

时序流数据支持下的实时数据处理技术应用研究

袁小勇¹

YUAN Xiaoyong

摘要

随着时代的发展,在许多信息化工作领域,数据种类、采集数据编目的复杂性逐渐增强,同时整体数据体量也在向着更大的方向发展。在这一背景下,对于训练数据处理的实时性、准确性要求自然也不断提升。传统的数据交互处理架构在面对新的、更加复杂的工作场景时,其传输与处理的性能瓶颈问题开始逐渐暴露出来,对其进行优化具有一定的现实意义。通过对相关文献进行查阅,结合 RocketMQ 消息中间件技术、Flink 流处理技术等,针对当前军事训练场景中多来源、高并发、高频率、大体量的数据实时传输、处理、回放需求,进行了一套功能完善的数据传输与处理系统设计,以期为复杂场景下的实时数据处理工作提供一定的理论与技术支持。

关键词

时序流数据;实时数据;处理技术;军事工程;信息处理

doi: 10.3969/j.issn.1672-9528.2024.08.042

0 引言

当前,我国信息化技术发展迅猛,为各行各业提供了全新的升级方向。在军事工程领域,提升军事活动的信息化水平逐渐成为当前军事变革的主要方向。在军事训练过程中,战场环境复杂性高,局势变化速度快,生成的情报数据数量很多。对情报数据、观测数据等数据进行有效处理,能够为指挥官带来巨大的信息优势,进而做出更加精准的临场决策,最终提升军事训练的综合质量^[1]。但是,在当前技术背景下,我国现有的军事训练信息系统仍旧采取着较为落后的信息传输交互手段,在实践中存在传输效率低、交互耦合性强、时序性效果差、数据传输与处理相对独立等缺陷,存在进一步改造与优化的空间与可能性。

1 研究背景与系统需求分析

1.1 研究背景

在当前的军事训练信息化平台中,普遍采用 RestFul 接口等传统方式进行信息交互,在实际工作情境中,存在数据交互耦合性差、实时性差等缺陷。在数据处理方面,各业务应用普遍采用 B/S、C/S 等传统处理架构进行数据处理,在实际工作情境中,存在处理实时性差、代码复用性低等缺陷。在态势评估等回放场景下,回放应用由系统独立完成,存在回放能力差等缺陷。综合上述内容可知,当前军事训练信息化平台的信息化水平存在比较明显的缺陷,为了提升军事训

练的信息化水平,需要在对需求进行正确分析的基础上,进行基于时序流数据的实时数据处理系统设计与实现^[2]。

1.2 系统需求分析

在系统建设之前,需要对系统需求进行全面分析。本文以现代军事训练信息系统为研究对象,从实时消息采集传输、业务数据实时处理、时序数据回放三方面对需求进行分析。

实时消息采集传输。在军事训练活动中,不同业务系统往往存在种类较多、差异性较强的数据交互需求,在信息交互过程中存在多主体数据处理情况。其具体需求内容如下。

(1) 业务数据实时交互传输需求。该需求是系统建设的根本内容。首先,机动中心负责接收各类采集设备、终端设备的回传数据,并进行汇聚、处理,最后传递给所属业务系统服务。其中,机动中心汇聚的数据可以分为弹道测量数据、设备状态监控信息、电磁监测数据等。这种宽泛的数据分类情况很容易增加数据处理难度,因此,在各类业务数据实时交互传输功能构建方面,应当提供更加细致的消息数据,才能满足灵活消息传输消费的业务需求。此外,在实践中,部分数据需要在指定终端之间进行定向传输,因此,在实时消息采集传输模块中,还应当提供数据的定向传输功能^[3]。

(2) 结构化数据实时同步传输需求。本文所研究的对象具有一定的特点,即训练场地面积较大、网络环境较为复杂,在这一背景下,无法保证全区位通信覆盖的通信环境,因此,需要实现结构化数据同步传输功能需求。具体而言,第一,移动数据中心通常承担着较高频率的数据交互作业需

