(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 105468458 B
(45)授权公告日 2019.04.19

(21)申请号 201510846500.1

(22)申请日 2015.11.26

(65)同一申请的已公布的文献号
申请公布号 CN 105468458 A

(43)申请公布日 2016.04.06

(73)专利权人 北京航空航天大学
地址 100191 北京市海淀区北京航空航天大学

(72)发明人 胡春明 赵云昌 杨任宇 沃天宇

(74)专利代理机构 北京同立钧成知识产权代理有限公司 11205
代理人 马爽 黄健

(51)Int.Cl.
G06F 9/50(2006.01)

(56)对比文件
CN 101090359 A,2007.12.19,
CN 101159699 A,2008.04.09,
CN 103593242 A,2014.02.19,
CN 103685563 A,2014.03.26,
田东等.网格资源提前预留中用户资源需求量预测模型.《华中科技大学学报(自然科学版)》.2006,第34卷3.
王文峰等.YarnPlus:基于Yarn的异构任务资源共享框架.《中国计算机大会》.2013,

审查员 刘瑛

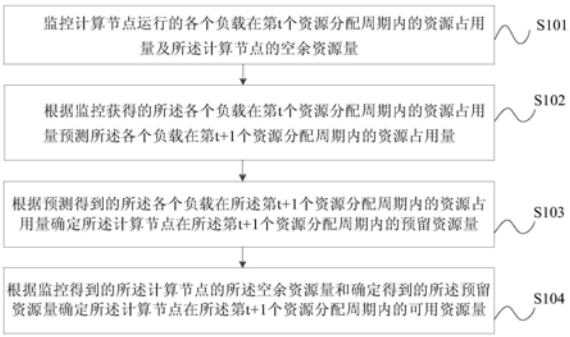
权利要求书3页 说明书7页 附图4页

(54)发明名称

计算机集群的资源调度方法及系统

(57)摘要

本发明提供一种计算机集群的资源调度方法及系统,所述方法包括:监控计算节点运行的负载在第t个资源分配周期内的资源占用量及计算节点的空余资源量;根据监控获得的负载在第t个资源分配周期内的资源占用量预测各个负载在第t+1个资源分配周期内的资源占用量;根据预测得到的负载在所述第t+1个资源分配周期内的资源占用量及监控得到的空余资源量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量;将所述计算节点在所述第t+1个资源分配周期内的可用资源量发送给资源管理器,使资源管理器根据所述可用资源量分配资源。本发明提供的计算机集群的资源调度方法及系统,能够提高所述计算节点的资源利用率及负载的服务质量。



CN 105468458 B

1.一种计算机集群的资源调度方法,其特征在于,包括:

监控计算节点运行的各个负载在第t个资源分配周期内的资源占用量及所述计算节点的空余资源量;t为大于等于1的整数;

根据监控获得的所述各个负载在第t个资源分配周期内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量;

根据预测得到的所述各个负载在所述第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的预留资源量;

根据监控得到的所述计算节点的所述空余资源量和确定得到的所述预留资源量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量;

将所述计算节点在所述第t+1个资源分配周期内的可用资源量发送给资源管理器,使所述资源管理器根据所述可用资源量分配资源;

其中,所述根据监控获得的所述各个负载在第t个资源分配周期内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量包括:

根据

$M_{t+1}=N_t+(1-\lambda)\triangle_t+\lambda\triangle'_{t-1}$

预测每个负载在所述第t+1个资源分配周期内的资源占用量;

其中,M_{t+1}表示每个负载在所述第t+1个资源分配周期内的资源占用量的预测值;N_t表示每个负载在所述第t个资源分配周期内的资源占用量的实际值;λ为遗忘系数,决定每个负载以前资源占用量数据对所述预测值的减小程度;$\triangle_t$为每个负载在所述第t个资源分配周期与所述第t-1个资源分配周期内资源实际使用量变动差值;$\triangle'_{t-1}$为所述第t个资源分配周期之前的所有资源分配周期内的资源实际使用量衰减值,$\triangle'_t=(1-\lambda)\triangle_t+\lambda\triangle'_{t-1}$。

2.根据权利要求1所述的方法,其特征在于,所述根据预测得到的所述各个负载在所述第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的预留资源量包括:

根据

$S_{t+1}=\sum B_j*\alpha$

确定所述计算节点在所述第t+1个资源分配周期内的预留资源量;

其中,$B=M_{t+1}$,B_j表示标号为j的负载在所述第t+1个资源分配周期内的资源占用量的预测值,j表示负载的标号,j为大于等于1的整数,S_{t+1}表示所述计算节点在所述第t+1个资源分配周期内的预留资源量,α表示需要为长时间运行的负载预留的资源的百分比。

3.根据权利要求2所述的方法,其特征在于,所述根据监控得到的所述计算节点的所述空余资源量和确定得到的所述预留资源量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量包括:

根据$A_{t+1}=T-S_{t+1}$确定所述计算节点在所述第t+1个资源分配周期内的可用资源量;其中,A_{t+1}表示所述计算节点在所述第t+1个资源分配周期内的可用资源量,T表示监控获得的所述计算节点的空余资源量。

4.一种计算机集群的资源调度方法,其特征在于,包括:

接收应用管理器发送的资源申请信息,所述资源申请信息包括在第t+1个资源分配周期内运行负载所需的资源量数据,t为大于等于1的整数;

