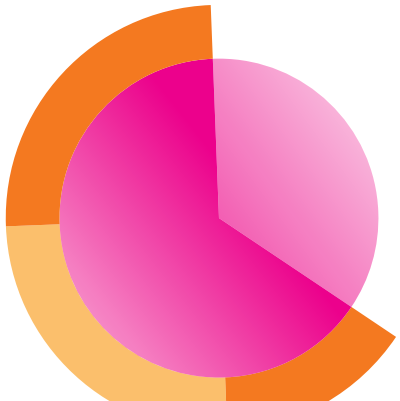
Le Guide essentiel des **données d'infrastructure**

splunk>
turn data into doing™

Données chronologiques. Données de flux. Dark data.

Nous savons tous qu'elles restent sous-exploitées et sous-estimées dans la plupart des entreprises du monde. Bien que les décisions axées sur les données fassent l'objet de discussions constantes, les entreprises de toutes les tailles ne parviennent pas à capturer et à exploiter efficacement les mines de données générées chaque jour, qu'elles proviennent des utilisateurs, de ressources professionnelles extérieures ou de leurs propres dispositifs réseau. En effet, la plupart des décideurs IT et métier estiment que **55 % de leurs données sont des dark data**, des informations dont vous ignorez l'existence ou que vous ne pouvez pas vraiment utiliser.

C'est un gaspillage considérable. Des renseignements essentiels sur votre IT, votre sécurité et vos activités se cachent au cœur de ces données. Les données contiennent les archives complètes de toute l'activité et de tous les comportements de vos clients, utilisateurs, transactions, applications, serveurs, réseaux, dispositifs mobiles et autres. Des informations cruciales sur les configurations, les API, les files de message, les résultats des diagnostics, les données des capteurs industriels, etc. :tout est là, il suffit d'y puiser l'information de la bonne façon.



Avec la bonne approche, les données permettent de facilement :

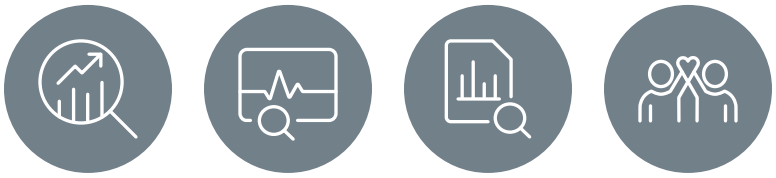
- prendre des décisions mieux informées sur tous les aspects de votre entreprise,
- administrer vos opérations plus efficacement,
- optimiser l'expérience des utilisateurs et des clients,
- détecter les traces de fraude, voire l'empêcher totalement,
- mettre au jour des désastres potentiels avant qu'ils ne se produisent,
- détecter les tendances cachées qui aideront votre entreprise à prendre une longueur d'avance sur la concurrence,
- transformer tous les utilisateurs des données en héros,
- et bien plus encore.

Le défi de l'exploitation des grands volumes de données que recueillent la plupart des entreprises réside dans le fait qu'elles sont générées dans un incroyable éventail de formats et que les outils traditionnels de supervision et d'analyse n'ont pas été conçus pour les gérer. Beaucoup d'outils sont incapables de gérer la variété des structures, des sources et des échelles temporelles des données. Et cela dépasse le simple cadre des données machine. Mais l'intérêt de puiser dans vos données est extraordinaire, et c'est là que Splunk intervient.

Avec Splunk, vous pouvez compter sur vos données pour toutes les questions, décisions et actions de votre entreprise, afin de produire des résultats pertinents. Contrairement aux autres plateformes, Splunk est capable de prendre les données de n'importe quelle source et d'orienter des actions concrètes pour le bien de l'entreprise, de la supervision de l'infrastructure IT et de la sécurité aux DevOps, en passant par la supervision et la gestion de la performance des applications.

Le concept Data-to-Everything en pratique

Utilisez les données pour :



Investiguer **Superviser** **Analyser** **Agir**

Les entreprises qui extraient le plus de valeur de leurs données sont celles qui parviennent à prendre des types de données disparates, à les enrichir et à en tirer des réponses. Mais ne pas savoir quelles données importer peut arrêter une entreprise avant même d'avoir commencé.

Familiarisez-vous avec les scénarios d'utilisation classiques de la sécurité, des opérations IT, de l'analyse commerciale, des DevOps, de l'Internet des objets (IoT), etc. et avec les types et les sources de données concernés pour prendre un bon départ.

Voici un exemple :

1. La commande d'un client n'a pas abouti
2. Le client a appelé le service d'assistance pour résoudre le problème
3. Après une trop longue attente au téléphone, il a abandonné et publié un tweet négatif sur l'entreprise

À quoi ressemblent les données machine ?

Figure 1 : Les données peuvent provenir de sources variées et, au premier abord, ressemblent souvent à du texte indéchiffrable.

Les données machine contiennent des informations stratégiques

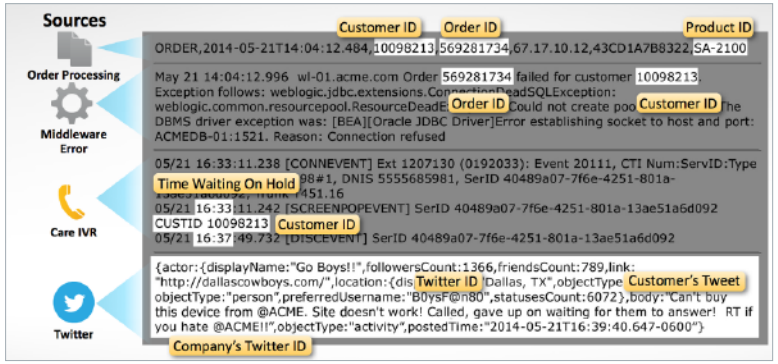


Figure 2 : La valeur des données machine se cache au cœur de ce texte apparemment aléatoire.

Les données machine contiennent des informations stratégiques

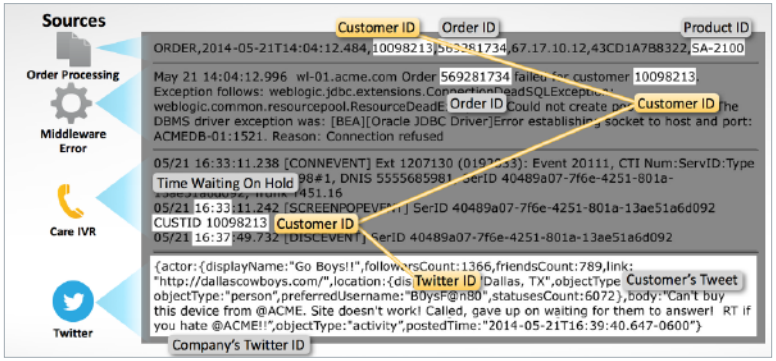


Figure 3 : En établissant des corrélations entre différents types de données, vous pouvez obtenir de précieux renseignements sur ce qui se passe dans votre infrastructure, visualiser les menaces de sécurité et même utiliser ces renseignements pour prendre de meilleures décisions business.

