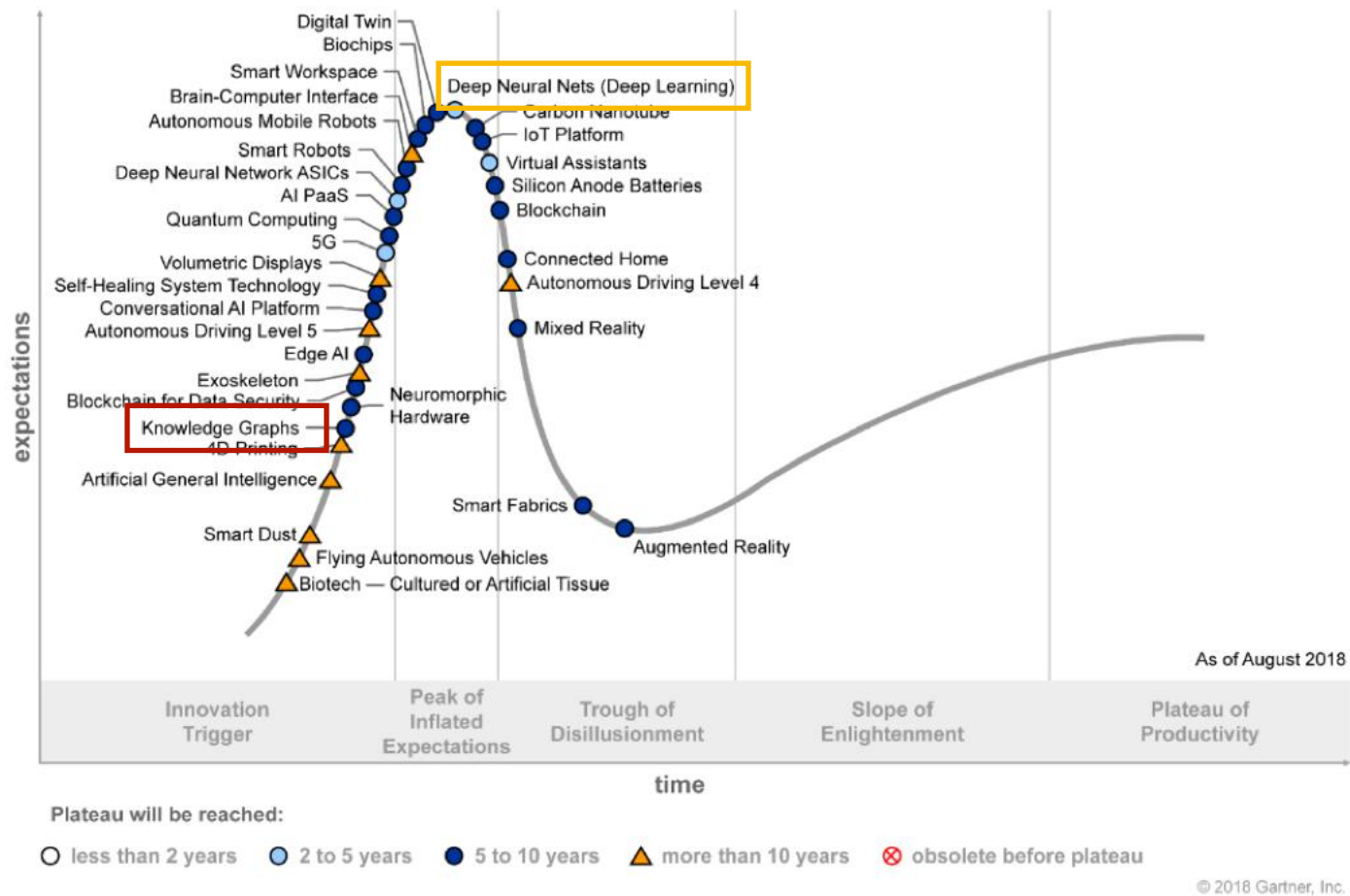




# 企业知识图谱云平台

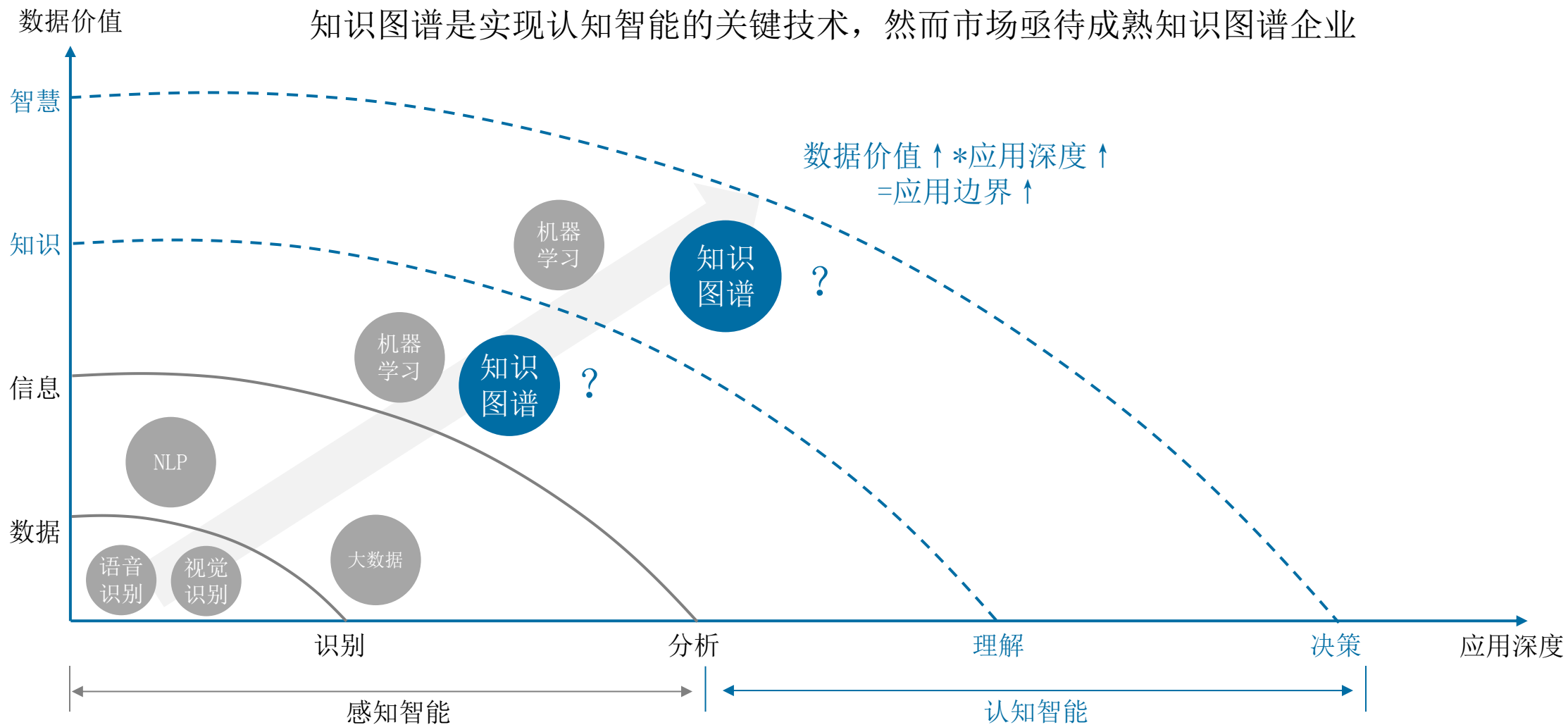
海知智能  
王宇 wy@ruiyi.ai  
201909

# 趋势：知识图谱在技术创新曲线中的位置



# 判断：知识图谱是AI突破认知智能的关键技术

AI已广泛实现感知层应用，突破认知层应用是AI发展的必然趋势；  
知识图谱是实现认知智能的关键技术，然而市场亟待成熟知识图谱企业



# 公司发展历程

四年多以来专注于知识图谱技术应用探索，商业化落地初见成效

## 启动

确定知识图谱为特点的语义理解发展方向  
结合语音交互形成原型：  
说说日历



2015

2016

## 落地

对接小米、百度等智能硬件语音交互开放平台，接入十万台设备  
企业智能化探索，医药、金融领域落地形成百万订单



2017

2018

## 爆发

银行为代表的金融行业、医药、汽车等产业快速增长  
智能硬件成为企业智能化新的流量入口

2019

2020



## 探索

面向智能机器人提供语义服务能力，问答机器人产品雏形



## 增长

覆盖国内几乎全部语音交互开放平台，接入设备数百万  
