

大数据时代电网结构化数据中心架构研究

在国家电网公司总体建设集中式数据中心的背景下,网省电力公司应迎合大数据时代的信息产业潮流趋势,参考IT领域内成功案例并运用成熟技术,对现有结构化数据中心整体架构进行调整,达到对公司数据资源和基础设施资源的集约化管理的目的。本文论述了网省公司现有结构化数据中心的架构缺陷与功能缺失和进行架构调整的必要性及可行性,同时介绍了一种结构化数据中心新架构模型,对其在如何提高网省电力公司整体资源利用率、系统运行可靠性和大数据量并行计算能力等关键指标的问题上进行了分析。

吴凯 倪家明/国网天津市电力公司信息通信公司

0 引言



吴凯/助理工程师

国家电网公司信息化建设在全世界IT领域具有代表性和自身的独特性,其整体信息化发展趋势与成效也倍受业界关注和研究。随着公司信息系统一级部署进程不断深入和公共数据资源池工作的提前开展,国家电网公司在数据规模与数据内容上对结构化数据中心进行了扩充与调整。集中式数据中心建设推动了网省公司信息系统一级部署进程,将原来部署在网省公司的如ERP系统和人资管控等大型系统剥离网省公司的数据中心体系。探索出一种更为高效、更为稳定和更具扩展性的新型数据中心架构与配套的运维模式来解决诸如部分系统历史数据的集中处理与利用和资源如何高效利用等问题,已成为网省公司信息专业人员共同研究的热题与难点。

关键词/Keywords

大数据 ·
结构化数据 ·
数据中心 ·
国家电网 ·

1 网省公司结构化数据中心的现状分析

1.1 结构化数据中心概念不准

结构化数据中心,不应被单纯地看

作单纯存取数据的一个或多个数据库所组成的系统。业界认可的结构化数据中心是一个包括多种前端应用程序,具有数据传输与处理功能的成熟组件,搭配具有高可用性与高处理能力数据库集群的软硬件组合。目前网省公司对于结构化数据中心的概念依然停留在以数据操作存储(Operational Data Store, ODS)、数据仓库(Data Warehouse, DW)、数据集市(Data Mart, DM)和ETL数据操作工具等概念为组合的狭义数据中心,其概念与规模较小,与业界成熟的结构化数据中心有巨大的差距。而网省公司的结构化数据中心又是在这种认识下以“数据中心系统”的概念与其他信息系统并存。在大数据时代这种对于数据中心的认知与定位明显是落后的。

1.2 结构化数据中心数据利用情况不佳

各系统间进行数据共享,多以一方原始数据直接推入“数据中心系统”,另一方直接读取的方式进行,整个共享过程显示出网省公司“数据中心系统”不具有结构化数据中心应有的按需分配与筛选计算的功能,而更像是一块仅供存取的硬盘。

网省公司结构化数据中心所存储的业务数据,由于业务能力和工作需求不明确,又存在着误存或存取过量的情况,导致整个数据共享过程数据准确性和数据利用率均出现不同程度的问题。

1.3 数据共享效率不高

目前网省公司各信息系统数据实现数据共享需要相关专业部室进行沟通协调与审批。一项数据共享工作,从需求方申请到提供方审批,再到信息管理部门批准开通安全策略,少则几天,多则

几周。这种低效的数据管理，无形中催生了保守的“数据壁垒”，这与国家电网公司一直倡导的“推进数据共享融合，消除数据孤岛”的发展方向相悖，并严重拖后了公司信息化建设与发展。

另外，由于长期未能得到充分的重视与利用，多数网省公司结构化数据中心的基础设施单薄、架构过于简单，同时由于数据库内数据杂乱和数据冗余程度较高造成了不必要的资源浪费。基于以上分析，可预见到未来网省公司将在支撑大型数据挖掘与分析工作中遇到规划困难与性能瓶颈。

2 研究新型结构化数据中心架构体系的必要性与可行性

2.1 必要性分析

结构化数据中心的新型架构需要弥补网省公司现有架构的缺陷和不足，其调整范围将覆盖网省公司所有相关业务系统和以数据仓库为主体的数据中心系统。

(1) 系统遗留数据需要集中存储和利用

人资类系统、财务类系统、ERP系统和规划类等系统遗留下来的业务数据，对于网省公司在查询、展示与分析公司历史业务状态时将发挥无法估量的作用。企业应尽可能利用数据中心自身的数据处理优势，延续企业历史数据的生命力并挖掘出对公司生产或管理上的价值。

(2) 网省公司需要提高硬件资源整体使用率

根据国家电网公司总部对网省公司系统硬件运行状态的调研结果不难发现，多数业务系统数据库CPU使用率在10%以下；系统应用服务器本地磁盘使用率也普遍偏低甚至不使用；由于估值保守而划分的存储资源的浪费情况更是普遍存在。网省公司必须进行信息专业整体性架构的优化与调整，实施配套的管理手段，才能根本解决基础设施资源浪费的问题。

