(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 104077182 B
(45)授权公告日 2017.04.26

(21)申请号 201410308882.8

(22)申请日 2014.06.30

(65)同一申请的已公布的文献号
申请公布号 CN 104077182 A

(43)申请公布日 2014.10.01

(73)专利权人 西安交通大学
地址 710049 陕西省西安市咸宁西路28号

(72)发明人 伍卫国 李谦 周夏心 黄舰航 张译之 王蕾

(74)专利代理机构 西安通大专利代理有限责任公司 61200
代理人 陆万寿

(51)Int.Cl.
G06F 9/46(2006.01)

(56)对比文件
CN 103677752 A,2014.03.26,
CN 103729480 A,2014.04.16,
US 8056079 B1,2011.11.08,
徐正光,陈雁,尹怡欣,胡长军,王珏.一种基于梯形自调度技术的集群任务调度的实现.《计算机工程》.2005,第31卷(第23期),第63-89页.

审查员 田静

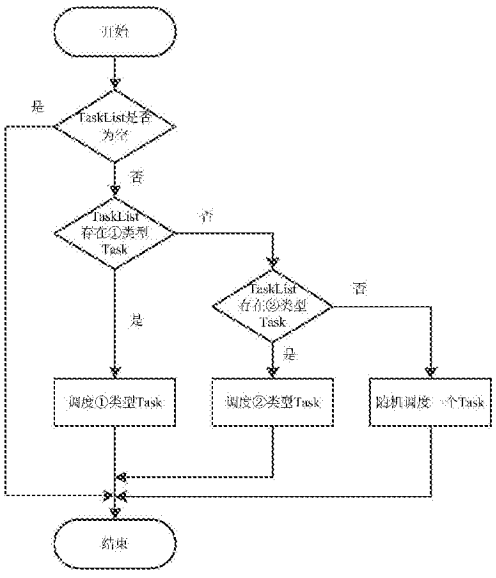
权利要求书1页 说明书5页 附图1页

(54)发明名称

一种同优先级任务调度策略

(57)摘要

本发明提供一种同优先级任务调度策略,定义两类任务,第一类任务是指该任务所对应的作业中有其他任务在已完成任务队列里,第二类任务是指该任务所对应的作业中有其他任务在正在执行任务队列里;在一个调度周期内,如果待调度任务队列中存在第一类任务,那么选择第一类任务进行调度;如果待调度任务队列中不存在第一类任务而存在第二类任务,那么选择第二类任务进行调度;如果待调度任务队列中不存在上述两类任务,那么随机选择一个任务进行调度,本发明缩短了作业的平均完成时间和作业的平均周转时间,提高了系统吞吐量。



CN 104077182 B

1.一种应用于集群作业管理系统的同优先级任务调度方法，其特征在于：包括以下步骤：定义两类任务，第一类任务是指该任务所对应的作业中有其他任务在已完成任务队列里，第二类任务是指该任务所对应的作业中有其他任务在正在执行任务队列里；在一个调度周期内，如果待调度任务队列中存在第一类任务，那么选择第一类任务进行调度；如果待调度任务队列中不存在第一类任务而存在第二类任务，那么选择第二类任务进行调度；如果待调度任务队列中不存在上述两类任务，那么随机选择一个任务进行调度。

2.根据权利要求1所述的方法，其特征在于：设n个作业具有相同的优先级，n个作业分别记为J_1，J_2，…，J_n，J_1，J_2，…，J_n被对应切分为x_1，x_2，…，x_n份任务，x_1＝x_2＝，…，＝x_n。

3.根据权利要求1所述的方法，其特征在于：由同一个作业切分得到的各个任务具有相同的单机执行时间。

一种同优先级任务调度策略

技术领域

[0001] 本发明属于计算机技术领域，具体涉及一种同优先级任务调度策略。

背景技术

[0002] 集群系统或称机群系统已经成为高性能计算机的主流计算平台，而集群作业管理则是保证集群高效运行的关键，也是集群应用的基础。同时，调度策略是集群作业管理系统的核心。一个好的调度算法不但可以减少作业的等待时间，缩短作业响应时间，还能够充分利用系统的资源，提高系统利用率。

