

(12) 发明专利申请

(10) 申请公布号 CN 102915257 A

(43) 申请公布日 2013. 02. 06

(21) 申请号 201210367653. 4

(22) 申请日 2012. 09. 28

(71) 申请人 曙光信息产业(北京)有限公司  
地址 100193 北京市海淀区东北旺西路 8 号  
中关村软件园 36 号楼

(72) 发明人 林霞

(74) 专利代理机构 北京安博达知识产权代理有限公司 11271  
代理人 徐国文

(51) Int. Cl.  
G06F 11/00(2006. 01)

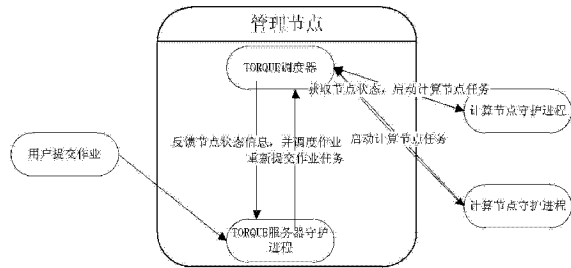
权利要求书 1 页 说明书 4 页 附图 1 页

(54) 发明名称

基于 TORQUE 的并行检查点执行方法

(57) 摘要

本发明提供了一种基于 TORQUE 的并行检查点执行方法,在 NFS 文件共享存储系统中对计算节点进行检查点操作,所述方法包括如下步骤:(1). 用户向 TORQUE 服务器守护进程 pbs\_server 提交作业,提交命令加入检查点请求,作业脚本中使用作业启动命令 chkp\_mpirun 启动任务;(2). TORQUE 服务器守护进程发送任务消息给 TORQUE 调度器 pbs\_sched, TORQUE 调度器 pbs\_sched 根据作业中指定的参数要求寻找计算节点;(3). 对计算节点进行检查点操作。本发明提供的基于 TORQUE 的并行检查点执行方法,运用改进的基于 TORQUE 的改进检查点技术,使得原来只支持单进程检查点的 TORQUE,现在也能够支持多进程检查点,并能够对节点故障自动处理,对进程进行迁移。



1. 一种基于 TORQUE 的并行检查点执行方法,在 NFS 文件共享存储系统中对计算节点进行检查点操作,其特征在于,所述方法包括如下步骤:

(1). 用户向 TORQUE 服务器守护进程 pbs\_server 提交作业;

(2). TORQUE 服务器守护进程发送任务消息给 TORQUE 调度器 pbs\_sched, TORQUE 调度器 pbs\_sched 根据作业中指定的参数要求寻找计算节点;

(3). 对计算节点进行检查点操作。

2. 根据权利要求 1 所述的方法,其特征在于,在所述步骤 1 中,用户通过 TORQUE 提交作业命令 qsub 提交作业脚本,作业脚本使用 MPI 进程启动命令 chkp\_mpirun 启动 MPI,在提交作业命令中加入作业检查点请求。

3. 根据权利要求 1 所述的方法,其特征在于,在所述步骤 2 中, TORQUE 调度器 pbs\_sched 通过计算节点守护进程 pbs\_mom 轮询各计算节点的状态,将结果返回给 TORQUE 服务器守护进程 pbs\_server。

4. 根据权利要求 1 所述的方法,其特征在于,在所述步骤 3 中, TORQUE 服务器守护进程 pbs\_server 将用户的检查点操作请求传送到相应的计算节点上并周期运行;其中,各个计算节点分别设置有独立检查点镜像。

5. 根据权利要求 4 所述的方法,其特征在于,周期进行检查点操作的各个计算节点将独立检查点镜像合并成全局检查点文件。

6. 根据权利要求 5 所述的方法,其特征在于,全局检查点文件存储在 NFS 文件共享存储系统中。

7. 根据权利要求 1 或者 4 所述的方法,其特征在于,在所述步骤 3 后,存在以下步骤 a:

(a). 在轮询各个计算节点时,如果发现某个计算节点故障,则将该计算节点上分配的所有未执行完的和未执行的任务根据全局检查点文件全部进行卷回恢复。

8. 根据权利要求 7 所述的方法,其特征在于,在所述步骤 a 中, TORQUE 调度器 pbs\_sched 将故障计算节点的所有任务号发回给 TORQUE 服务器守护进程 pbs\_server, TORQUE 服务器守护进程 pbs\_server 将读取 NFS 文件共享存储系统上离当前时间最近的全局进程映像文件,以卷回的方式重新提交任务。

