TechTarget 中国

亚马逊 Web 服务功能概览手册

亚马逊 Web 服务功能概览手册

根据 Gartner 去年后半年发布的报告来看 ,目前为止 ,亚马逊 Web 服务(AWS)仍旧是基础架构即服务(IaaS ）世界的主宰者。AWS 系统化的解决了主流企业应用面临的问题，2013 年的确很精彩，竞争者不断逼近，但是 AWS 仍旧占据绝对优势。AWS 现在对于开发合作伙伴生态环境想法多多。市场上有个特别有意思的比喻，说“亚马逊就是金刚，但是现在是金刚和他的小伙伴们。”现在基本上所有的主要的 IT 服务提供商和主要 IT 厂商都是充满了吸引力的合作伙伴候选。现在很多客户也认为如果你有一个精通 AWS 技能的人才，做事情就会快得多，在云端也会更加如鱼得水。

在这本技术手册中，我们将对亚马逊 Web 服务发展进行回顾，同时提供一些数据处理的功能和应用，以及其最新的功能应用，帮助读者对其有一个初步的了解，希望对大家有所帮助。另外如果您有任何意见或者建议欢迎致信 zhangpeiying@techtarget.com.cn。

亚马逊云产品概览

亚马逊 Web 服务在云计算世界的霸主地位导致了很多思考，比如未来的云世界会是什么样子？虽然很多人认为亚马逊 Web 服务将会模仿过去的微软和 IBM 这样的企业，我个人觉得并不会如此。虽然普遍的观念认为亚马逊将会主导大部分细分领域，但是其他的云提供商会在这些领域发现有利可图的生态环境。这部分中我们将简要回复亚马逊云产品的发展过程以及目前的价格战。

- 亚马逊云产品十年路
- AWS 首席数据科学家谈云价格战和大数据

亚马逊云数据处理

亚马逊并不像其他竞争的基础架构即服务提供商那样，亚马逊非常了解市场，知道并不能够仅仅是快速的分配虚拟机这么简单。一些 IaaS 提供商觉得整个市场更像是管理化的托管市场，只是在底层有不同类型的基础架构而已，所以他们也不知道从何开始自己的战斗。这部分中我们将介绍亚马逊 Web 服务的数据处理部分。

- Amazon Redshift：本地数据仓库替代品
- 亚马逊 Kinesis：何时使用亚马逊全新大数据处理服务
- 亚马逊数据流 PaaS：AppStream 和 Kinesis
- 资源密集应用开发利器：亚马逊 AppStream
- 云端大数据安全利器：亚马逊 DynamoDB 和 Accumulo 访问控制
- 亚马逊 Web 服务超级用户论战 DBaaS

亚马逊云服务的魅力

最让人感到欣慰的是 AWS 并没有倚仗过去的荣誉停滞不前，这家提供商立即开始在其核心的平台上增加新的服务和创新，从而解决早起采用者一些不断发展出现的问题。亚马逊一直都是云市场上的创新者，一直致力于解决客户的新问题。一些其他的竞争对手，尤其是小型厂商还在纠结下一步如何发展。这部分中我们将关注一些其他的功能，包括部分新发布的功能。

- 论亚马逊 WorkSpaces 的必要性
- Amazon S3 加密概述：如何确保 Amazon 云数据安全性
- Amazon 云：用 StarCluster 实现服务器集群管理自动化
- 亚马逊 CloudSearch 优于 DIY 搜索工具？
- AWS Auto Scaling：云服务成本与性能平衡利器

编者：亚马逊云产品十年路

十年前，亚马逊 Web 服务（AWS）并不是亚马逊公司发展蓝图上的重量级想法，而在当今的基础架构即服务的世界里，AWS 已经冠绝群雄。2004 年 11 月，亚马逊发布了第一款 Web 服务，这是一款称之为 Simple Queue Service 的消息队列产品，但是直到 2006 年，亚马逊才正式将 AWS 作为一项业务发布。现在，仅仅八年之后，亚马逊的云产品涵盖了计算、数据库、网络、支付、存储、应用和其他的服务。亚马逊的子公司运营在全球 190 多个国家，服务于无数的客户，拥有诸多数据中心。

亚马逊的客户从美国宇航局到诺基亚，从辉瑞制药到图钉网，从 Dropbox 到道琼斯股票，无所不涵盖。美国总统奥巴马 2012 年再次参加竞选也依赖于亚马逊云产品，而且该公司最近同 CIA 签署了一项交易。网飞公司在 AWS 上运行了差不多整个业务，尽管网飞的流媒体服务直接同亚马逊自己的视频产品竞争。2012 年 4 月，亚马逊发布了 AWS Marketplace，这是一个在线商店，现在为客户提供了来自其他厂商的 1100 多种 AWS 相关的软件产品。

同时，客户还是觉得不够。2013 年 11 月，在 Amazon re: Invent 会议上，参与票卖到脱销，吸引了 9000 名与会者来到拉斯维加斯参会 ,据统计还有 9000 名来自 57 个国家的参会者通过流媒体的方式参与此次会议。

亚马逊高级副总裁安迪·杰西告诉与会者，尽管亚马逊处于不断上升的阶段，但是 AWS 还会将自己看做一项年轻的业务。他随后发布数个新的 AWS 服务，最有名的就是 Amazon WorkSpaces，这是一款虚拟桌面服务，也标志着 AWS 进入了一个全新的企业级 IT 市场。会后，《华尔街日报》采访了安迪·杰西，标题为《Meet the Man Who Really Runs the Internet》。

AWS 吸引力

人们通常会将亚马逊云产品的普及归因于：亚马逊的品牌力量、AWS 的规模和范围、AWS 按需付费业务模式的灵活度和简易度。但是对于大多数客户而言，最大的吸引力就是价格。

这些吸引力可以直接转化为成本节省。2012 年，市场调研机构 IDC 调查了 11 个 AWS 客户，从而来判定使用亚马逊云产品对这些企业的长期财务影响。IDC 发现：平均而言，五年期投资回报率为 626%，投资回收期仅七个月。这些企业也表示平均五年的总体拥有成本节省 72%，而且赢得了生产力和正常运行时间。

但是对于其他客户，AWS 的真正价值是可扩展性。备受瞩目的案例即 Airbnb，这是一家在线房屋租赁市场，于 2009 年在 AWS 上发布，现在已成长为拥有 1100 万用户的企业，而且每天平均 15 万交易量。

背后的故事

那么 AWS 赚了多少钱呢？亚马逊可没公开说过；该公司从来没在其财报中明确表示 AWS 的收益。也许这种情况会发生变化，因为一些分析师认为，亚马逊最终会将 AWS 剥离，使其成为一个独立的公司。同时摩根斯坦利的分析师斯科特·德维特最近预估了 AWS 的当前市值，约为 250 亿美元。其他的分析师也预计 AWS 的市值在未来几年将会从 500 亿美元变为 1000 亿美元。还有一些分析师，包括亚马逊的杰西预测 AWS 可能最终会让亚马逊的零售业务黯然失色，亚马逊的零售业务在 2013 年第三季度的吸金 170 亿美元。

然而，亚马逊仍旧在纠结于盈利能力，这和它的历史有关系。在同样的第三季度，亚马逊丢失了 9%的市场占有率（410 万美元），而且和其预期的一样。该公司在 2013 年第四季度做的好一些，但是仍旧没能达到分析师的预期。分析师将这种现象称之为“亚马逊悖论”，尽管表现不佳，但该公司仍旧在吸引投资者。亚马逊的股票价格在 2013 年创造了两个新纪录，7 月份超过 300 美金一股，12 月份达到 400 美金一股。

这也并不是说 AWS 一直顺风顺水。一开始，该公司和其客户都遭受了数个重创，包括广为人知的 2012 年圣诞前夕发生的宕机，直接让网飞的流媒体服务受到影响。

随后 AWS 价格一路削减，从 2006 年到现在不少于 40 次。最近，AWS 仍在削减亚马逊简单存储服务和亚马逊弹性块存储的价格。虽然一开始降价是个大新闻，但是持续降价也就意味着不再具有差异性。

