

Archiving in the Cloud: What End Users Should Consider

Gartner RAS Core Research Note G00170510, Sheila Childs, Adam W. Couture, 4 September 2009, RA6 06312010

E-mail archiving has been offered as a hosted service for years. Many archiving service providers are promoting their offerings as cloud-computing applications and, although in some cases, things haven't changed beyond how products are marketed, organizations should expect to see updated solutions that take advantage of new cloud infrastructures to deliver services that begin to meet the definition of "cloud."

Key Findings

- The initial and ongoing capital and operational expenses associated with archiving can be significant. Cloud computing's promises of quick deployment using outsourced resources on a pay-for-use basis make it attractive for archiving.
- Archiving applications have been available as hosted service offerings for more than a decade. True cloud-computing archiving solutions deliver improved user access at lower costs by leveraging new Web technologies and storage architectures.
- Archiving solutions are increasingly being offered as part of a portfolio of services that can include e-mail security, high availability and e-mail hosting.

Recommendations

- Do not get caught up in the hype of cloud computing. Recognize that many so-called "cloud-archiving solutions" have been available for years as hosted services.
- When considering an archiving service, do an in-depth cost analysis comparing a cloud-computing approach versus an on-premises archiving infrastructure.
- Consider e-mail archiving and e-discovery as potential starting points for implementing archiving as a service.
- Ensure that organizational data management policies are sufficiently defined and fully understood as the basis for service selection and service-level agreement (SLA) negotiation with cloud archiving vendors.

ANALYSIS

Organizations have been taking advantage of archiving solutions offered via a software-as-a-service (SaaS) model for years. Early adopters often needed to implement archiving quickly to meet financial regulations or as part of a legal discovery activity. In addition, as a non-mission-critical application, with a high capital expense entry point, archiving has lent itself well to a deployment model that can leverage a third-party infrastructure with pay-as-you-grow pricing. E-mail archiving applications, in particular, have seen strong adoption, particularly in light of the current recession. This is largely due to requirements for e-discovery.

Gartner defines cloud computing as “a style of computing in which massively scalable IT-enabled capabilities are delivered ‘as a service’ to multiple customers using Internet technologies”. The emergence of cloud-based archiving solutions is, in many ways, an extension of the SaaS model for archiving. What’s changing is the availability of technologies to better support multitenancy, enhanced Web user access and storage infrastructures fine-tuned to store and manage large amounts of data at a low price point, making the model more attractive for many organizations that hadn’t considered it before.

The decision to deploy a cloud-based archiving solution enables a trade-off between upfront implementation costs for software/hardware and ongoing IT management vs. a service delivery model structured to ensure that these elements are transparent. The SLAs between organizations and their archiving service providers must reflect storage capacity, uptime, retention and disposition policies, search and data access, and data migration as parameters of the agreement. Many organizations are electing to deploy “hybrid models,” consisting of a combination of on-premises and hosted archives. This approach may be used to facilitate rapid access to recent data or, increasingly, to use one or more service providers as disaster recovery facilities for compliance data.

Archiving and Cloud-Computing Infrastructure

Archiving service solutions generally provide for cost reduction (storage, backup) and management of data for e-discovery and compliance. In addition to archiving e-mail and instant messages, organizations are beginning to look at archiving for files, images, application data and numerous other content types. Archive infrastructures must support high volumes of data at low cost, extended data retention and periodic high-volume searching. Archiving applications generally do not require low latency, given the nature of data access. New cloud infrastructure and delivery attributes that enable better cloud service delivery of archiving include:

- Changes to storage technologies – Some e-mail archiving service providers deploy storage from major vendors. Others leverage cloud-based utility storage which is based on commodity components and is massively scalable. In either case service providers are enhancing the suitability of their storage platform offerings by adding better management software, enhancing transparent provisioning capabilities and by making them more extensible and elastic.
- Changes to security models – One of the drawbacks to widespread adoption of cloud-based archiving is the perception by organizations that their data is not secure, or that it can be accessed by unauthorized parties. Secure, multitenant architectures (architectures with resources shared by multiple customers) that provide verification and access control, encryption and other methods of eliminating risk associated with off-site storage and data transfer by a third party have evolved to support this model.
- Changes to locations of physical data centers – Organizations that deploy cloud-based archiving solutions should not worry about back-end infrastructure, planning for scalability or capital investment. Improvements to SaaS models in the ability to write to multiple locations and in replication technology have enabled better access to data regardless of where it is located.
- Changes to storage pricing models – Price is an important element to enabling a successful cloud infrastructure. Using cloud-optimized hardware enables service providers to lower their cost of ownership by minimizing costs associated with vendor installation, configuration, maintenance and additional software. Service pricing models have traditionally been based upon a combination of per-seat plus per-gigabyte models. Recently, some enterprise archiving service providers have adopted pure per-seat models with unlimited capacity.

The cloud-archiving provider’s objective is to offer their customers a way to manage their data for long-term retention, e-discovery and compliance, frequently while improving application server management and performance. Cloud-archiving solutions should take advantage of the underlying infrastructure, while hiding complexity and technical details.

Should You Consider Cloud Computing for Archiving?

As archiving evolves to a large-scale IT service for the management of multiple content types, its suitability as a cloud application increases. Archiving is becoming a commoditized service, not unique to any one organization and not seen as a differentiator to the business, which also lends itself well to a cloud model. Because archiving generally requires significant upfront capital expense (predominately storage and software license and maintenance costs), cloud computing is attractive to organizations that are unwilling or unable to release limited capital resources for spending on low-value infrastructure projects. Cloud archiving is also appropriate if an organization simply needs an inexpensive storage tier for less frequently accessed data.

