Engineer

Analyse de données

repérer avec méthode les motifs et tendances des données

Ce que c’est

Transformer des faits enregistrés en réponses permettant d’agir est le rôle de l’analyse de données, souvent appelée simplement analyse. Dans le cadre consacré aux mégadonnées publié par le NIST, institut américain de normalisation, elle consiste à traiter méthodiquement les données, à les remodeler au besoin pour dégager des régularités, des liens entre points de données et des tendances temporelles, et tenter de prévoir la suite. Le NIST inscrit cette étape dans un *cycle de vie de l’analyse de données* qui transforme des données brutes en connaissances utilisables par une organisation : collecte, préparation, analyse, visualisation et mise à disposition.

C’est une forme de traitement des données, distincte du transactionnel par ses lectures : une application écrit une commande à la fois, tandis qu’une requête analytique lit des millions de lignes pour en renvoyer quelques-unes. Elle diffère aussi de l’ingénierie des données, qui construit les pipelines de traitement et le stockage dont elle se nourrit.

Le mot couvre des travaux très différents :

- **Rapports** : les mêmes chiffres livrés selon un calendrier, sous forme de PDF ou de courriel.
- **Tableaux de bord** : ces chiffres maintenus à jour dans un outil d’informatique décisionnelle comme Power BI.
- **Analyse ponctuelle** : une requête SQL répondant à une nouvelle question, souvent sur un entrepôt comme BigQuery.
- **Analyse géospatiale** : le même travail sur des données cartographiques dans un SIG, pour demander où autant que combien.
- **Applications analytiques** : un modèle ou une méthode statistique derrière un formulaire, pour des personnes qui n’écrivent pas de code.

- **Analyse Web et produit** : compter les visiteurs d'un site ou les utilisateurs d'un produit et leurs actions.

Cette dernière forme a un coût en confidentialité. C'est elle que visent les pages de confidentialité, cookies, crédits et sécurité de ce site lorsqu'elles disent « aucune analyse d'audience » : aucune balise d'analyse, aucun pixel de suivi, aucun cookie ni compte chez une société d'analyse d'audience. Le site se mesure néanmoins. Son serveur Web conserve un journal d'accès qu'un programme transforme en décomptes quotidiens et en classement mensuel des pages lues, sites référents et robots, conservés sous forme de nombres et jamais réunis en parcours d'un visiteur.

Utilisez le concept pour prendre une décision

Utilisez « Qu'est-ce que l'analyse de données ? » pour éclairer une décision réelle. Nous pouvons vous aider à relier ce concept au système que vous construisez ou achetez. Des tableaux de bord et des rapports présentant les quelques chiffres qui déterminent une décision, actualisés automatiquement.

Découvrir nos prestations Contactez-nous →

Vérifiez les hypothèses derrière le terme

Appuyez-vous sur l'explication et les réalisations liées pour formuler les questions de votre projet. Nous pouvons examiner les hypothèses avec vous avant le choix d'un outil. Une communication de marque éclairée par le public — une communication de marque éclairée par les réactions du public

Une communication de marque éclairée par le public Contactez-nous →

Pourquoi c'est important

Deux choix déterminent l'essentiel du coût de l'analyse : l'endroit où les requêtes s'exécutent et la collecte des données concernant les personnes.

Analyser sur la base transactionnelle est le point de départ le moins cher : les données sont actuelles et rien n'est copié, mais chaque rapport dispute mémoire et disque à l'application. Une réplique de lecture déplace cette charge hors du primaire au prix d'un second serveur. Un entrepôt comme BigQuery sépare stockage et calcul et, en tarification à la demande, facture chaque requête selon les octets lus, ce qui récompense les requêtes ciblées. Il nécessite un pipeline de traitement pour y copier les données.

La dépendance au fournisseur se cache dans les définitions. Un outil décisionnel conserve ses mesures dans son propre langage de formules et, un an plus tard, le sens de « client actif » réside dans un fichier que lui seul sait ouvrir. Des définitions conservées en SQL dans le dépôt survivent au changement d'outil.

