(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 106295403 A
(43)申请公布日 2017.01.04

(21)申请号 201610885912.0

(22)申请日 2016.10.11

(71)申请人 北京集奥聚合科技有限公司
地址 100085 北京市海淀区上地东路1号院5号楼9层901

(72)发明人 何良均 张翼 温宗臣 冯森林 李冰 张书凡 范卫卫 赵志华

(74)专利代理机构 北京和信华成知识产权代理事务所(普通合伙) 11390
代理人 胡剑辉

(51)Int.Cl.
G06F 21/62(2013.01)
G06F 21/60(2013.01)
H04L 9/06(2006.01)

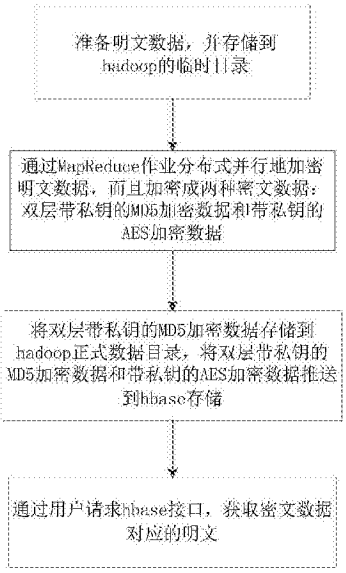
权利要求书2页 说明书5页 附图1页

(54)发明名称

一种基于hbase的数据安全处理方法及系统

(57)摘要

本发明公开一种基于hbase的数据安全处理方法，其既保证了存储到hadoop上的数据是加密的数据，也保证了用户需要明文时可以通过hbase接口调用获取。该方法包括：(1)准备明文数据，并存储到hadoop的临时目录；(2)通过MapReduce作业分布式并行地加密明文数据，而且加密成两种密文数据：双层带私钥的MD5加密数据和带私钥的AES加密数据；(3)将双层带私钥的MD5加密数据存储到hadoop正式数据目录，将双层带私钥的MD5加密数据和带私钥的AES加密数据推送到hbase存储；(4)通过用户请求hbase接口，获取密文数据对应的明文。还有基于hbase的数据安全处理系统。



CN 106295403 A

1.一种基于hbase的数据安全处理方法，其特征在于：该方法包括以下步骤：

(1)准备明文数据，并存储到hadoop的临时目录；

(2)通过MapReduce作业分布式并行地加密明文数据，而且加密成两种密文数据：双层带私钥的MD5加密数据和带私钥的AES加密数据；

(3)将双层带私钥的MD5加密数据存储到hadoop正式数据目录，将双层带私钥的MD5加密数据和带私钥的AES加密数据推送到hbase存储；

(4)通过用户请求hbase接口，获取密文数据对应的明文。

2.根据权利要求1所述的基于hbase的数据安全处理方法，其特征在于：所述步骤(2)中，双层带私钥的MD5加密数据的私钥设置成既包含数字、大小写字母，还包含了特殊可见字符的64位字符。

3.根据权利要求2所述的基于hbase的数据安全处理方法，其特征在于：所述步骤(2)中，带私钥的AES加密数据的私钥设置成包含数字和大小写字母的64位字符。

4.根据权利要求3所述的基于hbase的数据安全处理方法，其特征在于：所述步骤(3)中，存储到hbase里面，以MD5加密数据为rowkey，AES加密数据作为value存储；将MD5加密数据和AES加密数据做一个(key,value)映射，以方便后期数据解密。