En prenant toutes les données impliquées dans le processus, autrement dit en extrayant les informations des systèmes de traitement des commandes, du middleware, de réponse vocale interactive et celles de Twitter, une entreprise peut bénéficier d'une vue complète des problèmes de l'expérience client.

Données d'infrastructure

Ce manuel offre un aperçu de haut niveau sur la valeur que vous pouvez extraire des données générées par vos infrastructures physiques et virtuelles dans le cadre de leur fonctionnement normal. Ces données peuvent appuyer un large éventail de cas d'utilisation allant de la supervision des déploiements cloud à l'identification des tentatives d'infiltration, en passant par la réduction des vulnérabilités.

Certes, chaque entreprise a ses propres besoins et les sources de données varient selon les fournisseurs, les produits et les infrastructures. Néanmoins, ce guide vous indique où chercher les types de données machine capables d'enrichir vos cas d'utilisation dans les domaines de l'IT, de la sécurité, de l'IoT et des business analytics.

De nombreuses sources de données suggérées dans ce guide peuvent renforcer plusieurs cas d'usage : c'est d'ailleurs en grande partie ce qui fait la valeur considérable des données machine.

Sécurité et conformité

Opérations IT, livraison des applications et DevOps

Internet des objets

Business Analytics

Sommaire

Données des infrastructures virtuelles

AWS Services

Cas d'utilisation : Sécurité et conformité, Opérations IT

Exemples : CloudTrail, CloudWatch, Config, S3

AWS est l'infrastructure de cloud public la plus vaste et la plus employée. Elle fournit des services à la demande de puissance de calcul, de stockage, de base de données, de big data et d'applications, en appliquant une tarification à la consommation. AWS peut remplacer les infrastructures traditionnelles de serveurs virtuels d'entreprise, qui exécutent les logiciels sur des machines virtuelles (VM) distinctes, ou héberger des applications natives élaborées à partir d'une palette de services AWS. AWS propose un large éventail de services de gestion, d'automatisation, de sécurité, de réseau et de supervision qui permettent de déployer, redimensionner, mettre hors service, auditer et administrer son environnement AWS, ainsi que ses abonnements et ses applications hébergées.

Cas d'utilisation

Sécurité et conformité : Les données de sécurité des services AWS incluent les événements et les tentatives de connexion et de déconnexion, les appels d'API et les logs des pare-feux du réseau et des applications web.

Opérations IT : Les services AWS fournissent des types de données sur les systèmes et les services qui sont comparables à ceux d'une infrastructure informatique traditionnelle, et sont en grande partie regroupés par le service CloudWatch. Celui-ci assure la supervision des services, et présente des alertes et des tableaux de bord pour les indicateurs, les logs et les événements générés par d'autres ressources et applications AWS. Les événements et indicateurs classiques consignent la création et la mise hors service des instances, la consommation de CPU, le trafic réseau et l'utilisation du stockage.

Plateforme Google Cloud (GCP)

Cas d'utilisation : Sécurité et conformité, Opérations IT

Exemples : Stackdriver

GCP est une infrastructure de cloud public très répandue qui fournit des services à la demande de puissance de calcul, de stockage, de base de données, de big data et d'applications, en appliquant une tarification à la consommation. GCP peut remplacer les infrastructures traditionnelles de serveurs virtuels d'entreprise, qui exécutent les logiciels sur des VM distinctes, ou héberger des applications natives élaborées à partir d'une palette de services Azure. GCP propose un large éventail de services de gestion, d'automatisation, de sécurité, de réseau et de supervision qui permettent de déployer, redimensionner, mettre hors service, auditer et administrer son environnement GCP, ainsi que ses abonnements et ses applications hébergées.

Cas d'utilisation

Sécurité et conformité : Les données de sécurité des services GCP incluent les événements et les tentatives de connexion et de déconnexion, les appels d'API et les logs des pare-feux du réseau et des applications web.

Opérations IT : Les services GCP fournissent des types de données sur les systèmes et les services qui sont comparables à ceux d'une infrastructure informatique traditionnelle, et sont en grande partie regroupés par Stackdriver. Celui-ci assure la supervision des services, et présente des alertes et des tableaux de bord pour les indicateurs, les logs et les événements générés par d'autres ressources et applications GCP. Les événements et indicateurs classiques consignent la création et la mise hors service des instances, la consommation de CPU, le trafic réseau et l'utilisation du stockage.

Microsoft Azure

Cas d'utilisation : Sécurité et conformité, Opérations IT

Exemples : WADLogs, WADEventLogs, WADPerformanceCounter, WADDiagnostInfrastructure

Azure est une infrastructure de cloud public très répandue qui fournit des services à la demande de puissance de calcul, de stockage, de base de données, de big data et d'applications, en appliquant une tarification à la consommation. Azure peut remplacer les infrastructures traditionnelles de serveurs virtuels d'entreprise, qui exécutent les logiciels sur des VM distinctes, ou héberger des applications natives élaborées à partir d'une palette de services Azure. Azure propose un large éventail de services de gestion, d'automatisation, de sécurité, de réseau et de supervision qui permettent de déployer, redimensionner, mettre hors service, auditer et administrer son environnement Azure, ainsi que ses abonnements et ses applications hébergées.

Cas d'utilisation

Sécurité et conformité : Les équipes de sécurité peuvent utiliser les logs des services Azure pour effectuer des audits de conformité et attester du respect des politiques établies. Les données de logs sont également très précieuses lors des analyses et des investigations à la suite d'un incident. Elles permettent notamment d'identifier les tentatives d'accès non autorisé dans les logs d'accès, de tracer les ressources et les événements de modification des configurations et de localiser des vulnérabilités dans les hôtes et les pare-feux.

Opérations IT : Les services Azure fournissent des métriques et des logs détaillés permettant de superviser l'infrastructure sur toute la pile technologique : VM, conteneurs, stockage et services d'applications. Ces données sont utiles pour préserver la qualité de livraison des applications et les niveaux de services, pour mesurer le comportement des utilisateurs et l'utilisation des ressources, ainsi qu'à des fins de planification des capacités et de gestion des coûts.

DONNÉES DES INFRASTRUCTURES VIRTUELLES

Pivotal Cloud Foundry (PCF)

Cas d'utilisation : Opérations IT et DevOps

Exemples : Loggregator, PCF Healthwatch

Pivotal Cloud Foundry est une plateforme en tant que service (PaaS) reposant sur Cloud Foundry, une plateforme informatique open-source qui permet aux développeurs de déployer, administrer et redimensionner facilement des applications natives du cloud. Les entreprises peuvent ainsi gérer l'intégralité du cycle de vie des applications, de la création du paquet à l'exécution en passant par le déploiement, car Cloud Foundry prend en charge de nombreux frameworks cloud et langages d'applications. Avec PCF, l'installation et l'administration des applications natives du cloud sont simplifiées grâce à des fonctionnalités de gestion et de provisionnement de l'infrastructure, d'application des correctifs des OS, d'orchestration des conteneurs, de sécurité, et plus encore.

