

上海科学技术情报研究所
上海市前沿技术发展研究中心
TISC技术与创新支持中心

数字化城市

DIGITAL CITY BRIEFING

2023 年
第 4 期



云的基础设施化助推天空计算

编者按

新冠病毒大流行的遗产之一，就是交互协作的工作模式被继续保留下来，其中离不开云技术的发展。随着细分市场企业加速向云转移，多云管理的复杂性将助推云的公共设施化趋势。

目前，云计算向天空计算的形态转移这一趋势正发生在产、学、研的各个环节。美国加州大学伯克利分校对这一领域展开了初步探索，该校 Ion Stoica 等人在论文 *From Cloud Computing to Sky Computing* 中思考了该如何将目前的差异化云计算平台逐渐推动发展成为一名公共服务，并称其为“Sky Computing（天空计算）”。此外，该校成立的天空计算实验室推出了包括面向机器学习等批处理任务的天空计算编排系统 SkyPilot 和跨云数据传输系统 SkyPlane 等研究成果。美国其他高校相继成立天空计算实验室和技术中心；多地政府包括英国、新加坡、中国香港、印度等纷纷部署云基础设施服务；罗马尼亚初创公司则发布了企业的多云管理平台版本。



目录

技术3

 云的公共设施化催生天空计算3

 天空计算应对多云管理的复杂性7

产业11

 2023 年的云计算趋势11

政策15

 多地政府部署云基础设施服务15

 加州大学伯克利分校成立天空计算实验室17

 华盛顿大学艾伦学院成立云计算技术中心18

案例20

 加州大学伯克利分校推出天空计算开源框架20

 罗马尼亚初创公司发布企业多云管理平台版本22



云的公共设施化催生天空计算

2021年6月加利福尼亚大学伯克利分校计算机科学与电气工程教授Ion Stoica在展望云计算的未来时,思考了该如何推动目前的差异化云计算平台逐渐发展成为一项公共服务,称其为Sky Computing（天空计算）”在其论文 *From Cloud Computing to Sky Computing* 中, Ion Stoica 提出实现天空计算的障碍来自于经济层面,而不是技术层面,对等互惠则是实现天空计算的关键步骤。Stoica 从更长远视角对天空计算进行了设计,提出将其作为通用软件平台用于未来计算,并考虑技术趋势和市场力量如何在天空计算的实现中发挥关键作用。

正如这个词汇的含义,天空计算是云平台之上的一层,目标是实现云之间的互操作性。 本文将对 Stoica 的这篇论文及天空计算作简要介绍。

1. 计算机公用设施化的愿景

20 世纪 60 年代,人工智能之父约翰·麦卡锡提出了把计算能力作为一种像电话一样的公用事业提供给用户的理念,云计算由此起源。麦卡锡的预测颇有先见之明,云计算最终应运而生,作为一种新兴的资源使用和交付模式逐渐为学界和产业界所认知。然而,他在经济学方面的预测却与现实相去甚远。如今电话服务并没有成为一种公用设施,而是由互相竞争的供应商提供。同样地,现在也没有单一的公用计算设施,云计算也是差异化商品。并且,云计算市场已经离商品化越来越远,甚至演变成一系列彼此基本不兼容的专有平台,例如 AWS、微软 Azure、谷歌云等。

2. 云计算的历史

因特网的出现迅速带来了可以在全球范围内进行访问的多种服务,包括电子邮件、BBS 和游戏等。万维网的出现则带来了搜索、在线零售和社交媒体等新服务的爆炸式增长。为了应对由此产生的大规模服务,这些供应商必须构建数据中心,并设计自己的复杂分布式系统。许多公司都遵循了这条路线,例如雅虎、Google、eBay 和 Amazon 等。但是创建大规模的基础计算设施代价高昂,阻碍



了大多数公司进入互联网服务市场的脚步。

这种情况在 2006 年亚马逊推出 S3 和 EC2 时发生了变化。通过普及计算/存储访问，并且推广“即用即付”商业模式，亚马逊开启了云计算时代。再加上摩尔定律的终结（在本地集群中构建和扩展服务十分昂贵），重新推动了构建公用计算事业的设想。然而，商业趋势却将云计算推向了不同的方向。

在过去十几年中，云计算市场内出现了多个竞争对手。根据一份报告，AWS 现在仅拥有 32% 的市场份额，其次是微软 19%，谷歌 7%，阿里巴巴 6%，其他云平台（如 IBM 和甲骨文）瓜分了剩余 36% 左右的云计算市场。这种竞争导致价格越来越低，产品和服务种类不断增加。例如，仅 AWS 一家就提供了超过 175 种产品和服务。这些服务中有许多是专有服务，例如，每个云平台都有自己独特的用于管理集群的 API、对象存储、数据仓库、无服务产品等等。在某个云上开发的应用程序通常不能在不同的云上运行。直到 2021 年，没有一个底层云平台有一套任何人都可以使用的开放标准，云市场的竞争使离公用计算的愿景越来越远。因此现在要解决的问题是在计算机公用设施化方面取得进展。

3. 云计算与互联网的发展

尽管云计算和互联网在许多方面存在差异，但是互联网的公用设施化为公用计算提供了有用的历史经验。有 3 个关键的设计决策使互联网能够为大量异构技术（从以太网到 ATM 再到无线）和竞争公司提供统一的接口。第一个是掩盖技术差异的“兼容性”层。第二种是域间路由，它将互黏合在一起，使其对最终用户来说是一个网络。第三个是一组经济协议，称之为“对等”层，允许相互竞争的网络合作创建一个统一的网络。

