



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA PODNIKATELSKÁ

FACULTY OF BUSINESS AND MANAGEMENT

ÚSTAV INFORMATIKY

INSTITUTE OF INFORMATICS

NÁVRH DATOVÉHO SKLADU A DATOVÉ STRUKTURY BIGQUERY PRO VYBRANÝ PODNIK

DESIGN OF DATA WAREHOUSE AND DATA STRUCTURE IN BIGQUERY FOR SELECTED BUSINESS

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Michaela Šimková

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jan Luhan, Ph.D., MSc

BRNO 2024

Zadání diplomové práce

Ústav:	Ústav informatiky
Studentka:	Bc. Michaela Šimková
Vedoucí práce:	Ing. Jan Luhan, Ph.D., MSc
Akademický rok:	2023/24
Studijní program:	Informační management

Garant studijního programu Vám v souladu se zákonem č. 111/1998 Sb., o vysokých školách ve znění pozdějších předpisů a se Studijním a zkušebním řádem VUT v Brně zadává diplomovou práci s názvem:

Návrh datového skladu a datové struktury BigQuery pro vybraný podnik

Charakteristika problematiky úkolu:

Úvod
Cíle práce, metody a postupy zpracování
Teoretická východiska práce
Analýza současného stavu
Vlastní návrhy řešení
Závěr
Seznam použité literatury
Přílohy

Cíle, kterých má být dosaženo:

Navrhnout a implementovat datový sklad v cloudovém prostředí BigQuery pro následnou analýzu prostřednictvím Google Analytics 4 se zaměřením na automatizaci zpracování dat a jejich následnou vizualizační interpretaci.

Základní literární prameny:

LABERGE, R. Datové sklady - Agilní metody a business intelligence. 1. vyd. Praha: Computer Press, 2012. 352 s. ISBN 978-80-251-3729-1.

MARYŠKA, M.; POUR, J.; STANOVSKÁ, I. a ŠEDIVÁ, Z. Self Service Business Intelligence: Jak si vytvořit vlastní analytické, plánovací a reportingové aplikace. 1. vyd. Praha: Grada Publishing, 2018. 352 s. ISBN 978-80-271-0616-5.

NOVOTNÝ, O.; POUR, J. a SLÁNSKÝ, D. Business Intelligence: Jak využít bohatství ve vašich datech. 1. vyd. Praha: Grada, 2005. 256 s. ISBN 80-247-1094-3.

NOVOTNÝ, O.; POUR, J.; MARYŠKA, M. a BASL, J. Řízení výkonnosti podnikové informatiky. 1. vyd. Praha: Professional Publishing, 2010. 275 s. ISBN 978-80-7431-040-9.

SLÁNSKÝ D. Data a analytika pro 21. století. 1. vyd. Praha: Professional Publishing, 2018. 5 svazků. ISBN 978-80-88260-25-7.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2023/24

V Brně dne 4.2.2024

L. S.

doc. Ing. Miloš Koch, CSc.
garant

doc. Ing. Vojtěch Bartoš, Ph.D.
děkan

Abstrakt

Táto diplomová práca sa zaoberá návrhom a implementáciou dátového skladu a dátovej štruktúry v cloudovom prostredí BigQuery, s cieľom analýzy dát získaných z Google Analytics 4 a následnú automatizáciu spracovania týchto dát a ich vizualizačnú interpretáciu.

Teoretická časť práce prezentuje kľúčové koncepty a princípy týkajúce sa Business Intelligence, dátových skladov, webovej analytiky a nástrojov Business Intelligence v cloudovom prostredí, poskytujúc tak potrebný teoretický základ pre praktickú časť práce.

Praktická časť sa zaoberá návrhom štruktúry dátového skladu a konkrétnym návrhom logického a konceptuálneho modelu dátového trhoviska a jeho implementáciou v BigQuery. Opisuje požiadavky firmy, extrakciu dát z primárneho systému, tvorbu SQL dotazov na získanie historických dát a vytvorenie ETL procesov pre automatizované spracovanie nových dát. Práca tiež obsahuje vizualizáciu získaných dát pomocou nástroja Looker Studio, výsledkom čoho je interaktívny report umožňujúci vizuálne zobrazenie dát pre uľahčenie analýzy.

Cieľom tejto práce je ponúknuť efektívne riešenie pre spracovanie webových a aplikačných dát vybranej firmy v prostredí BigQuery, vďaka ktorému môže firma lepšie pochopiť nadobudnuté dáta a využiť poznatky pre strategické informované rozhodovanie.

Kľúčové slová

dátový sklad, dátová štruktúra, Business Intelligence, analytika, SQL, ETL proces, BigQuery, Google Analytics 4, Looker Studio

Abstract

This master's thesis focuses on the design and implementation of a data warehouse and data structure in the BigQuery cloud environment, with the aim of analyzing data obtained from Google Analytics 4 and subsequent automation of processing these data and their visual interpretation.

The theoretical part of the work presents key concepts and principles related to Business Intelligence, data warehouses, web analytics, and Business Intelligence tools in the cloud environment, thus providing the necessary theoretical basis for the practical part of the work.

The practical part deals with the design of the data warehouse structure and a specific design of the logical and conceptual model of the data marketplace and its implementation in BigQuery. It describes the company's requirements, extraction of data from the primary system, creation of SQL queries to obtain historical data, and creation of ETL processes for automated processing of new data. The work also includes visualization of the obtained data using the Looker Studio, resulting in an interactive report allowing visual display of data for easier analysis.

The objective of this work is to provide an efficient solution for processing web and application data of the selected company using BigQuery. This will enable the company to gain a better understanding of the acquired data, thereby facilitating strategic and informed decision-making.

Keywords

data warehouse, data structure, Business Intelligence, analytics, SQL, ETL process, BigQuery, Google Analytics 4, Looker Studio

Bibliografická citácia

ŠIMKOVÁ, Michaela. Návrh datového skladu a datové struktury BigQuery pro vybraný podnik [online]. Brno, 2024 [cit. 2024-05-13]. Dostupné z: <https://www.vut.cz/studenti/zav-prace/detail/160487>. Diplomová práce. Vysoké učení technické v Brně, Fakulta podnikatelská, Ústav informatiky. Vedoucí práce Jan Luhan.

Čestné prehlásenie

Prehlasujem, že predložená diplomová práca je pôvodná a spracovala som ju samostatne. Prehlasujem, že citácia použitých prameňov je úplná, že som vo svojej práci neporušila autorské práva (v zmysle zákona č. 121/2000 Sb., o práve autorskom a o právach súvisiacich s právom autorským).

V Brne dňa 12. 5. 2024

Bc. Michaela Šimková
(autor)

Pod'akovanie

Týmto chcem poďakovať vedúcemu mojej diplomovej práce, Ing. Janovi Luhanovi, Ph.D., MSc za jeho odborné vedenie, hodnotné rady a čas, ktorý mi venoval. Jeho trpezlivosť a odborný pohľad boli pre mňa veľmi cenné počas celej tvorby tejto práce.

Taktiež by som sa chcela poďakovať svojej rodine a priateľom, ktorí ma podporovali počas tvorby tejto práce.

Obsah

ÚVOD	10
1 CIEĽ PRÁCE, METÓDY A POSTUPY SPRACOVANIA	11
2 TEORETICKÝ ZÁKLAD BUSINESS INTELLIGENCE.....	13
2.1 Definícia a význam Business Intelligence.....	13
2.2 Komponenty Business Intelligence	14
2.2.1 Zdrojové systémy	15
2.2.2 Komponenty dátovej transformácie	15
2.2.3 Databázové komponenty	16
2.2.4 Analytické komponenty	18
2.3 Metodológie tvorby BI riešenia	20
2.3.1 Metóda podľa Kimballa	20
2.3.2 Metóda podľa Inmona	22
2.4 Návrh Business Intelligence riešenia.....	24
2.4.1 Prístupy k návrhu dátového skladu	25
2.4.2 Dimenzionálne modelovanie.....	26
2.5 Webová analytika	29
2.5.1 Životný cyklus zákazníka.....	31
3 NÁSTROJE A TECHNOLOGIE BI V CLOUDOVOM PROSTREDÍ	34
3.1 Google Analytics 4	34
3.1.1 Dátový model	35
3.1.2 Parametre jednotlivých entít.....	36
3.2 BigQuery	37
3.2.1 Štruktúra	38
3.2.2 Ukladanie a nahrávanie dát	39
3.2.3 Plánované dotazy.....	41
3.3 Looker Studio	42
4 ANALÝZA SÚČASNÉHO STAVU	44
4.1 Firma.....	44
4.2 Digitálne platformy.....	45
4.2.1 Webová stránka	45

4.2.2	Aplikácia	47
4.3	Požiadavky na riešenie	50
4.4	Súčasný stav	53
5	VLASTNÝ NÁVRH RIEŠENIA	54
5.1	Návrh architektúry dátového skladu	54
5.1.1	Primárny zdroj dát	55
5.1.2	Dočasné dátové úložisko	56
5.1.3	Dátový sklad	56
5.1.4	Dátové trhovisko	57
5.1.5	Vyhodnocovací nástroj	57
5.2	Navrhovaný model dátového trhoviska	58
5.2.1	Konceptuálny model	59
5.2.2	Logický dátový model	63
6	IMPLEMENTÁCIA NÁVRHU V BIGQUERY	68
6.1	Východiskový stav	68
6.2	Postup a plnenie dátového trhoviska	70
6.2.1	Naplnenie dátových tabuliek historickými dátami	70
6.2.2	Tvorba ETL procesu a plnenie tabuliek novými dátami	79
7	REPORTOVACÍ DASHBOARD	82
7.1	Napojenie dátového zdroja	82
7.2	Tvorba dashboardu	84
	ZÁVER	87
	Zoznam použitej literatúry	90
	Zoznam použitých obrázkov	96
	Zoznam použitých tabuliek	98
	Zoznam príloh	99
	Prílohy	100

ÚVOD

V dnešnej dobe digitálnej transformácie sa čoraz viac firiem obracia k využitiu riešení Business Intelligence (BI) na podporu a optimalizáciu svojich procesov. Tieto riešenia umožňujú firmám veľkým či naopak menším spoločnostiam, zhromažďovať, spracovávať a analyzovať obrovské množstvo dát, aby získali hlbšie pochopenie svojho podnikania a lepšie rozhodovacie schopnosti.

Na prvý pohľad sa môže zdať, že dátová analýza nie je tak kľúčová, ale opak je pravdou. Správne zhromaždené, spracované a interpretované údaje môžu výrazne prispieť k lepšiemu cieleniu marketingu, úspechu produktov a celkovému rozvoju biznisu.

Táto diplomová práca sa zaoberá návrhom a implementáciou BI riešenia pre digitálne platformy, ktorými sú webová stránka a aplikácia vybranej spoločnosti. Vďaka webovej analytike sa dáta o týchto platformách, zbierajú údaje nielen z tradičných primárnych zdrojov, ale aj z nástrojov ako Google Analytics 4, ktoré poskytujú detailný pohľad na aktivitu používateľa a jeho správanie na webe či v aplikácii. Tieto informácie umožňujú manažmentu a majiteľom produktov robiť informované rozhodnutia, ktoré môžu vylepšiť užívateľský zážitok na daných platformách, optimalizovať celkový produkt alebo lepšie využiť finančné prostriedky vynaložené do marketingových aktivít a to vďaka vyhodnoteniu dát týchto platforiem.

Webová analytika, ako súčasť BI, prináša oproti tradičnej analytickej praxi svoje špecifiká, zaoberá sa zbieraním, meraním a analýzou dát z webových stránok alebo aplikácií s cieľom porozumieť a optimalizovať zákaznícku cestu.

Autor práce, ktorý má skúsenosti s dátovou analýzou pre webové a mobilné aplikácie, a uvedomuje si, že efektívne BI riešenie na báze cloudových služieb je dlhodobým prínosom pre vyhodnotenie dát najmä ecommerce charakteru. Napriek možnosti pokročilého pohľadu nástroje webovej analytiky majú časové obmedzenie v uchovávaní dát, preto je riešenie exportu surových dát do cloudovej služby veľkým prínosom a to hlavne pre dlhodobé analýzy a predikcie. Cloudové služby poskytujú flexibilnú a škálovateľnú platformu pre BI riešenia, aj v prípade možnosti napojenia iných primárnych zdrojov. Toto riešenie môže byť prispôbené potrebám akejkoľvek firmy, bez ohľadu na jej veľkosť alebo odvetvie.

1 CIEĽ PRÁCE, METÓDY A POSTUPY SPRACOVANIA

Hlavným cieľom tejto práce je navrhnuť a implementovať dátový sklad v cloudovom prostredí BigQuery pre následnú analýzu prostredníctvom Google Analytics 4 dát so zameraním na automatizáciu spracovania dát a ich následnú vizuálnu interpretáciu pre vybraný podnik.

Pre dosiahnutie hlavného cieľa boli vytvorené nasledujúce **čiasťkové ciele**.

- Zozbieranie a definovanie požiadaviek na navrhované riešenie.
- Konceptualizácia štruktúry dátového skladu a tvorba dátového modelu dátového trhoviska.
- Implementácia navrhnutého dátového trhoviska, zahrňujúca vývoj procesov pre extrakciu, transformáciu a načítanie dát (ETL).
- Integrácia riešenia do vizualizačného softvéru a príprava interaktívneho dashboardu obsahujúceho kľúčové ukazovatele výkonnosti (KPI), vedľajšie ukazovatele, grafy a ďalšie vizuálne prvky.

Pre dosiahnutie hlavného a čiastkových cieľov je táto diplomová práca rozdelená do niekoľkých kapitol, ktoré sú logicky ucelené.

Druhá kapitola je venovaná Business Intelligence a jej teoretickému základu. V podkapitolách je vysvetlený jej význam, metodológie, hlavné komponenty, ktoré poslúžia v ďalších kapitolách ako základ pre návrh štruktúry dátového skladu. Taktiež je tu opísaná webová analytika a jej význam v BI a spôsob akými sa vyhodnocuje cesta potenciálneho zákazníka. Táto teória slúži ako vodítko pre ďalšie kapitoly.

Tretia kapitola je venovaná nástrojom a technológiám v cloudovom prostredí, tie boli vybrané špecificky na základe dlhodobej stratégie podniku. Je tu popísaná štruktúra týchto nástrojov, ich dátový model a spôsob akým sú informácie ukladané. Okrem toho tu možnosť nájsť aj popis funkcií jednotlivých nástrojov.

V štvrtej kapitole diplomovej práce je analyzovaný súčasný stav webovej stránky a mobilnej aplikácie. Sú tu charakterizované jednotlivé sekcie, v ktorých prebiehajú zákaznícke interakcie a následne poskytujúce dáta pre vyhodnocovanie. Ďalej je v tejto kapitole popísaný súčasný dátový sklad a sú sformulované požiadavky na jeho ďalší vývoj, vrátane určenia najdôležitejších metrik a ukazovateľov výkonnosti (KPI), ktoré sú kritické pre efektívne

riadenie a strategické rozhodnutia. Tieto požiadavky poslúžia ako základ pre návrhovú časť diplomovej práce.

Piata a šiesta kapitola sa venuje návrhu štruktúry dátového skladu a dátového modelu dátového trhoviska a následne jeho implementácií v BigQuery. V týchto kapitolách je prezentovaný dátový model pre dátové trhovisko a jeho konceptuálny a logický model, ktorý slúži ako podklad pre jeho realizáciu. Novovytvorený dátový sklad, ktorý zahŕňa aj komponenty východiskového stavu, zodpovedá požiadavkám stanoveným v kapitole analýzy stavu.

Po implementácií dátového trhoviska a finálneho dátového modelu sú dáta vizualizované prostredníctvom vizualizačného nástroja, čím je umožnená lepšia interpretácia a pochopenie získaných informácií. Táto implementácia je popísaná v siedmej kapitole.

2 TEORETICKÝ ZÁKLAD BUSINESS INTELLIGENCE

V nasledujúcich podkapitolách bude opísaná teória Business Intelligence. Bude definovaný pojem Business Intelligence (BI), jej základné princípy a aký je jej význam. Následne budú opísané hlavné komponenty a metodológie BI riešenia. V podkapitole bude taktiež opísaný základ návrhu Business Intelligence riešenia a teória dimenzionálneho modelovania. Nakoniec bude predstavená webová analytika a jej význam v BI, ktorá tvorí základ pre riešenie tejto práce. Cieľom tejto kapitoly je zhrnúť teoretické východisko pre tvorbu Business Intelligence riešenia pre dáta pochádzajúce z webovej stránky a aplikácie.

2.1 Definícia a význam Business Intelligence

Business Intelligence, podľa Novotného et al. (2005) predstavuje súbor procesov, aplikácií a technológií zameraných na efektívnu a cieľavedomú podporu firemných procesov. Tieto procesy napomáhajú analytickým a plánovacím aktivitám podnikov prostredníctvom multidimenzionálnych pohľadov na podnikové dáta. Aplikácie Business Intelligence pokrývajú široké spektrum funkcií podnikového riadenia, vrátane oblastí akými sú nákup, marketing, finančné riadenie, ľudské zdroje a mnoho ďalších. Business Intelligence možno chápať ako rozsiahly rámec, ktorý sa pohybuje od manažérskych aplikácií až po reporting, pričom jeho architektúra je zakotvená v informačných systémoch a informačných a komunikačných technológiách (IS/ICT).

Aplikácie Business Intelligence (BI), podľa Novotného et al. (2005), spĺňajú nasledujúce kľúčové princípy.

- Dáta sú organizované a uložené v multidimenzionálnej štruktúre, čo umožňuje rôzne úrovne analýzy a to na úrovni historickej, aktuálnej či prediktívneho pohľadu.
- Primárnym zdrojom dát pre BI aplikácie sú produkčné alebo transakčné systémy, ktoré podmieňujú kvalitu riešenia, pretože poskytujú vstupné dáta.
- Dáta sú v databáze archivované s časovými údajmi, čo umožňuje sledovať ich vývoj a trend v čase.
- Hierarchická štruktúra dimenzií umožňuje analýzu dát na rôznych úrovniach detailu, ich detailnú agregáciu.

Podľa Lokkena (2001) umožňuje Business Intelligence používateľom rýchlo pochopiť zložité informácie, aby mohli robiť lepšie a rýchlejšie rozhodnutia a tak efektívne dosiahnuť

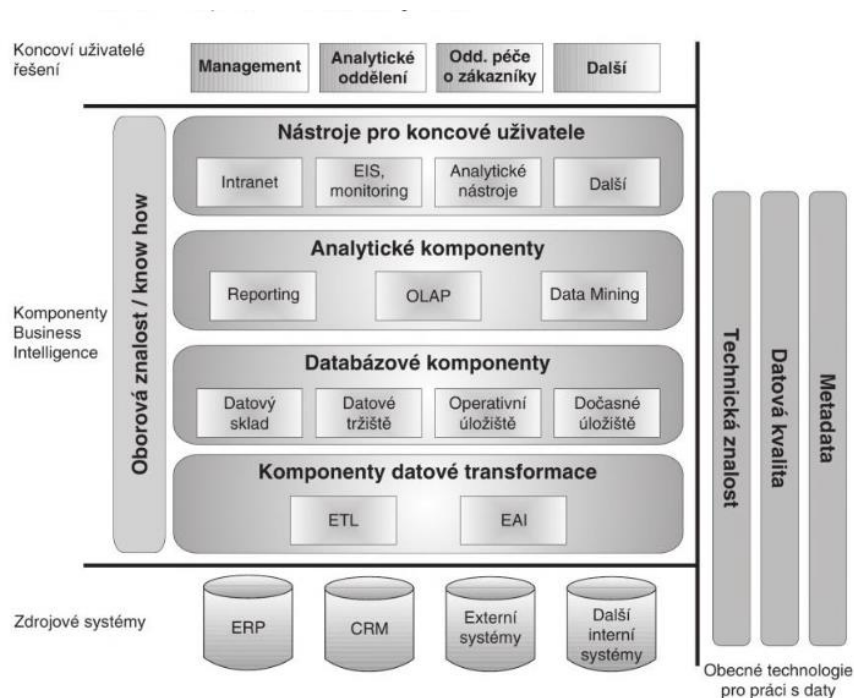
obchodné ciele. Kľúčové výhody, ktoré poskytuje, sú zvýšená efektívnosť a účinnosť organizácie. Niektoré riešenia totižto umožňujú rýchlejší dátový tok a ľahší prístup k informáciám v rámci organizácie. Iné, novšie riešenia sú založené na agresívnejšom prístupe, ktorý v určitých prípadoch vyžaduje zmenu existujúcich procesov a ich optimalizáciu.

