

O'REILLY®

Google Analytics od podstaw

Analiza wpływu biznesowego
i wyznaczanie trendów



Mark Edmondson

Helion 

Tytuł oryginału: Learning Google Analytics: Creating Business Impact and Driving Insights

Tłumaczenie: Lech Lachowski

ISBN: 978-83-8322-849-5

© 2023 Helion S.A.

Authorized Polish translation of the English edition of *Learning Google Analytics*
ISBN 9781098113087 © 2023 Mark Edmondson.

This translation is published and sold by permission of O'Reilly Media, Inc.,
which owns or controls all rights to publish and sell the same.

All rights reserved. No part of this book may be reproduced or transmitted in any
form or by any means, electronic or mechanical, including photocopying, recording
or by any information storage retrieval system, without permission from the Publisher.

Wszelkie prawa zastrzeżone. Nieautoryzowane rozpowszechnianie całości
lub fragmentu niniejszej publikacji w jakiegokolwiek postaci jest zabronione.
Wykonywanie kopii metodą kserograficzną, fotograficzną, a także kopiowanie
książki na nośniku filmowym, magnetycznym lub innym powoduje naruszenie
praw autorskich niniejszej publikacji.

Wszystkie znaki występujące w tekście są zastrzeżonymi znakami firmowymi
bądź towarowymi ich właścicieli.

Autor oraz wydawca dołożyli wszelkich starań, by zawarte w tej książce informacje
były kompletne i rzetelne. Nie biorą jednak żadnej odpowiedzialności ani za ich
wykorzystanie, ani za związane z tym ewentualne naruszenie praw patentowych
lub autorskich. Autor oraz wydawca nie ponoszą również żadnej odpowiedzialności
za ewentualne szkody wynikłe z wykorzystania informacji zawartych w książce.

Helion S.A.

ul. Kościuszki 1c, 44-100 Gliwice

tel. 32 230 98 63

e-mail: helion@helion.pl

WWW: <https://helion.pl> (księgarnia internetowa, katalog książek)

Drogi Czytelniku!

Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres

<https://helion.pl/user/opinie/goanod>

Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Printed in Poland.

- Kup książkę
- Poleć książkę
- Oceń książkę

- Księgarnia internetowa
- Lubię to! » Nasza społeczność

| | |
|---|-----------|
| Wstęp | 9 |
| 1. Nowa usługa Google Analytics 4 | 13 |
| Wprowadzenie do GA4 | 13 |
| Ujednoczenie analityki mobilnej i internetowej | 14 |
| Firebase i BigQuery — pierwsze kroki w chmurze | 15 |
| Wdrożenie GA4 | 15 |
| Porównanie Universal Analytics z GA4 | 16 |
| Model danych GA4 | 18 |
| Zdarzenia | 19 |
| Parametry niestandardowe | 20 |
| Elementy e-commerce | 21 |
| Właściwości użytkownika | 22 |
| Google Cloud Platform | 23 |
| Odpowiednie usługi GCP | 23 |
| Umiejętności kodowania | 24 |
| Wprowadzenie do GCP | 26 |
| Poruszanie się w górę piramidy bezserwerowej | 27 |
| Podsumowanie wprowadzenia do GCP | 29 |
| Wprowadzenie do przypadków użycia | 29 |
| Przypadek użycia — predykcyjna analiza zakupów | 30 |
| Przypadek użycia — segmentacja odbiorców | 32 |
| Przypadek użycia — prognozowanie w czasie rzeczywistym | 33 |
| Podsumowanie | 34 |
| 2. Architektura danych i strategia | 36 |
| Tworzenie środowiska do odniesienia sukcesu | 36 |
| Zaangażowanie zainteresowanych stron | 36 |
| Podejście oparte na analizie przypadków użycia i unikaniu statków kosmicznych | 37 |
| Demonstrowanie wartości biznesowej | 38 |

| | |
|--|-----------|
| Ocena dojrzałości cyfrowej | 38 |
| Ustalanie priorytetów przypadków użycia | 39 |
| Wymagania techniczne | 39 |
| Pozyskiwanie danych | 40 |
| Przechowywanie danych | 42 |
| Modelowanie danych | 45 |
| Porównanie wydajności modelu i wartości biznesowej | 45 |
| Zasada najmniejszego przenoszenia (danych) | 46 |
| Od wejść danych nieprzetworzonych do wyjść danych informacyjnych | 46 |
| Pomoc dla badaczy danych lub osób modelujących dane | 47 |
| Ustalenie wskaźników KPI modelu | 47 |
| Ostateczna lokalizacja modelu | 48 |
| Aktywacja danych | 49 |
| To nie musi być dashboard | 49 |
| Interakcja z użytkownikami końcowymi | 49 |
| Prywatność użytkowników | 51 |
| Przestrzeganie wyborów ochrony prywatności użytkowników | 52 |
| Prywatność uwzględniona w projekcie | 52 |
| Pomocne narzędzia | 53 |
| gcloud | 53 |
| System kontroli wersji (Git) | 54 |
| Zintegrowane środowiska programistyczne | 54 |
| Kontenery (w tym Docker) | 54 |
| Podsumowanie | 55 |
| 3. Pozyskiwanie danych | 56 |
| Rozbijanie silosów danych | 56 |
| Mniej znaczy więcej | 56 |
| Określanie schematu danych | 57 |
| Konfiguracja GA4 | 57 |
| Typy zdarzeń GA4 | 58 |
| Przechwytywanie zdarzeń GA4 za pomocą menedżera tagów | 61 |
| Konfiguracja niestandardowych pól | 64 |
| Modyfikowanie lub tworzenie zdarzeń GA4 | 66 |
| Właściwości użytkownika | 68 |
| Measurement Protocol v2 | 75 |
| Eksportowanie danych GA4 za pośrednictwem interfejsów API | 77 |
| Uwierzelnianie przy użyciu interfejsu Data API | 79 |
| Uruchamianie zapytań Data API | 80 |
| BigQuery | 82 |
| Połączenie GA4 z BigQuery | 82 |
| SQL BigQuery dla eksportów GA4 | 83 |

| | |
|--|------------|
| BigQuery dla innych źródeł danych | 85 |
| Publiczne zbiory danych BigQuery | 85 |
| Menedżer tagów po stronie serwera | 85 |
| Google Cloud Storage | 87 |
| Pamięć masowa oparta na zdarzeniach | 88 |
| Prywatność danych | 98 |
| Importy bazy danych CRM za pośrednictwem usługi GCS | 98 |
| Konfiguracja potoków CI/CD Cloud Build za pomocą GitHuba | 99 |
| Konfigurowanie GitHuba | 99 |
| Konfigurowanie połączenia GitHuba z Cloud Build | 99 |
| Dodawanie plików do repozytorium | 102 |
| Podsumowanie | 104 |
| 4. Przechowywanie danych | 105 |
| Zasady dotyczące danych | 105 |
| Dane uporządkowane | 106 |
| Zbiory danych dla różnych ról | 111 |
| BigQuery | 113 |
| Kiedy używać BigQuery? | 113 |
| Organizowanie zbiorów danych | 114 |
| Wskazówki dotyczące tablic | 116 |
| Pub/Sub | 117 |
| Konfigurowanie tematu Pub/Sub dla eksportu z GA4 do BigQuery | 118 |
| Tworzenie partycjonowanych tabel BigQuery z eksportu GA4 | 119 |
| Wysyłanie do Pub/Sub po stronie serwera | 121 |
| Firestore | 122 |
| Kiedy korzystać z Firestore'a? | 123 |
| Uzyskiwanie dostępu do danych Firestore'a za pośrednictwem API | 123 |
| GCS | 125 |
| Planowanie importów danych | 129 |
| Rodzaje importu danych — przesyłanie strumieniowe a zaplanowane dane wsadowe | 129 |
| Widoki BigQuery | 130 |
| Zaplanowane kwerendy BigQuery | 131 |
| Cloud Composer | 133 |
| Cloud Scheduler | 138 |
| Cloud Build | 139 |
| Strumieniowe przepływy danych | 146 |
| Pub/Sub do strumieniowego przesyłania danych | 147 |
| Apache Beam (Dataflow) | 147 |
| Przesyłanie strumieniowe za pośrednictwem Cloud Functions | 153 |

| | |
|---|------------|
| Ochrona prywatności użytkowników | 155 |
| Prywatność danych uwzględniona w projekcie | 156 |
| Data wygaśnięcia w BigQuery | 158 |
| Data Loss Prevention API | 159 |
| Podsumowanie | 159 |
| 5. Modelowanie danych | 160 |
| Model danych za pomocą GA4 | 160 |
| Raporty standardowe i eksploracje | 160 |
| Modelowanie atrybucji | 161 |
| Rozwiązywanie użytkownika i sesji | 163 |
| Modelowanie trybu zgody | 164 |
| Tworzenie grup odbiorców | 165 |
| Dane prognozowane | 165 |
| Trendy | 165 |
| Przekształcanie danych w trendy | 167 |
| Ustalanie zakresu dla wyników danych | 167 |
| Dokładność i korzyść przyrostowa | 170 |
| Wybór metody | 171 |
| Utrzymywanie aktualności potoków modelowania | 172 |
| Łączenie zbiorów danych | 173 |
| BigQuery ML | 175 |
| Porównanie modeli BigQuery ML | 175 |
| Wprowadzenie modelu do środowiska produkcyjnego | 176 |
| Interfejsy API uczenia maszynowego | 178 |
| Wprowadzenie ML API do środowiska produkcyjnego | 179 |
| Google Cloud AI — Vertex AI | 180 |
| Wprowadzenie Vertex API do środowiska produkcyjnego | 182 |
| Integracja z R | 183 |
| Przegląd możliwości | 183 |
| Docker | 185 |
| R w środowisku produkcyjnym | 187 |
| Podsumowanie | 188 |
| 6. Aktywacja danych | 189 |
| Znaczenie aktywacji danych | 190 |
| Odbiorcy GA4 i Google Marketing Platform | 191 |
| Google Optimize | 195 |
| Wizualizacja | 198 |
| Zapewnianie efektywności dashboardu | 199 |
| Opcje konfiguracji dashboardów w GA4 | 200 |
| Data Studio | 211 |

