

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号

WO 2020/168756 A1

(43) 国际公布日

2020 年 8 月 27 日 (27.08.2020)

(51) 国际专利分类号:
G06F 16/17 (2019.01)

(21) 国际申请号: PCT/CN2019/118288

(22) 国际申请日: 2019 年 11 月 14 日 (14.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910123928.1 2019 年 2 月 19 日 (19.02.2019) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。

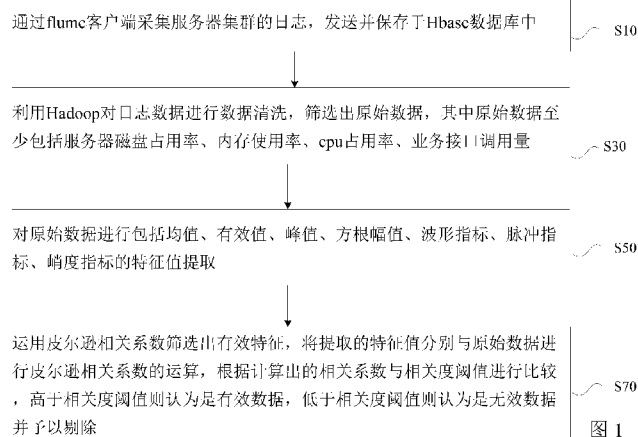
(72) 发明人: 吴超勇(WU, Chaoyong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。
陈仕财(CHEN, Shicai); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。

(74) 代理人: 北京鸿元知识产权代理有限公司(GRANDER IP LAW FIRM); 中国北京市朝阳区光华路 7 号汉威大厦东区 18A6 室, Beijing 100004 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,

(54) Title: CLUSTER LOG FEATURE EXTRACTION METHOD, AND APPARATUS, DEVICE AND STORAGE MEDIUM

(54) 发明名称: 集群日志特征提取方法、装置、设备及存储介质



S10 A flume client collects a log of a server cluster, and sends same and stores same in an Hbase database
S30 Use Hadoop to perform data cleaning on log data to screen out original data, wherein the original data at least comprises a server disk occupancy rate, a memory usage rate, a CPU occupancy rate and a service interface call volume
S50 Extract feature values, comprising the mean value, effective value, peak value, root amplitude, waveform index, impulse index and kurtosis index, of the original data
S70 Use a Pearson correlation coefficient to screen out valid features, respectively perform operation of the Pearson correlation coefficient on the extracted feature values and the original data, compare calculated correlation coefficients with a correlation threshold, regard correlation coefficients higher than the correlation threshold as being valid data, and regard correlation coefficients lower than the correlation threshold as being invalid data and remove same

(57) Abstract: A cluster log feature extraction method, and an apparatus, a device and a storage medium. The method comprises: a flume client collecting a log of a server cluster, and sending same to a database (S10); performing data cleaning on log data to screen out original data (S30); extracting feature values, comprising the mean value, effective value, peak value, root amplitude, waveform index, impulse index and kurtosis index, of the original data (S50); and respectively performing operation of a Pearson correlation coefficient on the extracted feature values and the original data, comparing calculated correlation coefficients with a correlation threshold, regarding

JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

correlation coefficients higher than the correlation threshold as being valid data, and regarding correlation coefficients lower than the correlation threshold as being invalid data and removing same (S70). Valid information of production data of each host in a server cluster can be effectively screened out, and feature values of the production data are extracted from the valid information, thereby facilitating failure prediction and failure classification of a production system and reducing the occurrence of production accidents.

(57) 摘要: 一种集群日志特征提取方法、装置、设备及存储介质, 通过 flume 客户端采集服务器集群的日志, 发送至数据库 (S10); 对日志数据进行数据清洗, 筛选出原始数据 (S30); 对原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取 (S50); 将提取的特征值分别与原始数据进行皮尔逊相关系数的运算, 根据计算出的相关系数与相关度阈值进行比较, 高于相关度阈值则认为有效数据, 低于相关度阈值则认为无效数据并予以剔除 (S70)。能有效筛选出服务器集群中各主机的生产数据的有效信息, 且从有效信息中提取出生产数据的特征值, 便于生产系统的故障预测和故障分类, 减少生产事故的发生。

集群日志特征提取方法、装置、设备及存储介质

本申请要求于2019年02月19日提交的中国专利申请号201910123928.1的优先权益，上述案件全部内容以引用的方式并入本文中。

技术领域

本申请涉及基架运维，具体地说，涉及一种集群日志特征提取方法、装置、设备及存储介质。

背景技术

在信息爆炸式增长的时代，文件大小和数据规模迈向 TB 级甚至 PB 级已成现实，集群存储系统节点数已达到 64 节点集群数目，管理如此庞大的集群系统已经成为数据中心所面临的严峻挑战。及时跟踪集群节点运行状态，精确定位节点出错信息变得尤为重要。在集群存储系统实际的运行中，目前常用一种集群存储系统日志管理方法，可以定时或实时发送系统日志，实现了日志的集中传输，但是没有对日志进行分析和处理，不能全局的了解整个集群存储系统的运行情况，不能快速的定位到错误信息。但是随着集群节点数的增多，对集群系统管理变得越来越复杂。从海量服务器数据中，抽取出能反映服务器性能的特征，精确定位集群节点的潜在故障，提前做好相应的性能检测显得尤为重要。

发明内容

为解决以上问题，本申请提供一种集群日志特征提取方法，应用于电子设备，包括以下步骤：通过 flume 客户端采集服务器集群的日志，发送至 Hbase 数据库，其中，flume 客户端通过多个 Agent 进程对应采集服务器集群中的每台服务器的日志，Agent 定时将对应的服务器上的日志数据收集并通过 API 接口发送到 Hbase 数据库；利用 Hadoop 对日志数据进行数据清洗，筛选出原始数据，其中原始数据至少包括服务器磁盘占用率、内存使用率、cpu 占用率、业务接口调用量；对原始数据进行包括均值、有效值、峰值、方根幅值、波

形指标、脉冲指标、峭度指标的特征值提取；运用皮尔逊相关系数筛选出有效特征，将提取的特征值分别与原始数据进行皮尔逊相关系数的运算，根据计算出的相关系数与相关度阈值进行比较，高于相关度阈值则认为是有效数据，低于相关度阈值则认为是无效数据并予以剔除。

