

# 云筑底座 数创价值

数据驱动型企业发展要务与构建路径



# 目录

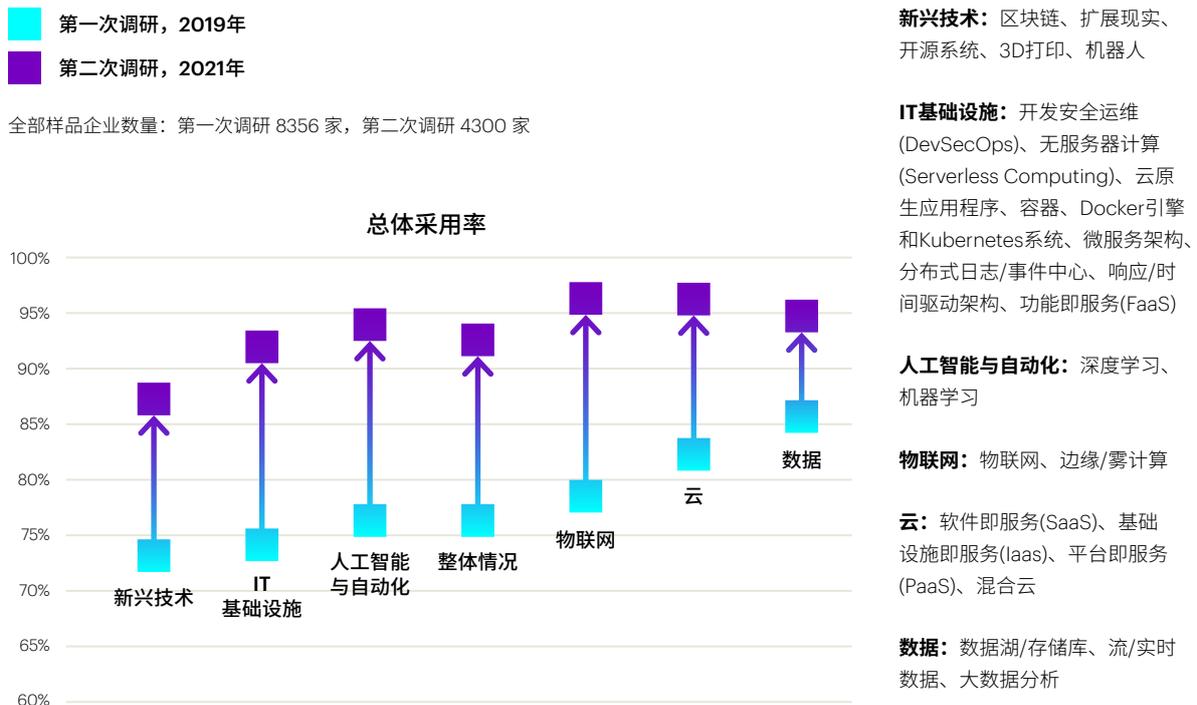
<b>01. 数字化转型提速，云释放数据价值</b>	03
<b>02. 数据驱动业务变革，实现业务战略价值</b>	06
<b>03. 建立现代化的云上数据资产</b>	07
<b>04. 制定“云上数据”战略，开启云数据现代化旅程</b>	11
<b>05. 亚马逊云科技云原生数据服务全景</b>	14
<b>06. 云数据现代化成功案例</b>	20
<b>07. 埃森哲与亚马逊云科技加速企业上云转型</b>	21

# 数字化转型提速， 云释放数据价值

根据埃森哲 2021 年研究调查，自新冠疫情爆发以来，领先企业加快了包括云计算和大数据在内的技术应用和投资周期。加大数字化转型投资力度的企业可以更快地提升系统成熟度和创新水平的能力，凭借技术投资带来的优势，在疫

情暴发后的一年中取得高于竞争对手四至五倍的营收增幅。由此可见，对诸如数据、云等关键技术的扩大投入能帮助企业加强韧性、提升效率，使其专注于增长，实现跨越式发展。

**图 1 新冠疫情加快技术应用**



来源：埃森哲未来系统系列报告《跨越发展，领军未来》<sup>1</sup>

工业和信息化部发布《“十四五”大数据产业发展规划》，将数据作为新时代重要的生产要素，是国家基础性战略资源。<sup>2</sup>《规划》要求企业围绕数据全生命周期关键环节，加快数据“大体量”汇聚，强化数据“多样化”处理，推动数据“时效性”流动，加强数据“高质量”治理，促进数据“高价值”转化，将大数据特性优势转化为产业高质量发展的重要驱动力，激发产业链各环节潜能。企业亟待将数据转变为前瞻性洞见及差异化资产，充分发挥数据的要素价值。

而数字化转型的关键催化剂——云计算能够使企业轻松测试和验证新想法，重构基础设施、应用程序和数据，加速释放企业上云的无限潜能。在全新的云环境中蓬勃发展、增效创新，是企业更需要关注的领域。在云的催化作用下，企业最终将蜕变为云原生企业，通过对基础设施、应用程序和数据进行现代化改造，最大程度地挖掘云环境的价值，加速数字化转型进程。

## 紧抓核心要素，打造数据资产

数据是企业运营变革和构建竞争优势的核心动力。数据对提高生产率及资源配置的效率、优化经济结构的作用凸显，已成为新型生产要素和重要资源。**到 2022 年，90% 的企业战略明确将数据作为关键的企业资产。**<sup>3</sup> 打造企业的数据资产有利于：

- 构建企业共通的数据语言：企业原有的分散的、不统一的数据使得数据分析者往往需要花费大量的时间和精力收集各类数据，并寻找共通的可以公共使用的数据，才能将其归类并分析应用。数据资产化可形成共同的“数据语言”，由此，企业的管理层可以更高效地对数据进行使用和管理。
- 转化为企业战略性资产：数据资产化之后，数据资产会渐渐成为企业的战略资产，企业将强化对其分析洞察，进而极大地提升企业核心竞争力。
- 加速数据资产价值提升：目前在缺乏标准规则的情况下，数据使用部门承担了较高的使

用成本，制约了数据资产的流动，但随着数据资产管理的完善，必然能加速数据资产交易的价值提升。

由此，以数据为企业核心资产，充分释放其在业务中价值，加速企业数字化转型步伐。

然而，许多企业都面临同样的难题：如何利用海量数据，打造差异化的竞争优势？虽然企业普遍承认数据的价值，并且正在使用更先进的技术来捕获数据，但正如 Gartner 所指出的，近 97% 的数据尚未被企业使用，超过 87% 的企业在商业智能和数据分析领域的成熟度偏低。<sup>4</sup> 传统的各种专有系统数据来源在业务逻辑和规则记录上存在差异，这阻碍了从数据来源到数据价值的处理流程。业务分析报告所需数据转换周期可能需要数天，严重影响到销量、生产等业务运营。为了克服这些挑战，企业需要借助云数据现代化战略从传统的数据架构升级为现代化、面向未来、灵活可扩展的云上数据架构。

## 基于云数据现代化战略赋能业务发展

数据存储云化已成为中国企业加“数”转型的主要发展趋势之一，与生活息息相关的人脸、指纹、摄像头、自动驾驶等传感器和 RFID 前端设备产生的数据，均通过网络传输到接收设备转换成数据后予以存储。数据存储已由原来的存储在独立服务器上，向存储在云（私有云或公有云）上转型，以降低建设费用、维护费用及安全

费用等，提高可维护性、可扩展性、安全性、利用率及负载的均衡性。

基于云打造现代数据战略，可以利用云的弹性按照业务变化及时响应业务对数据存取的需求。云是创新的催化剂，可以提高速度和敏捷性。将数据放在可扩展性、可信度和安全性极高的云端可以为企业带来明显收益，云数据战略已经成为企业现代数据资产管理、数据变现的目标。

云上提供了全面的数据存储和处理的工具，例如不同类型的数据库，可以更便捷地响应不同业务对数据的处理要求。基于云上数据服务消除存储和计算的限制，借助云原生工具强大功能，使企业能够打破现有数据局限，摆脱传统平台的束缚，避免业务中断。

数据在云上更容易构建一个打破数据孤岛的统一分析平台，实现数据分析的普惠化。无限存储和无限计算提高其延展性，且提供了自助式的获取数据和随时访问数据市场的服务，降低了成本和管理费用。