2.2 Komponenty Business Intelligence

Usporiadanie a komplexnosť jednotlivých komponentov v riešení BI sa odvíja od špecifických potrieb a situácie daného podniku. Avšak, podľa Novotného et al. (2005), existuje obecné rozdelenie do logických celkov, ktoré budú opísané v nasledujúcich podkapitolách. Na obrázku 1 možno vidieť rozdelenie BI komponent do piatich základných častí, ktorými sú:

- komponenty dátovej transformácie,
- databázové komponenty,
- analytické komponenty,
- nástroje pre konečných užívateľov
- odborná znalosť (know-how).

Tieto vrstvy by nemohli fungovať bez primárneho či zdrojového systému, ktorý poskytuje základnú bázu dát.



Obrázok 1 – Obecná koncepcia architektúry BI

(Zdroj: Novotný et al., 2005)

2.2.1 Zdrojové systémy

Zdrojové alebo primárne (či transakčné) systémy sú databázy alebo aplikácie, ktoré obsahujú pôvodné, surové dáta potrebné pre analytické úlohy. Podľa Novotného et al. (2005) sa nezaradzujú medzi BI aplikácie, no slúžia ako zdroj dát pre tieto aplikácie. Tieto systémy môžu byť veľmi rozmanité a môže sa jednať napríklad o CRM systémy (*Customer Relationship Management*), ERP systémy (*Enterprise Resource Planning*), finančné systémy, ďalej tam môžu patriť aj externé systémy, akými sú sociálne médiá, webové analytiky a ďalšie.

Dáta z týchto zdrojových systémov sú extrahované pomocou ETL procesov a následne transformované do vhodnej formy pre analýzu a uložené do dátového skladu. Tento proces bude vysvetlený v nasledujúcej podkapitole. Novotný et al. (2005) dodávajú, že kvalita a presnosť dát v zdrojových systémoch priamo ovplyvňuje kvalitu a presnosť BI analýz, čo je dôvodom pre neustále venovanie pozornosti pre riadenie a udržiavanie týchto systémov.

2.2.2 Komponenty dátovej transformácie

Komponenty dátovej transformácie predstavujú vrstvu zodpovednú za extrakciu, transformáciu, čistenie a nahrávanie dát. Tieto komponenty slúžia pre úpravu dát a pokrývajú oblasť zbierania a prenosu dát zo zdrojových systémov. Podľa Novotného et al. (2005) sa tu zaradzujú ETL a EAI systémy.

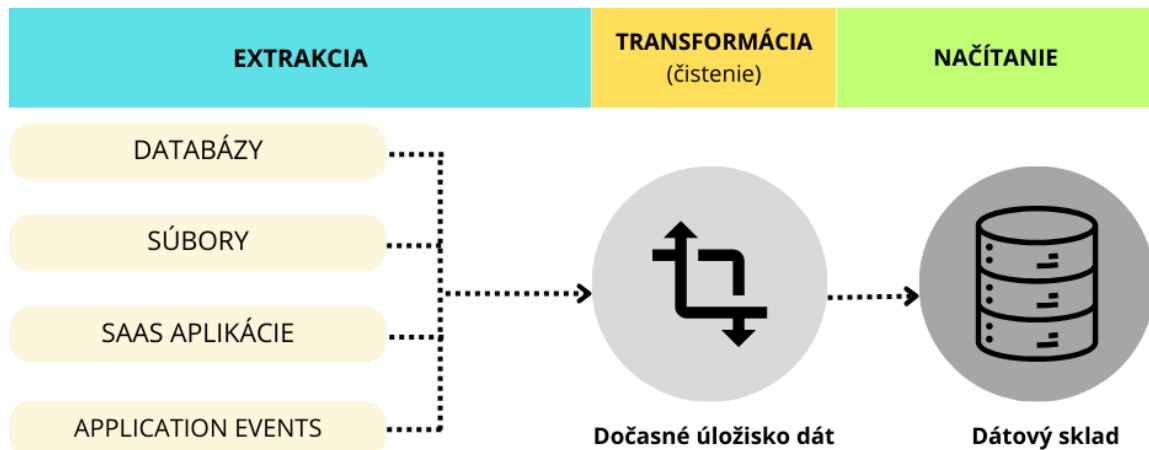
EAI systémy

EAI proces, z anglického *Enterprise Application Integration*, je na rozdiel od ETL procesu neviazaný na časový interval a funguje v reálnom čase. Pri EAI systémoch nie je potrebné využívať mnoho oddelených rozhraní medzi systémami, ktoré spolu komunikujú. Namiesto toho môžu aplikácie komunikovať cez jednu integračnú platformu, čo značne zjednodušuje procesy a zvyšuje efektivitu práce s dátami (Novotný et al., 2005).

ETL systémy alebo dátové pumpy

ETL proces je odvodený od anglického názvu *Extract, Transform and Load* a používa sa pre prenos dát medzi dvoma alebo viacerými systémami. Tento proces začína extrakciou dát zo zdrojového systému, nasleduje ich transformácia a čistenie na základe dátovej štruktúry cieľového systému a nakoniec sa tieto upravené dáta načítajú do cieľového systému – finálneho dátového skladu. Kimball (2004) navyše dodáva, že ETL systém je základným

krokom pri budovaní dátového skladu a má dve fázy – plánovanie, ktoré sa týka vytvárania stratégie a návrhu toho, ako bude ETL systém fungovať a ako bude štruktúrovaný. A fáza samotného toku dát, ktorá sa zaoberá procesom extrakcie, transformácie a načítania dát medzi rôznymi systémami. Tieto dve fázy sú úzko prepojené a vzájomne sa ovplyvňujú - rozhodnutia prijaté v jednej fáze ovplyvnia a formujú prácu v druhej fáze.



Obrázok 2 – Diagram ETL procesu

(Zdroj: Vlastné spracovanie)

2.2.3 Databázové komponenty

Tieto komponenty tvoria vrstvu pre ukladanie dát a obsahujú dáta, ktoré boli vyčistené, transformované a optimalizované pre analytické úlohy. Túto vrstvu tvoria dátové trhy, ktoré obsahujú dáta pre konkrétnu skupinu používateľov. Dátové trhy môžu byť buď odvodené z centrálného dátového skladu alebo samy tvoria dátový sklad, v závislosti od zvolenej koncepcie. Táto vrstva môže tiež obsahovať napríklad operačné dátové úložiská alebo dočasné úložiská dát (Novotný et al., 2005 a Dennis. 2017).

Dátový sklad

Dátový sklad alebo DWH, z anglického *Data Warehouse*, je podľa Billa Inmona – jedného zo zakladateľov tohto riešenia, integrovaný, subjektovo orientovaný, stály a časovo rozlíšený súhrn dát, usporiadaný na podporu potrieb manažmentu (Novotný et al., 2005 a Dennis. 2017).

- **Subjektívno orientovaný**, znamená, že dáta sú rozdeľované podľa ich typov a nie podľa aplikácie v ktorej vznikli.
- **Integrovaný** nesie vlastnosť, že dáta sú uložené v rámci celej organizácie, nie jednotlivých oddelení.

- **Stály**, znamená, že dátové sklady sú koncipované ako *read-only*. Dáta sú statické, vymazávajú sa administrátormi len v prípade značnej zastaranosti, či z legislatívnych dôvodov a existujú počas celej doby existencie DWH.
- Dátový sklad sa ďalej vyznačuje ako **časovo rozlíšený**, čo popisuje nutnosť dát byť časovo dimenzované, aby sa mohli vytvárať analýzy za konkrétne časové obdobia, popisuje Inmon (2002).

Bin Jiang reinterpretuje Inmonovu definíciu nasledovne: „*Dátový sklad je infraštruktúra založená na informačnej technológii pre organizáciu na integráciu, zber a prípravu dát na pravidelnej báze s cieľom uľahčiť analýzu.*” (Dennis. 2017).

Vzhľadom na to, že dátový sklad je viacrozmerná databáza, dáta a ich spôsob uloženia sa môžu líšiť, opisuje Tvrdíková (2008). Tri najčastejšie schémy, ktoré sa však najčastejšie využívajú sú nasledovné:

- **Hviezda (Star)**. Táto schéma je najčastejšie používaná pri prevode z relačnej do viacrozmerného usporiadania. Pozostáva z veľkej centrálnej tabuľky faktov a dopĺňujúcich tabuliek. Jej názov je odvodený od grafického zobrazenia, ktoré pripomína hviezdu.
- **Snehová vločka (Snowflake)** : Ide o variáciu schémy hviezdy, kde sú tabuľky normalizované, čo vedie k rozdeleniu dát do ďalších sekvenčných tabuliek. Hoci to môže viesť k oneskoreniu a menej prehľadnému grafickému zobrazeniu dát, jednou z hlavných výhod je úspora miesta v databáze dátového skladu. Efektívne využitie úložného priestoru môže výrazne ovplyvniť výkon a náklady na správu dátového skladu.
- **Súbor hviezd (Constellation)**: Táto schéma je kombináciou schém Star a Snowflake. V tomto modeli sa nachádza viacero tabuliek faktov, ku ktorým sú priradené tabuľky jednotlivých dimenzií. Tento prístup umožňuje väčšiu flexibilitu a komplexnosť pri modelovaní dát, pretože umožňuje vytvárať viacúrovňové a viacrozmerné štruktúry. Avšak táto zvýšená komplexita môže tiež znamenať vyššiu náročnosť na správu a analýzu dát.

Dátové trhovisko

Dátové trhovisko alebo DMA, z anglického *Data Mart*, je komponentou, ktorá je úzko spätá s dátovým sklado. Ich princíp je veľmi podobný, avšak zásadný rozdiel je v tom, že dátové trhovisko obsahuje špecifický segment dát pre určitý okruh používateľov, môže sa jednať

o oddelenia, pobočky či divízie. Vďaka špecifikácii a menšiemu obsahu dát, umožňujú rýchlejší prístup k dátam na analýzu. Zároveň umožňujú lepšiu kontrolu nad dátami, keďže každé oddelenie alebo skupina má svoje vlastné dátové trhovisko (Novotný et al., 2005 a Slánský, 2018).

Dočasné úložisko dát

Dočasné úložisko alebo DSA, z anglického *Data Staging Area*, sa používa na ukladanie dát extrahovaných z produkčných systémov a je najviac využívané počas procesu ETL. Jeho hlavnou úlohou je podporovať rýchlu a efektívnu extrakciu dát. Táto súčasť riešenia má široké uplatnenie v prípadoch zaťažených produkčných systémov, kde je potrebné transformovať dáta s minimálnym dopadom na výkon systémov, avšak nie je povinnou súčasťou riešenia BI (Novotný et al., 2005).

Dáta uložené v dočasnom úložisku sú detailné, nekonzistentné, neočistené dáta zo zdrojového systému, po ich spracovaní a prenose do cieľového dátového skladu sa tieto dáta z dočasného úložiska odstránia (Pour et al., 2018).

Operatívne dátové úložisko

Operatívne dátové úložisko, ODS, je rovnako ako DSA nepovinnou súčasťou BI riešenia. Slúži k integrácii a uloženiu operačných dát potrebných pre reporting a analýzu. Implementácia ODS je veľmi situačná, komentuje (Scheeps, 2008), a využíva sa pri potrebe analyzovať aktuálny snímok (z anglického *snapshot*) operačných dát potrebných pre rozhodovacie procesy. Tým sa líši od dátových skladov, ktoré naopak slúžia ako úložisko historických záznamov. ODS zabezpečuje vzhľad na dáta, ktoré sú aktuálne a dodané včas.

Rovnako ako DSA, obsahuje dáta bez histórie a meniace sa pri každom nahratí, no hlavný rozdiel možno vidieť v ich účele. ODS, na rozdiel od DSA, je budovaná práve pre to, aby obsahové dáta boli prístupné aj koncovým užívateľom pre potrebné analýzy či dotazy a to s minimálnym oneskorením (Novotný et al., 2005).

2.2.4 Analytické komponenty

Analytické komponenty tvoria, podľa Novotného et al. (2005), vrstvu pre analýzu dát, táto vrstva je tvorená klientskými aplikáciami v oblasti BI, ktorých cieľom je poskytnúť dáta používateľovi takým spôsobom, aby z nich získal potrebné informácie.

Pour et al. (2018) doplňuje ich princípy do troch bodov, a to:

- sú navrhované pre sledovanie manažérskych informácií, plnenie cieľov a iné firemné procesy,
- poskytujú nástroje pre online analýzy (napríklad analýzy trendov),
- sú jednoducho ovládateľné a vďaka grafickému výstupu poskytujú vysokú pridanú hodnotu.

V súčasnom poňatí sa vývoj analytických aplikácií ubera dvoma smermi, dopĺňa Pour et al. (2018), jedná sa o pokročilú analytiku v rámci CPM (*Corporate Performance Management*) alebo v podobe Business Analytics. V Business Analytics sa, okrem metódy riadenia nákladov, využívajú aj metódy dolovania dát, prediktívne modelovanie a rôzne iné pokročilé metódy.

Reporting

Reporting predstavuje komplexný systém informácií a ukazovateľov, ktoré charakterizujú činnosti spoločnosti vo vhodnej forme (Pour et al., 2018).

Realizuje sa dotazmi do databázy, ktorých výstupom môže byť tabuľka alebo iný prehľad dát. Takýto výstup je možné ďalej použiť vo vizualizačnom nástroji pre lepší pohľad na výsledné dáta. Reporty je možné ďalej rozlíšiť na statické, také ktoré neumožňujú žiadne zmeny v navigácii a sú bez filtrov, a dynamické, ktoré umožňujú filtrovanie dimenzií, zmeny úrovne detailov, či kontingenčné tabuľky. Reporty sa dajú taktiež rozdeliť na štandardné a ad hoc reporty (Novotný et al., 2005).

- **Štandardný reporting** sa vyznačuje tým, že predpripravené dotazy sa spúšťajú v určitých časových intervaloch.
- **Ad hoc reporting** je charakterizovaný tým, že na databázu sú formulované špecifické dotazy, ktoré sú explicitne vytvorené používateľom, väčšinou jednorazovo. Tieto reporty sa často nazývajú aj ako data discovery reporty (Novotný et al., 2005 a Slánský, 2018).

Systémy On-Line Analytical Processing (OLAP)

Systémy On-Line Analytical Processing (OLAP) sú kľúčovou súčasťou analytických aplikácií. Na rozdiel od reportovacích aplikácií sú analytické aplikácie flexibilnejšie, čo je vďaka tomu, že typicky umožňujú vykonávanie operácií ako *drill-down* (prehĺbenie analýzy), *drill-up* (zjednodušenie analýzy), *slice* (výber podmnožiny dát) alebo *dice* (výber viacerých podmnožín dát). Tieto aplikácie sú väčšinou prevádzkované nad databázami typu OLAP,

ktoré sú optimalizované pre rýchle a efektívne analytické spracovanie dát (Novotný et al., 2005).

Data Mining

Dolovanie dát, známe aj ako *Data Mining*, je proces získavania relevantných informácií z veľkých databáz, ktoré neboli predtým definé alebo známe. Tento proces využíva rôzne techniky, ako sú rozhodovacie stromy, neurónové siete, regresná analýza a mnohé ďalšie pokročilé metódy. Cieľom tohto procesu je identifikovať vzory, trendy alebo vzťahy v dátach, ktoré môžu poskytnúť cenné informácie alebo poznatky pre podnikanie alebo rozhodovanie (Pour et al., 2018).

2.3 Metodológie tvorby BI riešenia

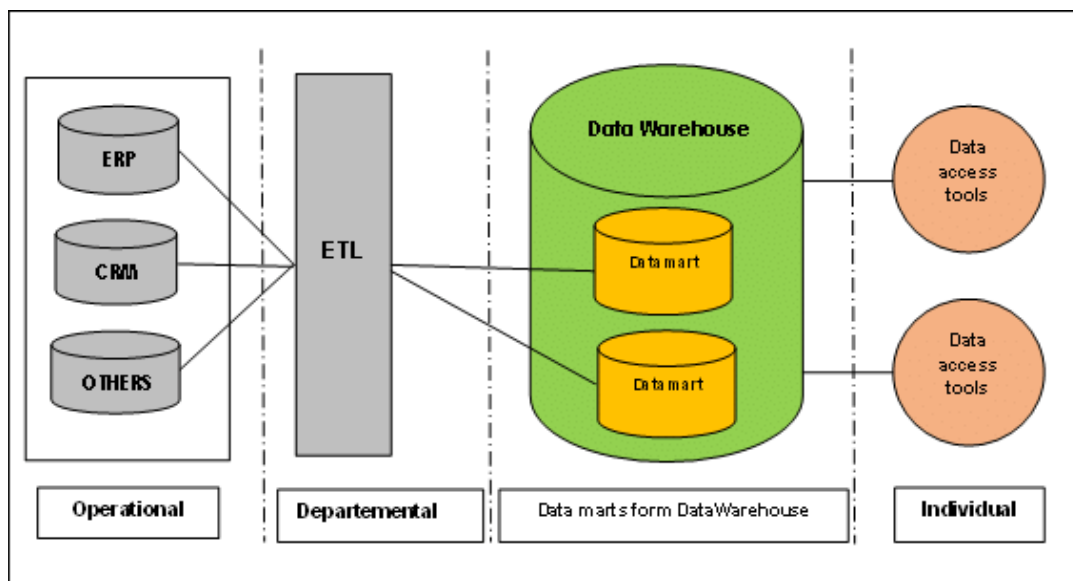
Dve najviac diskutované prístupy v tvorbe dátového skladu sú: metóda podľa Inmona a metóda podľa Kimballa. V priebehu rokov sa vedú odborné diskusie na tému, ktorý prístup k dátovému skladu je vhodnejší a efektívnejší pre podnikové prostredie. Napriek tomu stále neexistuje jednoznačný konsenzus, keďže oba prístupy prinášajú svoje špecifické výhody a zároveň aj určité nevýhody (Mukherjee, 2021).

Laberge (2012) dopĺňa, že obe tieto metodológie vyžadujú existenciu abstraktného podnikového dátového modelu pre dátový sklad a považujú dáta za cenné podnikové aktívum. Tento podnikový dátový model, ktorý nie je špecificky navrhnutý pre žiadnu konkrétnu aplikáciu, môže slúžiť ako spojivový prvok, ktorý udržiava podnikové aktíva pohromade. Tento model organizuje všetky podnikové dáta do jedného centralizovaného návrhu a komponenty tohto podnikového aktíva môžu byť využité v akýchkoľvek aplikáciách. Zároveň každá aplikácia má možnosť navrhnuť svoje vlastné dátové prostredie tak, aby čo najlepšie vyhovovalo jej špecifickým potrebám a operáciám.

2.3.1 Metóda podľa Kimballa

Kimballov prístup, alebo *bottom-up* prístup, k návrhu dátového skladu je založený na tvorbe menších databáz, tzv. dátových trhov, ktoré sú vytvorené na základe konkrétnych business požiadaviek a oddelené pre špecifické útvary podniku (Novotný et al., 2005 a Yessad & Labiod, 2016).

