

凌云驭势，重塑未来

云原生数据战略—助推数字化转型

翁建清

亚马逊云科技



数据 以惊人的速度增长

今天的数据中

90%

是在过去两年里创建的。

数据来源: [Worldwide IDC Global DataSphere Forecast, 2022-2026](#)



从数据中 获取价值很难

仅有

32%

的公司表示，他们能够从数据中
实现价值

数据来源: [Accenture study](#)

基于超过15年的技术优势与客户案例，率先提出了“云原生数据战略”

云原生数据战略

构建云原生数据战略——探索企业创新的“觉醒”一刻

人类科技的发展史来自于诸多关键的“觉醒”一刻，而无论对于个人还是组织，通往“觉醒”之路，都离不开“数据”的积累与分析
数据是企业的创新之源，成为数据驱动型组织可以让企业平均增加30%的年收入，然而，企业数据的处理与分析、企业构建数据驱动型的过程都面临众多挑战
亚马逊云科技基于自身在数据领域15年的深厚积累，助力**150万客户**成为了数据驱动型企业，提出“**云原生数据战略**”，帮助企业实现更好的决策与增长

云原生数据战略

建立面向未来的云原生数据基础设施

用合适的工具应对所有类型的工作负载
在大规模的情况下保持高性能
举重若轻应对繁重的重复性工作
提供可靠、可扩展的数据服务

实现高效、跨组织的数据一体化融合

应对数据的一体化融合，提供智能湖仓解决方案
通过ZERO-ETL集成实现数据的实时分析

通过机器学习工具，用数据驱动智能创新

提供能够满足大模型、机器学习工业化的AI/ML工具
提供面向众多应用场景的解决方案与产品服务
助力AI普惠，降低企业应用AI的门槛

数据治理与数据安全

应对数据的安全与治理，为企业数据创造安全的访问途径，并实现更好的数据治理与机器学习治理，与更精细的数据管理。

凌云驭势 重塑未来 —— 构建数据驱动型企业

亚马逊云科技为客户提供方法论，与全面、敏捷、安全以及高性价比的云原生数据服务，帮助客户构建自己的端到端的云原生数据战略；
打造数据驱动文化，让企业成为数据驱动型企业。

云原生数据基础设施

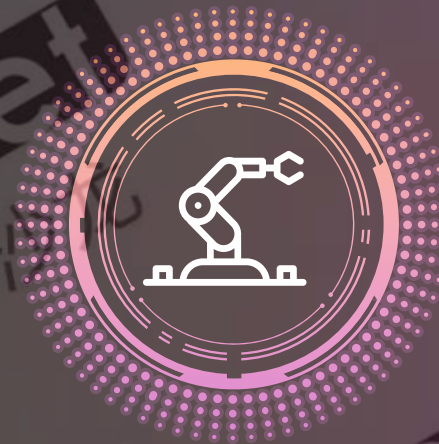
面向未来的云原生数据基础设施



用合适的数据服务应对
所有类型的工作负载



在大规模的情况
下保持高性能



举重若轻应对繁
重的重复性工作



提供可靠、可扩
展的数据服务

支持所有工作负载的工具

专门构建的云原生数据服务

94%

在亚马逊云科技顶级客户中

94% 的用户使用10种以上的数据库和分析服务

数据库

Amazon Relational Database Service (MySQL, PostgreSQL, MariaDB, Oracle, SQL Server)
Amazon Relational Database Service Custom
Amazon Aurora
Amazon DynamoDB
Amazon DocumentDB

Amazon Keyspaces
Amazon Neptune
Amazon Timestream
Amazon ElastiCache
Amazon MemoryDB for Redis
Amazon Database Migration Service

分析

Amazon Redshift
Amazon Athena
Amazon OpenSearch Service
Amazon EMR
Amazon Lake Formation
Amazon MSK

Amazon Kinesis Data Streams
Amazon Kinesis Data Firehose
Amazon Kinesis Data Analytics
Amazon Kinesis Video Streams
Amazon QuickSight
Amazon Glue

机器学习/人工智能

Amazon SageMaker
Amazon Comprehend
Amazon Polly
Amazon Rekognition
Amazon Textract
Amazon Transcribe
Amazon Translate
Amazon Fraud Detector
Amazon Personalize
Amazon Forecast
Amazon Augmented AI
Amazon DevOps Guru
Amazon CodeGuru
Amazon HealthLake
Amazon Kendra
Amazon Lex
Amazon Lookout for Equipment
Amazon Monitron
Amazon Panorama

数据一体化融合

ETL（提取、转换、加载）纷繁复杂



4.0 来了

Amazon Glue

简单、可扩展且无服务器



更快地集成数据



大规模自动化



无需管理服务器



基于Spark和Python构建



ETL开发者



Amazon Glue Studio



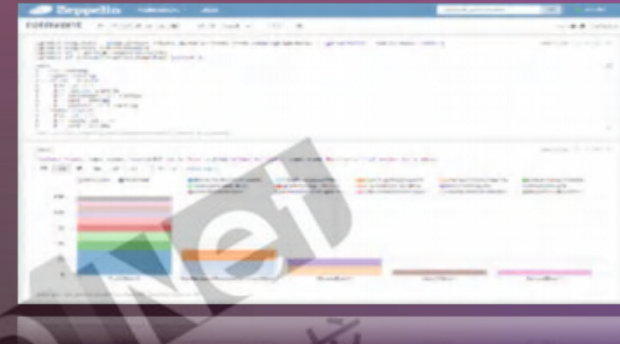
数据分析师



Amazon Glue DataBrew



数据工程师



Amazon Glue Notebooks

- ETL开发人员可以使用Glue基于Spark的ETL引擎处理PB的数据，而无需编写Glue Studio的任何代码
- 数据分析师可以使用AWS Glue DataBrew的250多个内置转换以可视化方式更快地准备数据
- 数据工程师可以利用Glue notebook进行交互式数据处理，并转换成任务，充分利用Spark和Python平台的强大功能

全新发布

Amazon Glue for Ray

易于扩展的Ray开源Python
并行学习框架，无服务器



在Amazon Glue for Ray实
现了Python并行和分布式计算



无需对基础设施进行管理或
调整



支持按需计费，以具有成本效
益的方式实现扩展

全新发布

Amazon Glue Data Quality

交付高质量数据

- 基于您的数据自动提供数据质量规则建议
- 通过持续的数据分析使数据保持高质量
- 控制数据湖和数据流水线中数据集的数据质量
- 采用现收现付计费成本效益，不锁定

能不能不做ETL，直接进行分析与ML？

客户正在寻求更易于迁移和分析大数据的方法

Zero ETL 的未来...

