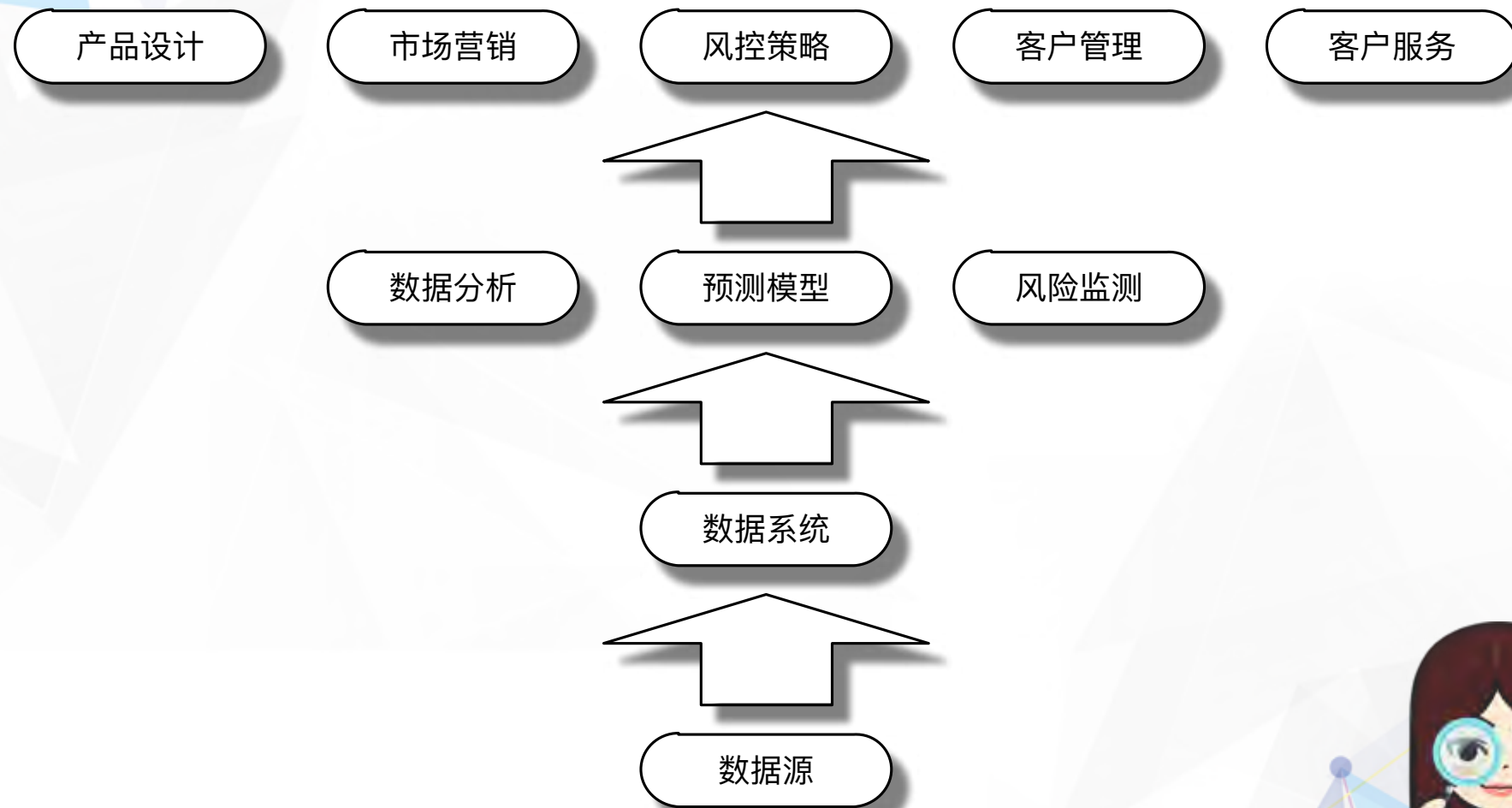


## 2. 理解风控

1. 信贷产品的组成部分
2. 什么数据是有用的数据
3. 个人借款风险的原因
4. 数据与风险的关系



## 2.1 信贷产品的组成部分



## 2.2 什么数据是有用的数据

- 可观查到并记录下来的都是数据。
- 被动收集与主动收集。
- 不同的数据有不同的价值。





## 2.3 个人借款风险

- 风险包括：欺诈风险和信用风险
- 欺诈风险：
  - 第三方欺诈：
  - 主动赖账：
- 信用风险：
  - 财务收入：
  - 支出管理：
  - 不良嗜好：





