

2020 亚马逊 云计算的轨迹线

通过这些重磅产品服务，
一窥2020年亚马逊云计算的技术脉络



观看亚马逊re:Invent 2020精彩视频
了解AWS最前沿的技术资讯



进入AWS开发人员中心
获取丰富的开发资源



快速上手
教你如何构建AWS应用程序



添加AWS小助手
加入技术交流群，共同成长



CONTENTS / 目录

人工智能

变身漫画道具玩了没？这样的AI人像特效机器学习服务帮你实现
机器学习项目复杂繁琐，该如何构建安全的机器学习环境？

数据智能

企业出海服务为何这么难？
不懂业务的开发人员，就不是好的数据科学家
数据湖火了，那数据仓库怎么办？

容器服务

AWS打造容器服务体系，Amazon EKS这块拼图有多重要？

无服务架构

开发运维配置繁杂，是时候给应用架构做减法了

开发人员相关资源

卷首语

时光飞快，一转眼已到了2021年。回望在人类历史长河中都会被牢记的2020年，亚马逊云计算也经历了非常多的不一般。

作为机器学习技术领导力的持续引领者，亚马逊云计算围绕着拳头产品 Amazon SageMaker 机器学习开发平台，进行了大量技术更新和迭代，并尤其强调“把机器学习交付到每一位开发者手中”。在刚过去的年度 re:Invent 大会期间，与机器学习相关的新产品，就发布了20多款之多。

“无数据不成云”，数据已是云计算的“基础工程”。对上，数据为人工智能等新型应用提供基础，而对下，数据如何在各种不同需求下都能平滑地“存、管、用”，这对底层计算与存储提出了新的要求。此外随着数据处理技术的不断成熟，如何让数据技术更好地契合业务应用需求，这些都是不断演进发展的课题。

云中一切的源头非计算莫属，因此计算作为基础，其重要性再怎样说都不为过。继早先的“虚拟化”，后来的“容器”等风潮之后，现在“云原生”成为了“当红炸子鸡”。池化、解耦、可迁移等，在这个路径演进中得以持续深入。这其实呼应了网络化应用形态从量变到质变的过程，即：过去以终端计算为中心，现在越发以云为中心来构建应用模型。这样的发展变化是深刻且不可逆的。又由于疫情的推动，企业上云的需求刚性达到了历史高位，直观地体现着市场对云计算基础的趋势性需求。

以下的文章选编，只是2020年 AWS 整体技术市场演进的“沧海一粟”，希望透过这些点滴的代表，为大家折射出这一段的发展，并更好地展望新的一年。

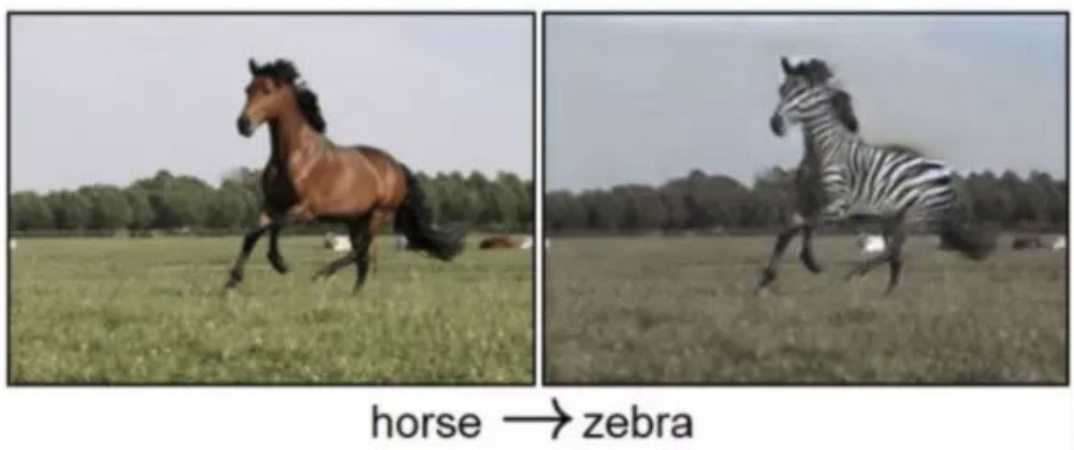
变身漫画道具玩了没？这样的AI人像特效机器学习服务帮你实现

作者 | 冯垚

近期，抖音上一款“变身漫画”的特效刷爆“我的关注”，二次元漫画样式的画风更是让大家欲罢不能。从明星到路人，从大朋友到小朋友纷纷参与其中，抖音App中“变身漫画”话题页显示约有1851万余人使用该道具。



如今，此类视频 / 图片风格转化特效，画面细节精致、色彩温和，颇受欢迎。其实它并非新鲜事物，通常此类风格转化是机器学习模型创造出的虚拟场景，主要采用生成对抗网络（GAN）技术。



生成对抗网络（GAN）是一种机器学习模型，由深度神经网络架构构成。生成对抗网络（GAN）由生成网络和判别网络组成，生成网络产生“假”数据，判别网络对生成数据进行真伪判断，通过卷积神经网络进行分析，在训练过程中，两个网络持续进行对抗，直至判别网络无法识别“假”数据，其被广泛应用于媒体、游戏、娱乐等领域。

生成对抗网络抗网络（GAN）可基于 PyTorch 机器学习框架构建，因涉及从构建到部署的整个机器学习流程，所以无法避免机器学习全流程中存在的诸多难题。首先，对于完整的机器学习项目而言，落地是一项纷繁复杂的工作且成本高昂。其次，整个项目从选择算法、优化算法到参数调节，影响模型准确性的参数有百万之多。最重要的是，需要开发者和数据科学家具有广而精的专业知识。

因此，一种更为友好的机器学习服务是如此重要。而云计算时代的到来，扫清了广大企业应用人工智能和机器学习的障碍，让简化机器学习项目成为可能，并且开发者又能得到庞大算力、数据存储等云能力的支持。根据 Gartner 预测，到2023年将有超过40%的开发团队将使用自动化机器学习服务。可见，机器学习从业者急需适用于整个机器学习工作流程的云端服务或平台。

AWS 全方位支持 AI 在云端生长

20多年来，Amazon Web Services (AWS) 在人工智能和机器学习方面进行的大量投入，获得了国际分析机构的高度认可。根据2020年 Gartner 公布的《云 AI 开发者服务魔力象限》显示，AWS 被评为该领域的领导者。



AWS 可根据用户的不同需求，为开发者和数据科学家提供全方位的机器学习和人工智能解决方案，包括三层的服务堆栈。其中，底层为机器学习的框架和基础架构，用户可以自由选择 PyTorch、TensorFlow、Apache MXNet、Chainer、Gluon、Horovod 和 Keras 等机器学习框架。顶层为训练好的人工智能服务，如视觉类、文字类、语音类等，用户可以直接调用相关服务，无需过多关注服务背后机器学习模型的复杂流程。



尤为重要的是中间层，AWS 基于机器学习的核心服务层推出了完全托管服务

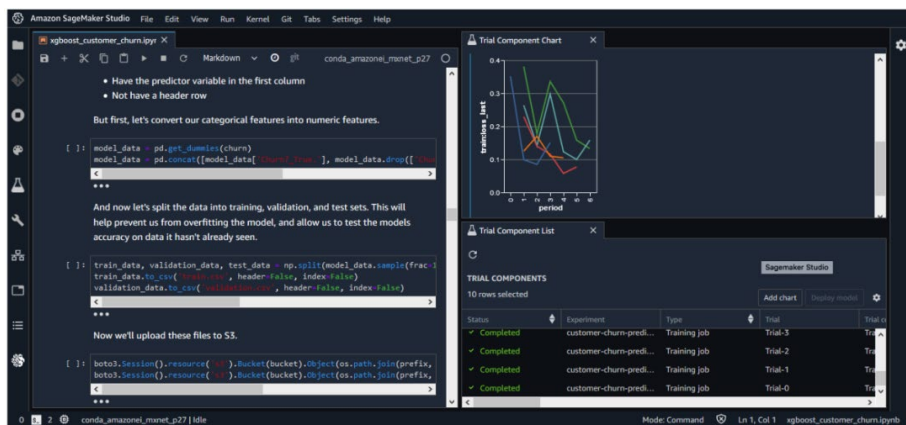
—— **Amazon SageMaker**。它更像是机器学习从业者的“工具包”，可集中提供用于机器学习的所有组件，贯穿机器学习项目从构建、训练到部署的整个工作流程，消除了传统机器学习复杂、昂贵、迭代的弊端，帮助开发人员和数据科学家更快速、高效地将机器学习模型投入实际应用中。

就中国市场而言，随着越来越多的企业和开发者意识到机器学习和人工智能的潜力，对于机器学习服务的需求愈发迫切。根据 IDC 预计，到2023年中国人工智能市场规模将达到979亿美元。AWS 适时而为，2020年4月30日 AWS 正式将 Amazon SageMaker 落地中国，在由西云数据运营的 AWS 中国（宁夏）区域和光环新网运营的 AWS 中国（北京）区域上线运营。

Amazon SageMaker 助力机器学习的黄金时代

宏观来看，机器学习项目整个过程是高度迭代的，需要多方协助。随着机器学习项目规模的增加，整个工作流程变得过于复杂，为此大多数机器学习项目中需要使用多种工具来满足各个阶段工程师的需求。

通常，大多数项目遵循数据科学家进行研究、实验及验证，开发者负责模型落地的流程，而最终实现多方协作并非易事。Amazon SageMaker 提供的 Amazon SageMaker Studio 适用于机器学习的完全集成式开发环境（IDE），其将所有用于机器学习的组件集中在一起，可以有助于提高整体项目的工作效率。



SageMaker Studio 为所有 Amazon SageMaker 功能和整个机器学习工作流

提供统一界面，使用者可以快速上传数据、创建新笔记本、训练和调优模型，帮助使用者全面掌控构建、训练和部署模型的每个步骤。

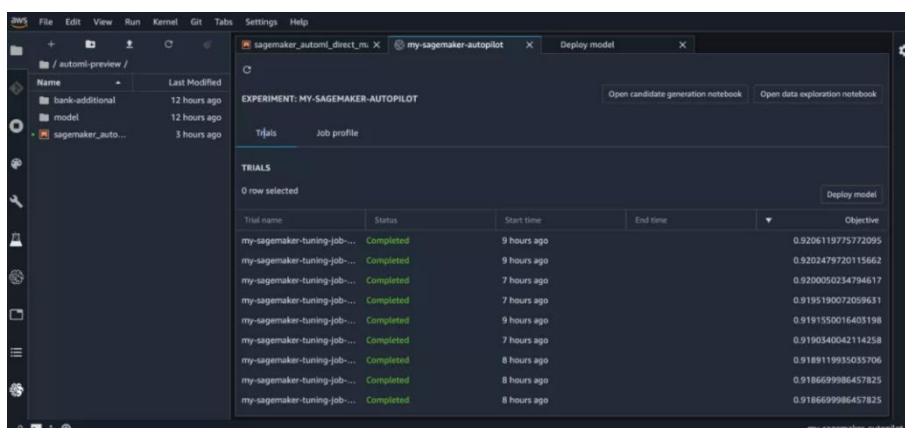
具体来看，机器学习项目工作流程分为以下几大类，AWS 可在各项流程中提供支持：

一、数据准备

机器学习项目的成功依靠优质的模型，优质的模型基于高质量的数据。Amazon SageMaker Ground Truth 可处理丢失或损坏的数据，帮助开发者快速构建和管理高度准确的训练数据集。并且，Ground Truth 可持续向人工制作的标签学习，以生成高质量的自动注释，同时可将数据标记成本降低70%左右。

二、构建

当数据进行了必要的转换和筛选处理，开发者又将面临如何选择算法与模型的困境。在设计机器学习模型时，虽然目前的一些机器学习自动化方法，已经实现了创建初始模型的功能，但只是停留于开发者的基础需求，无法满足开发者对于模型更深层次的需求，比如如何创建模型、改进模型以及获取相关数据等。



