

如何选择数据库产品

背景：如何为企业应用系统选择后台数据库产品？

答案：合适的才是最好的。

项目开始进行需求调研了，在谈到使用哪个数据库软件时，客户咨询 Smith，现在一般企业应用系统使用什么数据库呢？目前市场流行的是哪些数据库管理系统呢？Smith 回答：现在主流的中大型数据库产品有 Oracle、SQL Server、DB2、Sybase 等，也有一些小型的，如 Access、MySQL 等，具体应用系统该用什么数据库产品，不是单纯考虑产品的价格或者性能因素，而需要考虑很多综合因素，首先在了解数据库基础以后，还应该进一步了解这些数据库产品的特点吧。

1.1 数据库基础

1.1.1 数据库技术发展概述

(1) 数据库技术起源于 20 世纪 60 年代，与计算机技术的起源类似，首先应用于军事领域，以网络模型数据库系统规范报告 DBTG 为标志，使数据库系统开始走向规范化和标准化。

(2) 在 20 世纪 70 年代到 80 年代得到了比较迅速的发展，其中，影响至今的关系数据库系统基本理论得到了全面的研究与诠释，David Marker 所著的《关系数据库理论》一书，标志着数据库在理论上的成熟，同时数据库设计的基本过程基本稳定，分为需求分析、概念设计、逻辑设计和物理设计等。

(3) 从 20 世纪 80 年代以后的数十年来，数据库技术和计算机网络技术的发展相互渗透，相互促进，已成为当今计算机领域发展迅速、应用广泛的两大领域。

(4) 到 20 世纪 90 年代，随着基于 PC 的客户机/服务器计算模式和企业软件包的广泛采用，数据管理的变革基本完成。数据管理不再仅仅是存储和管理数据，而转变成用户所需要的各种数据管理的方式，数据库技术进入数据仓库时代。数据仓库技术是为了有效地把操作型数据集成到统一的环境中以提供决策型数据访问的各种技术和模块的总称，所做的一切都是为了用户更快更方便地查询所需要的信息，提供决策支持。

(5) 在近几年，随着云时代的来临，大数据也吸引了越来越多的关注，大数据的特点有 4 个层面：第一，数据体量巨大，从 TB 级别，跃升到 PB 级别；第二，数据类型繁多，如网络日志、视频、图片、地理位置信息等；第三，价值密度低，商业价值高，以视频为例，连续不

间断监控过程中,可能有用的数据仅仅有一两秒;第四,处理速度快,一秒定律。最后这一点也和传统的数据挖掘技术有着本质的不同。业界将其归纳为4个“V”——Volume, Variety, Value, Velocity。物联网、云计算、移动互联网、车联网、手机、平板计算机、PC以及遍布地球各个角落的各种各样的传感器,无一不是数据来源或者承载的方式。

1.1.2 基本概念与常用术语

1) 信息和数据

“数据”和“信息”这两个词常常被用做同义词,但是它们有不同的意义。

信息(Information)是现实世界在人们头脑中的反映,它以文字、数据、符号、声音、图像等形式记录下来,经传递和处理,为人们的生产、建设、管理等提供依据。

数据(Data)是指输入到计算机中并能被计算机进行处理的数字、文字、声音、图像、视频等符号。数据是对客观现象的表示,本身没有意义。

在计算机数据库中需要存储的是数据。信息通常是对数据进行结合、比较与计算的结果。

以测试成绩为例。如果一个班级中的每名同学都得到一个分数,则通过这些分数可以计算出班级平均分数,然后通过班级平均分数又可以计算出学校的平均分数。数据库软件可以将记录/存储的数据和统计数据转换成有用的信息。

数据:每名学生的测试分数都是一项数据。信息:班级的平均分数或学校的平均分数。

2) 数据库

数据库(Database)是依照某种数据模型来组织、存储和管理数据的仓库。同时这些数据以一定方式存储在一起、能为多个用户共享、具有尽可能小的冗余度、与应用程序彼此独立。从发展的历史来看,数据库是数据管理的高级阶段,它是由文件管理系统发展起来的。其主要特点是实现了数据共享、减少了数据的冗余度、提高了数据的独立性、数据实现了集中控制、增强了数据一致性和可维护性、提供了故障恢复等。

