

黄 志,林健玲,李 涛,等.基于“云+端”长序列结构化数据处理应用设计与实现[J].气象研究与应用,2024,45(2):29-36.  
HUANG Zhi, LIN Jianling, LI Tao, et al. Design and implementation of long sequential structured data processing application based on “cloud-edge”[J]. Journal of Meteorological Research and Application, 2024, 45(2): 29-36.

## 基于“云+端”长序列结构化数据处理应用设计与实现

黄 志, 林健玲\*, 李 涛, 黎伟川

(广西壮族自治区气象信息中心, 南宁 530022)

**摘要:**为提升长序列地面观测数据统计和处理支撑能力,加快关键业务平台“云+端”云化转型,基于“天擎”资源数据环境和云化融入标准,通过直连库方式搭建加工流水线定时任务处理流程,采用SpringBoot+Mybatis框架研制众创接口提供数据实时查询服务,数据处理流程监视和告警纳入“天镜”管理,实现对原Hadoop框架构建的业务平台的全面云化融入改造。云化应用表明,云化后节省硬件和存储资源,数据实时查询效率提高30%以上,开发模式简捷高效具有良好的拓展性,为相关海量结构化数据统计场景提供技术思路与借鉴。

**关键词:**气象大数据云平台(“天擎”);云+端;GBase;Mybatis;SpringBoot

**中图分类号:** P409 **文献标识码:** A **doi:** 10.19849/j.cnki.CN45-1356/P.2024.2.05

气象大数据云平台(China Meteorological Administration Data As A Service,英文简称CmaDaas,中文简称“天擎”)已于2021年底成为全国气象部门信息化的全新核心平台,“天擎”是一个庞大的、复杂的大型综合性气象业务系统,它是云计算、大数据、分布式计算、人工智能等各种技术的综合体,构建包含地面、高空、辐射、雷达、数值预报、行业数据等14大类的数据库,整合气象部门的各类数据资源和处理流程,具备数据汇聚、加工、服务和共享应用支撑全流程功能,全面提升数据交换和存储服务时效,能够实现数据管理、加工处理、应用服务的高度集约,致力于消除数据烟囱和信息“孤岛”,实现气象业务数据资源的汇聚互联共建共享,并具备全面支撑“云+端”的气象业务的能力,推动气象业务系统与气象大数据云平台的深度融合<sup>[1-3]</sup>,从而推进气象“云+端”业务技术体制改革和转型。

在“天擎”的“云+端”技术架构下,“云”的部分由基础硬件、虚拟化资源、软件算法等组成,提供系统云化所需的常规计算(算力)、存储(NAS)、中间件

服务(redis、rabbitMQ、Kafka等)等资源,以及满足不同业务应用场景需求的虚谷(分布式关系型数据库)、GBase(南大通用分析型列储存数据库)、Cassandra(开源分布式NoSQL数据库)等多种数据库选型,满足软硬件集约化的同时为融入系统提供高效、稳定、安全的专有网络运行环境,而“端”的部分则通过轻量化代码调用标准API接口实现交互<sup>[4]</sup>。“天擎”的DPL(Data Processing Line)加工流水线子系统是一个特色模块<sup>[3]</sup>,它以提升气象数据的加工处理能力、优化数据加工处理流程为主要目标,以容器化的方式部署应用,通过Kubernetes编排引擎实现按实际资源需求进行动态调度管理,并集成Prometheus实现常用定时任务调度、数据感知等处理模式应用场景,实现加工处理任务在流水线上进行算法部署、资源分配、执行策略配置、运行调度等相关功能,并将运行详情信息DI(DetailInformation)、事件信息EI(EventInformation)信息发送至气象综合业务实时监控系统(“天镜”),具备对整个加工流水线的任务进行综合监视和控制的能力<sup>[5]</sup>。“天擎”

收稿日期: 2023-12-15

基金项目: 广西气象科研计划项目(桂气科2021L02)

