

文章编号: 1671-251X(2022)04-0031-08

DOI: [10.13272/j.issn.1671-251x.2022010053](https://doi.org/10.13272/j.issn.1671-251x.2022010053)

面向煤矿安全监控的数据仓库关键技术

刘海强^{1,2,3}, 陈晓晶¹, 张兴华¹, 陈向飞¹

(1. 中煤科工集团常州研究院有限公司, 江苏 常州 213015;

2. 金陵科技学院 计算机工程学院, 江苏 南京 211169;

3. 中国矿业大学 信息与控制工程学院, 江苏 徐州 221116)

摘要: 针对煤矿安全监控系统因采用操作型数据存储方法而导致无法有效利用海量数据, 且数据分析能力较差等问题, 研究了面向煤矿安全监控的数据仓库关键技术。根据煤矿安全监控业务需求, 提出了煤矿安全监控数据仓库的功能结构, 设计了超限分析、调校分析、异常数据分析、测点网络中断分析和人员管理分析五大业务主题。采用事实星座模型建立了煤矿安全监控数据仓库的逻辑模型, 分主题设计了事实表和维度表, 采用 SQL Server 建立了数据仓库物理模型。根据煤矿安全监控数据仓库特点, 提出了数据抽取、转换和加载策略, 采用不同的数据抽取规则分主题进行数据抽取, 对不同来源的数据进行格式转换、清洗和排序, 在数据加载过程中进行预加载、加载和加载后处理操作。

关键词: 煤矿安全监控; 数据仓库; 瓦斯超限分析; 调校分析; 异常数据分析; 测点网络中断分析; 人员管理分析; 数据抽取、转换、加载

中图分类号: TD67

文献标志码: A

Key technologies of data warehouse for coal mine safety monitoring

LIU Haiqiang^{1,2,3}, CHEN Xiaojing¹, ZHANG Xinghua¹, CHEN Xiangfei¹

(1. CCTEG Changzhou Research Institute, Changzhou 213015, China; 2. School of Computer Engineering, Jinling Institute of Technology, Nanjing 211169, China; 3. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: Due to the adoption of operational data storage method, the coal mine safety monitoring system can't use massive data effectively and the data analysis capability is poor. In order to solve the above problems, this paper proposes the key technologies of data warehouse for coal mine safety monitoring. According to the business requirements of coal mine safety monitoring, the functional structure of coal mine safety monitoring data warehouse is proposed. Moreover, the five business subjects are designed, including overrun analysis, calibration analysis, abnormal data analysis, measuring point network interruption analysis and personnel management analysis. The logical model of coal mine safety monitoring data warehouse is established by using the fact constellation model. The fact table and dimension table are designed by subject. The physical model of data warehouse is established by using SQL Server. According to the characteristics of coal mine safety monitoring data warehouse, data extraction, conversion and loading strategies are proposed. The different data extraction rules are used to extract data by subject. The data from different sources are processed through format conversion, cleaning and sorting. In the process of data loading, pre-loading, loading and post-processing operations are carried out.

收稿日期: 2022-01-25; 修回日期: 2022-04-11; 责任编辑: 李明。

基金项目: 江苏省博士后科研资助计划项目(2021K399C); 江苏省高等学校基础科学(自然科学)面上项目(21KJD440001); 金陵科技学院博士科研启动基金项目(jit-b-202128)。

作者简介: 刘海强(1986—), 男, 山东泰安人, 讲师, 博士, 研究方向为矿山物联网, E-mail: hqiu86@163.com。

引用格式: 刘海强, 陈晓晶, 张兴华, 等. 面向煤矿安全监控的数据仓库关键技术[J]. 工矿自动化, 2022, 48(4): 31-37, 113.

LIU Haiqiang, CHEN Xiaojing, ZHANG Xinghua, et al. Key technologies of data warehouse for coal mine safety monitoring[J].

Journal of Mine Automation, 2022, 48(4): 31-37, 113.