利用云上的机器学习和人工智能工具，更便捷地进行创新。将宝贵的 IT 资源集中开发企业与众不同的业务及转变客户体验的应用程序，而不是将其用于管理基础设施和数据中心。

由此可见，企业基于云数据现代化战略进行业务创新已是必然趋势。



## 数据驱动业务变革， 实现业务战略价值

在提升业务、节约成本、灵活响应以及激发创新上，数据都有着十分重要的战略价值。例如，埃森哲正帮助某知名国际运动用品公司提升其数据能力。

### 降本增效

使用类似现收现付的新模式能够更简洁、更便宜，更快速地面面对经济不确定性，提高操作效率，降低成本。埃森哲帮助该运动用品公司将其数据仓库从本地数据库迁移至亚马逊云，借助亚马逊云科技强大的云能力，企业大幅提升数据处理能力，显著减少硬件及存储运维成本。

### 安全合规

构建和优化数据模型以灵活响应客户需求，同时确保数据安全及监管要求。为遵守各种数据安全及监管法律法规，在处理客户数据时，必须使用优化的隐私保护、数据安全和治理措施；利用新技术、新模型为未来的业务建立灵活的数据能力。埃森哲在帮助该运动用品公司数据上云的同时，对数据模型进行了优化及重构，使得数据更易于业务部门使用，对数据处理过程也强化其标准化和安全机制，以提升数据安全性及可治理性。

### 激发创新

通过打破孤岛和扩展数据访问来实现强大的业务智能决策，预测业务结果、优化和重塑流程，以敏捷的方式快速创造新产品和新的收入来源，从而缩短产品、服务的开发周期和上市时间，

提高整个企业的创新能力。例如，该运动用品公司基于重构和优化后的云上数据，已激发多种新的数据使用场景，优化其商业分析与预测能力。

### 业绩增长

利用数据作为核心资产，可深入理解客户需求，洞察客户购买行为和倾向，以此提供超个性化体验，提高市场占有率，不断加强和拓展客户关系，推动更高的收入增长及市场份额。埃森哲已帮助该运动用品公司使用亚马逊云科技各种大数据及高阶数据分析服务，包括机器学习在内的一系列高阶用户数据分析，提升企业营销能力。

## 制定云上数据战略

报告显示，全世界每天生成的数据量高达五百兆字节。<sup>5</sup>然而，能够利用自身和外部数据打造竞争优势的企业寥寥无几。**调查发现，84%企业缺乏云上数据平台的能力，<sup>6</sup>**他们往往沿用旧平台，缺乏敏捷且安全的数据平台和合理的数据架构，且难以保证数据质量，因此无法高效采集和处理各种数据、交付高质量洞察结论以及实现高度敏捷。

企业已经意识到基于云的现代化数据平台是实现向数据驱动型组织转变的关键举措。埃森哲的一项研究表明，高绩效企业与成为“数据驱动型企业”之间存在直接关联。数据驱动型企业将云作为催化剂，最大利用数据价值。

## 建立现代化的云上数据资产

数据驱动型组织将云作为催化剂，最大化发挥数据价值。他们将数据视为企业资产，根据数据的完整性和质量差异进行分类管理和变现，把数据作为创新、业务敏捷和关键业务决策的根本依据，通过人工智能和数据分析来提高效率。他们将基于云的数据和预测性分析作为重中之重，并强调利用洞察推动运营优化、前瞻性规划和持续学习。数据驱动型企业的盈利能力、客户获取率和客户留存率都优于竞争对手。

相较于部分从一开始就积极采用云计算的企业，大部分企业最初都通过本地计算进行数据分析和处理，随后才因特定需求将业务数据迁移至云端。但在今天的数字经济时代，本地数据架构的局限性日益凸显。

通过合理规划和实施，将企业数据迁移至云端可以带来诸多优势，精心设计的云上数据架构可为企业提供亟需的可扩展性、灵活性以及基于数据的可靠洞察，助力企业更快取得业务成果。

- 定制化的云上数据架构可实现实时地数据捕获，缩短数据处理时间最多可达 50% 以上；<sup>7</sup>
- 借助人工智能、机器学习等云端的智能数据服务，快速识别可变现的数据资产；
- 获得智能业务洞察，更快推动业务创新。

埃森哲的云数据现代化策略将帮助企业完成云上数据之旅，包含数据迁移上云、数据现代化和成长与创新三大支柱。

**数据迁移上云：**利用工具对数据上云进行发现与规划，然后对数据、代码以及数据应用 (Data Consumption) 进行自动化迁移，最后淘汰遗留的本地数据环境。

**数据现代化：**基于云为专门构建的数据产品开发可配置的可代码数据供应链，包括业务定义的数据载入、数据质量、数据产品规则、数据生命周期和数据可访问性。虽然云计算为服务现代化和运营转型提供了新的机会，但安全性和遵从性风险仍然是采用云计算的最大障碍。企业应采取云安全最佳实践，采取安全连接、安全着陆区配置策略，注重安全可使其成为云现代化的最大加速器。

**成长与创新：**在整个业务中实现 AI 模型的工业化和扩展。使用快速数据实验室实现精益实验和新想法的开发。借助数据文化和素养在整个公司范围内扩展以数据为主导的创新。

# 数据迁移上云

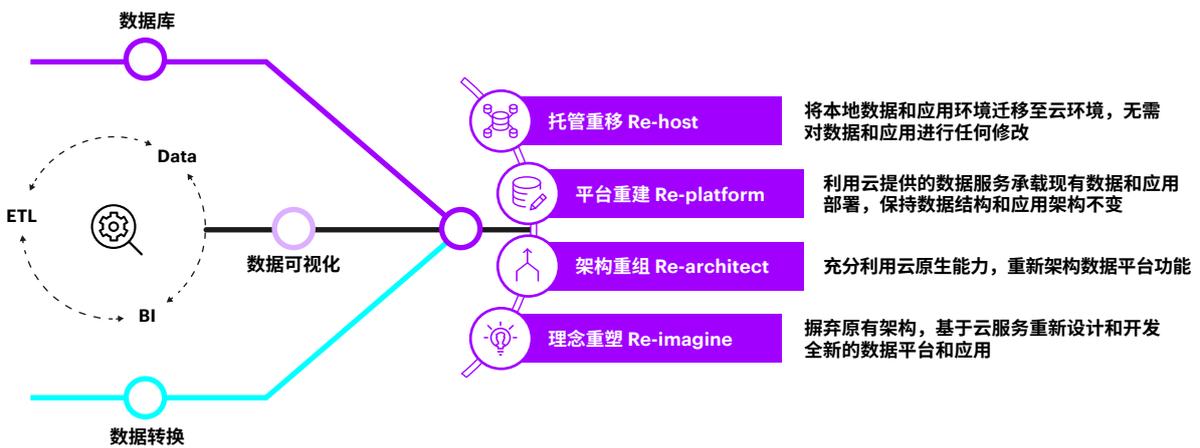
云数据现代化的首要环节是数据迁移上云，具体可参考 4 种迁移模式——托管重移 (Re-host)、平台重建 (Re-platform)、架构重组 (Re-architect) 和理念重塑 (Re-imagine)，每种方法都要评估相应策略和工作量。

- 托管重移是将本地数据和应用环境迁移至云环境，无需对数据和应用进行任何修改。

- 平台重建是利用云提供的数据服务承载现有数据和应用部署，保持数据结构和应用架构不变。
- 架构重组是充分利用云原生能力，重新架构数据平台功能。
- 理念重塑是摒弃原有架构，基于云服务重新设计和开发全新的数据平台和应用。

