CELENT

在 IBM Z16 上部署防范欺诈方案

减少银行、卡片和支付方面的损失

Neil Katkov

2022 年 4 月 5 日

本报告由 IBM 委托编制，IBM 委托 Celent 代表其设计制作并展开 Celent 调研。分析和结论由 Celent 独立作出，IBM 对报告内容没有编辑控制权。

目录

执行摘要

深度学习等人工智能 (AI) 领域的进步让欺诈检测得到显著改进。然而，由于欺诈检测系统的吞吐量和延迟限制，使用 AI 模型的大型银行和支付处理商通常只在一小部分交易中运行相关功能。结果导致许多欺诈性交易未受到监控和检测。

IBM Integrated Accelerator for AI 是 IBM 新型 Telum 大型机处理器的一部分，旨在为大规模和低延迟的实时工作负载运行推理。该芯片的设计可支持实时欺诈检测，即使在大批量银行、卡片或支付处理环境下也能游刃有余。

为了帮助银行和支付处理商了解这种创新在遏制欺诈操作中可发挥的潜在价值，Celent 进行了一项评估，以了解如果这些实体将 AI 推理应用于 100% 的交易，可能会减少多少欺诈损失。

IBM z16 大型机上基于 AI 的欺诈检测方案所具备的可量化优势：

减少行业欺诈损失		每家银行的损失减少额		卡片交易被拒减少率
美国	全球	一级美国银行	二级美国银行	
每 $100 美元 5.6 美分	每 $100 美元 2.0 美分	1.05 亿美元	1,800 万美元	46%

据 Celent 估计，理论上讲，如果将高级推理模型应用于在 IBM zSystems 大型机上运行的所有银行、卡片和支付交易，全球范围内可减少的欺诈损失约为 1,610 亿美元。其中，银行可避免 1,400 亿美元的损失，而卡片和支付方面可避免 210 亿美元的损失。仅在美国，银行欺诈损失就可能减少 440 亿美元，卡片和支付方面可能减少的损失可达 60 亿美元。

可以肯定的是，在大型机上采用人工智能推理来检测并预防欺诈操作确实存在障碍，例如模型治理问题、淘汰和替换成本、内部数据科学资源的可用性以及业务案例演示。

尽管如此，直接在大型机环境中运行高级 AI 模型是银行业界的一项具备强大创新意义的方案，据估计，全球有 70% 的交易金额在 IBM 大型机上运行。欺诈检测是 IBM 新功能的一个重要用例，它对提升企业盈利和客户体验都有显而易见的益处。

银行、卡片和支付中的高额欺诈成本

据估计，2021 年，全球银行、卡片和支付领域因欺诈而遭受的损失金额高达 3,850 亿美元。

银行和支付欺诈在零售和企业领域表现为多种形式。针对银行的欺诈包括账户接管、授权推送支付 (APP) 欺诈、发票欺诈，以及旨在触发非法转账或获取账户凭证的各种网络钓鱼和社会工程骗术。卡片和支付也容易受到账户接管和网络钓鱼以及包括合成 ID、破产欺诈和中间人欺诈在内的特定骗局攻击。

