

# 基于大数据的 信息系统运维智能化研究

## Research on Intelligent Information System O& M Based on Big Data

罗 砚(中国移动通信集团重庆有限公司,重庆 401121)  
Luo Yan(China Mobile Group Chongqing Co.,Ltd.,Chongqing 401121,China)

### 摘 要:

介绍了IT系统云化趋势下通过大数据技术挖掘运维数据价值、夯实基础运维、提升运维效率从而实现智能化的运维工作的方法。以性能动态基线为例,通过详细介绍动态基线的算法和实现逻辑,分析了大数据技术如何在运维工作中发挥强大的作用。通过海量数据分析得出数据的运行规律,进而预测运维数据的特点,发现IT系统隐性问题,运维人员根据分析结果主动出击预防故障。

### Abstract:

In the cloud trend, based on big data, IT system can mine O&M data value, consolidate basic O&M, improve O&M efficiency to realize intelligent O&M. It describes performance dynamic baseline algorithm and implementation logic, analyzes how big data plays a strong role in O&M. The operation rules of data is obtained from massive data analysis, to forecast the characteristic of O&M data and find hidden problems of IT system. Based on the above, O&M personnel can take proactive measures to prevent failures.

### Keywords:

Big data; Information system; Intelligent O&M

引用格式:罗砚. 基于大数据的信息系统运维智能化研究[J]. 邮电设计技术,2018(3):79-82.

### 关键词:

大数据;信息系统;智能运维

doi:10.12045/j.issn.1007-3043.2018.03.017

中图分类号:TN914

文献标识码:A

文章编号:1007-3043(2018)03-0079-04

## 0 前言

随着社会信息化水平的不断提升,信息系统在社会生产中的作用日益重要,同时信息系统的整体架构向着云化和分布式不断演进,对信息系统的运维质量提出了更高要求,如何通过主动服务提升运维效率,实现客户感知能力提升,是运维管理系统亟待解决的问题。通过构建大数据分析平台,可以有效利用运维管理系统的既有数据,进行深度分析挖掘,达到提升客户感知、提高运维智能化、加速故障定位、支撑辅助决策的目的,从而保障信息系统的运营服务质量。

## 1 云计算给信息系统运维带来的挑战

云计算技术正在以“能力分散化、管理集中化”的

方式革命性地改变信息系统的整体架构和运营模式。核心的虚拟化技术在信息系统中的推广使用给传统的运维工作带来了众多的挑战。各种基础设施,通用软硬件平台成几何数量增加,传统的监控方式、数据分析方式已经远远不能满足运维要求,主要存在以下问题。

a) 方式被动,传统的监控流程为:发现问题—发现问题—通知运维人员—运维人员解决问题,一旦产生告警,就意味着问题已经发生,不能防范于未然;对处理问题的运维人员的素质要求高,因为问题已经发生,如果不能在短时间内解决问题,就可能会造成严重的后果,这对运维人员的专业能力和心理都是一项严酷的考验。

b) 传统的运维手段在大数据环境下已经无法适应,在海量运维数据面前容易形成运维死角,无法保障运维质量。同时缺乏面向管理人员、业务人员,通过对海量运维数据进行多视角、多维度分析的工具,面对海

收稿日期:2017-12-21

量运维数据,运维人员无法有效开展分析工作。

基于以上问题,迫切需要改变传统的运维方式和手段,采用大数据技术提升运维智能化水平,减轻运维压力,提升运维质量。

## 2 基于大数据技术的智能运维研究

大数据技术是对海量数据进行实时分析和数据价值挖掘的技术。大数据技术在海量运维中的应用场景一直是业界的热点,本文仅以动态基线为例详细介绍大数据技术给运维工作自动化、智能化带来的变革。动态基线是基于大数据分析的一种预测性智能运维手段,与被动方式不同,它不是等待设备出现故障时,才产生告警信息。而是根据海量的历史数据分析出数据的运行规律,预测未来某个时间的正常数据范围(也就是基线),并且根据历史经验设置的容忍度,计算出数据能容忍出现的区间,当实时性能数据超过容忍区间时,就通知运维人员。因此,从严格意义上讲,动态产生的告警是一种预警,是通知运维人员当前的指标性能值已经超过了设定的容忍范围,需要关注和预处理。同时预测的结果也可以作为运维人员指标分析的有效辅助工具,协助运维人员定位问题。