毫无疑问，AWS 必须时刻保持警惕。VMware 2 月份在英国发布了其 vCloud 混合服务，EMC 也期望其成为主要的云服务提供商，但是并不是直接同 AWS 竞争。但是不到一周之后，VMware 宣布雇佣 AWS 高级技术布道者西蒙尼·布鲁诺兹作为 VMware 的新的副总裁以及混合云的首席技术官。

在最近 Gartner 发布的 IaaS 魔力象限中，研究人员发布了数个亚马逊云产品的警告，比如 AWS 的费用在每一个可选项目中是分开的。然而，研究人员也表示 AWS 目前为止仍旧是市场翘楚，其市场占有率仍旧非常之高，而且过去几年中，AWS 也一直致力于改善一些明显的缺陷。虽然市场上有很多出彩的竞争对手，但是 AWS 仍旧是市场老大。

（ 作者：TechTarget 中国高级编辑 张培颖 ）

AWS 首席数据科学家谈云价格战和大数据

亚马逊 Web 服务近年来频繁变动，从价格削减到紧密整合私有数据中心，再到调整企业市场，而且不断扩展其原来的开发者和创业公司客户。

在众多执行者中一马当先的就是 Matt Wood，他是亚马逊 Web 服务（AWS）的数据科学总经理。SearchCloudComputing 本周在 AWS 峰会上和他探讨了企业客户的热点问题。这里我们将谈谈云价格、云联盟、法规以及数据定位。

TechTarget 云计算：云价格是当下的热点话题，谷歌和亚马逊都在本周大幅降价。在每个人都可以免费得到一切之前价格究竟会有多低？

Matt Wood:我们一直都知道， 和我们的零售业务有一点像，云计算是一种大容量、低利润的游戏，而且这是一种我们非常适应的业务模式。

如果你回顾过去的八年，我们降价 42 次，这样做并没有任何真正的竞争压力。降价只是我们所做的一部分，是我们的组织发展的脉搏的而一部分，而且我们处于良性循环中……有越多客户采纳这个平台的地方，他们就使用得更多，而且因为我们能走出去和我们的厂商进行客制交易，我们走出去，并且利用规模经济，总体上我们最终是节省了成本。我们还可以从中获利。这也是一种合情合理的事情。但是我们选择将这些成本节省回馈给客户……我们一直在做这些事情，而且未来我们还会这样做。

TechTarget 云计算 :你的专长领域是数据科学和大数据分析。在这个领域里你是否看到了一些新的趋势？

Wood: 最大的趋势之一就是增益，而非取代，但是传统商业智能的增益则伴随更加实时的服务。而且是二者同时变得更加强大。

芬兰游戏公司 Supercell 就是个很好的例子……他们运作流行移动游戏，比如部落战争（COC），而且他们一天中有八百万人在 ios 上玩这个游戏。理想上，如果你是个游戏公司，你希望尽可能地捕捉价值。你想知道人们如何同游戏世界交互。你想知道你的游戏竞技表现如何。你想知道谁买了什么 ,以及谁和谁进行了对话，在什么情况下退出了游戏。你可以利用这些信息改善游戏。

通过收集所有数据有些已经实现了。你收集什么并没有限制，使用亚马逊实施管理流服务 Kinesis，就是一种实现途径，你可以将数据丢进去，并且以不同的样本率连接传感器，，用同样的数据流可以做不同的事情。

TechTarget 云计算：亚马逊将自己描述为“客户导向”。那么哪些特性和服务是客户目前寻求的呢？

Wood:他们会询问类似这样的事情，‘是否易于访问高价值、公共数据集？’这就要求我们要做很多。因为有很多数据。我们花费大量的时间识别，且同公共的可用数据工作，并且使其易于使用。

Common Crawl 就是个很好的例子，它会定期更新，是一个非常大的 Web 集，Web 上的每一页、下载和预计算，放到一个主页上，这就让 Hadoop 很容易运行。你不必自己来抓取，必须将原素材进行预计算所有的标签，并且移除 HTML 和类似的东西，这些都已经做好了。

你所得到的就是格式化的数据，易于用分布式的方式使用。你可以在冷启动中在不到十分钟的时间里查询数十亿网页。我们存储且托管这个数据是免费的，因为对于整个社区是有益的，然后我们确保其遵循了最佳的简单对象存储（S3）访问实践，所以这也很容易加速大型 Hadoop 集群，并且运行查询。

TechTarget 云计算：我们接触的一些客户认为云联盟很有前途，亚马逊如何看待这个问题呢？

Wood: 我们目前还没从客户那里听到这样的说法，但是这并不代表未来这不重要。我们从一些更大企业机构那里听到就是已经在基础架构上进行了大量投资。他们已经有了蓝图。我们和这些人谈论的时候，我们尝试引导他们并非一种选择，可以本地运行或者是在 AWS 上运行一切。我们过去的十八个月都在构建集成点，使其更易于让用户在有意义的地方运行工作负载。我们在他们的数据中心和我们的数据中心之间构建了直接的链接，我们提供私有存储选择，我们也有私有计算选择，我们识别了联合选择以及类似于 WorkSpaces 这样的事物 这是一种在后端同活动目录集成的事物。所有的这些集成点都帮助客户在其工作负载上做出正确的选择。

TechTarget 云计算：一些客户处于法规遵从原因需要将数据保存在某个特性区域，亚马逊能够签署法定协议，保证客户的数据不会离开具体的可用区或者具体的区域吗？

Wood: 实际上，你如果不选择数据存在哪里是没办法使用我们的平台的。客户必须做出谨慎的决定，确定其数据将会存在的区域。我们有这样的特定区域，每一个区域都有多种可用区，而且可用区内部都有数据中心。以 S3 服务为例，我们通过可用去得到镜像数据，但是我们不会在区域间做镜像数据。

原文链接：http://www.searchcloudcomputing.com.cn/showcontent_81368.htm

(来源：TechTarget 中国　作者：Beth Pariseau　译者：张培颖)

Amazon Redshift：本地数据仓库替代品

Amazon Redshift 是一个由 Amazon 网络服务（AWS）推出的数据仓库服务，尽管它可能会吸引一些用户，但是请注意它不同于企业内部版。作为企业内部数据仓库的一个替代产品，Redshift 深得用户的青睐，特别是在用户了解了其鲜明的服务特色，并将其用于提升业务优势之后更是如此。

Amazon Redshift 把数据仓库应用提升到了平台即服务（PaaS）产品的高度。这个数据仓库服务是基于 PostgreSQL 定制版的，这是一个具有其竞争对手商业关系型数据库管理系统所有功能的开源关系型数据库。多年以来，关系型数据库一直都支持服务器集群，但是其早期版本是难以实施和管理的。

Redshift 旨在解决过去被强加给数据库管理员（DBA）数据库集群那令人沮丧、费时的挑战的。数据库管理员使用 Amazon 控制面板来创建最多达 16 个计算节点的集群，其中每个节点都配置有 2TB 或 16TB 的存储器。

Redshift 用户可实现更低的存储成本

Redshift 是一个柱状的数据存储，因此当数据被存储在磁盘上时，它们是按列而不是按行进行排列的。这样就减少了当根据列选择数据时所需的输入输出操作数量，例如选择上月所有销售额大于 10000 单位的产品，它还允许实现更高效的数据压缩，从而最终实现用户存储成本的降低。

与所有畅销商品一样，Amazon Redshift 的价格也颇具吸引力，具体为 1000 美元每年每 TB。不要感到惊讶，这有利于这款产品迅速占领市场。1000 美元每年每 TB 的成本相当于在 2TB 节点上运行的预约价。如果你运行着一个小型数据仓库，那么你可能采用单节点的数据仓库。只有 2TB 的实例（即被称为 dw.hs1.xlarge 的服务器）适用于单节点配置；16TB 的实例（dw.hs1.8xlarge）是为集群保留的。

除了存储数据和执行查询的计算节点之外，你还需要一个群首节点。群首节点从客户端接收查询、制定运行计划、向计算节点发送查询并收集查询计算结果。Amazon 只根据计算节点进行收费；群首节点是不收取费用的。

定价是基于虚拟机规模的。2TB 节点的请求定价为每小时 0.85 美元，而 16TB 节点的价格则为 6.80 美元每小时。目前，Amazon Redshift 可供美国东部、美国西部和东欧（爱尔兰）等区域的用户使用。

