WiseClouds

Exploring the Database Forest

Robert Schneider – WiseClouds, LLC

Introduction

Today's database technology landscape is more dynamic and varied than ever before. What's driving all of this activity? First, enterprises are capturing, storing, and using immeasurably larger quantities of information, and they're retaining it for much longer, too. Secondly, there are many more data categories - such as video, images, and machine-generated metrics - in play now. Finally, there are bountiful new opportunities to take advantage of all this information.

Given all of this excitement and potential, it should come as no surprise that so many organizations are deploying an array of specialized information management technologies, rather than continuing to rely on a single relational database management (RDBMS) platform to meet all of their data-oriented needs. It's also no wonder that it can be confusing to determine what's the right tool for the job, and decide on the proper role for existing RDBMS infrastructure.

In this article - meant for information technology (IT) leaders and database administrators - I portray some of the most interesting trends that are shaking up the information management industry, along with the noticeable effects that these developments are having on IT. I begin by demonstrating that a 'one-size-fits-all' database platform isn't a good idea given today's complexities. Next, I describe several of the most intriguing new information management technology categories, followed by an overview of how integrated solutions such as SAP's Real Time Data Platform can provide a logical, unified approach. Finally, I point out that despite all of the fresh breakthroughs and noteworthy new solutions, transactional databases continue to remain at the heart of the enterprise's information processing responsibilities. This means that selecting an RDBMS will endure as a vital obligation.

Why a "One-Size-Fits-All" Database No Longer Makes Sense

Three independent but interrelated certainties present tremendous data management challenges - and opportunities - to every organization:

1. **There is more data than ever.** Gigabytes turn into terabytes, which then turn into petabytes. This information is being generated in support of transactional applications as well as business intelligence and analytics usage.
2. **There are more types of data.** Along with traditional relational records, enterprises now gather diverse categories such as images, video, text, and so on.

This information is frequently produced by Web, mobile, sensors, and so on, along with machine-to-machine streaming communication.

3. **There are more uses for all this data.** Users are clamoring for real-time applications that provide seamless, immediate integration with analytics, yet are massively scalable and deliver mission-critical OLTP performance.

Each of the above factors helps encourage organizations to employ fresh techniques for storing data and delivering the necessary processing power. Fortunately, there are now plentiful new offerings that help gather, manage, and extract benefits from all this information. These include specialized types of data repository along with solutions that magnify and extend the power of the relational databases that continue to serve as the primary information storage repositories for most enterprises.

It can be very puzzling to make sense of all these new information management offerings. Yet deadlines and responsibilities don't wait, which means that something must be done to promptly service user requests for access to the full spectrum of modern information. For many beleaguered IT organizations, the path of least resistance is to simply keep using existing RDBMS technology for every possible scenario. Unfortunately, this shortcut fails to recognize the unique needs imposed by these diverse information sources and highly focused applications. These requirements include:

- **Performance:** Relational platforms aren't necessarily optimized for the high throughput requirements of new applications and data types.
- **Targeted processing:** It's natural that effectively manipulating the often-enormous volumes of new data types will impose special processing demands.
- **Tooling:** Existing ancillary applications intended for relational databases may not be sufficient for visualizing, understanding, and managing new information categories.
- **Interfaces:** Relational databases have traditionally provided highly capable interfaces for accessing structured information. Unstructured and semi-structured data introduce additional prerequisites.
- **Monitoring:** Since these new applications and data types are often vital enterprise assets, IT leaders are demanding high caliber management and monitoring technology in support of these new solutions.
- **Storage Efficiency:** Relational databases typically have a high administrative overhead, expanding storage requirements to twice or even three times the original data size.

Forcing the RDBMS to perform chores that are outside its core capabilities is akin to jamming a square peg into a round hole. In contrast, leveraging the power of dedicated, best-of-breed offerings provides a much more elegant, logical outcome. However, it's critical to always remember two key facts that are pertinent no matter what technology has been chosen:

1. These diverse data platforms must be capable of working together: the whole is greater than the sum of the parts. I'll supply some examples of the value of this synergy later.
2. Relational databases still are the system of record for the transactional information that is so fundamental to the enterprise. This means that organizations should continue to be particularly selective and demanding when choosing an RDBMS.

Examples of New Information Management Technology

Startups and well-established technology suppliers alike are producing highly specialized solutions that help enterprises profit from the vast quantities of non-traditional information that are now ubiquitous. The upshot of all this innovation is that there has never been a broader selection of enterprise-grade information management offerings. In this next section, I briefly summarize some of the most prominent solutions.

Analytics platforms

As I illustrated earlier, most enterprises are seeking to capitalize on all of the data that they're capturing and maintaining. To make this possible, they're employing dedicated software to conduct real-time as well as delayed analysis. In some cases, these tools work on data contained in the relational database; in other cases, information must first be extracted and stored in a data warehouse or data mart. These analytics platforms provide powerful mechanisms for both technical and business users to help derive meaning from information. Columnar databases have proven to be especially popular, and they're converging with traditional row-based RDBMS to help satisfy analytical and online transaction processing (OLTP) imperatives.

In-memory databases

Accessing data from memory is generally at least ten times speedier than from disk. Thus, in-memory databases can deliver impressive performance enhancements for those applications that require the absolute fastest responsiveness without necessarily depending on long-term data persistence. Although some in-memory database products impose substantial barriers to easy deployment such as proprietary APIs, other offerings make it trivial to switch between in-memory and on-disk alternatives.

