维普资讯 http://www.cqvip.com

数据库异构集群的性能模型研究*)

王元珍　龚卫华
(华中科技大学计算机学院　武汉 430074)

摘　要　在OLTP应用中数据库集群是一种有效的并行处理方案，由于以前对数据库集群特别是异构情况下的性能评价不够完善，本文主要研究数据库异构集群的性能模型，分析了CPU和内存两种资源的异构带来性能影响，并给出了异构集群并行性的度量标准及系统有效性评估公式。最后，通过TPC-C实验表明数据库异构集群在OLTP处理中仍具有良好的可扩展性，次线性的加速比，以及高效费比的并行处理服务。

关键词　异构集群，并行处理，可扩展性

The Research of Performance Model on Database Heterogeneous Cluster

WANG Yuan-Zhen　GONG Wei-Hua
(College of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074)

Abstract　Database cluster is an efficient parallel processing system in On-Line Transaction Processing (OLTP), since the previous performance evaluation of database cluster especially for heterogeneous case was incomplete, this paper mainly discusses the performance model of database heterogeneous cluster, analyzes the performance impact of two heterogeneous resources such as CPU and RAM, and presents the metrics of measuring the parallelism of heterogeneous cluster and estimates the system efficiency. The results of TPC-C experiments show that the database heterogeneous cluster still has advantages of good scalability, asymptotic speedup, and high efficient parallel service.

Keywords　Heterogeneous cluster, Scalability, Parallel processing

1　引言

目前集群已被广泛地应用于Web服务[1]、集群计算(cluster computing)[2]、科学应用等许多领域，它们都是比较流行的研究热点，因为如今大量廉价机器和并不昂贵的高速网络使得集群成为提供高效费比服务的有效途径，而且集群是一种松耦合的结构，具有良好的扩展性。此外在对性能要求较高的OLTP应用中，集群也是通过并行处理来提高系统性能，虽然并行处理可以在基于主机的对称多处理机系统(SMP)和多机系统(MPP)上实现，现代数据库管理系统已经能够有效地利用对称多处理器资源，但是单机系统的系统负载能力总是有限度的，当系统负载达到极限时，数据库系统整体的运作效率就会严重下降。所以解决这个问题的根本方法，就是采用多机分布式并行处理(MPP)方案，而数据库集群[3]就是一种有效的系统。

近年来国内外大多数研究主要是基于同构集群的分析和应用，讨论集群系统的可扩展性和可用性[4,5]，以及负载平衡[6]等方面，文[7]中还评价了基于集群的网络服务器性能。但实际应用中我们很难集中软件和硬件结构都相同的集群一起来工作，异构集群会越来越普遍，而且异构集群与同构集群性能差异很大，在数据分布和负载平衡方面比同构集群也更加复杂，因此对异构集群的性能评价难度也比较大，虽然文[8]采用等效率(isoefficiency)模型给出了评价异构集群并行性能的效率和扩展性标准，文[9]讨论了提高异构集群加速比和资源利用率的并行技术，文[10]分析了异构平台上线性几何算法应用的扩展性度量，文[11]从理论上分析了异构集群的扩展性，以上对异构集群的研究都是针对数据或算法的并行处理，而对面向事务并行处理的讨论比较少。这里为了更具广泛性，我们可以把同构集群看作是异构集群的特例环境。

在数据库集群应用中不仅可以提供Web服务[12]例如电子商务(e-Commerce)，还可以提供面向事务处理的OLTP服务[13]，但这些应用都是基于同构的集群结构，而在数据库异构集群中数据分布是不均匀的，并且由于各异构节点的计算能力不同，所承担的负载也不相同，因此本文着重研究在针对OLTP应用中数据库异构集群的性能模型，讨论集群并行性的度量标准及系统的有效性评估。

2　数据库异构集群

在并行处理中SMP对称多处理是一种紧耦合的结构，虽然易于管理和配置，但是在可扩展性和性能上不及集群，而且规模也受到限制，数据库集群尤其是在处理OLTP应用时，具有完全的可扩展性、高有效性和性价比等优点。数据库集群是一组完整的自治的计算处理单元(节点)，每节点均有独自的CPU、内存以及磁盘等硬件资源，运行独立的操作系统和自治的数据库系统，通过高速专用网络或者商业通用网络互连，彼此协同计算，作为统一的数据库系统提供并行事务处理服务。

