

天云数据

人工智能第二波：AI重构商业流程

数据资产化 · 数据融合和流动 · 数据价值化

公司简介

About US

天云数据国内能够同时提供分布式计算平台产品和 AI 平台基础软件的科技厂商，拥有博士后工作站和国家级高新技术企业称号，并于 2016 首批进入中关村前沿科技企业重点计划。

公司在分布式数据库领域有自主产品，填补了联机事务等领域空白，并在多个大型银行核心交易系统部署验证。在人工智能方向领先于 BAT 发布了分布式 AI 平台，于 2016 年在大型股份制银行落地。该平台与科大讯飞一起获得了北美 ZDnet 评选的十大 AI 赋能平台奖项。

凭借分布式 AI 能力，获得国际一线机构 KPMG 评定的中国 Fintech50 强，天云自 2016 开始为金融机构提供数据模型深入信用风险欺诈等金融业务领域，为光大兴业银联等提供信用业务相关计算与数据科学模型，亚太 Asset 财经评选的 TripleA 金融科技领先奖。2017 年人工智能产业发展报告白皮书发布，天云数据同时上榜“机器学习深度学习平台”与“数据平台”；天云数据在中关村 5000 亿计划中名列大数据领域 TOP10；同时上榜中科院发布的人工智能企业排行榜百强；中关村前沿联盟企业；深圳人工智能学会会员单位、华夏基金会理事单位、中小银行联盟会员，国家统计局大数据平台企业、海关跨境电商协会会员单位。

天云旗下控股能源、军工、医药等领域的高科技公司：五季数据科技，致力于石油行业，基于 AI 的智能勘探开发技术服务；天云君安，面向军工、安全行业；医智生命科技，打造医药 +AI 的标杆企业。

重要客户

Main Customers



荣誉奖项

Honor & Awards

入选艾瑞2019中国人工智能产业研究报告
被中央电视台财经频道授予“实业创新典范”
荣选爱分析·2019中国数据智能创新企业50强
两度入选中国人工智能企业TOP100
多次荣获《中国大数据应用最佳实践案例奖》
荣获2017年度最具价值人工智能服务机构
天云数据荣选2018年大数据独角兽公司
天云数据入选2018企业服务产业独角兽榜单



英伟达深度学习及人工智能双创项目成员
2018中国人工智能金融服务企业排行榜 | 位居第七位
荣获2018年金融科技介甫奖最具价值大数据服务机构
荣获《2018智·创未来大会》首届“十大最具价值”科技创新奖
天云数据银行卡申卡反欺诈模型入选“机器之心”商用搜索
被中关村科技园区管理委员会授予“中关村高新技术企业”称号
荣获亚太“The Triple A Digital Awards 2016 Fintech of the Year”
数据中台助力银行数字化转型荣获《银行家》2019十佳金融科技创新奖
荣获英国2018 M&A Awards中的Big Data&Analytics Advisory Firm of the Year

多项业界权威认证

软件产品CMMI3级标准认证
北京新技术新产品(服务)认证

高新技术企业认证等业界权威认证
中关村前沿技术企业



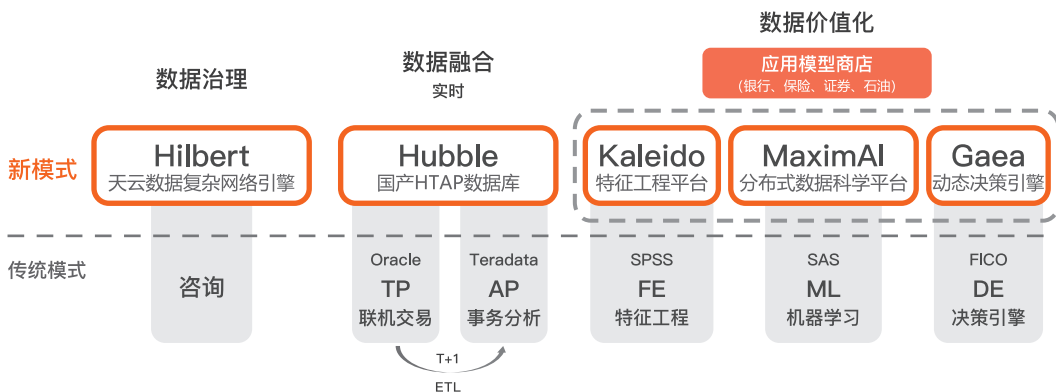
发展历程



建设人工智能底层能力，以AI PaaS平台推动产业化落地

天云数据成立于2013年5月，是一家着眼于人工智能领域的底层软件提供商，其核心业务是为客户提供AI PaaS化平台服务和支持数据高并发实时查询的数据库，辅助客户加速AI规模化、工程化落地部署。致力于推动中国产业数字化建设中软件国产化的进程，其业务在金融方面有较深的积累，同时在能源、医药、政府等行业也有广泛布局。

