Smart (enough) Systems LLC

Meeting Demands for Data Warehouse Performance

Expanding Operational Roles for Data Warehouses Require Dynamic Workload Management Capabilities

Sponsored Research

TERADATA
Raising Intelligence

By Neil Raden
Smart (enough) Systems LLC
February 2008

SMART (enough) SYSTEMS
DECISION MANAGEMENT SOLUTIONS

1415 Kenwood Road
Santa Barbara, CA 93109
info@smartenoughsystems.com

Table of Contents

1 Executive Summary 3

2 Optimization in Data Warehouse Performance 4

2.1 ODS: An Isolation Solution 5
2.2 Mixed Workload Alternative 6

3 New Challenges for Business Intelligence 7

3.1 Operational Intelligence 7
3.2 Enterprise Decision Management 7

4 The New Data Warehouse Roadmap 9

5 Dynamic Mixed Workload Management 10

5.1 How Does It Work 10

6 Benefits of Mixed Workload Management in Data Warehouses 12

6.1 Consolidation 12
6.2 Continuous operation 12
6.3 SLAs and Distributed Architecture 12
6.4 Leverage Physical Environment 12

7 Conclusion 13

8 About the Author 13

1 Executive Summary

Everything changes. Computers were once merely tabulating machines until programming languages were invented. The Internet was just a communications network before the Web emerged as the world-changing phenomenon it is. Now it's time for data warehousing to move on. Although they were once separate disciplines, applications for business analytics and business operations are converging, requiring new, dynamic approaches to collecting and delivering information.

Data warehousing is entering a steep trajectory. In the beginning, a data warehouse was a single, persistent database that gathered information at certain intervals from numerous internal sources. It was useful for shortening the report development backlog of IT organizations. It also took the pressure off the mainframe by relieving it of the rapidly increasing requests for extracts and queries needed to service those requests.

In short order, the data warehouse concept started to expand – first, to being an historical repository and then being directly available to non-programmers. Before long, it was an entire process composed of data integration, aggregation and integrated reporting and analytical tools. The scope grew to include people and data beyond the boundaries of the organization. New data arrived not every thirty days, but every thirty minutes or thirty seconds. The queries presented to the database became more complex and numerous.

Today, data warehouses must absorb new data in near real time and provide extraordinary performance, squeezing most of the latency[1] out of the process. Being able to provide fresher data enables the data warehouse, in turn, to participate in ongoing operational processes. But marrying the operational processes with data warehouses, while still supporting all of the legacy analytical functions, is not trivial. Balancing the mixed workload of high-volume simple queries, long running complex queries, near real-time updating and all of the other functions is a huge challenge.

No approach or methodology can remain static when the surrounding environment is changing rapidly. The prominence, first, of the Internet connecting in-house systems with the outside world, and more recently, whole new distributed architectures based on widely accepted standards, is causing all enterprise applications and business processes to be radically redrawn. The concept of an enterprise data warehouse surrounded by many dependent databases to serve various needs, both in terms of content and currency, is being replaced with a single warehouse structure that can do it all, not just for economy, but for fluidity and timeliness.

On the trajectory going forward is a complete revision of the enterprise computing architecture in favor of distributed, loosely coupled services, such as SOA (services oriented architecture). The consequence is unfathomable increases in the amount and diversity of data that can be captured and used. In addition, the newest members of the workforce have a facility with technology, in both their work and their lives, that is markedly different from their predecessors. This is already manifesting itself with greater frequency and complexity of queries presented to the data warehouse.

The implications are clear. A fast database and a few good DBAs no longer suffice. Data warehouse platforms must be able to perform all of the required tasks, simultaneously, without manual tuning. Intervention by skilled professionals, who are always in short supply, simply can't respond quickly enough. Their role is changing accordingly, from periodically adjusting the physical instance in reaction to changing workloads, to modeling the workload so that algorithmic processes can manage the workload in real time. This is a more realistic man-machine cooperative – humans defining the problem space and its variables and computers firing the rules dynamically, in real time, to manage the mixed workload with optimal performance.

