



(12) 发明专利

(10) 授权公告号 CN 110287172 B

(45) 授权公告日 2023.05.02

(21) 申请号 201910588013.8

G06F 16/27 (2019.01)

(22) 申请日 2019.07.01

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 105468735 A, 2016.04.06

申请公布号 CN 110287172 A

US 2013204948 A1, 2013.08.08

(43) 申请公布日 2019.09.27

CN 109299068 A, 2019.02.01

(73) 专利权人 四川新网银行股份有限公司

US 2013282668 A1, 2013.10.24

地址 610094 四川省成都市成都高新区吉

CN 109271365 A, 2019.01.25

泰三路8号1栋1单元26楼1-8号

丁晶. 基于Hadoop系统大数据平台在天津市

(72) 发明人 李烨

地震局的应用.《电子技术与软件工程》.2017,

(74) 专利代理机构 成都智言知识产权代理有限

秦东霞. 基于Hadoop的云平台设计与实现.

公司 51282

《智能计算机与应用》.2016,

专利代理师 徐金琼

(51) Int. Cl.

G06F 16/21 (2019.01)

G06F 16/25 (2019.01)

权利要求书1页 说明书4页 附图1页

(54) 发明名称

一种格式化HBase数据的方法

(57) 摘要

本发明公开了一种格式化HBase数据的方法,属于数据格式化领域,解决现有技术进行化HBase数据格式化,其操作较为繁琐,耗时较长的问题。本发明停止HBase集群的所有服务,同时保持HBase集群依赖的Zookeeper和Hadoop仍处于正常运行状态;在HBase集群上先删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点,在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录;删除后,启动HBase集群的所有服务,得到初始状态的HBase。本发明用于快速格式化HBase数据。



1. 一种格式化HBase数据的方法,其特征在于,如下步骤:

S1、停止HBase集群的所有服务,同时保持HBase集群依赖的Zookeeper和Hadoop仍处于正常运行状态;

S2、执行步骤S1后,在HBase集群上先删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点,再在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录;或在HBase集群上先删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录,再在HBase集群上删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点;或在HBase集群上先删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点,同时在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录;

其中在HBase集群上删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点的具体实现过程为:在HBase集群的配置文件hbase-site.xml的zookeeper.znode.parent标签中找到Zookeeper上存储HBase元数据的根节点,找到后,在Zookeeper上删除根节点及根节点下包含的所有子节点;在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录的具体实现过程为:在HBase集群的配置文件hbase-site.xml的hbase.rootdir标签中找到Hadoop上存储HBase数据的根目录,找到后,在Hadoop上删除根目录及根目录下包含的所有子目录;

S3、删除后,启动HBase集群的所有服务,即得到初始状态的HBase。

2. 根据权利要求1所述的一种格式化HBase数据的方法,其特征在于,处理器接收到格式化HBase数据的请求,停止HBase集群的所有服务,同时保持HBase集群依赖的Zookeeper和Hadoop仍处于正常运行状态;然后,处理器根据查询删除指令,处理器调用存储器中的查询和删除程序,在HBase集群上先删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点,再在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录;或处理器根据查询删除指令,处理器调用存储器中的查询和删除程序,在HBase集群上先删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录,再在HBase集群上删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点;或处理器根据查询删除指令,处理器调用存储器中的查询和删除程序,在HBase集群上先删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点,同时在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录;删除后,处理器启动HBase集群的所有服务,即得到初始状态的HBase。

一种格式化HBase数据的方法

技术领域

[0001] 一种格式化HBase数据的方法,用于快速格式化HBase数据,属于数据格式化领域。

背景技术

[0002] 数据格式化指删除系统中的全部数据和元数据,将系统还原到初始状态。当系统中的数据不再有用,或者系统状态异常时,通过执行数据格式化,可以快速使系统恢复到干净的可用的状态。

[0003] Zookeeper: ZooKeeper是一个分布式的,开放源码的分布式应用程序协调服务,是Google的Chubby一个开源的实现,是Hadoop和HBase依赖的重要组件,目前是Apache社区的顶级开源项目。它是一个为分布式应用提供一致性服务的软件,提供的功能包括:配置维护、域名服务、分布式同步、组服务等。

[0004] Hadoop: Hadoop包含一个分布式文件系统HDFS和分布式计算框架MapReduce,目前是Apache社区的顶级项目。Hadoop具有高容错性的特点,并且设计用来部署在低廉的硬件上,而且它提供高吞吐量来访问应用程序的数据,适合那些有着超大数据集的应用程序。

