VARONIS WHITEPAPER

Mastering the Information Explosion

CONTENTS

MASTERING THE INFORMATION EXPLOSION

BY OHAD KORKUS - EXECUTIVE VICE PRESIDENT OF ENGINEERING, CTO AND CO-FOUNDER

INTRODUCTION

This paper discusses the explosive growth of unstructured and semi-structured information in organizations of all sizes, and its ramifications on governance, protection and management practices.

UNSTRUCTURED DATA

Structured data repositories, like databases, have strictly defined data types, rules to enforce the existence of data and values, and rules to govern where in the database each particular type of data is stored.

In contrast, unstructured data is stored in continuously evolving, user-defined directory structures with few rules about what types of data is stored, where it is stored, and what each file contains. Unstructured data encompasses all distinct files – documents, images, spreadsheets, presentations, videos, and audio files – stored on file servers, NAS devices, and in semi-structured repositories like SharePoint.

80% OF YOUR DATA IS UNSTRUCTURED

Eighty percent of an organization's data is unstructured (Gartner 2010). Documents are being created constantly by virtually all members of an organization with access to a laptop or workstation, and saved on file servers and SharePoint servers, where they remain for long periods of time—often indefinitely. Unstructured data represents an enormous amount of organizational data inventory.

UNSTRUCTURED DATA GROWTH IS EXPONENTIAL

Not surprisingly, with so many individuals creating and storing files, the volume of unstructured data is growing at a phenomenal rate. Gartner estimates that in 5 years, unstructured data will grow by 650% – this roughly equates to 50% year over year growth.

A GREATER PORTION OF IT NEEDS TO BE MANAGED AND SECURED

As the data grows so does the scope of what it contains, and the potential ramifications associated with its loss, exposure, and misuse. As risks increase, they are naturally followed closely by new regulatory requirements, archive policies, intellectual property requirements, and personal confidentiality laws mandating additional protections. In The Digital Universe Decade – Are You Ready?, John Gantz and David Reinsel write, "The number of things to be managed is growing twice as fast as the total number of gigabytes [...] By 2020, almost 50% of the information in the Digital Universe will require a level of IT-based security beyond a baseline level of virus protection and physical protection. That's up from about 30% this year. And while the portion of that part of the Digital Universe that needs the highest level of security is small – in gigabytes and total files – that portion will grow by a factor of 100."

Data protection is necessary to safeguard an organization's customers, employees, business partners, and investors. It is fundamental in securing intellectual property and competitive edge, and for maintaining the organizational trust that allows it to properly function. Every organization has at least a modicum of customer information, employee information, product design documents, HR documents, legal documents, blue prints, images, audio and video files that relate to the business and its customers — most organizations have a formidable amount. This data must be protected and managed.

COMPLEXITY INCREASES WHILE HUMAN RESOURCES STAY THE SAME OR ARE REDUCED

> "Although the amount of information in the Digital Universe will grow by a factor of 44, and the number of containers or files will grow by a factor of 67 from 2009 to 2020, the number of IT professionals in the world will grow only by a factor of 1.4."
>
> – Source: IDC, The Digital Universe Decade – Are You Ready?

Digital collaboration is now the way organizations work, down to the most fundamental business processes – it is necessary to remain even minimally productive and competitive. Modern, data-driven organizations continually form collaborative, cross-functional teams that access fluid sets of digital resources. This increasing number of teams, combined with additional management and security requirements, cause the number of data containers – folders and SharePoint sites – to increase proportionately faster than the information we're storing in them. Each data set and container needs to be restricted to a list of team members and appropriate users. Each list of users represents a business decision made at a single point in time – these lists need to change in order to keep up with organizational changes and changes to the data's sensitivity. In order to properly maintain each list, a user with knowledge of the data and the organization needs to review it periodically.

Today, a single terabyte of data will often contain 50,000 folders, of which 2500, or 5%, are uniquely permissioned. Each of these 2500 folders represents an organizational decision – "who should have access to this data?" More containers will mean more decisions, and more maintenance. Unfortunately, the ever-increasing decisions and maintenance activities will need to be facilitated by the same (or fewer) staff members who struggle to keep up with the current load.

The increasing complexity widens an already sizable information gap between end users, data, and IT. Today, 91% of organizations have no process to identify a data owner (Source: Ponemon Institute Study, June 2008). Data containers without an owner can only be maintained on a best effort basis by IT personnel, who lack insight into who can and does access the data it contains. 77% percent of organizations already report that they lack adequate automation and visibility to manage and protect their data – in 5 years time they will be exponentially farther behind.

THE VARONIS METADATA FRAMEWORK

Five years later, IDC has affirmed the Varonis vision: "The fastestgrowing category in the Digital Universe is metadata, or data about data."

– Source: IDC, The Digital Universe Decade – Are You Ready?

Metadata is data about data. For example, when was it created? Who has access to it? Who has been accessing it? Is there a description or "tag" denoting its sensitivity? Metadata is critical for managing data — without it, these critical questions are unanswerable.