本申请还提供一种集群日志特征提取装置，包括：日志采集模块、数据清洗模块、特征提取模块、有效特征筛选模块，其中，日志采集模块用于通过 flume 客户端采集服务器集群的日志，发送至 Hbase 数据库，其中，flume 客户端通过多个 Agent 进程对应采集服务器集群中的每台服务器的日志，Agent 定时将对应的服务器上的日志数据收集并通过 API 接口发送到 Hbase 数据库；数据清洗模块用于利用 Hadoop 对日志数据进行数据清洗，筛选出原始数据，其中原始数据至少包括服务器磁盘占用率、内存使用率、cpu 占用率、业务接口调用量；特征提取模块用于对原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取；有效特征筛选模块运用皮尔逊相关系数筛选出有效特征：将提取的特征值分别与原始数据进行皮尔逊相关系数的运算，根据计算出的相关系数与相关度阈值进行比较，高于相关度阈值则是有效数据，低于相关度阈值则是无效数据，并予以剔除。

本申请还提供一种电子设备，该电子设备包括：存储器和处理器，所述存储器中存储有集群日志特征提取程序，所述集群日志特征提取程序被所述处理器执行时实现如下步骤：通过 flume 客户端采集服务器集群的日志，发送至 Hbase 数据库，其中，flume 客户端通过多个 Agent 进程对应采集服务器集群中的每台服务器的日志，Agent 定时将对应的服务器上的日志数据收集并通过 API 接口发送到 Hbase 数据库；利用 Hadoop 对日志数据进行数据清洗，筛选出原始数据，其中原始数据至少包括服务器磁盘占用率、内存使用率、cpu 占用率、业务接口调用量；对原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取；运用皮尔逊相关系数筛选出有效特征，将提取的特征值分别与原始数据进行皮尔逊相关系数的运算，根据计算出的相关系数与相关度阈值进行比较，高于相关度阈值则认为是有效数据，低于相关度阈值则认为是无效数据并予以剔除。

本申请还提供一种计算机非易失性可读存储介质，所述计算机非易失性可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令被处理器执行时，实现以上所述的集群日志特征提取方法。

本申请能有效筛选出服务器集群中各主机的生产数据的有效信息，且从有效信息中提取出生产数据的特征值，便于生产系统的故障预测和故障分类，减少生产事故的发生。

附图说明

通过结合下面附图对其实施例进行描述，本申请的上述特征和技术优点将会变得更加清楚和容易理解。

图 1 是本申请实施例的集群日志特征提取方法的流程示意图；

图 2 是本申请实施例的电子设备的硬件架构示意图；

图 3 是本申请实施例的集群日志特征提取程序的模块构成图；

图 4 是本申请实施例的日志采集模块的单元构成图；

图 5 是本申请实施例的特征提取模块的单元构成图；

图 6 是本申请实施例的数据清洗模块的单元构成图；

图 7 是 Flume 的 Agent 进程读取数据的示意图。

具体实施方式

下面将参考附图来描述本申请所述的集群日志特征提取方法、装置及存储介质的实施例。本领域的普通技术人员可以认识到，在不偏离本申请的精神和范围的情况下，可以用各种不同的方式或其组合对所描述的实施例进行修正。因此，附图和描述在本质上是说明性的，而不是用于限制权利要求的保护范围。此外，在本说明书中，附图未按比例画出，并且相同的附图标记表示相同的部分。

如图 1 所示，本实施例的集群日志特征提取方法，包括如下步骤：

步骤 S10，通过 flume（分布式的海量日志采集、聚合和传输系统）客户端采集服务器集群的日志，发送至 Hbase 数据库服务器。Flume 以 Agent 进程为最小的独立运行单位，一个 Agent 进程就是一个完整的数据收集工具。如图 7 所示，Agent 包含组件 Source（数据收集组件）、Channel（中转临时存储）、

Sink, 三者组建了一个 Agent, source 从服务器收集数据, 传递给 Channel, Channel 保存由 Source 组件传递过来的 Event (数据单元), Sink 从 Channel 中读取并移除 Event, 将 Event 传递到后台。Flume 通过多个 Agent 来对应各服务器收集日志数据。对应每一台服务器设置一个 Agent, 定时将对应的服务器上的日志数据收集并通过 API 接口发送到后台。

步骤 S30, 利用 Hadoop (分布式系统基础架构) 对日志数据进行数据清洗, 筛选出原始数据, 其中原始数据至少包括服务器磁盘占用率、内存使用率、cpu 占用率、业务接口调用量。

步骤 S50, 对原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取。

步骤 S70, 运用皮尔逊相关系数筛选出有效特征: 将提取的特征值分别与原始数据进行皮尔逊相关系数的运算, 根据计算出的相关系数与相关度阈值进行比较, 高于相关度阈值则认为是有效数据, 低于相关度阈值则认为是无效数据并予以剔除。

进一步地, 数据清洗中采用拉依达准则剔除具有粗大误差的数据, 包括以下步骤:

对日志数据 $x_1, x_2, \dots, x_n$, 计算其算术平均值 $\bar{X}$ 及剩余误差 $v_i = x_i - \bar{X} (i=1, 2, \dots, n)$, 其中, x_i 为单次 Agent 采集的日志数据;

$$\text{计算标准偏差 } S_x, S_x = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \right)^{1/2};$$

若日志数据中的 x_b 的剩余误差 $v_b (1 \leq b \leq n)$, 满足公式 $|v_b| = |x_b - \bar{X}| > 3S_x$

则认为 x_b 是含有粗大误差值的奇异值, 并剔除奇异值。

进一步地, 采用拉依达法则能有效地识别出生产数据的奇异值, 但对于剔除掉的数据则会产生空值。因此, 对识别出的日志数据的奇异值用中值替代, 实现对生产数据信息的预处理。其中所述中值是指将各个变量值 $x_1, x_2, \dots, x_n$ 按大小顺序排列起来, 形成一个数列, 处于变量数列中间位置的变量值就称为中值。

在一个可选实施例中, 对原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取, 其中,

有效值采用如下公式计算： $X_{\text{rms}} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$

峰值采用如下公式计算： $X_p = \max(x_i)$

方根幅值采用如下公式计算： $X_r = \left[\frac{1}{N} \sum_{i=1}^N \sqrt{|x_i|} \right]^2$

波形指标采用如下公式计算： $X_{\text{ws}} = \frac{X_{\text{rms}}}{|\bar{x}|}$

脉冲指标采用如下公式计算： $X_{\text{if}} = \frac{X_p}{|\bar{x}|}$

峭度指标采用如下公式计算： $X_{\text{kv}} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{X_{\text{rms}}^4}$

其中， x_i 为单次 Agent 采集的日志数据； N 为日志数据采集的次数； $\bar{x}$ 为采集的日志数据的算术平均值； X_{rms} 为采集的日志数据的有效值； X_p 为采集的日志数据的峰值； X_r 为采集的日志数据的方根幅值； X_{ws} 为采集的日志数据的波形指标； X_{if} 为采集的日志数据的脉冲指标； X_{kv} 为采集的日志数据的峭度指标。