- Prvý dátový trh sa vytvára na základe analytických potrieb konkrétneho oddelenia firmy. Identifikujú sa spoločné dimenzie s predpokladom, že budú využité aj pre ďalšie dátové trhy.
- Následné dátové trhy sa konštruujú tak, aby maximálne využívali už existujúce spoločné dimenzie, ktoré sa prípadne modifikujú podľa potreby.
- ETL (Extract, Transform and Load) komponenty sa vytvárajú nezávisle pre každý dátový trh. Keď sú dáta nahrané, nastáva fáza, kedy sa dáta načítajú do dimenzionálneho dátového skladu.



Obrázok 3 – Schéma tvorby dátového skladu podľa Kimballovej metodológie

(Zdroj: Yessad & Labiod, 2016)

Kimballov prístup k dátovému skladu zahŕňa vytvorenie „*conformed dimensions*”, ktoré majú rovnaký význam a hodnoty naprieč rôznymi faktovými tabuľkami alebo zdrojmi dát. Tieto dimenzie umožňujú konzistentnú analýzu dát v rámci celého dátového skladu, pretože umožňujú používateľom porovnávať a kombinovať fakty z rôznych obchodných procesov. Dimenzia dátumu tak môže byť použitá na analýzu predajov či o zákazníckom servise v čase. Hlavnou výhodou tohto prístupu je, že umožňuje rýchlejšie a flexibilnejšie spracovanie dát, keďže sa priamo zameriava na špecifické obchodné potreby. Avšak, prístup vyžaduje viac plánovania a koordinácie, pre zabezpečenie konzistencie dát naprieč rôznymi dátovými trhmi (Novotný et al., 2005 a Joshi, 2023).

Tabuľka 1 – Výhody a nevýhody Kimballovej metodológie

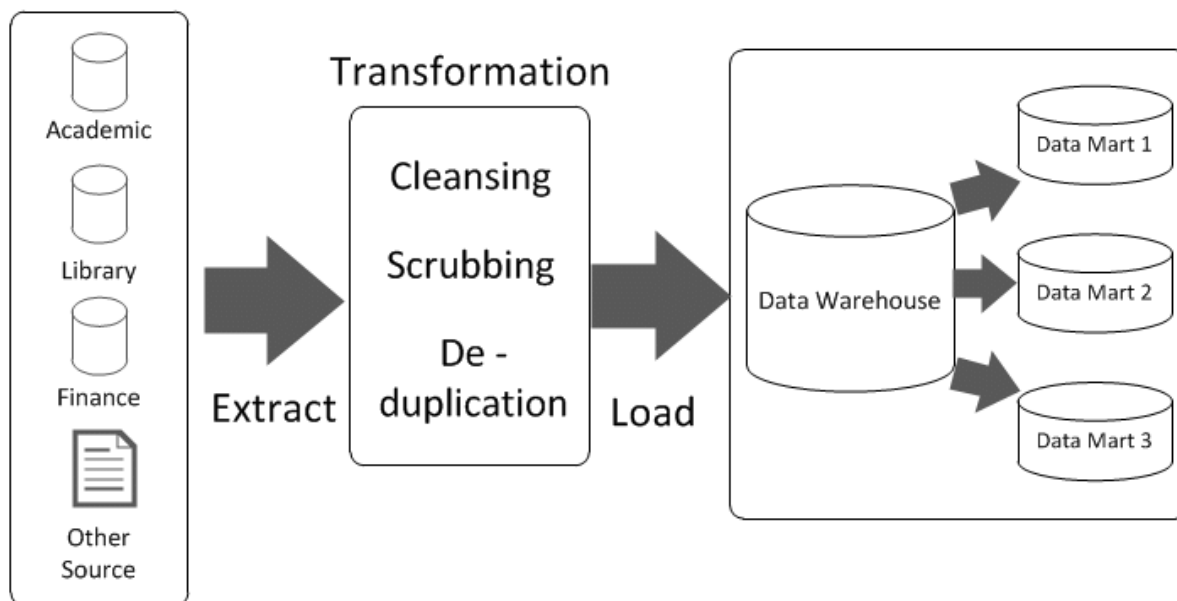
(Zdroj: Vlastné spracovanie podľa: Mukherjee, 2021)

Výhody	Nevýhody
Analytické úlohy vedia byť vykonané v najrýchlejšie možnom čase.	Údržba je tu náročnejšia v porovnaní s modelom Inmona.
Denormalizovaná forma dát uľahčuje a zrýchľuje dotazovanie, načítavanie a analýzu dát.	Nezávislé dátové trhoviská môžu viesť k prekrytiu a duplicitě dát, čo vedie k rozličnej interpretácii a nákladom.
Jednotlivé aplikácie poskytujú užívateľom všetko potrebné a nemusia prechádzať jednotlivými komponentami BI riešenia.	Denormalizované dáta môžu spôsobiť redundanciu a komplikovať pridávanie nových stĺpcov alebo funkcií do tabuľky faktov.
Na vytvorenie a správu tohto typu dátového skladu sú potrebné menšie zručnosti.	

2.3.2 Metóda podľa Inmona

Metodológia Inmona, známa aj ako normalizovaný alebo *top-down* prístup, bola vyvinutá Billom Inmonom a jeho kolegami. Tento prístup sa zameriava na vytvorenie jediného, centralizovaného dátového skladu, ktorý uchováva všetky podnikové dáta z rôznych zdrojov. Tieto dáta sú uložené v normalizovanej forme, čo znamená, že sú organizované do tabuliek tak, aby sa minimalizovala redundancia, opakovanie rovnakých údajov a maximalizovala integrita, čo je konzistencia a presnosť údajov (Joshi, 2023 a Novotný et al., 2005).

Dátový sklad slúži ako jednotný zdroj pravdy pre celú organizáciu. Je pravidelne aktualizovaný pomocou procesov ETL, ktoré získavajú nové údaje z rôznych zdrojov a integrujú ich do skladu. Z tohto centralizovaného dátového skladu potom môžu byť odvodené viaceré dátové trhy. Tieto dátové trhy sú špecificky upravené a optimalizované pre konkrétne analytické účely alebo potreby konkrétnych oddelení v rámci organizácie (Novotný et al., 2005).



Obrázok 4 – Schéma tvorby dátového skladu podľa Inmonovej metodológie

(Zdroj: Moscoso-Zea et al., 2018)

Vďaka tomuto prístupu je možné dosiahnuť konzistentnú a kvalitnú analýzu dát naprieč celou organizáciou. Táto metodológia je vhodná pre komplexné a detailné analýzy, keďže zabezpečuje konzistenciu a kvalitu dát naprieč rôznymi oblasťami a úrovňami detailnosti. V tabuľke 2 možno nájsť výhody a nevýhody tohto prístupu (Mukherjee, 2021).

Tabuľka 2 – Výhody a nevýhody Inmonovej metodológie

(Zdroj: Vlastné spracovanie podľa: Mukherjee, 2021 a Laberge, 2012)

Výhody	Nevýhody
Dátový sklad slúži ako jednotný zdroj dát pre celý podnik.	Môžu sa meniť zdrojové systémy a požiadavky užívateľov.
Tento prístup má veľmi nízku redundanciu dát, takže je menšia pravdepodobnosť nepravidielností pri aktualizácii dát.	Predbežná práca vyžaduje kvalifikovaných pracovníkov a veľa času.
Zjednodušuje obchodné procesy poskytovaním celkového prehľadu a rozdeľovaním dát podľa potrieb rôznych oddelení.	Primárne nastavenie je nákladné a časovo náročné.
Poskytuje väčšiu flexibilitu, pretože je ľahšie aktualizovať dátový sklad v prípade zmien v obchodných požiadavkách alebo zdrojových dátach.	Niektoré detailné dáta sa musia umiestňovať aj do dátových trhovísk.
Architektúra je dostatočne flexibilná pre podporu náročných analytických úloh. Dáta sú nie len agregované, ale detailné.	Keďže sa dáta používajú v normalizovanej forme, dotazovanie sa stáva ťažkým kvôli jednotlivým databázovým vrstvám.
Náklady na údržbu sú nižšie ako pri Kimballovom prístupe.	

2.4 Návrh Business Intelligence riešenia

Návrhová časť Business Intelligence riešenia predstavuje, podľa Novotného et al. (2005), základný kameň pre jeho následné modelovanie. Laberge (2012) dopĺňa, že pre návrh riešenia je však potrebné poznať plánované použitie dát.

Novotný et al. (2005) kategorizujú modely a ich návrh do troch hlavných úrovní: konceptuálnej, logickej a fyzickej. Konceptuálna a logická úroveň sú tvorené tromi kľúčovými činnosťami:

1. **Dimenzionálne analýzy a modelovanie.** Táto činnosť sa zaoberá identifikáciou a štruktúrovaním rôznych dimenzií dát, ktoré môžu byť analyzované v rámci Business Intelligence riešenia.
2. **Modelovanie dátového skladu.** Tu sa navrhuje štruktúra dátového skladu, ktorá efektívne a efektívne ukladá a spravuje dáta pre analýzu.
3. **Riešenie zmien v dimenziách.** Táto činnosť sa zaoberá aktualizáciou a údržbou dimenzií v reakcii na zmeny v podnikových dátach alebo požiadavkách na analýzu.

Tieto tri činnosti sú neoddeliteľnou súčasťou procesu návrhu a modelovania Business Intelligence riešenia a majú rozhodujúci vplyv na jeho úspešnosť a efektívnosť (Novotný et al., 2005 a Pour et al., 2018).

2.4.1 Prístupy k návrhu dátového skladu

Úspech celého projektu je daný správnym návrhom databáze. Podstatou návrhu je transformácia výsledkov dimenzionálnej analýzy do relačných databázových štruktúr. Návrh aj analýzu možno, podľa Novotného et al. (2005), rozdeliť do troch základných úrovní:

- **Konceptuálna.** V konceptuálnom modeli sa definujú základné entity a ich väzby v dátovom sklade.
- **Logická.** V logickom návrhu modelu sa jednotlivé entity transformujú do návrhu logickej štruktúry databázových tabuliek, vrátane štruktúry atribútov týchto tabuliek.
- **Fyzická.** Vo fyzickom modeli sa špecifikujú všetky potrebné technologické charakteristiky databázových tabuliek a ich väzieb.

The Data Warehouse Bus Architecture Matrix

Dátový sklad môže byť navrhnutý niekoľkými prístupmi, Novotný et al. (2005) spomína techniku pomocou matice nazývanú *The Data Warehouse Bus Architecture Matrix*, ktorá je založená na identifikácii dátových trhovísk a na určení všetkých možných dimenzií, ktoré sa tu vyskytujú. Tie sa tvoria pomocou matice, ktorá slúži ako podklad pre logický a fyzický model. Táto metóda bola prezentovaná R. Kimballom a je podmienená jeho prístupom, technika rozdeľuje postup do niekoľkých krokov, a to:

1. *identifikácia dátových trhovísk* a to najprv takých, ktoré pochádzajú z jedného zdroja, teda začať identifikovať od najjednoduchších trhovísk k tým zložitejším,

2. *identifikácia dimenzií* pre každé z potrebných dátových trhovísk a vytvorenie „**conformed dimensions**“, ktoré predstavujú také dimenzie, ktoré sa vyskytujú vo väčšine dátových trhovísk,
3. *detailný návrh jednotlivých fakt tabuliek*. Prvým krokom je výber dátového trhoviska, deklaráciu granularitu, následne sa vyberú dimenzie a následne sa vyberú fakty.

Tieto kroky sú však iteračné, to znamená, že každý krok je postupne prehodnocovaný, doplňovaný či modifikovaný. Okrem tejto metódy sa používa aj metóda zdrojového modelovania, *trriage* technika či cieľové modelovanie (Novotný et al., 2005).

2.4.2 Dimenzionálne modelovanie

Dimenzionálne modelovanie je technika návrhu databáz, ktorá sa často používa v systémoch pre Business Intelligence a dátových skladoch. Je to metóda, ktorá umožňuje efektívne organizovať a analyzovať veľké množstvá dát. V dimenzionálnom modelovaní sú dáta organizované do tabuliek, ktoré sú rozdelené na „fakty“ a „dimenzie“ (Pour et al., 2018).

- **Faktové tabuľky** obsahujú kvantitatívne údaje, ako sú predaj, počet kliknutí na webe alebo počet transakcií.
- **Dimenzionálne tabuľky** obsahujú opisné údaje, ktoré popisujú faktové údaje, čas (deň, mesiac, rok), geografická poloha (mesto, štát, krajina) alebo detaily o produktoch (názov produktu, kategória, dodávateľ).

Tento prístup umožňuje rýchle a flexibilné analýzy dát z rôznych uhlov pohľadu, čo je kľúčové pre Business Intelligence a rozhodovacie procesy v podnikaní (Novotný et al., 2005). Dimenzionálne modelovanie, ako ho opisuje Pour et al. (2018), je označované ako hrubé dimenzionálne modelovanie. Jeho hlavným cieľom je definovať obsah BI riešenia bez potreby technickej implementácie v konkrétnej databáze. Tento návrh modelovania sa dá rozdeliť do troch kľúčových častí alebo celkov:

1. **Dimenzie**
2. **Ukazovatele**
3. **Vzťahy medzi dimenziami a ukazovateľmi**

Každý z týchto celkov v rámci BI riešenia má určité charakteristiky, ktoré je možné prispôbiť alebo upraviť podľa konkrétnych potrieb. Hĺbka a detailnosť týchto charakteristík sú závislé od špecifických požiadaviek užívateľov a analytikov daného BI riešenia (Pour et al., 2018).

Ukazovatele (faktové tabuľky)

Ukazovatele, známe aj ako faktové tabuľky, obsahujú kvantitatívne hodnoty ukazovateľov, ktoré podnik sleduje. Tieto ukazovatele môžu zahŕňať rôzne metriky alebo merania, ako sú celkové predaje, počet návštevníkov webovej stránky, percento konverzií a podobne. Faktové tabuľky sú spojené s dimenzionálnymi tabuľkami, ktoré poskytujú kontext pre ukazovatele. Napríklad faktová tabuľka môže obsahovať informácie o celkových predajoch, zatiaľ čo dimenzionálna tabuľka môže obsahovať detaily o jednotlivých produktoch, ktoré boli predané (Laberge, 2012).

Pour et al. (2018) vysvetľujú, že úroveň granularity v kontexte dátových súborov odkazuje na úroveň detailnosti údajov, ktoré sú uložené. Nižšia granularita znamená menej detailné údaje. To môže obmedziť schopnosť analyzovať údaje na podrobnejšej úrovni, napríklad podľa konkrétneho dátumu, regiónu alebo značky. Na druhej strane, vyššia granularita poskytuje viac detailných informácií, čo umožňuje podrobnejšiu analýzu. Ak sú údaje uložené na úrovni jednotlivých transakcií, je možné analyzovať predaje podľa konkrétneho dňa, produktu alebo miesta predaja. Výber správnej úrovne granularity je dôležitý aspekt návrhu dátových súborov a systémov Business Intelligence. Okrem granularity sa v prípade ukazovateľov, podľa Novotného et al. (2005), určujú nasledujúce charakteristiky:

- *Jednoznačná identifikácia* (unikátne označenie ukazovateľa).
- *Plný názov* (kompletný názov ukazovateľa).
- *Obsahové vymedzenie a účel* (definícia a cieľ ukazovateľa).
- *Zdroje dát* (zdroj dát ukazovateľa).
- *Kalkulácia ukazovateľa* (spôsob výpočtu ukazovateľa).
- *Typ, formát* (formát a typ dát ukazovateľa).
- *Merná jednotka* (jednotka ukazovateľa).
- *Možnosť agregácie* (agregácia ukazovateľa).
- *KPI* (kľúčový ukazovateľ výkonnosti).
- *Väzby ukazovateľov na definované dimenzie* (spojitosť s definovanými dimenziami).

Dimenzie (dimenzionálne tabuľky)

Podľa Laberge (2012), dimenzia je tabuľka, ktorá poskytuje kontext pre metriky uložené v tabuľke faktov, ich kvalifikáciu. Dimenzie určujú granularitu faktov a umožňujú ich filtrovanie. V procese návrhu logického modelu, sú tabuľky označované ako entity a ich stĺpce sú považované za atribúty. Tieto entity dimenzií obsahujú atribúty, ktoré kvalifikujú a

popisujú danú dimenziu. Napríklad, v prípade dátového skladu pre obchodný reťazec by mohla byť jednou z dimenzií „*produkt*“. Atribúty tejto dimenzie by mohli zahŕňať údaje ako *názov produktu, kategória, dodávateľ a cena*. Tieto atribúty potom poskytujú kontext pre metriky uložené v tabuľke faktov, ako je napríklad celkový počet predaných jednotiek. Pri definovaní týchto dimenzií sa, podľa Novotného et al. (2005), berú do úvahy nasledujúce charakteristiky:

- *Identifikácia dimenzie* (unikátne označenie, ktoré identifikuje dimenziu).
- *Plný názov dimenzie* (kompletný a oficiálny názov dimenzie).
- *Obsah* (definícia a popis informácií obsiahnutých v dimenzii).
- *Typ* (klasifikácia dimenzie podľa jej charakteristiky alebo účelu).
- *Zdroj dát* (miesto alebo systém, odkiaľ pochádzajú dáta pre dimenziu).
- *Štruktúra prvkov* (organizácia alebo usporiadanie jednotlivých prvkov v rámci dimenzie).
- *Atribúty* (dodatkové vlastnosti alebo charakteristiky spojené s dimenziou).
- *Doplňujúce poznámky* (akékoľvek ďalšie informácie alebo poznámky týkajúce sa dimenzie).

Tieto charakteristiky predstavujú základný rámec a môžu byť upravené, doplnené alebo spresnené podľa konkrétnych potrieb podniku alebo v priebehu procesu dimenzionálneho modelovania (Pour et al., 2018).

Proces a princípy dimenzionálneho modelovania

Princípy a postupy dimenzionálneho modelovania sa môžu líšiť v závislosti od konkrétnej situácie podniku. Avšak, podľa Novotného et al. (2005), existujú štandardizované kroky, ktoré sa bežne dodržiavajú:

1. Výber a základné obsahové vymedzenie riešenej oblasti podnikového riadenia.
2. Navrhnutie všetkých relevantných dimenzií vrátane ich charakteristík.
3. Navrhnutie ukazovateľov, ich jednotlivých charakteristík a určenie úrovne granularity.
4. Riešenie vzťahov medzi dimenziami a ukazovateľmi.
5. Premietnutie riešenia do návrhu dimenzií tabuliek, návrhu tabuliek faktov a návrhu schémy.

Tieto kroky poskytujú rámec pre vytváranie efektívneho a prispôsobiteľného dimenzionálneho modelu, ktorý vyhovuje potrebám podniku.

Pri riešení vzťahov medzi tabuľkami, čo je štvrtým krokom, sa často využívajú dve základné schémy alebo hierarchie - hviezdicová schéma a schéma snehovej vločky, pokračuje Novotný et al. (2005). V centre týchto schém sa nachádza tabuľka faktov. Tá obsahuje sledované ekonomické hodnoty a ďalšie ukazovatele, ktoré sú identifikované pomocou kompozitného kľúča, tvoreného cudzími kľúčmi z dimenzionálnych tabuliek. Dimenzionálne tabuľky potom slúžia ako úložisko textových informácií, ktoré poskytujú kontext pre numerické hodnoty uložené v tabuľke faktov.

2.5 Webová analytika

Webová analytika je, podľa Jansen et al. (2022), proces zbierania, merania, analýzy a spracovania digitálnych dát s cieľom zlepšiť poznatky o správaní návštevníkov webových stránok. Je kľúčovou súčasťou Business Intelligence, konkurenčnej analýzy, benchmarkingu webových stránok, online reklamy, online marketingu a digitálneho marketingu, pretože na základe týchto meraní sú následne vykonávané obchodné rozhodnutia. Okrem webových stránok sem patria aj mobilné aplikácie, kde analytika funguje na veľmi podobnom princípe.

