Optimizing OLAP Workloads with the QLogic 10000 Series Adapter

Accelerate Query Execution and Reduce Report Run Times by Sharing Cached OLAP Data

The QLogic 10000 Series Adapter brings more critical OLAP data within reach, providing a deeper view of the enterprise never before available.

KEY FINDINGS

- The QLogic® FabricCache™ 10000 Series Adapter allows cached data cubes to be shared among OLAP report servers, providing faster query execution and enabling more comprehensive reporting.

- The View Selection Problem is aggravated by slow access to disk resources. QLogic FabricCache accelerates access to aggregate data, allowing deeper analysis from more data resources.

- The QLogic 10000 Series Adapter is the only enterprise solution that combines an industry-leading Fibre Channel Adapter with on-board caching intelligence and I/O management on a single card. The solution is transparent to the host server, OS, and the OLAP application, so no modifications to the OLAP environment are necessary.

INTRODUCTION

Analytical workloads can do some real heavy lifting. They sort through reams of data looking for very specific answers. Data is aggregated and presorted, making reporting and adhoc queries more efficient. Server side caching with the QLogic FabricCache 10000 Series Adapter technology is strategically positioned in the data path, allowing online analytical processing (OLAP) applications faster access to more data, which reduces the execution times for the critical reports that guide executive decision makers.

OLAP APPLICATIONS

Typical OLAP applications include business reporting and forecasting for sales, marketing, business process management, budgeting, financial reporting, and similar areas with new applications being discovered every day.

HISTORY

The first commercial product that performed OLAP queries was Express, which was released in 1970 by Information Resources. The term OLAP was coined by Edgar F. Codd in 1993. Codd has been described as “the father of the relational database”. OLAP experienced strong growth in the late 90s with dozens of commercial products coming to market. In 1998, Microsoft® released its first OLAP server - Microsoft Analysis Services, which drove wide adoption of OLAP technology and moved it into the mainstream.
Unlike relational databases, which had SQL as the standard query language, and widespread APIs such as ODBC, JDBC, and OLEDB, there was no such unification in the OLAP world for a long time. The first real standard API was OLE DB for an OLAP specification from Microsoft that appeared in 1997 and introduced the MDX query language. Multidimensional Expressions (MDX) is a query language for OLAP databases, much like SQL is a query language for relational databases. It is also a calculation language, with syntax similar to spreadsheet formulas.

Several OLAP vendors—both server and client—adopted it. In 2001, Microsoft and Hyperion announced the XML for Analysis specification, which was endorsed by most of the OLAP vendors. Because the specification used MDX as a query language, MDX became the de facto standard.

MINING FOR NUGGETS
As any seasoned prospector will tell you, gold is not found in large nuggets lying on the ground. It is discovered in very small particles embedded in large amounts of ordinary rock, mud, and other inert materials. Many tons of this pay dirt must be refined to locate a meaningful amount of gold. Such is the quandary faced by the modern day data miners.

FACING A MOUNTAIN OF DATA
Modern businesses create an enormous amount of data. There are 509,147 data centers worldwide creating 1 petabyte of data every 15 seconds. It takes some very large data moving machines to process that amount of data. The OLAP system is responsible for moving and refining the data that drives analytical reporting. The storage solution of an OLAP system faces many challenges that come from both the massive amounts of data and the many different types of I/O patterns produced through the many stages of processing and reporting.

DATA SOURCES
Data comes in many forms and from many sources. Before it can be used by the OLAP engines, it must pass through several stages of processing and refinement (see Figure 2).

ETL
The extract, transform, and load (ETL) process takes raw input data and organizes it into structured formats and naming conventions. This is important because many applications that feed data into the data warehouse use terms that are unique to that application. The terminology and formats must be changed to reflect something the data warehouse understands. For example, the CRM system may have a field called OrderDate. To CRM that means the date the customer reserved product. To the supply chain management system, OrderDate means the date product was ordered from a wholesale supplier. During ETL, the data in the different OrderDate fields are moved into the equivalent data warehouse fields so downstream operations can work with all the data regardless of its origin.

MULTIDIMENSIONAL DATABASES
Databases configured for OLAP use a multidimensional data model allowing for complex analytical and ad hoc queries with a rapid execution time. These cubes organize data so that each cell within a multidimensional structure contains aggregated data related to elements along each of its dimensions. Figure 3 shows a cube that describes sales for ABC Produce, Inc. The cube has dimensions for Region, Quarterly Sales, and Product.
Optimizing OLAP Workloads with the QLogic 10000 Series Adapter

The front slice of the cube in Figure 3 represents the total sales of apples as shown in Figure 4. Using just the second tier of the first slice returns the sales of apples for the southern region. Using the second rank of the second tier of the first slice returns the sales of apples for the southern region in the second quarter.

