# 基于知识图谱的
# 通用数据价值洞察平台

DataExa创始人　洪万福

# 知识图谱

DataExa
Thinking Intelligent

定义：实体及其之间的关系图。

本质：由知识组成的语义网络。

目的：让用户能够更快更简单的发现新的信息和知识。

# 知识图谱应用

搜索引擎、问答系统

金融反欺诈

反恐、情报分析

精准营销

# 知识图谱优势

数据模型： 更适合复杂、互联性、低结构化的数据。

信息检索： 精确度、完整性、关联度、智能化程度更高，互操作和用户体验更好。

分析能力：基于关系和图的分析能力更强。

更好地表征和计算这个多维的世界

# 知识图谱构建&应用

知识获取

知识验证

知识融合


知识利用

# 知识图谱——组成

互联网、移动互联网

引擎

语义分析　可视化构建　SNS分析

推荐引擎　图谱数据　机器学习

规则引擎　图数据库　并行计算

人机交互

可视化适配

分析平台

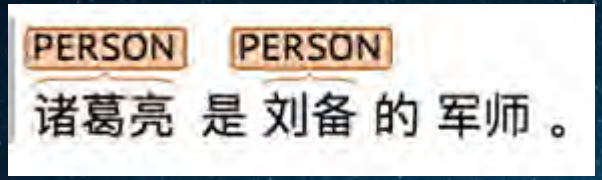
行业数据资源

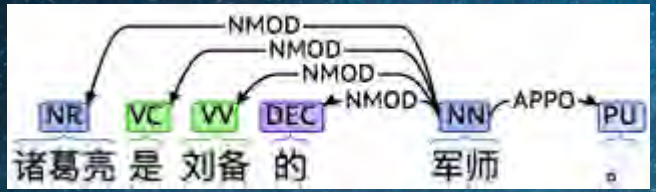
# 知识图谱——知识获取

DataExa
Thinking Intelligent

| 数据源 | 手段 | 知识类型 |
| --- | --- | --- |
| 数据库 | NLP | 实体 |
| 文本、语音 | 规则引擎 | 属性 |
| 图片 | 机器学习 | 标签 |
| 视频 | 人工 | 关系 |

