



SAP中国官方资深专家撰写，SAP HANA官方中文社区和HANA中国社区
联袂推荐，权威性毋庸置疑

详细讲解HANA的各个知识点及其使用方法，深入分析技术背后的本质，
以实践为导向，包含大量有价值的实用案例，注重阅读体验



刘刚 舒戈 著

SAP HANA 实战



机械工业出版社
China Machine Press



SAP HANA in Action

SAP HANA 实战

刘刚 舒戈 著



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

SAP HANA 实战 / 刘刚, 舒戈著. —北京: 机械工业出版社, 2013.11
(实战)

ISBN 978-7-111-44542-5

I. S… II. ①刘… ②舒… III. 企业管理—应用软件 IV. F270.7

中国版本图书馆 CIP 数据核字 (2013) 第 250886 号

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问 北京市展达律师事务所

本书由 SAP 中国官方资深专家撰写, SAP HANA 官方中文社区和 HANA 中国社区联袂推荐, 权威性毋庸置疑。内容全面, 详细讲解了 HANA 的各个知识点及其使用方法和技巧; 实战性强, 每个知识点都配有操作说明, 包含大量有价值的实用案例; 关注细节、分析透彻, 深入剖析了关键技术背后的本质和原理; 使读者知其然, 也知其所以然, 注重阅读体验, 目标是希望读者能“一看就会, 一用就通”。

全书共 13 章, 分为三个部分: 第一部分 (1~4 章) 主要介绍了内存计算技术产生的背景及其优势, HANA 的软件架构及其在架构上的创新, 以及 HANA 的应用场景; 第二部分 (5~10 章) 系统讲解了 HANA 的功能模块及其详细使用方法, 包括 HANA 相关的安装和配置, 不同的数据加载到 SAP HANA 的方式, 如何在 SAP HANA 工作台中进行信息建模, HANA 和商务智能的结合, 以及如何使用 BO、VI、Excel、Tableau 等不同的 BI 工具来访问 HANA 中的模型和数据; 第三部分 (11~13 章) 主要讲解了 SAP HANA 的应用开发 (ABAP、加速器和 R 语言), HANA 分布式集群架构, HANA 数据库系统管理 (数据管理、安全管理、审计监控管理、配置管理等) 等内容。

机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码 100037)

责任编辑: 姜 影

藁城市京瑞印刷有限公司印刷

2014 年 1 月第 1 版第 1 次印刷

186mm × 240mm • 29.25 印张

标准书号: ISBN 978-7-111-44542-5

定 价: 89.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzsjj@hzbook.com

Tape is Dead, Disk is Tape, Flash is Disk, RAM Locality is King.

磁带已经死了，磁盘已经落伍，闪存成为新存储，内存局部性才是王道。

By Jim Gray, 2006

前 言

为什么要写这本书

如果您曾经听说过 Jim Gray (吉姆·格雷), 那么相信您一定看到过本书开头的那句话, 作为 IT 史上最伟大的天才之一和数据库的奠基人, Jim Gray 的上述预言如今已经一一成为现实。

SAP HANA 是内存计算技术发展的一个重要里程碑, 它的出现宣示着高性能内存计算技术大范围的商用时代已经到来。与此同时, SAP 这家企业管理软件公司也正在凭借 SAP HANA 引领企业现代信息化建设走向一个新高度, 即“实时计算”。SAP 商务套件已经正式基于 SAP HANA, 而 BI/BW 商务智能平台也早已经基于 SAP HANA, 另外, SAP 云计算产品和移动应用也已经基于 SAP HANA, 同时, 其他合作伙伴已经基于 SAP HANA 开发了很多应用软件, 这使得在 SAP HANA 的应用方式上能看到无限种可能。

在撰写本书的过程中, 我们一直觉得非常幸运, 因为可以和 SAP HANA 如此一项伟大的新技术在中国地区共同成长。鉴于图书市场上 HANA 相关书籍较少, 为了让更多的人系统地了解并深入地掌握 SAP HANA 这一革命性创新产品, 同时也是让 SAP HANA 的客户更好地利用这一高性能实时数据计算平台, 于是开始撰写本书。我们将过去参与过的一些 SAP HANA 项目经验融入本书, 同时尽量用简洁明了的语言进行叙述, 以便让读者看得懂、学得会、用得通, 并希望起到一个抛砖引玉的作用。