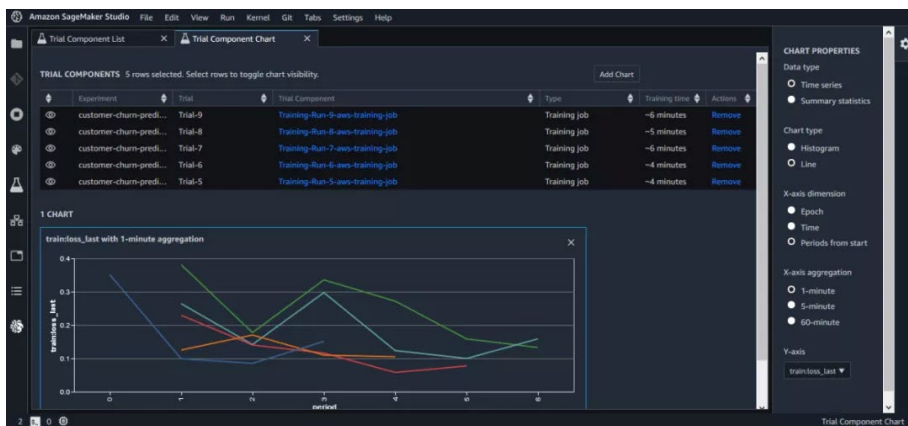
对于 AWS 而言，重视构建环节是夯实机器学习项目的关键。为此，Amazon SageMaker 提供自动构建模型的功能 Amazon SageMaker Autopilot，作为业内首个对机器学习模型保持控制和可见性的自动化机器学习功能，其可让开发者完全

掌控机器学习模型。一方面，对于缺乏经验的开发者 Amazon SageMaker Autopilot 可提供千余种算法、模型和数据。另一方面，经验丰富的开发者可将它用于开发基础模型，团队可以在此基础上进一步迭代。

同时针对开发者和数据科学家的特定需求，Amazon SageMaker 选择了最受欢迎工具之一——Jupyter Notebook。使用 Amazon SageMaker Notebook 可一键启动 Jupyter Notebook，使用者可以根据需求调整算力，从而快速实现构建模型并有助于增强团队内部的协作配合。

三、训练

训练机器学习模型是脆弱且困难的，通常需要进行多次迭代和不断调优。获得 NIPS 2017 年度时间奖的人工智能研究员 Ali Rahimi，在颁奖典礼上就曾抱怨过模型训练的脆弱性。假设一个机器学习模型要处理百万级别的记录数据，则至少需要数百次迭代。开发者该如何管理这些迭代的模型、训练数据和参数设置呢？



Amazon SageMaker Experiments 可以帮助开发者组织和跟踪机器学习模型的迭代。其将自动捕获输入参数、配置和结果，以可视化的形式帮助开发者管理迭代产生的结果。开发者也可自行保留实验的完整谱系，如出现模型偏离预期结果的情况，开发者可及时回溯和检查。

当解决了算法和模型的难题，整个项目的开发与运营成本自然会大幅降低。以“大宇无限”公司为例，采用 Amazon SageMaker 极大的简化了搭建机器学习

系统的构建、训练和部署流程。从始至终“大宇无限”无需顾虑开发成本与时间成本，算法工程师只需准备数据，估算约节省70%的训练成本。而根据 AWS 给出的资料显示，Amazon SageMaker 提供托管的 Spot 训练，最高可将训练成本降低90%。

此外，机器学习的发展仍然处于快速变化的进程中，开发者更需要一个开放的平台。在 Amazon SageMaker 上有着有非常开放的选择，特别是在中国市场。

(1) 对机器学习框架的开放：上文提到了 Amazon SageMaker 支持的众多机器学习框架，并且针对 TensorFlow、PyTorch、Apache MXNet 在内的主流框架进行了深度优化。例如，在 256GPUs 下，对 TensorFlow 的扩展效率可以做到90%，远高于开源平台的65%。值得一提的是，全球范围内85%的 TensorFlow 的工作负载跑在 AWS 上面。

(2) 对定制化的开放：Amazon SageMaker 已配备了多种内置算法，用户可以开箱即用来解决各类应用问题。但同时用户也可以将自己的算法或模型通过容器的方式用于 Amazon SageMaker，即使是早前的框架 Caffe 也都可通过自带容器的方式在 Amazon SageMaker 中使用。

(3) 对生态系统的开放：一个机器学习平台是否具有生命力，关键在生态系统。目前，AWS 已经与德勤、中科创达、伊克罗德等 APN 合作伙伴开展了合作，加速 Amazon SageMaker 在中国地区企业级市场的落地。

(4) 对开源社区的开放：考虑到目前很多机器学习负载已经容器化，Amazon SageMaker 为 Kubernetes 提供了运算符，并保持对众多开源社区组件的开放性。此外，AWS 还开源了 SageMaker 的部分实现，联合 Facebook 推出 TorchServe - 业内第一个针对 PyTorch 框架的 model server。

四、部署

机器学习始终因落地部署难而被诟病，如机器学习项目想要“开花结果”，落地部署环节是关键，但机器学习所具有的一些特性，却使得大规模部署十分困难。这个阶段开发者需要处理，编程语言兼容性、算力、可扩展以及模型质量等难题，

而对于完全托管的机器学习服务完全不成问题。

Amazon SageMaker 可实现一键式部署受训模型，以便开发者针对实时或批量数据生成预测，并提供适用于 Kubeflow Pipelines 的组件，保持模型的高效与准确。开发者只需关注应用程序为此终端节点提供的 API 调用，即可实现低延迟、高吞吐量推理。同样是“大宇无限”的案例，使用 Amazon SageMaker 服务后仅用了三个月的时间，“大宇无限”便完成了自身机器学习系统从0到1的突破。

未来 AI 将如何在云端生长

随着人工智能与机器学习的普及，新一轮技术浪潮正在开启。无论是创新企业还是传统企业都意识到需要利用人工智能与机器学习寻求突破和创新，根据 Gartner 在2019年对 CIO 的调查显示，2018年至2019年间，部署人工智能（AI）的企业或组织从4%增长到了14%。

可以说，AWS 已经为机器学习从业人员找到了更高效、更便捷、更迅速的解决方案。AWS 希望将 Amazon SageMaker 打造成通用的产品和服务，通过社区的形式把机器学习的能力交到每一位创建者的手中，真正让机器学习可以广泛地、深入地改变人们的生活。

机器学习项目复杂繁琐，该如何构建安全的机器学习环境？

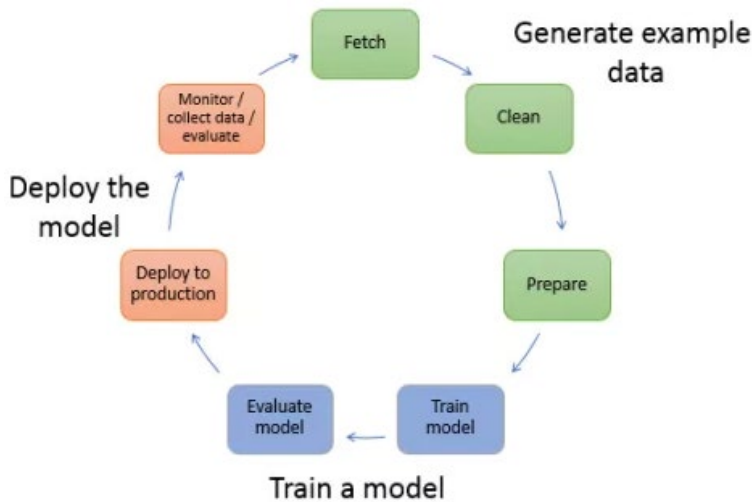
作者 | 冯垚

六一儿童节即将到来，但今年的六一却有些不同。随着疫情的缓解，多地明确了6月1日后，基础教育各学段需陆续开始返校复课，并按照安全要求严格做好人员管控和测温工作。当下疫情防控进入常态化，学校急需即可实现学生身份识别，又可保证学生身体情况正常的解决方案。目前，人脸识别测温一体化解决方案是解决学校疫情联防联控工作的最佳解决方案。

而这类高效、智能化的解决方案，背后需要大数据、人工智能、机器学习等新兴技术的有利支持。这场疫情所带来的“黑天鹅”效应，真正让人们意识到“科技防疫”的现实价值。

机器学习与大数据紧密相连，两者互相依存、互相促进。机器学习之所以在这些年有了长足的发展，备受追捧，其中最重要的一点便是有了大数据的支持。机器学习原本就是基于数据来训练产生价值的技术，当有了更多、更真实的数据，促使机器学习可快速提高模型的精确性。对于机器学习而言，大数据始终是不可或缺的一部分。

数据贯穿机器学习全流程，该如何保护？



机器学习模型的典型工作流程纷繁复杂，主要分为三部分，第一部分为生成事例数据，它又分为获取数据、清理数据和转换数据。第二部分是模型训练，其分为训练和评估两部分。第三部分是部署，包括监控、推理以及偏差识别。

将机器学习“化繁为简”是 Amazon SageMaker 的最终目标。由于机器学习项目整个工作流程极其复杂，外加各个环节的调试和设置，造成了机器学习项目落地难度较大。Amazon SageMaker 旨在帮助企业和开发者快速构建、训练和部署机器学习模型，用户只需提供数据便可以开始训练。

数据自始至终贯穿在机器学习的项目当中，必然成为重要的保护对象。随着语音、图像等非结构性数据识别的普及，机器学习执行用户指令的准确性变得至关重要，这也造成了数据源易被攻击者利用。这类攻击名为——对抗性攻击，攻击者通过对数据源的细微修改，造成机器学习做出错误决定，导致后续发生更多的错误行为。

使用 Amazon SageMaker 可协助用户构建安全的机器学习环境，满足用户的数据储存、传输、训练等多方面的安全防护需求。首先，如果想要为 Amazon SageMaker 打造安全的网络运行环境，就要构建一个安全的 VPC。主要包括以下几点：

第一点：网关设置。进行互联网网关和内网网关的设置，打造属于用户的私有子网。

第二点：打造跨可用区的多子网建设。目的是实现 Amazon SageMaker 高可用性。

第三点：引入 VPC Endpoint 概念。可以避免相关资源只连到互联网，降低更多资源直接暴露在互联网上的风险。

第四点：安全组设置。控制对整个 VPC 或安全组进出的访问。

其次，Amazon SageMaker 针对数据保护，主要分为加密保护静态数据和加密保护动态数据。一方面，Amazon SageMaker 通常使用 KMS 运管服务进行加密静态数据，满足这项加密的存储服务包括 S3、EBS、CodeCommit git repository 等。另一方面，传输中的所有网络间数据都支持 TLS 1.2 加密，对 Amazon SageMaker API 和控制台发出的请求全部在安全 (SSL) 连接中进行，确保机器学习模型项目和其他系统项目在传输中和静态下全部经过加密。

最后，对于 Jupyter Notebook 实例以及处理、训练和托管模型 Docker 容器，AWS 可允许开发者指定 AWS KMS 密钥。在默认状态下 Amazon SageMaker 使用的是 AWS SageMaker Service-Managed Key 对数据进行加密，不过开发者也可使用自己的密钥，制定 CMK 进行加密。

团队协作 人员权限等级需明确

由于机器学习的项目流程复杂、繁琐，对于开发者团队而言，访问权限控制急需明确。通常一个机器学习项目是由一个开发者团队来完成，但目前对于众多企业或团队，还是停留在如何应用机器学习的层面，并没有注意到机器学习使用的安全问题。站在开发者的角度，开发者团队需要明确哪些人员分别对应哪些安全级别，才能保证机器学习项目的安全运行。

例如一些初创型互联网公司，由于内部员工欠缺机器学习的操作经验，创建一台虚拟机后，直接登陆 Jupyter Notebook 便直接开始编写代码和运行代码。Jupyter Notebook 中包含的所有数据并没有实行明确的隔离等级保护，在这种情

