VARONIS

Fighting a Different Battle

Varonis' Unique Approach
to Cybersecurity

Introduction

As organizations become more data driven, they store more data in on-prem and cloud stores that employees access from anywhere with phones, tablets and laptops. The security perimeter is much less defined, and endpoints are fungible — very little data “lives” only on your phone or laptop these days.

This digital transformation has flipped the traditional security model that focused on perimeter and endpoint on its head. Instead of focusing on outside in, organizations are starting to think inside out, or data-first security.

Data protection is intuitively simple, but immensely complex.

Why is data protection intuitively simple?

I’d argue that if you can answer “yes” to these three questions, and you can answer yes continually, then your data is safe:

1. Do you know **where your important data is stored?**
2. Do you know that **only the right people have access to it?**
3. Do you know that **they’re using data correctly?**

Simple, right?

These are the three fundamental dimensions of data protection — importance, accessibility and usage. If you work in IT or IT security, you know that understanding these dimensions isn’t simple at all.

You probably also know that if you can’t answer yes to these questions, or answer them at all, they lead to other questions that have urgent ramifications for CISOs, compliance personnel, boardrooms and shareholders:

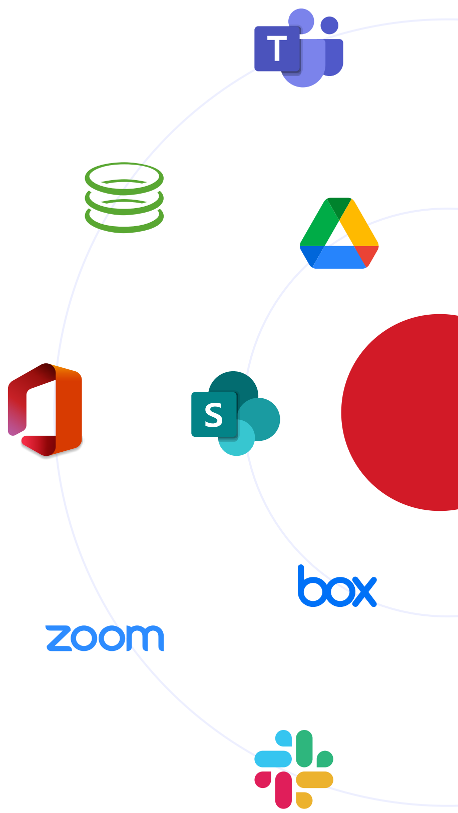
- Where is our sensitive and regulated data?
- Where is it overly accessible and most at risk?
- What is the blast radius for a compromised account or employee?
- How would we know if data was stolen, encrypted or deleted?
- Can we delete it?

The answers don’t come any easier as data keeps growing on prem and in the cloud, in applications and data stores that each have their own security models. It’s hard enough getting data protection right in one enterprise platform, let alone many at the same time.

Where is our data supposed to be?

The number of places we can put data has exploded over the past few years, and it's common for users to access their data across devices and endpoints. Endpoints now serve primarily as gateways to where the data really "lives," which is now usually in a cloud application.

Most now rely on a combination of cloud applications and infrastructure to function in addition to their on-premises infrastructure. It's becoming unusual that an organization doesn't sanction the use of a Microsoft 365, Box, Google Drive or a Slack for collaboration, a GitHub or Jira for source code, an AWS, Azure or Google Cloud to offload compute or storage, or a use a CRM solution like Salesforce.com.



Where is the important data?

Even in the realm of these sanctioned applications, the surface area is large and difficult to visualize and assess in terms of risk. Some organizations choose to focus their efforts by asking employees to tag files, or by using automation to identify or classify regulated or sensitive data with the hope of being able to prioritize data protection efforts.

It certainly makes sense to break an enormous problem down into smaller pieces, but the problem has become so large that even the smaller pieces can be overwhelming.

Most organizations are surprised at the number of sensitive files and records they find. Thousands of files here — thousands or tens of thousands there — and the list will be different tomorrow and the next day.

Those that arrive at this place without a clear plan of action can get stuck about what to do next. Some consider a brute force approach, like moving everything they find somewhere else, like to an on-premises store, deleting everything possible, or encrypting everything, thereby restricting it to employees only, or a small group that just inherited a large problem.

These don't really solve the core issue, however — ensuring that data is accessible to only the right people, also known as the principle of least privilege, and now, more popularly as Zero Trust.

To make sure access is correct for any data, sensitive or otherwise, you first need to be able to see who has access to it in the first place — that's almost always harder to answer than people realize, especially in the cloud.

Who has access to our important data? Who *should* have access?

