



(12)发明专利申请

(10)申请公布号 CN 107527121 A

(43)申请公布日 2017. 12. 29

(21)申请号 201710838665.3

(22)申请日 2017.09.18

(71)申请人 云南电网有限责任公司信息中心  
地址 650214 云南省昆明市云大西路云电科技园信息中心

(72)发明人 张榆 张雪坚 吕垚 向华伟

(74)专利代理机构 昆明大百科专利事务所  
53106

代理人 何健 李红

(51) Int. Cl.

G06Q 10/04(2012.01)

G06Q 10/06(2012.01)

G06Q 50/06(2012.01)

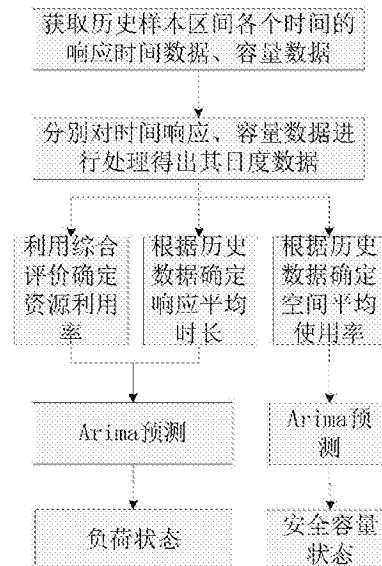
权利要求书1页 说明书9页 附图2页

(54)发明名称

一种电网的信息系统运行状态诊断预测的方法

(57)摘要

本发明提供了一种电网的信息系统运行状态诊断预测的方法,包括两部分内容:系统负载预测和安全容量预测。本发明以电网信息系统为研究对象,重点关注服务器、刀箱等设备的运行状态,通过这些历史数据与信息系统负载表现情况及信息系统所依赖的存储、数据库的容量进行大数据关联分析,寻找影响信息系统高负载的重要因素,建立信息系统高负载及安全容量预测模型,通过预测模型及时发现处于高负载运行的信息系统,从而及时采取资源的合理配置、硬件升级、软件功能优化等手段降低系统负载,防止系统故障。



1. 一种电网的信息系统运行状态诊断预测的方法,其特征在于,所述方法包括以下步骤:

(1) 系统负载的预测

- 1) 获取历史样本区间各个时间的系统资源指标、系统响应指标;
- 2) 分别对系统资源指标、系统响应指标数据进行处理得出其日度数据,建立指标体系;
- 3) 利用权值评价方法对系统资源相关指标进行权重确定并确定综合指标资源利用率;
- 4) 分别对资源利用率和系统响应时长采用arima算法进行预测;
- 5) 根据相关准则判读预测值的状态是否正常;

(2) 安全容量的预测

- 1) 获取历史样本区间各个时间的服务器指标、数据库指标;
- 2) 分别对服务器数据、数据库指标数据进行处理得出其日度数据,建立指标体系;
- 3) 利用arima算法对服务器指标和数据库指标进行指标预测;
- 4) 根据阈值判断各指标预警等级。

2. 根据权利要求1所述的一种电网的信息系统运行状态诊断预测的方法,其特征在于,针对系统负载预测和安全容量预测中步骤1) 相关数据的获取为5分钟一次。

3. 根据权利要求1所述的一种电网的信息系统运行状态诊断预测的方法,其特征在于,针对系统负载预测中步骤2) 中日度数据的获取方法为:CPU使用率=当日每次采集CPU使用率加和/288,磁盘使用率=当日每次采集的磁盘使用率加和/288,系统平均响应时长=当日每次采集响应时长加和/288;针对安全容量预测中步骤2) 中日度数据的获取方法为:磁盘已使用空间=当日每次采集的磁盘使用空间/288,文件系统已使用空间=当日每次采集的文件系统使用空间/288,表空间=当日每次采集的表使用空间/288。

