



(12)发明专利

(10)授权公告号 CN 103064670 B

(45)授权公告日 2016.12.07

(21)申请号 201210553873.6

(22)申请日 2012.12.18

(65)同一申请的已公布的文献号  
申请公布号 CN 103064670 A

(43)申请公布日 2013.04.24

(73)专利权人 清华大学  
地址 100084 北京市海淀区100084-82信箱

(72)发明人 黄永峰 罗芳 郑枫 陈跻

(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

(51)Int.Cl.  
G06F 9/44(2006.01)  
G06F 17/30(2006.01)

(56)对比文件

US 2011154341 A1,2011.06.23,  
CN 102739470 A,2012.10.17,

卞云龙.基于云计算平台的大规模流形学习  
算法研究.《中国优秀硕士学位论文全文数据  
库》.2012,

张烨.Jackrabbit封装Hadoop的研究及在内  
容管理系统中的应用.《中国优秀硕士学位论文  
全文数据库》.2011,

审查员 杨蕊

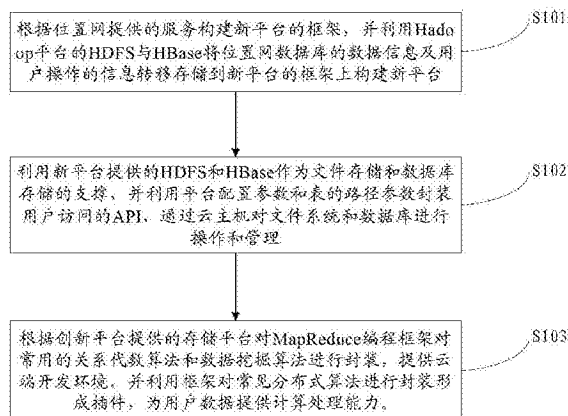
权利要求书2页 说明书8页 附图2页

(54)发明名称

基于位置网的创新平台数据管理方法及系  
统

(57)摘要

本发明提出一种基于位置网的创新平台数  
据管理方法及系统。其中,方法包括:根据位置网  
提供的服务构建创新平台的框架,并利用Hadoop  
平台的HDFS与HBase将位置网数据库的数据信息  
及用户操作的信息转移存储到创新平台的框架  
上构建创新平台;利用创新平台提供的HDFS和  
HBase作为文件存储和数据库存储的支撑,并利  
用平台参数封装用户访问的API,使用户通过云  
桌面实现对分布式文件系统和数据库的操作与  
管理;以及根据MapReduce编程框架对常用的关  
系代数算法和分布式数据挖掘算法进行封装并  
集成开发环境进行数据预处理。根据本发明实施  
的方法,通过位置网位置与服务数据的转移与计  
算,为位置网的构建提供了相关统计数据信息,  
还对该平台API进行扩展提供了数据计算操作与  
管理接口。



1. 一种基于位置网的创新平台数据管理方法,其特征在于,包括以下步骤:

S1:根据位置网提供的服务构建创新平台的框架,并利用Hadoop平台的HDFS与HBase将所述位置网数据库的数据信息及用户操作的信息转移存储到创新平台的框架上并进行数据挖掘统计构建创新平台;

S2:利用所述创新平台提供的HDFS和HBase作为文件存储和数据库存储的支撑,并利用平台配置参数和表的路径参数封装用户访问的API,通过云主机对文件系统和数据库进行操作和管理;以及

S3:根据所述创新平台提供的存储平台对MapReduce编程框架对常用的关系代数算法和数据挖掘算法进行封装,提供云端开发环境,并利用所述创新平台的框架对分布式算法进行封装形成插件,为用户数据提供计算处理能力,

所述步骤S1具体包括:

S11:在所述位置网提供的服务器平台上申请云服务器的使用;

S12:将所述位置网的位置和服务数据转移到创新平台的数据库中;

S13:对转移的所述位置网的位置和服务数据根据不同数据挖掘算法进行分析;以及

S14:将分析后的数据进行统计并显示;

所述步骤S2进一步包括:

S21:搭建支撑底层存储与计算的Hadoop平台,其中,所述Hadoop平台包括HDFS、MapReduce和HBase;

S22:整合Hadoop平台提供的所述HDFS和HBase相关的API;

S23:将所述HDFS和HBase相关的API进行二次开发和封装;以及

S24:调用封装后的所述HDFS和HBase相关的API生成数据存储管理界面对用户数据进行存储和管理。

2.如权利要求1所述的基于位置网的创新平台数据管理方法,其特征在于,所述步骤S3具体包括:

S31:对关系代数和数据挖掘算法采用MapReduce框架实现并进行封装形成插件;