第一作者: 黄志(1981—),硕士,高级工程师,主要研究方向:气象信息化、大数据、云计算研究与应用。E-mail:616646373@qq.com

\*通讯作者: 林健玲(1976—),硕士,高级工程师,主要研究方向:气象数据治理、气象信息化研究与应用。E-mail:haz4502396@163.com

的“算法”是指为解决一个特定问题、生成一类数据产品或实现一个业务逻辑所定义的有序指令集,包括可执行文件和众创接口两种形式,其中众创接口

意在解决“天擎”通用接口功能不能满足个性化、复杂业务需求的问题。省级气象大数据云平台(“天擎”)系统框架如图1所示。

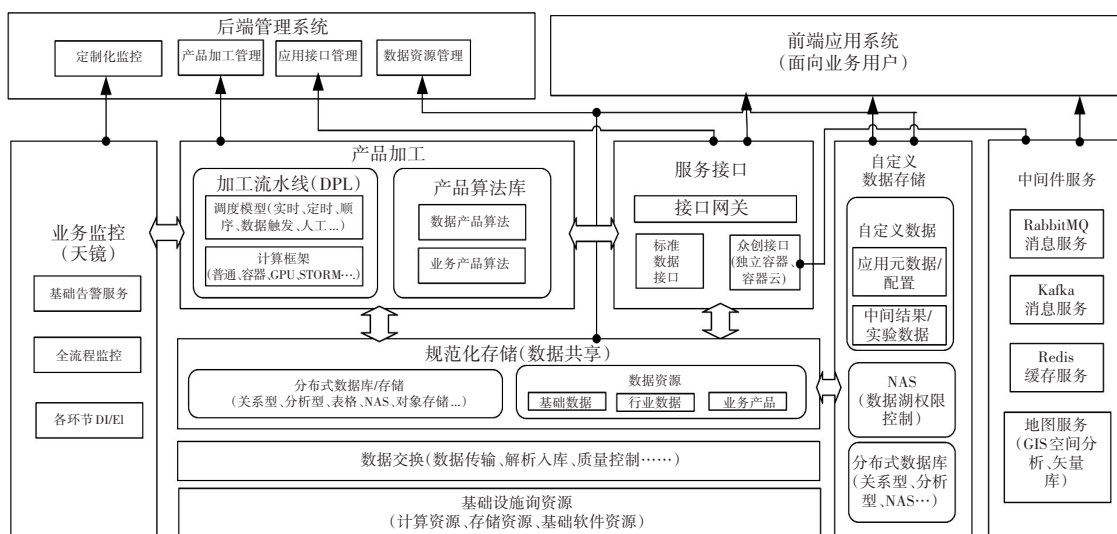


图1 省级气象大数据云平台(“天擎”)系统框架

全国各地都相继开展基于“天擎”的研究与应用,例如山东省基于省级“天擎”,研究系统的架构和功能;湖南省基于“天擎”实现一体化平台云化融入;福建省基于“天擎”实现福建气象综合业务平台云化融入,云化后平台的数据汇聚、加工和服务全流程进一步优化和规范,运行效率和访问速度显著提高;江西省基于“天擎”设计与开发“云+端”技术架构的雷达拼图系统,实现雷达资料的质量控制、网格化处理、数据缓存、组网拼接等实时快速处理<sup>[3-6]</sup>。目前省级“天擎”缺少多样化的长序列地面观测结构化数据统计接口和产品数据集,此类数据主要由本省基于Hadoop+Spark框架搭建的数字档案平台(以下简称“平台”)提供<sup>[7-8]</sup>,由于平台的数据处理流程复杂难以快速拓展和升级,对于多样化数据统计需求,数据服务支撑能力明显不足,严重影响本部门气候论证、防灾减灾工作的推进开展,如何满足多样性的数据需求并提升数据处理时效是目前亟待解决的问题。

为丰富本省“天擎”长序列地面观测数据统计接口和数据产品,加快推进重要业务系统的“云+端”融入。本文基于“天擎”资源环境和云化融入标准,完成对平台的数据环境迁移,实现“算法”重构与云化部署,并将DI/EI信息发送至“天镜”实现对

平台处理流程的实时监控,提升本省长序列数据统计服务支撑能力,为相关业务的“天擎”云化融入提供技术借鉴。

## 1 平台云化融入思路

平台运行环境是基于虚拟资源池的8台服务器组成的Hadoop分布式集群,操作系统和应用软件包括Linux、JDK、Spark、Nginx、Tomcat、redis等,数据集经ETL清洗和格式转换后存储于HDFS;数据处理流程基于Spark并行分布式框架构建,包含耗时较长的定时任务处理(类似OLAP)和实时统计查询(类似OLTP)2种方式,其中实时统计查询是平台的主要业务服务模式。平台的WEB端以SpringBoot微服务形式提供数据查询服务。此外,平台对所需的地面小时、日值等数据进行实时、定期归档、清洗并转换为parquet格式的大容量文件存储至HDFS<sup>[7]</sup>,平台运维流程复杂且数据孤岛效应明显。