4. 根据权利要求1所述的一种电网的信息系统运行状态诊断预测的方法,其特征在于,系统负荷预测针对步骤3) 中的综合评价法,采用熵值法对指标体系进行权重确定。

## 一种电网的信息系统运行状态诊断预测的方法

### 技术领域

[0001] 本发明属于电网信息系统运行状态预测技术领域,具体是涉及电网系统负荷和安全容量预测的方法。

### 背景技术

[0002] 随着云南电网公司信息化进程的发展,信息系统在日常运行时,会对底层软硬件造成负荷。IT监控系统每分钟要从这些信息系统采集上万个系统运行数据,这些数据包含了与信息系统运行相关的动环、网络、主机、存储、中间件、应用、运维等大量信息,任何一种资源负载过大,都可能会引起系统性能下降甚至瘫痪。针对目前存在的问题,本文以云南电网信息系统为研究对象,重点关注服务器、刀箱等设备的运行状态,通过这些历史数据与信息系统负载表现情况及信息系统所依赖的存储、数据库的容量进行大数据关联分析,寻找影响信息系统高负载的重要因素,建立信息系统高负载及安全容量预测模型,通过预测模型及时发现处于高负载运行的信息系统,从而及时采取资源的合理配置、硬件升级、软件功能优化等手段降低系统负载,防止系统故障。以此为基础,通过统计、评价、预测等大数据分析技术探索适合公司的信息系统及网络设备运行状态诊断的方法,解决公司目前被动应对信息系统的状态异常及故障的问题。通过建立基于运行状态的故障诊断模式,创新运维管理方式,不断提升信息系统运行水平,确保系统安全稳定运行。

### 发明内容

[0003] 本发明的目的是基于信息系统负荷和安全容量相关指标,采用数据挖掘的方法建立信息系统运行状态诊断预测模型,以克服当前系统的检修模式都是基于故障触发的事后处理方式所带来的不足,能够提前对信息系统的异常、故障进行预测,实现信息系统的事前维护,变系统故障的被动处理为主动预防。

[0004] 本发明的技术方案如下:

[0005] 一种电网的信息系统运行状态诊断预测的方法,包括以下步骤:

[0006] (1) 系统负载的预测

[0007] 1) 获取历史样本区间各个时间的系统资源指标、系统响应指标;

[0008] 2) 分别对系统资源指标、系统响应指标数据进行处理得出其日度数据,建立指标体系;

[0009] 3) 利用权值评价方法对系统资源相关指标进行权重确定并确定综合指标资源利用率;

[0010] 4) 分别对系统资源利用率和系统响应时长采用arima算法进行预测;

[0011] 5) 根据相关准则判读预测值的状态是否正常。

[0012] (2) 安全容量的预测

[0013] 1) 获取历史样本区间各个时间的服务器指标、数据库指标;

[0014] 2) 分别对服务器数据、数据库指标数据进行处理得出其日度数据,建立指标体系;

[0015] 3) 利用arima算法对服务器指标和数据库指标进行指标预测;

[0016] 4) 根据阈值判断各指标预警等级。

[0017] 所述的一种电网的信息系统运行状态诊断预测的方法,针对系统负载预测和安全容量预测中步骤1) 相关数据的获取为5分钟一次。

[0018] 所述的一种电网的信息系统运行状态诊断预测的方法,针对系统负载预测中步骤2) 中日度数据的获取方法为:CPU使用率=当日每次采集CPU使用率加和/288,磁盘使用率=当日每次采集的磁盘使用率加和/288,系统平均响应时长=当日每次采集响应时长加和/288;针对安全容量预测中步骤2) 中日度数据的获取方法为:磁盘已使用空间=当日每次采集的磁盘使用空间/288,文件系统已使用空间=当日每次采集的文件系统使用空间/288,表空间=当日每次采集的表使用空间/288。

[0019] 所述的一种电网的信息系统运行状态诊断预测的方法,系统负荷预测针对步骤3) 中的综合评价法,采用熵值法对指标体系进行权重确定。

[0020] 本发明通过选取系统相关性能指标建立指标体系,对指标数据采用数据挖掘的方法建立信息系统架构设计评价模型,有效解决如何让数据创造新的价值,提升运维管理水平,保障信息系统稳定运行。同时本发明也能为同类信息系统的建设带来参考,提升公司信息系统的建设管理能力,提高系统应用水平。

[0021] 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作简单地介绍。

#### 附图说明

[0022] 图1是信息系统运行状态诊断预测流程图;

[0023] 图2是系统资源负载状态评价过程图;

[0024] 图3是系统安全容量预测过程图。

#### 具体实施方式

[0025] S1、获取历史样本区间各个时间的系统资源指标、系统响应指标;