Bekavac a Pranicevic (2015) popísali webovú analytiku ako analýzu kvalitatívnych a kvantitatívnych dát na webovej stránke vedúcu k zlepšeniu užívateľského zážitku, ktoré vedie k účinnej a efektívnej realizácii plánovaných cieľov firmy, pričom kvantitatívne dáta odôvodňujú správanie užívateľov a vďaka kvalitatívnym možno pochopiť prečo sa používatelia správajú takým spôsobom.

V pojatí Business Intelligence, aplikácie webovej analytiky predstavujú špeciálnu funkcionality Business Intelligence a slúžia na meranie a analýzu ukazovateľov získaných z prevádzky webových aplikácií, ako sú napríklad elektronické obchody alebo informačné portály. Používatelia takýchto riešení sú najmä majitelia e-shopov, dodávatelia e-shopov alebo samotní tvorcovia webových stránok (Novotný et al., 2005).

Nástroje webovej analytiky umožňujú sledovať a analyzovať metriky, ako sú napríklad počet návštevníkov stránky, čas strávený na stránke, konverzné miery, cesty návštevníkov na webe a ďalšie. Kaushik (2009) ďalej vysvetľuje, že webová analytika nie je určená len pre samotnú tvorbu reportov, ale proces, ktorý vedie k postupnej optimalizácii webovej stránky. Tento proces sa skladá z piatich častí:

- **Vytýčenie cieľov.** Definícia cieľu webovej analytiky v rámci konkrétnych požiadaviek podniku.

- **Stanovenie kľúčových ukazovateľov výkonnosti (KPI).** Ktoré merateľné údaje vedia vecne opísať dosahovanie definovaných cieľov.
- **Meranie dát.** Meranie dát môže byť implementované niekoľkými spôsobmi. Druh implementácia by mal byť zvolený na základe situácie daného podniku.
- **Analýza dát.** Analýza dát v rámci webovej analytiky predstavuje proces preskúmania zhromaždených údajov s cieľom identifikovať trendy a vzorce. Tieto informácie môžu odhaliť potenciálne prekážky brániace konverziám, ktoré po odstránení môžu prispieť k lepšiemu dosahovaniu podnikových cieľov.
- **Implementácia zmien.** Kľúčovým krokom celého procesu je implementácia zmien, pretože bez ich využitia na optimalizáciu webu alebo procesov spoločnosti sú všetky zhromaždené údaje bezcenné. Pri realizácii zmien je dôležité sledovať ich vplyv na dosiahnutie cieľov a hodnotiť ich efektívnosť pomocou rôznych druhov testovania (Kaushik, 2009).

Aplikácie webovej analytiky sú teda dôležitou súčasťou moderných BI systémov, dodáva Novotný et al. (2005), poskytujú cenné informácie o správaní sa užívateľov na webe, ktoré môžu byť kľúčové pre úspech online podnikania. Aplikácie poskytujú nasledujúce druhy analýz:

- **Základné ukazovatele.** Poskytujú štatistické informácie o počte prístupov na jednotlivé podstránky, celkový počet prístupov a prístupy podľa typu prehliadača.
- **Chovanie návštevníkov.** Analýzy, ktoré pomáhajú porozumieť, ako sa návštevníci pohybujú po stránkach. Tieto informácie sú často kombinované s ďalšími dimenziami pre hlbšiu identifikáciu a analýzu správania užívateľov.
- **Analýza webového kanálu.** Poskytuje prehľad o výkonnosti komunikačného kanála, jeho výnosnosti, nákladoch na získanie nového zákazníka a pod.

Novotný et al. (2005) navrhuje integráciu údajov z webových analytických aplikácií s informáciami poskytnutými CRM systémom. Týmto spôsobom je možné získať komplexné informácie o jednotlivých zákazníkoch a ich interakciách. Taktiež odporúča zahrnutie finančných ukazovateľov a metrík, ktoré poskytujú pohľad na náklady a výkonnosť. Tento prístup umožňuje hlbšie pochopenie správania zákazníkov a ich interakcií s webovou stránkou, čo môže viesť k efektívnejšiemu cieleniu marketingových kampaní a lepšiemu riadeniu vzťahov so zákazníkmi. Okrem toho, začlenenie finančných ukazovateľov a metrík poskytuje podniku možnosť lepšie monitorovať a hodnotiť náklady a výkonnosť svojich online aktivít, čo môže prispieť k zlepšeniu celkovej efektívnosti a rentability.

2.5.1 Životný cyklus zákazníka

Podľa McDaniel (2023) je životný cyklus zákazníka marketingovou taktikou, ktorú firmy aplikujú na prilákanie, konverziu a udržanie zákazníkov v priebehu celého nákupného procesu. Táto stratégia sa silno zakladá na pochopení zákazníka a jeho interakcií v rámci marketingových a predajných procesov, čo umožňuje efektívne prispôbenie zážitku zákazníka.

Kirsch (2022) hovorí, že koncept životného cyklu zákazníka je spojený s marketingovým lievikom a modelom AIDA. Model AIDA však pokrýva len štyri fázy: *attention* (pozornosť), *interest* (záujem), *desire* (túžba), *action* (akcia).

Digital Marketing Institute (2019) uvádza, že na rozdiel od lievika, životný cyklus zákazníka popisuje etapy, ktorými zákazník prechádza vo vzťahu k značke, od prvého kontaktu až po konečný nákup a ďalšie interakcie. Zahrňuje aj fázy po nákupe, ako je retencia či vernosť. Tento model sa sústreďuje na dlhodobý vzťah a nie len na získavanie, ale aj na udržiavanie zákazníkov a maximalizáciu ich hodnoty v priebehu času. Oba modely sa využívajú na mapovanie cesty zákazníka, avšak lievik je viac orientovaný na proces vedúci k nákupu, kým životný cyklus zákazníka zohľadňuje celkový vzťah medzi zákazníkom a značkou.

Kushnazarov (2019) rozdeľuje model zákazníckeho životného cyklu do 5 štádií:

1. **dosah**
2. **akvizícia**
3. **konverzia,**
4. **retencia,**
5. **vernosť alebo advokácia.**



Obrázok 5 – Fázy zákazníkoveho životného cyklu

(Zdroj: Kushnazarov, 2019)

Každá fáza reprezentuje jedinečnú etapu v zákazníkovej ceste zákazníka, pričom každá z nich vyžaduje špecifický prístup a stratégie, dopĺňa Kirsch (2022). Optimalizácia zákazníkovej cesty je kľúčovým aspektom tejto stratégie. Ide o proces zisťovania toho, čo motivuje alebo naopak bráni vykonaniu konverzie jednotlivých návštevníkov. Pre efektívnu optimalizáciu je nutné premýšľať o tom, ako poskytnúť každému zákazníkovi to, čo potrebuje a chce v každej fáze, čo si vyžaduje hlboké porozumenie zákazníka a jeho potrieb. Práve to je možné dosiahnuť prostredníctvom webovej analytiky. Tá umožňuje identifikovať jednotlivé štádiá v marketingovom cykle zákazníka a lepšie pochopiť správanie a potreby zákazníkov v jednotlivých fázach (Bruce Clay Europe, 2019).

Dosah

Dosah, tiež označovaný ako fáza povedomia, predstavuje úvodnú etapu životného cyklu zákazníka. Fáza povedomia je kľúčovou súčasťou procesu získavania zákazníka a zahŕňa prilákanie zákazníkov na webovú stránku alebo do aplikácie, čo často predstavuje ich prvý priamy kontakt so značkou. Je dôležité si uvedomiť, že potenciálni zákazníci môžu vstúpiť do konverzného tunela aj bez priameho úsilia zo strany firmy, napríklad ak sami identifikujú potrebu produktu alebo služby, ktorú firma poskytuje. Cieľom firmy je byť na vrchole zákazníkoveho uváženia a asociovať svoju značku s budúcimi potrebami zákazníkov (Coffman, 2022 a Morgan, 2023).

Akvizícia

Fáza zvažovania nastáva v momente, keď sa potencionálni zákazník stáva oboznámeným so značkou, navštívi webovú stránku a začína uvažovať o jej produktoch alebo službách. Cieľom marketingu v tejto fáze je zvýšiť pravdepodobnosť, že zákazník sa rozhodne pre nákup práve danej značky. V tomto štádiu sa potencionálni zákazníci snažia lepšie spoznať značku a zistiť, čo ju odlišuje od konkurencie. Marketingové aktivity sú, v tejto fáze, zamerané na riešenie problémov alebo zdôraznenie prínosov, aby pomohli lepšie pochopiť, ako ich produkt alebo služba napĺňa zákaznícke potreby (Coffman, 2022 a Amaresan, 2022).

Konverzia

Kirsch (2022) uvádza, že konverzná fáza, niekedy nazývaná aj fáza akcie, nastáva, keď zákazník uskutoční konverziu (väčšinou nákup) alebo uskutoční žiadaný cieľ. Táto etapa teda predstavuje úspešné dokončenie špecifických aktivít návštevníkmi stránky, prinášajúca pozitívny prínos pre hospodárske výsledky firmy. Vedľajšie konverzie sa nazývajú mikrokonverzie a je nimi napríklad prihlásenie sa na odber noviniek, stiahnutie aplikácie, vyplnenie formuláru. Nastavenie makro a mikrokonverzii je čisto na business požiadavkách daného podniku. Príklad marketingovej techniky na zlepšenie konverzného pomeru, pomeru počtu konverzii ku návštevníkom, je vo webovej analytike použitie A/B testovania.

Retencia (udržiavanie)

Etapa udržiavania zákazníkov je kľúčovou fázou v procese budovania vzťahu so zákazníkom. Cieľom tejto etapy je nielen udržať existujúcich zákazníkov, ale aj podnietiť ich k opakovaným nákupom a interakciám so značkou alebo podnikom. V rámci retencie by sa firma mala sústrediť na stratégie zamerané na zvyšovanie hodnoty zákazníka (*Customer Lifetime Value* - CLV), čo zahŕňa motiváciu zákazníkov k častejším nákupom alebo realizácii nákupov s vyššou hodnotou (Amaresan, 2022 a Morgan, 2023).

Lojalita

Fáza lojality, alebo vernosti, označuje štádium, kedy sa zákazník nielen opakovane vracia k nákupu, ale vďaka svojim pozitívnym skúsenostiam odporúča značku ďalej. V tejto etape sa firmy snažia upevniť vzťah so zákazníkom prostredníctvom personalizovaných ponúk a vernostných programov. Táto fáza tiež zahŕňa monitorovanie spokojnosti zákazníkov a zhromažďovanie ich spätnej väzby, čo umožňuje ďalšie vylepšenie produktov a služieb (Coffman, 2022 a Morgan, 2023).

3 NÁSTROJE A TECHNOLOGIE BI V CLOUDOVOM PROSTREDÍ

Táto kapitola sa zaoberá konkrétnymi nástrojmi a technológiami, ktoré sú použité v návrhovej a implementačnej časti diplomovej práce. V jednotlivých podkapitolách je vysvetlený ich princíp a štruktúrovanie dát. Na základe požiadaviek podniku a jeho dlhodobej stratégie boli vybrané nástroje, ktoré využívajú Google Cloud platformu. Na meranie používateľov a ich správania je použitý nástroj Google Analytics 4, pre dátový sklad je použitý nástroj BigQuery a na tvorbu reportingovej vrstvy riešenia Looker Studio.

3.1 Google Analytics 4

Google Analytics 4, obdoba staršej verzie Google Universal Analytics, je online služba od spoločnosti Google a je jedným z najznámejších nástrojov pre webovú analytiku. Google Analytics je široko využívaný najmä majiteľmi webových stránok, marketingovými a online agentúrami, ktoré používajú dáta pre vyhodnotenie svojich marketingových aktivít (Edmondson, 2022).

Podľa Brunec (2017) je základný princíp tohto systému: „Zbieranie analytických údajov z meraného webu, ich následnom spracovaní podľa nastavenia definovaného používateľom a finálnom zobrazení vybraných údajov”. Služba umožňuje porovnávať v čase údaje rôznych metrik vychádzajúcich z správania návštevníkov na webe.

V kontexte Business Intelligence, Google Analytics 4 predstavuje pokročilý nástroj pre zber a analýzu dát, ktorý umožňuje podnikom hlbšie pochopiť svojich zákazníkov, ich správanie na webe a dynamiku online trhu. Na základe týchto informácií môžu spoločnosti robiť informované a strategické obchodné rozhodnutia. Hlavným cieľom je získať komplexný prehľad o aktivitách na webe, konkrétne sledovať návštevnosť a správanie užívateľov počas daného časového intervalu (Kolektív autorov, 2014).

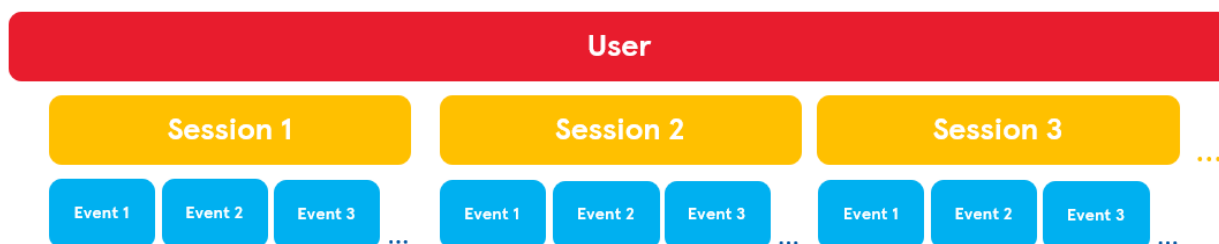
Vďaka verzii Google Analytics 4, ktorá vyšla v roku 2020, je možné sledovať a analyzovať dáta o používateľoch nielen na webovej stránke, ale aj v aplikácii a to prostredníctvom aplikačného dátového toku. Táto možnosť je dostupná prostredníctvom služby Firebase, ktorá a dá ľahko prepojiť s Google Analytics 4 (Burton, 2024).

Google Analytics 4 navyše umožňuje bezplatný export dát to prostredia cloudovej služby BigQuery, dodáva Pearce (2021), ktorá je vhodná na vyhodnocovanie nevzorkovaných dát a dáta nemajú časové obmedzenie ako v prostredí služby Google Analytics 4.

3.1.1 Dátový model

Dátový model Google Analytics 4 má takzvaný „*event-based*” formát zbierania dát. Tento model (z anglického *event*) zbiera dáta vo forme udalostí na webovej stránke alebo v aplikácii, ktoré sú na týchto platformách zaznamenané buď automaticky alebo vlastnou implementáciou. Každá udalosť má iné pomenovanie a môže predstavovať inú interakciu používateľa na webe, pričom môže mať veľké množstvo atribútov, nazývaných v prostredí tohto nástroja ako „parametre” (Mangold, 2021).

Aj keď je dátový model založený na udalostiach, najdôležitejšou zložkou modelu ostáva používateľ, ktorý je na vrchu hierarchie tohto modelu. Hierarchia je rozdelená do troch úrovní alebo entít, na ktorých čele je používateľ, ktorý spája všetky tieto úrovne. Ďalšími úrovňami sú relácia (z anglického *session*) a udalosť (z anglického *event*), uvádza Sopa (2023).



Obrázok 6 – Hierarchia entít v dátovom modeli GA4

(Zdroj: Vlastné spracovanie)

Každý používateľ môže mať priradený nekonečný počet relácií a každá z týchto relácií môže obsahovať veľké množstvo udalostí, ktoré sú spojené s danou reláciou. Relácia je definovaná ako nepretržitý úsek, ktorý používateľ strávil na webovej stránke, či v aplikácii a jej časové obmedzenie možno nastaviť v prostredí tohto nástroja, dodáva Sopa (2023).

Udalosti a ich druh rozdeľuje Google nástroj do štyroch kategórií: automatické udalosti, rozšírené, odporúčané a vlastné udalosti. Rozšírené udalosti je nutné zapnúť v nastaveniach ale je možné ich implementovať vlastným spôsobom. Odporúčané udalosti majú odporúčané parametre, jedná s najmä o ecommerce udalosti, ktoré majú produktové parametre, ktoré je pre vyhodnotenie v GA4 nutné posielat’ s preddefinovaným názvoslovím (Google Analytics Help, 2024). Základný prehľad a príklady týchto udalostí možno nájsť v tabuľke.

Tabuľka 3 – Prehľad základných udalostí a typov v GA4

(Zdroj: Vlastné spracovanie podľa: Google Analytics Help, 2024)

Názov udalosti	Druh udalosti	Podmienka spustenia
first_visit	automatická	Prvá návšteva webovej stránky.
first_open	automatická	Prvá návšteva aplikácie.
session_start	automatická	Keď používateľ začne novú reláciu.
page_view	rozšírená	Vždy, keď používateľ načíta webovú podstránku.
scroll	rozšírená	Používateľ prešiel 25%, 50%, 75% alebo 90% stránky.
user_engagement	rozšírená	Používateľ sa nachádzal na predchádzajúcej stránke aspoň 10 sekúnd.
add_to_cart	odporúčaná	Pridanie produktu do košíka.
form_submit	rozšírená	Používateľ vyplnil formulár.
click	rozšírená	Používateľ klikol na externý odkaz.
view_search_result	rozšírená	Používateľ zobrazil výsledky vyhľadávania.

3.1.2 Parametre jednotlivých entít

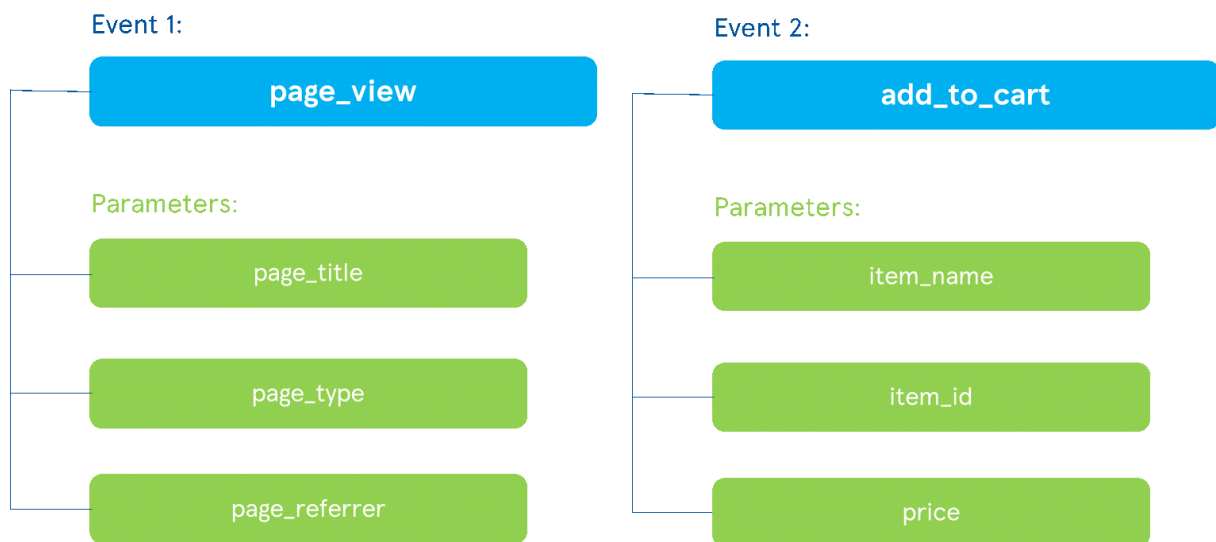
Parametre, ktoré možno zbierať prostredníctvom Google Analytics 4 môžu byť na úrovni používateľa alebo udalosti, pričom parameter na úrovni používateľa je nazývaný „*user property*“. Tento parameter je takmer nemenný, na rozdiel od parametra udalosti a nesie sa celou používateľskou cestou. Limit počtu parametrov, rovnako ako aj iných možností tohto nástroja je daná verziou, ktorú si používateľ tejto služby vyberie. Na výber je možnosť bezplatného nástroja, alebo platba či povýšenie vlastníctiev na verziu platenú, nazývanú Google Analytics 4 360 (Google Analytics Help, 2024).