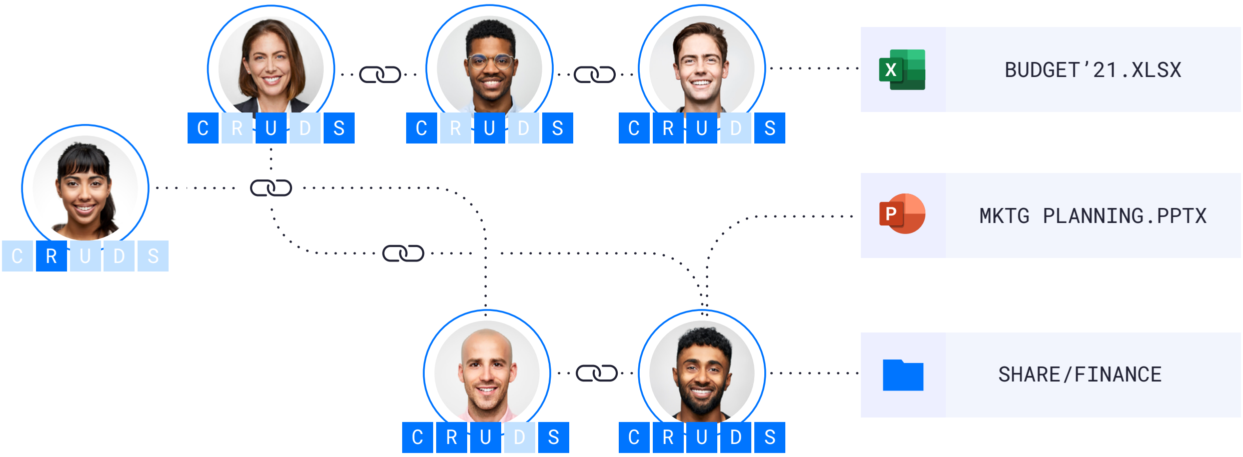
Without understanding what data is regulated or sensitive, it shouldn't come as a surprise that decisions about who should get access would have to be made without very important context. What often comes as a surprise is how hard it is to see who has access in the first place.

Access to data is handled by permissions, or access control lists. The logic is pretty uniform across applications and data stores:

- There's an **object,** like a file, folder or record.
- There are **digital identities** that correspond to users, accounts, and groups of users and accounts that can do stuff with those objects.
- There's a **description of what actions they can do,** like create, share, delete, etc.

Even though the logic is roughly the same whether you're talking about Slack, Box, SharePoint Online or on prem, or UNIX file systems, the implementations are all different:

- The objects are similar, but there are different object types depending on the application (e.g. files, sites, records, buckets).
- The users/accounts and groups are stored in multiple places — in the cloud, each data store usually has its own database of users and groups. Sometimes they connect to other accounts (like an Okta account), sometimes there are both personal and corporate accounts to track. Each of these applications can assign attributes to user and group objects, like title, role, and location.



What they can do is described differently in each application, even though they mostly net out to create, read, update, delete, and share.

On top of these differences, calculating effective rights for a given object or user can be very complex and varies greatly between stores. To determine effect rights on a given object, one must consider multiple attributes, including:

- **Object-specific permissions.** As mentioned above, each object has an access control list that lists user, group, or role entities. The range of possibilities is wide — in basic UNIX permissions, for example, there are 3 possible permissions (read, write & execute) for 3 users/groups (root, owner, group); in SharePoint online, there are 33 possible permissions that are grouped into 7 default levels (you can create additional ones), and these permission levels can be assigned to **many** users and groups on the objects.
- **Group relationships.** Groups may contain users or other "nested" groups. In order to determine effective permissions for an object or user, these relationships must be calculated. In some cases, groups in one directory service can refer to users and groups in other directory services, making this calculation more complex. For example, SharePoint Online has local groups that can contain users and groups in Azure AD.
- **Hierarchal inheritance.** In many data stores, permissions flow downward through the hierarchy, so all objects inside a folder will "inherit" access control entries from its parent or parents. Some stores allow you to stop inheritance on child objects, but not all. Box, for example, only supports adding access control entries on child objects, so a child object can never be less permissive than its parent(s).
- **Roles & role hierarchies.** Access may be granted to objects based on a role. Roles may contain other roles and are assigned differently in different applications. For example, in AWS, roles are usually assumed as needed, while roles in Salesforce tend to be assigned more statically.
- **System-wide settings.** Some settings affect access to all objects. In Google Drive, for example, the Link Sharing setting overrides all permissions and makes every newly created object accessible to the entire domain. In Salesforce, the "organization-wide defaults" (OWD's) set base-level access for all objects.