产品矩阵



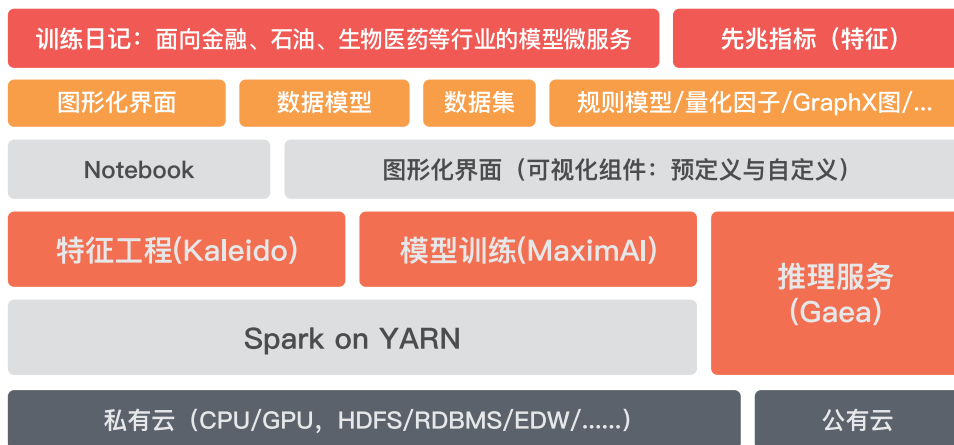
AI PaaS

AI能力平台化输出降低了入局门槛，推动商业化第二波爆发

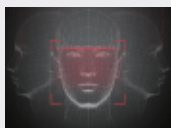
随着人工智能技术对传统产业的不断渗透，越来越多的企业对 AI 产生了需求，但自主组建一支 AI 技术团队，研发相关系统和应用对于大部分公司而言投入产出比并不高，而且难以达到“即插即用”的效果，因此通过云平台 PaaS 层输出 AI 能力的 AI PaaS 服务成为需求方向。结合产业化落地，AI PaaS 平台可分为三个阶段，既模型自动化生产、模型规模化生产和模型智能化生产，逐步实现去监督化生产。AI PaaS 又分公有云平台和私有云平台，二者在架构方面主体基本一致，只有在权限管理、资源管理和数据管理部分区分公有化和私有化，总体来看 AI PaaS 要满足模块化、分布式、资源共享、可拓展和环境分离五大特性，以满足不同量级用户的并发需求。

AI PaaS基础架构和商业价值

AI PaaS基础架构



AI第一波爆发：人机交互应用落地



视觉计算



自然语言处理

AI第二波爆发：重构企业商业流程

算法重构复杂商业流程 庞大数据资产盘活

- 无需数据搬家，可对海量数据进行快速探查和分析
- 无需机器学习知识，即可建立专家级模型
- 降低机器学习门槛，降低对数据科学家的依赖
- 一站式机器学习平台

天云大数据

提供数据中台能力，替代西方核心基础软件

银行核心业务系统是支持银行业务运营的关键系统，需要具备满足高并发、高性能、实时性、安全和稳定等能力要求，当前银行核心系统普遍采用国外 IOE 技术架构下的软硬件产品及服务，在移动互联网发展的冲击下已显乏力，且升级维护成本过高，面对这一痛点，天云提供国产分布式数据库软件，在金融场景中替代 A 类核心系统惯用的西方 IOE 架构，解决银行联机事务中系统减负问题。在银行数字化转型中，天云提供了较强的数据中台能力，实现了数据仓库的消费化，由 BI 向 AI 升级关键数据驱动的转型，能够替代西方核心软件，构建新型人工智能基础设施。