# 知识获取——NLP




- 命名实体识别
- 句法分析
- 文本挖掘（主题模型等）
- 关系提取

# 知识获取——规则

DataExa
Thinking Intelligent

## 正则表达式

```
/^(\w)+(\.\w+)*@(\w)+((\.\w+)+)$/
```

## 规则引擎

```
if user.consumption.total>=200000 then
    user.type='VIP'
```

## 网页信息抽取

XPath： *//div[@class=summary]/p*

CSS： *("div h1 a", "innerHtml")*

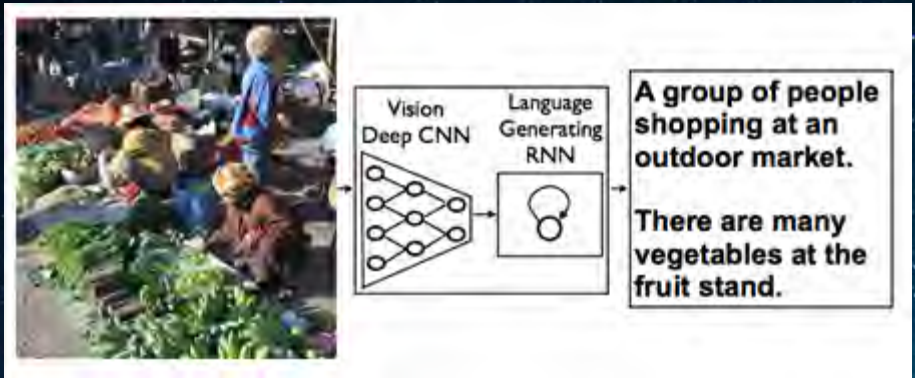
## 混合

NR师从NR => {2}-老师->{1}

# 知识获取——机器学习

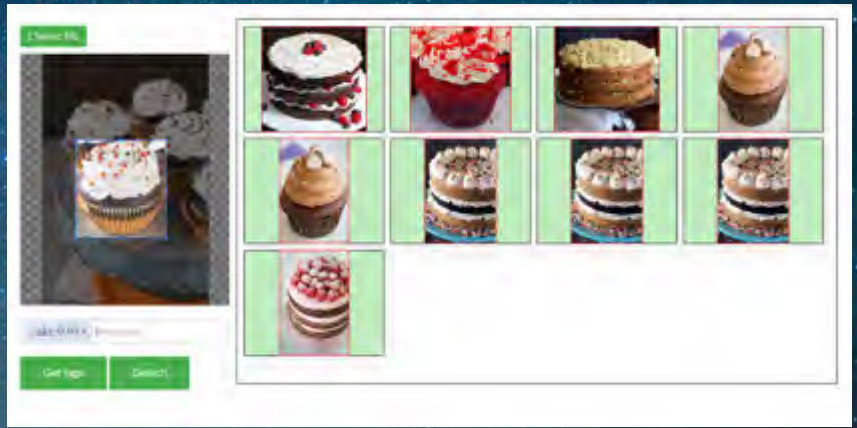
DataExa
Thinking Intelligent

# 知识获取——非结构化




图片内容解析

图片相似度



# 知识获取——词网


同义词
近义词
反义词
词向量
词距离
情感度
……

# 知识获取——Pipeline


# 知识融合

# 知识图谱——图综合分析

DataExa
Thinking Intelligent

四维分析
- 实体
- 关系
- 时间
- 空间

图分析
- 关系挖掘
- 路径提取
- 最短路径分析
- 关联度分析
- 出入度分析
- 社区发现
- 强连通分量
- ……

# 知识图谱——计算引擎

DataExa
Thinking Intelligent

## 自然语言

王菲的前夫的女儿

## 内置函数

relationship(诸葛亮).depth(2).limit(20)

## 图查询

```
m=[:]
g.v(1).out('likes').in('likes').out('likes').groupCount(m)
m.sort{-it.value}
```

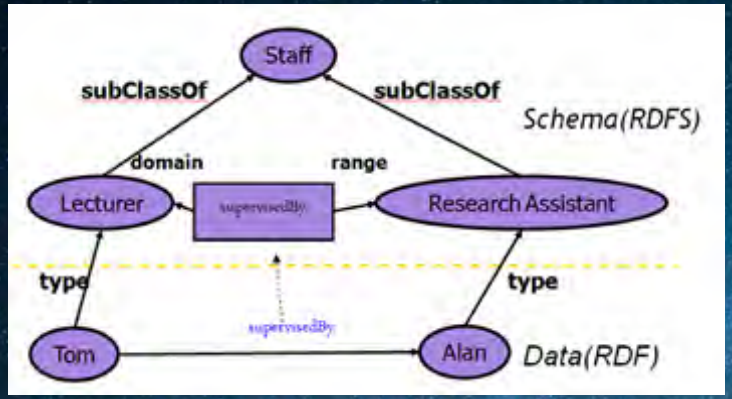
## 推理器

2016年A股主板发生的并购案

# 知识图谱——图谱模型

DataExa
Thinking Intelligent

RDF




PGM

# 知识图谱——数据存储

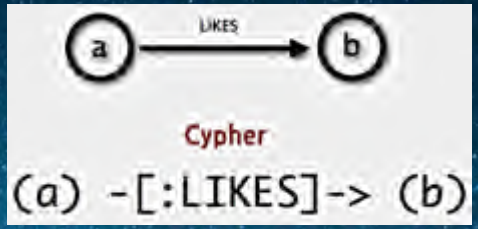
Neo4j

Titan

Jena

# 知识图谱——查询语言






# 知识图谱——逻辑推理

DataExa
Thinking Intelligent

## 机器学习

Representation Learning

## 规则引擎

if user.consumption.total>=200000 then
user.type='VIP'

