



Long Term Data Storage Database Archiving Functions

001101	10100	Craig S. Mullins
101011	010101	Mullins Consulting, Inc.
101010	010110	15 Coventry Court
11011001	1101	Sugar Land, TX 77479
01010110	100	
0010010	1100	
1011110	1101	

<http://www.craigsmullins.com>

INTELLIGENCE. INNOVATION. INTEGRITY



Author



This presentation was prepared by:

Jack Olson
CTO

NEON Enterprise Software, Inc.
11044 Research, Suite D300
Austin, TX 78730
Tel: 512-241-7335
E-mail: jack.olson@neonesoft.com

This document is protected under the copyright laws of the United States and other countries as an unpublished work. This document contains information that is proprietary and confidential to NEON Enterprise Software, which shall not be disclosed outside or duplicated, used, or disclosed in whole or in part for any purpose other than to evaluate NEON Enterprise Software products. Any use or disclosure in whole or in part of this information without the express written permission of NEON Enterprise Software is prohibited.
© 2006 NEON Enterprise Software (Unpublished). All rights reserved.



INTELLIGENCE. INNOVATION. INTEGRITY



Agenda



Emergence of Data Management Functions

The Long Term Data Storage Problem

Long Term Data Storage Solution Levels

Database Archiving Requirements



INTELLIGENCE. INNOVATION. INTEGRITY



Difference between DBA and DM



- Database Administration
 - Backup/Recovery
 - Disaster Recovery
 - Reorganization
 - Performance Monitoring
 - Application Call Level Tuning
 - Data Structure Tuning
 - Capacity Planning

Managing the database environment

- Data Management
 - Database Security
 - Data Privacy Protection
 - Data Quality Improvement
 - Data Quality Monitoring
 - Database Archiving
 - Data Extraction
 - Metadata Management

Managing the content and uses of data



INTELLIGENCE. INNOVATION. INTEGRITY



Database Administration Functions



- Very well defined tasks
- Very well defined Job Title and Description
- Overwhelming vendor support
- DBMS architectures fully supportive
- Functions fall entirely in IT
- Must be done well to support efficient operational environment



Mullins Consulting, Inc.
<http://www.craigsmullins.com>

INTELLIGENCE. INNOVATION. INTEGRITY



Database Management Functions



- Tasks definitions are emerging
- No standard Job Titles or Descriptions
- More aligned with business units than IT
- IT management has not been supportive (NMP)
- Executive management has not been supportive
- DBMS architectures built without consideration of DM
- Little Vendor Support
- Companies have accrued many penalties for not paying attention to DM requirements



Mullins Consulting, Inc.
<http://www.craigsmullins.com>

INTELLIGENCE. INNOVATION. INTEGRITY

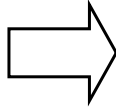


Emerging Data Management Drivers



Recent Regulations:

- Corporate Governance
- Data Privacy
- Data Retention
- Data Accuracy



Increasing Data Quality Costs



Increasing Data Volumes



Increasing uses/ users of data



More
Emphasis and
Spending on
Data Management
Functions



Significant
Tangible
Benefits

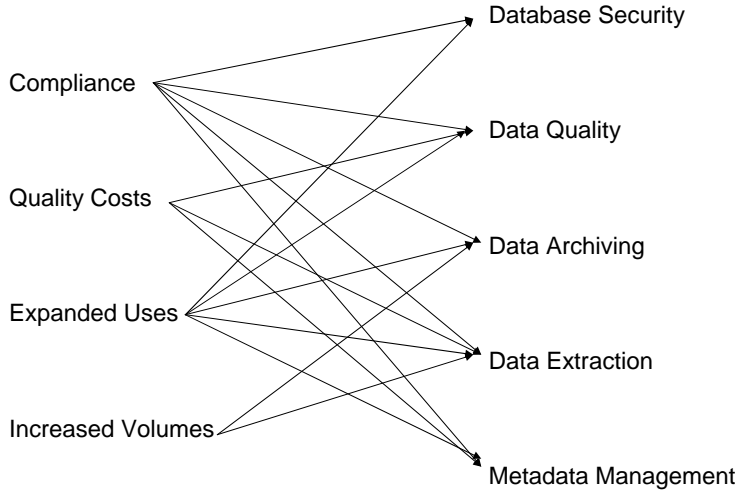


Mullins Consulting, Inc.
<http://www.craigsmullins.com>

INTELLIGENCE. INNOVATION. INTEGRITY



Drivers Impacts on Functions



Mullins Consulting, Inc.
<http://www.craigsmullins.com>

INTELLIGENCE. INNOVATION. INTEGRITY



Data Management Functions



- Database Security
 - Authorization Auditing
 - Access Auditing
 - Intrusion Detection
 - Replication Auditing
- Data Quality
 - Data Profiling
 - Data Quality Assessment
 - Data Quality Filtering
 - Data Profile Monitoring
- Data Archiving
 - Short term Reference Database
 - Long Term Database Archiving
- Data Extraction
 - Maintain privacy
 - Maintain Security
- Metadata Management
 - Complete Encapsulation
 - Change History Auditing



