

基于 JSON 的数据交换模型

张沪寅, 屈乾松⁺, 胡瑞芸

(武汉大学 计算机学院, 湖北 武汉 430000)

摘 要: 为解决传统基于 XML 的数据交换平台数据交换效率低以及通用性差的问题, 提出基于 JSON 的“中心+代理”的数据交换模型。采用 JSON 作为总线技术, 数据交换中心作为架构中心, 数据交换代理节点提供代理服务的星型结构模式, 在数据交换代理节点之间建立可靠的连接, 利用数据交换中心对 JSON 信息流进行上传、下载以及实时的监控。模拟实验结果表明, 用 JSON 取代 XML 作为信息流, 提高了数据交换的效率。

关键词: 数据交换; 可扩展标记语言; 信息流; 数据交换中心; 代理节点

中图法分类号: TP399 **文献标识号:** A **文章编号:** 1000-7024 (2015) 12-3380-05

doi: 10.16208/j.issn1000-7024.2015.12.042

Data exchange model based on JSON

ZHANG Hu-yin, QU Qian-song⁺, HU Rui-yun

(School of Computer, Wuhan University, Wuhan 430000, China)

Abstract: To solve the problem of the low data transfer rate and the poor versatility of traditional data-exchange model based on XML, a data exchange model based on JSON and “Center + Agent” was proposed. JSON was used as bussing technique, data exchange center was taken as the architecture center, and data exchange agent node provided the star-structure mode of agent services. A reliable connection among data exchange agent nodes was established, using data exchange center for uploading and downloading and real-time monitoring of JSON information flow. Results of simulation experiments show that using JSON to replace XML greatly improves the efficiency of data exchange.

Key words: data exchange; XML; information flow; data exchange center; data exchange agent node

0 引 言

目前国内外大部分数据交换系统^[1-4]都是基于可扩展标记语言 (extensible markup language, XML) 按照自有的标准来设计, 在数据交换效率、通用性与扩展性上较差。本文提出了一种基于 JavaScript 对象表示法 (JavaScript object notation, JSON) “中心+代理”的数据交换模型, 采用 JSON 作为总线技术, 数据交换中心作为架构中心, 数据交换代理节点提供代理服务的星型结构模式, 在提高数据交换效率的同时, 也具备很高的通用性与扩展性。

1 研究现状

传统的基于 XML 的异构数据交换具有良好的数据存储

格式、可扩展性、自描述性、高度结构化等优点, 成为了目前各系统间和系统内部信息共享的有效手段。文献 [1-4] 都是在 XML 的基础上对数据交换平台的研究, 分别在安全性与实用性做出了改进, 但基于 XML 数据交换也存在着某些不足, 如数据冗余性高、解析速率慢、浏览器不兼容等问题, 使得数据在网络传输中的速度较慢, 在客户端解析也较为麻烦。

JSON 是一种能够代替 XML 的轻量级数据交换格式, 不同的文献也给出了 JSON 在数据交换应用中的优势以及可行性。文献 [5] 给出了 JSON 与 XML 在传输效率方面的对比研究, 在扩展性和可读性上与 XML 相当, 但在数据传输效率和解析难度上远优于 XML; 文献 [6] 探讨了使用轻量级的 JSON 作为数据传输格式的可行性; 文献 [7]

收稿日期: 2014-12-20; 修订日期: 2015-03-05

基金项目: 国家自然科学基金项目 (61272454); 国家教育部博士点基金项目 (20130141110022); 武汉市科学技术局基金项目 (201302038)

作者简介: 张沪寅 (1962-), 男, 江苏苏州人, 博士, 教授, 研究方向为网络体系结构、高性能计算机、e-learning; +通讯作者: 屈乾松 (1991-), 男, 湖北汉川人, 硕士研究生, 研究方向为数据交换、认知网络; 胡瑞芸 (1991-), 女, 湖北荆门人, 硕士研究生, 研究方向为数据交换、认知网络。E-mail: 2012202110156whu.edu.cn

给出了在地理信息数据交换方面 JSON 的应用。

目前数据交换模型的架构设计多是采用“代理+服务器”的思想。文献 [8] 基于 SOA 的数据交换平台, 提出了“代理+服务器”的模式, 使系统层次结构清晰, 极大地降低了系统研发的难度; 文献 [9] 给出了在分布式系统中基于 SOA 平台技术的数据交换的研究, 在 JMS 的技术基础上实现了数据交换平台。