扫码移动阅读

Key words: coal mine safety monitoring; data warehouse; gas overrun analysis; calibration analysis; abnormal data analysis; measuring point network interruption analysis; personnel management analysis; data extraction, conversion and loading

0 引言

煤炭是关系我国国计民生的基础产业^[1]。煤矿生产中的不安全因素较多,环境恶劣,事故易发。近年来,随着开采深度进一步加大,煤矿安全隐患突出,安全事故尤其是重大、特大事故仍时有发生,安全形势依然严峻^[2]。

煤矿企业大多安装了安全监控系统,随着时间推移,存储了大量安全监控数据。现有煤矿安全监控系统大多采用操作型数据存储方式,主要面向事务处理,各种信息混合存储,与煤矿安全生产密切相关的重点信息不明确,导致信息无法有效利用。另外,系统大多只具备数据采集、存储和显示功能,数据分析能力较差,不能提供更好的决策支持。

数据仓库^[3]是大数据技术的重要组成部分,集成了各种形式的应用系统,以事物发展的角度组织数据。数据仓库是分析型数据存储方式,支持数据查询,能够为数据挖掘和决策分析提供良好的数据支持,逐渐成为大数据时代必然的数据管理技术,特别是基于 Hive 的数据仓库技术备受关注。

大数据和数据仓库技术被认为是实现智能矿山的关键技术,受到行业内研究人员的关注^[1,4]。张鹏^[5]提出了煤矿大数据架构,其中包括基于数据仓库的数据存储模型,但并没有研究实现数据仓库的具体方法。孟光伟^[6]建立了基于大数据技术的煤矿监管服务平台,提出使用 Spark 技术实现数据统计分析,使用 Hive 数据仓库分主题存储数据,但没有详细研究数据仓库的主题、逻辑模型等关键技术。张毅^[7]从企业管理的角度分析了煤矿数据仓库的优势,但与煤矿安全监控数据分析的差别较大。贾国兵^[8]将数据仓库应用于煤矿井下三维建模,开发了数据展示系统,但未研究数据仓库的实现技术。刘馨蕊^[9]研究了基于数据仓库的金属矿山信息集成系统,应用 Agent 方法建立了生产决策数据仓库,但使用的数据源主要是产量等数据,不适应安全监控分析需求。贾冬冬^[10]研究了基于数据仓库的冲击地压预警技术,从不同角度挖掘冲击地压数据特征。侯杰等^[11]研究了基于 OLAP(On-line Analytical Processing, 联机分析处理)技术的煤矿生产管理系统,但未研究数据仓库的建设方法。总体而言,数据仓库在矿业方面的应用还处于理论研究和实践摸索阶段,对需求

分析、主题设计和数据抽取方法等关键技术的研究较少。

本文研究面向煤矿安全监控的数据仓库关键技术,包括主题设计、模型设计、ETL(Extract–Transformation–Loading, 抽取、转换和加载)策略设计等,为煤矿安全监控数据挖掘和展示提供技术支撑。

1 数据仓库技术

数据仓库是面向主题的、集成的、相对稳定的、反映历史变化的数据集合。一方面,它可应用于决策支持,进行分析型数据处理;另一方面,它集成了多个不同结构的数据源,并根据主题进行重组和存储^[12]。

数据仓库具有以下 4 个特点:

(1) 面向主题。数据仓库中的数据是根据主题存储的,主题是用户进行分析和应用时关心的方面,多个操作型数据源的数据构成一个主题。

(2) 集成。数据仓库中的数据具有全局性和集成性,通过对分散的数据源中多种多样的数据进行预处理,建立统一的数据表达格式。

(3) 非易失。数据仓库中的数据主要面向分析查询,数据源中的数据经过抽取和清洗后进入数据仓库,被长期保留下来,很少修改和删除,只是定期加载和查询,不易丢失。

(4) 随时间变化。随着生产推进,生产过程中不断产生数据并被添加进数据仓库,因此,数据仓库是随时间变化的。汇总和分析这些数据能够预测某个问题的发展趋势。

数据仓库是一个多层次的体系结构,包括数据源层、数据存储与管理层、OLAP 与数据挖掘层、前端工具与应用层^[12],如图 1 所示。

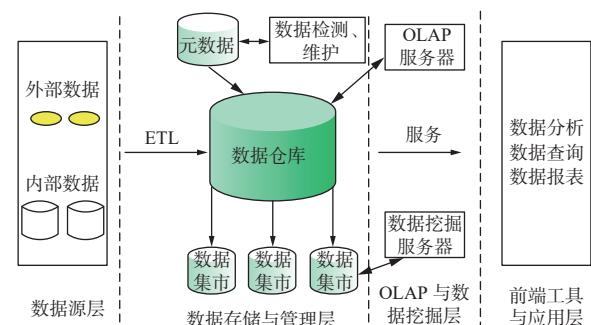


图 1 数据仓库体系结构

Fig. 1 Structure of data warehouse

(1) 数据源层。数据源包括内部数据和外部数据。内部数据是指关系型数据库中存储的数据和资料,本文主要是指在煤矿长期生产过程中,煤矿安全监控系统存储的安全监控数据、网络管理数据、配置管理数据和人员管理数据,这些数据可能存储于不同的部门、系统平台或物理位置。外部数据是与煤矿安全生产相关的法律法规和外部归类文档等。

(2) 数据存储与管理层。该层将从数据源层抽取的各类型数据进行转换和清洗,按照主题以确定的物理结构存储。数据仓库使用元数据描述数据及管理和维护数据所需的信息,根据元数据进行数据检测和维护。数据仓库根据不同部门的业务需求,建立特定的数据集市。数据集市之间互相配合,以满足用户的业务需求,组成一个完整的数据仓库。

(3) OLAP 与数据挖掘层。数据仓库使用 OLAP 进行数据挖掘。OLAP 提供切片、切块、钻取、旋转等数据分析方法,从多角度汇总和抽象数据,通过聚集查询技术提取用户所需信息。根据数据存储方式不同,OLAP 可分为关系型、多维型和混合型。关系型 OLAP 将细节数据和聚合后的数据均存储在关系型数据库中,数据量大时查询效率较低。多维型 OLAP 将细节数据和聚合后的数据均存储在多维数组中,查询效率较高,但生成多维数组所需时间较长。混合型 OLAP 将细节数据存储在关系型数据库中,将聚合后的数据存储在多维数组中,查询效率较关系型 OLAP 高,较多维型 OLAP 低,生成多维数组所需时间较关系型 OLAP 多,较多维型 OLAP 少。针对煤矿安全监控数据特点,本文采用混合型 OLAP 存储数据。

(4) 前端工具与应用层。该层应用可视化工具进行数据展示,为用户提供决策支持等业务功能,包括数据查询、数据分析、数据报表等。

2 煤矿安全监控数据仓库主题设计

数据仓库是面向主题的,其业务需求决定业务主题。针对煤矿安全监控的业务需求,设计了超限分析、调校分析、异常数据分析、测点网络中断分析和人员管理分析五大主题,见表 1。对应功能结构如图 2 所示。

2.1 超限及调校分析主题

超限是导致煤矿安全事故的重要原因,对超限数据的采集和分析是煤矿安全监控工作的重点。调校是重要的煤矿井下设备维护操作,目的是保障设备安全使用和监控数据准确。

煤矿安全监控系统记录了大量超限和调校信

表 1 煤矿安全监控数据仓库主题

Table 1 Subjects of coal mine safety monitoring data warehouse

主题	主题描述
超限分析	对超限测点配置、超限程度、超限原因、超限点的安全性等进行分析
调校分析	对调校测点配置、调校频率等进行分析
异常数据分析	对瓦斯浓度变化异常的时间、异常数据、异常原因、异常趋势等进行分析
测点网络中断分析	对测点配置及测点网络中断程度、原因等进行分析
人员管理分析	对人员工作职责和工作能力等进行分析