数据库集群不仅可以提供电子商务(e-Commerce)应用还可以提供OLTP服务，能够以较低的价格获得与一台大型主机运行的数据库系统一样的事务处理能力。在实际的数据

*)本课题获国家信息产业部电子发展基金项目。**王元珍**　教授，博士生导师，主要研究方向为分布式多媒体数据库、并行数据库、数据库安全。
龚卫华　博士研究生，主要研究方向为多数据库系统，数据库集群加速技术。

维普资讯 http://www.cqvip.com

库集群扩展过程中，异构配置的节点已经成为影响集群性能不可忽略的因素，数据库集群的异构性主要包含两个方面：软件配置异构性和硬件配置异构性，软件配置异构性主要是指由不同操作系统以及异构数据库而导致不同事务处理能力，而硬件配置异构性是指影响集群节点计算能力的各类资源不同主要包括 CPU 速度，RAM 大小和磁盘 IO，本文所主要讨论的异构仅指 CPU 计算能力和内存容量的不同而导致性能差异，并不考虑操作系统，网络连接类型及硬件组织的异构。文[14]和[15]中主要考虑了容易造成性能瓶颈的 CPU 和内存两种异构资源下的负载平衡策略，我们将进一步分析在此异构情况下数据库集群的性能。

数据库异构集群并行性的重要度量标准是可扩展性(scalability)和加速比(speed up)，可扩展性反映了事务处理规模与系统增长之间的关系，而加速比则体现了系统对给定规模的事务并行处理速度的提高，此外，系统的有效性也是描述集群系统使用效率的一个重要标准。

3 异构集群性能分析

数据库异构集群中影响整个系统性能的因素主要有以下几个方面：①节点数，②网络延迟，③整个数据库大小，④客户端数量，⑤事务类型，⑥单台机器的性能等。在网络延迟一定的情况下，异构系统容易形成的瓶颈是由最差的节点性能决定的，而单个节点的事务处理能力与 CPU 速度，内存(RAM)大小和磁盘 I/O 是紧密相关的，此外还有软件配置如数据库性能，数据库大小，并发用户数等方面的相关因素，但不在本文所讨论的范围之内。由于在数据库异构集群中 OLTP 负载的特点是大量频繁更新的短事务，因此主要消耗 CPU 和 I/O，而 I/O 的频度与内存大小相关，我们首先假设单机的数据库数据都在内存中，但随着数据库的增大内存与磁盘 I/O 间的数据交换次数也要增加，目前常见磁盘 I/O 的稳定速率一般在 30Mbps 到 45Mbps 之间，而 CPU 和内存的变化范围比较大，对单机系统性能的影响也比较显著，所以本文主要讨论 CPU 和 RAM 异构情况下的性能模型。

异构集群中单台机器的计算能力主要由 CPU 速度 V_i 和内存大小 R_i 决定，V_i 的单位是 MIPS，R_i 的单位是 Mb。为了便于比较集群中 p 个并行节点的性能指标，我们选定一个基准点，该基准点的 $\overline{V}=\frac{\sum_{i=1}^{p}V_i}{p}$，$\overline{R}=\frac{\sum_{i=1}^{p}R_i}{p}$，亦即基准点的 CPU 速度和内存大小都为异构集群中的平均值，节点的事务处理能力用吞吐量 $\overline{TP}=f(\overline{V},\overline{R})$ 表示，我们采用系数 H_{CPU} 和 H_{MEM} 来分别评价异构配置的 CPU 和内存：

$$H_{CPU}=\frac{\sum_{i=1}^{p}|\overline{V}-V_i|/\overline{V}}{p} \tag{1}$$

$$H_{MEM}=\frac{\sum_{i=1}^{p}|\overline{R}-R_i|/\overline{R}}{p} \tag{2}$$

可以看出在同构数据库集群中，各节点配置与基准点一致，计算能力也相同，$H_{CPU}=H_{MEM}=0$，对于异构数据库集群，H_{CPU} 和 H_{MEM} 的值越大说明各节点间异构性越大，各节点的计算能力差异也越大。