## 数学运算

order[2015-2016].sum

## 自定义

亲戚关系推理器

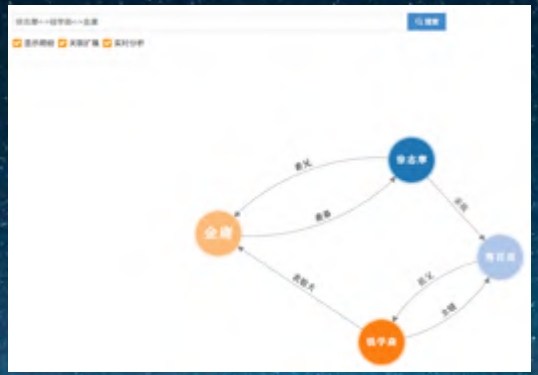
# 知识图谱——可视化适配

DataExa
Thinking Intelligent

根据图谱查询结果进行可视化展示的样式适配

## 实体检索

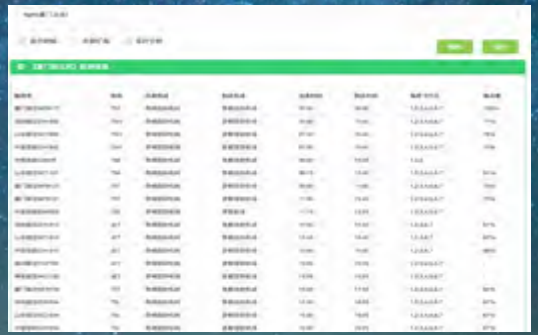
## 关系搜索



## 范围搜索

## 精准问答


## 列表结果



## 图表展示

# DataExa-Sati
# 语义计算平台

DataExa
Thinking Intelligent

## 实用的智能语义计算平台

DataExa-Sati是一个高效实用的语义计算平台，基于自然语言处理、图存储计算、问答系统等技术，提供一套成熟的行业知识图谱构建体系，深度挖掘海量非结构、半结构化数据背后隐藏的价值。

- 基于知识图谱的语义分析；
- 模糊语义识别，智能关联相关内容；
- 支持语境化的人机对话；
- 内置丰富的查询函数，简化复杂的查询逻辑；
- 查询结果输出和可视化适配，智能匹配用户最
- 丰富的数据展示，可视化手段

