



(12) 发明专利申请

(10) 申请公布号 CN 114185865 A

(43) 申请公布日 2022. 03. 15

(21) 申请号 202111515352.7

(22) 申请日 2021.12.13

(71) 申请人 国汽大有时空科技(安庆)有限公司
地址 246000 安徽省安庆市迎江区老峰镇
胜利路

(72) 发明人 杨宇锡

(74) 专利代理机构 南京苏高专利商标事务所
(普通合伙) 32204

代理人 柏尚春

(51) Int. Cl.

G06F 16/182 (2019.01)

G06F 16/178 (2019.01)

G06F 9/54 (2006.01)

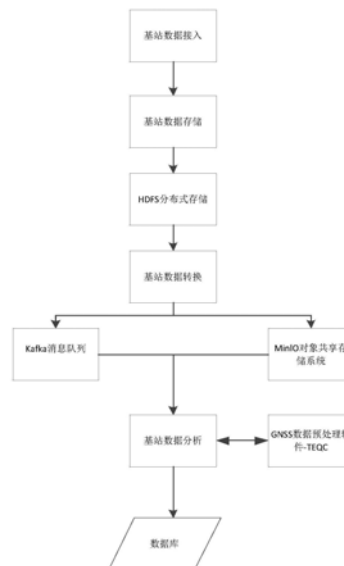
权利要求书1页 说明书3页 附图1页

(54) 发明名称

一种基于分布式存储的大规模基站数据存储与解析的方法及系统

(57) 摘要

本发明公开了一种基于分布式存储的大规模基站数据存储与解析的方法,包括基站数据存储,将流式的基站数据按基站和时间存储在分布式存储系统中;基站数据转换,将分布式存储系统中基站数据小文件合并,利用JAVA本地调用技术调用RTKLIB的C++动态库,将RTCM格式文件转换为RINEX格式文件,将RINEX格式文件上传到HDFS分布式存储系统和MinIO共享存储系统进行存储;基站数据分析,接受消息队列中的消息,解析消息内容,调用SHELL分析脚本,下载RINEX文件,调用GNSS数据预处理软件分析RINEX文件,输出分析结果并将分析结果存入数据库。本发明解决了基站数据存储水平拓展性差、数据转换的效率低的问题。



1. 一种基于分布式存储的大规模基站数据存储与解析的方法,其特征在于,包括以下步骤:

(1) 基站数据存储,将流式的基站数据按基站和时间存储在分布式存储系统中;

(2) 基站数据转换,将分布式存储系统中基站数据小文件合并,利用JAVA本地调用技术调用RTKLIB的C++动态库,将RTCM格式文件转换为RINEX格式文件,将RINEX格式文件上传到HDFS分布式存储系统和MinIO共享存储系统进行存储;

(3) 基站数据分析,接受消息队列中的消息,解析消息内容,调用SHELL分析脚本,下载RINEX文件,调用GNSS数据预处理软件分析RINEX文件,得出基站数据可用率,基站数据周跳比以及基站分频点的多路径数值,并将分析结果存入数据库。

2. 根据权利要求1所述的一种基于分布式存储的大规模基站数据存储与解析的方法,其特征在于,所述流式的基站数据,采用Flink流式计算引擎。

3. 根据权利要求1所述的一种基于分布式存储的大规模基站数据存储与解析的方法,其特征在于,所述分布式存储系统为HDFS分布式存储系统。

4. 根据权利要求1所述的一种基于分布式存储的大规模基站数据存储与解析的方法,其特征在于,所述消息队列,采用Kafka分布式发布订阅消息系统。

5. 一种基于分布式存储的大规模基站数据存储与解析的系统,其特征在于,包括:

(1) 基站数据存储模块:将流式的基站数据按基站和时间存储在分布式存储系统中;

(2) 基站数据转换模块:将分布式存储系统中基站数据小文件合并,利用JAVA本地调用技术调用RTKLIB的C++动态库,将RTCM格式文件转换为RINEX格式文件,将RINEX格式文件上传到HDFS分布式存储系统和MinIO共享存储系统进行存储;

(3) 基站数据分析模块:接受消息队列中的消息,解析消息内容,调用SHELL分析脚本,下载RINEX文件,调用GNSS数据预处理软件分析RINEX文件,得出基站数据可用率,基站数据周跳比以及基站分频点的多路径数值,并将分析结果存入数据库。

