



(12)发明专利

(10)授权公告号 CN 106686081 B

(45)授权公告日 2020.08.28

(21)申请号 201611248733.2

审查员 温丽丽

(22)申请日 2016.12.29

(65)同一申请的已公布的文献号

申请公布号 CN 106686081 A

(43)申请公布日 2017.05.17

(73)专利权人 北京奇虎科技有限公司

地址 100088 北京市西城区新街口外大街
28号D座112室(德胜园区)

(72)发明人 刘臻

(74)专利代理机构 北京市隆安律师事务所

11323

代理人 权鲜枝 何立春

(51)Int.Cl.

H04L 29/08(2006.01)

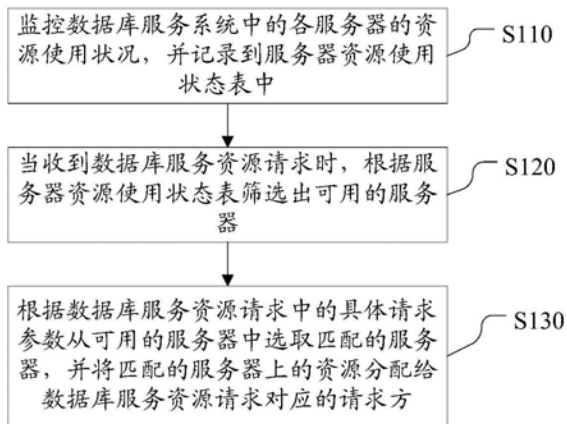
权利要求书2页 说明书10页 附图1页

(54)发明名称

一种数据库服务系统的资源分配方法和装置

(57)摘要

本发明公开了一种数据库服务系统的资源分配方法和装置。所述方法包括：监控数据库服务系统中的各服务器的资源使用状况，并记录到服务器资源使用状态表中；当收到数据库服务资源请求时，根据服务器资源使用状态表筛选出可用的服务器；根据数据库服务资源请求中的具体请求参数从可用的服务器中选取匹配的服务器，并将匹配的服务器上的资源分配给数据库服务资源请求对应的请求方。上述技术方案可以全面分析各个服务器的资源使用状态，实现数据库资源的合理分配，避免数据库资源的浪费。



1. 一种数据库服务系统的资源分配方法,包括:
监控数据库服务系统中的各服务器的资源使用状况,并记录到服务器资源使用状态表中;
当收到数据库服务资源请求时,根据所述服务器资源使用状态表筛选出可用的服务器;
根据所述数据库服务资源请求中的具体请求参数从可用的服务器中选取匹配的服务器,并将匹配的服务器上的资源分配给所述数据库服务资源请求对应的请求方;
该方法进一步包括:将服务器上的各项资源划分成多份,其中,每份中包含的各项资源之间的比例一致;
所述将匹配的服务器上的资源分配给所述数据库服务资源请求对应的请求方包括:将匹配的服务器上的资源以所划分的份为单位分配给所述数据库服务资源请求对应的请求方。
2. 如权利要求1所述的方法,其中,所述监控数据库服务系统中的各服务器的资源使用状况,并记录到服务器资源使用状态表中包括:
每隔固定时间,分别统计各服务器在过去的预设时间段内的平均资源使用率,并记录到服务器资源使用状态表中。
3. 如权利要求2所述的方法,其中,所述分别统计各服务器在过去的预设时间段内的平均资源使用率包括:
对于一个服务器,统计该服务器的如下一种或多种资源的平均使用率:
统计在过去的预设时间段内网卡的流入和流出流量的平均值占网卡标称值的百分比;
统计在过去的预设时间段内的内存的使用率;
统计在过去的预设时间段内的磁盘的使用率;
统计在过去的预设时间段内的负载率;
统计在过去的预设时间段内的CPU使用率。
4. 如权利要求1所述的方法,其中,所述根据所述服务器资源使用状态表筛选出可用的服务器包括:
为服务器的各项资源使用率分别设置阈值;
根据所述服务器资源使用状态表中所记录的各服务器的资源使用状况,筛选出各项资源使用率均低于相应阈值的服务器作为可用的服务器。
5. 如权利要求1所述的方法,其中,根据所述数据库服务资源请求中的具体请求参数从可用的服务器中的选取匹配的服务器包括:
从可用的服务器中选取剩余的可分配资源量最接近于所述数据库服务资源请求中的具体请求参数的服务器;
其中,所述剩余的可分配资源量是所述可用的服务器的理论的可分配的资源量。
6. 如权利要求1所述的方法,其中,根据所述数据库服务资源请求中的具体请求参数从可用的服务器中的选取匹配的服务器包括:
从可用的服务器中选取各项剩余可分配资源之间的比例与所述数据库服务资源请求中的具体请求参数之间的比例匹配的服务器。
7. 一种数据库服务系统的资源分配装置,包括:

监控记录单元,适于监控数据库服务系统中的各服务器的资源使用状况,并记录到服务器资源使用状态表中;

存储单元,适于保存所述服务器资源使用状态表;

