

ASO-S卫星工程科学应用系统的数据库设计*

葛蕴翊[†] 黎辉[‡] 黄宇 李敬伟 雷仕俊 藏京京 苏杨
封莉 李友平 李瑛 徐遵磊 李翔

(中国科学院紫金山天文台暗物质和空间天文重点实验室 南京 210023)

摘要 先进天基太阳天文台(Advanced Space-based Solar Observatory, ASO-S)卫星是我国首颗太阳观测卫星,主要观测太阳耀斑和日冕物质抛射以及产生它们的磁场结构. ASO-S卫星的科学应用系统是科学卫星工程的6大系统之一,它连接科学用户和卫星数据,为将卫星的科学数据转化为科学成果提供保障. 科学应用系统的数据库是连接软件与海量数据的枢纽,为科学数据生产和用户服务及运行提供数据层的支撑. 介绍了科学应用系统的数据库架构设计、数据库的选择以及数据库性能优化和表样例. 这里的数据库包括观测计划、工程参数、运维日志、科学数据、定标数据和特征事件识别等数据库. 这些数据库的建设将为ASO-S卫星工程科学应用系统的顺利运行提供数据支撑,也可以为未来其他科学卫星类似数据库的搭建提供参考和借鉴.

关键词 先进天基太阳天文台, 天文数据库, 方法: 分析
中图分类号: P114; **文献标识码:** A

1 引言

先进天基太阳天文台(Advanced Space-based Solar Observatory, ASO-S)^[1]卫星是中国科学院空间科学战略性先导专项立项的我国首颗太阳观测卫星. ASO-S卫星以观测太阳上两类最剧烈的爆发现象—太阳耀斑和日冕物质抛射(CME)以及产生它们的磁场结构作为主要科学目标. ASO-S卫星配置了3台载荷用以开展对太阳上“一磁两暴”的观测,它们分别是:全日面矢量磁像仪(FMG)^[2]、莱曼阿尔法太阳望远镜(LST)^[3]和太阳硬X射线成像仪(HXI)^[4]. HXI和FMG都只有一台仪器,而LST有3台仪器,即莱曼阿尔法全日面成像仪(SDI)、莱曼阿尔法日冕仪(SCI)和白光太阳望远镜(WST). 其中,SCI有两个探测器,即SCI莱曼阿尔法波段(SCIUUV)探测器和SCI白光波段(SCIWL)探测器;SDI和WST各只有一个探测器.

2020-02-18收到原稿, 2020-05-20收到修改稿

*国家自然科学基金项目(U1731241、11503089、11973012)及中国科学院空间科学战略先导专项(XDA15052200、XDA15320103、XDA15320301)资助

[†]geyunyi@pmo.ac.cn

[‡]nj.lihui@pmo.ac.cn

ASO-S卫星工程的科学应用系统^[5]是卫星工程的6大系统之一. 它连接卫星数据和科学用户, 是将观测数据生产成高级数据产品, 并最终转化为科学成果的关键. 科学应用系统的数据库是连接科学应用系统的数据生产软件和数据或数据文件的枢纽, 为科学应用系统的数据生产和用户服务以及运行提供数据层的支撑.

ASO-S卫星工程科学应用系统的主要功能包括制定观测计划、接收和管理观测数据、生产高级数据产品以及为用户提供数据和分析软件的浏览和下载等. 需要数据库提供数据和文件索引信息服务的功能包括: 观测计划管理、载荷状态监测管理、系统运行管理、数据生产、数据分析、用户服务、特征事件识别等. 数据库的建立有助于科学应用系统的高效运行和管理.

文章的结构如下: 我们在第2节介绍ASO-S卫星工程科学应用系统的数据库架构设计; 在第3节介绍数据库的选择; 在第4节介绍数据库性能优化; 在第5节中介绍数据库的表样例; 在第6节进行总结与展望.

2 架构设计

ASO-S卫星工程科学应用系统软件和数据库的架构设计要符合高内聚低耦合的原则^[6], 即相关性较强的功能放在同一模块中, 不同模块之间依赖关系较低. 为此, 使用这样的架构设计^[7]: 表现层(User Interface)、业务逻辑层(Business Logic)、数据访问层(Data Access)、数据库(Database)、数据文件(Data File). 我们稍后描述各层的功能. 这里需要强调的是, 只有数据访问层能够对数据库进行各项操作. 这样的架构设计符合高内聚低耦合要求^[8]. 图1为科学应用系统软件和数据库架构设计图.

