Information Management

CLOUD COMPUTING

Is It Worth the Risk and Expense?

Kirke Snyder, J.D., IGP

The "cloud" can refer to anything that's hosted remotely and delivered via the Internet. Today, nearly every type of core business function – from human resources to enterprise resource planning – is available via the cloud.

Companies like Amazon, Google, and Microsoft have found this service model very profitable. According to "Amazon Reveals Just How Huge the Cloud Is for Its Business," which appeared on *Wired.Com* on April 23, 2015, Amazon Web Services (AWS) earned $4.6 billion in revenue in 2014, an increase of 49% over the previous year.

The trend toward using cloud-based products and services is growing every year as more organizations are taking advantage of the benefits the cloud offers. For example, IT professionals are embracing third-party cloud computing and storage solutions as a way of supporting mobile workers, promoting multi-party collaboration, and avoiding upfront infrastructure costs.

However, many in-house lawyers are concerned about the potential for increased risk for data breaches and the increased cost and burden of e-discovery related to third-party cloud computing and data storage.

The November 2016 article in *The Sedona Conference Journal®* entitled "The Sedona Conference Commentary on Rule 34 and Rule 45 'Possession, Custody, or Control'"

sounds the alarm about these issues:

> … in today's digital world, the determination of whether and when information should be considered to be in a responding party's "possession, custody, or control" has become more complex. New technologies and organizational initiatives have further blurred the legal and operational lines of who actually "controls" data for purposes of preservation and production, and have multiplied the practical problems associated with preserving and producing data that a party does not directly control.

Security of Cloud-Based vs. On-Premise Data

The possibility of increased risk for cloud-based data depends upon many factors, but primarily it depends on which cloud solution is used.

A *private cloud*, which uses a proprietary architecture to deliver services to a single organization, is theoretically more secure than a *public cloud*, which delivers services through a computing infrastructure that is shared across multiple users, business units, or businesses.

A *single-tenancy solution*, an architecture in which each customer, or tenant, has its own software instance, is probably more secure than a *multi-tenancy architecture*, in which a single instance of a software application serves many customers.

A *hybrid solution*, which uses a mix of two platforms – an on-premises, private cloud and a third-party, public cloud service – with orchestration between the two platforms, could be less risky than a pure cloud solution, as it allows an organization to store its more-critical data on premise and its less-critical data in the cloud.

It should be noted that cloud vendors typically offer a much higher level of data center and virtual system security than most organizations can or will build out on their own. Furthermore, most security breaches are inside jobs, often from employees or other authorized users, and a cloud system offers greater protection from that than a traditional in-house data center.

No matter what data storage solution is implemented, though, the potential for human carelessness or error always exists.

Cost, Risks of Cloud-based E-discovery

Rule 26(a) of the U.S. Federal Rules of Civil Procedure allows for the discovery of "documents, electronically stored information, and tangible things" in the responding party's "possession, custody, or control." Similarly, Rule 34(a) and Rule 45(a) obligate a party responding to a document request or subpoena to produce "documents, electronically stored information, and tangible things" in that party's "possession, custody, or control." Yet, the rules are silent on what the phrase "possession, custody, or control" means, and case law across circuits is unclear and inconsistent as to its meaning.

The Sedona Conference's commentary mentioned above raises some potential red flags about information stored in the cloud:

The issues of who has possession, custody, or control in this age of electronic information is complicated by cost, burden, access, privacy, and contractual issues that simply did not exist in a world populated only by hardcopy documents.

The commentary goes on to say two primary issues could affect the cost of e-discovery of cloud-based information: the location of the data and who is managing the data (the company or a third party). Organizations must be sure they know the answers to the following questions before entering into a cloud agreement.

Where Will the Data Be Located?

It is common for data stored in the cloud to reside in more than one physical location, which can raise the question of which body of law is applicable and can increase the cost of preserving and collecting the data.

There is a growing legal concern about the ramifications of needing to collect information from cloud providers who store data in multiple jurisdictional locations, as different jurisdictions may have conflicting data protection laws. The obvious example is when data that is not

Cloud Computing Terminology

Software as a Service (SaaS) – SaaS is a software distribution model in which a third-party provider hosts applications and makes them available to customers over the Internet. Salesforce, Netsuite, and Office 365 are all examples of SaaS.

Platform as a Service (PaaS) – A PaaS provider hosts hardware and software tools on its own infrastructure and delivers them to users as a service. Typically geared toward software development, PaaS offers developers several advantages. For example, it frees them from having to install in-house hardware and software to develop or run a new application, allows them to change or upgrade operating system features frequently, and helps their development teams collaborate on projects.

Infrastructure as a Service (IaaS) – In an IaaS model, a third-party provider hosts hardware, software, servers, storage, and other infrastructure components on behalf of its users. IaaS providers also host users' applications and handle tasks like system maintenance, backup, and resiliency planning. Because IaaS customers pay on a per-use basis, typically by the hour, week, or month, or based on the amount of virtual machine space they use, they save the capital expense of deploying in-house hardware and software.