S32:封装集成基于Hadoop的分布式开发环境;以及

S33:根据MapReduce框架进行封装形成的插件和集成平台对用户数据进行处理。

3.如权利要求1或2所述的基于位置网的创新平台数据管理方法,其特征在于,所述HDFS和HBase作为底层的存储支持,以所述HDFS和HBase提供的API接口作为数据操作与管理接口的调用接口,并基于MapReduce框架的封装算法插件及开发接口提供用户数据的处理和计算。

4.一种基于位置网的创新平台数据管理系统,其特征在于,包括:

创建模块,用于根据位置网的提供服务构建创新平台的框架,并利用HDFS与HBase将所述位置网数据库的数据信息及用户操作的信息转移存储到创新平台的框架上构建创新平台;

封装模块,用于利用所述创新平台提供的HDFS和HBase作为文件存储和数据库存储的支撑,并利用平台配置参数和表的路径参数封装用户访问的API,通过云主机对文件系统和数据库进行操作和管理;以及

插件模块,用于根据所述创新平台提供的存储平台对MapReduce编程框架对常用的关系代数算法和数据挖掘算法进行封装,提供云端开发环境,并利用所述创新平台的框架对分布式算法进行封装形成插件,为用户数据提供计算处理能力;

所述创建模块具体包括:

申请单元,用于在所述位置网提供的服务器平台上申请云服务器的使用;

转移单元,用于将所述位置网的位置和服务数据转移到创新平台的数据库中;

分析单元,用于对转移的所述位置网的位置和服务数据根据不同数据挖掘算法进行数据分析;以及

统计显示单元,用于将分析后的数据进行统计并显示;

所述封装模块具体包括:

搭建单元,用于搭建支撑底层存储与计算的Hadoop平台,其中,所述Hadoop平台包括HDFS、MapReduce和HBase;

整合单元,用于整合Hadoop平台提供的所述HDFS和HBase相关的API;

封装单元,用于将所述HDFS和HBase相关的API进行二次开发和封装;以及

操作管理单元,用于调用封装后的所述HDFS和HBase相关的API生成数据存储管理界面对用户数据进行存储和管理。

5.如权利要求4所述的基于位置网的创新平台数据管理系统,其特征在于,所述插件模块具体包括:

插件单元,用于对关系代数和数据挖掘算法采用MapReduce框架实现并进行封装形成插件;

集成单元,用于封装集成基于Hadoop的分布式开发环境;以及

处理单元,用于根据MapReduce框架进行封装形成的插件和集成平台对用户数据进行处理。

6.如权利要求4或5所述的基于位置网的创新平台数据管理系统,其特征在于,所述HDFS和HBase作为底层的存储支持,以所述HDFS和HBase提供的API接口作为数据操作与管理接口的调用接口,并基于MapReduce框架的封装算法插件及开发接口提供用户数据的处理和计算。

## 基于位置网的创新平台数据管理方法及系统

### 技术领域

[0001] 本发明涉及云计算领域,特别涉及一种基于位置网的创新平台数据管理方法及系统。

### 背景技术

[0002] 位置网的服务器存储了大量的用户位置信息资源、网络分布位置信息资源、相关服务信息资源,提供了相关的服务扩展平台,因此我们在其提供的服务平台上构建了位置网创新平台。同时由于位置网采用MySQL数据库作为其位置数据存储支持,随着用户数量的增加,这个数据存储方式将不适用于海量数据的存储,因此需要采用一个可存储海量数据并具有良好的存取实时性的数据库作为位置数据的存储支持,因此我们可以考虑使用Hadoop的HBase分布式数据库来替代MySQL数据库,因为HBase数据库具有良好的可扩展性,实时性,适于位置数据信息的存储。位置网的日志文件也记载了大量的网站数据操作日志信息。这些日志信息从一定程度上反映了用户偏好,可通过对日志信息的分析获取如网站访问的时间段偏好,网站访问偏好,位置信息访问偏好等信息。因此采用基于MapReduce框架的相关数据挖掘算法对海量的获取相关统计信息,对于位置网的进一步构建能够起到一定的引导作用。

[0003] Hadoop是分布式的计算平台,是对Google的云计算平台的开源实现。主要由HDFS和MapReduce构成,HDFS主要用于实现对数据的分布式存储,MapReduce框架主要用于对大规模数据的分布式计算。HBase是基于HDFS与MapReduce框架的一个面向列的适合于非结构化数据存储的分布式数据库。由于Hadoop平台集分布式存储与计算功能于一体,而且便于布置在大量的廉价机器上及其具有易扩展、处理速度快、安全性高、易实现等特点,使得Hadoop平台无论是在工业界还是在学术界都受到了越来越多的重视,而对Hadoop平台的研发主要分为针对Hadoop平台本身的优化改进以及基于Hadoop平台的上层的应用开发。我们实现的基于Hadoop平台的云主机计算技术主要是基于Hadoop平台的上层应用开发,同时根据时间及性能需求实行相应的改进与优化。

