

**CLOUDERA**

# Cloudera Data Platform

现代化助力数据分  
析和人工智能

---

# 碎片化的数据版图

组织机构所具备的变革、创新和竞争的能力取决于您的团队如何有效地利用所掌控的每一块数据。到2022年, 90%的组织机构都会将信息视为至关重要的企业资产。数据分析也将被视为企业必不可少的能力。<sup>1</sup>

尽管人们已经充分认识到数据的重要性, 但是以多种方式访问和使用这些数据却并不是一件容易的事。如同大多数组织机构一样, 您可能会:

- 从越来越多的数据源中接收大量的数据
- 越来越多地在边缘捕捉数据, 并且必须实时提取和分析这些数据
- 将您的数据保存在众多难以集成的系统中
- 担心在整个组织机构中对访问数据的安全性
- 平均有两个公有云和两个私有云在使用

## 在数据的汪洋大海中探寻洞察力

如果数据在各种环境中被锁死使其难以访问或使用, 那么该数据蕴含的价值就会大打折扣。

这就是为什么要打破碎片化的数据版图以获得来自于各种数据、分析和人工智能工作负载的全部价值的原因, 因此我们需要一种全新的处理方法。该方法使您可以全面控制每一个环境和系统的整个数据生命周期, 并使您的业务团队能够做出更快捷、更智能的数据驱动决策。

# 90%

---

到2022年, 90%的组织机构都会将信息视为至关重要的企业资产。数据分析也将被视为企业必不可少的能力。<sup>1</sup>

<sup>1</sup> 《为什么数据和分析是数字化转型的关键》, Christy Pettey, Gartner咨询公司, 2019年3月8日。

## 企业数据云：一种全新的数据管理方法

消除数据竖井并实现各种数据集的全部潜力来推动业务价值的关键是企业数据云。企业数据云是一个大数据平台，可帮助您控制和管理所有环境（包括边缘、私有云和多个公有云）以及整个数据生命周期中的海量数据，以获取更深入的洞察力。借助企业数据云，您可充分发挥所有数据的全部分析潜能，采用最有价值和最具变革性的业务用例。

如需了解企业数据云可以为您做些什么，请[点击此处查看。](#)

## Cloudera Data Platform (CDP) 简介

CDP 是 Cloudera 公司发布的业界首个企业数据云产品。凭借其开放式的混合数据架构，它通过跨碎片化的IT环境无缝地连接数据来支持数据驱动的决策和预测性行动。它能够帮助您保护和治理整个数据生命周期，同时仍然可以基于开源特性为用户提供针对关键数据的快速而简便的访问，以进行预测性洞察和指导性决策。

通过使用 CDP，您可以拥抱看似相互对立的力量（多样化的环境、安全与治理需求以及不断演化的业务需求），这些力量在共享和分析数据时往往会相互牵制和抵消。

**支持顶尖的分析和机器学习功能。** CDP 大大地简化了 PB 级数据和多个工作负载的复杂管理，为组织机构收集、丰富、报告、服务和建模数据的方式扫清了障碍，有助于获得关于整个企业的更深刻洞察力。

**充分发挥混合云和多云环境的潜能。** CDP 可以将任何本地部署、混合和多云环境统一到统一平台中，该平台可以以一致的体验支持无限数量的并发工作负载。

**安全与治理尽在掌控中。** CDP 可以通过内置策略来管理共享数据的访问、活动和使用，从而确保数据和分析生命周期的安全与治理。这样，您的组织机构可以获得更强的敏捷性，可以更迅速地根据洞察力采取行动，同时保持数据安全并确保合规性。

**通过更快速、更便捷的系统来简化管理。** CDP 可以通过一个统一管理平台来全面了解整个数据生命周期，从而简化了基础架构的管理任务。

## CDP 的关键任务优势



适用于无限用例的  
顶尖分析和人工智能  
技术



一个平台统一混合和多  
云环境中的数据



内置的安全机制支持  
几乎无限的数据和成  
千上万的用户



面向所有工作负载的  
可扩展性、高性能和  
成本效率

# 充分发挥数据生命周期的全部潜能

通常, 由于分析引擎和数据科学工具在设计之初并非旨在协同工作, 因此分析工作负载会在各自的竖井中独立运行。

每一个数据竖井都存在着自身的安全、治理和控制需求。并且, 它们还具有自己的模式和数据目录。确保这些竖井保持同步的艰巨任务就落在了IT团队身上。所以IT不得不增加运营成本和开销, 并且孤立的系统也会阻碍数据用户, 因为这些系统无法支持多个分析工作负载针对同一数据集运行。

