(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
2018 年 4 月 5 日 (05.04.2018)

WIPO | PCT

(10) 国际公布号
WO 2018/059310 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

(21) 国际申请号: PCT/CN2017/102853

(22) 国际申请日: 2017 年 9 月 22 日 (22.09.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201610875245.8 2016年9月30日 (30.09.2016) CN

(71) 申请人: 阿里巴巴集团控股有限公司(ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层847号邮箱, Grand Cayman (KY)。

(72) 发明人; 及

(71) 申请人 (仅对*US*): 李国忠(LI, Guozhong) [CN/CN]; 中国浙江省杭州市余杭区文一西路969号3号楼5楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京国昊天诚知识产权代理有限公司(CO-HORIZON INTELLECTUAL PROPERTY INC.); 中国北京市朝阳区小关北里甲2号渔阳置业大厦B座605, Beijing 100029 (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, 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,

(54) **Title:** METHOD AND DEVICE FOR DATA PROCESSING AND QUERYING

(54) **发明名称**: 数据处理和查询方法及装置

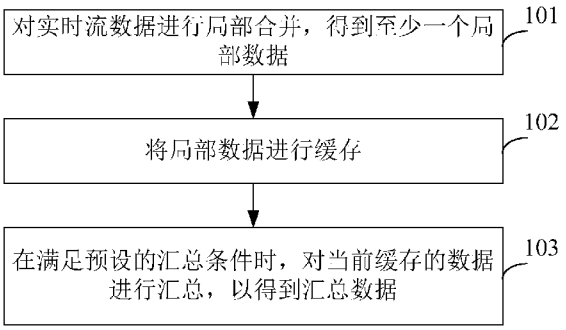


图 1

101 OBTAIN AT LEAST ONE PIECE OF LOCAL DATA BY PERFORMING LOCAL MERGING ON REAL-TIME STREAMING DATA
102 CACHE THE LOCAL DATA
103 PERFORM SUMMARIZATION ON THE CURRENTLY CACHED DATA TO OBTAIN SUMMARIZED DATA UPON SATISFYING A PRESET SUMMARIZATION CONDITION

(57) Abstract: A method and device for data processing and querying. The method comprises: obtaining at least one piece of local data by performing local merging on real-time streaming data (101); caching the local data (102); and upon satisfying a preset summarization condition, performing summarization on the currently cached data to obtain summarized data (103). Whenever the summarization condition is satisfied, the method and device perform summarization on the cached local data, thus reducing the amount of data stored at the physical storage and reducing resources occupied at the physical storage, and further reducing the requirement for performing real-time accumulation of the local data, thereby reducing pressure when calculating the final aggregation of data.

WO 2018/059310 A1

[见续页]

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)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：一种数据处理和查询方法及装置，通过对实时流数据进行局部合并，得到至少一个局部数据（101），将局部数据进行缓存（102），在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到汇总数据（103）。所述方法及装置每当满足汇总条件时，就对缓存的局部数据进行汇总，从而减少在落地存储的数据量，减轻了落地存储占用的资源，而且减轻了对局部数据进行实时累积的要求，减小了最终聚合的计算压力。

数据处理和查询方法及装置

技术领域

本发明属于数据处理领域，尤其涉及一种数据处理和查询方法及装置。

背景技术

在实时流数据落地存储的过程中，首先需要将一个时间间隔内产生的流数据需要进行一次局部合并，生成一个局部数据，然后将读取前面的累加数据，当前的流数据对应的局部数据与前面的累积数据再次进行累加，直到得到最后的累加数据。比如统计一分钟的销售额，可以每隔6秒算一个6s的销售总额，6s对应的销售总额就是一个局部数据，当到达第2次6s时，需要把第一次的6s对应的局部数据从落地的存储端中读取出来后，与第2个6s的局部数据进行累加。现有的实时流数据落地方法，整个落地过程需要内存对流数据的局部数据进行实时累积，计算压力较大，而且需要先读取一次数据再落地一次数据，耗费时间较长。

为了解决上述问题，现有技术中提出了一种多版本并发控制（Multi-Version Concurrency Control，简称MVCC）方案，在该方案中将每一个时间间隔内的流数据进行局部合并，生成一个局部数据，将该局部数据落地到存储端，例如，需要查询一天24小时内的销售总额时，从00：00开始每隔6s进行一个局部合并，得到一个局部数据，在到达设定的24：00时，这24小时内生成的所有局部数据个数为14400，而针对海量或者较长时间的流数据处理时，由于MVCC方案中的存储端中会存储了所有时间间隔对应的局部数据，导致存储端的存储压力较大。

发明内容

本发明提供一种数据处理和查询方法及装置，用于解决现有通过MVCC 方案进行流数据落地时由于存储端存储所有的局部数据，存储压力较大的问题。

为了实现上述目的，本发明提供了一种数据处理方法，包括：

对实时流数据进行局部合并，得到至少一个局部数据；

将所述局部数据进行缓存；

在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到汇总数据；其中所述当前缓存的数据至少包括所述局部数据。

为了实现上述目的，本发明提供了一种数据查询方法，包括：

接收客户端发送的查询请求；

查询接收到所述查询请求时的汇总数据以及局部数据；其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据；

将接收到所述查询请求时的汇总数据以及局部数据进行聚合，以得到聚合数据；

将所述聚合数据发送给所述客户端。

为了实现上述目的，本发明提供了一种数据查询方法，包括：

接收客户端发送的查询请求；

查询接收到所述查询请求时的汇总数据以及局部数据；其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据；

将接收到所述查询请求时的汇总数据以及局部数据发送给所述客户端。

为了实现上述目的，本发明提供了一种数据查询方法，包括：

向存储端发送查询请求，以使所述存储端根据所述查询请求查询接收到所述查询请求时的汇总数据以及局部数据；其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据；

接收所述存储端发送的所述汇总数据以及所述局部数据；

将所述汇总数据以及所述局部数据进行聚合得到聚合数据。

为了实现上述目的，本发明提供了一种数据处理装置，包括：

局部合并模块，用于对实时流数据进行局部合并，得到至少一个局部数据；

缓存模块，用于将所述局部数据进行缓存；

汇总模块，用于在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到汇总数据；其中所述当前缓存的数据至少包括所述局部数据。

为了实现上述目的，本发明提供了一种数据查询装置，包括：

接收模块，用于接收客户端发送的查询请求；

查询模块，用于查询接收到所述查询请求时的汇总数据以及局部数据；其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据；

聚合模块，用于将接收到所述查询请求时的汇总数据以及局部数据进行聚合，以得到聚合数据；

发送模块，用于将所述聚合数据发送给所述客户端。

为了实现上述目的，本发明提供了一种数据查询装置，包括：

接收模块，用于接收客户端发送的查询请求；

查询模块，用于查询接收到所述查询请求时的汇总数据以及局部数

据；其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据；

发送模块，用于将接收到所述查询请求时的汇总数据以及局部数据发送给所述客户端。

为了实现上述目的，本发明提供了一种数据查询装置，包括：

发送模块，用于向存储端发送查询请求，以使所述存储端根据所述查询请求查询接收到所述查询请求时的汇总数据以及局部数据；其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据；

接收模块，用于接收所述存储端发送的所述汇总数据以及所述局部数据；

聚合模块，用于将所述汇总数据以及所述局部数据进行聚合得到聚合数据。

本发明提供的数据处理和查询方法及装置，通过对实时流数据进行局部合并，以得到至少一个局部数据，将局部数据进行缓存，在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到汇总数据。本实施例每当满足汇总条件时，就对缓存的局部数据进行汇总，从而减少在落地存储的数据量，减轻了落地存储占用的资源，而且减轻了对局部数据进行实时累积的要求，减小了最终聚合的计算压力。

附图说明

图 1 为本发明实施例一的数据处理方法的流程示意图；

图 2 为本发明实施例二的数据查询方法的流程示意图；

图 3 为本发明实施例三的数据查询方法的流程示意图；

图 4 为本发明实施例四的数据查询方法的流程示意图；

图 5 为本发明实施例五的数据处理系统的示意图；

图 6 为本发明实施例六的数据处理装置的结构示意图；

图 7 为本发明实施例七的数据查询装置的结构示意图；

图 8 为本发明实施例八的数据查询装置的结构示意图；

图 9 为本发明实施例九的数据查询装置的结构示意图。

具体实施方式

下面结合附图对本发明实施例提供的数据处理和查询方法及装置进行详细描述。

实施例一

如图 1 所示，其为本发明实施例一的数据处理方法的流程示意图。本实施例中，执行主体为存储端。该数据处理方法包括以下步骤：

S101、对实时流数据进行局部合并，得到至少一个局部数据。

本实施例中，可以对实时流数据按照特定的分割方式进行分段，然后将每个分段内的流数据进行局部合并，得到至少一个局部数据。优选地，预设设置一个时间间隔，每当到达时间间隔时，存储端将该时间间隔内所有的流数据进行一个局部合并，即将该时间间隔内的所有流数据进行相加，得到该时间间隔内所有流数据对应的局部数据。比如计算每一分钟的实时销售额，设定每 5s 做一次合并，将该 5s 内的所有流数据进行一个相加汇总，生成一个局部的销售额，该局部的销售额就是局部数据，这样最后在存储端每一分钟就有 60/5=12 个局部的销售额。

S102、将局部数据进行缓存。

在获取到一个时间间隔的所有流数据对应的局部数据后，存储端可以对该局部数据进行缓存。

S103、在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到汇总数据。

其中，所述当前缓存的数据中至少包括局部数据。

本实施例中，预先设置有汇总条件，每当满足汇总条件时，存储端则可以触发对当前缓存的数据进行汇总计算，得到汇总数据。

一种可选地对缓存的数据进行汇总的方式为：

优选地，汇总条件可以为预设的时间长度，即每当到达该预设的时间长度后，存储端对当前已经缓存的所有数据进行一次汇总，得到一个汇总数据。此处需要说明，当初始满足预设的汇总条件时，当前缓存的数据均为局部数据，而从第二次满足预设的汇总条件开始，当前缓存的数据中包括上一次的汇总数据和当前汇总条件下的局部数据。其中当前汇总条件下的局部数据为并不包括上一次汇总条件下缓存的局部数据。例如，预设的时间长度可以为 1 小时，首先从 0 点开始当到达 1 点后，存储端可以将这一个小时内缓存的局部数据进行一次汇总，得到一个汇总数据，从 1 点开始当到达到 2 点后，存储端需要再进行一次汇总，而当前缓存的数据中不仅包括 1 点到 2 点一个小时内所有流数据按照预设的时间间隔生成的局部数据，还包括前一个小时汇总后的汇总数据，此时存储端需要将上一次的汇总数据以及当前缓存的局部数据进行一次汇总计算，得到最新的汇总数据作为当前的汇总数据。

此处需要说明，当预设的汇总条件为预设的时间长度时，从 0 点开始到达 1 点后，以每个 6s 对流数据进行一次局部合并为例，则预设的时间长度内包括 600 个局部数据。从 1 点开始，每次存储端进行下一次汇总之前，存储端中当前缓存的数据包括 1 个上一次汇总数据，以及按照预设的时间间隔 1 小时内生成 600 个局部数据，从而存储端中当前缓存的数据个数最多为 601，减少了存储端的存储压力。

可选地，汇总条件可以为预设的缓存数据的个数，即每当存储端当前缓存的数据个数到达预设的个数时，存储端就可以对已经缓存的数据进行一次汇总，得到一个汇总数据。此处需要说明，当初始满足预设的汇总条件时，当前缓存的数据均为局部数据，而从第二次满足预设的汇总条件开始，当前缓存的数据中包括上一次的汇总数据和当前缓存的局部数据。例如，预设的个数为 10，首先从缓存的数据个数为 0 开始，当缓存的数据个数第一次到达 10 个时，存储端中当前缓存的 10 数据均为局部数据，对 10 个局部数据进行汇总计算，以得到一个汇总数据。本实施例中，在汇总计算后，将经过汇总后的数据进行清除，缓存得到的汇总数据，也就是说，在汇总计算后存储端缓存的数据就只剩下 1 个汇总数据，当缓存到存储到的局部数据数量到达 9 个时，此时就可以满足预设的个数这一汇总条件，存储端则对 1 个汇总数据和当前缓存的 9 个局部数据进行汇总，得到一个新的汇总数据作为当前的汇总数据。

此处需要说明，当预设的汇总条件为缓存数据的个数时，从从缓存的数据个数为 0 开始开，以缓存的数据个数为 10 为例，则当存储端缓存的数据个数到达 10 个时，存储端进行一次汇总。在汇总计算后存储端缓存的数据就只剩下 1 个汇总数据，当缓存到存储到的局部数据数量到达 9 个时，此时存储端进行一次汇总，从而存储端中当前缓存的数据个数最多为 10，减少了存储端的存储压力。

进一步地，本实施例中在对当前的缓存数据进行汇总之后，需要将经过汇总的局部数据进行清除，将汇总数据进行缓存，等下一次满足汇总条件时，需要将汇总数据和下一汇总条件下的局部数据进行汇总。

另一种可选对当前缓存的数据进行汇总的方式为：

每当满足预设的汇总条件后，将该汇总条件下的局部数据进行汇总计算，得到一个该汇总条件下的汇总数据，该汇总条件下的汇总数据为

当前的汇总数据。在该汇总条件下的局部数据进行汇总后，将这些局部数据进行清除，只缓存该汇总条件下的汇总数据。优选地，预设的汇总条件为时间长度时，每当到达该时间长度时，就将该时间长度下的局部数据进行汇总，得到该时间长度下的汇总结果。在该汇总方式下，每个时间长度均生成一个汇总结果。

可选地，预设的汇总条件为缓存的局部数据的个数，每当缓存的局部数据的个数达到预设的个数时，将当前汇总条件下的这些局部数据进行汇总计算，得到当前汇总条件下的汇总数据作为当前的汇总数据。每当缓存的局部数据个数达到预设的个数，就可以得到一个汇总数据，这些汇总数据均缓存在存储端。

本实施例提供的数据查询方法，通过获取预设的时间间隔内所有流数据的局部数据，将局部数据进行缓存，在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到汇总数据。本实施例每当满足汇总条件时，就对缓存的局部数据进行汇总，从而减少在落地存储的数据量，减轻了落地存储占用的资源，而且减轻了对局部数据进行实时累积的要求，减小了最终聚合的计算压力。

实施例二

如图 2 所示，其为本发明实施例二的数据查询方法的流程示意图。本实施例中，执行主体为存储端。该数据查询方法包括以下步骤：

S201、接收客户端发送的查询请求。

在上述实施例一的基础之上，用户试图查询当前的聚合数据时，可以基于客户端向存储端发送查询请求，存储端可以设置一个查询接口，通过该查询接口接收该查询请求。

S202、查询接收到查询请求时的汇总数据以及局部数据。

在接收到客户端查询请求的时刻，存储端对当前缓存的数据进行查

询，以获取在接收到该查询请求时的汇总数据和局部数据。本实施例中，聚合和缓存时异步执行的，即在缓存局部数据的同时，可以对缓存数据进行聚合操作，从而可以保证用户可以实时地查询获取当前流数据的聚合数据。

S203、将接收到查询请求时的汇总数据以及局部数据进行聚合，以得到聚合数据。

S204、将聚合数据发送给客户端。

在获取到查询请求时的汇总数据以及局部数据后，为了使用户得到聚合数据，存储端需要对该查询请求时的汇总数据以及局部数据进行一次聚合操作，得到一个聚合数据，将该聚合数据发送给客户端。

在上述实施例一中的一种对缓存的数据进行汇总的情况下，接收到查询请求时的汇总数据为一个，然后将接收到查询请求时的汇总数据与查询请求时的局部数据进行聚合，得到聚合数据发送给客户端。

在上述实施例一中的另一种对缓存的数据进行汇总的情况下，每当满足汇总条件时就会产生一个汇总数据并将其缓存在存储端，因此接收到查询请求时的汇总数据为至少一个，将接收到查询请求时所有的汇总数据与查询请求时的局部数据进行聚合，得到聚合数据发送给客户端。

本实施例提供的数据查询方法，通过每当满足汇总条件时，存储端对缓存的数据进行汇总，从而减少落地存储的数据量，减轻了落地存储占用的资源，减轻了存储端的存储压力。而且存储端进行汇总时，不再需要对局部数据进行读取后再落到，节省时间。在用户需要对流数据进行聚合时，通过查询请求触发存储端对当前缓存的数据进行聚合，得到聚合数据，本实施例由于落地存储的数据量变少，从而降低存储端的聚合计算的压力。

实施例三

如图 3 所示，其为本发明实施例三的数据查询方法的流程示意图。本实施例中，执行主体为存储端。该数据查询方法包括以下步骤：

S301、接收客户端发送的查询请求。

在上述实施例一的基础之上，用户试图查询当前的聚合数据时，可以基于客户端向存储端发送查询请求，存储端可以设置一个查询接口，通过该查询接口接收该查询请求。

S302、查询接收到查询请求时的汇总数据以及局部数据。

在接收到客户端查询请求的时刻，存储端对当前缓存的数据进行查询，以获取在接收到该查询请求时的汇总数据和局部数据。本实施例中，聚合和缓存时异步执行的，即在缓存局部数据的同时，可以对缓存数据进行聚合操作，从而可以保证用户可以实时地查询获取当前流数据的聚合数据。

S303、将接收到查询请求时的汇总数据以及局部数据发送给客户端。

在获取到查询请求时的汇总数据以及局部数据后，为了使用户得到聚合数据，存储端可以将获取到查询请求时的汇总数据以及局部数据发送给客户端，以使客户端在本地对接收到查询请求时的汇总数据以及局部数据进行聚合，得到一个聚合数据。

在上述实施例一中的一种对缓存的数据进行汇总的情况下，接收到查询请求时的汇总数据为一个，然后将接收到查询请求时的汇总数据与查询请求时的局部数据发送给客户端。

在上述实施例一中的另一种对缓存的数据进行汇总的情况下，每当满足汇总条件时就会产生一个汇总数据并将其缓存在存储端，因此接收到查询请求时汇总数据为至少一个，将接收到查询请求时所有的汇总数据与查询请求时的局部数据发送给客户端。

S304、接收客户端返回的聚合数据。

其中， 聚合数据是由客户端对汇总数据以及局部数据进行聚合得到的。

本实施例提供的数据查询方法，在用户需要对流数据进行聚合时，通过查询请求从存储端获取当前缓存的数据，在客户端本地对存储端当前缓存的数据进行聚合，得到聚合数据。本实施例由于落地存储的数据量变少，从而减少了在查询时向客户端发送的数据量，减轻了客户端的计算压力。

实施例四

如图 4 所示，其为本发明实施例四的数据查询方法的示意图。本实施例中，执行主体为客户端端。该数据查询方法包括以下步骤：

S401、向存储端发送查询请求，以使存储端根据查询请求查询接收到查询请求时的汇总数据以及局部数据。

其中，汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，局部数据为对实时流数据进行局部合并得到的数据。

在上述实施例一的基础之上，当用户试图查询当前的聚合数据时，可以基于客户端向存储端发送查询请求，存储端可以设置一个查询接口，通过该查询接口接收该查询请求。

在接收到查询请求后，存储端根据该查询请求，查询接收到查询请求时的汇总数据和局部数据。其中，关于汇总数据以及局部数据的介绍，可参见上述实施例中相关内容的记载，此处不再赘述。

S402、接收存储端发送的汇总数据以及局部数据。

S403、将汇总数据以及局部数据进行聚合得到聚合数据。

在获取到查询请求时的汇总数据以及局部数据后，为了使用户得到聚合数据，存储端可以将获取到查询请求时的汇总数据以及局部数据发送给客户端，客户端接收该存储端发送的汇总数据和局部数据。

进一步地，客户端可以在本地对接收到查询请求时的汇总数据以及局部数据进行聚合，得到一个聚合数据。

在上述实施例一中的一种对缓存的数据进行汇总的情况下，接收到查询请求时的汇总数据为一个，然后将接收到查询请求时的汇总数据与查询请求时的局部数据发送给客户端。

在上述实施例一中的另一种对缓存的数据进行汇总的情况下，每当满足汇总条件时就会产生一个汇总数据并将其缓存在存储端，因此接收到查询请求时汇总数据为至少一个，将接收到查询请求时所有的汇总数据与查询请求时的局部数据发送给客户端。

进一步地，客户端在聚合得到聚合数据后，可以将该聚合数据发给存储端进行存储。

本实施例提供的数据查询方法，在用户需要对流数据进行聚合时，客户端可以通过查询请求从存储端获取当前缓存的数据，在客户端本地对存储端当前缓存的数据进行聚合，得到聚合数据。本实施例由于落地存储的数据量变少，从而减少了在查询时向客户端发送的数据量，减轻了客户端的计算压力。

实施例五

如图 5 所示，其为本发明实施例五的数据处理系统的示意图。该数据处理系统包括：实时流数据模块、流计算引擎模块、存储端。其中，实时流数据获取模块可以接收实时的流数据，并将流数据发送给流计算引擎模块中，在该流计算引擎模块中可以按照预设的时间间隔对流数据进行合并，得到该时间间隔对应的一个局部数据。流计算引擎模块将局部数据缓存到存储端。该存储端可以异地地合并多个局部结果，也就是说，存储端可以对缓存的数据进行汇总，得到汇总数据。关于对缓存的数据进行汇总的过程可参见上述实施例中相关内容的记载，此处不再赘

述。

进一步地，存储端可以为客户端提供一个查询接口，各种业务对应的客户端可以通过该查询接口，来实时地查询流数据的聚合数据。其中业务可以包括实时大盘、监控系统以及实时业务报表等。

为了更好地理解本发明实施例提供的数据处理方法，举例进行说明：

当前有 5 笔订单，其中订单就是流数据。

（20160113 13:42:00　10）

（20160113 13:42:10　20）

（20160113 13:42:20　30）

（20160113 13:42:35　40）

（20160113 13:42:50　50）

预设每 15s 由存储端对流数据进行一次合并，得到一个局部数据，此处局部数据为局部销售额，则一分钟就会产生 4 个局部销售额，分别为：0~15s、16~30s、31~45s 和 46~60s 的局部销售额，如下：

原始订单金额	局部销售额
10+20.	30
30	30
40	40
50	50

预设设置有一个汇总条件，如该汇总条件可以为存储端缓存的局部销售额的个数，例如将缓存的个数设置为 3 个，当缓存的局部销售额的个数到达 3 个时，存储端当前缓存的销售额为 30、30 和 40，在触发了汇总后，存储端则需要对 3 个局部销售额进行汇总，即当 40 进行缓存的时候触发了一次存储端的汇总，得到一个汇总结果 100，此时缓存端存储有一个汇总结果 100 和一个局部销售额 50。

当业务方基于客户端直接调用存储端的查询接口查询当前的实时销售额时，存储端的查询接口会做了最终一次的聚合操作，即将 100 和 50 做合并得到 150，然后该 150 再返回给客户端。

本实施例提供的数据处理方法，通过每当满足汇总条件时，存储端对缓存的数据进行汇总，从而减少落地存储的数据量，减轻了落地存储占用的资源，减轻了存储端的存储压力。而且存储端进行汇总时，不再需要对局部数据进行读取后再落到，节省时间。在用户需要对流数据进行聚合时，通过查询请求触发存储端对当前缓存的数据进行聚合，得到聚合数据，本实施例由于落地存储的数据量变少，从而降低存储端的聚合计算的压力。

实施例六

如图 6 所示，其为本发明实施例六的数据处理装置的结构示意图。该数据处理装置包括：局部合并模块 11、缓存模块 12 和汇总模块 13。

其中，局部合并模块 11，用于对实时流数据进行局部合并，得到至少一个局部数据。

本实施例中，局部合并模块 11 可以对实时流数据按照特定的分割方式进行分段，然后将每一段内的流数据进行合并，得到至少一个局部数据。优选地，预设设置一个时间间隔，每当到达时间间隔时，局部合并模块 11 将该时间间隔内所有的流数据进行一个合并，得到该时间间隔内所有流数据对应的局部数据。

缓存模块 12，用于将所述局部数据进行缓存。

在获取到一个时间间隔的所有流数据对应的局部数据后，缓存模块 12 可以对该局部数据做出一次持久化，即将该局部数据进行缓存。

汇总模块 13，用于在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到汇总数据。

本实施例中，预先设置有汇总条件，每当满足汇总条件时，汇总模块 13 则可以触发对当前缓存的数据进行汇总计算，得到汇总数据。

其中，当前缓存的数据可以均为局部数据。

可选地，当前缓存的数据中包括上次满足所述汇总条件时得到的汇总数据以及当前汇总条件下的局部数据。

关于对缓存的数据进行汇总的相关内容，可参见上述实施例中相关内容的记载，此处不再赘述。

进一步地，缓存模块 12，还用于在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到当前的汇总数据之后，清除经过汇总后的所述局部数据，以及缓存汇总得到的所述汇总数据。

缓存模块 12，具体用于将预设的时间间隔内的所有流数据进行合并，以得到局部数据。

本实施例中，汇总条件为预设的时间长度或者缓存的数据个数。

本实施例提供的数据处理装置可用于执行图 1 所示数据处理流程，具体原理可参见上述实施例一中相关记载，此处不再赘述。

本实施例提供的数据处理装置，通过获取预设的时间间隔内所有流数据的局部数据，将局部数据进行缓存，在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到汇总数据。本实施例每当满足汇总条件时，就对缓存的局部数据进行汇总，从而减少在落地存储的数据量，减轻了落地存储占用的资源，而且减轻了对局部数据进行实时累积的要求，减小了最终聚合的计算压力。

实施例七

如图 7 所示，其为本发明实施例七的数据查询装置的结构示意图。该数据查询装置包括：接收模块 21、查询模块 22、聚合模块 23 以及发送模块 24。

其中，接收模块 21，用于接收客户端发送的查询请求。

查询模块 22，用于查询接收到所述查询请求时的汇总数据以及局部数据。

其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据。

其中，汇总条件为预设的时间长度或者缓存的数据个数。

关于汇总数据和局部数据的介绍，可参见上述实施例中相关内容的记载，此处不再赘述。

聚合模块 23，用于将接收到所述查询请求时的汇总数据以及局部数据进行聚合，以得到聚合数据。

发送模块 24，用于将所述聚合数据发送给所述客户端。

本实施例提供的数据查询装置可用于执行图 2 所示数据处理流程，具体原理可参见上述实施例二中相关记载，此处不再赘述。

本实施例提供的数据查询方法，在用户需要对流数据进行聚合时，通过查询请求触发存储端对当前缓存的数据进行聚合，得到聚合数据，由于落地存储的数据量变少，从而降低存储端的聚合计算的压力。

实施例八

图 8 所示，其为本发明实施例八的数据查询装置的结构示意图。该数据查询装置包括：接收模块 31、查询模块 32 以及发送模块 33。

其中，接收模块 31，用于接收客户端发送的查询请求。

查询模块 32，用于查询接收到所述查询请求时的汇总数据以及局部数据。

其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得

到的数据。

发送模块 33，用于将接收到所述查询请求时的汇总数据以及局部数据发送给所述客户端。

进一步地，接收模块 31，还用于在将接收到所述查询请求时的汇总数据以及局部数据发送给所述客户端之后，接收所述客户端返回的所述聚合数据，所述聚合数据是由所述客户端对接收到所述查询请求时的汇总数据以及局部数据进行聚合得到的。

本实施例提供的数据查询装置可用于执行图 3 所示数据处理流程，具体原理可参见上述实施例三中相关记载，此处不再赘述。

本实施例提供的数据查询装置，在用户需要对流数据进行聚合时，通过查询请求从存储端获取当前缓存的数据，在客户端本地对存储端当前缓存的数据进行聚合，得到聚合数据。本实施例由于落地存储的数据量变少，从而减少了在查询时向客户端发送的数据量，减轻了客户端的计算压力。

实施例九

图 9 所示，其为本发明实施例九的数据查询装置的结构示意图。该数据查询装置包括：发送模块 41、接收模块 42 和聚合模块 33。

其中，发送模块 41，用于向存储端发送查询请求，以使所述存储端根据所述查询请求查询接收到所述查询请求时的汇总数据以及局部数据。

其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据。

接收模块 42，用于接收所述存储端发送的所述汇总数据以及所述局部数据。

聚合模块 43，用于将所述汇总数据以及所述局部数据进行聚合得到

聚合数据。

进一步地，发送模块 41，还用于在聚合模块 43 将所述汇总数据以及所述局部数据进行聚合得到聚合数据之后，将所述聚合数据发送给所述存储端进行存储。

本实施例提供的数据查询装置可用于执行图 4 所示数据处理流程，具体原理可参见上述实施例四中相关记载，此处不再赘述。

本实施例提供的数据查询装置，在用户需要对流数据进行聚合时，通过查询请求从存储端获取当前缓存的数据，在客户端本地对存储端当前缓存的数据进行聚合，得到聚合数据。本实施例由于落地存储的数据量变少，从而减少了在查询时向客户端发送的数据量，减轻了客户端的计算压力。

本领域普通技术人员可以理解：实现上述各方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成。前述的程序可以存储于一计算机可读取存储介质中。该程序在执行时，执行包括上述各方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

最后应说明的是：以上各实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述各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：其依然可以对前述各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各实施例技术方案的范围。

权 利 要 求 书

1、一种数据处理方法，其特征在于，包括：

对实时流数据进行局部合并，得到至少一个局部数据；

将所述局部数据进行缓存；

在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到汇总数据；其中所述当前缓存的数据至少包括所述局部数据。

2、根据权利要求 1 所述的方法，其特征在于，所述当前缓存的数据均为局部数据。

3、根据权利要求 1 所述的方法，其特征在于，所述当前缓存的数据中包括上次满足所述汇总条件时得到的汇总数据以及当前汇总条件下的局部数据。

4、根据权利要求 1-3 任一项所述的方法，其特征在于，所述在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到汇总数据之后，还包括：

清除经过汇总后的所述局部数据；

缓存汇总得到的所述汇总数据。

5、根据权利要求 4 所述的方法，其特征在于，所述对实施流数据进行局部合并，得到至少一个局部数据，包括：

将预设的时间间隔内的所有流数据进行合并，以得到所述局部数据。

6、根据权利要求 5 所述的方法，其特征在于，所述汇总条件为预设的时间长度或者缓存的数据个数。

7、一种数据查询方法，其特征在于，包括：

接收客户端发送的查询请求；

查询接收到所述查询请求时的汇总数据以及局部数据；其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到

的数据，所述局部数据为对实时流数据进行局部合并得到的数据；

将接收到所述查询请求时的汇总数据以及局部数据进行聚合，以得到聚合数据；

将所述聚合数据发送给所述客户端。

8、一种数据查询方法，其特征在于，包括：

接收客户端发送的查询请求；

查询接收到所述查询请求时的汇总数据以及局部数据；其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据；

将接收到所述查询请求时的汇总数据以及局部数据发送给所述客户端。

9、根据权利要求 8 所述的方法，其特征在于，所述将接收到所述查询请求时的汇总数据以及局部数据发送给所述客户端之后，还包括：

接收所述客户端返回的所述聚合数据，所述聚合数据是由所述客户端对接收到所述查询请求时的汇总数据以及局部数据进行聚合得到的。

10、一种数据查询方法，其特征在于，包括：

向存储端发送查询请求，以使所述存储端根据所述查询请求查询接收到所述查询请求时的汇总数据以及局部数据；其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据；

接收所述存储端发送的所述汇总数据以及所述局部数据；

将所述汇总数据以及所述局部数据进行聚合得到聚合数据。

11、根据权利要求 10 所述的方法，其特征在于，所述将所述汇总数据以及所述局部数据进行聚合得到聚合数据之后，包括：

将所述聚合数据发送给所述存储端进行存储。

12、一种数据处理装置，其特征在于，包括：

局部合并模块，用于对实时流数据进行局部合并，得到至少一个局部数据；

缓存模块，用于将所述局部数据进行缓存；

汇总模块，用于在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到汇总数据；其中所述当前缓存的数据至少包括所述局部数据。

13、一种数据查询装置，其特征在于，包括：

接收模块，用于接收客户端发送的查询请求；

查询模块，用于查询接收到所述查询请求时的汇总数据以及局部数据；其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据；

聚合模块，用于将接收到所述查询请求时的汇总数据以及局部数据进行聚合，以得到聚合数据；

发送模块，用于将所述聚合数据发送给所述客户端。

14、一种数据查询装置，其特征在于，包括：

接收模块，用于接收客户端发送的查询请求；

查询模块，用于查询接收到所述查询请求时的汇总数据以及局部数据；其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据；

发送模块，用于将接收到所述查询请求时的汇总数据以及局部数据发送给所述客户端。

15、根据权利要求 14 所述的装置，其特征在于，所述接收模块，还

用于接收所述客户端返回的所述聚合数据，所述聚合数据是由所述客户端对接收到所述查询请求时的汇总数据以及局部数据进行聚合得到的。

16、一种数据查询装置，其特征在于，包括：

发送模块，用于向存储端发送查询请求，以使所述存储端根据所述查询请求查询接收到所述查询请求时的汇总数据以及局部数据；其中，所述汇总数据为在满足预设的汇总条件时，对当前缓存的数据进行汇总得到的数据，所述局部数据为对实时流数据进行局部合并得到的数据；

接收模块，用于接收所述存储端发送的所述汇总数据以及所述局部数据；

聚合模块，用于将所述汇总数据以及所述局部数据进行聚合得到聚合数据。

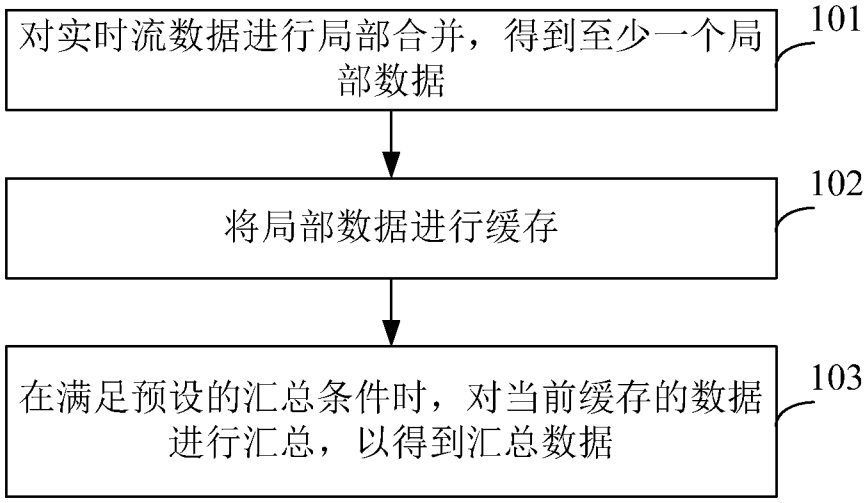

对实时流数据进行局部合并，得到至少一个局部数据
101
将局部数据进行缓存
102
在满足预设的汇总条件时，对当前缓存的数据进行汇总，以得到汇总数据
103

图 1

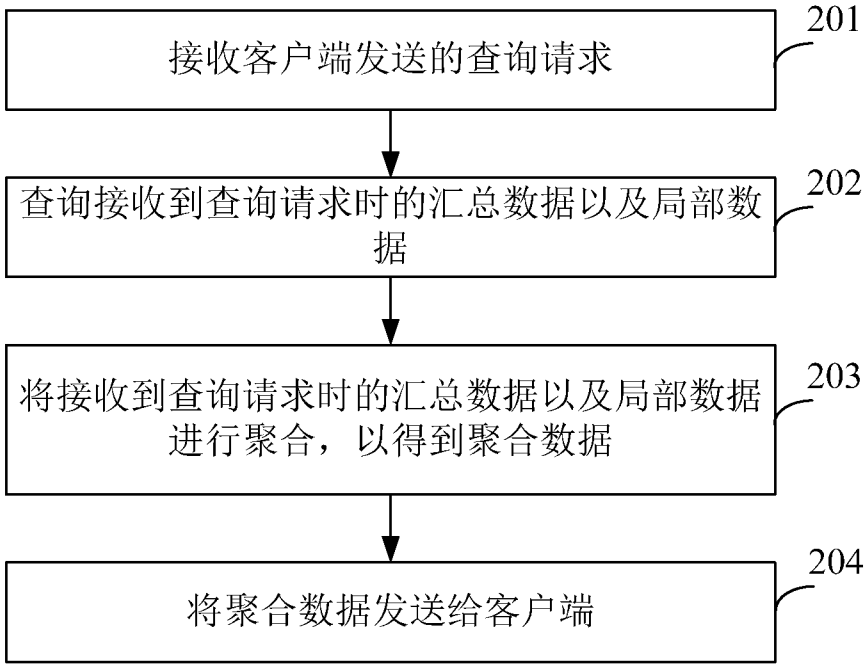

接收客户端发送的查询请求
201
查询接收到查询请求时的汇总数据以及局部数据
202
将接收到查询请求时的汇总数据以及局部数据进行聚合，以得到聚合数据
203
将聚合数据发送给客户端
204

图 2

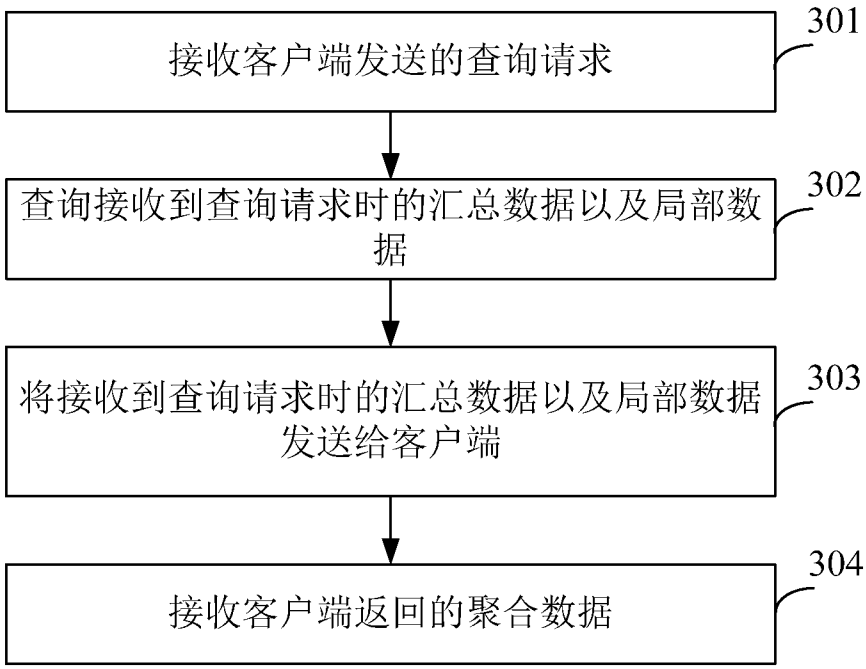

接收客户端发送的查询请求
301
查询接收到查询请求时的汇总数据以及局部数据
302
将接收到查询请求时的汇总数据以及局部数据发送给客户端
303
接收客户端返回的聚合数据
304

图 3

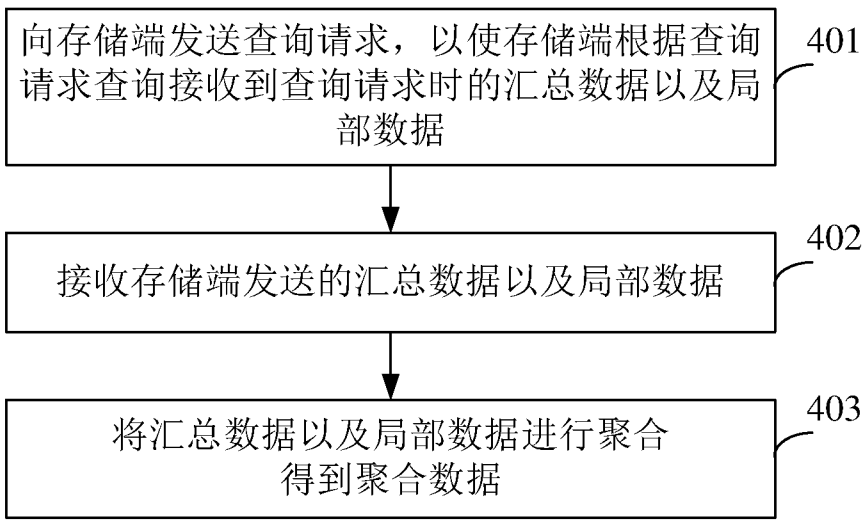

向存储端发送查询请求，以使存储端根据查询请求查询接收到查询请求时的汇总数据以及局部数据
401
接收存储端发送的汇总数据以及局部数据
402
将汇总数据以及局部数据进行聚合得到聚合数据
403

图 4

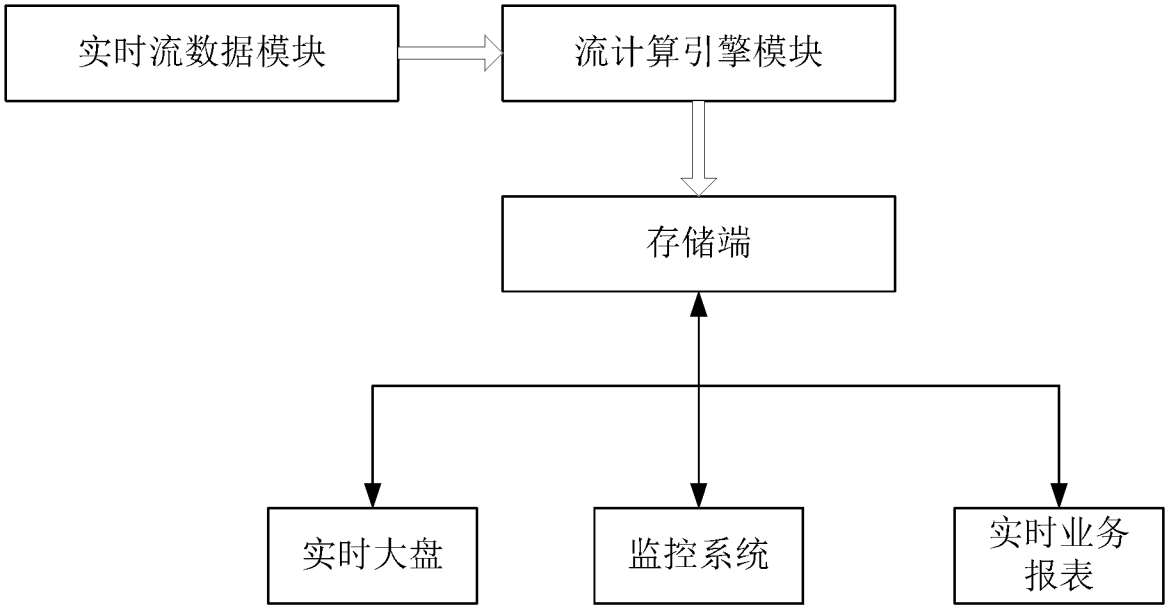

实时流数据模块
流计算引擎模块
存储端
实时大盘
监控系统
实时业务报表

图 5

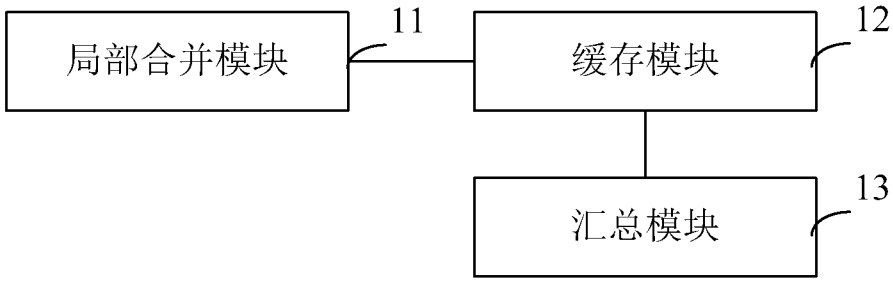

局部合并模块
11
缓存模块
12
汇总模块
13

图 6

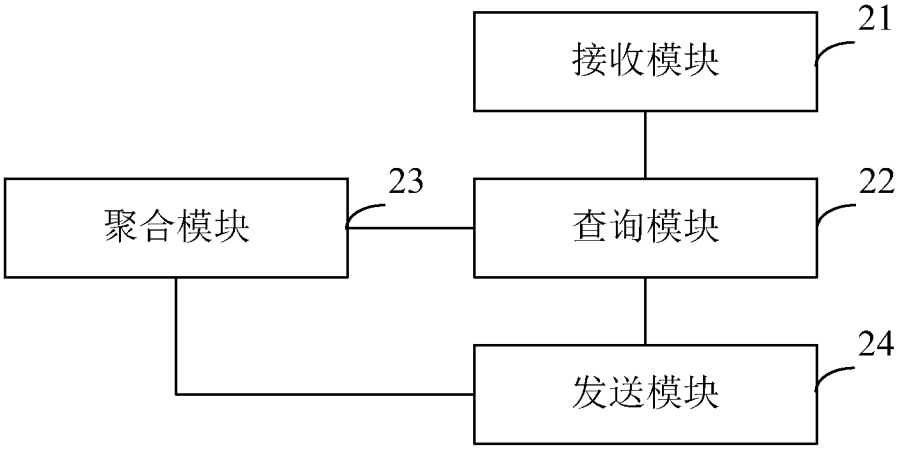

接收模块
21
聚合模块
23
查询模块
22
发送模块
24

图 7

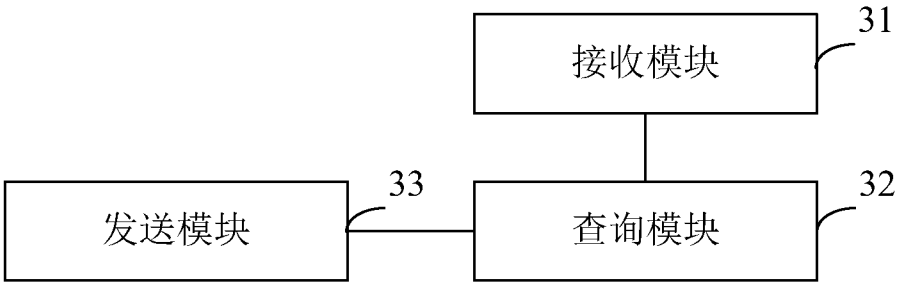


图 8

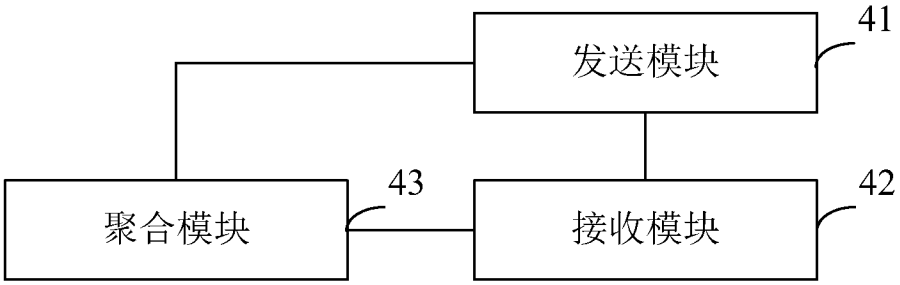


图 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/102853

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, EPODOC, WPI, CNKI: 数据, 流, 查询, 查找, 存储, 缓存, 合并, 汇总, 汇合, stream, flow, data, inquire, ql, storage, buffer, consolidation, union, collateral

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 105515842 A (COLASOFT CO., LTD.), 20 April 2016 (20.04.2016), description, paragraphs [0008]-[0026], and figures 1-2	1-16
A	CN 104216766 A (HUAWEI TECHNOLOGIES CO., LTD.), 17 December 2014 (17.12.2014), entire document	1-16
A	CN 101997701 A (ZTE CORP.), 30 March 2011 (30.03.2011), entire document	1-16
A	WO 2004068469 A2 (TACIT NETWORKS, INC. et al.), 12 August 2004 (12.08.2004), entire document	1-16

☐ 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
08 December 2017	27 December 2017
Name and mailing address of the ISA 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 SUN, Juan Telephone No. (86-10) 62413714

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/102853

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 105515842 A	20 April 2016	None	
CN 104216766 A	17 December 2014	WO 2016029738 A1	03 March 2016
CN 101997701 A	30 March 2011	None	
WO 2004068469 A2	12 August 2004	EP 1584036 A2	12 October 2005
		JP 2006516341 A	29 June 2006
		CN 1754155 A	29 March 2006
		AU 2004207357 A1	12 August 2004
		CA 2513503 A1	12 August 2004
		US 2004186861 A1	23 September 2004
		US 2007198685 A1	23 August 2007

国际检索报告

国际申请号 PCT/CN2017/102853

A. 主题的分类

G06F 17/30(2006.01)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,EPODOC,WPI,CNKI:数据，流，查询，查找，存储，缓存，合并，汇总，汇合，stream, flow, data, inquire, ql, storage, buffer, consolidation, union, collateral

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	CN 105515842 A (成都科来软件有限公司) 2016年 4月 20日 (2016-04-20) 说明书第[0008]-[0026]段，附图1-2	1-16
A	CN 104216766 A (华为技术有限公司) 2014年 12月 17日 (2014-12-17) 全文	1-16
A	CN 101997701 A (中兴通讯股份有限公司) 2011年 3月 30日 (2011-03-30) 全文	1-16
A	WO 2004068469 A2 (TACIT NETWORKS, INC. 等) 2004年 8月 12日 (2004-08-12) 全文	1-16

☐ 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
2017年 12月 8日	2017年 12月 27日
ISA/CN的名称和邮寄地址 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 传真号 (86-10)62019451	受权官员 孙娟 电话号码 (86-10)62413714

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2017/102853

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	105515842	A	2016年 4月 20日	无			
CN	104216766	A	2014年 12月 17日	WO	2016029738	A1	2016年 3月 3日
CN	101997701	A	2011年 3月 30日	无			
WO	2004068469	A2	2004年 8月 12日	EP	1584036	A2	2005年 10月 12日
				JP	2006516341	A	2006年 6月 29日
				CN	1754155	A	2006年 3月 29日
				AU	2004207357	A1	2004年 8月 12日
				CA	2513503	A1	2004年 8月 12日
				US	2004186861	A1	2004年 9月 23日
				US	2007198685	A1	2007年 8月 23日