

基于分布式数据集成企业门户系统的设计和实现

Design and Application of the Enterprise Information Portal System Based on Data - Integration of the Distributed Application

任 钢 (中国科学技术大学计算机科学技术系 安徽合肥 230026)
曾伟民 罗 宾 (湖北省电力公司 湖北武汉 430077)

摘 要: 在信息资源分布式网络化的趋势下,开放性和集成性成了企业信息化的重要发展方向。企业信息门户是近年 IT 领域的一项重要新技术,其特点就是可以对分散不同地域的数据加以集成和整合,实现统一的访问模式和管理。本文就结合湖北电力公司生产信息门户系统的实际开发情况,从数据集成,数据复制、数据聚合到企业信息门户的技术设计和实现方法都进行了详细的分析和说明。

关键词: 分布式数据集成 Oracle 高级数据复制 企业信息门户 门户小程序 BEA WebLogic 门户系统

1 引言

随着计算机技术和我国电力企业信息化的建设,地市电力企业内部已经建立了许多信息系统,并且有一定的信息化基础。然而由于这些系统存放在不同的服务器上,采用不同数据存放格式和访问模式,从而使这些系统成为了信息孤岛。供电地市局与省局的交互关系是一方面地市电力企业要为本单位提供信息,同时也要为自己的上级单位提供更多的信息服务,另一方面作为地市电力企业的上级单位省公司,要及时地了解下面地市公司的信息状况,要对下属单位的全部情况做汇总、统计、在线联机分析等操作。

湖北省电力公司就是处于这样的一个状况,现在全省 11 个地市公司都已经推广运行了湖北电网输变电生产管理系统,在各个地市都有比较完备齐全的业务数据信息。作为管理单位的省电力公司,要实时了解下属各个供电公司生产实时情况。尤其是一些比较紧急的缺陷处理、故障问题等有时效的数据。

企业信息门户是企业信息化建设的重要发展方向,数据复制是数据集成的最成熟方式。采用数据复制技术和 Portal 技术建立企业信息门户,不但可以为用户提供不同信息和应用程序的单一访问入口,并且可以实现了安全互动和一个可定制的接口等。本文就针对湖北电网生产门户系统的实现,来说明如何结合数据复制和 Portal 技术处理分布式业务应用系统,建

设企业信息门户。

2 关键技术

2.1 数据复制

数据复制,就是将数据库中的数据拷贝到另外一个或多个不同的物理站点上,从而保持源数据库与目标数据库中指定数据的一致性。按照数据复制的实时性,数据复制可分为同步数据复制和异步数据复制。

在分布式数据库系统中,由于数据复制和传输的使用比较频繁。几乎所有的大型数据库系统都提供了自己的数据复制解决方案和数据复制组件。

2.2 Portal 技术

Portal 即门户,是一种体系框架,是一个建立信息网站的集成环境,是一种附加在现有的企业管理信息系统之上的一层智能化较高 Web 应用。它能实现内容整合、单点登录、个性化以及查询分析等功能。

Portal 的组成包括三部分: Portal Server, Portlet Container 以及 Portlet 组件。Portal Server 就是指一个基于 Web 的系统,由各种不同来源或不同网站取得各式各样的信息且将这些信息放在网页之中组合而成的呈现平台。Portlet Container 是提供 Portlets 执行的环境,包含了许多 Portlets 并管理它们的生命周期。Portlet 组件是由 Portlet Container 所管理的 Web 组件,专门处理客户的 Request 并产生各种动态的信息内容。

2.3 Portlet 技术

Portlet 是 Portal 中的核心组件,用 Portlet API 来编写,是运行在 Portal 环境中的独立程序。Portlet 可以由任何支持 HTTP 协议进行通信的开发语言写成。Portlet 得到的信息内容通过 Portal 解释并将其显示在相应的页面视图中。

Portlet 一般基于两种标准进行开发,一种是 JSR168 Servlet 标准,保证 Java 服务器之间的互操作性。以便 Portal 可以对其进行访问。另一种标准是 WSRP,主要提供 Portal 和其他 Web 应用程序之间的通信功能。

2.4 Struts 框架

Struts 是 Apache jakarta 项目的一部分,Struts 是 MVC 的一种实现,它很好地结合了 JSP、Java Servlet、Java Bean、XML、Tag Lib 等技术,开发 Web 应用程序的开放源码的 Framework。

2.5 WSRP

WSRP (WEB SERVICES FOR REMOTE PORTALS) 是 OASIS Web Services for Remote Portlet 技术委员会在 2002 年制定的标准。其目的是开发一个允许门户之间、门户与其他集成应用和内容的 Web 应用之间,能够以即插即用的方式进行交互的 Web Service 标准。