况下极易给机器学习项目带来严重的安全隐患。

Amazon SageMaker 在面对此类问题时，明确了人员权限的等级。在 Amazon SageMaker 中即可保证基本权限配置的同时，又可便于内部协作。当创建者开启 Notebook 时，此 Notebook 是默认在他的 Amazon SageMaker VPC 中，如果其他人想访问，则需要创建者进行权限配置。其中 Amazon SageMaker 包含的弹性笔记本 Amazon SageMaker Notebooks 还可一键启用 Jupyter Notebook，便于开发者团队内部的协作办公。

Amazon SageMaker 更重要的创新是将编程和训练设置为松耦状态。由于 Amazon SageMaker 中编程和训练是分开的，特点是为每个用户提供单一容器或单机的支持，开发者便可使用极小的计算资源来开发程序。即达到了资源上的节省，同时还保证了数据方面的安全隔离。

而针对机器学习经验不足的企业或开发者团队，Amazon SageMaker 还给出了更详细的解决方案。主要包括以下四点：

第一点：使用 AWS 的托管策略。

第二点：针对对不同角色授予最低的权限，减少不必要权限的过度开放。

第三点：对相关的敏感操作启用多因子认证。

第四点：构建不同的策略条件，以及 API 的权限设置，增强机器学习安全性。

Amazon 也希望利用 SageMaker 来帮助企业快速将机器学习应用到实际生产中。天津华来科技已经开始使用这项服务，作为一家初创企业，自2015年成立以来，随着营业额的不断扩大，企业开始面临诸多难题。公司的开发团队应用 Amazon SageMaker，将人工智能和业界技术创新融合到产品和服务中，解决了数据处理与海外市场的开拓及业务转型的问题。同时 AWS 所提供的技术与隐私保护认证，避免了企业隐私泄露和潜在的安全风险。

责任共担模型 给机器学习创造更安全的环境

上文提到，大数据与机器学习相伴相生，此外云计算也对机器学习发展起到了决定性作用。云计算既提供大数据所需的存储，又能提供较高的计算能力，复杂的机器学习算法正迫切需要分布式计算这样的关键技术作为支撑。Amazon SageMaker 作为一款基于 AWS 的全面托管服务，可以帮助企业和开发者在云上快速构建、训练和部署模型。

对于企业和开发者而言，网络安全问题不可避免。当使用 Amazon SageMaker 时应用责任共担模型，整个端到端所有环节，包括负责运行、管理和控制从主机操作系统和虚拟层到服务运营所在设施的物理安全性的组件，均由 AWS 来负责，可大幅减轻客户的运营负担。

而面对攻击，需要构建动态的安全保障体系，日志则是体系建立的关键工具。Amazon CloudWatch 可实时监控 Amazon SageMaker，收集原始数据，并将数据处理为便于读取的指标。Amazon CloudWatch Logs 还可监控日志文件中的信息，便于使用者设置指定阈值，当指定指标达时通知使用者或采取措施。

此外，优秀的日志监控工具，需要高效的智能威胁检测服务帮助。Amazon GuardDuty 作为一项威胁检测服务，可持续监控在 AWS 云中发生的恶意活动和未经授权的行为，从而保护使用者的 AWS 账户和工作负载。同时 Amazon GuardDuty 自带机器学习引擎，可以尽快识别潜在的威胁并确定优先级别。可见，日志监控工具加智能威胁检测服务，可大幅缩短用户的响应时间。

放眼全球，Amazon SageMaker 作为商业化最为成熟的机器学习平台服务，在框架、集成性和生态系统的建设上已经非常完善，全球已有数万家企业采用 AWS 来运行机器学习算法。2020年4月30日，Amazon SageMaker 正式进入中国市场，在由光环新网运营的 AWS 中国（北京）区域和由西云数据运营的 AWS 中国（宁夏）区域正式开放。

企业出海服务为何这么难？

作者 | 冯垚

当前，新一轮科技革命席卷全球，新一代信息技术迅速发展，加速推进企业数字化转型。对我国而言，根据中国信通院发布的《中国数字经济发展白皮书（2020年）》显示，中国数字经济贡献不断增强，2019年我国数字经济增加值规模达到35.8万亿元，占 GDP 比重达到36.2%，正在形成数字产业化、产业数字化、数字化治理、数据价值化“四化”的数字经济框架。

随着我国数字经济规模不断扩张，我国拥有现象级的产品、丰富的人才、成熟的技术应用与商业模式，这正是我国互联网企业迈向全球的最佳时机。从国家层面来看，鼓励并提倡企业出海。2013年国家提出“一带一路”，促使更多企业开始走出国门开拓海外市场。2020年新基建迅速成为社会舆论的焦点，其更是将数字经济发展放大，有助于企业突破技术与产业瓶颈，利好企业出海。从企业侧来看，经营诉求的变化，出海成为最佳选择。大致可分为三类：

一、定位国际化的企业：此类企业多为第三代互联网公司，它们基于移动互联网诞生，一方面，资金和技术储备雄厚，易于出海。另一方面，部分企业的初心就是立足国际市场，乐于实践国际化。

二、成熟的企业：产品和技术成熟，但企业整体增速放缓，需要创新动力，希望通过出海保持企业发展动力。

三、初创企业：面对国内互联网下半场红利放缓，一些初创企业则选择转向海外的蓝海市场寻找机会。

“从模仿到创新，从内部扩张到争相出海，中国企业出海欲乘风破浪，数据库服务壁垒需破解！”

企业出海 数据先行

自2020年2月以来，新冠疫情在全球范围内爆发，全球经济尽显疲态，但移动互联网市场却十分火热。具体表现在疫情蔓延促使人们在线寻求娱乐和社交的热情上升。以如今全球范围内最受欢迎的短视频应用为例，根据 Sensor Tower 公布的数据显示，2020上半年中国短视频 / 直播应用，在海外下载量及收入 ToP20 的应用共获得12.7亿次下载，前五大市场分别为印度（42.5%），巴西（8.3%），印尼（6.2%），美国（5.2%）和墨西哥（3.3%），可见印度、美洲及东南亚地区已成为中国企业出海的热土。

目前，游戏、社交与娱乐应用处于利好之下，我国技术型企业出海恰逢时机，而“孤身”出海的中国企业，数据存储及数据安全则是首要问题。如今的数字经济时代，数据价值凸显，企业消费者分布在世界的各个角落，数据的存储和读取需要漂洋过海，这无疑会给出海企业的用户体验带来诸多挑战。同时“孤军奋战”的出海企业，势必面临高昂的基础设施投入，或许难以到达预期效果。而借助专业的第三方服务，则有助于出海企业的成本控制，并大幅提升数据储存与管理效率。

AWS 全球数据库服务赋能企业出海

根据2019年 Gartner 发布的《The Future of the Database Management System (DBMS) Market Is Cloud》报告指出数据库的未来是云上，其中 Amazon Web Services (AWS) 从2016年至2018年始终占据全球云数据库市场份额第一。AWS 作为全球云服务的先行者，可提供丰富、稳定、可靠的全球数据库服务。

AWS 基础设施覆盖全球，可快速帮助出海企业构建分布在全球的 Internet 规模应用程序，实现出海企业在世界各地以同样速度 and 安全性，便捷的使用相同资源。企业无需担心峰值等难题，AWS 数据库可自动适应峰值工作负载，从而实现最优的成本节省。AWS 可提供关系型数据库、非关系型数据库与数据仓库三种全

球性数据库服务。

1、关系数据库：Aurora Global Database (Amazon Aurora 全球数据库)

对于出海企业而言，业务的平稳运营是快速获客的必要条件，需要第三方服务提供可靠的数据灾备体系。Aurora Global Database 针对全球分布式应用程序而设计，可跨越多区域实现数据服务的灾难恢复，快速实现指向次区域的全局故障转移，帮助出海企业实现数据灾备。

其灾难恢复类解决方案，主要表现在以下两项指标：

- RTO（恢复时间目标）——即灾难发生后，需要多长时间才能恢复正常工作状态。
- RPO（恢复点目标）——灾难事件可能破坏的数据量。

Amazon Aurora 全球数据库以更优的恢复点目标（RPO）与恢复时间目标（RTO）处理区域整体故障，可以实现将 RPO 控制在5秒以内，最大程度保证数据的安全，RTO 保持在1分钟以内，缩短系统的停机时间，维持企业业务的正常运行。

如企业的主区域出现性能下降或中断，Amazon Aurora 全球数据库可保证在此区域内仍可继续运行，其会快速响应并将次区域提升为主区域，借助全局存储复制功能，新主区域可在1分钟之内执行读取 / 写入功能，从而大幅降低故障对应用程序正常运行的影响。

此外，海外市场的高延迟是各家企业都无法逃避的现实。就上文提到的娱乐和社交等应用为例，“内”与“外”用户体验尽可能保持一致始终是最大的挑战，无论是游戏的延迟，还是视频的加载延迟、卡顿等等都直接影响着用户体验。

Amazon Aurora 全球数据库除提供灾备功能之外，允许企业将指向主区域的数据读取操作快速转移至次区域，保证以本地级别延迟将内容交付，通常跨地区复制延迟小于1秒。其可在最多5个远程区域中建立 Amazon Aurora 集群的读取副本，访问位置更接近这些本地区域便为用户带来低延迟的访问，进而实现更佳体

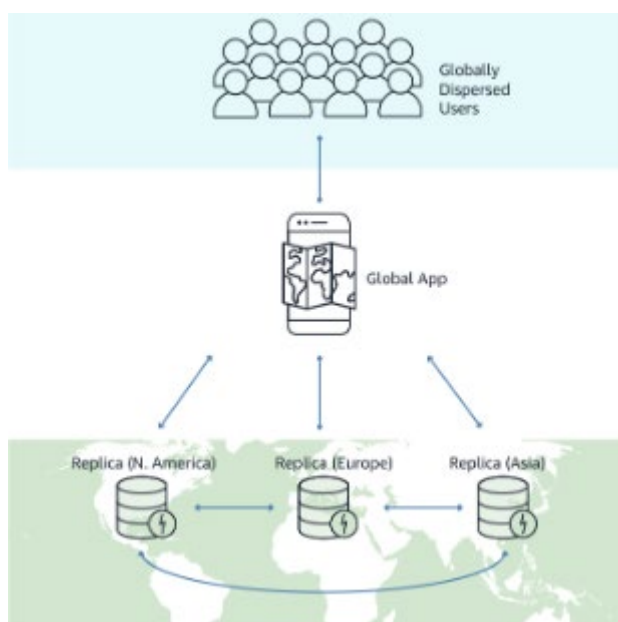
验与参与度。

除了为各区域用户提供更低的读取延迟，Amazon Aurora 全球数据库特有的写入转发功能，可让全局应用程序以更少的代码变更在远程区域内轻松执行写入操作，简化应用程序的开发流程，提高开发效率。首先，应用程序先将写入操作发送至本地读取集群，再由该集群以透明方式把写入转发至主区域，即可实现将写入操作发送至任意 Aurora 全球数据库远程集群。Amazon Aurora 兼容 MySQL 和 PostgreSQL 开源数据库引擎。

例如，虎牙直播面向海外市场的游戏直播平台 Nimo TV，虎牙直播利用 Amazon Aurora 全球数据库服务，在 AWS 亚太（新加坡）区域部署数据库，在其它区域建立副本，提高当地用户的使用体验。

2、Amazon DynamoDB Global Tables

Amazon DynamoDB 是一种键 / 值和文档的 NoSQL 数据库，可提供个位数的毫秒级性能，便于企业储存动态信息。Amazon DynamoDB Global Tables 基于 DynamoDB 的全球覆盖范围构建，可为出海企业提供一个多区域、多主控、完全托管的解决方案。



该数据库在企业的 AWS 区域中自动复制 Amazon DynamoDB 表，并采用多主复制确保所有地区的数据保持一致，实现全局分布式应用程序的本地读取 / 写入和全局数据访问。解决出海企业在区域之间复制和更新数据的难题，帮助企业将重心放在应用程序的业务逻辑等核心技术领域。

