IT@Intel White Paper

Intel IT
IT Governance and Cloud Computing
May 2013

Planning for eDiscovery in the Cloud

Intel IT and legal counsel continue to improve the processes, procedures, and capabilities for eDiscovery in the cloud. This enables Intel's cloud initiatives to move forward while working to address the legal obligation to preserve data potentially related to legal matters.

Executive Overview

Intel IT is implementing our technology roadmap for using hybrid cloud services. While cloud computing presents opportunities to reduce costs, increase data availability, and generate new sources of revenue, it also poses electronic discovery (eDiscovery) challenges.

- Cross-platform eDiscovery standards and requirements do not exist.
- Data export capabilities are often limited or nonexistent.
- Solution providers of eDiscovery applications are slow to introduce data collection capabilities because of cloud service providers' limited APIs.

To mitigate these challenges, we're developing best practices so that our IT eDiscovery team can locate, export, and manage electronically stored information in the hybrid cloud environment, while at the same time comply with applicable and governing privacy and data protection laws and agreements.

Some of these best practices are proactive in that they precede eDiscovery requests. Others are in response to an eDiscovery request.

- To prepare for eDiscovery requests for data stored in an Intel cloud environment, our ideal strategy is to look for cloud service providers who offer the data export capabilities we need, and then test those capabilities.
- When we receive an eDiscovery request for cloud data, we determine the best way to access and retrieve that data. We communicate our needs to the cloud service provider, export relevant metadata if required, and test the completeness and accuracy of the data collection.

Close collaboration between Intel IT and Intel's legal team strengthens Intel's ability to meet legal needs as they apply to our cloud computing initiatives. Intel IT and legal counsel continue to improve processes, procedures, and capabilities for eDiscovery in the cloud. This enables Intel's cloud computing initiatives to move forward while working to address the legal obligation to secure data potentially related to legal matters.

Steve Watson
Technical Solutions Engineer,
eDiscovery, Intel IT

Contents

IT@INTEL

The IT@Intel program connects IT professionals around the world with their peers inside our organization – sharing lessons learned, methods and strategies. Our goal is simple: Share Intel IT best practices that create business value and make IT a competitive advantage. Visit us today at www.intel.com/IT or contact your local Intel representative if you'd like to learn more.

BACKGROUND

Intel recognizes the potential of cloud computing to help drive new revenue, reduce costs, and increase the availability of data. Having successfully built an enterprise private cloud, we have now established a technology roadmap for the use of hybrid clouds—a mix of public and private clouds. While cloud computing presents many opportunities, it also poses challenges for Intel IT regarding the potential legal obligation to fulfill discovery requests for electronically stored information (ESI) or data stored in the cloud.

Electronic discovery (eDiscovery) has become increasingly important for enterprises (see sidebar). This process requires IT departments to prepare for eDiscovery requests for data stored in private, public, or hybrid clouds, while balancing employee privacy rights with corporate data security and eDiscovery concerns. Intel IT is taking a proactive approach to prepare for ESI data requests and to be able to respond promptly.

Cloud computing poses several new eDiscovery challenges.

- **Complexity.** Currently, no cross-platform standards exist for cloud-based eDiscovery. Each cloud service provider supports eDiscovery and data collection in its own way, especially outside of the cloud service provider's application. For example, one cloud service provider may offer a bulk export capability while another allows files to be downloaded or exported only one at a time. Some cloud service providers do not offer an API that enables data access outside their application; even for those that do, the APIs are often quite limited.
- **Export capability.** Not all cloud service providers consider data export capabilities when they design their products. Those that do provide some export capabilities may not have considered all contingencies, such as exporting specialized file types. Often, export capabilities are limited to a specific use case, such as a web browser, a client application, or specific networking protocols. Also, export capabilities may require user-specific permissions.

eDiscovery Becoming Increasingly Important to the Enterprise

Discovery—pretrial procedures involving the exchange of information between parties involved in a legal proceeding—has significantly changed in the last 10 years. As more companies use electronic methods to store and share data, electronic documents and electronic Discovery (eDiscovery) have taken an increasingly prominent role in litigation.