Cas d'utilisation

Opérations IT et DevOps : Les équipes des opérations utilisent les métriques PCF (pour la plupart consolidées via Loggregator Firehose) pour obtenir des informations sur l'état de santé du déploiement, les besoins de capacité et la santé des applications avant que les utilisateurs finaux ne soient affectés par une dégradation des performances. Comme PCF permet aux DevOps d'exécuter rapidement leurs applications sur le cloud de leur choix et de les faire évoluer sur demande, les données PCF sont cruciales pour donner aux équipes une visibilité de bout en bout sur l'intégralité du cycle de vie de chaque composant. Lorsque l'on administre des déploiements PCF à grande échelle, comprendre les performances implique de voir les dépendances qui lient les différentes couches de l'architecture d'applications, de conteneurs et du système dans son ensemble.

DONNÉES DES INFRASTRUCTURES VIRTUELLES

Logs de serveurs, données de configuration et indicateurs de performance VMware

Cas d'utilisation : Sécurité et conformité, Opérations IT

Exemples : vCenter, ESXi

VMware vSphere ESXi est la plateforme de virtualisation de serveurs d'entreprise la plus répandue. La plateforme de gestion VMware, qu'il s'agisse d'un produit vSphere ou d'un hyperviseur autonome, produit différents types de données qui peuvent être classées en quatre catégories principales :

- **Logs vCenter :** vCenter est le « centre de contrôle » de l'environnement vSphere. Les logs vCenter indiquent qui se connecte pour effectuer des modifications et qui sont les auteurs des modifications, et consignent les échecs d'authentification.
- **Logs ESXi :** tous les environnements vSphere comprennent au moins un hyperviseur ESXi : c'est le système qui héberge les machines virtuelles. Les logs ESXi contiennent des informations utiles pour dépanner les problèmes de matériel et de configuration.
- **Informations d'inventaire :** l'environnement vCenter trace la configuration de nombreuses entités de configuration : hyperviseurs, machines virtuelles, dépôts de données, clusters et plus encore. Outre la configuration de chaque élément, ses liens éventuels avec les autres entités sont également consignés. Ces informations ne sont pas représentées dans les fichiers de log des serveurs vCenter ni EXSi. Elles peuvent être consultées à l'aide du client vSphere ou des API vSphere, qui permettent d'extraire ces informations. Dans les deux cas, ces informations sont tirées des serveurs vCenter.
- **Informations de performance :** le serveur vCenter suit plusieurs indicateurs de performance pour chaque élément de configuration. La latence des dépôts de données, la consommation des CPU virtuels et physiques et plus de 100 autres indicateurs appartiennent à cette catégorie. Tout comme les informations d'inventaire, ces informations ne figurent pas dans les fichiers de log et doivent être visualisées à l'aide du client vSphere ou extraites à l'aide de l'API vSphere.

Cas d'utilisation

Sécurité et conformité : Le fait que les ressources virtuelles et le matériel physique de base soient dissociés peut rendre difficile la recherche des incidents, l'analyse des capacités, le suivi des changements et le rapport de sécurité. C'est dans les logs vCenter que l'on trouve l'une des applications de sécurité les plus courantes des données VMware. Les logs vCenter contrôlent l'activité des individus qui utilisent l'interface de vSphere pour réaffecter des autorisations utilisateur au sein de l'environnement VMware.

Opérations IT : Les équipes opérationnelles peuvent utiliser les données VMware pour mesurer la santé de l'environnement global des hyperviseurs ainsi que celle des systèmes d'exploitation hébergés sous-jacents. Pour les administrateurs, ces données facilitent la planification des capacités , et le dépannage des problèmes de performance persistants, comme les problèmes de latence dans les dépôts de données.

Ces données enregistrent également l'utilisation des ressources matérielles, une information qui peut permettre d'optimiser les déploiements de VM sur tout un pool de serveurs afin de maximiser la consommation des ressources sans qu'aucun serveur ne soit surchargé.

DONNÉES DES INFRASTRUCTURES VIRTUELLES

Données des infrastructures physiques

Sauvegarde

Cas d'utilisation : Opérations informatiques

En dépit de la réplication des données sur des systèmes, bases de données et dépôts de fichiers miroirs, la sauvegarde des données reste une fonction IT essentielle. Elle assure l'archivage à long terme des informations précieuses, dont beaucoup font l'objet d'exigences légales et réglementaires de conservation. Les sauvegardes servent également à conserver plusieurs versions d'images et de données système, permettant ainsi aux entreprises d'annuler rapidement des modifications, des suppressions accidentelles ou des corruptions de données, ou de rétablir une version fonctionnelle connue d'un système ou d'une base de données. Les logiciels de sauvegarde peuvent utiliser différents supports de stockage selon la fréquence d'accès aux données : disques externes ou bibliothèques de bandes virtuelles pour les données actives et bandes, disques optiques ou service cloud pour le stockage à long terme.

Cas d'utilisation

Opérations IT : Les systèmes de sauvegarde enregistrent en continu l'activité et l'état du système et consignent des informations comme l'historique des tâches, les conditions d'erreur, les cibles des sauvegardes et un manifeste détaillé des fichiers ou volumes copiés. Ces données permettent aux équipes opérationnelles de superviser l'état des systèmes, des logiciels et des tâches de sauvegarde, de déclencher des alertes en cas d'erreur et de faciliter le débogage des échecs de sauvegarde. Les équipes s'en servent aussi pour localiser l'emplacement de données spécifiques lorsqu'une restauration s'avère nécessaire.

Capteurs environnementaux

Cas d'utilisation : Internet des objets, Business Analytics

Exemples : Bosch Sensortec, Mouser Electronics, Raritan, Schneider Electric, TSI, Vaisala

Les capteurs environnementaux délivrent des informations sur la pression atmosphérique, l'humidité, la température ambiante et la qualité de l'air. Ils s'appliquent à toutes sortes de domaines : la lutte contre la pollution, la détection des gaz, la protection des datacenters contre la surchauffe.

Cas d'utilisation

Internet des objets : Les capteurs environnementaux sont une catégorie de compteurs intelligents qui ont été optimisés pour superviser l'environnement. Dans certains cas, comme dans un datacenter, les informations fournies par ces capteurs sont utilisées pour modifier automatiquement les réglages de température et de circulation de l'air.