Používateľské parametre

Používateľské parametre alebo „*user properties*” sa vzťahujú na samotného používateľa, môže sa jednať napríklad o používateľské ID, zdroj prvej návštevy (s informáciami napríklad o platenej reklame) či demografické a geografické parametre. Vďaka týmto parametrom možno vyhodnocovať zákazníka z dlhodobého hľadiska a zaradzovať ho do rôznych častí zákazníckeho cyklu (Sharma, 2024).

Parametre udalostí

Parametre udalostí Google rozdeľuje do skupiny automaticky generovaných, odporúčaných alebo vlastných. Automaticky generované parametre sú: *language*, *page_location*, *page_referrer*, *page_title*, *screen_resolution*. Pri odporúčaných udalostiach je možné posilať parametre s vopred definovaným názvom ako napríklad *item_id*, *item_name* alebo *price*, ako možno vidieť na obrázku 7, tie je následne možné vyhodnocovať v reportoch ako špeciálne produktové dimenzie či parametre (Google Analytics Help, 2024).



Obrázok 7 – Diagram udalostí a ich parametrov

(Zdroj: Vlastné spracovanie)

3.2 BigQuery

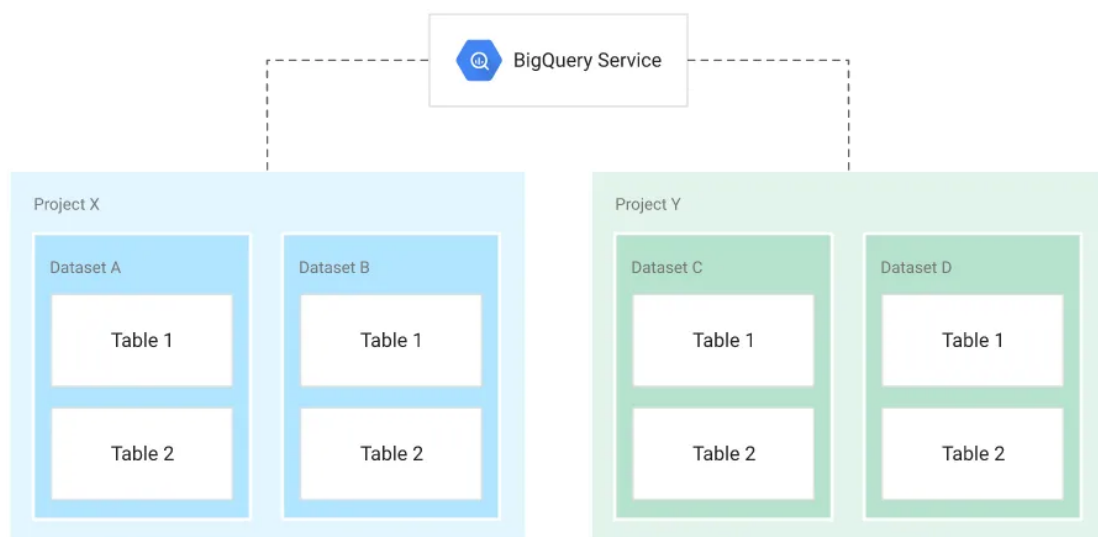
BigQuery je cloudový dátový sklad od spoločnosti Google, LLC so širokou škálou možností a integrovanou dotazovacou konzolou. Táto konzola je schopná spustiť SQL dotazy na terabajty dát za niekoľko sekúnd a petabajty za len niekoľko minút, tvrdí Google Cloud, (2024). BigQuery je súčasťou cloudových služieb Google Cloud Platform, preto je pomerne ľahké napojiť zdroje dát z Google primárnych zdrojov dát. Ako bolo spomínané

v predchádzajúcej kapitole, Google Analytics 4 je možné prepojiť s BigQuery a exportovať dáta bezplatne pri splnení určitých podmienok.

BigQuery umožňuje programátorom využívať knižnice pre jazyky C#, Go, Java, Node.js, PHP, Python a Ruby. Tieto knižnice poskytujú rôzne funkcie na prácu s dátami v BigQuery. Okrem toho, BigQuery poskytuje REST API, ktoré umožňuje programátorom komunikovať s touto službou pomocou HTTP požiadaviek. To znamená, že programátori môžu používať BigQuery bez ohľadu na to, aký programovací jazyk používajú (Google Cloud, 2024).

3.2.1 Štruktúra

BigQuery organizuje dátové tabuľky do jednotiek nazývaných datasety. Tieto datasety sú obmedzené na projekt v Google Cloud Platform (GCP). Rôzne úrovne, projekt, dataset a tabuľka, umožňujú logickú štruktúru informácií. Je možné vytvoriť viacero datasetov na jednu dátovú tabuľku, ktorá zahŕňa dáta z rozličných analytických oblastí. Navyše, pomocou úrovne projektu je možné oddeliť datasety podľa konkrétnych business potrieb (Thallam, 2020 a Bapna, 2019).



Obrázok 8 – Schéma štruktúry BigQuery

(Zdroj: Bapna, 2019)

Štruktúra cloudového nástroja BigQuery je rozdelená na niekoľko hlavných celkov, z ktorých každý plní určitú funkciu. Tieto jednotlivé celky sú navrhnuté tak, aby spolupracovali a umožnili efektívne spracovanie a analýzu veľkých objemov dát. Každý z týchto celkov má svoje špecifické vlastnosti a funkcie (Thallam, 2020). Tieto funkcie možno nájsť v tabuľke 4.

Tabuľka 4 – Prehľad významu kontrolných celkov v BigQuery

(Zdroj: Vlastné spracovanie podľa: Thallam, 2020)

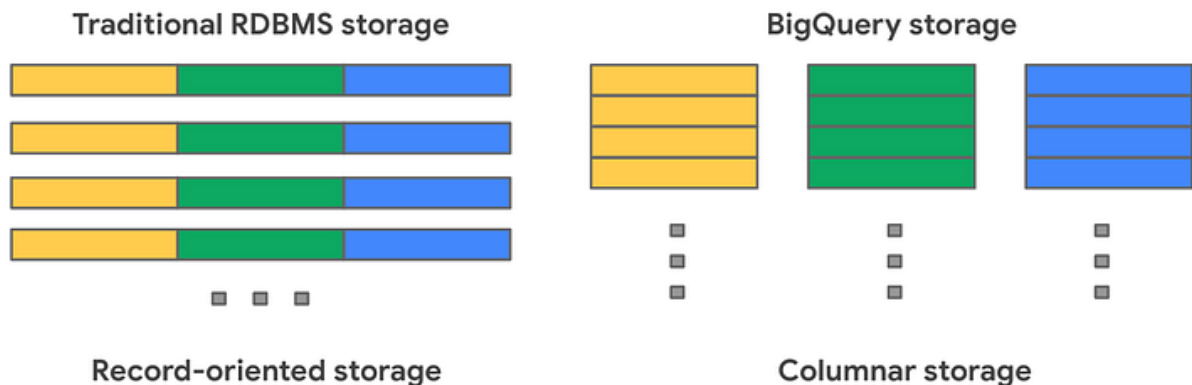
Projekty	
<ul style="list-style-type: none">• Sú základnou jednotkou a hlavným priestorom pre objekty ako dátové tabuľky, datasey, úlohy a iné.• Na tejto úrovni sa spravujú prístupy, užívateľské oprávnenia a role či samotná fakturácia.	
Dátové tabuľky	Datasey
<ul style="list-style-type: none">• Sú základné jednotky pre ukladanie dát v BigQuery. Každá tabuľka má definovanú schému, ktorá určuje typy dát v jednotlivých stĺpcoch.• Kolekcie stĺpcov a riadkov uložených v spravovanom úložisku.• Umožňujú kontrolu prístupu na úrovni tabuľky a stĺpca.	<ul style="list-style-type: none">• Sú kolekcie tabuliek, ktoré sú logicky pospájané popismi či štítkami.• Umožňujú kontrolu prístupu k úložisku na úrovni datasetu.• Definujú umiestnenie dát. viacregionálne alebo regionálne.
Pohľady	Úlohy
<ul style="list-style-type: none">• Virtuálne tabuľky definované SQL dotazom.• Pohľady neukladajú žiadne dáta, ale poskytujú štruktúrovaný pohľad na dáta založený na dotaze, ktorým sú definované.	<ul style="list-style-type: none">• Predstavujú akcie vykonané v BigQuery v mene užívateľa - načítanie dát, export dát, kopírovanie dát alebo dotazovanie dát.• Úlohy sa vykonávajú asynchrónne, môžete pokračovať v práci, zatiaľ čo BigQuery vykonáva tieto úlohy na pozadí.

V prípade odkazovania na tabuľku v príkazovom riadku, v SQL dotazoch alebo v kóde, konštrukcia názvoslovia a odkazovanie je podľa BigQuery štruktúry nasledujúca: *projekt.dataset.tabuľka*.

3.2.2 Ukladanie a nahrávanie dát

Tradičné relačné databázy, ako je napríklad MySQL, ukladajú dáta formou riadkov (sú orientované na záznamy). Naproti tomu, BigQuery využíva stĺpcové úložisko, pri ktorom je každý stĺpec uložený v samostatnom bloku súborov. Táto metóda je optimálna pre scenáre použitia typu OLAP (Online Analytical Processing), teda pre online analytické spracovanie.

Do tabuliek v BigQuery možno dáta jednoducho pridávať, aktualizovať ich alebo vymazávať existujúce hodnoty. Podporujú sa aj operácie zmeny dát (INSERT, UPDATE, MERGE, DELETE) a to bez akýchkoľvek obmedzení (Thallam, 2020).



Obrázok 9 – Diagram ukladania dát v relačných databázach a BigQuery

(Zdroj: Thallam, 2020)

BigQuery využíva pokročilé techniky a variácie stĺpcového úložiska, uvádza Thallam (2020), dáta ukladá vo vlastnom formáte, ktorý sa nazýva *Capacitor*. Tento formát prináša niekoľko výhod pre spracovanie veľkých objemov dát v dátových skladoch. Vďaka použitiu vlastného formátu môže BigQuery synchronizovať vývoj úložného systému s vývojom systému na spracovanie dotazov. Každý stĺpec v tabuľke je uložený v samostatnom bloku súborov a všetky stĺpce sú uložené v jednom súbore *Capacitor*, ktorý je komprimovaný a šifrovaný na disku.

Thallam (2020) a Faraz (2020) uvádzajú, že existuje niekoľko spôsobov, ako nahráť dáta do BigQuery v závislosti od zdrojov dát, formátov dát, metód nahrávania a prípadov použitia. Dáta možno nahráť cez dávkové nahrávanie (*batch ingestion*), streamovaním, prenosom dát (*data transfer service*), ukladaním výsledkov dotazov alebo integráciami rôznych partnerov. Azda najznámejšími konkrétnymi spôsobmi, ale sú cez:

- **CSV súbor.** Tento proces sa realizuje cez funkciu nahrávania pri vytváraní tabuľky. Stačí vybrať CSV súbor a BigQuery automaticky určí štruktúru tabuľky.
- **JSON súbor.** Postup je veľmi podobný ako pri CSV súboroch, avšak pri výbere formátu súboru sa zvolí JSON.
- **Google Cloud Storage,** ktorý je bezpečnou online službou pre ukladanie a prenos dát. Google Cloud Storage podporuje rôzne formáty súborov, takže možno nahráť dáta, ktoré najlepšie vyhovujú potrebám podniku.
- **Data Transfer Service,** ktorá umožňuje automatické nahrávanie dát z rôznych služieb Google, ako sú Google Ads alebo YouTube.

- **Cloud Client Libraries.** Pomocou týchto knižníc a API možno jednoducho sťahovať dáta z BigQuery priamo do potrebnej aplikácie v preferovanom jazyku.

Pri nahrávaní dát do BigQuery je poskytnutá možnosť vytvoriť novú tabuľku, pridať dáta do existujúcej tabuľky alebo prepísať existujúcu tabuľku, je však potrebné definovať schému tabuľky alebo jej časti. Pre podporované formáty dát je k dispozícii možnosť automatickej detekcie schémy (Thallam, 2020 a Faraz, 2020).

3.2.3 Plánované dotazy

Plánované dotazy (z anglického *scheduled queries*) sú možnosťou pre automatizáciu rutinných operácií s dátami v prostredí BigQuery. Silva (2023) opisuje, že používatelia si jednoducho vytvoria dotaz SQL, nastaví požadovaný interval (denne, týždenne alebo vlastný interval) a určia súbor dát, kde by mali byť výsledky uložené. BigQuery následne vykoná dotaz podľa definovaných požiadaviek. Zároveň tvrdí, že plánované dotazy sú užitočným nástrojom pre automatizáciu úloh, transformáciu dát, zníženie nákladov či údržbové operácie.

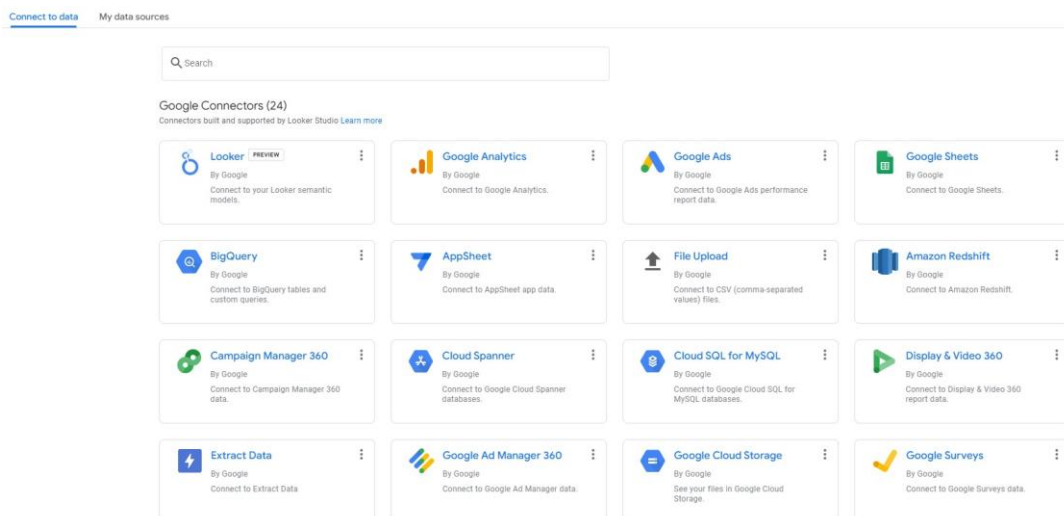
Obrázok 10 – Plánované dotazy a ich tvorba v BigQuery

(Zdroj: Vlastné spracovanie podľa: Google Cloud, 2024)

Naplánované dotazy a ich požiadavky definuje (Google Cloud, 2024) ako napísané v GoogleSQL a môže obsahovať príkazy jazyka pre definíciu dát (DDL) a jazyka pre manipuláciu s dátami (DML). Nutný je špeciálny typ povolenia pre druh takýchto operácií. Taktiež je nutné vyplniť *názov plánovaného dotazu, interval opakovania, dátum začiatku a čas spustenia, dátum ukončenia, destináciu a názov tabuľky, a nakoniec aj preferenciu zápisu* (prepísať alebo pridať). Príkaz „prepísať“ – znamená, že výsledky dotazu prepíšu dáta v tabuľke a „pridať“ – znamená, že výsledky dotazu budú pridané k dátam v tabuľke (Saha, 2023).

3.3 Looker Studio

Looker Studio, predtým nazývané Data Studio, je vizualizačný nástroj od spoločnosti Google, LLC pre bezplatné vizualizácie a zobrazenie dát. Ponúka niekoľko bezplatných konektorov, ktoré pomáhajú napojiť dátové zdroje do niekoľkých sekúnd. Taktiež podporuje širokú škálu konektorov tretích strán, ktoré pomôžu vizualizovať dáta z takmer každej marketingovej platformy (Nyberg, 2022).



Obrázok 11 – Prehľad zabudovaných konektorov v Looker Studiu

(Zdroj: Vlastné spracovanie)

Tento Business Intelligence nástroj využíva vlastný deklaratívny modelovací jazyk s názvom LookML, ktorý je odvodený z SQL. Okrem toho Looker poskytuje API (z anglického *Application Programming Interface*), ktoré umožňuje spojenie s inými aplikáciami a službami. Toto API umožňuje prístup k rôznym funkcionalitám Lookeru, vrátane manipulácie s dátami, tvorby a spustenia dotazov, správy užívateľských práv a mnohých ďalších. Princíp fungovania Lookeru je postavený na modelovaní dát. V LookML sa definujú modely, ktoré

mapujú štruktúru potrebných dát a vzťahy medzi nimi. Tieto modely potom slúžia ako základ pre vytváranie vizualizácií a pohľadov na dáta (Google Cloud, 2024). Looker tiež umožňuje vytváranie dashboardov a reportov, ktoré poskytujú jednoduchý prehľad o metrikách a KPI (z anglického *Key Performance Indicators*).

Looker ponúka robustné možnosti vizualizácie dát priamo v rámci svojej platformy. Umožňuje vytvárať rôzne typy grafov a diagramov, vrátane stĺpcových, riadkových, koláčových grafov, máp a ďalších, spomína (Nyberg, 2022). Užívateľ môže jednoducho vybrať požadovaný typ vizualizácie a priradiť k nemu relevantný dátový zdroj a konkrétne dáta. Tie možno zobrazovať v prostredí interaktívnych dashboardov, ktoré môžu obsahovať viaceré vizualizácie a poskytovať tak komplexný prehľad o dátach. Dashboardy sa dajú prispôbiť podľa potrieb užívateľa a je možné filtrovanie do detailov dát na základe rolovacieho filtra či iného filtra. Všetky tieto vizualizácie môžu byť zdieľané medzi používateľmi služieb Google alebo je možné ich exportovať pre ďalšie použitie, napríklad v reportoch alebo prezentáciách (Mangold, 2023).

4 ANALÝZA SÚČASNÉHO STAVU

V tejto kapitole bude predstavená firma, ktorá disponuje webovou stránkou a aplikáciou, ktoré zbierajú dáta, ktoré budú obsahom navrhovaného a implementovaného dátového skladu. Firma disponuje okrem online predajom aj predajom v kamenných predajniach, Avšak, dáta z týchto predajní nie sú súčasťou webovej analytiky a nebudú preto zahrnuté do tohto dátového skladu. Táto časť slúži ako základ pre navrhnutie vhodnej schémy dátového skladu v kapitole návrhu a implementácie riešenia.

4.1 Firma

Spoločnosť, o ktorej sa bude pojednávať v tejto práci, je obchodným reťazcom. Z dôvodu ochrany daného subjektu nebude v tejto práci spoločnosť explicitne menovaná a všetky údaje, ktoré by naznačili jej identitu budú zakryté. Táto firma sa postupne rozrástla a rozšírila svoju pôsobnosť do mnohých krajín po celom svete. Jej hlavnou činnosťou je prevádzkovanie samoobslužných obchodov, ktoré ponúkajú široký sortiment produktov, od potravinárskeho tovaru až po spotrebnú elektroniku a oblečenie (Webová stránka firmy, 2024).

V rámci svojho podnikania sa táto spoločnosť zameriava na poskytovanie kvalitných produktov za konkurencieschopné ceny, čo dosahuje vďaka efektívnej logistike, hromadnému nákupu a vlastným značkám produktov. Spoločnosť kladie dôraz spoločenskú zodpovednosť, čo sa prejavuje v jej prístupe k životnému prostrediu, etickom zdroji tovaru a podpore komunitných projektov (Interná dokumentácia, 2024).

