

文章编号: 1005-8451 (2012) 02-0001-04

JSON 数据交换格式在异构系统集成中的应用研究

谷方舟, 沈 波

(北京交通大学 通信与信息系统北京市重点实验室, 北京 100044)

摘 要: 针对现代计算科学中的异构系统集成问题以及其数据交换格式的选择问题, 本文分析了通用的XML技术的不足, 介绍了新一代的JSON数据交换格式的概念、特点及其对数据的表示形式, 研究对比了JSON技术的优越性, 并在此基础上以一种异构系统集成方案为研究对象, 探讨了在该领域使用轻量级的JSON技术作为数据传输格式的可行性。

关键词: JSON; 异构集成; 数据交换; 信息; XML

中图分类号: U29: TP393 **文献标识码:** A

Application study on JSON data exchange format in integration of Heterogeneous System

GU Fang-zhou, SHEN Bo

(Key Laboratory of Communication & Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China)

Abstract: According to the integration problems of Heterogeneous System in the modern computing science and the choice problem of its data exchange format, this paper analyzed the deficiency of general XML technology, and introduced the concept, characteristics and the way to represent the data for new generation of JSON data exchange format, studied and compared the superiority of JSON technology. And basing on that, the paper took a kind of integration solution of Heterogeneous System as research object, discussed the feasibility of the use of the lightweight JSON technology as the data transmission format in the field of the integration of Heterogeneous System.

Key words: JSON; heterogeneous integration; data exchange; information; XML

随着企业国际化、信息化的不断发展及深入, 在企业内部建设了多个应用信息系统, 并且各企业之间也会有大量的信息和数据的共享与交互。这些信息系统由于多种原因必然会存在着开发环境、运行平台、地理位置及所用数据库等的不同, 从而形成了企业内部与企业之间的异构数据源。这些分布式异构系统中的数据交换与信息的集成是最关键的, 这就是异构系统集成^[1]问题。

大量的数据信息存储于异构的数据源当中, 使得这些信息无法得到综合利用, 也就无法为企业的决策者提供全面的数据作参考以对企业的整体业务流程进行统一的管理。数据信息系统的交

换与集成, 是一个企业信息化的基础。

在融合这些异构数据源的过程中, 有一些关键性的问题还需要得到解决。例如, 一些数据库系统构建来自不同的数据库模型, 不同数据库中有相同语义的对象可能具有不同的名称, 而且这些对象可能由不同的单位进行度量。另外, 异构系统集成还存在识别冲突、表示冲突、域冲突等目前, 关于异构数据集成系统的研究涉及多种计算机技术, 如分布式对象技术、XML、面向对象技术等。

1 数据交换格式

1.1 XML

XML (The Extensible Markup Language, 可扩展标识语言)和HTML一样, XML同样来源于

收稿日期: 2011-05-04

基金项目: 中央高校基本科研业务费专项资金。

作者简介: 谷方舟, 在读硕士研究生; 沈 波, 副教授。

SGML(Standard Generalized Markup Language 通用标识语言标准),是其一个简化子集。XML是一种简单的、灵活的文本格式,是一套定义语义标记的规则^[2],这些标记将文档分成许多部件并对这些部件加以标识。

XML是标记语言,理解XML,首先要理解标记。HTML的标记(Markup),通俗地讲,它就是一种用来给文本添加标记的语言。在HTML里每个标志都是有确切含义的。例如,在HTML中,标签的含义是要求HTML浏览器将一段文本加粗表示,而标签<CENTER>的含义是告诉浏览器将这段文本在一行的中间显示。XML并非像HTML那样,提供了一组事先已经定义好了的标签,而是提供了一个标准,利用这个标准,可以根据实际需要定义新的置标语言,并为这个置标语言规定特有的一套标签。准确的说,XML是一种源置标语言,允许根据它所提供的规则,制定各种各样的置标语言。

1.2 XML 技术的不足

在异构系统集成领域,目前已有的各种实现方案中大多选择XML作为数据传输格式,XML在其中的应用也日臻成熟,提高了数据交换共享平台的通用性和扩展性,但存在着以下的问题:

(1)无论是服务器端生成或处理XML,还是客户端用JavaScript解析XML,都经常导致复杂的代码,较低的开发效率。

(2)XML文件繁冗,需要较长的传输时间来交换传输信息,影响了应用系统和网络的性能^[3]。

(3)XML文件为树状存储结构,虽然搜索效率高,但是插入和修改比较困难。

(4)大数据量低效率。XML的文本表现手法、标记的符号化会导致XML数据比二进制表现数据量增加,尤其当数据量很大的时候,效率就成为很大的问题。

(5)管理功能不完善。XML文档做为数据使用,没有数据库系统那样完善的管理功能。

(6)通信难。由于XML是源置标语言,任何人、公司和组织都可以利用它定义新的标准,这些标准间的通信就成了很大的问题。

1.3 JSON 数据交换格式

介绍JSON前,必须先了解JavaScript语言。JavaScript是一种面向对象的编程语言,主要用于

操作Web浏览器与Web内容的交互,当用户访问Web页面时,其完全由浏览器阅读解析和处理。从安全性的角度来讲,这是JavaScript语言的一个重要特点。所有使用JavaScript语言开发的应用只能在浏览器中运行使用^[4],同时使用该语言编写的任何文件也不能保存于浏览器之外的环境中。以上这些特性,使得JavaScript语言与其它计算机编程语言有着很大的不同。其它的Web进程既可以由服务器端的脚本语言管理(如asp网页)也可以由浏览器插件管理(如Flash网页),而这些服务器端或插件形式的应用都不具有JavaScript语言的安全性的限制。