[0005] HBase是非常流行的一个分布式的、面向列的NoSQL数据库,是Apache社区的顶级开源项目,其应用场景主要是海量数据的存储和高并发条件下固定条件的检索。在开发测试环境,当HBase中的数据不再有用,或HBase状态异常时,通过对HBase数据执行格式化,可以快速得到一个初始状态的HBase,即没有任何数据的HBase。HBase的运行依赖于Zookeeper和Hadoop,其元数据存储于Zookeeper上,数据存储于Hadoop上。HBase本身没有提供格式化的方法或工具,在公开的专利中没有检索到关于格式化HBase的专利,在互联网上也没有类似本文描述的格式化HBase的方法的详细介绍。容易想到的一种能实现相同目的的解决方案是,卸载原HBase集群,即需要删除HBase所有数据、元数据、软件包、配置文件等,重新搭建一套全新的HBase集群(需要在HBase集群每个节点重新安装软件包、重新设置配置文件),但这种方法操作较为繁琐,耗时较长。

发明内容

[0006] 针对上述研究的问题,本发明的目的在于提供一种格式化HBase数据的方法,解决现有技术中通过卸载原HBase集群,重新搭建一套全新的HBase集群来进行HBase数据格式化,其操作较为繁琐,耗时较长的问题。

[0007] 为了达到上述目的,本发明采用如下技术方案:

[0008] 一种格式化HBase数据的方法,如下步骤:

[0009] S1、停止HBase集群的所有服务,同时保持HBase集群依赖的Zookeeper和Hadoop仍处于正常运行状态;

[0010] S2、执行步骤S1后,在HBase集群上先删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点,再在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录;或在HBase集群上先删除Hadoop上存储HBase数据的根目录及

根目录下包含的所有子目录,再在HBase集群上删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点;或在HBase集群上先删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点,同时在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录;

[0011] S3、删除后,启动HBase集群的所有服务,即得到初始状态的HBase。

[0012] 进一步,所述步骤S2中,

[0013] 在HBase集群上删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点的具体实现过程为:在HBase集群的配置文件hbase-site.xml的zookeeper.znode.parent标签中找到Zookeeper上存储HBase元数据的根节点,找到后,在Zookeeper上删除根节点及根节点下包含的所有子节点;

[0014] 在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录的具体实现过程为:在HBase集群的配置文件hbase-site.xml的hbase.rootdir标签中找到Hadoop上存储HBase数据的根目录,找到后,在Hadoop上删除根目录及根目录下包含的所有子目录。

[0015] 进一步,处理器接收到格式化HBase数据的请求,停止HBase集群的所有服务,同时保持HBase集群依赖的Zookeeper和Hadoop仍处于正常运行状态;

[0016] 然后,处理器根据查询删除指令,处理器调用存储器中的查询和删除程序,在HBase集群上先删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点,再在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录;或处理器根据查询删除指令,处理器调用存储器中的查询和删除程序,在HBase集群上先删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录,再在HBase集群上删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点;或处理器根据查询删除指令,处理器调用存储器中的查询和删除程序,在HBase集群上先删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点,同时在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录;

[0017] 删除后,处理器启动HBase集群的所有服务,即得到初始状态的HBase。

[0018] 本发明同现有技术相比,其有益效果表现在:

[0019] 一、本发明通过删除HBase集群在Zookeeper上存储的全部元数据和Hadoop上存储的数据,简化了实现步骤,降低了操作的繁琐度,从而快速实现HBase数据格式化,即实现了计算机对内部对象进行处理的优化解决方案。

附图说明

[0020] 图1是本发明中先删除Zookeeper上存储的全部元数据,再删除Hadoop上存储的数据的流程示意图。

具体实施方式

[0021] 下面将结合附图及具体实施方式对本发明作进一步的描述。

[0022] 一种格式化HBase数据的方法,如下步骤:

[0023] S1、停止HBase集群的所有服务,同时保持HBase集群依赖的Zookeeper和Hadoop仍

处于正常运行状态；

[0024] S2、执行步骤S1后，在HBase集群上先删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点，再在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录；或在HBase集群上先删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录，再在HBase集群上删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点；或在HBase集群上先删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点，同时在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录；

[0025] HBase集群上删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点的具体实现过程为：在HBase集群的配置文件hbase-site.xml的zookeeper.znode.parent标签中找到Zookeeper上存储HBase元数据的根节点，找到后，在Zookeeper上删除根节点及根节点下包含的所有子节点；