接收计算节点发送的所述计算节点在所述第t+1个资源分配周期内的可用资源量;其中,所述计算节点在所述第t+1个资源分配周期内的可用资源量根据以下方式获得:监控所述计算节点中运行的各个负载在资源分配周期t内的资源占用量及空余资源量,根据所述各个负载在资源分配周期t内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量,根据所述计算节点在资源分配周期t内的空余资源量及预测得到的所述各个负载在第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量,其中,根据所述各个负载在资源分配周期t内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量通过以下方式获得:根据$M_{t+1}=N_t+(1-\lambda)\triangle_t+\lambda\triangle'_{t-1}$预测每个负载在所述第t+1个资源分配周期内的资源占用量;其中,M_{t+1}表示每个负载在所述第t+1个资源分配周期内的资源占用量的预测值;N_t表示每个负载在所述第t个资源分配周期内的资源占用量的实际值;λ为遗忘系数,决定每个负载以前资源占用量数据对所述预测值的减小程度;$\triangle_t$为每个负载在所述第t个资源分配周期与所述第t-1个资源分配周期内资源实际使用量变动差值;$\triangle'_{t-1}$为所述第t个资源分配周期之前的所有资源分配周期内的资源实际使用量衰减值,$\triangle'_t=(1-\lambda)\triangle_t+\lambda\triangle'_{t-1}$;

根据接收到的所述资源申请信息及所述计算节点在第t+1个资源分配周期内的可用资源量,向所述应用管理器分配在所述第t+1个资源分配周期内运行所述负载的资源。

5.根据权利要求4所述的方法,其特征在于,所述方法还包括:

接收所述计算节点发送的所述计算节点中运行的各个负载在资源分配周期t内的资源占用量;

记录每个资源分配周期内所述各个负载的资源占用量的变化量。

6.一种计算机集群的资源调度系统,其特征在于,包括:

计算节点、资源管理器及应用管理器,所述计算节点与所述资源管理器通信连接,所述应用管理器分别与所述计算节点及所述资源管理器通信连接;

所述计算节点用于,监控计算节点运行的各个负载在第t个资源分配周期内的资源占用量及所述计算节点的空余资源量,t为大于等于1的整数;根据监控获得的所述各个负载在第t个资源分配周期内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量,其中,所述根据监控获得的各个负载在资源分配周期t内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量,包括:

根据

$M_{t+1}=N_t+(1-\lambda)\triangle_t+\lambda\triangle'_{t-1}$

预测每个负载在所述第t+1个资源分配周期内的资源占用量;

其中,M_{t+1}表示每个负载在所述第t+1个资源分配周期内的资源占用量的预测值;N_t表示每个负载在所述第t个资源分配周期内的资源占用量的实际值;λ为遗忘系数,决定每个负载以前资源占用量数据对所述预测值的减小程度;$\triangle_t$为每个负载在所述第t个资源分配周期与所述第t-1个资源分配周期内资源实际使用量变动差值;$\triangle'_{t-1}$为所述第t个资源分配周期之前的所有资源分配周期内的资源实际使用量衰减值,$\triangle'_t=(1-\lambda)\triangle_t+\lambda\triangle'_{t-1}$;根据预测得到的所述各个负载在所述第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的预留资源量;及根据监控得到的所述计算节点的所述空余资源量和确定得到的所述预留资源量确定所述计算节点在所述第t+1个资源分配周期内

的可用资源量;并用于将所述计算节点在所述第t+1个资源分配周期内的可用资源量发送给所述资源管理器;

所述应用管理器用于,向所述资源管理器发送资源申请信息,所述资源申请信息包括在第t+1个资源分配周期内运行负载所需的资源量数据;

所述资源管理器用于,接收所述应用管理器发送的资源申请信息及所述计算节点发送的所述计算节点在所述第t+1个资源分配周期内的可用资源量,并根据所述资源申请信息及所述计算节点在第t+1个资源分配周期内的可用资源量,给所述应用管理器分配在所述第t+1个资源分配周期内运行所述负载的资源。

7.根据权利要求6所述的系统,其特征在于,所述应用管理器还用于,调用所述资源管理器分配的在所述第t+1个资源分配周期内运行所述负载的资源运行所述负载。

8.根据权利要求6或7所述的系统,其特征在于,所述计算节点与所述资源管理器通过心跳协议通信连接。

计算机集群的资源调度方法及系统

技术领域

[0001] 本发明涉及计算机技术领域，尤其涉及一种计算机集群的资源调度方法及系统。

背景技术

[0002] 计算机集群是一种计算机系统，通过一组松散集成的计算机软件和/或硬件连接起来高度紧密地协作完成计算工作。计算机集群系统中的单个计算机通常称为计算节点，通常通过局域网连接。计算机集群系统通过计算机集群资源管理器对计算机集群的资源进行监控与分配。

[0003] 现有技术中，常用的计算机集群管理系统资源管理器都是采用静态调度，即在每个负载运行前，由应用管理器向资源管理器进行资源申请，为要启动的负载申请CPU、内存等资源。资源管理器根据调度算法，选择能够满足负载所需资源的计算节点并将资源分配给所述负载。每个负载在其运行生命周期内占用的资源不变。

[0004] 但是，采用现有技术中的资源管理器对计算机集群的资源进行分配的方法，当负载占用的资源发生变化时，不能根据资源的变化情况动态地给负载分配资源，当负载占用的资源降低时，造成了资源浪费；当负载占用的资源上升时，无法保证服务质量。

发明内容

[0005] 本发明实施例提供一种计算机集群的资源调度方法及系统，用于解决现有技术中计算机集群资源调度方法不能根据资源的变化情况分配资源的问题。

[0006] 第一方面，本发明实施例提供一种计算机集群的资源调度方法，包括：

[0007] 监控计算节点运行的各个负载在第t个资源分配周期内的资源占用量及所述计算节点的空余资源量；t为大于等于1的整数；

[0008] 根据监控获得的所述各个负载在第t个资源分配周期内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量；

[0009] 根据预测得到的所述各个负载在所述第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的预留资源量；

[0010] 根据监控得到的所述计算节点的所述空余资源量和确定得到的所述预留资源量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量；

