ORACLE 数据库架构优化及数据治理研究

田宏胜,李佳华

(创发科技有限责任公司,湖南长沙 410126)

【摘要】在大数据时代,随着应用系统数据量的不断膨胀以及数据价值的逐渐显现,数据库架构优化及数据治理势在必行。文章从数据模型着手,实施自下而上的数据治理;同时通过监控、不断优化,加强数据模型的变更管控,逐步建立数据模型标准,规范 ORACLE 数据库架构,从整体上提升 ORACLE 数据库的运营效率。

【关键词】数据字典;索引;碎片整理;性能监控

【doi:10.3969/j.issn.2095-7661.2018.04.012】

【中图分类号】TP311.13　　【文献标识码】A　　【文章编号】2095-7661(2018)04-0040-03

Research on ORACLE database architecture optimization and data governance

TIAN Hong-sheng, LI Jia-hua

(Transfar Technology Co., Ltd, Changsha, Hunan, China 410126)

Abstract: In the era of big data, with the expanding of data size and the emerging of data value, it is necessary for the optimization of database architecture and management of data. Starting from data model, data governance is implemented from bottom to top. Meanwhile, by means of monitoring, keeping optimizing, and enhancing the change control over data model, a data model standard is gradually established to standardize Oracle database architecture and to improve Oracle database overall operating efficiency.

Keywords: data dictionary; index; defragmentation; performance monitoring

1 数据库应用现状

随着信息化系统的集中化管理,后台支撑的信息系统越来越复杂,承载的数据量也越来越庞大。系统的稳定、高效运行是支撑业务发展的关键。而新业务的不断发展、营销新模式的不断推出使信息系统处在不断的积木式迭代开发环境中。不断增长的业务数据,不断变化的数据字典,频繁的版本发布,这些都随时可能对数据库的性能造成冲击。

无论是 ORACLE 数据库的稳定维持, 还是数据库的优化,都不是能一蹴而就或者一劳永逸的,需要定期的维护,定期的观察。因此有必要制定一套合理可行的数据治理、数据库架构管理及优化流程,最大限度的来保证信息系统的数据变化、版本发布不会对业务系统的正常运行造成不利影响。

2 ORACLE 数据库数据治理及架构优化

ORACLE 数据库的数据治理及架构优化,是一项大规模的数据库精简工程。目的是提升数据的质量,保证业务系统良好的运行性能。从固化、提升、加强到可持续发展四个方面来稳定架构优化成果。

2.1 梳理物理模型

信息系统的模型规划,数据字典的设计一定要有前瞻性,要为系统的扩展留有余地。特别是部分系统是分期建设, 部分开发厂商出于成本与工期考虑,往往对系统的架构、数据字典的设计没有考虑系统的可扩展。系统上线后,因为业务的需要而频繁地对表增加字段、增加索引,最终导致数据字典的混乱,从而影响系统的正常运行。对于已经上线的系统可从以下几个方面来进行优化。

1)确定数据字典。信息系统提供商应该提供完整的数据字典结构,对数据库现有的用户、对象(表、索引、序列、存储过程等)提供详细的说明。存在的对象都要有据可查,同时每次版本发布更新数据字典结构

[收稿日期] 2018-09-07

[作者简介] 田宏胜(1975-),男,湖南古丈人,创发科技有限责任公司工程师,学士,研究方向:计算机软件。

后应及时更新原有数据字典文档,确保数据库中不存在业务系统不需要的对象[1]。

2)确定无用对象

确定无用索引。对于每一个表中创建的索引一定要严格管理，部分系统对于索引的创建比较随意,如曾有系统索引一个月未用到一次的比例占索引总数的 60%以上。因此有必要对上线系统使用的索引的使用情况进行监控。监控的语句为 alter index <INDEX_NAME> monitoring usage;监控的结果记录在 V$object_usage 里[2]。

确定无用表。对于表的使用,可通过编写触发器来进行。实际中可通过查看 dba_tab_modifications 及结合 v$access 视图来实现。(dba_tab_modifications 表记录所有表的 update、delete、insert 记录条数，但在 dbms_stats 后自动清除记录)。