Global Tables 简化了复杂的 DynamoDB 部署和管理全局可用表，企业只需在 AWS 管理控制台中选择需要复制数据的地区，数据库便会自动执行，保证应用程序通过现有的 DynamoDB API 和终端节点访问全局表。

目前 Amazon DynamoDB Global tables 已经在中国区上线，但 Global Tables 只能用于北京和宁夏之间或 global 区域之间的 DynamoDB 表同步，如果企业需要将中国区和 global 区域的 DynamoDB 表做双活复制同步，可通过 lambda、DynamoDB stream、Kinesis Stream 等托管服务实现。但需要企业自行构建解决方案，贴心的 AWS 也为此类企业提供了几项重要提示，如下：

- 如何捕获源端变化并持续复制到目标端。
- 如何避免循环复制。
- 如何处理同一个 item 的变更冲突。
- 如何处理网络延迟。

3、Amazon ElastiCache for Redis – Global Datastore

除 Amazon Aurora 全球数据库及 Amazon DynamoDB Global Table 之外，AWS 还提供 Amazon ElastiCache for Redis – Global Datastore 服务，其基于 Amazon ElastiCache for Redis。自 Amazon ElastiCache for Redis 推出以来拥有的低延迟流量特性，使其深受全球客户青睐，但随着企业服务需求的增加，需要更加灵活的云基础设施。为此，AWS 研发了可在区域之间复制 Amazon ElastiCache 集群的功能，正式推出了 Amazon ElastiCache for Redis – Global Datastore。

借助 Global Datastore，企业可在一个区域中写入 ElastiCache for Redis 集群，并使数据可从其他两个跨区域副本集群中读取，缩短缓存数据与用户之间的地

理距离，从而降低跨区域的网络延迟，提升数据的灾难恢复能力。全局数据存储设置相对简单、易操作，但需要提醒的一点是，主集群可以是新建集群，也可以是现有集群，其务必运行 Redis 5.06 或者更高版本。

数据的安全与合规

由于海外巨大的市场需求、活跃的用户群体与开放的消费习惯让中国企业出海得以乘风破浪，但与此同时数字经济消费潮也带来了全球性数据泄露事件的激增。根据2020年5月 Verizon 发布的2020年度《数据泄露调查报告》显示，在调查的15个行业中，28%的泄露涉及小型企业，72%的泄露事件涉及大型企业，其中58%的受害者遭到个人信息泄露。庆幸的是，81%的泄露在数天内得到了妥善处理。

随着人们对隐私数据保护意识的不断加强，众多国家及组织相继开始施行严格的安全法律法规，例如：《通用数据保护条例》（欧盟：GDPR）、《加利福尼亚州消费者隐私法案》（美国：CCPA）等，中国企业出海需要充分考虑到全球范围内隐私保护及合规方面的重要性。

但企业面对数据的一“紧”一“松”，确实左右为难。“紧”的一方面，数据助力企业出海，企业需要下大力气保证数据安全，限制数据的“活动范围”。“松”的一方面，企业需要数据为用户提供个性化的服务，出海企业更是需要数据在不同区域内的频繁传输和分析。作为业界的领导服务商，AWS 的云安全服务可为出海企业提供全面的数据合规和数据安全保护服务：

- 网络隔离

企业可在 AWS 提供的 Amazon Virtual Private Cloud (Amazon VPC) 服务中构建网络隔离区域运行各种资源。使用 Amazon VPC，可帮助企业将数据库隔离在企业自己的虚拟网络中，企业可使用 VPN 与本地 IT 基础设施进行连接。

- 资源级权限

AWS 全球数据库服务支持与 AWS Identity and Access Management (IAM) 集成，允许企业控制 AWS IAM 用户和组对特定资源（例如，数据库实例、数据库快照、数据库参数组、数据库事件订阅、数据库选项组）执行操作。此外，还可

配置 IAM 规则以确保特定访问者仅拥有只读访问权限，而只有管理员可创建、修改和删除资源。

- 加密

企业的所有数据完全处于静态加密之中，通过静态加密，企业可以构建符合加密合规性和相关法律法规要求的安全敏感型应用程序。同时，静态加密与 AWS KMS 集成创建和控制的密钥加密企业的数据库，有效降低出海企业在保护敏感数据时的复杂程度。

- 合规性

AWS 服务可支持包括 SOC 1、SOC 2、SOC 3、ISO、MTCS、C5、PCI、HIPAA 和 FedRAMP 等合规性计划。

不懂业务的开发人员，就不是好的数据科学家

作者 | 冯垚

当下，数字化转型浪潮席卷各行各业。就以刚过不久的七夕节为例，在这一天，看电影、聚餐、唱歌等等都成为了大家不可或缺的娱乐项目，由于线下等位人数过多，在线预约取号软件得到了广泛应用。而就在当天，某知名火锅店的排号系统却意外罢工，更是直接登上了微博热搜榜。

据了解，数据激增是导致系统崩溃的直接原因。不少“热心”网友在其官方微博留言：“什么鬼，软件崩了”、“赶紧修系统去！！”等等。好在，该火锅店迅速修复了系统 Bug，恢复了正常营业。可见，如何快速地、稳定地交付高质量应用，满足用户的多样化需求，已经成为企业必须要面对的关键问题，其直接影响到企业的利润及发展潜力。

其中，软件开发过程的改进是关键，正是 DevOps 开启了软件开发的新革命。它将开发、运营和业务等部门之间进行联通，实现了软件开发人员和 IT 运维技术人员有效沟通的工作方式。促使软件的构建、测试、发布更加迅速、可靠。

同样，在人工智能（AI）和机器学习（ML）领域，需要类似的改进方法进一步提高模型开发、训练和部署的效率，帮助企业实现机器学习的快速落地。

一、机器学习 / 人工智能如何轻松开发和运维？

Gartner 2020 年4月的一项调查显示，全球14%的 CIO 已经部署了人工智能，48%的 CIO 将在2019年或2020年部署人工智能。人工智能及机器学习的部署上升趋势表明，大多数企业认识到机器学习的变革优势，并正在从机器学习 / 人工智能中获得价值，但在实际生产中应用机器学习，企业仍然面临诸多挑战。

一方面，机器学习应用本身在开发工具（如 IDE，软件包 / 库和统计技术等）领域已经取得了长足的进步，但在部署和维护环节的关注度极低。机器学习项目同样遵循软件开发的流程——创建、训练、部署，再根据需求更迭版本。但机器学习系统与传统软件有着本质的区别，它由数据驱动，使得其不确定性很强。由于部署新模型的过程需要数周或数月，同时模型又时常变化，操作不当极易导致新模型无法投入生产。

另一方面，机器学习项目落地涉及企业众多环节，各个团队之间的协作是关键。具体到公司内部，基本涉及三类职责的人群：产品经理、应用开发、算法工程师 / 数据科学家。当算法模型创建完成后，应用开发会调用相应端口，而后续则需要建立机器学习平台或者运维团队的支持。对于各部门的团队来说，机器学习平台及其运维只是整个机器学习项目中的一部分，需要数据科学家、IT 等多部门能够通过监视和验证的手段管理机器学习。例如：数据科学家通常参与机器学习模型 / 算法的开发，但他们不负责生产流程和系统，所以实现协作和整合显得尤为重要。

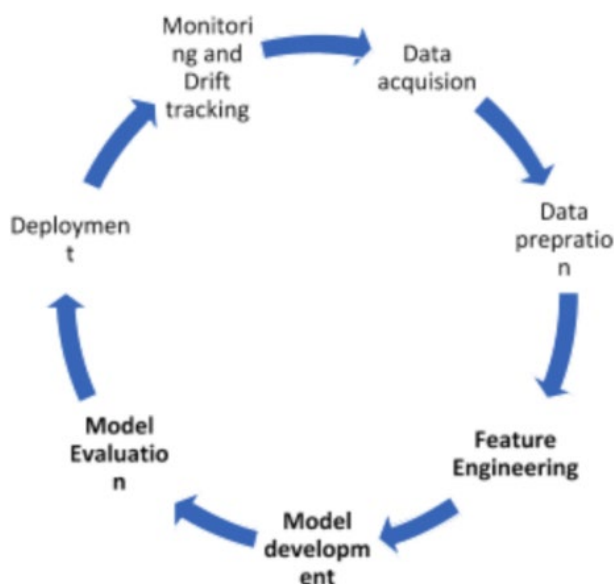
此外，机器学习架构设计方面也存在缺陷。目前诸多机器学习模型代码与应用代码为紧偶关系，当模型需要快速地更新、迭代、AB 测试时，机器学习团队依赖应用开发团队，使得平台运维团队、应用开发团队和机器学习团队相互抑制，直接影响到整个项目的进度。可见机器学习工作流程缺乏与通用软件工程相似的标准化流程，急需将数据科学、数据工程以及 DevOps 专业知识结合在一起。

由此，MLOps 应运而生，MLOps 基于提高 workflow 效率的 DevOps 原理和做法，将持续集成、持续交付和持续部署的理念应用到机器学习过程当中。其聚焦于机器学习新模型的持续训练、新模型的部署以及之后的技术工作，旨在保证机器学习项目快速、持续地试验和开发模型、快速地将模型部署到生产环境、减少模型衰减带来的影响。

总的来看，随着机器学习技术和实践与现有的软件交付系统和流程结合起来，MLOps 将获得迅速发展。不过，必须明确，MLOps 所具备的优势并不等于它是万能的，虽然 MLOps 改善了机器学习解决方案的质量和一致性，但整个方案相对复杂，企业自身无法独自实现，因此需要一个平台来处理这些频繁的事件。

二、Amazon SageMaker 赋能企业应用 AI/ML

为此，Amazon Web Services (AWS) 作为人工智能与机器学习先驱者之一，推出了完全托管端到端 MLOps 服务的 Amazon SageMaker 平台。Amazon SageMaker 可帮助开发人员和数据科学家快速地大规模构建、训练和部署机器学习 (ML) 模型，消除了机器学习过程中各个步骤的繁重工作，让开发高质量模型变得更加轻松。



在传统的机器学习项目中，如图所示，数据科学家会需要参与每个步骤，将不断更新的新模型投入到实际生产中。如部署周期是每月一次或每季度一次，传统流程还可满足需求。但如今，企业需要新功能、高可操作性的机器学习功能，传统的机器学习流程逐渐成为机器学习落地、规模化、生产化的阻力。

Amazon SageMaker 覆盖整个机器学习项目全阶段，将机器学习算法的开发过程、业务流程与部署过程进行了松耦，以全托管、自动化的方式提供监视、验证和管理机器学习模型等服务，加速企业建设 MLOps 的过程。

构建机器学习模型

Amazon SageMaker Autopilot 可帮助用户自动构建、训练和调优完全可视和可控的模型，即可自动检查原始数据、应用功能处理器、选择最佳算法集、训练和

调优多个模型、跟踪模型性能以及根据性能对模型进行排名。

训练机器学习模型

Amazon SageMaker Experiments 通过自动捕获输入参数、配置和结果并将其存储为案例，帮助用户组织和跟踪机器学习模型的迭代。此外用户还可使用 SageMaker Studio，以可视化的方式来浏览进行中的机器学习实验，根据实验特征搜索先前的实验、查看先前的实验及结果，以及直观地比较实验结果。

部署机器学习模型

用户可依靠 Amazon SageMaker 模型监控器来检测和修复概念偏差，让模型保持精确。当模型在生产环境中运行时，一些功能可能会出现偏差，这表明需要对模型进行再训练，模型监控器会自动检测已部署模型中的概念偏差，并提供详细的警报，帮助用户确定问题的根源。

