

来自工作现场的 DBA 手记

SQL 案例解析

以 SQL Server 与 Oracle 为例

适用版本 SQL Server 2008/2005/2000
Oracle 11g/10g/9i

 悦知文化
Delight Press

 ORACLE
Professional

徐桂祥 著

 清华大学出版社

SQL 案例解析

以 SQL Server 与 Oracle 为例

徐桂祥 著

清华大学出版社

北 京

SQL 案例解析：以 SQL Server 與 Oracle 為例

徐桂祥

精誠資訊股份有限公司-悅知文化, 2010.06

ISBN: 978-986-6348-42-6

本書為精誠資訊股份有限公司-悅知文化授權清華大學出版社於中國大陸(台港澳除外)地區之中文簡體版本。本著作物之專有出版權為精誠資訊股份有限公司-悅知文化所有。該專有出版權受法律保護, 任何人不得侵害之。

本書封面貼有清華大學出版社防偽標籤, 無標籤者不得銷售。

版權所有, 侵權必究。侵權舉報電話: 010-62782989 13701121933

圖書在版編目(CIP)數據

SQL 案例解析: 以 SQL Server 与 Oracle 为例/徐桂祥 著. —北京: 清华大学出版社, 2011.8

ISBN 978-7-302-25949-7

I. S… II. 徐… III. 关系数据库—数据库管理系统, SQL Server、Oracle IV. TP311.138

中国版本图书馆 CIP 数据核字(2011)第 113094 号

责任编辑: 王 军 李维杰

装帧设计: 康 博

责任校对: 蔡 娟

责任印制: 王秀菊

出版发行: 清华大学出版社

<http://www.tup.com.cn>

社 总 机: 010-62770175

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

地 址: 北京清华大学学研大厦 A 座

邮 编: 100084

邮 购: 010-62786544

印 装 者: 清华大学印刷厂

经 销: 全国新华书店

开 本: 185×230 印 张: 32 字 数: 656 千字

版 次: 2011 年 8 月第 1 版 印 次: 2011 年 8 月第 1 次印刷

印 数: 1~4000

定 价: 59.80 元



产品编号: 039298-01

推荐序

学习 SQL 语言时，获得 SQL 的正确观念是最基础、最简单的一步，而将观念转换成实际应用才是最困难的。因为必须要有实际的需求与足够的经验，才能顺利地将 SQL 知识由理论转换为真正可行的 SQL 命令，而这段转换期往往是最难熬的一段时间。如果能够成功通过这道关卡，便能海阔天空任你遨游，也表示你的 SQL 能力已经可以真正地解决工作上遇到的问题。

然而，并不是每个人都有能力或有足够的摸索时间可以自行突破此难关，通过本书可以协助读者缩短这段痛苦期，并快速上手。因为本书作者徐桂祥先生不但拥有丰富的实践经验与扎实的 SQL 理论基础，而且也十分愿意分享自己多年的工作经验。

本书通过大量经过验证且可行的范例，让 AP/DBA 以最短的时间熟悉 SQL 语言的各种运用与注意事项，以弥补理论与实际应用的差距。所以，每个 AP/DBA 都应该有一本作为案头参考书随时翻阅，通过借鉴作者多年的经验总结，帮助自己顺利地解决目前所遇到的 SQL 难题。

申建忠

精诚资讯知识产品事业部产品经理
恒逸教育培训中心资深讲师

作者序

本书从开始编写至完稿大约历经了一年半时间，主要设定读者群为工作一到两年的软件开发人员。本书同时探讨 Oracle 及 SQL Server 两种数据库，可作为已具有 Oracle 或 SQL Server 经验的读者学习另一种数据库的参考，书中范例由笔者多年从事半导体合格率分析软件的开发经验、银行客服系统开发时所碰到的一些问题，以及对朋友、同学、同事的咨询汇集而成，以逐步分解方式编写，对于已具有相当经验的读者，则建议直接参考所附的 SQL 命令。

本书受限于截稿因素，未能将 SQL 性能调校列入，而在书中所探讨的高级 SQL 应用，主要是用以提升整体运行效率。以综合日期报表为例，一般常见的开发方式是分别以周报、日报及月报等多组 SQL 来完成所有数据存取，常以应用程序端语言组合双 FOR 循环将数据进行转置，需耗费数秒甚至更长时间才能完成，本书希望向读者提供一个简单且性能好的方式来完成各种常见需求。

