

凝聚名家技术典范·分享成功IT之路



循序渐进DB2 DBA系统管理、运维 与应用案例(第3版)



牛新庄 著

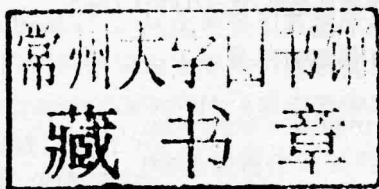
清华大学出版社

循序渐进 DB2

DBA 系统管理、运维与应用案例

(第 3 版)

牛新庄 著



清华大学出版社

北 京

内 容 简 介

DB2 数据库是 IBM 公司关系型数据库核心产品,在国内以及全球有着广泛的应用。针对 DB2 初学者,本书循序渐进地把 DB2 涉及的众多概念和知识介绍给大家。实例、数据库、表空间和缓冲池、数据迁移、备份恢复、DB2 基本监控方法、数据库参数、DBA 日常维护以及数据库常用工具都是本书关注的重点。在介绍这些数据库对象概念和用法的同时,作者尽可能从 DBA 日常工作的角度探究 DB2 数据库管理和运维工作,并提供了丰富的应用案例,对读者更有实践指导意义。在新的一版中,所有的内容、示例都基于 DB2 V10.5 版本进行了修订,并介绍了 DB2 V11.1 中的新功能、新特性。

本书适合 DB2 初学者、DB2 开发人员、准备参加 DB2 认证考试的读者以及 DB2 数据库管理人员学习和阅读。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

循序渐进 DB2 DBA 系统管理、运维与应用案例 / 牛新庄 著. —3 版. —北京:清华大学出版社, 2017
ISBN 978-7-302-48119-5

I. ①循… II. ①牛… III. ①关系数据库系统 IV. ①TP311.132.3

中国版本图书馆 CIP 数据核字(2017)第 208132 号

责任编辑:王 军 李维杰

装帧设计:牛静敏

责任校对:成凤进

责任印制:宋 林

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

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

印 装 者:清华大学印刷厂

经 销:全国新华书店

开 本:185mm×230mm 印 张:32.75 字 数:674 千字

版 次:2009 年 5 月第 1 版 2017 年 9 月第 3 版 印 次:2017 年 9 月第 1 次印刷

印 数:1~4000

定 价:98.00 元

序

关系型数据库已经走过整整半个世纪曲折而辉煌的历程，回顾 50 年的关系数据库发展史，我们心潮涌动，激情难抑。关系数据库始于 1970 年 IBM 公司研究员 E.F.Codd 博士，即“关系数据库之父”，发表的业界第一篇关于关系数据库理论的论文“A Relational Model of Data for Large Shared Data Banks”。从此关系型数据库如雨后春笋般四处萌发，各大厂商争先恐后，加入到关系型数据库的研发大潮中，而后再如大浪淘沙般去璞存真，时至今日，留下的是真正适用于客户、适应于潮流的关系型数据库产品。

IBM 公司作为关系数据库的推广先锋，为业界提供了一批优秀的数据库技术领域先驱科学家，他们所研发出的 DB2 数据库，经过近 50 年的发展，已经广泛应用于金融、电信、制造等多个行业，对日常的工作和生活带来了深远的影响。

在中国，DB2 数据库的兴起大概在 15 年前，当时国内鲜有应用，牛新庄先生作为国内首批接触研究 DB2 数据库的工程师，为业界提供了大量技术服务和专业培训，为 DB2 数据库的推广应用做出了积极杰出的贡献。同时，也为 IBM 公司反馈了很多很好的 DB2 研发建议，为关系型数据技术的长远发展贡献智慧。在数据库方向的精深造诣和丰富实践，就浓缩在他的 DB2 数据库著作中，为广大 IT 同仁授业解惑。

本套图书可以说是伴随着 DB2 数据库的成长，从第一版主讲 DB2 V8，到第二版的 DB2 V9，再到第三版的 DB2 V10。基本上每出一个版本我都会仔细品味，每个版本作者都很用心，都会删减对当前不适用的章节，加入很多新的功能和他近期实际经历的案例。经典性、权威性、实用性是本套书籍锁定的主要目标。

