

钻井数据中心建设与推广研究

◆ 冯 军

摘要: 本文对大庆钻探数据中心总体框架、部署结构、建设思路以及建设过程中的数据采集、数据标准建设、数据质量控制等核心问题进行了概要阐述。

关键词: 钻井数据中心; 架构; 数据标准; 数据质量

1. 引言

企业级数据中心承载着企业的核心计算、信息资源管理、信息资源服务等功能, 数据中心对企业可持续运营和发展, 以及提升企业核心竞争力的重要性日益加强。

中石油大庆钻探工程公司2008年启动了钻井工程技术数据中心建设。钻井工程技术数据中心建设是以支撑提升钻井工程技术管理水平和生产组织与决策的效率、降低钻井生产成本、持续钻井提速为目标, 从而提高钻井专业板块核心竞争能力。

本文对大庆钻探数据中心总体框架、部署结构、建设思路以及建设过程中的数据采集、数据标准建设、数据质量控制等核心问题进行了概要阐述, 可作为钻井专业信息化建设的借鉴。

2. 大庆钻探钻井数据中心建设

2.1 钻井信息化建设特点

石油钻井行业具有野外作业、施工地域广泛、施工环境复杂, 钻井、测井、固井等专项作业工作交叉, 地下不可预测因素经常导致施工工艺变化等特点, 从而形成了以钻井井型、井别、地区等不同而导致的多种钻井作业管理模式和施工工艺过程, 以及因不可预测因素引起的多种施工工艺变化处理流程。

钻井行业对信息化建设的影响具体体现在以下方面:

1) 施工地域广泛及施工环境复杂, 形成了传输网络链路形式的多样性以及传输性能的波动, 导致网络基础设施建设复杂, 要求软件系统对传输链路的变化具有更高的适应能力;

2) 钻井作业管理模式和施工工艺过程的多样性导致软件系统对钻井作业模式变化的适应能力要求高;

3) 施工工艺的变化导致软件系统对施工过程处理

的复杂度增高。

2.2 数据中心总体架构

2.2.1 数据中心构成

大庆钻探数据中心从构成上可分为数据中心组织管理系统和数据中心平台实体系统。

数据中心组织管理系统包括: 数据中心运营维护组织体系和安全与制度体系。

数据中心平台实体系统包括: 运营支撑平台、数据平台和应用平台三大平台。

数据中心逻辑架构如图1所示。

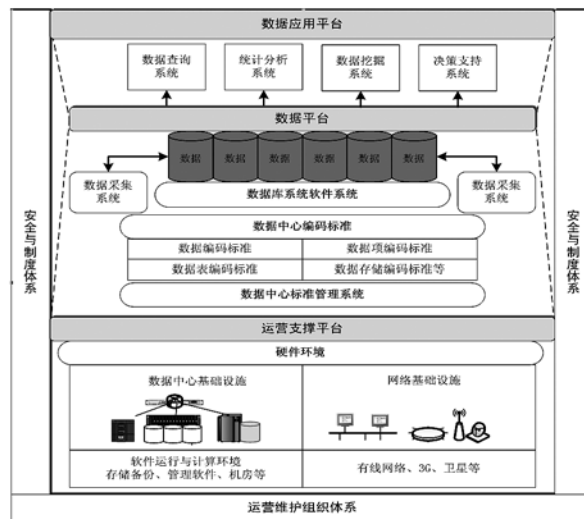


图1 企业数据中心逻辑架构

2.2.2 数据中心部署结构

数据中心部署可分为集中式部署和分散式部署, 考虑钻井行业特点、大庆钻探生产组织模式和组织机构地理分布及网络状况, 数据中心采用分散式部署(即主-分数据中心模式, 见图2所示)。

在各专业分公司建立分数据中心, 部署分公司的业务信息系统, 集成相关分公司的业务数据, 形成分公司层面的全局性数据应用。在大庆钻探层面建立主数据中心, 数据来源于分数据中心, 集成大庆钻探全局性数

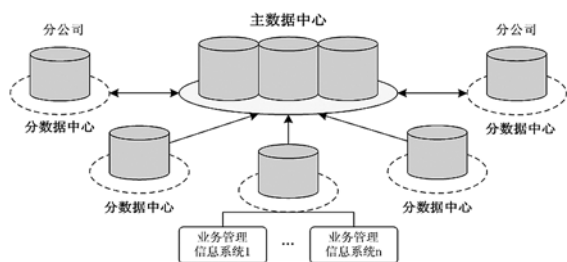


图2 主-分数据中心部署模式

据，形成数据集中化管理。

建立分数据中心，与主数据中心形成数据冗余，避免了数据中心集中式部署可能产生的风险，保证了数据安全性；分数据中心承载分公司的业务信息管理系统，避免了集中式部署可能产生的“单点故障”，保证了业务信息系统应用的时效性要求，也利于各分公司层面业务信息系统的差异化开发与数据应用；分数据中心的建立也有利于避免各分公司局域网络建设方面的差异而形成的系统应用问题。

