第 33 卷 第 9 期 计 算 机 工 程 2007 年 5 月
Vol.33 No.9 Computer Engineering May 2007
·开发研究与设计技术· 文章编号：1000—3428(2007)09—0263—03 文献标识码：A 中图分类号：TP399

面向多层次分布式商业应用的管理平台架构

刘丹军[1,2]，詹剑锋[2]，马 捷[2]，江 滢[1,2]
(1. 中国科学院计算技术研究所国家智能计算机研究开发中心，北京 100080；2. 中国科学院研究生院，北京 100039)

摘 要：对于被集中部署到机群环境中的应用服务来说，为了保障服务的负载均衡和高可用特性，通常会配备冗余的软硬件资源，并采用相应的管理系统[1]，帮助调配这些资源、维持稳定的服务质量。然而，商业分布式应用的规模日益庞大，如何让管理系统适应其复杂结构，该文提出了一种解决方案。该方案通过定义形式化模型，建立了一套描述复杂多层次应用结构、判定应用运行状况的方法，并在此基础上构建了具有广泛适应性的平台环境，使得部署于机群之上的复杂商业应用在此架构下得到统一的管理。

关键词：形式化模型；商业应用；应用管理；服务质量；机群

Management Platform Architecture for Multi-tier Distributed Commercial Applications

LIU Danjun[1,2], ZHAN Jianfeng[2], MA Jie[2], JIANG Ying[1,2]
(1. National Research Center for Intelligent Computing Systems, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080;
2. Graduate School, Chinese Academy of Sciences, Beijing 100039)

【Abstract】 In order to attain load balancing and high availability, large-scale commercial services are usually equipped with redundant resources and managed by certain management systems. However, the complex structures of services lay obstacles to the design and implementation of this kind of system. A formal model is defined, by which methods are created to describe the structures and status of the multi-tier applications and further build a platform environment. And most of complex commercial application services deployed in clusters can be governed under this kind of architecture.

【Key words】 Formal model; Commercial application; Application management; QoS; Cluster

随着机群技术的发展，集中式服务器管理又回归到商业信息服务领域中来，并产生了很多新的流行趋势，如服务器聚集[2]、服务器群组[3]、应用服务提供商模式[4]等。这些技术的共同点，就是使用同一机群内的软硬件资源，同时向外提供多种应用服务，以实现统一管理、减少TCO的目的。例如，某大型的公共数字图书馆，它允许用户通过认证后检索书目、在线浏览图书内容的同时，还提供流媒体服务，并生产和处理其它供未来使用数字产品。

在实际的部署中，为了保障服务质量，管理员常常要为各个服务配备冗余的节点资源，并通过负载均衡手段(添加请求分发器)和高可用工具利用这些资源来维持一定标准的服务质量。但是，由于商业应用有着较高的峰值-均值负载比，要按照负载的动态变化，将资源分配给负担最重的服务，这时就需要一套管理系统来保证资源的有效利用。这类系统在逻辑上提供了一个保证应用服务稳定可靠运行的基础环境，因此本文将其统称为“应用管理平台”。

然而，随着分布式商业应用的发展，多层次应用模式出现，应用结构愈发复杂，应用管理平台需要有效的手段来获知所要管理的应用的结构信息，判断其运行状态，并以此为基础实现上述管理功能。

1 多层次复杂应用结构和冗余资源配置

用户看到的某种服务，实际上可能是由运行于不同服务节点上的多个分布式应用组成的。它们之间存在复杂的层次依赖关系，合作完成对请求的应答。因此，要判断某种服务是否运行状况，必须事先获知该服务的组成结构，根据各部分的不同作用和各自的运行状态进行具体分析。

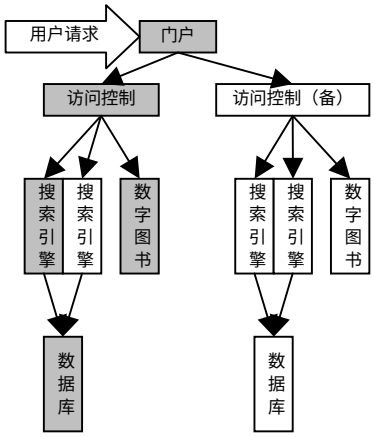

图 1 图书馆应用

另外为了维持稳定的服务质量，应用会使用冗余配置来做服务备份或负载分担，而且为了达到良好的效果，会在多

基金项目：国家“863”计划青年基金资助项目(2004AA616010)；“十五”科技攻关计划项目(2004BA811B09-1)
作者简介：刘丹军(1980－)，男，硕士生，主研方向：机群管理软件；詹剑锋、马 捷，博士、副研究员；江 滢，博士生
收稿日期：2006-05-28 **E-mail：**liudanj@ncic.ac.cn