## 基于 TORQUE 的并行检查点执行方法

### 技术领域

[0001] 本发明属于计算机领域,具体涉及一种基于 TORQUE 的并行检查点执行方法。

### 背景技术

[0002] 作业调度系统是高性能计算机系统基础的应用管理软件,其功能与性能直接关系到整个计算机系统的效率以及可靠性。但是在并行平台上对检查点技术并没能做到广泛的支持,而检查点的一个典型应用就是容错。

[0003] 检查点现有技术:Libckpt:集成在 Condor 操作系统中的检查点系统。由于缺乏内核的支持,它只能对有限的用户进程进行设置检查点,使得它在机群/作业管理系统中的应用变得困难。

[0004] SGI 的 Irix:内核级实现,提供了丰富的用户接口和实际应用。但是 Irix 不是一款开放源代码的系统。

[0005] Eckpt:基于 Linux 的检查点系统,未作存储优化,效率低下。

[0006] 现有的 TORQUE 虽然已经实现了检查点技术,但是只能对单进程的任务进行检查点设置,且无法支持对多进程进行检查点和进程的迁移。存在这些问题的原因如下:TORQUE 只是结合了 BLCR(Berkeley Lab's Linux Checkpoint/Restart 伯克利实验室实现的检查点与恢复技术)进行进程检查点。而 BLCR 自身无法支持分布式多进程任务;而且只在任务运行节点上存有检查点映像文件,其他节点无法利用该文件,也就无法做到进程迁移。

### 发明内容

[0007] 为克服上述缺陷,本发明提供了一种基于 TORQUE 的并行检查点执行方法,运用改进的基于 TORQUE 的改进检查点技术,使得原来只支持单进程检查点的 TORQUE,现在也能够支持多进程检查点,并能够对节点故障自动处理,对进程进行迁移。

[0008] 为实现上述目的,本发明提供一种基于 TORQUE 的并行检查点执行方法,基于 TORQUE 的并行检查点执行方法,在 NFS 文件共享存储系统中对计算节点进行检查点操作,其改进之处在于,所述方法包括如下步骤:

[0009] (1). 用户向 TORQUE 服务器守护进程 pbs\_server 提交作业;

[0010] (2). TORQUE 服务器守护进程发送任务消息给 TORQUE 调度器 pbs\_sched, TORQUE 调度器 pbs\_sched 根据作业中指定的参数要求寻找计算节点;

[0011] (3). 对计算节点进行检查点操作。

[0012] 本发明提供的优选技术方案中,在所述步骤 1 中,用户通过 TORQUE 提交作业命令 qsub 提交作业脚本,作业脚本使用 MPI 进程启动命令 chkp\_mpirun 启动 MPI,在提交作业命令中加入作业检查点请求。

[0013] 本发明提供的第二优选技术方案中,在所述步骤 2 中, TORQUE 调度器 pbs\_sched 通过计算节点守护进程 pbs\_mom 轮询各计算节点的状态,将结果返回给 TORQUE 服务器守护进程 pbs\_server。

[0014] 本发明提供的第三优选技术方案中,在所述步骤 3 中,TORQUE 服务器守护进程 pbs\_server 将用户的检查点操作请求传送到相应的计算节点上并周期运行;其中,各个计算节点分别设置有独立检查点镜像。

[0015] 本发明提供的第四优选技术方案中,周期进行检查点操作的各个计算节点将独立检查点镜像合并成全局检查点文件。

[0016] 本发明提供的第五优选技术方案中,全局检查点文件存储在 NFS 文件共享存储系统中。

[0017] 本发明提供的第六优选技术方案中,在所述步骤 3 后,存在以下步骤 a:

[0018] (a). 在轮询各个计算节点时,如果发现某个计算节点故障,则将该计算节点上分配的所有未执行完的和未执行的任务根据全局检查点文件全部进行卷回恢复。

[0019] 本发明提供的第七优选技术方案中,在所述步骤 a 中,TORQUE 调度器 pbs\_sched 将故障计算节点的所有任务号发回给 TORQUE 服务器守护进程 pbs\_server, TORQUE 服务器守护进程 pbs\_server 将读取 NFS 文件共享存储系统上离当前时间最近的全局进程映像文件,以卷回的方式重新提交任务。