Analyse commerciale : La collecte des données de capteurs environnementaux peut être utilisée par des applications dans le secteur du détail pour répondre à des questions comme « quel impact une météo médiocre peut-elle avoir sur la fréquentation d'un centre commercial ? »

Systèmes de contrôle industriel (ICS)

Cas d'utilisation : Sécurité et conformité, Internet des objets, Business Analytics

Exemples : ABB, Emerson Electric, GE, Hitachi, Honeywell, Rockwell Automation, Siemens, Toshiba

Dans le contexte d'un environnement de fabrication, les systèmes de contrôle industriel emploient des contrôleurs logiques programmables pour acquérir des données et exécuter des fonctions de supervision. Une grande partie de l'automatisation des processus d'un site de fabrication repose sur les systèmes de contrôle industriel.

Cas d'utilisation

Sécurité et conformité : Les systèmes de contrôle industriel (ICS) jouent un rôle crucial dans la livraison des services aux industries et aux municipalités du monde entier. Ces systèmes reposent sur une infrastructure IT traditionnelle, et même si elle est généralement séparée de l'IT de l'entreprise, les entreprises en pleine transformation numérique sont tentées d'apporter de la connectivité à ces systèmes, ce qui accroît leur exposition aux attaques. Ces systèmes sont généralement non supervisés sur le plan de la sécurité. Indépendamment du risque d'attaque ou d'infection des ICS, les données qu'ils produisent peuvent apporter de la visibilité et servir à analyser et identifier les activités malveillantes et les menaces potentielles. Cette visibilité permet aux entreprises de mesurer l'impact et le risque, et de les rapprocher des processus métier.

Internet des objets : Les données machine des ICS peuvent être utilisées pour obtenir une visibilité en temps réel sur la disponibilité des actifs stratégiques. Cela permet aux entreprises de détecter un problème, d'effectuer une analyse des causes profondes et d'appliquer des mesures préventives pour empêcher certains événements de se produire à nouveau à l'avenir. Les entreprises exploitent également les données machine des ICS pour sécuriser ces actifs stratégiques.

Business Analytics : Les entreprises peuvent appliquer des algorithmes de machine learning aux données machine créées par les systèmes de contrôle industriel pour gagner en productivité et en disponibilité. Les données d'ICS apportent également une visibilité sur les processus de fabrication complexes, ce qui facilite l'identification des goulets d'étranglement et l'élimination des pertes d'efficacité.

Mainframe

Cas d'utilisation : Opérations informatiques

Les mainframes sont les premiers ordinateurs d'entreprise : des systèmes centralisés volumineux abritant plusieurs processus, de la mémoire RAM et des contrôles I/O. En dépit de leurs 60 ans d'histoire, les mainframes sont encore largement utilisés pour les applications critiques, en particulier pour le traitement des transactions. Bien qu'ils utilisent généralement un OS propriétaire, les mainframes peuvent aussi être virtualisés pour exécuter Unix et Linux ou, grâce à des cartes de processeur complémentaires, Windows Server. Les mainframes sont appréciés pour leur fiabilité et leur sécurité à toute épreuve : ils comprennent en effet du matériel hautement redondant et des logiciels résilients soumis aux tests les plus rigoureux. C'est pour cela qu'ils séduisent les entreprises qui souhaitent regrouper leurs charges sur un petit nombre de systèmes et ont besoin de leur fiabilité et leur polyvalence.

Cas d'utilisation

Opérations IT : Comme les autres serveurs, les mainframes mesurent et consignent de nombreux paramètres système indiquant leur statut actuel, leur configuration et leur état général. Comme la plupart des sous-systèmes des mainframes sont redondants, les logs système signalent également les défaillances matérielles et les comportements anormaux qui n'ont causé aucune perturbation mais sont les signes avant-coureurs d'une interruption de service. Comme ils sont utilisés pour les applications critiques, les mainframes enregistrent souvent des données de performances des applications comme l'utilisation de la mémoire, le débit d'I/O et de transaction, la consommation du CPU et l'activité réseau.

Dispositifs médicaux

Cas d'utilisation : Internet des objets, Business Analytics

Exemples : Abbott Laboratories, Apple, Baxter, Boston Scientific, GE, Siemens, St. Jude Medical

Des unités de soins intensifs aux dispositifs corporels, tous les appareils médicaux génèrent de multiples types de données machine. En effet, tous les aspects des soins aux patients, dans l'hôpital comme à l'extérieur, peuvent être mesurés. Si l'objectif principal est de sauver des vies, l'objectif secondaire crucial est de faire baisser le coût de la santé en réduisant à la fois le nombre de visites potentielles à l'hôpital ainsi que la durée du séjour.

Cas d'utilisation

Internet des objets : La plupart des dispositifs d'un hôpital sont connectés à des applications de supervision locales. Mais il est possible de superviser les soins des patients à distance à l'aide de capteurs qui communiquent avec un dispositif corporel ou autre système de monitoring à domicile.

Business Analytics : Les données machine permettent également aux professionnels de santé d'analyser des données anonymes et les données des patients sur de vastes régions dispersées, par exemple pour déterminer dans quelle mesure certaines maladies affectent un groupe de personnes plus qu'un autre.

DONNÉES DES INFRASTRUCTURES PHYSIQUES

Protocoles de données métriques

Cas d'utilisation : Opérations IT, Livraison des applications, Internet des objets

Exemples : collectd, statsd

Les métriques sont des indicateurs générés par un processus exécuté sur un système qui fournit régulièrement une information statistique particulière comme la consommation de CPU. Les sources de données métriques génèrent, à intervalles réguliers, des indicateurs qui incluent généralement :

- Horodatage
- Nom de métrique
- Mesure (la donnée)
- Dimensions (qui décrivent le plus souvent l'hôte, le type d'instance ou d'autres attributs permettant de filtrer ou de trier les métriques)

Pour l'essentiel, les métriques sont générées par un démon (ou processus) qui s'exécute sur un serveur (OS), un conteneur ou une application. Chaque mesure est délivrée par un protocole réseau comme UDP ou HTTP à un serveur qui indexe et analyse cette information.

Les métriques sont d'une grande importance pour la supervision. Par exemple, de même qu'un moniteur cardiaque contrôle régulièrement le pouls d'un patient, les métriques apportent un éclairage sur les tendances et les problèmes qui affectent la performance et la disponibilité d'une infrastructure ou d'une application. En revanche, un moniteur cardiaque ne vous dira pas pourquoi le pouls d'un patient devient brutalement anormal : il faut d'autres outils pour identifier rapidement la cause du mal et stabiliser la victime. C'est la même chose avec les données machine. Une fois combinées avec d'autres sources de données – des logs, généralement – elles fournissent aussi bien des renseignements sur les problèmes eux-mêmes que sur leur cause.