Like many large corporations, Intel is regularly involved in civil court cases relating to intellectual property and other corporate legal matters. Therefore, we are proactive in the area of eDiscovery, with teams and processes in place to respond to changes in the legal and technology landscapes and remain in compliance with Intel's privacy and data security policies. We continue to expand and adjust our eDiscovery processes as technology advances. Intel's ability to meet its obligations in court may be jeopardized without a thorough review of eDiscovery obligations, including those in the area of cloud computing.

- **eDiscovery applications.** Providers of eDiscovery applications have been slow to introduce capabilities to interface with data stored in the cloud. This delay is partly because of limited collection and export capabilities. In addition, while some solution providers have introduced connectors to various cloud service providers, their applications are often limited by the cloud service provider's API or allowable connectivity options. For example, a cloud service provider may not offer any way to interact with the data except by using their proprietary tool, making it difficult to create forensic applications that work consistently across cloud service providers.

One study revealed that 95 percent of participating cloud service providers did not understand their clients' basic legal requirements or the value of supporting eDiscovery. Even more surprising, the same study reported that 100 percent of the general counsel and 99 percent of the IT management who participated in the study were not considering support for eDiscovery as a criterion for choosing a cloud service provider.[1]

The challenges that eDiscovery pose in the cloud environment, coupled with a widespread limited understanding of what eDiscovery is and why it is important, have compelled Intel IT to investigate potential eDiscovery hurdles and formulate best practices that are in full compliance with applicable and governing privacy and data protection laws and agreements. While it is usually possible to retrieve data from the cloud, the process varies significantly from the simple downloading of files from a desktop PC.

[1] Free Press Release. "2012 eDSG Study on How Cloud Service Providers Support eDiscovery and Information Governance." www.free-press-release.com/news-2012-edsg-study-on-how-cloud-service-providers-support-ediscovery-and-information-governance-1331158526.html

MITIGATING THE PRIMARY eDISCOVERY CHALLENGES ASSOCIATED WITH CLOUD COMPUTING

By taking a proactive approach to the deployment of cloud services and to responding to eDiscovery requests for cloud-based data, we have found it is possible to move ahead with cloud computing initiatives and address potential eDiscovery challenges that may arise.

We have found that we can achieve best results by seeking to understand what data we might need to collect and what methods might enable us to retrieve that data. Discussing our requirements with cloud service providers enables us to choose providers wisely and to craft agreements that maximize Intel's ability to respond to future eDiscovery requests.

Risk assessment is an important aspect of this process. Intel's legal team can identify which types of data are more likely to be relevant to eDiscovery requests and Intel IT's eDiscovery team can identify which types pose more or less risk associated with data collection and data export. Risk can be measured in terms of time and effort. Table 1 summarizes the types of questions associated with eDiscovery that we ask the legal team, the cloud service provider, and Intel IT staff.

When eDiscovery requests occur, we determine the best way to access the requested files. When exporting the files, we maintain close communication with the cloud service provider and attend to important details such as exporting relevant and required metadata in addition to the files themselves. After the files are exported, we test them for accuracy and compatibility with forensic tools and the legal team's needs.

Table 1. eDiscovery Discussion Topics by Role

Role	Questions We Ask
Intel's legal team	▪ What types of data are required? ▪ What format should the preserved data be in?
Prospective cloud service providers	▪ Are there export capabilities? ▪ Can we export data in bulk? ▪ Are there any file types that cannot be exported? ▪ Is there assistance with export activities? ▪ What tools exist to help with export? ▪ Where does the metadata reside?
Intel IT staff	▪ Can all necessary file types be exported? ▪ Is the export complete and accurate? ▪ Is the required metadata also exported? ▪ Is the exported data compatible with existing eDiscovery tools? ▪ Is the result of the export in a format the legal team can interpret and use?

Proactive Considerations

The best time to address eDiscovery in the cloud is during the process of choosing one or more cloud service providers and establishing the cloud environment, including teams, tools, and infrastructure. By proactively considering eDiscovery needs and potential problems, we avoid difficulties that might arise later. We also work with privacy, Human Resources, and litigation teams to define the processes regarding corporate and personal data distinction, collection, and disposal.