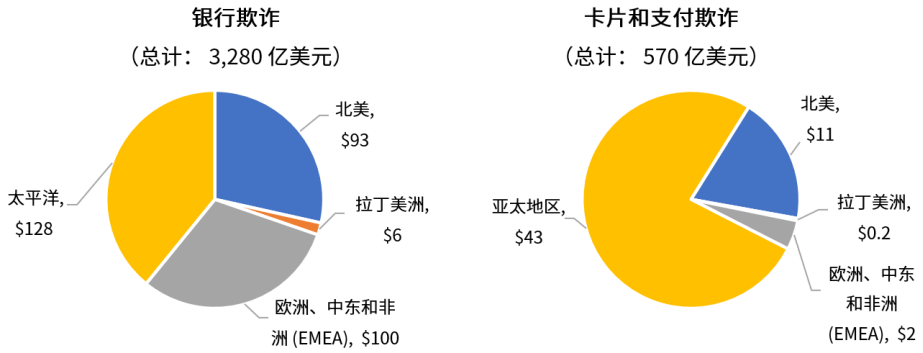
图 1：常见的银行和卡片欺诈骗局

银行欺诈		卡片欺诈	
	账户接管		应用程序欺诈
	APP 欺诈		透支欺诈
	支票欺诈		中间人攻击
	发票欺诈		网络钓鱼
	社会工程		合成 ID

资料来源：Celent

这些以及其他针对银行账户、卡片和支付的欺诈行为是金融机构严重关切的问题。据 Celent 估计，美国一级银行（总资产超过 1,000 亿美元）的年欺诈损失平均为 2.09 亿美元，二级银行（总资产在 500 至 1,000 亿美元之间）的年欺诈损失平均为 3,500 万美元。从行业层面来看，2021 年全球银行遭受的欺诈损失总计达到了 3,280 亿美元。另外，卡片和支付领域还遭受了 570 亿美元的欺诈损失。2021 年，欺诈给全球银行、卡片和支付行业造成的损失合计约 3,850 亿美元。

图 2：2021 年银行、卡片和支付类欺诈损失



资料来源：Celent 根据 BIS 交易数据和中央银行欺诈数据所做出的估算。

注：银行欺诈涉及转账、直接借记和支票。卡片和支付欺诈包括信用卡和借记卡、电子支付以及其他支付方式。

为了遏制欺诈，银行和支付处理商已通过检测系统和基于芯片的卡片安全设计与之进行了数十年的斗争，但欺诈者却通过设计基于新技术和社会工程的新伎俩，始终保持领先一步，导致损失持续攀升。

新冠疫情推高了欺诈数量。对银行而言，网络钓鱼和社会工程骗局是欺诈的一个重要来源，这些骗局利用有关大流行的焦虑和医疗需求大行其事。至于卡片交易，由于消费者避免了在网点和店内交易，疫情导致了数字银行和电商购物交易次数的增加。由于无卡交易 (CNP) 成为卡片欺诈的重灾区（约 65%），因此卡片欺诈的损失有所增加。

援兵即至：基于深度学习的欺诈模型

现在，人工智能（例如深度学习）的进步，为银行提供了更有效地打击欺诈的工具，具体方法是通过大规模分析数据来找到指向欺诈的模式，包括前所未见的新发类型。

深度学习是一种基于深度神经网络 (DNN) 的机器学习模型。DNN 由计算节点或神经元组成，并使用渐进式权重来加强节点之间的连接。节点采取多层排列，形成一个 深度 网络，增加了模型的容量和学习率。深度学习模型可在现有数据上进行训练，例如欺诈模型中的历史交易。然后在实时数据（例如实时交易）上执行经过训练的模型，以生成结论或推理结果。就欺诈模型而言，推理结果通常是一个得分，表示构成欺诈交易的可能性有多大。

Celent 根据行业对话和研究估计，基于深度学习模型的 AI 推理可以在现有欺诈模型的基础上将欺诈检测的准确性提高 60%。

然而，在大容量的大型机环境中，由于延迟、成本和客户摩擦等问题，用推理改善欺诈率的潜力非常有限。通常这些模型仅用于处理一小部分的交易，数量大概不到 10%。这意味着大约 90% 的可预防潜在欺诈行为仍未被检测到。这严重制约了银行利用 AI 技术挽回欺诈损失的能力。