最引人注目的数据用例需要多种分析引擎和多个流程协同工作, 并且无论数据位于何处, 都需要共享数据。

CDP可以在任何基础架构上运行 - 从内部私有云到多公有云, 或以私有云和公有云的任意组合构成的混合云。CDP使用户能够使用潜在不同的工具处理相同的数据集, 同时还遵循合规性规定, 而不会损害安全性。通过利用 CDP 授予的自助服务访问权限, 用户和群组可以浏览整个数据生命周期, 在数据所处的位置直接应用分析, 并在数据源与该数据进行交互以发现其所蕴含的价值。借助 CDP, 您的组织机构可以很好地管控从边缘到人工智能的整个数据生命周期, 转型为数据驱动型企业。



## CDP成功案例

“在AWS平台上运行的Cloudera Data Platform可以帮助我们进行客户细分数据管理, 分析阐释并将这些洞察力付诸行动, 使我们能更好地量身定制帮助客户发现、购买和使用我们的服务。”

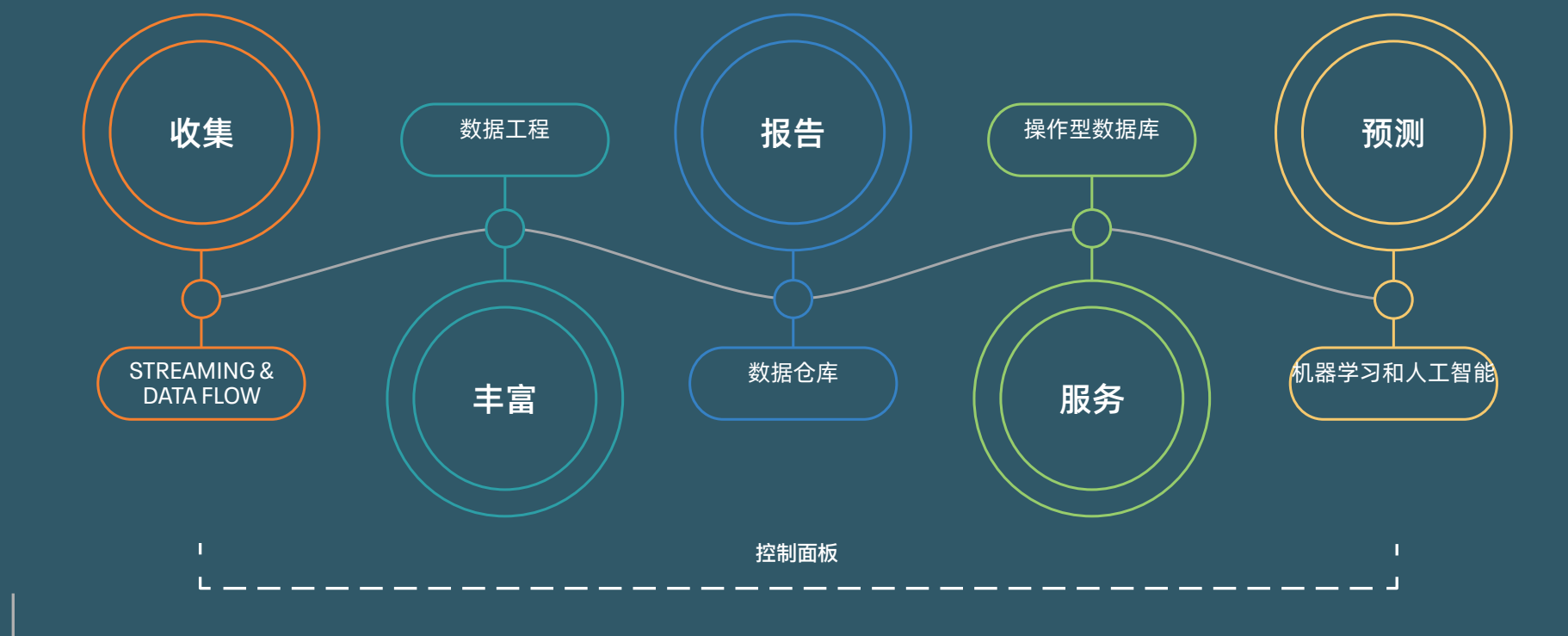
Gil Genio

环球电信 (Globe Telecom) 公司首席技术和信息官、首席战略官

[点击这里了解更多内容。](#)