数据库异构集群中的可扩展性是一个衡量并行处理的重要指标，虽然文[16]和[17]已分别采用增量分析方法和时标方法评价了集群的扩展性，这里本文将在异构集群中通过系统负载及响应时间等参数来分析数据库集群的并行性指标，可扩展性定义：一个系统是可扩展的如果吞吐率与系统大小的比率保持稳定，同时吞吐率的响应时间对于增长的系统能保持达到要求，系统可扩展性公式可表示为：$W=f_E(p)$，W 表示用户负载，p 是节点个数，$E=$常数，这里我们只要求 $E\geqslant\lambda$，λ 为达到扩展性要求的阈值。更进一步地，可扩展性在直观上反映为当集群系统从 p_1 台机器扩展到 p_2 台机器时，可扩展比为：

$$S_E=\Phi(p_1,p_2)=\frac{p_2*W_{p1}}{p_1*W_{p2}} \tag{3}$$

W_{p1} 表示在 p_1 台机器下的负载，由于 $TP=\frac{W}{T}$，TP 是事务处理吞吐率(每秒钟处理的事务数)，因此 $W_{p1}=\sum_{i=1}^{p1}(TP_i*T_i)$。当在同构集群环境下，各节点相同配置导致计算能力一样，这样负载也会均匀的分布在各节点上，系统扩展性可推导为：

$$S_E=\Phi(p_1,p_2)=\frac{p_2*W_{p1}}{p_1*W_{p2}}=\frac{p_2*p_1*(TP_{p1}*T_{p1})}{p_1*p_2*(TP_{p2}*T_{p2})}=\frac{TP_{p1}*T_{p1}}{TP_{p2}*T_{p2}} \tag{4}$$

也就是系统中单节点机器的负载当从 p_1 扩展到 p_2 时的比率，而对于异构集群由于配置的不同而导致负载的不均匀就不能依此计算扩展性。

在异构的集群环境下，加速比(speedup)也是衡量系统并行性的另一个重要标准，异构集群的加速比等于同一任务在最快处理能力的节点上单独运行时间 T_1 与在异构集群中 p 个节点上并行运行的时间的比值，亦即异构加速比公式为：

$$S(p)=\frac{T_1}{T_p}=\frac{Tc(1)}{T_{com}+Tc(p)} \tag{5}$$

$Tc(1)$ 假设在最快处理能力的节点只有计算时间，而 T_{com} 表示 p 个节点间的通信时间：$T_{com}=\alpha+\beta*Size$，其中 α 表示网络延迟，为了简单起见，我们假设集群中节点间连接长度相等网络类型也相同，因此 α 是固定的，另外 β 是传输单位数据所需的时间，那么通信开销就取决于数据传输量 Size。$Tc(p)$ 表示 p 个节点的并行处理时间，它由最慢机器的执行时间所决定的，取 $Tc(p)=\max_{i=1}^{p}Tc(i)$。

另一种改进的在异构环境下的理想加速比定义为：

$$S'_p=\frac{\sum_{i=1}^{p}V_i}{V_1}=\frac{\sum_{i=1}^{p}TP_i}{\max_{j=1}^{p}TP_j} \tag{6}$$

p 个处理机(节点)求解问题的运算速度/最快单机求解问题的运算速度，其中运算速度就是吞吐率 $TP=W/T$，W 为工作负载，T 表示响应时间。在同构环境下，理想的加速比就是节点数 p，然而在异构环境下理想加速比不是 p。

此外，异构集群系统中的有效性是用来衡量节点在并行计算中的使用率，也是评价集群系统性能的一个重要方面。文[11]中给出了有效性模型，它基于两点假设：一，对于给定的问题，系统加速比因为网络通信开销而不成线性增长，当并行系统中的处理器增加时系统有效性下降。二，对于给定的处理器数，更大规模的问题导致更高的加速比和有效性。所以异构数据库集群中 p 台机器并行处理问题规模为 N 时的系统有效性(Efficiency)为：

$$E(N,p)=\frac{Tseq(N)}{p*T(N,p)}=\frac{\sum_{i=1}^{p}t_i}{\frac{p}{p*T_p}}=\frac{\sum_{i=1}^{p}t_i}{p*p*T_p} \tag{7}$$

t_i 表示单个节点上的执行时间，而 T_p 表示并行执行时间，$T_p=T_{com}+\max_{i=1}^{p}t_i$，$E(N,p)$ 也称为并发有效性。在同构集群环境下，系统有效性：