3) 数据库管理系统

数据库管理系统(DBMS)是位于用户和操作系统之间的一层数据管理软件,它由系统运行控制程序、语言翻译程序和一组公用程序组成。其功能主要包括以下方面:数据的定义功能、数据的操纵功能、数据库的控制功能、数据库的维护功能、数据字典。现在常见的DBMS有MS SQL Server、Oracle、Sybase、DB2、MySQL、Access、Visual FoxPro、Informix、PostgreSQL。下面会对主要的产品进行介绍。

4) 数据库系统

数据库系统(DBS)是由数据库及其管理软件组成的系统。它是为适应数据处理的需要而发展起来的一种较为理想的数据处理系统,也是一个实际可运行的存储、维护和应用系统提供数据的软件系统,是存储介质、处理对象和管理系统的集合体。广义地讲,它包括硬件、操作系统、数据库、数据库管理系统、应用程序和各类相关人员(开发人员、数据库管理员、用户等)。

5) 数据模型

数据模型(Data Model)按不同的应用层次分成3种类型:分别是概念数据模型、逻辑数据模型、物理数据模型。

概念模型(Conceptual Data Model)是面向数据库用户的实现世界的模型,主要用来描述

世界的概念化结构，它使数据库的设计人员在设计的初始阶段，摆脱计算机系统及 DBMS 的具体技术问题，集中精力分析数据以及数据之间的联系等，与具体的数据管理系统无关。概念数据模型必须转换成逻辑数据模型，才能在 DBMS 中实现。

概念模型用于信息世界的建模，一方面，应该具有较强的语义表达能力，能够方便直接表达应用中的各种语义知识；另一方面，它还应该简单、清晰、易于用户理解。

在概念数据模型中有 E-R 模型、扩充的 E-R 模型、面向对象模型及谓词模型，目前最常用的是 E-R 模型，常用 Sybase 公司的 PowerDesign 工具来建立 E-R 图，Oracle 公司也推出了专门用于 E-R 建模的 Data Modeler 工具，这个工具的特点是方便好用、免费。

逻辑模型 (Logical Data Model) 是用户从数据库看到的模型，是具体的 DBMS 所支持的数据模型，如网状数据模型 (Network Data Model)、层次数据模型 (Hierarchical Data Model)、关系数据模型 (Relational Data Model) 等。此模型既要面向用户，又要面向系统，主要用于数据库管理系统的实现。

物理模型 (Physical Data Model) 是面向计算机物理表示的模型，描述了数据在存储介质上的组织结构，它不但与具体的 DBMS 有关，还与操作系统和硬件有关。每一种逻辑数据模型在实现时都有其对应的物理数据模型。DBMS 为了保证其独立性与可移植性，大部分物理数据模型的实现工作由系统自动完成，而设计者只设计索引、聚集等特殊结构。

层次模型、网状模型和关系模型是 3 种重要的数据模型。这 3 种模型是按其数据结构而命名的。前两种采用格式化的结构。在这类结构中实体用记录型表示，而记录型抽象为图的顶点。记录型之间的联系抽象为顶点间的连接弧。整个数据结构与图相对应。对应于树形图的数据模型为层次模型；对应于网状图的数据模型为网状模型。其中应用最广泛的是关系模型，关系数据库是指对应于一个关系模型的所有关系的集合。

关系数据库管理系统就是管理关系数据库的管理软件，简称 RDBMS。

1.1.3 关系数据库

关系模型的数据结构非常单一。在关系模型中，现实世界的实体以及实体间的各种联系均用关系来表示。关系数据库 (Relational Database) 是建立在关系数据库模型基础上的数据库，借助于集合代数等概念和方法来处理数据库中的数据。

在用户看来，关系模型中数据的逻辑结构是一张二维数据表。关系中的每一行称为一个元组，或一个记录；每一列称为一个属性，或者字段。对于每一个关系可以给它一个唯一标识这个关系的名称，称为关系名或者表名。对于每一列唯一标识该列的名称，称为属性名或字段名。

