

(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

(43) 国际公布日
2017 年 1 月 19 日 (19.01.2017)



(10) 国际公布号
WO 2017/008658 A1

- (51) 国际专利分类号:
G06F 11/08 (2006.01)
- (21) 国际申请号: PCT/CN2016/088519
- (22) 国际申请日: 2016 年 7 月 5 日 (05.07.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510412446.X 2015 年 7 月 14 日 (14.07.2015) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛英属开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。
- (72) 发明人: 李强 (LI, Qiang); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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))。

(54) Title: STORAGE CHECKING METHOD AND SYSTEM FOR TEXT DATA

(54) 发明名称: 一种文本数据的存储校验方法和系统

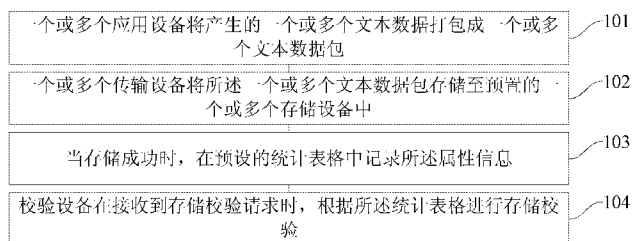


图 1

101 ONE OR MORE APPLICATION DEVICES PACKAGING ONE OR MORE PIECES OF GENERATED TEXT DATA INTO ONE OR MORE TEXT DATA PACKETS
102 ONE OR MORE TRANSMISSION DEVICES STORING ONE OR MORE TEXT DATA PACKETS IN ONE OR MORE PRE-SET STORAGE DEVICES
103 WHEN STORAGE SUCCEEDS, RECORDING THE ATTRIBUTE INFORMATION IN A PRE-SET STATISTICAL TABLE
104 WHEN A CHECKING DEVICE RECEIVES A STORAGE CHECKING REQUEST, PERFORMING STORAGE CHECKING ACCORDING TO THE STATISTICAL TABLE

(57) Abstract: Disclosed are a storage checking method and system for text data. The method comprises: one or more application devices packaging one or more pieces of generated text data into one or more text data packets (101), the text data packets comprising attribute information; one or more transmission devices storing one or more text data packets in one or more pre-set storage devices (102); when storage succeeds, recording the attribute information in a pre-set statistical table (103); and when a checking device receives a storage checking request, performing storage checking according to the statistical table (104). In the method, by packaging text data, the text data on which statistics need to be made are reduced by many times, so that a statistical quantity value is greatly reduced, thereby greatly reducing the checking processing capacity, reducing the performance consumption of a system, greatly improving the practicability during big data processing, and realizing the overall storage checking for big data in cloud.

(57) 摘要: 一种文本数据的存储校验方法和系统, 该方法包括: 一个或多个应用设备将产生的一个或多个文本数据打包成一个或多个文本数据包 (101); 所述文本数据包中具有属性信息; 一个或多个传输设备将所述一个或多个文本数据包存储至预置的一个或多个存储设备

中 (102); 当存储成功时, 在预设的统计表格中记录所述属性信息 (103); 校验设备在接收到存储校验请求时, 根据所述统计表格进行存储校验 (104)。该方法通过将文本数据进行打包, 将需要统计的文本数据缩小了很多倍, 进而大大减少了统计的量值, 进而大大降低了校验的处理量, 降低了系统的性能消耗, 大大增加了在大数据处理时的实用性, 实现了针对云中大数据的整体存储校验。

一种文本数据的存储校验方法和系统

本申请要求 2015 年 07 月 14 日递交的申请号为 201510412446.X、发明名称为“一种文本数据的存储校验方法和系统”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5

技术领域

本申请涉及计算机处理技术领域，特别是涉及一种文本数据的存储校验方法和一种文本数据的存储校验系统。

10 背景技术

随着云时代的来临，越来越多的平台在社交网络、电子商务、访问记录等来源中产生大数据，即量非常大的数据，例如，一天产生 100T 到 100P 之间，甚至更大的数据量，而生产这些数据的机器总数在 1 万到 100 万台之间，甚至更多。

