

云计算中心运维数据可视化研究

冯佳丽

(中国石化石油物探技术研究院 地球物理信息中心,江苏 南京 211103)

摘要:针对云计算中心运维数据分散、可视化程度低等问题,提出了建设基于 InfluxDB 和 Grafana 的云计算中心运维指标可视化平台。该平台从多个维度对运维数据进行展示,使用户能够全面了解云计算中心运行情况,提高运维管理工作的效率。

关键词: 运维数据;可视化;InfluxDB;Grafana

中图分类号: TP311

文献标识码: A

DOI: 10.19358/j.issn.2096-5133.2020.01.019

引用格式:冯佳丽.云计算中心运维数据可视化研究[J].信息技术与网络安全,2020,39(1):100-103.

Research on visualization of cloud computing center operating data

Feng Jiali

(Center for Geophysical Intelligence Research, Sinopec Geophysical Research Institute, Nanjing 211103, China)

Abstract: For the operating data of cloud computing center are scattered and poorly visualized, this paper proposed a visualization platform based on InfluxDB and Grafana. The platform realized a multi-dimensional display of operating data, which allows user to understand the running state of the cloud computing center, to improve the management efficiency.

Key words: operating data; visualization; InfluxDB; Grafana

0 引言

近年来,企业云计算中心规模及服务范围逐步扩大,运维工作量也随之急剧扩张,运维管理中一些薄弱环节逐渐暴露出监控功能单一、数据分散、可视化程度较低等问题^[1]。而大量的生产项目都依赖于云计算中心的软硬件资源,所以需要充分运用大量的运维数据,帮助运维人员提高效率和管理水平^[2-3]。

本文提出建设云计算中心运维数据可视化平

台,为应用用户、运维人员、管理人员提供直观、量化、灵活的运维数据服务,从而有助于做好项目研究、系统维护、发展决策等工作。

1 架构设计

建设运维数据可视化平台首先要梳理所需的运维指标数据项,进行数据采集、数据加工,而后开展图表样式设计、页面设计,逐步开发可视化页面。图1所示为运维数据可视化平台架构。

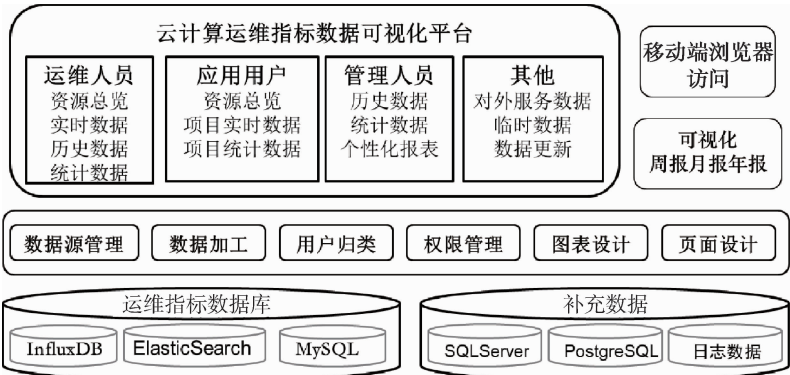


图1 运维数据可视化平台架构

运维最重要的基础就是监控,监控数据需要采集入库,再以生动直观的形式展现出来,为运维工作助力^[4]。以往是将监控数据存入关系数据库中,通过程序读取、处理数据,前端再采用图表插件(如 Highcharts)进行绘图^[5]。这种方式可定制性差,图表美观程度较低,开发周期长。本文中选用开源时序数据库及数据展示前端,快速开发可视化页面。

2 数据采集

云计算中心的运维指标数据属于时序数据,且采样间隔短、数据量大,根据新需求还需要增加数据项,故应选用灵活的时序数据库。

InfluxDB 是一款 Go 语言编写的开源分布式时序、事件和指标数据库^[6]。其主要用于存储基于时间序列的指标数据。InfluxDB 与传统的关系型数据库不同,它使用了 LSM Tree(图 2)存储模型,将大量的随机写转换为顺序写,从而极大地提升了数据写入的性能。且 LSM Tree 的数据排列模式使得读取大块连续的数据时效率也很高。由于时间序列数据的自身特性,通常不是读取单个 key 的值,而是一段范围的数据,所以运用 LSM Tree 模型显得非常合适。

