

成果与方法

# 基于 Oracle 的国土资源基础数据组织方法

史辉,冯永玉,张妍,张洪林

(山东省国土资源信息中心,山东 济南 250014)

**摘要:** Oracle 数据库管理系统已成为我国国家级和省级国土资源基础数据库建设的主流技术平台。该文就国家级和省级数据库的建设,提出了在 Oracle 数据库技术环境下,如何在物理上合理组织和布局各种类型的数据,如何设置和管理各衍生数据,如何实施数据完整性要求,其目的是实现国土资源基础数据的高效管理和可持续更新支持,更好地满足国土资源基础数据库之上各业务应用系统运转的要求。

**关键词:** 数据库物理设计;数据分区;国土资源管理

中图分类号:P208

文献标识码:A

## 0 引言

在国土资源行业内,国土资源基础数据库存在着业务和技术层面上的双重含义。在业务层面上,它是指通过国土资源调查、评价、汇交、统计、管理等产生的反映国土资源现状、利用情况、规划、环境影响等方面的数据集;在技术层面上,它是指长期储存在计算机内的、有组织的、可共享的国土资源数据,具有较小的冗余度、较高的数据独立性和易扩展性,并可为多种用户所共享。该文中的国土资源基础数据库主要指其在技术层面上的含义。

土地、矿产、地质、环境、测绘等业务基础数据库的建设是金土工程建设的一项重要内容。Oracle 数据库管理系统在海量数据管理、支持开放的空间数据组织、业务连续性保证等方面具有明显的优势;长期以来,Oracle 数据库占据数据库行业主流地位。为此,我国国家级、省级、甚至部分市级国土资源管理部门将 Oracle 数据库管理系统作为其基础数据库建设的首选技术平台。

面对规模庞大、空间特性显著、数据来源分散、共享需求强烈的国土资源基础数据环境,应淋漓尽致地发挥 Oracle 数据库管理系统的各种技术优势,建设实现各类基础数据有序管理和相互间有效联系、支持可持续数据更新、支持各项应用连续运行、

保证上下级数据内容一致的国土资源基础数据库。基于笔者对国土资源基础数据的认识,该文探讨了在 Oracle 数据库技术环境下,有效进行国土资源基础数据组织的有关技术方法。

## 1 构建基础数据库的一般过程

国土资源基础数据库设计分为概念设计、逻辑设计以及物理设计3个阶段。其中,概念设计和逻辑设计依据相关的国土资源基础数据标准,围绕数据实体划分、明确数据实体关联关系等工作而展开,其主要的阶段成果则是实体关系图,概念设计和逻辑设计工作与具体所使用的数据库管理平台相独立。物理设计则是将逻辑设计阶段形成的各设计元素映射到特定数据库平台下的物理数据结构上,使之真正满足数据管理、更新、共享的要求。完成数据库的全部设计后,就进入数据库部署和业务化运行阶段。下面讨论使用 Oracle 数据库管理系统进行国土资源基础数据库物理设计的有关技术方法。

## 2 各类型数据的组织

国土资源基础数据包括结构化的关系数据、半结构化的空间(矢量、栅格)数据和 XML 数据、非结构化的文档和媒体类型数据等。在 Oracle 数据库管理系统下,可采用相应的数据类型组织有关数据。

\* 收稿日期:2008-10-12;修订日期:2009-02-17;编辑:陶卫卫

作者简介:史辉(1968-),男,山东招远人,高级工程师,主要从事国土资源信息化、GIS研究工作。

## 2.1 结构化的关系数据组织

这类数据的组织在几乎所有的关系型数据库系统中是通用的,在 ORACLE 下,可采用 NUMBER, VARCHAR2, DATE, TIMESTAMP 等简单数据类型分别组织数值型、字符串型、日期型、时间戳型数据<sup>[1]</sup>。

## 2.2 空间数据组织

采用基于 ArcSDE for Oracle 的 ArcGIS Geodatabase 模型实现空间数据的组织。一个 Geodatabase 支持在 3 个层次上实现空间数据的分级管理: Feature, FeatureClass 和 FeatureDataset。其中 Feature 对应于一个空间数据图层的一个图斑; FeatureClass 用于管理具有相同几何数据类型和相同属性信息的一组 Feature; FeatureDataset 则用于实现在共享相同空间参照的一组 FeatureClass 的统一管理<sup>①</sup>; 基于 ArcSDE for Oracle 的 Geodatabase 模型,可以选择 SDEBinary、ESRI ST\_GEOMETRY 和 SDO\_GEOMETRY 几种方式实现空间几何数据的物理存贮。对于栅格型的空间数据(DRG, DEM),则基于 ArcGIS 固有格式组织,在 Oracle 数据库中以 BLOB 类型存储。