6. 根据权利要求5所述的一种基于分布式存储的大规模基站数据存储与解析的系统,其特征在于,所述流式的基站数据,采用Flink流式计算引擎。

7. 根据权利要求5所述的一种基于分布式存储的大规模基站数据存储与解析的系统,其特征在于,所述分布式存储系统为HDFS分布式存储系统。

8. 根据权利要求5所述的一种基于分布式存储的大规模基站数据存储与解析的系统,其特征在于,所述消息队列,采用Kafka分布式发布订阅消息系统。

一种基于分布式存储的大规模基站数据存储与解析的方法及系统

技术领域

[0001] 本发明涉及基站数据处理领域,尤其涉及一种基于分布式存储的大规模基站数据存储与解析的方法及系统。

背景技术

[0002] 信息时代的今天,各行各业的数据规模出现爆炸式增长,大数据是一种新的数据处理方式,具有更强的决策力、洞察发现力和流转优化能力来适应海量、高增长率和多样化的信息资产。大数据的战略意义不仅仅是对海量数据的管理,更多的在于对海量有意义数据的专业化处理,发掘数据价值,数据驱动企业成长。目前市场上主流的大数据技术都是基于Hadoop生态圈的一系列有关采集,存储,计算,查询,资源管理等功能的技术体系。下面是几种比较重要的技术框架:

[0003] HDFS(Hadoop Distributed File System,Hadoop分布式文件系统)是Hadoop体系中数据存储管理的基础。它是一个高度容错的系统,能检测和应对硬件故障,用于在低成本的通用硬件上运行。YARN是Hadoop最新的资源管理系统。负责多个应用程序的多个作业可以同时运行。

[0004] Spark和Flink是两种通用的分布式数据计算引擎,都能进行数据流式处理和批处理,区别是Spark更注重批处理,同时也将流处理看作一种特殊的批处理,Flink的理念是流批一体的概念,能做到真正的流式处理。

[0005] 大数据应用到地图中越来越广泛,一般涉及高精度地图和高精度定位。高精度地图,通俗来讲就是精度更高、数据维度更多的电子地图,能够为车辆感知提供超视距的路况、道路中的车道线、交通标志和环境信息,帮助自动驾驶汽车实现车道级规划决策,以地图匹配方式实现高精度局部定位,降低自动驾驶对昂贵传感器的依赖。高精度定位,根据场景以及定位性能的需求不同,自动驾驶行业有多种高精度定位方法。在大多数的车联网应用场景中,通常需要通过多种技术的融合来实现精准定位,包括GNSS(Global Navigation Satellite System)、无线电(例如蜂窝网、局域网等)、RTK(Real-time Kinematic)、惯性测量单元(Inertial Measurement Unit,IMU)、传感器。其中,使用最广泛的定位方法是融合GNSS/RTK&IMU。GNSS/RTK定位精度高,动态测量下能达到厘米级精度,不存在累计误差问题,但其定位更新频率低,并且信号容易被遮挡;IMU更新频率高,但有累计误差问题。将其结合起来实现优势互补,在GNSS/RTK定位间隔中使用IMU进行位移和方向变量的累加,用户接收机在进行GNSS观测的同时,也接收到基准站发出的改正数,并对其定位结果进行改正,从而提高定位精度,实现低延迟、高精度、高频率的实时定位。

[0006] 高精度地图结合高精度定位技术解决了感知和应用层面的难题。感知层面,当自动驾驶达到L3及以上,对精度和稳定性的要求越来越高。在极端场景例如恶劣天气,激光雷达和视觉传感器的识别感知作用会受到影响。高精度地图同时包含实时路况信息和原有3D模型,可以解决感知环节中传感器在雨雪、大雾天气里不适用的问题,在交互决策环节中对

地理数据进行修正,提高准确度,并且大量减少车载传感器的数目,降低整车成本。

[0007] 但是现在大部分对于基站数据的存储都是基于传统的数据库技术或者硬件磁盘。缺点在于,存储水平拓展性差,数据高可用和安全性不高;对于基站原始数据的转换现在是基于C++的基础动态库,未作任何封装和分布式管理,执行效率和可靠性保证低;对于基站数据完好性监测中的数据质量监控目前市面上有很多软件来做分析,但是都是属于手动执行的应用程序,自动化程度不高,不适用于大型高精度定位平台。