(3) 弥补现有架构所导致的功能性缺陷

网省公司现有的结构化数据中心受现有的并行计算与负载均衡能力所限，无法支撑用电信息采集等高热数据的高频度数据共享，造成结构化数据中心应有的价值与作用大打折扣。

(4) 网省公司信息专业需要适应IT产业潮流
通过关注近几年的专业期刊与学术讨论会不难发现，越来越多的电力信息从业人员希望并已经尝试将大数据、虚拟化和资源池等先进的技术领域新理念与新技术运用到电力行业中，意图从基础的运行与维护转向挖掘与创造信息化中的价值。而创造一种新的数据中心体系架构并转变运维模式，是为其创造条件的必要手段。

2.2 可行性分析

国家电网公司开展集中式数据中心建设促使网省公司的信息化体系结构将更为精简，这为进行结构化数据中心架构优化与调整带来了机遇，同时信息产业高速发展所催生出的新技术使架构调整更具可行性。

(1) 网省信息运行压力已经得到缓解

国网网公司推进一级部署工作，大型系统被剥离网省公司数据中心体系，网省公司对重要系统的运维压力得到缓解，使其得以拥有充裕的人力和时间去研究与尝试如何调整整体信息系统架构，提升其运行可靠性和安全性，进一步接近全年零故障的运行目标。

(2) 信息专业人员越发关注高端领域和技术

更多的电力行业信息专业人员在考虑如何调整结构化数据中心现有架构，能够将信息运维的重心偏向挖掘公司数据的价值的同时，还能够实现对有限基础设施资源的高效利用，达到集约化管理与精益化管理的目的。

(3) 公司运营监控中心需要高效数据中心支撑

国家电网公司于2012年成立了公司运营监控中心，历经两年的建设与运行，运营监控中心希望通过公司结构化数据中心调整架构，从流程上减少工作成本，跳过专业部室对数据共享的阻碍，直接获取或产出有价值的信息供其使用。

(4) 业界已有成熟的技术与成功案例

网省公司可借鉴IT行业中如阿里巴巴和百度等行业巨头的成熟案例与技术运用方案，其具有高集成度、高处理性能、高可靠性和优异的负载均衡能力等优秀特质的数据中心架构，已成为行业内的成功典范，并成为企业数据中心架构的良

好样板。

3 基于ORACLE RAC多节点数据库集群的结构化数据中心架构

为使网省公司结构化数据中心调整后更为符合业界主流架构模型,需要借鉴业界成功案例与典型经验为指导,对整体结构化数据架构、整体硬件部署结构进行调整,即重新规划数据在数据库中的存储结构,同时对于不同规模的数据库进行硬件资源的重新按需分配。

3.1 对整体数据架构进行调整

对业务系统数据库进行拆分,将主数据和维度数据等业务数据涉及的用户及实体与应用程序使用的支撑数据分离;于此同时,全面梳理数据仓库和ODS中现已存储的共享数据,进行数据模型去重和数据接口去重,按业务系统种类重新规划表空间的大小、命名规则等属性。

将业务系统原数据库中支撑应用程序的数据以保留或迁移的方式进行独立部署,而分解后得到的业务数据与一级部署后的系统遗留的业务数据则全部迁移至结构化数据中心多节点核心数据库表空间中,期间使用ETL数据转换技术,在统一异构数据库字符集依然能够保证数据一致性;同时业务系统进行应用程序或中间件数据源重新配置,保证业务系统的系统运行和业务功能均正常。

3.2 对硬件资源进行重新按需分配

对于数据总量较小并且数据增长缓慢的系统支撑库,为其分配如类似2路PC服务器的中低端服务器进行单独部署,数据文件使用本地硬盘支撑。

对于结构化数据中心的核数据库集群,如ODS和DW,则使用高性能的8路PC服务器或中高端小型机搭建多节点集群,每个节点的SGA分配量相比单库运行增加30%,SGA中的large pool、share pool和buffer cache三个参数内存分配量相比单库运行状态下均增加20%。

需要特别指出的是,出于RAC技术对操作系统苛刻要求,结合国家电网公司硬件国产化和未来可扩展性的要求,使用基于X86架构的PC服务器更具有优势。同时数据库操作系统采用兼容X86架构的红帽linux5或6版本的操作系统,其在电力信息行业近5年的运行中已表现出微软系列操作系统无法比拟的稳定性和可扩展性。

随着数据的迁移原业务系统数据库所挂的外接存储也必须被一并集中回收,并参考“云存储”模式供核心服务器所承载的数据库使用。

基于对数据架构、硬件资源分配和核心数据库结构的调整与优化,最终实现如下图所示的结构化数据中心的基本结构:系统支撑库独立部署,支撑应用服务运行;应用服务器产生的业务数据,集中在多节点核心数据库中实时存取;而

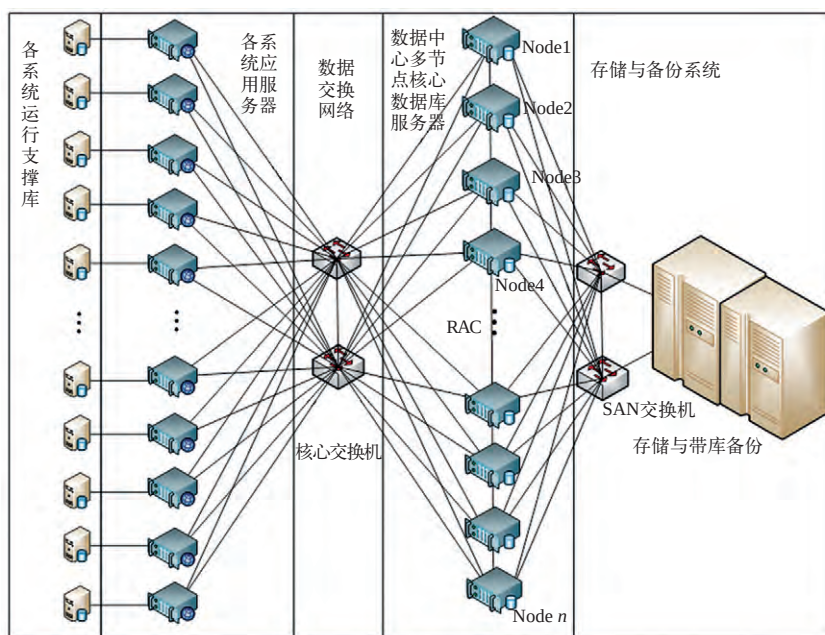


图 基于ORACLE RAC多节点数据库架构的结构化数据中心架构

多节点的核心数据库，发挥其突出的稳定性和并行计算能力，实现对业务系统的强力支撑。

4 新型架构优势分析

通过了解以上的新型结构化数据中心架构分析，不难发现，新架构首先解决了信息系统运行稳定性的问题。在目前各网省公司以保证业务系统正常运行为首要目标的前提下，对系统应用支撑库或资料库进行单独部署，使用独占资源支撑系统正常运行，减小了传统架构模式中由于应用程序对业务数据的高频存取而导致的数据库I/O性能降低进而对系统运行产生的不良影响。同时系统支撑库单独部署减小了系统复杂度，有利于网省公司提高系统检修效率和故障处理效率。

1) 新型架构模式有利于存储设备集中服务于单一数据库，抛弃了现有模式下以“库级”划分存储的模式，可实现信息专业人员以“表空间级别”来管理各业务数据增需，避免了现阶段普遍存在的由于保守估算而造成的存储资源的资源浪费，实现了对存储资源的精益化管理，提高资源使用率。

2) 借鉴淘宝网“20节点ORACLE数据库RAC集群”的成功经验与技术要点，新型架构中多节点ORACLE数据库集群具有极强的负载均衡能力，结构化数据中心核心数据资源的运行稳定性、可靠性和容灾能力得以大幅度提升。面对量级固定的网省公司信息业务量，多节点数据库集群的并行计算能力大幅提高了高端服务器资源的使用效率。

3) 新型架构要求全业务数据集中部署，有利于网省公司信息人员分析和统一业务数据模型，削弱各种因素造成的“数据壁垒”的不良影响，提高数据共享效率，降低数据分析与挖掘的难度，有利于网省公司进行“准大数据”量级下的数据分析与研究。

4) 由于系统支撑库仅依靠低端服务器便可满足，而核心数据库服务器使用的是高端服务器部署，作为原有业务系统数据库的中高端服务器可进行回收利用。无论是对现有结构化数据中心进行安全加固，还是批量下架解决机房拥挤问题，亦或是投入虚拟化硬件资源池建设，网省公司在硬件资源的调配上的灵活度将大大提高。

5 结束语

以上论述了网省电力公司现有信息主体结构的功能缺陷与概念缺失和新型架构在提高网省电力公司整体资源利用率、系统运行可靠性和大数据量并行计算能力关键点上的优势。但是目前还有诸如，如何由业务量与服务器计算能力共同计算核心数据库服务器节点数，如何设计与非结构化数据中心的接口等具体问题，必须在实施新架构调整与优化同时进一步研究和讨论的。

参考文献

- [1] 解林超，石佳，王仲锋，等．大数据时代对传统数据中心的影响及思考[J]．中国新通信，2014(2):38-39．
- [2] 宋亚奇，周国亮，朱永利．智能电网大数据处理技术现状与挑战[J]．电网技术，2013(4)：927-935．
- [3] 淘宝网构建20节点的RAC系统满足数据仓库海量数据处理需求[E/OL]．<http://docin.com/p-612551607.html>．

(收稿日期：2015-03-30) EA

作者简介

吴凯（1989-），男，助理工程师，主要研究方向为电力信息化、企业级数据中心规划。

倪家明（1982-），男，工程师，主要研究方向为智能电网信息化、物联网信息集成。