招行、中信、工行等银行为代表的金融知识图谱平台打造标杆案例



## 领导者

知识图谱平台这一细分领域的通过市场份额形成事实上的领导者

# 团队：知识图谱领域的创新领导者，兼具行业+技术+商业化背景



谢殿侠，联合创始人、CEO

- 曾在方正技术研究院（北京大学计算机科学技术研究所）从事文字处理技术开发，十年以上的技术解决方案经验以及行业理解和洞察
- 中文开放知识图谱联盟（OpenKG）发起人
- 北京大学力学系本科



王宇，联合创始人、常务副总裁

- 北京大学校友会副秘书长
- 北京大学金融校友联合会秘书长
- 北京大学创业训练营合伙人
- 在金融，媒体，法律等领域多年经验积累，擅长公共关系、市场及资源整合和组织管理等方面
- 北京大学法律硕士



丁力，联合创始人、CTO

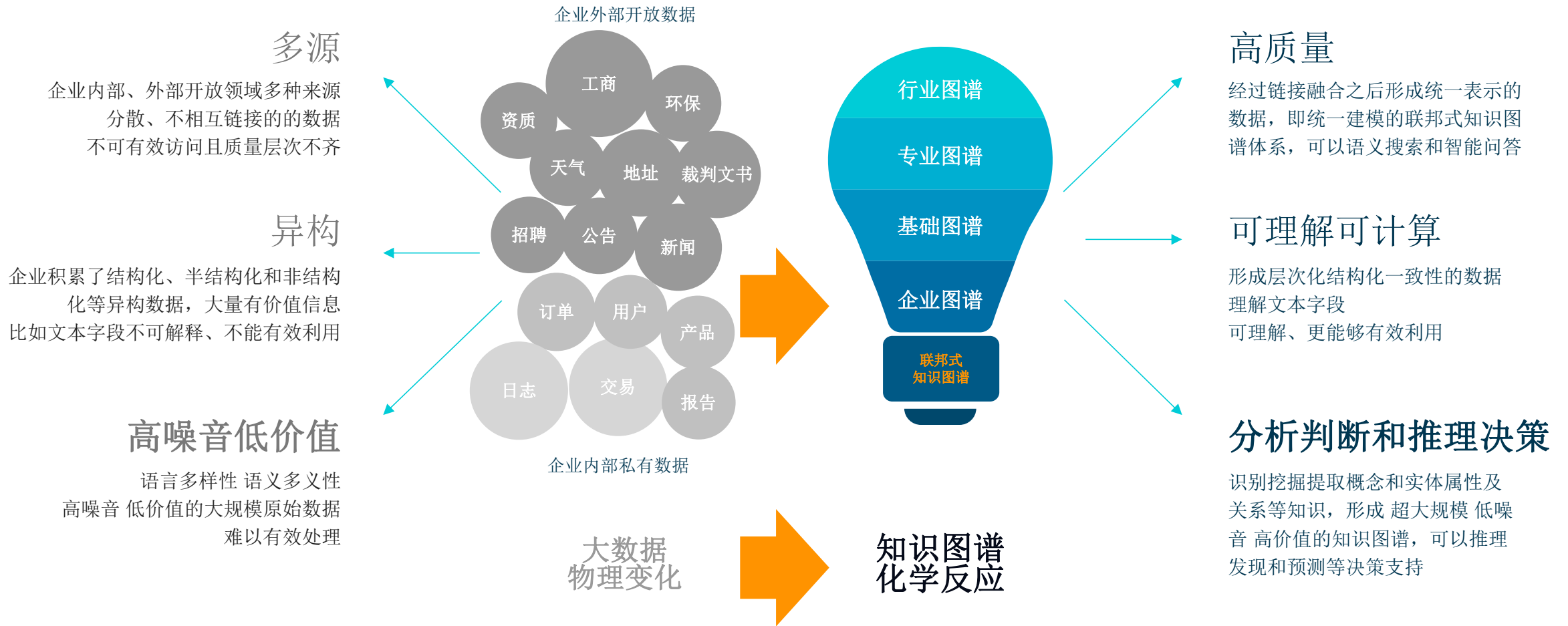
- 全球首款语义网（Semantic Web）搜索引擎作者，美国开放政府数据语义技术核心贡献者，参与知识图谱前身Semantic Web的创建
- 发表百篇论文，Google H-index>30；中文知识图谱规范cnSchema 牵头起草者
- 曾任高通科学家；斯坦福博士后，马里兰大学博士，北京大学计算机本科、硕士