L'analyse Web et produit ajoute une question juridique. L'outil hébergé habituel utilise une balise : un script dans chaque page dépose un cookie ou identifiant similaire dans le navigateur et signale chaque page vue au fournisseur. L'article 5(3) de la directive européenne ePrivacy conditionne au consentement le stockage ou la lecture d'informations sur l'appareil d'un visiteur, avec deux exceptions : une finalité exclusivement limitée à la transmission d'une communication sur un réseau, ou un stockage ou un accès strictement nécessaire pour fournir un service explicitement demandé par le visiteur. Dès que des événements peuvent être reliés à une personne, même par un identifiant de cookie ou une adresse IP, le RGPD s'applique aussi : base juridique, finalité annoncée, durée de conservation. Et le fournisseur est une autre entreprise qui possède désormais une trace des actions des visiteurs.

Le choix réel oppose donc un outil hébergé rapide à ajouter avec ses tableaux de bord, une collecte interne à construire et exploiter, et les décomptes issus des journaux que votre serveur écrit déjà, moins riches en réponses et beaucoup moins chargés en données.

À quoi cela ressemble en pratique

Pour une plateforme de réseau professionnel, l'entreprise a construit l'analyse produit dans PostgreSQL plutôt qu'une balise tierce : un schéma de 7 tables et 47 fonctions, des tables d'événements partitionnées selon 22 définitions de partitions, avec suppression automatique des partitions selon un calendrier et une identité visiteur pseudonymisée au moyen d'un sel renouvelé. Le client n'envoie rien avant le consentement, et les entonnoirs de conversion sont calculés dans la base près des données, ce que le NIST appelle la localité des données. La page indique le prix : le renouvellement du sel oblige à renoncer à l'analyse de cohortes à long terme, et aucun tableau de bord gratuit n'est fourni.

Côté rapports, un système de rapports Python récupérait les données, effectuait l'analyse et livrait le même rapport PDF selon un calendrier prévisible aux personnes prenant les décisions commerciales. La page attribue à ces rapports une augmentation de 400 % du chiffre d'affaires logiciel trimestriel.

Ce site constitue le troisième exemple. Sa page de confidentialité décrit la mesure : un journal remplaçant par des zéros la dernière partie de chaque adresse IP avant écriture, un site référent conservé uniquement comme nom et des lignes gardées environ trente jours. Un programme en produit un décompte quotidien et un classement mensuel, conservés indéfiniment comme nombres

sans rien permettant de reconstituer une visite. Sans balise, cookie ni tiers, la page ne montre aucun bandeau de consentement, en donnant pour raison « rien à quoi consentir ». Elle expose aussi la limite : tant qu'une ligne existe, un navigateur rare sur un petit réseau permet de cerner un visiteur plus précisément que le site ne le voudrait. Le compromis est moins de réponses et aucun entonnoir, contre aucun stockage sur l'appareil du lecteur ni envoi à autrui.

Où les problèmes surviennent

Deux outils annoncent deux nombres de visiteurs

L'outil hébergé donne un nombre pour le mois passé et le journal serveur le double, chacun mesurant correctement quelque chose de différent. Une balise ne compte que les navigateurs ayant exécuté son script et, derrière un bandeau, les visiteurs ayant accepté. Un journal compte chaque requête, robots compris, jusqu'à leur filtrage. Écrivez une fois la définition de chaque métrique en SQL, avec ses inclusions et exclusions.

