软件产品说明

(*Software Product Description*)

产品名称（中文）	TRS 全文检索网关
产品名称（英文）	TRS Gateway
产品编号	参见相关文件
版本	Version 4.6
发布日期	2011 年 4 月 14 日

一、产品描述

关系数据库（RDB）对于存储在“大对象”中的数据的内容进行检索和分析的效率低下，而在实际应用中往往需要对这些数据进行高效的检索和分析。

TRS 全文检索数据库是一种完备的文本型数据库系统，适合对各种结构化和非结构化的信息进行管理和查询，特别是在海量文本集合上实现了高效的全文检索功能。关系数据库中的大对象字段中的内容，使用 TRS 全文检索数据库可以实现高效检索和分析。

TRS 全文检索网关（TRS Gateway）是由北京拓尔思信息技术股份有限公司为实现关系型数据库的全文检索而推出的软件产品，该产品实现了关系型数据库与 TRS 全文数据库之间的数据自动迁移和同步更新，利用 TRS 全文检索数据库解决了关系数据库大对象的检索和分析效率问题，而上层应用则可以构架在由关系数据库和 TRS 全文数据库共同组成的数据库平台之上。

一直以来，虽然关系数据库厂商也宣称或推出了具有全文检索的产品，但内容检索是和本地语言密切相关的，TRS 领先的全文检索功能融合了中文自然语言处理的最新成果，包括中文按词索引、字索引的 BI-GRAM、基于语义辞典等语言学知识的智能检索，以及中文自动分类和自动摘要等领先技术，是中文全文检索的最佳选择，因此主流数据库厂商都选择 TRS 作为中文全文检索解决方案。

利用 TRS 全文检索网关软件，用户可以经过简单的配置操作，在关系数据库与 TRS 全文检索数据库之间建立映射关系，系统就可自动将关系型数据库中的数据导入 TRS 全文检索数据库，并自动保持同步更新，使用户在享有关系型数据库卓越的数据处理功能的同时，拥有 TRS 全文检索功能。

TRS 全文检索网关软件经过了千万量级文本数据的实际应用检验，具有良好的可靠性和稳定性。

二、产品特点

TRS 全文检索网关系统具有如下特点：

- 支持主流关系数据库：TRS 全文检索网关系统支持 Oracle, DB2, SQL Server, Sybase、MySQL 和人大金仓（KingbaseES）等关系数据库。
- 支持 RDBMS 中格式化文档的全文检索：TRS 全文检索网关内置格式化文档分析和过滤组件，能够自动对关系数据库大对象字段中存储的格式化文档，如 Word、Powerpoint、 Excel、 PDF 等文件进行全文检索。
- 数据同步和一致：TRS 全文检索网关充分利用关系数据库支持事务的特点，可以保证索引和数据的同步，从而保证查询的结果是完全正确的。
- 支持完全更新和增量更新：完全更新是把关系型数据库中数据一次性全部导入到 TRS 数据库中，不重复执行；增量更新是只对发生变化的数据进行数据同步，并以一定的时间周期循环执行。
- 支持多种类型的任务配置：支持从关系型数据库到 TRS 数据库的多种任务配置，包括（1）“RDBMS—>TRS”任务，实现将关系数据库中的数据向 TRS 数据库进行迁移；（2）“RDBMS—>TRS 格式文件”任务，实现将关系数据库中的数据向 TRS 格式文件进行迁移；（3）“TRS 格式文件—>TRS”任务，实现将 TRS 格式文件向 TRS 数据库进行迁移；（4）“TRS 优化”任务，实现对指定 TRS 数据库进行定时优化。
- 支持表和视图：关系数据库中的表和视图均可以作为同步操作的数据源。
- 自动化程度高：用户只需要按步跟随“任务创建向导”的提示就可以创建更新任务。
- 可设置的定时执行任务：对创建好的任务，用户可以设置其自动定时执行。如：用户可以设置一个增量更新任务每隔 5 分钟执行一次，即每隔 5 分钟将关系数据库表中数据的修改向 TRS 数据库中进行一次索引的更新。
- 易用性：界面友好，简单易用。
- 完善服务功能，使后台运行更稳健。
- 产品结构为 C/S 模式。
- 全面支持 32 位和 64 位操作系统。
- 支持 TRS Cluster 二级集群架构及其扩展，并支持对配置相同的多个集群服务器同时发送读写分离命令（IDLE 和 ORDER）。配置相同的集群，上层是通过硬件做到负载均衡。
- 支持 TRS 集群扩展，通过配置，多个任务共用临时表、触发器，达到一定规则的数据入一个 TRS 集群，另一个规则的数据入另一个 TRS 集群。
- 支持注册网关服务进程到系统服务，开放服务名，支持用户自定义，并提供更完善的服务管理工具。
- 支持低版本网关配置文件通过客户端上传，服务程序自动识别配置文件版本并进行转换，为当前版本所用。
- 提供命令行安装介质。