[0011] 将所述计算节点在所述第t+1个资源分配周期内的可用资源量发送给资源管理器，使所述资源管理器根据所述可用资源量分配资源。

[0012] 另一实施例中，所述根据监控获得的所述各个负载在第t个资源分配周期内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量包括：

[0013] 根据

[0014] $M_{t+1}=N_t+(1-\lambda)\triangle_t+\lambda\triangle'_{t-1}$

[0015] 预测每个负载在所述第t+1个资源分配周期内的资源占用量；

[0016] 其中，M_{t+1}表示每个负载在所述第t+1个资源分配周期内的资源占用量的预测值；

N_t表示每个负载在所述第t个资源分配周期内的资源占用量的实际值;λ为遗忘系数,决定每个负载以前资源占用量数据对所述预测值的减小程度;$\triangle_t$为每个负载在所述第t个资源分配周期与所述第t-1个资源分配周期内资源实际使用量变动差值;$\triangle'_{t-1}$为所述第t个资源分配周期之前的所有资源分配周期内的资源实际使用量衰减值,$\triangle'_t=(1-\lambda)\triangle_t+\lambda\triangle'_{t-1}$。

[0017] 另一实施例中,所述根据预测得到的所述各个负载在所述第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的预留资源量包括:

[0018] 根据

[0019] $S_{t+1}=\sum B_j*\alpha$

[0020] 确定所述计算节点在所述第t+1个资源分配周期内的预留资源量;

[0021] 其中,$B=M_{t+1}$,B_j表示标号为j的负载在所述第t+1个资源分配周期内的资源占用量的预测值,j表示负载的标号,j为大于等于1的整数,S_{t+1}表示所述计算节点在所述第t+1个资源分配周期内的预留资源量,α表示需要为长时间运行的负载预留的资源的百分比。

[0022] 另一实施例中,所述根据监控得到的所述计算节点的所述空余资源量和确定得到的所述预留资源量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量包括:

[0023] 根据$A_{t+1}=T-S_{t+1}$确定所述计算节点在所述第t+1个资源分配周期内的可用资源量;其中,A_{t+1}表示所述计算节点在所述第t+1个资源分配周期内的可用资源量,T表示监控获得的所述计算节点的空余资源量。

[0024] 第二方面,本发明实施例还提供一种计算机集群的资源调度方法,包括:

[0025] 接收应用管理器发送的资源申请信息,所述资源申请信息包括在第t+1个资源分配周期内运行负载所需的资源量数据,t为大于等于1的整数;

[0026] 接收计算节点发送的所述计算节点在所述第t+1个资源分配周期内的可用资源量;其中,所述计算节点在所述第t+1个资源分配周期内的可用资源量根据以下方式获得:监控所述计算节点中运行的各个负载在资源分配周期t内的资源占用量及空余资源量,根据所述各个负载在资源分配周期t内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量,根据所述计算节点在资源分配周期t内的空余资源量及预测得到的所述各个负载在第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量;

[0027] 根据接收到的所述资源申请信息及所述计算节点在第t+1个资源分配周期内的可用资源量,向所述应用管理器分配在所述第t+1个资源分配周期内运行所述负载的资源。

[0028] 另一实施例中,所述方法还包括:

[0029] 接收所述计算节点发送的所述计算节点中运行的各个负载在资源分配周期t内的资源占用量;

[0030] 记录每个资源分配周期内所述各个负载的资源占用量的变化量。

[0031] 第三方面,本发明实施例提供一种计算机集群的资源调度系统,包括:

[0032] 计算节点、资源管理器及应用管理器,所述计算节点与所述资源管理器通信连接,所述应用管理器分别与所述计算节点及所述资源管理器通信连接;

[0033] 所述计算节点用于,监控计算节点运行的各个负载在第t个资源分配周期内的资源占用量及所述计算节点的空余资源量,t为大于等于1的整数;根据监控获得的所述各个

负载在第t个资源分配周期内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量;根据预测得到的所述各个负载在所述第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的预留资源量;及根据监控得到的所述计算节点的所述空余资源量和确定得到的所述预留资源量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量;并用于将所述计算节点在所述第t+1个资源分配周期内的可用资源量发送给所述资源管理器;

[0034] 所述应用管理器用于,向所述资源管理器发送资源申请信息,所述资源申请信息包括在第t+1个资源分配周期内运行负载所需的资源量数据;

[0035] 所述资源管理器用于,接收所述应用管理器发送的资源申请信息及所述计算节点发送的所述计算节点在所述第t+1个资源分配周期内的可用资源量,并根据所述资源申请信息及所述计算节点在第t+1个资源分配周期内的可用资源量,给所述应用管理器分配在所述第t+1个资源分配周期内运行所述负载的资源。

[0036] 另一实施例中,所述应用管理器还用于,调用所述资源管理器分配的在所述第t+1个资源分配周期内运行所述负载的资源运行所述负载。

[0037] 另一实施例中,所述计算节点与所述资源管理器通过心跳协议通信连接。

[0038] 本发明提供的计算机集群的资源调度方法及系统,通过动态监控计算节点运行的各个负载在当前资源分配周期的资源占用量预测下一个资源分配周期内的资源占用量,根据预测得到的下一个资源分配周期的资源占用量及监控得到的所述计算节点的空余资源量确定所述计算节点在所述下一个资源分配周期内的可用资源量,并将所述计算节点在所述下一个资源分配周期内的可用资源量发送给资源管理器,使所述资源管理器可以根据所述可用资源量分配资源,从而提高所述计算节点的资源利用率及所述计算节点内运行的负载的服务质量。

附图说明

[0039] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图做一简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0040] 图1为实现本发明实施例计算机集群的资源调度方法的系统架构示意图;