# DataExa-Sati
# 产品特性

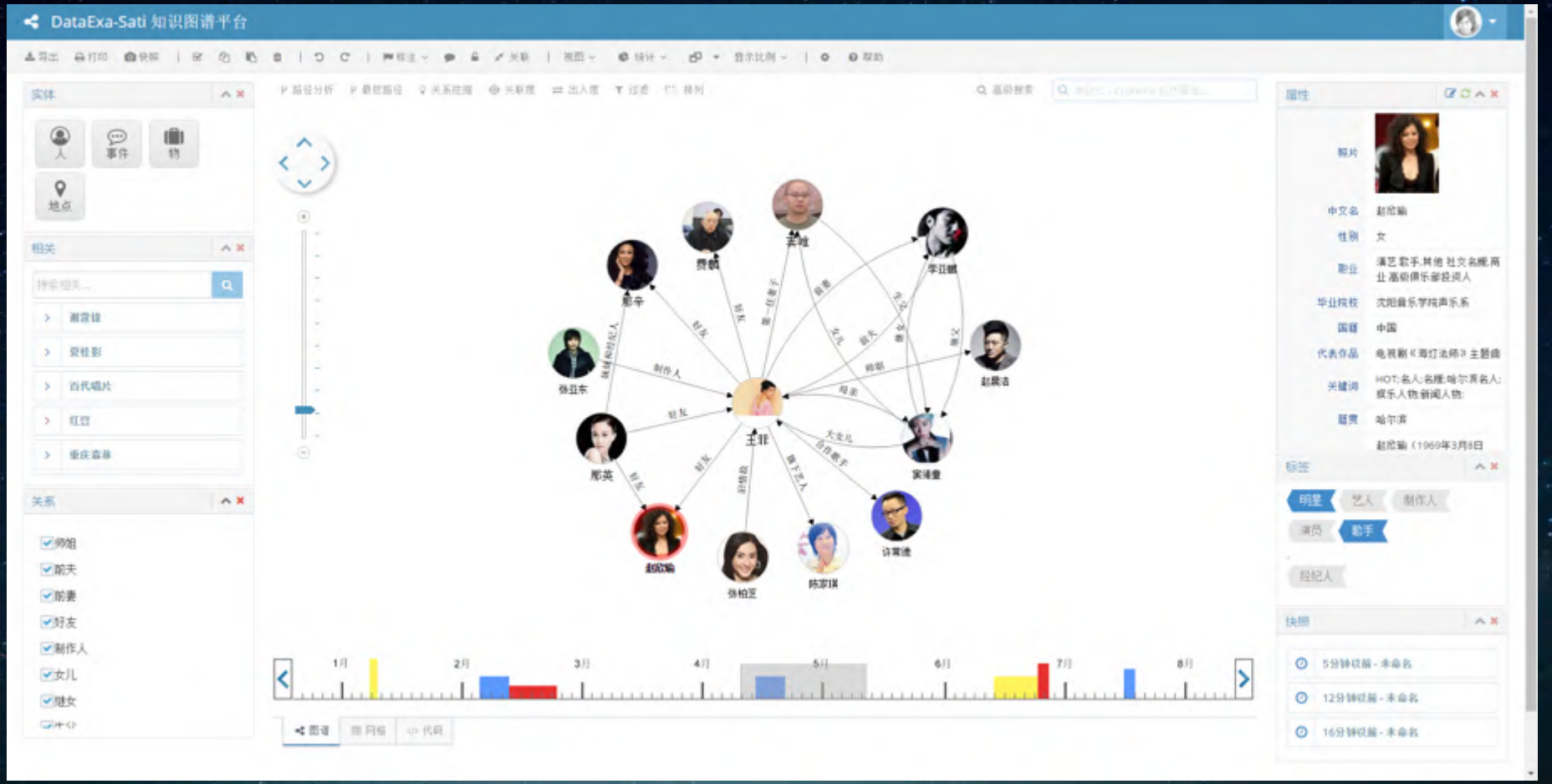
DataExa
Thinking Intelligent

## 特性

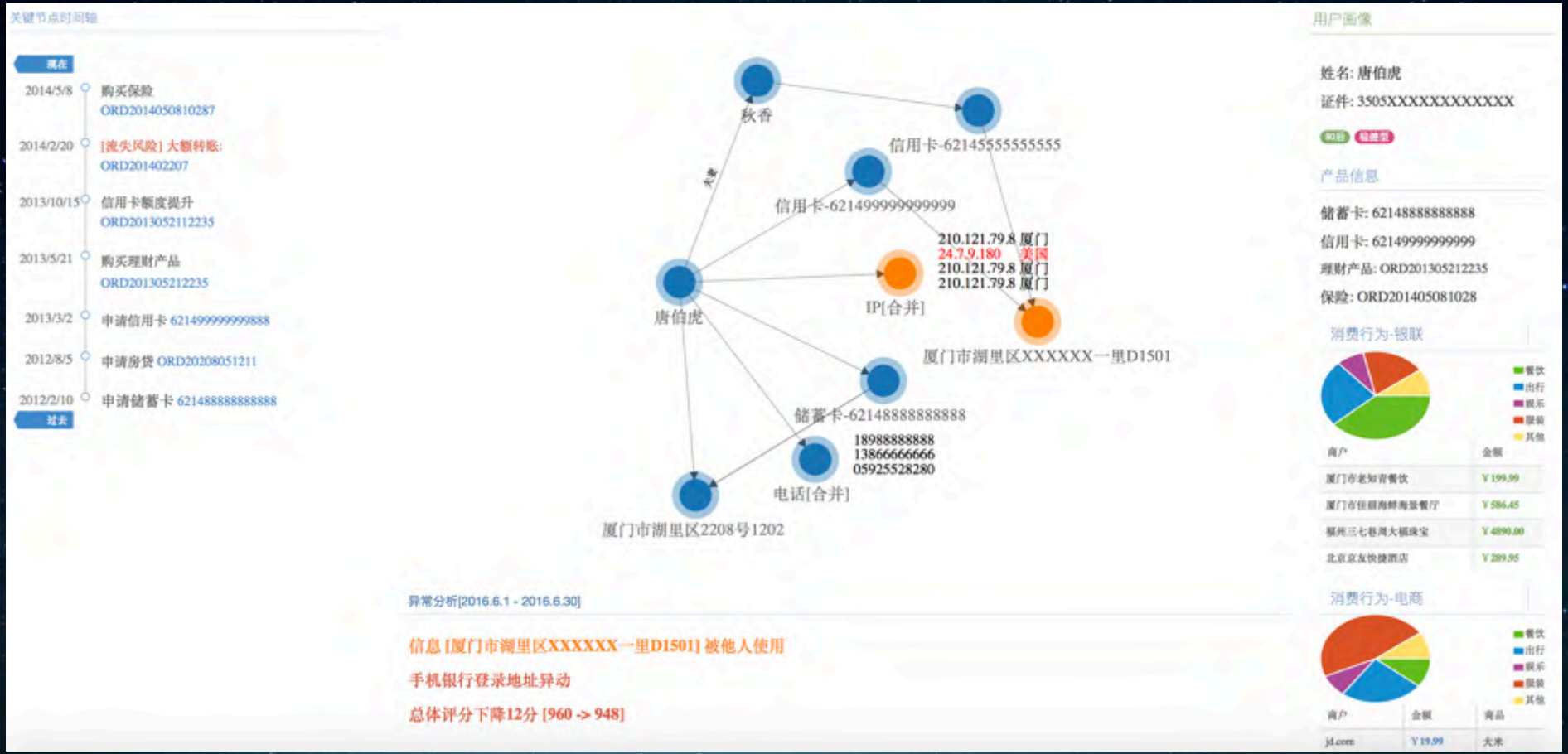
- 支持多源异构知识提取；
- 支持中文和英文；
- 基于图的存储和计算机制；
- 面向领域的词网和图谱自定义管理；
- 模糊语义识别，智能关联相关内容；
- 内置丰富的查询函数，简化复杂的查询逻辑；
- 快速分析数据实体在时间、空间下的关联性和因果性；
- 丰富的数据展示和可视化手段；

