



(12)发明专利

(10)授权公告号 CN 106649552 B

(45)授权公告日 2020.05.12

(21)申请号 201610975235.1

(51)Int.Cl.

(22)申请日 2016.11.07

G06F 16/215(2019.01)

(65)同一申请的已公布的文献号

审查员 汤明达

申请公布号 CN 106649552 A

(43)申请公布日 2017.05.10

(73)专利权人 湖北省农村信用社联合社网络信息中心

地址 430071 湖北省武汉市武昌区水果湖街小洪山中区70号

(72)发明人 高天翔 杨志云 付毅 汪源
陈广涛 杨欣 周新 刘麟
杨建设 吴宇光 陈胜

(74)专利代理机构 北京清亦华知识产权代理事务所(普通合伙) 11201

代理人 张大威

权利要求书1页 说明书4页 附图2页

(54)发明名称

常态化数据清理方法

(57)摘要

本发明公开了一种常态化数据清理方法,包括以下步骤:S1、构造数据容器转换系统,此系统用于实现以日为单位的数据转换加工、查询、删除和调整操作;S2、数据分离,将数据容器转换系统中的数据进行数据分离,得到分离后的数据;S3、数据加工,将分离后的数据依据业务逻辑进行数据加工,得到加工后的数据;S4、数据整合,将加工完毕的数据进行整合,以保证数据的完整性;S5、实现自动维护系统。该方法减少了人员参与作业的人力成本,且降低了维护过程中实施人员的误操作风险。



1. 一种常态化数据清理方法,其特征在于,包括以下步骤;

S1、构造数据容器转换系统,此系统用于实现以日为单位的数据转换加工、查询、删除和调整操作;

S2、数据分离,将数据容器转换系统中的数据进行数据分离,得到分离后的数据;

S3、数据加工,将分离后的数据依据业务逻辑进行数据加工,得到加工后的数据;

S4、数据整合,将加工完毕的数据进行整合,以保证数据的完整性;

S5、实现自动维护系统;

所述步骤S1包括:

S101、创建一个分区表,其表定义与现有的物理表相同;

S102、从现有的物理表中导出数据;

S103、将导出的数据加载到创建的分区表中;

S104、对分区表建立索引和约束;

S105、删除物理表;

S106、重命名分区表。

2. 根据权利要求1所述的常态化数据清理方法,其特征在于,所述步骤S2包括:

将待加工清理区间内的数据从分区表中进行分离,分离出小表。

3. 根据权利要求2所述的常态化数据清理方法,其特征在于,所述步骤S3包括:

对分离出的小表依据业务逻辑进行数据加工,得到加工完毕的数据,加工过程中加入各类数据有效性检查步骤,加工完成后提取有效数据,清理无效数据;并使用自动的维护脚本检查,在数据提取后自动核对数据完整性,避免数据混乱。

4. 根据权利要求3所述的常态化数据清理方法,其特征在于,所述步骤S4包括:

将加工完毕的数据加入至分区表中,保证数据的完整性。

5. 根据权利要求3所述的常态化数据清理方法,其特征在于,所述步骤S5包括:

在TWS作业流中加入自动维护脚本;

设定报警机制,具体为:

检查脚本是否执行成功,如果执行失败则短信通知值班人员;

检查分离小表是否成功,如果执行失败则短信通知值班人员;

检查数据导入是否成功,如果执行失败则短信通知值班人员。

常态化数据清理方法

技术领域

[0001] 本发明属于数据库清理技术领域,具体涉及一种数据库的常态化数据清理的方法。

背景技术

[0002] 数据库清理即对数据库的数据进行清除,传统方法数据库清理一般通过人工清理的方式,将历史数据从现有数据库剥离。传统人工清理方式主要有DELETE和Truncate两种。

[0003] 其中DELETE方式是编程中实现删除功能的命令,用于创建一个删除查询,可从列在FROM子句之中的一个或多个表中删除记录。这种方式执行时会记录数据库日志,在误操作时允许进行数据回滚(部分数据库产品支持闪回)及数据安全提升。但是它的不足在于该方法由两部分进行,1、查询筛选所需目标;2、记录日志后再删除所选目标。在海量数据表中,如果查询目标效率低下,将直接影响删除执行时长,尤其在数据量超过一定量级后,删除执行效率极低,那么表对象中会执行加锁动作,这样就会导致生产业务系统因不能支持长时间锁表而造成业务中断。

[0004] Truncate方式是使ddl操作立即生效,原数据不放到rollback segment(回滚段,内存中的回收站)中,不能回滚;语句缺省情况下将空间释放到minextents个extent(自动进行升高水位动作,释放表空间容量);truncate会将高水位线复位(回到最开始)。这种方式只清除表中的所有数据,不删除表的结构,所以执行速度快,占用系统开销小。但是它的不足在于无法进行数据筛选,而且整表数据清除,不记日志,不能回滚,不允许误操作,安全性低。虽然Truncate通常与select insert into(查询插入,将需要的数据加工后放入新表)一起使用。把所选数据转储后进行原表Truncate可以达到快速清理数据的目的。但是他仍然需要依赖select的效率。

发明内容

[0005] 本发明旨在解决现有技术中存在的技术问题之一,为此,本发明的一个目的在于提供了一种常态化的数据清理方法,可以有效解决查询时间长,操作无法回滚,安全性低,且全部依赖人工操作等问题。

[0006] 需要说明的是,本发明是基于发明人的下列发现而完成的:

[0007] 根据本发明的一个方面,本发明提供了一种常态化数据清理方法,包括以下步骤;

[0008] S1、构造数据容器转换系统,此系统用于实现以日为单位的数据转换加工、查询、删除和调整操作;

[0009] S2、数据分离,将数据容器转换系统中的数据进行数据分离,得到分离后的数据;

[0010] S3、数据加工,将分离后的数据依据业务逻辑进行数据加工,得到加工后的数据;

[0011] S4、数据整合,将加工完毕的数据进行整合,以保证数据的完整性;

[0012] S5、实现自动维护系统。

[0013] 另外,根据本发明上述实施例一种常态化数据清理方法,还可以具有如下附加的

技术特征：

[0014] 根据本发明的实施例,所述步骤S1包括:

[0015] S101、创建一个分区表,其表定义与现有的物理表相同;

[0016] S102、从现有的物理表中导出数据;

[0017] S103、将导出的数据加载到创建的分区表中;

[0018] S104、对分区表建立索引和约束;

[0019] S105、删除物理表;

[0020] S106、重命名分区表。

[0021] 根据本发明的实施例,所述步骤S2包括:

[0022] 将待加工清理区间内的数据从分区表中进行分离,分离出小表。

[0023] 根据本发明的实施例,所述步骤S3包括:

[0024] 对分离出的小表依据业务逻辑进行数据加工,得到加工完毕的数据,加工过程中加入各类数据有效性检查步骤,加工完成后提取有效数据,清理无效数据;并使用自动的维护脚本检查,在数据提取后自动核对数据完整性,避免数据混乱。

[0025] 根据本发明的实施例,所述步骤S4包括:

[0026] 将加工完毕的数据加入至分区表中,保证数据的完整性。

[0027] 根据本发明的实施例,所述步骤S5包括:

[0028] 在TWS作业流中加入自动维护脚本;

[0029] 设定报警机制,具体为:

[0030] 检查脚本是否执行成功,如果执行失败则短信通知值班人员。

[0031] 检查分离小表是否成功,如果执行失败则短信通知值班人员。

[0032] 检查数据导入是否成功,如果执行失败则短信通知值班人员。

[0033] 将S1-S4步骤流程作为自动维护脚本加入TWS自动作业流程中,并设定报警机制,从而实现自动维护系统。使用自动维护系统后,维护人员无需值守作业。减少了人员参与作业的人力成本,且降低了维护过程中实施人员的误操作风险。

[0034] 实施本发明提供的常态化数据清理方法,具有以下有益效果:

[0035] 1、减少了人员参与作业的人力成本。

[0036] 2、降低了维护过程中实施人员的误操作风险。

[0037] 3、降低了服务器资源开销。

[0038] 4、提高了自动维护作业的安全性。

[0039] 5、提高了数据清理速度。

附图说明

[0040] 图1为步骤S1构造数据容器转换系统的流程示意图;

[0041] 图2为步骤S2数据分离流程示意图;

[0042] 图3为步骤S3数据加工流程示意图;

[0043] 图4为步骤S4数据整合流程示意图。

具体实施方式

[0044] 下面详细描述本发明的实施例。下面通过参考附图描述的实施例是示例性的，仅用于解释本发明，而不能理解为对本发明的限制。

[0045] 如图1所示，常态化数据清理方法包含如下步骤：

[0046] S1、构造数据容器转换系统；

[0047] S2、进行数据分离；

[0048] S3、进行数据加工；

[0049] S4、完成数据整合；

[0050] S5、实现自动维护系统。

[0051] 在本发明所述的常态化数据清理方法中，所述步骤S1包括：

[0052] S101、建一个新的分区表，其表定义与现有的物理表相同；

[0053] S102、从当前的物理表中导出数据；

[0054] S103、将数据加载到新创建的分区表中；

[0055] S104、对分区表建立必要的索引和约束；

[0056] S105、删除物理表；

[0057] S106、重命名新的分区表。

[0058] 此步骤实现了基于分区表的分区分离功能，将时间业务数据以日为单位进行逻辑分区，并将数据装入可自由分离的分区表中。通过此方法可构造出一个将装载数据的物理表转换为分区表的系统，此系统可实现以日为单位的数据转换加工、查询、删除、调整等操作。

[0059] 在本发明所述的常态化数据清理方法中，所述步骤S2包括：

[0060] 将待加工清理区间内的数据(以日为单位)从分区表(大表)中进行分离。分离后小表数据量小便于加工清理，且加工清理数据时，不影响分区表(大表)继续对业务系统提供服务。

[0061] 在本发明所述的常态化数据清理方法中，所述步骤S3包括：

[0062] 对分离出的小表依据业务逻辑进行数据加工(过程中加入各类数据有效性检查步骤)，提取有效数据，清理无效数据。并使用自动的维护脚本检查，在数据提取后自动核对数据完整性，避免数据混乱。

[0063] 在本发明所述的常态化数据清理方法中，所述步骤S4包括：

[0064] 将加工完毕的数据加入至分区表(大表)中，保证数据的完整性。

[0065] 在本发明所述的常态化数据清理方法中，所述步骤S5包括：

[0066] 1、在TWS作业流中加入自动维护脚本；

[0067] 2、设定报警机制，检查脚本是否执行成功，如果执行失败则短信通知值班人员。

[0068] 检查分离小表是否成功，如果执行失败则短信通知值班人员。

[0069] 检查数据导入是否成功，如果执行失败则短信通知值班人员。

[0070] 在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不一定指的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任何

的一个或多个实施例或示例中以合适的方式结合。

[0071] 尽管已经示出和描述了本发明的实施例,本领域的普通技术人员可以理解:在不脱离本发明的原理和宗旨的情况下可以对这些实施例进行多种变化、修改、替换和变型,本发明的范围由权利要求及其等同物限定。

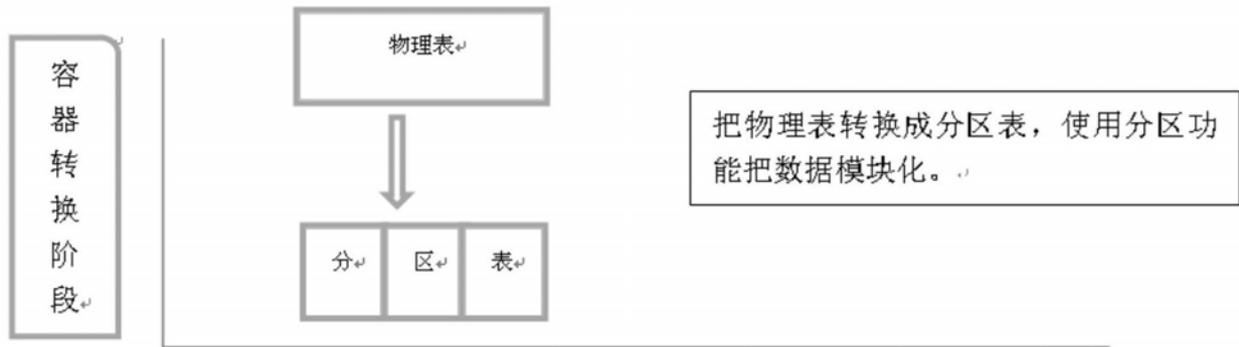


图1



图2

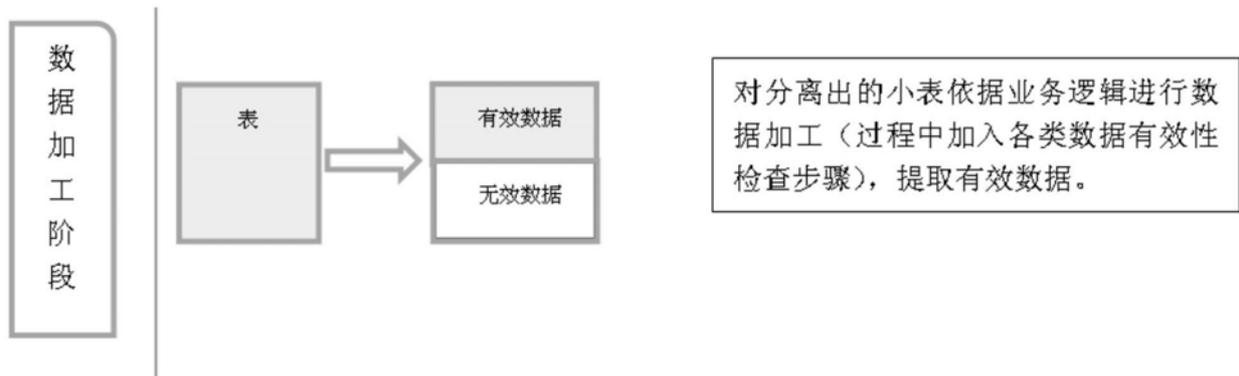


图3

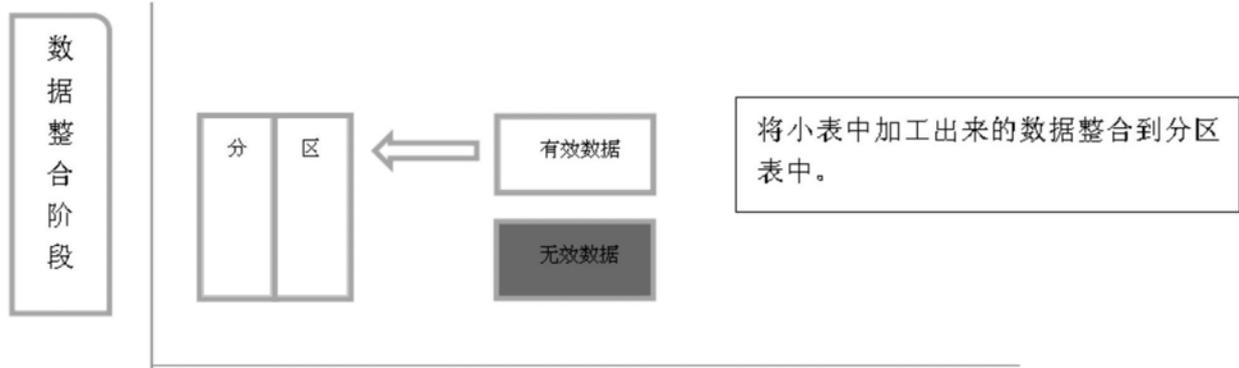


图4