

# 软件打包“容器”好用

## 集装箱化技术帮助科学家加速实验进程并提高可重复性

2015年,遗传学家 Guy Reeves 试图安装一个被称为“银河”的免费软件系统,以便使他的生物信息学项目取得进展。然而,在经过一两天的挫败后,他开始向 IT 部门的成员寻求帮助。他们帮其安装了 Docker——一种模拟电脑环境的技术。这使 Reeves 得以利用同他需要的任何东西一起打包的特殊版本“银河”。为此,Docker 也被称为容器。

来自德国马普学会进化生物学研究所的 Reeves 是采用这种容器的诸多科学家之一。随着科学越来越多地向数据密集型发展,更多软件正在被编写以便从这些数据中汲取知识。不过,极少有研究人员拥有时间和计算机技术来充分利用它。容器,即把软件代码以及运行它的计算环境放在一起的程序包能缩小这一差距。它们能帮助研究人员利用更加广泛的软件,加速实验进程,并且提高可重复性。

容器实质上是轻量级、可配置的虚拟机——操作系统及其硬件的虚拟版本,可以让软件开发者共享其计算环境。研究人员利用它们分配复杂的科学软件系统,并因此允许别人在和初始开发者利用的相同的条件下运行软件。在这种情况下,容器可消除计算生物学可变性的一个来源。不过,美国加州大学戴维斯分校生物信息学家 C. Titus Brown 介绍说,虚拟机器是相对资源密集型的,且没有那么灵活,容器则是紧凑、可配置的。尽管配置底层的容器化软件可能比较麻烦,但容器可被修改,从而根据用户的需求增加或移除一些工具。Brown 表示,正是这种灵活性提升了它们的知名度。

### 将任何东西打包

实验室制造的工具极少能拿来就用。它们通常以必须接受处理和配置脚本或者编程源代码的形式存在。大多数软件需要额外的工具和程序库,但用户可能并未安装。即便用户能获得软件并运行它,计算环境的不同也会无意中改变其性能并且影响可重复性。容器通过将运行按所需软件要求配置的计算环境的关键要素打包进一个轻量级的虚拟盒子,减少了这种复杂性。它们并未改变运行软件需要的资源——如果某个工具需要大量内存,那么它的容器也需要。不过,它们使得软件更容易被使用,结果也更容易重复。

取决于所使用的软件,容器可在 Windows、Mac OS X、Linux 上或者云中运行。Docker、Singularity 和 rkt 是比较流行的容器。它们可将来自单一流程到诸如“银河”等复杂环境的任何东西打包。这些工具能彼此交互,比如共享数据或者创建管道。由于每个应用都呆在自己的盒子里,因此即便是通常相互冲突的工具,也能和谐地运行。

Docker 利用的是被称为镜像的可执行软件包。它包括被包含的工具以及开发者的计算环

“5~10年内,科学家将不再需要担心下载和配置软件问题,工具也将被简单地容器化。”

图片来源:Project Twins



境。为创建 Docker 镜像,开发者需要创建配置文件以及如何下载和构建全部所需工具的指令。随后,他们运行该文件,以创建可执行软件包。用户需要做的全部事情是获取软件包并且运行它。

位于加州伯克利的国家能源研究科学计算机中心计算机专家 Deborah Bard 已帮助很多研究人员在实验室的超级计算机上安装了他们的软件。她回忆说,安装用于望远镜模拟和分析的复杂软件通道之前需要三四天。利用容器则将这一时间缩减到几个小时。“你可以把时间用在做科研上,而不是弄清楚编译器的版本。”Bard 表示。

### 获得容器

尽管计算机专家拥有容器平台的多个选择,但在 2013 年发起的开源项目——Docker 可能在科学家中最流行。它拥有大型预建容器注册表以及竞争者无法企及的活跃网络社区。不过,很多高性能计算系统的管理员会排除对 Docker 的使用,因为它需要高级别的管理访问权限才能运行。这种类型的访问可能允许用户复制或损害系统上的任何东西。一种基于付费的企业版本的附加组件使用户得以回避这一要求,但对于免费的社区版本来说,这种组件是无法获得的。不过,他们可以利用诸如 Shifter 等不要求全部权限或根访问但仍支持 Docker 镜像的不同的容器化工具。

