论分布式存储系统架构设计

2016 年 6 月，我所在的公司成功中标某省环保厅的"污染源智能监控平台"建设项目。我作为项目的系统架构师和技术负责人，主要完成了需求的分析，系统的架构设计等工作。该项目是基于互联网，对全省各类污染源（水、气、声）进行实时监控并可视化显在电子地图上进行分析、预警。由于系统覆盖范围广（全省污染源企业），数据量大（污染源实时数据），所以在数据库设计方面，我们采用了分布式数据库设计，在省环保厅信息中心机房设置中心（全局）数据库，全省数据都要汇总到此数据库中，各市环保局设置数据库服务器，只存储本市辖区企业信息和污染源数据，通过 Oracle Stream 实现各市数据库与省厅全局数据库的同步。我从数据库选型，分布式数据库设计，应用数据集成和测试，以及分布式数据库部署等方面开展工作，经过 8 个月的努力，平台顺利开发完成并上线，通过近一年的用户反馈，整个系统运行稳定，达到了预期的目标和要求。

随着工业的飞速发展，工业排放的废水、废气、城市流域水环境以及噪声、废弃物对环境的污染日益严重，而相关监测和信息管理技术仍停留在早期的模式，环境监测数据还需要人工操作设备进行样本采集，由于环境污染中污染源数据总是不断变化，环境管理机构对环境污染数据掌控的实时性要求无法达到；各监测站和分局统计上报的数据也存在格式和标准不一致问题；环境数据是空间数据，具有强烈的空间性，而这些数据库都不能很好地表达空间数据，空间分析的功能也极其有限。"污染源智能监控平台"设置数据 GIS 展示功能，将污染源各类信息可视化显示在电子地图上，通过分析信息的空间分布，监测不同时段的污染数据变化，实现对空间信息及其他各类信息的有效管理，使大量抽象、枯燥的数据变得更加直观和易于理解，对污染源进行切实有效的管理，为政府有关职能部门对污染源管理提供科学依据。系统包括：实时视频监控模块、在线数据查询模块、数据 GIS 展示模块、超标企业预警模块、统计分析模块、手机端等。

鉴于该项目业务复杂、流程多，系统要涉及污染企业、省市区三级环保部门，项目工期短等特点，既要按期上线运行，又要实现基于分布式数据库的系统建设，我从数据库选型，数据库设计，应用数据集成和测试，以及分布式数据库部署等方面开展如下工作。

一、数据库选型

现在的主流数据库都可以按分布式进行部署，例如 SQL Server，Oracle，DB2 等。由于以前给环保厅做的项目的数据库平台都是选择的 Oracle，经过大量的实践检验，其多用户并发处理、性能、数据自动备份等都表现良好，我们公司有大量的基于 Oracle 数据库平台的分布式应用系统开发经验，所以经过我们与建设方的沟通和内部讨论，我们决定在该项目中继续采用 oracle 数据库来实现。在空间数据引擎方面，ArcSDE 由于其支持超大数据集，对海量数据可以进行高效的管理。故系统选用使用最广泛的空间数据库引擎 ArcSDE。　　二、

二、分布式数据库设计

由于涉及的污染企业众多，数据库采用分布式设计，各市环保局信息中心设置数据库服务器，省环保厅信息中心设置全局数据库服务器，并对数据进行完整性和一致性检查。即方便各市环保局管理本辖区污染企业，查询相关数据，又可以进行冗余备份，提高了系统可靠性、可用性，使系统易于扩充，提高了局部应用的效率，减少数据的网络传输。数据库分布在多个城市，为了保证数据的一致性及完整性，通过 Oracle Stream 实现各市数据库与省厅全局数据库的同步，Oracle Stream 利用高级队列技术，通过解析归档日志，将归档日志解析成 DDL 及 DML 语句，实现数据库之间的同步。通过使用 Stream 的技术，对归档日志的挖掘，可以在对主系统没有任何压力的情况下，实现数据表或数据对象的同步。

三、应用数据集成和测试

由于系统需要用到污染企业基础数据和其它应用系统支持，如水源地管理系统等，以支持本系统的审批、预警和统计分析。为了保证相关数据表的完整性和一致性，我们采用数据库的二阶段提交（2 phase commit），由于 Oracle 是一个非常成熟的数据库平台，又由于每次事务都不大，所以运行到现在没有没有出现过数据不一致问题。我们还开发一个同步程序，由省中心数据库服务器发起，对省厅与各市数据库进行比对，该程序可按需要执行，也可以定期执行。由于 Oracle 的出色表现和我们在其它项目中使用过分布式数据库设计，架构设计相当成熟稳定，所以测试的过程中我们选择了一个市作为试点，进行了完整的流程和数据测试，包括企业在线污染源监测数据的获取，数据同步等。测试通过调试没有问题后，在全省进行了系统的上线部署和完整测试。