[0020] 与现有技术比,本发明提供一种基于 TORQUE 的并行检查点执行方法,基于 TORQUE 的改进检查点技术,可以使得 TORQUE 对运行于其上的,无论是单进程还是多进程的任务进行检查点,并在节点故障时自动进行卷回恢复,避免了由用户发现节点失效并重新提交作业的不便,以及重启作业不及时导致的系统利用率低下,从而提高了调度系统的可靠性;而且,经实验测试,运用改进的基于 TORQUE 的改进检查点技术,使得原来只支持单进程检查点的 TORQUE,现在也能够支持多进程检查点,并能够对节点故障自动处理,对进程进行迁移,从而减少了因节点故障造成的计算资源浪费。

## 附图说明

[0021] 图 1 为基于 TORQUE 的并行检查点执行方法的流程示意图。

## 具体实施方式

[0022] 涉及的关键技术术语的说明

[0023] 1TORQUE Tera-scale Open-source Resource and Queue manager 开源的万亿级计算资源与队列管理器

[0024] 2 检查点在一个事务结束而另一个事务即将开始的时候,对系统状态的一次快照。

[0025] 3 映像文件检查点文件,保存了进程执行的所有信息。

[0026] 4 进程迁移在不同的处理器上移动一个正在执行进程的能力,这些处理器通过网络互相连接而不是局部共享内存。

[0027] 5NFS Network File System 网络文件系统

[0028] 如图 1 所示,一种基于 TORQUE 的并行检查点执行方法,基于 TORQUE 的并行检查点执行方法,在 NFS 文件共享存储系统中对计算节点进行检查点操作,所述方法包括如下步骤:

[0029] (1). 用户向 TORQUE 服务器守护进程 pbs\_server 提交作业;

[0030] (2). TORQUE 服务器守护进程发送任务消息给 TORQUE 调度器 pbs\_sched, TORQUE 调

度器 pbs\_sched 根据作业中指定的参数要求寻找计算节点；

[0031] (3). 对计算节点进行检查点操作。

[0032] 在所述步骤 1 中,用户通过 TORQUE 提交作业命令 qsub 提交作业脚本,提交命令中需要加入作业检查点请求,作业脚本中使用 MPI 进程启动命令 chkp\_mpirun 启动 MPI,或者可以直接在客户端以命令行形式提交 MPI 作业,方法为在命令行执行 MPI 启动命令 chkp\_mpirun。

[0033] 在所述步骤 2 中,TORQUE 调度器 pbs\_sched 通过计算节点守护进程 pbs\_mom 轮询各计算节点的状态,将结果返回给 TORQUE 服务器守护进程 pbs\_server。

[0034] 在所述步骤 3 中,TORQUE 服务器守护进程 pbs\_server 将用户的检查点操作请求传送到相应的计算节点上并周期运行;其中,各个计算节点分别设置有独立检查点镜像。周期进行检查点操作的各个计算节点将独立检查点镜像合并成全局检查点文件。全局检查点文件存储在 NFS 文件共享存储系统中。

[0035] 在所述步骤 3 后,存在以下步骤 a:

[0036] (a). 在轮询各个计算节点时,如果发现某个计算节点故障,则将该计算节点上分配的所有未执行完的和未执行的任务根据全局检查点文件全部进行卷回恢复。在所述步骤 a 中,TORQUE 调度器 pbs\_sched 将故障计算节点的所有任务号发回给 TORQUE 服务器守护进程 pbs\_server, TORQUE 服务器守护进程 pbs\_server 将读取 NFS 文件共享存储系统上离当前时间最近的全局进程映像文件,以卷回的方式重新提交任务。

[0037] 通过以下实施例对基于 TORQUE 的并行检查点执行方法做进一步描述。

[0038] 基于 TORQUE 的并行检查点执行方法,包括:

[0039] 一、NFS 文件系统 (Network File System 网络文件系统)

[0040] 本方法采用了 NFS 共享存储来存储映像文件,各进程在为自己设置检查点时,其检查点文件实际上是被 NFS 发送到挂载在管理结点上的稳定存储阵列中,存储阵列的可靠性一般由其自身来保证 (RAID-Redundant Arrays of InexpensiveDisks 磁盘阵列机制)。用 NFS 共享文件系统提供的服务,可以使得所有的计算节点都能访问到该存储空间,为进程的迁移以及节点故障后在其他节点恢复提供了有利条件。

