北斗卫星导航系统数据分析与应用专题·论文

基于大数据技术的北斗导航信息服务性能提升

曹纪东*, 唐桂芬, 刘晓萍, 孙伟杰

北京卫星导航中心, 北京 100094

*联系人, E-mail: jncs313@139.com

收稿日期: 2015-03-27; 接受日期: 2015-04-30; 网络出版日期: 2015-05-29

摘要 随着全球卫星导航系统的迅速发展, 导航信息服务在各行业领域得到了大范围的应用. 面对导航信息服务数据量的快速指数级增长, 传统的数据管理模式出现了存储效率降低、查询检索时间过长、缺乏数据智能分析方法等一系列问题, 已经不能满足各项应用对卫星导航信息服务的需求. 本文针对现有北斗卫星导航系统数据管理模式中出现的问题, 研究并提出了基于大数据的北斗导航信息服务性能提升技术. 试验结果表明, 采用本文的大数据技术, 北斗信息服务的数据存储效率能够提高 2.5 倍以上, 用户查询等待时间减少 90% 左右, 并具有初步的数据智能分析能力.

关键词 北斗导航信息服务, 大数据, 性能提升

PACS: 91.10.Fc, 93.85.Bc, 07.05.Kf, 29.85.+c

doi: 10.1360/SSPMA2015-00119

1 引言

卫星导航系统可发送高精度、全天时、全天候的导航、定位和授时信息, 是当今国民经济和国防建设不可或缺的重要空间基础设施, 该系统在我国海洋渔业、物流、运输、旅游、抢险、水文、水利、地质勘探等众多领域应用非常广泛, 已形成庞大的卫星导航产业.

同时, 美国 GPS 系统自建成以来, 美国军方、政府及商业组织均投入了大量的资金进行相关信息服务系统的建设, 逐步形成了网络化的位置报告系统、基于位置的信息处理与分发系统、GPS 系统完好性等导航保障信息服务系统, 进一步深化和拓展了 GPS 的应用领域^[1].

在我国, 较为典型的卫星导航信息服务应用系统包括中国交通通信中心研制的北斗基础信息服务系统、北斗星通公司研发的北斗船位监控指挥管理系统、神州天鸿公司研发的三峡水情数据采集系统等. 主要信息种类需求为其下属用户的位置信息和通信信息, 用户总数近 90 万, 平均每秒提供近 200 次服务, 在交通监控管理、海洋渔业指挥管理、水文监测等方面发挥了巨大的作用.

随着近年来卫星导航应用的快速发展, 卫星导航系统需要管理的数据量明显增加, 传统的数据管理技术已经无法满足卫星导航信息服务的需求, 较为突出的问题包括:

(1) 使用单一存储设备, 硬件的访问速度限制导致数据存储效率降低, 特别是数据查询服务对数据

引用格式: 曹纪东, 唐桂芬, 刘晓萍, 等. 基于大数据技术的北斗导航信息服务性能提升. 中国科学: 物理学 力学 天文学, 2015, 45: 079510
Cao J D, Tang G F, Liu X P, et al. The performance enhancement of COMPASS navigation information services based on big data technology (in Chinese). Sci Sin-Phys Mech Astron, 2015, 45: 079510, doi: 10.1360/SSPMA2015-00119

存储效率的负面影响尤为明显;

(2) 数据库表设计不合理, 数据入库与数据表索引间的矛盾无法解决, 导致数据检索时间过长, 用户查询等待时间逐渐增加;

(3) 缺乏数据智能分析方法, 大量数据仅仅是被动地等待用户, 而不是主动综合分析, 无法产生更大的效用.

因此, 需要结合大数据管理技术, 研究新的卫星导航数据管理方法, 适应卫星导航信息服务的新需求.

2 基于大数据技术的性能优化方法

大数据管理技术通常包括结构化数据、云计算、服务器集群、分布式存储技术、并行数据库设计、智能分析技术等^[2]. 结合卫星导航信息服务存在的问题, 下面从优化存储架构、提高查询速度、开发数据综合分析能力 3 方面进行研究.

2.1 设计分布式存储架构

卫星导航信息服务需要同时满足数据高速写入和用户快速查询这两方面需求. 硬盘级的读和写是一对天生的矛盾, 对于传统的单套存储系统来说, 只能尽量平衡两方面的需求, 或者仅满足一方需求^[3]. 因此, 需要设计能够同时满足硬盘级读写需求的存储架构.

分布式存储架构的设计如图 1 所示.

(1) 建立集中存储中心, 并统一数据存储格式规

图 1 (网络版彩图) 分布式存储架构设计图

Figure 1 (Color online) Distributed storage structure design.