5.根据权利要求4所述的基于hbase的数据安全处理方法，其特征在于：所述步骤(4)中，解密的时候hbase接口执行：

(a)接受用户解密请求，获取用户传递过来的数据密文；

(b)将数据密文作为rowkey，查询hbase中该rowkey对应的value，这里的value就是AES加密的数据；

(c)拿到AES加密的数据，并根据私钥对密文数据进行AES解密，解密算法为：

私钥.数据密文AES.decrypt，这里的私钥64位包含数字和大小写字母的字符；

(d)将AES解密的数据返回给用户。

6.一种基于hbase的数据安全处理系统，其特征在于：该系统包括：

明文数据准备模块，其配置来准备明文数据，并存储到hadoop的临时目录；

数据加密模块，其配置来通过MapReduce作业分布式并行地加密明文数据，而且加密成两种密文数据：双层带私钥的MD5加密数据和带私钥的AES加密数据；

数据存储模块，其配置来将双层带私钥的MD5加密数据存储到hadoop正式数据目录，将双层带私钥的MD5加密数据和带私钥的AES加密数据推送到hbase存储；

数据解密模块，其配置来通过用户请求hbase接口，获取密文数据对应的明文。

7.根据权利要求6所述的基于hbase的数据安全处理系统，其特征在于：所述数据加密模块中，双层带私钥的MD5加密数据的私钥设置成既包含数字、大小写字母，还包含了特殊可见字符的64位字符。

8.根据权利要求7所述的基于hbase的数据安全处理系统，其特征在于：所述数据加密模块中，带私钥的AES加密数据的私钥设置成包含数字和大小写字母的64位字符。

9.根据权利要求8所述的基于hbase的数据安全处理系统，其特征在于：所述数据存储模块中，存储到hbase里面，以MD5加密数据为rowkey，AES加密数据作为value存储；将MD5加密数据和AES加密数据做一个(key,value)映射，以方便后期数据解密。

10.根据权利要求9所述的基于hbase的数据安全处理系统，其特征在于：所述数据解密

模块中，解密的时候hbase接口执行：

(a)接受用户解密请求，获取用户传递过来的数据密文；

(b)将数据密文作为rowkey，查询hbase中该rowkey对应的value，这里的value就是AES加密的数据；

(c)拿到AES加密的数据，并根据私钥对密文数据进行AES解密，解密算法为：

私钥.数据密文AES.decrypt，这里的私钥64位包含数字和大小写字母的字符；

(d)将AES解密的数据返回给用户。

一种基于hbase的数据安全处理方法及系统

技术领域

[0001] 本发明涉及大数据处理和数据安全的技术领域，尤其涉及一种基于hbase的数据安全处理方法，以及基于hbase的数据安全处理系统。

背景技术

[0002] hbase是一个分布式的、面向列的开源数据库，该技术来源于Fay Chang所撰写的Google论文“Bigtable：一个结构化数据的分布式存储系统”。就像Bigtable利用了Google文件系统(File System)所提供的分布式数据存储一样，hbase在Hadoop之上提供了类似于Bigtable的能力。hbase是Apache的Hadoop项目的子项目。hbase不同于一般的关系数据库，它是一个适合于非结构化数据存储的数据库。另一个不同的是hbase基于列的而不是基于行的模式。

[0003] 现有的大数据数据存储过程中，并没有对数据进行加密处理，用户如果有权限便可以随便查看所有hadoop上面的数据，而且hadoop的权限管理相对简单，所以hadoop上面存储的数据存在很大的安全隐患。中国专利申请(申请号：CN201510710555.X)提供了一种实现大数据安全的方法，可以一定程度地解决这种安全隐患。

[0004] 但是，目前还没有非常好的大数据安全解决方案，所有的数据基本都是明文存储到hadoop上，或者以文体文件格式存储，或者以二进制文件存储格式(RCFILE/ORCFILE/HFILE/PARQUET)存储，除了hadoop有限的权限管理来对用户的数据操作进行约束外，还没有很好的解决方案来解决大数据的安全问题。

发明内容

[0005] 为克服现有技术的缺陷，本发明要解决的技术问题是提供了一种基于hbase的数据安全处理方法，其既保证了存储到hadoop上的数据是加密的数据，也保证了用户需要明文时可以通过hbase接口调用获取。

