SIFT

Next-Generation Cloud Data Migration

The Core Question

How do you move files to the cloud **safely**, and ensure ongoing data protection?

Problem #1: File Assessment

Only once we know the contents of a document should we consider moving it to a public cloud location! Now if only we lived in a world where everyone tagged their documents with useful and accurate metadata. Unfortunately, that simply isn't the case – and it's all too easy for users to classify documents incorrectly. Even beyond classification levels, documents may contain privileged or sensitive information, including employee or customer personally identifiable information (PII). It's vital that document content be analyzed against a list of organizational keywords, which may include static keywords as well as pattern-based data, like social security numbers or IP addresses.

Problem #2: File Protection & Access

Once documents are moved to the cloud, it's equally important they be continuously monitored for changes. It's all too easy for a cloud-hosted document to be updated with sensitive information, potentially exposing the organization to risk and liability. Even as documents must be protected, it's equally important that organizations retain access to valuable data. Without proper metadata tagging, sophisticated security technologies like DRM, DLP, and ABAC become completely ineffective. Additionally, untagged and unsearchable documents (like pictures and video) may wind up lost down deep content "gravity wells," hidden from enterprise search tools.

The Risks

There are a number of issues that complicate the problem of migrating data to the cloud:

Complexity. Complex multi-cloud environments are especially vulnerable to misconfiguration.

Decentralization. Documents are stored in different locations — including local hard drives, file shares, web portals, and email.

Classification. Most organizations have not implemented a formal data classification system, which makes it difficult to manage documents appropriately.

Collaboration. Companies can only function when information is shared, however controlling data access is always difficult.

Human Error. Despite the best policies and procedures, and the best of intentions, people make mistakes.

Document Types. Files are stored in many different formats, most of which require special software to open.

Auditing. It is difficult to track where and when documents are transferred, and hard to detect anomalous behavior.

Keyword Tagging. Most files are not tagged correctly with relevant keywords, complicating both enterprise search and data protection.

Data Patterns. Some customer data (like SSNs or credit card numbers) cannot be identified by a single static keyword.

Searchability. Many file formats are not text-searchable, such as picture files, scanned PDFs, and video.

MIGRATE

Publish Files Based
On Keyword Content

MANAGE

Make All Files
Tagged & Searchable

MONITOR

Identify and Recover
From Data Spills

The Solution

SIFT™, from Aerstone Labs, is automated cloud data migration software. It's designed to identify keywords in files of any kind, based on a centrally-maintained list, and then move tagged files to the appropriate cloud or network destination. SIFT™ helps protect an organization from accidentally exposing sensitive data, while making all files properly discoverable. SIFT™ can be used as a stand-alone portal, or integrated seamlessly with existing document management systems. Once configured with a relevant set of keywords, SIFT™ evaluates and tags files with useful and specific metadata, then moves the files to where they belong, **safely** and **securely**.

The Features

- Move files to the cloud safely and securely, by scanning file content and validating cloud enclave suitability.
- Reliable human review for files that are ambiguous based on keyword content.
- Support for a wide range of content file types, including most Microsoft Office and Adobe documents.
- A patented OCR pre-processing algorithm, to support scanning for text in pictures and PDF documents.
- An advanced machine learning capability, to identify specific shapes in pictures and video files.
- A modular design, which easily allows extending support to additional document types.
- Highly recursive file scanning, including processing of zipfiles and OLE-embedded objects.
- A RESTful API, for inline deployment with existing enterprise document management processes.
- Customizable scanning rules, to support assessing documents against a centralized list of keywords.
- Full support for both static and pattern-based keywords, based on industry-standard RegEx patterns.
- Tag file metadata with discovered keywords, to support enterprise search, as well as security solutions like digital rights management (DRM), data loss prevention (DLP), and attribute-based access control (ABAC).
- Monitor cloud locations for spilled data against enclave keyword rulesets.
- Highly customizable auditing and historical reporting, with drill-down capability.

Take Control Of Your Cloud Data!

AERSTONELABS

For more information, or to arrange SIFT™ demo,
please visit our website at www.aerstonelabs.com