资源分配单元,适于当收到数据库服务资源请求时,根据所述服务器资源使用状态表筛选出可用的服务器;根据所述数据库服务资源请求中的具体请求参数从可用的服务器中选取匹配的服务器,并将匹配的服务器上的资源分配给所述数据库服务资源请求对应的请求方;

所述资源分配单元,进一步适于将服务器上的各项资源划分成多份,其中,每份中包含的各项资源之间的比例一致;进而将匹配的服务器上的资源以所划分的份为单位分配给所述数据库服务资源请求对应的请求方。

8.如权利要求7所述的装置,其中,

所述监控记录单元,适于每隔固定时间,分别统计各服务器在过去的预设时间段内的平均资源使用率,并记录到服务器资源使用状态表中。

9.如权利要求8所述的装置,其中,

所述监控记录单元,适于对于一个服务器,统计该服务器的如下一种或多种资源的平均使用率:

统计在过去的预设时间段内网卡的流入和流出流量的平均值占网卡标称值的百分比;

统计在过去的预设时间段内的内存的使用率;

统计在过去的预设时间段内的磁盘的使用率;

统计在过去的预设时间段内的负载率;

统计在过去的预设时间段内的CPU使用率。

10.如权利要求7所述的装置,其中,

所述资源分配单元,适于为服务器的各项资源使用率分别设置阈值;根据所述服务器资源使用状态表中所记录的各项服务器的资源使用状况,筛选出各项资源使用率均低于相应阈值的服务器作为可用的服务器。

11.如权利要求7所述的装置,其中,

所述资源分配单元,适于从可用的服务器中选取剩余的可分配资源量最接近于所述数据库服务资源请求中的具体请求参数的服务器;

其中,所述剩余的可分配资源量是所述可用的服务器的理论的可分配的资源量。

12.如权利要求7所述的装置,其中,

所述资源分配单元,适于从可用的服务器中选取各项剩余可分配资源之间的比例与所述数据库服务资源请求中的具体请求参数之间的比例匹配的服务器。

一种数据库服务系统的资源分配方法和装置

技术领域

[0001] 本发明涉及计算机技术领域,具体涉及一种数据库服务系统的资源分配方法和装置。

背景技术

[0002] 在计算机技术领域中,根据业务需求需要在数据库系统中选择一定的服务器为该业务配置相应的数据库资源。现有技术中,在进行数据库的选择的时候,通过人工的方法,根据服务器的负载情况、空闲资源等状态进行选择可用的服务器。但是,通过人工方法进行选择,存在人为因素,不能全面分析各个服务器的状态,会造成服务器中的数据库资源的分配不合理,导致数据库资源的浪费。

发明内容

[0003] 鉴于上述问题,提出了本发明以便提供一种克服上述问题或者至少部分地解决上述问题的一种数据库服务系统的资源分配方法和装置。

[0004] 依据本发明的一个方面,提供了一种数据库服务系统的资源分配方法,包括:

[0005] 监控数据库服务系统中的各服务器的资源使用状况,并记录到服务器资源使用状态表中;

[0006] 当收到数据库服务资源请求时,根据所述服务器资源使用状态表筛选出可用的服务器;

[0007] 根据所述数据库服务资源请求中的具体请求参数从可用的服务器中选取匹配的服务器,并将匹配的服务器上的资源分配给所述数据库服务资源请求对应的请求方。

[0008] 可选地,所述监控数据库服务系统中的各服务器的资源使用状况,并记录到服务器资源使用状态表中包括:

[0009] 每隔固定时间,分别统计各服务器在过去的预设时间段内的平均资源使用率,并记录到服务器资源使用状态表中。

[0010] 可选地,所述分别统计各服务器在过去的预设时间段内的平均资源使用率包括:

[0011] 对于一个服务器,统计该服务器的如下一种或多种资源的平均使用率:

[0012] 统计在过去的预设时间段内网卡的流入和流出流量的平均值占网卡标称值的百分比;

[0013] 统计在过去的预设时间段内的内存的使用率;

[0014] 统计在过去的预设时间段内的磁盘的使用率;

[0015] 统计在过去的预设时间段内的负载率;

[0016] 统计在过去的预设时间段内的CPU使用率。

[0017] 可选地,所述根据所述服务器资源使用状态表筛选出可用的服务器包括:

[0018] 为服务器的各项资源使用率分别设置阈值;

[0019] 根据所述服务器资源使用状态表中所记录的各服务器的资源使用状况,筛选出各

项资源使用率均低于相应阈值的服务器作为可用的服务器。

[0020] 可选地,根据所述数据库服务资源请求中的具体请求参数从可用的服务器中的选取匹配的服务器包括:

[0021] 从可用的服务器中选取剩余的可分配资源量最接近于所述数据库服务资源请求中的具体请求参数的服务器。

[0022] 可选地,根据所述数据库服务资源请求中的具体请求参数从可用的服务器中的选取匹配的服务器包括:

[0023] 从可用的服务器中选取各项剩余可分配资源之间的比例与所述数据库服务器资源请求中的具体请求参数之间的比例匹配的服务器。

[0024] 可选地,所述将匹配的服务器上的资源分配给所述数据库服务资源请求对应的请求方包括:

[0025] 根据匹配的服务器上已分配资源的数量,统计剩余的可分配资源量;

[0026] 从所述剩余的可分配资源量中划分相应数量的资源分配给所述数据库服务资源请求对应的请求方。

[0027] 可选地,该方法进一步包括:将服务器上的各项资源划分成多份,其中,每份中包含的各项资源之间的比例一致;

[0028] 所述将匹配的服务器上的资源分配给所述数据库服务资源请求对应的请求方包括:将匹配的服务器上的资源以所划分的份为单位分配给所述数据库服务资源请求对应的请求方。

[0029] 根据本发明的另一方面,提供了一种数据库服务系统的资源分配装置,其中,该装置包括:

[0030] 监控记录单元,适于监控数据库服务系统中的各服务器的资源使用状况,并记录到服务器资源使用状态表中;

[0031] 存储单元,适于保存所述服务器资源使用状态表;

[0032] 资源分配单元,适于当收到数据库服务资源请求时,根据所述服务器资源使用状态表筛选出可用的服务器;根据所述数据库服务资源请求中的具体请求参数从可用的服务器中选取匹配的服务器,并将匹配的服务器上的资源分配给所述数据库服务资源请求对应的请求方。

[0033] 可选地,所述监控记录单元,适于每隔固定时间,分别统计各服务器在过去的预设时间段内的平均资源使用率,并记录到服务器资源使用状态表中。

[0034] 可选地,所述监控记录单元,适于对于一个服务器,统计该服务器的如下一种或多种资源的平均使用率:

[0035] 统计在过去的预设时间段内网卡的流入和流出流量的平均值占网卡标称值的百分比;

[0036] 统计在过去的预设时间段内的内存的使用率;

[0037] 统计在过去的预设时间段内的磁盘的使用率;

[0038] 统计在过去的预设时间段内的负载率;

[0039] 统计在过去的预设时间段内的CPU使用率。

[0040] 可选地,所述资源分配单元,适于为服务器的各项资源使用率分别设置阈值;根据

所述服务器资源使用状态表中所记录的各服务器的资源使用状况,筛选出各项资源使用率均低于相应阈值的服务器作为可用的服务器。

[0041] 可选地,所述资源分配单元,适于从可用的服务器中选取剩余的可分配资源量最接近于所述数据库服务资源请求中的具体请求参数的服务器。

[0042] 可选地,所述资源分配单元,适于从可用的服务器中选取各项剩余可分配资源之间的比例与所述数据库服务器资源请求中的具体请求参数之间的比例匹配的服务器。

[0043] 可选地,所述资源分配单元,适于根据匹配的服务器上已分配资源的数量,统计剩余的可分配资源量;从所述剩余的可分配资源量中划分相应数量的资源分配给所述数据库服务资源请求对应的请求方。

[0044] 可选地,所述资源分配单元,进一步适于将服务器上的各项资源划分成多份,其中,每份中包含的各项资源之间的比例一致;进而将匹配的服务器上的资源以所划分的份为单位分配给所述数据库服务资源请求对应的请求方。

[0045] 根据本发明的技术方案实时监控数据库服务系统中的各服务器的资源使用状况,当收到数据库服务资源请求时,先根据监控的服务器资源使用状态筛选出可用的服务器;但是选出可用的服务器后并不意味着该数据库资源请求都可以使用,还需要根据数据库服务资源请求中的具体请求参数从可用的服务器中选取匹配的服务器,然后将匹配的服务器上的资源分配给数据库服务资源请求对应的请求方,自动实现根据服务器资源使用状态,并结合数据库服务资源请求的具体请求参数从数据库服务系统选择最优的服务器,以供数据库服务资源请求对应的请求方使用。可见,本发明可以全面分析各个服务器的资源使用状态,实现数据库资源的合理分配,避免数据库资源的浪费。

[0046] 上述说明仅是本发明技术方案的概述,为了能够更清楚了解本发明的技术手段,而可依照说明书的内容予以实施,并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂,以下特举本发明的具体实施方式。

附图说明

[0047] 通过阅读下文优选实施方式的详细描述,各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的,而并不认为是对本发明的限制。而且在整个附图中,用相同的参考符号表示相同的部件。在附图中:

[0048] 图1示出了根据本发明一个实施例的数据库服务系统的资源分配方法的流程图;

[0049] 图2示出了根据本发明一个实施例的数据库服务系统的资源分配装置的结构示意图。

具体实施方式

[0050] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例,然而应当理解,可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反,提供这些实施例是为了能够更透彻地理解本公开,并且能够将本公开的范围完整的传达给本领域的技术人员。

[0051] 图1示出了根据本发明一个实施例的数据库服务系统的资源分配方法的流程图。如图1所示,该方法包括:

[0052] 步骤S110,监控数据库服务系统中的各服务器的资源使用状况,并记录到服务器资源使用状态表中。

[0053] 这里监控数据库服务系统中的各服务器的资源使用状况,是监控各服务器的多种资源项目的使用状况,例如,内存的使用状况、磁盘的使用状况、CPU的使用状况、繁忙程度(load值)、是否过保等状况信息,并且将各个监控项目的状况记录到一个服务器资源使用状态表中。从各个方面记录各服务器的状况,以及共记录中分析服务器的整体状态,以便在接收到数据库服务资源请求时进行服务器的最合理的选择。

[0054] 步骤S120,当收到数据库服务资源请求时,根据服务器资源使用状态表筛选出可用的服务器。

[0055] 这里的根据服务器资源使用状态表筛选出可用的服务器是针对服务器自身而言,可用的服务器是指可能再接受资源需求的服务器。例如,服务器A的CPU使用率已经达到80%,就不可能再接收任何资源需求了;而服务器B的CPU使用率才30%,那么先不考虑数据库服务资源请求的具体参数如何,该服务器B是存在接收资源需求的可能的,因此服务器B被视为可用的服务器。

[0056] 步骤S130,根据数据库服务资源请求中的具体请求参数从可用的服务器中选取匹配的服务器,并将匹配的服务器上的资源分配给数据库服务资源请求对应的请求方。

[0057] 当根据服务器资源使用状态表筛选出可用的服务器后,并不是都满足数据库服务资源请求,例如可用的服务器的磁盘空间还剩下100G,而数据库服务资源请求需要150G的磁盘空间,那么这个服务器是不符合该请求的要求的。所以,还需要再根据数据库服务资源请求中的具体请求参数从可用的服务器中选取匹配的服务器。这里的匹配的服务器,一方面是服务器各项资源符合数据库服务资源请求中的具体请求参数的要求,例如,数据库服务资源请求中的磁盘空间是100G,那么需要从可用的服务器中选择出剩余磁盘空间最接近100G的服务器;另一方面,为了最合理的利用资源,选择的服务器的各项剩余可分配资源之间的比例必须与数据库服务器资源请求中的具体请求参数之间的比例相符合。例如,数据库服务器资源请求中的CPU使用率是30%、网卡的流入和流出流量为20兆,那么,就需要从可用的服务器中选择出剩余CPU使用率和网卡的流入和流出流量之间的比例为30%:20兆的服务器,如果选择的服务器的CPU的剩余使用率是50%,而网卡的是20兆,那么这里的CPU的剩余使用率50%相比较30%来说。所出来的20%就是资源的浪费。所以为了最合理的利用资源,需要综合上述两个方面,才能从可用的服务器其中选择出与数据库服务资源请求相匹配的服务器。例如,为满足上述数据库服务资源请求,选择出的服务器必须同时满足剩余磁盘空间大于100G以及剩余CPU使用率和网卡的流入和流出流量之间的比例为30%:20兆。

[0058] 可见,本发明自动实现根据服务器资源使用状态,并结合数据库服务资源请求的具体请求参数从数据库服务系统选择最优的服务器,以供数据库服务资源请求对应的请求方使用,也就是说,本发明可以全面分析各个服务器的资源使用状态,实现数据库资源的合理分配,避免数据库资源的浪费。

[0059] 数据库服务系统中的各服务器的资源使用状况是存在上下浮动的,例如当前时间的服务器的CPU使用率是70%,下一秒就变成了60%。为了合理记录数据库服务系统中的各服务器的资源使用状况,在本发明的一个实施例中,监控数据库服务系统中的各服务器的

资源使用状况,并记录到服务器资源使用状态表中包括:每隔固定时间,分别统计各服务器在过去的预设时间段内的平均资源使用率,并记录到服务器资源使用状态表中。例如,设定固定时间为1小时,那么每隔1小时,分别统计各服务器在过去的1小时内的平均资源使用率,并记录到服务器资源使用状态表中。

[0060] 在一个具体的例子中,过去的1小时内,服务器的CPU使用率分别是20分钟的时间是60%,20分钟的时间是30%,20分钟的时间是70%。那么过去的1小时内,服务器的CPU平均使用率约是53%。

[0061] 在本发明的一个实施例中,分别统计各服务器在过去的预设时间段内的平均资源使用率包括:对于一个服务器,统计该服务器的如下一种或多种资源的平均使用率:统计在过去的预设时间段内网卡的流入和流出流量的平均值占网卡标称值的百分比;统计在过去的预设时间段内的内存的使用率;统计在过去的预设时间段内的磁盘的使用率;统计在过去的预设时间段内的负载率;统计在过去的预设时间段内的CPU使用率。

[0062] 需要说明的是,上述列举的各资源是本发明的一个实施例,但是本发明针对统计的资源类型不作具体限定。

[0063] 根据数据库服务资源请求中的具体请求参数选取匹配的服务器之前,先从众多的服务器中选择出可用的服务器。在本发明的一个实施例中,步骤S120中的根据服务器资源使用状态表筛选出可用的服务器包括:为服务器的各项资源使用率分别设置阈值;根据服务器资源使用状态表中所记录的各项服务器的资源使用状况,筛选出各项资源使用率均低于相应阈值的服务器作为可用的服务器。

