TeraSolve
Technical Overview

INTELLIGENT ANALYTICAL APPLICATION SERVICE

WHITE PAPER

information/edge
SEE • ANALYSE • CONTROL

TERASOLVE

WHAT IS TERASOLVE?

TeraSolve is an Intelligent Analytic Application Service that enables organisation to perform complex analytic operations to drive considered actions, through rapid, dynamic, empirical insight, into information at the level of detail they require, the dynamic performance they expect and with the flexibility they demand.

TeraSolve enables organisations to apply computationally complex business rules and queries, on massive volumes of data, across multiple business dimensions to deliver business insight, drive innovative strategy, plan and manage precise business execution or meet multifaceted regulatory and governance requirements such as Basel II, International Accounting Standards (IAS) and Sarbanes Oxley. Organisations can apply TeraSolve to a broad range of enterprise scale applications including Risk Management (Financial), Portfolio Profitability Management (Financial), Store Planning (Retail), Product Profitability (Consumer Goods) and Revenue Assurance (Telecommunications).

TeraSolve represents the unique culmination of several innovative advances in analytical application technology, that embrace open standards and permit complex multi-dimensional analysis at a level of granularity and performance that until now has not been possible.

HOW IS TERASOLVE DIFFERENT FROM OLAP TECHNOLOGY?

OLAP technology is widespread and this is testimony to the business benefits that organisations gain by using it. However, existing OLAP product limitations have seen its use restricted to small datasets that have been unable to scale to enterprise levels.

TeraSolve is an Intelligent Analytical Application Service technology that offers, for the first time ever, the proven benefits of OLAP functionality with massive scalability and superior performance thus enabling analytical applications that would not have been feasible using OLAP technology. Thus, TeraSolve offers OLAP functionality without being an OLAP technology.

TeraSolve differs from OLAP technologies because:

■ It boosts a Data Warehouse's analytical performance. TeraSolve's Intelligent Analytical Application Service can be embedded within existing business intelligence infrastructures such as Data Warehouses. Enabling the Data Warehouse to provide a high performance, massively scalable, analytical application service, that users can interact with to read/write, compute and report multi-dimensionally, directly within the Data Warehouse, to derive business insight.

■ TeraSolve enhances existing Data Warehouses. There are a number of different industry accepted approaches to Data Warehousing that organisations have been used to implement Data Warehousing applications. All these can benefit from the addition of a TeraSolve Intelligence Analytical Application Service. Two of these approaches are discussed below, including describing how TeraSolve can enhance these solution architectures.

THE RELATIONAL 'ONLY' DATA WAREHOUSE

In this approach a Data Warehouse primarily uses relational database constructs such as 3rd normal form or 'star' schema's table structures to provide users with a view of data from directly within the Data Warehouse. This approach suffers from several problems:

■ Multi-dimensional data is stored in a series of two dimensional tables that are linked to create a 'star' schema. A multi-dimensional view is only possible using 'table joins' which, require significant SQL processing in the relational database and lead to poor end user query performance as well as challenges in achieving the desired row/column format.

■ This type of Data Warehouse offers a read only view of multi-dimensional data. Users are not able to interact with the data and update it especially at aggregated levels.

■ All computations are limited to what is possible with SQL and since even basic multi-dimensional business rules require complex SQL a user is not able to define them. Indeed many multi-dimensional business rules cannot be coded in SQL irrespective of how conversant the user is with SQL.

By introducing TeraSolve to the solution, users will be able to access data rapidly and easily, including complex calculations, from any dimensional perspective.

THE 'HUB & SPOKE' DATA WAREHOUSE

In this approach a Data Warehouse uses a central relational data warehouse and a number of 'spun off' smaller loosely coupled relational or OLAP based, 'data marts'. This approach also suffers from several problems:

■ All data marts are not only loosely coupled to the core Data Warehouse but also loosely coupled to one another. This means that it is difficult and resource intensive to maintain and update the independent multi-dimensional data and structures.

■ The OLAP based data marts typically store data in proprietary databases and not in proven relational databases hence incurring stability and integrity issues.

■ The OLAP based data marts are not capable of scaling to either large numbers of users or large data volumes thus imposing limits on the types of multi-dimensional business applications they are suitable for. Typically, they are only able to perform analysis on aggregate levels of data. As a consequence many small OLAP data marts are created, which are then required to be loaded and maintained, typically with significant data duplication.