[0041] 图2为本发明实施例计算机集群的资源调度方法的流程示意图;

[0042] 图3为本发明另一实施例计算机集群的资源调度方法的流程示意图;

[0043] 图4为本发明实施例计算机集群的资源调度系统的结构示意图;

[0044] 图5为现有技术中计算节点运行负载时的CPU资源申请量和实际CPU利用率的对比图;

[0045] 图6为现有技术中计算节点运行负载时的内存资源申请量和实际内存利用率的对比图。

具体实施方式

[0046] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例

中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0047] 本发明提供一种计算机集群的资源调度方法，用于对计算机集群的资源进行动态调配，以提高计算机集群系统的资源利用率及服务质量。图1为实现本发明实施例计算机集群的资源调度方法的系统架构示意图。请参阅图1，所述系统包括计算节点10，资源管理器20及应用管理器30。所述计算节点10包括监控模块11、可用资源确定模块12、远程通讯模块13及容器管理通讯模块14。所述资源管理器20包括资源分配模块21、远程通讯模块22及应用管理通讯模块23。所述计算节点10的通讯模块13及所述资源管理器22的通讯模块13通过心跳协议实现通讯连接。所述资源管理器20的所述应用管理通讯模块23通过应用管理协议与所述应用管理器30通讯，向所述应用管理器30分配资源，所述应用管理器30与所述计算节点10的容器管理通讯模块14通过容器管理协议通讯，调用所述计算节点10内的资源。

[0048] 图2为本发明实施例计算机集群的资源调度方法的流程示意图。请参阅图2，所述方法有计算节点执行，包括：

[0049] S101：监控计算节点运行的各个负载在第t个资源分配周期内的资源占用量及所述计算节点的空余资源量；t为大于等于1的整数；

[0050] 具体地，所述计算节点内设置有监控模块，用于监控所述计算节点运行的各个负载在第t个资源分配周期内的资源占用量及所述计算节点的空余资源量。

[0051] S102：根据监控获得的所述各个负载在第t个资源分配周期内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量；

[0052] 所述根据监控获得的所述各个负载在第t个资源分配周期内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量包括：

[0053] 根据

[0054] $M_{t+1}=N_t+(1-\lambda)\triangle_t+\lambda\triangle'_{t-1}$

[0055] 预测每个负载在所述第t+1个资源分配周期内的资源占用量；

[0056] 其中，M_{t+1}表示每个负载在所述第t+1个资源分配周期内的资源占用量的预测值；N_t表示每个负载在所述第t个资源分配周期内的资源占用量的实际值；λ为遗忘系数，决定每个负载以前资源占用量数据对所述预测值的减小程度；$\triangle_t$为每个负载在所述第t个资源分配周期与所述第t-1个资源分配周期内资源实际使用量变动差值；$\triangle'_{t-1}$为所述第t个资源分配周期之前的所有资源分配周期内的资源实际使用量衰减值，$\triangle'_t=(1-\lambda)\triangle_t+\lambda\triangle'_{t-1}$。具体地，由于每个集群中运行的负载不同，每个负载的资源变动曲线也不同，λ的值是根据负载资源变动的统计结果人工指定的。针对每个负载，可以先以单机操作的方式运行一段时间，选择不同的λ值，看看预测值和实际值的差异，选择使得预测和实际差异最小的那个λ值，使得预测尽可能的贴近实际变化。

[0057] S103：根据预测得到的所述各个负载在所述第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的预留资源量；

[0058] 所述根据预测得到的所述各个负载在所述第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的预留资源量包括：

[0059] 根据

[0060] $S_{t+1}=\sum B_j*\alpha$

[0061] 确定所述计算节点在所述第t+1个资源分配周期内的预留资源量；

[0062] 其中，$B=M_{t+1}$，B_j表示标号为j的负载在所述第t+1个资源分配周期内的资源占用量的预测值，j表示负载的标号，j为大于等于1的整数，S_{t+1}表示所述计算节点在所述第t+1个资源分配周期内的预留资源量，α表示需要为长时间运行的负载预留的资源的百分比。假设α定为0.1，比如我们预测出下一个周期，所有的长时间运行的负载需要内存资源为2GB，则需要为他预留2GB*0.1＝200MB内存。同样α是个人工指定值，过大造成资源浪费，过小会影响负载运行的服务质量，这个需要根据对每个集群进行模拟得到。

[0063] S104：根据监控得到的所述计算节点的所述空余资源量和确定得到的所述预留资源量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量；

[0064] 所述根据监控得到的所述计算节点的所述空余资源量和确定得到的所述预留资源量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量包括：

[0065] 根据$A_{t+1}=T-S_{t+1}$确定所述计算节点在所述第t+1个资源分配周期内的可用资源量；其中，A_{t+1}表示所述计算节点在所述第t+1个资源分配周期内的可用资源量，T表示监控获得的所述计算节点的空余资源量。

[0066] S105：将所述计算节点在所述第t+1个资源分配周期内的可用资源量发送给资源管理器，使所述资源管理器根据所述可用资源量分配资源。

[0067] 本发明实施例提供的计算机集群的资源调度方法，通过动态监控计算节点运行的各个负载在当前资源分配周期的资源占用量预测下一个资源分配周期内的资源占用量，根据预测得到的下一个资源分配周期的资源占用量及监控得到的所述计算节点的空余资源量确定所述计算节点在所述下一个资源分配周期内的可用资源量，并将所述计算节点在所述下一个资源分配周期内的可用资源量发送给资源管理器，使所述资源管理器可以根据所述可用资源量分配资源，从而提高所述计算节点的资源利用率及所述计算节点内运行的负载的服务质量。