[0003] 在集群作业管理系统中，作业往往是大规模的，这些作业一般被切分成若干个任务后才能被调度和执行。在这里，调度器的调度单位是任务而不是作业。这些任务按照任务调度器的调度策略被分派到各个计算节点进行执行。一般情况下，这些任务被分配到不同的计算节点进行执行。每个任务执行完成之后整个作业即执行完成。

[0004] 集群作业管理系统中任务调度的目标一般分为基于用户性能的目标和基于系统性能的目标。基于用户性能的调度目标一般包括作业的完成时间、作业的周转时间等。基于系统性能的调度目标一般包括资源利用率、系统吞吐量等。作业的完成时间是指作业中第一个任务开始执行到最后一个任务执行完成所经历的时间，即作业的执行时间。作业的周转时间是指作业从提交到全部完成的时间跨度，不仅包括执行时间还含有排队时间。资源利用率是指资源的忙闲程度或者资源使用百分比。系统吞吐量是指单位时间内集群系统完成的作业个数。

[0005] 在基于优先级的任务调度系统中，每个作业都有一个优先级参数，这个作业被切分成的各个任务的优先级与这个作业的优先级相同；不同的作业可以有相同的优先级。这样，在系统中，经常会出现很多个任务具有相同优先级的情况。在传统基于优先级的调度系统中，优先级高的任务先于优先级低的任务得到执行，而对于优先级相同的任务，则没有一个高效的调度策略，而只是随机或按顺序选取一个任务进行执行。可能出现这种情况，即某个作业的部分任务早已执行完毕，但是由于其他一些同优先级的非此作业的任务一直在执行，这个作业的其余任务隔了很长时间才得到执行，从而大大增加了整个作业的周转时间。

发明内容

[0006] 为了克服上述现有调度策略的缺点，本发明的目的在于提供一种同优先级任务调度策略，解决同优先级任务的高效调度问题。

[0007] 为了达到上述目的，本发明采取的技术方案为：

[0008] 定义两类任务，第一类任务是指该任务所对应的作业中有其他任务在已完成任务队列里，第二类任务是指该任务所对应的作业中有其他任务在正在执行任务队列里；在一个调度周期内，如果待调度任务队列中存在第一类任务，那么选择第一类任务进行调度；如果待调度任务队列中不存在第一类任务而存在第二类任务，那么选择第二类任务进行调度；如果待调度任务队列中不存在上述两类任务，那么随机选择一个任务进行调度。

[0009] 设n个作业具有相同的优先级，n个作业分别记为$J_1, J_2, \cdots, J_n, J_1, J_2, \cdots, J_n$被对应切分为$x_1, x_2, \cdots, x_n$份任务，$x_1 = x_2 = , \cdots, = x_n$。

[0010] 由同一个作业切分得到的各个任务具有相同的单机执行时间。

[0011] 本发明的有益效果是：

[0012] 本发明提出的同优先级任务调度策略，针对在优先级调度中经常会出现很多个任务具有相同优先级的情况，对于优先级相同的任务，首先选择存在已完成任务的作业优先调度，再选择正在执行任务的作业进行调度，如果不存在以上两类作业，最后随机选择一个作业调度，这样可以在不影响资源利用率的前提下尽量减少作业的完成时间、作业的周转时间，提高系统吞吐量。

附图说明

[0013] 图1是本发明所述同优先级任务调度策略的算法流程图。

具体实施方式

[0014] 下面结合附图和实施例对本发明作详细描述。

[0015] (一)集群系统中同优先级任务调度建模与分析

[0016] 作业(Job)：作业是用户请求资源的单位。

[0017] 任务(Task)：任务是作业通过切分后的一组子作业。一个作业被分为若干个任务。

[0018] 设n个作业具有相同的优先级，n个作业分别记为$J_1, J_2, \cdots, J_n$。n个作业被对应分为$x_1, x_2, \cdots, x_n$份任务。这些任务的优先级也是相同的。则任务集合为$T_{1,1}, T_{1,2}, \ldots, T_{1,x_1}, T_{2,1}, \ldots, T_{n,x_n}$，任务个数为$x_1+x_2+x_3+\cdots+x_n$。n个作业中任务的单机执行时间分别为$t_{1,1}, t_{1,2}, \ldots, t_{1,x_1}, t_{2,1}, \ldots, t_{n,x_n}$。