亚马逊云科技各项服务之间的深度集成



其实，多年来，我们通过深化服务之间的集成，已经在Zero ETL中取得了一定成果，这些亚马逊云科技各项服务之间的深度的数据集成正在帮助企业执行分析和机器学习，且**无需移动数据**。

全新发布

[GA]

Amazon Redshift 流式注入功能

将流数据注入数据仓库，以进行
实时分析，实现实时数仓



Amazon Kinesis
Data Streams



Amazon MSK



Amazon
Redshift



全新发布

Amazon Aurora Zero-ETL to Amazon Redshift

- 交易数据的近实时分析和机器学习
- 整合来自多个Aurora数据库的数据
- 自动连续更新
- 无服务器——无需管理基础设施



Zero-ETL



全新发布

Amazon Redshift auto-copy from s3

可将文件从S3自动化复制到
Amazon Redshift

现已可供预览

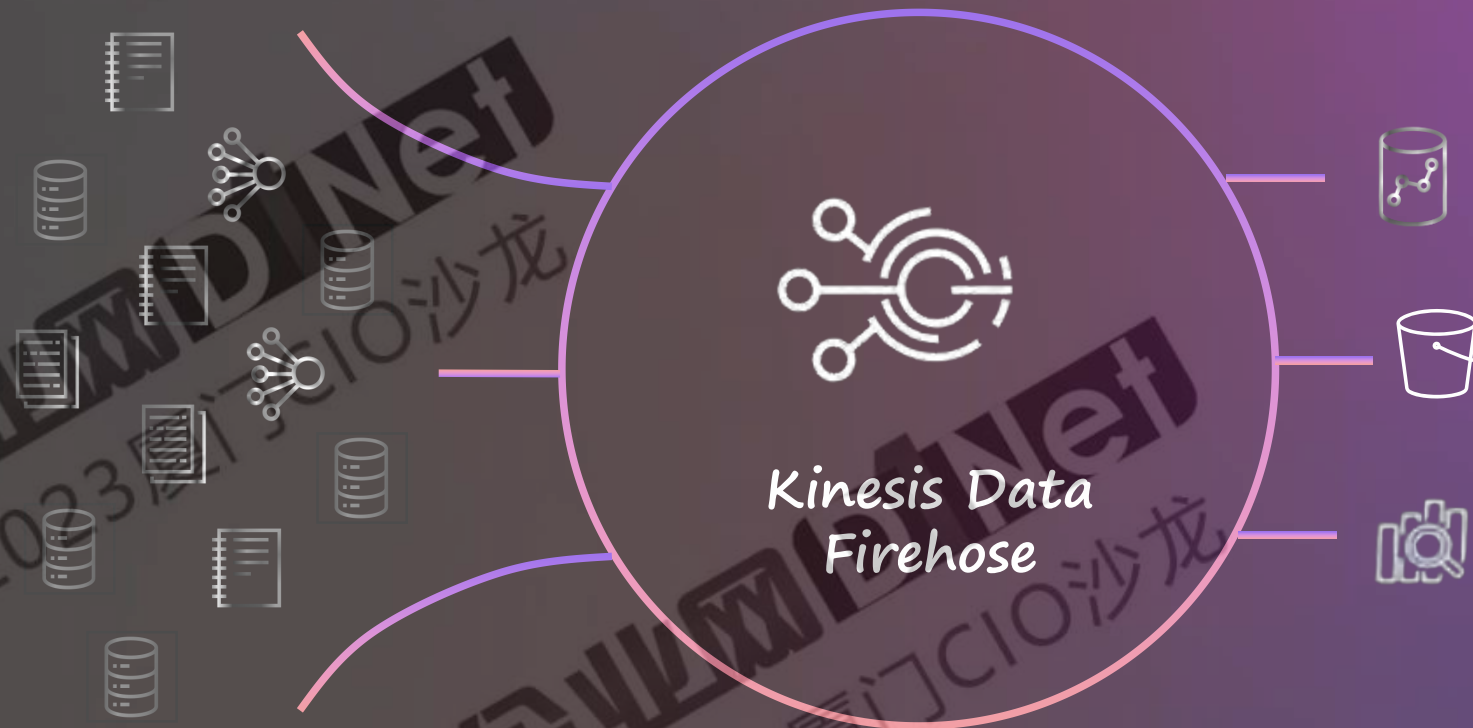
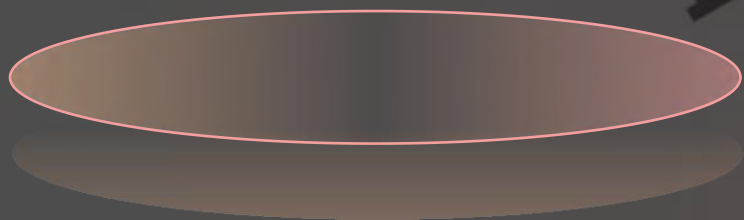
轻松创建并维护简单的数据摄取管道