[0068] 图3为本发明另一实施例计算机集群的资源调度方法的流程示意图。请参阅图2，所述方法由资源管理器执行，包括：

[0069] S201：接收应用管理器发送的资源申请信息，所述资源申请信息包括在第t+1个资源分配周期内运行负载所需的资源量数据，t为大于等于1的整数；

[0070] 具体地，当用户需要运行一项负载(例如：运行应用程序)时，通过所述应用管理器提交负载运行申请。所述应用管理器计算运行所述负载所需要的资源量数据，并将所述资源量数据以资源申请信息的方式提交给所述资源管理器。

[0071] S202：接收计算节点发送的所述计算节点在所述第t+1个资源分配周期内的可用资源量；其中，所述计算节点在所述第t+1个资源分配周期内的可用资源量根据以下方式获得：监控所述计算节点中运行的各个负载在资源分配周期t内的资源占用量及空余资源量，根据所述各个负载在资源分配周期t内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量，根据所述计算节点在资源分配周期t内的空余资源量及预测得到的所述各个负载在第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量；

[0072] S203：根据接收到的所述资源申请信息及所述计算节点在第t+1个资源分配周期

内的可用资源量，向所述应用管理器分配在所述第t+1个资源分配周期内运行所述负载的资源。

[0073] 具体地，所述资源管理器接收到所述计算节点发送的所述计算节点在所述第t+1个资源分配周期内的可用资源量之后，根据贪婪算法选择能够运行所述负载的计算节点，并将所述计算节点的资源分配给所述应用管理器。

[0074] 本发明实施例提供的计算机集群的资源调度方法，通过动态地接收计算节点发送的所述计算节点在所述第t+1个资源分配周期内的可用资源量，并根据接收到的所述应用管理器提交的第t+1个资源分配周期内运行负载所需的资源量数据信息，可以动态地向所述应用管理器分配在所述第t+1个资源分配周期内运行所述负载的资源，从而提高所述计算节点的资源利用率及所述计算节点内运行的负载的服务质量。

[0075] 进一步地，为了便于根据所述计算节点在各个资源分配周期内的资源可用量的变化情况合理地分配资源，所述方法还包括：

[0076] 接收所述计算节点发送的所述计算节点中运行的各个负载在资源分配周期t内的资源占用量；

[0077] 记录每个资源分配周期内所述各个负载的资源占用量的变化量。

[0078] 本发明实施例还提供一种计算机集群的资源调度系统。图4为本发明实施例计算机集群的资源调度系统的结构示意图。请参阅图4，所述系统包括：

[0079] 计算节点10、资源管理器20及应用管理器30，所述计算节点10与所述资源管理器20通信连接，所述应用管理器20分别与所述计算节点10及所述资源管理器30通信连接；具体地，所述计算节点10与所述资源管理器20通过心跳协议通信连接。

[0080] 所述计算节点10用于，监控计算节点运行的各个负载在第t个资源分配周期内的资源占用量及所述计算节点的空余资源量，t为大于等于1的整数；根据监控获得的所述各个负载在第t个资源分配周期内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量；根据预测得到的所述各个负载在所述第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的预留资源量；及根据监控得到的所述计算节点的所述空余资源量和确定得到的所述预留资源量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量；并用于将所述计算节点在所述第t+1个资源分配周期内的可用资源量发送给所述资源管理器20；

[0081] 所述应用管理器30用于，向所述资源管理器20发送资源申请信息，所述资源申请信息包括在第t+1个资源分配周期内运行负载所需的资源量数据；

[0082] 所述资源管理器20用于，接收所述应用管理器30发送的资源申请信息及所述计算节点10发送的所述计算节点10在所述第t+1个资源分配周期内的可用资源量，并根据所述资源申请信息及所述计算节点10在第t+1个资源分配周期内的可用资源量，给所述应用管理器30分配在所述第t+1个资源分配周期内运行所述负载的资源。

[0083] 本发明实施例提供的计算机集群的资源调度系统，通过动态监控计算节点运行的各个负载在当前资源分配周期的资源占用量预测下一个资源分配周期内的资源占用量，根据预测得到的下一个资源分配周期的资源占用量及监控得到的所述计算节点的空余资源量确定所述计算节点在所述下一个资源分配周期内的可用资源量，并将所述计算节点在所述下一个资源分配周期内的可用资源量发送给资源管理器，使所述资源管理器可以根据所

述可用资源量分配资源，从而提高所述计算节点的资源利用率及所述计算节点内运行的负载的服务质量。

[0084] 进一步地，为了便于对负载的运行情况进行管理，所述应用管理器30还用于，调用所述资源管理器20分配的在所述第t+1个资源分配周期内运行所述负载的资源运行所述负载。

[0085] 本发明实施例提供的计算机集群的资源调度系统，用以执行上述方法实施例中的计算机集群的资源调度方法，其实现原理及技术效果类似，在此不再赘述。

[0086] 图5为现有技术中计算节点运行负载时的中央处理器(Central ProcessingUnit，CPU)资源申请量和实际CPU利用率的对比图。请参阅图4，其中直线1表示计算节点的CPU资源申请量，曲线2表示计算节点的实际CPU利用率。现有技术中，一个长时间作业的计算节点的CPU的实际利用率只有申请量的35％，并且在大部分相邻的多个监控周期中CPU的占用率变化不大，完全可以部署部分短时间任务。在现有的资源调度方法中，因为计算节点的CPU资源已经被申请占用，导致该部分资源浪费。而采用本发明实施例计算机集群的资源调度方法，使得未实际使用的部分资源得到充分利用，其效率可以比原来提高1.6倍。

[0087] 图6为现有技术中计算节点运行负载时的内存资源申请量和实际内存利用率的对比图。请参阅图6，其中直线3表示计算节点的内存资源申请量，曲线4表示计算节点的实际内存利用率。现有技术中，一个长时间作业的计算节点的实际内存使用量为申请量的47％，通过本发明实施例计算机集群的资源调度方法对资源进行实时动态调度，实际内存利用率可以提高1.1倍。