感谢悦知文化出版此书，以及编辑 Vicky 的催促，使得本书得以顺利完稿，也感谢家人的支持。期望本书对读者能有所帮助，本书范例虽是以 SQL Server 及 Oracle 数据库为主，但对于 MySQL、DB2、Teradata 等其他数据库，稍作修改同样也可以顺利执行。

徐桂祥

第 1 章：数据库的使用

本章将简要说明关系型数据库的基本来由及其使用的查询语言 SQL(Structured Query Language)，对于系统设计人员经常面临的，应该选择用程序还是用数据库的问题，在此提出一些想法及概念。接着，提供 SQL Server 2000/2005/2008 数据库系统重要 SQL 命令的说明与介绍，并在文档中分别标示适用的版本，以供数据库管理人员、程序开发人员等参考。此外，本章还介绍 Oracle 数据库的一些概念及说明。

第 2 章：SELECT 查询语句

首先介绍如何使用 SELECT 查询命令进行相关的查询，接着按照 SELECT 语法的基本结构，分别介绍 SELECT、FROM、WHERE、GROUP BY 及 ORDER BY 的概念及用法。本章所探讨的内容较为基础，有经验的读者可直接略过。

第 3 章：基本 SQL 用法

延续“第 2 章：SELECT 查询语句”，将探讨多数据源的方法、特定条件(或组合)进行分支处理，以及数据汇总或分组后的数据筛选方法，最后介绍以集合运算(SET)实现 FULL OUTER JOIN 的方法，以及应用 NULL 特性处理数学运算的除零错误问题。为顾及 SQL 语法的可读性及后续可维护性，将说明如何在 SQL 命令中以断行、缩排、空格、空白行、括号及注释等编排风格，从而使 SQL 查询结构及隐含运算逻辑等能更清楚地展现出来。

第 4 章：数据操作处理

介绍如何通过 INSERT、UPDATE 及 DELETE 等数据操作命令(Data Manipulation Language, DML)变更数据表中的数据。此外，将说明如何使用数据库中提供的序号(Sequential Number)功能，在无明确主键(Primary Key)、主键太长或包含太多字段时使用。当建立主从(Master-Detail)关系数据时，需先取得主要数据的流水序号。在此将以 C#语言为例，说明如何在应用程序中取得存入数据的自动编号(识别值)，以供后续使用。

第 5 章：数据类型

如何在众多数据类型中选择适当的类型，需考虑后续数据使用上的问题，合理的扩展空间以及避免不必要的空间浪费。设计时，请审慎选用数据类型，因为类型一旦确定以后往往难以再变更。本章将介绍常用的三大数据类型，字符、数值及日期的特性、概念、适用范围、所需的存储空间等，并探讨类型选择应注意的事项。数据库对于未知(Unknown)、不明、未定义或类似意义的数据以 NULL 表示，本章将详加说明此特殊数据值以避免误用，以及如何利用其特性在特定运算中使用。

第 6 章：字符操作函数的使用

首先，将对比介绍常用的字符串函数，整理出两种数据库在函数上的差异及对应关系。在 LIKE 子句中使用百分比符号(%)、下划线(_)等通用字符(Wildcard)，用于局部字符串的模糊比对。此外，以字符串特征(Pattern)利用正则表达式(Regular Express, 简称 regexp 或 RE)实现弹性好且功能强大的字符串模糊对比。对于如何判别是否为数字及中文字符串，SQL Server 提供了 ISNUMERIC 函数，可用来判断是否符合数值定义；Oracle 并未提供类似函数，它运用正则表达式及 TRANSLATE 函数进行判断。因“中文”涉及的范围实在太过于广泛，这里只提供一种简易且基本的判断方法。

笔者整理了过去所处理的问题，编写成“6.5：筛选特定字符串前缀”，探讨如何剔除半导体生产中的实验生产批号(LOT)，若批号的前 2 位或前 3 位符合实验批号定义，则将其予以删除。在“6.6：固定长度字符串”介绍如何产生银行、证券等相关行业与大型主机进行沟通的固定长度、格式的字符串数据。最后介绍在 TFT-LCD 生产过程中，如何对玻璃(Glass)基板中各面板(Panel)进行检测，测试结果以字符串(如 OOXOXX)存储于数据表中，据此计算并筛选出合格率较低的基板。

