基于 TCP 多连接通信实时并发数据处理技术研究

周钦强,谭鉴荣,伍光胜,李建勇
ZHOU Qin-qiang,TAN Jian-rong,WU Guang-sheng,LI Jian-yong

广东省气象计算机应用开发研究所,广州 510080
Computer Application & Development Institute of Guangdong Meteorology,Guangzhou 510080,China
E-mail:zqq0256@hotmail.com

ZHOU Qin-qiang,TAN Jian-rong,WU Guang-sheng,et al.Research on real-time and concurrent DP system based on TCP multi-connections.Computer Engineering and Applications,2007,43(18):246-248.

Abstract: Automatic weather stations are usually distributed at various regions widely and each of them transfers real-time observational data to DP center synchronously.A high-speed data processing system is designed for this given application in this paper.The designing project and realizing method of software are discussed in detail.In particular,the buffer which acquisition threads put data into and processing thread takes data from concurrently is discussed emphatically.It has indicated that the system implements high speed and accurate results,high stability and reliability.

Key words: multi-connection;real-time;concurrence;Data Processing(DP)

摘 要:针对自动气象观测站分布地区广、资料实时同步触发上传的特点,深入研究了基于 TCP 多连接通信的实时并发数据处理系统,详细介绍了该系统的工作原理和软件设计方案,尤其对决定系统时效性能的并发数据缓存设计给予了深入细致的阐述。该系统能满足对周期性大批量"浪涌"数据进行实时快速处理的要求。实际运行结果表明,该系统在自动气象观测站实时数据处理过程中资料处理及时准确,运行稳定可靠。

关键词:多连接;实时;并发;数据处理
文章编号:1002-8331(2007)18-0246-03 **文献标识码**:A **中图分类号**:TP311.52

1 引言

随着 GPRS 通信技术的成熟以及社会和经济的发展对气象灾害预报提出的更高更迫切的要求,自动气象站 GPRS 通信方式实现了观测数据实时同步高密度的传输。自动气象站在一般天气过程中的数据观测周期为 6 分钟,以保持与雷达探测资料同步对比分析;在如台风、暴雨、冰雹、雷暴等特殊天气过程中,加密观测周期为 1 分钟,所有自动气象站观测资料同步并发传送,因此,采集中心需要接收处理周期性实时大批量的"浪涌"数据,快速准确地处理所有观测资料对于资料数据分析和气象预报具有重要意义。

本文以 J2SE5.0 为核心,利用多线程并发处理技术构建了一个实时性强、界面友好、操作方便的自动气象站实时并发数据处理系统,系统运行稳定可靠,资料处理及时准确。

2 系统方案设计

2.1 系统工作原理

由于各地区自动气象站分布广泛,并且各地区站点数量基本均衡, 所以每个地区所辖站点具有专用的资料传输通道(TCP),即地区与传输连接一一对应,因此数据处理系统接收每个地区观测资料即为本地区所有自动站周期上传的观测数据流,如图 1 所示。

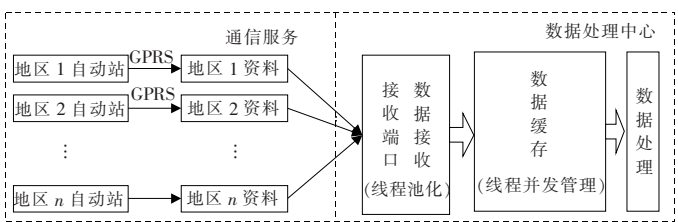

图 1 数据处理系统工作原理图

数据处理系统使用线程池化技术[4]为每一个地区动态分配一条独立的资料接收线程,该线程将接收到的观测数据存储到与其对应的地区数据缓存, 从而避免了线程的频繁创建和销毁,节约系统资源,保证数据处理系统高效接收自动气象站大批量同步上传的资料。同时,系统启动数据处理线程,循环从每个地区数据缓存区读取所接收的自动站观测数据,解析处理完成后,刷新该地区数据缓存。因此,数据接收线程与处理线程可能会同时访问数据缓存并分别对数据缓存进行读写操作,存在访问竞争,因此引入多线程并发管理(加锁)机制。