3 系统设计与实现

3.1 企业的现状和目标

湖北省电力公司的现状是:省公司下属各个地市公司都运行一套完整的、版本统一的输变电生产管理系统,拥有大量的输变电生产事务数据,同时在每个地市公司都有针对本地数据库的 Web 综合查询系统。输变电生产管理系统采用基于 PowerBuilder + EAServer (Sybase Enterprise Application Server) 的 C/A/S 三层结构模式 (非 J2EE 平台),地市公司 Web 综合查询系统采用 IIS (Internet Information Server) + ASP 技术的 B/S 结构查询模式,数据库系统都是 Oracle。省公司与地市公司是通过广域网连接。省公司数据中心采用 Oracle 数据库,省公司的应用服务器是支持 J2EE 规范的 BEA WebLogic Server。

湖北省电力公司建设信息门户的目的是:省公司通过企业信息门户系统迅速、全面、准确、及时地了解下面各个地市公司的生产情况,对于一些关键业务还

要细节到字段程度。

3.2 系统总体架构模式

针对湖北电网生产信息系统的现状,采用如下解决方案:数据集成方式采用 Oracle 高级复制机制,业务展现方式采用 BEA WebLogic Server 的 Portal 门户。总体实现结构如图 1。

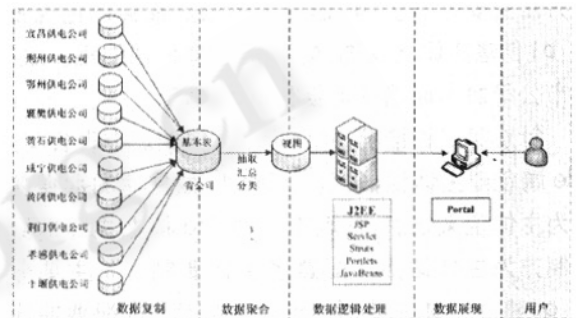


图 1 湖北电网生产门户系统总体框架图

方案说明:按照一定的时间间隔,把各个地市供电公司电力生产数据复制到省公司电力生产数据库。省公司的综合查询、分析和统计都是在省公司电力生产数据库上进行。

3.3 基于数据复制的数据集成方案

Oracle 高级数据复制有三种模式:分别是多主站复制 (如图 2);快照复制 (如图 3);混合复制 (如图 4)。

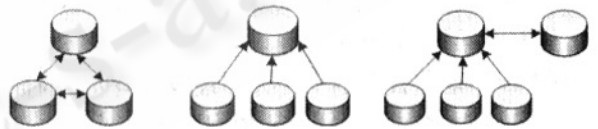


图 2 多主站复制 图 3 快照复制 图 4 混合复制

根据湖北现在全省的生产管理模式,地市公司为数据的生产者,省公司为接受地区数据的消费者,省公司所有的地市公司的生产运行一线数据都是由各地市公司提供给省公司,省公司数据库只能被动接受各地市公司提供数据。数据复制都是基于异步方式的单向传递。各个地市公司业务表的数据库模型是完全一样的。

对于地市公司的数据复制模式,也按照 Oracle 复制模式进行区别对待。主要是按照三种不同类型的数据进行数据集成。第一种是实时性比较强的关系数

据,第二种是实时性比较弱的关系数据,第三种是大文本数据(如 BLOB、CLOB 或 Long Row)。

针对第一种情况,采用数据快照(又称实体化视图)。各地市公司 Oracle 数据库的数据复制方为主体站点(数据提供方),省公司 Oracle 数据库的数据复制方为实体化视图站点(数据接收方)。配置实体化视图方法主要分三步:a)建立主体站点或实体化视图站点;b)创建实体化视图;c)创建实体化视图组。这样,地市公司数据能迅速地复制到省公司。

针对第二种情况,采用多主体复制。在省公司 Oracle 端创建主体组,省公司 Oracle 数据库的数据复制方为主体定义站点,各地市公司 Oracle 数据库的数据复制方为主体站点。配置多主体复制方法主要分三步:a)建立主体站点;b)创建主体组;c)配置冲突解析。为什么要在省公司创建主体组并把省公司 Oracle 数据库的数据复制方规定为主体定义站点?这是由于在多主体复制中,各个主体站点是对等的,只有创建省公司主体组的站点相对于其他地市公司主体站点来说,权限更大一点,它能够控制整个复制组是静默还是正常运行。而且只有该站点能够控制复制组中的复制对象。其他的主体站点只能删除自己的复制组中的对象,而该删除并不会传播到主体定义站点。因此,创建主体组的站点经常被称作主体定义站点,在一个复制组中只能有一个主体定义站点。