To visualize access, all these attributes and functional relationships must be pre-calculated and normalized across data stores and applications. Without this kind of automation, determining who has access to an object or what user or account has access to (the effective blast radius in an attack) is impossibly time consuming, impairing day-to-day tasks that range from incident response to troubleshooting to audit reporting.

Is understanding access activity any easier than understanding permissions?

It is not.

When considering data security, there are several types of events that pertain directly to data protection.

- **Data access events.** The most security-relevant activities involve direct interaction with data — when users create, read, change/update, delete, or share data. Unfortunately, each application and data store has its own way of recording (or not) how users directly interact with data. In Salesforce logs, for example, data access activity does include which object was accessed.
- **Access control changes and configuration changes** that affect the accessibility of data are also highly relevant. Access control changes are also reported differently and are incomplete without knowledge of the user and groups that they reference. For example, many systems that log permissions only log that an access control list (ACL) was changed — not which entries were changed. Additionally, changes to the objects referenced in the ACL may not be recorded by the file system or application — these may need to be recorded in the directory service (e.g. Azure Active Directory). Configuration changes are similarly complex, even with respect to data accessibility. GPO changes in Active Directory can affect all sorts of important things, like password policies and endpoint functionality. GitHub, for example, records changes to the accessibility of code repositories, but does not indicate what the changes were.
- **Authentication events** can provide meaningful context about which users connected to the application or data store, from where, and with what kind of authentication (e.g. single or multi-factor). Authentication events vary between directory services and applications.
- **Perimeter events.** In on-prem infrastructure, perimeter signals from DNS, VPN gateways and proxies provide insight into unusual connections into and out of the environment. Events from perimeter devices are voluminous and non-uniform — it can be tempting to start pulling in telemetry from many places, but you must be careful not to break the ratio of signal to noise. What is most practical is telemetry that's relevant from the data perspective, like DNS to see infiltration and web proxy to see exfiltration. See **5 Ways Your SIEM Is Failing You** for more details.

Because data stores and applications describe these events so differently, it is very difficult to answer questions across them. Just to understand what data an employee accessed on a given day, or what access control changes an administrator made become research projects instead of simple queries.

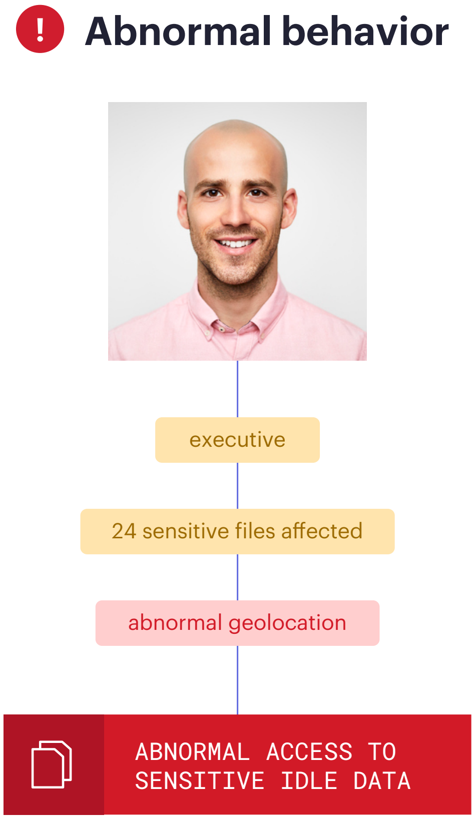
What about alerting?

Without a uniform, normalized event stream, rule-based alerting is challenging, and behavior-based alerting is either limited to a single application, or off the table entirely. When it becomes to modeling behavior to build profiles, events also need to be enriched for AI to have a meaningful set of facets to evaluate. For example, if you wanted to create a simple threshold-based alert to fire when someone deleted, updated or accessed more than 1000 files or objects in a 5-minute period, without a reliable, uniform event stream, you'd probably have to create an alert for every application. If you wanted to have that alert fire when the total number of events exceeded 1000 file operations in a 5-minute period across all stores, you're already into some pretty advanced queries.

If you want to get a little more advanced, like getting an alert if 1,000 sensitive files are accessed in a 5-minute period across data stores, you're going to need to enrich the events with file sensitivity information before your alert logic processes. You can start to see how important clean, enriched events are in threshold-based alerts — they're even more important for AI-driven alerts.

Clean, enriched event streams are key to building behavioral baselines, or peace-time profiles, that AI can evaluate for deviations. These data-centric profiles yield alerts with very high signal to noise ratios. For example, when an executive that normally accesses a dozen of files a week, very few of which are important, starts accessing dozens of important files neither they nor their peers access, maybe from a device never associated with that user, or from a location they don't normally visit — this deserves immediate attention.

