

SAP HANA

SAP HANA® 卓越性能

极速响应、轻松扩展 打造实时商务智能



The Best-Run Businesses Run SAP™



目录

4 简介

5 测试环境

数据库模式

测试数据

系统配置

设置

查询

9 测试结果

基线测试

吞吐量

即席历史查询

结果综述

真实体验

12 总结

13 附录

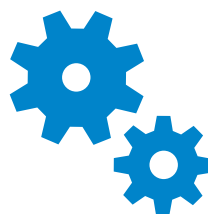
简介

借助SAP HANA®应用软件，企业可以实时分析海量数据，进而优化业务运营。该软件无需依赖专用的扩展组件，可运行于经济实惠的通用硬件上。无需进行任何调优，SAP HANA直接带来高性能感受

100 TB性能测试旨在证明SAP HANA的高效性和可扩展性；企业运用大型数据库中的数据进行运营分析时，SAP HANA能够帮助企业轻松提升实时商务智能的分析性能。这100 TB¹数据集的数据格式与从SAP ERP应用系统里抽取的，用于SAP NetWeaver® Business Warehouse（SAP NetWeaver BW）组件分析的数据格式完全相同。²

本文将介绍该测试环境，并提供和分析测试结果。

借助SAP HANA®应用软件，企业可以实时分析海量数据，进而优化业务运营。



测试环境

性能测试环境旨在搭建一个销售与分销（SD）智能商务查询环境，为大量的结构化查询语言（SQL）和业务用户提供支持。

数据库模式

图1中的星形模式设计显示SD测试数据环境和每个表的基数。此设计中不使用手动调优结构；没有索引、物化视图、汇总表或其他为实现更快查询性能添加的冗余结构。

测试数据

测试数据包括一个大的数据表和几个较小的维度表，如图1中所示。仅在数据表中就有1千亿条记录，代表5年的SD数据。数据使用“Customer_ID”在16个节点进行平均哈希分区，然后在每个节点，将数据按月份分区。这样，每个节点会有60个分区，每个分区约有1.04亿条记录。（另请参见图2和“加载”部分。）

系统配置

测试系统配置（图2）是一个配有8 TB RAM的IBM X5服务器的16节点群集。每台服务器有：

- 4个CPU，每个CPU有10个内核，每个内核有2个超线程，合计：
 - 40个内核
 - 80个超线程
- 512GB的RAM
- 3.3TB的磁盘存储

设置

这些性能测试没有使用结果的预缓存结构或任何手动调优结构，因此，这能够验证SAP HANA“加载即能查询”的能力。一种完全没有调优结构（内部或外部）的设计对建立可持续发展的实时

商务智能环境非常重要；这种设计不仅能够加快实施过程，还为即席查询提供持续的灵活性，同时省去了调优结构的维护成本。

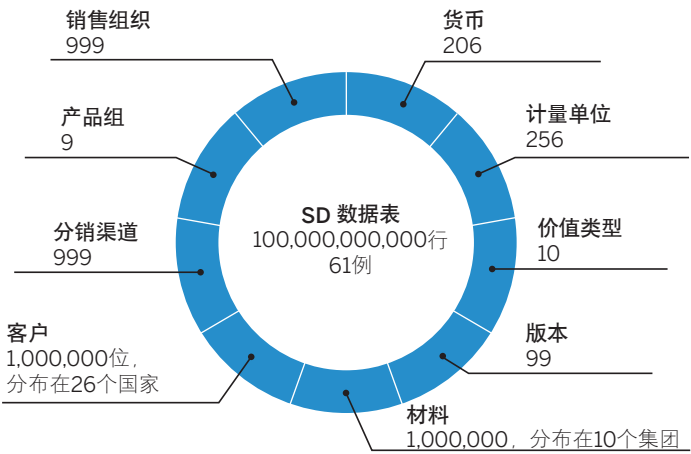
加载

使用SAP HANA中的“IMPORT”（“导入”）命令，可以并行完成加载，该命令是一个单独的SQL语句，用来指定要加载的文件。加载过程将自动在所有节点并行处理，并使用每个数据表定义的先分布后分区的模式³（在这种情况下，指哈希分布和按月分区）。测量的加载速度为每分钟1600万条记录，或每个节点每分钟100万条记录。这种加载性能足以在短短六分钟内加载1亿条记录（代表一个工作日的活动）。