# 数据生命周期的管理和保护

CDP 使您的组织机构可以在整个数据和分析生命周期中对其数据进行控制。您可以消除数据竖井, 同时保护, 管理, 管理和自动化整个生命周期过程。



**CLOUDERA**  
**SDX** 元数据 / 模式 / 迁移 / 安全 / 治理

# 面向多功能分析和人工智能的CDP服务



## 数据流管理 (DataFlow and Streaming)

CDP 具有高度可扩展的实时流分析引擎, 该引擎可摄取、整理和分析数据以获取关键洞察力和立即可付诸行动的情报。



## 数据工程 (Data Engineering)

CDP 使您能够丰富、转换和清理数据, 以轻松地创建、执行和管理端到端数据管道。



## 数据仓库 (Data Warehouse)

CDP 具有自动扩展性、高度并发性以及经济高效性, 因为它可以从任何地方 (结构化、非结构化和边缘数据) 摄取海量数据。



## 操作型数据库 (Operational Database)

CDP 具有实时、可扩展的数据库, 使您能够在统一平台中处理传统结构化数据和非结构化 (和半结构化) 数据。



## 机器学习 (Machine Learning)

CDP 为数据科学团队提供了针对业务数据以及端到端机器学习工作流所需的工具和计算资源的简单而直观访问。



## CDP 成功案例

# 23PB

的非结构化数据得到了使用, 帮助制药公司快速将新药推向市场

一家制药公司利用快捷、全面的分析和机器学习技术来加快 FDA 的审批速度, 并将新药更快速地推向市场。

[点击这里了解更多内容。](#)

# 跨环境实现互操作性的混合云

许多组织机构都在最大限度地发挥多个公有云结合其私有云的优势。公有云及私有云的这种组合, 即混合云环境, 能够带来灵活性、敏捷性、高性能和成本效益等优势。它使您可以根据每个工作负载的需求自由抉择每个工作负载将在哪里运行。例如, 您可以在公有云中运行间歇/批处理工作负载, 而在私有云中运行更为持续的工作负载。

通常, 在整个混合云环境中的多处地方都可以找到任务关键型数据。分析这些数据非常重要, 但是面临一个棘手的问题: 数据竖井会导致分析不完整和不准确。其中部分原因是由于混合云本身缺乏在不同环境中管理工作负载的集中式处理方法。确保整个混合云中的数据治理和安全性也是十分费时、费力的工作, 会导致合规风险。

当您能够针对各种业务用例在相同的多元数据集上运行多种分析和机器学习模型时, 您将从数据中挖掘出最大的价值。但是工作负载需要一种共享数据集的方式, 并且需要在不增加管理混合云复杂性的情况下做到这一点。



## CDP 成功案例

# 40%

的销售队伍效率提升

一家快速消费品公司维护着一个跨应用程序和外部数据源的多结构化数据的集中存储库, 以获得单一洞察力信源。

[点击这里了解更多内容。](#)

CDP使所有这一切成为了可能。它大大地简化了在私有云和公有云环境中管理PB级数据和多个工作负载的复杂性,从而为您更好地利用混合云扫清了以数据为中心的工作负载带来的障碍。凭借CDP,您可以:



### 更好地控制资源

如果资源受到本地集群容量的限制,则可以调配资源将工作负载快速爆发到云端。



### 非常便捷地将工作负载转移到任何地方

将选定的数据工作负载迁移到不同的云中,而开发人员无需重写代码。



### 以更大的灵活性和敏捷性进行部署

可以在公有云中快速部署新的用例,一旦明晰了性能和增长情况,可将其迁移到私有云中。



### 确保您的工作负载保持合规

使用云原生数据平台以满足数据中心内的数据本地化规章制度。

无论您需要如何使用混合云,CDP均可确保您的工作负载能够以其最高的性能和成本效率运行。

[请在Cloudera.com上开始免费试用。](https://www.cloudera.com)