作为一个纯技术型产品, 无论什么行业, 或者何种规模的企业用户, 都可以基于 SAP HANA 平台支撑其各种各样的企业应用。例如, 日本野村综合研究所 (Nomura Research Institute) 基于 HANA 开发的东京实时出租车 GPS 信息最佳路径导航系统; MKI 利用 SAP HANA 加速基因序列检测从而使个性化医疗服务成为现实; 友都八喜实时商场 POS 交易数据的分析系统; iOS 的社交应用软件 (My Runway、Call Back 等)。

本书特色

让读者“绝对看得懂, 一定学得会”是作者在写作本书时对自己的要求, 并力求将所有的相关

的技术尽量用简洁的语言讲解清楚。考虑到有很多读者没有 SAP 系统或 HANA 相关的背景知识，所以本书中的一些章节会提供尽可能多的前提说明，使读者能够轻松地读完并掌握相关知识。

- 涵盖各种基础的理论，使读者知其然并能知其所以然。例如，列存储怎样才能实现高性能 OLTP、数据读取的时候为什么可以不加锁。

- 知识点均结合实践操作。即使是输入用户名的操作，也会讲解这个用户名的作用。

- 精选极具价值、实用性强的示例。

本书包含 SAP HANA 实施项目中会用到的重要技术，包括系统架构、硬件估算、系统安装、数据准备、系统建模和开发管理、数据应用、系统管理等，并且在所有的重要过程中，都提供了“手把手”式的介绍和帮助，让读者在一个较轻松的状态下就能够掌握 SAP HANA 常用的知识点。本书尽量做到讲解技术知识时不枯燥，并以“通俗、好懂、易操作”为最终的追求目标。

本书很多章的开头列出了一些技术名词，并且给出简洁、易懂的描述，这样做是便于在后续介绍到这些术语时读者能够快速理解。因为全书讲解的都是与 SAP HANA 相关的技术，所以对其他的数据处理产品的讲解并不多，即使有提及或评论，也是根据作者自身的知识和理解来对比的。如果读者想要了解相关细节，那么可以参考这些产品的官方文档和书籍。

因为 SAP HANA 是一个技术平台型产品，往往作为各种规模及各种类型应用系统的底层平台，另外，本书讲解的技术知识点都是通用的，所以本书并未涉及任何关于项目管理和实施方法等话题。

如何阅读本书

和学习其他 IT 技术一样，在学习 SAP HANA 的过程中，很多内容之间其实是有相互依赖关系的。举个最简单的例子，如果不会安装产品，就只能等着专门的技术顾问安装好，然后才能开始学习建模。例如，一开始学习如何在 SAP HANA 中创建数据库表，或者如何浏览数据库表中的数据，是需要做一些用户权限设定的。又如，在学习创建视图时，其实需要预先定义 Content Vendor 和创建 Package，这其实是和 SAP HANA 应用内容的生命周期管理相关的内容。很多实际操作都会跨不同的内容领域，但是如果分开讲解，那么读者一上手就会立即遇见困难，所以在每章涉及动手操作的地方，都会尽量讲解和当前实践操作相关的要点和提示信息，确保读者可以一路顺畅地学习下去。

本书一共分为三个部分，每个部分包含若干章，每一章分别侧重讲解 SAP HANA 中某一个技术点或体系架构的一个部分。

第一部分包含第 1 ~ 4 章，主要讲解了内存技术产生的背景，SAP HANA 软件架构及 SAP HANA 在软件架构上的创新，如何克服过去的旧的内存计算技术，如何将列存储、行存储，以及内存计算技术和多核大内存的硬件架构进行融合，如何在新硬件架构下创造出新一代的高性能内存计

算技术，还讲解了 SAP HANA 典型的应用场景。

第二部分包含第 5 ~ 10 章，主要讲解 SAP HANA 的操作性知识，包括软件安装、各种不同的数据加载到 SAP HANA 的方式，以及如何在 SAP HANA 工作台中进行信息建模（属性视图、分析视图、计算视图）等。同时，还介绍了 SAP HANA 和商务智能的结合，并且使用 BO、VI、Excel、Tableau 等不同的 BI 工具来访问 SAP HANA 中的模型和数据。

第三部分包含第 11 ~ 13 章，主要讲解 SAP HANA 的应用开发（ABAP、加速器和 R 语言）、HANA 分布式集群架构、HANA 数据库系统管理（数据管理、安全管理、审计监控管理、配置管理等）等内容。