建立主数据中心，一方面能够满足大庆钻探全局性数据应用的需求，另一方面能够承担大庆钻探层面各分公司之间的数据交换与共享枢纽的功能。

2.3 数据中心建设总体思路

在充分考虑了钻井行业特点、大庆钻探实际情况和业务需求的前提下，提出了数据中心建设的基本思路是：在数据中心总体规划指导下，以业务需求为核心，以“数据平台建设”和“数据应用平台”为建设主线，分步骤分阶段地建设“数据平台”和“数据应用平台”。在此过程中，以“数据平台”和“数据应用平台”的阶段性需求为驱动，逐步建立和完善“数据标准平台”和“运营支撑平台”，从而保证应用系统与运行环境的平衡发展。

数据中心总体建设思路如图3所示。

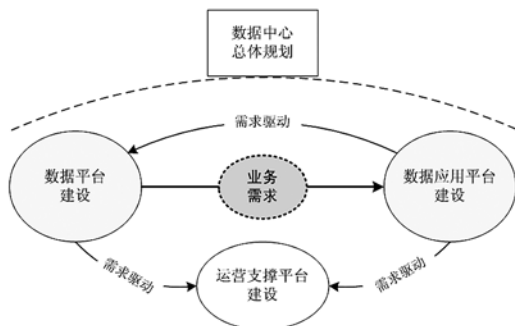


图3 数据中心建设思路

依据总体思路，大庆钻探数据中心建设一期建设是以“数据平台”建设为核心。

2.4 数据平台建设

数据平台建设的核心内容是数据规划及数据标准建设和数据采集系统建设，见图4所示。

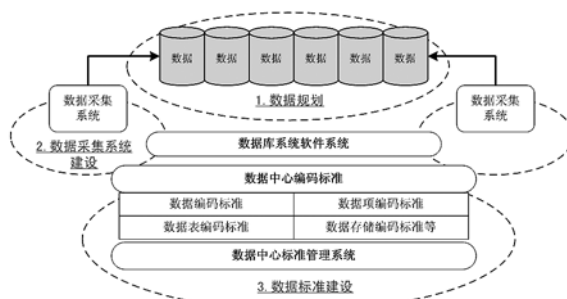


图4 数据平台建设内容

2.4.1 数据规划思路与内容

遵循信息资源管理（IRM）理论的基本原则，以面向业务主题为数据规划指导，分为三个阶段。

第一阶段，针对钻井工程职能域，确定核心流程，进行核心流程分析，形成业务模型、功能模型和用户视图。第二阶段，对用户视图进行分析，建立信息编码标准、数据分类体系和基础表结构，形成概念数据库模型。通过与用户讨论修正，形成最终概念数据库模型。第三阶段，按概念数据库模型在大庆钻探总部数据库服务器上建立钻井工程技术主题数据库。

2.4.2 数据标准建设

在数据标准建设方面，大庆钻探根据数据中心建设的需要，建立了大庆钻探企业级的数据分类体系和信息编码体系。

数据规划过程分为业务主线和信息主线两条主线展开。其中，业务主线规划是以大庆钻探当前的业务工作活动为核心，以实现各类用户视图为目标，满足大庆钻探信息处理的需要；信息主线规划是以大庆钻探合理的业务信息分类为核心，面向不同业务主题，建立各类业务主题数据库，满足日常业务管理、决策分析，以及数据资源的共享。

业务主线和信息主线的对应关系如图5所示。

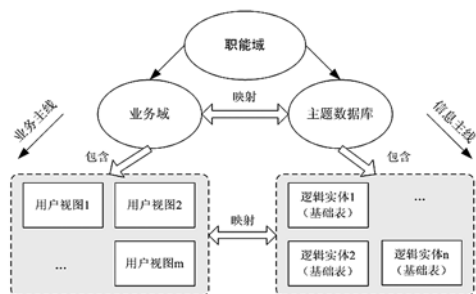


图5 业务主线和信息主线的对应关系

1) 信息分类体系

根据大庆钻探工程公司目前状况,业务主线的数据分类上述遵从信息资源规划理论体系要求,分为专业(板块)、职能域、业务过程、业务活动、用户视图、数据元素六个树状层次表示。

