



(12) 发明专利申请

(10) 申请公布号 CN 119149528 A

(43) 申请公布日 2024. 12. 17

(21) 申请号 202410751789.8

G06F 16/2455 (2019.01)

(22) 申请日 2024.06.12

G06F 16/28 (2019.01)

(30) 优先权数据

G06N 20/00 (2019.01)

18/334,595 2023.06.14 US

(71) 申请人 埃森哲环球解决方案有限公司

地址 爱尔兰都柏林

(72) 发明人 S·沙玛 R·马尔霍特拉

P·R·萨万特

J·阿布舍克·库玛 A·库玛

G·亚达夫

(74) 专利代理机构 北京市金杜律师事务所

11256

专利代理师 吴岩琨

(51) Int. Cl.

G06F 16/22 (2019.01)

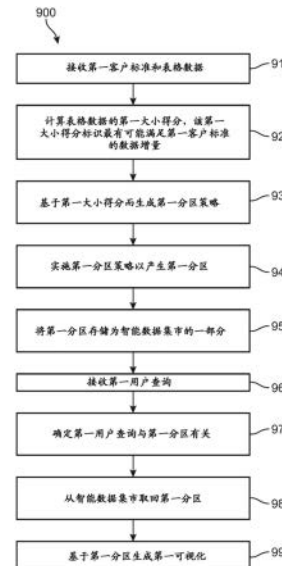
权利要求书4页 说明书15页 附图9页

(54) 发明名称

用于基于云的智能分析和可视化生成的系统和方法

(57) 摘要

本公开实施例涉及用于基于云的智能分析和可视化生成的系统和方法。一种优化用于数据存储和刷新事件的资源的数据处理和分析系统。用于数据分析系统的分区器模块可接收第一客户标准和包括表格数据的第一客户数据集并且计算用于生成分区策略的得分。所选的分区策略可被实施以产生可被存储在智能数据集市中的聚合数据。然后,数据可视化平台可访问这些分区,以智能、以便对用户的数据分析和可视化生成请求做出智能、动态的响应。通过提供数据(尤其是大数据)的同步划分和智能刷新,数据可以从后端移动至前端,用户点击次数最少并且性能延迟最小。



1. 一种用于为基于云计算的基础设施处理和对数据进行分区的计算机实施的方法,所述方法包括:

在针对数据分析系统的分区器模块处接收第一客户标准和包括表格数据的第一客户数据集;

在所述分区器模块处计算所述表格数据的第一大小得分,所述第一大小得分标识最有可能满足所述第一客户标准的第一数据增量;

在所述分区器模块处,基于所述第一大小得分而生成第一分区策略;

在所述数据分析系统处实施所述第一分区策略,以在所述表格数据中产生聚合数据的集合,所述集合包括第一分区;

将所述聚合数据的集合作为智能数据集的部分存储在所述数据分析系统中;

经由与数据可视化平台相关联的第一用户界面接收第一用户查询;

确定所述第一用户查询与所述第一分区有关;

经由所述数据可视化平台从所述智能数据集取回所述第一分区;以及

针对所述数据可视化平台并且经由生成式人工智能AI模块基于所述第一分区来生成第一可视化。

2. 根据权利要求1所述的方法,其中实施所述第一分区策略还包括:在基本上同步的过程中在所述表格数据中创建多个分区。

3. 根据权利要求1所述的方法,还包括:

接收指示针对第一类型数据的第一刷新率的第二客户标准;

通过对所述智能数据集的参考,确定仅所述第一分区的第一部分包括所述第一类型数据;以及

基于所述第二客户标准,以所述第一刷新率仅刷新所述第一分区的所述第一部分。

4. 根据权利要求3所述的方法,还包括:

接收指示针对不同于所述第一刷新率的第二类型数据的第二刷新率的第三客户标准;

通过对所述智能数据集的参考,确定仅所述第一分区的第二部分包括所述第二类型数据;以及

基于所述第三客户标准,以所述第二刷新率仅刷新所述第一分区的所述第二部分。

5. 根据权利要求1所述的方法,还包括:

在所述分区器模块处计算针对所述表格数据的第二大小得分,所述第二大小得分标识也可能满足所述第一客户标准的第二数据增量;

在所述分区器模块处,基于所述第二大小得分而生成第二分区策略;以及

经由针对所述数据分析系统的第二用户界面将所述第一分区策略呈现为第一选项以及将所述第二分区策略呈现为第二选项。

6. 根据权利要求5所述的方法,还包括:经由所述第二用户界面接收对所述第一选项的选择,并且其中实施所述第一分区策略是响应于接收到对所述第一选项的所述选择。

7. 根据权利要求1所述的方法,还包括:

在所述分区器模块处计算针对所述表格数据的第二大小得分,所述第二大小得分标识比所述第一数据增量更不可能满足所述第一客户标准的第二数据增量;

在所述分区器模块处,基于所述第二大小得分而生成第二分区策略;以及

经由针对数据分析系统的第二用户界面将所述第二分区策略呈现为拒绝的选项以供审查。

8. 一种非暂时性计算机可读介质,其存储软件,所述软件包括由一个或多个计算机可执行的指令,所述指令在执行时,使所述一个或多个计算机通过以下方式为基于云计算的基础设施处理和对数据进行分区:

在针对数据分析系统的分区器模块处接收第一客户标准和包括表格数据的第一客户数据集;

在所述分区器模块处计算针对所述表格数据的第一大小得分,所述第一大小得分标识最有可能满足所述第一客户标准的第一数据增量;

在所述分区器模块处,基于所述第一大小得分而生成第一分区策略;

在所述数据分析系统处实施所述第一分区策略以在所述表格数据中产生聚合数据的集合,所述集合包括第一分区;

将所述聚合数据的集合作为智能数据集的部分存储在所述数据分析系统中;

经由与数据可视化平台相关联的第一用户界面接收第一用户查询;

确定所述第一用户查询与所述第一分区有关;

经由所述数据可视化平台从所述智能数据集取回所述第一分区;以及

针对所述数据可视化平台并且经由生成式人工智能AI模块基于所述第一分区生成第一可视化。

9. 根据权利要求8所述的非暂时性计算机可读介质,其中所述指令还使所述一个或多个计算机以基本上同步的过程在所述表格数据中创建多个分区。

10. 根据权利要求8所述的非暂时性计算机可读介质,其中所述指令还使所述一个或多个计算机:

接收指示第一类型数据的第一刷新率的第二客户标准;

通过对所述智能数据集的参考,确定仅所述第一分区的第一部分包括所述第一类型数据;以及

基于所述第二客户标准,以所述第一刷新率仅刷新所述第一分区的所述第一部分。

11. 根据权利要求10所述的非暂时性计算机可读介质,其中所述指令还使所述一个或多个计算机:

接收指示针不同于所述第一刷新率的第二类型数据的第二刷新率的第三客户标准;

通过对所述智能数据集的参考,确定仅所述第一分区的第二部分包括所述第二类型数据;以及

基于所述第三客户标准,以所述第二刷新率仅刷新所述第一分区的所述第二部分。

12. 根据权利要求8所述的非暂时性计算机可读介质,其中所述指令还使所述一个或多个计算机:

在所述分区器模块处计算针对所述表格数据的第二大小得分,所述第二大小得分标识也可能满足所述第一客户标准的第二数据增量;

在所述分区器模块处,基于所述第二大小得分而生成第二分区策略;以及

经由针对所述数据分析系统的第二用户界面将所述第一分区策略呈现为第一选项以及将所述第二分区策略呈现为第二选项。

13. 根据权利要求12所述的非暂时性计算机可读介质,其中所述指令还使所述一个或多个计算机经由所述第二用户界面接收对所述第一选项的选择,并且其中实施所述第一分区策略是响应于接收到对所述第一选项的所述选择。

14. 根据权利要求8所述的非暂时性计算机可读介质,其中所述指令还使所述一个或多个计算机:

在所述分区器模块处计算针对所述表格数据的第二大得分,所述第二大得分标识比所述第一数据增量更不可能满足所述第一客户标准的第二数据增量;

在所述分区器模块处,基于所述第二大得分而生成第二分区策略;以及

经由针对数据分析系统的第二用户界面将所述第二分区策略呈现为拒绝的选项以供审查。

15. 一种用于为基于云计算的基础设施处理和对数据进行分区的系统,所述基础设施包括一个或多个计算机和存储指令的一个或多个存储设备,所述指令在被所述一个或多个计算机执行时可操作以使所述一个或多个计算机:

在针对数据分析系统的分区器模块处接收第一客户标准和包括表格数据的第一客户数据集;

在所述分区器模块处计算针对所述表格数据的第一大小得分,所述第一大小得分标识最有可能满足所述第一客户标准的第一数据增量;

在所述分区器模块处,基于所述第一大小得分而生成第一分区策略;

在所述数据分析系统处实施所述第一分区策略以在所述表格数据中产生聚合数据的集合,所述集合包括第一分区;

将所述聚合数据的集合作为智能数据集市的部分存储在所述数据分析系统中;

经由与数据可视化平台相关联的第一用户界面接收第一用户查询;

确定所述第一用户查询与所述第一分区有关;

经由所述数据可视化平台从所述智能数据集市取回所述第一分区;以及

针对所述数据可视化平台并且经由生成式人工智能AI模块基于所述第一分区生成第一可视化。

16. 根据权利要求15所述的系统,其中所述指令还使所述一个或多个计算机以基本上同步的过程在所述表格数据中创建多个分区。

17. 根据权利要求15所述的系统,其中所述指令还使所述一个或多个计算机:

接收指示针对第一类型数据的第一刷新率的第二客户标准;

通过对所述智能数据集市的参考,确定仅所述第一分区的第一部分包括所述第一类型数据;以及

基于所述第二客户标准,以所述第一刷新率仅刷新所述第一分区的所述第一部分。

18. 根据权利要求17所述的系统,其中所述指令还使所述一个或多个计算机:

接收指示针对不同于所述第一刷新率的第二类型数据的第二刷新率的第三客户标准;

通过对所述智能数据集市的参考,确定仅所述第一分区的第二部分包括所述第二类型数据;以及

基于所述第三客户标准,以所述第二刷新率仅刷新所述第一分区的所述第二部分。

19. 根据权利要求15所述的系统,其中所述指令还使所述一个或多个计算机:

在所述分区器模块处计算针对所述表格数据的第二大小得分,所述第二大小得分标识也可能满足所述第一客户标准的第二数据增量;

在所述分区器模块处,基于所述第二大小得分而生成第二分区策略;以及

经由针对所述数据分析系统的第二用户界面将所述第一分区策略呈现为第一选项以及将所述第二分区策略呈现为第二选项。

20.根据权利要求19所述的系统,其中所述指令还使所述一个或多个计算机经由所述第二用户界面接收对所述第一选项的选择,并且其中实施所述第一分区策略是响应于接收到对所述第一选项的所述选择。