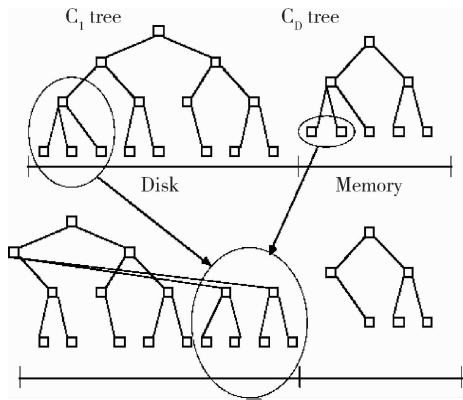


图 2 LSM Tree 遍历合并数据示意图

InfluxDB 时序数据库提升了应对时序数据的处理能力,相较于关系型数据库它的存储空间大大降低,但时间序列函数优越的查询性能远超过关系型数据库^[7]。

运维指标数据通过服务器上部署好的 PAgent 采集组件进行数据收集,多个 PAgent 数据统一发送至 PCollect 进行汇总。tower 组件通过订阅的形式从 PCollect 管道中获取数据,tower 将获取的原数据写入 ES 数据库中,同时也会根据管理员配置的聚合规则在管道中实时聚合数据,再将聚合数据写入 ES、InfluxDB、MySQL 数据库中,如图 3 所示。

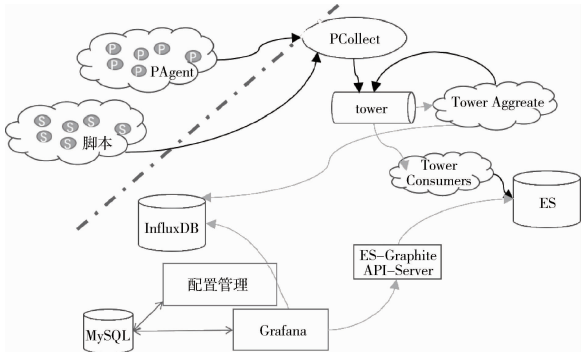


图 3 运维数据采集流程

以 CPU 监控数据为例(如图 4 所示),PAgent 根据配置启动 CPU 采集模块,异步读取 idle、iowait、nice、softirq、sys、user 等 CPU 利用情况数据(数据间隔 30 s),经过数据格式转换,将数据异步发送给 PCollect。PCollect 接收 CPU 信息并进行处理、分析,通过实时数据计算平均值、计算聚合数据(5 分钟、1 小时、1 天),并存储在数据库中。

3 运维指标数据可视化平台建设

本文中运维指标数据可视化平台采用快速迭代开发模式,利用开源的图形化工具 Grafana,将部分用户需求快速转化为可视化界面,再根据用户反馈逐步优化页面。

Grafana 是使用 Go 和 JS 开发的图形编辑器,用于可视化大型数据的开源程序^[8]。它的功能非常强大,可以把它理解为一个时间序列数据库的展示层(前端)^[9]。在 Grafana 中,每个组织都是完全隔离的,有自己的 DataSource(数据库)和 Dashboards(仪表盘),每个 Dashboard 又可以有多个 Panel(用来展示某一具体的测量值)。如图 5 所示,Grafana 从 DataSource 中获取数据,在某个 Dashboard 的某个 Panel 中输入查询条件生成图表。通过 SQL 命令生成一个查询,从而将 Metrics 结果进行展示。Dashboard 中可显示不同 Metric 数据源中的数据。

采集到的运维指标数据主要存储于名为 para-data_dev 的 InfluxDB 数据库中,其中包含集群 CPU 利用率、GPU 利用率、内存利用率、磁盘利用率、网络吞吐量、存储占用率、读写速率、项目占用空间、项目机时、交换机端口数、收发速率等。要使这些数据能够在运维指标数据可视化平台中展示,首先需要配置数据源,如图 6 所示,设置数据库名称、数据库类型、地址、访问方式等参数,将 para-data_dev 数据库设为当前默认数据源。