[0004] Hadoop本身也为HDFS提供了网页显示和简单的操作功能,但是提供的操作功能很少,不实用,而且网页的显示形式不适合于普通用户的使用,用户体验差。而基于tomcat和Web-Dav虽然已经能够将Hadoop文件系统mount到本地,但是这也需要用户进行下载、配置安装相关软件,这对于某些用户(如数据分析人员)来说显然是繁琐的,而且它现在提供的功能还不完善。同时Hadoop平台提供的HBase分布式数据库提供了不同语言的访问接口及可通过hbase shell命令进行操作,但是这些方式都使得数据库不适合于普通用户的使用。

### 发明内容

[0005] 本发明的目的旨在至少解决上述的技术缺陷之一。

[0006] 为达到上述目的,本发明一方面的实施例提出一种基于位置网的创新平台数据管理方法,包括以下步骤:S1:根据位置网提供的服务构建创新平台的框架,并利用Hadoop平

台的HDFS与HBase将所述位置网数据库的数据信息及用户操作的信息转移存储到创新平台的框架上构建创新平台；S2：利用所述创新平台提供的HDFS和HBase作为文件存储和数据库存储的支撑，并利用平台配置参数和表的路径参数封装用户访问的API，通过云主机对文件系统和数据库进行操作和管理；以及S3：根据所述创新平台提供的存储平台对MapReduce编程框架对常用的关系代数算法和数据挖掘算法进行封装，提供云端开发环境，并利用所述框架对常见分布式算法进行封装形成插件，为用户数据提供计算处理能力。

[0007] 根据本发明实施例的方法，通过位置网位置与服务数据的转移与计算学习，为位置网的进一步构建提供了相关统计数据信息，还对该平台API进行扩展提供了相应的数据操作管理与计算接口，同时提供了程序开发平台满足用户开发需求。

[0008] 本发明的一个实施例中，所述步骤S1具体包括：S11：在所述位置网提供的服务器平台上申请云服务器的使用；S12：将所述位置网的位置和服务数据转移到创新平台的数据库中；S13：对所述转移数据根据不同数据挖掘算法进行数据分析；以及S14：将所述分析数据进行统计并显示。

[0009] 本发明的一个实施例中，所述步骤S2具体包括：S21：搭建支撑底层存储与计算的Hadoop平台，其中，所述Hadoop平台包括HDFS、MapReduce和HBase；S22：整合Hadoop平台提供的所述HDFS和HBase相关的API；S23：将所述API进行二次开发和封装；以及S24：调用所述封装的API生成数据存储管理界面对用户数据进行存储和管理。

[0010] 本发明的一个实施例中，所述步骤S3具体包括：S31：对关系代数和数据挖掘算法采用MapReduce框架实现并进行封装形成插件；S32：封装集成基于Hadoop的分布式开发环境；以及S33：根据所述插件和集成平台对用户数据进行处理。

[0011] 本发明的一个实施例中，所述HDFS和HBase作为底层的存储支持，以所述HDFS和HBase提供的API接口作为数据操作与管理接口的调用接口，并基于MapReduce框架的封装算法插件及开发接口提供用户数据的处理和计算。

[0012] 为达到上述目的，本发明的实施例另一方面提出一种基于位置网的创新平台数据管理系统，包括：创建模块，用于根据位置网的提供服务构建创建平台的框架，并利用HDFS与HBase将所述位置网数据库的数据信息及用户操作的信息转移存储到创新平台的框架上构建创新平台；封装模块，用于利用所述创新平台提供的HDFS和HBase作为文件存储和数据库存储的支撑，并利用平台配置参数和表的路径参数封装用户访问的API，通过云主机对文件系统和数据库进行操作和管理；以及插件模块，用于根据所述创新平台提供的存储平台对MapReduce编程框架对常用的关系代数算法和数据挖掘算法进行封装，提供云端开发环境，并利用所述框架对常见分布式算法进行封装形成插件，为用户数据提供计算处理能力。

[0013] 根据本发明实施例的系统，通过位置网位置与服务数据的转移与计算学习，为位置网的进一步构建提供了相关统计数据信息，还对该平台API进行扩展提供了相应的数据操作管理与计算接口，同时提供了程序开发平台满足用户开发需求。

