

面向调度主站一体化的数据网格设计与实现*

于振波¹, 牛文楠², 鄢安娜¹, 郭子健², 陶文伟², 吴润泽¹, 张海波¹

(1. 华北电力大学 电气与电子工程学院, 北京 102206;

2. 深圳供电局, 广东 深圳 518001)

摘要: 高性能电力应用致力于搭建用于调度数据的统一管理、集中处理平台。数据网格易于海量数据共享的特点使其适用于实现调度主站一体化。构建了调度主站数据网格的基本架构, 利用网格软件 Globus Toolkit 建立网格环境, 在此基础上配置数据库访问中间件 OGSA-DAI 来建立面向调度主站一体化的数据网格实验平台。

关键词: 调度主站一体化; 数据网格; 网格服务; 统一访问

中图分类号: TP311

文献标识码: A

文章编号: 1674-7720(2012)11-0092-03

Design and implementation of data grid oriented dispatch master station integration

Yu Zhenbo¹, Niu Wennan², Yan Anna¹, Guo Zijian², Tao Wenwei², Wu Runze¹, Zhang Haibo¹

(1. School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China;

2. Shenzhen Power Supply Bureau, Shenzhen 518001, China)

Abstract: The high-performance power applications are committed to build a unified management and centralized processing platform for scheduling data. Data grid is suitable to realize the integration of the dispatch master station, because it is easy to share massive data. This paper used OGSA-DAI to construct the experiment platform of data grid oriented dispatch master station integration based on the grid environment built by Globus Toolkit.

Key words: dispatch master station integration; data grid; grid service; unified access

电力计算需要在分布式区域建立数据交换与共享, 以实现电网的整体分析计算, 提高实时监控性能与分布式计算的能力。而数据网格通过统一分享平台, 实现对于资源的统一管理和动态组合, 为用户提供能够完成高性能计算的一体化平台。电力数据网格致力于建立一种高性能、支持分布式电力计算的协作环境, 满足电力系统新型并行与分布式协作型业务的需要, 提高信息系统的资源利用率^[1-2]。

OGSA-DAI 是目前技术较为成熟的数据网格中间件, 建立在 Globus Toolkit 的网格环境上, 通过屏蔽下层操作来实现对异构数据库的统一化访问^[3]。利用数据网格便于异构资源管理和共享的功能, 实现分布式数据资源管理和共享。

1 数据网格

1.1 数据网格环境

Globus 是一种用于构建计算网络的开放体系结构、开放标准的项目, 对资源管理、安全与信息服务及数据管理等网格计算的关键理论进行研究并提供了基本的机制和接口, 适合大型网格系统运行的大型应用程序的开发。Globus Toolkit 工具包来源于 Globus 项目, 是一个开放源码的网格的基础平台, 基于开放结构、开放服务资源和软件库, 并支持网格和网格应用, 目的是为构建网格应用提供中间件服务和程序库。

1.2 中间件

OGSA-DAI 中间件完全由 Java 编写, 因而适用于各种系统平台。通过 Web service 进行数据访问, 即屏蔽服务内部的配置, 将查询动作以服务的方式执行, 为构建

* 基金项目: 国家自然科学基金项目(50807014)

高层次的数据网格服务提供开发工具。

2 调度主站网格架构

数据网格适用于解决数据密集型电力计算应用中方便高效使用分布式数据资源的问题,其内容主要集中在广域、异构、分布环境下如何对数据进行管理;如何从调度主站的各种异构数据资源中获取数据;并通过地域分布的协作和处理,从数据源中获取有用信息。面向调度主站的数据网格架构分为数据服务源、数据网格中间件和调度应用三个层次,如图1所示。

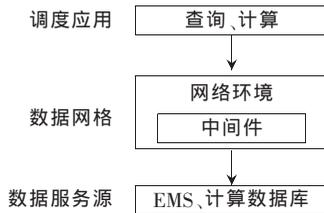


图1 调度主站网格架构

数据网格可以理解为是调度应用同底层存储系统之间的综合服务代理,它对应用程序提供一致的管理以及统一的数据结构,屏蔽操作的复杂性和网络上存储系统的异构性。该架构以数据网格为基础,建立输网调度自动化系统、电能计量自动化系统的数据一体化查询和管理平台。

3 调度主站网格设计实现

网格中间件,在整个查询的过程中涉及到多个模块的应用。整个软件的实现是建立在 Globus Toolkit 以及在此软件平台上的 OGSA-DAI 软件包。其实现包括实现数据网格中间件软件配置和数据管理应用接口编程。

3.1 中间件配置

中间软件配置是数据网格的实现基础和开发平台。首先安装 Globus Toolkit,重点是配置 Globus 的认证机制,正确分配用户证书;之后需要将 OGSA-DAI 作为 Globus Toolkit 中的新服务,配置到网格环境下的服务容器中,最后将数据库作为数据服务源配置到新服务下。中间件模块如图2所示。