关系数据完整性有 4 种类型：实体完整性、域完整性、引用完整性和用户定义完整性。

(1) 实体完整性将行定义为特定表的唯一实体。实体完整性强制标识表的符列或主键的完整性 (通过索引、UNIQUE 约束、PRIMARY KEY 约束或 IDENTITY 属性标识)。

(2) 域完整性是指给定列的输入有效性。强制域有效性的方法有：限制类型 (通过数据类型) 格式 (通过 CHECK 约束和规则) 或可能值的范围 (通过 FOREIGN KEY 约束、CHECK 约束、DEFAULT 定义、NOT NULL 定义和规则)。

(3) 在输入或删除记录时，引用完整性保持表之间已定义的关系，引用完整性基于外键与主键之间或外键与唯一键之间的关系 (通过 FOREIGN KEY 和 CHECK 约束)，引用完整性确

保键值在所有表中一致。这样的一致性要求不能引用不存在的值,如果键值更改了,那么在整个数据库中,对该键值的所有引用要进行一致的更改。

(4) 用户定义完整性使用户得以定义不属于其他任何完整性分类的特定业务规则。所有的完整性类型都支持用户定义完整性 (CREATE TABLE 中的所有列级和表级约束、存储过程和触发器)。

1.2 主流的数据库管理系统

主流的数据库软件产品是根据权威市场调查中的产品市场占有率来评价的,目前,商品化的数据库管理系统以关系型数据库为主导产品,在国内的数据库产品市场,甚至是全球的数据库产品市场上,有几大主流的关系型数据库产品,下面将一一简单介绍。

1. Oracle

Oracle 数据库系统是美国 Oracle (甲骨文) 公司提供的以分布式数据库为核心的一组软件产品,是目前最流行的大型通用关系型数据库之一。它的一个主要特点是能在所有主流操作系统平台上运行 (包括 Windows、Linux、UNIX 等),支持所有的工业标准,提供给用户一个高性能、低成本的完整解决方案,并且一直在不断地根据信息技术的需求研究开发与之配套的新功能,新版本之间完全向下兼容,版本移植和平台的移植风险较少,受到各大企业客户的青睐并得到广泛的应用。目前市场使用的 Oracle 数据库版本有 Oracle 8、Oracle 8i、Oracle 9i、Oracle 10g、Oracle 11g 等,最近推出的新产品是 Oracle 12c,从版本后缀特点可以看出该数据库产品的技术发展路线。

2. Sybase

Sybase 公司在 1987 年推出了 Sybase 数据库,它也是一款支持多操作系统平台的通用关系型数据库。早年 Sybase 是最流行的数据库产品之一,于是 Sybase 与微软形成了战略伙伴关系,微软购买了 Sybase 数据库产品技术,并且在此基础上迅猛发展,形成了目前市场最主流的关系型数据库产品之一——SQL Server 数据库。

3. SQL Server

上面提到,微软的 SQL Server 数据库是购买了 Sybase 公司的 Sybase 数据库之后发展起来的大型通用关系型数据库产品,1995 年正式发布的 SQL Server 6.0 是一款小型商业数据库,1998 年发布了基于 Web 的 SQL Server 7.0 数据库,2000 年发布了企业级的 SQL Server 2000 数据库,因为其功能全面、简单易用,此版本是目前市场上应用比较广泛的版本;为了进一步与微软开发套件集成,2005 年发布了引入了 .NET Framework 的 SQL Server 2005,并在此基础上,于 2008 年推出了最新版本 SQL Server 2008。SQL Server 数据库只支持 Windows 操作系统服务器,没有开放性,而且在业界,Windows 操作系统的安全性并不是很好,但是因为提供了简单易用的图形化操作界面,深受一些用户的喜爱,在很多高校或者培训结构,一般作为入门的数据库系统学习。

4. DB2

DB2（注意，其中的 2 并非产品版本号，而是产品名称）是 IBM 公司开发的一个适用于多平台、大型的关系型数据库产品，因为其具有良好的开放性和并行性，DB2 在企业级的应用最为广泛，在全球的 500 家最大的企业中，几乎 50% 以上用 DB2 数据库服务器，目前广泛应用于金融、电信、保险等较为高端的领域，尤其在金融系统备受青睐。1996 年发布了 DB2 v2 并改名为 DB2 UDB（DB2 通用数据库）。2008 年发展到 v9，将数据库领域带入到 XML 时代。