| | |
|---|------------|
| Looker | 214 |
| Zewnętrzne narzędzia do wizualizacji | 216 |
| Zagregowane tablice umożliwiające podejmowanie decyzji opartych na danych | 217 |
| Buforowanie danych i zarządzanie kosztami | 217 |
| Tworzenie marketingowych interfejsów API | 218 |
| Tworzenie mikrousług | 218 |
| Aktywatory zdarzeń | 220 |
| Integracje Firestore'a | 223 |
| Podsumowanie | 227 |
| 7. Przypadek użycia — predykcyjna analiza zakupów | 228 |
| Tworzenie uzasadnienia biznesowego | 228 |
| Ocena wartości | 229 |
| Szacowanie zasobów | 230 |
| Architektura danych | 230 |
| Pozyskiwanie danych — konfiguracja GA4 | 231 |
| Przechowywanie danych i projektowanie prywatności | 232 |
| Modelowanie danych — eksportowanie odbiorców do Google Ads | 233 |
| Aktywacja danych — testowanie wydajności | 235 |
| Podsumowanie | 237 |
| 8. Przypadek użycia — segmentacja odbiorców | 238 |
| Tworzenie uzasadnienia biznesowego | 238 |
| Ocena wartości | 239 |
| Szacowanie zasobów | 240 |
| Architektura danych | 243 |
| Pozyskiwanie danych | 243 |
| Konfiguracja przechwytywania danych GA4 | 244 |
| Eksporty BigQuery GA4 | 245 |
| Przechowywanie danych — transformacje zbiorów danych | 247 |
| Modelowanie danych | 248 |
| Aktywacja danych | 249 |
| Konfigurowanie importów GA4 za pośrednictwem GTM SS | 252 |
| Eksportowanie odbiorców z GA4 | 255 |
| Testowanie wydajności | 256 |
| Podsumowanie | 257 |
| 9. Przypadek użycia — prognozowanie w czasie rzeczywistym | 259 |
| Tworzenie uzasadnienia biznesowego | 259 |
| Potrzebne zasoby | 260 |
| Architektura danych | 261 |

| | |
|--|------------|
| Pozyskiwanie danych | 262 |
| Konfiguracja GA4 | 262 |
| Przechowywanie danych | 264 |
| Hosting aplikacji Shiny w Cloud Run | 264 |
| Modelowanie danych | 266 |
| Aktywacja danych — dashboard w czasie rzeczywistym | 270 |
| Kod R dla aplikacji Shiny czasu rzeczywistego | 270 |
| Uwierzytelnianie GA4 za pomocą konta usługi | 271 |
| Poskładanie wszystkiego razem w aplikacji Shiny | 275 |
| Podsumowanie | 279 |
| 10. Kolejne kroki | 281 |
| Motywacja — jak dowiedziałem się, co znajdzie się w książce? | 281 |
| Pomoce naukowe | 283 |
| Proszenie o pomoc | 286 |
| Certyfikacja | 287 |
| Uwagi końcowe | 287 |

Nowa usługa Google Analytics 4

W tym rozdziale przedstawię nową usługę Google Analytics 4 (GA4) i zbadam przyczyny, które doprowadziły do jej opracowania. Sprawdzę, jakie braki znalazło Google w poprzedniku tej usługi, Universal Analytics, i jak GA4 ma wzmocnić te obszary dzięki przyjęciu nowego modelu danych.

Pokażę również, w jaki sposób integracja GA4 z platformą GCP (ang. *Google Cloud Platform*) zwiększa funkcjonalność tej usługi, i zaprezentuję pierwsze przypadki użycia, które pomogą zilustrować nowe możliwości GA4, abyś mógł (mogła) rozpocząć pracę z własnymi projektami danych.

Wprowadzenie do GA4

Usługa Google Analytics 4 została wydana na podstawie wersji beta i wprowadzona jako nowe Google Analytics na początku 2021 r. Jej nazwa beta „App+Web” została zastąpiona przez Google Analytics 4.

Kluczowe różnice między GA4 i Universal Analytics podkreślone w poście wprowadzającym GA4 (<https://oreil.ly/kj6TL>) to funkcjonalności uczenia maszynowego, ujednolicony schemat danych dla aplikacji internetowych i mobilnych oraz projektowanie oparte na prywatności.

Google planowało wydanie GA4 przez wiele lat przed jego publicznym ogłoszeniem. Po wydaniu Google Analytics stało się najpopularniejszym systemem analityki internetowej, jednak w 2021 r. jego projekt nadal odzwierciedlał cele projektowe z poprzednich 15 lat. Chociaż ta platforma była przez lata ulepszana przez wyznaczony zespół Google Analytics, istniało kilka nowoczesnych wyzwań, które były trudniejsze do realizacji: użytkownicy prosili o widoki pojedynczych klientów dla aplikacji internetowych i mobilnych zamiast wysyłania danych do dwóch oddzielnych właściwości, Google Cloud był liderem w technologiach uczenia maszynowego, ale uczenie maszynowe nie było łatwe do zintegrowania z modelem danych GA, a prywatność użytkowników stawała się coraz większym problemem, który wymagał ściślejszej kontroli nad tym, gdzie przepływają dane analityczne.

Kiedy usługa Google Analytics została uruchomiona po raz pierwszy w 2005 r., wprowadziła branżę analityczną w konsternację, gdyż oferowała w pełni funkcjonalną bezpłatną wersję tego, co wcześniej było dostępne tylko w płatnych produktach dla przedsiębiorstw. Ponieważ Google zdało sobie sprawę, że im więcej webmasterów będzie miało dane na temat swojego ruchu, tym

bardziej prawdopodobne będzie, że zainwestują oni w AdWords (teraz Google Ads), Google Analytics było inwestycją korzystną dla wszystkich, która dała każdemu możliwość usłyszenia głosu swoich użytkowników podczas przeglądania ich witryn.

Gdy nastał rok 2020 krajobraz analityczny był już zupełnie inny. Konkurencja wprowadziła produkty analityczne z prostszymi modelami danych, które mogły działać z różnymi źródłami danych i były bardziej dostosowane do uczenia maszynowego i prywatności (niezbędnej funkcjonalności użytkownika). Możliwe było użycie chmury, aby uczynić system analityczny bardziej otwartym i zapewnić większą kontrolę specjalistom ds. analityki. Konkurencyjne rozwiązania analityczne mogły być uruchamiane nawet na własnej infrastrukturze chmurowej Google'a, co zmieniło ekonomikę budowania lub kupowania. Idealne rozwiązania analityczne miały mieć rozsądne wartości domyślne dla osób szukających szybkiego rozpoczęcia, ale miały być zarazem bardziej konfigurowalne i skalowalne, aby zaspokoić potrzeby odważniejszych klientów.

Ujednoczenie analityki mobilnej i internetowej

Chociaż w momencie premiery poprzednia nazwa „App+Web” została zastąpiona przez GA4, znacznie lepiej oddawała ona odmienną charakterystykę tej platformy.

Aż do końca 2019 r. Google Analytics dla aplikacji mobilnych (dla systemów Android i iOS) miało własny system analityczny, odrębny od analityki internetowej. Te pakiety SDK wykorzystywały inny model danych, który był bardziej odpowiedni do analizy aplikacji, gdzie pojęcia takie jak odsłony stron, sesje i użytkownicy oznaczały nieco inne rzeczy, przez co nie można było ich łatwo porównać z danymi internetowymi. Użytkownicy, którzy odwiedzili zarówno aplikację mobilną, jak i internetową, zwykle nie byli powiązani.

Model danych GA4 wykorzystuje konfigurowalną strukturę opartą wyłącznie na danych, która została przyjęta przez aplikacje mobilne. Universal Analytics nakładało ograniczenia na to, kiedy dane mogły być łączone (tzw. zakresy danych), co oznaczało, że marketingowcy musieli zastanowić się, czy ich dane będą pasować do konkretnych zakresów, takich jak użytkownik, sesja lub zdarzenia. Zakresy zostały z góry określone przez firmę Google, więc było się zmuszonym do przyjęcia jej modelu danych. Dzięki podejściu GA4 opartemu wyłącznie na zdarzeniach zyskujesz większą elastyczność w określaniu, jak mają wyglądać Twoje dane.

Kiedy pod koniec 2019 r. stare pakiety SDK Google Analytics dla urządzeń mobilnych zaczęły odchodzić w zapomnienie, Google zachęcało użytkowników do przechodzenia na pakiety SDK Firebase'a. Firebase został opracowany jako kompletne doświadczenie programistów mobilnych dla iOS-a i Androida z zintegrowanym mobilnym pakietem SDK do tworzenia aplikacji mobilnych od podstaw, również z uwzględnieniem analityki aplikacji internetowych. Nowe GA4 reprezentowało dodatkowy strumień danych: nowy strumień aplikacji internetowych. Korzystanie przez strumienie aplikacji iOS-a, Androida i internetowych z tego samego systemu oznacza, że mamy teraz naprawdę połączony sposób zapewniania analityki cyfrowej opartej na wszystkich tych źródłach.

Firestore i BigQuery — pierwsze kroki w chmurze

Dla wielu marketingowców GA4 jest wprowadzeniem do nowych produktów chmurowych, które są integralną częścią działania tej platformy: Firestore'a i BigQuery.

Firestore i BigQuery to produkty dostępne w ramach GCP, szerokiej usługi oferowanej przez Google'a dla wszystkich rodzajów usług w chmurze. W tej książce skoncentrowano się na tych produktach, które są częścią tej chmurowej oferty związanej z analityką danych, ale pamiętaj, że są one tylko podzbiorem całej platformy chmurowej.

Firestore to obszerny framework rozwoju oprogramowania mobilnego, który obejmuje teraz platformę Google Analytics. Programiści mobilni używają go również, aby zapewnić aplikacjom mobilnym bezserwerowość dzięki przydatnym funkcjonalnościom, takim jak zdalna konfiguracja do zmieniania kodu wdrożonych aplikacji bez konieczności ponownego publikowania ich w sklepie z aplikacjami, interfejsy API uczenia maszynowego (np. modelowanie predykcyjne), uwierzytelnianie, alerty mobilne i integracje z systemem reklamowym Google'a. Firestore to podzbiór usług platformy GCP, które w niektórych przypadkach stanowią rebranding bazowego produktu GCP, np. Firestore Cloud Functions są tym samym co GCP Cloud Functions.

BigQuery można uznać za jeden z klejnotów GCP; jest uważany za jeden z najbardziej atrakcyjnych produktów tej platformy w porównaniu z odpowiednikiem działającym u innych dostawców usług w chmurze. BigQuery to baza danych SQL dostosowana do analitycznych obciążeń roboczych, która była jedną z pierwszych dostępnych bezserwerowych baz danych. Obejmuje innowacje, takie jak model cenowy zapewniający niskobudżetowe przechowywanie danych przy naliczaniu opłat na żądanie dla zapytań oraz błyskawiczny silnik zapytań działający na Dremelu, który oferuje w niektórych przypadkach stukrotne przyspieszenie w porównaniu z MySQL-em. Użytkownicy GA360 mogą być z nim zaznajomieni, ponieważ jedną z jego funkcjonalności było eksportowanie surowych, niepróbkowanych danych do BigQuery — ale tylko wtedy, gdy zakupiło się licencję GA360 (to było moje wprowadzenie do chmury!). W GA4 eksportowanie do BigQuery będzie dostępne dla wszystkich, co jest ekscytujące, ponieważ sam BigQuery jest bramą do reszty usług GCP. BigQuery to jeden z tematów przewodnich w tej książce.

Wdrożenie GA4

Ta książka nie jest wyczerpującym przewodnikiem dotyczącym implementowania GA4; więcej informacji na ten temat znajdziesz w zasobach wymienionych w rozdziale 10. Omówię jednak w niej typowe konfiguracje, które dadzą pełen obraz, od gromadzenia danych po wartość biznesową.

