

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(43) 国际公布日
2017年1月19日 (19.01.2017)

WIPO | PCT

(10) 国际公布号
WO 2017/008604 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
 - (21) 国际申请号: PCT/CN2016/085487
 - (22) 国际申请日: 2016年6月12日 (12.06.2016)
 - (25) 申请语言: 中文
 - (26) 公布语言: 中文
 - (30) 优先权:
201510417386.0 2015年7月15日 (15.07.2015) CN
 - (71) 申请人: 中兴通讯股份有限公司 (ZTE CORPORATION) [CN/CN]; 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。
 - (72) 发明人: 刘丽霞 (LIU, Lixia); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦中兴通讯股份有限公司转交, Guangdong 518057 (CN)。文韬 (WEN, Tao); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦中兴通讯股份有限公司转交, Guangdong 518057 (CN)。王志坤 (WANG, Zhikun); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦中兴通讯股份有限公司转交, Guangdong 518057 (CN)。王东 (WANG, Dong); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦中兴通讯股份有限公司转交, Guangdong 518057 (CN)。刘海军 (LIU, Haijun); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦中兴通讯股份有限公司转交, Guangdong 518057 (CN)。
 - (74) 代理人: 北京安信方达知识产权代理有限公司 (AFD CHINA INTELLECTUAL PROPERTY LAW OFFICE); 中国北京市海淀区学清路8号B座1601A, Beijing 100192 (CN)。
 - (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
 - (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。
- 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))
 - 发明人资格(细则 4.17(iv))
- 本国际公布:
- 包括国际检索报告(条约第 21 条(3))。

(54) Title: DATA PROCESSING METHOD, DEVICE AND SYSTEM

(54) 发明名称: 数据处理方法、装置和系统

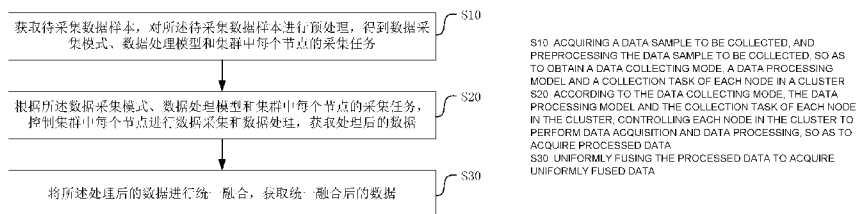


图 1

(57) Abstract: A data processing method, comprising: acquiring a data sample to be collected, and preprocessing the data sample to be collected, so as to obtain a data collecting mode, a data processing model and a collection task of each node in a cluster (S10); according to the data collecting mode, the data processing model and the collection task of each node in the cluster, controlling each node in the cluster to perform data acquisition and data processing, so as to acquire processed data (S20); and uniformly fusing the processed data to acquire uniformly fused data (S30). A data processing device and a data processing system.

(57) 摘要: 一种数据处理方法, 包括: 获取待采集数据样本, 对所述待采集数据样本进行预处理, 得到数据采集模式、数据处理模型和集群中每个节点的采集任务(S10); 根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务, 控制集群中每个节点进行数据采集和数据处理, 获取处理后的数据(S20); 将所述处理后的数据进行统一融合, 获取统一融合后的数据(S30)。一种数据处理装置和数据处理系统。



WO 2017/008604 A1

数据处理方法、装置和系统

技术领域

5 本发明实施例涉及但不限于计算机信息处理领域，尤其涉及一种数据处理方法、装置和系统。

背景技术

10 数据已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素，大数据时代已经降临。大数据的爆炸式增长在大容量、多样性和高增速方面，考验着现代企业的数据处理和分析能力，同时，也为企业带来了获取更丰富、更深入和更准确地洞察市场行为的大量机会。当前，最重要的现实是对大数据进行处理分析，只有通过处理分析才能获取很多智能的、深入的、有价值的信息。

15 相比于传统的数据，大数据呈现了 4V+1O 特征，表现为数据量大（Volume）、种类繁多（Variety）、时效性高（Velocity）、价值密度低（Value）和数据在线（Online）。相关技术中常规的大数据处理工具有两类：传统的 ETL（Extract-Transform-Load，抽取、转换、加载）工具和基于 Hadoop（Hadoop Distributed File System，分布式文件系统）的大数据采集工具。

20 用户对于大数据处理的高效性和融合的集成度要求越来越高。然而相关技术中常规的大数据处理工具对于大数据这种多源、异构、海量的数据，在处理方面存在欠缺，数据处理过程人工参与度高，数据处理不精细、效率低，而且对不同的数据也没有进行统一的融合，集成度不高，导致获取的最终数据可利用价值和实用性不高。

25 发明内容

以下是对本文详细描述的主题的概述。本概述并非是为了限制权利要求的保护范围。

本发明实施例可以解决大数据处理效率低、融合度低的技术问题。

为此，本发明实施例提供一种数据处理方法，所述数据处理方法包括以下步骤：

5 获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中每个节点的采集任务；

根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务，控制集群中每个节点进行数据采集和数据处理，获取处理后的数据；

将所述处理后的数据进行统一融合，获取统一融合后的数据。

10 可选的，其中，所述获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中每个节点的采集任务，包括：

控制所述集群中每个节点读取待采集数据，获取待采集数据样本和集群中所有节点的负载情况；

15 评估所述待采集数据样本，获取质量评估报告，所述质量评估报告包括：数据类型、数据编码、数据冗余率、数据稀疏性和数据传输速度；

根据所述质量评估报告和所述集群中所有节点的负载情况，适配数据采集模式和数据处理模型，分配所述集群中每个节点的采集任务。

20 可选的，其中，所述根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务，控制集群中每个节点进行数据采集和数据处理，获取处理后的数据，包括：

控制所述集群中每个节点根据对应的采集任务，以所述采集模式进行数据采集；

25 控制所述集群中每个节点根据所述数据处理模型处理采集到的数据，获取处理后的数据。

可选的，其中，所述将所述处理后的数据进行统一融合，获取统一融合后的数据，包括以下至少之一：

- 5 根据预设的数据规整模型，对处理后的数据进行数据规整，得到统一融合后的数据；根据预设的数据集成模型，对处理后的数据进行数据集成，得到统一融合后的数据；根据预设的数据建模模型，对处理后的数据进行数据建模，得到统一融合后的数据。

可选的，所述将所述处理后的数据进行统一融合，获取统一融合后的数据之后，所述方法还包括：

- 10 分类存储所述统一融合后的数据。

可选的，所述获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中每个节点的采集任务之前，所述方法还包括：

- 15 配置数据采集接口；

所述获取待采集数据样本，对所述待采集数据样本进行预处理的步骤包括：

通过所述数据采集接口获取待采集数据样本，对所述待采集数据样本进行预处理；

- 20 所述控制集群中每个节点进行数据采集和数据处理的步骤包括：

控制集群中每个节点通过所述数据采集接口进行数据采集，控制集群中每个节点对采集到的数据进行数据处理。

可选的，其中，所述配置数据采集接口，包括：

- 25 获取数据采集接口配置参数；

根据所述配置参数进行所述数据采集接口的连接测试；

若所述连接测试成功，则将包含有所述配置参数的数据采集接口配置文

件向集群中的其他节点进行分发，配置所述集群中的每个节点，并控制所述集群中的每个节点连接所述数据采集接口；

若所述连接测试不成功，则转入步骤：获取数据采集接口配置参数。

5 此外，本发明实施例还提供一种数据处理装置，包括：

适配模块，设置为获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中每个节点的采集任务；

数据处理模块，设置为根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务，控制集群中每个节点进行数据采集和数据处理，获取
10 处理后的数据；以及

统一融合模块，设置为将所述处理后的数据进行统一融合，获取统一融合后的数据。

可选的，其中，所述适配模块包括：

15 样本采集单元，设置为控制所述集群中每个节点读取待采集数据，获取待采集数据样本和集群中所有节点的负载情况；

数据评估单元，设置为评估待采集数据样本，获取质量评估报告，所述质量评估报告包括：数据类型、数据编码、数据冗余率、数据稀疏性和数据传输速度；以及

20 适配单元，设置为根据所述质量评估报告和所述集群中所有节点的负载情况，适配数据采集模式和数据处理模型，分配所述集群中每个节点的采集任务。

可选的，其中，所述数据处理模块包括：

25 数据采集单元，设置为控制所述集群中每个节点根据对应的采集任务，以所述采集模式进行数据采集；以及

数据处理单元，设置为控制所述集群中每个节点根据所述数据处理模型

处理所述采集到的数据，获取处理后的数据。

可选的，其中，所述统一融合模块包括数据获取单元和以下一个或多个单元：数据规整单元、数据集成单元、数据建模单元，其中：

5 所述数据规整单元，设置为根据预设的数据规整模型，对处理后的数据进行数据规整；

所述数据集成单元，设置为根据预设的数据集成模型，对处理后的数据进行数据集成；

10 所述数据建模单元，设置为根据预设的数据建模模型，对处理后的数据进行数据建模；

所述数据获取单元，设置为获取统一融合后的数据。

可选的，所述数据处理装置还包括分类存储模块，其设置为分类存储所述统一融合后的数据。

15

可选的，所述数据处理装置还包括接口配置模块，其设置为配置数据采集接口；

20 所述适配模块获取待采集数据样本，对所述待采集数据样本进行预处理，包括，所述适配模块通过所述数据采集接口获取待采集数据样本，对所述待采集数据样本进行预处理；

所述数据处理模块控制集群中每个节点进行数据采集和数据处理，包括，所述数据处理模块控制集群中每个节点通过所述数据采集接口进行数据采集，控制集群中每个节点对采集到的数据进行数据处理。

25 可选的，其中，所述接口配置模块包括：

参数获取单元，设置为获取数据采集接口配置参数；

连接测试单元，设置为根据所述配置参数进行所述数据采集接口的连接

测试；以及

文件分发单元，设置为若所述连接测试成功，则将包含有所述配置参数的数据采集接口配置文件向集群中的其他节点进行分发，配置所述集群中的每个节点，并控制所述集群中的每个节点连接所述数据采集接口；

- 5 所述参数获取单元还设置为，若所述连接测试不成功，则获取所述数据采集接口配置参数。

此外，本发明实施例还提供一种数据处理系统，包括适配处理器、数据处理器和数据统一融合处理器，其中：

- 10 所述适配处理器，设置为获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中每个节点的采集任务；

- 15 所述数据处理器，设置为根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务，控制集群中每个节点进行数据采集和数据处理，获取处理后的数据；

所述数据统一融合处理器，设置为将所述处理后的数据进行统一融合，获取统一融合后的数据。

可选的，所述数据处理系统还包括：

- 20 数据采集接口适配器，设置为配置数据采集接口；
数据分类存储器，设置为分类存储所述统一融合后的数据；以及
系统管理器，其中，所述系统管理器包括：

- 25 数据管理模块，设置为管理所述数据采集接口配置文件、数据采集接口的连接测试结果、待采集数据样本、质量评估报告和集群中所有节点的负载情况；

资源管理模块，设置为管理所述待采集数据样本、数据采集、数据处理、数据统一融合和数据存储，及集群资源消耗情况，控制所述集群中所有节点

的负载均衡;

模型管理模块, 设置为管理数据采集接口配置模板、数据评估模板、数据采集模板、数据处理模型、数据规整模型、数据集成模型和数据建模模型; 以及

5 过程监控模块, 设置为监控数据采集接口配置、读取待采集数据样本、数据评估、数据处理、数据统一融合和数据存储过程, 并记录、处理数据采集接口配置、读取待采集数据样本、数据评估、数据处理、数据统一融合和数据存储过程中的完成状态、中间错误和异常。

10 本发明实施例通过对异构数据根据样本进行分析, 自动适配数据采集模式和数据处理模型, 采集数据并进行处理数据, 调整集群中所有节点的负载均衡, 并对有内在关联性的各种异构数据进行统一融合, 有效解决了相关技术中大数据处理效率低、融合度低的问题, 提高了数据处理效率和数据融合度, 提升了最终数据的可利用价值和实用性, 为大数据分析和价值挖掘提供了数据质量保障。

在阅读并理解了附图和详细描述后, 可以明白其他方面。

附图概述

图 1 为数据处理方法第一实施例的流程示意图;

20 图 2 为数据处理方法第二实施例的流程示意图;

图 3 为数据处理方法第三实施例的流程示意图;

图 4 为数据处理方法第四实施例的流程示意图;

图 5 为数据处理方法第五实施例的流程示意图;

图 6 为数据处理方法第六实施例的流程示意图;

25 图 7 为数据处理方法第七实施例的流程示意图;

图 8 为数据处理装置第一实施例的功能模块示意图;