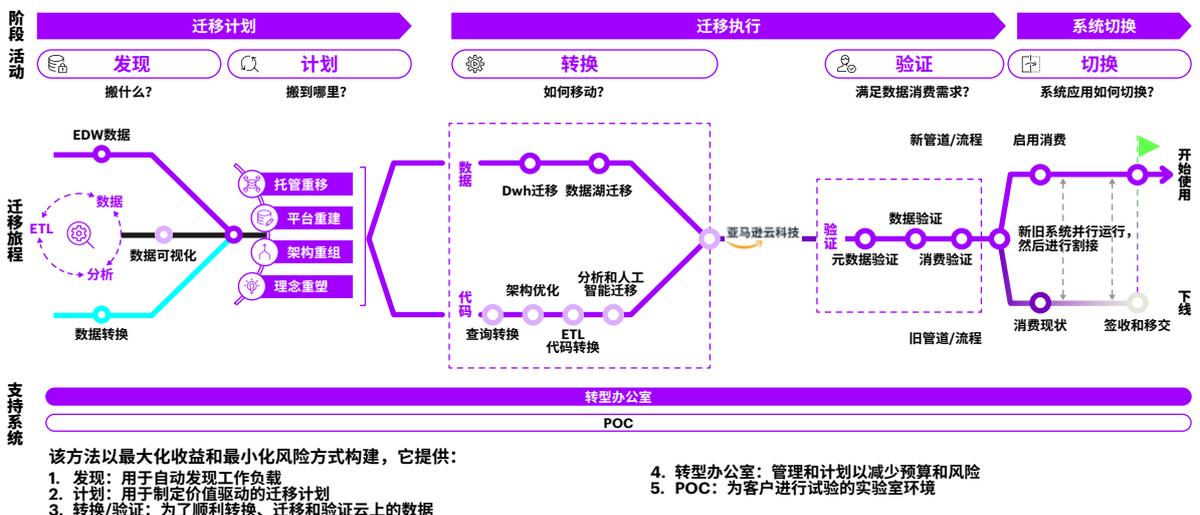
## 图 2 数据云迁移模式

借助 4R 迁移模式，将有助于企业明确其基于云的数据架构实现的方式。云上数据，可以按照 3 个阶段、5 项关键活动展开。



来源：埃森哲

## 图 3 数据云迁移实施路径



该方法以最大化收益和最小化风险方式构建，它提供：

1. 发现：用于自动发现工作负载
2. 计划：用于制定价值驱动的迁移计划
3. 转换/验证：为了顺利转换、迁移和验证云上的数据

4. 转型办公室：管理和计划以减少预算和风险
5. POC：为客户进行试验的实验室环境

来源：埃森哲

### 阶段一：计划

- **发现：**对本地的数据、代码以及数据消费应用进行分析，为制定迁移路线图提供洞察及详细依据。
- **计划：**充分考量各种可能场景，为数据上云制定合适的路线图、项目计划及预算。

### 阶段二：执行

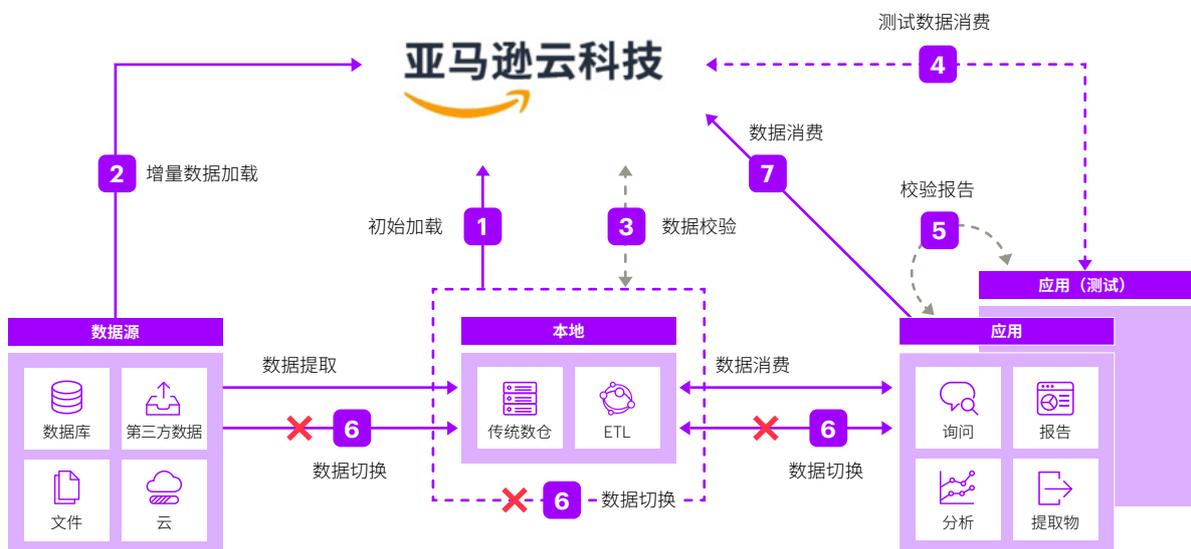
- **转换：**根据当前和目标架构，尽可能自动化地迁移数据，代码以及数据消费应用。
- **验证：**利用各种工具，检查源数据的准确性和质量，帮助实现数据迁移的“最后一公里”，实现数据协调、测试和迁移后校验。

### 阶段三：切换

- **切换：**明确定义切换步骤，指导和管理迁移切换环节，降低切换风险。

图 4 数据上云系统切换步骤

从简单到复杂、从点到面、从外围到核心，逐步建立并优化云数据平台。



来源：埃森哲

## 数据现代化

数据现代化为现代数据产业的设计和规划提供了工具驱动的方法，全面的现代化咨询以数据成为资产为目标，助力智能企业并以此盈利。数据现代化包括数据策略和架构、数据基础、数据产品工厂、数据产品商务、数据安全。

- **数据策略和数据架构：**为可扩展、功能丰富的现代云数据平台建立设计、架构、业务案例和计划，跨数据架构、治理和运营模式的数据战略，以建立现代的、基于云的企业数据平台。
- **建立数据基础：**建立云数据基础，包括基础设施、数据安全和摄取框架，通过 API 进行的数据和元数据集成，以及用于治理和平台流程的工作流编排，推动数据平台和集成的可扩展性。
- **创建数据产品工厂：**开发用于创建具有透明的业务规则、数据质量规则和业务条款规范的数据产品的工厂。行业数据管道自动创建从摄取到转换的数据产品，而企业拥有数据作为战略性资本来源。轻松构建和管理所有企业数据产品。

## 成长与创新

企业利用大数据、自动化和人工智能来提高生产率，推动业务增长，但也确保为下一代工人创造新的机会。同时致力于数字化培训，包括通过“实践”学习数字化，并向员工灌输数字化思维方式，帮助员工提高使用数字技术技能。利用云平台原生服务构建 AI 与数据科学平台，并助力其实现独特的创新成果，保持企业可持续成长，要特别强调以下因素：

- **数据素养：**在组织的各个方面将数据和洞察力的使用制度化，为了创建和维护数据驱动的文化，数据必须可供用户使用，并且必须培训如何使用数据。可以通过文化和变革管

- **启用数据产品商务：**为数据消费者、生产者和平台管理员创建单一、集中的用户体验。自动配置数据、工具和沙箱的用户。及时提供可信赖的数据进行分析。
- **云数据安全：**现代化数据安全可以通过多种方式实现安全保障。
  - **快速：**使用云服务提供商 (CSP) 的本地加速器，可以在几分钟或几小时内内部署安全能力和控制。
  - **无摩擦：**在现有解决方案、业务流程和运营团队中嵌入安全性。
  - **可扩展：**通过应用自动化和自修复过程，减少手动步骤，打破增加员工数量的资源模式，使组织能够扩展。
  - **先发制人：**建立先发制人的控制，从一开始就阻止意外或恶意安全事件的发生。
  - **成本效益：**从一开始就做好安全准备，避免重复工作带来的额外成本。

理、学习和发展、数据教育以及自助服务工具和技术来促进数据素养。

- **快速数据实验室：**快速灵活地构思和试验创新想法，获得弹性资源。通过快速数据实验和原型设计体验成功，从而以最小投资、最低风险快速实现规模扩展。
- **工业 AI：**加速高级分析的步伐，支持业务 AI 需求。利用诸如对话式 AI 的一些 AI 平台，扩展企业的工业 AI 水平，需要大量的 AI 实践以及涵盖数据、人才、流程和技术资产的方法框架和自动化能力。

# 制定“云上数据”战略，开启云数据现代化旅程

“知识就是力量”，数据能够给一个企业的员工带来力量。如果只有少数人拥有数据访问权限，这少数人就代表了全部的洞察和决策能力，而企业中的其他人则会陷入盲目，形成发展瓶颈。