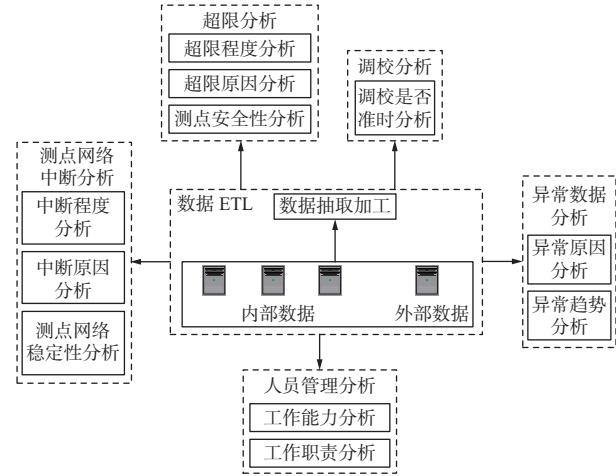


图 2 煤矿安全监控数据仓库功能结构

Fig. 2 Functional structure of coal mine safety monitoring data warehouse

息,二者混合存储在一起。在数据仓库建设过程中,有必要建立超限和调校分析主题,将二者提取出来单独存储,为数据挖掘提供基础。

超限分析主题可实现超限程度分析、超限原因分析和测点安全性分析。调校分析主题可实现调校是否准时分析。

2.2 异常数据分析主题

煤矿生产过程中,安全状态下瓦斯浓度变化一般是平稳的,且数值较小。当井下环境变化或设备出现故障时,瓦斯浓度监测值会出现异常,如监测值突然变大或变小等。这种异常变化可能是瓦斯突出的前兆,是煤矿安全监控工作需重点关注的信息。因此,有必要设计瓦斯浓度异常数据分析主题,根据瓦斯浓度变化规律,将瓦斯浓度异常数据提取出来,为分析异常原因、异常数据变化趋势、超限预警提供数据支持。

异常数据分析主题可实现异常原因分析和异常趋势分析。

2.3 测点网络中断分析主题

煤矿井下安装有大量传感器,采集数据通过通信网络传输至地面服务器进行存储和显示。如果通

信网络出现故障,则监控数据无法上传至地面,该情况下盲目生产容易引发安全事故。因此,有必要建立测点网络中断分析主题,提取测点网络断线的时间、持续时间、恢复时间及该测点的配置信息等进行主题存储。

测点网络中断分析主题可实现中断程度分析、中断原因分析和测点网络稳定性分析。

2.4 人员管理分析主题

煤矿安全监控系统中存储了大量的员工操作数据,包括对故障的处理、对安全事故的处理、对设备的检修和维护等。分析这些数据能发现员工工作过程中存在的问题,评估和考核员工的工作能力,以增强员工的责任感和紧迫感,提升企业的管理水平,从而保证安全生产,减少因人为因素引发的事故。因此,有必要建立人员管理分析主题,为煤矿企业管理提供决策依据。