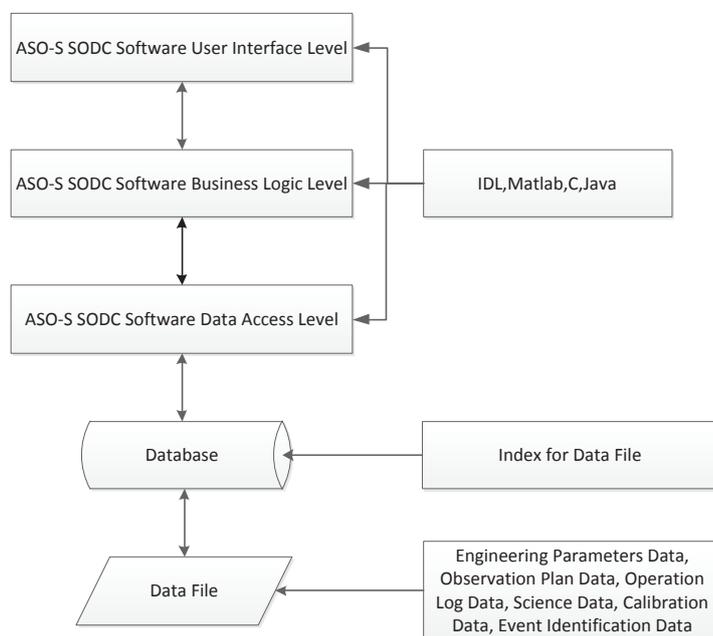


图 1 ASO-S科学应用系统软件和数据库架构设计图

Fig. 1 The software and database architecture design chart of the ASO-S science operation and data center

在科学应用系统软件和数据库架构设计中,表现层位于最上层,与用户直接交互,向用户输出数据和接收用户的数据输入.业务逻辑层在构架中位于表现层与数据层中间位置,它是表现层与数据层的桥梁.业务逻辑层接收到表现层的用户指令或数据后,按照既定的业务逻辑处理后,发送给数据访问层;将数据访问层返回的数据按照既定的业务逻辑处理完后,将结果反馈到表现层中给用户查看.数据访问层负责访问数据库,实现数据库的增加、删除、改写和查询等操作,并将操作结果反馈到业务逻辑层.该架构设计禁止进行跨层交互.每层只能和上一层以及下一层进行交互.表现层访问数据文件需要发送指令给业务逻辑层,业务逻辑层再发送指令给数据访问层,数据访问层去访问数据库从而定位数据文件,最终对数据文件进行访问.

科学应用系统软件和数据库使用这样的架构设计在开发效率、维护效率、数据安全性上具有明显优势.(1)项目结构清楚、分工明确,从而大大提高开发速度和效率;(2)在后期维护中,有利于各层代码的复用,提高维护效率,降低维护成本;(3)避免了表现层直接访问数据访问层,表现层只和业务逻辑层有联系,提高了数据安全性.

从业务角度考虑,系统的数据库^[9]划分为:观测计划数据库、工程参数数据库、运维日志数据库、科学数据数据库、定标数据数据库和特征事件识别数据库.其中观测计划管理功能使用观测计划数据库;载荷状态监测管理功能使用工程参数数据库;系统运行管理功能使用运维日志数据库;数据生产功能和数据分析功能使用科学数据数据库和定标数据数据库;特征事件识别功能使用特征识别数据库;用户服务功能使用科学数据数据库和特征事件识别数据库.图2为在对数据库进行增加、删除、改写和查询操作的业务架构图.

根据ASO-S卫星数据生产规划和流程,科学应用系统需要为ASO-S卫星上3个载荷的6个探测器不同级别的数据分别创建子库.考虑到LST的4个探测器的数据各自为独立的一组数据,每组数据又各自分为0级、Q0级、1级、1.5级、2级、3级、Q1级、E级,共需要32个子库.HXI数据分为0级、Q0级、0.5级、1级、1.5级、2级、Q1级、E级,需要8个子库.FMG数据分为0级、Q0级、0.5级、1级、1.5级、2级、Q1级、E级,也需要8个子库.这样科学数据数据库一共要创建48个子库,为数据生产功能和数据分析功能服务.

观测计划数据库为观测计划管理功能服务,包含长期观测计划、用户申请观测计划、定标观测计划以及计划执行结果等子库;工程参数数据库为载荷状态监测管理服务,包含遥测数据、数传工程参数和辅助数据等子库;运维日志数据库为系统运行管理功能服务,包含存储服务器状态、网站状态和自动化运行的任务状态等子库;定标数据数据库为数据生产功能和数据分析功能提供服务,包含3个载荷各种定标数据子库;特征事件识别数据库为特征事件识别功能服务,包含暗条、CME、日珥、耀斑、磁中性线以及X射线源等子库;此外,用户服务还需要科学数据数据库和特征事件识别数据库提供相应支撑.

