

智能交通公用信息平台中分布式数据库构建研究

深圳市城市交通规划研究中心 彭珂珂 董吉军 文锦添

【摘要】 本文通过分析智能交通公用信息平台的构成、信息来源，展望了在平台上构建分布式数据库的可能性，并对于具体技术给予讨论和研究。对于智能交通公用信息平台这种大型系统来说，采用分布式数据库设计无疑是一条切实可行的道路。

【关键词】 智能交通公用信息平台，分布式，数据库，DBS

【Abstract】 Based on the analyses of the architecture and the sourcing nature of the information for the intelligent traffic information platform, this article elaborates the possibility of constructing a distributed database system (DBS) for the platform. It also produces a discussion on various detailed technologies for the system. For a large system of intelligent transport information platform for public information and planning support, it was no doubt that the adoption of a distributed database system design would be a practical way.

【Key words】 Intelligent Traffic Information Platform for Public use, Distributed, Database System, DBS

1 引言

随着社会经济的进一步发展，私人车辆的急剧增加带来了各种日益严重的交通问题（交通拥堵、环境污染等）。为解决这些交通问题，我国一直大力开展交通信息化建设，但是以往建立的信息系统隶属于不同部门和单位，建成之后彼此信息不能共享，数据格式难以统一，存储了大量重复冗余的数据，形成了众多的信息孤岛。随着经济和信息化建设的进一步发展，更多的单位需要更多的复合型数据服务，而这种服务往往涉及到更大范围内的业务和数据，因此国家在“十一五”信息化建设中，特别强调建立公用信息平台的重要性，在交通领域，智能交通公用信息平台恰恰是为了解决上述问题而提出建立的一个综合性信息服务平台。

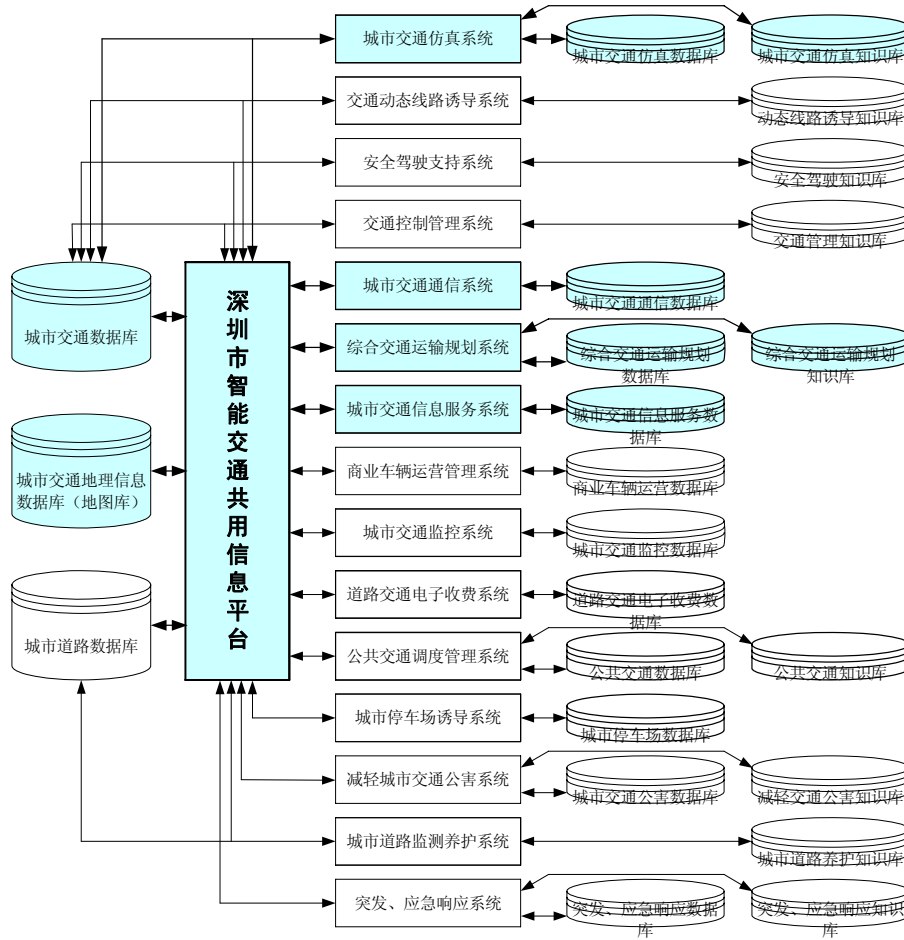
构建公用信息平台的数据库通常采用两种方式，一种是集中式，一种是分布式，根据当前技术发展趋势，分布式数据库系统在越来越多的系统中发挥着至关重要的作用。本文根据深圳市具体交通情况，分析讨论关于深圳智能交通公用信息平台中的关键技术—分布式数据库的构建。