1 Latency is defined here as the time between events, such as the availability of new data to load into the data warehouse and the availability of the data for reporting.

2 Optimization in Data Warehouse Performance

All optimization problems are hard by their nature, and data warehouse workload management is no exception. The usefulness of an optimization model is determined by how well it identifies the many variables of the problem and how effectively it captures the important dependencies between them. But in most Information Technology departments, there is hardly enough time to put applications into production and keep them running. Thinking through optimization is a luxury most organizations cannot afford. The alternative is a combination of hardware safety buffers, governors, and workload isolation.

In database systems, when it isn't possible to optimize a process, one alternative is to provide excess capacity. If a system typically peaks at 65% capacity, a buffer of 35% provides some assurance that a spike in usage will not cause poor performance. This is an expensive alternative and not completely reliable.

Governors limit access to the resources, either through restriction to the data resources, time limits on queries or even limiting the time of day access is available. For most organizations today, this is unacceptable.

Isolation, which is separating different processes on different platforms, is used widely in IT organizations for managing workloads. Typically, performance and usage characteristics of individual applications are much easier to understand in isolation, allowing for more accurate planning. Also, mixing multiple applications on the same platform can and often does result in unpredictable combinations that cause performance problems that are neither foreseen nor easily resolved. This is even true in data warehousing itself, where many enterprise data warehouse (EDW) implementations separate the EDW from the data marts[2] that actually drive the business intelligence (BI) operations. This actually combines both isolation and the excess capacity game plan where each subject area gets a server of its own that also requires optimization. Each instance also needs its own performance buffer of excess hardware, increasing overall costs, complexity, and multiple points of failure.

Isolation works very well when taken from a systems perspective for a single application. However, data warehouses exist to serve the needs of people in organizations, and they are inherently multi-use. Furthermore, one system at a time is an approach that no longer works. There are real drawbacks to the isolation approach in data warehousing:

- **Expensive:** A constellation of servers and database licenses are required to provide the various reporting, analysis, data mining, operational inquiries, and ad-hoc exploration by subject area. Combined with the maintenance and tuning required for all of them, this can be a very costly proposition.
- **Interfaces:** Every application that is dependent on the EDW requires interfaces such as ETL routines, scripts and even manual entry and checking. Every one of those hand-offs costs time and money.
- **Inconsistency:** As data flows from an EDW to myriad downstream applications, the potential for error creeps into every process. These inconsistencies can be caused by programmatic errors, timing differences, or misunderstandings caused by incorrect or incomplete metadata.
- **Latency:** Potentially the greatest drawback of all, isolation vastly increases latency in the process. Not only do downstream databases have to wait for upstream processes to complete, a failure in an upstream process can cause a cascade of failures throughout the system. Reducing latency in applications is an increasingly important goal in BI applications. Isolation works against latency reduction.

2 There is no clear, unambiguous definition for "data mart." Data marts can often be completely independent collections of data that are not dependent on a data warehouse. More commonly, they are subsets and/or derivations/aggregations of portions of the EDW.

2.1 ODS: An Isolation Solution

The "classic" data warehouse operated in only two mutually exclusive modes: loading and querying. When it was shut down for incoming queries, bulk updating occurred at regular intervals (often monthly, but never more frequently than daily). The data warehouse schemas were designed to provide a decision support view of the subject areas. For example, historical data was arranged in a way to support trend analysis and actual-versus-variance reporting. When loading was complete, the data warehouse was available to process queries.

Initially, this was a giant leap forward. But in time, organizations found that the rich reporting capabilities of the data warehouse environment were sorely needed for reporting more up-to-date, operational information. This posed a dilemma because the database schemas devised to support BI were optimized for bulk updating and read-only queries. Adding more timely data, especially operational data that was not integrated into the data warehouse models, posed a serious problem.

