

# 基于 Oracle 9i 技术的气象资料服务系统存储管理设计与实现

## Storage Management of Meteorological Data Service System Design and Implement Based on ORACLE 9i Technology

**摘要:** 气象资料服务系统是一个面向气象业务和科研人员在线提供常用资料的网上气象资料服务。本文结合气象数据的特点和 Oracle9i 的技术特性阐述了一种气象数据的在线存储和管理的方法。

**关键词:** 资料服务 数据库 存储管理

邓 莉 (国家气象中心信息中心 100081)

国家气象中心气象资料室保存有我国最完整和系统的气象数据信息, 如何使气象部门很好地利用这一宝贵的数据资源, 极大地提高其利用率, 一直是我国气象工作者和广大科技人员的心愿, 但是长期以来, 国家在政策上缺少信息共享的法律法规, 尚未形成基础数据共享的良好机制, 目前国家气象中心各业务服务系统无法在业务上实时获得所需的基础历史数据, 而且所需的基础资料和数据处于分散的、部门所有的状态, 各个业务和服务系统各自建立许多内容重复的数据库, 既增加数据库的建立和管理成本, 加重通信负担, 又不能及时获得数据的支持。这在一定程度上影响了中心业务服务工作的发展。在此情况下, 我们集中了中心的各个业务服务系统共同所需的数据资源, 并对资料进行细致的加工、处理、统计的基础上, 建立了气象资料共享服务数据库, 及时和方便地提供正确的数据服务。

### 1 数据分析

气象资料服务系统提供了 14 种数据产品, 包括有:

- (1) 中国地面观测资料 (1951-今);
- (2) 中国地面累年值 (1971-2000年);
- (3) 全球地面气候标准值 (1961-1990年);
- (4) 中国天气现象资料 (1951-今);
- (5) 中国天气现象累年值 (1971-2000);
- (6) 中国高空探测规定层资料 (1954-今);
- (7) 中国高空探测规定层资料累年值 (1971-2000年);
- (8) 全球地面天气报资料 (1980-今);
- (9) 全球高空天气报资料 (1980-今);

- (10) 全球船舶天气报资料 (1980-今);
- (11) 中国农业气象旬月资料 (1991-今);
- (12) 地面天气图资料 (Micaps格式);
- (13) 高空天气图资料 (Micaps格式);
- (14) GMS 卫星云图资料。

14 种数据产品中既包括有常规气象观测站点资料, 也有图形图象资料。从资料的表象看既有结构化数据, 又有非结构化数据。因此在进行数据库设计和规划时我们需要从两个方面进行考虑: 第一是大数据量气象资料的存储方法, 第二是结构化数据与非结构化数据的集中统一管理方法。

从资料服务系统涉及的资料看, 资料种类非常丰富, 时间序列长, 数据量庞大, 其中中国地面观测资料包含了我国自建站以来的常规使用资料; 中国高空探测规定层资料是目前我国处理的最完整、时间最长的一份有很高使用价值的资料; 全球地面、高空和船舶资料是目前为止提供在线服务的时间跨度最长的资料。14 种资料初始数据量超过 150GB, 并且每月增量达 1.5GB。如何高效实现数据存储管理是一个非常突出的问题。

对于这些资料, 其应用方式是多样的。例如, 一些常规的站点气象观测资料, 需要为用户提供基于要素、时间、空间等多项条件组合的检索, 显然, 对这些类型资料可以直接转换为二维或多维数据关系结构进行管理; 而对天气图、卫星云图资料, 无法用二维或多维数据结构表征, 这些资料应用需求不是很复杂, 使用频率不高, 但是总体数据量大, 文件数量多, 导致数据管理难度增大, 如何把这两种表征的数据进行统一管理是需要我们解决的问题。

## 2 系统选型

通过对数据的分析,以及对目前市场上的数据库产品的深入调查,我们最终选择了ORACLE 9i数据库管理系统,作为资料服务的数据库管理和运行平台。采用IBM xServer 235 服务器, XEON 2.0双处理器, 300GB磁盘阵列作为系统硬件运行平台。