数据中台服务

算法重构复杂商业流程



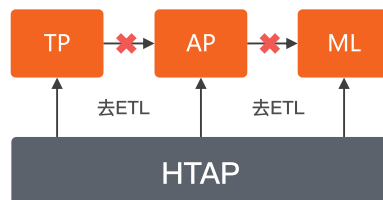
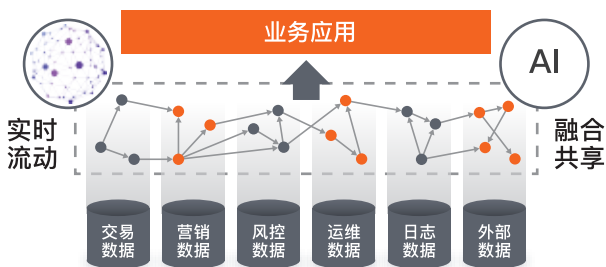
全量数据+机器学习算法建模=精准预测模型，事前预防，快速响应，快速迭代



数据驱动业务问题求解，全量数据建模取代传统专家经验与抽样建模，以数据而非专家为核心

天云AI PaaS实现BI向AI升级关键数据驱动的转型替代西方核心软件，构建新型人工智能基础设施

庞大数据资产盘活



HTAP数据库：支撑数据融合与数仓消费化实现“数据搬家”转向“代码搬家”

天云HTAP国产分布式数据库替代A类核心系统惯用的西方IOE架构解决银行联机事务中系统减负问题

数据中台三部曲

2011年5月，麦肯锡发表研究报告《大数据：下一个创新、竞争和生产率的前沿》，八年，大数据从概念诞生到行业落地，整个数据行业由最初漫无目的的基础设施兴建，逐渐发展到基于中台而驱动决策的数据智能时代。

数据中台具有坚实的基础平台能力，距离业务更近，能更快速的响应业务和应用开发的需求，可追溯，更精准，重在以数据驱动为中心。在数据量爆发式增长、涌现大量新业务场景的大型金融机构之中，数据中台可以加强数据融合和流动，支撑混合交易处理，大幅提升业务响应能力。

现代企业的业务模式正在经历从流程驱动到数据驱动转变的商业重构，从 BI 向 AI 升级。数据不再是业务系统的副产品，而成为业务系统的核心资产，成为企业运营决策和业务发展的驱动力。数据中台从诞生至今，经历了数据资产化、数据融合与流动、数据价值化的三部曲。

数据中台三部曲

流程驱动到数据驱动

自动化
数据治理

ADG

数据
融合和流动

HTAP

提供
人工智能化服务

AI PaaS

ADG

第一阶段：数据资产化、自动化数据治理

数据中台的第一步是发现识别和标识度量数据资产，传统方法是通过人工咨询的方式实现数据治理，回答数据在哪里，从而解决新的应用升级和应用的上云容灾等等数据关联等问题。但传统上基于业务咨询执行的主数据治理 MDM 技术很难满足快速多变的数据治理结构，现在出现的复杂网络（比如天云数据 Hilbert）等知识表示的前沿技术应用的 ADG（Automatic Data Governance），加速和自动化数据治理工作，可以建立统一的关系视图，建立数据自治的数字资产网络。

ADG 商业价值

- 为集成开发环境简单性、可靠性、一致性和自定义性的提供保障。
- 将数据治理纳入系统或应用开发过程，进而企业可以实现更全面，可持续的数据治理计划。
- 向企业中的人员提供均可以使用的数据视图。
- 支持更广泛的业务需求及应用需求。
- 数据血缘和影响分析在治理和风险评估中体现了重要价值，为数据治理提供更直观的数据视图。

ADG 如何构建完整数据视图

在应用调用时，常常通过三种方式对数据进行加工或使用，其中包括数据字典的引用，通过 schema 建立主键、外键及关联关系，及函数与存储过程对数据进行的转化更新。在 ADG 中，将这三种方式经过不同的拆分与处理，最终生成数据间的完整视图。