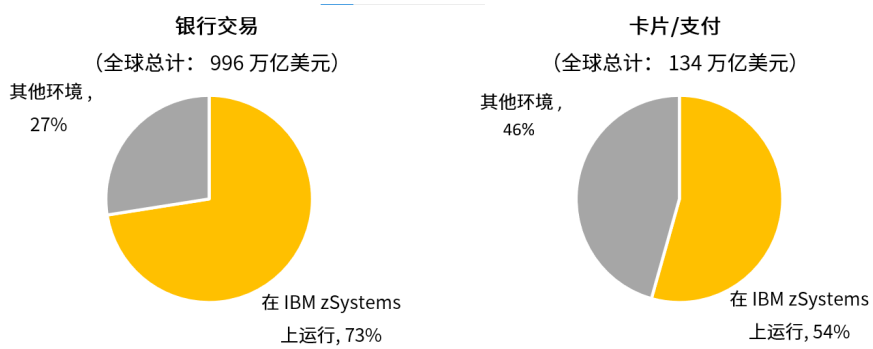
而如今，100% 的银行和卡片交易全部可以采取高级模型进行检测，其中所涉及的延迟和成本障碍或许已不再是个需要担忧的问题。新型 IBM z16 Telum 处理器包含一个 AI 加速器，这是 IBM zSystems 的首创，可以直接在芯片上实时运行 AI 模型。这一创新之举巨幅提升了吞吐量和加快了响应时间，首次确保几乎所有交易都可通过基于深度学习的欺诈检测模型完成检测。

现有欺诈检测的局限性

对于大型机环境，典型的欺诈检测技术和操作方法包括在平台外系统上针对选定的交易运行欺诈检测，和/或在交易后运行欺诈检测。这极大地限制了银行和支付处理商在所有交易中运行高级 AI 模型的能力。

许多大型银行和支付处理商均在大型机计算环境中运行其核心系统。据 IBM 估计，全球前 50 家银行中有 45 家在 IBM zSystem 大型机上运行其系统。大多数主要的卡片和支付处理商也在该平台上运行。据 Celent 估计，全球范围内有 70% 的银行、卡片和支付交易总额在 IBM zSystems 环境中运行。

图 3：IBM zSystems 上的银行、卡片和支付交易总额



资料来源：Celent

对于某些交易类型，可以容忍核心系统和平台外检测系统之间出现延迟。但是，对于应用于实时交易（例如实时支付、卡片交易和数字银行交易）的数据密集型 AI 推理例程，延迟使得在大容量环境下将所有交易提交 AI 检测平台进行检测的做法并不切实际。当核心系统交易从大型机发送到平台外检测系统进行实时分析时，接收检测结果的响应时间在 50 到 80 毫秒之间 ┗ 而交易在此期间处于等待状态。这会减慢交易的批准时间，从而造成客户摩擦，尤其是对于卡片交易类型。

更为重要的是，高延迟可能会导致无法通过平台外欺诈检测系统检测所有交易。核心系统和检测软件之间的延迟可能会使得核心系统过晚收到检测结果，从而导致实时交易超时。因此，一些银行仅在交易后运行深度学习模型来检测欺诈行为。

因此，银行仅能实时发送一小部分交易（不到 10%）至其欺诈检测引擎进行处理。这种方法会导致严重的后果。深度学习模型现在可将检出率大幅提升约 60%。然而，银行并没有全面获益，因为他们仅通过这些模型运行部分交易样本。这意味着还有更高比例的欺诈没有被被检测到，从而增加欺诈所带来的损失。随着欺诈成为金融类犯罪的焦点问题，如果银行无法将其所有交易进行反欺诈检测，银行也可能会面临监管风险。

一级美国银行存在的遗留问题

美国一家一级银行在 IBM zSystems 平台上运行其核心系统，同时还部署了一个基于 AI 的平台外欺诈检测系统。由于成本和延迟问题，银行仅通过 AI 系统检测风险极高的交易。为方便客户，大多数交易都是通过基于规则的评分系统进行审批，然后在事后进行交易后分析。由于无法在所有交易上运行模型，AI 的优势受到严重制约，这意味着 AI 并没有充分发挥其潜力。

在大型机上利用 AI 推理来减少欺诈损失

IBM 为其 IBM z16 大型计算机开发了一款处理器，该处理器采用了一个 AI 加速器，目的是直接在芯片上大规模运行高级推理计算。据 Celent 估计，新的 IBM z16 处理器可以支持对几乎所有交易进行基于深度学习的欺诈检测，从而有可能在全球范围内减少 1,610 亿美元的银行、卡片和支付欺诈损失。

