



(12) 发明专利

(10) 授权公告号 CN 114138477 B

(45) 授权公告日 2022.06.03

(21) 申请号 202111406061.4
 (22) 申请日 2021.11.24
 (65) 同一申请的已公布的文献号
 申请公布号 CN 114138477 A
 (43) 申请公布日 2022.03.04
 (73) 专利权人 中国人民解放军军事科学院战争
 研究院
 地址 100091 北京市海淀区厢红旗路东门
 外一号
 (72) 发明人 李华莹 刘必欣 刘怡静 薛源
 张捷 刘丽 杨超
 (74) 专利代理机构 北京中创云知识产权代理事
 务所(普通合伙) 11837
 专利代理师 徐辉
 (51) Int.Cl.
 G06F 9/50 (2006.01)
 G06F 11/34 (2006.01)
 G06Q 10/04 (2012.01)
 (56) 对比文件
 CN 106686136 A,2017.05.17

CN 113190351 A,2021.07.30
 CN 109766182 A,2019.05.17
 CN 103945548 A,2014.07.23
 CN 110535894 A,2019.12.03
 US 2016164797 A1,2016.06.09
 US 2020167203 A1,2020.05.28
 CN 102355736 A,2012.02.15
 李华莹等.面向软件生态的资源定位技术.
 《计算机与现代化》.2020,(第03期),第24-28页.
 S. K. Mishra等.Time efficient dynamic
 threshold-based load balancing technique
 for Cloud Computing.《2017 International
 Conference on Computer, Information and
 Telecommunication Systems (CITS)》.2017,第
 161-165页.

(续)

审查员 苗文娟

权利要求书4页 说明书11页 附图1页

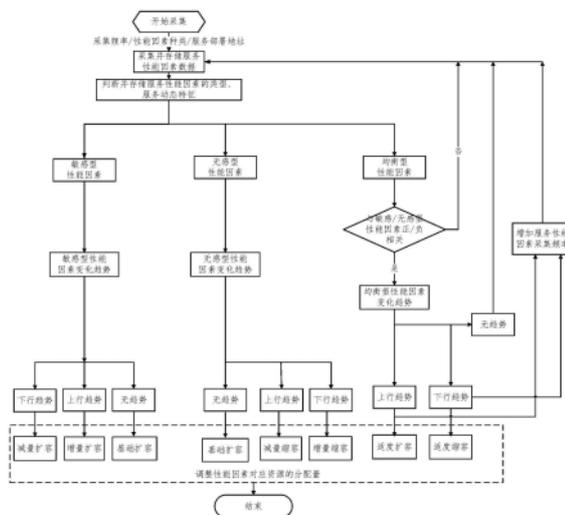
(54) 发明名称

一种信息系统运行态服务资源分配方法

(57) 摘要

本发明公开一种信息系统运行态服务资源分配方法,采集并存储服务的运行状态信息;基于各运行状态信息,判断该运行状态信息对应性能因素的类型;判断服务各个性能因素之间的相关性,作为服务动态特征;基于一段时间内的运行状态信息,预测服务性能因素变化趋势;对于基于敏感型、无感型的性能因素基于变化趋势调整对应资源的分配量;对于均衡型的性能因素基于与其相关的性能因素的变化趋势,判断均衡型的性能因素的变化趋势,调整对应资源的分配量。本发明实现了信息系统运行态服务资源的精准自动化分配,无需依赖操作人员的经验,在满足服务需求的同时,提高了资源利用效率。

CN 114138477 B



[接上页]

(56) 对比文件

Bixin Liu等.QoS aware service composition with multiple quality constraints.《NPC'05: Proceedings of the 2005 IFIP international conference on

Network and Parallel Computing》.2005,第123-131页.

李丽娜等.大规模流数据处理中代价有效的弹性资源分配策略.《吉林大学学报(工学版)》.2020,第50卷(第05期),第1832-1843页.

1. 一种信息系统运行态服务资源分配方法,其特征在于,包括:

采集并存储服务的运行状态信息,包括CPU占用率、内存使用率、磁盘使用率、网络流量占用率和并发访问占用率;

基于各运行状态信息,判断该运行状态信息对应性能因素的类型为敏感型、无感型或均衡型;

判断服务各个性能因素之间的相关性,作为服务动态特征;

基于一段时间内的运行状态信息,预测服务性能因素变化趋势;

对于基于敏感型、无感型的性能因素基于变化趋势调整对应资源的分配量;对于均衡型的性能因素基于与其相关的性能因素的变化趋势,判断均衡型的性能因素的变化趋势,调整对应资源的分配量;

服务的动态特征包括:因素正相关、因素负相关和因素无相关;

判断该运行状态信息对应性能因素的类型,包括:

计算该性能因素采集数据的平均值:

$$\bar{A} = \frac{\sum_{i=0}^n C_i}{n}$$

其中 C_i 表示某种服务性能因素的第 i 个采集值, n 表示采集周期内的采集数量;计算某种服务性能因素的类型 T :

$$T = \begin{cases} T_+, \bar{A} > R_{max} \\ T_-, \bar{A} < R_{mix} \\ T_0, R_{mix} < \bar{A} < R_{max} \end{cases}$$

式中, T_+ 表示敏感性, T_- 表示无感型, T_0 表示均衡型, R_{max} 表示该服务性能因素均值的上限阈值, R_{min} 表示该服务性能因素均值的下限阈值;

预测服务性能因素变化趋势,包括:

对一段时间内的某运行状态信息进行拟合,计算预测时间内拟合曲线的斜率值:

$$\varphi = F'(t), t \in (t_1, t_2)$$

上式中, $F(t)$ 表示曲线拟合表达式, t_1 表示预测起始时间, t_2 表示预测的结束时间;

变化趋势为:

$$L = \begin{cases} L_+, \varphi > +|\sigma| \\ L_0, |\varphi| > |\sigma| \\ L_-, \varphi < -|\sigma| \end{cases}$$

其中 L 表示服务性能因素的变化趋势; L_+ 表示上行趋势; L_0 表示无变化趋势; L_- 表示下行趋势; σ 为经验阈值;

调整对应资源的分配量,包括:对敏感类型的性能因素资源分配方案,如果其变化趋势是无变化,应当按基础调节量对资源进行扩容;如果其变化趋势是上行,应当增大调节量对资源进行扩容;如果其变化趋势是下行,应当减小调节量对资源进行扩容;

对无感类型的性能因素资源分配方案,如果其变化趋势是无变化,应当按基础调节量

对资源进行扩容;如果其变化趋势是下行,应当增大调节量对资源进行扩容;如果其变化趋势是上行,应当减小调节量对资源进行扩容;

对均衡类型的性能因素资源分配方案,如果其变化趋势是无变化,不对其资源进行调节;如果其变化趋势是上行,应当增大调节量对资源进行扩容,如果其变化趋势是下行,应当减少调节量对资源进行扩容。

2. 根据权利要求1所述的信息系统运行态服务资源分配方法,其特征在于,判断服务各个性能因素之间的相关性,包括:

利用斯皮尔曼秩关系因子定义服务 s_i 的性能因素 x 和 y 在采集周期时间窗 j 内的关系值:

$$c_{i,j}(x, y) = \frac{\sum_{k=1}^w (m'_{i,x,(j-1) \times w+k} - \overline{m'_{i,x,j}})(m'_{i,y,(j-1) \times w+k} - \overline{m'_{i,y,j}})}{\sqrt{\sum_{k=1}^w (m'_{i,x,(j-1) \times w+k} - \overline{m'_{i,x,j}})^2 \sum_{k=1}^w (m'_{i,y,(j-1) \times w+k} - \overline{m'_{i,y,j}})^2}}$$

式中, $m_{i,x,k}$ 和 $m_{i,y,k}$ 分别表示服务 s_i 在采集周期时间窗 j 内采集的性能因素 x 和 y 的第 k 个数据; w 表示在采集周期时间窗 j 内一共采集了 w 个性能因素数据;将一个时间窗 j 中采集的服务 s_i 的性能因素 x 和 y 的所有采集数据 $m_{i,x,k}$ 和 $m_{i,y,k}$ 按从大到小的顺序排序后,得到排序序列 $V'_{i,x,j}$ 和 $V'_{i,y,j}$; $m'_{i,x,(j-1) \times w+k}$ 和 $m'_{i,y,(j-1) \times w+k}$ 分别为性能因素 x 的采集值 $m_{i,x,(j-1) \times w+k}$ 和性能因素 y 的采集值 $m_{i,y,(j-1) \times w+k}$ 在排序序列 $V'_{i,x,j}$ 和 $V'_{i,y,j}$ 中的秩次; $\overline{m'_{i,x,j}}$ 和 $\overline{m'_{i,y,j}}$ 分别为排序序列 $V'_{i,x,j}$ 和 $V'_{i,y,j}$ 的平均秩次;

服务动态特征计算公式如下:

$$w_{i,j}(x, y) = \begin{cases} W_+, c_{i,j}(x, y) >+ L_{\max} \\ W_0, |c_{i,j}(x, y)| < L_{\min} \\ W_-, c_{i,j}(x, y) <- L_{\max} \end{cases}$$

上式中, $w_{i,j}(x, y)$ 表示服务 s_i 在采集周期时间窗 j 内采集的性能因素 x 和 y 的动态特征, W_+ 表示性能因素 x 和 y 正相关, W_0 表示性能因素 x 和 y 无相关, W_- 表示性能因素 x 和 y 负相关, L_{\max} 表示服务特征判断上限阈值, L_{\min} 表示服务特征判断下限阈值。

3. 根据权利要求1或2所述的信息系统运行态服务资源分配方法,其特征在于,对于敏感型,变化趋势为无变化的性能因素对应资源的分配量调整方式为:

$$C_0(t+1) = C(t) + \left| \bar{A} - R_c \right| \cdot C(t)$$

其中, $C_0(t+1)$ 为敏感型,变化趋势为无变化的性能因素当前资源分配量, $C(t)$ 为该性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_c 为敏感型性能因素的经验阈值;

对于敏感型变化趋势为上行的性能因素对应资源的分配量调整方式为:

$$C_+(t+1) = C(t) + \left| \bar{A} - R_c \right| \cdot \left(1 + \frac{x'(t+1) - x(t)}{x(t)} \right) C(t)$$

其中, $C_+(t+1)$ 为敏感型变化趋势为上行的性能因素当前资源分配量, $C(t)$ 为该性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_c 为敏感型性能因素的经验阈值, $x'(t+1)$ 为预测的 $t+1$ 时刻的该性能因素数据, $x(t)$ 为当前采集的该性能因素数据;

对于敏感型变化趋势为下行的性能因素对应资源的分配量调整方式为：

$$C_{-}(t+1) = C(t) + \left| \bar{A} - R_c \right| \cdot \left(1 + \frac{x'(t+1) - x(t)}{x(t)} \right) C(t)$$

其中, $C_{-}(t+1)$ 为敏感型变化趋势为下行的性能因素当前资源分配量, $C(t)$ 为该类型性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_c 为敏感型性能因素的经验阈值, $x'(t+1)$ 为预测的 $t+1$ 时刻的该性能因素数据, $x(t)$ 为当前采集的该性能因素数据。

4. 根据权利要求1或2所述的信息系统运行态服务资源分配方法, 其特征在于, 对于无感型, 变化趋势为无变化的性能因素对应资源的分配量调整方式为：

$$D_0(t+1) = D(t) - \left| \bar{A} - R_D \right| \cdot D(t)$$

其中, $D_0(t+1)$ 为无感型, 变化趋势为无变化的性能因素当前资源分配量, $D(t)$ 为该性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_D 为无感型性能因素的经验阈值；

对于无感型, 变化趋势为下行的性能因素对应资源的分配量调整方式为：

$$D_{-}(t+1) = D(t) - \left| \bar{A} - R_D \right| \cdot \left(1 + \frac{x(t) - x'(t+1)}{x(t)} \right) D(t)$$

其中, $D_{-}(t+1)$ 为无感型, 变化趋势为下行的性能因素当前资源分配量, $D(t)$ 为该性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_D 为无感型性能因素的经验阈值, $x'(t+1)$ 为预测的 $t+1$ 时刻的该性能因素数据, $x(t)$ 为当前采集的该性能因素数据；

对于无感型, 变化趋势为上行的性能因素对应资源的分配量调整方式为：

$$D_{+}(t+1) = D(t) - \left| \bar{A} - R_D \right| \cdot \left(1 + \frac{x(t) - x'(t+1)}{x(t)} \right) D(t)$$

其中, $D_{+}(t+1)$ 为无感型, 变化趋势为上行的性能因素当前资源分配量, $D(t)$ 为该性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_D 为无感型性能因素的经验阈值, $x'(t+1)$ 为预测的 $t+1$ 时刻的该性能因素数据, $x(t)$ 为当前采集的该性能因素数据。

5. 根据权利要求1或2所述的信息系统运行态服务资源分配方法, 其特征在于, 对于均衡型的性能因素基于与其相关的性能因素的变化趋势, 判断均衡型的性能因素的变化趋势, 包括：

如果所有与均衡型性能因素呈正相关的敏感型和无感型性能因素的变化趋势均为上行趋势, 并且所有与均衡型性能因素呈负相关的敏感型和无感型性能因素的变化趋势也均为下行趋势或无变化趋势, 则认为该均衡型性能因素的变化趋势为上行趋势；

如果所有与均衡型性能因素呈负相关的敏感型和无感型性能因素的变化趋势均为下行趋势, 并且所有与均衡型性能因素呈正相关的敏感型和无感型性能因素的变化趋势均为上行趋势或无变化趋势, 则认为该均衡型性能因素的变化趋势为上行趋势；

如果所有与均衡型性能因素呈正相关的敏感型和无感型性能因素的变化趋势均为下行趋势, 并且所有与均衡型性能因素呈负相关的敏感型和无感型性能因素的变化趋势均为上行趋势或无变化趋势, 则认为该均衡型性能因素的变化趋势为下行趋势；

如果所有与均衡型性能因素呈负相关的敏感型和无感型性能因素的变化趋势均为上行趋势,并且所有与均衡型性能因素呈正相关的敏感型和无感型性能因素的变化趋势也均为下行趋势或无变化趋势,则认为该均衡型性能因素的变化趋势为下行趋势;