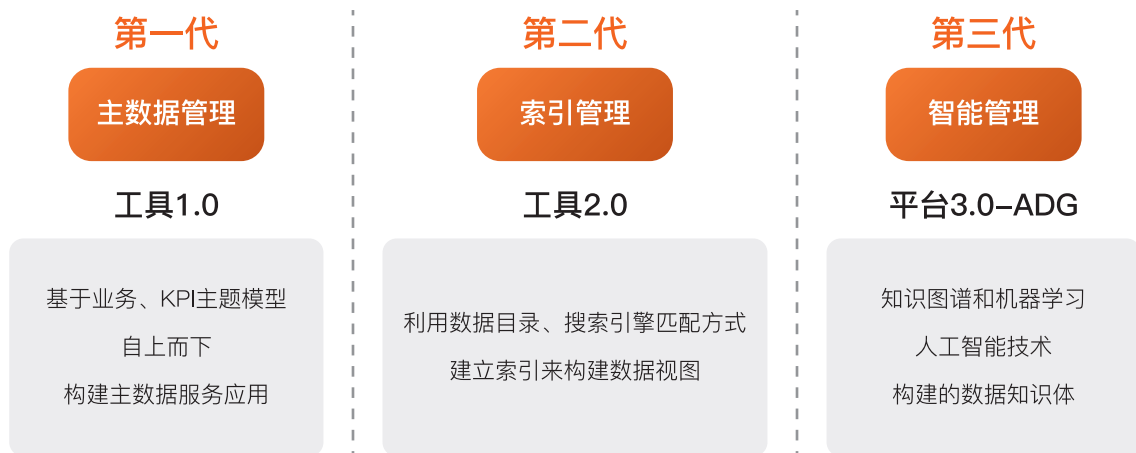
首先，通过对数据字典的定义及使用，创建适用于目标的数据体系的参考数据，其中涉及了可使用的元数据，数据类型及取值范围。

其次通过对 schema、PK、FK 等关联性的分析，对于原始数据库、表及字段的定义进行全面分解整合，通过主外键等关系找出表间存在的关联关系。与此同时，ADG 根据不同的使用方式将其进行划分，最终生成数据库间、表间及字段间的七种数据关系。

再次，利用函数和存储过程发现数据转化更新的演变过程。根据两者的可重用性、共享性及可移植性，探索及还原了数据产生、演变、使用及消亡的生命周期。

最终通过以上三步的数据间关系分析，ADG 从不同视角实现了对现有数据系统较为完整分解与转化，并且通过所捕获的信息构建了数据完整视图。

数据治理发展趋势的三个代际



第三代：数据治理的核心能力

I、**字符 -【词向量】**：用于字符型数据相似性分析。通过计算字符串转换次数，来进行相似度比较。其中转换操作包括替换、插入、删除等。通过计算操作次数，进而获得字符型数据间的相似程度。

II、**数值 - 数据冗余分析【相关系数】**：用于数值型数据相关性，对当前数据表中的数据进行数值比较计算，计算数据表中数据冗余程度及数据分布情况。

III、**主体构建【机器学习】**：利用机器学习的方式优化系统中的主数据。ADG 从多个数据源获取数据，对已知数据进行分析与匹配，通过机器学习的方式，在构建数据实体。在此期间，为确保准确性，还通过自动化和协作的方式进行指定的数据校验。

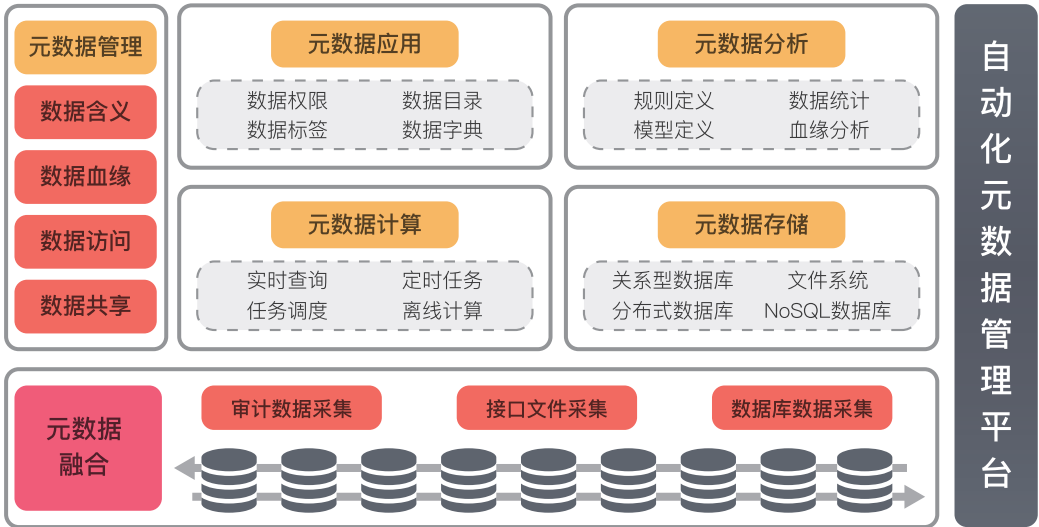
IV、**数据目录【人工智能】**：ADG 通过自动化和协作的方式对数据系统构建了数据目录，其通过业务元数据将技术元数据进行丰富，通过数据间的关联关系、数据间的相似性进行数据编目。