James Handler，顾问

- 全球首款语义网（Semantic Web）的创始人之一
- 伦斯勒理工学院数据探索与应用研究所的所长；TWRC 计算机、网络和认知科学领域的教授；RPI-IBM中心主任，通过分析、学习和语义学研究健康
- 1995年富布莱特基金会奖金的获得者；美国人工智能学会AAAI、英国计算机学会BCS、电气与电子工程师学会IEEE和美国科学发展学会AAAS的院士
- 在语义网、人工智能、基于代理的计算和高性能处理等领域撰写了超过400多篇技术论文和书籍。  
Google H-index 84

# 痛点：数据资产价值化—从大数据（bigData）到智慧数据（smartData）



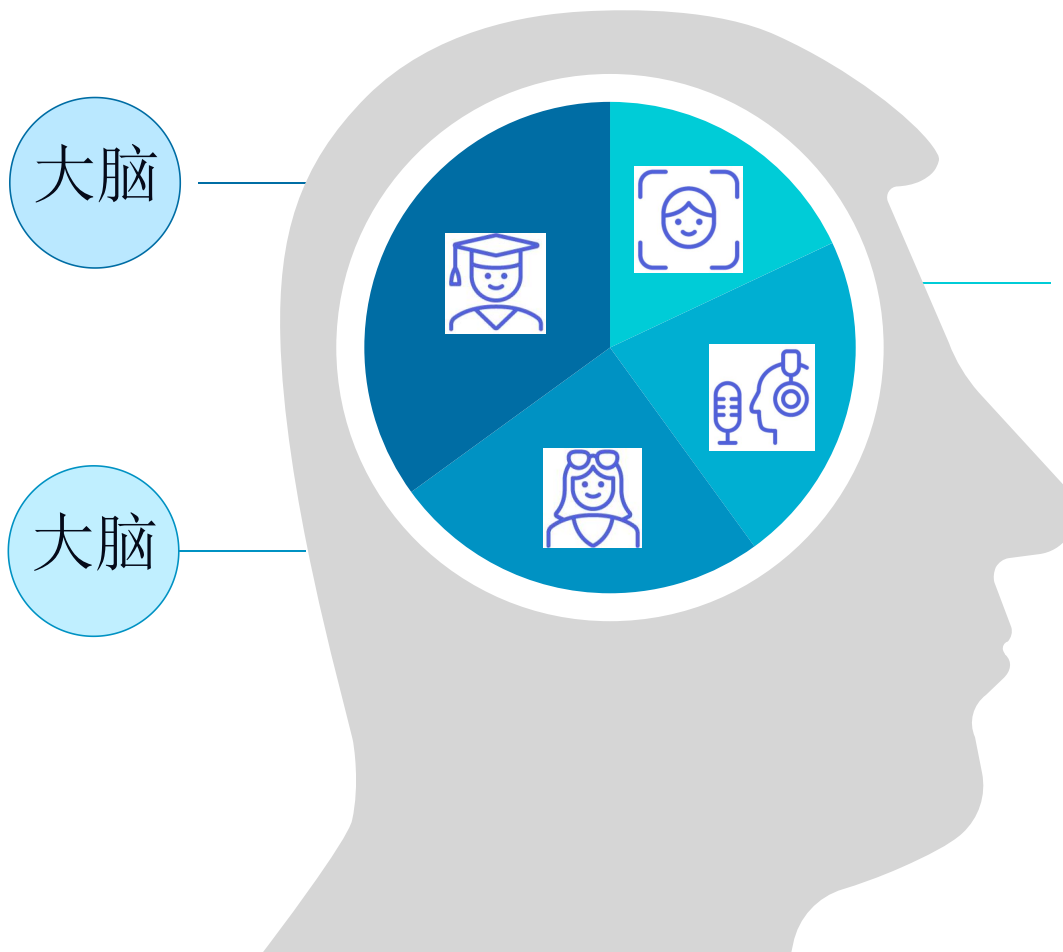
# 需求：每家企业都能有人工智能知识大脑

## 知识图谱

通过实体和关系这样的知识点而不是字符串构成的图谱网络来让机器理解和表达真实世界；  
知识图谱构建更有知识的大脑

## 机器学习

通过数据和模型训练而不是直接编程让机器具备学习和自动改进优化的能力；  
机器学习构建更聪明的大脑



大脑

大脑

眼

口耳

## 计算机视觉

数字识别、人脸识别、图像识别等应用到安防、金融等领域

## 自然语言处理

语音识别、自然语言理解等应用到智能硬件、智能客服、机器翻译、推荐以及营销等领域

# 应用：知识图谱四大场景



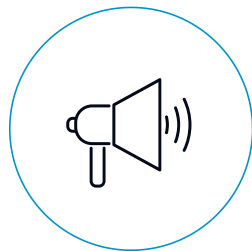
## 智能问答

从基于文本匹配、分类的智能客服，到掌握了领域知识的智能顾问，提供从售后为主到售前导购、咨询和专业服务的问答机器人，提升用户体验



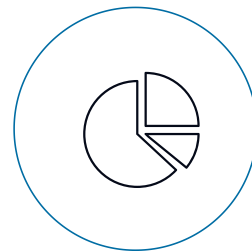
## 洞察发现

发现人物、机构、地点、产品、交易等关联关系，从而发现潜在销售线索、市场机会等



## 风险预警

事件图谱叠加产业链图谱可以形成主动式风险预警，发现更多潜在风险并可解释、可追溯



## 决策分析

不仅依赖与内部、结构化数据，而是融合外部开放领域、文本等非结构化半结构化数据构建知识图谱，类似于人的理解、判断、推理、分析和预测等决策机制

专业知识工作者部分工作自动化

# 市场规模：从数据仓库到知识图谱，仅金融和医药行业规模即达200亿

## 全球金融科技投入增长迅猛

- 2018年，美国各大银行按照收入的10%投入Fintech，**摩根大通**按年收入的10%、净利润的40%投入到Fintech，即80-100亿美金<sup>1</sup>。