Verifying Data Export Capabilities

When choosing a cloud service provider, Intel IT prefers that data export capabilities be precisely described in the service contract, although this is not always possible currently, because of limitations in our internal processes and in the cloud providers' processes.[2]

We strive to take this proactive approach because if the export capabilities are not sufficient, complying with an eDiscovery request can be costly in terms of time and effort. For example, suppose a company receives a request for data and hasn't discussed bulk export capabilities with its cloud service provider. It is possible the company may have to download thousands of files one at a time within a one- to two-week period—a very labor-intensive undertaking.

We explore the following aspects of data export with a potential cloud service provider:

- What sort of bulk export capabilities exist, either through the cloud service provider's proprietary environment or through an external API
- Whether the cloud service provider will assist with data export
- Which data types can be exported
- For those data types that cannot be exported, what workarounds exist
- The location of the metadata associated with the files and how to export that metadata
- Whether version control capabilities exist for exported files
- Whether our eDiscovery tools support the data types that are created during export

Additional considerations include determining how the export can be organized, such as by custodian, date, or geographical location.

Testing Collection Capabilities

We do not simply accept the cloud service provider's claims about data export and collection. We strive to test the exportability of each data type. For example, we discovered that one cloud service provider's contract and documentation stated that we could export data using the administrative console. Our testing revealed that certain data types could not be exported. Had we relied on the documentation instead of performing testing, and then received an eDiscovery request that included certain data types that existed on that cloud service provider's servers, our ability to swiftly comply with the request would have been hindered. When discussing eDiscovery capabilities with a cloud service provider, we establish that the capabilities meet Intel's expectations.

During our testing, we investigate whether our existing eDiscovery tools and processes are compatible with the exported data. If not, we may need to write custom scripts or attempt to collaborate with additional third-party solution providers. We also strive to verify that the format of the exported results is acceptable to the legal team. For example, data from a wiki page or web page might be exported as a comma-separated values (CSV) file.

eDiscovery Request Compliance Considerations

In addition to the preventive steps we take when choosing a cloud service provider, we have also established best practices that help us more efficiently comply with eDiscovery requests.

Determining the Best Data Collection Method

When we need to retrieve data stored in the cloud, we have found that determining how end users access the data is a good starting point for determining the best way to retrieve the data. For example, if the end user accesses the data through a web browser, that data may be best extracted using the web browser to interact with the back-end or administrative consoles. Or if the primary user interface is an installed client application, we explore the possibility of accessing the content using APIs or other application-based means.

In some instances, tools may exist for retrieving certain types of cloud data. For example, a vector-based image format or export to a .PDF file can be useful for retrieving data presented primarily though a web browser. These formats tend to preserve the look and feel of the original website.

After we have determined the best technical way to collect the data, we perform extensive testing to verify that the data is exported accurately and completely.

[2] Other terms for service contract include service-level agreement (SLA), terms of service, or licensing agreement.

Communicating with the Cloud Service Provider

We have found that cloud service providers often are not familiar with eDiscovery terms and requirements. Therefore, when communicating with cloud service providers about eDiscovery, we ask specific questions. Instead of asking, "Do you support eDiscovery?" we might ask, "What happens if we want to get a copy of our data?" However, simply stating "We need a copy of the data" may yield incomplete results, which can introduce delays and duplication of effort. Instead, we are more explicit about how the data should be exported and ask probing questions, such as "How do you track file creation and modification?"

Also, we use terminology that the cloud service provider can understand, and we ask questions several ways. For example, the cloud service provider may be more familiar with the term "logging" than with the term "metadata."

Exporting Metadata

One easily overlooked aspect of eDiscovery in the cloud is the potential need to extract metadata. When a file is uploaded to the cloud, file metadata is sometimes stored in a database table. This metadata may include a file creation timestamp, the user account that uploaded the file, the source URL, and the date and time the file was uploaded. When a file is exported, its metadata may not accompany the file. If the legal team needs the metadata, two separate downloads—one of the file and one of the metadata—may be required.

