# 人工智能和业务就绪数据之旅始于信息架构

探索以治理和编目为核心的可信分析基础

IBM **Analytics**
Unified Governance & Integration

# 引言

## 目录



## 关键要点

- 合规性可以鼓励企业实施持续且有益的数据治理策略
- 机器学习在很大程度上实现了治理与整合活动的自动化，克服了大量数据带来的困难和人类能力的限制
- 无论是在本地环境还是多云环境中，数据治理都行之有效

各类企业中的数据数量和种类都在迅速倍增。在多云环境中，从物联网、社交媒体到移动设备、虚拟现实实施和光学跟踪等一系列数据源传入的信息流都在以指数级激增。虽然企业乐意投资于人工智能 (AI)，但大多数企业都没有进行尽职调查来了解它们的数据，或确保拥有必要的高质量数据，以便从人工智能解决方案中获益。很多企业的数据都不可访问、不可靠，抑或是不符合数据隐私和保护规则。

像《通用数据保护条例》(GDPR)、《加州消费者隐私法案》(CCPA)、巴西《通用数据保护法》(LPGD) 这样的全球性法规都专注于个人客户和员工数据。尽管不合规可能会导致严重的后果，比如降低生产率，或者损害品牌价值，但这些类型的法规也为企业实现转型和创建以数据为导向的全新业务模式创造了契机。为了履行隐私义务和保护个人信息，企业首先应发现各种类型的数据并加以分类。在收集或正确使用客户数据方面有困难的企业，可能会遇到一些亟待解决的问题。为了应对这一挑战，企业纷纷实施可管控的信息架构，在承认各类法规的同时，继续支持企业在数据推动下实现高绩效和创新。

# 可信分析基础的构建块

将数据隐私条例看作是实现数据基础架构现代化的义务和机会，将会带来显著的效益。这样做可以鼓励企业实施数据治理策略，进而创造全新业务模式，最终发掘数据驱动的洞察。统一治理与整合 (UGI) 计划应用于数据 — 无论是结构化还是非结构化数据，无论是公共云还是私有云中的数据。实施 UGI 以实现合规性，这本身就非常重要。此外，它的价值还会影响企业的其他领域，特别是对数据科学家的人工智能模型的监管。

当企业使用数据治理来给予数据信任时，用户知道这些数据来自于高质量的数据源。他们知道这些数据在整个企业中是如何使用的，也知道数据将如何增强分析项目。无论分析工具有多先进，只有拥有可信的数据，分析活动才能高效开展。可信的业务就绪数据似乎会带来无穷无尽的效益。通过分析，还可能提出新的产品设计和营销方案，有助于改善销售、供应链或客户服务计划。分析甚至可以揭示运营效率低下的环节，如果相应地消除这些低效环节，就可以提升企业敏捷性并增加利润。

企业中的数据和 AI 治理实施包含如下几个构建块。

### 数据发现和质量

企业可能并不知道自己拥有大量数据。数据治理的第一步就是盘点企业数据。先关注特定项目中的数据集，然后扩展到其他业务用例，实现更广的企业覆盖。冗余、过时或琐碎的数据（ROT）不仅存储和管理成本高昂，而且会干扰决策和运营。ROT 数据还会使合规工作更加困难，并且严重影响分析工作。数据必须满足并保持一定的质量标准，才能确保顺利运用于下游任务。

### 编目

发现数据并进行概要分析后，对其进行编目，使用元数据标记来标识数据类型、用途、所有权、数据沿袭等。由于某些行业中的企业有共同的需求，所以预构建的行业模型可以通过使用现成的业务术语和分类法加快编目过程。随着机器学习技术的发展，几小时内就可以自动映射业务术语来构建企业目录。依托 UGI 提供的编目基础，企业能够管理他们的 AI 模型、Notebook 和其他数据源，创建企业知识中央存储库。这样的基础是企业中许多数据用户的资源，包括数据工程师、数据管理员以及分析师、数据科学家和营销人员等业务线用户。

### 数据迁移、转换和同步

多个来源的数据可以根据需要以物理方式或虚拟方式轻松地集成、转换并与其他系统共享。这一过程将结构化和非结构化数据汇集在一起，并利用 Apache Atlas 和 Hadoop 等开源技术进行整合。通过创建自动化的数据流和同步，有助于确保数据湖、数据仓库、数据集市和影响点解决方案中拥有最新数据。随着数据量的增加，复制功能也与时俱进，支持低延迟的大容量复制。企业可以根据需要使用虚拟化，而无需移动数据。