3)整合系统数据。通过整合系统规格数据、流程数据、历史数据,使系统架构清晰明了。应用系统厂家应该合理规划系统架构,使系统规格数据、流程数据有效整合,同时做好历史数据的分离工作,从而使存放有效期内数据的业务表不至过于庞大,不会随着时间的推移而使系统的性能下降。

4)清理无用对象。通过技术、管理等各种手段确定的数据库对象(如表、索引、存储过程等),维护和应用系统人员进行统一确认,对确实不需要的对象进行清理[3]。数据库本身不应该存放大量的附件数据以及图像数据,存放在操作系统磁盘上更加合理,以减少数据库容量和备份时间。尽量避免 DBlink 的使用,采用 Web 服务的方式替代。同时从管理的角度来保证信息系统的正常运行，主要从这几个方面来保证:版本发布定时审阅、稽核是否需要增删改;确定的数据字典发布版本定期更新;存档系统数据字典,保持系统优化效果和架构持续合理。

2.2 优化物理模型

随着时间的推移,数据库中存储的数据量越来越庞大,大量的历史数据没有及时从运行系统中清理出去,运行系统数据库的不断增、删、改等操作导致数据的碎片化较为严重,同时也会伴随较多垃圾数据的产生,这些都是威胁业务系统正常运行的因素。因此需要定期进行历史数据分离、数据清理和碎片整理等三方面的工作[4]。

2.2.1 历史数据分离

为保证系统的良好运行,要有一套行之有效的机制保证超过时限的业务数据能及时地从运行库中迁移到历史库中。为保证历史数据的迁移不影响业务系统的正常运行,信息系统厂家在进行架构规划、数据字典设计中一定要充分考虑到这一点。如存放通话详单的表一定要按时间分区(或按月单独存放在一个表中)，以便于进行历史数据的分离和运行库中过期数据的删除工作。

2.2.2 定期清理数据

根据数据库中存放数据表的不同特性进行定期的清理工作。信息系统厂家根据表的业务特性提供详细清单，确定其可进行 drop、truncate 还是 delete 操作,并提出定期操作的时限要求。

Drop 操作：对于业务系统运行不再需要的表进行此操作，操作语句:drop table owner. tbale_name purge; Truncate 操作：对于系统运行中不再需要其数据,但要保留其数据结构的表进行此操作,操作语句:truncate table ower.table_name;Delete 操作：对于业务系统表,部分数据不再需要的进行此操作,操作语为 delete from owner.table_name where <condition>。

2.2.3 定期整理碎片

随着数据的不断增、删、改等操作,表和索引会存在较严重的碎片化,导致这些对象占用大量的存储空间,系统访问这些数据增加 IO 访问量,从而影响系统的性能。

1)索引。对于索引,通过语句 analyze index index_name validate structure;(结果存放在 index_stats 中)来确认索引的碎片化程度。

情况一:SQL> analyze index abcd validate structure;SQL> select (del_lf_rows_len/lf_rows_len)*100 as index_usage From index_stats Where index_name = 'ABCD';如果删除的行数占总行数的 30%,即 del_lf_rows / lf_rows > 0.3。情况二:如果"height"大于 3。情况三:如果索引的行数(LF_rows)远远小于 LF_BLKS。

以上三种情况需要进行索引重建或者 rebuild 操作。索引重建:先删除原有碎片化严重的索引,再重新创建该索引。rebuild 索引:不删除原索引,使用语句 alter index index_name rebuild parallel <degree> online,以达到消除碎片的目的。

2)表。对于那些进行频繁 DML(insert、update、delete) 操作的表，一般会存在严重的碎片化。通过 dbms_stats 分析后查看 dba_tables 可以确定表碎片化程度。在 ASSM 下,block 的空间使用分为 free space:0-25%、25-50%、50-75%、70-100%和 Full 这 5 种情况，使用 dbms_space.space_usage 对需要统计的 table 汇总这 5 种类型的 block 数量。当 table HWM 下的 blocks 的状态大多为 free space 较高值时，可以考虑合并 HWM 下的 blocks，将空闲的 block 释放，降低 table 的 HWM。

对于表的整理可以通过三种方式实现，方法一，对表进行 move 操作，同时局部索引也需重建，语句

为:alter table table_name move tablespace tablespace_name。方法二,通过 expdp/impdp 操作来进行碎片整理。方法三,整理行迁移(dba_tables.chain_cnt)数目较大的表。

