THE BUSINESS CASE FOR IN-MEMORY DATABASES

By Elliot King, PhD

Research Fellow,
Lattanze Center for Information Value
Loyola University Maryland

Abstract

Creating a true real-time enterprise has long been a goal for many organizations. The efficient use of appropriate enterprise information is always a central element of that vision. Enabling organizations to operate in real-time requires the ability to access data without delay and process transactions immediately and efficiently. In-memory databases, (IMDB) which offer much faster I/O than on-disk database technology deliver on the promise of real-time access to data. Case studies demonstrate the value of real-time access to data provided by in-memory database systems. Organizations are increasingly recognizing the value of incorporating real-time data access with appropriate applications.

In-memory databases, an established technology, have traditionally been used in telecommunications and financial applications. Now they are being successfully deployed in other applications. The overall increases in data volumes which can slow down on-disk database management systems have driven this shift. Additionally, increased computer processing power and main memory capacities have facilitated more ubiquitous in-memory databases which can either standalone or serve as a cache for on-disk databases—thus creating a hybrid infrastructure.

Introduction: The Real-Time Enterprise

For the last decade, the real-time enterprise has been a strategic objective for many organizations and has been the stimulus for significant investment in IT. Building a real-time enterprise entails implementing access to the most timely and up-to-date data, reducing or eliminating delays in transaction processing and accelerating decision-making at all levels of an organization. It requires streamlining information management in a way that provides a competitive advantage.

The excitement generated by the promise of building a real-time enterprise is easy to understand. Real-time access to data, information processing and process automation, and decision-making offers companies the prospect of a deep organizational transformation that will make them better able to compete in complex, global, always-open market places. In situation after situation, real-time capabilities offer specific, measurable benefits to organizations. Real-time information infrastructure can:

- Cut the costs of business transactions
- Automate critical processes
- Accelerate and improve forecasting and decision-making
- Innovate more quickly and with increased agility
- Improve market awareness and responses to changing conditions

In short, real-time access to data makes organizations smarter, more efficient and more nimble--qualities that are essential to success in today's economy.

But creating a real-time enterprise is not a trivial task. Perhaps the most significant barrier is cultural. Senior managers have to be able to envision how real-time applications will improve competitiveness and bolster the

bottom line. Sometimes, moving towards an effective real-time infrastructure requires reorganizing human resources and functions.

An additional barrier is budgetary. Organizations must commit to investing in the technology needed to achieve their goals. This investment can take the form of new hardware, software and networks that integrate infrastructure components. It can also take the form of developing the expertise necessary to implement efficient best practices.

Finally there can be a technical barrier. Different technologies are needed to implement real-time solutions for different business activities. Nonetheless, those technologies may have to be integrated to maximize the benefits of the investment. Technology selection is a critical element when implementing real-time solutions

In-Memory Databases

In-memory databases can be a significant technological component in the development of a real-time enterprise solution. Historically, the process to establish a real-time enterprise has been focused on improving the performance of many elements of the overall computer infrastructure. Unfortunately, performance improvements in storage systems have conspicuously lagged those in other areas such as processing power and memory. Storage system performance has emerged as a factor in low latency, high volume data applications.

In-memory database systems were developed in the 1990s to address those applications for which writing and retrieving data from a disk was not fast enough--particularly in applications for which the most critical function is high throughput. In-memory databases are fully relational, and provide standard SQL functionality, with all data under management being stored in the computer's main memory. These databases are also called main memory databases. This approach is faster than caching, used by some disk-based database systems. It is also a more effective solution than creating a RAM disk to manage data. With a RAM disk, regular caching and I/O operations continue even though they are redundant. In-memory databases require only a single data transfer.

When data is stored in-memory, the database architecture is optimized. Buffer pools and data structure management algorithms for minimizing the I/O used by disk-based systems are eliminated. The net result is that in-memory databases need fewer CPU instructions to execute requests. This leads to much faster response times.