压缩

数据压缩在数据加载过程中进行。SAP HANA的压缩率⁴比SD模式的压缩率高出20倍；将100 TB SD数据集压缩成一个3.78 TB的SAP HANA数据库，但是群集中的每个节点只被占用了236GB的RAM（见图2）。在100 TB的数据集中，大数据表占用了85TB，大数据表经压缩后所占用的空间不到每个节点可用RAM的一半。

图1：SD模式



查询

查询套件共有20个不同的SQL查询，包括11个基础查询外加一系列时间间隔（月、季度等）。所选查询代表在原始形式数据（即没有索引、调优或采用其他非规范化方法以避免连接，这些都是传统数据库中的惯例）上运行的混合工作负载环境。它们涵盖一系列商务智能活动，从部门到企业级别，包括：

- 常规报告
- 迭代查询（向下钻取）
- 等级
- 年同比分析

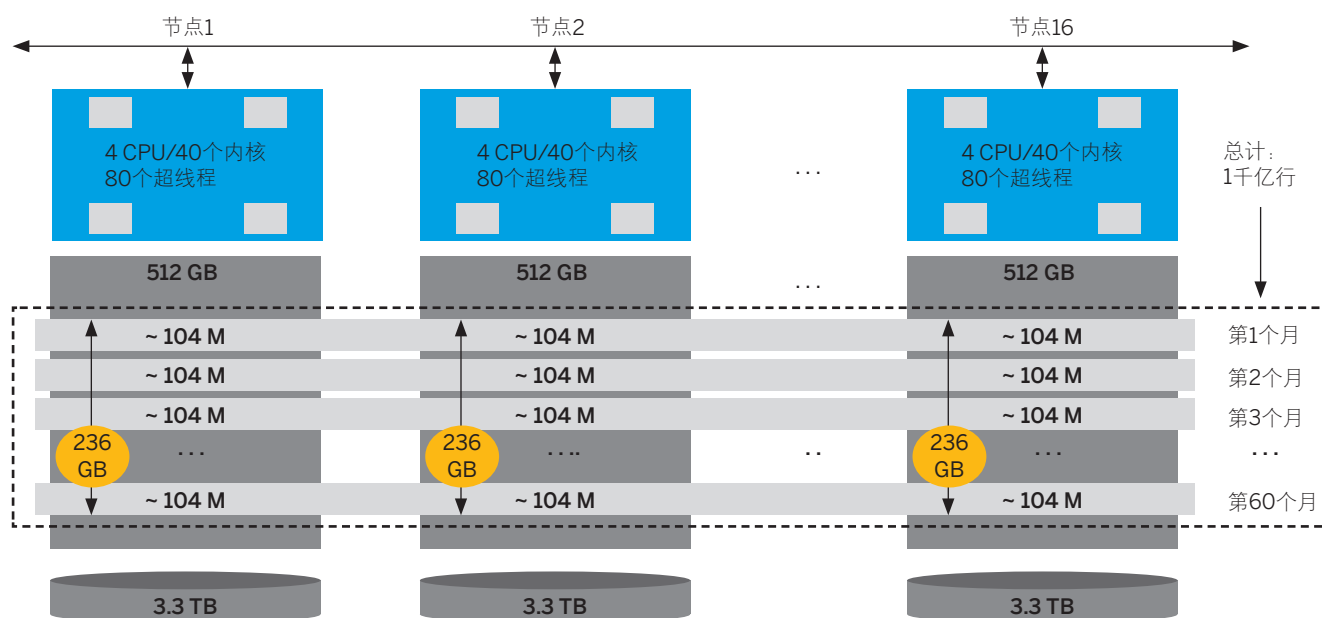
产生的查询范围从中度复杂到非常复杂，包括SQL构建，例如：

- 多连接 (Multiple Join)
- 在列表中 (in list)
- 事实表对事实表连接 (fact to fact joins)
- 子查询
- 相关子查询 (CSQ)
- 全部合并 (Union all)

查询分组为商务智能用途的三个普通类别:

- 报告——计算一段时间内各种材料或客户或两者的业务绩效衡量
- 向下钻取——迭代用户启动的查询，用以收集给定的单个或一组材料或客户（或二者兼有）的详细信息
- 分析——跨材料或客户（或二者兼有）进行定期深入历史分析

图2: 测试配置





下表“查询说明”记录每个查询的业务说明、SQL结构、范围和限定符以及时间周期变化。如果用多个时间期间介绍一个查询，每个时间期间都作为不同的查询运行；例如，查询R1作为四个不同的查询运行，涵盖了月度、季度、半年度和年度不同的时间周期，以反映其对应数据量发生变化后的性能表现（年度数据是月度数据量的12倍）。粗体项目表示查询复杂性要素。