[0006] 本发明的技术方案是：这种基于hbase的数据安全处理方法，该方法包括以下步骤：

[0007] (1)准备明文数据，并存储到hadoop的临时目录；

[0008] (2)通过MapReduce作业分布式并行地加密明文数据，而且加密成两种密文数据：双层带私钥的MD5加密数据和带私钥的AES加密数据；

[0009] (3)将双层带私钥的MD5加密数据存储到hadoop正式数据目录，将双层带私钥的MD5加密数据和带私钥的AES加密数据推送到hbase存储；

[0010] (4)通过用户请求hbase接口，获取密文数据对应的明文。

[0011] 本发明根据数据的特点和应用需求，将非常重要的数据存储到hadoop正式数据目录之前进行加密处理，同时通过hbase提供数据解密接口，拿到数据的明文，这样既保证了存储到hadoop上的数据是加密的数据，也保证了用户需要明文时可以通过hbase接口调用获取。

[0012] 还提供了一种基于hbase的数据安全处理系统，该系统包括：

[0013] 明文数据准备模块，其配置来准备明文数据，并存储到hadoop的临时目录；

[0014] 数据加密模块，其配置来通过MapReduce作业分布式并行地加密明文数据，而且加密成两种密文数据：双层带私钥的MD5加密数据和带私钥的AES加密数据；

[0015] 数据存储模块，其配置来将双层带私钥的MD5加密数据存储到hadoop正式数据目录，将双层带私钥的MD5加密数据和带私钥的AES加密数据推送到hbase存储；

[0016] 数据解密模块，其配置来通过用户请求hbase接口，获取密文数据对应的明文。

附图说明

[0017] 图1所示为根据本发明的基于hbase的数据安全处理方法的流程图。

具体实施方式

[0018] 如图1所示，这种基于hbase的数据安全处理方法，该方法包括以下步骤：

[0019] (1)准备明文数据，并存储到hadoop的临时目录；

[0020] (2)通过MapReduce作业分布式并行地加密明文数据，而且加密成两种密文数据：双层带私钥的MD5加密数据和带私钥的AES加密数据；

[0021] (3)将双层带私钥的MD5加密数据存储到hadoop正式数据目录，将双层带私钥的MD5加密数据和带私钥的AES加密数据推送到hbase存储；

[0022] (4)通过用户请求hbase接口，获取密文数据对应的明文。

[0023] 本发明根据数据的特点和应用需求，将非常重要的数据存储到hadoop正式数据目录之前进行加密处理，同时通过hbase提供数据解密接口，拿到数据的明文，这样既保证了存储到hadoop上的数据是加密的数据，也保证了用户需要明文时可以通过hbase接口调用获取。

[0024] 另外，所述步骤(2)中，双层带私钥的MD5加密数据的私钥设置成既包含数字、大小写字母，还包含了特殊可见字符的64位字符。

[0025] 另外，所述步骤(2)中，带私钥的AES加密数据的私钥设置成包含数字和大小写字母的64位字符。

[0026] 另外，所述步骤(3)中，存储到hbase里面，以MD5加密数据为rowkey，AES加密数据作为value存储；将MD5加密数据和AES加密数据做一个

[0027] (key,value)映射，以方便后期数据解密。

[0028] 另外，所述步骤(4)中，解密的时候hbase接口执行：

[0029] (a)接受用户解密请求，获取用户传递过来的数据密文；

[0030] (b)将数据密文作为rowkey，查询hbase中该rowkey对应的value，这里的value就是AES加密的数据；

[0031] (c)拿到AES加密的数据，并根据私钥对密文数据进行AES解密，解密算法为：

[0032] 私钥.数据密文AES.decrypt，这里的私钥64位包含数字和大小写字母的字符；

[0033] (d)将AES解密的数据返回给用户。

[0034] 本领域普通技术人员可以理解，实现上述实施例方法中的全部或部分步骤是可以通过程序来指令相关的硬件来完成，所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，包括上述实施例方法的各步骤，而所述的存储介质可以是：ROM/RAM、磁碟、光盘、存储卡等。因此，与本发明的方法相对应的，本发明还同时包括一种基于hbase的数据安全处理系统，该系统通常以与方法各步骤相对应的功能模块的形式表示。使用该方法的系统包括：