个层次上进行这种配置，这也对管理系统提出了挑战。

例如图书馆应用可以近似表示为图 1。其中灰色填充表示该部分应用已在提供服务，白色填充表示为冗余资源，仍在空闲。这就是一个多个层次的，各组成部分存在依赖关系，又配备了冗余资源的应用服务。

对该实际应用进行简化可得如图 2 中所示的抽象表现。A_1 和A_2表示的应用具有相同的功能，B_i, C_i, D_i (i=1,2)也分别是表示功能地位相同的应用。

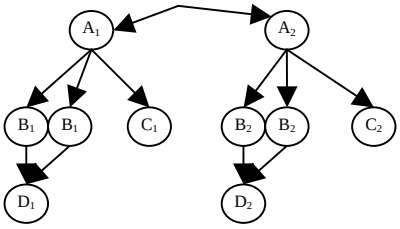

图 2 多层次应用及冗余资源的抽象表现

A向外提供服务，依赖于B和C，而B又依赖于D。B_1和B_2互为冗余资源。A_1~D_1这 4 种应用合作向外提供一种服务，A_2~D_2也完成相同的功能，这两组也互为冗余资源。管理平台不仅要了解ABCD之间的结构关系，还要清楚上述的两个层次的冗余配置，以便在下列情况时作出准确的判断：

(1)当部分应用失效时，是否认定无法向外提供服务。一个B_1(或一个B_2)失效所导致的结果同C_1(或C_2)应用失效的结果是不同的；

(2)当服务质量恶化时，如何判断性能瓶颈的所在位置，以便进行负载均衡。例如，图 2 中$A_1B_1C_1D_1$提供的服务过载时，是要启动新的B_1还是启动整个的A_2~D_2来分担负载？为了让管理系统就需要事先了解应用的具体结构，本文通过定义形式化模型给出了一套解决方案。

2 多层次应用的抽象形式化模型

大多数由中央服务器提供的应用服务都遵循相同的模式，就是客户机/服务器模式。几乎所有复杂的应用服务都可以分解为这种两点依赖关系(如图 2 中A_1和C_1)。依据这种模式，我们定义两种逻辑模型实体来描述多层次应用的结构和冗余资源配置。

定义 1 服务单元

若应用服务或应用服务的组合满足：(1)向外提供某种特定的服务供他方(可以是实际用户或其它服务)使用；(2)具有单一的服务访问点，则称其为“服务单元”。

最简单的服务单元，是在某个服务器节点上运行的向外提供服务的进程。服务单元是可以自身递归嵌套的，多个服务单元可以根据客户机/服务器的关系，形成一个有向无环图的结构[5]，提供一个单一的服务访问点，组成新的服务单元。

定义 2 服务池

服务池是用于描述冗余资源配置的。它由两部分构成：一个请求分发器和具有相同服务功能的若干服务单元。

分发器是唯一的服务访问点，负责将外界的服务请求分发到后端某个服务单元上；后端的各个服务单元的功能作用都是相同的，其中每个都向外提供完整服务，这里称之为“服务实例”(简称“实例”)。服务池提供服务，也具备单一服务访问点，符合定义 1，所以可以看作一种特殊的服务单元。因此，服务单元和服务池可以相互递归嵌套。

在文献[5]中已经推定，具有确定的稳定拓扑结构的复杂应用及其冗余的资源配置都可以用上面模型进行形式化的归纳划分。

图 3 给出了图 2 中的应用按照上述模型描述得出的两种划分方案。图中的阴影三角代表请求分发器，椭圆代表服务单元/服务池。图 3(a)中的应用由 3 个服务池组成了一个服务单元，而图 3(b)则是一个服务池，由 1 个分发器和 2 个服务实例组成，而每个服务实例又是由 1 个服务池和 3 个服务单元组成。这两种划分都是逻辑上的，与实际部署不完全对应。