JSON^[5]是(Java Script Object Notation, JavaScript 对象符号)一种轻量级的数据交换格式。JSON采用与编程语言无关的文本格式,但是也使用了类C语言(包括C, C++, C#, Java, JavaScript, Perl, Python等)的习惯,这些特性使JSON成为理想的数据交换格式。虽然,JSON是一种用来存储数据的JavaScript语言符号,但是更严格的说,JSON不是一种编程语言,其实是一种直接定义部分文档对象模型的方法。像JavaScript语言的其他部分一样,JSON可以直接用来编写Web页面的某些部分。JSON文件就是普通的ASCII文本文件,使用普通的文字编辑器就可以创建。与机器代码不同的是,JSON永远不需要编译和执行。

JSON建构于两种结构:

(1)“名称/值”对的集合。不同的语言中,它被理解为对象(object),纪录(record),结构(struct),字典(dictionary),哈希表(hash table),键列表(keyed list),关联数组(associative array)。

(2)值的有序列表(An ordered list of values)。在大部分语言中,它被理解为数组(array)。

下面给出JSON的简单示例:

```
{ "firstName": "Daniel" }
```

这是JSON最典型的结构,相当于:

```
firstName = Daniel
```

JSON还可以创建包含多个名称/值对的记录,比如:

```
{ "firstName": "Daniel", "lastName": "Teddy",  
  "email": "aaaa" }
```

当需要表示一组值时,JSON不但能够提高可读性,而且可以减少复杂性。例如,假设要表示一

个人名列表,在XML中,需要许多开始标记和结束标记,如果使用典型的“名称/值”对,还必须建立一种专有的数据格式。而如果使用JSON,就只需将多个带花括号的记录分组在一起:

```
{ "people": [  
  { "firstName": " Daniel ", "lastName": " Teddy  
", "email": "aaaa" },  
  { "firstName": "Jason", "lastName": "Hunter",  
"email": "bbbb" },  
  { "firstName": "Elliott", "lastName": "Harold",  
"email": "cccc" } ] }
```

在这个示例中,只有一个名为people的变量,值是包含3个条目的数组,每个条目是一个人的记录,其中包含名、姓以及电子邮件地址。

1.4 JSON 的应用优势

与XML相比,JSON具有更好的简单性和灵活性,非常易于阅读和编写,同时也易于机器的解析和生成,这是因为JSON使用了数据分隔符,它的优点可能并不那么明显,但它简化了数据访问。使用这些数据分隔符时,JavaScript引擎对数据结构(如字符串、数组、对象)的内部表示恰好与这些符号相同,这将开创一条比DOM技术更为便捷的数据访问途径。

JSON的另一个优点是它的非冗长性。在XML中,打开和关闭标记是必需的,这样才能满足标记的依从性;在JSON中,所有这些要求只需通过一个简单的括号即可满足。在包含有大量字段的数据交换中,传统的XML标记将会延长数据交换时间,而通过简单的字节数比较发现,对于等效的JSON和XML有效负载,前者总是小于后者,因而JSON比XML具有更高的线上传输效率。

2 应用实例

2.1 城市轨道交通信息交换平台建设需求

城市地铁在建设过程中和开通运营后,各级相关单位必然会使用到多个应用信息系统,而由于时间关系以及获得方式等的不同,各个子系统在数据格式、数据存储和数据处理方法上都存在着较大的差异,使得城市轨道交通信息不能被立体化、全方位管理以及综合利用^[6]。图1显示了地铁各信息系统与城市轨道交通指挥中心的结构关

系。从图1中可以看出,位于上层的各信息系统之间相互独立,缺乏信息资源的共享与交互,无法完成联动协调以及对数据进行深层次综合挖掘的要求。依据以上分析,城轨交通信息交换平台的建设应该满足以下几点功能要求:

(1) 平台应该对已有的信息系统具有很好的兼容性。(2) 能够完成接入平台系统之间的信息交互。(3) 平台和各信息系统之间完全解耦。

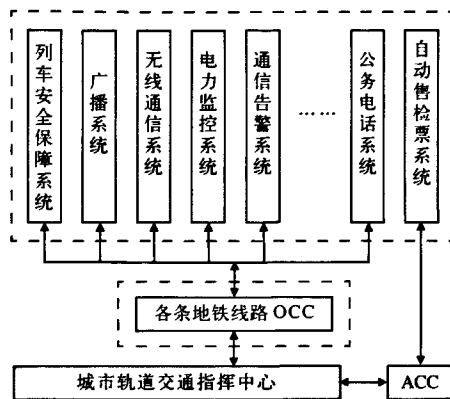


图1 城市轨道交通指挥中心信息系统

2.2 城轨交通信息交换平台建设方案

本文的研究对象城轨信息交换平台采用XML作为各个接入信息系统的中间信息交互格式,以ESB技术为EAI(Enterprise Application Integration)核心技术路线,以JMS(Java Message Service)消息中间件为系统核心架构技术,实现了系统异步消息通信机制。

图2是城轨交通信息交换平台的结构图,平台主要由3部分组成。第1部分是外部信息系统的接入部分。对于已有信息系统的接入采用适配器的方案解决,而对于新建的系统则直接在系统中嵌入这部分功能。第2部分是平台系统的基础服务管理,这一部分包括系统注册、事物管理以及日志记录功能。第3部分是该交换平台的核心,按分层的结构来组织。顶层为系统服务层,负责接收接入信息系统的请求,并作响应处理。第2层为数据协议层,提供对需要交换的使用不同协议的数据的支持。第3层为数据格式转换层,通过XSLT技术完成各个信息系统不同数据格式的转换工作。

3 JSON 数据交换格式在交换平台中的应用

接入平台的信息系统之间进行数据交换,所

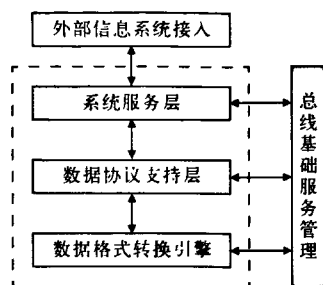


图2 城轨交通信息交换平台结构

传输的各种信息均使用的是XML文件格式,使用JSON技术对交换平台的数据传输技术进行改造,使得最终的各种信息流完全变为JSON文件格式的交互。应用系统与平台间的交互消息分为4种类型:注册消息、注册响应消息、数据请求消息和请求响应消息。下面给出交换平台的数据请求消息的JSON格式片段:

```

{
  "ApplicationId": "ApplicationId",
  "DestinationId": "DestinationId",
  "SessionID": "SessionID",
  "RequestID": "RequestID",
  "Requests": [
    .....
  ]
}

```

由以上可以看出,对于需要交换的各种数据或者信息,平台系统均将其包含在这4种JSON消息中进行传输。出于安全性考虑,任何信息系统要接入平台交换数据都要先发送注册消息,通过合法性验证后,平台发送注册响应消息允许系统接入。当开始进行数据交换阶段时,参与交换的信息系统通过数据请求消息和请求响应消息完成最终的数据交换。下面给出在服务器端输出JSON格式消息的java代码片段:

```

interface JSONGenerator
{
  String getJSONMessage(Information
information);
}

```

其中,Information为一个维护消息体信息的类。这个接口里只有一个方法,用来实际生成JSON格式的消息,每种生成JSON消息的类都要继承这个接口并实现这个方法,当调用这个方法

时,就可以通过接口调用,从而达到与实现的具体方法解耦的目的。

```

public Class Information implements java.io.
Serializable
{
  private java.lang.String applicationId;// 消息
发送源 ID
  private java.lang.String destinationId;// 消息
目的地 ID
  .....
}

```

4 结束语

本文介绍了JSON技术以及它在异构系统集成领域应用方面的优势,并且结合具体的应用实例探讨了JSON技术在特定背景下的应用可行性。本文只给出了JSON在异构系统集成方面的应用,而在利用AJAX的应用程序中JSON的应用效果更为突出,并且关于利用JSON进行数据缓存及JSON在J2EE中的其他应用还有待进一步研究。当然,关于JSON的应用才刚刚起步,但是JSON特有的优势必将使得其应用越来越广泛。

参考文献:

- [1] 彭 斌,赵曦滨,雍建平,顾 明.基于消息机制的异构系统集成方案[J].计算机应用研究,2005(8):43.
- [2] Maarouf,M.Y.;Chung,S.M.. XML Integrated Environment for Service-Oriented Data Management[C]; 2008, 20th IEEE International Conference on Tools with Artificial Intelligence; 2008, 361-368.
- [3] Bayu Erfianto, Ahmad Kamil Mahmood. Abdullah Sani Abdul Rahman; Modeling Context and Exchange Format for Context-Aware Computing[C]; The 5th Student Conference on Research and Development SCORED, 2007.
- [4] 朱印宏. JavaScript征途[M]. 北京:电子工业出版社,2009.
- [5] Prochazka, M.; Holub, P.; Hladka, E.. jSon: Network of Active Elements with Peer-to-Peer Control Plane[C]. 2009 Eighth International Conference on Networks; 2009, 208.
- [6] 孙 章,蒲 琪.城市轨道交通概论[M].北京:人民交通出版社,2010.

责任编辑 徐侃春