如果均衡型性能因素不是上行趋势或下行趋势则为无变化趋势。

6. 根据权利要求5所述的信息系统运行态服务资源分配方法,其特征在于,对于均衡型,变化趋势为无变化的性能因素不对其资源进行调节;

对于均衡型,变化趋势为上行的性能因素对应资源的分配量调整方式为:

$$E_{+}(t+1) = C(t) + \frac{|\bar{A} - R_{E_{+}}|}{2} E(t)$$

其中, $E_{+}(t+1)$ 为均衡型,变化趋势为上行的性能因素当前资源分配量, $E(t)$ 为该性能因素上一次的某类资源分配量, \bar{A} 表示该性能因素采集数据的平均值, $R_{E_{+}}$ 为均衡型性能因素在上行趋势的经验阈值;

对于均衡型,变化趋势为下行的性能因素对应资源的分配量调整方式为:

$$E_{-}(t+1) = C(t) - \frac{|\bar{A} - R_{E_{-}}|}{2} E(t)$$

其中, $E_{-}(t+1)$ 为均衡型,变化趋势为下行的性能因素当前资源分配量, $E(t)$ 为该性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, $R_{E_{-}}$ 为均衡型性能因素在下行趋势的经验阈值。

7. 根据权利要求5所述的信息系统运行态服务资源分配方法,其特征在于,对于均衡型,变化趋势为上行或下行的性能因素,增加数据的采集频率。

8. 根据权利要求1或2所述的信息系统运行态服务资源分配方法,其特征在于,对于每种服务的所有性能因素添加标签,标示各个性能因素的类型和变化趋势。

一种信息系统运行态服务资源分配方法

技术领域

[0001] 本发明涉及信息系统集成服务技术领域,尤其涉及一种信息系统运行态 服务资源分配方法。

背景技术

[0002] 随着虚拟化、云平台、容器技术的应用,信息系统由服务化向微服务化 转变,服务按业务功能进行拆分,直至单个服务的功能和职责单一,不可再 拆分为止,从而保障单个服务能独立部署、灵活扩缩容,提高资源的利用率。

[0003] 然而,信息系统微服务转型过程中,存在服务拆分粒度较大,致使单个 服务仍具有多种功能和职责。因此,服务在不同的应用场景下呈现出复杂的 资源需求特征。针对此类服务的资源分配,往往忽略服务运行状态信息,多 采用简单的人工预设策略,以简单比例关系对计算、网络、存储等资源进行 扩缩容操作,扩缩容粒度较大,极易造成资源的浪费和闲置。

[0004] 服务在不同应用场景和不同业务特征条件下,如何依据服务运行状态、资源需求特征,为服务进行资源精细化分配是亟待解决的问题。

发明内容

[0005] 本发明目的在于提供一种信息系统运行态服务资源分配方法及装置,解 决以往资源分配采用简单的人工预设策略,造成资源的浪费和闲置,难以应 对不同应用场景和不同业务特征条件下服务的差异性资源分配需求的问题。

[0006] 为达到上述目的,本发明提供了一种信息系统运行态服务资源分配方法, 包括:

[0007] 采集并存储服务的运行状态信息,包括CPU占用率、内存使用率、磁盘 使用率、网络流量占用率、并发访问占用率;

[0008] 基于各运行状态信息,判断该运行状态信息对应性能因素的类型为敏感 型、无感型或均衡型;

[0009] 判断服务各个性能因素之间的相关性,作为服务动态特征;

[0010] 基于一段时间内的运行状态信息,预测服务性能因素变化趋势;

[0011] 对于基于敏感型、无感型的性能因素基于变化趋势调整对应资源的分配 量;对于均衡型的性能因素基于与其相关的性能因素的变化趋势,判断均衡 型的性能因素的变化趋势,调整对应资源的分配量。

[0012] 进一步地,判断该运行状态信息对应性能因素的类型,包括:

[0013] 计算该性能因素采集数据的平均值:

$$[0014] \quad \bar{A} = \frac{\sum_{i=0}^n C_i}{n}$$

[0015] 其中 C_i 表示某种服务性能因素的第 i 个采集值, n 表示采集周期内的采 集数量;计

算某种服务性能因素的类型T:

$$[0016] \quad T = \begin{cases} T_+, \overline{A} > R_{max} \\ T_-, \overline{A} < R_{mix} \\ T_0, R_{mix} < \overline{A} < R_{max} \end{cases}$$

[0017] 式中, T_+ 表示敏感性, T_- 表示无感型, T_0 表示均衡型, R_{max} 表示该服务性能因素均值的上限阈值, R_{min} 表示该服务性能因素均值的下限阈值。

[0018] 进一步地, 判断服务各个性能因素之间的相关性, 包括:

[0019] 利用斯皮尔曼秩关系因子定义服务 s_i 的性能因素x和y在采集周期时间窗j内的关系值:

$$[0020] \quad c_{i,j}(x, y) = \frac{\sum_{k=1}^w (m'_{i,x,(j-1) \times w+k} - \overline{m'_{i,x,j}})(m'_{i,y,(j-1) \times w+k} - \overline{m'_{i,y,j}})}{\sqrt{\sum_{k=1}^w (m'_{i,x,(j-1) \times w+k} - \overline{m'_{i,x,j}})^2 \sum_{k=1}^w (m'_{i,y,(j-1) \times w+k} - \overline{m'_{i,y,j}})^2}}$$

[0021] 式中, $m_{i,x,k}$ 和 $m_{i,y,k}$ 分别表示服务 s_i 在采集周期时间窗j内采集的性能因素x和y的第k个数据; w表示在采集周期时间窗j内一共采集了w个性能因素数据; 将一个时间窗j中采集的服务 s_i 的性能因素x和y的所有采集数据 $m_{i,x,k}$ 和 $m_{i,y,k}$ 按从大到小的顺序排序后, 得到排序序列 $V'_{i,x,j}$ 和 $V'_{i,y,j}$; $m'_{i,x,(j-1) \times w+k}$ 和 $m'_{i,y,(j-1) \times w+k}$ 分别为性能因素x的采集值 $m_{i,x,(j-1) \times w+k}$ 和性能因素y的采集值 $m_{i,y,(j-1) \times w+k}$ 在排序序列 $V'_{i,x,j}$ 和 $V'_{i,y,j}$ 中的秩次; $\overline{m'_{i,x,j}}$ 和 $\overline{m'_{i,y,j}}$ 分别为排序序列 $V'_{i,x,j}$ 和 $V'_{i,y,j}$ 的平均秩次;

[0022] 服务动态特征计算公式如下:

$$[0023] \quad w_{i,j}(x, y) = \begin{cases} W_+, c_{i,j}(x, y) > +L_{max} \\ W_0, |c_{i,j}(x, y)| < L_{min} \\ W_-, c_{i,j}(x, y) < -L_{max} \end{cases}$$

[0024] 上式中, $w_{i,j}(x, y)$ 表示服务 s_i 在采集周期时间窗j内采集的性能因素x和y的动态特征, W_+ 表示性能因素x和y正相关, W_0 表示性能因素x和y无相关, W_- 表示性能因素x和y负相关, L_{max} 表示服务特征判断上限阈值, L_{min} 表示服务特征判断下限阈值。

[0025] 进一步地, 预测服务性能因素变化趋势, 包括:

[0026] 对一段时间内的某运行状态信息进行拟合, 计算预测时间内拟合曲线的斜率值:

$$[0027] \quad \varphi = F'(t), t \in (t_1, t_2)$$

[0028] 上式中, F(t)表示曲线拟合表达式, t_1 表示预测起始时间, t_2 表示预测的结束时间;

[0029] 变化趋势为:

$$[0030] \quad L = \begin{cases} L_+, \varphi > +|\sigma| \\ L_0, |\varphi| > |\sigma| \\ L_-, \varphi < -|\sigma| \end{cases}$$

[0031] 其中L表示服务性能因素的变化趋势; L_+ 表示上行趋势; L_0 表示无变化趋势; L_- 表

示下行趋势; σ 为经验阈值。

[0032] 进一步地,对于敏感型,变化趋势为无变化的性能因素对应资源的分配量调整方式为:

$$[0033] \quad C_0(t+1) = C(t) + |\bar{A} - R_C| \cdot C(t)$$

[0034] 其中, $C_0(t+1)$ 为敏感型,变化趋势为无变化的性能因素当前资源分配量, $C(t)$ 为该性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_C 为敏感型性能因素的经验阈值。

[0035] 进一步地,对于敏感型变化趋势为上行的性能因素对应资源的分配量调整方式为:

$$[0036] \quad C_+(t+1) = C(t) + |\bar{A} - R_C| \cdot \left(1 + \frac{x'(t+1) - x(t)}{x(t)}\right) C(t)$$

[0037] 其中, $C_+(t+1)$ 为敏感型变化趋势为上行的性能因素当前资源分配量, $C(t)$ 为该性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_C 为敏感型性能因素的经验阈值, $x'(t+1)$ 为预测的t+1时刻的该性能因素数据, $x(t)$ 为当前采集的该性能因素数据。

[0038] 进一步地,对于敏感型变化趋势为下行的性能因素对应资源的分配量调整方式为:

$$[0039] \quad C_-(t+1) = C(t) + |\bar{A} - R_C| \cdot \left(1 + \frac{x'(t+1) - x(t)}{x(t)}\right) C(t)$$

[0040] 其中, $C_-(t+1)$ 为敏感型变化趋势为下行的性能因素当前资源分配量, $C(t)$ 为该类型性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_C 为敏感型性能因素的经验阈值, $x'(t+1)$ 为预测的t+1时刻的该性能因素数据, $x(t)$ 为当前采集的该性能因素数据。

[0041] 进一步地,对于无感型,变化趋势为无变化的性能因素对应资源的分配量调整方式为:

$$[0042] \quad D_0(t+1) = D(t) - |\bar{A} - R_D| \cdot D(t)$$

[0043] 其中, $D_0(t+1)$ 为无感型,变化趋势为无变化的性能因素当前资源分配量, $D(t)$ 为该性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_D 为无感型性能因素的经验阈值。

[0044] 进一步地,对于无感型,变化趋势为上行的性能因素对应资源的分配量调整方式为:

$$[0045] \quad D_-(t+1) = D(t) - |\bar{A} - R_D| \cdot \left(1 + \frac{x(t) - x'(t+1)}{x(t)}\right) D(t)$$

[0046] 其中, $D_-(t+1)$ 为无感型,变化趋势为上行的性能因素当前资源分配量, $D(t)$ 为该性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_D 为无感型性能因素的经验阈值, $x'(t+1)$ 为预测的t+1时刻的该性能因素数据, $x(t)$ 为当前采集的该性能因素数据。

素数据。

[0047] 进一步地,对于无感型,变化趋势为上行的性能因素对应资源的分配量调整方式为:

$$[0048] \quad D_+(t+1) = D(t) - \left| \bar{A} - R_D \right| \cdot \left(1 + \frac{x(t) - x'(t+1)}{x(t)} \right) D(t)$$

[0049] 其中, $D_+(t+1)$ 为无感型,变化趋势为上行的性能因素当前资源分配量, $D(t)$ 为该性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_D 为无感型性能因素的经验阈值, $x'(t+1)$ 为预测的 $t+1$ 时刻的该性能因素数据, $x(t)$ 为当前采集的该性能因素数据。

[0050] 进一步地,对于均衡型的性能因素基于与其相关的性能因素的变化趋势,判断均衡型的性能因素的变化趋势,包括:

[0051] 如果所有与均衡型性能因素呈正相关的敏感型和无感型性能因素的变化趋势均为上行趋势,并且所有与均衡型性能因素呈负相关的敏感型和无感型性能因素的变化趋势也均为下行趋势或无变化趋势,则认为该均衡型性能因素的变化趋势为上行趋势;

[0052] 如果所有与均衡型性能因素呈负相关的敏感型和无感型性能因素的变化趋势均为下行趋势,并且所有与均衡型性能因素呈正相关的敏感型和无感型性能因素的变化趋势均为上行趋势或无变化趋势,则认为该均衡型性能因素的变化趋势为上行趋势;

[0053] 如果所有与均衡型性能因素呈正相关的敏感型和无感型性能因素的变化趋势均为下行趋势,并且所有与均衡型性能因素呈负相关的敏感型和无感型性能因素的变化趋势均为上行趋势或无变化趋势,则认为该均衡型性能因素的变化趋势为下行趋势;

[0054] 如果所有与均衡型性能因素呈负相关的敏感型和无感型性能因素的变化趋势均为上行趋势,并且所有与均衡型性能因素呈正相关的敏感型和无感型性能因素的变化趋势也均为下行趋势或无变化趋势,则认为该均衡型性能因素的变化趋势为下行趋势;

[0055] 如果均衡型性能因素不是上行趋势或下行趋势则为无变化趋势。

[0056] 进一步地,对于均衡型,变化趋势为无变化的性能因素不对其资源进行调节。

[0057] 进一步地,对于均衡型,变化趋势为上行的性能因素对应资源的分配量调整方式为:

$$[0058] \quad E_+(t+1) = C(t) + \frac{\left| \bar{A} - R_{E_+} \right|}{2} E(t)$$

[0059] 其中, $E_+(t+1)$ 为均衡型,变化趋势为上行的性能因素当前资源分配量, $E(t)$ 为该性能因素上一次的某类资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_{E_+} 为均衡型性能因素在上行趋势的经验阈值。

[0060] 进一步地,对于均衡型,变化趋势为下行的性能因素对应资源的分配量调整方式为:

$$[0061] \quad E_-(t+1) = C(t) - \frac{\left| \bar{A} - R_{E_-} \right|}{2} E(t)$$

[0062] 其中, $E_-(t+1)$ 为均衡型,变化趋势为下行的性能因素当前资源分配量, $E(t)$ 为该

性能因素上一次的资源分配量, \bar{A} 表示该性能因素采集数据的平均值, R_E 为均衡型性能因素在下行趋势的经验阈值。

[0063] 进一步地,对于均衡型,变化趋势为上行或下行的性能因素,增加数据的采集频率。

[0064] 进一步地,对于每种服务的所有性能因素添加标签,标示各个性能因素的类型和变化趋势。

[0065] 本发明的上述技术方案具有如下有益的技术效果:

[0066] (1) 本发明实现了信息系统运行态服务资源的精准自动化分配,无需依赖操作人员的经验,在满足服务需求的同时,提高了资源利用效率。

[0067] (2) 本发明通过对服务动态特征的识别,实现了对不同特征服务对资源的差异性分配需求;通过对服务运行参数变化率的预测,实现了对服务各类资源的更精细化的分配能力。

[0068] (3) 本发明提升了对分布式部署服务的资源利用效能,实现对服务部署资源的精细化管控能力。

附图说明

[0069] 图1是运行态服务资源分配流程示意图。

具体实施方式

[0070] 为使本发明的目的、技术方案和优点更加清楚明了,下面结合具体实施方式并参照附图,对本发明进一步详细说明。应该理解,这些描述只是示例性的,而并非要限制本发明的范围。此外,在以下说明中,省略了对公知结构和技术的描述,以避免不必要地混淆本发明的概念。

[0071] 提供一种信息系统运行态服务资源分配方法,结合1,具体步骤包括:

[0072] (1) 采集并存储服务的运行性能信息

[0073] 服务的运行性能由服务在运行态中的资源分配情况与外部并行请求数的匹配程度决定。因此定义表征服务运行性能因素分别为:CPU占用率、内存使用率、磁盘使用率、网络流量占用率、并发访问占用率等。

[0074] 在信息系统运行过程,根据预先设置的采集频率、采集的性能因素种类和服务的部署地址,持续采集服务的性能因素数据。将实时采集的服务性能因素数据存储在数据库中,存储的内容包括:采集时刻、采集值。

[0075] (2) 判断服务性能因素的类型

[0076] 依据不同服务在运行态中,对性能因素的不同敏感度,将服务的性能因素分为:敏感型、无感型和均衡型。同一服务的不同性能因素具有不同的类型,例如:服务A在某运行态的CPU占用率一直比较高,同时内存使用率一直较低,而磁盘使用率在正常范围内,则称服务A为CPU占用敏感型,内存使用无感型,磁盘使用均衡型。

[0077] 判断为敏感型的性能因素,说明该性能因素在当前运行态下的资源紧张,需要增加相应的资源分配,进行扩容;判断为无感型的性能因素,说明该性能因素在当前运行态下的资源空闲,需要减少相应的资源分配,进行缩容;判断为均衡型的性能因素,说明该性

能因素在当前运行态下的资源占用合理，不需要调整相应的资源分配。

[0078] 对于划分粒度较大的服务，由于具备多种功能，随着业务场景的变化其性能因素类型可能发生变化。

[0079] 判断服务性能因素类型的具体过程如下：

[0080] 首先，计算数据库中存储的某一采集周期内的CPU占用率、内存使用率、磁盘使用率、出流量、入流量、并发访问数等服务性能因素历史数据集的平均值。计算公式如下：

$$[0081] \quad \bar{A} = \frac{\sum_{i=0}^n C_i}{n}$$

[0082] 式中， \bar{A} 表示服务某种性能因素采集周期内数据的平均值，n表示采集周期内的采集数量， C_i 表示某种性能因素的第i个采集值。

[0083] 然后，将上一步计算得到的服务各性能因素历史平均值与上限阈值及下限阈值进行比较，得到性能因素的类型。计算公式如下：

$$[0084] \quad T = \begin{cases} T_+ , \bar{A} > R_{max} \\ T_- , \bar{A} < R_{mix} \\ T_0 , R_{mix} < \bar{A} < R_{max} \end{cases}$$

[0085] 式中，T表示服务某性能因素类型， T_+ 表示敏感性， T_- 表示无感型， T_0 表示均衡型， R_{max} 表示该服务性能因素均值的上限阈值， R_{min} 表示该服务性能因素均值的下限阈值。不同性能因素的 R_{max} 和 R_{min} 可通过用户经验进行设置， R_{max} 的建议取值范围为60%-80%， R_{min} 的建议取值范围为10%-20%。

[0086] 进一步地，将服务性能因素类型的判断结果以服务的性能因素类型标签进行存储。服务性能因素类型标签由服务的所有性能因素的类型组成，例如：CPU占用类型标签、内存使用类型标签、磁盘使用类型标签等。所有服务性能因素类型标签的取值包括：敏感型、无感型和均衡型三种。所有性能因素类型标签作为服务的元数据存储于数据库表中。

[0087] (3) 判断服务动态特征

[0088] 通过对服务不同性能因素的相关性分析，判断服务的动态特征。服务的动态特征包括：因素正相关、因素负相关和因素无相关。对于服务S，如果性能因素y的采集值随性能因素x的值采集增加而增加，则称服务S的性能因素x与性能因素y正相关；如果性能因素y的采集值随性能因素x的采集值增加而降低，则称服务S的性能因素x与性能因素y负相关；如果性能因素x和性能因素y之间没有相关性，则称服务S的性能因素x与性能因素y无相关。

[0089] 利用斯皮尔曼秩关系因子定义服务 s_i 的性能因素x和y在采集周期时间窗j内的关系值。即：

$$[0090] \quad c_{i,j}(x,y) = \frac{\sum_{k=1}^w (m'_{i,x,(j-1) \times w + k} - \overline{m'_{i,x,j}})(m'_{i,y,(j-1) \times w + k} - \overline{m'_{i,y,j}})}{\sqrt{\sum_{k=1}^w (m'_{i,x,(j-1) \times w + k} - \overline{m'_{i,x,j}})^2 \sum_{k=1}^w (m'_{i,y,(j-1) \times w + k} - \overline{m'_{i,y,j}})^2}}$$

[0091] 式中， $m_{i,x,k}$ 和 $m_{i,y,k}$ 分别表示服务 s_i 通过步骤(1)在采集周期时间窗j内采集的性能因素x和y的第k个数据；w表示在采集周期时间窗j内一共采集了w个性能因素数据；将一

个时间窗 j 中采集的服务 s_i 的性能因素 x 和 y 的所有采集数据 $m_{i,x,k}$ 和 $m_{i,y,k}$ 按从大到小的顺序排序后, 得到排序序列 $V'_{i,x,j}$ 和 $V'_{i,y,j}$; $m'_{i,x,(j-1) \times w+k}$ 和 $m'_{i,y,(j-1) \times w+k}$ 分别为性能因素 x 的采集值 $m_{i,x,(j-1) \times w+k}$ 和性能因素 y 的采集值 $m_{i,y,(j-1) \times w+k}$ 在排序序列 $V'_{i,x,j}$ 和 $V'_{i,y,j}$ 中的秩次; $\overline{m'_{i,x,j}}$ 和 $\overline{m'_{i,y,j}}$ 分别为排序序列 $V'_{i,x,j}$ 和 $V'_{i,y,j}$ 的平均秩次。

[0092] 当排序序列 $V'_{i,x,j}$ 和 $V'_{i,y,j}$ 中不存在相同秩次的参数值时, 上式还可进一步简化为:

$$[0093] \quad c_{i,j}(x,y) = 1 - \frac{6 \sum_{k=1}^w d_{i,(j-1) \times w+k}^2}{w(w^2 - 1)}$$

[0094] 其中, $d_{i,(j-1) \times w+k}$ 表示在排序序列 $V'_{i,x,j}$ 和 $V'_{i,y,j}$ 中参数值的秩次之差, 即:

$$[0095] \quad d_{i,(j-1) \times w+k} = m'_{i,x,(j-1) \times w+k} - m'_{i,y,(j-1) \times w+k}$$