人员管理分析主题可实现人员工作能力分析和工作职责分析。

3 煤矿安全监控数据仓库模型设计

3.1 逻辑模型

数据仓库逻辑模型设计即关系模型设计。数据仓库的常用设计模型包括雪花模型、星型模型和事实星座模型等^[13]。

根据煤矿安全监控数据仓库的特点,本文采用事实星座模型。在该模型中,每个主题设计 1 个事实表。事实表具有自身维度和共享维度,表之间通过共享维度交换信息。

事实星座模型根据主题设计数据表(包括事实表和维度表)的关系模式,把信息组织到结构中^[14]。煤矿安全监控数据仓库的逻辑模型如图 3—图 7 所示。

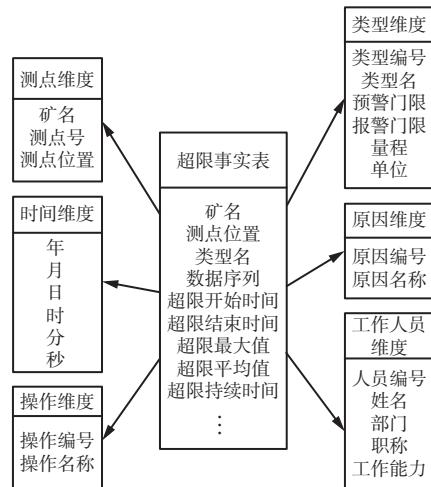


图 3 超限分析主题逻辑模型

Fig. 3 Logical model of overrun analysis subject

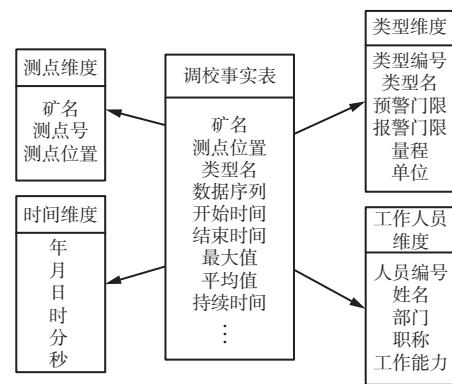


图 4 调校分析主题逻辑模型

Fig. 4 Logical model of calibration analysis subject

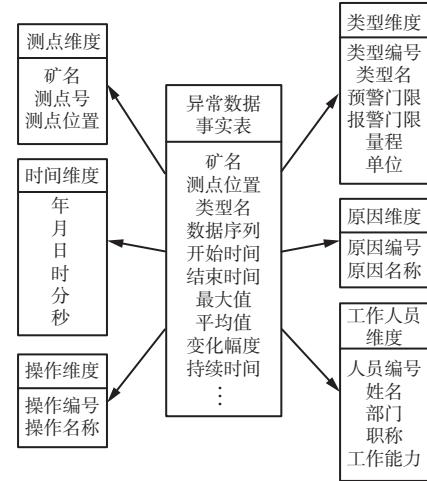


图 5 异常数据分析主题逻辑模型

Fig. 5 Logical model of abnormal data analysis subject

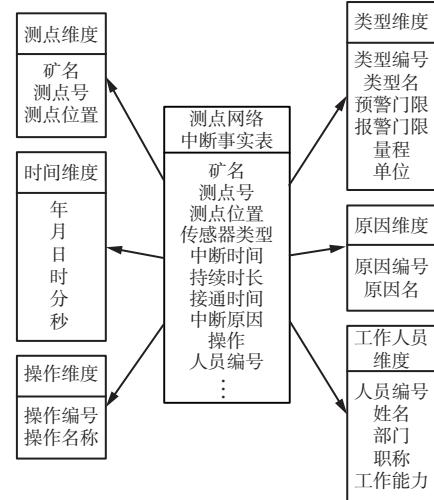


图 6 测点网络中断分析主题逻辑模型

Fig. 6 Logical model of analysis subject for measuring point network interrupt

为了描述细节信息,煤矿安全监控数据仓库将维度划分为:①类型维度,包括甲烷、CO、风速、温度、湿度等。②原因维度,为事故或故障原因,包括放炮、通风故障、线路问题等。③操作维度,为工作人员对故障所做的操作,包括断电、检修、记录等。

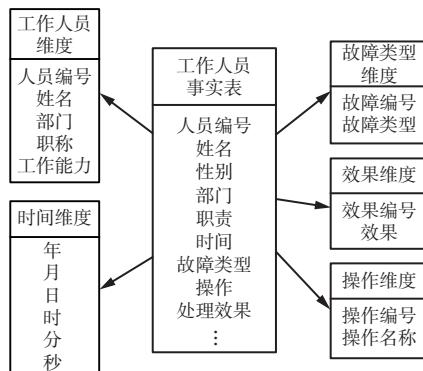


图 7 人员管理分析主题逻辑模型

Fig. 7 Logical model of personnel management analysis subject

④ 故障类型维度, 为各种故障, 包括超限、异常、上传网络中断等。⑤ 效果维度, 为工作人员对故障的处理效果, 包括优、良、中、差等。⑥ 测点维度, 为某位置的 1 个或多个测点。⑦ 时间维度, 为某具体时间点或时间段。⑧ 工作人员维度, 为某 1 位或几位员工的信息, 包括编号、部门、职称等。

根据用户权限和关心的细节设计煤矿安全监控数据仓库的粒度^[15]。时间粒度分为年、月、日、时、分、秒; 监管粒度分为集团公司、矿、工作面、测点等。

3.2 物理模型

物理模型是数据的存储结构, 对应逻辑模型中的主题维度^[16]。由逻辑模型可知, 煤矿安全监控数据仓库涉及的维度主要有时间、测点、类型、工作人员、原因、操作、故障类型、效果。

物理模型使用 SQL Server 建立表的主键和外键聚簇索引, 根据应用情况建立非聚簇索引, 以获取更强的查询性能。以超限分析主题为例, 其事实表和维度表见表 2—表 7。其他主题物理模型设计与此类似。

4 煤矿安全监控数据仓库 ETL 策略

煤矿安全监控数据仓库将不同数据源中的数据集中到数据仓库数据集, 需对数据进行 ETL 操作。煤矿安全监控数据仓库的数据来源、格式存在很大差异, 如在历史数据库中, 安全监控数据源分布于集团公司下属企业, 网络安全管理数据源位于网络管理部门。ETL 需最大程度地保证数据一致性, 提高数据质量。