2.2 系统性能要求

由于各地区自动气象站的资料上传周期一致并且上传时间触发点统一,因此在每一个资料传输时间触发点,所有自动站同步上传观测资料,这就要求系统必须具备接收处理大批量

基金项目:广东省科技计划项目(No.2005B60401024)。
作者简介:周钦强(1979-),男,硕士,主要从事气象装备研发工作;谭鉴荣(1971-),男,高工,主要从事气象装备研发、管理工作。

"浪涌"数据的能力,在下一观测周期数据"浪涌"之前必须将该观测周期接收到的数据处理完成,观测数据才不会排队等待接收处理最终导致系统因负荷过重而崩溃,保证资料处理的及时准确。

3　系统软件设计

系统软件设计基于数据处理时效性和吞吐量要求,利用线程池化技术使各地区观测数据接收效率更高,同时利用J2SE5.0多线程并发特性util.concurrent[5]API实现资料数据缓存设计和数据接收与处理线程的并发管理机制。

3.1　系统软件组成

整个系统软件包括系统初始化、界面模块、数据缓存模块、数据接收模块、数据处理模块和数据统计模块。

(1)系统初始化:系统启动后自动读取系统配置文件进行初始化设置。配置文件包括自动站站点信息子文件:站号、地区代号、网络地址等。

(2)数据接收模块:所有地区数据接收线程的管理(线程池)和每个地区自动站观测数据的接收存储任务。

(3)数据缓存模块:最能体现系统整体性能要求的核心模块,包括所有地区每个自动站观测数据缓存读写任务,着重于数据接收和处理线程并发读写数据缓存的管理机制。

(4)数据处理模块:读取并发缓存数据、自动站数据解码和数据产品编码。

(5)数据处理统计模块:经过系统处理的所有地区自动站来报统计。

(6)界面显示模块:数据处理、参数设置、状态监控、数据统计、系统日志和使用帮助。

3.2　软件开发难点分析

经过以上分析可知,整个系统是以自动站数据的接收、处理为根本,因此及时接收所有地区自动站数据与准确快速解析自动站数据间的优化协作是系统软件设计的重点和难点,下面将围绕这几个方面进行阐述。

(1)自动站数据接收线程

为了及时准确地接收所有地区自动站的同步"浪涌"数据,数据接收线程池为每个地区自动站分配一条数据接收线程执行该地区的数据接收与存储任务,程序流程图如图2所示。

(2)自动站数据处理线程

数据处理线程负责所有地区自动站资料数据的处理任务:遍历每个地区自动站观测数据缓存区,根据通信协议和自动站数据格式对自动站数据解码,根据数据产品格式编码数据产品并存储,程序流程图如图3所示。

(3)数据缓存并发读写机制

由于数据接收线程与处理线程可能会同时访问数据缓存并分别对数据缓存进行读写操作,存在访问竞争,因此,为了使系统及时准确地接收处理观测数据,要求每个地区数据接收线程和系统数据处理线程对数据缓存区的并发读写操作均以最小等待时间工作,也就是说,数据处理线程与不同地区自动站观测数据接收线程协作最优化以完成所有自动站观测资料的接收处理任务。并发缓存数据结构设计如下,如图4所示。

①地区自动站与对应的数据缓存

自动站所属地区与对应的缓存区一一对应,即Region(i)—Buffer(i)。其中,Region(i)为地区(i)标识,Buffer(i)为地区(i)的数据缓存区。