针对第三种情况,不复制数据。原因如下:数据量太大,复制时间长,系统开销大,影响正常的系统使用,比如其中一个地市公司的一个大字段表就有 4G 的数据。所以这种大数据、长时间的复制过程没有必要进行。对于这些数据的查询可以通过其他方式来实现,本文在后面的查询模式中有说明。

3.4 基于 VIEW 的数据展现模式

在省公司数据库系统中,分别给不同的地市公司创建了对应的 Oracle 用户。下面一个地市公司的数据只能复制给省公司的一个 Oracle 用户。这样可以保证数据的唯一性和一致性,减少复制过程中数据冲突。

复制到省公司的数据都是业务基本表数据,这些数据并不适合联机分析统计。原因有二:①由于基本表在数据库设计上主要是为了实现新增、删除和修改的事务处理,一般遵循 3NF,数据表之间关联比较复杂。②有些基本表的数据量比较大,复制上来还要叠

加,数据成几何级增长,比如,一个地市公司业务表有 1 万条数据,那么 11 个地市公司就有 11 万条数据。因此,要对这些复制上来的原始数据进行优化处理。当数据复制完成后,就调用相应的触发器进行操作,这些操作包括抽取、清洗、分类、汇总和统计操作。这样,当用户在提取数据的时候,不会因为重新进行数据处理而感到延时。同时,引入了视图进行汇总。数据表视图是由 SQL 查询语句从一个或多个表中建立的虚拟表。引入数据表视图的目的是:①允许用户访问视图而不直接访问基表;②限定从数据表中检索数据内容;③可以集成多个基表而不用在程序中写 SQL 语句,开发维护方便。从而达到优化查询的目的。

数据处理按照两种方式进行,一种是基表直接转化为视图,如图 5,这种情况主要针对基表数据量不大或者基表数据模型不复杂。还有一种是基表先生成二次表,再转化为视图,如图 6。这种情况主要针对基表数据量比较大(上万条记录)或者基表数据模型比较复杂(50 个字段以上,而且还有复杂的主外键约束)。

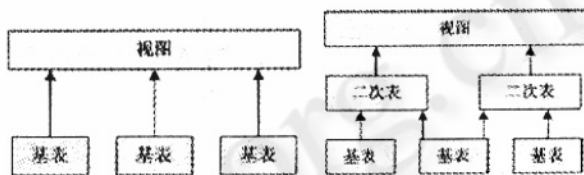


图 5 基表之接转化成视图 图 6 基表间接转化成视图

用户查询模式一般也是按照省公司汇总、地市公司汇总、地市公司数据列表,地市公司数据明细的查阅顺序来实现。这种模式可以支持上卷,下钻,切片,切面等手段来查阅和挖掘数据。在数据展现的过程中分别用图表,曲线等直观方式来展示。对于比较特殊要求的数据,在首页上就直接显示数据明细列表并进行提醒。比如,对于设备缺陷、故障、事故等信息,就直接显示,并根据不同的处理状态(发现状态,按时消除状态,未消除状态)用不同的颜色标志出来。当希望查阅到最初始记录性数据时,通过参数传递,可以直接访问到地市公司 Web 综合查询网站的相关记录的信息明细。比如在查阅电力设备台帐的时候,用户希望能看到设备台帐的相关图片,部件构造,这些数据都按照大文本形式保存在下面各个供电公司的数据库中,并没

有通过数据复制到省公司数据中心。我们通过参数传递(如链接的 URL 和设备编号等),就可以直接链接到地市公司 Web 综合查询网站的设备网页上,有针对性地浏览设备台帐的相关图片,部件构造文档等大文本信息。

3.5 结合 Struts 的 Portlet 方案

在 WebLogic Portal 中,Portlet 是门户 Web 应用程序的基本组成部分,可以将 Portlet 看作呈现应用程序、信息和业务流程的窗口。Portlet 可以彼此通信,它们可以使用 Java 控件,并可添加到通过应用程序确定用户路径的 Java 页面流中。一个页面可以有多个 Portlet。一个 Portlet 也可以有多个实例。Portlet 本身可以使用 JSP、页面流或可选支持文件,并且可以生成符合 JSR 168 标准的 Portlet,以实现 Portlet 兼容性。Portlet 可以使用现有的 Web 应用程序和内容(ASP、JSP、HTML、XML,等等)。

WebLogic Portlet 有五种类型:JSP/HTML Portlet; Java Portlet; Page Flow Portlet; Struts Portlet; Web Service Portlet。在本系统中,为了实现通用的页面控制,我们选用 Struts 作为 Portal 的开发框架,运用 Struts 的 MVC 框架来实现表现层的管理和调度。当创建 WebLogic Portlet,我们选择 Struts Portlet, Struts Portlet 是经典的 MVC(模型-视图-控制器)模型的 Web 应用。利用控制器来分离模型和视图,达到一种层间松散耦合的效果,这样,应用程序功能划分清晰,开发人员可以分别集中开发业务逻辑模块、程序控制模块和数据显示模块,而且,这还能提高系统灵活性、复用性、可移植性、易扩展性和可维护性。