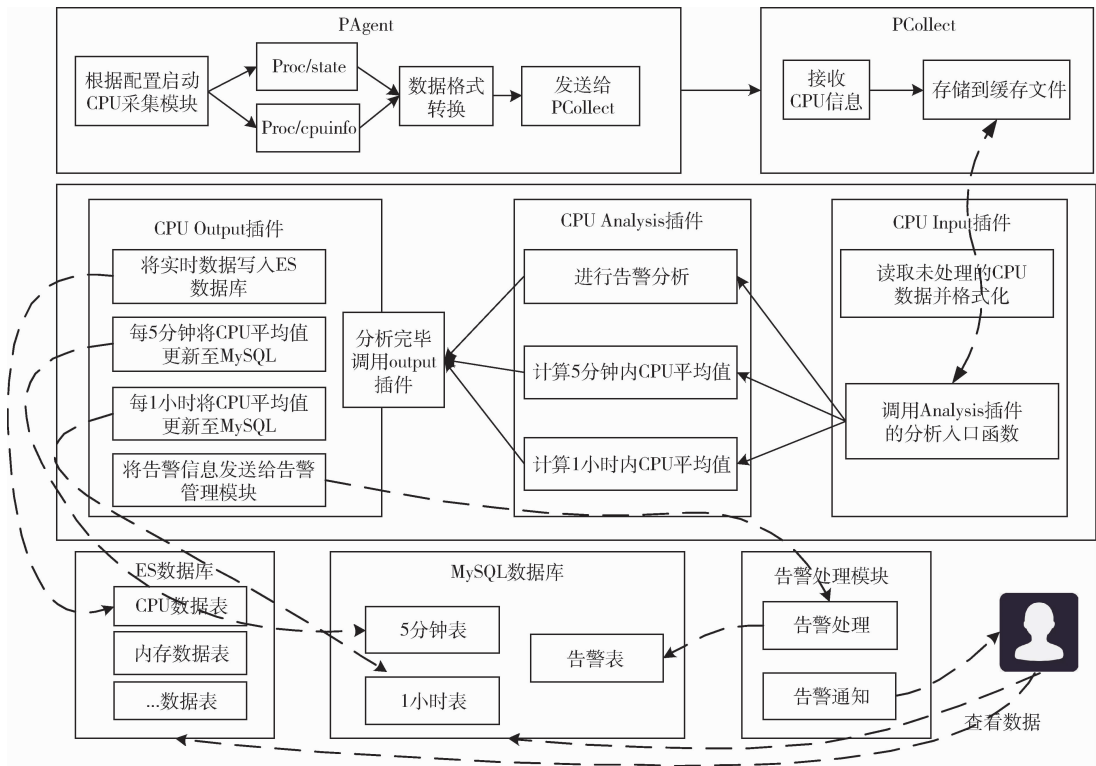


图 4 数据采集示例

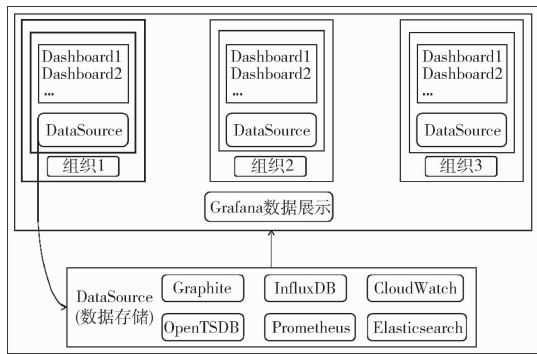


图 5 Grafana 架构

Edit data source

Name	paradata_dev	Default	☒
Type	influxDB		

HTTP settings

URL	
Access	proxy

图 6 数据源配置

本文中的云计算中心拥有十几套集群,共计 2 000多个节点,每台主机又有多项监控指标。要展示每个节点的运维数据,需要配置无数的页面。而模

板化变量查询功能极大地方便了配置监控图表。数据查询、图表标题等都可使用变量,通过页面交互操作作为变量赋值,使得一个可视化页面可以千变万化。

以 CPU 集群节点利用情况为例,集群的 CPU 利用率信息都存储在 load-cpu 表中,如图 7 所示。可根据 tagKey(包括集群、CPU 名、节点名、节点 id)进行数据筛选、分组,获取 fieldType(各项 CPU 利用率值)。

通过 Templating,可将主要的 tagKey 设置为参数模板,实现多个集群、多个节点在同一个页面进行展示。如图 8 所示,构建了两个参数模板 \$ cluster、\$ host,且前者可作为后者的查询参数,实现二级模板。

形成的参数模板会自动显示在该平台可视化页面的左上方,如图 9 所示,通过 cluster 下拉列表可选择想要查看的集群,然后再通过 host 下拉列表选择该集群中指定的节点。

基于参数模板,通过类 SQL 语句,实现对某个集群中特定节点的数据抽取。

```
SELECT "user" AS "user", "sys" AS "sys", "irq" AS "irq", "softirq" AS "softirq", "iowait" AS "iowait", "nice" AS "nice" FROM "load-cpu" WHERE ( "cluster" = ~ / ^ $ cluster $ / AND "hostname" = ~ / ^ $ host $ / AND "cpu_name" = 'cpu') AND $ timeFilter
```

	fieldKey	fieldtype
	active_cpu_count	float
	count	float
	cpunum	float
	format	string
	guest	float
	guest_nice	float
	hostid	string
	idle	float
	iowait	float
	ip	string
	irq	float
	mmid	string
	nice	float
	softirq	float
	steal	float
	suboid	string
	sys	float
	user	float
tagKey		
cluster		
cpu_name		
hostname		
tag_hostid		