Firma je známa svojím lojálnym programom, ktorý zákazníkom ponúka výhody a zľavy, a tiež svojimi iniciatívami zameranými na zlepšenie zákazníckeho zážitku, ako sú samoobslužné pokladne alebo online nakupovanie prostredníctvom webovej stránky. Táto webová stránka ponúka okrem prebiehajúcich akcií a sortimentu ponúkaného v predajni aj informácie o tipoch a trikoch pre využitie tovaru, ktorý si zákazník môže nakúpiť. Jej úspech je zakladaný na stratégii dlhodobého rastu, inovácií a adaptácie na meniace sa trhové podmienky, čo jej umožňuje udržať si silnú pozíciu na trhu aj v tvári intenzívnej konkurencie. (Interná dokumentácia, 2024).

4.2 Digitálne platformy

Webová stránka a mobilná aplikácia firmy sú navrhnuté tak, aby poskytovali užívateľsky prívetivé prostredie a jednoduchú navigáciu. Obe platformy sú štruktúrované do logických sekcií, vrátane domovskej stránky s aktuálnymi ponukami, sekcie s akčnými ponukami a kupónmi, produktových podstránok s detailnými informáciami a košíka pre správu vybraných položiek.

Dizajn oboch platforiem je zameraný na atraktivnosť a intuitívnosť, čo umožňuje zákazníkom ľahko nájsť požadované produkty alebo informácie a efektívne realizovať nákup. Analytické údaje z týchto platforiem poskytujú cenné informácie o správaní sa zákazníkov, ktoré môžu byť využité na optimalizáciu služieb a zlepšenie zákazníckej skúsenosti. Aktualizácia a optimalizácia týchto platforiem pre rôzne zariadenia, najmä pre mobilné, je nevyhnutná pre dosiahnutie maximálnej efektívnosti (Kaushik, 2009).

4.2.1 Webová stránka

Webovú stránku firmy môžeme rozdeliť do niekoľkých sekcií ktoré sú navrhnuté tak, aby uľahčili navigáciu a zlepšili zákaznícku skúsenosť.

- **Domovská stránka** webovej platformy poskytuje prehľadný a užívateľsky prívetivý dizajn, ktorý zákazníkov víta s aktuálnymi ponukami, zľavami a propagačnými akciami.
- **Sekcia s akčnými ponukami:** V tejto sekcii možno nájsť ponuky na nasledujúci týždeň, produkty v zľave a súťaže či akcie. Pri každom produkte sa nachádza link pre jeho zakúpenie.
- **Hlavné kategórie produktov:** Na webovej stránke sú produkty organizované do logických kategórií, Každá kategória je ďalej rozčlenená na podkategórie, čo umožňuje zákazníkom ľahko nájsť konkrétny typ produktu.
- **Produktové podstránky:** Každý produkt má samostatnú podstránku s detailnými informáciami o produkte. Každá podstránka obvykle obsahuje obrázky produktu, podrobný popis, cenu a informácie o dostupnosti. Nachádza sa tu možnosť pridať produkt do košíka alebo ho uložiť do zoznamu obľúbených produktov.

- **Košík a objednávkový formulár:** Sekcia košíka na webovej stránke je podstránka, kde zákazníci môžu prezerat' a upravovat' položky, ktoré si vybrali na nákup. Keď zákazník pridá produkt do košíka, tento sa automaticky aktualizuje a zobrazí počet položiek a celkovú cenu nákupu. Zákazníci môžu ľahko pridat' alebo odstrániť položky, meniť ich množstvo a vidieť jednotlivé ceny produktov aj celkovú sumu nákupu. V košíku je tiež možné uplatniť rôzne zľavy alebo použiť kódy z vernostných programov. Po revízii a potvrdení obsahu košíka môže zákazník pokračovat' k pokladni a dokončiť proces nákupu v objednávkovom formulári. V sekcii objednávkového formulára na webovej stránke zákazníci zadávajú svoje kontaktné a dodacie údaje, vyberajú preferovaný spôsob platby a doručenia. Po vyplnení všetkých potrebných informácií môže zákazník skontrolovať a potvrdiť svoju objednávku.
- **Vyhľadávacia lišta:** Umiestnená v hornej časti stránky, vyhľadávacia lišta umožňuje užívateľom rýchlo vyhľadať produkty alebo informácie pomocou kľúčových slov.
- **Bannery:** Dynamické bannery zobrazujúce najnovšie kampane a špeciálne ponuky sa pravidelne obmieňajú na hlavnej stránke a upútavajú pozornosť zákazníkov.
- **Informačné sekcie:** Webová stránka tiež obsahuje sekcie s informáciami o spoločnosti, vrátane jej histórie, hodnôt a spoločenskej zodpovednosti. Ďalej sú tu uvedené informácie o službách zákazníkom, vrátane možností doručenia, vrátenia tovaru a kontaktných údajov.
- **Vernostný program:** Oddiel venovaný vernostnému programu poskytuje informácie o výhodách, ktoré môžu zákazníci získať pri nákupe, a o tom, ako sa môžu do programu zapojiť.
- **Prihlasovacia sekcia:** Užívatelia majú možnosť vytvoriť si účet alebo sa prihlásiť do existujúceho, čo im umožňuje sledovať históriu objednávok, spravovat' osobné údaje a využívať personalizované ponuky, ktoré sú súčasťou spomínaného vernostného programu.
- **Sekcia pre mobilné aplikácie:** Webová stránka, ktorá propaguje mobilné aplikácie, ktoré zákazníkom umožňujú pohodlné nakupovanie a prístup k službám aj na cestách.
- **Nápoveda a podpora:** Sekcia so často kladenými otázkami a podpornými materiálmi pomáha riešiť bežné problémy a poskytuje užitočné rady pre zákazníkov.

- **Pätička stránky:** V dolnej časti stránky sa nachádzajú odkazy na právne informácie, zásady ochrany osobných údajov, podmienky používania stránky a ďalšie dôležité dokumenty.

Celkovo je webová stránka navrhnutá tak, aby poskytovala bohatý a zároveň prehľadný obsah, ktorý zákazníkom umožňuje ľahko sa orientovať a efektívne hľadať potrebné informácie či uskutočňovať nákupy. Náhľad štruktúry stránky možno nájsť na obrázku 12.



Obrázok 12 – Wireframe webovej stránky firmy

(Zdroj: Vlastné spracovanie)

4.2.2 Aplikácia

Aplikácia je navrhnutá, aby bola rovnako ako web, používateľsky prívetivá a aby sa v nej používateľ nestratil. Dizajn je obdobný ako pri webovej stránke a sekcie sú rozdelené podobnou logikou.

- **Domov:** Hneď v úvodnej sekcii sú zákazníkom prezentované atraktívne zľavy a špeciálne ponuky, ktoré môžu využiť pri svojom nákupe. Tieto ponuky sú obvykle zvýraznené a sú doplnené o vizuálne prvky, ako sú grafiky alebo animácie, aby pritiahli pozornosť užívateľa. Okrem aktuálnych zliav sa tu nachádza aj výber

hlavných kategórií, ktorý je umiestnený v prehľadnej forme, vo forme zoznamu, umožňujúci zákazníkovi rýchlo sa orientovať a vybrať potrebné produkty. V tejto sekcii môžu zákazníci jednoducho pristupovať k nastaveniam svojho účtu, či už ide o kontrolu vernostných bodov, zmenu osobných údajov alebo nastavenia notifikácií. Nachádza sa tu aj pohyblivá bannerová oblasť, ktorá pravidelne zobrazuje rôzne aktuálne kampane, novinky alebo informácie s pridanou hodnotou.

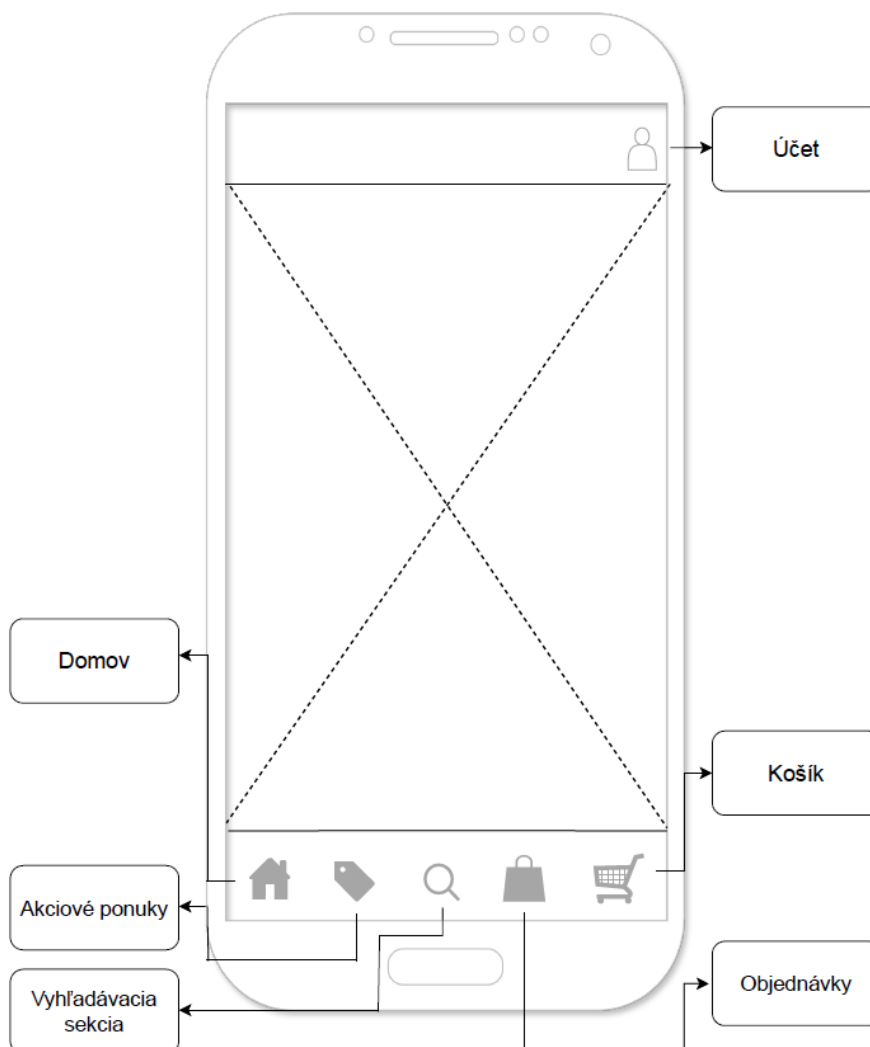
- **Akčné ponuky a kupóny:** Sekcia je určená pre správu a uplatnenie kupónov, ktoré aplikácia ponúka. Užívatelia môžu svoje kupóny aktivovať priamo v aplikácii alebo v kamennej predajni a tak okamžite získať zľavy na vybrané produkty alebo celú objednávku. Taktiež sa tu nachádza aktuálna ponuka zľavnených výrobkov vo forme letáku, ktorý možno jednoducho prelistovať.
- **Vyhľadávacia sekcia:** Vyhľadávanie, ktoré je prístupné pomocou vyhľadávacej lišty umožňuje zákazníkovi hľadať konkrétne produkty, kategórie alebo informácie na stránke. Výsledky vyhľadávania sú zobrazené pod sebou a zákazník môže produkt jednoducho pridať do košíka alebo preskúmať ďalšie podrobnosti.
- **Objednávky:** Táto sekcia je určená pre správu a prehľad predchádzajúcich a aktuálnych objednávok užívateľa. Zákazníci tu môžu sledovať stav svojej objednávky, zobraziť si históriu nákupov a získať podrobnosti o každej objednávke, vrátane jednotlivých položiek, ceny a stavu doručenia.
- **Košík:** Táto sekcia ukazuje počet položiek a celkovú cenu vybraných produktov potencionálnej objednávky, ktoré si zákazník vybral akčným tlačidlom „pridať do košíka“. V tejto sekcii môžu zákazníci upravovať svoj výber, meniť množstvo produktov, alebo odstrániť položky, ktoré si kúpiť neprajú. Po skontrolovaní obsahu košíka môžu zákazníci prejsť k pokladni a dokončiť objednávku. Pre dokončenie objednávky je nutné sa prihlásiť alebo registrovať.

Stránky alebo sekcie s akčnými ponukami a produktové stránky s nákupným košíkom a procesom dokončenia nákupu (z anglického názvu *checkout*) predstavujú kľúčové sekcie pre úspech firmy a jej online obchodné aktivity. Tieto stránky sú navrhnuté tak, aby upútali pozornosť a prilákali zákazníkov prostredníctvom atraktívnych zliav a špeciálnych akcií, čo je základom pre generovanie predaja a zvyšovanie zákaznickej angažovanosti (Interná dokumentácia, 2024).

Produktové stránky poskytujú podrobné informácie o jednotlivých položkách, vrátane popisu, ceny a dostupnosti, pričom sú navrhnuté pre jednoduchú navigáciu a intuitívne vedenie zákazníka k nákupu. Nákupný košík umožňuje užívateľom pohodlne pridávať a upravovať

vybrané produkty, zatiaľ čo proces dokončenia nákupu, známy aj ako „checkout“, je rovnako kritickou komponentou. Tento proces je navrhnutý s ohľadom na užívateľa (*user-friendly*) a zabezpečuje hladký prechod od výberu produktov k ich finálnej kúpe produktov.

Význam týchto stránok pre podnik spočíva v ich priamom vplyve na konverzie a ich množstvo a teda celkovú efektívnosť online predaja, ktorú je možno pozorovať na nastavených kľúčových ukazovateľoch výkonnosti danej spoločnosti. Analytika na týchto stránkach predstavuje najdôležitejšie nápovede o chovaní sa zákazníka v rámci zákazníckej ceste. Preto je dôležité, aby boli tieto stránky pravidelne aktualizované a optimalizované pre rôzne zariadenia, a to najmä pre mobilné zariadenia, aby sa zabezpečila maximálna efektívnosť a zákaznícka spokojnosť (Matomo, 2024).



Obrázok 13 – Wireframe aplikácie firmy

(Zdroj: Vlastné spracovanie)

4.3 Požiadavky na riešenie

Pre navrhované riešenie boli stanovené požiadavky, ktoré zahŕňajú sledovanie metrík a rôznych ukazovateľov, ktoré sa týkajú výkonnosti webovej stránky a aplikácie, ako aj analýzu správania jej používateľov. Požiadavky sú odvodené zo základných princípov analýzy životného cyklu zákazníka, definovaného v súlade s obchodnými potrebami spoločnosti. V súčasnosti sa spoločnosť intenzívne venuje digitalizácii svojich služieb a aktívne zdokonaľuje svoje online marketingové aktivity. Tieto kroky sú neoddeliteľnou súčasťou stratégie zameranej na posilnenie dostupnosti služieb a budovanie lojality u zákazníkov. Jej cieľom je posilniť značku online, a to prostredníctvom inovatívnych digitálnych iniciatív, ktoré pomôžu vylepšiť konkurencieschopnosť a zvýšiť povedomie o značke medzi zákazníkmi.

Prvou metrikou, ktorá bude jedna zo zásadných ukazovateľov je **počet návštevníkov**. Počet návštevníkov či používateľov je nutné vedieť agregovať aj podľa zapojenia, to znamená vedieť zaradiť pod atribút nový návštevník alebo vracajúci sa návštevník, t.j. ktorý prichádza na web opakovane. Sledovanie počtu návštevníkov a pozorovanie jeho trendov v priebehu času v podobe denného priemeru a aj ako kumulatívneho súčtu za stanovené časové obdobie pomáha, podľa Foy (2022), rozumieť a posúdiť účinnosť marketingových kampaní, účinnosť optimalizácie obsahu a zlepšenie užívateľskej skúsenosti.

Druhá metrika, ktorá je nevyhnutná pre určenie dosiahnutia firemných cieľov prostredníctvom online marketingu, spočíva v sledovaní **počtu nákupov**. Tento ukazovateľ predstavuje jeden z kľúčových faktorov v oblasti marketingu a predaja, umožňujúci firme sledovať zákaznícky dopyt a jeho vývoj v čase. Využitie tejto informácie môže mať významný vplyv na ziskovosť firmy. Identifikácia faktorov, ktoré prispievajú k zvyšovaniu počtu nákupov, umožňuje firme tieto faktory ďalej rozvíjať a zlepšovať, čím sa zvyšuje ich ziskovosť. Aby firma bola schopná úspešne vyhodnotiť svoju výkonnosť v oblasti online marketingu, je nutné vedieť agregovať počet nákupov aj podľa počtu produktov v objednávke, ich kategóriu, cenu, ID či názov produktu. Taktiež dôležité je sledovať, ktorý marketingový kanál je zdrojom pre túto konverziu. Tento postup umožňuje firme vyhodnotiť úspešnosť jednotlivých produktov a akcií, a tým lepšie plánovať svoje budúce marketingové stratégie (Foy, 2022). Porovnanie prvej a druhej metriky pomôže ukázať ďalší dôležitý ukazovateľ, ktorým je **konverzná miera**. Tento ukazovateľ ukazuje percento návštevníkov webovej stránky, ktorí sa stali zákazníkmi alebo vykonali požadovanú akciu, ktorá je nákup.

Tretou požiadavkou pre meranie sú **tržby**. Sledovanie tržieb v ecommerce je považované za nevyhnutné z dôvodu hodnotenia výkonnosti podniku, plánovania a predpovedania budúcich výsledkov. Porovnávanie s konkurenciou môže byť realizované prostredníctvom tržieb, čo umožňuje identifikáciu oblastí na zlepšenie. Identifikácia trendov v predajoch môže byť dosiahnutá sledovaním tržieb, čo naznačuje sezónne trendy a príležitosti na zvýšenie predajov. Efektivita marketingových kampaní môže byť tiež monitorovaná prostredníctvom sledovania tržieb, ktoré poskytujú indikáciu úspešnosti kampane. Tržby je opäť nutné vedieť vyhodnotiť na top produkty, čiže také, ktoré sú najviac predávané a počet nákupov, ktoré dané tržby priniesli (Foy, 2022).

Priemerná hodnota objednávky je jedným z ďalších kľúčových ukazovateľov v ecommerce, pretože poskytuje hlbší pohľad na správanie zákazníkov a efektivitu stratégií cenotvorby. Táto metrika bude taktiež potrebná pre správne vyhodnotenie marketingových aktivít.

Zapojenie užívateľa na webe je kľúčové pre úspech v ecommerce. Podporuje zákaznícku lojalitu, motivuje k opakovaným nákupom a vedie k stabilnejšiemu príjmu. Zapojení užívatelia sú ochotnejší poskytnúť spätnú väzbu a zdieľať svoje skúsenosti, čo pomáha pri vylepšovaní produktov a služieb. Navyše, zapojenie často vedie k vyššej miere konverzie a lepšiemu hodnoteniu vo vyhľadávačoch. Preto je sledovanie a optimalizácia zapojenia užívateľa neoddeliteľnou súčasťou stratégie ecommerce firmy (Kaushik, 2009).

Jedným z dôležitých metrik zapojenia je **čas, počas ktorého sú používatelia aktívni na webe či v aplikácii**. Dá sa sledovať v agregovanej podobe ako priemerná hodnota za celú webovú stránku v rámci určitého časového intervalu alebo na jednotlivé stránky či sekcie.

Ďalšia dôležitá informácia poskytuje **návštevnosť jednotlivých podstránok webovej stránky alebo sekcií v aplikácii**. Ktorá z týchto podstránok či sekcií bola najviac navštevovaná či ktorá bola vstupnou stránkou na ktorej používateľ inicioval svoj príchod, taktiež nazývaná ako *Landing Page*. Pri obidvoch dimenziách je možné vďaka Google Analytics 4 vedieť rozlíšiť názov stránky (*Page Title*) či predchádzajúcu stránku (*Page Referrer*) a mnoho iných dát, ktoré pomáhajú detailnejšej analýze.