为了实现计算机公用设施化的愿景，应用程序应该能够在任何云平台上运行（即一次编写，随处运行）。此外，用户不必管理单个云上的部署，也不用面临从一个云迁移到另一个云的各种障碍。简而言之，开发人员构建多云应用程序应该像构建在单个云上运行的应用程序一样容易。Stoica 称之为天空计算。使用这个术语是因为公用计算意味着基础计算设施是公用设施，而天空计算指的是在由多个不同云平台组成的基础设施上构建公用计算的愿景。

Stoica 认为互联网解决的这三个设计问题正是从当前的云计算中创建天空



计算所需的思想。天空计算需要一个兼容层来掩盖低级技术差异，一个云间层将作业路由到正确的云，以及一个对等层，允许云之间就如何交换服务达成协议。

4. 兼容性

实现天空计算愿景的第一步是提供一个云兼容层，通过抽象出云计算服务，使在该层之上开发的应用程序无须更改即可在不同的云上运行。简而言之，兼容层是一组可以构建应用程序的接口或 API，然后通过云平台的一组接口将该兼容性层移植到每个云。

该层可以类比于互联网中的 IP 层，然而与 IP 层不同的是，云兼容层明显更广泛且定义不明确，因为云平台向应用程序提供了丰富的一组服务，包括计算、存储、数据传输等。因此，云兼容层更像一个操作系统（例如 Linux），可以管理计算机资源并向应用程序提供 API。

虽然每个云平台都提供一组专有的低级接口，但目前大多数用户主要与更高级别的管理和服务接口交互。其中一些是专有的，但已经有越来越多的接口受到开源软件（OSS）的支持。此外，涌现出大量由 OSS 创建者创立的公司，以在多个云上提供托管服务。如果企业的应用程序使用这些基于多云 OSS 的产品，从一种云切换到另一种云就能变得相对容易。

兼容层可以从这些 OSS 解决方案中构建出来。例如 Cloud Foundry，是一种开源多云应用平台，支持主要的云供应商以及内部部署集群。以及 RedHat 的 Open Shift，是一个基于 Kubernetes 的多云和内部部署平台。因此，从纯粹的技术角度来看，实现一个能够广泛使用的兼容层是很容易的。问题是市场是否会支持这一发展。因为虽然兼容层对用户有明显的好处，但它自然会使云平台商品化，这可能不符合供应商的利益。

5. 云间层

兼容层只是实现天空计算愿景的第一步。即使兼容层允许用户在不同的云上运行应用程序，用户仍然需要决定在哪个云上运行应用程序，需要在不同云之间进行性能/成本权衡。这类似于要求互联网用户为其域间流量明确选择 AS 路径，为了解决这个问题，Internet 使用 BGP 来做出 AS 级别的路由决策。同样，天空计算应该实现一个云间层，从用户中抽象出云供应商；也就是说，用户不需要知



道应用程序在哪个云上运行。

通过云间层，用户可以指定有关其作业应在何处运行的策略。这些策略允许用户表达他们对性能、可用性和成本之间权衡的偏好。此外，用户可能希望避免他们的应用程序在竞争对手运营的数据中心上运行，或者留在某些国家/地区以遵守相关的隐私法规。并且云间层也可以实现更安全的应用。

Stoica 相信在兼容层之上实施云间层基本没有技术限制。因为，从性能的角度来看，跨云移动作业与跨数据中心移动同一云平台中的作业非常相似。一旦应用程序被设计为跨多个数据中心运行，剩下的跨云问题可以通过以下 3 个功能来解决：（1）OSS 服务的统一命名方案；（2）目录服务，允许云供应商注册他们的服务，并允许应用程序根据他们的喜好选择服务；（3）跨云计费机制。

服务命名方案 为了识别在特定云上运行的服务实例，需要一个命名方案来识别该实例。一种方案是利用 DNS 来命名这些服务实例。此外，需要将元数据与每个这样的服务实例相关联。此类元数据应包含如何调用服务、云供应商的名称、位置、软件或 API 版本、硬件类型等。此外，可能还需要添加动态信息，如定价、负载和可用性。

目录服务 需要特定服务的应用程序必须找到满足其要求和偏好的服务实例，这需要目录服务。每个云供应商可以通过提供名称和元数据信息将服务发布到该目录。此外，云供应商应定期更新其动态元数据，例如负载和定价。反过来，应用程序应该请求能够表达其偏好和要求的特定服务。收到此类请求后，目录服务将返回满足这些要求和偏好的实例。

记账和计费 通过天空计算，用户的应用程序可以运行在多个云中的一个云上，甚至同时运行在多个云上，每个云平台都必须对所使用的资源进行核算。如果每个云平台都进行计费，那么用户需要在每个云上都有账户。有一种替代方案，让每个用户都与第三方签约并累计费用，然后将每个用户的付款分配回各个云平台。

尽管仍有许多细节有待解决，但实现合理有效的云间层并没有不可逾越的技术障碍。与兼容层一样，问题在于市场是否能产生云间层。



6. 对等层

云间层旨在最能满足其需求的云（或多个云）上运行作业。如果作业涉及大型数据集，需要将数据移动到将发生计算的云中。目前大多数云平台都有定价政策，将数据传入云中比将数据移出要便宜得多。例如，将数据提取到 AWS 是免费的，而将数据从 AWS 传输出去可能需要 0.05~0.09 美元/GB。作者将这种定价形式称为“数据引力”定价，它驱动用户在当前所在的云平台中处理数据。尽管如此，对于一些计算资源比数据传输成本高得多的工作，将数据从一个云移动到另一个云仍然是最划算的选择。

目前，导出数据的定价政策与数据可能要去的云无关。Stoica 提出了“互惠数据对等”方案，迄今为止还没有看到对此类方案的探索。在这种方案下，云间可以通过建立高速连接互相免费传递数据，使数据传输又快又便宜，降低两个同级云之间的数据重力，并实现更大的工作流动自由。

