

# 气象数据仓库及多维数据报表的设计与实现

**摘要：**本文概述了基于 SQL Server 的气象数据仓库以及气象数据报表的设计与实现，旨在将数据仓库技术应用于气象研究领域，利用联机分析处理和多维报表展示获取有益数据，用 SQL Server Reporting Services 平台生成气象报表，并以网页形式显示，为气象研究工作提供支持。

**关键字：**气象数据仓库，联机分析处理，气象数据报表。

**中图分类号：**TP311.13

**文献标识码：**A

## 1. 引言

随着通信技术和计算机技术的发展，每天在气象通信网上传输庞大的气象数据资料，对这些资料的存储、处理和应用的的问题显得越来越突出。将每个气象站的气象数据信息整合和汇总，有利于气象工作人员的分析及决策。同时气象工作人员对数据分析的要求越来越高，需要一个具有高度自动化、智能化、集成的联机分析处理系统。本文论述了基于 SQL Server 平台及气象数据库的数据仓库、数据报表的设计与创建过程，涉及 SQL Server 的组件 Server Analysis Services、SQL Server Reporting Services 的应用等，并给出一套较完整的从气象数据仓库到气象数据多维报表展示的方法。

## 2. 系统总体架构

### 2.1. 系统总体规划设计流程

数据库技术在气象领域广泛应用，通过计算机可以保存大量且结构复杂的数据。全国气象站每天都有大量的数据入库保存。随着科技的发展，气象数据在种类和数量上不断则增长，现有数据库管理系统已经不足以满足发展要求。数据仓库系统和数据仓库技术的应用不仅可以满足大容量的存储需求，而且它还可以在多年历史数据记录的基础上进行分析，得出有益于气象分析的结果，本系统设计实现流程如图 1 所示。

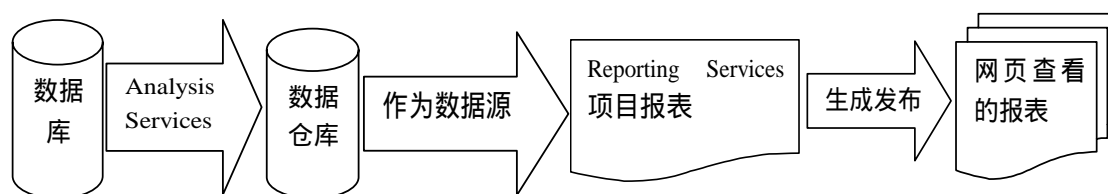


图 1 报表服务器实现总体规划流程

### 2.2. 气象数据仓库系统

数据仓库是一个面向主题 (Subject Oriented)、集成 (Integrated)、反映历史变化 (Time Variant)、相对稳定的 (Non-Volatile) 的数据集合，是支持管理部门的决策过程。可以以常用观测要素统计资料为主题，以星形模式建立多维数据模型，模型的核心是事实表，围绕事实表的是维度表。<sup>[1]</sup>简单多

维数据模型包括一个事实数据表,每个记录包含区站号、时间、降水量、气温、相对湿度、风速、日照等字段。维度表可以以时间、区站号建立,每个维度被划分为多个层次,例如:时间维度层次结构的顶层可以是年,下一层是季,然后是月,最后位于层次结构底层的是天。

粒度是数据仓库中保存数据的细化或综合程度的级别。细化程度越高,粒度级就越小;相反,细化程度越低,粒度级就越大。数据的粒度影响存放在数据仓库中的数据量的大小并且影响数据仓库所能回答的查询类型,在数据量大小与查询的详细程度之间要做出权衡<sup>[2]</sup>。例如对于气象信息中的站点高度信息,因站点分布不均匀,气象站的高度属性也不均匀,会出现某些数值段的站点信息较集中。而有的数值段几乎没有任何信息。这样的信息分布可能不符合某些分析方法的要求,可以将高度维度离散化为每 20 个值一段,这样就可以方便的按区域进行气象信息的统计了。

