



(12) 发明专利申请

(10) 申请公布号 CN 103713955 A

(43) 申请公布日 2014. 04. 09

(21) 申请号 201310752071. 2

(22) 申请日 2013. 12. 31

(71) 申请人 曙光云计算技术有限公司

地址 100193 北京市海淀区东北旺西路 8 号  
院 36 号楼 5 层

(72) 发明人 马玉国 李亦民 张平

(74) 专利代理机构 北京德恒律治知识产权代理  
有限公司 11409

代理人 章社杲 孙征

(51) Int. Cl.

G06F 9/50 (2006. 01)

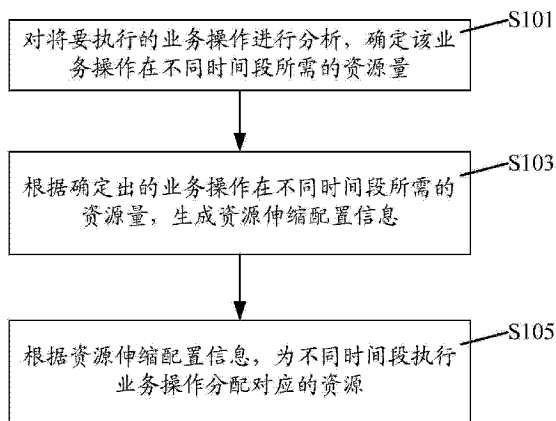
权利要求书2页 说明书5页 附图1页

(54) 发明名称

资源动态分配的管理方法和装置

(57) 摘要

本发明公开了一种资源动态分配的管理方法和装置,该资源动态分配的管理方法包括:对将要执行的业务操作进行分析,确定该业务操作在不同时间段所需的资源量;根据确定出的业务操作在不同时间段所需的资源量,生成资源伸缩配置信息;根据资源伸缩配置信息,为不同时间段执行业务操作分配对应的资源。本发明通过预先对将要执行的业务操作进行分析,确定该业务操作在不同时间段所需的资源量,然后,根据该资源量,生成资源伸缩配置信息,并根据该资源伸缩配置信息,为不同时间执行业务操作分配对应的资源,从而实现了在不同的时间段,采用不同的伸缩策略进行资源的动态分配,进而有效的提高了资源池的使用效率,使得资源池更具有弹性。



1. 一种资源动态分配的管理方法,其特征在于,包括:  
对将要执行的业务操作进行分析,确定该业务操作在不同时间段所需的资源量;  
根据确定出的所述业务操作在不同时间段所需的资源量,生成资源伸缩配置信息;  
根据所述资源伸缩配置信息,为不同时间段执行所述业务操作分配对应的资源。
2. 根据权利要求1所述的管理方法,其特征在于,根据确定出的所述业务操作在不同时间段所需的资源量,生成资源伸缩配置信息包括:  
将确定出的所述业务操作在不同时间段所需的资源量,作为该对应时间段执行所述业务操作的默认资源分配量;  
并且,根据不同时间段的所述默认资源分配量之间的差异关系,生成资源伸缩配置信息。
3. 根据权利要求2所述的管理方法,其特征在于,进一步包括:  
预先为业务操作的不同时间段配置多个备份资源分配量,其中,所述备份资源分配量大于所述默认资源分配量。
4. 根据权利要求3所述的管理方法,其特征在于,进一步包括:  
在执行所述业务操作时,实时检测所述业务操作在不同时间段所需的实际资源。
5. 根据权利要求4所述的管理方法,其特征在于,进一步包括:  
根据所述业务操作在每个时间段所需的实际资源,判断该时间段的所述默认资源分配量是否满足所述实际资源;  
在判断结果为否的情况下,将该时间段中满足所述实际资源的备份资源分配量作为当前的资源分配量,并根据该资源分配量为该时间段内的所述业务操作分配资源。
6. 一种资源动态分配的管理装置,其特征在于,包括:  
分析模块,用于对将要执行的业务操作进行分析,确定该业务操作在不同时间段所需的资源量;  
生成模块,用于根据确定出的所述业务操作在不同时间段所需的资源量,生成资源伸缩配置信息;  
执行模块,用于根据所述资源伸缩配置信息,为不同时间段执行所述业务操作分配对应的资源。
7. 根据权利要求6所述的管理装置,其特征在于,所述生成模块包括:  
第一配置模块,用于将确定出的所述业务操作在不同时间段所需的资源量,作为在对应时间段执行所述业务操作的默认资源分配量;  
信息生成模块,用于根据不同时间段的所述默认资源分配量之间的差异关系,生成资源伸缩配置信息。
8. 根据权利要求7所述的管理装置,其特征在于,进一步包括:  
第二配置模块,用于预先为业务操作的不同时间段配置多个备份资源分配量,其中,所述备份资源分配量大于所述默认资源分配量。
9. 根据权利要求8所述的管理装置,其特征在于,进一步包括:  
实时检测模块,用于在执行所述业务操作时,实时检测所述业务操作在不同时间段所需的实际资源。
10. 根据权利要求9所述的管理装置,其特征在于,进一步包括:

判断模块,用于根据所述业务操作在每个时间段所需的实际资源,判断该时间段的所述默认资源分配量是否满足所述实际资源;

第三配置模块,用于在判断结果为否的情况下,将该时间段中满足所述实际资源的备份资源分配量作为当前的资源分配量,并根据该资源分配量为该时间段内的所述业务操作分配资源。

## 资源动态分配的管理方法和装置

### 技术领域

[0001] 本发明涉及计算机领域,具体来说,涉及一种资源动态分配的管理方法和装置。

### 背景技术

[0002] 目前,现有的 Paas (Platform-as-a-Service, 平台即服务) 平台基本都支持资源池的自动伸缩(即资源的自动申请和释放)策略。现有自动伸缩策略有两种,一种是采用定值伸缩值来实现资源的自动伸缩,另一种是采用设定的伸缩率来实现资源的自动伸缩。

[0003] 然而,在实际应用时,不管是第一种自动伸缩策略,还是第二种自动伸缩策略,均存在着不足之处,例如,伸缩动作不够灵活。使用不够方便,且当用户对系统不够熟悉时,设定的值存在过大或过小的情况,进而影响了资源池的使用效率。

[0004] 针对相关技术中的上述问题,目前尚未提出有效的解决方案。

### 发明内容

[0005] 针对相关技术中的上述问题,本发明提出一种资源动态分配的管理方法和装置,能够实现在不同的时间段,采用不同的伸缩策略进行资源的动态分配,从而有效的提高了资源池的使用效率,使的资源池更具有弹性。

[0006] 本发明的技术方案是这样实现的:

[0007] 根据本发明的一个方面,提供了一种资源动态分配的管理方法。

[0008] 该资源动态分配的管理方法包括:

[0009] 对将要执行的业务操作进行分析,确定该业务操作在不同时间段所需的资源量;

[0010] 根据确定出的业务操作在不同时间段所需的资源量,生成资源伸缩配置信息;

[0011] 根据资源伸缩配置信息,为不同时间段执行业务操作分配对应的资源。

[0012] 其中,在根据确定出的业务操作在不同时间段所需的资源量,生成资源伸缩配置信息时,可将确定出的业务操作在不同时间段所需的资源量,作为在对应时间段执行业务操作的默认资源分配量;并根据不同时间段的默认资源分配量之间的差异关系,生成资源伸缩配置信息。

[0013] 此外,该资源动态分配的管理方法还包括:预先为业务操作的不同时间段配置多个备份资源分配量,其中,备份资源分配量大于默认资源分配量。

[0014] 另外,该资源动态分配的管理方法还包括:在执行业务操作时,实时检测业务操作在不同时间段所需的实际资源。

[0015] 此外,该资源动态分配的管理方法还包括:根据业务操作在每个时间段所需的实际资源,判断该时间段的默认资源分配量是否满足实际资源;在判断结果为否的情况下,将该时间段中满足实际资源的备份资源分配量作为当前的资源分配量,并根据该资源分配量为该时间段内的业务操作分配资源。

