SYMASE DYNAMIC ARCHIVE
MAXIMIZING THE PERFORMANCE OF YOUR OLTP SYSTEM
DYNAMIC ARCHIVE AGENDA

- Data Life Cycle
- Data Growth
- Sybase Vision of Dynamic Archive
- The Architecture
- Typical Resets
- Data Life Cycle Assessment
- Summary
- Outstanding Q & A
Sybase Dynamic Archive supports an environment of total data lifecycle management, enabling enterprises to reduce the size of application production databases – improving application performance, speeding up maintenance activities, and reducing costs associated with storage and maintainence.
DATA GROWS BY NATURE OF BUSINESS PROCESS

80% of Data is **INACTIVE**

125% Data Growth Fueled by Compliance, Upgrades

Infrastructure Costs Growing

<table>
<thead>
<tr>
<th>Customers</th>
<th>Products</th>
<th>Business</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transactions</th>
<th>Applications</th>
<th>Data</th>
<th>Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Diagram showing data growth and inactive data](Image)
Data proliferation and the need to retain data for extended periods of time is causing explosive growth in production databases, resulting in:

- Drawn-out, resource-intensive maintenance cycles impacting SLAs
- Sluggish application performance
- Runaway hardware and storage costs

<table>
<thead>
<tr>
<th>Complications</th>
<th>Repercussions</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Regulatory Compliance requirements for access to years’ worth of data</td>
<td>✓ Unsatisfactory Customer Service</td>
</tr>
<tr>
<td>✓ Need for duplicate copies of production database throughout the organization</td>
<td>✓ Increased Inefficiencies</td>
</tr>
<tr>
<td>➢ on average, 15 copies of production database are maintained</td>
<td>✓ Higher Cost</td>
</tr>
<tr>
<td>➢ for backup, recovery, reporting, IT initiatives, etc.</td>
<td>✓ Reduced Revenues</td>
</tr>
<tr>
<td></td>
<td>✓ Legal/Compliance Implications</td>
</tr>
</tbody>
</table>
Enables customers to intelligently and automatically, manage
business transactional data through its lifecycle – by offloading
inactive data from the production system, while keeping archived
data fully accessible online.

Benefits:

- **Maintenance & Performance**
  - Simplifies maintenance operations
  - Improves performance of the application by 30% to 80%

- **Costs**
  - Reduces hardware & storage expenditures

- **Compliance**
  - Enables compliance with regulatory requirements for long-
term storage, fast access and data integrity

- **Manage Lifecycle of Information**
  - Supports implementation of an environment of total
  information lifecycle management by archiving at business
  transaction level
HOW SYBASE DYNAMIC ARCHIVE WORKS

Application

Production

Phase I: Only ASE
Phase II: More Source Choices
HOW SYBASE DYNAMIC ARCHIVE WORKS

Application

Active
Production
Inactive

Step 1: **Identify** Inactive Data

Phase I: Only ASE
Phase II: More Source Choices
HOW SYBASE DYNAMIC ARCHIVE WORKS

Phase I: Only ASE
Phase II: More Source Choices

Step 1: Identify Inactive Data

Step 2: Relocate Inactive Data to Archive using Rep Server

Application

Active

Production

Inactive

Archive

Sybase ASE
HOW SYBASE DYNAMIC ARCHIVE WORKS

Step 1: Identify Inactive Data

Step 2: Relocate Inactive Data to Archive using Rep Server

Application

Phase I: Only ASE

Active

Production

Phase II: More Source Choices

Active

Sybase ASE

Archive

• Large amounts of data retained online
• Seamless access to archive data
• In initial release, archive is a superset containing both active and inactive data
• Can be deployed as reporting instance of the application DB
**HOW SYBASE DYNAMIC ARCHIVE WORKS**

**Step 1:** Identify Inactive Data
- Large amounts of data retained online
- Seamless access to archive data
- In initial release, archive is a superset containing both active and inactive data
- Can be deployed as a reporting instance of the application DB

**Step 2:** Relocate Inactive Data to Archive using Rep Server

**Step 3:** Retain user access without changing application interfaces
- Identical access to production or archive data

**Production**
- Phase I: Only ASE
- Phase II: More Source Choices

**Application**

**Active**

**Archive**

**Sybase ASE**
What does it mean to you, if your production OLTP database where 30 to 80 percent smaller?

What does it mean to business users?

What does it mean to hardware and system administrators?

