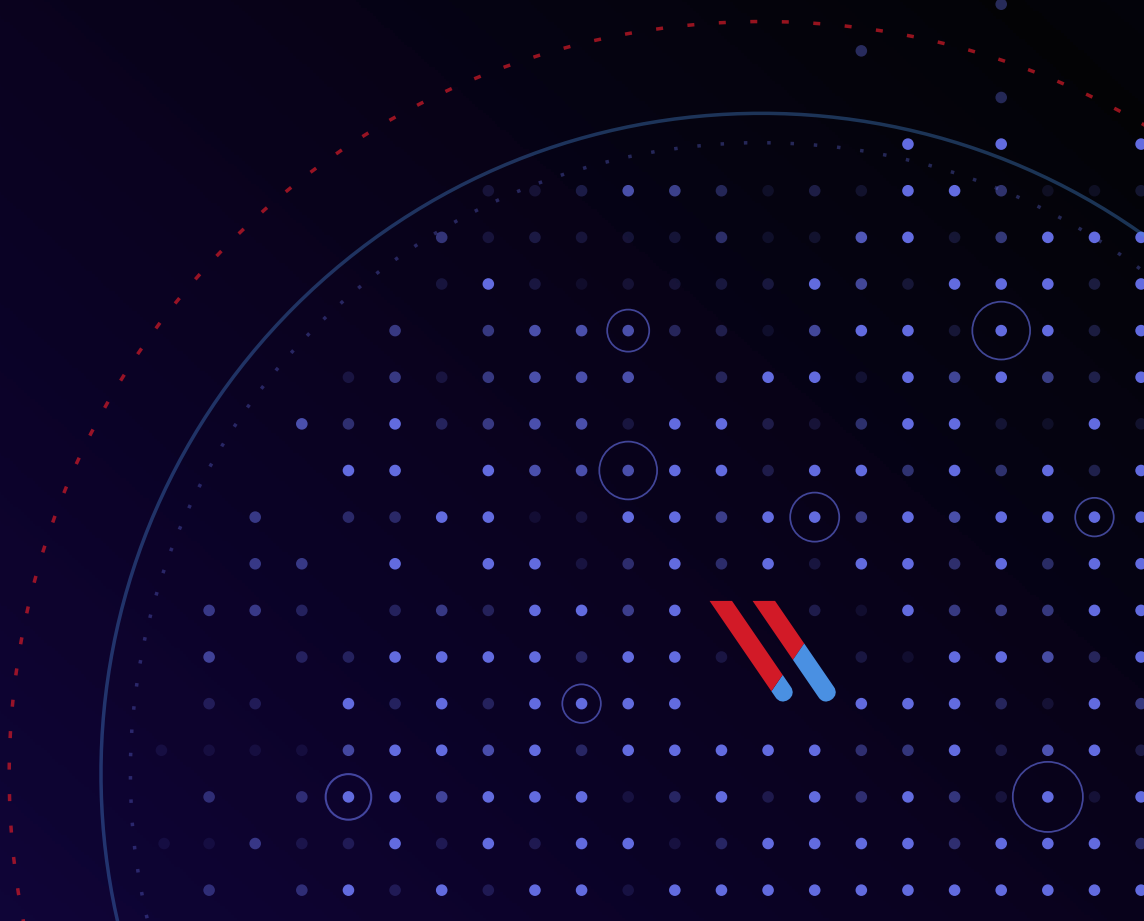
VARONIS

Changer de cheval de bataille

L'approche unique de Varonis sur la cybersécurité



Introduction

Alors que les entreprises utilisent de plus en plus de données dans leurs activités, elles en stockent davantage que ce soit sur site ou dans le cloud et leurs employés y ont accès sur leurs téléphones, tablettes et ordinateurs portables, quel que soit l'endroit. Le périmètre de sécurité est bien plus flou et les terminaux sont facilement interchangeables : aujourd'hui, très peu de données se limitent à un seul appareil.