The solution was a sort of half-way data warehouse structure called an Operational Data Store (ODS). An ODS exists as a separate data model and physical schema within the data warehouse environment. Typically, there are multiple ODSs because each one is composed of transactional data from a single operational system. Logically, the ODSs often do not relate to the rest of the EDW at all; they are more or less just a separate application within an overall data warehouse environment and accessed with the same tools. A common misconception is that data flow to an ODS first, in smaller batches and more frequent intervals than the data warehouse, but are integrated with the EDW at each update cycle. In fact, this rarely happens. Instead, operational data is usually pulled from the operational systems twice, then transformed and integrated separately based on the data models of the EDW and ODS. This process duplicates work and often leads to inconsistencies due to both timing differences and separate logic to maintain.

ODS was a solution for providing functionality that can now be delivered more efficiently. For example, a common ODS application was providing live data for customer touch points such as a call center or website. The idea was that the data needed to interact with a single customer was too low-level for a data warehouse, and needed to be more current than daily update cycles. Another application was reporting data marts for single functions such as Accounts Payable or Work In Progress. Attempts to create master data files, such as product masters, were also commonly assigned to ODSs because there was no other alternative.

The rationale for ODSs has faded. Today, many EDWs provide the same intra-day updates as well. The operational data originally designated for the ODS can now flow easily into the EDW. In fact, fine-grained data usually carries interesting attributes that are the fodder of advanced analytics, like data mining and predictive modeling. The amount of data that today's databases can ingest and manage is orders of magnitude greater than a decade ago when the ODS idea was in full bloom. It is no longer necessary to duplicate ODS data in the EDW in order to get comprehensive real-time reports or complete customer touch point insights.

All of the drawbacks of an isolation optimization solution are endemic in the ODS approach: expense, interfaces, inconsistency, and latency. The real-time, fine-grained operational data handled by an ODS can now be accommodated in an EDW, *provided there are adequate mixed workload management capabilities*. The solution is a database platform that can provide predictable performance and scalability while handling a variety of user types, query types, and loads.