发明内容

[0008] 发明目的:为了解决基站数据存储水平拓展性差、数据转换的效率低的问题,本发明提供一种基于分布式存储的大规模基站数据存储与解析的方法及系统,利用大数据技术处理基站数据的存储,转换,分析。

[0009] 技术方案:一种基于分布式存储的大规模基站数据存储与解析的方法,包括以下步骤:

[0010] (1) 基站数据存储,将流式的基站数据按基站和时间存储在分布式存储系统中;

[0011] (2) 基站数据转换,将分布式存储系统中基站数据小文件合并,利用JAVA本地调用技术调用RTKLIB的C++动态库,将RTCM格式文件转换为RINEX格式文件,将RINEX格式文件上传到HDFS分布式存储系统和MinIO共享存储系统进行存储;

[0012] (3) 基站数据分析,接受消息队列中的消息,解析消息内容,调用SHELL分析脚本,下载RINEX文件,调用GNSS数据预处理软件分析RINEX文件,得出基站数据可用率,基站数据周跳比以及基站分频点的多路径数值,并将分析结果存入数据库。

[0013] 所述流式的基站数据,采用Flink流式计算引擎。

[0014] 所述分布式存储系统为HDFS分布式存储系统。

[0015] 所述消息队列,采用Kafka分布式发布订阅消息系统。

[0016] 一种基于分布式存储的大规模基站数据存储与解析的系统,包括:

[0017] (1) 基站数据存储模块:将流式的基站数据按基站和时间存储在分布式存储系统中;

[0018] (2) 基站数据转换模块:将分布式存储系统中基站数据小文件合并,利用JAVA本地调用技术调用RTKLIB的C++动态库,将RTCM格式文件转换为RINEX格式文件,将RINEX格式文件上传到HDFS分布式存储系统和MinIO共享存储系统进行存储;

[0019] (3) 基站数据分析模块:接受消息队列中的消息,解析消息内容,调用SHELL分析脚本,下载RINEX文件,调用GNSS数据预处理软件分析RINEX文件,得出基站数据可用率,基站数据周跳比以及基站分频点的多路径数值,并将分析结果存入数据库。

[0020] 所述流式的基站数据,采用Flink流式计算引擎。

[0021] 所述分布式存储系统为HDFS分布式存储系统。

[0022] 所述消息队列,采用Kafka分布式发布订阅消息系统。

[0023] 有益效果:与现有技术相比,本发明具有如下显著优点:

[0024] 1、对于基站数据存储水平拓展性差的问题:基站数据的存储主要采用分布式存储系统HDFS,该存储系统基于Hadoop生态,支持水平式扩展,具有副本机制,完全能实现数据的高可用性,同时数据的安全性和加密也能得到保证。

[0025] 2、对于基站数据的转换效率问题：首先，利用JAVA的JNI技术来调用底层C++的RTKLIB动态库，在上层JAVA程序中进行进一步封装和分布式管理；同时，利用Kafka消息队列集合分布式JAVA转换程序进行并行调用，实现同时对多个基站的数据进行转换处理，转换效率大大提升。

[0026] 3、对于基站数据的分析自动化程度差的问题：用SHELL脚本对Linux版本的TEQC的执行程序进行了封装，利用Kafka消息队列和分布式JAVA程序来调用分析脚本，完全实现数据分析完全自动化的流程。

附图说明

[0027] 图1为一种基于分布式存储的大规模基站数据存储与解析的方法的流程图。

具体实施方式

[0028] 下面结合附图对本发明的技术方案作进一步说明。

[0029] 实施例1：

[0030] 如图1所示，一种基于分布式存储的大规模基站数据存储与解析的方法，包括以下步骤：

[0031] (1) 基站数据存储，将流式的基站数据按基站和时间存储在分布式存储系统中；

[0032] (2) 基站数据转换，将分布式存储系统中基站数据小文件合并，利用JAVA本地调用技术调用RTKLIB的C++动态库，将RTCM格式文件转换为RINEX格式文件，将RINEX格式文件上传到HDFS分布式存储系统和MinIO共享存储系统进行存储；

[0033] (3) 基站数据分析，接受消息队列中的消息，解析消息内容，调用SHELL分析脚本，下载RINEX文件，调用GNSS数据预处理软件分析RINEX文件，得出基站数据可用率，基站数据周跳比以及基站分频点的多路径数值，并将分析结果存入数据库。

