

文档数据库与关系数据库的比较 PDF转换可能丢失图片或格式，建议阅读原文

[https://www.100test.com/kao\\_ti2020/141/2021\\_2022\\_\\_E6\\_96\\_87\\_E6\\_A1\\_A3\\_E6\\_95\\_B0\\_E6\\_c29\\_141425.htm](https://www.100test.com/kao_ti2020/141/2021_2022__E6_96_87_E6_A1_A3_E6_95_B0_E6_c29_141425.htm)

摘要：随着应用领域的不断拓展和多媒体技术,人们发现关系数据库的许多限制和不足,因而数据库技术进入了“后关系数据库时代”。文档数据库由此应运而生。本文概要地从数据格式、数据库结构和WEB发布数据三个方面比较了文档数据库与关系数据库的异同,同时区别了文档数据库与过去存储数据的文件系统的不同。关键词：数据库 Internet 文档数据库 WEB数据库服务器

一. 关系数据库的优势和目前面临的问题 从60年代末开始,数据库技术经历了层次数据库、网状数据库和关系数据库而进入数据库管理系统(DBMS)阶段至今,数据库技术的研究也不断取得进展。80年代,关系数据库成为发展的主流,几乎所有新推出的DBMS产品都是关系型的。关系型数据库在计算机数据管理的发展史上是一个重要的里程碑,这种数据库具有数据结构化、最低冗余度、较高的程序与数据独立性、易于扩充、易于编制应用程序等优点,目前较大的信息系统都是建立在结构化数据库设计之上的。来源：www.examda.com

然而,随着网络技术和软件技术的飞速发展,特别是Internet和Intranet技术的发展,使得非结构化数据的应用日趋扩大。关系数据库从1970年发展至今,虽功能日趋完善,但对数据类型的处理只局限于数字、字符等,对多媒体信息的处理只是停留在简单的二进制代码文件的存储。然而,随着用户应用需求的提高、硬件技术的发展和Intranet/Internet提供的多彩的多媒体交流方式,用户对多媒体处理的要求从简单的存储上升为识

别、检索和深入加工,正是用户呼唤出"通用"数据库服务器来处理占信息总量70%的声音、图像、时间序列信号和视频等复杂数据类型。据有关数据,1996年,全球数据库市场的销售额已接近40亿美元,预计到2000年,数据库市场销售额将达到80亿美元。巨大的增长潜力来自复杂数据类型的处理需求,使处理复杂数据类型的"超媒体"数据库将成为各公司投资研发的重点。

二. 全新的文档数据概念 从1989年起, Lotus通过其群件产品Notes提出了数据库技术的全新概念-"文档数据库", 文档数据库区别于传统的其它数据库, 它是用来管理文档。在传统的数据库中,信息被分割成离散的数据段, 而在文档数据库中, 文档是处理信息的基本单位。一文档可以很长、很复杂、可以无结构, 与字处理文档类似。文档数据库与五、六十年代管理数据的文件系统不同, 文档数据库仍属于数据库范畴。首先, 文件系统中的文件基本上对应于某个应用程序。当不同的应用程序所需要的数据有部分相同时, 也必须建立各自的文件, 而不能共享数据, 而文档数据库可以共享相同的数据。因此, 文件系统比文档数据库数据冗余度更大, 更浪费存储空间, 且更难于管理维护。其次, 文件系统中的文件是为某一特定应用服务的, 所以, 要想对现有的数据再增加一些新的应用是很困难的, 系统不容易扩充。数据和程序缺乏独立性。而文档数据库具有数据的物理独立性和逻辑独立性, 数据和程序分离。来源: [www.examda.com](http://www.examda.com) 文档数据库也不同于关系数据库, 关系数据库是高度结构化的, 而Notes的文档数据库允许创建许多不同类型的非结构化的或任意格式的字段, 与关系数据库的主要不同在于, 它不提供对参数完整性和分布事务的支持, 但和关系数据库也不是相互排斥的, 它们之

间可以相互交换数据,从而相互补充、扩展。三. 关系数据库和文档数据库的异同 下面从三个方面比较两种数据库的异同:

1. 格式文本/多媒体 在信息时代,所有信息大体上可以分为两类:一类信息能够用数据或统一的结构加以表示。例如,成本、费用、人员、工资以及员工业绩等,都可以用数字或文字来描述或表达。这类信息具有相同的层次或网络结构,我们称之为结构化数据;而另一类信息根本无法用数字或者统一的结构表示,例如,图象、声音等,我们称之为非结构化数据。这些非结构化数据既可以是一段包含大量附加信息(如排版信息)的文字信息,也可能是一段声音、图象,甚至是影像。对于结构化数据和非结构化数据的关系如图所示,非结构化数据包括结构化数据,但又不止是结构化数据;结构化数据属于非结构化数据,是非结构化数据的特例。一般来说,在人们的认识之中,关系数据库最主要的特征就是数据的结构化。然而,随着网络技术和软件技术的飞速发展,特别是Internet和Intranet技术的发展,使得非结构化数据的应用日趋扩大。关系数据库是高度结构化的,这种数据结构化使关系数据库具有冗余度最低、程序与数据独立性较高、易于扩充、易于编制应用程序的特点。但是,随着应用领域的不断拓展,为满足应用对数据处理不断"苛刻"的要求,人们开始发现关系数据库的许多限制和不足。因为文档数据库的基本元素就是文档本身,而数据库中的文档可以同时包含结构化的和非结构化的信息,所以,文档数据库能够存储和管理类似文档这样的非结构化数据。特别是,Notes的对象库是一个非常理想的商业信息存储器,可用于高效地存储、传播、分配和管理这类信息。这类信息通常具有丰富的数据类型

，如表格（可以从某个关系数据库或电子表软件中得到的）、格式化文本、WWW的页面、图形、OLE对象、或扫描的图象以及传真件、声频或视频信号这样的多媒体信息。2. 在结构上 关系数据库是应用数学方法来处理数据库数据的，其数据模型也是建立在数学概念基础上的，在关系模型中，数据在用户观点下的逻辑结构就是一张二维表。而Lotus Notes是一个文档数据库管理系统，文档数据库的基本元素就是文档。这里的文档和关系数据库中的记录相似。Notes文档的结构是由表单（form）定义的，而表单由一组各式各样的字段域组成。域（Field）、域（Domain）和字段（Field）在关系数据库中，域（Domain）是属性值的集合，如：大于0小于150的正整数，长度小于25的字符串集合等等。而在文档数据库中，域（Field）的概念出现在表单对象中，域是表单上存储数据的单个元素，域决定了一个独立的文档能包含什么数据。虽二者的中文译文相同，但其实际意义和作用相差千里。反而在文档数据库中域的概念与关系数据库中的字段（Field）相类似。条目（Item）和属性（Attribute）在关系数据库中，有属性（Attribute）的概念，即二维表中的每一列称为一个属性，给每一列起一个名称即属性名。而在文档数据库中，就没有“属性”，文档拥有的是条目（Item），条目是指存储于文档中的任意数据部分。每一个条目代表文档中一段数据，在用户界面中，是通过表单中的域来显示文档中的条目的。二者虽命名不同，但笔者认为从“属性”的角度更容易理解文档数据库中条目的意义。视图（View）在关系数据库和文档数据库中，均有视图的概念。在关系数据库中，视图是指从一个或几个基本表（或视图）导出的表。视图

和基本表不同，视图是一个虚表，即视图所对应的数据不实际存储在数据库中，数据库中只存储视图的定义（存在数据字典中）。在文档数据库中，用户通过Notes视图浏览文档。视图是很方便的目录，用户可以从其中看到关于文档的概要信息和文档的状态，然后存取特定的Notes文档。文档数据库中文档可以在视图中显示起全部或部分内容，通过视图，用户可以看一组文档的关键域，并可按某一准则对显示的信息进行分类和排序。实质上，二者逻辑上非常相似，即视图所对应的数据均不实际存储在数据库中，数据库中只存储视图的定义。100Test 下载频道开通，各类考试题目直接下载。详细请访问 [www.100test.com](http://www.100test.com)