Brown 解释说,对根访问的要求是 Docker 被广泛采用的最大障碍。很多学者在由其所属机构或政府管理的高性能计算集群上运行生物信息学工具。“当然,他们并未拥有大多数系统的管理权限。”虽然每年要为亚马逊网络服务提供的云计算时长支付约 5 万美元,但 Brown 表示,这仅占据了其计算工作的 1/3。其他的则在密歇根州立大学的计算集群中开展,而他并未拥有最高权限。于是,Brown 为自己的软件工具创建了 Docker 容器,但自己很少能利用它们。

研究人员可通过 Docker 平台自身的主机服务或者诸如 BioContainers、Dockstore 等容器寄存器,获取 Docker 镜像。曾是 Dockstore 寄存器技术主管的加州大学研究人员 Brian O'Connor 建议科学家浏览容器寄存器以寻找对其项目起作用的工具,而非试图重复发明已经存在的东西。

不过,旨在简化该流程的网络服务——“代码海洋”首席执行官 Simon Adar 表示,真正让底层 Docker 软件正确运行非常具有挑战性。“它太过技术化,并且原本是为开发者部署复杂系统而设计的。”在今年 2 月启动的“代码海洋”创建了被 Adar 称为“计算胶囊”的服务。它把代码、数据、结果和 Docker 容器本身包含在一起。研究人員上传他们的代码和数据,然后在网络浏览器上执行该软件,或者分享给其他人。Adar 将这一过程比作在网上共享视频。该公司甚至提供可使用用户在网页中嵌入可执行代码的小部件。

### 提高可重复性

参与欧洲核子研究中心 ATLAS 项目(寻找新的基本粒子)的纽约大学物理学家 Kyle Cranmer 介绍说,尽管围绕可重复性的大多数争论集中在数据和代码上,但计算环境本身也起到了很大作用。“这真的很重要。”Cranmer 表示。例如,一项对结构分析工具在不同计算环境下的性能进行的研究发现,操作系统的选择产生了相当大的影响。

不过,乔治·华盛顿大学机械与航空航天工程师 Lorena Barba 认为,只有封装的工具质量高,容器才会表现得更好。“如果研究人员将糟糕的代码塞进容器中并将其分享开来,我们注定会失败。”Brown 则表示,没有来自资助机构和期刊的压力,容器化技术本身不可能让研究人员突然欣然接受计算的可重复性。

伊利诺伊大学香槟分校统计学家 Victoria Stodden 表示,的确很少有研究人员在使用容器。部分原因在于缺少需求或者意识,他们可能并不具备相关计算机技能也是重要原因。

不过,这一切正在悄悄地发生改变。石溪大学生物信息学家 Jonas Almeida 介绍说,诸如谷歌、微软等公司已开始运行容器中的一些软件。大规模生物信息学项目也开始采用容器化技术。Almeida 预测,5~10年内,科学家将不再需要担心下载和配置软件问题,工具也将被简单地容器化。“这是不可避免的。”Almeida 说。

(宗华编译)

# 引力波“探路者”初战告捷

## 研究者希望能提前发射正式卫星

欧洲引力波猎人正在大肆庆祝。7月1日,一颗卫星完成了其技术测试任务,以探索在静止的太空中测量引力波。去年,该卫星的“表现”已经超出预期。而此次成功让欧洲空间局(ESA)批准给予全面感知宇宙重大事件的项目优先权,这些事件难以在地球上被发现。

这个名为“LISA 探路者号”的探测器发射于 2015 年年底,全名为激光干涉空间天线,是由 ESA 斥资 10 亿欧元设计的空间观测平台。其接下来的任务是跟踪宇宙中最大的天体——包括巨大黑洞的合并以及星系的碰撞——发生时产生的时空涟漪。

该项目科学家、ESA 天体物理学家 Paul McNamara 表示,它以 1000 系数的精度确定了目标,平息了那些对其潜力表示怀疑的批评声。“人们曾认为这是无法完成的任务。”McNamara 说。