## 到 2019 年，具备自助服务能力的商业用户的分析输出量将赶超专业数据科学家。

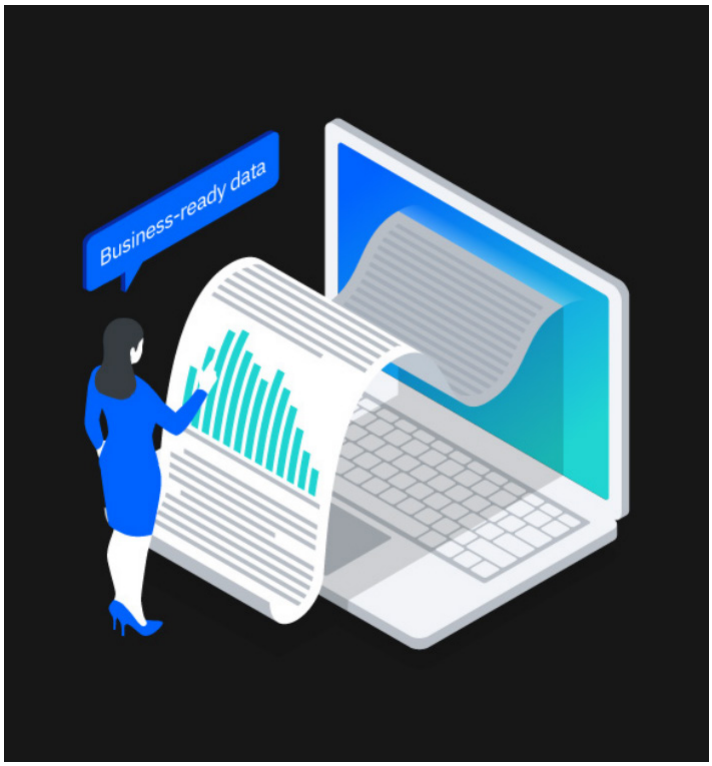
### 主数据管理

据 Gartner 预测，到 2019 年，具备自助服务能力的商业用户的分析输出量将赶超专业数据科学家。企业需要依靠全面、可信且统一的关键实体（如客户、产品和帐户）视图，这一点非常重要。现代主数据管理 (MDM) 实施伴随着基于分析图的探索、高度精确的匹配引擎、选择匹配算法的数据优先方法和机器学习支持的管理过程。此外，MDM 解决方案还具有灵活的自助服务访问、治理工具和用户友好型仪表板等功能。

### 数据隐私和保护

企业必须主动保护他们的战略资产和敏感信息资产。数据生命周期管理贯穿从创建到处置的整个流程，可采取记录管理、诉讼数据管理和归档存储等做法。现在，通过运用认知学习技术来处理企业文档和历史记录，可以根据企业的相关背景自动识别风险。

为保障业务运营与合规而治理的数据已为业务准备就绪，可随时用于任何决策、改进或创新。随着数据量的增加，复制功能也与时俱进，支持低延迟的大容量复制。企业可以根据需要使用虚拟化，而无需移动数据。



# 机器学习加速治理流程

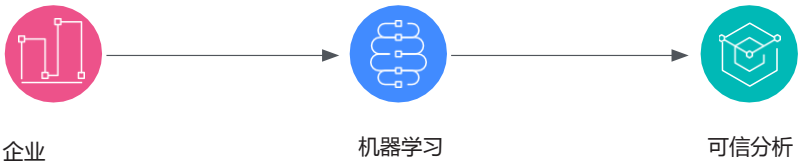
在近年来技术进步的推动下，机器学习现已增强了人类智能，弥补了人类能力的巨大短板。机器学习可以大规模实现治理与整合计划的自动化，克服海量数据带来的困难，在整个企业范围建立健全的数据治理体系。比方说，企业拥有两万个数据项，通常需要由六人组成的团队花上六个月时间对这些数据项进行人工分类，推进分析工作以可靠可信的方式开展。而在机器学习的帮助下，同样的过程只需数天乃至数小时即可完成，具体时间的长短取决于数据资产的总量。这种级别的加速使得治理过程不再背负高昂的成本。机器学习可以让合规义务更易于管理，并且也为高效分析计划铺平了道路。