与传统欺诈模型相比，深度学习算法的计算强度更高。随着银行针对欺诈实施基于深度学习的 AI 推理计算，他们在管理这些最关键的工作负载方面遇到了一些困难。在平台外系统上执行检测时，检测响应时间可长达 80 毫秒以上，吞吐率在每秒 1,000-1,500 笔交易 (tps) 范围内。

由于存在种种延迟和吞吐量限制条件，银行在等待检测结果时会遇到交易超时的问题。上述问题和其他一些问题导致银行仅向其检测引擎发送一小部分交易（不到 10%）。

大型机上的深度学习

基于信用卡欺诈深度学习模型的 32 个 IBM Telum 芯片整合至单个服务器运行，每秒可提供高达 350 万次推理，平均响应时间为 1.2 毫秒。

资料来源：IBM 微基准测试，2021 年 8 月

免责声明：性能结果是从 IBM 内部测试中推算而来。

IBM 为其 IBM z16 大型计算机开发了一种加速器，可直接在芯片上运行 AI 推理模型。IBM 表示，在大型机上运行 AI 模型的吞吐量和性能改进足以支持对几乎所有交易进行实时欺诈分析，即使在大容量的银行、卡片或支付处理环境中也能胜任。

此外，分析可在几乎不影响交易处理时间的情况下完成。IBM 表示，IBM Integrated Accelerator for AI 是其新 Telum 处理器的一部分，可在大型机上运行 AI 模型，每个推理请求的响应时间非常快，仅为 1.2 毫秒。就信用卡欺诈检测这一具体用例而言，早期的基准测试表明，32 个 Telum 芯片的配置即可支持每秒高达 350 万次的推理。

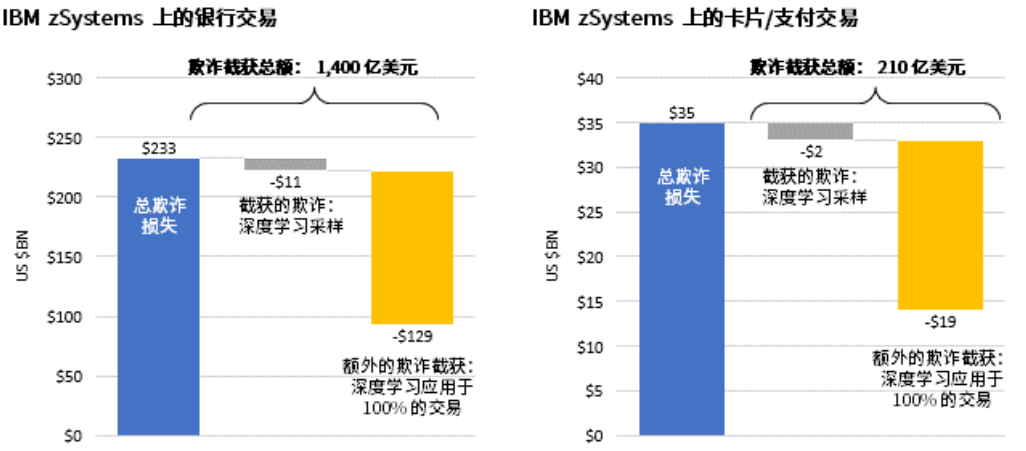
这足以支持峰值交易流，确保银行和支付处理商可以通过深度学习模型运行几乎所有交易。

银行以及卡片和支付处理商可以针对所有交易运行高级模型，以此充分发挥现代推理技术的潜力。据 Celent 估计，将高级推理模型应用于所有交易可能会大幅减少欺诈损失，全球每 100 美元交易减少 2.0 美分（2.0 个基点）。