[0019] 为了简化这一模型，在这里做两个假定。假定每个作业被分为相同的份数x，则任务个数为n*x，任务集合为$T_{1,1}, T_{1,2}, \cdots, T_{1,x}, T_{2,1}, \cdots, T_{n,x}$。另外，假定同一作业中的各个任务的单机执行时间都相同。这样，建立了任务调度的简化版模型。

[0020] (二)集群系统中一个调度周期内同优先级任务调度策略如下：

[0021] 定义两类Task，第一类是指Task所对应的Job中有其他Task在已完成任务队列里的Task，第二类是指Task所对应的Job中有其他Task在正在执行任务队列里的Task。

[0022] 定义待调度任务队列TaskList＝$\{T_1, T_2, \cdots, T_v\}$，其中，$T_1, T_2, \cdots, T_v$具有相同的优先级，在待调度任务队列中，如果存在第一类Task，那么选择此Task进行调度；如果不存在第一类Task而存在第二类Task，那么选择第二类Task进行调度；如果上述两类Task都不存在，那么随机选择一个Task进行调度。参见图1，即：

[0023]

TaskList={T_1，T_2，…，T_v}为等待调度的相同优先级的任务队列；

If (TaskList 非空){

If(TaskList 中存在 T_i，T_i所对应的 Job 中有其他 Task 在已完成任务队列里)

选择 T_i 执行；

Else

if(TaskList 中存在 T_i，T_i所对应的 Job 中有其他 Task 在正在执行任务队列里)

选择 T_i 执行；

Else

随机选择一个 Task 执行；

}

[0024] (三)集群系统中同优先级任务调度策略分析

[0025] 假设J1，J2，J3，J4四个作业的优先级相同，这4个作业分别被分解成4个任务，分别为$T_{1,1}$，$T_{1,2}$，$T_{1,3}$，$T_{1,4}$，$T_{2,1}$，$T_{2,2}$，$T_{2,3}$，$T_{2,4}$，$T_{3,1}$，$T_{3,2}$，$T_{3,3}$，$T_{3,4}$，$T_{4,1}$，$T_{4,2}$，$T_{4,3}$，$T_{4,4}$。假定每个任务的单机执行时间都为t。另外假定在调度过程中集群中只有一台空闲机器。

[0026] 按照本发明的调度策略(上文(二))，任务的执行顺序是同作业多个不同任务挨在一起先后执行，其中一种执行顺序为$T_{1,1}$，$T_{1,2}$，$T_{1,3}$，$T_{1,4}$，$T_{2,1}$，$T_{2,2}$，$T_{2,3}$，$T_{2,4}$，$T_{3,1}$，$T_{3,2}$，$T_{3,3}$，$T_{3,4}$，$T_{4,1}$，$T_{4,2}$，$T_{4,3}$，$T_{4,4}$。这样，J1，J2，J3，J4完成时间分别为4t，8t，12t，16t。作业平均完成时间为10t。

[0027] 按照传统的调度策略，每次随机取一个任务调度，平均完成时间最长情况下，其中一种执行顺序为$T_{1,1}$，$T_{2,1}$，$T_{3,1}$，$T_{4,1}$，$T_{1,2}$，$T_{2,2}$，$T_{3,2}$，$T_{4,2}$，$T_{1,3}$，$T_{2,3}$，$T_{3,3}$，$T_{4,2}$，$T_{1,4}$，$T_{2,4}$，$T_{3,4}$，$T_{4,4}$。这样，J1，J2，J3，J4完成时间分别为13t，14t，15t，16t。作业平均完成时间为14.5t。平均完成时间最短情况即符合本发明调度策略的情况，完成时间分别为4t，8t，12t，16t。平均完成时间为10t。

[0028] (四)集群系统中同优先级任务调度策略实验验证

[0029] 实验一，4个同优先级作业，每个作业被分成4个任务，共16个任务的情况下。

[0030] 假设J1，J2，J3，J4四个作业的优先级相同，这4个作业分别被分解成4个任务，分别为$T_{1,1}$，$T_{1,2}$，$T_{1,3}$，$T_{1,4}$，$T_{2,1}$，$T_{2,2}$，$T_{2,3}$，$T_{2,4}$，$T_{3,1}$，$T_{3,2}$，$T_{3,3}$，$T_{3,4}$，$T_{4,1}$，$T_{4,2}$，$T_{4,3}$，$T_{4,4}$。假定集群中只有一台空闲机器，调度周期为1s，每个任务的单机执行时间约为60s。