运用皮尔逊相关系数筛选出有效特征，具体说，是将以上特征值分别与原始数据进行皮尔逊相关系数的运算，根据计算出的相关系数与相关度阈值来比较，高于相关度阈值则认为是有效数据，低于相关度阈值则认为是无效数据，需要予以剔除，从而可筛选出有效的数据。例如，相关度阈值为 0.7，方根幅值与原始数据的相关系数为 0.2，则表明方根幅值为无效数据，峭度指标与原始数据的相关系数为 0.85，则认定峭度指标为有效数据。其中，皮尔逊相关系数的公式如下：

$$\rho = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

其中， x_i 为单次 Agent 采集数据值； y_i 为单次 Agent 采集数据中提取的某一特征值； $\bar{x}$ 是日志数据 $x_1, x_2, \dots, x_n$ 的算数平均值； $\bar{y}$ 是 $y_1, y_2, \dots, y_n$ 的算数平均值； N 为日志数据采集的次数。

在一个可选实施例，Flume 包括多个第一层级 Agent 和一个第二层级

Agent，每个第一层级 Agent 分别对应的采集一个服务器的日志数据，多个第一层级 Agent 采集的日志数据汇集至第二层级 Agent，并由第二层级 Agent 传输至 HDFS（分布式文件系统）中。

参阅图 2 所示，是本申请电子设备的实施例的硬件架构示意图。本实施例中，所述电子设备 2 是一种能够按照事先设定或者存储的指令，自动进行数值计算和/或信息处理的设备。例如，可以是智能手机、平板电脑、笔记本电脑、台式计算机、机架式服务器、刀片式服务器、塔式服务器或机柜式服务器(包括独立的服务器，或者多个服务器所组成的服务器集群)等。如图 2 所示，所述电子设备 2 至少包括，但不限于，可通过系统总线相互通信连接的存储器 21、处理器 22、网络接口 23。其中：所述存储器 21 至少包括一种类型的计算机可读存储介质，所述可读存储介质包括闪存、硬盘、多媒体卡、卡型存储器(例如，SD 或 DX 存储器等)、随机访问存储器(RAM)、静态随机访问存储器(SRAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、可编程只读存储器(PROM)、磁性存储器、磁盘、光盘等。在一些实施例中，所述存储器 21 可以是所述电子设备 2 的内部存储单元，例如该电子设备 2 的硬盘或内存。在另一些实施例中，所述存储器 21 也可以是所述电子设备 2 的外部存储设备，例如该电子设备 2 上配备的插接式硬盘，智能存储卡(Smart Media Card, SMC)，安全数字(Secure Digital, SD)卡，闪存卡(Flash Card)等。当然，所述存储器 21 还可以既包括所述电子设备 2 的内部存储单元也包括其外部存储设备。本实施例中，所述存储器 21 通常用于存储安装于所述电子设备 2 的操作系统和各类应用软件，例如所述集群日志特征提取程序代码等。此外，所述存储器 21 还可以用于暂时地存储已经输出或者将要输出的各类数据。

所述处理器 22 在一些实施例中可以是中央处理器(Central Processing Unit, CPU)、控制器、微控制器、微处理器、或其他数据处理芯片。该处理器 22 通常用于控制所述电子设备 2 的总体操作，例如执行与所述电子设备 2 进行数据交互或者通信相关的控制和处理等。本实施例中，所述处理器 22 用于运行所述存储器 21 中存储的程序代码或者处理数据，例如运行所述的集群日志特征提取程序等。

所述网络接口 23 可包括无线网络接口或有线网络接口，该网络接口 23

通常用于在所述电子设备 2 与其他电子设备之间建立通信连接。例如，所述网络接口 23 用于通过网络将所述电子设备 2 与推送平台相连，在所述电子设备 2 与推送平台之间建立数据传输通道和通信连接等。所述网络可以是企业内部网(Intranet)、互联网(Internet)、全球移动通讯系统(Global System of Mobile communication, GSM)、宽带码分多址(Wideband CodeDivision Multiple Access, WCDMA)、4G 网络、5G 网络、蓝牙(Bluetooth)、Wi-Fi 等无线或有线网络。

可选地，该电子设备 2 还可以包括显示器，显示器也可以称为显示屏或显示单元。在一些实施例中可以是 LED 显示器、液晶显示器、触控式液晶显示器以及有机发光二极管(Organic Light-Emitting Diode, OLED)显示器等。显示器用于显示在电子设备 2 中处理的信息以及用于显示可视化的用户界面。

需要指出的是，图 2 仅示出了具有组件 21-23 的电子设备 2，但是应理解的是，并不要求实施所有示出的组件，可以替代的实施更多或者更少的组件。

包含可读存储介质的存储器 21 中可以包括操作系统、集群日志特征提取程序 50 等。处理器 22 执行存储器 21 中集群日志特征提取程序 50 时实现如下步骤：

步骤 S10，通过 flume（分布式的海量日志采集、聚合和传输系统）客户端采集服务器集群的日志，发送至 Hbase 数据库服务器。Flume 以 Agent 组件为最小的独立运行单位，一个 Agent 组件就是一个完整的数据收集工具。Flume 通过多个 Agent 来对应各服务器收集日志数据。对应每一台服务器设置一个 Agent，定时将对应的服务器上的日志数据收集并通过 API 接口发送到后台。

步骤 S30，利用 Hadoop（分布式系统基础架构）对日志数据进行数据清洗，筛选出原始数据，其中原始数据至少包括服务器磁盘占用率、内存使用率、cpu 占用率、业务接口调用量。

步骤 S50，对原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取。

步骤 S70，运用皮尔逊相关系数筛选出有效特征，将提取的特征值分别与原始数据进行皮尔逊相关系数的运算，根据计算出的相关系数与相关度阈值进行比较，高于相关度阈值则认为有效数据，低于相关度阈值则认为无效数据并予以剔除。

需要说明的是，本申请之电子设备的具体实施方式与上述集群日志提取

方法的具体实施方式大致相同，在此不再赘述。

在本实施例中，存储于存储器 21 中的所述集群日志特征提取程序可以被分割为一个或者多个程序模块，所述一个或者多个程序模块被存储于存储器 21 中，并可由一个或多个处理器(本实施例为处理器 22)所执行，以完成本申请。例如，图 3 示出了所述集群日志特征提取程序的程序模块示意图，该实施例中，所述集群日志特征提取程序 50 可以被分割为日志采集模块 501、数据清洗模块 502、特征提取模块 503、有效特征筛选模块 504。其中，本申请所称的程序模块是指能够完成特定功能的一系列计算机程序指令段，比程序更适合于描述所述集群日志特征提取程序在所述电子设备 2 中的执行过程。通过所述程序模块的具体功能实现集群日志特征提取方法。