A TeraSolve enabled Data Warehouse can avoid the explosion of lots of independent OLAP cubes by integrating the TeraSolve Intelligence Analytical Services to the central "Hub". TeraSolve can simplify the solution architecture by provide a single integrated repository that will represent all the required OLAP data marts with the dependent business dimensions only maintained once.
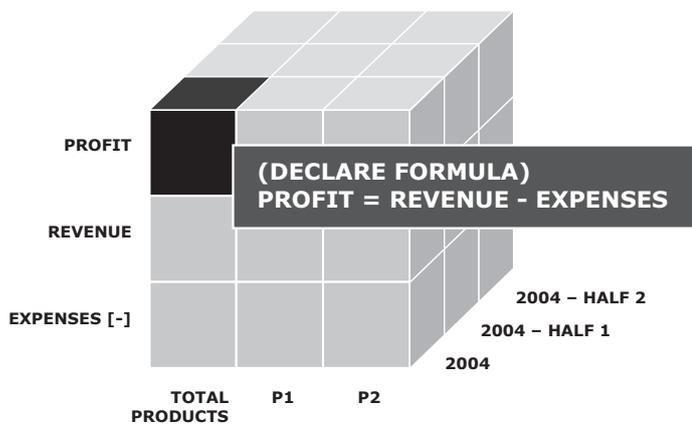
- TeraSolve offers unique and compelling analytic flexibility through advanced functionality such as 'multiple business structures' which for the first time allow organisations
to truly model, compare and contrast evolving business models as they originally were historically or as they will be in the future to empirically understand and manage the impact of change. Thus organisations can both:

  - Preserve previous business models (from both a data and more importantly structure perspective) for later comparison with the current business model e.g. If I compare my current organisation structure with my organisation structure as it was 6 months ago what has changed and what is the impact of this change to my bottom line?

  - Or model future business models (from both data and more importantly structure perspective) to compare them with the current business model and assess the likely impact of change going forward e.g. What would be the impact to my bottom line of deleting, moving or splitting my cost centres going forward?

- TeraSolve exploits the latest, open standards based, application server technology (J2EE application servers) and relational database technology. This means that TeraSolve:

  - Is platform independent and can operate seamlessly across a loosely coupled disparate infrastructure.

  - Distributes computations across multiple clustered servers making it massively scalable.

  - Inherently manages failover making it significantly more suitable for complex, large scale, mission critical analytic application requirements.

  - Benefits from robust and reliable data storage in market leading relational database technology.

  - Seamlessly integrates with market leading systems infrastructures.

  - Provides extremely fast, interactive query and update, performance regardless of the volume of data to be processed.

- TeraSolve uses patented, intelligent computation algorithms that:

  - Support 'declarative' definition of business rules by users. This means that users simply declare required business rules, the TeraSolve computation engine determines:

    - All intermediate multi-dimensional computations that are required to produce numerically correct results for the required business rule.

    - Precedence across multiple computations to guarantee a numerically correct answer.

An example how TeraSolve's computation engine would resolve a simple business rule is provided below:
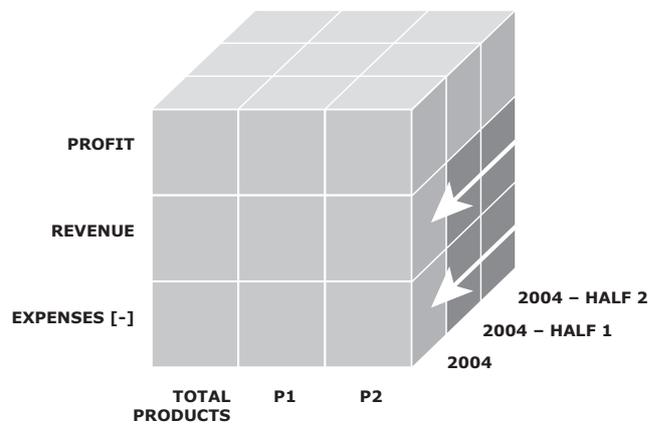
The business rule **[Profit = Revenue – Expenses]** is defined and declared.