省公司的 Portal 结构就是这样:Portal→多个 Portlet→多个 Struts Action→多个 JSP 页面。如图 7 所示。

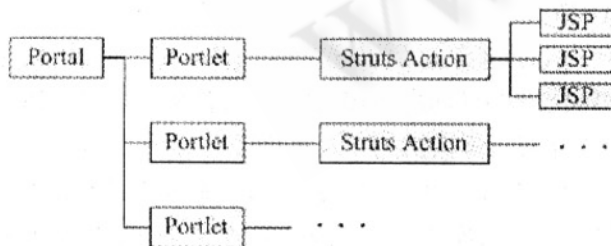


图 7 Portal 引用 Struts 结构模式图

在这样的结构模式下 Portlet 还采用直接面对 JSR168 Portlet 的 MVC 框架开发。Portlet 可以在任意符

合 JSR168 规范的 Portlet 容器间轻松的移植,这对移植性要求较高的系统是非常有价值的。

3.6 支持 WSRP

BEA WebLogic Portal 支持 Web Services for Remote Portlets 协议(WSRP),我们在创建 Portlet 的时候,就已经把新建的 Portlet 信息映射到 wsrp-producer-config.xml 文件中,这样,其他门户系统可以使用远程 Portlet 访问生产门户的本地 Portlet。将来湖北省电力公司做全省电力 EIP 系统时,生产门户系统可以作为生产者把 Portlets 直接提供给作为消费者的全省 EIP 系统。这样可以减少重复开发,资源可以再次利用。

但是在生成支持 WSRP 的 Portlets 要注意几点:① 用户请求和会话过程中,Portlet 之间不要依赖 Request 属性。② Portlet 使用 SSL 提供安全性。③ 在 Portlet 中创建链接时,用 WebLogic Portal 提供的 Tag 和 API 去创建 URL。

4 结束语

本文以湖北电网生产门户系统为案例,详细地介绍了利用 Oracle 高级复制技术把各个分布式数据源数据汇总到数据中心,在数据中心采用视图、触发器等 SQL 操作,形成查询的虚拟数据,并通过企业信息门户技术发布出来,满足用户的业务分布处理,数据集中查询、分析和决策的业务需要。

与一般数据集成的企业门户系统相比,具有如下特点:① 区别对待数据的复制,一方面保证了数据的时效性,另一方面也提高了系统的性能。② 采用多 Oracle 用户的数据复制,数据一致性得到了保证,并且配置和维护简单。③ 采用了 MVC 框架设计,数据模型、用户视图和控制逻辑分开。模型、用户交互和视图等部分都可以设计为可插入式组件,使得程序设计的过程更清晰,提高了可复用程度。④ 采用了 Struts 结构,具有较强的移植性。可以支持不同的 Portal 平台。⑤ 采用 WSRP 标准,支持 Portlets 复用和跨平台使用。但是,由于把大量的业务表复制过来,本系统省局数据库有一定的冗余数据。

现在,湖北电网生产门户系统已经完全投入运行,达到了预期设计目的和湖北电网生产的需求。实践证明,这种方式具有很好的安全性、实用性、复用性、可移植性和可维护性。
(下转第 5 页)

参考文献

- 1 盖九宇,张忠能,肖鹤. 分布式数据库数据复制技术的分析与应用. 计算机应用与软件,2005,(2).
- 2 武小平,胡启平. 基于 ORDB 的分布式空间数据异步更新模型研究. 计算机应用研究,2003,(4).
- 3 李军茹. Oracle 分布式系统数据复制技术. 程序员: CSDN 开发高手,2004,(11).
- 4 何一伟,陈勇,王文胜. ORACLE 的分布式数据更新机制的研究. 现代电子工程,2004,(3).
- 5 谷震离. 关系数据库查询优化方法研究. 微计算机信息,2006,(5):162-164.
- 6 甘邯,刘振元,陈学广,吕灏. 基于 Portlet 的大型水电工程业主企业信息门户. 华中科技大学学报:城市科学版,2005,(3).
- 7 胡一栋,陈福生,朱炜. 基于 Struts 框架的 Portal 研究和应用. 计算机工程与科学,2006,(8).
- 8 陈毓林,许舒人,宋靖宇,万淑超. 一个 Portlet 协作框架的分析与设计. 计算机工程,2006,(11).
- 9 施美雅,罗笑. 基于 JSR168 的 Portal Server 容器开发. 微型电脑应用,2005,(8).
- 10 徐碧云,王志坚,张少柏. 企业信息门户关键技术研究. 计算机应用研究,2005,(6).