图 9 为数据处理装置第二实施例的功能模块示意图;

- 图 10 为数据处理装置第三实施例的功能模块示意图；
图 11 为数据处理装置第四实施例的功能模块示意图；
图 12 为数据处理装置第五实施例的功能模块示意图；
图 13 为数据处理装置第六实施例的功能模块示意图；
5 图 14 为数据处理装置第七实施例的功能模块示意图；
图 15 为数据处理系统第一实施例的结构示意图；
图 16 为数据处理系统第二实施例的结构示意图。

本发明的实施方式

- 10 本发明实施例的主要解决方案是：通过数据采集接口获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中每个节点的采集任务；根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务，控制集群中每个节点进行数据采集和数据处理，获取处理后的数据；将所述处理后的数据进行统一融合，获取统一融合后的
15 数据。

由于相关技术数据处理过程人工参与度高，数据处理不精细、效率低，而且对分散的、有内在关联的数据也没有进行统一的融合，融合度不高，导致获取的最终数据可利用价值和实用性不高。

- 20 本发明实施例提供一种解决方案，能够自动适配数据采集模式和处理类型，对多源异构海量数据进行统一融合，从而提高数据处理效率和数据融合度，提升了最终数据的可利用价值和实用性，为大数据分析和价值挖掘提供了数据质量保障。

- 25 参照图 1，数据处理方法第一实施例提供一种数据处理方法，所述数据处理方法包括以下步骤：

步骤 S10，获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中每个节点的采集任务；

控制集群中每个节点采集待采集数据，获取待采集数据样本和集群中所有节点的负载情况，待采集数据样本包括预设数量的待采集数据，集群中所有节点的负载情况包括数据采集过程中每个节点的数据采集速率、性能、响应情况、负载能力，用于了解待采集数据样本采集过程中每个节点的资源负载情况。

5

对待采集数据样本进行预处理，检测与判定待采集数据基本质量属性，获取质量评估报告。

根据质量评估报告和集群中所有节点的负载情况，决定数据采集模式、适配数据处理模型、分配集群中每个节点的采集任务。

10

根据不同的数据类型适配不同的数据采集模式，不同的数据采集模式包括全量、增量、非实时、实时，例如：动态流式数据适配增量实时采集模式，结构化数据适配全量非实时采集模式，半结构化数据及非结构化数据适配全量非实时采集模式，数据采集模式也可以是其他采集模式，可根据实际需要灵活设置。

15

根据不同的数据类型适配数据处理模型，不同的数据类型如流式数据、批量数据、结构化数据、半结构化数据和非结构化数据进行不同的数据处理，预置数据处理模型用于流式数据、非结构化图片数据、非结构化音视频数据、非结构化文本数据、半结构化数据、结构化数据的处理，可直接适配调用，支持模型复用和修订。

20

数据处理模型包括：动态流式数据为数据抽取、数据过滤、数据关联、数据转换、数据分类的顺次处理步骤；结构化数据为数据抽取、异常清洗的顺次处理步骤；半结构化数据及非结构化数据为数据抽取、数据解析、异常清洗、数据分割、特征提取的顺次处理步骤。需要说明的是，可以直接适配数据处理模型，也可以对数据处理模型进行微调使用，也可以根据实际处理

25

需求对数据处理步骤和模型进行灵活设置。

集群中每个节点的采集任务的分配，实现了对集群中每个节点负载情况的动态调整，消除或减少系统中所有节点负载不均衡的现象，提高数据采集和数据处理效率。

步骤 S20, 根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务, 控制集群中每个节点进行数据采集和数据处理, 获取处理后的数据;

5 控制集群中每个节点根据已分配的对应的采集任务 (即每个节点根据已分配的本节点的采集任务), 以适配的采集模式进行数据采集, 例如: 动态流式数据以增量实时采集模式进行数据采集, 结构化数据以全量非实时采集模式进行数据采集, 半结构化数据及非结构化数据以全量非实时采集模式进行数据采集, 获取采集到的数据。

10 控制集群中每个节点以适配的数据处理模型对本节点采集到的数据分别进行数据处理, 例如: 根据数据处理模型对动态流式数据顺次进行数据抽取、数据过滤、数据关联、数据转换、数据分类的处理步骤; 根据数据处理模型对结构化数据顺次进行数据抽取、异常清洗的处理步骤; 根据数据处理模型对半结构化数据及非结构化数据顺次进行数据抽取、数据解析、异常清洗、
15 数据分割、特征提取的处理步骤, 实现对各种不同的异构数据按照自身数据特性进行对应的采集和处理, 获取处理后的数据。

步骤 S30, 将所述处理后的数据进行统一融合, 获取统一融合后的数据。

对采集及处理后的数据进行数据规整、数据集成、数据建模, 其中: 数
20 据规整和数据集成针对有内在关联性的数据进行统一格式规整和有效集成, 数据建模提供数据模型设计、数据索引构建功能; 这样可将分散的、有内在联系的各种数据进行关联, 获取统一融合后的数据。

在本实施例中, 通过对异构数据根据样本进行分析, 自动适配数据采集
25 模式和数据处理模型, 采集数据并处理数据, 调整集群中所有节点的负载均衡, 并对有内在关联性的各种异构数据进行统一融合, 有效解决了相关技术中大数据处理效率低、融合度低的问题, 提高了数据处理效率和数据融合度, 提升了最终数据的可利用价值和实用性, 为大数据分析和价值挖掘提供了数

据质量保障。

可选地，参照图 2，数据处理方法第二实施例提供一种数据处理方法，基于上述图 1 所示的实施例，步骤 S10 包括：

5 步骤 S11、控制所述集群中每个节点读取待采集数据，获取待采集数据样本和集群中所有节点的负载情况；

控制集群中每个节点采集预设数量的待采集数据，获取待采集数据样本，和数据采集过程中集群中所有节点的负载情况，待采集数据样本包括预设数量的待采集数据，集群中所有节点负载情况包括数据采集过程中每个节点的
10 数据采集速率、性能、响应情况、负载能力。

步骤 S12、评估所述待采集数据样本，获取质量评估报告，所述质量评估报告包括：数据类型、数据编码、数据冗余率、数据稀疏性和数据传输速度；

15 控制集群中每个节点对各自采集到的待采集数据样本进行数据质量评估，包括集群中每个节点采集到的待采集数据样本的数据类型、数据编码、数据冗余率、数据稀疏性和数据传输速度；然后将集群中每个节点数据质量评估的数据汇总、整理成为最终的质量评估报告，最终的质量评估报告是最终
20 的待采集数据样本的数据基本质量属性的检测与判定，包括最终的待采集数据样本的数据类型、数据编码、数据冗余率、数据稀疏性和数据传输速度。
或，

获取集群中每个节点采集到的待采集数据样本，汇总采集到的待采集数据样本，对汇总后的待采集数据样本进行数据基本质量属性的检测与判定，包括数据类型、数据编码、数据冗余率、数据稀疏性和数据传输速度，获取
25 质量评估报告。

步骤 S13、根据所述质量评估报告和所述集群中所有节点的负载情况，适配数据采集模式和数据处理模型，分配所述集群中每个节点的采集任务。

根据质量评估报告中的数据类型和数据传输速率适配数据采集模式，根据质量评估报告中的数据编码、数据冗余率、数据稀疏性适配数据处理模型，根据质量评估报告中的数据传输速度和集群中所有节点的负载情况分配集群中每个节点的采集任务。

5

在本实施例中，获取待采集数据样本，通过对待采集数据样本的预处理，得到质量评估报告和集群中所有节点的负载情况，并根据质量评估报告和集群中所有节点的负载情况，自动适配数据采集模式和数据处理模型，实现了各种异构数据包括各种动态的流式数据、静态的结构化数据、半结构化数据和非结构化数据按照其自身的数据特性进行更加合理有效的数据采集及数据处理；合理分配采集任务，使集群中每个节点按照相应的采集任务进行数据采集，实现了集群中所有节点的负载平衡，提升了集群处理能力。

可选地，参照图 3，数据处理方法第三实施例提供一种数据处理方法，基于上述图 1 所示的实施例，步骤 S20 包括：

步骤 S21、控制所述集群中每个节点根据对应的采集任务，以所述采集模式进行数据采集；

控制集群中每个节点根据已分配的对应的采集任务，以适配的采集模式进行数据采集，例如：动态流式数据以增量实时采集模式进行数据采集并缓存，结构化数据以全量非实时采集模式进行数据采集，半结构化数据及非结构化数据以全量非实时采集模式进行数据采集，非结构化数据会细分为文本数据、音频数据、视频数据、图片数据，不同类型的非结构化数据的采集方法不同，例如文本数据会进行文本数据采集、音频数据会进行音频数据采集、视频数据会进行视频数据采集、图片数据会进行图片数据采集，获取采集到的数据。

步骤 S22、控制所述集群中每个节点根据所述数据处理模型处理采集到的数据，获取处理后的数据。

控制集群中每个节点根据适配的数据处理模型对本节点采集到的数据分别进行数据处理，例如：根据数据处理模型对动态流式数据顺次进行数据抽取、数据过滤、数据关联、数据转换、数据分类的处理步骤；根据数据处理模型对结构化数据根据数据处理模型顺次进行数据抽取、异常清洗的处理步骤；根据数据处理模型对半结构化数据及非结构化数据根据数据处理模型顺次进行数据抽取、数据解析、异常清洗、数据分割、特征提取的处理步骤，完成数据处理，获取处理后的数据。非结构化数据可细分为文本数据、音频数据、视频数据、图片数据，不同类型非结构化数据处理方法不同，如根据数据处理模型对文本数据会进行文本数据采集、文本结构及编码解析、重复及逻辑异常清洗、中文分词、特征提取的处理步骤。

数据处理过程中产生的中间数据、元数据和进行数据处理后的结果数据，根据采集模式的不同选择合适的存储介质进行数据存储或缓存，即待采集数据的原始数据经过数据抽取、数据解析及异常清洗之后的结果数据、经过数据分割后的结果数据、经过特征提取后的结果数据都会存入分布式文件系统或分布式数据库；增量流式数据的中间数据采用内存数据库进行缓存，结果数据先进行文件缓存再存入分布式数据库。

在本实施例中，自动根据不同的数据类型，使用合适的数据采集模式进行数据采集，根据适配的数据处理模型对不同类型的数据进行特定的数据处理，实现了根据各种异构数据自身数据特性自动进行合适、有针对性的数据采集和处理，提升了数据处理效率和处理后的数据质量，使处理前的数据经过处理成为可操作的统一的数据，便于数据融合。

可选地，参照图 4，数据处理方法第四实施例提供一种数据处理方法，基于上述图 1 所示的实施例，步骤 S30 包括：

步骤 S31、根据预设的数据规整模型，对处理后的数据进行数据规整；

根据预设的数据规整模型，对处理后的不同数据进行统一的数据规整，包括：统一的格式转换、统一编码、数据修正和缺失填充，获取统一规整的

数据。

步骤 S32、根据预设的数据集成模型，对处理后的数据进行数据集成；

5 根据预设的数据集成模型，对处理后的不同数据或数据规整后的数据进行数据集成，包括：数据关联、数据合并、数据分组、数据汇总，形成统一的常用的基础的数据集合。

步骤 S33、根据预设的数据建模模型，对处理后的数据进行数据建模；

10 根据预设的数据建模模型，对处理后的不同数据、或数据规整后的数据、或数据集成后的数据进行模型设计、索引构建，以形成后续更高级别数据分析、数据挖掘的基础数据支撑。

以上步骤 S31、S32 和 S33 可以执行其中的任意一个步骤，也可以执行任意多个步骤。

15 步骤 S34、得到统一融合后的数据。

对处理后的数据，根据数据处理后的情况，进行以下一种或多种操作：数据规整的统一融合操作，数据集成的统一融合操作，数据建模的统一融合操作。可按顺次进行数据规整、数据集成和数据建模完成统一融合，也可按顺次进行数据规整、数据集成完成统一融合，或其他统一融合操作完成统一融合，可根据实际统一融合的需要进行数据规整、数据集成或数据建模的统一融合操作。最后获取统一融合后的数据。

20

在本实施例中，对处理后的数据进行数据规整、数据集成、数据建模的统一融合操作，实现了自动对分散的、有内在关联性的各异构数据的关联，
25 提高了数据融合度，提升了最终数据的可利用价值和实用性，为后续的大数据分析、数据价值挖掘提供基础数据支撑和数据质量保证。

可选地，参照图 5，数据处理方法第五实施例提供一种数据处理方法，基于上述图 1 至图 4 中所示任一实施例（本实施例以图 1 为例），步骤 S30 之后还包括：

步骤 S40、分类存储所述统一融合后的数据。

- 5 对不同类型的数据分类存储，将统一融合后的数据存储于分布式数据仓库，将特定数据存储于专用数据库，例如：将交通、气象等时空数据存储于时空数据库，将社交网络等图数据存储于图数据库。而将采集到的半结构化数据、非结构化数据的原始数据存储于分布式文件系统；将采集到的结构化数据的原始数据存储于分布式数据库；将半结构化数据及非结构化数据处理
- 10 后的数据存储于分布式数据库；将流式数据处理在实时数据库进行。

在本实施例中，自动适配数据采集模式和数据处理模型，根据数据处理模型处理采集到的数据，对处理后的数据进行统一融合，分类存储统一融合后的数据，便于后续数据的大数据的分类查询和分析处理，提升了最终数据

15 的实用性和调取便捷性。

可选地，参照图 6，数据处理方法第六实施例提供一种数据处理方法，基于上述图 1 所示实施例，步骤 S10 之前还包括：

步骤 S50、配置数据采集接口。

- 20 对不同数据源的数据采集接口进行统一配置，并对配置进行连接测试，连接成功后再进行配置文件的统一分发，控制集群中每个节点连接不同数据源的数据采集接口，通过配置成功的数据采集接口进行数据采集；若连接测试失败，重新进行数据采集接口配置。

- 25 步骤 S10 中获取待采集数据样本，对所述待采集数据样本进行预处理包括，通过所述数据采集接口获取待采集数据样本，对所述待采集数据样本进行预处理。

控制集群中每个节点通过数据采集接口采集待采集数据，获取待采集数

据样本和集群中所有节点的负载情况，待采集数据样本包括预设数量的待采集数据，所有节点的负载情况包括数据采集过程中每个节点的数据采集速率、性能、响应情况、负载能力，用于了解待采集数据样本采集过程中每个节点的资源负载情况。

- 5 对待采集数据样本进行预处理，检测与判定待采集数据基本质量属性，获取质量评估报告。

10 步骤 S20 中控制集群中每个节点进行数据采集和数据处理包括，控制集群中每个节点通过所述数据采集接口进行数据采集，控制集群中每个节点对采集到的数据进行数据处理。

15 控制集群中每个节点根据已分配的对应的采集任务，以适配的采集模式通过数据采集接口进行数据采集，例如：动态流式数据通过数据采集接口以增量实时采集模式进行数据采集，结构化数据通过数据采集接口以全量非实时采集模式进行数据采集，半结构化数据及非结构化数据通过数据采集接口以全量非实时采集模式进行数据采集，获取采集到的数据。

20 控制集群中每个节点以适配的数据处理模型对本节点采集到的数据分别进行数据处理，例如：根据数据处理模型对动态流式数据顺次进行数据抽取、数据过滤、数据关联、数据转换、数据分类的处理步骤；根据数据处理模型对结构化数据顺次进行数据抽取、异常清洗的处理步骤；根据数据处理模型对半结构化数据及非结构化数据顺次进行数据抽取、数据解析、异常清洗、数据分割、特征提取的处理步骤，实现对各种不同的异构数据按照自身数据特性进行对应的采集和处理。

25 在本实施例中，统一配置数据采集接口，兼顾静态的、动态的，结构化的、半结构化的、非结构化的异构数据采集接口配置，实现了多源数据采集时，数据采集接口的统一配置，为多源异构数据采集做好了准备，从而可以控制集群中每个节点通过数据采集接口采集多源异构数据，提升了数据采集速率。

可选地，参照图 7，数据处理方法第七实施例提供一种数据处理方法，基于上述图 6 所示实施例，步骤 S50 包括：

步骤 S51、获取数据采集接口配置参数；

- 5 向用户给出预置的接口配置模板，包括：文件系统接口配置模板、数据库接口配置模板和网络接口配置模板，供用户选择，根据用户选定的接口配置模板获取模板中的配置参数；或，

- 10 向用户给出预置的接口配置模板，包括：文件系统接口配置模板、数据库接口配置模板和网络接口配置模板，用户选择接口配置模板后，可根据实际情况对模板中的配置参数进行修改；或，

用户自定义设置数据采集接口配置参数，可根据实际需要灵活设置。

数据采集接口配置参数包括数据采集接口的关键接口参数，例如：网络地址、端口号和路径。

- 15 步骤 S52、根据所述配置参数进行所述数据采集接口的连接测试；

根据获取的数据采集接口配置参数控制集群中主节点对数据采集接口进行连接，测试根据获取的配置参数，是否能成功连接数据采集接口。若根据获取的配置参数，成功连接数据采集接口，则连接测试成功；若根据获取的配置参数，未成功连接数据采集接口，则连接测试不成功。

20

步骤 S53、若所述连接测试成功，则将包含有所述配置参数的数据采集接口配置文件向集群中的其他节点进行分发，配置所述集群中的每个节点，并控制所述集群中的每个节点连接所述数据采集接口；若所述连接测试不成功，则转入步骤 S51。

- 25 若连接测试成功，则将获取的数据采集接口配置参数固化为数据采集接口的配置文件，将配置文件向集群中的其他节点进行分发，并控制集群中每个节点根据获取的数据采集接口配置参数连接数据采集接口。

若连接测试不成功，则通知用户未成功连接数据采集接口，由用户更换接口配置模板，或修改对模板中的数据采集接口配置参数，或自定义设置数据采集接口配置参数，重新获取数据采集接口配置参数。

5 在本实施例中，通过获取数据采集接口配置参数，连接测试获取正确可用的数据采集接口配置参数，用以连接数据采集接口，完成了数据采集接口的配置，实现了对多源数据采集接口的统一自动化配置及连接，为数据采集做好了准备和支持，提升了数据采集和处理效率。

10 参照图 8，数据处理装置第一实施例提供一种数据处理装置，所述数据处理装置包括适配模块 100，数据处理模块 200 以及统一融合模块 300，其中：

 适配模块 100，设置为获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中每个节点的采集任务；

15 适配模块 100 控制集群中每个节点采集待采集数据，获取待采集数据样本和集群中所有节点的负载情况，待采集数据样本包括预设数量的待采集数据，集群中所有节点的负载情况包括数据采集过程中每个节点的数据采集速率、性能、响应情况、负载能力，用于了解待采集数据样本采集过程中每个节点的资源负载情况。

20 适配模块 100 对待采集数据样本进行预处理，检测与判定待采集数据基本质量属性，获取质量评估报告。

 根据质量评估报告和集群中所有节点的负载情况，适配模块 100 决定数据采集模式、适配数据处理模型、分配集群中每个节点的采集任务。

25 根据不同的数据类型适配不同的数据采集模式，不同的数据采集模式包括全量、增量、非实时、实时，例如：动态流式数据适配增量实时采集模式，结构化数据适配全量非实时采集模式，半结构化数据及非结构化数据适配全量非实时采集模式，数据采集模式也可以是其他采集模式，可根据实际需要灵活设置。

 根据不同的数据类型适配数据处理模型，不同的数据类型如流式数据、

批量数据、结构化数据、半结构化数据和非结构化数据进行不同的数据处理，预置数据处理模型用于流式数据、非结构化图片数据、非结构化音视频数据、非结构化文本数据、半结构化数据、结构化数据的处理，可直接适配调用，支持模型复用和修订。

- 5 数据处理模型包括：动态流式数据为数据抽取、数据过滤、数据关联、数据转换、数据分类的顺次处理；结构化数据为数据抽取、异常清洗的顺次处理；半结构化数据及非结构化数据为数据抽取、数据解析、异常清洗、数据分割、特征提取的顺次处理。需要说明的是，可以直接适配数据处理模型，也可以对数据处理模型进行微调使用，也可以根据具体处理需求对数据处理
- 10 和模型进行灵活设置。

集群中每个节点的采集任务的分配，实现了对集群中每个节点负载情况的动态调整，消除或减少系统中所有节点负载不均衡的现象，提高数据采集和数据处理效率。

- 15 数据处理模块 200，设置为根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务，控制集群中每个节点进行数据采集和数据处理，获取处理后的数据；

数据处理模块 200 控制集群中每个节点根据已分配的对应的采集任务，以适配的采集模式进行数据采集，例如：动态流式数据以增量实时采集模式

20 进行数据采集，结构化数据以全量非实时采集模式进行数据采集，半结构化数据及非结构化数据以全量非实时采集模式进行数据采集，获取采集到的数据。

数据处理模块 200 控制集群中每个节点以适配的数据处理模型对本节点采集到的数据进行数据处理，例如：根据数据处理模型对动态流式数据顺次

25 进行数据抽取、数据过滤、数据关联、数据转换、数据分类的处理；根据数据处理模型对结构化数据顺次进行数据抽取、异常清洗的处理；根据数据处理模型对半结构化数据及非结构化数据顺次进行数据抽取、数据解析、异常清洗、数据分割、特征提取的处理，实现对各种不同的异构数据按照自身数据特性进行对应的采集和处理，获取处理后的数据。

统一融合模块 300，设置为将所述处理后的数据进行统一融合，获取统一融合后的数据。

5 统一融合模块 300 对采集及处理后的数据进行数据规整、数据集成、数据建模，数据规整和数据集成针对有内在关联性的数据进行，数据建模提供数据模型设计、数据索引构建功能，将分散的、有内在联系的各种数据进行关联，获取统一融合后的数据。

10 在本实施例中，适配模块 100 通过对异构数据根据样本进行分析，自动适配数据采集模式和数据处理模型，数据处理模块 200 采集数据并处理数据，调整集群中所有节点的负载均衡，统一融合模块 300 对有内在关联性的各种异构数据进行统一融合，有效解决了相关技术中大数据处理效率低、融合度低的问题，提高了数据处理效率和数据融合度，提升了最终数据的可利用价值和实用性，为大数据分析和价值挖掘提供了数据质量保障。

15

可选地，参照图 9，数据处理装置第二实施例提供一种数据处理装置，基于上述图 8 所示的实施例，适配模块 100 包括样本采集单元 110、数据评估单元 120 以及适配单元 130，其中：

20 样本采集单元 110，设置为控制所述集群中每个节点读取待采集数据，获取待采集数据样本和集群中所有节点的负载情况；

样本采集单元 110 控制集群中每个节点采集预设数量的待采集数据，获取待采集数据样本，和数据采集过程中集群中所有节点的负载情况，待采集数据样本包括预设数量的待采集数据，集群中所有节点的负载情况包括数据采集过程中每个节点的数据采集速率、性能、响应情况、负载能力。

25

数据评估单元 120，设置为评估待采集数据样本，获取质量评估报告，所述质量评估报告包括：数据类型、数据编码、数据冗余率、数据稀疏性和数据传输速度。

数据评估单元 120 控制集群中每个节点对各自采集到的待采集数据样本进行数据质量评估，包括集群中每个节点采集到的待采集数据样本的数据类型、数据编码、数据冗余率、数据稀疏性和数据传输速度；然后数据评估单元 120 将集群中所有节点的数据质量评估的数据汇总、整理成为最终的质量
5 评估报告，最终的质量评估报告是最终的待采集数据样本的数据基本质量属性的检测与判定，包括最终的待采集数据样本的数据类型、数据编码、数据冗余率、数据稀疏性和数据传输速度。或，

数据评估单元 120 获取集群中每个节点采集到的待采集数据样本，汇总采集到的待采集数据样本，对汇总后的待采集数据样本进行数据基本质量属
10 性的检测与判定，包括数据类型、数据编码、数据冗余率、数据稀疏性和数据传输速度，获取质量评估报告。

适配单元 130，设置为根据所述质量评估报告和所述集群中所有节点的负载情况，适配数据采集模式和数据处理模型，分配所述集群中每个节点的
15 采集任务。

适配单元 130 根据质量评估报告中的数据类型和数据传输速率适配数据采集模式，根据质量评估报告中的数据编码、数据冗余率、数据稀疏性适配
数据处理模型，根据质量评估报告中的数据传输速度和集群中所有节点的负载情况分配集群中每个节点的采集任务。

20

在本实施例中，样本采集单元 110 获取待采集数据样本，数据评估单元 120 通过对待采集数据样本的预处理，得到质量评估报告和集群中所有节点的负载情况，适配单元 130 根据质量评估报告和集群中所有节点的负载情况，自动适配数据采集模式和数据处理模型，实现了各种异构数据包括各种动态
25 的流式数据、静态的结构化数据、半结构化数据和非结构化数据按照其自身的数据特性进行更加合理有效的数据采集及数据处理；合理分配采集任务，使集群中每个节点按照相应的采集任务进行数据采集，实现了集群中所有节点的负载平衡，提升了集群处理能力。

可选地，参照图 10，数据处理装置第三实施例提供一种数据处理装置，基于上述图 8 所示的实施例，数据处理模块 200 包括数据采集单元 210 和数据处理单元 220，其中：

5 数据采集单元 210，设置为控制所述集群中每个节点根据对应的采集任务，以所述采集模式进行数据采集；

10 数据采集单元 210 控制集群中每个节点根据已分配的对应的本节点的采集任务，以适配的采集模式进行数据采集，例如：动态流式数据以增量实时采集模式进行数据采集并缓存，结构化数据以全量非实时采集模式进行数据采集，半结构化数据及非结构化数据以全量非实时采集模式进行数据采集，非结构化数据可细分为文本数据、音频数据、视频数据、图片数据，不同类型的非结构化数据的采集方法不同，例如文本数据会进行文本数据采集、音频数据会进行音频数据采集、视频数据会进行视频数据采集、图片数据会进行图片数据采集，获取采集到的数据。

15

 数据处理单元 220，设置为控制所述集群中每个节点根据所述数据处理模型处理所述采集到的数据，获取处理后的数据。

20 数据处理单元 220 控制集群中每个节点根据适配的数据处理模型对本节点采集到的数据分别进行数据处理，例如：根据数据处理模型对动态流式数据顺次进行数据抽取、数据过滤、数据关联、数据转换、数据分类的处理；根据数据处理模型对结构化数据根据数据处理模型顺次进行数据抽取、异常清洗的处理；根据数据处理模型对半结构化数据及非结构化数据根据数据处理模型顺次进行数据抽取、数据解析、异常清洗、数据分割、特征提取的处理，完成数据处理，获取处理后的数据。非结构化数据可细分为文本数据、25 音频数据、视频数据、图片数据，不同类型非结构化数据处理方法不同，如根据数据处理模型对文本数据会进行文本数据采集、文本结构及编码解析、重复及逻辑异常清洗、中文分词、特征提取的处理。

 数据处理过程中产生的中间数据、元数据和进行数据处理后的结果数

据，数据处理单元 220 根据采集模式的不同选择合适的存储介质进行数据存储或缓存，即待采集数据的原始数据经过数据抽取、数据解析及异常清洗之后的结果数据、经过数据分割后的结果数据、经过特征提取后的结果数据存入分布式文件系统或分布式数据库；增量流式数据的中间数据采用内存数据库进行缓存，结果数据先进行文件缓存再存入分布式数据库。

在本实施例中，数据采集单元 210 自动根据不同的数据类型，使用合适的的数据采集模式进行数据采集，数据处理单元 220 根据适配的数据处理模型对不同类型的数据进行特定的数据处理，实现了根据各种异构数据自身数据特性自动进行合适、有针对性的数据采集和处理，提升了数据处理效率和处理后的数据质量，使处理前的数据经过处理成为可操作的统一的数据，便于数据融合。

可选地，参照图 11，数据处理装置第四实施例提供一种数据处理装置，基于上述图 8 所示的实施例，统一融合模块 300 包括数据获取单元 340 和以下一个或多个单元：数据规整单元 310、数据集成单元 320、数据建模单元 330，其中：

数据规整单元 310，设置为根据预设的数据规整模型，对处理后的数据进行数据规整；

根据预设的数据规整模型，数据规整单元 310 对处理后的不同数据进行统一的数据规整，包括：统一的格式转换、统一编码、数据修正和缺失填充，获取统一规整的数据。

数据集成单元 320，设置为根据预设的数据集成模型，对处理后的数据进行数据集成；

根据预设的数据集成模型，数据集成单元 320 对处理后的不同数据或数据规整后的数据进行数据集成，包括：数据关联、数据合并、数据分组、数据汇总，形成统一的常用的基础的数据集合。

数据建模单元 330，设置为根据预设的数据建模模型，对处理后的数据进行数据建模；

5 根据预设的数据建模模型，数据建模单元 330 对处理后的不同数据、或数据规整后的数据、或数据集成后的数据进行模型设计、索引构建，以形成后续更高级别数据分析、数据挖掘的基础数据支撑。

数据获取单元 340，设置为获取统一融合后的数据。

10 对处理后的数据，根据数据处理后的情况，数据获取单元 340 控制数据规整单元 310 进行数据规整的统一融合操作，数据集成单元 320 进行数据集成的统一融合操作，数据建模单元 330 进行数据建模的统一融合操作，可按顺次进行数据规整、数据集成和数据建模完成统一融合，也可按顺次进行数据规整、数据集成完成统一融合，或其他统一融合操作完成统一融合，可根据实际统一融合的需要进行数据规整、数据集成或数据建模的统一融合操作。
15 数据获取单元 340 获取统一融合后的数据。

20 在本实施例中，数据获取单元 340 控制数据规整单元 310、数据集成单元 320 和数据建模单元 330，对处理后的数据进行统一融合，实现了自动对分散的、有内在关联性的各种异构数据的关联，提高了数据融合度，提升了最终数据的可利用价值和实用性，为后续的大数据分析、数据价值挖掘提供基础数据支撑和数据质量保证。

25 可选地，参照图 12，数据处理装置第五实施例提供一种数据处理装置，基于上述图 8 至图 11 中所示任一实施例（本实施例以图 8 为例），所述数据处理装置还包括分类存储模块 400：

分类存储模块 400，设置为分类存储所述统一融合后的数据。

分类存储模块 400 对不同类型的数据分类存储，将统一融合后的数据存储于分布式数据仓库，将特定数据存储于专用数据库，例如：分类存储模块

400 将交通、气象等时空数据存储于时空数据库；分类存储模块 400 将社交网络等图数据存储于图数据库。分类存储模块 400 将采集到的半结构化数据、非结构化数据的原始数据存储于分布式文件系统；分类存储模块 400 将采集到的结构化数据的原始数据存储于分布式数据库；分类存储模块 400 将半结构化数据及非结构化数据进行处理后的数据存储于分布式数据库；分类存储模块 400 将流式数据处理在实时数据库进行。

在本实施例中，适配模块 100 自动适配数据采集模式和数据处理模型，数据处理模块 200 根据数据处理模型处理采集到的数据，统一融合模块 300 对处理后的数据进行统一融合，分类存储模块 400 分类存储统一融合后的数据，便于后续数据的大数据的分类查询和分析处理，提升了最终数据的实用性和调取便捷性。

可选地，参照图 13，数据处理装置第六实施例提供一种数据处理装置，基于上述图 8 所示实施例，所述数据处理装置还包括接口配置模块 500：

接口配置模块 500，设置为配置数据采集接口。

接口配置模块 500 对不同数据源的数据采集接口进行统一配置，并对配置进行连接测试，连接成功后接口配置模块 500 再进行配置文件的统一分发，控制集群中每个节点连接不同数据源的数据采集接口；若连接测试失败，接口配置模块 500 重新进行数据采集接口配置。

所述适配模块 100 获取待采集数据样本，对所述待采集数据样本进行预处理，包括，所述适配模块 100 通过所述数据采集接口获取待采集数据样本，对所述待采集数据样本进行预处理。

适配模块 100 控制集群中每个节点通过数据采集接口采集待采集数据，获取待采集数据样本和集群中所有节点的负载情况，待采集数据样本包括预设数量的待采集数据，集群中所有节点的负载情况包括数据采集过程中每个节点的数据采集速率、性能、响应情况、负载能力，用于了解待采集数据样

本采集过程中每个节点的资源负载情况。

适配模块 100 对待采集数据样本进行预处理，检测与判定待采集数据基本质量属性，获取质量评估报告。

5 所述数据处理模块 200 控制集群中每个节点进行数据采集和数据处理，包括，所述数据处理模块 200 控制集群中每个节点通过所述数据采集接口进行数据采集，控制集群中每个节点对采集到的数据进行数据处理。

10 数据处理模块 200 控制集群中每个节点根据已分配的对应的采集任务，以适配的采集模式通过数据采集接口进行数据采集，例如：动态流式数据通过数据采集接口以增量实时采集模式进行数据采集，结构化数据通过数据采集接口以全量非实时采集模式进行数据采集，半结构化数据及非结构化数据通过数据采集接口以全量非实时采集模式进行数据采集，获取采集到的数据。

15 数据处理模块 200 控制集群中每个节点以适配的数据处理模型对本节点采集到的数据进行数据处理，例如：根据数据处理模型对动态流式数据顺次进行数据抽取、数据过滤、数据关联、数据转换、数据分类的处理；根据数据处理模型对结构化数据顺次进行数据抽取、异常清洗的处理；根据数据处理模型对半结构化数据及非结构化数据顺次进行数据抽取、数据解析、异常清洗、数据分割、特征提取的处理，实现对各种不同的异构数据按照自身数据特性进行对应的采集和处理。

20

在本实施例中，接口配置模块 500 统一配置数据采集接口，兼顾静态的、动态的，结构化的、半结构化的、非结构化的异构数据采集接口配置，实现了多源数据采集时，数据采集接口的统一配置，为多源异构数据采集做好了准备，从而适配模块 100 和数据处理模块 200 可以控制集群中所有节点通过
25 数据采集接口采集多源异构数据，提升了数据采集速率。

可选地，参照图 14，数据处理装置第七实施例提供一种数据处理装置，基于上述图 13 所示实施例，接口配置模块 500 包括参数获取单元 510、连接

测试单元 520 以及文件分发单元 530，其中：

参数获取单元 510，设置为获取数据采集接口配置参数；

5 参数获取单元 510 向用户给出预置的接口配置模板，包括：文件系统接口配置模板、数据库接口配置模板和网络接口配置模板，供用户选择，根据用户选定的接口配置模板获取模板中的配置参数；或，

参数获取单元 510 向用户给出预置的接口配置模板，包括：文件系统接口配置模板、数据库接口配置模板和网络接口配置模板，用户选择接口配置模板后，可根据实际情况对模板中的配置参数进行修改；或，

10 用户自定义设置数据采集接口配置参数，参数获取单元 510 获取用户自定义设置的数据采集接口配置参数。

数据采集接口配置参数包括数据采集接口的关键接口参数，例如：网络地址、端口号和路径。

15 连接测试单元 520，设置为根据所述配置参数进行所述数据采集接口的连接测试；

20 连接测试单元 520 根据获取的数据采集接口配置参数控制集群中主节点对数据采集接口进行连接，测试根据获取的配置参数，是否能成功连接数据采集接口，若根据获取的配置参数，成功连接数据采集接口，则连接测试单元 520 确认连接测试成功；若根据获取的配置参数，未成功连接数据采集接口，则连接测试单元 520 确认连接测试不成功。

文件分发单元 530，设置为若所述连接测试成功，则将包含有所述配置参数的数据采集接口配置文件向集群中的其他节点进行分发，配置所述集群中的每个节点，并控制所述集群中的每个节点连接所述数据采集接口。

25 若连接测试成功，则文件分发单元 530 将获取的数据采集接口配置参数固化为数据采集接口的配置文件，将配置文件向集群中的其他节点进行分发，并控制集群中每个节点根据获取的数据采集接口配置参数连接数据采集接口。

所述参数获取单元 510 还设置为, 若所述连接测试不成功, 则获取所述数据采集接口配置参数。

5 若连接测试不成功, 则参数获取单元 510 通知用户未成功连接数据采集接口, 由用户更换接口配置模板, 或由用户修改对模板中的数据采集接口配置参数, 或由用户自定义设置数据采集接口配置参数, 参数获取单元 510 重新获取数据采集接口配置参数。

10 在本实施例中, 通过参数获取单元 510 获取数据采集接口配置参数, 连接测试单元 520 连接测试获取正确可用的数据采集接口配置参数, 文件分发单元 530 连接数据采集接口, 完成了数据采集接口的配置, 实现了对多源数据采集接口的统一自动化配置及连接, 为数据采集做好了准备和支持, 提升了数据采集和处理效率。

15 参照图 15, 数据处理系统第一实施例提供一种数据处理系统, 包括适配处理器 A、数据处理器 B 和数据统一融合处理器 C, 其中:

适配处理器 A, 设置为获取待采集数据样本, 对所述待采集数据样本进行预处理, 得到数据采集模式、数据处理模型和集群中每个节点的采集任务;

20 适配处理器 A 包括样本采集模块 A1, 数据评估模块 A2 和适配模块 A3, 其中:

样本采集模块 A1 设置为获取待采集数据样本, 数据评估模块 A2 设置为对所述待采集数据样本进行预处理, 适配模块 A3 设置为适配数据采集模式和数据处理模型, 分配集群中每个节点的采集任务。

25 数据处理器 B, 设置为根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务, 控制集群中每个节点进行数据采集和数据处理, 获取处理后的数据。

数据处理器 B 包括数据采集模块 B1 和数据处理模块 B2, 其中:

数据采集模块 B1 设置为根据适配的数据采集模式和集群中每个节点的采集任务采集数据,数据处理模块 B2 设置为根据适配的数据处理模型对采集到的数据进行数据处理,获取处理后的数据。

- 5 数据统一融合处理器 C, 设置为将所述处理后的数据进行统一融合, 获取统一融合后的数据。

数据统一融合处理器 C 包括数据规整模块 C1、数据集成模块 C2、数据建模模块 C3 和数据获取模块 C4, 其中:

- 10 数据规整模块 C1 设置为根据预设的数据规整模型, 对处理后的数据进行数据规整; 数据集成模块 C2 设置为根据预设的数据集成模型, 对处理后的数据进行数据集成; 数据建模模块 C3 设置为根据预设的数据建模模型, 对处理后的数据进行数据建模; 数据获取模块 C4 设置为控制据规整模块 C1、数据集成模块 C2 和数据建模模块 C3 进行数据统一融合, 获取统一融合后的数据。

- 15 在本实施例中, 适配处理器 A 通过对异构数据根据样本进行分析, 自动适配数据采集模式和数据处理模型, 数据处理器 B 采集数据并处理数据, 调整集群中所有节点的负载均衡, 数据统一融合处理器 C 对有内在关联性的各种异构数据进行统一融合, 有效解决了相关技术中大数据处理效率低、融合度低的问题, 提高了数据处理效率和数据融合度, 提升了最终数据的可利用
20 价值和实用性, 为大数据分析和价值挖掘提供了数据质量保障。

可选地, 参照图 16, 数据处理系统第二实施例提供一种数据处理系统, 基于上述图 15 所示的实施例, 所述系统还包括数据采集接口适配器 D、数据分类存储器 E 和系统管理器 F, 其中:

- 25 数据采集接口适配器 D, 设置为配置数据采集接口;

数据采集接口适配器 D 包括参数获取模块 D1、连接测试模块 D2 和文件分发模块 D3, 其中:

参数获取模块 D1 设置为获取数据采集接口配置参数; 连接测试模块 D2

设置为测试获取的数据采集接口配置参数是否能够对数据采集接口进行连接；文件分发模块 D3 设置为若连接测试成功，将包含有所述配置参数的数据采集接口配置文件向集群中的其他节点进行分发，配置所述集群中的所有节点，并控制所述集群中的所有节点连接数据采集接口。

5

数据分类存储器 E，设置为分类存储所述统一融合后的数据；

数据分类存储器 E 包括分布式文件系统 E1、分布式数据库 E2、分布式数据仓库 E3 和专用数据库 E4，其中：

10 数据分类存储器 E 将统一融合后的数据存储于分布式数据仓库 E3，将特定数据存储于专用数据库 E4，例如：将交通、气象等时空数据存储于专用数据库 E4，将社交网络等图数据存储于专用数据库 E4。

15 数据分类存储器 E 将采集到的半结构化数据、非结构化数据的原始数据存储于分布式文件系统 E1；将采集到的结构化数据的原始数据存储于分布式数据库 E2；将半结构化数据及非结构化数据进行处理后的数据存储于分布式数据库 E2；将流式数据处理在专用数据库 E4 进行。

系统管理器 F，所述系统管理器 F 包括数据管理模块 F1、资源管理模块 F2、模型管理模块 F3 以及过程监控模块 F4，其中：

20 数据管理模块 F1，设置为管理所述数据采集接口配置文件、数据采集接口的连接测试结果、待采集数据样本、质量评估报告和集群中所有节点的负载情况；

25 数据管理模块 F1 存储并分发数据采集接口配置文件；存储并向用户反馈数据采集接口的连接测试结果；存储或缓存待采集数据样本；评估待采集数据样本，获取并存储质量评估报告；获取并存储集群中所有节点的负载情况，便于其他模块查询或调取。

资源管理模块 F2，设置为管理所述待采集数据样本、数据采集、数据处理、数据统一融合和数据存储，及集群资源消耗情况，控制所述集群中所有

节点的负载均衡;

资源管理模块 F2 设置为根据待采集数据样本的采集过程、数据质量的评估过程、数据采集过程、数据处理过程、数据统一融合过程和数据存储过程中集群资源消耗情况,控制集群中所有节点的负载均衡。

5

模型管理模块 F3, 设置为管理数据采集接口配置模板、数据评估模板、数据采集模板、数据处理模型、数据规整模型、数据集成模型和数据建模模型;

10 模型管理模块 F3 设置为数据采集接口配置模板的存储和修改;数据评估模板的管理;数据采集模板的存储、适配、调用和修改;数据处理模型的存储、适配、调用和修改;数据规整模型的存储和调用;数据集成模型的存储和调用;数据建模模型的构建、存储和调用。

15 过程监控模块 F4, 设置为监控数据采集接口配置、读取待采集数据样本、数据评估、数据处理、数据统一融合和数据存储过程,并记录、处理数据采集接口配置、读取待采集数据样本、数据评估、数据处理、数据统一融合和数据存储过程中的完成状态、中间错误和异常。

过程监控模块 F4 设置为监控数据处理系统每个模块的运行状态、资源使用情况,记录并处理数据处理系统每个模块运行时错误、异常。

20

在本实施例中,系统管理器 F 监控适配处理器 A、数据处理器 B、数据统一融合处理器 C、数据采集接口适配器 D 和数据分类存储器 E 的运行状态、数据处理情况和负载情况,对每个模块运行进行过程控制,集中管理数据处理系统中各类数据,实现了多源数据接口的统一配置,异构数据采集方式和处理模型的自动适配,对异构数据的统一融合和分类存储,提高了数据处理效率和融合度,提升了最终数据的可利用价值和实用性,为大数据分析和价值挖掘提供了数据质量保障。

以上仅为本发明的可选实施例，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的权利要求范围内。

本领域普通技术人员可以理解上述方法中的全部或部分步骤可通过程序来指令相关硬件（例如处理器）完成，所述程序可以存储于计算机可读存储介质中，如只读存储器、磁盘或光盘等。可选地，上述实施例的全部或部分步骤也可以使用一个或多个集成电路来实现。相应地，上述实施例中的各模块/单元可以采用硬件的形式实现，例如通过集成电路来实现其相应功能，也可以采用软件功能模块的形式实现，例如通过处理器执行存储于存储器中的程序/指令来实现其相应功能。本发明实施例不限制于任何特定形式的硬件和软件的结合。

工业实用性

本发明实施例通过对异构数据根据样本进行分析，自动适配数据采集模式和数据处理模型，采集数据并进行处理数据，调整集群中所有节点的负载均衡，并对有内在关联性的各种异构数据进行统一融合，有效解决了相关技术中大数据处理效率低、融合度低的问题，提高了数据处理效率和数据融合度，提升了最终数据的可利用价值和实用性，为大数据分析和价值挖掘提供了数据质量保障。

权 利 要 求 书

1、一种数据处理方法，包括以下步骤：

获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中每个节点的采集任务；

5 根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务，控制集群中每个节点进行数据采集和数据处理，获取处理后的数据；

将所述处理后的数据进行统一融合，获取统一融合后的数据。

2、如权利要求 1 所述的数据处理方法，其中，所述获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中每个节点的采集任务，包括：

控制所述集群中每个节点读取待采集数据，获取待采集数据样本和集群中所有节点的负载情况；

评估所述待采集数据样本，获取质量评估报告，所述质量评估报告包括：数据类型、数据编码、数据冗余率、数据稀疏性和数据传输速度；

15 根据所述质量评估报告和所述集群中所有节点的负载情况，适配数据采集模式和数据处理模型，分配所述集群中每个节点的采集任务。

3、如权利要求 1 所述的数据处理方法，其中，所述根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务，控制集群中每个节点进行数据采集和数据处理，获取处理后的数据，包括：

20 控制所述集群中每个节点根据对应的采集任务，以所述采集模式进行数据采集；

控制所述集群中每个节点根据所述数据处理模型处理采集到的数据，获取处理后的数据。

4、如权利要求 1 所述的数据处理方法，其中，所述将所述处理后的数据进行统一融合，获取统一融合后的数据，包括以下至少之一：

25 根据预设的数据规整模型，对处理后的数据进行数据规整，得到统一融合后的数据；

根据预设的数据集成模型，对处理后的数据进行数据集成，得到统一融合后的数据；

根据预设的数据建模模型，对处理后的数据进行数据建模，得到统一融合后的数据。

5 5、如权利要求 1 至 4 中任一项所述的数据处理方法，所述将所述处理后的数据进行统一融合，获取统一融合后的数据之后，所述方法还包括：

分类存储所述统一融合后的数据。

6、如权利要求 1 所述的数据处理方法，所述获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中
10 每个节点的采集任务之前，所述方法还包括：

配置数据采集接口；

所述获取待采集数据样本，对所述待采集数据样本进行预处理的步骤包
括：

15 通过所述数据采集接口获取待采集数据样本，对所述待采集数据样本进行预处理；

所述控制集群中每个节点进行数据采集和数据处理的步骤包括：

控制集群中每个节点通过所述数据采集接口进行数据采集，控制集群中每个节点对采集到的数据进行数据处理。

20 7、如权利要求 6 所述的数据处理方法，其中，所述配置数据采集接口，包括：

获取数据采集接口配置参数；

根据所述配置参数进行所述数据采集接口的连接测试；

25 若所述连接测试成功，则将包含有所述配置参数的数据采集接口配置文件向集群中的其他节点进行分发，配置所述集群中的每个节点，并控制所述集群中的每个节点连接所述数据采集接口；

若所述连接测试不成功，则转入步骤：获取数据采集接口配置参数。

8、一种数据处理装置，包括：

适配模块，设置为获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中每个节点的采集任务；

5 数据处理模块，设置为根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务，控制集群中每个节点进行数据采集和数据处理，获取处理后的数据；以及

统一融合模块，设置为将所述处理后的数据进行统一融合，获取统一融合后的数据。

9、如权利要求 8 所述数据处理装置，其中，所述适配模块包括：

10 样本采集单元，设置为控制所述集群中每个节点读取待采集数据，获取待采集数据样本和集群中所有节点的负载情况；

数据评估单元，设置为评估待采集数据样本，获取质量评估报告，所述质量评估报告包括：数据类型、数据编码、数据冗余率、数据稀疏性和数据传输速度；以及

15 适配单元，设置为根据所述质量评估报告和所述集群中所有节点的负载情况，适配数据采集模式和数据处理模型，分配所述集群中每个节点的采集任务。

10、如权利要求 8 所述数据处理装置，其中，所述数据处理模块包括：

数据采集单元，设置为控制所述集群中每个节点根据对应的采集任务，以所述采集模式进行数据采集；以及

20 数据处理单元，设置为控制所述集群中每个节点根据所述数据处理模型处理所述采集到的数据，获取处理后的数据。

11、如权利要求 8 所述数据处理装置，其中，所述统一融合模块包括数据获取单元和以下一个或多个单元：数据规整单元、数据集成单元、数据建模单元，其中：

25 所述数据规整单元，设置为根据预设的数据规整模型，对处理后的数据进行数据规整；

所述数据集成单元，设置为根据预设的数据集成模型，对处理后的数据进行数据集成；

所述数据建模单元，设置为根据预设的数据建模模型，对处理后的数据进行数据建模；

所述数据获取单元，设置为获取统一融合后的数据。

12、如权利要求 8、9、10 或 11 所述的数据处理装置，所述数据处理装置还包括分类存储模块，其设置为分类存储所述统一融合后的数据。

13、如权利要求 8 所述数据处理装置，所述数据处理装置还包括接口配置模块，其设置为配置数据采集接口；

所述适配模块获取待采集数据样本，对所述待采集数据样本进行预处理，包括，所述适配模块通过所述数据采集接口获取待采集数据样本，对所述待采集数据样本进行预处理；

所述数据处理模块控制集群中每个节点进行数据采集和数据处理，包括，所述数据处理模块控制集群中每个节点通过所述数据采集接口进行数据采集，控制集群中每个节点对采集到的数据进行数据处理。

14、如权利要求 13 所述数据处理装置，其中，所述接口配置模块包括：参数获取单元，设置为获取数据采集接口配置参数；

连接测试单元，设置为根据所述配置参数进行所述数据采集接口的连接测试；以及

文件分发单元，设置为若所述连接测试成功，则将包含有所述配置参数的数据采集接口配置文件向集群中的其他节点进行分发，配置所述集群中的每个节点，并控制所述集群中的每个节点连接所述数据采集接口；

所述参数获取单元还设置为，若所述连接测试不成功，则获取所述数据采集接口配置参数。

15、一种数据处理系统，包括适配处理器、数据处理器和数据统一融合处理器，其中：

所述适配处理器，设置为获取待采集数据样本，对所述待采集数据样本进行预处理，得到数据采集模式、数据处理模型和集群中每个节点的采集任务；

所述数据处理器，设置为根据所述数据采集模式、数据处理模型和集群中每个节点的采集任务，控制集群中每个节点进行数据采集和数据处理，获取处理后的数据；

5 所述数据统一融合处理器，设置为将所述处理后的数据进行统一融合，获取统一融合后的数据。

16、如权利要求 15 所述的数据处理系统，所述数据处理系统还包括：

数据采集接口适配器，设置为配置数据采集接口；

数据分类存储器，设置为分类存储所述统一融合后的数据；以及

系统管理器，其中，所述系统管理器包括：

10 数据管理模块，设置为管理所述数据采集接口配置文件、数据采集接口的连接测试结果、待采集数据样本、质量评估报告和集群中所有节点的负载情况；

15 资源管理模块，设置为管理所述待采集数据样本、数据采集、数据处理、数据统一融合和数据存储，及集群资源消耗情况，控制所述集群中所有节点的负载均衡；

模型管理模块，设置为管理数据采集接口配置模板、数据评估模板、数据采集模板、数据处理模型、数据规整模型、数据集成模型和数据建模模型；以及

20 过程监控模块，设置为监控数据采集接口配置、读取待采集数据样本、数据评估、数据处理、数据统一融合和数据存储过程，并记录、处理数据采集接口配置、读取待采集数据样本、数据评估、数据处理、数据统一融合和数据存储过程中的完成状态、中间错误和异常。

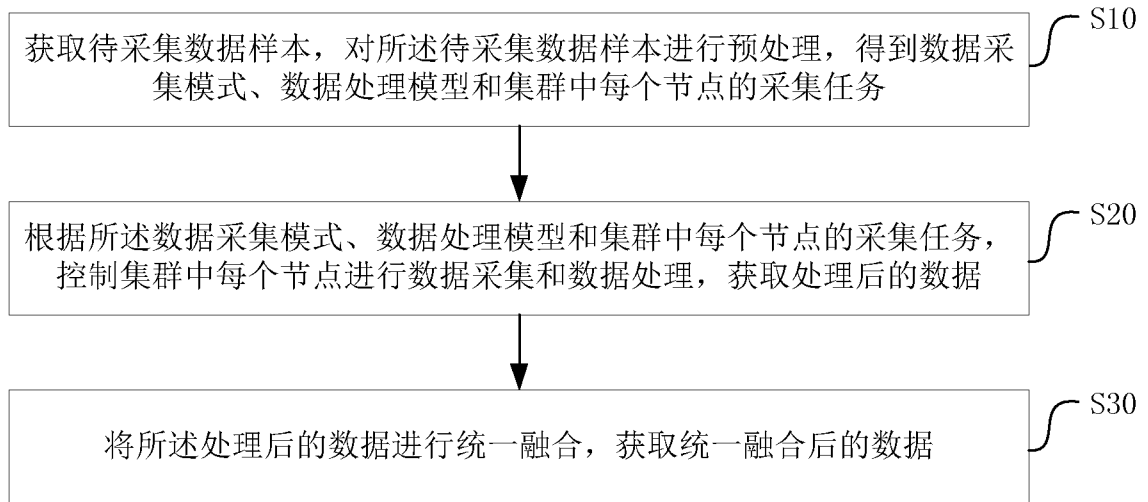


图 1

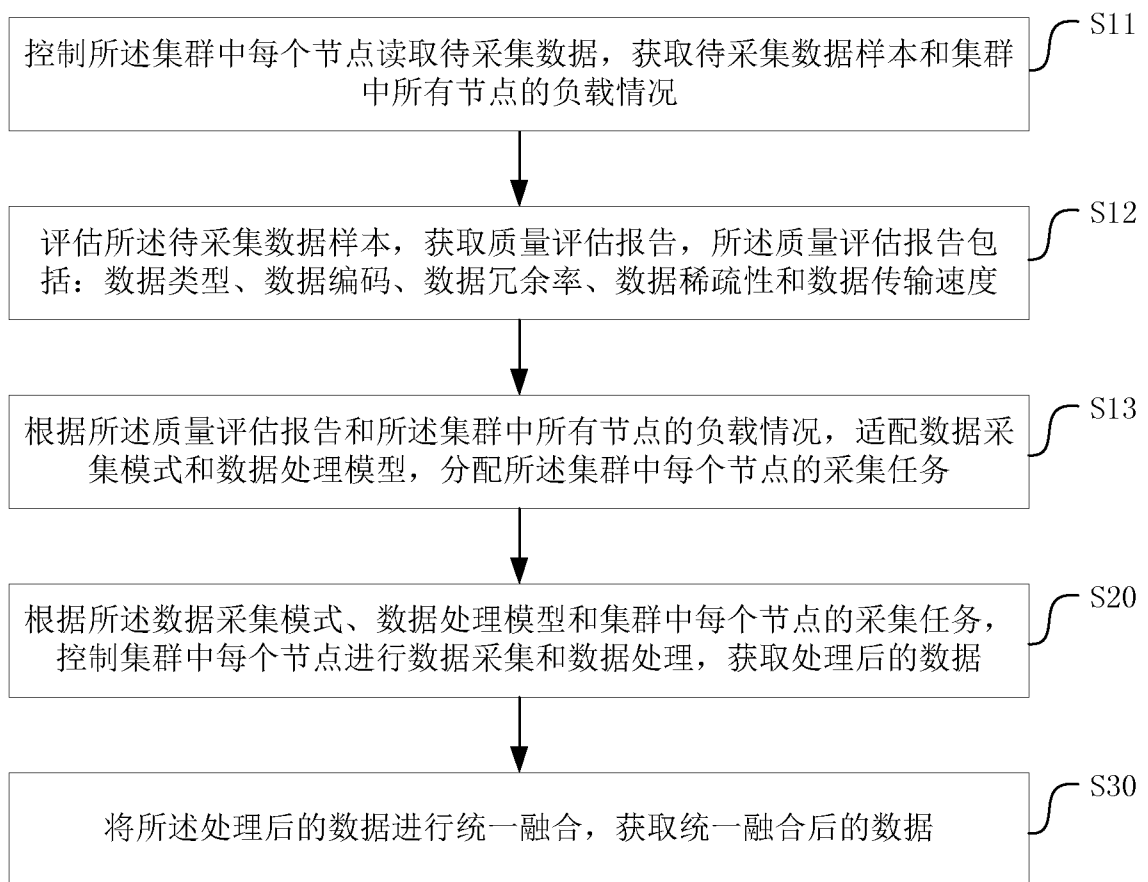


图 2

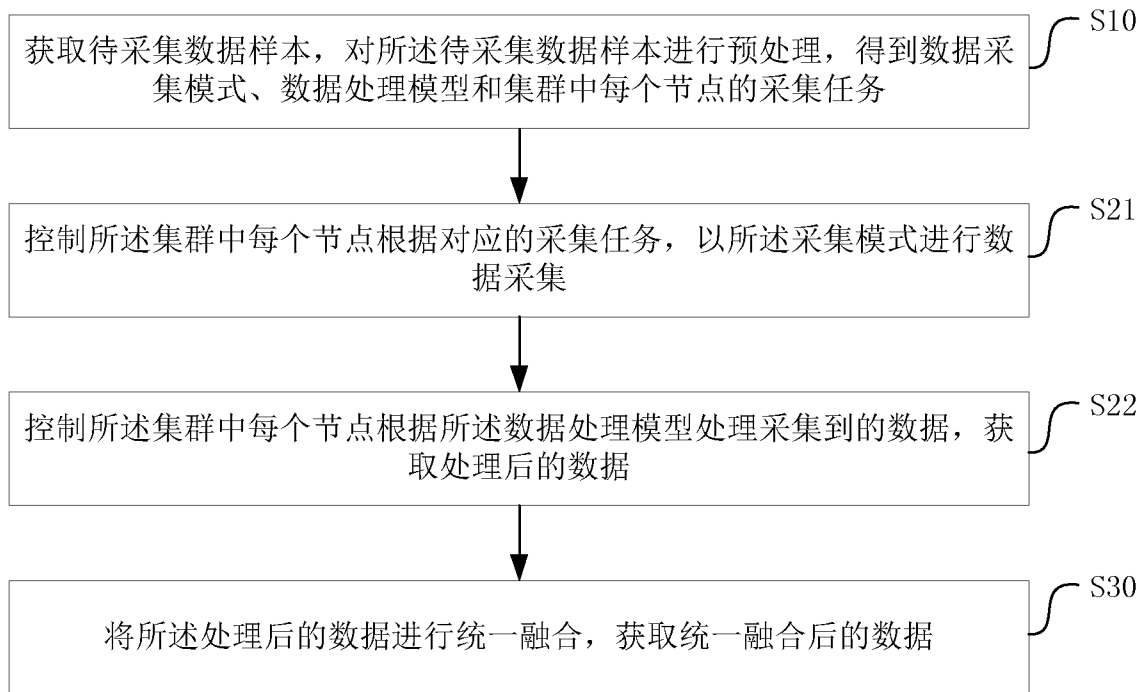


图 3

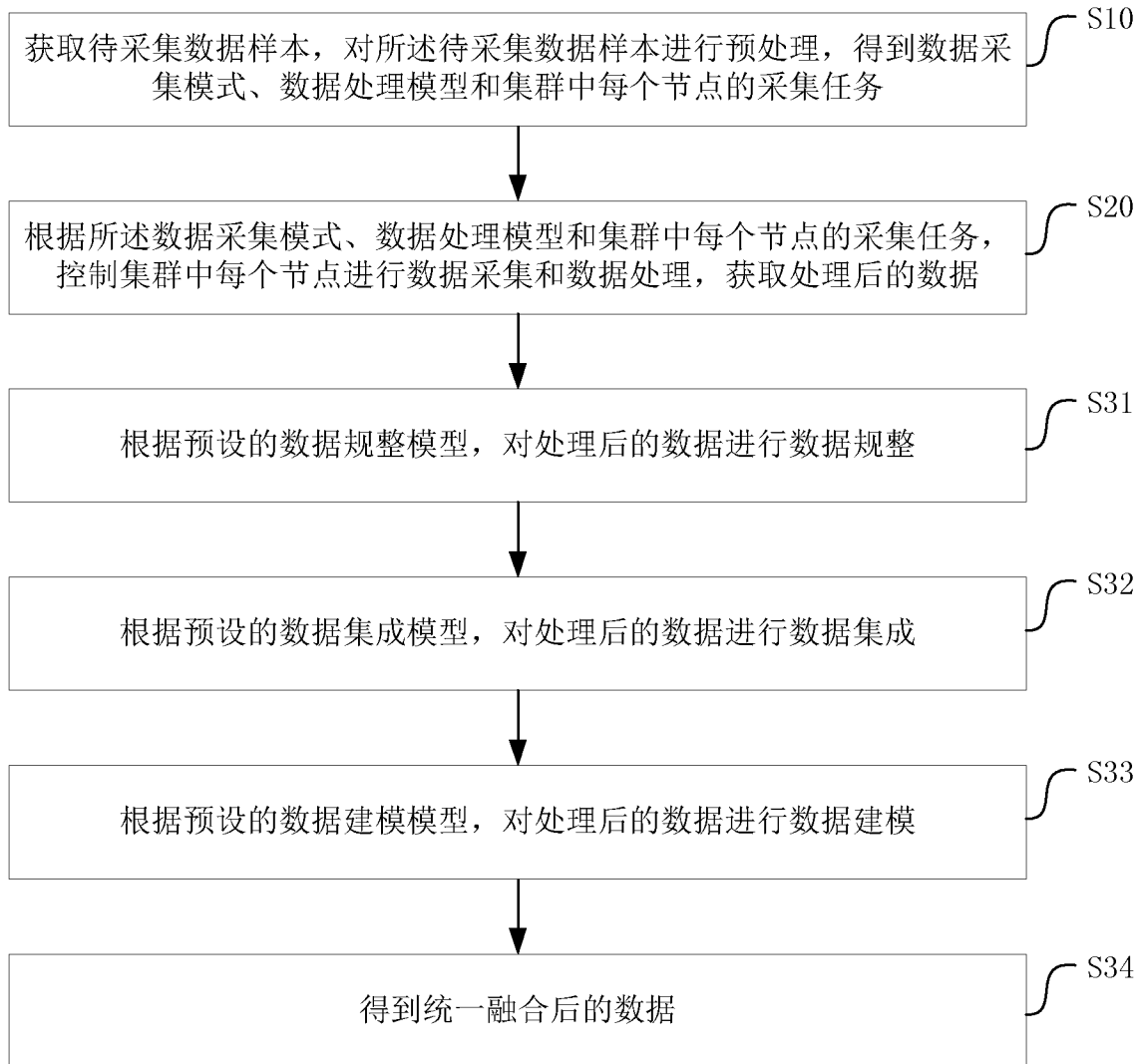


图 4

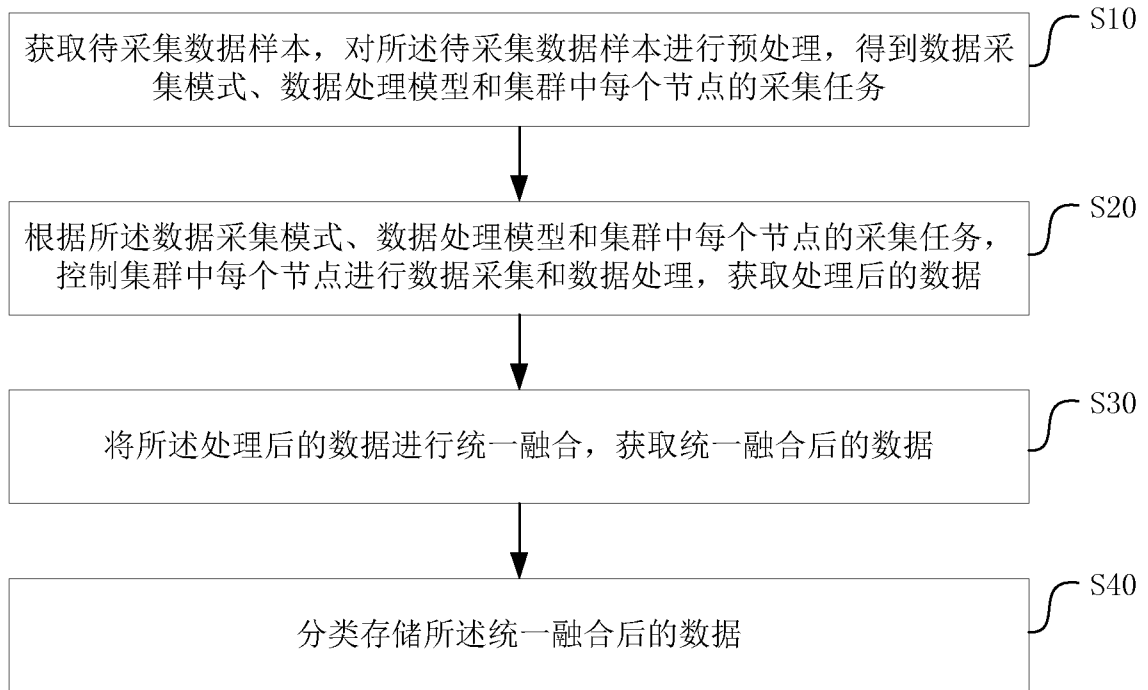


图 5

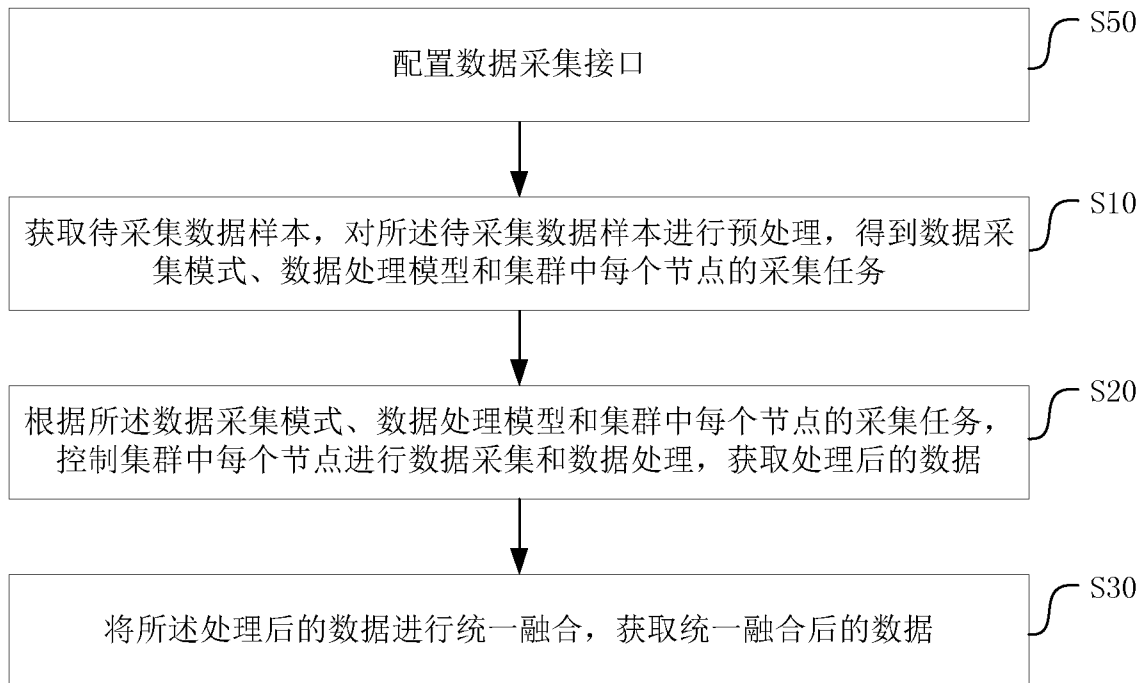


图 6

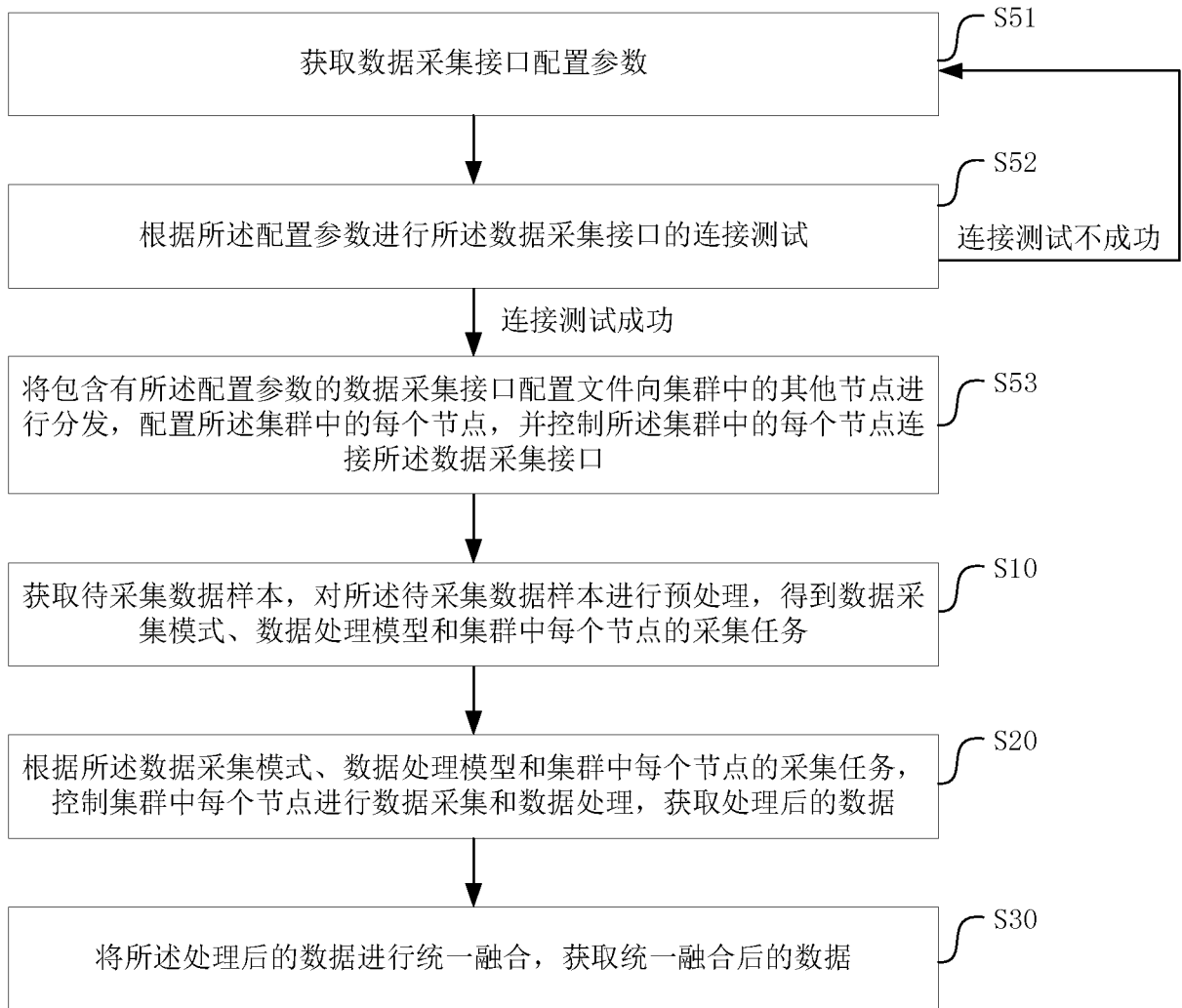


图 7

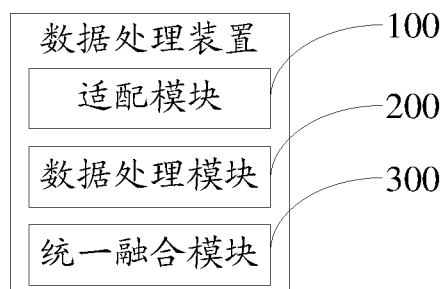


图 8

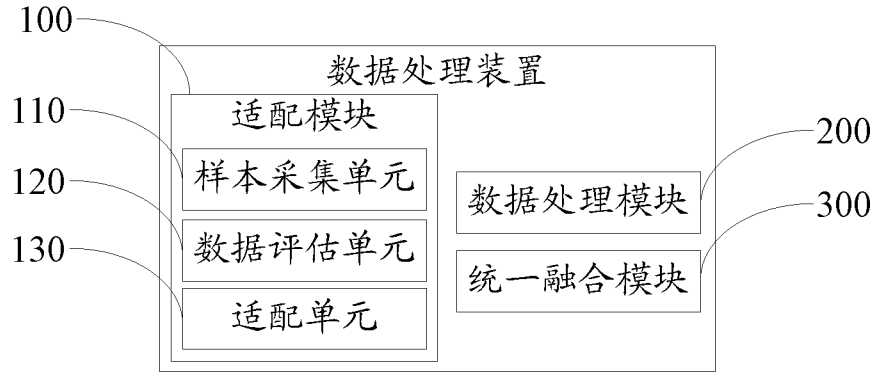


图 9

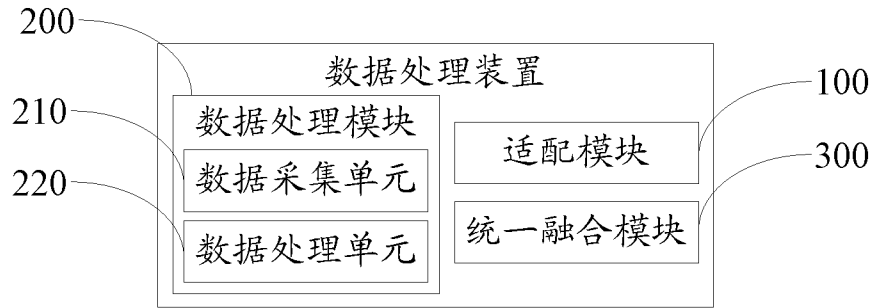


图 10

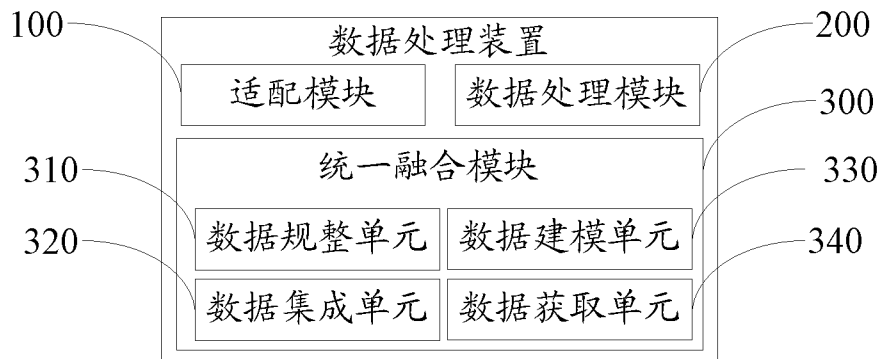


图 11

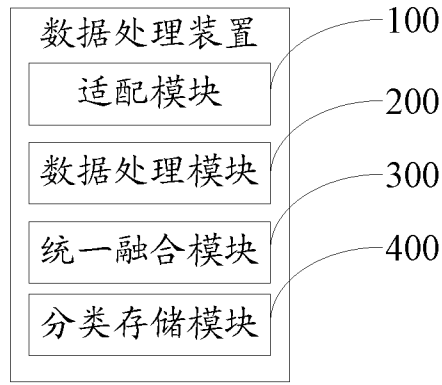


图 12

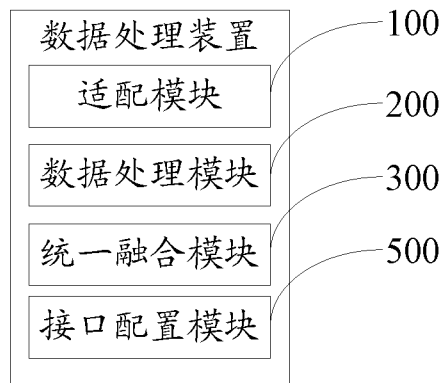


图 13

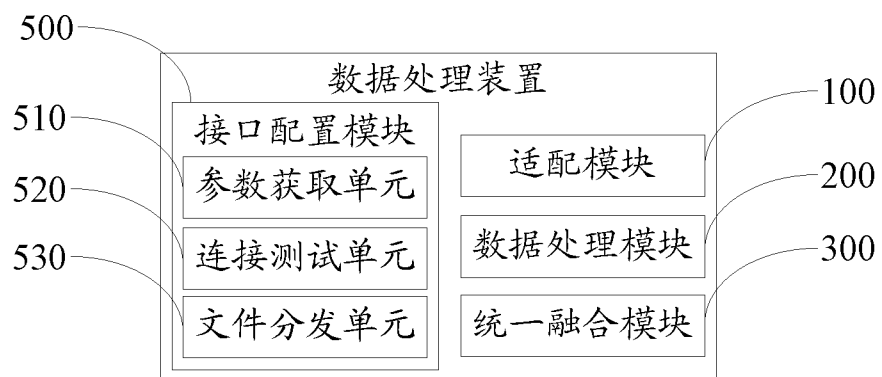


图 14

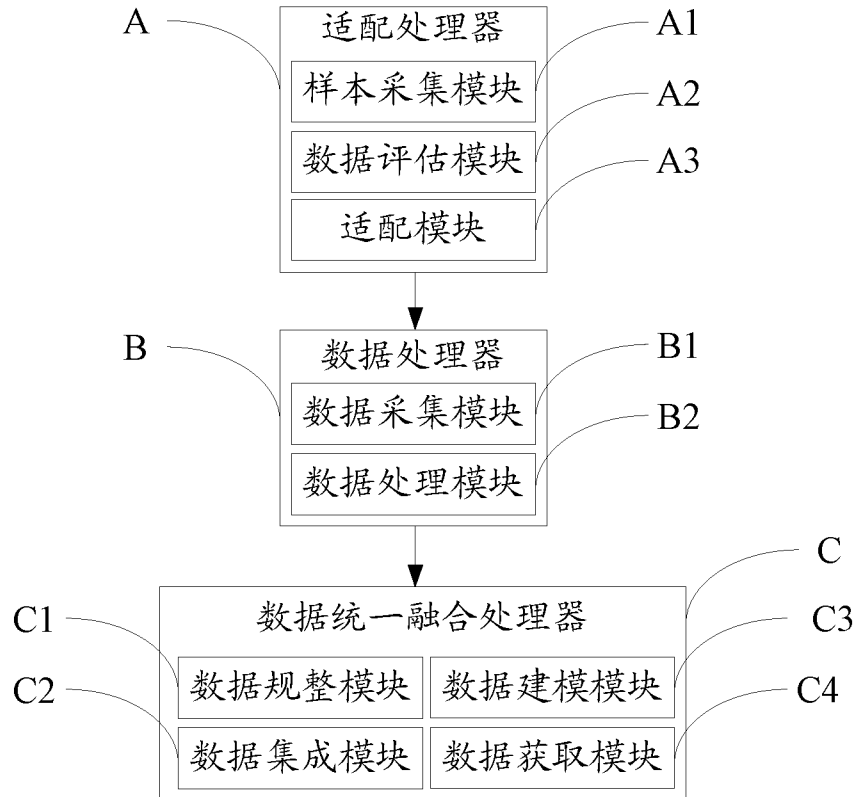


图 15

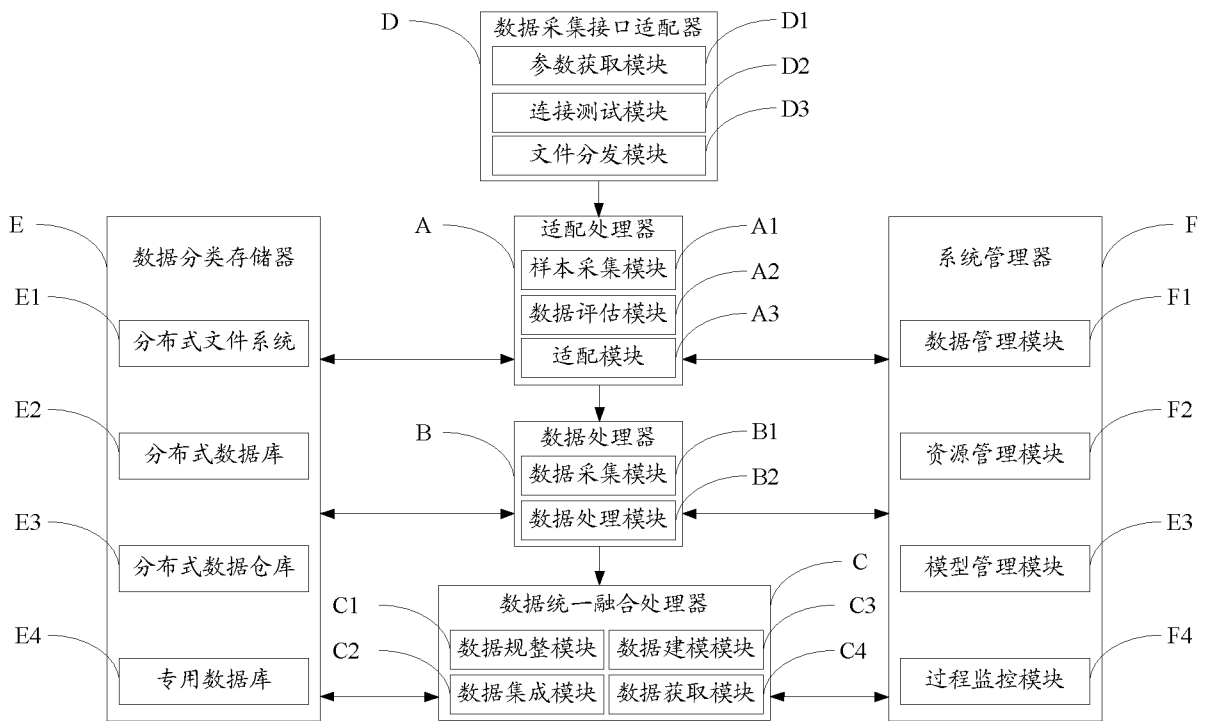


图 16

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/085487

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

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

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

CNPAT, CNKI, WPI, EPODOC: data, process+, sample, collect+, pre-process+, cluster, node

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 104021194 A (INSPUR (BEIJING) ELECTRONIC INFORMATION INDUSTRY CO., LTD.), 03 September 2014 (03.09.2014), description, paragraphs 0080-0150	1-16
A	CN 104765765 A (HANGZHOU BANGSUN FINANCIAL INFORMATION TECHNOLOGY CO., LTD.), 08 July 2015 (08.07.2015), the whole document	1-16
A	CN 103944777 A (GCI SCIENCE & TECHNOLOGY CO., LTD.), 23 July 2014 (23.07.2014), the whole document	1-16
A	CN 103023970 A (COMPUTER NETWORK INFORMATION CENTER, CHINESE ACADEMY OF SCIENCES), 03 April 2013 (03.04.2013), the whole document	1-16
A	US 6701324 B1 (INTERNATIONAL BUSINESS MACHINES CORPORATION), 02 March 2004 (02.03.2004), the whole document	1-16

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“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</p> <p>“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</p> <p>“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</p> <p>“&” document member of the same patent family</p>
---	---

Date of the actual completion of the international search
15 August 2016 (15.08.2016)

Date of mailing of the international search report
30 August 2016 (30.08.2016)

Name and mailing address of the ISA/CN:
State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao
Haidian District, Beijing 100088, China
Facsimile No.: (86-10) 62019451

Authorized officer
LIU, Xuelian
Telephone No.: (86-10) **62414012**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/085487

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104021194 A	03 September 2014	None	
CN 104765765 A	08 July 2015	None	
CN 103944777 A	23 July 2014	None	
CN 103023970 A	03 April 2013	CN 103023970 B	22 July 2015