用于基于云的智能分析和可视化生成的系统和方法

技术领域

[0001] 本公开总体上涉及云中复杂数据的分析,更具体地说,涉及大量数据的高效并且智能分区和实时可视化见解的生成。

背景技术

[0002] 数据分析是日益发展的领域,它从原始信息源中获取见解。数据分析技术可有所不同,但它们典型地用于收集、处理和分析数据以供人们消费和决策。数据分析可用于揭示因大量未分析数据而丢失或“不可见”的趋势和指标。然而,以这种方式使用该信息可优化过程并提高系统的整体效率,否则将无法获得该信息。

[0003] 传统上,许多可视化仪表板主要侧重于提供高级摘要,如细分概况、群组分析和其他类似指标。为了获得竞争优势,利益相关者(包括直线经理)寻求通过深入研究报告中的较低等级数据来揭示更深层次分析见解的工具。然而,在‘大数据’的上下文中(具有数万亿表格行和海量存储器),可视化报告生成变得滞后,尤其是在响应前端刷新请求时。这造成了不必要的高延迟仪表板并阻碍了对重要见解的访问。

[0004] 本领域需要一种解决上述缺点的系统和方法。

发明内容

[0005] 公开了一种优化用于数据存储和刷新事件的资源的数据处理和分析系统。所提出的系统和方法描述了用于生成分区策略的动态和自动化过程,该分区策略用于在云中管理和维护大数据的实施方式。该系统和方法通过自动评分和标识最适合给定项目的用户需求的最佳分区大小来解决上述问题。在一个示例中,来自多个源的数据可以被集成在云不可知和数据格式不可知的过程中,允许系统创建新的度量并且建立关系的集合,这些关系支持为可视化工具提供动力的表格语义数据模型的构建。在一些实施例中,所提出的数据分析框架采用高度优化和自动化的分区加载技术,该技术涉及将数据分割为更小的单元。此外,自定义数据刷新技术可有选择性地只刷新必要的分区,加上智能数据集市(“mart”)能力,随着冗余的消除,所需的计算时间大大减少。

[0006] 一方面,本公开为基于云计算的基础设施提供了一种处理和对数据进行分区的计算机实施方法。该方法包括在数据分析系统的分区器模块处接收第一客户标准和包括表格数据的第一客户数据集的第一步骤,以及在分区器模块处计算表格数据的第一大小得分的第二步骤,第一大小得分标识最有可能满足第一客户标准的第一数据增量。此外,第三步骤包括在分区器模块处基于第一大小得分而生成第一分区策略,并且第四步骤包括在数据分析系统处实施第一分区策略以产生表格数据中的聚合数据的集合,这些集合包括第一分区。该方法的第五步骤包括将集合数据作为智能数据集市的部分存储在数据分析系统中。第六步骤包括经由与数据可视化平台相关联的第一用户界面接收第一用户查询,并且第七步骤包括确定第一用户查询与第一分区有关。此外,该方法包括经由数据可视化平台从智能数据集市取回第一分区的第八步骤,以及为数据可视化平台并且经由生成人工智能(AI)

模块基于第一分区生成第一可视化的第九步骤。

[0007] 另一方面,本公开提供了一种存储软件的非暂时性计算机可读介质,该软件包括可由一个或多个计算机执行的指令,这些指令在执行时,促使一个或多个计算机通过执行以下各项为基于云计算的基础设施处理和对数据进行分区:(1)在数据分析系统的分区器模块处接收第一客户标准和包括表格数据的第一客户数据集;(2)在分区器模块处计算表格数据的第一大小得分,该第一大小得分标识最有可能满足第一客户标准的第一数据增量;(3)在分区器模块处基于第一大小得分而生成第一分区策略;(4)在数据分析系统处实施第一分区策略以产生表格数据中的聚合数据的集合,该集合包括第一分区;(5)将聚合数据的集合作为智能数据集的部分存储在数据分析系统中;(6)经由与数据可视化平台相关联的第一用户界面接收第一用户查询;(7)确定第一用户查询与第一分区有关;(8)经由数据可视化平台从智能数据集取回第一分区;以及(9)针对数据可视化平台并且经由生成式人工智能(AI)模块基于第一分区生成第一可视化。

[0008] 在又一方面,本公开提供了一种用于为基于云计算的基础设施生成应用程序架构的系统,该系统包括一个或多个计算机和存储指令的一个或多个存储设备,这些指令在被一个或多个计算机执行时可操作以使一个或多个计算机:(1)在数据分析系统的分区器模块处接收第一客户标准和包括表格数据的第一客户数据集;(2)在分区器模块处计算表格数据的第一大小得分,该第一大小得分标识最有可能满足第一客户标准的第一数据增量;(3)在分区器模块处基于第一大小得分而生成第一分区策略;(4)在数据分析系统处实施第一分区策略以产生表格数据中的聚合数据的集合,该集合包括第一分区;(5)将聚合数据的集合作为智能数据集的部分存储在数据分析系统中;(6)经由与数据可视化平台相关联的第一用户界面接收第一用户查询;(7)确定第一用户查询与第一分区有关;(8)经由数据可视化平台从智能数据集取回第一分区;以及(9)为数据可视化平台并且经由生成式人工智能(AI)模块基于第一分区生成第一可视化。

[0009] 本领域普通技术人员在研究以下附图和详细说明后,将会明白本公开的其他系统、方法、特征和优点。所有这类附加的系统、方法、特征和优点均旨在包括在本说明书和本公开内容中,包括在本公开的范围,并且受到所附权利要求的保护。

[0010] 虽然对各种实施例进行了描述,但该描述旨在进行示例性说明,而非限制性说明,对本领域普通技术人员而言,清楚的是,实施例范围内可能有更多实施例和实施方式。尽管在附图中示出了许多可能的特征组合,并在详细描述中进行了论述,但是所公开特征的许多其他组合也是可能的。除非特别定义,否则任何实施例的任何特征或元素可与任何其他实施例中的任何其他特征或元素结合使用或替代任何其他特征或元素。

[0011] 本公开包括并考虑了与本领域普通技术人员已知的特征和元素的组合。已公开的实施例、特征和元素也可与任何常规特征或元素组合,以形成权利要求所定义的独特发明。任何实施例的任何特征或元素也可与来自其他发明的特征或元素相结合,以形成如权利要求所定义的另一不同的发明。因此,将会理解,本公开中示出和/或论述的任何特征可单独地被实施或以任何合适的组合被实施。因此,除了根据所附权利要求及其等同物之外,实施例不受限制。此外,在所附权利要求的范围内可进行各种修改和改变。

附图说明

[0012] 参考以下附图和说明可更好地理解本发明。附图中的组件不一定按比例绘制,而是重点在于说明本发明的原理。此外,在附图中,相同的附图标记在不同的视图中表示相应的部分。

[0013] 图1是智能数据分区和分析过程的实施例的概述;

[0014] 图2是根据一个实施例的数据分析系统的高级视图;

[0015] 图3是描绘根据实施例的数据分析系统的智能数据集市的特征的示意图;

[0016] 图4是描绘根据实施例的数据分析系统的分区器模块的评分过程的示意图;

[0017] 图5呈现了根据实施例的可由分区器模块产生的分区策略生成的一些示例;

[0018] 图6是根据一个实施例的同步分区操作的示例;

[0019] 图7是描绘根据一个实施例的为用户仪表盘生成可视化的方法的示意流程图;

[0020] 图8为描绘可用于实施本文所描述的系统和/或方法的示例环境和组件的示意图;
以及

[0021] 图9是呈现根据一个实施例的处理和对数据进行分区的方法的流程图。

具体实施方式

[0022] 可视化仪表盘是数据分析的组成部分,尤其是在处理大量数据时。一般而言,仪表盘以持续聚合的较小数据的高级摘要形式呈现信息,提供关键见解的快速概述。然而,要让仪表盘发挥最大效用,它必须考虑到既能容纳更大范围的数据‘大图’又能容纳只能在粒度等级检测到的细微差别的见解。由于常规可视化仪表盘的设计和生依赖于汇总数据,因此基于汇总数据进行详细分析和探索粒度模型仍然具有挑战性。例如,常规的前端仪表盘不允许用户深入数据,只能生成肤浅的分析见解。此外,这类仪表盘无法接收越来越常见的大规模数据。由于传统的可视化和报告工具试图处理大数据以响应前端刷新操作,因此仍然存在频繁的滞后。此外,个人身份信息(PII)或受保护健康信息(PHI)等敏感客户数据,无法在没有暴露风险的情况下获取,因为数据必须通过导入模式或直接查询模式输入至报告工具中。

[0023] 尽管有这种上下文,但几乎每层业务决策均越来越有望得到数据的支持,在一些情况下,还有人工智能(AI)驱动的见解。迫切需要一种无代码的方法来进行数据分析和可视化仪表盘生成,以提供数据安全性并且使用户能够执行高效的分析,而无需依赖编码专业知识,从而释放有价值的数据驱动的见解,否则这些见解会在大规模数据存储体中丢失。

[0024] 所提出的系统和方法针对粒度级数据的自动化、智能化监管,以生成可利用云中托管的企业级数据模型的全面、高级、高质量仪表盘。如下面将更详细论述的,实施例可以创建智能数据集市,该智能数据集市以期望的粒度等级保存最相关的信息,并且以不同的聚合等级组织。来自多个来源的数据可被集成至与云无关和与数据格式无关的流程中,允许系统创建新的指标并且建立关系集合,这些关系支持为可视化工具提供动力的表格语义数据模型的构建。在一个示例中,所提出的分析框架采用了高度优化和自动化的分区加载技术,该技术涉及将数据分割为更小的单元。此外,随着冗余的消除,选择性地只刷新必要分区的自定义数据刷新技术以及智能数据集市功能显著减少了所需的计算时间。在一些实施例中,系统可以使用定制的企业数据模型战略性地摄取大数据,该模型智能地将数据分

割为较小的单元并且实现预定的定制自动刷新,而不会影响数据粒度和数据质量。这种方法已被证明可提供更快的处理速度并且提供安全的数据集成。此外,这类粒度级数据分析使企业能够在不依赖编码专业知识的情况下执行工具提示分析,从而为最终用户提供有效且高效的自助工具。所需见解的周转时间显著缩短,这提高了仪表板的可用性,并且为最终用户提供了来自最新数据的见解,促进及时做出明智的决策。此外,一些实施例可在产生仪表板时结合生成式人工智能(AI)技术,以另外减少(压缩)后端数据大小和处理复杂性。

