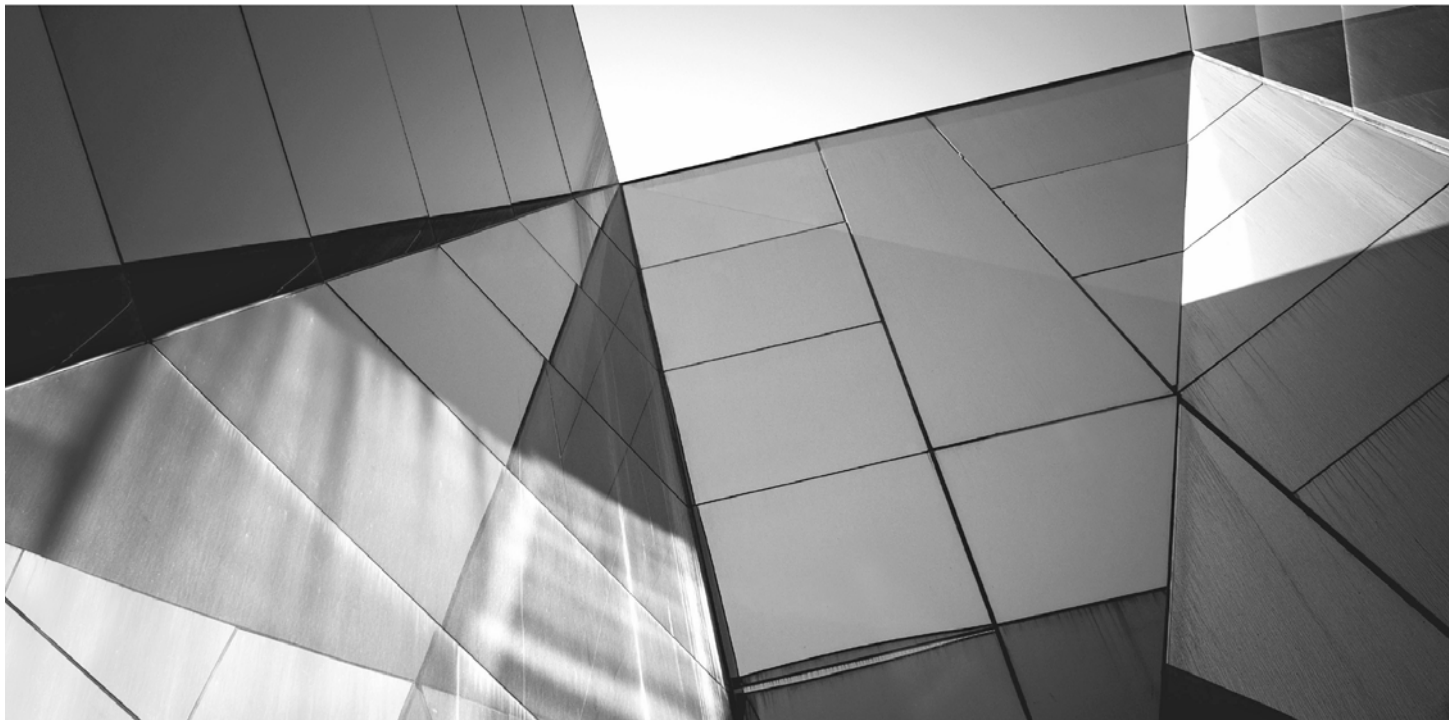




第 1 章

概述与环境准备

20 世纪 70 年代关系模型理论的提出，距今已经 40 多年了。在此期间，关系型数据库从无到有、从小到大、从弱到强地发展起来。时至今日，市面上已经有 Oracle 公司的 Oracle 产品、微软的 SQL Server、IBM 的 DB2 等多种关系型数据库产品，此外还有开源的 MySQL、与大数据相关的 HBase、常用于缓存数据的 Redis、NoSQL 数据库 MongoDB 等。目前数据库领域已经是百家争鸣的格局了。尤其是近十年随着大数据的兴起和非规范化数据(如视频、图片等)的急剧增加，新兴数据库不断出现。这其中，国产数据库(例如武汉达梦、人大金仓、湖南上容