[0014] 本发明的一个实施例中，所述创建模块具体包括：申请单元，用于在所述位置网提供的服务器平台上申请云服务器的使用；转移单元，用于将所述位置网的位置和服务数据转移到创新平台的数据库中；分析单元，用于对所述转移数据根据不同数据挖掘算法进行数据分析；以及统计显示单元，用于将所述分析数据进行统计并显示。

[0015] 本发明的一个实施例中，所述封装模块具体包括：搭建单元，用于搭建支撑底层存

储与计算的Hadoop平台,其中,所述Hadoop平台包括HDFS、MapReduce和HBase;整合单元,用于整合Hadoop平台提供的所述HDFS和HBase相关的API;封装单元,用于将所述API进行二次开发和封装;以及操作管理单元,用于调用所述封装的API生成数据存储管理界面对用户数据进行存储和管理。

[0016] 本发明的一个实施例中,所述插件模块包括:插件单元,用于对关系代数和数据挖掘算法采用MapReduce框架实现并进行封装形成插件;集成单元,用于封装集成基于Hadoop的分布式开发环境;以及处理单元,用于根据所述插件和集成平台对用户数据进行处理。

[0017] 本发明的一个实施例中,所述HDFS和HBase作为底层的存储支持,以所述HDFS和HBase提供的API接口作为数据操作与管理接口的调用接口,并基于MapReduce框架的封装算法插件及开发接口提供用户数据的处理和计算。

[0018] 本发明附加的方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本发明的实践了解到。

## 附图说明

[0019] 本发明上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解,其中:

[0020] 图1为根据本发明一个实施例的基于位置网的创新平台数据管理方法的流程图;

[0021] 图2为根据本发明一个实施例的步骤S101的流程图;

[0022] 图3为根据本发明一个实施例的基于位置网的创新平台数据管理系统的框架图;

[0023] 图4为根据本发明一个实施例的创建模块的框架图;

[0024] 图5为根据本发明一个实施例的封装模块的框架图;以及

[0025] 图6为根据本发明一个实施例的插件模块的框架图。

## 具体实施方式

[0026] 下面详细描述本发明的实施例,实施例的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。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,而不能解释为对本发明的限制。

[0027] 图1为本发明实施例的基于位置网的创新平台数据管理方法的流程图。如图1所示,根据本发明实施例的基于位置网的创新平台数据管理方法,包括以下步骤:

[0028] 步骤S101,根据位置网提供的服务构建创新平台的框架,并利用Hadoop平台的HDFS与HBase将位置网数据库的数据信息及用户操作的信息转移存储到创新平台的框架上构建创新平台。

[0029] S201,在位置网提供的服务器平台上申请云服务器的使用。在利用位置网提供的服务平台上,为用户显示了云端服务器信息,如内核、内存、硬盘的基本信息及其使用情况,使用户通过相关参数进行云服务器的使用申请。申请成功后用户将获取申请的云主机的登陆名及密码,利用该用户名及密码即可到平台上登陆进行创新平台。该创新平台是以相关web技术作为前台支持,以Hadoop框架作为后台存储与计算支持实现的。

[0030] 在本发明的一个实施例中,在位置网服务平台上搭建创新平台框架时,computerInfo提供了云服务器的主机信息,如cpuInfo,memInfo,diskInfo信息,用户可通

过相关appPara参数来申请使用创新平台。在利用位置网位置服务及日志数据信息构建创新平台过程主要涉及了如表1的参数。

[0031] 表1

[0032]

字段名称	字段属性	备注
computerInfo	主机信息	提供服务器相关信息
appPara	申请参数	通过appPara申请平台
attribute	数据属性	位置数据属性信息
posData	数据源	位置数据
userLog	日志数据	网站操作日志数据
webData	信息源	网站来源
opTime	操作时间	统计用户操作时间段
opType	操作类型	统计用户操作类型

[0033] S202,将位置网的位置和服务数据转移到创新平台的数据库中。位置网的位置信息数据采用MySQL数据库进行存储。在构建创新平台基础服务中,将源数据库存储的数据进行转移,采用HBase分布式数据库,定义需要的存储结构与字段,实现源数据的转移存储。位置网的其他信息数据,如用户操作日志信息等,是以文件格式进行存储的,转移时采用HDFS作为主要存储支持,其中,网络日志数据是由log文件格式进行存储,则可以采用简单的API将其转移到HDFS中存储,然后根据需求再决定是否将HDFS上存储的日志数据进行解析转存储到HBase数据库中。

[0034] S203,对转移数据根据不同数据挖掘算法进行数据分析。在对位置网位置与服务等数据信息进行转移后,根据应用需求,基于Hadoop分布式系统基础架构对转移的信息数据采用不同的数据挖掘算法进行数据分析,解析出位置网的相关数据信息与用户信息,用于位置网的进一步构建。