云下的很多业务产生的数据，如网站产生的 pv（浏览量）日志，一般需要进行实时
15 存储，检验数据的完整性，保证数据挖掘等处理的准确性。

目前，虽然有某些系统提供了消息的校验机制，校验存储操作是否成功，但是，这种检测机制仅仅是局部的检测机制，无法应用于云时代大数据的存储校验。

发明内容

20 鉴于上述问题，提出了本申请实施例以便提供一种克服上述问题或者至少部分地解决上述问题的一种文本数据的存储校验方法和相应的一种文本数据的存储校验系统。

为了解决上述问题，本申请实施例公开了一种文本数据的存储校验方法，包括：

一个或多个应用设备将产生的一个或多个文本数据打包成一个或多个文本数据包；
所述文本数据包中具有属性信息；

25 一个或多个传输设备将所述一个或多个文本数据包存储至预置的一个或多个存储设备中，当存储成功时，在预设的统计表格中记录所述属性信息；

校验设备在接收到存储校验请求时，根据所述统计表格进行存储校验。

优选地，所述一个或多个应用设备将产生的一个或多个文本数据打包成一个或多个文本数据包的步骤包括：

30 当产生的文本数据的大小与预设的大小阈值匹配时，将产生的文本数据打包成文本

数据包；

或者，

在当前时间超过预设的时间阈值时，将产生的文本数据打包成文本数据包。

优选地，所述统计表格包括第一统计表格，所述第一统计表格包括存储时间；所述

5 传输设备具有传输设备标识；所述属性信息包括应用设备标识、产生时间；

所述在预设的统计表格中记录所述属性信息的步骤包括：

查找所述应用设备标识对应的第一统计表格；

判断是否具有未存储成功的特征文本数据包；所述特征文本数据包的产生时间小于当前存储成功的文本数据包的产生时间；

10 若不具有，则在所述第一统计表格中，将所述产生时间更新至所述传输设备标识和所述应用设备标识对应的存储时间中。

优选地，所述统计表格包括第二统计表格，所述第二统计表格包括分区时间段、存储行数；所述属性信息包括产生时间、包数据行数；

所述在预设的统计表格中记录所述属性信息的步骤包括：

15 在所述第二统计表格中，查找所述第一统计表格对应的、所述产生时间所属的分区时间段；

将所述包数据行数累加至所述分区时间段对应的存储行数中。

优选地，所述校验设备根据所述统计表格进行存储校验的步骤包括：

在所述应用设备标识对应的第一统计表格中，查找值最小的存储时间；

20 确认产生时间小于所述存储时间的文本数据包已存储完成。

优选地，所述校验设备根据所述统计表格进行存储校验的步骤包括：

统计落入指定的校对时间段的分区时间段的存储行数，获得第一总行数。

优选地，所述校验设备根据所述统计表格进行存储校验的步骤还包括：

从所述存储设备中读取在所述校对时间段中存储的文本数据包的第二总行数；

25 当所述第一总行数与所述第二总行数相等时，确认所述校对时间段对应的文本数据包未丢失；

当所述第一总行数与所述第二总行数不相等时，确认所述校对时间段对应的文本数据包至少部分丢失。

优选地，所述属性信息包括产生时间，所述存储设备中包括一个或多个存储分区；

30 所述一个或多个传输设备将所述一个或多个文本数据包存储至预置的一个或多个存

储设备中的步骤包括：

将所述文本数据包存储至在预置的存储设备中、所述产生时间对应的存储分区中。

本申请实施例还公开了一种文本数据的存储校验系统，所述系统包括一个或多个应用设备、一个或多个传输设备、一个或多个存储设备和校验设备；

5 其中，所述应用设备包括：

文本数据打包模块，用于将产生的一个或多个文本数据打包成一个或多个文本数据包；所述文本数据包中具有属性信息；

所述传输设备包括：

10 文本数据包存储模块，用于将所述一个或多个文本数据包存储至预置的一个或多个存储设备中；

属性信息记录模块，用于在存储成功时，在预设的统计表格中记录所述属性信息；

所述校验设备包括：

存储校验模块，用于在接收到存储校验请求时，根据所述统计表格进行存储校验。

优选地，所述文本数据打包模块包括：

15 第一打包子模块，用于在产生的文本数据的大小与预设的大小阈值匹配时，将产生的文本数据打包成文本数据包；

或者，

第二打包子模块，用于在当前时间超过预设的时间阈值时，将产生的文本数据打包成文本数据包。

20 优选地，所述统计表格包括第一统计表格，所述第一统计表格包括存储时间；所述传输设备具有传输设备标识；所述属性信息包括应用设备标识、产生时间；

所述属性信息记录模块包括：

表格查找子模块，用于查找所述应用设备标识对应的第一统计表格；

25 特征文本数据包判断子模块，用于判断是否具有未存储成功的特征文本数据包；若不具有，则调用时间更新子模块；所述特征文本数据包的产生时间小于当前存储成功的文本数据包的产生时间；

时间更新子模块，用于在所述第一统计表格中，将所述产生时间更新至所述传输设备标识和所述应用设备标识对应的存储时间中。

30 优选地，所述统计表格包括第二统计表格，所述第二统计表格包括分区时间段、存储行数；所述属性信息包括产生时间、包数据行数；

所述属性信息记录模块包括：

分区时间段查找子模块，用于在所述第二统计表格中，查找所述第一统计表格对应的、所述产生时间所属的分区时间段；

5 存储行数累加子模块，用于将所述包数据行数累加至所述分区时间段对应的存储行数中。

优选地，所述存储校验模块包括：

存储时间查找子模块，用于在所述应用设备标识对应的第一统计表格中，查找值最小的存储时间；

10 存储完成确认子模块，用于确认产生时间小于所述存储时间的文本数据包已存储完成。

优选地，所述存储校验模块包括：

存储行数统计子模块，用于统计落入指定的校对时间段的分区时间段的存储行数，获得第一总行数。

优选地，所述存储校验模块还包括：

15 第二总行数读取子模块，用于从所述存储设备中读取在所述校对时间段中存储的文本数据包的第二总行数；

未丢失确认子模块，用于在所述第一总行数与所述第二总行数相等时，确认所述校对时间段对应的文本数据包未丢失；

20 丢失确认子模块，用于在所述第一总行数与所述第二总行数不相等时，确认所述校对时间段对应的文本数据包至少部分丢失。

优选地，所述属性信息包括产生时间，所述存储设备中包括一个或多个存储分区；

所述文本数据包存储模块包括：

分区存储子模块，用于将所述文本数据包存储至在预置的存储设备中、所述产生时间对应的存储分区中。

25

本申请实施例包括以下优点：

本申请实施例的应用设备将产生的文本数据打包成文本数据包，由传输设备存储至存储设备中，在存储成功时，记录属性信息，由校验设备依据统计的属性信息按照存储校验对存储情况进行校验，通过将文本数据进行打包，将需要统计的文本数据缩小了很多倍，进而大大减少了统计的量值，进而大大降低了校验的处理量，降低了系统的性能

30

消耗，大大增加了在大数据处理时的实用性，实现了针对云中大数据的整体存储校验。

本申请实施例在第一统计表格中更新存储时间，通过对比各存储时间，得出值最小的存储时间，实现了海量的文本数据的存储持久化校验。

本申请实施例在第二统计表格中累积存储行数，通过累加所需分区时间段的存储行数，实现了海量的文本数据的存储数量校验。

本申请实施例通过对比基于传输设备统计的第一总行数和基于存储设备统计的第二总行数，实现了海量的文本数据的存储丢失校验。

附图说明

图 1 是本申请的一种文本数据的存储校验方法实施例的步骤流程图；

图 2 是本申请的一种文本数据的存储校验系统实施例的结构框图。

具体实施方式

为使本申请的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本申请作进一步详细的说明。

云平台可以向广大用户提供大数据的处理和存储能力，用户在云平台中的很多业务，如某网站的 pv（浏览量）日志，需要写入云平台网络侧设备。

假设该网站的前端设备为 1000 台，每台设备实时产生 access.log 日志，若将这 1000 台机器的 access.log 日志实时写入云平台网络侧设备中，现在业务对数据的质量要求已经越来越高，很多业务不允许丢失任何数据，这个时候需要有机制确保数据的完整性。

在存储时，一般需要确定日志一条不丢，同时确保所有设备在某一个时间前的数据都成功写入云平台网络侧设备，或者准确给出当前所有机器某一时刻前的 access.log 都被成功写入了云平台网络侧设备。

目前的方法中无法端到端的提供数据一条不丢的能力，例如，Flume 系统虽然提供了消息的 ack 机制，但这个 ack 机制仅仅是局部的确认机制，无法确认 Flume 系统本身成功了写入了多少条数据，无法确认在某个时间段的数据在存储时是否发生丢失，也无法判定最后在云平台网络侧设备上的整体数据的条数是多少。