在S3中创建新文件后持续摄取数据

自动化数据加载，无需工程资源
简化了湖到仓之间的数据联动

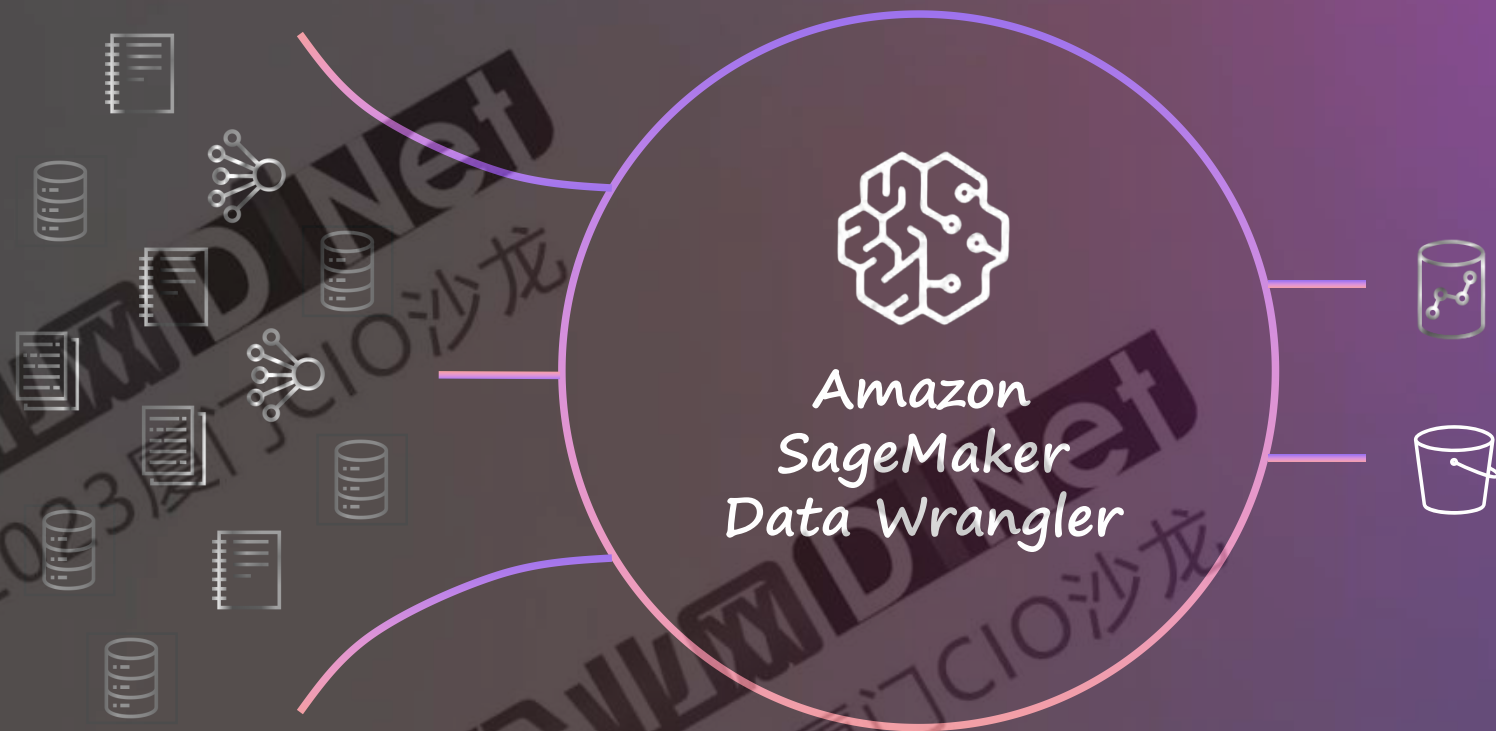
丰富的第三方数据源
已有的本地数据源
如何无缝连接？

20多个来自亚马逊云科技和第三方来源的流数据



从SaaS服务和第 三方来源导入数据， 用于机器学习

您可以快速访问和选择来自各种热门数据源
以及 40+ 个其他第三方来源的数据，生成
数据洞察理解数据质量，并应用于机器学习。



可视化的数据准备和特征工程服务

在SaaS服务、数据湖和数据仓库之间移动数据



零代码数据集成服务

全新发布

Amazon AppFlow

现在共支持

50多个Connector

DATADOG

DYNATRACE

● FACEBOOK ADS

SLACK

AMAZON ATHENA

ZENDESK

SALESFORCE

● SAP

● GOOGLE ADS

SNOWFLAKE

SERVICENOW

● LINKEDIN ADS

INFOR NEXUS

JIRA CLOUD

MARKETO

SINGULAR

AMAZON S3

● GOOGLE ANALYTICS 4

UPSOLVER

● INSTAGRAM ADS

AMAZON REDSHIFT

AMAZON EVENTBRIDGE

● GOOGLE SEARCH CONSOLE

AMAZON LOOKOUT

● FACEBOOK PAGE INSIGHTS

AMAZON HONEYCODE

● AND MORE

数据驱动智能创新



5 年前



如今

56% 的受访者表示在至少一项功能中采用人工智能技术

超过10万 名客户使用亚马逊云科技人工智能/机器学习服务