[0026] 结合业务分析,本发明系统负荷评价预测选取的指标主要有CPU使用率、内存利用率、磁盘使用率、系统评价响应时长等;安全容量预测选取的指标有磁盘已使用空间、文件系统已使用空间、表空间已使用空间。

[0027] 下表1和表2分别为负荷预测评价模型和安全容量预测模型拟采用的初始指标:

模型	数据分类	指标	计量单位	备注
[0028] 系统 负载 评价 模型	系统资源 指标	CPU使用率	%	
		内存利用率	%	
		磁盘使用率	%	
		网络使用率	%	

模型	数据分类	指标	计量单位	备注
[0029]	系统应用指标	系统平均响应时间	ms	由于该指标是反应给客户系统性能的最直接指标，因此优先选择该指标。

[0030] 表1

[0031] 在正

对象名称	部件名称	指标名称	指标说明	备注
[0032]	服务器	磁盘	磁盘空间使用率	磁盘已使用空间占总空间的比率 正常:W<60% 注意:W>=60%&&W<85% 异常:W>=85%&&W<95% 严重:W>=95%
			磁盘已使用空间	磁盘已使用空间大小
			磁盘剩余空间	磁盘剩余空间大小
	文件系统	文件系统使用率	文件系统已使用空间占总空间的比率 正常:W<50% 注意:W>=50%&W<60% 异常:W>=60%&W<70% 严重:W>=70%	
		文件系统已使用空间	磁盘已使用空间大小	
		文件系统剩余空间	文件系统剩余空间大小	
数据库	表空间	表空间使用率	表空间已使用空间占总空间的比率 正常:W<50% 注意:W>=50%&W<60% 异常:W>=60%&W<70% 严重:W>=70%	
		表空间使用率	表空间已使用空间大小 正常: W<85% 一级: W>=85%&W<90% 二级: W>=90%&W<95% 三级: W>95%	

[0033] 表2

[0034] 正常情况下指标频率时间为每5分钟采取一次。

[0035] S2、分别对系统资源指标、系统响应指标,服务器指标、数据库指标数据进行处理得出其日度数据,建立指标体系;

[0036] 针对系统负载预测中步骤2)中日度数据的获取方法为:CPU使用率=当日每次采集CPU使用率加和/288,磁盘使用率=当日每次采集的磁盘使用率加和/288,系统平均响应时长=当日每次采集响应时长加和/288;针对安全容量预测中步骤2)中日度数据的获取方法为:磁盘已使用空间=当日每次采集的磁盘使用空间/288,文件系统已使用空间=当日每次采集的文件系统使用空间/288,表空间=当日每次采集的表使用空间/288,空间使用率=已使用空间/总容量。

[0037] 以云南电网综合管理系统为例,采取并计算获得负荷预测和容量预测相关数据的日度数据,历史数据样本区间定位2017年5月20日至2017年6月20日。

[0038] S3、利用权值评价方法对系统资源相关指标进行权重确定并确定综合指标资源利用率。

[0039] S31、服务单元资源利用率评估

[0040] 将虚拟机CPU使用率、内存利用率、磁盘使用率作为虚拟机资源使用情况的度量指标,记衡量资源使用情况的指标有p项,单元内有m个虚拟机,虚拟机总数为n,共采集了t个时刻的指标数据。

[0041] (1) 熵值法权值确定

[0042] 样本数据选择。样本数据为当前时刻全部虚拟机的采集指标,样本容量为n,即样本包含n个虚拟机的指标数据,指标数量为p。模型最终输入如下:

$$[0043] \quad X = \{X_1, X_2, \dots, X_p\} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}_{n \times p}$$

[0044]  $X_i = (x_{1i} \ x_{2i} \ \cdots \ x_{ni})$ ,  $i = 1, 2, \dots, p$

[0045] 其中 $X_i$ 为指标i的样本值, $x_{ij}$ 为第i个虚拟机第j个指标的取值。

[0046] 数据的标准化处理。不同指标的数量级和量纲可能会有所不同,需要进行标准化处理以消除数量级和量纲的影响。

[0047] a) 对于正向指标,即越大越好的指标,作如下转换:

$$[0048] \quad x'_{ij} = \frac{x_{ij} - \min(x_{1j}, \dots, x_{nj})}{\max(x_{1j}, \dots, x_{nj}) - \min(x_{1j}, \dots, x_{nj})}, (i = 1, \dots, n; j = 1, \dots, p)$$

