1.背景

互联网的迅速发展，为电子商务兴起提供了肥沃的土壤。2014年，中国电子商务市场交易规模达到13.4万亿元，同比增长31.4%。其中，B2B电子商务市场交易额达到10万亿元，同比增长21.9%。这一连串高速增长的数字背后，不法分子对互联网资产的觊觎，针对电商行业的恶意行为也愈演愈烈，这其中，最典型的就是黄牛抢单囤货和商家恶意刷单。黄牛囤货让广大正常用户失去了商家给予的优惠让利；而商家的刷单刷好评，不仅干扰了用户的合理购物选择，更是搅乱了整个市场秩序。

京东作为国内电商的龙头企业，在今天遭受着严酷的风险威胁。机器注册账号、恶意下单、黄牛抢购、商家刷单等等问题如果不被有效阻止，会给京东和消费者带来难以估量的损失

互联网行业中，通常使用风控系统抵御这些恶意访问。在技术层面上来讲，风控领域已逐渐由传统的“rule-base”（基于规则判断）发展到今天的大数据为基础的实时+离线双层识别。Hadoop，Spark等大数据大集群分布式处理框架的不断发展为风控技术提供了有效的支撑。

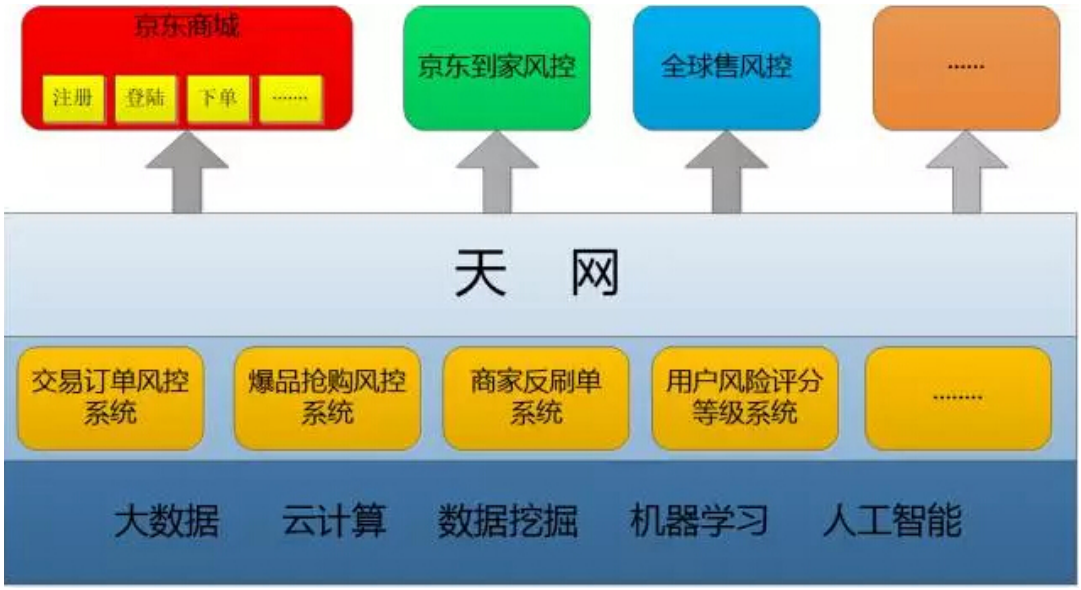
2.什么是“天网”

在此背景下，京东风控部门打造“天网”系统，在经历了多年沉淀后，“天网”目前已全面覆盖京东商城数十个业务节点并有效支撑了京东集团旗下的京东到家及海外购风控相关业务，有效保证了用户利益和京东的业务流程。

“天网“作为京东风控的核心利器，目前搭建了风控专用的基于spark的图计算平台，主要分析维度主要包括：用户画像，用户社交关系网络，交易风险行为特性模型。

其系统内部既包含了面向业务的交易订单风控系统、爆品抢购风控系统、商家反刷单系统，在其身后还有存储用户风险信用信息及规则识别引擎的风险信用中心（RCS）系统，专注于打造用户风险画像的用户风险评分等级系统。