It is also important to realize that export of the data may alter file creation dates. When we test export capabilities, we verify that the export API queries the secondary table containing the metadata so that the true file creation timestamp is preserved; otherwise, the file might be stamped with the time and date of the export instead. In cases where the metadata is unavailable, we document what steps were taken to export the data so that the chain of custody can recognize the limitation of the exported date field and document the actual date of creation.

Collecting Data

Once we have exported the data associated with an eDiscovery request, we check for quality and accuracy. We examine the original data in the cloud application and the exported data to verify that we have received what we expected. For example, we verify that the hash function[3] matches the original file and the exported file or, in the case of web-browser-based collections, we verify that the data in the web browser matches the collected image format.

Our process identifies important timelines and milestones and the people responsible for each step. For example, we determine how long the collection of data will take and whether this amount of time fits with the eDiscovery request timeline. If the service-level agreement (SLA) contractually obligates the cloud service provider or a third party to perform the data collection, we review the level of accuracy and turnaround time specified in the cloud service provider and third-party SLAs. In the case of large amounts of data, we determine whether there is a size threshold at which the data must be shipped rather than downloaded.

Once the data collection is complete, we verify that the format of the export is compatible with existing tools and satisfactory to the legal team. We inform Intel's legal team that the exported data may be in a different file format, or presented in a different form, than the attorneys are accustomed to seeing.

[3] A cryptographic hash function is a way of encoding data so that changes to that data are easily detected. These changes could be intentional or may result from dropped bits during upload or download.

Choosing a Cloud-based Collaboration Tool that Supports eDiscovery

As more applications and business groups take advantage of cloud computing at Intel, we have found that cloud-based collaboration tools can help increase productivity and team communication. These sorts of tools can also enable content synchronization, enabling employees to access up-to-date content from a variety of devices.

The IT eDiscovery team strives to ensure that any cloud-based collaboration tool we choose supports eDiscovery when needed by providing the following capabilities:

- **Enterprise management.** The cloud service provider supports the ability to control which devices can connect and provide access to data even if the local copy has been wiped from a device, such as when a device has been returned and reprovisioned to a new user.
- **Data retention policies.** The service-level agreement should clearly state how long the cloud service provider keeps data after it is deleted from a device and whether it keeps previous versions (revision history).
- **Data leakage prevention.** It should be possible to log which people access files—and folders—stored in the cloud.

Generally speaking, Intel IT encourages employees to think about the data classification of a file; if the file contains sensitive information, it is typically not suitable for cloud storage. We are also exploring data anonymization, which is the process of obscuring data stored in the cloud to protect the data and the privacy of our employees, customers, and suppliers. For more on Intel IT's work with data anonymization see "Enhancing Cloud Security Using Data Anonymization."

For example, attorneys may be familiar with seeing data from a collaborative web-based tool in a neatly designed interface. However, this data may be exported as code. Similarly, a table of data from a spreadsheet application may be exported as a CSV file.

CONCLUSION

Intel IT is taking a proactive approach that enables Intel to take advantage of the benefits of hybrid cloud solutions—such as reduced costs, increased data availability, and new potential sources for revenue—in a secure manner. But along with these benefits come challenges. One of those challenges is preparing for eDiscovery in the cloud environment.

By its very nature, data stored in the cloud is outside of the physical control of the corporation. Intel IT and Intel's legal team strive to find ESI—on corporate file shares, laptops, desktop PCs, or in the cloud. We have developed best practices for planning for eDiscovery in the cloud environment, while remaining in compliance with applicable and governing privacy and data protection laws and agreements. By implementing these best practices, we can move ahead with cloud computing initiatives while effectively addressing potential eDiscovery challenges that may arise related to cloud-based data.

FOR MORE INFORMATION

Visit www.intel.com/it to find white papers on related topics.

- "Enhancing Cloud Security Using Data Anonymization"
- "Successful eDiscovery in a Bring-Your-Own-Device Environment"

ACRONYMS

CSV	comma-separated values
eDiscovery	electronic discovery
ESI	electronically stored information
SLA	service-level agreement

For more information on Intel IT best practices, visit www.intel.com/it.

Printed in USA ♻ Please Recycle 0513/ERAI/KC/PDF 328919-001US