5. MySQL

MySQL 在开发源代码世界和 Web 团体社区中都是非常有名的流行数据库系统，也是免费数据库软件的代表，因为其免费、面向大众、通俗的解决方案，吸引了大量的企业用户和开发用户，多应用在小规模的 Web 网站应用系统上，但是不太适合大规模、要求高可靠性的应用系统。

6. 国产数据库

前面介绍的都是国外的数据库产品，也在应用市场上占据了主要位置，这里介绍几大国产数据库产品。

OpenBASE 是东软集团有限公司软件产品事业部推出的我国第一款自主知识产权的商品化数据库管理系统，该产品由东软集团有限公司软件产品事业部研发并持有版权。十多年来，OpenBASE 已逐渐形成了以大型通用关系型数据库管理系统为基础的产品系列。2007 年东软正式推出了 OpenBASE 的最新版本 OpenBASE v6.0。在新版本中对 OpenBASE 之前各版本的功能进行了扩充和优化，功能、性能、可用性和可靠性方面都得到了较大提高。目前该产品已广泛应用于办公自动化、医院、房地产、多媒体教学、电子商务、信息安全等数十个领域。

DM（达梦）数据库有限公司是从事数据库管理系统研发、销售和服务的专业化公司，其前身华中理工大学（现华中科技大学）达梦数据库与多媒体研究所成立于 1992 年，是国内最早从事数据库研究的科研机构。DM 数据库是大型通用数据库管理系统，从 1988 年起，先后有 6 代数据库产品问世，从第一个自主知识产权的数据库管理系统 DM1 到最新的 DM6，目前广泛应用在消防、物流与财务系统中。

图 1-1 是一个全球权威调查机构于 2008 年所调查的各大数据库产品的市场占有率分布图，仅供读者参考。

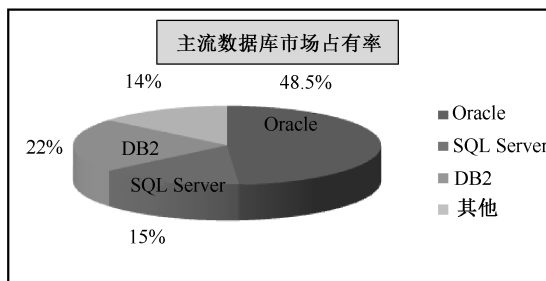


图 1-1 数据库产品市场占有率

总结：客户听了 Smith 的介绍以后，虽然对几种主流的数据库产品有了一个大致印象，

可是他对于如何为他的应用系统选择合适的数据库系统还是没有具体的思路,于是 Smith 为客户分析了选择数据库产品的几个主要指标。

1.3 选择数据库软件产品的要素

我们的目标是选择一个最合适的数据库产品,通常从以下几个方面来考虑。

1. 性能

考虑性能时一般要面向大量操作和用户的应用系统,在大型数据库产品上考虑,但是大型数据库产品 Oracle、DB2、SQL Server 这 3 款数据库中哪一个是性能最好、处理速度最快、可扩展性最好的呢?这个问题在数据库领域也被讨论和争论很多年了,每个公司也会投入大量的资金来证明自己是最优的,在过去的这些年所获得的性能上的改进说明,这 3 个公司的中任何一个都有足够的力量来提供比大多数应用程序所需要的更好的速度、可扩展性以及性能,而且一套数据库产品的最佳性能需要最优的硬件、平台和维护人员辅佐。所以对于大多数的企业来讲这些性能已经足够了。

2. 平台

因为每个数据库产品与操作系统平台结合度情况都不一样,所以企业对于平台的选择也会对数据库产品的选择有所影响。在数据库产品中,IBM 的 DB2 毫无疑问占领了大型机的市场;SQL Server 致力于为 Windows 操作系统提供最优解决方案,不支持其他平台;Oracle 支持多平台操作系统,并且提供较低风险的移植。所以企业用户可以结合目前的现有资源或者企业的发展目标定位数据库的平台,从而选择相应的数据库软件。