基于“天擎”云化资源和数据环境,对平台的云化融入设计流程如图2所示,平台的云化融入分为6个部分进行,其中1—5部分为“云”化建设内容,第6部分为“端”应用,具体如下:

1)平台的物理资源和软件架构全部迁移和融入“天擎”云化资源环境。

2) 基于“天擎”GBase 数据库集群具备的大规模并行计算、海量数据压缩、智能索引等优势, 特别适用于解决大容量高价值密度结构化大数据存储、计算和分析的应用场景<sup>[9]</sup>, 支持多表关联扫描+分组统计、分组汇总等复杂查询计算, 将平台 HDFS 数据环境切换至“天擎”GBase 的地面数据环境, 消除数据孤岛保证数据的完整一致性。

3) 依托“天擎”容器化部署理念, 平台基于 Spark 的定时处理流程转换为“天擎”JDBC 直连 GBase 库的处理模式部署于 DPL; 而平台基于 SpingBoot+SparkSql 的实时查询则采用 SpringBoot+Mybatis 微服务构建 RESTful 风格众创接口, 整体实现对平台数据处理流程(“算法”)的“天擎”融入, 是平台云化融入的关键环节。

4) 平台的数据产品通过“天擎”数据注册融入分布式虚谷数据库, 平台的用户管理、元数据配置等数据信息等迁移至“天擎”自定义数据存储。

5) 数据处理、接口查询等日志信息(DI)以及运行告警等信息(EI)则发送至天镜进行展示<sup>[5]</sup>。

6) 平台 Web 端通过对接“天擎”标准接口或众创接口实现查询服务。

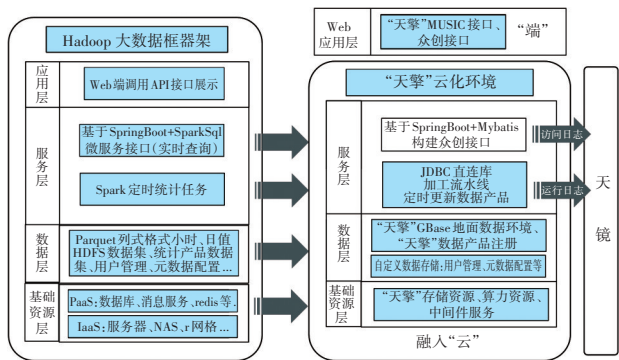


图2 云化融入设计框图

## 2 平台“天擎”云化关键技术研究实现

### 2.1 SQL 设计思路与总结

按照气候论证数据统计规范, 遵循 SQL 设计语法, 对平台的数据统计处理流程重新进行整理分析, 应用聚合、转换、数学、日期、字符串等函数, 通过使用开窗、游标、临时表等方式, 重构 SQL 查询统计逻辑。对平台部分数据统计与处理 SQL 逻辑设计语法总结如下表 1 所示。

表1 SQL 逻辑设计语法

数据统计与处理	SQL 语法设计
建站至今逐年四季的最高、最低、累计降水等前 10 名以及出现时间(气象的春季开始时间为每年 3 月, 夏、秋、冬季延后类推)	统计日期前移 2 个月, 将四季区间转换为 1-12 月, 如春季“extract (month from ADD_MONTHS(D_DATETIME, -2)) >=1 and extract (month from ADD_MONTHS(D_DATETIME, -2)) <=3, 结合 ROW_NUMBER() over(partition by 站号, year order by 统计要素 desc)”进行排名排序
建站至今历年/累年 1-3-6-12-24 小时累计降水极值前 10 名及出现时间、日累计降水极值前 10 名及出现时间	用开窗函数构建各 N 小时滚动极值子表, 例如统计过去 3 小时累计降水: “sum(小时降水字段) over(partition by 站号字段 order by d_datetime desc rows between current row and 3 following)”, 通过 join 各 N 小时滚动极值子表, 结合 ROW_NUMBER() 函数进行排名排序, 以站号关联拼接输出结果
最长连续有、无降水日数以及开始和结束日期(气温类似)	日值表中挑选无(有)降水日, 以站号字段进行分组, 按日期(D_DATETIME)进行正序排序, 构建子表 a; 在子表 a 中, 用排序后的日期减去对应的排序数字, 如果结果相同的则表明日期为连续。核心语法: “D_DATETIME - row_number() over (partition by 站号 order by D_DATETIME) nums”, 以 nums 进行分组选出最大值
累年四季及年各风向频、累年各月最多、次多风向频率	使用 case when 对风向进行落区判断并转换为对应的 16 个风向方位
污染系数、气候风险、气候宜居禀赋等	根据行标、地标数据标准构建 SQL 逻辑统计