# DataExa-Sati 产品模块

# DataExa-Sati一览

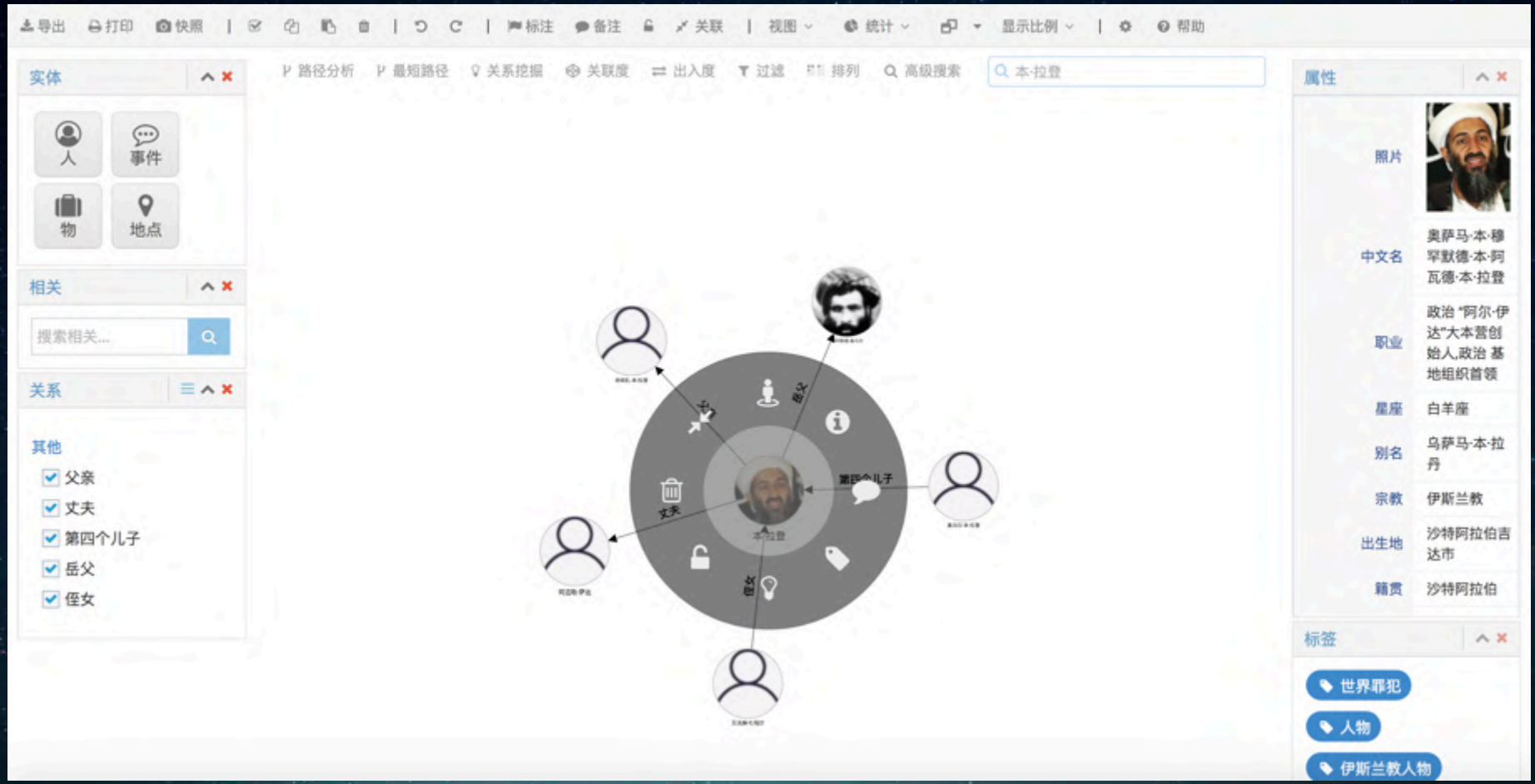


# 行业图谱应用——金融

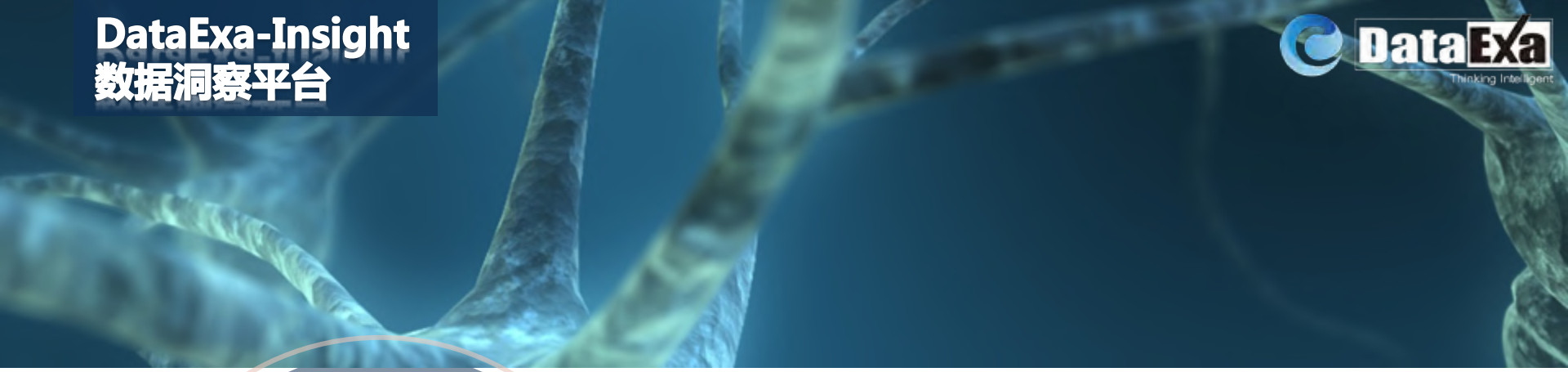

关键节点时间轴
现在
2014/5/8 购买保险 ORD2014050810287
2014/2/20 [流失风险] 大额转账: ORD201402207
2013/10/15 信用卡额度提升 ORD2013052112235
2013/5/21 购买理财产品 ORD201305212235
2013/3/2 申请信用卡 621499999999888
2012/8/5 申请房贷 ORD20208051211
2012/2/10 申请储蓄卡 621488888888888
过去
秋香
信用卡-62145555555555
信用卡-621499999999999
210.121.79.8 厦门
24.7.9.180 美国
210.121.79.8 厦门
210.121.79.8 厦门
IP[合并]
唐伯虎
厦门市湖里区XXXXXX一里D1501
储蓄卡-62148888888888
18988888888
13866666666
05925528280
电话[合并]
厦门市湖里区2208号1202
异常分析[2016.6.1 - 2016.6.30]
信息 [厦门市湖里区XXXXXX一里D1501] 被他人使用