What does it mean to your developers?
# TYPICAL RESULTS

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Typical Results</th>
</tr>
</thead>
</table>
| Improved Application Performance             | • Reduces production database by average of 40% (or 100GB for 500 GB database)  
• Increases business application performance by 30%–80%*  
• Increases database performance, due to less I/O, etc. |
| Decreased Time for Maintenance Operations    | Faster backup/restore, update statistics, DBCC (Database Consistency Checker) and rebuilding of indexes                                         |
| Storage Cost Savings                         | • Eliminates an average of 2.5TB of storage space across the enterprise*  
• Storage cost savings: $625K (estimated at $500/GB by Gartner)* |
| Hardware Cost Savings                        | • Reduces hardware expenditures previously needed for performance tuning  
• Hardware savings are typically 20% of storage savings                                           |
| Easier to Search Archive without Manual Processes | • Searches archive without changing application interfaces  
• Eliminates costly manual searches of archive (saves two weeks resource cost for each search) |
| Regulatory Compliance                        | Online archiving of business transaction data retains data integrity for compliance requirements  
Improves regulatory reporting by providing transparent access to production and archived data |

*Based on 250GB production database with 40% reduction in size & 25 copies of production database*
Sybase Dynamic Archive solution:

- **Automatically stores** the right data *(active or inactive)* in the right place *(production or archive)*, based on its business value.

- Reduce maintenance windows, lowers storage costs and improves application performance while still maintaining easy online access to both production and archived data.

Information Managed  ►  Information Unwired  ►  People Unwired
Questions?
Comments?
What else do you need to know?

For more information:
http://www.sybase.com/dynamicarchive
Dynamic Archive

A Deeper Look

Returning Your OLTP System to Optimal Performance By Solving the Data Explosion Problem

November 2004
Agenda

Archiving Overview
- Reasons for Archiving
- Types of Archiving
- Archiving Benefits, Challenges & Issues

Dynamic Archive Architectures
- Distributed Archive
- Physical Archive w/ Replication

Dynamic Archive Process
- Defining Archive Rules
- Managing the Runtime Archive Process
- Analyzing & Forecasting Archive Requirements

“Under the Covers”
- How it works (for real)
- Establishing Application Transparency
- Unarchiving Data

Dynamic Archive Product Roadmap
**META:** Data is growing at a rate of 125% per year, yet up to 80% of this data remains inactive in production systems, where it cripples performance.

**GARTNER:** In 2004, enterprises will manage 30 times more data than in 1999.

**BERKLEY:** 65% of the average IT budget is spent on data growth.
The Impact of Unmanaged Data Growth

Data growth impacts many areas:

**APPLICATION SUPPORT**
- Performance, scalability
- Increased backup window
- Existing resources tied up in managing current application growth - no resources for new initiatives
- Application obsolescence causes maintenance problems

**DATABASE MANAGEMENT**
- Performance, scalability
- Difficulty maintaining Service Level Agreements (i.e. completing nightly-maintenance tasks on time)
- Longer backups, upgrade times

**FUNCTIONAL MANAGEMENT**
- Slower reporting runtimes
- Increased critical process runtimes

**INFRASTRUCTURE PERSONNEL**
- Increased data center costs (hardware and storage)
- Use of alternate solutions increases access cost – like tape drives
- Difficult to support new initiatives
“Offline” Archives
- Tape, Historical Backups
- Chief Benefit: Easy
- Issues
  - Restoral Nightmare
  - No access

“Near” Line Archives
- HSM, CAS
  - File extraction based
- Chief Benefit:
  - Cheaper Storage
  - Restoring discrete data
- Issues:
  - Storage Tracking
  - Extremely limited access

“Online” Archives
- RDBMS
- Chief Benefit: Access
Let’s Talk CAS & Archiving

Content Access Storage (CAS)
- EMC Centera
- Use SATA w/ SCSI RAID
- Files are MD5 hashed
- Apps must use API
  - no hierarchical file system implementation

Archiving
- Great for end-of-life data!!
- Archive tool must support CAS vendor API

The problem
- Extracted data is put in XML file and stored.
- Querying then is a problem – application must be able to display XML
- If query spans XML files, all bets are off for performance
  - i.e. cross file joins, cross file aggregations, etc.
Archiving Benefits

- **Improved OLTP performance**
  - Reduced I/O’s due to smaller indexes/fewer levels
  - Reduced data cache usage
  - Reduced LRU-MRU promotions
  - Fewer row scans for queries

- **Reduced Maintenance**
  - Faster backups, upd stats and reorg
  - Faster schema mods/app changes
  - Faster recovery on reboot

- **Reduced Active Storage**
  - Inactive data can be moved to cheaper disks (SATA RAID)

- **Retain Data Access**
Archiving Challenges

**Archive Integrity**
- Avoiding orphaned data
- Identifying candidate business transactions
  - Rules
  - Retention Policy
  - Schema

**Archive Access**
- Accessing archived data
- Tracking archived data

**Archive Management**
- Maintaining archive rules
- Monitoring space projections
- Impact Analysis/Forecasting

OLTP -> Archive
Common Archive Issues

Retention Policy & Data Life Cycle
- When is data eligible for archiving?
- How long does data have to be online?
- What happens after the online archive time expires?

Data Recovery
- How do you bring data back online from an archive?
- How dynamic is the process?
  - Consider the scenario of archiving ‘old’ customers
    - Someone who hasn’t purchased anything in 3 years
  - Now, an ‘old’ customer submits a new order
    - If treated as a new customer, does this impact the business ability to determine success of attracting returning customers?
    - If treated as an old customer, does the data recovery happen fast enough to complete the current transaction?
    - Do app changes due to business rules now change other data or void changes attempting to make
      - Named account → Geographic Model
      - Signature limits changes due to business constraints.
Common Archive Issues

**Schema Changes**
- Reflecting the schema change in archived data
- What happens when offline data with a different schema is restored/accessed?
  - Which application versions are used?

