



(12) 发明专利

(10) 授权公告号 CN 106933206 B

(45) 授权公告日 2021.07.06

(21) 申请号 201610885292.0

(51) Int.Cl.

(22) 申请日 2016.10.10

G05B 19/418 (2006.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 106933206 A

(56) 对比文件

(43) 申请公布日 2017.07.07

WO 2014145801 A2, 2014.09.18

(30) 优先权数据

CN 101477568 A, 2009.07.08

62/239,620 2015.10.09 US

CN 101606184 A, 2009.12.16

15/274,244 2016.09.23 US

CN 104035392 A, 2014.09.10

(73) 专利权人 费希尔-罗斯蒙特系统公司

CN 103020078 A, 2013.04.03

地址 美国德克萨斯州

CN 104662490 A, 2015.05.27

(72) 发明人 M·J·尼克松 A·T·恩弗

CN 102375449 A, 2012.03.14

N·H·贝尔 J·B·基德

CN 104885077 A, 2015.09.02

P·R·穆斯托尼

CN 103491059 A, 2014.01.01

(74) 专利代理机构 永新专利商标代理有限公司

CN 103562809 A, 2014.02.05

72002

CN 103631910 A, 2014.03.12

EP 2936341 A1, 2015.10.28

代理人 曹雯

审查员 赵立莎

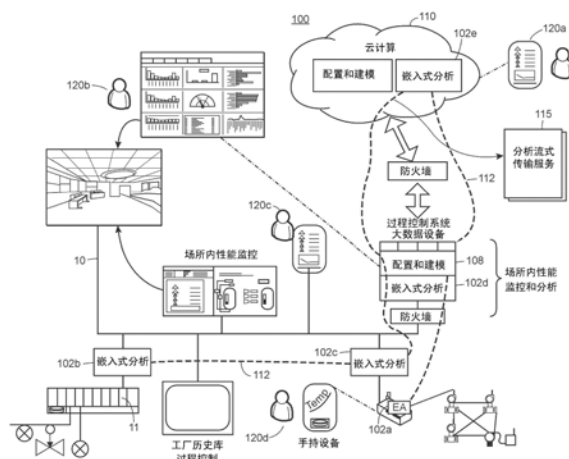
权利要求书4页 说明书83页 附图36页

(54) 发明名称

分布式工业系统中的独立于源的查询

(57) 摘要

提供了用于使用独立于源的标准化查询来请求和提供过程工厂数据的技术。请求设备生成标准化查询以从一个或多个数据源(例如,相关数据库或非相关数据库)获得数据。查询使用并不依赖于数据源的标准化格式,请求可以被生成为JSON文件。标准化查询不可直接用于任何数据源。相反,一旦接收到标准化查询,数据设备就生成一个或多个特定于源的查询。特定于源的查询使用对每个数据源原生的语法来获得数据。在某些实例中,所接收到的数据必须被进一步处理以调整不同的采样时间或采样率,例如通过插值法。来自所有数据源的得到的数据可以在返回到请求设备之前被聚合到数据帧中。



1. 一种用于从多个电子数据源中访问与过程工厂的操作相关联的过程工厂数据的系统,包括:

第一电子数据源和第二电子数据源,所述第一电子数据源和所述第二电子数据源均包括有形的、非暂时性储存介质,所述储存介质储存与所述过程工厂的所述操作相关联的所述过程工厂数据;

处理器,所述处理器通信地耦合到所述第一电子数据源和所述第二电子数据源;

程序存储器,所述程序存储器通信地耦合到所述处理器,并包括有形的、非暂时性的储存介质,所述储存介质储存指令,当所述指令被所述处理器执行时,使得所述处理器进行以下操作:

接收标准化查询,所述标准化查询使用标准化的查询格式;

基于所述标准化查询生成特定于第一数据源的查询,其中,所述特定于第一数据源的查询使用与所述第一电子数据源相关联的第一查询格式;

基于所述标准化查询生成特定于第二数据源的查询,其中,所述特定于第二数据源的查询使用与所述第二电子数据源相关联的第二查询格式;

使得特定于所述第一数据源的查询被执行以从所述第一电子数据源获得具有第一采样率的第一数据集;

使得特定于所述第二数据源的查询被执行以从所述第二电子数据源获得具有第二采样率的第二数据集;以及

通过对齐所述第一采样率和所述第二采样率来生成包括所述第一数据集和所述第二数据集的聚合的数据帧。

2. 根据权利要求1所述的系统,其中,所述标准化查询包括所述第一电子数据源的第一指示和所述第二电子数据源的第二指示。

3. 根据权利要求2所述的系统,其中,所述第一指示标识所述第一查询格式并且所述第二指示标识所述第二查询格式。

4. 根据权利要求2所述的系统,其中,所述第一指示和所述第二指示被置于对数据列的指示之前。

5. 根据权利要求1所述的系统,其中,所述标准化查询格式使用与所述第一查询格式的查询语法和所述第二查询格式的查询语法两者不同的查询语法。

6. 根据权利要求5所述的系统,其中,所述标准化查询格式的所述查询语法无法直接执行来从所述第一电子数据源获得所述第一数据集或从所述第二电子数据源获得所述第二数据集。

7. 根据权利要求1所述的系统,其中,所述标准化查询包含在Java脚本对象表示(JSON)文件中。

8. 根据权利要求1所述的系统,其中,所述标准化查询包括对时间范围的指示,所述时间范围标识与所述第一数据集和所述第二数据集中的数据条目相对应的时间段。

9. 根据权利要求8所述的系统,其中,所述时间范围包括多个时间段,并且其中,所述多个时间段中的至少两个被所排除的时间段分隔开,针对所述所排除的时间段未查询数据。

10. 根据权利要求9所述的系统,其中,所述时间范围包括多个时间段,并且其中,所述时间范围标识与所述第一数据集相关联的第一时间段以及与所述第二数据集相关联的第

二时间段。

11. 根据权利要求1所述的系统,其中,所述第一电子数据源是关系型数据库,并且所述第二电子数据源是非关系型数据库。

12. 根据权利要求1所述的系统,其中,所述标准化查询包括所聚合的数据帧的采样率的指示。

13. 根据权利要求1所述的系统,其中,对齐所述第一采样率和所述第二采样率包括增加与获得期望的采样率所需要的、未采样的时间相对应的数据点。

14. 根据权利要求1所述的系统,其中,所增加的数据点是在时间上正好在所增加的数据点之前的经采样的数据的给定值。

15. 根据权利要求1所述的系统,其中,对齐所述第一采样率和所述第二采样率包括移除经采样的数据点。

16. 根据权利要求1所述的系统,其中,所述第一数据集和所述第二数据集各自包括多个数据点,所述多个数据点包含由过程工厂内的一个或多个现场设备所测量的信息。

17. 一种有形的、非暂时性计算机可读储存介质,所述储存介质储存用于从电子数据源获得与过程工厂的操作相关联的时间序列数据的指令,当所述指令被计算机系统的处理器执行时,使得所述计算机系统进行以下操作:

接收所述电子数据源的指示,所述数据将从所述电子数据源获得;

接收时间范围的指示,所述数据将在所述时间范围内获得,其中,所述指示以独立于所述电子数据源的标准化查询格式来表示;

接收将被获得的所述数据的特性的指示,其中,所述指示以独立于所述电子数据源的所述标准化查询格式来表示;

确定与所述电子数据源相关联的特定于数据源的查询格式;

基于所述时间范围的指示和将被获得的数据的特征的指示来生成遵守特定于数据源的格式的特定于所述数据源的查询;

使得所述特定于数据源的查询被执行,以从所述电子数据源选择所述数据;以及
从所述电子数据源接收所述数据。

18. 根据权利要求17所述的有形的、非暂时性计算机可读储存介质,其中,所述时间范围的指示和所述特征的指示是在使用所述标准化查询格式的标准化查询中接收的。

19. 根据权利要求17所述的有形的、非暂时性计算机可读储存介质,还储存指令,所述指令使得所述处理器进行以下操作:

接收一格式的指示,所述数据将以所述格式被获得;以及

将从所述电子数据源接收到的所述数据转换成所指示的格式。

20. 根据权利要求17所述的有形的、非暂时性计算机可读储存介质,其中,所述标准化查询格式使用与所述电子数据源所使用的原生语法不同的语法。

21. 根据权利要求17所述的有形的、非暂时性计算机可读储存介质,其中,所述标准化查询格式使用无法直接执行以从所述电子数据源获得所述数据的语法。

22. 根据权利要求17所述的有形的、非暂时性计算机可读储存介质,其中,使得所述处理器生成所述特定于数据源的查询的所述指令使得所述处理器将所指示的时间范围和所指示的将被获得的数据的特征映射到所确定的特定于数据源的查询格式。

23. 根据权利要求17所述的有形的、非暂时性计算机可读储存介质,其中,使得所述处理器使所述特定于数据源的查询被执行的所述指令使得所述处理器向所述电子数据源的执行引擎发送所述特定于数据源的查询,其中,所述执行引擎使用所述特定于数据源的查询来查询所述电子数据源。

24. 根据权利要求17所述的有形的、非暂时性计算机可读储存介质,其中,所述时间范围的指示包括多个开始时间和多个结束时间。

25. 根据权利要求17所述的有形的、非暂时性计算机可读储存介质,其中,所述时间范围的指示包括针对将被获得的数据的采样率。

26. 根据权利要求17所述的有形的、非暂时性计算机可读储存介质,还储存使得所述处理器生成数据帧的指令,所述数据帧包括从所述电子数据源接收到的数据,其中,所述数据帧包括多个数据点,每个数据点与所述时间范围内的时间相关联。

27. 根据权利要求26所述的有形的、非暂时性计算机可读储存介质,其中,每个数据点与所述时间范围内的、与开始时间之后的整数倍的采样周期对齐的或者为所述开始时间的时间相关联。

28. 根据权利要求17所述的有形的、非暂时性计算机可读储存介质,其中,所述将获得的数据的特征的指示包括关于所述过程工厂的操作的一种或多种类型的测量结果的指示。

29. 根据权利要求28所述的有形的、非暂时性计算机可读储存介质,其中,所述一种或多种类型的测量结果的指示包括一个或多个标签、别名、以及与所述数据相关联的数据类型。

30. 根据权利要求17所述的有形的、非暂时性计算机可读储存介质,其中,所述电子数据源是大数据储存节点。

31. 一种用于从电子数据源提供数据的方法,包括:

从数据请求实体接收具有标准化查询格式的标准化查询,其中,所述标准化查询格式与所述电子数据源使用的特定于数据源的查询格式不同;

从所述标准化查询提取查询参数;

基于所提取的查询参数来生成具有所述特定于数据源的查询格式的特定于数据源的查询;

执行所述特定于数据源的查询以从所述电子数据源获得所述数据;以及

向数据接收实体提供所获得的数据。

32. 根据权利要求31所述的方法,其中,所述数据请求实体包括分析程序内的查询块。

33. 根据权利要求31所述的方法,其中,所述标准化查询使用无法直接执行以从所述电子数据源获得所述数据的语法。

34. 根据权利要求31所述的方法,其中,提取所述查询参数包括确定(i)所述数据的时间周期和数据参数,所述时间周期具有开始时间和结束时间,所述时间参数指示将从所述电子数据源获得的数据的类型,以及(ii)采样率。

35. 根据权利要求34所述的方法,其中,向所述请求实体提供所获得的数据包括返回与所述时间周期相关联的数据点集,其中每个数据点都包括时间戳,所述时间戳为所述开始时间或者在所述开始时间之后的整数倍的所述采样率的周期。

36. 根据权利要求35所述的方法,其中,通过使用所述电子数据源中的与一个或多个添

加的数据点的每个时间戳之前最接近的时间相关联的数据条目的值,来将一个或多个数据点添加至所述数据点集,其中时间戳指示所述开始时间之后的整数倍的所述采样率的周期。

37. 根据权利要求34所述的方法,其中,由所述数据参数指示的所述类型是一种类型的测量结果或者来自一种测量设备的测量结果,其中,所述测量设备是位于过程工厂内的现场设备。

38. 根据权利要求34所述的方法,其中,所述数据参数还指示所述将获得的数据的标签、被提供给所述数据请求实体的所述数据的别名、以及被提供给所述数据请求实体的所述数据的数据格式类型。

39. 根据权利要求31所述的方法,其中,向所述数据接收实体提供所述数据包括向所述数据接收实体发送包含所获得的数据的数据帧,并且其中,所述标准化查询包括所述数据帧的格式的指示,并且其中,所述数据帧根据所述数据帧的格式的指示而被格式化。

40. 根据权利要求31所述的方法,其中,所述数据接收实体是与所述数据请求实体相同的实体。

分布式工业系统中的独立于源的查询

[0001] 相关申请的交叉引用

[0002] 本申请享有与2015年10月9日递交的、标题为DISTRIBUTED INDUSTRIAL PERFORMANCE MONITORING AND ANALYTICS的美国临时申请No.32/239,620的优先权,故通过引用方式将其全部内容明确地并入本文。本申请与以下申请有关:(i)于2013年3月4日递交的、标题为“BIG DATA IN PROCESS CONTROL SYSTEMS”的美国申请No.13/784,041,(ii)于2014年2月6日递交的、标题为“COLLECTING AND DELIVERING DATA TO A BIG DATA MACHINE IN A PROCESS CONTROL SYSTEM”的美国申请No.14/174,413,(iii)于2014年8月11日递交的、标题为“SECURING DEVICES TO PROCESS CONTROL SYSTEMS”的美国申请No.14/456,763,(iv)于2014年3月17日递交的、标题为“DATA MODELING STUDIO”的美国申请No.14/216,823,(v)于2014年1月31日递交的、标题为“MANAGING BIG DATA IN PROCESS CONTROL SYSTEMS”的美国申请No.14/169,965,(vi)于2014年3月14日递交的、标题为“DETERMINING ASSOCIATIONS AND ALIGNMENTS OF PROCESS ELEMENTS AND MEASUREMENTS IN A PROCESS”的美国申请No.14/212,411,(vii)于2014年3月14日递交的、标题为“DISTRIBUTED BIG DATA IN A PROCESS CONTROL SYSTEM”的美国申请No.14/212,493,(viii)于2014年10月6日递交的、标题为“STREAMING DATA FOR ANALYTICS IN PROCESS CONTROL SYSTEMS”的美国申请No.14/506,863,(ix)于2014年10月6日递交的、标题为“REGIONAL BIG DATA IN PROCESS CONTROL SYSTEMS”的美国申请No.14/507,188,(x)于2014年10月6日递交的、标题为“DATA PIPELINE FOR PROCESS CONTROL SYSTEM ANALYTICS”的美国申请No.62/060,408以及(xi)于2014年10月6日递交的、标题为“AUTOMATIC SIGNAL PROCESSING-BASED LEARNING IN A PROCESS PLANT”的美国申请No.14/507,252,故通过引用方式将上述申请的全部公开内容明确地并入本文。

技术领域

[0003] 概括地说,本公开内容涉及过程工厂和过程控制系统,更具体地说,涉及由过程工厂和过程控制系统生成的实时数据的实时性能监控和分析。

背景技术

[0004] 如那些用于化学、石油或工业或其它过程控制工厂中以制造、提炼、变换、生成或者产生物理材料或产品的分布式过程控制系统,典型地包括一个或多个过程控制器,该过程控制器经由模拟、数字或组合的模拟/数字总线、或经由无线通信链路或网络来通信地耦合到一个或多个现场设备。该现场设备(其可以是例如阀、阀定位器、开关和变送器(例如,温度、压力、液位和流速传感器)),位于过程环境中并且通常执行物理或过程功能,例如打开或关闭阀、测量诸如温度或压力之类的过程和/或环境参数等等以控制在过程工厂或系统内执行的一个或多个过程。智能现场设备,例如遵循公知的Fieldbus协议的现场设备,还可以执行控制计算、报警功能以及在控制器内通常实现的其它控制功能。该过程控制器(典型地也位于工厂环境中)接收指示由现场设备获得的过程测量结果的信号和/或其它关于

现场设备的信息,并且执行控制器应用,控制器应用例如运行作出过程控制决定的不同的控制模块,基于所接收到的信息来生成控制信号,以及与正在诸如HART[®]、WirelessHART[®]和FOUNDATION[®]Fieldbus现场设备之类的现场设备中执行的控制模块或块进行协调。该控制器中的控制模块通过通信线路或链路发送控制信号给现场设备,由此来控制过程工厂的至少一部分的操作,例如以控制在该工厂内运行或执行的一个或多个工业过程的至少一部分。例如,控制器和现场控制设备控制由过程工厂的过程控制系统控制的过程的至少一部分。

[0005] 来自现场设备和控制器的信息通常通过数据高速通路或通信网络而可用于一个或多个其它硬件设备,例如操作员工作站、个人计算机或计算设备、数据历史库(historian)、报表生成器、集中式数据库、或其它集中式管理计算机设备,这些硬件设备典型地放置在控制室或远离较恶劣的工厂环境的其它位置。这些硬件设备中的每一个通常跨过程工厂或跨过程工厂的一部分而集中。这些硬件设备运行应用,该应用可以例如使操作员能够执行关于控制过程和/或操作过程工厂的功能,例如改变过程控制例程的设置、修改控制器或现场设备内的控制模块的操作、查看过程的当前状态、模拟过程的操作以便培训员工或测试过程控制软件、保留和更新配置数据库等。硬件设备、控制器和现场设备所使用的的数据高速通路可以包括有线通信路径、无线通信路径、或者有线和无线通信路径的组合。

[0006] 举例而言,由Emerson过程管理(Emerson Process Management)销售的DeltaV[™]控制系统包括多个应用,这些应用储存在位于过程工厂内的不同地方的不同设备内,并且由这些设备执行。位于一个或多个工作站或计算设备中的配置应用使用户能够创建或改变过程控制模块,以及经由数据高速通路将这些过程控制模块下载到专用分布式控制器。通常,这些控制模块由通信地互连的功能块组成,其是面向对象的编程协议的对象,所述对象在控制方案内基于到控制方案的输入来执行功能,并且将输出提供给控制方案内的其它功能块。配置应用也可以允许配置设计者来创建或改变操作员界面,这些界面被查看应用用于向操作员显示数据,以及使用户能够改变设置,诸如过程控制例程内的设定点。每个专用控制器以及在某些情况下一个或多个现场设备储存和执行相应的控制器应用,该应用运行被分配和下载至其的控制模块,以便实现实际的过程控制。可以在一个或多个操作员工作站上(或在与操作员工作站和数据高速通路通信连接的一个或多个远程计算设备上)执行的查看应用经由数据高速通路从控制器应用接收数据并且向过程控制系统设计者、操作员、或使用用户界面的用户显示该数据,并且可以提供多种不同视图的任意一种,例如操作员视图、工程师视图、技术人员视图等。数据历史库应用典型地储存在数据历史库设备中并由其执行,该数据历史设备收集和储存跨数据高速通路提供的一些或全部数据,同时配置数据库应用可以在另一个附接到数据高速通路的计算机中运行,以便储存当前过程控制例程配置及其相关联的数据。替代地,可以将配置数据库置于与配置应用相同的工作站中。

[0007] 在过程工厂或过程控制系统中,当发生异常状况或故障的迹象时(例如,当生成报警时,或当发现过程测量或致动器具有过量变化时),操作员、仪表技术人员或过程工程师通常使用分析工具结合他或她关于受系统控制的该过程以及其通过该系统的流动路径的知识来尝试确定可能已促成异常状况或故障(fault)的迹象的产生的上游测量结果和过程变量。例如,操作员可以随着时间已从过程控制设备(例如,现场设备、控制器等)的输出捕获的历史数据日志(log)馈送到DeltaV[™]批量分析产品或连续数据分析工具以尝试确定各

个过程变量和/或测量结果对于异常或故障状况的贡献。通常,用户决定哪些历史数据日志和/或其它时间序列数据要馈送到分析工具中并基于他或她关于该过程的知识来识别候选的上游因素(例如,测量结果、过程变量等)。随后,这些数据分析工具使用主成分分析(PCA)或其它分析技术(例如偏最小二乘法、线性回归等)来确定候选的上游因素中的哪些因素影响下游预测的质量参数。因此,分析工具所提供的输出的准确性和有效性基于或受限于用户的知识,并且因此可能不会提供对异常状况或故障的来源的完整或正确的洞察。

[0008] 另外,这种分析通常从过程离线执行,并且因此,当执行分析时,该过程可能改变或移动。例如,典型的过程工厂通常每天执行特定分析的一个或两个循环(例如,数据收集和分析循环),并且仅在已执行了分析之后的某些时间之后是在工厂中分析的结果以及开发和实现的任何规范性动作。因此,不仅分析结果的准确性可能是可疑的,而且从其开发的规范性动作可能不是最优的或者可能不再适用于当前执行的过程。

[0009] 此外,当前已知的过程控制工厂和过程控制系统的架构受有限的控制器和设备存储器、通信带宽以及控制器和设备处理器能力的强烈影响。例如,在当前已知的过程控制系统架构中,控制器中动态和静态非易失性存储器的使用通常被最小化或至少小心地管理。结果,在系统配置期间(例如,先验),用户通常必须选择控制器中的哪些数据要归档或保存,其将被保存的频率,以及是否使用压缩,故控制器被相应地配置有该有限的规则集。因此,在故障排除和过程分析中可能是有用的数据经常不被归档,并且如果其被收集,则有用的信息由于数据压缩而可能已经丢失。

[0010] 还进一步地,工业或过程控制工厂的数据集的规模已稳步增长到目前的数据处理分析应用是不够的。通常,已知的分析技术仅尝试从数据中提取值,但没有解决从中将提取值的特定大小的数据集,并且值得注意的是,没有以无缝方式对非常大的数据集(例如由工厂生成的所有过程数据)进行操作。此外,已知的分析技术通常不能处理流式传输或流式数据。

[0011] 上文所讨论的当前已知的过程工厂监控和分析以及过程控制系统的局限性和其它局限性可能不合希望地在过程工厂或过程控制系统的操作和优化中表现出来,例如在工厂操作、故障排除、和/或预测建模期间。一般来说,利用已知的监控和分析工具,使用实时的、当前的工业过程性能数据的实时分析是不可能的。

发明内容

[0012] 本文公开了用于分布式工业过程性能监控和/或分析的技术、系统、装置、部件和方法。所述技术、系统、装置、部件和方法可以应用于工业过程控制系统、环境、和/或工厂,其在本文中可互换地称为“自动化”、“工业控制”、“过程控制”、或“过程”系统、环境、和/或工厂。通常,这种系统和工厂以分布式方式提供对操作为制造、提炼、变换、生成、或者产生物理材料或产品的一个或多个过程的控制。通常,所述技术、系统、装置、部件、和方法包括以分布式方式将数据监控和/或数据分析引擎(在本文中还可互换地称为“分布式数据引擎”、“数据引擎”、或“DDE”)嵌入在一起协调操作以控制工业过程的设备内。例如,分布式数据引擎可以被制造到与其它设备一起操作以控制在过程工厂或过程控制系统中执行的过程(例如,诸如现场设备、控制器、I/O卡等的过程控制设备),和/或分布式数据引擎可以设备本地耦合或直接地耦合到这种设备。另外的数据引擎可以被嵌入在或制造到包括在过程

工厂中的其它设备中,诸如通信节点、工作站或其它操作员接口、服务器等中。在一些配置中,数据引擎连接到过程工厂内的各种通信链路或以其它方式连接到过程工厂,以便具有在执行或控制工厂内的过程期间传输的实时数据的视图或窗口。

[0013] 数据在各种嵌入式数据引擎之间流式传输,例如,通过使用一个或多个数据通信信道和网络,这些数据通信信道和网络通常存在于常规通信系统外部,在分布式控制系统(DCS)、可编程逻辑系统(PLS)以及过程控制安全和健康监控系统(SHM)中找到这些常规通信系统。为了避免与这种常规已知的过程控制通信网络和链路的混淆,用于在数据引擎之间流式传输通信的通信信道和网络在本文中可互换地称为“数据分析通信信道”、“数据分析信道”、“数据分析通信网络”或“数据分析网络”。流式数据可以包括由数据引擎查看或观察的实时数据。例如,当分布式数据引擎连接到常规的过程控制系统的通信链路(并且因此在查看正在穿越链路的数据)时,数据引擎可以经由数据分析网络向一个或多个其它数据引擎流式传输正在穿越该常规通信链路的数据的副本。在另一个示例中,当分布式数据引擎被制造到或嵌入到过程控制设备中时,由数据引擎使用数据分析网络来流式传输的分析数据可以包括由设备接收、生成或以其它方式处理的数据的副本。另外地或替代地,流式数据可以包括与在设备处本地地执行的一个或多个分析相对应的数据,例如分析结果、规范性动作等。该架构允许本地地绑定分析服务以及接近数据源或者甚至在数据源处提供分析服务,同时提供较大规模的分析,由此提供及时的结果和优化,同时使带宽使用和整个系统的处理循环最小化,如下文更详细解释的。

[0014] 在一个方面中,一种分布式工业过程监控和分析系统包括嵌入在操作为控制过程的过程工厂内的多个分布式数据引擎(DDE)。DDE中的每个DDE都耦合到过程工厂内的相应的一个或多个数据源,该一个或多个数据源分别生成由于该过程受到控制而得到的数据。另外,DDE中的每个DDE储存由与每个DDE耦合的相应的一个或多个数据源生成的数据。该系统还包括数据分析网络,该数据分析网络支持在多个DDE之间流式传输分析数据,并且还支持传输针对储存在多个DDE处的数据的查询。

[0015] 在另一个方面中,一种方法包括:通告包括在过程控制工厂的数据分析网络中的集群(cluster)的存在,其中该过程控制工厂进行操作以控制过程。该方法还包括:注册数据源,该数据源响应于通告并生成由于工厂对过程的控制而引起的连续数据。另外,该方法包括:接收由该数据源生成的连续数据,并且经由数据分析网络将由该数据源生成的连续数据中的至少一些流式传输到消费者。该方法的一个或多个部分可以例如由集群来执行。

[0016] 在又一个方面中,一种方法包括:通告包括在过程控制工厂的数据分析网络中的集中式集群的存在,其中该过程控制工厂进行操作以控制过程。该方法还包括:注册本地集群,该本地集群响应于通告并且被配置为储存作为控制过程的结果而生成的流式传输的连续数据。该本地集群嵌入在过程工厂中并且通信地耦合到一个或多个数据源,其中在进行操作以控制所述过程期间,所述一个或多个数据源生成所述连续数据中的至少一部分。另外,该方法可以包括:从该本地集群接收流式传输的连续数据中的至少一部分,对从该本地集群接收的流式传输的连续数据执行一个或多个数据分析功能,以及经由数据分析网络将一个或多个数据分析功能的输出流式传输到数据消费者。该方法的一个或多个部分可以例如由集中式集群来执行。

[0017] 在另一个方面中,该系统包括绑定到连续数据流的数据分析模块。该连续数据流

具有由包括在过程工厂中的数据源实时生成的内容,其中该过程工厂进行操作以控制过程,并且该连续数据流的内容是作为用于控制所述过程的操作的结果而生成的。该系统还包括用户界面,该用户界面显示由在该连续的数据流上实时地操作的数据分析模块实时生成的连续的输出。显示在该用户界面上的连续的输出包括一个或多个预测值的连续更新。

[0018] 在又一个方面中,一种用于工业过程控制的性能监控和分析系统包括平台,该平台提供一组用户控制和画布,该一组用户控制和该画布使得用户能够创建表示数据模型的数据图。该平台还使得用户能够评估或编译数据图以在输入数据集上执行,从而生成输出数据。该输入数据集包括在线过程工厂对过程进行控制而得到的时间序列数据(以及可选地,其它数据)。如本文所使用的术语“时间序列数据”通常指代通常由一个或多个数据源在一段时间间隔上生成的一系列数据点、值或集合。

[0019] 该数据图包括通过一组线互连的一组数据块,数据经由该一组线在数据块之间进行传输。所述一组数据块中的每个数据块都对应于相应的数据操作,包括零个、一个或多个输入连接器,并且包括零个、一个或多个输出连接器。相应的输入数据经由所述输入连接器在每个数据块处被接收,并且相应的数据块经由所述输出连接器提供由于每个数据块对相应的输入数据执行其相应的数据操作而得到的相应的输出数据。此外,所述输入连接器包括在一组数据块的第一数据块中,包括在所述一组数据块中的至少一个数据块的相应的数据操作包括数据分析功能,并且该数据图的不同部分异步地和独立地可编译和可执行。在一些实施例中,例如,当没有配置输入连接器和输出连接器时,可以由数据块内部地处理输入和输出数据。

[0020] 在一个方面中,一种用于在过程控制环境中执行实时分析的方法包括:在图形用户界面上创建第一图。该第一图表示适用于使得处理器操纵由所述过程控制环境先前生成和储存的数据的第一编程指令。该方法还包括:编译由该第一图表示的第一编程指令,执行经编译的由该第一图表示的第一编程指令,以创建第一结果输出,以及对第一结果输出进行评估,以确定用于创建第一结果输出的模型的预测值。另外,该方法包括:根据该第一图自动地生成第二图。该第二图表示第二编程指令,第二编程指令适用于使得处理器进行以下操作:绑定到实况数据源,并至少使用用于创建第一结果输出的模型来对从所述实况数据源接收到的数据进行评估。此外,该方法包括:编译由该第二图表示的第二编程指令,以及执行经编译的第二编程指令,以预测所述过程控制环境的操作的方面。

[0021] 在另一个方面中,一种用于在过程控制环境中执行实时分析的系统包括:多个过程控制设备,所述多个过程控制设备在过程工厂中进行操作;以及控制器,所述控制器经由控制网络通信地耦合到所述多个过程控制设备。该系统还包括:大数据储存节点,所述大数据储存节点包括对所述过程控制环境的操作的数据进行储存的有形的、非暂时性储存介质;处理器,所述处理器通信地耦合到所述大数据储存节点;以及程序存储器,所述程序存储器通信地耦合到所述处理器。所述程序存储器可以包括储存指令的有形的、非暂时性储存介质,所述指令在被所述处理器执行时使得所述处理器进行以下操作:显示包括编辑画布的图形用户界面;以及接收对多个图形编程元素的选择以便放置在所述编辑画布上,从而创建第一图。该第一图可以表示第一编程指令,第一编程指令适用于使得所述处理器操纵储存在所述大数据储存节点上的数据。另外,当所述指令在被所述处理器执行时,可以使得所述处理器执行经编译的第一编程指令,以创建第一结果输出,并且基于第一结果输出

来创建预测模型。当所述指令在被所述处理器执行时,还可以使得所述处理器根据该第一图自动地生成第二图。该第二图表示第二编程指令,第二编程指令适用于使得所述处理器进行以下操作:绑定到实况数据源,并至少使用用于创建第一结果输出的模型来对从所述实况数据源接收到的数据进行评估。当所述指令在被所述处理器执行时,还可以使得所述处理器编译由该第二图表示的第二编程指令,并且执行经编译的第二编程指令,以预测所述过程控制环境的操作的方面。

[0022] 在另一个方面中,一种用于在过程控制环境中执行数据分析的分析服务包括块定义库。该库可以储存在有形的、非暂时性介质上,并且储存在该库中的每个块定义可以都包括:(i) 用于执行关于所述过程控制环境中数据的动作的目标算法,以及(ii) 一个或多个块属性。该分析服务还包括:机器可读指令集,所述机器可读指令集储存在所述有形的、非暂时性介质上,所述机器可读指令集在被处理器执行时适用于:(1) 经由显示器向用户呈现画布;(2) 向所述用户呈现该块定义库;(3) 接收对一个或多个对应的块定义的一个或多个选择;(4) 将与一个或多个块定义相对应的一个或多个块放置在所述画布上以创建包括所述一个或多个块的模块;(5) 促进针对以下各项中的任意一项或多项的、对一个或多个属性的配置:(i) 所述一个或多个块,(ii) 所述模块,或(iii) 所述一个或多个块以及所述模块;以及(6) 引起对以下各项的评估:(i) 所述块中的一个或多个块或(ii) 所述模块。

[0023] 另外,该分析服务包括在处理器上运行的执行服务,所述执行服务被配置为:(1) 接收用于评估所述一个或多个块或者所述模块的请求;(2) 编译与所述一个或多个块或者所述模块相对应的指令;(3) 创建用于执行经编译的指令的一个或多个作业进程并使得所述一个或多个作业进程执行所述经编译的指令;以及(4) 接收所述一个或多个作业进程的结果。该分析服务还包括通信网络,所述通信网络将执行所述一个或多个作业进程的一个或多个处理器耦合到一个或多个分布式数据引擎。每个分布式数据引擎都可以包括对在所述过程控制环境中所生成的数据进行储存的有形的、非暂时性储存介质。

[0024] 事实上,由过程工厂的操作生成的数据或关于过程工厂的操作的数据可以储存在多个数据储存器中,诸如关系型或非关系型数据库。这些数据储存器可能使用各种数据结构和查询机制,以使得需要不同的查询语法来访问不同数据储存器中的数据。本文描述了标准化查询,以用于促进对使用各种格式的数据储存器的数据访问。利用标准化查询格式的标准化查询包含用于访问数据储存器中的数据所需要的信息,但该标准化查询可以不直接执行来获得此类数据。事实上,基于标准化查询来生成特定于数据源的查询。这可以包括:从标准化查询提取查询参数,以及利用与特定数据源相关联的、特定于数据源的查询格式来生成一个或多个特定于数据源的查询。可以执行特定于数据源的查询以访问其相应的数据源并从其相应的数据源选择数据,所述数据可以随后被格式化以生成数据帧,所述数据帧以任何期望的格式来呈现在标准化查询中指示的数据。这可以包括对齐由特定于数据源的查询获得的、针对数据的采样率、数据属性、或其它数据特性。在一些实施例中,来自利用不同的特定于数据源查询格式的多个数据源的数据可以组合成聚合数据帧以供在过程控制或分析中进一步使用。

[0025] 在另一个方面中,描述了用于从电子数据源获得时间序列数据的方法、系统和储存用于从电子数据源获得时间序列数据的指令的计算机可读介质,包括:接收所述电子数据源的指示,所述数据将从所述电子数据源获得;接收时间范围的指示,所述数据将在所述

时间范围内获得;接收将被获得的所述数据的特性的指示;确定与所述电子数据源相关联的特定于数据源的查询格式;基于所述时间范围的指示和将被获得的数据的特征的指示来生成遵守特定于数据源的格式的特定于所述数据源的查询;使得所述特定于数据源的查询被执行,以从所述电子数据源选择所述数据,和/或从所述电子数据源接收所述数据。时间范围的指示和数据特性的指示以独立于所述电子数据源的所述标准化查询格式来表示,并且该电子数据源可以是关系型数据库或非关系型数据库。

[0026] 该方法包括使用标准化查询格式的标准化查询。时间范围的指示和数据特性的指示在所述标准化查询中被接收。电子数据源的指示也可以在标准化查询中接收。所述标准化查询进一步指示用于所述数据或文件的格式以返回所述数据,譬如,Java脚本对象表示(JSON)。所述标准化查询也可以是JSON格式文件。标准化查询格式可以使用与由所述电子数据源所使用的原生语法(native syntax)不同的语法。由所述标准化查询格式使用的语法无法直接执行以从所述电子数据源获得所述数据。

[0027] 所述时间范围的指示包括至少一个开始时间和至少一个结束时间。所述时间范围的指示类似地包括针对将被获得的数据的采样率。所述将获得的数据的特征的指示包括关于所述过程工厂的操作的一种或多种类型的测量结果的指示。所述一种或多种类型的测量结果可以包括来自所述过程工厂中的一个或多个现场设备的测量值。所述一种或多种类型的测量结果的指示包括一个或多个标签、别名、以及与所述数据相关联的数据类型。

[0028] 该方法还包括接收一格式的指示,所述数据将以所述格式从所述电子数据源接收,并且所述数据可以以所指示的格式从所述电子数据源接收。在一些实施方式中,一格式的指示可以被接收,所述数据将以所述格式被获得,并且将从所述电子数据源接收到的所述数据转换成所指示的格式。

[0029] 特定于数据源的查询格式可以基于电子数据源的类型而确定。产生所述特定于数据源的查询可以包括将时间范围指示和将被获得的数据的特征映射到所确定的特定于数据源的查询格式。使得所述数据源查询被执行可以包括向电子数据源的接口发送特定于数据源的查询,从而所述接口使用所述特定于数据源的查询来查询所述电子数据源。

[0030] 数据帧可以被生成并包括接收自所述电子数据源的数据。所述数据帧可以包括多个数据点。每个数据点与所述时间范围中的时刻相关联。每个数据点还可以与所述时间范围内的、与开始时间之后的整数倍的采样周期对齐的或者为所述开始时间的时刻相关联。

[0031] 在另一方面,描述了一种方法、系统以及储存用于从电子数据源获得时间序列数据的指令的计算机可读介质,包括:从数据请求实体接收使用标准化查询格式的标准化查询;从所述标准化查询提取查询参数;基于所提取的查询参数来生成具有所述特定于数据源的查询格式的特定于数据源的查询;执行所述特定于数据源的查询以从所述电子数据源获得所述数据,和/或向数据接收实体提供所获得的数据。所述标准化查询格式与由电子数据源使用的特定于数据源的查询格式不同。所述电子数据源是关系型数据库或非关系型数据库。所述数据请求实体可以是分析程序中的查询块,譬如本文讨论的数据分析工作室。数据接收实体可以是与数据请求实体相同的实体。所述标准化查询使用无法直接执行以从所述电子数据源获得所述数据的语法。

[0032] 从所述标准化查询提取所述查询参数可以包括确定所述数据的时间周期和数据参数。所述时间周期可以具有开始时间和结束时间,并且所述数据参数可以指示将从所述

电子数据源获得的数据的类型或特性。从所述标准化查询提取所述查询参数还可以包括确定采样率。

[0033] 将所获得的数据提供至所述请求实体可以包括：返回与所述时间周期相关联的数据点集。每个数据点都包括时间戳，所述时间戳指示与所述数据点相关联的一个或多个值的时刻。该时间戳可以被限制为开始时间或是在所述开始时间后的采样率的周期的整数倍。为了获得完整的时间序列，利用时间戳，一个或多个数据点可以被添加到所述数据集，所述时间戳指示在所述开始时间后的采样率的周期的整数倍。所添加的数据点可以使用电子数据源中的数据条目的值，所述数据条目的值与所述一个或多个添加的数据点的每个时间戳之前最接近的时间相关联。所述数据点集可以基于标准化查询中指定的格式而被格式化。

[0034] 由数据参数指示的数据的类型或特性可以一种类型的测量结果或者来自一种测量设备的测量结果。所述数据的类型或特性还可以指示特定的测量设备，其可以是位于过程工厂中的现场设备。所述数据参数还可以指示将被获得的数据的标签、被提供给所述数据请求实体的所述数据的别名、和/或被提供给所述数据请求实体的所述数据的数据格式类型。

[0035] 向所述数据请求实体提供所述数据包括向所述数据请求实体发送包含所获得的数据的数据帧。所述标准化查询可以包括所述数据帧的格式的指示。所述数据帧根据所述数据帧的格式的指示而被格式化。所述标准化查询可以将被提供到所述数据接收实体的数据帧指示为JSON文件。

[0036] 在另一方面中，描述了一种方法、系统和存储用于从多个电子数据源来访问过程工厂的指令的计算机可读介质，包括：接收接收标准化查询，所述标准化查询使用标准化的查询格式；基于所述标准化查询生成特定于第一数据源的查询；基于所述标准化查询生成特定于第二数据源的查询；使得特定于所述第一数据源的查询被执行以从所述第一电子数据源获得第一数据集；使得特定于所述第二数据源的查询被执行以从所述第二电子数据源获得第二数据集；以及生成包括所述第一数据集和所述第二数据集的聚合的数据帧。所述数据的第一数据集和第二数据集均包括由过程工厂中的一个或多个现场设备测量的信息的多个数据点。

[0037] 所述第一特定于数据源指示可以使用与第一电子数据源相关联的第一查询格式，以及所述第二特定于数据源的查询可以使用与第二电子数据源相关联的查询格式。所述标准化查询可以包括所述第一电子数据源的第一指示和所述第二电子数据源的第二指示。在一些实施例中，所述第一电子数据源可以是关系型数据库，所述第二电子数据源可以是非关系型数据库。所述第一指示可以标识所述第一查询格式，所述第二指示可以标识所述第二查询格式。在一些实施例中，所述第一指示和所述第二指示被置于对数据列的指示之前。

[0038] 所述标准化查询格式可以使用与所述第一查询格式的查询语法和所述第二查询格式的查询语法两者不同的查询语法。所述标准化查询格式的所述查询语法无法直接执行来从所述第一电子数据源获得所述第一数据集或从所述第二电子数据源获得所述第二数据集。例如，所述标准化查询包含在Java脚本对象表示(JSON)文件中。

[0039] 所述标准化查询包括对时间范围的指示，所述时间范围标识与所述第一数据集和

所述第二数据集中的数据条目相对应的时间段。所述时间范围可以包括多个时间周期,所述多个时间周期可以通过被所排除的时间段分隔开,针对所述所排除的时间段未查询数据。所述时间范围还可以标识与所述第一数据集相关联的第一时间周期和与所述第二数据集相关联的第二时间周期。所述标准化查询包括所聚合的数据帧的采样率的指示。

[0040] 第一数据集可以包括具有第一采样率的数据,并且第二数据集可以包括具有第二采样率的数据。在这种实例中,生成聚合数据帧可以包括对齐第一和第二采样率。对齐第一和第二采样率可以增加与获得期望的采样率所需要的未采样的时刻相对应的数据点。这种增加的数据点是在时间上直接在所增加的数据点之前的经采样的数据的给定值。对齐第一和第二采样率可以类似地包括移除经采样的数据点。

[0041] 用于实现这些方法的系统还可以包括一个或多个处理器,所述一个或多个处理器通信地连接到一个或多个数据源和一个或多个程序存储器,所述一个或多个程序存储器储存计算机可读指令,所述计算机可读指令在由所述一个或多个处理器执行时使得计算机系统执行上文所描述的系统功能中的部分或全部。计算机可读介质可以储存类似的计算机可读指令,所述计算机可读指令可以由计算机系统的一个或多个处理器执行,从而使得所述计算机系统执行上文所描述的系统功能中的部分或全部。可以实现这些方法、系统、或计算机可读介质的另外或替代的方面,如下文更全面详细描述。

[0042] 另外,公开了一种新颖的频率分析的分析技术,并且该分析技术可以由本文所描述的系统、方法、装置和技术中的任何一种来提供。该新颖的频率分析的分析技术可以分析流式数据以在过程工厂或过程控制系统内提供早期警告故障检测。更具体地说,该频率分析的分析技术可以创建与所识别的故障、异常、性能降低、目标性能水平、不期望的状况、和/或期望的状况的领先指标相对应的新的过程变量集,和/或可以通过对流式传输的过程数据执行滚动FFT来确定新过程变量的时间序列数据。滚动FFT可以将流式传输的过程数据从时域转换到频域,其中在频域中,可以确定新的过程变量集的值。所确定的新过程变量的值可以被变换回到时域以用于监控。在所监控的时域数据内检测到领先指标的存在可以引起预测故障、异常、性能下降、和/或其它不期望的状况的指示被产生和/或呈现,并且可以导致过程工厂的操作的变化,从而避免、防止、和/或减轻不期望的扰乱、事件和/或状况的影响。类似地,在所监控的时域数据内检测到领先指标的存在可以引起所预测的期望的状况(诸如目标性能水平)的指示被产生和/或呈现,并且在一些情况下,连同预测期望的状况发生的时间或时间间隔被产生和/或呈现。

附图说明

[0043] 图1描绘了示出包括分布式、嵌入式工业过程性能监控和/或分析引擎的示例系统过程工厂或过程控制系统的框图;

[0044] 图2A描绘了用于分布式工业过程性能监控/分析的示例系统的框图;

[0045] 图2B描绘了与示例过程控制系统相关的示例分布式工业过程性能监控分析系统的布置的实施例;

[0046] 图3描绘了示例工业过程性能监控和/或分析引擎的框图;

[0047] 图4A描绘了示出各种数据块特征和互连的示例数据图;

[0048] 图4B描绘了示例数据分析工作室(Data Analytics Studio)用户界面;

- [0049] 图4C描绘了数据块定义库、数据块定义、数据模块、数据块、数据块实例和线(wire)之间的示例关系；
- [0050] 图4D描绘了数据模块的草拟(draft)或创建的示例；
- [0051] 图4E描绘了在上面正创建数据模块的数据工作室(Data Studio)画布的放大区域；
- [0052] 图4F-图4H描绘了针对不同类型的数据块的示例数据块定义模板；
- [0053] 图4I描绘了对数据图的评估、编译或部署的示例流程；
- [0054] 图4J描绘了支持标准和定制的可视化的呈现的示例架构；
- [0055] 图4K描绘了其中创建合成的数据块的示例场景；
- [0056] 图4L描绘了示例离线数据图；
- [0057] 图4M描绘了从图4L中的示例离线数据图变换所生成的在线数据图；
- [0058] 图4N-1和图4N-2描绘了示例数据分析仪表板(Data Analytics Dashboard)用户界面；
- [0059] 图4O描绘了控制和优化过程工厂性能的控制回路的高级框图；
- [0060] 图4P描绘了用于提供本地化(localized)数据分析服务的示例方法；
- [0061] 图4Q描绘了用于提供本地化数据分析服务的示例方法；
- [0062] 图5A根据本说明书描绘了分析服务的示例架构的框图；
- [0063] 图5B描绘了更详细示出在图5A中描绘的架构的示例数据服务实体的框图；
- [0064] 图5C根据目前描述的实施例，描绘了示出用于呈现针对放置在编程画布上的块的属性对话框的方法的流程图；
- [0065] 图5D描绘了示例编程画布，其中示例离线图被配置在该编程画布上；
- [0066] 图5E描绘了与图5D的离线图相对应的示例在线图；
- [0067] 图5F描绘了针对模块的示例属性对话框；
- [0068] 图5G描绘了针对离线模块的块的示例属性对话框，在该对话框中，属性与离线和在线操作两者有关；
- [0069] 图6A描绘了示例分析查询系统的框图；
- [0070] 图6B描绘了使用标准化查询的示例分析配置的框图；
- [0071] 图6C描绘了示例标准化查询；
- [0072] 图6D描绘了示例标准化查询方法的流程图；
- [0073] 图6E描绘了示例标准化查询方法的流程图；
- [0074] 图7A描绘了在一时间间隔期间提炼厂的火炬系统的示例过程流程测量的图；
- [0075] 图7B描绘了在另一个时间间隔期间图7A中的示例过程流程测量的图；
- [0076] 图7C描绘了提炼厂的示例过程流程数据的图，其中该提炼厂在该数据中包括火炬事件；
- [0077] 图7D描绘了根据将滚动FFT应用于图7C中的过程流程数据而确定的时间序列数据的图；
- [0078] 图7E描绘了使用PCA来分析图7D中火炬事件附近的频域数据的结果的图；
- [0079] 图7F描绘了使用PCA来分析图7D中另一个火炬事件附近的频域数据的结果的图；以及

[0080] 图7G描绘了用于在过程工厂中提供早期故障检测的示例方法的流程图。

具体实施方式

[0081] 本文中公开了用于分布式工业性能监控和分析的技术、系统、装置、部件和方法。概括地说,所述新颖的性能监控和分析技术提供了关于操作为实时控制一个或多个工业过程的工业过程工厂、过程控制环境和/或过程控制系统的知识发现和可行动的知识。通常,由这种过程工厂或控制系统执行的工业过程在过程工厂或控制系统中受控制,从而制造、提炼、变换、生成或产生物理材料或产品。这种过程工厂或控制系统的示例包括炼油厂、造纸厂、化工制造、制药、食品加工和分销等。

[0082] 在任何过程控制系统、工厂或环境中的最重要的关注是其安全性、可靠性和经济效益。安全操作对于任何过程控制系统或工厂尤为重要,因为不受控制的过程、故障、失效、和/或可防止的人为失误可能导致爆炸、火灾、危险化学品的释放、对环境的损害、设备损失、和/或人员生命损失。设备和过程操作的可靠性对于维持和/或提高工厂的经济效益和盈利能力来说是重要的。此外,通过调整工厂的操作以提高性能,可以提高工厂的经济效率和盈利能力。

[0083] 过程控制系统中发生的故障可能会影响其安全性、可靠性和/或经济效益。当事件相结合以创建过程控制系统并未被设计为进行处理的情形(例如原料组成的未知变化、设备退化、设备失效、以及异常(或错误)的用户操作动作)时,在该过程控制系统中通常发生故障。故障的其它示例包括非计划的设备停机、释放化学品到大气中,压力泄放阀的提升、设备的温度漂移、以及性能退化。当然,故障不限于系统性或大规模的事件;故障可能是具有或可能导致经济、安全和/或环境影响的任何异常事件或事件的组合。

[0084] 通常,当(例如,由过程控制系统自动地)检测到故障时,在操作员界面处生成警报。然后,操作员可以尝试诊断故障的根本原因,并采取纠正动作。因此,在管理故障时的一些重要因素包括早期故障检测和减少假警报,其可以包括例如及时和可靠的故障检测、对故障的根本原因的诊断、以及执行纠正动作,该纠正动作使得是故障源的仪表、控制器、过程、和/或设备回到正常操作。

[0085] 已知的数据监控和分析工具尝试将过程工厂保持在正常工作区域内,并提供了第一道防线抵御故障的发生。但是,这样的工具通常关注于单个回路或过程工厂内的非常有限的范围。它们不能解决典型的现代过程控制工厂的全部(甚至是一个很大的一部分),其中在典型的现代过程控制工厂中,以非常快的速率观察成千上万的过程变量。此外,已知的数据监控和分析工具并不能解决不寻常的事件,如使设备性能退化、移除现场设备、或由每个工具的即时、有限的范围外部的情形造成的操作状况的大变化。

[0086] 另一方面,本文中所公开的该新颖的分布式工业过程性能监控与分析技术、系统、装置、部件、和方法能够解决(例如,监控和/或分析)任何大小范围的工业过程的监控和分析,从整个过程工厂下至单个回路或甚至单个设备。事实上,在一些配置中,本文中所公开的新颖技术能够同时解决多个过程工厂(例如,单个公司或甚至不同公司拥有和经营的多个炼油厂,如果数据是可用的话)。一般来说,本文中所公开的该新颖的性能监测与分析技术发现和提供指示的早期检测和/或可在过程工厂和控制系统中发生的可能的故障的事先警告的知识,从而允许足够的时间来采取规范或纠正动作以防止故障发生。在某些情形下,

本文中所公开的新颖技术还发现和提供规范性可行动知识,以防止可能的故障发生和/或限制其发生的影响。此外,本文中所公开的新颖技术可以发现或提供指示对工厂效率的可能提高的知识,以及发现或提供可行动知识以实现效率提高。

[0087] 具有分布式工业过程监控和分析的示例过程控制系统

[0088] 如上所讨论的,包括或支持本文中所描述的该新颖的分布式工业过程监控和分析技术中的至少某些技术的过程工厂、过程控制系统、或过程控制环境操作为实时控制一个或多个工业过程。因此,过程工厂或控制系统可以在其中包括一个或多个有线或无线过程控制设备、部件、或元件,其执行物理功能(例如开启或闭合阀、测量温度、压力、和/或其它过程和/或环境参数等)以控制在过程工厂或系统内执行的过程。过程工厂或过程控制系统可以包括例如一个或多个有线通信网络和/一个或多个无线通信网络。过程工厂或控制系统可以包括集中式数据库,诸如连续的、批量的和其它类型的历史数据库。

[0089] 举例说明,图1是示例过程工厂或过程控制环境5的详细框图,示例过程工厂或过程控制环境5包括或支持本文中所描述的分布式工业过程监控和分析技术中的任何一个或全部。过程控制系统5包括分布式工业过程监控和分析系统的多个分布式数据引擎,该系统包括在过程控制工厂或环境5中、与过程控制工厂或环境5相集成、或受过程控制工厂或环境5支持(在后续章节中提供对分布式工业过程监控与分析系统的更完整描述)。工业过程监控和分析系统的分布式数据引擎中的每一个分布式数据引擎直接或间接连接(例如,物理连接或无线连接)到过程工厂5的一些部件或部分。例如,分布式数据引擎可以被嵌入在或制造到过程工厂5的特定设备或节点中,从而形成单一的实体,数据引擎可以附接或耦合到工厂5的特定设备或节点,或者数据引擎可以附接或耦合到工厂5的常规通信链路。在图1中,圆圈圈起的“DDE”表示相应的示例分布式数据引擎。

[0090] 如先前所讨论的,分布式数据引擎可以被嵌入在过程控制设备中,该过程控制设备的主要功能是自动生成和/或接收过程控制数据,以执行功能来在过程工厂环境5中实时控制过程。例如,相应的数据引擎可以被嵌入或制造到过程控制器、现场设备和I/O设备中。在过程工厂环境5中,过程控制器接收指示由现场设备获得的过程测量结果的信号,处理该信息以实现控制例程,并生成控制信号,该控制信号通过常规的有线或无线通信链路或网络被发送到其它现场设备以控制工厂5中的过程的操作。通常,至少一个现场设备执行物理功能(例如,开启或闭合阀、增加或降低温度等)以控制过程的操作,并且一些类型的现场设备通过使用I/O设备来与控制器通信。过程控制器、现场设备、和I/O设备可以是有线或无线的,并且任何数量和组合的有线和无线过程控制器、现场设备和I/O设备可以包括在过程工厂环境或系统5中,并且可以均包括相应的分布式数据引擎。

[0091] 例如,图1示出了过程控制器11,过程控制器11经由输入/输出卡26和28通信地连接到有线现场设备15-22,并且经由无线网关35和过程控制数据高速通路或骨干10(其可以包括一个或多个有线和/或无线通信链路,并且可以使用任何期望的或适当的或通信协议(诸如举例来说,以太网协议)来实现)通信地连接到无线现场设备40-46。在一个实施例中,控制器11使用除了骨干10之外的一个或多个通信网络来通信地连接到无线网关35,例如通过使用任何数量的其它有线或无线通信链路,其支持一个或多个通信协议,例如,符合Wi-Fi或其它IEEE 802.11的无线局域网协议、移动通信协议(例如,WiMAX、LTE、或其它ITU-R兼容的协议)、蓝牙®、HART®、WirelessHART®、Profibus、FOUNDATION®

Fieldbus等。骨干10和这些其它通信网络是“常规”的过程控制通信网络的示例,如本文所提及的。

[0092] 控制器11(举例而言,其可以是由Emerson过程管理公司销售的DeltaV™控制器)可以操作为使用现场设备15-22和40-46中的至少一些来实现批量过程或连续过程。在一个实施例中,除了通信地连接到过程控制数据高速通路10之外,控制器11还使用与例如标准4-20mA设备、I/O 26、28和/或任何智能通信协议(例如**FOUNDATION®** Fieldbus协议、**HART®** 协议、**WirelessHART®** 协议等)相关联的任何期望的硬件和软件,来通信地连接到现场设备15-22和40-46中的至少一些。在图1中,控制器11、现场设备15-22和I/O卡26、28是有线设备,并且现场设备40-46是无线现场设备。当然,有线现场设备15-22和无线现场设备40-46可任何其它期望的标准或协议,诸如任何有线或无线协议,包括未来开发的任何标准或协议。

[0093] 图1的过程控制器11包括处理器30,处理器30实现或监督一个或多个过程控制例程38(例如,例程储存在存储器32中)。处理器30被配置为与现场设备15-22和40-46以及与通信地连接到控制器11的其它节点通信。应该注意的是,本文中所描述的任何控制例程或模块(包括质量预测和故障检测模块或功能块)的部分可以由不同的控制器或其它设备实现或执行(如果如此期望的话)。同样,将在过程控制系统5内实现的、本文中所描述的控制例程或模块38可以采用任何形式,包括软件、硬件、固件等。控制例程可以用任何期望的软件格式实现,诸如使用面向对象编程、梯形逻辑、顺序功能图、功能块图,或使用任何其它软件编程语言或设计范式。控制程序38可以储存在任何期望类型的存储器32中,诸如随机存取存储器(RAM)或只读存储器(ROM)。同样,控制程序38可以被硬编码到例如一个或多个EPROM、EEPROM、专用集成电路(ASIC)、或任何其他硬件或固件元件。因此,控制器11可以被配置为以任何期望的方式来实现控制策略或控制例程。

[0094] 在一些实施例中,控制器11使用通常所称的功能块来实现控制策略,其中每个功能块是整体的控制例程的对象或其它部分(例如,子例程),并且结合(经由被称为链路的通信方式)其它功能块来操作以在过程控制系统5内实现过程控制回路。基于控制的功能块典型地执行输入功能、控制功能或输出功能中的一种以执行过程控制系统10中的某种物理功能,该输入功能是例如与变送器、传感器或其它过程参数测量设备相关联的功能,该控制功能是例如与执行PID、MPC、模糊逻辑等等控制的控制例程相关联的功能,该输出功能控制某个设备(例如阀)的操作。当然,也存在混合类型和其它类型的功能块。功能块可以储存在控制器11中并由其执行,这通常是在这些功能块被用于、或被关联于标准4-20mA设备和某些类型的智能现场设备(例如**HART®**设备)的情况下,或者可以储存于现场设备本身中并由现场设备本身实现,这可以是在利用**FOUNDATION®** Fieldbus设备的情况下。控制器11可以包括一个或多个控制例程38,其可以实现一个或多个控制回路,并且可以通过执行功能块中的一个或多个功能块来执行。

[0095] 有线现场设备15-22可以是任何类型的设备,诸如传感器、阀、变送器、定位器等,而I/O 26和28可以是遵循任何期望的通信或控制器协议的任何类型的I/O设备。在图1中,现场设备15-18是通过模拟线路或者组合的模拟和数字线路与I/O卡26通信的标准4-20mA设备或者**HART®**设备,而现场设备19-22是使用**FOUNDATION®** Fieldbus通信协议

通过数字总线与I/O卡28通信的智能设备(例如**FOUNDATION®**Fieldbus现场设备)。但是在一些实施例中,有线现场设备15、16和18-21中的至少一些和/或大数据I/O卡26、28中的至少一些另外地或替代地使用过程控制数据高速通路10和/或通过使用其它适当的控制系统协议(例如,Profibus、DeviceNet、Foundation Fieldbus、ControlNet、Modbus、HART等)来与控制器11(和/或其它大数据节点)通信。

[0096] 如图1中所示,控制器11、I/O卡26和28以及有线现场设备15-16、18-21中的每个都包括如圆圈圈起的“DDE”所指示的、相应的嵌入的分布式数据引擎,该分布式数据引擎经由一个或多个数据分析通信信道和/或网络(在图1中未示出)与其它分布式数据引擎通信。

[0097] 在图1中,无线现场设备40-46经由常规的无线过程控制通信网络70,使用诸如**WirelessHART®**协议之类的无线协议来通信。此类无线现场设备40-46可以直接与过程控制数据分析通信网络112中的也被配置为无线地通信(使用例如该无线协议或另一个无线协议)的一个或多个其它设备或节点进行通信。为了与不被配置为无线地通信的一个或多个其它节点进行通信,无线现场设备40-46可以利用连接到过程控制数据高速通路10或另一个常规过程控制通信网络的无线网关35。

[0098] 无线网关35可以提供对无线通信网络70的各种无线设备40-58的访问。具体而言,无线网关35提供在无线设备40-58、有线设备11-28、和/或过程控制工厂5的其它节点或设备之间的通信耦合。例如,无线网关35可以通过使用过程控制数据高速通路10和/或通过使用过程工厂5的一个或多个其它常规通信网络来提供通信耦合。

[0099] 类似于有线现场设备15-22,无线网络70的无线现场设备40-46可以在过程工厂5内执行物理控制功能,例如开启或闭合阀或获得过程参数的测量结果。然而,无线现场设备40-46被配置为使用网络70的无线协议来通信。因此,无线现场设备40-46、无线网关35、和无线网络70的其它无线节点52-58是无线通信分组的生产者和消费者。

[0100] 在一些场景中,无线网络70可以包括非无线设备。例如,图1中的现场设备48可以是旧式(legacy)4-20mA设备,并且现场设备50可以是常规的有线**HART®**设备。为了在网络70内通信,现场设备48和50可以经由无线适配器或在该处的历史化或52b连接到无线通信网络70。无线适配器52a、52b可以支持其它通信协议,诸如**Foundation®**Fieldbus、PROFIBUS、DeviceNet等。此外,无线网络70可以包括一个或者多个网络接入点55a、55b,该一个或者多个网络接入点可以是与无线网关35进行有线通信的单独物理设备,或者可以与无线网关35作为集成设备来提供。无线网络70还可以包括用于从一个无线设备向无线通信网络70内的另一无线设备转发分组的一个或者多个路由器58。无线设备40-46和52-58可以通过无线通信网络70的无线链路60和/或经由过程控制数据高速通路10相互通信并且与无线网关35通信。

[0101] 如图1中所示,适配器52a、接入点55a和路由器58中的每一个都包括如圆圈圈起的“DDE”所指示的、相应的嵌入的分布式数据引擎,该分布式数据引擎经由一个或多个数据分析通信信道和/或网络(在图1中未示出)与其它分布式数据引擎通信。

[0102] 在一些实施例中,过程控制工厂5包括一个或多个其它无线接入点72,其使用其它无线协议来通信,诸如符合Wi-Fi或其它IEEE 802.11的无线局域网协议、移动通信协议(例如WiMAX(全球微波互联接入)、LTE(长期演进)、或其它ITU-R(国际电信联盟无线电通信部

门)兼容协议)、短波无线电通信(例如近场通信(NFC)和蓝牙)、或其它无线通信协议。通常,此类无线接入点72允许手持或其它便携式计算设备(例如,用户接口设备73)在不同于无线网络70并支持与无线网络70不同的无线协议的相应的常规无线过程通信网络上通信。例如,无线或便携式用户接口设备73可以是由过程工厂5内的操作员使用的移动工作站或诊断测试设备。在一些场景中,除了便携式计算设备之外,一个或多个过程控制设备(例如,控制器11、现场设备15-22、或无线设备35、40-58)还使用接入点72所支持的无线协议来通信。

[0103] 在一些实施例中,过程控制系统5包括通向即时过程控制系统5的外部的系统的一个或多个网关75、78。通常,此类系统是过程控制系统5生成或者进行操作的信息的消费者或者供应者。例如,过程控制工厂5可以包括网关节点75,以通信地连接即时过程工厂5与另一个过程工厂。另外地或替代地,过程控制工厂5可以包括网关节点78,以通信地连接即时工程工厂5与外部公共的或私有的系统,诸如实验室系统(例如,实验室信息管理系统或者LIMS)、操作员巡回数据库、材料处理系统、维护管理系统、产品库存控制系统、生产调度系统、天气数据系统、装运和处理系统、封装系统、互联网、另一提供方的过程控制系统、或者其它外部系统。

[0104] 如图1所示,接入点72、用户接口设备73和网关75每一个都包括如圆圈圈起的“DDE”所指示的、相应的嵌入的分布式数据引擎,该分布式数据引擎经由一个或多个数据分析通信信道和/或网络(在图1中未示出)与其它分布式数据引擎通信。

[0105] 要注意的是,虽然图1仅示出了包括在过程工厂5中的单个控制器11与有限数量的现场设备15-22和40-46、无线网关35、无线适配器52、接入点55、路由器58、以及无线过程通信网络70,但这仅是说明性和非限制性的实施例。任何数量的控制器11可以被包括在过程控制工厂或系统5中,并且任何数量的控制器11可以与任何数量的有线或无线设备和网络15-22、40-46、35、52、55、58和70通信以控制工厂5中的过程。

[0106] 示例分布式工业过程性能监控/分析系统架构

[0107] 图2A包括用于分布式工业过程性能监控/服务的示例系统100的框图,该系统还在本文中可互换地称为数据分析系统(DAS)。DAS 100可以结合例如图1中的过程工厂5来操作,并且为方便说明在本文中是如此论述的。然而,要理解的是,DAS 100的至少一部分可以结合除过程工厂5之外的其它过程工厂和/或过程控制系统来操作。

[0108] 通常,DAS 100支持本地化性能监控和/或分析,同时支持用于过程工厂环境的大规模(例如,全系统和/或跨过程工厂5的多个设备或节点)性能监控、数据挖掘、和数据分析。为此,系统100包括多个分布式数据引擎(DDE),在图2A中用附图标记102a-102e以及在图2B中用附图标记102f-102h指示DDE的示例。如本文所使用的,附图标记“102x”指代DDE 102a-102h中的任何一个或多个。图2A中示出的分布式数据引擎的至少一些对应于图1中示出的分布式数据引擎。例如,图2A中的数据引擎102a是包括在图1的无线网关35中的数据引擎,并且图2A中的数据引擎102b是嵌入在图1的控制器11中的数据引擎。另外,系统100包括图1中未示出的其它分布式数据引擎102。例如,分布式数据引擎102c连接到数据高速通路10,分布式数据引擎102d嵌入在过程工厂5的集中式大数据设备108,并且分布式数据引擎102e嵌入在大数据云节点110中,其中大数据云节点110可以服务过程控制工厂5的数据分析需求(以及在一些配置中,还可以服务其它过程控制工厂)。当然,系统100并不仅限于五个数据引擎102a-102e或八个数据引擎102a-102h,但还可以包括任何数量的分布式数据引

擎,这些数据引擎中的至少一些被嵌入在或制造到其相应的数据源(例如,过程工厂5的相应的过程控制设备)和/或这些数据引擎中的至少一些以其它方式连接到过程工厂5的某种其它数据源(例如,部件、部分等)。

[0109] 如上所提到的,通常分布式数据引擎并不通过使用常规过程控制通信网络中的任何一个网络(例如,图1的骨干10、无线网络70、在I/O卡28与设备19-22之间的有线连接等)来与其它分布式数据引擎通信,但是在一些配置中,一些分布式数据引擎102x可以通过使用常规过程控制通信网络来向其它数据引擎传输至少一些信息。但是,一般来说,数据引擎102x通过使用与常规过程控制网络分开和不同的一个或多个数据分析通信网络212来与其它数据引擎102x通信。分析数据在网络112的分布式数据引擎或节点之间进行流式传输,例如,通过使用分析流式传输服务、流式传输和/或排队协议,和/或通过使用支持流式传输的消息代理(messaging broker)或系统115,诸如定制流式传输源、Flume、HDFS、ZeroMQ、Kafka、Microsoft Message Bus、MQTT、AMQP、RabbitMQ等,诸如在前述的标题为“STREAMING DATA FOR ANALYTICS IN PROCESS CONTROL SYSTEMS”的美国申请No.14/506,863中描述的专门定制的历史对象通信协议或另一个适当的通信协议。数据分析通信网络112的多个分支在图2A中用虚线表示,但为了清楚起见,没有完整描绘网络112。因此,在一个实施例中,数据分析通信网络112的至少一部分可以在某种意义上是在工厂5内的常规过程控制通信网络之上的叠加网络。例如,包括DDE 102x和数据分析网络112的物理设备中的至少一些可以位于过程工厂5的场所内,并且在一些情况下,可以位于过程工厂5的设备之中。

[0110] 在另一个实施例中,数据分析通信网络112中的至少一部分可以实现在常规过程控制通信网络旁边或大致平行于常规过程控制通信网络而实现,如图2B中所示。在图2B中,过程工厂5包括通信地连接到常规过程控制通信网络的各种过程控制设备或部件,例如,以帮助控制工厂5内的一个或多个过程。此类过程控制设备/部件中的三个各自具有嵌入在其中并与其相关联的相应的DDE 102f、102g和102h,并且DDE 102f、102g和102h中的每一个都通信地连接到数据分析网络112。然而,大多数支持和/或包括数据分析网络112的设备没有位于过程工厂5的设备之中,而是相反,可以远程地位于企业数据中心处。

[0111] 在另一个实施例中,数据分析通信网络112的至少一部分可以实现为过程工厂5内的逻辑网络。在该实施例中,例如,可以在相同的物理通信链路上传输过程控制数据和分析数据两者,该物理通信链路在逻辑上具有独立链路的表现(apppearance),例如,一个常规过程通信链路和一个数据分析通信链路。

[0112] 不管数据分析网络112的实现如何,然而数据分析网络112与过程控制系统5内的设备或节点处的常规过程控制通信网络相交,所述设备或节点经由常规过程控制通信网络向过程控制系统5内的其它设备或节点传送信息,并且分布式数据引擎(例如,数据引擎102a、102b、102f、102g、102h)嵌入在其中,和/或借助于连接到常规过程控制通信网络以便具有针对在上面传输数据的窗口的分布式数据引擎向过程控制系统5内的其它设备或节点传送信息。

[0113] 数据分析网络112可以使用任何期望的通信技术和/或通信协议。数据分析网络112的至少一些部分可以使用有线技术,并且数据分析网络112的至少一些部分可以使用无线技术。但是,一般来说,数据分析网络112可以支持数据分组传输协议和/或技术,诸如IP或以太网上的其它异步数据分组协议。

[0114] 在较高的级别或层处,数据分析网络112可以使用具有流式传输能力的协议、处理实时数据馈送的消息代理、和/或处理实时数据馈送115的消息传送系统、诸如Apache Kafka。如图2A所示,数据分析网络112所提供的流式传输能力115允许分布式性能监控/分析系统或DAS 100本地地绑定各种分析服务以及接近数据源提供各种分析服务(例如,接近或位于其中嵌入分布式数据引擎102x的、过程工厂5的设备或节点,诸如图1中的节点11、18、26、28、72、35、52a、55a、72)。同时,系统100的流式传输能力115还允许系统100提供较大规模的预测和优化,因为仅特定的源生成的数据(其对于支持更广泛或更高级别的数据分析是必要的)需要传输给接收方数据引擎。例如,如果数据引擎102d正在仅对由过程工厂5的控制器中的每个控制器生成并在特定配置变化在每个控制器处被实例化之后的一小时内捕获的特定的数据执行特定数据分析,则嵌入在控制器11处的数据引擎102b仅流式传输所需要的数据到引擎102d(例如,在特定配置变化被行实例化之后的该一小时期间,仅流式传输由控制器11生成的输出数据到数据引擎102d,但不流式传输由控制器11生成的其它输出数据)。

[0115] 如图2所示,分布式性能监控/分析系统或DAS 100的部分本地地设置在过程控制环境或工厂5内。例如,分布式数据引擎102a、102b、102c和102d位于过程工厂5的场所内。还如图2A中所示,系统100的其它部分被远程地设置。例如,数据引擎102e位于云(cloud)110中。其它数据引擎102x可以被设置在一个或多个远程位置处(未示出),诸如在对多个过程工厂或过程控制系统进行服务的服务器组(server bank)处。当然,在一些配置中,系统100可以省略本地地设置的数据引擎,或者可以省略远程地设置的数据引擎。也就是说,系统100可以完全在对象过程控制环境5的场所内(例如,本地地)、完全在对象过程控制环境5的场所外(例如,远程地)、或通过使用场所内和场所外分布式数据引擎的组合(例如,本地地和远程地两者)来提供性能监控和分析(例如,经由一个或多个分布式数据引擎)。

[0116] 用于工业控制系统的示例类型的性能监控和分析

[0117] 通常,工业控制系统的性能监控和分析包括:采集由对象控制系统生成的数据(例如,当系统正在操作控制一个或多个过程时,实时生成的数据),并且对所采集的数据中的至少一些执行一个或多个分析功能或操作,以确定关于控制系统执行得如何的知识,以及在一些情况下,以确定可以采取来提高系统的性能的动作。用于工业过程控制系统的数据分析可以被认为是三种通用类别:描述性分析、预测性分析和规范性分析。下面对这些类别的分析进行一般性论述。

[0118] 描述性分析允许用户发现在过程控制系统或工厂内发生了什么、何时发生、多久发生一次、以及因发生了什么而引起的什么问题。通常,对从监控过程系统或工厂得到的数据(例如,后验)执行描述性分析,并且该描述性分析可以使用诸如基本计算、缩放、和标准统计之类的技术。

[0119] 规范性分析允许用户优化过程控制系统或工厂内的操作。例如,规范性分析允许用户回答诸如以下的问题:什么是最佳的回答?在给定的不确定度下什么是最佳的结果?什么是显著不同和更佳的选择?预测性分析可以识别、监控和控制工业过程控制工厂和系统中的过程操作的重要质量变量或重要指标。另外,预测性分析可以识别:在给定一组未来输入或因果条件的情况下过程操作的重要质量变量或重要指标将发生什么。预测值随后可以由规范性分析用于生成规范性动作。通常,规范性分析使用诸如基于约束的优化和多目标

优化之类的技术,然而,已知的数据监控和分析工具是初步的并严重限于其范围。

[0120] 例如,利用已知的数据监控和分析工具,用户和操作人员可以通过使用基于约束的优化工具来优化工厂的单件设备或单元,该基于约束的优化工具对从监控该单件设备或单元得到的数据(例如,后验)进行操作。然而,由于过程控制系统或工厂包括多件设备和过程单元,因此整体过程可能由于多个单元之间的过程交互而仍然远未达到最优。用于交互单元的多目标优化技术或跨过程控制系统的一部分的协调优化在已知的数据和分析工具可获得(例如,基于第一原理的建模),但由于增加的过程复杂度,这种协调优化技术是困难的并且通常花费太长时间来执行(如果它们甚至可能执行的话)来成为有用的。例如,已知的多目标优化工具对从监控多件设备或单元得到的数据(例如,后验)进行操作,以识别最优操作点并且确定用于使当前操作点移动(maneuver)到最优点的轨迹。然而,数据收集的每个循环以及对其的计算可能花费很长的时间,以致于在识别规范性动作的时候,过程已移动远离其操作,以致于所识别的规范性动作不再是最优的,或者可能甚至是无效的、适得其反的、和/或危险的。此外,当用户尝试通过限制输入到多目标优化工具中的变量的数量来缩小或缩减计算的规模和范围(以及因此计算所需要的时间)时,用户作出关于哪些变量被选择为输入的人为判断,这不仅限制范围,而且可能是不正确的,并且因此可能产生误导或不正确的结果。

[0121] 此外,已知的数据监控和分析工具通常不能按需提供重要质量变量的标识和值(例如,由于可用的测量技术的限制)并且不能及时提供相关的结果。例如,当前使用的、用于过程控制系统的预测性分析的技术包括离线技术(例如,在实验室中)或在线技术(例如,通过分析器),这两种技术具有其相应的缺点。重要质量变量的离线确定是不令人期望的,因为在实验室测试中引起的显著延迟将使任何输出对于当前过程状况而言呈现为次优的或者甚至无效的(例如,测量的信号不能用作反馈信号)。使用在线分析器来确定重要质量变量也是不那么令人期望的,因为已知的在线分析器是有问题的、昂贵的,并需要频繁的高成本的维护。此类限制可能对产品的质量、废品的产生、以及操作的安全性具有严重影响。

[0122] 此外,对于大规模监控情形,常规的、集中式性能监控方案具有严重限制。适当的性能监控系统应当具有故障容忍能力、操作可靠性、以及经济效益,所有这些在集中式性能监控方案应用于大规模监控情形时都受到损害。具体而言,集中式监控器同时照顾过程的所有单元故障。一旦在一单元中发现故障,集中式监控器在用于检测在相同时间段发生的、来自其它单元的另外故障的能力上变得受限制,因为集中式监控器在其计算中使用所有测量的变量,并且当一个变量不可用或一个通信信道阻塞时,这个监控系统可能停止运转。另外,集中式监控器的实时能力受变量的最低采样率所限制,这降低了监控系统的效率。此外,工厂中的不同变量可以具有非常不同的动力学尺度(甚至在同一单元内),并且单个监控方案通常不应用于整个过程或单元的所有变量,尤其是,单个监控方案对于当变量彼此影响(例如,跨交互的过程单元)时的情形来说是次优的。更进一步地,当在过程控制系统中存在地理分布时(例如,各种过程单元之间存在长距离,诸如在石油管线中),通常每个单元装备单独的性能监控器,从而提供本地的(而非整体的)性能监控,以及传输延迟、数据丢失、以及电池受限问题。

[0123] 不同于当前在使用中的初步的、单变量的、和/或集中式预测监控和分析技术,并且不同于描述性和规范性分析的后验本质,本文中所描述的用于工业控制系统和工厂的、

新颖的分布式性能监控和分析系统、方法、装置、部件和技术允许用户监控在过程控制系统或工厂中在任何时间点当前在发生什么,以及基于目前数据来预测接着或后续可能发生什么。本文中所公开的新颖技术允许用户回答诸如以下的问题:“现在在发生什么”、“接着将发生什么”、“如果这些趋势继续会怎样?”,具有很少或不具有关于过程数据的特性的用户或人为偏见。

[0124] 本文中所描述的用于工业控制系统的分布式性能监控和分析的新颖技术通常包括平台(例如,图2A中示出的系统架构100)和在该平台之上或结合该平台来操作的应用。平台使用利用设备和过程知识来增强的数据驱动的方法,以支持例如在过程控制系统或工厂(例如,过程控制系统5)正在操作为控制一个或多个过程之前、期间和之后对该过程控制系统或工厂的估计、监控和数据挖掘。在一些实现方式中,平台可以支持多个过程控制系统或工厂。

[0125] 在平台上或结合平台来操作的性能监控/分析应用可以包括:例如,推断的测量、设备监控、故障检测、过程预测、因果性、其它监控应用、和/或其它分析应用。可以由应用使用的技术包括数据挖掘、优化、预测建模、机器学习、仿真、分布式状态估计等。因此,性能监控/分析应用可以用于监控、预测、和诊断任何数量的、过程控制系统5的任何部分的性能退化和故障,包括在诸如仪表、设备、控制、用户交互、和过程之类的方面中。

[0126] 因此,本文中所描述的分布式工业性能监控和分析技术可以包括推断机制,该推断机制根据容易获得的过程测量值来提供对重要过程变量的连续在线估计。为了在延长的时间段上维持数据模型,系统还可以支持用于监控、调整和增强数据模型的能力。在最简单的情况下,可使用的方法包括取消调整(detune)模型(例如,将正常操作范围增加至预测 $\pm 3s$)。更复杂的示例包括模型切换和模型自适应。

[0127] 事实上,作为研究和原型测试程序的部分,发明者使用本文中所描述的技术来建立原型工业过程性能数据监控和数据分析框架,以解决在存在非线性、非高斯分布的情况下软传感器开发问题,而不管数据采样、过程约束、和/或其它因素如何。在这些原型和研究中,执行对由用于特殊化学品过程、批量过程、和烃过程的多个软传感器生成的数据的监控和分析,并且所得到的知识可以用于提高所获得的过程。我们的工业伙伴已报告了显著的经济利益和环境利益。在后续章节中描述我们的研究和原型的示例。

[0128] 现在返回图2A,分布式工业性能监控和分析系统或DAS 100提供用于在支持针对过程工厂环境(例如,图1的过程工厂5)的大规模数据挖掘和数据分析的同时支持本地化性能监控所需要的工具、配置和接口。例如,系统100可以包括:用于配置和开发数据分析模型的用户界面应用(例如,数据分析工作室)、用于执行模型的运行时(runtime)引擎(其可以整体地或部分地对由过程控制系统生成的实时数据进行操作)、以及用于显示分析结果的相同用户界面应用或另一个用户界面应用(例如,运行时仪表板)。系统100可以对多种类型的数据源进行操作或支持多种类型的数据源,包括如先前所提到的实时值(例如,实时的连续值)、事件收集、批量数据收集、操作员巡回数据、LIMS数据、诸如供应链和/或天气数据之类的外部数据、以及与过程控制系统相关联的任何其它类型的数据,包括结构化和非结构化数据两者。系统100可以提供一组“开箱即用”的标准分析,诸如描述性统计信息、直方图、相关图等。另外,系统100可以为用户提供结构化环境以创建期望的分析以及查看源数据和通过应用多个不同数据分析而生成的n-阶数据两者,以识别过程控制系统的各个数据集内

的隐含关系、和/或预测产品能力、质量和其它特性。通常,系统100可以提供关于过程控制系统的知识发现以及可行动知识(例如,以提高或优化过程控制系统的性能、纠正和/或预测故障、增加可靠性、提高安全性、增加经济效益等),其可以被表示为数据模型。

[0129] 如图2A中所示,系统100可以是高度分布式的,例如,可以包括多个分布式数据引擎102x。嵌入式数据引擎102x的网络112可以互连多个(例如,数百个、数千个、甚至数百万个)过程控制系统信息的传感器和源。如先前所提到的,数据引擎102x聚集(cluster)在过程控制系统的设备和节点(例如,制造设备、过程控制设备、通信节点、材料处理系统、实验室系统、工厂的用户、以及甚至过程本身)附近、在其处或在其内。因此,分布式工业性能和分析系统100的嵌入式本质紧密地耦合到物理过程控制工厂。然而,将数据引擎102x嵌入和集成到过程控制系统设备和节点的制造结构中可能使得数据引擎102对于端用户几乎不可见。通常,数据引擎102x是小型的、无线连接到网络112中、带宽受限的,并且在诸如有限的能量和需要足够的散热之类的物理约束下操作,因为其可能嵌入到诸如温度传感器或其它类型的传感器之类的过程控制设备中。

[0130] 如先前所提到的,分布式数据引擎102x在网络112上通过使用流式传输协议和/或排队协议来与其它数据引擎102x互联。每个分布式数据引擎102x可以支持一个或多个数据监控和/或数据分析应用。应用的连接的集群的绝对数量需要使用统计上正确的(而不是确定性的)算法来进行资源核算(account)、故障检测和纠正、系统管理等,并且每个集群可以托管本地化需求感兴趣的功能。

[0131] 因此,分布式工业性能监控和分析系统100可以增加过程控制工厂的安全性、效率和生产率。例如,系统100可以精确地控制过程参数,从而降低了过程制造的总成本。此外,将系统100集成到过程控制环境5中可以通过严格的过程控制和实时质量保证来实现更佳的产品质量和更少的废品;由于可编程子系统而实现更灵活、快速配置的生产线;实现系统健康监控,这引起更有效的、预防的和低成本的维护;由于更佳的监控和控制而实现更安全的工作环境;以及实现更佳的部件组装技术,诸如通过使用智能RFID标签,仅举不计其数的益处中的几个例子。

[0132] 此外,分布式工业性能监控和分析系统100可以将用户和过程控制系统5之间的人机交互提高到生成过程工厂5内的实时的、人加机器控制回路的程度。例如,改进的人机交互可以通过确保不存在操作员/维护/可靠性错误,以及通过减少事故,来提高质量和生产率。此外,通过由系统100提供的先进的控制、冗余的、智能报警、自诊断和维修,可以实现过程控制工厂5的服务的可用性、可靠性和连续质量。

[0133] 示例工业性能数据监控/数据分析引擎

[0134] 现在转到分布式工业性能分布式数据监控和分析引擎102x,通常,分布式数据引擎102x可以是数据分析通信网络112中的收集、观察、获取、接收、处理、储存、缓存和/或分析所有或大多数过程控制相关的数据(例如,大数据)的节点,该过程控制相关的数据由一个或多个数据源(例如,由其中嵌入数据引擎102x的设备或部件或由数据引擎102x所耦合或所附接的部件)生成或以其它方式观察到。在一些情形下,分布式数据引擎102x可以生成另外的数据(例如,其执行的分析的结果),和/或可以发送或转发选定的数据到数据分析网络112的其它节点。如本文中可互换地所使用的术语“过程控制大数据”、“过程大数据”和“大数据”通常指代由包括在过程控制系统或工厂(例如,过程控制系统或工厂5)中并与过

程控制系统或工厂相关联的设备和/或部件生成、接收、和/或观察的所有(或几乎所有)数据,特别是,当过程控制系统或工厂正在实时执行以控制一个或多个过程时生成、接收、和/或观察的所有(或几乎所有)数据。在一个实施例中,由包括在过程工厂5中并与过程工厂5相关联的所有设备生成、创建、在设备处接收、或以其它方式观察的所有数据(包括所有过程数据和所有分析数据)被收集和储存为数据分析通信网络112内的大数据。在过程工厂和过程控制环境中,大数据的这种收集和分析对于提高安全性、可靠性和经济效益是重要的,因为时间的维度和特定数据点的存在或忽略可能是关键的。例如,如果特定的数据值没有在这个时间间隔内被传送到过程工厂的接收方部件,则过程可能会变得不受控制,这可能会导致火灾、爆炸、设备的损失、和/或人员生命的损失。此外,在过程工厂内和/或过程工厂外部操作的不同部件、实体、和/或过程之间的多个和/或复杂的基于时间的关系可能会影响操作效率、产品质量、和/或工厂安全。

[0135] 由分布式数据引擎102x生成、收集、观察、获取、接收、储存、缓存、处理、分析、和/或转发的过程控制大数据可以包括已被直接用于控制工厂5内的过程的数据或根据控制工厂5内的过程而生成的数据,例如,由诸如控制器、输入/输出(I/O)设备和现场设备生成或使用的实时和配置数据。另外地或替代地,数据引擎102x可以生成、收集、观察、分析、储存、接收、获取、缓存、和/或转发与在过程工厂5内传送和路由这种第一阶过程控制数据和其它数据有关的数据,例如,与工厂5中的数据分析通信网络112和/或其它通信网络的网络控制有关的数据、指示带宽的数据、网络接入尝试、诊断数据等。此外,一些数据引擎102x可以生成、收集、观察、储存、缓存、获取、接收或转发数据,该数据指示通过处理和/或分析其已收集的过程控制大数据而在过程控制数据分析通信网络112内已学习到的结果、学问、和/或信息。通常,这样的分析结果、学问、和/或学习到的信息是根据由一个或多个数据引擎102x执行的分析来生成的。

[0136] 因此,分布式数据引擎(例如,DDE 102x)在本文中可互换地称为“大数据设备”、“大数据设备节点”或“设备节点”。在大多数情况下,大数据设备节点的分布式数据引擎102x包括多核硬件(例如,多核处理器),以用于实时发送和接收大数据(例如,经由流式传输),并且在一些实施例中,以用于缓存和/或储存实时大数据以便在过程控制数据分析通信网络112上进行后续流式传输或其它传送。因此,分布式数据引擎102x还包括存储器(例如,高密度存储器),以用于缓存和/或储存大数据。可以由数据引擎102x发送、接收、流式传输、缓存、收集、储存、获取、接收、缓存、和/或以其它方式观察的实时数据的示例可以包括诸如测量数据、配置数据、批量数据、事件数据、和/或连续数据之类的过程控制数据。例如,可以收集与配置、批量配方、设定点、输出、速率、控制动作、诊断、警报、事件和/或它们的变化相对应的实时数据。实时数据的其它示例可以包括过程模型、统计信息、状态数据、网络、和工厂管理数据、以及分析结果。可以在前述的美国专利申请No.13/784,041、No.14/174,413和No.14/212,493中找到对可以与本文所描述的技术中的任何技术或所有技术一起使用的各种类型的示例大数据设备及其部件的描述,但要理解的是,本文中所描述的技术中的任何技术或所有技术可以与其它适当的大数据设备一起使用。

[0137] 典型地,分布式数据引擎102x通常对已由过程工厂或系统5内一个或多个数据源生成或提供的和/或以其它方式与过程工厂或系统5的实时操作相关联的大数据进行操作。例如,DDE 102x可以对数据进行收集和加时间戳,该数据是在嵌入DDE 102x的设备处接收

的和/或由该设备生成,或者该数据是在DDE 102x对应的数据源处接收的和/或由该数据源生成。所收集的数据可以(至少临时地)储存在数据引擎102x的本地存储器中。在一些情形下,可以使用专门的历史对象通信协议(诸如在前述的标题为“STREAMING DATA FOR ANALYTICS IN PROCESS CONTROL SYSTEMS”的美国申请中No.14/506,863所描述的)或另一个适当的通信协议或诸如Kafka之类的消息传送系统来传送或流式传输数据中的至少一些到一个或多个其它数据引擎102x。

[0138] 在一些实施例中,一个或多个数据引擎102x支持在包括实时连续值、事件收集、批量数据收集、操作员巡回数据、和/或其它数据的多维数据上的大规模数据挖掘和数据分析。分布式数据引擎102x可以被配置为对其本地收集的数据和/或其它DDE 102x收集的数据执行一个或多个数据分析。例如,DDE 102x可以包括对结构化数据(例如,储存在存储器中的时间序列和表数据、关系型和/或非关系型数据库、或流式传输的数据)以及非结构化数据(例如,pdf)进行操作的各种工具。另外,DDE 102x可以支持任何一个或多个期望的目标环境并执行其中相应的目标代码(例如,Java、C#、R脚本、Python脚本、**Matlab®**脚本、Statgraphics等)。DDE 102x可以执行学习算法(例如,偏最小二乘回归、主成分分析等)、分类技术(例如,随机森林、模式识别等),和/或其它数据分析,以便生成诸如预测产品能力、质量和/或其它期望的特性之类的结果和/或有用信息。本地分析的结果可以储存在数据引擎102x的本地存储器处,并且本身可以被视为另外的大数据。此外,一些数据引擎102x可以包括用于配置和开发模型的界面、用于执行模型的运行时引擎、以及用于在用户界面处显示结果的仪表板。例如,分布式数据引擎102x可以以诸如在前述的标题为“DATA PIPELINE FOR PROCESS CONTROL SYSTEM ANALYTICS”的美国申请No.62/060,408中论述的和/或如在本公开内容的稍后章节中论述的方式来配置有益于运行时的分析。所显示的结果可以包括标准的描述性统计信息、直方图、相关图、和/或其它可识别各种数据集内的隐含关系的数据表示。

[0139] 在一些情况下,分布式数据引擎102x被嵌入在、制造到、直接耦合到主机数据源(例如,图2A中示出的数据引擎102a、102b和102c)、或以其它方式与该主机数据源驻留在一起。在一些情况下,分布式数据引擎102x可以是过程控制数据分析通信网络112的单独大数据节点(例如,图2A中示出的数据引擎102d和102e)。也就是说,在这些情况下,数据引擎102x没有嵌入在过程控制系统或工厂5的数据源中或没有与该数据源驻留在一起,但可以其它方式观察已由一个或多个数据源生成的数据(例如,当分布式数据引擎耦合到常规过程通信链路时,诸如数据引擎102c)。无论是嵌入式的或是单独的,数据引擎102x分析已由一个或多个数据源本地生成和/或提供的大数据,以发现或学习知识。可以将该学习到的知识储存在数据引擎102x处,由数据引擎102x本地地对该知识进行操作,和/或将该知识作为大数据提供或发送给其它数据引擎102x,例如接收方大数据节点。另外地,分布式数据引擎102x可以将其已知或储存的大数据的一部分提供给分析网络112的其它数据引擎102x和/或其它节点(例如,本地或远程用户界面节点)。

[0140] 图3包括示例分布式工业过程性能数据监控和/或数据分析引擎150的简化框图,该引擎的实例可以被包括在图2A中的过程控制数据分析通信网络112中(例如,分布式数据引擎102x)。参考图3,示例分布式数据引擎150包括用于缓存、储存、和/或历史化大数据的大数据储存区域155、一个或多个大数据设备接收器160、以及一个或多个大数据设备请求

服务者165。大数据设备接收器160中的每个被配置为接收和/或观察来自一个或多个数据源168的数据。在一个示例中,经由通向常规过程控制通信网络(例如数据高速通路10、Fieldbus网络、WirelessHART网络等)的网络接口,大数据设备接收器160接收和/或观察在常规过程控制通信网络上传播的数据。另外地或替代地,大数据设备接收器160可以经由本地接口从其对应的一个或多个数据源168接收数据,例如从被嵌入或制造有DDE 150或DDE 150本地地、直接地或以其它方式紧密地所耦合的过程控制设备或其它类型的设备。另外地或替代地,大数据设备接收器160可以例如经由数据分析网络接口175接收大数据分组。所接收到的大数据分组可能已从另一个DDE 150流式传输和/或可能已由与数据引擎150驻留在一起的大数据源生成。不管所接收到/观察的数据的一个或多个源168如何,大数据设备接收器160对所接收到的/观察的数据分组和/或消息进行处理,以获取其中携带的实质数据和时间戳,并将该实质数据和时间戳储存在数据引擎150的大数据储存区域155中,例如,作为时间序列数据和可选地还作为元数据。如本文中所使用的术语“元数据”通常指代关于数据的数据,诸如关于数据的统计信息、关于数据的类别信息、概要信息、描述性信息、定义等。大数据储存节点区域155可以包括多个本地和/或远程物理数据驱动器或储存实体,诸如RAID(独立磁盘冗余阵列)储存器、固态储存器、云储存器、高密度数据储存器、和/或任何其它适当的数据储存技术,该储存技术适用于数据组或数据中心储存器,对于其它节点而言表现为单个或单体的逻辑数据区域或实体,并且可以被配置为本地地储存和/或历史化大数据。

[0141] 大数据设备请求服务者165中的每个都被配置为访问储存在大数据设备储存区域155中的时间序列数据和/或元数据,例如,每次发生诸如本地或远程数据分析应用、用户界面应用、或另一个应用之类的请求实体或应用的请求时。例如,大数据设备请求服务者165可以在任何给定的时间间隔将储存在大数据设备储存区域155中的选定数据发布到该选定数据的订阅者。在另一个示例中,大数据设备请求服务者165可以每次发生远程执行数据分析应用的请求时,获取储存在本地大数据设备储存区域155内的数据。将在后续章节中论述访问储存在DDE 150处的数据的另外示例。可以以多种数据格式存储这种数据,包括关系型或非关系型数据库或其它数据结构。在一些实施例中,标准化查询格式可以用于访问这些数据源中的任何数据源中的数据。

[0142] 在一些实施例中,分布式数据引擎150包括一个或多个大数据分析器170以对储存在存储器155中的大数据的至少一部分执行相应的数据分析和/或学习。本地分析和/或学习的执行可以是响应于由用户或由另一个节点生成的命令或指令而执行的。另外地或替代地,本地分析和/或学习的执行可以以自动和/或自主方式来执行而无需使用来自用户或其它节点的任何输入来发起和/或执行学习分析。例如,数据分析和/或学习可以以诸如先前所论述的方式、以诸如在前述的标题为“DATA PIPELINE FOR PROCESS CONTROL SYSTEM ANALYTICS”的美国申请No.62/060,408中所论述的方式、或者以某种其它适当的方式来执行。在一个实施例中,大数据分析器170单独地或共同地对所储存的数据执行大规模数据分析(例如,数据挖掘、数据发现等)以发现、检测或学习新的信息或知识。数据挖掘通常涉及检查大量的数据以提取新的或先前未知的感兴趣的数据或模式(诸如不寻常的记录或多组数据记录)的过程。大数据分析器170还可以对所储存的数据执行大规模数据分析(例如,机器学习分析、数据建模、模式识别、预测分析、相关分析等)以预测、计算或识别所储存的数

据内的隐含关系或推断。

[0143] 在一个实施例中,多个大数据分析器170(和/或至少一个大数据分析器170的多个实例)可以并行地和/或合作地操作以分析储存在分布式数据引擎150的大数据储存区域155中的数据和/或分析储存在其它分布式数据引擎102x的一个或多个其它大数据储存区域中的数据。此外,作为一种类型的合作的数据分析和学习,多个大数据分析器170可以彼此之间共享、交换或传输经计算的参数和模型信息。多个大数据分析器170可以共同驻留同一大数据节点上,或者可以驻留不同大数据节点上。可以在前述的标题为“DATA PIPELINE FOR PROCESS CONTROL SYSTEM ANALYTICS”的美国专利申请No.62/060,408中找到可以与本文所描述的技术中的任何技术或所有技术一起使用的合作的数据分析的示例,但是任何适当的合作的数据分析技术可以与本公开内容的任何方面或所有方面一起使用。由大数据分析器170执行的分析的结果可以储存在大数据设备储存区域155中,和/或可以被返回给请求实体或应用。

[0144] 在一个实施例中,大数据接收器160、大数据设备请求服务者165和/或大数据分析器170的至少一部分包括在或实现在一个或多个集成电路、半导体、芯片或其它适当的硬件上。例如,执行频谱分析的大数据分析器170可以由包括在大数据节点中的集成电路芯片来实现,诸如在前述的标题为“AUTOMATIC SIGNAL PROCESSING-BASED LEARNING IN A PROCESS PLANT”的美国申请14/507,252中所描述的。在一个实施例中,大数据接收器160、大数据设备请求服务者165和/或大数据分析器170的至少一部分包括储存在存储器上并可由运行在分布式数据引擎150上的过程执行的计算机可执行指令。例如,大数据接收器160、大数据设备请求服务者165和/或大数据分析器170的至少一些部分包括储存在一个或多个非暂时性有形存储器或数据储存设备上的相应的计算机可执行指令,并且所述指令可由一个或多个处理器执行以执行相应的大数据功能中的一个或多个。

[0145] 在一些实施例中,至少一些大数据分析器170未包括在分布式数据引擎150中,而是替代地与分布式数据引擎150共同驻留在同一主机数据源设备或部件上并且与数据引擎150通信连接。例如,数据引擎150(包括储存区域155、接收器160和服务者165)可以由第一计算机可执行指令集来实现,并且大数据分析器170可以由半导体芯片或由第二计算机可执行指令集来实现,第二计算机可执行指令集可以与第一计算机可执行指令集储存在相同的非暂时性有形存储器或数据储存设备上或其可以不与第一计算机可执行指令集储存在相同的非暂时性有形存储器或数据储存设备上。在一些实施例中,大数据分析器170没有包括在数据引擎150中,并且没有与数据引擎150共同驻留同一主机数据源设备或部件上,但是尽管如此仍与数据引擎150通信连接。例如,大数据分析器170可以位于数据分析网络112的另一个节点上,该节点不是DDE 150,例如,用户接口节点或分析服务器。

[0146] 图3中的分布式数据引擎150还包括一个或多个网络接口175,该一个或多个网络接口175被配置为允许数据引擎150在数据分析网络112上发送和接收有效载荷大数据,以及与数据分析网络112的其它数据引擎和节点通信(例如,信令传送或其它类型的通信)。例如,数据引擎150可以通过使用网络接口175来订阅由另一个节点发布的一种类型的数据,由发布节点生成并被数据引擎150订阅的该类型的数据可以经由网络接口175来接收。

[0147] 如先前所论述的,可以由分布式数据引擎102x生成、收集、观察、获得、接收、储存、缓存、处理、分析和/或流式传输各种类型的数据,诸如过程相关的数据、工厂相关的数据、

以及其它类型的数据。过程相关的数据可以包括当过程在过程工厂5中正在受到控制时生成的连续、批量、测量和事件数据(并且,在一些情况下,这些数据指示过程的实时执行的影响)。此外,过程相关的数据可以包括过程定义、布置或设置数据,诸如配置数据和/或批量配方数据、与配置相关联的数据、执行和结果相对应的数据等。

[0148] 工厂相关的数据(诸如与过程工厂5有关的数据,但其可以不是由直接对过程工厂中的过程进行配置、控制或诊断的应用生成)可以由分布式数据引擎102x生成、收集、观察、获得、接收、储存、缓存、处理、分析和流式传输作为大数据。工厂相关的数据的示例包括振动数据、汽阱数据、指示与工厂安全性相对应的参数的值的数据(例如,腐蚀数据、气体检测数据等),指示与工厂安全性相对应的事件的数据、与机器的健康相对应的数据、与工厂中的资产(诸如工厂装置和/或设备)相对应的数据、与装置、机器和/或设备诊断的配置、执行和结果相对应的数据、以及对于诊断和预报有用的数据。

[0149] 其它类型的数据,包括与过程工厂5的过程控制大数据网络骨干和各种通信网络有关的数据高速通路业务和网络管理数据、用户相关的数据(诸如与用户业务、登录尝试、查询和指令有关的数据)、文本数据(例如,日志、操作程序、手册等)、空间数据(例如,基于位置的数据)以及多媒体数据(例如,闭路电视、视频剪辑等),可以由数据引擎102x生成、收集、观察、获得、储存、缓存、处理、分析和/或流式传输作为大数据。

[0150] 在一些实施例中,动态测量和控制数据可以由分布式数据引擎102x自动地生成、收集、观察、获得、储存、缓存、处理、分析和/或流式传输作为大数据。动态测量和控制数据的示例包括指定过程操作中的变化的数据、指定操作参数的变化的数据、过程和硬件警报和事件的记录(诸如下载或通信失败等)。此外,当检测到变化时或当初始地将控制器或其它实体添加到数据分析通信网络112时,可以默认地自动收集诸如控制器配置、批量配方、警报和事件之类的静态数据。

[0151] 此外,在一些场景中,当检测到元数据的变化时,在分布式数据引擎102x中捕获描述或识别动态控制和测量数据的至少一些静态元数据。例如,如果影响模块或单元中的测量和控制数据(其必须由控制器发送)的控制器配置发生变化时,则由数据引擎102x自动地捕获相关联的元数据的更新。另外地或替代地,可以由数据引擎102x自动地捕获与用于对来自外部系统或源的数据(例如,天气预报、公共事件、公司决策等)、监视数据、和/或其它类型的监控数据进行缓冲的特殊模块相关联的参数。

[0152] 在一些情形下,在分布式数据引擎102x中自动地捕获由端用户创建的添加的参数。例如,端用户可以在模块中创建特殊的计算或者可以向单元添加需要收集的参数,或者端用户可能想要收集默认情况下不传送的标准控制器诊断参数。可以以与默认参数相同的方式来传送端用户可选地配置参数。

[0153] 用于工业性能监控/分析的用户界面应用

[0154] 如先前所提到的,数据分析系统或DAS 100可以包括一个或多个用户界面应用,经由该一个或多个用户界面应用来进行数据相关的人机交互。在图2A中用附图标记120a-120d描绘这些用户应用的示例性实例的存在。本公开内容的本章节和图4A-图4Q更详细描述了用于工业性能监控/分析的用户界面应用,例如,该用户界面应用可以由图1的过程工厂或系统5、图2A的工业性能监控/分析系统100、和/或图3的监控和分析引擎150来提供或结合图1的过程工厂或系统5、图2A的工业性能监控/分析系统100、和/或图3的监控和分析

引擎150来操作。然而,本文所描述的用于工业性能监控/分析的用户界面应用可以由用于过程控制系统的其它工业性能监控和分析系统来提供或结合用于过程控制系统的其它工业性能监控和分析系统来操作。但是,为容易论述,下文同时参照图1、图2和图3来论述工业性能监控/分析用户界面应用。另外,为容易阅读,工业性能监控/分析用户界面应用在本文中使用大写来引用,以区分通用和/或其它用户界面应用,并且在本文中还可互换地称为“数据分析用户应用”、“数据分析用户界面应用”、“数据分析应用”、“DDE用户界面应用”或“用户界面应用”。

[0155] DDE用户界面应用提供界面,用于用户与分布式工业过程性能监控/分析系统或DAS 100交互以定义结构、查询数据、建立和评估草拟的数据模型。当草拟的数据模型被最终确定时,DDE用户界面应用允许将数据模型下载到运行时引擎并部署为结合在线过程控制系统来操作。可以经由DDE用户界面应用的运行时仪表板来访问和监控部署的数据模型(还称为执行的或在线数据模块)。DDE用户界面应用还能够生成与执行数据模型相对应的警报和通知。

[0156] 具体而言,DDE用户界面应用使得用户能够创建、查看和修改数据模型,这些数据模型中的每个模型定义将对输入数据集执行或实施的一个或多个数据分析(例如,描述性、预测性、和/或规范性分析)。在离线模式中草拟和创建数据模型(例如,当数据模型与在线或操作过程控制系统中的实况数据源断开连接时),并且在该模式中,数据模型在本文中称为“离线数据模块”。通常,离线数据由离线数据图定义,用户经由DDE用户界面应用,通过选择一组“块”或“数据块”,并在图上以期望的方式利用一组“线”互连这些块,来创建离线数据图。每个数据块包括输入端,经由该输入端,数据被接收到块中。每个数据块还表示或定义要由每个数据块对其输入数据执行的特定功能、动作、算法和/或操作,从而生成可以由一个或多个输出端向其它块提供的输出数据。每个数据块能够被单独地评估,以使得块及其互连线的离线图表示被编译成可执行代码并被执行,并且在离线数据图上呈现每个数据块的执行的结果。如将在稍后章节中论述的,将数据块编译成代码以及其后续的执行可以跨越不同目标环境和位置来分布式地执行。

[0157] 类似地,还能够在整体上对离线数据图进行评估。离线数据图的评估包括编译数据块和线、执行经编译的代码以跨越线传输输入数据和输出数据以及执行如由离线数据图的块和线的特定配置定义的功能、动作、算法和/或操作。同样类似于单独的数据块,离线数据图的编译以及经编译的离线数据图的执行可以跨不同目标环境和位置来分布式地执行。

[0158] 此外,DDE用户界面应用使得用户能够将离线数据模块变换成“在线数据模块”,以使得数据模型的在线数据模块绑定到或以其它方式接收由在线过程控制系统的实况数据源生成的实况(例如,流式传输)数据,对实况数据执行其中定义的一个或多个数据分析,并且将输出提供给用户界面、历史库或其它应用。例如,由数据模型生成的输出数据可以包括与在其中控制的过程工厂和/或过程相对应的描述性、预测性、和/或规范性信息或数据。

[0159] 具体而言,用户可以将特定数据模型的在线数据图变换成特定数据模型的在线数据图。类似于离线数据图,在线数据图包括通过一组线互连的一组数据块,并且由于离线和在线数据图通常对应于相同数据模型,因此由在线数据图定义的功能、动作、算法和/或操作对应于离线数据图的定义的功能、动作、算法和/或操作。然而,块和互连中的至少一些在离线数据图和在线数据图之间有所不同,主要(但不必仅)为了适应将在线数据模块连接到

在线过程工厂。

[0160] 同样类似于离线数据图,在线数据图可以在整体上编译成与数据模型的在线数据模块相对应的可执行代码。在线数据图的特定块的编译导致了将特定块与在线过程工厂内的相应数据源和数据消费进行绑定的定义。对线数据模块的部署实例化这些绑定,并且使得可执行文件运行或执行,从而使在线数据模块与在线过程工厂相集成,以使得在线数据模块结合在线过程工厂的操作来执行。在线数据图的编译以及所得到的在线数据模块的绑定和执行可以跨不同目标环境和位置来分布式地执行。

[0161] 在特定有用的实施例中,数据模型的在线数据模块接收因对在工厂中执行的在线过程的进行中的控制而由过程工厂的设备和部件生成的连续、实时数据并且至少部分地对该数据进行操作。例如,当过程工厂和过程控制系统正在操作为控制过程时,在线数据模块对由在线过程工厂和/或由包括在工厂中的过程控制系统生成的连续、实时时间序列数据进行操作。数据模型对连续、实时数据流进行持续地操作(例如,通过执行在其上如数据模型所定义的一个或多个数据分析功能或其它功能(如果有的话)),并且连续地生成实时的结果或输出数据流,该实时的结果或输出数据流可以显示在用户界面处(例如,作为滚动线形图或其它描绘)并且可以持续地反映数据模型的输出的当前值、实时值。在一个示例中,由数据模型生成并在用户界面处显示的数据输出包括至少一个预测值及其在时间上的变化。但是,一般来说,出于对过程控制工厂或环境5的故障检测、预测和规范化的目的,数据模型能够对由过程工厂或环境5内的多种数据源生成的大量数据进行操作。

[0162] 在一个实施例中,DDE用户界面应用是基于网页的并且通过网页浏览器来访问,以使得各种平台(例如,Apple Macintosh、Linux、Windows等)以及各种用户在各种计算设备处可以使用该应用的不同实例(有时候同时地使用)。然而,DDE用户界面应用不限于基于网页的实现方式,并且可以使用任何其它适当的实现方式(其是独立于平台的并且可以同时扩展到多个用户和/或实例)。

[0163] 此外,虽然本文将DDE用户界面应用描述为应用于单个过程工厂或环境5,其中DDE用户界面应用的多个实例同时执行,但该配置仅是说明性的,并不意在进行限制。例如,在一些配置中,DDE用户界面应用可以应用于在不同位置和对不同过程独立地进行操作的多个过程控制工厂或环境。例如,石油公司的多个炼油厂可以使用由服务器或计算机组提供的单个DDE用户界面应用,以使得DDE用户界面应用的一些实例在提炼厂A处执行以及一些实例在提炼厂B处执行。

[0164] 无论如何,如先前所论述的,DDE用户界面应用将可以由数据块对相应的输入数据执行的功能、动作、算法和/或操作抽象为块或数据块。块或数据块的示例是由其相应的功能命名的,例如,加载、清除、操纵、分析、可视化等。每个数据块可以加载或获得相应的输入数据,对所获得的输入数据执行其一个或多个相应的功能、动作、算法、和/或操作,以及生成一个或多个相应的结果或相应的输出数据。多个数据块可以如所期望的进行互连(例如,串行、并行、多对一、一对多等)以形成表示或定义草拟的数据模型或离线数据模块的离线数据图,并且可以通过使得离线数据模块(或其部分)应用于一组一个或多个数据集或离线数据资源或对一组一个或多个数据集或离线数据资源进行操作来执行或评估该草拟的数据模型/离线数据模块。例如,在离线数据模块的执行或评估期间,将离线数据图编译成可执行代码,获得指定的数据集并将其加载到离线模块中或供离线模块使用,并且执行经编

译的离线模块的互连的块中的每一个以对其相应的输入执行相应的操作并在其相应的输出提供相应的结果,从而产生由与正在开发或创建的数据模型相对应的整体离线数据模块生成的一个或多个分析输出或结果。在一个实施例中,在离线评估期间,可以从诸如测试工具、环境或工厂之类的流式传输的离线数据源接收输入数据,而不是从文件将数据集加载到草拟的数据模型中。此外,如下文将更详细解释的,可以在开发离线数据模块或草拟的数据模型时逐步地对其进行评估。在DDE用户界面应用中,用于评估离线数据块和离线数据模块的平台或基础架构是数据分析服务,将在稍后章节中对其进行描述。

[0165] 在已评估了离线数据模块(或其部分)之后,用户能够查阅和细察评估的结果,并且可以相应地修改并重新评估草拟的模型(或其部分)直到达到了满意的结果为止。用户可以最终确定草拟的数据模型/离线数据模块,将其变换成在线数据模块,并且执行或部署在线数据模块以对由于过程控制环境或工厂5的实时操作而生成的实时、在线、或流式、实况数据进行操作。例如,执行或部署的在线数据模块或数据模型可以连续地对由于过程工厂的在线操作而连续地生成的实时数据进行操作,并且在线数据模块或模型自身可以连续地生成其执行的实时输出数据或结果,该实时输出数据或结果可以用户界面上进行显示和连续地更新。可以在DDE用户界面应用的仪表板处监控所部署的在线数据模块的在线执行或操作及对应的结果,也在稍后章节中对其进行更详细描述。

[0166] 如先前所论述的,在DDE用户界面应用内,数据块、数据模块(无论在线还是离线)、以及其部分可以被视觉地和/或图形地表示给用户作为在由DDE用户界面应用提供的工作空间或画布上的图或数据图。通常,每个数据图包括通过一组线(其可以用线条表示)来互连的一组数据块(例如,其可以用二维形状表示)。DDE用户界面应用通常包括两种操作模式(但是,在一些实施例中,可以包括更多或更少数量的模式),这两种操作模式都使用数据图来表示数据模块和包括在数据模块中的数据块。

[0167] 这些操作模式中的一种在本文中可互换地称为“数据分析工作室”、“数据工作室”、或“工作室”。通常,用户使用数据工作室来设计和开发数据分析模型。另外,数据工作室允许用户将离线数据模块变换成在线数据模块,以及将最终确定的离线模块部署为对应的在线数据模块。数据工作室提供工作区或绘图画布,在该工作区或绘图画布上用户能够通过创建和连接块来将数据模型开发成数据图,数据图是数据模型或其部分的视觉、图形表示。当数据模型在数据工作室内部被开发时,其可以称为草拟的数据模型或离线数据模块,并且可以被认为是草图或原型。离线数据模块图可以比其得到的在线数据模块包括更多数量的数据块和/或连接,因为在离线模块中所包括的块和连接中的一些可以用于分析和查看模块的不同部分处的数据,例如,以便分析和检查数据模型的特定部分是否正在如所期望的评估和/或提供足够有用的和/或预测的数据。通常,数据模块的离线图是工作流的表示,该工作流(1)探索和清除原始数据集,以及(2)可以针对诸如分类、回归、聚集、维度缩减等的期望的分析操作来被建立、调整、以及评估。可以在开发离线的数据模块时对其进行逐步地或持续地评估。离线数据模块或数据模型的评估过程的结果经由数据工作室来传送给用户。

[0168] 在一些情况下,离线数据模块的执行或评估可能花费很长的时间段来完成。在这些情形下,离线数据模块的评估过程的状态和进度可以经由DDE用户界面应用的另一种操作模式来传送给用户,该另一种操作模式在本文中可互换地称为“数据分析仪表板”、“分析

仪表板”、或简称“仪表板”。仪表板通常由用户用于查看和/或管理以下的模块：(i) 在被评估的离线数据模块(例如,草拟的数据模型),和/或(ii) 在线数据模块(例如,最终确定的数据模型,其已被部署为数据模块)。仪表板还表示有待评估的草拟的数据模型以及使用相应的图或视觉、图形表示的部署的在线数据模块。部署的数据模块称为“在线的”,因为该部署的模块正在对由于过程控制系统或工厂5的在线或运行时操作而生成的实时数据执行操作。通常,数据模块的在线图是工作流的表示,该工作流被绑定到过程控制系统或工厂5内的一个或多个数据源(例如,流式传输数据源)以进行实时的描述、预测和/或规范化、和/或在运行时期或在运行时持续地调整数据模型。部署的或在线的数据模块可以持续地执行直到其从仪表板显式地终止。

[0169] A. 数据模块图

[0170] 数据模块(无论离线还是在线)的示例图通常包括多个数据块和将各个块互连在一起以形成一个或多个数据流动管道的线。如先前所提到的,数据块通常是对用户想要应用于数据集的功能或操作的抽象。例如,特定的块可以加载来自磁盘上的数据储存或文件的数据集,另一个特定的块可以替换数据集中的所有缺失的值(例如,在与另一个参数/变量具有测量值的时间相对应的时间处不具有测量值的那些值),又一个特定的块可以执行随机森林分析等。通常,不同数据块对应于可以对数据集执行的不同功能或操作,因此,各个数据块各自可以具有对应的类型或名称,例如“加载”、“查询”、“填充”、“列”、“PCA(主成分分析)”、“PLS(偏最小二乘)”、“探索”、“写入”等。

[0171] 每个块可以不具有属性或具有多个相应的属性。块的属性集(对于一些块来说其可以是空集)分别对应于其块类型,使得相同块类型的所有实例具有相同的属性集。对于一些块,可以由DDE用户界面应用提供属性的默认值,并且对于一些块,可以允许用户输入和/或修改属性值中的一个或多个。数据块定义了相应的属性和数据块类型的任何默认属性值(以及对于一些块,还有其它信息)。数据块定义储存在由DDE用户界面应用提供的数据块定义库中。通常,数据块定义库可用于DDE用户界面应用的所有开放实例,使得多个数据分析用户或工程师可以同时使用库所提供的资源来开发和/或探索数据。

[0172] 每个块具有零个、一个、或多个输入连接器,其指定从一个或多个其它块或数据源流入该块的数据(如果有的话)。另外,每个块具有零个、一个、或多个输出连接器,其指定流出该块(以及潜在地经由一个或多个接收方块的相应的输入连接器流入到该一个或多个接收方块)的数据(如果有的话)。通过线在数据图上表示各个块的输入端与输出端之间的互连。任何类型的数据可以沿着任何线流动或传输,从简单的比例因子值到各自包含针对目标代码的数百万个值的数据帧。

[0173] 另外,每个块具有状态。例如,当块被首先创建时,该块处于“配置”或“配置中”或“未配置”状态。在块已被配置之后,该块转移到“已配置”状态。当离线块正在被评估时,其处于“评估”状态。在离线块已被评估之后,该块转移到评估“成功的”状态或评估“失败的”状态。如果离线块正在进行配置或绑定到数据源和/或数据消费者,则其状态是“配置中”或“绑定中”。在在线块已被部署到过程控制系统之后并正在对实况数据进行操作,其状态是“已部署”或“执行中”。当然,其它状态也是可能的。通常,每个块包括指示其当前状态的视觉指示器(indicator)。另外,每个数据模块在整体上可以具有指示其当前状态的指示器。

[0174] 图4A示出了描绘了各种可能的数据块特征和互连的示例数据图200。该示例数据

图200包括加载数据 (LoadData) 块202a,其经由线205连接到填充NaN (FillNaN) 块202b。加载数据块202a的表示包括其当前块状态的指示208a、其当前块类型或名称的指示210a、块的评估的结果的指示212a、以及输出连接器215a,通过该输出连接器215a,作为加载数据块的评估的结果而生成的数据中的至少一些数据经由线205被传送到接收方 (在该场景中,传送到填充NaN块202b)。

[0175] 填充NaN块202b包括输入连接器218b,输入连接器218b接收通过线205从加载数据块202a流动的数据。类似于加载数据块202a,填充NaN块202b包括其当前块状态的指示208b、其当前块类型或名称的指示210b、其评估的结果的指示212b、以及输出连接器215b,通过该输出连接器215b,作为加载数据块的评估的结果而生成的数据中的至少一些数据经由线220传送到接收方或消费者 (未示出)。

[0176] 注意,在图4A中,加载数据块202a不具有输入连接器,因为加载数据块202a没有接收从数据图200的任何其它数据块输出的数据。相反,加载数据块202a可以被配置为从一个或多个数据源加载或获得一个或多个输入数据集。所述数据源可以包括离线数据源,诸如数据文件、数据源 (例如,Seeq系统)、关系或非关系数据库,和/或所述数据源可以包括在线或流式传输数据源,诸如由分布式数据引擎202x生成的数据流。

[0177] B. 数据分析工作室

[0178] 如先前提及的,DDE用户界面应用的操作模式中的一种操作模式是数据分析工作室。数据分析工作室可以由用户用来设计、开发、查看、和探索数据模型。图4B描绘了由数据分析工作室240呈现的示例性用户界面,其包括导航条242和工作空间或绘图画布245。导航栏242提供控制和指示器,经由该指示器,用户能够管理离线和在线数据模块,例如,通过允许用户执行动作,例如创建新的离线数据模块、识别当前在画布245中打开并示出的离线或在线模块、容易地查看当前在画布245中打开和示出的数据模块的状态 (例如,离线或在线)、保存/储存当前在画布245中打开和示出的离线数据模块、将离线模块转变成在线数据模块、在查看数据模块的离线和在线数据图之间切换、评估离线数据模块、部署在线数据模块、浏览至其它数据模块、以及其它这些模块管理功能。因此,数据分析工作室在其上包括众多用户控制和指示器248a-248n,例如但不限于:

[0179] • 模块导航控制248a以使得用户能够找到并浏览至其它数据模块;

[0180] • 当前在画布245上打开的数据模块的标识符248b;

[0181] • 指示当前在画布245上打开的数据模块的视图是离线视图还是在线视图的一个或多个指示器248c、248d;

[0182] • 一个或多个控制248e、248f,经由该一个或多个控制248e、248f,用户可以在当前在画布245上打开的数据模块的在线视图与离线视图之间进行切换;

[0183] • 用户控制248g,经由该用户控制248g,用户可以查看和/或定义当前在画布245上打开的数据模块的属性;

[0184] • 用户控制248h,经由该用户控制248h,用户可以保存当前打开的数据模块;

[0185] • 用户控制248i,经由该用户控制248i,用户可以评估当前打开的数据模块的至少一部分;

[0186] • 用户控制248j,经由该用户控制248j,用户可以部署当前打开的数据模块;

[0187] • 指示器248k,该指示器248k指示当前打开的模块的操作状态;和/或

[0188] • 一个或多个其它用户控制和/或指示器(未示出)。

[0189] 图4B还示出了用户控制248m,经由该用户控制248m,用户可以查看数据块定义库(在图4B中未示出)、从数据块定义库中选择块定义、和/或将块定义添加到数据块定义库。在图4B中示出的实施例中,库用户控制248m被示出为位于画布245上,然而,在其它实施例中,所述控制248m可以位于导航栏242上或者在任何期望的位置处。

[0190] 事实上,在示例的数据分析工作室240中示出的用户控制和指示器的数量、类型、位置/定位、形成因素、布置等等仅仅是许多可能的实施例中的一个。可以包括更多或更少数量和/或类型的用户控制和/或指示器。可以利用这些用户控制和/或指示器的不同位置/定位,以及不同的形成因素、布置、等等。在某些实施例中,可以省略导航栏242,并且可以提供用户访问控制的另一个机制(例如,弹出式窗口、下拉菜单、等等)。

[0191] 1. 数据分析工作室—离线数据模块

[0192] 在数据分析工作室240中,绘图画布或工作空间245是经由其可以显露、定义、和评估离线数据模块的区域。例如,在绘图画布或工作空间245上,用户能够访问DDE用户界面应用的块定义库(例如,经由用户控制248m),选择储存在其中的各数据块定义、以及互连(例如,连线在一起)所选定的块定义以形成数据模型的数据图。此外,在绘图画布工作空间245上,用户能够编辑特定的数据块实例的属性;评估草拟的数据模型的一部分,包括查看评估的进展及结果;和/或执行与离线数据模块相关的其它动作。

[0193] 如先前所讨论的,可用在数据图中的每个数据块是基于数据块的类型的块定义的。也就是说,给定类型的块的特定实例具有根据给定类型的块定义的一组属性,但是特定实例可以例如以与对象类别和对象类别的实例类似的方式与相同的给定类型的块的其它实例不同。此外如先前所讨论的,块定义储存在块定义库中,该块定义库经由用户控制248m来进行访问(例如,切换到查看或隐藏),用户控制248m可以位于画布245上。

[0194] 在图4C中示出了数据块定义库、数据块定义、数据模块、数据块、数据块实例、与线之间的示例关系250的例示。如图4C中示出的,通过DDE用户界面应用提供的块定义库252而使块定义可用。在某些实施例中,对特定块定义和/或对其它数据分析库资源(和/或对它们的部分)的不同访问权限(例如,只读、读写、等等)可以被给予不同的用户和/或用户组。

[0195] 储存在库252中的块定义255可以用于创建在数据模块260内使用的块255的实例258。可以针对块实例258具体定义块定义255的一个或多个属性262。块实例258被配置为或者被设计为包括一个或多个输入连接器265,经由该输入连接器265,数据被接收到块实例258中,并且块实例258可以被配置为或设计为包括一个或多个输出连接器270,经由该输出连接器270,输出数据(例如,由于对在经由输入连接器265接收到的输入数据进行操作的块实例258进行评估得到的数据)被提供给例如另一个块实例258,被写入或储存到用户界面、等等。块实例258的每个输入连接器265可以通过数据模块260的一个或多个线272接收数据,并且块实例258的每个输出连接器270可以通过数据模块260的一个或多个线272提供数据。数据模块260的每个线272可以提供特定块实例258的特定输出连接器270与另一个特定块实例258的特定输入连接器265之间的互连,由此使得数据能够在两个数据块实例之间传输。

[0196] 在实施例中,块定义通过功能或动作的类别被组织在库252内。类别可以包括子类别,子类别可以包括子子类别、等等。至例示性的而非限制性的示例中,块定义库252包括至

少五个块定义类别：数据源、过滤器、变换器、可视化、和数据消费者。

[0197] 包括在数据源类别中的块定义通常定义与数据源有关的各种动作，该数据源提供了数据模块对其进行操作的数据。例如，数据源类别中的块定义可以包括“创建数据集”以生成输入数据集、“加载数据集”以从离线或在线数据源加载或获得现有的数据集、“加载DB”以从数据库（例如，过程控制数据库或分析数据库）加载或获得数据、“保存数据集”以将输入数据集历史化到长期数据存储器中（例如，在其创建之后，例如用于测试输入数据集）、等等。此外，数据源类别中的某些可以包括一个或多个子类别。例如，创建数据集和加载数据集类别可以针对特定的数据集格式和/或其中生成数据集的环境各自包括相应的子类别，例如，CSV（逗号分隔值）、树莓派、Seeq、等等。通常，离线数据源块被配置为获得或访问静态数据源，例如数据文件、过程控制数据库、分析数据库、等等。另一方面，在线数据源块通常被配置有使在线数据源块的特定实例与过程工厂内的一个或多个在线数据源相关的绑定定义。对经配置的、在线数据源的编译使绑定实例化，从而在线数据源块接收由绑定到该在线数据源块的一个或多个在线数据源生成的流式传输的数据。

[0198] 数据源类别可以包括与从一个或多个数据源（例如，“查询数据源”或“查询”）搜索或获得选定数据相关的块定义。查询块可以对离线数据源和/或在线数据源进行操作。通常，查询块定义允许数据模块查询、获得、或请求数据的特定类型或特性（例如，如由列、标签、或其它适当的标识符所指示的）和/或查询、获得、或请求仅在特定的时间间隔期间生成的数据，该特定的时间间隔可以是不相交的时间间隔。再进一步，查询块能够查询任何类型的数据而不考虑格式和/或捕获或储存数据的环境。在本申请的后面部分中提供了查询块的另外的细节。

[0199] 包括在过滤器类别中的块定义通常定义了可以在数据集上执行的各种过滤技术。例如，过滤器类别中的块定义可以包括“削波(clip)”、“高通”、“低通”、“SGF”（例如，用于Savitsky-Golay滤波）、“指数型”、“平均型”、“微波”、等等。

[0200] 包括在数据块定义库252的“变换器”类别中的块定义通常定义操作、分析、和/或以其它方式变换数据集的内容或经过滤的数据集的内容的各种技术。例如，变换器类别可以包括与清洁的输入数据集相对应的子类别数据块定义，例如“填充NaN”（例如，以填充在不是数字的数据集的条目中）、“移除异常值”（例如，以移除异常数据）、“校正不良值”（例如，以校正在输入数据集内检测到的不良数据）、“排除不良值”（例如，以排除检测到的不良数据）、等等。此外，变换器类别可以包括与操纵输入数据集的内容相对应的子类别数据块定义，例如“缩放”、“移动”、“分离”、“合并”、“中心正常”、“降低采样”、“时间偏移”、“列”等等。此外，在某些实施例中，变换器类别可以包括与对齐输入数据集内的数据相对应的子类别块定义，例如“时间延迟”、“滞后校正”、等等。

[0201] 变换器类别可以包括与分析输入数据集以获得知识和学习它们的内容相对应的子类别数据块定义，例如“灵敏度”、“集群”、“随机森林”、“CBP”（例如，有条件的贝叶斯概率分析）、“K均值”、“傅里叶变换”、“快速傅里叶变换”、“PLS”、“PCA”、等等。变换器类别中的某些子类别可以包括一个或多个子子类别。例如，变换器类别中的“PCA”子类别可以包括与各种主成分分析技术（例如，“PCA_NIPALS”（PCA和非线性迭代偏最小二乘法）、“PCA_SCD”（PCA和奇异值分解）、“PCA_Test”等等）相对应的子子类别数据块定义。

[0202] 可视化类别的块定义通常定义了用于呈现由数据模块生成的输出的各种技术。例

如,可视化类别可以包括与图形和/或另外的视觉表示相对应的子类别,例如“线条图”、“图表”、“柱状图”、“散点图”、“直方图”、“数据网格”、“数据云”、“动画”、等等。可视化类别可以包括与准备输出数据以用于特定的图形表示相对应的子类别,例如“圆形SigFig (RoundSigFig)”。

[0203] 数据消费者类别的块定义通常定义了与数据模块所生成的输出数据的各消费者或接收者相关的动作。例如,数据消费者类别可以包括子类别“写”,其使得输出数据写入或保存到文件、数据库、或其它静态储存区域。子类别“写”转而可以具有与指示输出数据可以写入或保存的地方或输出数据可以如何写入或保存的各种类型的位置、数据库、数据库格式、文件格式、等等相对应的相应的子类别。数据消费者类别可以包括子类别“公布”,其使得输出数据被流式传输(例如,经由分析数据网络112)或者以其它方式发布至或传送至数据的订户或接收者(例如,至分布式数据引擎102x、至在线数据模块、至由过程控制系统执行的应用、至用户界面、等等)。在某些实施例中,为了将离线的公布块变换成其在线版本,公布数据块被配置并被绑定至实时数据消费者(例如,至具有对数据模型的输出的订阅的在线数据模块、执行应用、过程控制设备、元件或部件、等等)。在实施例中,数据消费者类别可以包括子类别“转换到控制”,其使得输出数据转换成传统的过程控制通信网络理解的格式的信号,并使得信号经由传统的过程控制通信网络传送到过程控制实体、元件、设备、或部件以影响或引起过程工厂5中的变化。子类别“转换到控制”可以包括例如与各种传统的过程协议格式(例如,HART、PROFIBUS、无线HART、等等)和/或过程工厂5内的各种目标(是变化的信号(例如,控制回路输入、控制回路配置、消息优先级、参数值、等等)的潜在接收者)相对应的子子类别。例如,转换到控制块可以被配置为绑定至过程工厂5内的用于接收转换到控制块的输出并相应地修改其操作的特定的过程控制实体、元件、设备、或部件。

[0204] 数据块定义库252可以在其中包括其它数据块定义,其中每个都可以属于或者可以不属于类别。例如,“探索”块允许用户在离线模式中探索由多个其它数据块生成的输出数据。例如,探索数据块可以在其输入连接器处接收多个不同的数据线,并且可以生成视觉表示,该视觉表示以比较的方式包括输入数据的两个组,例如通过随着时间对齐两个输入数据集并将一个线条图覆盖在另一个线条图的顶部、通过创建直方图,其中每个相应的数据集的信息紧挨着其它数据集的信息显示、等等。探索块允许用户操纵其输出的可视化,例如通过增加/减小x轴和/或y轴的缩放、对显示在图表中的统计数据进行分类和/或过滤、等等。应当指出,探索块是可能不具有在线配对物的块的示例,这是因为其功能主要是允许用户探索和了解由草拟的数据模块生成的输出数据。

[0205] 对于用户开发新的数据模型,用户可以调用数据块定义库252(例如,通过经由用户控制248m进行切换)。用户可以通过将期望的数据块的定义从库252拖放到画布245的期望位置上,来将期望的数据块添加到画布245(以及由此将数据块添加到在画布245上描绘的开发下的数据模块)。(当然,拖放仅仅是指示选择、定位特定的数据块、以及将特定的数据块并入数据模块中的一种方式,并且如此做的许多其它适当的方式是可能的)。在拖放操作之后,所选定的块定义与草拟的数据模块相关联,并且创建和命名针对该草拟的数据模型的该数据块的相应实例。在图4D中所示出的示例性场景中,草拟的模块A1在被开发的过程中,并且在数据工作室画布245上示出其对应的草拟的、离线的数据图。如由指示器248c所示出的,模块A1具有“离线”状态。用户已经将块A1-B1、A1-B2、以及A1-B3添加到草拟的模

块A1,并已经使用线A1-W1和A1-W2连接它们。用户已经经由控制248m调用库252,从库中选择块B4,并将块B4拖放到画布245上(如用虚线指示的),并处于将块B4的实例命名为“A1-B4”的过程中。在块B4已经被命名的实例中,用户可以用线将A1-B4互连到草拟的模块A1的一个或多个其它数据块。例如,用户可以使光标悬停在A1-B4的输入连接器上方,并点击以创建至选定的输入连接器的新的线连接。用户随后可以点击画布245上的另一个块实例的期望的输出连接器以创建将所选定的输入连接器互连至期望的输出连接器的新的线,由此互连两个块实例。在另一个示例中,用户可以将光标悬停在另一个块实例的输出连接器上方,点击以创建至所选定的输出连接器的新的线连接,并点击A1-B4的期望的输入连接器以创建互连线。当然,可以使用任何适当的用户控制机制以创建块实例之间的线互连。

[0206] 在数据工作室画布245上,用户能够针对特定的块实例修改块的给定属性的值。为了示出,图4E描绘了在上面开发草拟的数据模块B1的数据工作室画布245的一部分。包括在草拟的模块B1中的两个块实例已经位于画布245上并互连,即,B1-填充NaN和B1-PCA。B1-填充NaN从数据源(在图4E中未示出)接收其输入数据集,并且B1-PCA将两个输出“模型”280a和“分数”280b提供给相应的消费者或接收块(也未示出)。用户已经指示他或她期望修改与B1-PCA块实例相对应的块定义属性的值(例如,通过悬停在B1-PCA块实例上方、通过在B1-PCA块上双击、或者通过使用任何其它适当的用户控制机制)。响应于激活的用户控制,包括与PCA块定义相对应的所定义的属性列表的窗口282已经出现(例如,出现为浮动式对话框、弹出式窗口、或某种其它适当的表示)。用户随后可以经由窗口282根据期望修改B1-PCA的各个块属性的值。

[0207] 如先前所讨论的,每个数据块(如果有的话)的属性由储存在块定义库252中的其相应的块定义来定义。在图4F-图4H中示出了块定义和它们的相应属性的例示性的(但不是限制性的)示例。应当理解,块定义库252可以包括在图4F-图4H中示出的示例性块定义中的一个或多个块定义、不包括这些示例性块定义、和/或其它块定义。此外,图4F-图4H中示出的每个示例性块定义的属性的默认值也可以与本文中所描述的示例不同。

[0208] 在图4F中,由DDE用户界面应用呈现“加载DB”数据块285a的数据块定义模板,例如,由于用户将加载DB块定义从库252拖到画布245上并随后指示他或她期望查看和/或修改加载DB数据块285a的块属性值,以创建块285a的特定实例。加载DB块285a在数据模型的离线图内执行的动作或功能包括加载(或者以其它方式获得其内容)对其操作数据模型的特定数据集。例如,可以从数据库或文件加载或获得特定的输入数据集。用户能够通过将期望的名称、指示器、或值输入到数据库名称属性字段285b和数据集属性字段285c中(例如,经由下拉菜单、搜索和选择(例如,与过程控制系统相关联的过程控制数据库或其它数据库)、自由形式的文本输入、等等),来指示将被加载到数据模型中的特定的输入数据集。例如,用户可以从下拉菜单中选择期望的数据库名称285b,并且对期望的数据库285b的选择引起为数据集字段285c提供对应的下拉菜单,该下拉菜单仅包括所选定的/识别的数据库285b的数据集285c。加载DB块285a还包括绘图属性字段285d,在该实施方式中,其是布尔(boolean)标志,该布尔标志的值指示当加载DB块实例被评估时,来自数据库285b的数据集285c的绘图是否将生成/呈现。如果绘图285b被设置为“是”,则在块评估之后,所加载的数据集285c的绘图将显示于在画布245上示出的加载DB块实例的图形表示上,例如,在加载DV块实例的查看块结果区域212中。对于加载DB块285a,绘图属性285d是可选的,并且它的值

默认为“否”(例如,当评估块时,将不会生成/呈现加载的输入数据集的绘图)。在用户已经将所有期望值输入到加载DB数据块285a的属性字段285b、285c、和285d中时,用户可以保存经修改的属性值,由此配置加载DB数据块模板285a以创建将在用户正创建/开发的数据模型中使用的加载DB数据块285a的特定实例。用户可以命名特定的实例,例如,通过将期望的名字(例如,“加载DB-1”)输入到在画布245上示出的块实例的块名字字段210中。

[0209] 图4G示出了由DDE用户界面应用呈现的“列”数据块286a的数据块定义模板。列数据块286a的特定实例从如由选定列属性字段286b的其属性值指定的加载的/获得的数据集(在本文中被称为数据“列”、或者由“标签”所引用的数据)提取或获得选定组的数据。例如,列数据块286a经由其输入连接器,例如,从加载DB数据块285a的实例接收数据集(或其某种指示,例如指针、指示器、或对数据集的其它引用)。列数据块286a的选定列属性字段286b允许用户选择一列或多列、标签、或者用户期望由数据模型的其它块进行操作的输入数据集的其它部分。例如,如在图4G中示出的,用户处于滚动遍及数据集A的标签和高亮的期望标签的列表的过程中。通常(但不是必须),选定的列或标签是包括在输入数据集中的总的列或标签的子集。用户可以保存列或标签选择286b(以及其它块属性,如果存在的话),由此配置列数据块286a以创建特定的实例,例如,将在用户正创建/开发的数据模型中使用的特定实例。用户可以命名列286a的特定实例,例如,通过将期望的名称输入到其块名称字段210中。

[0210] 图4H示出由DDE用户界面应用呈现的“PCA_NIPALS”数据块287a的数据块定义模板。PCA_NIPALS数据块287a对一组数据执行主成分分析(PCA)和非线性迭代偏最小二乘操作。例如,PCA_NIPALS数据块287a可以经由其输入连接器接收由列数据块286a的实例确定的数据的列或标签(或者它们的指示器或对它们的引用)。在另一个示例中,PCA_NIPALS数据块287a可以经由其输入连接器接收已经被加载DB 285a数据块加载以便由数据模型使用的整个数据集(或者它们的指示器或者对它们的引用)。PCA_NIPALS块287a的模板允许用户选择PCA的部件的数量287b和/或置信水平287c(如果期望的话)。属性字段287b、287c两者对于PCA_NIPALS块287a是可选的,并且可以被设置为默认值(例如,其在PCA_NIPALS的块定义中进行定义)或零。在图4H中示出的示例性场景中,用户已经指示PCA_NIPALS的当前实例将被配置为生成两个部件287b和至少95%的置信等级287c。PCA_NIPALS数据块287a模板可以被配置(例如,具有或不具有经修改的属性值)以创建将在用户正创建/开发的数据模型中使用的特定实例,并且用户可以命名特定的实例,例如,通过将期望的名称输入到其块名称字段210中。

[0211] 某些类型的数据块是对由先前的块生成的输出数据进行操作的消费者或接收数据块。这些类型的消费者数据块可能需要先前的块被评估以用于对消费者数据块进行配置。例如,如果加载DB块485a的实例被连线为为列块486a的实例提供数据,则对加载DB块实例485a的评估将提供用户从中可以选择特定的列或标签以配置列块实例486a的数据集。

[0212] 现在转到对离线数据块和数据模块的评估,离线数据模块可以在其正被开发时被重复评估,并且不需要完全完成来进行评估。因此,对离线数据模块的评估可以异步执行,从而用户能够评估草拟的数据模块的部分、基于评估修改草拟的数据模块、重评估、再次修改、再次重评估、等等。例如,用户可以针对数据模块创建初始数据块、评估单个数据块、随后将第二数据块连接到第一数据块、仅评估第二数据块(或将第一数据块和第二数据块两

者作为整体进行评估)、添加四个更多的数据块、仅评估新添加的数据块(或者将所有的六个数据块作为整体进行评估)、等等。也就是说,当期望数量的数据块实例和/或相关联的线已经被添加到离线或草拟的数据模块的作业数据图表时,可以对在画布245上表示的模块(或者其部分)进行评估,例如,通过激活评估导航栏242上的用户控制248i。例如,如果用户期望评估呈现在画布245上的全部块和线,则用户可以简单地评估用户控制248i。另一方面,如果用户期望仅评估呈现在画布245上的块和线的子集,则用户可以选择期望的块和/或线(例如,通过点击、套捉(lasso)、或其它适当的机制)并随后激活控制248i以使得选定组的块和线被评估。

[0213] 图4I示出了可以如何评估离线数据图标288的示例。由于离线数据图表288正在被创建或开发,因此对在数据工作室画布245上示出的其块和互连的表示被储存到传输文件或文档290(在本文中也称为“配置文件”)中,其可以具有诸如JSON(Java脚本对象表示)或任何其它期望的格式之类的轻量数据交换格式。因此,图形的、离线数据图表288被绑定到传输储存文件或文档290。当用户指示他或她期望离线数据图表288或者其部分被评估(例如,通过激活用户控制248i)时,传输文件或文档290被编译成目标执行环境的语言,由此针对数据图表288生成可执行代码292。在实施例中,传输文件或文档290可以被划分成多个部分,每个部分都被编译成可在不同的目标环境中执行的不同的目标语言(例如,不同目标语言的可执行代码292的多个块或段可以根据传输储存文件或文档290生成)。在已经生成离线数据图表288的可执行代码292之后,创建过程以管理和协调代码292的执行,代码292可以或可以不跨不同的目标环境、和/或可以或可以由不同的处理器执行。

[0214] 当离线数据图表285的各个块被执行或评估时,过程可以向数据工作室发送反馈。例如,基于来自过程的反馈,数据工作室更新离线图表285的每个数据块的相应的状态指示器208以指示块是否正在编译、正在评估、已经成功完成评估、还是尚未成功完成评估(例如,失败)。事实上,通常,贯穿图4I中所示出的编译和执行过程,反馈可以返回到数据工作室并在用户界面上指示。例如,基于来自过程的反馈,数据工作室更新模块状态指示器248k和/或离线状态指示器248c以指示模块作为整体是否正在经受编译、经受评估、已经成功完成评估、还是尚未成功完成评估。

[0215] 在DDE用户界面应用中,用于评估离线数据块、图表、和模块的基础结构和动作由数据分析服务来提供。例如,评估用户控制248i的激活调用数据分析服务以执行对离线数据块、图或当前正在画布245上开发的模块(或者其部分)的评估,例如,其可以利用多个不同的后端平台(例如,多种不同的目标语言、编译器、处理器、和/或目标环境)。因此,DDE用户界面应用避免用户必须参与用于执行数据块、图、或模块的后端平台(或甚至具有其知识)。也就是说,用户可以利用数据工作室和仪表板特征(例如,数据块定义库252、画布245、和相关联的用户控制),以不仅使用数据图来设计或计划(engineer)数据模块,而且还实时并独立于后端平台评估、测试、和部署它们的设计,由数据分析服务来“在遮盖下”管理这些设计。此外,DDE用户界面应用架构在不影响面对用户的数据模块设计、评估、测试、和部署功能的情况下有利地允许增加或删除另外的后端平台和/或部件。

[0216] 在后面的部分中提供数据分析服务及其对评估的支持的更详细描述。

[0217] 参考图4A,如先前所讨论的,对离线模式下的数据模块的评估可以是迭代的过程。随着用户针对草拟的数据模块添加和配置块,用户可以评估所添加的块、并且每个块的状

态由其相应的状态图标(例如,在图4A中示出的指示器208)来反映。在其中对块的评估不成功或失败的情形下,用户可以获得关于失败的细节(例如,通过选择或点击“失败”状态指示器208),从而用户能够解决导致失败的任何问题。在其中对块的评估成功的情形下,用户可以查看块评估的结果,例如,通过选择或点击块的相应的查看块结果图标212。在实施例中,当用户选择“查看块结果”212时,模态或可视化对话可以出现(例如,出现为弹出式窗口、浮动对话框、或其它适当的格式),并且用户可以查看并探索各种可视化以看到数据块结果。

[0218] 事实上,当数据块执行或评估时,数据块可以储存可能对于辅助用户确定对传送至其(例如,经由线和输入连接器)的数据进行操作的数据块的一个或多个影响有用的任何类型的一个或多个结果。所储存的一个或多个结果的类型通常是特定于块的(以及在某些情形下,特定于块的特定实例),并可以由块开发者来定义或确定。这些结果可以经由模态或可视化对话窗口向用户示出。

[0219] 除了特定于块和/或特定于块实例的结果以外,DDE用户界面应用可以提供可适用于多于一个(以及在某些情形下,绝大多数或甚至所有)数据块的标准可视化。例如,当块被执行时,可以收集关于在块的执行结束时的数据状态的各种统计信息的标准集合,从而对于数据集的每列、每个标签、或每个部分,可以计算平均值、标准偏差和其它这种统计信息并连同得到的数据集储存。当呈现特定块实例的可视化对话时(例如,经由相应的用户控制212的激活),从数据储存设备获取针对每列、每个标签、或每个部分的标准统计信息的经计算的集合,并呈现给用户。用户随后可以选择感兴趣的列/标签/部分并请求数据工作室生成相应的图表或表示所述列/标签/部分的统计信息的其它视觉格式(例如,线形图、散点图、直方图、数据网格、数据汇总网格、示出数据分布的经计算的统计信息和直方图、等等)。在实施例中,统计信息和数据集单独进行储存,这是因为将统计信息独立于得到的数据集进行储存有利地允许DDE用户界面应用仅将所需要的数据量加载到浏览器中。

[0220] 在图4J中示出了用于呈现标准的和定制的可可视化的示例性架构300。在示例性架构300中,“块1”302的数据块代码执行,并生成得到的数据集305、经计算的标准统计信息/其它可视化308、和特定于块1的结果310,将它们储存到由DDE用户应用界面管理的本地或远程储存区域312中。在数据工作室实例315(例如,浏览器窗口)处,一旦用户选择显示在块1图形上的“查看块结果”用户控制212,块1的经计算的统计信息308(例如,标准集合和/或任何定制的可可视化)就被加载318到数据工作室实例315,并且用户能够选择感兴趣的期望的列、标签、或部分。一旦用户选择块1的期望的列/标签/部分,对应的数据就被加载320到数据工作室实例315,以使用户查看和探索。

[0221] 随着时间过去,当对多个数据模块进行配置时,块配置和使用的通常模式可以发生在例如相同的过程控制系统5内和/或跨具有多个过程控制系统的企业发生。如果识别到这些共性,可能期望对一组单独的数据块进行分组以形成包括(encapsulate)它们的共同行为的新的块,例如,合成的块。在某些场景下,可能期望定义定制的、单一的数据块,例如当发现一组特定的属性值随着时间重复时,或者当用户期望定义定制的数据操作或功能时。定制的和/或合成的数据块可以被创建并储存在库252中,从而它们在其它数据模块中可以使用。在图4K中示出了示例性场景330,该示例性场景330示出了合成的数据块的创建。在数据图332中,用户选择具有特定关系(例如,“填充NaN”和“缩放”)335的两个单一的数据块以形成新的合成块。使用模态对话框、弹出式窗口、或者由数据工作室提供的其它适当的

界面,用户用期望的名称“填充&缩放”配置或定义新的合成块,并将新的“填充&缩放”合成块保存到数据块定义库252(附图标记338)。在已经定义并保存“填充&缩放”之后的某个时间,其可以代替使用单独的“填充NaN”和“缩放”数据块而被并入到另一个数据图339中。

[0222] 图4L描绘了示例性离线数据体340,其示出了以上所描述的特征和原理中的某些特征和原理,并且其在本文中是同时参照图4A、图4B、和图4F-图4H来描述的。在图4L中,已经由用户在数据工作室的用户画布245上创建示例性离线数据图340。具体来说,用户已经调用块定义库252,例如通过激活画布245上的控制248m,并且已经将加载DB块模板285a拖放到画布245上。此外,用户已经将加载DB块285a的特定实例命名为“加载DB4M”(附图标记342a),并且已经将加载DB4M块实例342a配置为从数据库M中加载数据集4,例如,通过在加载DB4M块342a的字段285c和285b中设置属性值。此外,尽管在图4L中未明确描绘,但用户已经将加载DB4M 342a的绘图属性值285d设置为“是”,从而一旦对块加载DB4M342a进行评估,加载的数据集4的绘图就被生成并呈现在加载DB4M块342a的查看块结果字段212a中。

[0223] 在用户配置加载DB4M块实例342a之后,用户将两个接收者列块实例342b、342c连接到加载DB4M块实例342a的输出。例如,用户将列数据块模板286a的两个不同的实例拖放到画布245上,并相应地将该实例命名为“列X”(附图标记342b)和“列Y”(附图标记342c)。此外,用户已经通过使用互连线将列X块342b和列Y块342c中的相应输出中的每个输出连接到加载DB4M块342a的输出。

[0224] 用户也已经基于对加载DB4M块342a的评估来配置列X块342b和列Y块342c。具体来说,用户首先评估加载DB4M块342a(例如,通过在画布245上选择块342a的图像并激活“评估”用户控制248i),由此使得加载DB4M块342a进行编译并执行以从数据库M加载或获得数据集4。由于加载DB4M块342a的绘图属性被设置为“是”,因此对加载DB4M块342a的评估还使得加载的数据集4的绘图显示在例如加载DB4M块342a的查看块结果字段202a中(在图4L中未示出)。使用该绘图,用户查看并探索数据集4内的数据的各列、标签、或子集,并随后将列X块342b配置为接收在数据集4中被打标签或标记为“X”的数据的子集或列,并且将列Y块342c配置为接收在数据集4中被打标签或标记为“Y”的子集或列(例如,通过针对列块342b、342c中的每个块在列块模板286a的属性字段286b中相应地选择适当的子集、列、或标签指示器)。因此,对列X块342b的评估仅得到来自数据集4的、被特征化、打标签、或标记为“X”以在块342b处加载或访问的数据,并且对列Y块342c的评估仅得到来自数据集4的、被特征化、打标签、或标记为“Y”以在块342c处加载或访问的数据。

[0225] 用户还已经在离线数据图340上拖曳、放下、并配置偏最小二乘(PLS)块。具体来说,用户已经命名PLS块实例342d“PLS4M_X_Y”,并且已经配置PLS4M_X_Y块342d以在其相应的输入处接收列X块342b的输出和列Y块342c的输出。PLS4M_X_Y块342c对列X 342b和列Y 342c提供的数据操作或执行偏最小二乘函数、动作、算法、或操作,并且结果(例如,基于列X与列Y数据之间的关系所生成的相应的PLS模型)经由PLS4M_X_Y块342d的输出被提供给探索4M_X_Y块342e。与其它块342a-342d类似,探索4M_X_Y块342e是用户已经拖放到画布245上、命名、并配置的相应的探索块定义的实例。

[0226] 在离线图340中,除了342d以外,探索4M_X_Y块342e被配置为接收由PLS4M_X_Y块342d生成的输出,用户已经将探索4M_X_Y块342e配置为还经由线342f接收列Y块342c的直接输出作为输入。这种配置允许探索4M_X_Y块342e使得用户能够探索PLS4M_X_Y块342c的

输出与列Y块342c的输出之间的关系,例如,通过呈现一个或多个可视化和/或统计信息。例如,通常,PLS4M_X_Y块342c的输出包括一个或多个预测值。探索4M_X_Y块342e允许用户将包括在PLS4M_X_Y块342c的输出中的一个或多个预测值与列Y块342c的值进行比较,例如以判断PLS4M_X_Y模型342d中的列X数据342b的使用是否足够预测列Y数据342c。

[0227] 应当指出,在该离线的数据图中,PLS4M_X_Y块342d和探索4M_X_Y块342e在其自身配置可以完成之前都不需要评估其一个或多个紧接着的块。也就是说,对PLS4M_X_Y块342d和探索4M_X_Y块342e中的每个块的配置可以独立于其它块的评估、并在任何时间执行。然而,由于离线数据图340表示一个或多个数据管道或数据流,因此通常在可以评估下游数据块之前(当然,除非接收者或下游数据块接收仅针对单独的块评估目的而设置的测试或虚设输入数据),对每个下游数据块的评估需要评估其一个或多个上游数据块。事实上,在某些场景中,用户将整个离线数据图340作为整体进行配置,和/或可以将整个离线数据图340作为整体评估而不是在逐个块或逐个部分的基础上评估。

[0228] 2. 数据分析工作室-在线数据模块

[0229] 现在转到图4B,在数据模型的离线数据图已经完成并在数据工作室中评估为用户的满意度之后,离线数据图可以被转变或转换成其等同的在线形式。为了将离线数据图变换成其在线形式,用户可以在导航栏242上选择或激活在线切换或用户控制248f,由此使得数据模型的在线数据图显示在画布245上。某些离线数据块可以具有在线配对定义(例如,“加载”数据块的离线和在线版本),其它离线数据块可以不需要不同的在线配对物,然而包括在线数据模块中(例如,“填充NaN”块或“PCA_NIPALS”块),而另外的离线数据块可以在在线数据模块中省略(例如,“探索”块)。在将离线图变换成其在线配对物期间执行的基础结构和动作由数据分析服务来提供。例如,用户控制248f的激活调用数据分析服务以执行将离线数据图变换成其对应的在线数据图。在下面对数据分析服务的描述中提供了数据分析服务如何执行这种变换。

[0230] 当在数据工作室的画布245上呈现数据模块的在线数据图时,用户可以配置在线数据模块。通常,配置在线数据模块包括与控制过程相关的实时过程控制数据,该实时过程控制数据将被获得为在线数据模块的输入数据,例如,通过定义使在线数据模块与输入数据的对应的数据源相关的绑定。此外,配置在线数据模块可以包括指示由在线数据模块生成的输出数据将被提供的一个或多个位置和/或消费者(例如,通过定义绑定)(例如,一个或多个消费者DDE和/或应用、数据储存设备、文件、历史库、过程控制设备、例程、元件、部件、等等)。在实施例,以类似于上面关于离线数据模块所讨论的方式类似的方式来执行数据块的配置和/或修改、它们的相应属性、和在线数据模块的互连。

[0231] 如果用户识别关于在线数据模块的疑问或问题或者以其它方式期望修改在线数据模块,则用户可以将在线数据模块翻转或切换回其离线表示,例如,通过利用离线切换或用户控制248d,并且用户可以如上面所描述地进一步修改和评估数据模型的离线数据图。随后,当用户将离线数据图翻转或切换回其在线表示时,经修改的离线数据图变换成数据模型的对应的在线数据图。

[0232] 图4M示出了与图4L的示例性离线数据图340相对应的在线数据图345,并且下面同时参照图4A、图4B、图4F-图4H、和图4L来讨论图4M。在图4M中所描绘的示例性场景中,用户已经决定图4L中的离线图340根据预期或期望定义了数据模型,并且数据模型准备用于部

署到在线过程控制系统中。因此,工作室在查看离线数据图340(例如,如图4L中示出的)时,用户已经激活数据工作室的在线切换248f,由此将离线数据图340变换成对应的在线数据图345,如在图4M中示出的,该在线数据图345呈现在数据工作室的画布245上。这种变换由DDE用户界面应用的数据分析服务来执行,并且在后面的部分中更详细地描述由数据分析服务利用的特定动作、机制、和架构来执行这种变换和其它变换。

[0233] 应当指出,尽管在线数据图345和离线数据图340表示相同的数据模型,但两个数据图345、340的块和线的集合不同。例如,在加载DB4M块342a的输出处起始的、横贯通过列Y块342e,并在探索4M_X_Y块342e的输入处终止的离线数据图340的数据管道从在线数据图345中省略,这是因为该数据管道在离线数据图340中用于测试和验证离线PLS4M_X_Y块342d,并且在PLS4M_X_Y块342d的功效已经被证明为用户的满意度的现在的在线环境中并不是必须的。然而,在某些实施例中,用户可以选择在在线数据图345中包括该验证的数据管线,例如,如果用户期望在在线环境中具有对在线数据模型的连续不断的测试和验证。

[0234] 离线数据图340与在线数据图345之间的另一个差异是加载DB4M块。在离线数据图340中,加载DB4M块342a被配置为从数据库M的离线数据源数据集4中加载数据。另一方面,在在线数据图345中,加载DB4M块347a已经变换成可以连接到在线数据源(例如,过程工厂5内的在线流式传输的数据源)的块。用户可以将加载DB4M块347a配置为绑定至期望的在线数据源,并且对经配置的加载DB4M块347a的编译实例化该绑定。

[0235] 在在线数据图345中,列X块347b通常与其离线形式342b类似或等同。然而,如先前所讨论的,PLS4M_X_Y块347c的在线形式仅接收对于PLS4M_X_Y块347c在在线环境中进行操作是必要的输入,即,列X块347b的输出。

[0236] 此外,在线数据图345并不具有对于离线的探索4M_X_Y块342e的配对物,这是因为在离线环境中的探索4M_X_Y块342e用于为用户提供草拟的数据图340能多大程度上实现用户的目标的视图。但是,在在线数据图345中,PLS4M_X_Y块347c的输出被提供给“写”数据块347d,该“写”数据块347d是一种类型的数据消费者数据块。写数据块347d使得在线环境中的PLS4M_X_Y块347c的实时输出被写到文件或数据库,如在写块347d的配置中识别的。当然,如果用户这样期望,则可以通过将PLS4M_X_Y块347c的输出与其它数据消费者块的输入互连来将PLS4M_X_Y块347c的输出另外地或替代地提供给数据模型的一个或多个其它数据消费者。例如,PLS4M_X_Y块347c可以连接到公布块,从而PLS4M_X_Y块347c的输出被公布至数据分析网络112并且可由其它数据分析应用使用。在另一个示例中,PLS4M_X_Y块347c可以连接到转换到控制块,从而PLS4M_X_Y块347c引起在线过程控制系统的变化。例如,转换到控制块可以被配置为将PLS4M_X_Y块347c的输出转换成用于过程工厂5内的过程控制功能、回路、或应用的输入信号,并使得输入信号被传送到对应的过程控制产业、元件、设备、或部件。

[0237] 3. 数据模块的部署

[0238] 通常,在线数据模块并不是异步评估(对于离线数据模块是可能的)。相反,在线数据模块首先被配置(例如,绑定至实时数据源和/或一个或多个数据消费者/接收者),并随后作为整体被部署为连续执行并与在线过程控制系统进行交互。用户可以将数据模块部署到在线过程控制系统中,例如,通过点击或激活导航栏242上的“部署”用户控制248j,由此通知数据分析服务处理对在线数据模块的部署。在对数据分析服务的描述中更详细地描述

了用于部署数据模块的基础结构和动作。但是,通常,当部署在线数据模块时,在线数据模块遵循与关于正被评估的离线数据模块的编译和执行过程(例如,如在图4I中所示出的)类似的编译和执行过程。然而,在线数据模块的一个或多个目标环境可以与其对应的离线数据模块的一个或多个目标环境不同,并且在线数据模块被绑定至过程控制系统或工厂内的实时数据源。

[0239] 在实施例中,DDE用户界面应用支持“一次点击式”部署。也就是说,当离线数据模块已经被配置为用户的满意度时,用户不需要将数据模块转换为其在线表示并对针对负责从实时数据源读和写数据的在线表示的数据块的绑定进行配置。相反,在“一次点击式”部署期间,一旦用户发起,就由DDE用户界面应用自动确定和配置针对对象数据模块的任何需要的绑定。在示例性场景中,用户可以分配数据模块以在特定的DCS控制器上执行,例如,通过使用由数据工作室提供的对话框或窗口,该数据工作室获取关于DCS系统的层级的信息。用户可以浏览层级并相应地分配数据模块。DDE用户界面应用(例如,数据工作室和/或数据分析服务)随后利用所分配的过程控制系统元件(在该场景中,特定的DCS控制器)以自动确定必要的绑定。在对数据分析服务的描述中更详细描述了用于执行一次点击式部署的示例性基础结构和动作。

[0240] 在在线数据分析模块已经被成功部署之后,通过其执行所生成的任何得到的值(例如,描述性的、预测性的、和/或规范性的)可以向用户示出,例如,经由DDE用户界面应用的数据仪表盘。在实施例中,得到的值包括至少一个预测性的值,并且由过程控制系统5针对预测值所应用的一个或多个标签或过程控制元件生成的对应的实时的、实况(live)的一个或多个数据值也可以结合预测值来示出。

[0241] C. 数据分析仪表盘

[0242] 如先前所讨论的,DDE用户界面应用包括第二操作模式,数据分析仪表盘。仪表盘通常由用户用来查看和/或管理模块,该模块为(i)正被评估的离线数据模块(例如,草拟的数据模块)、和/或(ii)已经被部署到在线过程控制系统或工厂中的在线数据模块。使用视觉和其它表示,仪表盘为用户提供与评估离线数据模型相对应的状态和其他信息的视图,以及所部署的在线数据模块和正由在线数据模块生成的实时或实况信息的视图。通常,数据分析仪表盘模式为用户提供了界面,经由该界面,用户可以监控和管理在线和离线数据模块。用户还可以从仪表盘终止对任何离线数据模块的评估和对任何在线数据模块的执行。通常,但不是必须,显示在数据分析仪表盘上的信息中的至少某些信息通过DDE用户界面应用的数据分析服务被提供给仪表盘350。

[0243] 图4N-1描绘了由DDE用户界面应用的数据分析仪表盘呈现的示例性用户界面350。示例性仪表盘用户界面350的第一部分352提供了当前的数据分析概要信息,例如对当前正由DDE用户界面应用监控的在线数据模块355的数量的指示以及对实况数据流358的指示,在线数据模块355正对该实况数据流358进行操作。如先前所讨论的,从一个或多个数据源接收实况数据流358,该一个或多个数据源连续生成源自控制一个或多个工业过程的在线过程控制工厂或环境5的实时数据,并且执行的在线数据模块355已经被配置为接收实况数据流358。数据分析概要信息还包括对其评估当前正被DDE用户界面应用监控的离线数据模块360的数量的指示。在某些实施例中,仪表盘用户界面350包括与所监控的数据分析模块(不管是在线的还是离线的)相关的一个或多个其它通知或其它信息362。

[0244] 当用户点击或以其它方式激活概要控制355、358、360、362中的一个时,与该选择相对应的另外的总结信息呈现在仪表板350的第二部分365上。在图4N-1中,用户已经选择查看用于当前执行在线数据模块355(此刻存在两个,即,用附图标记368指示的图2、以及用附图标记370指示的图3)的总结信息。对于各自执行在线数据模块360、370,仪表板350的第二部分365指示其相应的概要信息,例如模块372的名称、模块被部署的起始时间375、模块的输出随着最近的时间间隔的趋势指示378(其可以是图形的,如图4N-1中示出的)、最新的预测结果或一个或多个预测值380、和/或其它信息(未示出)。特别要指出的是,在仪表板350上,对每个在线数据模块368、370的趋势指示378连续更新以反映相应的数据模块368、370的连续生成的输出,由此允许用户实时监控在线数据模块的执行及其连续生成的结果。

[0245] 此外,数据分析仪表板350允许用户点击与任何数据模块360、378相关联的任何总结信息以查看更多的细节,例如,在弹出式窗口中、在对话框中、或者在其它适当的查看机制(未示出)中。例如,当用户点击图2 368的名称时,仪表板250可以呈现图2 368的在线数据模块图和对图2 368所绑定的一个或多个特定的数据流358的指示。当用户点击图2 368的开始信息时,可以呈现对图2 368的创建者的指示、部署图2 368的用户、和/或图2 368的拥有者。点击图2 368的趋势378信息允许用户操纵并探索图2 368的输出数据趋势,例如,通过延伸或缩短所查看的时间间隔,将趋势信息转换成另一种类型的可视化(例如,条形图、饼形图、散点图、等等)、应用和/或获得关于随着给定间隔的趋势数据的统计信息(例如,平均值、滚动平均值、最大值、最小值、等等)、等等。类似地,点击图2 368的最后的预测结果380信息允许用户操纵并探索最新预测的一个或多个值,例如,通过查看最后的预测结果的时间和多久一次作出预测、将最后的预测结果与当前的实况数据值进行比较、等等。

[0246] 当用户选择用户控制360来查看当前执行的离线数据模块(在图4N-1的示例性场景中不存在)时,可以在仪表板350的第二部分365中呈现与离线数据模块相对应的另外的总结信息。通常,针对离线数据模块的总结信息与在线数据模块的总结信息类似,然而,代替指示相关联的实况数据流,针对离线数据模块的总结信息可以指示由离线数据模块在上面执行的一个或多个离线数据源(例如,测试工具、设备、和/或环境)生成的一个或多个测试数据文件或数据流。此外,针对离线数据模块的总结信息可以指示当前状况(status)、状态、和/或其执行进展。

[0247] 在某些场景中,用户可以配置和/或部署来自仪表板(而不是来自数据工作室)的离线数据模块。例如,用户可以监控对来自仪表板的离线数据模块的评估,并且离线数据模块成功地完成其评估。用户可以激活仪表板(未示出)的用户控制,经由该用户控制,用户可以配置数据模块以引用在上面执行在线数据模块的一个或多个实况数据源和由该一个或多个实况数据源生成的数据。在线数据模块绑定至一个或多个数据源的这种配置与以上描述的配置类似,除了从仪表板发起而不是从数据工作室发起以外。或者,用户可以经由仪表板利用针对离线数据模块的一次点击式部署。

[0248] 当用户选择用户控制358来查看与当前绑定至执行在线数据模块(在图4N-1的示例性场景中存在三个)的在线数据流相关联的信息时,在仪表板350的第二部分365中呈现与在线数据流相对应的另外的总结信息(未示出)。针对每个实况数据流的总结信息可以包括数据流的标识符、对正从中接收数据流的数据源的指示、对对应的过程控制系统标签或实况数据源的其它传统过程控制系统标识符的指示、关于对数据流的一个或多个订阅和/

或数据流的公布的信息、对当前在实况数据流上执行的一个或多个在线数据模块的指示、实况数据流的连续更新的可视化(例如,线条图、条形图、散点图、等等、和/或它们的基本统计信息)、和/或其它信息。

[0249] 特别要指出的是,在图4N-1中所描绘的示例性场景中,数据流概览358指示存在当前正由在线数据模块操作的三个实况数据流,而在线数据模块概览355指示仅存在两个执行的在线模块。因此,两个执行的在线模块368、370中的一个在线模块对实况数据流中的两个数据流进行操作。用户可以通过进行导航以查看每个在线数据模块368、370的细节(这将提供对相应的模块368、370所绑定的数据流的指示)、或者通过进行导航以查看每个数据流358的细节(这将提供对执行正在对每个数据流358进行操作的在线数据模块的指示),来发现这种关系。

[0250] 图4N-2描绘了可以由DDE用户界面应用的数据分析仪表板呈现的示例性用户界面382。在实施例中,用户界面382被包括在在图4N-1中示出的用户界面350的第二部分365中。在实施例中,用户界面382是全屏视图。

[0251] 用户界面382包括一个或多个片(tile)或卡385a-385j,其中每个都与相应的执行的在线数据模块相对应,并且可以共同被称为卡385x的托盘或组(pallet or deck)。托盘中的卡385x中的每个卡都可以包括与针对图4N-1的执行的在线数据模块360、370所显示的信息类似的信息,例如,模块的名称、模块被部署的起始时间、模块的输出随着最近的时间间隔的趋势指示、最新的预测结果或一个或多个预测值、和/或其它信息。尽管与针对执行的在线数据模块360、370的显示类似,但对卡385x中的每个卡的趋势指示连续更新以反映其相应的在线数据模块的连续生成的输出,由此允许用户实时监控其执行和连续生成的结果。

[0252] 卡385x中的每个卡都是可重新设置尺寸的和可移动的,如由用户期望的。在实施例中,当特定的卡385x处于焦点内(in focus)时,该特定的卡的尺寸增加。卡385x可以手动开始聚焦,例如当用户点击该特定的卡385x、在该特定的卡385x上方悬停、等等时。另外地或替代地,卡385x可以自动开始聚焦,例如当特定的数据模块需要迫切关注时(例如,预测值跨阈值、数据模块自身需要用户输入或用户干预、在卡385x中排序的情况严重性、等等)。可以在用户界面382上指示所需要的关注的其它自动生成的指示,例如,通过高亮或向特定的卡385x提供某种其它处理、根据卡385的组的相应的紧迫性设置卡385x的组的尺寸、将特定的卡385x带到托盘的前景或前方、以关注的优先级的顺序布置卡385x、等等。在实施例中,用户可以定义在哪些特定情形中特定的处理将自动应用于卡385x。

[0253] 现在回到储存在DDE用户界面应用的数据块定义库252(例如,如在图4C中示出的)中的数据块定义255,可以由DDE用户界面应用在其启动或可视化期间自动发现储存在库252中的数据块定义。在实施例中,使用轻量数据交换格式(例如JSON文件或文档、或其它适当的格式)来定义数据块定义。特定的数据块的定义文件或文档包括块定义名称并详细说明定义特定的数据块的块属性和连接器。在某些情形下,特定的块定义可以为离线表示提供定义并为在线表示提供定义。

[0254] 由于DDE用户界面应用可以连续地并动态地发现块定义,因此另外的块定义(例如,对于定制的数据块、合成的数据块、和/或其它数据块)能够在DDE用户界面应用正在执行的任何时间被添加到库252。在实施例中,由实用应用程序(在本文中被称作“数据块定义

包装器”)来生成另外的块定义。例如,当用户保存定制的或合成的数据块时,可以调用数据块定义,如先前关于图4K所描述的。当块定义包装器执行时,其创建了另外的块定义的实例,并使得定义了另外的块定义的对应的JSON文档或类似物被创建和储存。例如,另外的块定义的所创建的实例生成JSON文档并将其储存在库252中。参考图4K作为示例性场景,当用户保存合成的块定义338时,块定义包装器执行以使得针对合成的块定义338的JSON文档或类似物被生成或储存。

[0255] 在数据块(不管是离线的还是在线的)的配置期间,DDE用户界面应用可能需要执行某种功能(例如,商业逻辑),以便实现帮助用户成功地配置和使用数据块所需要的配置体验。由于块定义由DDE用户界面应用动态地发现,因此这种功能自身并不嵌入在DDE用户界面应用中。因此,在实施例,当DDE用户界面应用被初始化时,还可以动态地加载特定于特定数据块的任何相关联的功能。随后,对与特定的数据块相对应的动态加载的功能的引用被保持在特定数据块的块定义中,并且当特定块的块定义的实例被配置时由DDE用户界面应用来使用。在本申请的后面部分中提供了关于块定义和相关联的功能的另外的细节。

[0256] 与过程控制系统集成的工业过程监控和分析系统

[0257] 如上面对图1-图3和图4A-图4N-2的讨论中看到的,分布式工业过程性能监控和/或分析系统或DAS 100提供了支持用于工业过程工厂的数据分析配置、数据建模、数据模型执行、数据可视化、和流式服务的平台。如先前所讨论的,尽管监控和分析系统100的各个节点被嵌入在过程控制系统或工厂5内,但是通常,分析平台100的数据通信网络112存在于过程工厂5的现有DCS、PLS、健康监控、和其它系统外部(并独立与这些系统)。在示例性配置中,数据分析网络112覆盖在过程工厂5的顶部,由此将网络112与过程工厂5中的传统的过程控制通信网络10、70、等等中的大多数或全部过程控制通信网络分隔开。传统的过程控制通信网络继续传输过程控制数据以控制、管理、和/或实施(administrate)对工厂5的过程和操作的控制,而数据分析网络112传输分析数据(其可以包括通过传统的过程控制通信网络传输的数据的副本),以发现关于过程工厂5的操作的知识,例如,描述性知识、预测性知识、和/或规范性的知识。数据分析系统100的架构允许数据分析系统100容易地被添加到或集成到现有的、传统的过程工厂5中。

[0258] 由数据分析系统100发现的关于过程工厂5的指示可以用于修改过程工厂5的参数、装备、操作、和其它部分以提高工厂5的性能并降低故障、失效、和其它不期望的情况的出现。有利地,数据分析系统100实时地发现与过程工厂的当前的、在线操作联系的关于过程工厂5的知识,从而所发现的知识是描述性的、预测性的、和/或规定过程工厂5的当前操作。因此,数据分析系统100和过程工厂5形成控制回路以控制和优化过程工厂5的性能。

[0259] 在图40中示出了控制和优化过程工厂5的性能的控制回路的高级别框图400。如在图40中示出的,在线过程工厂5通过传统的过程控制通信网络(例如,网络10、70、等等)生成、操作、以及传输各种类型的控制数据402,以控制一个或多个工业过程并生成过程工厂输出405。分布式工业过程性能监控/分析系统100经由嵌入的分布式数据引擎102x连接到过程工厂5并与过程工厂5集成,该分布式数据引擎102x捕获由于过程工厂5控制一个或多个过程以产生输出405所生成的分析数据408。如先前所讨论的,所捕获的分析数据408可以包括第一级或较高级的过程控制数据402的副本,以及由DDE 102x本地执行的和/或由经由数据分析系统100的数据工作室开发的离线和在线数据分析模块执行的分析功能的结果/

输出。在某些情形下,所捕获的分析数据408包括经由数据工作室生成的和/或由DDE102x自主生成的新的分析功能。

[0260] 数据合理系统或平台100接收、生成、传输分析数据408、并对分析数据408进行操作以生成分析输出410。分析输出410可以包括关于过程工厂5的所发现的知识,例如描述过程工厂5的当前操作的知识、给定过程工厂5的当前操作,预测故障、失效、时间间隔、性能、事件、等等的出现的知识、和/或给定过程工厂5的当前操作,规定可以被采取以减轻当前工厂操作的不期望的特征和/或以减轻不期望的预测的故障、失效、时间间隔、性能、时间、等等的出现的概率的一个或多个规范性的动作的知识。在某些实施例中,所发现的知识410包括关于分析数据408自身的知识。例如,所发现的分析知识410可以包括可用于评估不同的离线数据模块的最佳的离线数据集、已经由数据工程师重复使用并被定义为新的数据分析用户程序的一系列步骤、被定义为新的数据分析技术的数据块的新的配置、等等。事实上,在对集成到炼油过程工厂中的原型数据分析系统100的测试期间,由控制回路400发现用于工业过程工厂的新的数据分析技术,并且该新的数据分析技术在以下后面的部分中进行解释。

[0261] 包括在分析输出410中的所发现的知识中的至少某些知识可以被提供给人机界面(HMI) 412,例如数据分析用户界面应用。经由HMI 412,用户可以探索由数据分析系统100生成的关于过程工厂5的所发现的知识410。在某些情形下,对所发现的知识410的探索包括用户利用数据工作室来创建并执行另外的数据分析模块以生成另外的所发现的知识410以用于另外的探索。在某个点处,基于所发现的知识410的主体,用户可以对在线过程工厂5的一个或多个值、参数、装备、部件、控制回路、和/或其它当前操作作出一个或多个变化418,由此优化过程工厂5的性能和输出405和/或防止或阻止故障、失效、和其它不期望的情况的出现。

[0262] 在某些情形下,包括在分析输出410中的所发现的知识中的至少某些知识被直接提供给过程工厂5以影响或实现一个或多个规范性的变化,如在图40中用箭头420所表示的。例如,所发现的知识410可以包括规范性的动作,其包括对设定点的变化、对控制器的配置的变化、对过程控制消息的优先级的变化、或者对某个其它值、参数、配置的变化、等等。数据分析系统100可以自动并直接下载或者以其它方式使得变化在过程工厂5内实现,而不需要任何用户干预418,由此基于所发现的知识410来自动优化过程工厂5的性能和输出405和/或防止或阻止故障、失效、和其它不期望的情况的发生。应当指出,当分析系统100基于过程工厂5的当前操作来处理当前的分析数据408并生成所发现的知识410时,控制回路400的该分支自动更新在线过程工厂5以实时地改进其当前操作。因此,与需要几小时甚至几天来计算、确定、和实现规定动作的现有的分析技术相反,由分析系统100生成的规范性的动作针对工厂5的当前操作情况而优化,并立即被应用于过程工厂5,并且在某些情形下,甚至在任何用户注意到逼近不期望的或期望的情况之前被应用。

[0263] 当然,对于过程工厂5内的规定动作作为整体的最佳实施方式,控制回路400需要自主分支420和HMI控制分支418两者。例如,某些规范性的动作需要人为干预,例如安装替换阀、传感器、或其它部件。

[0264] 如先前所讨论的,数据分析平台100使得分析数据能够在多个嵌入的分布式数据引擎(DDE) 102x之间流式传输。可以流式传输由于控制过程而由过程工厂5生成的任何类型

的一阶数据。例如,包括在由过程控制设备(例如,控制器、I/O卡、现场设备、等等)生成的信号中的过程控制数据(为了控制过程的目的,其通常经由传统的过程控制通信网络进行传输)也可以在数据分析网络112中流式传输,以用于数据分析的目的。因此,包括在过程控制信号、传感器信号、等等中的内容或数据可以被复制和/或重新产生并通过数据分析网络112进行流式传输。此外,所生成的和经由传统的过程控制通信网络传输的二阶(或者较高阶)过程控制数据(例如,将在过程控制数据库或历史库内被历史化的数据、过程控制配置、过程控制用户界面命令、过程控制网络管理信号、以及由于控制过程而在传统上经由传统的过程控制通信网络发送和接收的任何其它信号)可以在数据分析网络112中流式传输,以用于数据分析的目的。此外,数据分析网络112可以流式传输已经对传统的过程控制数据和/或对经计算的分析数据执行的分析功能或技术的经计算的数据或结果。

[0265] 因此,数据分析平台100部分地通过本地绑定至过程工厂5的数据源、和/或通过提供接近于过程工厂5内的数据源的数据分析服务或功能,而同时提供较大规模的预测和优化。仅有对于支持外部循环预测和优化来说是必要的数据需要被传输到分析系统100的较高级/更有能力的分析服务器和云平台。

[0266] 如先前所讨论的,数据分析系统100的多个DDE 102x中的某些被嵌入到过程工厂5内的物理设备中。例如,多个DDE 102x中的某些各自与相应的过程控制设备(例如,控制器11;I/O卡26、28;现场设备15-16、18-21、42A、42B、44、等等)和/或与相应的网络路由和管理设备(例如,网关35、75、78;适配器52a;路由器58;接入点52a、72;等等)集成。在实施例中,代替被集成到过程工厂5的单一设备或部件中,DDE可以是作为潜望镜(scope)、监视器、或窥视器(voyeur)的附接到传统的过程通信链路或网络(例如,网络10、70,PLS网络;安全健康监控系统、等等)的独立式设备,并且可以复制/重新产生经由传统的过程通信链路或网络传输的所选定的过程控制数据,以便为了数据分析的目的通过数据分析网络112进行流式传输。

[0267] 在某些场景中,多于一个DDE可以形成本地的数据分析集群,数据源可以向该数据分析集群注册和加入,并且经由该数据分析集群,由经注册的数据源生成的分析数据可以被接收到数据分析网络112中作为分析数据。在示例中,一组DDE形成本地集群并通告本地集群的存在。本地集群出现为单一的节点或数据分析网络112的单个DDE实例150,尽管本地集群包括经由数据分析网络112互连的多个DDE。一旦在过程工厂5的场所内新的数据源被激活或启动,数据源就扫描或寻找本地集群通告。数据源可以检测本地集群的通告并向本地集群注册,由此加入数据分析网络112。随后,经注册的数据源可以从分析系统100(例如,经由它们的本地集群)请求带宽,并可以向它们的本地集群、向数据分析网络112的另一个节点、或直接向分析系统100公布数据。由于本地集群出现为分布式数据引擎150的实例,因此本地集群经由相应的大数据接收器160本地地接收所生成的分析数据(例如,从经注册的源和/或其它源),并将本地生成的分析数据储存在相应的本地永久存储器155中。此外,本地集群可以例如经由相应的大数据请求服务者(servicer) 165向请求的应用(例如,向另一个数据分析应用、向数据模块、向用户界面、向数据工作室、向另一个应用、等等)提供所储存的分析数据中的至少某些分析数据。在某些情形下,本地集群可以例如经由相应的大数据请求分析器170来对已经从经注册的数据源接收到的所储存的数据和/或对在数据分析网络112内流式传输的其它数据执行相应的本地数据分析功能。本地数据分析功能的结果

或输出可以本地地储存在相应的大数据储存器155中、和/或可以例如经由本地集群的相应的一个或多个数据分析网络接口175来流式传输到本地集群外部的一个或多个DDE,例如,流式传输到另一个本地DDE或本地集群、集中式数据集群、或流式传输到订阅应用。

[0268] 在某些实施例中,本地集群利用更多集中式集群来注册它们自身,该集中式集群接收并储存较高级的分析数据并执行相应的、较高级的数据分析功能。例如,特定的集中式集群(例如,区域性集群)可以从工厂5的特定区域内的所有本地集群接收流式传输的数据,并且可以对其执行数据分析。在另一个示例中,特定的集中式集群可以从工厂5内的特定类型的配置的所有控制循环接收流式传输的数据,并可以对其执行数据分析。类似于本地集群,集中式集群具有数据分析网络112中的DDE 150的单个节点或实例的外观。

[0269] 因此,分布式工业过程性能监控和分析系统100在各种本地化的层级或级别处提供DDE和/或DDE集群的能力允许故障检测和预测接近于过程工厂5的数据源来执行,由此提供与在线数据源生成的实时数据相对应的实时故障检测和预测。这种多层级或集群的架构还允许遍及系统100的分析数据的储存,并且事实上,遍及过程工厂5,这是因为分析系统100包括嵌入在其中的部分。此外,数据分析系统100提供访问已经被储存在任何DDE处的能力,并为分析配置、诊断、和监控性能、故障、和失效提供集中式系统。

[0270] 在实施例中,一个或多个集中式集群为用户或系统管理者提供集中式位置,以便为来自多个DDE 102x的分析数据的备份或历史化建立规则。在示例中,一个或多个集中式集群位于过程工厂5的场所内,并将过程工厂5作为整体来服务。在另一个示例中,一个或多个集中式集群位于或托管在云中,并将过程工厂5作为整体来服务。在某些配置中,提供过程工厂5的备份或历史化的一个或多个集中式集群还服务其它过程工厂。通常,如先前所讨论的,分析数据永久地储存在每个DDE或集群处,例如储存在其相应的大数据储存区域155中。在某些情形下,在DDE处接收到的或者由DDE所生成的某些分析数据可以在DDE或集群处缓存(例如,基于诸如订户数量、公布或传输频率等之类的一个或多个标准)以进行更快的访问,当通常来说,所有本地分析数据被移动到DDE或集群处的永久储存器155。周期性地或者以预定义的时间,在每个DDE处的永久储存的分析数据可以被备份、保存、或历史化到长期数据分析储存器中,例如,在可经由数据分析网络112访问的集中式分析数据历史库处。例如,数据分析历史库可以是全工厂的数据分析历史库、和/或可以被托管在云中。但是,通常,数据分析历史库是任何过程控制数据历史库是独立的和分离的。经由访问一个或多个集中式集群的用户界面,用户或系统管理者可以定义一组规则,其指示每个特定的DDE储存区域155将多久被备份一次、备份何时发生、分析数据将在何处被历史化、等等。

[0271] 在实施例中,一个或多个集中式集群为用户或系统管理者提供集中式位置,以便为分析数据在集群、节点、和/或数据分析系统100的DDE 102x中的复制建立规则。例如,该组复制规则可以定义哪些特定的分析数据将在数据分析网络112的哪些特定的节点中进行复制、以及特定的分析数据将在特定的节点中多久更新一次。例如,该组复制规则可以定义由节点A执行的分析功能A的结果在节点B-D处多久被复制一次(例如,通过流式传输),从而所复制的结果对节点B-D是可用的以用于执行它们的相应的数据分析功能。与历史化和备份规则类似,该组复制规则可以由用户或系统管理者经由访问一个或多个集中式集群的用户界面来定义。例如,系统管理者可以利用用户界面来管理该组复制规则,并且该组复制规则可以储存在一个或多个集中式集群处。提供分析数据复制的管理的一个或多个集中式集

群可以是提供分析数据备份/历史化的管理的相同组的集群,或可以是不同组的集群。

[0272] 在实施例中,一个或多个集中式集群提供位置服务、流式服务(例如,图1中的流式服务115)、数据字典、和/或可以由本地集群、其它集中式集群、和/或单独的DDE 102x利用以在它们之间建立分析数据交换的其它功能。提供位置服务、流式服务、数据字典、和/或其他功能的一个或多个集中式集群可以是提供分析数据复制的管理和/或分析数据备份/历史化的管理的相同组的集群,或者可以是不同组的集群。系统管理者或用户可以利用用户界面来访问一个或多个集中式集群,以管理服务、字典、和用于在数据分析网络112的集群与节点之间建立和提供分析数据交换的其它功能。

[0273] 图4P描绘了提供本地化的数据分析服务的示例性方法450。在实施例中,分布式数据引擎102x或DDE的集群执行方法450的至少一部分。通常,方法450的至少一部分可以由分布式工业性能监控和/或分析系统(例如系统100)的任何部分来执行。事实上,方法450可以使用本文中所描述的系统、装置、设备、和/或技术中的任何一个的任何部分来实现。但是,为了讨论的简单起见,下面关于分析系统100的一个或多个DDE 102x的集群描述了方法450,该集群出现为单一的DDE实例150,但是这仅仅是许多实施例中的一个实施例,而并非表示限制。

[0274] 示例性方法450包括由集群通知集群的存在(框452)。集群包括在过程控制工厂5的数据分析网络112中,该过程控制工厂5是在线的并操作为控制过程,并且集群通告可以通过数据分析网络112、通过一个或多个其它网络、经由短距离无线协议(例如蓝牙、NFC、等等)或经由任何其它适当的通知手段来传输。

[0275] 方法450还包括由集群从数据源接收对集群的通知的响应。例如,来自数据源的响应可以是对数据源的请求以向集群注册,并且方法450可以包括向集群和/或向数据分析网络112注册数据源(框455),由此将数据源加入数据分析网络112。数据源可以是包括在过程工厂5中的并由于控制过程的在线过程工厂5生成数据的设备或部件。在示例中,数据源生成源自受控制的过程的连续的、时间序列数据。

[0276] 此外,方法450包括由集群接收由数据源生成的数据(框458)。例如,数据源可以经由数据分析网络112将其生成的数据流式传输至集群。另外地或替代地,集群可以订阅由数据源生成和公布的特定数据。在某些实施例中,集群可以针对特定的数据查询数据源,例如通过使用在本申请的后面部分中所描述的查询机制。当数据源是通常经由一个或多个传统的过程控制通信网络(例如,网络10、70)传输过程控制数据的设备或部件时,数据源可以继续如平常那样通过一个或多个传统的过程控制通信网络传输过程控制数据,而同时还将所传输的过程控制数据的内容的副本或复制品经由数据分析网络112流式传输到集群。

[0277] 一旦接收到由数据源生成的数据,集群就可以缓存所接收到的数据中的至少某些数据,例如,当所接收到的数据满足一个或多个缓存条件时。不论所接收到的数据中的任何数据是否被缓存,但是所接收到的数据的全部被本地储存到集群的永久储存器155中。

[0278] 方法450包括由集群经由数据分析网络112将由数据源所生成的数据中的至少某些数据流式传输到数据消费者(框460)。数据消费者可以例如是另一个分析节点或集群、数据分析应用、用户界面和/或用户界面应用、在线数据分析模块、另一个应用、和/或数据库或数据历史库。数据消费者可以绑定至数据源和/或绑定至包括由数据源生成的数据中的至少某些数据的特定数据流。期望被数据消费者接收的特定数据通常是时间序列数据,并

可以是连续的数据。

[0279] 数据消费者可能已经针对由数据源生成的特定数据查询集群,例如,通过利用在本申请的后面部分中所描述的查询机制,并且流式传输(框460)可以响应于由数据消费者生成的查询。在实施例中,数据消费者可以订阅由数据源生成的并由数据源或者由集群公布的特定数据。在任何速率下,集群从其缓存或者从其本地永久数据存储器155获取已经由数据源生成的期望的数据,并将该数据经由数据分析网络112流式传输到数据消费者(框460)。

[0280] 在某些实施例中,方法450包括由集群对数据源生成的数据中的至少一部分执行一个或多个数据分析功能或技术(在图4P中未示出)。一个或多个分析功能或技术可以包括描述性分析、预测性分析、和/或规范性的分析。在例如每次出现在线数据分析模块、另一个数据分析应用、或用户的请求时,一个或多个分析功能或技术可以由集群来执行。在另一个示例中,一个或多个分析功能或技术可以由集群自动发起或执行。一个或多个分析功能或技术执行的结果可以返回到请求方、和/或可以本地储存在永久数据存储器155中的集群处。在某些实施例中,对已经由向集群注册的多个数据源生成的、并且已经储存(以及可选地缓存)在集群处的数据执行一个或多个分析功能或技术。

[0281] 由于集群接收由过程工厂5内的数据源所生成的一阶数据,并接近于过程工厂5内的这些数据源而被放置,因此集群在本文中被称作“本地”集群。在某些实施例中,方法450包括检测由集中式集群(例如,区域性集群、地区(area)集群、全工厂集群、等等)生成的通告的本地集群。本地集群可以向集中式集群注册其自身,并可以将储存在其缓存和/或其永久数据存储器155中的本地分析数据中的至少某些数据流式传输至集中式集群。例如,被流式传输到集中式集群的分析数据可以包括由数据源生成的数据和/或由本地集群执行的一个或多个分析功能的结果。每次出现由集中式集群发起的请求时,本地集群可以将数据流式传输到集中式集群,和/或本地集群可以公布由集中式集群订阅的数据。

[0282] 在某些实施例中,方法450包括对储存在集群的本地永久存储器155中的数据进行历史化或备份(在图4P中未示出)。数据备份可以由集群发起,或者数据备份可以由集中式分析数据历史库发起。在实施例中,集群和集中式分析数据历史库可具有发布者/订阅者关系。在实施例中,将从集群备份的数据的定时和/或数量可以由系统100的系统管理者来定义。将被备份的数据可以从集群传输和/或流式传输到分析数据历史库,以用于长期储存和备份。

[0283] 图4Q描绘了提供集中式数据分析服务的示例性方法470。在实施例中,分布式数据引擎102x或DDE 102x的集群执行方法470的至少一部分。通常,方法470的至少一部分可以由分布式工业性能监控和/或分析系统(例如,系统100)的任何部分来执行。事实上,方法470可以使用本文中所描述的系统、装置、设备、和/或技术中的任何一个的任何部分来实现。但是,为了讨论的简单起见,下面关于系统100的一个或多个DDE 102x的集群描述了方法470,该集群出现为单一的DDE实例150,但这仅仅是许多实施例中的一个实施例,而并非表示限制。

[0284] 示例性方法470包括由集中式集群通知集中式集群的存在(框472)。集中式集群与图4P中的本地集群不同,这是因为数据源通常不直接向集中式集群注册。相反,在某些情形下,本地集群或其它类型的集群向集中式集群注册,并且集中式集群对从这些其它集群接

收到的数据执行数据分析功能。但是与本地集群类似,集中式集群包括在过程控制工厂5的数据分析网络112中,该过程控制工厂5在线操作以控制过程,并且集中式集群通知可以通过数据分析网络112传输、流式传输、和/或广播。如本文中所使用的“集中式”集群通常指代其它集群向其注册的集群。集中式集群可以服务控制回路(例如,“控制回路”集群)、过程工厂的地区(例如,“地区”集群)、过程工厂的区域(例如,“区域性”集群)、整个过程工厂(“全工厂”集群)、多个过程工厂(“企业”集群)、等等。通常,集中式集群包括经由数据分析网络112互连的多个DDE,该多个DDE具有作为单个、单一DDE实例150的外观。

[0285] 方法470还包括由集中式集群从另一个集群接收对集中式集群通知的响应。例如,来自另一个集群的响应可以是另一个集群向集中式集群注册的请求,并且方法470可以包括向集中式集群注册另一个集群(框475)。注册的集群可以是诸如图4P中的本地集群之类的本地集群,或者注册的集群可以是另一个集中式集群。

[0286] 此外,方法470包括在集中式集群处从经注册的集群接收流式传输的数据(框478)。例如,经注册的集群可以经由数据分析网络112将其储存的和/或缓存的数据中的一部分流式传输到集中式集群。因此,流式传输的数据通常包括时间序列数据,其可以是例如连续的数据。另外地或替代地,集中式集群可以订阅由经注册的集群公布的特定数据。在某些实施例中,集中式集群针对特定数据查询经注册的集群,例如,通过使用在本申请的后面部分中所描述的查询机制。流式传输的数据可以包括由经注册的集群生成的数据、和/或可以包括从数据源和/或从其它集群流式传输到经注册的集群的数据。

[0287] 在某些实施例中,集中式集群缓存其接收的流式传输的数据中的至少某些数据,例如,当所接收到的数据满足一个或多个缓存条件时。不管在集中式集群处接收到的数据中的任何数据是否被缓存,所接收到的数据中的全部都本地储存到集中式集群的持久储存器155中。

[0288] 在框480处,方法470包括对所接收到的、流式传输的数据(例如,储存在其缓存和/或其永久储存区域155中的数据)的至少一部分执行一个或多个数据分析功能或技术。一个或多个分析功能或技术可以包括描述性分析、预测性分析、和/或规范性的分析。例如,每次发生在线数据分析模块、另一个数据分析应用、或用户的请求时,可以由集中式集群执行一个或多个分析功能或技术。在另一个示例中,可以由集中式集群自动地和/或周期性地执行一个或多个分析功能或技术。一个或多个分析功能或技术的执行的结果或输出可以返回到请求方,和/或可以本地储存在集中式集群的永久数据储存器155中。在某些实施例中,对已经由多个经注册的集群所生成的、以及已经在集中式集群处储存(以及可选地缓存的)数据执行一个或多个分析功能或技术。

[0289] 方法470包括由集中式集群经由数据分析网络112将一个或多个分析功能或技术的结果或输出中的至少某些流式传输到一个或多个数据消费者(框482)。一个或多个数据消费者可以包括例如另一个分析节点或集群、数据分析应用、用户界面和/或用户界面应用、在线数据分析模块、另一个应用、和/或数据库或数据历史库。一个或多个数据消费者可以绑定到集中式集群,绑定到特定的分析结果/输出、和/或绑定到包括特定的分析结果/输出的特定数据流。

[0290] 一个或多个数据消费者可能已经针对特定的分析结果/输出请求集中式集群,例如,通过利用本申请的后面部分中所描述的查询机制,并且流式传输(框482)可以响应于由

一个或多个数据消费者生成的查询。另外地或替代地,数据消费者可以订阅由集中式集群公布的特定的分析结果/输出。在任何速率下,集中式集群从其缓存或者从其本地的永久数据存储器155获取期望的数据,并经由数据分析网络112将期望的数据流式传输到一个或多个数据消费者。在某些情形下,一旦通过一个或多个数据分析功能或技术生成所期望的分析结果/输出数据,就流式传输该期望的分析结果/输出数据。

[0291] 在某些实施例中,方法470包括向集中式集群注册一个或多个另外的本地集群,并从该一个或多个另外的本地集群接收相应的分析数据(在图4Q中未示出)。在这些实施例中,执行一个或多个数据分析功能或技术(框480)可以包括对在来自多个本地集群的集中式集群处接收到的分析数据执行数据分析功能或技术。

[0292] 在某些实施例中,方法470包括对储存在集中式集群的本地永久存储器中的数据(例如,储存在大数据储存区域155中的数据)进行历史化或备份(在图4Q中未示出)。数据备份可以由集中式集群发起、或者数据备份可以由集中式分析数据历史库发起。在示例中,位于集中式分析数据历史库中的集中式集群可以具有发布者/订阅者关系。将从集中式集群备份的数据的定时和/或数量可以由该集中式集群处的或者另一个集中式集群处的系统100的系统管理者来定义。例如,系统管理者可以利用用户界面来管理针对系统100内的备份和/或其部分的一组规则,并且该组规则可以储存在集中式集群处。将被备份的数据可以从集中式集群152和分析数据历史库进行传输和/或流式传输,以用于长期储存和备份。

[0293] 另外地或替代地,在某些实施例中,方法470包括为分析数据在数据分析系统100的集群、节点、和/或DDE 102x中的复制提供一组规则。例如,该组复制规则可以定义哪些特定的分析数据将在数据分析网络112的哪些特定节点中进行复制,以及特定的分析数据将在特定的节点中多久更新一次。例如,该组复制规则可以定义由节点A执行的分析功能A的结果在节点B-D处多久复制一次,从而所复制的结果对于节点B-D是可用的,以用于执行它们的相应的数据分析功能。与历史化和备份规则类似,该组复制规则可以由系统100的系统管理者在一个或多个集中式集群处定义。例如,系统管理者可以利用用户界面来管理该组复制规则,并且该组复制规则可以储存在它们的一个或多个集中式集群处。

[0294] 在某些实施例中,方法470包括检测由另一个集中式集群生成的通告的集中式集群。例如,通知的集群可以位于过程工厂5的场所内的区域性集群、地区(area)集群、或全工厂集群。在某些情形下,通知的集群可以托管在云中,并且可以为工厂5的一部分、整个工厂5、或多个工厂提供分析服务。集中式集群可以向通知的集群注册其自身,并且可以将储存在其缓存和/或其永久数据存储器155中的分析数据流式传输到通知的集群。例如,集中式集群可以流式传输集中式集群已经执行的分析功能的结果中的至少某些结果。集中式集群可以将数据流式传输到集中式集群向其注册的通知的集群,和/或集中式集群可以公布由通知的集群订阅的数据。

[0295] 在某些实施例中,方法470包括提供位置服务、流式服务(例如,图1中的流式服务115),和/或可以由其它本地或集中式集群以及单独的DDE102x使用以在它们之间建立分析数据交换的数据字典(在图4Q中未示出)。工业性能监控/分析的服务

[0296] 由数据工作室使离线数据图到在线数据图的变换、离线和在线块和模块两者的执行、以及由图所表示的分析的功能由工业性能监控/分析系统(例如数据分析系统(DAS)100)所提供的分析服务来完成。分析服务在本文中可互换地被称为“服务”、“多个服务”、或

“分析服务”。本文中所描述的分析服务中的任何部分或全部部分可以结合图1、图2、图3、和/或图4A-图4Q中的任何部分或全部部分来操作和/或结合在相应地与图1、图2、图3、和/或图4A-图4Q对应的本公开内容的部分中所描述的任意数量的特征和/或技术来操作。

[0297] 通常,分析服务能够编译、执行、和/或变换离线图(例如,在数据工作室和/或仪表板中)并能够编译、执行、和创建在线图(例如,在数据工作室和/或仪表板中)。分析服务有助于探索大量数据集-大数据-而并不必须需要一种或多种特定类型的数据的知识而且并不需要分析服务的用户进行执行分析的低级别编程。在图5A中在系统100的一部分的背景下示出了分析服务500的示例性框图架构。更具体来说,分析服务500是多个计算机实现的过程中的一个过程,其中每个计算机实现的过程都与各软件实体相关联(例如,储存在一个或多个有形的、非暂时性计算机介质上的计算机可执行指令和数据)。如下面将更详细描述,分析服务500经由由系统100中的各实体产生和/或储存的数据在很大部分中与较广泛的系统100进行交互。

[0298] 如以上所描述的,在实施例,中,DDE用户界面应用是基于网页的应用。图5A描绘了在这种实施例中实现的分析服务500的实施例;也就是说,利用基于网页的DDE用户界面应用。在其最广泛的意义上,以此方式实现的分析服务500包括客户端侧、服务器侧、和后端计算机实现的过程。例如,如在图5A中所描绘的,分析服务500包括网页客户端进程502、网页服务器进程504、执行服务过程506、以及一个或多个作业过程508。各过程502、504、506、508可以在一个或多个处理器上执行,在一个或多个工作站或服务器中执行、在一个或多个物理和/或地理位置中执行。也就是说,尽管网页客户端进程502的一个实例可以在远离网页服务器进程504在上面执行的服务器的工作站上执行,而网页客户端进程502的另一个实例可以在网页服务器进程504在上面执行的相同服务器上执行,事实上,可以在相同处理器上运行。作为另一个示例,执行服务过程506可以在过程控制环境的局限下的工作站上运行,而一个或多个作业过程508可以在远离过程控制环境而放置的分布式过程环境的一个或多个处理器(例如,服务器农场)上执行。简言之,分析服务500被设计为足够灵活,从而各过程502、504、506、508是否在相同处理器、工作站、处理器库(bank)上执行、在一个或多个虚拟机中执行、和/或在相同的物理或地理位置中执行不存在要求。有利地,架构不仅有助于对由分析服务500提供的数据分析的远程访问、和利用遥远的处理器和计算资源以执行分析的能力,而且还允许分析服务500跨多个过程工厂消耗和/或分析数据,然而,它们的地理位置可能不同。

[0299] 网页客户端进程502根据不与通常理解的原理如此不同的原理、但根据以上所描述的对DDE用户界面应用的描述来进行操作。例如,网页客户端进程502包括多个软件实体,该实体包括例如:查看实体510,其向用户呈现DDE用户界面应用(例如,数据分析工作室240的画布245和用户控制248a-n、数据分析仪表板、等等);视图模型实体512,其管理应用数据的转换和转化,例如块状态到用户界面可以利用的形式,并且其生成提供来自在线数据模块和/或以其它方式指示在线数据模块的操作的视图;应用实体514,其是被下载到网页客户端进程502并驻留在客户端工作站或设备上的软件实体(即,基于网页的应用),用户使用该客户端工作站或设备来与DDE用户界面应用进行交互;以及数据服务实体516,该数据服务实体516将数据传递至DDE用户界面应用以及从DDE用户界面应用传递数据。

[0300] 数据服务实体516接收例如从响应于用户输入和请求而执行的各作业返回的数

据。如以上所描述的,并在下面另外详细描述,DDE用户界面应用可以在用于发现不同数据之间的关系(或已知关系的数据)的探索模式下(例如,在数据分析工作室中)、或者在用于执行实时(或接近实时)的、连续的分析以预测过程控制环境中的至少一个值的预测模式下(例如,在仪表板中),请求各分析在来自过程控制环境(在某些情形下,当前正由过程控制环境生成的)的数据上运行。如本文中所使用的,短语“实时的”“或实时”表示“足够同时发生以便对于调节或者以其它方式影响过程工厂的当前操作保持有用”。例如,当一个或多个块被置于画布245上,并且这些块中的一个或多个块被执行时,该执行的结果可以经由服务实体516(例如,使用来自网页服务器进程504的web socket协议连接)返回到应用实体514,并经由查看实体510向用户显示。

[0301] 同时,数据服务实体516可以经由其它协议(例如,HTTP)与网页服务器进程504进行通信,以请求和接收对于应用实体514的操作来说是必要的各种类型的信息。作为示例,当应用实体514首先执行时,其可以经由数据服务实体516、使用HTTP协议从块定义库252请求和接收最新的块定义255。或者,应用实体514可以最初仅接收可用的块定义255的类别和列表,当用户选择时,应用实体514可以经由数据服务实体516、使用HTTP协议请求和接收由用户选择的特定块定义255。

[0302] 现在转到网页服务器进程504,一个或多个服务器通信实体518与网页客户端进程502进行通信。如上面描述的,一个或多个服务器通信之间的通信可以实现诸如HTTP、HTTPS、和web socket协议之类的协议,但是所描述的特定协议不应当被认为是限制性的,这是因为可以实现适用于应用的任何其它协议。除了与网页客户端进程502进行通信以外,服务器通信实体518可以从执行根据被置于画布245上的一个或多个块(例如,图示的编程元素)所请求的指令(例如,数据分析、数据加载、数据过滤、等等)的一个或多个过程接收数据。可以从过程经由例如服务器通信实体518中的web socket(未示出)接收数据,并且数据也经由web socket协议被传输到网页客户端进程502。

[0303] 数据服务实体520向网页客户端进程502提供除了从执行与各块相关联的指令的过程接收到的数据以外的数据。例如,数据服务实体520可以从块定义库252获取块定义255并将该块定义255传输到应用实体514和/或可以将关于块执行的状态信息从执行服务过程506传输到应用实体514,从而状态可以由查看实体510来显示。

[0304] 转到图5B,更详细地描绘了数据服务实体520。具体来说,数据服务实体520包括块定义库252中的块定义255的副本(或对其指针)和一组数据插件523。数据插件523中的每个数据插件是一组计算机可执行指令,该计算机可执行指令与可以从应用实体514调用以执行特定作业过程的接口相符。插件523中的每个插件都独立于应用实体514,并且当应用实体514遭遇指定加载的插件523的元数据时,该插件由应用实体514加载。插件523可以包括特定于应用和/或分析服务500在其中操作的操作环境的逻辑。也就是说,用于实现分析服务500的过程控制环境的该组插件523可以与用于实现分析服务500的金融服务的一组插件不同(记住,分析服务500有助于大量数据集-大数据的探索-而不必须需要一种或多种特定类型的数据的知识并且不需要分析服务500的用户进行执行分析的低级别编程),并且可以事实上与用于实现分析服务500的不同的过程控制环境的一组插件不同。

[0305] 块定义255与应用实体514的分离允许应用实体514取决于客户的需求来使用和部署不同的块定义。可以在不影响应用实体514的下层操作的情况下创建、添加、修改、和/或

移除块定义,这是因为特定于应用和/或操作环境的逻辑并未被硬编码到应用实体514中。通过与应用实体514分开提供插件523和块定义255,应用实体514可以以类似的功能进行操作,而不管对于特定应用所需要的特定块定义255和插件523。

[0306] 再次参考图5A,如上面所描述的,应用实体514操作为有助于在离线图(例如,图4L中的离线图340)的画布245上的创建。为了创建离线图,用户选择一个或多个块来放置在画布245上(例如,通过拖拉),并且将块“连线”在一起,从而数据在块之间进行传输。这些块中的每个块是储存为块定义库252中的块定义255中的一个块定义的模板块的实例。当用户激活用户控制248m时,应用实体514使得查看实体510显示块定义255的可用类别和/或可用的块定义255,用户可以选择块以放置在画布245上。在实施例1中,块定义库252储存在由执行服务过程506或网页服务器进程504中的一个或两者共享的存储设备中,并且块定义库252中的元数据(例如,块定义名称、块定义类别、等等)被传输给应用实体514,其中只有当块被选择为被放置在画布245上时,块定义才被传输给应用实体514。或者,当在网页客户端中实例化应用实体514时,块定义库252作为一个整体被传输给应用实体514,并且贯穿网页客户端进程502的执行仍然驻留在网页客户端上。块定义库252的原本(master copy)可以存在于过程控制环境内的数据库(未示出)中、可经由网络访问的云存储设备中、工作站中、等等。

[0307] 块定义255可以根据每个块的功能来进行分类。尽管图5A描绘了在若干描绘的类别521中的每个类别中划分的块定义255,但是图5A并非旨在建议被分类成任何特定类别521的块定义255必须储存在一起,或者必须与其它块定义255分开储存。尽管在某些实施例中可能是这种情形,但在其它实施例中,块定义255并非以任何特定的顺序进行储存,而是每个可以被标记有元数据,该元数据可以指示块定义所属的类别。用于任何给定的块定义255的元数据还可以通过示例而不是限制的方式指示输入和输出的数量、对于块的执行所需要的属性、对于每个属性的UI控制类别、何时启用/禁用对每个属性的控制(即,在什么情况下)、与属性相关的插件信息、属性的默认值、等等。

[0308] 在许多实施例(但不是所有实施例)中,块包括在可以编译和执行块之前可以或必须指定的属性。由于块定义255并非必须被构建到应用实体514中,因此当应用实体514被实例化时,用户从中可以选择特定块的属性的值可以不预定和/或提供给应用实体514。图5C是描绘了用于针对被放置在画布245上的块呈现属性框的方法551的流程图。当块被放置在画布245上时,应用实体514从数据服务实体520(或者在实施例中,从数据库529)获取相对应的块定义255(框553)。其后,应用实体514可以接收命令以针对被放置在画布245上的块显示属性框(框555),例如,当用户在该块上双击时。

[0309] 当应用实体514接收命令以示出块的属性时(框555),应用实体514逐个属性解析块定义。对于每个属性,应用实体514获取属性的元数据(框557),解析属性的元数据以查找插件信息(框559)。如果针对属性,应用实体514找到插件信息(框561),则应用实体514发现并加载所指定的插件(例如,从数据服务实体520中的插件523中)(框563)。应用实体514调用由属性的元数据所执行的插件,以获得所需要的属性值(框565)。另一方面,如果应用实体514未找到针对属性的插件信息(框561),则应用实体514在属性的元数据中找到属性的默认的一个或多个值(框567)。

[0310] 无论如何,已经调用由属性的元数据所指定的插件以获得属性值(框565)或发现

属性的一个或多个默认值(框567),应用实体514找到属性的元数据中的UI控制类型(例如,文本框、下拉选择、单选按钮、复选框、等等)(框569)并将该一个或多个值和控制添加到属性框(框571)。如果在块定义中呈现另外的属性(框573),则应用实体514获取下一个属性的元数据(框557),并且方法继续,直到不存在更多的另外的属性(框573)。当在框定义中未发现另外的属性时,应用实体514为用户显示属性对话框(框575)以设置属性。

[0311] 在某些实施例中和/或对于某些块定义255,块的一个或多个第一属性可以以未激活或禁用状态进行显示,直到块的一个或多个第二属性被配置。一旦配置一个或多个第二属性,应用实体514就可以重新执行方法551中的一个或多个步骤和/或可以重新执行与第一属性相关联的插件523中的一个或多个插件。

[0312] 举例来说,假设用户将“加载”块放置在画布245上。应用实体514获取块定义255。在实施例中,加载块的块定义255为应用实体514提供关于块的元数据,并且具体来说,指示加载块具有三个属性:块名称、所选定的数据库(数据库名称)和所选定的数据集(数据集名称)以从所选定的数据中拉取。当触发属性配置对话框时(例如,通过在加载块上双击),应用实体514需要确定显示什么。应用实体514可以确定在块名称属性的元数据中不存在插件信息,而相反找到默认值,例如“加载<##>”。应用实体514还可以找到在块名称属性的元数据中所指示的该块名称的控制类型指定文本框。应用实体解释默认值(例如,“加载01”),并将其放置到对话框中作为文本框。

[0313] 解析块定义的应用实体514找到数据库名称属性的元数据。检查数据库名称属性的元数据,应用实体514找到属性的插件信息,并且相应地,发现、加载、和调用所指定的超贱。所指定的插件(针对特定应用和/或操作环境(例如针对特定的过程控制环境)所编写的)被编程为发现数据库名称的列表(和相关联的位置),可以从该数据库加载数据集。已经接收到数据库名称的列表,应用实体514解析数据库名称属性的元数据,并确定其应当在下拉选择控制中进行显示。应用实体514将下拉选择控制添加到属性对话框,其中,数据库名称的列表由插件返回。

[0314] 解析框定义的应用实体514找到数据设置名称属性的元数据。该数据设置名称属性的元数据指定插件并指示数据设置名称属性是下拉徐峥控制,而且还指示控制必须被禁用,直到由用户选择数据库名称。在加载块的块定义中未找到更多的属性,应用实体514显示对话框。一旦用户从数据库名称下拉选择控制中选择数据库名称中的一个名称,应用实体514就知道(根据数据块定义)调用在数据设置名称属性的元数据中所指定的插件,将所选定的数据库名称传递到所调用的插件作为论证。插件返回在所选定的数据库名称中可用的数据集的名称,并且用可用的数据集的名称填充数据集名称属性的下拉选择控制,再次显示对话框,这次使得控制启用而不是禁用。用户随后可以选择数据集名称来完成数据块的配置。

[0315] 当然,一旦块被放置在画布245上,在块被配置为指定任何所需要的属性之后,可以执行块(即,可以执行在块定义中所指定的编程指令),从而任何下游块(即,具有与块的输出“连线”的输入的任何块)将具有有效输入,从该有效输入产生有效输出。参照图4L中示出的离线数据图340,例如,加载DB4M块342a必须在块列X 342b和列Y 342c可以被配置之前进行评估或执行,这是因为对块列X 342b和列Y 342C的配置是基于来自加载DB4M块342a的输出的用户选择的。另一方面,并另外参照图4L,尽管PLS4M_X_Y块342d能够独立于块列X

342b和列Y 342c的评估来配置,但块列X 342b和列Y 342c必须在PLS4M_X_Y块342d可以基于被加载到离线数据图340中的输入数据集而被评估之前进行评估。

[0316] 对块和/或模块的执行和评估是异步的。这允许用户在块或模块被评估时继续修改该块和模块。例如,用户可以开始一个块的执行,而同时继续编辑和/或配置一个或多个其它块。有利地,在选择和/或配置和/或执行下游的块之前,可以对可单独编译的和执行的每个块进行评估,并探索结果。因此,特定块的期望输出的先验知识并不是必须的,甚至在看到任何特定块的结果之前并非必须要知道在数据流中的下个块将是什么。

[0317] 离线或在线图的任何执行涉及执行服务过程506和作业过程508的功能。总的来说,如先前所描述的,任何过程508可以执行服务过程506和/或网页服务器进程506在上面执行的一个或多个处理器(其可以是或者可以不是一个或多个相同的处理器)上执行、在网页客户端进程502在上面执行的一个或多个处理器上执行、和/或在一个或多个单独的处理器(例如,在云计算平台上的处理器)上执行、以及可以在配置在一个或多个处理器上的一个或多个虚拟机器上执行。此外,每个作业过程508都可以在不同的目标环境中执行,例如,通过在配置用于目标环境的专用虚拟机器中执行每个作业过程508。作业过程508中的每个通常表示待执行的代码的一部分—在某些实例中,特定的作业过程508将执行与特定块相关联的目标脚本/代码,而在其它实例中,与单个块相关联的指令将被划分成更小的脚本/代码段,以便在多个作业过程508中执行。

[0318] 具体来说,当与单个块相关联的指令被划分成小的脚本/代码段,以便由不同的作业过程执行时,不同的作业过程可以进行合作以完成某种形式的并行化。如本文中所使用的,“并行计算”用于表示将任务划分成较小的单元并并行执行它们的通常实践;“多线程处理”表示软件程序运行多个线程的能力(在资源是可用的情况下);以及“分布式计算”表示跨多个物理或虚拟机器扩散处理的能力。原则上,分布式计算可以在没有限制的情况下向外扩张。鉴于这些概念,应当理解的是,在分布式平台上执行指令与在分布式模式下执行指令并不相同;在没有支持分布式处理的明确的指令的情况下,指令将本地执行。

[0319] 使任务并行化的能力是任务自身的定义中固有的属性。某些任务易于并行化,这是因为由每个作业过程执行的计算独立于所有其它作业过程,并且所期望的结果的集合是来自每个作业过程的结果的简单组合;我们将这些任务称为“不易并行(embarrassingly parallel)”。

[0320] 某些任务需要更多的努力来并行化。具体来说,这些“线性并行”的任务是其中由一个作业过程执行的计算独立于由另一个作业过程执行的计算的那些任务,但是期望的结果的集合是来自每个作业过程的结果的线性组合。例如,“平均值计算”块可以由计算分布式数据库中的一组值的平均值的指令来定义,在该分布式数据库中,跨多个物理存储设备储存单组数据。举一个较具体的示例,假设平均值计算块跨一系列300个生物反应器计算平均温度,每个生物反应器储存相关联的控制器的存储器中的温度数据。可以通过将300个生物反应器的一组温度值分配给多个作业过程中的每个、并随后计算“重要的(grand)”平均值作为由多个作业过程中的每个确定的平均值的加权平均值,来将平均值计算块的执行并行化。

[0321] 而其中数据必须以有意义的方式组织的任务较难来并行化。“数据并行的”任务是其中由每个作业过程执行的计算独立于所有其它作业任务的那些任务,只要每个作业过程

都具有“正确的”作业块。这也许包括例如针对上面的相同的300个生物反应器中的每个计算独立的温度平均值(而不是跨它们的所有300的平均值),假设在生物反应器中不存在交叉效应。如果这些作业过程中的每个都具有生物反应器中的单个生物反应器的温度数据,则这些任务可以被并行化。

[0322] 然而,为了这种任务是数据平行的,必须在与任务对齐的块中组织数据。如果在以上场景中的300个生物反应器中的每个生物反应器都将其温度数据储存在单个对应的数据库或位置中,则可能是这种情形。然而,如果所有300个生物反应器的数据都储存在单个数据库中,并且不以任何方式在数据库内组织,则在任何类型的分析可能发生之前,数据将需要重新组织。数据的重新组织可以在本文中所描述的系统中由被构造为执行分析的图中的其它块(例如,由数据过滤块、查询块、等等)来进行处理。

[0323] 创建和管理作业过程508以及作业过程在各处理器中的并行化和分布是执行服务过程506的责任。当执行模块或块时,不管模块还是块被执行为在线图或离线图的部分,模块或者块的配置从网页服务器进程504被发送到执行服务过程506。在实施例中,配置被发送为JSON(Java脚本对象表示)文件,然而,配置文件可以利用适合于独立于语言的、异步浏览器/服务器通信的任何数据格式。配置文件包括其中将执行由图602表示的分析(或者其部分)的环境的本地源代码,以及对于其执行所需要的环境属性和变量的值。

[0324] 由于块定义并不需要块在任何特定的环境中执行,因此一旦对块(或者包括该块的模块)进行配置就选择目标环境,新的目标环境可以被添加到在分析服务中可用的那些环境,而不需要重新编写应用、块定义、或甚至储存的或执行的块和/或模块。

[0325] 类似地,新的块定义255可以在任何时候被添加到块定义库252,而不影响先前所构建的模块中的任何模块或改变分析服务500作为整体或者具体来说应用实体514的操作中的任何操作。再次参照图5B,在实施例中,由算法定义包装器525完成新的块定义的创建和添加。在算法定义包装器525中,用户或开发者可以使用定义生成器527来通过开发算法创建定义。定义生成器527生成定义255'并将该定义255'储存在块定义库252中,例如储存在529中。如将在说明书的后面部分中变得更清楚,每个块定义255都可以具有对应的数据插件523,并且数据插件523可以储存在对于分析服务500已知的位置(例如,数据库531)中。数据服务实体520可以在应用实体514被初始化/实例化时获取块定义525和数据插件523,并可以将它们提供给应用实体514。

[0326] 或者,代替创建新的块定义,在某些实施例中,用户定义的块定义可能不具有相关联的算法,而是可以包括用户可以将用户自己的代码放置于其中的属性。也就是说,块定义255中的一个块定义可以允许用户将用户自己的算法输入到块中作为属性。

[0327] 如本文中所使用的,术语“参数”和“多个参数”分别指代包括在配置文件中并通过配置文件的方式作为参数传递到执行服务的一个或多个块/一个或多个模块的一个或多个属性。在配置文件中作为参数传递的属性可以包括举例来说文件路径、数据集名称、待选择的列、置信度和在PCA分析的结果中所期望的部件数量、以及编译器实体526编译可执行代码以及作业过程508执行经编译的指令所需要的任何其它信息。也就是说,并非对于执行需要块或模块的所有属性一块或模块的某些属性例如可以指定或者与其它方式涉及该块或模块如何显示在数据分析工作室中,并且对于如何收集、操纵、或处理/分析数据没有影响。

[0328] 在实施例中,配置文件包括块和/或模块的定义,并且还包括对于编译和执行块/

模块所需要的任何属性。在替代的实施例中,配置文件仅包括块的身份和所需要的配置参数,并且块定义是从存储器(例如,从块定义库252所获取的)。不管怎样,配置参数可以根据块定义而不同。某些块可以不具有配置参数,而其它块可具有一个、两个、或多个配置参数。此外,配置参数可以是必需的或可选的。例如,数据加载块(加载数据集的块)或查询块(在数据集中搜索特定数据的块)可能需要制定待加载或查询的数据的位置的数据路径。然而,列选择块可具有作为默认选择的“所有列”,并因此可能不需要对列的子集的选择的特定的、前瞻性的配置。

[0329] 配置文件中的其它参数可以包括环境参数。举例来说但并非限制,配置文件可以包括:指定例如作业流程或者其部分将离线执行还是在线执行的执行模式;目标执行环境(例如,Python、PySpark、本地Spark、等等);执行将是本地的还是分布式的;分布式环境的配置;作业流程或者作业流程的部分将绑定的流式传输的数据源的指示;创建新作业、或者绑定并继续执行持久作业的选项;跨持久作业缓存数据的选项;在在线图的执行中包括验证变量以用于评估执行模型的准确性的选项;或者其它环境变量。

[0330] 块、模块、和/或环境参数的配置可以以多种方式中的任何方式来发生。如上面所描述的,例如,许多参数根据在相应的块或模块中设置的属性而被传递到配置文件。在构建图的过程中,在可以配置和/或执行下游块之前用户配置和执行特定的块常常(但不总是)是必须的。在用户正使用第一块来加载数据和下游的第二块来从数据中选择列的情况下是这种情形—加载块必须在可以执行加载块之前进行配置(例如,具有从中加载数据的路径、对加载的数据的选择、等等),并且必须在数据中可用的列对于第二块是已知的之前执行,从而用户可以选择列。

[0331] 替代地或者另外地,一旦请求块或模块的执行,一个或多个参数就可以添加到配置文件。例如,在用户并未指定块或模块的属性中的一个属性的情况下可以是这种情形。在一个或多个属性被编程为仅在一旦请求执行时才被指定的情况下,也可以是这种情形。在实施例中,例如,数据分析工作室提示用户指定何时选择块或模块以用于执行、用于执行的目标环境和/或在上面执行块或模块的目标处理器或处理器库。此外,在某些实施例中,配置文件中的一个或多个参数可以从数据分析工作室环境自身传递。诸如优选的执行环境、对其中处理可以/必须发生(例如,阻止或需要云执行)的物理位置的限制、等等之类的参数可以针对从应用内执行的所有块/模块而在应用自身中指定。一旦请求任何特定块或模块的执行,这些参数就被传递到配置文件。

[0332] 无论如何,仍然参考图5A,作业监听器实体522接收具有用于待执行的一个或多个块或一个或多个模块的配置文件的执行请求、并创建与请求执行所针对的一个或多个块和/或一个或多个模块相对应的一个或多个作业管理器实体524。作业管理器实体524接收配置文件,并根据所指定的目标执行环境,将配置文件发送到图编译器实体526中的一个或多个以便被编译成可由所指定的目标环境中的处理器执行的指令。图编译器实体526可以被设计为使得图编译器实体526中的每个可操作为将配置文件编译成若干所指定的目标环境中的任何目标环境,或者可以被设计为使得图编译器实体526中的每个可操作为编译仅指定一个特定的目标环境的配置文件。在任一种情况下,作业编译器实体526的输出是可由所指定的目标环境中的处理器执行的计算机可读指令。

[0333] 将配置文件发送给图编译器实体526的作业管理器实体524将经编译的输出从图

编译器实体526发送到作业执行器实体528。作业执行器实体528判断哪个或哪些资源(例如,处理器资源)可用于根据与包括用于作业的目标环境、作业是否应当是分布式的、作业应当是本地的还是远程的相关的所指定的配置参数以及任何/或者其它变量(包括但不限于输入数据的一个或多个源和输出数据的一个或多个目的地)来执行经编译的输出,并随后使用所识别的资源来创建作业过程508。每个作业过程508都在特定的目标环境中执行经编译的输出以实现特定的目标算法(即,由经编译的指令所指定的)。在实施例中,作业过程508中的某些或全部作业过程可以由嵌入在分布式数据引擎150中的大数据分析器170来执行。

[0334] 因此,以下场景中的至少每个场景在该灵活的架构中是可能的:以跨多个处理器的分布式方式执行为多个作业过程508的单个块或模块;在单个处理器上连续执行为多个作业过程508的单个块或模块;在单个处理器上执行为单个作业过程508的单个块或模块;以跨多个处理器的分布式方式执行为对应的多个作业过程508的多个块、等等。此外,尽管通常预期到的是,与单个块或模块相对应的作业过程508将在相同的目标环境(例如,Python、PySpark、等等)中执行,但是并不存在用于模块中的不同块的所有过程、或者对象中的不同模块在单个目标环境中操作的要求。在某些实例中,例如,特定的目标环境可以较好地适合于完成与模块中的一个块相关联的目标算法,而模块中的其它块可以在不同的目标环境中更有效地执行,因此,块可以被配置为在相应的属性不同的目标环境中进行指定,最终结果是与不同块相关联的作业过程508将在不同的环境中执行。

[0335] 又进一步,期望的是,在许多实例中,与特定模块或对象相关联的块将在作业过程508上执行,这些作业过程508全都是分布式的或者全都是本地的。然而,鉴于以上描述,应当明显的是,可以通过本地执行模块或对象的一个或多个块而同时在一个或多个远程处理器(例如,在云计算环境中)上执行模块或对象的其它块来获得效率。例如,考虑其中第一块执行查询以从远离分析服务500而设置的数据库(或其它数据储存器)获取指定数据并且第二块对由查询所返回的数据执行分析的图。在数据库本地的处理器处(例如,在数据库处的嵌入式DDE中)创建作业过程508以对嵌入式DDE本地的数据执行查询比对以远离数据库的处理器执行查询可能更有效。当数据库在另一个地理区域中时(例如,当用户查询除了其中用户当前坐落的过程控制环境以外的过程控制环境的数据时),这可能尤其是这种情形。随后可以根据相同作业过程(和处理器)中的、相同处理器中的不同作业过程中的、或者不同处理器中(例如,在本地处理器中)的不同作业过程中的第二块来对查询结果进行分析。

[0336] 作业执行器实体528还保持对作业过程508的追踪,尤其在分布式处理应用中,并保持对数据依赖性的追踪。因此,当每个作业过程508完成时,输出数据返回到作业执行器实体528。作业执行器528判断该数据中的某些数据还是全部数据被传递到另一个作业过程508和/或传递回网页服务器进程504以用于由网页客户端进程502在应用实体514中进行显示。在作业过程508的执行期间,作业执行器实体528可以接收可被传递回网页服务器进程504以用于在应用实体514中进行显示的状态信息。作业过程508还可以使预测性数据返回到作业执行器实体528,尤其当作业过程508正在执行在线图(或者在线图的一部分)时,该在线图连续执行直到被明确停止。当生成预测性数据时,可以使该预测性数据返回到作业执行器实体528,并且该预测性数据可以被传递到网页服务器进程504以用于在应用实体514中进行显示,而同时记住,预测性数据可以同时被写回根据在线图中的块中的一个块的

过程控制环境中的位置(例如,DDE储存位置、执行控制功能的控制器、等等)。在实施例中,一个或多个状态和/或预测性数据通过在作业创建期间建立的反向通道流式传输通信信道来在作业过程508与作业执行器实体528之间传递以允许作业过程508彼此通信和/或与作业执行器实体528进行通信。

[0337] 通过以下示例,将变得显而易见的是,在某种程度上,数据经过分析服务器500的流动至少依赖于所配置的准确的块和/或模块、针对块或模块中的每个所配置的属性、以及正在执行的图是离线图还是在线图。

[0338] 示例1-创建离线图

[0339] 图5D是描述示例的画布600的示意图,其中,画布600上设有经配置的离线图602,经配置的离线图602包括块604a-604g和线606a-606h。再参考图5A,画布600将通过网页客户端进程502显示,并且尤其是通过查看实体510与应用实体514来显示。应用实体514将经由服务实体516接收数据(包括块定义库252),数据服务实体516经由服务器通信实体518来与网络服务器进程540通信。

[0340] 一般来说,有一些方法来构造如图5D所绘出的离线图602。首先,用户可以触发该块定义触发件248m来打开块定义库252(参见图4D)。从块定义库252,用户可以一次选择块604a至604g中的每一个,并且把块604a-604g布置到画布600上。然后,用户可以布置线606a至606h来连接各个块:将线606a布置到块604a的输出端608a和块604b的输入端610b之间;将线606b布置在块604b的输出端608b和块604c的输入端610c之间;将线606c布置到块604b的输出端608b和块604d的输入端610d之间;将线606d布置到块604b的输出端608b和块604e的输入端610e之间;将线606e布置到块604c的输出端608c和块604f的输入端610f2之间;将线606f布置到块604d的输出端608d和块604g的输入端610g之间;将线606g布置到块604d的输出端608d和块604f的输入端610f1之间;以及将线606h布置到块604f的输出端608f1和块604g的输入端610g之间。在对块604a至604g以及线606a至606h进行如此布置后,离线图602将保持未配置。

[0341] 可替换地,用户可以一次一个地选择和布置块604a-604g,并且因为每个块均被布置在画布600上,用户可以对块604a-604g中的每一个块配置属性,可选地(并且在大多数情况下)在每个块的属性被配置后,执行块604a-604g中的每一个块。

[0342] 通过点击块并且在用于该块的块属性窗口612中指定与块相关联的属性,用户可以一次一个地配置块604a-604g中的每一个(无论块604a-604g在属性配置之前是否被布置并且连线,或者块604a-604g被一次一个地布置并且每个块的属性在布置下一个块之前被配置)。在图5D中,块604A(查询块)被选择并且显示了该查询块604a的块属性窗口612。在图5D中所描绘的示例画布600中,用于查询块604a的块属性窗口612包括查询属性输入区614,在其中用户可以输入查询。虽然对查询语言的详细描述留在说明书的后面部分进行讨论,但对于当前的目的,已经足够来假设根据查询块604a的执行,输入到查询属性输入区614的查询可以指定将被查询的数据的位置。在与块相关的配置文件中,块属性被存储为参数。配置文件可以驻留在执行网页客户端进程502的网页客户端上,或者可以位于执行网络服务器进程504的网络服务器上。

[0343] 在配置查询块604a的属性后,用户通常在布置下一个下游块和/或开始配置该下一个下游块(在此示例中,为块604b)的属性之前执行执行块604a。为了执行该块604A,用户

将激活评估用户控件248i。这会导致与块604A相关联的配置文件经由网络服务器进程504的数据服务实体520传送到该执行服务器进程506。执行服务进程506将接收该配置文件(例如,作为JSON文件),并且作业监听实体522将创建作业管理实体524来处理执行请求。作业管理实体524将接合适当的图编译器实体526来将该配置文件编译为适用于由配置文件中的参数指定的目标环境的可执行的指令(即经编译的查询代码)。图编译器实体526返回包括经编译的代码的输出文件,并且也可能会返回指示适用于经编译的代码的目标环境的元数据。(可选地,作业管理实体524可以在编译过程中保留目标环境的知识。)作业管理实体524将输出文件传送到作业执行实体528(该作业执行实体528根据与目标环境相关的配置文件的参数来确定该代码是否应该本地或远程地执行,该代码是否应该在分布式系统中执行等等),并且选择和保全将要执行包含在输出文件中的指令的处理器和存储器。然后,作业执行实体528创建一个或多个作业进程508来执行输出文件。

[0344] 当执行输出文件完成(在离线块或离线模块的情况下)时,或者作为结果是可用的(对于在线模块)时,结果被返回到作业执行实体528。因为,在这种情况下,没有其它挂起的执行(因为该图是一个离线图并且刚执行的块(查询块604A)是唯一被配置的块),结果经由服务器通信实体508返回给网络服务器进程504,并且最终返回到网页客户端进程502上的应用实体514。输出的数据可能包括由该查询返回的所有数据,但也可能包括关于查询数据的各种统计信息(元数据),譬如(但不限于)由查询返回的数据的每列的平均值、标准差以及中值。在实施方案中,元数据和结果数据彼此分开存储。这有利于允许用户查看关于结果数据的信息,而不必加载所有结果数据,这可能是时间和/或资源(如计算机存储器)密集的地方。用户可以通过点击在该块上的查看块结果控件212(参见图4A)来查看返回的数据和/或关于该返回的数据的元数据。默认视图可包括包含在结果中的每一个变量的结果的元数据,并且可以允许用户选择结果数据中的一个或多个选定部分来加载,并且如果需要的话,无需加载所有的数据进行查看(例如,以图形、图表、表格等方式)。此外,该结果返回给应用实体514使得能够对下一个下游块(在此实施方案中为块604b)的属性进行配置,借助于线606a,块604可以得到数据和/或元数据。

[0345] 应当注意的是,执行块(例如,块604a)可能不会在所有情况下均需要配置下一个下游块(如块604b)。块是否在能够配置和/或执行的下游块之前必须被配置和/或执行将取决于数据相关性,通过本公开内容的整体考虑,这将是能理解的。

[0346] 在执行期间还应当注意的是,作业进程508能够与各种分布式数据引擎150通信以接收或访问存储在大数据存储器155中的数据。在作业进程508在嵌入式大数据分析器170中执行的实例中,数据分析器170可以被通信地耦合(直接地或经由一个或多个大的数据请求服务者165)到大数据存储器155。可选地,在作业进程508在不是嵌入式大数据分析器170的一部分的处理器中执行的实例中,该作业进程508可以经由网络耦合到DDE,并且经由大数据请求服务者165请求数据。

[0347] 每个在离线图602中的块604b-604g中的每一个的属性配置和执行以相同的方式进行,并且不需要再很详细地阐述。

[0348] A. 使用离线图的探索和建模

[0349] 与离线图一起工作的DDE用户界面应用的用户通常希望做以下两件事中的一件或两件:(1)探索各种数据集以寻找并发现各种数据集和/或事件之间的关系;以及(2)以实时

实现和连续的预测能力的目标,创建这些关系的模型。后者的完成通常需要用在前者的时间花费的测度。也就是说,在许多(尽管不是全部)的情况下,用户探索数据集来查找使用各种分析工具(主成分分析,傅里叶分析,回归分析,等等)的数据之间的关系,当一关系被发现为呈现具有预测值,该用户将该分析实现为模型并且将该模型用在该过程的实时、连续分析中。

[0350] 在实施方案中,分析服务500可包括自动进行数据的分析(不管是实时数据或历史化数据或两者)的工具,以发现异常现象、该过程环境中的异常情况、数据之间的关系和/或特定数据集的特征。在一个实施方案中,块定义库252包括块255,该块255着眼于数据并发现和/或突出显示异常现象(譬如,在查核的数据集的上下文中的数据中的非常规的尖峰或凹陷)。离线图中的异常发现块能够接收历史化数据为输入,以查找历史化数据的特征,该历史化数据的特征可能是有利于使用例如包含PLS、PCA和其它类型分析的技术中一些技术来进一步探索。可选地或另外地,在线图中的异常发现块能接收历史化数据和实时数据二者,并比较该历史化数据和实时数据,以查找相对于历史化数据的实时数据中的异常现象。在任一情况下,该异常发现块的输出可以包括异常事件发生的一个或多个时间、异常数据的一个或多个来源(例如,其测量和/或设备与该异常相关联)等等。

[0351] 在一个实施方式中,块定义库252包括自主地发现各种数据之间的关系的块255。关系发现块可以随机或半随机选择数据集(在时间上对齐),并且可以对该数据集运行几个不同的分析来查找吸引用户的关系,尤其是因果关系或预测的关系。当然,数据集真正的随机选择不太可能生成许多有用的结果。然而,数据集的选择可以通过,例如,暗示可能存在的数据之间的关系的物理位置、工厂的层次结构,或其它标记来管理。

[0352] 在一些实施方案中,关系发现块可以与异常发现块配合以发现数据之间的关系。作为示例,该发现异常块可输出各种数据集中的各种异常发生的时间。在实施方式中,异常发现块存储异常事件的时间,当不同数据集中的异常事件在同一时间发生或时间上密集地发生时,异常发现块中的一个输出可以是数据源的列表,该列表对应于涉及在时间上彼此接近的异常事件的数据集,并且异常发现块的另一个输出可以是异常现象发生的时间的列表。关系发现块可以接收作为输入的异常发现块的输出,并且可装载相关的时间段的相关数据集,可能利用其它数据集和/或更大的时间段(例如,在异常发生的时间之前或之后延长更长的时间),以将各种类型的分析的应用到数据,并寻求数据之间的因果关系或至少是预测关系。

[0353] 另外,关系发现块的一个或多个属性可能指示用户希望发现的特定的数据集(例如,特定过程变量的值)以发现数据中的关系。关系发现模块可以自主地查找其它数据的关系和指定的数据集之间的关系(物理位置、工厂层级、逻辑关系等相关),对数据执行不同的分析直到一个或多个因果或预测关系被发现并符合一定的标准(例如,95%的预测置信,0.75的相关值,等等)。

[0354] 当然,任何自动块(譬如,关系发现块或异常发现块)被设想为在在线图中或在离线图中是可适用的。

[0355] 如上所述,用户将在画布245上创建离线图,譬如画布600上的图602。图的创建通常如上所述地进行,利用在画布600上布置的一个或多个块604以及一次一个块地配置的每个块604的属性,在配置下一块之前执行每个块604。通过点击与块604相关的数据汇总用户

控件,用户可以精密地查看在图中的任何一点的数据,以看到关于块中数据的各种统计数据(例如,均值、中位数、标准差等),或者可以(经由导线606)连接到块604的输出608,块604用于探索数据(例如,图5D中的块604e)。假设用户已经执行离线图中的块直到并包含与探索块连接的块,用户将能够使用该探索块来查看在各个图中的数据,看到与该数据相关联的元数据等。例如,在离线图602中,探索块604e将允许用户查看在数据已经由块604b处理后的数据。类似地,探索块604g接收来自PLS块604f和来自列块604d中的数据。在后一种情况下,探索块604g可以允许用户将PLS模型的输出(其中可以包括变量的预测值)可视化,并将该输出与用来创建PLS模型的这些变量的实际值(来自块604d)进行比较。通过观察数据的图,用户可以能够确定PLS模型是否精确地表示了该过程,如此,其具有该过程中的一些预测值。

[0356] 在实施方案中,与块255相关联的数据汇总用户控件将特定于块的类型和功能。例如,加载数据的块(例如,查询块或负载块)可以被编程为当其被激活时,数据汇总用户控件导致显示表示通过该查询加载或返回的数据的各种统计数据。在实施方案中,对于包括在该数据中的每个数据源,该数据可以包括平均值、中值、标准差、最大值和最小值。相比之下,对数据执行分析(例如,PLS或PCA分析)的块可以在汇总用户控件被激活时显示不同的数据。该分析块可以显示一个或多个R平方值、用于PLS和/或PCA分析的系数、差异值、观察计数(例如,对于特定于数据源,包括多少个时间序列值)以及可定制的图形,以允许用户选择查看哪一个数据(即,从哪些数据源选择)。在实施方案中,探索块也将根据连接它(多个)数据输入的(多个)数据输出来以不同的方式运行。也就是说,探索块中显示的和/或能够获得并显示的数据的格式和类型可以取决于生成该探索块的输入的块的类型。

[0357] 参考图5D所示,离线图602的功能(这仅仅是许多可能性中的一个图形配置)现在将进行一般化的描述。离线图602以查询块604开始,查询块604的目的是发现特定的数据集并加载它以进行分析。仅提供为示例,该特定的数据集可以是与特定过程变量相关的历史化过程控制数据,该特定过程变量表示在三个具体时间的异常差异。然而,在一般情况下,查询块604a可以查询任何存储的数据,包括但不限于,大数据存储器155中的任何数据、存储和/或缓存在一个或多个数据源168中的任何数据、存储在外部数据源的数据,该外部数据源譬如天气数据源、供应链数据源、可交付的跟踪数据源等。在实施方案中,数据库或文件存储器(未示出)存储描述与该分析服务相关联的在线和/或离线图的文档(例如,JSON文档),在这种情况下,查询块604a可以被配置为查询关于所执行的计算类型、在计算中使用到的一个或多个数据源、结果的质量等的数据库或文档。后一种类型的查询可以有利地构建图形的历史或知识基础。查询块604a可以被配置为选择其它过程变量和某个数量的其它过程变量的值,并且甚至可以配置为选择采样率和/或特定过程变量和其它过程变量的值(例如,仅在对应于三个事件中的每个事件的前一小时到后一小时的时间段中选择)。也就是说,该查询可以选择不相交的时间段,并且发现在这些时间段中生成的用于任意数目的变量的数据,这些中的至少一个是用户所希望预测的。

[0358] 一旦执行,由查询块604a获取的数据可以由填充块604b被使用。在该示例的图602中,对应于在该变量没有值但其它变量中的一个变量有值的时刻,填充块604b可为该变量填充数据。也就是说,当所查询的值中的一个的采样的速率比其它的查询的值更频繁的进行采样时,填充块604b可以插入用于较少采样变量的值来匹配较高频率采样的变量的频

率。在实施方案中,填充块604b可以推断以较低频率采样的变量的值。例如,如果较高频率采样的变量的四个值在较低频率采样的变量的样本之间发生,则块604b可能会发现较低频率采样的变量的两个连续值(例如,1.0和6.0)之间的差,除以四,并用连续较大或较小的值(例如,2.0,3.0,4.0和5.0)填充进“缺失(missing)”的四个值,从而对于以较高频率采样的变量的每个值均具有对应的较低频率采样值的值。(当然,这不是在实时分析中填充值的合理的方法,因为后续的值还尚未得知。)在其它实施方案中,对于较低频率采样值,填充块604b可以简单地使用较高频率采样值来填充进“缺失”的值。在上面使用的例子中,缺失的值将分别以值1.0来填充。

[0359] 在发现数据并插入值来创建没有缺失的的数据集后,填充块604b的输出被提供给3个块:探索块604e,第一列块604c和第二列块604d。探索块604e已经在前文进行了描述。第一和第二列块604c和604d分别操作以将变量(列)从数据拉取出来。数据存储在表中,表中的每一列表示一个变量,每列的行表示在不同时间(即,时间序列数据)的相应的变量的值。例如,第一列块604c可以选择所有列(即,所有的变量)包括除了表示异常变化的过程变量之外的变量的数据,同时,第二列块604d可以例如选择包括表示异常变化的过程变量的数据的列。

[0360] 列块604c和604d中的每一个块的输出被发送到PLS块604f。块604f的输入端610f2可被配置为接收独立的、解释的或输入变量的值,在实施方式中,块604f的输入端610f1还可配置为接受相关的、经解释的或输出变量的值。虽然模型块(例如,离线图602中的PLS块604f)的输入的特定布置和/或功能在图602中示出,但取决于块的功能、输入的类型、输入的数量,输入的布置和功能可以是不同的。

[0361] 块604f的输出端608f1被线连接到探索块604g。因此,探索块604f接收由PLS块604f输出的值以及表示异常变化的变量的值。用户使用探索块604g能够看到块604f的输出如何在任意给定的时间与因变量相比较,并且确定预测值。虽然块604f的输出端608f2在图5D中并未连接到其它块,但说明了一些块和线的属性。具体地说,输出端608f2是对象输出端。该对象输出端适用于输出代码;在此示例中,例如,该对象输出端输出如所编程的灵敏度模型,包括模型、数据和输出。当然,这意味着连接到输出端608f2的线能够通过线携带模型。例如,连接该线到新的块的输入端能够利用整个图的上游来对该块进行编程以创建包括所有功能与数据的块。

[0362] 因此,现在可以显而易见的是,用户可迭代地修改离线图602并且探索该结果来构建具有给定变量的所需的预测值的图。在这样做时,用户可以使用不同的数据集、相同的数据集的不同的子集、不同的建模/分析技术等等。换句话说,用户可以查询或加载不同的数据集(例如,替换或修改或添加到离线图602中的块604a),可以通过尝试预测不同的变量来对数据进行不同地分段(例如,通过与改变在每个块604c和604d中所选择的列),可以尝试不同类型的分析(PCA,PLS等)和/或特定分析的不同的属性(例如,通过替换或修改块604f)等。

[0363] 鉴于上面的示例,可以明确的是DDE用户界面应用和随之的分析服务在很多的收集大量数据的系统中具有优越的实用性。关于过程控制系统,过程控制环境相对近地扩大所收集并存储的数据的量来包括(在一些情况下)在环境中生成的所有数据。然而,过去的系统曾生成重要数据,这些系统仅使用用于过程控制系统的数据中的一小部分,例如,将之

送入过程控制算法并且在操作员工作站上显示其中的一些;存储相对少的数据以用于以后的分析和/或使用。例如,每秒采样一次的变量的值可以每分钟仅存储一次,以平衡具有受限的储存器和网络容量的因果分析的需要。此外,所存储的数据经常被压缩,这导致了不可靠的数据。

[0364] 与此相反,当前的系统存储在工厂中所生成的数据。在实施方案中,样本每记录一次,系统便存储每个变量的值一次,并且还可以存储之前从未历史化的数据,譬如,操作员的输入、操作员到操作员的消息、视频等。此处描述的DDE的用户界面应用和分析服务合作以有利于探索除了其它数据(例如,来自其它工厂的数据、气象数据等)的所有数据,并且有利于发现之前未被发现的或无法发现的关系。由于发现各种过程值之间的关系、过程值的相互关系、过程值之间的关系,设备工程师、操作员和维修人员能够更好地设计、构建、操作和维护过程工厂,这转而使得过程工厂更便宜、更高效、更易于操作和维护,生产出更好的产品、对环境更少的负面影响,并且对人员和周围社区更加安全。

[0365] B. 创建在线图

[0366] 当然,所发现的关系可以为了过程的改善操作而被使用的途径是通过使用所发现的关系来执行实时连续预测。具体地,在该过程的历史数据中发现过程值和其它数据之间的一个或多个关系,从而一组值能够被用来预测其它的值(或一些事件的发生),所发现的关系能够用来查看来自过程的实时数据来预测相同的值(或相同的事件的发生)。如下所述,DDE的用户界面应用和分析服务有利于所发现的关系的使用以执行预测分析。

[0367] DDE的用户界面的应用包括允许其将离线图(譬如离线图602)转换到在线图(即,使用至少一个实时值来预测工厂操作的某方面的图)的功能。如上所述,在线图与离线图不同的地方在于其绑定到至少一个实时数据源(而不是纯粹的历史化数据),并且提供了实时的、连续预测输出,该输出可以被查看、存储,和/或用在控制算法中以触发报警、警报和/或影响过程工厂的操作中的变化。

[0368] 只要离线图包括至少一个模型生成块,DDE用户界面应用的用户便能够激活在线切换用户控件248f,并且离线图将被自动转换为在线图,并显示在画布245上。再参考图5A,在实施方案中,通过网络服务器进程504的图变换器实体530,完成离线图到在线图的转换。

[0369] 在最简单的实施方案中,图变换器实体530定位模型生成块(例如,离线图602中的PLS块604f),移除所有输出线(如线606h),增加了写入块,并连接线模型生成块的输出与写入块的输入端之间的线。该写入块通常将输出自该模型的值写入到数据储存位置、过程控制算法(例如,在过程控制装置中的控制器或功能块中执行控制算法),和/或仪表盘。图变换器实体530还用对应的在线数据加载块替换离线数据加载块,而不是加载一批历史化诗句,该对应的在线数据加载块绑定到来自过程控制环境的至少一个的实时数据源。

[0370] 然而,在各种其它实施方案中,图变换器实体530可适用于不止仅替换离线图中的两个块。在一些实施方案中,图变换器实体530主动地移除来自离线图的块来创建在线图。例如,如果模型块需要两个输入以创建模型(例如,一系列的X变量和变量Y),将在不同的块处生成两个输入。然而,当实施开发的模型(即,位于在线图中)时,如果该模型仅具有一个输入,则在先前提供数据到另一输入的块将不再是必要的,并且可以被移除。作为另一个例子,在创建在线图时,图变换器实体530可以移除探索块(例如,图5D中的探索块604)。

[0371] 在一些实施方案中,图变换器530可以引用架构库(未显示),该架构库可以提供每

个块类型的离线和在线架构、哪个架构定义了与块的在线和离线版本相关联的输入和输出、它访问什么数据源等等。在替代实施方案中,块定义库252中的每个块定义255限定了用于该块的在线和离线架构。

[0372] 在一些实施方案中,图变换器实体530可以被配置成在离线图变换到在线图的期间执行或不执行可选的操作。例如,一些可选的操作可以包括,但不限于,标记预测值以便其能作为过程控制算法的输入,输出仪表盘上的预测值的连续曲线,并将预测值写入到分布式数据引擎。在一些情况下,由图变换器实体530生成的输出中的至少一些可被提供以用于分析和/或控制系统的可视化(譬如面板,趋势图标等)。

[0373] 在实施方案中,数据加载块(加载、查询等)所绑定的实时源、模型块输出所投递的位置和/或与输出数据相关联的虚拟标记在某种程度上均是可配置的,给定的实施例实现每个选项。在实施方式中,每个实施方式在驻留在网页客户端进程502中的应用实体514的环境中配置,例如,被配置为模块属性和/或块属性。例如,如在图5F中所示的,对于模块的模块属性可以包括数据投递位置以及实时数据绑定位置。在图5F中,当用户激活模块属性控件632时,模块属性对话框630针对该模块而显示。模块属性对话框630包括模块名称栏位634,其允许用户输入该模块的名称。模块属性对话框630还包括投递栏位636,以允许用户指定由在线图在执行期间输出的数据被写(即,投递)到哪里。数据被投递的位置可以是存储位置,在该位置处,数据被历史化,过程控制环境中的控制模块使用所投递的值作为控制算法的输入来控制过程工厂的操作,或者两者。用户控件636a可以打开允许用户通过查看图像形式的可使用的位置(例如,以树形形式,目录结构,可选位置的列表)来选择该数据将要被投递的位置的资源管理器窗口。类似地,栏位638允许用户将绑定的数据路径指定为在在线图的执行期间的在线数据源。同样,用户控件636a打开资源管理器窗口以允许用户选择数据将被接收的位置。

[0374] 可替代地,离线块可以包括与该块的离线版本相关的属性,并且该块的在线版本可以包括用于该块的在线版本类似的属性。例如,加载块可具有该块的离线版本中的“源路径”(例如,\DDE_path\device_tag)属性,还可能有该块的在线版本中的“装置标签”(例如,\unit1\device_tag)属性,这使得加载块能够对应于存储在离线块的源路径中的数据的源读出的实时数据。同样,在线图中的写入块可以具有指定输出数据应该被写入的位置的属性。

[0375] 作为另一个替代方案,此时参照图5G,在离线图640中的块均可以具有用于离线和在线操作的配置。作为示例,加载块642可被选择,并且可以具有相关联的块属性对话框644。块属性对话框644包括块名称栏位646,其允许用户输入该块的名称。块属性对话框644还包括离线源栏位648,其允许用户指定在离线图中将要加载的数据在哪里可以被发现。要加载的数据的存储的位置可以是存储器位置,数据在其中被历史化。用户控件648a在资源管理器窗口可以打开允许用户选择来自哪里数据将被加载,通过以图形格式查看可用的位置(例如,以树形形式、目录结构、可选位置的列表)。类似地,栏位649允许用户将绑定的数据路径指定为在在线图的执行期间的在线数据源。同样,用户控件649a打开资源管理器窗口,以允许用户选择该数据将从哪个位置被接收。当然,控件648、648a、649、649a可以是在块定义中指定的任何控件,并且如前所述地可以调用适当的插件以执行。

[0376] 在实施方案中,通过数据源模块532来管理:一方面,指向存储的、历史化的数据的

路径(离线块通过该路径可以获取/查询数据)之间的关系;另一方面,对应的实时数据源(在线块可以绑定到该实时数据源)。数据源管理器模块532可以被集成为网络服务器进程502的一部分,可以被包括为图变换器实体530的一部分,或者可以仅仅是一个查询表,当离线图转换到在线图时,图变换器实体530将参照该查询表。在至少一些实施方案中,数据源管理器模块532与大数据请求服务商165相配合来根据在线块的要求来设置实时数据流。

[0377] 图5E示出了对应于图5D中的离线图602的在线图622。在线图622包括块624a、624b、624c、624f和624h,以及线626a、626b、626e和626j。其中,块和/或线对应于离线图602的块和/或线,与块相关联的字母是相同的。

[0378] 在线图622的操作和执行以前述的与离线图602大体上同样的方式进行。在线图622,一旦所有的属性被配置(和存储在配置文件中的相关联的参数),可通过接合部署用户接口控件248j来“部署”。激活控件248j将把配置文件发送到执行服务进程506,在那里它会根据配置文件中的参数而被编译。然而,在执行为一个或多个作业进程508时,作业进程508将从大数据接收器接收实时数据,例如或直接通过由大数据请求服务者165之一来设置的流来接收,而不是经由数据分析网络界面175从大数据储存器155下载或查询数据。同样,作业进程508可以将输出数据返回到大数据储存器155和/或一个或多个数据源(例如,控制器、其它DDE等),而不是将输出数据发回到作业执行器528来经由网络服务器进程504的服务器通信实体518来仅发送到应用实体514以显示在仪表板上。

[0379] 用于工业性能监控/分析的查询语言

[0380] 与运行有工业性能监控/分析系统100的过程控制系统或环境相关联的数据源通常提供时间序列数据,虽然其它类型的数据也可以被使用(例如,在一个或多个过程工厂5中分别执行的来自多个批次的截面数据)。时间序列数据可包括过程工厂5(包括现场设备15-22和40-46)内的各种类型的测量设备的各种类型的数据测量结果。数据源能够以它们格式从通常已知到专有格式的宽范围内进行变化,例如,OSIsoft PI,DeltaV历史库,SEEQ,FF3,和/或手动捕获的电子表格中的格式。一些数据源可以包括关系型数据库,而其它的数据源可以包括非关系(NoSQL的)数据库。进一步,数据源可以不是数据库,不使用文件目录或文档(例如,XML文档)的文本来存储数据。除了查询语法的差异,由于数据存储方式的差异,各种数据源可能需要根本上不同的查询结构。例如,文档导向的非关系型数据库(譬如Mongo)基于文档存储数据,而不是将数据存储在与通过关系型数据库(如MySQL)中的SQL查询是可用的表中。因此,除了格式和语法上的不同,对存储在不同类型的数据源中的数据的查询采用不同结构和规则。如图6A所示,在此所描述的工业性能监控/分析系统和技术(并且,尤其是DDE用户界面应用)利用标准化查询来与数据源702a-702f中的每一个相接,而不是使用每个不同的数据源的原生查询机制来访问存储在其中的数据。

[0381] 图6A示出了经由查询执行服务706通信地连接到DDE用户界面应用的请求块708的不同数据源702a-702f以及数据提供者704a-704d的框图。数据源702a-702f中的每一个均是电子数据源,以计算机可读格式存储数据。某些数据源702e和702f可接入查询执行服务706,而无需中间数据提供者704,譬如通过内部或网络数据连接。其它数据源702a-702d经由一个或多个数据提供者704a-704d来与查询执行服务706连接。数据提供者704a-704d可被配置为访问、搜索、筛选、读取和/或写入去往或来自相应的数据源702a-702d。在一些实施方案中,数据提供者704a-704d可以接收利用数据源702a-702d的相应的特定于数据源的

查询格式的特定的数据源的查询。在其它实施方案中,数据提供者704a-704d可被配置成使用来自查询块708的定制的查询语言来接收标准化查询709,并且将标准化查询转换成利用目标数据源702a-702d的特定查询机制的相应的特定于数据源的查询格式。数据提供者704a-704d或数据源702e-702f可以包括处理并执行对特定的数据源702的特定于数据源的查询的执行引擎。执行引擎可以是数据源702本身的一部分,或者它可以与数据源702相关联的数据提供者704的一部分(例如,与数据源702b相关联的数据提供者704b)。每个数据源702可以是数据库或文档存储,譬如,包括包含了数据的多个JSON文件的文档。在一些实施方案中,数据源702a-702f甚至可能包括实时数据源,譬如由过程工厂5的现场设备15-22和40-46所生成的数据流。这种实时数据源可以通过解析获得的数据并且保留或提取具体请求的数据的一部分来查询。

[0382] 数据源702a-702f可以被配置为存储在任何已知的或以后开发的格式的数据。数据源702a、702b和702e被显示为数据库以表示存储在关系型数据库中的数据。这种关系型数据库的例子包括MySQL或其它在表中存储数据并利用数据的SQL格式的查询的数据库。数据源702c、702d和702f被显示为文件或数据输入集以表示非关系数据存储,譬如NoSQL数据库或非数据库数据源。非关系型数据库的例子包括面向文档的数据库,譬如基于文档而不是表来存储数据的MongoDB或CouchDB。因为使用非关系型数据库来储存或搜索非常大量的数据或复杂的数据集可以更有效,因此,这种数据库经常用于大数据分析。由于非关系型数据库中的数据通常不是以表格格式(其形成SQL数据库的基础)来布置,因此,查询这种非关系型数据库需要不同的技术和不同的语法。由于SQL查询被广泛使用和众所周知的,其也经常用于数据存储和分析。通过使用用于标准化查询709的标准化查询格式,本发明允许用户访问具有相同查询格式的SQL或NoSQL中的数据,其中,标准化查询709被转换成利用特定于数据源的格式的特定于数据源的型查询。标准化查询格式进一步允许来自多个不同类型的数据源702的数据被组合成具有一致的数据结构和格式的数据集。因此,查询块708可以作为超级连接器,使用标准化查询709来连接具有不同的数据结构、格式和查询机制的数据源702。

[0383] 该查询块708可以是数据分析工作室中实现的块,如上所述,用来指定要从数据源702a-702f获得的数据。例如,查询模块708可以是具有一个或多个块属性的配置文件,包括指定标准化查询709的属性。标准化查询709可以由查询块708引用的独立的文件(诸如JSON文件)。可选地,查询块708可以是包含或指示标准化查询709的任何对象,譬如,在任意一个分布式数据引擎150中操作的过程或例程。查询块708接收标准化查询709(譬如,通过用户选择或输入),并且标准化查询709使用标准化查询格式。通过利用标准化查询的格式,对数据源702a-702f的查询无需需要数据源702a-702f中特定类型或结构的数据的知识的用户、技术人员或数据请求实体。标准化查询格式可以由一些数据源使用的现有的查询格式,也可以是不被任何数据源使用的不同的查询格式。在后一种情况下,利用标准化查询格式的语法的标准化查询不能直接执行或实现以从数据源702a-702f获得数据。标准化查询可以指定数据源702a-702f中的一个或多个(或数据提供者704a-704d中的一个或多个),从数据源702a-702f中的一个或多个(或数据提供者704a-704d中的一个或多个)获得数据。可替代地,该查询块708可以包括指定一个或多个数据源702a-702f的属性。

[0384] 查询执行服务706从查询块708接收标准化查询709,从而查询数据源702a-702f中

的一个或多个。从查询块708接收该标准化查询709可以包括：接收包含指示从数据源702a-702f获得的数据的查询参数的文件。致使查询数据源702a-702f可以包括提取查询参数并基于查询参数来生成一个或多个特定于数据源的查询。然后，每个特定于数据源的查询可以由执行服务706来执行并或发送到数据源702e-702f或数据提供者704a-704d以被执行以获得由查询参数所指示的数据。该查询参数可以指示从一个或多个数据源702获得的特定数据，其可以是测量数据（譬如来自现场设备15-22和40-46的测量值），由测量数据计算或以其它方式获得的测度，和/或关于数据源702或存储在其中的数据的元数据。例如，该元数据可以包括的存储在数据源702中的数据的类型、源、或质量，包括对数据执行的计算类型。该元数据在开发过程模型或来自系统的数据源702的知识库是有用的。

[0385] 查询执行服务706可以是上面讨论的执行服务进程506。查询执行服务706可以包括分析服务500的任意部件，包括数据服务520、作业监听器522、作业管理器524、作业执行器528，或数据源管理器532。例如，查询块708可以由应用514使用来自用户的输入来创建，包括存储在JSON文件中的标准化查询709。在接收到用户请求或其它触发事件时，应用514可能使得标准化查询709从或通过网络服务器进程504传输到执行服务进程506（其中执行服务进程506可以充当执行服务706）的作业监听器522。作业监听器522可能使得一个或多个作业管理器524接收标准化查询709，这将进一步使得一个或多个作业执行器528将标准化查询709转换成与一个或多个作业进程508相关联的一个或多个特定于数据源的查询。然后，作业进程508可以导致一个或多个数据源702（即，数据源168或大数据储存器155）被使用特定于数据源的查询来查询。通过查询所述一个或多个数据源702而获得的数据然后可由作业进程508、执行服务进程506、数据服务520和/或应用514来接收。

[0386] 在一些实施方案中，查询执行服务706将标准化查询709转换到利用数据源702e或702f原生的特定于数据源的格式的特定于数据源的查询。在替代实施方案中，查询执行服务706可确定一个或多个数据源702a-702f来查询，并且向一个或多个数据提供者704a-704d提供标准化查询709以被转换成特定于数据源的查询。数据提供者704可以与数据源702分开，或者可以与数据源702组合。数据提供者704a和704d被示出为分别通信地连接到数据源702a和702d，但仍然是分开的。与此相反，数据提供者704b和704c被示出为分别包含数据源702b和702c或与数据源702b和702c相组合。例如，数据提供者704b可以是服务器（譬如数据历史库）或数据库接口程序，在其上数据源720b被存储为数据库。作为另一示例，数据提供者704a可以类似地是服务器或数据库接口程序，该接口程序连接到外部数据源702a，譬如，存储在与服务器通信连接的外部存储设备中的数据库。作为又一示例，数据提供者704a-704d可以是作业进程508，该作业进程508包括当它们从来自查询执行服务706的标准化查询709所接收到标准化查询709或查询参数服务时将标准化查询709转换成特定于数据源的查询的算法。

[0387] 为了说明标准化查询在过程控制和分析中的使用，图6B示出了数据分析工作室中的查询块708的框图。查询块708与从一个或多个数据源702a-702f获得过程数据的标准化查询709相关联。例如，用户可以在标准化查询格式中将标准化查询709定义到查询块708的栏位。标准化查询709可以与查询块708一起存储，或者它可以被存储在单独的文件中。例如，查询块708可以是配置文件，该配置文件包括存储有包含标准化请求709的JSON格式文件的标识符的查询属性。当一个事件发生（譬如，时间、过程控制状态或用户选择的发生），

标准化查询709可以从查询块708被标识并发送到查询执行服务706以从一个或多个数据源702a-702f获得所请求的数据。由标准化查询709指示的数据或由与该数据相关的信息(例如,摘要信息,验证元数据等)来指示的数据可以返回给查询块708或另一数据接收实体。其它数据接收实体可以包括数据分析工作室中的另一个块、作业进程508、查询执行服务706、数据服务520、应用514、程序存储器,或任何其它服务、例程、进程或可以进一步分析、显示、或存储与过程工厂5相关联的数据的设备。如以上所讨论的,在一些实施方案,该过程数据可以经由一个或多个的数据提供者704来得到。所获得的数据可在查询块708、查询执行服务706或其它数据接收实体处以所期望的格式来接收,或可以被查询块708、查询执行服务706或其它数据接收实体格式化。例如,查询执行服务706可通过使得标准化查询709转换成一个或多个特定于数据源的查询并且该源特定查询的结果以所希望的方式被格式化来生成数据帧。在一些实施方案中,该格式化的数据帧可以被返回到查询块708。从一个或多个数据源702获得的数据可以包括由特定于数据源的查询获取的多个数据点,其数据点可以对应于处理变量,譬如在过程工厂5中的测量或从该测量所获得的值。该数据点可以以时间序列的方式表示点,数据点具有与每个点相关联的时间戳。另外,该数据点可以表示与位置、进程批次或其它标识特征相关联的横截面数据。通过执行标准化查询获得的数据可以被称为数据帧以表示其可变结构,其可以使用或不使用表格格式。

[0388] 查询块708可以提供使用对填充块710的标准化查询来获得的数据帧,这根据与填充块710相关联的规则来填充数据集中的空项或非数字(not-a-number, NaN)项。执行该操作是为了审查数据或检查数据帧中的错误的的数据或误差的指示,以及为了增加达到所需的采样率所需要的数据点,如在别处和此处所述的。然后,数据帧可以用于进一步的分析和/或过程控制。如图所示,通过自变量块712和因变量块714,一个或多个数据集可以从该数据帧中选取。例如,自变量块712可从接收自填充块710的数据帧中选择的一种或多种类型的数据,数据对应于数据匹配指定的参数或特征(例如,压力或温度测量)或数据列(其中,数据帧对应于一个数据表)。因变量块714可以类似地用于从接收自填充块710的数据帧中选择一种或多种类型的数据。虽然,示出了两个块712和714,但是可以使用任何数量的类似的块。然后,分析块716可接收由自变量块712和因变量块714所选择的数据,如块之间的连接器所示。分析块716可以执行其被配置的任何类型的分析。例如,该分析块716执行偏最小二乘(PLS)分析来确定自变量块712的数据对因变量块714的数据的影响。探索块718可被连接到分析块716和因变量块714以测试查询块708以及其它块和连接的适当的配置。探索块718可生成可被用户存储或由查看的图形、表格或文本输出。尽管前面的描述将块708-718讨论为执行特定的动作,可以理解的是这些块可以反而使得这些动作由分析服务500的网络服务器进程504、执行服务进程506和/或作业进程508的来执行,如在其它地方所讨论的。

[0389] 为了进一步论证标准化查询709的使用,图6C示出了利用示例性的标准化查询格式的示例性查询。该示例性查询作为利用符合JSON的格式以最大化跨平台的兼容性,但任何其它格式也可以被使用(例如,基于XML的格式,基于CSV格式,等等)。示例性查询以查询名称开始,接着是用于返回数据帧(被指定为JSON所有)的格式指示和查询格式的版本的指示。“timeSelector”阵列指示多个时间帧,该多个时间帧指示一时间范围,对于该时间范围,数据通过“startTime”和“endTime”对象来被返回。在示例性查询中,开始和结束时间已经以协调通用时间中的日历数据和时钟时间的方式来规定的,由字母“T”所间隔开。如由示

例性查询所指示的,每个时间帧可以通过的排除的时间段(数据不被查询)来分离,这对应于时间帧中的一个时间帧的结束时间和另一个时间帧的开始时间之间的时间段。继“timeSelector”阵列之后,“sampleRateSecs”对象指定以秒计算的采样率,其指定每单位时间将获得多少数据点。示例性查询的最后部分是一个“columns”阵列,其指示了将从数据源702获得的数据的参数(即,将要获得的过程变量)。示例性查询指定四类或列的数据以返回。四列中的每一列由指定数据源和/或数据源的过程变量(即“FT630B/DENS.CV”,“PT615/WIRED_PSIA.CV”,“TT6079/INPUT_1CV”,和“630.molefrac.c5”)的“tag”对象来标记。四列中的每一列包括用于标记数据帧中返回的数据的“alias”对象、用于指定数据的数据类型(例如,浮点值、长整型值、文本串等)的“dataType”对象、“renderType”对象(例如,值、计数等),以及指定返回的数据的格式(即,“0.###”指示由小数后具有三位的浮点值构成的格式)的“format”的对象。

[0390] 与列相关联的“tag”对象可以定义数据源702,从数据源702可以获得数据。可替换地,在示范性查询不明确指定数据源702时,数据源702a-702f可以通过块708的其它属性来制定。在利用标准化查询格式的类似的查询中,数据源702可以通过对象明确地在查询中指出。该指示可包括指向源对象的路径或指示数据源类型(如MongoDB、CouchDB、SQL等)的值。当合适时,数据源702的指示还可以指示另外相关联的数据提供者704。可替换地,数据源管理器532可以提供数据提供者704的指示或由数据源702使用的结构或格式。在一些实施方案中,该源的指示可以进一步包括数据源702的将要获得的数据的每个参数或列的单独的指示。因此,利用标准化查询格式的查询可以从多个不同的数据源702来获得数据。

[0391] 如上所述,标准化查询709被转换成特定于数据源的查询,特定于数据源的查询可以针对每个目标数据源702而被执行。该转换可以通过查询执行服务706来执行,或者可以由每个目标数据提供者704a-704d来执行。图6D示出了示例性的查询方法740,在该方法中,查询执行服务706将标准化查询709转换成一个或多个特定于数据源的查询,并且图6E描述了示例的查询方法760,在该方法中,数据提供者704将标准化查询709转换为源特定的查询。

[0392] 方法740可以始于接收利用标准化查询格式的标准化查询(框742)。标准化查询可以通过查询块708、查询执行服务706或数据提供者704a-704d来接收。该查询可以包括任何上述特征或元件。该查询可以经由直接输入而接收自用户,或者可以接收自程序块或例程,其可以生成和/或存储基于用户输入的标准化查询。标准化查询可以在符合JSON的文件中构造或以符合JSON的格式来构造,或者其可以使用任何其它方便的语言或语法来构造。标准化查询可利用一些数据源702a-702f可执行的语法来获得数据(例如SQL),或者标准化查询可利用数据源702a-702f无法直接执行的语法来获得数据(例如,图6C中所示的示例性查询)。标准化查询可以标识将要获得的数据(例如,数据的时间帧和参数)、数据源702和/或数据将被提供的格式,其中,数据获得自数据源702(例如,数据源702a-702f或数据提供者704a-704d)。如果标准化查询709在查询块708处被接收,其然后可以在触发事件发生时被发送到查询执行服务706,譬如,用户选择一选项以运行该查询或运行由分析服务500中的其它对象、例程、块、进程、服务或功能对查询数据的请求。

[0393] 收到标准化查询709后,查询执行服务706(或数据提供者704)可以确定从中可以获得数据的一个或多个数据源702(框744)。在一些实施方案中,该确定可以包括接收或确

定一个或多个数据源702的指示。该指示可以是特定于每个数据特性(例如,过程变量或列)或者可以是对于整个查询是一般的。该指示可以被包括作为标准化查询709中的对象或标签,或者它可以通过查询块708的另外的属性来指示。例如,用于数据列或数据类型的数据源702的指示可被添加到指定数据列或数据类型的对象或标签之前。可替换地,查询块708或数据提供者704可与一个或多个数据源702以其它方式相关联。这种关联可能受与标准化查询709分离的一个或多个数据源702的指示的接收影响。例如,当查询块708通过接收对数据源702(或数据提供者704)的选择时,经由数据源属性,查询块708可与一个或多个数据源702(或与一个或多个数据提供者704)相关联。同样,当数据提供者704a-704d被配置时,数据提供者704a-704d可以与一个或多个数据源702a-702d相关联,无论是固有地或通过配置过程中所创建的关联。

[0394] 使用关于一个或多个数据源和标准化查询709的信息,特定于数据源的查询被生成以从一个或多个数据源702获得数据(框746)。该源特定的查询可以由查询执行服务706生成并且被发送到每个数据源702a-702f以基于在框744处所标识的数据源来进行查询(直接地或经由数据提供者704a-704d)。可替换地,查询执行服务706可发送标准化查询709以通过与数据源702a-702d相关联的一个或多个数据提供者704a-704d中的每一个来将标准化查询709转换为特定于数据源的查询,基于在框744处所标识的数据源,数据将从数据源702a-702d获得。例如,查询执行服务706可能导致一个或多个作业进程508如同数据提供者704来转换标准化查询709。无论查询执行服务706或数据提供者704是否从标准化查询709生成特定于数据源的查询,该特定于数据源的查询必须利用可以由数据提供者704或数据源702来执行以获得存储在数据源702中的数据的特定于数据源的格式。特定于数据源的查询格式可以是用于查询或以其它方式访问存储在电子数据源(譬如SQL、MongoDB、CouchDB)中的数据的任何已知的或以后开发的格式或语法。

[0395] 生成所述特定于数据源的查询可以包括标识标准化查询709中的查询参数。该查询参数可包括与数据的时间帧或其它特性相关联的参数,譬如由“timeSelector”阵列所指定的时间帧、由“sampleRateSecs”对象指定的采样率,以及由前文所讨论的示例性标准化查询中的“列”阵列(特别是“标签”对象)指定的数据参数。这些查询参数可以包括与类型或测量、测量设备的类型、或特定的测量设备(譬如现场设备15-22和40-46)相关联的过程变量的指示。一个或多个时间帧可以被指定以用于要被查询的每个数据源702,并且不同的数据源或数据源内具有不同的参数(例如,不同的列,不同的过程变量等)的数据可以与之相关联的不同的时间帧。所标识的查询参数可以由查询执行服务706或数据提供者704提取自标准化查询709,并转换成特定于数据源的查询的各方面。所生成的特定于数据源的查询从而包含代表标准化查询709中指定的实质性参数的查询语言,该实质性参数被包括在特定于数据源的查询格式的特定于数据源的查询中。

[0396] 在一个或多个时间帧和采样率被指定的实施方案中,特定于数据源的查询可以被生成以提供与在由每个时间帧指定的时间段中的特定时间处的多个数据点相关联的数据。每个数据点可具有指示与数据点相关联的指定的时间的的时间戳,时间戳是在与时间帧相关联的开始时间处或之后的采样率的周期的整数倍,并且在与时间帧相关联的结束时间处或之前开始。如图6C所示,多个时间帧可以通过标准化查询来指定。在该实施例中,可以生成特定于数据源的查询以获得在具有对应于多个时间帧中的每一个时间帧中的采样率的时

间戳的数据点处的数据。在某些情况下,针对所有时间帧中的一些,分离的特定于数据源的查询可以被生成。如果一个或多个时间帧为使用采样率来制定,多个数据点可以利用时间戳来获得,但是时间戳可以采用时间帧内的任何值。因此,时间戳将指示每个数据点测量(或在潜在的过程值被测量)的时间。

[0397] 然后,查询执行服务706或数据提供者704将使得一个或多个特定于数据源的查询被执行以查询标识的数据源702(框748)。这可以包括将数据特定源的查询从查询执行服务706发送到一个或多个的数据提供者707a-704d和/或发送到一个或多个数据源702a-702f。类似地,在一些实施方案中,数据提供者704可将特定于数据源的查询发送到数据源,以获得所请求的数据。在进一步的实施方案中,特定于数据源的查询可以由接收数据源702的执行引擎来执行。在其它实施方式中,数据提供者704可以通过访问和操作存储在数据源702中的数据来执行一个或多个数据源指定的查询的。基于一个或多个特定于数据源的请求的执行,从一个或多个数据源返回的数据可以被发送到数据接收实体或由数据接收实体接收,该数据接收实体可包括数据提供者704、查询执行服务706或查询块708。因此,数据接收实体从一个或多个数据源702获得所请求的数据(框750)。

[0398] 在采样率被指定的实施方案中,获取数据可以包括确保该数据包括与具有时间戳的数据点,其中,时间戳与由时间帧和采样率指示的采样点中的每一个相匹配。这可以为了每个时间帧和采样率来完成。如上所述,这些采样点对应于在该时间帧的开始时间处或之后的采样率的周期的整数倍,并且在时间帧的结束时间处或之前发生。如果存储在数据源702中的数据点不包括针对一个或多个与样本点相关联的时间的数据点,具有对应于采样点的时间戳的另外的数据点将被增加。该另外的数据点可以被分配以与数据源702中的数据条目的值相关联的值,该值在时间上在时间戳之前且最为接近。例如,数据源702可以包括时间的数据点7:01:5500(具有值V1),7:02:0500(具有值V2),以及7:02:5500(具有值V3),但在标准化查询中指示的时间帧可以是7:02:0000(开始时间)到7:03:0000(结束时间),并且采样率的周期可以是0:00:5000(对应于每半秒采样一次的采样率)。在该实例中,所获得的数据点将具有时间戳7:02:0000,7:02:5000,和7:03:0000的,这些将分别具有值V1、V2和V3。因此,最新的前值V2(测量于7:02:0500)被用于具有时间戳7:02:5000的数据点,即使另外的值(时间7:02:5500处的V3)在时间上更为接近后续的时间戳。另外,在所需采样时间之间时间生成的过量的数据点可能会被从所获得的数据中移除或删除。因此,在所获得的数据中的每个数据点(在这样的调整或对齐之后)将具有与由标准化查询指示的每个时间帧中的开始时间处或之后的采样率的周期的整数倍相关联的时间戳。可以类似地使用将所获得的数据中的数据点对齐的其它方式。

[0399] 其中,多个数据源702在相同的标准化查询中被指出,每个数据源702可以具有具有不同的时间戳的数据点,并且与该时间戳相关联的测量可以以不同的采样率发生。例如,第一数据源702a和第二数据源702d可由标准化查询来指示。数据源702a可以是存储具有第一采样率(例如,每秒测量一次)的数据点的第一数据源,并且数据源702d可以是存储具有第二采样率(例如,每秒测量四次)的数据点的第二数据源。如以上所讨论的,根据标准化查询中指定的参数,所获得的数据可以通过添加或移除在由标准化查询指定的采样率的周期的整数倍的时间处的数据点来对齐。

[0400] 当数据接收实体获得来自一个或多个数据源702的数据时,数据接收实体的可以

根据与标准化查询709相关联的格式参数或指示来格式化数据(框752)。格式化参数或指令可以包括在标准化查询709中,可以与标准化查询格式相关联,或者可以在查询块708的特有的属性中指定。在一些实施方案中,这可以包括通过将所得到的数据的格式调整为匹配所期望的格式来进一步从所获得的数据生成数据帧。其中,多个数据源702被查询,生成数据帧可以进一步包括将从每个数据源702所获得的数据组合以生成聚合的数据帧。例如,其中第一数据集是从第一数据源702a所查询的,第二数据集是从第二数据源702d所查询的,组合第一和第二数据集的聚合的数据帧可以被生成。

[0401] 该方法760可以开始于从数据请求实体接收标准化查询709(框762)。数据请求实体(譬如查询块708或查询执行服务706)可以通过发送标准化查询709到数据提供者704来请求来自一个或多个数据源702的数据。数据提供者704可以与一个或多个数据源702相关联,或是可以被配置为经由网络连接从各个数据源702来获取数据。可替换地,数据请求实体的数据可以接收来自数据分析系统100内的另一个数据块、软件例程、过程或服务的数据的标准化查询709。另外,数据接收实体(其同样可以是查询块708、查询执行服务706、作业进程508、数据服务520、应用程序514、程序存储器,或任何其它可以进一步分析、显示或存储与过程工厂5相关联的数据的服务、例程、过程、或装置)可以通过标准化查询709或其它方式来指示。在一些情况下,数据接收实体可以是与数据请求实体相同的实体,或数据请求实体可以与数据接收实体分开。为了清楚起见,以下讨论的示例性方法760将假设数据提供者704a作为数据请求实体和数据接收实体以从查询执行服务706接收标准化查询709,其中,标准化查询请求来自数据源702a的标准化查询709。这样做是为了更好地示出了该方法的显着特征,并且不旨在限制本公开内容的范围。熟悉本领域技术的人员能够理解的许多可替换的配置可利用对如此述方法进行微小和一般的调整来被容易地创建。

[0402] 当将数据提供者704a接收到标准化查询时,数据提供者704a从该标准化查询709提取查询参数(框764)。该查询参数可包括与时间相关联的参数(譬如,要获得的数据的时间帧、采样率等),数据类型或特征(例如,过程变量、表内的列、测量、基于该测量的计算值等),或将要从其中获得数据的数据源702a(例如,数据库的指示、指向其的路径、或其中的表)。提取查询参数可包括基于标准化查询709内的对象、阵列或元件来确定一个或多个参数。在一些实施方式中,数据提供者704还可进一步提取指示数据如何被返回给查询执行服务706的参数(即,用来提供所请求的数据到数据请求实体的格式、结构、时序或协议)。数据提供者704a可以将所提取的查询参数存储在易失性或非易失性存储器中以用来生成一个或多个特定于数据源的查询和/或格式化获得的数据以提供到数据请求实体。

[0403] 然后,数据提供者704a可以基于所提取的查询参数来生成特定于数据源的查询(框766),其中特定于数据源的查询利用与数据源702a相关联的特定于数据源的查询格式。例如,数据源702可以是使用MongoDB的数据结构或格式的非关系型数据库,在这种情况下,数据提供者704a生成利用MongoDB的查询语法的特定于数据源的查询以获得提取自标准化查询709的查询参数所指示的数据。作为可替换的示例,数据源702a可以是利用MySQL的关系型数据库,在这种情况下,数据提供者704a生成利用SQL查询的语法的特定于数据源的查询以获得由提取自标准化查询709的查询参数来指示的数据。为了生成特定于数据源的查询,数据提供者704a可以应用标准化查询格式和特定于数据源的查询格式之间的映射。该映射可以包括对语法或格式的调整,将标准化查询709中表达的请求参数转换到特定于数

据源的请求格式中的实质上等同的参数。在一些实施方案中,提取查询参数和生成特定于数据源的查询可被组合,从而数据提供者704a直接地将有标准化查询指示的查询参数映射到特定于数据源的查询的参数或元素。如上所述,可以生成特定于数据源的查询以返回具有时间戳的数据点,该时间戳对应于由标准化查询709所指示的每个时间帧中的开始时间处或之后的采样率的周期的整数倍。

[0404] 一旦特定于数据源的查询已经生成,数据提供者704a可以执行特定于数据源的查询,以从数据源702获得所请求的数据(框768)。在一些实施方案中,数据提供者704a可发送执行特定于数据源的查询的请求到数据源702a,并且数据源702a可以执行特定于数据源的查询并将结果数据返回给数据提供者704a。可替换地,数据提供者704a可作为用于数据源702a的截面或执行引擎,在这种情况下,数据提供者704a可以根据特定于数据源的查询格式的规则来执行特定于数据源的查询以访问、分析和选择存储在数据源702a中的数据。无论特定于数据源的查询是由数据提供者704a来执行还是由数据源702a来执行,都是数据提供者704a获得执行特定于数据源的查询所生成的数据。

[0405] 在一些实施方案中,数据提供者704a可基于标准化查询709来格式化通过执行查询而得到的数据(框770)。因此,数据提供者704a可处理所获得的数据以格式化所获得的数据,以对应于在标准化查询709中所表达的数据要求。这可以包括从所获得的数据生成的数据帧,包括将数据格式规则和数据别名应用到所获得的数据,数据帧包括具有容易被查询执行服务706所使用的格式的所获得的数据。例如,标准化查询709可使用JSON文件来指定数据将提供在文档中。类似地,标准化查询可指定数据将以表格格式返回到数据请求实体或数据值将被格式化为小数点后有三位的浮点数字。虽然这些格式化可以在标准化查询709中被指定,但部分或全部的格式化要求可以由标准化查询格式(或其版本)来指定。在这种方式下,利用标准化查询格式(或其版本)的标准化查询将一直以一致的格式返回数据。如上所述,格式化所获得的数据可以包括调节或对齐数据点的时间以具有对应于由标准化查询709所指示的每个时间帧内的开始时间处或之后的采样率的周期的整数倍的时间戳。这还可以包括增加具有合适的时间戳的数据点或移除在所期望的采样时间之间产生的数据点。

[0406] 一旦所获得数据被格式化,数据提供者704a将向数据接收实体提供格式化的数据(框772)。数据提供者704a可将具有所需格式或其它任何已知的格式的数据帧发送到查询执行服务706以供进一步分析。在进一步的实施方案中,当数据提供者704a向查询执行服务706发送获得自数据源特定查询的执行的未格式化的或部分格式化的数据时,获得的数据的格式中的一些或全部可以替代地通过查询执行服务706来执行。当数据由数据请求实体接收并格式化时,格式化的数据可以在过程工厂控制或分析中,如其它地方所讨论的。在一些实施方案中,查询执行服务706还可以将该数据帧提供到另一数据接收实体,譬如,查询块708或作业进程508。

[0407] 一般地,数据源702a-f、数据提供者704a-d、查询块708中的任意一部分或所有部分以及用来从可变格式的数据源702a-f获取数据的查询语言,可以与图1、2、3、4A-4Q和/或图5A-5G中的任意一部分或所有部分结合操作,和/或与任何数量的分别对应于图1、2、3、4A-4Q和/或5A-5G的在本公开内容的部分中描述的特征和/或技术结合操作。

[0408] 用于早期警报故障检测的频率分析的分析技术

[0409] 新的数据分析技术或功能(例如,可以由分布式工业过程性能监控/分析系统或DAS100提供的)是用于过程控制系统或工厂(譬如,过程工厂5)中的早期故障检测频率分析的分析技术或功能。所述的新频率分析的分析技术或功能在此被称为“滚动傅立叶变换(rolling fast Fourier transform)”或“滚动FFT”,并且可结合任何的系统、架构、方法和此处所描述的技术来使用。例如,用于滚动FFT的数据块定义可被存储在数据块定义库252,并能够供用户在离线执行的和/或在线执行的数据模块中使用。另外地或替代地,一个或多个分布式数据引擎102x可以对在一个或多个DDE102x处获得的数据执行滚动FFT。然而,滚动FFT并不限定为仅与此处描述的系统、架构、方法和/或技术结合使用,其还可以与生成时间序列数据的任何系统、结构、方法和/或技术结合使用。如将在下面所示的,滚动FFT分析技术或功能是描述性和预测性分析的组合。

[0410] 通常,在过程控制系统或工厂中,如果提供未来过程工厂行为领先指示的过程数据能够被发现(优选地在允许预防或减轻动作发生的时间帧中),异常、故障、性能的减小和/或其它不期望的或不可取的情况可以被防止(或它们的影响可以被最小化)。该过程数据可包括测量数据,例如,压力、温度、材料通过工厂的流速,以及用于装置部件的类似的信息。该过程数据还可包括,例如,过程流程流的化学组成和装置的通/断状态。一般地,可以应用滚动FFT的过程数据可以包括任何在线和离线的时序数据,在线和离线的时序数据由于控制过程工厂中的过程的而生成,并且以任何合适的方式获得,例如,通过采样,接收数据流,读自数据库、数据文件、数据源(例如,Seeq系统)或者历史库,查询等。

[0411] 在过程工厂查找领先指标的技术之一包括分析所述过程数据随着时间的行为。该行为可以因很多原因而改变,但是在某些情况下,这些改变可与工艺失常相关联,该工艺失常导致异常、故障、性能减弱,和/或其它情况,并且该改变同样地可以被认为是这些情况的领先指标。此处描述的滚动FFE技术将对应于该领先指标的第一组时间序列数据转换到频域,并且随后基于频域数据生成第二组时间序列数据,其将被监控并用来预测过程工厂中的异常、故障、性能减弱和/或情况的发生。

[0412] 通常,来自过程工厂的过程数据可被收集、接收或以其它方式获得以用作频率分析技术的输入。该数据可以是与在此过程工厂中发现的存在的过程信号相关的任意数据譬如:温度、流量、压力、组成和/或由于操作该过程工厂以控制过程而生成的其它连续信号。通常,对所获得的过程数据执行FFT变换以通过固定的窗口(例如,特定数量的数据点)来确定其中的重要频率的幅值,这通常基于2的幂(例如, $2^{10}=1024$)。现代计算方法允许数据的这个窗口具有用户定义的长度,然而,该长度经常受计算机可用存储器的量的限制。FFT中的样本数目和采样频率必须满足奈奎斯特准则(每感兴趣的最高频率至少具有两个样本)。此外,传统的FFT应该在所期望的数据信号的周期性行为的若干个循环上运行。

[0413] 虽然FFT应用到过程数据的很多常规的应用中,但是它是假设感兴趣的信号并不随时间改变。然而,滚动FFT不受这个假设的限制。实际上,“滚动”类型的FFT有利地能够捕获信号随时间的变化,以确定信号随时间的变化何时发生。具体地,滚动FFT包括在由感兴趣的信号或变量(譬如,测量、感测到的值、或作为过程工厂的结果而生成的其它信号)所产生的数据的窗口上执行FFT,并且该窗口的频率的幅值(例如,峰值)。然后,在时间上将窗口向前移动一个样本,并且再次执行FFT,结果被记录或保存。该过程持续到时间序列数据的末尾。因为对数据集的每个采样时间均执行了FFT(除了在一些情况下,第一个n-1样本,其

中 n 是窗口中样本的数量),所以,包括感兴趣的一个或多个频率的幅值(例如,峰值)的一个或多个新的时间序列数据集被创建或生成。每个感兴趣的频率可以对应于过程工厂的相应的新过程变量,该新过程变量产生对应于感兴趣的频率的幅值(例如,峰值)。由每个新过程变量所生成的时间序列数据可被存储、监控和/或分析来预测过程工厂中的可能的异常、故障或其它情况。

[0414] 因而,对于感兴趣的信号或变量,对应于预测不期望的过程工厂情况的一个或多个新的过程变量可在过程工厂中创建、生成和使用。感兴趣的信号或变量可以是单个信号或测量点,或者可以是信号的组合(例如,相加或其它组合)或单个信号/点的测量值的组合(例如,相加或其它组合)。感兴趣的变量可以由用户来定义,或者可以是自动地由的数据块、数据模块和/或数据分析功能来确定。

[0415] 如以上所讨论的,确定的一组新时间序列数据集的过程涉及将感兴趣的信号或变量从时域变换到频域,获得频域数据,并对所获得的频域数据进行变换以确定对应于新的过程变量的时间序列数据。转换或变换回到时域是重要的,因为这样就可以允许新时间序列数据集连通原始过程数据信号和/或其它时间序列过程数据一起来被查看。因此,新的时间序列数据可以通过使用相同的分析技术(例如,统计分析、PCA分析,标准差等)来进行探索,该分析技术可以用在由感兴趣的原始信号或变量来产生的数据上,和/或可以用在其它过程数据上,并且在一些情况下,该新时间序列信号可以连同原始过程数据和/或其它过程数据来一起进行分析。例如,新时间序列数据和原始/其它过程数据可以被提供为变压器数据块521的输入。

[0416] 为了说明示例的滚动FFT分析技术和它的好处,考虑一个示例场景,在该场景中,滚动FFT被应用于检测过程工厂(譬如炼油厂或化工厂)中潜在的火炬事件的问题。一般情况下,一些过程工厂具有火炬系统,该火炬系统收集来自各个过程单元的多余蒸汽并在排放之前对所收集的过量的蒸汽进行燃烧以去除有毒化学物质。对过量蒸汽或气体的这种燃烧通常被称为“燃尽”或“火炬事件”。在某些情况下,火炬气被压缩并再利用为原料或燃料气体,而不是被烧掉基因反弹。然而,当超过火炬气压缩机的容量时,火炬事件将发生从而使多余的材料通过火炬燃烧器排放。通常情况下,火炬的事件的数量和/或频率是由环境和/或其它类型的规定所管理的。虽然一些计划中的火炬事件是被允许的且是必要的,但当计划外的火炬事件发生过于频繁时,过程工厂供应商或公司可能会被罚款。因此,过程工厂供应商或公司期望能够根据当前的操作情况预测计划外的火炬事件将或可能发生,并且能够以足够多的前置时间来获得该预测,在该前置时间中,操作员可以采取减缓行动以防止计划外的火炬事件的发生。

[0417] 这是一个棘手的问题,因为现代炼油厂和化工厂是具有许多互联的单元(例如,1000甚至更多的互联单元)的复杂的系统,并且每个单元本身可被视为一个大的过程工厂。通常情况下,这些单元被连接到公司的火炬系统。由于这些单元中的任意一个单元均可以是蒸汽的潜在来源(其必须要由火炬系统处理),因此,可能很难监控哪个(些)单元接近火炬条件。此外,一旦火炬确实发生,不能立即明确哪个单元应负责。

[0418] 此处描述的滚动FFT技术可以用来处理这种情况。图7A显示由炼油厂生成的示例过程数据,其中,滚动FFT应用于该示例过程数据。该示例过程数据包括在特定时间间隔期间从炼油厂的火炬系统的单元所获得的过程流程测量或信号1100。可以看到的,过程流程

信号1100在本质上是周期性信号,其以大约一天为周期(例如,因为该周期性可对应于每日加热和冷却循环)。此外,示例过程数据包括指示在相同的特定时间间隔上火炬系统的火炬压力的另一信号1102。火炬压力是在火炬系统中收集的蒸汽的测量,并且该测量可以被获得,例如,从压缩机或包含过量的气体和蒸汽的容器的传感器处获得。在图7A可以注意到,信号数据1100和1102在时间上被对齐并且缩放,从而它们随着时间的推移的行为和关系很容易显现。在一个实施方式中,一个或多个数据块、数据模块和/或一个或多个DDE102x接收信号数据1100和1102,并且执行这两个信号1100和1102的时间对其,从而信号1100和1102能够以图7A中所示的时间对齐的方式来显示。

[0419] 图7B示出了对应不同时间间隔的相同的信号1100和1102,在不同的时间间隔器件,信号1102中的火炬事件发生(例如,对应于信号1102的火炬压力增大并且超过压力限制,从而导致火炬事件1104)。参阅图7B,显而易见的是,过程流程信号1100的行为在火炬事件1104之前改变。例如,该过程流程信号1100的周期行为在火炬事件1104发生的二至三天前发生改变。作为火炬事件1104的领先指标,过程流程信号1100中该变化的标识可以例如通过使用数据分析技术(譬如,PCA、互相关、PLS回归等)来确定。该变化/领先指标可以被捕获、标识或定义;对应于所标识的变化/领先指标的新信号或过程变量可以被产生、定义或创建;以及新过程变量的时间序列输出可以通过利用滚动FFT来确定。因此,针对领先指标的存在,通过监控新过程变量的时间序列数据(该时间序列数据是通过滚动FFT来确定的),将确定即将发生火炬,并且相应的警告可被呈现给用户(例如,工程师、过程操作员等),从而可以采取预防性的和/或减轻的动作。在一个实施方式中,新定义的过程变量可以结合在炼油厂的过程控制系统中(例如,通过分配一个标签和/或以其它使得过程控制数据库能标识新的过程变量的方法),并且可以使用传统过程控制监控设备和技术来被监控。因此,对应于该新的过程控制变量的状态和其它信息可以利用限制、告警、趋势图等来呈现给用户,仅呈现为过程信息的任何其它项。

[0420] 在一个实施方案中,被用来预测火炬事件的过程流信号是个别流程或测量的组合(例如,和、加权平均或其它适当的组合)。图7C示出了一实施方案,其中滚动FFT分析所施加的过程信号对应于独立的过程流程信号1106的和的变化。在一示例中,该和信号1106可以通过合计单个流程信号的测量来创建。生成和1106或其它希望的组合减少了所需创建和分析的新信号的数目,这转而又可以简化整体分析。然而,对使用单个处理信号或一些信号的组合并没有限制。图7C还同时显示了火炬压力信号1102,该火炬压力信号1102包括在所显示的时间间隔期间的火炬事件1108和1110。注意的是,信号1106和1102是按时间对齐的,所以相互关系很容易显现。过程流程1106的和行为中的变化能够在火炬事件1108之前被看到,但对于火炬事件1110,汇总的过程流程1106中的变化在火炬事件1110之后发生。流程1102和1106之间的关系的另外的分析被需要,以通过执行另外的分析功能等来确定汇总的过程流程1106的是否具有行为(如果是,则如何行为)。

[0421] 图7D示出了应用滚动FFT技术到信号1106的结果1112,其将过程流程信号1106的和从原始的时域转换到频域,然后再回到时域。具体地,图7D中的信号1112对应于信号1106感兴趣的特定频率(例如,对应于每天四个周期的第四频率)的幅值(例如,时间序列数据)。该对应于每天四个周期的第四频率已被定义为新的过程变量,该过程变量的时间序列峰值已经从对应的频域数据捕获,并在时域中显示为信号1112以及火炬压力信号1102。注意到,

信号1112和1102是时间对齐的,从而相互关系很容易显现。如图7D可知,信号1112的特定的峰值1113与火炬事件1108相关联。数据1112看起来嘈杂,但是包含了可以被假阳性的事件(例如,事件1115a、1115b)。然而,假阳性并不过分有关,因为它们可能用于咨询能力(advisory capacity),和/或可以表示被避免的“接近”的火炬事件。

[0422] 为进一步处理并清洁信号数据1112,附加的数据分析技术可被执行。例如,在图7E中,PCA已应用到对应于每天四个周期的第四频率的频域数据来确定其第一主成分,以及所述第一主成分的时间序列值被示为信号1118。特别地,图7E示出了围绕火炬压力信号1102的火炬事件1108的时间间隔中的第一主成分1118的行为的放大图。注意到,信号1118和1102是时间对齐的,从而它们之间的关系很容易地显现。可以在图7E中看到,第一主成分1118的值的尖峰发生在火炬事件1108很久之前,并且随后的第一主成分1118的值显著地减少。为捕捉到这一尖峰,衰减滤波器可被用来将信号1118保持一段时间。在一个实施方案中,衰减滤波器可以由数据块定义来定义,并且对应的允许该衰减滤波器的使用的数据块属性将被针对每个应用来进行不同地调整。例如,衰减滤波器块的可配置的属性可定义信号增强的速率和/或信号衰减的速率。在一些实施方案中,额外的主成分可以被确定性并且用于增加技术的灵敏度。

[0423] 图7F示出了图7D的另一部分的放大视图。具体而言,图7F更详细地示出了围绕火炬压力信号1102的火炬事件1110的时间中的第一主成分1118的行为。在图7F中,对应于火炬事件1110的第一主成分1118的值比火炬事件1108的第一主成分1118的值要小,但是,相较于基准线,在值上仍然存在很大的差异。

[0424] 使用滚动FFT分析以检测过程控制数据中的领先指标的技术可以被用在离线分析和数据模型构建的工作上。然而,一旦新的过程变量(例如,频率、主成分,和/或感兴趣的其它更高阶的数据)被标识和定义,可以对来自在线过程工厂的流式传输的数据执行滚动FFT分析。例如,滚动FFT可以被定义为一个数据块并被存储在块定义库252中,从而滚动FFT数据块的实例可引入到在线数据模块中并且绑定到过程工厂中的在线源。因此,在其上可操作滚动FFT分析的实时流式传输的数据可以提供即将发生的情况的实时预测/警告给用户或工厂操作员。在一些实施方案中,用户可以定义一个组合数据块以包括与其它分析功能块(例如,PCA,PLS和/或其它分析功能块)互联的滚动FFT数据块。此外,用户可以定义将多个过程信号组合(例如,求和、计算加权平均等)为单个信号以用作滚动FFT数据块的输入和/或包括滚动FFT数据块的组合块的输入的数据块。

[0425] 因此,滚动FFT数据分析技术是一种示例性的描述性分析技术,其也可用于离线过程数据和用于在线过程数据。

[0426] 在一些实施方案中,对应于该滚动FFT分析技术的新的过程变量可以被引入该过程工厂。例如,新的过程变量可以被定义、标识(例如,通过分配相应的过程控制标签),并被存储在过程工厂的过程控制数据库中。在一个实施方式中,由新的过程变量生成的时间序列数据可作为控制功能或控制块的输入,该控制功能或控制块操作以控制过程工厂中的过程的一部分,或作为过程工厂中的变化的触发器。

[0427] 此外,除了检测潜在的火炬事件,其它状况(滚动FFT可以被应用在其中以防止非期望的情况或事件)包括潜在泄压阀升降的警告、潜在压缩机喘振的警告、即将发生的泵故障的警告、即将发送的阀故障的警告、流动不稳定的警告,该流动不稳定譬如在蒸馏塔中的

涌入、在旋转粉碎磨机中的故障、石油和天然气井的生产不稳定的故障等。在某些情况下，滚动FFT可应用于防止所述过程工厂内的或甚至整个过程工厂内的一个或多个实体的非希望的性能的测量。

[0428] 图7G示出了用于提供在过程工厂和过程控制系统中的早期故障检测的示例的示例方法1200的流程图。方法1200的一个或多个部分可以例如由块定义库252中的数据块来执行。方法1200的一个或多个部分可以由工业过程性能监控/分析系统100的一个或多个部分来执行，例如，由一个或多个DDE102x来执行，由离线数据模块执行，由在线数据模块执行等等。方法1200的一个或多个部分可以由系统、设备和除了此处所描述的数据分析系统100的装置来执行。在一些实施方案中，该方法1200可以包括更多、更少或与不同于本文描述的步骤。

[0429] 在框1202处，方法1200可包括接收或获得过程信号或作为过程工厂控制过程的结果而产生的数据的初始集。过程信号的初始集可以包括通过操作以控制过程工厂中的控制过程的一个或多个过程控制设备所生成的时域数据，和/或可以包括作为控制该过程的结果而由过程工厂中的一个或多个其它部件、设备或实体来生成的时域数据（例如，过程工厂中的各种数据源生成的一阶过程数据）。例如，过程信号的初始集可以包括随时间推移的过程测量的值，譬如，温度值、流量值、压力值、组分值、状态值等。在一些实施方案中，所获得的过程信号或数据的初始集可以包括作为过程工厂控制过程的结果而产生的随时间推移的第二或更高阶的过程数据，譬如：诊断结果、一系列操作员或用户的动作、链接或过程通信网络的可用带宽、数据分析的结果等。例如，所获得的过程信号的初始集可以包括离线过程的数据和/或在线过程数据。

[0430] 在框1204处，方法1200可以包括基于所获得的过程信号的初始集，确定在过程工厂中正在发生（或已经发生）的异常、故障、性能降低或其它不期望的/不可取的情况的领先指标的内的处理设备。例如，领先指标可以是一个或多个过程控制信号的行为中的变化，该一个或多个过程控制信号的行为的变化在异常、故障、性能降低和/或其它情况（譬如，特定过程控制信号的特定频率的峰值（例如，图7B所示的尖峰1104）。在一个实施方案中，领先指标可以通过在一个或多个所获得的过程信号初始集上使用一个或多个统计性分析技术来确定，譬如，PCA、PLS回归、聚类或互相关等。在实施方案中，数据分析系统100的一个或多个离线和/或在线数据块和/或数据模块可以在一个或多个所获得的过程信号上操作以确定异常、故障、性能降低和/或其它情况的一个或多个领先指标。

[0431] 在框1206处，方法1200可以包括创建、定义或者生成对应于领先指标的一组一个或多个新的过程变量。在一个实施方式中，对应于在特定信号的行为中的特定变化的新的过程变量可以被创建、定义或生成。例如，可能产生有领先指标的感兴趣的频率（例如，图7C中的信号1106的频率）可以被标识、和创建/生成为新的进程变量。在一些实施方案中，生成一个或多个新过程变量集（框1206）包括定义/标识/生成对应于领先指标（例如，图7C中的信号1106的改变）的第一新过程变量，以及随后对由第一新过程变量（独自或与由其它过程变量所生成的其它时间序列数据的组合）生成的时间序列执行一个或多个分析技术来确定与该领先指标相关联的另一个新过程变量。例如，参考图7D，PCA被应用到对应于图7C中的信号1106的频域数据以确定第一主成分，并且第一主成分被标识/定义为另一个新过程变量，其值随着时间的推移由信号1112在图7C中表示。

[0432] 在框1206的一些实现中,数据分析系统100的一个或多个离线和/或在线的数据块和/或的数据模块可以操作来标识和创建/定义/生成对应于领先指标的一个或多个新的过程变量。在某些情况下,一个或多个新的过程变量可以过程工厂或过程控制系统中被标记和/或存储。

[0433] 在框1208,方法1200可以包括获取一后续组的过程信号(例如,以类似于框1202的方式),并在其上执行滚动FFT以确定对应于一个或多个新过程变量的时间序列数据。例如,该后续组的过程信号可以包括离线和/或在线信号。为了对其执行滚动FFT,可以对该后续组的过程信号的数据的第一窗口执行FFT,并且其中的频率的峰振幅的值可以被存为时间序列数据,可以对数据的第二窗口执行FFT,并且其中的频率的峰振幅的值可以被存储为存储为时间序列数据等等,从而生成包含在后续组过程信号中的感兴趣的一个或多个频率的相应的时间序列数据。例如,当特定过程信号的感兴趣的特定频率对应于新的过程变量,滚动FFT可以被应用到随后的过程信号集,以获得随着时间推移的随后的过程信号集中的感兴趣的特定频率的行为,例如,获得包括由新过程变量随时间生成的峰值的时间序列数据。在分析技术被执行以标识新过程变量的情况下(例如,新的过程变量对应于与如以上讨论的信号1106相对应的频域数据的第一主成分),一个或多个分析技术可以在框1208处执行以确定该新过程变量的时间序列数据。框1208可以由数据分析系统100的一个或多个部分来执行,譬如,在实施例中,由一个或多个在线和/或离线的数据块和/或数据模块来执行。

[0434] 在框1210处,方法1200可以包括针对领先指标的存在,监控由一个或多个新的过程变量生成的时间序列数据(例如,如在框1208处所得到的)。例如,当感兴趣的特定频率的峰值超出幅值和/或持续时间阈值时,可以检测到领先指标的存在。在一个实施方式中,新过程变量通过相应的标签或其它指示器来标识,并且并入该过程工厂,并且因此,其时间序列数据值的监控(框1215)可以通过故障检测、报警处理和/或过程工厂的其它监控应用来执行。另外地或可替代地,可通过数据分析系统100来执行时间序列数据的监控(框1215)以检测领先指标的存在。

[0435] 在框1212处,方法1200可以包括在生成指示,故障、异常、事件、性能减弱,非期望情况,和/或期望的情况等被基于在框1210处的监控期间所发现的检测到的领先指标的存在而被预测。在一些情况下,框1212包括生产时间帧的指示,在该时间帧中,所述的情况被预测为发生。基于所获得的时间序列数据中的领先指标的存在,例如,趋势图、报警、警报和/或其它视觉或听觉指示器可以被生成并由用户界面呈现,以警告用户或操作员即将发生的异常、故障、性能减弱,和/或过程工厂中的其它情况。

[0436] 在一些实施方案中(未示于图7G),该方法1200可以包括基于检测到的领先指标的存在,使信号(譬如控制信号或其它指示变化的信号)被生成并提供给在线过程工厂。例如,数据分析系统100可以基于监控数据中的检测到的领先指标的存在来自动地生成一个或多个控制信号(框1210),并可以自动地将该一个或多个控制信号提供到一个或多个控制器以改变的过程控制工厂的至少一部分的行为。其它可被提供到在线过程工厂以影响其中的变化的信号包括,例如,指示(例如,设备、装置、例程或应用等的)参数、值、配置和/或状态的改变的触发器或信号,或指示过程工厂中执行或连同过程工厂执行的应用的改变的触发器或信号。

[0437] 在一些实施方案中(未示于图7G),方法1200可以包括使得一个或多个信号被生成

并提供给数据分析系统100。例如,方法1200可以包括向数据分析系统100提供新过程变量和它们相应的标识符的指示、所确定的领先指标的指示、执行在时间序列数据(和从中生成的更高阶数据,譬如,用在其上的各种分析技术的输出)以确定另外的新过程变量、由该新过程变量生成的监控的时间序列数据、其中的领先指标的存在。通常,方法1200可以包括使得通过执行方法1200而生成的任何数据被提供给数据分析系统100。在一个实施方式中,提供给数据分析系统100的一个或多个信号可以是流式传输的数据。

[0438] 当然,方法1200不限于框1202至1212。在一些实施方案中,另外的块可以通过方法1200执行,和/或框1202-1212中的一些框可以从方法1200移除。另外,方法1200的实施例可以与图1、2、3、图4A-4Q、5A-5G和6A-6E中的任意或所有的部分一起操作,并且/或利用本公开内容其它部分的中描述的任意数目的特征和/或技术。

[0439] 鉴于上述情况,熟悉的过程控制系统在时间上的演变(例如,从控制和其它处理被或多或少地集中的处理的集中式过程控制系统向下一代分布式控制系统(DCS)的演变,在DCS中,控制和其它处理被分布在遍布工厂的多个控制器中)的人将认识到新分布式分析系统(Distributed Analytics System,DAS)100和此处描述的相关联的工业过程性能监控和分析方法、设备、装置、部件和技术提供了到下一代的过程控制性能监控和分析的类似的飞跃。例如,不同于在过程控制系统的操作员工作站处执行、在分析设施离线地执行、或通过使用基本的分析工具来实行执行定向的、狭隘聚焦的分析,分布式分析系统100和此处描述的相关联的方法、设备、装置、元件和技术允许全面的性能监控和实时分析,该性能监控和实时分析广泛分布在工厂中甚至以公司和/或企业的级别(类似于分布式控制系统(参见例如图2)的性能监控和实时分析的方式),从而与该过程工厂相关的实时描述、预测和命令是一直可用的。另外,该模块,DAS100的分布式架构使得性能和分析功能如所期望的被与过程工厂和/或过程控制系统的数据源接近或远离地嵌入,并且允许DDE102在如期望地或需要时被添加到DAS100(例如,当工厂5扩张时,为了更多的本地化处理能力,在工厂5的特定区域或部位提供分析等)。另外,如以上所述的,DAS100允许数据模块的设计、定义、部署和监控被根据它们所执行的环境和平台来划分或分离,从而将数据工程师或用户从必须关系分析和数据模块的编码和编译中彻底释放出来。此外,这种分离允许具有更大的灵活性和无缝迁移到不同的目标环境。

[0440] 当在软件中实现时,此处描述的任意的应用、服务和引擎可以被存储在任何有形的、非临时性计算机可读存储器上,譬如存储于磁盘、光盘、固态存储器设备、分子存储储存设备或其它储存介质,存储在计算机或处理器的RAM或ROM中,等等。虽然此描述的示例性系统被公开为在其它部件中包括执行在硬件中的软件和/或固件,应该注意的该系统仅仅是说明性的,并且不应当被认为是限制性的。例如,可以设想这些硬件、软件和固件部件中的一个或多个可以独占在硬件中、独占在软件中,或以在硬件和软件的任意组合中。因此,虽然此描述的示例性系统被描述为实现在一个或多个计算机设备的处理器上执行的软件,本领域的普通技术人员能够容易理解该些提供的示例并不是实现该系统的唯一途径。

[0441] 因此,虽然本发明已经参考特定实施方案进行了描述,其旨在是说明性的而不是限制本发明。对于本领域的普通技术人员,对本申请所公开的实施例进行改变、增加或删除并不会脱离本发明的精神和范围。

[0442] 任何特定实施方式的特定的特征、结构和/或特性可以以任何合适的方式组合,

和/或一个和/或多个其它实施方式来组合,包括利用或不利用其它特征的使用来使用所选择的特征。此外,可做出许多修改以使特定的应用、情形和/或材料适应于本发明的实质范围和精神。应当理解的是,根据此处的教导,此处描述和/或示出的本发明的实施例的其它变型和/或修改是可能的,并应被视为本发明的精神或范围的一部分。此处描述的本发明的特定方面是示例性的方面。

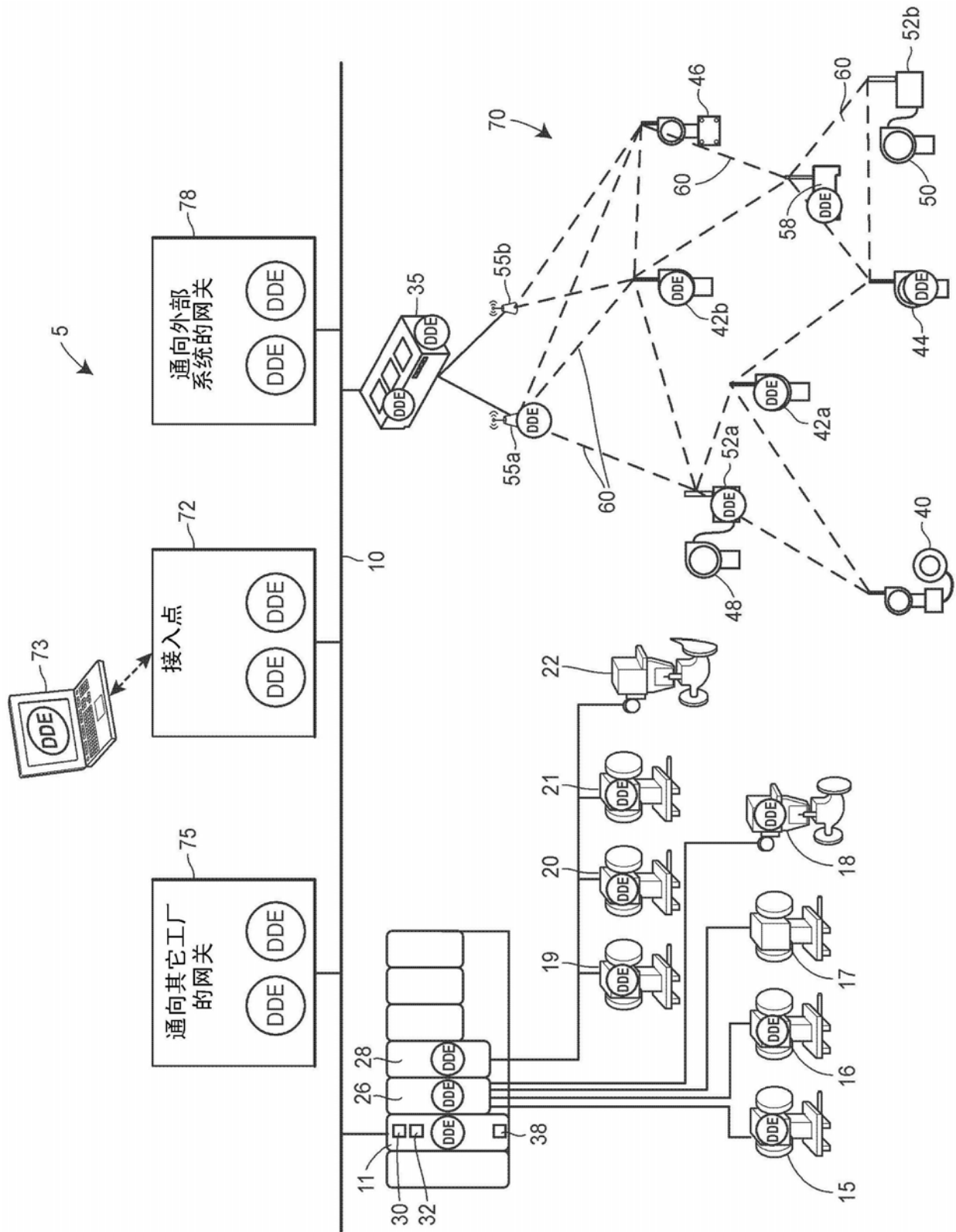


图1

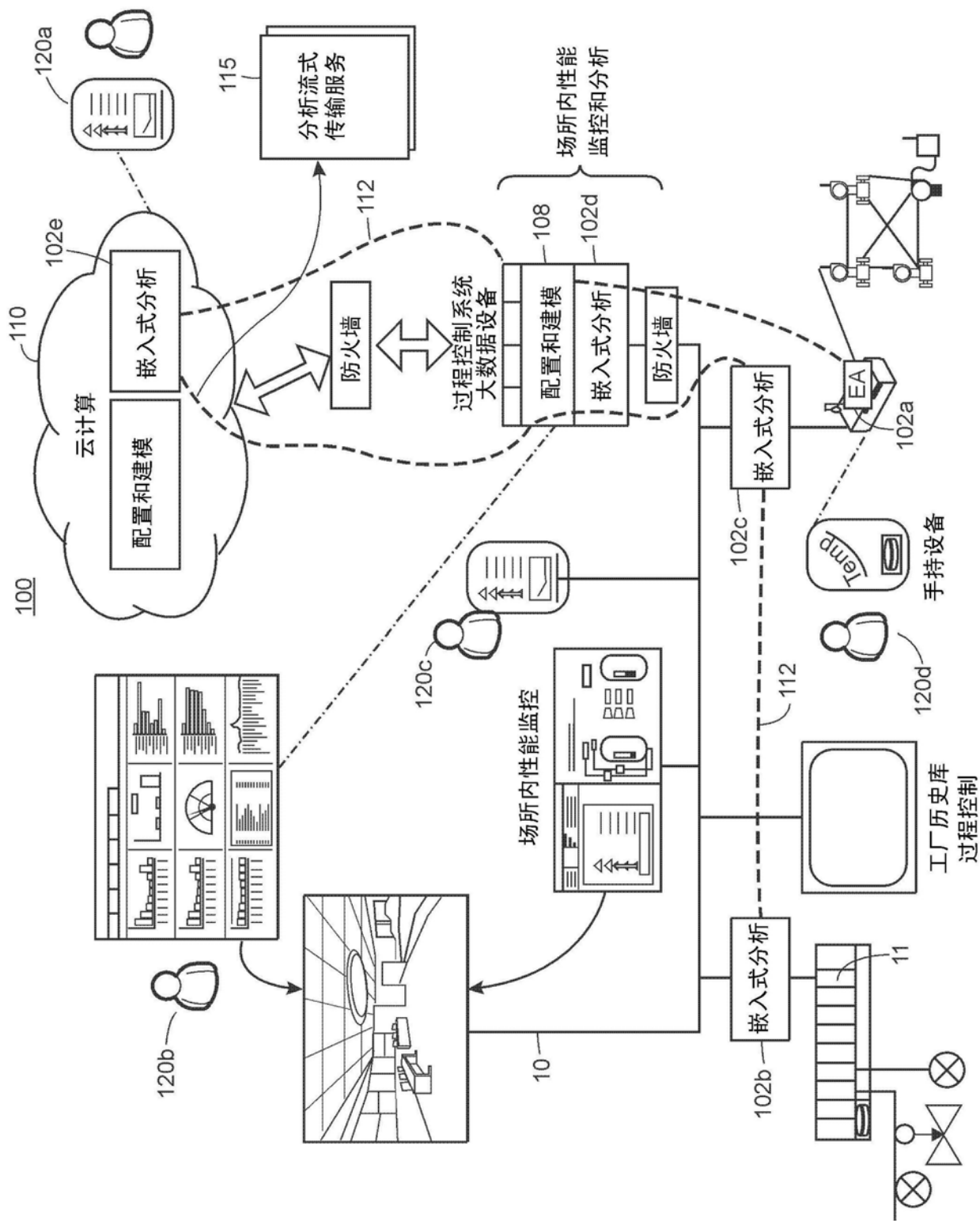


图2A

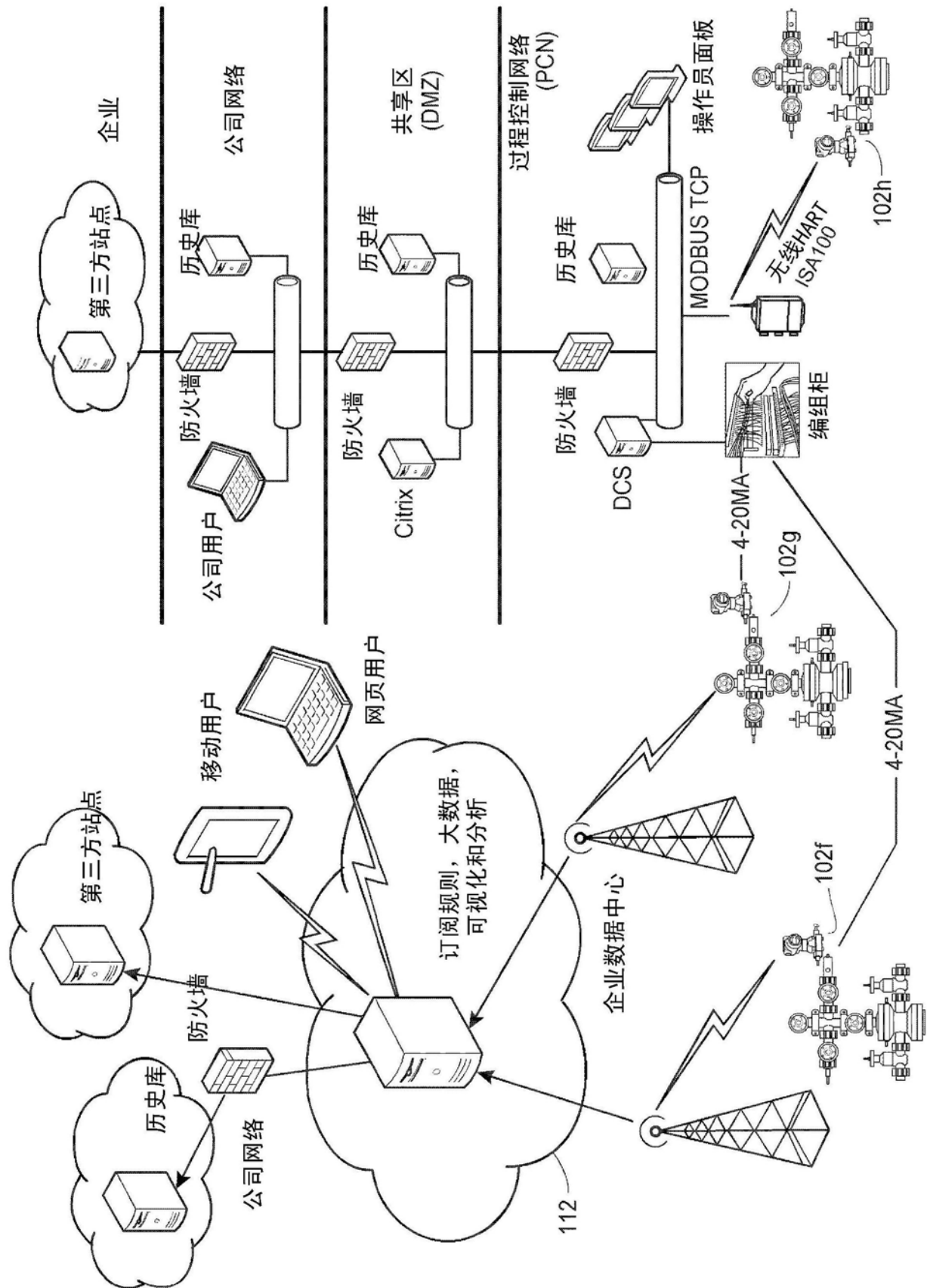


图2B

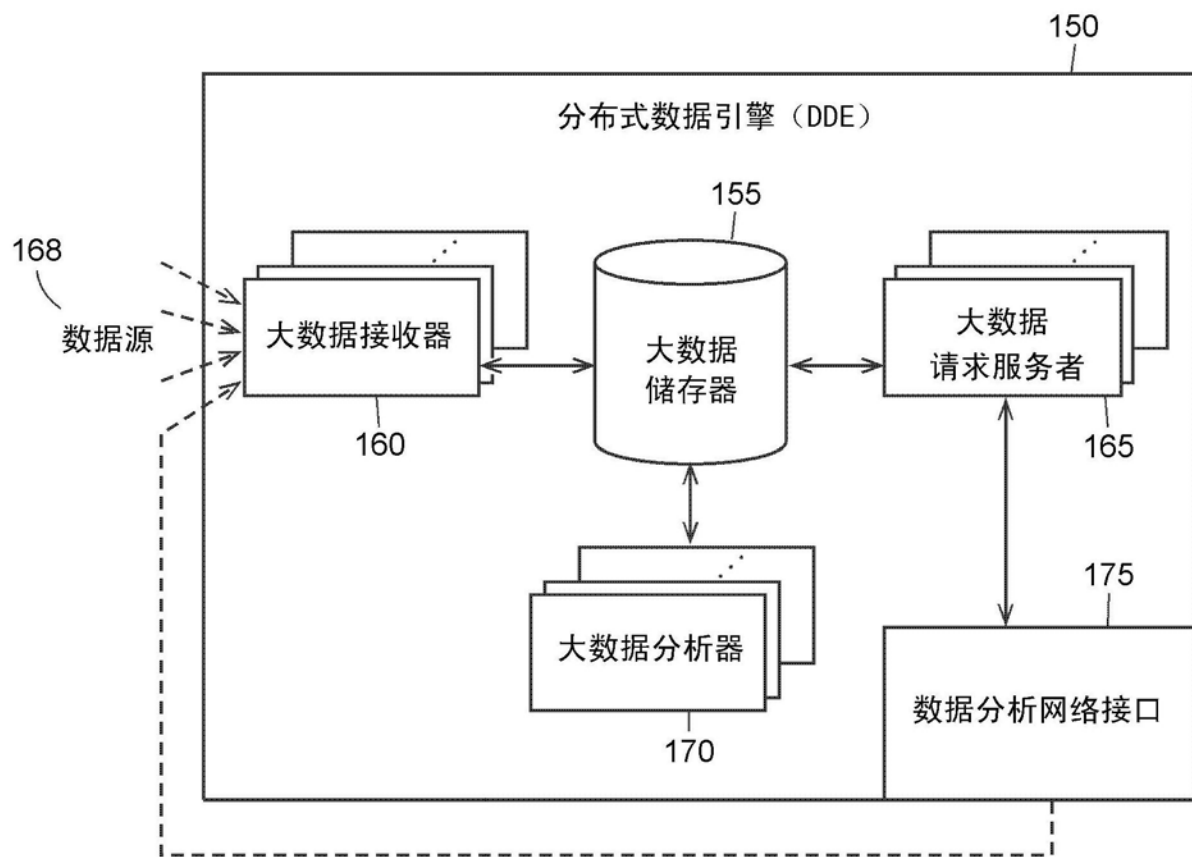


图3

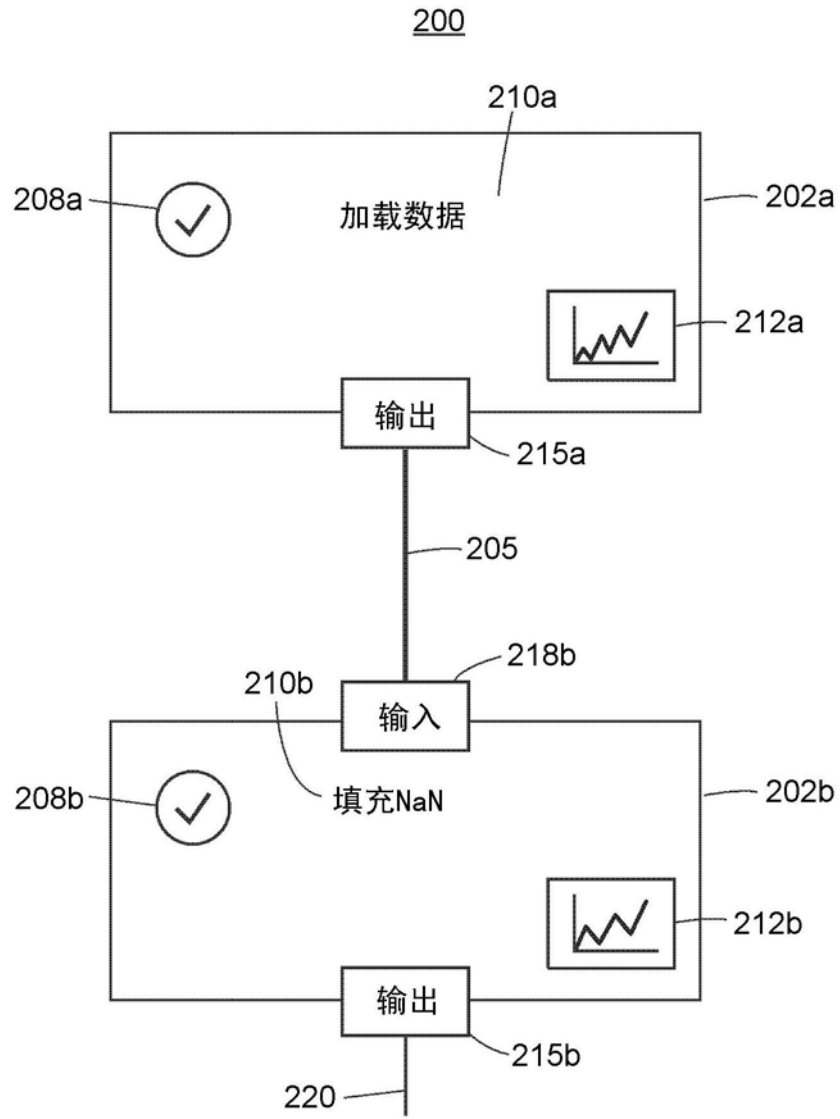


图4A

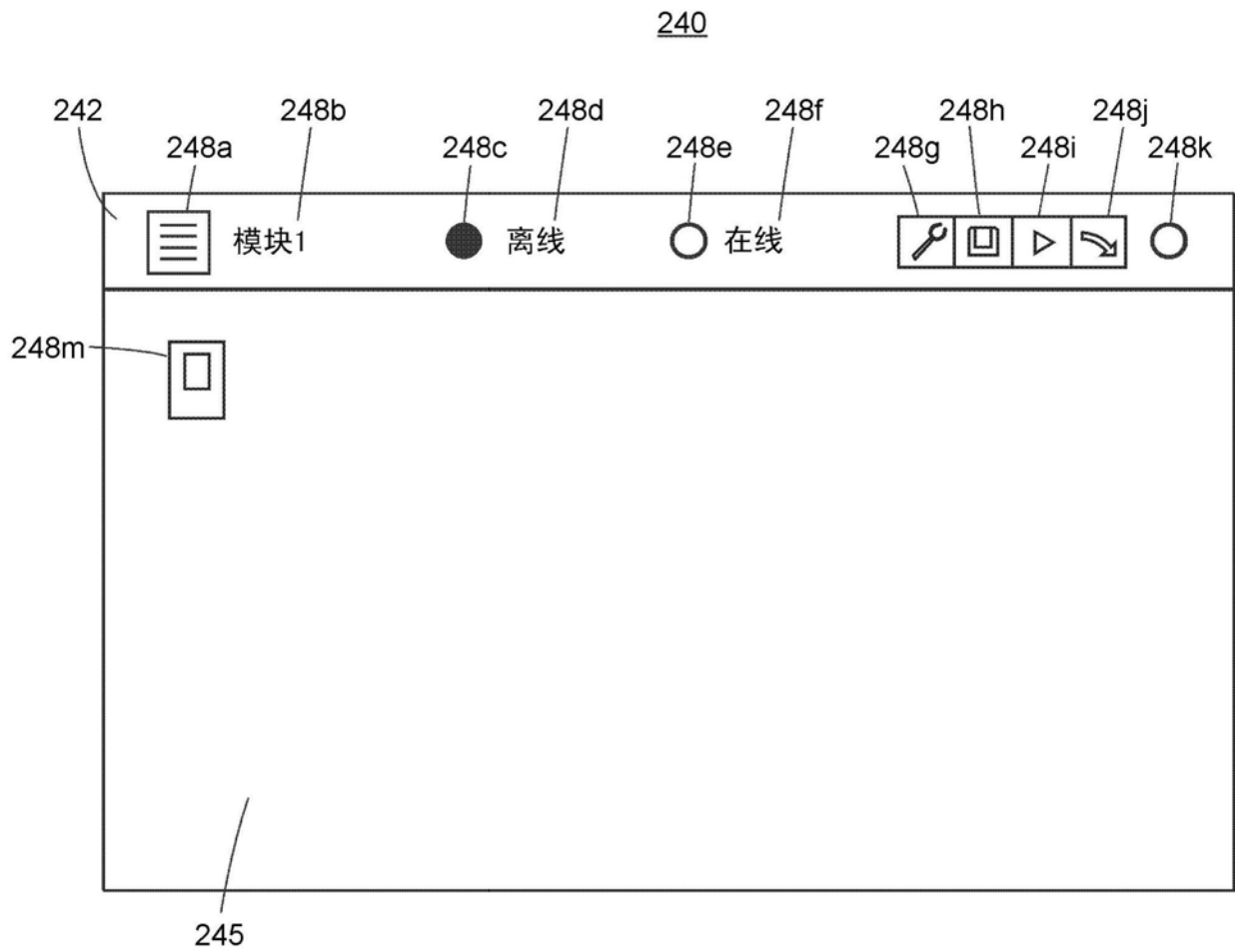


图4B

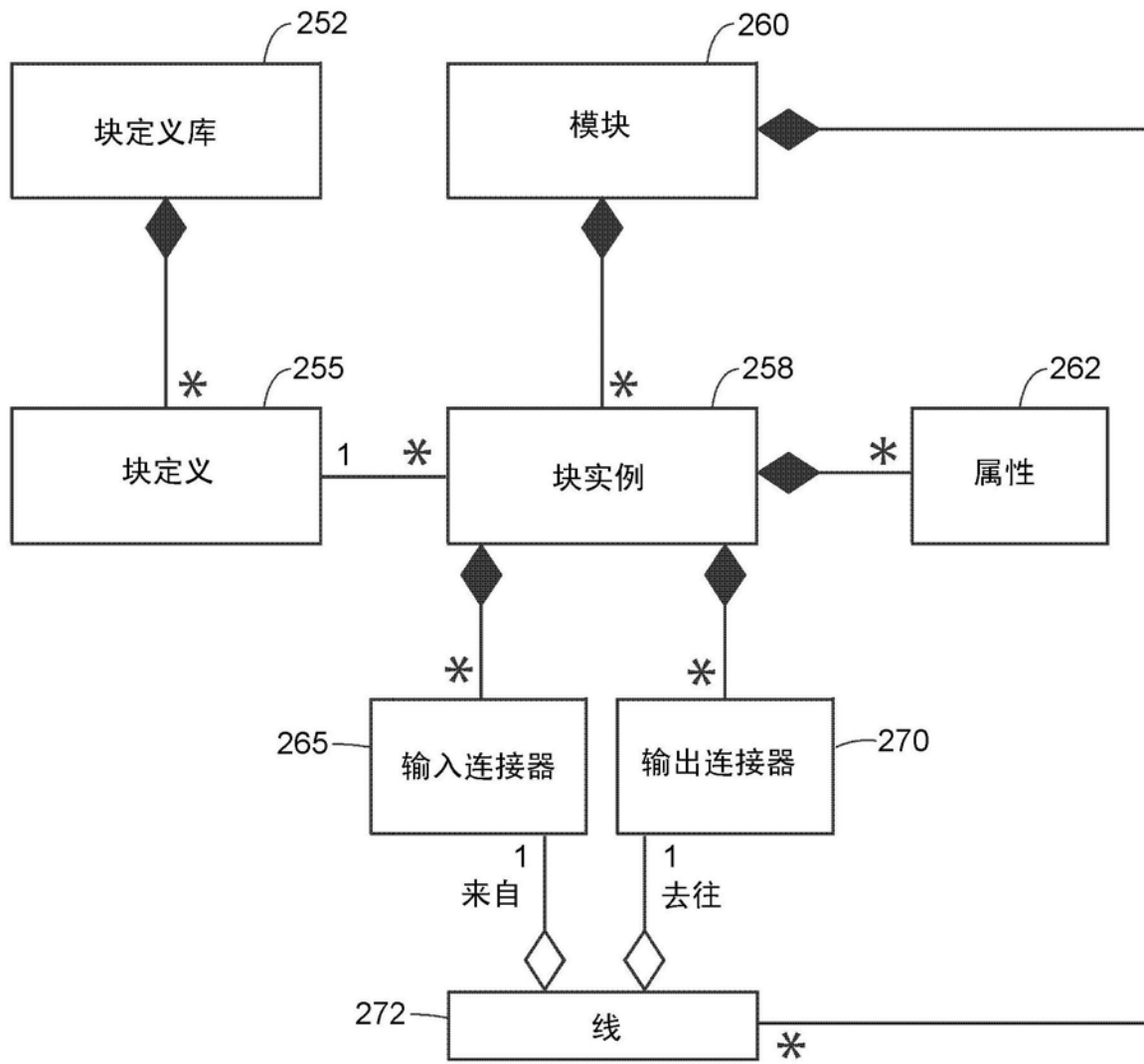
250

图4C

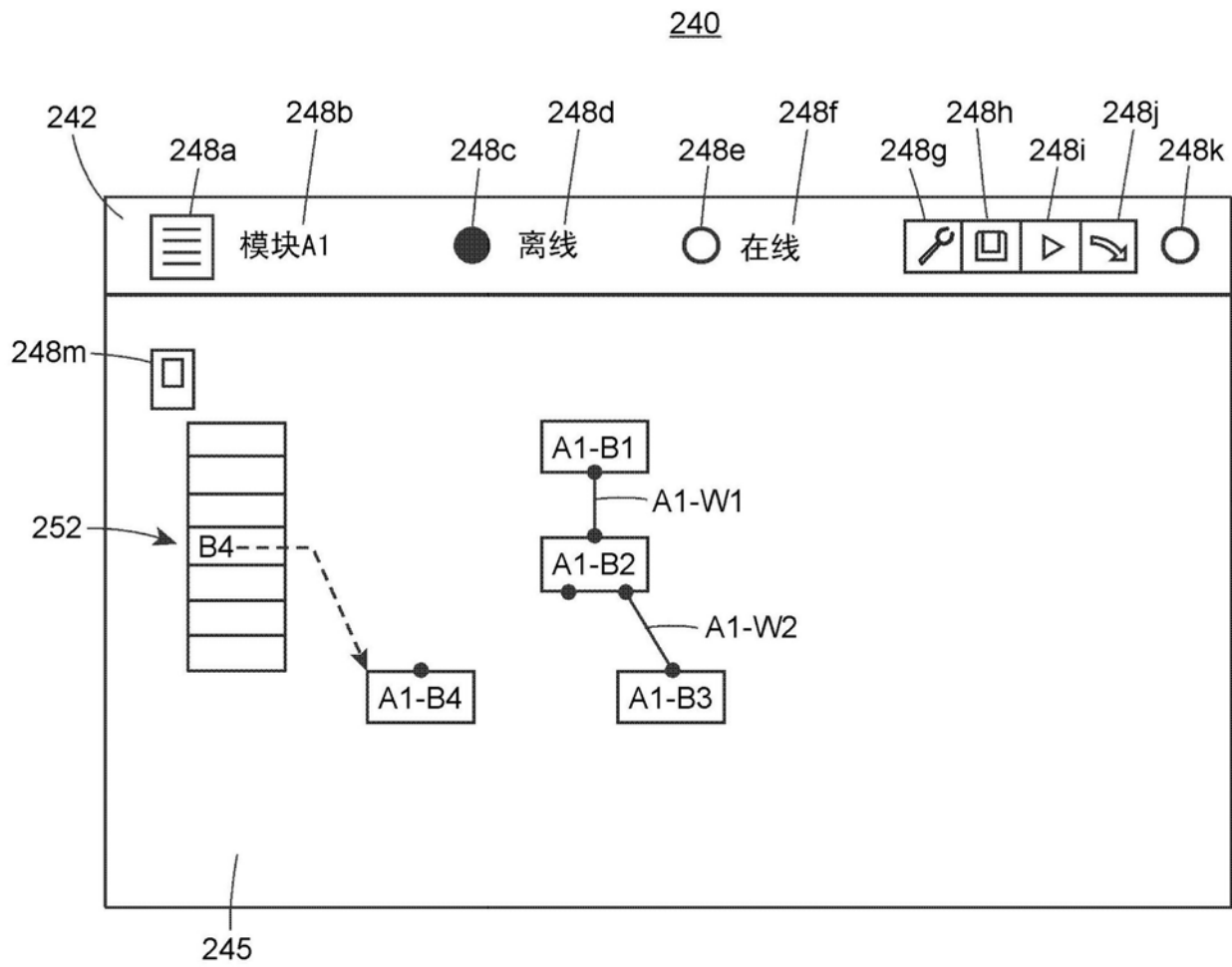


图4D

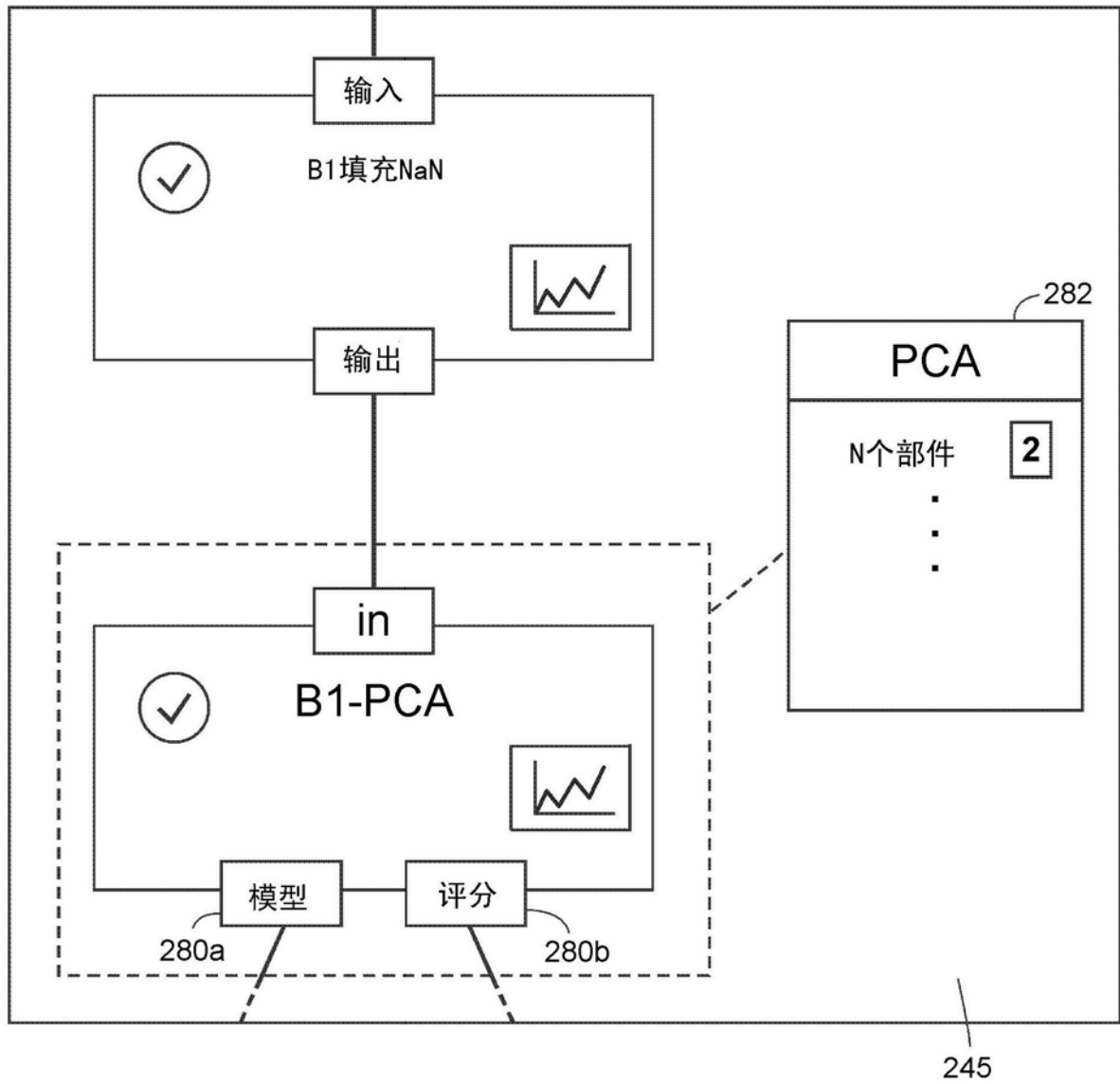


图4E

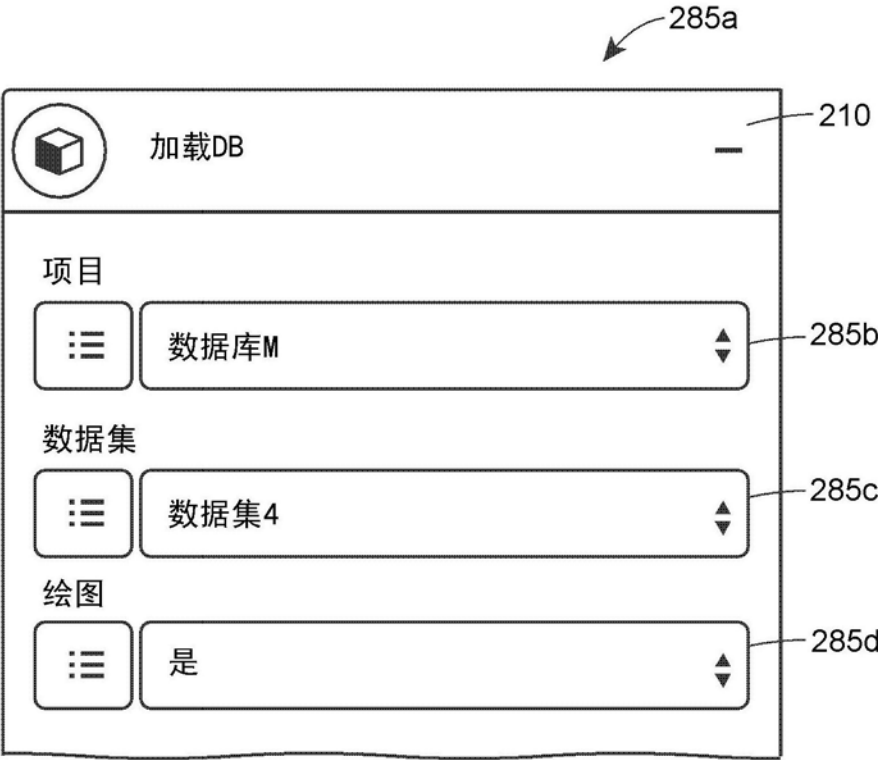


图4F

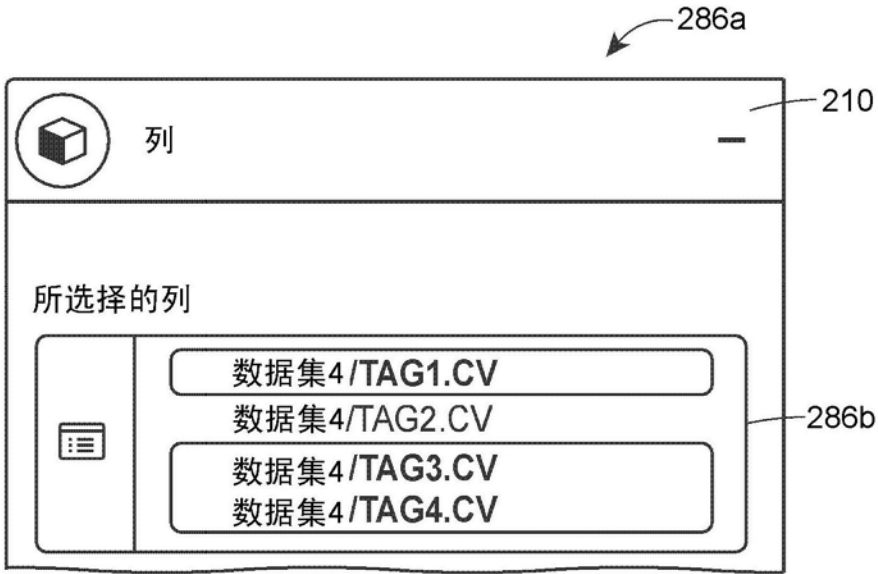


图4G

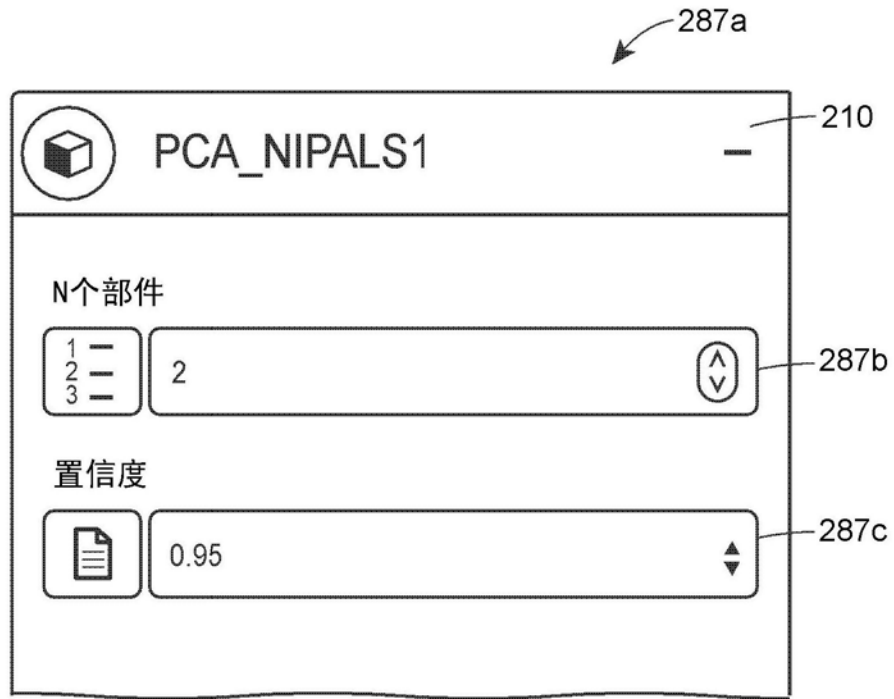


图4H

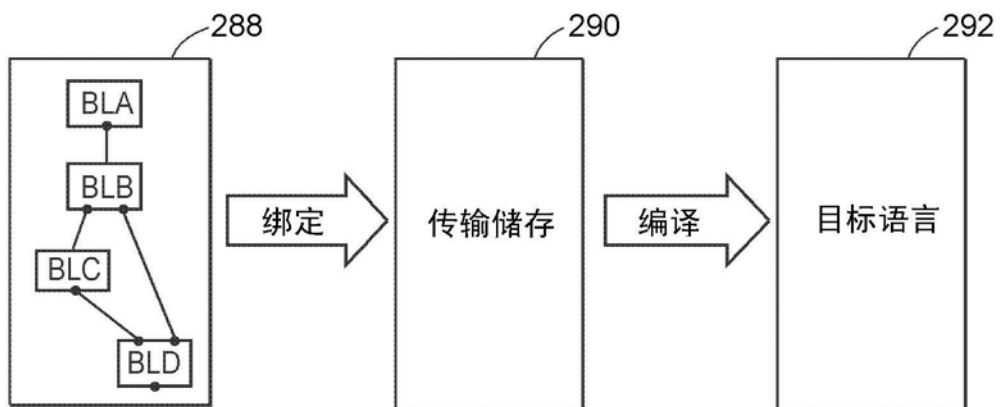


图4I

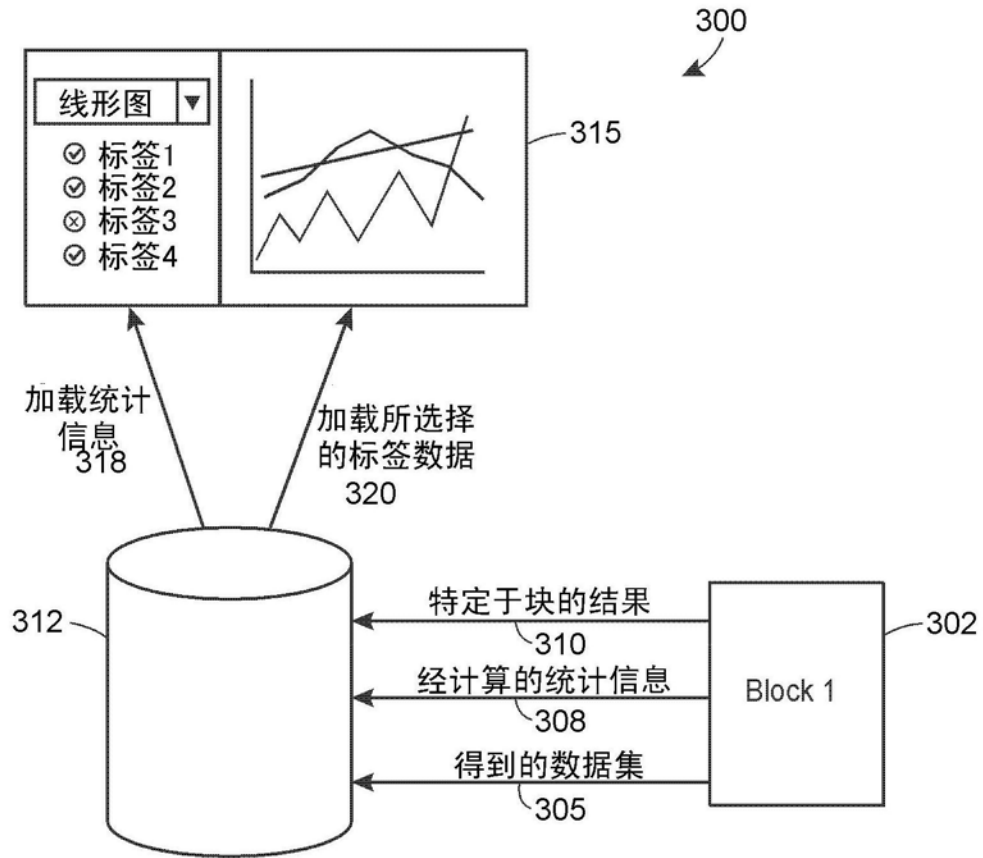


图4J

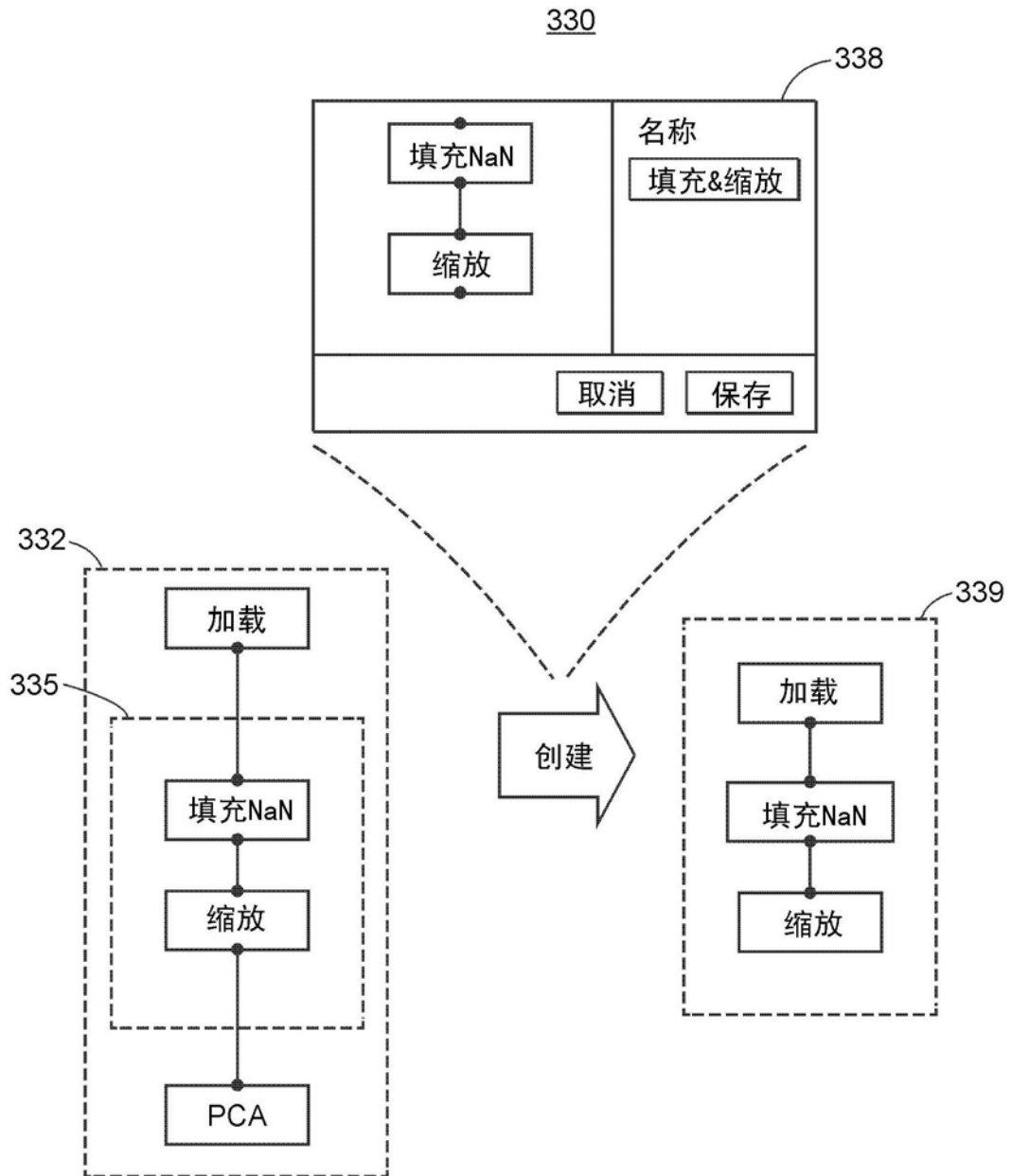


图4K

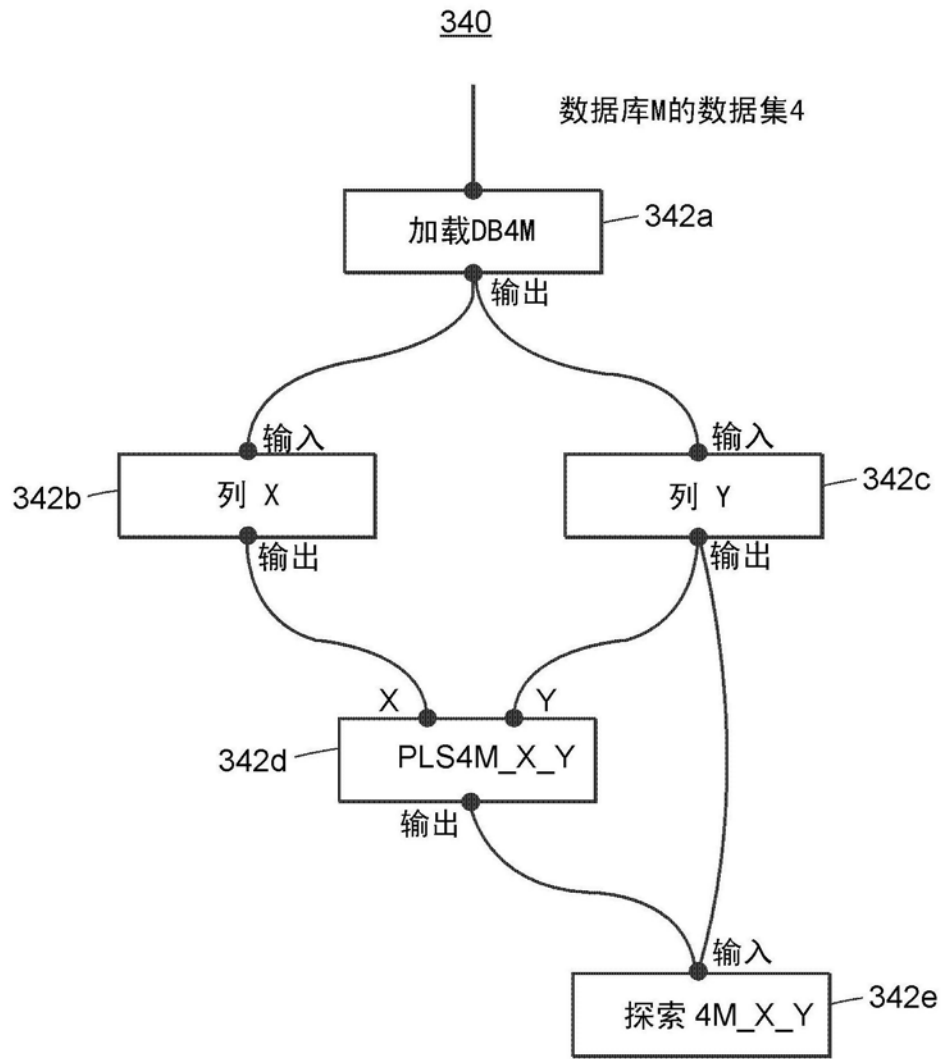


图4L

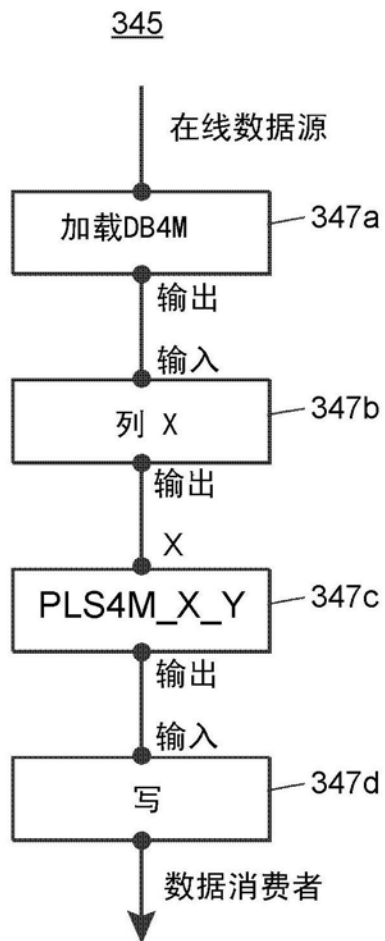


图4M

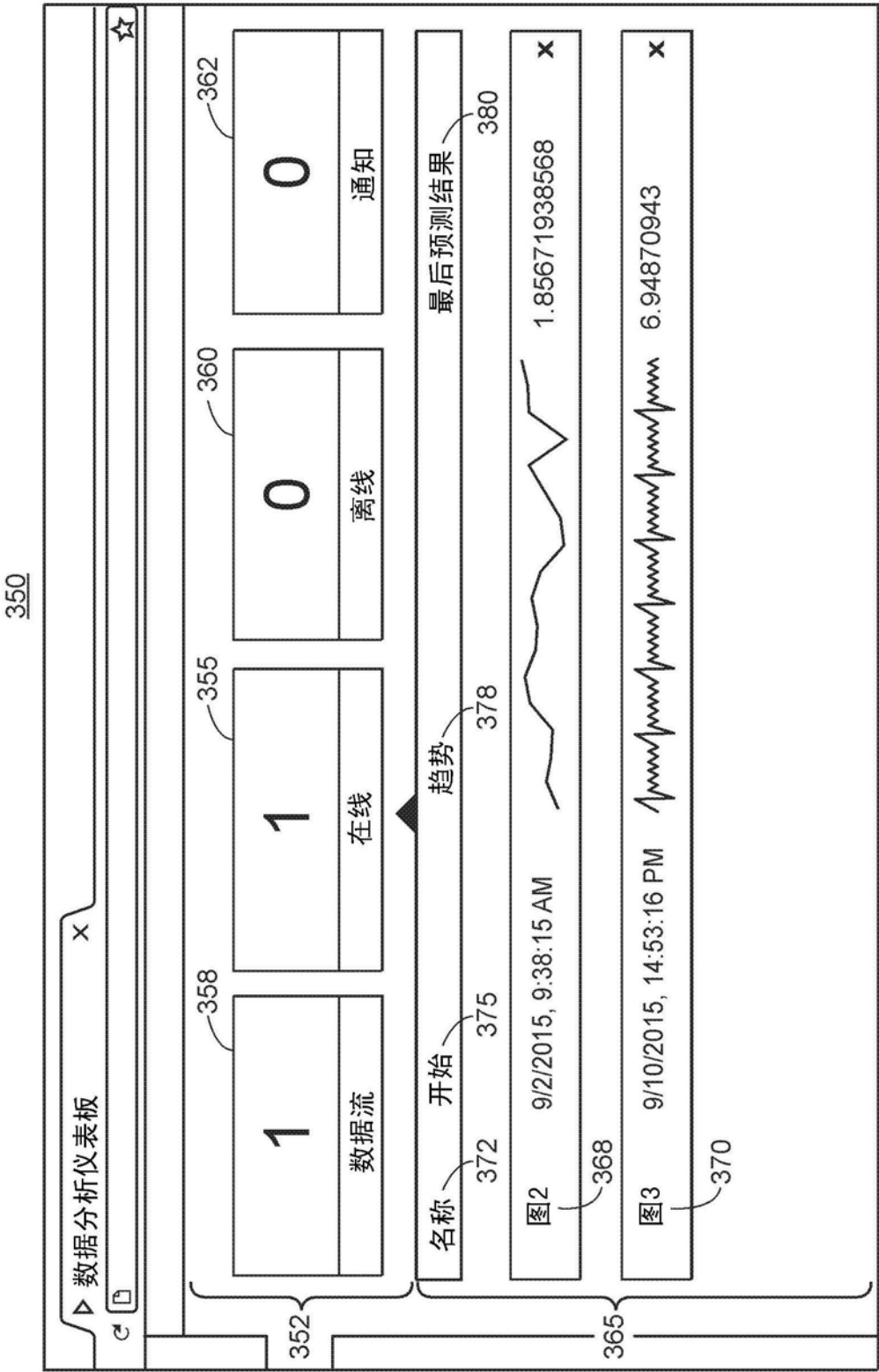


图4N-1

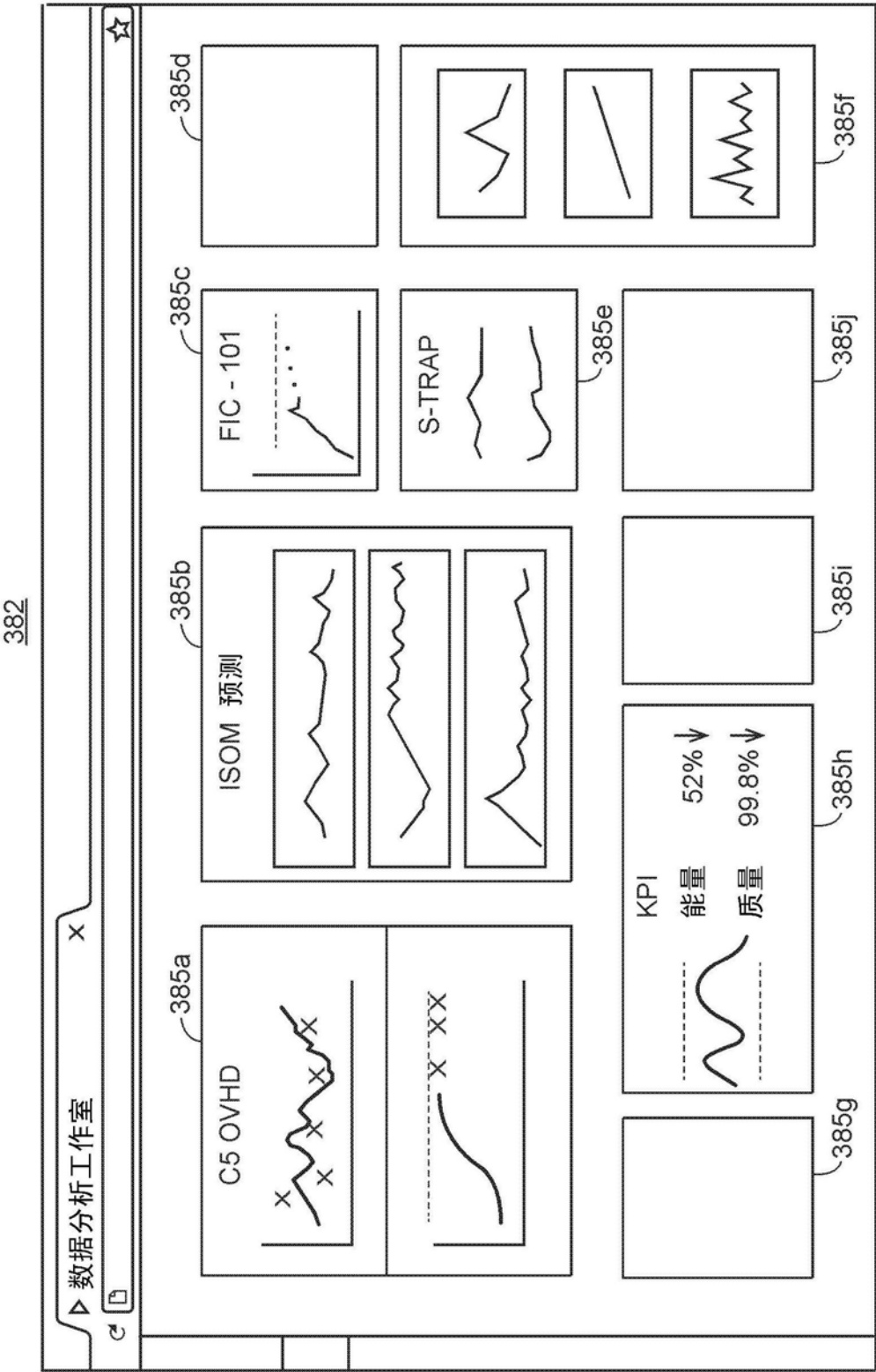


图4N-2

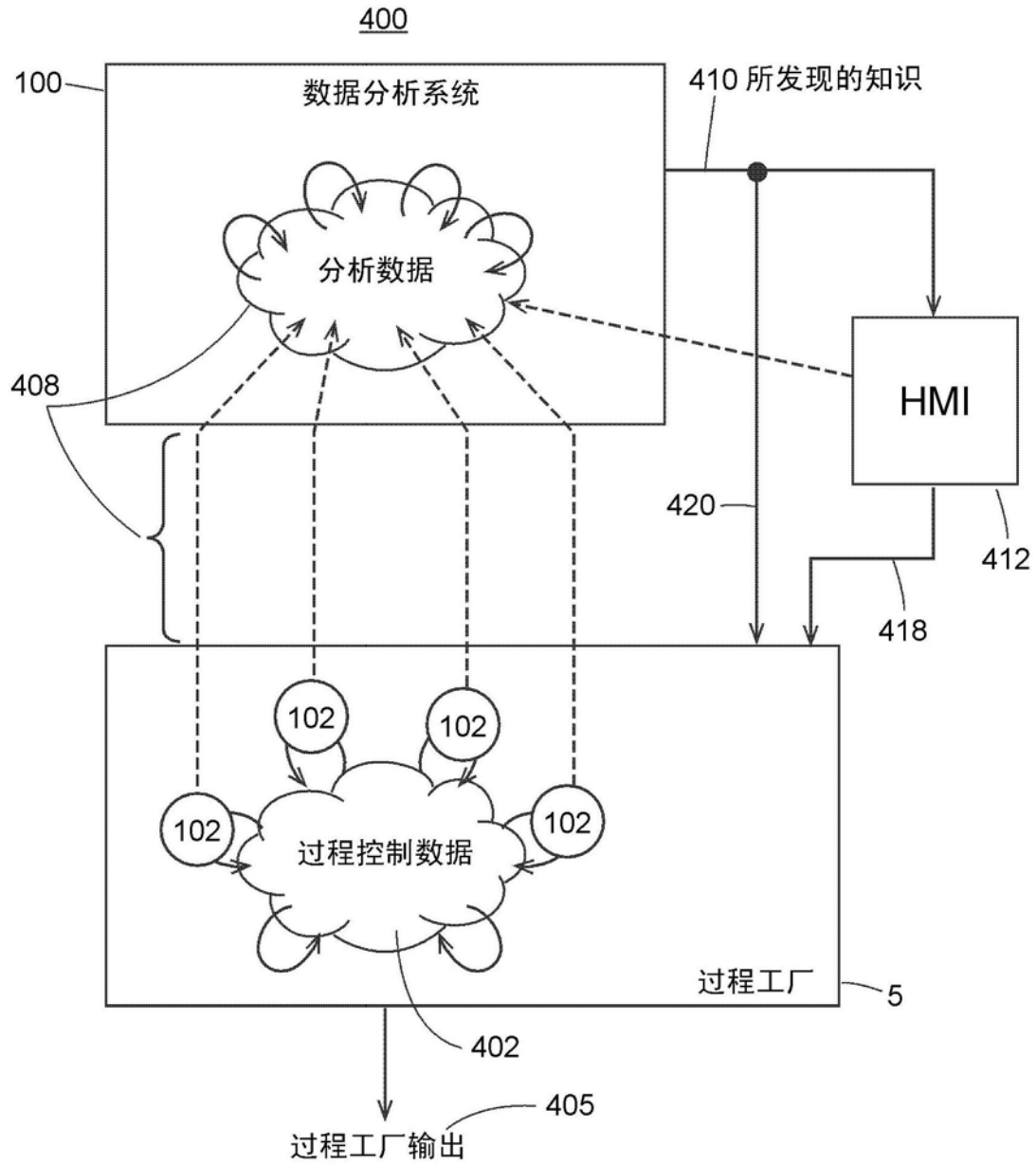


图40

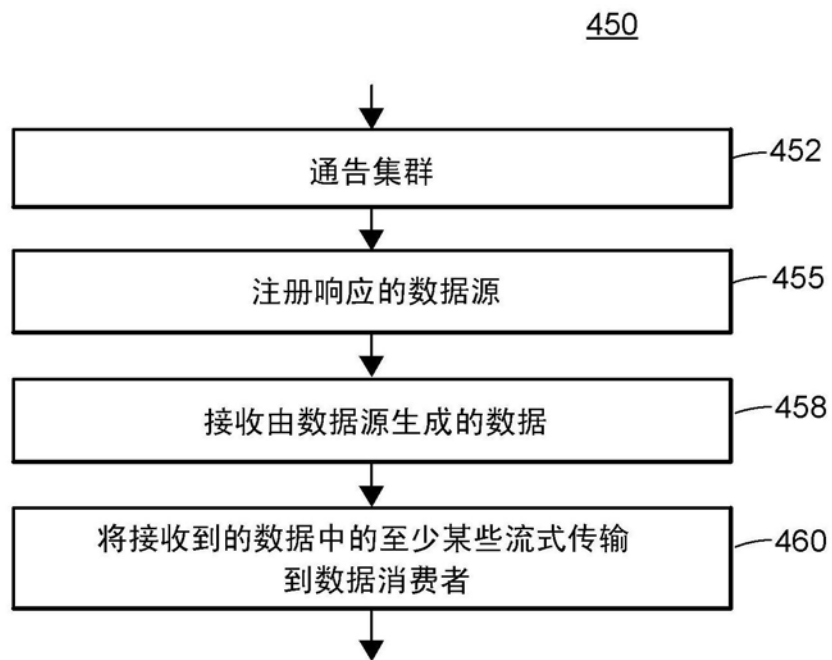


图4P

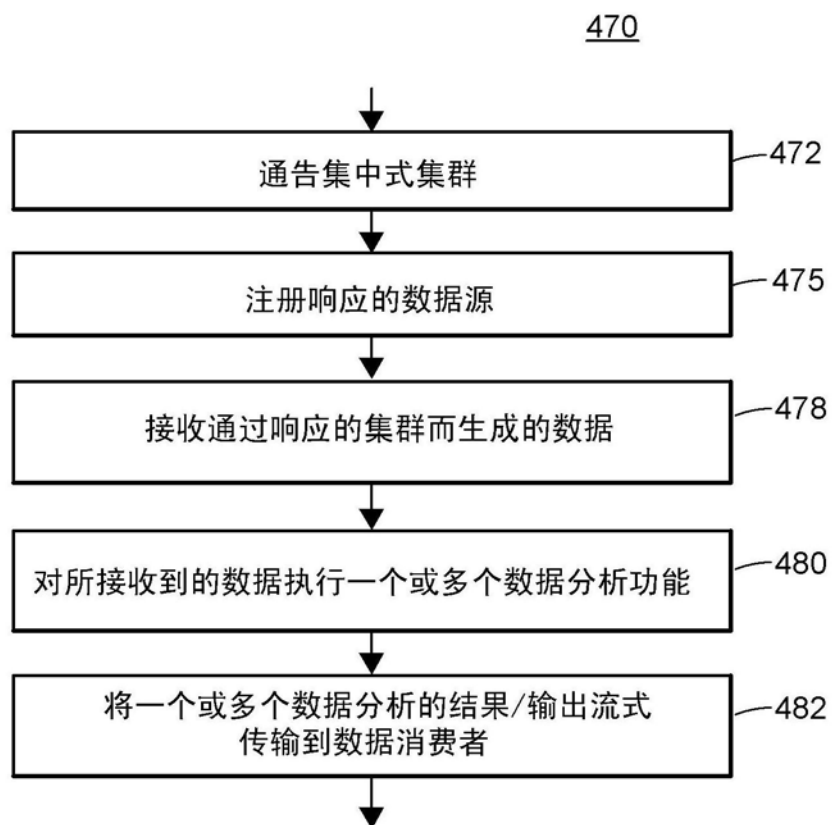


图4Q

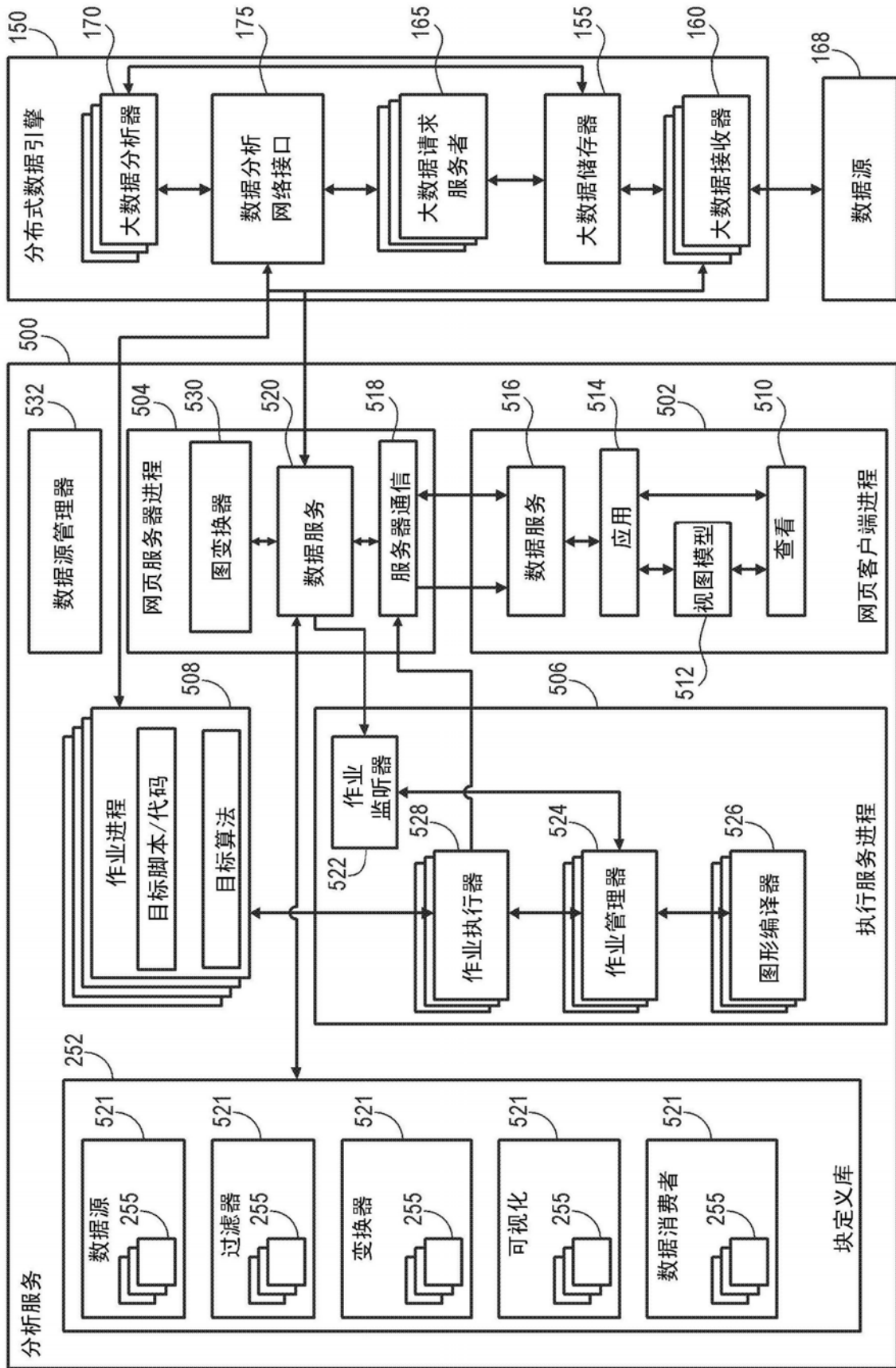


图5A

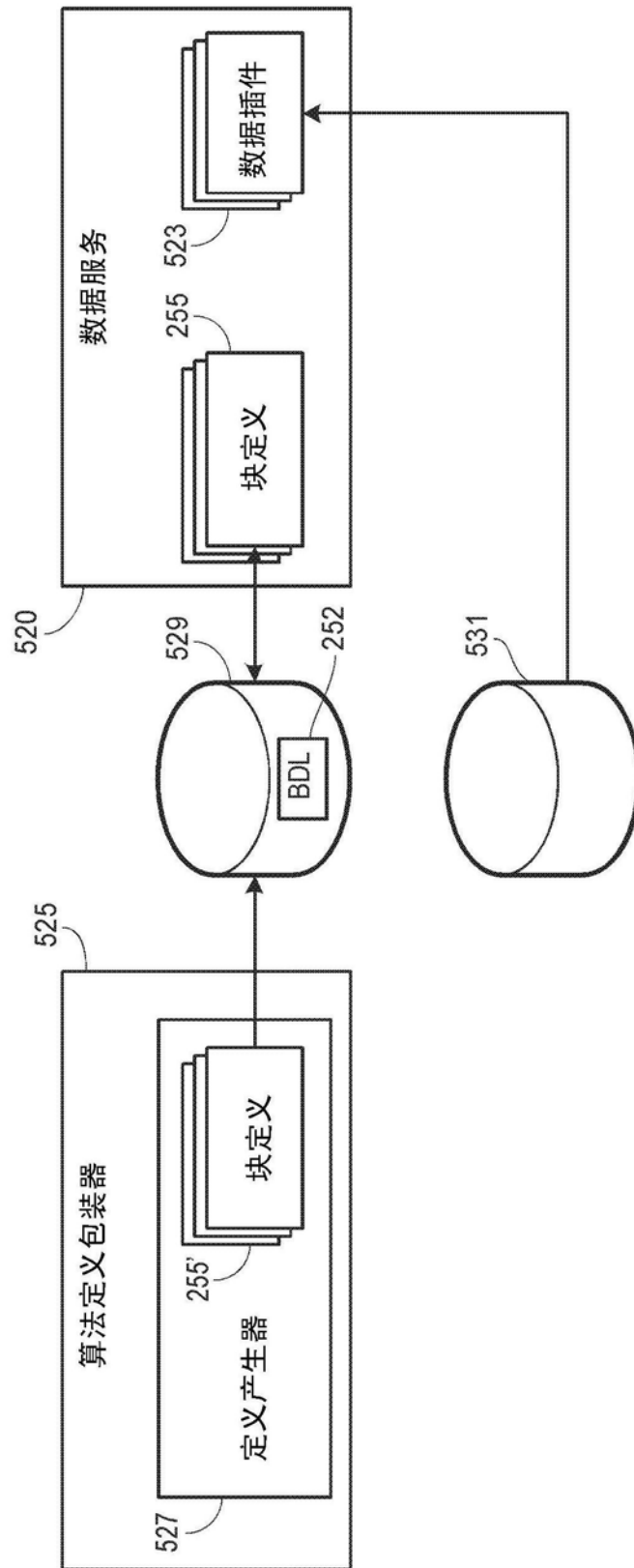


图5B

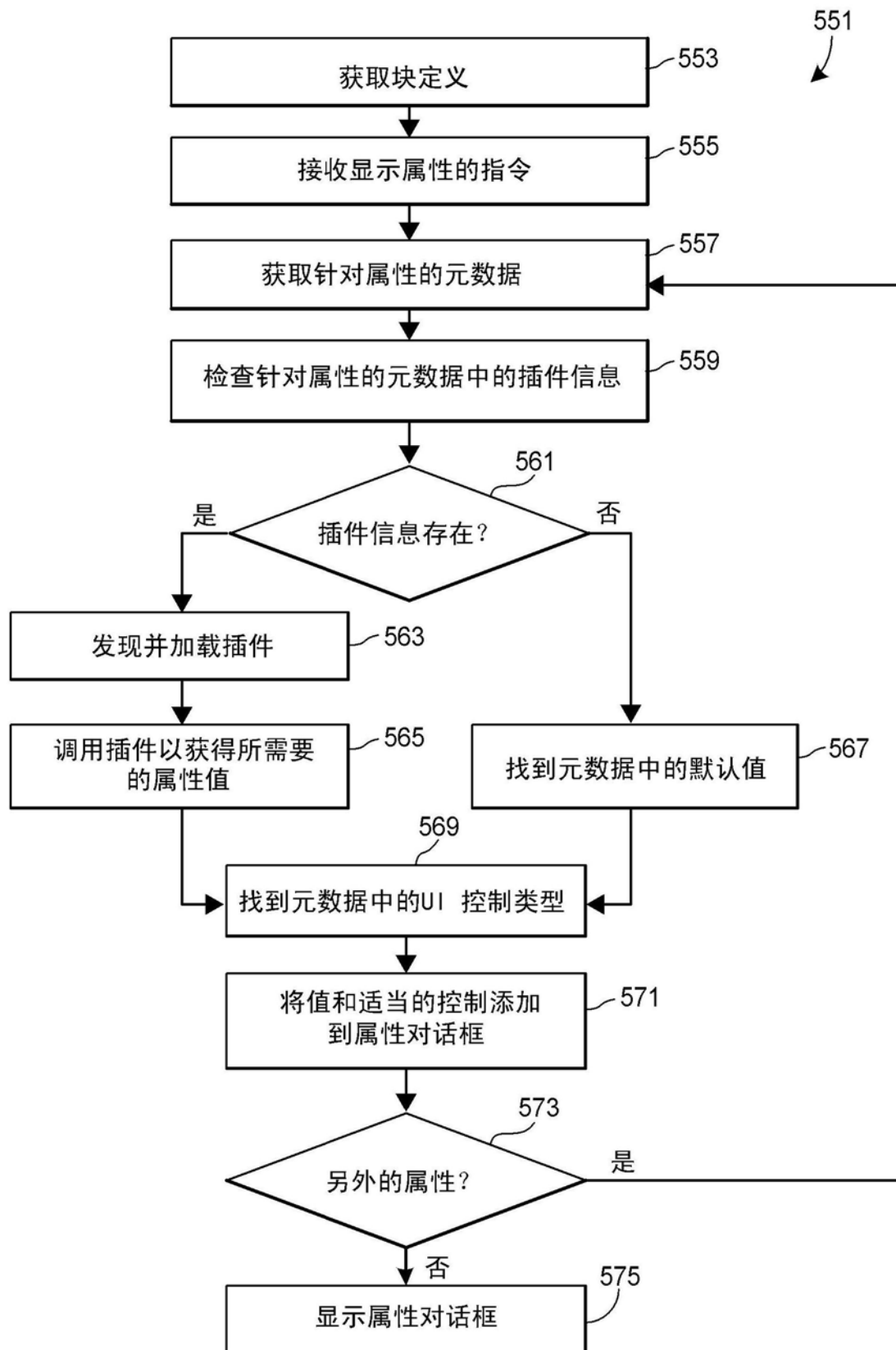
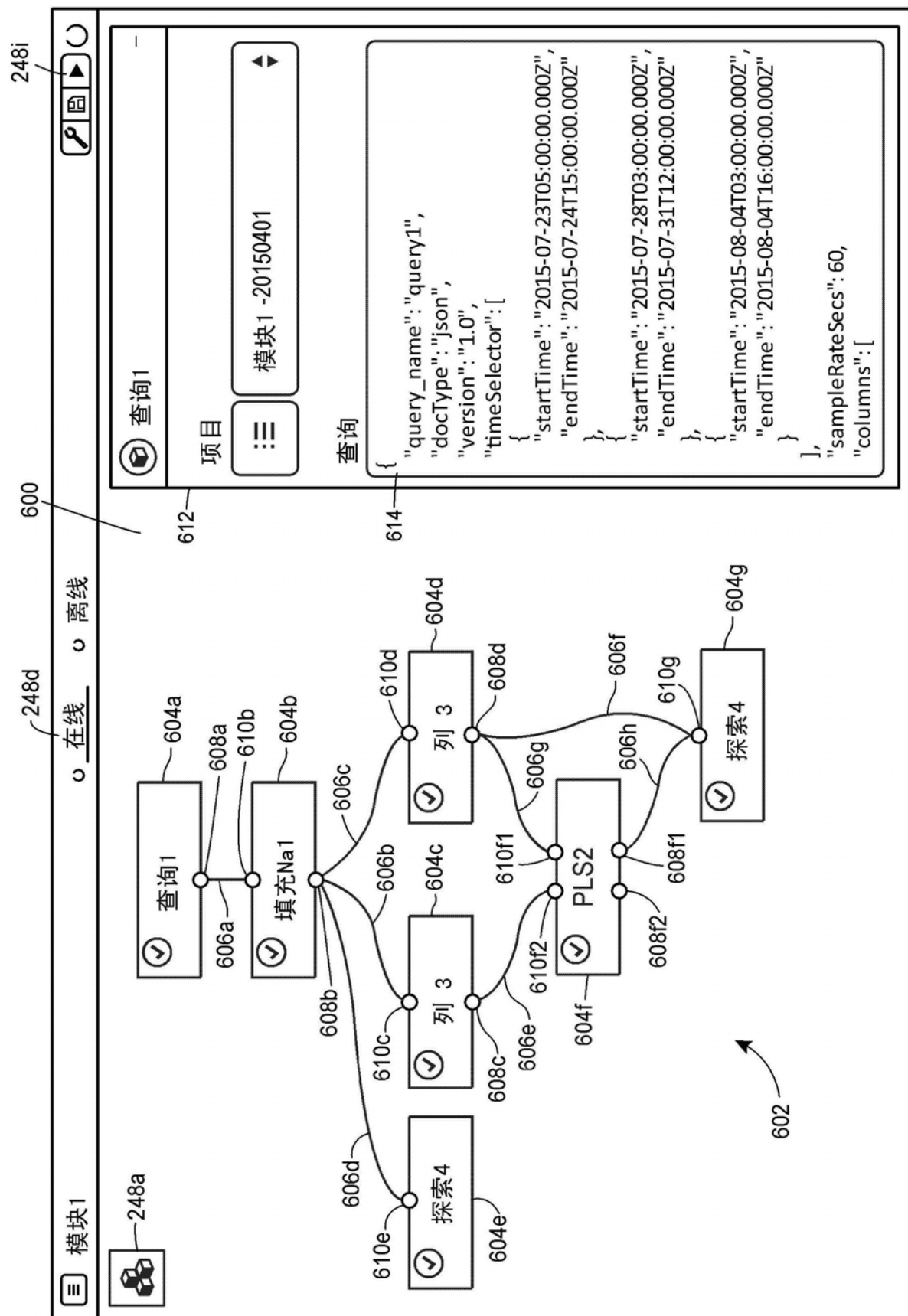


图5C



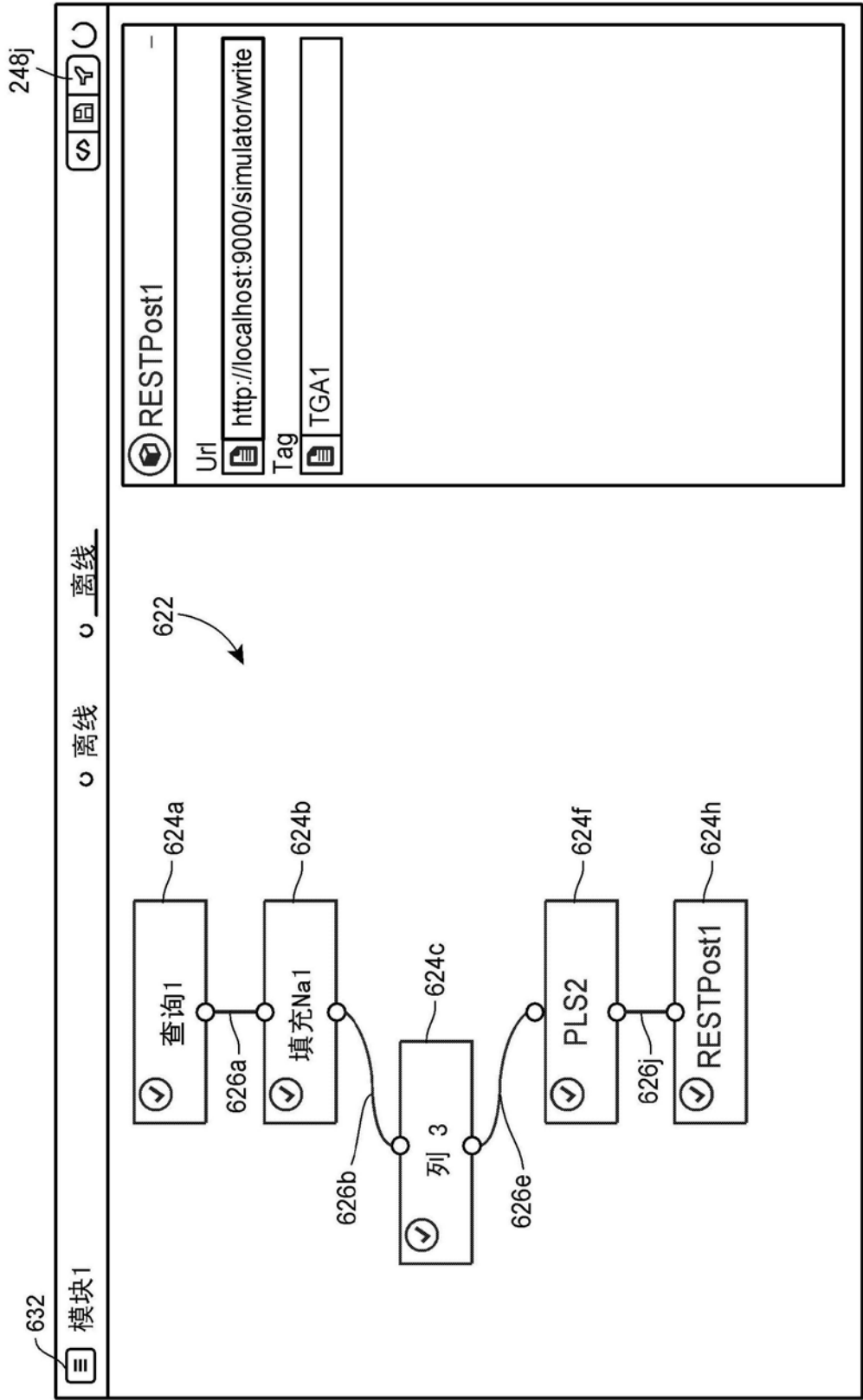


图5E

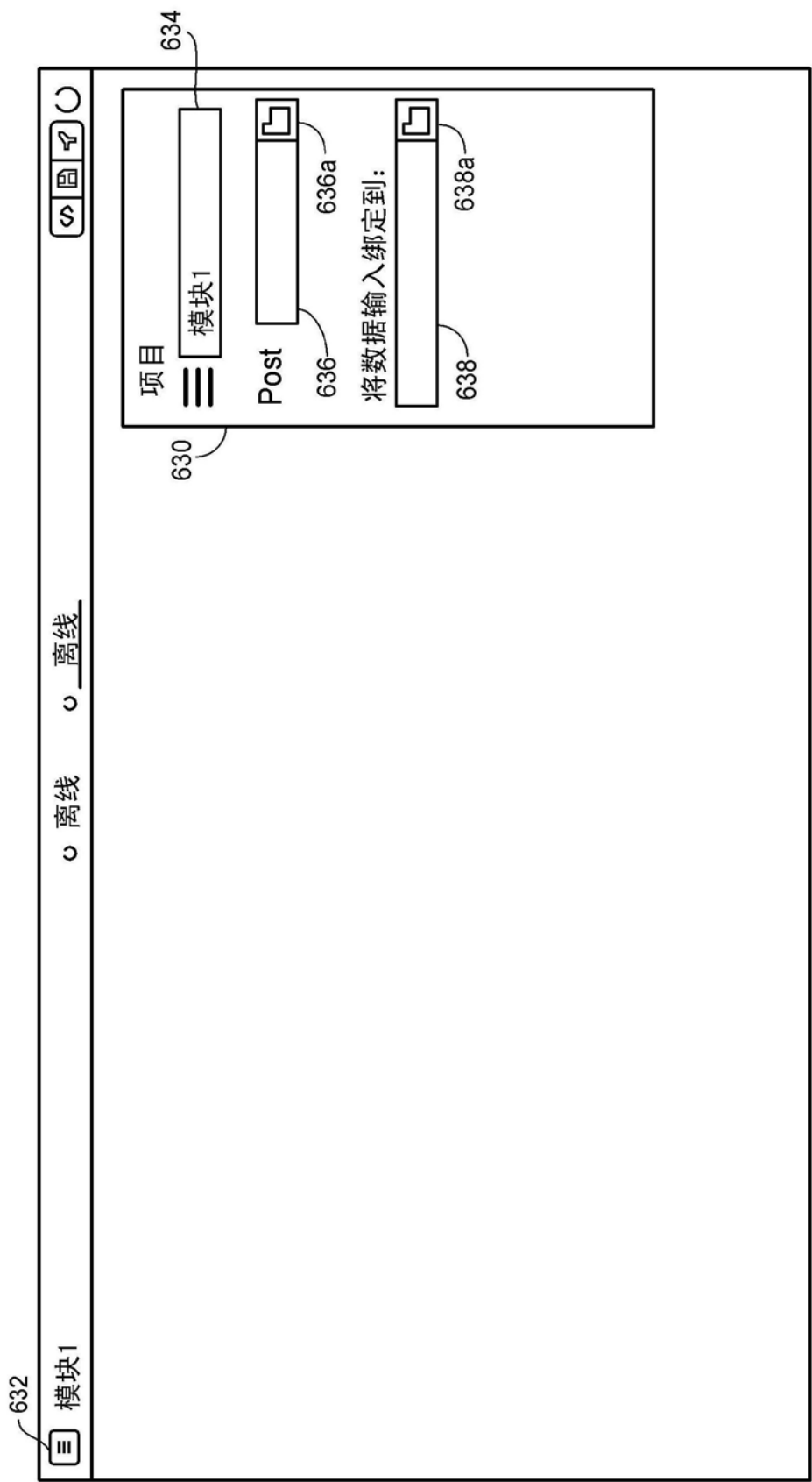


图5F

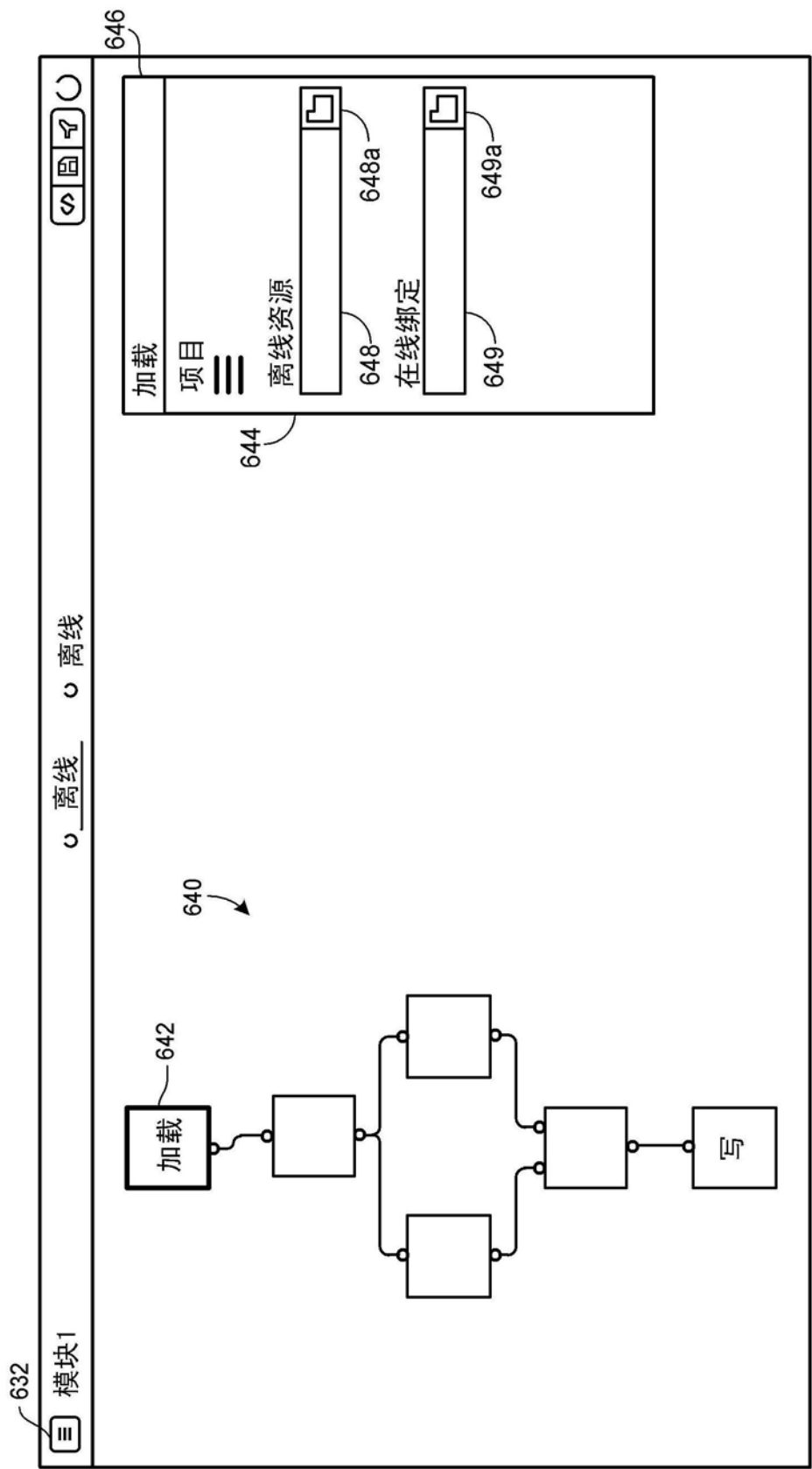


图5G

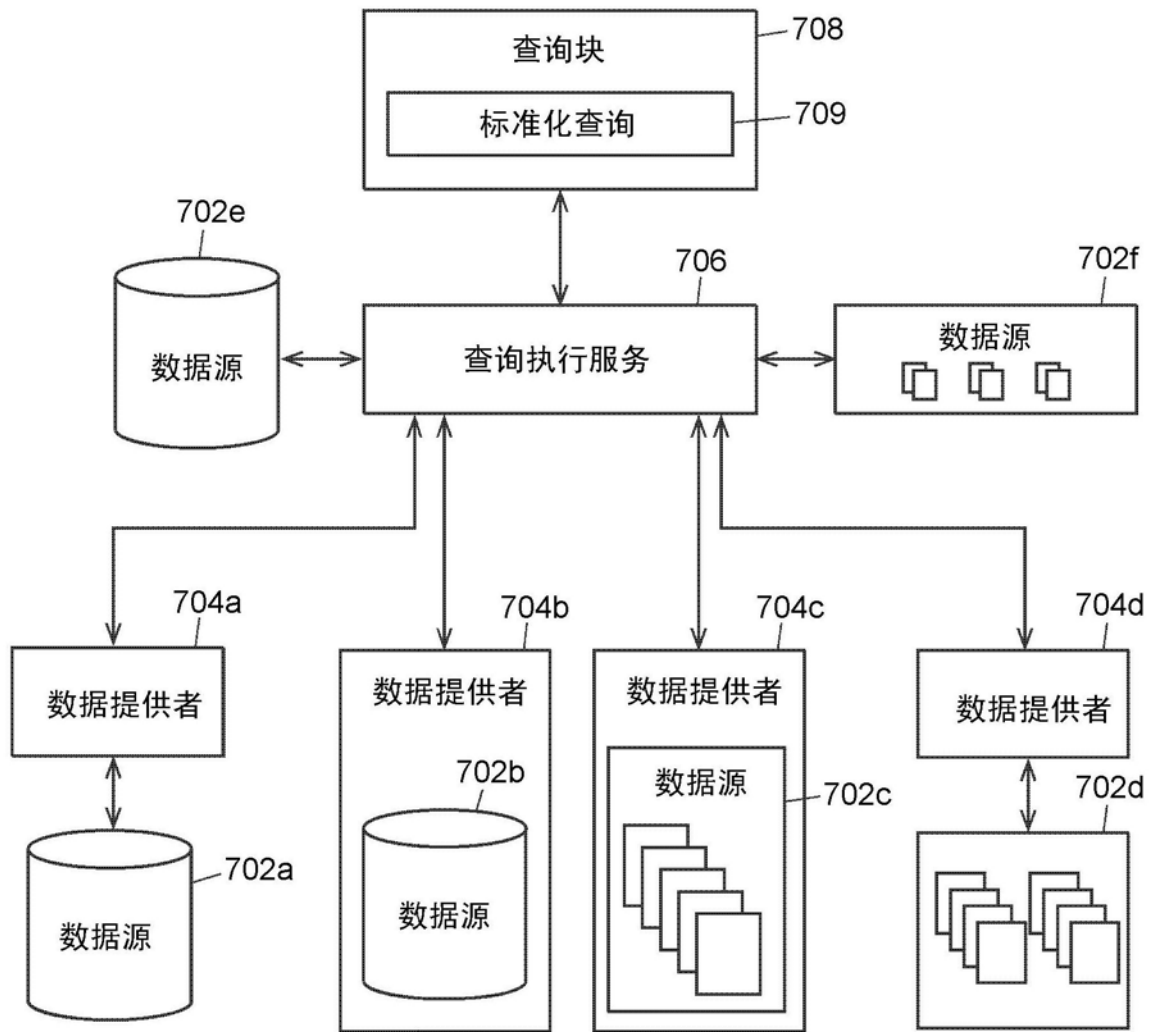


图6A

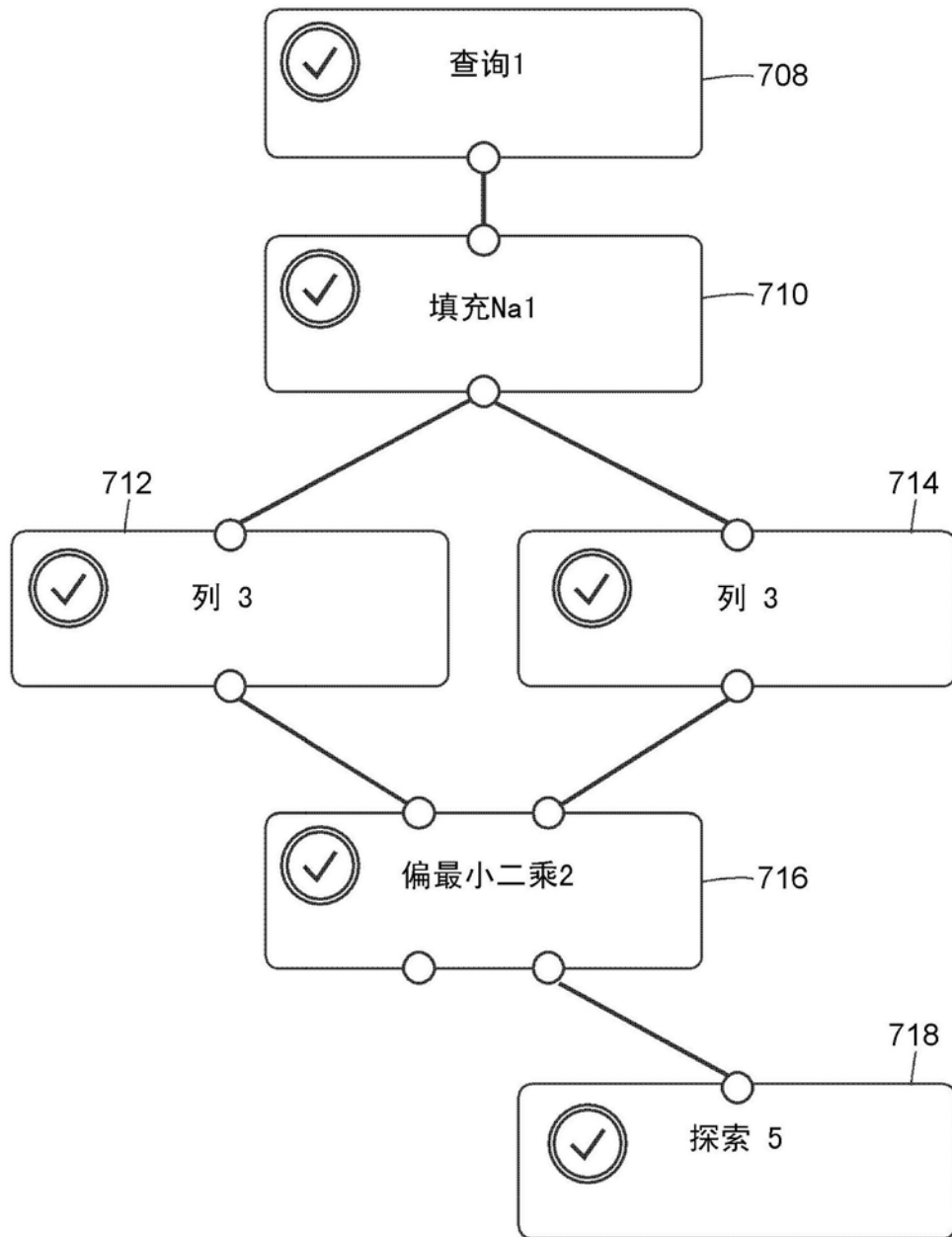


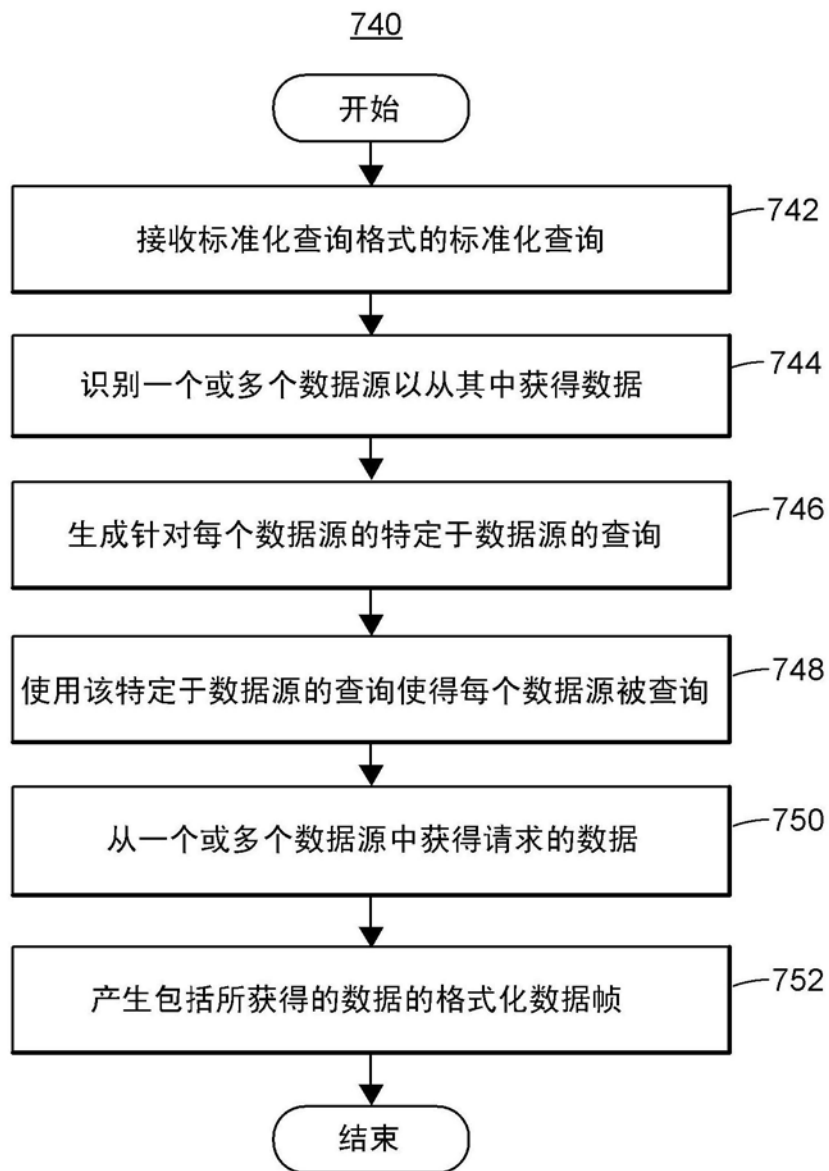
图6B

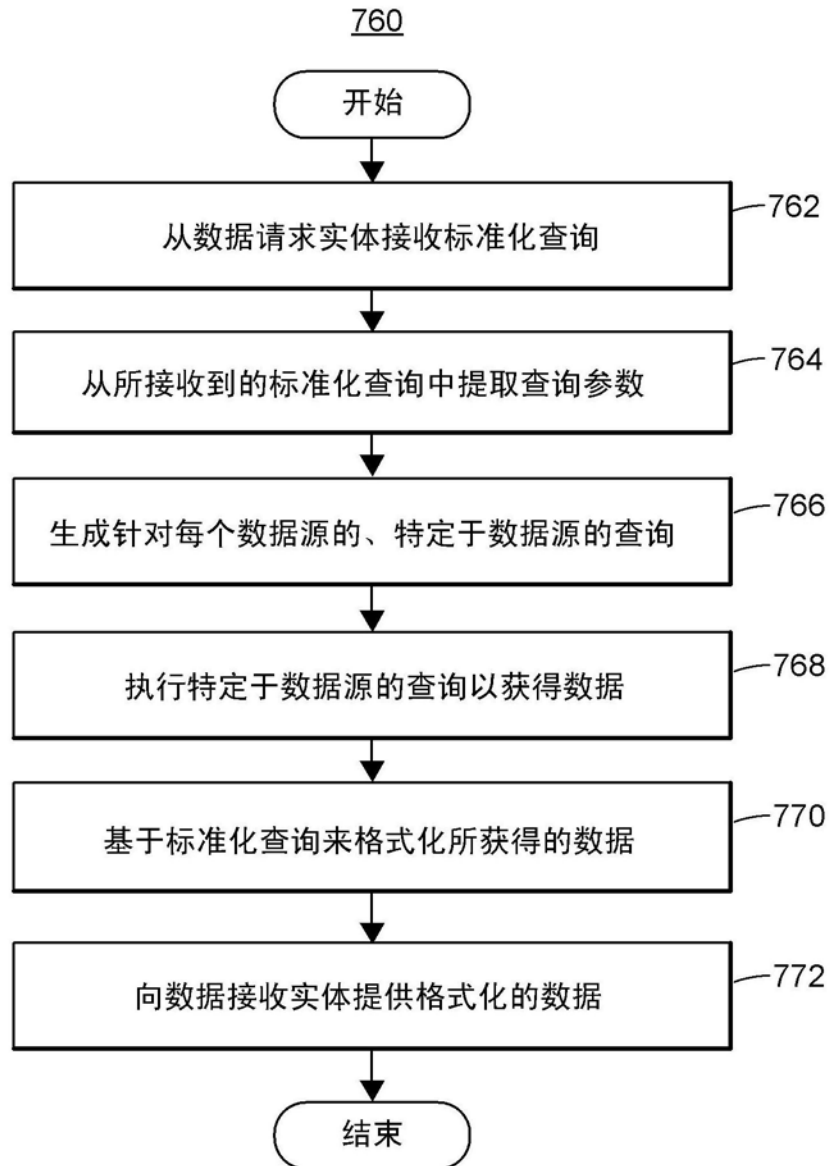
```

{
  "query_name": "query1",
  "docType": "json",
  "version": "1.0",
  "timeSelector": [
    {
      "startTime": "2015-07-23T05:00:00.000Z",
      "endTime": "2015-07-24T15:00:00.000Z"
    },
    {
      "startTime": "2015-07-28T03:00:00.000Z",
      "endTime": "2015-07-31T12:00:00.000Z"
    },
    {
      "startTime": "2015-08-04T03:00:00.000Z",
      "endTime": "2015-08-04T16:00:00.000Z"
    }
  ],
  "sampleRateSecs": 60,
  "columns": [
    {
      "tag": "FT630B/DENS.CV",
      "alias": "FT630B/DENS.CV",
      "dataType": "Float",
      "renderType": "VALUE",
      "format": "0.###"
    },
    {
      "tag": "PT615/WIRED_PSIA.CV",
      "alias": "PT615/WIRED_PSIA.CV",
      "dataType": "Float",
      "renderType": "VALUE",
      "format": "0.###"
    },
    {
      "tag": "TT6079/INPUT_1.CV",
      "alias": "TT6079/INPUT_1.CV",
      "dataType": "Float",
      "renderType": "VALUE",
      "format": "0.###"
    },
    {
      "tag": "630.molefrac.c5",
      "alias": "630.molefrac.c5",
      "dataType": "Float",
      "renderType": "VALUE",
      "format": "0.###"
    }
  ]
}

```

图6C





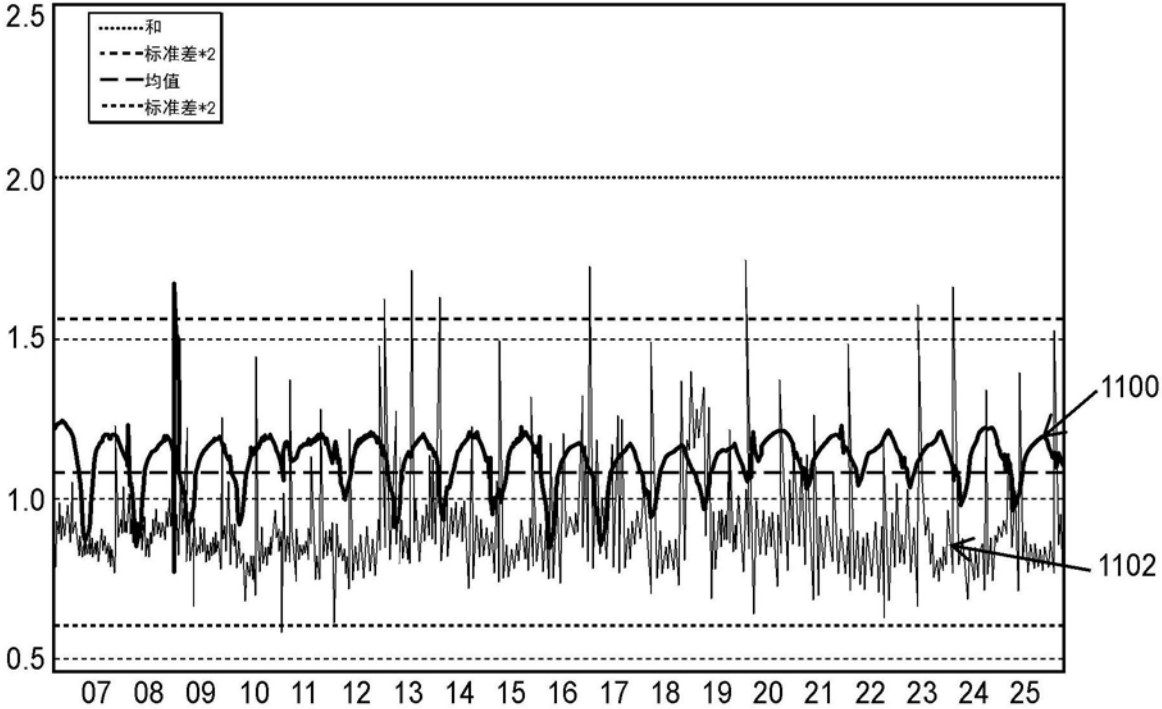


图7A

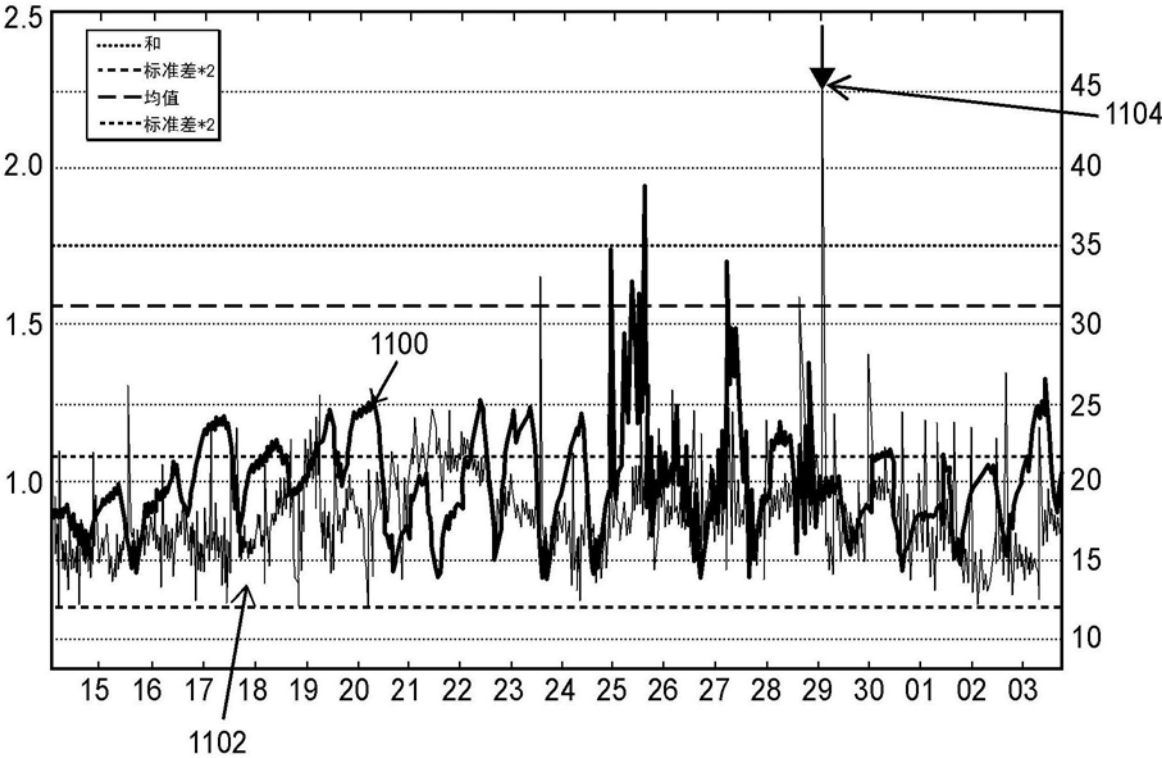


图7B

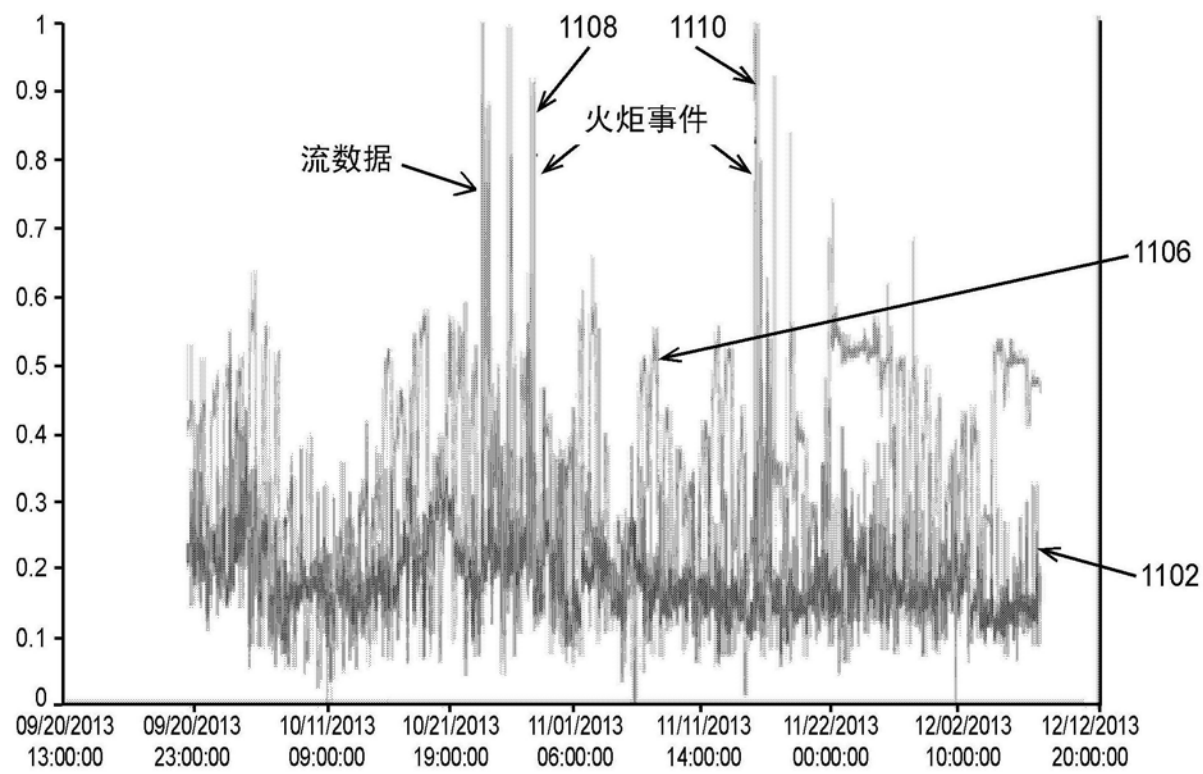


图7C

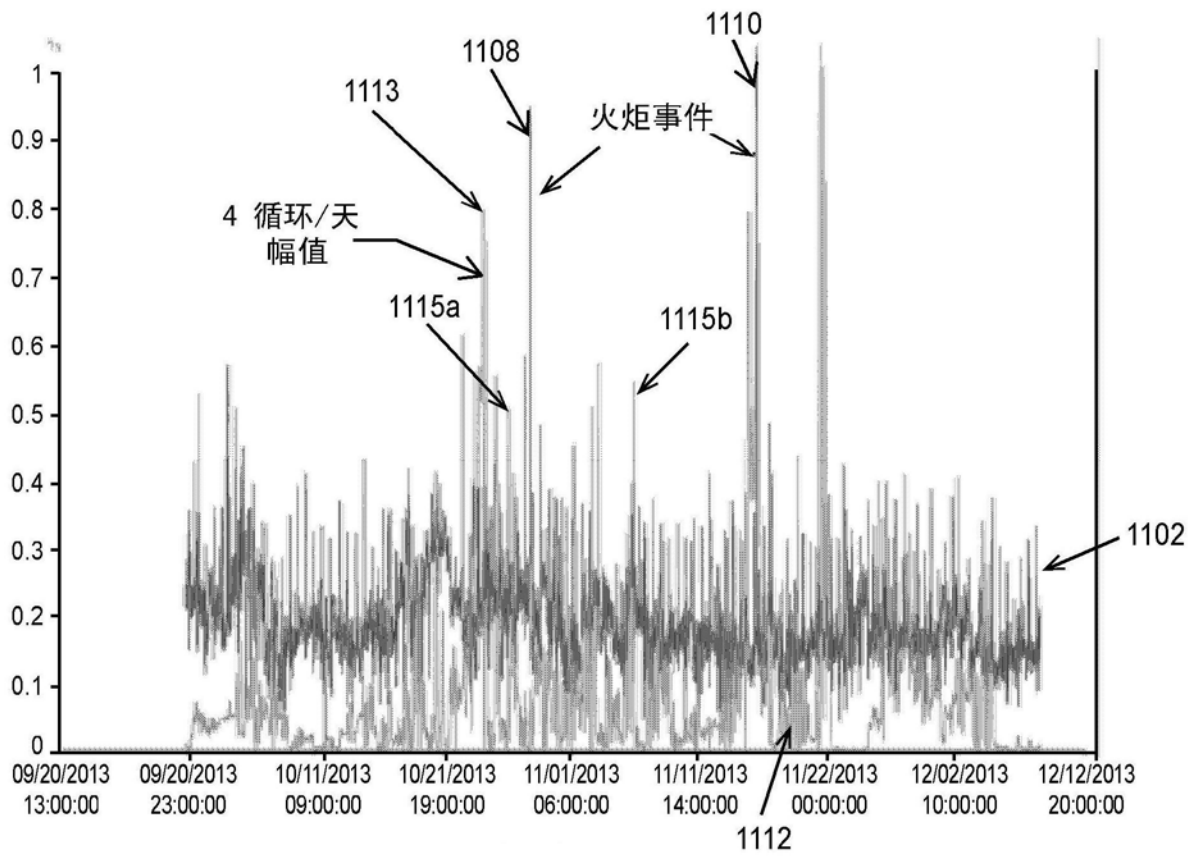


图7D

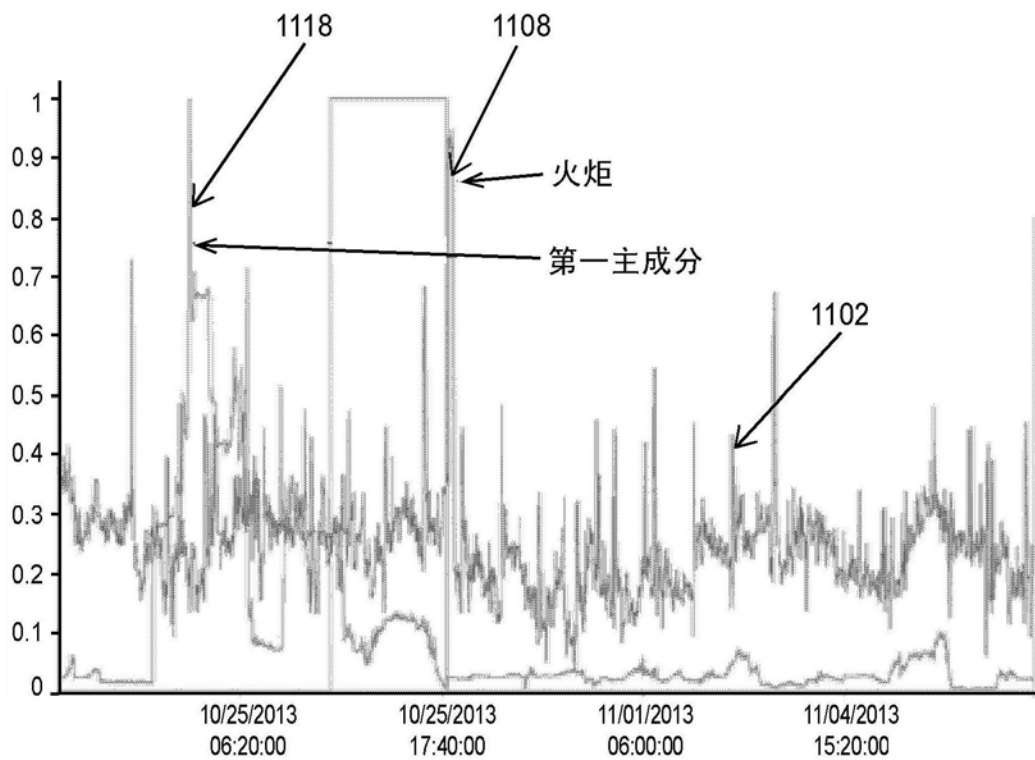


图7E

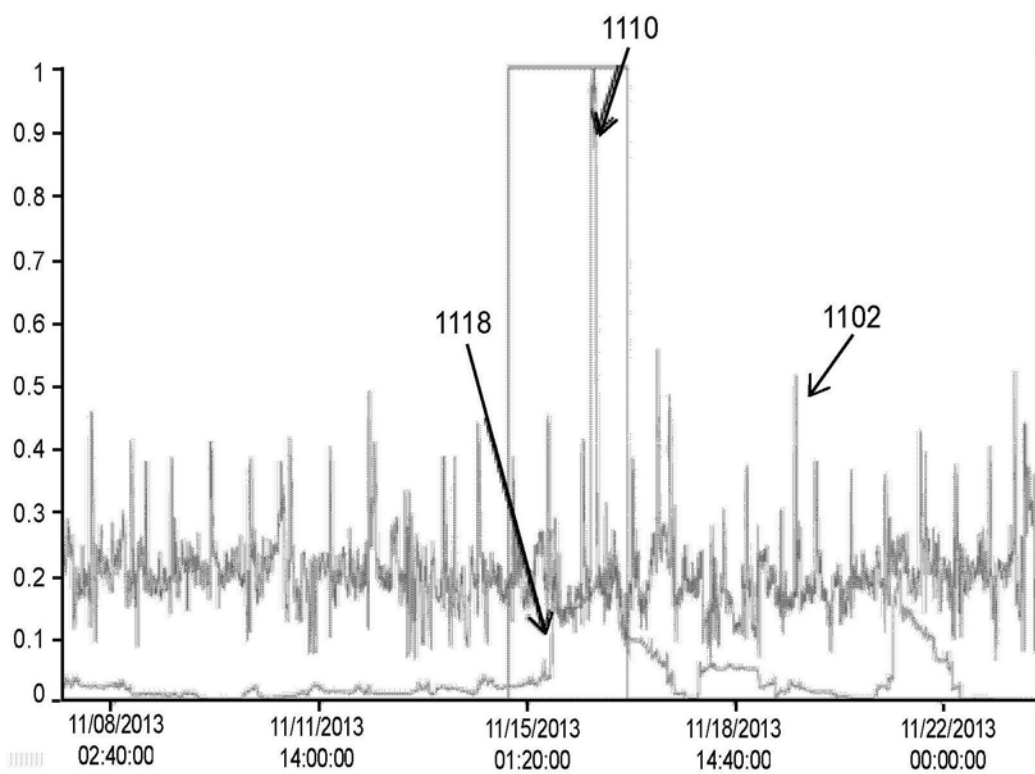


图7F

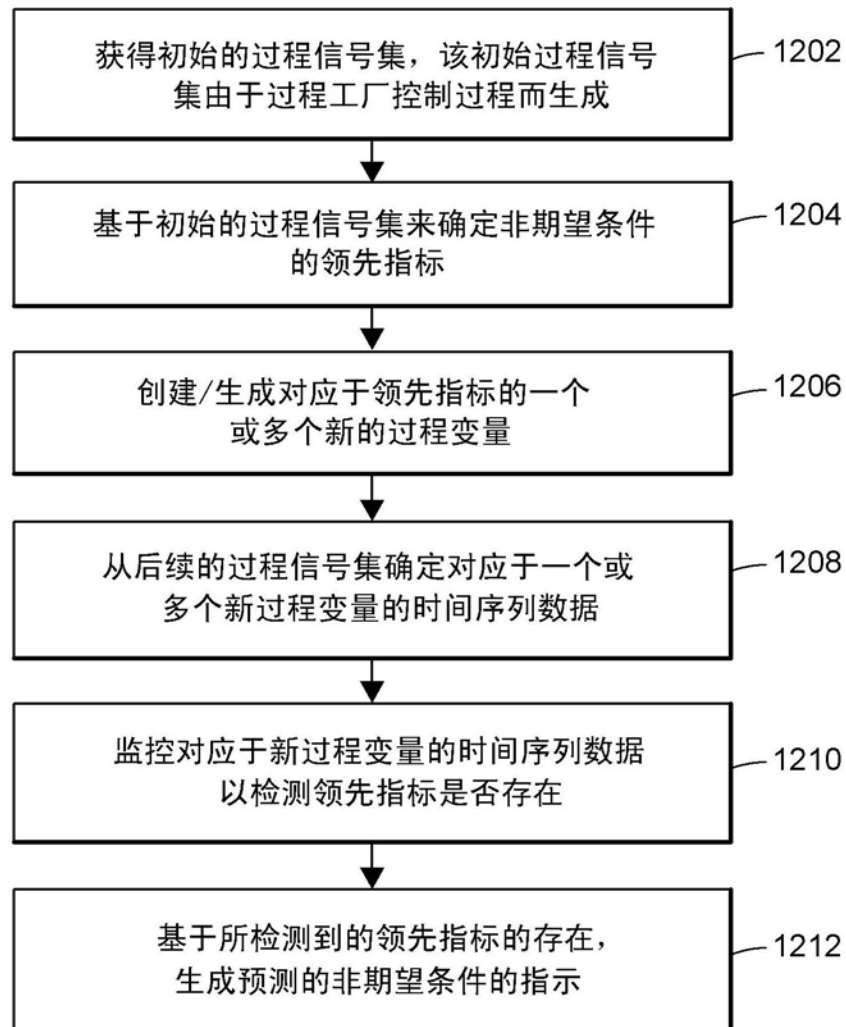
1200

图7G