Cette transformation numérique a bouleversé le modèle traditionnel de sécurité qui se concentrait principalement sur le périmètre et le terminal. Au lieu de chercher à se protéger de l'extérieur, les entreprises commencent à penser dans l'autre sens et à s'intéresser à la sécurité qui passe par les données.

La protection des données est simple en théorie, mais immensément complexe en réalité.

Pourquoi la protection des données est-elle simple en théorie ?

Si vous répondez Oui aux trois questions suivantes et que vous pouvez le faire en permanence, alors vos données sont effectivement protégées :

1. Savez-vous **où sont stockées vos données importantes ?**
2. Êtes-vous sûr **que seules les personnes appropriées y ont accès ?**
3. Êtes-vous sûr **qu'elles utilisent ces données correctement ?**

Simple, n'est-ce pas ?

Ce sont les trois dimensions fondamentales de la protection des données : l'importance des données, leur accessibilité et leur utilisation. Si vous travaillez dans l'informatique ou la sécurité de l'information, vous savez que la compréhension de ces dimensions est loin d'être aussi simple.

Vous savez aussi probablement que si vous ne pouvez pas répondre par l'affirmative à ces questions ou que vous ne pouvez pas y répondre du tout, que d'autres questions surgissent, qui ont des conséquences urgentes pour les RSSI, le personnel en charge de la conformité, le conseil d'administration et les actionnaires :

- Où se trouvent vos données sensibles et soumises aux réglementations ?
- Où sont-elles trop exposées et le plus à risque ?
- Quel est le rayon d'exposition pour un compte compromis ou un employé piraté ?
- Comment saurions-nous si nos données étaient dérobées, chiffrées ou supprimées ?
- Pouvons-nous les supprimer ?

Les réponses à ces questions ne sont pas plus simples à mesure que les données s'accumulent sur site et dans le cloud, dans des applications et des dépôts de données, qui ont tous leur propre modèle de sécurité. Il est déjà suffisamment difficile d'implémenter correctement la protection des données dans une seule plateforme d'entreprise, alors comment faire avec plusieurs plateformes en même temps ?

Où sont censées se trouver nos données ?

Depuis quelques années, le nombre d'endroits dans lesquels nous accumulons des données a explosé et les utilisateurs ont l'habitude d'accéder à leurs données sur plusieurs appareils et terminaux. Les terminaux sont aujourd'hui principalement des passerelles qui nous permettent d'accéder aux données, là où elles se trouvent, c'est-à-dire généralement dans une application cloud.

La plupart des entreprises dépendent aujourd'hui d'une combinaison d'applications et d'infrastructure cloud, en plus de leur infrastructure sur site. Il est rare aujourd'hui qu'une entreprise n'autorise pas l'utilisation de Microsoft 365, Box, Google Drive ou Slack pour la collaboration, de GitHub ou de Jira pour le code source, d'AWS, Azure ou Google Cloud pour libérer des ressources de calcul ou du stockage ou l'utilisation d'une solution de CRM comme Salesforce.com.

Où se trouvent les données importantes ?

Même au sein de ces applications approuvées, la zone de risque est vaste et difficile à visualiser et à évaluer. Certaines entreprises choisissent de concentrer leurs efforts et demandent à leurs employés de marquer les fichiers par des balises, ou utilisent des solutions d'automatisation pour identifier ou classer les données réglementées ou sensibles, dans l'espoir de pouvoir prioriser les mesures de protection des données.

Fractionner un problème colossal en parties plus petites et gérables tombe sous le sens, mais le problème est si vaste aujourd'hui que même ses plus petites parties représentent un challenge.

La plupart des entreprises sont surprises par la quantité de fichiers et de données sensibles qu'elles découvrent. Des milliers de fichiers par-ci, des dizaines ou des centaines de milliers de fichiers par-là, et la liste change d'un jour à l'autre.

Ceux qui arrivent à ce stade de l'analyse sans avoir de plan d'action précis peuvent se retrouver un peu bloqués. Certains envisagent une approche pour le moins radicale, comme déplacer tout ce qu'ils trouvent dans un autre lieu, comme un dépôt de données sur site, supprimer tout ce qui est superflu ou tout chiffrer, pour restreindre les fichiers aux employés uniquement ou à un groupe plus petit qui se retrouve alors avec un gros problème sur les bras.