保留实例可以降低你的成本，但是用户从 AWS 直接购买需签订为期一年或三年的使用承诺书。用户也可以通过 Amazon 市场的另一个客户处购买。销售实例的客户自行定价，并确定市场上所提供机器实例的类型。

由于 Amazon Redshift 刚刚问世，你可能还无法马上找到很多的数据仓库实例。另外，如果你能够总是在你的合同中销售所有未用和不必要的机时，那么你可能会决定购买一个保留实例。

数据仓库节点的价格包括了计算节点上的存储成本以及用于备份应用 Amazon 简单存储服务（S3）上的等量存储资源。如果你在你的数据仓库中存储了超过存储量的数据，那么你会需要对超出的存储资源按标准 S3 价格支付费用。

通过 Amazon Redshift 进行数据维护

Redshift 的性能恰与其并行运行的能力一致。在集群中查询是跨节点分布的，因此每个节点都会完成整个工作量中的一小部分。不要因为跨节点的数据分布而错误估计了工作量。默认情况下，Redshift 将使用 Round-Robin 算法来实现集群中跨所有节点的数据分发。如果你选择基于关键节点的数据分布，那么你应当仔细选择这个关键节点以避免在节点子集中的瓶颈。

所有的数据库都需要一定程度的维护，而 Redshift 将会执行一些最常见的维护任务，其中包括执行备份操作和为软件打补丁等。数据库管理员们将仍然需要监控与数据库设计和数据负载相关的性能表现。RedShift 通过在表中删除现有的列和添加新的列来执行更新操作。这将有助于提升运行性能，但也会导致存储碎片的出现和增多。数据库管理员需要定期运行 VACUUM 命令以便于回收未使用的存储空间。数据库管理员还需熟悉用于检查查询执行计划的 ANALYZE 命令，这是分析查询运行缓慢原因的一个关键技术。

原文链接：http://www.searchcloudcomputing.com.cn/showcontent_73594.htm

(来源：TechTarget 中国　作者：Dan Sullivan　翻译：滕晓龙)

亚马逊 Kinesis：何时使用亚马逊全新大数据处理服务

亚马逊最近发布了 Amazon Kinesis，这是一种针对大量数据快速处理的管理服务。你对于这项服务有什么看法，在哪些地方最有用？

Amazon Kinesis 是实时处理数据流的一个管道。类似于 Apache Storm，但是完全由亚马逊 Web 服务托管和扩展。Kinesis 提供了完整的几近实时的数据处理解决方案（十秒以内）。Kinesis 并没有真正提供什么新概念，但是这个管理版本可以看做是具体用例的完整套件。

一些案例包括：

- 实时处理日志数据（发现可能的错误并在问题发生时向 IT 人员发送警报）
- 接收应用用例的实时分析
- 有人在 Twitter、Facebook 或者 Google+提及具体的公司时，设置实时通稿警报
- 使用具体的关键词针对内容监控实时新闻源，然后将这个内容交付给移动设备

Kinesis 所做的这种类型的处理与其他亚马逊服务所能做的处理一样。你已经使用了亚马逊的简单队列服务、简单通知服务和自动扩展容量到处理实时数据流（实际上，我的公司也在这么做）。亚马逊 Kinesis 真正的优势在于能够更轻松地从头开始构建新的服务，为整个流程提供完整的管理服务。

最终，亚马逊 Kinesis 对于处理大数据问题非常有用，比如：

- 处理日志数据
- 针对具体术语或者关键短语处理社交媒体流
- 股票价格趋势识别
- 分析实时销售统计

即使你没有大量数据要处理，这种类型的架构能帮助处理容错以及扩展。因此有很多因素可以来使用亚马逊 Kinesis，或者只要你担心任何数据流的实时处理问题都可以。

原文链接：http://www.searchcloudcomputing.com.cn/showcontent_79661.htm

(来源：TechTarget 中国　作者：Chris Moyer　翻译：张培颖　)

资源密集应用开发利器：亚马逊 AppStream

亚马逊 Web 服务最近公布了亚马逊 AppStream 的通用版本，这项服务可以让开发者构建复杂的应用。哪些应用能够从 AppStream 开发中获益？对于 AppStream 开发是否还有别的建议？

亚马逊 AppStream 类似 AWS Elastic Beanstalk。但是 Web 应用不是目标市场，AppStream 的目标市场是桌面应用，尤其是游戏。

AppStream 旨在让你构建基于 Windows 的游戏，这些游戏使用高性能的图形库（比如 DirectX 和 OpenGL），都要求快速的图形处理单元（GPU）、高内存或者其他高度资源密集的处理。AppStream 旨在让你编写一个代码基，就可以让应用运行在 Mac、Windows 或者移动设备上。

随着移动游戏的不断增长，开发者面临的最大问题之一就是尝试覆盖到市场上的每一种可能的设备。虽然这些设备共享不同的屏幕大小、操作系统以及硬件规格，但是拥有相当同意的用户体验需求。

移动游戏的另一个大问题在于如何能够让游戏的大量计算或者图形显示逻辑适应设备，而且至少需要 512MB 随机访问内存（RAM）。同时，开发者希望完全利用设备，这就需要十倍的 RAM（比如高端平板电脑）。AppStream 将图形和计算工作负载卸载到云端，你可以构建你的应用，然后在终端用户的设备上进行渲染。

如果你的应用并不如 Web 应用做得好，你应该只使用 AppStream，比如：

- 跨平台(包括移动)游戏
- 多人游戏
- 图形密集游戏(比如图形设计程序)
- 你希望使用 Windows 库构建的其他跨平台应用

你不应该使用 AppStream，如果：

•一个简单的 Web 应用就能工作

•你需要支持离线模式

•你不想在应用中使用基于 Windows 的库

•你的应用不需要密集图形或者 CPU 或者网络利用

AppStream 的主要缺点：需要完全的连接。不幸的是，蜂窝网络并不提供完全的融合，而且如果用户离开这个范围，就不能使用这个应用。AppStream 适合永远在线的额设备。也会不管你做什么都是用数据，因此用户必须有一个更高级的数据计划，才能充分利用它。折中的做法就是开发者要确保其用户得到最佳体验，不管用户在什么设备上。

AppStream 可能并不适合每一个人，但是可以肯定的是适用于游戏领域。如果你正在构建下一个伟大的游戏，可以看看 AppStream 是否能够帮助你获得更多的受众，扩展你的平台支持。

原文链接：http://www.searchcloudcomputing.com.cn/showcontent_81377.htm

(来源：TechTarget 中国　作者：Chris Moyer　翻译：张培颖）

亚马逊数据流 PaaS：AppStream 和 Kinesis

云计算的未来很可能是在增加托管服务的基础上，通过 RESTful 访问网络友好的应用程序编程接口，从而实现特殊的提升以解决企业用户和开发人员的问题。亚马逊公司的缓存技术、数据库管理软件即服务以及虚拟桌面服务都是这一发展趋势的一种具体表现，这些技术或服务可让开发人员部署机器镜像，而这些机器镜像绝非是简单的基础设施即服务（IaaS），它们是整合了基于云计算的特殊功能的镜像。AppStream 和 Kinesis 就是这样的两个服务，它们的目的就是为了让云计算成为完全不同的东西——即可以支持庞大的数据流和具有挑战性的图形用户界面演示的应用程序。另外，AppStream 和 Kinesis 都是具有重大现实意义的，但同时它们也是极具吸引力的。结合其他的亚马逊公司网络服务（AWS）发展，它们将有可能改变云计算的游戏规则。

在线数据生成并不是大数据的唯一形式，但是它是具有特殊意义的一种数据形式，因为数据源都是分布的而且由于数据源众多的原因数据量是潜在庞大的存在。在很多应用程序中，开发人员被迫开发大规模的汇聚网络来收集数据，开发弹性处理框架来处理数据，同时调整适应数据量的变化，进而同时实现整个架构的低延迟和高可用性。这是一个相当艰巨的任务，因为任何公司都可能会有一个数据流密集型的应用程序。