### 2.1 动态基线计算流程

a) 从数据源中提取历史数据,作为基线计算的样本数据。

b) 针对基线规则中的数据粒度,对数据进行平滑处理。

c) 运用排序和概率分布算法对平滑后数据进行计算。

d) 生成上下基线数据,用于预测数据正常波动的区间范围。

图1示出的是动态基线计算流程。

### 2.2 数据平滑处理

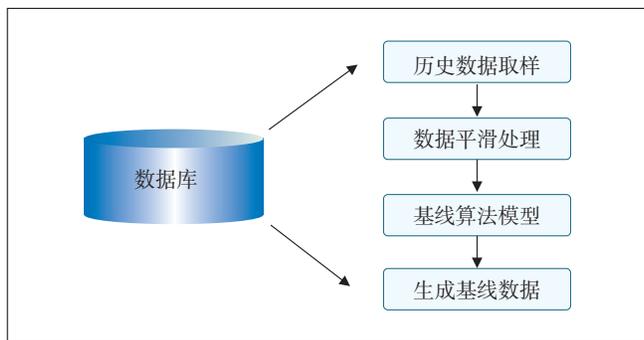


图1 动态基线计算流程示意图

由于数据获得的方式分为主动采集和被动采集,而主动采集又分为整点采集和等间隔采集,这就导致采集上来的时间并不一定是整点数据,而在进行数据计算时,又受到原动态基线获取整点数据计算的限制,所以就需要将采集的历史数据先进性平滑处理,将时间都换算为整点数据,方便动态基线的计算。

### 2.3 使用标准差(均方差)的动态基线概率算法

首先对预处理后的有效数据进行排序,假定共有  $N$  个,分别记为  $X_1 \sim X_N$ 。

假定有效数据的  $Y\%$ (对应基线的样本置信度)为可以接受而不用产生告警的指标值,滑动排序数据的窗口(共  $N \times Y\%$  个数据),计算该窗口中数据的均方差(标准差)。

计算该窗口中样本的平均值作为期望值  $E(X)$ 。

计算窗口各个样本点对于数学期望的偏离程度,单个偏离是  $X - E(X)$ 。为消除符号影响,一般取  $(X - E(X))^2$ 。

求方差,即窗口中所有样本点偏离平方的均值,记为

$$D(X) = E[(X - E(X))^2]$$

求标准差或均方差  $\sqrt{D(X)}$ 。

取均方差最小的窗口中的数据,以其最大值作为基线的上限,最小值作为基线的下限。

## 3 基于概率算法的动态基线计算演示

### 3.1 历史数据取样

下面以计算 2012-04-16T20:00:00 的基线数据为例。基线样本数据空间设定为 30 天,即从数据库中抽取从 2012 年 3 月 17 日至 2012 年 4 月 15 日,连续 30 天同一时间点(20:00)的历史数据作为样本数据(见表 1)。

### 3.2 数据平滑处理

如果连续 30 天的历史数据都存在,那么无需做数据平滑处理。如表 1 所示,2012-04-06T20:00:00 的数据不存在,则需要做数据平滑处理,即取该时间点前后相邻时间的 2 条数据计算出平均值,把平均值作为该时间点的值,如图 2 所示。

平滑后得到 2012-04-06T20:00:00 的数据为 89。

### 3.3 基线数据计算

动态基线数据计算主要分成两大步。

#### 3.3.1 5 区间取数去杂法

表1 历史数据样例

时间	历史性能数据	时间	历史性能数据
2012-03-17T20:00	56	2012-04-01T20:00	78
2012-03-18T20:00	91	2012-04-02T20:00	79
2012-03-19T20:00	79	2012-04-03T20:00	67
2012-03-20T20:00	34	2012-04-04T20:00	66
2012-03-21T20:00	54	2012-04-05T20:00	69
2012-03-22T20:00	51	2012-04-06T20:00	
2012-03-23T20:00	53	2012-04-07T20:00	63
2012-03-24T20:00	63	2012-04-08T20:00	87
2012-03-25T20:00	39	2012-04-09T20:00	81
2012-03-26T20:00	38	2012-04-10T20:00	94
2012-03-27T20:00	24	2012-04-11T20:00	99
2012-03-28T20:00	39	2012-04-12T20:00	73
2012-03-29T20:00	41	2012-04-13T20:00	103
2012-03-30T20:00	41	2012-04-14T20:00	102
2012-03-31T20:00	39	2012-04-15T20:00	79