美国的欺诈率高于全球平均水平，每 100 美元为 9.3 美分，而全球平均水平为 3.7 美分。因此，在美国，每 100 美元可减少 5.6 美分的欺诈损失。这相当于每 2,375 美元的交易可为银行节省 1.33 美元。

据 Celent 估计，理论上，若将当前在 IBM zSystems 上运行的所有交易提交深度学习模型分析，可在全球范围内减少 1,610 亿美元的欺诈损失。其中，银行可以避免 1,400 亿美元的欺诈损失；卡片和支付方可以避免 210 亿美元的欺诈损失。仅在美国，银行减少欺诈损失的潜力为 440 亿美元，卡片和支付处理商减少欺诈损失的潜力则为 60 亿美元。

图 4：使用深度学习模型减少欺诈损失的潜力



资料来源：Celent

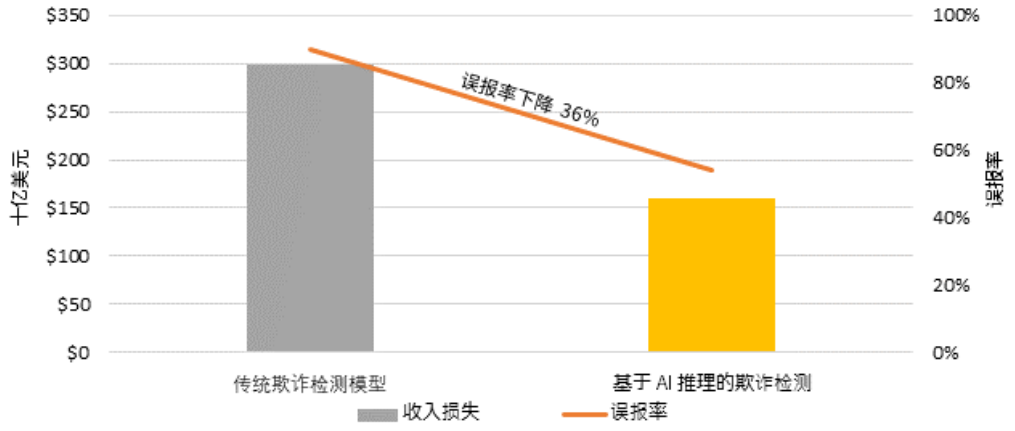
据 Celent 估计，对于使用 IBM z16 的一级银行而言，通过高级推理模型运行所有交易（与仅将 AI 模型应用于约 10% 交易的当前最佳实践相比）可以额外减少 1.05 亿美元的欺诈损失。二级银行可以额外避免 1,800 万美元的损失。通过高级模型运行所有交易也将反过来改进模型本身。更多的交易会产生更多的数据来训练模型，从而提高欺诈检测的准确性。

控制误报以减少客户流失

传统的反欺诈模型具有非常高的误报率，通常占所有警报交易的 90% 或更高，这会导致银行拒绝合法交易。居高不下的误报率和交易被拒率不仅会造成客户摩擦，还会导致实打实的经济损失，因为客户只需取出另一张信用卡或借记卡即可进行购买。据 Celent 估计，信用卡交易的减少使该行业在全球损失了 2,980 亿美元的费用收入。

银行将欺诈检测例程局限于部分交易样本的另一原因是，我们需要在反欺诈工作与最大限度地减少客户摩擦之间找到平衡。当检测软件错误地将合法交易标记为欺诈时，就会出现误报。深度学习模型的高准确性可以显著改善业内居高不下的误报率。这反过来会减少交易被错误拒绝的几率，从而改善客户体验并减少因客户流失而造成的收入损失。这也意味着银行可在对所有交易进行欺诈检测的同时，减少因客户摩擦而造成的损失。

图 5：深度学习模型降低误报率



资料来源：Celent

应用于每笔卡片交易的深度学习模型可以将误报率降至 55% 左右。虽然这个数字仍然很高，但这已有可能在全球范围内减少 1,370 亿至 1,610 亿美元的卡费损失。