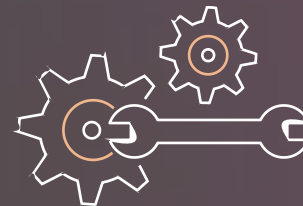
推动机器学习创新的六大关键趋势



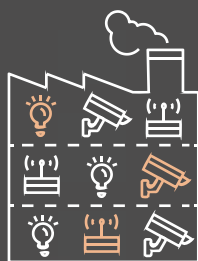
模型
成熟度



数据
增长



机器学习
产业化



机器学习驱动
的用例



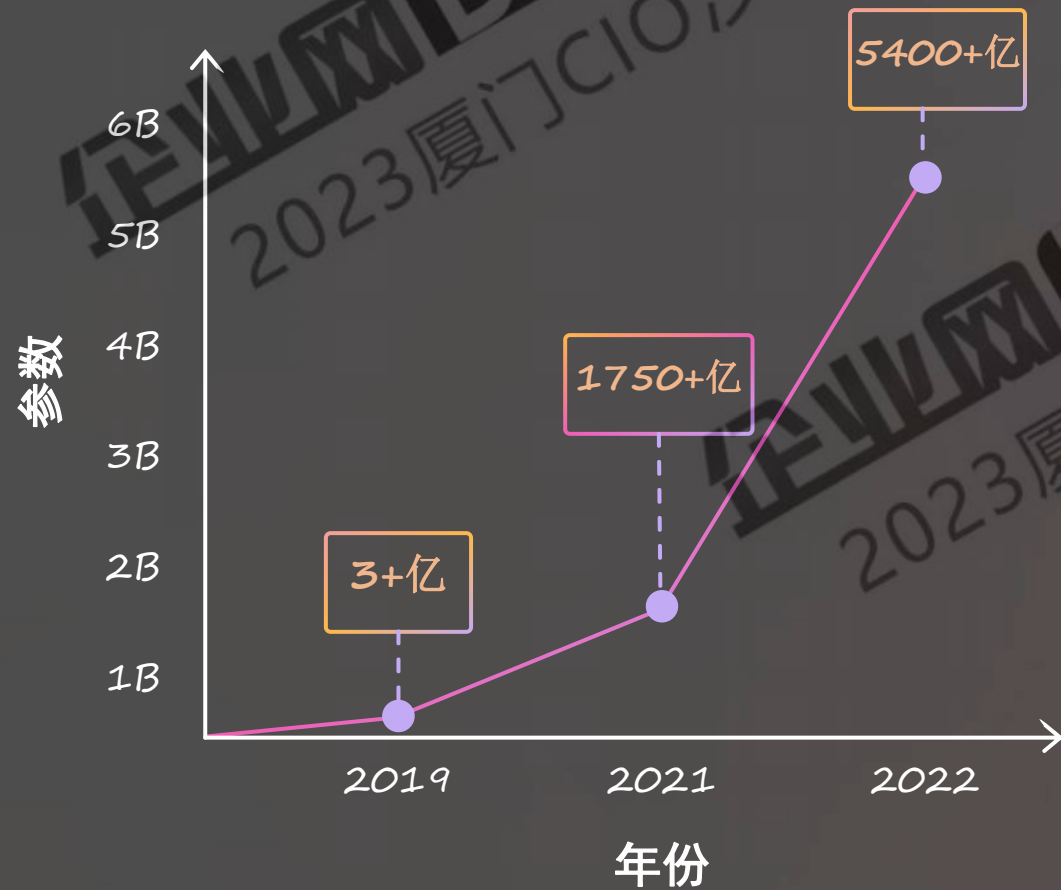
负责任的
人工智能



机器学习
普惠化

趋势1：模型成熟度

基础模型的兴起



2019 - 2022年

用参数数量来衡量，模型大小增加

1,600+倍

Stable Diffusion 1.0 引起的 AIGC 热潮

第一个走向全球的基础模型



花样野狼



公主的色彩装



猫骑士

Stable Diffusion 2.0

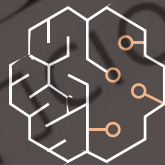
无限的创造力



利用Amazon SageMaker JumpStart 一键部署 Stable Diffusion



在您的亚马逊云科技账户中进行大规模测试和部署



Jupyter 笔记本和代码示例



针对您的场景进行模型调优

企业网DINet
2023厦门CIO沙龙

趋势2：数据增长

趋势2：数据增长趋势



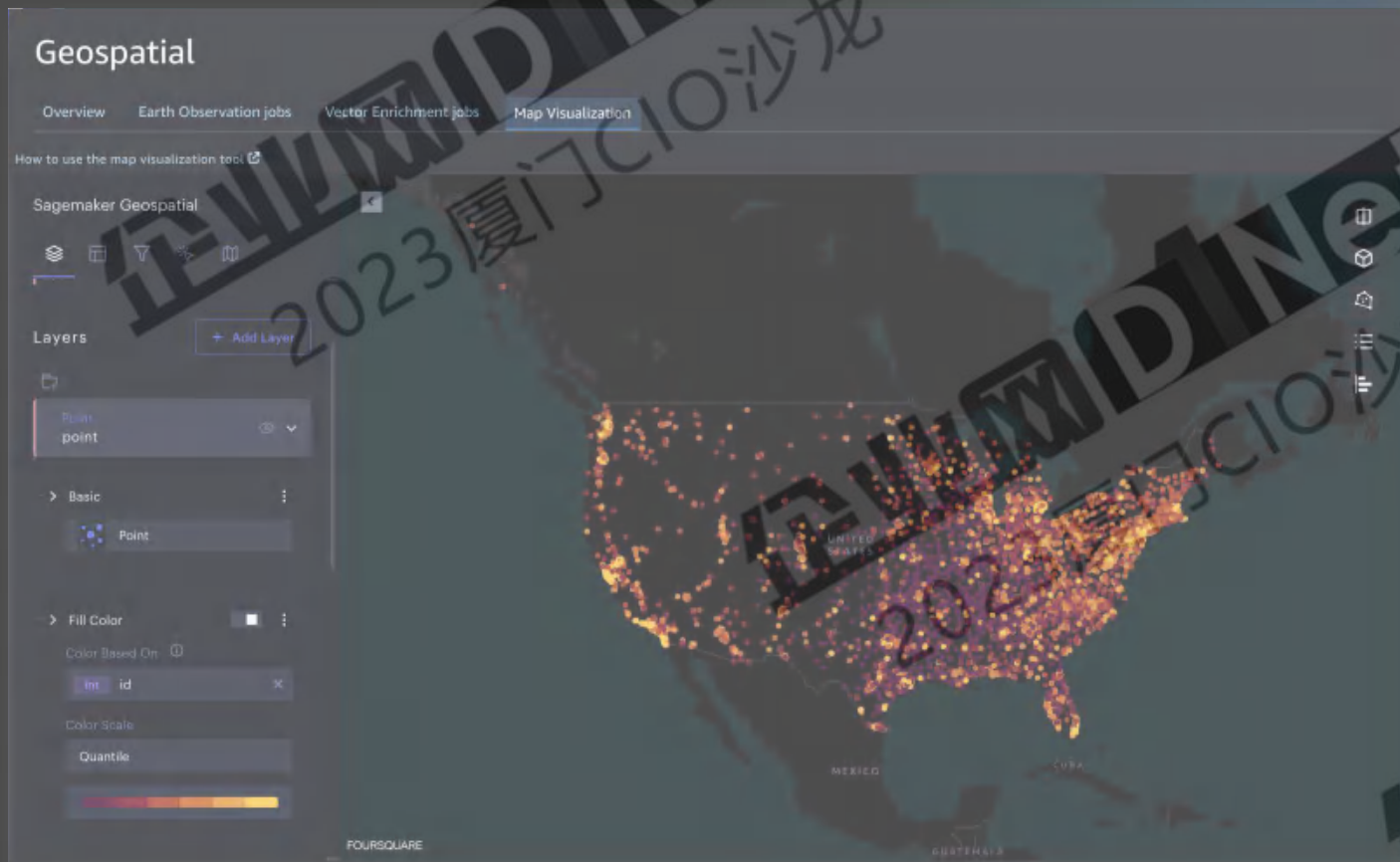
(40.44641777896232, -80.01560258865356)
匹兹堡 07:11:36

最新推出！

Amazon SageMaker 的地理空间功能

使用地理空间数据构建、训练和
部署机器学习模型预览

预览版



趋势3： 机器学习产业化



数以万计的客户使用 Amazon SageMaker 推动机器学习产业化

3M Science.
Applied to Life.