## 2.2 平台云化数据处理流程设计

上述工作完成SQL的逻辑设计,平台的“天擎”云化融入关键是“算法”的融入,如何将SQL统计逻辑嵌入平台云化处理流程是本文重点实现的工作(图3)。平台的数据处理流程包含定时处理和实时查询2种模式。在平台“天擎”云化过程中采用JDBC的直连库执行原生SQL的方式代替平台定时处理模式,此模式主要针对需要定时更新统计的数据产品;采用SpringBoot+Mybatis构建众创接口的方式代替平台实时查询模式,具体融入流程分为三个步骤:基于JDBC直连库方式数据统计处理流程设计与构建、基于SpringBoot+Mybatis架构的众创接口设计与构建、镜像制作、“算法”DPL部署与众创接口发布。部分数据处理流程云化前后处理方式如下表2所示。

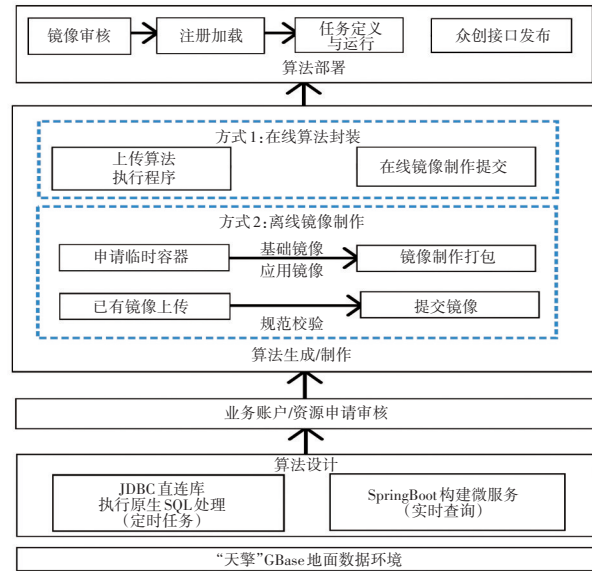


图3 平台数据处理云化流程

表2 部分数据处理流程云化前后处理方式对比

数据处理流程	原处理方式	云化处理方式
建站至今累年平均-最高-最低气温-极大风速风向以及出现时间前10名(数据集)	Spark(定时任务)	JDBC(DPL定时任务)
建站至今逐日最高、最低、平均气温-极大风速以及出现时间前10名(数据集)	Spark(定时任务)	JDBC(DPL定时任务)
建站至今逐月/四季最高气温、最低温度前10名四季、最高气温以及出现时间前10名(数据集)	Spark(定时任务)	JDBC(DPL定时任务)
建站至今历年/累年1-3-6-12-24小时累计降水极值前10名及出现时间、日累计降水极值前10名及出现时间(数据集)	Spark(定时任务)	JDBC(DPL定时任务)
历年暴雨、大暴雨和特大暴雨日数、历年4-9月、5-8月、6-8月、年总降水量、暴雨过程等	SpringBoot+SparkSql (实时查询)	SpringBoot+Mybatis (众创接口实时查询)
能见度、日变化、污染系数、气候风险、气候宜居禀赋等	SpringBoot+SparkSql (实时查询)	SpringBoot+Mybatis (众创接口实时查询)
累年、历年、四季及年各风向频、累年各月最多、次多风向频率	SpringBoot+SparkSql (实时查询)	SpringBoot+Mybatis (众创接口实时查询)
累年四季及年各风向频、累年各月最多、次多风向频率	平台无此接口	SpringBoot+Mybatis (众创接口实时查询)
污染系数、气候风险、气候宜居禀赋等	平台无此接口	SpringBoot+Mybatis (众创接口实时查询)