本书面向的读者

- 对高性能内存计算感兴趣的读者。
- DBA 和 SAP 业内的咨询顾问，例如 ABAP 顾问、BASIS 顾问、BI 顾问、BW 顾问等。
- 希望向 SAP HANA 这一领域发展的读者，当然这部分读者需要具备一定的基础知识才能更好地理解某些章节，例如需要了解 SQL、EDW 建模，以及其他一些数据库基础知识、基本程序设计等。
- SAP 合作伙伴或基于 SAP HANA 进行应用开发的软件公司和咨询公司等。

勘误和支持

和 SAP 过去推出的任何一款软件产品都不同的是，SAP HANA 发布之后，从市场的拥有份额和客户的满意度上来看，它的发展速度极大地超出所有人的想象。当然，SAP HANA 也在不断变化，不断地推出新的功能，本书的很多章节都经历了系统功能的变化，所以进行了多次的修订。若你阅读本书时发现自己所使用的产品功能和本书介绍的有差异，欢迎提出反馈意见。

由于不同版本的 SAP HANA 在功能上略有差异，本书不可能面面俱到，望读者见谅。如果你在阅读本书的过程中，遇到了技术理解上的相关问题，可以利用下面列出的电子邮件地址将问题发送给作者，或者加入到本书的新浪微群中，参与关于 SAP HANA 特性和技术的讨论、分享你的读书心得，同时也可以进行问题反馈。

电子邮件：sap.hana.china@gmail.com

新浪微群：1040765（微群名：HANA）

目 录

前言

第 1 章 话说内存计算 1

- 1.1 企业应用系统架构和限制因素 2
- 1.2 难以实现的实时计算 3
 - 1.2.1 信息、延迟和价值 4
 - 1.2.2 OLTP 和 OLAP 合理分离 6
- 1.3 软件架构与硬件发展 6

第 2 章 实时数据平台 9

- 2.1 旧瓶颈和新问题 10
 - 2.1.1 内存计算也有麻烦 12
 - 2.1.2 多核 CPU 时代 13
 - 2.1.3 打破 FSB 瓶颈 13
 - 2.1.4 高效利用 NUMA 15
 - 2.1.5 CAP 的偏向性 16
- 2.2 SAP HANA 演化史 17
- 2.3 掀起 SAP HANA 的盖头 20
- 2.4 SAP HANA 的创新 24
 - 2.4.1 SAP HANA 的“看家本领” 25
 - 2.4.2 行存储、列存储以及历史表 28
 - 2.4.3 并发控制和一致性 33
 - 2.4.4 数据更新 37

- 2.4.5 聚合或 SQL 42
- 2.4.6 压缩比例 44
- 2.4.7 数据分区 47
- 2.4.8 最小化传输数据 50
- 2.4.9 并行处理 50

第 3 章 SAP HANA 架构介绍 52

- 3.1 SAP HANA 及相关外围系统 52
 - 3.1.1 SAP HANA 系统 54
 - 3.1.2 Index Server 56
 - 3.1.3 XS Server 58
 - 3.1.4 SAP HANA 的开发接口 59
 - 3.1.5 分布式 SAP HANA 系统 61
- 3.2 SAP HANA 技术路线图 63
- 3.3 一体化设备的竞争 65
 - 3.3.1 不同的创新 65
 - 3.3.2 内存计算与缓存数据 67
 - 3.3.3 混搭和简约 68
 - 3.3.4 SAP HANA 与 Teradata 及 Oracle 系列产品的对比 69
- 3.4 实时计算的应用价值 70