ETL 程序与数据源通过网络连接, 使用 ADO (ActiveX Data Objects, ActiveX 数据对象) 和 ODBC (Open Database Connectivity, 开放数据库连接) 连接数据源和数据仓库^[14], 通过抽取程序完成数据抽取, 在程序处理模块完成转换和清洗, 然后将数据加载至数据仓库, 如图 8 所示。

表 2 超限分析主题事实表

Table 2 Fact table of overrun analysis subject

字段名称	字段描述	数据类型
MineName	矿名	Varchar(20)
TpName	测点号	Varchar(10)
Location	测点位置	Varchar(20)
Type	类型名	Varchar(10)
StartTime	超限开始时间	datetime
EndTime	超限结束时间	datetime
MaxValTime	超限最大值时间	datetime
MaxVal	超限最大值	real
LastTime	超限持续时间	real
AverageVal	超限平均值	real
Data	数据序列	Varchar(500)
Reason	超限原因	Varchar(500)
Staffnumber	值班人员编号	Real
Operation	操作	Varchar(500)

表 3 超限分析主题测点维度表

Table 3 Monitoring point dimension table of overrun analysis subject

字段名称	字段描述	数据类型
MineName	矿名	Varchar(20)
TpName	测点号	Varchar(10)
Location	测点位置	Varchar(20)

表 4 超限分析主题类型维度表

Table 4 Type dimension table of overrun analysis subject

字段名称	字段描述	数据类型
Typebh	类型编号	Int
Type	类型名	Varchar(10)
WarnGate	预警门限	real
AlarmGate	报警门限	real
MaxR	量程最大值	real
MinR	量程最小值	real
Units	单位	Varchar(10)

表 5 超限分析主题原因维度表

Table 5 Reason dimension table of overrun analysis subject

字段名称	字段描述	数据类型
Reasonbh	原因编号	Int
Reason	原因名称	Varchar(20)

4.1 数据抽取

数据抽取是从数据源中抽取数据, 通过一定规则判断是否为所需数据, 若是则添加进数据仓库。煤矿安全监控数据仓库按主题进行数据抽取, 流程如图 9 所示。

表 6 超限分析主题操作维度表

Table 6 Operation dimension table of overrun analysis subject

字段名称	字段描述	数据类型
Operationbh	操作编号	Int
Operation	操作名称	Varchar(20)

表 7 超限分析主题工作人员维度表

Table 7 Personnel dimension table of overrun analysis subject

字段名称	字段描述	数据类型
Peoplebh	人员编号	Int
Name	姓名	Varchar(10)
Department	部门	Varchar(10)
Title	职称	Varchar(20)
Capacity	工作能力	Varchar(50)

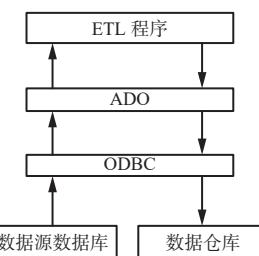


图 8 煤矿安全监控数据仓库 ETL 实现流程

Fig. 8 ETL realization flow of coal mine safety monitoring data warehouse

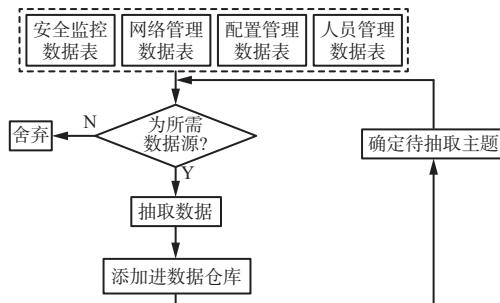


图 9 煤矿安全监控数据仓库数据抽取流程

Fig. 9 Data extraction flow of coal mine safety monitoring data warehouse

超限和调校数据主要来源于煤矿安全监控系统报警数据表。从该表中抽取这 2 种数据即是对这 2 种数据进行分离。根据调校时瓦斯浓度变化规律(本文为先降至 0, 再快速上升, 然后快速下降, 最大值约为 2.0×10^{-6} , 整个过程持续约 1 min), 得到调校数据分离标准为最大值 2.0×10^{-6} 且持续 1 min。

