Data Management Challenges in Very Large Enterprises

Michael L. Brodie

Chief Scientist
Verizon Communications
USA
Michael.Brodie@verizon.com

1. Panelists[1]

- Pat Selinger, IBM Fellow and VP, Data Management Architecture and Technology, IBM
- James Hamilton, Architect, Microsoft SQL Server, Microsoft
- Adam Bosworth, VP, Engineering, BEA Systems Inc
- Hans-Peter Steiert, Research & Technology, DaimlerChrysler AG

2. Panel Outline

Very large enterprises have approximately a petabyte of operational data stored in over 1,000 data repositories supporting over 5,000 applications. Data storage volumes grow in excess of 50% annually. Repositories for decision support systems, which often contain replicated data, grow at twice to three times as fast as databases used for online transaction processing (OLTP). OLTP workloads are growing at over 60% per year. This growth is expected to continue for some time due to new Web-based systems, increased accesses to existing systems and the introduction of new sources of data, new workloads, and, new (e.g., XML-based) access requirements. While dealing with massive growth, large enterprises must also address unpredictable or elastic access demand of constantly evolving Web-based systems, increased storage complexity, new storage technologies (e.g., network data storage over IP, storage utilities), and more conventional but increasingly complex data and storage management challenges (e.g., backup, recovery).

While data management and data storage technologies continue to make impressive advances, there is only so much they can do in the face of the predicted growth rates. Very large enterprises are attempting to identify and address the drivers of data growth. A leading candidate is integration. Recent analyst studies conclude that over 40% of IT budgets are devoted to the integration of new and existing systems and databases. Technology advances often manifest in new systems and databases rather than in improvements and enhancements to existing systems. Consequently, very large enterprises operate their businesses with 1,000's of systems and databases ranging in age from 6 weeks to 30 years. Operational efficiencies require that these systems be integrated. The Web's potential of universal access adds increased urgency to these challenges. As a result, very large enterprises deal not only with massive data and workload growth and the attendant management activities, but also with massive integration challenges and costs.

Solution providers continue to offer significant advances to deal with specific data storage and data management problems (e.g., availability, robustness, performance) and are beginning to turn their attention to the integration challenge.

Current solutions tend not to map directly to the problems of very large enterprises. Solutions are seldom comprehensive and are product or vendor specific. Three approaches to address the problem of integrating component solutions into an enterprise solution are standards, consultants, and integrated product suites. Standards, such as those for Web Services, are intended to provide common specifications for all products in a domain so that different vendor products can be readily integrated. Consultants are intended to be vendor neutral while bringing a wealth of experience and knowledge to multi-vendor problems. A third approach is for vendors to integrate their products into tightly integrated product suites. Each approach has severe limitations. More than ever very large enterprises require solution providers to assist with their massive data management challenges.

The panel will identify the dominant data management challenges facing very large enterprises from the perspective of problem owners and will explore the solutions being offered by leading solution owners. It will discuss specific VLDB challenges and how the solutions address the challenges.

---

[1] Confirmed at publication time.