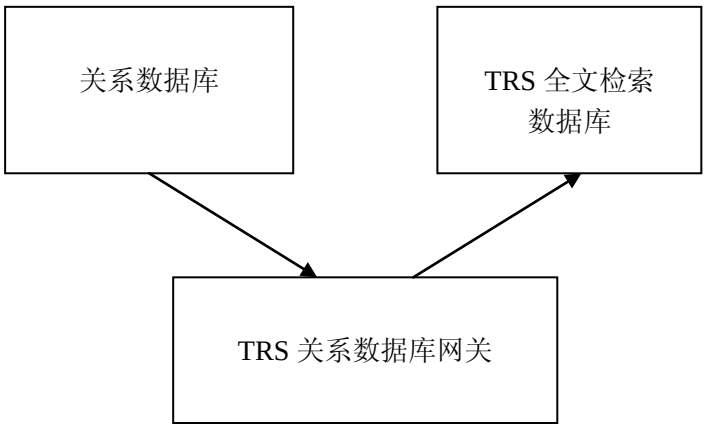
三、系统组成

TRS 全文检索网关分为客户端程序与服务器端程序。客户端程序只能部署在 Windows 环境下，用于配置网关任务。服务器端程序可部署在 Windows 和 Linux 下，用于网关任务的运行。

为了实现数据同步的任务，对于每一种关系数据库，在安装 TRS 全文检索网关系统的机器上，必须安装该种关系数据库的客户端程序（或数据库），以使 TRS 全文检索网关系统能够正常访问该种关系数据库。