当前的 LISA 计划于 2034 年发射,将是激光干涉引力波观测台(LIGO)的空间类似物。目前的美国 LIGO 由两条分别长达 4 公里并且互相垂直的干涉臂构成,首次检测到两个黑洞合并发出的引力波。而 LISA 探路者号由 3 个置于边长为数百万千米的三角形的三个顶点上的探测器组成,以便使其对更长的引力波较为敏感,例如由更大的黑洞碰撞产生的波纹。

最终的空间观测站将试图通过这 3 个 LISA 探测器在自由落体运动中漂移时对激光的反弹,发现空间的拉伸和压缩现象。因为这些物体将受到其他所有外部力量的保护,理论上只有引力波能扰乱它们自由落体的同步性——这一干扰将会影响激光频率。

LISA 探路者号的目标是展示这样的变化能在失重状态下被测量,精度可达 1 皮米。高精度推进器将调整 LISA 探路者号的路线,以便探测器内部的两个实验立方体能够自由下落,而不受



LISA 探路者号

图片来源:Manuel Pedoussaut, 2015/ESA

外部轨道运行的干扰。

这两个物体均为质量 2 公斤、用金和铂打造而成的立方体。当它们围绕拉格朗日点 1 (L-1) 运行并进入微重力环境时,探路者号将打开这些微型推进器。而这些推进器的推力十分微弱,需要 1000 个这样的推进器才能推得地面上的一张纸。它们将确保探测器在其上空盘旋,同时不让两个立方体接触到它的边缘。而它们在相对运动中出现的任何偏差都可以极其精确地被测量到。

英国伦敦帝国学院天体物理学家、带领团

队设计并建造该探测器保护机制的 Tim Sumner 指出,作为宇宙空间中的高精度实验室,LISA 探路者号和 ESA 此前的任何一项任务都不同。

2016 年 2 月,在 LIGO 宣布了首个探测结果 2 周后,LISA 探路者号也宣布取得了一个胜利。它虽然没有发现引力波,却表示能探测到为 1 皮米的 1/100 的微小运动。自那以来,探测器的性能被提升到另一个量级。

到今年 6 月初,LISA 探路者号几乎用完了所有推进器燃料,而地面指挥中心使用剩余燃料将其推离现有轨道,并进入最终的太阳轨道。

7月1日,LISA 探路者号停止收集数据,7月18日,它将永久休眠。

美国宇航局喷气推进实验室物理学家 William Klipstein 表示,LISA 探路者号取得了“胜利”。他参与了 LISA 研发,但未涉足 ESA 的探路者号任务。他提到,它的性能“消除了最后一个主要技术障碍,推动了 ESA 进行长期引力波项目”。

6月20日,ESA 科学项目委员会一致同意将 LISA 列为第三大科技项目。德国马普学会引力物理学研究所所长 Karsten Danzmann 表示,人们已经等待这一决定很长时间了,不过,LISA 探路者号的成功和 LIGO 发现引力波让该决定几乎不存在疑问。

虽然该决议并非最终版本,但它意味着产业合作者目前将参与到详细方案设计和成本核算中。一旦这些工作完成,ESA 将决定是否“采纳”该任务,并为其划拨经费。该项目组也希望美国能提供重要部件。美国曾是 LISA 任务的平等合作方,但 2011 年为减少开支,美国缩减了其参与度。

在进行 LISA 任务前,ESA 已经选择了其他两个大型项目:计划 2022 年发射的木星卫星探测器和 2028 年发射的 X 射线天文台。而 LISA 计划 2034 年升空。但 LISA 探路者号首席科学家、意大利特兰托大学物理学家 Stefano Vitale 等人则希望其发射时间能提前。

LISA 项目在 2000 年被 ESA 批准,然后开始征集计划书,Vitale 团队的设计脱颖而出,最终被该机构采纳。其最初计划在 2006 年发射,但时间一再推迟。“我认为以前完全低估了其中的难度。”Vitale 说,“但这也是我们建造 LISA 探路者号的原因。”

而 Vitale 等人希望提前发射 LISA 的原因之一是这样一来,它能在目前主要科学家退休之前开始返回数据。

(张章编译)