在此基础上, 本文提出的基于 JSON “中心+代理”的数据交换模型是一个全新的模型, 将要交换的数据转换成 JSON 信息流, 在数据交换代理节点之间建立可靠的连接, 对信息流进行传输, 数据交换中心对信息流进行上传、下载以及实时监控, 系统层次清晰, 在保证通用性的同时也大大提高了数据交换效率。

2 数据交换模型的介绍

模型在总体上以 JSON 作为总线技术, 数据交换中心 (data exchange center, DEC) 作为架构中心, 数据交换代理节点 (data exchange agent node, DEAN) 提供代理服务的设计结构。整个体系结构为星型结构, DEC 处于中心位置, 它是实现数据交换的中心, 它通过标准的 Web Service 接口为每个 DEAN 提供服务。数据交换的过程就是, 各个应用系统以 DEAN 作为代理接口与 DEC 进行消息与数据的交换, 所有消息与数据的交换均为 JSON 信息流。模型的总体结构如图 1 所示。

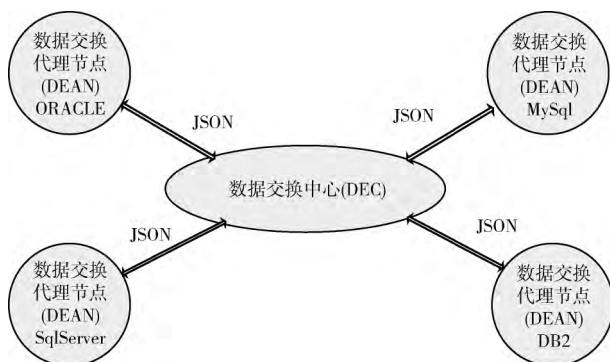


图 1 数据交换模型总体架构

DEAN 是整个数据交换模型的主要组成部分, 它是连接各个应用系统的桥梁, 每一个 DEAN 负责一种数据库系统的代理。它的主要功能是捕获应用系统的交互事件、封装信息流、与 DEC 交互。这里, 信息流包括消息与数据两部分, 消息则包括数据交互双方代理节点的信息、数据交互的操作信息等, 数据则是要上传或下载的数据。信息流的组成如图 2 所示。



图 2 信息流的组成

信息流的通用 JSON 格式为: { “MESSAGE”: [], “DATA”: [] }。

DEC 是整个数据交换模型的核心, 主要功能是信息流的解析、信息的应答交互管理、数据的上传/下载、服务监控等。信息流的解析是指 DEC 根据信息流的消息 (头部) 建立起两个要交互代理节点的数据传输通道, 并确定传输的方式 (比如数据插入、更新等), 这样可以保证数据传输的可靠性。信息的应答交互管理是基于“请求+应答”的模式, 在数据交换的前期, 数据发送方以“请求”的方式向数据接收方发起数据交换的连接, 收到接收方的“应答”后, 开始进行信息流的传输, 发送方上传数据, 接收方下载数据。服务监控是指 DEC 要对交换过程进行监控, 记录数据传输的日志, 当数据传输发生瘫痪时可以及时恢复, 保证数据传输的安全性。功能流程如图 3 所示。

因此 DEC 的整体行为就像一个虚拟的中心数据库, 同时又像一个交换机, 信息流以 JSON 在可靠的信道中传输, 从源节点到目的节点。整个数据交换的底层实现对各应用节点是透明的, 该结构耦合性低, 并且很容易扩展为层次的雪花型结构, 构建为多级的数据交换中心结构, 以支持更大范围的广域方案。当增加一种数据库类型时, 只需要建立一个对应的代理节点连接到 DEC 即可。

3 数据交换流程与实现

数据交换的流程包括: 建立连接、信息流封装、数据上传/下载、服务监控、会话结束。

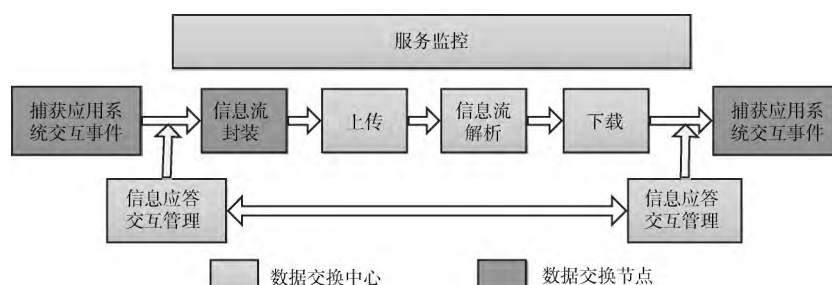


图 3 功能流程

3.1 建立连接

应用系统间建立数据交换的连接采用的是“请求+应答”的模式，数据交换的发送方通过对应的 DEAN 向 DEC 发送一条请求的 ASK 信息流，DEC 通过与数据接收方的交互获得应答请求的 ANSWER 信息流并返回给发送方，此时建立连接完成。如图 4 所示。