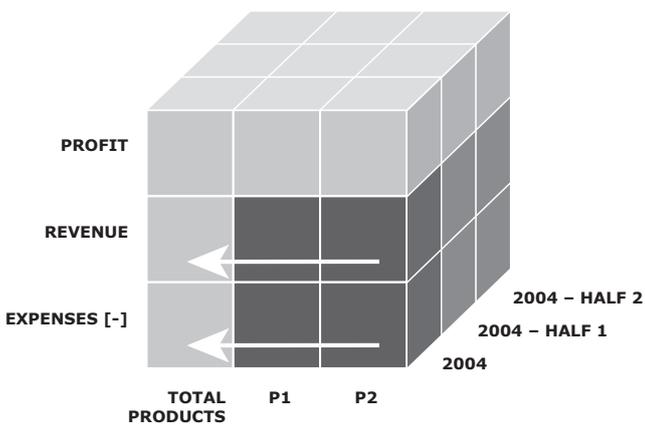TeraSolve when required automatically calculates steps 2 to 4.
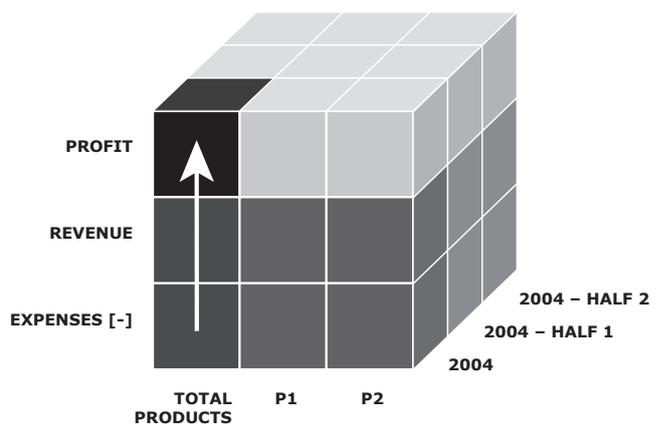
DECLARATIVE CALCULATION

(DECLARE FORMULA)
PROFIT = REVENUE - EXPENSES

PROFIT
REVENUE
EXPENSES [-]
2004 – HALF 2
2004 – HALF 1
2004
TOTAL PRODUCTS   P1   P2

(AUTOMATIC STEP 2) AGGREGATE TIME

PROFIT
REVENUE
EXPENSES [-]
2004 – HALF 2
2004 – HALF 1
2004
TOTAL PRODUCTS   P1   P2

(AUTOMATIC STEP 3) AGGREGATE PRODUCTS

PROFIT
REVENUE
EXPENSES [-]
2004 – HALF 2
2004 – HALF 1
2004
TOTAL PRODUCTS   P1   P2

(AUTOMATIC STEP 4) CALCULATE PROFIT

PROFIT
REVENUE
EXPENSES [-]
2004 – HALF 2
2004 – HALF 1
2004
TOTAL PRODUCTS   P1   P2

This approach is different to the majority of OLAP technologies that rely on a user 'procedurally' defining business rules, all intermediate multi-dimensional computations and the precedence of computations to derive a numerically correct answer. The user effectively has to build in the computation intelligence.

- The internal algorithms within TeraSolve:

  - Are designed to process complex multi-dimensional business rules typically associated with interactive, end-user read/write, modelling requirements. The intelligence built into the computation engine autonomously determines what computations will be required and only computes those that are required. This unique capability allows the computation engine to adapt to multi-user computation requirements in real time.

  - Adapt computation performance based on workload, type of computation and computation complexity. Thus ensuring optimal computation performance and response for all users. Why is query and computation speed critical? Because study after study has shown that slow query performance leads to low user acceptance. The whole attraction of OLAP technology is its speed over relational technologies when used for analytic applications. TeraSolve offers at least an order of magnitude faster performance than current OLAP technologies.

  - Are optimised for high performance processing of complex multi-dimensional business rules, across massive data volumes, at an atomic level. Current OLAP technologies are not capable of processing complex business rules across massive data volumes.