## 3 存储管理设计与实现

在Oracle9i产品中,提出了数据分区存储管理技术,它可以将大表分成可以管理的小块,从而避免对每个大表只能作为一个单独的对象进行管理。分区(partitioning)是一种“分而置之”的技术,为大量数据提供了可伸缩的性能。Oracle支持范围分区、列表分区、复合分区等更多的分区技术。其中范围分区,是按照一定范围条件划分的方式把数据分成小的分区。范围分区最典型的应用例子就是对时间或区域进行分区。范围分区可以方便地把数据按月、按年进行分区,或按区域进行分区。由于气象数据本身的特征要素就有一个时间和区域的概念,在使用资料时我们通常会调用最近一个月的降水资料;某天历史同期的所有气温资料;西北和华北地区某段时间的沙尘天气过程资料等,数据的存放和查询往往都与时间和区域相关。另外分区有分区忽略的功能,所有和本次查询无关的分区都不会参与运算,大大减少查询所需的时间。

结合Oracle分区技术,我们对每类资料进行了分析和统计,在数据库设计中采用范围分区技术,对频繁使用的资料,大数据量的资料实行分区管理的原则。对于具体实施,以下举几个典型的例子来说明。

(1) 中国地面资料:从使用频率上看,查询调用某个站点或某一区域多年资料的应用频率比较高,我们在数据表设计时采用区域分区管理这类资料,分区条件为台站号。我国的台站号编码从50XXX到59XXX,区域划分为10个区域,从50区到59区,建数据表中,对应的字段名为V01000。以下为中国地面定时观测数据表的创建脚本:

```
CREATE TABLE "CLIMATE"."A0101_TB" ("V01000" CHAR(5 byte)
NOT NULL, "V04001" NUMBER(4) NOT NULL, "V04002" NUMBER
(2) NOT NULL, "V04003" NUMBER(2) NOT NULL, "V04004" NUMBER
(2) NOT NULL, "V10004" NUMBER(5), "V12004" NUMBER(5), "V13004"
NUMBER(5), "V13003" NUMBER(5), "V11001" NUMBER(5), "V11002"
NUMBER(5), "V20010" NUMBER(5), "V20311" NUMBER(5), "V12231"
NUMBER(5))
```

```
PARTITION BY RANGE ("V01000")
```

```
(PARTITION "STA_50" VALUES LESS THAN ('51000')
TABLESPACE "SURF_HOUR_50",
```

```
PARTITION "STA_51" VALUES LESS THAN ('52000')
TABLESPACE "SURF_HOUR_51",
```

```
.....
PARTITION "STA_59" VALUES LESS THAN ('60000')
TABLESPACE "SURF_HOUR_59"
) ENABLE ROW MOVEMENT;
```

(2) 全球高空资料:从使用频率上看,查询调用某时次多年资料的使用频率比较高,我们在数据表设计时采用时间分区管理这类资料,也就是以年为分割条件,每5年建一个分区。创建脚本如下:

```
CREATE TABLE "CLIMATE"."B310101_TB" ("V01000" CHAR(5 byte),
"V05001" NUMBER(5), "V06001" NUMBER(6), "V01201" NUMBER
(4), "V04001" NUMBER(5), "V04002" NUMBER(5), "V04003" NUM-
BER(5), "V04004" NUMBER(5), "V07004" NUMBER(6), "V10003" NUM-
BER(6), "V12001" NUMBER(5), "V12003" NUMBER(5), "V11011" NUM-
BER(5), "V11012" NUMBER(5), "F07004" NUMBER(5), "F12001"
NUMBER(5), "F12003" NUMBER(5), "F11011" NUMBER(5), "F11012"
NUMBER(5))
```

```
PARTITION BY RANGE ("V04001")
```

```
(PARTITION "HIGH_GLO1980_4" VALUES LESS THAN (1985)
TABLESPACE "HIGH_GLO01_1980_4",
```

```
.....
PARTITION "HIGH_GLO1995_9" VALUES LESS THAN (2000)
TABLESPACE "HIGH_GLO01_1995_9",
```

```
PARTITION "HIGH_GLO2000_4" VALUES LESS THAN (2005)
TABLESPACE "HIGH_GLO01_2000_4")
```

```
ENABLE ROW MOVEMENT;
```

以上只列举了有代表性的两种资料的分区存储设计方法,通过对每类资料的应用特征进行统计分析,我们制定了详细的分区规划。总的来看,采用分区技术可以很容易对数据进行精确控制,实现大数据量表的管理。管理员可以指定每个分区的存储属性,分区在宿主文件系统中的放置情况,这样增加了对超大型数据库的控制力度。分区技术有多种方法,掌握每一种数据分区方法的特点,合理使用合适的数据分区方法,对于构筑海量、高性能的数据库平台,会奠定扎实的基础。