V、**图数据库**：通过使用图数据库作为后端存储的一部分，更好的将数据实体与关系进行存储及使用，提高了数据服务效率。

ADG 所解决的问题

- 数据可信度：提升数据的质量、准确性和安全性。
- 数据可访问性：对于不同人员数据人员，即数据管理者、使用者、运维人员，获取正确数据。
- 体现数据价值：对数据系统进行统一管理，增加数据使用的能力。

ADG 产品功能概览



ADG产品优势



HTAP

第二阶段：数据的融合和流动

传统架构中，TP\AP\ML 分属不同的应用系统，数据无法融合应用，只能进行数据搬家。

新的分布式 HTAP 架构中，可以实现三者的无缝融合：同一个数据计算平台，同时面向联机事务处理，进行密集分析和数据挖掘，TP\AP\ML 融合计算场景，对核心 Oracle 数据库等关键计算组件实现替代，从而不需要每日跑批 ETL。

HTAP 系统优势

打破了事务处理和分析之间的“墙”

大数据体系下数据实时处理和分析能力

融合架构同时支撑 TP 和 AP

凭借 8 年来在分布式计算集群的沉淀，及多家大型银行核心交易系统的实践验证，天云数据推出了百万原生代码规模的 HTAP 数据库产品：Hubble，已入选 WIKI 百科定义 HTAP 数据库：



多源异构

Mysql、Oracle、Hbase、ES、Hive 等都可以作为 Hubble 的数据源，支持跨数据源查询



超高并发

支持上万用户在线实时高并发修改与查询



SQL 标准

全部 SQL 的标准支持，即提供高并发有提供 OLAP 分析



实时处理

业务的实时数据分析处理

Hubble: 使用灵活，无需太多额外的成本，对传统应用的迁移，只需要替换 jdbc 使用的驱动即可。自诞生起就基于 hadoop 之上，天生支持海量存储，支持对 TB 甚至 PB 级的数据。