第 7 章：日期函数的使用

日期这一数据类型对使用频率极高的时间区间报表(日报、周报及月报)有相当重要的影响。本章将首先介绍类型转换及取得日期部分信息等重要功能，然后探讨日期运算，可以“日期±天(浮点数)”方式实现日期时间的延后或提前，使用日期减日期的方式得到两个日期的时间差，但 Oracle 与 SQL Server 回传结果的数据类型却不相同，在“7.3：日期运算概念”、“7.4：日期加减运算”这两节将分别介绍两种数据库在日期延后或提前、时间差函数方面的差异。本章最后提供各种常用特定日期(如月初、月底、年初、年底、季初等)的 SQL 命令。因许多公司的生产、运营绩效周报表中，周的定义并非与数据库默认值相同，特别是提供的如何取得下一个周几的日期的方法，所以可以此为分组条件进行汇总，产生数周(如 4 周或 8 周)整体的运营状况报表，具体内容请参阅“7.7：找出下个周五”、

“10.3: 周报”这两节。

第 8 章: 数值函数的使用

本章介绍 SQL Server 与 Oracle 两种数据库中常用数值函数间的差异。SQL Server 在数学运算上的特殊限制, 可能造成运算时发生溢位或舍位, 在此一一整理数值运算的问题及解法。将应用余数用以判断数值是否为整数, 利用 SIGN 正负值判断函数将数据区分成三个区域, 用以判断两者前后关系, 以定存单是否提前解约、当月解约或延后解约来说明函数的其他用途。

第 9 章: 高级 SQL 用法

本章的重点在于提供一些原本考虑以应用程序或存储过程开发的需求, 尝试直接以一组 SQL 命令来完成。以非直观(觉)的思考模式, 将商业逻辑进行拆解与合并, 归纳整理为创造(数据列)、转换(字段)、连续分组三大类型, 并在本章一一呈现。详述常见的各种形式的案例, 如数据转置、累计加总、名次、所得税计算、商品组合、求职网(薪资谋合)、职务代理及组织图等常见需求, 将逐步分析设计概念及方法, 导引读者另一种思考及解决问题的模式。本章值得细细品读。

第 10 章: 日期区间报表

探讨几种常见的日期区间报表, 分别以日报、月报、周报、小时区间报表、科学园区的四班二轮特殊工时报表, 以及完整日期报表等分别说明。周报部分将提供如何通过函数及特殊运算模式, 以一组 SQL 方式产生数周(如 4 周或 8 周)的运营报表, 减少数据库存取资源耗损, 有效提升效率。小时区间报表则提供每三小时、五小时的状况监控报表, 科学园区为应对 24 小时生产上的需要, 采用四班制一天二轮(四班二轮)的工时制度, 以 07:30 交班时间为一天的开始, 在此提供如何通过时间标准化方法来简化报表产生, 并以日、月、周报等为例。另外, 对于一些异常状况造成停产, 或其他原因造成某时段数据空缺的现象, 管理者希望以整月完整各天的运营详细状况, 在完整日期报表中以三种样式分别说明。

第 11 章: 存储过程

存储过程(Stored Procedure)是将一个或多个 SQL 语句组装成一个逻辑单元, 存储在数据库端可以供数据库端或应用程序端调用。本章介绍程序(Procedure)及函数(Function)的建立, 并说明如何将程序加密以保护商业及运算逻辑。本章的重点在于如何建立用户自定义聚合函数(User-Defined Aggregate Functions), 以分组字符串连接为例, 说明数据库聚合函数的概念与机制。SQL Server 需以 CLR 开发相关组件, Oracle 则采用 PL/SQL 中面向对象

程序设计的开发方法，实际开发过程中的概念与采用一般程序开发类似，以此范例建议开发人员需反复思考如何直接应用数据库中的内建命令及功能，相信能实现不少过去认为无法应用 SQL 命令完成的工作。