US 6701324 B1	02 March 2004	None	

<p>A. 主题的分类</p> <p>G06F 17/30 (2006.01)i</p> <p>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</p>																				
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G06F</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNPAT, CNKI, WPI, EPODOC: 数据, 处理, 样本, 采集, 预处理, 集群, 节点, data, process+, sample, collect+, pre-process+, cluster, node</p>																				
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 104021194 A (浪潮北京电子信息产业有限公司) 2014年 9月 3日 (2014 - 09 - 03) 说明书第0080-0150段</td> <td>1-16</td> </tr> <tr> <td>A</td> <td>CN 104765765 A (杭州邦盛金融信息技术有限公司) 2015年 7月 8日 (2015 - 07 - 08) 全文</td> <td>1-16</td> </tr> <tr> <td>A</td> <td>CN 103944777 A (广州杰赛科技股份有限公司) 2014年 7月 23日 (2014 - 07 - 23) 全文</td> <td>1-16</td> </tr> <tr> <td>A</td> <td>CN 103023970 A (中国科学院计算机网络信息中心) 2013年 4月 3日 (2013 - 04 - 03) 全文</td> <td>1-16</td> </tr> <tr> <td>A</td> <td>US 6701324 B1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2004年 3月 2日 (2004 - 03 - 02) 全文</td> <td>1-16</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 104021194 A (浪潮北京电子信息产业有限公司) 2014年 9月 3日 (2014 - 09 - 03) 说明书第0080-0150段	1-16	A	CN 104765765 A (杭州邦盛金融信息技术有限公司) 2015年 7月 8日 (2015 - 07 - 08) 全文	1-16	A	CN 103944777 A (广州杰赛科技股份有限公司) 2014年 7月 23日 (2014 - 07 - 23) 全文	1-16	A	CN 103023970 A (中国科学院计算机网络信息中心) 2013年 4月 3日 (2013 - 04 - 03) 全文	1-16	A	US 6701324 B1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2004年 3月 2日 (2004 - 03 - 02) 全文	1-16
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
A	CN 104021194 A (浪潮北京电子信息产业有限公司) 2014年 9月 3日 (2014 - 09 - 03) 说明书第0080-0150段	1-16																		
A	CN 104765765 A (杭州邦盛金融信息技术有限公司) 2015年 7月 8日 (2015 - 07 - 08) 全文	1-16																		
A	CN 103944777 A (广州杰赛科技股份有限公司) 2014年 7月 23日 (2014 - 07 - 23) 全文	1-16																		
A	CN 103023970 A (中国科学院计算机网络信息中心) 2013年 4月 3日 (2013 - 04 - 03) 全文	1-16																		
A	US 6701324 B1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2004年 3月 2日 (2004 - 03 - 02) 全文	1-16																		
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																				
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&” 同族专利的文件</p>																				
<p>国际检索实际完成的日期</p> <p>2016年 8月 15日</p>		<p>国际检索报告邮寄日期</p> <p>2016年 8月 30日</p>																		
<p>ISA/CN的名称和邮寄地址</p> <p>中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>		<p>授权官员</p> <p>刘雪莲</p> <p>电话号码 (86-10)62414012</p>																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/085487

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	104021194	A	2014年 9月 3日	无	
CN	104765765	A	2015年 7月 8日	无	
CN	103944777	A	2014年 7月 23日	无	
CN	103023970	A	2013年 4月 3日	CN	103023970 B 2015年 7月 22日
US	6701324	B1	2004年 3月 2日	无	