为中小微企业提供一站式金融服务解决方案！



下面，我们将从用户可以直接感知的前端业务风控系统和后台支撑系统两部分对天网进行剖析：

3.前端业务风控系统

1交易订单风控系统

交易订单风控系统主要致力于控制下单环节的各种恶意行为。该系统根据用户注册手机，收货地址等基本信息结合当前下单行为、历史购买记录等多种维度，对机器刷单、人工批量下单以及异常大额订单等多种非正常订单进行实时判别并实施拦截。

目前该系统针对图书、日用百货、3C产品、服饰家居等不同类型的商品制定了不同的识别规则，经过多轮的迭代优化，识别准确率已超过99%。对于系统无法精准判别的嫌疑订单，系统会自动将他们推送到后台风控运营团队进行人工审核，运营团队将根据账户的历史订单信息并结合当前订单，判定是否为恶意订单。从系统自动识别到背后人工识别辅助，能够最大限度地保障订单交易的真实有效性。

2 爆品抢购风控系统

在京东电商平台，每天都会有定期推出的秒杀商品，这些商品多数来自一线品牌商家在京东平台上进行产品首发或是爆品抢购，因此秒杀商品的价格会相对市场价格有很大的优惠力度。

但这同时也给黄牛带来了巨大的利益诱惑，他们会采用批量机器注册账号，机器

抢购软件等多种形式来抢购秒杀商品，数量有限的秒杀商品往往在一瞬间被一抢而空，一般消费者却很难享受到秒杀商品的实惠。针对这样的业务场景，秒杀风控系统这把利剑也就顺势而出。

在实际的秒杀场景中，其特点是瞬间流量巨大。即便如此，“爆品抢购风控系统”这把利剑对这种高并发、高流量的机器抢购行为显示出无穷的威力。目前，京东的集群运算能力能够到达每分钟上亿次并发请求处理和毫秒级实时计算的识别引擎能力，在秒杀行为中，可以阻拦98%以上的黄牛生成订单，最大限度地为正常用户提供公平的抢购机会。

3商家反刷单系统

随着电商行业的不断发展，很多不轨商家尝试采用刷单、刷评价的方式来提升自己的搜索排名进而提高自家的商品销量。随着第三方卖家平台在京东的引入，一些商家也试图钻这个空子，我们对此类行为提出了 “零容忍”原则，为了达到这个目标，商家反刷单系统也就应运而生。

商家反刷单系统利用京东自建的大数据平台，从订单、商品、用户、物流等多个维度进行分析，分别计算每个维度下面的不同特征值。通过发现商品的历史价格和订单实际价格的差异、商品SKU销量异常、物流配送异常、评价异常、用户购买品类异常等上百个特性， 结合贝叶斯学习、数据挖掘、神经网络等多种智能算法进行精准定位。

而被系统识别到的疑似刷单行为，系统会通过后台离线算法，结合订单和用户的信息调用存储在大数据集市中的数据进行离线的深度挖掘和计算，继续进行识别，让其无所遁形。而对于这些被识别到的刷单行为，商家反刷单系统将直接把关联商家信息告知运营方做出严厉惩罚，以保证消费者良好的用户体验。

前端业务系统发展到今天，已经基本覆盖了交易环节的全流程，从各个维度打击各种侵害消费者利益的恶意行为。

4.后台支撑系统

天网作为京东的风控系统，每天都在应对不同特性的风险场景。它可能是每分钟数千万的恶意秒杀请求，也可能是遍布全球的黄牛新的刷单手段。天网是如何通过底层系统建设来解决这一个又一个的难题的呢？让我们来看一看天网的两大核心系统：风险信用服务(RCS)和风控数据支撑系统（RDSS）。

1风险信用服务

风险信用服务（RCS）是埋藏在各个业务系统下的风控核心引擎，它既支持动态规则引擎的高效在线识别，又是打通沉淀数据和业务系统的桥梁。它是风控数据

层对外提供服务的唯一途径，重要程度和性能压力不言而喻。

1.1 RCS的服务框架

RCS作为天网对外提供风控服务的唯一出口，其调用方式依赖于京东自主研发的服务架构框架JSF，它帮助RCS在分布式架构下提供了高效RPC调用、高可用的注册中心和完备的容灾特性，同时支持黑白名单、负载均衡、Provider动态分组、动态切换调用分组等服务治理功能。