Exemples de protocoles de données métriques

Collectd : Collectd est un protocole qui consiste à exécuter un agent sur un serveur configuré pour mesurer des attributs spécifiques, puis transmettre ces informations à une destination définie. Collectd est un moteur de mesure extensible qui permet de recueillir un large éventail de données. Aujourd'hui, collectd est essentiellement employé pour assurer la supervision du cœur des infrastructures, notamment pour obtenir des renseignements sur la charge, l'utilisation de la mémoire, le taux d'I/O et le stockage des serveurs et autres composants d'infrastructure. Collectd fait partie de la communauté open-source ; pour en savoir plus sur ce protocole, rendez-vous sur **http://collectd.org**.

Statsd : c'est un démon réseau qui s'exécute sur node.js. Il jouit d'une grande popularité auprès des administrateurs Windows et des experts en performance des applications, entre autres. Statsd fournit des capacités permettant de délivrer des métriques de façon groupée. De plus, bien qu'il utilise la méthode UDP, moins fiable, beaucoup d'administrateurs apprécient la simplicité de son déploiement. Similaire à collectd, statsd est dédié à la collecte de métriques concernant essentiellement l'utilisation et les performances des applications et de leurs composants. Le protocole les envoie ensuite, via le réseau, à un outil capable de recueillir et d'analyser ces informations.

Cas d'utilisation

Opérations IT et livraison des applications : Les protocoles de données métriques fournissent des données d'utilisation, de performance et de disponibilité sur les systèmes d'exploitation, les dispositifs de stockage, les applications et autres composants de l'infrastructure IT. Les métriques sont particulièrement utiles pour les activités de supervision des équipes des Opérations IT et de Gestion des applications, où les tendances permettent souvent de localiser les problèmes. Une fois les anomalies détectées à l'aide de tendances et de seuils, d'autres sources de données sont souvent corrélées pour déterminer la cause profonde du problème.

Internet des objets : En devenant de plus en plus intelligents, les dispositifs connectés produisent toujours plus de métriques à des fins de supervision à distance. Les protocoles de données métriques offrent un moyen efficace de recueillir l'état et la performance de ces dispositifs.

DONNÉES DES INFRASTRUCTURES PHYSIQUES

Logs de correctifs

Cas d'utilisation : Sécurité et conformité, Opérations IT

Il est indispensable de maintenir les systèmes d'exploitation et les applications à jour en appliquant les correctifs de débogage et de sécurité les plus récents, afin d'éviter les interruptions imprévues, les défaillances aléatoires et les failles de sécurité. Bien que les applications et les systèmes d'exploitation commerciaux intègrent souvent un système de mise à jour, certaines entreprises utilisent des logiciels indépendants de gestion des correctifs pour harmoniser l'application des correctifs sur leur parc logiciel et développer des mises à jour pour les applications internes et personnalisées.

Un logiciel de gestion des correctifs conserve un inventaire des correctifs à l'aide d'une base de données des mises à jour disponibles, qu'il rapproche de la liste des logiciels installés d'une organisation. Il permet également de planifier l'installation des correctifs, de réaliser des tests post-installation et de valider et documenter les configurations système requises ainsi que les procédures de correction.

Cas d'utilisation

Sécurité et conformité : Les équipes de sécurité peuvent consulter les logs de correctifs pour superviser les mises à jour et identifier les actifs potentiellement à risque, parce qu'un correctif n'a pas été bien appliqué ou que sa version est obsolète.

Opérations IT : Les équipes opérationnelles utilisent les logs de correctifs pour vérifier que les mises à jour planifiées ont été appliquées correctement et dans les temps, pour identifier les systèmes et les applications non corrigés et pour être alertées en cas d'erreur dans le processus d'installation. Corréler les erreurs et les logs de correctifs peut indiquer si une erreur est due à un correctif.

Lecteurs de cartes physiques

Cas d'utilisation : Sécurité et conformité

La plupart des entreprises emploient des systèmes automatisés pour sécuriser l'accès physique à leurs locaux. Historiquement, il s'agit de simples bandes magnétiques fixées aux badges des employés ; toutefois, les sites qui ont des exigences de sécurité plus strictes emploient parfois des lecteurs biométriques ou des clés numériques. Quelle que soit la technologie utilisée, le système compare l'identité d'une personne à une base de données et active les portes quand elle est autorisée à entrer dans une zone particulière. Comme ce sont des systèmes numériques, les lecteurs de badges enregistrent des informations comme l'identifiant de l'utilisateur, la date et l'heure d'entrée, et éventuellement une photo de chaque tentative d'accès.

Cas d'utilisation

Sécurité et conformité : Pour les équipes de sécurité IT, les données provenant des lecteurs de cartes fournissent, pour les sites physiques, les mêmes types d'informations d'accès que les logs des pare-feux réseau. Ces données peuvent être exploitées pour détecter les tentatives d'infraction et être corrélées aux logs des systèmes et du réseau pour identifier les menaces internes potentielles et fournir une vision globale de la situation. Elles permettent également de détecter les accès qui ont lieu à une heure ou dans un lieu inhabituel, ou dont la durée est anormale.

DONNÉES DES INFRASTRUCTURES PHYSIQUES

Systèmes de point de vente (POS)

Cas d'utilisation : Sécurité et conformité, Internet des objets, Business Analytics

Exemples : IBM, LightSpeed, NCR, Revel Systems, Square, Toshiba, Vend

Les systèmes de point de vente (POS) sont essentiellement associés aux transactions générées dans un point de vente au détail. Toutefois, avec l'apparition des solutions POS mobiles, on les voit apparaître dans des contextes temporaires tels que des kermesses ou des fêtes d'établissement scolaire.

Un système POS classique comprend une caisse enregistreuse installée sur un PC ou un système intégré, un écran, une imprimante à tickets de caisse, un affichage, un lecteur de codes-barres et un lecteur de carte de crédit/débit. Les données machine générées par les systèmes POS fournissent aux entreprises des renseignements en temps réel sur les articles vendus, la trésorerie générée par transaction et les méthodes de paiement utilisées.

Cas d'utilisation

Sécurité et conformité : Les systèmes POS sont naturellement employés pour des transactions financières et sont fréquemment la cible d'attaques car ils contiennent des données de compte bancaire, de paiement et autres informations financières. Comme les informations sur les transactions de POS sont extrêmement précieuses et recherchées par les malfaiteurs, et parce que les POS peuvent être utilisés comme point d'entrée dans le réseau, ils doivent impérativement être protégés. D'autre part, les systèmes POS sont généralement non supervisés et le système d'exploitation sous-jacent, ainsi que sa supervision et ses mises à jour, échappent le plus souvent au champ de vision de l'IT, ce qui ajoute encore à la complexité de la sécurité. La visibilité et l'analyse des systèmes POS et de leurs données peuvent apporter des renseignements vitaux pour la protection des informations financières, la détection de la fraude et la sécurisation des vulnérabilités.