[0096] 在每个采集周期的时间窗内, 对任意两个性能因素采集序列进行排序, 然后利用斯皮尔曼秩关系因子计算任意两个性能因素值之间的关系。当性能因素关系值趋近于 -1 或 +1 时, 表明这两种性能因素之间有强关系; 当性能因素关系值趋近于 0 时, 表明这两种性能因素之间关系较弱。因此, 对于任意服务 s_i , 在采集周期时间窗 j 内采集了 1 个性能因素, 其任意两个性能因素的关系值构成 1×1 的关系值矩阵, 即:

$$[0097] \quad C_{i,j} = \begin{pmatrix} c_{i,j}(1,1) & c_{i,j}(1,2) & \cdots & c_{i,j}(1,l) \\ c_{i,j}(2,1) & c_{i,j}(2,2) & \cdots & c_{i,j}(2,l) \\ \vdots & \vdots & \ddots & \vdots \\ c_{i,j}(l,1) & c_{i,j}(l,2) & \cdots & c_{i,j}(l,l) \end{pmatrix}$$

[0098] 上式表示一个对称矩阵, 即 $c_{i,j}(x,y) = c_{i,j}(y,x)$ 。并且对于任意参数 x , 其都与自身完全相关, 即 $c_{i,j}(x,x) = 1$ 。

[0099] 服务动态特征计算公式如下:

$$[0100] \quad w_{i,j}(x,y) = \begin{cases} W_+, & c_{i,j}(x,y) > +L_{\max} \\ W_0, & |c_{i,j}(x,y)| < L_{\min} \\ W_-, & c_{i,j}(x,y) < -L_{\max} \end{cases}$$

[0101] 上式中, $w_{i,j}(x,y)$ 表示服务 s_i 在采集周期时间窗 j 内采集的性能因素 x 和 y 的动态特征, W_+ 表示性能因素 x 和 y 正相关, W_0 表示性能因素 x 和 y 无相关, W_- 表示性能因素 x 和 y 负相关, L_{\max} 表示服务特征判断上限阈值, L_{\min} 表示服务特征判断下限阈值。 L_{\max} 和 L_{\min} 可通过用户经验进行设置, L_{\max} 的建议取值范围为 $0.8 \sim 1$, L_{\min} 的建议取值范围为 $0 \sim 0.2$ 。

[0102] 将 $w_{i,j}$ 代入矩阵 $c_{i,j}$, 得到服务的动态特征关系矩阵 $W_{i,j}$, 即:

$$[0103] \quad W_{i,j} = \begin{pmatrix} w_{i,j}(1,1) & w_{i,j}(1,2) & \cdots & w_{i,j}(1,l) \\ w_{i,j}(2,1) & w_{i,j}(2,2) & \cdots & w_{i,j}(2,l) \\ \vdots & \vdots & \ddots & \vdots \\ w_{i,j}(l,1) & w_{i,j}(l,2) & \cdots & w_{i,j}(l,l) \end{pmatrix}$$

[0104] 最后, 将服务动态特征的判断结果以服务的动态特征标签进行存储。服务动态特征标签由所有服务性能因素两两之间的动态关系组成, 例如: CPU 占用与内存使用动态关

系,CPU占用与磁盘使用动态关系等。所有服务动态特征标签的取值包括:正相关、负相关和无相关三种。所有动态特征标签作为服务的元数据存储于数据库表中。

[0105] (4) 预测服务性能因素变化趋势

[0106] 当前应用场景下服务性能因素运行状态的变化趋势包括:上行趋势、下行趋势、无变化趋势。

[0107] 通过对数据库中存储的某一段时间内的CPU占用率、内存使用率、磁盘使用率、网络出/入流量、并发访问数等服务性能因素历史数据集的曲线拟合,预测服务各性能因素在未来一定时间内的大小。曲线拟合方法计算量小,适用于实时的在线数据预测。

[0108] 通过对未来一定时间内预测参数值的变化率计算,得到当前应用场景下服务某项性能因素的变化趋势预测。当预测参数值变化率为正,则该服务性能因素处于上行趋势;当预测参数变化率为负,则该服务性能因素处于下行趋势。变化趋势的计算公式如下:

$$[0109] \quad L = \begin{cases} L_+, \varphi > +|\sigma| \\ L_0, |\varphi| > |\sigma| \\ L_-, \varphi < -|\sigma| \end{cases}$$

[0110] 上式中,L表示服务性能因素的变化趋势; L_+ 表示上行趋势; L_0 表示无变化趋势; L_- 表示下行趋势; σ 为经验阈值,建议取值为1; φ 表示通过对数据库中存储的某一段时间内的CPU占用率、内存使用率、磁盘使用率、网络出/入流量、并发访问数等服务性能因素历史数据集的拟合曲线在预测时间内的斜率值,计算公式如下:

$$[0111] \quad \varphi = F'(t), t \in (t_1, t_2)$$

[0112] 上式中,F(t)表示拟合曲线, t_1 表示预测起始时间, t_2 表示预测的结束时间。

[0113] 在该步骤中,仅计算步骤(2)判断为敏感/无感类型的性能因素。针对敏感/无感型性能因素在判断其运行状态变化趋势时,由其自身变化趋势决定,忽略因性能因素间的正/负相关性所带来的影响。

[0114] 而对于步骤(2)判断为均衡型的性能因素,若通过步骤(3)判断与敏感和无感类型性能因素具有因素正相关或因素负相关,则可通过动态特征的正或负相关性推导出其变化趋势。具体推导过程如下:

[0115] (a) 所有与均衡型性能因素呈正相关的敏感型和无感型性能因素的变化趋势均为上行趋势,并且所有与均衡型性能因素呈负相关的敏感型和无感型性能因素的变化趋势也均为下行趋势或无变化趋势,则认为该均衡型性能因素的变化趋势为上行趋势;

[0116] (b) 所有与均衡型性能因素呈负相关的敏感型和无感型性能因素的变化趋势均为下行趋势,并且所有与均衡型性能因素呈正相关的敏感型和无感型性能因素的变化趋势均为上行趋势或无变化趋势,则认为该均衡型性能因素的变化趋势为上行趋势;

[0117] (c) 所有与均衡型性能因素呈正相关的敏感型和无感型性能因素的变化趋势均为下行趋势,并且所有与均衡型性能因素呈负相关的敏感型和无感型性能因素的变化趋势均为上行趋势或无变化趋势,则认为该均衡型性能因素的变化趋势为下行趋势;

[0118] (d) 所有与均衡型性能因素呈负相关的敏感型和无感型性能因素的变化趋势均为上行趋势,并且所有与均衡型性能因素呈正相关的敏感型/无感型性能因素的变化趋势也均为下行趋势或无变化趋势,则认为该均衡型性能因素的变化趋势为下行趋势;

[0119] (e) 所有与均衡型性能因素呈正或负相关的敏感型和无感型性能因素的变化趋势均为无变化,则认为该均衡型性能因素的变化趋势为无变化趋势;