亚马逊公司的 Kinesis 就是创建了一个可以调集弹性网络服务以处理分布式或单一流式大容量数据流的网络服务。Twitter 是最经常被提出的例子，但是在企业用户的眼中它可能会降低 Kinesis 的使用率。几乎所有的金融贸易应用程序，大部分的大规模事务处理，以及特别是机器-机器和互联网数据传输等应用都适合采用这一数据流模式。

Kinesis 让开发人员能够定义任意数量的数据源点，然后可以把这些数据源与以一种弹性的方式托管在亚马逊公司云计算的处理相关联。源点和主机处理之间的联系就是一个流，一个流可以被定义包括任意数量的源点，并与任意数量的处理相关联。通常情况下，为实现较高的可靠性，用户可以跨亚马逊公司的可用区域来复制这些流。这样做的结果就是出现了一个供开发关键数据流应用程序使用的架构，而这个架构对于除了公共云计算以外的任何云计算模式都完全没有任何实际的意义。因此，Kinesis 需要买家和亚马逊公司跳出“迁移至云计算”的模式而进入一个开发云计算的模式。

显而易见，Kinesis 并不是买家们能够在财务上轻松负担的一项服务，但是它却是专为需要实现预期数据处理性能和高度弹性特性的应用程序而设计开发的。与专用的私有基础设施相比，Kinesis 可能是一个便宜货，尤其是当数据容量的可变性相当高时。甚至有应用程序使用 Kinesis 作为从网络商店提高后端交易的框架或者

一组进行交易处理系统的网络应用程序。随着 Kinesis 的进一步发展，它将会以更简单的形式、更低廉的成本和更广泛的应用程序出现。一个关于公共云计算的大真理在于，即便是在单个应用程序和用户使用承诺较小的情况下它也能够非常高效地使用资源。当亚马逊公司做好准备时，Kinesis 将是一个大众市场的工具。

亚马逊公司的另一个服务，AppStream 似乎是其 Kindle 接口在内容渲染亚马逊使用的产品化。几乎所有的游戏应用程序、众多的视频制作、甚至图形模拟和显示应用程序所共同面临的挑战之一就是，它们的运行非常非常地依赖于高性能图形处理单元。这类技术很少被用于大部分的商务笔记本电脑、平板电脑或者智能手机。AppStream 在应用程序和设备之间的云计算中创建了一个代理，这个代理可发挥图形处理单元（GPU）的图形渲染功能，并把渲染计算结果发送至一个简单的显示数据流，而这个数据流将是几乎所有现代计算机、平板电脑或者智能手机都能够轻松处理的。

AppStream 最明显的优势就是它能够让图形计算密集型应用程序在不具备特殊 GPU 工具的设备上正常运行。当然，对于游戏应用这是相当重要的，而且这一优势可能会再次打动企业，这事实上可能就是亚马逊公司的最大客户们。在医疗保健与金融行业中，通过行业工程与设计，以及其他设计数据图形渲染或图形描绘的电路和结构，AppStream 可以为平板电脑和智能手机打开使用图形应用程序的大门。

AppStream 的第二个优势就是，它可以把数据传送至在多个平台上运行的客户浏览器或应用程序，而不需要修改底层的应用程序。虽然亚马逊公司提供的材料让这一连接并不明显，但 AppStream 很可能与亚马逊公司的 WorkSpaces 虚拟桌面结合以满足企业用户需求，或者甚至与 WorkSpaces 以及 Kinesis 一起对大型复杂数据流进行虚拟化。在诸如医疗保健和金融贸易这样的行业中，这样的组合的叠加价值就将较为明显了。

也许，这就是 Kinesis 和 AppStream 的真正价值所在。亚马逊公司正在努力打造一个部署在云计算中的软件服务社区，并将其提供给用户和开发人员用于开发他们自己的应用程序。这使得亚马逊公司已经超越了云计算；它是在为开发人员进行应用程序开发创建了一个分布式的软件操作系统。随着时间的推移，这不仅会鼓励更多的亚马逊公司合作伙伴参与到他们的网络中来，而且这也使得亚马逊公司成为了一个独特的基于云计算的应用程序平台，而不仅仅是另一个 IaaS 厂商。当然，这似乎是亚马逊公司的长期计划。

原文链接：http://www.searchcloudcomputing.com.cn/showcontent_80162.htm

(来源：TechTarget 中国　作者：Tom Nolle 翻译：张培颖　)

云端大数据安全利器：
亚马逊 DynamoDB 和 Accumulo 访问控制

随着云计算中多个用户访问机密信息以便于执行数据分析工作的情况越来越多，安全性控制也变得日益重要了。NoSQL 数据库是大型数据集应用中最受欢迎的数据库选择，但是 NoSQL 数据存储的早期版本并未重视数据安全性方面的功能。目前，大数据管理员们不仅可以利用 NoSQL 的优点，而且还仍然保持对某些能够访问云计算中数据子集人员的控制权限。

根据数据库中数据的不同模式，IT 管理员们和云计算专家们能够实施各种不同粗细粒度的 NoSQL 访问控制措施。在你无意中泄漏个人数据之前，可以考虑采用 NoSQL 的内置数据访问控制措施。

确保 NoSQL 数据存储的安全性有三种不同的方法，而其中每一种方法都有其各自的优缺点：Accumulo 基于单元的访问控制、亚马逊 DynamoDB 使用 AWS 的身份访问管理（IAM）策略以及 MarkLogic 的隔离间控制与执行权限。

Accumulo 数据存储

Accumulo 是一个基于谷歌 Big Table、采用键值的分布式数据存储。这个开源选项是由美国国家安全局开发，并于 2011 年发布的。Accumulo 是一个可在 Hadoop 环境中运行的 Apache 项目，它具有 Big Table 中所没有的功能，其中被包括了基于单元的访问控制。

Accumulo 的键包括了一个指定安全性标签的可见属性，例如管理员、财务或管理员等。因为每一个键都与一个值相关，相当于关系型表格中的一行，所以基于键的访问控制也就限制了用户能够查询或操作的行的集合。然后，用户就会被分配指定某种安全性标签的授权，它可被组合在逻辑表达式中以便于创建所需的访问控制。例如，一个财务部的经理就会被分配“经理”和“财务”两个标签。

应用程序和数据库管理员们将确定标签组以及是如何根据企业的安全性政策具体实施它们的。

用户授权就存储在受信任的第三方授权服务器上。当一个应用程序执行操作（例如一次查询或一次更新）时，用户授权就会从第三方被找回并发送回 Accumulo。因此，当执行一个操作时，将把用户授权集与数据单元进行比较，那么对于特定用户所能访问的行集合是有限的。

亚马逊 DynamoDB

亚马逊的 Dynamo 数据库（DB）是增长速度最快的亚马逊网络服务（AWS）产品之一。DynamoDB 是一个提供自动化的可扩展性和配置 IOPS 的键-值数据存储服务。对于开发人员和更喜欢使用一个托管服务来管理他们自己的 NoSQL 数据库的应用程序管理员来说，亚马逊的 DynamDB 是一个不错的选择。对于已经投入时间和资源建立 IAM 策略的 AWS 用户来说，这是特别有吸引力的，因为他们可以精细地控制对保存在亚马逊 DynamoDB 中数据的访问。

为了在亚马逊 DynamoDB 中保持细粒度的访问控制，管理员们必须在 IAM 策略中指定条件。条件可以允许或拒绝对键-值数据存储中特定项和属性的访问。这一模式限制了对特定值或行的访问，例如与特定客户账户相关的数据，这样一来客户就只能查看他们自己的数据了。它还允许应用程序管理员们为访问特定属性而定义规则。例如，一个策略可以指定只有甘愿而非顾客可以查看数据库中的一个类别属性。

MarkLogic

基于文档的 NoSQL 数据库（例如 MarkLogic）可以扩展基于角色的访问控制以便于把角色和文档归入不同的隔离间。隔离间可允许访问控制检查和标准的基于角色的控制。如果一个文档被分配了一个隔离间，那么只有被分配了相同隔离间的用户才能够访问这个文档。

MarkLogic 还提供了对对操作执行的访问控制。这个数据库包括了一组可处理数据管理、安全性以及其他管理操作的预定义执行权限。数据库管理员们能够通过创建执行权限来控制执行查询的能力。在查询的定义中包含了一个执行权限，并被分配给角色。只有被分配了拥有必要执行权限的角色的用户才能够执行查询操作。