Istnieją zasadniczo trzy sposoby konfigurowania przechwytywania danych ze stron internetowych: *gtag.js*, *analytics.js* i GTM (ang. *Google Tag Manager*), czyli menedżer tagów. W prawie wszystkich przypadkach polecam implementowanie przechwytywania za pomocą menedżera tagów, co omówię szerzej w rozdziale 3. Powodem tego jest elastyczność i możliwość oddzielenia pracy obiektu warstwy danych (*dataLayer*) od konfiguracji analitycznej, co minimalizuje nakłady pracy programistycznej wymagane w ramach witryny HTML-owej. Zasoby programistów są najbardziej efektywne przy wdrażaniu uporządkowanej warstwy danych dla menedżera tagów, albowiem obejmują wszystkie potrzeby śledzenia, a nie tylko tagi GA4 lub Google'a. Wszelkie

dotatkowe zmiany w konfiguracji śledzenia można potem wprowadzać w interfejsie internetowym menedżera tagów bez konieczności ponownego poświęcania cennego czasu na programowanie każdej drobnej edycji.

Dzięki wprowadzeniu menedżera tagów po stronie serwera (ang. *Server Side* — SS) możliwe konfiguracje mogą obejmować również bezpośrednie integracje z Google Cloud i systemami backendowymi oraz modyfikacje żądań i odpowiedzi wywołań HTTP, zapewniając najwyższą elastyczność.

Porównanie Universal Analytics z GA4

Mówi się, że GA4 jest ewolucją swojego poprzednika, Universal Analytics (od czasu wydania GA4 nazywanego GA3), ale czym właściwie się różni?

Gdy ludzie dowiadują się o GA4, z reguły pytają najpierw, czy różni się ono na tyle, by powinni rozważyć zmianę usługi. Pytają również, dlaczego mieliby zadawać sobie trud zmiany narzędzi, przeszkolenia i ponownego uczenia się nowego systemu, jeśli stary działał dobrze przez ostatnie 15 lat. Jest to kluczowe pytanie, które przeanalizuję szczegółowo w tym punkcie.

Do tej kwestii odnosi się także specjalny temat pomocy Google'a (<https://oreil.ly/G0ePW>).

Nowy model danych

Pierwsza duża zmiana dotyczy samego modelu danych, który omówię w podrozdziale „Model danych GA4”.

Universal Analytics był bardzo skoncentrowany na wskaźnikach strony internetowej, gdzie można było łatwiej zdefiniować takie koncepcje jak użytkownicy, sesje i odsłony stron. Koncepcje te były jednak trudniejsze do zdefiniowania dla innych źródeł danych, np. aplikacji mobilnych i odwołań do serwera. Często oznaczało to konieczność uwzględnienia obejść lub zignorowania niektórych wskaźników w raportach, gdy dane pochodziły z określonych źródeł. Oznaczało to również, że niektóre wskaźniki nie współdziałały ze sobą zbyt dobrze lub nie można było ich kwerendować.

GA4 przechodzi od narzuconego schematu danych do czegoś o wiele swobodniejszego: teraz wszystko jest zdarzeniem. Ta elastyczność pozwala łatwiej definiować własne wskaźniki, ale użytkownicy, którzy nie chcą wchodzić na ten poziom szczegółowości, mają do dyspozycji także domyślne automatyczne typy zdarzeń, zapewniające im kilka znajomych wskaźników.

Poza tym oznacza to, że teraz możliwe jest automatyczne zbieranie określonych danych, które wcześniej trzeba było konfigurować osobno, np. kliknięć linków. Z tego względu poprawne implementacje GA4 powinny wymagać mniejszego doświadczenia, co pomaga obniżyć barierę wejścia dla nowych użytkowników analityki cyfrowej. Mniej istotna staje się specjalistyczna wiedza, taka jak znajomość różnicy między wskaźnikiem sesji a wskaźnikiem odwołań.

Bardziej elastyczne podejście do wskaźników

Zdarzenia GA4 można modyfikować po ich wysłaniu (<https://oreil.ly/rtmxb>). Pozwala to korygować błędy śledzenia lub normalizować zdarzenia (np. „sprzedaż” w porównaniu z „transakcją”) bez konieczności modyfikowania skryptów śledzenia — jest to o wiele łatwiejsze do zrobienia.

Przy tworzeniu niestandardowych definicji dla własnych zdarzeń nie ma predefiniowanego schematu, który trzeba zapamiętać. Wystarczy utworzyć zdarzenie z opcjonalnymi parametrami i zarejestrować je w interfejsie GA4, aby zaczęło się ono pojawiać w raportach.

Eksportowanie z BigQuery

Eksportowanie z BigQuery, które było wcześniej funkcjonalnością GA360, jest teraz dostępne, nawet jeśli nie płaci się za korporacyjną wersję GA4. Firebase Analytics dla urządzeń mobilnych ma tę funkcjonalność od początku, a ponieważ GA4 jest tylko dodatkiem do tego, analityka aplikacji internetowych również ją ma.

Jest to przełom, albowiem zazwyczaj najtrudniejszą częścią projektu danych jest uzyskanie dostępu do skrywanych przez aplikacje surowych danych w taki sposób, aby można było łatwo z nimi pracować. Dzięki eksportom BigQuery w GA4 wystarczy wypełnić tylko kilka formularzy internetowych, aby uzyskać przepływ danych w czasie zbliżonym do rzeczywistego, gotowy do analizy za pomocą SQL-a BigQuery.

Ponieważ BigQuery jest tak bardzo zintegrowane z resztą platformy GCP, oznacza to również, że jest ściśle zintegrowane z resztą stosu danych GCP, np. usługami Pub/Sub, Dataflow i Data Studio. Pozwalają one przesyłać dane potokiem bezpośrednio z BigQuery, a skoro jego interfejsy API są otwarte, jest to popularne źródło lub ujście danych także dla wielu zewnętrznych usług.

Wszystko to oznacza, że odwieczny problem silosów danych, w których potrzebne dane są zablokowane za murami baz danych o różnych regułach i politykach firmy, można teraz rozwiązać przez wysyłanie wszystkiego do jednego miejsca docelowego — do BigQuery. W ten sposób można zacząć łączyć sprzedaż i marketing lub łatwiej pobierać przydatne zewnętrzne dane, takie jak prognoza pogody. Z mojego doświadczenia wynika, że przeniesienie wszystkich użytecznych danych w jedno miejsce ma największy wpływ na cyfrową dojrzałość klienta, ponieważ usuwa jedną z najczęstszych przeszkód: „Jak pozyskać dane?”.

Brak próbkowania — wszystko jest w czasie rzeczywistym

Motywacją do eksportowania BigQuery w GA360 było to, że był to jeden ze sposobów uzyskania niepróbkowanych danych. Teraz ma to zastosowanie również do GA4. O ile w internetowym interfejsie użytkownika limity próbkowania zostały poprawione, o tyle bazowe dane są zawsze niepróbkowane i dostępne w czasie rzeczywistym. Jeśli kiedykolwiek będziesz potrzebować niepróbkowanego eksportu, to wiedz, że jest on dostępny za pośrednictwem BigQuery lub za pomocą bezpłatnego Data API. Dzięki temu zapewnianie danych dla określonych przypadków użycia, które wymagają wysokiej dokładności i źródeł danych analitycznych w czasie rzeczywistym, nie wymaga już płacenia za GA360.

Prywatność i dane analityki cyfrowej

Użytkownicy są obecnie znacznie bardziej świadomi wartości swoich danych, a prywatność stała się gorącym tematem w branży. Istnieje przekonanie, że użytkownicy potrzebują możliwości w pełni świadomego wyrażania zgody na to, gdzie ich dane będą wykorzystywane, a obowiązkiem witryny

internetowej jest zdobycie zaufania i prawidłowa wycena tych danych. W tym zakresie pomocny jest Google Consent Mode, czyli tryb uzyskiwania zgody, który umożliwia usunięcie plików *cookie* i przechowywanych przez nie identyfikatorów osobowych, aby nie były one dostępne dla Google Analytics, dopóki użytkownik nie wyrazi na to zgody. Jednak dane nieosobowe mogą być nadal przydatne, a GA4 oferuje sposób modelowania tego, jak wyglądałyby sesje danych i konwersje, gdyby 100% użytkowników wyraziło zgodę na podanie swoich danych. Ponieważ najczęściej nowi klienci najprawdopodobniej nie będą jeszcze ufać Twojej witrynie lub nie wyrażą zgody, może to być cenna informacja, która pomoże Ci poprawić wydajność.

Kiedy GA4 jest odpowiedzią?

Poniżej podsumowałem funkcjonalności, które są nowe w stosunku do Universal Analytics, aby pomóc Ci znaleźć odpowiedzi na często zadawane pytania dotyczące zmian, jakie wprowadzono wraz z GA4:

- Jak można zintegrować dane analizy cyfrowej z GCP, aby działały poza usługami GA4 (o tym właśnie jest ta książka!)?
- Jak ujednoczyć śledzenie użytkowników we wszystkich cyfrowych właściwościach, również w aplikacjach mobilnych i witrynach internetowych?
- Jak można łatwiej tworzyć niestandardowe implementacje analityczne z wartościami innymi niż domyślne?
- W jaki sposób można uzyskać dostęp do cyfrowych danych analitycznych, aby przesyłać je do modelu uczenia maszynowego?
- Jak możemy szanować wybory dotyczące prywatności, ale nadal mieć pewne dane na temat wydajności prowadzonej strony internetowej?

W tym punkcie omówiłem, dlaczego warto korzystać z GA4 i jakie są kluczowe różnice tej usługi w porównaniu z Universal Analytics. Podstawowym źródłem tych zmian jest sposób, w jaki GA4 rejestruje swoje dane w nowym modelu danych, który opiszę szerzej w następnym podrozdziale.

Model danych GA4

To właśnie model danych odróżnia GA4 od Universal Analytics. Ten nowy model danych umożliwia GA4 oferowanie bardziej zaawansowanych funkcjonalności. W tym podrozdziale przyjrzymy się dokładniej modelowi danych i jego działaniu.

Oto kluczowe elementy tego modelu danych:

Prostota

Wszystko jest zdarzeniem tego samego typu. Na dane nie są narzucane żadne arbitralne relacje.

Prędkość

Biorąc pod uwagę prostszy model danych — zredukowane przetwarzanie zdarzeń pozwala robić wszystko w czasie rzeczywistym.

Elastyczność

Zdarzenia mogą mieć dowolną nazwę, którą ogranicza tylko dozwolona liczba znaków (domyślnie 500). Do każdego zdarzenia można dołączać parametry, aby dostosowywać jego metadane.

Teraz zagłębimy się w szczegóły i zbadamy składnię tworzenia zdarzeń GA4.

Zdarzenia

W GA4 zdarzenia są niepodzielną jednostką przechwytywania danych. Każda czynność wykonywana przez użytkownika na danej stronie internetowej powoduje wysłanie zdarzenia do serwerów Google'a zgodnie z Twoją konfiguracją.