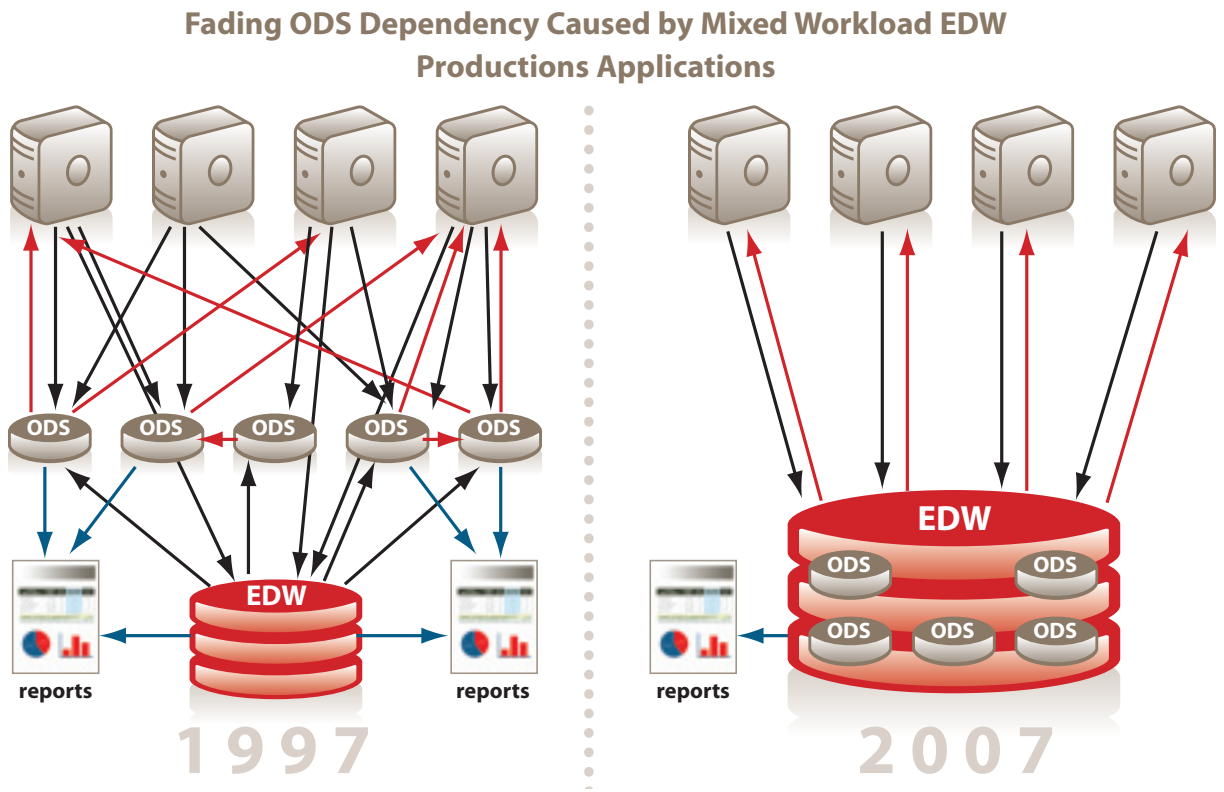

Figure 1. The left-hand side of this figure shows the complicated instances of many Operational Data Stores versus the simplified nature of the ODS within the enterprise data warehouse. Both the ODS and the EDW generate reports.

2.2 Mixed Workload Alternative

The alternative to isolation is dynamic mixed workload management. The problem is that mixed workload management is also an optimization problem, and a hard one at that. All relational databases have query optimizers, but they are limited to optimizing a single query – they find the most efficient way to resolve one query without considering the optimization of the entire system. It's not unlike finding the fastest route home without considering the traffic, weather, or time of day. However, when a data warehouse serves only a single purpose, such as production reporting or feeding data marts, query optimizers can often suffice. The system workload is more or less predictable, and the optimization schemes generated by the query optimizers are more or less alike. Systems tuners and DBAs can configure a database to perform well and predictably in these cases because of the homogeneity of the workload.

The difficulty arises when a single data warehouse has to meet the needs of a variety of applications. This requires thoughtful and purposeful design, not on the part of the DBAs, but rather the designers and engineers of the database platform itself. These solutions, to meet the new challenges for BI, are only just now emerging after years of development.

3 New Challenges for Business Intelligence

Businesses used to operate as standalone enterprises, and the information they consumed was largely internal. But businesses are externalizing their transactions, operations, and even their staffs, creating demand for new approaches to data warehousing to support new kinds of analytical applications. Two concepts in particular that are driving the reinvention of BI are Operational Intelligence and Enterprise Decision Management.

3.1 Operational Intelligence

Operational systems have been, until recently, comfortably dumb. They process transactions quickly, apply relatively straightforward rules reliably, and manage the data they gather and create effectively and securely. Systems in the BI world are considerably smarter, imbued with things like flexible hierarchies, time intelligence, and moderate to complex modeling capabilities.

But largely due to the expanded business opportunities created by the Internet, these separate but equal worlds can no longer exist oblivious to each other. Once a business process extends beyond the familiar confines of one's own business, even the simplest transaction systems need to be a little smarter. Terms don't mean the same thing, understanding who is on the other end of a transaction and their relationship to you becomes crucial. Most important of all, decisions based on analysis such as stock levels versus forecast, propensity to purchase, or fraud probability become crucial. A large percentage of operational transactions can benefit from injecting a little or a lot of fact-based decision support.

Operational intelligence requires different analytical models from those most closely associated with BI today. It's easy to utilize existing approaches to measure what happened in a business process. However, only a convergence in business process modeling, business process management, and business intelligence can deliver true operational intelligence that can detect process events as they occur while drawing reasonable conclusions about the situation in real time. This is the frontier of data warehousing today, and it will be commonplace very shortly.