但是，多数情况下，企业管理者并不会刻意隔离数据和信息。造成这种问题的原因，往往是因为组织缺乏一套有效的数据战略，数据基础设施无法满足员工能够在适当的时间，安全地访问和分析数据，并且利用有效的工具去得出正确的业务洞察。

## 成为数据驱动型企业需要三步：

### 1. 将数据迁移至云端，构建现代化数据基石

在云中，企业可以通过互联网获取存储、数据库、分析和机器学习(ML)等 IT 资源，而无需购买、拥有或维护物理数据中心和服务器。通过云，企业可以一键扩缩存储容量，且仅需为自己使用的存储容量付费(价格实际上比企业自己存储数据更低)，同时，与其他方案相比，企业还可以获得更多的数据工具。

对于首次进军云端的企业来说，提升和转移(Lift and Shift)策略是一种常见的举措。例如，在企业的托管设施中运行的 Oracle 数据库迁移到云上运行。这种提升和转移策略可能是移动到云端的最简单方式。企业的架构图看起来仍然相似，团队的职责没有改变。然而，企业并没有得到云的全部好处。

亚马逊云科技最近委托 IDC 进行的一项研究考虑了使用亚马逊关系数据库服务(Amazon RDS)的客户。研究发现，三年内，这些客户的数据库操作成本降低了 39%，投资回报率为 264%。对于大多数公司来说，数据库管理是一项无差别的繁重工作。企业的数据是公司拥有的

最有价值的资产之一，但数据库的管理并不是企业提供价值的地方。

在数据库领域，过去的几十年里，客户不得不与昂贵的、专有的、具有高锁定性的、实施惩罚性许可条款的传统商业数据库厂商打交道。由于传统商业数据库供应商的一些做法，我们看到许多客户试图尽可能快地转移到开源数据库，如 MySQL、PostgreSQL，然而开源数据库往往无法满足客户对于性能和扩展的一些需求，数十万客户通过迁移到托管型数据库服务实现了节省，这些服务融合了开源数据库的灵活性和低成本以及商用数据库稳健的企业功能集。

同时，在过去的十年里，数据库已经发生了很大的变化。传统的应用程序，如 ERP、CRM 和电子商务大多使用关系数据库。但是，构建应用程序的方法和应用程序需求本身已经发生了变化，开发人员越来越多地使用微服务体系结构构建应用程序，并选择专门为满足应用程序特定需求而构建的关系数据库和非关系数据库，如键值类型数据库和文档数据库。

## 2. 将数据作为资产，发掘数据的价值

数据驱动型企业将数据视为整个企业的资产，而不像传统企业那样，将其视为个别部门的财产。确保整个企业的员工而非仅仅数据科学家或开发人员能够安全、简单地访问数据，有助于提高业务绩效并促进创新。

在传统数据架构下，对数据访问进行管理，程序上通常非常复杂，因为用户可以从许多不同的位置对数据发起访问，而这些都是需要规避的安全风险。为此，企业需要授予、追踪和撤销员工对不同数据存储库的访问权限，以及为了遵守日益复杂的法规要求，审查谁拥有对哪些数据的访问权限，这是一项极其繁重的工作。实际上，许多规模较大的企业都设有专门的部门来处理这些事务。

企业需要利用这些工具打通数据孤岛、提高数据对需求方的可访问性，从而“解放”自己的数据。在云中，企业可以将数据移动至任意地点，包括中央数据资源库——“数据湖”、专门构建的数据库，以及专为特定分析目的（如数据仓库或日志数据分析等）建立的极具成本效益的分析工具。这样，企业中的用户便能以安全、受控的方式，实现对数据的无缝发现、访问和分析，而不受数据存储位置的限制。

要实现赋能员工、优化决策的目标，不仅需要向员工授予数据访问权限，同时也要为他们提供数据洞察。例如，许多亚马逊云科技的客户都使用商业智能(BI)工具，以便所有员工和利益相关方(而不仅仅是技术人员)都能简单、快速地找到对自己最有用的数据洞察。同时，我们的许多BI工具都利用机器学习(ML)进行预测，以进一步帮助用户做出决策。

因此，企业正在加速迁移到云，基于“智能湖仓”方法实现数据库、数据湖、数据分析、机器学习之间的“统一数据”。这样，企业可以打破孤岛，使数据可以跨多种服务投入使用。

### 3. 利用云上的人工智能与服务， 利用数据进行业务变革和创新

最后，企业需要基于这些数据，使用分析工具、人工智能和机器学习进行创新。机器学习是我们这一代最具颠覆性的技术之一。从预测制造问题到定制医疗体验，企业正以全新的创造性方式，利用机器学习进行创新，并确立自身的竞争优势。麦肯锡近期进行的一项调查显示，使用 AI 的企业中，有 63% 的受访企业表示采用 AI 让业务单位的收入得到了提高。<sup>8</sup> 同样明显的是，采用者相信今后人工智能技术将能持续为企业带来更多价值和优势。根据 Gartner 的数据，到 2024 年年底，75% 的企业将从 AI 试点转为具体实施。<sup>9</sup>

机器学习之旅并不总是一帆风顺。要通过机器学习实现成功，不仅需要出色的技术，还需要确保企业上下朝着正确的目标努力。为了确定并实现这些目标，需要对各种流程、管理方式和文化做出大量的改变。在后面几节中，我们将探讨组织该如何克服通常会阻碍进步的常见挑战，以及应该采取哪些妥善措施，才能以高效且可持续的方式实施机器学习。

要发挥机器学习的全部业务潜力，需要在团队组织、目标和愿景方面进行文化变革。为了让机器学习在企业中实现全面普及，业务和技术团队必须共同努力并共享相同的优先级。为此，企业最高层级必须从一开始就全力支持机器学习实施工作，由执行负责人制定目标，并对有助于机器学习实施的技术和流程进行投入。在制定机器学习计划时，管理层必须做到目光长远。高管们必须坚定目标，但在实现企业目标的方式上可以灵活处理。犯错在所难免，但是，通过专注于长期愿景并不懈努力，企业可以吃一堑，长一智，并将这些经验教训应用到整个企业的机器学习文化中。

许多企业仍然不确定在哪些领域应用人工智能和机器学习，才能产生有效的影响。企业

需要在业务价值与快速取得成效之间找到平衡。由孤立的数据工程师团队来构建一个未知业务价值的试点 (PoC, Proof-of-Concept) 项目，不太可能在企业中产生足够的影响力。如果希望能够快速拿到企业在人工智能和机器学习领域的资源投入，企业应该展示的是人工智能如何能解决当前所面临的实际业务问题。同时，需要找到一个可以充分利用已有数据并且在 6 个月以内可以完成的使用案例。检验采用人工智能和机器学习的方式是否比传统的方式可能会更理想，更有成效。这些潜在的收益可能是企业获得新的业务增长方向、运营效率的增加和成本的降低、或是终端客户体验得以改善等。要满足上述所有条件，最好的方法就是让技术专家和业务领域专家在机器学习项目中携手合作。技术专家可以执行可行性评估，而业务领域专家可以确保解决方案能够解决实际业务问题，并产生实际影响。

同时，也必须注重工程师团队的培训，助其掌握相应专业技能。企业越来越明显地意识到机器学习技能缺口，即技术与企业内部 IT 专家能否充分利用这些技术的能力之间所存在日益扩大的差距。O'Reilly 《2021 年企业中的 AI 采用》报告调查了超过 3500 位业务领导者，调查结果表明，AI 面临的巨大挑战是缺乏熟练人才和招聘困难，有 19% 的受访者认为这是一个很大的障碍。<sup>10</sup> 要弥合机器学习的这一缺口，就必须同时做好培训和招聘工作。然而现实情况是，在机器学习变革即将到来的今天，数据科学家供不应求。这就要求那些想要充分利用机器学习的组织首先将资金投入到了人才培养之中。尽管没有一种机器学习技能缺口解决方案能放之四海而皆准，但还是有一些行之有效的方法可以最大限度地提高现有员工的能力，从而在引入或借用预先培训的专家人才方面降低投资。