[0026] 在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录的具体实现过程为：在HBase集群的配置文件hbase-site.xml的hbase.rootdir标签中找到Hadoop上存储HBase数据的根目录，找到后，在Hadoop上删除根目录及根目录下包含的所有子目录。

[0027] 在上述查找删除中，采用手动的方式在HBase集群的配置文件hbase-site.xml进行查找并根据查找结果进行删除，即通过肉眼查看找到相应内容并给出删除指令进行删除；或通过程序接收到查到指令后，自动在HBase集群的配置文件hbase-site.xml进行查找并根据查找结果进行删除，其中，查找Zookeeper上存储HBase元数据的根节点的程序：即编写XML解析程序（如调用DOM4J等常用的XML解析库），从hbase-site.xml中找出<name>zookeeper.znode.parent</name>标记对应的<value>...</value>标记的值，然后执行该程序，进行查找。查找Hadoop上存储HBase数据的根目录的程序：编写XML解析程序（如调用DOM4J等常用的XML解析库），从hbase-site.xml中找出<name>hbase.rootdir</name>标记对应的<value>...</value>标记的值，然后执行该程序，进行查找；删除的程序为：删除Zookeeper上的节点可采用zkCli.sh脚本，或者Zookeeper的删除节点的Java API或其他语言的API等；删除Hadoop上的目录可采用hdfs dfs-rm-r<目录>或hadoop fs-rm-r<目录>这两种hadoop自带的命令均可，或者使用Hadoop的删除目录的Java API或其他语言的API。

[0028] S3、删除后，启动HBase集群的所有服务，即得到初始状态的HBase。

[0029] 实现格式化的数据流如下：

[0030] 处理器接收到格式化HBase数据的请求，停止HBase集群的所有服务，同时保持HBase集群依赖的Zookeeper和Hadoop仍处于正常运行状态；

[0031] 处理器根据查询删除指令，处理器调用存储器中的查询和删除程序，在HBase集群上先删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点，再在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录；或处理器根据查询删除指令，处理器调用存储器中的查询和删除程序，在HBase集群上先删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录，再在HBase集群上删除Zookeeper上存储HBase元数据的根节点及根节点下包含的所有子节点；或处理器根据查询删除指令，处理器调用存储器中的查询和删除程序，在HBase集群上先删除Zookeeper上存

储HBase元数据的根节点及根节点下包含的所有子节点,同时在HBase集群上删除Hadoop上存储HBase数据的根目录及根目录下包含的所有子目录;

[0032] 删除后,处理器启动HBase集群的所有服务,即得到初始状态的HBase。

[0033] 以上仅是本发明众多具体应用范围中的代表性实施例,对本发明的保护范围不构成任何限制。凡采用变换或是等效替换而形成的技术方案,均落在本发明权利保护范围之内。

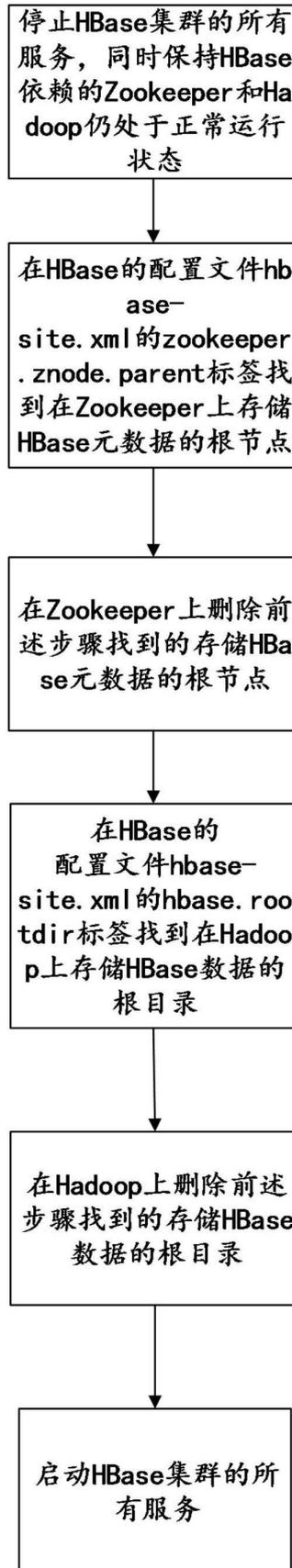


图1