第 4 章 SAP HANA 应用场景 72

- 4.1 数据集市和实时报表 73
- 4.2 企业级数据仓库 75

4.3	SAP 应用加速	81	6.2.2	使用 IMPORT 命令从 服务器端加载数据	126
4.4	创新应用平台	82	6.3	使用 SLT	127
4.5	高性能数据库	85	6.3.1	基本原理	127
第 5 章 SAP HANA 基础知识		86	6.3.2	安装和配置	129
5.1	实施 SAP HANA 前的准备 工作	86	6.3.3	SLT 场景演示	130
5.1.1	HANA 项目的一些共同 关注点	87	6.3.4	SLT 基本管理	140
5.1.2	技能要求和工作范围	88	6.4	使用 DS	149
5.2	软件安装	89	6.4.1	演示环境介绍	150
5.2.1	SAP HANA 硬件准备	90	6.4.2	方法 1: 全量加载	150
5.2.2	SAP HANA 软件准备	91	6.4.3	方法 2: 增量加载—— 表比较	163
5.2.3	安装 SAP HANA 数据库	92	6.4.4	方法 3: 增量加载—— 基于时间戳	170
5.2.4	安装 SAP HANA 工作台	98	6.4.5	方法 4: 增量加载—— 数据库 CDC	184
5.2.5	安装 SAP HANA 客户端	100	6.5	数据加载方式小结	194
5.3	SAP HANA 工作台的使用	102	第 7 章 SAP HANA 建模入门		
5.3.1	初次使用 SAP HANA 工作台	103	7.1	建模准备	196
5.3.2	添加 SAP HANA 系统	105	7.1.1	分解 TPC-H	196
5.3.3	建模工具	108	7.1.2	建模用户授权	197
5.3.4	管理工具	111	7.1.3	新建 Package	200
5.3.5	快速创建数据库表	113	7.2	属性视图	201
5.3.6	使用 SQL 编辑器	114	7.2.1	属性视图——客户 主数据	202
5.4	本书使用的数据库表	115	7.2.2	属性视图——供应商 部件	216
第 6 章 加载数据到 SAP HANA		118	7.3	分析视图	218
6.1	关于 Schema	119	7.3.1	步骤 1: 创建分析视图	218
6.2	使用 IMPORT 命令	123	7.3.2	步骤 2: 添加事实表	219
6.2.1	使用 IMPORT 向导工具 从本地加载数据	123	7.3.3	步骤 3: 添加属性视图	221

7.3.4	步骤 4: 为视图添加新列	223	10.1.2	连接到 SAP HANA	262
7.3.5	步骤 5: 指定分析指标	224	10.1.3	探索和分析数据	264
第 8 章	SAP HANA 建模进阶	226	10.2	SAP BusinessObjects Explorer	265
8.1	计算视图	226	10.2.1	步骤 1: 配置 Explorer 和 HANA 之间的连接	265
8.1.1	基本工作界面	226	10.2.2	步骤 2: 创建信息空间	272
8.1.2	计算操作	229	10.2.3	步骤 3: 开始分析数据	275
8.2	示例 1: 计算视图——可视化方式	233	10.3	SAP Web Intelligence	276
8.3	SQLScript	237	10.3.1	在 HANA 信息模型上构建 Universe	277
8.3.1	表类型	238	10.3.2	Web Intelligence 数据展现	284
8.3.2	存储过程	239	10.3.3	扩展话题: 在传统 BI 语义层还是 HANA 数据库中建模	288
8.3.3	计算引擎函数	240	10.4	Excel	288
8.4	示例 2: 计算视图——脚本方式	241	10.4.1	HANA 建模中的 Hierarchy	289
第 9 章	HANA 内容生命周期管理	245	10.4.2	使用 Excel 数据透视表分析 HANA 数据	292
9.1	创建 DU	246	10.5	Tableau	296
9.2	创建 Package	251	10.5.1	使用 Tableau Desktop 连接 SAP HANA	296
9.3	导入和导出功能简介	253	10.5.2	Tableau Desktop 数据分析	300
9.3.1	演示 1: 导出 DU	255	第 11 章	SAP HANA 应用开发	303
9.3.2	演示 2: Developer Mode 下的导出	256	11.1	ABAP 和 SAP HANA 开发	304
9.4	生成技术文档	258	11.1.1	SAP HANA 对 ABAP 开发的影响	304
第 10 章	SAP HANA 与商务智能的结合	259	11.1.2	ABAP 访问 HANA 的	
10.1	SAP Visual Intelligence	261			
10.1.1	创建 VI 文档	262			

准备工作	307	介绍	354
11.1.3 配置连接	308	12.3.1 单节点 HANA 服务器	354
11.1.4 三种 ABAP 访问 SAP HANA 的方式	312	12.3.2 分布式 HANA 服务器 配置	354
11.2 SAP HANA 加速器	323	12.3.3 HANA 集群的灾备恢复 方案	355
11.2.1 Basis Kernel 的支持	323	12.4 演示: HANA 分布式架构	358
11.2.2 加速的原理	324	12.4.1 步骤 1: 在 HANA 服务器和存储服务器 之间配置 NFS	359
11.2.3 准备工作和配置步骤	325	12.4.2 步骤 2: 安装 HANA 主服务器	363
11.2.4 为 ABAP 程序创建 XML 配置文件	326	12.4.3 步骤 3: 安装 HANA 工作服务器	366
11.2.5 将 ABAP 程序配置为 加速模式	328	12.4.4 步骤 4: 安装 HANA 备用服务器	367
11.2.6 SAP 标准加速器方案	330	12.5 HANA 分布式架构的文件 结构	370
11.2.7 加速器方案的适用性	333	12.6 演示: 模拟服务器停机	372
11.3 R 和 SAP HANA	334	12.6.1 HANA 系统的最初 状态	372
11.3.1 配置一个在线软件 资源库	335	12.6.2 测试 1: 工作服务器 停机	373
11.3.2 安装 R 语言所需的 编译器	339	12.6.3 测试 2: 主服务器停机	375
11.3.3 安装 R 语言	341	第 13 章 SAP HANA 系统管理	377
11.3.4 安装 R 服务器	342	13.1 启动和停止	378
11.3.5 配置 SAP HANA Server 和 R 的集成	343	13.2 备份、恢复和升级	379
11.3.6 演示: SAP HANA SQLScript 调用 R	346	13.2.1 备份	381
第 12 章 SAP HANA 分布式 架构实战	351	13.2.2 恢复	387
12.1 HANA 分布式架构介绍	351	13.2.3 升级	390
12.2 HANA 分布式架构的硬件 方案介绍	352	13.3 内存使用管理	391
12.3 惠普公司的 HANA 服务器			