1. 贵州经贸职业技术学院 贵州都匀 558000

求, 为了提升数据传输稳定性, 结构化数据应当具有高效、轻量的特性。第二, 在项目深化建设过程中, 可能存在数据库表增删、修改等变化, 为了减少后续开发、维护成本, 结构化数据应当具有灵活配置的特性。

业务数据实时处理。在军事训练领域, 涉及的数据具有种类多、输入频率高、处理实时性要求高等特点。在传统训练信息化系统中, 数据预处理主要由各自业务服务后端独立完成, 存在代码复用性差、开发成本高等缺陷。总之, 业务数据实时处理的具体需求内容如下。

(1) 业务数据实时预处理需求。在军事训练过程中, 业务数据种类相对较多, 可能由于业务应用负载过高、网络波动等原因产生各种异常数据, 包括但不限于空值、重复值等, 因此, 业务数据实时预处理应当首先能够完成各种异常数据的初步过滤工作。其次, 为了保证业务系统高效运行, 预处理模块应当满足高效、实时的需求。

(2) 弹道数据实时融合处理需求。在军事训练过程中, 存在导弹飞行试验, 弹道数据的测量与计算是军事训练的重要内容, 其输入信息包括测量雷达、光学经纬仪等多种设备的测量数据, 因此, 在处理过程中, 需要对多种设备进行融合处理, 这也是业务数据处理的重要需求^[4]。

时序数据回放。态势回放也是军事训练领域的重要工作内容, 其具体概念为: 对训练内容进行分析回顾。与其他工作相比, 态势回放对于时序性的要求相对较高, 其中, 时序性主要指的是各个数据的时间顺序, 需要对其进行厘清。总之, 时序数据回放的具体需求如下。

(1) 时序精确性需求。第一, 在回放过程中, 应当厘清不同事件的先后次序。第二, 在回放过程中, 应当厘清不同事件数据之间的间隔时间。

(2) 通用性。即系统可支撑多业务系统的通用性回放需求。

(3) 灵活性。在回放过程中, 应当满足暂停、恢复、倍速等功能需求^[5]。

2 相关技术分析

2.1 RocketMQ 消息中间件

Apache RocketMQ 是一款优秀的消息中间件产品, 在 Kafka 基础上进行了适当优化, 能够在大量消息应用背景下, 提供海量的数据吞吐能力。

从性质上分析, Apache RocketMQ 属于典型的高性能小的队列, 采用生产者-消费者模式进行功能实现, 其基本逻辑为: 首先, 由生产者向代理服务器中的消息队列中写入数据; 其次, 由消费者从消息队列中读取数据, 最终完成消息发布与订阅。其核心概念主要包括以下六个方面。

(1) Topic。消息主题, 能够帮助工作人员对消息进行标识与分类, 具体可以实现一级类别分类。

(2) Tag。消息标签, 能够在某业务主体下对消息进行进一步分类, 可以实现二级类别分类。

(3) Message。消息, 即消息传递载体, 表现形式为字节数组。

(4) Queue。队列, 能够在 Broker 中实现对消息数据的转存储, 通常分布在多个 Broker 服务器中。

(5) Group。分组, 可以对某一类生产者 (Producer)、消费者 (Consumer) 进行标识, 进而提升并发处理能力。

(6) Offset。消息队列的偏移量, 能够实现对服务端无界消息队列中的定长的消息记录的有效定位。

2.2 Flink 流处理技术

Apache Flink 属于开源的分布式流处理架构, 摒弃了微批处理模式 (Micro-batching), 采用了流式处理模式 (Native Streaming), 将实时数据进行拟态, 看作连续事件流产生的无界数据。该技术具有处理能力强、吞吐效率高等重要优势, 在开源社区中的应用愈发广泛。

Flink 流处理架构具有一定的特殊性。Flink 主要包括作业管理器、资源管理器、任务管理器、分发器四个核心组件, 不同组件在流处理程序中的职责存在一定的差异。