Internet des objets : Traditionnellement, les systèmes POS n'étaient ni connectés ni même gérés sur un réseau privé dédié. Mais avec l'avènement de l'IoT, ils sont désormais connectés directement à des plateformes cloud qui facilitent considérablement leur administration à distance depuis un emplacement central. Il n'est plus nécessaire d'envoyer du personnel IT sur place pour mettre les systèmes à jour un par un. C'est un atout de poids car une panne de POS peut créer de longues files d'attente qui sont désagréables pour les clients et peuvent entraîner des pertes de revenus. Une expérience négative peut facilement inciter les clients à faire leurs achats ailleurs dans un secteur où la concurrence est intense.

Business Analytics : Les systèmes POS contiennent des informations sur les articles vendus, les méthodes de paiement et le rythme de vente. Ces données peuvent permettre de superviser les revenus en temps réel, ce qui permet d'informer les stratégies d'interaction avec les clients, de suivre les ventes des articles par rapport à leur emplacement dans la boutique et de détecter les transactions potentiellement frauduleuses en temps réel. Ce type d'analyse de big data peut avoir un impact puissant sur les opportunités de vente croisée et additionnelle. Les données de POS apportent en outre de la visibilité sur l'expérience du client, indiquant par exemple quels coupons sont les plus utilisés ou quels articles sont fréquemment vendus ensemble. Enrichies de données de géolocalisation, elles peuvent aussi délivrer des renseignements précieux en termes d'analyse par site.

DONNÉES DES INFRASTRUCTURES PHYSIQUES

RFID/NFC/BLE

Cas d'utilisation : Internet des objets, Business Analytics

Exemples : Alien Technology, BluVision, CheckPoint Systems, Gimbal, MonsoonRF, Radius Networks, STMicroelectronics, TAGSYS RFID, ThingMagic

Les deux grandes méthodes sans fil employées actuellement par les entreprises pour suivre les objets et interagir avec les clients dans les boutiques impliquent deux types différents de technologies de communication. La plus connue est la RFID, ou optimisation des radio-fréquences, qui implique l'utilisation d'étiquettes capables de stocker des données, comme des informations sur le produit ou les marchandises à charger dans un conteneur.

Parallèlement à cela, les entreprises adoptent des solutions de connectivité sans fil Bluetooth à faible énergie (BLE), capables de transmettre des signaux à d'autres appareils. Le BLE est essentiellement utilisé dans des balises, par exemple pour informer les consommateurs d'une promotion via leur smartphone ou informer les fans des différentes activités en cours lors d'un événement sportif.

Cas d'utilisation

Internet des objets : La RFID est certainement l'une des toutes premières applications de l'IoT. Déployées à la place des lecteurs de codes-barres traditionnels, les étiquettes RFID sont utilisées partout, de l'expédition au suivi des animaux d'élevage. Les déploiements IoT permettent de capter les données RFID de façon à simplifier le suivi des événements affectant tout ce qui peut porter une étiquette. Les renseignements provenant de la RFID contribuent à l'amélioration globale de la chaîne logistique, du traitement des commandes et de la gestion de l'inventaire.

Le BLE, quant à lui, permet d'interagir plus directement avec les clients lorsqu'ils se déplacent à un endroit précis : il produit donc des données utilisables pour optimiser l'expérience client.

Business Analytics : Qu'il s'agisse de suivre les stocks à l'aide d'étiquettes RFID ou d'interagir avec les clients et les employés au fil de leurs déplacements, de nouvelles catégories d'applications d'analyse exploitent les données générées par ces dispositifs pour produire des renseignements commerciaux exploitables en quasi-temps réel. Les détaillants peuvent ensuite utiliser ces données dans différents scénarios d'utilisation, notamment pour veiller à ce que les stocks soient situés au plus près des sites où les clients sont les plus susceptibles de les vouloir.

DONNÉES DES INFRASTRUCTURES PHYSIQUES

Données de capteurs

Cas d'utilisation : Sécurité et conformité, Opérations IT, Internet des objets, Business Analytics

Exemples : Valeurs binaires et numériques : état de commutateur, température, pression, fréquence, débit, provenant des courtiers MQTT, AMQP et CoAP, ou d'un collecteur d'événements HTTP

L'équipement industriel, les capteurs et autres dispositifs sont souvent équipés de processeurs embarqués et d'outils réseau qui leur permettent d'enregistrer et de transmettre un large éventail d'informations sur les conditions de fonctionnement. Quel que soit l'appareil concerné, les données offrent un haut degré de détails sur les paramètres de performance et les anomalies qui peuvent indiquer des problèmes plus vastes, comme la défaillance imminente d'un équipement ou un problème touchant un autre système. L'agrégation et la corrélation de données provenant de multiples dispositifs et sous-systèmes produit une image complète des performances de l'équipement, du système, de l'usine ou des installations.

Cas d'utilisation

Sécurité et conformité : Les données des capteurs peuvent protéger les actifs stratégiques et les systèmes industriels contre les menaces de cybersécurité, en offrant une visibilité sur les performances du système ou les points de bascule susceptibles de mettre en danger les machines ou les personnes. Elles peuvent aussi servir à produire des rapports de conformité obligatoires.

Opérations IT : Les facteurs environnementaux font partie des paramètres supervisés en priorité par les équipes d'exploitation : température, humidité, débit d'air et, dans un datacenter, régulation de la tension. Chaque serveur et équipement réseau envoie ces mesures qui, une fois corrélées, peuvent mettre en lumière des problèmes touchant le site ou une interruption de service imminente.

Cas d'utilisation supplémentaires

Maintenance préventive et Gestion du cycle de vie des actifs : Les données de capteurs fournissent des renseignements sur le déploiement et l'utilisation des actifs, ainsi que sur la consommation des ressources. Les données opérationnelles peuvent aussi être exploitées au bénéfice d'une approche proactive et à long terme de la gestion, la maintenance et la performance des actifs.

Supervision et diagnostics : Les capteurs de supervision contribuent à vérifier que l'équipement de terrain fonctionne comme prévu, par exemple en supervisant et en suivant les interruptions imprévues des dispositifs et des systèmes. Ces données servent aussi à comprendre la cause d'une panne d'un appareil pour améliorer son efficacité et sa disponibilité, et à identifier les anomalies et les problèmes au moment de la production ou du déploiement des dispositifs.

DONNÉES DES INFRASTRUCTURES PHYSIQUES

Logs de serveurs

Cas d'utilisation : Sécurité et conformité, Opérations IT, Livraison des applications

Les systèmes d'exploitation des serveurs enregistrent en permanence une variété de données de fonctionnement, de sécurité, d'erreurs et de débogage : librairies système chargées au démarrage, processus d'application ouverts, connexions réseau, systèmes de fichiers montés, utilisation de la mémoire système, etc. Le niveau de détail est configurable par l'administrateur système, mais il y a suffisamment d'options pour obtenir une image complète de l'activité du système tout au long de son service. Selon le sous-système, les logs des serveurs seront utiles aux équipes chargées du système, du réseau, du stockage ou de la sécurité.

