The Interactive Analysis Method of Industrial Control Big Data Based on DAG and RDD*

ZHANG Yongqing*, ZHANG Huazheng
(Jicheng Electronics Co., Ltd., Jinan Shandong 250100, China)

Abstract: In order to meet the rapidly changing needs for the analysis of industrial control big data, a technical method to build the interactive analysis system for the industrial control big data based on DAG and RDD is proposed. DAG is used as the model to describe data calculation process, and the interactive analysis system for industrial control data is realized by constructing the calculation engine based on DAG and RDD. The practice shows that, using the system, data analysts can carry out data statistical analysis and exploration interactively by themselves. Compared with existing methods like SQL, BI, or Spark programming, the approach utilizing visual DAG modeling coupled with Spark RDD execution eliminates the need for users to learn SQL scripting, understand data storage structures, or master programming languages. Users can freely query and analyze data through drag-and-drop operations based on their business needs, achieving a "what-you-see-is-what-you-get" experience. The big data analysis system developed with this method enables business personnel to perform data analysis independently. It also plays an active role in driving business innovation through data mining and analysis.

Key words: interactive data analysis; big data; DAG; RDD
EEACC: 0290Z **doi**: 10.3969/j.issn.1005-9490.2026.03.021

基于 DAG 与 RDD 的工控大数据交互式挖掘分析方法*

张永庆*, 张华正
(积成电子股份有限公司, 山东 济南 250100)

摘 要: 为了满足对工控大数据的自主分析计算需求, 提出了一种基于有向无环图与弹性分布式数据集的工控大数据交互式分析方法。根据该方法采用可视化绘制 DAG 图构建数据计算过程模型, 在此基础上通过基于 DAG 与 RDD 的数据分析计算引擎, 实现交互式数据分析功能。实践表明, 根据该分析方法构建的交互式数据分析系统, 让数据分析人员可以通过可视化绘制数据计算流程图的方式自助进行数据统计与探查分析。与现有的数据分析方法相比, 采用可视化 DAG 建模+Spark RDD 执行的分析方法, 既不需要用户学习 SQL 脚本的编写方法, 也不需要用户了解数据的存储方式与结构, 更不需要学习编程语言, 即可按照用户自己的分析需求通过拖拽的方式自由查询与分析, 所见即所得, 显著降低数据分析操作难度, 减少学习成本并提高工作效率。基于该方法构建的大数据交互式分析系统可以满足普通业务人员自主进行数据分析计算的需要, 同时也可在通过数据挖掘分析驱动业务创新等方面发挥积极作用。

关键词: 交互式数据分析; 大数据; DAG; RDD
中图分类号: TP391 **文献标识码**: A **文章编号**: 1005-9490(2026)03-0626-07

随着工业互联网技术的发展, 工业控制(工控)行业的数字化程度越来越高, 在实际的运行和生产过程中, 产生了庞大而繁杂多样的数据信息[1-5], 引入"工业大数据"技术, 通过对大数据的挖掘分析发现内在的规律和隐藏在海量数据中的知识, 提升自动化控制智能化水平, 在企业运营、业务模式等方面实现创新, 是企业数字化转型的迫切需求[2,7]。当前, 大数据分析[2,6,20]技术在工控行业各个领域已经有了大量的成功应用案例, 如设备故障预测、设备运行管理、客户服务等[3]。而这些大数据分析应用通常是针对特定的业务目的采用专用的算法开发而成[4], 难以满足业务人员对数据的自主探索分析需求, 而且由于业务需求变化迅速, 专用分析应用也难以满足复杂多变的分析需求。因此, 面向企业的数据分析人员, 提供方便易用的交互式数据分析工具, 实现对大数据的自主分析探索具有重要的意义。

在大数据分析方面, 基于 Spark 编程的方式实现数据查询并进行各种计算处理, 最终得出所需要的分析结果, 是一种常见的实现方式[9-10]。但是这种分析方法既要求对大数据的存储结构非常了解, 还需要熟练掌握 Spark 编程技术, 对于一般的业务人员来讲难度很高无法掌握, 他们需要方便易用的

