



中国云体系产业创新战略联盟
China Cloud System Pioneer Strategic Alliance

云计算战略联盟技术标准

HB/P-2020-0002

大规模网络信息服务平台的海量实时数据分布式处理 结构及其流程技术标准

The Technology Standards of Distributed Processing
Structure and Workflow of Massive Real-time Data in
Information Service Platform of Large Scale Network

编制单位：同济大学、东华大学

发布时间：2020-09-01





前 言

《大规模网络信息服务平台的海量实时数据分布式处理结构及其流程技术标准》由以下4部分构成：

- 第1部分：范围；
- 第2部分：规范性引用文件；
- 第3部分：海量实时数据分布式处理系统的框架；
- 第4部分：海量实时数据分布式处理的工作流程。

本标准按照GB/T 1.1-2009给出的规则起草。

本标准由同济大学提出。

本标准由信息技术标准化技术委员会（SAC/TC180）归口。

本标准负责起草单位：同济大学

本标准参加起草单位：东华大学

本标准主要起草人：蒋昌俊、章昭辉、丁志军、喻剑、闫春钢、张亚英





引 言

随着云计算技术的普及，如何迅速有效的处理各行各业实时产生的海量数据，并从中提取所需的信息，成了大规模网络信息服务平台的关键问题之一。然而，现有的实时数据采集系统、数据处理系统种类繁多，缺乏统一的数据处理系统结构及其工作流程的统一标准。本规范正是围绕这一内容，提出了一个能动态适应输入数据规模的海量实时数据分布式处理的统一标准规范。

大规模网络信息服务平台动态的海量实时数据分布式处理系统及其流程技术标准

1 范围

本标准规范了大规模网络信息服务平台的海量实时数据分布式处理结构及其流程，对其进行统一的名称规范和定义说明，并为大规模网络信息服务平台其他各项标准的编制提供参照。

本标准适用于所有大规模网络信息服务平台相关组织及其设计、研制、发行、管理、维护的产品、系统等，为信息服务平台的海量实时数据分布式处理提供参照性指标规范。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 1.1-2009 标准化工作导则

ISO/IEC 23006-4-2013 信息技术 多媒体服务平台技术 第4部分:基本服务

ISO/IEC 23006-1-2013 信息技术 多媒体服务平台技术 第1部分:架构

GA/T 739.1-2007 公安请求服务平台应用规范 第1部分:应用服务描述

ISO/IEC 23006-5-2013 信息技术 多媒体服务平台技术 第5部分:服务聚合

GB/T 25470-2010 制造业信息化共性技术资源服务平台功能规范

GA/T 739.2-2007 公安请求服务平台应用规范 第2部分:请求服务应用接口

GA/T 1038.3-2012 消防公共服务平台技术规范 第3部分:信息交换接口

GB/T 30290.1-2013 卫星定位车辆信息服务系统 第1部分:功能描述

GB/T 29746-2013 实时交通信息服务数据结构

AS 3965-1991 信息技术 开放系统和互联 公共管理信息服务定义

GB/T 29841.4-2013 卫星定位个人位置信息服务系统 第4部分:终端通用规范

3 海量实时数据分布式处理系统的框架

3.1 概述

所谓分布式处理结构，传统意义上就是指客户端/服务器端这样的双层体系结构。在这种体系结构中，应用分为两部分：一部分是由多个用户共享所需的功能和信息，即服务器端部分；另一部分是每个用户所需要的功能和信息，即客户端部分。客户端通过专用接口向服务器端发送请求，服务器端收到请求后，激活相应的服务并作相应处理，然后将处理信息返回给客户端。