Public vs. Private Cloud – A *public cloud,* such as those from Amazon Web Services or Google Compute Engine, delivers computing services to multiple organizations; its computing infrastructure is shared across different users, business units, or businesses. A *private cloud* is dedicated to a single organization and uses a proprietary architecture. They have similar advantages, including scalability and self-service, but a public cloud's shared computing environment isn't suitable for some businesses, such as those with mission-critical workloads, security concerns, uptime requirements, or management demands.

Hybrid Cloud – The term *hybrid cloud* implies a cloud computing environment that uses a mix of on-premises, private cloud and third-party, public cloud services with orchestration between the two platforms. By allowing workloads to move between private and public clouds as computing needs and costs change, hybrid cloud gives businesses greater flexibility and more data deployment options. For example, an enterprise can deploy an on-premises private cloud to host sensitive or critical workloads while using a third-party, public cloud provider to host less-critical resources.

Service Level Agreement (SLA) – In addition to setting system reliability standards, an SLA spells out parameters for issues such as data ownership, security requirements, and maintenance schedules. An organization should make sure its SLA has a clause that explicitly states that it can export its data from the provider, including how often and in what type of format it may access your data. It's common for SLAs to also stipulate that the vendor will help migrate data for an appropriate fee.

Multi-tenancy – Multi-tenancy is an architecture in which a single instance of a software application serves multiple customers, or tenants. Multi-tenancy can be economical because software development and maintenance costs are shared. It can be contrasted with single-tenancy, an architecture in which each customer has its own software instance and may be given access to code. With a multi-tenancy architecture, the provider needs to make updates only once. With single-tenancy architecture, the provider must touch multiple instances of the software to make updates.

otherwise subject to European Union (EU) data protection limitations becomes subject to these laws because the cloud provider elects to store that information on servers in the EU.

Another example of this conflict is the U.S. requirement that an entity must "voluntarily" disclose to a requesting party its employees' e-mail without their knowledge or consent, which runs directly against the EU and other countries' trend of increasing the protection of individuals' information.

Another issue is when potentially relevant cloud-based data is preserved "in place," as that calls into question its physical location for production or inspection. Is the data being produced at the locations of dozens of servers around the world or at the cloud provider's headquarters location? The subpoena service location and the controlling body of law may be unclear.

Organizations, then, must always be certain about where their data will be stored to avoid these types of potential conflicts.

Who Manages the Data?

Data stored in the cloud may be easily accessed by a greater number of people than if it were stored onsite. *Easy access*, though, shouldn't be confused with *easy to collect*. An organization's ability to search and collect its own data may be limited by the cloud vendor's search utilities.

For example, a cloud vendor's method of storing data may retain all or most of the primary components of the original metadata, but still not meet the demanding party's requirements for retrieving it. If the data manager's storage protocols are not compatible with the search tools used by the data-owning organization's own shared servers and computers, this may affect the organization's ability to collect the data it needs.

Recommendations for Minimizing Risk, Cost

Organizations must be proactive to minimize the risks and costs

inherent with cloud computing. Here are some recommended actions to take.

Consider encryption. Before storing data in a cloud server consider encrypting sensitive data.

Do due diligence. Before finalizing the contract, understand the provider's costs, policies, and protocols for responding to a subpoena.

Take a test drive. Load a reasonable volume of unstructured data into the cloud application or platform. Include user-created documents along with e-mail, attachments, .zip, and .pst file types. Test and confirm your ability to encrypt sensitive data and to search and identify potentially relevant materials. Understand exactly how you will protect those files from alteration or deletion when preserving in place in the cloud or when exporting the files to a secure legal hold server.

If involved in a legal hold scenario, consider the following protocols.

Collect completely. The idea of taking a set of keyword search terms and running them against the cloud data to reduce the amount of data to be collected may be tempting. However, because new keywords will arise as data moves through the review process, minimize the risk and cost of cloud-based e-discovery by collecting the complete set of potentially relevant documents.

Use the right tools. Many data analysis utilities that can index and search a complete collection of e-discovery documents are available to rent or buy. These utilities can perform simple keyword searches, cluster documents by concepts, and create timelines to help reveal "who knew what and when they knew it." Compare the utilities' features; some can perform this analysis directly on the cloud storage platform, while others require data collection prior to analysis.

Document fully. Regardless of the type of collection, document the collection process: what data was collected, who collected it, and when and how it was collected. If possible, create an audit trail that compares information about the source data and collected data. Capture the corresponding metadata and hash values – an "electronic fingerprint" – of the source data and the collected data to prove that the collection was forensically sound and legally defensible.

Be Aware and Prepared

Favorable economics for cloud computing will continue to drive its growth and adoption rates, so organizations and the courts must prepare for evidentiary challenges associated with identifying, preserving, and collecting cloud data for e-discovery. While cloud data storage doesn't automatically raise the risk and costs associated with e-discovery, it could. It's best to be aware of the potential risk and cost issues and be prepared to mitigate them.

About the Author: Kirke Snyder J.D., IGP, has more than 25 years of experience helping clients solve litigation e-discovery, records management, and information security related issues. Located in Denver, Colorado, the Certified Information Governance Professional has worked with more than 200 law firms, corporate law departments, and government agencies. He has a master's degree in legal administration and a juris doctor degree. Snyder can be contacted at *kirke@LIC.consulting*.