如果写完云存储后，再进行 count 等操作，显然效率非常低。

因此，提出本申请实施例的核心构思之一，通过将数据进行打包，对数据包进行统

计，以便进行存储校验。

参照图 1，示出了本申请的一种文本数据的存储校验方法实施例的步骤流程图，具体可以包括如下步骤：

- 5 步骤 101，一个或多个应用设备将产生的一个或多个文本数据打包成一个或多个文本数据包；

需要说明的是，本申请可以应用于云平台中，即计算机集群，如分布式系统等。

以某个分布式系统为例，该分布式系统可以分为以下几部分：

- 10 分布式系统底层服务：提供分布式环境下所需要的协调服务、远程过程调用、安全管理和资源管理的服。这些底层服务为上层的分布式文件系统、任务调度等模块提供支持。

- 15 分布式文件系统：提供一个海量的、可靠的、可扩展的数据存储服务，将集群中各个节点的存储能力聚集起来，并能够自动屏蔽软硬件故障，为用户提供不间断的数据访问服务；支持增量扩容和数据的自动平衡，提供用户空间文件访问 API（Application Program Interface，应用程序编程接口），支持随机读写和追加写的操作。

任务调度：为集群系统中的任务提供调度服务，同时支持强调响应速度的在线服务（Online Service）和强调处理数据吞吐量的离线任务（Batch Processing Job）；自动检测系统中故障和热点，通过错误重试、针对长尾作业并发备份作业等方式，保证作业稳定可靠地完成。

- 20 集群监控和部署：对集群的状态和上层应用服务的运行状态和性能指标进行监控，对异常事件产生警报和记录；为运维人员提供整个分布式系统以及上层应用的部署和配置管理，支持在线集群扩容、缩容和应用服务的在线升级。

在云平台（如分布式系统）中，应用设备可以为在应用服务运行中可产生文本数据的设备，如服务器。

- 25 需要说明的是，该文本数据为时序数据，按照时间顺序生成，如 pv 日志、access.log 日志、系统运行日志等日志数据。

在具体实现中，应用设备可以通过若干种方式对文本数据进行打包：

在一种方式中，当产生的文本数据的大小与预设的大小阈值匹配时，将产生的文本数据打包成文本数据包；

- 30 在此方式中，可以基于数据量的维度进行打包。

应用设备将产生的文本数据按照大小阈值进行打包，则可以将需要统计的文本数据缩小了很多倍，进而大大减少了统计的量值，同时，文本数据包中还可以设置属性信息进行过滤等其他实时增值处理的打标操作。

例如，假设平均每条文本数据的大小为 1k，若一个文本数据包的阈值为 512K，即一个文本数据包的大小为 512K，若产生的文本数据为 512 亿条，应用在先的技术进行统计则需要统计 512 亿次，而经过打包后，512 亿条文本数据变成 1 亿个文本数据包，则需要统计的文本数据的量缩小达到 512 倍。

在另一种方式中，在当前时间超过预设的时间阈值时，将产生的文本数据打包成文本数据包。

在此方式中，可以基于时间的维度进行打包。

由于文本数据是时序数据，因此，在云端网络侧设备中也可以按时间分区存储。

例如，若文本数据按照小时分区，在云端网络侧设备中有 24 个分区，名字为 00、01……23，则 00:00:00~00:59:59 产生的文本数据存储于 00 分区，01:00:00~01:59:59 的文本数据存储于 01 分区，其他时间产生的文本数据的存储方式也类似。

经过打包后，文本数据包也可以按照产生时间存储于相应的分区中，一般情况下，需要保证文本数据包落入分区的正确性，比如，某个文本数据包的落入 00 分区，在该文本数据包中一般不包含 01:00:00~01:59:59 产生的文本数据，否则 01:00:00~01:59:59 产生的文本数据也会被放到 00 分区中，会造成文本数据的漂移（分区不准确），也是一种数据质量故障。

当然，文本数据的漂移通常不需要百分之百的避免，但不能幅度太大，若以 5 分钟等较短的时间阈值进行打包，可以有效防止文本数据的漂移，将写错分区的情况控制在一个可接受的误差范围内。

当然，上述打包方式可以同时使用，例如，假设大小阈值为 512K，时间阈值为 5 分钟，对 13:00:00-13:04:59 的文本数据进行打包，生成 3 个文本数据包，分别为 A1（第一条文本数据的产生的时间为 13:00:00，大小为 512K）、A2（大小为 512K）、A3（大小为 402K，最后一条文本数据的产生的时间为 13:04:59）。

此外，上述打包方式只是作为示例，在实施本申请实施例时，可以根据实际情况设置其他打包方式，本申请实施例对此不加以限制。另外，除了上述打包方式外，本领域技术人员还可以根据实际需要采用其它打包方式，本申请实施例对此也不加以限制。

需要说明的是，文本数据包可以进行压缩，以在网络传输中为了节省带宽，也可以

不进行压缩，本申请实施例对此不加以限制。

在文本数据打包成功时，可以对文本数据包配置属性信息，即文本数据包中具有属性信息。

在具体实现中，文本数据包的结构中除了文本数据本身之外，还可以设置称之为属
5 性的数据，而这些属性都有相应的名字，用于存储相应的属性信息，该属性信息可以包
括应用设备标识、产生时间、包数据行数。

其中，应用设备标识（HostName）为产生该文本数据包中文本数据的应用设备的标识，即一个唯一确定的应用设备的信息，如应用设备 ID、主机地址。

产生时间（FileTime）为产生该文本数据包中文本数据的时间，一般情况下，可以
10 以该文本数据包中第一条文本数据的产生的时间作为该文本数据包的产生时间。

包数据行数（LineCount）为文本数据包中所有文本数据的行数；若在数据库中，可以使用 `select count(1) from table_name` 指令来统计文本数据包中的包数据行数；若在分布式的数据库中或者分布式环境下，可以使用 Map Reduce 对文本数据进行条数扫描，累加至包数据行数。

15 在文本数据包打包成功时，则可以将该文本数据包发送至一个或多个传输设备。

需要说明的是，在云平台（如分布式系统）中，一般具有 ack 等机制保证传输的成功性，不发生丢包，若该文本数据包发送传输设备失败，则继续重新发送，直至发送成功。

步骤 102，一个或多个传输设备将所述一个或多个文本数据包存储至预置的一个或
20 多个存储设备中；

在云平台（如分布式系统）中，传输设备可以为传输数据（如文本数据包）至处理节点（如存储设备）的设备，存储设备可以为存储数据（如文本数据包）的设备。

在本申请实施例中，云平台提供了用于存储数据的 API（Application Program Interface，应用程序接口），由传输设备调用该 API，写入文本数据包。