For those unfamiliar with the technology, two concerns about in-memory database are generally raised—persistence of the data, or will the data be lost if there is a sudden power outage, and the size of the database under management. At least four approaches to the persistence issue are in use including periodically generating snapshots or checkpoint images of the data, transaction logging, backing up to non-volatile RAM and hybrid approaches with data written to disk as well as residing in memory.

As computers' main memory capacity have increased, benchmark studies have shown that in-memory databases can manage more than a terabyte of data without any apparent limitations to further scaling. The size of the database is limited only by the size of the memory. As the price of main memory has plunged, the limits on the amount of data in-memory databases can handle have essentially evaporated.

In short, in-memory databases are tuned for performance and throughput several times faster than standard disk-based database systems. They represent lower complexity, lower cost solutions that provide high performance and throughput for real-time applications with high data volume requirements along with data persistence, recovery and transactional capabilities.

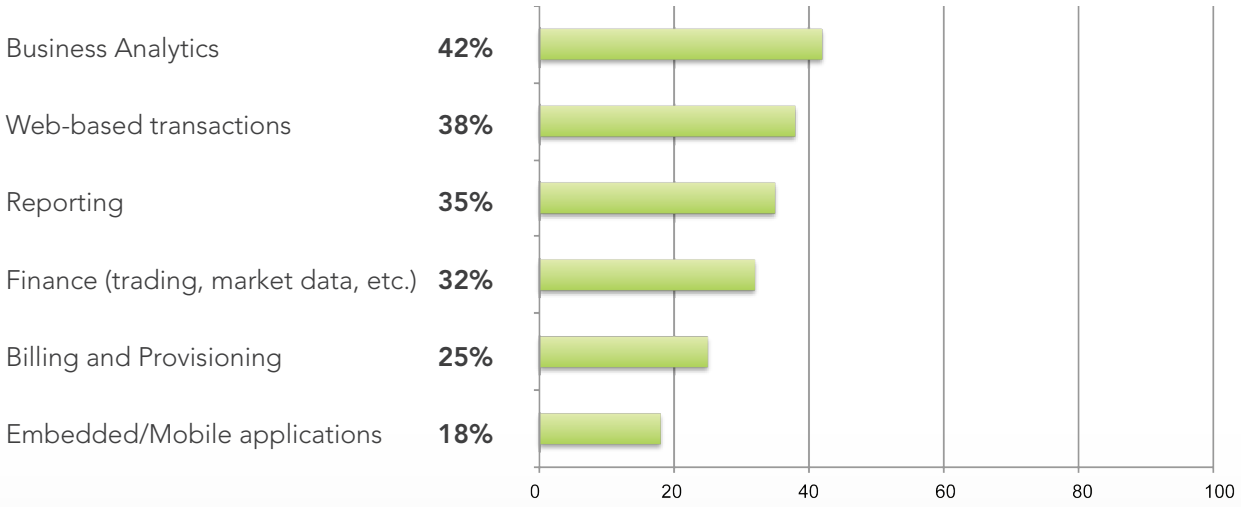
The New Hybrid Approach

Almost since they were first developed, in-memory databases have been used in conjunction with on-disk databases to create a hybrid database infrastructure that takes advantage of the faster execution speeds of in-memory databases. Most database vendors achieve this with an architectural concept known as a dual-engine DBMS by integrating their in-memory engine with a disk-based engine. However, there is a small number of database vendors that offer a true hybrid architecture in which both in-memory and on-disk objects are handled by a single engine. With true hybrid databases, the tables with large volumes of data reside on disk, and the smaller tables that are frequently accessed reside in RAM. A single engine is responsible for processing all data objects. Hybrid database management systems strike a balance between performance, cost, persistence and form factor.

Typical Applications

Since their introduction, in-memory database systems have typically been associated with applications in specific industries. However as Figure 1 shows, the use of in-memory databases is starting to broaden.