[0016] 根据本发明的另一方面,提供了一种资源动态分配的管理装置。

[0017] 该资源动态分配的管理装置包括:

[0018] 分析模块,用于对将要执行的业务操作进行分析,确定该业务操作在不同时间段所需的资源量;

[0019] 生成模块,用于根据确定出的业务操作在不同时间段所需的资源量,生成资源伸缩配置信息;

[0020] 执行模块,用于根据资源伸缩配置信息,为不同时间段执行业务操作分配对应的资源。

[0021] 其中,生成模块包括:第一配置模块,用于将确定出的业务操作在不同时间段所需的资源量,作为在对应时间段执行业务操作的默认资源分配量;信息生成模块,用于根据不同时间段的默认资源分配量之间的差异关系,生成资源伸缩配置信息。

[0022] 此外,该资源动态分配的管理装置还包括:第二配置模块,用于预先为业务操作的不同时间段配置多个备份资源分配量,其中,备份资源分配量大于默认资源分配量。

[0023] 另外,该资源动态分配的管理装置还包括:实时检测模块,用于在执行业务操作时,实时检测业务操作在不同时间段所需的实际资源。

[0024] 此外,该资源动态分配的管理装置还包括:判断模块,用于根据每个时间段所需的实际资源,判断该时间段的默认资源分配量是否满足实际资源;第三配置模块,用于在判断结果为否的情况下,将该时间段中满足实际资源的备份资源分配量作为当前的资源分配量,并根据该资源分配量为该时间段内的业务操作分配资源。

[0025] 本发明通过预先对将要执行的业务操作进行分析,确定该业务操作在不同时间段所需的资源量,然后,根据该资源量,生成资源伸缩配置信息,并根据该资源伸缩配置信息,为不同时间执行业务操作分配对应的资源,从而实现了在不同的时间段,采用不同的伸缩策略进行资源的动态分配,进而有效的提高了资源池的使用效率,使得资源池更具有弹性。

[0026] 此外,本发明还通过为不同时间段配置多个备份资源分配量,并同时实时检测业务操作在不同时间段所需的实际资源,当默认资源分配量不满足所需的实际资源时,将多个备份资源分配量中满足所需实际资源的备份资源分配量作为当前的资源分配量,并根据该资源分配量为该时间段内的业务操作分配对应的资源,从而实现了根据业务操作实际所需的资源,动态调整资源分配量,以满足实际的需求,进而避免了设定的资源分配量过大或过小的问题发生。

## 附图说明

[0027] 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,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅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的前提下,还可以根据这些附图获得其他的附图。

[0028] 图 1 是根据本发明实施例的资源动态分配的管理方法的流程示意图;

[0029] 图 2 是根据本发明实施例的资源动态分配的管理装置的结构示意图。

## 具体实施方式

[0030] 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,所描述的实施例仅仅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所获得的所有其他实施例,都属于本发明保护的范围。

[0031] 根据本发明的实施例,提供了一种资源动态分配的管理方法。

[0032] 如图 1 所示,根据本发明实施例的资源动态分配的管理方法包括:

[0033] 步骤 S101,对将要执行的业务操作进行分析,确定该业务操作在不同时间段所需的资源量;

[0034] 步骤 S103,根据确定出的业务操作在不同时间段所需的资源量,生成资源伸缩配置信息;

[0035] 步骤 S105,根据资源伸缩配置信息,为不同时间段执行业务操作分配对应的资源。

[0036] 其中,在根据确定出的业务操作在不同时间段所需的资源量,生成资源伸缩配置信息时,可将确定出的业务操作在不同时间段所需的资源量,作为在对应时间段执行业务操作的默认资源分配量;并根据不同时间段的默认资源分配量之间的差异关系,生成资源伸缩配置信息。