13.3.1	内存使用监控	392	13.5.3	备份相关配置	426
13.3.2	SAP HANA 内存限制 配置	394	13.5.4	安全策略配置	427
13.3.3	内存操作	395	13.6	系统监控	428
13.4	表分区管理	397	13.6.1	SAP HANA 工作台	428
13.4.1	如何查看表分区	398	13.6.2	DBACockpit 工具	436
13.4.2	单层分区的创建	400	13.7	审计	438
13.4.3	多层分区的创建	402	13.7.1	激活审计对象	439
13.4.4	多 HANA 节点下的 分区	404	13.7.2	创建审计策略	439
13.4.5	分区后的基本操作	406	13.8	安全管理	441
13.4.6	演示——分区前后的 性能差异	411	13.8.1	用户管理	442
13.5	系统参数配置	413	13.8.2	认证管理	445
13.5.1	Delta Merge	414	13.8.3	权限管理	447
13.5.2	数据加载	422	附录 A	SAP HANA 应用商店	452
			附录 B	学习环境	454
			附录 C	SAP HANA 学习资源	456

第 1 章

话说内存计算

内存计算 (In-Memory Computing) 对于很多读者来说并不是一个新词汇, 简而言之, 就是使用 CPU 去计算和管理保存在内存中的数据。在内存计算技术的发展过程中, 最广泛的应用就是在数据库技术方面。内存数据库是在超越传统磁盘数据库技术的基础之上, 利用内存计算技术而发展起来的一种应用。本书所谈及的内存计算技术主要是和内存数据库相关的, 接下来我们会先了解一些内存计算技术发展的大背景、企业应用架构的典型瓶颈, 以及在业务中难以实现数据的实时计算等内容。

信息技术在不断创新, 而且每一次的技术突破都伴随着大量新应用的涌现, 无论是硬件技术发展推动软件更新换代, 还是软件设计架构促使硬件技术革新, 总之这两者之间总是互为动力, 而且是信息化技术得以在过去几十年迅猛发展的最主要原因。例如, x86 架构的处理器推动个人计算机的普及, 使得如今桌面应用软件如此丰富, 又如, 高性能、低功耗的芯片让 iOS、Android 等可以在手持设备上顺畅运行。

同样, 从传统磁盘技术的数据库系统进化到缓存型内存数据库, 再到全内存计算数据库技术, 也经历了漫长的过程。在过去, 虽然内存计算产品已经有 TimesTen、sqlite、Altibase、eXtremeDB、fastDB 等, 当然也不乏很多开源的内存计算软件, 但是, 由于一些软件厂商受限于其已有的市场、已有的客户和已有的应用, 采取的并非是积极主动和大力创新的态度, 因此内存计算技术在过去的这些年并未得到迅速发展和广泛推广, 当然, 这与很多厂商的市场战略也不无关系。现在, SAP HANA 的推出正在让整个市场发生着巨大的改变, 那就是所有的软件厂商都在积极追随内存计算发展的脚步。

SAP HANA 一开始就从软件架构的创新上摆脱过去“逐步进化型”创新模式, 将多种技术简洁而有效地组合起来, 融合应用, 彻底摆脱单一架构技术所存在的技术短板, 从而将内存计算技术推到了一个全新的高度, 而且大大超越过去的内存计算技术, 成为最先进内存计算技术的领导者。表 1-1 是本章会涉及的一些技术名词, 供读者快速参考。

通过几十年的发展, SAP 公司已经从德国沃尔多夫小镇的乡镇型企业发展到德国市值第一的公司, 从 R2 到 R3, 在过去的这些年中, SAP HANA 是 ERP 产品以外, 另一个让 IT 业