[0031] 按照传统调度策略，运行10次。每个作业的完成时间(秒s)如表1所示。

[0032] 表1传统策略每个作业完成时间(秒)

[0033]

J1	674	609	977	733	918	979	857	732	734	978
J2	612	854	795	855	976	673	795	978	551	673
J3	794	980	917	490	733	428	980	916	980	856
J4	979	392	672	977	611	917	612	612	917	613

[0034] 这样，作业J1，J2，J3，J4的完成时间平均值分别为819.1s，776.2s，807.4s，730.2s。这四个作业完成时间平均值为783.225s。

[0035] 按照本发明提出的调度策略，运行10次。每个作业的完成时间(秒s)如表2所示。

[0036] 表2本发明策略每个作业完成时间(秒)

[0037]

J1	731	980	490	489	977	247	734	246	489	247
J2	487	249	248	978	735	980	488	736	978	733
J3	244	732	978	731	246	492	245	979	735	491
J4	979	487	734	245	491	735	978	489	490	979

[0038] 这样，作业J1，J2，J3，J4的完成时间平均值分别为563s，661.2s，587.3s，660.7s。这四个作业完成时间平均值为618.05s。

[0039] 实验二，4个同优先级作业，每个作业被分成4个任务，共16个任务的情况下。

[0040] 假设J1，J2，J3，J4四个作业的优先级相同，这4个作业分别被分解成4个任务，分别为$T_{1,1}$，$T_{1,2}$，$T_{1,3}$，$T_{1,4}$，$T_{2,1}$，$T_{2,2}$，$T_{2,3}$，$T_{2,4}$，$T_{3,1}$，$T_{3,2}$，$T_{3,3}$，$T_{3,4}$，$T_{4,1}$，$T_{4,2}$，$T_{4,3}$，$T_{4,4}$。假定集群中只有一台空闲机器，调度周期为1s，$T_{1,1}$，$T_{1,2}$，$T_{1,3}$，$T_{1,4}$的单机执行时间约为30s，$T_{2,1}$，$T_{2,2}$，$T_{2,3}$，$T_{2,4}$的单机执行时间约为40s，$T_{3,1}$，$T_{3,2}$，$T_{3,3}$，$T_{3,4}$的单机执行时间约为50s，$T_{4,1}$，$T_{4,2}$，$T_{4,3}$，$T_{4,4}$的单机执行时间约为60s。

[0041] 按照传统调度策略，运行10次。每个作业的完成时间(秒s)如表3所示。

[0042] 表3传统策略每个作业完成时间(秒)

[0043]

J1	740	521	738	698	473	461	460	616	738	412
J2	709	675	614	739	688	738	429	358	501	688
J3	633	634	573	430	740	654	741	553	614	739
J4	552	737	707	624	596	367	690	739	707	515

[0044] 这样，作业J1，J2，J3，J4的完成时间平均值分别为585.7s，613.9s，631.1s，623.4s。这四个作业完成时间平均值为613.525s。

[0045] 按照本发明提出的调度策略，运行10次。每个作业的完成时间(秒s)如表4所示。

[0046] 表4本发明策略每个作业完成时间(秒)

[0047]

J1	738	125	739	534	123	533	495	367	289	125
J2	613	290	165	163	533	410	371	737	164	740
J3	205	495	369	739	741	739	205	433	493	330

J4	450	740	614	409	370	245	738	244	736	434

[0048] 这样，作业J1，J2，J3，J4的完成时间平均值分别为406.8s，418.6s，474.9s，498s。这四个作业完成时间平均值为449.575s。

[0049] 实验结果分析

[0050] 通过分析实验一和实验二的实验结果可以发现，与传统的随机调度相比较，本发明提出的调度策略在不影响资源利用率的前提下缩短了作业的平均完成时间和作业的平均周转时间，提高了系统吞吐量。显然，本发明提出的调度策略比传统的随机调度更适用于同优先级情况下的任务调度。