本气象数据仓库是使用 SQL Server Analysis Services 创建的,它为用户提供联机分析处理(OLAP)和数据挖掘功能。基于关系表的气象数据仓库存储方式为星型模型,因为这种存储方式比较成熟,适用于软件开发,查询速度快。相关的维度和粒度如下:

时间维。粒度水平可分为年、月、旬、日等。

区站维。粒度水平可分为具体区站信息、地区、经度、纬度等。

### 2.3. 以数据仓库为基础的数据展示

SQL Server Reporting Services 可以使用 Analysis Services 的联机分析处理结果通过自定义的组合,实现多维分析展示。用户可自定义维度、维度限制、统计方式等,并将结果通过网络发布。此过程中,Reporting Services 提供了很友好的图形用户界面,可以为用户简化大量操作。

Reporting Services 体系结构是一组集成的组件。该体系结构具有多层,其中包括应用程序层、服务器层和数据层。由于该体系结构是模块化的,且具有可伸缩性,也就是说 Reporting Services 是支持联机操作的。

### 2.4. 报表格式、查看方式多样,符合日常报表使用习惯

报表格式采用常用的矩阵报表格式,这样方便查看报表信息,直观地了解气象的变化。工作人员查看报表方式多样,可以在部署设置报表的行和列来进行查询,例如在报表中可把时间信息作为行,区站信息作为列,报表的整体架构符合日常报表的习惯。

## 3. 创建气象数据仓库并生成多维数据集

针对于气象数据库,设计多维的气象数据仓库。创建的气象数据仓库,设计适合 OLAP 的星型模型来存储数据。在该星型模型中,主要构建时间维、区站维维表。事实表与维度表的星型数据模型见图 2 所示。



图 2 事实表与维度表的星型数据模型

如图 2 所示左边表示日值气象数据情况，其中气象数据包括降水量、风速、气压、湿度、云量、日照、温度等，其他两表表示与之相关的信息，如时间、地点。每个维表都有自己的属性，维表与事实表之间通过维关键字相关联。通过这一模型，工作人员可以进行各种组合查询，同时数据仓库根据故障诊断的主题建立故障数据集市。

气象数据库用于存储大量的气象数据，它的设计是面向事务的，其中要尽量避免冗余。气象数据仓库面向主题，它在设计时必须要有意的加入一些冗余数据。气象数据库内存储的一般为当前工作需要使用到的气象数据，而气象数据仓库中存储的是历史数据。这表示，气象数据库与气象数据仓库存储的气象内容不同，而且气象数据仓库容量比气象数据库要大得多，存储的气象数据也要多。

气象数据信息是气象报表服务器系统中一类重要的数据，对这些数据进行深入分析，获得气候变化与不同时间和地点之间的联系，可以辅助调度人员进行决策和预报计划，从而保证现有的技术能更好的为气象应用领域服务。

气象数据仓库中拥有大量的气象数据信息，同时一个气象数据可能伴随多个维度数据，往往有时并不知道哪些信息是有用的，而只是将所有的数据都存入数据仓库中，这样加大了信息的处理量，影响了查询气象信息的效率。本文采用参数化查询报表信息，达到了减少数据处理量的目的。

在已经建立的气象数据仓库基础上，气象工作人员可以直接对该数据仓库进行一些简单的查询，例如统计、更新数据信息等。在气象数据仓库中，数据处理、存储和数据挖掘都是后台运行的，而直接面向工作人员的工作界面就是联机分析处理系统(OLAP)。<sup>[3]</sup> OLAP 是数据仓库系统的主要应用，支持复杂的分析操作，侧重决策支持，并且提供直观易懂的查询结果。在数据仓库

环节中已经设计出了适合 OLAP 的星型模型来存储数据，该数据仓库模型是基于多维数据的，该模型将数据转化成数据立方体(data cube)形式。

OLAP 在气象仓库中是共享多维信息的、针对气象数据问题进行联机数据访问和分析。它通过对气象信息各种情况的观察形式进行快速、稳定一致和交互性存取，允许气象管理决策人员对数据进行深入观察。

