A Best Practice to Manage Application Data Growth for High Performance Application Environment and IT Infrastructure Cost Reduction
Introduction

Mission-critical ERP, CRM, or SCM application environments grow larger and more complex as they retain more historical business transactions for longer periods. Data growth is one of the biggest challenges for IT organizations because it degrades application performance, reduces application availability, and escalates IT infrastructure costs. As a result, it diminishes enterprises’ ability to close the books on time, gain real-time financial visibility, or provide a single, global view of all customer interactions.

This paper provides an overview of a new practice called Application Data Management (ADM). ADM has delivered proven and measurable results to Global 1000 organizations – up to 70% application performance improvement and several million dollars of annual savings in IT infrastructure costs.

In today’s economy, enterprises cannot continue to invest more and more in IT infrastructure to accommodate application data growth. Upgrading servers and adding storage only provide temporary relief. Continued data growth leads to more frequent, larger, and costlier upgrades. This “hardware treadmill” only postpones the inevitable: a permanent solution that addresses the root cause, unmanaged data growth, across all applications and databases.

Application Data Management (ADM) brings a new lifecycle approach to managing data growth. ADM is a combination of processes and technologies that manage business transactions from the moment the transactions are created to the time they are no longer needed.

Application Data Management:
- Provides the ability to define access, retention, and service level requirements for a particular business transaction at different points in its lifecycle.
- Enables enterprises to monitor and forecast application data growth,

---

META Group projects the annual application data growth rate at 125% through 2007 and up to 80% of this application data is inactive.¹

to establish, enforce, and refine data retention policies, to relocate historical transactions to online archives, and to create subsets of the production environment.

- Allows enterprises to leverage the latest innovations in server and storage technologies such as Linux on Intel and ATA to reduce costs while delivering consistent service levels and meeting complex requirements of the real time enterprise and compliance.

**The Problem: Unmanaged Data Growth**

**Key Contributors to Data Growth**

Application data growth is not simply a function of ongoing business activities. There are other strategic initiatives that significantly impact data growth.

- **Consolidation:** Consolidation of regional or departmental applications into a single global instance or integration of merged or acquired businesses results in a smaller number of larger application databases.
- **Compliance:** There are more than 10,000 regulations in the United States alone, some with retention periods as long as 30 years. Increasingly stringent data retention regulations such as Sarbanes-Oxley have caused maintaining data longer.
- **Application Upgrades:** Upgrades to new application releases such as Oracle 11i or PeopleSoft 8.x can add significant amounts of data due to enhanced data models to support additional capabilities.
- **Multiplier Effect:** Enterprises manage multiple copies of the production environment for development, testing, staging, and training purposes, requiring considerable additional storage.

![Figure 1: Multiplier Effect](image)

---

2 Peter A. Gerr, Brian Baineau, Patrick C. Gordon, “Compliance: The effect on information management and the storage industry,” ESG Impact Report, May 2003
Cost of Unmanaged Data Growth

Data growth has ripple effects on IT and throughout the enterprise. The most critical are:

- **Performance Degradation**: As data grows, applications and underlying databases must process more data, overloading database servers. Performance deteriorates. Users must work around performance problems often by limiting access.

- **Escalating Server and Storage Costs**: Although raw storage pricing is falling 30-40% per year\(^3\), the application data growth rate of 125% quickly outpaces the decreases in storage prices. Raw storage prices also do not accurately reflect the true cost of storage. As one CIO observes, “Each time I buy storage, it requires more software, backup, array and bandwidth” not to mention resources to plan, test, and tune the upgraded configuration and downtime during the upgrade.

- **Reduced Application Availability**: With more data to back up, restore, and replicate, maintenance windows increase, making it more challenging to meet 24x7 availability requirements. Another domino effect is longer downtime required for application upgrades.

- **Limited Ability to Support New Business Initiatives**: Because of the size of the production database, some IT organizations limit the number of development and test databases. The complexity of scheduling and managing these environments leads to missed project dates, undermining IT’s ability to support multiple business initiatives or deliver new application functionality on time and on budget.

---

3 Fred Moore, *Storage Manifesto* (Louisville, CO: Storage Technology Corporation, 2002), 8
4 Craig Stanley, “Through the Storage Looking Glass and What Measurement Found There,” Gartner PlanetStorage, June 2003
Application Data Management (ADM)

ADM is a combination of processes and technologies that manage business transactions from the moment the transactions are created to the time they are no longer needed. ADM provides the ability to define access, retention, and service level requirements for a particular business transaction at different points in its lifecycle.