[0037] 此外,该资源动态分配的管理方法还包括:预先为业务操作的不同时间段配置多个备份资源分配量,其中,备份资源分配量大于默认资源分配量。

[0038] 另外,该资源动态分配的管理方法还包括:在执行业务操作时,实时检测业务操作在不同时间段所需的实际资源。

[0039] 此外,该资源动态分配的管理方法还包括:根据业务操作在每个时间段所需的实际资源,判断该时间段的默认资源分配量是否满足实际资源;在判断结果为否的情况下,将该时间段中满足实际资源的备份资源分配量作为当前的资源分配量,并根据该资源分配量为该时间段内的业务操作分配资源。

[0040] 根据本发明的实施例,还提供了一种资源动态分配的管理装置。

[0041] 如图 2 所示,根据本发明实施例的资源动态分配的管理装置包括:

[0042] 分析模块 21,用于对将要执行的业务操作进行分析,确定该业务操作在不同时间段所需的资源量;

[0043] 生成模块 22,用于根据确定出的业务操作在不同时间段所需的资源量,生成资源伸缩配置信息;

[0044] 执行模块 23,用于根据资源伸缩配置信息,为不同时间段执行业务操作分配对应的资源。

[0045] 其中,生成模块 22 包括:第一配置模块(未示出),用于将确定出的业务操作在不同时间段所需的资源量,作为在对应时间段执行业务操作的默认资源分配量;信息生成模块(未示出),用于根据不同时间段的默认资源分配量之间的差异关系,生成资源伸缩配置信息。

[0046] 此外,该资源动态分配的管理装置还包括:第二配置模块(未示出),用于预先为业务操作的不同时间段配置多个备份资源分配量,其中,备份资源分配量大于默认资源分配量。

[0047] 另外,该资源动态分配的管理装置还包括:实时检测模块(未示出),用于在执行业务操作时,实时检测业务操作在不同时间段所需的实际资源。

[0048] 此外,该资源动态分配的管理装置还包括:判断模块(未示出),用于根据每个时间

段所需的实际资源,判断该时间段的默认资源分配量是否满足实际资源;第三配置模块(未示出),用于在判断结果为否的情况下,将该时间段中满足实际资源的备份资源分配量作为当前的资源分配量,并根据该资源分配量为该时间段内的业务操作分配资源。

[0049] 为了方便理解本发明的上述技术方案,以下从技术原理的角度对本发明的上述技术方案进行说明。

[0050] 在实际应用时,实践表明,对于一个资源池来说,资源池的使用一般不会是一个线性的过程,而是开始增长比较快,到一定使用量后,增长开始放缓,最后基本保持一个比较平稳的使用状态。因此,对资源池伸缩的策略,可以设定一个参考曲线,设定曲线的 X 轴为时长,一般定位 12 个月,Y 轴定义为资源池伸缩的比率,这样资源池在不同的时段将使用不同的伸缩策略,这将极大的提高资源池的使用效率,使的资源池更有弹性。而在预测数据更改时,伸缩服务可以使用新的参数进行伸缩。

[0051] 在实际应用时,本发明的上述技术方案可通过以下步骤来实现,具体如下:

[0052] 1) 定义一个数据表,保存资源池伸缩率的预测数据,例如:

[0053] Table(Id, PolicyId, MonthKey, Value),

[0054] 其中,表的数据包括策略 ID、月份、伸缩率值。

[0055] 2) 定义一个类,用来读取资源池伸缩率数据,例如,

[0056]

```

Public class ReadData {

    Float[] myData=dao.getShinkData(policyId){

        .....

    };

}

```

[0057] 3) 定义一个资源池类,用来保存资源池资源实例。

[0058] 4) 定义一个监控类,用来监控资源池状态,并读取当前资源池的伸缩率。当资源池实例数发生变动时,可通过该监控类读取当前阶段设定的伸缩比率,按照预先设定的伸缩曲线,读取对应的伸缩值,进行伸缩操作。

[0059] 由此可见,借助于本发明的上述技术方案,通过预先对将要执行的业务操作进行分析,确定该业务操作在不同时间段所需的资源量,然后,根据该资源量,生成资源伸缩配置信息,并根据该资源伸缩配置信息,为不同时间执行业务操作分配对应的资源,从而可以在不同的时间段,采用不同的伸缩策略进行资源的动态分配,进而有效的提高了资源池的使用效率,使得资源池更具有弹性。