既支持 OLTP 又支持 OLAP，一站式的解决联机查询问题，无需部署众多的数据库组件，只需要使用 Hubble，就可以涵盖大部分的业务需求。

Hubble 技术能力

去 ETL 的混布融合计算框架，无需再每天跑批迁移数据；

全面兼容 SQL 服务，无需在 Hbase,ES,Spark 等大数据组件原生 API 上开发；

支持多元异构数据库的联邦计算 (FederationQuery)，无需在 Oracle/MySQL 等传统数据库之间做全量数据同步；

面向机器学习的分布式计算框架，原生支撑 AI Paas 服务。

Hubble 生态定位

TP

AP



Hubble
- HTAP数据库 -

impala



HUBBLE：迈向数据新世界过程中更加易用和简单

混布能力是对生态的组件的又一次升级

让客户通过标准统一服务入口实现生态中不同组件的能力

联邦计算



联邦计算解决了不同数据源之间数据即席分析难题，同时减化了传统数据搬迁到统一存储池中再进行数据分析的过程，大大提高了即席交互性的服务能力。

优势

标准访问入口

客户统一入口实现各种源系统数据的访问

灵活性强

可自由编写 SQL 实现数据访问

支持多种数据源

可灵活支持对数据源的接入，适用性强

效率高

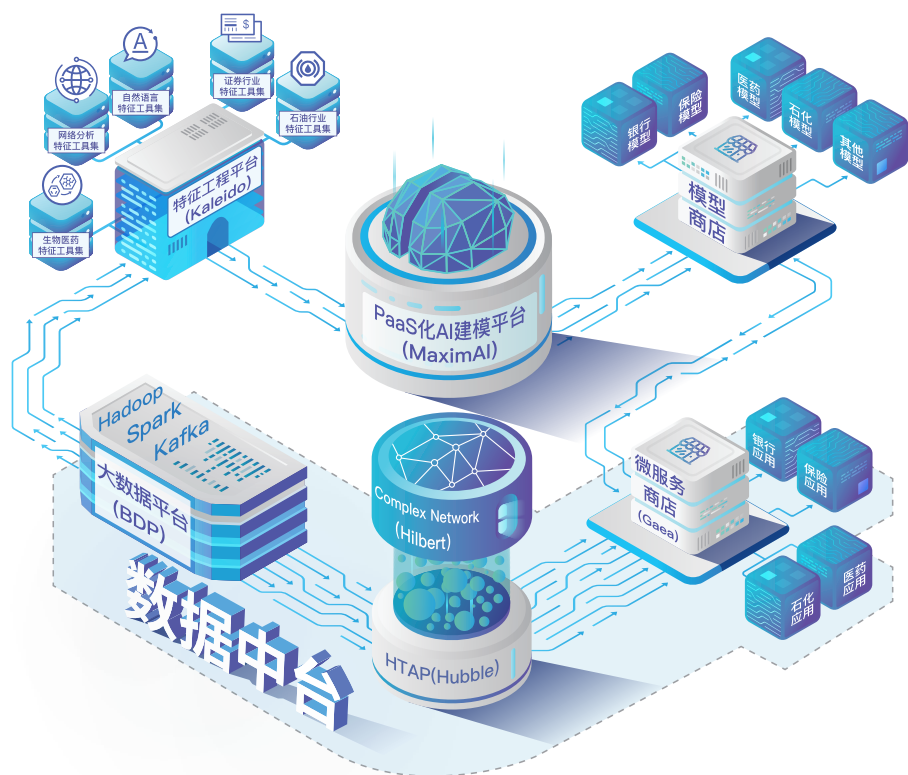
无需将数据完全搬迁，即可以现有数据的分析探查

AI PaaS

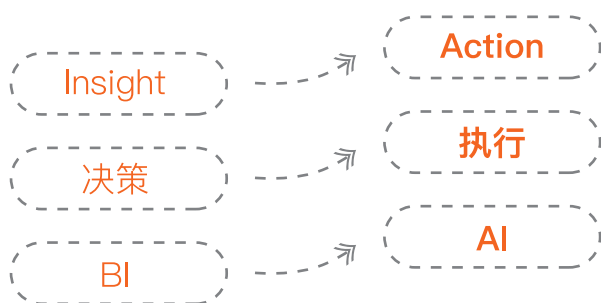
第三阶段：提供人工智能化服务

企业要获得敏锐嗅觉和洞察能力,数据是基础。数据已经被视为当今时代新的生产资料,没有数据,机器学习、深度学习和人工智能等技术也无的放矢;反过来,利用数据中台技术实现数据治理,完成数据质量提升、数据融合之后,掌握一个能够有效地挖掘数据、洞察数据的工具则是企业拥抱 DT 时代的关键。只有凭借这样一个灵活、高效、易用的工具,企业才能够真正唤醒沉睡的数据,获取到从数据中涌现出的新业务逻辑、新商业模式,基于此改变企业的业务模式和商业模式,唯有如此才能够将沉睡的数据变现为真金白银,实现数据应用的闭环。

通过数据中台可以提升数据价值,并使数据资产变现,催生新的利润增长点。数据中台中包含的 AI PaaS 平台,支持原始数据的特征工程和模型孵化训练以及模型的上线实时运行,正是帮助企业实现数据应用最后一环的关键。



AI PaaS 平台并不针对特定应用场景提供解决方案，而是提供一组丰富的特征工程算子和人工智能模型训练算法以及一个模型上线环境，便于企业利用这个几乎无所不包的工具箱和其数据对复杂多变的业务进行灵活的、响应及时的建模和模型应用。企业利用高阶算法对数据进行升维，抽取海量数据中的特征并描述出来，进而重构一个新的商业世界，这正是利用 AI PaaS 平台能够获得相对传统的专家经验方法得竞争优势的原因所在，也是企业在 DT 时代实现数字化转型的核心。



在过去流程驱动的模式下，企业往往会从行业经验和规则中总结出“最佳实践”作为标杆，而在数据驱动的模式中，其知识构造则由算法从数据中提炼而来，可以用更丰富的特征表达能力把人类没法抽象的复杂事务量化出来，不再受限于人类的大脑。