Based on these requirements, ADM allows enterprises to set and enforce data retention policies to automatically relocate historical transactions out of the production environment into online archives. Unlike offline archives, online archives retain real-time access to relocated data, critical for business intelligence and compliance requirements.

This process improves performance of the production environment using existing server and storage infrastructure and avoids costly hardware upgrades. Less data in the production environment also means less data to be replicated, backed up or restored, thus reducing costs for disaster recovery (DR) or business continuance (BC) initiatives.

Figure 2: Changes in access, retention, and service level requirements throughout the transaction’s lifecycle

OuterBay
Fall 2003
ADM also enables a more cost-effective and streamlined development processes by creating fully functional subsets of the production environment for development and test projects. This dramatically reduces the costs associated with the multiplier effect and leads to more effective utilization of storage assets.

Furthermore, ADM provides the ability to monitor and forecast data growth in the entire application environment for effective capacity planning and to take advantage of the most cost-effective server and storage classes.

ADM includes technologies to monitor application resources, to model, establish, and enforce data retention policies, to relocate transactions to online archives, and to create relationally-intact subsets of the production environment.

**ADM Implementation Phases**

Implementing ADM includes the following phases:

1. Analysis and Planning
2. Implementation and Automation
3. Review and Refinement

**Analysis and Planning:**

The goal of this phase is to understand the current state of your application environment – data growth and breakdown of open (or active) vs. closed (or inactive) transactions – using Application Resource Management (ARM) technologies. The first step is to monitor current data allocation and analyze data growth by application, application module, storage class, and IT or business function. ARM tools can also forecast storage requirements and assess performance impact of continued data growth.

Next, examine the business, application, and legal requirements for different business transactions. These requirements help determine data retention policies – when transactions become eligible for online archiving, either the live archive or encapsulated archive.
For the development and test environment, determine which transactions are necessary for specific development and test purposes by creating *subsetting criteria*.

Also critical in this phase is to establish service level agreements (SLAs) for user community for access to active and inactive transactions. These SLAs are response time driven and will dictate the appropriate storage class for online archives. For example, the online archives can reside on an enterprise-class storage system, redeployed as a result of better storage utilization, or on a more cost-effective ATA system.

Finally, model the cost impact of adopting the data retention policies and subsetting criteria.

---

**Modeling Example (Initial Implementation):**

<table>
<thead>
<tr>
<th>Before ADM:</th>
<th>After ADM:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total storage consumption = 5.4 TB</td>
<td>Total storage consumption = 2.25 TB</td>
</tr>
<tr>
<td></td>
<td>54% reduction</td>
</tr>
<tr>
<td></td>
<td>Storage savings = $1,575,000</td>
</tr>
</tbody>
</table>

Production database\(^1\) = 300 GB

Number of clones to support production\(^1\) = 3
Number of clones for development and test\(^2\) = 10

Inactive data (estimate) = 50% of Production
Subset for dev & test (estimate) = 25% of Production

\(^1\)Note: this model is based on a TCO co GB that is $250 less per GB than Gartner TCO Report: “Through the Storage Looking Glass,” by Craig Stanley.

---

**Matching the Right Storage Class with the Right Data at the Right Time**

Until recently, storage classes for business-critical applications have been limited to enterprise-class storage systems such as EMC Symmetrix for high I/O performance and tapes for backup and archive.

With advances in storage technologies, more and more storage classes are available at different price points. Two such additions are low cost ATA-based storage systems and content addressed storage (CAS) specifically designed for fixed content.
**Implementation and Automation**
This phase implements data retention, archiving, and subsetting strategies defined in the previous phase. The goals for this phase are system setup, implementation of the data retention policies and subsetting criteria, and user validation. Enforce the data retention policies as well as create subsets of the production environment for development and testing, using a combination of ARM, database archiving, and database subsetting technologies. Ensure users have access both active and inactive transactions through the same application interfaces using SLAs as guides.