Les rapports ralentissent l'application

Chaque lundi, l'application ralentit pendant que le rapport hebdomadaire parcourt des tables entières dans la base transactionnelle. Déplacez-le vers une réplique ou un entrepôt, ou agrégez chaque nuit les lignes brutes dans une petite table de synthèse sur laquelle exécuter des fonctions de fenêtre :

```
SELECT day,
       signups,
       sum(signups) OVER (ORDER BY day
                          ROWS BETWEEN 6 PRECEDING AND CURRENT ROW)
         AS last_7_days
FROM daily_counts
ORDER BY day;
```

Le bandeau disparaît, le suivi reste

Une équipe remplace les cookies par un produit « sans cookies », retire le bandeau et continue à suivre les visiteurs. L'article 5(3) concerne l'accès aux informations de l'appareil, quel que soit le mécanisme. Les lignes directrices du CEPD sur son périmètre technique précisent que les empreintes issues d'en-têtes HTTP ou le suivi par seule adresse IP peuvent aussi en relever, tout en rappelant que le fait de relever de cette règle ne signifie pas à lui seul que le consentement est nécessaire. Jugez un outil selon sa capacité à relier une requête à la suivante, pas selon le dépôt d'un cookie.

Un consentement rejeté par les régulateurs

Un bandeau bloque la page jusqu'à acceptation, considère le défilement comme accord ou rend le retrait plus difficile que l'acceptation. Les lignes directrices du CEPD indiquent qu'un mur de cookies empêche un consentement libre et que le défilement n'est jamais un acte affirmatif clair. L'article 7(3) du RGPD exige un retrait aussi facile que l'accord. Donnez autant de poids au refus qu'à l'acceptation et n'envoyez rien avant la réponse.

Des événements pseudonymes conservés sans limite

La table d'événements croît de millions de lignes par mois et ses identifiants utilisateurs hachés sont considérés comme anonymes. Le considérant 26 du RGPD traite comme personnelles les données pseudonymisées que des informations supplémentaires permettraient d'attribuer à quelqu'un. Le considérant 30 cite les adresses IP et identifiants de cookies parmi les identifiants en ligne dont les traces, combinées à d'autres informations, peuvent profiler et identifier une personne. Fixez la conservation avant le premier événement, partitionnez par temps pour supprimer un mois expiré d'un bloc et gardez les chiffres à long terme sous forme d'agrégats sans identifiants.

Pour aller plus loin

- NIST SP 1500-1r2, Big Data Interoperability Framework: Definitions — définitions de l'analyse, de son cycle de vie et de la localité des données à la base de cet article.
- Directive 2002/58/CE consolidée — article 5(3), règle de consentement pour le stockage ou la lecture sur l'appareil du visiteur, avec ses deux exceptions.
- Règlement (UE) 2016/679, le RGPD — définition du consentement, article 7 sur son retrait et considérants 26 et 30 sur les données pseudonymisées et identifiants en ligne.
- Lignes directrices 05/2020 du CEPD sur le consentement — murs de cookies, défilement et retrait, avec exemples détaillés.
- Lignes directrices 2/2023 du CEPD sur l'article 5(3) — pourquoi suivi par IP, empreintes et pixels de suivi peuvent relever de la même règle que les cookies.
- Tutoriel PostgreSQL : fonctions de fenêtre — classements et totaux cumulés sur des lignes liées, calculés dans la base.

Glossaire

Choisissez selon vos contraintes réelles

Présentez un jeu de données représentatif, les questions auxquelles il doit répondre et le goulot d'étranglement actuel. Nous pouvons convenir de la manière de mesurer la qualité, le débit et la reprise.

Consulter les tarifs Contactez-nous →

Rendez la terminologie concrète

Demandez comment « Qu'est-ce que l'analyse de données ? » s'applique à votre organisation. Un exemple concret tiré de votre système actuel suffit pour engager la discussion.

Découvrir nos prestations

welcome@engineer.company +45 50 10 61 84 Copenhague, Danemark

Emportez cette page avec vous

Fiche de contact

Créé avec soin et amour par D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

Droit d'auteur © 2025 – aujourd'hui · Engineer ApS · Tous droits réservés.

Imprimé depuis https://engineer.company/fr/glossary/analytics/