面对每分钟千万级别的调用量，RCS结合JSF的负载均衡、动态分组等功能，依据业务特性部署多个分布式集群，按分组提供服务。每个分组都做了跨机房部署，最大程度保障系统的高可用性。

1.2 RCS动态规则引擎的识别原理

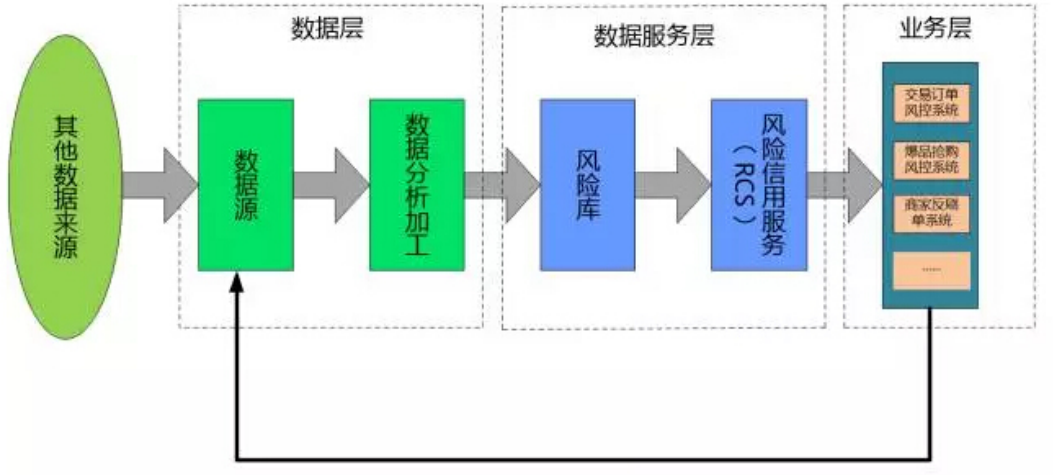
RCS内部实现了一套自主研发的规则动态配置和解析的引擎，用户可以实时提交或者修改在线识别模型。当实时请求过来时，系统会将实时请求的数据依据模型里的核心特性按时间分片在一个高性能中间件中进行高性能统计，一旦模型中特性统计超过阀值时，前端风控系统将立刻进行拦截。

而前面我们所说的高性能中间件系统就是JIMDB，它同样是自主研发的，主要功能是基于Redis的分布式缓存与高速Key/Value存储服务，采用“Pre-Sharding”技术，将缓存数据分摊到多个分片（每个分片上具有相同的构成，比如：都是一主

一从两个节点）上，从而可以创建出大容量的缓存。支持读写分离、双写等I/O策略，支持动态扩容，还支持异步复制。在RCS的在线识别过程中起到了至关重要的作用

1.3 RCS的数据流转步骤

风险库是RCS的核心组件，其中保存有各种维度的基础数据，下图是整个服务体系中的基本数据流转示意图：



1) 各个前端业务风控系统针对各个业务场景进行风险识别，其结果数据将回流至风险库用户后续离线分析及风险值判定。

2) 风险库针对业务风控识别进过数据进行清洗，人工验证，定义并抽取风控指标数据，经过此道工序风险库的元数据可以做到基本可用。

3) 后台数据挖掘工具对各来源数据，依据算法对各类数据进行权重计算，计算结果将用于后续的风险值计算。

4) 风险信用服务一旦接收到风险值查询调用，将通过在JIMDB缓存云中实时读取用户的风控指标数据，结合权重配置，使用欧式距离计算得出风险等级值，为各业务风控系统提供实时服务。

1.4 RCS的技术革新与规划

进入2015年以后，RCS系统面临了巨大的挑战。首先，随着数据量的不断增大，之前的处理框架已无法继续满足需求，与此同时不断更新的恶意行为手段对风控的要求也越来越高，这也就要求风控系统不断增加针对性规则，这同样带来不不小的业务压力。