# 云上数据服务及人工智能突破 既往局限，推动业务创新

基于现代化云数据架构，实现数据驱动型企业，需要通过云上数据服务集成整个业务链条数据，为由业务专家、数据科学家和技术专家组成的多学科团队提供便利的分析体验，突破历史积

累的技术约束和局限，并通过更为高级的预测，识别、优化等人工智能服务，推动业务创新。亚马逊云科技为满足上述需求，专门构建了云上数据服务和人工智能。

## 借助亚马逊云科技数据服务进行数据架构的现代化改造

亚马逊云科技通过对数据基础设施进行现代化改造，企业可以从本地数据存储转移到基于云的数据库、存储和分析环境中。为了部署支持当今业务和 IT 需求的数据和分析基础设施，企业需要对以下四个主要组成部分进行现代化改造。

### 专用数据库

随着数据量从 GB、TB 增加到数 PB 甚至更高，传统的数据库解决方案和工具现已无法满足要求。典型的案例就是关系数据库曾经是默认选择，如今只不过是更经济地构建高弹性、可扩展的应用程序的选择之一。与其使用 All-In-One 的一体式数据库，不如考虑“根据业务目的选择专门构建的数据库”的理念。选择合适的专用数据库取决于三个因素。

**应用程序工作负载：**了解存储的数据类型和访问模式。这些可分为三类：事务型（适用于大量并发应用程序）、分析型（汇总和总结大量数据，每次查询都要对许多行进行操作）和缓存型（适用于需要提高加载速度以缩短终端用户响应时间的读取量较大的工作负载）。

**性能和规模要求：**不仅要考虑数据库的速度，还要考虑它如何服务于终端用户。开发人员

应询问数据库是面向客户，还是仅供内部使用。从地理上看，如果数据库离用户更近，那么可以缩短响应时间。

**数据类型：**了解实体的类型及其关系。

- 关系数据库将数据规范化为单独的数据表。
- 键值或宽列数据库专为扩展而设计，数据分散在多个存储节点上。
- 文档数据库适用于大型记录，这些记录是针对频繁访问的数据以异构方式组合而成。
- 图数据库强调数据之间的关系，以发现关系。这方面的示例包括社交网络关系或欺诈检测服务。
- 内存数据库专为缓存、会话管理、排行榜和地理空间应用程序而设计。
- 时序数据库非常适合物联网 (IoT) 应用程序、DevOps 和工业遥测。
- 分类账数据库非常适合记录系统和银行交易系统。

**图 5 亚马逊云科技专门构建数据库引擎概览**

亚马逊云科技提供超过 15 个专用引擎来支持各种数据模型，包括关系、键值、文档、内存、图、时间序列、宽列和分类账数据库。

**亚马逊云科技专门构建的数据库服务**

	 <b>关系数据</b> 参照完整性 ACID 事务 写时模式验证	 <b>键值数据</b> 高吞吐 读写低延时 无限扩展	 <b>文档数据</b> 文档级别存储 可基于任何属性 快速查询	 <b>内存数据</b> 按键 key 查询 达到微秒级 延迟	 <b>图数据</b> 快速轻松地 创建和导航 数据之间的 关系	 <b>时序数据</b> 收集、存储 和处理时间 序列数据	 <b>分类账数据</b> 保存对应用 数据的所有 更改的完整、 不可篡改和 可验证的记录	 <b>宽列数据</b> 可扩展、 高可用、托管的 Apache Cassandra 兼容的服务
<b>亚马逊云科技服务</b>	Amazon Aurora, Amazon RDS	Amazon DynamoDB	Amazon DocumentDB	Amazon ElastiCache, Amazon MemoryDB	Amazon Neptune	Amazon Timestream	Amazon OLDB	Amazon Keyspaces, Managed Cassandra
<b>适用场景</b>	传统应用 ERP CRM	高流量 Web 应用 电子商务系统 游戏应用程序	内容管理 个性化 移动应用	排行榜 实时分析 缓存	欺诈检测 社交网络 推荐引擎	物联网应用 事件跟踪 工业遥测	系统记录 供应链 医疗保健 银行事务	设备维护、 队列管理和 路线优化的 大规模工业 应用程序

来源：亚马逊云科技

**数据仓库**

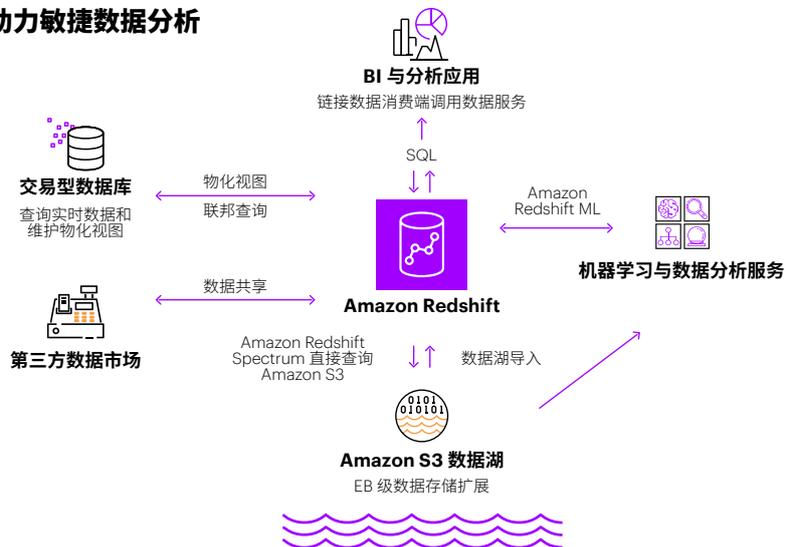
本地数据仓库系统也无法跟上数据量的增长和复杂性的增加。维护、更新和保护僵化的架构需要大量投资，这些架构无法为企业提供充分利用自身数据的机会。迁移到云数据仓库可以消除这些限制。企业可以对数据仓库中的数 PB 数据运行查询，并延伸到数据湖中。现代数据仓库可自动执行大多数常见的管理任务（比如管理、监控和扩展），让 IT 团队能够专注于更高效、更有价值的工作。它还提供快速查询性能，提高 I/O 效率，并随着您的性能和容量需求的变化而

扩大或缩小。一种简单的策略就是选择将现有的数据仓库迁移到云端，但更好的方法可能是先从当前基础设施不能很好地满足要求的场景开始。

Amazon Redshift 专注于通过适用于所有人的简单分析方法在数秒内从数据获得洞察。无需考虑管理数据仓库基础设施。提供跨数据库、数据湖、数据仓库和第三方数据集交叉分析的能力，协助企业分析不同来源数据，并且通过自动优化提高查询速度，获得比其他大规模云数据仓库高达 3 倍的性价比。

**图 6 云原生数据仓库 Redshift 助力敏捷数据分析**

**打破数据孤岛，跨存储服务、数据库和数据湖的数据分析**



来源：亚马逊云科技

## 存储

本地存储可能既昂贵又复杂，具有成本高昂的硬件更新周期并需要迁移数据来支持系统升级。此外，如果企业的数据困在各个孤岛或其他“泥潭”中，则很难从数据中获得洞察。企业将存储转移到云上，就可以随时调整并按需付费，降低总体拥有成本，且不必受限于硬件更新。同时，它还可以使企业保持敏捷，提供无限的可扩展性，同时消除数据孤岛。

## 分析系统

在本地管理开源分析软件(如 Apache Hadoop/Spark、Elasticsearch 和 Apache Kafka)需要维持一个专门的专家团队来管理硬件和软件配置，修补和备份，性能调整和优化以及针对未来增长进行容量规划。而迁移到云中的托管分析可以节省时间，降低成本，并显著提高工作效率。

# 借助“智能湖仓”架构实现统一数据分析与敏捷创新

为了适应未来发展，企业必须制定数据策略。为了加速部署任何重大业务举措(从打造新的客户体验到开辟新的收入来源)，领导层必须能够快速获取专业见解、洞察真相，利用信息把握机遇并预测未来趋势，以及尽早抓住机遇采取行动。这就要求我们的解决方案可以：