第一，本套书籍涵盖了 DB2 的几乎所有功能，是业界最大规模的系统梳理与总结。从理论知识到最佳实践，无所不包，无所不有，同时还总结了几十个最佳案例。

第二，本套书籍理论讲解深入浅出，案例多样详实，无论是对零基础还是拥有多年 DBA 工作经验的人都非常适用。

第三，精品的价值在于传世久远，经典的意义在于常读常新。我认为，只有被广泛阅读，受到大家喜爱、接受的作品，才具有经典的资质与意义。希望作者继续努力，将本套书籍打磨成关系数据库的经典图书。

其中第一本书《循序渐进 DB2 DBA 系统管理、运维与应用案例(第 3 版)》，是 DB2 学习的入门书籍。该书包含了从入门到中级阶段的知识技能的介绍，全面展示了 DB2 的主要功能和日常的工作技巧，尤其是实例和案例部分，这部分内容从实际出发为读者列举了常见的案例场景和处理办法，非常实用。在新的一版中，所有的内容、示例都基于 DB2 V10.5 版本进行了修订，并介绍了 DB2 V11.1 中的新功能、新特性。

第二本书《DB2 高级管理、系统设计与诊断案例(第 3 版)》是 DB2 学习的高级进阶，该书从 DB2 体系结构入手，介绍了 DB2 各个内部组件的层次与功能、内存内部结构、存储内部结构、锁和并发原理等。在理解 DB2 内部基本原理的基础上进一步介绍了 DB2 的高级功能，包括分区功能、高级压缩功能等。此外，系统介绍了 OLTP 和 OLAP 系统的设计方法和管理技术、高可用和容灾方案，以及集群技术，其中涉及 HADR、DPF 和 pureScale 技术，以及地理上分离的 pureScale 集群(GDPC)技术。该书还介绍了 DB2 各种监控和诊断方法，通过精选的诊断案例使读者在学习知识的同时积累实践经验。

第三本书《DB2 数据库性能调整和优化(第 3 版)》专门介绍 DB2 性能调整和优化，从 DB2 数据库性能有关的基础知识和原理入手，从数据库所处的运行环境(OS、存储等)开始介绍，并对 DB2 的进程和内存进行深入讲解。在全面了解性能相关知识后，开始逐步展开，从设计到监控，从配置参数到调优工具，从锁和并发到优化器统计信息，最后列出了几个完整的性能调优案例以增加技术理解。

祝愿每一位读者能有所得、有所悟，成长为新一代的数据技术专家，也祝愿牛新庄先生在数据技术领域这条康庄大道上走得更宽更远。

IBM 前大中华区总经理

王天义

IBM 大中华区高级顾问

前 言

自 1999 年左右我开始从事数据库有关的技术工作到现在已近 20 年时间，此期间信息科技飞速发展，从无纸化办公和数据大集中到移动互联和大数据、人工智能、云计算等信息技术改变了生活，并颠覆了传统商业模式。信息科技的发展离不开数据处理技术的进步，在这一轮信息化浪潮中，数据处理技术也发生了翻天覆地的变化，对企业经营发展和对外服务的意义越来越重要。一方面，传统企业级数据库的能力，在原有的道路和方向上不断地持续提升演进，以满足企业市场不断迸发的各类需求。另一方面，互联网场景孕育的各种新兴的数据处理技术亦不断涌现，例如 NewSQL、NoSQL、Hadoop 等大数据处理技术，这些技术成为传统数据库产品的必然补充，同时也对传统数据库产品产生了一定的冲击。但是以我长期从事企业数据处理相关工作的经验看，在企业级市场尤其是金融企业市场里面，传统数据库产品的能力依然是解决企业主要业务需求的不二选择。因此，传统数据库技术的研究和应用仍然是信息科技工作的重点。