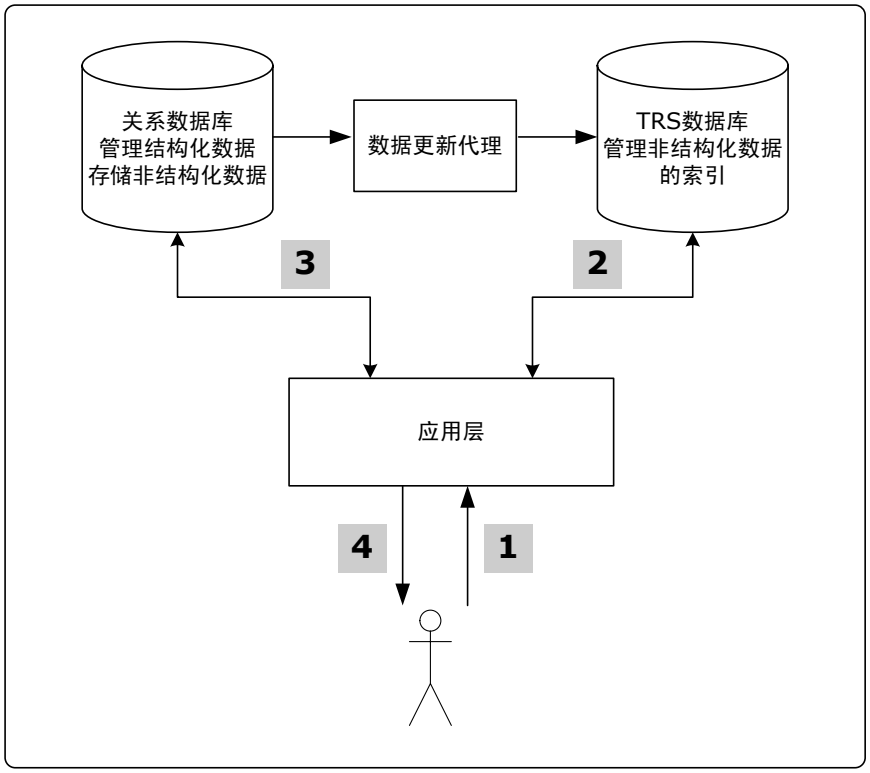
四、体系结构

TRS 全文检索网关系统自身结构比较简单，它相当于关系数据库和 TRS 全文检索数据库的客户端程序，并在两者之间实现数据和索引的同步。



TRS 关系数据库网关示意图

在典型的应用中，TRS 全文检索网关系统的定位在于将关系数据库处理结构化数据的优势和 TRS 处理非结构化数据的优势结合起来，同时在应用层进行无缝集成。其实现原理如下图所示：



典型应用的系统结构示意图

TRS 全文检索网关的作用是数据更新代理，保证关系数据库中的数据发生变化时，数据所对应的全文索引可以及时更新，考虑到数据量和性能的影响，数据更新代理必须能够实现索引的增量更新。在这种系统结构下，实现非结构化数据的检索应用包括如下步骤：

1. 用户提交非结构化数据的检索请求；
2. 应用层将用户检索条件提交给 TRS 数据库进行检索，TRS 数据库返回命中记录的主键信息；
3. 应用层根据 TRS 返回的主键信息在关系数据库中找到相应命中的记录；
4. 将最终检索结果以用户期望的表现形式返回给最终用户。

五、运行环境

TRS 全文检索网关运行环境：

- Microsoft Windows XP/NT/2000/2003/2008/Vista/7
- Linux 2.4、2.6 内核
- 256MB 以上内存
- 500MB 硬盘空间用于安装文件及临时空间，出错日志需要另外的空间，大数据量迁移时需保证系统的临时目录下有足够的空间
- 相应的数据库客户端软件

TRS 全文检索网关所支持的数据库及版本：

- Oracle 8/8i 及以上版本
- DB2 5.2 及以上版本
- Microsoft SQL Server 2000 及以上版本
- Sybase 15 及以上版本
- MySQL 5.0.3 及以上版本
- KBE V6.1 及以上版本
- TRS Server 4.5 及以上版本
- TRS Cluster 2.0.B3302 及以上版本

六、相关产品

TRS 全文数据库服务器	提供各类文本信息和数据的内容管理以及全方位检索功能，特别是提供高效的中英文全文检索功能。
关系型数据库	提供数据源

七、其它注意事项

运行大数据量（一次迁移数据>100,000 条）的任务时，需要关注：

- 关系数据库的表空间是否足够。
- 关系数据库的 LOG 文件空间是否足够：RDBMS 的数据库表空间或 LOG 文件空间不足可能导致本应用程序运行过程中出现长时间无工作进度，在系统中所占 CPU 为 0%的情况。LOG 文件空间不足的情况更为普遍，尤其是 SQL SERVER 数据库更是如此。
- 安装本应用程序的机器上的指定文件目录是否足够：这种情况在 RDBMS—>TRS 且 TRS 目标字段中包含大量 DOCUMENT/BIT 类型字段时尤其需要注意，通常在此种情况下，系统临时目录所在硬盘需要保证有>100M 的空间。