These challenges were clearly approaching prior to Varonis' inception in 2005. Manual management and protection practices were already ineffective, and far from efficient. Ongoing, scalable data protection and management required a new kind of technology designed to handle ever-increasing volume and complexity.

We realized that four types of metadata were critical for data governance:

- User and Group Information – From Active Directory, LDAP, NIS, SharePoint, etc.
- Permissions information – Knowing who can access what data in which containers
- Access Activity – Knowing which users do access what data, when and what they've done
- Sensitive Content Indicators – Knowing which files contain items of sensitivity and importance, and where they reside

Furthermore, we knew that the number of functional relationships between these metadata entities would grow exponentially.

For example, on a given day, a single folder may be accessible by 3-5 groups, each containing 10-20 members, with some overlap between each group. The folder may contain dozens of subfolders and thousands of files, many containing sensitive data. Now, start the clock. Over a single month, dozens of users in these groups will access thousands of files, usually several times a day. Content will be created, deleted, and change. Permissions and group memberships will change.

Now suppose you want to answer a hypothetical question, "if I had removed access for a single group of users on this container several months ago, who would have lost access, and what actual activity would have been affected by this change? The number of relationships required to answer these two questions for a single folder can exceed 100,000 – in a single month of activity.

These relationships accumulate relentlessly in today's environments. In future environments, new data will be created, new servers will be added, new platforms will be created, and eventually other types of metadata will be desirable to collect and analyze, further increasing the number of functional relationships. The magnitude of the present and future data sets represented a formidable computing challenge – one that needed to be solved using standard computing infrastructure components – specialized hardware would make any solution unfeasible.

In order to meet this computing challenge, we needed to create a robust framework to collect, process, analyze and present metadata to IT and data owners that could scale to present and future requirements, using standard equipment. We constructed the Varonis Data Governance Suite as that framework.

The Varonis Metadata Framework non-intrusively collects critical metadata about unstructured and semi-structured data, generates metadata where existing metadata is lacking (e.g. its file system filters and content inspection technologies), pre-processes it, normalizes it, analyzes it, stores it, and presents it to IT administrators in an interactive, dynamic interface. Once data owners are identified, they are empowered to make informed authorization and permissions maintenance decisions that are then programmatically executed – with no

IT overhead or manual backend processes.

With these metadata streams collected, synthesized, processed, and presented intelligently by the Varonis framework, organizations now regularly answer the numerous pressing questions that arise in data governance:

- Who has access to a data set?
- Who has been accessing it?
- What other data have they been accessing?
- Who is the likely data owner?
- Who has unnecessary permissions to each data set?
- Which data is sensitive?
- Where is my sensitive data overexposed and how do I fix it?
- What data is unused?

Since its release in 2006, Varonis has been collecting, processing, analyzing, and presenting metadata for its over 700 customers, automating their data governance processes, increasing visibility and data intelligence, and helping them meet and even get ahead of their increasing data management and protection demands.

SUMMARY

Data and its associated protection and management requirements are growing at an extraordinary pace. Without a scalable framework to manage and protect metadata, organizations will fail to keep up with their already overwhelming data related tasks. Manual processes have already proven to be ineffective; one-dimensional reporting tools are not constructed to accommodate increasing data volumes or multi-dimensional functional relationships, and will present only an ineffectual slice of the data governance picture. IT organizations urgently require an automated solution for data governance that can meet their management requirements quickly, effectively, and completely.

The Varonis Metadata Framework has successfully transformed ineffective, manual data governance processes to automated, precise, decision-enabling workflows for over 700 organizations. With Varonis, organizations manage and protect their growing data sets and complexities more efficiently and effectively. They can begin reducing significant risks from exposed sensitive data on day 2 of implementation, and can see a full return on investment within 3 months of deployment.

> "How will we protect the information we need to protect? If the amount of information in the Digital Universe is growing at 50% a year or so, the subset of information that needs to be secured is growing almost twice as fast. The amount of UNPROTECTED yet sensitive data is growing even faster."
>
> – Source: IDC, The Digital Universe Decade – Are You Ready?

ABOUT VARONIS

Varonis is the leading provider of software solutions for unstructured, human-generated enterprise data. Varonis provides an innovative software platform that allows enterprises to map, analyze, manage and migrate their unstructured data. Varonis specializes in human-generated data, a type of unstructured data that includes an enterprise's spreadsheets, word processing documents, presentations, audio files, video files, emails, text messages and any other data created by employees. This data often contains an enterprise's financial information, product plans, strategic initiatives, intellectual property and numerous other forms of vital information. IT and business personnel deploy Varonis software for a variety of use cases, including data governance, data security, archiving, file synchronization, enhanced mobile data accessibility and information collaboration.

Free 30-day assessment:

WITHIN HOURS OF INSTALLATION

You can instantly conduct a permissions audit: File and folder access permissions and how those map to specific users and groups. You can even generate reports.

WITHIN A DAY OF INSTALLATION

Varonis DatAdvantage will begin to show you which users are accessing the data, and how.

WITHIN 3 WEEKS OF INSTALLATION

Varonis DatAdvantage will actually make highly reliable recommendations about how to limit access to files and folders to just those users who need it for their jobs.