基于“天擎”的集约化设计理念,申请“天擎”DPL的1个计算节点匹配相应的CPU和内存资源,采用JAVA语言以JDBC直连库执行原生SQL的处理方式,对14种需要定时统计更新的数据产品进行算法程序开发,按统计逻辑和统计数据量封装为2个可独立运行的jar包,通过在线镜像制作方式完成DPL定时任务部署(即2条加工流水线);对涉及降水、温度、气压、湿度、能见度、风、日照、蒸发、春播

期等要素的多个统计值实时查询则通过SpringBoot构建微服务,通过在线镜像制作以独立容器形式部署、注册和发布众创接口。平台云化数据处理流程如图3所示。

## 2.3 基于SpringBoot+Mybatis架构的众创接口设计与实现

DPL+JDBC的设计与实现相对比较简单,本文重点阐述众创接口的设计过程。考虑到实时查询

统计业务是平台的主要业务功能模块,如何将SQL统计逻辑嵌入并封装为众创接口是平台云化的技术关键点。采用SpringBoot+Mybatis框架进行众创接口的研发, SpringBoot继承Spring框架原有的优秀特性,简化Spring应用开发过程,在集成大量软件框架基础上很好地解决依赖包的版本冲突,以及引用的不稳定等问题。SpringBoot可以创建独立的Spring应用程序,并且基于其Maven或Gradle插件,创建可执行的JARs和WARs,并内嵌Tomcat或Jetty等Servlet容器,提供定制化的指标、健康检查和外部配置,是目前“天擎”众创接口的主要采用的开发框架<sup>[9]</sup>。

Mybatis 是一个开源的 Java 持久化框架,它用

于简化数据库访问和操作的过程,Mybatis 与其它持久性框架最大的不同是,Mybatis 强调使用 SQL,而其它框架(例如 Hibernate)通常使用自定义查询语言,即 HQL(Hibernate 查询语言)或 EJB QL(Enterprise JavaBeans 查询语言)。Mybatis 的核心思想是将 SQL 语句与 Java 代码解耦,通过将 SQL 语句和参数绑定到预定义的查询语句,它提供一种将数据库操作与 Java 对象之间进行映射的方式,使得开发人员可以使用简单的 XML 配置或注解来定义 SQL 查询、插入、更新和删除等操作,而无需编写冗长的 jdbc 和 SQL 嵌套的原生代码<sup>[10-11]</sup>。SpringBoot+Mybatis 开发设计框图如图 4 所示。

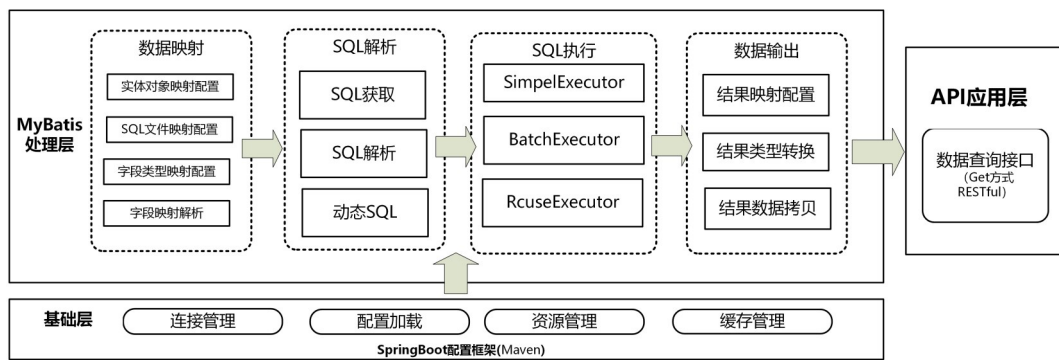


图4 SpringBoot+Mybatis 开发设计框图

文中的SpringBoot的代码设计分为逻辑控制层Controller、业务层Service和数据访问层。通过RestController和RequestMapping对不同统计接口按业务逻辑进行划分,并在一个微服务中封装成多个RESTful风格接口。目前,广西“天擎”容器云提供可承载独立容器的运行环境,“天擎”的众创接口支持在线镜像制作和离线镜像上传两种部署方式,本文采用的是在线镜像制作的方式,即直接在线上传ZIP格式压缩包完成部署<sup>[12-14]</sup>。图5为SpringBoot处理流程图。