范, 提高数据处理效率.

卫星导航系统及全部相关系统的数据全部按照标准格式统一集中存储. 这样能够简化数据处理流程, 节约大量数据解析和转换的时间成本^[4]. 经测试, 每包数据(50 KB)的平均处理时间从原来的 21 ms 降为 10 ms 左右. 如图 2 所示.

(2) 将用于数据存储和查询的磁盘阵列从物理层面上分离, 提高磁盘写入效率.

分别建立高速盘阵区和低速盘阵区(相对低速). 其中, 高速磁盘阵列采用固态硬盘, 仅用于数据写入, 避免数据查询动作影响磁盘的写入效率; 低速磁盘阵列采用光纤硬盘, 分为两个区, 一个用于数据检索与查询, 另一个用于定制服务数据的综合分析处理.

高速盘阵和低速盘阵间的数据同步可以通过磁盘阵列自带的同步软件完成, 依据目前每天 50 GB 的数据增量测试, 整个同步时间能够在 10 min 内完成, 由于进行数据同步时低速盘阵需要同时进行写操作, 而此时的用户查询又需要读操作, 为避免影响用户访问速度, 数据同步工作可以在每天凌晨 2:00 进行.

(3) 采用大容量内存, 增大缓冲区, 提高数据库写入效率.

数据库的写入效率与数据库服务器的可分配内存成正比^[5], 当前的主流服务器均能够配备 TB 级内存, 增大数据库的可使用内存, 能够显著提高数据库的入库效率.

2.2 优化数据查询服务性能

数据查询服务性能主要依赖于数据库的性能^[5].

图 2 (网络版彩图) 数据包处理时间比较图

Figure 2 (Color online) Data package processing time comparison.

在数据库设计方面, 根据数据存储和查询的不同需求, 采用并行数据库设计, 分别建立存储数据库和查询数据库, 并部署在不同盘阵上, 如图 3 所示.

(1) 存储数据库部署在高速盘阵上, 按照每天一个表的方法设计, 表名后缀从 1-31, 表示 1-31 天, 如观测数据表名为 OBS_1 至 OBS_31, 其中包括各卫星、各观测站、各接收机的数据, 不建立索引, 存储时按照数据块集中写入磁盘. 这样可以最大程度地保障数据写入速度^[3].

(2) 查询数据库部署在低速盘阵上, 按照每小时一个表的方法设计, 表名根据数据类型和时间进行区别, 如观测数据的表名为“OBS_SAT*_STA*_RCV*_Y*_M*_D*_H*”, 表示“观测数据_卫星号_观测站号_接收机号_年_月_日_时”, 这样可以将查询数据分类进一步细化, 尽量满足用户快速访问的需求.

(3) 设计数据管理策略, 存储数据库的表中仅保留一个月的数据, 每天清除编号为下一天的数据. 例如: 数据库中目前存储的数据为 3 月 10 日至 4 月 9 日, 系统正向后缀编号为 9 的数据表中写入数据, 当 4 月 9 日 23:50 时, 清除编号为 10 的数据表中的数据, 即清除 3 月 10 日的数据, 为写入 4 月 10 日的数据做准备.

查询数据库根据数据种类不同, 设计不同的管理策略, 如: 精密定轨数据保留最近 15 天, 观测数据保留 1 年, 钟差数据保留 2 年, 电离层数据长期保留等. 到期的数据写入对应的数据文件, 再删除数据表, 当需要查询时, 直接以文件方式提供下载.

数据文件命名规则与表名相对应, 如上文的

“OBS_SAT*_STA*_RCV*_Y*_M*_D*_H*” 数据库表中的数据写为数据文件后, 文件名即为“OBS_SAT*_STA*_RCV*_Y*_M*_D*_H*.txt”.

(4) 在夜间数据同步时, 同时完成数据在两类数据库表间的数据转换.

如图 4 所示, 在进行数据同步时, 将存储数据库表中的数据依次读出, 按照数据要素分别写入查询数据库中对应的表, 同时实现数据同步和数据转换.

2.3 开发导航服务数据智能分析能力

数据智能分析是大数据的重要特征之一^[6], 通过数据分析, 从海量数据中提炼出用户的隐形需求, 是卫星导航信息服务的价值所在.

(1) 数据关联性分析, 主要用于向用户提供其可能需要的数据服务.

由于现有卫星导航应用领域的快速拓展, 其数据关联性分析主要的难点并不在于数据组合变化的多样性, 而是如何在现有已知数据关联的情况下进行合理的扩展, 既要满足数据检索的扩展需求, 又要避免因为部分个体的访问而轻易地变更数据关联, 从而向多数用户推送无用信息而造成的资源浪费.