BI tools have not penetrated much of the organization yet because they are oriented toward analysts, which usually implies after-the-fact reporting of the business events. This is still a useful and valuable effort. But BI as such has much more to offer to a wider population, hence the term operational BI.

Beyond just reporting and alerting people about operations, there is also a need for fully-unattended systems that are capable of making huge numbers of decisions with accuracy – an emerging area called Enterprise Decision Management[3] (EDM).

3.2 Enterprise Decision Management

EDM is not a product or a technology; it is a framework for using existing technologies. The goal of EDM is to unify the analytical and operational systems of an organization, imbuing the operational systems with the "smarts" of BI and bringing analytical processes into operations. EDM is an approach for automating and improving high-volume operational decisions. Focusing on *operational decisions*, it develops *decision services* using *business rules* to automate those decisions, adds analytic insight to these services using *predictive analytics,* and allows for the ongoing improvement of decision making through *adaptive control and optimization.*

3 For a detailed description of EDM, see *Smart (Enough) Systems* by James Taylor and Neil Raden, Prentice Hall, 2007.

The data warehouse is an integral part of the EDM architecture. It provides the data and the powerful resources to analyze the data at the formative steps of EDM. The result of this analysis is the creation of predictive models that drive the creation and continuous enhancement of business rules. Some business rules do not derive from these models, such as policies and statutes, but the ones that drive competitive advantage usually do. This places the data warehouse in a more direct and immediate path of the formation and execution of business tactics than ever before.

EDM also includes adaptive control mechanisms that continuously test the results of the decision services, not only by analyzing the outcomes, but by running tests against real-time events, constantly testing scenarios and algorithms for better approaches. The data warehouse is a critical component in this step as well.

Enterprise Data Management Architecture

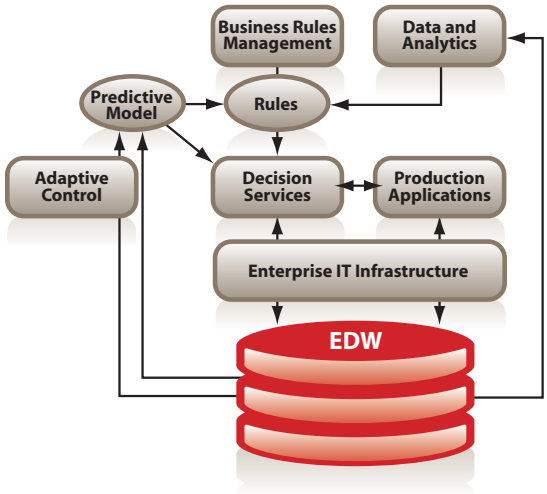

Figure 2. This diagram is borrowed from Smart Enough Systems.

4 The New Data Warehouse Roadmap

Data warehouses are now expected to perform complex analytics, reach into operational systems in real time to provide up-to-the-minute snapshots, run cooperatively with operational systems, and provide historical and analytical context to decisions – all in addition to their original role. The implication is that the data warehouse has to simultaneously support many types of queries, such as:

The strategic query: Strategic queries are large, resource-intensive queries designed to answer complex questions such as why something happened or which alternatives are best.

The analytical query: An insurance company that allows the risk managers of its client companies access to claims data (sufficiently anonymized) for questions such as, "Compare the frequency and severity of our workers compensation claims to other mid-sized manufacturing companies by type of injury and location."

The operational query: If we're under allocation, should we ship to this customer, or put it on back order? Or, what's the next-best offer we can prescribe when the consumer contacts the call center? These are the queries most often used by EDM and Operational BI applications.

Reporting query: Sales by quarter by major category, page breaks on store, this year versus last year.

In addition to queries, a data warehouse must balance its workload between these different kinds of queries and loading data, while at the same time ensuring that performance and response times remain consistent and predictable. In the past, service level agreements (SLAs) for data warehouses were rarely established so the performance of queries was never negotiated. Now that data warehouses are expected to continuously load data and integrate with operational systems, they have to adopt fairly aggressive performance standards.

