**Why enterprise data archiving is critical in a changing landscape**

**Ovum white paper for Informatica**

**SUMMARY**

**Catalyst**

The most successful enterprises manage data as a strategic asset. They have complete visibility across operations and interactions with customers, business partners, and other stakeholders, and they offer complete transparency to satisfy auditors and regulators. Managing data strategically involves, not only having full access to online data, but also in-context views of historical data that must be available when needed. Yet, the silo'ed archival practices common across most enterprises yields only a partial picture at best and no information at worst. As organizations implement version upgrades to their application portfolio, modify database schema to support emerging business activities, or evolve IT infrastructure to handle new patterns of load and data usage, the ability to deploy an application archival strategy that retains appropriate levels of access to the archive data falls in jeopardy if the wrong tool is chosen for the job. For data-driven organizations, archival processes can no longer be taken for granted.

**Ovum view**

For most organizations, data archiving has become a common point solution for email and for selected, strategic enterprise applications. With archival processes common for helping organizations deprecate non-current data to reduce cost and processing overhead for online systems, numerous paths for archiving data for long term storage have emerged. Silo'ed archival approaches are jeopardized by ongoing changes to the IT environment and the reality that data and applications are growing more intertwined. Today, no single enterprise application contains a critical mass of "the truth." For instance, customer records are likely to transcend the ERP, CRM, email, content management, and e-commerce systems. For data-driven organizations, application data archival should be a core building block of an enterprise information lifecycle management and data integration strategy especially in heterogeneous application IT landscapes.
Key messages

- Today’s most successful enterprises are managing data as strategic asset.
- Enterprises are storing more data and more kinds of data than ever about their products, customers, and operations across heterogeneous applications.
- Data, applications, platforms, and deployment options are proliferating – and becoming more intertwined across most enterprises.
- Archiving has traditionally been implemented as point solutions for email, databases, and applications, respectively.
- Because of the growing interdependency of data across different applications and platforms, an enterprise archiving approach is becoming increasingly essential.
- Enterprise archiving requires the right tools for the job.

ARCHIVING IS MORE ESSENTIAL THAN EVER

Data in context is gold

The most successful enterprises are those that view data as a strategic asset, and realize that the value of integrated data is greater than the sum of its parts. They understand that customer information is more valuable when the data from the CRM system is correlated with other customer interactions that occur when they request content from the website, correspond through email, or make relevant comments on social networks. The same applies to operations that span multiple systems, from B2B trading networks to warehouse systems, logistics providers, and so on. Having a complete, in context picture is even more critical when organizations must conduct e-discovery for interactions that cross multiple system domains – including production and archived data.

IT environments are moving targets – and so is data

Unchecked data growth exacts a toll. According to a 2011 Ovum survey, 85% of respondents cited ballooning data sets as the cause of application performance problems.

Not surprisingly, the IT environment in most organizations has become a moving target. Data models evolve to support new business ventures, scheduled version upgrades of enterprise applications, or new regulatory mandates. Many are looking to cloud deployment for pilots, new ventures, or as backup support for handling planned or unplanned traffic spikes; others are looking to virtualization to improve infrastructure utilization. As data volumes expand, organizations revisit the underlying architectures of servers, database platforms, networks, and operating systems to ward off potential performance bottlenecks.
Application Data Archiving can no longer be taken for granted

Most organizations look to some form of archiving to keep data and operating costs from overwhelming their online systems. Archiving is intended to preserve enterprise application data in context for long periods, while backups only provide periodic snapshots of the data that was maintained on a specific date. As data, platforms, and deployment options proliferate, enterprises can no longer take archiving for granted. Archiving can provide the critical missing link that reconciles past history with the present, and for providing the necessary context for satisfying increasingly stringent compliance mandates for data retention.