Behavioral analytics requires relevant, enriched, reliable behavior to analyze. That's why security technologies that analyze unreliable, noisy streams inevitably fail.


Abnormal behavior
executive
24 sensitive files affected
abnormal geolocation
ABNORMAL ACCESS TO SENSITIVE IDLE DATA

Without three dimensions, **you really fall flat**

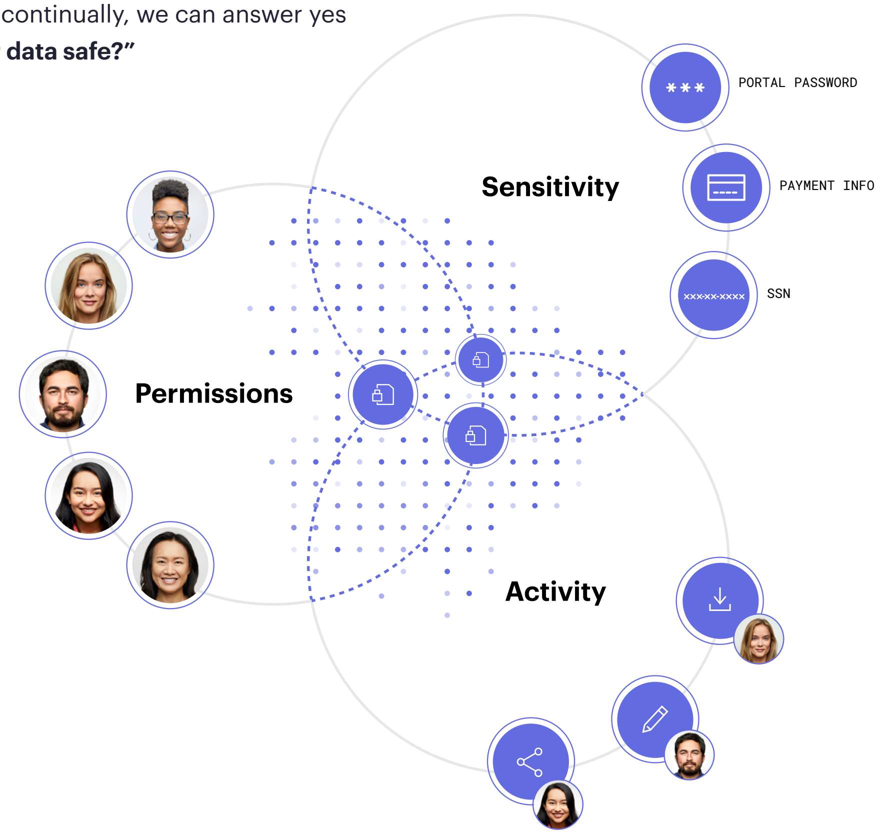
When you can't see any of the three dimensions we've covered so far — sensitivity, permissions, and activity — many start out thinking they can get by with one or two. If we explore the different combinations, however, we'll quickly see how much more powerful it is to have all the dimensions together.

As I mentioned previously, if you only know what data is sensitive, you won't know where it's concentrated, or exposed without the dimension of permissions. Without activity, you'd never know how to safely fix any exposure you found, or whether sensitive data was being stolen or even whom you should speak with about it. If you only have activity, you will be able to see what data has been taken after a breach, or even alert on some behavioral deviations, but you won't know how sensitive the data was, who else was in a position to access it, or whether it's incorrectly exposed to everyone in the company (or the internet) in the first place.

When it comes to data protection, each of these dimensions are needed to answer the critical questions that started this paper:

1. Do you know **where your important data is stored?**
2. Do you know that **only the right people have access to it?**
3. Do you know that **they're using data correctly?**

When we can answer yes to these questions continually, we can answer yes to the most important question of all: **"Is our data safe?"**



Ready to experience the Varonis difference?

Reduce your risk without taking any. Contact our team to learn what will be covered in your **free** data risk assessment.

Contact us

ABOUT VARONIS

Varonis is a pioneer in data security and analytics, fighting a different battle than conventional cybersecurity companies. Varonis focuses on protecting enterprise data on premises and in the cloud: sensitive files and emails; confidential customer, patient and employee data; financial records; strategic and product plans; and other intellectual property.

The Varonis Data Security Platform detects insider threats and cyberattacks by analyzing data, account activity and user behavior; prevents and limits disaster by locking down sensitive and stale data; and efficiently sustains a secure state with automation. With a focus on data security, Varonis serves a variety of use cases including governance, compliance, classification, and threat analytics. Varonis started operations in 2005 and has thousands of customers worldwide — comprised of industry leaders in many sectors including technology, consumer, retail, financial services, healthcare, manufacturing, energy, media, and education.