The progression of technology solutions and the relentless search for more efficiency in operations are the driving forces behind this new roadmap for data warehousing.

5 Dynamic Mixed Workload Management

Workload management in most database systems is often simply managing the running and queued jobs by their pre-assigned priorities, and weighing that against in-process utilization. True dynamic mixed workload management in data warehousing has to assign resources based on the type of work being performed and dynamically reassign resources on an almost instantaneous basis, making judgments based on the model and the rules.

Any useful optimization scheme operates as a closed loop:

- A model of the expected workload is developed.
- Rules are developed that apply that model to the environment as it operates determining, among other things, resource entitlements and priorities.
- Workload metrics are gathered for analysis to determine if the model is achieving its goals, in this case, service level agreements, uptime percentages, turnaround times and any metrics the operators of the system find useful in demonstrating compliance and diagnosing problems.
- Metrics are evaluated and modifications are made to both the model and the rules, if necessary, or are used as a basis for suggested modifications in the physical environment.

Why is this necessary? If systems are scalable and computing resources are plentiful and relatively inexpensive, wouldn't a better solution be to just add more resources when problems arise? The answer is no, because without understanding the problem, patching it with more computers is rarely a sustainable solution. Operating systems are not aware of the internal processes of a database platform, only what resources it is using. Only the database platform can understand the characteristics of the work in progress and in queue. The data warehouse, in this case, has to perform its own dynamic workload management and, as mentioned above, that workload has become decidedly mixed and will continue to be more so. Database management software, in this environment, is woefully incomplete if it cannot understand the characteristics of a unit of work and weigh it against its model to determine how to assign (or dynamically reassign) the resources. Finally, databases need to begin to exhibit characteristics of autonomic systems – detecting and addressing problems without assistance from console operators.

5.1 How Does It Work?

A good example of EDW workload management in action is the Priority Scheduler feature of Teradata.[4]

Teradata uses a concept called allocation groups. This is a fairly coarse-grained model – allocation groups can be set up for finance, marketing, or sales people. When queries or load jobs are submitted, they are identified as belonging to an allocation group that has a pre-assigned "relative weight" (i.e., priority) inside the RDBMS. This is relatively easy to administer as there are generally fewer than a dozen groups. This works well with a wide range of queries. However, always classifying a user into the same priority grouping limits the user to one priority, even when they may have widely varying queries to submit.

A more active process, like the Teradata Active Systems Management, has the advantage of classifying tasks on the actual query being requested. This fine grained classification means a user (or another application) can submit a very complicated query then turn around and submit a one second query with each of the two tasks assigned to a different priority queue. Furthermore, dynamic workload management helps to prevent runaway queries or system

4 For a more complete technical description of the Teradata Mixed Workload offerings see Teradata Active Systems Management: **www.teradata.com/t/page/145613/**
Introduction to Priority Scheduler: **www.teradata.com/t/page/87006/**