[0025] 出于介绍目的,图1中描绘了所提出的数据分析可视化系统(“系统”)的实施的一个非限制性示例。在不同的实施例中,客户可能期望存储在其基于云的系统(诸如基于云的提取、转换和加载(ETL)和数据集成服务企业(“数据集成企业”))中的数据被利用并且准备用于使用所提出的实施例以进行分析。在图1中,数据可以由数据集成企业100沿自动管线移动和刷新至数据建模器,诸如在云中提供企业级数据模型的平台即服务(PaaS)。因此,所提出的系统可被集成至现有的数据集成企业100中,诸如但不限于Azure Data Factory®、AWS Glue®、IBM InfoSphere DataStage®、Matillion ETL®、Apache NiFi®、Talend Big Data Platform®、Pentaho Data Integration®、SnapLogic Intelligent Integration Platform(IIP)®、MuleSoft Anypoint Platform®等。

[0026] 例如,当数据被接收到数据集成企业100中时,系统在阶段A执行数据大小和速度的初始评估。第一模块110对接收到的数据执行数据转换和聚合并且在阶段B产生包括数据集市最终设计的第一输出。第一模块110可包括例如用于构建、部署、共享和维护规模企业级数据解决方案的工具,诸如但不限于Databricks®、MongoDB Atlas®、Oracle Database®、Amazon Redshift®、Spark®、Domino®、Confluent®、Amazon EMR®、Db2®、DataStax Enterprise®、Redis Enterprise Cloud®、CDP Data Hub®、Couchbase Server®等。

[0027] 在一些实施例中,最终设计作为一个或多个CSV文件输出。用于数据存储的第二模块120接收并且存储第一输出,并且还启动表格模型中的同步分区创建过程,以产生与客户的PaaS 160共享的第二输出(阶段C)。在不同的实施例中,第二模块120可以包括用于结构化和非结构化数据的基于云的储存库,诸如但不限于Azure Data Lake Storage Gen2®、Oracle Database®、Amazon Redshift®、Db2®、CDP Data Hub®、Databricks Lakehouse Platform®、EDB Postgres Advanced Server®等。

[0028] PaaS还可以包括第三模块130和第四模块140。在这种情况下,第二输出由第三模块130接收,该第三模块使用该数据来开发PaaS的模型,并且可以将该模型部署至云。在不同的实施例中,第三模块130可以包括可用于创建网站、web应用程序、web服务和移动应用程序的开发环境或代码编辑器工具,诸如但不限于Microsoft Visual Studio®、Xcode®、Eclipse®、NetBeans®、Android Studio®、Qt®、Firebase®、IntelliJ IDEA®、OutSystems®等。

[0029] 此外,在阶段D,对数据执行使用数据编排器的分区智能刷新过程,产生与第四模块140共享的第三输出。在不同的实施例中,第四模块140可以包括用于配置、管理和经营

SQL服务器实例的软件应用,诸如但不限于Microsoft SQL Server Management Studio®、MySQL WorkBench®、Teradata Vantage®、Oracle Database®、Amazon Relational Database Service (RDS)®、Db2®、PostgreSQL®、SAP HANA Cloud®、Toad For Oracle®等。在一些实施例中,智能刷新基于客户请求的刷新频率。因此,在系统的初始设计期间,客户可指示数据刷新和分区更新应当发生的频率或时间间隔,并且该偏好将被实施至系统中,以允许自动刷新发生以匹配客户的偏好。在其他情况下,客户可指示他们对于一种类型的数据报告的第一刷新率(例如,每周刷新)以及对于另一种类型的数据报告的不同刷新率(例如,每月刷新)的偏好,并且系统可在智能数据中创建以不同刷新周期率工作的两种不同类型的数据表。在分区被系统确定为在客户的期望刷新率下太重或太大的情况下(基于内部分区评分机制),系统可自动将刷新率转换至较低等级(更频繁的刷新率)以确保分区得分允许数据分析无延迟地发生。因此,在一些实施例中,第四模块140可插入分区,刷新表格,并且能够访问用于自助服务和企业商业智能(BI)和数据可视化的所选平台的数据,此处被示出为第五模块150,该第五模块可用于基于数据而生成关键见解。在不同的实施例中,第五模块150可以包括用于整理、管理和分析来自各种来源的数据的工具,诸如PowerBI®、Datapad®、Tableau Desktop®、Sisense®、QlikSense®、Zoho Analytics®、Cyfe®、Holistics®、Ubiq®和IMP Cognos Dashboard®等。

[0030] 现在转至图2,图示了所提出的数据分析系统200的架构的实施例,以向读者提供上下文。架构200可被理解为提供通过系统的高级关键组件的简化流程。最初,接收要被分析的客户数据210,并且针对客户210做出第一决定220,以确定是否应当实施所推荐的系统来进行分析。如果数据不需要实时分析,也不是计算量大的(大数据),则可使用传统的数据加载技术来生成充分覆盖客户需求的可视化仪表盘270。然而,如果客户需要实时大数据分析,则客户数据210可以被移动至系统的第一阶段,其中智能数据集市模块230开始处理客户数据210并且产生基于CSV的最终分区。智能数据集市模块230的操作将在下面参考图3、图4和图5另外论述。

[0031] 智能数据集市模块230的输出随后被传送至同步分区创建模块240,该模块使用逻辑序列分区将优化的分区策略应用于数据。一般而言,分区将表分割为多个逻辑部分。分区的工作原理是将表分割为多个逻辑分区对象。单独的分区,每个分区均含有唯一的数据段,然后可独立于其他分区顺序或并行地进行增量处理,或者完全排除在处理操作之外。因此,分区将需要频繁处理(刷新)的数据部分与处理频率较低的数据部分分割。换句话说,每个分区均可以独立于其他分区进行处理(刷新)。在许多情况下,诸如对于表格数据,将表的单个分区分割为多个分区可更好地利用可用资源进行处理。如将在本文所描述的,所提出的系统实现了有效的模型设计和处理策略,该策略利用分区来消除不必要的处理器负载和内存消耗,同时确保数据被足够频繁地刷新以反映来自数据源的最新数据。在模型创作期间为模型限定的分区在已部署的模型中是重复的。

[0032] 在一些实施例中,可以将分区数据传送至刷新分区模块250,该模块采用数据编排来刷新每个分区。刷新的分区可以被提供给调度刷新模块260,该模块确保分区被维持在它们指定的刷新率(依据客户的请求)。因此,调度刷新模块260可以自动使分区按照每个数据流的指定调度刷新。该最新数据然后被系统用来生成期望的可视化仪表盘270。可视化仪表

板270可以基于同步分区而更新,该同步分区减少了需要通过智能过滤来搜索的数据量,该智能过滤仅搜索与所需见解相关的数据单元(分区)。此外,刷新也可以更有效,因为只需要从较大的数据集中刷新相关分区,从而减少处理时间。在不同的实施例中,可视化仪表盘270为终端用户提供了使用独特的提示命令基于粒度数据而创建支持自助服务的视觉效果的能力,在一些情况下通过结合生成式AI来创建定制视觉呈现。

[0033] 应当理解,系统智能数据集市的创建和可用性实现了在多个等级将大数据自动分区为多维数据单元的独特功能性。因此,智能数据集市支持从数据仓库层到可视化层的无缝数据连接件。在一些实施例中,系统通过构建分区的逻辑序列来利用后端的表格模型的创建,这进而有助于促进分区等级的同步调度刷新事件。这类方法可确保在通过前端创建用户查询时,系统不需要参考大数据表,而是在后端将数据分解为更小的单元(通过智能数据集市中的创新分区),从而实现无缓冲取回过程并且显著加快可视化生成的响应时间。

[0034] 现在参考图3、图4和图5,提供了关于智能数据集市模块230所执行的分析的特征的附加细节。在图3中,示出了智能数据集市模块230与客户的现有数据生态系统(“现有数据”)300的示例的集成。在不同的实施例中,现有数据300可以包括客户的现有数据语义层。在一个示例中,这类现有数据300可以驻留在具有常规一致刷新机制的遗留系统或云数据湖中,并且具有或不具有数据网格架构。应当理解,所提出的数据分析系统不需要客户生态系统的任何特定格式或模板,因为它可以容易地跨各种各样的现有语义数据源连接或集成(例如,该系统是格式不可知的)。

[0035] 在一些实施例中,现有数据300可以包括语义层,该语义层具有可以被馈送至标准数据集市的多个维度和表格。一般而言,数据集市是指面向主题的数据库,它通常是企业数据仓库的分区部分。数据集中保存的数据子集典型地与某个特定的单位相关联,如制造、销售等。数据集市可以通过允许在更短的时间内访问数据仓库中的相关信息来加速流程,从而形成一种经济高效的方式来快速获得可操作的见解。将数据从(多个)源数据库传输值一个或多个目的数据库(多个)/数据仓库(多个)可以涉及提取转换加载(ETL)过程,该过程负责数据的每次移动和处理,以将数据以集成的形式加载至目的数据库/数据仓库。

[0036] 在一个示例中,现有数据300可以被接收到较大数据分析系统200的初始处理生态系统390中。在第一阶段310,机器学习算法便于收集附加的客户输入,这些输入将部分用于确定如何组织数据。在一些实施例中,客户可以描述他们期望查看刷新(更新)数据的频率/频度,例如每月一次、每周一次、每天一次等。在一个实施例中,客户可以为每个等级的数据选择刷新率。因此,客户输入将用于确定诸如最近的详细信息的期望等级粒度:每月、每周、每天、每小时等,以及报告中所需的其他特征

[0037] 现有数据300和相关客户输入随后被传送至两个单独的模块,包括分区器模块330和主标记模块340,其可以自动准备用于构建智能数据集市的数据。在一个示例中,分区器模块330和主标记模块340用作摄取客户数据并且生成数据中的各种特征或项目的概述的工具,诸如唯一值和/或多列的计数(例如,Microsoft PowerBI®中的distinctcount等),以及主关键字分析,以确定表中符合条件并且适合作为表的主关键字的一列或多列,详细说明可行的分区选项。

[0038] 在一些实施例中,分区器模块330包括算法,算法被配置成基于客户输入和数据的准确性,跨多个数据层级提出一个或多个分割策略。从这些所提出的策略中,客户可以选择

适合他们需求的策略,或者可根据客户输入中指示的偏好自动进行选择。转至图4,论述了关于分区器模块330的操作的附加细节。一般而言,客户数据410(例如,图3的现有数据300)可由分区器模块330接收,并且基于所涉及的表的类型沿两条路径转移。因此,基于事实的数据420可以沿第一路径路由,而基于DIM的数据430可以沿第二路径路由。

