

网格、云、SaaS、SOA、WebX.0……，其本质是互联网计算

韩燕波 王桂玲 刘晨 王菁

中科院计算技术研究所

关键词：互联网计算，云计算，网格计算，SaaS，SOA，WebX.0

近年来，云计算已经成为国内IT领域最热门的概念之一，一时间吸引了工业界、学术界、行业和个人用户的广泛关注，像是又发现了新大陆。回头看看可以发现，在IT领域这样的现象还是很常见的，概念一个个闪过，激起一波波热情，但并非每个都能带来实质性的进步。迄今，已经舶来了很多概念和英文缩写，大都描绘了很美好的愿景，其中也不乏厂商推销用语。与云计算紧相关的概念有很多，包括：效用计算（Utility Computing）、网格计算（Grid Computing）、元计算（Meta Computing）、按需计算（On-demand Computing）、软件即服务（SaaS）、平台即服务（PaaS）、基础设施即服务（IaaS）、数据密集型超算（DISC）、互联网数据中心（Internet-based Data Center）、资源池化（Resource Pooling）、资源虚拟化（Resource Virtualization）、虚拟组织（VO）、面向服务的体系结构（SOA）、面向资源的体系结构（ROA）、自主计算（Autonomic Computing）、对等计算（Peer-to-Peer Computing）、普适计算（Pervasive Computing）、Web2.0、语义网（Semantic Web）、Web服务、SOAP、RESTful服务、Mashup、OpenAPI、MapReduce等等。这些概念往往体现不同视角（理想和愿景视角，系统和平台视角，标准化视角，构造和使用视角，商业模式视角，产品和工业界具体实践视角等等），但多半并没有精确的定义，起源也不尽相同。其科学性、合理性、还有背后的本质问题和原理常缺乏考究和归纳。特别是，在实际应用价值这一至关重要的指标上，常常因为盲目拓展了应用面或领域，或因为对实现的周期、过程和复杂度估计不足，难以及时得到实用，频频会让人们有失所望。相当多的研究和探索工作只是围绕这些概念的解析和技术实现展开，忽视实用目标和实际价值，先搭建概念和技术体系再寻找应用，难以取得重要突破。那么，云计算又是一波概念而已呢，

还是的确存在巨大潜力？只是工业界噱头还是确有特定的科学问题和核心技术？其应用价值及学科根基何在？现在的成功案例往往是问题特定、领域特定的，离广泛实用还有多远呢？这些是我们亟待思考的问题。本文是作者做这方面思考时产生的随笔，在此和读者分享，肯定存在不妥和偏颇之处，欢迎批评指正。详尽的讨论请见《互联网计算的原理和实践》¹一书。

人类社会和科学技术都在快速发展，各种新概念层出不穷。笔者认为，光解释和比较概念意义不大，概念之间肯定存在区别，即使同一概念，随着时间的推移，内涵和外延都会发生变化，重要的是能够透过纷繁的表面现象，把握事物的本质。在 IT 领域，很多新概念往往是标识一组目标、做法或思路以及相关技术的整体概念（Umbrella Term）。这类概念难以精确定义，会随人们的认识、环境和边界条件以及技术的发展而演变。以云计算为例，人们最常问到的问题之一就是，它和网格计算真有本质区别吗？由于起步的应用领域、客观条件和主观认识都不同，仅仅拿网格计算的早期工作与云计算的当前发展相比较没什么意义，更重要的是对其本质问题和共性基础的认识。两者的愿景没有本质区别，基本思想均来自早期的效用计算。即使没有云计算的发展，网格也迟早需要考虑服务提供和多租户等问题。同样，如果工业界同行广泛沿用网格概念的话，先锁定在更“亲民”的特定问题域并在其基础上拓展 SaaS 能力，我们可能正在面对商业网格、数据网格、服务网格和个人网格等。

说到底，网格、云、SaaS、SOA、Web X.0... ..兴起的根本动力是互联网计算。互联网(Internet)即广域网、局域网及单机按照一定的通讯协议组成的逻辑一体的计算机网络。今天，连接在互联网上的计算设备、存储设备能力有了大幅提升，数据资源呈指数级增长，互联网上的各种服务资源日益丰富，互联网的经典使用环境——万维网（Web）也已经不再是单纯的内容平台，而是朝着提供更强大、丰富的用户交互和体验能力方向发展。互联网（包括正在兴起的移动互联网）和 Web 已成为构造、运维、使用各类分布式应用系统不可或缺的基础环境，正在演变为迄今人类最大的计算平台。