[0049] b) 对于负向指标,即越小越好的指标,作如下转换:

$$[0050] \quad x'_{ij} = \frac{\max(x_{1j}, \dots, x_{nj}) - x_{ij}}{\max(x_{1j}, \dots, x_{nj}) - \min(x_{1j}, \dots, x_{nj})}, (i = 1, \dots, n; j = 1, \dots, p)$$

[0051] 为了便于理解起见,记标准化处理后的数据仍为 $x_{ij}$ 。

[0052] 计算第j项指标下第i个虚拟机占该指标的比重:

[0053] 
$$P_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}} \quad (j = 1, 2, \dots, p)$$

[0054] 计算第j项指标的熵值:

[0055] 
$$e_j = -k \sum_{i=1}^n P_{ij} \ln(P_{ij})$$

[0056] 其中k>0, ln为自然对数, e<sub>j</sub>>0, 式中常数k与样本容量n有关, 一般另:

[0057] 
$$k = \frac{1}{\ln(n)}$$

[0058] 计算第j项指标的差异系数g<sub>j</sub>=1-e<sub>j</sub>, 对于第j项指标, 指标值x<sub>ij</sub>的差异越大, 对方案评价的作用越大, 熵值就越小, 则:g<sub>j</sub>越大指标越重要。

[0059] 求各个指标的权数:

[0060] 
$$W_j = \frac{g_j}{\sum_{j=1}^p g_j}, \quad j = 1, 2, \dots, p$$

[0061] 用熵权法计算得到的指标权重向量记为W:

[0062] 
$$W = \{w_1, w_2, \dots, w_p\}$$

[0063] (2) 计算资源利用率

[0064] 记 $l_{ij}^k$ 为第k个虚拟机第i个指标在时间点j的取值, 则虚拟机指标数据集合为:

[0065] 
$$L_k = \{l_{ij}^k : i = 1, \dots, p; j = 1, \dots, t\}$$

[0066] L<sub>k</sub>表示虚拟机k上指标的采集数据集合。下表3为虚拟机指标数据集合。

[0067]

指标	虚拟机 1	虚拟机 2	...	虚拟机 m
CPU 使用率	$l_{1j}^1$	$l_{1j}^2$	...	$l_{1j}^m$
内存利用率	$l_{2j}^1$	$l_{2j}^2$	...	$l_{2j}^m$
磁盘使用率	$l_{3j}^1$	$l_{3j}^2$	...	$l_{3j}^m$
网络使用率	$l_{4j}^1$	$l_{4j}^2$	...	$l_{4j}^m$

[0068] 表3

[0069] 单元的数据集合为单元内所有虚拟机的数据集合, 记为:

[0070] 
$$L = \{L_k : k = 1, \dots, m\}$$

[0071] 不同的任务对资源的要求不同, 同时虚拟机的配置性能也可能不同, 因此引入权重因子, 记δ<sub>k</sub>为虚拟机k的权重, 则虚拟机权重的集合可以表示为:

[0072] 
$$\theta = \{\delta_k : k = 1, \dots, m\}$$

[0073] 用熵权法进行权重的计算, 得到指标权重向量如下:

[0074]  $W = \{w_i : i = 1, \dots, p\}$

[0075] 定义虚拟机k在j时刻的负载,并描述虚拟机实时的负载状态:

[0076] 
$$f_k = \sum_{i=1}^p (w_i \times l_{ij}^k)$$

[0077] 定义虚拟机k在指标i上的负载能力为 $c_i^k$ ,  $c_i^k$ 表示虚拟机k在指标i的取值上限(如:CPU使用率小于75%正常,则该值为75%),则指标i负载能力 $R_i$ 为:

[0078] 
$$R_i = \sum_{k=1}^m (c_i^k \times \delta_k)$$

[0079] 指标i上的负载能力 $c_i^k$ ,可以根据各单位的管理规定或经验给出一个合理的范围[a, b]。

[0080] 单元在时间点j时指标i的负载率为实际负载对负载能力的占比,计算公式为:

[0081] 
$$S_i = \frac{\sum_{k=1}^m (\delta_k \times l_{ij}^k)}{R_i}$$

[0082] 其中 $\sum_{k=1}^m (\delta_k \times l_{ij}^k)$ 表示虚拟机在时刻j对资源i的实际使用率,即实际负载。

[0083] 最终可以得到单元在时刻j对资源的使用率为:

[0084] 
$$S = \sum_{i=1}^p (w_i \times S_i)$$

[0085] 分析计算结果S取值,首先S是一个非负实数,反应的资源使用状况为:

[0086]  $S > 1$ : 过载

[0087]  $S = 1$ : 负载正常

[0088]  $S < 1$ : 空闲

[0089] 实际应用中,资源使用情况是在一定区间内波动,不可能总是达到资源100%利用,因此可以根据业务情况划定区间进行分级,定义业务系统资源正常使用区间为[a, b],其中b为上限,a为下限:

[0090]  $S > b$ : 过载

[0091]  $S \in [a, b]$ : 正常

[0092]  $S < a$ : 空闲

[0093] 下表4为服务单元资源使用率

[0094]

指标	单元1	单元2	...	单元n
资源使用率	$S_1$	$S_2$	...	$S_n$

[0095] 表4

[0096] S32、业务系统资料使用率评估;

[0097] 在该业务分析中,各个单元中的服务器配置相同,因此系统级资源使用率的整体



情况用单元的平均使用率表示。

[0098] 记单元数量为N,单元h的使用率为 $S_h$ ,系统资源使用率记为Rate,则:

[0099] 
$$Rate = \frac{\sum_{h=1}^N S_h}{N}$$

[0100] 所述的一种电网信息系统架构评价的方法,所述的综合评价法为熵值法。以云南电网综合管理系统为例,各指标权重如下表3:

[0101]

指标名	Cpu使用率	内存使用率	磁盘/文件系统使用率
权重值	0.2	0.12	0.68

[0102] 表5

[0103] S4、分别对系统资源利用率和系统响应时长采用arima算法进行预测;

[0104] 对于时间序列指标建立差分自回归滑动平均模型

[0105] 第一步,根据时间序列的散点图、自相关函数和偏自相关函数图以ADF单位根检验其方差、趋势及其季节性变化规律,对序列的平稳性进行识别。

[0106] 第二步,对非平稳序列进行平稳化处理。如果数据序列是非平稳的,并存在一定的增长或下降趋势,则需要对数据进行差分处理,如果数据存在异方差,则需对数据进行技术处理,直到处理后的数据的自相关函数值和偏相关函数值无显著地异于零。

[0107] 第三步,根据时间序列模型的识别规则,建立相应的模型。若平稳序列的偏相关函数是截尾的,而自相关函数是拖尾的,可断定序列适合AR模型;若平稳序列的偏相关函数是拖尾的,而自相关函数是截尾的,则可断定序列适合MA模型;若平稳序列的偏相关函数和自相关函数均是拖尾的,则序列适合ARMA模型。

[0108] 第四步,进行参数估计,检验是否具有统计意义。

[0109] 第五步,进行假设检验,诊断残差序列是否为白噪声。

[0110] 第六步,利用已通过检验的模型进行预测分析。

[0111] 记需要预测指标数为p,指标i历史数据序列为:

[0112]  $X_i = \{x_{t,i}, x_{t-1,i}, \dots, x_{t-n,i}\}, i = 1, 2, \dots, p,$

[0113] 对序列 $X_i$ ,根据ARIMA预测t+1时刻的指标为 $x_{t+1,i}$ ,最终得到t+1时刻指标估计向量为

[0114] 
$$\hat{X}_{t+1} = [\hat{x}_{t+1,1}, \dots, \hat{x}_{t+1,p}]$$

[0115] 其中 $\hat{x}_{t+1,i}$ 为第i个指标的预测值。

[0116] S5、根据相关准则判读预测值的状态是否正常。

[0117] S51、资源利用率

[0118] 记指标X在时刻t的预测值为 $\hat{x}_t$ ,样本标准差为s,通常通过加减三个标准差构造基线,结果为 $[\hat{x}_t - 3s, \hat{x}_t + 3s]$ 。资源利用率状态分级说明如下表6:

[0119]

