

一种用于联机分析处理的数据处理方法和装置

本申请要求 2018 年 02 月 26 日提交的申请号为 No.201810160009.7 的中国申请的优先权，通过引用将其全部内容并入本文。

技术领域

5 本发明涉及数据处理技术领域，具体涉及一种用于联机分析处理的数据处理方法和装置。

发明背景

当今互联网是海量数据的时代，无线网络的快速升级、终端设备的广泛使用、保险业务的线上开展，都方便了用户进行保险查看、购买、编辑等操作；以及为了
10 了方便对数据、用户、保险进行数据分析和决策，从而将这些数据进行电子化存储，使得保险数据内容、数据量都变得越来越大。

线上保险业务数据主要存储在业务数据库中，为了能够适应创建、修改、编辑、删除等快速操作，线上数据库主要是使用 OLTP 模式，即为了支持事务的快速响应。

15 为了能够使得数据能够最大限度的发挥价值，数据分析人员、业务人员、决策人员会对这些海量的保险数据进行分析，从中发现用户特点、客群特点、保险业务特点，从而为数据运营和决策提供数据依据，这个数据分析模式则是通过 OLAP（联机分析处理系统）的方式来实现数据的快速多维分析查询的。

由于线上事务数据和线下分析数据使用的差异，需要将数据从 OLTP 转移到
20 OLAP 模式，现有的数据处理方式是通过将各个不同类型的数据根据保单号码 ID、用户等信息拼接在一起，然后进行 OLAP 的数据多维分析。

传统的互联网数据拼接方案是按照天、周或者月对数据进行批量处理即可，然后就可以进行数据的多维分析；但是这种方案并不适合于大数据量下的保险数据，因为相对于传统数据，保险金融数据在一定时间范围内会对数据进行修改，
25 而这个数据的修改也需要反映到 OLAP 的数据仓库中，也就是需要通过回溯将数据变动和 OLAP 的数据仓库之前的数据进行拼接、合并处理，当变动的数据量较大的时候，批量处理的方式会消耗较多不必要的计算资源。

发明内容

有鉴于此，本发明提供一种用于联机分析处理的数据处理方法和装置，以在
30 大数据量下的业务数据变动频繁时，能够实现对数据的快速拼接，同时减少不必

要的数据拼接计算量。所述技术方案如下：

第一方面，提供了一种用于联机分析处理的数据处理方法，所述方法包括：

根据选定的分区字段，对联机分析处理系统的增量数据进行分区划分，得到多个增量数据分区；

5 根据所述分区字段，对所述联机分析处理系统的存量数据进行分区划分，得到多个存量数据分区；以及

将具有相同分区标识的所述增量数据分区的数据和所述存量数据分区的数据整合。

结合第一方面，在第一种可能的实现方式中，所述将具有相同分区标识的所述增量数据分区的数据和所述存量数据分区的数据整合包括：

对各个所述增量数据分区中的数据与对应的所述存量数据分区中的数据进行拼接操作和/或合并操作。

结合第一方面的第一种可能的实现方式，在第二种可能的实现方式中，所述方法还包括：

15 构建所述多个存量数据分区的数据索引；基于所述多个存量数据分区的数据索引，对各个所述增量数据分区中的数据与对应的所述存量数据分区中的数据进行拼接操作和/或合并操作。

结合第一方面至第一方面的第二种任意一种可能的实现方式，在第三种可能的实现方式中，所述分区字段包括时间字段、业务字段和自定义字段中的一种或多种组合。

结合第一方面的第二种可能的实现方式，在第四种可能的实现方式中，所述构建所述多个存量数据分区的数据索引包括：

以二叉树或散列方式构建所述数据索引。

结合第一方面的第二种可能的实现方式，在第五种可能的实现方式中，所述基于所述多个存量数据分区的数据索引，对各个所述增量数据分区中的数据与对应的所述存量数据分区中的数据进行拼接操作和/或合并操作包括：

针对各个所述增量数据分区中的数据，按照所述数据索引，在与所述各个增量数据分区对应的所述存量数据分区中进行查询；

若查询成功，则对该增量数据分区中的数据和与其相关联的所述存量数据分区中的数据进行拼接操作。

结合第一方面的第二种或第一方面的第五种可能的实现方式，在第六种可能的实现方式中，所述基于所述多个存量数据分区的数据索引，对各个所述增量数据分区中的数据与对应的所述存量数据分区中的数据进行拼接操作和/或合并操作包括：

针对各个所述增量数据分区中的数据，按照所述数据索引，在与所述各个增量数据分区对应的所述存量数据分区中进行查询；

若查询失败，则对该增量数据分区中的数据和与其相关联的所述存量数据分区中的数据进行合并操作。

- 5 结合第一方面至第一方面的第六种任意一种可能的实现方式，在第七种可能的实现方式中，所述将具有相同分区标识的所述增量数据分区的数据和所述存量数据分区的数据整合步骤之后，所述方法还包括：

通过转换加载方式，将整合后的数据导出至指定位置，以进行所述联机分析处理系统的数据分析；和/或

- 10 输出用于指示数据已更新的提示信息至下游数据处理单元。

第二方面，提供了一种用于联机分析处理的数据处理装置，所述装置包括：

分区划分模块，用于根据选定的分区字段，对联机分析处理系统的增量数据进行分区划分，得到多个增量数据分区；

- 15 所述分区划分模块，还用于根据所述分区字段，对所述联机分析处理系统的存量数据进行分区划分，得到多个存量数据分区；

数据处理模块，用于将具有相同分区标识的所述增量数据分区的数据和所述存量数据分区的数据整合。

- 20 结合第二方面，在第一种可能的实现方式中，所述数据处理模块进一步配置为：对各个所述增量数据分区中的数据与对应的所述存量数据分区中的数据进行拼接操作和/或合并操作。

结合第二方面的第一种可能的实现方式，在第二种可能的实现方式中，所述装置还包括：

- 25 数据索引构建模块，用于构建所述多个存量数据分区的数据索引；所述数据处理模块基于所述多个存量数据分区的数据索引，对各个所述增量数据分区中的数据与对应的所述存量数据分区中的数据进行拼接操作和/或合并操作。

结合第二方面至第二方面的第二种任意一种可能的实现方式，在第三种可能的实现方式中，所述分区字段包括时间字段、业务字段和自定义字段中的一种或多种组合。

- 30 结合第二方面的第二种可能的实现方式，在第四种可能的实现方式中，所述数据索引构建模块具体用于：

以二叉树或散列方式构建所述数据索引。

结合第二方面的第二种可能的实现方式，在第五种可能的实现方式中，所述数据处理模块包括：

- 35 查询子模块，用于针对各个所述增量数据分区中的数据，按照所述数据索引，在与所述各个增量数据分区对应的所述存量数据分区中进行查询；

拼接操作子模块，用于若所述查询子模块查询成功，则对所述增量数据分区中的数据 and 与其相关联的所述存量数据分区中的数据进行拼接操作。

结合第二方面的第五种可能的实现方式，在第六种可能的实现方式中，所述数据处理模块还包括：

- 5 合并操作子模块，用于若所述查询子模块查询失败，则对该增量数据分区中的数据 and 与其相关联的所述存量数据分区中的数据进行合并操作。

结合第二方面至第二方面的第六种任意一种可能的实现方式，在第七种可能的实现方式中，所述装置还包括：

- 10 数据导出模块，用于通过转换加载方式，将整合后的数据导出至指定位置，以进行所述联机分析处理系统的数据分析。

结合第二方面至第二方面的第七种任意一种可能的实现方式，在第八种可能的实现方式中，所述装置还包括：

信息输出模块，用于在整合后输出用于指示数据已更新的提示信息至下游数据处理单元。

- 15 本发明实施例提供了用于联机分析处理的数据处理方法和装置，通过分别对联机分析处理系统的增量数据和存量数据进行分区划分，将具有相同分区标识的所述增量数据分区的数据和所述存量数据分区的数据整合，由此使得数据整合处理过程能够快速地进行，同时也避免了在全量数据上的不必要的操作，极大地简化了数据的计算量；另外，通过实现对数据的快速整合，从而帮助业务人员能够
20 基于完整维度的保险数据，进行数据和业务探索。

附图简要说明

- 25 为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 是本发明实施例提供的一种用于联机分析处理的数据处理方法流程图；

图 2 是本发明实施例提供的一种用于联机分析处理的数据处理方法流程图；

图 3 是本发明实施例提供的分区结果和分区下的数据索引构建结果；

图 4 是本发明实施例提供的一种用于联机分析处理的数据处理装置框架图。

实施本发明的方式

为使本发明的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的

实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

5 本发明实施例提供了一种用于联机分析处理的数据处理方法，该方法能够应用于大数据量下的保险数据变动频繁场景，实现对数据的快速拼接，同时能够减少不必要的数据库拼接计算量；此外，本发明实施例提供的方法还可以应用于大数据量下的其他业务数据，本发明对具体的应用场景不加以限定。

图1是根据本发明实施例示出的一种用于联机分析处理的数据处理方法流程图，参照图1所示，该方法包括如下步骤：

10 101、根据选定的分区字段，对联机分析处理系统的增量数据进行分区划分，得到多个增量数据分区。

102、根据分区字段，对联机分析处理系统的存量数据进行分区划分，得到多个存量数据分区。

15 103、将具有相同分区标识的所述增量数据分区的数据和所述存量数据分区的数据整合。

图2是根据本发明实施例示出的一种用于联机分析处理的数据处理方法流程图，参照图1所示，该方法包括如下步骤：

201、根据选定的分区字段，对联机分析处理系统的增量数据进行分区划分，得到多个增量数据分区。

20 其中，由于增量数据在和存量数据进行拼接的时候，通常是采用一个相同数据类型进行拼接的，比如交易ID、唯一ID等信息，同时存量数据的变动会随着时间的推移，变动会越来越少，因此可以选定在数据创建后不会发生变化的字段作为分区字段对增量数据和存量数据进行分区划分，其中，选定的分区字段可以包括时间字段、业务字段和自定义字段中的一种或多种组合，时间字段可以是交易数据的创建时间，业务字段可以是渠道、地域等字段。

具体的，该过程可以包括：

对新加入到联机分析处理系统中的增量数据进行ETL处理，包括数据清洗、转换，以将增量数据转换成联机分析处理所需的存储方式；

30 选定在数据创建后不会发生变化的字段作为分区字段，对ETL处理后的增量数据按照分区字段进行分区划分，得到多个不同的增量数据分区。

此外，在对增量数据和存量数据进行分区划分时，可以通过对数据分区进行抽象，以便能够支持更多分区方式的选择。比如，在以时间字段作为分区字段时，可以对时间分区进行抽象，根据数据任务的周期处理方式，分别调整为更详细的分区组织，即：如果数据周期处理方式是按年、月、周、天，则对应的数据分区方式也是按年、月、周、日进行分区。

此外，在对增量数据和存量数据进行分区划分时，还可以增加更细粒度的分区方式，根据业务的实际需要合并详细粒度分区字段。比如，如果实际业务中渠道、地域对应的交易数据是不重叠的，那么可以采用时间字段+业务字段[渠道、地域]作为分区字段，分别对增量数据和存量数据进行分区划分。

- 5 202、根据分区字段，对联机分析处理系统的存量数据进行分区划分，得到多个存量数据分区，并构建多个存量数据分区的数据索引。

其中对存量数据的存储可以通过大数据平台来进行存储，以方便进行快速横向扩展，比如可以通过 Hadoop 进行大数据的存储。

- 10 其中，采用与分区划分增量数据相同的分区字段对存量数据进行分区划分，以得到多个存量数据分区；

其中，构建多个存量数据分区的数据索引，该过程可以包括：

采用二叉树进行构建各个存量数据分区的数据索引表，其中，该数据索引表具有二叉树数据结构，数据索引表中包括数据标识与存量数据之间的映射关系。

- 15 二叉树是一种特殊的树，在二叉树中每个节点最多有两个子节点，一般称为左子节点和右子节点，并且二叉树的子树有左右之分，其次序不能任意颠倒。二叉树是递归定义的。

除此之外，还可以采用散列（hash）方式构建多个存量数据分区的数据索引。

本发明实施例对具体的数据索引构建过程不加以限定。

- 20 本发明实施例中，可以是使用 Hbase 的索引方式进行数据定位和更新操作，也可通过 Mysql 数据库进行索引的建立和更新，也可以通过小的本地文件数据库进行索引建立和数据更新，比如 levelDB。

203、依据是否具有相同分区标识，建立多个增量数据分区和多个存量数据分区之间的对应关系。

其中，多个增量数据分区和多个存量数据分区分别具有各自的分区标识；

- 25 具体的，本发明实施例对具体的建立过程不加以限定。

- 图 3 示出了本发明实施例的数据分区结果和分区下的数据索引构建结果。在图 3 中，以创建时间作为分区字段，分别对联机分析处理系统的增量数据和存量数据进行分区划分，得到多个不同的增量数据分区和多个不同的存量数据分区，其中，多个增量数据分区和多个存量数据分区均包括{day}、{day-1}、...{day-n}、
30 {day-n-1}...，对相同分区标识的增量数据分区和存量数据分区可以建立对应关系；另外，在图 3 中可以看出，在存量数据的各个分区下的数据索引与各个分区一一对应。

- 本发明实施例中，通过依据是否具有相同分区标识，建立多个增量数据分区和多个存量数据分区之间的对应关系，以便后续在相同分区上对数据进行拼接、
35 合并处理，进而能够减少不必要的拼接尝试，避免不必要的拼接计算量。

204、基于多个存量数据分区的数据索引，对各个增量数据分区中的数据与对应的存量数据分区中的数据进行拼接操作和/或合并操作。

具体的，该过程可以包括：

a、针对各个增量数据分区中的数据，按照数据索引，在与各个增量数据分区对应的存量数据分区中进行查询。

具体的，该过程可以包括：

确定各个增量数据分区中的数据的数据标识，将该数据标识在数据索引表中进行查询，若该数据标识存在于数据索引表中，则确定数据索引表中该数据标识对应的存量数据。

b、若查询成功，则对增量数据和与其相关联的存量数据进行拼接操作，若查询失败，则对增量数据进行合并操作。

示例性的，用户 user1 在 2017-10-10 这一天产生了交易数据，且对应的交易数据 ID 为 ID_user1，创建时间为 2017-10-10，该交易数据作为存量数据存储到联机分析处理系统的数据仓库中，通过时间字段进行对该存量数据进行分区，那么该存量数据被划分到 2017-10-10 分区；若用户 user1 由于业务原因在 2017-12-15 对交易数据 ID_user1 进行修改，修改的内容是交易数据 ID 以外其他任何数据，比如修改了联系方式、增加的身份信息等，此时的交易数据作为增量数据传输到联机分析处理系统中，通过时间字段进行对该增量数据进行分区，那么该增量数据被划分到 2017-10-10 分区。显然，对于交易数据 ID_user1 是可以在 2017-10-10 分区中查询到的，那么进行数据拼接处理，即将修改数据的 2017-10-10 分区和历史存量数据的 2017-10-10 分区各自的数据拼接操作，具体的拼接是按照交易数据 ID 进行处理，这里的交易数据 ID 是按照 ID_user1 的 ID 值，通过数据拼接处理将修改的联系方式、增加的身份信息存储到历史数据中。

可选的，在步骤 204 之后，本发明实施例提供的方法还包括：

通过转换加载方式，将拼接操作和/或合并操作后的数据导出至指定位置，以进行所述联机分析处理系统的数据分析；和/或

输出用于指示数据已更新的提示信息至下游数据处理单元。

本发明实施例提供的用于联机分析处理的数据处理方法，在实际应用中，在进行针对 OLAP 进行数据处理时，对数据进行分区组织，可以按照创建时间对数据进行分区组织这种常见的分区组织方式；在对存量数据使用 HBase 作为存储方案，可以通过 rowkey 的方式能够对在索引上进行排序查找，从而进行数据的快速定位；另外由于 Hive 具有 SQL 友好、支持多种查询引擎的特定，可以使用 Hive 作为全量历史数据的基础存储，从而基于 Hive 构建能够更便捷地进行多维 OLAP 分析；另外，对于每次批量数据的更新，进行逐个分区的处理，对每个分区处理的时候，通过 rowkey 定位和读取需要更新的数据，并进行数据的更新，同时将数

据写回到 HBase 中；最后通过对变动的数据直接导出分区数据到 Hive 中即可，由于导出不涉及到数据的更新合并操作，所以可以比较快速的进行。

本发明实施例提供一种用于联机分析处理的数据处理方法，通过分别对联机分析处理系统的增量数据和存量数据进行分区划分，并在分区下构建数据索引，
5 以及在相同分区上进行数据拼接、合并处理，由于通过索引的方式对数据进行查找和相应数据的更新，只查找和更新需要更新的数据，由此使得数据拼接、合并处理过程能够快速地进行，同时也避免了在全量数据上的不必要的操作，极大地简化了数据的计算量；另外，通过实现对数据的快速拼接，从而帮助业务人员能够基于完整维度的保险数据，进行数据和业务探索。本发明实施例提供的方法在
10 保险金融数据场景中具有很大的优势，尤其是针对保险场景创建数据较多、变动频繁的时候，数据的快速拼接、合并操作优势更加明显。

图 4 是根据本发明实施例示出的一种用于联机分析处理的数据处理装置框架图，参照图 4 所示，该装置 3 包括：

分区划分模块 31，用于根据选定的分区字段，对联机分析处理系统的增量数
15 据进行分区划分，得到多个增量数据分区；

分区划分模块 31，还用于根据分区字段，对联机分析处理系统的存量数据进行分区划分，得到多个存量数据分区；

数据索引构建模块 32，用于构建多个存量数据分区的数据索引；

对应关系建立模块 33，用于依据是否具有相同分区标识，建立多个增量数据
20 分区和多个存量数据分区之间的对应关系；

数据处理模块 34，用于基于多个存量数据分区的数据索引，对各个增量数据分区中的数据与对应的存量数据分区中的数据进行拼接操作和/或合并操作。

可选的，分区字段包括时间字段、业务字段和自定义字段中的一种或多种组合。

25 可选的，数据索引构建模块 32 具体用于：

以二叉树或散列方式构建数据索引。

可选的，数据处理模块 34 包括：

查询子模块 341，用于针对各个增量数据分区中的数据，按照数据索引，在
与各个增量数据分区对应的存量数据分区中进行查询；

30 拼接操作子模块 342，用于若查询子模块 341 查询成功，则对增量数据和与其相关联的存量数据进行拼接操作；

合并操作子模块 343，用于若查询子模块 342 查询失败，则对增量数据进行合并操作。

可选的，装置 3 还包括：

35 数据导出模块 35，用于通过转换加载方式，将拼接操作和/或合并操作后的数

据导出至指定位置，以进行联机分析处理系统的数据分析；

信息输出模块 36，用于输出用于指示数据已更新的提示信息至下游数据处理单元。

5 本发明实施例提供了一种用于联机分析处理的数据处理装置，该装置通过分别对联机分析处理系统的增量数据和存量数据进行分区划分，并在分区下构建数据索引，以及在相同分区上进行数据拼接、合并处理，由于通过索引的方式对数据进行查找和相应数据的更新，只查找和更新需要更新的数据，由此使得数据拼接、合并处理过程能够快速地进行，同时也避免了在全量数据上的不必要的操作，极大地简化了数据的计算量；另外，通过实现对数据的快速拼接，从而帮助业务
10 人员能够基于完整维度的保险数据，进行数据和业务探索。本发明实施例提供的装置在保险金融数据场景中具有很大的优势，尤其是针对保险场景创建数据较多、变动频繁的时候，数据的快速拼接、合并操作优势更加明显。

上述所有可选技术方案，可以采用任意结合形成本发明的可选实施例，在此不再一一赘述。

15 需要说明的是：上述实施例提供的用于联机分析处理的数据处理装置在执行用于联机分析处理的数据处理方法时，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能模块完成，即将设备的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。另外，上述实施例提供的用于联机分析处理的数据处理装置与用于联机分析处理的数据
20 处理方法实施例属于同一构思，其具体实现过程详见方法实施例，这里不再赘述。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

25 以上所述仅为本发明的较佳实施例，并不用以限制本发明，凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

以上仅为本发明的较佳实施例而已，并非用于限定本发明的保护范围。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权利要求书

1、一种用于联机分析处理的数据处理方法，其特征在于，所述方法包括：

根据选定的分区字段，对联机分析处理系统的增量数据进行分区划分，得到多个增量数据分区；

5 根据所述分区字段，对所述联机分析处理系统的存量数据进行分区划分，得到多个存量数据分区；以及

将具有相同分区标识的所述增量数据分区的数据和所述存量数据分区的数据整合。

2、根据权利要求1所述的方法，其特征在于，所述将具有相同分区标识的所述增量数据分区的数据和所述存量数据分区的数据整合包括：

对各个所述增量数据分区中的数据与对应的所述存量数据分区中的数据进行拼接操作和/或合并操作。

3、根据权利要求2所述的方法，其特征在于，所述方法还包括：

15 构建所述多个存量数据分区的数据索引；基于所述多个存量数据分区的数据索引，对各个所述增量数据分区中的数据与对应的所述存量数据分区中的数据进行拼接操作和/或合并操作。

4、根据权利要求1至3任意一项所述的方法，其特征在于，所述分区字段包括时间字段、业务字段和自定义字段中的一种或多种组合。

20 5、根据权利要求3所述的方法，其特征在于，所述构建所述多个存量数据分区的数据索引包括：

以二叉树或散列方式构建所述数据索引。

6、根据权利要求3所述的方法，其特征在于，所述基于所述多个存量数据分区的数据索引，对各个所述增量数据分区中的数据与对应的所述存量数据分区中的数据进行拼接操作和/或合并操作包括：

25 针对各个所述增量数据分区中的数据，按照所述数据索引，在与所述各个增量数据分区对应的所述存量数据分区中进行查询；

若查询成功，则对该增量数据分区中的数据和与其相关联的所述存量数据分区中的数据进行拼接操作。

- 7、根据权利要求3或6所述的方法，其特征在于，所述基于所述多个存量数据分区的数据索引，对各个所述增量数据分区中的数据与对应的所述存量数据分区中的数据进行拼接操作和/或合并操作包括：

针对各个所述增量数据分区中的数据，按照所述数据索引，在与所述各个增量数据分区对应的所述存量数据分区中进行查询；

若查询失败，则对该增量数据分区中的数据和与其相关联的所述存量数据分区中的数据进行合并操作。

- 8、根据权利要求1至7任意一项所述的方法，其特征在于，所述将具有相同分区标识的所述增量数据分区的数据和所述存量数据分区的数据整合步骤之后，所述方法还包括：

通过转换加载方式，将整合后的数据导出至指定位置，以进行所述联机分析处理系统的数据分析；和/或

- 输出用于指示数据已更新的提示信息至下游数据处理单元。

9、一种用于联机分析处理的数据处理装置，其特征在于，所述装置包括：

分区划分模块，用于根据选定的分区字段，对联机分析处理系统的增量数据进行分区划分，得到多个增量数据分区；

- 所述分区划分模块，还用于根据所述分区字段，对所述联机分析处理系统的存量数据进行分区划分，得到多个存量数据分区；以及

数据处理模块，用于将具有相同分区标识的所述增量数据分区的数据和所述存量数据分区的数据整合。

10、根据权利要求9所述的装置，其特征在于，所述数据处理模块进一步配置为：

- 对各个所述增量数据分区中的数据与对应的所述存量数据分区中的数据进行拼接操作和/或合并操作。

11、根据权利要求10所述的装置，其特征在于，所述装置还包括：

数据索引构建模块，用于构建所述多个存量数据分区的数据索引；所述数据处理模块基于所述多个存量数据分区的数据索引，对各个所述增量数据分区中的数据与对应的所述存量数据分区中的数据进行拼接操作和/或合并操作。

12、根据权利要求 9 至 11 任意一项所述的装置，其特征在于，所述分区字段
5 包括时间字段、业务字段和自定义字段中的一种或多种组合。

13、根据权利要求 11 所述的装置，其特征在于，所述数据索引构建模块具体用于：

以二叉树或散列方式构建所述数据索引。

14、根据权利要求 11 所述的装置，其特征在于，所述数据处理模块包括：

10 查询子模块，用于针对各个所述增量数据分区中的数据，按照所述数据索引，在与所述各个增量数据分区对应的所述存量数据分区中进行查询；

拼接操作子模块，用于若所述查询子模块查询成功，则对该增量数据分区中的数据和与其相关联的所述存量数据分区中的数据进行拼接操作。

15、根据权利要求 14 所述的装置，其特征在于，所述数据处理模块还包括：

15 合并操作子模块，用于若所述查询子模块查询失败，则对该增量数据分区中的数据 and 与其相关联的所述存量数据分区中的数据进行合并操作。

16、根据权利要求 9 至 15 任意一项所述的装置，其特征在于，所述装置还包括：

20 数据导出模块，用于通过转换加载方式，将整合后的数据导出至指定位置，以进行所述联机分析处理系统的数据分析。

17、根据权利要求 9 至 16 任意一项所述的装置，其特征在于，所述装置还包括：

信息输出模块，用于在整合后输出用于指示数据已更新的提示信息至下游数据处理单元。

25