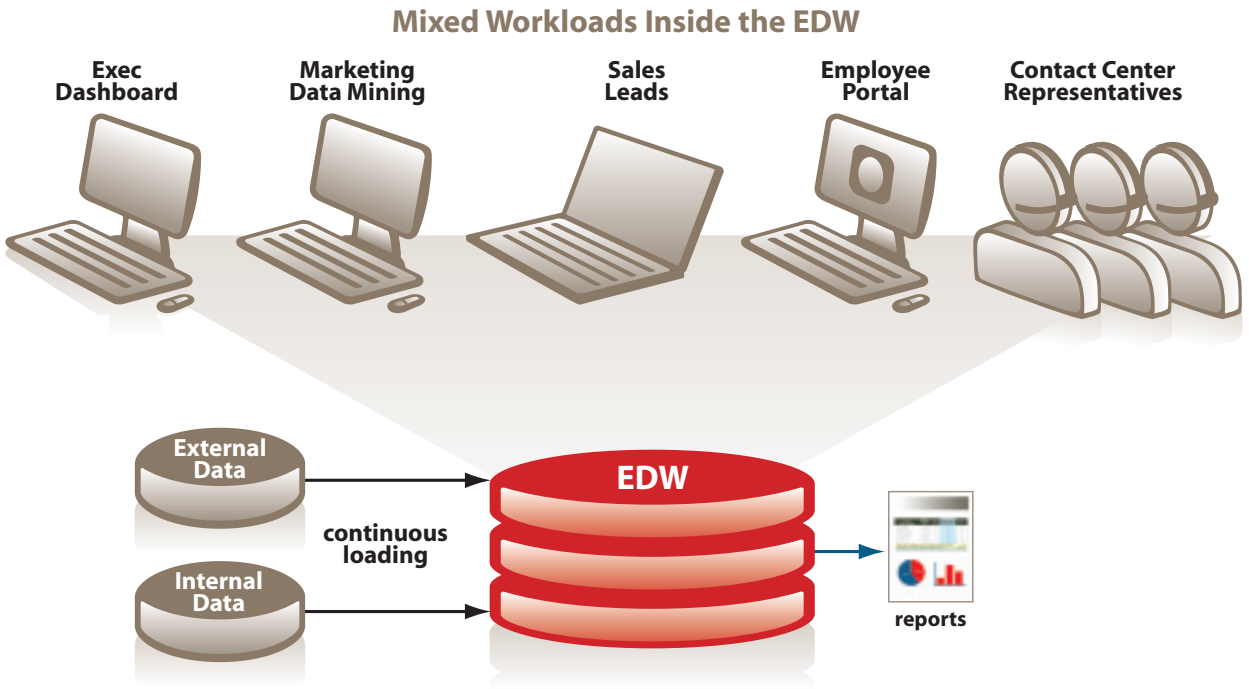

Figure 3. This figure shows reports, data loading, dashboards, and tactical requests happening at the same time in the EDW. True dynamic mixed workload management in a data warehouse must assign resources based on the type of work being performed and dynamically reassign resources almost instantaneously based on the model and the rules.

overloads with rules that can filter jobs and even throttle them up or down. The rules are set up by the system administrator and maintained over time as conditions change. The rules can, among other actions, limit the number of concurrent queries allowed from a single allocation group, either absolutely or based on the time of day, restrict the size of answer sets, or stop queries from initiating a full table scan. This automated exception handling does what a team of experts cannot – make changes hundreds of times per second. And because it's a dynamic optimization, the workload manager can allocate 100% of the resources to low priority tasks when nothing else is running, then switch resource allocations when high priority tasks arrive.

Another technique is allowing certain tasks a "fast path" through the system. This is crucial for those high-priority tasks with stringent service level goals. The workload manager has to analyze each database request before and during execution, assigning priorities based on the rules. But if a "fast path" job appears in the queue, the workload manager can assign the needed resources at the highest priority and set lower priorities for reporting and data.

6 Benefits of Mixed Workload Management in Data Warehouses

Rationalizing and redesigning functional environments such as data warehouses has to be driven by some type of rationale. In other words, there has to be a significant goal or reward to offset the cost and the risk of the effort. Mixed workload management offers significant performance and efficiency benefits and opportunities in consolidation, continuous operation, and conformance within the rapidly-evolving world of standards-based, distributed applications (SOA).

6.1 Consolidation

Mixed workload management facilitates consolidation of resources on a number of levels. At the very least, the many separate database servers and interface programs between them can be consolidated into a single image if the data warehouse platform can service all of the various requirements from one location. In addition, the database schemas themselves, both logical and physical, can be consolidated into a single model. This eliminates the need for costly maintenance and enhancement activities in an environment with many strong dependencies. A significant portion of maintenance budget can be exhausted in researching, developing, and testing changes in complicated, workload isolation environments.