原文链接：http://www.searchcloudcomputing.com.cn/showcontent_80503.htm

(来源：TechTarget 中国　作者：Dan Sullivan• 译者：滕晓龙)

亚马逊 Web 服务超级用户论战 DBaaS

高级亚马逊 Web 服务用户更喜欢自我管理运行在亚马逊弹性计算云上的数据库，而不是数据库即服务产品，至少现在看是这样的。

上周，AWS 超级用户在线活动群组创立会议的演示中，关注超级用户如何在 AWS 上运行数据库。大多数演讲者表示他们在弹性计算云（EC2）上运行类似 Cassandra 和 MySQL 这样的自我管理数据库，而不是使用亚马逊的数据库即服务（DBaaS）平台，比如关系型数据库服务（RDS）以及 DynamoDB。

然而，一些 IT 专家在此次活动中也表示有过 DBaaS 体验，而且一些仍旧在自我管理和 DBaaS 选择中保持中立态度。

美国加州一家提供在线社交学习平台的公司 Edmodo，在将其 MySQL 操作从 EC2 上自我管理实例转移到 RDS 时，收获颇多。该公司的运营总监 Jack Murgia 表示："在我们决定从 RDS 退出时，我们学到了更多。"

2011 年春天，Edmodo 为两百万用户服务，而且得到了风投公司的重大投资，他们将这笔资金用来雇佣 Murgia，从而同拥有十个开发者的团队共同工作。Murgia 介绍："基本上我走过一扇门，就有一个数据库，一个熟练员工和一个苦工。"这些都运行在 EC2 上。Murgia 进来后，人员配置上并没有一个数据库管理员。

跟着亚马逊 RDS 一路走来，其提供了一次使用 MySQL 管理繁忙的初创企业的机会。2011 年秋天，该公司完成了到 RDS 的迁移。Murgia 谈到 RDS 时说："我们能够通过点击一些按钮，创建开发和质量保证环境，随着那年秋天负载的不断增加，用鼠标点几下就可以读取副本，改变域名系统记录。"

但是在多有效区域故障恢复的时候，RDS 部署遇到了障碍。

"我们发现多 AZ 故障恢复在绝大部分时间都是失败的，" Murgia 说，"有时候即便是计划中的故障恢复我们发现复制都是失败的，而且那个时候唯一的选择就是提出新的副本。"

主数据库有八个副本，每一个新的副本用时大约一小时，这也意味着 Edmodo 再一次服务于用户之前，要有一整天的宕机时间。因此公司重组了，准备转到 RDS 上一个单独的熟练员工，如果有什么失败了就要计划

新的副本。RDS 在 2013 年六月开始提供服务水平协议（SLA），这也让 Edmodo 寻求一种继续使用服务的方式。

但是随着 Edmodo 不断发展，该公司从外包公司引入了 DBA，在 2011 年到 2013 年雇佣了更多的系统管理员。那时候，该公司拥有了内部运行 EC2 上自服务数据库的技能，逐渐离开 RDS，转向自服务 MySQL 环境。

Murgia 说："我们的双手被 RDS"黑盒"绑架。"如果 Edmodo 管理自己的 MySQL 和副本，IT 团队可以促进副本到一个精通的员工，将所有的副本指向这个精通的员工并且重新启动并运行。相反，该公司致力于恢复数据库时，没有基础设施的控制就会出现失败。

"这是一种不得不做的妥协，" Murgia 说道，"可能你不具备技能，可能你只是一个小的初创公司，但是随着你开始获得这些技能，并且开始提升性能和可用性标准，这就会成为问题。"

超级用户活动会议上的另一个演讲来自 Stackdriver 的 IT 专家，这是一家位于美国波士顿的公司，提供 AWS 监控即服务。这家公司通过 Cassandra 集群迎来了一个转折点，而且考虑了两种替代方案：扩展现有的集群或者部署亚马逊 DynamoDB DBaaS。

"我们有非常繁重的工作负载，涉及数以亿计的数据点，而且 Cassandra 对于各种写操作过多的工作负载有很好的支持，" Joey Imbasciano 说道，他是 Stackdriver 的云平台工程师，"Cassandra 中的建模时间系列数据设计模式也是众所周知的，因此我们知道我们不会有任何问题。"

Cassandra 另一个吸引人的特性就是能够以编程的方式删除数据，这样就可以让数据库保持一种可管理的规模，而且无需人工介入。Stackdriver 也考虑了 MySQL 和 RDS，但是感觉 NoSQL 更适合自身的数据集。该公司也在部署 Cassandra 的 18 个月前就关注 DynamoDB。

"那时候，厂商锁定是我们尽力去避免的，" Imbasciano 说道，"此外，我们做了一点成本估算，并且发现那时候使用 Dynamo 的成本要稍高一点。"

Stackdriver 开始是三节点的 Cassandra 闭环，现在已经增长为 36 节点，随着其继续增长，该公司会再一次关注 DynamoDB。"优势很明显，" Patrick Eaton 说道，他是 Stackdriver 的架构师，"焦点就是自动化。升级是自动化的。亚马逊的全天候支持人员处理浙西额事情，他们可以在你需要时进行扩展。"

Eaton 补充：“此外，我们看到 AWS 一直在削减价格，因此从常量的角度来看，我们的价格实际上会随着时间的推移而变得更加便宜。”然而，该公司的 Dynamo 部署时间序列数据第一次部署还是要比继续使用 Cassandra 更贵。

“成本模型相当复杂，基于这些抽象的工作量，他们称之为‘写单元’和‘读单元’，这是一种请求率和数据规模以及一致性模型的结合，”Eaton 说，“原型阶段很难评估这种模型的持续成本。”根据 Stackdriver 的计算，Cassandra 持续管理价格为每月大约 3000 美元。主要集群成本为每月 12500 美元。在目前的 Cassandra 部署中，更小的集群预警成本大约为 1300 美元。

使用 Dynamo 作为主要集群，Stackdriver 的预估成本大约为存储和写单独计算 22000 美元。另一方面预警集群 DynamoDB 成本约为 600 美元。Eaton 表示：“成本节省或者成本实际取决于工作负载类型，不能在一种综合的状态中对比这些替代选择。”

截至新闻发布亚马逊未发表任何评论。

原文链接：http://www.searchcloudcomputing.com.cn/showcontent_81131.htm

(来源：TechTarget 中国　作者：Beth Pariseau　翻译：张培颖)

论亚马逊 WorkSpaces 的必要性

很多企业已经报告，支持员工使用个人电脑的成本要远远高于维护个人电脑本身所带来的成本。虚拟桌面基础设施（VDI）或桌面即服务都是针对集中化应用程序进行托管从而降低这些相关支持成本的，但是这些技术似乎已被移动设备和 BYOD 政策所压倒，同时用户也希望享受到云计算所提供的更多应用程序和资源灵活性。现在，亚马逊的 WorkSpaces 有望成为正确的答案。

从历史上来看，虚拟桌面应用程序都只是在服务器上创建了一个固定的“个人电脑实例”，并将其连接至一个在远程 PC 上的瘦客户端。在很多方面，这与服务器整合中的虚拟化方法相当类似，而且它确实具有提高硬件使用率和实现服务与支持集中化的优势。虽然虚拟桌面的应用是逐年递增的，但是其发展速度最快的时期被认为是在 2007-2010 年期间。似乎从当前模式中受益的大部分用户已经有所转变。

VDI 应用步伐放缓的最大原因可能就是可移动性。用户转移至智能手机和平板电脑、转移至不被传统 VDI 所支持的平台，因此很多用户在进行工作期间就不得不定期地在这些移动设备和传统个人电脑之间来回转换。为了对整个用户和软件的支持成本产生更大影响，企业需要一种超越桌面应用的虚拟化方法，一个可容纳所有技术选项的方法。

云计算还是 VDI 应用中的一个因素，因为很多公司都把桌面虚拟化和服务器虚拟化联系在一起，而现在他们会更多地把后者视为云计算迁移的一次机遇。云计算提供了资源的弹性和潜在的弹性改善，这些都是标准的 VDI 所难以做到的，而且它还很好地配合了可移动性的 IT 趋势。