Cependant, ces mesures ne résolvent pas le problème principal, à savoir garantir que les données ne puissent être accessibles que par les bonnes personnes. Un principe connu sous le nom de politique du moindre privilège, et plus communément modèle « Zero Trust » aujourd'hui.

Pour garantir que l'accès soit approprié pour chaque donnée, qu'elle soit sensible ou non, vous devez premièrement déterminer qui y a accès, ce qui est probablement encore plus difficile à dire, en particulier quand elles se trouvent dans le cloud.

Qui a accès à nos données importantes ? Qui *devrait* y avoir accès ?

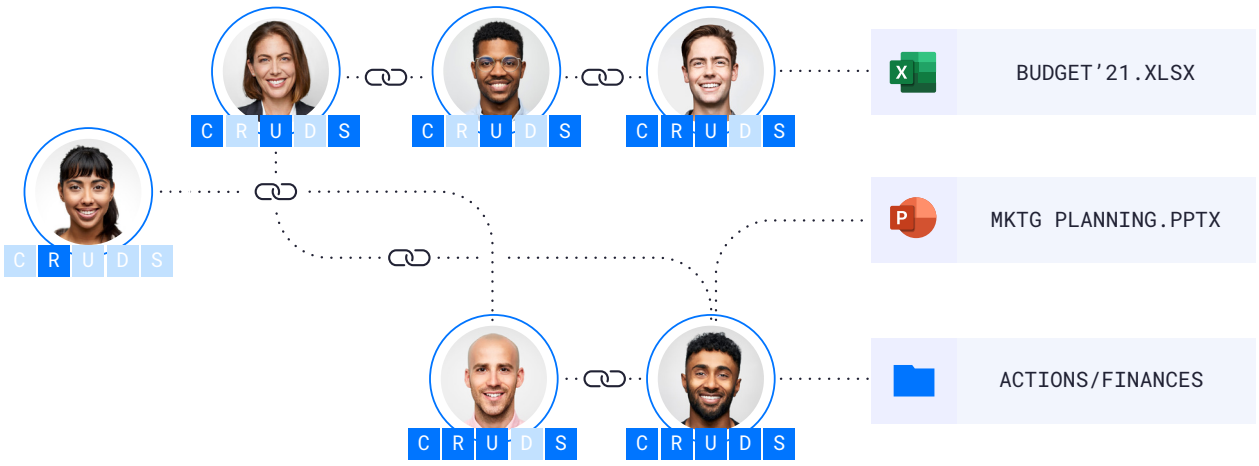
Quand on ne sait pas quelles données sont réglementées ou sensibles, on comprend facilement que décider quels utilisateurs devraient y avoir accès est un choix qui se fait alors à l'aveuglette. Mais il est parfois surprenant de constater à quel point il est difficile de savoir qui a accès à ces données.

L'accès aux données est accordé par des autorisations ou des listes de contrôles d'accès. La logique est similaire pour toutes les applications et les dépôts de données :

- Nous avons un **objet,** comme un fichier, un dossier ou un enregistrement.
- Puis nous avons des **identités numériques** qui correspondent à des utilisateurs, des comptes et des groupes d'utilisateurs et de comptes qui peuvent agir sur ces objets.
- Enfin, nous avons une **description de ces actions,** comme la création, le partage, la suppression, etc.

Même si la logique est plus ou moins identique, qu'il s'agisse de Slack, de Box, de SharePoint Online, de données sur site ou de systèmes de fichiers UNIX, toutes les implémentations sont uniques :

- Les objets sont similaires, mais les types d'objet sont différents selon l'application (par ex. fichiers, sites, enregistrements, compartiments).
- Les utilisateurs/comptes et les groupes sont stockés dans plusieurs endroits : dans le cloud, chaque dépôt de données possède généralement sa base de données d'utilisateurs et de groupes. Parfois ils sont reliés à d'autres comptes (comme un compte Okta), parfois il faut surveiller à la fois un compte personnel et un compte professionnel. Chacune de ces applications peut attribuer des caractéristiques aux utilisateurs et groupes, comme un titre, un rôle ou un lieu.