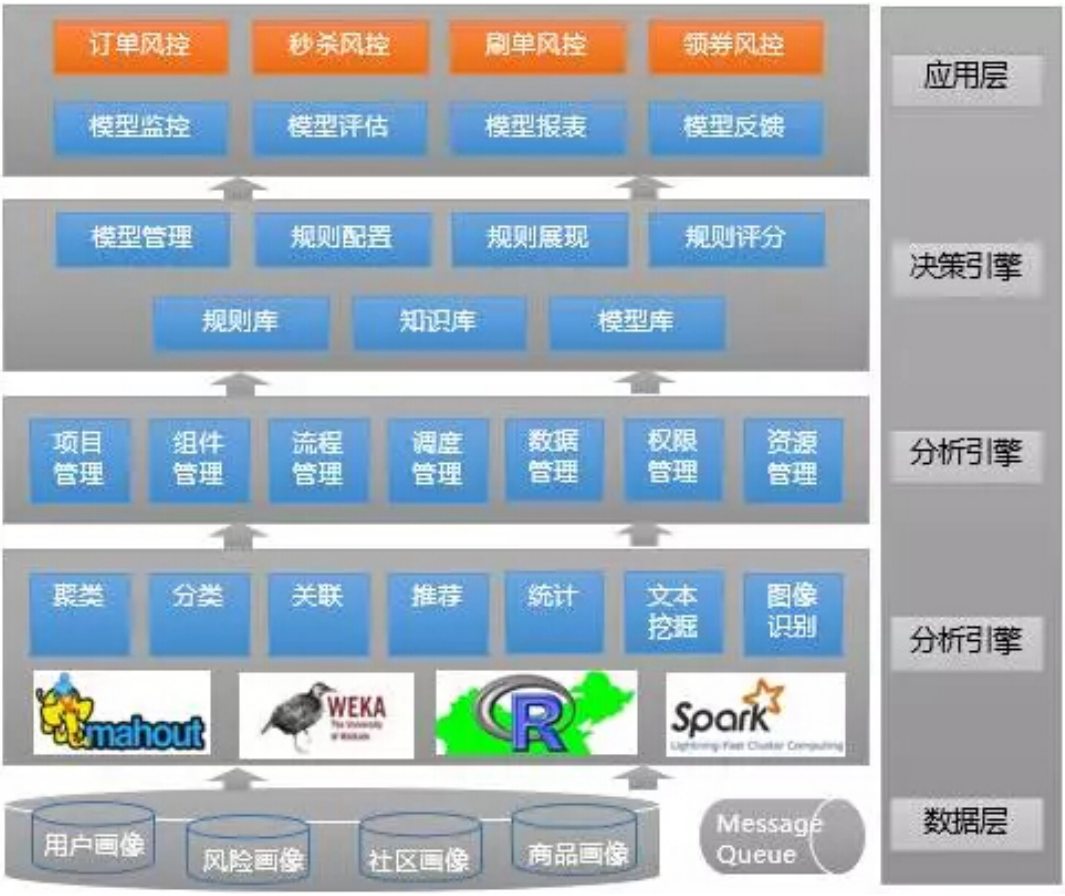
面对这样的挑战，RCS更加密切地加强了和京东大数据平台的合作。在实时识别数据的存储方面，面对每天十几亿的识别流水信息，引入了Kafka+Presto的组

合。通过Presto对缓存在Kafka一周之内的识别数据进行实时查询。超过1周的数据通过ETL写入Presto的HDFS，支持历史查询。在RCS识别维度提升方面，目前已经与京东用户风险评分等级系统打通流程，目前已拿到超过1亿的基于社交网络维度计算的风险等级，用于风险信用识别。在风险等级的实时计算方面，已经逐步切换到大数据部基于Strom打造的流式计算计算平台JRC。