$$EF(p)=S(p)/p \tag{8}$$

因为理想的加速比就是节点数 p，然而在异构环境下理

维普资讯 http://www.cqvip.com

想加速比不是 p，所以异构环境下的有效性可定义为：

$$EF(p)=S(p)/S'_p \qquad (9)$$

以上分析了数据库异构集群的可扩展性，加速比和有效性等性能度量标准，下面我们通过实验来评价针对 OLTP 应用的数据库集群性能的各方面指标。

4 实验结果及性能分析

TPC-C 作为联机事务处理(OLTP)的性能基准测试标准，在国际上已经取得了广泛的认可，它模拟了非常接近现实的商业应用环境。提供数据库服务的异构集群环境主要由 15 台异构节点采用 TCP/IP 通信协议，以 100Mbps 交换机连接组成快速以太网，其异构节点的 CPU 和内存配置依次如表 1 所示。

表 1 数据库异构集群节点硬件配置

节点数	CPU	内存(MB)	计算能力(Cap.)
4	PⅣ 1.6G	512	27
1	PⅣ 1.6G	256	26
4	CⅢ 1.0G	384	22
3	PⅣ 2.0G	128	25
3	AMD 800M	512	21
Total=15	$\bar{V}$=1.36 G	$\bar{R}$=384	$\bar{C}$=22

表 1 中异构节点的计算能力是以每种 CPU 和内存的配置在事务处理到达饱和点时所容纳的仓库数作为衡量标准，由(1)式和(2)式分别计算得 $H_{CPU}=0.306$ 和 $H_{MEM}=0.311$，由此可见节点异构配置与基准点差异较小，集群中的异构节点按表 1 中列出的配置顺序增加，分别测出每种配置的最大事务处理能力，单机的数据库处理能力有限，从表中得出，综合性能最快的节点 TPC-C 测试事务处理能力最多运行 27 个仓库就会形成系统资源瓶颈。当扩展到多台数据库集群时，系统虽然能并行运行更多数量的仓库，但集群中站点由于异构性以及通信开销的增大在饱和点时所运行的仓库总数量并不是成比例线性增加的，因此我们从可扩展性、加速比以及有效性等方面系统地评价异构数据库集群的并行处理性能。

当异构数据库集群的扩展性 $S_E \geqslant 1$ 时，我们就认为该集群系统是线性扩展的，通过 TPC-C 工具测得在各节点数下的负载，由式(3)计算出可扩展比曲线如图 1 所示，由图可知，系统开始在同构节点下扩展比上升较快，随着异构节点的增加，扩展比上升比较平稳，但一直都大于 1，因此异构数据库集群仍是一个线性扩展的过程。

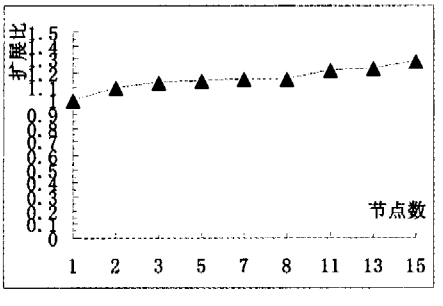

图 1 异构数据集群扩展比

异构集群的加速比也是衡量系统性能的重要指标，为了更直观地与同构集群比较，我们假设有相当于异构计算能力的同构集群由表 1 中与平均计算能力的基准点相同配置的 15 个节点组成，经过实验分析由式(6)计算出数据库异构集群的加速比，图 2 中给出了同构和异构集群以及理想情况下的加速比变化过程，从图中可以看出理想的加速比是一个线性加速的过程，在 8 个节点以内的同构和异构集群由于网络开销比较小，所以与理想的加速曲线相一致，总体上在同构环境下有较理想的加速比曲线，异构环境下的加速比曲线接近于同构环境，它们都是一个次线性的加速过程。