- 大部分银行在金融科技的投入规模达到**营收的1%**。以招商银行为例，2019年招行科技投入从收入的1%大幅提升至收入的3.5%，金额达到80亿<sup>2</sup>。
- 中信银行、中国银行、光大银行等都提出了明确的金融科技投入研发的资金比例，范围在**营收总额的1%或者净利润的一到两个百分点**<sup>3</sup>。全球范围内领先银行投入税前利润的17%-20%用户数字化银行、金融科技银行的转型和创新<sup>4</sup>。

发展路径

赋知识：  
小B快速渗透

金融：20亿

价值：200亿人民币

中大型金融机构数量2万<sup>5</sup>\*平均客单价10万<sup>6</sup>=20亿

医药：42亿

医药资质企业数量42万<sup>5</sup>\*平均客单价1万<sup>6</sup>=42亿

探索中

赋智能：  
大B竖立标杆

金融：110亿

价值：60亿人民币

大型金融机构数量7333<sup>5</sup>\*平均客单价150万<sup>6</sup>=110亿

医药：14亿

大型医药企业数量282<sup>5</sup>\*平均客单价500万<sup>6</sup>=14亿

✓ 已落地

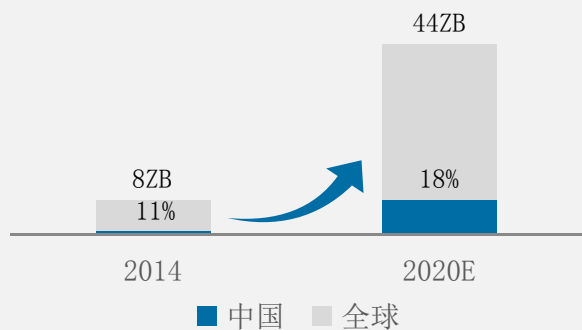
垂直行业复制

注：1、市场规模测算逻辑：企业数量\*需求比例\*平均客单价 2、在金融和医药领域的知识图谱需求比例为100% 3、各阶段估值逻辑：1) 赋智能：124亿\*20%市占率\*10%利润率\*25pe=62亿人民币；2) 赋知识：62亿\*20%市占率\*20%利润率\*50pe=124亿人民币  
数据来源：1. “2017金融安全与科技高峰论坛”；2. 海知智能采集客户信息；3. 中国贸易金融网；4. 麦肯锡《全球数字化银行的战略实践与启示》；5. 国家统计局，wind；6. 客单金额由公司提供

# 时机：基础设施和场景都已成熟，知识图谱应用正处于爆发临界点

## 1 数据：数据数量暴增，高质量数据已实现大规模开放

中国数据规模相对全球呈爆发式增长<sup>1</sup> 高质量数据实现大规模开放



- 工商数据开放：2016年12月，工商总局开放企业名称数据库<sup>2</sup>
- 裁判文书开放：2013年7月，中国裁判文书公开查阅<sup>3</sup>
- 2012年以来，国内各省市积极加入建设数据开放平台