查询说明

BI 查询	业务说明	SQL结构	范围和限定符	时间周期
R1	月度销售与分销报告，用于： - 材料给定范围 - 客户给定范围	星形模式查询，带有到维度的多个连接，具有group by和order by	- 材料范围 - 客户范围	四种变化： 1个月 3个月 6个月
R2	用于给定客户列表的月度销售与分销报告	星形模式查询，带有到维度的多个连接，具有group by和order by	使用“in list”定义的客户范围	1个月
R3	月度销售与分销报告，用于： - 材料给定范围 - 给定客户国家	星形模式查询，带有到维度的多个连接，具有group by和order by	- 材料范围 - 客户国家	1个月
R4	前100位客户报告用于： - 一个或几个材料组 - 一个或几个客户国家	星形模式查询，带有到维度的多个连接，具有group by和order by	- 一个或几个材料组 - 一个或几个客户国家	用于两种变化的一个月数据： 1个国家和1个产品组 3个国家和2个产品组
R5	前100位客户报告用于： - 给定材料组 - 给定客户国家	星形模式查询，带有到维度的多个连接，具有group by和order by	- 材料组 - 客户国家	两种变化： 3个月 6个月

大部分查询达到亚秒响应时间；即使最复杂的查询（涉及整整五年的数据）也可在不到四秒时间内完成。



BI 查询	业务说明	SQL结构	范围和限定符	时间周期
D1	销售与分销报告用于： • 单独客户 • 给定材料组	星形模式查询， 带有到维度的 多个连接，具有 group by和order by	• 材料范围 • 客户范围	四种变化： • 1个月 • 3个月 • 6个月
D2	销售与分销报告用于： • 单独客户 • 单独材料	星形模式查询， 带有到维度的 多个连接，具有 group by和order by	使用 “in list” 定义的客户范围	1个月
D3	销售与分销报告用于： • 单独材料 • 给定客户国家	星形模式查询， 带有到维度的 多个连接，具有 group by和order by	• 材料范围 • 客户国家	1个月
D4	销售与分销报告用于： • 单独客户 • 材料给定范围	星形模式查询， 带有到维度的 多个连接，具有 group by和order by	• 一个或几个材料组 • 一个或几个客户国家	用于两种变化的一个月数据： • 1个国家和1个产品组 • 3个国家和2个产品组
A1	前100位客户年同比 (YOY) 分析 分析报告： • 为当前时间期间确定的前 100位客户 • 在给定历史时期为之前确 定的前100位客户计算的商 务衡量结果 • 计算的最终结果集，结合 用于前100位客户的当前和 历史商务衡量结果	星形模式查询，带有相关 子查询的fact-to-fact连接， 包括到纬度的多连接，然后 group by,再order by	• 材料范围 • 客户国家	三种变化： • 1个月带YOY对比 • 3个月带YOY对比 • 6个月带YOY对比
A2	用于前100位客户的YOY趋势 报告： • 为去年（或多年）相同时间 期间前100位客户 • 为去年（或多年）相同时间 期间计算的前100位客户 • 计算的最终结果集，结合 数年用于前100位客户的商 务衡量结果	几个星形模式子查询，每个有 多个连接，具有group by和 order by，通过union all运算进 行合并	• 材料组 • 客户国家	三种变化： • 3个月，超过2年 • 6个月，超过2年 • 3个月，超过5年

测试结果

该测试测量了查询的响应时间和每小时的查询吞吐量。查询首先在单一流（single stream）中运行，以测量基准查询时间，然后在10、20和25个流中运行，以测量吞吐量（按每小时查询数计算）和在不同工作负载环境中的查询响应时间。⁵ 多个流（multiple streams）测试中各个流的查询提交顺序随机化⁶，并在每次查询之间插入10毫秒的“思考”时间，模拟多用户即席BI查询环境。“附录”中列出单独的运行时间。