这意味着，利用 AI PaaS 平台提供的算法能力，可以定义更多以前无法定义的“最佳实践”，创造出新的商业模式。

利用 AI PaaS 平台提供的
机器学习与深度学习等算法模型
进行业务“演练”和预测

可以回答“会发生什么”的问题

通过数据中台提供的
“数据 + 算力 + 算法”能力

将预测的结果反馈回业务系统中形成闭环信息流

对业务模式进行干预和优化

可以回答的是“最好发生什么”的问题

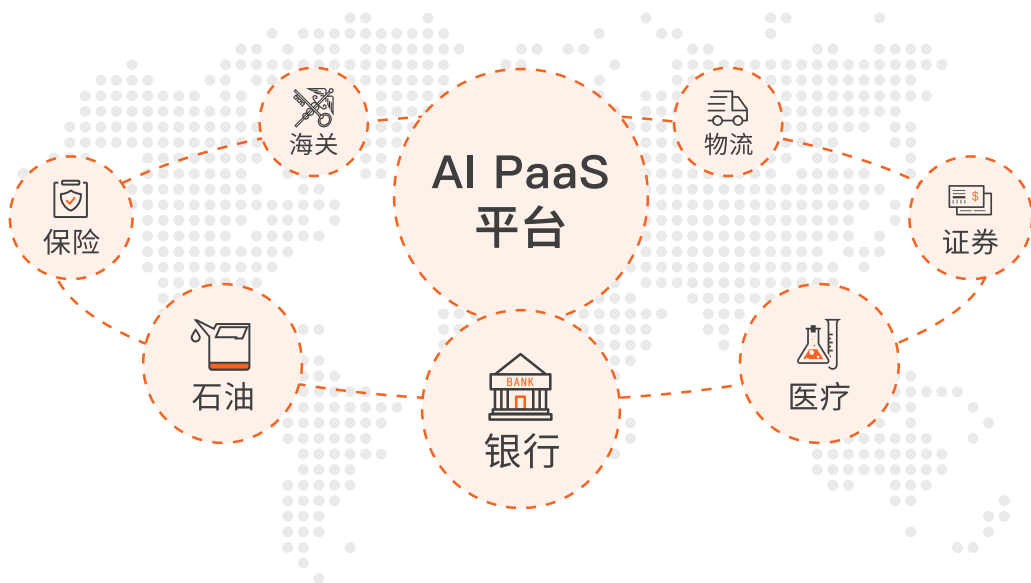
凭借数据中台尤其是其 AI PaaS 平台持续的挖掘出的全新的“最佳行业实践”，商业流程被重构，新的商业模式和利润增长点有望不断涌现。

以金融行业为例，基于其自身庞大的数据资源和天云数据 AI PaaS 平台的算力基础，许多金融机构正在天云数据的帮助下通过简单的“数据输入”和“数据输出”操作实现风险管控、智能投顾、大数据挖掘、个性化营销等等。

在某证券机构利用天云 AI PaaS 平台进行建模，在短短 2 周内就将成功预测异常交易账户的准确率提高到了 90% 以上；

在证券行业监管层面，天云数据基于资产管理产品数据，利用 AI PaaS 平台生产的风险传导模型，成功帮助机构梳理了万亿规模资产的风险传导，最大化体现了数据资产价值。

能源行业，在经历了十几年智能油田的布局之后，行业中拥有了大量的物联网传感器，实时记录着地面、地下机器的油压、温度、工况等高维度、高密度、高时效性的数据。凭借天云 AI PaaS 平台，所有数据都可以被解读，其价值也得到了最大化。甚至可以通过数据的分析和预测了解地底下的情况，而不再需要真正把设备送到地下。



除此之外，AI 方法的引入也给许多过去在 IT 投入并不大的行业带来了新的机会，比如生物制药、工业制造等等。与金融、运营商不同，这些行业的数据来自于“机器”的生产，这里的“机器”包括大量的传感器、智能设备、医疗仪器等无处不在的设备。AI PaaS 平台有望帮助这些行业实现知识再生产的一次规模性变革，创造出跟多的机会，应用 AI 机会好的企业有望实现弯道超车。



BEAGLEDATA

天云融创数据科技(北京)有限公司
beagledata technology (beijing) Co.,Ltd.



010-59625658



www.beagledata.com



北京市海淀区北四环西路67号中关村国际创新大厦503
北京市朝阳区东四环中路78号大成国际A1座9A08-9A09
上海市黄浦区河南南路16号中汇大厦二层
广州市南沙区丰泽东路106号