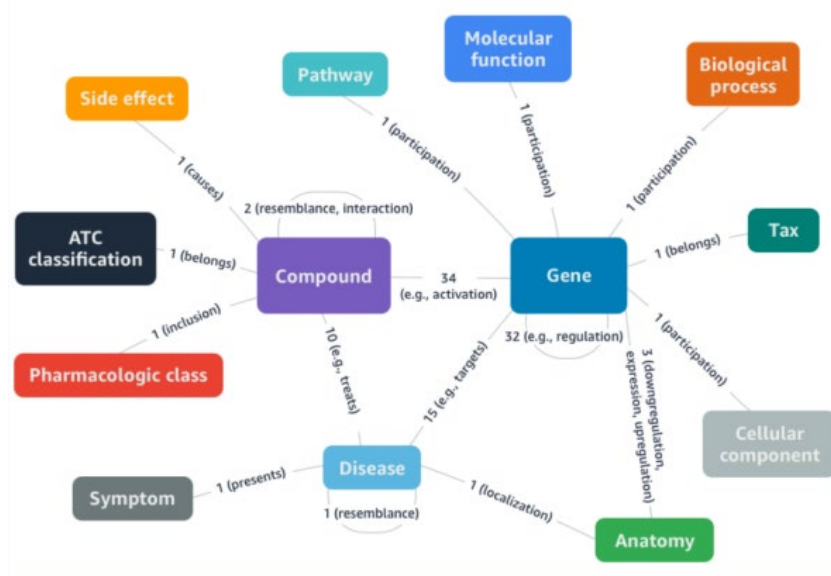
与此同时，AWS 针对中国市场的特点，为开发者提供十分开放的使用环境，帮助用户在 AI 时代实现智能转型。

开放性

AWS 不断积极参与众多开源社区项目。Amazon SageMaker 对包括 TensorFlow、PyTorch、Apache MXNet 在内的主流框架进行了深度优化。例如：在全球范围内使用 TensorFlow 框架的机器学习项目约85%的负载跑在 AWS 平台上，其针对 TensorFlow 进行优化后能够实现同时跨百余个 GPU 提供近线性扩展效率，为开发者解决了在云端运行时的大量处理开销。根据实际案例，256 GPU 扩展效率可达到90%，训练时间从30分钟缩短到14分钟，开发者可在更短时间内训练更准确、更专业的模型。



此外，在今年4月 AWS 和 Facebook 合作推出了 PyTorch 模型服务库 TorchServe。借助 TorchServe，PyTorch 用户可以更快地将其模型应用于生产，而无需编写自定义代码：除了提供低延迟预测 API 之外，TorchServe 还为一些最常见的应用程序嵌入了默认处理程序，例如目标检测和文本分类。目前 Amazon SageMaker 支持的深度学习框架包括：TensorFlow、PyTorch、Apache MXNet、Chainer、Keras、Gluon、Horovod、Scikit-learn、Deep Graph Library、Deep Java Library 等。



值得关注的是，AWS 在重视技术创新发展的同时，一样关注当下存在的社会问题。在疫情期间，上海 AWS 人工智能实验室发布了基于深度学习的有关制药领域的知识图谱计算库 Deep Graph Library。通过调用该库，能帮助人们更快、更有效地从现有药物中查找针对新冠病毒有潜在治疗效果的药物。

数据湖火了，那数据仓库怎么办？

作者 | 冯垚

未来，人类将面临着三大问题：

生物本身就是算法，生命是不断处理数据的过程；

意识与智能的分离；

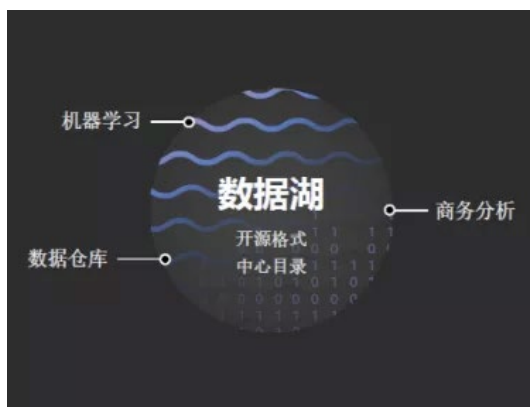
拥有大数据积累的外部环境将比我们自己更了解自己；

这是《未来简史》中提出的三个革命性观点。一本书短短百页，让我们看到了世界颠覆性的变化，从计算机，到互联网，再到大数据、人工智能，所有的变化都在以一种肉眼可观却又无法捕捉的状态悄然发生着，而推动变化发生的背后，则是数据价值的提升。

如果把数据比喻成“油田”，要想充分挖掘其价值，首先需要将数据进行“开采 / 储存”即数据的收集储存，其次进行“精炼”即数据的挖掘和分析，最终实现数据创造更多价值。以如今常见的电商为例：电商企业通过收集用户的相关数据，再利用数据分析技术，对用户的偏好进行分析，而后进行相关商品推荐，从而提高用户的购买效率；此外电商企业还可建立预测模型，针对特定人群进行预测，在不同阶段适时调整销售方式，提高用户对产品的满意度，从而提高销售额。

原先，企业通常依靠昂贵且私有的本地数据仓库解决方案来存储和分析数据，由于模型范式的要求，底层数据无法做到多样变化，导致企业业务不能随意变迁。同时随着互联网 / 移动互联网的爆发，数据量从 TB 到 PB 到 EB 级，数据类型更是涵盖结构化数据、非结构化数据、半结构化数据，并且用户对地域性、及时性的要求愈发苛刻，使得传统的数据仓库方案需要更新。

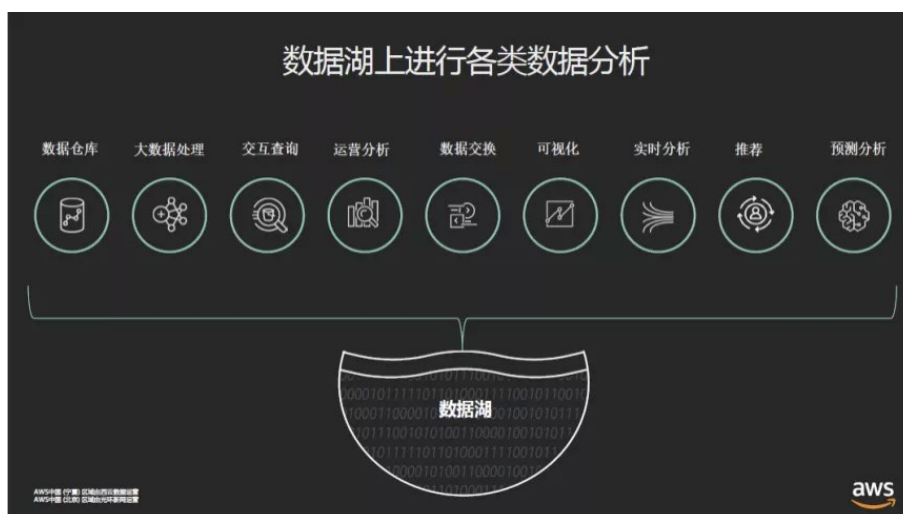
如今，有了云端，高弹性和可扩展的计算与存储，使得数据的储存与分析更容易解决，可以说云端数据解决方案已经成为大势所趋。一方面，分布式架构与开源体系，可以适应当下快速的数据变化。另一方面可集成更多新技术服务，比如与机器学习结合，实现更多预测性分析。而分布式储存、多种文件格式、多种引擎和元数据服务，这也逐渐形成了数据湖的基础。



1 AWS 数据湖的技术革新之路

数据湖的概念最早在2006年提出，其主要概念是将数据湖定义为一个中心数据存储的容器。数据可以很轻松进入数据湖，它可以存储结构化、非结构化和半结构化的数据，并支持数据量的快速缩放，灵活地适应上层数据应用的变化，最终实现海量数据的存储和查询分析。

而真正将数据湖概念推而广之的便是 AWS (Amazon Web Services)。AWS 很早便开始推动有关数据湖的技术演进，2009年 AWS 推出了 Amazon Elastic MapReduce (EMR) 数据湖架构，以跨 EC2 实例集群自动配置 HDFS；2012年又继续推出了云端 MPP 架构的数据仓库云服务 Amazon Redshift；随后 AWS 逐渐将数据湖核心转向 Amazon S3。



随着大数据技术的发展，计算能力成为关键，实现计算和存储的分离，所带来的弹性扩展和成本优势逐渐显现。而云服务天生具有存储和计算分离的特性，AWS 的云优势慢慢凸显。最终 AWS 数据湖将大数据和云计算结合在一起，形成了一个存储和多个引擎 / 服务的经典数据湖搭配。这里，我们将结合 AWS 整体的分析服务来向开发者们解释，AWS 是如何帮助开发者 / 企业构建数据湖环境，进而高效使用数据的。

• 快捷的数据查询引擎

在 AWS 上，Amazon S3 对象存储服务由于其高可用性、高持久性、可扩展性和数据格式兼容性等特点，成为了建设数据湖的首选。而 AWS 还提供了交互式查询方式可以直接查询 S3 中的数据，Amazon Athena 便是一种交互式查询服务。

它可以使用标准 SQL 分析 Amazon S3 中的数据，Athena 简单易用，只需指向开发者存储在 S3 中的数据，定义架构即可开始查询，它无需执行复杂的 ETL 作业来为数据分析做准备，开发者可以轻松实现分析大规模数据集。

• 如何解决元数据格式多样的问题？

由于数据湖可以按任何格式存储，因此无需将其转换为预先定义的数据结构，

使用数据湖的主要挑战之一便是查找数据并了解数据结构和格式。AWS Glue 则可帮助开发者抽取、转换和加载数据，并可在不同数据存储之间可靠地移动数据。此外 Glue 作为一项完全托管服务，它会像“爬虫”一样对数据湖里的海量数据自动爬取，自动生成数据目录，该数据目录是所有数据资产的永久元数据存储。存入目录后，数据便可立即供 ETL 搜索、查询和使用。

值得一提的是，Athena 可与 AWS Glue 数据目录进行集成，实现开箱即用，帮助开发者能够跨各种服务创建统一的元数据存储库、抓取数据源以发现架构，并使用新的和修改后的表与分区定义填充数据目录，以及维护架构版本控制。

- 如何快速构建数据湖？

不难看出，数据湖是一个高效、快速的数据存储 / 分析理念，但同时它还具有相当高的复杂度。在设置和管理数据湖时，涉及大量极为耗时的复杂手动任务，包括加载不同来源的数据、监控数据流、设置分区、打开加密和管理密钥、定义转换作业并监控其操作、将数据重新组织成列格式等。

面对解决此类难题，开发者可使用 AWS Lake Formation 服务，它简化了数据湖的创建和管理工作，缩短了数据湖的构建时间，可在几天内实现建立安全的数据湖。而且可喜的是，就在2020年10月25日，AWS Lake Formation 已在由光环新网运营的 AWS 中国（北京）区域正式推出。

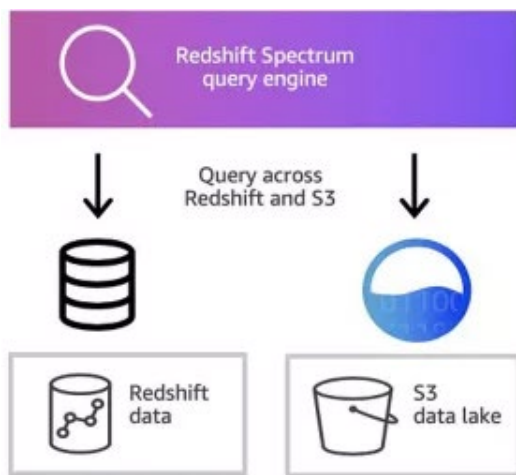
Lake Formation 建立在 AWS Glue 中可用的功能之上。开发者只需手动定义数据源，制定要应用的数据访问和安全策略。Lake Formation 会自动帮助开发者从数据库和对象存储中收集并按目录分类数据，再将数据移动到新的 Amazon S3 数据湖。最终，用户可通过选择不同的分析和机器学习服务，利用这些数据集实现多样化服务。

2 湖仓新模式：数据湖 + 数据仓库 = Lake House

综上所述，大数据的时代，开源技术体系的设计确实让云端产品或开源组件构成大数据整体解决方案逐渐兴起，比如数据湖，但并不代表着数据仓库会被淘汰，双方存在必要的联系。一方面，通过上云的方式，持续增强数据仓库的核心能力，