在论文中，Stoica 提出了天空计算这一构想，介绍了从云计算转变为天空计算之前必须克服的一些挑战，这将使人们更接近麦卡锡的计算机公用设施化概念。其中一些挑战纯粹是技术性的，并且很可能是可以实现的。然而，为了使经济激励发挥作用，天空计算需要云供应商采用互惠数据对等策略，以便工作负载可以在这个庞大的云集合中轻松迁移。本文旨在确定这是组织云市场的一个关键步骤，并希望为实现天空计算这一愿景创造动力。

参考文献：

- [1] STOICA I, SHENKER S. From cloud computing to sky computing[J/OL].<https://sigops.org/s/conferences/hotos/2021/papers/hotos21-s02-stoica.pdf>.
- [2] 达坦科技. 从云计算到天空计算[EB/OL]. (2023-01-20)[2023-06-02].<https://segmentfault.com/a/1190000043350594>.

天空计算应对多云管理的复杂性

在云计算技术发展的初期，相较以往，开发人员不再受限于本地服务器，他们可以发挥想象力开发新产品和服务。由于不受繁琐的资源申请流程的限制，软件工程师只需一次点击就可以获得所需要的资源。随着供应商在其平台添加了越来越多的高级功能，如机器学习等，云迅速变成了满足所有需求的“一站式商店”。



然而，目前企业逐渐发现云服务同样存在一些局限。随着开发人员对最新的云服务和工具的热情高涨，支撑业务的平台也越来越多。云工具形成了一个复杂的网络，实现了互联互通，但经常也带来不必要的冗余。

采用多云战略的企业在不断增加，即使用多个云环境和供应商的组合服务。然而，虽然多云战略至少理论上可以提供更专业的能力和更优的价格，但由于在不同专有平台、服务和接口界面上混合操作较为复杂，设计应用和管理工作负载很有挑战。这意味着许多企业难以发挥云投资的所有价值，包括自助服务、广泛的网络访问、敏捷弹性、资源池化以及服务计量等。为简化多云管理，一些企业正在转向位于其迅速发展的多云架构之上的抽象和自动化层。它们也被称为天空计算。

1. 多云带来的复杂性催生天空计算

如今，不论是否自愿，绝大多数企业都在使用多种“平台即服务”的工具。多达 85% 的企业正在使用两个及其以上云平台，有 25% 的企业使用至少 5 个云平台。这种情况在短期内不太可能改变。无论在哪个云平台中，解决方案团队都想用最合适的工具来工作。他们不想被受制在某一供应商的“围墙花园”内。此外，他们也利用供应商之间的竞争来获得更优的服务条款。对大多数企业来说，将大多数业务整合到一个云供应商中不太可行，在可预见的未来，多云很可能会成为主流。

但是，许多处于多云环境中的企业并非有意为之。他们临时引入了新的服务，却没有整体性的战略来处理冗余和安全等问题。同时维护多个安全配置和数据存储使多云管理变得复杂。技术领导者希望消除这些复杂性，否则企业无法利用云服务降低成本或提高运营效率。多云的复杂性为企业带来了一些问题，例如为不需要的服务付费、安全漏洞以及难以找到合适的人来处理问题。

2. 天空计算让集中控制多云成为可能

精明的企业领导者不会任由自己被多云和科技发展的乱象迷惑。相反，他们积极寻找方法，通过管理多云来获得运营收益并消除多云的复杂性。

这个方法就是天空计算，即构建一个兼容层来提供对存储和计算、人工智能、数据、安全、运营、治理以及应用开发和部署等通用服务的访问。该兼容层逻辑



上位于企业的各种云平台之上，通过 API 调用各云平台的本地技术标准，这样可以让应用程序在集中控制下仍具有云服务供应商所提供的良好安全防护。天空计算通过一个公共界面实现了这一点，让管理员能够集中控制多个云。计算领域专家 Ion Stoica 和 Scott Shenker 在计算机协会(Association of Computing Machine) HotOS 会议上发表的文章称，天空计算的兼容层可以围绕 API 建立。这使得兼容层可以向每个单独的云接口发送指令。他们把云兼容层比作计算机的操作系统，管理计算机的资源，并将 API 接口开放给各应用程序。

3. 天空计算减少企业的安全风险

通过在各云平台之上的一个抽象和自动化层，企业就不再需要那么多专业人才。云开发人员可以培养更多的通用技能，而不用专注于特定的云平台。

天空计算还可以提高安全性。一般来说，每个云平台都有完备的安全标准，独立工作时也能出色完成。问题往往在企业混合搭配多个云服务时出现。由于要管理多个云平台，进行必要的安全设置变得更困难。开发人员可以通过天空计算在兼容层设置统一的安全配置，再通过各云平台的本地接口执行，从而解决上述问题。

取消不必要的云服务可以让企业减少安全风险，加强用户隐私，降低成本，以更少的代价获得更多的效益。企业不再过度追求员工的专业化，员工可以把精力集中于解决更高层次的、更普遍的问题。

4. 天空计算在商业层面的挑战

单纯从技术角度而论，天空计算非常有意义。但从商业角度看就没那么简单了。单纯从技术角度而言，广泛可用的兼容层是很容易实现的，问题在于能否得到市场的支持，因为虽然兼容层对用户有明显的好处，但它也会导致云服务的商品化，这可能不符合云供应商的利益。

天空计算的另一潜在问题是，企业需要自己构建它。目前，很少有供应商把天空计算作为一种服务来提供。因此，开发团队需要自己动手建立每个连接和最终的接口。这是用复杂的方案来应对复杂性，但最终的结果应当更加精简。