## 3 算力：技术突破实现图谱大规模构建的工程能力

### 重大突破

数据存储与检索

图数据库2014年起高速发展至今并趋于成熟  
传统数据库完成图数据库的兼容<sup>5</sup>

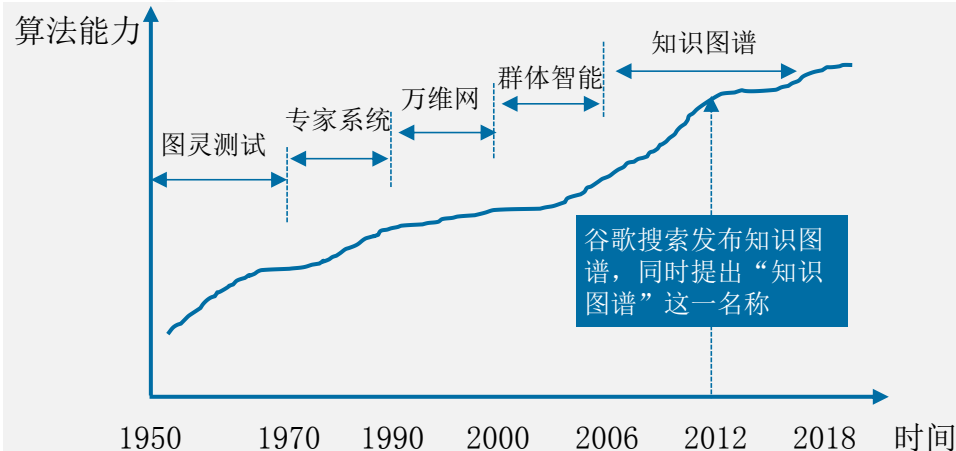
大数据计算

Hadoop降低数据分析存储成本至1/30  
Spark进一步提升Hadoop性能20倍<sup>6</sup>

云计算

中国云计算市场规模在2019年将达到4300亿<sup>7</sup>  
容器技术实现占用资源少跨平台部署

## 2 算法：知识图谱算法不断成熟<sup>4</sup>



## 4 场景：企业对知识图谱的重视大幅提升

企业认知

知识图谱作为核心技术独立开展项目

知识图谱作为大数据项目的一个功能模块实现

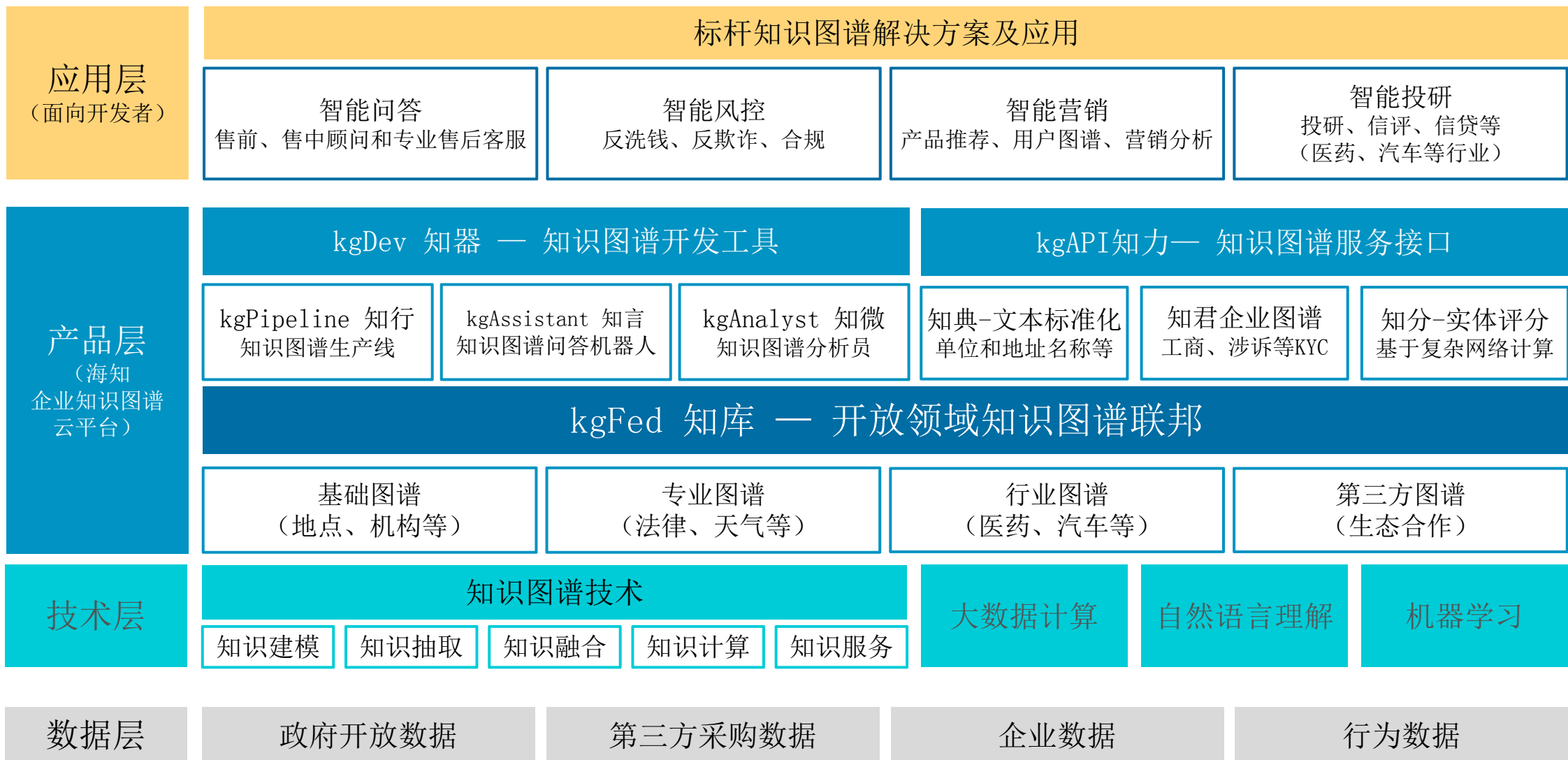
企业投入

知识图谱主导项目投入达到AI经费的10%<sup>8</sup>

知识图谱没有专项支出

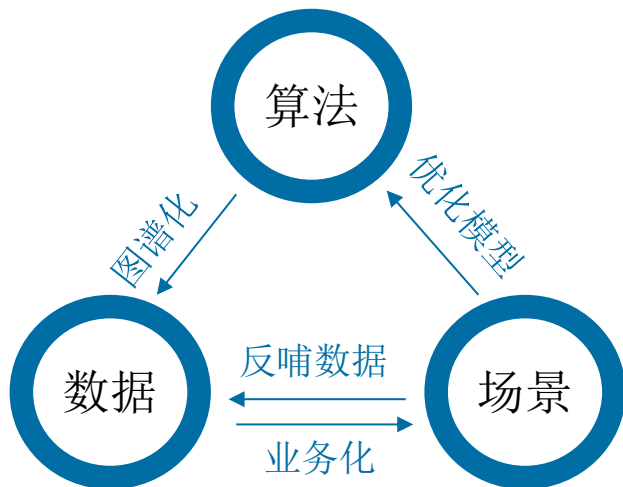
数据来源：1. IDC数据；2. 中央人民政府网站；3. 中国裁判文书网；4. 《2018知识图谱发展报告》；5. Gartner；6. 《大数据行业深度报告》华菁证券；7. 《云计算发展三年行动计划(2017-2019年)》；8. 公司采集客户信息