2 智能交通公用信息平台

2.1 平台构成

深圳市智能交通公用信息平台是整个深圳市 ITS 实施的“中心”与“枢纽”，它承担着向各业务系统提供公共信息和中转信息的功能。智能交通系统所覆盖的各业务系统，将信息按照一定的接口和编码规则传输给城市智能交通公用信息平台，同时各业务系统再根据自己的需求，从城市智能交通公用信息平台中获取所需的公共信息。平台采用层次划分的方法对用户授予不同的获取权限，同时又确保数据传输过程的安全性及已有的信息实现共享，发挥信息共享的作用，从而提高信息利用率。

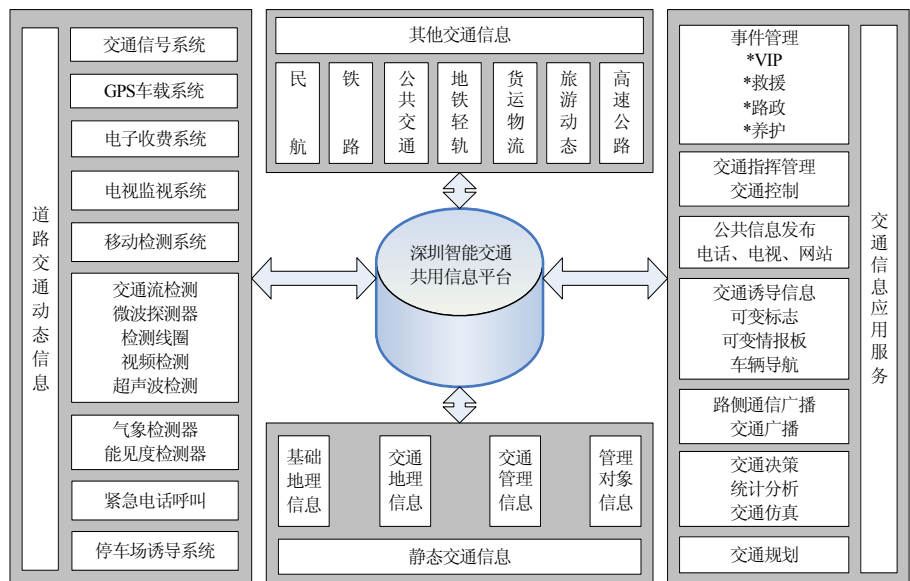
智能交通公用信息平台的构成如下图所示。



2.2 信息来源

按照信息论划分，整个智能交通系统按信息流程可简化为信息采集、信息传输、信息处理、信息服务四大部分。智能交通公用信息平台的信息来源非常复杂，涉及到地理信息、道路信息、经济人口数据、各业务系统数据等。