#### 2.4 基于“天镜”的运行监控设计

根据“天擎”对云化系统信息监控要求,将平台云化后的众创接口的访问日志、DPL数据处理日志等信息纳入“天镜”进行监控。“天镜”系统以ElasticSearch数据库(分布式、高扩展、高实时的搜索与数据分析引擎)提供URL接口地址供DI和EI信息写入,按照“天镜”的EI/DI标准规范,发送的DI/EI信息采用JSON格式,信息结构主要由信息类型(type)、信息名称(name)、信息内容(message)、信息

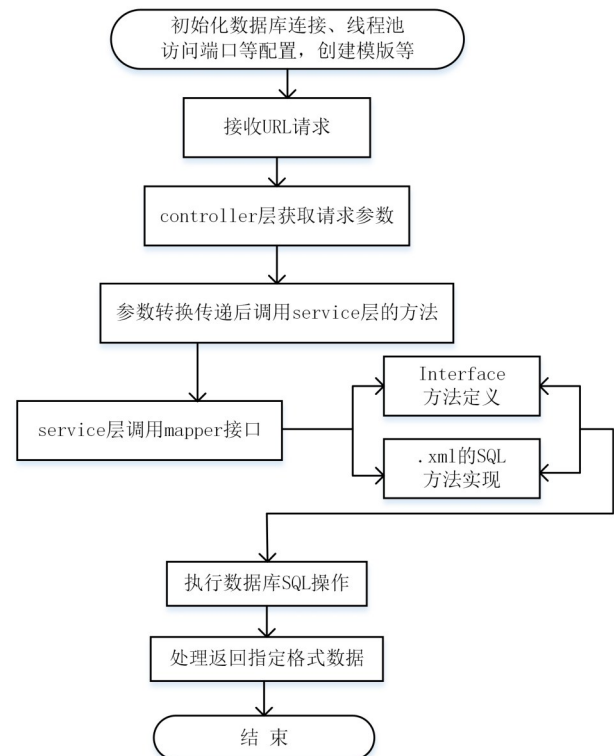


图5 SpringBoot处理流程图

时间(occur\_time)、信息字段(fields)等关键部分组成,以众创接口访问日志为例其信息字段fields主要包括访问用户、调用接口的用户IP、资料编码、访问时间等内容。在众创接口、DPL开发流程中嵌入DI发送模块,众创接口会将每次接口查询日志、DPL在定时数据处理信息以POST方式提交至“天镜”RESTful接口,由“天镜”自动完成DI信息的解析入库,通过配置监控策略,实现基于“天镜”的数据处理流程监视和异常告警服务<sup>[5-6]</sup>。

### 3 融入效果

将平台云化后(简称A模式)与云化前的Hadoop+Spark(简称H模式)两种模式的总体应用情况进行比对。首先在定时任务的运行性能方面,H模式采用Quartz框架进行任务调度,需要整合Spring且不能直接调用业务层Service,任务执行无完善的负载均衡机制,存在单点故障和时间同步问题,偶有任务调度失败情况;A模式定时处理任务基于“天擎”DPL完成,从部署运行至今DPL的管理调度没有出现失败的情况;采用直连库、指定时间区间的迭代式更新统计处理方式也更为高效,具体指标如表3所示。

结合A模式和H模式的集群实际运行现状,A

模式的GBase集群由6个节点组成,其中的3个管理/计算节点负责任务调度和数据处理,另外3个为数据节点(不参与计算),而H模式集群有7个分布式计算节点,算力资源远超A模式,因此本文将H模式的计算节点调试为3个,在A模式和H模式算力基本一致的情况下进行众创接口实时查询处理效率测试。