图2 中间件软件配置

OGSA-DAI 在网格环境的服务容器中添加专门用来和数据库交互的网格服务,可以通过这个服务来查询数据库的数据,并返回查询的结果,在这个过程中实现的是将数据库注册到服务中,以便在访问能够得知都能访问的数据的目录。单个服务可以对应多个数据源,通过 Resource ID 来进行识别。

网格服务管理和发现模块是由 Globus Toolkit 实现的,调节各个服务的执行,具体的调用管理由 Globus Toolkit 自动实现。

3.2 数据管理接口

设计目的是将数据网格的基本管理功能封装起来的中间件软件模块,提供简单、方便的数据调用和管理

接口,为上层应用提供统一、共享的数据资源。而实现调度主站一体化,需要给应用提供异构电力数据库数据的统一查询,以及提供对底层的数据库的统一管理。调度主站数据网格的数据管理接口模块如图3所示。



图3 数据管理接口架构

(1) 元数据提取

元数据也就是用来描述数据的数据。在数据源注册的过程中,除需要向平台提交数据库的类型、服务器IP、端口号、数据库名称、只读用户与口令等参数外,可以通过元数据提取模块来提取各个数据库中的信息,包括包含的表、字段名等,以便能够建立所谓的虚拟数据库,方便定位到所需的服务和数据源。

(2) 服务定位和数据库连接

在访问数据库之前,当然要连接到数据库。这个过程可以分为两个部分:服务定位和数据库连接。

① 服务定位:如果建立了虚拟数据库,这可以通过程序在虚拟数据库中查找到所需的服务和相对应的数据库源。而在简单的情况下,可以直接使用 URL 来连接数据服务,Resource ID 来确定需要访问的数据库。

② 数据库连接:部署数据源时同时也部署了 JDBC,它是一种用于执行 SQL 语句的 Java API,可以为多种关系数据库提供统一访问,是由一组用 Java 语言编写的类和接口组成。当确定了服务和数据源后软件就通过 JDBC 实现对数据库的连接。

(3) SQL 查询

SQL 查询是软件最关键的部分,也是软件的目的。首先涉及到的就是对于查询语句的分析。如果访问单个数据库,则直接将 Sql 语句添加到执行文档,然后发送给数据库直接执行查询就可以。当 Sql 语句涉及到多个数据源时,就需要对 Sql 语句进行分析、拆分,然后再发送给对应的数据源。在整个执行的过程中,由于数据是以流的形式传送,也需要数据流化和存储模块的参与。

(4) 结果处理

对于执行结果的处理涉及到数据集成、数据格式转换、查询结果重组这三个模块。

① 数据集成:输配一体化的数据管理包括 EMS、DMS 以及计量数据库的统一访问。在某些查询中需要来自不同数据库的结果集成到一张表中。软件中采用的方法是连接到一个具有多个数据资源的数据服务在其中一个数据库中建立新表,将涉及到的数据库的查询结果汇总到这张表上。

②数据格式转换:最后的查询结果是以响应文档反应的。不同的应用客户端可能会对结果的形式有不同的要求,而软件的数据格式转化模块中的 XSITransform 活动可以实现将 xml 的响应文档转换为任何形式的文本文档,比如转换为 HTML 以便在浏览器中显示。

③查询结果重组:这个过程可以说就是将集成和格式转换后的查询结果以客户端要求的形式通过网格服务接口传送到相应的应用。

3.3 接口程序执行流程

OGSA-DAI 是实现异构数据库访问的软件包,它提供了许多编译好的 class 程序以便在编程时通过 import 调用。可以说整个编程的过程就是涉及到的 class 程序的调用和执行,而对于数据库的连接则是采用 JDBC 驱动实现的。接口程序流程图如图 4 所示。

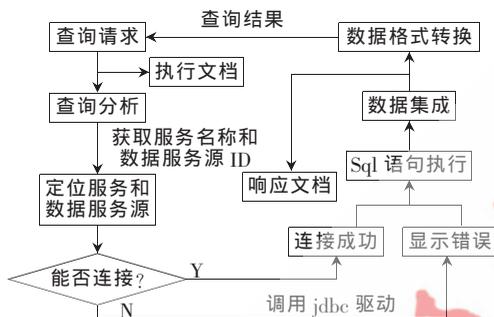


图 4 接口程序流程图

调度主站一体化数据涉及到 EMS 和计量数据库,平台设计的目的是要实现数据库的统一访问。假如需要查询某个变电站的所有信息,首先是接口接收到用户或者应用程序的查询需求信息,通过调用程序形成执行文档;然后是中间件,对执行文档进行分析,根据数据库的元数据判断是否需要将 Sql 语句进行拆分,也就是说,如果这个变电站的信息分别保存在多个数据库中,就需要对 Sql 语句进行逻辑拆分,在确定了 Resource ID 后连接数据库,执行完 Sql 语句后可生成响应文档,响应文档中涉及到模块的完成情况;最后是针对不同的应用而对结果做相应的处理,包括数据流的存储、数据集成以及数据格式的转换,以需要的形式将查询的变电站所有信息传回客户端或者电力应用。