5. 天空计算可能只是过渡性方案



云技术的发展历史表明，天空计算可能只是一个过渡性的解决方案。为了管理好百花齐放的数据中心、数据库和操作系统等，企业在过往的经验可归结为整合化、集中化、标准化和合理化，而非通过中间件或者编排引擎。提高许可证利用率一直是整合的重点，大多数企业都有管理云资源利用率的政策。

此外，云服务通常被卖给 IT 部门的各个团队。很可能一旦 IT 部门将云服务集中到天空计算下，总会发现需要纳入管理的新的平台。

最终取代天空计算的可能是一种更具策略性的方法，它借鉴了天空计算的集中化和控制模式，同时开发人员可以继续像现在一样自由选择最适合的工具。这种策略性的天空计算可以管理云凭证的供应，只将资源分配给具备有效商业案例和技术知识的用户，以有效利用云资源而不造成复杂的情况。

自助服务一直是 IT 领域的终极颠覆能力。任何集中云资源的方式都必须满足终端用户对敏捷性的需求。过去，集中管理通常意味着垂直系统。业务线用户绕过 IT 管理部门以摆脱繁琐的流程，这些流程拖慢了解决业务问题的速度。但是，自动化工具让 IT 职能服务更加简单，而不会拖慢业务动作，并且有助于控制多云复杂性。

企业和技术领导者应尽可能运用天空计算降低多云的复杂性，消除因维护多云而产生的安全和冗余问题。

参考文献：

[1] Deloitte Insights. Tech Trends 2033[R]. 2023.

2023 年的云计算趋势

过去几年剧烈变化的社会经济环境对云技术的采用产生了持久影响。新冠病毒全球大流行和各国政府形形色色的居家令促使各个公司加速了数字化转型，其中就包括转向基于云的应用程序，用以支持远程工作人员，并迅速推出新的云服务以留住客户。

2022 年，数字化转型计划仍处于快车道，而且 2023 年的云计算趋势也没有放缓的迹象。Gartner 公司预测，全球公共云支出将从 2022 年的 4 903 亿美元增长 20.7%，至 5 918 亿美元。预计到 2025 年，关键细分市场中将有超过一半的企业 IT 支出转移到云端。

1. 2023 年的 7 大云计算趋势

尽管经济衰退迫在眉睫，但企业无力中断数字化转型计划。这是建立在云计算技术的基础上的。公共云的弹性和可扩展性降低了创新的财务风险，同时在公司最需要的时候帮助实现业务敏捷性。

Gartner 预测到 2025 年，企业在公共云服务上的支出将超过传统 IT 解决方案。以下是可以预期的一些云计算趋势：

- 为混合动力工作提供动力；
- 云的可持续性；
- 人工智能的采用促进了云增长；
- 企业拥抱主权云（Sovereign Cloud）；
- XaaS 的成熟；
- FinOps 呈上升趋势；
- 云原生策略取得进展。

尽管人们重新回到线下办公，但这个过程并没有一蹴而就。组织在新冠病毒大流行期间开发的远程和混合工作模式将被继续保留下来，他们会继续使用云技术来支持这些灵活的工作环境。Gartner 估计，随着公司从传统的客户端计算解决方案转向基于订阅的虚拟桌面服务，2023 年的全球桌面即服务支出将达到 32



亿美元。

用于通信、项目管理、视频会议、文件共享、知识管理等基于云的工具仍将是公司的优先事项,因为它们希望随时随地支持员工的相互协作和生产力的提高。长期的混合工作模式正在推动对基于云的生产力技术以及协作工具的需求。

2. 2023 年的云计算趋势：可持续发展目标

2023 年,预计会有更多公司利用云的效率来实现其可持续发展目标。

根据 IDC 的一项市场调查,83%的受访者认为可持续性 IT 购买决策中最重要的标准。IDC 预计到 2025 年,85%的组织将可持续效率提高 35%。

由于云服务提供商聚合其客户的资源,CSP 可以实现小型公司无法实现的规模经济。与在本地运行应用程序或工作负载相比,在云中运行应用程序或工作负载不仅效率更高,而且在云中运行整个数据中心产生的碳排放量低于在本地运行的数据中心。

3. 2023 年,采用 AI 是必须具备的趋势

随着组织寻求提高运营效率和推动创新,企业对人工智能的采用一直在稳步增长。云正在帮助他们从 AI 投资中获得最大收益。

麦肯锡的一项关于人工智能的调查发现,64%的“高绩效者”,即那些最深刻地意识到人工智能的采用会产生的影响力的人,他们在公共云或混合云上运行人工智能,而在其他组织中这一比例为 44%。

高绩效者还使用公共云来访问更多人工智能功能,包括自然语言语音和面部识别等。根据这项研究,近三分之二的公司计划在 2023 年加大对人工智能的投资。与私有数据中心相比,公有云基础设施具备人工智能和机器学习应用所需的充足的数据存储和计算能力。

4. 企业拥抱主权云和行业云

为满足越来越多的数据隐私要求,2023 年主权云解决方案和行业特定云的开发和采用将得到普及。

主权云(Sovereign Cloud)的概念并不新鲜,但随着地缘政治环境的变化而受到越来越多的关注。主权云在特定国家或地区运行,支持当地管理机构的数据



保护标准。

Capgemini 的一项研究表明，来自美国、英国和德国等 10 个国家/地区 71% 的企业领导者认为，他们的组织应当采用主权云服务来确保合规性。这对于云服务提供商来说既是一个市场机会，也是对他们受到监管组织保护的一种保障。据 Forrester 的调查，向主权云解决方案的转变由德国金融服务提供商和其他欧洲国家领导。

此外，为满足行业特定需求而量身定制的行业云也将增长，包括专为零售、电信和公用事业设计的专用应用程序。据 Gartner 称，到 2027 年，超过一半的企业将使用行业云解决方案。