AGCO

amazon

Atlas Van Lines

AUTODESK

avis budget group

BUNDESLIGA

Celgene

Capital One

Cerner

CONDÉ NAST

CONVOY

Domino's

F1

FICO

Formosa Plastics

Frame.io



GoDaddy

lyft

HEARST

ifood

Infoblox

intuit

INVISTA

Lenovo

Liberty Mutual
INSURANCE



NOVARTIS

Roche

NuData Security
mastercard



THOMSON REUTERS

tinder

verizon

arlo

coinbase

freshworks

INTERCOM

grammarly

AstraZeneca

SIEMENS

EULER HERMES

T-Mobile



mercado
libre

Georgia-Pacific

NatWest

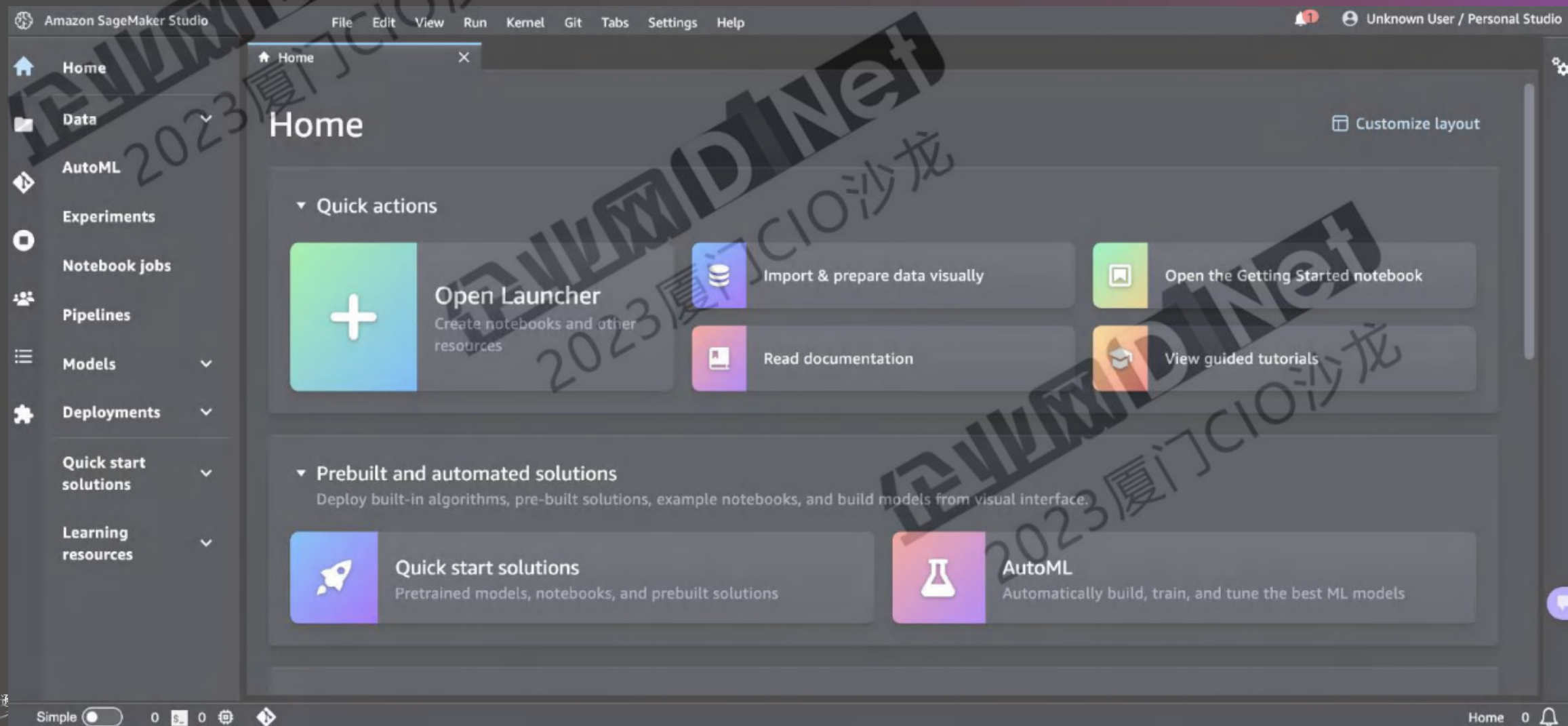
持续升级的软件工具，助力机器学习产业化

全新发布

下一代SageMaker Studio Notebooks

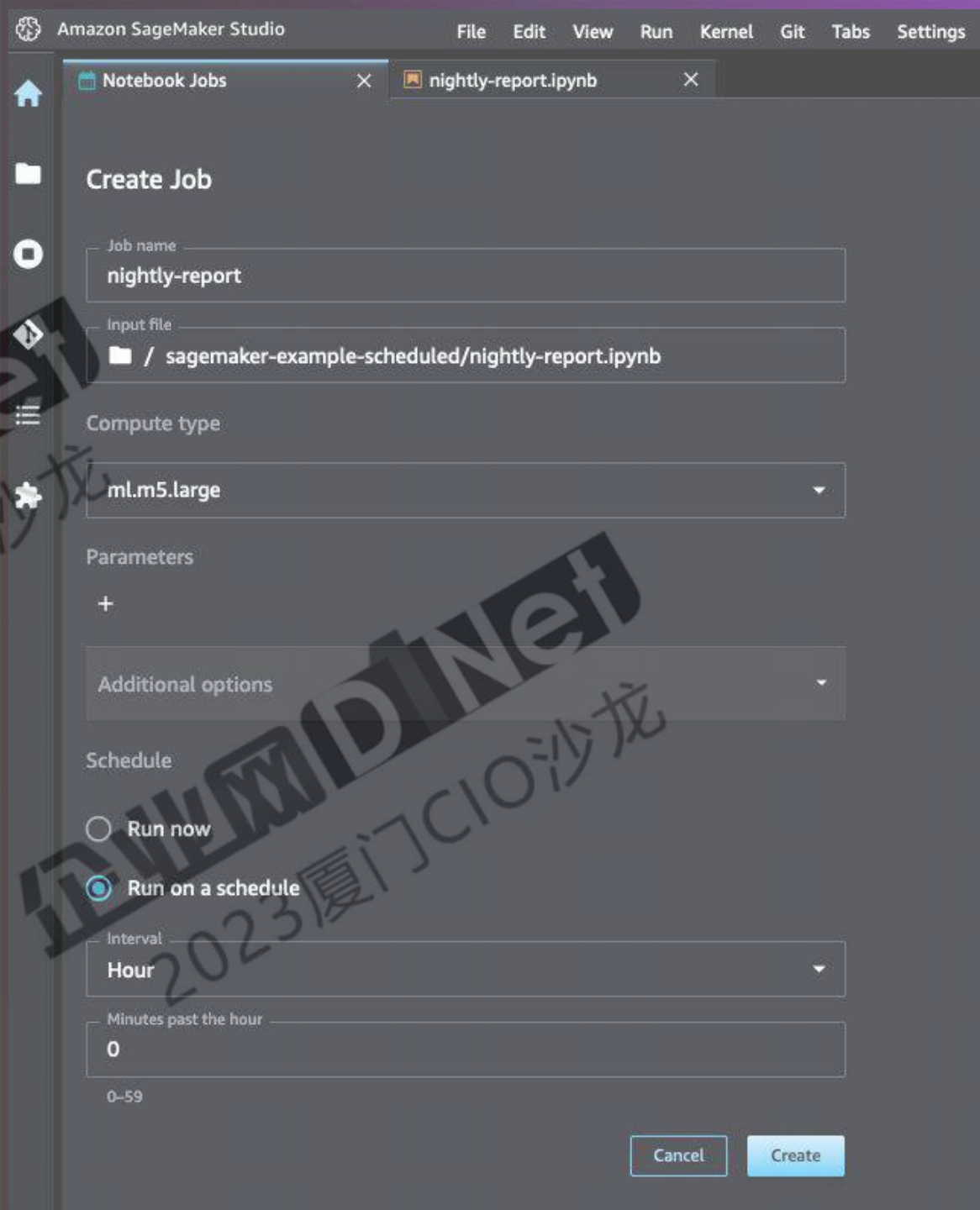
现已可用

内置数据准备、实时协作和自动化部署功能



自动化笔记本部署

- 定期自动运行笔记本
- 管理预定笔记本任务
- 将笔记本部署从数天缩短到数小时



趋势4： 机器学习驱动的业务场景





智能联络中心



预测



智能文档处理



个性化



异常检测



工业设备监控