手机银行登录地址异动
总体评分下降12分 [960 -> 948]
用户画像
姓名: 唐伯虎
证件: 3505XXXXXXXXXXX
产品信息
储蓄卡: 62148888888888
信用卡: 62149999999999
理财产品: ORD201305212235
保险: ORD201405081028
消费行为-银联
餐饮
出行
娱乐
服装
其他
商户 金额
厦门市老知青餐饮 ¥ 199.99
厦门市佳丽海鲜海鲜餐厅 ¥ 586.45
福州三七巷周大福珠宝 ¥ 4890.00
北京京友快捷酒店 ¥ 289.95
消费行为-电商
餐饮
出行
娱乐
服装
其他
商户 金额 商品
jd.com ¥ 19.99 大米

# 行业图谱应用——反恐

DataExa
Thinking Intelligent



# DataExa-Insight
# 数据洞察平台



## 企业级的可视化机器（深度）学习平台

DataExa-Insight是一个简单易用的大规模机器学习平台，通过集成行业成熟的机器学习框架，提供可视化建模、配置化参数、流程化操作、模板化任务等功能，大大降低客户进行数据价值洞察的成本。

- 关注业务全局，避免价值孤岛；
- 建立全景业务图谱，深度挖掘潜在价值；
- 无代码、灵活的可视化建模和分析流程设计；
- 强大的分布式智能计算引擎；
- 包括自然语言、推理引擎、机器学习等；
- 丰富的数据展示、可视化手段。