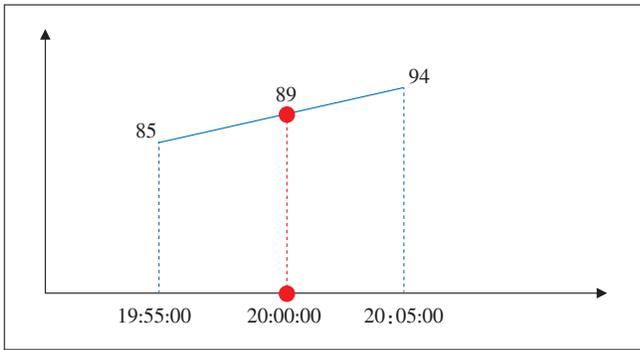


图2 数据平滑计算示意图

经过数据平滑处理后,将得到的样本数据按从小到大排序,为保证所有数据都为非异常情况下产生的,需要剔除掉最小和最大的数,得到数据为:[24,34,38,39,39,39,41,41,51,53,54,56,63,63,66,67,69,73,78,79,79,79,81,87,89,91,94,99,102,103]。

确定5个区间范围,103/5=21,即[0~21],[21~42],[42~63],[63~84],[84~105]。

将剔除异常后剩余的28个数分配到这5个区间中去(区间最小值≤N<区间最大值)。

区间1=[ ]。

区间2=[34,38,39,39,39,41,41]。

区间3=[51,53,54,56]。

区间4=[63,63,66,67,69,73,78,79,79,79,81]。

区间5=[87,89,91,94,99,102]。

区间分配后可以看出,区间4中的数据个数最多,所以取区间4以及上下相邻区间3和区间5的数据,如果上下相邻区间没有数据,可不取,得到21个数据。

### 3.3.2 使用标准差(均方差)的动态基线概率算法

对去杂处理后剩余的21个数据进行排序,得到[51,53,54,56,63,63,66,67,69,73,78,79,79,79,81,87,89,91,94,99,102]。

通过置信度计算滑动窗口大小,即纳入滑动窗口的有效数据的个数。在概率算法中,样本置信度为所选取的历史样本数据的可信程度,推荐使用默认值0.8(即80%的为最接近实际的真实数据,20%为杂质数据)。滑动窗口大小=有效计算数据个数×置信度=21×0.8=16.8取整为16。

滑动排序数据的窗口大小为16,计算各窗口中数据的均方差。

第1窗口为第1个数到第16个数,其均方差为:10.96。

第2窗口为第2个数到第17个数,其均方差为:11.00。

第3窗口为第3个数到第18个数,其均方差为:10.96。

第4窗口为第4个数到第19个数,其均方差为:10.82。

第5窗口为第5个数到第20个数,其均方差为:10.89。

第6窗口为第6个数到第21个数,其均方差为:11.48。

根据计算的滑动窗口的均方差结果如表2所示。

表2 滑动窗口的均方差

滑动窗口范围	窗口均差值
1~16	10.96
2~17	11.00
3~18	10.96
4~19	10.82
5~20	10.89
6~21	11.48

其中均方差最小为第4窗口的10.82,因此取该窗口中的最大值94作为基线的上基线值,最小值56作为基线的下基线值。

得到上基线值为94,下基线值为56之后,为允许实际数据在基线上下一定范围内进行波动(即容忍度),按默认容忍度20%计算,得到:

$$\text{上容忍度} = \text{上基线值} \times (1 + 20\%) = 112.8$$

$$\text{下容忍度} = \text{下基线值} \times (1 - 20\%) = 44.8$$

### 3.4 实际预测效果

图3是某虚拟机的CPU利用率性能趋势,其中蓝线表示实时性能数据,蓝色区域为本文介绍的基线算法计算得出的安全基线带,上容忍线和下容忍线所示区域,作为基线计算的一个补充,用于排除偶尔产生的数据小范围波动。当实际数据超过了安全范围,则可理解为存在异常波动,需要引起运维人员的关注,为运

维人员提供了一种有别于传统监控的预警装置,同时通过动态基线工具,也向管理人员、业务人员,提供了一种通过对海量运维数据进行多视角、多维度分析的工具。只需要在工具中完成指标类型、采集频率、分析周期等相关配置,系统就可自动产生分析结果,是一种高效、直观运营分析工具。