近年来传统数据库产品在不断改进升级，以支持更快的处理能力和更高的可用性，满足不同场景下的用户需求。DB2 作为一款主流数据库产品，在这些方面也都进步明显，例如 Purescale 集群技术、跨数据中心的 GDPC 技术、列存储的 BLU 技术等创新功能就表现不俗，满足了特定业务场景需求，给企业带来了很大的价值提升。特别是 GDPC 技术，帮助企业搭建关键业务系统同城对等全双活生产架构，为最终用户提供高等级容灾的连续服务，对企业对外服务的提升意义非凡，也使数据库从业者们领略了 DB2 产品创新的精华。

基于 DB2 产品的演进以及近些年的思考和实践，我重新梳理了之前编写的第 2 版的 3 本 DB2 系列技术图书，对其进行了大篇幅的修改和重写，力图对近些年实践的精华和 DB2

产品的新趋势进行总结。在此奉献给各位数据库从业的同仁，在技术的路上共勉。

由于本人水平有限，时间有限，书中不免有这样或者那样的错误，希望广大读者朋友不吝赐教指正！

最后，感谢我的家人和同事在本书重写过程中的帮助，谢谢你们！

牛新庄

目 录

第 1 章 DB2 介绍和安装	1	3.1.4 数据库目录	63
1.1 DB2 数据库概述	1	3.2 设计表空间	66
1.1.1 DB2 发展历史	1	3.2.1 创建表空间	66
1.1.2 DB2 版本和平台支持	5	3.2.2 维护表空间	69
1.1.3 DB2 产品组件和功能	9	3.2.3 表空间设计注意事项	76
1.2 DB2 数据库的安装与配置	12	3.2.4 prefetchsize 大小选择	81
1.2.1 DB2 在 Windows 上的安装	13	3.2.5 文件系统(CIO/DIO)和裸设备	82
1.2.2 DB2 在 Linux/UNIX 上的安装	21	3.2.6 设置 OVERHEAD 和 TRANSFERRATE	85
1.3 DB2 数据库的体系结构	23	3.2.7 优化 RAID 设备上表空间的性能	86
第 2 章 创建实例	31	3.2.8 合理设置系统临时表空间	88
2.1 实例	31	3.3 缓冲池	89
2.1.1 实例的概念	31	3.3.1 缓冲池的使用方法	89
2.1.2 创建实例	32	3.3.2 缓冲池和表空间之间的关系	90
2.1.3 实例目录	35	3.3.3 维护缓冲池	91
2.1.4 实例的相关命令	39	3.3.4 缓冲池的设计原则	93
2.1.5 DB2INSTANCE 变量介绍	43	3.4 多温度存储器	96
2.1.6 删除实例	44	3.4.1 存储器组	96
2.1.7 配置实例	45	3.4.2 表空间与存储器组	100
第 3 章 创建数据库和表空间	47	3.5 本章小结	101
3.1 创建数据库	47	第 4 章 访问数据库	103
3.1.1 DB2 数据库存储模型	49	4.1 访问 DB2	103
3.1.2 表空间管理类型	51		
3.1.3 创建数据库	54		