[0064] 例如,设置内存的使用率预设阈值为40%,以及设定CPU的使用率的预设阈值是30%,只有满足内存的使用率小于40%,同时满足CPU的使用率小于30%的服务器才被视为是可用的服务器。

[0065] 当选择出可用的服务器后,再根据数据库服务资源请求中的具体请求参数从可用的服务器中的选取匹配的服务器,具体包括:从可用的服务器中选取剩余的可分配资源量最接近于数据库服务资源请求中的具体请求参数的服务器。

[0066] 这里的剩余的可分配的资源量值的是该服务器的理论的可分配的资源量,并不是实际的剩余的可分配资源量,因为,通常情况下请求方进行资源请求时会根据该请求运行后的最大使用情况进行资源的请求,但是实际运行中,该请求的资源使用情况要小于其请求的资源的大小,这就造成了服务器的实际的剩余的可分配资源量大于理论的可分配的资源量。例如,服务器的磁盘空间是300G,其中已经承担了一个业务,该业务请求的资源100G,但是在该业务的实际运行中其占用的磁盘空间只有50G,那么这是该服务器的理论剩余的可分配的资源量是200G,实际剩余的可分配的资源量是250G。但是。因为已该服务器已经承担的业务会不断的更新,可能会达到其最初申请的资源量,例如上述例子中的最初申请的100G的磁盘空间,所以在根据一个新的数据库服务资源请求中的具体请求参数从可用的服务器中的选取匹配的服务器的时候,要根据服务器中的理论剩余的可分配资源量进行选择,例如上述例子中要根据服务器的200G的剩余的可分配资源量进行选择。

[0067] 那么当确定服务器的理论剩余的可分配的资源量后,就可以进行匹配了。例如,数据库服务资源请求中的磁盘空间是100G,那么为了最合理利用资源需要从可用的服务器中选择出理论的剩余的可分配的磁盘空间最接近的100G的服务器(比如110G)。

[0068] 上述实施例是服务器各项资源符合数据库服务资源请求中的具体请求参数的要求,为了最大化的利用数据库资源,根据数据库服务资源请求中的具体请求参数从可用的服务器中的选取匹配的服务器的时候还需要根据服务器的各项剩余可分配资源之间的比例必须与数据库服务器资源请求中的具体请求参数之间的比例相符合。在本发明的一个实施例中,步骤S130中的根据数据库服务资源请求中的具体请求参数从可用的服务器中的选取匹配的服务器包括:从可用的服务器中选取各项剩余可分配资源之间的比例与数据库服务器资源请求中的具体请求参数之间的比例匹配的服务器。

[0069] 为了最合理的利用资源,选择的服务器各项剩余可分配资源之间的比例必须与数据库服务器资源请求中的具体请求参数之间的比例相符合。例如,数据库服务器资源请求中的磁盘空间是50G、网卡的流入和流出流量为20兆,那么,就需要从可用的服务器中选择出剩余磁盘空间和网卡的流入和流出流量之间的相对比例为50G:20兆的服务器,如果选择的服务器的剩余磁盘空间是60G,而网卡的是20兆,那么这里的剩余磁盘空间是60G相比较50G来说,多出来的10G就是资源的浪费的。为了最合理的;利用资源,选择的服务器的剩余磁盘空间是50G,而网卡的是20兆即可。

[0070] 选择匹配的服务器的时候是根据服务器资源使用状态表中的记录进行选择,而是实际进行资源分配的时候,还需要进行可分配资源的计算,然后再进行分配。在本发明的一个实施例中,步骤S130中的将匹配的服务器上的资源分配给数据库服务资源请求对应的请求方包括:根据匹配的服务器上已分配资源的数量,统计剩余的可分配资源量;从剩余的可分配资源量中划分相应数量的资源分配给数据库服务资源请求对应的请求方。

[0071] 那么为了便于资源的分配,在本发明的一个实施例中,图1所示的方法进一步包括:将服务器上的各项资源划分成多份,其中,每份中包含的各项资源之间的比例一致,当接收到数据库服务资源请求后,根据数据库服务资源请求个具体参数之间的比例关系,以及已经分份的服务器的份数信息直接在可用的服务器中进行选择。

[0072] 例如,一个服务器的内存为100G,磁盘空间是500G,均分成10份,那么每份包含10G内存和50G的磁盘空间,

[0073] 步骤S130中的将匹配的服务器上的资源分配给数据库服务资源请求对应的请求方包括:将匹配的服务器上的资源以所划分的份为单位直接分配给数据库服务资源请求对应的请求方,不需要再进行统计剩余的可分配资源量的过程。上述例子中,当一个数据库服务资源请求中的内存和磁盘空间的比例关系符合10G内存和50G的磁盘空间比例关系的时候,无需再进行该服务器中的剩余资源的统计,将该份10G内存和50G的磁盘空间资源直接分配给该数据库服务资源请求。