## 2.3 XML 数据组织

随着 XML 数据的广泛使用,使用数据库管理系统统一管理 XML 数据和其他类型数据正成为一种技术趋势。Oracle 对管理 XML 数据提供很好的支持,提供 XMLTYPE 数据类型用于组织 XML 数据,并对 XML 数据的操作提供了数据库级的支持<sup>②</sup>。

## 2.4 文档和媒体数据组织

Oracle 的 InterMedia, 提供了与文档、图片、音频、视频或其他不同类型媒体相应的对象类,支持文档、图片、音频、视频或其他不同类型媒体数据的存贮、管理和查询,使文档及媒体数据与普通属性数据的管理和使用基于同一平台。同时,InterMedia 还通过内置的元数据提取机制为媒体数据的访问提供了便利的手段。对于其他形式的非结构化数据,可以使用 Oracle 的 BLOB, CLOB, FLOB 等数据类型进行组织。

# 3 数据物理布局

物理上,数据是要存储到磁盘上的。如何将现

有的和未来的数据合理布局到物理磁盘上是数据库物理设计的一项重要内容。保证数据的并发存取能力和易于管理是进行这类设计的主要考虑因素。

## 3.1 使用 Oracle 分配物理空间

在 Oracle 数据库中,组织数据的物理结构有数据块(Data Block)、范围(Extents)。数据块是 Oracle 数据库的最小数据单位。由于大部分国土资源基础数据库以批量写入、批量读取为主,兼顾 OLTP 和 OLAP 型应用的支持要求,使用较大的数据块。另外,鉴于事务级的基础数据增删改操作较少,笔者倾向于在数据库块中分配极少量的空余空间用于事务操作,并对重复度高的数据进行压缩存储。

随着数据库的运行,需要分配新的磁盘空间来容纳新的数据。Oracle 按照一次就分配若干数据块的策略来获得这部分空间。这若干个块就是一个范围(Extent)。由于国土资源基础数据多以批量加载为主,应基于每次增加的数据量估算范围(Extent)所含的块数,提高磁盘空间的利用效率并减少 I/O 负担。

## 3.2 基于分区的数据布局

国家级和省级国土资源基础数据往往是通过通过对市县提交的资料数据整合而成,各业务主题的数据规模是相当庞大的,一个数据表往往含有几百万甚至上千万条的数据记录;并且,这些数据每年都在随数据变更工作和日常管理而增加。在 Oracle 中,可有 2 种布局方法应对这一问题。

(1)按照所属行政区和年份划分数据表单元。通过命名机制来标识各表中所含的行政区特定年份的数据。如土地利用现状数据的地类图斑表,就可用 DLTB\_<行政区代码 1>\_<年份 1>、...、DLTB\_<行政区代码 n>\_<年份 m> 分别容纳各行政区各年份的地类图斑数据,并根据管理需要分别为每个表分配相应的物理存储。这种方式组织数据的好处是数据的布局与数据汇交模式一致,简化数据管理工作。然而将行政区代码和年份硬编码到表的命名中,难以对依据数据库标准开发的应用程序提供数据库透明支持,这样的数据库也难以大范围推广部署。

为了解决应用透明支持问题,可使用联合视图

① ESRI 公司, Building a Geodatabase。

② Oracle 公司, Oracle XML DB 开发者指南。

(UNION VIEW),即基于上述物理数据表,创建合并了所有行政区各年份数据的逻辑视图;应用程序直接存取的视图。但应用程序每次存取这些视图时,都要进行特别耗费计算资源的联合操作,数据库性能会大大降低;随着新数据的纳入,也需要进行视图重建等维护工作。

(2)使用分区进行数据布局。这种布局方式物理上与上述的数据布局相同,但逻辑上又与相关的数据库标准保持一致,满足数据管理工作的需要,支持应用程序对数据库的透明存取,支持大范围推广部署。

数据分区技术类似于分卷的档案组织,即将一个数据量庞大的数据表依据一定的规则人为分成多个分卷——分区(Partition),每个分卷又可以再进行分卷——子分区(Subpartition),并基于管理需要,为每个分区或子分区分配相应的物理存储。上述的地类图斑表就可以以年份作为主分区依据,即同一年份的地类图斑数据处于一个主分区中;以行政区代码作为子分区的依据,即同一年份的同一行政区的地类图斑数据处于一个子分区中(图1)。