4.2	DB2 CLP 处理程序	104	5.2.2	选择合适的约束类型	142
4.2.1	DB2 CLP 简介	104	5.2.3	使用 not null with default	145
4.2.2	DB2 CLP 设计	104	5.2.4	生成列及应用案例	145
4.2.3	DB2 CLP 命令选项	106	5.2.5	自动编号和标识列应用 案例	146
4.2.4	设置 DB2_CLPPROMPT 以定制 DB2 CLP	110	5.2.6	使用 not logged initially 特性	147
4.3	配置 DB2 服务器的 TCP/IP 通信	114	5.2.7	使用 append on 特性	148
4.3.1	在服务器上更新 services 文件	115	5.2.8	数据、索引和大对象分开 存放	149
4.3.2	在服务器上更新数据库 管理器配置文件	116	5.2.9	设置 pctfree	149
4.3.3	设置 DB2 服务器的通信 协议	117	5.2.10	表的 locksize	150
4.3.4	查看服务器通信端口的 状态	117	5.2.11	表的 volatile 特性	150
4.4	配置客户机至服务器通信	117	5.2.12	表维护相关命令	151
4.4.1	客户机至服务器通信概述	117	5.2.13	表设计高级选项	155
4.4.2	深入了解 DB2 节点目录、 数据库目录	118	5.3	索引设计	160
4.4.3	使用 CLP 配置客户机到 服务器通信的案例	124	5.3.1	索引的优点	160
4.5	实际生产中连接数据库的 各种方式	128	5.3.2	索引类型	161
4.6	案例: 数据库连接问题诊断	131	5.3.3	索引结构	164
4.7	本章小结	134	5.3.4	理解索引的访问机制	167
第 5 章	创建数据库对象	135	5.3.5	创建集群索引	170
5.1	模式	135	5.3.6	创建双向索引	171
5.1.1	模式的概念	135	5.3.7	完全索引访问 (index access only)	172
5.1.2	系统模式	137	5.3.8	创建索引示例	173
5.1.3	设置和获得当前模式	137	5.3.9	索引总结	179
5.1.4	模式和用户的区别	138	5.4	使用序列提高性能	181
5.2	表设计	139	5.4.1	应用程序性能和序列	181
5.2.1	选择合适的数据类型	139	5.4.2	序列的设计原则	182
			5.4.3	维护序列	183
			5.4.4	比较序列与标识列	187
			5.5	视图	189
			5.5.1	视图的类型	189
			5.5.2	创建 with check option 视图	193

5.5.3	维护视图	194	6.4.7	使用 CURSOR 文件类型 移动数据	236
5.6	表表达式	195	6.4.8	提高 LOAD 性能	238
5.6.1	嵌套的表表达式	195	6.4.9	LOAD 失败恢复	242
5.6.2	公用表表达式	195	6.4.10	LOAD 和 IMPORT 的 比较	245
5.7	触发器设计	197	6.5	数据移动的性能问题	246
5.7.1	触发器的类型	197	6.6	db2move 和 db2look	248
5.7.2	触发器创建示例	199	6.6.1	数据库移动工具—— db2move	248
5.7.3	触发器设计总结	201	6.6.2	DB2 DDL 提取工具 ——db2look	250
5.8	例程	202	6.6.3	利用 db2move 和 db2look 移动数据的案例	251
5.9	本章小结	203	6.6.4	带 COPY 操作的 db2move 实用程序	254
第 6 章	数据移动	205	6.7	本章小结	259
6.1	数据移动格式	205	第 7 章	数据库备份与恢复	261
6.1.1	定界 ASCII 文件格式	206	7.1	恢复的概念	261
6.1.2	非定界 ASCII 文件格式	206	7.1.1	崩溃恢复(Crash Recovery)	265
6.1.3	PC/IXF 文件格式	207	7.1.2	灾难恢复(Disaster Recovery)	266
6.1.4	工作表文件格式	207	7.1.3	版本恢复(Version Restore)	266
6.1.5	游标	207	7.1.4	前滚恢复(RollForward Recovery)	267
6.2	EXPORT	208	7.2	DB2 日志	269
6.2.1	EXPORT 概述	208	7.2.1	日志文件的使用	270
6.2.2	导出数据	208	7.2.2	日志类型	272
6.2.3	导出数据示例	211	7.2.3	日志相关配置参数	275
6.3	IMPORT	212	7.2.4	数据库日志总结	276
6.3.1	IMPORT 概述	212	7.2.5	DB2 日志的建议设置	278
6.3.2	导入数据	212	7.3	数据库和表空间备份	280
6.3.3	导入示例	217	7.3.1	数据库备份	280
6.4	LOAD	219	7.3.2	表空间备份	282
6.4.1	LOAD 概述	219			
6.4.2	装入数据	220			
6.4.3	装入示例	225			
6.4.4	在线 LOAD	229			
6.4.5	监控 LOAD 进度	232			
6.4.6	LOAD 期间和之后的表空间 状态	233			