Oto pojedyncze zdarzenie:

```
{"events": [{"name": "book_start"}]}
```

Już samo zliczanie zdarzeń "book_start" dostarcza przydatnych informacji, takich jak liczba osób, które zaczęły czytać książkę, czy średnia liczba przeczytanych książek dziennie.

Aby zapewnić powiązanie określonej kolekcji zdarzeń z jednym użytkownikiem, trzeba nadać im wspólny identyfikator. W GA4 oznacza to wysłanie również identyfikatora `client_id`, który jest pseudonimem znajdującym się zwykle w pliku *cookie* GA4. Z reguły na identyfikator składa się liczba losowa z dołączonym znacznikiem czasu, który określa moment utworzenia zdarzenia:

```
{"client_id": "1234567.1632724800", "events": [{"name": "book_start"}]}
```

Powyższa linia to minimalne wymagane dane dla zdarzeń wysyłanych na Twoje konto GA4.



Znaczniki czasu są zwykle podawane w formacie czasu uniksowego (epoki Uniksa), czyli jako liczba sekund, które upłynęły od północy 1 stycznia 1970. Oznacza to, że pliki *cookie* z przykładowym znacznikiem 1632724800 będą wskazywały poniedziałek, 27 września 2021, 08:39:56 czasu środkowoeuropejskiego letniego — właśnie w tym momencie zacząłem pisać to zdanie.

Zaprezentowane przykłady zostały utworzone z wykorzystaniem protokołu pomiarowego v2 (ang. *Measurement Protocol*), który jest jednym ze sposobów wysyłania zdarzeń. O wiele bardziej powszechną metodą jest budowanie i tworzenie tych zdarzeń za pomocą skryptów śledzenia GA4 na stronie internetowej lub w aplikacji mobilnej (iOS-a lub Androida). Warto jednak wiedzieć, co robi taki skrypt.

To samo zdarzenie wysłane z trackera internetowego z użyciem funkcji `gtag()` będzie wyglądało tak:

```
gtag('event', 'book_start')
```

Dostarczaniem identyfikatorów `client_id` przez pliki *cookie* zajmuje się biblioteka JavaScriptu dla GA4, wystarczy więc podać niestandardową nazwę zdarzenia.

Gdy korzystasz ze skryptów śledzenia GA4, ta biblioteka pomaga Ci uniknąć konfigurowania powszechnych typów zdarzeń, ponieważ zapewnia automatycznie zbierane zdarzenia (<https://oreil.ly/fe6V8>). Obejmuje to takie przydatne zdarzenia jak odsłony stron, wyświetlenia filmów, kliknięcia, pobieranie plików i przewijanie stron. Samo w sobie stanowi to już przewagę nad Universal Analytics: w GA4

standardem jest to, co wcześniej trzeba było konfigurować. Mniej konfiguracji oznacza szybsze implementacje i mniejsze ryzyko popełniania błędów. Aby użyć tych automatycznych zdarzeń, możesz za pomocą ulepszonych ustawień pomiaru wybrać, które z nich włączyć (<https://oreil.ly/NHRpH>).

Istnieją również **zdarzenia zalecane**. Są to zdarzenia implementowane, ale zgodne z zalecaną strukturą nazewnictwa Google'a. Są one bardziej dostosowane do określonej witryny internetowej i zawierają rekomendacje dotyczące branż, takich jak podróże, e-commerce lub strony internetowe z ofertami pracy. Warto się ich trzymać również dlatego, że przyszłe raporty mogą na podstawie tych konwencji nazewnictwa ujawniać nowe funkcjonalności. Do ogólnych zdarzeń rekomendowanych (<https://oreil.ly/JZo7Q>) należą m.in. logowanie użytkownika, dokonanie zakupu i udostępnienie treści.

Ponieważ te automatyczne i zalecane zdarzenia są znormalizowane, jeśli zbierasz własne niestandardowe zdarzenia, upewnij się, że nie powielasz ich nazw, aby uniknąć kolizji i nieporozumień. Mam nadzieję, że dostrzegasz elastyczność tego systemu pod względem próby zapewnienia standaryzacji z rozsądnymi ustawieniami domyślnymi, dzięki którym nie trzeba wymyślać koła na nowo dla każdej implementacji.

Parametry niestandardowe

Samo liczenie zdarzeń nie wystarcza jednak do utworzenia użytecznego systemu analitycznego. Każde zdarzenie może mieć wiele parametrów (lub nie mieć żadnego), które zapewniają dodatkowe informacje na jego temat.

Zdarzenie logowania daje np. liczbę logowań na stronie, ale prawdopodobnie wymagane będzie podzielenie tego na sposób logowania użytkownika: za pomocą e-maila lub portalu społecznościowego. W takim przypadku do określenia metody logowania dla rekomendowanego zdarzenia login sugerowany jest również parametr `method`:

```
gtag('event', 'login', {
  'method': 'Google'
})
```

Przy zastosowaniu podstawowego protokołu pomiarowego będzie to miało taką postać:

```
{
  "client_id": "a-client-id",
  "events": [
    { "name": "login",
      "params": {
        "method": "Google"
      }
    }
  ]
}
```

Zwróć uwagę, że dodałem tablicę `params` z dodatkowymi informacjami.

Elementy e-commerce

Szczególną klasą parametrów niestandardowych jest `items`, która stanowi kolejną tablicę zagnieżdżoną w tych parametrach i zawierającą wszystkie informacje o danym elemencie. E-commerce (handel elektroniczny) reprezentuje zazwyczaj najbardziej skomplikowane strumienie danych, ponieważ ze sprzedają jest powiązanych wiele elementów, działań i danych.

Zasady są jednak w dużej mierze takie same: w tym przypadku parametrem niestandardowym jest tablica zawierająca niektóre rekomendowane pola, takie jak `item_id`, `price` i `item_brand`:

```
{
  "items": [
    {
      "item_id": "SKU_12345",
      "item_name": "jeggings",
      "coupon": "SUMMER_FUN",
      "discount": 2.22,
      "affiliation": "Google Store",
      "item_brand": "Gucci",
      "item_category": "pants",
      "item_variant": "Black",
      "price": 9.99,
      "currency": "USD"
    }
  ]
}
```

W połączeniu z rekomendowanymi zdarzeniami e-commerce, takimi jak zakup i kilka innych parametrów, pełny ładunek zdarzenia będzie wyglądał tak:

```
{
  "client_id": "a-client-id",
  "events": [{
    "name": "purchase",
    "params": {
      "affiliation": "Google Store",
      "coupon": "SUMMER_FUN",
      "currency": "USD",
      "items": [{
        "item_id": "SKU_12345",
        "item_name": "jeggings",
        "coupon": "SUMMER_FUN",
        "discount": 2.22,
        "affiliation": "Google Store",
        "item_brand": "Gucci",
        "item_category": "pants",
        "item_variant": "Black",
        "price": 9.99,
        "currency": "USD",
        "quantity": 1
      }],
      "transaction_id": "T_12345",
      "shipping": 3.33,
      "value": 12.21,
      "tax": 1.11
    }
  ]
}
```

Chociaż powyższy kod reprezentuje jedno z najbardziej złożonych zdarzeń wysłanych do GA4, mam nadzieję, że docenisz prostotę bazowego modelu. Używając wyłącznie zdarzeń i parametrów, można skonfigurować GA4 do przechwytywania złożonych interakcji, które mają miejsce na stronie internetowej.

Właściwości użytkownika

Oprócz danych na poziomie zdarzeń możliwe jest także skonfigurowanie danych na poziomie użytkownika (<https://oreil.ly/hrmQv>). Są to dane powiązane z rejestrowanym `client_id` (identyfikatorem klienta) lub `user_id` (identyfikatorem użytkownika). Można ich użyć do ustawienia segmentu klienta lub preferencji językowych.



Pamiętaj, aby szanować wybory dotyczące prywatności użytkowników. Jeśli dodasz informacje do konkretnego użytkownika, przepisy takie jak RODO wymagają uzyskania zgody użytkownika na gromadzenie jego danych w określonym celu.

Przesyłanie właściwości użytkownika jest bardzo podobne do przesyłania zdarzeń, ale oprócz wszelkich zdarzeń, które chcesz wysłać, używasz pola `user_properties`:

```
{
  "client_id": "a-client-id",
  "user_properties": {
    "user_type": {
      "value": "bookworm"
    }
  },
  "events": [
    {
      "name": "book_start",
      "params": {
        "title": "Learning Google Analytics"
      }
    }
  ]
}
```

Oto zastosowanie funkcji `gtag()`:

```
gtag('set', 'user_properties', {
  'user_type': 'bookworm'
});

gtag('event', 'book_start', {
  'title': 'Learning Google Analytics'
});
```

W tym podrozdziale przyjrzeliśmy się różnym sposobom wysyłania zdarzeń GA4, takim jak protokół pomiarowy i funkcja `gtag()`, oraz składni wysyłania zdarzeń z parametrami i właściwościami użytkownika. Przejdziemy teraz do przetwarzania tych zdarzeń pochodzących z GA4 przez integrację z GCP.

Google Cloud Platform

GCP można teraz mocno osadzić w systemie GA4 za pośrednictwem istniejących systemów analizy danych. Platforma oferuje wysoce skalowalne usługi w czasie rzeczywistym i z funkcjonalnością uczenia maszynowego, za które płacisz tylko wtedy, gdy z nich korzystasz. Pozwala jednocześnie uniknąć konieczności wykonywania nudnych prac związanych z konserwacją, bezpieczeństwem i aktualizacjami. Twoja firma może skupić się na tym, w czym jest ekspertem, a chmura zajmuje się zadaniami, które nie są związane z podstawową działalnością. Dzięki modelowi płatności *pay as you go* chmury małe zespoły mogą tworzyć usługi, które wcześniej wymagałyby znacznie większej liczby pracowników i zasobów IT.

W tym podrozdziale przyjrzymy się usługom GCP, z których najprawdopodobniej skorzystasz podczas integracji z GA4. Poznasz umiejętności i role, których Twój zespół będzie potrzebował, aby skorzystać z tych narzędzi, i dowiesz się, jak zacząć, jak zarządzać kosztami i jak wybrać odpowiednią dla siebie usługę w chmurze.

Odpowiednie usługi GCP

W tej książce skupiam się przede wszystkim na usługach aplikacji danych GCP, ale jest to szeroki wachlarz usług, które są stale aktualizowane. Pełen przegląd usług znajdziesz w książce *Data Science on the Google Cloud Platform* Valliappy Lakshmanana (O'Reilly).

Opisane poniżej kluczowe usługi w chmurze są wykorzystywane w przypadkach użycia, które omówię dalej w książce, i są niezbędne w mojej ogólnej pracy. Dostępnych jest wiele różnych usług w chmurze, a wybór odpowiedniej może być wręcz oszałamiający, gdy dopiero zaczynasz z nich korzystać. Proponuję przyrzeć się wyróżnionym tutaj usługom jako przydatnym na początek.