In most enterprises, application data archiving is conducted in piecemeal fashion, either through point tools or through various means such as periodic backups or image snapshots. Few organizations understand the costs and disconnects of conducting archiving on an ad hoc basis. As data platforms, applications – and deployment options – proliferate and become more intertwined, organizations that value data must take an enterprise approach to archiving that can be applied across business applications.

**CHOOSE THE RIGHT TOOL FOR THE JOB**

Point tools present partial pictures, at best

There is no lack of tooling for retaining historical data; there are many approaches that are driven by applications, data integration, and infrastructure management providers. Examples include:

- **Application-specific tools** – There are numerous point tools that archive individual applications, with email and enterprise applications being the most common targets. Their chief drawback is that they only work for that specific application and cannot be applied across multiple applications or systems in a data center. Additionally, many do not provide optimizations that can reduce the footprint of an archive, thereby missing out on potential savings in application licensing and infrastructure costs. In the case of tools provided by ISVs of packaged applications, the archiving tools make the data no longer accessible from the native application requiring the need to reload or restore the archive data into production systems in order to view or report on the data.

- **Extract/Load.Transform tools (ETL)** – These tools are used for physically migrating and transforming data between sources and targets; they are not designed to accommodate business rules to satisfy archive eligibility and cannot support the complexities of business entity and application data models. ETL tools are also not designed to provide native application access to archive data – this would require significant development and custom integration code.

- **Backup/Snapshot tools** – As their names imply, backups and snapshots make copies of a static copy of the data or the image of a platform (as constructed through log files) of a single point in time. The only metadata carried by these tools is date, time, and origin. They do not reflect changes to data or underlying environments that occurred
subsequent to the backup. Backups are not accessible to the business user unless restored in an environment that was like production at the time the backup was taken.

- Enterprise content management tools (ECM) – Some ECM systems integrate with archive storage platforms, but they are restricted to unstructured content like documents and images and lack optimized storage and access required for structured or relational application data.

- Hierarchical storage management (HSM) – These tools provide at best a coarse picture of data, moving it based on arbitrary date or lineage criteria. More importantly, they work only at the OS or file system tier, not the application tier, and they are incompatible with databases that store their data artifacts in tables rather than individual files.

None of these point approaches offer purpose-built workflows designed specifically to archive complex business application data while retaining native application access or similar reporting views for business users. Aside for some specific ECM features, none contain the metadata that is essential for several purposes: making the data searchable, and representing the interrelationships of data to source systems, system configurations, or policies and rules that govern access. Furthermore, any changes to the current environment, such as OS version, virtual machine configuration, or database schema, can make these data stores created by alternative approaches to application data archive unreadable.

**Enterprise Archival – the solution for data-driven enterprises**

For organizations that are serious about competing with better data, archiving takes its place alongside data quality, data transformation, and data migration as an enterprise strategy that can improve system performance while lowering costs. Purpose-built solutions designed for application archiving offer online access to archive data via native application user interfaces, integrated retention management and enforcement, as well as extreme scalability and performance as data volumes eligible for archiving continue to increase.

Enterprise data archiving solutions offer a metadata-driven approach that avoids the inconsistencies of point solutions and preserves the context of data, even as the underlying IT environment changes. The metadata can carry any artifact, parameter, descriptor, or rule about the data such as:

- Data models or table constraints of the underlying data store;
- Specifications for reading and writing to/from proprietary data or table formats;
- Relationships of data to other data and/or applications;
- Relationships between source tables and business entities;
- Business rules governing how data is represented or accessed; and
- Handling and revision status of relevant attachments.
Furthermore, an enterprise archival solution can automate key tasks in the data lifecycle, such as automated purging of online stores after archival; maintaining audit trails and appropriate levels of end user access; enforcing retention and disposal policies; and managing legal holds. A comparison between enterprise archiving and other data retention approaches is shown in Table 1.