采用2种方式进行测试,首先针对A模式众创接口的运行性能进行测试,通过设定一定的并发数(如50并发),在一定的站点数和时间区间查询条件下,对部署于“天擎”容器云的众创接口运行参数进行弹性扩容,将虚拟微服务实例从1个扩容至10个、容器内运行的CPU从1核扩容至4核、内存从256M扩容至4G进行测试,测试结果表明,随着A模式的众创接口运行服务实例个数、CPU核数以及内存大小的运行参数变化,其查询统计时效波动变化很小。其次,设置A模式众创接口的运行服务实例(5个)、CPU(2核)和内存(1G)为定参状态与H模式进行并发测试,测试结果表明,在不同的并发条件下A模式的实时数据统计查询效率整体要优于H模式,效率整体提升30%以上。具体对比指标如表4所示。

结合SpringBoot处理流程和上述2种测试方式

表3 A模式与H模式定时任务运行性能比对

对比选项	A模式	H模式
海量数据处理时效	毫秒、秒、分钟	秒、分钟
数据处理方式	智能索引过滤掉无关数据,只读取时间区间、指定站点数据	无索引,每次读取需要进行全表扫描
定时调度	无调度失败记录	偶尔调度异常

表4 A模式与H模式实时数据查询效率比对

统计类型	数据量	统计时间区间/年	A模式 耗时(s)/10并发	A模式 耗时(s)/50并发	H模式 耗时(s)/10并发	H模式 耗时(s)/50并发
单站暴雨过程	18.3万条	20	2.9	6.8	4.5	11.8
单站暴雨过程	36.6万条	40	3.9	8.2	6.3	14.2
单站暴雨过程	64万条	70	4.7	10.7	8.2	17.8
累年各月平均 气温、总降水量	50个站 73万条	40	5.5	12.4	9.2	22.2
累年各月平均 气温、总降水量	100个站 146万条	40	7.3	15.3	11.8	25.8
累年各月平均 气温、总降水量	200个站 292万	40	10.2	21.2	16.7	35.7



的测试结果,可以发现 A 模式的数据处理耗时主要集中在微服务的执行数据库 SQL 操作处理环节,所以众创接口的运行实例数、CPU 核数以及内存大小对查询效率影响甚小,而 GBase 集群的计算节点总数和实时运行负荷则会直接影响众创接口的查询效率,如何提升优化 GBase 集群运行效率、以及优化 SQL 逻

辑结构是后续的重点应用研究方向。

最后在应用快捷方面,两种模式都支持大批量数据导入,A 模式在数据智能索引、处理流程的修善、系统的部署运维等方面都比 H 模式要快捷简单。表 5 为 A 模式与 H 模式各应用指标的综合比情况。

表 5 A 模式与 H 模式综合应用指标比对

对比选项	A 模式	H 模式
索引机制	通过列式存储、智能索引过滤掉无关数据,只读取目标数据,磁盘 IO 和 CPU 资源消耗小	无索引,每次执行即使无命中数据都需要进行全表扫描,磁盘 IO 和 CPU 资源消耗大
数据读写方式	可以支持任意对数据集合的更新和删除。	一般只支持一次写和多次读,不支持对数据进行实时大批量修改
SQL 支持与调优	与传统关系型数据库接口语言保持不变,容易拓展和掌握	SQL 兼容性差,仅是语言实现兼容,其算法调优复杂多样
模糊、精准、复杂查询	能够充分应对各种长序结构化数据查询场景	数据文件块随机分布于 HDFS,无法做到数据的本地运算,Spark 是内存计算引擎,当数据量大内存溢出时性能急剧下降
处理流程部署难易程度	直接修改 SQL 逻辑,SQL 代码可复用,调试快捷、支持快速启停,部署耗时短	Spark 对 SQL 兼容性较差,代码复用性差,算法迭代调优复杂多样,需重启集群,部署耗时较长

## 4 结论与讨论

平台“天擎”云化融入实现对数据、算法的集约化管理和共享。云化融入研发取得主要结论如下:

(1) SpringBoot+Mybatis 微服务接口开发模式是逻辑松耦合的设计思路,Mybatis 框架通过映射文件分离 Java 代码与 SQL 语句,提供简捷的 SQL 逻辑设计封装模式,与 H 模式处理方式相比能简化大量程序代码,提高数据持久层的开发效率。

(2) 对于海量长序列结构化数据统计处理,A 模式的 OLTP 和 OLAP 综合能力要优于 H 模式;A 模式的服务应用基于容器云部署和管理,资源分配和管控实现弹性扩容和自动化运维,运行更稳定故障率更低。