图7 load-cpu 数据示例

Templating

Variables Help

```

$cluster SHOW TAG VALUES from "load-cpu" WITH KEY = "cluster"
$host SHOW TAG VALUES from "load-cpu" WITH KEY = "hostname" WHERE "cluster" =~ /^$cluster$/

```

图8 构建参数模板

cluster sgibdw host sgibdw220

图9 参数模板列表

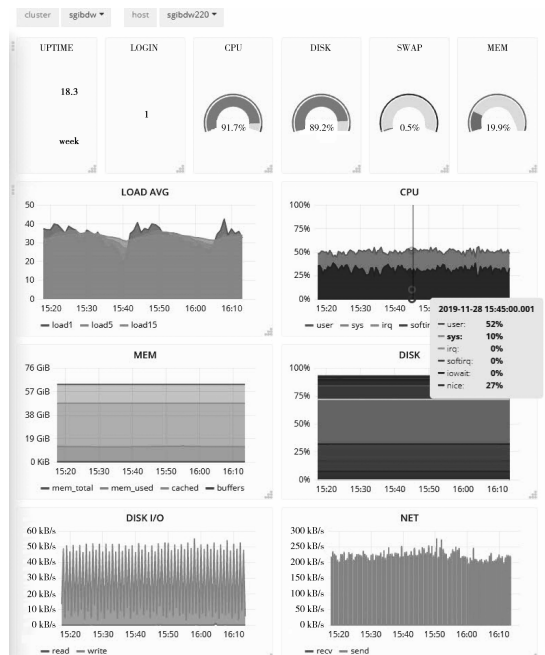


图11 集群节点运维数据可视化

以上 SQL 语句可查询该节点的系统 CPU 消耗、用户态 CPU 消耗等。其中使用的 \$ cluster 参数会根据一级列表中选取的集群名进行填充, \$ host 根据二级列表中的节点名进行填充,此外, \$ timeFilter 参数根据页面中选定的时间范围进行填充。从而形成完整的查询代码,对 InfluxDB 中的数据进行抽取,实现 CPU 利用率可视化,如图 10 所示为 sgibdw 集群 sgibdw220 节点的 CPU 利用数据展示。

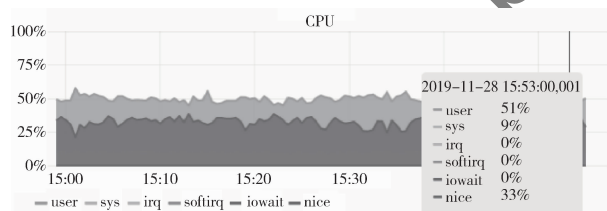


图10 集群节点 CPU 利用数据展示

以此类推,编写不同的类 SQL 语句,结合模板参数,从数据源中抽取 CPU 利用率、磁盘利用率、内存利用率、网络收发速率等时序数据,形成集群节点详细运维数据可视化页面,通过页面左上方的模板选择指定的节点,切换页面展示内容,如图 11 所示。

4 结论

本文提出的运维指标可视化平台,对云计算中心的运维指标数据进行了规范化、集中化、图形化管理,用丰富直观的可视化页面,从多个维度展示了云计算中心集群、存储、网络设备、项目、机时等

数据,使用户能够全面了解云计算中心运行情况,提高运维管理工作的效率。

参考文献

- [1] 赖作财,吴华娟.基于云计算的智慧机房建设与运维管理研究[J].电子世界,2018(2):194-195.
- [2] 祁文坤.大数据及可视化平台在电力企业中的应用[J].电子世界,2016(24):138-139.
- [3] 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.
- [4] 杨秀云,郭磊,刘露,等.云计算数据中心移动运维管理系统的设计与实现[J].电子技术与软件工程,2017(14):199-200.
- [5] 和荣,肖海力.基于 Nagios 的监控数据分析展示平台[J].科研信息化技术与应用,2017,8(5):36-43.
- [6] 柴亚刚.基于时序数据库的分部署网络波动监控系统[J].中国传媒科技,2018(3):36-37.
- [7] 潘庆和.一种通用分布式数据抓取系统的设计与实现[J].哈尔滨商业大学学报(自然科学版),2016,32(3):307-312.
- [8] 陈军.基于 Flume 的分布式日志聚合系统的研究[J].科技视界,2017(11):77,114.
- [9] 杜翔军.基于 Zabbix 技术整合第三方插件的自动化监控系统研发[J].计算机产品与流通,2017(9):149.

(收稿日期:2019-09-23)

作者简介:

冯佳丽(1987-),女,硕士,工程师/副主任师,主要研究方向:信息化系统开发与云计算中心运维管理。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《信息技术与网络安全》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《信息技术与网络安全》编辑部
中国电子信息产业集团有限公司第六研究所