## 4 数据管理设计与实现

资料服务系统面向气象业务和科研人员,地面、高空、船舶、农气等站点观测资料,其应用十分频繁,不仅要求响应时间快,也要对资料的选择和查阅有很大的灵活性,而且在很多时候会出现多用户对同一类数据的并发检索;科研人员则需要对较长时间序列和多种相关资料进行统计与分析,因此,对这些类资料直接存放在数据库中进行管理,可以充分发挥数据库的优点,达到很好的应用效果。

对于天气图、卫星云图这些非结构化数据文件的管理, Oracle技

术提供两种数据类型blob和bfile进行管理:

(1) Blob方式: 数据表中存放元数据等结构化数据, 并以blob方式将其具体的非结构化数据存入数据库中, 文件对操作系统不可见。

(2) Bfile方式: 数据表中存放元数据等结构化数据, 具体的文件则以文件形式存放在操作系统上, 文件的存储、访问、管理通过bfile方式进行, bfile中存放的是相当于一个访问文件路径的指针。虽然文件对操作系统是可见的, 但是是由数据库对其进行管理。

从应用需求来看, 天气图资料、卫星云图资料等数据量大的资料、其使用频度低, 而且不需要复杂查询, 不需要文件内容处理, 可以不直接存储在数据库里, 而采用文件方式存储在数据库外部进行管理, 因此Bfile方式管理非结构化数据比较适合我们的应用。

在设计上需要在数据库内建立这些非结构化数据的元数据信息, 把管理的文件目录和有关数据信息纳入到数据库管理中, 在数据库内建立相应的索引机制, 文件内容本身存储在数据库外部, 便于数据的查询检索和元数据的统一管理。具体实施如下:

(3) 地面天气图 (Micaps) 资料:

存放目录: surfmic/year/month/ 一个月一个子目录存放资料

目录命名约定: surfmic: 为地面天气图存放目录

year: 为实际年份, 如2000, 2001

month: 为实际月份, 如01, 02.....

文件名: yymmddhh.000

文件命名约定: yy: 两位年份

mm: 两位月份

dd: 两位日期

hh: 两位小时

表 1 索引信息表: surmicbf\_tb

序号	中文名	字段名	数据类型
1	资料类别	Typeinfo	Char(30)
2	年	V04001	Smallint
3	月	V04002	Tinyint
4	日	V04003	Tinyint
5	时	V04004	tinyint
6	文件名	fname	CHAR ( 30 )
7	文件指针	F_po	BFILE

其中资料类别约定为: SURFACE\_PLOT

(4) 高空天气图 (Micaps) 资料:

存放目录: highmic/year/month/level 一个月一个层面一个子目录存放资料

目录命名约定: highmic: 为高空天气图存放目录

year: 为实际年份, 如2000, 2001

month: 为实际月份, 如01, 02.....

level: 为高空规定层面, 如1000, 925, 850

文件名: yymmddhh.000

文件命名约定: yy: 两位年份

mm: 两位月份

dd: 两位日期

hh: 两位小时

表 2 索引信息表: highmicbf\_tb

序号	中文名	字段名	数据类型
1	资料类别	Typeinfo	Char(30)
2	年	V04001	Smallint
3	月	V04002	Tinyint
4	日	V04003	Tinyint
5	时	V04004	tinyint
6	文件名	fname	CHAR ( 30 )
7	文件指针	F_po	BFILE

其中资料类别约定为: HIGH\_PLOT\_层次

层次共11层: 1000、925、850、700、500、400、300、250、200、150、100

卫星云图的数据管理设计与天气图相似, 在此不再详述。

我们通过以上各类结构化数据和非结构化数据的设计和把对数据库内数据的管理和对存放在文件中数据的管理很好地结合起来, 统一在Oracle数据库中, 实现了系统整体数据的集中统一管理。

## 5 总结

气象数据具有种类多, 格式复杂, 数据量庞大的特点, 在气象资料服务系统的建设过程中, 为实现高效的各类业务及科研用户的访问支持, 我们对每类资料的应用方式进行了系统分析, 设计了存储管理方案, 目前气象资料服务系统已投入业务运行, 为气象业务和科研用户提供资料服务。同时本系统的建立对正在建设中的国家级存储检索系统海量数据库设计起到很好的借鉴作用。

### 参考文献

- 1 Dave Ensor & Ian Stevenson Oracle Design O'Reilly & Associate, Inc
- 2 Anand Sdkoli & Rama Vepuri Oracle9i for Windows Handbook The McGraw-Hill Companies, Inc
- 3 Oracle9i Database Online Document Release 2(9.2)
- 4 Oracle 9i 数据库高级管理, 电子工业出版社。