

VARONIS WHITEPAPER

Mastering Big Data

CONTENTS

OVERVIEW	3
WHAT DOES IT MEAN TO HARNESS THE POWER OF BIG DATA?	4
BIG DATA ANALYSIS AND STRUCTURED DATA	5
BIG METADATA—INSTRUMENTING UNSTRUCTURED DATA FOR DATA ANALYSIS	6
THE VARONIS METADATA FRAMEWORK	7
VARONIS PRODUCTS	8

MASTERING BIG DATA

BY YAKI FAITELSON - CEO, PRESIDENT AND CO-FOUNDER

OVERVIEW

Last year we released a white paper, “Mastering the Information Explosion,” that discussed how quickly data is growing, how the pace of collaboration is accelerating, and how manual methods of managing and protecting data are insufficient. Therefore, effective data governance requires harnessing the power of metadata through intelligent automation. It is not surprising that industry experts are now saying that the same kind of automation is necessary for more than good governance— in order to harness the power of “Big Data,” you’ll need to analyze and look for patterns in how and when these massive amounts of data are used, who uses it, in what sequence, and what it contains in order to effectively run a data-driven organization.

In “Big Data is Only the Beginning,” Gartner states: “Extreme information management challenges will exacerbate the difficulty of information sharing and will fuel the demand for an overall metadata management capability in enterprises.”



WHAT DOES IT MEAN TO HARNESS THE POWER OF BIG DATA?

Big data analytics has already turned entire industries on their head. For example, High Frequency Trading (HFT) has completely changed the dynamics of institutional investing. In HFT, trades are executed in microseconds based on huge amounts of information that is processed within seconds of its arrival.

As much as 70% of all trades now are HFT², and is now critical to many firms. HFT has become so effective that it is somewhat controversial and regulators have investigated HFT tactics that might be used to gain an unfair advantage. On the other hand, there is debate over findings from the SEC and CFTC that HFT contributed to volatility during the “flash crash” of May 6, 2010³. Regardless of the veracity of this association, it would be wise to consider that Big Data Analysis, if used incorrectly, may lead to a flurry of wrong decisions made very quickly.

A quick internet search for “big data analytics” will return almost 3 million results linking to articles that discuss potential and present success in many fields and verticals—from astrophysics, to healthcare, to finance, to public policy, to retail. There is a lot of excitement and a sense of urgency among executives to make sure their organizations will be ready to compete: In “Executive Advisory: CEO and Senior Executive Survey, 2011; Detailed Report,” “data-driven decision making” was the technology contribution regarded by CEOs as delivering the most strategic value to the business.⁴”



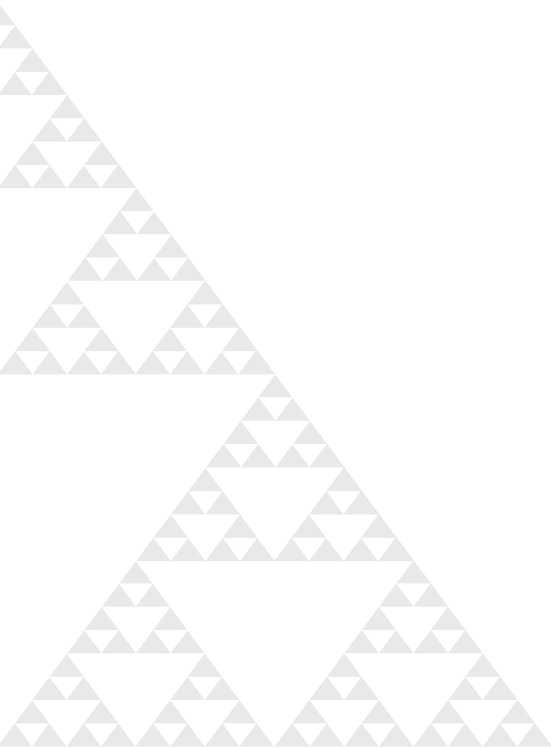
BIG DATA ANALYSIS AND STRUCTURED DATA

So far, Big Data analytics has mostly centered on information stores where there is ample metadata to analyze, like websites with extensive logs of activity, and structured data repositories (databases), where transactions are straightforward to track and analyze. In situations where metadata is available, the challenge truly is about volume and technique—how to process lots of information quickly enough and analyze it effectively to test assumptions, answer questions quickly, detect changes, and understand patterns.

However, Gartner points out, “Businesses and technologists are realizing that there is even more potential value in evaluating other types of data, some that currently exist in the enterprise and some new types of data. Many organizations have stored data for years and have never attempted to analyze it or look for patterns, simply because the business appetite for doing so didn’t exist.”

Examples of data include spreadsheets, presentations, images, audio files, video files, blueprints, and designs. The data most often resides in unstructured repositories, like file shares.

Unstructured data repositories often don’t have much existing metadata to analyze. There is usually no record of activity, no strict connection to the creators and owners of the data, and no catalog or index of what all the data contains. Ironically, this is where the most (and biggest) data actually lives: many studies show that more than 80% of organizational data is stored in unstructured repositories.



BIG METADATA – INSTRUMENTING UNSTRUCTURED DATA FOR DATA ANALYSIS

In “Mastering the Information Explosion,” we compare the digital revolution to the transportation revolution; just as more cars and airplanes necessitated traffic lights and air traffic control, more data and collaboration necessitated automated controls for making sure data is correctly accessible and correctly used.

Without automated controls, organizations have found it impossible to identify and track data owners, perform entitlement reviews to manage permissions and maintain a least privilege model, audit data access, spot abuse, and identify stale data. Automated data governance controls are now like the traffic lights in a big city—if you turn them off, everyone needs to drive very slowly or they crash.