将数据仓库实现现代化。另一方面，数据仓库和数据湖，是大数据架构的两种设计方式，两者功能可以相互补充，这意味着双方需要实现交互和共享数据。

为了实现湖仓的交互，在2019年 AWS re:Invent 大会上，AWS 提出了运行数据仓库和数据湖业务正在形成一种新的模式，即“Lake House”。AWS Lake House 中遵循“ELT”范式（提取，加载，转换），当从本地数据仓库迁移到 Redshift 时，开发者可使用已有的针对 ELT 优化的 SQL 工作负载，无需从头开始将关系和复杂的 SQL 工作负载重写为新的计算框架。



Amazon Redshift 和数据湖之间的无缝互操作性

AWS Lake House 模型中 Redshift 作为首选的转换引擎，实现了高效地加载、转换和扩充数据。Amazon Redshift Spectrum 是 Amazon Redshift 的一项功能，（提示：避免到 console 中搜索 spectrum）AWS 选择开发者熟悉的 SQL 语言，也旨在帮助更多开发者轻松实现查询数据。

不仅如此，新的 Redshift 还具有数据湖导出功能。该功能可将数据写回到数据湖中，目前支持 Apache Parquet、ORC、JSON 和 CSV 格式，以 Parquet 格式为例（一种用于分析的高效开放式列式存储格式），与传统文本格式相比，Parquet 格式的卸载速度最多快2倍，而在 S3 中占用的存储空间最多减少6倍。

此外，Redshift 的 RA3 实例类型允许开发者独立扩展 Redshift 数据存储和计算需求，帮助开发者以较低的价格来管理数据和工作负载的组合。并且

Redshift 可通过自动扩展额外的瞬态容量来处理并发查询并保持一致的性能，从而完成处理工作量高峰。

当数据在数据湖和 Redshift 之间开始顺畅移动，这种灵活性使开发者在存储数据时可以在成本和性能之间选择最佳的折中方案。当前已经有大量的企业和机构都开始采用 AWS 的数据湖和数据分析云服务。其中 FOX Corporation (FOX 公司) 作为世界娱乐行业的巨头之一，每天需要面对大规模的提取、优化、转换和聚合多方来源的事务型事件，数据量达到十亿量级。Amazon Redshift 支撑了其数据仓库和数据湖中查询实时数据，见证了数据 PB 级的快速增长。同时帮助 FOX 公司在保持成本不变的情况下，工作负载提升了10倍。

在中国区域内，欣和作为一家大型食品生产企业，在多品牌、全方位的业务发展规划下，对大规模数据分析和处理提出了更高要求。而原先传统数据仓库系统和人才短缺又限制了欣和的进一步发展，为此欣和选择使用 AWS 云平台来搭建数据湖。AWS 数据湖的高稳定性和安全性，实现了欣和数据仓库的高可用和高可扩展，使欣和各业务系统间的底层数据相连通，并通过调用、分析，为企业业务发展提供强有力的支撑，帮助欣和真正实现数字化。

总结来看，选用 AWS Lake House 可帮助开发者实现以下目标：

- 高效、低成本的数据存储
- 独立可扩展的计算能力，能够进行大规模并行处理
- 标准 SQL 转换
- 并发扩展灵活地执行 SQL 查询

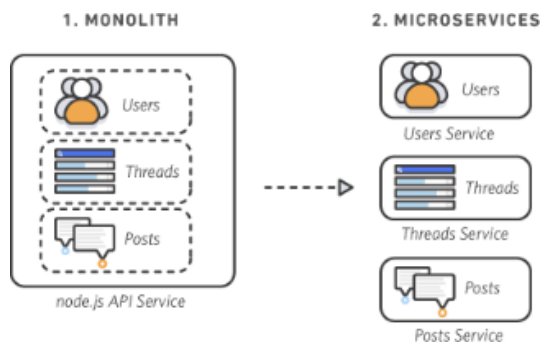
可见，随着产品集、架构模式的继续发展，数据湖与数据仓库的协同运行将会更加频繁，AWS 基于 Redshift Spectrum 提出的 Lake House 也将会在 AWS 数据湖架构中继续发挥关键作用。同时随着 Amazon Redshift 的更多服务在中国区域推出，AWS 更是希望吸引更多中国的大数据开发者，来了解 AWS 数据湖的解决方案，了解 AWS 。未来，AWS 坚信与传统的数据仓库和数据分析解决方案相比，湖仓新模式等云端方案将会为用户释放更大的数据价值。

AWS打造容器服务体系，Amazon EKS这块拼图有多重要？

作者 | 冯垚

近年来，随着云计算、大数据、物联网、人工智能等技术的驱动，企业数字化转型正在全球范围内普及。转型的浪潮正在重塑企业的业务战略，根据 IDC 2019 年数据显示，未来数年内企业或组织将会有超过50%的信息技术支出用于数字化转型，外加上2020年新冠疫情的影响，数字化转型更将成为企业或组织的必经之路。

对于企业数字化转型而言，应用就是生产力。原先如何落地实施数字化转型始终考验着企业或组织，而如今随着数字业务灵活性的加速，新架构、集成技术和模式已经从前沿转变为主流应用。对于开发者而言，由于企业或组织对于灵活、低成本、弹性伸缩的架构越发倚重，单体应用在面对新需求时已经显得力不从心。



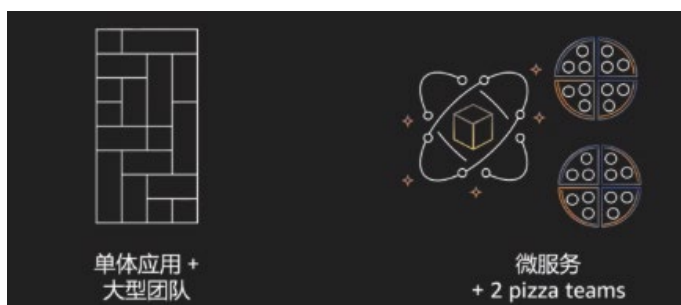
正因如此，微服务作为一种开发的基础架构和组织方法，如今备受追捧。微服务架构具有敏捷、灵活扩展、轻松部署、技术自由、可重复使用的代码、弹性的优势。其实，微服务架构的概念本身并不是新事物，但为何始终没能真正流行？归根结底，是由于传统的本地化部署使微服务落地变得较为复杂。随着云计算的风靡，真正“解放了”微服务，更好的扩展性、伸缩性与可靠性让微服务的普及成为了必

然。

而提到云计算，Amazon 当然是这个时代的开创者。众所周知2006年 AWS 正式上线了 EC2、S3 等核心产品，预示着云计算时代的开始。但鲜为人知的是，早在2001年 AWS 就已经开始了现代化应用架构的实践。当时 Amazon 采用了一套成熟的架构“单体应用程序 +Oracle 数据库”，但随着电商平台的繁荣，产品从上市到发布的时间成为了企业发展迫在眉睫的瓶颈。

AWS 的微服务实践

Amazon 提出了经典的 Two pizza teams 概念。Two pizza teams 含义为两块披萨饭量的小组，代表着这个团队足够小，同时单个团队仅负责相应的微服务，有充分的权限来进行相应产品的开发，并通过解耦来提升产品开发的敏捷性。而不同团队间利用 DevOps 技术来构建一系列小型且灵活的应用团队，进一步加快团队对于产品的开发过程和节奏。



Two pizza teams 将所有应用程序的 API 保持一致，并且拥有不同服务的所有权，这也就实现了面向服务的 SOA 架构，而它的本质同样是微服务。随着 Amazon 电商平台逐渐转向微服务，AWS 将不同的代码库重构为小型的、集中的、单用途的服务，实现了单一的代码库重构为面向服务的体系架构。如今，在 Amazon 官网上已有上万个微服务在支撑整个电商平台的运行，并且每年的部署次数超过六千万次。

可见 AWS 率先实践微服务，让其奠定了当今的电商与 IT 领域的行业地位。而微服务之所以能成为当下的主流架构，AWS 认为，一方面，在云时代微服务满足了开发者对于应用程序特定功能的开发需求，成为了现代化应用的最佳选择；另

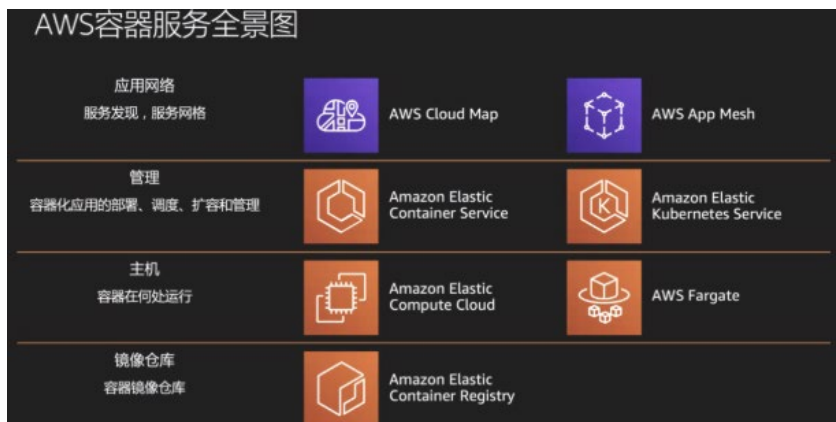
一方面，容器为微服务提供了最理想的实现方式，帮助开发者实现了敏捷开发、持续集成与持续发布。

AWS 建设容器服务体系

容器在应用交付的过程中有着先天的优势。自2016年后随着 Docker 热潮的逐渐冷却，容器技术逐渐进入下一阶段：如何管理大规模的容器化部署、编排和管理。而如今，当初 Swarm、Kubernetes 以及 Mesos 三分天下的场景早已不在，Kubernetes 遥遥领先坐稳了霸主地位。根据 sysdig 发布的2019年容器使用报告显示，已有77%的用户正在使用 Kubernetes 技术。

Kubernetes 的大获全胜，使得 Kubernetes 的实际应用不断增多，其存在的 IP 分配、安全合规、使用场景等问题也层出不穷，自此各大云厂商将精力投入到 Kubernetes 平台的研发当中，并不断完善。与此同时，由于 Kubernetes 既可在本地运行，也可在云平台运行，如何在两种生产环境中来构建 Kubernetes 的应用，如何利用生产环境中的特性来帮助企业进行数字化转型，也给各大云厂商提出了不小的挑战。

Amazon 作为领先的云厂商，积极拥抱容器化，围绕容器服务打造产品体系。2015年，AWS 率先推出了 Amazon ECS 全托管容器调度管理服务，其为 AWS 原生 Docker 容器解决方案。随后，为了进一步支持开源的 Kubernetes 的发展，相继推出了 Amazon EKS 与 AWS Fargate。AWS 的容器产品涵盖从底层容器镜像仓库到上层应用网络，支持容器化的全生命周期。



众多的容器产品组合，可以为企业或组织量身定做在云端的容器服务和容器管理。首先最底层 AWS 提供 Amazon ECR 容器镜像仓库的服务；其次，容器可以运行在两种位置上，其一在 Amazon EC2 虚拟机上构建容器服务，其二选择 AWS Fargate 构建无服务器的容器服务；接着，在容器的管理层，AWS 提供 Amazon ECS 和 Amazon EKS 两种方式。最后，在最上层 AWS 支持不同的服务网格和服务发现，包括 Amazon Cloud Map 与 AWS App Mesh。

EKS 正在成为体系核心

虽然 AWS 基于开源 Kubernetes 的容器服务起步较晚，但其发展速度却是最快的，各国开发者对于 AWS 的容器服务认可度也是越来越高。根据 Nucleus Research 在2019年进行的调查显示，82%的云端 Kubernetes 工作负载运行在 AWS 云服务之上。