As mentioned before, designing the data warehouse to accommodate the most detailed data from the start, and allowing for periodic updates on an intra-day basis as the organization requires, can eliminate all of the ODS structures and associated programs.

6.2 Continuous operation

Managing the resources to service queries, load new information and any other activities with a dynamic, model-based approach to workload management also increases the likelihood that SLAs can not only be met, but can be negotiated based on realistic expectations. It is also more likely that resources can be leveraged fully and system expansions predicted accurately without the need for expensive buffers of excess hardware or restrictions on usage.

6.3 SLAs and Distributed Architecture

When update cycles were scheduled monthly or weekly, or even overnight, faults were curable by alerting system operators to manually intervene. With data load cycles disappearing and data warehouses expected to provide 24 by 7 performance, advance warnings to performance bottlenecks coupled with programmatic suggestions for resolving them are absolutely necessary to maintain SLAs. Where previously DBAs had the leisure to study logs and query system tables to analyze the problem, time has been compressed to the point where this is no longer desirable. Like EDM systems, the data warehouse must be automatically governed by rules.

6.4 Leverage Physical Environment

Computer hardware may be relatively inexpensive compared to five or ten years ago, based on capacity and throughput, but that doesn't mean it's cheap. In fact, the cost per gigahertz of CPUs, gigabytes of memory, or terabytes of storage may have fallen precipitously, but the processing volume has risen to match. In fact, in some cases – notably disk drive reads – the hardware progress has not been nearly as dramatic as portrayed. Optimizing hardware investments is still a very critical function of a well-oiled data warehouse environment.

While adding physical resources to a computing environment is a less painful, less risky, and less time-consuming process than it used to be, latent hardware resources are still something to avoid. Dynamic optimization capabilities, especially in mixed workload data warehouses, can provide for better utilization of hardware and avoid still costly investments in upgrades.

7 Conclusion

Mixed workload management is an optimization solution to a difficult computer science problem. It removes several technical barriers to having one database performing multiple tasks as efficiently as possible.

Blended analytical and operational processing means more timely access and use of information. Ramping up the audience for BI, both behind and beyond the firewall implies scalability and diversity and lots of it. A dynamic mixed workload capability in data warehousing can create value in two directions – cost savings and enablement of other high-value applications that have not been possible before. Operational Intelligence and Enterprise Decision Management are two examples. Managing a mixed workload effectively is the first step in closing the loop between analysis and action that has been open since analytical software appeared.

Dynamic mixed workload management is a departure from the original concept of the warehouse, which was based on isolation. Isolation is a technique that works well when there is time to spare, but in today's environment, there is none.

8 About the Author

Based in Santa Barbara, CA, Neil Raden is an active consultant and widely published author and speaker and is the co-author of the recently published ***Smart (Enough) Systems*** (Prentice Hall, 2007). He is the co-founder of **Smart (enough) Systems LLC**, which provides research, consulting, systems integration, and implementation services in Decision Management/Automation, Operational Business Intelligence, Data Warehousing, Analytics, and Information Integration for clients worldwide. He welcomes your comments at **nraden@hiredbrains.com**.

Smart (enough) Systems LLC. helps clients automate and improve the decisions underpinning their day-to-day business operations.

Smart (enough) Systems LLC. is the only full-service, vendor neutral company focused on the enterprise decision management marketplace, providing research, advisory services, and implementation.

The company's singular focus on the automation and improvement of operational decisions allows it to bring multiple classes of technology together, such as information integration, data warehousing/business intelligence, analytics & predictive modeling, business rules management, and business process automation. Our solutions solve business problems such as business agility, operational business intelligence, analytic competition, and business process management.

1415 Kenwood Road
Santa Barbara, CA 93109
info@smartenoughsystems.com
Visit us at **www.smartenoughsystems.com**

© 2008 Smart (enough) Systems, LLC.
Duplication only as authorized in writing by Smart (enough) Systems LLC. Teradata is a registered trademark of Teradata Corporation and/or its affiliates in the U.S. and worldwide.