Figure 1: Most popular applications for in-memory databases.



(Check all that apply)

Business analytics and business intelligence are clearly growing as important applications for in-memory databases. Each of the most popular use cases demonstrates the technology's strengths.

Telecommunications—Quality of service and extremely low latency are of paramount importance to telecommunications providers. In fact, those are two of the primary aspects on which different service providers compete. With that in mind, telecommunications service providers need to execute many customer-facing operations in as close to real time as possible. These include account authentication, management of pre-paid calling plans including Web-access to call histories, number of minutes used and funds remaining, as well as a host of other functions. Indeed, in-memory databases have been used to expand network capacity by providing faster access to subscriber information and configuration settings.

One of the most interesting uses of in-memory databases in telecommunications is managing roaming operations. Roaming management is complex. The network must identify a caller; determine the home location, the visitor location and the billing information. With data loaded in memory, these processes can be executed efficiently and the caller experiences the expected quality and timeliness of service.

Financial Applications—In-memory databases have been used in a wide range of financial and trading applications. In most trading systems, trades have to meet compliance regulations and other rules before they can be executed. Compliance and reference data is read repeatedly during trade execution but compliance and reference data does not change frequently. Since how fast the data can be read has a direct impact on how quickly the trade can be executed, storing compliance and rules information in an in-memory database dramatically accelerates the entire trade process.

In-memory databases are also used in a wide range of other finance related applications. In recent years, particularly in the wake of the current financial crisis, the technology has been deployed as part of fraud detection and risk management applications. In a fraud detection application, in-memory databases have been used for authentication, authorization and transaction monitoring. In some implementations, transactions are reviewed for multiple types of fraud with no discernible impact on the speed of the transaction. For risk management, in-memory databases can be used to aggregate data from multiple systems to better ascertain total exposure to risk.

Business Analytics and Business Intelligence— the use of in-memory database technology in business analytics and business intelligence applications has been growing rapidly, in response to organizations' needs to make good decisions in real time based on the latest and most up-to-date information.

A wide range of business intelligence applications using in-memory databases are emerging. One example is an application in which field sales personnel have access to manufacturing information to insure that they don't promise more than what the operations team can deliver. An in-memory system can also be used by a sales person to calculate the impact of a proposed customer discount on profitability. And in-memory databases can be deployed as a part of a payment processing system to determine appropriate payment options based on historical and real-time data.

Web Application Support— another use of in-memory databases is use in support of existing applications, particularly Web-based applications. While some applications do not require the high performance that in-memory databases offer, an increasing number do.
For example, both Google and Yahoo reportedly keep their indices in memory. Obviously response time is critical for both.

Once enterprises can incorporate real –time data into their analytical infrastructure, the applications seem almost endless. For example, integrated into smart meters, in-memory databases could help forecast demand. An application based on in-memory database technology could insure that retailers never run out of inventory. And an in-memory approach would allow reports to be generated daily or even hourly without impacting the data warehouse.

Everything changes when data is moved off the disk. Data is available without a lag time to all levels of an organization. Not only do in-memory database systems allow companies to execute operations more efficiently, they also can enable as of yet unseen business practices and cost savings. And that is the true promise of the real-time enterprise.

Case Studies

Samsung Fire & Marine Insurance

Founded in 1952 and offering auto, injury, fire, marine, technology, overseas casualty, title, and long-term health insurance Samsung Fire & Marine Insurance is the largest casualty insurance company in Korea. The company competes in a newly deregulated market that allows foreign competition. Samsung itself has entered foreign markets including China, Europe, Indonesia, Japan and Vietnam. Moreover, the company has moved into the financial investment sector, offering asset management services, loans and other financial products.

As the company has expanded its operations and offerings it has had to comply with an ever-increasing set of regulations. All data transactions that are processed by the company must contain relevant historical information. Timely access to real-time data became essential. The company monitors more than 2.5 million transactions hourly. Delayed analysis of this transactional data creates risks, but timely access clearly could create new opportunities. The company's existing enterprise database system simply could not deliver the performance needed in the changing operational environment.