[0034] 所述流式的基站数据，采用Flink流式计算引擎。

[0035] 所述分布式存储系统为HDFS分布式存储系统。

[0036] 所述消息队列，采用Kafka分布式发布订阅消息系统。

[0037] 实施例2：

[0038] 一种基于分布式存储的大规模基站数据存储与解析的系统，包括：

[0039] (1) 基站数据存储模块：将流式的基站数据按基站和时间存储在分布式存储系统中；

[0040] (2) 基站数据转换模块：将分布式存储系统中基站数据小文件合并，利用JAVA本地调用技术调用RTKLIB的C++动态库，将RTCM格式文件转换为RINEX格式文件，将RINEX格式文件上传到HDFS分布式存储系统和MinIO共享存储系统进行存储；

[0041] (3) 基站数据分析模块：接受消息队列中的消息，解析消息内容，调用SHELL分析脚本，下载RINEX文件，调用GNSS数据预处理软件分析RINEX文件，得出基站数据可用率，基站数据周跳比以及基站分频点的多路径数值，并将分析结果存入数据库。

[0042] 所述流式的基站数据，采用Flink流式计算引擎。

[0043] 所述分布式存储系统为HDFS分布式存储系统。

[0044] 所述消息队列，采用Kafka分布式发布订阅消息系统。

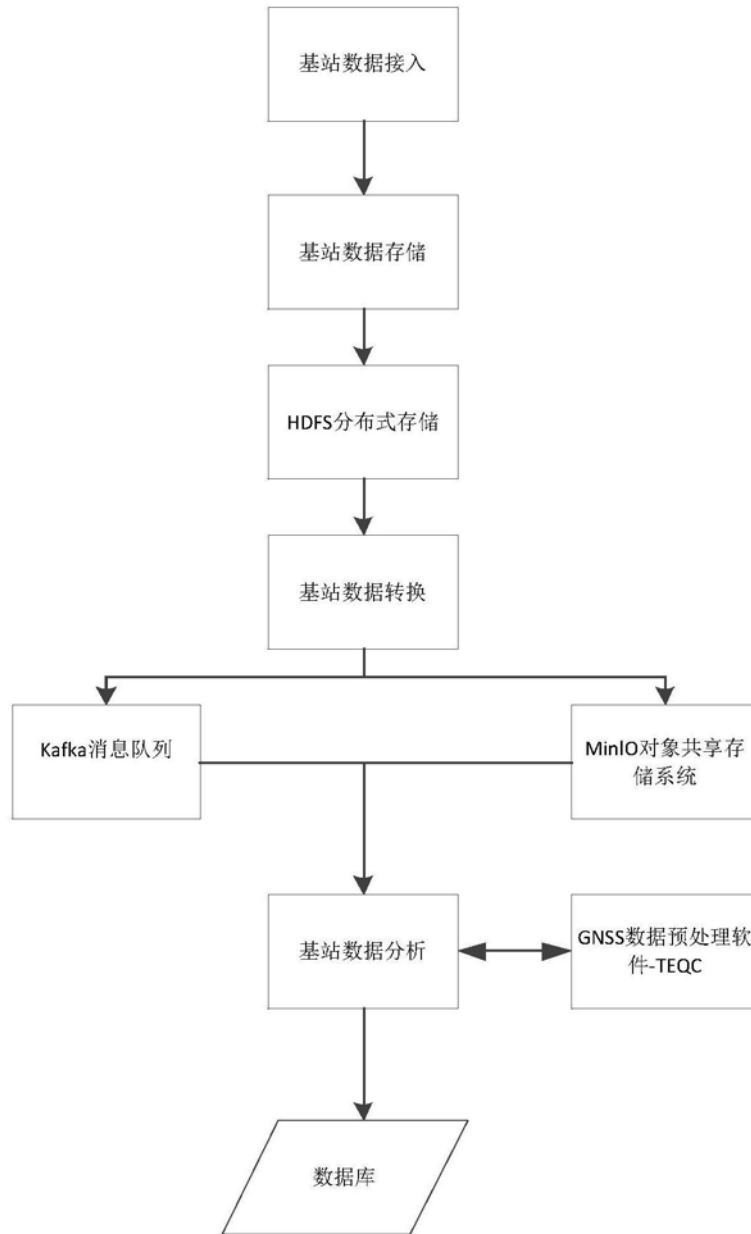


图1