更少的误报还会带来其他益处。欺诈分析师需要减少警报，从而降低交易后调查的成本。在声誉利益方面，减少客户摩擦和挫折感将增强商誉和客户信任度。

高级模型还可以改进对可能提示为疑似洗钱行为的检测。美国《银行保密法》、欧盟《反洗钱指令》和其他法规将银行的反洗钱计划 (AML) 置于监管机构的严格审查之下。美国监管机构频频指责银行的反洗钱计划不到位，对一些银行的罚款超过 10 亿美元。反洗钱业务的误报率也非常高，通常超过 95%，这给银行带来了沉重的运营负担。此外，反洗钱监控流程通常在交易后进行，这使银行面临更大的风险。而将基于 AI 的模型用于反洗钱操作流程，可提高反洗钱行为检测的准确性并减少误报，以帮助解决此类问题。

未来路径

我们的分析指出，在高达 100% 的交易上运行深度学习模型可带来显著的、可量化的好处。IBM 表示，其新型加速器支持对 IBM z16 大型机上运行的交易进行深度学习模型分析，即使在极高容量的环境中也能如此。然而，对于正在实现技术跨越的银行和支付处理商，仍有许多因素需要考虑。

在银行以及卡片和支付处理商权衡在大型机上实施基于深度学习的欺诈检测模式而带来的优势时，Celent 建议他们考虑以下问题：

- **模型治理。**监管机构和内部审计师需要围绕欺诈模型进行强有力的治理。这意味着 AI 模型必须透明且可解读。虽然 AI 平台供应商总体上正在摆脱 黑盒 方法，但 AI 模型的治理仍然是一项复杂的工作。
- **监管阻力。**监管者对基于规则的传统检测感到满意，但对先进的深度学习技术并不太熟悉。在向前推进的过程中，在某些情况下，银行、数据科学家及其供应商可能需要向监管机构宣传先进 AI 技术的有效性和可靠性。
- **更换费用。**许多机构已经部署了基于人工智能的欺诈检测系统。这些公司需要开发将检测转移到大型机的业务用例，包括决定是否以某种形式保留现有系统（例如，支持交易后分析或较小的业务线），或完全废弃它们。
- **数据科学资源。**IBM 的 Integrated Accelerator for AI 针对多种模型进行了优化，包括使用 Pytorch 和 TensorFlow 等开源框架构建的模型。尽管它尚未被证明可支持打包的欺诈检测软件，但我们预计，一些欺诈供应商最终会推出可以在加速器上运行的软件包。无论哪种情况，将基于 AI 的检测迁移到 IBM z16 的机构都需要数据科学能力来开发并支持先进的欺诈深度学习模型，无论是在内部还是通过专业模型提供商。

金融机构将需要仔细考虑这些因素，并对 IBM 的新型 AI 加速器进行尽职调查。尽管如此，在减少因欺诈和拒绝交易所造成的损失以及减少摩擦和改善客户体验方面，其潜在益处是不言而喻的。运行 IBM zSystems 的公司最好能仔细研究将欺诈检测转移到大型机可获得的益处。

利用 CELENT 的专业知识

如果您认为此报告对贵公司有价值，您可以考虑与 Celent 合作开展定制的分析和调研活动。我们的经验组合以及我们在编写本报告时获得的知识可帮助您简化策略的创建、改进或执行过程。

对金融机构的支持

我们支持的典型项目包括：

供应商的筛选和选择。我们针对贵公司以及贵公司的业务进行特定的研究和发现，可更好地了解你们的独特需求。然后，我们为选定的供应商创建和管理定制的 RFI，以帮助贵公司快速准确地选择供应商。

商业实践评估。我们花时间评估贵公司的业务流程和要求。根据我们对市场的了解，我们会识别出潜在的流程或技术限制，并提供清晰的洞察分析，帮助贵公司实施行业最佳实践。

IT 和业务战略创建。我们从贵公司的管理团队、一线业务人员和 IT 员工以及客户那里收集观点和建议。然后，我们会根据贵公司的目标分析你们当前所处的位置，以及当前所拥有的机构能力和技术。如有必要，我们会帮助你们重新制定技术和业务计划，以满足短期和长期需求。

