Engineer

Datenanalyse

die systematische Suche nach Mustern und Trends in Daten

Was es ist

Aufgezeichnete Fakten in Antworten zu verwandeln, die jemand als Handlungsgrundlage nutzen kann, ist die Aufgabe der Datenanalyse, häufig auch Analytics genannt. Im Big-Data-Rahmenwerk des US-Normungsinstituts NIST bedeutet Analyse, Daten systematisch zu bearbeiten und bei Bedarf umzuformen, um Muster, Zusammenhänge zwischen Datenpunkten und zeitliche Entwicklungen sichtbar zu machen und Vorhersagen zu versuchen. NIST ordnet diesen Schritt in einen größeren *Datenanalyse-Lebenszyklus* ein, der Rohdaten in nutzbares Wissen für eine Organisation überführt: sammeln, vorbereiten, analysieren, visualisieren und zugänglich machen.

Es ist eine Form der Datenverarbeitung, die sich von der transaktionalen durch das Gelesene unterscheidet: Eine Anwendung schreibt jeweils einen Auftrag, eine analytische Abfrage liest Millionen Zeilen für eine Handvoll Ergebnisse. Sie unterscheidet sich auch vom Data Engineering, das die Pipelines und Speicher baut, aus denen die Analyse liest.

Der Begriff umfasst sehr unterschiedliche Arbeit:

- **Berichte:** dieselben Kennzahlen nach Zeitplan als PDF oder E-Mail zustellen.
- **Dashboards:** diese Kennzahlen in einem Business-Intelligence-Werkzeug wie Power BI aktuell halten.
- **Ad-hoc-Analyse:** eine SQL-Abfrage für eine zuvor ungestellte Frage schreiben, häufig gegen ein Warehouse wie BigQuery.
- **Geodatenanalyse:** dieselbe Arbeit mit Kartendaten in einem GIS, wobei neben der Menge auch der Ort gefragt ist.
- **Analytische Anwendungen:** ein Modell oder statistisches Verfahren hinter einem Formular für Menschen, die keinen Code schreiben.

- **Web- und Produktanalyse:** die Besucher einer Website oder Nutzer eines Produkts zählen und erfassen, was sie dort tun.

Die letzte Art kostet Privatsphäre. Sie ist gemeint, wenn die Datenschutz-, Cookie-, Credits- und Sicherheitsseiten dieser Website von „keiner Webanalyse“ sprechen: kein Analyse-Tag, kein Zählpixel, kein Cookie und kein Konto bei einem Analyseunternehmen. Die Website misst sich trotzdem. Ihr Webserver führt ein Zugriffsprotokoll. Ein Programm macht daraus tägliche Zählungen und monatliche Ranglisten gelesener Seiten, verweisender Websites und Crawler. Die Angaben bleiben Zahlen und werden nie zu einem Weg eines einzelnen Besuchers zusammengesetzt.

Das Konzept für eine Entscheidung nutzen

Nutzen Sie „Was bedeutet Datenanalyse?“ für eine konkrete Entscheidung. Wir können helfen, das Konzept mit dem System in Verbindung zu bringen, das Sie entwickeln oder einkaufen. Dashboards und Berichte zeigen die wenigen Zahlen, von denen eine Entscheidung tatsächlich abhängt, und aktualisieren sich ohne Aufforderung.

Unsere Leistungen entdecken Kontakt aufnehmen →

Die Annahmen hinter dem Begriff prüfen

Nutzen Sie die Erklärung und die verlinkten Arbeiten, um Fragen für Ihr Projekt zu formulieren. Wir können die Annahmen gemeinsam mit Ihnen prüfen, bevor ein Werkzeug ausgewählt wird. Markenkommunikation anhand von Publikumsreaktionen — Markenkommunikation anhand von Publikumsreaktionen

Markenkommunikation anhand von Publikumsreaktionen Kontakt aufnehmen →

Warum es wichtig ist

Zwei Entscheidungen bestimmen den größten Teil der Kosten: wo die Abfragen laufen und wie Daten über Menschen erhoben werden.

Die Analyse auf der Transaktionsdatenbank ist der billigste Einstieg: Die Daten sind aktuell, und nichts wird kopiert. Dafür konkurriert jeder Bericht mit der Anwendung um Speicher und Platte. Eine Lesereplik nimmt diese Last vom Primärserver, zum Preis eines zweiten Servers. Ein Warehouse wie BigQuery trennt Speicher und Rechenleistung und berechnet bei nutzungsabhängigen Abfragen die gelesenen Bytes. Das belohnt eng begrenzte Abfragen, benötigt aber eine Pipeline zum Hineinkopieren der Daten.