项目来源: 公司产品研发计划项目
收稿日期: 2025-09-26 **修改日期**: 2025-11-12

交互式数据分析工具。当前,常见的交互式数据分析工具主要有两种类型,一种是基于结构化查询语言(Structured Query Language,SQL)的数据查询工具[12],另一种是商业智能(Business Intelligence,BI)软件。基于 SQL 的数据库查询工具可以支持数据查询人员通过编辑并执行 SQL 脚本的方式进行交互式数据查询,但是此类工具需要数据分析人员全面了解数据库表并熟练掌握 SQL 语言,难度较大。一般的业务人员通常不懂数据库技术,既不了解数据在数据库中是如何存储的,也不会用 SQL 编写数据查询与分析的脚本,因此基于 SQL 的数据库查询工具无法满足业务人员自主进行数据分析的需求。BI 软件虽然支持数据分析人员通过鼠标拖拽的方式选择需要的主题表、维度和指标,采用多种展示形式对数据进行多维度分析[8],但是存在以下几个问题:首先,需要预先通过编写 SQL 的方式创建数据集,而编写 SQL 难度较大,普通用户一般很难掌握,因此需要运维人员预先把常用的数据集创建出来,但是在遇到新的分析需求时往往需要等待运维人员先创建数据集再进行数据分析,需求响应比较慢。其次,BI 软件的定位主要面向数据的可视化展示与报表定制,难以实现复杂的数据计算逻辑,无法满足业务分析人员复杂多变的数据探索与计算需要。最后,BI 软件的拖拽式分析功能是面向数据集或数据表的,不是面向业务对象的,难以描述业务对象与数据之间的复杂关联关系,操作比较繁琐,不够直观,一般的用户学习比较困难,通常需要由运维人员预先通过利用 BI 将数据查询与分析页面定制好,用户利用预制的数据分析页面进行查询或分析,而预制的分析页面难以满足突发性、急迫性的复杂多变的分析需求,因此,BI 软件也无法满足用户自主进行数据查询与分析的需要。

有向无环图(Directed Acyclic Graph,DAG)是描述系统进行过程的有效工具,在拓扑排序、关键路径计算、任务调度、大数据、区块链等领域应用广泛[11]。Spark 是 UC Berkeley AMPLab 所开源的内存计算框架,其核心是弹性分布式数据集(Resilient Distributed Dataset,RDD)与 DAG 优化技术。是当前应用最为广泛的大数据计算框架,在提高大数据计算性能方面具有明显的优势,尤其适用于计算密集型、交互式、迭代式计算的场景[12,19]。文献[13]、[17]、[18]对基于 DAG 模型的高效并行任务调度算法进行了研究,文献[14]和文献[15]分别提出了分布式系统下以及云计算环境下的基于 DAG 图的任务调度策略。文献[16]提出了一种基于 DAG 模型描述数据计算过程的思路。在借鉴上述研究成果的基础上,本文提出了一种基于 DAG 与 RDD 的工控大数据交互式分析方法。根据该分析方法,采用 DAG 作为描述数据计算过程的模型,通过设计可视化的数据分析界面,提供对象化数据查询节点、各种原子化的数据处理节点和各种结果输出节点,用户可以通过可视化拖拽的方式选择各种节点并绘制成数据分析 DAG 图,实现数据查询与计算流程的自由编排,满足数据查询与分析的需要。同时,基于 DAG 与 RDD 实现计算引擎,解决了自主进行大数据分析计算的灵活性和效率问题。利用此方法开发的大数据交互式分析系统,计算灵活,操作简单,用户既不需要了解数据的存储方式与结构,也不需要学习 SQL 或 Spark 编程语言,可由数据分析人员利用可视化分析界面完全自主地进行交互式数据查询与挖掘分析,有效降低数据分析的门槛,提高数据分析效率。

1　需求场景及解决方案

普通业务人员自主进行大数据分析面临的主要困难包括:①不了解数据的底层存储细节;②不掌握 SQL 数据查询语言;③数据计算逻辑比较复杂,难以自由定义计算过程;④等待时间长,交互性差。因此,面向普通业务人员的交互式数据分析系统应满足以下要求:①提供方便的数据选择功能,屏蔽底层数据库细节,不需要分析人员了解数据在数据库中的存储位置和存储格式。②提供灵活的数据计算过程定义工具,让数据分析人员可通过简单操作实现对计算过程的自由编排。③应提供丰富的算子,且可以不断扩展,以满足业务数据分析人员常用的数据分析需求。④应提供足够的计算性能,以满足交互式分析对响应速度的要求。⑤计算状态与计算结果应可以即席查看。