提升客户体验

更好、更快地做出决策

改善业务运营

集成人工智能的几个案例

全新发布

Amazon Transcribe
Call Analytics
实时功能

提高代理商生产力并改善客户体验
普遍可用

Amazon Transcribe Call Analytics将人工智能集成到客服中心，能从数百万个电话中进行通话后和实时通话分析，很大程度提升客服效率和满意度



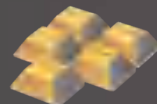
提高**83%**处理电话的效率

Wix

将客户情绪可见性提高至**100%**

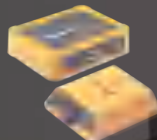
talkdesk

将大型食品物流企业的电话
平均处理时间缩短了**24%**



Monitor

传感器



Monitor

网关



Monitor

应用程序

- 端到端预测性维护解决方案
- 帮助您跟踪和检查潜在故障
- 无需机器学习技能

Baxter

500H

预估减少当机时间

7M

生产元件影响

10K

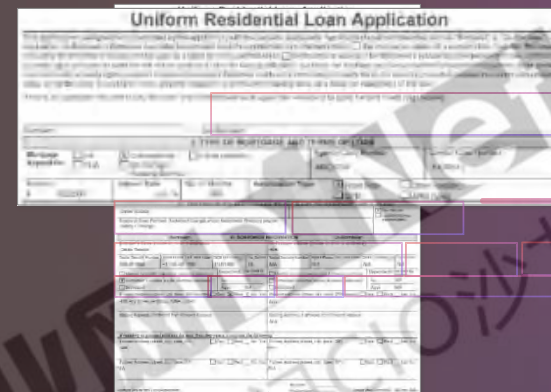
病人影响

全新发布

利用Amazon Textract
分析借贷

加快处理借款文档
广泛可用

原始文件



输出信息

文档类别：固定居所贷款申请
借款人姓：JORGE
借款人姓：SOUZA
借款人中名：
共同借款人姓：GEORGINA
共同借款人姓：DAVIS
共同借款人中名：AMY
按揭类别及贷款条款：
贷款类型：普通
代理案例编号：ABC1234
贷方案例编号：XYZ6543
贷款金额：552500 美元
利率：3.555%
贷款月数：360
分期偿还类型：固定利率