Samsung wanted an application offering a "real time dashboard" for senior management and an underlying system that could efficiently monitor and report on the high volume of transactions. It implemented the Altibase Hybrid Database (ALTIBASE HDB™) database management solution. ALTIBASE HDB is a true hybrid solution that combines an in-memory DBMS with a disk-based DBMS for processing large amounts of data in a single relational database. This structure allows the system to handle both the transient "hot data" and permanent "cold data" within a single instance. Cold data is stored on the disk resident resources, while in-memory resources handle the hot data.

The new monitoring system based on ALTIBASE HDB reduces costs and improves efficiency. Data is collected continuously and the hybrid on-disk and in-memory storage provides faster transaction speeds and analysis. The solution has reduced CPU utilization by 30 to 50 percent per second and delivers transaction speeds of 32,000 TPS (transactions per second), which made the real-time dashboard feature possible.

China Unicom

China Unicom is the world's third largest mobile communications provider. With roots as a wireless paging and GSM provider, it now offers 3G digital cellular services nationwide in China and a wide range of services within 31 provinces, municipalities and autonomous regions. These services include nationwide GSM mobile networks, long-distance, local calling, data communications, Internet services and IP telephony. Additionally it operates a mobile network in Macau and, as of October 2010, had over 163 million GSM subscribers, 99 million local access subscribers and 46 million broadband subscribers.

During the past decade, the demand for China Unicom's mobile communications and Internet services has grown dramatically. The increased demand had a significant impact on the company's existing billing infrastructure. Pre-paid and post-paid services had to be integrated and real-time processing for credit limit management was required. With a disk-based database management system, it could take hours to reconcile the available credit on a pre-paid calling card. Clearly this circumstance required a real-time processing solution.

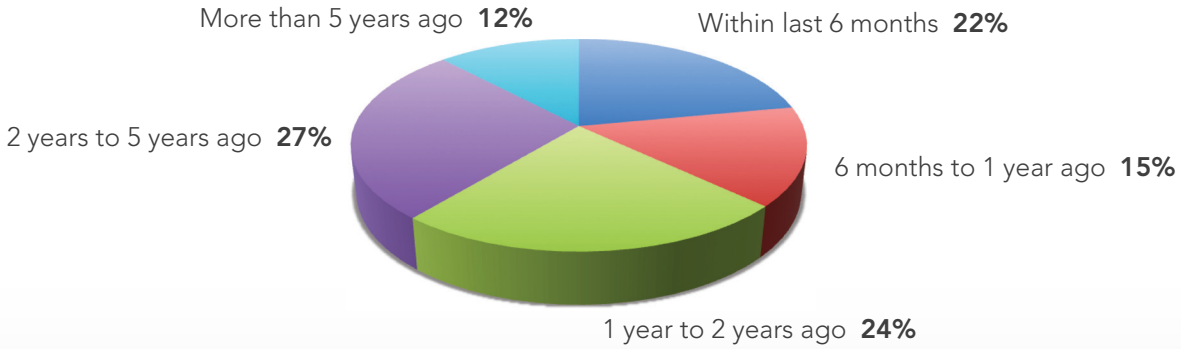
The ALTIBASE HDB with a highly available topology including primary Altibase servers augmented with an Active-Standby configuration was implemented. All operations related to billing are now processed in the Active servers while Standby servers are used as a backup in case the Active system becomes unavailable.
Billing reconciliation of pre-paid calling cards now takes minutes rather than the hours that it previously required.

As a result, the China Unicom billing system has experienced increased performance and unparalleled reliability processing greater volumes of data with increased system performance. The system now allows retrieval of real-time balances as well as real-time available credit inquiries. In addition, the average age for China Unicom subscriber accounts receivables has decreased considerably.