OGSA-DAI 通过面向文档的接口支持与数据服务源的交互过程如图 5 所示。调度应用程序不需要直接操作数据服务源,而是通过向数据服务发送一个执行文档,然后数据服务将文档转送给代表实际数据源的数据服务源。数据服务源解析执行文档,并且执行文档中指定的动作。数据源再创建描述请求结果的响应文档,并且通过数据服务将响应文档返还到调度程序^[4]。

4 数据网格实验系统

调度主站数据网格实验平台利用局域网模拟分布在不同地理位置的数据库或者文件服务器,并通过网格环境将各个服务器作为统一资源进行管理,实现异构数

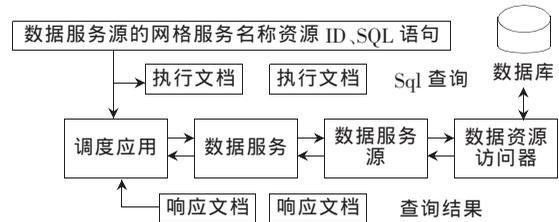


图 5 OGSA-DAI 执行流程

据库的访问,并为上层的电力应用提供统一数据查询工作。平台的结构包括软件配置和硬件两个部分,如图 6 所示。作为实验平台,硬件主要是数据库服务器并通过网络互联,软件配置主要是网格运行环境和为应用程序实现远程、异构数据库安全访问而开发的业务应用中间件。



图 6 数据网格实验系统结构

数据库服务器模拟网格节点,并安装配置 Globus Toolkit 网络运行中间件,具有安全服务认证功能。远程访问的主机则作为客户机通过安全证书认证后具有访问数据网格节点权限。在网格环境内部可以实现在客户主机上统一访问位于远程不同网络上的异构数据库,并将计量数据库、EMS 数据库安装在不同数据网格节点上。在网格外部,面向调度主站的接口可实现对网格节点数据库服务器统一的数据访问。实验平台可完成复杂、具有不同标识的输配电数据查询功能,通过向接口程序提交查询要求,通过 OGSA-DAI 数据库访问中间件的 SQL 语句处理后可以在为调度应用提供统一数据结构的各个数据库中存储的数据。经过查询,系统已得到欲查询的一次设备的相关实时量测数据,最后通过数据显示模块,系统将量测配置信息及相关的实时量测数据分别显示在界面上,供用户查看参考。

在数据网格环境下,开发了基于 OGSA-DAI 的数据网格中间件以实现底层异构数据管理与访问,利用数据网格中间件,建立分布式数据体系的统一通用的基本信息框架。将调度主站的各自动化系统与实时数据库真正统一在一起,直接查出各数据平台上涉及到的所有相关实时量测数据功能。OGSA-DAI 能够十分有效地屏蔽数据库的下层管理,用于异构数据的管理与访问,实现广泛的分布式数据资源管理和共享。

参考文献

- [1] DOBRZELECKI B, KRAUSE A, HUME A C, et al. Integrating distributed data sources with OGSA-DAI DQP and VIEWS[C]. UK e-Science All Hands Meeting, 2010.
- [2] LIU X H, SHI Y M, XU Y B, et al. Heterogeneous database integration of EPR system based on OGSA-DAI [C]. 2nd International Conference on High Performance Computing and Applications, 2009.
- [3] 罗清地, 蒋秀凤. 网格环境下基于 OGSA-DAI 的异构数据集成研究[J]. 计算机技术与发展, 2009(3): 144-147.
- [4] MARIA S, JEMAL H. An agent architecture for managing data resources in a grid environment [J]. Future Generation Computer Systems, 2009(7): 295-303.
- [5] PACITTI E, VALDURIEZ P, MATTOSO M. Grid data management: open problems and new issues [J]. Journal of Grid Computing, 2007, 5(3): 273-281.
(收稿日期: 2011-12-05)

作者简介:

于振波, 男, 1975 年生, 高级工程师, CISSP, 主要研究方向: 网络和信息系统宽带技术分析。

牛文楠, 男, 1981 年生, 工程师, 主要研究方向: 电力调度自动化及配网自动化的运行管理。

鄢安娜, 女, 1987 年生, 硕士研究生, 主要研究方向: 电力数据网格。

