Evolving Solutions Disruptive Technology Series
Manage Data Growth through Database Archiving

Presenter – Dave Welch
Manager, World Wide Optim CoE Leader
Host - Michael Downs, Solution Architect, Evolving Solutions
Manage Data Growth through Database Archiving

InfoSphere Optim Solutions

Dave Welch
Manager, World Wide Optim CoE
Costs, performance and compliance continue to be a struggle for today’s organizations

Increasing Costs
Buying more storage is not a “cheap” fix when you add the operational burden

Poor Application Performance
Waning application performance due to growing data volumes

Manage Risk & Compliance
The “keep everything” strategy can impact disaster recovery and data retention & disposal compliance
Organizations have been increasingly challenged with successfully managing data growth

<table>
<thead>
<tr>
<th>Increasing Costs</th>
<th>Poor Application Performance</th>
<th>Manage Risk &amp; Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>3-10x</strong></td>
<td><strong>80%</strong></td>
<td><strong>50%</strong></td>
</tr>
<tr>
<td>Cost of managing storage over the cost to procure&lt;sup&gt;a&lt;/sup&gt;</td>
<td>The time DBA’s spend weekly on disk capacity issues&lt;sup&gt;c&lt;/sup&gt;</td>
<td>of firms retain structured data for 7+ years&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>$1.1 billion</strong></td>
<td><strong>250 hours</strong></td>
<td><strong>57%</strong></td>
</tr>
<tr>
<td>Amount organizations will have spent in 2011 on storage&lt;sup&gt;b&lt;/sup&gt;</td>
<td>The amount of time needed to run “daily” batch processes&lt;sup&gt;d&lt;/sup&gt;</td>
<td>of firms use Back-up for data retention needs&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

(a) Merv Adrian, IT Market Strategies, “Data Growth Challenges Demand Proactive Data Management”, November 2009
(c) Simple-Talk, “Managing Data Growth in SQL Server”, January 2010
(d) IBM Client Case Study: Toshiba TEC Europe; archiving reduced batch process time by 75%
(e) IDC Quick Poll Survey 2011, “Data Management for IT Optimization and Compliance”, November 2011
What is archiving?

Data Archiving is an intelligent process for **moving** inactive or infrequently accessed data that still has **value**, while providing the ability to **search and retrieve** the data.
Database archiving

For data retention, litigation hold and defensible disposal

- Ensure information is retained and archived based on value and duration of value, containing risk
- Ensure litigation hold requirements are upheld
- Ensure defensibility of the consistent disposal of data records based guidelines and regulations

- Control IT costs while supporting business needs
- Oversee IT management and budget, including storage costs associated with data growth
- Balance data management costs and performance, while ensuring data retention, litigation hold and defensible disposal needs are met
Archiving questions to consider

What data should I be saving, for how long and for what reasons?

How am I going to find the data when I need it?

How do I ensure I preserve data for e-discovery and audit needs?

What do I do with the data when I no longer need it?

What is the most appropriate solution to meet my archiving needs?

What is the cost/benefit analysis to support an archiving solution acquisition?
Effectively archive and manage data growth with InfoSphere Optim

**Reduce Costs**
- **Reduce** hardware, software, storage & maintenance costs of enterprise applications

**Improve Performance**
- **Improve** application performance & streamline back-ups and upgrades

**Minimize Risk**
- **Support** data retention regulations & safely retire legacy/redundant applications

- Intelligently archive data to improve application performance and support data retention
- Discover & identify data record types to archive across heterogeneous environments
- Capture & store historical data in its original business context to maintain referential integrity
- Support data retention policies automatically & consistently across the enterprise
- Ensure application-independent access of archived data via multiple access methods
- Support for custom & packaged ERP applications and data warehouses in heterogeneous environments
You can’t govern what you don’t understand

- Complex data relationships within and across sources
- Historical and reference data for archiving
- Test data needed to satisfy test cases
- Sensitive data identification
Discover & define business objects across heterogeneous databases & applications

Referentially-intact subsets of data across related tables & applications, including metadata.

Overall historical “snapshot” of business activity, representing an application data record – e.g. payment, invoice, customer

Federated access to related business objects across the enterprise

CRM on Oracle database

Custom Inventory Mgmt on DB2

ERP / Financials on DB2
Discovery

Accelerate project deployment by automating discovery of your distributed data landscape

Requirements

- Define business objects for archival and test data applications
- Discover data transformation rules and heterogeneous relationships
- Identify hidden sensitive data for privacy

Benefits

- Automation of manual activities accelerates time to value
- Business insight into data relationships reduces project risk
- Provides consistency across information agenda projects
A detailed look of the complete business object
Archive a complete business object and retain integrity

<table>
<thead>
<tr>
<th>Customer Number</th>
<th>Customer Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>8675309</td>
<td>John Smith</td>
</tr>
<tr>
<td>5025202</td>
<td>Jane Jones</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Order Number</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>50505</td>
<td>Product A</td>
</tr>
<tr>
<td>306959</td>
<td>Product B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Customer Number</th>
<th>Order Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>8675309</td>
<td>$1,056</td>
</tr>
<tr>
<td>5025202</td>
<td>$5,690</td>
</tr>
</tbody>
</table>

**Production Databases**

**Archive Files**
Leverage cost-effective storage alternatives

Current Data
1-2 years

Active Historical
3-4 years

Online Archive
5-6 years

Offline Archive
7+ years

Production Database

Archive Reporting Database

Compressed Archives

Non DBMS Retention Platform
ATA File Server
EMC Centera
IBM RS550
HDS

Offline Retention Platform
CD
Tape
Optical
Manage archived data as part of a big data strategy

Derive maximum value from archived data

Social Media
Website
ERP
CRM
InfoSphere Optim Archive

Big data platform
Provide application-independent access to the data

Third-Party Report Tools

InfoSphere Data Explorer

InfoSphere BigInsights

Application

ODBC/JDBC

XML

Ensure access & restore capabilities based on functional user and business process requirements.
Perform ad-hoc searches across InfoSphere Optim archive files

- Use web-based search engine powered by InfoSphere Data Explorer
- Search structured data with simple keywords
- Quickly identify where archive data is located
Leverage a variety of methods to access archive data

View archive records via the native enterprise application

Leverage third-party reporting tools, such as IBM Cognos

As well as analytic solutions such as InfoSphere BigInsights
Monitor and audit data archive access

Real time monitoring and protection with InfoSphere Optim & InfoSphere Guardium

- Production
  - Historical
  - Current
  - Monitor
  - Archive
  - Retrieve

- Data Archives
  - Reference Data
  - Historical Data

Universal Access to Application Data

- Application
- ODBC / JDBC
- XML
- Report Writer
Conclusion

Data Growth Management reduces hardware, storage & maintenance costs of enterprise applications.

Data Growth Management improves application and data warehouse performance, streamlining back-ups and upgrades.

Support data retention regulations & safely retire legacy/redundant applications by intelligently archiving historical data.
Thank you
Presentation Download
http://www.evolvingsol.com/events-and-education

Webinar Replay available on the Evolving Solutions YouTube Channel

Evolving Solutions
Michael Downs, Solution Architect
michael.d@evolvingsol.com
612-805-5579
Twitter: @emeldi

IBM Optim
Dave Welch
Manager, World Wide Optim CoE Leader
dewelch@us.ibm.com