[0035] 在本发明的一个实施例中,首先根据源位置数据的属性字段attribute设置HBase的存储字段,然后将posData和userLog数据进行数据转移到创新平台的HBase和HDFS中进行存储,再根据不同需求采用对应的数据挖掘算法对移植的数据进行分析,获取webData、opTime、opType等相关数据的统计信息,最后对该统计信息采用趋势图在开发平台上显示出来方便管理查看。对数据操作与管理过程中,主要实现了如下操作:文件操作,主要包括upload、download、delete、create、read、write等操作;数据库操作,主要包括create、scan、put、get、delete、alert等操作;文件与数据库管理,主要包括setComputer、setMode、setOwn、setShared、setVisible等操作。

[0036] S204,将分析数据进行统计并显示。将位置数据经过计算后形成的统计结果,例如,webData、opTime、opType等相关数据的统计信息,采用趋势图在开发平台上显示出来方便管理查看。根据其数据输出格式,进行二次处理生成相应的动态趋势变化图或表单在创新平台上显示出来。

[0037] 步骤S102,利用创新平台提供的HDFS和HBase作为文件存储和数据库存储的支撑,并利用平台配置参数和表的路径参数封装用户访问的API,通过云主机对文件系统和数据库进行操作和管理。

[0038] 搭建支撑底层存储与计算的Hadoop平台,其中,Hadoop平台包括HDFS、MapReduce和HBase。然后,整合Hadoop平台提供的HDFS和HBase相关的API,并对其进行二次开发和封装。调用封装的API生成数据存储管理界面对用户数据进行存储和管理。

[0039] 在本发明的一个实施例中,首先结合数据管理相关参数如UserHome、para\_Cpu, para\_Mem,para\_Disk等来为用户配置对应的创新平台环境,然后以Hadoop平台自身提供的操作API、相关配置参数(如Conf,FileSystem)为基础,结合数据操作参数如filePath及tableName等实现文件及数据库操作接口。在管理过程中,提供了数据的各项属性设置,用户可通过对para\_Mode,para\_Own,isShared,isVisible等参数对数据及用户自身属性进行制定和修改。

[0040] 在本发明的一个实施例中,当后台的API函数获取到传递的参数后则执行相关调用操作,与Hadoop平台进行交互,并最终将获取的返回结果回馈到用户端,完成一次数据操作。提供的数据操作主要包括对文件系统HDFS和对HBase数据库的操作与管理。测试结果表明当对Hadoop集群进行有效的监控管理,通过云端对平台进行相关数据操作具有较好的稳定性和实时性。

[0041] 在本发明的一个实施例中,基于云桌面利用Hadoop平台API的图形界面方便用户的使用,并为用户屏蔽后台的操作,同时采用对异构数据的组织和管理机制扩展前台的功能,丰富用户的文件管理功能。

[0042] 在本发明的一个实施例中,用户在执行相关操作后,创新平台会向后台提交相关操作参数,操作主要涉及的参数如表2所示。

[0043] 表2

[0044]

操作类型	字段名称	字段属性	备注
适用于所有Hadoop平台基本操作	Conf	平台配置参数	用于获取平台配置信息
	FileSystem	HDFS 文件系统	访问 HDFS 文件系统参数
	HTable	HBase 数据库	访问 HBase 数据库参数
数据操作	filePath	文件路径	对文件进行各种操作参数
	tableName	表名	用于对数据库操作参数
	Column	列名	用于对数据库的操作参数
数据管理	UserHome	用户根目录	创建独立的用户根目录
	para_Mode	文件属性设置	指定文件读写执行属性
	para_Own	文件属主设置	用于设置文件的属主及属组

[0045]

	isShared	文件共享设置	设置文件是否共享特性。
	isVisible	用户可见设置	是否对其他用户可见

[0046] 步骤S103,根据创新平台提供的存储平台对MapReduce编程框架对常用的关系代



数算法和数据挖掘算法进行封装,提供云端开发环境,并利用框架对常见分布式算法进行封装形成插件,为用户数据提供计算处理能力。

[0047] 首先,对关系代数和数据挖掘算法采用MapReduce框架实现并进行封装形成插件。然后,封装集成基于Hadoop的分布式开发环境。之后,根据插件和集成平台对用户数据进行处理。