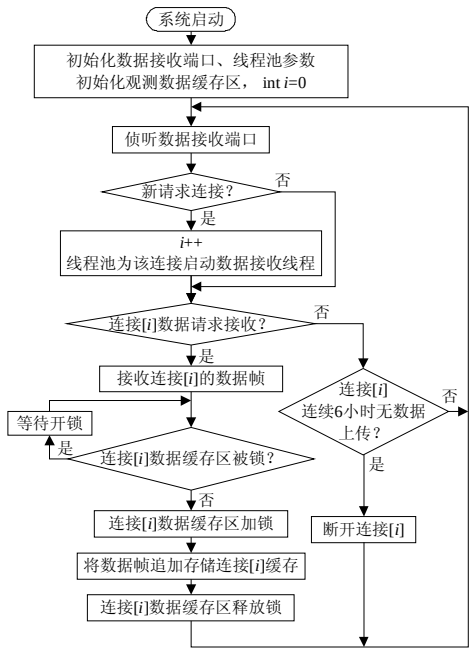

图2　数据接收线程程序流程图

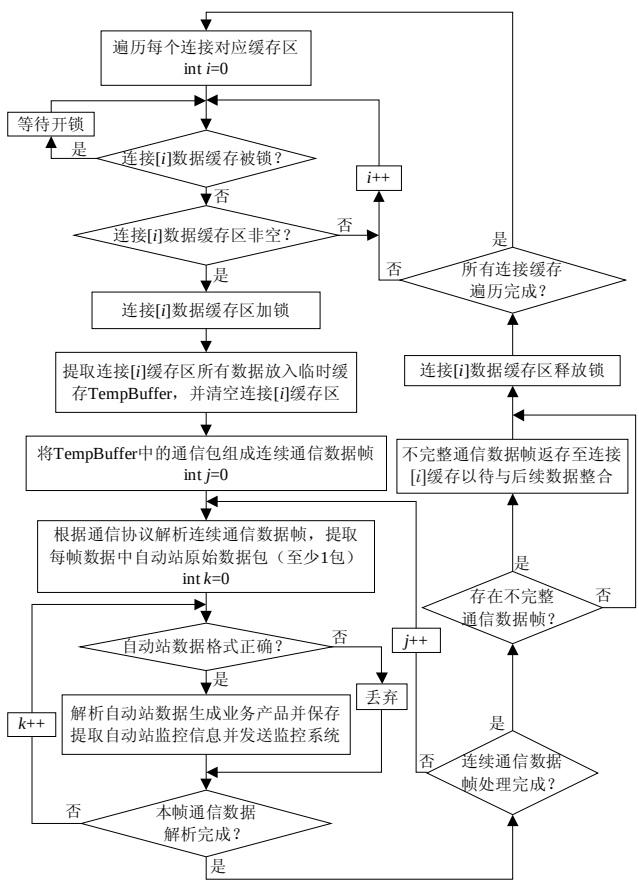

图3　数据处理线程程序流程图

②不同地区自动站观测数据接收线程

由于为每个地区分配独立的数据缓存区,不同地区的数据接收线程可同步进行数据存储操作,即RcvThread(1),RcvThread(2),…,RcvThread(i),…,RcvThread(n)可以同时操作分别对应的数据缓存区Buffer(1),Buffer(2),…,Buffer(i),

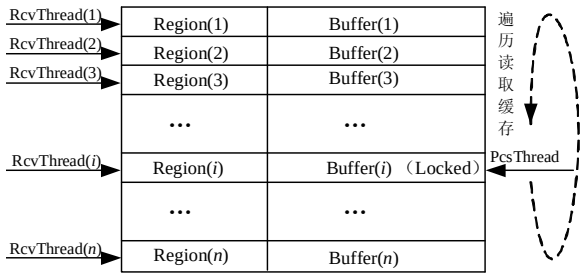

图 4 并发缓存数据结构

…Buffer(n)。因此,当地区(i)数据接收线程 RcvThread(i)存储数据时,其它地区数据接收线程不必排队等候存储,这样保证了所有自动站“浪涌”数据能够及时高效地并行存入对应的缓存区而不会丢失数据。