[0035] 明文数据准备模块，其配置来准备明文数据，并存储到hadoop的临时目录；

[0036] 数据加密模块，其配置来通过MapReduce作业分布式并行地加密明文数据，而且加密成两种密文数据：双层带私钥的MD5加密数据和带私钥的AES加密数据；

[0037] 数据存储模块，其配置来将双层带私钥的MD5加密数据存储到hadoop正式数据目录，将双层带私钥的MD5加密数据和带私钥的AES加密数据推送到hbase存储；

[0038] 数据解密模块，其配置来通过用户请求hbase接口，获取密文数据对应的明文。

[0039] 另外，所述数据加密模块中，双层带私钥的MD5加密数据的私钥设置成既包含数字、大小写字母，还包含了特殊可见字符的64位字符。

[0040] 另外，所述数据加密模块中，带私钥的AES加密数据的私钥设置成包含数字和大小写字母的64位字符。

[0041] 另外，所述数据存储模块中，存储到hbase里面，以MD5加密数据为rowkey，AES加密数据作为value存储；将MD5加密数据和AES加密数据做一个(key,value)映射，以方便后期数据解密。

[0042] 另外，所述数据解密模块中，解密的时候hbase接口执行：

[0043] (a)接受用户解密请求，获取用户传递过来的数据密文；

[0044] (b)将数据密文作为rowkey，查询hbase中该rowkey对应的value，这里的value就是AES加密的数据；

[0045] (c)拿到AES加密的数据，并根据私钥对密文数据进行AES解密，解密算法为：

[0046] 私钥.数据密文AES.decrypt，这里的私钥64位包含数字和大小写字母的字符；

[0047] (d)将AES解密的数据返回给用户。

[0048] 以下将详细说明本发明。

[0049] 第一部分：数据加密

[0050] 数据进入集群之前是明文，这些明文数据暂时存储到hadoop的临时目录下，然后执行MapReduce加密作业，对这些明文数据进行分布式并行加密处理，加密的时候输出两种加密方式的数据，以下是两种加密方式：

[0051] (1)、双层带私钥的MD5加密

[0052] MD5加密是非对称加密，加密之后无法解密，虽然不能解密，但还是可以通过类似撞库的方法撞出明文来，特别如果要加密的数据取值范围是已知的情况下，比如qq号等，为了解决这个问题，在对数据进行MD5加密的时候进行双层MD5加密，同时加上一个足够安全和复杂的私钥，加密算法如下：

[0053] MD5(“私钥”+MD5(“数据明文”))

[0054] 为了使得加密的数据足够安全，不被撞库撞出来，私钥要足够复杂，不便于记忆和丢失，这里的私钥设置成既包含数字、大小写字母，还包含了特殊可见字符组成的64位字符，比如：

[0055] deLQh+-Fa2KKXKOop3p/mj2UvryQzlaARR7ZhI1n3Q8P6pvjc8V2ZT5OoDCcSj-9

[0056] (2)、带私钥的AES加密

[0057] AES加密是对称加密，也即加密的数据是可以解密的，同MD5加密一样，在使用AES加密时也需要设置一个私钥，这个私钥不是为了防撞库的，这个私钥是在对数据密文进行解密的时候要用到，以下是加密算法：

[0058] AES.encrypt(“私钥”,”数据明文”)

[0059] 其中的私钥也是非常重要，需要设置得足够复杂，这里我们把也私钥设置成既包含数字和大小写字母的64位字符，比如：

[0060] zZxCe83v1aSmk9kN4AOEqAfuckztC5tt831trkASsbs4gYOZaFchqCsbjldVBaGSw

[0061] 第二部分:数据存储

[0062] 上一部分中数据被加密成两种密文，一种是双层带私钥的MD5加密数据，一种是带私钥的AES加密数据，这两种数据都有各自不同的存储目的。MD5加密的数据存储到hadoop里面，在组织内部所有人员使用的这种数据都是MD5加密的，保障数据在组织内部流转的时候是安全的。带私钥的AES加密数据将和双层带私钥的MD5加密数据一起被推送到hbase，存储到hbase里面，hbase存储表结构如下：