3. 价格

价格是目前一些中小企业需要重点考虑的问题,也是一个很复杂的因素。因为这个价格不仅仅是指数据库产品的购买价格,还包括服务器配置、产品系统的维护、个人许可、额外工具、开发成本以及技术支持等费用,特别是人力资源成本是企业需要提前预知的长期成本。

以 SQL Server 数据库系统和 Oracle 数据库系统来说,从数据库服务器的要求方面考虑,Oracle 数据库系统的要求高过 SQL Server,故硬件成本较高;从人力成本的角度考虑,根据应用市场情况分析,一个 Oracle 数据库管理员的成本远远高出 SQL Server 数据库管理员(目前市场上也出现很多维护服务外包公司,专门提供服务人员为企业提供高质量的系统管理和维护服务),故人力成本较高;从软件本身的购买价格而言,对于具有一定规模的企业级应用需求,Oracle 数据库软件产品的价格也是高过 SQL Server 的。

4. 可用资源

部署一套完整的用户解决方案,需要有一系列的配套支持资源,如服务器(应用程序服务器、数据库服务器)和人力资源,人力资源包括系统操作用户和系统维护人员,其中维护人员包括应用程序维护人员以及数据库管理员。故在资源系统中,与数据库相关的是数据库服务器

和数据库管理员,对于一般系统而言,会考虑现有的服务器是否能够满足新系统的需求?现有的维护人员是否能够满足新系统的需求?如果原系统的相关配套在满足系统现有需求的基础上还有一定的可发展空间,则是可以考虑沿用原有的数据库系统的。

5. 建议方案

如果是一般的小型 Web 网站应用,使用目前免费的 MySQL 是一个可以考虑的选择,因为其简单易用,维护成本等综合成本相对较低;如果是数据应用处理稍大,但也不是海量数据,对数据库的可靠性和稳定性要求不是很高,同时希望维护成本不高,可以考虑选择 SQL Server;如果要实现的是较高端的企业应用,并且需要处理的并发数据量较大,同时对数据库的可靠性、安全性和可扩展性有很高的要求,那么 Oracle 应当是一个不错的选择,但应用 Oracle 的维护成本相对较高。

客户综合了各种因素,并且组织网上购物系统甲方项目组相关人员专门召开了一个邀请 Smith 参加的小型讨论会,综合企业实际情况与需求,会议研究决定,选择 Oracle 公司的 Oracle 11g 作为本系统的数据库软件。因此,Smith 决定再次针对 Oracle 数据库产品对客户进行更详细的介绍。

思考与提高:2009年,电子商务创新领跑者阿里巴巴首次提出“去 IOE”战略模式,“去 IOE”指的是摆脱对 IT 部署中原有的 IBM 小型机、Oracle 数据库以及 EMC 存储的过度依赖。“去 IOE”的过程也是技术发展的过程,如以淘宝为基础形成了优秀的 MySQL 数据库团队,也建立了自己开发数据库 OceanBase 的团队。当时考虑“去 IOE”并不仅仅出于对成本的考量,成本只是最容易看得到的东西,最重要的是满足企业未来长期发展的需要。互联网时代,每一家企业都必须拥抱互联网,传统 IT 架构的软硬件已经无法满足企业在这方面的发展。阿里巴巴的成功经验表明原来依赖 IBM、Oracle 和 EMC 的系统是可以构建在 Commodity PC 上的,这为大多数企业在云计算上搭建 IT 系统扫清了障碍,使它们可以彻底拥抱云计算,拥抱互联网。

1.4 Oracle 11g 数据库产品

Oracle 数据库是 Oracle 公司出品的历史比较悠久的十分优秀的关系型数据库管理系统。当前,Oracle DBMS 以及相应的开发工具和其他相关产品几乎在全世界各个工业领域中都会用到。无论是大型企业中的数据仓库应用,还是中小型企业中的联机事务处理业务,都可以找到成功使用 Oracle 数据库的典范。