(1) 作业管理器, 主要负责作业控制执行工作。在实践中, 该组件在从分发器发配到应用之后才正式运行, 且与流处理应用对应, 执行资源管理统筹、任务分配等重要管理工作任务。

(2) 资源管理器, 主要负责资源调度工作, 能够完成闲置管理器的有效回收、作业管理器资源高效分配等工作任务。

(3) 任务管理器, 主要负责具体的子任务执行工作。插槽是 Flink 流处理引擎中的基本单位, 任务管理器可以以其为基本单位对并行子任务进行控制与管理。

(4) 分发器, 主要负责作业分发调度工作, 能够提供公用入口, 辅助流处理器作业进行任务转交。

3 基于时序流数据的实时数据处理系统设计与实现

3.1 系统总体设计

在对需求进行分析的背景下, 即可进行系统总体设计, 表 1 为系统分层架构设计思路。在该结构设计策略背景下, 本文采取不同技术进行不同模块的设计与实现, 其中, 采用 RocketMQ 进行实时消息采集传输模块设计与实现, 采用 Flink 进行业务数据实时处理模块设计与实现, 设计时序回放队列进行时序数据回放模块设计与实现。

表 1 系统分层架构设计思路

序号	层级	结构	工作内容
1	基础设施与服务层	包括业务服务、采集设备、观测设备、同步数据库	工作内容为相关数据的收集与存储，并将相关数据传输到下一层级
2	传输与处理层	包括数据传输层、数据处理层	数据传输层的工作内容包括定向消息、结构化数据同步等。处理层的工作内容包括去除重复内容、去除空值、弹道数据的融合
3	回放层	包括回放层	工作内容包括回放队列、历史消息存储
4	训练领域各业务子系统	包括导弹飞行试验子系统、导控子系统、态势子系统、其他子系统	

3.2 实时消息采集传输模块设计与实现

结合 RocketMQ 技术，进行实时消息采集传输模块设计，其具体结构如图 1 所示。基于图 1，实时消息采集传输模块的实现策略如下。

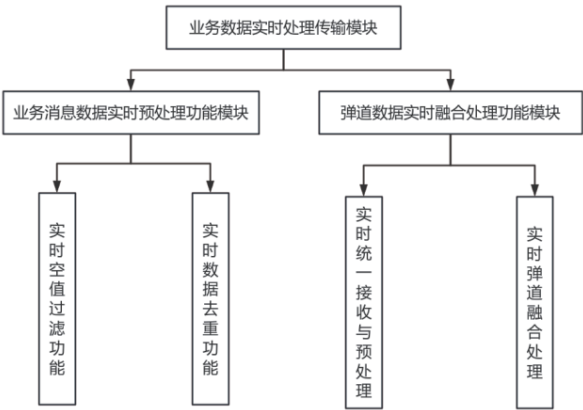


图 1 实时消息采集传输模块的具体结构

业务消息数据实时传输功能实现，具体策略如下。

- (1) 多级主题消息实现。在该方面，军事训练领域存在业务流程复杂、业务维度多等特点，需要对数据进行进一步细分。以常见采集数据为例，可将其分为采集数据 ID、设备 ID、设备所属单位 ID、回传数据类别、数据内容、采集时间。在上述需求场景下，基于 Java 开发语言，对 RocketMQ 模块中的生产者、消费者进行封装整合，定义为 Client 接口，从而为应用开发提供统一生产、消费入口。在功能实现方面，基于开源 Apache RocketMQ 中的内容完成功能实现^[6]。
- (2) 定向消息传输实现。在该方面，可以利用 RocketMQ 的消息 Tag 筛选过滤功能，以终端唯一标识为依据，