# 产品体系：企业知识图谱云平台及应用



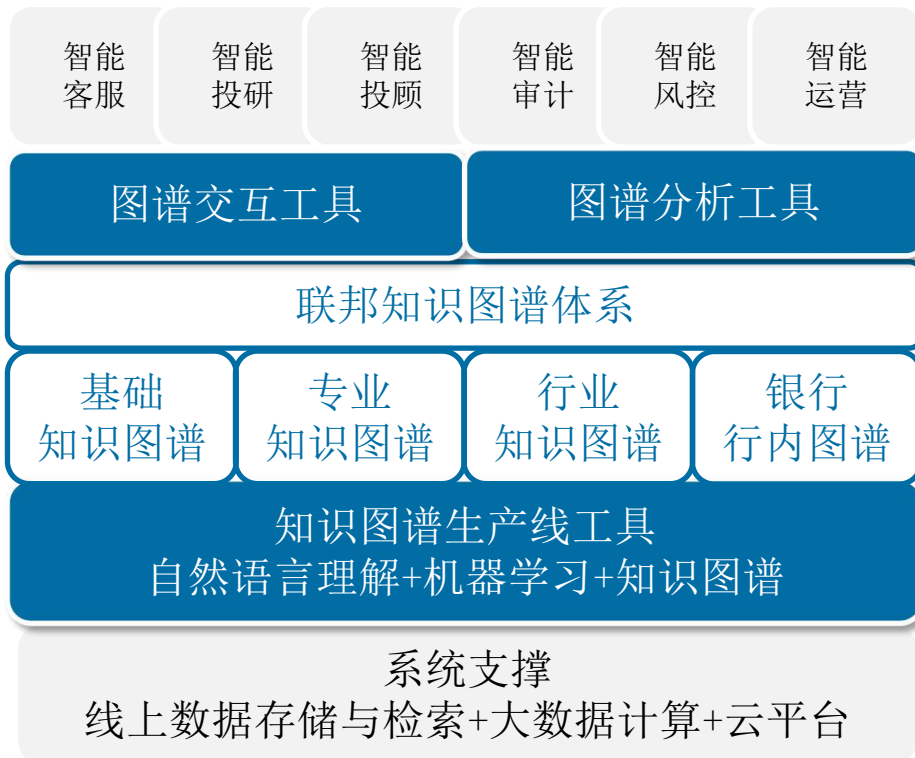
# 落地：已成功将知识图谱技术落地金融领域，形成增长小飞轮

- 多源异构数据结构化
- 高噪音数据价值提取
- 超大规模图谱处理能力

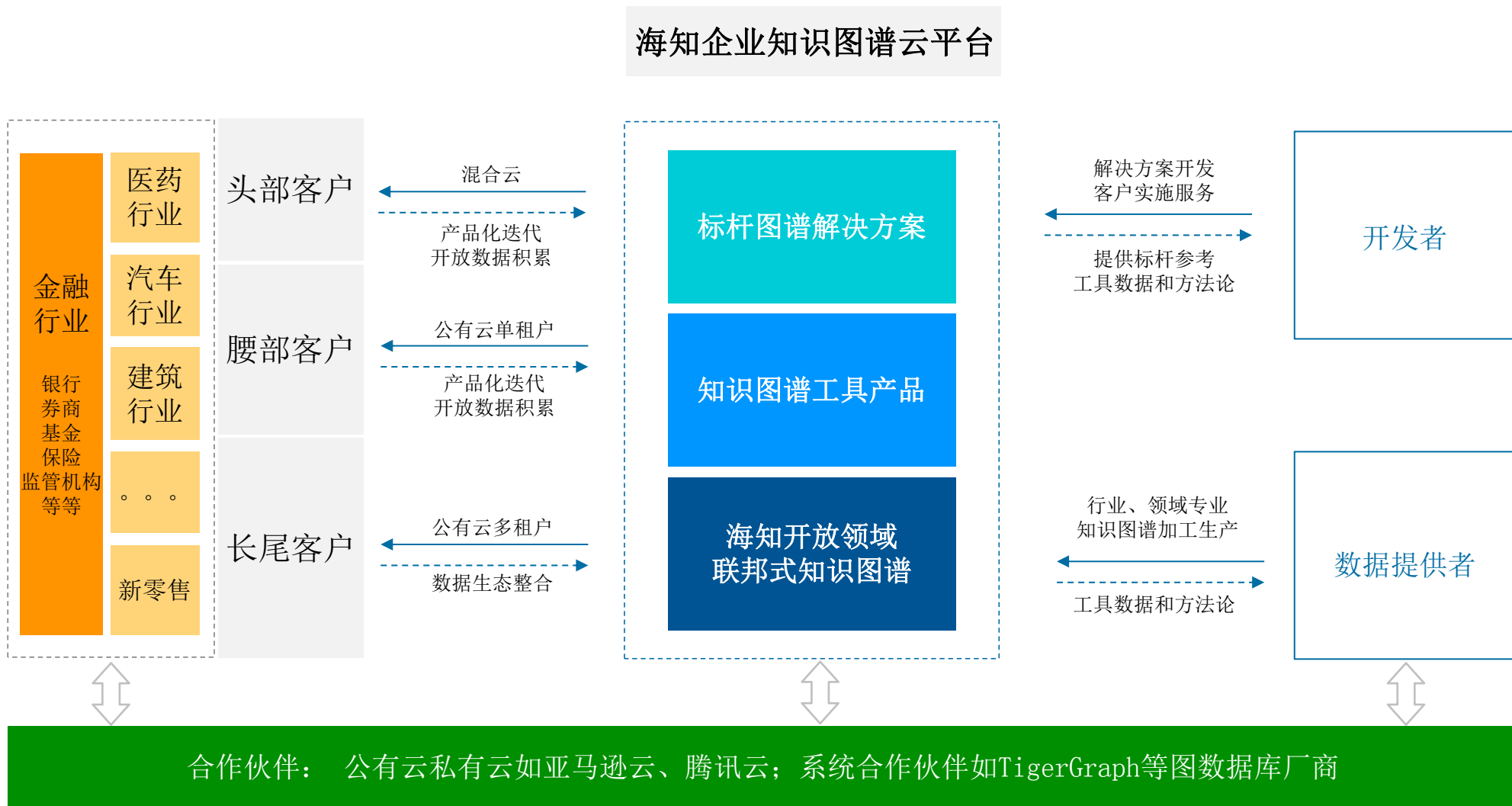


- 数据积累领先同行2年
- 数据价值优于数据供应商（构建高质量图谱，计算更多关系）
- 落地能力领先同行，已经抢占地头部银行客户，形成应用场景优势
- 服务能力领先缺乏应用场景的互联网大公司