[0088] 综上所述，长时间作业的计算节点的CPU和内存资源抖动非常大，在现有技术的调度模式下，如果按照负载运行所需资源的峰值进行申请，虽然可以保证服务质量，但是将严重降低节点的实际资源利用率。采用本发明实施例提供的计算机集群的资源调度方法可以有效提高系统资源利用率。

[0089] 本领域普通技术人员可以理解：实现上述各方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成。前述的程序可以存储于一计算机可读取存储介质中。该程序在执行时，执行包括上述各方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0090] 最后应说明的是：以上各实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述各实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分或者全部技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的范围。

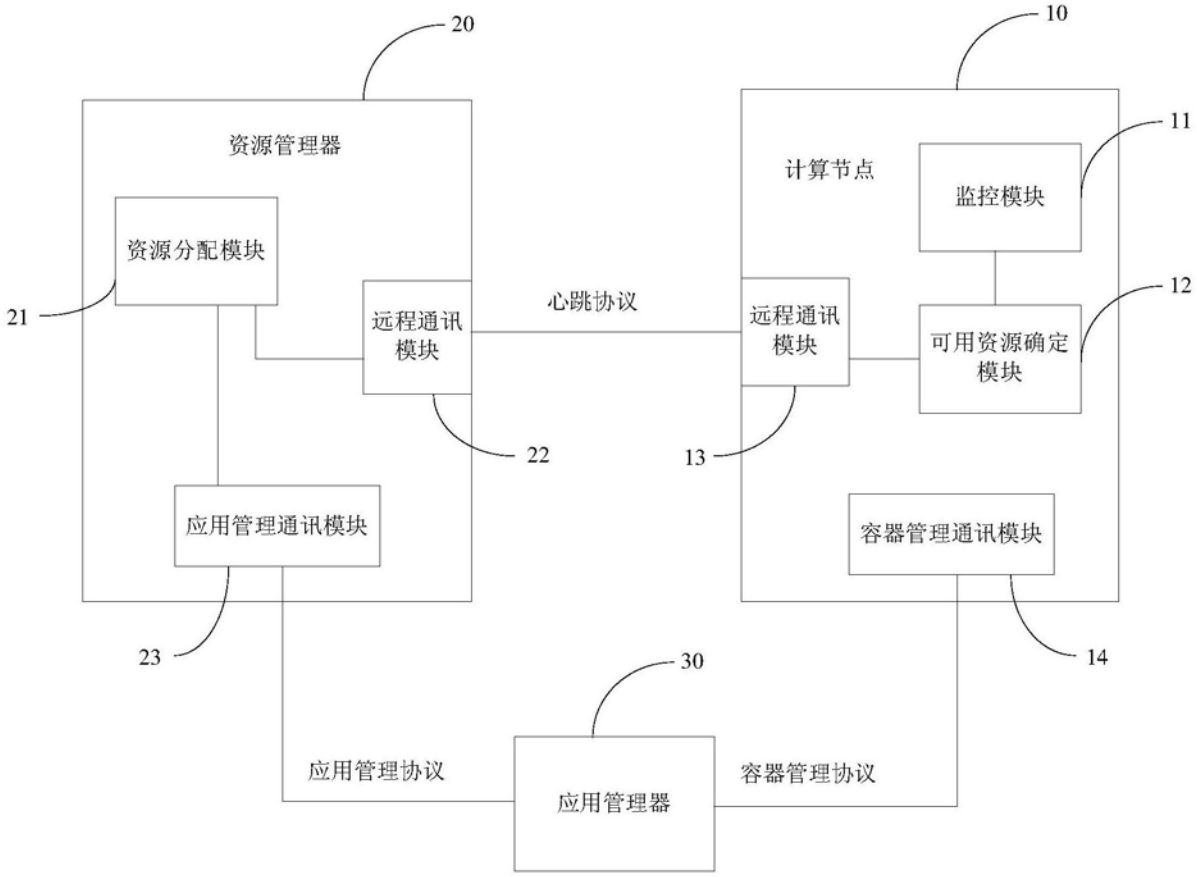

20
资源管理器
资源分配模块
21
远程通讯模块
心跳协议
22
应用管理通讯模块
23
10
计算节点
监控模块
11
远程通讯模块
可用资源确定模块
12
13
容器管理通讯模块
14
30
应用管理协议
应用管理器
容器管理协议

