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

Cloud computing strategic alliance technical standards

HB/P-2020-0002

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

Drafted by: Tongji University, Donghua University

Release date: September 1, 2020

Preface

"Massive real-time data distributed processing structure and process technical standards for large-scale network information service platforms" consists of the following 4 parts:

——Part 1: Scope;

——Part 2: Normative references;

——Part 3: The framework of massive real-time data distributed processing system;

——Part 4: Workflow of distributed processing of massive real-time data.

This standard was drafted in accordance with the rules given in GB/T 1.1-2009.

This standard was proposed by Tongji University.

This standard is under the jurisdiction of the Information Technology Standardization Technical Committee (SAC/TC180).

The organization responsible for drafting this standard: Tongji University

Participated in the drafting of this standard: Donghua University

The main drafters of this standard: Jiang Changjun, Zhang Zhaohui, Ding Zhijun, Yu Jian, Yan Chungang, Zhang Yaying

Introduction

With the popularization of cloud computing technology, how to quickly and effectively process the massive data generated in real time in various industries and extract the required information from it has become one of the key issues for large-scale network information service platforms. However, the existing real-time data acquisition systems and data processing systems are of various types, and they lack a unified data processing system structure and a unified standard for its work flow. This specification is based on this content and proposes a unified standard specification for distributed processing of massive real-time data that can dynamically adapt to the scale of input data.

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

1 Scope

This standard regulates the distributed processing structure and flow of massive real-time data of a large-scale network information service platform, provides a unified name specification and definition description, and provides a reference for the compilation of other standards for a large-scale network information service platform.

This standard applies to all large-scale network information service platform-related organizations and their products and systems designed, developed, issued, managed, and maintained, and provides reference index specifications for the distributed processing of massive real-time data of the information service platform.

2 Normative references

The following documents are indispensable for the application of this document. For dated reference documents, only the dated version applies to this document. For undated references, the latest version (including all amendments) applies to this document.

GB/T 1.1-2009 Standardization Guidelines

ISO/IEC 23006-4-2013 Information Technology Multimedia Service Platform Technology Part 4: Basic Services

ISO/IEC 23006-1-2013 Information Technology Multimedia Service Platform Technology Part 1: Architecture

GA/T 739.1-2007 Application Specification of Public Security Request Service Platform Part 1: Application Service Description

ISO/IEC 23006-5-2013 Information technology Multimedia service platform technology Part 5: Service aggregation

GB/T 25470-2010 The functional specification of common technology resource service platform for manufacturing informatization

GA/T 739.2-2007 Public Security Request Service Platform Application Specification Part 2: Request Service Application Interface

GA/T 1038.3-2012 Technical Specification for Fire Protection Public Service Platform Part 3: Information Exchange Interface

GB/T 30290.1-2013 Satellite Positioning Vehicle Information Service System Part 1: Function Description

GB/T 29746-2013 Real-time traffic information service data structure

AS 3965-1991 Information Technology, Open System and Interconnection, Public Management Information Service Definition

GB/T 29841.4-2013 Satellite Positioning Personal Location Information Service System Part 4: Terminal General Specification

3 The framework of a massive real-time data distributed processing system

3. 1 Overview

Dstributed processing structure, in the traditional sense, refers to a two-layer architecture such as the client side/server side. In this architecture, the application is divided into two parts: one is the function and information required by multiple users to share, that is, the server-side part; the other part is the function and information required by each user, that is, the client-side part . The client side sends a request to the server side through a dedicated interface. After receiving the request, the

server side activates the corresponding service and performs corresponding processing, and then returns the processing information to the client side.

3. 2 Architecture of a massive real-time data distributed processing system

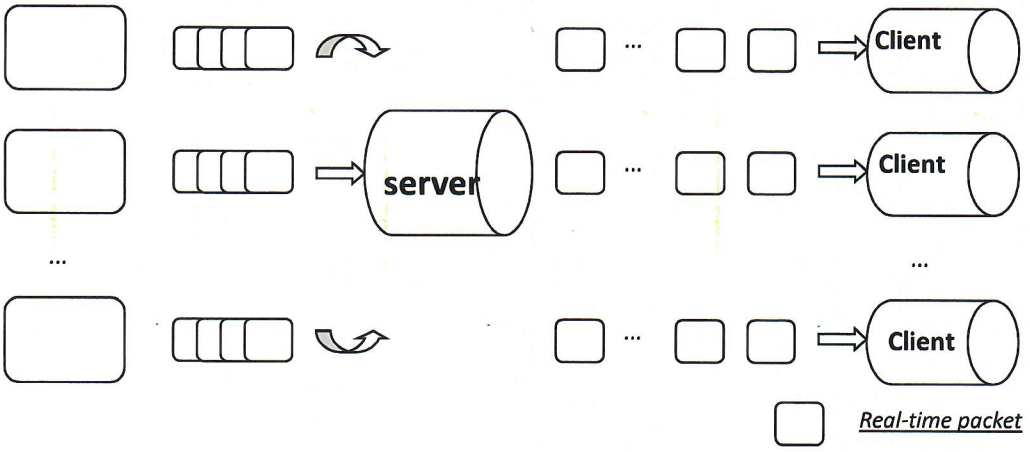

Figure 1 Architecture of a massive real-time data distributed processing system

The architecture of distributed processing of massive real-time data is shown in Figure 1. It is mainly divided into two modules: the server module responsible for receiving multi-source real-time data and grouping and distributing it, and the client module responsible for unpacking and processing real-time data.