具体方法为, 设事先提供的相关数据使用量为 X_n , n 为数据类别个数, 求出平均数 M , 我们称 X_n 占 M 的百分比为数据关联值, 分别用 C_n 表示, 如式(1)和(2)所示.

$$M = \frac{\sum_{i=1}^n x_i}{n}, \quad (1)$$

$$C_n = \frac{X_n}{M} \times 100\%. \quad (2)$$

当 C_n 低于某一指定阈值时, 该数据被认定不与当前数据存在关联关系.

图 3 (网络版彩图)并行数据库设计图

Figure 3 (Color online) Parallel database design.

图 4 (网络版彩图)数据库表转换示意图

Figure 4 (Color online) Database table transition sketch map.

(2) 导航应用前瞻性分析是通过用户对使用导航数据的需求情况, 分析统计出卫星导航将来可能发展的应用方向.

就卫星导航数据分析而言, 主要目的是为用户提供定制服务的基础上, 根据用户需要快速完成相关数据的综合处理, 在返回处理结果的同时, 在数据库中按类别记录用户的服务需求, 并定期进行分析导航数据的使用方向, 为卫星导航应用开拓思路.

各类应用具有不同的统计分析方法. 以城市交通路况预警为例, 对用户城市驾车行驶轨迹按时间段进行统计, 能够得出城市交通拥堵路段的概略时间^[7], 从而发布交通拥堵预警, 并在用户路径规划服务中及时进行标注, 提醒用户注意.

具体分析方法为, 以通常车流量较多的时段 7–22 时为统计区间, 分别统计某相同路段每小时的车流量 D , 取平均数 M , 分别计算 D_n 与 M 的差占 M 百分比 C_n , 我们称为增量比. 如式(3)和(4)所示.

$$M = \frac{\sum_{i=1}^n d_i}{n}, \quad (3)$$

$$C_n = \frac{d_n - M}{M} \times 100\%. \quad (4)$$

当增量比 C_n 大于某一指定阈值时, 该小时为车流量高峰时段.

3 试验结果

3.1 数据存储效率试验

分别采用传统数据管理模式和大数据管理模式, 模拟在用户随机查询数据库的同时, 测试数据在两种模式下写入数据库的速度, 分别进行 60 组, 每次 10 min 取平均值. 测试结果见图 5.

经统计, 在传统数据管理模式, 平均入库效率为 6.1 MB/s, 且波动较大, 最大值 8.2 MB/s, 最小值仅 3.1 MB/s; 而采用大数据管理技术后, 平均入库效率为 15.5 MB/s, 效率平稳, 最大值 16.8 MB/s, 最小值 14.2 MB/s, 结果表明, 采用大数据管理技术能够将数据入库效率提升至原来的 2.5 倍.

3.2 用户查询等待时间试验

分别采用传统数据管理模式和大数据管理模式, 模拟在数据写入磁盘阵列的同时, 测试用户在两种

图 5 (网络版彩图) 两种数据管理模式数据入库效率对比图

Figure 5 (Color online) Database entrance rate comparison between two data management modes.

模式下的查询等待时间, 分别发出 20 组数据查询请求, 在每组查询参数中, 数据种类和时间各不相同, 测试结果见图 6.

经统计, 在传统数据管理模式, 用户的平均查询等待时间为 27 s, 而且波动较大; 而采用大数据管理技术后, 用户的平均查询等待时间缩短至 2.8 s, 最大值不超过 5 s. 结果表明, 采用大数据管理技术能够将用户查询等待时间减少 90% 左右.

3.3 验证导航服务数据智能分析能力

(1) 数据关联性分析能力试验.

以精密定轨数据查询为试验对象, 当用户进行精密定轨数据检索时, 系统应当同时提供轨道机动信息、激光数据、监测站监测接收机等等效距离误差 MERE(Monitoring Equivalent Range Error)、广播星历等相关信息的下载目录, 我们称之为关联数据, 便于用户使用.

同时, 系统记录用户对这些关联数据的使用情

图 6 (网络版彩图) 两种数据管理模式用户查询等待时间对比图

Figure 6 (Color online) User query waiting time comparison between two data management modes.

况, 根据用户使用次数的统计结果动态调整数据间的关联关系.

表 1 是精密定轨关联数据一段时间内的用户使用次数统计. 指定数据关联阈值为 30%.

依据式(1), 可以计算出平均使用次数 M 为 70.25, 依据式(2), 得出上述 4 类信息的数据关联值依次为 149.5%, 108.2%, 125.3% 和 17%.

从计算结果可以看出, 轨道机动信息、激光数据、监测站 MERE 的数据关联值均大于 30%, 广播星历的数据关联值小于 30%, 则应当取消精密定轨数据和广播星历数据间的关联关系.