3.2 海量实时数据分布式处理系统的架构

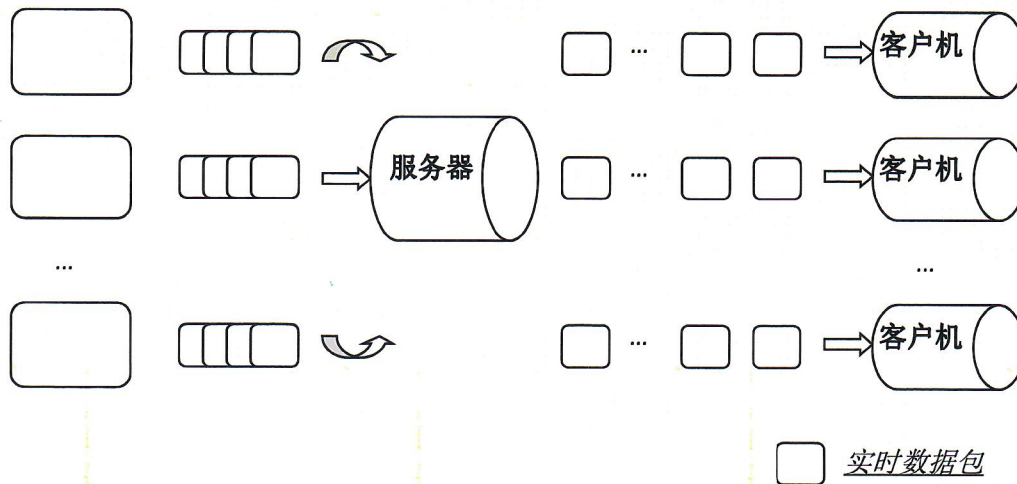


图1 海量实时数据分布式处理系统的架构

海量实时数据分布式处理的架构如图1所示，主要分为两个模块：负责接收多源实时数据并组包分发的服务器模块和负责解包处理实时数据的客户机模块。

服务器模块，包含两个子模块：接收子模块S-I和分发子模块S-II（如图2 服务器模块构成图）。S-I主要负责和数据源建立通信，接收实时数据并组包放入全局包数据共享池。S-II主要负责和客户机端建立通信，从全局包数据共享池中取出包数据并发送，然后等待客户机端发回的反馈信息。从功能上看，服务器端相当于一个整流器，把从数据源接收到的原始的“毛糙”的实时数据流加工成“稳恒”的易处理的包数据流，然后依次分发给各客户机端。所谓“毛糙”是指由于各数据源的原始数据采集速率不同，导致服务器端接收到的各条数据时间间隔参差不齐，而“稳恒”是指原始数据经处理后，被打包成同一时间间隔包数据流。根据各客户机端发回的反馈信息和实时实时数据接收情况，对系统进行动态调整（增减客户机端节点数、改变全局包数据共享池大小以及调整实时数据包发送时间间隔）。

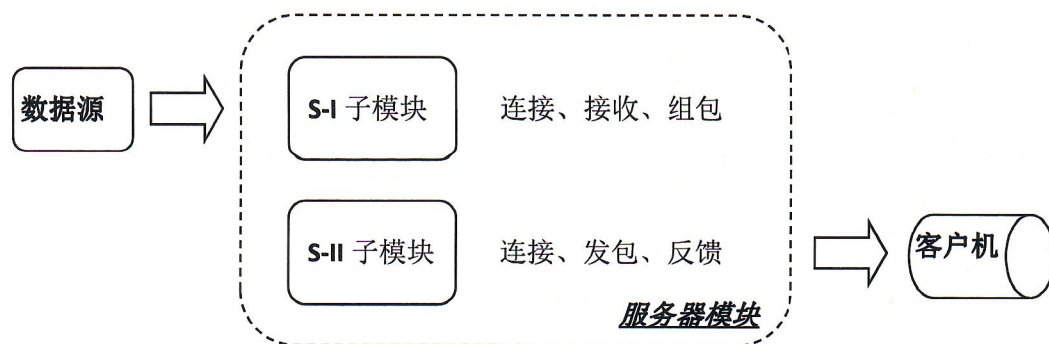


图2 服务器模块构成图

客户机模块：各客户机端连向同一服务器端，接收统一格式的实时数据包，解包然后对数据进行相应处理，提取所需信息，再将处理情况反馈给服务器端，为调整系统运行参数提供决策依据。