25 在实际应用中，传输设备可以采用若干种分配策略为文本数据包分配存储设备，本申请实施例对此不加以限制。

例如，分配策略为哈希分配（ $\text{hash}(x)\%N$ ），即计算文本数据包的 hash（哈希）值，分配到 $\text{hash}(C)\%N$ 对应的存储设备上。

又例如，分配策略为随机分配，取一个随机数，然后将文本数据包分配到
30 $\text{random}(C)\%N$ 对应的存储设备上。

在本申请的一种优选实施例中，存储设备中可以包括一个或多个存储分区，每个存储分区可以存储某一段时间段的文本数据包，该时间段可以由本领域技术人员根据实际情况进行设定，如 1 小时、1 天等等，本申请实施例对此不加以限制。

因此，在存储时，可以查找文本数据包所属的存储分区，将该文本数据包存储至在
5 预置的存储设备中、该产生时间对应的存储分区中。

需要说明的是，在云平台（如分布式系统）中，一般具有 ack 等机制保证传输的成功性，不发生丢包，若该文本数据包发送存储设备失败，则继续重新发送，直至发送成功。

步骤 103，当存储成功时，在预设的统计表格中记录所述属性信息；

10 若文本数据包存储成功，则可以对该文本数据包进行属性信息的记录，已进行相应的存储校验。

在本申请的一种优选实施例中，所述统计表格可以包括第一统计表格，所述第一统计表格可以包括存储时间，则在本申请实施例中，步骤 103 可以包括如下子步骤：

子步骤 S11，查找所述应用设备标识对应的第一统计表格；

15 子步骤 S12，判断是否具有未存储成功的特征文本数据包；若不具有，则执行子步骤 S13；

其中，所述特征文本数据包的产生时间小于当前存储成功的文本数据包的产生时间；

子步骤 S13，在所述第一统计表格中，将所述产生时间更新至所述传输设备标识和所述应用设备标识对应的存储时间中。

20 在本申请实施例中，在云平台向广大用户提供大数据的处理和存储能力等情况中，用户可以租用云平台中的部分应用设备，即用户标识（如用户 ID）与应用设备标识具有关联关系，因此，同一用户的应用设备产生的文本数据包通常进行统一统计，不同用户具有不同的第一统计表格。

在第一统计表格中，同一传输设备标识和同一应用设备标识对应的存储时间一般有一个（即一对一关系）。

例如，用户甲分配的第一统计表格为 Test1，其应用设备包括应用设备 application_1、应用设备 application_2，传输设备包括传输设备 transmission_1、传输设备 transmission_2，则第一统计表格的示例可以如表 1 所示：

表 1

第一统计表格	传输机器标识	应用设备标识	存储时间
Test1	transmission_1	application_1	13:00:00
Test1	transmission_1	application_2	15:00:00
Test1	transmission_2	application_1	14:00:00
Test1	transmission_2	application_2	14:00:00

该存储时间是不断刷新的，可以表征已存储的，由某个传输设备传输的、某个应用设备所产生的文本数据包的最新时间。

假设，由传输设备 transmission_1 将应用设备 application_1 产生的某个文本数据包成功存储至存储设备，该文本数据包的产生时间为 13:00:03，则第一统计表格的修改示例可以如表 2 所示：

表 2

第一统计表格	传输设备标识	应用设备标识	存储时间
Test1	transmission_1	application_1	13:00:03
Test1	transmission_1	application_2	15:00:00
Test1	transmission_2	application_1	14:00:00
Test1	transmission_2	application_2	14:00:00

此外，为了保证文本数据包的时序性，可以在文件数据包存储成功时，检测是否具有未存储成功的特征文本数据包，若具有，则可以换成当前存储成功的文本数据包的产生时间，待检测到不具有未存储成功的特征文本数据包时，更新至存储时间。

例如，传输设备将上述文本数据包 A1、A2、A3 传输至存储设备，若 A1、A2 尚未存储成功，而 A3 已存储成功，则 A3 的产生时间不更新至第一统计表格中的存储时间，待 A1、A2 存储成功时，更新至第一统计表格中的存储时间。

在本申请的另一种优选实施例中，所述统计表格包括第二统计表格，所述第二统计表格包括分区时间段、存储行数，则在本申请实施例中，步骤 103 可以包括如下子步骤：

子步骤 S21，在所述第二统计表格中，查找所述第一统计表格对应的、所述产生时间所属的分区时间段；

子步骤 S23，将所述包数据行数累加至所述分区时间段对应的存储行数中。

在本申请实施例中，也可以对同一用户的应用设备产生的文本数据包进行统一统计。在第二统计表格中，第一表格对应的分区时间段（PartitionTime）、存储行数一般有

一个（即一对一关系），该分区时间段可以由本领域技术人根据实际情况进行设定，如 1 小时、15 分钟等等，存储行数为累计值，可以表征在该分区时间段已存储的某个用户的应用设备所产生的文本数据包的数据行数。

例如，用户甲分配的第一统计表格为 Test1，其分区时间段设置为 15 分钟，用户乙分配的第一统计表格为 Test2，其分区时间段设置为 10 分钟，则第二统计表格的示例可以表 3 所示：

表 3

第一统计表格	PartitionTime	lineCount
Test1	12:30:00	1500
Test1	12:45:00	1600
Test1	13:00:00	1300
.....		
Test2	13:00:00	1200
Test2	13:10:00	1100
Test2	13:20:00	1500

假设，当前某个文本数据包成功存储至存储设备，该文本数据包的产生时间为 13:00:03，属于 13:00:00 的分区时间段，行数为 50，则第二统计表格的修改示例可以如表 4 所示：

表 4

第一统计表格	PartitionTime	lineCount
Test1	12:30:00	1500
Test1	12:45:00	1600
Test1	13:00:00	1350
.....		
Test2	13:00:00	1200
Test2	13:10:00	1100
Test2	13:20:00	1500

步骤 104，校验设备在接收到存储校验请求时，根据所述统计表格进行存储校验。

在云平台（如分布式系统）中，校验设备可以为后端设备，其提供 API，让用户调

用存储校验请求，对该用户的应用设备产生的文本数据包的存储情况进行校验。

本申请实施例的应用设备将产生的文本数据打包成文本数据包，由传输设备存储至存储设备中，在存储成功时，记录属性信息，由校验设备依据统计的属性信息按照存储校验对存储情况进行校验，通过将文本数据进行打包，将需要统计的文本数据缩小了很多倍，进而大大减少了统计的量值，进而大大降低了校验的处理量，降低了系统的性能消耗，大大增加了在大数据处理时的实用性，实现了针对云的整体存储校验。

