(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2020 年 11 月 12 日 (12.11.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/223997 A1

(51) 国际专利分类号:
G06F 8/20 (2018.01)

(21) 国际申请号: PCT/CN2019/087192

(22) 国际申请日: 2019 年 5 月 16 日 (16.05.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910367895.5 2019年5月5日 (05.05.2019) CN

(71) 申请人: 东 北 大 学 (**NORTHEASTERN UNIVERSITY**) [CN/CN]; 中国辽宁省沈阳市和平区文化路三巷11号, Liaoning 110819 (CN)。

(72) 发明人: 宋杰 (**SONG, Jie**); 中国辽宁省沈阳市和平区文化路三巷11号, Liaoning 110819 (CN)。

李祥弘 (**LI, Xianghong**); 中国辽宁省沈阳市和平区文化路三巷11号, Liaoning 110819 (CN)。 张一川 (**ZHANG, Yichuan**); 中国辽宁省沈阳市和平区文化路三巷11号, Liaoning 110819 (CN)。 徐纯发 (**XU, Chunfa**); 中国辽宁省沈阳市和平区文化路三巷11号, Liaoning 110819 (CN)。 李锋 (**LI, Feng**); 中国辽宁省沈阳市和平区文化路三巷11号, Liaoning 110819 (CN)。

(74) 代理人: 沈阳东大知识产权代理有限公司 (**SHENYANG DONGDA INTELLECTUAL PROPERTY AGENCY CO., LTD**); 中国辽宁省沈阳市和平区文化路三巷11号, Liaoning 110819 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,

(54) **Title:** DATA ANALYSIS SOFTWARE ARCHITECTURE DESIGN METHOD CAPABLE OF IMPLEMENTING GLOBAL CONFIGURATION OF STORAGE, CALCULATION AND DISPLAY

(54) 发明名称: 一种存算显全局可配置的数据分析软件架构设计方法

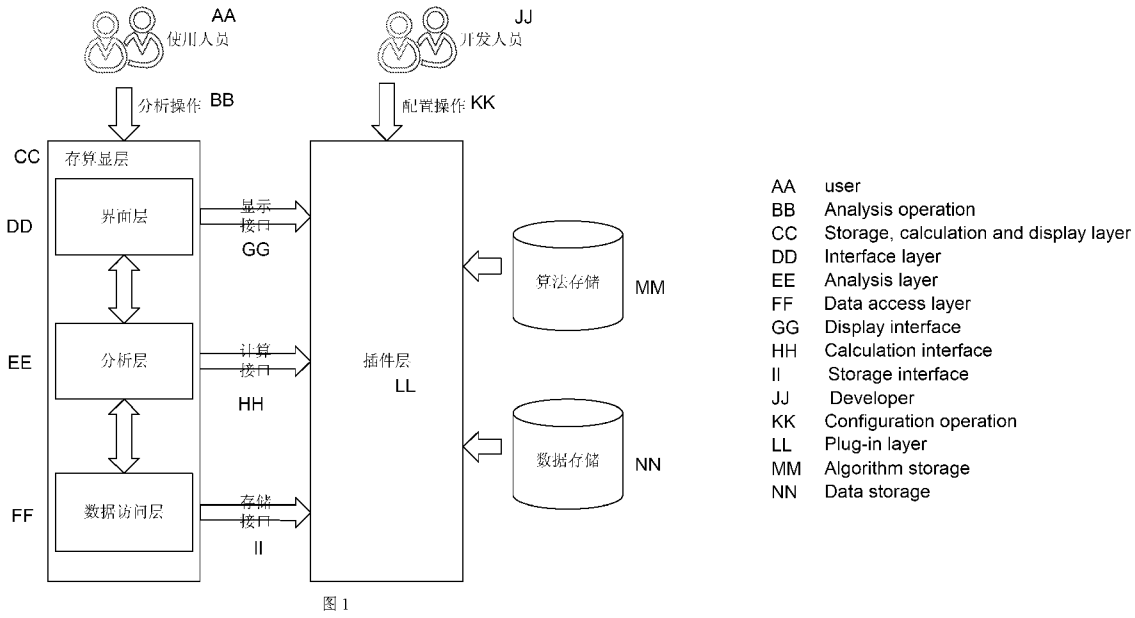

图 1

(57) **Abstract:** A data analysis software architecture design method capable of implementing global configuration of storage, calculation and display, relating to the technical field of software system architecture. According to the method, data analysis software architecture is divided into an interface layer, an analysis layer, a data access layer, and a plug-in layer. The data access layer, the analysis layer, and the interface layer are together named a storage, calculation and display layer. The interface layer provides an interactive visual

WO 2020/223997 A1

[见续页]

CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

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

根据细则4.17的声明：

— 关于申请人有权申请并被授予专利(细则4.17(ii))

本国际公布：

— 包括国际检索报告(条约第21条(3))。

interface for a software user. The analysis layer is responsible for performing algorithmic data analysis and processing the service logic of a software system. The data access layer obtains data from a data storage medium according to data analysis requirements and transmits the data to the analysis layer. The plug-in layer provides a software developer with a mode for configuring data and algorithms, parses the configuration of the data and algorithms, and provides an interface for the parsing result. According to the method, the designed software architecture satisfies the original intention of the design of software architecture which is friendly to use, has good expansibility, and is high in maintainability and high usability.

(57) 摘要：一种存算显全局可配置的数据分析软件架构设计方法，涉及软件系统架构技术领域。该方法将数据分析软件架构分为界面层、分析层、数据访问层和插件层，并将数据访问层、分析层、界面层三层的总和命名为存算显层；界面层为软件使用人员提供交互式操作的可视化界面；分析层负责执行算法数据分析，以及处理软件系统的业务逻辑；数据访问层根据数据分析需求从数据存储介质中获取数据并传入到分析层；插件层提供软件开发人员配置数据和算法的方式，解析上述数据和算法的配置，并提供解析结果的接口。所述方法使设计的软件架构满足使用友好、扩展性好、维护性强、可用性高的软件架构设计的初衷。

一种存算显全局可配置的数据分析软件架构设计方法

技术领域

[0001] 本发明涉及软件系统架构技术领域，尤其涉及一种存算显全局可配置的数据分析软件架构设计方法。

背景技术

[0002] 如今软件规模越来越庞大、功能越来越复杂，软件架构的设计随着也越来越受到软件从业人员的重视。软件架构作为软件的骨架，是决定软件质量的重要原因。一个软件在架构设计阶段出现的失误，很难在实际开发阶段予以彻底弥补，最终会造成开发和维护过程中大量资源的浪费，甚至导致软件的彻底失败。

[0003] 现有的数据分析软件原本大多是为特定分析目的而设计的，这使得数据分析软件在功能上只支持特定的算法种类与特定的数据参数。现有的软件架构设计方法所设计的软件架构不能轻易扩展数据分析所使用的算法、不能支持随着数据变化与分析目的变化造成的算法参数的动态变化。当分析目的和分析方法需要改变时，现有的软件难以扩展，常常需要开发一款新软件以用于新需求下的数据分析，从而使得原有软件被放弃，在满足数据分析人员的动态变化的分析需求的过程中造成了人力物力财力的大量浪费。

发明概述

技术问题

问题的解决方案

技术解决方案

[0004] 本发明要解决的技术问题是针对上述现有技术的不足，提供一种存算显全局可配置的数据分析软件架构设计方法，使用该方法设计出一种能够让软件开发人员灵活扩展数据与算法，且对软件使用人员使用友好的，扩展性好、维护性强、可用性高的数据分析软件架构。

[0005] 为解决上述技术问题，本发明所采取的技术方案是：一种存算显全局可配置的数据分析软件架构设计方法，将数据分析软件架构分为界面层、分析层、数据

访问层和插件层四个层次，其中将数据访问层、分析层、界面层三层的总和命名为存算显层；所述界面层为软件使用人员提供交互式操作的可视化界面；软件使用人员选择数据分析的算法和数据传送至分析层进行分析计算，并在可视化界面显示分析层返回的数据分析结果；所述分析层负责执行算法数据分析，以及处理软件系统的业务逻辑；所述数据访问层根据数据分析需求从数据存储介质中获取数据并传入到分析层，为分析层提供数据服务；所述插件层提供软件开发人员配置数据和算法的方式，解析上述数据和算法的配置，并提供解析结果的接口；所述接口用于界面层的显示接口调用、分析层的计算接口调用、以及数据访问层的存储接口调用。

[0006] 优选地，所述插件层配置两种外部存储类型：数据存储和算法存储；所述数据存储为数据分析所需数据集的外部存储介质；所述算法存储为数据分析执行时调用的外部算法代码包，其形式是输入数据集和算法参数并返回分析结果的函数或函数集的代码集合；软件开发人员将用于数据分析的数据存储和算法存储配置于插件层中，配置的具体数据格式和语义在软件开发设计阶段定义，开发人员在配置时遵循设计阶段定义的配置方式；所述数据存储的配置用于描述数据所存放的文件系统中的文件，描述数据文件或数据库中每个文件或数据库表的文件或表索引号和访问路径这些数据文件或表元数据，以及在文件或数据库表中每个属性及其子属性的属性索引号、属性名、属性数据类型这些数据属性元数据；所述算法存储的配置用于描述算法插件的基本信息，给出可显示在界面层的算法名、算法路径、算法输入参数约束、算法输出参数约束这些算法参数元数据；除了上述元数据外，开发人员根据具体情况扩展元数据；当软件部署并运行后，插件层将自动解析上述配置，作为界面层的显示接口、分析层的计算接口和数据访问层的存储接口的返回结果。

[0007] 优选地，所述数据访问层通过调用存储接口访问插件层；数据访问层在获得由界面层产生，由分析层传递的数据索引的情况下，调用存储接口获取插件层解析的相关数据索引的存储位置和存储数据类型这些数据存储配置信息；数据访问层根据存储接口返回结果访问外部数据存储，获取所需数据并根据存储数据类型保存为相应数据集对象，并将数据集对象作为数据分析的数据集返回至分

析层。

[0008] 优选地，所述分析层通过调用计算接口访问插件层；分析层在获取了界面层传入的算法索引和算法参数后，调用计算接口获取插件层解析的相关算法的访问路径、输入参数约束、输出参数约束这些算法存储配置信息；分析层根据计算接口返回的结果，唤起分析层内的数据分析引擎，传入根据计算接口返回结果封装后的算法参数和数据访问层返回的数据集至数据分析引擎中，执行算法存储中的算法包代码，完成数据分析计算，并将分析结果根据输出参数约束返回到界面层。

[0009] 优选地，所述界面层通过调用显示接口访问插件层；当软件架构使用人员访问可视化界面时，界面调用显示接口获取插件层解析的所有算法存储配置信息，构建动态的算法选择界面，软件架构使用人员从该界面选择算法，输入算法输入参数；同时调用显示接口获取插件层解析的所有数据存储配置信息，构建动态的数据集选择界面，软件架构使用人员从该界面选择数据集；软件架构使用人员获得的分析结果的可视化界面也由显示接口获取的算法输出参数约束为基础构建，结果的形式包括图像、表格、文字和分析过程日志。

发明的有益效果

有益效果

[0010] 采用上述技术方案所产生的有益效果在于：本发明提供的一种存算显全局可配置的数据分析软件架构设计方法，使设计的软件架构存算显全局可配置，在软件架构使用人员的需求发生变更时，开发人员可将新的数据存储或算法存储配置于插件层中，此时无需更改存算显层的代码，软件的相应模块即可通过调用存储接口、计算接口、显示接口，获取插件层相应的配置解析结果，从而访问新增的数据存储和算法存储，快速实现新需求下的数据分析，满足使用友好、扩展性好、维护性强、可用性高的软件架构设计的初衷。

对附图的简要说明

附图说明

[0011] 图1为本发明实施例提供的软件架构示意图；

[0012] 图2为本发明实施例提供的软件架构模块示意图；

[0013] 图3为本发明实施例提供的数据插件配置器配置实例；

[0014] 图4为本发明实施例提供的算法插件配置器配置实例；

[0015] 图5为本发明实施例提供的软件架构核心类示意图；

[0016] 图6为本发明实施例提供的软件架构的数据分析流程示意图。

发明实施例

本发明的实施方式

[0017] 下面结合附图和实施例，对本发明的具体实施方式作进一步详细描述。以下实施例用于说明本发明，但不用来限制本发明的范围。

[0018] 本实施例采用本发明的一种存算显全局可配置的数据分析软件架构设计方法设计一种数据分析软件架构。

[0019] 一种存算显全局可配置的数据分析软件架构设计方法，将数据分析软件架构分为界面层、分析层、数据访问层和插件层四个层次，如图1所示，其中将数据访问层、分析层、界面层三层的总和命名为存算显层；所述界面层为软件使用人员提供交互式操作的可视化界面；软件使用人员选择数据分析的算法和数据传送至分析层进行分析计算，并在可视化界面显示分析层返回的数据分析结果；所述分析层负责执行算法数据分析，以及处理软件系统的业务逻辑；所述数据访问层根据数据分析需求从数据存储介质中获取数据并传入到分析层，为分析层提供数据服务；所述插件层提供软件开发人员配置数据和算法的方式，解析上述数据和算法的配置，并提供解析结果的接口；所述接口用于界面层的显示接口调用、分析层的计算接口调用、以及数据访问层的存储接口调用。

[0020] 所述插件层配置两种外部存储类型：数据存储和算法存储；所述数据存储为数据分析所需数据集的外部存储介质，其形式包括但不限于各类数据库或数据文件系统；所述算法存储为数据分析执行时调用的外部算法代码包，其形式是输入数据集和算法参数并返回分析结果的函数或函数集的代码集合；软件开发人员将用于数据分析的数据存储和算法存储配置于插件层中，配置的具体数据格式和语义在软件开发设计阶段定义，开发人员在配置时遵循设计阶段定义的配置方式；所述数据存储的配置用于描述数据所存放的文件系统中的文件，描述数据文件或数据库中每个文件或数据库表的文件或表索引号和访问路径这些数

据文件或表元数据，以及在文件或数据库表中每个属性及其子属性的属性索引号、属性名、属性数据类型这些数据属性元数据；所述算法存储的配置用于描述算法插件的基本信息，给出可显示在界面层的算法名、算法路径、算法输入参数约束、算法输出参数约束这些算法参数元数据；除了上述元数据外，开发人员也可根据具体情况扩展元数据；当软件部署并运行后，插件层将自动解析上述配置，作为界面层的显示接口、分析层的计算接口和数据访问层的存储接口的返回结果。

[0021] 所述数据访问层通过调用存储接口访问插件层；数据访问层在获得由界面层产生，由分析层传递的数据索引的情况下，调用存储接口获取插件层解析的相关数据索引的存储位置和存储数据类型这些数据存储配置信息；数据访问层根据存储接口返回结果访问外部数据存储，获取所需数据并根据存储数据类型保存为相应数据集对象，并将数据集对象作为数据分析的数据集返回至分析层。

[0022] 所述分析层通过调用计算接口访问插件层；分析层在获取了界面层传入的算法索引和算法参数后，调用计算接口获取插件层解析的相关算法的访问路径、输入参数约束、输出参数约束这些算法存储配置信息；分析层根据计算接口返回的结果，唤起分析层内的数据分析引擎，传入根据计算接口返回结果封装后的算法参数和数据访问层返回的数据集至数据分析引擎中，执行算法存储中的算法包代码，完成数据分析计算，并将分析结果根据输出参数约束返回到界面层。

[0023] 所述界面层通过调用显示接口访问插件层；当软件架构使用人员访问可视化界面时，界面调用显示接口获取插件层解析的所有算法存储配置信息，构建动态的算法选择界面，软件架构使用人员从该界面选择算法，输入算法输入参数；同时调用显示接口获取插件层解析的所有数据存储配置信息，构建动态的数据集选择界面，软件架构使用人员从该界面选择数据集；软件架构使用人员获得的分析结果的可视化界面也由显示接口获取的算法输出参数约束为基础构建，结果的形式包括图像、表格、文字和分析过程日志。

[0024] 本实施例中，采用本发明的软件架构设计方法设计的数据分析软件架构在存算显层共有5个模块，分别为参数选择模块、结果展示模块、数据分析引擎、业务

逻辑模块、数据访问模块；在插件层有4个模块，分别为数据插件配置器、算法插件配置器、数据插件解析器、算法插件解析器；存算显层分别通过显示接口、计算接口、存储接口访问插件层，如图2所示。

[0025] 本实施例使用文件系统中的文件作为软件架构中的数据存储，使用可执行的R语言算法包作为软件架构中的算法存储。文件系统中有若干文件，文件内数据的形式是二维表，其中横坐标是若干数据属性，纵坐标为日期，一行数据表示某个日期下若干属性的值；通过确定文件名、文件内属性列名和日期的方式可以确定该属性在某个日期内的值。R语言算法包是用R语言实现的数据分析算法基本功能的代码文件，当服务器部署了R执行环境后，输入算法包内的函数所需的参数后，即可输出分析结果。

[0026] 本实施例在插件层中，实现了用于配置数据存储的数据插件配置器，其描述形式为XML文件。在数据插件配置器中，使用category标签描述文件系统中的文件信息，包括文件索引号、文件名、文件路径、文件内数据起始日期和数据结束日期（本实施例中，文件数据一行表示一天，上下行之间为连续不间断的日期，故定义第一数据行的日期为起始日期，最后一个数据行的日期为结束日期，数据插件解析器即可解析出文件的行数）；使用attribute标签描述文件内每列属性的信息，包括属性索引号、属性名、列数和属性数据类型。category的子标签attributes是该文件下所有attribute标签所描述的数据属性的集合。

[0027] 数据插件配置器对于“操业参数”文件（文件索引号21）的配置实例如图3所示，该文件内的起始时间和结束时间表示文件第一行数据和最后一行数据的对应日期，每个attribute标签代表文件每列数据的索引、名称、列数和数据类型。

[0028] 本实施例在插件层中，实现了用于配置算法存储的算法插件配置器，其描述形式为XML文件。在算法插件配置器中，使用algorithm标签描述R算法包的算法信息，包括算法包索引号、算法包名、算法包调用函数名、算法包路径、算法包依赖库、返回图表类型。algorithm的子标签parameters是该算法所有算法参数的集合，算法参数的描述标签为parameter，包括算法参数索引号、算法参数类型和参数约束。若算法参数是选择型，则在parameter标签下加上option子标签描述具体选项；而输入型的算法参数则无需option子标签。本实施例中，将算法存储

所含的R语言算法包进行了归类，分为“描述统计”、“统计分析”和“数据挖掘”三大类算法，以RPackage标签描述；在三大类算法下也通过approach标签分别划分为若干小类。而algorithms是approach标签下的子标签，表示描述R算法包的algorithm标签的集合，在界面层，软件使用人员能通过大类标签快速定位和选择所用算法。例如在本实施例的界面上，界面顶部是算法选择区，上述三大类算法是算法选择区的一级菜单，approach标签为二级菜单，algorithm标签为三级菜单，在选择算法时鼠标点击三级菜单进入相应具体三级算法的算法参数输入界面。

[0029] 算法插件配置器对于两种“时间序列分析”算法的配置实例如图4所示，其中单变量指数预测模型（算法索引号“3_5_1”）需要引入第三方R库“lubridate”，输出图像类型为散点图，输入限制为“只能选择一个数据属性”，模型预测月数为不小于1的整数，不得为空；单变量ARIMA预测模型（算法索引号“3_5_2”）需要引入第三方R库“tseries”和“forecast”，输出图像类型为散点图，输入限制为“只能选择一个数据属性”，模型预测月数为不小于1的整数，不得为空。

[0030] 本实施例在存算显层中，服务器端使用Java Web技术开发，服务器部分的核心类图如图5所示。在分析层中，AttributePrase类和RPackagePrase类分别为架构中的数据插件解析器和算法插件解析器。数据插件解析器和算法插件解析器在软件服务器启动时解析上述用XML描述的数据插件配置器和算法插件配置器，反序列化为内存对象。AttributePrase类解析的对象保存在CategoriesPool类中，RPackagePrase类解析的对象保存在AlgorithmsPool类中，其余模块调用插件配置信息只需调用上述类的对象的成员方法。AttributesServlet类和RPackageServlet类属于本实施例中软件架构的界面层，作用是当软件架构使用人员访问客户端时，这两个类分别将CategoriesPool对象和AlgorithmsPool对象的数据封装后发送之客户端，从而完成参数选择模块的功能。DatasetPool类和DatasetFactory类属于数据访问层的数据访问模块。DatasetFactory类的作用是读取文件系统中的数据，其对数据的定位与CategoriesPool对象存在依赖关系。DatasetFactory类读取的数据以DatasetPool类的对象形式保存。AnalyzeCore类是数据分析过程中调用数据分析引擎前解析客户端参数、获取数据分析所需算法和数据集并将引擎的算法分析结果封装并返回到客户端的业务逻辑模块类。数据分析引擎的功能由REngine类完

成，接受AnalyzeCore对象的算法参数和数据集，调用R执行算法包的算法，并将执行结果简易封装为Java对象返回到AnalyzeCore中。

[0031] 本实施例中，数据分析流程的核心业务流程如图6所示。软件使用人员打开客户端后，客户端自动从AttributesServlet类和RPackageServlet类获取配置信息，并根据信息自动构建界面的参数选择模块。软件使用人员在客户端中首先选择算法，随后搜索并选择加入数据集的数据属性（即选择数据文件内数据的列的范围），接着输入或选择算法参数（输入框或选择框由RPackageServlet传来的算法信息生成），最后选择数据的时间范围（即选择数据文件内数据的行的范围）并点击分析按钮。上述信息随即传到服务器端，由ServiceServlet类接收并转发到AnalyzeCore类。AnalyzeCore类解析客户端传来的信息，首先根据信息调用DatasetPool对象中的数据，而DatasetPool对象又通过DatasetFactory类的方法从文件系统获取数据集。AnalyzeCore类获得数据后将数据集和算法信息作为参数调用REngine类，REngine类根据传入的算法索引，调用服务器中部署的R环境执行相应的R算法包，开始数据分析过程。数据分析结束后，REngine类将分析结果（成功的分析结果或失败的错误代号）返回AnalyzeCore类，AnalyzeCore进行封装后由ServiceServlet类通过HTTP协议返回分析结果到客户端。

[0032] 最后应说明的是：以上实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述实施例所记载的技术方案进行修改，或者对其中部分或者全部技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明权利要求所限定的范围。

权利要求书

[权利要求 1] 一种存算显全局可配置的数据分析软件架构设计方法，其特征在于：将数据分析软件架构分为界面层、分析层、数据访问层和插件层四个层次，其中将数据访问层、分析层、界面层三层的总和命名为存算显层；所述界面层为软件使用人员提供交互式操作的可视化界面；软件使用人员选择数据分析的算法和数据传送至分析层进行分析计算，并在可视化界面显示分析层返回的数据分析结果；所述分析层负责执行算法数据分析，以及处理软件系统的业务逻辑；所述数据访问层根据数据分析需求从数据存储介质中获取数据并传入到分析层，为分析层提供数据服务；所述插件层提供软件开发人员配置数据和算法的方式，解析上述数据和算法的配置，并提供解析结果的接口；所述接口用于界面层的显示接口调用、分析层的计算接口调用、以及数据访问层的存储接口调用。

[权利要求 2] 根据权利要求1所述的一种存算显全局可配置的数据分析软件架构设计方法，其特征在于：所述插件层配置两种外部存储类型：数据存储和算法存储；所述数据存储为数据分析所需数据集的外部存储介质；所述算法存储为数据分析执行时调用的外部算法代码包，其形式是输入数据集和算法参数并返回分析结果的函数或函数集的代码集合；软件开发人员将用于数据分析的数据存储和算法存储配置于插件层中，配置的具体数据格式和语义在软件开发设计阶段定义，开发人员在配置时遵循设计阶段定义的配置方式；所述数据存储的配置用于描述数据所存放的文件系统中的文件，描述数据文件或数据库中每个文件或数据库表的文件或表索引号和访问路径这些数据文件或表元数据，以及在文件或数据库表中每个属性及其子属性的属性索引号、属性名、属性数据类型这些数据属性元数据；所述算法存储的配置用于描述算法插件的基本信息，给出可显示在界面层的算法名、算法路径、算法输入参数约束、算法输出参数约束这些算法参数元数据；除了上述元数据外，开发人员根据具体情况扩展元数据；当软件部署并运行后，

插件层将自动解析上述配置，作为界面层的显示接口、分析层的计算接口和数据访问层的存储接口的返回结果。

[权利要求 3] 根据权利要求1所述的一种存算显全局可配置的数据分析软件架构设计方法，其特征在于：所述数据访问层通过调用存储接口访问插件层；数据访问层在获得由界面层产生，由分析层传递的数据索引的情况下，调用存储接口获取插件层解析的相关数据索引的存储位置和存储数据类型这些数据存储配置信息；数据访问层根据存储接口返回结果访问外部数据存储，获取所需数据并根据存储数据类型保存为相应数据集对象，并将数据集对象作为数据分析的数据集返回至分析层。

[权利要求 4] 根据权利要求1所述的一种存算显全局可配置的数据分析软件架构设计方法，其特征在于：所述分析层通过调用计算接口访问插件层；分析层在获取了界面层传入的算法索引和算法参数后，调用计算接口获取插件层解析的相关算法的访问路径、输入参数约束、输出参数约束这些算法存储配置信息；分析层根据计算接口返回的结果，唤起分析层内的数据分析引擎，传入根据计算接口返回结果封装后的算法参数和数据访问层返回的数据集至数据分析引擎中，执行算法存储中的算法包代码，完成数据分析计算，并将分析结果根据输出参数约束返回到界面层。

[权利要求 5] 根据权利要求1所述的一种存算显全局可配置的数据分析软件架构设计方法，其特征在于：所述界面层通过调用显示接口访问插件层；当软件架构使用人员访问可视化界面时，界面调用显示接口获取插件层解析的所有算法存储配置信息，构建动态的算法选择界面，软件架构使用人员从该界面选择算法，输入算法输入参数；同时调用显示接口获取插件层解析的所有数据存储配置信息，构建动态的数据集选择界面，软件架构使用人员从该界面选择数据集；软件架构使用人员获得的分析结果的可视化界面也由显示接口获取的算法输出参数约束为基础构建，结果的形式包括图像、表格、文字和分析过程日志。

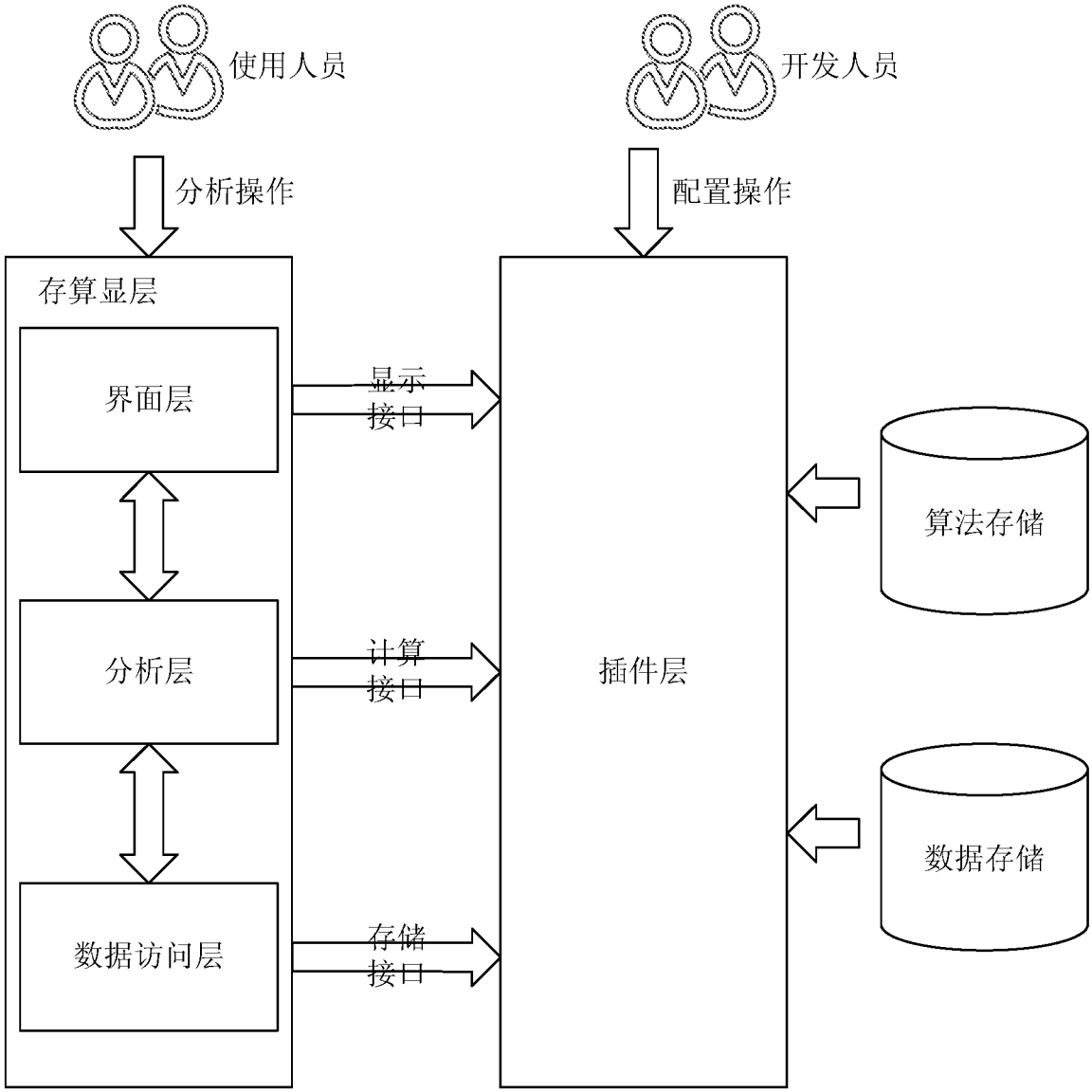
使用人员
开发人员
分析操作
配置操作
存算显层
界面层
分析层
数据访问层
显示
接口
计算
接口
存储
接口
插件层
算法存储
数据存储

图 1

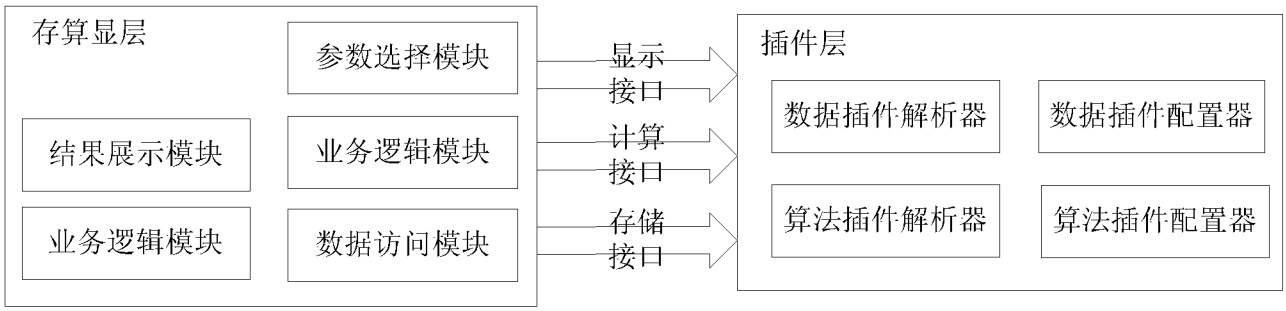
存算显层
参数选择模块
结果展示模块
业务逻辑模块
数据访问模块
显示
接口
计算
接口
存储
接口
插件层
数据插件解析器
数据插件配置器
算法插件解析器
算法插件配置器

图 2

```
<dataset>
    <category id="21" name="操业参数" file="Operating_parameters.txt" dataFrom="2011-01-01" dataTo="2017-12-31">
        <attributes>
            <attribute id="21_1" name="P0" col="1" type="double" />
            <attribute id="21_2" name="料批" col="2" type="double" />
            <attribute id="21_3" name="BV" col="3" type="double" />
            <attribute id="21_4" name="O2" col="4" type="double" />
            <attribute id="21_5" name="BP" col="5" type="double" />
            <attribute id="21_6" name="K值" col="6" type="double" />
            <attribute id="21_7" name="TP" col="7" type="double" />
            <attribute id="21_8" name="BT" col="8" type="double" />
            <attribute id="21_9" name="BH" col="9" type="double" />
            <attribute id="21_10" name="n_co" col="10" type="double" />
        </attributes>
    </category>
</dataset>
```

图 3

```
<RPackages>
    <RPackage index="3" name="数据挖掘">
        <approach index="3_5" name="时间序列分析">
            <algorithms>
                <algorithm index="3_5_1" name="单变量指数预测模型" method="TimeSeries" fileLocation="TimeSeries.R" libs="lubridate" ChartType="plot">
                    <parameters>
                        <parameter index="3_5_1_1" name="" type="attr_limit" value="属性只能选择一个">
                            <options>
                                <option index="3_5_1_1_1" name="" type="integer" value="=,1" />
                            </options>
                        </parameter>
                        <parameter index="3_5_1_2" name="预测月数" type="input" value="数量至少为1，必须填写">
                            <options>
                                <option index="3_5_1_2_1" name="" type="null_limit" value="true" />
                                <option index="3_5_1_2_2" name="" type="int_limit" value=">=,1" />
                            </options>
                        </parameter>
                    </parameters>
                </algorithm>
                <algorithm index="3_5_2" name="单变量ARIMA预测模型" method="ARIMATimeSeries" fileLocation="ARIMATimeSeries.R" libs="tseries,forecast" ChartType="plot">
                    <parameters>
                        <parameter index="3_5_2_1" name="" type="attr_limit" value="属性只能选择一个">
                            <options>
                                <option index="3_5_2_1_1" name="" type="integer" value="=,1" />
                            </options>
                        </parameter>
                        <parameter index="3_5_2_2" name="预测月数" type="input" value="数量至少为1，必须填写">
                            <options>
                                <option index="3_5_2_2_1" name="" type="null_limit" value="true" />
                                <option index="3_5_2_2_2" name="" type="int_limit" value=">=,1" />
                            </options>
                        </parameter>
                    </parameters>
                </algorithm>
            </algorithms>
        </approach>
    </RPackage>
</RPackages>
```

图 4

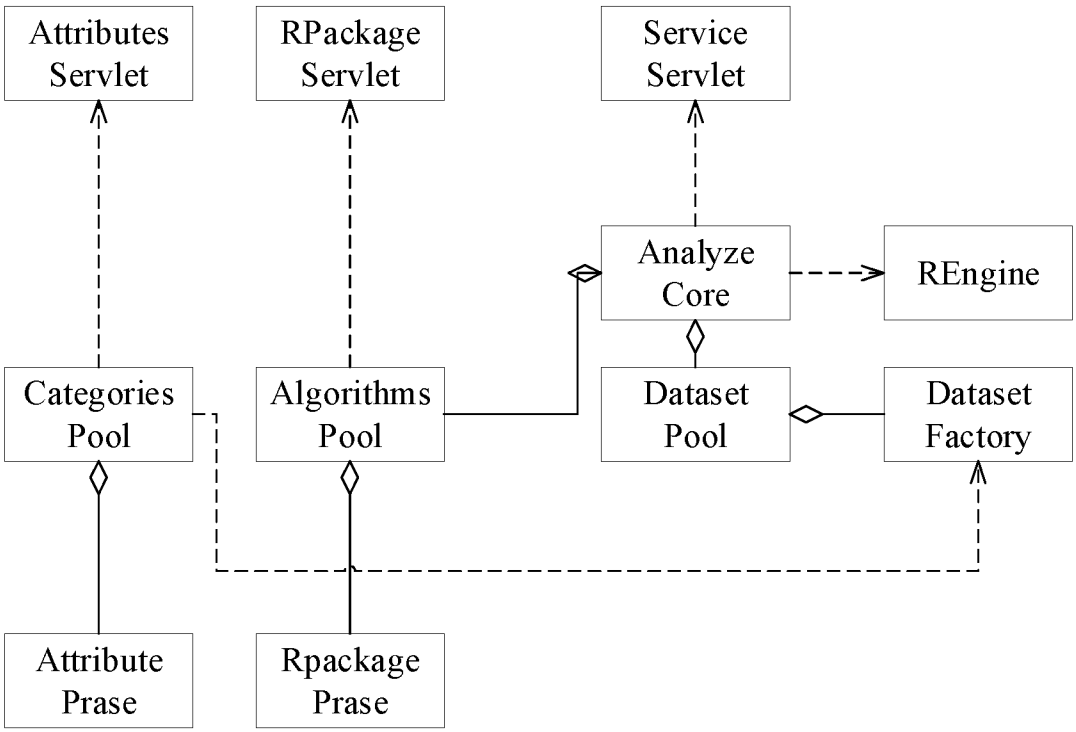
Attributes Servlet
RPackage Servlet
Service Servlet
Analyze Core
REngine
Categories Pool
Algorithms Pool
Dataset Pool
Dataset Factory
Attribute Prase
Rpackage Prase

图 5

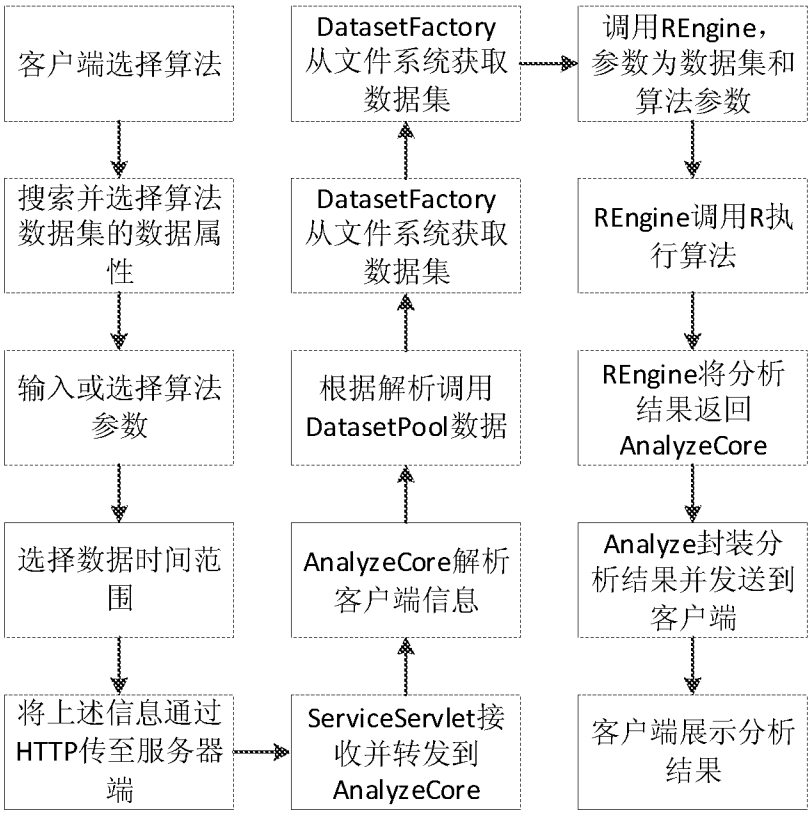
客户端选择算法
搜索并选择算法数据集的数据属性
输入或选择算法参数
选择数据时间范围
将上述信息通过HTTP传至服务器端
ServiceServlet接收并转发到AnalyzeCore
AnalyzeCore解析客户端信息
根据解析调用DatasetPool数据
DatasetFactory从文件系统获取数据集
DatasetFactory从文件系统获取数据集
调用REngine，参数为数据集和算法参数
REngine调用R执行算法
REngine将分析结果返回AnalyzeCore
Analyze封装分析结果并发送到客户端
客户端展示分析结果

图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/087192

A. CLASSIFICATION OF SUBJECT MATTER

G06F 8/20(2018.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

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

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

CNPAT, CNKI, WPI, EPODOC, IEEE: 界面层, 访问层, 分析层, 逻辑层, 展示层, 存储层, 处理层, 插件, 应用层, 数据, 展现层, 数据层, 表示层, 架构, 框架, 扩展, software, application, framework, architecture, layer, plug in

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	陈玉坤 (CHEN, Yukun). "中国优秀硕士学位论文全文数据库 信息科技辑 (China Master's Theses Full-Text Database, Information Science & Technology)" *基于插件架构的数据挖掘系统平台研究与设计 (Research and design of a data mining system platform based on plugin architecture),* 15 October 2009 (2009-10-15), ISSN: 1674-0246, chapters 4.1.2, 4.3.1, 5.1-5.2, pages 29-30, 35, 37-39	1-5
A	WO 2015100646 A1 (BEIJING FEINNO COMMUNICATION TECH CO., LTD.) 09 July 2015 (2015-07-09) entire document	1-5
A	CN 109325620 A (SHANGHAI MUBAI NETWORK TECHNOLOGY CO., LTD.) 12 February 2019 (2019-02-12) entire document	1-5
A	US 2017371647 A1 (QUARKDATA SOFTWARE CO., LTD) 28 December 2017 (2017-12-28) entire document	1-5

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 January 2020	**23 January 2020**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/087192

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
WO 2015100646 A1	09 July 2015	CN 105745620 A	06 July 2016
CN 109325620 A	12 February 2019	None	
US 2017371647 A1	28 December 2017	CN 105959302 A	21 September 2016

国际检索报告

国际申请号
PCT/CN2019/087192

A. 主题的分类

G06F 8/20(2018.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNPAT,CNKI,WPI,EPODOC,IEEE:界面层，访问层，分析层，逻辑层，展示层，存储层，处理层，插件，应用层，数据，展现层，数据层，表示层，架构，框架，扩展，software, application, framework, architecture, layer, plug in

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
X	陈玉坤. ″“中国优秀硕士学位论文全文数据库 信息科技辑”″ 基于插件架构的数据挖掘系统平台研究与设计, 2009年 10月 15日 (2009-10-15), ISSN：1674-0246, 第4.1.2、4.3.1、5.1-5.2章节，第29-30、35、37-39页	1-5
A	WO 2015100646 A1 (北京新媒传信科技有限公司) 2015年 7月 9日 (2015-07-09) 全文	1-5
A	CN 109325620 A (上海木白网络科技有限公司) 2019年 2月 12日 (2019-02-12) 全文	1-5
A	US 2017371647 A1 (QUARKDATA SOFTWARE CO., LTD) 2017年 12月 28日 (2017-12-28) 全文	1-5

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2020年 1月 14日	2020年 1月 23日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 孙国辉 电话号码 86-(10)-53961538

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/087192

检索报告引用的专利文件	公布日 (年/月/日)	同族专利	公布日 (年/月/日)
WO 2015100646 A1	2015年 7月 9日	CN 105745620 A	2016年 7月 6日
CN 109325620 A	2019年 2月 12日	无	
US 2017371647 A1	2017年 12月 28日	CN 105959302 A	2016年 9月 21日