4 海量实时数据分布式处理的工作流程



由于在分布式多机环境下处理海量实时数据，各个节点（包括一个服务器端和多个客户端）的运行具有空间上的独立性和时间上的并行性，因此海量实时数据分布式处理的工作流程按节点功能可分为服务器端工作流程和客户端工作流两部分。

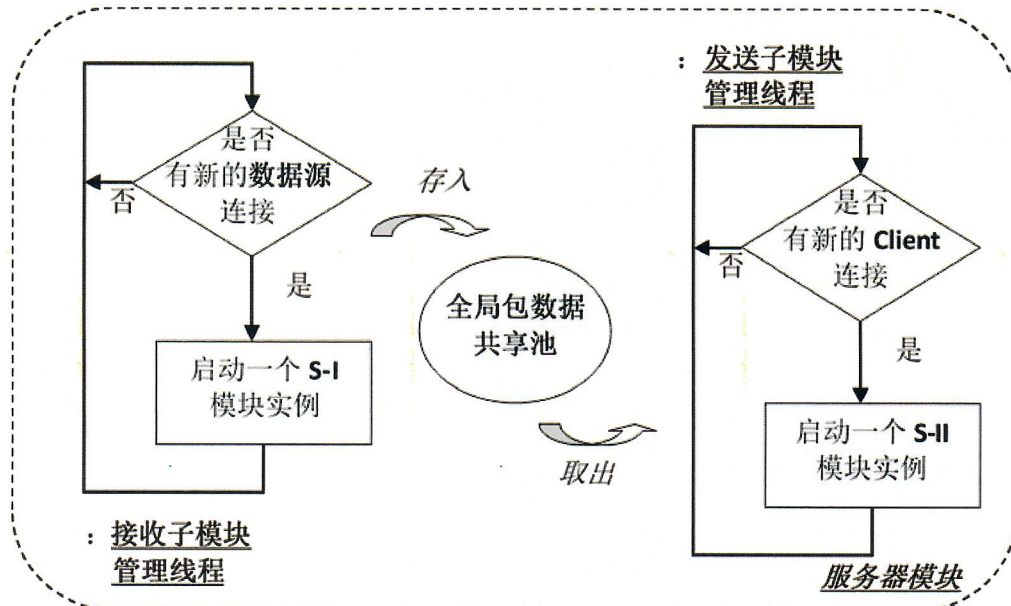


图3 服务器模块工作流程

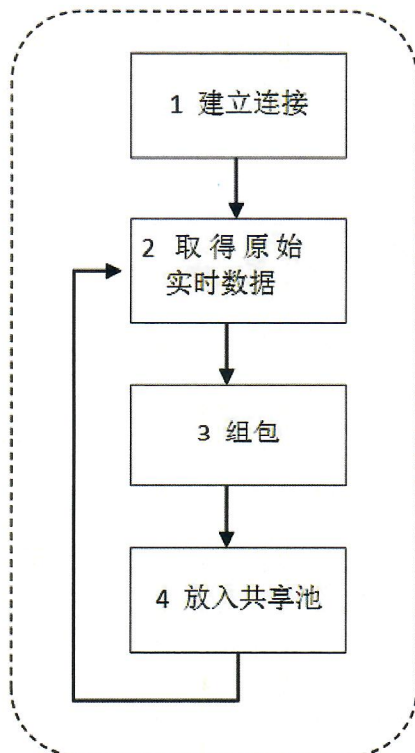


图4 S-I子模块工作流程

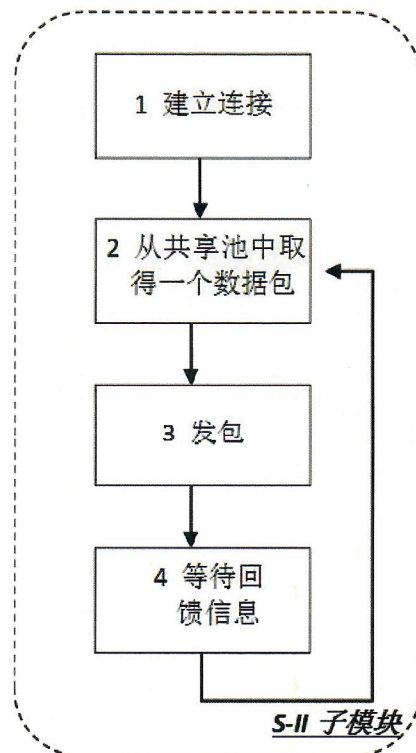


图5 S-II子模块工作流程

服务器端工作流程如图3所示，分为以下几个步骤：



(1) 根据实际数据流量，设置模块配置文件中的各运行参数（客户机端节点数，全局数据包共享池大小以及实时数据包发送时间间隔）；

(2) 根据配置文件初始化模块并启动；

(3) 模块将开启两个管理线程，一个负责为每个新连接的数据源初始化一个实时数据接收子模块实例（参见图4 S-I子模块流程图），另一个则为每个新连接的客户机端初始化一个数据包发送子模块实例（参见图5 S-II子模块流程图）；

(4) 根据各客户机端实时返回的包处理回馈信息和实时数据接收情况，调整系统运行参数，提高系统运行效率。

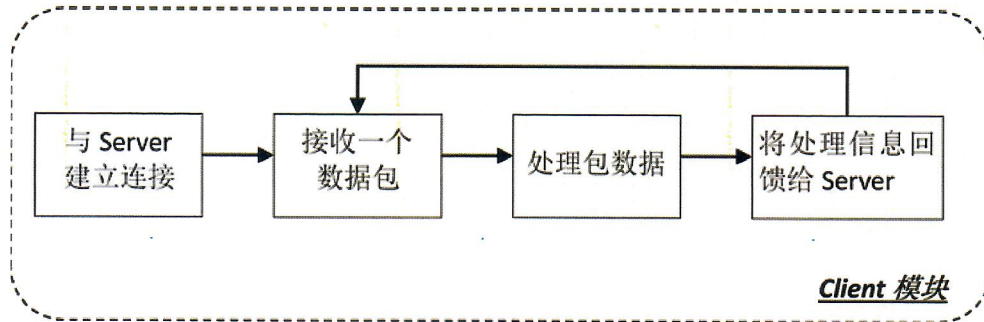


图6 客户机模块技术流程图

客户机端工作流程如图6所示，分为以下几个步骤：

- (1) 根据实际应用环境，设置模块配置文件中的各运行参数（服务器端IP地址和连接端口）；