1.4.1 Oracle 公司介绍

Oracle 公司是世界上最大的数据库厂商,向遍及大约 145 个国家的用户提供数据库、工具和应用软件以及相关的咨询、培训和支持服务。Oracle 公司总部设在美国加利福尼亚州的红木城,全球员工超过 40000 名。Oracle 公司 1989 年正式进入中国市场,成为第一家进入中国的世界软件巨头,标志着刚刚起飞的中国经济信息化建设得到了 Oracle 公司的积极响应,Oracle 首创的关系型数据库技术也从此开始服务于中国用户。

Oracle 公司的产品线包括数据库、应用服务器、开发工具包、电子商务套件以及产品的培

训认证 (OCA、OCP、OCM)。下面主要针对数据库产品进行介绍。

1.4.2 Oracle 数据库产品发展阶段

1979年,Oracle公司首先推出了基于SQL标准的第一代关系数据库产品,并取名为Oracle;之后十来年来经历了多次更新与发展,到Oracle 8版本时,已经成为一套比较稳定的大型关系型数据库系统了,这个版本也为支持Internet、网络计算等奠定了基础;

从Oracle 8之后,为了适应Internet技术的飞速发展,Oracle公司于1999年推出世界上第一个Internet数据库Oracle 8i,这一版本中添加了大量为支持Internet而设计的特性,同时,这一版本为数据库用户提供了全方位的Java支持;

2001年,Oracle又推出了新一代Internet电子商务基础架构Oracle 9i,在Oracle 9i的诸多新特性中,最重要的就是Real Application Clusters (RAC);

2004年,Oracle公司发布了最新数据库产品——Oracle数据库10g,它与同日发布的Oracle应用服务器10g和Oracle企业管理器一起构成了全球首个面世的集成式网格计算架构软件——Oracle网格计算,这一版本的最大的特性就是加入了网格计算的功能;

在此基础上的3年之后,Oracle公司顺势发布了11g,Oracle 11g提供了高性能、伸展性、可用性、安全性,并能更方便地在低成本服务器和存储设备组成的网络上运行,Oracle 11g也是Oracle公司30年来发布的最重要的数据库版本,根据用户的需求实现了信息生命周期管理(Information Lifecycle Management)等多项创新;

值得一提的是,随着近两年云计算技术的白热化发展,Oracle公司的战略是提供广泛的软件和硬件产品以及服务等组合来支持公有云、私有云和混合云,从而让客户能够选择最适合自己的方式。故甲骨文公司于2013年重磅发布了Oracle Database 12c。

1.4.3 Oracle 认证体系

由于Oracle产品的特殊性,作为全球最大的数据库厂商,Oracle在行业中有着不容置疑的地位,而数据库又是整个IT行业中的关键和核心应用,特别是大型企业级数据库,更是高端中的高端。因此,获得Oracle认证成为很多数据库技术人员的职业追求。

Oracle大学(Oracle University)是Oracle公司专门负责培训业务的部门,它为从事Oracle领域的工作人员,包括数据库管理员、开发人员以及管理人员设计了一系列课程。这些专业的课程能使受训者更快地掌握Oracle技术。

Oracle的最新认证体系包括3个层次:主要面向高校学生的Oracle操作专员认证(OCA),Oracle专业认证(OCP)和Oracle专家级认证(OCM),具体如图1-2所示。

Oracle专员认证:这项较初级的认证是Oracle专为那些仅通过OCP中两项考试的人员设计的初级技能水平考试,是使用Oracle产品的基础。要获得OCA证书,需要通过的考试如图1-3所示,可以选择IZO-007、IZO-047、IZO-051任意一门,加上一门IZO-042,考试通过即可以获得OCA的认证证书,此证书适合高校在校学生。