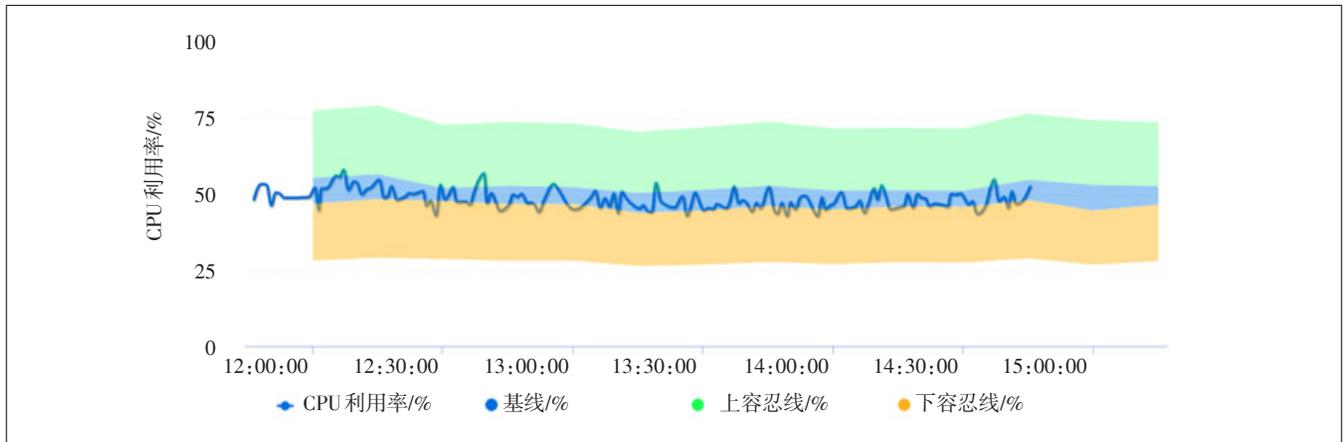


图3 实际预测效果

通过大数据平台不仅提升了运维质量,减少了信息系统发生故障的概率,而且提高了系统的运行效率,保证了系统的正常运行,提升了用户体验,降低用户投诉率,提升了客户感知以及满意度。

#### 4 结束语

本文以动态基线为例详细介绍了大数据技术给运维管理带来的变化,通过将大数据技术引入运维管理系统,将各类运维数据通过大数据平台进行数据挖掘,提升数据价值,还适应未来架构、夯实基础运维。同时利用大数据技术搭建预测算法模型,通过各类运维数据预估系统运行状态,达到提前发现,提前处理,快速定位,从而提升整体运维水平。并且通过大数据技术搭建预测算法模型反馈的结果,不断闭环优化算法模型,使得预测算法达到最优,以达到提升客户感知、提高运维效率、加速故障定位、支撑辅助决策的目的,从而保障了信息系统的运营服务质量。

#### 参考文献:

[1] 陈威. 移动通信运维大数据应用价值探讨研究[J]. 信息通信, 2016(10):59-60.  
[2] 钟成原,和健. 基于大数据异常分析的运维数据分析初探[J]. 移动通信,2016(12):113-114.  
[3] 马丽,汪吉珊. 基于云计算的通信网络智能运维服务管理系统的

设计与应用[J]. 移动通信,2015,39(23):85-91.

[4] 李强. 移动通信运营商办公云[J]. 信息通信,2014(11):240-240.  
[5] 匡华,王翔,王欣明,等. 移动通信运营商办公云[J]. 移动通信, 2014,38(1):61-64.  
[6] 钟湘琼. 基于云计算的大数据智能运维系统设计[J]. 信息通信, 2016(2):146-147.  
[7] 常文旭. 掌上综合运维的支撑研究[J]. 网络运维与管理,2015 (6):40-41.  
[8] 赵健. 电源专业综合运维管理支撑平台的研究[J]. 通信电源技术,2014,31(2):36-38.  
[9] 李学龙,龚海刚. 大数据系统综述[J]. 中国科学:信息科学,2015, 45(1):1-44.  
[10] 何克抗. 大数据面面观[J]. 电化教育研究,2014(10):8-16.  
[11] 刘智慧,张泉灵. 大数据技术研究综述[J]. 浙江大学学报(工学版),2014,48(6):957-972.  
[12] 宫夏屹,李伯虎,柴旭东,等. 大数据平台技术综述[J]. 系统仿真学报,2014,26(3):489-496.  
[13] 徐英超. 运维数据的大数据分析和前瞻性展望[J]. 综合运输, 2015(11):94-99.

#### 作者简介:

罗砚,毕业于天津大学,工程师,硕士,主要从事中国移动IT中心运维管理系统建设和运维工作。