## 2.4 数据与风险的关系

- 统计模型适合个人信贷决策。
- 相关性比因果关系更重要。
- 数据整理 -> 数据整合 -> 数据特性
- 原始裸数据 -> 简单特征变量 -> 复杂特征变量
  - 信用卡原始交易记录（时间、地点、商家、金额等）
  - 简单变量（不同品类、时间段、地点的消费金额和频次）
  - 简单特征变量的组合特征（在深夜购买游戏点卡的消费金额和频次，等等）
- 模型应用。



# 大纲

1. 我们是谁
2. 理解风控
3. 理解用户
4. 风控系统的架构设计



# 3. 理解用户

1. 风控政策人员
2. 数据科学家
3. 产品开发工程师



# 3.1 风控政策人员

- 特点
  - 对金融市场的风险有丰富经验。
  - 有一定统计分析的背景，复杂算法的能力不足。
  - 对负责的产品理解深刻，但对其他产品认识有限。
- 需求
  - 灵活方便的控制风控政策的执行。
  - 执行历史的BI报表和报警机制。
  - 简单高效的政策分析和回测。
  - 研究成果的跨团队共享。



## 3.2 数据科学家

- 特点
  - 对数据挖掘和机器学习算法有深入的理解。
  - 对数据有很好的感觉。
  - 有一定的编程能力。
- 需求
  - 完整清晰的数据定义和数据流。
  - 高性能的分布式计算集群和基础设施。
  - 多人协作以及跨团队协作的能力。
  - 能轻松将研究成果应用到业务系统中。



## 3.3 产品开发工程师

- 特点
  - 专业的系统开发能力。
  - 开发任务排期紧张。
  - 关注系统稳定性和性能。
- 需求
  - 对接工作尽量简单。
  - 有完善的文档和测试环境。
  - 有系统级别的监控报警。
  - 有专人负责对接工作和问题联调。