OLAP OPERATIONS

OLAP tools enable users to analyze multidimensional data interactively from multiple perspectives. OLAP consists of three basic analytical operations: consolidation (roll-up), drill-down, and slicing and dicing.

- **Consolidation** involves the aggregation of data that can be accumulated and computed in one or more dimensions.
- **Drill-down** allows users to navigate from a higher aggregated level back down through the details. For instance, users viewing the sales by region can drill down into the individual products that make up that region’s sales.
- **Slicing and dicing** is a feature whereby users can take out (slicing) a specific set of data of the OLAP cube and view (dicing) the slices from different viewpoints. Data viewed from different angles gives a broader perspective of a problem unlike other models.

AGGREGATIONS

Complex queries are resource intense and can saturate the I/O capacity of storage architecture. OLAP cubes can return an answer in around 0.1% of the time required for the same query on a relational database. The most important mechanism in OLAP that allows it to achieve such performance is the use of aggregations. Aggregations are built from the data warehouse or other cubes by changing the granularity on specific dimensions and aggregating data along these dimensions. The number of possible aggregations is determined by every possible combination of dimension granularities. Theoretically, the combination of all possible aggregations contains the answers to every possible combination of dimension granularities. Unfortunately, provisioning enough resources to build a cube for every scenario is impractical.

VIEW SELECTION PROBLEM

The problem of deciding which aggregations (views) to calculate is known as the view selection problem. The objective of view selection is typically to minimize the average time to answer OLAP queries as well as minimize the time it takes to refresh data. Many companies are forced to forego access to useful aggregations due to limitations in their storage solutions. Others suffer through painful delays while on-demand requests for queries assemble data from less than optimal sources.

QLOGIC FABRICCACHE ACCELERATES OLAP REPORTING

QLogic FabricCache Adapters bring shared server-based caching to the OLAP storage solution, simplifying deployment and management while delivering scalable performance gains to I/O-hungry OLAP applications. The cubes that drive the OLAP solutions contain data in high demand. As this high priority data travels through the QLogic 10000 Series Adapter, it becomes captured by the high-speed SSD cache. Subsequent requests for the data are lightning quick because they no longer have to traverse the storage fabric and wait for a spinning disk (see Figure 5). Even complex queries and ad hoc requests see enormous performance gains when the QLogic FabricCache 10000 Series Adapter is added to the OLAP solution.

Figure 3. An OLAP Sales Cube for ABC Produce, Inc.

Figure 4. Selecting Specific Dimensions Returns Useful Information.

One dimension returns the total sales of apples.

Two dimensions returns the sales of apples in the southern region.

Three dimensions returns the sales of apples in the southern region during Q2.
Optimizing OLAP Workloads with the QLogic 10000 Series Adapter

The QLogic 10000 Series Adapters are transparent to the data warehouse, the OLAP cubes, and the OLAP reporting tier. The performance gains from this solution are available without having to modify any data structures or existing reports. The OLAP applications perform at levels never before realized, paving the way for more online cubes and aggregate structures. Much faster access to the data lifelines that drive the modern enterprise is now possible.

**SHARING CACHED CUBES ACROSS CLUSTER NODES**

The QLogic 10000 Series Adapters introduce an industry first: sharable cache. Unlike standard Host Bus Adapters, QLogic 10000 Series Adapters communicate with each other as both initiators and targets, using the Fibre Channel infrastructure. This feature allows the FabricCache cluster to share cached OLAP cubes across multiple server nodes (see Figure 6).

This breakthrough technology allows more OLAP cubes to be online. Each online OLAP cube is available in the hyper-accelerated state only possible through SSD caching. Bringing more accelerated cubes within reach of the OLAP report servers allows a whole new level of depth and detail. Deeper analysis uncovers richer business insights. Predictive analytics provide crystal clear forecasting. Enhanced executive guidance controls costs, unlocks revenue potential, and drives profits skyward.

**SUMMARY**

The modern OLAP infrastructure is a key resource of executive decision makers. Distilling ever-growing mountains of data into actionable information requires revolutions in storage technology. The QLogic FabricCache Adapters are a breakthrough in enterprise application acceleration, integrating transparent, server-based, shared caching into the SAN. With QLogic FabricCache Adapters, key aggregate data is brought closer to the OLAP reporting engine, optimizing query execution and accelerating report production. The shared cache capability brings even more OLAP data within reach, providing deeper insights into the state of the enterprise than was ever before possible.
Optimizing OLAP Workloads
with the QLogic 10000 Series Adapter