本申请还提供一种集群日志特征提取装置，包括：日志采集模块 501、数据清洗模块 502、特征提取模块 503、有效特征筛选模块 504。

其中，日志采集模块 501 用于通过 flume（分布式的海量日志采集、聚合和传输系统）客户端采集服务器集群的日志，发送至 Hbase 数据库服务器。Flume 以 Agent 组件为最小的独立运行单位，一个 Agent 组件就是一个完整的数据收集工具。Flume 通过多个 Agent 来对应各服务器收集日志数据。对应每一台服务器设置一个 Agent，定时将对应的服务器上的日志数据收集并通过 API 接口发送到后台。

数据清洗模块 502 用于利用 Hadoop（分布式系统基础架构）对日志数据进行数据清洗，筛选出原始数据，其中原始数据至少包括服务器磁盘占用率、内存使用率、cpu 占用率、业务接口调用量。

特征提取模块 503 用于对原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取。

有效特征筛选模块 504 运用皮尔逊相关系数筛选出有效特征，将提取的特征值分别与原始数据进行皮尔逊相关系数的运算，根据计算出的相关系数与相关度阈值进行比较，高于相关度阈值则认为是有效数据，低于相关度阈值则认为是无效数据并予以剔除。

在一个可选实施例中，如图 6 所示，数据清洗模块 502 包括拉依达准则判定单元 5021，拉依达准则判定单元 5021 采用拉依达准则剔除具有粗大误差的数据，包括以下步骤：

对日志数据 $x_1, x_2, \dots, x_n$, 计算其算术平均值 $\bar{X}$ 及剩余误差 $v_i = x_i - \bar{X} (i=1, 2, \dots, n)$, 其中, x_i 为单次 Agent 采集数据值;

计算标准偏差 S_x , $S_x = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \right)^{1/2}$;

若数据 x_b 的剩余误差 v_b ($1 \leq b \leq n$), 满足下式 $|v_b| = |x_b - \bar{X}| > 3S_x$

则认为 x_b 是含有粗大误差值的奇异值, 并剔除该奇异值。

进一步地, 数据清洗模块 502 还包括奇异值替换单元 5022。采用拉依达法则能有效地识别出生成数据的奇异值, 但对于剔除掉的数据则会产生空值。奇异值替换单元 5022 对识别出的日志数据的奇异值用中值替代, 实现对生产数据信息的预处理。其中所述中值是指将各个变量值 $x_1, x_2, \dots, x_n$ 按大小顺序排列起来, 形成一个数列, 处于变量数列中间位置的变量值就称为中值。

在一个可选实施例中, 如图 5 所示, 特征提取模块 503 包括均值提取单元 5031、有效值提取单元 5032、峰值提取单元 5033、方根幅值提取单元 5034、波形指标提取单元 5035、脉冲指标提取单元 5036、峭度指标提取单元 5037。分别对原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取, 其中,

有效值采用如下公式计算: $X_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$

峰值采用如下公式计算: $X_p = \max(x_i)$

方根幅值采用如下公式计算: $X_r = \left[\frac{1}{N} \sum_{i=1}^N \sqrt{|x_i|} \right]^2$

波形指标采用如下公式计算: $X_{ws} = \frac{X_{rms}}{|\bar{X}|}$

脉冲指标采用如下公式计算: $X_{if} = \frac{X_p}{|\bar{X}|}$

峭度指标采用如下公式计算: $X_{kv} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})^4}{X_{rms}^4}$

其中, x_i 为单次 Agent 采集的日志数据; N 为日志数据采集的次数; $\bar{X}$ 为采集的日志数据的算术平均值; X_{rms} 为采集的日志数据的有效值; X_p 为采集的日志数据的峰值; X_r 为采集的日志数据的方根幅值; X_{ws} 为采集的日志数据

的波形指标； X_{if} 为采集的日志数据的脉冲指标； X_{kv} 为采集的日志数据的峭度指标。

运用皮尔逊相关系数筛选出有效特征，具体说，是将以上特征值分别与原始数据进行皮尔逊相关系数的运算，根据计算出的相关系数与相关度阈值来比较，高于相关度阈值则认为有效数据，低于相关度阈值则认为无效数据，需要予以剔除，从而可筛选出有效的数据。例如，相关度阈值为 0.7，方根幅值与原始数据的相关系数为 0.2，则表明方根幅值为无效数据，峭度指标与原始数据的相关系数为 0.85，则认定峭度指标为有效数据。其中，皮尔逊相关系数的公式如下：

$$\rho = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{X})^2 \sum_{i=1}^N (y_i - \bar{Y})^2}}$$

其中， x_i 为单次 Agent 采集数据值； y_i 为单次 Agent 采集数据中提取的某一特征值； $\bar{X}$ 是日志数据 $x_1, x_2, \dots, x_n$ 的算数平均值； $\bar{Y}$ 是 $y_1, y_2, \dots, y_n$ 的算数平均值； N 为数据采集的次数。

在一个可选实施例中，如图 4 所示，日志采集模块 501 还包括 Agent 设置单元 5011，用于针对 Flume 进行包括多个第一层级 Agent 和一个第二层级 Agent 的设置，每个第一层级 Agent 分别对应的采集一个服务器的日志数据，多个第一层级 Agent 采集的日志数据汇集至第二层级 Agent，并由第二层级 Agent 传输至 HDFS 中。

需要说明的是，本申请之集群日志特征提取装置的具体实施方式与上述集群日志特征提取方法、电子设备的具体实施方式大致相同，在此不再赘述。

此外，本申请实施例还提出一种计算机非易失性可读存储介质，所述计算机可读存储介质可以是硬盘、多媒体卡、SD 卡、闪存卡、SMC、只读存储器(ROM)、可擦除可编程只读存储器(EPROM)、便携式紧致盘只读存储器(CD-ROM)、USB 存储器等等中的任意一种或者几种的任意组合。所述计算机非易失性可读存储介质中包括集群日志特征提取程序等，所述集群日志特征提取程序 50 被处理器 22 执行时实现如下操作：

步骤 S10, 通过 flume 客户端采集服务器集群的日志, 发送至 Hbase 数据库服务器。Flume 以 Agent 组件为最小的独立运行单位, 一个 Agent 组件就是一个完整的数据收集工具。Flume 通过多个 Agent 来对应各服务器收集日志数据。对应每一台服务器设置一个 Agent, 定时将对应的服务器上的日志数据收集并通过 API 接口发送到后台。

步骤 S30, 利用 Hadoop 对日志数据进行数据清洗, 筛选出原始数据, 其中原始数据至少包括服务器磁盘占用率、内存使用率、cpu 占用率、业务接口调用量。