(2) 车流量分析预警能力试验.

模拟某路段一天中每小时车流量, 数据见表 2. 指定增量比为 30%.

车流量分析统计如图 7 所示. 应用式(3)和(4), 计算可得出 12, 13, 14, 18, 19, 20 时的增量比分别为 33.6%, 37.3%, 31.2%, 43.7%, 68.8%, 60.2%.

从计算结果可以得出, 该路段 12–14 时, 18–20 时为拥堵时间, 与图 7 态势基本吻合.

4 结论

试验结果表明, 在卫星导航信息服务中应用大

表 1 精密定轨关联数据统计表

Table 1 Related data statistics table of precise orbit determination

关联数据	用户使用次数
轨道机动信息	105
激光数据	76
监测站 MERE	88
广播星历	12

表 2 某路段车流量数据统计表

Table 2 Analysis and statistics table of the cars flux on some road

时间	车流量	时间	车流量	时间	车流量
00:00	1287	08:00	5463	16:00	12352
01:00	421	09:00	9785	17:00	12462
02:00	175	10:00	8684	18:00	19328
03:00	164	11:00	11541	19:00	22705
04:00	181	12:00	17976	20:00	21548
05:00	205	13:00	18476	21:00	12624
06:00	546	14:00	17648	22:00	8759
07:00	3275	15:00	13621	23:00	3254

图 7 (网络版彩图)某路段车流量分析统计图

Figure 7 (Color online) Analysis and statistics of the cars flux on some road.

数据技术, 采用分布式存储架构和并行数据库设计, 能够有效提高数据存储效率 2.5 倍, 并减少 90% 的用户查询等待时间.

同时, 开发适当的导航服务数据智能分析能力, 对我国北斗卫星导航产业应用将能够起到较为积极的推动作用.

参考文献

- 曹会超. GPS 新的应用领域. 测绘地理信息发展动态, 2012, 49: 17
- Li X L, Gong H G. A survey on big data systems (in Chinese). Sci Sin Inf, 2015, 45: 1–44 [李学龙, 龚海刚. 大数据系统综述. 中国科学: 信息科学, 2015, 45: 1–44]
- Zhang B, Chen J R, Le J J. Overview on big data management technology research (in Chinese). Comput Appl Softw, 2014, 41: 1–6 [张滨, 陈吉荣, 乐嘉锦. 大数据管理技术研究综述. 计算机应用与软件, 2014, 41: 1–6]
- Wu G J, Wang S P, Chen M, et al. Massive structured data oriented storage and retrieve system (in Chinese). J Comput Res Dev, 2012, 49(Suppl): 1–5 [吴广君, 王树鹏, 陈明, 等. 海量结构化数据存储检索系统. 计算机研究与发展, 2012, 49(Suppl): 1–5]
- 杨旭, 汤海京, 丁刚毅. 数据科学导论. 北京: 北京理工大学出版社, 2014. 112
- Yan X F, Zhang D X. Big data research (in Chinese). Comput Technol Dev, 2013, 23: 168–172 [严霄凤, 张德馨. 大数据研究. 计算机技术与发展, 2013, 23: 168–172]
- Fang W, Zheng Y, Xu J. Big data: Conceptions, key technologies and application (in Chinese). J Nanjing Univ Inform Sci Technol (Nat Sci

Edit), 2014, 6: 405–419 [方巍, 郑玉, 徐江. 大数据: 概念、技术及应用研究综述. 南京信息工程大学学报(自然科学版), 2014, 6: 405–419]

The performance enhancement of COMPASS navigation information services based on big data technology

CAO JiDong^{*}, TANG GuiFen, LIU XiaoPing & SUN WeiJie

Beijing Satellite Navigation Center, Beijing 100094, China

With the rapid development of the global satellite navigation system, navigation information services have been widely applied in various industries. Facing with the rapid exponential growth in the amount of navigation information services data, traditional data management mode merged a series of questions such as reducing storage efficiency, long query retrieval time and lacking intelligent data analysis methods, which has been unable to fully meet the satellite navigation services needs of the application. According to the existing problems of COMPASS satellite navigation system data management mode, the performance enhancement technology of COMPASS navigation information services based on COMPASS navigation big data was studied and proposed in this paper. The experiment results show that the data storage efficiency can be increased by more than 2.5 times and the user query latency can be reduced by about 90% through using big data technology in this paper, which has preliminary data intelligence analysis capabilities.

COMPASS navigation information services, big data, performance enhancement

PACS: 91.10.Fc, 93.85.Bc, 07.05.Kf, 29.85.+c

doi: 10.1360/SSPMA2015-00119