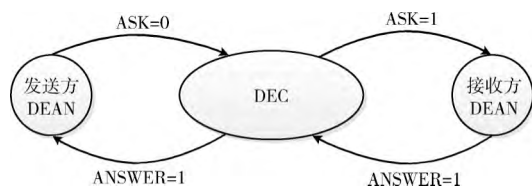


图 4 数据交换建立连接过程

ASK 信息流的消息部分主要包括发送方代理节点 SEND_ID、接收方代理节点 RESERVER_ID、信号量 ASK（初始化为 0，表示发送请求），若 DEC 找到对应的接收方 DEAN 的服务，DEC 将信号量 ASK 置为 1 并发送到对应的代理节点，接收方 DEAN 收到 ASK=1 的信号量后，发送一个 ANSWER=1 的信号量给 DEC，DEC 再转发给发送方 DEAN，发送方 DEAN 收到 ANSWER=1 的信号量，则知道建立连接成功。若 DEC 未找到对应的接收方 DEAN，则直接返回 ANSWER=0 的信号量，告诉发送方建立连接失败。

例如：ASK 信息流的 JSON 表示片段为：

```

{"MESSAGE": [{"SEND_ID": "ORACLE1001", "RESERVER_ID": "MYSQL1002", "ASK=0"}]}
"DATA": [{"}]
  
```

3.2 信息流封装

在数据交换中信息流的封装包括消息的封装与数据的封装，消息包括发送方代理节点 SERD_ID、接收方代理节点 REVERSER_ID、业务类型 BUSSINESS_TYPE、上传时间 UPLOAD_DATE 等。数据的封装是发送方代理节点通过调用 JDBC 得到源数据的数据集 ResultSetMetaData，将其转化为 JSON 格式数据，具体方法如下：

```

Public String resultSetToJson (ResultSet rs) {
    While(rs.next) { //遍历 ResultSet 里的每一条记录
        JSONObject writeJSONByData(); //将数据生成
        为 JSON 格式
        JSONObject writeJSONByInformation(); //将信息
        流生成 JSON 格式
    }
}
  
```

信息流封装好后，将其传输到 DEC 处理。

3.3 数据上传/下载

数据的上传/下载实质就是发送方将源数据上传到缓冲池，DEC 将源数据转换为目的数据，提供给接收方下载的过程。采用缓冲池来实现，缓冲池是一个固定大小的队列，当缓冲池未满时，发送方将源数据放入缓冲池，缓冲池满时，发送方将处于阻塞状态，等待缓冲池的资源，当超过最长等待时间时发送方给服务监控系统发送一个异常消息，交由服务监控系统处理。如图 5 所示。



图 5 数据上传/下载过程

采用缓冲池的目的是为了避免同时处理大量数据的交换，每次只对缓冲池里的数据进行交换，缓解了系统的压力，同时也有利于数据的恢复，当数据交换过程出错时，只需要对当前缓冲池里的数据进行恢复即可。

数据交换的过程：传统基于 XML 数据交换采用的是 XML 相关技术，如 XSLT、XPath 技术等。而 JSON 格式目前没用相关的技术对其进行转换，本文采用文献 [10] 提到的方案，用普通表格来存储各种数据库类型的映射关系，如需增加一种数据库类型，添加一行即可，该方案具有很强的可扩展性，见表 1。

表 1 数据库中常见的数据类型映射

Oracle	Mysql	SQL Server	DB2
number	smallint	bit	smallint
date	datetime	datetime	datetime
number	integer	Integer	integer
float	float	float	float
float	real/double	real	real
varchar2	varchar	varchar	varchar
char	char	char	char
clob	text	text/clob	clob/dclob
blob	blob	integer/blob	blob

下面给出数据类型转换的伪代码：

```

Filed fields=JSONObject toFields (jsonObject); //将
源数据转换为 Fileds 对象
for (int i=1; i<fields.length; i++) { //遍历数据
类型 Fileds
  
```

获取列名即 fields.name，判断是否有命名冲突，如果有重命名。

获取列数据类型 fields.type，查询表 1，获的相应的目的数据类型。

生成目的数据类型 filesds。

}

JSONObject writeJSON (filesds); //将目的数据类型转化为 JSON 格式

3.4 服务监控

服务监控是指对交换过程进行监控, 记录数据传输的日志, 当数据传输发生瘫痪时可以及时恢复, 保证数据传输的安全性。服务监控包括两个部分: 日志记录与异常处理。日志记录保存了系统在运行过程中的重要信息, 异常处理保证了系统在出现异常时能及时恢复。