步骤 S50, 对原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取。

步骤 S70, 运用皮尔逊相关系数筛选出有效特征, 将提取的特征值分别与原始数据进行皮尔逊相关系数的运算, 根据计算出的相关系数与相关度阈值进行比较, 高于相关度阈值则认为是有效数据, 低于相关度阈值则认为是无效数据并予以剔除。

本申请之计算机非易失性可读存储介质的具体实施方式与上述集群日志特征提取方法以及电子设备2的具体实施方式大致相同, 在此不再赘述。

以上所述仅为本申请的优选实施例, 并不用于限制本申请, 对于本领域的技术人员来说, 本申请可以有各种更改和变化。凡在本申请的精神和原则之内, 所作的任何修改、等同替换、改进等, 均应包含在本申请的保护范围之内。

权 利 要 求 书

1. 一种集群日志特征提取方法，应用于电子设备，其特征在于，包括以下步骤：

通过 flume 客户端采集服务器集群的日志，发送至 Hbase 数据库，其中，flume 客户端通过多个 Agent 进程对应采集服务器集群中的每台服务器的日志，Agent 定时将对应的服务器上的日志数据收集并通过 API 接口发送到 Hbase 数据库；

利用 Hadoop 对日志数据进行数据清洗，筛选出原始数据，其中原始数据至少包括服务器磁盘占用率、内存使用率、cpu 占用率、业务接口调用量；

对原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取；

运用皮尔逊相关系数筛选出有效特征：将提取的特征值分别与原始数据进行皮尔逊相关系数的运算，根据计算出的相关系数与相关度阈值进行比较，高于相关度阈值则是有效数据，低于相关度阈值则是无效数据，并予以剔除。

2. 根据权利要求 1 所述的集群日志特征提取方法，其特征在于，在数据清洗过程中，采用拉依达准则剔除具有粗大误差的数据，包括以下步骤：

对日志数据 $x_1, x_2, \dots, x_n$ ，计算其算术平均值 $\bar{X}$ 及剩余误差 $v_i = x_i - \bar{X} (i=1, 2, \dots, n)$ ，其中， x_i 为单次 Agent 采集的日志数据；

计算标准偏差 S_x ，
$$S_x = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \right)^{1/2}；$$

若日志数据中的 x_b 的剩余误差 v_b ($1 \leq b \leq n$)，满足公式 $|v_b| = |x_b - \bar{X}| > 3S_x$

则确定 x_b 是含有粗大误差值的奇异值，并剔除奇异值。

3. 根据权利要求 2 所述的集群日志特征提取方法，其特征在于，对日志数据的奇异值用中值替代，其中所述中值是指将各个日志数据 $x_1, x_2, \dots, x_n$ 按大小顺序排列，处于中间位置的值称为中值。

4. 根据权利要求 2 所述的集群日志特征提取方法，其特征在于，原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取，其中，

有效值采用如下公式计算： $X_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$

峰值采用如下公式计算： $X_p = \max(x_i)$

方根幅值采用如下公式计算： $X_r = \left[\frac{1}{N} \sum_{i=1}^N \sqrt{|x_i|} \right]^2$

波形指标采用如下公式计算： $X_{ws} = \frac{X_{rms}}{|\bar{x}|}$

脉冲指标采用如下公式计算： $X_{if} = \frac{X_p}{|\bar{x}|}$

峭度指标采用如下公式计算： $X_{kv} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{X_{rms}^4}$

其中， x_i 为单次 Agent 采集的日志数据；N 为数据采集的次数； $\bar{x}$ 为采集的日志数据的算术平均值； X_{rms} 为采集的日志数据的有效值； X_p 为采集的日志数据的峰值； X_r 为采集的日志数据的方根幅值； X_{ws} 为采集的日志数据的波形指标； X_{if} 为采集的日志数据的脉冲指标； X_{kv} 为采集的日志数据的峭度指标。

5. 根据权利要求 2 所述的集群日志特征提取方法，其特征在于，皮尔逊相关系数的公式如下：

$$\rho = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

其中， x_i 为单次 Agent 采集的日志数据； y_i 为单次 Agent 采集数据中提取的某一特征值； $\bar{x}$ 是日志数据 $x_1, x_2, \dots, x_n$ 的算数平均值； $\bar{y}$ 是 $y_1, y_2, \dots, y_n$ 的算数平均值；N 为日志数据采集的次数。

6. 根据权利要求 1 所述的集群日志特征提取方法，其特征在于，

Flume 包括多个第一层级 Agent 和一个第二层级 Agent，每个第一层级 Agent 分别对应的采集一个服务器的日志数据，多个第一层级 Agent 采集的日志数据汇集至第二层级 Agent，并由第二层级 Agent 传输至 HDFS 中。

7. 一种集群日志特征提取装置，其特征在于，包括：日志采集模块、数据

清洗模块、特征提取模块、有效特征筛选模块，

其中，日志采集模块用于通过 flume 客户端采集服务器集群的日志，发送至 Hbase 数据库，其中，flume 客户端通过多个 Agent 进程对应采集服务器集群中的每台服务器的日志，Agent 定时将对应的服务器上的日志数据收集并通过 API 接口发送到 Hbase 数据库；

数据清洗模块用于利用 Hadoop 对日志数据进行数据清洗，筛选出原始数据，其中原始数据至少包括服务器磁盘占用率、内存使用率、cpu 占用率、业务接口调用量；

特征提取模块用于对原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取；

有效特征筛选模块运用皮尔逊相关系数筛选出有效特征：将提取的特征值分别与原始数据进行皮尔逊相关系数的运算，根据计算出的相关系数与相关度阈值进行比较，高于相关度阈值则是有效数据，低于相关度阈值则是无效数据，并予以剔除。

8.根据权利要求 7 所述的集群日志特征提取装置，其特征在于，数据清洗模块包括拉依达准则判定单元，所述拉依达准则判定单元采用拉依达准则剔除具有粗大误差的数据，包括以下步骤：

对日志数据 $x_1, x_2, \dots, x_n$ ，计算其算术平均值 $\bar{X}$ 及剩余误差 $v_i = x_i - \bar{X} (i=1, 2, \dots, n)$ ，其中， x_i 为单次 Agent 采集数据值；

计算标准偏差 S_x ，
$$S_x = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \right)^{1/2}；$$

若数据 x_b 的剩余误差 v_b ($1 \leq b \leq n$)，满足下式 $|v_b| = |x_b - \bar{X}| > 3S_x$

则认为 x_b 是含有粗大误差值的奇异值，并剔除所述奇异值。

9.根据权利要求 8 所述的集群日志特征提取装置，其特征在于，数据清洗模块还包括奇异值替换单元，所述奇异值替换单元对日志数据的奇异值用中值替代，其中所述中值是指将各个日志数据 $x_1, x_2, \dots, x_n$ 按大小顺序排列，处于中间位置的值称为中值。