以及山东浪潮的 K-DB 等)也开始逐步进入相对成熟阶段,整个数据库市场更是新产品不断涌现,各种新特性争相亮相。然而拨云见日之后,Oracle 数据库依然是关系型数据库产品中最强大与突出的。

关于 Oracle 数据库的悠久历史,笔者在这里不再赘述。只想讨论一下与另外众多数据库相比,它的优势和特色:

- 功能和性能足够强大
- 开放性与可研究性

以上两个特点,应该是 Oracle 数据库最与众不同的地方了。首先,功能强大,性能优异,无论与其他任何关系型数据库产品相比,Oracle 数据库都在功能与性能上高出一筹。其次,Oracle 数据库足够开放,当然不是指开源数据库可以让用户自己修改源代码的这种开放性,而指在网络上,已有足够多的技术社区和相关论坛,以及众多的博客来提供相关的各种资料,从而使得任何人都可以极便捷地得到 Oracle 的相关信息与技术细节,并进行研究。再者,国内不少城市,例如北京、杭州、上海等地,已经有相当多的线下技术分享与沙龙活动。技术爱好者们完全可以就近参加这些活动,互相学习共同提高。

Oracle 数据库发展到现在,已然经历了多个版本变化。目前市面上,在企事业单位的实际应用中,以 Oracle 11g 为主流版本,尤以 11.2.0.4 为甚。但在笔者最近接触到的项目中,采用 12.1 版本的应用也日渐增多。但也有不少系统仍旧采用 10g 版本。因为 2013 年底,Oracle 公司就已经宣布不再对 10g 版本的数据库提供官方技术支持。因此对于仍旧采用老版本数据库的企业来说,还是有一定风险的。本书以 11.2.0.4 为主要环境,来演示众多实际生产系统中的经典实验案例,并在最后使用 12.1.0.2 版本来研究 12c 中的一些重要新特性。

1.1 数据库、数据仓库与大数据

众所周知,现在是大数据时代。这个自诞生到现在刚好 10 年的概念,至今已然席卷全球,成为各行各业讨论的热门话题,对当今人类社会的诸多方面都产生了重要影响。就在几年前,笔者的一个师姐(一位哲学博士后),有段时间还曾跟笔者聊到大数据等话题。大数据的影响之深,波及行业之广,由此可知。

那么,可曾有人想过,大数据究竟是怎样流行起来的呢?它的前世今生又是如何?笔者并非大数据方面的专家,但恰好有幸亲历了从数据库到数据仓库,再到大数据的行业技术发展过程,故在此略加阐述。

我们知道,数据库用来存储大量数据,并处理高并发的用户访问,同时保证数据的完整性和可恢复性。数据库是企业的数据量达到一定程度后,所需的一种管理数据的技术。但再进一步呢?

一家公司可能有多个业务系统,有用于管理人力资源的,有用于支持销售的,有用于控制工厂生产的,等等。一般情况下,这些系统都有自己私有的数据库。这些数据库之间,虽然会

有数据交互和访问，但往往是零星的。那么，请设想以下需求：当公司发展 to 一定程度，公司总经理忽然想到，能否将这些类似于信息孤岛的诸多业务系统打通，使得可以在整个公司的层面上来查看数据？例如，要是想知道过去数年间公司销售业绩的增长曲线图，并分析它与公司产品策略、生产排产计划等之间有何种关系，那该如何处理？

这便是产生数据仓库的社会需求。数据仓库的概念在1990年由 Inmon W.H 在他的著作 *Building the Data Warehouse* 中提出。其完整概念如下：

数据仓库(Data Warehouse)是一个面向主题的(Subject Oriented)、集成的(Integrated)、相对稳定的(Non-Volatile)、反映历史变化(Time Variant)的数据集合，用于支持管理决策(Decision Making Support)。

数据仓库的出现，使得公司高层有了能够从全局把握公司全部数据的工具，从而可以在更高层次上洞察影响公司业务发展的诸多因素之间的关系，并因此做出能够影响公司战略发展的重要决策。于是自 20 世纪末至 21 世纪的前 10 年，算得上是数据仓库发展的黄金时期了。笔者也曾经在刚刚毕业的时候，参与了天津电力数据中心的建设项目。