日志记录的过程: ①发送方在上传一条源数据到缓冲池前, 要填写一个待发的日志信息到日志系统中。②DEC 根据日志系统中待发消息记录来对缓冲池中的源数据进行数据交换。③数据交换完成后, 要向日志系统中发送一条交换成功的日志消息, 然后通知接收方去下载目的数据。④接收方下载完数据后, 向日志系统发送一条接收成功的信息。

异常处理就是当服务监控系统收到一条异常消息时, 会根据日志系统里的消息记录进行相应的处理, 例如: 重新对缓冲池里的数据进行数据交换, 或者通知发送方重新发送一条源数据等。

3.5 会话结束

当发送方的源数据全部上传完成后, 会发送一条请求结束会话的消息 FIN MESSAGE 到日志系统, DEC 收到发送方的 FIN MESSAGE 后暂不做处理, 让发送方处于等待结束状态 WAIT CLOSED, 并等待接收方下载完成后通知接收方数据已经全部下载, 此时, 接收方也发送一条 FIN MESSAGE 到日志系统请求结束会话, 当 DEC 收到接收方的 FIN MESSAGE 后向双方同时发送结束会话的通知 FIN ACK, 并断开连接, 此时会话结束。

4 实验与分析

实验对某局在线考试系统中的数据集进行数据交换, 记录相应的数据交换时间, 实验环境为: Eclipse、Oracle 数据库、Mysql 数据库、Firefox 浏览器。实验描述了模型中单点对单点的数据交换过程 (ORCALE 代理节点到 MYSQL 代理节点的数据交换), 分别用 XML 和 JSON 作为传输流进行实验对比。利用 web services 实现数据交换中心与数据交换代理节点的主要功能, 包括 DataToXml、DataToJson、数据流的传输、数据格式转换等, 分别建立简单的基于 XML 与 JSON 的数据交换平台。经过分析知道, 数据交换总时间主要包括: 数据封装时间、数据在网络中传输时间、数据转换时间, 实验主要记录这 4 类数据的对比。针对 3 种不同大小的数据集, 得到的实验结果如图 6 所示。

由图 6 (a) 可以看出, 在封装时间上 XML 与 JSON 差

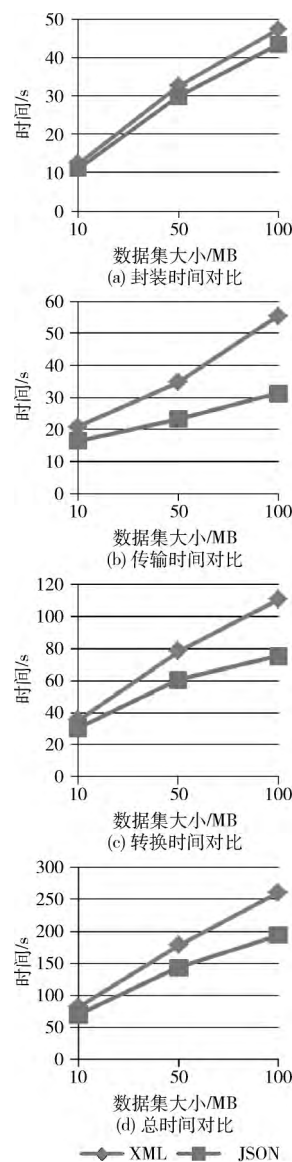


图 6 实验结果

别不大, 这与处理的方式有关, 封装 XML 的技术相对于 JSON 比较成熟。由图 6 (b) 可以看出, 在传输时间上, JSON 要远优于 XML, 差别也非常明显, 这也体现了 JSON 作为轻量级数据格式在网络传输中的优势。由图 6 (c) 可以看出, 在转换时间上, JSON 也要优于 XML, 这是由于 JSON 的解析要比 XML 更加简单。由图 6 (d) 可以看出, 在总时间上, JSON 也要优于 XML, 并随着数据集的增大, 优势越明显。

5 结束语

本文通过建立“中心+代理”的数据交换模型, 用 JSON 信息流来传输要交换的数据, 既提高了数据交换的效率, 又具有较高的通用性, 适合于各种不同数据库系统之间的数据交换, 同时也具有较强的扩展性, 当有新的数据