界和企业客户如此激动且兴奋的 SAP 产品。从 2010 年全球发布 SAP HANA 之后，SAP 没有任何一款产品得到客户如此之多的喜爱和关注，就连当年推出 NetWeaver 全线产品之时也没有达到如此之高的风靡程度。截止 2013 年 6 月，在全球范围内已经有超过 1500 家客户选择了 SAP HANA，而且用户增长速度越来越快。从正式推出后的这两年不到的时间来看，这样的成绩的确说明 SAP HANA 是一款企业用户非常喜欢的产品。

表 1-1 术语表

术语或缩写	解 释
OLTP	Online Transaction Processing, 联机事务处理系统的简称
OLAP	Online Analytical Processing, 联机分析处理系统的简称
SAP HANA	SAP 公司推出的基于新一代内存计算技术的高性能实时数据平台
BI	商务智能软件, 以 SAP BI、Oracle Hyperion、IBM Cognos 为典型代表, 通常与数据挖掘、支持企业管理的业务运营和决策联系在一起
RDBMS	关系型数据库管理系统, 以 MS SQL Server、SAP Sybase ASE、IBM DB2、Oracle 和 MySQL 为典型代表
列式数据库	列式数据库是相对传统的行存储数据库而言的, 是以列存储相关技术为设计架构进行数据存储的关系型数据库, 主要适合批量数据处理和即时查询
内存计算	内存计算是指为了消除磁盘 I/O 在应用系统中的数据吞吐瓶颈, 而将应用系统的数据部分或全部放在内存中以供访问, 从而提升系统性能的一种技术
写优化	专门为提升数据修改的性能而设计的。行存储数据库基本上都是这样的架构
读优化	为了提升数据查询和计算的性能而设计的, 一般列式数据库使用较多, 被视为读优化的数据库

1.1 企业应用系统架构和限制因素

毫无疑问, 传统关系型数据库管理系统 (RDBMS) 是为联机事务型应用而设计的, 而且这一经典的架构被使用如此之久。几十年前设计的到今天依然存在的 Oracle、DB2、MS SQL Server 等数据库产品, 虽然在不断更新换代, 但是你会发现数据库在任何一个 IT 发展时代都是特别重要的。过去的数据库产品只处理少量的企业业务信息, 然而今天的企业每时每刻都会产生大量的业务信息数据, 虽然传统数据库在过去这些年一直在不断发展, 但是这些基于传统磁盘技术的数据库在应对目前的海量数据处理时已经力不从心。简而言之, 无法实时地计算海量数据已经成为制约企业信息化的瓶颈, 这些信息处理瓶颈直接限制了企业信息化的持续发展。

从 SAP 发布全球第一款 R 系列的产品到现在 SAP HANA 所领导的高性能实时数据计算技术, R1 (R 代表 RealTime, 即实时的意思)、R2 和 R3 不断在强调实时计算, 这期间已经跨越了 40 年的时间, 无论是 20 世纪 80 年代基于 ABAP 虚拟机的 R2, 还是经典的企业级三层架构应用 (展现层、应用层、持久层, 如图 1-1 所示), 再或者是成为企业移动计算技术的领导者, 拥有业界最先进的内存计算技术, SAP 公司在企业级信息化架构上一一直保持其固有的领导地位和先进性。

SAP 公司在 40 年前就已经认识到底层传统 RDBMS 数据库的数据吞吐将是一个瓶颈,

它将成为一种限制性的因素，所以 SAP 的企业管理软件一开始就是基于三层架构（展现层、应用层、数据库持久层）而实现的。

SAP 将企业的应用逻辑和业务处理都建立在应用层（底层数据库性能有瓶颈，无法支持大量业务操作），业务流程非常灵活且容易配置，并且应用层服务器可以灵活扩展多个节点，从而增强应用服务器的内存和计算能力（例如，可以设置 SAP 系统应用层的透明表在内存中为行级缓存、区域或全表缓存），这样可以减少对于底层数据库的磁盘读取依赖，并且利用应用层缓存来降低瓶颈，提升整个应用系统的性能。除此之外，SAP 系统还对底层数据库做抽象处理，保持其灵活性，让应用层的软件系统可以适用于所有主流的 RDBMS，而非局限于某一种数据库产品。

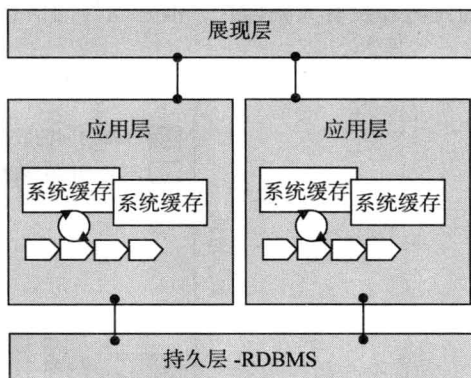


图 1-1 SAP 应用软件三层架构

随着传统数据库软件系统不断升级，例如使用缓存型数据库和硬件加速，也在解决部分问题，但是在软件核心架构上无法有效利用新硬件架构优势。随着企业业务的快速增长，以及结构化和非结构化数据的海量增加，大数据处理和实时计算已日益迫切，传统数据库已经无法满足企业对 IT 技术的要求。总而言之，SAP HANA 更多地摒弃了传统 RDBMS 在技术上的一些束缚和限制，具备了传统 RDBMS 所不具备的核心功能，即无限扩展的并行化、内存计算、OLTP+OLAP 的融合等。

1.2 难以实现的实时计算

在过去，企业的业务系统往往也承载了大量业务分析需求，服务器资源在一定程度上被大量数据查询占用了，这对系统性能造成严重的负面影响，例如影响用户的使用、响应变慢、后台程序占用大量服务器资源、数据库的吞吐瓶颈频发等。伴随着系统的多年运行，系统中也不可避免地堆积着大量的历史数据。对于企业核心业务而言，两三年以上的业务数据属于“冷数据”，对这些数据的访问频率并不是很高，但是这些历史数据往往是重要的企业资产，为企业商务智能的实现提供重要的数据支撑，因此企业无法仅仅简单地数据进行归档。大量历史数据的存在也是影响核心业务系统性能的重要因素之一。

为了降低在企业应用系统中的处理大数据的影响，我们将一部分业务需求在 BI 应用系统中实现，其他的业务应用则继续在当前应用系统中供终端用户继续使用。一个很重要的问题就是，鉴于基于传统 RDBMS 的应用系统的性能瓶颈（磁盘 I/O 是最主要的瓶颈），我们将一部分数据拿出来，转移到另外一个还是基于传统 RDBMS 的 BI 应用系统中，这其实是有一点自相矛盾的，因为 BI 应用系统底层的数据库随着计算数据量的增大，也会很快陷入性能处理瓶颈，不过很多的应用就是这么做的，现实情况也的确如此。

无论是 OLTP 还是 OLAP，企业应用系统的数据量随着业务发展会增长得越来越快，但

是从图 1-2 中我们可以看出，企业日常访问的信息量随着数据量的增大而增大，占总数据量的比例却没有大幅增长，但是同样的信息会割裂开来，有很多个副本和临时计算结果。

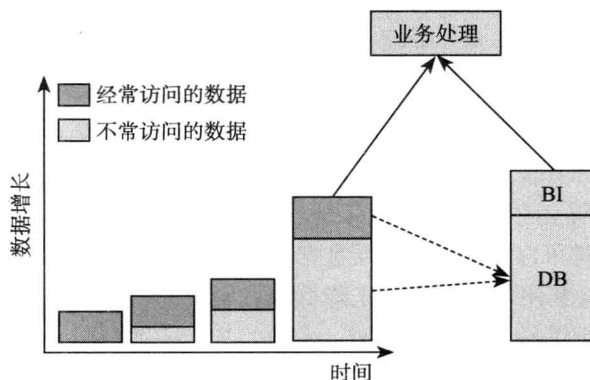


图 1-2 应用系统被“合理”割裂

1.2.1 信息、延迟和价值

例如，在应用系统中，一般对当年和过去一年的销售数据分析较多，而对 3 年或 5 年前的数据分析较少。即使是这样比较少的数据访问，基于传统 RDBMS 构建的应用系统依然无法高效地处理查询请求，在海量数据中对极少的一些数据做处理依然会有很长的延时。解决这一问题，除了为了应对业务部门的需求而不断维护和创建数据库索引，进行数据库调优，并增加新的服务器节点之外，对于加大缓存，依然没有一个有效的手段。

对于有海量数据分析需求的企业用户来说，需要基于海量数据分析做出商业决策。例如，零售企业需要实时知道自己的可用库存水平，对 POS 流水数据进行分析，实时制定促销政策等；金融企业需要实时计算海量数据，了解不同的投资组合所带来的预期收益率，根据这些实时分析结果制定决策，进而在金融市场获得收益。如果无法在有效的时间内获得信息，信息延迟所带来的影响将导致数据无法转变成有用的信息。