The parallel continues: automobile and airplane movements are being tracked and analyzed at scale so we can route our cars with traffic-aware GPS, law enforcement can catch speeders by looking at automated toll-booth records, and air traffic management can make better use of airspace. The vehicles themselves are more sophisticated, providing more information about the status of each component and their overall health so that safety and efficiency increase.

In “Pattern-Based Strategy: Getting Value From Big Data,” Gartner writes, “Business leaders place a high priority on the role of technology to deliver meaningful data to the organization, so that they can make better decisions based on facts, rather than assumptions.”⁶

Just as metadata framework technology is now necessary for organizations to manage and protect data residing in unstructured and semi-structured repositories, organizations will find that without analyzing metadata, it is impossible to get maximum leverage from their data, to understand its value, identify data sets, correlate them with users, projects, and owners, understand how and when they use their data, where it should be stored, and how they can use it to collaborate more effectively.

The same metadata intelligence and information leverage will be used to enhance business processes along many vectors, optimizing workflows, connecting disparate teams, and discovering new patterns. The good news is that organizations don’t need to change the data storage platforms or their workflows in order to take advantage of this intelligence—they need to instrument the existing data stores they already use, and use metadata framework technology to normalize, synthesize, and analyze that metadata.



THE VARONIS METADATA FRAMEWORK

Ongoing, scalable data protection and management require technology designed to handle an ever-increasing volume and complexity—a metadata framework.

Four types of metadata are critical for data governance:

- User and Group Information – from Active Directory, LDAP, NIS, SharePoint, etc.
- Permissions and other File System Information – knowing which users and groups are listed on ACLs, access time stamps, file counts and sizes
- 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

The Varonis Metadata Framework non-intrusively collects this critical metadata, 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 through a web-based interface—that are then executed—with no IT overhead or manual backend processes.

The Varonis Data Governance Suite will scale to present and future requirements using standard computing infrastructure, even as the number of functional relationships between metadata entities grows exponentially. As new platforms and metadata streams emerge, they will be seamlessly assimilated into the Varonis framework, and the productive methodologies it enables for data management and protection.

VARONIS PRODUCTS

The Varonis Data Governance Suite enables secure collaboration by providing a complete metadata framework and integrated product suite for the governance of unstructured data on file servers, NAS devices and semi-structured data on SharePoint and Exchange servers. Organizations can effectively and automatically audit, manage data access control, ownership identification, classification, entitlements and authorization processes for their critical data. Digital collaboration boundaries can be safely expanded while significantly increasing IT workforce productivity for daily data protection and management tasks.

- Varonis DatAdvantage for Windows, UNIX/Linux, SharePoint, Exchange, Directory Services aggregates user, permissions, data and access event information from business critical platforms such as file systems, NAS devices, SharePoint and Exchange. Key functionalities include bi-directional visibility into permissions, a complete audit trail of access events, modeling of changes, and a patented recommendations feature that enables the removal of excess permissions. Varonis DatAdvantage ensures only the right people have access to the right data at all times.
- Varonis DataPrivilege provides a web interface that brings IT, data owners, and users together in a self-service portal for automating access authorization, entitlement reviews, policy enforcement (e.g. ethical walls), and compliance reporting. This ensures that permissions are kept current by enabling the people in the organization with knowledge about the data to authorize, review, and manage access without disrupting existing business processes.
- Varonis IDU Classification Framework provides visibility into the content of data across file systems and SharePoint sites to identify which data is sensitive, and integrates this metadata into the Varonis Metadata FrameworkTM. Classification information is then presented in the DatAdvantage interface and in its reports, enabling actionable intelligence for data governance, including a prioritized list of those folders with the most exposed permissions and containing the most sensitive data, who has access to that data, who is using it, who owns it, and how to effectively limit access.

¹ Gartner, *Big Data is Only the Beginning*, by Mark A. Beyer, Anne Lapkin, Nicholas Gall, Donald Feinberg, Valentin T. Sribar

² <http://www.dailyfinance.com/2010/10/12/what-you-shouldnt-buy-in-the-investment-market-the-pundits-ad/>

³ <http://online.wsj.com/article/SB10001424052748704029304575526390131916792.html>

⁴ Gartner, *Pattern-Based Strategy: Getting Value From Big Data*, by Yvonne Genovese, Stephen Prentice

⁵ Gartner, *Pattern-Based Strategy: Getting Value From Big Data*, by Yvonne Genovese, Stephen Prentice

⁶ Gartner, *Pattern-Based Strategy: Getting Value From Big Data*, by Yvonne Genovese, Stephen Prentice

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.

WORLDWIDE HEADQUARTERS

1250 Broadway, 31st Floor, New York, NY 10001 **T** 877 292 8767 **E** sales@varonis.com **W** www.varonis.com

UNITED KINGDOM AND IRELAND

Varonis UK Ltd., Warnford Court, 29 Throgmorton Street, London, UK EC2N 2AT **T** +44 0207 947 4160 **E** sales-uk@varonis.com **W** www.varonis.com

WESTERN EUROPE

Varonis France SAS 4, rue Villaret de Joyeuse, 75017 Paris, France **T** +33 184 88 56 00 **E** sales-france@varonis.com **W** sites.varonis.com/fr

GERMANY, AUSTRIA AND SWITZERLAND

Varonis Deutschland GmbH, Welscherstrasse 88, 90489 Nürnberg **T** +49 (0) 911 8937 1111 **E** sales-germany@varonis.com **W** sites.varonis.com/de