基线测试

正如您在图3中的基准结果中看到的，大部分查询达到亚秒响应时间；即使最复杂的查询（涉及整整五年的数据）也可在不到四秒时间内完成。

报告和选项钻取查询（267毫秒到1.041秒）证明SAP HANA在汇总数据方面的卓越能力。例如，其中最长的运行R1-4仅仅用了一秒

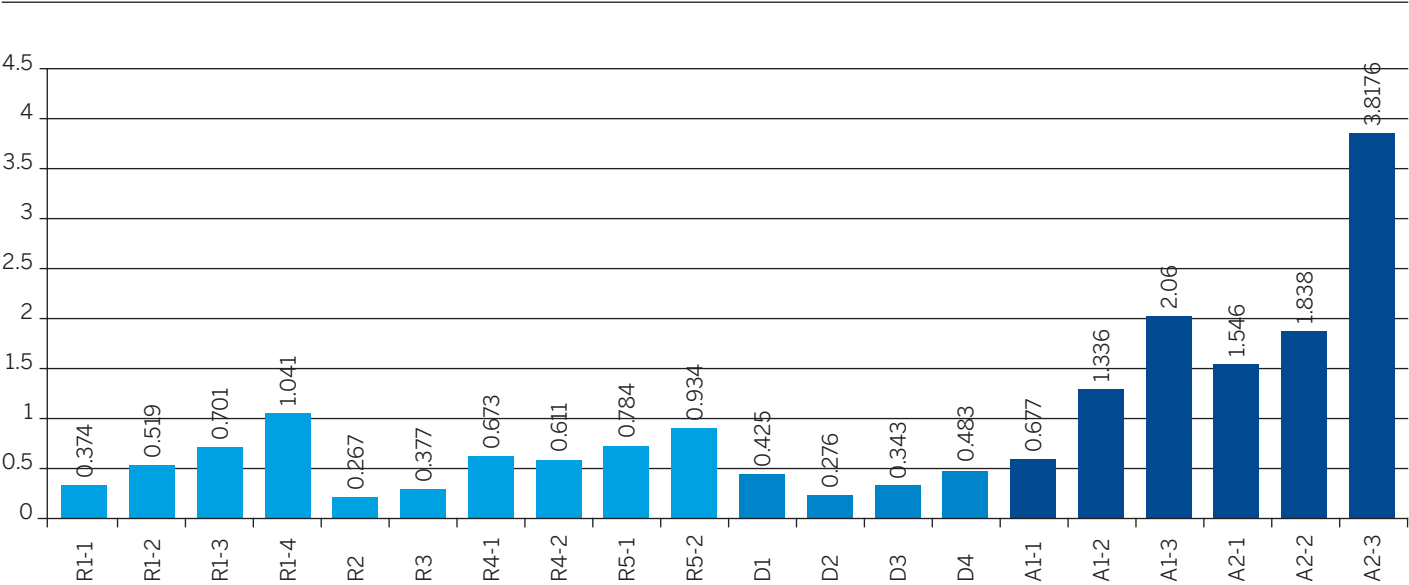
多（1.041），相对于其月度运行R1-1，数据量增加了12倍，处理时间却只增加了2.8倍。

向下钻取查询（276到483毫秒）证明SAP HANA对即席连接的强大支持，因此为用户提供了“切片和切块”的无限能力，却无需让技术人员提供支持它的索引（传统数据库需要这样做）。

分析查询（677毫秒到3.8秒）跨越滑动时间窗口（年）高效执行了复杂连接（实际数据连接、子查询、CSQ、全部合并）和分析。一到六个月日期范围里（A1-1到A2-2）的查询能在两秒或更短时间内运行。查询A2-3分析全部五年的日期范围，运行时间不到四秒。全部分析查询时间完全在时间范围内，这样帮助实现迭代的思维性分析。

总之，基准测试表明SAP HANA针对给定的查询，随着数据量的增加实现了有效的扩展（超过线性）。

图3：基准性能（单位：秒）





吞吐量

吞吐量测试在表“吞吐量测试”中汇总，并表明在面对不断增加的混合BI工作负载时，SAP HANA可有效扩展。

吞吐量测试	
测试案例	业务说明
1个流	6,282
10个流	36,600
20个流	48,770
25个流	52,212

对于25个流，平均查询响应时间不到三秒，仅高于基准2.9倍（参见“附录”），该基准是SAP HANA管理并发和混合工作量的卓越内部效率和能力的一个指标。

用估算的每位用户每小时查询数除以每小时总查询数，可以得出BI用户并发的粗略估计值。例如，每小时25个流的52,212次查询除以20（每三分钟一次查询的平均用户率），得出报告、向下钻取和分析查询类型混合的2,610个并发BI用户的合理估计值。