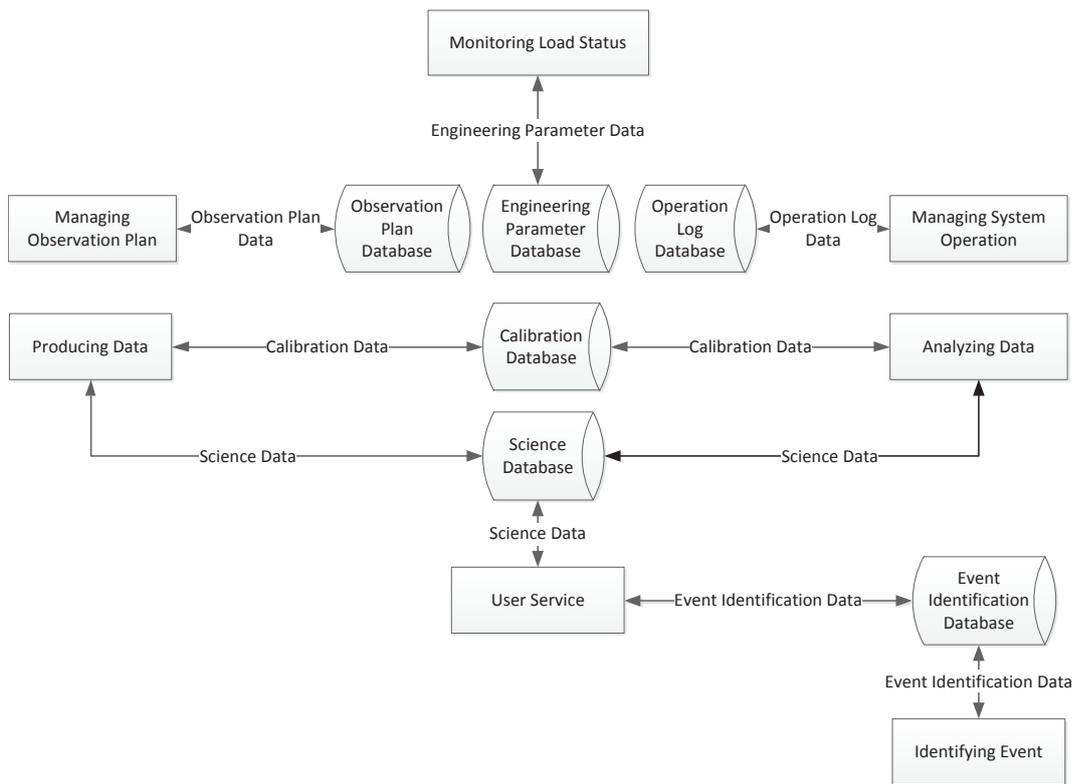


图 2 ASO-S科学应用系统的数据库业务架构图

Fig. 2 The database business architecture of the ASO-S science operation and data center

3 数据库的选择

根据ASO-S卫星工程的科学应用系统工作的特点, 为科学研究提供服务的数据库必须满足以下需求: 必须保证存储的数据准确, 数据库必须稳定, 必须满足事务的ACID特性^[10] (Atomicity、Consistency、Isolation、Durability), 满足一定的性能指标等要求. 只有满足ACID特性, 才能保证对科学应用系统数据库的事务性操作的正确性. 基于前3点需求, 科学应用系统的数据库还是适合使用关系型数据库.

根据评估, 仅科学数据数据库每天约存储8万条左右的数据, 一个月就是240万条数据, 4 yr共存储约1.2亿条左右的数据, 某些表最后可能会存储几百万条数据. 针对这样的大容量数据, 选择数据库主要从性能和稳定性方面考虑. 目前关系型数据库软件常用的有Oracle、MySQL和MariaDB. 这3款数据库软件易用性都较好, 帮助资料也容易寻找, 性能和稳定性方面Oracle数据库最优. 因此优先选择Oracle, 其次选择MySQL和MariaDB.