(3) 基于“天擎”数据环境,实现对平台多种数据产品集约化存储管理,并提供“天擎”标准数据接口;消除本地自建数据环境,原部署的 8 台服务器缩减为 1 台 WEB 端应用服务器,为资源池节省存储和物理资源。

(4) 构建 2 条 DPL 加工流水线,实现对 14 种产品数据集的定时统计更新,按气象要素和统计处理逻辑封装 4 个“天擎”众创接口涵盖气候论证统计产品 50 多种,并将数据处理 DI/EI 日志对接“天镜”实

现自动运维监控。气候禀赋、气候风险等新增数据产品统计时效由原来的人工处理耗时 3 d 提升至秒级响应,大幅提升数据服务支撑能力,也为其他海量长序列关系型数据统计处理提供技术思路和方案借鉴。

### 参考文献:

- [1] 赵芳,熊安元,张小纓,等.全国综合气象信息共享平台架构设计技术特征[J].应用气象学报,2017,28(6):750-757.
- [2] 刘洋,黄志,徐娟,等.气象大数据云平台监控告警系统[J].计算机系统应用,2023,32(3):86-94.
- [3] 黄志,黄珩,梁维亮,等.基于“天擎”DPL 的业务融入设计与应用初探[J].气象研究与应用,2022,43(1):73-77.
- [4] 戚云枫,曾小团,梁苑苑,等.广西网格预报系统融入“天擎”的实践与思考[J].气象研究与应用,2022,43(2):111-116.
- [5] 袁正国,李显风,邓卫华,等.基于“云+端”架构的江西省雷达拼图系统设计[J].科学技术与工程,2022,22(18):7773-7779.
- [6] 余永城,王笑,魏夏潞.福建气象综合业务平台融入“天擎”技术方案设计及实现[J].气象科技,2022,50(5):653-659.
- [7] 黄志,苏传程,苏晓红.大数据环境下 Spark 性能优化分析研究与应用[J].气象科技,2022,50(1):51-58.

- [8] 黄志,詹利群,任晓炜,等.Hadoop 环境下基于 Spark 海量自动站数据查询统计的初探[J].气象科技,2019,38(1):768-772,871.
- [9] GBase 8a MPP Cluster技术白皮书[Z].<https://www.gbase.cn/download/gBase-8a?category=DOCUMENT>.
- [10] 欧阳宏基,葛萌,程海波.Mybatis 框架在数据持久层中的应用研究[J].微型电脑应用,2023,39(1):73-75.
- [11] 张旭刚,张昕,高若寒.基于 Spring Boot 与 Mybatis 框架构建动态读写分离模型[J].微型电脑应用,2021,37(2):84-86.
- [12] 纪元,郑卫波,王梓,等.基于容器的安全接入虚拟化[J].计算机与现代化,2022(9):106-110.
- [13] 邵瑛,徐斌.基于微服务的通用性系统日志评估方案及应用[J].计算机系统应用,2022,31(1):124-131.
- [14] 陆伟强.基于微服务的民机工业软件架构设计[J].计算机与现代化,2023(7):73-78.

## Design and implementation of long sequential structured data processing application based on “cloud-edge”

HUANG Zhi, LIN Jianling\*, LI Tao, LI Weichuan

(1.Guangxi Meteorological Information Center, Nanning 530022, China)

**Abstract:** In order to enhance the statistical and processing support capability of long sequential ground observation data, and to accelerate the cloud-based transformation of key business platforms in the “cloud-edge” mode, a timed task processing flow of processing pipeline is built by directly connecting to the library based on the resource data environment and the cloud integration standard of the provincial meteorological big data cloud platform (“CMADaaS”), and the “SpringBoot+Mybatis framework” is adopted to develop a collaborative interface to provide real-time data query service, and the full-process business monitoring and alarm functions for data processing are incorporated into the “TianJing” management, achieving comprehensively integrating the original platform into the cloud environment. The results of cloud application indicate significant savings in hardware and storage resources, with real-time query efficiency improved by over 30%. The development model is simple, efficient, and highly extensible, providing technical guidance and reference for related massive structured data statistical scenarios.

**Key words:** CMADaaS; cloud-edge; GBase; Mybatis; SpringBoot