The server module contains two sub-modules: receiving sub-module S-I and distributing sub-module S-II (as shown in Figure 2 server module composition diagram). S-I is mainly responsible for establishing communication with data sources, receiving real-time data and grouping packages into the global package data sharing pool. S-II is mainly responsible for establishing communication with the client side, taking out the packet data from the global packet data sharing pool and sending it, and then waiting for the feedback information from the client side. From a functional point of view, the server side is equivalent to a rectifier, which processes the original "rough" real-time data stream received from the data source into a "stable" and easy-to-process packet data stream, and then distributes it to each client side in turn . The so-called "rough" refers to the fact that the original data collection rate of each data source is different, resulting in the uneven

time interval of each piece of data received by the server, while "stable" refers to the original data after being processed and packaged into the same time Interval packet data flow. According to the feedback information sent back by each client and the real-time real-time data reception, the system is dynamically adjusted (increasing or decreasing the number of client-side nodes, changing the size of the global packet data sharing pool, and adjusting the time interval for sending real-time data packets).

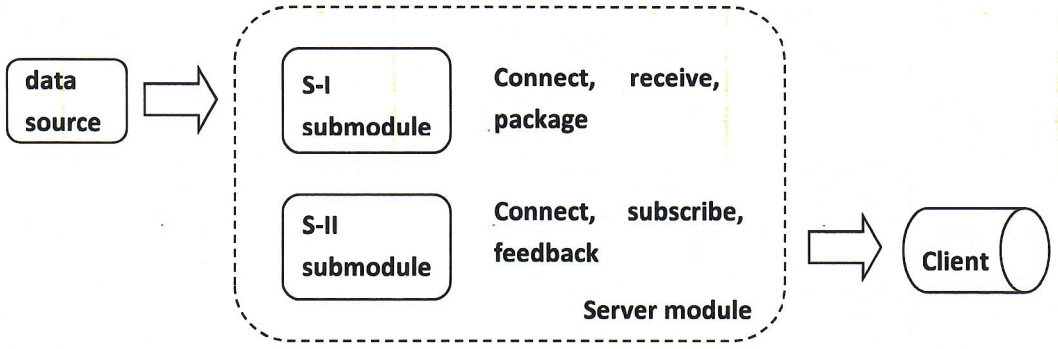

Figure 2 Server module composition diagram

Client module: each client connects to the same server, receives the real-time data packets in a uniform format, unpacks and processes the data accordingly, extracts the required information, and then feeds the processing conditions back to the server to adjust system operating parameters Provide a basis for decision-making.

4 Workflow for distributed processing of massive real-time data

Due to the processing of massive real-time data in a distributed multi-machine environment, the operation of each node (including a server and multiple client terminals) has spatial independence and temporal parallelism. The workflow of the distributed processing of massive real-time data can be divided into two parts: server-side workflow and client-side workflow according to node functions.