# 58%

## 的企业正在向混合IT环境转型。<sup>3</sup>

<sup>3</sup> 《未来是一种混合模式》演讲, 451 Research, 2020年。



# 在加强治理和控制的同时，支持对数据的自助访问

您希望员工、合作伙伴和客户可以访问您的数据，以用于各种各样的业务用例。但是，如果没有企业数据云，在孤立的系统之间维护安全、治理和控制将是一场无望取胜之仗：您的私有云和公有云环境使用不同的工具、功能和流程用于保护数据安全。典型的混合云管理方法不允许 IT 和信息安全团队在这些环境中应用统一的综合策略。此外，IT 团队还可能被迫限制在遗留系统和专有锁定的技术和厂商中。

要能在这些限制条件下授予对数据的访问权限，您的IT团队必须创建和管理相同数据集的多个副本，并且在大量的访问请求到达前能妥善地应对。这是一项相当复杂的任务，耗时耗力并且成本昂贵。同时，您的IT团队仍然需要保持合规性。能够做到这一点的唯一方法是限制用户对这些有价值的数据的访问。但因而造成的结果是：延迟的业务洞察力，增加的运营开销以及只会带来更大安全风险的盲点。

如果您想让更多的用户充分地利用分析和机器学习的优势，从而可以更好地进行决策，且能安全地根据他们可以查看和使用的数据进行适当的限制，则需要提供一个一致且共享的治理与安全模式，能支持跨私有云和多个公有云环境进行各种分析。



## CDP 成功案例

# 10x

增长的数据量也能轻松处理

一家全球性的物流公司使用CDP架构来提升可扩展性和可访问性，同时更好地确保合规性并能探索新的创收机会。

[点击这里了解更多内容。](#)

## 采用共享数据体验 (SDX) 实现安全与治理的综合性方法

SDX是 CDP 体系架构的基本组成部分。这项始终在线的服务使您可以通过一个统一的界面来控制任何环境中各项部署的安全与治理。

### 一次应用, 处处执行。

通过SDX, CDP 可一次定义数据访问策略, 然后在所有工作负载以及在部署它们的混合云和多云环境中无缝地实施这些策略。这样就无需创建同一数据的多个副本, 从而降低了成本和复杂性。

### 跟踪和审核数据访问和使用情况。

借助SDX, 您可以通过统一的平台范围内的操作快速识别和管理敏感数据, 并有效地满足监管要求。此外, 您还可以在所有分析和部署的整个数据生命周期中有效地证明合规性。

### 让分析师非常便捷地访问数据集。

SDX 支持强大的数据目录功能, 可在该企业数据云服务中搜索、组织、保护和治理数据。这为数据分析人员提供了一个统一的控制界面视图, 他们可以在其中管理和发现他们有权访问的数据 (通过元数据进一步增强了该功能)。通过应用于操作、社交和业务的上下文环境, SDX 可创建受信任的可复用数据集供数据分析人员快速找寻, 以便他们更快速地专注于关键工作。

数据管理员可以使用 SDX 中的数据目录功能来浏览、搜索和标记数据湖的内容, 创建和管理授权策略 (按照文件、表、列、行等), 识别用户已经访问的数据, 并访问特定数据集的血缘沿革。

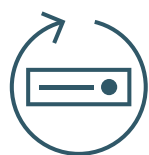
A large graphic of the number '90%' in a dark blue, sans-serif font. The '9' and '0' are filled with a diagonal hatching pattern, while the '%' symbol is solid dark blue.