PENNYMAC

将文档处理时间从数小时缩短到**数分钟**

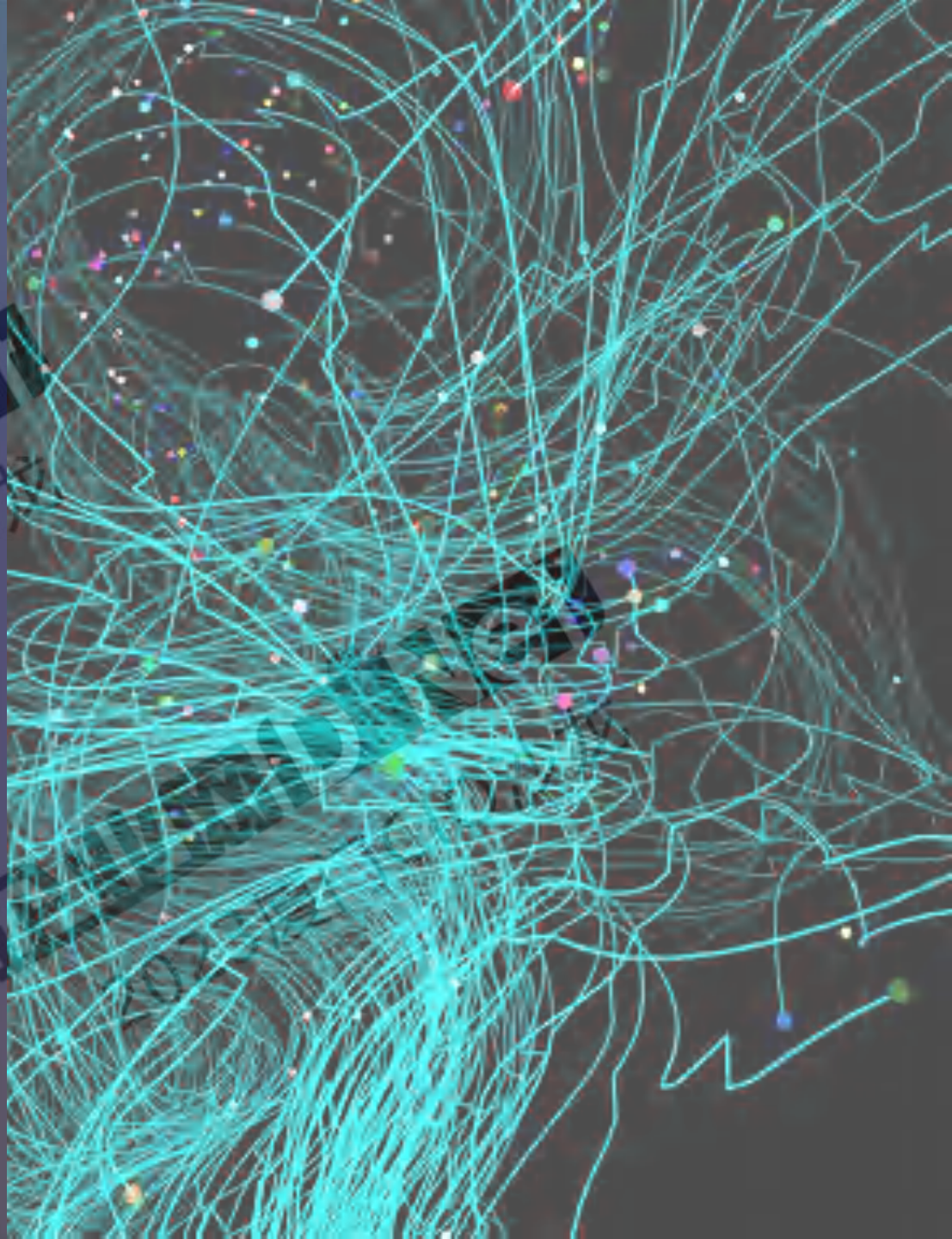
Elevance
Health

自动对**90%**的文档进行分类，以用于索赔处理

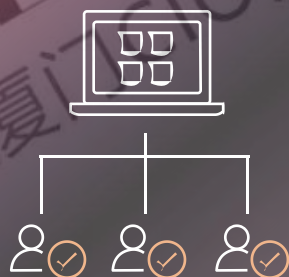
bluevine

每天自动处理数以**万计**的贷款申请

趋势5： 负责任的人工智能



什么是负责任的人工智能？



公平性



可解释性



鲁棒性



隐私性和安全性



治理



透明度

帮助负责任地构建机器学习服务和功能



隐私与安全

Amazon Macie

Amazon Lake Formation

Amazon CodeWhisperer

Amazon Glue DataBrew

Amazon security/identity services



公平性

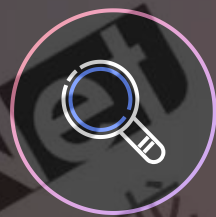
Amazon SageMaker Clarify

Amazon SageMaker Model Monitor

Amazon SageMaker Data Wrangler

Amazon SageMaker Autopilot

Amazon Augmented AI (A2I)