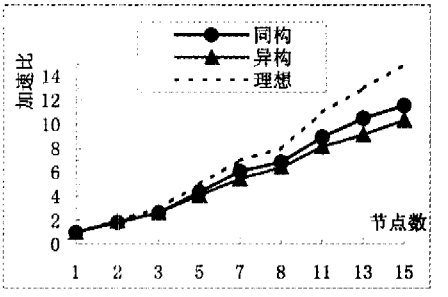

图 2 异构数据库集群与同构集群的加速比曲线

根据以上 TPC-C 实验测得在理想及异构情况下的加速比曲线，我们可以进一步评价数据库异构集群在提供 OLTP 服务时的系统有效性，由式(8)和(9)可得出理想和异构情况下的系统效率变化如图 3 中所示，在最好情况下系统使用效率为 1，理想状况下的系统有效性一直在 0.9 左右波动，而在异构情况下，系统有效性虽然随着异构节点的增多而缓慢下降，但仍接近于理想情况下的有效性，而且异构系统仍保持了较高的系统效率。因此，异构数据库集群系统在并行 OLTP 应用中具有较高的效费比。

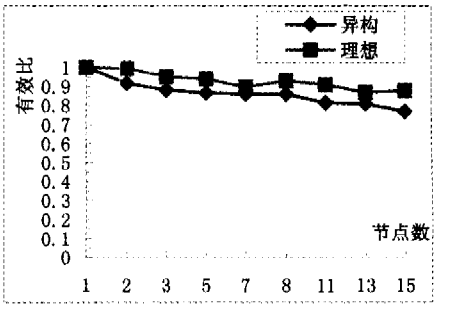

图 3 数据库异构集群系统有效性

以上 TPC-C 测试模拟了现实中 OLTP 商业应用，实验结果表明异构数据库集群在 OLTP 应用中具有良好的可扩展性，较理想的加速比，并且也达到了较高的系统使用效率指标。

结论 在 OLTP 商业环境中数据库集群是理想的数据库并行处理方案，其优点是不仅增加了 CPU 数，增大了内存，而且还增加了并行 I/O 带宽，该并行处理方式有效实现了高加速比和性价比。

本文分析了 OLTP 应用中数据库异构集群的性能模型，主要讨论了 CPU 和内存两种资源的异构带来性能影响，给出了评价异构系统并行性的三种重要指标，其中可扩展性描述了事务处理规模与系统增长之间的比例关系，而加速比则体现了系统对于给定规模的事务在系统规模增长后并行处理速度的提高，另外有效性也反映了节点在并行计算中的使用效率，从而完善了以前数据库集群研究中比较单一的性能衡量方法。

(下转第 142 页)

维普资讯 http://www.cqvip.com

方法类似于香农保密系统无条件安全的分析。它便于了解一个安全系统中各个部分之间熵的关系。而后一种方法是利用了相关熵和假设检验的概念，不仅可以用来定义无条件安全隐藏系统，还可以用来分析 θ-安全系统。与香农熵相比，相关熵不方便反映系统各个部分之间的关系。但它们对隐藏的模型的研究和安全隐写系统的设计都有指导意义，都反映出了设计一个无条件安全隐藏系统应该注意的几个方面。

(1)一个安全的隐写系统应该能抵抗被动攻击。即不但不能让攻击者检测到嵌入消息的内容，而且不能检测到秘密消息存在。这才算一个安全的隐写系统。

(2)载体信息 C 不能暴露给攻击者，即必须满足(6)式。

(3)隐写系统的安全性不能依赖于隐写算法。而应该依赖于密钥 K 的保密性。要满足(7)式，避免对密钥进行攻击。

(4)应尽可能地保证隐文消息 S 和载体消息 C，在统计概率分布上的不可区分。对于无条件安全隐写系统，要保证其严格的统计概率分布上的不可区分。

结论 信息隐藏是信息安全领域中一个新兴学科。本文从信息论的观点出发，对信息隐藏系统的安全性进行了定义和分析。指出了设计一个安全信息隐藏系统应注意的几个问题，对隐藏系统的安全性研究和设计有一定的指导意义。

参 考 文 献

1 Simmons G J. The prisoners' problem and the subliminal channel. In: Advances in Cryptology: Proc. of Crypto 83, Plenum Press, 1984. 51～67

2 ollner J Z, Federrath H, Klimant H, et al. Modeling the security of steganographic systems. In: 2nd Intl. Workshop on Information Hiding, LNCS, Springer, 1998. 344～354

3 Anderson R J, Petitcolas F A P. On The Limits of Steganography. IEEE Journal of Selected Areas in Communications, 1998, 16(4): 474～481

4 Shannon C E. Communication theory of secrecy systems. Bell System Technical Journal, 1949, 28(10): 656～715

5 P--tzmann B. Information hiding terminology. In: First Intl. Workshop on Information Hiding., LNCS 1174, springer, 1996