这就是催生亚马逊的 WorkSpace 的综合原因。它是一个融合了云计算应用程序的元素、亚马逊绑定的云计算服务以及一个把远程用户设备与托管“设备实例”相链接的新方法的基于云计算的应用程序托管战略。有些人可能会把 WorkSpace 视为“云计算中的 VDI”，但是事实上远非如此。需要特别指出的是，它是一个移动世界中的 VDI。

首先也是最重要的一点，与重点关注 Windows PC 的传统 VDI 不同，WorkSpaces 的设计初衷就是与客户端无关。虽然 WorkSpace 是基于 Windows Server 的并包含了传统的办公应用程序，但是可以通过定制的 WorkSpaces 客户端把它们链接至各种各样的设备。这些设备协调了应用程序的 GUI（实际上就是指 Windows 7）和客户端设备的具体细节。

亚马逊使用了 Teradici 公司的 PcoIP 协议在云计算中 WorkSpaces 实例与客户端之间创建了一个安全的链接，同时因为这个协议仅仅只承载了 GUI 数据而不是底层的应用程序数据，还有安全信息本来就不太暴露，以及采取了针对应用程序的加密和身份验证（通过 Active Directory）保护。

数据压缩和客户端连接只承载 GUI 数据的事实意味着，WorkSpaces 的性能通常是良好的。虽然有一些报告表明，分支机构连接速度可能太慢以至于无法处理在那里产生的服务请求，用户更可能发现拥堵的公共 WiFi 存在着问题。这些连接性问题也将影响到任何其他的 VDI 或远程数据访问应用程序。

在其最初形式中，由于受限于 Windows 应用程序，WorkSpaces 还体现出了一个显著的优势，即通过 Android 和 iOS 客户端以及通过 Windows PC 支持 Windows 应用程序。对于在他们日常活动中混合使用平板电脑和 PC 的用户来说，这可能是一个真正的福音。最大的红利可能还没有来到，因为亚马逊公司将在两个方向上提供 WorkSpaces（而且还是规划中的东西）。

一个就是使用更多的亚马逊云计算服务以便于加速 PCoIP 连接。亚马逊选择在云计算中为 Kindle 使用网页预处理，而这个功能很可能会被用于一个自定义的 Kindle 客户端以便于让 Kindle 成为一个更好、更具有竞争力的平板电脑平台。

事实上，有很多人都认为 WorkSpaces 在很大程度上是瞄准着为 Kindle 平板电脑开发出一个业务市场。这就意味着，其他的 AWS 服务（包括了内容缓存和流程处理）能够在之后被整合在 WorkSpaces 中。

第二个方法就是生成连接至客户端的 PCoIP 连接，以便于令其成为一个任何云计算应用程序的演示界面而不仅仅是一个 Windows 应用程序。这意味着，亚马逊将提供一个高效的云计算托管应用程序架构，并可通过 WorkSpaces 客户端被用于融合的移动设备或者（潜在的）任何的笔记本电脑或台式电脑。

这将会允许开发人员开发与平台无关的应用程序的能力，并通过全方位的移动和固定硬件把它们提供给企业，从而进一步巩固亚马逊在企业级市场中的巨头地位——这是一个平台供应商而不仅仅是一个云计算供应商。

现在 亚马逊在他们的虚拟 CPU 性能和安装的软件中提供了四个包(例如“增加的”包中包括了 Office)。用户可以自定义地在他们的 WorkSpaces 中安装他们自己的 Windows 服务器兼容软件。与所包含的软硬件成本相比，包的价格是极具竞争力的。

亚马逊很有可能会随着时间的推移而进一步深层次地开发这些包，例如增加其他深受用户欢迎的软件包。这个包供应的方式也是亚马逊整合其他 AWS 平台服务的方式，并最终把 WorkSpaces 发展成为兑现其设计承诺的“云计算中的应用程序架构”。

亚马逊并没有承诺对 WorkSpaces 做成任何具体的改进，但这样的改善很快来到似乎是板上钉钉的事了。目前产品对 Windows 的关注并没有将其与来自于诸如 Citrix 或 VMware 这样的供应商的其他 VDI 战略区分开来，它并没有充分利用亚马逊云计算的力量。从来没有一个能够错失市场潜力，亚马逊似乎也不太可能发布诸如 WorkSpaces 这样的东东而不计划让其成为毋庸置疑的领导者。

原文链接：http://www.searchcloudcomputing.com.cn/showcontent_80289.htm

(来源：TechTarget 中国　作者：Tom Nolle　翻译：滕晓龙)

Amazon S3 加密概述：如何确保 Amazon 云数据安全性

随着越来越多的企业使用公共云和混合云部署，同时越来越多的敏感数据被存储在云服务厂商（CSP）的环境中，因此企业也在不断地积极寻求更好的方法来保护他们在云中的信息。当然，企业所采用的最普遍的控制措施之一就是他们已经习惯于使用的方法：加密。

在云中对数据进行加密意味着数据拥有两种状态：传输状态和存储状态。

Amazon 公司的简单存储服务（S3）是目前最著名的云存储服务之一，它能够整合 Amazon 公司其他的云功能和产品。它还提供了多种加密功能，企业用户可以使用这个功能用于保护存储在 S3 环境中的敏感数据。

本文介绍了 S3 中所提供的加密类型，以及使用这些功能的相关方法。此外，我们将把 Amazon S3 所提供加密方法与其他云厂商的同类产品进行比较，并总结出几条在 S3 或其他云存储环境中使用加密技术的要点。

Amazon S3 加密简析

在云中对数据进行加密意味着数据拥有两种状态：传输状态和存储状态。首先，为了在 Amazon 环境中对用于发送和接收的数据进行加密，S3 允许用户通过 HTTPS 协议进行连接。这是云厂商们所提供的一个相当标准的选项，所有的云厂商都需要支持基于 SSL 加密技术的连接以便于保护在传输过程中的敏感数据。

保护存储状态的数据则是另一个问题了；很少有公共云厂商会提供加密功能的支持。Amazon 公司实际上向 S3 用户提供了两种加密方法以保护存储状态的数据。其中较简单的一个方法是服务器端加密（SSE），该方法允许 Amazon 在它的基础设施中管理加密密钥。SSE 所采用的高级加密标准（AES）使用了 256 位的密钥，这被认为是一个安全的密钥长度。Amazon 会使用一个唯一的密钥对每个 S3 对象进行加密，然后使用一个主密钥来对这个唯一的密钥进行加密，而这个主密钥是定期更换的（通常至少为每月一次）。

针对一个特定 S3 对象建立 SSE 是可选的，它也可在单个对象水平轻松建立。有一个“一刀切“的政策要求所有发送至 S3 的数据都必须进行加密处理。具体示例如下：

{

```
"Version":"2013-05-17",
"Id":"PutObjPolicy",
"Statement":[{
    "Sid":"DenyUnEncryptedObjectUploads",
    "Effect":"Deny",
    "Principal":{
      "AWS":"*"
    },
    "Action":"s3:PutObject",
    "Resource":"arn:aws:s3:::SensitiveBucket/*",
    "Condition":{
      "StringNotEquals":{
        "s3:x-amz-server-side-encryption":"AES256"
      }
    }
  }
]
}
```

为了成功地把任何数据发送至这个 S3 中，要求必须包括有“x-amz-server-side-encryption”这样的文件头信息。可通过 Amazon 公司基于 REST 的 API 以及使用 Amazon 的软件开发包（SDK）（其中也包括了可实现相同功能的 API）完成开发工作。客户也可通过标准 Amazon 网络服务管理控制台委托实现 SSE 功能。在 2013 年五月，Amaozn 宣布该公司的 Elastic MapReduce 大数据分析服务正式使用了 S3 SSE。

在 S3 中对存储状态数据进行加密的第二个选择是用户使用 Amazon 公司提供的客户端加密工具来创建和管理他们自己的密钥。采用这种方法意味着在把数据发送至 S3 之前就已经完成了数据加密工作。客户端加密可使用 Amazon 公司的 Java SDK 进行部署，尤其是 S3 加密客户端，它使用了一个被称为“信封加密”的方法。客户端创建一个一次性使用的对称加密密钥来对数据进行加密；然后使用用户自己的密钥来对这个密钥进行加密。然后这个被加密的“信封密钥”与加密数据被一起上传至 S3 ,其中密钥作为元数据也被存储在 S3 中。