- TeraSolve uses patented, intelligent algorithms to automatically determine how data is to be stored and managed. This capability allows:

  - TeraSolve to adapt storage characteristics based on storage availability, workload and the performance characteristics required. OLAP solutions require users to define at design time what data is to be stored in multi-dimensional structures (MOLAP), in relational database structures (ROLAP), or a hybrid of multi-dimensional / relational database structures (HOLAP). If these decisions are not correct performance and scalability are adversely affected. In effect, OLAP technologies expect designers to predict the future based on vague assessments of future usage.

  - TeraSolve utilises 'pure sparse data storage' that is it only stores data that is required for computation purposes, it does not store empty or null values as is common with OLAP technology. It also uses the principle that multi-dimensional data tends to cluster into groups of 'co-dependent data' to determine how data is stored for optimal performance. This, in conjunction with additional data compression, minimises data storage requirements, reduces cost and optimises performance.

■ TeraSolve decouples the multi-dimensional structure of the business application from the data that supports it. Thus the multi-dimensional business model can be modified, on a real time basis, while TeraSolve continues to service users. This capability is critical for the support of complex large scale analytic requirements that are constantly evolving.

In summary, TeraSolve offers significant advantages over existing OLAP technologies including:

■ Ease of use; TeraSolve schema's visually reflects a multi-model, which is acknowledged as the easiest way for an end-user to understand business problems.

■ Unique and compelling analytic flexibility through advanced functionality such as 'multiple business structures,' which allow organisations to truly model, compare and contrast evolving business models, as they originally were in the past or as they will be in the future, to empirically understand and manage the impact of change.

■ Performance; TeraSolve has a speed advantage of at least 10 times over MOLAP technology. This advantage is evidenced in calculation speeds that can be as much as 50 to 100 times faster with complex data sets and large numbers of concurrent users. MOLAP in turn is accepted as being at least an order of magnitude faster than ROLAP. This makes TeraSolve 20 to 50 times faster than current ROLAP solutions and even faster under certain circumstances.

■ Massive scalability; TeraSolve's multi-threaded technology exploits application server based multi-processing environments giving it massive scalability. OLAP technologies are significantly less advanced and typically do not scale to more than a few processors.

■ Rapid deployment; 50% of all BI projects fail and a vast number of these failures attempt to use ROLAP technology which was never intended to support analytic applications. TeraSolve leads to rapid deployment over a ROLAP approach since significant time savings are gained by not having to optimise the ROLAP schema for faster performance.

■ Cost effectiveness; ROLAP solutions are the most expensive analytic solutions to build and maintain. Even, MOLAP solutions are even more expensive to build and maintain.

■ Real world analytic applications; TeraSolve can handle terabytes of raw data which is comparable with current ROLAP solutions and massively larger than MOLAP technologies which can, at best, handle a few gigabytes. TeraSolve therefore enables true enterprise analytic applications to be built.

■ High availability; TeraSolve stores all of its data in the underlying relational database thereby utilising the robustness, stability and data integrity of a relational database. MOLAP technologies store data in proprietary array based databases which are similar to file based systems which relational databases have superseded due to the inherent instability of the file based systems.

■ Support for open standards; TeraSolve supports industry standard application servers, relational databases and query and reporting tools. An organisation can therefore leverage the skill base that exists in its standard operating environment. There is no need to learn a new proprietary database technology.

■ Hardware platform independence; TeraSolve can run on any platform for which leading application server and relational database technology is available. In addition, it has been certified on Linux, adding a further level of hardware platform independence.

HOW DOES TERASOLVE FIT WITHIN AN ENTERPRISE INFORMATION ARCHITECTURE?
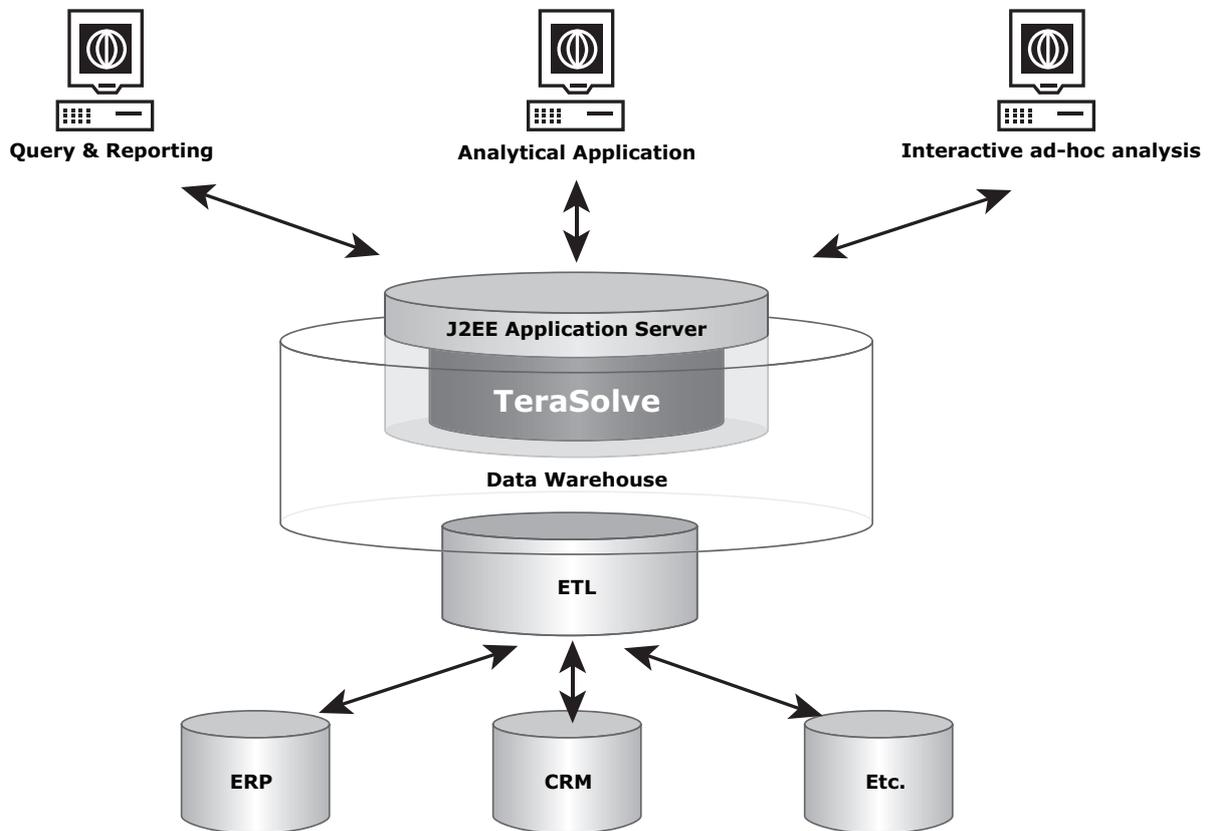
TeraSolve fits seamlessly within an enterprise information architecture as the analytical engine of the Data Warehouse. Consistent with modern informational architectures, it exploits Application Server technology both for computation processing and to provide access to the Data Warehouse.

TeraSolve captures massive volumes of business data from within a Data Warehouse, computes complex multi-dimensional business rules from it and then stores results in the Data Warehouse for further update, query, reporting or ad-hoc analysis.

Users can interact with data either through industry standard Query & Reporting tools, analytic applications or even the ubiquitous spreadsheet for ad-hoc analysis.

HOW IS TERASOLVE USED?

Users define the characteristics of a enterprise multi-dimensional business application within the TeraSolve service. Such characteristics include dimensions, hierarchies within dimensions, user constraints for data access, 'declarative' business rules and data inputs. This information is then used to construct the multi-dimensional structure and compute the necessary business rules. Users are then able to either update one or more of these characteristics (e.g. business rules) or the input/computed numeric data. TeraSolve uses its patented intelligent algorithms to autonomously assess the impact of these updates on the multi-dimensional structure and only re-computes numeric data affected by the updates.

Using these principles of definition and then update, users are able to interact with the multi-dimensional structures and data to derive business insight.

HOW DOES TERASOLVE WORK?