File/object storage

For a variety of technical and operational rationales, relational databases have generally been perceived as unsuitable for storing unstructured data and digital content. In response, a robust

ecosystem of vendors have created combined hardware and software solutions meant to hold and provide access to mammoth amounts of these information types, particularly:

- Videos
- Images
- Documents
- Scanned data

Key/value databases

These solutions are an excellent choice for storing data that doesn't require - or conform to - the table/column-oriented schemas imposed by a relational database. Key/value products are regularly used for capturing, storing, and querying fine-grained name/value pairs such as data from device monitoring, timestamps, metadata, and so on. There are many key/value offerings on the market today using a variety of storage approaches such as in-memory, disk-based, and so on. In addition, customers can opt for cloud-based as well as on-premise configurations.

Document stores

Data that can be expressed in consistent rows and columns are ideal for RDBMS. However, there are other types of information that are far less structured and frequently don't even adhere to a formalized schema. Document stores are great for these more ad-hoc information representations, which are often encoded in XML, JSON, and so on.

Graph databases

These increasingly popular technologies are designed to express bonds among a limitless set of elements. Through their structure and related toolsets, they let users and applications traverse these relationships very quickly and get the answers to some very complex queries. In fact, graph databases form the foundation of massively scalable social networks such as Facebook and LinkedIn.

Message queues

Although they're not necessarily considered data management platforms per se, message queues play a central role in some of the largest and most multifaceted information topologies. They are designed to foster communication among disparate processes and applications. To make this possible, they commonly implement both asynchronous and synchronous messaging, with copious delivery, queuing, routing, failover, and other configuration settings available for administrators and developers.

Tying It All Together

Every one of the specialized information management solutions that I just itemized introduces its own unique collection of challenges for IT. First, simply trying to juggle all of these technologies is a headache for even the most sophisticated, resource-rich IT organization. Each one has distinct performance, tooling, management, and monitoring profiles; each also tends to require very specific (and non-transferable) developer and administrator skills.

Fortunately for those enterprises that can surmount these hurdles, there are plentiful opportunities to synthesize traditional relational data with some of the new information categories that I've already noted. This commonly delivers fresh, innovative insights, especially when pairing structured and unstructured information. Let's look at three different examples of what can result when data is wisely blended:

- **Analyzing advertising performance.** In this scenario, daily sales figures are stored in a relational database. Meanwhile, click trails through a website are maintained in a key/value store. Tying these two silos together helps produce better awareness of advertising campaign efficacy.
- **Improving the e-commerce experience.** Inventory and stock levels are held in a relational database, while product images, instructional videos, and other documentation are maintained in file/object store. Linking these systems results in a better customer experience and higher revenue.
- **Ensuring regulatory compliance.** Financial transactions are maintained in relational database, while customer and employee interactions and relationships are tracked in a graph database. Unifying these systems results in enhanced compliance and reduced risk.

Regardless of the exact components in the overall architecture, ideally IT will employ an integrated solution to help conquer the challenges and deliver on the opportunities. Information will be presented in real-time, while administrative headaches will be soothed. Developers and users will have access to optimal tooling and interfaces.

Naturally, the major enterprise information management vendors have been eagerly observing all of this activity. All-inclusive new solutions are now reaching the market, with SAP's Real Time Data Platform (RTDP) serving as an example of one of the most comprehensive and well thought-out of these offerings. It recognizes that while the specialized information siloes that I've been portraying so far bestow a number of distinct processing rewards, they also increase cost and complexity because of redundancy, data proliferation, and other management overheads.

The SAP RTDP employs a unified, holistic strategy that pairs the speed and scalability of in-memory computing with a common framework that cleverly - and transparently - blends data

regardless of where it has been persisted: in the cloud, in memory, or on disk. Customers can elect to deploy the SAP RTDP one component at a time, or all at once.

SAP appreciated that disconnected information management toolsets and hand-cobbled processes simply won't suffice for enterprise-grade deployments. Thus, it designed the RTDP to amalgamate all data-related operations through mechanisms such as:

- Common information modeling
- Pictorial views of the enterprise's entire information architecture
- A unified metadata language
- Transparent integration among online transaction processing (OLTP), analytics, and complex event processing (CEP)
- A single administrative and monitoring environment

But it's vital to remember that even though integrated solutions will increasingly become important, relational databases will endure at the heart of the enterprise for many years to come. RDBMS platforms have been the most popular choice for managing enterprise information for 30 years for assorted reasons such as:

- The value and meaning of all this data is based on its relations, such as among customers, orders, inventory, and so on.
- Transactional integrity (ACID compliance) offered by RDBMS is crucial.
- There's an enormous installed base of applications and RDBMS developer expertise.

All of this means that IT organizations will need to balance the needs of new, non-traditional data applications while continuing to support – and improve – their existing relational infrastructure. For all of these reasons, properly evaluating relational databases will continue to be a crucial obligation. I'll discuss this important duty in an upcoming article.

Conclusion

Vastly bigger amounts of relational data, combined with unstructured, semi-structured, binary, and other contemporary data - all accessed by entirely new applications - are now an inescapable reality for the IT organization, and it's likely that these trends will continue to accelerate.

Although there are many interesting, highly focused technologies on the market that can help deduce knowledge from all of this new information, it's essential to recognize that the RDBMS will retain its role as the system of record for the most important enterprise data. This also means that identifying the right RDBMS will remain a vital responsibility for every enterprise.

About the Author

Robert D. Schneider is a Silicon Valley–based technology consultant and author. He has provided database optimization, distributed computing, and other technical expertise to a wide variety of enterprises in the financial, technology, and government sectors.

He has written six books and numerous articles on database technology and other complex topics such as cloud computing, Big Data, data analytics, and Service Oriented Architecture (SOA). He is a frequent organizer and presenter at technology industry events, worldwide. Robert blogs at http://rdschneider.com.