到2023年, 全球500强企业  
中90%的公司都会  
将分析治理整合到更广泛  
的数据和分析治理计  
划中。<sup>4</sup>

<sup>4</sup> 《2020-2024年数据和分析领域100个预测》, Gartner, 2020年3月20日。

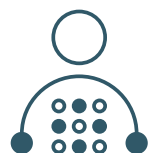
# 一站式管控

可通过一个统一的控制台来管理、监控和编排所有 CDP 服务, 以实现数据管理、工作负载分析和数据移动功能。



## Replication Manager

在 CDP 中的环境之间复制、迁移、获取快照和恢复数据。系统管理员和数据管理员可以在数据湖之内或之间移动、复制、备份、拷贝和恢复数据。该操作可以出于备份、灾难恢复或迁移的目的, 或者需要在另一个虚拟环境中进行开发/测试。



## Workload Manager

深入了解工作负载所需的工具。通过 Workload Manager, 可以查看有关作业的相关统计数据以及可用于对失败的作业进行故障排除或对慢速作业进行优化的信息。这意味着您可以确保 CDP 中每个工作负载都能达到最佳性能。



## Management Console

仅使用一个仪表盘即可流畅地简化用户、环境和服务的管控、管理和调配。该管理控制台 (Management Console) 使您可以从一个统一的控制界面中管理云和本地资源, 并同时治理用户及其访问权限。