5.风控数据支撑系统

风控数据支撑系统是围绕着京东用户风险评分等级系统搭建起来的一整套风控数据挖掘体系。

1RDSS的核心架构



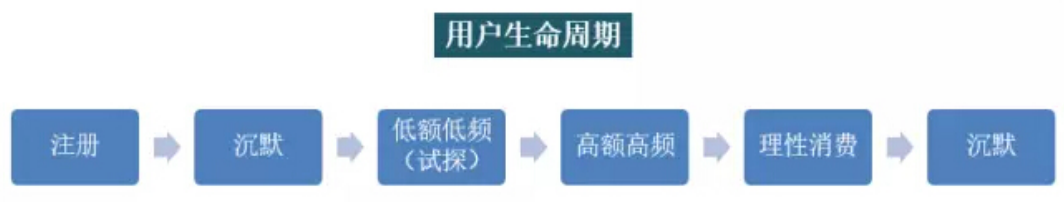
1) 数据层

如图所示，数据层负责数据的抽取、清洗、预处理。目前ETL程序通过JMQ、Kafka、数据集市、基础信息接口、日志接入了超过500个生产系统的业务数据，其中包括大量的非结构化数据。通过对数据的多样性、依赖性、不稳定性进行处理，最终输出完整的、一致性的风控指标数据，并通过数据接口提供给算法引擎层调用。这一层最关键的部分是在对风控指标数据的整理。指标数据质量的好坏

直接关联到系统的最终输出结果。目前指标的整理主要从以下三个维度开展：

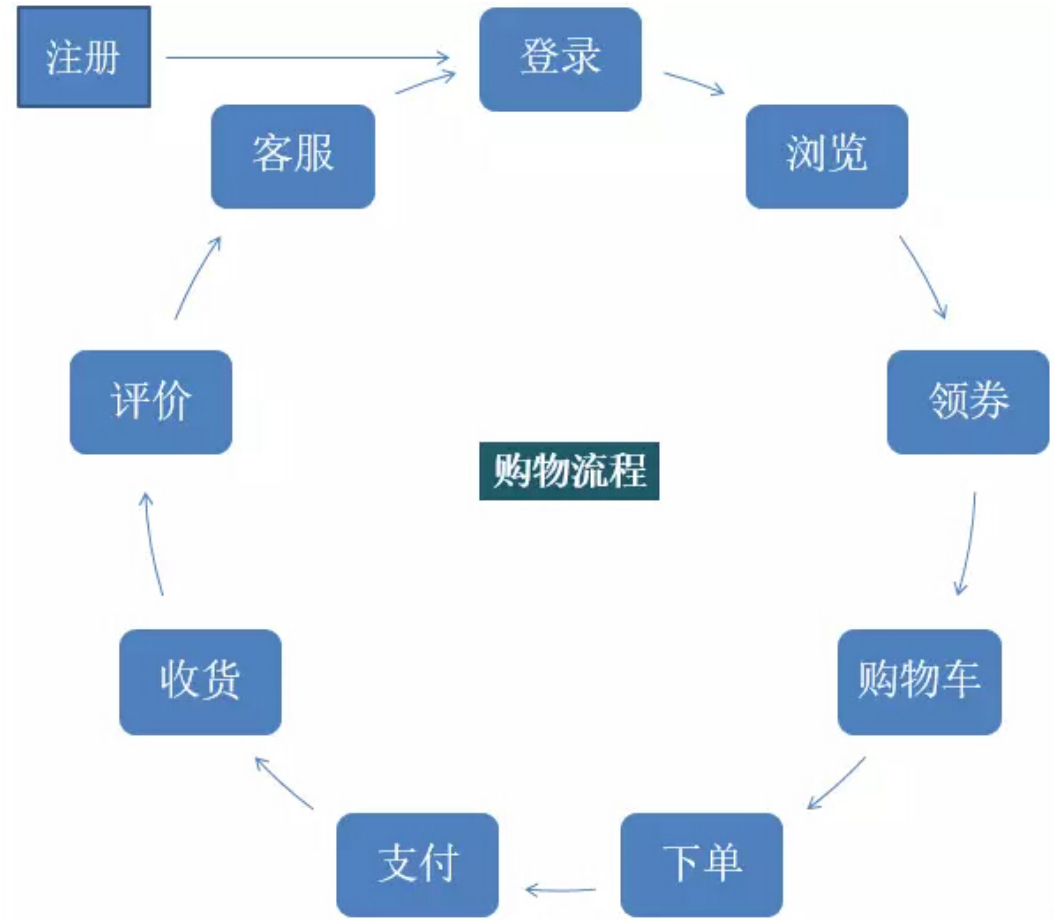
a) 基于用户生命周期的指标数据整理

对于电商业务而言，一个普通用户基本上都会存在以下几种粘性状态，从尝试注册，到尝试购买；从被深度吸引，到逐渐理性消费。每一种状态总是伴随着一定的消费特征，而这些特征也将成为我们捕获用户异常行为的有利数据。