Cas d'utilisation

Sécurité et conformité : Les logs de serveurs regroupent les données des sous-systèmes de sécurité : événements des pare-feux locaux, tentatives de connexion et erreurs d'accès aux fichiers. Les équipes de sécurité savent exploiter ces données pour identifier les tentatives d'infraction, tracer les infiltrations réussies et corriger les vulnérabilités. Superviser les logs de serveurs qui consignent les accès aux fichiers, les authentifications et l'utilisation des applications peut contribuer à sécuriser les composants d'infrastructure.

Opérations IT et livraison des applications : Les logs de serveurs fournissent un enregistrement détaillé de l'état général du système et des informations sur l'heure exacte des erreurs et des situations anormales, qui sont inestimables lorsque l'on recherche la cause profonde des problèmes d'un système.

Compteurs intelligents

Cas d'utilisation : Internet des objets, Business Analytics

Exemples : ABB, GE, Google, eMeter, IBM, Itron, Schneider Electric, Siemens

Les compteurs intelligents enregistrent la consommation d'électricité, d'eau ou de gaz naturel afin de traiter et partager les informations en continu. Généralement, les compteurs intelligents assurent une communication bidirectionnelle en temps réel, ce qui permet d'ajuster une forme de jauge.

Cas d'utilisation

Internet des objets : Les compteurs intelligents sont déployés sur les systèmes stratégiques des grandes sociétés de services publics d'électricité, de gaz et d'eau, notamment. Ces systèmes sont la sève des infrastructures et la moindre interruption de service peut avoir des résultats catastrophiques. La supervision en temps réel des compteurs intelligents permet aux entreprises de mieux analyser les interruptions à distance. Il est tout aussi important de protéger les appareils des tentatives de manipulation pouvant ouvrir la voie à des attaques malveillantes ou des failles.

Les sociétés d'électricité et d'eau tirent parti des capteurs intelligents pour suivre toutes sortes de paramètres, des réserves de pétrole à la qualité de l'approvisionnement en eau.

Business Analytics : Un large éventail de secteurs applique des analyses aux données recueillies par les compteurs intelligents pour optimiser leurs services. Une société de pétrole ou de gaz, par exemple, n'a plus besoin d'envoyer physiquement un technicien sur place pour lire un compteur. Le fournisseur sait déjà quelle quantité de carburant a été consommée et combien il en reste.

À l'avenir, les compteurs intelligents seront utilisés partout, des systèmes de contrôle de la circulation aux systèmes de défense conçus pour protéger les infrastructures critiques. L'agrégation des données de ces compteurs intelligents offre à ces compagnies des informations stratégiques sur la demande. Ces entreprises fortement encadrées doivent respecter des SLA définis en cas de pic de demande et les données machine des compteurs intelligents apportent de la visibilité sur la réactivité des systèmes.

Stockage

Cas d'utilisation : Opérations IT

Exemples : EMC, Netapp, IBM, Amazon EBS

Dans les datacenters, le stockage est généralement mis en œuvre de deux manières : intégré aux serveurs et partagé à l'aide de divers protocoles de stockage en réseau, ou via un dispositif de stockage dédié qui regroupe des capacités utilisées par différentes applications. Celles-ci accèdent aux données par le biais d'un protocole de partage de fichiers de type SAN (réseau de stockage) ou LAN Ethernet. L'activité d'un stockage interne en serveur est généralement enregistrée dans des logs système, mais les dispositifs de stockage sont dotés de contrôleurs internes ou de processeurs qui exécutent un OS optimisé pour le stockage et consignent pléthore de données de fonctionnement, d'erreurs et d'utilisation. Comme de nombreuses entreprises possèdent plusieurs dispositifs de ce type, les logs sont souvent regroupés par un système de gestion du stockage qui produit des rapports agrégés sur l'activité et les capacités.

Cas d'utilisation

Opérations IT : Les logs du stockage partagé enregistrent l'état général du système (matériel et logiciel), les conditions d'erreur (défaillance d'un contrôleur, d'une interface réseau ou d'un disque) et l'utilisation (capacité utilisée par volume et fichier, et accès aux volumes). Collectivement, les informations peuvent avertir les équipes opérationnelles en cas de problème, de capacités insuffisantes ou de goulets d'étranglement au niveau des performances.

Téléphonie

Cas d'utilisation : Opérations IT

Exemples : Cisco Unified Communications Manager, Shoretel, Twilio

Les communications d'entreprise en temps réel ne sont plus limitées aux appels vocaux fournis par les services de téléphonie traditionnels (POTS). Aujourd'hui, la voix, la vidéo, la messagerie texte et les conférences web sont des applications IP délivrées sur les réseaux de l'entreprise. Contrairement aux applications client-serveur ou web traditionnelles, la téléphonie et les autres outils de communications ont des besoins stricts en termes de qualité du réseau, de latence et de pertes de paquets, si bien que la qualité et la fiabilité du service est bien plus sensible à l'état du réseau et à la réactivité des serveurs. La téléphonie traditionnelle a conditionné les utilisateurs à entendre une tonalité dès qu'ils décrochent le combiné et les a rendus intolérants aux moindres bruits, échos et autres problèmes typiques de la téléphonie sur IP. C'est pourquoi les systèmes et l'infrastructure sous-jacente doivent être soigneusement supervisés et gérés pour garantir la qualité et la fiabilité attendues.

Cas d'utilisation

Opérations IT : Comme pour la VoIP, les logs de téléphonie fournissent une vue d'ensemble de l'état du système ainsi que des données de dépannage et d'utilisation similaires à celles d'autres applications réseau. Ils consignent notamment la source, la destination, l'heure et la durée des appels voix/vidéo, des conférences web et des messages textes, des indicateurs de la qualité des appels (pertes de paquets, latence, fidélité audio/compression, etc.), les conditions d'erreur et la participation des utilisateurs aux conférences web. En rapprochant les enregistrements des adresses sources et cibles de téléphone d'une base de données des employés comme AD ou LDAP et d'une base de données DHCP, l'entreprise peut associer les registres d'appels à des ID utilisateurs et les adresses IP à des emplacements physiques (des informations utiles à des fins de dépannage et de facturation). Les logs peuvent également mettre en lumière des segments du réseau victimes de congestion ou d'autres problèmes de performance, indicateurs d'un problème matériel ou de la nécessité d'une mise à niveau.