四、分布式数据库部署

为了提高系统的可用性和负载均衡能力，省厅全局数据库采用 Oracle RAC，形成一个具有最高可用性（RAC+Data Guard）、安全性（数据安全）的整体解决方案。RAC 数据库由多个服务器节点组成，每个服务节点上面都有自己独立的 OS、ClusterWare、Oracle RAC 数据库程序等，每个节点都有自己的网络监听器。ClusterWare 是一个集群软件，主要用于集群系统管理，Oracle RAC 数据库程序用于提供 Oracle 实例进程，以提供客户端访问集群系统，监听服务主要用于监控自己网络端口的信息，，所有的服务和程序提供操作系统都去访问一个共享存储，最终完成数据的读写。对于各市局数据库服务器，采用主从分离技术，主数据库处理事务性增、改、删操作（INSERT、UPDATE、DELETE）操作，而从数据库处理 SELECT 查询操作。数据库复制把事务性操作导致的变更同步到其他从数据库。将读操作和写操作分离到不同的数据库上，避免主服务器进行写操作时的性能瓶颈，将不影响查询应用服务器的查询性能，提高并发。而且数据拥有多个容灾副本，提高数据安全性，同时当主服务器故障时，可立即切换到其他服务器，提高系统可用性、可靠性。

分布式数据库系统的安全性，也必须引起我们的重视。分布式数据库系统本身具有一套保证自身安全的身份认证机制，但要实现数据在网络环境下的传输安全，还需要采取一些其它的措施手段。我们采用的是远程数据连接建立在 VPN 基础之上，VPN 通过隧道、加密和认证等技术，在公共网络上建立一个虚拟专用网，由于各市的地理位置分散，网络状况差别很大，各市数据库服务器部署在 NAT 转换的私网内通过公用网络互联。综合考虑各因素，在 UNIX 上构建基于 PPTP 协议的 VPN 网络环境，借助防火墙技术，加强 VPN 网络安全，VPN 服务器部署在省厅内网，在防火墙上 VPN 服务器内网 IP 地址映射成公网 IP 地址，有效保证了数据在传输过程的安全。

2017 年 2 月，整个项目顺利完成，通过近一年的用户反馈，整个系统运行稳定，达到了预期的目标和要求，受到了用户的好评。但也存在些不足，比如在与其它系统交互数据时，因为码表不一致的原因，导致部分数据同步失败，后来我们及时更新数据同步程序，解决了这个问题。通过这个项目，我也学习到了不少知识增长了数据库架构设计经验。

三、总结

2017年2月，整个项目顺利完成，通过近一年的用户反馈，整个系统运行稳定，达到了预期的目标和要求，受到了用户的好评。但也存在些不足，比如在与其它系统交互数据时，因为码表不一致的原因，导致部分数据同步失败，后来我们及时更新数据同步程序，解决了这个问题。通过这个项目，我也学习到了不少知识，增长了数据库架构设计经验。

市的地理位置分散，网络状况差别很大，各市数据库服务器部署在NAT转换的私网内通过公用网络互联。综合考虑各因素，在UNIX上构建基于PPTP协议的VPN网络环境，借助防火墙技术，加强VPN网络安全，VPN服务器部署在省厅内网，在防火墙上VPN服务器内网IP地址映射成公网IP地址，有效保证了数据在传输过程的安全。

服务器进行写操作时的性能瓶颈，将不影响查询应用服务器的查询性能，提高并发。而且数据拥有多个容灾副本，提高数据安全性，同时当主服务器故障时，可立即切换到其他服务器，提高系统可用性、可靠性。

分布式数据库系统的安全性，也必须引起我们的重视。分布式数据库系统本身具有一套保证自身安全的身份认证机制，但要实现数据在网络环境下的传输安全，还需要采取一些其它的措施手段。我们采用的是远程数据连接建立在VPN基础之上，VPN通过隧道、加密和认证等技术，在公共网络上建立一个虚拟专用网，由于各

数据库程序等，每个节点都有自己的网络监听器。ClusterWare是一个集群软件，主要用于集群系统管理，Oracle RAC数据库程序用于提供Oracle实例进程，以提供客户端访问集群系统，监听服务主要用于监控自己网络端口的信息，所有的服务和程序提供操作系统都去访问一个共享存储，最终完成数据的读写。对于各市局数据库服务器，采用主从分离技术，主数据库处理事务性增、改、删操作（INSERT、UPDATE、DELETE）操作，而从数据库处理SELECT查询操作。数据库复制把事务性操作导致的变更同步到其他从数据库。将读操作和写操作分离到不同的数据库上，避免主

分布式数据库设计，架构设计相当成熟稳定，所以测试的过程中，我们选择了一个市作为试点，进行了完整的流程和数据测试，包括企业在线污染源监测数据的获取，数据同步等。测试通过调试没有问题后，在全省进行了系统的上线部署和完整测试。

四、分布式数据库部署

为了提高系统的可用性和负载均衡能力，省厅全局数据库采用Oracle RAC，形成一个具有最高可用性(RAC+Data Guard)、安全性(数据安全)的整体解决方案。RAC数据库由多个服务器节点组成，每个服务节点上面都有自己独立的OS、ClusterWare、Oracle RAC

三、应用数据集成和测试

由于系统需要用到污染企业基础数据和其它应用系统支持，如水源地管理系统等，以支持本系统的审批、预警和统计分析。为了保证相关数据表的完整性和一致性，我们采用数据库的二阶段提交（2 phase commit），由于Oracle是一个非常成熟的数据库平台，又由于每次事务都不大，所以运行到现在没有没有出现过数据不一致问题。我们还开发一个同步程序，由省中心数据库服务器发起，对省厅与各市数据库进行比对，该程序可按需要执行，也可以定期执行。由于Oracler 的出色表现和我们在其它项目中使用过