Long Term Database Archiving



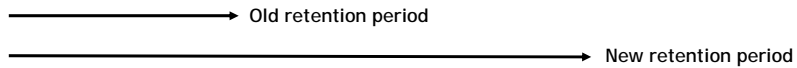
Why Is This Important



External Regulations
Internal needs for analytic applications

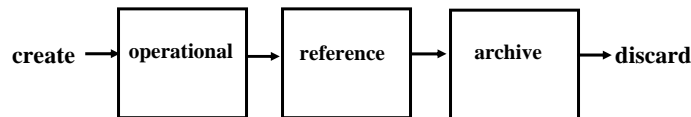


We need to keep more data: a lot more data
for more years: a lot more years
We need to preserve original content and meaning



INTELLIGENCE. INNOVATION. INTEGRITY

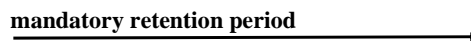
Data Retention Stages



needed for
completing
business
transactions

needed for
reporting
or
expected
queries

no expected
needs for
business
transactions
or reference



INTELLIGENCE. INNOVATION. INTEGRITY

What are Choices



- Keep Data in Operational Database
- Store Data in UNLOAD files
- Move Data to a Parallel Reference Database (RDB)
- Move Data to a Database Archive (DBA)



What's Wrong with Keeping in Op DB



- Too Much Data
 - Slows down everything
 - Transaction processing
 - Report generation
 - Extract routines
 - Recovery/disaster recovery
 - Requires frontline DASD for all data
 - May not fit
- (Major Problem 1)
 - Modifying old data for metadata changes
- (Major Problem 2)
 - How to handle old data for major application re-engineering
- (Major Problem 3)
 - Data susceptibility to unauthorized changes



Major Problem 1: Modifying Old Data For Metadata Changes



- Metadata changes are done often
 - Adding a field
 - Changing the length or precision of a field
 - Changing the encoding scheme within a field
 - Adding a new segment type/ table
- Since only 1 metadata definition applies to the operational database
 - Must modify old data to match new definition
 - Data stored may not be correct values for current context
- Accumulation of changes over the years is substantial
- At some point, the old data becomes unreliable in content since you cannot separate what is true from what is not



Major Problem 2: How to Handle Old Data For Application Re-engineering



- Major application re-engineering happens every so many years
 - Re-design and build application over
 - Move to a packaged application
 - Change host platform
 - Change DBMS platform
- Old data probably does not conveniently match new definitions
- Must either change old data to match new definitions
 - Lose authenticity
- Or, save old data separately for retention period
 - Requires preservation of old systems
 - Requires preservation of old applications
 - Requires preservation of old DBMS versions
 - Requires preservation of old metadata



Major Problem 3: Data Susceptibility to Unauthorized Changes



- Operational Systems allow many users to have access to data for update and delete
 - System administrators
 - Application stewards
 - Application data entry and update staff
- This leaves opportunities for
 - Unauthorized changes or deletes by corporate personnel
 - Changes or deletes by external mischief makers
 - Sabotage from external hackers
- The longer the data is kept in the operational systems, the more time it is exposed to these threats



How about data UNLOADs



- Need to have a place to bring it back to
 - Don't want in operational database
 - Can't access in unload format
- Need to search for data
 - Have no indexes or scope limiting parameters
- Need to manage storage media over time (bit rot)
- (Major Problem 1)
 - Modifying old data for metadata changes
- (Major Problem 2)
 - How to handle old data for major application re-engineering
- (Major Problem 3)
 - Data susceptibility to unauthorized changes



What's Wrong with Keeping in a Reference Database



- **Too Much Data**
 - Requires frontline DASD for all data
 - May not fit
- **Archiving activities (append) get slower and slower**
- **(Major Problem 1)**
 - Modifying old data for metadata changes
- **(Major Problem 2)**
 - How to handle old data for major application re-engineering
- **(Major Problem 3)**
 - Data susceptibility to unauthorized changes



Can I Use File Archiving Functions?