Pre vybranú firmu je nutné vedieť vyhodnotiť niekoľko ciest používateľov skrz čiastkové ciele až po vybranú konverziu. Sledovanie priebehu pre hlavnú konverziu, nákup, nazývané tiež ako **nákupný lievik**, poskytuje značnú výpovednú hodnotu pre vyhodnotenie marketingových aktivít. Predajným lievikom je proces návštevy stránky, pozretia si produktu

a pridania si produktu do košíka prostredníctvom CTA tlačidla. Následne prechod objednávkovým formulárom až po najdôležitejší cieľ, ktorý je nákup alebo objednávka. Objednávkový proces je rozdelený do niekoľkých krokov s ďalšími čiastkovými cieľmi a pre dokončenie objednávkového procesu je nutné sa prihlásiť alebo registrovať. Tento proces nastáva aj na webovej stránke aj v aplikácii.

Ďalší lievik, ktorý je potrebné odmerať je **prechod registračným formulárom pre prihlásenie do vernostného programu**. Používateľ je najprv uvítaný s možnosťou zapojiť sa do vernostného programu prostredníctvom rôznych komunikačných kanálov. Pre registráciu je vytvorená špecifická podstránka na webovej stránke a v mobilnej aplikácii sekcia, ktorá poskytuje informácie, čo toto členstvo ponúka. Tu je používateľovi poskytnutý registračný formulár, kde sú vyžadované určité údaje. Po úspešnom vyplnení a odoslaní formulára je používateľ registrovaný do vernostného programu, ktorý je hlavným cieľom tohto lievika.

Menej dôležitý funnel je **návšteva akčných ponúk od prvej strany ponuky až po poslednú**. Používateľ je najprv upozornený na existenciu akčných ponúk prostredníctvom hlavnej stránky webovej stránky alebo v aplikácii. Po kliknutí na tieto ponuky je zákazník presmerovaný na podstránku s detailmi o danej akcii. Tu má možnosť prechádzať medzi rôznymi produktmi v rámci ponuky.

Všetky tieto metriky a ukazovatele je nutné vedieť monitorovať kumulatívne ako priemernú hodnotu počas určeného časového intervalu a jeho progresiu v priebehu času. Okrem spomenutého je nevyhnutné zapracovať technické, geografické a marketingové dimenzie. S týmito aspektmi bude možné sledovať distribúciu používateľov podľa jednotlivých zariadení, oblastí alebo posudzovať, ktoré platformy privádzajú na web alebo do aplikácie najväčší počet používateľov, nákupov, prípadne určiť, ktoré produkty prinášajú najväčšie tržby.

Všetky tieto požiadavky je nevyhnutné brať do úvahy pri koncipovaní a vytváraní dátového úložiska. Novo vytvorené dátové centrum by malo zahrňovať historické údaje minimálne od začiatku roka 2024.

4.4 Súčasný stav

V súčasnosti je v prevádzke plne funkčná produkčná verzia webovej stránky a natívnej aplikácie, pre ktoré prebieha meranie a zber dát prostredníctvom Google Analytics 4 360. Meranie aplikačných dát je nastavené skrz Firebase a prepojenie s Google Analytics 4. Obidva dátové toky sú smerované do toho istého účtu, ktorý je platenou verziou Google Analytics 4 - Google Analytics 4 360.

Paralelne bol vytvorený projekt v rámci prostredia BigQuery, kde sú Google Analytics 4 dátové toky integrované a je nastavený export dát, či už na dennej alebo nepretržitej báze. Google Analytics 4 je primárnym zdrojom dát (DA), dočasné úložisko (DSA) a zároveň dátový sklad sa nachádza v prostredí BigQuery. Tieto komponenty poslúžia ako základ pre riešenie návrhu dátového skladu, ktorý spĺňa požiadavky spomínané v predchádzajúcej kapitole.

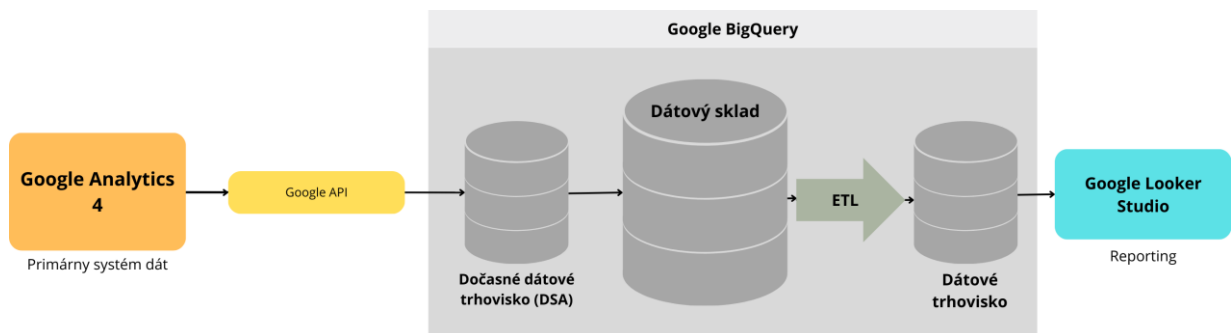
Cieľom riešenia tejto diplomovej práce je teda navrhnúť spôsoby ďalšieho spracovania a ukladania dát na základe vyššie popísaných požiadaviek. Pod spracovaním a ukladaním dát sa rozumie návrh a implementácia dátového trhoviska a dátových púmp (ETL procesy). Následne je potrebné novo vytvorené dátové trhovisko pripojiť do vizualizačného nástroja a vytvoriť reportovací dashboard, ktorý odráža vyššie uvedené požiadavky na riešenie.

5 VLASTNÝ NÁVRH RIEŠENIA

Nasledujúce podkapitoly sa budú zaoberať návrhom štruktúry dátového skladu pre BI riešenie webovej stránky a aplikácie firmy, ktorá je predmetom tejto práce. Prvá podkapitola poskytuje prehľad o základnej štruktúre dátového skladu. V ďalších podkapitolách je tento návrh podrobne rozpracovaný pomocou konceptuálneho a logického modelu. Tento návrh bude slúžiť ako východiskový bod pre implementáciu samotného riešenia, ktorá bude opísaná v nasledujúcej kapitole.

5.1 Návrh architektúry dátového skladu

Navrhovaná architektúra štruktúry dátového skladu je tvorená nasledujúcimi komponentami, ktoré sú znázornené na schéme nižšie:



Obrázok 14 – Návrh dátového skladu

(Zdroj: Vlastné spracovanie)

Prvky, ktoré tvoria základ systému a nie sú priamo súčasťou novej implementácie, zahŕňajú Google Analytics 4 ako primárny systém a zdroj dát. Ďalšie prvky, ktoré sa už nachádzajú v prostredí BigQuery sú dočasné úložisko dát, ktoré čerpá dáta skrz Google API (*Application Programming Interface*), a samotný dátový sklad, ktorý uchováva historické dáta. Pravá strana schémy zobrazuje prvky, ktoré boli navrhnuté a implementované v rámci tejto diplomovej práce. Medzi tieto prvky patria **ETL procesy, dátové trhovisko a vrstva pre reportovanie v nástroji Looker Studio.**

V nasledujúcich podkapitolách budú vysvetlené všetky časti riešenia, ktoré sú súčasťou pôvodného stavu či novej implementácie, pretože všetky tieto komponenty a prvky spolu tvoria celkové navrhované riešenie.

5.1.1 Primárny zdroj dát

Primárnym zdrojom dát je Google Analytics 4, ktorý tvorí základ pre meranie dát o užívateľoch a ich správaní na webovej stránke či v aplikácii. Implementácia tohto nástroja bola špecificky navrhnutá a prispôbena business požiadavkám tejto firmy. Navyše firma využíva platenú verziu tohto nástroja Google Analytics 4 360, ktorá má menej limitované funkcie.

Samotná služba Google Analytics 4 poskytuje možnosť vyhodnotenia si rôznych metrik a parametrov, avšak pri veľkých objemoch dát, čo je prípadom spomínaného podniku, sa možno stretnúť so vzorkovaním a domodelovanými dátami. Tento nástroj preto umožňuje export dát do BigQuery. V rámci integrácie Google Analytics 4 (GA4) a BigQuery je dôležité zdôrazniť, že dáta exportované do BigQuery sú nevzorkované a uchovávané v ich surovej (raw) forme. Tento aspekt je kritický, najmä keď sa zvažuje rozsah a presnosť dátových analýz.

Pred začatím exportu je nevyhnutné mať vytvorený projekt v Google Cloud Platform, kde bude BigQuery služba použitá a projekt musí mať povolené BigQuery API. Užívateľ určí frekvenciu exportu. Export dát z Google Analytics 4 (GA4) do BigQuery môže prebiehať v dvoch základných režimoch: buď ako denný export alebo ako priebežný (real-time) export.

- Pri dennom exporte sú dáta zasielané do BigQuery raz denne a zahŕňajú informácie z predchádzajúceho dňa do tabuľky o názve **events_YYYYMMDD**.
- Pri priebežnom exporte sú dáta posielané do BigQuery tabuľky s názvom **events_intraday_YYYYMMDD** kontinuálne počas dňa, čo umožňuje užívateľom prístup k dátam takmer v reálnom čase. Následne sú po skončení dňa uložené do tabuľky **events_YYYYMMDD** (Sharma, 2023).

Pre spoločnosť zvolíme priebežný (*realtime*) export a to kvôli prípadným ad hoc analýzám pri problémoch, napriek tomu, že denný export je pre bežnú agendu postačujúci.

Taktiež sa nastaví dodatočné možnosti, ako napríklad export určitých používateľských dát, ktoré obsahujú navyše od štandardných eventov aj dáta o publikáciách a predikciách ku konkrétnemu používateľovi. Toto však nie je potrebné pre vybranú firmu a zbytočne by bol zaťažovaný limit neplateného exportu dát.

Po nastavení exportu, začne Google Analytics 4 automaticky posielat' dáta do BigQuery podľa zvolenej frekvencie, čo bolo pre vybranú firmu 29.5.2023. Dáta sú exportované v

„event-based“ formáte, kde riadok v tabuľke predstavuje jednu udalosť preddefinovanú primárnym zdrojom dát a atribúty danej udalosti sú vnorené do dát pod udalosťou, ako je možné vidieť aj na obrázku 15.

Row	event_date	event_timestamp	event_name	event_params.key	event_params.value.string_value	event_params.value.int_value	event_params.value.float_value
1	20201031	1604099118066484	first_visit	page_location	https://www.ga4bigquery.com/	<i>null</i>	
				ga_session_number	<i>null</i>	1	
				page_title	Learn to utilize the Google Analytics data export in BigQuery	<i>null</i>	
				ga_session_id	<i>null</i>	1604099117	
2	20201031	1604099118066484	session_start	ga_session_id	<i>null</i>	1604099117	
				page_location	https://www.ga4bigquery.com/	<i>null</i>	
				ga_session_number	<i>null</i>	1	
				page_title	Learn to utilize the Google Analytics data export in BigQuery	<i>null</i>	
3	20201031	1604099118066484	page_view	page_title	Learn to utilize the Google Analytics data export in BigQuery	<i>null</i>	
				page_location	https://www.ga4bigquery.com/	<i>null</i>	
				ga_session_id	<i>null</i>	1604099117	
				entrances	<i>null</i>	1	
				ga_session_number	<i>null</i>	1	
				session_engaged	0	<i>null</i>	

Obrázok 15 – Náhľad štruktúry exportu dát do BigQuery

(Zdroj: GA4BigQuery, 2020)

5.1.2 Dočasné dátové úložisko

Spomínané „intraday“ tabuľky slúžia ako dočasné úložisko dát a poskytujú aktuálne informácie o užívateľskej interakcii. Posledná časť názvu tabuľky YYYYMMDD reprezentuje dátum, na ktorý sa dáta vzťahujú. Tieto tabuľky sú vytvorené automaticky pri nastavení integrácie BigQuery do Google Analytics 4 a sú súčasťou už spomínaného predchádzajúceho stavu (Sharma, 2023).

Po ukončení dňa, obvykle po polnoci v časovej zóne zvolenej v GA4 vlastníctve, čo je pre vybranú spoločnosť časová zóna GMT+1, sa dáta z intraday tabuliek presunú do denných tabuliek, a dočasná tabuľka sa resetuje (vymaže) pre zber dát na nasledujúci deň.

5.1.3 Dátový sklad

Následne sú dáta usporiadané v tabuľke, ktorej názov sleduje formát events_YYYYMMDD, kde YYYYMMDD opäť označuje konkrétny dátum. Tento formát názvu umožňuje jednoduchú identifikáciu a správu dát podľa dátumu ich vzniku. Tieto tabuľky tvoria historický dátový sklad a sú v nich uložené historické dáta za jednotlivé dni od počiatočného nastavenia exportu dát z Google Analytics 4.

Prvý export pre vybranú firmu, ktorý možno vidieť v rozhraní BigQuery, nastal na konci mája 2023, konkrétne je uložený v dátovej tabuľke *events_20230529*.

5.1.4 Dátové trhovisko

Dátové trhovisko, ktoré predstavuje pridanú komponentu systému, pozostáva z jedného alebo viacerých súborov dát, známych ako datasety. Tieto datasety obsahujú rôzne tabuľky naplnené dátami, ktoré prešli procesom spracovania. Kontinuálne spracovanie a príprava dát je realizovaná cez mechanizmus známy ako ETL, v prostredí BigQuery je tento proces označovaný ako „Scheduled queries“, jedná sa o tzv. naplánované dotazy.

Tvorby tabuliek možno rozdeliť na dve hlavné časti. Ako prvé je potrebné zostaviť SQL dotaz, keďže tento jazyk je predvolený v BigQuery konzole. Po zostavení sa tento dotaz spustí a výsledkom sú dáta, ktoré sú dočasne uložené v tabuľke prechodného charakteru. Táto prechodná tabuľka poskytuje možnosť prenosu alebo skopírovania dát do vopred pripraveného datasetu, v dôsledku čoho vzniknú dátové tabuľky vyplnené historickými dátami. Pre naplnenie datasetu je nutné určiť časový interval, ktorý má byť zastúpený historickými dátami. V tomto prípade sú vybrané dáta od začiatku roka 2024.

Druhým krokom je vytvorenie funkčnej dátovej pumpy alebo takzvaného ETL procesu. Tu sa využije predchádzajúci SQL dotaz s definovanou podmienkou určujúcou časový rozsah spracovávaných dát, ktoré majú byť prenesené do cieľovej tabuľky. Nastavenie tohto procesu sa realizuje cez používateľské rozhranie BigQuery, kde užívateľ určí názov procesu, čas spustenia, frekvenciu aktivácie, ako aj cieľový dataset a dátovú tabuľku. Užívateľ má tiež možnosť rozhodnúť, či sa existujúce dáta v cieľovej tabuľke prepíšu, alebo sa k nim pridajú nové dáta.

5.1.5 Vyhodnocovací nástroj

Poslednou komponentou, vyhodnocovacou a reportingovou vrstvou, je nástroj Looker Studio. Looker Studio je služba Google, ktorá má zabudovaný neplatený konektor pre dáta z BigQuery, čo znamená, že si možno jednoducho zvoliť datasety, ktoré boli pre spoločnosť vytvorené prostredníctvom niekoľkých kliknutí v rozhraní Looker Studia. Tieto datasety a ich dáta možno následne vizualizovať či kombinovať s inými zdrojmi podľa business požiadaviek na reporting a vyhodnocovanie návštevnosti webu či aplikácie. Ďalej je tam veľmi ľahké dopočítavanie metrík, či tvorenie dodatočných dimenzií, ktoré sú agregované. Napríklad

nákupná konverzná miera, ktorá sa dá ľahko vytvoriť ako podiel počtu celkových nákupov a návštevníkov webovej stránky či aplikácie.

Pri výbere tohto nástroja pre spoločnosť bol dôležitý najmä aspekt flexibility, škálovateľnosti, jednoduchosti použitia, ceny a ľahkého napojenia na Google Analytics 4, ktoré je vďaka Google API rýchle a spoľahlivé.

Napriek tomu, že sa v tejto práci nespomína o prepojení dát z iných systémov, do budúcnosti sa plánuje pridávať dáta z reklamných systémov akými sú napríklad Google Ads či Google Search Console, ktoré majú taktiež neplatený konektor v tomto nástroji.

5.2 Navrhovaný model dátového trhoviska

Dáta, ktoré sú nevyhnutné pre vytvorenie riešenia požadovaného spoločnosťou, boli rozvrhnuté do 14 dátových tabuliek, ktoré tvoria základný koncept návrhu dátového trhoviska. Ich štruktúra je navrhnutá tak, aby zabezpečila efektívne spracovanie a analýzu informácií. Každá tabuľka obsahuje špecifické dáta relevantné pre konkrétnu oblasť riešenia a sú medzi sebou prepojené prostredníctvom jedinečných identifikátorov. Pre tabuľky je spoločným identifikátor primárny kľúč – `user_id`. Prepojenia tabuliek umožňujú komplexné prehľady a prípadné analýzy, ktoré sú potrebné pre optimalizáciu procesov spoločnosti. Dátové tabuľky sú navrhnuté tak, aby boli flexibilné a mohli sa prispôbiť akýmkoľvek budúcim zmenám alebo rozšíreniam požiadaviek.