在本申请的一种优选实施例中，步骤 104 可以包括如下子步骤：

子步骤 S31，在所述应用设备标识对应的第一统计表格中，查找值最小的存储时间；

子步骤 S32，确认产生时间小于所述存储时间的文本数据包已存储完成。

在本申请实施例中，该存储校验请求可以用于校验哪个时间点之前的文本数据包存储完成（即持久化）。

该存储校验请求可以包括用户信息（如用户 ID）、第一统计表格标识、第一校验标识等参数。

其中，用户信息可以用于对存储校验请求进行鉴权，当鉴权通过时，允许进行存储校验。

第一统计表格标识为标识第一统计表格的信息，如名称、ID 等等。

第一校验标识为表示进行校验哪个时间点之前的文本数据包存储完成的信息。

在实际应用中，各传输设备（传输设备标识表征）存储成功的各应用设备（应用设备标识表征）产生的文件数据包的产生时间，都小于或等于值最小的存储时间，在值最小的存储时间与其他存储时间之间可能存在尚未存储的文件数据包，因此，可以以值最小的存储时间表征已完成存储的文本数据包时间点。

例如，如表 2 所示，值最小的存储时间为 13:00:03，transmission_1 存储成功的 application_1、application_2 产生的文件数据包的产生时间都小于或等于 13:00:03，transmission_1 可能存在产生时间在 13:00:03-15:00:00 的文本数据包尚未存储成功；

transmission_2 存储成功的 application_1、application_2 产生的文件数据包的产生时间都小于 13:00:03，transmission_1 可能存在产生时间在 13:00:03-14:00:00 的文本数据包尚未存储成功；

相对地，用户甲的应用设备在 13:00:03 之前产生文本数据包均已存储完成。

本申请实施例在第一统计表格中更新存储时间，通过比对各存储时间，得出值最小的存储时间，实现了海量的文本数据的存储持久化校验。

在本申请的一种优选实施例中，步骤 104 可以包括如下子步骤：

子步骤 S41，统计落入指定的校对时间段的分区时间段的存储行数，获得第一总行数；

在本申请实施例中，该存储校验请求可以用于校验哪个时间段之间的已存储的文本数据包 5 的行数。

该存储校验请求可以包括用户信息（如用户 ID）、第一统计表格标识、第二校验标识等参数、校验时间段。

其中，第二校验标识为表示进行校验哪个时间段之间已存储的文本数据包的行数的信息；

10 校对时间段用于统计该段时间已存储的文本数据包的行数。

在实际应用中，对落入指定的校对时间段的分区时间段的存储行数进行汇总，则可以获得该校验时间段的总的行数（即第一同行数）。

对于统计的第一总行数：

一、业务的有了直观的统计报表，对于业务的增值服务或者逻辑处理上提供直观的 15 数据依据；

二、确保数据的完整性提供期望数据。

文本数据落到云平台的存储设备上后，是否有没有受损，质量如何，可以通过统计第一总行数进行对比，来确定业务的所有文本数据落到云平台的存储设备的质量情况，如果质量不满足业务的需求，则需要纠正。

20 需要说明的是，进行行数统计的校验该段时间一般小于表征文本数据包存储完成的时间点，否则，统计尚未存储完成的时间段，可能导致统计失去意义。

例如，如表 4 所示，若用户甲需要对 12:30:00-13:00:00（校验时间段）之间文件数据包进行统计，则可以将分区时间段（PartitionTime）为 12:30:00、12:45:00 的存储行数（lineCount）进行汇总，统计出第一总行数为 3100（行）。

25 若用户甲需要对 12:30:00-13:15:00（校验时间段）之间文件数据包进行统计，由于表征文本数据包存储完成的时间点为 13:00:03，即在 13:00:03-13:15:00 之间可能存储尚未存储成功的文件数据包，所统计的第一总行数可能并非真实的行数。

本申请实施例在第二统计表格中累积存储行数，通过累加所需分区时间段的存储行数，实现了海量的文本数据的存储数量校验。

30 子步骤 S42，从所述存储设备中读取在所述校对时间段中存储的文本数据包的第二

总行数；

子步骤 S43，当所述第一总行数与所述第二总行数相等时，确认所述校对时间段对应的文本数据包未丢失；

子步骤 S44，当所述第一总行数与所述第二总行数不相等时，确认所述校对时间段
5 对应的文本数据包至少部分丢失。

在本申请实施例中，为了进一步校验在存储中是否发生丢失，可以将存储设备统计的第二总行数与传输设备统计的第一总行数进行比对，若两者相等，则可以表示未发生丢失，若两者不相等，则可以表示发生丢失。

本申请实施例通过对比基于传输设备统计的第一总行数和基于存储设备统计的第二
10 总行数，实现了海量的文本数据的存储丢失校验。

需要说明的是，对于方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请实施例并不受所描述的动作顺序的限制，因为依据本申请实施例，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员
15 也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作并不一定是本申请实施例所必须的。

参照图 2，示出了本申请的一种文本数据的存储校验系统实施例的结构框图，所述系统可以包括一个或多个应用设备 210、一个或多个传输设备 220、一个或多个存储设备
20 230 和校验设备 240；

其中，所述应用设备 210 可以包括如下模块：

文本数据打包模块 211，用于将产生的一个或多个文本数据打包成一个或多个文本数据包；所述文本数据包中具有属性信息；

所述传输设备 220 可以包括如下模块：

25 文本数据包存储模块 221，用于将所述一个或多个文本数据包存储至预置的一个或多个存储设备 230 中；

属性信息记录模块 222，用于在存储成功时，在预设的统计表格中记录所述属性信息；

所述校验设备 240 可以包括如下模块：

30 存储校验模块 241，用于在接收到存储校验请求时，根据所述统计表格进行存储校

验。

在本申请的一种优选实施例中，所述文本数据打包模块 211 可以包括如下子模块：

第一打包子模块，用于在产生的文本数据的大小与预设的大小阈值匹配时，将产生的文本数据打包成文本数据包；

5 或者，

第二打包子模块，用于在当前时间超过预设的时间阈值时，将产生的文本数据打包成文本数据包。

在本申请的一种优选实施例中，所述统计表格可以包括第一统计表格，所述第一统计表格可以包括存储时间；所述传输设备可以具有传输设备标识；所述属性信息可以包
10 括应用设备标识、产生时间；

所述属性信息记录模块 222 可以包括如下子模块：

表格查找子模块，用于查找所述应用设备标识对应的第一统计表格；

特征文本数据包判断子模块，用于判断是否具有未存储成功的特征文本数据包；若不具有，则调用时间更新子模块；所述特征文本数据包的产生时间小于当前存储成功的
15 文本数据包的产生时间；