当原始数据量较大的时候，经过 SSAS 分析的数据可能并不适合直接进行展示。这些数据的维度密度不均匀导致用户不能很快发现其中变化规律。如图 3 展示的是没有经过离散化的 SSAS 分析结果，其中维度粒度很小，使数据之间差距不明显。

		LONGITUDE ▾									
		75. 2300	75. 2500	75. 4000	75. 9500	75. 9800	76. 1700	76. 7300	76. 7800	77. 2700	77. 2700
LATITUDE ▾	V13241	V13241	V13241	V13241	V13241	V13241	V13241	V13241	V13241	V13241	V13241
16. 5300											
16. 8300											
18. 2300											
18. 5000											
18. 6500											
18. 7500											
18. 7700											
18. 8000											
19. 0300											
19. 1000											
19. 2300											
19. 2700											
19. 3700											
19. 5200											

维度密度不均匀,当维度密度大的时候经常出现空白的统计结果。

图 3 SSAS 查询结果

为了使数据更容易被用户理解，在进行展示之前需要将数据进行分组离散化。这个过程是从数据库开始，按照原始数据大小进行离散化。离散化的每一个分组成为桶，离散化的方式按桶可分为：桶内数值等差分组、桶内数据条数等量分组、自动分组、聚类分析方法分组。其中除桶内数值等差分组外，其余分组方法 SSAS 已经提供，只需要在维度表中设置需要离散化的维度属性的相关属性即可。

关于桶内数值等差分组，可以在数据库的维度表中插入若干字符列，并通过 SQL 查询将这些列的值修改为响应的离散化值。

SQL 语句样例：

```
update dbo.staioninf
set
HEIGHT_5000=HEIGHT-HEIGHT%5000,
HEIGHT_2000=HEIGHT-HEIGHT%2000,
HEIGHT_1000=HEIGHT-HEIGHT%1000
```

离散结果如下图所示。

		LONGITUDE 10 Name ▾						
		70 - 79	80 - 89	90 - 99	100 - 109	110 - 119	120 - 129	130 - 139
LATITUDE 10 Name ▾	V13241	V13241	V13241	V13241	V13241	V13241	V13241	V13241
10 - 19					1230455.89999996	1171029.39999997		
20 - 29		442985.200000009	3335500.19999992	20404625.0000003	31137550.1000022	2926392.39999998		
30 - 39	273108.800000018	209486.499999999	3058815.49999996	16497499.0999987	28244054.1999994	7603560.29999967		
40 - 49	271831.10000001	2737826.10000003	573143.400000015	557992.600000018	7008762.89999961	24081187.4999991	2294453.89999	
50 - 59						1465965.59999981		
总计	544939.900000044	3390297.80000004	6967459.09999983	38690572.599999	67561396.8000013	36077105.7999985	2294453.89999	

图 4 SSAS 离散化结果展示

#### 4. 气象数据报表展示设计

在 Reporting Services 中，分析人员负责创建包括聚合数据、表达式和参数的报表。分析人员还负责为单位内的其他人员设计报表模型，以便将数据组织成便于创建报表的有意义的实体。在实际工作中，气象数据展示的报表运用了 SQL Server Reporting Services 的相关功能。SQL Server Reporting Service 中所开发出的报表简单易用，而且对于用户还提供了很多个性化的操作方式，包括按不同格式提供导出，提取、缩放、查询等。报表格式采用常用的矩阵报表格式，这种报表格式、查看方式多样，直观的了解气象的变化，符合日常报表使用习惯。