10.根据权利要求 8 所述的集群日志特征提取装置，其特征在于，特征提取模块包括均值提取单元、有效值提取单元、峰值提取单元、方根幅值提取单元、波形指标提取单元、脉冲指标提取单元、峭度指标提取单元，分别对

原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取，其中，

有效值采用如下公式计算： $X_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$

峰值采用如下公式计算： $X_p = \max(x_i)$

方根幅值采用如下公式计算： $X_r = \left[\frac{1}{N} \sum_{i=1}^N \sqrt{|x_i|} \right]^2$

波形指标采用如下公式计算： $X_{ws} = \frac{X_{rms}}{|\bar{x}|}$

脉冲指标采用如下公式计算： $X_{if} = \frac{X_p}{|\bar{x}|}$

峭度指标采用如下公式计算： $X_{kv} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{X_{rms}^4}$

其中， x_i 为单次 Agent 采集的日志数据；N 为日志数据采集的次数； $\bar{x}$ 为采集的日志数据的算术平均值； X_{rms} 为采集的日志数据的有效值； X_p 为采集的日志数据的峰值； X_r 为采集的日志数据的方根幅值； X_{ws} 为采集的日志数据的波形指标； X_{if} 为采集的日志数据的脉冲指标； X_{kv} 为采集的日志数据的峭度指标。

11.根据权利要求 8 所述的集群日志特征提取装置，其特征在于，皮尔逊相关系数的公式如下：

$$\rho = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

其中， x_i 为单次 Agent 采集的日志数据； y_i 为单次 Agent 采集数据中提取的某一特征值； $\bar{x}$ 是日志数据 $x_1, x_2, \dots, x_n$ 的算数平均值； $\bar{y}$ 是 $y_1, y_2, \dots, y_n$ 的算数平均值；N 为日志数据采集的次数。

12.根据权利要求 7 所述的集群日志特征提取装置，其特征在于，日志采集模块还包括 Agent 设置单元，用于针对 Flume 进行包括多个第一层级 Agent

和一个第二层级 Agent 的设置,每个第一层级 Agent 分别对应的采集一个服务器的日志数据,多个第一层级 Agent 采集的日志数据汇集至第二层级 Agent,并由第二层级 Agent 传输至 HDFS 中。

13. 一种电子设备,其特征在于,该电子设备包括:存储器和处理器,所述存储器中存储有集群日志特征提取程序,所述集群日志特征提取程序被所述处理器执行时实现如下步骤:

通过 flume 客户端采集服务器集群的日志,发送至 Hbase 数据库,其中,flume 客户端通过多个 Agent 进程对应采集服务器集群中的每台服务器的日志,Agent 定时将对应的服务器上的日志数据收集并通过 API 接口发送到 Hbase 数据库;

利用 Hadoop 对日志数据进行数据清洗,筛选出原始数据,其中原始数据至少包括服务器磁盘占用率、内存使用率、cpu 占用率、业务接口调用量;

对原始数据进行包括均值、有效值、峰值、方根幅值、波形指标、脉冲指标、峭度指标的特征值提取;

运用皮尔逊相关系数筛选出有效特征:将提取的特征值分别与原始数据进行皮尔逊相关系数的运算,根据计算出的相关系数与相关度阈值进行比较,高于相关度阈值则是有效数据,低于相关度阈值则是无效数据,并予以剔除。

14. 根据权利要求 13 所述的电子设备,其特征在于,

数据清洗中采用拉依达准则剔除具有粗大误差的数据,包括以下步骤:

对日志数据 $x_1, x_2, \dots, x_n$, 计算其算术平均值 $\bar{X}$ 及剩余误差 $v_i = x_i - \bar{X} (i=1, 2, \dots, n)$, 其中, x_i 为单次 Agent 采集数据值;

计算标准偏差 S_x , $S_x = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \right)^{1/2}$;

若日志数据中的 x_b 的剩余误差 $v_b (1 \leq b \leq n)$, 满足公式 $|v_b| = |x_b - \bar{X}| > 3S_x$,

则认为 x_b 是含有粗大误差值的奇异值,并剔除所述奇异值。

15. 根据权利要求 14 所述的电子设备,其特征在于,

对日志数据中的奇异值用中值替代,其中所述中值是指将各个日志数据 $x_1, x_2, \dots, x_n$ 按大小顺序排列,处于中间位置的值称为中值。

16. 根据权利要求 13 所述的电子设备,其特征在于,Flume 包括多个第一层级 Agent 和一个第二层级 Agent,每个第一层级 Agent 分别对应的采集一个

服务器的日志数据，多个第一层级 Agent 采集的日志数据汇集至第二层级 Agent，并由第二层级 Agent 传输至 HDFS 中。

17. 一种计算机非易失性可读存储介质，其特征在于，所述计算机非易失性可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令被处理器执行时，实现权利要求 1 所述的集群日志特征提取方法。

18. 根据权利要求 17 所述的计算机非易失性可读存储介质，其特征在于，在数据清洗过程中，采用拉依达准则剔除具有粗大误差的数据，包括以下步骤：

对日志数据 $x_1, x_2, \dots, x_n$ ，计算其算术平均值 $\bar{X}$ 及剩余误差 $v_i = x_i - \bar{X} (i=1, 2, \dots, n)$ ，其中， x_i 为单次 Agent 采集的日志数据；

计算标准偏差 S_x ， $S_x = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \right)^{1/2}$ ；

若日志数据中的 x_b 的剩余误差 v_b ($1 \leq b \leq n$)，满足公式 $|v_b| = |x_b - \bar{X}| > 3S_x$

则确定 x_b 是含有粗大误差值的奇异值，并剔除奇异值。

19. 根据权利要求 18 所述的计算机非易失性可读存储介质，其特征在于，对日志数据的奇异值用中值替代，其中所述中值是指将各个日志数据 $x_1, x_2, \dots, x_n$ 按大小顺序排列，处于中间位置的值称为中值。

20. 根据权利要求 17 所述的计算机非易失性可读存储介质，其特征在于，Flume 包括多个第一层级 Agent 和一个第二层级 Agent，每个第一层级 Agent 分别对应的采集一个服务器的日志数据，多个第一层级 Agent 采集的日志数据汇集至第二层级 Agent，并由第二层级 Agent 传输至 HDFS 中。