[0074] 图2示出了根据本发明一个实施例的数据库服务系统的资源分配装置的结构示意图。如图2所示,该数据库服务系统的资源分配装置200包括:

[0075] 监控记录单元210,适于监控数据库服务系统中的各服务器的资源使用状况,并记录到服务器资源使用状态表中。

[0076] 这里监控数据库服务系统中的各服务器的资源使用状况,是监控各服务器的多种资源项目的使用状况,例如,内存的使用状况、磁盘的使用状况、CPU的使用状况、繁忙程度(load值)、是否过保等状况信息,并且将各个监控项目的状况记录到一个服务器资源使用状态表中。从各个方面记录各服务器的状况,以及共记录中分析服务器的整体状态,以便在

接收到数据库服务资源请求时进行服务器的最合理的选择。

[0077] 存储单元220,适于保存服务器资源使用状态表。

[0078] 资源分配单元230,适于当收到数据库服务资源请求时,根据服务器资源使用状态表筛选出可用的服务器;根据数据库服务资源请求中的具体请求参数从可用的服务器中选取匹配的服务器,并将匹配的服务器上的资源分配给数据库服务资源请求对应的请求方。

[0079] 这里的根据服务器资源使用状态表筛选出可用的服务器是针对服务器自身而言,可用的服务器是指可能再接受资源需求的服务器。例如,服务器A的CPU使用率已经达到80%,就不可能再接收任何资源需求了;而服务器B的CPU使用率才30%,那么先不考虑数据库服务资源请求的具体参数如何,该服务器B是存在接收资源需求的可能的,因此服务器B被视为可用的服务器。

[0080] 当根据服务器资源使用状态表筛选出可用的服务器后,并不是都满足数据库服务资源请求,例如可用的服务器的磁盘空间还剩下100G,而数据库服务资源请求需要150G的磁盘空间,那么这个服务器是不符合该请求的要求的。所以,还需要再根据数据库服务资源请求中的具体请求参数从可用的服务器中选取匹配的服务器。这里的匹配的服务器,一方面是服务器各项资源符合数据库服务资源请求中的具体请求参数的要求,例如,数据库服务资源请求中的磁盘空间是100G,那么需要从可用的服务器中选择出剩余磁盘空间最接近100G的服务器;另一方面,为了最合理的利用资源,选择的服务器的各项剩余可分配资源之间的比例必须与数据库服务器资源请求中的具体请求参数之间的比例相符合。例如,数据库服务器资源请求中的CPU使用率是30%、网卡的流入和流出流量为20兆,那么,就需要从可用的服务器中选择出剩余CPU使用率和网卡的流入和流出流量之间的比例为30%:20兆的服务器,如果选择的服务器的CPU的剩余使用率是50%,而网卡的是20兆,那么这里的CPU的剩余使用率50%相比较30%来说。所出来的20%就是资源的浪费。所以为了最合理的利用资源,需要综合上述两个方面,才能从可用的服务器其中选择出与数据库服务资源请求相匹配的服务器。例如,为满足上述数据库服务资源请求,选择出的服务器必须同时满足剩余磁盘空间大于100G以及剩余CPU使用率和网卡的流入和流出流量之间的比例为30%:20兆。

[0081] 可见,本发明自动实现根据服务器资源使用状态,并结合数据库服务资源请求的具体请求参数从数据库服务系统选择最优的服务器,以供数据库服务资源请求对应的请求方使用,也就是说,本发明可以全面分析各个服务器的资源使用状态,实现数据库资源的合理分配,避免数据库资源的浪费。

[0082] 数据库服务系统中的各服务器的资源使用状况是存在上下浮动的,例如当前时间的服务器的CPU使用率是70%,下一秒就变成了60%。为了合理记录数据库服务系统中的各服务器的资源使用状况,在本发明的一个实施例中,监控记录单元210,适于每隔固定时间,分别统计各服务器在过去的预设时间段内的平均资源使用率,并记录到服务器资源使用状态表中。例如,设定固定时间为1小时,那么每隔1小时,分别统计各服务器在过去的1小时内的平均资源使用率,并记录到服务器资源使用状态表中。

[0083] 在一个具体的例子中,过去的1小时内,服务器的CPU使用率分别是20分钟的时间是60%,20分钟的时间是30%,20分钟的时间是70%。那么过去的1小时内,服务器的CPU平均使用率约是53%。

[0084] 在本发明的一个实施例中,监控记录单元,适于对于一个服务器,统计该服务器的如下一种或多种资源的平均使用率:统计在过去的预设时间段内网卡的流入和流出流量的平均值占网卡标称值的百分比;统计在过去的预设时间段内的内存的使用率;统计在过去的预设时间段内的磁盘的使用率;统计在过去的预设时间段内的负载率;统计在过去的预设时间段内的CPU使用率。