第 12 章：异质数据源

企业组织内系统因功能、负责单位不同，而将各种数据分散在不同数据库、异质数据库或文件之中，开发时往往需要将各种来源的数据进行整合连接。

本章将探讨如何连接异质数据源，SQL Server 中建立链接服务器(Linked Server)以连接另外的 SQL Server 数据库或 Oracle 数据库，另外以 Oracle 为主建立数据库链接(Database Link)，连接另外的 Oracle 数据库，以 Gateway 方式连接 SQL Server。SQL Server 可采用 OPENROWSET 命令，通过 OLE DB Provider 连接至各种数据库及文件，提供异质偶发数据查询操作，还可供各种异质数据库(含文件)进行连接(Join)。本章最后介绍 Oracle 的 SQL*Loader 工具，此工具提供了可变长度文件(如 csv)或固定长度文件的导入功能。

第 13 章：特殊应用

本章主要探讨一些特殊应用，SQL 命令的局限在于难以进行循序处理，这里介绍三种方法，使 SQL 命令实现近似程序语言中的循环语句功能。另外还介绍数据库的随机数功能，除了取得随机数数值外，也可用于随机抽选及各群体随机抽选等应用。另一项影响效率的重大问题是：一般常使用 IN 命令进行多数据值筛选，当 IN-LIST 中指定的数据值过多时，除了可能超过数据库所允许的数量上限外，还会造成执行效率低下。本章将介绍如何通过将 IN-LIST 转换成暂存数据表或其他方法，克服此问题。

范例的下载与使用

为方便读者配合各章节范例进行练习，本书范例将以 SQL Server 及 Oracle 分成两个文件：SQLTips-MSSQL.zip 和 SQLTips-ORA.zip。请到悦知网站下载：<http://www.delightpress.com.tw/bookothers/sknd00002.html>。

Oracle 范例的使用方式

Oracle 提供的数据库备份文件为 sqltips.dmp。为避免不必要的困扰，建议建立新的 Schema 后再 Import，请至 Command Mode，执行 Imp 指令即可将数据还原。操作指令如下：

```
C:\Users\Harry>IMP SQLTIPS2/admin
Import: Release 11.1.0.6.0 - Production on 星期三 6月 2 21:31:20 2010
Copyright (c) 1982, 2007, Oracle. All rights reserved.

联机至: Oracle Database 11g Enterprise Edition Release 11.1.0.6.0 - Production
With the Partitioning, OLAP, Data Mining and Real Application Testing options

汇入档: EXPDAT.DMP > C:\sqltips.dmp
请输入插入缓冲区的大小(下限为 8192) 30720>
经由传统路径, 由 EXPORT:V11.01.00 建立的汇出档
已在 ZHT16MSWIN950 字符集与 AL16UTF16 NCHAR 字符集中完成了汇入
仅列出汇入档的内容 (yes/no): no >
对象已存在, 建立对象的错误被略过不予处理 (yes/no): no >
汇入授权 (yes/no): yes >
汇入表格数据 (yes/no): yes >
汇入整个汇出档 (yes/no): no >
使用者名称: SQLTIPS

输入表格 (T) 或分割表 (T:P) 名称. 空的清单表示使用者所有的表格
输入表格 (T) 或分割表 (T:P) 名称或 . 若完成的话:

. 正在将 SQLTIPS 的对象汇入 SQLTIPS 中

(以下省略)
```