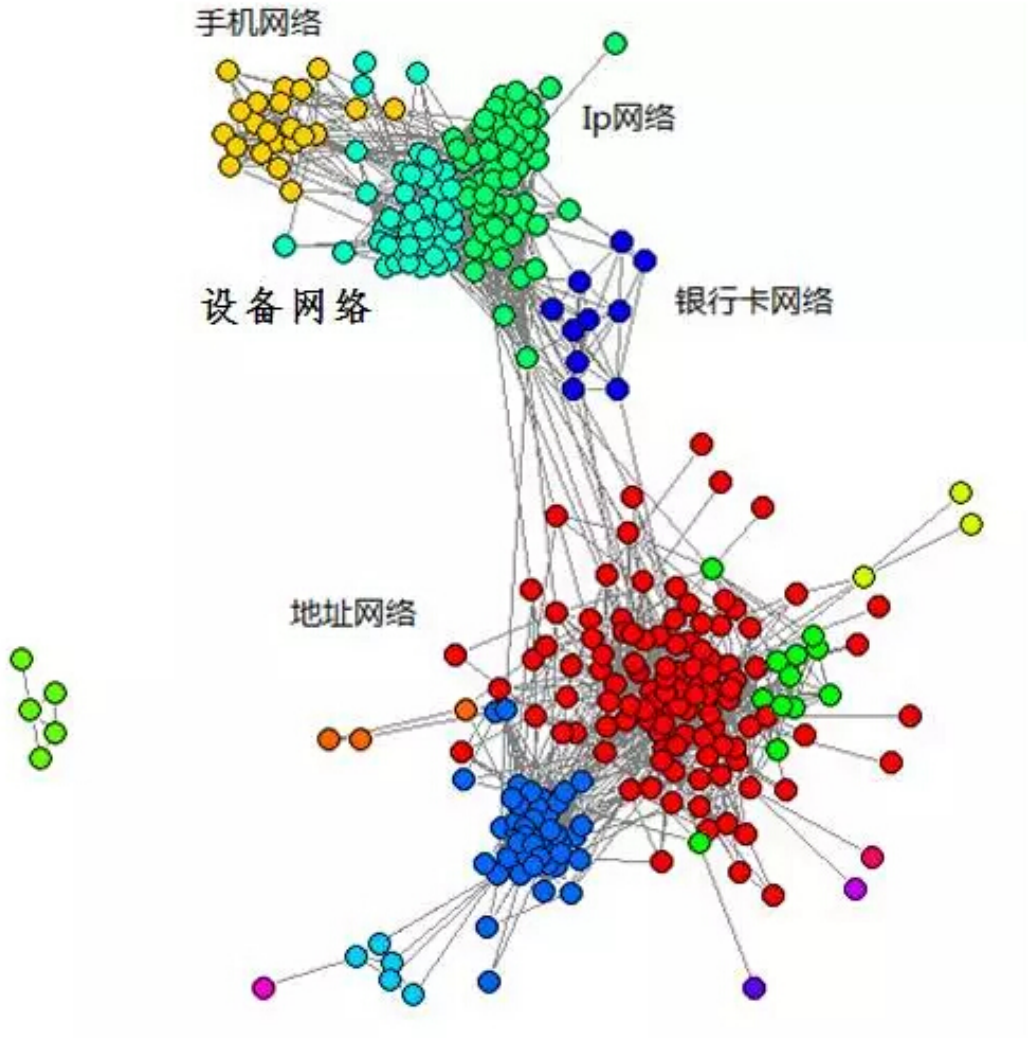
b) 基于用户购买流程的风控指标数据整理

对于一般用户而说，其购买习惯具有相当的共性，例如，通常都会对自己需求的商品进行搜索，对搜索结果中自己感兴趣的品牌进行浏览比较，几经反复才最终做出购买决定。在真正购买之前还要找一下相关的优惠券，在支付过程中也会或多或少有些停顿。而对于黄牛来说，他们目标明确，登录之后直奔主题，爽快支付，这些在浏览行为上的差异也是我们寻找恶意用户的有利数据。