Ce que chaque utilisateur ou groupe peut faire est décrit différemment dans chaque application, même si les actions se résument souvent à créer, lire, modifier, supprimer et partager.

Outre ces différences, le calcul des autorisations réelles sur un objet ou un utilisateur donné peut être très complexe et varie fortement entre chaque dépôt de données. Pour déterminer les autorisations réelles sur un objet donné, il faut prendre en compte plusieurs attributs, comme :

- **Les autorisations spécifiques à l'objet.** Comme nous venons de le mentionner, chaque objet dispose d'une liste de contrôle d'accès qui énumère les utilisateurs, les groupes ou les rôles. L'éventail de possibilités est vaste. Dans les systèmes UNIX de base, il existe 3 autorisations possibles (lecture, écriture et exécution) pour 3 utilisateurs/groupes (racine, propriétaire, groupe). Dans SharePoint Online, il existe 33 autorisations possibles, regroupées en 7 niveaux par défaut (vous pouvez en créer d'autres) et ces niveaux d'autorisation peuvent être attribués à **plusieurs** utilisateurs et groupes sur les objets.
- **Les relations entre les groupes.** Les groupes peuvent contenir des utilisateurs ou d'autres groupes « imbriqués ». Pour déterminer les autorisations réelles pour un objet ou un utilisateur, ces relations doivent être calculées. Dans certains cas, les groupes d'un service d'annuaire peuvent faire référence à des utilisateurs ou des groupes d'autres services d'annuaire, ce qui rend ce calcul d'autant plus complexe. Ainsi, SharePoint Online possède des groupes locaux qui contiennent des utilisateurs et des groupes d'Azure AD.
- **Les paramètres hérités de la structure hiérarchique.** Dans de nombreux dépôts de données, les autorisations découlent de la hiérarchie de haut en bas. Tous les objets au sein d'un dossier « héritent » des entrées de contrôle d'accès du dossier ou des dossiers parents. Certains dépôts de données vous permettent d'empêcher le transfert des contrôles d'accès aux objets « imbriqués », mais pas tous. Box, par exemple, prend seulement en charge l'ajout d'entrées de contrôles d'accès sur les objets imbriqués, ce qui signifie qu'un objet imbriqué ne pourra jamais avoir plus d'autorisations que son ou ses dossiers parents.
- **Les rôles et les hiérarchies entre rôles.** Les accès aux objets peuvent être accordés en fonction d'un rôle donné. Les rôles peuvent contenir d'autres rôles et sont attribués différemment selon l'application. Par exemple, dans AWS, les rôles peuvent être changés selon les besoins, tandis que les rôles dans Salesforce sont attribués de façon plus statique.
- **Les paramètres du système global.** Certains paramètres affectent l'accès à tous les objets. Dans Google Drive, par exemple, le partage de lien écrase toutes les autorisations et fait en sorte que tous les nouveaux objets créés soient accessibles sur le domaine tout entier. Dans Salesforce, les paramètres par défaut dans toute l'organisation définissent un accès de base pour tous les objets.

Pour visualiser l'accès, tous ces attributs et ces relations fonctionnelles doivent être pré-calculés et standardisés à travers tous les dépôts de données et les applications. Sans ce genre d'automatisation, déterminer qui a accès à un objet ou quel utilisateur ou compte y a accès (le rayon d'exposition réel dans une attaque) est incroyablement fastidieux et impacte les tâches quotidiennes qui vont de la réponse aux incidents au dépannage en passant par les rapports d'audit.

Est-il plus facile de comprendre l'activité autour des accès que de comprendre les autorisations ?

Pas du tout.

En matière de sécurité des données, plusieurs types d'événements relèvent directement de la protection des données.