AWS 顺势而为，2020年3月13日，将 Amazon EKS 在宁夏西部云数据技术有限公司运营的 AWS 中国（宁夏）区域和北京光环新网科技股份有限公司运营的 AWS 中国（北京）区域正式商用，更好的为中国开发者提供 Kubernetes 平台和应用程序的容器化的服务。

Amazon EKS 从结构上来看，EKS 平台不仅仅是一项完全托管的 Kubernetes 服务，更像是 AWS 容器产品的集成“者”。

第一，EKS 集群可以选择在 Amazon EC2 虚拟机上构建容器服务，或者使用 AWS Fargate 构建无服务器的容器服务。使用 AWS Fargate 时，开发者无需预置和管理服务器，可以保证每个应用程序按需付费，并通过设计隔离应用程序提高安全性。

第二，EKS 与诸如 Amazon CloudWatch、Auto Scaling 组、AWS Identity and Access Management (IAM) 和 Amazon Virtual Private Cloud (VPC) 之类的服务深度集成，为开发者带来监控、扩展和负载平衡应用程序的无缝体验。

第三，EKS 与 AWS App Mesh 集成，并提供 Kubernetes 原生体验，以使用服务网格功能并为应用程序带来丰富的可观察性、流量控制和安全功能。

Amazon EKS 从设计理念上来看，Amazon EKS 区别于其他 Kubernetes 服务，从设计之初便定位于企业级、完全托管的 Kubernetes 的解决方案，旨在帮助开发者在云端轻松运行 Kubernetes。此外，EKS 提供 Kubernetes 原生的体验，始终保持和社区同步更新，并且积极将改进意见反馈给 Kubernetes 开源社区。

更多的，开发者使用 Amazon EKS 框架时，还将获得：

- **更全面的安全服务：**Amazon EKS 集成了 AWS 的安全服务，实施 AWS 安全责任共担模式，并结合多项技术保证 EKS 服务的安全运行。例如：IAM 进行身份和验证的管理、CloudTrail 进行 Kubernetes 日志的管理、ACM 进行证书的管理、KMS 进行密钥的管理等。
- **更简便的管理方式：**Amazon EKS 可以帮助 Kubernetes 集群进行节点预配置和全生命周期管理。EKS 负责来进行 Amazon EC2 节点的创建和管理，同时构建起自动伸缩的节点组 Auto Scaling Group，并且同时执行健康检查，减轻开发者在后端的管理压力。
- **更高可用性：**Amazon EKS 提供托管的控制平面，控制平面包含至少两个 API 服务器的节点和跨三个可用区运行的三个 etcd 的节点，保证 EKS Master Node 的高可用构建。同时 Amazon EKS 会自动检测并且替换掉状况不佳的控制平面的实例，以消除单点故障。

Amazon EKS 可以较好的满足开发者对于现代化应用构建的需求。然而，不同于传统的生产资源的集群管理工具，Kubernetes 项目其最大的价值，乃在于它的 API 兼容能力和以此为基础“控制器”模式。在这个体系的指导下，使得 Amazon EKS 可以满足丰富的支持工具。

- **机器学习工具：**Kubeflow 是一套 Kubernetes 的机器学习工具，Amazon EKS 可与 Kubeflow 进行整合，来构建深度学习平台。需要强调的是，在 Amazon SageMaker 中同样也提供一系列 Kubernetes 的 Operator 相关工具，可方便开发者将 EKS 和 Kubeflow、SageMaker 进行无缝连通。更是有助于容器开发人员和机器学习开发人员，构建相关的机器学习容器化实践。
- **企业迁移：**Amazon EKS 提供一系列迁移的最佳实践和方法，帮助开发者将基

于 Kubernetes 的容器平台向 EKS 进行迁移。对于企业或组织而言，集群迁移可以将工作委托给 AWS 处理，无需自行管理 Kubernetes 控制平面，有效的减轻企业的数字化转型难度。

- **数据分析平台：**在大数据领域，Amazon EKS 可基于 Spark 构建数据分析平台，利用 Kubernetes 原生模式来构建 Spark 数据分析平台。另外，EKS 还可利用 R 语言平台，基于 Shiny 和 EKS 构建 R 应用平台，帮助开发者进行数据分析应用的构建。

在落地实践上，对于 Amazon EKS 来说同样不成问题。一方面 EKS 保持了 Kubernetes 平台的速度、稳定性、可扩展性以及平台的良好集成度，另一方面搭配 AWS 云服务的加持，使其在全球广泛应用，包括金融、社交媒体、电子商务等。而如此庞大的用户基数，也正是 EKS 得以不断优化的关键。

AWS 脚踏实地，遵循“优良的平台架构是企业转型升级的保证”的理念，深耕于企业的数字化转型与应用现代化领域。随着 Amazon SageMaker、Amazon EKS 等产品在中国区正式商用，AWS 也希望通过技术的力量帮助更多的中国企业渡过难关，成功实现企业的现数字化转型，加速新经济的落地。

开发运维配置繁杂，是时候给应用架构做减法了

作者 | Cherry

“Less is more”是路德维希·密斯·凡德罗在建筑领域提出的观点，近些年来，这一观点不断被用于生活中的其他领域。在软件开发世界中，也有对“Less is more”这一观点的架构理念，这就是如今逐渐盛行的“Serverless 架构”。

十多年前，主流的应用架构都是单体应用，当时的开发者们既要关注所需的计算、存储资源，还要关注最底层的服务器等资源，同时当企业业务规模开始激增，导致开发和运维难度更大。随着容器技术的衍生及应用，虽然用户可以从对基础服务器关注中抽离出来，但其投入的运维精力依然绕不开的是与业务相关的 CPU、内存、网络等资源。

如今在资源投入、架构理念、架构模式向越来越精简，越来越“物尽其用”的演进中，Serverless 可以说是“Less is more”的最佳实践。它让开发人员不再操心运行所需的资源，精简自己的开发流程，将关注点重点放在产品代码、业务逻辑上，同时只需为实际消耗的资源付费。它使得程序开发架构中只保留了重要的、有价值的资源；其余的资源要么从开发主体中精简剔除，要么隐藏在选择性可见的界面中，用户随用随取。

1 Serverless 是“速度”与“激情”的再现

Serverless 是随着云计算、虚拟化的延伸发展历程演进而来的。有人说，未来将是 Serverless 的天下。那么，Serverless 究竟有哪些优势，使得它受到开发者们如此高的重视呢？

- **节省维护成本，可实现自动伸缩**

首先，Serverless 是一个基于云的服务，服务提供者帮助处理了服务器端的基础 IT 工作，比如把云部署从 x86 机器码（99%的云计算使用 x86 指令集）提升到了高级语言层面、管理操作系统、数据库版本升级等等。因而开发者们只要编写代码并部署它即可，不需要处理任何后端服务器的任务。

同时，相比于传统的非 Serverless 架构，这种架构模式带来的另一大优势是，开发者无需为过度配置或意外的负载峰值提前做好分配计划。因此，在企业级架构侧常常会遇到的服务伸缩性等问题，Serverless 也可以做到自动伸缩，或方便开发者对容量进行简单的手动设置。

- **节省人工成本，复杂工作都可以交给机器**

一方面，Serverless 有相对标准的编程环境，减少了对服务器和容器部署所耗费的操作成本。另一方面，在所有的应用程序架构中，Serverless 应用程序拥有的代码量最少，且恰当的 Serverless 架构在相互依赖性上较少。

对于开发者来说，这意味着更少的开发逻辑，用更少的代码来定义开发、测试、部署、运维。另外从应用程序角度来看，无服务器的功能基本上是一种外部服务，它不需要紧密集成到应用程序的容器生态系统中。

- **缩短交付时间与周期，节省开发成本**

随着产品及软件版本迭代周期的速度越来越快，一些云厂商在面向客户的咨询调研中发现，越来越多的客户已不满足于缩短开发与测试的周期，而是需要更短的交付周期——从新产品或功能的概念化到以 MVP 部署到生产环境的整个时间。

在应对该问题的解决方案上，Serverless 提供了巨大的作用。部分客户在使用该架构及应用程序后，能实现在几天时间内完成项目的部署。

总的来说，Serverless 可以称得上是当前各类新架构中“激情与速度”的再现——在降低人工成本、降低风险、降低基础设施成本、提高扩展性、缩短交付时间上，都形成了绝对的杠杆力。目前，Serverless 的适用场景非常广泛，或者说它或将成为大多数交付链中的一部分。

不过，必须要提的一点是，Serverless 所具备的优势并不等于它是万能的。很多开发者基于对 Serverless 优势的理解，容易陷入“它是容器替代方案”的认知误区。而实际上，Serverless 与容器针对的是不同的用户需求。

2 AWS Serverless 的基础技术革新之旅

1. Lambda 开启 Serverless 商业化进程

Serverless 商业化进程的真正开启，起源于 AWS 在2014正式推出的 AWS Lambda 计算服务。随后，各大巨头也都相继推出了相关服务，遂而将 Serverless 的市场竞争推向白热化，Serverless 是云服务商提供云服务能力的试金石，如何兑现向客户承诺的 Serverless 构建能力，需要云服务商的众多云服务能力作为支撑。

Lambda 的诞生，可以说是云计算技术的一次跃进式发展。正如上文所说，让开发者从对虚拟机、服务器机群容量、集群扩展这些细碎的关注点中抽离出来，Lambda 帮助其真正实现了按需执行、按需计费、按需自动弹性扩展和高可用能力。

值得一提的是，一些人更喜欢用缩写 FaaS (Function as a Service, 函数即服务) 来描述 Lambda 这类技术，对于无服务器技术来说，FaaS 只是无服务器技术和架构中必须提供的众多能力中的一种。但 Lambda 是 FaaS 的典型代表，它允许用户仅仅上传代码而无需提供和管理服务器，由它负责代码的执行、高可用扩展，支持从别的 AWS 服务或其他 Web 应用直接调用等。

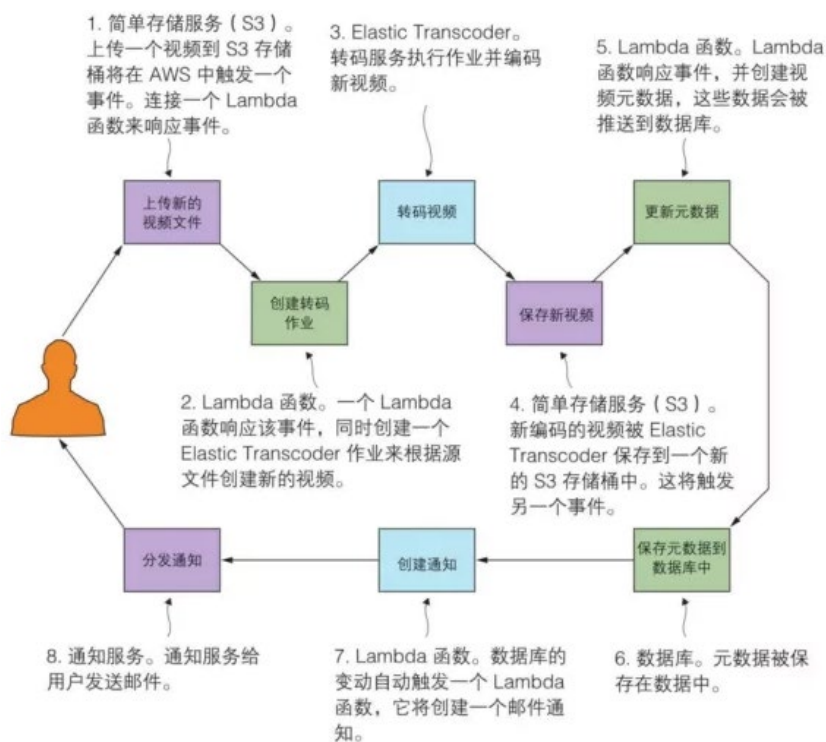
Lambda 能和大量的 AWS 服务进行整合。这里，我们将 AWS Lambda 放在若干个实际应用场景中，来向开发者们解释，基于它，能构建哪些内容，并如何和 AWS 的其他服务进行联动应用，加速开发。