# DataExa-Insight 特性

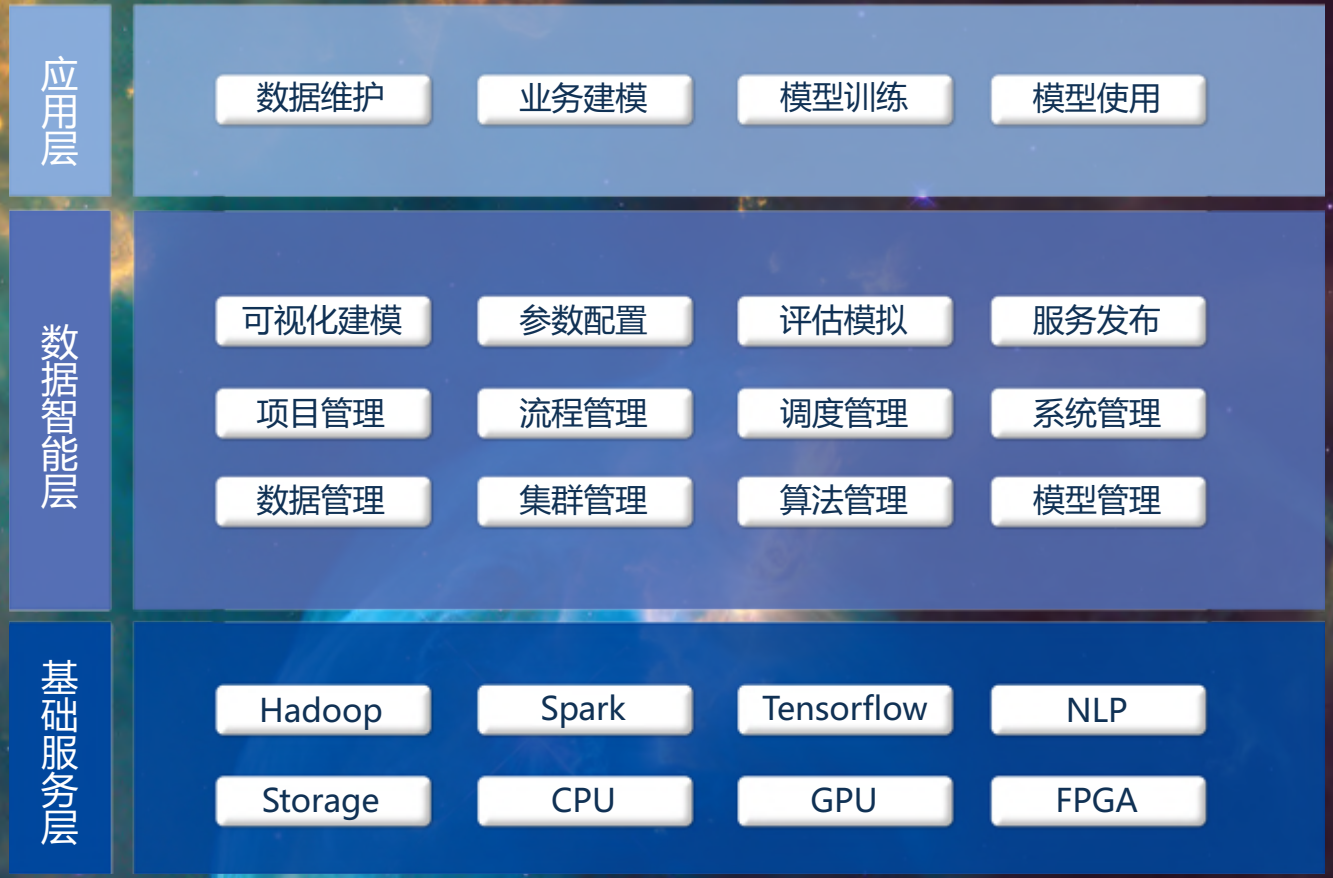
DataExa
Thinking Intelligent

## 特性

- 支持多源异构数据集成；并行化分布式的数据处理组件，支持清理、替换、组合、采样、去重、拆分等数据预处理操作 ；
- 支持Spark、Tensorflow多个机器学习深度学习（或混合）平台；
- 超细粒度的配置化算子；提供多个分类、聚类、回归、主题模型、推荐算法，同时支持前沿的深度学习、在线学习、贝叶斯推荐等算法 ；
- 可视化和模板化的机器学习操作，降低使用难度；提供在线模拟功能 ；
- 支持YARN、Mesos、GPU等计算集群；
- 细粒度的权限体系和安全审核控制；
- 机器学习全过程细粒度监控；
- 多类型的机器学习服务发布组件；

# DataExa-Insight
# 产品模块

DataExa
Thinking Intelligent



# 宇宙的本质是计算。

# 感谢观看，敬请指导!

DataExa
Thinking Intelligent