4 数据库性能优化

ASO-S卫星设计寿命为4 yr, 在设计寿命期内的预估数量超过4 PB. 数据种类多, 数据量大是ASO-S卫星数据的特点, 也是ASO-S卫星工程科学应用系统的数据库创建和维护工作的难点. 针对4 yr共存储约1.2亿条数据容量的数据库, 增加、删除、改写和查询等操作的性能是一个必须认真考虑的问题. 为此, 我们拟采用如下一些提高性能的方法.

(1)硬件方面: 在ASO-S卫星科学应用系统拟建万兆带宽内部网络条件下, 数据库的硬件瓶颈主要为硬盘的输入输出(IO)速率^[11]. 需要选择平均物理寻道时间和平均旋转延迟时间小一些的硬盘;

(2)软件方面: 针对ASO-S卫星的科学数据数据库每月数据有240万条数据的特点, 需要有针对性地采取优化措施. 根据公开的《阿里巴巴Java开发手册》的建议, 单表行数不宜超过500万行, 单表容量不宜超过2 GB, 否则可能对性能有较大影响. 为了提高数据库的响应速度, 我们拟将科学数据数据库的单表行数控制在100万行以内. 因此需要按照月份创建科学数据数据库的归档表, 每两个月就将科学数据数据库的本月以前的48个子库数据移动到归档表中. 用户从数据库读取科学数据的时候会根据输入的日期参数定位到对应的归档表里再读取相应的数据. 除了这种按月归档的方法外, 还计划针对科学数据数据库使用分区表技术. 对表进行分区后, 逻辑上仍然是一张完整的表, 只是将表数据存放到多个分开的物理空间. 这样查询数据时, 不需要扫描整张表. 从用户的角度来看, 使用上没有任何区别, 对分区后的表进行增加、删除、改写和查询操作时, 无需任何修改.

5 数据库的表样例

数据库的关系型模型内容根据不同的软件应用分为不同的2维表. 因为涉及业务太多, 所以这里仅列出一些2维表的设计作为样例, 包括科学数据数据库的部分参数表(表1)、特征事件识别数据库的部分参数表(表2)和工程参数数据库的部分参数表(表3). 在ASO-S卫星正式运行后, ASO-S科学应用系统的数据下载网站上会提供对应各种参数的详细说明的技术文档.

表 1 科学数据数据库: 1级SDI/LST数据部分参数表
 Table 1 Science data database: some parameters of SDI/LST level 1 data

Serial Number (SN)	Table Column Name	Whether Level 1 keywords	Whether Searched	Data Type
1	simple	yes		char
2	bitpix	yes		number
3	naxis	yes		number
4	naxisn	yes	yes	number
5	extend	yes		char
6	dsun_obs	yes		number
7	rsun_obs	yes		number
8	filename	yes		char
9	bunit	yes		char
10	content	yes		char
11	comprssn	yes		char
12	date	yes		date
13	telescop	yes		char
14	instrume	yes	yes	char
15	camera	yes	yes	char
16	date_obs	yes	yes	date
17	time_obs	yes		date
18	exptime	yes	yes	char
19	wavelnth	yes		number
20	tempdet	yes		char
21	aec_mode	yes	yes	char
22	acs_mode	yes	yes	char
23	lvl_num	yes	yes	number
24	xcen	yes	yes	number
25	ycen	yes	yes	number
26	quality	yes		number
27	ff_name	yes		char
28	ff_date	yes		date
29	dc_model	yes		char
30	dc_date	yes		date

表 2 特征事件识别数据库: 耀斑识别结果的部分参数表
 Table 2 Event identification database: some parameters of automatic detection of solar flares

SN	Table Column Name	Data Type
1	date	date
2	start_time	date
3	flare_id	char
4	active_region_number	char
5	x_coordinate_at_central_location	number
6	y_coordinate_at_central_location	number
7	minimum_x_coordinate	number
8	minimum_y_coordinate	number
9	maximum_x_coordinate	number
10	maximum_y_coordinate	number
11	area	number
12	maximum_intensity	number
13	minimum_intensity	number
14	mean_intensity	number
15	total_intensity	number
16	wst_detected_or_not	char
17	wst_start_time	date
18	wst_peak_time	date
19	wst_end_time	date
20	goes_class	char
21	goes_start_time	date
22	goes_peak_time	date
23	goes_end_time	date
24	hxi_flare_event_or_not	char
25	hxi_start_time	date
26	hxi_peak_time	date
27	hxi_end_time	date

表 3 工程参数数据库: 载荷管理器部分工程参数表
Table 3 Engineering Parameter Database: some engineering parameters of payload management

SN	Table Column Name	Data Type
1	system time code seconds	number
2	system time code millisecond	number
3	engineering parameter count	number
4	data injection packet receiving count	number
5	program upload packet receive count	number
6	program upload error packet serial number	number
7	data injection package checksum	number
8	load data processor program upload packet forwarding count	number
9	front electrical engineering configuration parameter block receive count	number
10	forwarding count fee configuration parameter block	number
11	negative direction fee configuration parameter block receive count	number
12	direct instruction forwarding count a power supply box	number
13	direct instruction forwarding count b power supply box	number
14	direct instruction forwarding count c power supply box	number
15	direct instruction forwarding count d power supply box	number
16	load data processor indirect instruction forwarding count	number
17	negative direction fee indirect instruction receive count	number
18	x direction power supply box instruction executed count	number
19	y direction power supply box instruction executed count	number
20	current monitoring threshold configuration count	number
21	load manager indirect instruction execution count	number
22	load manager injection event pending execution count	number
23	load manager injection event executed count	number
24	correct receiving count of load operating parameters	number
25	inaccurate receiving count of load operating parameters	number

6 总结与展望

本文介绍了ASO-S卫星工程科学应用系统的数据库设计. ASO-S卫星工程科学应用系统数据库为系统的科学运行、数据管理、数据生产和分析以及用户服务等功能提供数据层支撑. 根据ASO-S卫星科学应用系统的需求, 拟创建的数据库包括: 观测计划数据库、工程参数数据库、运维日志数据库、科学数据数据库、定标数据数据库和特征事件识别数据库等. 其中, 科学数据数据库包括48个子库, 对应不同的载荷和数据级别.

在实际创建和运维ASO-S卫星工程科学应用系统数据库的工作中,还可能根据需要对这些数据库进行针对性的设计和结构优化,以保证ASO-S科学应用系统的各应用层对数据库进行增加、删除、改写和查询等操作时,系统的性能和稳定性不受大量数据的影响.同时针对数据库的稳定性,从硬件角度选择性能优和质量好的存储和计算设备,从软件角度选择适当的数据库备份策略.

本文提出的针对ASO-S卫星科学应用系统的数据库架构设计及性能优化方法也可以为其他科学卫星的科学应用系统设计提供参考,具有复用价值.

致谢 感谢中国科学院紫金山天文台甘为群研究员在本论文撰写过程中给予的建议和帮助.感谢审稿人对本文的认真审阅及提出的宝贵意见.

参考文献

- [1] Gan W Q, Ding M D, Huang Y, et al. RAA, 2019, 19: 155
- [2] Deng Y Y, Zhang H Y, Yang J F, et al. RAA, 2019, 19: 157
- [3] Li H, Chen B, Feng L, et al. RAA, 2019, 19: 158
- [4] Zhang Z, Chen D Y, Wu J, et al. RAA, 2019, 19: 160
- [5] Huang Y, Li H, Gan W Q, et al. RAA, 2019, 19: 164
- [6] 王雷, 王智广. 计算机工程与设计, 2017, 38: 1808
- [7] 江洪波. 信息通信, 2016, 12: 205
- [8] 陈焕彬. 电脑与信息技术, 2016, 24: 45
- [9] 陈丙秋, 杨明, 姜碧涛. 天文学报, 2011, 52: 297
- [10] 秦靖, 刘存勇. Oracle从入门到精通. 北京: 机械工业出版社, 2011: 258-269
- [11] 柳遵梁, 潘敏君, 应以峰. Oracle数据库性能优化方法论和最佳实践. 北京: 机械工业出版社, 2016: 243-254

The Database Management Analysis for the ASO-S Science Operation and Data Center

GE Yun-yi LI Hui HUANG Yu LI Jing-wei LEI Shi-jun ZANG Jing-jing
SU Yang FENG Li LI You-ping LI Ying XU Zun-lei LI Xiang

(Key Laboratory of Dark Matter and Space Astronomy, Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023)

ABSTRACT The Science Operations and Data Center (SODC) of the Advanced Space-based Solar Observatory (ASO-S) satellite project connects the mission and users, and is responsible for the science operations, the data production, management, distribution, and archiving. SODC heavily relies on various databases with high stability and performance. This article introduces the function architecture and performance optimization of the databases in the frame of the SODC software via several examples with detailed realizations of the database tables. The proposed databases for detailed design include observation plan database, engineering parameters database, operation log database, science data database, calibration data database, and featured event identification database. This article can be a reference for establishing similar databases of other scientific satellite missions.

Key words Advanced Space-based Solar Observatory (ASO-S), astronomical databases, methods: analytical