[0048] 在本发明的一个实施例中,提供的相关数据计算插件是基于Hadoop平台实现的,主要集成了一些典型的分布式数据挖掘算法,基于MapReduce框架进行实现并形成插件作为集成开发环境的一部分提供给开发者使用,免去了重复开发开销;集成开发平台的实现是采用对MapReduce框架进行填充的方式进行功能实现和扩展的,MapReduce框架是为了简化并行程序的开发复杂度而设计的,因此我们在上面增加了一层图形开发界面的封装,使得开发人员可自定义配置参数如sourceFilePath,dstFilePath,FileInputFormat,FileOutputFormat等,实现自己的Map及Reduce类,从而实现程序的分布式计算功能。

[0049] 在本发明的一个实施例中,在当用户使用开发的集成框架进行编程时,会调用后台集成的数据计算中间件,用于将用户的代码嵌套在中间件中,然后自编译运行,调用平台提供的计算框架执行数据计算,用户在开发时可能会调用开发的计算插件,此时插件也会调用对应基于MapReduce框架实现的算法和底层的计算框架进行相关计算,而且开发过程可能是涉及了MapReduce框架的循环调用的,故为提供循环计算性能,采用了Hadoop++相关的优化策略。在集成计算过程中,开发者需要对如下表3计算参数根据需求进行适当配置。

[0050] 表3

[0051]

字段名称	字段属性	备注
InputFormat	输入数据格式	分片分记录方式

[0052]

OutputFormat	输出数据格式	与InputFormat对应
DataType	计算的数据类型	Key-Value类型

[0053] 在本发明的一个实施例中,MapReduce计算模型因其并行化的工作方式,在数据挖掘和机器学习算法领域有广泛的应用。传统的AFOPT算法、Aprior算法、协同过滤算法、局部加权线性回归算法、素贝叶斯分类算法等数据挖掘方面有关算法,都可以通过改进并行化策略,使其适用于MapReduce计算框架,并在开发平台上将实现的计算接口提供给用户开发使用。

[0054] 根据本发明实施例的方法,通过位置网位置与服务数据的转移与计算学习,为位置网的进一步构建提供了相关统计数据信息,还对该平台API进行扩展提供了相应的数据操作管理与计算接口,同时提供了程序开发平台满足用户开发需求。

[0055] 图3为根据本发明一个实施例的基于位置网的创新平台数据管理系统的框架图。如图3所示,根据本发明实施例的基于位置网的创新平台数据管理系统包括创建模块100、封装模块200和插件模块300。

[0056] 具体地,创建模块100用于根据位置网的提供服务构建创建平台的框架,并利用HDFS与HBase将位置网数据库的数据信息及用户操作的信息转移存储到创新平台的框架上构建创新平台。

[0057] 图4为根据本发明一个实施例的创建模块的框架图。如图4所示,创建模块100包括申请单元110、转移单元120、分析单元130和统计显示单元140。

[0058] 申请单元110用于搭建Hadoop平台,其中,Hadoop平台包括HDFS、MapReduce和HBase。在利用位置网提供的服务平台上,为用户显示了云端服务器信息,如内核、内存、硬盘的基本信息及其使用情况,使用户通过相关参数进行云服务器的使用申请。申请成功后用户将获取申请的云主机的登陆名及密码,利用该用户名及密码即可到平台上登陆进行创新平台。该创新平台是以相关web技术作为前台支持,以Hadoop框架作为后台存储与计算支持实现的。

[0059] 转移单元120用于整合Hadoop平台提供的HDFS和HBase相关的API。位置网的位置信息数据采用MySQL数据库进行存储。在构建创新平台基础服务中,将源数据库存储的数据进行转移,采用HBase分布式数据库,定义需要的存储结构与字段,实现源数据的转移存储。位置网的其他信息数据,如用户操作日志信息等,是以文件格式进行存储的,转移时采用HDFS作为主要存储支持,其中,网络日志数据是由log文件格式进行存储,则可以采用简单的API将其转移到HDFS中存储,然后根据需求再决定是否将HDFS上存储的日志数据进行解析转存储到HBase数据库中。

[0060] 分析单元130用于将API进行二次开发和封装。在对位置网位置与服务等数据信息进行转移后,根据应用需求,基于Hadoop分布式系统基础架构对转移的信息数据采用不同的数据挖掘算法进行数据分析,解析出位置网的相关数据信息与用户信息,用于位置网的进一步构建。

[0061] 统计显示单元140用于调用封装的API生成数据存储管理界面对用户数据进行存储和管理。将位置数据经过计算后形成的统计结果,例如,webData、opTime、opType等相关数据的统计信息,采用趋势图在开发平台上显示出来方便管理查看。根据其数据输出格式,进行二次处理生成相应的动态趋势变化图或表单在创新平台上显示出来。

[0062] 封装模块200用于利用创新平台提供的HDFS和HBase作为文件存储和数据库存储的支撑,并利用平台配置参数和表的路径参数封装用户访问的API,通过云主机对文件系统和数据库进行操作和管理。