## CDP成功案例

# \$200K

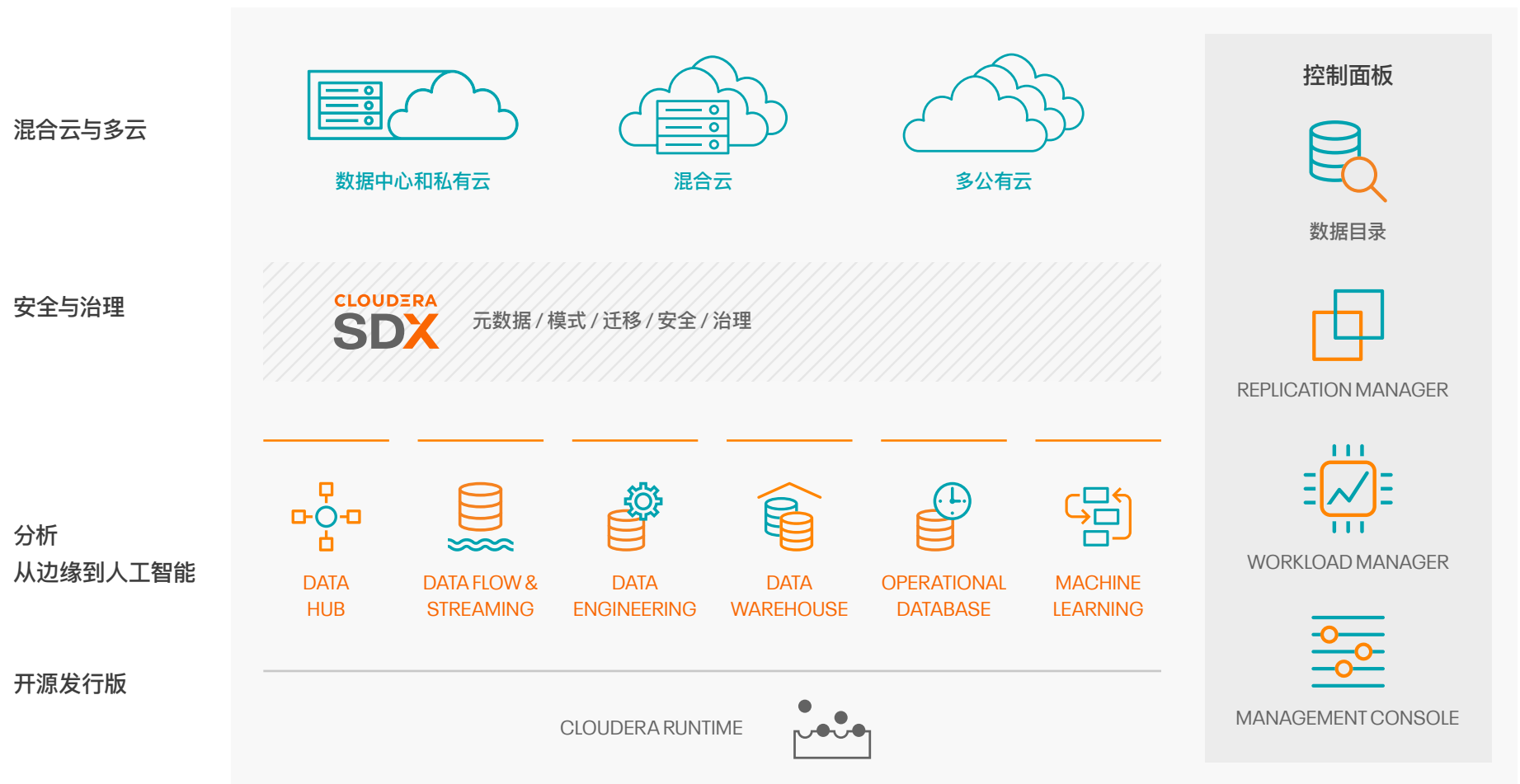
预计的每月运营支出减少金额

一家全球性的消费信用卡公司在遵守相关法规并最小化成本的同时, 可以更好地管理其敏感的信用数据。

[点击这里了解更多内容。](#)

# 利用完整的企业数据云解决方案 加速分析和人工智能

通过CDP, 您可以在任何环境中控制整个数据和分析生命周期, 并具备完全的安全性和控制力。



# CDP 的优势的显而易见的



## 顶尖的分析和人工智能技术

CDP 为您带来了无与伦比的数据管理和分析功能, 这些功能可以在整个混合环境中的任何地方应用, 而无需在此过程中牺牲其易用性、功能、性能等。



## 扩展安全性不成问题

通过其 SDX 层, CDP 可以非常方便地扩展到成千上万的用户, 并且所有用户都可以在治理团队确定的不同策略中工作。这样可确保组织机构能够得益于数据驱动的决定, 而又不会增加违规风险。



## 跨混合环境的整体视图

CDP 是一种容器化、可扩展的轻便型服务平台, 可通过集中式控制台在不同的工作负载和云之间进行部署。您可从一个单一的控制界面中管理、监视和编排所有部署, 无论这些部署是否在您的私有云还是多个公有云中。此外, 您还可以非常便捷地创建、提供和删除服务。



## 增强 IT 控制能力并降低复杂性

随着 CDP 技术性能的不不断提升, 其使用和管理也变得更加直观。IT 团队可以通过统一的可见性更有效地控制混合环境, 而用户可以通过自助服务访问关键数据并使用自己选择的工具。



## 针对每个工作负载优化的资源

CDP 的设计旨在提高效率, 同时控制成本。它将计算资源和数据资源独立分开, 从而为您提供更多的部署选项, 并能够以所需的准确的性能级别将工作负载扩展到 PB 级数据。



## 基于开源的构建

CDP 使您彻底摆脱了供应商锁定的束缚, 您可以充分利用与现有系统和应用程序集成的开源生态系统。在这个 100% 开源的平台上, 您既可以访问高度定制化的服务, 也可使用现成的用户服务体验。

---

# 在数据的汪洋大海中捞针。 请立即试用CDP。

业务转型需要创新性的数据分析和人工智能技术。凭借 CDP, 您现在可以通过PB级的混合数据架构来满足该需求, 可以非常灵活地使用私有云和公有云产品。让您的数据带您走得更快更远。

[请在Cloudera.com上开始免费试用。](https://www.cloudera.com)

## 关于Cloudera

在 Cloudera, 我们相信数据可以使今天的不可能, 在明天成为可能。我们使人们能够将复杂的数据转换为清晰而可行的洞察力。Cloudera 为任何地方的任何数据从边缘到人工智能提供企业数据云平台服务。在开源社区不懈创新的支持下, Cloudera 推动了全球最大型企业的数字化转型历程。

如需了解更多信息, 请访问 Cloudera 公司中文网站: [cn.cloudera.com](https://cn.cloudera.com)

## 资料来源

- 1 《为什么数据和分析是数字化转型的关键》, Christy Pettey, Gartner, 2019年3月8日。
- 2 Flexera公司发布的《2020年云计算报告》。
- 3 未来是一种混合模式》演讲, 451 Research, 2020年。
- 4 《2020-2024年数据和分析领域100个预测》, Gartner, 2020年3月20日。

---

[cn.cloudera.com](https://cn.cloudera.com)

电话: +86-21-62369001

地址: 上海长宁区延安西路 2299 号世贸大厦 26 楼 2612 室

© 2021 Cloudera 公司版权所有。Cloudera 和 Cloudera 徽标是 Cloudera 公司在美国和其他国家的注册商标。所有其他商标均为其各自所有者的财产。资料如有更改, 恕不另行通知。

**CLouDERA**