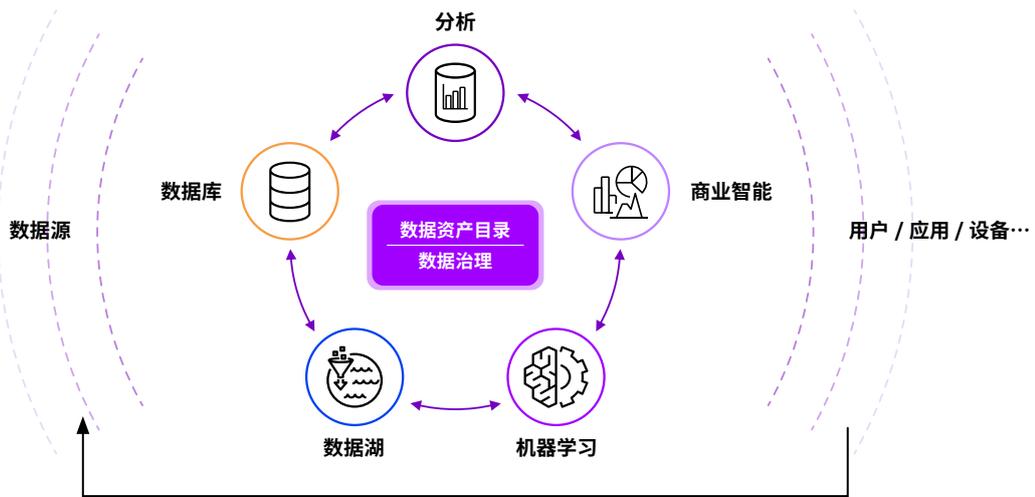
- 以开放格式存储任意数量的数据；
- 打破彼此隔离的数据孤岛；
- 赋予员工借助偏好的工具或技术运行分析或机器学习的能力；
- 采用合适的安全和数据监管控制措施，管理特定数据的访问者。

在数据架构上采用千篇一律的方法会导致各方面性能互相制约。因此，亚马逊云科技的智能湖仓架构并不仅仅是将数据湖与数据仓库集成，而是通过一种方法将数据湖、数据仓库和专用存储全方位整合起来，实现统一的监管和轻松的数据移动。不仅解决了数据引力问题，还可满足合

规性、安全和监管要求。它为广泛深入的专用数据服务集合提供了一种架构，可以支持组织的企业再造、创新和分析目标。它还提供了高性能、经济高效且可扩展的数据湖存储解决方案。亚马逊云科技的智能湖仓方法提供了下列五种支持：

1. 可用于快速构建可扩展的数据湖；
2. 支持专用数据和机器学习服务，根据企业的使用案例提供高性能；
3. 确保企业可以无缝地在数据湖和专用数据服务之间移入、移出数据，以及围绕周边移动数据；
4. 始终遵守数据访问的安全、监控和管理的合规性要求；
5. 提供经济高效的性能和可扩展性。

图 7 云原生数据之旅

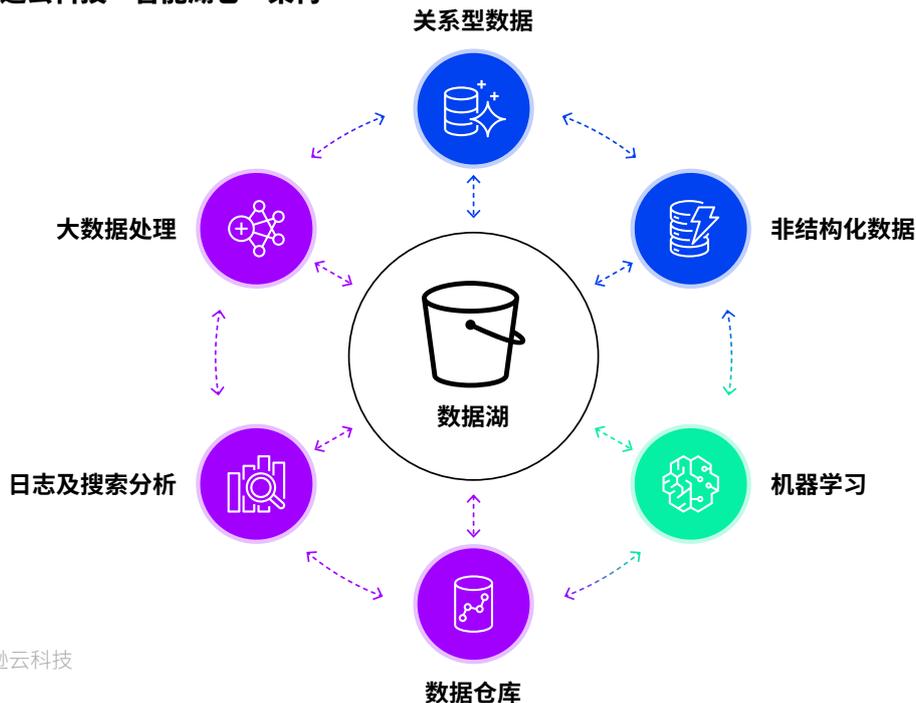


来源：亚马逊云科技

Amazon S3 解决方案中提供了数据湖，可供企业在任意规模下存储和检索任意类型的数据。它具有无与伦比的持久性、可用性和可扩展性，最佳的安全性、合规性和审计功能；同时能以最低成本获得最快性能，且拥有最多的数据导入方法。企业可以使用 Amazon S3 手动构建数据湖。不过，Amazon Lake Formation 是一种更为简单、直接的解决方案，其开发的目的是为了消除自行构建过程中的手动任务。它能帮助企业在数天（而非数月）内在云中启动安

全的数据湖。所有手动和费时的任务均由 Lake Formation 自动完成并进行优化。它从数据库和对象存储收集数据并建立目录，将数据移动到 Amazon S3 数据湖，使用机器学习算法对数据进行清理和分类，并保护对敏感数据的访问。由于其易用性、持久性、可用性和可扩展性，Amazon S3 和 Amazon Lake Formation 是为智能湖仓方法建立数据湖的理想解决方案。

图 8 亚马逊云科技“智能湖仓”架构



来源：亚马逊云科技

当数据存储在多个不同的系统中时，企业需要具备在所有服务和数据存储之间灵活自由地移动数据的能力。Amazon Redshift 和 Amazon Athena 均支持联合查询，此功能可跨运营数据库、数据仓库和数据湖，对其中存储的数据运行查询，从而跨多个数据源提供见解，无需移动数据，也无需设置和维护复杂的提取、转换和加载（ETL, extract, transform, and load）管道。利用 Amazon Redshift 数据湖导出功能，企业可以将数据从数据仓库以开放格式分载到数据湖中，直接可供分析使用。企业还可以通过数据功能系列的一个关键组成部分 Amazon Glue，利用亚马逊智能湖仓架构在所有服务和存储之间无缝移动并存储数据。Amazon Glue 是一种无服务器数据集成服务，让企业可以轻松地准备数据用于分析、机器学习和应用程序开发。Amazon Glue 提供了数据集成所需的全部功能，这样企业只需数分钟即可获得见解。

为了让开发人员、业务分析师和数据科学家能够打破数据孤岛，并可通过安全可控的方式发现、收集和分析数据，通过 Amazon Lake Formation 等功能，包括可自动发现数据、标记数据并对数据进行编目的数据目录，还提供了简单的方法来集中定义和管理安全性、监管与审计策略，而所有这些都集中在一个方法中实现。这样，企业可以面向合适的用户，在合适的时间提供细粒度控制的数据访问，并高效地满足其法规监管与合规性要求。

利用依托亚马逊云科技的多种专用服务组合打造的智能湖仓架构，企业可以面向未来打造产品、引领行业的发展、带着无限革新的可能性走向成功的彼岸。

## 利用亚马逊云上机器学习工具推动业务创新

机器学习可以推动企业创新，实现业务变革，甚至为企业带来革命性的影响。在推动机器学习的创新之前，企业要预先对业务目标和衡量标准作出规划，从而制定好衡量成功的 KPI。相较于在给定时间范围内实现目标财务 ROI，如果企业能通过业务敏捷性、竞争优势和风险承受能力方面衡量机器学习计划的影响，则能取得更大的成功。迭代和学习机器学习的潜力得到证明后，下一步就是从试点过渡到生产，这可能包括将机器学习功能集成到更大的 IT 系统中。此举通常比试点过程更为费时，并且会因整个系统的复杂程度和生产部署的规模而有所差异。成功的机器学习计划不仅仅需要适当的工具，一个整合数据存储、安全、分析服务的综合性平台，以及用于训练和部署的计算资源更为重要。在云上实现机器学习的项目则可带来一系列好处，包括速度、可扩展性、灵活性、弹性、安全性和成本节省等等。