异常数据可根据瓦斯浓度监测值是否突然变化来判断, 对异常数据抽取需要对异常状态和数据是否属于同一异常状态进行判断。

对异常状态的判断主要考虑相邻数据间的变化斜率指标 k_v 。设瓦斯浓度变化时间序列为 $T =$

$\{t_1, t_2, \dots, t_n\}$, 数值序列为 $Y = \{y_1, y_2, \dots, y_n\}$, 其中 t_m, y_m 分别为第 m 个数据采样时刻及对应采样值, $m=1, 2, \dots, n, n$ 为采样点数, 则 $k_v = |y_m - y_{m-1}| / |t_m - t_{m-1}|$ 。设置斜率门限为 K , 当 $k_v > K$ 时, 则判定当前数据和前一数据为异常数据。 K 根据煤矿安全生产规定和煤矿具体安全生产情况来确定。

根据 2 个相邻异常数据的时间间隔判断数据是否属于同一异常状态。设异常数据时间序列为 $T' = \{t'_1, t'_2, \dots, t'_N\}$, 数据序列为 $Y' = \{y'_1, y'_2, \dots, y'_N\}$, t'_M, y'_M 分别为第 M 个异常数据采样时刻及对应采样值, $M=1, 2, \dots, N, N$ 为异常数据个数, 则相邻时间间隔 $\Delta T = t'_M - t'_{M-1}$ 。设属于同一异常的时间间隔门限为 T_s , 若 $\Delta T < T_s$, 则 y'_M 和 y'_{M-1} 属于同一异常状态, y'_M 和 y'_{M-1} 都计入本次异常状态, 否则 y'_{M-1} 计入本次异常状态, y'_M 计入下一次异常状态。

测点网络中断数据主要来源于煤矿安全监控系统的安全监控数据表和配置管理数据表。安全监控数据断线和接通时状态字段设置为特殊值, 可根据该字段值对断线数据进行抽取。

人员管理数据主要是工作人员的操作, 可抽取安全监控数据表、网络管理数据表、配置管理数据表和人员管理数据表中的操作信息和个人信息。人员管理以 1 个工作人员对应 1 个表的方式来组织, 将一段时间内该人员的操作添加进该表, 以便对该人员的工作情况和工作能力进行分析。

4.2 数据转换

煤矿安全监控数据源分散, 存储格式不同, 需整理且转换为相同格式才能进行统一存储, 该工作在数据转换阶段完成。煤矿安全监控数据仓库的数据转换过程包括格式转换、数据清洗和数据排序。

(1) 格式转换分为字段转换、数值转换和表示方法转换。字段转换是针对同一对象在不同数据源的数据表中使用不同字段表示的情况, 如测点类型字段在报警数据表中可能为 `Stype`, 在超限主题事实表中可能为 `Type`, 在数据转换过程中对这些字段进行统一表示。数值转换是针对不同系统中数据存储采用不同的单位或数据类型的情况, 在数据转换时进行单位和数据类型的统一, 如将瓦斯浓度数据从 `float` 型转换为 `real` 型。表示方法转换即统一数据的表示格式, 如将日期和时间格式统一为 `yyyy-mm-dd hh:mm:ss`。

(2) 数据清洗包括重复记录清洗和错误数据清洗。重复记录清洗是针对不同系统中存储的重复数据进行判断, 只抽取 1 条记录, 同时判断抽取的数据是否已存在于目标数据仓库中, 对重复记录舍弃。

错误数据清洗是针对抽取数据中存在的明显错误,如数据类型不对、数值为空、日期超过当前日期等进行弥补,无法弥补的则舍弃。错误数据弥补方式:①将其记录到特定文件,等待数据仓库管理人员处理。②赋一个缺省值,如一段时间内的平均值、最可能的预测值等。

(3) 数据排序是针对瓦斯超限、网络通断等有时间顺序的数据,存入数据仓库时根据时间戳进行数据排序。

4.3 数据加载

数据转换后需将其加载进数据仓库,过程包括预加载、加载和加载后处理。预加载是指从数据库中找到需添加、更新的记录及添加的目标地址,将其写入预处理文件,之后由加载程序加载到数据仓库中。加载是指调用批量加载工具添加和更新数据,并记录加载日志。加载后处理主要是删除不需要的冗余文件和临时表。