- **NO**
- **Data is not kept in databases in discreet files that contain only the records you need to archive now.**
- **Using file archiving retains dependencies on systems/ applications, and DBMSs for future retrieval.**



So, What's the Best Solution



Database Archiving

- A separate data store designed for long term data retention
- Functions for
 - Comprehensive Data Storage
 - System/DBMS/application independence
 - Metadata independence
 - Direct query access
 - Authenticity management
 - Continuous storage management
 - Archive Administration
- Data Archivists: Staff to design and manage database archiving applications



Database Archiving



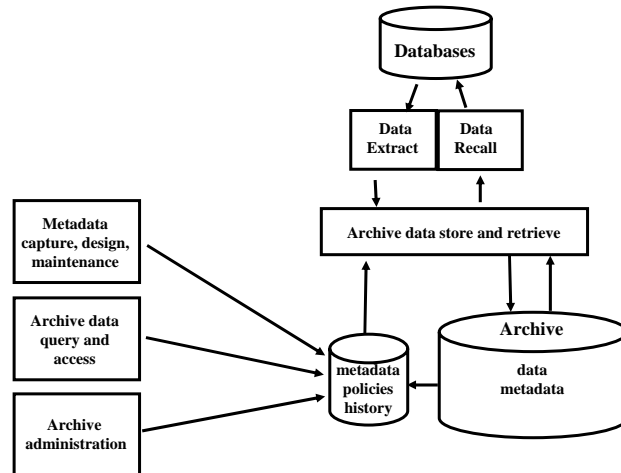
Database Archiving:

The process of removing selected data records from operational databases that are not expected to be referenced again and storing them in an archive database where they can be retrieved if needed.

Differs from Storage Archiving which handles files, not logical records.



Components of a Database Archiving Solution



Comprehensive Data Storage



- Encapsulation
 - Data
 - Metadata
- Unlimited Capacity
- Partitions that allow no update on previously stored data when adding new data
- Indexes for access
- Scoping indexes for partition selection
- Reliable representation of data

Independence from system\DBMS\application



- Ability to access data directly from the archive
- Ability to move to a DBMS different from one it was created on
- Implementation on multiple platforms
- Copy facilities for moving to newer archive platforms

What are the chances that you will have the matching systems, applications, and database systems when you need to look at the data many years from now.



Mullins Consulting, Inc.
<http://www.craigsmullins.com>

INTELLIGENCE. INNOVATION. INTEGRITY

Independence from metadata



- Enhanced metadata
 - Accurate
 - Data encoding explanations
 - Data semantics explanations
- Must be stored in the archive
 - No COPYBOOKS
 - XML in industry standard format
- Each partition has own metadata which can vary from previous partition
- Design to eliminate need for DBMS PROCEDURES
- Transformation of data to more standard forms
 - UNICODE
 - Universal dates

You must be able to access and understand the data from nothing more than the archive itself: data + archive metadata



Mullins Consulting, Inc.
<http://www.craigsmullins.com>

INTELLIGENCE. INNOVATION. INTEGRITY

Direct Query Access



- SQL like capabilities regardless of data source
 - Ad hoc queries
- Support of report generators
- Resolution of metadata differences between archive partitions
- Read minimum number of partitions to satisfy each access request



Authenticity Management



- No support of update or delete functions
- .. With exception of system controlled data discard function
- Use of encryption
- Use of checksums or equivalents to detect mischief
- Offsite backups for replacement of lost or damaged partitions
- Never modify data: no rollups to new metadata definitions
- Retain ability to produce original input bit-for-bit



Continuous Storage Management



- Pushdown of aged data to cheaper devices
- Multiple backups at multiple sites with registration
- Recopy of data to avoid media rot
- Recopy of media to avoid device or media obsolescence



Data Archivist on staff



- Full time job(s)
- Requires education in archiving principles
- Lots of work to be done
 - Collecting, validating, and improving metadata
 - Classifying data for archiving
 - Designing archive data structures
 - Developing archive processes
 - Storage capacity planning
 - Develop policies for ARCHIVE and DISCARD
 - Monitoring archive activities





Intelligent Solutions for Enterprise Data. Depend On It.

Archive Administration



- Control authorizations to archive functions
- Control security of datasets
- Maintain logs of all activities
- Maintain audit trails of all access and extract activities
- Monitor for need to perform storage management functions
- Integrate archive applications into application development change management system