- 实现消息的定向传输。
- 结构化数据同步传输功能实现，具体策略如下。
- (1) 在本次研究中，项目采用“一个固定数据中心，多个移动分数据中心”的策略进行部署，为了保证数据同步准确确定，决定结合 Flink 技术进行同步传输核心实现。首先，在数据采集方面，决定使用 Flink-CDC 对数据库中的变更数据进行捕获，实现逻辑包括开启数据库、创建流执行环境、创建 Flink-MySQL-CDC 数据源、自定义数据解析器，经过上述流程，可以将处理后的变更数据发送到 RocketMQ 指定消息主题中；其次，在数据传输阶段，基于 RocketMQ 的消息实时传输模块，负责将上述数据快速投递到负责持久化的 Flink 应用中；最后，在同步持久化阶段，需要基于 Flink 消费 RocketMQ 中的同步数据，并采用流处理解析处理，将持久化写入目标数据库中^[7]。

- (2) 灵活可配置的同步功能实现。在项目建设过程中，关系型数据表可能由于各种意外因素而出现变动情况。针对此情况，可以结合 Flink 流处理技术，将同步数据主流、配置规则广播进行留链接处理，即可实现优化功能。其基本实现流程为：第一，利用 Flink-CDC 对同步规则变化进行监控；第二，利用 Flink 采集同步数据，其具体流程包括实现规则广播流处理方法、连接数据主流与规则广播流、实现主流数据处理方法、将筛选后的同步数据持久化到目标数据库中。

3.3 业务数据实时处理模块设计与实现

结合 Flink 技术，进行业务数据实时处理模块设计，其具体结构如图 2 所示。基于图 2，业务数据实时处理模块的实现策略如下。

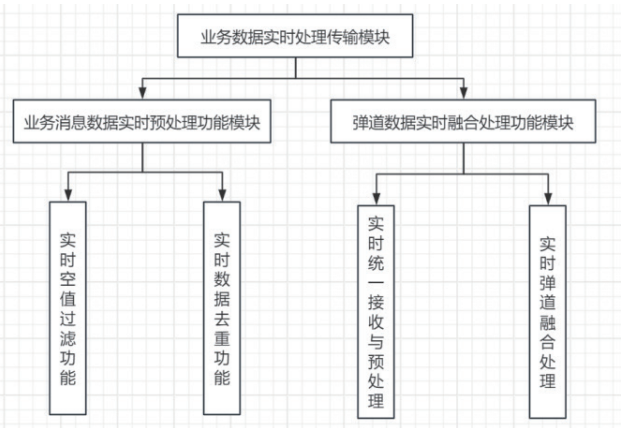


图 2 业务数据实时处理模块的具体结构

业务数据实时预处理功能实现，具体策略如下。

- (1) 时空空值过滤功能实现。基于 Flink 技术，按照获取流执行环境、创建 Flink 数据流、将数据转化为 JSON 对象的逻辑对正确测量数据、异常测量数据进行区分。随后，对

于正确测量数据,应当输出到下游处理算子;对于异常测量数据,应当写入数据库中,并进行记录备份。

(2) 实时数据去重功能实现。对于重复数据,需要依靠测量数据唯一标识 `uuid`,具体去重流程应当为:创建消费者、将数据转化为 JSON 对象、初始化 `bitmap`、根据重复判断策略对数据进行处理分流。如果数据为非重复数据,应当添加 `unique` 标签,输出到主数据流;如果数据为重复数据,则应当添加 `duplicate` 标签,写入数据库中^[8]。

弹道数据实时融合处理功能实现,具体策略如下。

(1) 弹道数据融合处理方法。在实践中,主要测量设备通常包括多个雷达测量站、光学经纬仪测量站。在计算过程中,需要结合坐标初值计算、多站融合计算、弹道平滑计算、弹道预测计算等内容进行计算。