Poznasz następujące usługi (w przybliżonej kolejności przydatności):

BigQuery

Jak już wspomniałem, usługa BigQuery będzie odgrywała dużą rolę zarówno jako miejsce docelowe, jak i źródło analiz oraz obciążeń roboczych związanych z danymi. Oferuje nawet możliwości modelowania za pomocą BigQuery ML.

Cloud Functions

Cloud Functions jest spoiwem dla usług i umożliwia uruchamianie w środowisku bezserwowym małych fragmentów kodu, takich jak kod Pythona.

Pub/Sub

Pub/Sub to system kolejkowania komunikatów, który gwarantuje, że każdy komunikat zostanie dostarczony co najmniej raz. Zapewnia skalowanie, które może obsłużyć wysyłanie przez jego kolejkę całego internetu.

Cloud Build

Cloud Build to narzędzie do ciągłej integracji i ciągłego dostarczania (ang. *continuous integration/continuous delivery* — CI/CD), które pozwala uruchamiać wsadowe kontenery Dockera

w odpowiedzi na aktualizacje wysyłane z GitHuba. To ukryty koń pociągowy kilku moich rozwiązań.

Cloud Composer/Airflow

Cloud Composer/Airflow to orkiestrator pozwalający tworzyć skomplikowane współzależne przepływy danych z uwzględnieniem planowania.

Dataflow

Dataflow to rozwiązanie wsadowe i strumieniowe dla danych w czasie rzeczywistym, które jest dobrze zintegrowane z wieloma usługami GCP.

Cloud Run

Usługa Cloud Run jest podobna do Cloud Functions, ale umożliwia uruchamianie kontenerów Dockera zawierających dowolny kod.

Zwykle to, czego potrzebujesz, możesz utworzyć na kilka sposobów, a różnice mogą być subtelne. Zalecam jednak, abyś był(a) pragmatyczny(-na) i najpierw zbudował(a) coś działającego, a potem to zoptymalizował(a) i wybrał(a) tę usługę, która może się lepiej sprawdzić przy późniejszym uruchamianiu. Codzienny import danych możesz mieć uruchomiony np. na zaplanowanym zapytaniu BigQuery, ale w miarę jak Twoje potrzeby będą stawały się coraz bardziej złożone, lepszym narzędziem do koordynowania importu może okazać się Cloud Composer.

Nie są to jednak narzędzia typu „wskaz i kliknij”. W celu osiągnięcia zamierzonego rezultatu niezbędne jest kodowanie, dlatego w następnym punkcie omówię umiejętności, jakich potrzebujesz, aby wykorzystać możliwości tych narzędzi.

Umiejętności kodowania

Jednym z najbardziej zniechęcających aspektów stosowania tych integracji może być to, że wymagają one umiejętności, które — jak mogłoby się wydawać — mają tylko programiści komputerowi. Być może uważasz się za osobę „nietechniczną”.

Kiedyś myślałem tak samo. Pamiętam, że na początku swojej kariery stwierdziłem, że „nie znam JavaScriptu” i czekałem sześć tygodni, aż programista będzie wolny i wprowadzi na stronie internetowej pięcioliniowy kod. Kiedy w końcu znalazłem czas i nabrałem ochoty, zacząłem sam próbować, przy czym popełniałem po drodze wiele błędów. Dowiedziałem się ponadto, że profesjonaliści również robią wiele błędów, a jedyną różnicą jest to, że mają motywację, aby się nie poddawać. Zdałem sobie także sprawę, że wiele rzeczy, które robiłem w Excelu, było w rzeczywistości bardziej skomplikowanych i pracochłonnych, niż gdybym używał narzędzia bardziej odpowiedniego do tego zadania. Rozwiązywanie zadania w Excelu wymagało większego wysiłku intelektualnego niż zrobienie tego np. w R.

Jeśli więc masz do tego upodobanie, zachęcam, byś spróbował(a). Jeżeli jest to trudne, to niekoniecznie dlatego, że nie masz talentu — te rzeczy na początku są każdemu obce. W niektórych przypadkach kodowanie może się wydawać czynnością niezwykle kapryśną, gdzie wystarczy przegapić jeden średnik, aby coś poszło źle. Jednak gdy nauczysz się tej jednej rzeczy, następna będzie już nieco łatwiejsza. Najpierw byłem zaawansowanym użytkownikiem Excela. Potem

nauczyłem się Pythona i JavaScriptu, zakochałem się w R, musiałem nauczyć się doceniać SQL i bash, a teraz bawię się językiem Go. Natura programowania polega na tym, że gdy się uczysz i doskonalisz swoje umiejętności, Twoje wcześniejsze kody wydają Ci się okropne. Jest to naturalne; ważne, aby móc spojrzeć wstecz i zobaczyć postęp. Kiedy uda Ci się zbudować coś działającego, jest to doświadczenie, które powoli gromadzisz, aż 10 lat później piszesz o tym książkę.

Dla mnie sposobem na szlifowanie umiejętności było również open source, ponieważ udostępnianie kodu i uzyskiwanie informacji zwielokrotniało wszelkie moje doświadczenia związane z uruchamianiem danego kodu. Dlatego jestem tak wdzięczny za każdą informację zwrotną, którą dziś otrzymuję na GitHubie lub inną drogą. Kod przedstawiony w tej książce jest dostępny także w repozytorium GitHuba, które będę się starał aktualizować i oczyszczać z błędów.



Jeśli po przeczytaniu jakichś fragmentów mojego kodu będziesz mieć przemyślenia i pomysły, jak można zrobić to lepiej, skontaktuj się ze mną! Ciągłe się uczę.

Przypadki użycia opisane w tej książce obejmują przykłady kodów dla następujących języków:

JavaScript

Jest to język niezbędny do śledzenia stron internetowych z wykorzystaniem HTML-a i jest najczęściej używany do przechwytywania danych za pomocą tagów. Jest również szeroko stosowany w menedżerze GTM do tworzenia niestandardowych szablonów.

Python

Bardzo popularny język, obsługiwany przez wiele platform. Python jest przydatny, ponieważ można go uznać za drugi na liście najlepszych języków do uniwersalnych zastosowań. Jest również szeroko stosowany w uczeniu maszynowym, chociaż prawdopodobnie nie będziesz tego potrzebować, chyba że pracujesz nad zaawansowanymi wdrożeniami.

R

Chociaż można radzić sobie wyłącznie z Pythonem, społeczność R w zakresie nauki o danych sprawia, że w mojej opinii jest to najlepszy język dla takich zastosowań. Jego biblioteki i społeczność open source umożliwiają wszystko, od pobierania danych aż po aktywację danych za pośrednictwem interaktywnych dashboardów i raportów. Moje podejście do przepływów danych jest w dużej mierze schedą po R, więc wpływa on na moje projekty, nawet jeśli nie używam bezpośrednio tego języka.

bash

Podczas interakcji z serwerami w chmurze, które najprawdopodobniej będą korzystać z systemów opartych na Linuksie, takich jak Ubuntu lub Debian, będziesz musiał(a) używać powłoki bash, a nie interfejsu graficznego, takiego jak w serwerach windowsowych. Znajomość programowania w wierszu poleceń bash jest przydatna także podczas pracy z bardzo dużymi plikami, których nie da się łatwo importować do innych języków. Pewnej wiedzy na temat skryptów powłoki, z których najbardziej popularną jest bash, wymagają również narzędzie gcloud i inne CLI.

SQL

W większości przypadków nieprzetworzone dane, z którymi będziesz pracować, będą znajdować się w bazie danych, a SQL to najlepsza metoda ich wyodrębniania. SQL wprowadza również pomocny sposób myślenia o obiektach danych.

Chociaż możliwe jest kopiowanie i wklejanie swojej drogi do zwycięstwa, naprawdę polecam analizowanie kodu linia po linii i przynajmniej zrozumienie, co robi każda jego sekcja.

Przyjąłem założenie, że Ty lub Twój zespół macie pewne umiejętności kodowania, przejdziemy więc teraz do tego, jak zacząć korzystać z GCP i wdrożyć pierwszy kod w chmurze.

Wprowadzenie do GCP

Platforma GCP jest głównym elementem działalności firmy Google i ma przepływy całkowicie odrębne od Google Analytics, po których musisz nauczyć się nawigować.

Możesz zacząć od wykorzystania bezpłatnego, ale powinieneś (powinnaś) wiedzieć, że w przypadku czegoś poważnego będziesz musiał(a) dodać kartę płatniczą w celu korzystania z chmury. Możesz jednak przez kilka miesięcy korzystać z możliwości użytkowania poprzez dostępne kupyony dla początkujących użytkowników.

Przez pierwsze logowanie poprowadzi Cię strona startowa Google'a (<https://oreil.ly/9e6Hn>).



Nawet jeśli masz już jakiś projekt Google Cloud, warto utworzyć nowy dla przykładów zawartych w tej książce, aby mieć pewność, że będą aktywowane za pomocą najnowszych wersji interfejsów API. Najprawdopodobniej będziesz musiał(a) aktywować m.in. Google Analytics Reporting API, Google Analytics Admin API i Cloud Build API i sprawdzić, czy interfejs BigQuery API jest domyślnie aktywny.

Koszty chmury

Chmura oferuje nieograniczone możliwości, ale wiąże się z kosztami. Dostępnych jest wiele bezpłatnych limitów na usługi w chmurze, ale musisz być świadomy(-ma), że należy mieć oko na koszty, ponieważ mogą one szybko się sumować. Spotykałem się z przypadkami, gdy codzienne zaplanowane zapytanie SQL-a BigQuery wykorzystywało znacznie więcej danych niż oczekiwano, a użytkownik pojechał akurat na urlop. Kiedy wrócił kilka tygodni później, okazało się, że wykonywanie zadania kosztowało tysiące dolarów! Jeszcze gorszym scenariuszem byłoby przypadkowe opublikowanie poufnych kluczy uwierzytelniania. Co najmniej trzy razy spotkałem się z tym, że klucze zostały pobrane przez boty, aby uruchomić maszyny wydobywania bitcoinów z intensywnym wykorzystaniem GPU, z których każda generuje koszty liczone w tysiącach dolarów.

Chociaż bezpłatne poziomy usługi są zwykle odpowiednie do eksperymentowania, a modele cenowe są z reguły bardzo korzystne, warto najpierw sięgnąć po kalkulator cen GCP (<https://oreil.ly/XOWeS>) lub uruchomić ograniczone wersje swoich aplikacji, aby ocenić koszty działania środowiska produkcyjnego. Koszty usług mogą w dużym stopniu wpłynąć na to, z której aplikacji w chmurze należy korzystać.

Powinieneś (powinnaś) także proaktywnie konfigurować alerty rozliczeniowe i chronić klucze uwierzytelniające.