## 科学线人

全球科技政策新闻与解析

## 加拿大基础研究面临衰退风险



图片来源:Rebecca Schley

一份日前发布的报告称,由于资助缩减以及误入歧途的政府和拨款委员会政策,加拿大科学事业面临着沦为垃圾债券的风险。

报告作者特别谴责了资助重点从基础科学领域的逐步转移。这使得该国不断壮大的科学界只剩下极少数人聚焦基础研究。

“取消对基础研究的资助,改变了科学研究如何在加拿大开展的真正本质。”来自全球青年学会的报告称。它敦促加拿大政府在来年向该国 3 个拨款委员会注入至少 3.52 亿美元,并且呼吁让关于基础研究的联邦开支直接同该国科研生态系统中在职研究人员的数量联系起来。

在加拿大总理贾斯汀·特鲁多领导的自由党政府中担任科学部长的 Kirsty Duncan 表示,该报告有助于凸显开展关于科学和科学资助的全国性讨论的必要性。不过,她并不打算支持报告提到的任何特定建议,理由是解决这些问题没有权宜之计。“我们不得不现实一些。”Duncan 说,“在此前由史蒂芬·哈珀领导的保守党政府的管理下,科学已被掏空了 10 年,科学家也被裁员。不过,平衡科研界的需求和加拿大人民的需求是本届政府的职责。”

这份名为《重塑加拿大在基础研究领域的竞争力:长凳上的注视》的报告,给基础研究的粉丝带来了足够多的发人深省的发。

加拿大 3 个拨款委员会——自然科学与工程研究理事会、加拿大卫生研究院以及社会科学及人文研究委员会为基础研究提供的资助占总支出的份额从 2005 年的 67% 降至 2015 年的 48%。

和此项研究同时开展的对 1303 名加拿大研究人员进行的调查显示,2006~2015 年,“仅开展基础研究的研究人员比例大幅下跌,从 24% 降至 1.6%”。

报告首席作者、维多利亚大学生物学副教授 Julia Baum 表示,对基础研究的支持遭到侵蚀,已迫使科学家“缩小可被质疑的问题范畴并且降低其工作质量”。

(宗华)

## 也门将大规模测试霍乱疫苗



6月,在萨那的一家临时医院,也门儿童被怀疑患上霍乱。图片来源:AFP/Getty Images

试想在一个面积和西班牙相当、饱受战争蹂躏的贫困国家,有 2000 余万人受到一种在城市间快速传播的致命疾病的威胁。你可以支配 100 万份疫苗。你会试图保护谁?

这是在也门迅速扩散的霍乱疫情作斗争的公共卫生专家和国际小组面临的问题。近日,一个管理全球霍乱疫苗适度储备的小组决定向也门分发 100 万份疫苗。这几乎占到该小组现有存货量的一半,疫苗接种则定于 7 月初开始。这将是针对霍乱疫苗的最大规模测试之一。不过,世界卫生组织(WHO)霍乱专家 Dominique Legros 表示,在哪里展开此项测试仍处于争论中。

医务人员在大面积紧急事态中处理这种疫苗——包含被杀死的细菌并且只能口服——的经验非常有限。与此同时,全球储备仅在 4 年前创建,并且主要被用于较小规模的行动。“如何利用疫苗阻止疫情扩散?我们仍在学习中。”来自全球疫苗和免疫联盟的 Melissa Ko 表示。该机构为霍乱疫苗储备提供资助。专家已决定通过给予一个剂量而非推荐的两个剂量舒缓疫苗供应问题,但从哪里开始接种仍在讨论中。一些人警告说,由于疫苗的疗效有限,即便是供应充足,也可能仅对大规模疫情产生有限的影响。

霍乱只是降临到也门的最新悲剧。自 2015 年起,效忠于总统 Abdrabbuh Mansour Hadi 的部队开始同反叛者作斗争。后者自 2012 年被废黜的前总统 Ali Abdullah Saleh 结成同盟。他们控制了该国首都萨那以及北部地区,Hadi 则控制了南部地区,包括港口城市亚丁。双方的冲突导致上千人丧生,摧毁了该国的大部分基础设施,并且令卫生系统陷入瘫痪。据 WHO 估计,1450 万人无法正常获取到清洁饮用水,从而为霍乱弧菌创造了理想环境。

也门的霍乱病例在今年 4 月开始急剧增加,据估计已达到 20 万例,导致 1300 人死亡。其中,大部分人口位于该国西部地区。目前,每天有 5000 起新发病例。

(徐徐)