5. XaaS 云产品在 2023 年发展成熟

“一切即服务”（XaaS）的日益成熟也将有助于推动 2023 年的云投资。XaaS 是指通过互联网向客户提供的范围广泛的产品和数据，根据研究公司 IMARC Group 统计，全球 XaaS 市场在 2021 年达到了 1 986 亿美元，该公司预计到 2027 年，XaaS 市场将达到 6 241 亿美元。

IMARC 的估计还包括存储即服务、安全即服务和网络即服务。对于最终用户，XaaS 的好处包括灵活性、可扩展性、速度和通过灵活的支付模式节省的成本。也就是说，用户只需为使用的东西付费，而不是预先支付许可费用和现场设备。

XaaS 模型在专业服务行业中不断发展，在该行业中，公司将数字功能、数据或资产与以人为本的服务打包在一起。该模型允许专业服务提供商扩展他们的产品，同时使组织能够为结果而不是花费的时间付费。

将专业服务公司过渡到 XaaS 商业模式并不容易，但 2023 年将有更多新来者和传统公司开发和改进他们的产品。

6. FinOps 成为顶级云计算趋势

虽然组织 2023 年将在云上花费更多，但他们也将密切监控这些支出以努力控制成本，尤其是在通货膨胀继续增加经营成本的情况下。这将推动对 FinOps 的投资。根据 IDC 的数据，到 2023 年，80% 的使用云服务的组织将建立起来。

FinOps 是企业内部为管理云计算成本而开发的一种实践。由于按需云模型



的成本会波动，因此这是了解整个组织的云消费情况的一种方式。

Forrester 预计组织将转向第三方云成本管理和优化平台，以更好地了解他们在多云部署中的云支出。

7. 云原生策略呈上升趋势

云原生策略涉及一种现代软件开发方法，包括微服务、容器、声明式 API 和服务网格。

根据 Forrester 的 2022 年基础设施云调查，40%的公司将在 2023 年采取云原生优先战略，因为他们希望在降低成本的同时提高敏捷性和效率。

Forrester 预计会有更多企业采用云原生策略，因为他们越来越多地选择在容器而不是传统虚拟机中运行工作负载，因为容器在利用 AI/ML、物联网和 5G 等技术时运行效率更高。越来越多地采用容器也将推动组织对其应用程序开发流程进行现代化改造。

组织已经过了争论迁移到云的价值阶段。无论是利用 AI 和 ML 等先进技术，获得竞争优势，还是只是在经济动荡时期利用安全港，组织都不得不迁移到云。

参考文献：

[1] ZANDE V J. Cloud computing trends 2023: Top predictions, stats, and growth drivers [EB/OL]. (2022-11-09) [2023-06-02]. <https://www.the-future-of-commerce.com/2022/11/09/cloud-computing-trends-2023/>.

多地政府部署云基础设施服务

如今，云计算已成为一项帮助各种规模和级别的政府实现数字化转型目标的重要技术。不过，密集的人工智能应用和安全隐私要求对于单个本地云提供商来说成本过高且技术过于复杂。必须由两个或以上的云提供商在日益普遍的混合云解决方案中进行任务分担，才能解决这样的挑战。

在这种情况下，政府转向创新的云解决方案则非常有意义。为了实现下一阶段的数字化转型，使各国从全球数字经济中获益，政府必须在可行的情况下将重点从依赖遗留的 IT 系统转移到利用更多基于云的解决方案来管理其数据，并且越来越多地使用通过多个客户共享的基础设施提供的公共云服务。

1. 英国——“云优先”政策

英国政府采用了“云优先”政策，即在将技术与公共服务集成时首先考虑和评估公共云解决方案，然后再寻求其他替代方案。政府各部门有权根据安全性、灵活性和物有所值选择最适合其需求的云解决方案。

2. 新加坡——私有云“G-cloud”

新加坡为政府创建了一个私有云，即“G-cloud”，以满足政府的安全和治理要求，为公有云和私有云解决方案可以一起运行的混合解决方案铺平了道路。如果公共云或 G-cloud 不能满足新加坡各部委的需求，可以选择更合适的解决方案。

3. 中国香港——政府云基础设施服务（GCIS）

中国香港特区政府推出的云基础设施服务（GCIS）是新一代政府云基础设施服务，它取代了政府云（GovCloud）、中央计算机中心虚拟化基础设施（CCC VI）和电子政府基础设施服务（EGIS）。GCIS 运用现代云技术，提供安全、可靠和可扩展的 IT 基础设施，配备灵活的应用程序开发工具，以促进决策局和部门（B/D）灵活开发和交付数字政府服务。GCIS 提供四种主要类型的服务，即基础设施即服务（IaaS）、平台即服务（PaaS）、数据库即服务（DBaaS）和容



器即服务（CaaS）。

GCIS 的优势包括：通过规模经济和资源共享节约成本；通过简化采购和系统实施以及按需提供服务来节省时间；提高敏捷性，满足决策局 / 部门的动态需求，支持智慧政府发展；产生对各类 IT 专业工作和服务的需求，并通过加强云计算相关专业技能和商业模式促进本地 IT 行业的发展。

4. 印度——创建强大而安全的云生态系统

在印度，云计算作为政府机构和部门为其选民提供更好服务的战略工具正在获得发展势头。印度政府推出了多项计划和政策来促进云采用，例如国家云计划、GI 云（MeghRaj）项目和云服务聚合和管理平台（CSAMP）。这些努力旨在创建一个强大而安全的云生态系统，以支持各种政府利益相关者的需求，包括中央和州政府、地方机构、公共部门企业和公民。