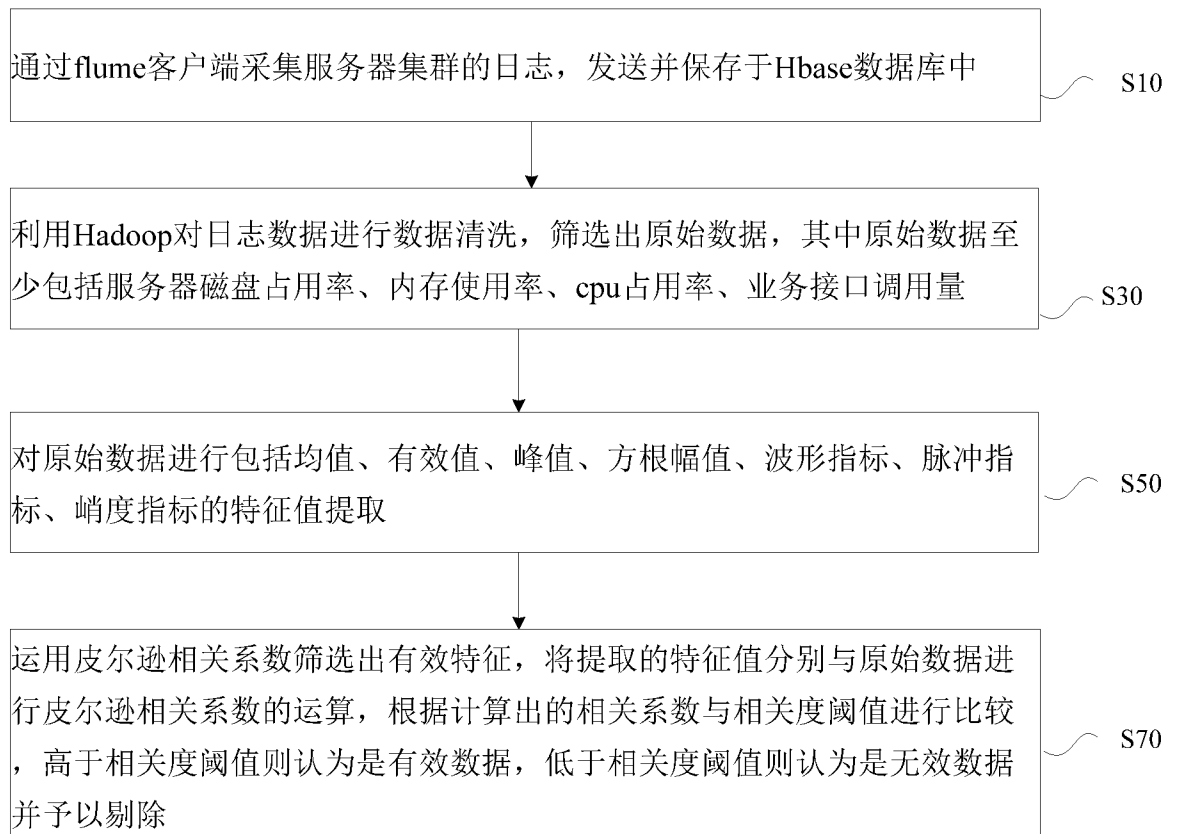


图 1

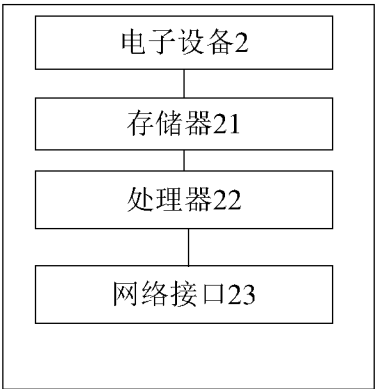


图 2

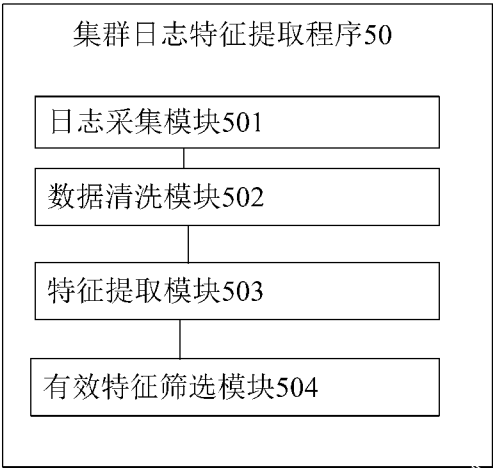


图 3



图 4

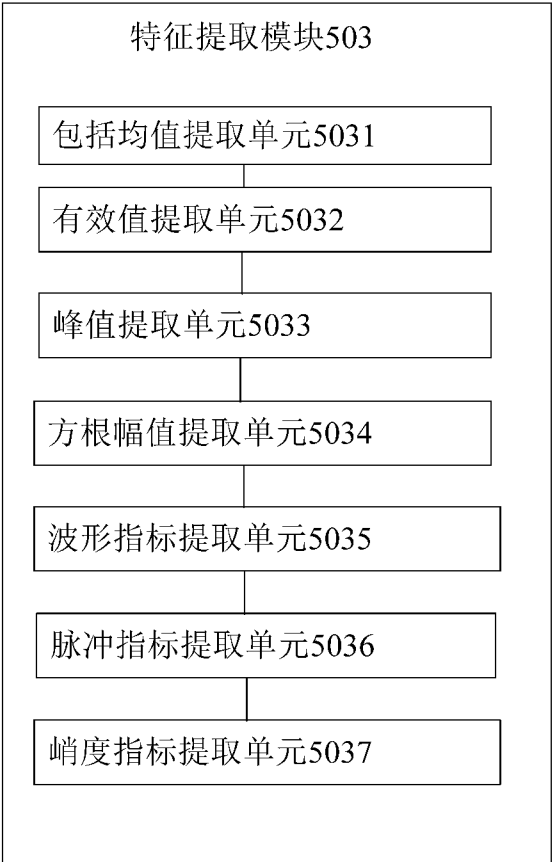


图 5

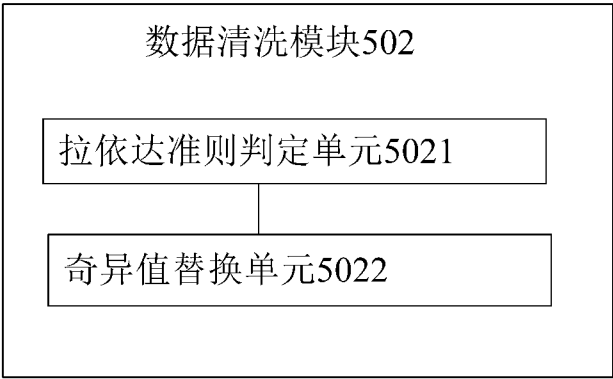


图 6

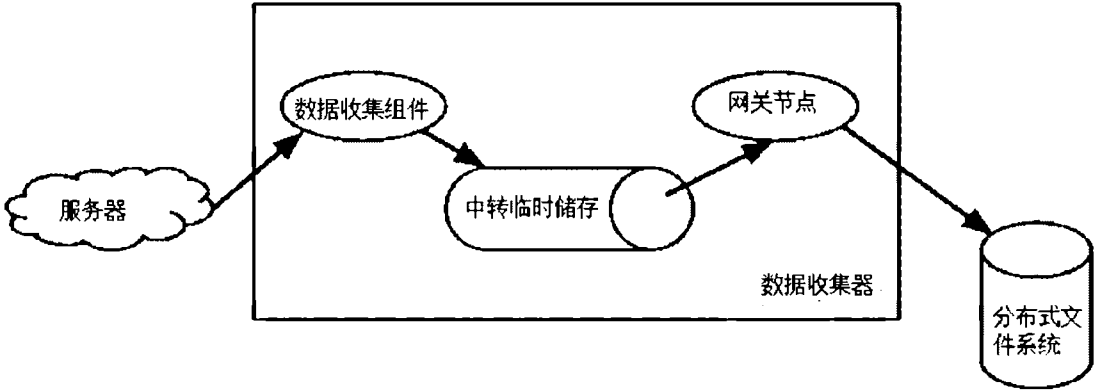


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/118288

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/17(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC, IEEE: 集群, 日志, 皮尔逊相关系数, 阈值, 剔除, 有效, 无效, 数据, 清洗, 峰值, 均值, 方根, flume, agent, hbase, hadoop, cluster, record, pearson correlation coefficient, threshold, eliminate, effective, invalid, data, cleanout, peak, mean value, root

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109992569 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 09 July 2019 (2019-07-09) claims 1-10	1-20
Y	CN 106570151 A (PHICOMM (SHANGHAI) CO., LTD.) 19 April 2017 (2017-04-19) description, paragraphs [0022]-[0033]	1-20
Y	CN 104036025 A (BLUEDON INF SAFETY TECH CO., LTD.) 10 September 2014 (2014-09-10) claim 3	1-20
Y	CN 106845799 A (CHINA ELECTRIC POWER RESEARCH INSTITUTE et al.) 13 June 2017 (2017-06-13) claims 5-7	1-20
A	CN 107092592 A (ZHEJIANG HONGCHENG COMPUTER SYSTEM CO., LTD.) 25 August 2017 (2017-08-25) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance
 “E” earlier application or patent but published on or after the international filing date
 “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 “O” document referring to an oral disclosure, use, exhibition or other means
 “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
 “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
 “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
 “&” document member of the same patent family

Date of the actual completion of the international search

18 January 2020

Date of mailing of the international search report

12 February 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/118288

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 7356550 B1 (TAIWAN SEMICONDUCTOR MANUFACTURING COMPANY) 08 April 2008 (2008-04-08) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/118288

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	109992569	A	09 July 2019	None	
CN	106570151	A	19 April 2017	None	
CN	104036025	A	10 September 2014	None	
CN	106845799	A	13 June 2017	None	
CN	107092592	A	25 August 2017	None	
US	7356550	B1	08 April 2008	None	

国际检索报告

国际申请号

PCT/CN2019/118288

A. 主题的分类 G06F 16/17 (2019.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																							