SQL Server 范例的使用方式

SQL Server 提供了 SQLTips.bak 文件，请先建立新的数据库后再 Restore 开启 Microsoft

SQL Server Management Studio 即可将数据还原。操作方法如下：

Step01: 右击“Database”，选择“Restore Database”，如图 0-1 所示。

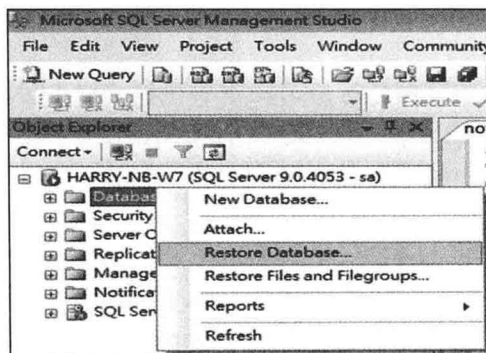


图 0-1

Step02: 右击“Database”，选择“Restore Database”，打开“Restored Database”窗口，如图 0-2 所示。

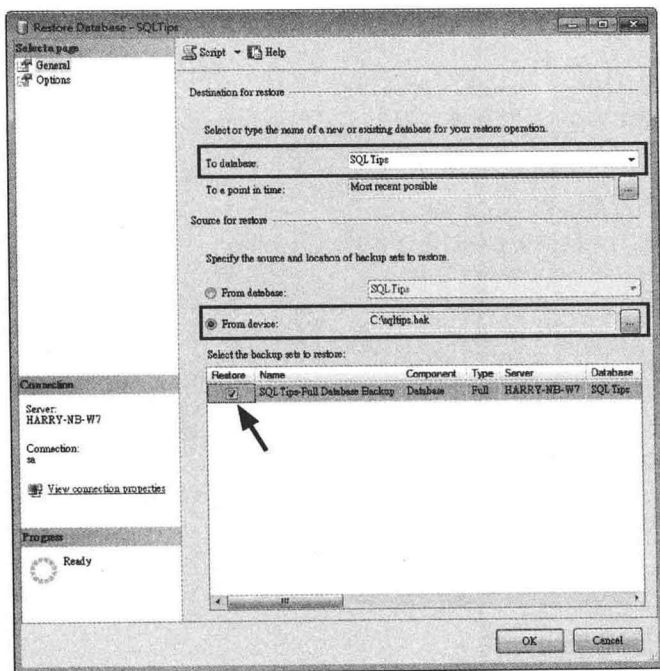



图 0-2

在“To database”下拉选择文本框中输入或选择还原的数据名称，即备份的文件名称(如SQLTips)，同时指定要还原的备份数据的来源和位置，选中“From device”并单击“”按钮，依指示选择所存放的备份文档，并于下方的“Select the backup sets to restore”区域选择要还原的备份，即可还原数据库。

目 录

第 1 章 数据库的使用	1
1.1 什么是 SQL	2
1.2 程序语言或数据库	3
1.3 SQL Server 概述	5
1.4 Oracle 概述	7
第 2 章 SELECT 查询语句	9
2.1 基本 SELECT 子句	10
2.2 字段选择(SELECT)	12
2.3 指定数据源(FROM)	14
2.4 指定查询限定条件(WHERE)	15
2.4.1 比较运算符	15
2.4.2 以逻辑运算符组合多个条件	17
2.4.3 列举项目(IN / NOT IN)	19
2.4.4 NULL 选取	20
2.4.5 范围限制(BETWEEN...AND)	21
2.4.6 字符模糊匹配	22
2.5 查询结果的排序(ORDER BY)	23
2.6 数据分组(GROUP BY)	26
第 3 章 基本 SQL 用法	29
3.1 SQL 编写风格	30
3.2 数据表连接的使用	33
3.2.1 内连接	34
3.2.2 外连接	37
3.2.3 交叉连接	39
3.3 子查询	41
3.3.1 嵌套子查询	42
3.3.2 相关子查询	43

3.4	集合的使用	45
3.4.1	并集(UNION)	47
3.4.2	交集(INTERSECT)	49
3.4.3	差集(MINUS)	49
3.5	CASE 表达式	51
3.5.1	简单型 CASE 子句	51
3.5.2	搜索式 CASE 子句	52
3.6	HAVING 子句的使用	55
3.7	DISTINCT 语法的使用	57
3.8	特殊 SQL 排序法	59
3.8.1	CASE 表达式	60
3.8.2	字符串位置函数(INSTR、CHARINDEX)	67
3.8.3	COUNT 函数(聚合函数)	68
3.9	如何取得两数据集中非重复的数据	69
3.9.1	FULL OUTER JOIN	71
3.9.2	区域去除	72
3.10	如何避免除零错误(Avoid Diving by Zero)	76
第 4 章	数据变更处理	79
4.1	新增—INSERT	80
4.1.1	数据输入(INSERT...VALUES)	80
4.1.2	数据来自其他表(INSERT...SELECT)	82
4.2	序列值的使用	83
4.3	取得当前自动编号值	89
4.3.1	SCOPE_IDENTITY()	89
4.3.2	OUTPUT	91
4.3.3	RETURNING	92
4.4	更新数据—UPDATE	93
4.5	删除数据—DELETE	96
4.5.1	删除特定数据	96
4.5.2	使用连接删除数据(DELETE With JOIN)	97
4.5.3	删除重复的数据	98
4.5.4	删除大量数据	100
4.5.5	清空数据表(TRUNCATE TABLE)	101