印度政府实施了基于云计算技术的项目。例如，印度南部安得拉邦政府实施了一个基于云的解决方案来管理其土地记录。Bhudhaar 项目涉及将土地记录及其在云平台上的存储数字化，防止物理土地记录被篡改和出错。又如印度政府的 National Health Stack（NHS），这是一个基于云的健康信息交换平台，旨在提高该国医疗保健服务的可及性、可负担性和质量。该平台利用云计算在利益相关者（包括患者、医疗保健提供者、研究人员和政策制定者）之间安全地存储和共享电子健康记录、远程医疗咨询和其他健康相关数据。

参考文献：

- [1] The World Bank. Cloud Services Advance Digital Transformation for Governments[EB/OL]. (2022-06-10)[2023-06-05]. <https://www.worldbank.org/en/news/feature/2022/06/07/cloud-services-advance-digital-transformation-for-governments>.
- [2] OGCIO. Government Cloud Infrastructure Services (GCIS) [EB/OL]. [2023-06-05]. https://www.ogcio.gov.hk/en/our_work/infrastructure/e_government/gcis/index.html.
- [3] DELVADIYA D. The benefits of cloud computing for government[EB/OL]. (2023-04-13) [2023-06-05]. <https://www.tutorialspoint.com/the-benefits-of-cloud-computing-for-government>.



加州大学伯克利分校成立天空计算实验室

在新冠病毒全球大流行发生之前，云计算已经在增长，它通过互联网提供数据存储和视频流服务。在大流行期间，随着社交距离变得必不可少、社交中断变得频繁，远程工作成为许多国家多个行业（包括医疗保健、教育和商业等）的常态，该技术得到进一步发展。在大流行后的世界中，云的使用可能会继续增长。据 Gartner 估计，到 2025 年，关键细分市场中将有超过一半企业的 IT 支出转移到云端，超过传统 IT 支出。

然而，问题仍然是企业如何从向云的过渡中进行价值的最大化发挥。云计算仍然是一个分散的行业。换句话说，它在成本和价值、定制（客户和服务之间的适当匹配）和可扩展性方面缺乏透明度。该技术还会提供冗余产品。例如，在不同公司提供的服务类型方面没有清晰的市场意识，企业尝试云计算服务将是一个耗时的过程。这有点像花几个小时与不同的保险公司通电话，试图找到最好的汽车保险费率。此外，资产（在本例中为数据）的转移仍然具有挑战性，阻碍了使用不同云的团队之间的协作。此外，对数据安全和隐私的日益关注，以及随之而来的地方政府监管的增加，将为跨不同云的服务、跨越不同国家和大洲的高效计算和分析带来更多挑战。因此，需要降低使用云计算的成本，同时提高兼容性和安全性，以便将该技术更广泛地应用于更多应用，包括自动驾驶汽车和远程医疗保健等。

加州大学伯克利分校最近推出的天空计算实验室是一个旨在加速向云计算过渡的研究实验室。天空计算实验室并不满足于寻找逐步增加云计算采用率的方法，而是旨在为互联云计算构建新的基础架构，以解决行业面临的许多问题，例如成本高、兼容性差和可用性低。

实验室希望将现有的云计算碎片化市场转变为公用事业市场，让云计算变得像互联网一样互联易用，最终演变为天空计算。

实验室的研究人员设想天空计算有几个组成部分。首先，中介服务或“互联云代理”将充当互联程度更高的云计算的关键。每个中介都有一个目录，其中包含有关云服务、兼容性、定价和性能的信息。同时，工作接收界面将收集用户需求的描述，以便根据使用什么云和何时使用来创建优化的用户特定计划。



此外，天空计算将提高兼容性，因为代理服务可以将数据移动或复制到指定位置、分配资源以及跨云执行计算计划并排除故障。这些服务将具有验证用户身份的能力，确保数据安全。

未来，实验室期望跨云协议能够实现数据的自由交换和许多不同云的连接。

根据实验室的说法，天空计算最终将扩展到更多数量和范围内对使用云服务感兴趣的消费者，从而让用户对于使用适合其需求的硬件变得更加容易，而不管最初哪个云拥有该应用程序。用户将能够跨云聚合资源以运行他们无法在单个云账户上访问的大型工作负载。此外，天空计算将帮助企业遵守不同国家的数据主权规则，更轻松地保护他们的数据。

天空计算的愿景正在成为现实。实验室的团队正在测试用于机器学习工作负载的早期版本的 **intercloud broker** 原型。研究人员在题为“云端之上的天空”的白皮书中阐述了他们的愿景。实验室提出的建议无异于一场重要技术生态系统的革命。

参考文献：

[1] HE S. UC Berkeley's Sky Computing Lab aims to connect the clouds[EB/OL]. (2022-07-27)[2023-06-02]. <https://www.engineering.com/story/uc-berkeley-sky-computing-lab-aims-to-connect-the-clouds>.

华盛顿大学艾伦学院成立云计算技术中心

目前，艾伦学院在华盛顿大学建立了一个新中心，旨在催化下一代云计算技术。新建立的中心即 FOCI，作为艾伦学院的云基础设施，FOCI 将在学术界和工业界之间建立更牢固的合作伙伴关系，使基于云的系统在安全性、可靠性、性能和可持续性方面达到新的高度。

第一代云颠覆了传统计算，但专注于类似的工程建模，这是许多新技术的典型特征，现在云计算正处于更彻底转型的风口浪尖，该中心将通过培养研究人员和从业者之间更紧密的伙伴关系来解决新出现的瓶颈并探索新机遇，从而帮助开创一个新时代。