为了满足上述需求,提出如下实现方法:①通过建立统一元数据,实现对异构大数据的业务对象化描述,屏蔽底层数据库细节,以解决业务数据分析人员查询数据方便性的问题。②采用 DAG 模型描述数据计算处理过程,并提供可视化的数据计算过程建模工具,业务数据分析人员通过鼠标拖拽的方式选择需要的数据和算子,采用绘制 DAG 图的方式自由定义数据计算过程。③通过对各种常用的数据分析计算过程进行分解、归纳、提炼、抽象,创建丰富的算子,以满足常用的业务数据统计分析需要,并通过算子注册管理功能满足算子的自由扩展需要。④计算过程中采用 RDD 实现步骤间中间结果集传递,基

于 Spark 内存计算框架实现高效的 DAG 图调度计算引擎,以满足计算性能的要求。

2 关键技术

2.1 基于 DAG 图的数据计算过程模型

2.1.1 基于 DAG 图的数据计算过程描述原理

有向图是指由节点以及节点之间的有向边组成的图结构,DAG 图是指没有环的有向图。

对于数据计算处理过程而言,主要涉及通过查询生成数据集以及对数据集进行的一系列转换计算,且一般不存在数据计算后回到原点的情况。因此,本文采用 DAG 图对数据计算过程进行建模是可行的。

①DAG 图

A. DAG 图定义:将一个数据计算处理过程用一个 DAG 图表示,记为:$G(V,E)$。

其中,V 为图中的顶点集合,记为 $V=\{v_i\}$,顶点 v_i 表示数据计算处理过程中第 i 个步骤,如获取数据集或对数据集进行的各种转换处理等。根据其作用的不同,顶点分为数据节点、算子节点两种类型,即:①数据节点:是指数据查询节点,通过查询数据库或文件生成数据集并向后输出。②算子节点:是指数据处理节点,对输入数据集进行计算处理,计算结果以新的数据集输出。E 为图中边的集合,记为 $E=\{e_{ij}\}$,边 e_{ij} 表示自第 i 个步骤流出并流入到第 j 个步骤的数据流。

B. 节点的入度与出度:以某节点为头,终止于该节点的边的数目称为该顶点的入度,以某节点为尾,起始于该节点的边的数目称为该节点的出度。

②约束条件

根据数据分析计算处理过程的现实情况,数据计算 DAG 图应满足如下约束条件:①至少存在 1 个入度为 0 的节点。②至少存在 1 个出度为 0 的节点。③不存在入度为 0 的算子节点。④当且仅当节点数为 1 时,可存在孤立节点,且节点类型必须为数据节点。

例:设备最高负载率的计算过程由以下步骤组成:

步骤 1:从设备描述表中读取设备的额定容量。

步骤 2:从设备的量测历史表中读取用电负荷的有功、无功功率曲线数据。

步骤 3:将上述 2 个数据集,按照设备标识关键字进行连接。

步骤 4:针对步骤 3 得到的数据集中每个设备各时间点的量测数据,利用有功、无功功率和额定容量计算各时间点的负载率。

步骤 5:针对步骤 4 得到的数据集,按照设备标识分组,计算其负载率的最大值。

上述计算过程可以用图 1 的 DAG 图表示:

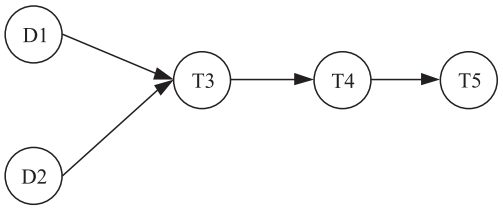

图 1 数据计算流程 DAG 图示例

图 1 中,D1、D2 表示数据节点,分别描述上述的步骤 1、步骤 2,T3、T4、T5 为算子节点,分别描述上述的步骤 3 至步骤 5 的数据计算过程。图中带箭头的连线,为 DAG 图的边,表示数据流动的方向,从流出节点指向流入节点。

2.1.2 数据计算过程的 DAG 模型

数据计算过程的 DAG 模型采用 XML 文件进行描述,包含其中的数据节点、算子节点、拓扑连接关系的描述。下面对 DAG 图以及其中的节点、节点间连接关系模型的描述信息进行说明。