云厂商加密的比较

我们已经详细介绍了 Amazon S3 中所提供的加密功能，但是这些加密功能与其他云存储厂商所提供的加密功能相比，孰优孰劣呢？很多云存储厂商都坚持遵循了相同的加密标准，但其中大部分都没有达到 Amazon S3 加密那种程度的灵活性。例如，Rackspace 在它的云备份产品中提供了服务器端 256 位的加密功能，而 Dropbox 和 SpiderOak 也都提供了 256 位的 AES 加密功能。SpiderOak 则稍有不同，它总是使用客户端加密方法（也被称为零知识的安全性）而不是更为传统的服务器端加密方法。

其他大部分的主要云厂商们（其中包括了 Verizon Terremark 和 Savvis）也向他们的云存储客户们提供了数据加密功能。Terremark 为备份和冗余业务提供了自动加密功能，为托管平台与数据提供了多种托管和协管加密功能，在其 CloudSwitch 混合云中提供了客户自我管理的加密产品。Savvis 在它的云存储加密功能中使用了 SafeNet，向客户公开了密钥管理和 API 集成。

对于比较不同云厂商加密功能的企业来说，这里有几点需要考虑：

•确定所有的云加密选项都尽可能支持基于标准的加密方法和最高的密钥长度/强度（AES-256 是行业标准）。

•确定厂商是否提供了访问加密功能的 API，因为这可能是软件即服务（SaaS）和平台即服务（PaaS）环境和应用程序集成的关键要求。

•确实厂商是否同时支持服务器端和客户端的密钥管理。虽然在客户端加密方法中密钥管理的责任在客户自己的手中，但是它更适合大部分具备安全意识的企业。如果厂商只支持服务器端的加密方法，那么就必须考虑应如何管理和保护内部密钥。

•了解其他可能可用的加密选项，例如 Amazon 公司于近期推出的 CloudHSM 服务，该服务使用了一个基于硬件的加密密钥存储设备。在某些应用中，这被证明是更好的选择。

当谈及保护云中数据时，加密功能理所当然地被视为最重要的安全控制措施之一。虽然 S3 提供了多种加密功能选项，而更多的云存储厂商也在跟风，但是我敢断言这不会持续很长时间。

原文链接：http://www.searchcloudcomputing.com.cn/showcontent_74011.htm

(来源：TechTarget 中国　作者：Dave Shackleford　翻译：滕晓龙)

Amazon 云：用 StarCluster 实现服务器集群管理自动化

云计算技术使我们能够为计算密集型任务快速部署大量的服务器。你是否需要转换、筛选以及分析数以 TB 计的数据呢？你是否有成千上万的媒体文件需要被转换至不同的格式？

针对这些数据密集型的大任务，可选择在 Amazon 云中通过跨大型服务器集群发布大工作负载来实现。建立和管理这样一个服务器集群是一项极具挑战性的任务，但是 StarCluster 开源工具可通过在 Amazon 弹性云计算（EC2）中自动化运行众多的繁琐程序，可使服务器集群的创建和管理工作得到显著地简化，且耗时更少。StarCluster 是在 GNU 宽通用公共许可证(LGPL)下发布的。

StarCluster 软件可以让用户使用简单的命令行程序创建服务器集群。服务器集群中包含了一台单个的主服务器、多台工作服务器以及弹性块存储（EBS）卷。当你输入一个创建服务器集群的命令时，StarCluster 会完成如下的工作：

- 初始化虚拟机实例
- 配置一个新的安全组
- 定义一个用户友好的主机名（如，node001）
- 创建一个非管理员的用户帐号
- 为密码登陆配置 SSH
- 定义跨集群的网络文件系统（NFS）文件共享
- 配置 Oracle 网格引擎排队系统以实现跨服务器集群的任务管理

在 Amazon 云中使用 StarCluster

StarCluster 是使用 Python 语言编写的，因此你可以使用 easy_install 命令从 Python Package Index 进行安装。Linux 和 Mac 的用户可能已经安装了 Python ，但是 Windows 用户可能需要先行安装 Python 2.7 和

Python 安装工具。一旦 StarCluster 工具安装完毕，你就需要定义一个配置文件，其中应包含一些服务器集群和 Amazon 云账户的相关基本信息，例如访问 ID 和密钥位置等。StarCluster 创建了一个默认模板，你可以在此基础之上进行进一步编辑以指定特定账户的信息。

虽然 StarCluster 使用了一个服务器集群的默认小配置，但是你可以通过命令行方式或在配置文件中指定机器的容量、EBS 卷以及服务器的数量等参数。你还可以指定托管你服务器集群的 Amazon 网络服务（AWS）区域。StarCluster 的配置参数包括集群规模、在工作服务器节点上运行的 Amazon 机器镜像(AMI)、工作服务器节点实例类型、主服务器节点 AMI，以及服务器集群所连接和 NFS 共享的 EBS 卷。

将为服务器集群创建一个安全组，它可让你为整个集群制订防火墙规则。因为防火墙规则属于 EC2 核心功能的一部分，所以可通过 AWS 管理控制器或配置文件进行管理。

使用 Oracle 网格引擎实现精简高效的管理

Oracle 网格引擎（即 Sun 网格引擎）是一个排队系统，它可在集群中精简高效地执行任务管理。例如，如果你申请对数以千计的文件进行转换，那么你就可以使用 Oracle 网格引擎对所有这些文件进行任务调度，即在资源可用时按队列顺序管理文件转换处理。

使用 Oracle 网格引擎的另一个优势在于，它能够在服务器集群中实现跨服务器的工作负载平衡。根据你对引擎的具体配置设置，它会在服务器集群中增加或删除节点以适应实际需求的变化。Oracle 网格引擎还会监控队列中的任务和错误，并维护服务器集群中与任务状态相关的其他有用信息。

存储 StarCluster 的输出数据

一般而言，在集群中运行的任务都会产生数据；而输出数据中的部分（即便不是所有的）可能需要在集群关闭之后还是可用的。因为这些服务器集群都是由 Amazon 虚拟机构成的，所以你需要制订把相关数据存储在简单存储服务（S3）或 EBS 卷中的计划。如果你正在使用 S3 对象存储，那么你的任务就可以对 S3 进行数据读写，就好象你的应用程序是在服务器集群以外运行一样。StarCluster 的配置文件有一个可选项，可用于指定 EBS 卷的卷 ID 和载入点。在集群关闭后，这些存储卷都会与其他的 Amazon EC2 实例相连,所以任何写至集群内 EBS 卷的数据都可供这些实例使用。

StarCluster 支持对工作节点使用现场实例 ,它可通过可选的命令行方式为现场实例指定一个特定的服务器类型。此外，一个 spothisory 命令还可显示在之前三十天内一个实例类型的当前价、最高价和平均价。

StarCluster 是专为科学计算研究而开发的，它通常是一般科学计算工具中的标准配置组件。诸如 NumPy 这样的 Python 数值计算工具包以及 ATLAS 这样的优化线性代数工具包都非常擅长数据分析。StarCluster 的插件支持通用工具，如 Hadoop 和 MySQL。

原文链接：http://www.searchcloudcomputing.com.cn/showcontent_77485.htm

(来源：TechTarget 中国　作者：Dan Sullivan　翻译：滕晓龙)

亚马逊 CloudSearch 优于 DIY 搜索工具？

大数据对于企业商务智能的价值无可估量。然而，为了这种价值要挖掘大量非结构化文本数据，这意味着至少需要一个基本的搜索服务，有时候是更加高级的文本分析方法。

使用亚马逊 Web 服务（AWS）的云管理者和开发者现在可以实施自己的搜索服务器，使用流行的开源工具，比如 Lucene 和 Solr；或者使用亚马逊 CloudSearch。在决定使用搜索即服务和 DIY 搜索时，有一些问题需要考虑。