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06F 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNPAT, CNKI, WPI, EP0D0C, IEEE: 集群, 日志, 皮尔逊相关系数, 阈值, 剔除, 有效, 无效, 数据, 清洗, 峰值, 均值, 方根, flume, agent, hbase, hadoop, cluster, record, pearson correlation coefficient, threshold, eliminate, effective, invalid, data, cleanout, peak, mean value, root																							
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 109992569 A (平安科技深圳有限公司) 2019年 7月 9日 (2019 - 07 - 09) 权利要求1-10</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 106570151 A (上海斐讯数据通信技术有限公司) 2017年 4月 19日 (2017 - 04 - 19) 说明书第[0022]-[0033]段</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 104036025 A (蓝盾信息安全技术有限公司) 2014年 9月 10日 (2014 - 09 - 10) 权利要求3</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 106845799 A (中国电力科学研究院 等) 2017年 6月 13日 (2017 - 06 - 13) 权利要求5-7</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 107092592 A (浙江鸿程计算机系统有限公司) 2017年 8月 25日 (2017 - 08 - 25) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 7356550 B1 (TAIWAN SEMICONDUCTOR MANUFACTURING COMPANY) 2008年 4月 8日 (2008 - 04 - 08) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 109992569 A (平安科技深圳有限公司) 2019年 7月 9日 (2019 - 07 - 09) 权利要求1-10	1-20	Y	CN 106570151 A (上海斐讯数据通信技术有限公司) 2017年 4月 19日 (2017 - 04 - 19) 说明书第[0022]-[0033]段	1-20	Y	CN 104036025 A (蓝盾信息安全技术有限公司) 2014年 9月 10日 (2014 - 09 - 10) 权利要求3	1-20	Y	CN 106845799 A (中国电力科学研究院 等) 2017年 6月 13日 (2017 - 06 - 13) 权利要求5-7	1-20	A	CN 107092592 A (浙江鸿程计算机系统有限公司) 2017年 8月 25日 (2017 - 08 - 25) 全文	1-20	A	US 7356550 B1 (TAIWAN SEMICONDUCTOR MANUFACTURING COMPANY) 2008年 4月 8日 (2008 - 04 - 08) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 109992569 A (平安科技深圳有限公司) 2019年 7月 9日 (2019 - 07 - 09) 权利要求1-10	1-20																					
Y	CN 106570151 A (上海斐讯数据通信技术有限公司) 2017年 4月 19日 (2017 - 04 - 19) 说明书第[0022]-[0033]段	1-20																					
Y	CN 104036025 A (蓝盾信息安全技术有限公司) 2014年 9月 10日 (2014 - 09 - 10) 权利要求3	1-20																					
Y	CN 106845799 A (中国电力科学研究院 等) 2017年 6月 13日 (2017 - 06 - 13) 权利要求5-7	1-20																					
A	CN 107092592 A (浙江鸿程计算机系统有限公司) 2017年 8月 25日 (2017 - 08 - 25) 全文	1-20																					
A	US 7356550 B1 (TAIWAN SEMICONDUCTOR MANUFACTURING COMPANY) 2008年 4月 8日 (2008 - 04 - 08) 全文	1-20																					
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																			
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																						
国际检索实际完成的日期 2020年 1月 18日		国际检索报告邮寄日期 2020年 2月 12日																					
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 田民丽 电话号码 86-(10)-53961355																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/118288

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	109992569	A	2019年 7月 9日	无	
CN	106570151	A	2017年 4月 19日	无	
CN	104036025	A	2014年 9月 10日	无	
CN	106845799	A	2017年 6月 13日	无	
CN	107092592	A	2017年 8月 25日	无	
US	7356550	B1	2008年 4月 8日	无	