Die Bindung an ein Werkzeug versteckt sich in den Definitionen. Ein BI-Werkzeug speichert Kennzahlen in seiner eigenen Formelsprache. Ein Jahr später steht die Bedeutung eines aktiven Kunden in einer Datei, die nur dieses Werkzeug öffnet. Definitionen als SQL im Repository überstehen einen Werkzeugwechsel.

Web- und Produktanalyse ergänzt eine rechtliche Frage. Das übliche gehostete Werkzeug arbeitet über ein Tag: ein Skript auf jeder Seite, das ein Cookie oder einen ähnlichen Bezeichner im Browser setzt und jeden Seitenaufruf an den Anbieter meldet. Artikel 5 Absatz 3 der EU-ePrivacy-Richtlinie verlangt für das Ablegen oder Auslesen von Informationen auf dem Gerät eines Besuchers eine Einwilligung, mit zwei Ausnahmen: Speicherung oder Zugriff ausschließlich zur Übertragung einer Nachricht über ein Netz sowie Speicherung oder Zugriff, die unbedingt erforderlich sind, um einen ausdrücklich angeforderten Dienst zu erbringen. Sobald Ereignisse einer Person zugeordnet werden können, auch über eine Cookie-Kennung oder IP-Adresse, gilt zudem die DSGVO: eine Rechtsgrundlage, ein benannter Zweck und eine Aufbewahrungsfrist. Außerdem besitzt nun ein weiteres Unternehmen Aufzeichnungen über das Verhalten Ihrer Besucher.

Die tatsächliche Wahl liegt damit zwischen einem schnell ergänzten gehosteten Werkzeug mit Dashboards, einer selbst gebauten und betriebenen Erfassung auf eigenen Systemen und Auswertungen aus ohnehin geschriebenen Serverprotokollen. Letztere beantworten weniger Fragen und halten deutlich weniger fest.

Wie es in der Praxis aussieht

Für eine berufliche Netzwerkplattform wurde Produktanalyse direkt in PostgreSQL gebaut, statt ein fremdes Tag einzubinden: ein Schema mit 7 Tabellen und 47 Funktionen, Ereignistabellen mit 22 Partitionierungsdefinitionen, die sich nach Zeitplan bereinigen, sowie Besucheridentitäten, die mit wechselndem Salt pseudonymisiert werden. Der Client sendet erst nach der Einwilligung. Funnels werden in der Datenbank neben den Daten berechnet, was NIST Datenlokalität nennt. Die Seite benennt den Preis: Wechselnder Salt verhindert langfristige Kohortenanalysen, und kostenlose Dashboards sind nicht dabei.

Auf der Berichtsseite holte ein Python-Berichtssystem die Daten, führte die Analyse aus und lieferte denselben PDF-Bericht nach einem verlässlichen Zeitplan an die kaufmännischen Entscheider. Die Seite schreibt dieser Berichterstattung eine Steigerung des vierteljährlichen Softwareumsatzes um 400 % zu.

Diese Website ist das dritte Beispiel. Ihre Datenschutzseite beschreibt die Messung: ein Zugriffsprotokoll, das vor dem Schreiben den letzten Teil jeder IP-Adresse auf null setzt, die Herkunftswebsite nur als Namen und Zeilen mit einer Lebensdauer von etwa dreißig Tagen. Ein Programm erstellt daraus eine tägliche Zählung und eine monatliche Rangliste. Beide bleiben unbegrenzt als Summen erhalten, ohne etwas, das sie wieder zu einem Besuch verbinden könnte. Ohne Tag, Cookie oder Dritte gibt es kein Einwilligungsbanner; die Begründung lautet, dass es nichts gibt, in das einzuwilligen wäre. Auch die Grenze wird offengelegt: Solange eine Zeile existiert, grenzt ein seltener Browser in einem kleinen Netz einen Besucher stärker ein, als der Website lieb ist. Weniger Antworten und keine Funnels sind der Preis dafür, nichts auf dem Gerät des Lesers zu speichern und nichts an andere zu senden.

Wo es schiefgeht

Zwei Werkzeuge melden zwei Besucherzahlen

Das gehostete Werkzeug zeigt für den letzten Monat eine Zahl, das Serverprotokoll doppelt so viel. Beide können bei unterschiedlichen Gegenständen recht haben. Ein Tag zählt nur Browser, die sein Skript ausgeführt haben, und hinter einem Banner nur zustimmende Besucher. Ein Protokoll zählt jede Anfrage einschließlich Crawlern, bis jemand sie herausfiltert. Schreiben Sie die Definition jeder Kennzahl einmal in SQL auf, einschließlich der ein- und ausgeschlossenen Fälle.