**人工分类**

**需要 6 个月时间完成**


**机器学习分类**

**数天或数小时即可完成**



# 单一基础满足多种用途

如果存在可管控的基础，那么它可以跨业务单元并在整个企业中使用，如下面的常用示例所描述的那样。

### 可管控的数据湖

各行各业的许多企业都已开展了大数据预测性分析项目。关键的第一步是将海量的结构化和非结构化数据存储到数据湖中。

很多企业已经在他们的数据湖中使用了 Hadoop 或手动编码解决方案。由于缺乏治理结构，再加上缺乏用于管理数据标准、业务术语、沿袭、使用和质量等的策略，数据沼泽可能由此而形成。当用户既不理解也不信任他们的数据时，就会发生这种情况。企业逐渐意识到，他们的数据湖需要相应的策略和治理才能成功。

在可管控实施当中，数据湖中的数据可以和业务术语对应起来，不仅便于数据用户理解，而且还可以在整个企业中保持一致。将此类数据提供给用户，几乎可以加快任何自助服务数据科学、数据探索或人工智能项目实现价值的速度，并提供敏捷性。这种访问为支持多云环境、本地架构和各种数据源奠定了必要的基础。

### 应用现代化

企业正在不断加大对应用现代化领域的投入，致力于提升效率、降低成本并获得竞争优势。根据企业的具体情况，应用程序现代化的理念可以通过多种方式体现出来。首要考虑事项包括测试数据管理、数据虚拟化和连接。企业现在使用敏捷方法来测试和开发，同时实现虚拟数据访问。他们还依赖于使用灵活的整合功能连接业务应用和数据。

### 全方位客户数据视图

员工必须拥有关于客户、产品或其他实体的可信、最新且准确的单一视图主信息，这一点很重要。错误或过时的信息可能会破坏与客户的互动，影响客户的信任，最终导致客户流失或增加供应链成本。通过 UGI 流程获得的数据有助于加强客户互动，使之成为提升信任度、品牌忠诚度和公平性以及提高供应链敏捷性的有效途径。

### 企业数据仓库优化

在企业数据仓库 (EDW) 的助力下，优化架构代表了数据访问、存储、准备、治理和分析方式的升级和显著转变。最有效的优化方法之一是卸载提取、转换、加载 (ETL) 作业，以及不再使用的数据和探索性模型中所需的数据。这一过程不仅降低了成本，还使数据能够与可管控的数据湖等环境中的其他数据类型组合，从而支持动态数据探索。

### 法规合规

对于监管要求而言，可信的分析基础有助于推动和加速合规管理。最重要的是，要保护个人数据，首先要对个人数据进行定义，这样企业才能确定自己所拥有的个人数据。基本数据目录包含用于数据质量、丰富和分析的治理规则，以及合规策略。

# 结语

在企业经历数字化转型之际，企业领导逐渐认识到：对整个数据和 AI 模型（无论是在本地还是多云环境中）实施治理将会产生显著效益。通过关注核心治理实践，企业正在积极准备数据和人工智能，不仅用于分析处理和洞察挖掘，还用于做好合规准备。虽然数据量庞大，但机器学习和人工智能实践可在数据映射、编目、海量数据匹配以及保持数据质量等任务中增强人类智能。

目光远大的企业领导明白，花些时间建立稳固的 UGI 基础，会在今天和不久的将来带来显著的回报。他们认识到，如果将数据治理作为推进业务优化、开拓创新以及数据与 AI 计划合规的引擎，企业就将如虎添翼，无往不胜。采用涵盖数据操作管理（从创建到使用）的解决方案，这一点至关重要。只有经济高效地找准范围、相应扩展并适当共享，才能精简这些操作。

# 了解更多信息

IBM 统一治理与整合解决方案可以帮助您构建可信分析基础，从而推动人工智能规模化发展。

**访问网站**

ibm.com/unified-governance-integration

**探索更多内容**

深入了解认知数据治理，探究 IBM 解决方案如何在机器学习和人工智能的支持下实现业界领先。阅读白皮书。

**与专家交流**

与思想领导者、杰出工程师以及统一治理与整合领域专家互动，他们曾与成千上万的客户合作，制定了成功的数据、分析和人工智能策略。安排 30 分钟的咨询。


IBM Corporation New Orchard Road Armonk, NY 10504
美国出品 2019 年 10 月