7.3.3	增量备份	282
7.3.4	检查备份完整性—— db2ckbkp	283
7.4	数据库和表空间恢复	285
7.4.1	数据库恢复	285
7.4.2	表空间恢复	288
7.4.3	增量恢复	289
7.4.4	增量恢复检查——db2ckrst	290
7.4.5	重定向恢复	291
7.4.6	恢复已删除的表	294
7.5	数据库和表空间前滚	298
7.5.1	数据库前滚	298
7.5.2	表空间前滚	299
7.6	RECOVER 实用程序	302
7.7	恢复历史文件	306
7.8	数据库重建	309
7.8.1	数据库重建的概念	309
7.8.2	使用表空间备份重建可 恢复数据库	309
7.8.3	只使用部分表空间备份 重建可恢复数据库	312
7.8.4	使用包含日志文件的在线 备份重建数据库	314
7.8.5	使用增量备份映像重建 可恢复数据库	314
7.8.6	使用重定向选项重建 可恢复数据库	315
7.8.7	重建不可恢复数据库	315
7.8.8	数据库重建的限制	316
7.9	监控备份、复原和恢复进度	316
7.10	备份、恢复和复原期间的 表空间状态	318
7.11	优化备份、复原和恢复 性能	318

7.12	备份恢复最佳实践	320
------	----------	-----

第 8 章	SQL 基础知识	323
8.1	简单查询入门	323
8.1.1	SELECT 和 FROM	324
8.1.2	WHERE	324
8.1.3	ORDER BY	324
8.1.4	GROUP BY 和 HAVING	325
8.2	搜索条件	326
8.2.1	谓词种类	326
8.2.2	基本谓词	326
8.2.3	量化谓词	327
8.2.4	BETWEEN、EXISTS 和 IN 谓词	328
8.2.5	LIKE 谓词	328
8.2.6	NULL 谓词	329
8.3	数据操作语言	329
8.3.1	INSERT	329
8.3.2	DELETE	330
8.3.3	UPDATE	331
8.3.4	MERGE	332
8.4	多表查询	334
8.4.1	JOIN 连接	334
8.4.2	集合运算	335
8.5	高性能的 SQL 语句	337
8.5.1	高效 SQL 的准则	337
8.5.2	提高插入性能的准则	339
8.5.3	复杂查询的准则	340
8.5.4	索引的注意事项	341
8.6	本章小结	342
第 9 章	DB2 基本监控方法	343
9.1	监控工具概述	343
9.2	快照监视器	345
9.3	利用表函数监控	350

9.4	性能管理视图及案例	352	11.2.4	包含列组统计信息的 RUNSTATS	422
9.5	快照监视器案例	358	11.2.5	包含 LIKE STATISTICS 的 RUNSTATS	422
9.5.1	监控案例 1——动态 SQL 语句	358	11.2.6	包含统计信息配置文件 的 RUNSTATS	422
9.5.2	监控案例 2——通过表函数 监控	361	11.2.7	带有抽样的 RUNSTATS	423
9.5.3	编写快照监控脚本	362	11.2.8	带有系统页级抽样 的 RUNSTATS	424
9.5.4	db2pd 及监控案例	363	11.2.9	收集统计信息的其他可供 选择的方法	424
9.5.5	事件监视器及监控案例	370	11.2.10	RUNSTATS 总结	426
9.5.6	db2mtrk 及监控案例	375	11.3	碎片整理	426
9.6	本章小结	377	11.3.1	表重组(REORG)	426
第 10 章	运行数据库必须考虑的数据库 设置	379	11.3.2	索引重组	434
10.1	数据库配置参数概述	379	11.3.3	重组表和索引的成本	440
10.2	通信设置	381	11.3.4	合理设计以减少碎片 生成	441
10.3	内存有关的设置	383	11.3.5	启用表和索引的自动 重组	442
10.4	锁有关的设置	389	11.4	碎片整理案例分析	443
10.5	日志相关的配置	394	11.4.1	执行表、索引检查是否 需要做 REORG	443
10.6	自动维护相关的配置	399	11.4.2	表和索引碎片整理	445
10.7	监控相关的配置	400	11.5	案例：生成碎片检查、统计 信息更新、碎片整理和 REBIND 脚本	445
10.8	安全相关的设置	402	11.6	重新绑定程序包	446
10.9	供参考的 DB2 上线前设置	402	11.7	DB2 健康检查	447
10.10	本章小结	405	11.7.1	查看是否有僵尸实例 进程	448
第 11 章	DBA 日常运行维护	407	11.7.2	检查数据库是否一致	448
11.1	统计信息更新	407	11.7.3	查找诊断日志以判断是否 有异常	449
11.1.1	统计信息的重要性	408			
11.1.2	减小 RUNSTATS 对系统 性能影响的策略	415			
11.1.3	DB2 自动统计信息收集	416			
11.2	统计信息更新案例分析	418			
11.2.1	RUNSTATS 更新示例	418			
11.2.2	收集分布式统计信息	419			
11.2.3	包含频率和分位数统计 信息的 RUNSTATS	420			