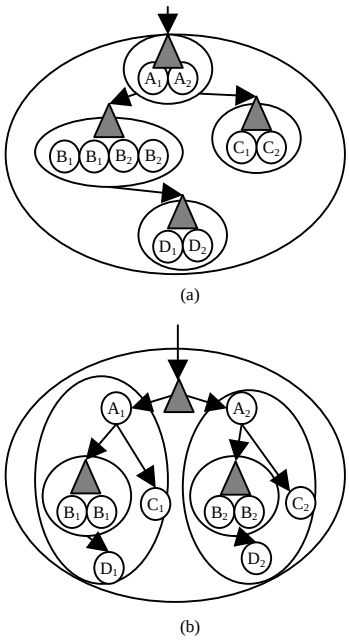

图 3 按模型划分得出的两种方案

由此例也可以看出，采用上文定义的模型进行描述不一定会得出唯一的划分结果。不同的组织形式各有优缺点，比如图 3(a)的结构比图 3(b)的结构简单，但对分发器的可用性要求高，而且耦合过于紧密，不容易定位错误和及时修复。

3 复杂应用运行状态的判断

依据上面的模型定义，可以用两种实体概念来描述应用和资源。但还要在此基础上对应用运行状况和负载信息进行量度，才能够作出判断，替换故障部分，均衡应用负载，保障服务质量。

对于最简单的服务单元来说，使用操作系统提供的标准接口(系统调用或 API)，即可获知相关的状态信息。

简单服务单元可以构成复杂的服务单元。对于一般的复杂服务单元(非服务池)来说，其各组成部分都是不可替代的，组成部分的运行状态直接决定了服务整体的状态。依据这种关系进一步分析可知，各组成部分的负载大小直接影响整体的服务能力，服务单元整体的剩余服务能力(由其服务能力和其上当前负载得出的表征)同其结构中可提供的剩余的服务能力最小的部分是成正比的，如果用图形面积表示服务能力大小，那么此关系用公式 $R=\bigcap_{i=1}^{n}R_i=\bigcap_{i=1}^{n}[S_i\cdot(1-L_i)]$表示，其中R是整体的剩余服务能力的图形，$R_i$表示各个部分的剩余服务能力的图形，$S_i$为各个部分的总服务能力的图形，$L_i$是用百分数表示的当前负载的大小。因此在计算负载时，可以用负载最大的子部分的负载来近似表征。因为如果一个部分过载的话，即使其它部分还剩余服务能力，服务单元整体也无法向外提供及时的服务。

对服务池运行状况的判断同服务单元有较大差别。在可用性方面：(1)如果分发器失效，整个服务池无法向外提供服务；(2)分发器后端的服务实例中，单个服务单元失效不影响服务池服务的可用性。如果分发器工作正常，当且仅当所有“实例”全部失效，服务池才无法向外提供服务。而在服务能力方面，服务池的服务能力是组成它的所有的服务单元的服务能力的总和。如果服务池中有n个相同功能的服务单元SU_1~SU_n，负载为L_1~L_n，按照服务能力设置权值W_1~W_n，权值和能力成正比，越高表明可同时应答的用户请求越多，那么该服务池的负载为

$$L = \frac{\sum_{i=1}^{n} L_i W_i}{\sum_{i=1}^{n} W_i}$$

为了充分利用资源，在多数情况下，服务池中的服务实例并未全部参与服务，因为其中部分作为冗余资源也可能被分配给别的服务池使用或停用以减少消耗。对于这样的空闲节点资源，如果可以随时参与服务，那么要将其负载记为 0 代入公式计算，可以得到对服务池负载的近似量度。

4 应用服务高可用与多级负载均衡

根据分析所得的应用运行状况，管理平台可以有针对性地进行资源调度和调整，最终实现服务对用户请求的自反馈和自适应。调整主要体现在两方面。

(1)高可用：服务池中的多个实例都可以独立地向外提供相同的服务，所以可以用空闲的服务单元来替换故障的应用服务。这里的“实例”服务单元可以包含多个子服务单元。例如图 3(b)代表的应用，如果一个B_1失效，可以切换到位于同一个服务池的另一个B_1上，但是如果服务单元中的 2 个B_1都失效，或A_1(或C_1或D_1)失效，则A_1~D_1组成的应用不能提供服务，必须用A_2~D_2组成的应用来替换。而至于分发器，由于实现方式不同高可用实现也不同，这里略。