数据加载设计主要是数据追加方法的设计。数据追加包括直接追加、覆盖追加和更新追加。直接追加是直接将数据添加进数据表;覆盖追加是当新抽取的数据包含所有历史数据时,直接覆盖历史数据;更新追加是针对需要进行状态更新的数据,对历史状态数据和当前状态数据进行比较,更新数据后追加。

根据煤矿安全监控数据仓库的数据特点选择数据追加方式。配置管理数据和人员管理数据量少、变化小,可将之前的数据覆盖,可采用覆盖追加策略;安全监控数据和网络管理数据要记录连续的变化状态,需比较当前状态数据和历史状态数据,将更新的数据添加至数据仓库,因此采用更新追加策略。

5 结论

(1) 对面向煤矿安全监控的数据仓库进行了功能分析,根据煤矿安全监控业务需求,建立了超限分析、调校分析、异常数据分析、测点网络中断分析和人员管理分析五大业务主题。

(2) 设计了煤矿安全监控数据仓库的逻辑模型和物理模型。针对煤矿安全监控数据仓库的特点,使用事实星座模型建立了逻辑模型,分主题设计了事实表和维度表,设计了逻辑模型的维度和粒度。

(3) 设计了煤矿安全监控数据仓库的ETL策略。针对数据来源和业务需求,分主题进行数据抽取,设计了不同主题数据抽取的规则;对形式各异数据进行数据格式转换、清洗和数据排序;设计了数据

加载策略,分为预加载、加载和加载后处理3个阶段,在数据追加时根据数据特点,采用覆盖追加和更新追加结合的方式。

参考文献(References):

- [1] 王国法,王虹,任怀伟,等.智慧煤矿2025情景目标和发展路径[J].煤炭学报,2018,43(2):295-305.
WANG Guofa, WANG Hong, REN Huawei, et al. 2025 scenarios and development path of intelligent coal mine[J]. Journal of China Coal Society, 2018, 43(2): 295-305.
- [2] 新华社.国家矿山安监局:2020年全国煤矿无重特大瓦斯事故[EB/OL].(2021-01-08)[2022-01-03].http://www.xinhuanet.com/2021-01/08/c_1126961922.htm.
Xinhua News Agency. China State Administration of Work Safety: there is no heavy and large gas accident in China in 2020[EB/OL]. (2021-01-08) [2022-01-03]. <http://www.xinhuanet.com/2021-01/08/1126961922.htm>.
- [3] DZEMYDIEN D, MASKELIUNAS S, RADZEVIUS V. Approach of ensuring interoperability of multi-dimensional data warehouses for monitoring of water resources[J]. *Journal of Environmental Engineering and Landscape Management*, 2021, 29(1): 9-20.
- [4] 何敏.智慧矿山定义探讨[J].工矿自动化,2017,43(9):12-16.
HE Min. Discussion on definition of wisdom mine[J]. Industry and Mine Automation, 2017, 43(9): 12-16.
- [5] 张鹏.智能矿山大数据体系建设探索[J].工矿自动化,2021,47(增刊1):21-23, 44.
ZHANG Peng. Exploration on construction of big data system for intelligent mine[J]. Industry and Mine Automation, 2021, 47(S1): 21-23, 44.
- [6] 孟光伟.基于大数据技术的区域煤矿监管数据服务平台设计[J].工矿自动化,2021,47(10):97-102, 109.
MENG Guangwei. Design of a regional coal mine supervision data service platform based on big data technology[J]. Industry and Mine Automation, 2021, 47(10): 97-102, 109.
- [7] 张毅.基于数据仓库原理的煤矿精细化管理研究[D].徐州:中国矿业大学,2014.
ZHANG Yi. Study on coal mine management based on the principle of data warehouse[D]. Xuzhou: China University of Mining and Technology, 2014.
- [8] 贾国兵.基于数据仓库的煤矿井下真三维建模与综合应用系统研发[D].沈阳:东北大学,2015.
JIA Guobing. Development of real 3D modeling and integrated application system in coal mine downhole data based on the data warehouse[D]. Shenyang: Northeastern University, 2015.

(下转第113页)