[0041] 二、并行检查点执行流程:

[0042] 1. 通常并行任务的实现都是基于 MPI (Message Passing Interface 消息传递接口) 而实现的,本方法在执行作业命令时不再使用 MPI 自身的 mpirun (MPI 启动命令),而是使用本方法提供的 chkp\_mpirun 命令 (本方法提供的 MPI 进程启动命令),具体实现流程如下:

[0043] 2. 用户可以直接使用 chkp\_mpirun 命令在命令行启动 MPI 可执行程序,用户也可以通过 qsub 命令 (TORQUE 的提交作业命令) 提交作业脚本,作业脚本中使用 chkp\_mpirun 命令启动 MPI 可执行程序,并在提交命令中加入作业检查点请求;pbs\_server (TORQUE 在管理节点上的守护进程) 接收到用户脚本之后,发送任务消息给 pbs\_sched 调度器 (TORQUE 在管理节点上的调度器),调度器根据脚本文件中指定的参数要求寻找合适的节点。调度器通过 pbs\_mom (TORQUE 在计算节点上的守护进程) 轮询各计算节点的状态,将结果返回给 pbs\_server, pbs\_server 再将用户的任务程序由网络传送到相应的计算节点,并在该节点上运行,同时在该节点上周期进行检查点,启动任务的节点会将个计算节点上的独立检查

点镜像合并成全局一致的检查点镜像,从而达到单一系统映像效果。并将这个周期性产生的检查点文件存放在 NFS 共享存储上。

[0044] 3. 实现自动卷回恢复也就是实现进程迁移,需要管理节点定时轮询各计算节点是状态。如果发现节点故障,则将该节点上分配的所有未执行完的和未执行的任务全部进行卷回恢复。这是通过修改 pbs\_sched 调度器的定时执行函数来实现的。pbs\_sched 调度器会将故障节点的所有任务号发回给 pbs\_server, pbs\_server 将读取共享存储上的离当前时间最近的全局进程映像文件,以卷回的方式重新提交任务。

[0045] 需要声明的是,本发明内容及具体实施方式意在证明本发明所提供技术方案的实际应用,不应解释为对本发明保护范围的限定。本领域技术人员在本发明的精神和原理启发下,可作各种修改、等同替换、或改进。但这些变更或修改均在申请待批的保护范围内。

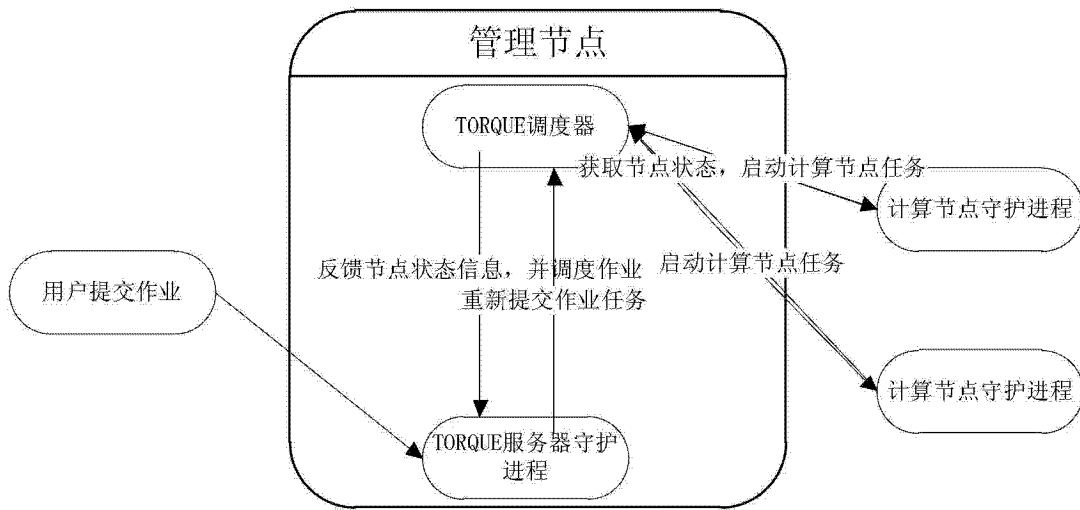


图 1