# 大纲

1. 我们是谁
2. 理解风控
3. 理解用户
4. 风控系统的架构设计



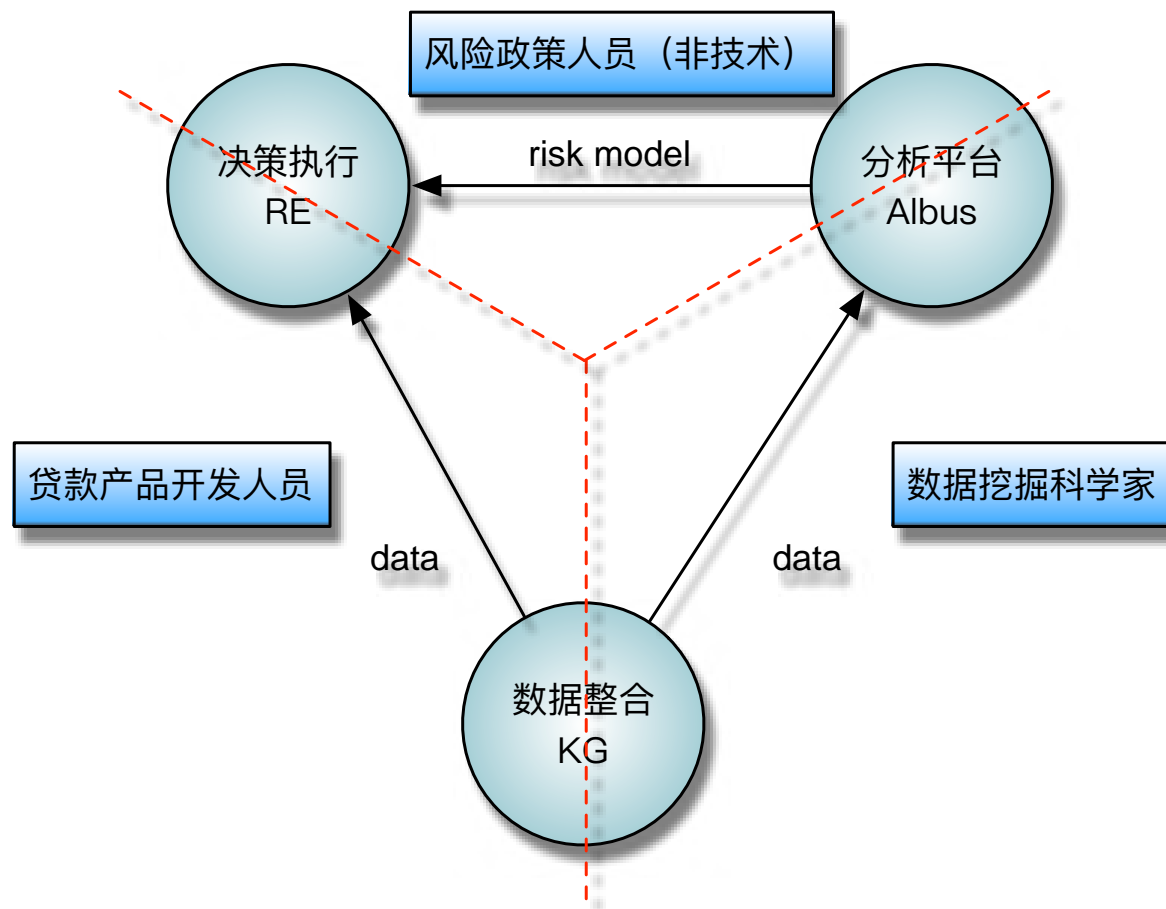


# 4. 风控系统的架构设计

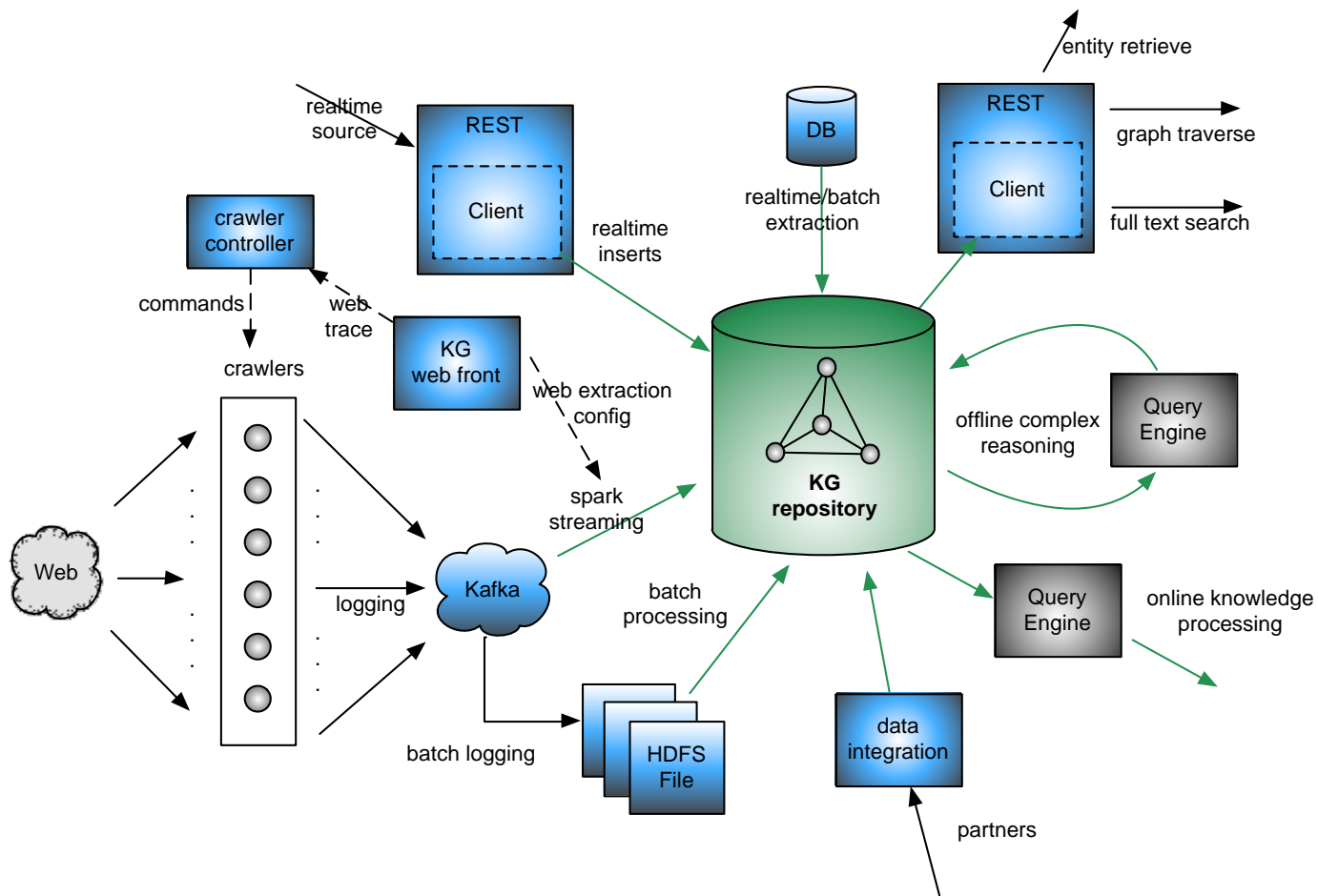
1. 总体结构
2. 数据整合部分 – 知识图谱
3. 政策执行部分 – 决策引擎
4. 实验分析部分 – ALBUS