(2)负载均衡：首先，平台可以根据服务池中的不同实例的负载，调整分发器的分发策略；其次，平台可以自适应地调整参与服务的实例的数量，某个应用服务承受请求高峰时，就把更多的冗余节点资源分配给其使用。加之冗余资源是出于不同层次上的，所以平台可以在多个级别进行调整。一般来说，在调整时应该遵循开销最小、对系统整体影响最小的原则，在提高服务能力的同时启动尽量少的应用(服务单元)。所以，应用管理平台应该按以下顺序进行负载均衡：1)调整分发器的分发策略；2)调整层次较低的服务池(如图 3(b)中的 B)；3)调整较高层次的服务池。每次调整只能进行这三者中的一个，只有在顺序靠前的调整无效时，才会进行其后的调整，直至所有的冗余资源均被使用。

5 评价与应用

着眼于服务高可用和资源有效使用的管理系统是近年来的研究热点[6,7]，本文通过利用形式化模型进行描述和划分，使应用管理平台可以了解集中部署于机群内的复杂应用的结构和冗余资源配置，对服务运行状况和资源使用情况进行判断和分析，针对不同的情况作出相应的调整和资源调度，从而保证了服务质量的稳定。依据这一原理实现的管理平台，可以发挥如下作用：

(1)根据机群内各应用服务结构和负载状况，自适应地调整冗余资源的分配，提高资源使用率和应用服务质量；

(2)找出可以停用的空闲资源，减少消耗；

(3)对于大规模应用服务，可以在实际部署之前按照比例进行小规模测试，根据系统自动调整所得的判断，对资源需求进行定量推演。

依照上述建模和分析，我们在Phoenix的基础上实现了一套应用管理平台原型系统BAMP(Basic Application Management Platform)。Phoenix[8]是由中科院计算所国家智能计算机中心开发的一体化机群管理中间件平台。BAMP利用了Phoenix提供的基础服务，其架构如图 4：BAMP从配置服务中获取应用结构和资源配置等静态信息，从“数据公告服务”和“事件服务”获取应用运行时的动态信息，按照第 3 节中的算法进行分析判断，再按照第 4 节的原则进行调整，保证其管理的各种应用服务的服务质量。

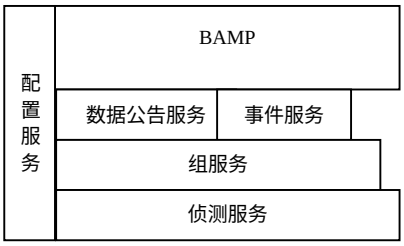

图 4 BAMP 系统结构

BAMP 系统已成功地应用于某大型数字图书馆项目的资源配置测试和推演。该数字图书馆的全文搜索引擎按照访问频度进行了分类，对热点数据配置冗余资源。项目进行了小规模的性能测试，BAMP 对冗余资源进行了管理，得出的运行数据被用于对实际系统部署的推演。BAMP 还可以用于大型数据中心的日常系统监控和管理。

参考文献

1 Ye Q, Xiao L, Meng D, et al. AppManager: A Powerful Service-based Application Management System for Cluster[C]//Proc. of IEEE ICPPW'02. 2002.

2 Eastwood M. Worldwide Server Consolidation Forecast and Analysis, 2002-2006[R]. IDG Research, 2002-09: 1-9.

3 Matsubara K, Parker A, Castro A L A, et al. Managing AIX Server Farm[R]. IBM Redbook SG24-6606-00, 2002-06.

4 Harney J. Application Service Providers(ASPs): A Manager's Guide[M]. Addison-Wesley Professional, 2002-01: 3-20.

5 Liu D, Ma J, Zhan J, et al. A Formal Method for Modeling and Managing Large-scale Distributed Applications[C]//Proc. of the 17th IASTED PDCS. 2005-11.

6 Appleby K, Fakhouri S, Fong L, et al. Océano–SLA Based Management of a Computing Utility[C]//Proc. of the 7th IFIP/IEEE Symposium on Integrated Network management, Seattle, Washitong. 2001: 855-868.

7 Jiang Y, Meng D, Zhan J, et al. Adaptive Mechanisms for Managing the High Performance Web-based Applications[C]//Proc. of the 8th International Conference on High Performance Computing in Asia Pacific Region. 2005-11.

8 Meng D, Zhan J, Wang L, et al. Fully Integrated Cluster Operating System[J]. Journal of Computer Research and Development, 2005, 42(6).