[0085] 需要说明的是,上述列举的各资源是本发明的一个实施例,但是本发明针对统计的资源类型不作具体限定。

[0086] 根据数据库服务资源请求中的具体请求参数选取匹配的服务器之前,先从众多的服务器中选择出可用的服务器。在本发明的一个实施例中,资源分配单元230,适于为服务器的各项资源使用率分别设置阈值;根据服务器资源使用状态表中所记录的各服务器的资源使用状况,筛选出各项资源使用率均低于相应阈值的服务器作为可用的服务器。

[0087] 例如,设置内存的使用率预设阈值为40%,以及设定CPU的使用率的预设阈值是30%,只有满足内存的使用率小于40%,同时满足CPU的使用率小于30%的服务器才被视为是可用的服务器。

[0088] 选择出可用的服务器后,资源分配单元230,适于从可用的服务器中选取剩余的可分配资源量最接近于数据库服务资源请求中的具体请求参数的服务器。

[0089] 这里的剩余的可分配的资源量值的是该服务器的理论的可分配的资源量,并不是实际的剩余的可分配资源量,因为,通常情况下请求方进行资源请求时会根据该请求运行后的最大使用情况进行资源的请求,但是实际运行中,该请求的资源使用情况要小于其请求的资源的大小,这就造成了服务器的实际的剩余的可分配资源量大于理论的可分配的资源量。例如,服务器的磁盘空间是300G,其中已经承担了一个业务,该业务请求的资源100G,但是在该业务的实际运行中其占用的磁盘空间只有50G,那么这是该服务器的理论剩余的可分配的资源量是200G,实际剩余的可分配的资源量是250G。但是,因为已该服务器已经承担的业务会不断的更新,可能会达到其最初申请的资源量,例如上述例子中的最初申请的100G的磁盘空间,所以在根据一个新的数据库服务资源请求中的具体请求参数从可用的服务器中的选取匹配的服务器的时候,要根据服务器中的理论剩余的可分配资源量进行选择,例如上述例子中要根据服务器的200G的剩余的可分配资源量进行选择。

[0090] 那么当确定服务器的理论剩余的可分配的资源量后,就可以进行匹配了。例如,数据库服务资源请求中的磁盘空间是100G,那么为了最合理利用资源需要从可用的服务器中选择出理论的剩余的可分配的磁盘空间最接近的100G的服务器(比如110G)。

[0091] 上述实施例是服务器各项资源符合数据库服务资源请求中的具体请求参数的要求,为了最大化的利用数据库资源,根据数据库服务资源请求中的具体请求参数从可用的服务器中的选取匹配的服务器的时候还需要根据服务器的各项剩余可分配资源之间的比例必须与数据库服务器资源请求中的具体请求参数之间的比例相符合。在本发明的一个实施例中,资源分配单元230,适于从可用的服务器中选取各项剩余可分配资源之间的比例与数据库服务器资源请求中的具体请求参数之间的比例匹配的服务器。

[0092] 为了最合理的利用资源,选择的服务器的各项剩余可分配资源之间的比例必须与数据库服务器资源请求中的具体请求参数之间的比例相符合。例如,数据库服务器资源请求中的磁盘空间是50G、网卡的流入和流出流量为20兆,那么,就需要从可用的服务器中选

择出剩余磁盘空间和网卡的流入和流出流量之间的相对比例为50G:20兆的服务器,如果选择的服务器的剩余磁盘空间是60G,而网卡的是20兆,那么这里的剩余磁盘空间是60G相比较50G来说,多出来的10G就是资源的浪费的。为了最合理的;利用资源,选择的服务器的剩余磁盘空间是50G,而网卡的是20兆即可。

[0093] 选择匹配的服务器的时候是根据服务器资源使用状态表中的记录进行选择,而是实际进行资源分配的时候,还需要进行可分配资源的计算,然后再进行分配。在本发明的一个实施例中,资源分配单元230,适于根据匹配的服务器上已分配资源的数量,统计剩余的可分配资源量;从剩余的可分配资源量中划分相应数量的资源分配给数据库服务资源请求对应的请求方。

[0094] 那么为了便于资源的分配,在本发明的一个实施例中,资源分配单元230,进一步适于将服务器上的各项资源划分成多份,其中,每份中包含的各项资源之间的比例一致,当接收到数据库服务资源请求后,根据数据库服务资源请求个具体参数之间的比例关系,以及已经分份的服务器的份数信息直接在可用的服务器中进行选择。那么,选择出匹配的服务器后进而将匹配的服务器上的资源以所划分的份为单位分配给数据库服务资源请求对应的请求方,不需要再进行统计剩余的可分配资源量的过程。

[0095] 例如,一个服务器的内存为100G,磁盘空间是500G,均分成10份,那么每份包含10G内存和50G的磁盘空间,当一个数据库服务资源请求中的内存和磁盘空间的比例关系符合10G内存和50G的磁盘空间比例关系的时候,无需再进行该服务器中的剩余资源的统计,将该份10G内存和50G的磁盘空间资源直接分配给该数据库服务资源请求。