即席历史查询

性能测试的主要目的是证明SAP HANA在侧重分析业务运营的常规实时BI查询环境中的性能。但是，分析所有存储数据量的即席历史查询偶尔也可能需要运行（并且其效果不会太差）。除了查询A2-3（包括在流测试中），认为查询R1-1、R4-2和R5-1与五年时间窗口潜在相关并重新提交，这次用于整个五年日期范围。表“即席历史查询”显示SAP HANA超过线性的可扩展性，以三倍或超过三倍的响应速度运行五倍的数据。

总体而言，这些查询全部在几秒钟内运行完成，进一步凸显了SAP HANA在海量数据方面支持实时BI的能力，无需调优，也不用大量硬件或专用硬件组件。

结果汇总

这些性能测试证明，随着工作负载在容量（最高100 TB原始数据）、复杂性（带复杂连接建构的查询以及不到两秒钟得出的大量直接结果）和并发（25个流吞吐量代表约2,600位活动的用户）方面的增加，SAP HANA同样实现了实时商务智能查询的性能。

即席历史查询

时间段	R1-1（单位：秒）	R1-1（单位：秒）	R5-1（单位：秒）
1年	1.1	1.4	1.1
2年	1.8	2.4	1.8
3年	2.6	3.4	2.4
4年	3.3	4.4	3.0
5年	4.1	5.4	3.6



SAP HANA可提供超过20倍的压缩率

SAP HANA性能基于全面的深思熟虑的设计元素，包括诸如：

- 一种基于内存的设计
- 智能内部数据结构（例如，内在的列存储、高级压缩和强大的分区）
- 一个智能、基于成本的优化器，可将这些智能数据结构融入有效的查询计划中
- 智能执行查询计划的查询执行器有效的利用各类内部机制（例如，高级算法、CPU中的多级缓存优化和超线程）。

SAP HANA的设计还有很多，这里无需一一说明。有关SAP HANA技术特性的详细信息，请参阅SAP HANA技术概述和用于下一代业务应用程序和实时分析的SAP HANA。

真实体验

这些测试在一个模拟环境中运行以分离SAP HANA性能⁷；不过，您可期望自己的BI查询性能在SAP HANA上显著优于在您现有的传统数据库管理系统上——或许好上数千倍。最重要的是，SAP客户推荐确认SAP HANA性能。下面是一些具体案例。

“我们看到系统速度大幅提升，以及对客户和产品最详细内容的分析能力的提高。”——高露洁

“[取]代我们企业范围的Oracle数据集市，在处理我们最复杂的货运成本计算方面实现超过20,000倍的速度提升。我们之前在Oracle上运行的独立移动应用程序目前在SAP HANA上运行。我们的2,000位本地销售代表现在能与实时数据交互，并有能力进行快速促销决策以改进销售。”——农夫山泉

“我们的内部技术对比证明，SAP HANA的性能超过传统的基于磁盘系统408,000倍。”——三井情报株式会社

“凭借SAP HANA，我们有无限的机会大大改善我们的企业数据仓库解决方案，并显著减少数据延迟和提高速度，我们通过SAP HANA上的SAP NetWeaver BW在45秒内可返回查询结果，而在传统的基于磁盘的数据库上，等待20分钟从SAP NetWeaver BW获得的仍然是空的结果。”——上海大众

总结

实时BI查询环境可支持客户随时拟定的新查询。只有基础数据库平台始终保持卓越性能，这种思维般的速度能力才能实现。

这些严谨的性能测试及测试结果证明了，在无需调优，无需大量

硬件或专用的硬件组件的情况下，SAP HANA依然能够实现一流的性能和可扩展性，并为必须支持大量分析工作、处理海量数据和拥有数以千计并发用户的企业实现实时商务智能。

SAP HANA能够实现一流的性能和可扩展性，并为必须支持大量分析工作、处理海量数据和拥有数以千计并发用户的企业实现实时商务智能



脚注

1. 100 TB = 压缩前的原始数据大小
2. 这些测试不涉及SAP NetWeaver BW。
3. 如“CREATE TABLE”命令的定义。
4. 压缩率主要取决于实际被压缩数据的特性。个别结果可能存在差异。
5. 一个单独查询流代表支持潜在数以百计并发用户的应用程序连接。
6. 这样做是防止不同的流在完全相同时间提交的查询，同时帮助确保实际工作量混合（例如，向下钻取比季度报告更频繁地发生）。
7. 这些性能测试旨在分离SAP HANA系统级性能，独立于SAP软件环境中可能使用的各种应用程序。因此，这些测试结果只能用作一般指导原则，由于平台的可用性、应用程序和数据不同，每次实施的结果也存在差异。