互联网计算的一个重要目标是构建使能资源优化利用、跨管理域资源共享和应用集成的 IT 基础设施，把分散、无序的资源变成为有序的、逻辑一体化的共

¹韩燕波 王桂玲 王菁 刘晨，互联网计算的原理和实践，科学出版社，2010.3

享资源，并按社会基础设施（类似水、电）的商业模式运作。互联网计算即是网格计算、云计算、SaaS、SOA 等各种计算模式的“超集”，也是它们的共性支撑。

两项重大需求促动了互联网计算：一个是 IT 资源的优化利用，另一个是跨管理域的资源共享和应用集成。

- 当前 IT 行业有一种普遍的现象：硬件性能快速提升、成本不断下降；软件相应的发展速度较为缓慢，但其占用的硬件资源却越来越多；用户体验没有得到合理的提升。几十年来，硬件系统的发展始终遵循着摩尔定律预测的速度，每十八个月，CPU、存储、网络带宽等 IT 产品的性能会翻一番。这些提升的硬件性能被越来越复杂的软件占用。以个人计算机为例，主流操作系统占用内存、CPU 和硬盘资源越来越多，但是软件的功能通常并没有本质的提升。在行业应用领域，平台软件十分庞大、复杂，重复投资现象严重，企业在 IT 上的全部投资，仅用到了其中很小部分的功能；用户购买的上线新系统中，往往很多功能模块用户在其它系统中已经付费过。软件从业人员众多，系统换代频繁、版本众多，IT 系统已经成为浪费严重的“烧钱机”。人们把这种现象戏称为“安迪-比尔定律(Andy and Bill's Law)”：硬件的进步被以微软为首的软件开发商吃掉，牟取利益；膨胀的软件又迫使用户升级机器，以英特尔为首的各硬件厂商只有不断提升性能，才能获取利润；硬件提升带来的好处又进一步被新的软件吃掉。在这个循环过程中，最终对用户造成了极大的浪费。这种现象的背后其实是一种以经济利益为导向的“资源消耗型”产业发展模式。针对“资源消耗型”发展模式，人们不禁会思考一个问题：硬件的性能会无限制的得到提升吗，摩尔定律会永远有效吗？假如硬件性能提升的空间是有限的，20 年之后 IT 行业会不会面临资源枯竭的危机？我们何时才能够走出上述怪圈？站在整个 IT 产业可持续发展的角度来看，另外一种思路是，可否摆脱以经济利益为导向的“资源消耗型”发展模式，以用户为中心审视当今软件业进步的瓶颈，思考如何恰当、充分地利用硬件和网络基础设施发展带来的好处，提高 IT 行业的资源利用率。互联网计算是一条出路，其社会价值极其巨大。
- 跨域集成问题是一个困扰人们多年的难题，经历了从点到点集成到基于

软总线的集成平台的发展历程。随着互联网技术的发展以及应用的深化，应用软件逐渐转移到互联网这一开放、动态、难控的网络计算平台上进行开发和运行，跨域集成面临着更大的挑战。随着 SOA、网络化构件、SaaS、Web2.0 等新兴互联网应用架构、模式与技术的发展，互联网上的资源越来越多的以服务形式对外提供，通过对网络上封装各类资源的服务的共享和集成来构造和支撑应用软件正逐渐成为一种新兴的、主要的方式，在互联网之上正逐渐形成一个以聚合和协同为核心的计算环境和互联网软件新形态，体现出分布、开放、动态、快速演变、个性化等特征，并使能用户参与和聚众效应。

上述发展在体系结构、系统边界、运营方式、管控原理和使用模式等方面对分布式系统带来一些明显甚至是质的变化，可伸缩性、可用性、在线演化等特征被进一步突出，会形成不同于传统分布式系统的新的技术、方法和理论体系，为网格、云计算的根基和学科基础。相对于各种昙花一现的概念“噱头”，发展互联网计算才能保障云计算等的根本价值体现。下面，让我们从不同侧面来讨论互联网计算的特点：

- 从科学问题角度，泛在、开放、异构、自治、多样化、不确定性和动态变化等是互联网计算固有的特征，这些注定了互联网计算与传统分布式计算的本质不同。人们对互联网内在规律、机理及其可利用的价值等方面的探索方兴未艾。技术的进步以及应用的激增和普及暴露出很多新的现象和矛盾。例如，人们已观察到，当相关联的互联网资源集合达到一定规模后，会带来聚集效应，也即能够产生新的增值能力和吸附更多相关资源的能力随之得到加强。有效地利用这些规律反过来又会进一步推动技术创新、应用创新和社会进步。对互联网计算背后本质问题和学科基础的探究有重要的意义。互联网计算重点关注和研究的问题包括：互联网基本抽象和计算模型；互联网分布式系统的体系结构；互联网环境下数据、计算、服务等资源的建模、虚拟化、管控及性质保障相关理论；互联网资源空间的划分和社区模型；互联网服务的基本构成和演化机理，核心抽象和演算的（数学）模型；编程模型和应用模式；此外，安全、信任等方面的问题也是互联网计算必须考虑的基本问题。