• 数据处理与操作

Lambda 和 AWS 服务非常适用于构建用于处理数据的事件驱动管道。开发者可以使用 AWS Lambda 执行代码以响应数据更改、系统状态变化或用户操作等触发器，AWS 中的 S3、Amazon DynamoDB、Kinesis、SNS 和 CloudWatch 等服务，都可以作为 Lambda 的直接触发“机关”。

在数据处理管道中，许多用户会遇到数据上传后需要得到立即处理的场景，例

如需要将视频从一种格式转换成另一种格式，或者即时调整图像大小以匹配不同设备。Lambda 则可以实现实时创建缩略图、转换视频代码、聚合和筛选数据等，并且可以由 S3 或 Kinesis 触发。



一个模拟数据处理流中，Lambda 在各环节中的作用示意

• 实时数据流处理

很多 AWS 用户会使用 Lambda 和 Kinesis 处理实时流数据，从而跟踪应用程序活动、处理事务处理顺序等。其中，Kinesis 服务可以对数据（如日志、系统事件、用户点击等）的摄入进行处理，Lambda 函数则可以对数据流中的新记录做出反应，并能快速处理、保存或丢弃数据。

Lambda 和 Kinesis 的组合很适合会产生大量需要被分析、汇总并存储数据的应用程序。在应用程序产生的大量数据中，Lambda 可以随负载自动扩展和缩减，月度处理数据点可达百亿级。

- 后端

Lambda 还被用于构建无服务器后端，以处理 Web、移动、物联网（IoT）和第三方 API 请求。在很多客户场景中，可能会通过无服务器架构将前端直接连接到数据库，允许前端与服务进行安全通信，这里面只要通过 API Gateway，即可调用 Lambda 函数，Lambda 函数可以执行自定义任务并与其他服务通信。

2. Fargate 与 Firecracker 的诞生——Lambda 在“进化”

Lambda 所具备的丰富特性和应用场景的背景，让其成为一度流行于 FaaS 届的、可以称得上完美的方案。实际上，Lambda 当然也存在一些缺点与问题。例如迁移难度大、自动扩展性差、应用语言种类较少、计算规模受限、冷启动（函数未被运行一段时间后需要重新启动容器运行，而造成的函数调用被延迟）、不断膨胀的代码库维护等。

直至2017年年底的 AWS re:Invent 大会上，AWS 宣布针对容器的无服务器计算引擎推出 AWS Fargate，云计算技术尤其是 Serverless 架构和应用的演进，才算真正迎来了一次新的机遇点。

Fargate 不仅可以抽象出运行容器的服务器，还可以提供服务器编排的抽象，作为容器的免编排计算。

这也意味着，当 K8s 等容器编排工具的使用度越来越高，乃至成为开发中的一项“基础设施”时，开发者们可以将创建和管理容器的事情交给云服务商（Fargate）来处理，就好像今天的服务器虚拟化一般，容器也越来越“隐形”。

此外，相比于 Lambda 在自动伸缩、灵活定制资源等特征，Fargate 还可以通过与其他 AWS 服务（包括 Amazon CloudWatch Container Insights）的内置集成获得开箱即用的可观测性。Fargate 可以让开发者通过具有开放式界面的大量第三方工具来收集指标和日志，从而监控应用程序。

随后2018年的 AWS re:Invent 大会上，AWS 又开源了 Firecracker——AWS 容器安全沙箱的基础组件。它是 AWS 针对无服务器计算设计的虚拟化技术（利用 KVM 的新虚拟化技术，专门用于创建和管理多租户容器以及基于函数的服务）。目前，Firecracker 已为 Lambda 和 Fargate 在内的多个高容量 AWS

服务提供支持。Firecracker 诞生的内因，也是 Lambda 演进的结果。

从 Lambda 到 Fargate，再到 Firecracker，显示了 AWS 在 Serverless 架构等"基础服务"方面的革新能力。对于用户而言，这些服务的提供，正在让开发者逐步对其带来的安全、高性能、低开销等特性感知更加明显。

3 更多服务及工具，帮助开发者更高效地上手 Serverless

当然，除了 Lambda、Fargate 这类计算类服务外，AWS 可提供与之相关各个维度的一系列完全托管的服务。开发者可以使用这些托管服务构建和运行无服务器应用程序，从而解决一些特定问题。这里，我们列出了一份服务清单：

存储类	Amazon S3（对象存储工具） Amazon EFS（弹性文件存储工具）
数据存储类	Amazon DynamoDB（快速灵活的 NoSQL 数据库服务） Amazon Aurora Serverless（按需自动扩展配置的数据库） Amazon RDS 代理（高度可用的数据库代理）
API 代理类	Amazon API Gateway（处理并发 API 调用和流量管理等）
应用程序集成类	Amazon SNS（全托管的消息收发服务） Amazon SQS（全托管的消息队列服务） AWS AppSync（使用 GraphQL 的托管服务） Amazon EventBridge（无服务器事件总线服务）
编排类	AWS Step Functions（创建并运行一系列检查点和事件驱动的工作流）
分析类	Amazon Kinesis（流处理数据平台） Amazon Athena（交互式查询服务）

以上分类及工具清单来源于 AWS 官网 (<https://aws.amazon.com/cn/serverless/>)

有了 AWS 上述服务的支持，开发者无需为后端组件（如计算、数据库、存储、流处理、消息排队等）预置、维护和管理服务器。同时，应用程序的容错能力和可用性也可以变得更强。

此外，AWS 及合作伙伴生态系统也在开发者工具上提供了多样化使用组合，包括框架、软件开发工具包、IDE 插件和监控解决方案等。

例如框架层面，AWS 兼容了 AWS SAM（用简单方式定义 Lambda 函数、API、数据库以及事件源映射）、Apex、Chalice 等近十款 AWS 自研、开源或

第三方的框架供开发者使用。持续集成和部署层面，AWS CodePipeline、AWS Serverless Application Model、AWS CodeBuild 等一系列工具可以帮助开发者自动化构建、测试和部署无服务器应用程序。监控及日志记录与诊断层面，也有 Amazon CloudWatch 和 AWS X-Ray 等辅助进行函数性能监控或故障排除。

归纳来看，无论是扩充提供不同的服务还是丰富的开发者工具，AWS 都是尽可能地帮助开发者在应用 Serverless 架构的过程中，降低其遇到不同场景下处理复杂问题的难度，从而让为“高效”而生的 Serverless 技术能更高效的让开发者上手，更高效的解决问题，从而带来更高效地用户体验。

最后要提的是，**Serverless 是利用云的要素帮助用户实现价值交付的颠覆式创新。**

因为用户价值交付涉及方法论、开发者工具、应用交付体系、商业模式设计等多个维度，所以 Serverless 是顶层设计的产物。它并不是任何企业在任何场景下都必须“跟风”应用的时髦技术，毕竟它从真正诞生到至今应用，还只有短短6年而已。开发者们一定要选最合适，而非最流行的架构方式。

而一旦当你下定决心全面应用 Serverless，也一定要在这项新兴技术得到普及之前，学会借助实用的服务或工具来应对复杂问题，进而帮助你更快地创建高效、高性能的新架构及软件系统，让你的“酷想法”更快成真。

开发人员相关资源

Amazon SageMaker 是一项完全托管的服务，可为每位开发人员和数据科学家提供快速构建、训练和部署机器学习（ML）模型的能力。SageMaker 消除了机器学习过程中的每个步骤的繁重工作，开发人员能够更轻松地开发高质量模型。

- 入门

[使用 Amazon SageMaker 构建、训练和部署机器学习模型](#)

- 博客

提升效率：

- [Amazon SageMaker Processing](#) – 完全托管的数据处理和模型评估
- [Amazon SageMaker Experiments](#) – 组织、跟踪和比较机器学习训练
- [Amazon SageMaker 模型监控器](#) – 完全托管的机器学习模型自动化监控
- [Amazon SageMaker Autopilot](#) – 在完全控制和可见的情况下，自动创建高质量的机器学习模型
- [Amazon SageMaker Debugger](#) – 调试机器学习模型
- [使用 Amazon SageMaker Operator 简化 Kubernetes 上的机器学习推理](#)

优化成本：

- [通过使用 Amazon SageMaker 多模型终端节点节省推理成本](#)
- [托管 Spot 训练：最高可节省 90% 的 Amazon SageMaker 训练作业成本](#)
- [使用 Amazon SageMaker 降低机器学习的总体拥有成本并提高工作效率](#)

- [SageMaker 不到1美元！1小时内训练自己的中文词向量](#) [Amazon SageMaker BlazingText 一切皆有可能](#)

SageMaker 和其他 AWS 服务的集成：

- [使用 Amazon Athena 从您的 SageMaker 笔记本运行 SQL 查询](#)
- [隆重推出 Amazon SageMaker Operators for Kubernetes](#)
- [AWS 如何将机器学习送到每位开发者和 BI 分析师手中](#)
- [使用适用于 Amazon SageMaker 的 AWS Step Functions Data Science SDK 自动执行模型重新训练和部署](#)

边缘智能：

- [使用 Jetson nano 结合 AWS 机器学习和 IoT 实现边缘智能—上篇](#)
- [使用 Jetson nano 结合 AWS 机器学习和 IoT 实现边缘智能—下篇](#)
- [预测性维护在工业物联网的应用](#)

自定义模型开发：

- [发布 Amazon Rekognition 自定义标签](#)

相关发布系列：

- [Amazon SageMaker 现已推出：Deep Graph Library](#)
- [Amazon SageMaker Studio：首个全集成的机器学习开发环境](#)
- [使用 Amazon SageMaker 运行分布式 TensorFlow 训练](#)
- **网络研讨会**
 - [AWS 机器学习平台——Amazon SageMaker 详解](#)
 - [让人人皆为人工智能的开发者——AWS 机器学习服务综](#)

- [AI 基础架构资源的演进与选择——CPU、GPU、ARM 和 ASIC \(AWS Inferentia\)](#)
- [Amazon Prime Video 如何做视频推荐](#)

• 客户案例

- 新媒体— AWS 案例研究: [大宇无限 \(SageMaker\)](#)
- 新媒体— AWS 案例研究: [虎牙直播 \(SageMaker\)](#)
- 游戏— AWS 案例研究: [嘉谊互娱 \(SageMaker\)](#)
- 智能制造— AWS 案例研究: [天津华来科技有限公司 \(SageMaker\)](#)

• 视频

[10分钟视频教程:使用Amazon SageMaker训练汽车型号图像识别的模型](#)

[10分钟视频教程: 使用 BlazingText快速训练中文词向量](#)

[10分钟视频教程: 使用 Amazon Sagemaker训练因子分解机模型并应用于推荐系统](#)

[10分钟视频教程: 使用AWS Sagemaker XGBoost模型进行房价预测](#)

[10分钟视频教程: 利用DeepAR进行时间序列预测](#)

[将您的模型迁移到 Amazon SageMaker](#)

[使用 Amazon SageMaker 内置算法加速模型开发](#)

[使用 Amazon SageMaker 托管您的模型](#)

[使用 Amazon SageMaker 托管的 Jupyter Notebook 实例](#)

[使用 Amazon SageMaker 超参自动调优](#)

[使用 Amazon SageMaker 训练模型](#)

[使用 Amazon SageMaker 进行分布式训练](#)

[使用 Amazon SageMaker 进行特征工程](#)

使用 Amazon SageMaker 超参自动调优

在 Amazon SageMaker 上使用深度学习框架 (TensorFlow、MXNet、PyTorch 等)

其他 AI 资源: <https://amazonaws-china.com/cn/ai/aiml-resources/>

Amazon Aurora 存储和可靠性

- [Aurora 存储概述](#)
- [集群卷包含的内容](#)
- [Aurora 存储如何自动调整大小](#)
- [Aurora 数据存储的计费方式](#)
- [Amazon Aurora 可靠性](#)

Amazon EKS 集群

- [简介](#)
- [功能](#)
- [常见问题](#)
- [创建集群](#)