Tabuľka 5 – Prehľad navrhovaných dátových tabuliek dátového trhoviska

(Zdroj: Vlastné spracovanie)

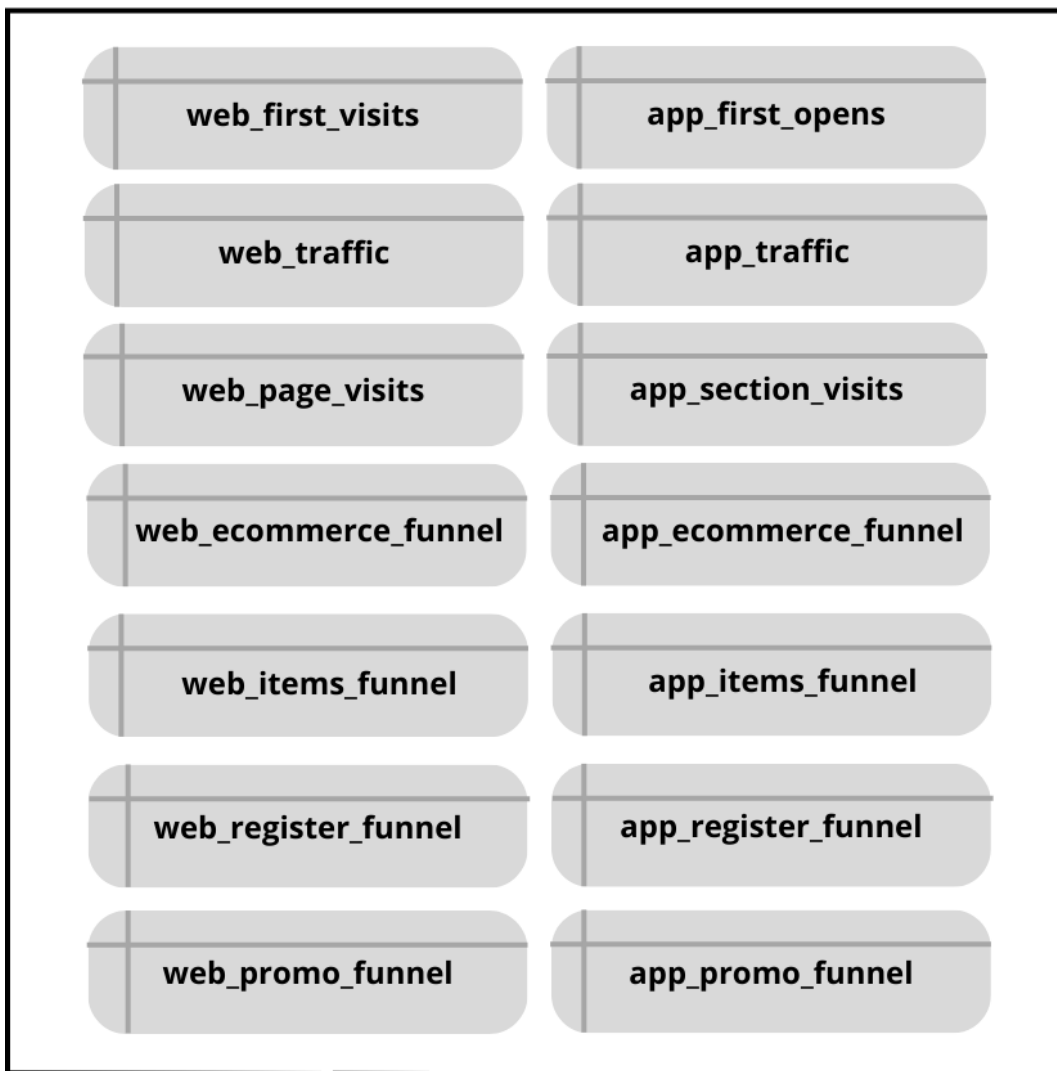
Názov tabuľky	Popis
web_first_visits	Dátová tabuľka uchováva záznamy o prvých návštevách webovej stránky.
app_first_opens	Dátová tabuľka uchováva záznamy o prvých otvoreníach webovej aplikácie a ich inštalovaní.
web_traffic	Dátová tabuľka uchováva záznamy o návštevách webovej stránky.
app_traffic	Dátová tabuľka uchováva záznamy o návštevách aplikácie.
web_page_visits	Dátová tabuľka nesie dáta o návšteve jednotlivých podstránok na webovej stránke.

app_section_visits	Dátová tabuľka nesie dáta o návšteve jednotlivých sekcií aplikácie.
web_ecommerce_funnel	Dátová tabuľka uchováva dáta o používateľoch, ktorý vytvorili objednávku na webovej stránke.
app_ecommerce_funnel	Dátová tabuľka uchováva dáta o používateľoch, ktorý vytvorili objednávku v aplikácii.
web_items_funnel	Dátová tabuľka uchováva dáta o produktoch v rámci nákupného lievika na webovej stránke.
app_items_funnel	Dátová tabuľka uchováva dáta o produktoch v rámci nákupného lievika v aplikácii.
web_register_funnel	Dátová tabuľka uchováva dáta o používateľoch, ktorý vyplnili registráciu do vernostného programu.
app_register_funnel	Dátová tabuľka uchováva dáta o používateľoch, ktorý vyplnili registráciu do vernostného programu.
web_promotion_funnel	Dátová tabuľka uchováva dáta o používateľoch, ktorý otvorili akčnú ponuku.
app_promotion_funnel	Dátová tabuľka uchováva dáta o používateľoch, ktorý otvorili akčnú ponuku.

5.2.1 Konceptuálny model

Konceptuálny dátový model dátového trhoviska identifikuje kľúčové entity a mapuje ich základné vzťahy. Tento model bol vytvorený s ohľadom na logickú štruktúru dát, pričom sa nezaobrá technickými detailmi ako sú dátové typy atribútov. Jeho primárnym účelom je poskytnúť prehľadný pohľad na organizáciu dát a interakcie medzi nimi. V rámci tohto modelu je možné jednoducho identifikovať a pochopiť hlavné prvky dátového trhoviska. Grafické znázornenie tohto konceptuálneho dátového modelu nájdete na obrázku 16. Okrem toho, tento model slúži ako základ pre ďalšie fázy návrhu a implementácie riešenia.

DÁTOVÉ TRHOVISKO



Obrázok 16 – Schéma konceptuálneho dátového modelu dátového trhoviska

(Zdroj: Vlastné spracovanie)

Dátové tabuľky v dátovom trhovisku možno charakterizovať nasledovne:

web_first_visits

Dátová tabuľka o prvých návštevách webovej stránky je organizovaná v riadkoch, pričom každý riadok reprezentuje jedinečnú návštevu. Atribúty tejto tabuľky potom poskytujú ďalšie detaily o týchto návštevách.

app_first_opens

Dátová tabuľka o prvých návštevách a inštaláciách aplikácie je usporiadaná v riadkoch, kde každý riadok predstavuje jedinečnú udalosť - buď prvú návštevu alebo inštaláciu aplikácie. Atribúty v rámci každého riadku potom poskytujú podrobnosti o týchto udalostiach.

web_traffic a app_traffic

Dátové tabuľky uchovávajú záznamy o používateľských návštevách webovej stránky alebo aplikácie, pričom každý riadok reprezentuje jedinečnú návštevu používateľa a atribúty sú stĺpcami. Jedná sa však o agregované dáta za celú dobu návštevy.

web_page_visits a app_section_visits

Dátová tabuľka je usporiadaná v riadkoch, pričom každý riadok reprezentuje užívateľskú návštevu stránky alebo sekcie a nesie dáta o návšteve jednotlivých podstránok na webovej stránke alebo sekciách aplikácie. Atribúty, ktoré sú stĺpcami, poskytujú podrobnosti o týchto návštevách a konkrétnych podstránkach.

web_ecommerce_funnel a app_ecommerce_funnel

Dátové tabuľky uchovávajú dáta o krokoch používateľov, ktorí boli súčasťou ecommerce lievika na webovej stránke alebo v aplikácii, počínajúc zobrazením produktu v kategórií, zobrazením daného produktu na produktovej podstránke, následným pridaním do košíka, prezretím košíka, prejdéním objednávkového formulára, platbou a finálnym vytvorením objednávky na webovej stránke alebo v aplikácii, tento lievik je zobrazený na obrázku 17. Dátové tabuľky sú organizované v riadkoch, pričom každý riadok reprezentuje jeden z týchto krokov. Atribúty tejto tabuľky potom poskytujú ďalšie detaily o daných udalostiach.

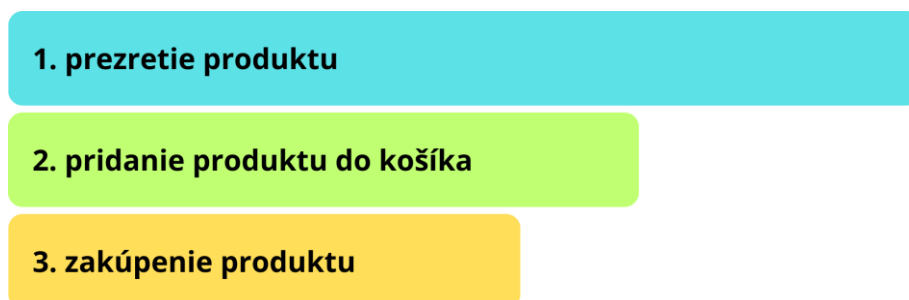


Obrázok 17 – Schéma nákupného lievika a jeho jednotlivé fázy

(Zdroj: Vlastné spracovanie)

web_items_funnel a app_items_funnel

Dátové tabuľky obsahujú agregované dáta o jednotlivých produktoch, ktoré spoločnosť ponúka. Každý riadok predstavuje záznam o produkte, ktorý bol súčasťou interakcie používateľa, to znamená interakciou prostredníctvom udalosti, a v stĺpcoch sa nachádzajú produktové atribúty. Pre jednoduchšie pochopenie je každá udalosť tohto lievika zobrazená na obrázku 18.

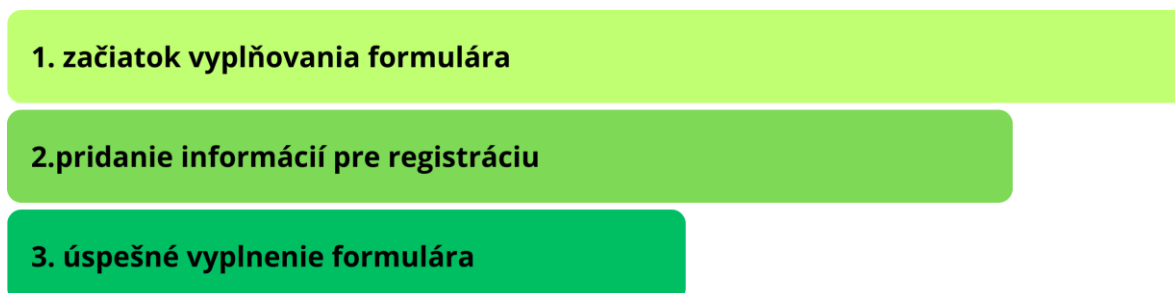


Obrázok 18 – Schéma nákupného lievika v rámci produktov

(Zdroj: vlastné spracovanie)

web_register_funnel a app_register_funnel

Dátové tabuľky obsahujú záznamy o používateľoch, ktorí navštívili registračnú stránku a začali proces registrácie do vernostného programu na webovej stránke alebo v aplikácii, sled krokov je znázornený na obrázku 19. Každý riadok v tejto tabuľke reprezentuje návštevu podstránky s danou udalosťou v rámci procesu registrácie, od prvého kroku - prechodu na registračnú stránku alebo sekciu, cez vyplňovanie registračného formulára, až po finálne dokončenie registrácie. Atribúty, stĺpce, v rámci každého riadku potom poskytujú podrobnosti o týchto krokoch.



Obrázok 19 – Schéma registračného lievika

(Zdroj: Vlastné spracovanie)

web_promo_funnel a app_promo_funnel

Dátové tabuľky obsahujú záznamy o používateľoch, ktorí na webovej stránke alebo v aplikácii interagovali s akčnou ponukou. Každý riadok v tejto tabuľke reprezentuje krok akčnej ponuky a interakcie s ňou, od prvého kroku - otvorenia ponuky, po prelistovanie na do polovice až po finálne prelistovanie celej ponuky. Tieto konkrétne udalosti boli vybrané ako kľúčové marketingovým tímom a sú súčasťou dát zbieraných prostredníctvom Google Analytics 4. Atribúty, ktoré sú stĺpcami, poskytujú podrobnosti o týchto krokoch.

1. prezretie prvej strany akčnej ponuky

2. prezretie polovice akčnej ponuky

3. prezretie poslednej strany akčnej ponuky

Obrázok 20 – Schéma lievika prezerania akčnej ponuky

(Zdroj: Vlastné spracovanie)

5.2.2 Logický dátový model

Logický dátový model, na rozdiel od konceptuálneho, poskytuje detailnejší pohľad na organizáciu a štruktúru dát, ktorý stojí na základoch navrhovaného konceptuálneho modelu. Tento model zahŕňa technické aspekty, ako sú dátové typy atribútov a vzťahy medzi entitami. Jeho cieľom bude poskytnúť presný a komplexný pohľad na to, ako sú dáta spoločnosti v trhovisku organizované a spracované. Časť jeho vizualizácie je zobrazená na obrázku 21, zvyšok logického modelu, chýbajúce tabuľky pre aplikačné dáta, možno nájsť v prílohe A. Jednotlivé dátové tabuľky obsahujú atribúty, ktoré sú špecifické pre druh tabuľky a účel dát a spĺňajú požiadavky zadávateľa. Niektoré názvy týchto atribútov sú zmenené z pôvodného vopred definovaného názvu z exportu Google Analytics 4 a ich názov je zvolený podľa toho ako sa budú volať vo finálnom stave implementácie.

Celkový logický dátový model tak predstavuje jasné a presné usporiadanie dát, ktoré je pripravené na nasledujúcu fázu – implementáciu o ktorej sa pojednáva v nasledujúcej kapitole tejto práce.

web_first_visits	
event_date	DÁTUM
user_pseudo_id	TEXT
user_id	TEXT
session_id	ČÍSLO
page_title	TEXT
page_location	TEXT
page_location_stripped	TEXT
page_referrer	TEXT
platform	TEXT
device	TEXT
os	TEXT
country	TEXT
city	TEXT
first_traffic_source	TEXT
first_traffic_medium	TEXT
first_traffic_campaign	TEXT

web_page_visits	
event_date	DÁTUM
user_pseudo_id	TEXT
user_id	TEXT
user_logged	BOOLEAN
session_id	ČÍSLO
page_title	TEXT
page_location	TEXT
page_location_stripped	TEXT
page_referrer	TEXT
engagement_time_sec	ČÍSLO
platform	TEXT
device	TEXT
os	TEXT
country	TEXT
city	TEXT
traffic_source	TEXT
traffic_medium	TEXT
traffic_campaign	TEXT

web_traffic	
event_date	DÁTUM
event_name	TEXT
user_pseudo_id	TEXT
user_id	TEXT
session_id	ČÍSLO
user_logged	BOOLEAN
session_engaged	TEXT
engagement_time_sec	ČÍSLO
landing_page	TEXT
landing_page_stripped	TEXT
page_referrer	TEXT
platform	TEXT
device	TEXT
os	TEXT
country	TEXT
city	TEXT
traffic_source	TEXT
traffic_medium	TEXT
traffic_campaign	TEXT

web_ecommerce_funnel	
event_date	DÁTUM
event_name	TEXT
step_number	TEXT
user_pseudo_id	TEXT
user_id	TEXT
user_logged	BOOLEAN
session_id	ČÍSLO
page_title	TEXT
page_location	TEXT
page_location_stripped	TEXT
page_referrer	TEXT
platform	TEXT
device	TEXT
os	TEXT
country	TEXT
city	TEXT
traffic_source	TEXT
traffic_medium	TEXT
traffic_campaign	TEXT
transaction_id	TEXT
purchase_revenue	ČÍSLO
total_item_quantity	ČÍSLO
shipping_value	ČÍSLO
unique_items	ČÍSLO

web_register_funnel	
event_date	DÁTUM
event_name	TEXT
step_number	TEXT
user_pseudo_id	TEXT
user_id	TEXT
user_logged	BOOLEAN
session_id	ČÍSLO
page_title	TEXT
page_location	TEXT
page_location_stripped	TEXT
page_referrer	TEXT
platform	TEXT
device	TEXT
os	TEXT
country	TEXT
city	TEXT
traffic_source	TEXT
traffic_medium	TEXT
traffic_campaign	TEXT

web_items_funnel	
event_date	DÁTUM
event_name	TEXT
step_number	TEXT
item_id	TEXT
item_name	TEXT
item_brand	TEXT
item_category	TEXT
item_category2	TEXT
item_category3	TEXT
price	ČÍSLO
quantity	ČÍSLO
item_revenue	ČÍSLO
transaction_id	TEXT
user_pseudo_id	TEXT
user_id	TEXT
session_id	ČÍSLO
user_logged	BOOLEAN
page_location	TEXT
page_title	TEXT
page_referrer	TEXT
page_location_stripped	TEXT
platform	TEXT
device	TEXT
country	TEXT
city	TEXT
traffic_source	TEXT
traffic_medium	TEXT
traffic_campaign	TEXT
item_views	ČÍSLO
item_add_to_carts	ČÍSLO
item_purchases	ČÍSLO

web_promo_funnel	
event_date	DÁTUM
event_name	TEXT
step_number	STRING
user_pseudo_id	TEXT
user_id	STRING
user_logged	BOOLEAN
session_id	ČÍSLO
page_title	TEXT
page_location	TEXT
page_location_stripped	TEXT
page_referrer	TEXT
platform	TEXT
device	TEXT
os	TEXT
country	TEXT
city	TEXT
traffic_source	TEXT
traffic_medium	TEXT
traffic_campaign	TEXT

Obrázok 21 – Schéma čiastočného logického dátového modelu dátového trhoviska pre webové dáta (1/2)

(Zdroj: Vlastné spracovanie)

V dátových tabuľkách možno nájsť viaceré atribúty, tieto atribúty budú pre jasnejší prehľad vysvetlené v logických celkoch.

Atribúty, ktoré sú súčasťou takmer všetkých tabuliek sú:

- **event_date** – Časový údaj, ktorý špecifikuje, kedy používateľ vykonal danú interakciu.
- **event_name** – Udáva názov špecifickej interakcie, ktorú používateľ vykonal v rámci návštevy webovej stránky alebo aplikácie.
- **user_pseudo_id** – Unikátny identifikátor používateľa vygenerovaný BigQuery.
- **user_id** – Unikátny identifikátor používateľa posielaný priamo podnikom.
- **session_id** - Unikátny identifikátor relácie. Relácia predstavuje nepretržitý časový úsek, ktorého čas vypršania možno nastaviť v prostredí Google Analytics 4 a počas ktorého sa používateľ nachádza na stránke. V angličtine sa relácia označuje ako „*session*“.
- **user_logged** - Informácia o tom, či je používateľ prihlásený alebo nie.
- **platform** – Špecifikuje aký typ platformy bol použitý pre jednotlivé interakcie, t.j. web alebo aplikácia.
- **device** – Uvádza typ zariadenia, na ktorom používateľ vykonal danú akciu.
- **os** – Špecifikuje typ operačného systému, z ktorého používateľ vykonal danú akciu v rámci prechodu produktovým formulárom.
- **country** – Uvádza informáciu o tom, z ktorej krajiny používateľ vykonal danú akciu.
- **city** – Nesie informáciu, z ktorého mesta používateľ vykonal danú akciu.
- **traffic_source** – Uvádza názov zdroja návštevy, ktorým sa používateľ dostal na webovú stránku či do aplikácie v rámci relácie.
- **traffic_medium** – Jedná sa o dodatočný atribút ku **traffic_source**, prináša informáciu o spôsobe, akým sa používateľ dostal na webovú stránku alebo do aplikácie v rámci relácie. Príkladom je napríklad platená návštevnosť (reklama) alebo referencia na inej webovej stránke.
- **traffic_campaign** – Rozširuje atribút **traffic_source** o názov kampane, ktorou sa používateľ dostal na webovú stránku alebo do aplikácie v rámci špecifickej relácie.
- **user_logged** – Nesie informáciu o tom, či je používateľ prihlásený do svojho účtu alebo nie.

Následne sa jednotlivé atribúty môžu líšiť aj na základe platformy z ktorej prichádzajú.

Atribúty špecifické pre dátový tok z webovej stránky:

- **page_title** – Špecifikuje názov stránky, ktorú používateľ navštívil.
- **page_location** – Špecifikuje webovú adresu stránky, ktorú používateľ navštívil.
- **page_referrer** – Odkazujúca webová stránka, z ktorej používateľ prišiel.
- **page_location_stripped** – Webová adresa stránky bez parametrov URL, ktorú používateľ navštívil.
- **landing_page** – Prvá stránka, ktorú používateľ navštívil počas svojej relácie.
- **landing_page_stripped** – Prvá stránka bez parametrov URL, ktorú používateľ navštívil počas svojej relácie. Tento atribút umožňuje vyhodnotiť vstupné stránky bez filtrovania dodatočných parametrov.

Atribúty špecifické pre dátový tok z aplikácie:

- **screen_class** – Názov triedy obrazovky či sekcie v aplikácii.
- **screen** – Názov obrazovky či sekcie, ktorú používateľ navštívil v aplikácii.
- **previous_screen** – Názov predchádzajúcej obrazovky či sekcie, ktorú používateľ navštívil v aplikácii.

Pre tabuľky súvisiace s produktmi (*items*), ako sú `web_items_funnel` a `app_items_funnel`, boli pridané aj nasledujúce atribúty:

- **item_id** – Unikátny identifikátor položky alebo produktu.
- **item_name** – Názov položky alebo produktu.
- **item_brand** – Značka položky alebo produktu.
- **item_category** – Kategória, do ktorej patrí položka alebo produkt.
- **price** – Cena položky alebo produktu.
- **quantity** – Množstvo položky alebo produktu.
- **item_revenue** – