



(12) 发明专利

(10) 授权公告号 CN 102646136 B

(45) 授权公告日 2014. 04. 02

(21) 申请号 201210099036. 0

(22) 申请日 2012. 04. 06

(73) 专利权人 北京西塔网络科技股份有限公司  
地址 100080 北京市海淀区丹棱街 16 号海  
兴大厦 B 座 F102

(72) 发明人 祝守宇

(74) 专利代理机构 北京超凡志成知识产权代理  
事务所(普通合伙) 11371  
代理人 吴开磊

(51) Int. Cl.  
G06F 17/30(2006. 01)

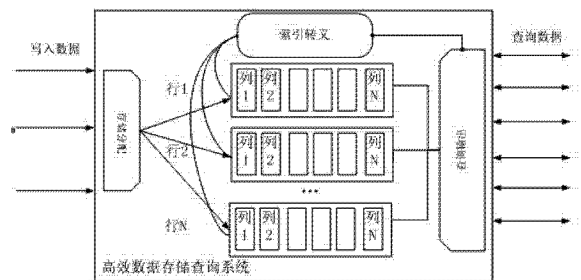
(56) 对比文件  
CN 101311930 A, 2008. 11. 26, 全文.  
CN 101968806 A, 2011. 02. 09, 全文.  
US 5471478 A, 1995. 11. 28, 全文.  
杨栋. Facebook 数据仓库揭秘 :RCFile  
高效存储结构. 《http://www. csdn. net/  
article/2011-04-29/296900》. 2011, 全文.

审查员 王平

权利要求书1页 说明书2页 附图1页

(54) 发明名称  
一种高效数据存储查询的方法及系统

(57) 摘要  
本发明公开一种高效数据存储查询方法, 实现移动互联网用户数据记录高速写入、大容量存储及快速提取, 并依据上述方法发明一种高效数据存储查询系统。



1. 一种高效数据存储查询的实现方法,其特征在于将移动互联网用户数据记录的用户关键信息快速存储,所述用户关键信息包括用户手机号码、客户端 IP 地址,存储时将每个用户的多条记录存放在一起,根据客户端 IP 地址将数据分为块和行,客户端 IP 地址前 24 位相同的数据都写入到相同的行块中,客户端 IP 地址后 8 位分行,分行的原则是保证一个用户的数据都保存一个行块中,然后在行块中划分列,每一列就是一个数据库表空间的一个字段,将手机号码、客户端 IP 地址转义成数字或数字组合作为索引,实现存储数据的快速查询定位,所有数据支持一次写入多次读取。

## 一种高效数据存储查询的方法及系统

### 技术领域

[0001] 本发明涉及一种实现移动互联网用户数据业务记录存储的方法及系统,具体涉及技术领域的是索引转义、行列二维存储及用户数据业务记录存储所涉及的系统和方法。

### 背景技术

[0002] 随着移动互联网和智能终端的快速普及,使用移动互联网数据业务的用户越来越多,移动互联网检测需要保存用户上网记录越来越多。传统关系型数据库成本高,写入速度慢,查询时间长,难以满足日益庞大的数据存储。

### 发明内容

[0003] 本发明的主要目的是提供一种高效数据存储查询的方法,实现移动互联网用户数据记录高速写入、大容量存储及快速提取。

[0004] 本发明的主要目的是提供一种高效数据存储查询的系统,实现移动互联网用户数据记录高速写入、大容量存储及快速提取。

[0005] 为了实现上述目的,本发明公开一种高效数据存储查询方法,该方法将移动互联网用户数据记录的用户手机号码、Client IP 等用户关键信息快速存储,存储时采用行列二维存储方式将每个用户的多条记录存放在一起,支持将手机号码、Client IP 等信息自动转义成数字或数字组合索引,实现存储数据的快速查询定位,数据只有一次写入过程,可多次读取。

[0006] 为了实现上述目的,本发明公开一种高效数据存储查询系统,该系统将移动互联网用户数据记录的用户手机号码、Client IP 等用户关键信息快速存储,存储时采用行列二维存储方式将每个用户的多条记录存放在一起,支持将手机号码、Client IP 自动转义成数字或数字组合索引,实现存储数据的快速查询定位,数据只有一次写入过程,可多次读取。

[0007] 本发明实施例的高效数据存储查询系统能够在保证数据存储质量的同时,降低数据存储成本,提高数据存储和查询效率。支持每秒 5 万条 500 字节长度的记录插入,30 亿条记录查询时间小于 10 秒。能有效解决当前海量数据存储成本高、容量小、存取速率低问题。

[0008] 附图说明:

[0009] 此处所说明的附图用来提供对本发明的进一步理解,构成本申请的一部分,并不构成对本发明的限定。在附图中:

[0010] 图 1 高效数据存储查询系统实现原理图

[0011] 具体实施方式:

[0012] 为使本发明的目的、技术方案和优点更加清楚,下面结合附图对本发明的具体实施例进行详细说明。在此,本发明的示意性实施例及其说明用于解释本发明,但并不作为对本发明的限定。

[0013] 高效数据存储查询系统按照行列二维存储方式保存数据。系统根据 Client IP 将数据分为块和行,Client IP 前 24 位相同的数据都写入到相同的行块中,Client IP 后 8 位

分行,分行的原则是保证一个用户的数据都保存一个行块中。然后在行块中划分列,每一列就是一个数据库表空间的一个字段,由于一个用户存在多条数据记录,很多信息是相同的,包括手机号码、Client IP、URL 等,所以按列存储可以对数据做压缩存储,可以大大减小硬盘 I/O 及存储容量。

[0014] 高效数据存储查询系统由图一所示,本系统输入数据来自移动互联网用户使用数据业务记录,系统输入信息包含用户号码、Client IP、URL 等数据信息。

[0015] 系统接收到写入数据,按照上述所讲的行列二维存储统一分配存储空间,保证同一用户的信息存放在相邻位置,同时将用户号码、Client IP 等信息转义成索引存储在系统中,便于以后的数据精确查询。

[0016] 当用户查询记录时,系统都交予查询输出功能模块处理。该模块根据用户提交的查询条件到索引表中查询,目前支持的索引包括用户号码、Client IP、用户位置等,系统在设计时充分考虑过现场实际应用中的各种可能的索引条件,因此能满足现场的大部分应用。本系统的索引都是数字或数字组合方式,可根据索引快速定位到对应的信息记录中并取出。

[0017] 本系统适用于海量移动互联网的用户数据存取,可以很好的解决传统关系型数据库成本高,写入速度慢,查询时间长问题。

[0018] 应当理解的是,上述针对具体实施例的描述较为详细,并不能因此而认为是对本发明专利保护范围的限制,本发明的专利保护范围应以所附。

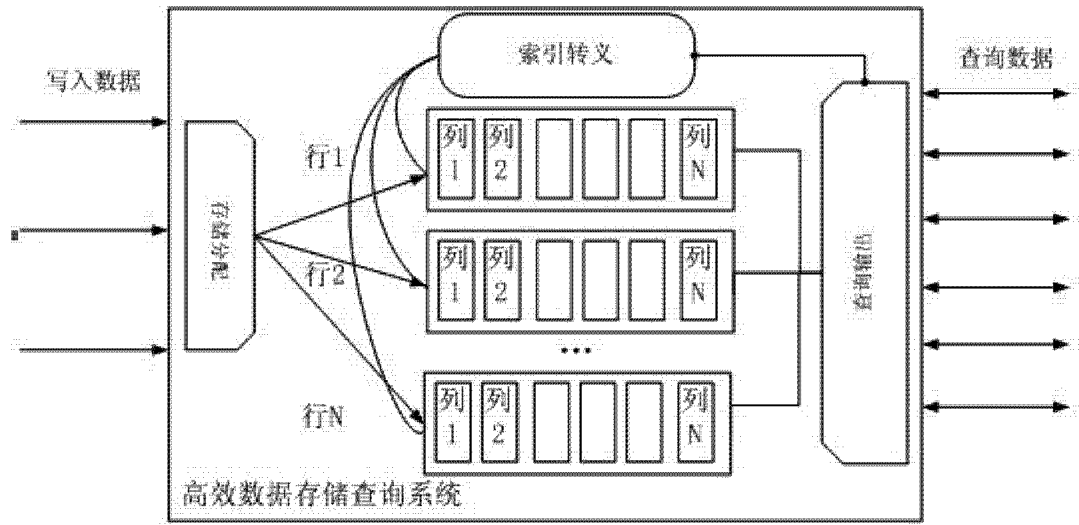


图 1