第5章 数据类型	103
5.1 基本数据类型	104
5.1.1 字符类型	104
5.1.2 数值类型	106
5.1.3 日期类型	111
5.2 数据类型的选择	113
5.3 类型转换	115
5.4 NULL 概述	116
第6章 字符串函数的使用	121
6.1 字符串函数介绍	123
6.1.1 字符串连接	123
6.1.2 截取子串	124
6.1.3 字符位置	127
6.1.4 获取字符串长度(字符个数)	128
6.1.5 删除空格(特定字符)	129
6.1.6 字符串替换	130
6.1.7 字符串填补	134
6.1.8 变换字符/字符 ASCII 码	135
6.1.9 变换大小写	136
6.2 字符串模糊比较	136
6.3 是否为数值字符串	141
6.4 中文字判断	143
6.5 筛选特定字符串前缀—以半导体实验生产批号为例	146
6.6 固定长度字符串—以特定字符补足长度	148
6.7 特定字符出现次数—以面板检测结果为例	149
第7章 日期函数的使用	155
7.1 格式转换	157
7.2 取得部分日期信息	161
7.3 日期运算的概念	164
7.4 日期加减运算	168
7.5 取得两个日期/时间的差	174
7.6 取得特定日期	176

7.7	找出下个周五	181
第 8 章	数值函数的使用	185
8.1	数值函数	186
8.1.1	余数(MOD)和商数	187
8.1.2	常用的数值运算函数	187
8.1.3	次方、平方根、自然对数、对数、EXP(e)	191
8.1.4	圆周率(PI)、三角函数(Sin、Cos、Tan)	192
8.2	转换函数	193
8.3	SQL Server 数学运算时的注意事项	198
8.4	判断是否为整数—余数的应用	203
8.5	判断两个日期前后顺序—SIGN 函数的应用	204
第 9 章	SQL 用法进阶	207
9.1	数据行转置—Row To Column	210
9.1.1	通过 SQL 命令实现转置	210
9.1.2	使用 PIVOT 命令实现转置	214
9.2	数据列转置—Column To Row	220
9.3	员工专长明细	227
9.4	累计汇总	230
9.5	汇总数据	238
9.6	数据记录的排序——传回各种书籍订单订购数量排名	249
9.7	每个部门中薪水最高的员工	253
9.8	机群分组合并状态	257
9.9	个税计算	266
9.10	商品组合	270
9.11	求职网	279
9.12	Defect Overlap—缺陷重叠分析	284
9.13	职务代理关系表/系统异常处理人员顺序表	289
9.14	组织图—树型结构展开	294
第 10 章	日期区间报表	307
10.1	日报	309
10.1.1	日期类型	309

10.1.2	字符类型	312
10.2	月报	314
10.2.1	日期类型	314
10.2.2	字符类型	317
10.3	周报	318
10.3.1	周的特殊定义	319
10.3.2	运算差异分析	321
10.4	小时区间报表	323
10.4.1	3小时区间报表	324
10.4.2	5小时区间报表	328
10.4.3	利用序列产生报表	332
10.5	科学园区四二轮报表	334
10.5.1	四二轮报表—日报	337
10.5.2	四二轮报表—月报	340
10.5.3	四二轮报表—周报	345
10.5.4	综合日期报表	349
10.6	完整日期区间报表	361
10.6.1	完整日期报表—未转置	362
10.6.2	完整日期报表—以产品转置	365
10.6.3	完整日期报表—以日期转置	368
第 11 章	存储过程	371
11.1	存储过程	373
11.1.1	程序	373
11.1.2	函数	377
11.1.3	错误来源	380
11.1.4	加密	381
11.2	触发器	383
11.3	游标的使用	394
11.4	程序包	403
11.5	用户自定义聚合函数	407
第 12 章	异质数据源	421
12.1	分布式查询—Database Link	422