Cloud archiving is well-suited to organizations that embrace new technologies and, in particular, those comfortable with Internet-based service consumption models. Organizations must also understand their service requirements for information access and retention management, and be comfortable with security models used by their archiving service providers. Many organizations are not willing or able to deploy a model that treats this IT service as a utility. These organizations may be better suited for on-premises archiving, or a hybrid solution consisting of a combination of on-premises and cloud-archiving services.

One of the key evaluation criteria for cloud-archiving services is whether SLAs make sense for your requirements and how well the provider can deliver on those SLAs. Cloud archiving must offer you a way to expand required storage capacity at an affordable price point; however, to be a true cloud enterprise archiving solution, payment must be for the archiving application service, not simply for physical storage used. There should be policies and SLAs associated with data retention and disposition (how long will data be retained, and how many retention periods can the provider offer?). This is important in that you will need to ensure that your provider has the ability to update media and data formats as the data ages, for the retention period specified.

There should also be SLAs for data access. While archived data may not be accessed frequently, organizations should think through their tolerance for less-than-immediate access and see that their provider can meet their requirements. Also, SLAs for the search production of electronically stored content (for e-mail and other data types) should be built into the service contract, particularly in environments with significant e-discovery activity.

Organizations evaluating cloud-based archiving solutions should also keep in mind the following:

- Security – Archiving requires the same strong security elements as other cloud-based applications. Most service providers encrypt customer data, making encryption key management an important consideration. In addition, archived data, in many cases, must be “tamper-proof” to comply with industry regulations, particularly in the financial services or broker-dealer markets.
- Data availability – The service must provide support for e-discovery and regulatory investigation, regardless of the age or location of the data. This is a fundamental use case for archiving solutions, and it does not change with a cloud implementation.
- Market maturity – As with any exciting new technology and IT services model, there are many vendors looking to capitalize on the hype. Due to the nature of the use cases, archived data must be retained for a long time. Organizations should ensure that they feel extremely comfortable that their service provider will be in business for the long haul.
- Migrations – Organizations must be comfortable that their service provider has the ability to continue to provide access to their data for as long as is needed. If you decide to switch providers, then how is this done and at what cost? A focus on ensuring conformance to any industry standards that support transportability of data is important. Also, in the absence of industry standards, you should closely evaluate the import/export capabilities offered by providers and ensure that those mechanisms can be used effectively if migration becomes necessary.
- Pricing – Organizations should work with their vendors to understand how their SaaS/cloud models will be priced and how that will change over time. How will the vendor account for long-term fluctuations (drops) in storage costs? Will those savings be passed on to you?

From SaaS to Cloud: Providers of Archiving Services

While many archiving SaaS offerings have been built from the ground up to deliver the scalability, reduced complexity and cost benefits associated with cloud-based utility storage, some are further along than others in embracing this model. Gartner sees a number of vendors enhancing their SaaS offerings by refreshing their infrastructures and improving their access to data archived in the cloud (via Web browsers, mobile devices, etc.). For archiving vendors, the cloud offers the ability to direct engineering resources to provide comprehensive archiving features and functions, as they do not have to focus on base access or the storage element of their archiving solutions – that is delivered through the cloud infrastructure component. This results in functionality improvements and savings that can be passed on to customers.

Most of the vendors offering established archiving services today are focused on e-mail. Zantaz (acquired by Autonomy in 2007) was the first service provider to deliver an e-mail archiving service, and, since then, the market has grown to include dozens of vendors. Gartner anticipates this growth will continue. Vendors that have focused specifically on building proprietary and/or private clouds to deliver their services (using multiple data centers to support failover and disaster recovery) include Google (Google Message Discovery, Postini-powered services), Microsoft (Exchange Hosted Archive, Exchange Online Dedicated Service), and Dell (MessageOne), to name a few.

Many vendors, including some of those listed above, are offering a hybrid model consisting of on-premises archiving products and cloud archiving. For example, Mimosa Systems and LiveOffice have just announced a partnership offering called LiveOffice CloudMerge for NearPoint, which combines Mimosa's on-premises solution with LiveOffice's cloud offering. This type of configuration enables organizations to manage more frequently accessed archived data for some period of time on-site, before moving that same data off-site, enabling them to keep costs low. Iron Mountain offers several archiving services, and has recently launched a new offering called Virtual File Store (VFS) that leverages a hybrid model built using a new cloud-optimized storage infrastructure to further reduce costs.

Increasingly, new archiving service providers are opting out of building their own cloud infrastructures and are leveraging cloud-based infrastructure utilities from Amazon, Nirvanix and others. Sonian (Sonian Archive SA2) utilizes Amazon's EC2 virtual server with S3 cloud storage, one of the first archiving products to take advantage of this architecture. Moonwalk uses Amazon S3 as a storage target for its Hierarchical Storage Management product (a

form of archiving), and Clearpace (RainStor) uses Amazon S3 and EC2 to move inactive data from databases or event logs to the hosted storage platform (for archiving and application retirement).

Given the availability and maturity of available offerings, organizations should feel comfortable that they can deploy solid SaaS-based archiving solutions, particularly for e-mail archiving. Based on the hype surrounding the evolution to cloud, however, organizations should make an effort to understand the types of services their vendor is offering. Organizations should also consider whether a combined cloud-based and on-premises archive infrastructure best meets their needs initially, or even indefinitely. Expect to see continued growth and more new players, and continued innovation in this area.