TeraSolve uses a 3-tier service architecture:

| CLIENT TIER | APPLICATION SERVICE TIER | DATABASE TIER |
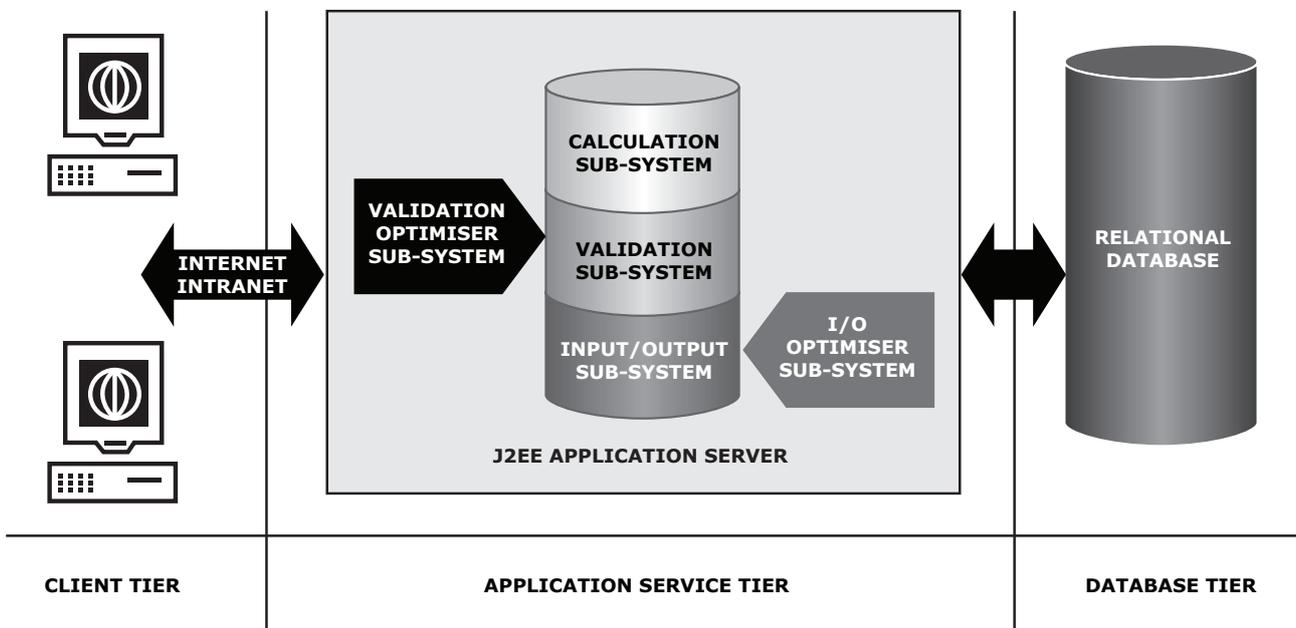
CLIENT TIER

The client tier is the set of web native tools that provide a Graphical User Interface for users to define, maintain and administer TeraSolve applications. These tools also provide a straight forward mechanism for loading data into the database at whatever granularity and frequency the multi-dimensional business application requires, including the ability to alter the multi-dimensional business application structure automatically based on the input data. Unlike current OLAP tools data load is not a batch process in TeraSolve and can be performed whenever needed with all users logged in and working (this includes the multi-dimensional application structure changes).

Reporting tools are also part of the client tier; TeraSolve supports all main stream analysis standards such as OLE DB, JOLAP and XMLA. This means that any industry standard query and reporting tool that conforms to open standards can connect to and report from the TeraSolve database. If the analysis tool implements XMLA or JOLAP then reporting can be performed with no TeraSolve drivers on the client machine at all.

APPLICATION SERVICE TIER

TeraSolve exploits the next generation J2EE enterprise application platform. This platform offers a new level of abstraction which delivers a number of benefits including:

■ Platform neutrality; J2EE is both hardware and operating system independent.

■ Application server clustering means J2EE inherently offers:

   ■ Load balancing; J2EE applications distribute work across multiple application servers. Thus TeraSolve applications can scale and perform far beyond the limits of current OLAP technologies.

   ■ Fail-over redundancy; J2EE applications keep parallel copies of critical data in different physical machines.Thus TeraSolve application can survive any single machine's untimely demise without missing a beat.

■ Adherence to open standards.

■ Direct, reliable, high performance integration with modern infrastructures such as Data Warehouses.

■ Operation as a service.

The core TeraSolve engine runs within a modern J2EE enterprise application framework. The TeraSolve engine has three main sub-systems:

■ **Calculation sub-system;**

The Calculation sub-system is responsible for intelligently managing the aggregation process. Aggregation may involve simply summing together data in a multi-dimensional structure, or may extend to the evaluation of complex multi-dimensional business rules. It communicates through the Validation sub-system to determine what data is required to be re-computed.