系统资源利用率	资源利用状态
---------	--------

Rate ∈ [X <sub>t</sub> -3s, X <sub>t</sub> +3s]	正常
Rate > X <sub>t</sub> +3s	过载
Rate < X <sub>t</sub> -3s	空闲

[0120] 表6

[0121] S52、系统响应时间

[0122] 能够表现系统应用状态的核心指标为系统平均响应时间。记平均响应时间的阈值为T,当平均响应时间超过T则认为响应时间超长。系统响应时间分级说明如下表7:

[0123]

系统平均响应时间	状态
大于T (>1500)	服务繁忙
小于T (<=1500)	服务正常

[0124] 表7

[0125] S53、系统负荷评价状态

[0126] 结合资源使用率和系统响应时间,可得最终的业务系统负荷状态,如下表8:

平均响应时间 (状态)	资源利用率 (状态)	业务系统负载评价
超长	过载	系统繁忙
正常	过载	系统繁忙
超长	正常	系统异常
正常	正常	系统正常
超长	空闲	系统异常
正常	空闲	系统空闲

[0127]

[0128] 表8

[0129] S54、空间使用率

[0130] 将预测值与磁盘总量进行比较,获得预测的磁盘使用率。使用率阈值设置及预警级别如下表9:

[0131]

预测已使用空间率	预警等级
85%	I
90%	II
95%	III

[0132] 表9

[0133] 系统符合状态及安全容量状态结果分别如下表10、11:

[0134]

系统名称	日期	状态预测
人力资源管理系统	2017-05-20	系统正常
财务管理系统	2017-05-20	系统正常
资产管理系统	2017-05-20	系统正常

人力资源管理系统	2017-05-20	系统正常
财务管理系统	2017-05-20	系统正常
资产管理系统	2017-05-20	系统正常
人力资源管理系统	2017-05-20	系统正常
人力资源管理系统	2017-05-20	系统正常
财务管理系统	2017-05-20	系统正常
资产管理系统	2017-05-20	系统正常
...	...	...

[0135] 表10

[0136]

日期	预警等级
2017-05-20	正常
2017-05-21	正常
2017-05-22	正常
2017-05-23	正常
2017-05-24	正常
^...	...

[0137] 表11

[0138] 以上所述实施方式仅仅是对本发明的优选实施方式进行了描述,并非对本发明的范围进行限定,在不脱离本发明精神和范围的前提下,本领域普通技术人员对本发明的技术方案作出的各种变形和改进,均应落入本发明的权利要求书确定的保护范围内。