11.7.4	检查数据库备份完整性、 日志归档是否正常	449	11.8.19	查找新创建的对象	458
11.7.5	维护实例目录和数据库 目录的权限	451	11.8.20	查找无效对象	458
11.7.6	查看磁盘空间	451	11.8.21	检查表空间状态	458
11.8	数据库监控	453	11.8.22	检查表状态	459
11.8.1	监控工具	453	11.8.23	查找需要 REORG 的表和 索引	459
11.8.2	计算数据库的大小	454	11.8.24	查找需要 RUNSTATS 的 表和索引	461
11.8.3	监控表的物理大小	455	11.8.25	定期清理 db2diag.log 文件	461
11.8.4	监控单个索引的大小	455	11.8.26	查找异常增长的表空间 和表	462
11.8.5	监控数据库实用工具的 进度	455	11.8.27	数据库维护总结	462
11.8.6	监控数据库 crash recovery 进度	455	第 12 章	数据库常用工具	465
11.8.7	监控 catalog cache 命中率	455	12.1	解释工具	465
11.8.8	监控 package cache 命中率	455	12.1.1	Visual Explain (可视化解释)	465
11.8.9	监控排序溢出率	456	12.1.2	db2expln	473
11.8.10	监控正在 REORG 的表	456	12.1.3	db2exfmt	476
11.8.11	监控缓冲池命中率	456	12.1.4	各种解释工具的比较	477
11.8.12	监控高成本应用程序	456	12.1.5	如何从解释信息中获取 有价值的建议	478
11.8.13	监控正在执行的时间 最长的 SQL 语句	456	12.2	索引设计工具(db2advis)	479
11.8.14	监控 SQL 准备和预编译 时间最长的 SQL 语句	457	12.2.1	DB2 Design Advisor(db2advis)	479
11.8.15	监控执行次数最多的 SQL 语句	457	12.2.2	DB2 Design Advisor (db2advis)案例讲解	480
11.8.16	监控执行时间最长的 SQL 语句	457	12.3	基准测试工具 db2batch	483
11.8.17	监控排序次数最多的 SQL 语句	457	12.3.1	db2batch	483
11.8.18	监控引起锁等待的 SQL 语句	457	12.3.2	db2batch 基准程序测试 分析示例	485
			12.4	数据一致性检查工具	486
			12.4.1	db2dart 及案例	486
			12.4.2	inspect 及案例	487

12.5	db2look	488	13.2.2	兼容性增强	499
12.5.1	db2look 概述	488	13.3	安全性功能增强	500
12.5.2	利用 db2look 构建模拟 测试数据库	490	13.3.1	行和列访问控制(RCAC) 支持按列组织的表	500
12.6	其他工具	492	13.3.2	企业加密码钥管理	501
12.6.1	db2bfd	492	13.4	pureScale 功能增强	503
12.6.2	db2_kill 和 db2nkill	493	13.5	可管理性和性能增强	504
12.6.3	db2tbst	493	13.5.1	可管理性增强功能	504
12.7	本章小结	494	13.5.2	简易大版本升级	505
第 13 章	DB2 V11 新特性	495	13.5.3	联合增强功能	506
13.1	分区数据库(DPF)支持 BLU	495	13.5.4	高可用性、备份、日志记录、 弹性和恢复增强功能	506
13.1.1	DB2 with BLU Acceleration 基本介绍	495	13.5.5	性能增强	507
13.1.2	BLU 的用法	497	13.6	备份增强	508
13.1.3	DB2 V11 支持在 MPP 中 使用 BLU	498	13.6.1	基本介绍	508
13.2	BLU 增强	499	13.6.2	备份比较	508
13.2.1	性能增强	499	13.6.3	恢复场景	509