Jednak mimo tych wszystkich moich ostrzeżeń firmy były generalnie zaskoczone, jak małe są koszty chmury w stosunku do jej wartości. Firmy przechowujące swoje dane w BigQuery zazwyczaj mają na początku rachunki mniejsze niż 100 dolarów miesięcznie, a koszty rosą tylko wtedy, gdy utworzą dobry przypadek użycia, który zwiększa wartość wielu czynników. Zaokrągliłem w górę do 100 dolarów, ale w rzeczywistości jest bardziej prawdopodobne, że będzie to około 5 dolarów, dopóki nie będziesz mieć gotowych swoich aktywnych przypadków użycia. Klientom podaję jednak z reguły tę pierwszą kwotę, są więc mile zaskoczeni, jeśli koszty okazują się mniejsze!

Czynniki wpływające na koszty to ilość przesyłanych danych, czas obliczeniowy i stopień przybliżenia działania aplikacji do czasu rzeczywistego. Oszczędności kosztów chmury są zwykle spowodowane rozliczaniem tylko po wykonaniu rzeczywistej pracy zamiast płać stałą opłatę za usługi. Jednak duży wpływ mają na to również usługi, z jakich korzystasz. Ponieważ najczęściej istnieje wiele sposobów rozwiązania konkretnego problemu, zazwyczaj istnieje jakaś metoda zrobienia tego w środowisku lokalnym i tańszy sposób korzystania z chmurowych technologii bezserwerowych, które omówię w następnym punkcie.

Poruszanie się w górę piramidy bezserwerowej

Prawdziwe odblokowanie mocy chmury wiąże się z ewolucją myślenia o tym, jak radzić sobie z problemami IT z wykorzystaniem jej mocnych stron. Pierwszym krokiem firm w chmurze zwykle jest przyjęcie modelu *lift-and-shift*, w którym po prostu replikują w chmurze to, co działało lokalnie, np. zastępują lokalną bazę danych MySQL chmurowym serwerem uruchamiającym MySQL-a. Kolejną strategią jest „przeń i ulepszaj”, co polega np. na umieszczeniu bazy danych MySQL w Google Cloud SQL, czyli zarządzanej instancji MySQL-a.

Model *lift-and-shift* przynosi jednak tylko niewielkie korzyści w porównaniu z pełnym potencjałem chmury. Aby firma mogła osiągnąć prawdziwą transformację cyfrową, musi uwzględnić wyższe metausługi zbudowane na fundamentach obliczeń i pamięci masowej, ze świadomością tego, że takie postępowanie będzie musiało wiązać ją nieco bardziej z usługą danego dostawcy chmury.

Dostawcy chmury zachęcają do korzystania z tych usług, zachwalając je i przekonując, że *pozbądź się* zasobów IT niezbędnych do utrzymywania, aktualizowania i rozwijania zbudowanych usług, a zamiast tego *inwestujesz* w korzystanie z aplikacji budowanych na żądanie na bazie infrastruktury chmurowej. Niewątpliwie właśnie ten model umożliwił mi napisanie tej książki, albowiem bez chmury obliczeniowej tworzenie własnych usług byłoby znacznie bardziej skomplikowane i ograniczałoby możliwość eksperymentowania z rozwiązaniami. Gdy zasoby IT są skutecznie outsourcingowane, potrzebne są znacznie mniejsze zespoły, aby osiągnąć te same rezultaty.

Przykładem tego jest BigQuery. Zbudowanie własnej usługi BigQuery wymagałoby zainwestowania w ogromne farmy serwerów pozostających w gotowości, gdy będziesz potrzebować zasobów dla „dużego zapytania”. Jednak bezczynność serwerów również generuje koszty. Gdy dla tego samego zapytania korzystasz z usługi BigQuery, zasoby te są kupowane online w razie potrzeby i w efekcie płacisz tylko za czas ich działania.

Pomoże to zilustrować schemat piramidy bezserwerowej, który pokazałem na rysunku 1.1. Przedstawia on niektóre usługi oraz kompromisy związane z wyborem usługi niezbędnej do uruchomienia własnego przypadku użycia.



Rysunek 1.1. Hierarchia piramidy GCP

Na dolnym poziomie są maszyny wirtualne i pamięć masowa, będące w zasadzie chmurowymi wersjami komputerów działających lokalnie. Jeśli chcesz mieć pełną kontrolę nad konfiguracją, możesz je odpalić z pewnymi zaletami chmury, takimi jak kopie zapasowe, bezpieczeństwo i aktualizacje. Ta warstwa jest czasami określana jako **infrastruktura jako usługa** (ang. *Infrastructure as a Service* — IaaS).

Na następnym poziomie masz usługi, które uruchamiają maszyny wirtualne i pamięć masową, ale ukrywają szczegóły ich działania, musisz więc przejmować się tylko niezbędnymi konfiguracjami. Przykładem jest usługa App Engine, a ta warstwa jest określana również mianem **platformy jako usługi** (ang. *Platform as a Service* — PaaS).

Powyżej jest kolejny poziom abstrakcji działający na bazie ekwiwalentu PaaS. Usługi te są zwykle bardziej oparte na rolach, więc dostępne są usługi takie jak hurtownie danych analitycznych (BigQuery). Nazywamy to **bazą danych jako usługą** (ang. *DataBase as a Service* — DBaaS).

A jeszcze wyżej możesz mieć usługi, które eliminują część konfiguracji, aby zapewnić jeszcze większą wygodę. Często wystarczy podać jedynie kod, który trzeba uruchomić, lub dane, które mają zostać przekształcone. Przykładem jest usługa Cloud Functions: nie musisz wiedzieć, w jaki sposób funkcja wykonuje swój kod — musisz jedynie określić, w jaki sposób chcesz go uruchomić. Określamy to mianem **funkcji jako usługi** (ang. *Functions as a Service* — FaaS).

Mając to na uwadze, możesz ocenić, gdzie powinna zostać uruchomiona Twoja aplikacja. Usługi w górnej części piramidy zazwyczaj mają większy koszt uruchomienia, ale jeśli musisz zmieścić się w określonych kosztach wdrożenia, nadal stanowią ogromne oszczędności. Ponieważ musisz posiadać lub skalować więcej infrastruktury, możesz rozważyć przesunięcie się w dół piramidy, aby mieć większą kontrolę.

Dla przypadków użycia przedstawionych w tej książce przyjąłem założenie, że mają znajdować się one na jak najwyższym poziomie piramidy. Usługi te są zwykle najnowszymi osiągnięciami i szybko można rozpocząć z nich korzystać, a ponadto zapewniają skalę, która pozwoli Ci obsłużyć pierwszy miliard użytkowników.

I naprawdę jest to teraz w zasięgu ręki — przy wyborze usługi należy wziąć pod uwagę intensywność jej wykorzystywania, co może obejmować globalną skalę Google'a. Być może jeszcze tego nie potrzebujesz, ale warto to rozważyć na wypadek, gdybyś musiał(a) przeprojektować aplikację, gdyby okazała się nadspodziewanie udana.

W tym miejscu pomocne jest bycie na samej górze piramidy (jak pokazałem na rysunku 1.1), ponieważ usługi te zwykle mają autoskalowalne alokowanie zasobów. Należy je ograniczyć, aby uniknąć kosztownych błędów, ale zasadniczo, jeśli masz pieniądze, powinieneś (powinnaś) spodziewać się podobnej wydajności zarówno dla tysiąca użytkowników, jak i dla miliarda. Idąc w dół hierarchii, nadal masz pewne opcje, ale musisz być bardziej zaangażowany(-na) w konfigurację dotyczącą tego, kiedy i gdzie należy zastosować daną skalę.

Podsumowanie wprowadzenia do GCP

To było błyskawiczne streszczenie powodów, dla których chmura jest tak potężna, i omówienie sposobów zastosowania jej mocy do implementacji GA4. Pokazałem, jak chmura daje Ci do dyspozycji zasoby, które zaledwie kilka lat temu wymagałyby dużego zespołu IT. Wspomniałem także o koncepcjach modeli bezserwerowych w porównaniu z modelami *lift-and-shift*, abyś wiedział(a), jak się z tym obchodzić. Będzie to wymagało rozszerzenia Twoich ról cyfrowych o języki kodowania, które pomogą włączyć takie usługi i zapewnić, że inwestowanie w te umiejętności uczyni z Ciebie bardziej skutecznego marketingowca cyfrowego. Większość tej książki dotyczy tego, jak zastosować to w praktyce, i znajdziesz w niej przykładowe przypadki użycia dla rzeczy, które możesz zrobić od razu.

Wprowadzenie do przypadków użycia

W tej książce wprowadzę wszystkie koncepcje i technologie istotne dla integracji z GA4, ale teoria i planowanie to nie wszystko. Umiejętności, które omówię w tej książce, rozwinąłem tak naprawdę podczas implementowania aplikacji. Po drodze popełniałem błędy, ale były one często najcenniejszymi doświadczeniami edukacyjnymi, ponieważ po debugowaniu, dlaczego coś poszło nie tak, zaczynasz lepiej rozumieć, jak zrobić to dobrze.

Aby pomóc Ci od razu rozpocząć Twoją podróż, w kilku kolejnych rozdziałach wprowadzę wszystkie elementy składowe niezbędne do budowania własnych aplikacji. Rozdziały 7., 8. i 9. są natomiast poświęcone technicznym przypadkom użycia opisującym cały cykl życia aplikacji danych GA4 i zawierającym przykłady kodu. Omówię w nich budowanie uzasadnień biznesowych, wymagania techniczne i wybór wykorzystywanych technologii. Jeśli będziesz na bieżąco pracował(a) ze wszystkimi przykładami, pod koniec książki powinieneś (powinnaś) mieć działającą integrację.



W praktyce możesz przypadkowo pominąć niektóre kroki, więc będziesz musiał(a) wrócić i uważnie przeczytać to, co przegapiłeś(-łaś). Ponadto zanim wdrożysz konkretny przypadek użycia, może się okazać, że technologie nieznacznie się zmieniły, i będziesz musiał(a) go zaktualizować.

Jest mało prawdopodobne, aby nawet doskonale zaimplementowany przykład pasował dokładnie do potrzeb Twojej firmy lub tego, czemu powinien nadać priorytet. Przypadki użycia obejmują moje doświadczenia z typowymi problemami klientów, ale Twój własny będzie niewątpliwie nieco inny. Ponieważ najprawdopodobniej będziesz musiał(a) dostosować przypadki użycia do własnych potrzeb, ważne jest, by zrozumieć nie tylko to, co robić, ale także dlaczego trzeba robić to w taki sposób, a nie inny. Będziesz wtedy mógł (mogła) dostosować proces, by lepiej pasował do Twoich priorytetów.

