

Structured and Unstructured Data Retention Solution

Keith Richard RajECKi, Senior Director, Industry Solutions; Oracle Corporation; Redwood Shores, CA/USA

Abstract

Data Retention solutions are increasingly important as the world becomes more digital. These solutions help museums, national libraries, governments, educational institutions, and corporations world-wide capture and manage institutional assets. Data Retention provides a means of organizing institutional data while providing greater levels of access and controls.

Today's Data Retention solutions must be extensible to meet the evolving business needs of Big Data and Cloud Computing. While Hierarchical Storage Solutions and tiered storage provide the most reliable and cost effective solution, access and performance requirements are also important design factors. If architected incorrectly, data searching and retrieval can be time consuming and costly. Traditional tape-only archival methods simply do not meet the access requirements of many of today's repositories and long-term archives. Likewise, storing all the data on disk requires greater administration and is more costly. The proposed Data Retention solution provides a proven solution with a balance between disk, tape, and cloud storage to support long term archiving of structured and unstructured data.

Data Retention Challenges

Many organizations struggle with various aspects of data retention from policy and process to technology. Oftentimes these challenges create barriers to a successful data retention strategy. The Data Retention Solution for Structured and Unstructured Data was developed to eliminate the obstacles to successful data retention.

In order to develop this solution, I pulled together a team of cross-functional advisers consisting of librarians, archivists, administrators, developers, and technologists. We sought to minimize the points of complexities without minimizing their significance in the solution design. We evaluated many data types across many industries. We determined that at a very basic level, data is data. The policies, procedures, and technology may be different depending on the type of data, access and preservation requirements. We needed to develop a solution that provided the most flexible architecture and supported open standards for interoperability. This solution needed to be flexible enough to store very large amounts of very small files, small numbers of very large files, and every combination.

We also discovered that this solution needed to be easy to use, easy to set up, and easy to administer. As with any multi-component solution, management complexities can lead to data loss and increased costs. This solution integrates technology into an appliance in order to improve manageability of the technical infrastructure. The storage API's can be access from any application or Oracle database with the Data Retention Appliance providing the management of the data retention through configurable preservation policies.

Key Benefits

The Data Retention Solution provides a turn-key system to preserve nearly any data with high service levels and scalability. It provides the following major benefits:

- Higher service levels — The solution is designed to optimize service levels with redundant components and automated failover using storage virtualization and cluster technologies.
- Reduced cost — The ease of implementation and minimized administration combined with tiered storage capability reduce the cost of the solution environment.
- Faster time to delivery — Accelerated deployment of proven and tested configurations in an appliance.

It should be specially noted that each component of the solution can be individually upgraded or expanded to meet storage capacity or throughput requirements, while maintaining data security, integrity and protection against technology obsolescence. SAM-QFS enable storage capacity upgrades with minimal, if any service interruption but certainly without compromise to data integrity or security. The tape archive module is especially flexible, seamlessly and non-disruptively upgradable.

Performance

40+ PB total storage

Up to 4 PB of disk cache per file system

Large files sizes >5 GB/s

MBPS solution – sequential

Ex: HPC applications, medium to large file size data collection

Small (100KB files), ingest rates of up to 10,000 IOPS

Ex: Content Management applications

Conclusion

The solution development approach of working with a cross-functional team of advisers to establish the functional and technical requirements enabled us to develop a comprehensive solution to manage data integrity and authenticity.

The Data Retention Solution provides a near turn-key approach for any organization to manage and protect their data. The software and hardware components of this have been integrated and tested to reduce cost and bring data retention to the masses. This solution can support nearly any application through a standard set of storage API's. The appliance approach also simplifies administration with pre-tested components that significantly reduce upgrades. The appliance model enables scalability through additional nodes in the cluster. These additional nodes are organized into a system that supports the long-term viability of the stored data. The simple user interface enables distributed data administration to the data owner without requiring technical skills. Providing data owners with the ability to control the policy based distribution of data by applying the retention policy further supports fiscal responsibility and

sustainability. The incorporation of industry standard storage ensures the ongoing management, access, and security of the data retention solution.

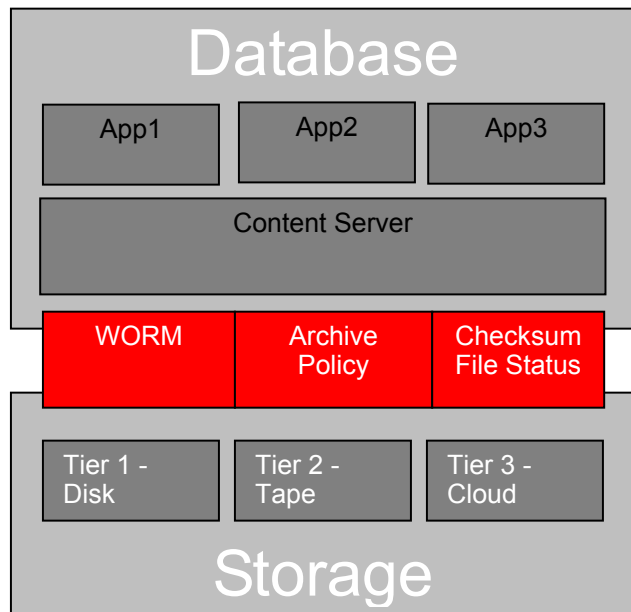


Figure 1. Logical Architecture

Author Biography

Keith Rajecki is Senior Director of Oracle's Industry Business Solutions for Global Education and Research, responsible for driving strategy and industry solutions across the organization. He leads the Global Education and Research industry solution development, and integrally works with customers and partners to drive strategic initiatives. With more than 18 years in the education and research industry, he brings deep global experience leading innovation in Education and Research industry solutions.