- **Événements d'accès aux données.** Les activités les plus pertinentes en matière de sécurité sont celles qui impliquent une interaction directe avec les données, c'est-à-dire quand les utilisateurs créent, lisent, modifient/mettent à jour, suppriment ou partagent des données. Malheureusement chaque application et dépôt de données a sa propre façon d'enregistrer (ou non) les interactions directes des utilisateurs avec les données. Dans les logs de Salesforce par exemple, cela inclut le nom de l'objet auquel l'utilisateur a eu accès.
- **Les changements de contrôle d'accès et de configuration** qui affectent l'accessibilité des données sont aussi extrêmement pertinents. Les changements de contrôle d'accès sont également enregistrés différemment et ne sont pas complets quand ils n'indiquent pas à quel utilisateur ou groupe ils font référence. Par exemple, de nombreux systèmes qui enregistrent les autorisations enregistrent seulement le fait qu'une liste de contrôle d'accès a été modifiée et non quelles entrées ont été modifiées spécifiquement. De plus, les changements apportés aux objets auxquels la liste fait référence ne sont peut-être pas enregistrés par le système de fichier ou l'application. Ils devront possiblement être enregistrés dans le directory service (p. ex. Azure Active Directory). Les changements de configuration sont tout aussi complexes, même en matière d'accessibilité. Les changements sur les GPO (objets de politique de groupe) dans Active Directory peuvent affecter toutes sortes de choses, comme les politiques en matière de mots de passe ou la fonctionnalité des terminaux. GitHub par exemple, enregistre les changements apportés à l'accessibilité des référentiels de code, mais ne mentionne pas la nature des changements.
- **Les événements d'authentification** peuvent fournir du contexte utile : quels utilisateurs se sont connectés à l'application ou au dépôt de données, à partir de quel terminal et avec quelle forme d'authentification (un seul ou plusieurs facteurs). Les événements d'authentification varient entre services d'annuaire et applications.
- **Les événements du périmètre.** Dans une infrastructure sur site, les signaux du périmètre provenant des DNS, des passerelles VPN et des proxies donnent des informations sur les connexions inhabituelles entrantes et sortantes. Les événements survenant sur les appareils du périmètre sont volumineux et non uniformes. Il peut être tentant de commencer à enregistrer la télémétrie de plusieurs endroits différents, mais il faut faire attention à ne pas perturber le ratio signal/bruit. Un relevé télémétrique pertinent par rapport aux données serait bien plus pratique, comme les DNS pour détecter les infiltrations et les proxies Web pour détecter les exfiltrations. Voir l'article **SIEM : découvrez ses 5 défauts...** pour plus de détails.

Les dépôts de données et les applications décrivent ces événements de manière si différente qu'il est très difficile de répondre aux questions de manière uniforme. Loin d'être une simple requête, comprendre simplement à quelles données un employé a eu accès pour une journée donnée ou quels changements de contrôles d'accès un administrateur a effectués devient vite un projet ambitieux.

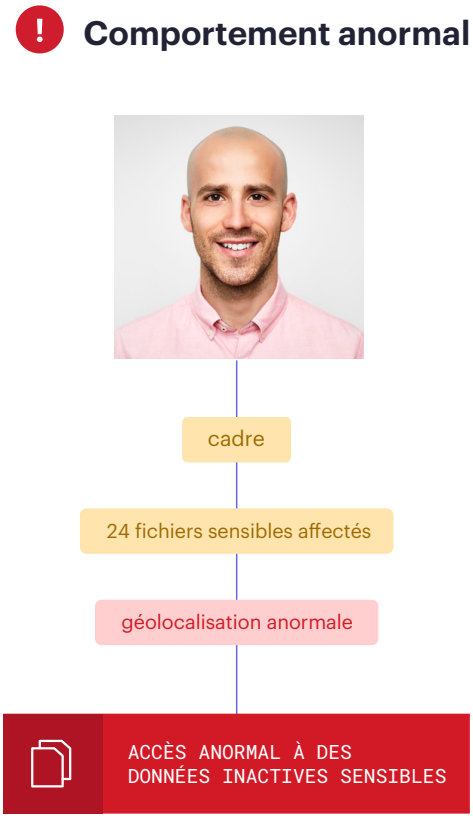
Qu’en est-il des systèmes d’alerte ?