- (2) 根据模块配置文件初始化模块并启动；
- (3) 运行客户机模块。



参 考 文 献

- GB/T 1.1-2009 标准化工作导则
- JR/T 0096.6-2012 中国金融移动支付 联网联合 第6部分:安全规范
- GB 4943.1-2011 信息技术设备 安全 第1部分:通用要求
- JR/T 0097-2012 中国金融移动支付 可信服务管理技术规范
- ISO/IEC 23006-4-2013 信息技术 多媒体服务平台技术 第4部分:基本服务
- ISO/IEC 23006-1-2013 信息技术 多媒体服务平台技术 第1部分:架构
- GA/T 739.1-2007 公安请求服务平台应用规范 第1部分:应用服务描述
- ISO/IEC 23006-5-2013 信息技术 多媒体服务平台技术 第5部分:服务聚合
- GB/T 25470-2010 制造业信息化共性技术资源服务平台功能规范
- GA/T 739.2-2007 公安请求服务平台应用规范 第2部分:请求服务应用接口
- GA/T 1038.3-2012 消防公共服务平台技术规范 第3部分:信息交换接口
- GB/T 30290.2-2013 卫星定位车辆信息服务系统 第2部分:车载终端与服务中心信息交换协议
- GB/T 30290.1-2013 卫星定位车辆信息服务系统 第1部分:功能描述
- GB/T 29746-2013 实时交通信息服务数据结构
- AS 3965-1991 信息技术 开放系统和互联 公共管理信息服务定义
- GB/T 29841.4-2013 卫星定位个人位置信息服务系统 第4部分:终端通用规范