DONNÉES DES INFRASTRUCTURES PHYSIQUES

Transport

Cas d'utilisation : Internet des objets, Business Analytics

Exemples : Boeing, BMW, Ford, GE, General Motors, Daimler-Benz, John Deere, Volkswagen

Des véhicules de toutes les tailles et tous les types produisent des quantités massives de données chaque jour, données qui permettent d'obtenir une visibilité en temps réel sur l'état de santé et les performances d'un actif, mais aussi d'informer la maintenance prédictive. Armé de ces données, un constructeur d'avions ou de voitures peut appliquer un plan de maintenance axé sur les données plutôt que dans les règles préétablies.

Ces informations servent ensuite à améliorer la disponibilité et la fiabilité, et ainsi prolonger le cycle de vie d'un véhicule qui n'a pas beaucoup été utilisé ou, à l'inverse, remplacer des composants usés prématurément.

Cas d'utilisation

Internet des objets : Les constructeurs de véhicules installent des capteurs sur tous les composants mécaniques et électroniques qu'ils emploient. Ils obtiennent ainsi une vue unifiée des différents composants, ce qui leur permet d'identifier et de diagnostiquer les problèmes opérationnels, mais aussi de superviser, suivre et éviter les interruptions non planifiées. Tout cela contribue à garantir le bon fonctionnement de l'équipement. Ces données permettent aussi de détecter les anomalies et les écarts par rapport au comportement normal afin de les corriger, ce qui améliore encore la disponibilité, la fiabilité et la durée de vie des équipements.

Business Analytics : Parce que les constructeurs ont accès aux données machine, ils exploitent l'analyse de façon inédite et transforment fondamentalement leurs modèles commerciaux. Au lieu de vendre un véhicule, ils préfèrent de plus en plus les louer sur la base de l'utilisation réelle. Plus un véhicule peut être utilisé longtemps entre deux réparations, plus le service de leasing devient rentable. Pour offrir ce service de façon économique, la clé réside dans l'analyse avancée, qui est appliquée à toutes les données agrégées qui sont collectées.

Dispositifs corporels

Cas d'utilisation : Internet des objets, Business Analytics

Exemples : ARM, Intel, Lenovo, Microsoft, Samsung

Des montres intelligentes servant de coach sportif aux dispositifs médicaux qui permettent aux médecins de superviser à distance des statistiques vitales, les dispositifs corporels ont prouvé qu'ils étaient durablement installés. Ils représentent même l'un des aspects les plus emblématiques de l'Internet des objets.

Cas d'utilisation

Internet des objets : La dernière génération de montres intelligentes ne se contente plus de se synchroniser avec nos téléphones : elle utilise les systèmes de géolocalisation et les API pour offrir à ses utilisateurs une expérience d'application optimale intégrant à la fois leur localisation et l'heure de la journée.

De nouvelles catégories entières de dispositifs corporels feront leur apparition à l'avenir, et ils exploiteront aussi bien les applications de réalité virtuelle via un casque que des capteurs intégrés aux nouvelles collections de vêtements.

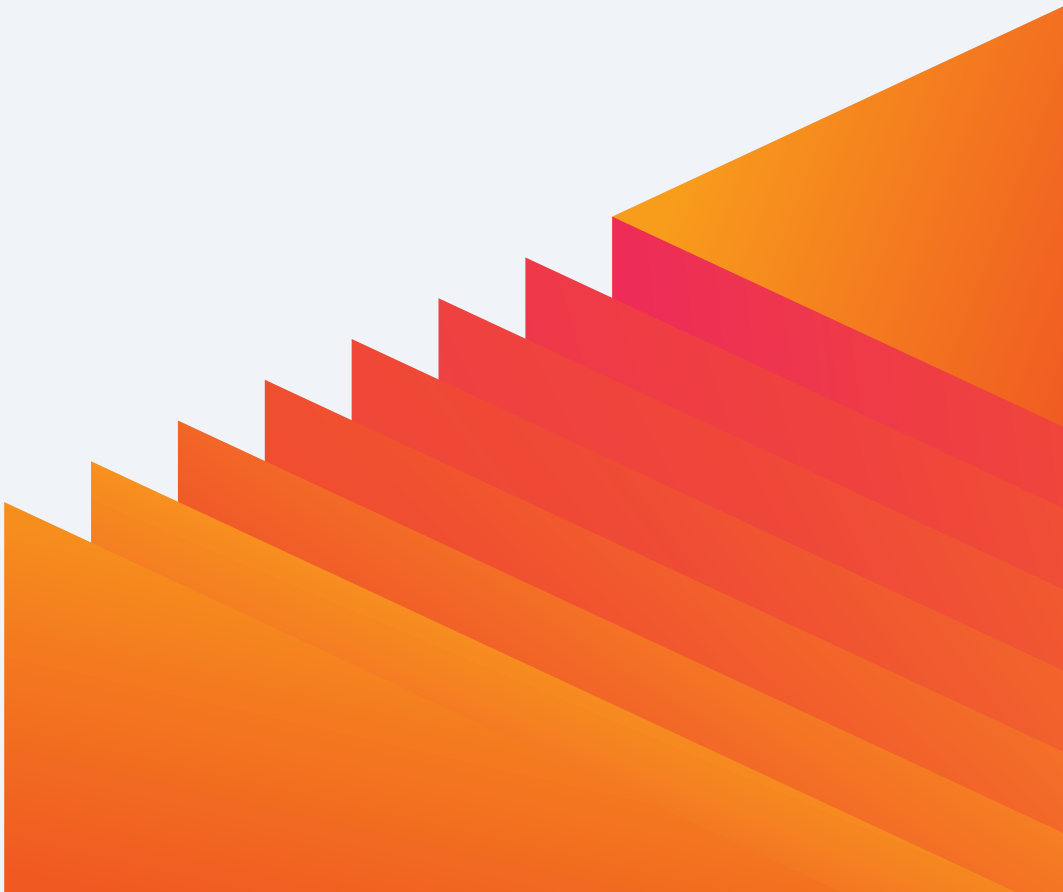
Business Analytics : Le public accepte de plus en plus facilement de partager ses données via des dispositifs corporels et beaucoup d'utilisateurs profitent directement de la puissance de l'analyse. Les développeurs d'applications optimisées pour les dispositifs corporels délivrent des recommandations touchant aussi bien l'amélioration de l'espérance de vie que l'emplacement d'un restaurant. L'analyse des données des dispositifs corporels peut améliorer l'expérience utilisateur et favoriser l'innovation. Un responsable produit peut en effet examiner la façon dont les consommateurs interagissent avec son appareil pour élaborer de meilleures fonctionnalités.

DONNÉES DES INFRASTRUCTURES PHYSIQUES

À propos de **Splunk.**

Splunk transforme les données en actions grâce à la plateforme Data-to-Everything™. La technologie Splunk est conçue pour investiguer, superviser, analyser et exploiter les données à toutes les échelles. Rejoignez des milliers d'utilisateurs passionnés en essayant Splunk gratuitement.

Essai gratuit



splunk>

20-13476-SPLK-Essential-Guide-to-Data-Infrastructure-Data-104_FR