**Business Structure Changes**
- The inevitable re-orgs (business, geo-political, etc.)...
  - How is archived data mapped to new structure?
- New licensing models
  - Do constraints need to change (license type?)
  - What happens to old values in lookup tables
Dynamic Archiving Overview

Developers
- Design/Modeling Tool
  - Application Transactions
  - Entity/Relation Mapping
  - Archival Rules, etc.

Users
- User Application

DBA’s
- Monitoring Tool

OLTP or Combined Access View

OLTP

Archival/Retrieval Engine

Archive
Distributed Archive

- Available Q1’05
- ASE → ASE
- Active Data in OLTP
- Inactive Data in Archive

Data Movement

- Controlled by Archive Process
- Periodic (Monthly, Quarterly)

Archived Data

- Transaction Details Only
- “Read-Only” Data or Data that is “Static” after Archival
- Result: Some infrequently accessed data must remain in OLTP
Archive process identifies candidate rows and related information for archival based on a configurable batch size. Rows identified are listed in control tables in OLTP database.

Archive process copies rows to archive database using proxy tables/array inserts in the archive database.

Archive process removes rows from OLTP system and clears control tables. *(Process repeats until all rows archived)*
Physical w/ Replication Archive

- **Physical w/ Replication**
  - Available Now
  - ASE → ASE
  - Active Data in OLTP
  - Active + Inactive Data in Archive

- **Data Movement**
  - Replication Server
  - Near-Real Time
  - Archive process only removes data from OLTP

- **Archived Data**
  - Transaction Details + Business Objects
  - Useful for reporting
Replicated Archive: How It Works

1. Archive process identifies candidate rows and related information for archival based on a configurable batch size. Rows identified are listed in control tables in OLTP database.

2. Archive process verifies rows are in the archive database (pre-populated by Replication Server).

3. Archive process removes rows from OLTP system and clears control tables. *(Process repeats until all rows archived)*
Advantages & Disadvantages

**Distributed**

**Advantages**
- Less Storage for Archive
- Simpler Maintenance
- Less Complex Implementation

**Disadvantages**
- May leave more inactive data in the OLTP system
- Reporting: Distributed Query Optimization
  - union-in-view
  - Result merging
- Longer Archive processing time due to data movement

**Physical/Replicated**

**Advantages**
- Useful for reporting
  - Offload OLTP
- Greater application transparency
- Reduced active/OLTP storage

**Disadvantages**
- More storage required for archive
- More complicated administration if not already replicating (WS)
- May require re-educating users that archive ≠ inaccessible
Help design an application package (Apps Pack aka business transaction)

1. Discover
2. Validate and
3. Generate

Key building blocks of a business transaction

1. Relationship (App Transaction Type)
2. Constraints
3. Retention Policy
Developer Edition provides several ways to discover data schemas and relationships including:

- Connecting to Source Schemas
- Extracting from Export files
- Import from DDL

Developer Edition registers all the required tables in the repository to enable further discovery for relationships.

- Rule discovery from integrity constraints, field names, and SQL code

View Models in Visio
Developers Edition provides several ways to validate that the relationships and data are correct including:

- Interactively checking the rules against live data
- Generate a batch program to validate rules against live data

Validate ATTs and Constraints

Validate Data in Tables is correct prior to archiving

**CHECK SUMMARY REPORT**

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PROCESSED</th>
<th>ERRORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORDERS</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>ORDER_INQUIRY</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>ORDER_ITEMS</td>
<td>665</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>774</td>
<td>1</td>
</tr>
</tbody>
</table>

SMARTCHECK chk001 COMPLETED ON 20030915 AT 08:18:08
Industrial strength automation / workflow engine that provides dynamic work discovery, parallel task execution and distributed deployment

- Support complete automation of large, complex 24x7 tasks
- High Performance
- Change tolerant:
  - Run time task discovery
  - Application code version checking
  - Data parity checking
  - Data parity correction
- Fault tolerant
  - Restart without intervention
  - Exception logging
- Parallel task execution
- Multi-level dependency framework
- Self auditing
- GUI Console
Runtime Components - Console
(Dynamic Archive Cycle using Replication Server Configuration)
Data growth monitor
- Detects - application data growth and program performance
- Diagnoses - business area growth, correlates application performance, and drills to related modules and tables
- Forecasts - application growth with policy management and previews of policy impact

Supports Multiple Data sources
- Data Collection Snapshots
- Automated growth statistics collection
- ATT Row Counts
- Manual data entry

Enterprise Dashboard
- Instance Drilldown reports
- Custom Appspack integrated with Developers Edition
Application Resource Monitor

- Analyze current data allocation and growth activity
- Forecast data growth across applications, databases, business processes, etc.
- Model the impact of time based data retention policies without affecting production