库系统接入时只需要添加对应的代理节点即可。模型实现的数据交换过程也类似于网络通信的过程,包括建立连接、数据交换、日志记录和结束会话,易于理解。本文将 JSON 作为数据交换平台的总线是一个创新,将 JSON 与其它技术相结合其优势会更加突出,如 AJAX、FLEX 技术、缓存技术等,这将会是接下来的研究方向,相信在不久的将来 JSON 在数据交换方面的应用将更加广泛和成熟。

参考文献:

- [1] HU Nengfa, TANG Weiping. General heterogeneous data exchange model based on XML [J]. Computer Engineering and Design, 2010, 31 (8): 1743-1745 (in Chinese). [胡能发, 唐为萍. 基于 XML 的通用异构数据交换模型 [J]. 计算机工程与设计, 2010, 31 (8): 1743-1745.]
- [2] HAO Shaohua, HAN Xie. Heterogeneous relational database integration model based on XML [J]. Computer Engineering and Design, 2010, 31 (24): 5285-5288 (in Chinese). [郝少华, 韩燮. 基于 XML 技术的异构关系数据库集成模型 [J]. 计算机工程与设计, 2010, 31 (24): 5285-5288.]
- [3] SUN Teng, SUN Anjian. Data exchange platform for e-government integration based on XML [J]. Computer Applications and Software, 2012, 29 (5): 188-190 (in Chinese). [孙腾, 孙安健. 基于 XML 数据交换的电子政务系统集成 [J]. 计算机应用与软件, 2012, 29 (5): 188-190.]
- [4] Bojańczyk M, Kołodziejczyk L A, Murlak F. Solutions in XML data exchange [C] //Proceedings of the 14th International Conference on Database Theory. ACM, 2011: 102-113.
- [5] GAO Jing, DUAN Huichuan. Research on data transmission efficiency of JSON [J]. Computer Engineering and Design, 2011, 32 (7): 2267-2270 (in Chinese). [高静, 段会川. JSON 数据传输效率研究 [J]. 计算机工程与设计, 2011, 32 (7): 2267-2270.]
- [6] GU Fangzhou, SHEN Bo. Application study on JSON data exchange format in integration of heterogeneous system [J]. Railway Computer Application, 2012, 21 (2): 1-4 (in Chinese). [谷方舟, 沈波. JSON 数据交换格式在异构系统集成中的应用研究 [J]. 铁路计算机应用, 2012, 21 (2): 1-4.]
- [7] HAN Min, FENG Hao. Research of the method of geographic information data exchange based on JSON [J]. Science of Surveying and Mapping, 2010, 35 (1): 159-161 (in Chinese). [韩敏, 冯浩. 基于 JSON 的地理信息数据交换方法研究 [J]. 测绘科学, 2010, 35 (1): 159-161.]
- [8] YANG Li, LIAN Dongben. Design of data exchange platform based on SOA [J]. Computer Systems & Applications, 2011, 20 (5): 30-33 (in Chinese). [杨丽, 廉东本. 基于 SOA 的数据交换平台设计 [J]. 计算机系统应用, 2011, 20 (5): 30-33.]
- [9] LI Yanchun, LI Xin, JIAO Wenbin. Design and implementation of data exchange platform for distributed information system [J]. Computer Engineering and Design, 2012, 33 (7): 2640-2645 (in Chinese). [李艳春, 李新, 焦文彬. 分布式信息系统中数据交换平台设计与实现 [J]. 计算机工程与设计, 2012, 33 (7): 2640-2645.]
- [10] SHEN Fengxian, ZHU Qiaoming, LIU Zhao, et al. Migration method of database management system based on JDBC and XML [J]. Computer Engineering and Design, 2008, 29 (20): 5376-5378 (in Chinese). [沈凤仙, 朱巧明, 刘钊, 等. 基于 JDBC 和 XML 的数据库系统迁移方法 [J]. 计算机工程与设计, 2008, 29 (20): 5376-5378.]

(上接第 3327 页)

- [8] Tang Xinmin, Wang Yuting, Han Songchen. Aircraft taxi route planning for a-SMGCS based on discrete event dynamic system modeling [C] //Second International Conference on Computer Modeling and Simulation, 2010.
- [9] LE Xuemei, YANG Kai. Research of technical of route planning automatically in A-SMGCS system [J]. Modern Electronic Engineering, 2008: 43-45 (in Chinese). [勒学梅, 杨恺. A-SMGCS 中自动路由规划关键技术研究 [J]. 现代电子工程, 2008: 43-45.]
- [10] WANG Chong. Research on taxiing route planning of surface aircrafts and 3D simulation system [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012 (in Chinese). [王翀. 机场场面航空器滑行路由规划及三维仿真研究 [D]. 南京: 南京航空航天大学, 2012.]