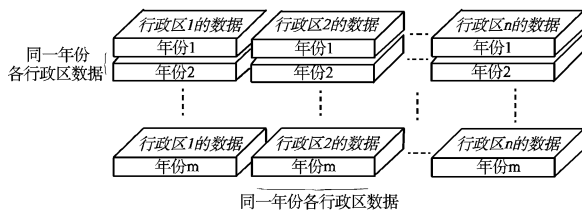


图1 使用分区的地类图斑数据布局

由图1可见,每个子分区对应于方法(1)中的各个表,所不同的是,在分区的布局方案中,对应用程序而言,只用一个数据表就容纳了所有年份所有行政区的数据,同时又保持了基于行政区和年度粒度的数据管理能力。

## 4 衍生数据的设置和管理

衍生数据是指国土资源数据库标准没有规定,但为了更好地支持数据管理和应用运行而由数据库管理系统创建或数据库构建者主动生成的数据。在Oracle数据库管理系统下,各种索引则是典型的由系统形成的衍生数据,数据库快照(即Materialized View:实体化视图)和数据库表则可用于存储服务于特定应用目的的衍生数据。

### 4.1 索引的设置和管理

索引用于提高应用系统检索数据的性能。在构建国土资源基础数据库时,设计人员往往是为实施唯一性约束而创建B-tree索引,却忽略了创建索引的最重要目的是提高数据库性能。Oracle数据库针对数据及其应用特点,提供了多种索引类型。就国土资源基础数据库而言,最有用的有B-tree索引、位图(Bitmap)索引。

#### 4.1.1 B-tree索引

B-tree索引用于提高事务处理的数据检索速度和实施唯一性约束。国土资源基础数据库有事务处理的主数据库,如矿业权数据库;有事务处理的支持数据库,如土地利用现状数据库(其内容不是记录事务处理的结果,而是记录专业的调查评价结果)。所以,在非事务处理的基础数据库上建立B-tree索引,往往没有意义,甚至会起反作用。因为B-tree索引数据是要大量占用存储空间的,并且在数据加载时,系统要消耗一定资源去同步维护B-tree索引数据。

#### 4.1.2 Bitmap索引

对于字段值重复度较大的字段,在其上创建位图索引,可以提高基于这些字段作检索条件的数据查询性能。如在地类图斑的地类代码字段上就可以建立这样的位图索引。位图索引的基本做法是:针对枚举型字段的各个可能的枚举值,建立相应的由二进制位构成的位图(如有3个枚举值的字段就有3个位图)。数据表中的每个记录在位图中对应1个二进制位,如记录中的字段取值等于位图对应的枚举值,则该二进制位被置为“1”,否则置为“0”。通过建立位图二进制位的顺序位置与记录的物理地址之间的映射关系,就可快速确定字段值等于或不同于给定枚举值的记录所处的物理地址集。

由于位图索引实际上并不存储字段值,位图中使用1个二进制位对应1个记录,所以位图索引一般都是很小的,并且可以基于几个位图进行与、或、非、异或的位逻辑运算,在不对记录值进行存取和判断的情况下,依据位逻辑运算的结果,就可以过滤掉绝大多数不符合检索条件的记录。在土地利用现状数据库中,像“行政区代码”、“界线类型”、“地类代码”等许多属性都是取值重复度很大的字段,应在其上创建相应的位图索引。

不管是何种索引,索引数据都要占据物理磁盘

空间。所以,与数据表一样,也应通过设置合适的存储参数、适当使用分区方法来布局索引数据。

### 4.2 实体化视图的设置和管理

实体化视图即数据库快照,它是存储了查询结果的数据库视图。实体化视图通过将大量耗时的数据库检索、计算操作的结果预存起来,应用程序直接使用这些存储的结果,极大提升应用程序的性能。然而,这又会带来 2 个问题,一是实体化视图如何适时体现其所引用的数据库的数据变化,使其内容得到相应的及时更新,使应用程序获得及时的数据支撑;二是应用程序能透明地使用实体化视图,而不需为此进行编码修改、测试等工作。为此,Oracle 数据库管理系统分别提供了实体化视图刷新 (Materialized View Refresh) 机制和查询重写 (Query Rewrite) 机制。

实体化视图刷新是指当数据库表的内容发生变化时,其上的实体化视图内容也按照其定义发生相应的变化。Oracle 数据库提供了按需的 (On Demand) 刷新/即时的 (On Commit) 刷新,快速的刷新 (Refresh Fast)/完全的刷新 (Refresh Complete)/基于分区追踪的刷新 (PCT Refresh) 等可选机制,满足实体化视图各种情况下的刷新要求。查询重写是利用实体化视图为应用程序提供透明支持的关键 (图 2)。