可解释性

Amazon SageMaker Clarify

Amazon SageMaker Model Monitor

Amazon SageMaker Model Registry

Amazon SageMaker ML Lineage Tracking



Amazon SageMaker Role Manager

在几分钟内定义最低权限



Amazon SageMaker Model Cards

记录、检索和共享必要的模型信息



Amazon SageMaker Model Dashboard

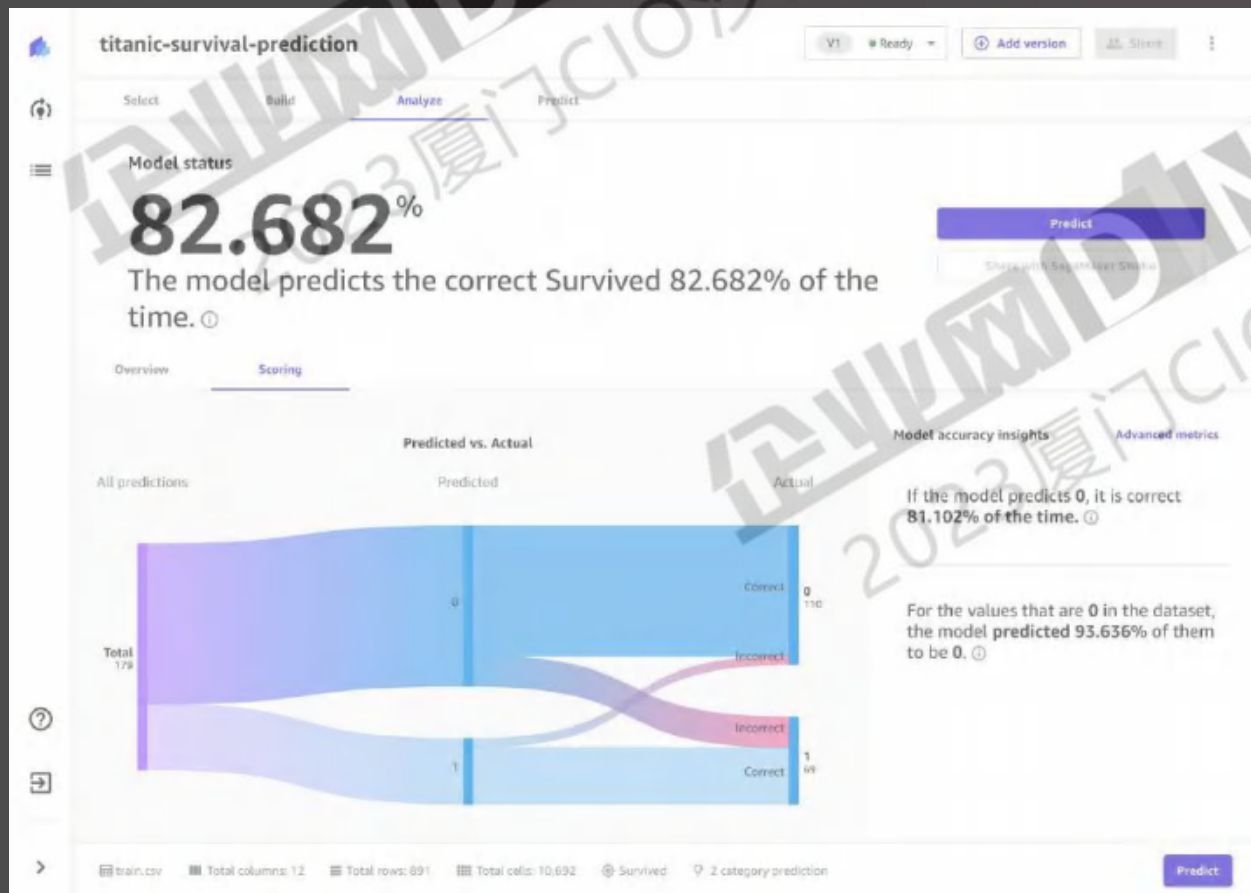
在统一视图中监控模型性能

趋势6: 机器学习普惠化



Amazon SageMaker Canvas

构建机器学习模型并生成准确预测——无需代码



SAMSUNG

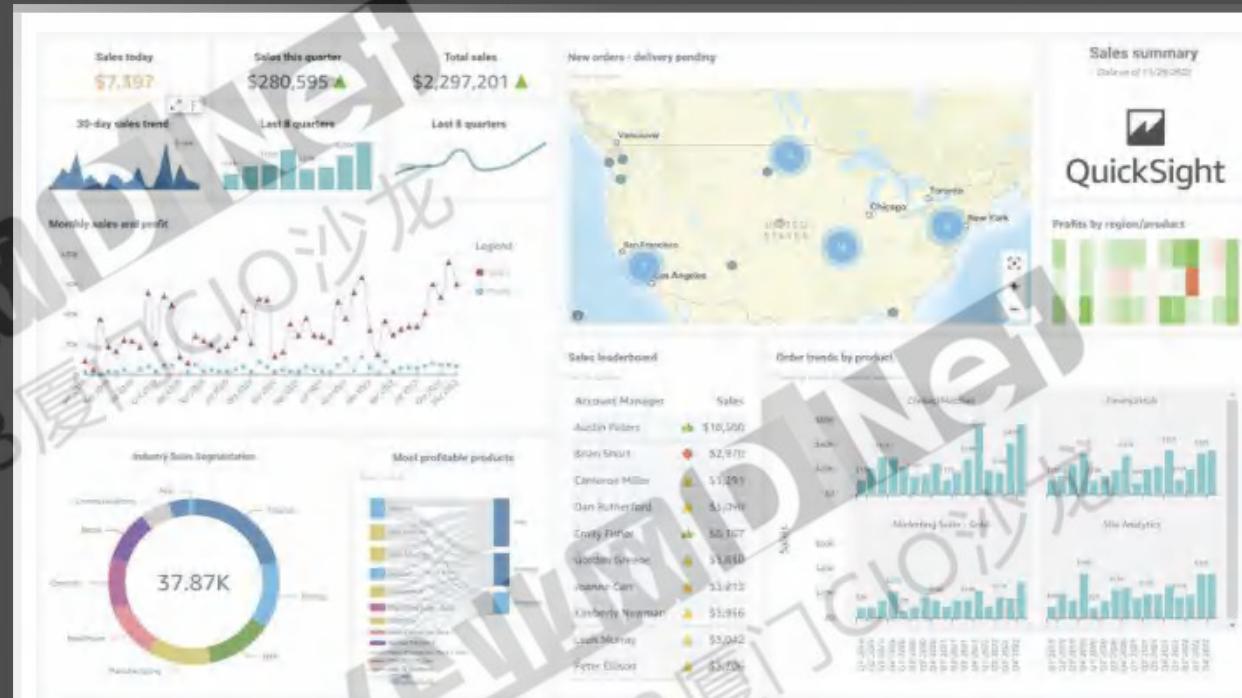
3M

SIEMENS

Amazon QuickSight

为云构建的可扩展、可嵌入的基于
ML的商业智能服务

用自然语言咨询数据问题，并可获
得可视化答案



亚马逊云科技推出的低代码机器学习

Amazon SageMaker

准备

数据探索、数据准备和特征工程

构建

模型开发

训练和调整

模型训练和调整

部署和管理

实时或批量预测：模型部署、托管和监控

Amazon SageMaker Data Wrangler
有助于更加快速、直观地为机器学习聚合和准备数据

Amazon SageMaker Autopilot
AutoML功能可自动准备数据，并根据您的表格数据集构建、训练和调整最佳机器学习模型

Amazon SageMaker JumpStart
一个模型中心，其中汇集了预训练和易调整的最先进模型，可用于计算机视觉和自然语言处理任务

数据治理与数据安全

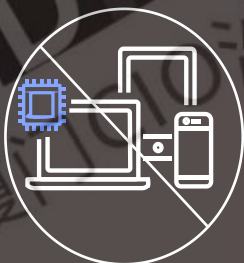
客户面临的挑战



众多数据中
难以找到所需数据



分析师
访问数据费力耗时



访问数据
的工具纷繁复杂



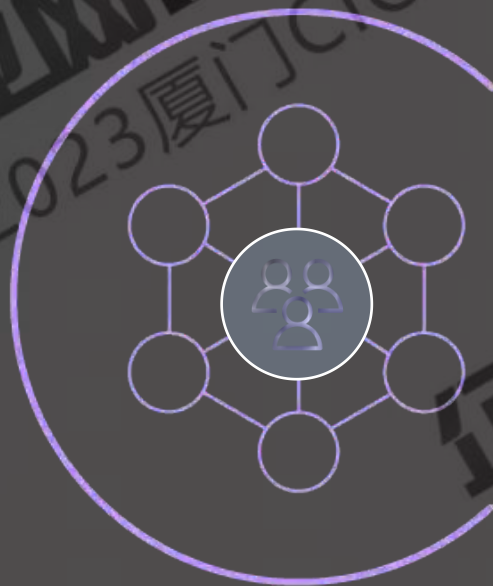
没有简单
的协同环境



数据治理
隐藏在各自工具中

通过治理实现数据普惠化

DATAZONE将人、数据、工具融为一体释放数据的力量



Data producers

数据生产团队



*Amazon
DataZone*
数据管理团队



Data consumers

数据消费团队

Amazon DataZone

利用DataZone实现数据可安全地，发现、访问、协同、治理

全新发布

Preview



企业范围的 业务数据目录

通过**业务语言**方式，让用户能够更快、更容易地找到和理解数据

通过丰富的元数据和**业务语言**，梳理编目整个企业、云上、本地或**SaaS**应用的**任何数据类型**



自助服务门户

体验集成的数据环境，通过个性化页面**促进探索**和**推动创新**

以**自助方式**使用数据和工具的同时，简化**跨职能的协同**



简化数据的访问

创建基于业务用例数据项目，**对团队、工具和数据**进行**分组**

无缝地使用您选择的工具，而无需在服务之间切换时进行无差别的繁重操作



受治理的数据共享

自动化工作流，以促进生产者 and 消费者之间更**安全地共享数据**，以确保正确的用户出于正确的目的访问正确的数据

高效地**审计**谁在使用数据，出于什么业务场景，并**监控**跨项目组和业务线的**使用情况和成本**

Thank you!