◆ 专业(板块):指工程公司的业务专业领域,如:钻井专业、物探专业等。

◆ 职能域:指专业板块下属的专业职能领域,如:钻井专业的工程技术、生产运行、物资管理、人力资源、财务管理等职能领域。

◆ 业务过程:指相关职能领域下属的某一业务过程,如钻井工程技术的单井设计过程。

◆ 业务活动:指业务过程涵盖的业务处理活动,如钻井工程技术的设计业务过程中的钻井工程设计活动。

◆ 用户视图:指业务活动涵盖的用户数据表单,包括单证、台帐(账册)、分析报表、统计报表等。

◆ 数据元素:指用户数据表(视图)中涵盖的数据项。

信息主线的信息分类遵从上述信息资源规划理论体系要求,分为专业(板块)、职能域、业务主题库、基础表、基础表数据元素、数据六个树状层次表示。

◆ 业务主题库:围绕企业核心业务主题而非业务过程建立的数据库,具有实体特征,如单井数据库。业务主题库由一系列基础表组成。

◆ 基础表:围绕业务主题库建立的存储数据表。基础表具有原子性、演绎性、规范性(满足3—NF)、稳定性等特征。

◆ 数据元素:基础表的数据元素是不可再划分的基础数据元素,如井号、钻压等。

2) 信息分类编码体系

信息分类编码体系标准必须同时满足业务主线和信息主线的需要

◆ 业务主线的编码对象确定围绕着用户视图展开;

◆ 信息主线编码对象围绕着主题数据库展开。

按照信息资源规划理论要求,结合信息资源规划总体框架,确定业务主线的编码对象和信息主线的编码对象。

围绕用户视图规划,大庆钻探业务主线编码对象包括:专业板块、职能域、业务过程、业务活动、用户视图、用户视图元素六类。

围绕主题数据库规划,大庆钻探信息主线编码对象包括:专业板块、职能域、主题数据库、基础表、基础

表数据元素、基础表数据七类。

其中:专业板块、职能域作为共同的编码对象处理,具有统一的编码规则。

信息分类编码的范围包括大庆钻探涉及的全部专业板块。制定专业板块和职能域的信息分类编码标准是信息分类编码实施的首要工作范围,下级次的信息分类编码实施根据大庆钻探工程公司的实际需要,确定当期的具体范围。

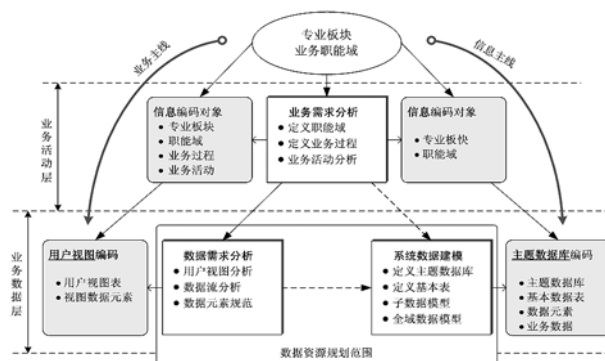


图6 信息码对象导出过程

大庆钻探的信息分类编码体系框架规定了系统信息分类编码层次及范围,编码对象及编码对象的编码范围、编码方式三方面及相互关系,见图7所示。

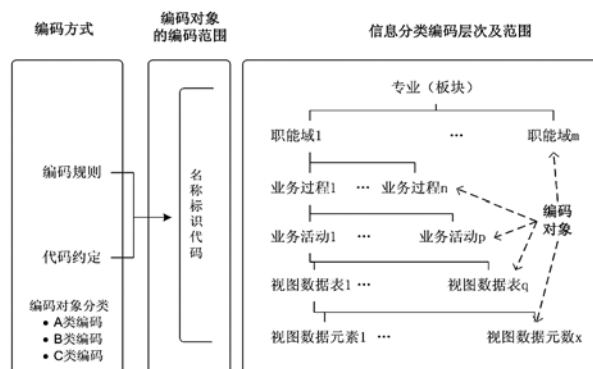


图7 业务主线信息分类编码体系框架

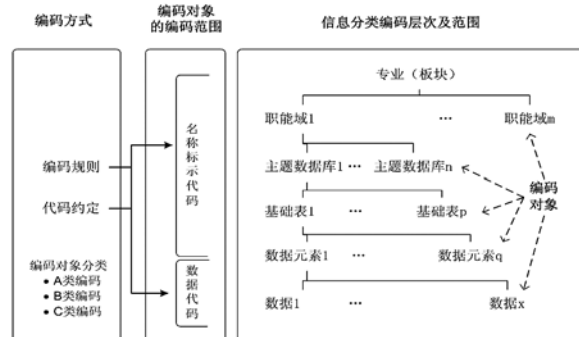


图8 信息主线的信息分类编码体系框架

在此信息编码体系基础上,定义了大庆钻探信息编码对象的编码规则和代码表。

2.5数据采集系统建设

数据采集系统是数据中心数据建设的核心和难点。数据质量控制是数据采集的关键。在数据采集系统建设上，大庆钻探遵循“业务需求驱动”的原则，即把业务系统建设作为数据中心数据采集的基础，通过满足业务岗位需求来完成数据中心的数据采集，实现工作中的业务过程和数据中心数据采集的“融合”。为此，在数据中心建设过程中，把各成员单位的业务信息系统建设也纳入到数据中心建设的统一规划之中。