**OVUM ANALYSIS: INFORMATICA’S DATA ARCHIVE SOLUTION**

Having positioned itself as an independent data integration solution provider, Informatica has adopted a platform approach to providing solutions for integration across the lifecycle of data. Starting from its ETL roots, Informatica has steadily expanded its platform to encompass data quality, B2B data integration, cloud integration, master data management, ultra-low latency.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Enterprise Archive</th>
<th>Application-specific backup/archive</th>
<th>ETL</th>
<th>Backup/Snapshot</th>
<th>ECM</th>
<th>HSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managed, multi-application long-term data retention</td>
<td>Archives specific apps</td>
<td>Transform &amp; migrate data</td>
<td>Static copy of data at specific date/time</td>
<td>Access &amp; retention of documents or web content</td>
<td>Manages data retention based on utilization rate</td>
<td></td>
</tr>
<tr>
<td>Native Format</td>
<td>Yes</td>
<td>Partial support</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preserves all key metadata</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Multi-app/database support</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Optimize footprint</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Ensure access when environment changes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Access w/ out need to restore</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Rule-driven access workflow</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

Key:

- Yes
- Partial support
- No

Source: Ovum
messaging and event processing. With Information Lifecycle Management, Informatica has extended the coverage of its data integration platform to cover the management of structured and unstructured data under long-term storage.

Like the rest of its data integration solutions, Informatica has adopted a platform approach to archival, providing a metadata layer that is the key to retaining data in context—and to making the archive process consistent, repeatable, and traceable while maintaining complete access to the archive data through native application interfaces, through standard reporting APIs, and via eDiscovery tools. Informatica Data Archive provides tools that:

- Offer a common, centralized platform to manage archiving policies across applications in a heterogeneous data center when data resides either on premises, in the cloud, or both.
- Manage the key parameters for archiving and maintaining appropriate levels of access to the data, covering the underlying platform, associated business rules, and interrelationships with other applications and data sources;
- Automate partitioning to improve application performance while reducing the cost of archival, and purging of archived data from online data stores;
- Maintain native application access to archive data for business users eliminating the need to restore data from the archive;
- Enforce policy-driven archival practices for supporting compliance; and
- Analyze application performance, data growth and usage trends for supporting planning of archival strategy.
- Enable access to archived data without the need for restoration onto target environments that may have subsequently changed.

Informatica Data Archive is targeted for organizations with critical data spread across multiple sources and platforms.

**RECOMMENDATIONS FOR ENTERPRISES**

When planning a strategy for enterprise archival, ask the following questions:

- What is the rate of growth of online data across your organization? How visible is the rate of growth?
- Does your organization have a reliable tool for tracking growth and usage of data?
- Does your organization have adequate control of the costs of maintaining online data - including platform and software licensing costs?
- Does your organization have adequate control over the costs of migrating, archiving and managing data in long-term storage?
Does your organization rely on multiple tools for archiving data to long-term storage? If yes, how are retention policies enforced across multiple applications?

Is mission-critical data “owned” by multiple applications and business users? Does it encompass structured and unstructured data?

Is your organization subject to stringent data retention policies? How well has it been able to demonstrate compliance?

Do business users require access to aged data for longer periods of time? Is it important for them to be able to view archive data from native applications or via standard reporting interfaces?

Has your legal department requested eDiscovery of application data? If yes, was data easily available and accessible during the search phase?

Author

Tony Baer, Principal Analyst, Ovum IT Enterprise Software

Ovum Consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Ovum’s consulting team may be able to help you. For more information about Ovum’s consulting capabilities, please contact us directly at consulting@ovum.com.

Disclaimer

All Rights Reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the publisher, Ovum (an Informa business).

The facts of this report are believed to be correct at the time of publication but cannot be guaranteed. Please note that the findings, conclusions, and recommendations that Ovum delivers will be based on information gathered in good faith from both primary and secondary sources, whose accuracy we are not always in a position to guarantee. As such Ovum can accept no liability whatever for actions taken based on any information that may subsequently prove to be incorrect.