[0063] rowkey:双层带私钥的MD5加密数据

[0064] value:带私钥的AES加密数据

[0065] 这种存储方式是将MD5加密数据和AES加密数据做一个(key,value)映射，以方便后期数据解密。

[0066] 第三部分:数据解密

[0067] 数据经过加密和存储，平时数据流转过程中，组织内部使用的都是hadoop上面那份MD5加密数据，但是当这种加密数据随其他和业务相关的数据在不同的业务场景下经过处理之后，是需要将这些数据以明文的方式对外提供，这个时候就需要对密文数据进行解密了，解密的方法是将MD5加密的密文作为参数调用hbase接口，获取对应的明文信息。这里解密的时候hbase接口需要做以下工作：

[0068] (1)、接受用户解密请求，获取用户传递过来的数据密文

[0069] (2)、将数据密文作为rowkey，查询hbase中该rowkey对应的value，这里的value也就是AES加密的数据

[0070] (3)、拿到AES加密的数据，并根据私钥对密文数据进行AES解密，解密算法如下：

[0071] AES.decrypt(“私钥”,”数据密文”)

[0072] 这里的私钥既是上文的64位包含数字和大小写字母的字符：

[0073] zZxCe83v1aSmk9kN4AOEqAfuckztC5tt831trkASsbs4gYOZaFchqCsbjldVBaGSw

[0074] (4)、将AES解密的数据返回给用户

[0075] 至此，以上完成了从数据加密，数据存储，在到数据解密的整个过程，解决了大数据安全存储问题。

[0076] 本发明的有益效果如下：

[0077] 1.数据hadoop存储使用双层带私钥的MD5加密。

[0078] 2.数据解密调用hbase接口，而且hbase中存储的数据时AES加密的密文。

[0079] 以上所述，仅是本发明的较佳实施例，并非对本发明作任何形式上的限制，凡是依据本发明的技术实质对以上实施例所作的任何简单修改、等同变化与修饰，均仍属本发明

技术方案的保护范围。

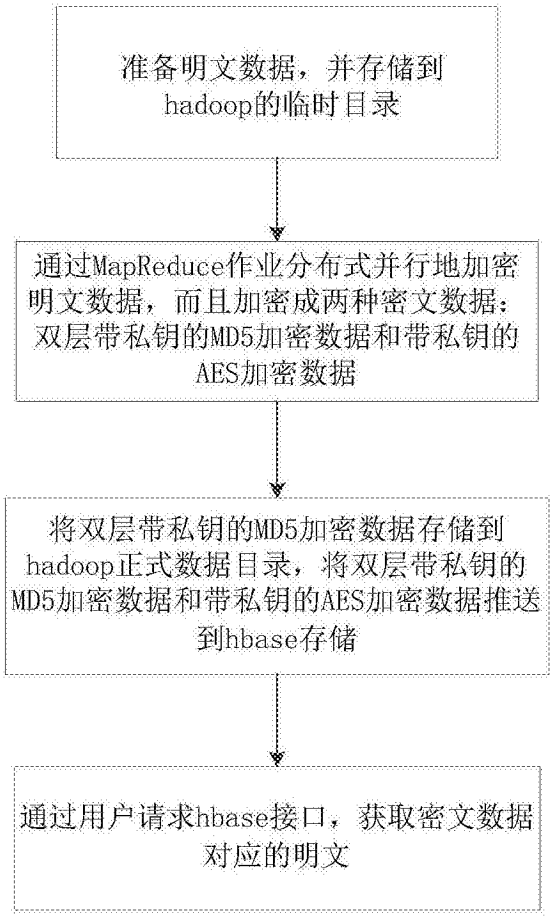

准备明文数据，并存储到hadoop的临时目录
通过MapReduce作业分布式并行地加密明文数据，而且加密成两种密文数据：双层带私钥的MD5加密数据和带私钥的AES加密数据
将双层带私钥的MD5加密数据存储到hadoop正式数据目录，将双层带私钥的MD5加密数据和带私钥的AES加密数据推送到hbase存储
通过用户请求hbase接口，获取密文数据对应的明文

图1