Sans un flux d’événement standardisé et uniforme, les alertes basées sur des règles sont difficiles à paramétrer et celles basées sur des comportements sont soit limitées à une seule application soit totalement irréalisables. Quand il s’agit de modéliser les comportements pour créer des profils, les événements doivent également être enrichis pour que l’intelligence artificielle puisse avoir un ensemble concret de facettes à évaluer. Par exemple, si vous voulez créer une alerte simple basée sur un seuil, qui se déclenche quand quelqu’un supprime, modifie ou accède à plus de 1 000 fichiers ou objets en moins de 5 minutes, sans flux d’événement fiable et uniforme, il faudrait probablement créer une alerte pour chaque application. Et si vous voulez que cette alerte se déclenche quand le nombre total d’événements dépasse le millier d’opérations sur des fichiers en moins de 5 minutes parmi tous les dépôts de données, la requête est déjà bien sophistiquée.

Si vous voulez aller plus loin, comme obtenir une alerte si un millier de fichiers sensibles sont ouverts en moins de 5 minutes parmi tous vos dépôts de données, il faudra enrichir les événements avec des informations sur la sensibilité des fichiers avant de lancer la logique de l’alerte. Vous voyez donc l’importance d’avoir des événements propres et enrichis dans les alertes basées sur des seuils. Et c’est encore plus valable dans le cas des alertes déclenchées par IA.

Les flux d’événements propres et enrichis sont déterminants pour créer des profils comportementaux en « temps normal » que l’IA peut ensuite comparer aux comportements inhabituels. Ces profils centrés sur les données génèrent des alertes dont le rapport signal/bruit est très élevé. Par exemple, quand un cadre dirigeant qui accède normalement à une dizaine de fichiers par semaine dont quelques-uns seulement sont importants, consulte plusieurs dizaines de ces derniers auxquels ni lui ni ses collègues n’accèdent, potentiellement à partir d’un appareil qui n’a jamais été associé à cet utilisateur ou depuis un lieu inhabituel, on comprend que cela mérite attention.

Pour fonctionner, l’analyse comportementale nécessite des comportements pertinents, enrichis et fiables. C’est pourquoi les technologies de sécurité qui analysent des flux hétérogènes et peu fiables sont vouées à l’échec.



Sans ces trois dimensions, **vous tournez en rond**

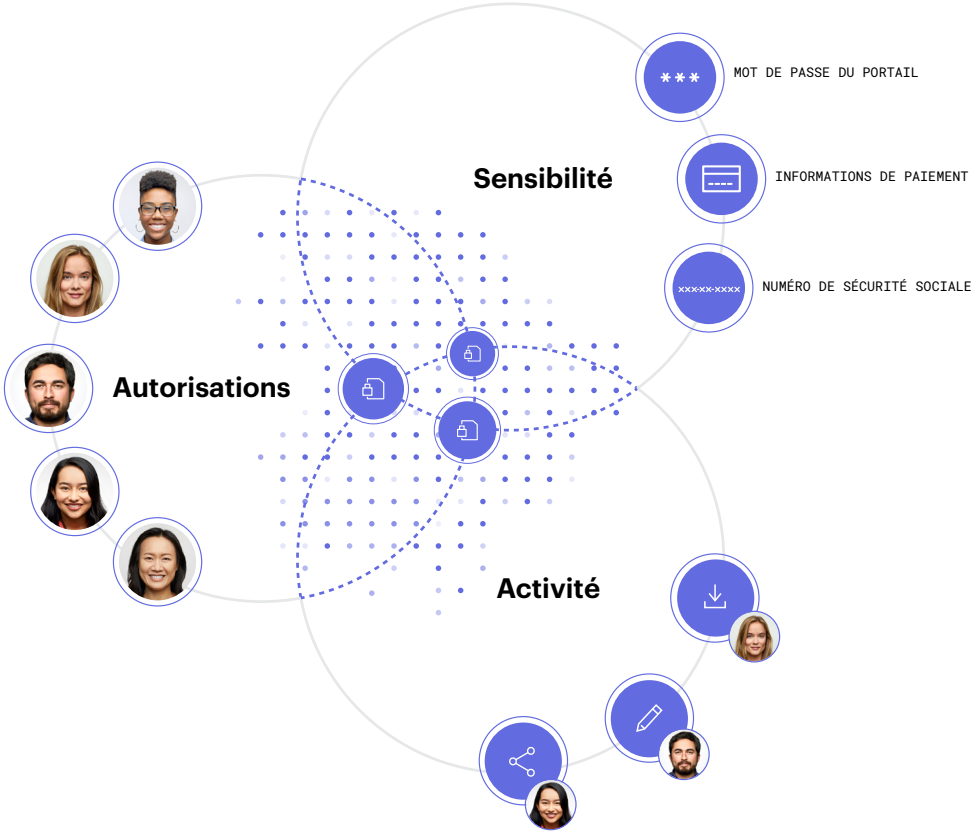
Sans visibilité sur les trois dimensions que nous venons de voir – le niveau de sensibilité, les autorisations et les activités – beaucoup ont tendance à penser qu'ils peuvent s'en sortir avec une ou deux des dimensions. Si nous explorons les différentes combinaisons cependant, nous verrons rapidement qu'il est bien plus puissant d'avoir une visibilité sur toutes ces dimensions en même temps.