①DAG 图

DAG 模型描述的信息包括 DAG 图形的标识、名称、功能说明、创建时间以及图中包含的节点列表和节点间拓扑连接关系信息,如表 1 所示:

②数据节点

基于统一元数据实现对象化数据查询节点,使数据分析人员能够仅根据所掌握的业务知识,即可选择所需要的数据对象。数据节点模型描述所查询的数据对象以及过滤条件,描述的信息如表 2 所示:

③算子节点

算子节点对输入数据集进行计算处理,其计算结果以新的数据集输出。模型描述信息如表 3 所示:

④连接关系

连接关系描述节点间的连接与数据流向,采用邻接表的形式描述,描述信息如表 4 所示:

表 1 DAG 图模型

属性名称	含义
ID	节点标识
Name	名称
Description	功能描述
CreateTime	创建时间
NodeList	节点列表
ConnectionList	连接关系列表

表 2　数据节点模型

属性名称	含义
ID	节点标识
Name	名称
ObjClass	实体对象类型标识
ObjPropList	实体对象属性列表
ObjFliter	实体对象过滤条件
DataObjType	数据对象类型
DataObjPropList	数据对象属性列表
DataObjFliter	数据对象过滤条件
Coordinate	在图中的坐标

表 3　算子节点模型

属性名称	含义
ID	节点标识
Name	名称
Alias	别名
Operatorype	算子类型标识
CalcType	计算类型 (单记录/数据集/全部数据)
Description	功能描述
Implementationclass	实现类
ResourceLocation	资源存储位置 (jar 文件在 HDFS 的存储路径)
InputPorts	输入端口配置信息
OutPorts	输出端口配置信息
ParamLists	配置参数列表
Coordinate	在图中的坐标

表 4　节点连接模型

属性名称	含义
FromNode	起点 Node 标识
ToNode	终点 Node 标识

2.2　数据计算过程可视化建模

让数据分析人员根据数据分析计算的需要,自由选择所需要的数据查询节点与算子节点,并按照计算逻辑的需要进行编排,生成数据计算 DAG 图文件,是实现交互式数据分析的关键途径。

DAG 图可视化绘制工具由前端界面和后端服务构成。前端界面实现数据 DAG 图的绘制与展示,后端业务逻辑接口服务实现数据流图模型的保存与查询,采用 Java 和 SpringBoot 等技术实现。前后端之间通过 Ajax 实现信息交互。

2.2.1　数据计算过程 DAG 图建模过程

①构建数据流图对象与画布:创建画布,并建立存放数据流图对象的模型数据,添加对拖拽、单击、双击、移动等各种操作事件的监听方法。

②拖拽式绘制节点:从节点树拖拽数据节点或算子节点到画布上释放时,通过拖放事件监听方法,根据拖拽的节点类型,绘制节点的图标,同时添加节点数据对象,自动设置节点数据对象与坐标、图标形状、背景色,并设置节点的事件响应函数。

③节点信息设置:通过节点的双击事件响应函数,调出节点的设置界面,设置查询节点的查询条件或数据处理节点的处理逻辑。设置的信息保存在节点的模型对象中。

④绘制连接线:在一个节点图标上按住鼠标左键并移动到另一个节点图标上方释放,即触发连线响应函数,从起始节点向终止节点绘制一条连接线,箭头指向终止节点,同时生成一个连接关系数据对象。

⑤数据流图对象存储:在绘制过程中,数据流图中的节点与连接关系数据暂存在内存数据对象中,执行保存操作后,存储到数据库表中。

2.2.2　节点设置界面

①数据节点设置界面:选择对象类型,根据该类对象的关联对象、对象属性以及有关的数据对象等元数据信息,生成该对象的查询节点界面,支持按照各级上层关联对象以及本级对象属性过滤查询工控对象与属性、选择数据对象及属性并设置过滤条件。

②算子节点设置界面:根据算子类型注册信息所描述的算子节点的参数列表以及指定的设置界面类型,生成算子节点设置界面。

2.3　数据计算 DAG 图计算引擎

DAG 图计算引擎由 DAG 图执行服务和服务接口构成。服务接口负责与前端交互,采用 Restful 接口实现,包括执行计算、执行状态查询、计算结果查询等接口方法。执行服务由 DAG 图调度服务和各种节点执行器构成,负责根据接收的执行命令,对前端提交的 DAG 图 xml 文件进行解析、校验并转成可以由 Spark 执行的 SparkJob,提交 Spark 执行。节点间采用 RDD 实现数据流的传输。最后计算结果与计算过程状态缓存在 Redis 中。

2.3.1　DAG 图调度执行服务

①接收执行命令,解析 DAG 图的配置信息。

②对 DAG 图的合法性进行校验。若存在错误,则将执行状态置为错误并将状态与错误信息写入 Redis,执行结束。

③若 DAG 图校验合法,则对 DAG 图的节点进行排序。

④根据各节点的类型与配置信息,创建每个节点的执行器、状态监听器、数据行监听器。

⑤生成每个节点的结果集,并根据边的设置信息将输出结果集与下级节点的数据行监听器绑定。

⑥按照节点排序结果,依次启动各节点执行器,并将执行状态置为正在执行。

⑦利用节点状态监听器,监听每个节点的执行状态,并将执行状态与日志信息输出至 Redis。

⑧判断是否有节点出现严重错误,若出现严重错误,则终止所有节点执行线程,将执行状态置为错误等并结束。否则判断各节点执行线程是否全部完成,若所有节点执行完毕,则将最后 1 个节点的输出结果集写入到 Redis,将执行状态置为完成并结束。

上述执行过程可用图 2 表示:

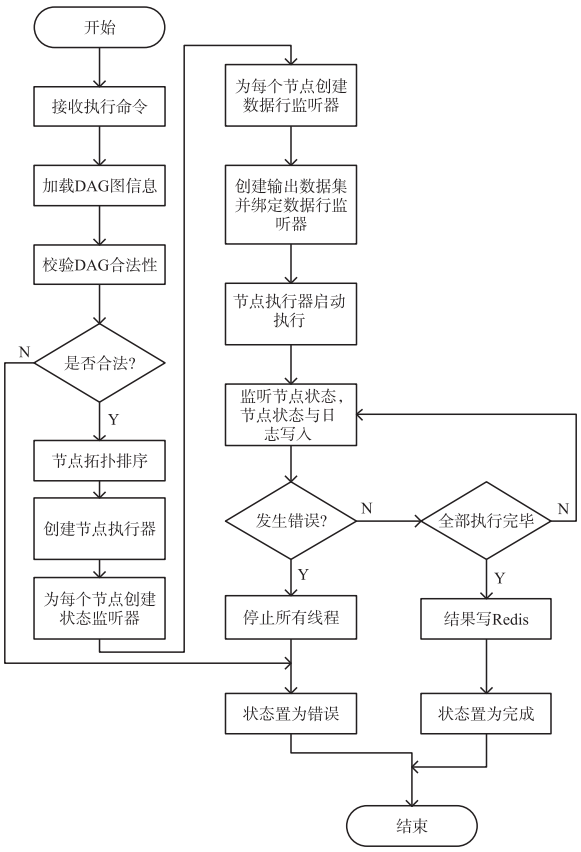

图 2 DAG 图调度执行过程

2.3.2 数据节点执行器

数据节点执行器解析查询节点配置信息,根据所查询对象的类型标识与属性集、对象查询过滤条件、查询的数据对象及查询过滤条件的组合,通过元数据解析,确定数据库类型与位置,生成数据库所支持的查询命令,根据查询命令生成 SparkJob,提交 Spark 执行,实现数据查询。数据查询节点执行器的主要执行过程如图 3 所示:

2.3.3 算子节点执行器

算子节点执行器由计算逻辑执行线程、输入数据行监听器、结果输出器构成。输入数据行监听器负责监听输入数据流的变化,计算逻辑执行线程负责根据设置的计算逻辑,对新的输入数据进行计算,结果输出器负责按照设置,将计算结果生成新的数据流,向后传递。算子节点的主要执行过程如图 4 所示:

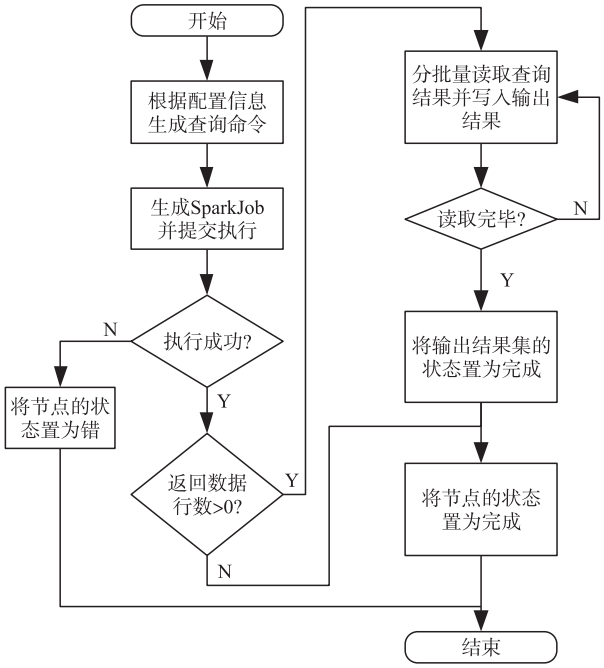

图 3 数据查询节点执行过程

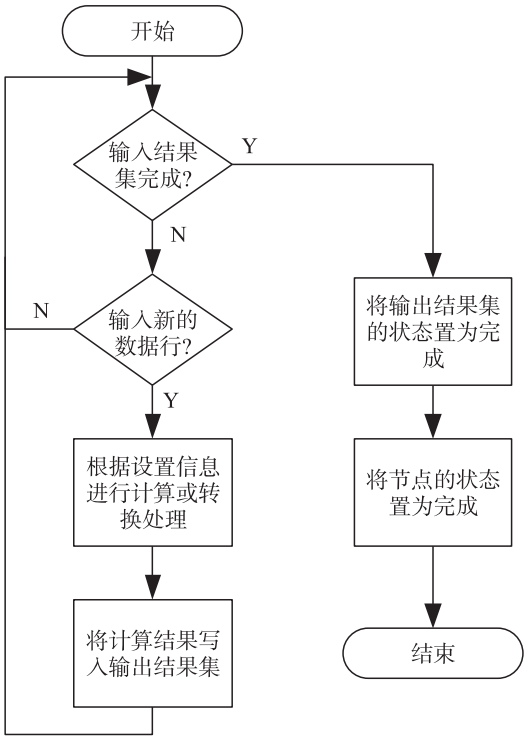

图 4 算子节点执行过程

2.3.4 RDD 在 DAG 数据流图执行过程中的状态与传递过程

在执行数据查询节点时,解析节点的配置信息并生成数据库查询命令,据此创建弹性分布式数据集 RDD。当一个查询节点涉及到多个异构数据库查询时,生成多个 RDD。查询节点生成的 RDD 作

为后面数据处理节点的输入。

当后面有连接的数据处理节点时,根据数据处理节点的配置信息生成相应的数据转换操作命令,在由前面传入的 RDD 上执行转换操作,生成新的 RDD。例如,对于数据过滤节点,执行 Filter( )转换操作,对于计算器节点,执行 Map( )操作,对于数据集连接,执行 Join( )操作,对于分组统计节点、越限统计节点、极值统计节点,执行 Groupbykey( )操作。数据处理节点执行得到的 RDD 作为后面的数据处理节点或输出节点的输入。

当执行到表输出、E 格式文件输出、Excel 文件输出等输出节点时,触发 RDD 行动操作,开始执行对应的查询、转换计算并将结果输出到指定位置。如 SaveToTable( )、SaveToEFile( )、SaveToExcelFile( )。当执行到图表输出节点或数据流图最后一个节点时,触发 RDD 执行 SaveToRedis( )操作,将结果数据输出到 Redis 中,以便前端页面从缓存中分页获取数据。

3　基于 DAG 与 RDD 的大数据交互式分析方法应用实践

3.1　系统实施方案

根据此分析方法,以某大型工业控制生产企业的工控大数据平台为基础设计开发了工控大数据交互式分析系统,并已投入使用,在工控数据综合分析应用与共享服务方面发挥了重要作用。

在该项目中,构建了 8 个物理节点的 Hadoop 大数据平台,汇集了 120 TB 的运行数据,并在此基础上开发了基于 DAG 与 RDD 的交互式分析系统。为了满足行业内常用的数据统计分析需要,系统预置了一组算子,如表 5 所示:

表 5　预置算子节点

名称	功能
字段计算器	定义对一个数据集中字段间的计算公式,计算结果生成新的字段
记录排序	设置对一个数据集按照指定字段进行排序
字段选择	从一个数据集中选择所需要的字段集合
列转行	按照关键字,将一条记录中的多个字段转换成多条记录中的数据
转行列	按照关键字,将多条记录中指定字段的值转换为新的数据集中多个列
数据集连接	将 2 个数据集进行连接
数据集合并	将 2 个数据集合并为 1 个数据集
数据集差集	从 1 个数据集中删除与另一个数据集中的关键字相同的集合
条件分组统计	按照设定的过滤条件和分组维度字段,分组统计
极值统计	按设定的维度对指定字段统计最大最小值
越限统计	按设定的限值统计指定字段的越限次数与越限比例

利用此交互式数据分析系统,业务分析人员可以采用自由绘制数据计算流程图的方式对数据进行自助式探索分析。交互式数据分析界面如图 5 所示:

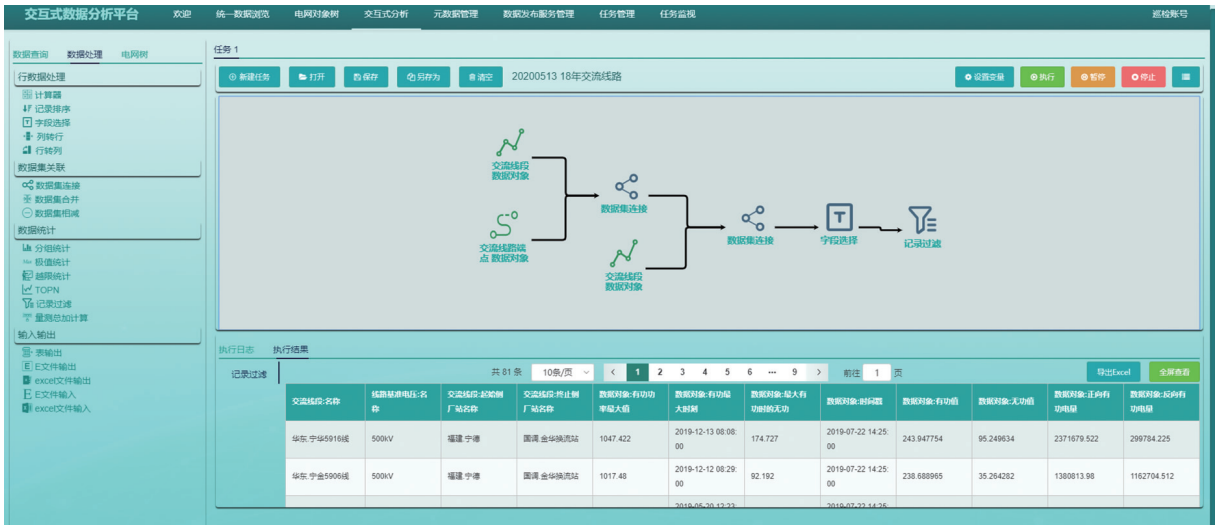

图 5　交互式数据分析界面示例

3.2　应用效果

通过提供基于 DAG 图的数据计算过程建模工具,让业务分析人员经过简单培训后可以根据自己的需求自由定义数据计算过程,满足自主进行数据分析的需求。性能测试结果表明,从存储 80 亿条数据的表中抽取 3 亿条数据进行交互式计算统计分析,耗时约 120 s,对 200 万条数据进行分析计算,耗时约 8 s,基本可以满足交互式数据分析的需要。

从试点系统的应用效果来看,仍有些需要解决的问题。一是数据计算的流程需要数据分析人员自

己根据计算要求分解并定义,对于普通的业务分析人员仍具有一定的难度,需要提供更加智能化的解决方案。二是系统预置的算子需要足够丰富,否则数据分析人员在做数据分析时,会遇到无法满足要求的问题,影响使用效果,需要对行业的数据分析需求进行尽可能全面的梳理分析并完善预置算子库。最后,在大数据量计算的情况下,等待的时间还比较长,影响用户体验,需要进一步提高计算速度。后续还需要对分析方法进行进一步的优化改进,以提高其易用性与实用性。

4 结语

针对业务分析人员自主进行大数据分析计算的需求,提出了基于 DAG 与 RDD 的工控大数据交互式分析方法,并据此方法实现了工控大数据交互式分析系统。实际应用效果表明,该交互式分析方法可以解决传统数据统计分析应用存在的不足,在保证数据分析计算灵活性与性能的同时降低数据分析操作难度,满足普通业务人员自主进行数据分析计算的需要,在通过数据分析驱动业务创新等方面可发挥积极作用。但是,对于普通业务分析人员来讲,根据数据分析需求绘制成数据计算流程 DAG 图仍具有一定的难度。未来,通过引入人工智能与自然语言处理技术,实现基于语音或自然语言文本的智能化交互式数据分析将是一个值得深入研究的方向。

参考文献:

[1] 杨艳华,姚立纲. 基于时间递推建模及交叉熵算法求解柔性作业车间调度问题[J]. 计算机集成制造系统,2021,27(6):1703-1713.

[2] 林静怀. 基于大数据平台的电网运行指标统一管控方案[J]. 电力系统保护与控制,2018,46(4):165-170.

[3] 林静怀. 基于云计算的电网调度控制培训仿真系统设计[J]. 电力系统自动化,2017,41(14):164-170.

[4] 陈郑平,梅峥,林静怀,等. 面向调控业务的多级一体化数据交换系统设计与实现[J]. 电力系统自动化,2020,44(4):195-201.

[5] 侯婷婷,李云婷. 中国制造业企业数字化转型内涵和方法浅析[J]. 科技与创新,2020(10):125-126.

[6] 杨艳华,姚立纲. 基于 SDG 简化模型的工控系统数据网故障溯源方法[J]. 山东大学学报(理学版),2020,55(11):46-57.

[7] 李晓云. 企业数字化转型路径初探[J]. 科学与信息化,2020(10):168-169.

[8] 梁建交. 工业大数据——制造企业数字化转型的重点方向[J]. 信息安全与通信保密 2020(4):72-81.

[9] 冷喜武,陈国平,白静洁,等. 智能电网监控运行大数据分析系统总体设计[J]. 电力系统自动化,2018,42(12):160-166.

[10] Silva J, Pineda Lezama O B, Romero L, et al. Risk Analysis of Using Big Data in Computer Sciences[J]. Procedia Computer Science,2019,160(1):532-537.

[11] Lin H Y, Yang S Y. A Cloud-based Energy Data Mining Information Agent System Based on Big Data Analysis Technology[J]. Microelectronics Reliability, 2019, 97(6):66-78.

[12] Mohammad H A, Vahid T V, Behnam B. Evaluation of big data frameworks for analysis of smart grids[J]. Journal of Big Data,2019,6(1):860-871.

[13] Wang K, Liu N, Sadooghi I, et al. Overcoming Hadoop Scaling Limitations through Distributed Task Execution[C]//2015 IEEE International Conference on CLUSTER Computing, Chicago, IL, USA, IEEE, 2015:236-245.

[14] 李聪颖,王瑞刚,梁小江. 基于 Hadoop 的交互式大数据分析查询处理方法[J]. 计算机技术与发展,2016,26(8):134-137,142.

[15] 基于 Power BI 工具的大数据分析在医疗设备多维数据分析中的应用研究[J]. 中国医学装备,2020,17(5):169-173.

[16] 程学旗,靳小龙,王元卓,等. 大数据系统和分析技术综述[J]. 软件学报,2014,25(9):1889-1908.

[17] 华强盛. 基于 DAG 模型的高效并行任务调度算法[D]. 长沙:中南大学,2004.

[18] 张银娟. 云计算环境下基于复杂 DAG 图的任务调度策略研究[D]. 扬州:扬州大学,2016.

[19] 李巧巧. 面向负载均衡的 Spark 任务划分与调度策略研究[D]. 长沙:湖南大学,2017.

[20] 殷荣. 基于 DAG 模型的离线数据处理引擎的设计与实现[D]. 哈尔滨:哈尔滨工业大学,2016.

[21] 周小明,陈郑平,陶蕾,等. 基于电力调控大数据平台的指标展示应用研究[J]. 电力信息与通信技术,2016,14(12):15-19.

张永庆(1970—),男,硕士,高级工程师,主要从事电力调度自动化、大数据与人工智能应用方向的技术研究、系统设计与开发工作,zhangyongqing@ ieslab.cn。