[0096] 综上所述,根据本发明的技术方案实时监控数据库服务系统中的各服务器的资源使用状况,当收到数据库服务资源请求时,先根据监控的服务器资源使用状态筛选出可用的服务器;但是选出可用的服务器后并不意味着该数据库资源请求都可以使用,还需要根据数据库服务资源请求中的具体请求参数从可用的服务器中选取匹配的服务器,然后将匹配的服务器上的资源分配给数据库服务资源请求对应的请求方,自动实现根据服务器资源使用状态,并结合数据库服务资源请求的具体请求参数从数据库服务系统选择最优的服务器,以供数据库服务资源请求对应的请求方使用。可见,本发明可以全面分析各个服务器的资源使用状态,实现数据库资源的合理分配,避免数据库资源的浪费。

[0097] 需要说明的是:

[0098] 在此提供的算法和显示不与任何特定计算机、虚拟装置或者其它设备固有相关。各种通用装置也可以与基于在此的示教一起使用。根据上面的描述,构造这类装置所要求的结构是显而易见的。此外,本发明也不针对任何特定编程语言。应当明白,可以利用各种编程语言实现在此描述的本发明的内容,并且上面对特定语言所做的描述是为了披露本发明的最佳实施方式。

[0099] 在此处所提供的说明书中,说明了大量具体细节。然而,能够理解,本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中,并未详细示出公知的方法、结构和技术,以便不模糊对本说明书的理解。

[0100] 类似地,应当理解,为了精简本公开并帮助理解各个发明方面中的一个或多个,在上面对本发明的示例性实施例的描述中,本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而,并不应将该公开的方法解释成反映如下意图:即所要求保

护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说,如下面的权利要求书所反映的那样,发明方面在于少于前面公开的单个实施例的所有特征。因此,遵循具体实施方式的权利要求书由此明确地并入该具体实施方式,其中每个权利要求本身都作为本发明的单独实施例。

[0101] 本领域那些技术人员可以理解,可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件,以及此外可以把它分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元中的至少一些是相互排斥之外,可以采用任何组合对本说明书(包括伴随的权利要求、摘要和附图)中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述,本说明书(包括伴随的权利要求、摘要和附图)中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

[0102] 此外,本领域的技术人员能够理解,尽管在此所述的一些实施例包括其它实施例中包括的某些特征而不是其它特征,但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。例如,在下面的权利要求书中,所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

[0103] 本发明的各个部件实施例可以以硬件实现,或者以在一个或者多个处理器上运行的软件模块实现,或者以它们的组合实现。本领域的技术人员应当理解,可以在实践中使用微处理器或者数字信号处理器(DSP)来实现根据本发明实施例的数据库服务系统的资源分配装置中的一些或者全部部件的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序(例如,计算机程序和计算机程序产品)。这样的实现本发明的程序可以存储在计算机可读介质上,或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下下载得到,或者在载体信号上提供,或者以任何其他形式提供。

[0104] 应该注意的是上述实施例对本发明进行说明而不是对本发明进行限制,并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中,不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中,这些装置中的若干个可以是通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

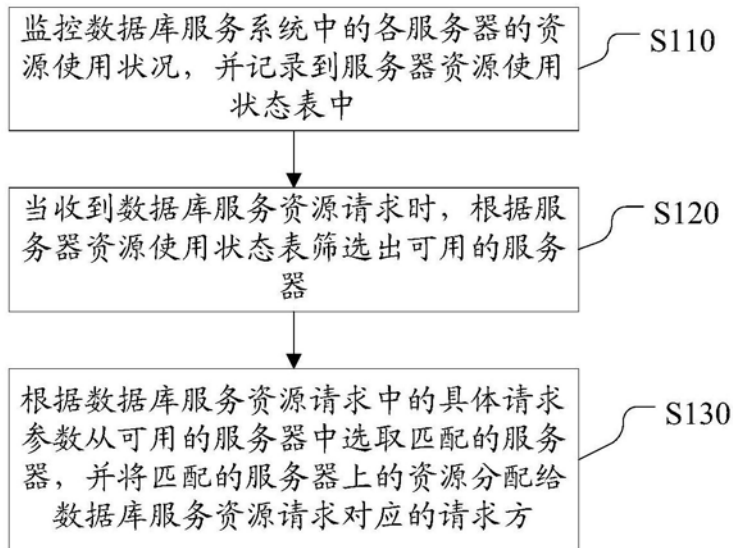


图1

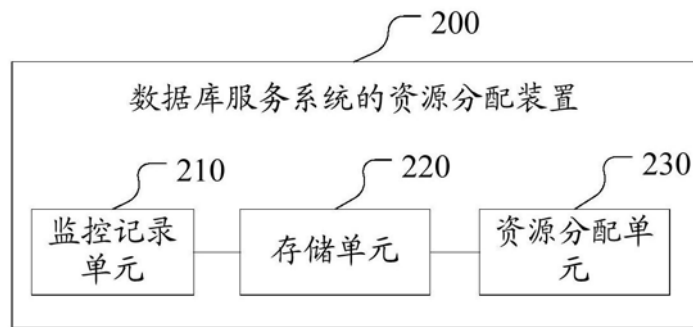


图2