## 银行知识图谱云平台



# 商业模式：短期大客户探索落地场景，中期产品工具复制，长期图谱数据生态服务



# 客户优势：服务头部客户，成功打造行业标杆，进入规模化复制阶段

## 行业客户

金融



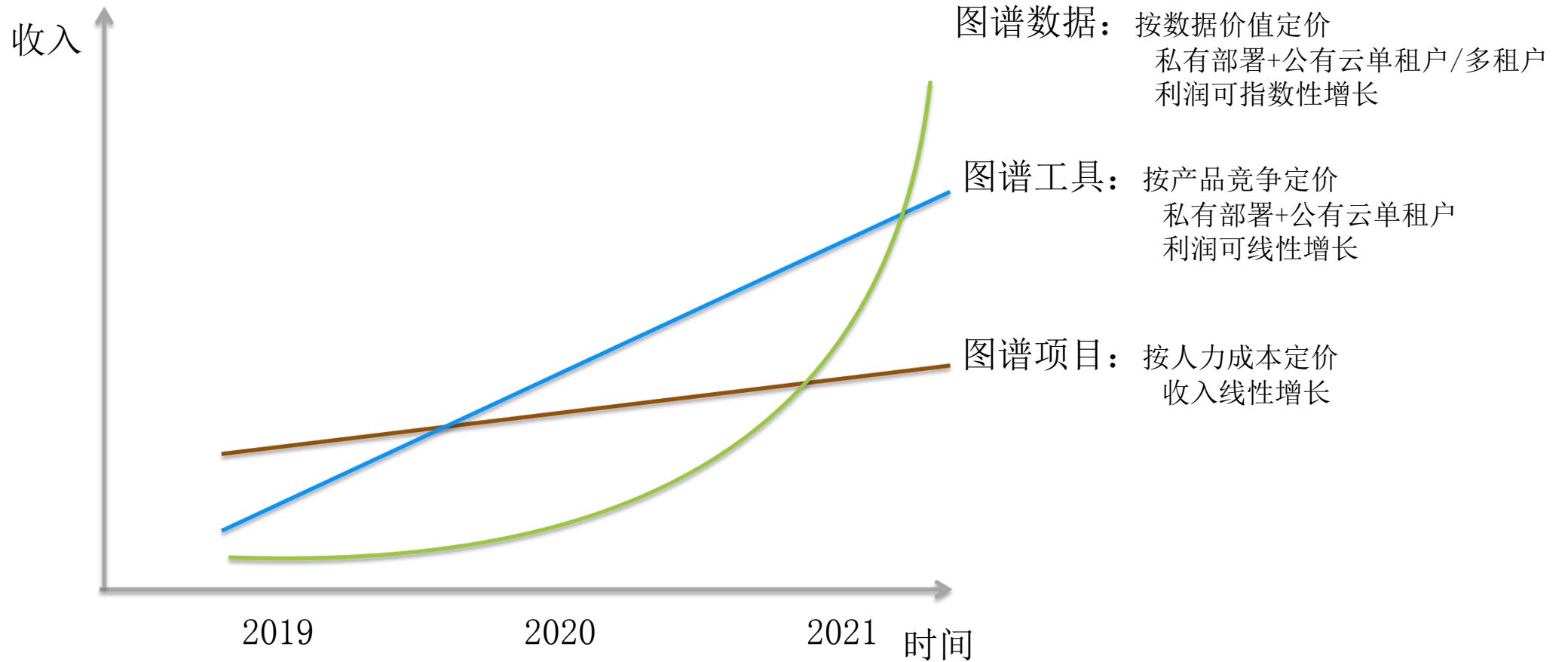
其他



## 合作伙伴



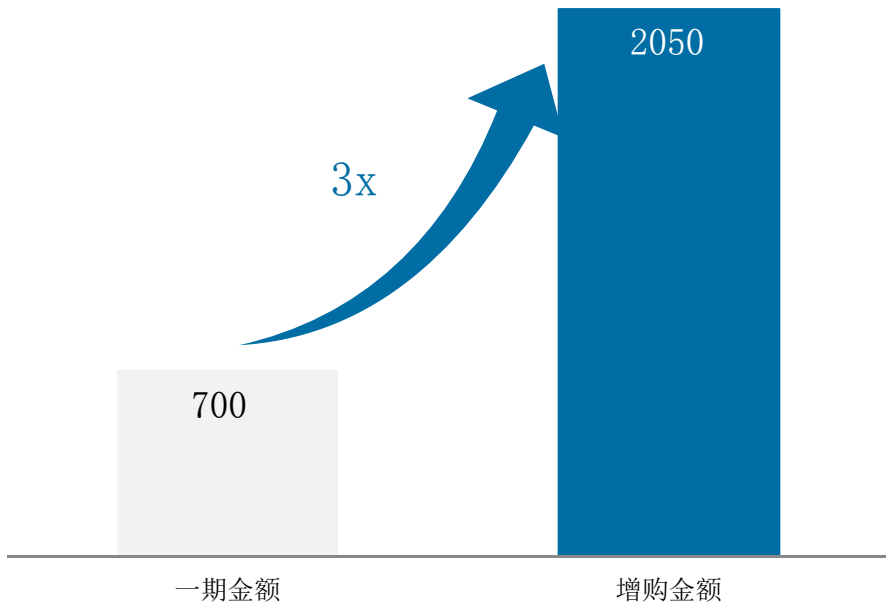
# 盈利模式：解决方案启动、产品工具植入、图谱数据持续服务的三级火箭



# 成长：进入爆发式增长的前夜

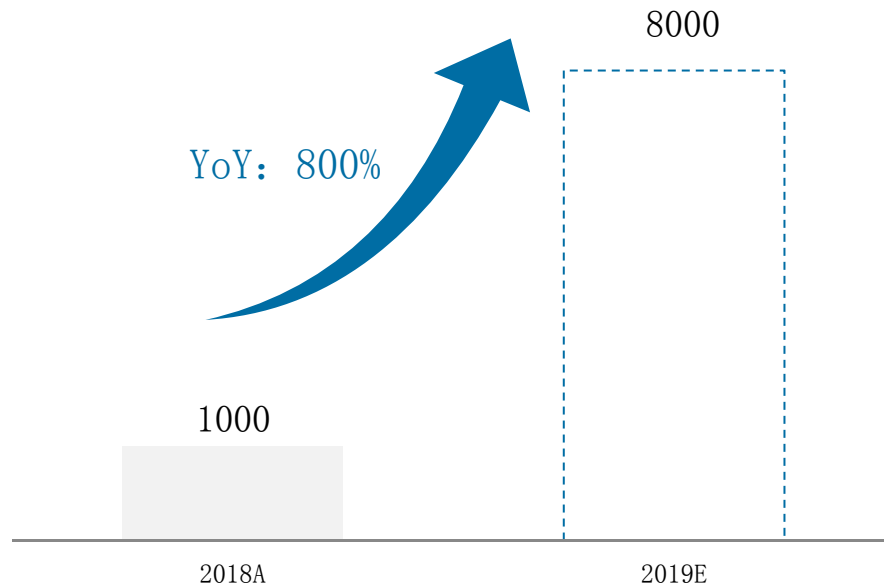
## 典型客户增购模型（万元）

合同金额



## 合同金额（万元）

合同金额



- 增购驱动因素：1. 图谱平台持续迭代升级，2. 更多业务场景标杆应用，3. 图谱数据扩大更新
- 增购情况：典型客户在2019年上半年增购金额较2018年全年已签约合同金额预计增长200%

- 合同增长：2018年签约合同金额一千万，2019年预计签约八千万，800%增长率
- 增长驱动：头部银行标杆客户深入铺开，银行为核心的金融行业开始复制；医药、汽车产业开始系统性复制



# OpenKG发起人

中文开放知识图谱（简称OpenKG.CN）

旨在促进中文知识图谱数据的开放与互联，  
促进知识图谱和语义技术普及和广泛应用



微软亚洲研究院

## 开展活动和业务

1. 培育开放社区

2. 组织社区活动

3. 建中文开放社区规模知识图谱系统 OpenBase (开放CC0协议)

4. 起草中文知识图谱规范 cnSchema (海知智能牵头)

经 典 部 分 知 识 库

📦 Zhishi.me = 逾1000万实体 - 10亿三元组

📦 CN-DBPedia = 逾1000万实体 - 8000多万三元组

📦 NLPRE的语义知识图谱 = 1000万实体

📦 Belief Engine = 5000万三元组

📦 PRU-PIE知识库 = 70万 实体 - 900万三元组

📦 ConceptNet中文部分 = 2800万三元组

📦 WikiData中文部分 = 60万 Items

# 海知智能：中文知识图谱领域的Palantir+Kensho



=



+



联邦式知识图谱平台

基于知识图谱的行业应用解决方案

基于知识图谱的智能投研服务平台

2018年合同收入10亿美元<sup>1</sup>

2018年被S&P Global收购

估值410亿美元<sup>2</sup>

收购价5.5亿美金<sup>3</sup>

数据来源：1. 云栖社区《起底硅谷最神秘、估值最高的大数据公司：Palantir》；2. 腾讯网《神秘数据公司Palantir或明年上市 估值达410亿美元》；3、4. 网络公开资料