DB2 介绍和安装

DB2 LUW(IBM DB2 Database for Linux、UNIX and Windows, 本书后续章节中, 统一简称为 DB2)是 IBM 于 1983 年推出的第一款面向大型企业的商业化关系数据库管理系统。在 20 世纪 80 年代初, DB2 的发展重点放在大型的主机平台, 从 80 年代中期到 90 年代初, DB2 已发展到中型机、小型机以及微机平台。DB2 的诞生不仅促进了与关系数据库概念相关的数学和科学的发展, 还创造性地开发出一种极具影响力的全新软件类型。今天, DB2 已经发展成为 IBM 信息管理(Information Management, IM)软件组合的重要组成部分。在 IBM 信息按需应变策略和体系结构中, DB2 扮演数据基础服务平台的重要角色, 并且已经发展成同时支持传统关系数据和 XML 的混合型数据服务器。传承 IBM 数据库的优良传统并具有突破性的数据库产品 DB2 LUW V10.5 已经于 2013 年 6 月份问世, 它在原有 V10.1 版本的基础上, 增加了众多革命性的技术(BLU Acceleration), 使 DB2 LUW 产品在多个领域实现了突破。

DB2 数据库产品及解决方案广泛应用在金融、电信、制造、零售、保险等行业及政府机关, 以数据库技术创新帮助客户实现更大价值, 以技术创新推动商业模式的变革和不断发展。

本章主要讲解如下内容:

- DB2 数据库概述
- DB2 数据库的安装和配置
- DB2 数据库的体系结构

1.1 DB2 数据库概述

1.1.1 DB2 发展历史

我们都知道 DB2 是关系型商用数据库的一种, 那么在开始学习 DB2 之前, 先来了解一

下数据库的发展历史。

1. 数据库的发展历史

在没有数据库之前，人们靠什么来记录数据呢？最早是靠文件，但是用文件记录有很多缺点，例如不易保存和共享等，而数据库的出现可以解决这些问题。数据库的历史可以追溯到 50 多年前，当时计算机开始广泛地应用于数据管理，对数据的共享提出了越来越高的要求。传统的文件方式已经不能满足人们的需要。能够统一管理和共享数据的数据库管理系统(DBMS)应运而生。数据模型是数据库系统的核心和基础，各种 DBMS 软件都是基于某种数据模型的。所以通常也按照数据模型的特点将传统数据库系统分成网状数据库、层次数据库和关系数据库 3 类。最早出现的是网状 DBMS，1961 年通用电气公司(General Electric)的 Charles Bachman 成功地开发出世界上第一个网状 DBMS，也是第一个数据库管理系统——集成数据存储(Integrated Data Store, IDS)，奠定了网状数据库的基础，并在当时得到了广泛的发行和应用。IDS 具有数据模式和日志的特征，但它只能在 GE 主机上运行，并且数据库只有一个文件，数据库所有的表必须通过手工编码来生成。网状数据库模型对于层次和非层次结构的事物都能比较自然地模拟，在关系数据库出现之前，网状 DBMS 要比层次 DBMS 用得普遍。在数据库发展史上，网状数据库占有重要地位。层次 DBMS 是紧随网状 DBMS 而出现的。最著名、最典型的层次数据库系统是 IBM 公司在 1968 年开发的 IMS(Information Management System)——一种适合其主机的层次数据库。这是 IBM 公司研制的最早的大型数据库系统产品。从 20 世纪 60 年代末产生起，如今已经发展到 IMS V10，提供对群集、N 路数据共享、消息队列共享等先进特性的支持。这个具有 50 年历史的数据库产品在如今的 WWW 应用连接、商务智能应用中仍扮演着新的角色。目前国内的 4 大银行在主机上仍然在使用 IMS 数据库。