6 常迥. 信息理论基础. 北京：清华大学出版社，2001

7 冯登国. 密码学导引. 北京：科学出版社，1999

8 Cachin C. An Information Theoretic model for Steganography. In: Proc. of the Second Intl. Workshop on Information Hiding. LNCS 1525, Springer, 1998. 306～318

(上接第 108 页)

以上 TPC-C 测试的结果表明数据库异构集群在 OLTP 处理中仍具有良好的可扩展性，次线性的加速比，而且还能提供高效费比的并行处理服务。同时我们也发现异构节点的计算能力在集群中考虑网络开销时会有一定的影响，因此异构节点间的负载是不均匀的，在一定的网络环境下，异构节点的数量与系统性能存在一个最佳平衡点。

目前，我们只假设机器的 CPU 和内存硬件配置的异构，对于软件包括数据库在内的异构我们需要再做进一步的研究。

参 考 文 献

1 邸烁，郑纬民，王鼎兴，沈美明. 并行 WWW 服务器集群请求分配算法的研究. 软件学报，1999，10(7)

2 Baker M, Apon A, Buyya R, Jin H. Cluster computing and applications. Encyclopedia of Computer Science and Technology. New York: Marcel Dekker, Aug. 2001, 45

3 Gancarski S, Naacke H, Pacitti E, Valduriez P. Parallel processing with autonomous databases in a cluster system. In: Proc. of on the Move to Meaningful Internet Systems, DOA, CoopIS and ODBASE Confederated Intl. Conf. 2002. Oct. 2002

4 Fox A, Gribble S D, Chawathe Y, et al. Cluster-based scalable network services. ACM SIGOPS Operating Systems Review, 1997, 32(5)

5 Carrera E V, Bianchini R. Efficiency vs. Portability in Cluster-Based Network Servers. In: Proc. of the eighth ACM SIGPLAN symposium on Principles and practices of parallel programming, June 2001

6 Shen Kai, Yang Tao, Chu Lingkun. Cluster load balancing for fine-grain network services. In: Proc. of the Intl. Parallel and Distributed Processing Symposium, April 2002

7 Carrera E V, Bianchini R. Evaluating Cluster-Based Network Servers. In: Proc. of the Ninth IEEE Intl. Symposium on High Performance Distributed Computing(HPDC'00), Aug. 2000

8 Pastor L, Orero J L B. An efficiency and scalability model for heterogeneous clusters. In: Proc. of the 2001 IEEE Intl. Conf. on Cluster Computing, 2002. 427～434

9 Cuellar G D, Salzar D A. A parallelization technique that improves performance and cluster utilization efficiency for heterogeneous clusters of workstations. In: Proc. of the IEEE Intl. Conf. on Cluster Computing, 2002. 275～283

10 Kalinov A. Scalability analysis of matrix-matrix multiplication on heterogeneous clusters. In: The Third Intl. Workshop on Parallel and Distributed Computing, July 2004. 303～309

11 Bosque J L, Perez L P. Theoretical scalability analysis for heterogeneous clusters. In: IEEE Intl. Symposium on Cluster Computing and the Grid, April 2004. 285～292

12 Teodoro G, Tavares T, Coutinho B, et al. Load balancing on stateful clustered web servers. In: Proc. of the 15th Symposium on Computer Architecture and High Performance Computing, Nov. 2003

13 Gunther N J. Issues facing commercial OLTP applications on MPP platforms. IEEE 1994

14 蒋江，张民选，廖湘科. 异构集群系统中一种基于资源的负载平衡算法的设计与模拟. 小型微型计算机系统，2003，24(4)

15 Xiao Li, Zhang Xiaodong, Qu Yanxia. Effective load sharing on heterogeneous networks of workstations. In: Proc. of the 14th Intl. Parallel and Distributed Processing Symposium, May 2000. 431～438

16 Wang Min, Ding Weiqun, Li Hong. E-differentiation for analyzing scalability of parallel algorithms on parallel architectures. In: Proc. of Intl. Conf. on Information, Communications and Signal Processing(ICICS), Singapore, Sept. 1997

17 Ji Yongchang, An Hong, Ding Weiqun, Chen Guoliang. A scalability metric for algorithm-machine on NOW and MPP. IEEE 2000