这种转变在很大程度上是由机器学习、边缘计算、5G 和其他新兴技术的兴起推动的。这些新工作负载的需求——包括为其应用提供动力所需的能量呈指数



增长——将要求研究人员从头开始重新思考整个计算堆栈。

当下的公司和消费者正在寻求以更低的成本在云中获得更高级别的安全性、可靠性和性能，不仅仅是金钱成本，还有环境成本。有一段时间，由于摩尔定律，研究人员在能源效率方面取得了进展，但现在收益已经放缓甚至逆转。例如，在美国，计算能源需求的增长速度是太阳能和风能的两倍。因此，研究人员需要全面考虑硬件—软件接口，以及如何使云计算具有可持续性、弹性和安全性。

艾伦学院的 FOCI 研究团队热衷于探索的领域之一是能源感知云计算，它可以在功耗和性能之间进行权衡，同时使云应用程序能够抵御中断。它的另一个潜在的调查途径涉及如何开发系统以有效管理在分解存储和新兴机器学习应用程序等环境中使用的各种硬件加速器，同时最大限度地减少延迟、确保公平和满足多维资源需求以及其他挑战。

同时，FOCI 中心还成立了一个技术顾问委员会，以应对这些挑战。该委员会由来自云计算公司阿里巴巴、思科、谷歌、微软和 VMware 的代表组成，这些公司都是云计算领域的重要推动者和推动者。他们的意见将有助于指导该中心根据当前趋势对现实世界的影响进行研究，他们可以预测在 5~10 年的时间范围内会出现什么问题，以及如何在实践中应用解决方案。中心研究人员将利用这些实用的见解来追求宏大的、开放式的想法，更好地利用系统、计算机体系结构、网络、机器学习、数据科学、安全等方面的专业知识。

行业界知道痛点和技术趋势，学术界擅长探索性协作工作，通过将两者结合在一起，FOCI 中心不仅会产生令人信服的解决方案，还将为构建下一代系统的学生提供教育。

参考文献：

[1] Allen School News. “The sky’s the limit”: Allen School launches new FOCI Center at the UW to shape the future of cloud computing [EB/OL]. (2022-10-14)[2023-06-02].<https://news.cs.washington.edu/2022/10/14/the-skys-the-limit-allen-school-launches-new-foci-center-at-the-uw-to-shape-the-future-of-cloud-computing/>.

案例

加州大学伯克利分校推出天空计算开源框架

如今，失控的云计算成本会扼杀机器学习和数据科学项目，许多组织出于不同目的使用多个公共云来节省资金。然而，多云方法会显著增加复杂性，因为并非每个人都是云基础架构专家。

1. SkyPilot——在多个云上运行机器学习

为了解决这个问题加州大学伯克利分校天空计算实验室的研究人员推出了 SkyPilot，这是一个开源框架，可用于在任何云或多个云上运行 ML 和数据科学批处理作业，具有一个与云无关的界面。

SkyPilot 使用一种算法来确定哪个云区域或服务提供商对于给定项目最具成本效益。该程序会考虑工作负载的资源需求（是否需要 CPU、GPU 或 TPU），然后自动确定哪些位置（区域/地区/云）有可用的计算资源来完成作业，然后再将其发送到成本最低的选项来执行。

该解决方案自动化了在云上运行工作负载的一些更具挑战性的方面。SkyPilot 的开发者表示，如果发生容量或配额错误，该程序可以可靠地配置集群并自动故障转移到其他位置，它可以将用户代码和文件从本地或云存储桶同步到集群，并且可以管理作业排队和执行。研究人员称，这会大大降低成本。

SkyPilot 开发者在一篇博文中表示，多云和多区域策略的增长趋势促使团队构建了 SkyPilot，并将其称为“跨云代理”。开发者指出，组织正在战略性地选择多云方法以获得更高的可靠性，避免云供应商锁定更强的谈判杠杆。

为了节省成本，SkyPilot 利用了类似硬件资源的云提供商之间的巨大价格差异。开发者举了 Nvidia A100 GPU 的例子，以及 Azure 目前如何提供最便宜的 A100，Google Cloud 和 AWS 对相同的计算能力收取 8% 和 20% 的溢价。对于 CPU，有些价格差异可能会超过 50%。

专用硬件也是货比三家的一个原因，因为许多云提供商现在为不同的工作负载提供定制选项。例如，Google Cloud 提供用于 ML 训练的 TPU，AWS 提供用于 ML 推理的 Inferentia 和用于 CPU 工作负载的 Graviton 处理器，而 Azure 提供



用于机密计算的 Intel SGX 代码。这些专用资源的稀缺也是使用多个云的原因，因为高端 GPU 经常不可用且等待时间长。

不管采用多云有什么好处，通常都会增加复杂性，伯克利团队在使用公共云运行机器学习、数据科学、系统、数据库和安全方面的项目时就体验到了这一点。开发者指出，使用一个云已经够难了，但使用多个云会加重最终用户的负担，而 SkyPilot 的开发人员旨在减轻这种负担。

积极开发了一年多，开发者表示，该项目已在伯克利天空计算实验室被 10 多个组织用于用例，包括 GPU/TPU 模型训练、分布式超参数转换和 CPU 现场批处理作业实例。开发者表示，用户报告的好处包括 GPU 实例的可靠配置、集群上的多个作业排队以及同时运行数百个超参数试验。

2. Skyplane——优化云提供商之间大型数据集的传输

加州大学伯克利分校 RISELab 的研究人员最近发布了 Skyplane，这是一种开源工具，用于优化云提供商之间大型数据集的传输，从而减少传输时间和成本。Skyplane 目前支持 AWS、谷歌云和 Azure 上的对象存储之间的传输。

开发人员认为，云中的数据传输缓慢且昂贵。使用典型的 CLI 工具（如 `aws s3 cp` 或 `rsync`）进行传输会导致传输速度低至 20MB/s（低于美国平均宽带速度）。此外，云传输非常昂贵。由于云的惊人出口费用，复制 100GB 数据集的成本可能高达 14 美元。