智能交通公用信息平台的信息来源如右图所示。



2.3 分布的数据和应用

由以上分析看出，公用信息平台构成是比较复杂的，数据来源是多源的，数据源采用了不同数据库构建技术。当前网络通讯技术已经有了突飞猛进的发展，传统的网络已经在向万兆速率甚至亿兆逼近，以往难以解决的通信问题早已经成为历史。而分布式数据库技术恰恰是分布式技术与数据库技术的结合。分布式数据库是物理上分散在计算机网络各结点上，而逻辑上属于同一个系统的数据集合。它具有数据的分布性和数据库间的协调性两大特点。采用分布式数据库系统的好处在于：适应分布式的组织结构，降低费用；提高系统的可靠性和可用性；充分利用数据库资源，提高现有集中式数据库的利用率；系统伸缩性高，可以逐步扩展处理能力和系统规模。

3 分布式数据库原理

3.1 定义

分布式数据库是由多组数据组成的，这些数据分布在计算机网络的计算机上，网络中的每个节点具有独立处理的能力（称为场地自治），可以执行局部应用。同时，每个节点也能通过网络通信子系统执行全局应用。所谓局部应用就是指应用仅涉及到局部或者自身数据库的应用。所谓全局应用就是指应用中涉及到两个或两个以上节点的数据库。每一个场地就是一个独立的数据库系统，它有自己的数据库、自己的用户、自己的 CPU，执行局部应用，具有高度的自治性。

3.2 特点

(a)不强调集中控制，具有一个以全局数据库管理员为基础的分层控制结构，每个局部数据库管理员都具有高度的自主权。

(b)强调数据独立性和分布式透明性。所谓分布式透明性就是在编写程序时好象数据没有被分布一样，因此把数据进行转移不会影响程序的正确性。但程序的执行速度会有所降低。

(c)与集中式数据库系统不同，数据冗余在分布式系统中被看作是所需要的特性，其原因在于：首先，如果在需要的节点复制数据，则可以提高局部的应用性。其次，当某节点发生故障时，可以操作其它节点上的复制数据，因此这可以增加系统的有效性。当然，在分布式系统中对最佳冗余度的评价是很复杂的。

3.3 数据存储

分布式数据存储可以从数据分配和数据分片两个角度考察。在分布式数据库中，数据存储通过以下三种途径实现：(a)复制：系统维护关系的几个完全相同的副本，这些副本存储在不同的结点上。(b)分片：关系被划分为几个片段，各个片段存储在不同的结点上。(c)复制+分片：关系被划分为几个片段，系统为每个片段维护几个副本。

数据分配是指数据在计算机网络各场地上的分配策略。包括：(a)集中式：所有数据均安排在同一场地上。(b)分割式：所有数据只有一份，分别被安置在若干个场地。(c)全复制式：数据在每个场地重复存储。(d)混合式：数据库分成若干相交的子集，每一子集安置在一个或多个场地上，但是每一场地未必保存全部数据。

数据分片是指数据存放单位不是全部关系，而是关系的一个片段，也就是关系的一部分。可分为如下三种：**(a)水平分片**：按一定的条件把全局关系的所有元组划分成若干不相交的子集，每个子集为关系的一个片段。**(b)垂直分片**：把一个全局关系的属性集分成若干子集，并在这些子集上进行投影运算，每个投影为垂直分片。**(c)混合型分片**：将水平分片与垂直分片方式综合使用则为混合型分片。

数据分片应遵循的原则为：**(a)完备性条件**：必须把全局关系的所有数据映射到各个片段中，绝不允许发生属于全局关系的某个数据不属于任何一个片段的情况。**(b)重构条件**：划分所采用的方法必须确保能够由各个片段重建全局关系。**(c)不相交条件**：要求一个全局关系被划分后得到的各个数据片段互相不重叠。