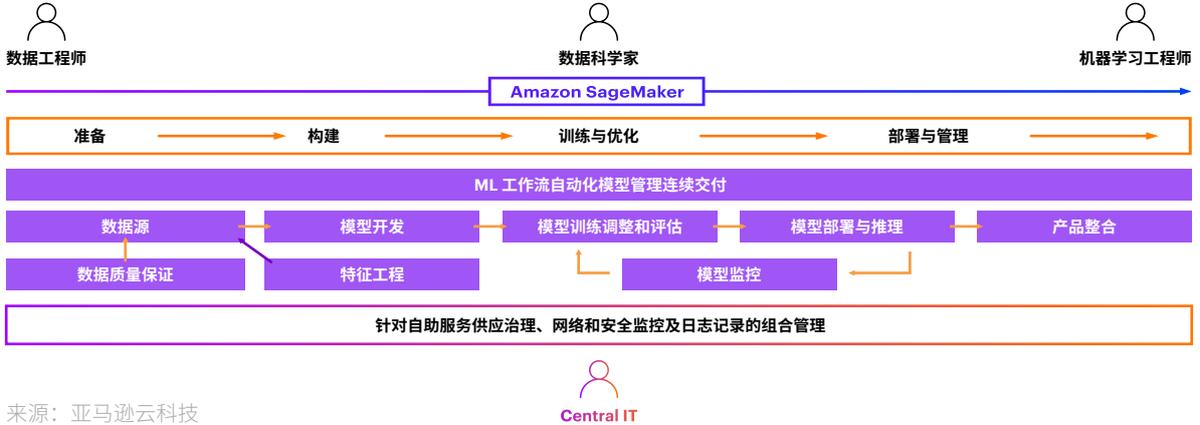
亚马逊云科技能够提供丰富的高性能 CPU 和 GPU 处理器类型，这对于进行大规模训练和在生产环境中进行部署至关重要。使用云端数据湖和存储还能确保企业轻松使用和管理数据，实现无缝衔接、可重复、可扩展的机器学习计划。

同时，亚马逊云科技拥有广泛且深入的机器学习和 AI 服务，并已与 1 万多个客户合作，帮助他们成功实施机器学习。

Amazon SageMaker 让所有开发人员和数据科学家都能快速构建、训练和部署机器学习模型，无需数据工程师或者开发运营人员支持。Amazon SageMaker 是一种完全托管的服务，涵盖整个机器学习工作流程，包括标注和准备数据、选择算法、训练模型、为部署而对算法进行调整和优化、做出预测以及采取行动等。只需花费较少精力和较低成本，企业即可更快地将模型投入生产。

图 9 借助 Amazon SageMaker 轻松触达机器学习

Amazon SageMaker



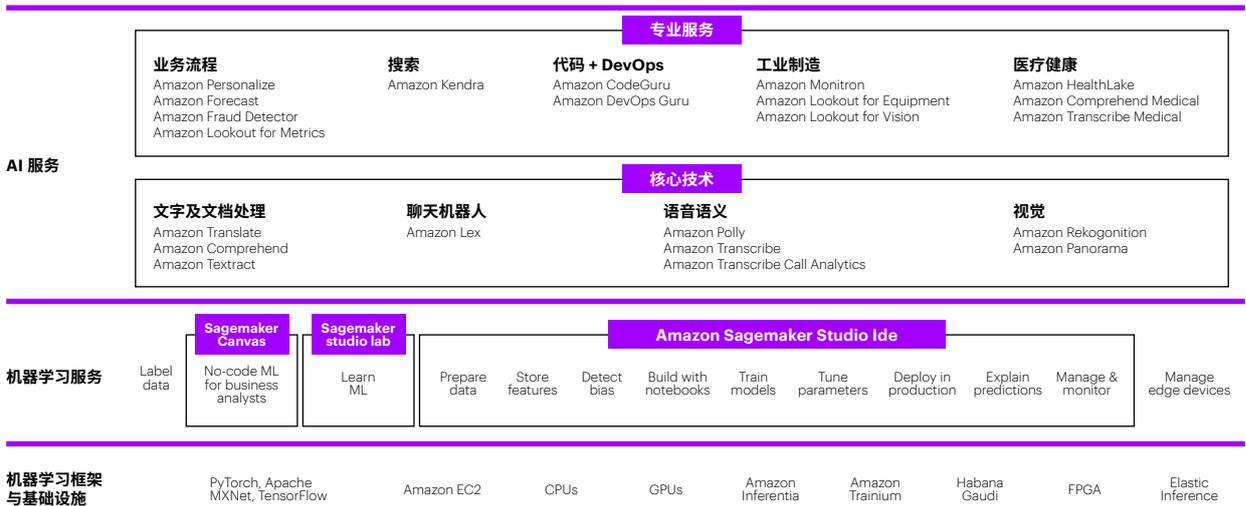
亚马逊云科技预先训练的 AI Services (人工智能服务) 能够为企业的应用程序和工作流, 提供开箱即用的智能体验。人工智能服务可以轻松地与企业的应用程序集成, 帮助应用于常见的使用案例相关用途, 如个性化推荐、对企业的联

系中心进行升级换代、提高安全性、提升客户互动水平。亚马逊云科技 AI Services 会持续学习和进化, 这样的 API 会为企业带来出色的质量和准确性。最重要的是, 使用亚马逊云科技 AI Services 不需要具备机器学习经验。

图 10 丰富的 AI Services 赋能业务创新

机器学习规模化, 驱动业务创新

亚马逊云科技 AI/ML 服务, 提供广泛、完整的机器学习能力



来源：亚马逊云科技

Amazon ML Solutions Lab 将实践教育研讨会与咨询专业服务相结合, 可帮助企业“逆向推理”应对业务挑战, 然后逐步开发基于机器学

习的解决方案。企业能够将自己在此过程中学到的知识, 用于组织的其它方面, 一旦发现业务机遇, 就能立即将机器学习应用其中。

## 云数据现代化成功案例

### 全球领先制药公司

埃森哲联合亚马逊云科技帮助全球领先制药公司成功践行云数据现代化旅程。助力其构建云上现代化数据底座，汇聚企业各域数资产；完善数据治理体系，定义统一数据规范；提升数据服务能力，为业务分析师、数据科学家提供自助式数据服务，进而帮助该企业实现了数据分析和业务创新的可持续的正向循环和相互促进，加速其数字化转型进程。

### 某劳动行政部门

由于新冠疫情，某就业行政部门接到的公众咨询请求激增三倍以上，现有人员应接不暇，无法及时响应公众需求。埃森哲与亚马逊云科技快速响应，仅用五天就协助该部门推出人工智能（AI）虚拟助手。AI助手利用亚马逊云科技自然语言处理及机器学习服务，以及埃森哲高级客户体验解决方案，将人工智能、数据和分析完美整合到客户体验洞察中，像真实工作人员一样理解客户请求并快速准确回复，帮助该部门改进流程、缩短对话并聚焦问题。该部门还将AI助手快速扩展到该机构其他服务中使用。至此，AI助手已经成功服务数百万用户，回答近千万个问题，及时有效的解决了极端异常情况下的公众服务瓶颈。



# 埃森哲与亚马逊云科技 加速企业上云转型

云的潜能是巨大的，它为企业创造了更多机遇和无限可能。埃森哲与亚马逊云科技携手于2015年成立事业部，在全球累计构建了175项联合资产，拥有2万多名持有24000项亚马逊云科技认证的云专家。埃森哲在亚马逊云科技方面的加大投入将进一步促进埃森哲云优先战略的落地，带来更多资源优势。

## 价值共创

从战略规划、云迁移、云原生、安全合规到云管理和运营，埃森哲提供端到端的全栈式、定制化云服务，助力企业跻身“云优先”行列，加速数字化转型与创新，获取可持续的差异化价值。