- 从相关学科基础的角度，中间件技术和分布系统工程、构件技术和软件工程、信息检索技术和互联网信息服务等领域的发展也极大地促进了互联网应用发展。特别是，面向服务的计算（SOC）的发展标识了分布式系统和软件集成领域技术一个里程碑式的进步。服务呼应了人们对据有自治、开放、平台无关等一些期待的特征的网络化构件抽象，而面向服务计算方式倡导关注分离、多方参与和松散耦合的系统设计思想，可使分布式应用具有更好的复用性、灵活性和可增长性。服务计算为互联网计算提供重要支撑。
- 从系统和基础设施的角度，把互联网看作为一个巨大的虚拟计算机的话，它的“操作系统”等系统软件该是什么样子呢？互联网上有那么多节点，能否指定和建立一些像局网中的数据库服务器一样的专用节点呢？还有，已有的很多计算中心在互联网环境下该如何生存和发展呢？这些问题都可归结为互联网计算基础设施相关问题。如同我们人类社会许多其它领域一样，规模的壮大促动了标准化和基础设施建设的发展。实际上，早在上世纪 60 年代，图灵奖获得者 John McCarthy 就提出了效用计算的愿景，即像用水、电一样获得“按需计算”的能力。这种愿景一直是 IT 领域的一大诉求，随着网络和信息处理技术的快速发展，该愿景会离我们越来越近。不论是早期的 ASP（Application Service Provider）还是近年流行起来的网格计算和云计算，其实质追求都无出其右。总结和归纳互联网计算基础设施相关的原理和方法比单独讨论某一类特定模式更有助于理解和把握关键问题。
- 从应用的角度，在“互联”、“开放”、“共享”和“协作”为主旋律的互联网计算环境下，软件呈现出网络化、服务化、虚拟化、集成化的发展趋势，应用系统的形态也在发生质的变化，不再以固化、独有的形式出现，会包含越来越多的“不为所有，但为所用”的服务构件。这些变化为突破困惑了人们多年的软件生态链条不够优化、成本高昂等问题带来机会，也凸现出若干科学问题。互联网计算与传统分布式计算的一个重要区别就在于公用服务和运营，出现了软件即服务及平台即服务为代表的新兴服务供给模式。当前，怎么利用互联网平台做计算和构造软

件是仍是一个有巨大意义的挑战性问题的。

- 从设计、技术实现和性质保障角度，互联网计算需要寻找更适合的工程方法和技术。例如，对互联网计算来说，可伸缩性的保障并不能完全照搬传统分布式系统的做法，这是因为，在互联网计算中，对系统进行扩展可以采取的方法更加复杂和丰富、对成本的计算与采用软件即服务的运营和使用模式相关、系统的“大小”也是一个多维、复杂的度量属性等。传统的系统一旦建好便很少变化，但是对于互联网分布式系统来说，开放性和动态性是其固有的特性，服务的规模、用户的需求都难免不断变化，因此，在系统设计之初，特别是对那些对可用性要求比较高的互联网分布式系统来说，就必须将这些变化的因素考虑在内，以便在运行时响应开放环境和需求的动态变化而进行升级和动态调整，以免系统因自身升级原因而造成重大损失。高的可用性和可靠性是任何计算机系统所追求的性质，而对软件即服务模式下的互联网分布式系统来说，由于租户程序的运行环境几乎完全依赖于运营商的系统，一旦发生系统停机不可访问的意外情况，对租户造成的经济损失是巨大甚至是无法估计的，因此，互联网级分布式系统对可用性和可靠性的要求更为迫切和严格。

网格、云、SaaS、SOA、Web2.0 的本质是互联网计算。虽然从名字上看，云计算只是比拟而来，Web2.0 过于模糊，SaaS 体现商业模式，SOA 不够达意，但其优势在于生逢其时。使能互联网计算的客观条件逐渐成熟，这包括互联网本身、网上资源、普及程度、标准化程度、工业界和学术界的共同关注、网格等前期相关工作研究基础和相关技术发展、人们主观认识的变化、以及需求迫切性（尤其是，与贯彻标准、提升效率、降低计算成本和落实绿色 IT 有关的需求）等诸多因素。只要抓住问题本质，同时注重实际应用，相信互联网计算蓬勃发展之日已为时不远。

韩燕波，博士，中科院计算技术研究所研究员。主要研究兴趣为分布式系统、互联网服务、业务流程管理和协同以及跨管理域的行业应用集成。
邮件地址：yhan@ict.ac.cn。



王桂玲，博士，中科院计算技术研究所助理研究员。主要研究兴趣为分布式系统、软件集成、服务组合和Mashup技术等。邮件地址：wangguiling@software.ict.ac.cn。



刘晨，博士，中科院计算技术研究所助理研究员。主要研究兴趣为分布式系统、互联网服务、语义网。邮件地址：liuchen@software.ict.ac.cn。



王菁，博士，中科院计算技术研究所助理研究员。主要研究兴趣为分布式系统、面向服务的计算以及跨管理域的行业应用集成。邮件地址：wangjing@ict.ac.cn