(2) 基于 Flink 的弹道融合处理实现。第一,需要对项目场景进行抽象凝练。第二,测量数据流接收与合流阶段,其实现逻辑为数据源创建、数据流获取、插入水位线逻辑、基于 API 实现雷达与光学测量数据合流。第三,分区与坐标初值计算阶段实现,按照数据流分流、分区、坐标初值计算的逻辑进行。在初值计算过程中,应当利用测站位置信息表中的未读数据进行坐标转换处理,即可获取带有坐标初值的 `MeasureDateExt` 结果。第四,基于 Flink 时间滚动窗口,对同一采样周期的测量数据进行融合计算。第五,采用状态化编程,进行弹道预测实现^[9]。

3.4 时序数据回放模块设计与实现

结合设计时序回放队列,进行时序数据回放模块实现,策略如下。

(1) 历史业务消息数据持久化存储。可以利用 MySQL 关系型数据库对消息进行存储,存储应当分离进行。具体而言,首先,将消息全局唯一标识、消息生产时间等结构性信息拆分为消息记录信息进行存储;其次,将数据大小不定的消息内容数据进行聚合处理,并将其进一步存储在本地文件中。

(2) 分批次历史业务消息数据预加载。当数据量较多时,可以对数据进行分批次加载。一般而言,可以以 60 s 为一个阶段,对完整回放任务进行划分^[10]。

(3) 回放的灵活控制接口实现。可以采用 Spring Boot 实现回放应用的搭建,并对外提供 RestFul 灵活控制接口,以满足轻量化控制需求。核心接口一般应当包括创建回放任务、开始、暂停、恢复、结束、设置倍速等内容,其接口类型均为 Post,以创建回放任务为例,其入口参数为回放任务名称、开始时间、结束时间,返回信息为回放任务编号,含

义为创建回放任务。

4 结论

本文以现代军事训练场景为研究对象,结合 RocketMQ 技术、Flink 流处理技术等内容,进行了分析研究,最终完成了数据传输处理系统的设计与实现,在数据传输、数据处理、数据回放等方面均具有比较明显的技术优势。

参考文献:

- [1] 朱本科,高丙朋,蔡鑫.药品库温度监测中基于分层分簇无线传感器网络实时数据融合策略研究[J/OL].中国测试,1-7[2024-04-13].<http://kns.cnki.net/kcms/detail/51.1714.TB.20240321.1647.007.html>.
- [2] 王冠军,吴静.轨道交通中 5G 通信对实时数据传输的效率提升[C]//第五届电力工程与技术学术交流会议论文集.广州:广东省国科电力科学研究院,2024:2.
- [3] 李佳益,马智亮,陈礼杰,等.基于数据融合的建筑机器人定位研究综述[C]//第九届全国 BIM 学术会议论文集.北京:中国图学会建筑信息模型(BIM)专业委员会,2023:6.
- [4] 李海燕,饶国华,赖芸,等.工程测量技术助力矿山环境保护与可持续发展的研究[J].中国金属通报,2023(9):22-24.
- [5] 王沙沙.实时数据处理轻量化应用程序的设计与实现[J].漯河职业技术学院学报,2023,22(4):28-32.
- [6] 白宇晨.基于民航西北空管局数据平台的实时数据仓库技术选型分析[J].网络安全技术与应用,2023(2):107-109.
- [7] 冯勇,李微,袁雅涵.流式计算技术在山东省非考核地面气象自动站数据实时处理中的应用[J].数字技术与应用,2022,40(9):149-153.
- [8] 何小兵,陈学锋,陈俊,等.多源异构数据实时主动处理技术在城市轨道交通供电智能运维中的应用研究[J].城市轨道交通研究,2022,25(9):16-22.
- [9] 何伟,王浩,蔡忠平.基于 Kafka 的公交 GPS 数据实时通讯系统的设计与应用[J].城市公共交通,2022(6):34-37+41.
- [10] 刘智博.基于时序流数据的实时数据处理技术应用研究[D].北京:中国电子科技集团公司电子科学研究院,2022.

【作者简介】

袁小勇(1988—),男,贵州习水人,本科,讲师,研究方向:计算机应用。

(收稿日期:2024-05-16)