[0120] (f) 对于其他情况的均衡型性能因素,或者因为与其他性能因素的正负相关性复杂,同时存在与其他性能因素相关性而带来的上行、下行趋势,无法通过服务的动态特性判断其变化趋势;或者不存在与之正相关或负相关的敏感型和无感型性能因素。因此可认为其无变化趋势,从而减少预测计算的整体工作量。

[0121] 进一步地,将服务的所有性能因素运行趋势以服务性能因素运行趋势标签进行存储。服务性能因素运行趋势标签由服务的所有性能因素的运行趋势组成,例如:CPU占用运行趋势标签、内存使用运行趋势标签、磁盘使用运行趋势标签等。所有服务性能因素运行趋势标签的取值包括:上行趋势、下行趋势和无变化趋势三种。所有服务性能因素运行趋势标签作为服务的元数据存储在数据库表中。

[0122] 进一步地,增加均衡型变化趋势为上行或下行的性能因素的采集频率

[0123] 通过步骤(2)判断某服务存在敏感/无感型的性能因素,若根据步骤(3)判断的服务动态特征中存在与该敏感/无感型性能因素正/负相关的均衡型性能因素,且根据步骤(4)预测到该均衡型性能因素具有上行或下行趋势,则认为该均衡型性能因素可能转变为敏感/无感型性能因素,应增加对该均衡型性能因素在采集周期中的采集数量,增强对该均衡型性能因素的监控力度。

[0124] (5) 调整资源的分配量

[0125] 根据服务性能因素的类型、动态特征以及变化趋势,生成对服务部署资源的精确分配方案。形成以下分配准则:

[0126] (a) 对敏感类型的性能因素资源分配方案,如果其变化趋势是无变化,应当按基础调节量对资源进行扩容,其扩容量计算公式如下:

$$[0127] \quad C_0(t+1) = C(t) + |\bar{A} - R_c| \cdot C(t)$$

[0128] 其中, $C_0(t+1)$ 为当前调整的某类性能因素资源分配量, $C(t)$ 为该类型性能因素上一次的资源分配量, \bar{A} 表示该性能因素历史采集数据集的平均值, R_c 为敏感型性能因素的经验阈值,建议取值范围为60%~80%。

[0129] (b) 对敏感类型的性能因素资源分配方案,如果其变化趋势是上行,应当增大调节量对资源进行扩容,其扩容量计算公式如下:

$$[0130] \quad C_+(t+1) = C(t) + |\bar{A} - R_c| \cdot \left(1 + \frac{x'(t+1) - x(t)}{x(t)}\right) C(t)$$

[0131] 其中, $C_+(t+1)$ 为当前调整的某类性能因素资源分配量, $C(t)$ 为该类型性能因素上一次的资源分配量, \bar{A} 表示该性能因素历史采集数据集的平均值, R_c 为敏感型性能因素的经验阈值,建议取值范围为60%~80%, $x'(t+1)$ 为预测的t+1时刻的性能因素数据, $x(t)$ 为当前采集的性能因素数据。因为性能因素当前的运行趋势为上行,因此 $x'(t+1) > x(t)$, $C_+(t+1) > C_0(t+1)$ 。 $C_0(t+1)$ 为变化趋势为无变化的敏感类型性能因素当前调整的分配方案。

[0132] 例如,服务A和B,各有一个类型相同的敏感型性能因素 x_A 和 x_B ,并且 x_A 和 x_B 的历史采集数据集的平均值也相同。若 x_A 当前的变化趋势是无变化, x_B 当前的变化趋势是上行,则

x_B 的资源调节量将大于 x_A 的资源调节量。

[0133] (c) 对敏感类型的性能因素资源分配方案, 如果其变化趋势是下行, 应当减小调节量对资源进行扩容, 其扩容量计算公式如下:

$$[0134] \quad C_-(t+1) = C(t) + \left| \bar{A} - R_c \right| \cdot \left(1 + \frac{x'(t+1) - x(t)}{x(t)} \right) C(t)$$

[0135] 其中, $C_-(t+1)$ 为当前调整的某类性能因素资源分配量, $C(t)$ 为该类型性能因素上一次的资源分配量, \bar{A} 表示该性能因素历史采集数据集的平均值, R_c 为敏感型性能因素的经验阈值, 建议取值范围为60%~80%, $x'(t+1)$ 为预测的 $t+1$ 时刻的性能因素数据, $x(t)$ 为当前采集的性能因素数据。因为性能因素当前的运行趋势为上行, 因此 $x'(t+1) < x(t)$, $C_-(t+1) < C_0(t+1)$ 。 $C_0(t+1)$ 为变化趋势为无变化的敏感类型性能因素当前调整的分配方案。

[0136] 例如, 服务A和B, 各有一个类型相同的敏感型性能因素 x_A 和 x_B , 并且 x_A 和 x_B 的历史采集数据集的平均值也相同。若 x_A 当前的变化趋势是无变化, x_B 当前的变化趋势是下行, 则 x_B 的资源调节量将小于 x_A 的资源调节量。

[0137] (d) 对无感类型的性能因素资源分配方案, 如果其变化趋势是无变化, 应当按基础调节量对资源进行缩容, 其缩容量计算公式如下:

$$[0138] \quad D_0(t+1) = D(t) - \left| \bar{A} - R_D \right| \cdot D(t)$$

[0139] 其中, $D_0(t+1)$ 为当前调整的某类性能因素资源分配量, $D(t)$ 为该类型性能因素上一次的资源分配量, \bar{A} 表示该性能因素历史采集数据集的平均值, R_D 为无感型性能因素的经验阈值, 建议取值范围为10%~20%。

[0140] (e) 对无感类型的性能因素资源分配方案, 如果其变化趋势是下行, 应当增大调节量对资源进行缩容, 其缩容量计算公式如下:

$$[0141] \quad D_-(t+1) = D(t) - \left| \bar{A} - R_D \right| \cdot \left(1 + \frac{x(t) - x'(t+1)}{x(t)} \right) D(t)$$

[0142] 其中, $D_-(t+1)$ 为当前调整的某类性能因素资源分配量, $D(t)$ 为该类型性能因素上一次的资源分配量, \bar{A} 表示该性能因素历史采集数据集的平均值, R_D 为无感型性能因素的经验阈值, 建议取值范围为10%~20%。因为性能因素当前的运行趋势为下行, 因此 $x(t) > x'(t+1)$, $D_-(t+1) < D_0(t+1)$ 。 $D_0(t+1)$ 为变化趋势为无变化的无感类型性能因素当前调整的分配方案。

[0143] 例如, 服务A和B, 各有一个类型相同的无感型性能因素 x_A 和 x_B , 并且 x_A 和 x_B 的历史采集数据集的平均值也相同。若 x_A 当前的变化趋势是无变化, x_B 当前的变化趋势是下行, 则 x_B 的资源调节量将大于 x_A 的资源调节量。

[0144] (f) 对无感类型的性能因素资源分配方案, 如果其变化趋势是上行, 应当减小调节量对资源进行缩容, 其缩容量计算公式如下:

$$[0145] \quad D_+(t+1) = D(t) - \left| \bar{A} - R_D \right| \cdot \left(1 + \frac{x(t) - x'(t+1)}{x(t)} \right) D(t)$$