[0060] 此外,借助于本发明的上述技术方案,通过为不同时间段配置多个备份资源分配量,并同时实时检测业务操作在不同时间段所需的实际资源,当默认资源分配量不满足所需的实际资源时,将多个备份资源分配量中满足所需实际资源的备份资源分配量作为当前的资源分配量,并根据该资源分配量为该时间段内的业务操作分配对应的资源,从而可以实现根据业务操作实际所需的资源,动态调整资源分配量,以满足实际的需求,进而避免了

设定的资源分配量过大或过小的问题发生。

[0061] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。



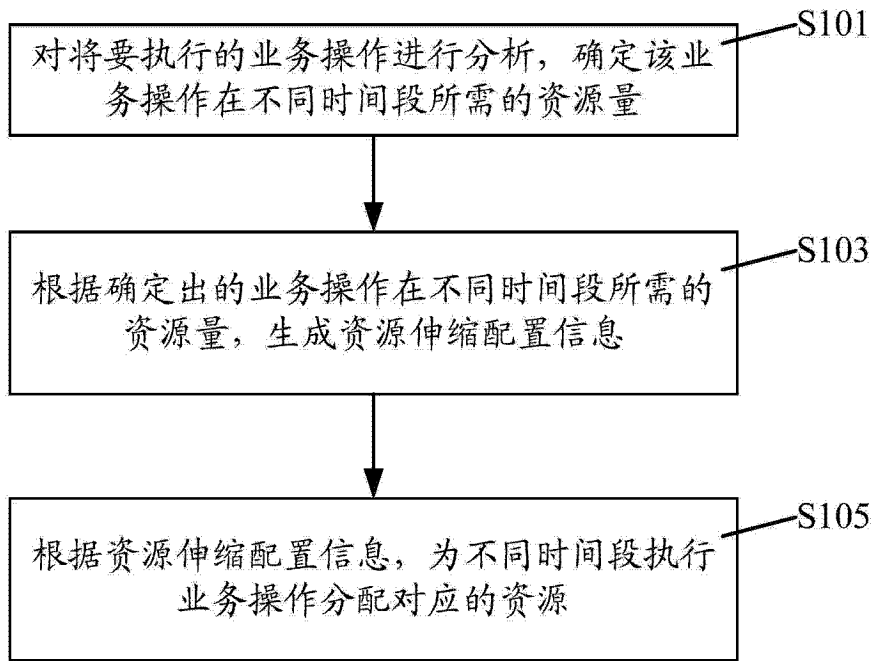


图 1

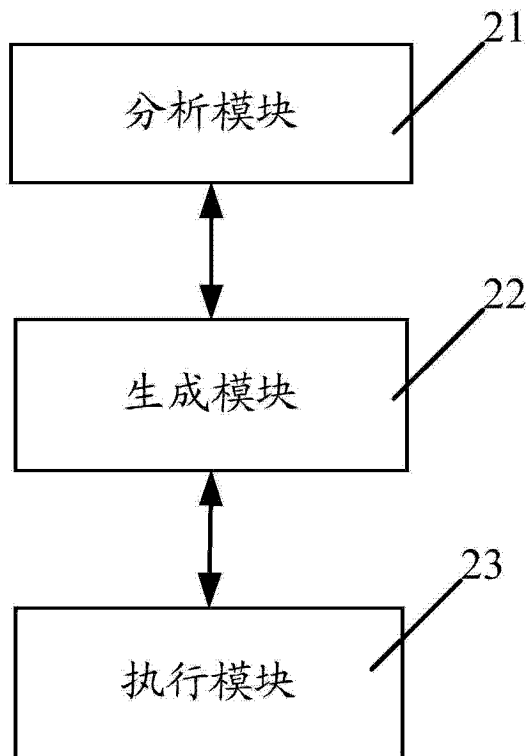


图 2