时间更新子模块，用于在所述第一统计表格中，将所述产生时间更新至所述传输设备标识和所述应用设备标识对应的存储时间中。

在本申请的另一种优选实施例中，所述统计表格可以包括第二统计表格，所述第二统计表格可以包括分区时间段、存储行数；所述属性信息可以包括产生时间、包数据行
20 数；

所述属性信息记录模块 222 可以包括如下子模块：

分区时间段查找子模块，用于在所述第二统计表格中，查找所述第一统计表格对应的、所述产生时间所属的分区时间段；

存储行数累加子模块，用于将所述包数据行数累加至所述分区时间段对应的存储行
25 数中。

在本申请的一种优选实施例中，所述存储校验模块 241 可以包括如下子模块：

存储时间查找子模块，用于在所述应用设备标识对应的第一统计表格中，查找值最小的存储时间；

存储完成确认子模块，用于确认产生时间小于所述存储时间的文本数据包已存储完
30 成。

在本申请的另一种优选实施例中，所述存储校验模块 241 可以包括如下子模块：

存储行数统计子模块，用于统计落入指定的校对时间段的分区时间段的存储行数，获得第一总行数。

在本申请的另一种优选实施例中，所述存储校验模块 241 还可以包括如下子模块：

5 第二总行数读取子模块，用于从所述存储设备中读取在所述校对时间段中存储的文本数据包的第二总行数；

未丢失确认子模块，用于在所述第一总行数与所述第二总行数相等时，确认所述校对时间段对应的文本数据包未丢失；

10 丢失确认子模块，用于在所述第一总行数与所述第二总行数不相等时，确认所述校对时间段对应的文本数据包至少部分丢失。

在本申请的另一种优选实施例中，所述属性信息可以包括产生时间，所述存储设备 230 中可以包括一个或多个存储分区；

所述文本数据包存储模块 221 可以包括如下子模块：

15 分区存储子模块，用于将所述文本数据包存储至在预置的存储设备 230 中、所述产生时间对应的存储分区中。

对于系统实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

20 本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。

本领域内的技术人员应明白，本申请实施例的实施例可提供为方法、装置、或计算机程序产品。因此，本申请实施例可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请实施例可采用在一个或多个其中包含有计
25 算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

在一个典型的配置中，所述计算机设备包括一个或多个处理器 (CPU)、输入/输出接口、网络接口和内存。内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器(RAM) 和/或非易失性内存等形式，如只读存储器(ROM)或闪存(flash RAM)。内存
30 是计算机可读介质的示例。计算机可读介质包括永久性和非永久性、可移动和非可移动

媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括非持续性的电脑可读媒体(transitory media)，如调制的数据信号和载波。

本申请实施例是参照根据本申请实施例的方法、终端设备(系统)、和计算机程序产品的流程图和 / 或方框图来描述的。应理解可由计算机程序指令实现流程图和 / 或方框图中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理终端设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理终端设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理终端设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理终端设备上，使得在计算机或其他可编程终端设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程终端设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

尽管已描述了本申请实施例的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例做出另外的变更和修改。所以，所附权利要求意欲解释为包括优选实施例以及落入本申请实施例范围的所有变更和修改。

最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者

终端设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者终端设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者终端设备中还存在另外的相同要素。

5

以上对本申请所提供的一种文本数据的存储校验方法和一种文本数据的存储校验系统，进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

10

权 利 要 求 书

1、一种文本数据的存储校验方法，其特征在于，包括：

一个或多个应用设备将产生的一个或多个文本数据打包成一个或多个文本数据包；
所述文本数据包中具有属性信息；

5 一个或多个传输设备将所述一个或多个文本数据包存储至预置的一个或多个存储设备中，当存储成功时，在预设的统计表格中记录所述属性信息；

校验设备在接收到存储校验请求时，根据所述统计表格进行存储校验。

2、根据权利要求 1 所述的方法，其特征在于，所述一个或多个应用设备将产生的一个或多个文本数据打包成一个或多个文本数据包的步骤包括：

10 当产生的文本数据的大小与预设的大小阈值匹配时，将产生的文本数据打包成文本数据包；

或者，

在当前时间超过预设的时间阈值时，将产生的文本数据打包成文本数据包。

3、根据权利要求 1 或 2 所述的方法，其特征在于，所述统计表格包括第一统计表格，所述第一统计表格包括存储时间；所述传输设备具有传输设备标识；所述属性信息包括应用设备标识、产生时间；

所述在预设的统计表格中记录所述属性信息的步骤包括：

查找所述应用设备标识对应的第一统计表格；

15 判断是否具有未存储成功的特征文本数据包；所述特征文本数据包的产生时间小于当前存储成功的文本数据包的产生时间；

若不具有，则在所述第一统计表格中，将所述产生时间更新至所述传输设备标识和所述应用设备标识对应的存储时间中。

4、根据权利要求 1 或 2 所述的方法，其特征在于，所述统计表格包括第二统计表格，所述第二统计表格包括分区时间段、存储行数；所述属性信息包括产生时间、包数据行数；