图1

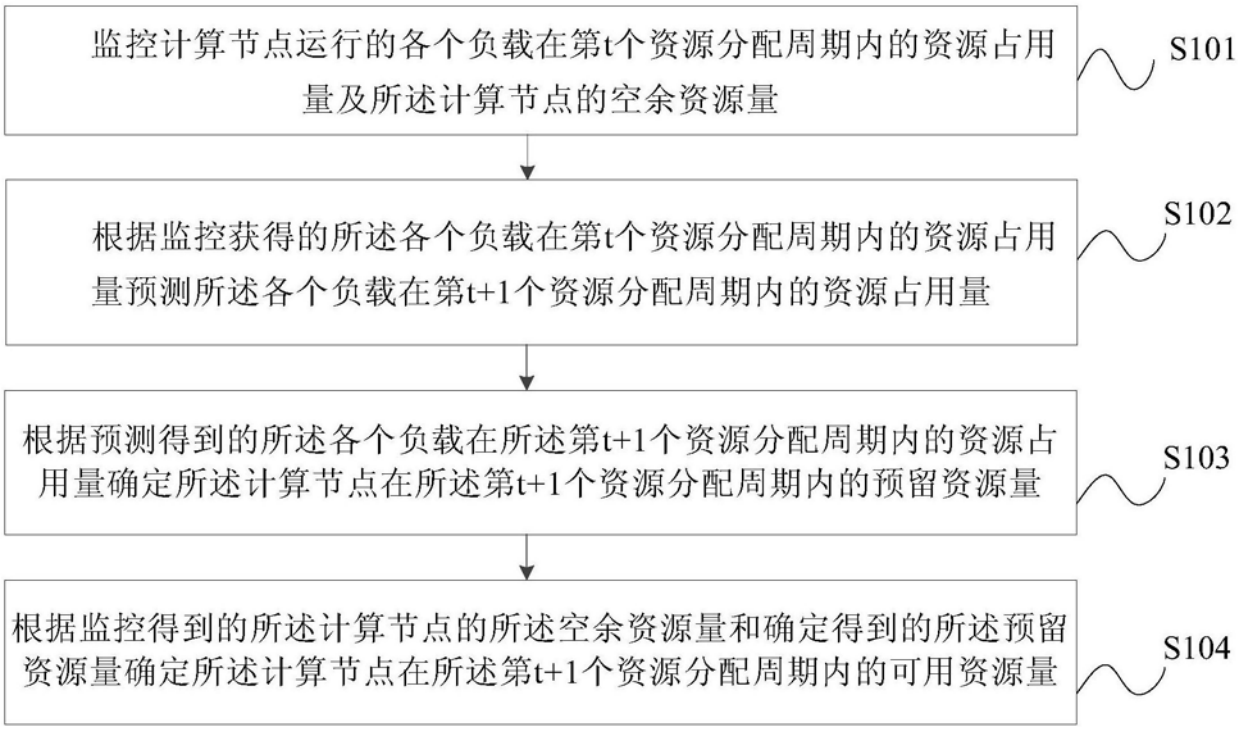

监控计算节点运行的各个负载在第t个资源分配周期内的资源占用量及所述计算节点的空余资源量
S101
根据监控获得的所述各个负载在第t个资源分配周期内的资源占用量预测所述各个负载在第t+1个资源分配周期内的资源占用量
S102
根据预测得到的所述各个负载在所述第t+1个资源分配周期内的资源占用量确定所述计算节点在所述第t+1个资源分配周期内的预留资源量
S103
根据监控得到的所述计算节点的所述空余资源量和确定得到的所述预留资源量确定所述计算节点在所述第t+1个资源分配周期内的可用资源量
S104

图2

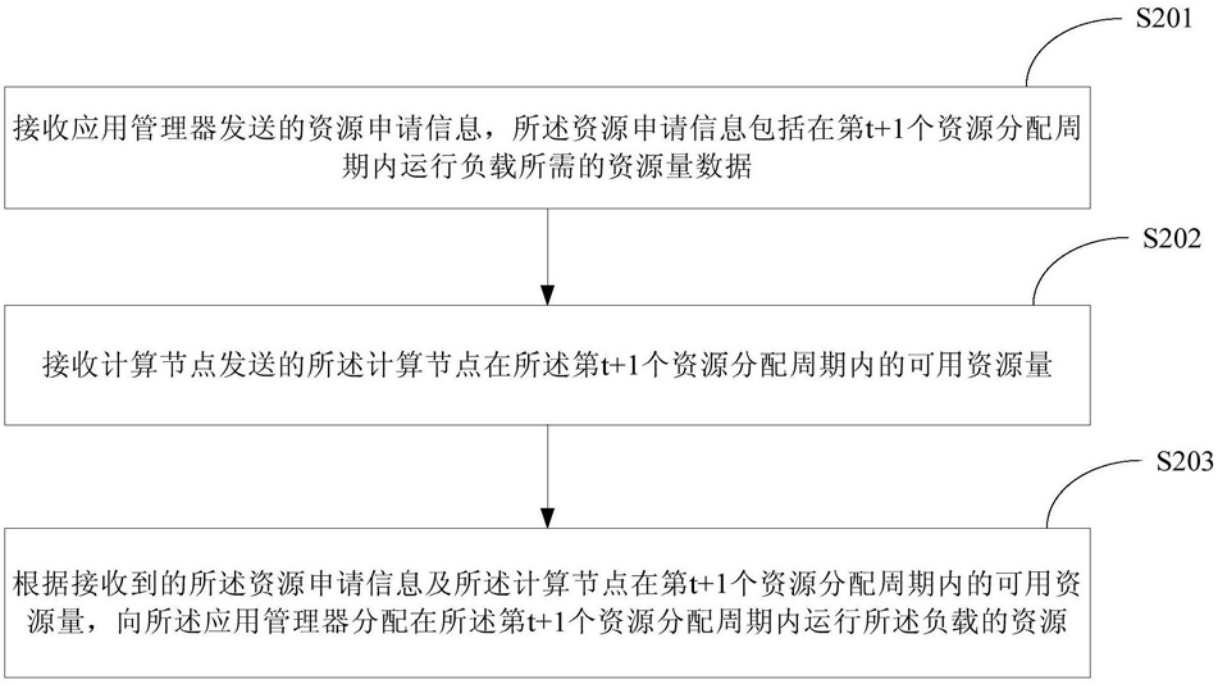

S201
接收应用管理器发送的资源申请信息，所述资源申请信息包括在第t+1个资源分配周期内运行负载所需的资源量数据
S202
接收计算节点发送的所述计算节点在所述第t+1个资源分配周期内的可用资源量
S203
根据接收到的所述资源申请信息及所述计算节点在第t+1个资源分配周期内的可用资源量，向所述应用管理器分配在所述第t+1个资源分配周期内运行所述负载的资源