2.5.1基于业务信息系统的数据采集

数据中心的数据采集采用与业务信息系统融合的方式。由于成员单位的业务系统数据有严格的业务流程控制和具体业务应用，数据的及时性和准确性有所保证，因此能够保证数据中心数据的准确性和及时性，从而保证数据中心的数据质量。数据采集模式如图9所示。

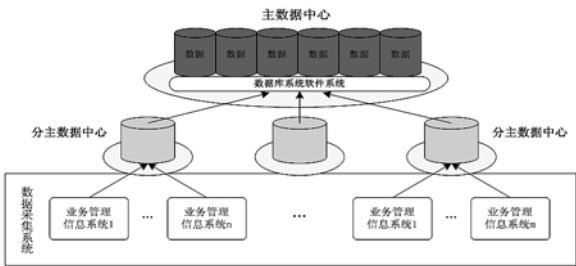


图9

2.5.2数据采集质量控制

1) 井基础数据的单入口控制

对于单井的核心基础数据（如井号、井位坐标、设计井深、井别等），通过建立单井基础数据管理系统，实现单井的核心基础数据的单入口采集，保证全系统单井核心基础数据的一致性。

2) 标准化数据的应用控制

数据中心数据标准化管理除了数据标准规范建立之外，另一项重要的工作是标准的应用控制，以保证数据中心数据统计分析的准确性。

大庆钻探数据中心数据标准的应用和控制采用分级分类的方式，按组织范围分为大庆钻探层面、分公司层面两个层面，按重要程度分为A（关键）、B（重要）、C（一般）三类数据项，如A类数据项包括井号、井别、区块等。

A类数据项进行大庆钻探层面的控制，实现大庆钻

数据类别 控制范围	A类 (如：井别、区块)	B类	C类
大庆钻探层面	√		
分公司层面		√	√

探范围的规范统一；B、C类数据项进行分公司层面的

控制，实现分公司层面的规范统一。

分类数据标准控制项由系统流程进行，即A类数据项的数据增、删、改通过系统流程由大庆钻探层面通过系统统一完成，B、C类数据项的数据增、删、改通过系统流程由分公司层面通过系统统一完成。

3) 数据的逻辑正确性控制

业务信息系统采集的数据质量控制通过建立逻辑规则库，进行数据合法性、数据之间逻辑关系（例如：实际井深≤设计井深+50）等合法性与正确性校验，实现页面级、数据实体表内、数据实体表间的数据校验。

4) 数据审核流程控制

由系统提供数据实际业务过程中的数据审核流程，进行人工数据审核，进行数据提交和驳回处理。

通过以上四个方面，实现数据采集过程中的全方位质量控制。

2.5.3数据采集系统的业务模式

由于石油钻井野外施工作业的特点，物理网络具有不确定性，为此数据采集系统需要支持多种业务应用模式。

数据中心数据采集系统支持的业务模式见下表所示。

子系统	数据采集系统					
	井基础数据管理子系统	施工设计子系统	井史管理子系统		钻井班报子系统	
应用模式	科室在线模式	科室在线模式	科室在线模式	在线-离线混合模式	钻井队离线模式	钻井队离线模式
系统功能	井基础数据采集功能	井位运行功能	全部功能	全部功能	部分功能	全部功能
使用用户	井基础数据采集功能	井位运行功能	全部功能	全部功能	部分功能	全部功能
各钻井分公司机关科室	地质部门	●				
	规划计划部门	●				
	生产运行部门	●				
	相关设计管理部门		●			
	相关技术资料管理部门			●	●	
小队 钻井	技术员				●	●
	场地工					
	泥浆工					

注释：●——为可选择系统功能分配点

3. 结论

大庆钻探目前已按照数据中心建设总体规划方案和建设思路，完成了数据平台建设，数据标准化建设，开发了钻井井史采集系统、钻井班报采集系统、井基础数据采集系统等数据中心采集系统，实现了3万多口井的钻井工程技术数据采集。通过标准化数据流程控制、数据逻辑校验控制、数据审核流程控制等数据质量控制机制，保证了数据中心数据的质量，为数据中心数据应用开发奠定了基础。通过开发多种业务应用模式的数据采集系统，很好地适应了各分公司的业务模式和应用需求，保证了系统的可实施性。

（作者单位：大庆钻探工程公司）