[0063] 图5为根据本发明一个实施例的封装模块的框架图。如图5所示,封装模块200包括搭建单元210、整合单元220、封装单元230和操作管理单元240。

[0064] 搭建单元210用于搭建Hadoop平台,其中,Hadoop平台包括HDFS、MapReduce和HBase。

[0065] 整合单元220用于整合Hadoop平台提供的HDFS和HBase相关的API。

[0066] 封装单元230用于将API进行二次开发和封装。

[0067] 操作管理单元240用于调用封装的API生成数据存储管理界面对用户数据进行存储和管理。

[0068] 在本发明的一个实施例中,首先结合数据管理相关参数如UserHome、para\_Cpu, para\_Mem, para\_Disk等来为用户配置对应的创新平台环境,然后以Hadoop平台自身提供的操作API、相关配置参数(如Conf, FileSystem)为基础,结合数据操作参数如filePath及tableName等实现文件及数据库操作接口。在管理过程中,提供了数据的各项属性设置,用户可通过对para\_Mode, para\_Own, isShared, isVisible等参数对数据及用户自身属性进行

制定和修改。

[0069] 在本发明的一个实施例中,当后台的API函数获取到传递的参数后则执行相关调用操作,与Hadoop平台进行交互,并最终将获取的返回结果回馈到用户端,完成一次数据操作。提供的数据操作主要包括对文件系统HDFS和对HBase数据库的操作与管理。测试结果表明当对Hadoop集群进行有效的监控管理,通过云端对平台进行相关数据操作具有较好的稳定性和实时性。

[0070] 插件模块300用于根据创新平台提供的存储平台对MapReduce编程框架对常用的关系代数算法和数据挖掘算法进行封装,提供云端开发环境,并利用框架对常见分布式算法进行封装形成插件,为用户数据提供计算处理能力。

[0071] 图6为根据本发明一个实施例的插件模块的框架图。如图6所示,插件模块300包括插件单元310、集成单元320和处理单元330。

[0072] 插件单元310用于对关系代数和数据挖掘算法采用MapReduce框架实现并进行封装形成插件。

[0073] 集成单元320用于封装集成基于Hadoop的分布式开发环境。

[0074] 处理单元330用于根据插件和集成平台对用户数据进行处理。

[0075] 在本发明的一个实施例中,提供的相关数据计算插件是基于Hadoop平台实现的,主要集成了一些典型的分布式数据挖掘算法,基于MapReduce框架进行实现并形成插件作为集成开发环境的一部分提供给开发者使用,免去了重复开发开销;集成开发平台的实现是采用对MapReduce框架进行填充的方式进行功能实现和扩展的,MapReduce框架是为了简化并程序的开发复杂度而设计的,因此我们在上面增加了一层图形开发界面的封装,使得开发人员可自定义配置参数如sourceFilePath,dstFilePath,FileInputFormat,FileOutputFormat等,实现自己的Map及Reduce类,从而实现程序的分布式计算功能。

[0076] 在本发明的一个实施例中,在当用户使用开发的集成框架进行编程时,会调用后台集成的数据计算中间件,用于将用户的代码嵌套在中间件中,然后自编译运行,调用平台提供的计算框架执行数据计算,用户在开发时可能会调用开发的计算插件,此时插件也会调用对应基于MapReduce框架实现的算法和底层的计算框架进行相关计算,而且开发过程可能是涉及了MapReduce框架的循环调用的,故为提供循环计算性能,采用了Hadoop++相关的优化策略。

[0077] 根据本发明实施例的系统,通过位置网位置与服务数据的转移与计算学习,为位置网的进一步构建提供了相关统计数据信息,还对该平台API进行扩展提供了相应的数据操作管理与计算接口,同时提供了程序开发平台并对常用的算法进行封装满足用户开发需求。

[0078] 应当理解,本发明的系统实施例中的各个模块和单元的具体操作过程可与方法实施例中的描述相同,此处不再详细描述。

[0079] 尽管上面已经示出和描述了本发明的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本发明的限制,本领域的普通技术人员在不脱离本发明的原理和宗旨的情况下在本发明的范围内可以对上述实施例进行变化、修改、替换和变型。