Comme nous l'avons mentionné plus haut, si vous savez uniquement lesquelles de vos données sont sensibles, vous ne saurez pas où elles sont concentrées ni exposées sans la dimension des autorisations. Sans visibilité sur les activités, vous ne sauriez pas comment corriger toute exposition détectée ni si des données sensibles ont été dérobées ni même à qui en parler. Si vous disposez uniquement des informations sur l'activité, vous pourrez voir quelles données ont été dérobées après une faille ou déclencher une alerte sur certains comportements inhabituels, mais vous ne saurez pas si les données étaient sensibles, qui d'autre a pu y accéder ni si elles étaient exposées accidentellement à tout le personnel de l'entreprise (ou sur Internet).

En matière de protection des données, chacune de ces dimensions est nécessaire pour répondre aux questions essentielles énumérées au début de ce livre blanc :

1. Savez-vous **où sont stockées vos données importantes ?**
2. Êtes-vous sûr **que seules les personnes appropriées y ont accès ?**
3. Êtes-vous sûr **qu'elles utilisent ces données correctement ?**

Quand vous êtes en mesure de répondre Oui à ces questions en permanence, vous pouvez répondre Oui à la plus importante question de toutes : **« Vos données sont-elles en sécurité ? »**



VARONIS

Prêt à découvrir Varonis par vous-même ?

Réduisez votre risque sans en prendre aucun. Contactez notre équipe pour découvrir ce que nous aborderons dans notre évaluation des risques sur vos données, **gratuitement**.

Contactez-nous

À PROPOS DE VARONIS

Varonis est un pionnier de la sécurité et de l'analyse des données, et mène un autre combat que les entreprises de cybersécurité classiques. Varonis protège les données d'entreprise conservées sur site et dans le cloud : fichiers sensibles et e-mails ; données confidentielles sur les clients, patients et employés ; dossiers financiers ; plans stratégiques et produit ; et autre propriété intellectuelle.

La plate-forme de sécurité des données Varonis détecte les menaces internes et les cyberattaques en analysant les données, l'activité des comptes et le comportement des utilisateurs. Il prévient et limite les catastrophes en verrouillant les données sensibles et obsolètes et maintient un état sécurisé grâce à l'automatisation. Axé sur la sécurité des données, Varonis répond aux besoins de différents cas d'utilisation tels que gouvernance, conformité, classification et analyse des menaces. Fondée en 2005, la société Varonis compte des milliers de clients dans le monde— parmi eux figurent des leaders de nombreux secteurs : technologie, grande consommation, vente au détail, services financiers, santé, fabrication, énergie, médias et éducation.