关系数据库

关系数据库的由来：网状数据库和层次数据库已经很好地解决了数据的集中和共享问题，但是在数据独立性和抽象级别上仍有很大欠缺。用户在对这两种数据库进行存取时，仍然需要明确数据的存储结构，指出存取路径，而后来出现的关系数据库较好地解决了这些问题。1970 年，IBM 的研究员 E.F.Codd 博士在刊物 *Communication of the ACM* 上发表了一篇名为“A Relational Model of Data for Large Shared Data Banks”的论文，提出了关系模型的概念，奠定了关系模型的理论基础。尽管之前在 1968 年 Childs 已经提出了面向集合的模型，然而这篇论文被普遍认为是数据库系统历史上具有划时代意义的里程碑。Codd 的心愿是为数据库建立优美的数据模型。后来 Codd 又陆续发表多篇文章，论述了范式理论和衡量关系系统的 12 条标准，用数学理论奠定了关系数据库的基础。IBM 的 Ray Boyce 和 Don Chamberlin 将 Codd 关系数据库的 12 条准则的数学定义以简单的关键字语法表现出

来,里程碑式地提出了 SQL(Structured Query Language)语言。关系模型有严格的数学基础,抽象级别比较高,而且简单清晰,便于理解和使用。但是当时也有人认为关系模型是理想化的数据模型,用来实现 DBMS 是不现实的,尤其担心关系数据库的性能难以接受,更有人视其为当时正在进行中的网状数据库规范化工作的严重威胁。为了促进对问题的理解,1974年 ACM 牵头组织了一次研讨会,会上开展了一场分别以 Codd 和 Bachman 为首的支持和反对关系数据库两派之间的辩论。这次著名的辩论推动了关系数据库的发展,使其最终成为现代数据库产品的主流。而 Oracle 公司在 1979 年开发了第一个商用 SQL 关系数据库管理系统。关系数据库系统以关系代数为坚实的理论基础,经过几十年的发展和实际应用,技术越来越成熟和完善,代表产品有 Oracle、DB2、SQL Server 以及 Informix、Sybase 等等。

面向对象数据库

随着信息技术和市场的发展,人们发现关系数据库系统虽然技术很成熟,但其局限性也是显而易见的——它能很好地处理所谓的“表格型数据”,却对技术界出现的越来越多的复杂类型的数据无能为力。20世纪90年代以后,技术界一直在研究和寻求新型数据库系统。但在什么是新型数据库系统的发展方向的问题上,产业界一度是相当困惑的。受当时技术风潮的影响,在相当一段时间内,人们把大量的精力花在研究“面向对象的数据库系统(Object-Oriented Database)”或简称“OO数据库系统”。值得一提的是,由美国 Stonebraker 教授提出的面向对象的关系数据库理论曾一度受到产业界的青睐。而 Stonebraker 本人也在当时被 Informix 花大价钱聘为技术总负责人。然而,数年的发展表明,面向对象的关系数据库系统产品的市场发展情况并不理想。理论上的完美性并没有带来市场的热烈反应。其不成功的主要原因在于,这种数据库产品的主要设计思想是企图用新型数据库系统来取代现有的数据库系统。这对许多已经运用数据库系统多年并积累了大量工作数据的客户,尤其是大客户来说,是无法承受因为新旧数据间的转换而带来的巨大工作量及巨额开支的。另外,面向对象的关系数据库系统使查询语言变得极其复杂,从而使得无论是数据库的开发商还是应用客户,都视其复杂的应用技术为畏途。

混合数据库

IBM 经过多年的积累和持续创新,在 2006 年年底率先推出了第一个直接支持 XML 的混合数据服务器——IBM DB2 V9(代号为 Viper)。

IBM DB2 提供了与以前版本非常不同的体系结构,它通过提供新的查询语言、新的存储技术、新的索引技术和支持 XML 数据及其固有层次结构的特性,使得 IBM DB2 成为 IBM 的第一个“混合型”(即多结构)数据库管理系统。除了支持表数据模型外,DB2 还支持 XML 文档和消息中固有的层次化数据模型。用户可以在表中自由地混合存储传统 SQL