③数据处理线程与不同地区自动站观测数据接收线程

当数据处理线程 PcsThread 读取地区 Region(i)对应的数据缓存区 Buffer(i)时,该地区的数据接收线程 RcvThread(i)必须等待数据处理线程完成对该缓存区的操作释放 Buffer(i)锁后才能进行存储操作,而其他地区的数据接收线程 RcvThread(1),RcvThread(2),…,RcvThread(n)($n\neq i$)仍可对其分别对应的数据缓存 Buffer(1),Buffer(2),…,Buffer(i),…Buffer(n)($n\neq i$)进行存储操作;反之,当数据接收线程 RcvThread(i)对其数据缓存区 Buffer(i)进行存储操作时,如果数据处理线程 PcsThread 恰好读取该数据缓存区 Buffer(i),则 PcsThread 必须等待 RcvThread(i)完成相应的存储操作释放 Buffer(i)锁后才能读取该缓存区的数据,如图 4 中所示。

(4)并发缓存数据结构 ConcurrentHashMap[6]

J2SE5.0 并发管理机制中提供了符合上述条件的并发缓存数据结构 ConcurrentHashMap。ConcurrentHashMap 是 Doug Lea 的 util.concurrent[5]包的一部分,它提供比 Hashtable(或者替代方案 Collections.synchronizedMap)更高程度的并发性。它为不同的 hash bucket(桶)使用多个写锁和使用 Java 内存管理(JMM)的不确定性来最小化锁被保持的时间——或者根本避免获取锁,使得并发应用程序有着非常好的吞吐量。Hashtable 或者 synchronizedMap 的可伸缩性的主要障碍是它使用了一个 map 范围(map-wide)的锁,为了保证插入、删除或者检索操作的完整性,必须保持这样一个锁,而且有时候甚至还要为了保证迭代遍历操作的完整性保持这样一个锁。这样一来,只要锁被保持,就从根本上阻止了其他线程访问 Map,即使处理器有空闲也不能访问,这样大大地限制了并发性。ConcurrentHashMap 摒弃了单一的 map 范围的锁,取而代之的是由 32 个锁组成的集合,其中每个锁负责保护 hash bucket 的一个子集,具有 32 个独立的锁意味着最多可以有 32 个线程可以同时修改 map。

3.3 软件实现

Eclipse[1] 是一个开放源码的软件开发平台,其图形 API:SWT/JFace 开发的界面美观,响应速度快。系统采用 Java 语言基于 Eclipse3.1 平台开发实现,界面简洁友好,响应速度快,并且充分利用 J2SE5.0 并发管理机制很好地实现了系统设计中所阐述的开发重点和难点问题。系统实现界面如图 5 所示。

图 5 数据处理中心软件实现

由图5 可以看出,在 22 个地区约 600 个自动气象观测站每 6 分钟上传一次观测数据的业务运行环境中,系统运行稳定可靠,数据处理结果及时准确,统计平均来报率达到 99.4%,报文处理及时率达到 100%,很好地实现了各个地区自动气象站大批量周期性“浪涌”观测数据的采集处理任务。

4 结束语

本文针对自动气象观测站分布地区广、资料实时同步上传的业务应用背景,分析并实现了基于 TCP 多连接通信的实时并发数据处理系统,重点分析了决定系统整体时效性能要求的多线程并发管理机制和并发数据缓存设计。它能高效及时地采集处理所有地区自动气象站周期性同步上传的大批量观测数据,运行稳定可靠,数据处理结果准确,满足了后续业务产品的数据需求,为气象数据分析和预报提供了有力的保障。

(收稿日期:2006 年 11 月)

参考文献:

[1] 陈刚.Eclipse 从入门到精通[M].北京:清华大学出版社,2005.

[2] 唐勇,胡振宁,李泽滔.利用 C++Builder 进行多线程实时数据处理的研究[J].贵州工业大学学报:自然科学版,2000,107(6):89-90.

[3] 罗续业.海滨自动观测系统数据处理技术研究[J].国家海洋局海洋技术研究所,1997.

[4] Oaks S,Wong H.Java threads[M].3rd ed.[S.l.]:O'Reilly,2004.

[5] Doug Lea.JAVA 并发编程—设计原则与模式[M].2 版.北京:中国电力出版社,2004.

[6] Goetz B.ConcurrentHashMap and CopyOnWriteArrayList offer thread safety and improved scalability[EB/OL].[2003].http://www-128.ibm.com/developerworks/.