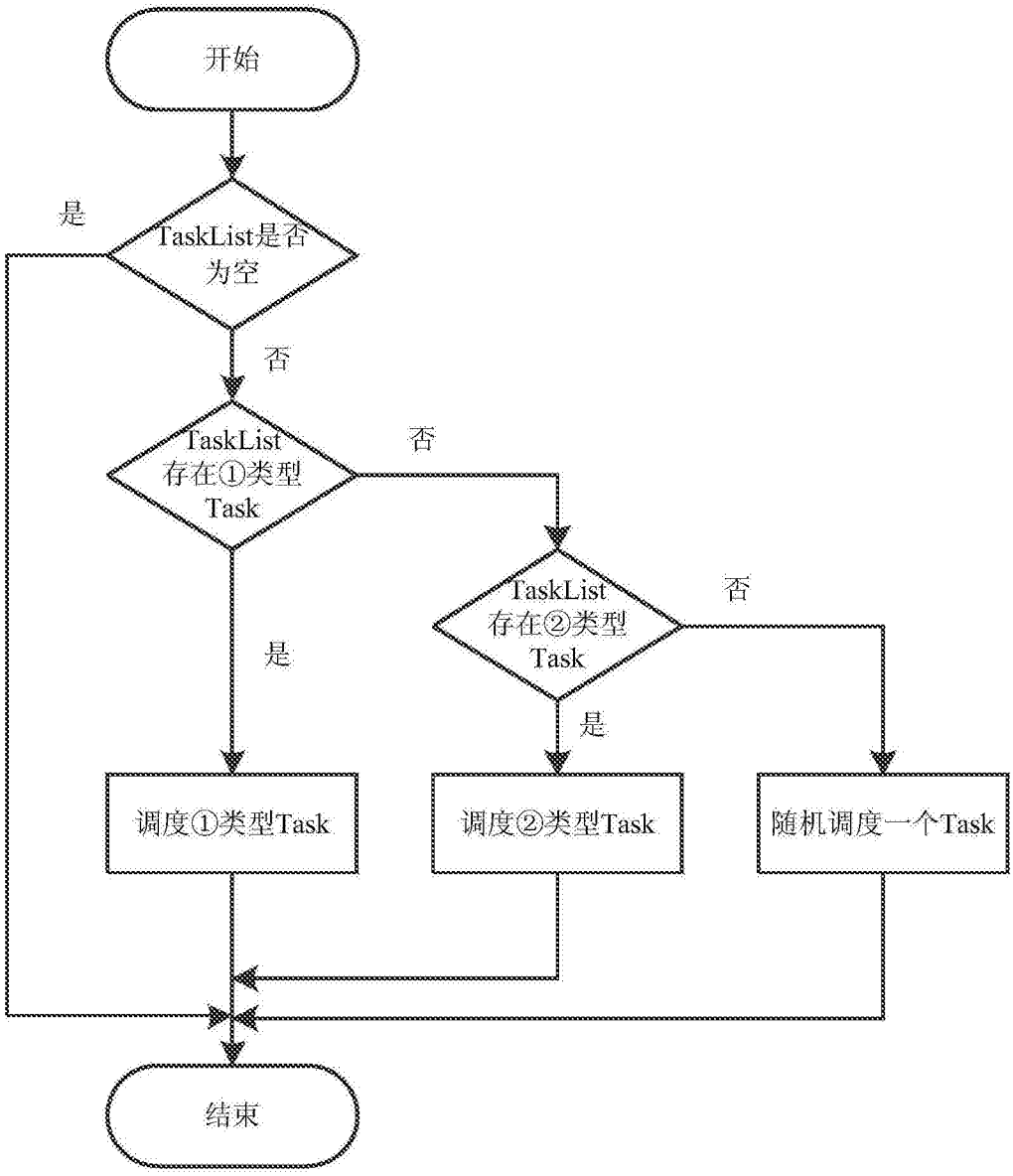

开始
是
TaskList是否为空
否
TaskList存在①类型Task
否
是
TaskList存在②类型Task
否
是
调度①类型Task
调度②类型Task
随机调度一个Task
结束

图1