



**STORAGE** Most organisations could benefit from database optimisation, but there is no one approach to unlocking the savings offered by improved data flow

## Optimise your data delivery system

**Arif Mohamed**

[arif.mohamed@rbi.co.uk](mailto:arif.mohamed@rbi.co.uk)

Whether your database is Oracle, SQL Server, DB2, MySQL, or Ingres, you want it to perform as well as it can.

But getting the best performance out of your database system – termed database optimisation – is a multi-disciplined practice. It means getting everything right: the database configuration, the hardware infrastructure and the data itself, which should be organised and indexed in the most efficient way.

Mike Thompson, principal research analyst at Butler Group, said that database optimisation starts with having the correct database design methodology. This means understanding the major purpose and usage of the database.

For some, this will mean using the database to insert and delete records, for others there is going to be a heavy query load, said Thompson.

“Most databases are query intensive, where you are looking for the best route to retrieving the data. Cost optimisation tools can help with this, because they do not just look at the database, but also the underlying infrastructure – the hit on the processor, I/O subsystems and disc spindles,” said Thompson.

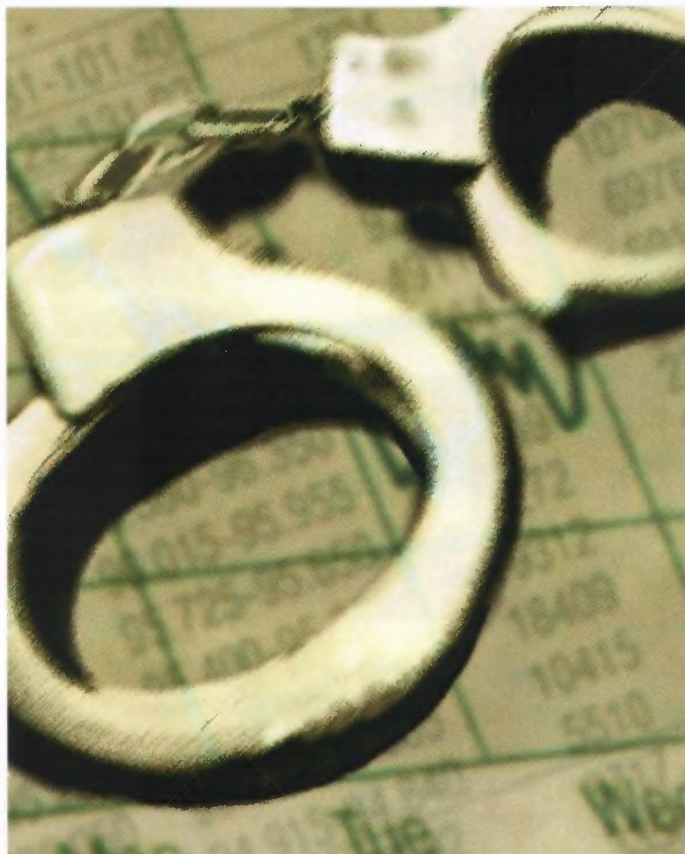
He said that it is also important to get the indexing methodology right, which is where data indexes are queried rather than the data itself, speeding up the query process. Indexing is an art in itself, and there are tools available, such as CopperEye, which can create efficient database indexes, said Thompson.

He added that there are many ways to optimise the database, but users should at least ensure their database is compacted, with both the data and data logs up-to-date, and make sure that they are not holding a huge amount of data that is not needed and will slow down queries.

### Built-in database optimisation

Commercial databases increasingly come with optimisation technologies built in as standard.

For example, the Microsoft SQL Server database has a high degree of self optimisation and tuning, said Matthew Stephen, technical solu-



**Unlock your data: making data easily retrievable can deliver big time and cost savings**

### KEY POINTS

- ▶ Start optimisation by analysing what your database is used for
- ▶ You must also get the indexing methodology right, says Butler Group
- ▶ Some firms are using datawarehouse appliances to speed up data delivery
- ▶ Getting it right can cut the burden on the helpdesk and speed data querying

**Database optimisation means getting everything right: the database configuration, the hardware infrastructure and the data itself, which should be organised and indexed in the most efficient way**

tion professional at Microsoft. “The smaller database market always requires a high degree of optimisation, and a lot of places do not have a lot of hands-on database administrators,” he said.

SQL Server includes data compression technology to speed up database activities, the ability to shrink and grow databases easily, and dynamic memory management. It also comes with optimisation tool suites such as the Database Tuning Advisor – a set of developer tools that Microsoft uses internally.

David Hobbs-Mallyon, SQL Server product manager at Microsoft, said the tools allow users to carry out complex “what if?”

analyses on indexes, potential performance improvement figures for indexes and index views, and they can suggest new database structures. “This is useful particularly for migration proof of concepts where you want new index strategies,” he said.

IBM, Oracle and Sybase databases also have optimisation tools. For example, IBM sells a suite of tools to improve the performance and management of its DB2 and IMS databases. These include “autonomic” features (an IBM term for self-diagnosis and automatic fixing). IBM said it updates the tools as new releases of the databases become available and advises users to check for the latest tool releases.

### Beyond your database supplier

For many organisations, using the suppliers’ own tools to optimise their database will not be enough. Many are turning to other technologies to boost the performance of their database searches and queries.

Retailer Tesco.com optimised its databases by using software-based compression technology to provide faster Microsoft SQL Server backups.

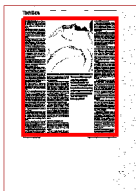
As a result, it significantly reduced its back-up times on the SQL Server databases, which hold information about all of the available products, customer billing, user favourites, shopping basket and delivery data.