3.4 系统体系结构

分布式数据库的体系结构分为 4 级：全局外模式、全局概念模式、分片模式和分布模式。**全局外模式**：全局应用的用户视图，是全局概念模式的子集。**全局概念模式**：定义了分布式数据库中所有数据的逻辑结构。**分片模式**：定义片段以及定义全局关系与片段之间的映像。这种映像是一对多的，即每个片段来自于一个全局关系，而一个全局关系可分成多个片段。**分布模式**：片段是全局关系的逻辑部分，一个片段在物理上可以分配到网络的不同节点上。分布模式根据数据分配策略的选择定义片段的存放场地。

分布式数据库的分层体系结构的三个特征是：**(a)数据分片和数据分配概念的分离**，“数据分布独立性”。**(b)数据冗余的显式控制**。**(c)局部 DBMS 的独立性**。

3.5 分布透明性

分布透明性是指用户不必关心数据的逻辑分片，不必关心数据物理位置分配的细节，也不必关心各个场地上的数据库数据模型。分布透明性可归入物理独立性的范围。分布透明性包括 3 个层次：分片透明性、位置透明性和局部数据模型透明性。

3.6 系统面临的问题

集中式系统主要目标是减少对磁盘的访问次数。对于分布式系统，压倒一切的性能目标是使通过网络传送信息的次数和数据量最小。

4 智能交通公用信息平台数据库构建

4.1 系统数据库现状

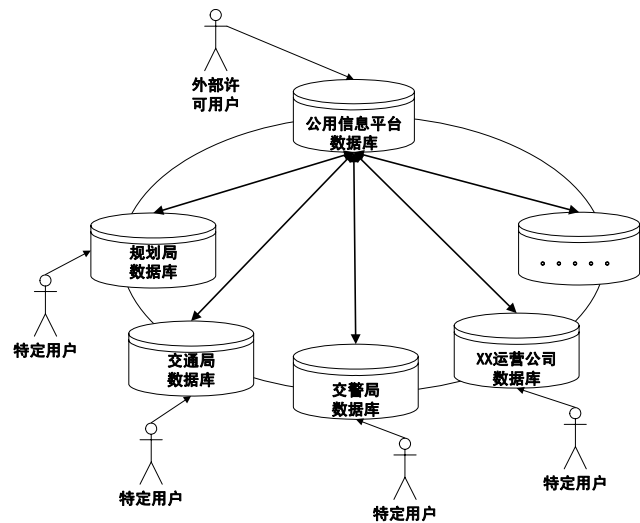
目前深圳市智能交通公用信息平台下已基本建成的数据库有：城市道路数据库、城市交通数据库、道路交通地理信息数据库等。但由于历史原因，这些数据库位于不同的地理位置和由不同的单位管理。比如交通警察局管理监控录像和 CCTV 数据，交通局管理公交和交通运营数据，规划局拥有道路网络、规划网络图、交通地理信息、规划方案数据，这些数据都是通过各单位自己建立的系统收集和处理，并且各个系统都是建立在不同的运行环境下，同时在系统设计时彼此之间没有进行充分的协调，因此未能形成良好的数据信息共享条件。

因此，要构建公用信息平台数据库必须考虑一下问题：**a) 业务重组**；**b) 信息源、存储和共享方式**；**c) 操作权限的控制**；**d) 数据库的异构模型**；**e) 网络规划**等问题。

4.2 构建原则

自下而上和自上而下两个原则的结合。对于现有的已经存在处于运行中的数据库系统要采取第一个原则，将这些数据库组合成一个物理上分布，逻辑上整合成一体的数据网络，主要构建思路：建设平台中心数据库作为整个系统数据核心点，采取星形+环形的连接方式，保证各业务库均与中心数据库连通以及业务库之间能形成一个巨大的闭环。

智能交通公用信息平台的分布式数据库具体构建示意如右图。



4.3 数据库职责说明

公用信息平台数据库为中心数据库，主要职责如下：

(a)制定数据规范和接口标准。

(b)对各个业务数据库数据进行抽取、融合和备份；同时向各个业务数据库发送基础编码类数据信息。

(c)为外部授权许可用户提供综合类信息服务。

各个业务数据库主要职责如下：

(a)负责原有业务的运营管理。

(b)同时根据相关规则，将本地数据进行抽取并上传给中心数据库。

4.4 数据库存储信息说明

根据系统的功能需求，智能交通公用信息平台中的数据及其使用情况分为两类：

第一类：中心数据库。基本地理数据（地形图、红线、宗地、法定图则、街坊图形），基本交通数据（路网、交通流量、基本交通仿真模型等），深圳基本情况数据（人口、经济、车辆增长等），基本权限、系统配置数据、各业务数据库使用频率较高的数据备份（根据业务需要和发展情况进行扩充）等。

第二类：业务数据库。在各单位现有的数据库系统基础上，在满足所在单位的业务处理的前提下，进行优化调整和扩充，以满足整体信息系统的需要。

在中心数据库失效的情况下，业务数据库仍旧可以连通，以保证正常业务的基本使用。通过客户端修改的重要数据，如地形图、法定图则要定期更新到中心数据库中去，以保证正常的系统运转。

4.5 分布式处理说明

4.5.1 数据同步

原则上原有系统的业务处理基本保持不变，增加数据库间的数据同步和一致性处理功能，具体如下：

对于第一类数据，在中心数据库中建立基础数据表，在分枝数据库中建立对应的快照和视图，对这些数据的修改在中心系统端进行，业务数据库通过视图修改这些数据，通过快照查询这些数据。

对于第二类数据，数据存放在业务数据库中，在中心数据库中分别建立对应的业务系统的数据库连接，中心数据库通过视图创建或修改这些数据，通过快照查询这些数据；业务系统的相应客户端直接对业务数据进行操作。

4.5.2 数据查询

任何一个客户端在得到授权的情况下，可以对多项业务进行复合查询，查询界面相当复杂。主要有以下几种情况：

①局部信息查询：一般只针对业务性查询而言，即特定的用户通过业务系统进入业务数据库，仅获取业务数据库的信息，该项查询方式和集中数据库方式功能一致。相当于原有的数据库查询功能。

②整体信息查询：用户直接连接到公用平台中心数据库系统，获取数据信息，中心通过分布式数据库同步机制向用户提供信息，这些对用户来说是透明的。

通过局部信息查询方式以确保现有查询方式，以满足单一业务数据库系统的特殊需要；而整体信息查询功能则大大提高了数据共享和综合信息服务能力，从而提高整个智能交通公用信息平台对外信息综合服务水平。

5 结论和建议

对于智能交通公用信息平台这种既有非空间数据又有空间数据的大型系统来说，采用分布式数据库设计无疑是一条可行的道路。但是要想真正做好这样的系统，还得在系统规划、数据库构建方面作进一步深入研究，才能真正地将分布式数据库的优点发挥出来。

参考文献

- [1] M.Tamer Ozsü & Patrick Valduriez, Principles of Distributed Database Systems(Second Edition), 清华大学出版社/Prentice Hall, 2002
- [2] Andrew S. Tanenbaum, Maarten Van Steen, Distributed systems:principles and paradigms, Tsinghua University/ Prentice Hall, 2002
- [3] Jie Wu, 分布式系统设计, 机械工业出版社, 2000
- [4] Sarah Brown 著, Object Design of a Distributed Client/Server System, IEEE, 1998
- [5] Kiril Boyanov 著,Parallel and Distributed Processing, ELSEVIER Science Publishers, 1991
- [6] Hackathorn, Richard 著, Emerging Architectures for Database Connectivity, InfoDB, January, 1991
- [7] Elmasri, Ramez and Navathe, Shamkant B 著, Fundamentals of Database Systems, 2nd Ed. Benjamin, 1994
- [8] 萨师焯, 王珊, 数据库系统概论(第三版), 高等教育出版社, 2000
- [9] Date C J, An Introduction to Database System(Ed.6), Addison-Wesley, 1995