## 顶尖团队：核心团队是知识图谱领域的资深实践者，兼具行业+技术+商业化背景

- 核心团队具备十年以上知识图谱领域的技术行业经验和资源积累
- 领先的全球视野，团队成员参与了知识图谱前身语义网Semantic Web的创建，并牵头起草中文知识图谱规范

## 市场机会：知识图谱是人工智能实现行业应用的关键基础设施，具备百亿级市场规模

- 知识图谱是人工智能在认知智能阶段的驱动核心技术，亟需出现专业服务者
- 在金融、政府、医药等知识密集领域，知识图谱应用市场达到200亿，足以孕育出独角兽企业

## 竞争优势：领先的数据积累能力和落地能力，奠定行业领导者地位

- 领先同行2年的数据积累和应用，筑造了多领域知识图谱融合应用的壁垒
- 领先的技术实力、产品能力和落地能力，打造业务内领先的数据+算法+应用融合的能力

## 商业势能：已在金融领域验证模式，服务于行业头部客户，进入商业爆发期

- 已竖立标杆客户：产品成功在多个大型商业银行应用，如招商银行、中信银行、工商银行，即将进入规模化复制
- 客户付费能力强：客户增购潜力巨大，付费意愿强，单客户价值可达千万以上

# 案例：服务头部商业银行，首个总行级别的知识图谱云平台

## 某大型股份制商业银行

### ■ 客户背景

- 从商业银行到Fintech银行战略转型  
业务上全员创新探索智能化升级
- 数据仓库积累了海量历史数据
- 各部门自行采购外部数据

### ■ 需求痛点

- 总行级数据治理统一规划管理
- 行内数仓海量数据，有效利用率低
- 外部数据不能有效共享
- 认识到知识图谱价值缺乏有效方法论以及标杆项目
- 大量业务和技术人员缺乏数据建模和图谱算法能力

### ■ 预算投入

- 2018年Fintech预算22亿
- 知识图谱投入占比可达10%
- 未来计划投入增速30%

## 海知智能解决方案：总行级联邦式知识图谱平台

### ■ 咨询服务

- 银行联邦知识图谱建设规划

### ■ 解决方案

- 应用项目：智能客服、智能投研、智能营销、智能风控等
- 产品工具：知行kgPipeline、知言kgAssistant、知微kgAnalyst

### ■ 客户价值

- 图谱数据：地点、机构、基金、医药行业等
- 建设面向总行级联邦式知识图谱平台，提供数据治理工具基础
- 打造若干业务场景的标杆知识图谱应用，探索业务创新可能
- 知识转移知识图谱能力到技术部门、业务部门，降低人员门槛、缩短设计开发周期和提升价值
- 最终提升风险管理能力、增加精准营销效果以及提高运营效率

### ■ 海知价值

- 合同订单：咨询和标杆项目类；产品工具类；数据类
- 产品迭代：结合落地场景，抽象需求形成工具产品
- 数据积累：结合实际需求，积累开放领域相关高质量图谱数据