SQL Server Reporting Service 报表是为了满足气象工作人员需求而设计的，可制作符合效益、可靠的报表来提供最大的研究成果。

经过 Reporting Services 部署后，打开预览选项卡，等待系统执行完毕就可以查看已发布的报表。报表参数是自定义的，一般会把时间维度设成参数对象，本项目中还涉及了城市、区县、纬度值等。当然，可以根据用户的需要来定义所需的参数对象，如图 5 所示定义的报表参数。



图 5 查询 20-20 时降水量参数

图 6 查询粒度离散化后 20-20 时降水量的报表

Reporting Services 部署后生成网页报表，打开 IE 浏览器，地址为之前部署的项目 URL 键入的 <http://localhost/ReportServer>，点击要查询的 20-20 时降水量报表，执行模糊查询，如图 6 所示。

#### 5. 结束语

本文构建了基于数据仓库系统、OLAP 联机分析处理的决策支持系统平台，为决策支持系统在气象数据仓库的建立及应用提供了有效的手段，以 SQL Server Reporting Services 作为展示平台生成气象数据报表，针对目前气象数据，进行系统查询、分析，对其数据源进行加工提取，帮助气象工作人员方便快捷高效的查询所需气象数据信息。本文只给出了气象数据仓库的设计方案，进行了联机分析处理，过程中多使用的是常用维度设计。如果与用户进行进一步沟通后，按照用户的需求进行设计，那么得出的分析结果将更贴近工作实际，为用户提供更好支持。

#### 参考文献：

[1] 王红霞,王志伟,王培红. 数据仓库在气象上的应用研究[J]. 山西: 电脑开发与应用, 2006. 10-12.

[2] 王红霞,朱喜林,马季兰,王志伟. 气象数据仓库建立及数据统计与挖掘[J]. 山西: 太原理

- 工大学学报, 2006. 101 – 103.
- [3] 谭晓光. 数据仓库技术在天气预报决策中的应用探讨[J]. 北京: 应用气象学报, 2006. 325-332.
- [4] 范增禄. 省市级历史气象资料共享系统建设[A]. 河北: 信息技术在气象领域的开发应用论文集. 118-123.
- [5] 陈宝学, 俞经善, 关宏伟. 数据挖掘技术应用于天气预报的可行性研究[J]. 黑龙江: 应用科技, 2004. 48 – 50.
- [6] 邱声春. 数据挖掘和数据融合技术在天气预报和气象服务中的应用研究[J]. 山西: 山西气象, 2007. 34 – 36.
- [7] 焦飞, 黄天文, 何华庆. 数据挖掘技术在气温长期变化趋势预测中的应用[J]. 广东: 广东气象, 2006. 33 – 35, 39.
- [8] 林杰斌, 刘明德, 陈湘. 数据挖掘与OLAP 理论与务实[M]. 北京: 清华大学出版社, 2002.
- [9] 刘同明, 夏祖勋, 解洪成. 数据融合技术及其应用[M]. 北京: 国防工业出版社, 2000.
- [10] 张德龙, 张立, 苏和. 数据挖掘技术在气象信息存储中的应用[J]. 内蒙古: 内蒙古气象, 2008. 34-35.
- [11] Claudia Imhoff, Nicholas Gallempo, Jonathan G. Geiger. Mastering Data Warehouse Design: Relational and Dimensional Techniques[M]. Wiley, 2003.8
- [12] Joy Mundy. The Microsoft Data Warehouse Toolkit[M]. WILEY Wiley Publishing, Inc. 2007.8

# **The Design and Implementation of Meteorological Data Warehouse and Multidimensional Data Report**

## **Abstract**

**The article discusses the design and implementation of meteorological data warehouse as well as the meteorological data report based on the SQL Server. The purpose is applied the data warehouse technology in the meteorology research area. Using on-line analysis processing and the multidimensional report , we can gain some beneficial data. The system product meteorology report using SQL Server Reporting the Services platform, which can publish the report form being able to check with Web page form. The research provides the technology support for**

**meteorological phenomena studies.**

**Keywords: Meteorological data warehousing, online analytical processing, meteorological data report.**