## 致谢

### 编写指导

#### 崔东博

埃森哲大中华区企业技术创新事业部  
董事总经理、亚马逊云科技服务主管

### 亚马逊云科技

#### 邱胜

大中华区市场部总经理

### 经验共享

在过去14年中，埃森哲与亚马逊携手成功为客户提供了1000多个联合解决方案，覆盖16个行业，并根据不同行业和客户企业特点，打造了多项上云典范。

### 人才共建

在全球累计构建了175项联合资产，拥有2万多名持有24000项亚马逊云科技认证的云专家。

### 主编人员

#### 张为普

埃森哲大中华区企业技术创新事业部  
数据服务主管

#### 梁健明

埃森哲大中华区企业技术创新事业部  
数据服务数据迁移专家

#### 李昕

亚马逊云科技高级产品营销专家

#### 谢开浪

亚马逊云科技产品营销专家

#### 吴佳敏

亚马逊云科技产品营销专家

## 参考资料:

1. P. 多尔蒂(P. Daugherty)、B. 戈什(B. Ghosh)等人(无日期)。埃埃森哲未来系统系列报告《跨越发展, 领军未来》: <https://www.accenture.com/cn-en/insights/digital-transformation-index>
2. 工信部发布《“十四五”大数据产业发展规划》进一步强调数据要素价值。(2021年11月30日)。工业和信息化部关于印发“十四五”大数据产业发展规划的通知。(2021年11月15日)。中国政府网: [http://www.gov.cn/zhengce/zhengceku/2021-11/30/content\\_5655089.htm](http://www.gov.cn/zhengce/zhengceku/2021-11/30/content_5655089.htm)
3. 埃森哲公司在2019年对美国5个行业的190家公司的高管进行了调查研究。埃森哲: <https://www.accenture.com/us-en/insights/technology/data-driven-reinvention>
4. L. 巴斯蒂安(L. Sebastian) (2018年)。在迷宫中找到出路(Navigating The Labyrinth) : [https://books.google.ca/books?id=P4paDwAAQBAJ&pg=PT20&lpg=PT20&dq=gartner+data+unused+97&source=bl&ots=9INMWEJUoU&sig=ACfU3U2C4kWfnNrrl2L2X9fWakObTR3Pzw&hl=en&sa=X&redir\\_esc=y#v=onepage&q=gartner%20data%20unused%2097&f=false](https://books.google.ca/books?id=P4paDwAAQBAJ&pg=PT20&lpg=PT20&dq=gartner+data+unused+97&source=bl&ots=9INMWEJUoU&sig=ACfU3U2C4kWfnNrrl2L2X9fWakObTR3Pzw&hl=en&sa=X&redir_esc=y#v=onepage&q=gartner%20data%20unused%2097&f=false) 2018
5. T. 斯塔克(T. Stack) (2018年2月5日)。物联网(IoT)数据继续呈指数增长。谁在使用这些数据, 如何使用? Cisco: <https://blogs.cisco.com/datacenter/internet-of-things-iot-data-continues-to-explode-exponentially-who-is-using-that-data-and-how>
6. S. 然(S. Jain) (2020年11月19日)。创新数据驱动(Data-driven reinvention)。埃森哲: <https://www.accenture.com/us-en/insights/technology/data-driven-reinvention>
7. J. 斯瓦米(J. Swamy)、A. 雷(A. Ray) (2018年)。云上数据智能推动涡轮增压业务。埃森哲: [https://www.accenture.com/\\_acnmedia/PDF-87/Accenture-CVE-Turbocharge-A4-Brochure-v4-web.pdf#zoom=50](https://www.accenture.com/_acnmedia/PDF-87/Accenture-CVE-Turbocharge-A4-Brochure-v4-web.pdf#zoom=50)
8. 麦肯锡调查: 人工智能对企业收入和规模的影响差异。(2019年11月25日)。网易: <https://www.163.com/dy/article/EUQPKQ8K0511S8FT.html>
9. 2022年数据市场的演变: 大数据趋势。(2022年1月12日)。腾讯: <https://page.om.qq.com/page/OCGH0aKxyxQ2N3ffRddPX6nw0>
10. 《2021年企业中的AI采用》(2021年4月19日)。Oreilly: <https://www.oreilly.com/radar/ai-adoption-in-the-enterprise-2021>

## 关于埃森哲

埃森哲公司注册于爱尔兰，是一家全球领先的专业服务公司，在数字化、云计算与网络安全领域拥有全球领先的能力。凭借独特的业内经验与专业技能，以及翘楚全球的卓越技术中心和智能运营中心，我们为客户提供战略 & 咨询、互动体验、技术和智能运营和 Accenture Song 等全方位服务，业务涵盖 40 多个行业，以及企业日常运营部门的各个职能。埃森哲是《财富》全球 500 强企业之一，目前拥有约 72.1 万名员工，服务于 120 多个国家的客户。我们秉承“科技融灵智，匠心承未来”的企业使命，致力于通过引领变革创造价值，为我们的客户、员工、股东、合作伙伴与整个社会创造美好未来。

埃森哲在中国市场开展业务 35 年，拥有一支约 2 万人的员工队伍，分布于多个城市，包括北京、上海、大连、成都、广州、深圳、杭州、香港和台北等。作为可信赖的数字化转型卓越伙伴，我们正在更创新地参与商业和技术生态圈的建设，帮助中国企业和政府把握数字化力量，通过制定战略、优化流程、集成系统、部署云计算等实现转型，提升全球竞争力，从而立足中国、赢在全球。

详细信息，敬请访问埃森哲公司主页 [www.accenture.com](http://www.accenture.com) 以及埃森哲大中华区主页 [www.accenture.cn](http://www.accenture.cn)。

## 关于亚马逊云科技

亚马逊云科技 (Amazon Web Services) 是全球云计算的开创者和引领者，超过 15 年以来一直以不断创新、技术领先、服务丰富、应用广泛而享誉业界。亚马逊云科技可以支持几乎云上任意工作负载。亚马逊云科技目前提供超过 200 项全功能的服务，涵盖计算、存储、网络、数据库、数据分析、机器人、机器学习与人工智能、物联网、移动、安全、混合云、虚拟现实与增强现实、媒体，以及应用开发、部署与管理等方面；基础设施遍及 26 个地理区域的 84 个可用区，并计划新建 8 个区域和 24 个可用区。全球数百万客户，从初创公司、中小企业，到大型企业和政府机构都信赖亚马逊云科技，通过亚马逊云科技的服务强化其基础设施，提高敏捷性，降低成本，加快创新，提升竞争力，实现业务成长和成功。

亚马逊云科技从 2013 年起进入中国，致力于在中国长期投资和发展。2016 年 9 月，由光环新网运营的亚马逊云科技中国 (北京) 区域正式商用。2017 年 12 月，由西云数据运营的亚马逊云科技中国 (宁夏) 区域正式上线。2019 年 4 月，亚马逊云科技亚太 (香港) 区域正式上线。目前中国是全球除美国以外唯一有 3 个亚马逊云科技区域的国家。此外，亚马逊云科技还建立了上海人工智能研究院、深圳及台北物联网实验室。

亚马逊云科技在中国的愿景是，作为全球云计算的开创者和引领者，利用与生俱来的创新文化，赋能客户的重塑，加速客户全球业务拓展，加强本地人才培养，从而促进行业转型，助力数字经济的可持续发展，并让全社会共同受益。

如果您有任何问题，欢迎拨打热线电话：

亚马逊云科技海外区域：**1010 0866**

亚马逊云科技中国 (宁夏) 区域 - 由西云数据运营：**1010 0966**

亚马逊云科技中国 (北京) 区域 - 由光环新网运营：**1010 0766**

### 免责声明：

本研究报告由埃森哲撰写和制作。报告仅作为研究内容介绍之用。未得到埃森哲的书面许可，文中内容不得采取任何形式进行复制。尽管我们对所依据的信息和资料保持高度谨慎，但无法对其中的准确性和完整性做出绝对保证，请勿绝对化地加以利用。本报告并非埃森哲受托所作。文中所述观点有可能在未经知会的情况下进行调整。报告内容亦非根据任何公司所处独特环境而提供的具体咨询建议，也不作为任何投资建议。