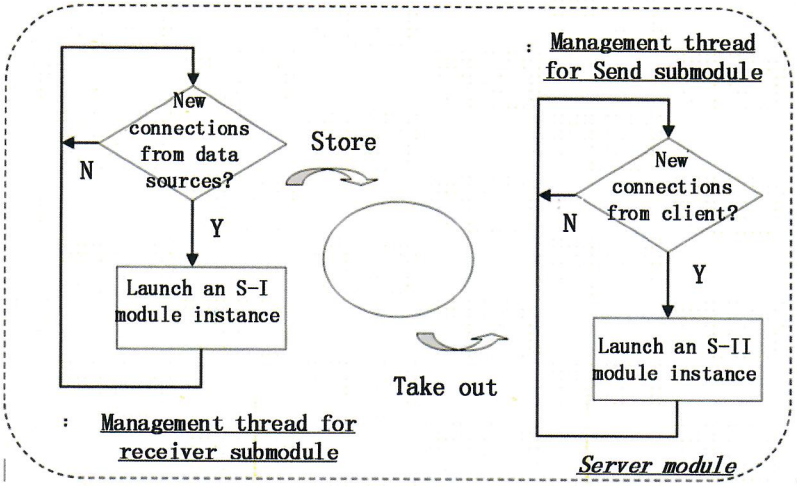

Figure 3: Server module workflow

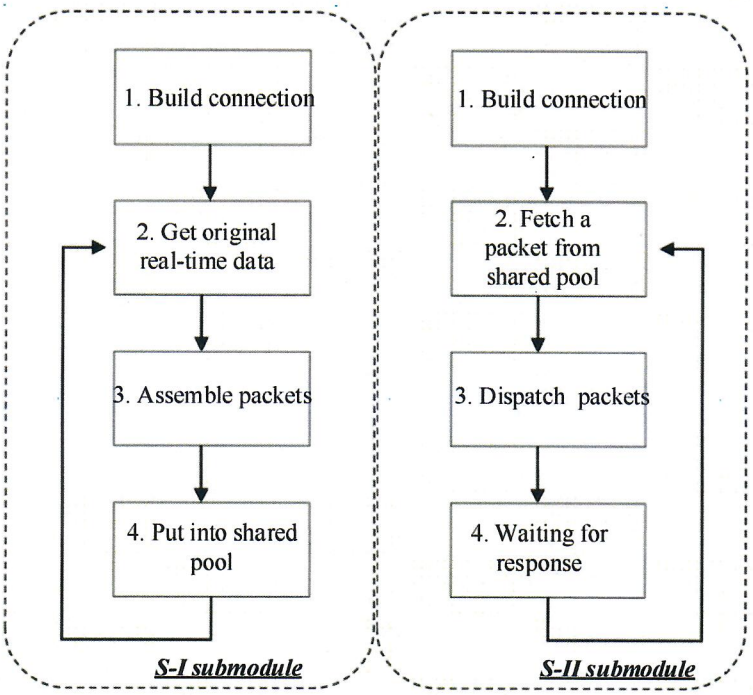

Figure 4 S-I sub-module workflow Figure 5 S-II sub-module workflow

The server-side workflow is shown in Figure 3, which is divided into the following steps:

(1) According to the actual data flow, set the operating parameters in the module configuration file (the number of client nodes, the size of the global packet data sharing pool, and the time interval for sending real-time data packets);

(2) Initialize the module and start it according to the configuration file;

(3) The module will start two management threads, one is responsible for initializing a real-time data receiving submodule instance for each newly connected data source (see Figure 4 SI submodule flowchart), and the other is for initializing each newly connected client side An example of packet data sending sub-module (see Figure 5 S-II sub-module flowchart);

(4) According to the real-time packet processing feedback information and real-time data reception status returned by each client, the system operating parameters are adjusted to improve the system operating efficiency.

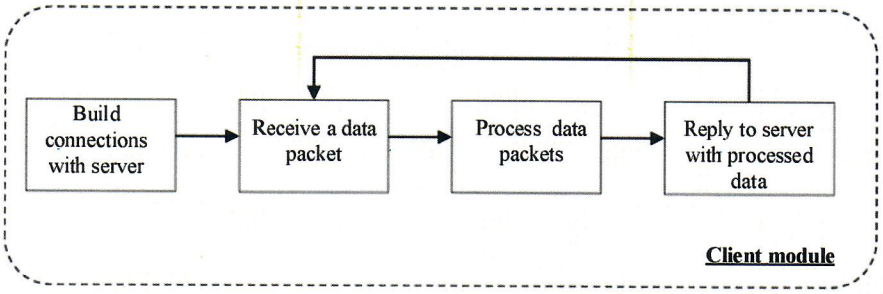

Figure 6: Technical flowchart of client module

The client-side workflow is shown in Figure 6, which is divided into the following steps:

(1) According to the actual application environment, set the operating parameters (server-side IP address and connection port) in the module configuration file;

(2) Initialize and start the module according to the module configuration file;

(3) Run the client module.

References

[1] GB/T 1.1-2009 Standardization Guidelines

[2] JR/T 0096.6-2012 China Financial Mobile Payment Networking Joint Part 6: Security Specifications

[3] GB 4943.1-2011 Information Technology Equipment Safety Part 1: General Requirements

[4] JR/T 0097-2012 China Financial Mobile Payment Trusted Service Management Technical Specification

[5] ISO/IEC 23006-4-2013 Information Technology Multimedia Service Platform Technology Part 4: Basic Services

[6] ISO/IEC 23006-1-2013 Information Technology Multimedia Service Platform Technology Part 1: Architecture

[7] GA/T 739.1-2007 Application Specification of Public Security Request Service Platform Part 1: Application Service Description

[8] ISO/IEC 23006-5-2013 Information technology Multimedia service platform technology Part 5: Service aggregation

[9] GB/T 25470-2010 The functional specification of common technology resource service platform for manufacturing informatization

[10] GA/T 739.2-2007 Public Security Request Service Platform Application Specification Part 2: Request Service Application Interface

[11] GA/T 1038.3-2012 Technical Specification for Fire Protection Public Service Platform Part 3: Information Exchange Interface

[12] GB/T 30290.2-2013 Satellite positioning vehicle information service system Part 2: On-board terminal and service center information exchange protocol

[13] GB/T 30290.1-2013 Satellite Positioning Vehicle Information Service System Part 1: Function Description

[14] GB/T 29746-2013 Real-time traffic information service data structure

[15] AS 3965-1991 Information Technology, Open System and Interconnection, Public Management Information Service Definition

[16] GB/T 29841.4-2013 Satellite Positioning Personal Location Information Service System Part 4: Terminal General Specification