**Review and Refinement**
The goal of this phase is to review adequacy of user acceptance, performance, and budget achievement. Examine storage consumption in the production and development and test environments and performance levels. ARM tools with policy preview capabilities enable what-if analysis by modeling more aggressive data retention policies. Refine the data retention policies and subsetting criteria accordingly. Determine the next round of application modules or applications to apply ADM.

**ADM Example: ADM for Sales Order Transactions**

**Figure 3:** Relocation of sales order data from production to live archive to encapsulated archive
ADM Benefits

ADM provides a permanent solution for unmanaged data growth. It enables future-proof IT infrastructure by allowing continuous improvement and leveraging new innovations in server, database, and storage technologies. It has been adapted across industries with measurable results:

- **Application performance and stability:** By relocating inactive business transactions from the production environment, Sun Microsystems gained a 40% system-wide performance improvement.\(^5\)

- **Lower server and storage costs:** ADM extends the life of existing server and storage infrastructure investments, defers costly server or storage upgrades, and reduces costs for DR/BC initiatives. Applied Materials saved over $2 million in storage costs in the first year of implementing ADM and approximately $1 million thereafter.\(^6\)

- **Automated regulatory compliance:** By establishing and enforcing corporate-wide data retention policies for both online and long-term archives, Tektronix has met financial and legal reporting requirements in 27 countries.\(^7\)

- **Accelerated delivery of new initiatives:** ADM enables streamlining development and test environments to deliver development projects for new business initiatives or new application functionality on time and on budget.

More importantly, by providing high performance, stable application environments, ADM frees resources to enable enterprises to focus on business processes – closing the books on time, gaining real-time financial visibility, processing payroll more efficiently, improving the reliability of materials management, providing a centralized view of all customer interactions, etc.

New advances in server and database infrastructure, such as Linux on Intel and Oracle Real Application Clusters (RAC), will play an important role in ADM helping enterprises to further reduce costs while meeting rigorous demands of mission-critical applications. By hosting online archives and development instances on Linux, ADM enables enterprises to leverage innovations in Linux and commodity hardware and to migrate to more effective IT infrastructure.

---

5 OuterBay, “Consistent Performance: The Key to Competitive Business Advantage, A Case Study: Sun Microsystems,” Fall 2001


7 Lois Hughes, “27 Countries on a Single Instance! But What About Performance?” Gartner PlanetStorage, June 2003
OuterBay ADM 3.0 integrates technologies critical to managing application data growth. Its extensible, enterprise-wide architecture supports leading application packages, custom applications, databases, platforms, storage classes, and data types. The components of OuterBay ADM 3.0 include:

- **Application Resource Monitor™** analyzes and forecasts data growth as well as enables creating, modeling, and enforcing aggressive data retention policies.
- **Instance Generator™** creates smaller, relationally intact subsets of the production database.
- **LiveArchive™** identifies and relocates inactive data out of the production environment to a separate database or an online archive file (or encapsulated archive).
- **Developers Edition™** extends support to custom or customized applications and databases.
OuterBay ADM: A Proven Solution

OuterBay’s ADM solution has delivered immediate benefits to leading Global 1000 enterprises in various industries that include high technology, manufacturing, and financial services:

- Performance improvements - over 70% at Tektronix
- Decrease in production database sizes - up to 60% at Waterpik
- Reduction in development and test database sizes - up to 85%
- Storage cost savings - over $2 million in year one at Applied Materials
- Reduction in Oracle 11i upgrade downtime - over 40% at Parsons Brinkerhoff

These benefits extend to application environments of all sizes. Enterprises with databases as small as 40 GB implemented ADM as an integral part of managing new application environments. Other enterprises with databases as large as 1 TB implemented OuterBay’s solution because they were prompted by concerns on performance, storage costs, downtime caused by application upgrades, and daily operational difficulties.

Next Steps

- **Start now:** Data growth and associated performance and availability issues creep up over time. Whether you have just upgraded your storage infrastructure or believe that you have enough headroom to handle future data growth, it’s the right time to start planning. Take a proactive approach.
- **Start with one application:** Find an application module with a significant data growth rate. Implement ADM, monitor data growth and performance, and refine data retention policies. Then expand across other modules or applications.
- **Choose a proven solution:** The right solution must demonstrate that it optimizes efficiency (cost/risk) and effectiveness (service levels) with proven results. OuterBay brings a combination of proven and application vendor certified technology and well-developed implementation methodology. Call OuterBay today for an application analysis workshop.