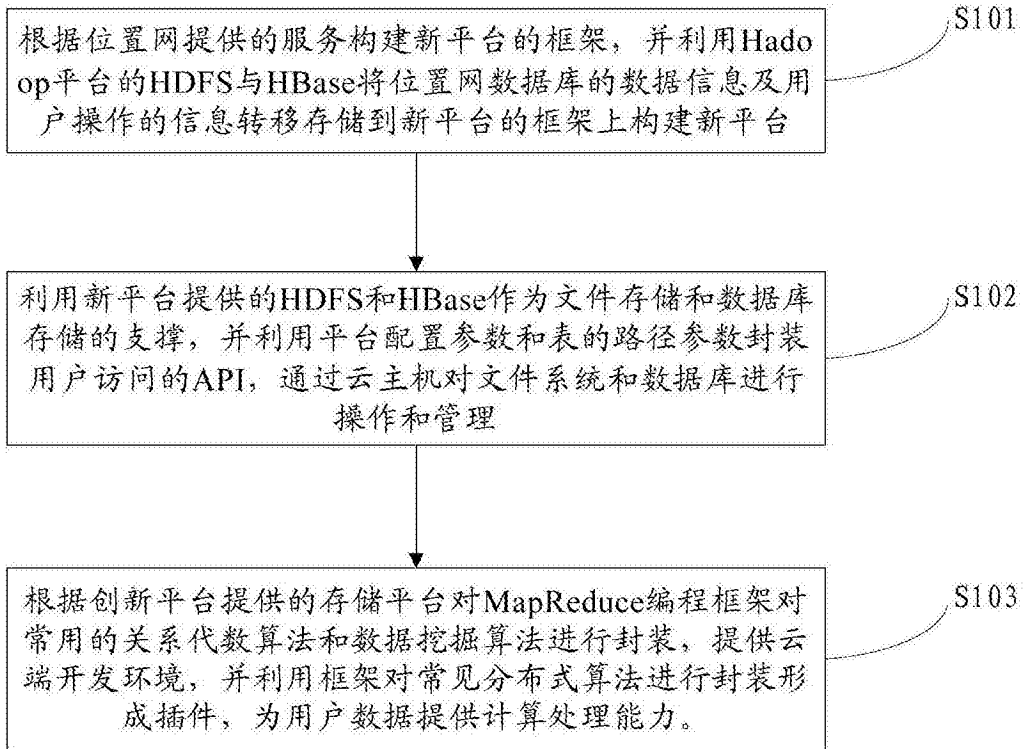


图1

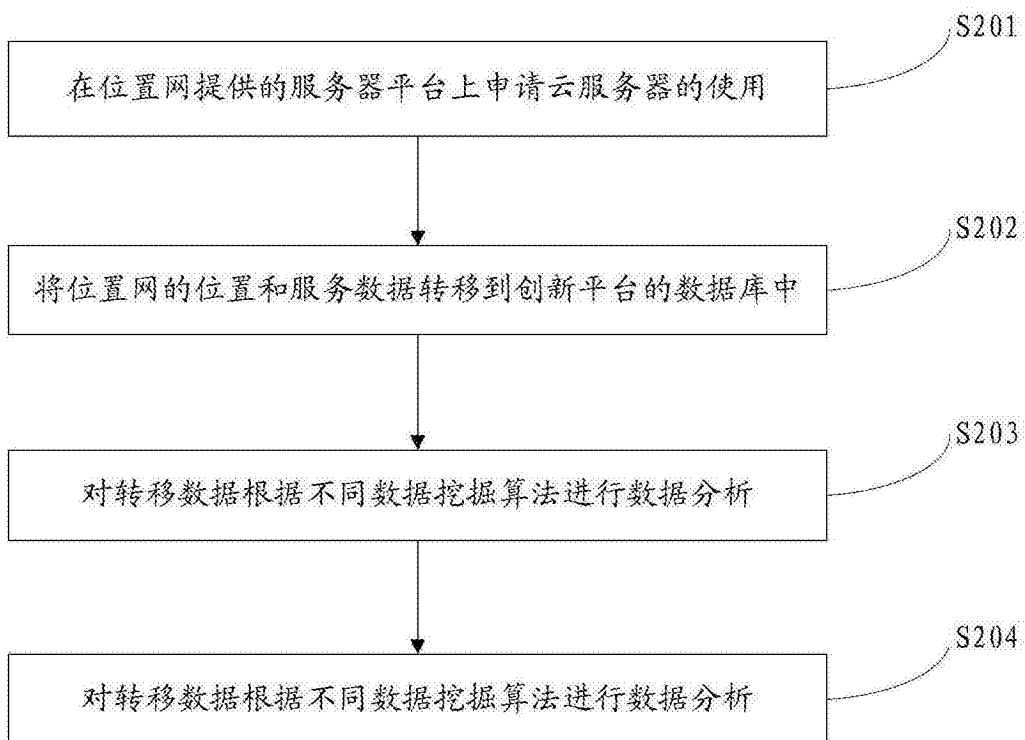


图2



图3



图4



图5



图6