所述在预设的统计表格中记录所述属性信息的步骤包括：

在所述第二统计表格中，查找所述第一统计表格对应的、所述产生时间所属的分区时间段；

将所述包数据行数累加至所述分区时间段对应的存储行数中。

30 5、根据权利要求 3 所述的方法，其特征在于，所述校验设备根据所述统计表格进

行存储校验的步骤包括：

在所述应用设备标识对应的第一统计表格中，查找值最小的存储时间；

确认产生时间小于所述存储时间的文本数据包已存储完成。

6、根据权利要求 4 所述的方法，其特征在于，所述校验设备根据所述统计表格进

5 行存储校验的步骤包括：

统计落入指定的校对时间段的分区时间段的存储行数，获得第一总行数。

7、根据权利要求 6 所述的方法，其特征在于，所述校验设备根据所述统计表格进行存储校验的步骤还包括：

从所述存储设备中读取在所述校对时间段中存储的文本数据包的第二总行数；

10 当所述第一总行数与所述第二总行数相等时，确认所述校对时间段对应的文本数据包未丢失；

当所述第一总行数与所述第二总行数不相等时，确认所述校对时间段对应的文本数据包至少部分丢失。

8、根据权利要求 1 或 2 或 5 或 6 或 7 所述的方法，其特征在于，所述属性信息包
15 括产生时间，所述存储设备中包括一个或多个存储分区；

所述一个或多个传输设备将所述一个或多个文本数据包存储至预置的一个或多个存储设备中的步骤包括：

将所述文本数据包存储至在预置的存储设备中、所述产生时间对应的存储分区中。

9、一种文本数据的存储校验系统，其特征在于，所述系统包括一个或多个应用设
20 备、一个或多个传输设备、一个或多个存储设备和校验设备；

其中，所述应用设备包括：

文本数据打包模块，用于将产生的一个或多个文本数据打包成一个或多个文本数据包；所述文本数据包中具有属性信息；

所述传输设备包括：

25 文本数据包存储模块，用于将所述一个或多个文本数据包存储至预置的一个或多个存储设备中；

属性信息记录模块，用于在存储成功时，在预设的统计表格中记录所述属性信息；

所述校验设备包括：

存储校验模块，用于在接收到存储校验请求时，根据所述统计表格进行存储校验。

30 10、根据权利要求 9 所述的系统，其特征在于，所述文本数据打包模块包括：

第一打包子模块，用于在产生的文本数据的大小与预设的大小阈值匹配时，将产生的文本数据打包成文本数据包；

或者，

第二打包子模块，用于在当前时间超过预设的时间阈值时，将产生的文本数据打包成文本数据包。

11、根据权利要求 9 所述的系统，其特征在于，所述统计表格包括第一统计表格，所述第一统计表格包括存储时间；所述传输设备具有传输设备标识；所述属性信息包括应用设备标识、产生时间；

所述属性信息记录模块包括：

10 表格查找子模块，用于查找所述应用设备标识对应的第一统计表格；

特征文本数据包判断子模块，用于判断是否具有未存储成功的特征文本数据包；若不具有，则调用时间更新子模块；所述特征文本数据包的产生时间小于当前存储成功的文本数据包的产生时间；

15 时间更新子模块，用于在所述第一统计表格中，将所述产生时间更新至所述传输设备标识和所述应用设备标识对应的存储时间中。

12、根据权利要求 9 所述的系统，其特征在于，所述统计表格包括第二统计表格，所述第二统计表格包括分区时间段、存储行数；所述属性信息包括产生时间、包数据行数；

所述属性信息记录模块包括：

20 分区时间段查找子模块，用于在所述第二统计表格中，查找所述第一统计表格对应的、所述产生时间所属的分区时间段；

存储行数累加子模块，用于将所述包数据行数累加至所述分区时间段对应的存储行数中。

13、根据权利要求 11 所述的系统，其特征在于，所述存储校验模块包括：

25 存储时间查找子模块，用于在所述应用设备标识对应的第一统计表格中，查找值最小的存储时间；

存储完成确认子模块，用于确认产生时间小于所述存储时间的文本数据包已存储完成。

14、根据权利要求 12 所述的系统，其特征在于，所述存储校验模块包括：

30 存储行数统计子模块，用于统计落入指定的校对时间段的分区时间段的存储行数，

获得第一总行数。

15、根据权利要求 14 所述的系统，其特征在于，所述存储校验模块还包括：

第二总行数读取子模块，用于从所述存储设备中读取在所述校对时间段中存储的文本数据包的第二总行数；

5 未丢失确认子模块，用于在所述第一总行数与所述第二总行数相等时，确认所述校对时间段对应的文本数据包未丢失；

丢失确认子模块，用于在所述第一总行数与所述第二总行数不相等时，确认所述校对时间段对应的文本数据包至少部分丢失。

16、根据权利要求 9 或 10 或 13 或 14 或 15 所述的系统，其特征在于，所述属性信息包括产生时间，所述存储设备中包括一个或多个存储分区；

所述文本数据包存储模块包括：

分区存储子模块，用于将所述文本数据包存储至在预置的存储设备中、所述产生时间对应的存储分区中。

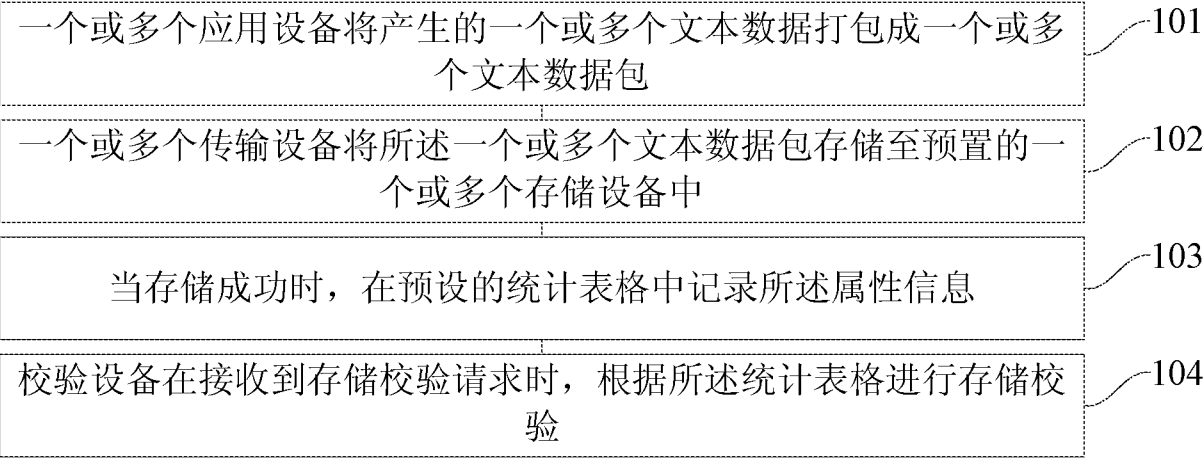


图 1

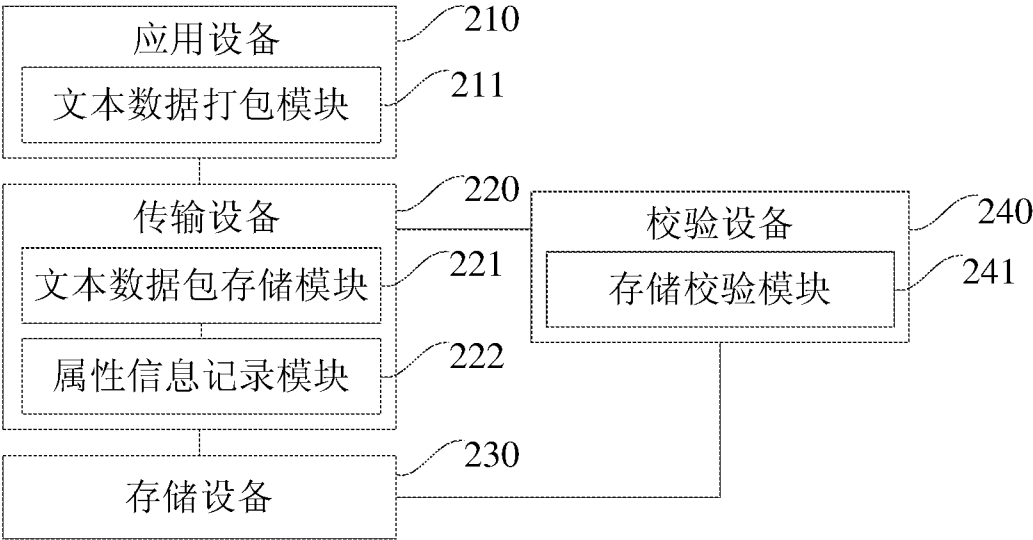


图 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/088519

A. CLASSIFICATION OF SUBJECT MATTER

G06F 11/08 (2006.01) 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)

WPI, EPODOC, CNKI, CNPAT, IEEE: text data, attribute, success, line number, file, data packet, size, time, verify, storage, integrity, lost, table, index, line

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 104202168 A (INSPUR ELECTRONIC INFORMATION INDUSTRY CO., LTD.), 10 December 2014 (10.12.2014), description, paragraphs [0006]-[0007]	1, 2, 8-10, 16
Y	CN 103401934 A (GUANGZHOU VIPSHOP INFORMATION TECHNOLOGY CO., LTD.), 20 November 2013 (20.11.2013), description, paragraph [0033]	1, 2, 8-10, 16
A	CN 103699851 A (HANGZHOU NORMAL UNIVERSITY), 02 April 2014 (02.04.2014), the whole document	1-16
A	CN 102413313 A (SONY CORPORATION), 11 April 2012 (11.04.2012), the whole document	1-16
A	CN 103268460 A (BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS), 28 August 2013 (28.08.2013), the whole document	1-16
A	US 8799334 B1 (EMC CORPORATION), 05 August 2014 (05.08.2014), the whole document	1-16

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 13 September 2016 (13.09.2016)	Date of mailing of the international search report 28 September 2016 (28.09.2016)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer MA, Chunli Telephone No.: (86-10) 62413712

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2016/088519

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104202168 A	10 December 2014	None	
CN 103401934 A	20 November 2013	None	
CN 103699851 A	02 April 2014	None	
CN 102413313 A	11 April 2012	None	
CN 103268460 A	28 August 2013	None	
US 8799334 B1	05 August 2014	US 9230114 B1	05 January 2016
		US 8346742 B1	01 January 2013

A. 主题的分类 G06F 11/08 (2006.01) 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)) WPI, EPDOC, CNKI, CNPAT, IEEE: 文本数据, 数据包, 属性, 大小, 时间, 校验, 存储, 完整, 成功, 丢失, 表格, 索引, 行数, file, data packet, size, time, verify, storage, integrity, lost, table, index, line																							
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>Y</td> <td>CN 104202168 A (浪潮电子信息产业股份有限公司) 2014年 12月 10日 (2014 - 12 - 10) 说明书第[0006]–[0007]段</td> <td>1, 2, 8–10, 16</td> </tr> <tr> <td>Y</td> <td>CN 103401934 A (广州唯品会信息科技有限公司) 2013年 11月 20日 (2013 - 11 - 20) 说明书第[0033]段</td> <td>1, 2, 8–10, 16</td> </tr> <tr> <td>A</td> <td>CN 103699851 A (杭州师范大学) 2014年 4月 2日 (2014 - 04 - 02) 全文</td> <td>1–16</td> </tr> <tr> <td>A</td> <td>CN 102413313 A (索尼公司) 2012年 4月 11日 (2012 - 04 - 11) 全文</td> <td>1–16</td> </tr> <tr> <td>A</td> <td>CN 103268460 A (北京航空航天大学) 2013年 8月 28日 (2013 - 08 - 28) 全文</td> <td>1–16</td> </tr> <tr> <td>A</td> <td>US 8799334 B1 (EMC CORPORATION) 2014年 8月 5日 (2014 - 08 - 05) 全文</td> <td>1–16</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	Y	CN 104202168 A (浪潮电子信息产业股份有限公司) 2014年 12月 10日 (2014 - 12 - 10) 说明书第[0006]–[0007]段	1, 2, 8–10, 16	Y	CN 103401934 A (广州唯品会信息科技有限公司) 2013年 11月 20日 (2013 - 11 - 20) 说明书第[0033]段	1, 2, 8–10, 16	A	CN 103699851 A (杭州师范大学) 2014年 4月 2日 (2014 - 04 - 02) 全文	1–16	A	CN 102413313 A (索尼公司) 2012年 4月 11日 (2012 - 04 - 11) 全文	1–16	A	CN 103268460 A (北京航空航天大学) 2013年 8月 28日 (2013 - 08 - 28) 全文	1–16	A	US 8799334 B1 (EMC CORPORATION) 2014年 8月 5日 (2014 - 08 - 05) 全文	1–16
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
Y	CN 104202168 A (浪潮电子信息产业股份有限公司) 2014年 12月 10日 (2014 - 12 - 10) 说明书第[0006]–[0007]段	1, 2, 8–10, 16																					
Y	CN 103401934 A (广州唯品会信息科技有限公司) 2013年 11月 20日 (2013 - 11 - 20) 说明书第[0033]段	1, 2, 8–10, 16																					
A	CN 103699851 A (杭州师范大学) 2014年 4月 2日 (2014 - 04 - 02) 全文	1–16																					
A	CN 102413313 A (索尼公司) 2012年 4月 11日 (2012 - 04 - 11) 全文	1–16																					
A	CN 103268460 A (北京航空航天大学) 2013年 8月 28日 (2013 - 08 - 28) 全文	1–16																					
A	US 8799334 B1 (EMC CORPORATION) 2014年 8月 5日 (2014 - 08 - 05) 全文	1–16																					
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
* 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 2016年 9月 13日		国际检索报告邮寄日期 2016年 9月 28日																					
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (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 马春黎 电话号码 (86-10) 62413712																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/088519

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	104202168	A	2014年 12月 10日	无	
CN	103401934	A	2013年 11月 20日	无	
CN	103699851	A	2014年 4月 2日	无	
CN	102413313	A	2012年 4月 11日	无	
CN	103268460	A	2013年 8月 28日	无	
US	8799334	B1	2014年 8月 5日	US 9230114 B1	2016年 1月 5日
				US 8346742 B1	2013年 1月 1日