The Business Case for In-Memory Databases

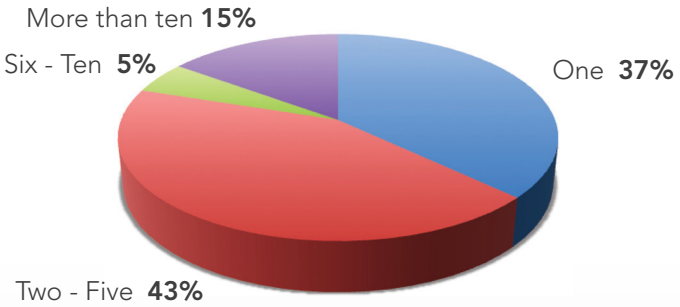
As figures 2, 3 and 4 show, the use of in-memory databases is clearly growing. The reasons for the growth are clear. Companies using the technology are installing additional databases for new applications. The amount of data under management is growing but the number of applications that incorporate in-memory databases is expanding as well.

Figure 2: When did you install your first in-memory database?



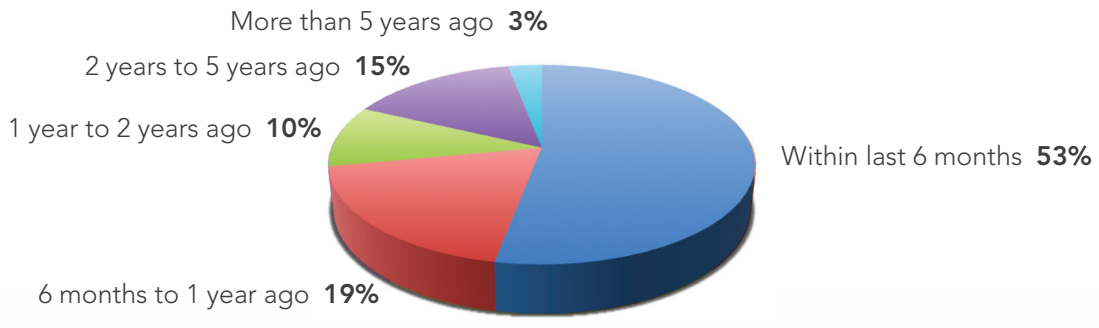
The increased use is being driven by three technical factors and one significant business imperative. The technical drivers are the spread of 64-bit computing, which makes more of the main memory addressable, multi-core servers, and most significantly, the rapidly falling price of RAM. One gigabyte of RAM once cost $150; now it costs $35 or less. Taken together, the economics of moving to in-memory databases for many different kinds of workloads has improved dramatically.

Figure 2: How many in-memory databases do you currently use?



As for the business imperative, the more companies can operate in real-time, the more efficient they become. And the more companies can improve transaction speeds, access data faster and get the answers they need without delay, the better than can compete in today's global marketplace.

Figure 4: When did you install your latest in-memory database?



Conclusion

Three key factors are driving the growth of the use of in-memory database technology. Companies need to be able to manage larger volumes of data, and do this more quickly. In-memory databases enable them to achive both at once. In addition, by accelerating execution times, in-memory databases offer a quantifiable return on investment. As a result, they are playing an increasingly important role in the drive towards the real-time enterprise.

The new generation of hybrid in-memory/on-disk databases offer serveral added advantages when implementing real time solutions. They solve the problem of data persistence without compromising execution speeds, delivering the major benefits of both technologies. In short, they can play a major role in the effort to create the real-time enterprise.

Altibase Technology

ALTIBASE HDB is a true hybrid relational DBMS that delivers extreme speed while supporting large data sets. ALTIBASE HDB reliably supports real-time applications and allows information managers to pick and choose between in-memory and on-disk data storage models.

ALTIBASE DSM is data event middleware that filters, analyzes and distributes high-volume data streams in real time. Users can register queries with ALTIBASE DSM to run continuously, so that when conditions are met, subscribers receive real-time notifications.