当时的数据仓库技术，依然是基于传统的关系型数据库构建而成。底层的数据存储仍是诸如 Oracle 的数据库。但是为了处理比原有数据库中更大量的数据(数据库中的数据量，我们称为大量；数据仓库中的数据量，我们称为海量)，分区技术、并行技术等逐渐涌现，大型甚至超大型数据库以及数据中心也应运而生。那时，不少公司和政府机构都在忙于建立大型数据中心，并对其管理的海量数据进行分析和挖掘，期望能从中获得有价值的信息。

但由于当时条件所限，很多企业的历史数据积累不够，数据质量也堪忧，数据挖掘与分析的相关算法尚不完善，种种不利因素，使得当时虽然有不少企业投身于数据仓库研究与搭建，但最终能得到充分利用的，其实不多。

再往后，大数据出现了。

从数据中挖掘有价值的信息，这一由数据仓库提出的理念，被大数据完整继承下来，并进一步发扬光大。大数据采用了全新的架构和搭建技术，一开始就是为处理海量数据而生。它不但能处理传统的规范化数据，就连非规范化数据也照单全收。这顺应了近年来电商和社交类网站的兴起这一趋势，因此大数据就这样在极短时间内流行起来。而作为大数据前身的数据仓库，便就这样被大数据无情取代。

虽然现在还有一些公司在采用数据仓库技术，但是现在，确实已经是大数据的天下了。

1.2 数据库技术在大数据中的地位与价值

那么问题来了，大数据如此流行，那数据库在这样的时代，其地位又将如何？

其实很简单，大数据时代，数据库技术依然是核心技术之一。无论多大的数据量，只要涉及存储及访问处理，其原理和实现方式都是极接近的。并且 Oracle 公司本身，也在拥抱大数据

技术,他们已经提供了 Oracle 数据库和大数据/Hadoop 的应用接口。可以在 Oracle 中存储数据,然后在 Hadoop 中进行分析处理。

因此,即使是在大数据称雄天下的时代,数据库技术依然是不可或缺的。

1.3 相关技术

数据库是基于操作系统和存储之上的应用软件。因此,在实际使用和管理 Oracle 数据库系统时,也往往会涉及部分操作系统和存储的相关技术。当然,若是从整个应用系统的角度看,与数据库相关的技术就更多,如网络、中间件等。

本书旨在为数据库技术初学者或零基础人员提供入门级别的技术指导与动手实验,因此不涉及网络和中间件这样的技术。但是数据库基本的安装配置与管理,会涉及部分操作系统的相关知识,因此如果读者有一定的 Linux 基础,则学习效果更好。毕竟,本书的实验都是基于 Red Hat 6.4 搭建的环境来完成的。

当然,在本书后续内容中涉及 Linux 的地方,笔者也会详细介绍相关的知识。

1.4 本书内容与架构说明

本书的内容以 Oracle 数据库中典型的实验为主,涵盖 SQL、数据库配置管理、备份恢复、性能优化等数据库知识领域。每部分都有具体的实验内容及相关知识说明。读者可以根据这些实验一步步进入 Oracle 数据库的知识殿堂。

因此建议读者能够按照本书的章节内容,认真完成书中的每个实验。这些实验是笔者和诸多同行在多年工作经验中的实战总结,具有较高的实际应用价值。另外,在阐述如何去完成这些实验的同时,也会解释它们能够解决实际中的何种问题,适用于什么样的业务场景,以及相关的知识点等等。

1.5 实验环境准备

至此,Oracle 的相关知识已经介绍完毕,可以开始准备后续实验的环境了。可扫描本书封底的二维码,访问本书的支持网站,通过相关内容或链接,获得学习本书所需的安装介质说明和相关技术资料:

- Oracle 11gR2 安装介质说明
- Oracle 11gR2 官方文档
- Oracle 11gR2 安装手册
- Red Hat 6.4 64 位安装介质说明

请下载或按照说明获取上述所有资源，这些内容在本书的后续实验中都将用到。在下章的实验之前，我们首先需要按照文档“Oracle 11gR2 安装手册”依次完成虚拟机创建、Red Hat 6.4 操作系统安装与配置，以及数据库软件安装和监听创建工作。该安装文档的内容足够详细，请按步骤认真操作即可。