Skyplane 在云提供商之上创建了一个覆盖网络，因此它可以绕过拥挤的云网络链接。它还通过在许多流水线 TCP 连接和多个 VM 上条带化数据传输来利用 LZ4 压缩和并行性。

在过去十年里，伯克利研究人员优化了数据密集型工作负载性能的研究，促成了 Apache Spark 和 Ray 系统的产生。这些系统的涉及是基于数据集主要存在于单个云的单个区域中的情况。目前研究方面越来越多的瓶颈则是云区域和云提供商之间的数据传输。研究人员认为，跨任何云的数据传输都应该更快、更便宜和通用。

在成本方面，Skyplane 比现有工具便宜 3.8 倍。Skyplane 分析了跨区域的云网络成本和吞吐量，借鉴了麻省理工学院 RON（弹性覆盖网络）的思想来识别



跨区域和云提供商的最佳传输路径，并对最终用户透明。

例如，要在 AWS 的两个区域之间移动单个大文件，Skyplane CLI 命令是：

```
$ skyplane cp s3://srcbucket/64gb.tar.gz s3://dstbucket/
```

该语法类似于 AWS CLI，但在并行性和成本降低方面有显著改进。

Skyplane 并不是运行高吞吐量传输的唯一选项。开发人员可以使用 AWS DataSync 或 GCP Storage Transfer Service 等数据传输服务，但它们会收取服务费，并且通常经过优化以将数据移动到提供商的云中，缺乏可移植性。

Skyplane 目前仅通过默认为 AWS 的 awscli 和 gsutil 谷歌云来支持本地到云的传输，并支持 Oracle Cloud、IBM Cloud 和 Alibaba Cloud 的计划。

参考文献：

[1] HAMPTON J. UC Berkeley launches skypilot to help navigate soaring cloud costs [EB/OL]. (2022-12-12)[2023-06-06]. <https://www.datanami.com/2022/12/12/uc-berkeley-launches-skypilot-to-help-navigate-soaring-cloud-costs/>.

[2] LOSIO R. Open source skyplane targets faster and cheaper data transfers between clouds [EB/OL]. (2022-10-16)[2023-06-06]. <https://www.infoq.com/news/2022/10/skyplane-data-transfer-cloud/>.

罗马尼亚初创公司发布企业多云管理平台版本

目前，罗马尼亚初创公司 Kubeark 正在挖掘天空计算的潜力，部署可持续解决方案，为其产品寻找全球市场。这家初创公司最近发布了其最新的企业多云管理平台版本，努力满足各行各业组织的需求。

根据 Kubeark 的数据，85%的企业目前使用多达 5 个云平台，这使得天空计算成为实现跨不同云和基础设施的兼容性和互操作性的选择。Kubeark 认为这个领域的趋势实际上是人们想要从这些云中获得不同的功能和不同的东西。例如，谷歌在人工智能、机器学习等领域是最好的，微软的 Azure 则被广泛用于数据存储……因此最大的挑战是需要让人们在想要运行的云上接受培训。

天空计算将事情提升到一个新的水平。它的目标是减少人们在所有云中需要做的事情的复杂性和工作量，以便充分利用这些云。这就是 Kubeark 的用武之地，它提供了管理各种云、软件和基础设施的工具，并以一种高效的方法来管理各行各业的公司如何存储他们的数据。



在 21 世纪前十年的后期，直至现在，人们都倾向于认为一切都会迁移到云中。但事实上这是有前提条件的。有些东西会迁移到云中，当人们需要非常快速地联系世界各地的其他人时，此时需要大量的计算能力，但有些敏感的东西仍然会保留在公司内部，比如员工记录，或者学校记录等等。所以，现在有大量软件、云、流行的基础设施，这需要一个工具来实际管理一切，以便交付软件。

Kubeark 主要针对两大类客户。一是软件供应商或软件创建者，包括传统软件和新的机器学习模型创建者/平台。他们需要按照最终用户的需要分发他们的软件，想通过将一切都转移到 SaaS 中来解决这个问题，因此软件即服务，但在许多情况下，最终用户想要不同的东西。这就是 Kubeark 可以帮助他们的地方，包括在 SaaS 业务模型中，还有在客户端。二是跨银行、电信或医疗保健等行业的“深度技术大企业”。Kubeark 也有很多合作伙伴，但主要是来自云提供商的合作伙伴，Kubeark 帮助他们将所有这些工作负载带到他们的云中，还有外包商或为最终用户完成工作的人，以便维护软件、部署软件、使用软件，等等。

同时，Kubeark 还在寻找全球客户群。他们主要关注欧洲、美国、中东。同时，Kubeark 的发展重点不再是一定数量的客户，而是更多地开发产品并了解如何将其应用于各个行业。Kubeark 希望更多地接触某些行业的某些客户，以便了解如何将技术应用于一定的空间。

在人才方面，这家初创公司将目光投向了罗马尼亚，在那里公司找到了所需的人才，看到了日益发展的初创企业文化。罗马尼亚有越来越多的“创业”文化，尤其是开发人员，他们希望转向创业公司。这家初创公司目前拥有一支由 28 人组成的团队，另外还有大约 10~12 名顾问委员会成员。

2022 年第四季度 Kubeark 在由 Credo Ventures 领投的种子轮融资中筹集了 280 万美元。这进一步证明了投资者对公司和整个天空计算领域提供的解决方案的兴趣。

参考文献：

[1] FODOR S. Kubeark: The Romanian-founded startup's take on sky computing and sustainability [EB/OL]. (2023-05-25 [2023-06-02]. <https://www.romania-insider.com/kubeark-bogdan-nedelcov-may-2023>.