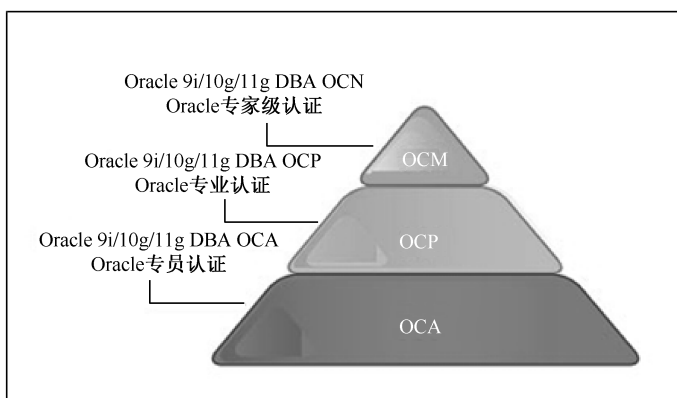


图 1-2 Oracle 数据库认证体系

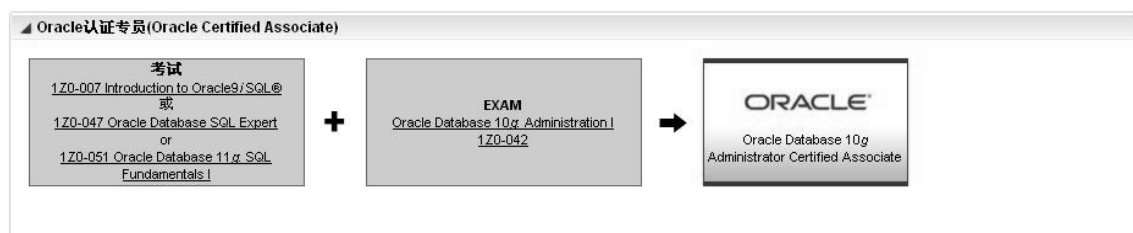


图 1-3 OCA 认证流程

Oracle 专业认证：要取得进阶的 OCP，必须先通过 OCA 的两科考试，以及至少参加过一门由 Oracle 认证讲师教授的课堂训练课程或线上讲师训练课程，并通过 OCP 一科考试 (1Z0-043) 才能拥有 OCP 认证，如图 1-4 所示，以证实在 Oracle 数据库管理领域内的熟练程度。通过这种考试之后，说明此人可以管理大型数据库，或者能够开发可以部署到整个企业的强大应用系统。获得 OCP 证书后，认证者将有机会申请更高的职位，并增强老板对自己的信任和支持。

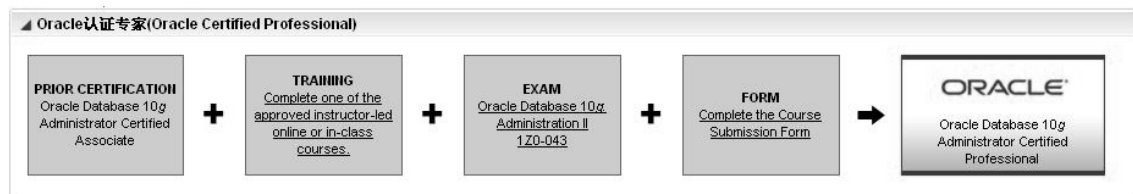


图 1-4 OCP 认证流程

Oracle 专家级认证：这项新的 Oracle 认证要求参试人员必须参加 Oracle 大学的 Oracle 培训，它是 20 世纪 90 年代前 Oracle 专家认证的一个分支，OCM 要求参试人员必须获得 OCP 认证，参加 Oracle 大学的两门高级课程，通过预先测试，并通过 Oracle 试验室的实践测试，具体如图 1-5 所示。

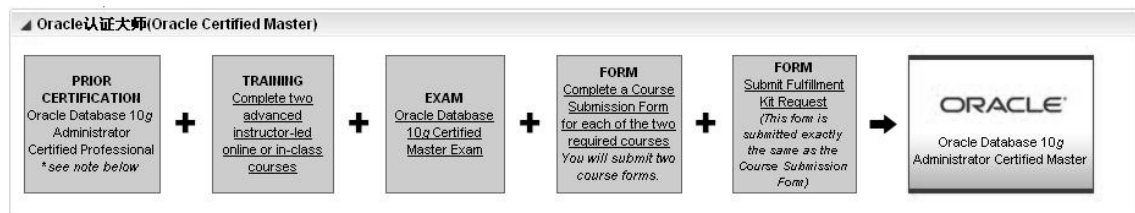


图 1-5 OCM 认证流程

以上信息根据技术的发展和市场需求随时发生更新，想要获取更多详情和最新情况，请参考 Oracle 大学的官方网站 (<http://education.oracle.com/>) 上的认证栏目，其中有最新的考证通知、要求及学习资料。