Pomimo indywidualnych wymagań niektóre typowe tematy można ze sobą powiązać w odniesieniu do sposobu podejścia do tych projektów. W rozdziale 2. omówię framework, który łączy wszystkie udane projekty integracji danych, nad którymi pracowałem. Przypadki użycia będą zgodne z tym frameworkiem, abyś nabrał(a) praktyki w jego stosowaniu. Cztery główne obszary to pozyskiwanie, przechowywanie, modelowanie i aktywacja danych. Głównym motorem tego wszystkiego jest jednak pytanie stawiane przed nami przez konkretny przypadek użycia, ponieważ jeśli próbujemy znaleźć dla jakiegoś problemu rozwiązanie, które tak naprawdę nie pomoże naszej firmie, całe przedsięwzięcie nie będzie tak efektywne, jak można by sobie życzyć. Znalezienie odpowiedniego problemu do rozwiązania jest ważne dla Twojej firmy, dlatego w rozdziale 2. zaproponuję również kilka pytań, na które powinieneś (powinnaś) sobie odpowiedzieć, aby ten problem zdefiniować.

Przypadki użycia pozwolą Ci skoncentrować się wyłącznie na praktycznej pracy związanej z implementacją. Najlepszym sposobem nauki będzie śledzenie przykładów oraz ich implementowanie, a nie tylko samo czytanie. Przykłady te mogą również stanowić pewien punkt odniesienia, gdy będziesz implementować własne przypadki użycia, ponieważ często możesz ponownie wykorzystać aspekty jednego rozwiązania w drugim. We wszystkich przypadkach użycia opisanych w tej książce GA4 stanowi źródło pozyskiwania danych. Staram się ponadto wykorzystywać w nich kilka różnych technologii, aby objąć szeroki zakres zastosowań.

Przypadek użycia — predykcyjna analiza zakupów

Pierwszy przypadek użycia, omówiony w rozdziale 7., jest podstawą, która pomoże Ci przyzwyczaić się do ogólnego podejścia wspólnego także dla bardziej złożonych przypadków użycia opisanych dalej w książce. Będziemy używać tylko jednej platformy — GA4. Te same zasady mają zastosowanie również do bardziej angażujących przypadków użycia, ale powinno to ponadto pokazać, że możliwe jest zastąpienie GA4 innymi aplikacjami, jeśli lepiej zaspokajają Twoje potrzeby. Ten przypadek wykorzystuje kilka nowych funkcjonalności GA4, w tym uczenie maszynowe i eksporty odbiorców.

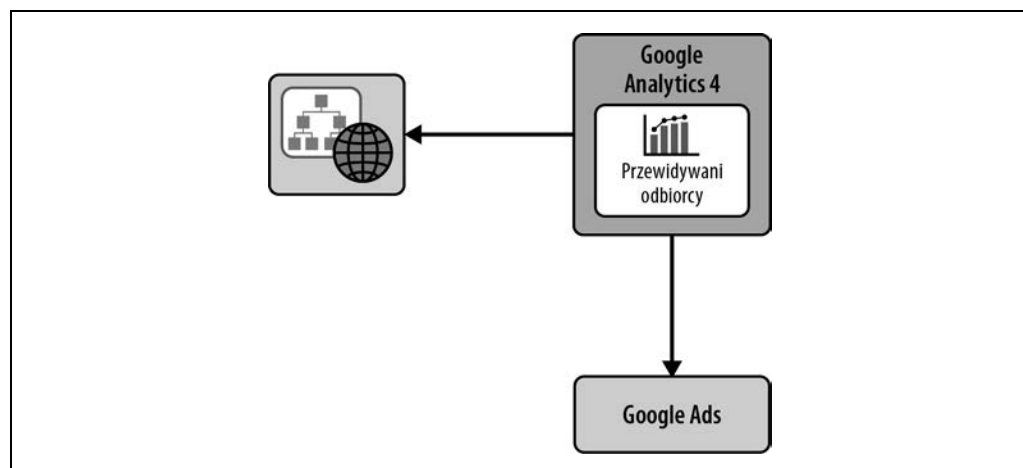
W predykcyjnej analizie zakupów stosowane jest modelowanie do przewidywania, czy użytkownik w przyszłości coś kupi. Można to wykorzystać do zmiany treści witryny lub strategii reklamowej dla określonych użytkowników. Jeśli prawdopodobieństwo, że użytkownik dokona zakupu,

przekracza 90%, być może powinniśmy np. wstrzymać marketing skierowany do tego użytkownika, gdyż zadanie zostało już wykonane. Analogicznie, jeżeli prawdopodobieństwo zakupu wynosi poniżej 30%, być może powinniśmy uznać danego użytkownika za przegraną sprawę. Ustanowienie takich polityk oznacza, że możesz zmienić alokację budżetu, aby kierować marketing tylko do 60% użytkowników, którzy mogą kupować lub nie. Powinno to ponadto obniżyć koszt zdobycia jednego nabywcy (ang. *Cost Per Acquisition* — CPA) i potencjalnie zwiększyć przychody ze sprzedaży.

Użyjemy GA4, aby wykonać następujące czynności:

- gromadzenie danych ze strony internetowej, w tym zdarzeń konwersji;
- zapisywanie wszystkich niezbędnych danych;
- zapewnienie modelowania danych przy użyciu wskaźników predykcyjnych, takich jak zakup i prawdopodobieństwo;
- eksportowanie do Google Ads w celu aktywacji przy użyciu odbiorców GA4.

Proces ten ilustruje prosty schemat architektury danych pokazany na rysunku 1.2.



Rysunek 1.2. Architektura danych dla przypadku użycia przewidywanych odbiorców

Do tego celu nie będzie wymagane żadne programowanie, a cała konfiguracja zostanie wykonana w interfejsie użytkownika.

Wskaźniki predykcyjne to funkcjonalność zintegrowana z GA4 i bezpośrednio wykorzystująca możliwości Google’a w zakresie uczenia maszynowego, aby naprawdę zmienić sposób działania Twojej firmy. Twoja witryna musi jednak spełniać określone kryteria, aby kwalifikować się do korzystania z funkcjonalności wskaźników predykcyjnych, co daje Ci mniejszą kontrolę nad tym, kiedy ta funkcjonalność może być stosowana. Nawet jeśli nie możesz używać wskaźników predykcyjnych, nadal możesz korzystać z własnych danych i samodzielnie zbudować model, a następnie skorzystać z integracji Google Ads. Omówię to w następnym punkcie.

Przypadek użycia — segmentacja odbiorców

Przypadek użycia dla segmentacji odbiorców opisany w rozdziale 8. pokazuje, jak lepiej zrozumieć zagregowane zachowanie klientów. Jakie wspólne trendy lub zachowania możesz wybrać, aby lepiej obsłużyć dany segment? Ile masz typów klientów? Czy znalezione segmenty oparte na danych odpowiadają założeniom Twojej firmy?

Takie projekty segmentacji były niegdyś wykorzystywane do personalizacji wiadomości marketingowych dla danych użytkowników. Niektórzy klienci mogą zostać np. określani jako bardziej skłonni do kupowania produktów powiązanych, możesz więc ograniczyć wiadomości marketingowe tylko do tych klientów, aby obniżyć koszty kampanii i uniknąć niepotrzebnych wiadomości kierowanych do klientów, którzy mogą się nimi denerwować.

Możesz segmentować według wielu różnych kryteriów. Skuteczną metodą poprzedzającą czasy internetu jest model RFM, który analizuje zachowania związane z czasem od ostatniego zakupu (ang. *recency*), częstotliwością zakupów (ang. *frequency*) oraz wartością zakupów (ang. *monetary*) dla użytkowników i segmentów z podobnymi wynikami w poszczególnych sektorach. Dzięki bogactwu dostępnych obecnie danych możesz tworzyć inne modele z setkami pól. Model, który wybierzesz, będzie w dużej mierze regulowany wymaganiami biznesowymi Twojego przypadku użycia, a także względami dotyczącymi prywatności. Prywatność jest tutaj ważna, ponieważ konieczne może być uzyskanie zgody użytkowników na umieszczanie ich danych w modelach. Jeśli nie uzyskasz takiej zgody, klient może się denerwować, jeśli będą do niego kierowane reklamy.

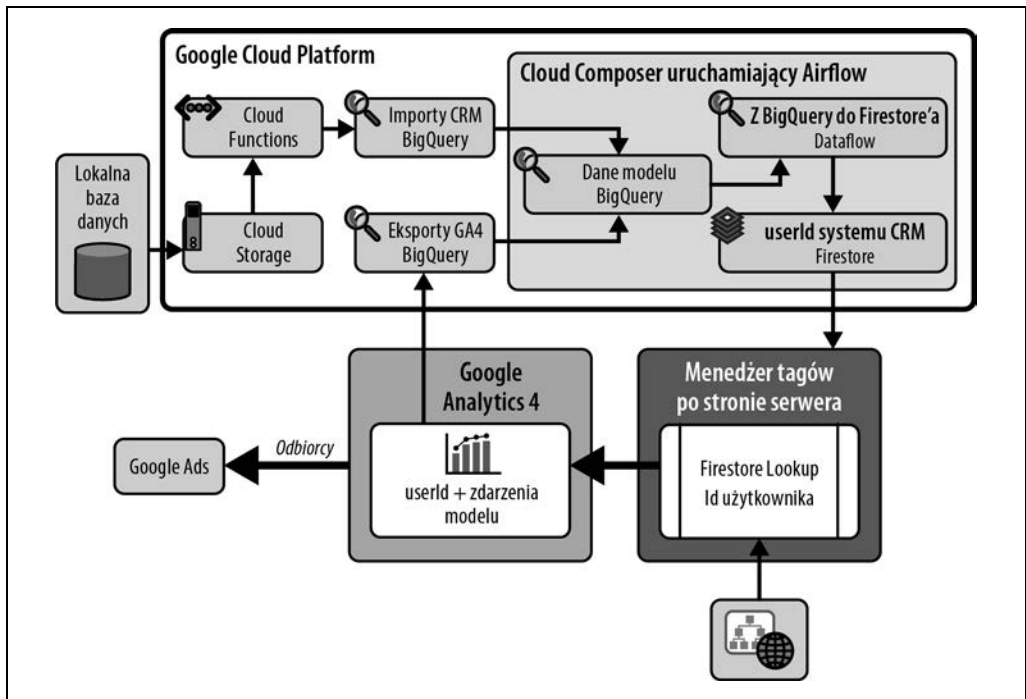
Wykorzystując ten przykład, będziemy chcieli obniżyć nasze koszty związane z Google Ads. W tym kontekście Google Ads zajmie się aktywacją danych, tam będziemy bowiem wysyłać dane, aby zmienić zachowanie użytkowników. Naszym uzasadnieniem biznesowym jest obniżenie kosztów, a także uzyskanie wyższej sprzedaży, jeśli możemy lepiej dostosować nasz przekaz marketingowy.

Chcielibyśmy wykorzystać posiadane przez nas dane dotyczące zachowania klienta w witrynie internetowej i historii jego zakupów, aby określić, czy powinniśmy wyświetlać mu określone reklamy. W tym celu użyjemy następujących usług:

- GA4 i naszej bazy danych zarządzania relacjami z klientami (ang. *Customer Relationship Management* — CRM) jako źródła danych;
- Cloud Storage i BigQuery jako pamięci masowej dla danych;
- BigQuery do tworzenia segmentów;
- Firestore do wysyłania tych segmentów do naszych użytkowników GA4 w czasie rzeczywistym;
- menedżera tagów po stronie serwera do wzbogacenia danych GA4;
- GA4 Audiences do wysyłania tych segmentów do Google Ads.