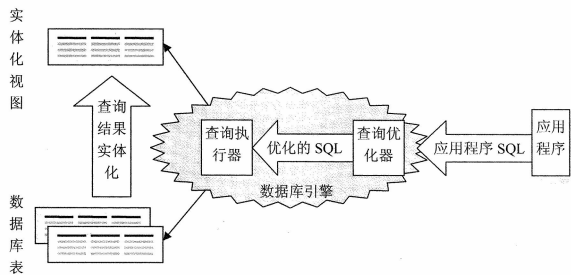


图 2 查询重写机制

当应用程序向数据库引擎发出查询请求时,查询优化器在对 SQL 语句进行优化时,发现有能满足此 SQL 要求的实体化视图可用,查询优化器改写应用程序所提交的 SQL 语句,由“从数据库表中获得结果”改为“从实体化视图中获得结果”,优化后的 SQL 语句交由查询执行器去执行。由此可见,数据库引擎已将应用程序针对数据库表的查询改为针对实体化视图的查询,达到对应用程序透明支持的目的。

在土地利用现状数据库中,一个利用实体化视图具有代表性的例子就是实现基于行政区的土地利用分类面积统计,其伪代码是:

```
CREATE MATERIALIZED VIEW land_area_sum AS
SELECT 行政区名称,地类名称,SUM(地块面积)
FROM 地类图斑表
GROUP BY ROLLUP("权属单位","街坊(村)","街道
(乡、镇)","区(县)","市","省")
ROLLUP("土地二级分类","土地一级分类")
```

当应用程序检索特定省、市的各土地分类面积时,数据库会利用查询重写机制使用上述建立的实体化视图,而应用程序除了感受到其性能有极大提升外,数据库已经为其作了查询的重定向工作。

实体化视图数据都要占据物理磁盘空间。所以,与数据表一样,也应通过设置合适的存储参数、适当使用分区方法来布局实体化视图数据。

## 5 数据完整性实施

国土资源基础数据主要有 2 种产生方式:一是作为国土资源事务管理的结果,如采矿权数据、探矿权数据、土地登记数据等;二是作为对有关部门提交的数据进行整合而形成的结果,如土地利用现状数据、矿产资源登记统计数据等。这 2 种方式产生的结果建立相应的基础数据库,实施数据完整性的目的和策略是有很大区别的。对于作为事务处理结果的基础数据库,实施数据完整性的目的是保证在事务处理过程中事务的完整性和数据的正确性。在 Oracle 数据库中,它是通过定义声明性的完整性约束规则和触发器,并使这些规则和触发器处于启用 (Enable) 状态而实施的。在事务数据作为正式结果写入库之前,数据库使用规则和触发器进行检查;对违背的数据,应用系统要么修改数据、要么回滚事务。显然,实施启用状态的完整性规则是要消耗计算资源的。

对于基于数据整合结果而建的基础数据库。数据完整性的保证是数据整合工作的一项主要目标,理想情况下整合成的数据就已经满足数据完整性的要求了。在这样的基础数据库中实施数据完整性似乎显得多此一举。但 Oracle 还使用完整性约束规则实施数据完整性,但规则的状态是验证 (Validate) 或停用 (Disable)。声明这些规则的目的不是实施检查,而是为 Oracle 数据库引擎提供数据检索优化

的依据。由于不实施真正的数据检查,也就谈不上消耗与此有关的数据资源。更重要的是,整合数据所占用的时间窗(Time Window)可以变得更小,这对提高基础数据库之上应用系统的连续性是有很大好处的。

## 6 结论

在 Oracle 数据库平台技术环境和国家与省级国土资源的数据环境下,应对国土资源基础数据库进行合理的物理设计,使用恰当的结构组织结构化、

半结构化和非结构化数据,对数据实体特别是大规模数据实体进行合理的物理布局,设置必要的衍生数据并进行合理布局,正确实施数据完整性约束要求。这样才能在满足数据标准的同时,提高基础数据的可管理性和应用支撑能力。

## 参考文献:

- [1] 萧文龙,陈怡如. Oracle10g 数据库最佳入门[J]. 北京:机械工业出版社,2007.

# Method on Organizing Basic Land and Resource Datas Based on Oracle

SHI Hui, FENG Yong - yu, ZHANG Yan, ZHANG Hong - lin

(Shandong Department of Land and Resources, Shandong Jinan 250014, China)

**Abstract:** Oracle database management system has become the mainstream technology platform of national and provincial land and resources database construction. Many countermeasures are put forward in this paper, such as how to organize and layout various types of datas physically and reasonably, how to set up and manage the derivative datas, and how to implement the datas integrity requirements. It will realize effective management and sustainable update support of basic land and resources database. Thus, it will meet the business application system requirements.

**Key words:** Physical design of database; data division; land and resource management