亚马逊 CloudSearch 的搜索即服务

亚马逊 CloudSearch 是一种基于云的搜索服务，企业可以将这个应用集成到索引文件中，响应搜索查询。和其他的 AWS 服务一样，亚马逊管理服务器实现，而非使用者。亚马逊 CloudSearch 提供了免费的文本搜索，以及一些更加高级的功能，比如分面搜索和自定制相关性排序。

分面搜索。分面搜索可以让应用用户通过使用文档分类刚要缩小搜索的文档范围。比如，一个文档注册库可能根据多个面或者字段分类文档，比如创建日期、文档类型或者关键话题。

自定制相关性排序。默认情况下，搜索索引中的所有字段都被认为是平等相关的，这也并非总是最佳的权重模式。然而，相关字段权重允许开发者权衡一些字段（比如关键字）的重要性，来确定文档的相关度，最终，在结果集中排列文档。

除了为应用开发者和管理者提供核心搜索服务，亚马逊 CloudSearch 会根据需求扩展。也在内存中维护了搜索索引来减少延迟。

用 Solr 和 LuceneDIY 搜索

亚马逊服务通常在运行企业自己的服务时具备成本竞争力；然而，如果你愿意承担由于用程序管理开销导致的潜在的高成本风险，从而获得更大的控制权和更多的功能的话，你可以看一下第三方的工具。比如，开源搜索平台 Apache Solr 是一种免费的平台，包括支持高级文本搜索功能、线性扩展性、几近实时的索引和扩展插件架构。Solr 也支持更加高级的文本分析操作，比如

单词拆分、正则表达式和听起来不错的过滤器。这个开源平台也包括支持国际化，对于拥有全球用户群的应用而言是一项重要的功能。

使用 Solr 的另一个优势是访问具体的应用可以减少你自己的开发者需求。以 LucidWorks 为例，提供了附件来执行命名实体识别；用 Drools 整合，开源业务规则引擎；调整搜索指针改善搜索结果质量和排序。

Lucene 是一个基于 Java 的搜索和索引服务，也是另外一种选择，但是提供的功能比 Solr 少。实际上，Solr 是基于 Lucene 的，但是增加了搜索和管理功能。

对比 CloudSearch 和 DIY 的成本

亚马逊 CloudSearch 的收费基于搜索实例的大小、文档批量上传、文档索引操作的数量和数据传输量。搜索实例的成本范围为：小实例每小时 0.1 美元到双倍超大实例每小时 1.1 美元

如果搜索服务需要持续较长时间，你可能需要考虑对比亚马逊 CloudSearch 成本和预留实例价格，而非按需价格。预留实例的一到三年承诺有效。

Number of Documents	Document Size (KB)	Queries per Day	Updates per Day	Re-indexes per Month	Instance Size	Cost per Mont
10,000	5	1,000	100	1	Small	$73.25
50,000	10	2,500	500	1	Small	$73.67
125,000	10	5,000	500	1	Large	$286.65
250,000	10	5,000	500	1	Large	$287.82
1,000,000	10	5,000	500	1	Extra Large	$411.95

图 1 不同场景成本

亚马逊 CloudSearch 成本如图所示。亚马逊 CloudSearch 的成本收到文档注册库的高度影响，决定了搜索实例大小。评估运行自己的搜索服务的成本，比如 Lucene 或者 Solr 服务器，由于管理成本的多变性更加困难，但是我们可以评估运行实例的成本，对比在亚马逊 CloudSearch 上的运行情况。使用按需价格和假定实例每天运行 24 小时，每月运行三十天，通用小型实例的成本为 43.2 美元，大型实例的成本为 172.8 美元，超大型实例的成本为 345.6 美元。DIY 实例和亚马逊 CloudSearch 成本之间的差异并不明显。在用例查询大型实例中，DIY 节省的成本可能少于管理员两个小时的成本。

亚马逊 CloudSearch 可以让开发者针对基于云的应用快速实施搜索功能。服务包括支持基本的搜索操作，以及一些比 DIY 方法更具成本竞争优势的更加高级的性能。对于需要更多高级需求的用户，管理自己的服务的额外支出等价于高级搜索和文本分析带来的好处。

原文链接：http://www.searchcloudcomputing.com.cn/showcontent_80098.htm

(来源：TechTarget 中国　作者：Dan Sullivan　翻译：张培颖)

AWS Auto Scaling：云服务成本与性能平衡利器

基于即付即得（pay-as-you-go）的云计算模型有助于开发人员和架构师设计出能够一直运行于最佳资源分配下的应用程序。给服务器过高的配置来满足业务需要意味着花费太多冤枉钱，这个错误的另一个极端则更糟糕，即给服务器过低的配置导致应用程序的性能不足。而应用程序要求的变化使得对服务器配置的选择变得更加困难，因为最佳的资源组合不是一成不变的。

使用 AWS 服务来应对这些变化的一种方法是创建一个负载均衡服务器集群，然后根据需要添加或删除服务器。你可以自己管理这样一种集群，除此之外,你还可以尝试 AWS Auto Scaling 服务。该服务在避免过度配置的条件下保持足够的性能，同时也能够减少一些管理费用。该服务要达到的最佳应用场景是工作负载变化显著变化时，AWS Auto Scaling 能够轻松地跨多个服务器分配负载。

AWS Auto Scaling 使用 CloudWatch（亚马逊提供的一种监控工具）来提供所需的性能数据，完成伸缩服务建议。每隔五分钟，CloudWatch 都将从服务器和其他 AWS 资源处，免费地收集性能统计数据，包括 CPU 使用率、磁盘使用情况和数据传输情况等。（额外付费的话将可获得每分钟一次的性能指标收集服务。）系统管理员可根据这些测量信息规定添加或删除服务器的配置策略。例如，一项配置策略指示当 CPU 平均利用率超过 70%时，将启动一个额外的虚拟实例。

通过实现自动伸缩组协助制定配置策略

使用 AWS Auto Scaling 服务的第一步是实现一个自动伸缩组，即在逻辑上统一管理的一组亚马逊弹性计算云（EC2）实例。每一个组均被指定了最小和最大实例数，而实际数值则由基于 CloudWatch 的测量值做出的配置策略来决定。当需要手动干预时，通过 AWS 提供的 ExecutePolicy 命令行，系统管理员便可以无需等待触发条件，直接执行一项策略。

通过自动伸缩组的应用，企业在维护应用程序性能和开销的许多相关工作中得到了帮助。

自动伸缩组能够跨越可用性区域（一种能够支撑应用程序高可用性要求的特性）提供服务。这些可用性区域处于 AWS 范围内，例如，美国东部（北维吉尼亚）或欧盟地区（爱尔兰），通过不同的基础架构隔离故障

实例与正常实例。如果在某个可用性区域中发生了故障，AWS Auto Scaling 将在同一个地理区域内的某个功能区启动一个新的实例。

自动伸缩组能够通过负载均衡器配置集群内服务器间的工作负载。亚马逊的弹性负载均衡服务提供了一个到你的应用程序的所有流量的单点访问。当使用负载均衡器时，可以引用负载均衡测量指标（比如请求等待时间）以及 EC2 实例测量指标来制定自动伸缩策略。

除了响应变化的负载，Auto Scaling 还支持其他伸缩选项，包括持续确保当前实例的性能、手动伸缩以及基于排程的伸缩。

应对 AWS Auto Scaling 的潜在问题

完成一项自动伸缩策略的指定动作需要花费一定的时间，然而，AWS 在一次触发后，实现了一个冷却周期，用于防止当对触发器最初的响应还在继续的时候，执行了对触发器第二次响应所引起的一系列事件。冷却周期始于执行策略动作。

针对 Web 服务器和一些应用程序服务器而言，当应用程序负载可以分布在多个服务器之上时，自动伸缩不无裨益。某些系统（如关系型数据库系统）可以配置运行于集群中，但其中存在诸多缺点：商业版关系型数据库可能通过收取额外授权费用来提供集群支持，这些授权费甚至会超过因在不同数量的小型服务器运行数据库而拒绝一台大型服务器所节省下来的费用。此外还要考虑管理数据库服务器集群与管理单个服务器的不同开销。

原文链接：http://www.searchcloudcomputing.com.cn/showcontent_80716.htm

(来源：TechTarget 中国　作者：Dan Sullivan 翻译：高珂)

本期电子书由 TechTarget 出品