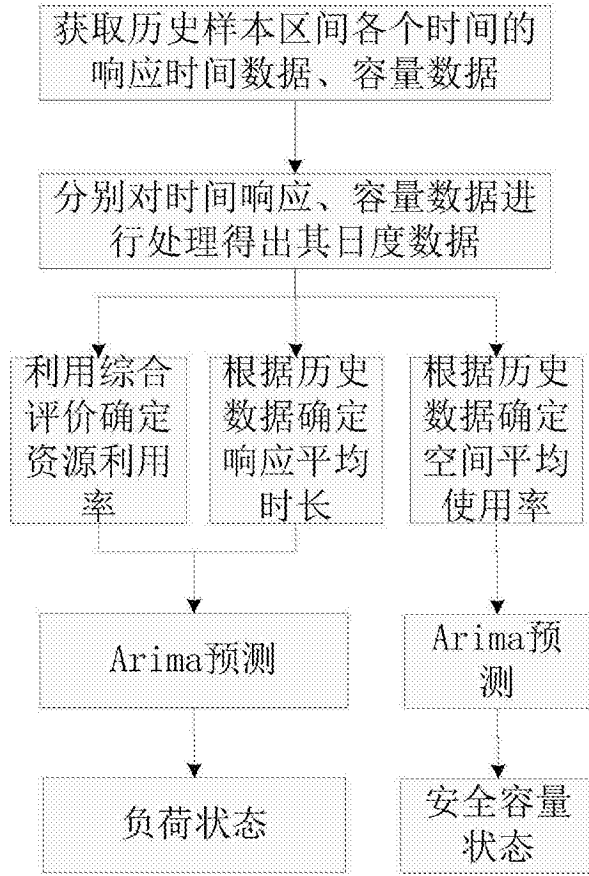


图1

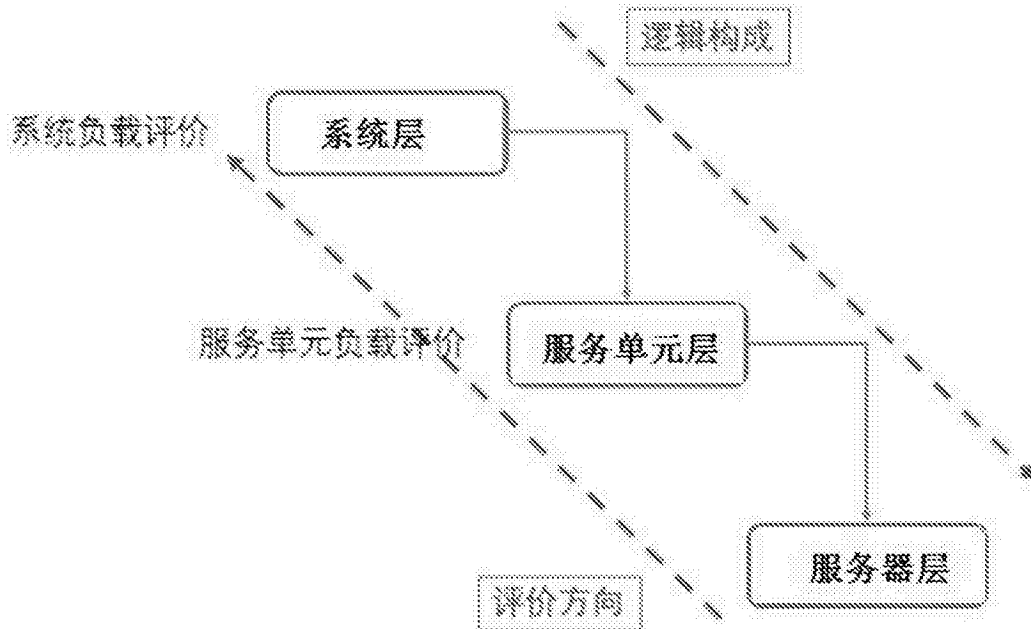


图2

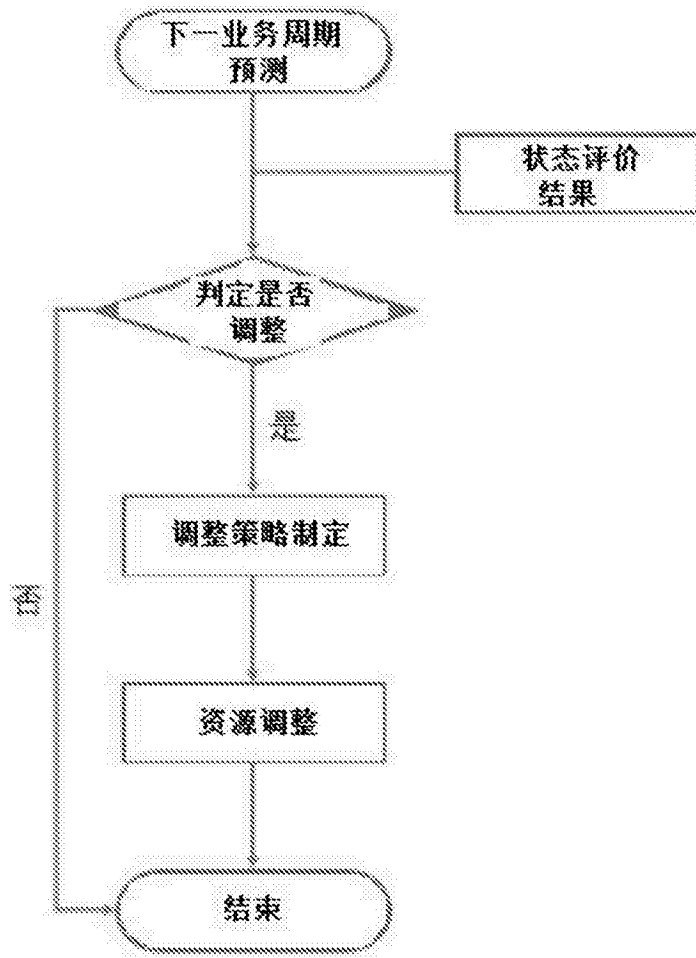


图3