Berichte bremsen die Anwendung

Jeden Montag wird die Anwendung langsam, während der Wochenbericht ganze Tabellen auf der Transaktionsdatenbank scannt. Verlagern Sie ihn auf eine Replik oder ein Warehouse. Oder verdichten Sie Rohzeilen jede Nacht in eine kleine Aggregattabelle und führen darauf Fensterfunktionen aus:

```
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;
```

Das Banner verschwindet, das Tracking bleibt

Ein Team ersetzt Cookies durch ein „cookieless“-Produkt, entfernt das Banner und verfolgt Besucher weiter. Artikel 5 Absatz 3 betrifft den Zugriff auf Informationen auf dem Gerät, unabhängig vom Mechanismus. Die EDPB-Leitlinien zum technischen Geltungsbereich erklären, dass auch Fingerprinting aus HTTP-Headern oder die Verfolgung von Navigation allein über IP-Adressen darunterfallen können. Zugleich erinnern sie daran, dass die Anwendbarkeit der Regel allein noch keine Einwilligungspflicht bedeutet. Beurteilen Sie ein Werkzeug danach, ob es eine Anfrage mit der nächsten verknüpft, und nicht danach, ob es ein Cookie setzt.

Eine Einwilligung, die Aufsichtsbehörden ablehnen würden

Ein Banner blockiert die Seite bis zur Zustimmung, wertet Scrollen als Einwilligung oder erschwert den Widerruf gegenüber der Zustimmung. Nach den EDPB-Leitlinien wird eine Einwilligung hinter einer Cookie-Wall nicht freiwillig erteilt, und Scrollen ist niemals eine eindeutige bestätigende Handlung. Artikel 7 Absatz 3 der DSGVO verlangt, den Widerruf so einfach wie die Erteilung zu machen. Geben Sie Ablehnung dasselbe Gewicht wie Zustimmung und senden Sie vor der Antwort nichts.

Pseudonyme Ereignisse bleiben für immer

Die Ereignistabelle wächst monatlich um Millionen Zeilen, und ihre gehashten Nutzerkennungen gelten als anonym. Erwägungsgrund 26 der DSGVO behandelt pseudonymisierte Daten, die mit Zusatzinformationen einer Person zugeordnet werden könnten, als personenbezogen. Erwägungsgrund 30 nennt IP-Adressen und Cookie-Kennungen als Online-Kennungen, deren Spuren zusammen mit anderen Informationen Profilbildung und Identifizierung ermöglichen können. Legen Sie die Frist vor dem ersten Ereignis fest. Partitionieren Sie nach Zeit, damit ein abgelaufener Monat vollständig entfällt, und bewahren Sie langfristige Zahlen als Aggregate ohne Kennungen auf.

Weiterführende Lektüre

- NIST SP 1500-1r2, Big Data Interoperability Framework: Definitions — die Definitionen von Analyse, Datenanalyse-Lebenszyklus und Datenlokalität, auf denen dieser Artikel aufbaut.
- Richtlinie 2002/58/EC, konsolidiert — Artikel 5 Absatz 3 zur Einwilligung für das Speichern oder Lesen auf Besuchergeräten, einschließlich beider Ausnahmen.

- Verordnung (EU) 2016/679, die DSGVO — die Einwilligungsdefinition, Artikel 7 zum Widerruf sowie Erwägungsgründe 26 und 30 zu pseudonymisierten Daten und Online-Kennungen.
- EDPB-Leitlinien 05/2020 zur Einwilligung — Cookie-Walls, Scrollen und Widerruf mit ausgearbeiteten Beispielen.
- EDPB-Leitlinien 2/2023 zu Artikel 5 Absatz 3 — weshalb IP-Tracking, Fingerprinting und Zählpixel derselben Regel wie Cookies unterliegen können.
- PostgreSQL-Tutorial: Fensterfunktionen — Rangfolgen und laufende Summen verwandter Zeilen, berechnet innerhalb der Datenbank.

Glossar

Für Ihre tatsächlichen Rahmenbedingungen auswählen

Bringen Sie einen repräsentativen Datensatz, die Fragen, die er beantworten muss, und den aktuellen Engpass mit. Wir können vereinbaren, wie Qualität, Durchsatz und Wiederherstellung gemessen werden.

Preise ansehen Kontakt aufnehmen →

Die Begriffe praktisch nutzbar machen

Fragen Sie, wie „Was bedeutet Datenanalyse?“ auf Ihre Organisation zutrifft. Ein konkretes Beispiel aus Ihrem bestehenden System reicht aus, um das Gespräch zu beginnen.

Unsere Leistungen entdecken

welcome@engineer.company +45 50 10 61 84 Kopenhagen, Dänemark

Nehmen Sie diese Seite mit

Kontaktkarte

Mit Sorgfalt und Liebe gestaltet von D.

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

Copyright © 2025 – heute · Engineer ApS · Alle Rechte vorbehalten.

Gedruckt von https://engineer.company/de/glossary/analytics/