对于信息的时效性及其价值，从图 1-3 和图 1-4 中就能看到明显的差异，从信息产生到最后采取行动的时间越长，这条信息带来的价值就越少。所有的业务决策都是要经过分析的，而非通过直觉来做出决定。和传统的信

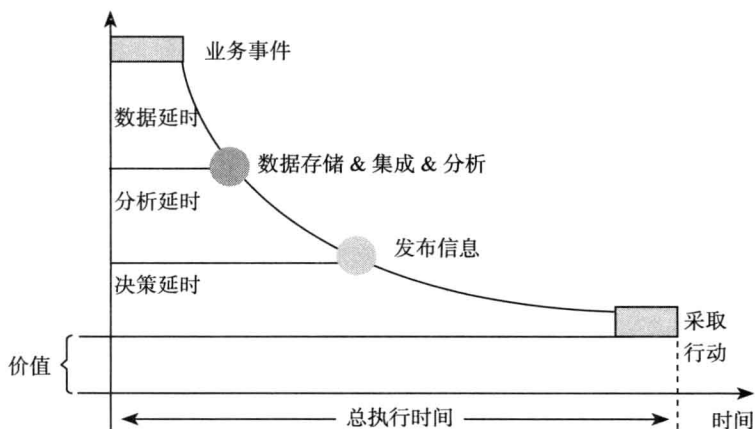


图 1-3 Richard D. Hackathorn——信息延迟导致信息的价值低下

息系统相比，秒级的应用能够大大提升我们对于信息中包含的业务价值的掌控和专注力，而不是让有价值的信息到最后采取行动的时候已经变得毫无意义，这才是速度提升的意义所在。

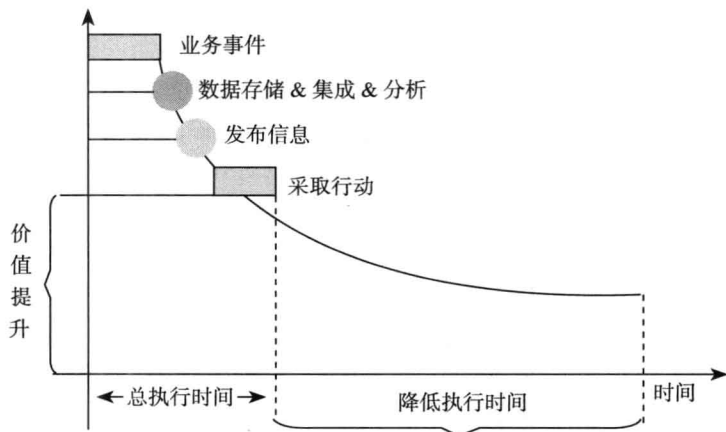


图 1-4 Richard D. Hackathorn——信息及时处理带来更多价值

图 1-5 是 IDC 在 2011 年对于全球数字信息增长做出的预测分析，在数据持续翻倍增长的前提下，对于数据进行有效挖掘和管理，继而从中获益将会变得越来越困难，而且系统对于数据请求的相应速度的性能问题会变得越来越突出，同时，数据的增长可达到每 18 个月翻一番的程度。在这样的现实情况下，对于希望从海量数据中获取价值，将数据变成信息的过程就会变得异常困难。在现有的信息系统架构下，的确很难做到“从数据矿山实时提炼出金子”。

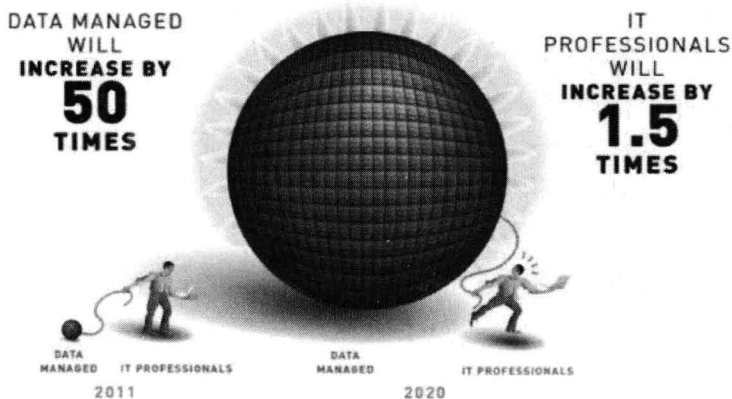


图 1-5 IDC 2011 数据增长和管理需求