[0146] 其中, $D_+(t+1)$ 为当前调整的某类性能因素资源分配量, $D(t)$ 为该类型性能因素

上一次的资源分配量, \bar{A} 表示该性能因素历史采集数据集的平均值, R_D 为无感型性能因素的经验阈值, 建议取值范围为10%~20%。因为性能因素当前的运行趋势为下行, 因此 $x'(t+1) > x(t)$, $D_+(t+1) < D_0(t+1)$ 。 $D_0(t+1)$ 为变化趋势为无变化的无感类型性能因素当前调整的分配方案。

[0147] 例如, 服务A和B, 各有一个类型相同的无感型性能因素 x_A 和 x_B , 并且 x_A 和 x_B 的历史采集数据集的平均值也相同。若 x_A 当前的变化趋势是无变化, x_B 当前的变化趋势是上行, 则 x_B 的资源调节量将小于 x_A 的资源调节量。

[0148] (g) 对均衡类型的性能因素资源分配方案, 如果其变化趋势是无变化, 不对其资源进行调节。

[0149] (h) 对均衡类型的性能因素资源分配方案, 如果其变化趋势是上行, 应当适度增大调节量对资源进行扩容, 其扩容量计算公式如下:

$$[0150] \quad E_+(t+1) = C(t) + \frac{|\bar{A} - R_{E+}|}{2} E(t)$$

[0151] 其中, $E_+(t+1)$ 为当前调整的某类性能因素资源分配量, $E(t)$ 为该类型性能因素上一次的某类资源分配量, \bar{A} 表示该性能因素历史采集数据集的平均值, R_{E+} 为均衡型性能因素在上行趋势的经验阈值, 建议取值范围为60%~80%。

[0152] (i) 对均衡类型的性能因素资源分配方案, 如果其变化趋势是下行, 应当适度减少调节量对资源进行扩容, 其缩容量计算公式如下:

$$[0153] \quad E_-(t+1) = C(t) - \frac{|\bar{A} - R_{E-}|}{2} E(t)$$

[0154] 其中, $E_-(t+1)$ 为当前调整的某类性能因素资源分配量, $E(t)$ 为该类型性能因素上一次的资源分配量, \bar{A} 表示该性能因素历史采集数据集的平均值, R_{E-} 为均衡型性能因素在下行趋势的经验阈值, 建议取值范围为10%~20%。

[0155] 综上所述, 本发明公开一种信息系统运行态服务资源分配方法, 采集并存储服务的运行状态信息; 基于各运行状态信息, 判断该运行状态信息对应性能因素的类型; 判断服务各个性能因素之间的相关性, 作为服务动态特征; 基于一段时间内的运行状态信息, 预测服务性能因素变化趋势; 对于基于敏感型、无感型的性能因素基于变化趋势调整对应资源的分配量; 对于均衡型的性能因素基于与其相关的性能因素的变化趋势, 判断均衡型的性能因素的变化趋势, 调整对应资源的分配量。本发明实现了信息系统运行态服务资源的精准自动化分配, 无需依赖操作人员的经验, 在满足服务需求的同时, 提高了资源利用效率。

[0156] 应当理解的是, 本发明的上述具体实施方式仅仅用于示例性说明或解释本发明的原理, 而不构成对本发明的限制。因此, 在不偏离本发明的精神和范围的情况下所做的任何修改、等同替换、改进等, 均应包含在本发明的保护范围之内。此外, 本发明所附权利要求旨在涵盖落入所附权利要求范围和边界、或者这种范围和边界的等同形式内的全部变化和修改例。

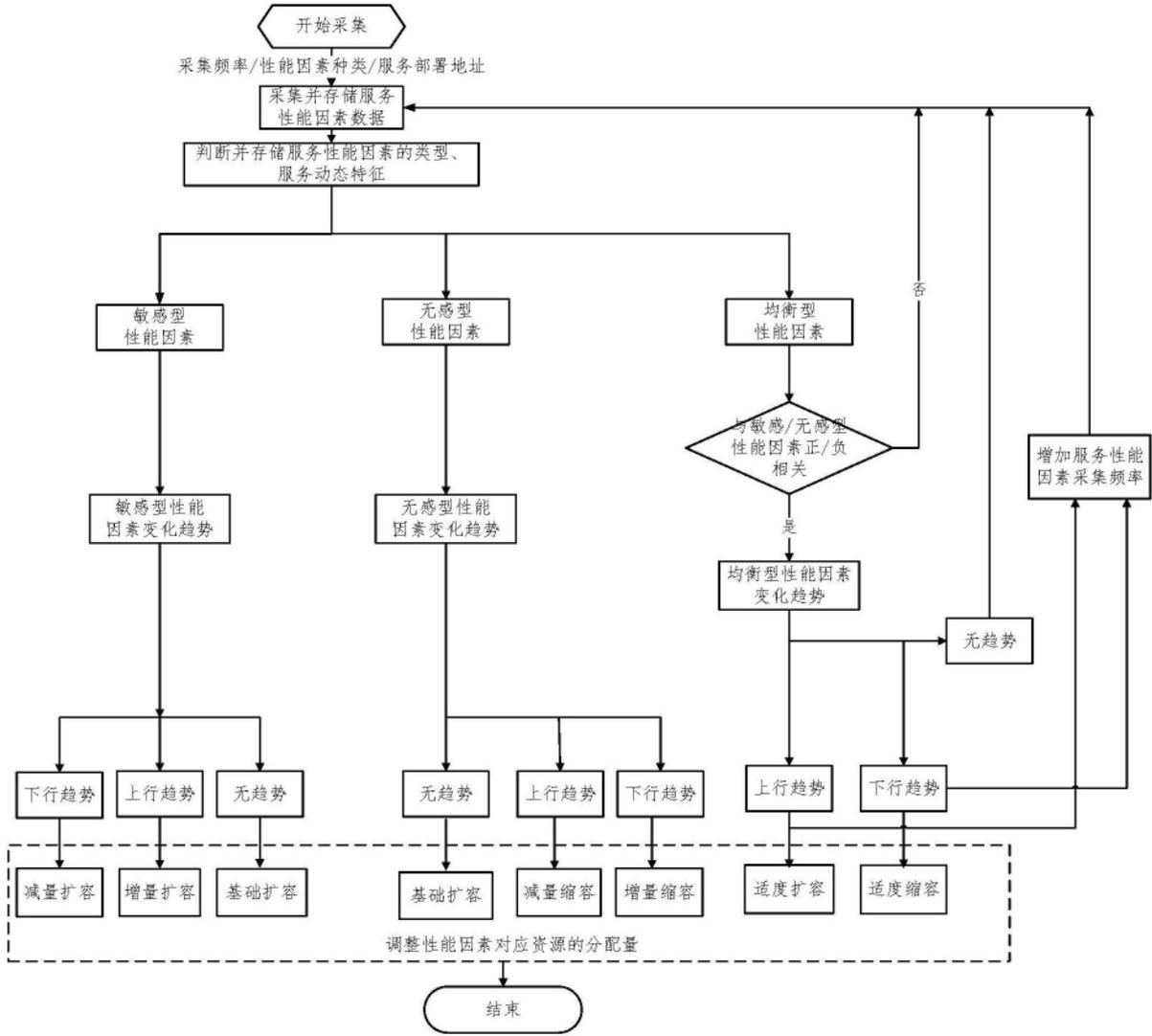


图1