Interakcje między tymi usługami pokazałem na rysunku 1.3.

Po drodze zapewnimy również przestrzeganie wyborów dotyczących prywatności i zadamy o to, aby żadne dane osobowe nie były eksportowane ani przesyłane tam, gdzie nie są potrzebne.



Rysunek 1.3. Architektura danych dla przypadku użycia segmentacji odbiorców

Technologie, których będziemy używać do następujących usług, zostaną omówione szczegółowo w odpowiednich rozdziałach dalej w książce:

- GA4 do pomiarów wtryny internetowej;
- produkcyjna baza danych dla historii zakupów użytkowników;
- importy za pośrednictwem Cloud Storage, Pub/Sub i Cloud Functions;
- BigQuery do tworzenia modeli segmentacji;
- Cloud Composer do planowania aktualizacji;
- Cloud Storage, Pub/Sub i Cloud Functions do importowania segmentów do GA4;
- GA4 do tworzenia odbiorców.

Będziesz potrzebować umiejętności kodowania w Pythonie i SQL-u oraz wykonywania prac konfiguracyjnych w ramach GA4, konsoli Google Cloud i Google Ads. Musimy również upewnić się, że zbieramy w GA4 odpowiednie dane, abyśmy mogli połączyć aktywność internetową z danymi CRM w sposób zgodny z zasadami zachowania prywatności.

Przypadek użycia — prognozowanie w czasie rzeczywistym

Przypadek użycia opisany w rozdziale 9. dotyczy tworzenia aplikacji prognozowania w czasie rzeczywistym. Gdy firmy rozpoczynają swoją przygodę z analityką, często najpierw pytają o analitykę w czasie rzeczywistym, ale zwykle tracą zainteresowanie tą kwestią, jeżeli odkrywają, że nie

mogą reagować na ten strumień danych w czasie rzeczywistym. Jeśli jednak masz taką zdolność, jest to ekscytujący projekt, ponieważ możesz zobaczyć natychmiastowe korzyści.

Dobrym przykładem tego przypadku użycia jest zachowanie wydawcy, który reaguje na wydarzenia w czasie rzeczywistym w ciągu dnia przy wyborze historii do opublikowania lub promowania. W firmie, w której kliknięcia i wyświetlenia oznaczają przychody, trafienie w wiral z mediów społecznościowych może wyrzucić wpływ na duży biznes. Uzyskanie tego trafienia wymaga wielokrotnych prób, edycji i promocji na stronach głównych oraz utrzymywania stałych kanałów w czasie rzeczywistym dla popularnych tematów i nastrojów z mediów społecznościowych. Przypadek użycia, który opiszę tu szczegółowo, pokazuje, jak korzystać z tego strumienia danych analitycznych i na podstawie tych informacji prognozować ruch. Możesz dokonywać tych prognoz za pomocą grup odbiorców GA4, które zostały skonfigurowane w celu identyfikacji różnych segmentów Twoich klientów.

Ten przypadek użycia zademonstruje użycie Dockera do obsługi uruchamiania rozwiązania w Cloud Run dla dashboardu — pakietu Shiny języka R do tworzenia aplikacji internetowych. Kluczowym powodem korzystania z Dockera jest to, że kod działający w kontenerach możesz zastąpić kodem napisanym w dowolnym innym języku, takim jak Python, Julia lub innym przyszłym języku do nauki o danych. Oto role danych dla tego projektu:

- pozyskiwanie danych za pośrednictwem interfejsów API;
- przechowywanie danych w aplikacji;
- modelowanie danych w języku R;
- aktywacja danych za pośrednictwem dashboardu R Shiny.

Do realizacji tego przypadku użycia potrzebujemy następujących usług:

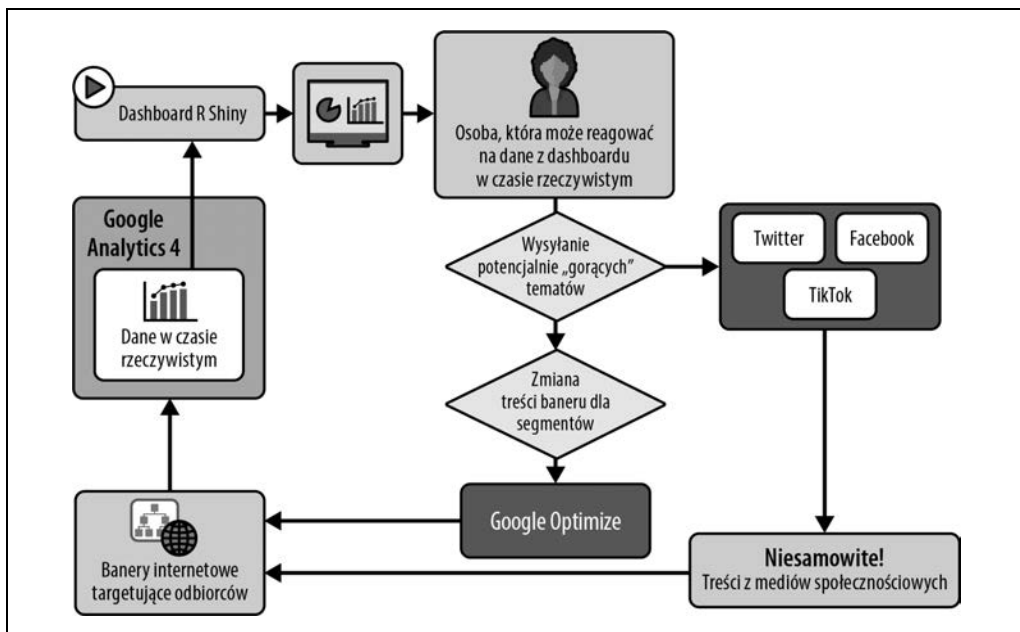
- GA4 do zbierania strumienia zdarzeń internetowych w czasie rzeczywistym;
- Cloud Run do uruchomienia dashboardu;
- GA4 Audiences, aby otrzymać przydatny segment do prognozowania.

Wzajemne relacje tych zasobów pokazałem na rysunku 1.4.

Będziemy wykorzystywać umiejętności kodowania w R do tworzenia kanałów i modelowania w czasie rzeczywistym, a ponadto wykorzystamy wizualizację do utworzenia dashboardu.

Podsumowanie

W tym rozdziale przedstawiłem główne sposoby wykorzystania GA4 w celu ulepszenia implementacji analiz cyfrowych. Wyjaśniłem, dlaczego w ogóle powstała usługa GA4 oraz czym różni się i w czym jest lepsza od Universal Analytics dzięki nowemu, prostszemu modelowi danych. Omówiłem, w jaki sposób integracja tej usługi z GCP otwiera Twoje cyfrowe analizy na zupełnie nowy świat aplikacji obejmujących usługi takie jak Firebase i BigQuery. Chociaż korzystanie z tych nowych usług w chmurze wymaga także nowych umiejętności, takich jak kodowanie, zyskujemy również



Rysunek 1.4. Dane w czasie rzeczywistym są pobierane z GA4 i tworzona jest prognoza, która pomaga w ustalaniu priorytetów treści w mediach społecznościowych i na banerach na stronie za pośrednictwem Google Optimize

większą dostępność tych technologii niż w przeszłości. Oferta architektury bezserwerowej umożliwiła wyeliminowanie wielu rutynowych prac związanych z konfigurowaniem i skalowaniem usług obliczeniowych. Ogólnym zaleceniem na początku jest dążenie do korzystania z usług na jak najwyższym poziomie tej architektury, aby utrzymać jak najniższą barierę wejścia.

Chociaż technologia jest już dostępna, sposób podejścia do niej i jak najlepsze jej wykorzystanie są kluczowymi umiejętnościami, których marketingowcy cyfrowi niezaznajomieni z chmurą mogą nie posiadać. Dlatego w rozdziale 2. przedstawię ogólny framework i strategię tworzenia udanych projektów analizy danych, które można wielokrotnie wykorzystywać. Opracujemy role dla pozyskiwania danych, ich gromadzenia, modelowania i aktywacji z perspektywy strategii, gotowe do wykorzystania w praktycznych rozdziałach poświęconych implementacji.

PROGRAM PARTNERSKI

— GRUPY HELION —



1. ZAREJESTRUJ SIĘ
2. PREZENTUJ KSIĄŻKI
3. ZBIERAJ PROWIZJĘ

Zmień swoją stronę WWW w działający bankomat!

Dowiedz się więcej i dołącz już dzisiaj!

<http://program-partnerski.helion.pl>

GRUPA
Helion 

Przygotuj się na cyfrowy marketing przyszłości!

Google Analytics 4 wyznacza nowe zasady marketingu cyfrowego: obecnie chodzi nie tyle o raportowanie zdarzeń z przeszłości, ile o aktywację danych przez łączenie danych online i offline ze wszystkich strumieni, aby zapewnić kompleksowe informacje marketingowe. Taka analityka cyfrowa umożliwi szybsze podejmowanie decyzji. Umiejętność korzystania z GA4 staje się więc nieodzowna.

Oto zwięzły i praktyczny przewodnik po usłudze GA4 i jej integracji z chmurą. Szczególnie skorzystają z niego analitycy danych, biznesu i marketingu. Opisano tu wszystkie istotne kwestie dotyczące tego nowego, potężnego modelu analitycznego. Szczególną uwagę poświęcono bardziej zaawansowanym funkcjonalnościom GA4. Zaprezentowano architekturę GA4, strategię danych, a także informacje dotyczące pozyskiwania, przechowywania i modelowania danych. W książce znalazło się również omówienie typowych przypadków użycia dla aktywacji danych i instrukcji przydatnych podczas implementacji tych przypadków. Co istotne, poszczególne zagadnienia zostały zilustrowane praktycznymi przykładami kodu.

Ta książka pomaga nie tylko zrozumieć dane, ale także wygenerować przewagę konkurencyjną!

Melinda Schiera, strateg do spraw analityki

Najciekawsze zagadnienia:

- integracja Google Cloud z GA4
- przeprowadzanie integracji GA4
- przechwytywanie danych GA4 i realizacja przypadków użycia
- projektowanie przepływów danych
- dostosowywanie przypadków użycia do potrzeb biznesowych

Mark Edmondson pracuje w analityce cyfrowej od ponad 15 lat, jest autorem kilku pakietów języka R dotyczących interfejsów API Google. Często występuje na międzynarodowych konferencjach branżowych. Zajmuje się również uczeniem maszynowym i przetwarzaniem w chmurze. Mieszka w Kopenhadze w Danii.

Helion
helion.pl
HELION SA
ul. Kościuszki 1c
44-100 Gliwice
tel. 32 230 99 63
helion@helion.pl

KOD KORZYŚCI
Sięgnij po więcej! ▶



ISBN 978-83-8322-849-5



Cena: 69,00 zł