对供应商的支持

我们的服务可帮助贵公司改进产品和服务。例如：

产品和服务战略评估。我们帮助评估贵公司在功能、技术和服务方面的市场地位。我们的战略研讨会将帮助贵公司锁定正确的客户，并将你们的产品与他们的需求相匹配。

市场讯息和相关评估。根据我们在贵公司潜在客户方面所掌握的丰富经验，我们将对你们的营销和销售材料进行评估，包括你们的网站和任何相关的营销工具。

与本主题相关的 CELENT 研究

Remaking Risk:A Taxonomy of Regtech

2021 年 10 月

Technology Trends Previsory:Risk，2022 年版

2021 年 10 月

IT and Operational Spending in AML-KYC:2021 年版

2021 年 12 月

IT and Operational Spending on Fraud:2021 年版

2021 年 2 月

Innovation In Risk:A Snapshot Through the Lens of Model Risk Manager 2021

2021 年 4 月

Fino Payments Bank:Remote Implementation of Enterprise-Wide Fraud Management During the Pandemic

2021 年 3 月

Swedbank:Modernizing Card Fraud Management and Improving Customer Experience

2021 年 3 月

版权声明

对于某金融机构应如何执行其战略，本报告并不能取代量身定制的专业建议。本报告并非投资建议，亦不应作为此类建议的依据或替代专业会计师、税务、法律或财务顾问的咨询建议。Celent 已尽一切努力使用可靠、最新和全面的信息和分析结果，但不就所提供的一切信息给出任何明示或暗示的保证。本报告的全部或部分内容所依据的，由其他方所提供的信息据信是可靠的，但并未经过验证；此类信息在准确性方面不作任何保证。公共信息以及行业和统计数据来自我们认为可靠的渠道；但是，我们对此类信息的准确性或完整性不作任何表述，并且我们是在未经进一步验证的情况下接受了此类信息。

Celent 不负责更新本报告中的信息或结论。对于因受本报告所含信息或本文所述任何报告或信息来源的影响而采取或未采取某种行动，从而导致的任何损失，或任何间接的、特殊的或类似的损害，Celent 不承担任何责任，即使已被告知此类损害的可能性。

本报告没有第三方受益人，我们不对任何第三方承担任何责任。本报告所表达的意见仅对本报告所述目的有效，其有效性仅截至本报告发布之日。

对于市场条件或法律法规的变化，我们不承担任何责任，也不承担修改本报告以反映本报告日期之后发生的变化、事件或条件的义务。

如需获取更多信息，请联系 info@celent.com 或：

Neil Katkov nkatkov@celent.com

美洲

美国
99 High Street, 32nd Floor
Boston, MA 02110-2320

+1,617,424.3200

美国
1166 Avenue of the Americas
New York, NY 10036

+1,212,345.8000

美国
Four Embarcadero Center
Suite 1100
San Francisco, CA 94111

+1,415,743.7800

巴西
Rua Arquiteto Olavo Redig
de Campos, 105
Edifício EZ Tower – Torre B – 26º andar
04711-904 – São Paulo

+55 11 3878 2000

欧洲、中东和非洲 (EMEA)

瑞士
Tessinerplatz 5
Zurich 8027

+41.44.5533.333

法国
1 Rue Euler
Paris 75008

+33 1 45 02 30 00

意大利
Galleria San Babila 4B
Milan 20122

+39.02.305.771

英国
55 Baker Street
London W1U 8EW

+44.20.7333.8333

亚太地区

日本
Midtown Tower 16F
9-7-1, Akasaka
Minato-ku, Tokyo 107-6216

+81.3.6871.7008

香港
Unit 04, 9th Floor
Central Plaza
18 Harbour Road
Wanchai

+852 2301 7500

新加坡
138 Market Street
#07-01 CapitaGreen
Singapore 048946

+65 6510 9700