TeraSolve uses the industry standard scripting language JavaScript for the definition of business rules within the system. JavaScript is a widely accepted industry standard for scripting and has the benefits of being both easy and natural to pick up and yet able to express the most complex and functional computations businesses may need.

■ **Validation sub-system;**

In a multi-dimensional structure each individual data item is referred to as a 'cell'. The number of aggregated cells in a multi-dimensional structure is combinatory and therefore can be massive. Storing all cells, in anything but the most trivial of schemas, is generally physically impossible on commercially available hardware. The Validation sub-system decides what aggregated cells to store. Conceptually there exists a "Persistence Curve" that runs across the entire multi-dimensional structure. The Validation sub-system decides from the curve whether to store a cell or not. Generally, cells above the curve are not stored while those below are. Aggregated cells below the curve are flagged as either valid or invalid. A stored aggregated cell is made invalid when any of its dependent input cells within a dimensional hierarchy are updated. It will be made valid when the Calculation sub-system next aggregates the cell to its correct value.

The position of the persistence curve changes the performance characteristics of TeraSolve. Generally a low lying curve within the multi-dimensional structure will improve write performance but decrease read performance. Conversely a high lying curve will improve read performance but decrease write performance. TeraSolve automatically tunes the persistence curve to optimise performance.

This system has a patent pending; NZ Patent Application No 527534.

■ **Input/Output sub-system;**

The Input/Output sub-system is responsible for read/write access to an individual cell within the multi-dimensional structure. For each cell it stores cell data and the indexes to each cells physical storage location. The indexes provide a loosely coupled mapping between the multi-dimensional structure and the physical cell storage locations.

Two main advantages are realised by separating the structure from the cell storage are:

■ Cell storage can be reorganised independently of the structure to give optimal performance for specific cell accesses.

■ The structure can be modified without having to take the database offline for restructuring.

This system has a patent pending; NZ Patent Application No 527535.

In addition to the three main sub-systems the TeraSolve engine uses two further optimiser sub-systems:

- **Validation optimiser sub-system;**

The Validation optimiser sub-system continually adjusts sections of the Persistence Curve, moving it up or down as required within the multi-dimensional structure depending on actual system usage and performance.

- **Input/Output optimiser sub-system;**

The Input/Output optimiser sub-system continually gathers statistics from individual cell accesses and uses these to:

  - Rearrange stored cells for optimal performance.

  - Reorganise the index order for optimal performance.

DATABASE TIER

Relational databases are designed to store data in a tabular form which is completely unsuitable for storage of multi-dimensional formats. However, they do offer storage of data in customisable binary formats. TeraSolve uses this feature of modern relational databases to store multi-dimensional data in an optimised binary format and then applies its patented intelligent indexing and look up algorithms to retrieve data resulting in massively faster performance than the native relational database could deliver.

TeraSolve use of relational database technology for storage and retrieval of data ensuring that applications benefit from:

- Reliable storage technology with full backup and recovery built in.

- Strong performance through optimal relational database storage.

- Platform independence.

- Strong relational database skills in most organisations.

SUMMARY

In summary, TeraSolve:

■ Is a true quantum leap in analytic application technology that handles real-world analytic flexibility through advanced functionality e.g. 'multiple business structures,' massive data volumes and enterprise-wide scalability requirements.

■ Offers several orders of magnitude performance advantage over OLAP technologies.

■ Supports open standards thereby enabling organisations to leverage in-house skills and stick to their standard operating environment.

■ Offers cost effective, rapid, low risk deployment with low ongoing running costs.

■ Can be deployed on most hardware platforms and Linux environments leading to even greater efficiencies.

CONTACT DETAILS

For more details please contact your nearest office:

UNITED KINGDOM
42 Langham Street
London
W1W 7AT
T. +44 207 255 2633

AUSTRALIA
Level 39
Citigroup Centre
2 Park Street
Sydney
New South Wales
T. +61 2 9004 7851

NEW ZEALAND
Level 2
268 Manukau Rd
Epsom
PO Box 8239
Auckland
T. +64 9 366 0710

UNITED STATES
12021 Wilshire Boulevard Suite 223
Los Angeles
CA 90025
T. +1 310 656 0476