c) 基于用户社交网络的风控指标数据整理

基于用户社交网络的指标数据是建立在当前风控领域的黑色产业链已经逐渐成体系的背景下的。往往那些不怀好意的用户总会在某些特征上有所聚集，这背后也就是一家家黄牛，刷单公司，通过这种方式可以实现一个抓出一串，个别找到同伙的效果。



2) 算法引擎层

算法引擎层集合了各种数据挖掘算法，在系统内被分门别类的封装成各种常用的分类、聚类、关联、推荐等算法集，提供给分析引擎层进行调用。

3) 分析引擎层

分析引擎层是风控数据分析师工作的主要平台，数据分析师可以在分析引擎层依据业务设立项目，并且在平台上开展数据挖掘全流程的工作，最终产出风控模型和识别规则。

4) 决策引擎层

决策引擎层负责模型和规则的管理，所有系统产出的模型及规则都集合在这里进行统一管理更新。

5) 应用层

应用层主要涵盖了决策引擎层产出模型和规则的应用场景，这里最重要的就是风

险信用服务(RCS)，其主要职能是对接底层数据，对外层业务风控系统提供风险识别服务。

而在模型和规则投入使用之前必须要经过我们另外一个重要的系统也就是风控数据分析平台（FBI），因为所有的模型和规则都先将在这个平台中进行评估，其输入就是所有规则和模型的产出数据，输出就是评估结果，评估结果也将反馈到决策引擎层来进行下一步的规则，模型优化。

2RDSS之用户风险评分等级系统

京东用户风险评分等级系统是天网数据挖掘体系孵化出的第一个数据项目。其主要目的在于将所有的京东用户进行分级，明确哪些是忠实用户，哪些又是需要重点关注的恶意用户。其实现原理是依赖前面所描述的社交关系网络去识别京东用户的风险程度。而这种方式在整个数据领域来说都是属于领先的。京东用户风险评分等级系统一期已经产出1亿数据，目前已经通过RCS系统对外提供服务。根据识别结果评估，识别忠实用户较RCS风险库增加37%，识别的恶意用户较RCS风险库增加10%。

目前，京东用户风险评分等级系统已经实现：

1) 数据层基于社交网络的维度产出50余个风险指标。

2) 通过PageRank、三角形计数、连通图、社区发现等算法进行点、边定义，并识别出数十万个社区网络。

3) 通过经典的加权网络上的能量传播思想，计算上亿用户的风险指数。

5.结语

凡是过去，皆为序曲，京东风控正在打造一套数据定义一切的超级风控计算框架。这套风控框架将统一风控模型管理(数据模型，识别模型，规则引擎)、统一风控服务管理(JRC，PRESTO，Streaming)、统一风控数据管理(HDFS，HBASE，Kafka)，并将横跨云计算、大数据、人工智能，针对瞬息万变的电商交易风险智能调整风控策略实时处理。

关于作者

张帅，京东成都研究院高级研发工程师，毕业于西华大学，2012年加入京东风控研发部，参与多个风控业务和数据核心系统的研发。

陈诚，京东成都研究院数据产品经理，四川大学硕士，参与多个风控天网系统和数据相关业务系统的研发。

孟勐，京东成都研究院高级经理，电子科技大学硕士，主要负责京东风控天网系统后台和数据处理、数据挖掘、决策支持等相关业务系统研发。

感谢杜小芳对本文的审校。

给InfoQ中文站投稿或者参与内容翻译工作，请邮件至editors@cn.infoq.com。也欢迎大家通过新浪微博（@InfoQ，@丁晓昀），微信（微信号：InfoQChina）关注我们。