识别行迁移:SQL>analyze table TABLE compute statistics (会覆盖统计信息);SQL>@?/rdbms/admin/utlchain.sql;SQL>analyze table table_name list chained rows into chained_rows。

消除行迁移可通过 move、expdp/impdp、将迁移行插入临时表再删除后回填的方法(不适用于有外键引用的表)。

2.3 增强架构变更管理

由于企业营销手段的多样化、业务变更频繁,信息支撑系统会经常进行版本发布,而版本发布就会有数据库结构的变更。因此在版本发布前一定要加强架构变更管理,保证数据结构的变更不会破坏系统整体架构。同时加强对新增或变更 SQL 语句的审核,要在测试平台进行测试,同时加强这些 SQL 语句执行总效率的评估和上线后的性能跟踪。

2.4 实施日常性能监控

通过 AWR、ADDM 对数据库进行性能监控,及时发现消耗资源的 SQL 语句,通过业务及技术手段来优化这些 SQL 语句,以保证系统持续健康的运行。

1)定期检查消耗资源过高的 SQL。通过操作系统监控及时发现 CPU 的异常增长情况,对比近期 CPU 消耗情况,消耗资源过高的 SQL 可以通过 ORACLE 的 AWR 报告分析出来,及时优化[5]。

2)定期分析特殊的业务表。虽然 oracle 10g 以上版本是数据库后台自动进行表分析的,但是只有在数据表的数据量改变在 10%以上才会自动分析,对于大表,自动分析有时需要较长时间才会执行。带有时间索引的 SQL,如果不及时分析,将产生错误的执行计划,影响系统稳定。因此,根据业务需求,增加分析特殊业务表的定时任务是必要的。

3 总结

通过数据库架构优化和数据治理,基本可以达到 ORACLE 数据库系统所需容量有效控制、系统物理模型有效精简、系统碎片有效整理的效果,使得整个数据库系统性能提升、系统响应速度提高。同时系统运行效率较高,维护费用就会相对降低,一定程度上可节省系统维护成本。

【参考文献】

[1]吴锋珍.高校图书管理系统的数据库设计[J].湖南邮电职业技术学院学报,2016(2):36-39.

[2]韦壹刚.Oracle 数据库日常维护与优化建议[J].信息与电脑(理论版),2018(8):115,117.

[3]常正青.基于 Oracle 数据库系统的优化与性能调整研究[J].电子世界,2017(17):195.

[4]韦德昌.Oracle 数据库应用系统的性能优化[J].数码世界,2017(4):14-15.

[5]孟津平.浅谈 Oracle 数据库 SQL 调整与优化[J].数字技术与应用,2018(3):227,229.

(上接第 27 页)

投诉量最高的为:J 社区、火车站、R 医院。公众场所的投诉量居高不下,处理一个类似的热点区域往往能够同时处理几十个甚至上百个同类用户的投诉,此类投诉对 NPS 影响较大。因此加强公众场所、住宅小区的建设应该是今后建站重点考虑的方面。

5 结束语

进行基于大数据挖掘的客户投诉热点分析,通过挖掘投诉大数据价值,可以对运营商在网络、服务等多个环节进行预警,从客观角度给运营商提供了一个新的视角,审视运营商的服务是否达到客户的需求。运营商可以结合网管告警、话务统计、客户投诉等多维数据来源,实现网络的维护闭环处理,一方面完成投诉、NPS 等多种考核指标的提升,另外一方面可以提高网络投资效益,做到精准投资、有效投资,具备较大的推广价值。

【参考文献】

[1]聂晶.Python 在大数据挖掘和分析中的应用优势[J].广西民族大学学报(自然科学版),2018(1):77-79.

[2]卢中宁,张保威.一种基于改进 TF-IDF 函数的文本分类方法[J].河南师范大学学报,2012(6):158-160.

[3]曾雄辉.MR 大数据分析支撑 LTE 800M 精准规划建设研究[J].湖南邮电职业技术学院学报,2017(4):14-17.

[4]蔡志成.基于 4G 环境的数据挖掘在移动通信网络优化中的运用研究[J].通讯世界,2017(8):81-82.

[5]王森.4G 环境下移动通信网络优化中数据挖掘的应用研究[J].网络安全技术与应用,2016(12):96-98.