图3

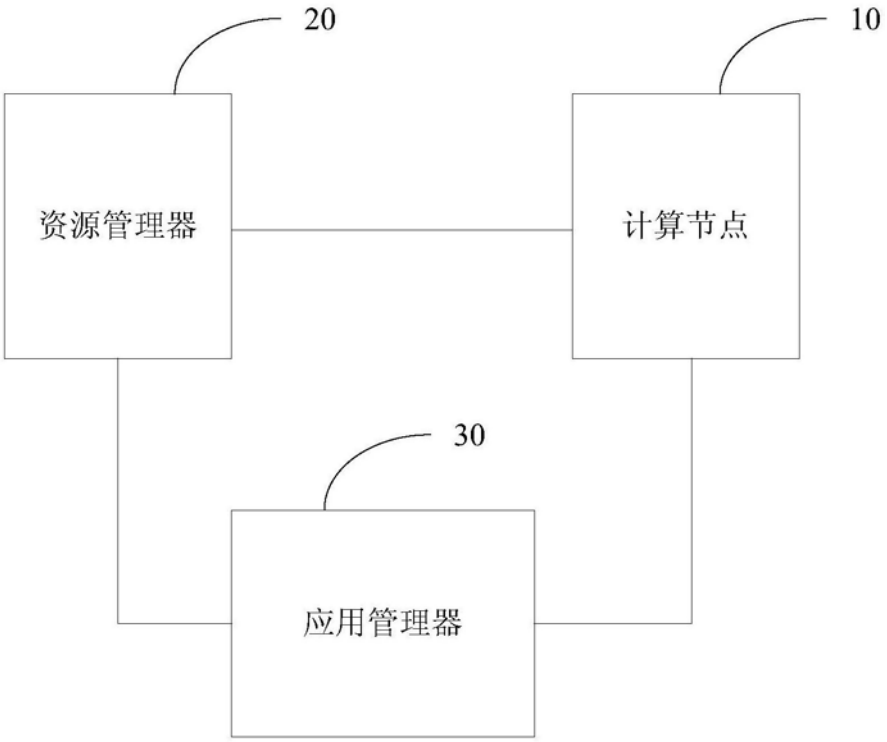

20
10
资源管理器
计算节点
30
应用管理器

图4

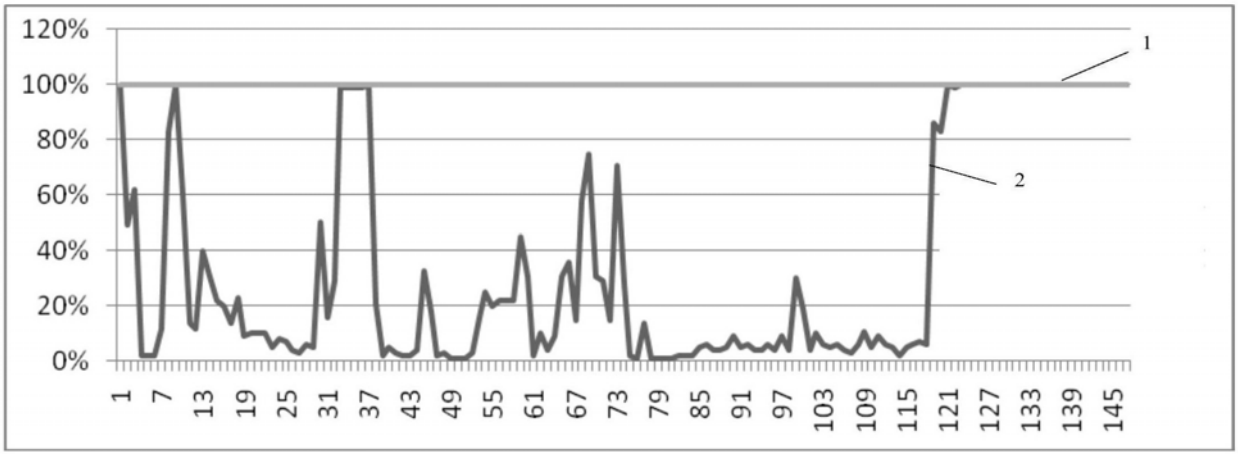

120%
100%
80%
60%
40%
20%
0%
1
7
13
19
25
31
37
43
49
55
61
67
73
79
85
91
97
103
109
115
121
127
133
139
145
1
2

图5

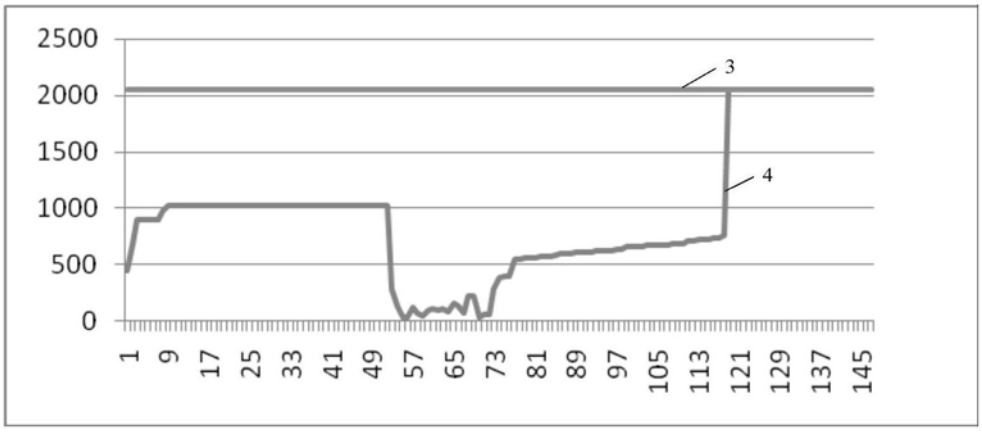

2500
2000
1500
1000
500
0
1
9
17
25
33
41
49
57
65
73
81
89
97
105
113
121
129
137
145
3
4

图6