[0039] 一般而言,基于事实的数据420包括事实表,并且基于DIM的数据430包括维度表。对于读者来说,这两个可以被理解为数据仓库中模式的关键组件。事实表含有组合了不同维度表的属性的记录件。这些记录件允许用户分析其业务的不同方面,这有助于决策和改进业务。维度表为事实表中记录的度量提供上下文和上下文信息。因此,事实表和维度表的区别在于维度表含有事实表中的度量所基于的属性。此外,当使用多个事实表时,可以使用“事实星座(constellation)模式”来组织它们。事实表典型地有两种类型的列:含有事实的列和用作链接至维度表的外关键字的列。事实表的主关键字通常是由表中所有外关键字组成的组合关键字。每个事实表均有关键字,该关键字由连接至该事实表的所有维度表的主关键字构成。串联关键字就是这样一种专门标识事实表行的关键字。事实表可以保存各种类型的度量值,诸如可加性度量值、非可加性度量值和部分可加性度量值等,并且在数据仓库中存储重要信息。因此,事实表对于评估维度属性非常有用,因为它们提供了可作为独立变量的附加值。事实表的粒度是描述其层的常用方法。事实表的粒度是指可以限定事实的最基本等级。例如,销售事实表的粒度可能是“商店每天的副产品销售量”,这意指表中的每个条目均由日期、产品和商店唯一标识。相反,术语“维度表(dimension table)”是指与任何可量化事件相关的一组数据,是维度建模的基础。维度表有一个充当主关键字的列,允许唯一标识每个维度行或记录件。该关键字用于将维度表链接至事实表。代理关键字是系统生成的关键字,通常用于唯一标识维度表中的行。

[0040] 在不同的实施例中,出于分区器模块330的目的,仅基于事实的数据420用于生成加权得分。如图4所示,基于事实的数据420连同指定的偏好470(例如,标识用于分区的潜在列的列表以及等待时间的权重)由评分子模块490摄取。在第一阶段440中,计算第一大得分。如果提供了列列表,则通过检查所有组合的CSV文件大小来计算大小得分。如果未提供列列表,则通过检查分类列中的组合来计算第一大得分。在第二阶段450中,通过检查由强力创建的分区的集合中的每个分区覆盖的数字列的数目,然后另外计算聚合的最小、最大和标准偏差来计算第二大得分。这些得分将共同用于生成加权分区器得分460,该得分可以用于评估分区策略。换句话说,分区器模块330可以确定主关键字、排序列等的所有可能组合。并且基于这些组合,计算所有不同组合的不同长度的数据的序列集,并且为每个序列集分配评分。客户以基于得分最高的列的组合来选择是在存储大小方面进行优化还是在同步加载方面进行优化。

[0041] 为了清晰起见,图5中描绘了可经由分区器模块330生成的策略选项的非限制性示例。如上所述,分区器模块330可以接收客户数据,诸如输入数据概述510。在这种情况下,输入数据概述510包括描述较大数据集的两个表格。在涉及数据诊断技术的初始筛选520之后,系统可以标识所有可行分区。在本示例中,阴影框表示基于客户所需的详细程度被标识为最可行或最佳分区的数据类别。然后对筛选的数据进行评分(例如,参见图4)。得分应用于数据以生成不同的多级分区,每个分区表示策略,该策略寻求优化分区数目和每个分区中的数据。

[0042] 在该示例中,分区策略560示出了第一推荐,该第一推荐包括在商店级分区的第一选项和在日等级分区的第二选项。在第二推荐中,第一个选项是在子类别等级进行分区,并且第二个选项是在日等级进行分区。在不同的实施例中,基于客户的偏好可以包括附加等级。此外,在一些实施例中,还可以呈现被拒绝的选项570,这些选项基于所计算的得分被认为对于所提供的设定或云结构不可行。应当理解,最终评分和策略选项不需要人工干预。在一个实施例中,然后将具有最高得分的分区策略选项标识为最有可能向用户提供期望的数据粒度同时维持效率的顶级策略。利用这些基于客户需求的优化分区策略选项,系统可以更容易地开发和构建智能数据集市。

[0043] 如之前图3中所述,与分区器模块330一起,现有数据也由主标记模块340处理。一般而言,主标记模块340实施基于适合于各种等级的主标记的多个选项生成得分的算法,用户可以选择一个算法并且继续,或者系统可以自动地选择最高得分并且继续。在不同的实施例中,主标记模块340可分析现有数据并且标识客户的主标记(或主关键字)。简单地为了说明的目的,下面示出了主标记模块340可以用来聚合数据的过程的一个示例。在此示例中,接收到由表1表示的第一数据的集合,用于两个客户(a1和a2)在不同日期的商店访问以及他们的消费历史。

[0044]

客户	日期	商店	消费
a1	4/4/2023	s1	26
a1	3/17/2023	s1	3
a1	3/9/2023	s4	99
a2	4/3/2023	s2	24
a2	2/27/2023	s2	96
a2	3/26/2023	s5	29

[0045] 表1

[0046] 主标记模块340可以基于客户的商店应用主标记,并且输出列出相同信息的新表(表2),但在这种情况下,商店数据已被更改为仅标识每个客户的主商店(客户a1的S1和客户a2的S2)。

[0047]

客户	日期	主商店	消费
a1	4/4/2023	S1	26
a1	3/17/2023	S1	3
a1	3/9/2023	S1	99
a2	4/3/2023	S2	24
a2	2/27/2023	S2	96
a2	3/26/2023	S2	29

[0048]

[0049] 表2

[0050] 然后,该表数据可以由主标记模块340通过参考主商店进行汇总,以生成第三表(表3),该第三表允许数据大小的显著减少(在这种情况下,大约40-50%)。

[0051]

客户	主商店	消费
----	-----	----

a1	S1	128
a2	S2	149

[0052] 表3

[0053] 因此,在不同的实施例,系统的主标记特征能够灵活地在可视化层执行快速聚合。然后构建的智能数据集市基于分区器模块330和主标记模块340的输出,并且由此可以集成来自广泛且不同范围的大数据源(例如,以十亿和万亿计的行)的数据,以实现粒度(例如,客户、商店、邮政编码等)等级的实时见解。开发的智能数据集市还支持分区的逻辑排序以及动态数据的同步刷新,可完全支持仪表板或报告层,并且数据跨更多等级分布和分类。此外,智能数据集市可以用于生成不同粒度等级的不同类型的报告,同时还支持单元服务功能性和即席分析。因此,如果客户期望重新设计其报告的参数,而不是返回至原始数据,他们可以直接访问智能数据集市并且运行简单的查询来取回所需的见解,因为智能数据集市充当了极大地简化所有查询的单一真实来源。智能数据集市可以维护和表达客户预先存在的数据,并且继续执行刷新以确保数据是最新的。为了平衡频繁刷新的需求和处理效率,系统可以基于已经使用分区器和主标记聚合的数据而执行刷新,使得处理的数据量减少,同时像所有关键性能指标(KPI)这样的大部分见解保持完整和准确。在一些实施例中,使用所提出的系统的自动数据刷新将仅刷新(重写和归档)已经变得过时的特定分区,并且用新的分区数据替换它(而不是每当有新的增量数据加载至后端时低效地刷新所有分区)。

[0054] 在不同的实施例中,智能数据集市建成后,数据分析系统可以基于所选的分区策略而自动地开始分区过程,从而将客户数据加载至分区中。如先前在图2中介绍的,数据分析系统200可以使用同步分区创建模块240和刷新分区模块250来创建和维护分区的逻辑序列作为表格模型。为了比较和更好地说明拟议系统的有效性,下面提供了两个表格。在表4中,标识了用于数据集的常规手动分区创建方法的一些特征,而在表5中,示出了用于相同数据集的所提出的系统的同步分区创建方法的特征。

	在 SSMS 手动创建分区
	1 分区加载工作
	脚本中每个分区的 4 个变化。 在 SSMS 中再次加载脚本。 然后等待分区完成加载并从下一个分区开始。
[0055]	总点击次数= 14 次点击+附加字符串 1 (打开编辑器) + 1 (复制脚本) + 1 (下载脚本) + 2 (在当前脚本中搜索分区名称) + 4 (查找和替换名称) + 1 (复制脚本) + 4 (在 SSMS 中加载脚本, 运行并且保存)
	对于 12 个月= $12 * 14 = 168$ 次点击和等待时间
	对于 104 个分区= $14 * 104 = 1456$ 次点击和等待时间
	创建一个分区的平均时间= ~3 分钟
	对于 104 个分区= $3 * 1456 = 4368$ 分钟= ~72 小时

[0056] 表4

[0057]	同步分区创建
--------	--------

[0058]	1 分区加载工作
	将蓝图传送至自动化实用程序并且 它将为 SSMS 返回完整的编排脚本
	总点击次数= 3 次点击 1 (复制蓝图) + 1 (将蓝图传送至实用程序) + 1 (在 SSMS 中加载脚本)
	对于 12 个月= 3 次点击
	对于 104 个分区= 3 次点击
	创建一个分区的平均时间= ~3 分钟
	对于 104 个分区= ~3 分钟

[0059] 表5

[0060] 因此,应当理解,后端使用基于智能数据集市的表格建模的同步分区创建框架可以表示远比常规设计更加高效和简化的过程。此外,由于所有分区均是同步加载的,因此所提出的系统与所选择的分区的实际数目无关,从而允许对多个等级的大量分区进行有效、智能的管理。在不同的实施例中,数据分析系统然后可经由调度刷新模块260(在图2中引入)执行调度智能分区刷新,以确保每个分区中的数据是最新的。图6中描绘了用于更新分区600的这种方法的简化示例。在此示例中,应当理解,分区是基于早期模块的结果按逻辑顺序排列的,并且与表格模型紧密关联。

[0061] 在第一操作610中,示出初始列,该初始列表示已按日期分区的数据(按月分区,从2022年2月至2023年1月)。该列下方是一个较小的列,反映需要合并至现有数据中的更新数据。作为参考,这些分区将被标记为M1、M2、M3、……Mn,其中的数据将被标记为{M1}、{M2}、{M3}、……{Mn}。在第二操作620中,系统执行分区的第一次刷新。为了简单起见,仅示出了13个条目(M1-M13),但是在其他实施例中,该列表的数目可以大得多。在第一次刷新中,将数据增量(M13)追加至现有列表中。然而,系统不仅会添加新数据,还会替换旧分区以节省存储器。在这种情况下,存在被带入集合的数据{M13}的新分区M13,系统将使用它来自动地替换最早的先前数据。因此,通过第三操作630的第二次刷新,数据{M13}已被插入并重写了原始的{M1}数据,使得M1现在含有{M13}。此外,旧数据{M1}被删除并且归档(由列下方的独立M1单元格表示),从而优化存储、降低成本并减少数据泄露。此外,归档过程确保有适应系统故障的数据备份(即维护第二个历史版本),如果需要,可以用于恢复模型。在一些实施例中,系统的算法管理历史并且决定哪个分区需要被更新

[0062] 换句话说,可以发生交换,其中新数据内容替换旧数据内容,但仅限于该特定分区。同时,可以将下一更新的数据集(M14)移入循环中。这种方法适应了分区名称和分区数

目不变的性质,一旦部署了表格模型,就不能对其进行修改。在第四操作640中,系统从第二刷新移动至第三刷新。在该示例中,当前M1数据(即,现在由{M13}组成)保持完整并且被认为是最新的。相反,前一阶段附加的M14的新数据现在与由M2所持有的数据互换。换句话说,{M2}被{M14}所取代。与此同时,下一更新的数据集(M15)可被移动至循环中,而M2的原始数据被归档。

[0063] 在第五操作650中,进行第四次刷新。在这种情况下,当前M2数据(即,现在由{M14}组成)保持完整并且被认为是最新的。取而代之的是,在前一阶段附加的M15的新数据现在与由M3所持有的数据互换。换句话说,{M3}被{M15}所取代。与此同时,下一更新的数据集(M16)可以被移动至循环中,而M3的原始数据被归档。该过程在第六操作660中总结,其中呈现了来自第一操作610的基本相同的列,但是前三个分区的内容已经被更新,同时这三个分区的原始数据已经被安全地归档。

[0064] 尽管为读者清晰起见,图6中的操作按顺序呈现,但是应当理解,在不同实施例中,在系统本身的背景下,这些操作可基本上同步的方式发生(“同步刷新”)。这种方法与更耗时并且容易出现人为错误的常规数据刷新机制形成鲜明对比,常规数据刷新机制依赖于逐步的顺序,通过以手动定义的间隔写入和对数据进行分区的一系列步骤将数据从数据仓库移动至可视化层。所提出的实施例结合了同步刷新,例如通过在层之间(以及在分区的每个等级上)自动移动数据(每次需要刷新时)以及在新数据到达时自动添加数据,实现了更快、无错误的执行过程来刷新大量数据。换句话说,所提出的系统采用了分区刷新技术,该技术显著减少了在粒度等级维护最新资源所需的计算时间。在一个示例中,当系统通过中间表格模型工具在每个刷新周期(例如,用户选择的分区策略所请求的每个新周)将新数据从数据仓库层移动至可视化层时,系统可复制在表格模型上的对应智能数据集市中创建和维护的单独分区(例如,第104分区层)。该系统的高效同步刷新工具可以自动执行整个过程并且跨所有分区层(例如,103个分区层)移动数据,以便向第104个分区层添加新数据。这类自动化的方法——降低了错误率——在大数据中常规上是不可行的。

[0065] 如前所述,在不同的实施例中,所提出的系统可以采用生成式AI技术产生存储在智能数据集市中的聚合数据的可视化。在一些实施例中,生成式AI被配置成基于查询而过滤数据并以自动生成与查询相关的见解。因此,通过为生成式AI提供智能数据集市作为其来源(其中的信息经过高度组织),可以快速开发可视化以提供表示最新可用信息的见解。例如,在图7中,在第一操作710中,用户可以经由数据分析仪表盘访问系统的前端并且提交理解数据的特定方面/特征的请求。在第二操作720中,系统调用生成式AI,该生成式AI作为响应——在第三操作730中——查询智能数据集市。基于可用数据,然后可以自动生成各种可视化。

[0066] 在一些实施例中,生成式AI可以包括基于变换器的深度学习神经网络架构。例如,领域特定语言模型可包括自回归语言模型(例如,生成预训练转换器(GPT)语言预测模型),该模型使用深度学习来产生响应于查询的文本(例如,产生对问题的回答和/或产生对评论的回复,等等)。在一些实施例中,可视化可以通过将数据传送至机器学习模型来生成,诸如生成预训练转换器(GPT),其是由OpenAI创建的开源人工智能(参见Alec Radford等人对<https://OpenAI.com/blog> (2023)的发帖,包括语言模型是无监督的多任务学习者,其全部内容通过引用并入)。在一些实施例中,机器学习模型可以在数据集中训练,诸如来自单词

加扰 (PAWS) 数据集中的释义反义词。

[0067] 图8是根据一个实施例的智能数据分析系统814(或系统814)的环境800的示意图。环境800可以包括能够执行所公开的方法的多个组件。例如,环境800包括用户设备804、计算/服务器系统808、数据可视化平台826以及可以包括智能数据集市的数据库890。环境800的组件可以通过网络802彼此通信。例如,用户设备804可经由网络802从数据库890中取回信息。在一些实施例中,网络802可以为广域网(“WAN”),诸如因特网。在其他实施例中,网络806可以为局域网(“LAN”)。

[0068] 如图8所示,系统814的组件可以被安装在计算系统808中,计算系统808可具有存储器812和处理器810。处理器810可以包括位于单个设备上的单个设备处理器,或者它可以包括位于一个或多个物理设备上的多个设备处理器。存储器812可以包括任何类型的存储,其可以物理地位于一个物理设备上,或者位于多个物理设备上。在一些情况下,计算系统808可以包括用于托管该系统的一个或多个服务器。

[0069] 虽然图8示出了一个用户设备,但是应当理解,可以使用一个或多个用户设备。例如,在一些实施例中,系统可包括两个或三个用户设备。在一些实施例中,用户设备可以为用户使用的计算设备。例如,用户设备804可以包括智能电话或平板计算机。在其他示例中,用户设备804可以包括膝上型计算机、台式计算机和/或其他类型的计算设备。用户设备可用于输入、处理和显示信息。参考图8,环境800可以还包括数据库890,该数据库存储已处理的分区数据。该数据可由系统814的其他组件取回。如上所述,系统814可以包括分区器模块818、主标记模块820、分区模块822和同步刷新模块824。这些模块/组件中的每个均可以用于执行本文所描述的操作。

[0070] 图9是图示为基于云计算的基础设施处理和对数据分区的方法900的实施例的流程图。方法900包括在数据分析系统的分区器模块处接收第一客户标准和包括表格数据的第一客户数据集的第一步骤910,以及在分区器模块处计算表格数据的第一大小得分的第二步骤920,第一大小得分标识最有可能满足第一客户标准的第一数据增量。此外,第三步骤930包括在分区器模块处基于第一大小得分而生成第一分区策略,并且第四步骤940包括在数据分析系统处实施第一分区策略以产生表格数据中的聚合数据的集合,这些集合包括第一分区。方法900的第五步骤950包括将集合数据作为智能数据集市的一部分存储在数据分析系统中。第六步骤960包括经由与数据可视化平台相关联的第一用户界面接收第一用户查询,并且第七步骤970包括确定第一用户查询与第一分区有关。此外,方法900包括经由数据可视化平台从智能数据集市取回第一分区的第八步骤980,以及为数据可视化平台并且经由生成人工智能(AI)模块基于第一分区生成第一可视化的第九步骤990。

[0071] 在其他实施例中,该方法可以包括附加的步骤或方面。在一个示例中,实施第一分区策略还包括在基本上同步的过程中在表格数据中创建多个分区。在一些实施例中,该方法还包括以下步骤:接收指示第一类型数据的第一刷新率的第二客户标准;通过对智能数据集市的参考,确定仅第一分区的第一部分包括第一类型数据;以及基于第二客户标准,以第一刷新率仅刷新第一分区的第一部分。在另一个实施例中,该方法还包括以下步骤:接收指示不同于第一刷新率的第二类型数据的第二刷新率的第三客户标准;通过对智能数据集市的参考,确定只有第一分区的第二部分包括第二类型数据;以及基于第三客户标准,以第二刷新率仅刷新第一分区的第二部分。

[0072] 在不同的实施例中,附加步骤可以包括在分区器模块处计算表格数据的第二大小得分,第二大小得分标识也可能满足第一客户标准的第二数据增量;在分区器模块处,基于第二大小得分而生成第二分区策略;以及经由数据分析系统的第二用户界面将第一分区策略呈现第一选项以及将第二分区策略呈现为第二选项。在一个示例中,该方法还包括经由第二用户界面接收对第一选项的选择,并且其中实施第一分区策略是响应于接收到对第一选项的选择。在另一示例中,该方法还包括在分区器模块处计算表格数据的第二大小得分,该第二大小得分标识比第一数据增量更不可能满足第一客户标准的第二数据增量;在分区器模块处,基于第二大小得分而生成第二分区策略;以及经由数据分析系统的第二用户界面将第二分区策略呈现为拒绝的选项以供审查。

[0073] 就本申请而言,“接口(interface)”可以被理解为通过客户应用程序向应用程序用户传达内容的机制。在一些示例中,接口可以包括可以经由本地应用用户界面(UI)呈现给用户的弹出窗口、控件、可驱动接口、交互式按钮/选项或可通过本地应用程序UI示出给用户的其他对象,以及特定应用程序固有的用于呈现与这些本地控件相关联的内容的机制。此外,术语“致动(actuation)”或“致动事件(actuation event)”是指与特定输入或经由接口使用应用程序相关联的事件(或特定的事件序列),其可触发应用程序的显示的变化。此外,“本地控件(native control)”是指用于通过客户应用程序向应用程序用户传送内容的机制。例如,本地控件可包括可经由本地应用程序UI、触摸屏访问点、菜单项或可通过本地应用程序UI向用户示出的其他对象呈现给用户的可致动或可选择的选项或“按钮”、较大接口的片段以及特定应用程序本地的用于呈现与这些本地控件相关联的内容的机制。术语“资产(asset)”是指可与本机应用程序中的本机控件相关联地呈现的内容。作为一些非限制性示例,资产可以包括可致动弹出窗口中的文本、与按钮或其他原生应用程序对象的交互点击相关联的音频、与用户界面相关联的视频或其他这类信息呈现。

[0074] 应当理解,图中所示的文本、图像和特定应用程序特征仅用于说明目的,绝不限制应用程序传达或接收信息的方式。此外,在其他实施例中,一个或多个选项或其他字段和文本可不同地出现和/或可在与客户系统相关联的屏幕(一个或多个)上的任何其他地方显示或生成,包括与用户界面隔开、邻近或围绕用户界面。换句话说,附图仅呈现接口的一种可能布局,并且不以任何方式限制任何公开特征的呈现布置。

[0075] 实施例可以包括存储软件的非暂时性计算机可读介质(CRM),该软件包括可由一个或多个计算机执行的指令,这些指令在这类执行时,可促使一个或多个计算机执行所公开的方法。非暂时性CRM可指在短时间内或在有电源的情况下存储数据的CRM,诸如存储设备或随机存取存储器(RAM)。例如,非暂时性计算机可读介质可包括存储组件,诸如硬盘(例如,磁盘、光盘、磁光盘和/或固态硬盘)、光盘(CD)、数字多功能盘(DVD)、软盘、盒式磁带和/或磁带。

[0076] 为了提供另外的背景,在一些实施例中,本文所描述的一些过程可被理解为在系统架构中操作,该系统架构可以包括位于不同位置处的多个虚拟局域网(VLAN)工作站,这些工作站通过专用虚拟服务器与主数据中心通信,该专用虚拟服务器诸如用于用户界面的网络服务器、用于数据处理的应用程序服务器、用于数据存储的数据库等。一般而言,虚拟服务器是一种在硬件组件(例如,服务器)上执行的虚拟机(VM)。在一些示例中,可在一个或多个服务器上部署多个虚拟机。

[0077] 在不同实施例中,系统可以至少部分被托管在供给现成的可扩展性和安全性的云计算环境中。云计算环境可以包括例如托管文档处理管理服务的环境。云计算环境可提供计算、软件、数据访问、存储等服务,这些服务不需要最终用户知道托管策略管理服务的系统(一个或多个)和/或设备(一个或多个)的物理位置和配置。例如,云计算环境可包括一组计算资源(统称为“计算资源”,单独称为“计算资源”)。可设想,本公开的实施方式可利用适当的云提供方(例如,由AmazonTM提供的AWS、由GoogleTM提供的GCP、由MicrosoftTM提供的Azure等)来实施。

[0078] 上述方法、设备和处理可通过多种不同方式和多种不同的硬件和软件组合实施。例如,实施方式的全部或部分可为包括指令处理器的电路装置,诸如中央处理单元(CPU)、微控制器或微处理器;或者被实施为专用集成电路(ASIC)、可编程逻辑器件(PLD)或现场可编程门阵列(FPGA);或者被实施为包括离散逻辑或其他电路组件的电路装置,包括模拟电路组件、数字电路组件或两者;或它们的任意组合。

[0079] 尽管已描述了本发明的各种实施例,但是该描述旨在进行示例性说明,而非限制性说明,并且对于本领域的普通技术人员来说,清楚的是,在本发明的范围内还有更多的实施例和实施方式是可能的。因此,除了根据所附权利要求及其等同物之外,本发明不受限制。此外,在所附权利要求的范围内可进行各种修改和改变。

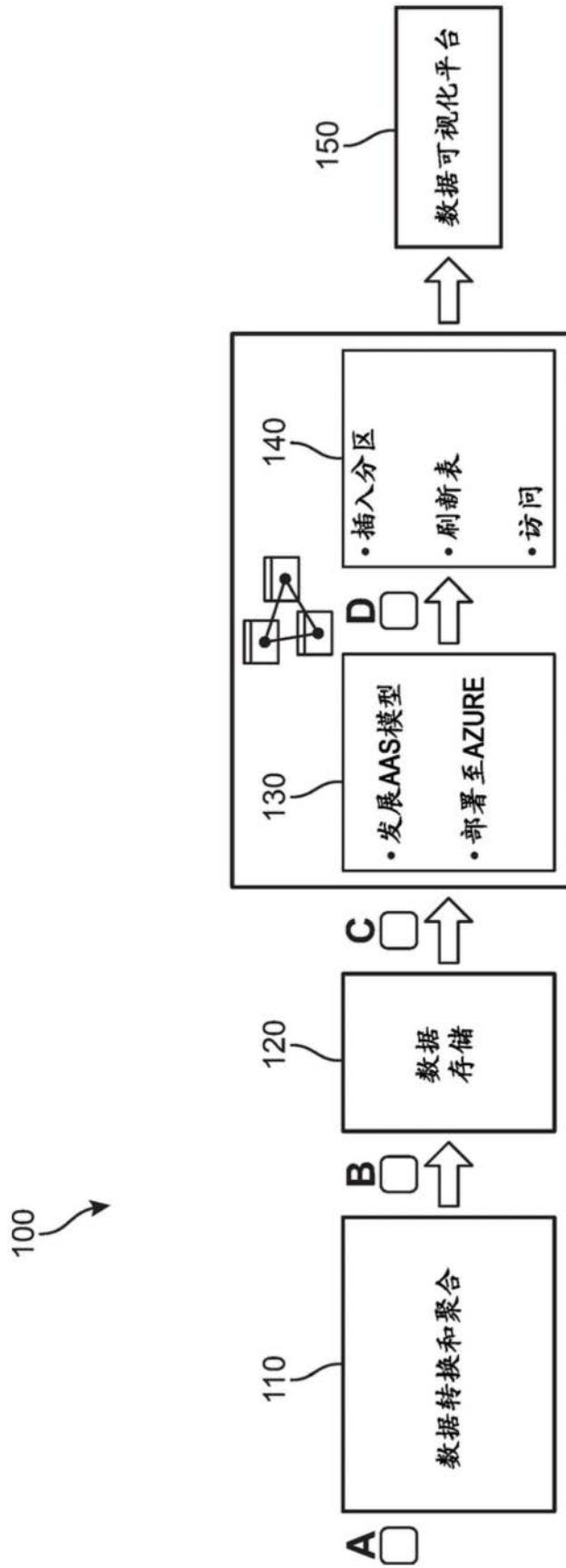


图1

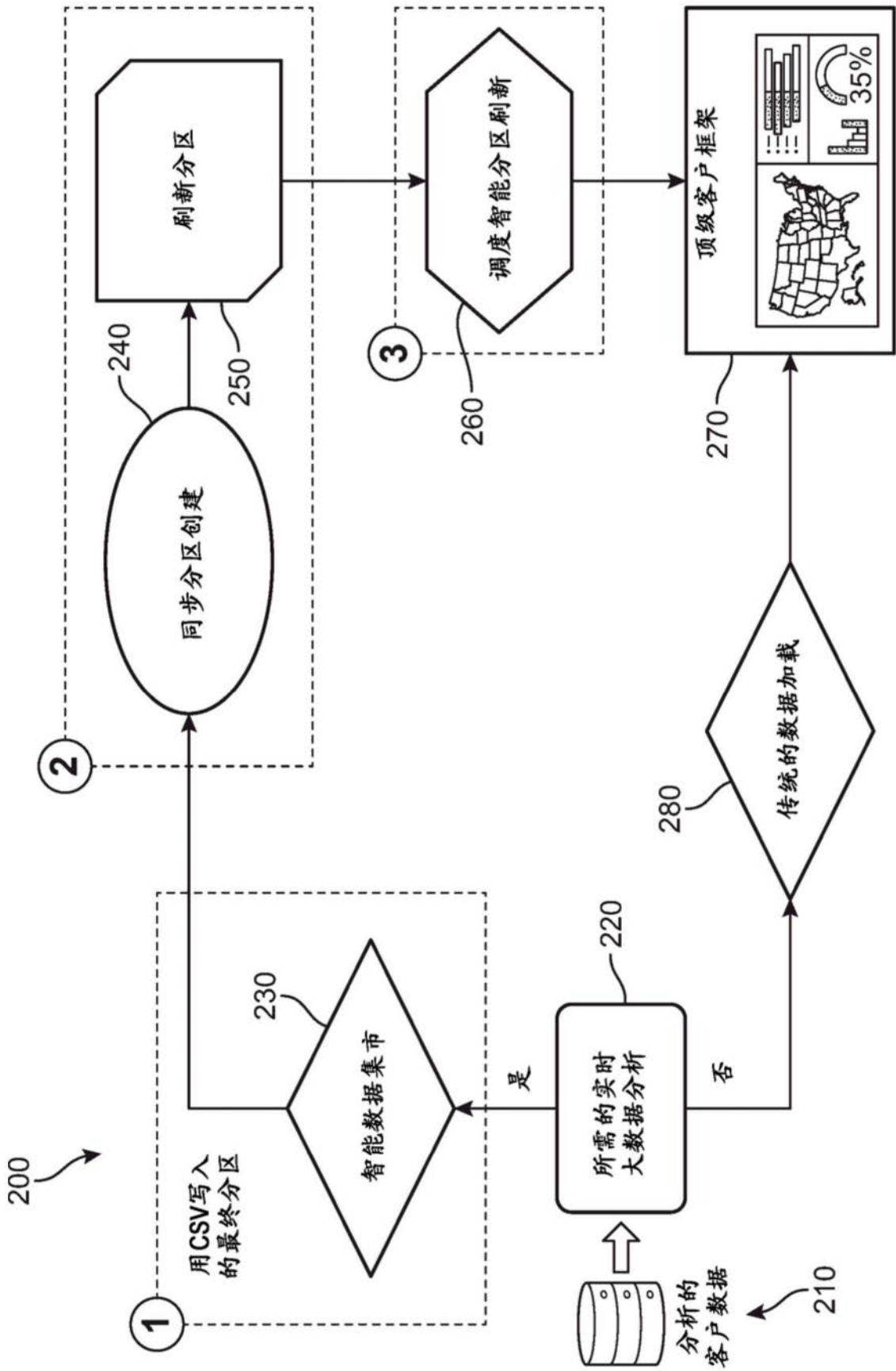


图2

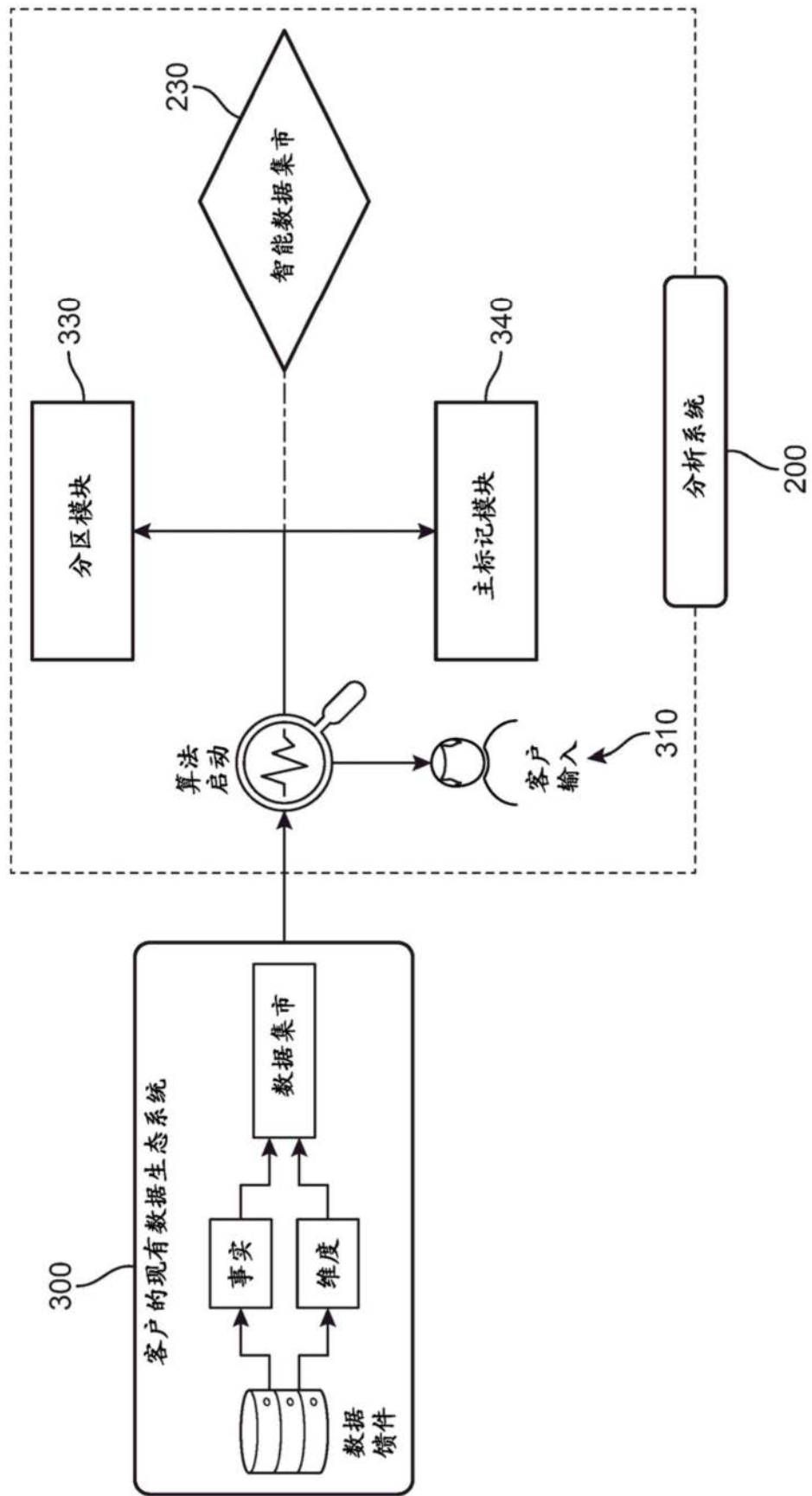


图3

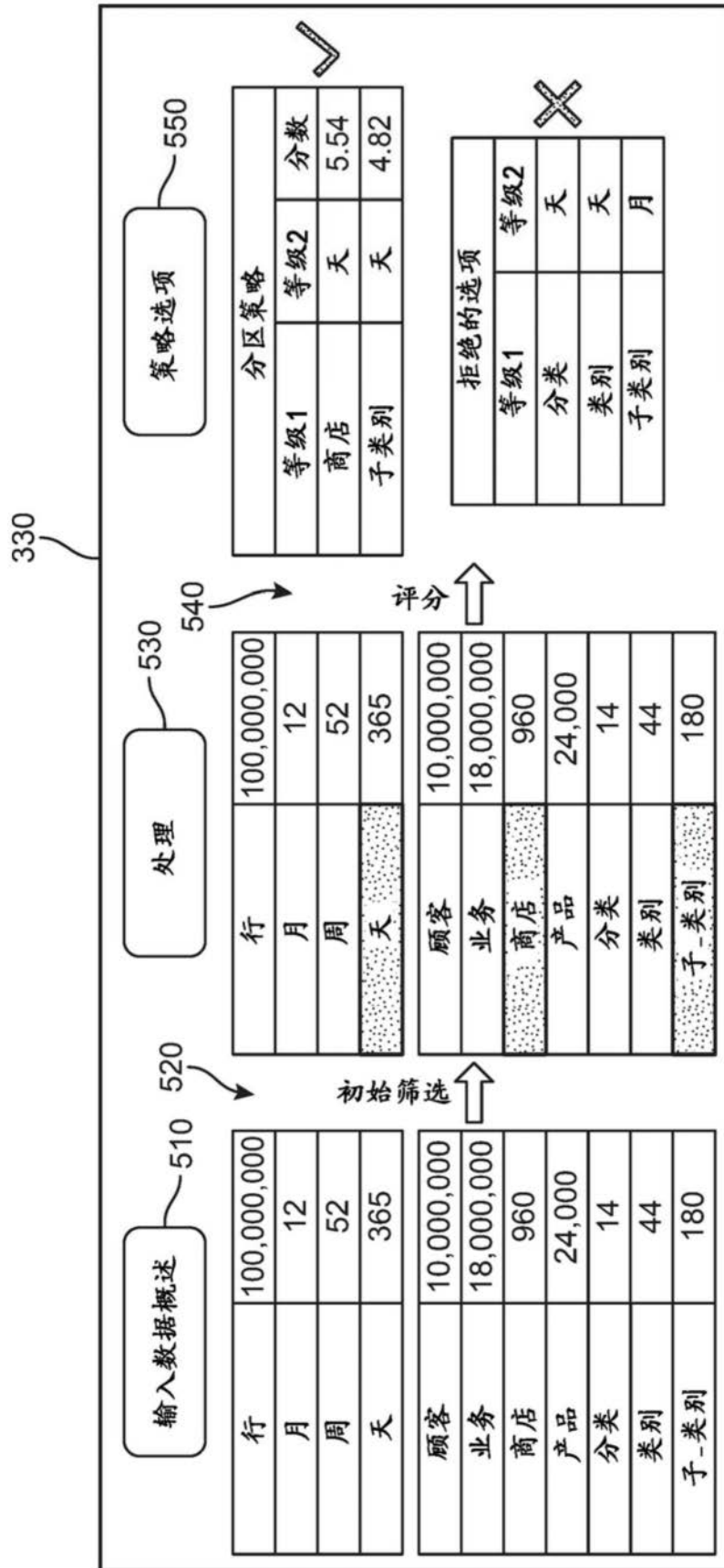


图5

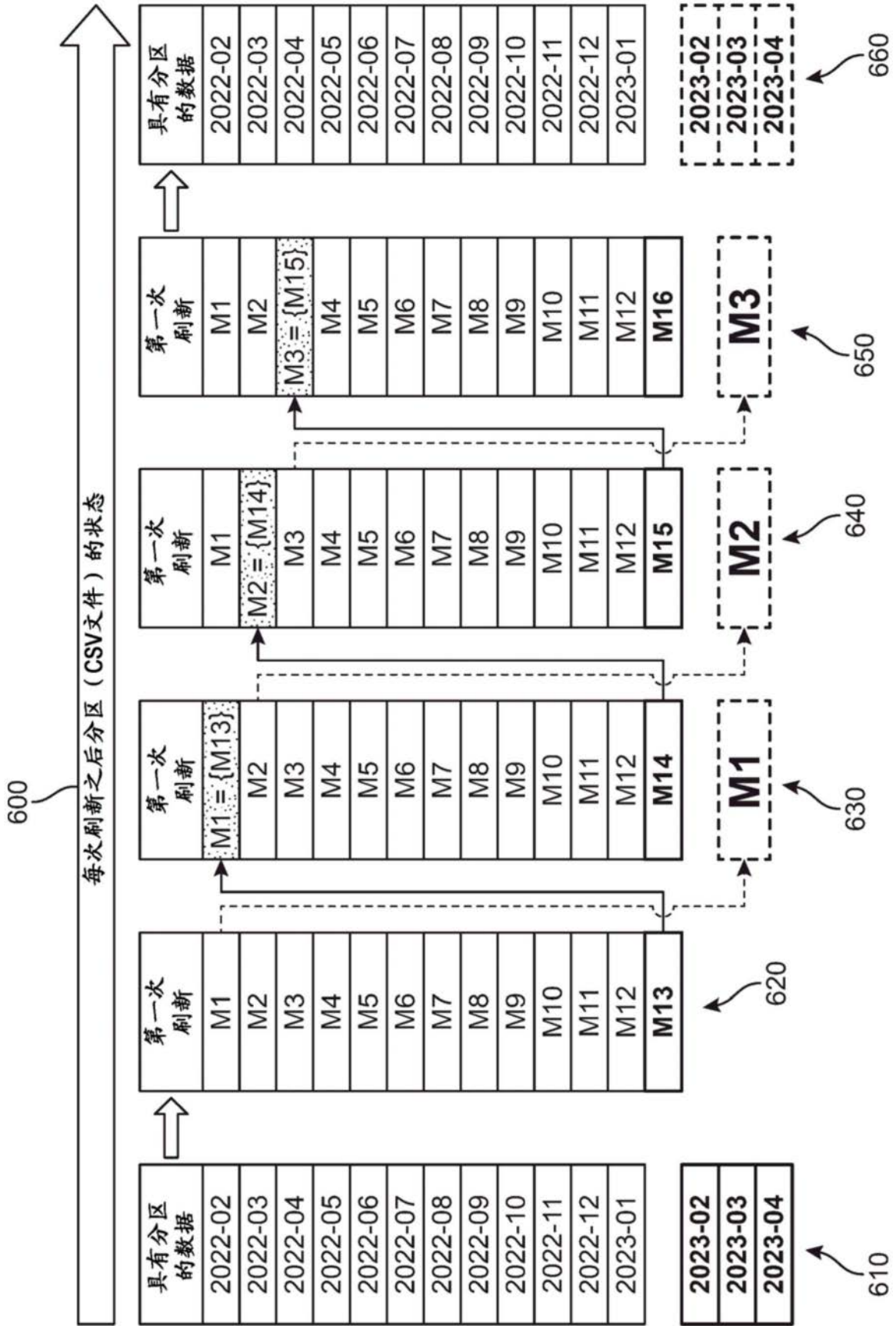


图6

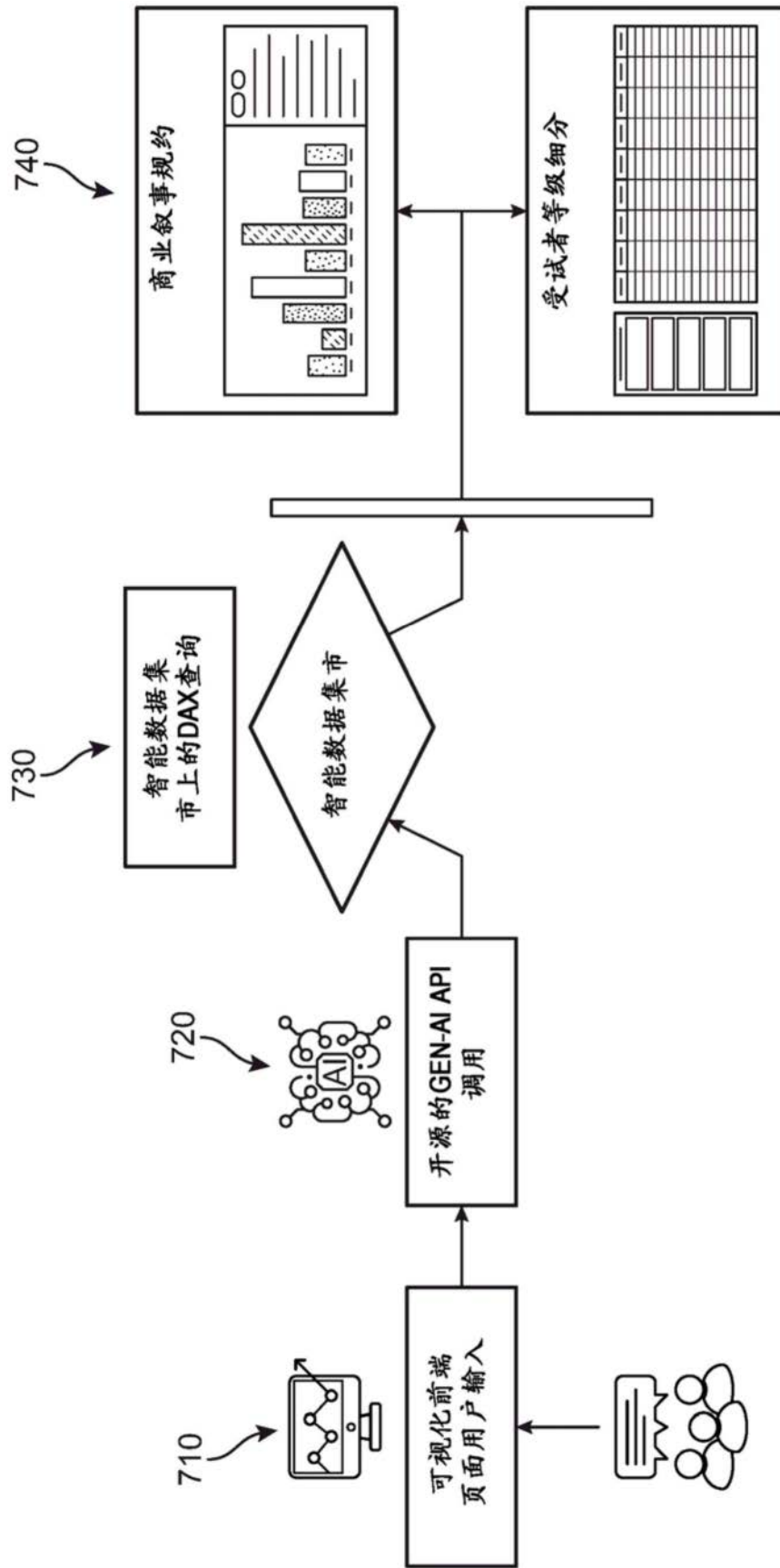


图7

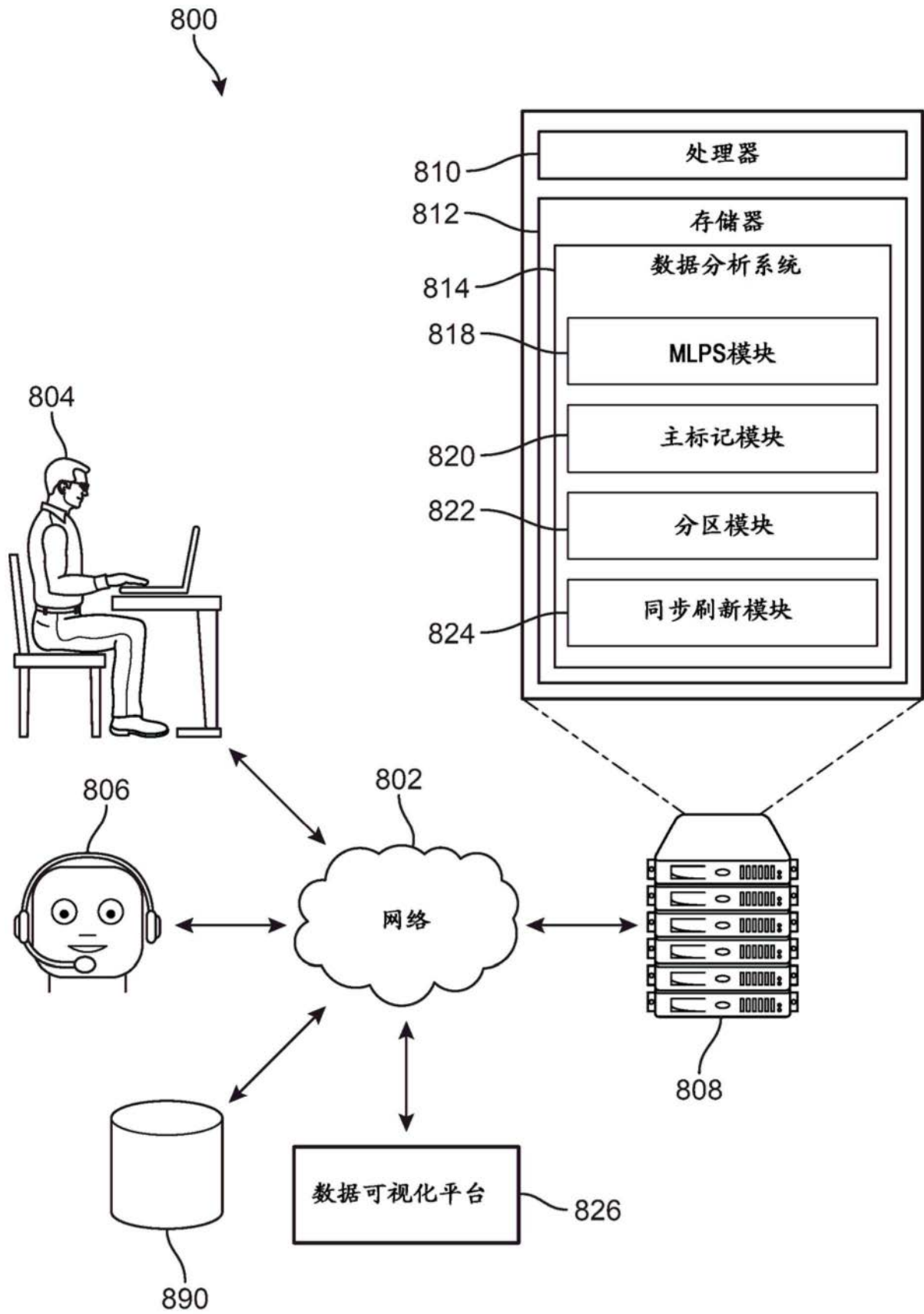


图8

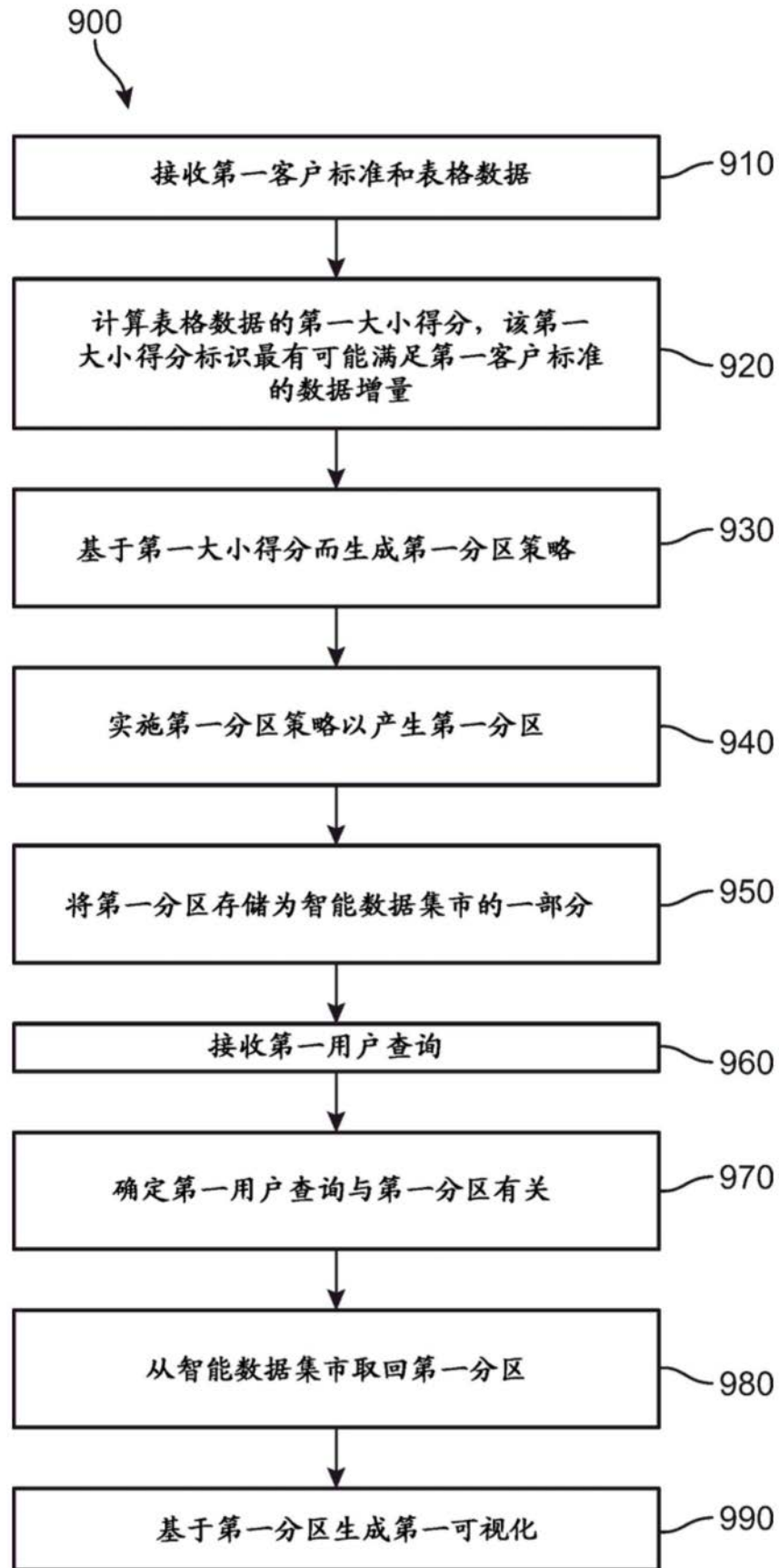


图9