Tesco.com’s 120Gbyte SQL Server database was taking 59 minutes for each back-up, with the firm’s database administrators performing a number of back-ups in each 24-hour period. As the volume of SQL Server data grew, the time taken to perform the back-up increased to nearly five hours, which was within Tesco.com’s eight-hour back-up window.

However, the firm wanted to reduce back-up times, and decided to implement Quest Software’s Litespeed compression technology. This cut the back-ups to 18 minutes, said Chris Howell, IT manager for operations and infrastructure at Tesco.com.

“This improvement means that we have ample time to deal with any back-up problems to ensure that Tesco.com is always open,” he said.

→ continued on p28



→ continued from p26

As well as faster back-ups, Tesco.com managed to cut its back-up sizes by 62% by using Litespeed - from 73Gbytes of storage to 28Gbytes.

Many organisations have optimised the data on their databases to improve database performance. One way to do this is to meta-tag the data, which can be done in a variety of ways.

The Norwegian government has meta-tagged its data to optimise searches of 30,000 web documents, which consist of more than 300,000 web pages from 17 of its ministries.

It is using NetworkedPlanet's topic maps-based platform to tag the data with simple categories, such as health, law, or welfare, so that content is easier for website visitors to find. NetworkedPlanet's software is designed for Microsoft .net environments.

Davies Arnold Cooper, an international dispute resolution and real estate law firm, took a similar route to make the two million documents it held more searchable.

Stuart Cowell, head of IT at Davies Arnold Cooper said that for years the firm found it a problem to search across its many databases of client and legal data. It uses an interwoven content database as its primary document repository, but also accesses databases such as PLC Law, Lexis Nexis and Westlaw, as well as government websites.

The firm wanted a "one-stop shop" interface from which to search all the databases, but search categorisation was a problem, said Cowell.

Davies Arnold Cooper issued a tender for a database search tool, and recently piloted a Microsoft SQL Server-based system from Recommind, an enterprise search firm that specialises in legal e-discovery.

"We had a large litigation case with a 30Gbyte hard disc with numerous data and information on it. Recommind managed to copy it, and in three days had it fully indexable. It saved the business a lot of time and cost.

"It is going to make lawyers' and secretaries' lives easier, and hopefully reduce the number of calls to IT helpdesk from people saying they cannot find anything," said Cowell.

Another popular indexing tool is CopperEye Greenwich, used by Vodafone, Orange, Kingston Communications and MessageLabs.

It is designed for large volumes of data, up to hundreds-of-terabytes in size, and uses fast real-time indexing to improve database performance. Rather than using a relational



**Hardware: companies are using appliances to reduce their database management**

database, it requires data to be held in a flat file format on standard off-the-shelf servers.

Online security service provider MessageLabs uses CopperEye Greenwich to track and trace billions of e-mail records on behalf of its customers. The technology indexes the data and provides fast query access to billions of rows of e-mail tracking data, which is stored in flat files on low-cost storage.

This has meant that the supplier has not needed to implement a costly and complex database or datawarehouse architecture, said chief operating officer Carmen Carey.

MessageLabs said the traditional approach to developing a track-and-trace system for its data would have been to load it from the flat file into a relational database or datawarehouse for access. But the large quantity of data, and the need to simultaneously support both loading and fast queries, would have been costly.

#### **The hardware approach**

Another way to see database performance improve is to switch hardware platforms, which may sound like major surgery but can yield results.

Virgin Media found that moving its database over to a dedicated datawarehouse server was the best

**Thompson: "Many people are seeing appliance technology as the way forward. These appliances have dedicated computing and storage resources and many of them now contain their own autonomic healing systems"**

way to optimise its data querying and analysis operations.

The firm is currently trialling the Netezza Performance Server datawarehouse appliance, which it has found performs queries several times faster than its previous database system.

The Netezza appliance is a single piece of hardware that combines a database, server and storage. It can perform detailed queries and analyses on large volumes of stored data because of its dedicated computing resources. Other users of Netezza include Debenhams, Amazon and Carphone Warehouse.

Virgin Media previously outsourced the data analysis process, but the Netezza appliance has in-built business intelligence software, allowing Virgin Media to analyse data in-house. It is expecting to make a rapid return on investment, said Paul Froggatt, a manager in IT application support at Virgin Media.

Froggatt said the firm had so far migrated two of its largest telephone datawarehouses to the Netezza appliance. Virgin Media carried out a four-week trial in November, moving 6Tbytes of data onto Netezza. It recorded a 252% performance improvement on data queries, using dashboard software from Business Objects and Informatica to mine the data.

In mid-February it migrated the second database, which was used by the revenue assurance department. "We have almost recouped the costs of the system by saving money on calls and margins. The parallel architecture of the Netezza system allows the revenue assurance guys to run queries that they never could have dreamed of running. They can get more granular and complex reports over longer date ranges," said Froggatt.

"For example, they can take real call data for the past four to six months and perform 'what if?' scenarios, for example, adjusting our tariffs to see what revenues could be generated on nine billion calls an hour," said Froggatt.

Thompson said that more and more businesses were turning to appliances. "Many people are seeing appliance technology as the way forward. These appliances have dedicated computing and storage resources and many of them now contain their own autonomic healing systems. This takes away a lot of the database management overhead," he said.

**Unearth the benefits of buried data**

→ [www.computerweekly.com/224787](http://www.computerweekly.com/224787)

**Database market now worth £8bn**

→ [www.computerweekly.com/224884](http://www.computerweekly.com/224884)