附录

查询名称	测试用例*	平均运行时间（秒）
R1-1	1个流	0.374301625
	10个流	0.7104476375
	20个流	1.11287894375
	25个流	1.263543115
R1-2	1个流	0.5193995
	10个流	0.8852610125
	20个流	1.2852265875
	25个流	1.55048987
R1-3	1个流	0.70110375
	10个流	1.098921025
	20个流	1.6092578375
	25个流	1.85689055
R1-4	1个流	1.0411995
	10个流	1.5377812
	20个流	2.24272095
	25个流	2.53941444
R2	1个流	0.2671005
	10个流	0.4740847375
	20个流	0.765801125
	25个流	0.94418651

查询名称	测试用例*	平均运行时间（秒）
R3	1个流	0.3772025
	10个流	0.687762475
	20个流	1.03632401875
	25个流	1.28864399
R4-1	1个流	0.67338575
	10个流	1.15797815
	20个流	1.7928188625
	25个流	2.081943225
R4-2	1个流	0.61149025
	10个流	1.0192868125
	20个流	1.48342026875
	25个流	1.751548475
R5-1	1个流	0.784670625
	10个流	1.3914058
	20个流	2.017145075
	25个流	2.37001971
R5-2	1个流	0.9343955
	10个流	1.6150219
	20个流	2.326104475
	25个流	2.70342093

查询名称	测试用例*	平均运行时间（秒）
D1	1个流	0.4254266875
	10个流	0.67727600625
	20个流	0.9883220375
	25个流	1.1591933925
D2	1个流	0.2762340625
	10个流	0.49984061875
	20个流	0.7925614875
	25个流	0.9211584
D3	1个流	0.3439448125
	10个流	0.6137608375
	20个流	0.969213509375
	25个流	1.1464113225
D4	1个流	0.48321675
	10个流	0.88516171875
	20个流	1.345513390625
	25个流	1.56276527
A1-1	1个流	0.67762625
	10个流	1.0925558
	20个流	1.77672375
	25个流	2.03975081

查询名称	测试用例*	平均运行时间（秒）
A1-2	1个流	1.3362615
	10个流	2.0676603
	20个流	3.0319986125
	25个流	3.62353857
A1-3	1个流	2.0605925
	10个流	3.15405455
	20个流	4.469254975
	25个流	5.01461024
A2-1	1个流	1.5460765
	10个流	2.7397760625
	20个流	3.99316446875
	25流	4.90314213157894
A2-2	1个流	1.838777
	10个流	3.256667375
	20个流	4.5519178125
	25个流	5.32557834210526
A2-3	1个流	3.816785
	10个流	6.371394
	20个流	9.88785475
	25个流	11.9014413333333

*10、20和25个流的运行过程已反映查询之间10毫秒的思考时间。



The Best-Run Businesses Run SAP™

(12/11) 版权所有©2012SAPAG。保留所有权利。

SAP, R/3, SAPNetWeaver, Duet, PartnerEdge, ByDesign, SAPBusinessObjectsExplorer, StreamWork, SAPHANA和本文提及的其它SAP产品和服务及其各自标识均为SAPAG在德国和其它国家的商标或注册商标。

BusinessObjects和BusinessObjects标识、BusinessObjects、CrystalReports、CrystalDecisions、WebIntelligence、Xcelsius和本文提及的其它BusinessObjects产品和服务及其各自标识均为BusinessObjectsSoftwareLtd.的商标或注册商标。BusinessObjects是一家SAP公司。

Sybase和AdaptiveServer、iAnywhere、Sybase365、SQLAnywhere和本文提及的其它Sybase产品和服务及其各自标识均为SybaseInc.的商标或注册商标。Sybase是一家SAP公司。

Crossgate、m@gicEDDY、B2B360®和B2B360®服务是CrossgateAG在德国和其它国家的注册商标。Crossgate是一家SAP公司。

本文提及的所有其它产品和服务名称分别是其各自公司的商标。本文包含的数据仅供参考。各国的产品规格可能不同。

上述资料如有变更，恕不另行通知。上述资料由SAPAG及其关联公司（统称“SAPGroup”）提供，仅供参考。SAPGroup对其不做任何陈述或保证。对于其中的错误或疏漏不承担任何责任。对SAPGroup产品和服务所做的保证仅为这类产品和服务随附的明示保证声明中的保证（如有）。本文中的任何信息均不构成额外保证。