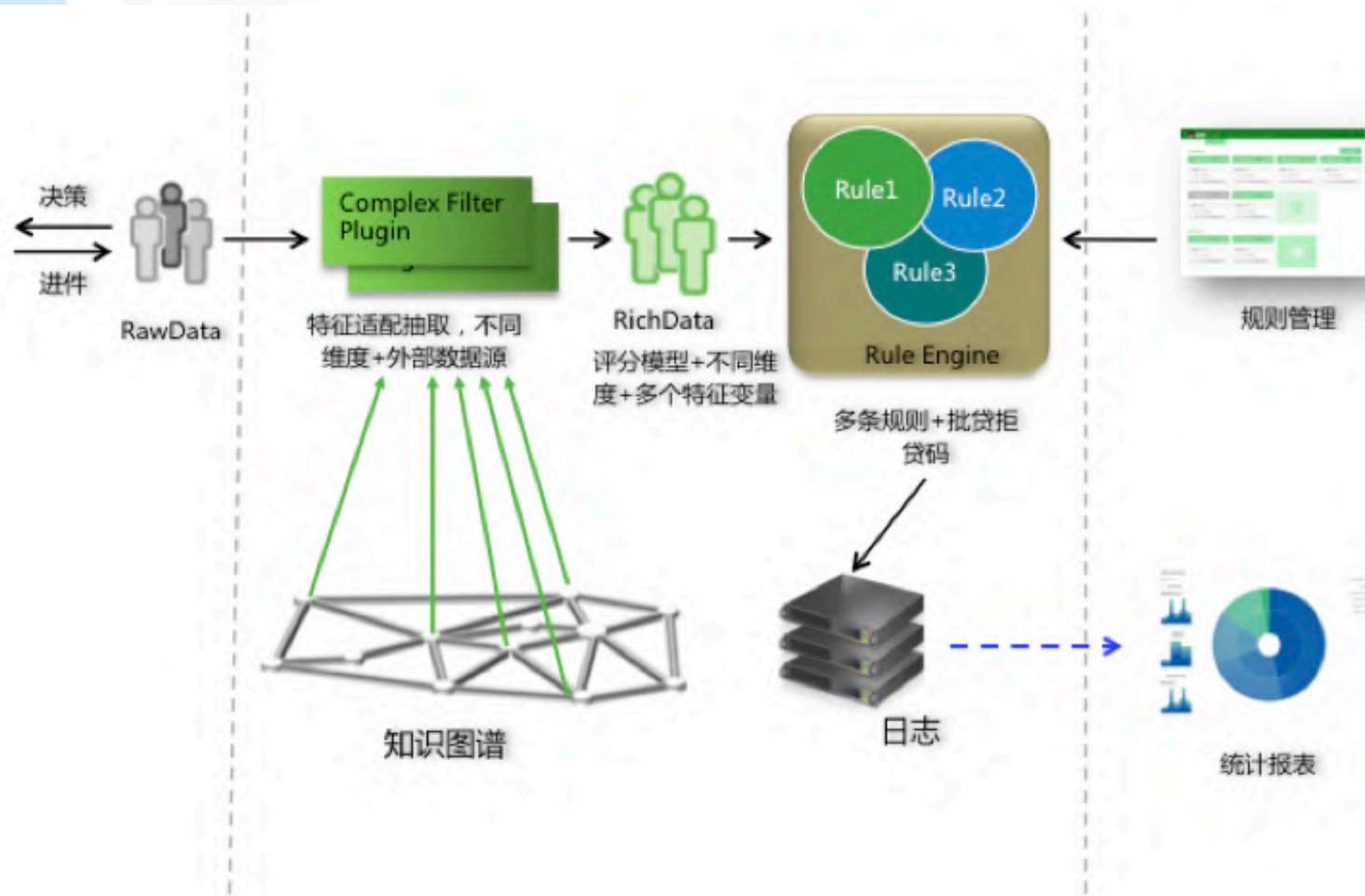
# 4.1 总体结构



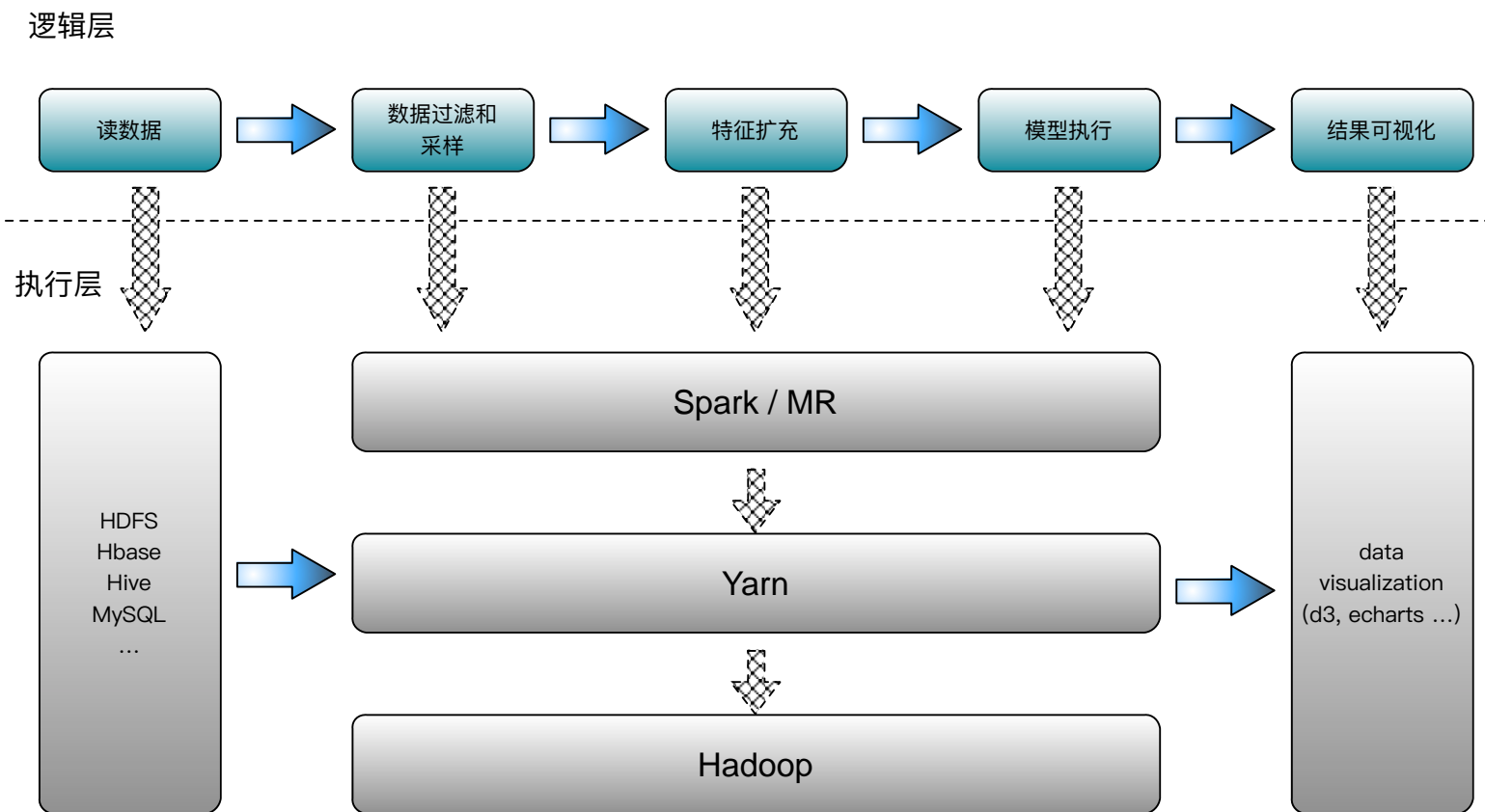
# 4.2 数据整合部分 - 知识图谱



# 4.3 政策执行部分 – 决策引擎



# 4.4 实验分析部分 – ALBUS



# 大纲

1. 我们是谁
2. 理解风控
3. 理解用户
4. 风控系统的架构设计





# 姨搜

## 信贷场景下全流程数据风控体系



侯松

songhou@creditease.cn

2016年10月28日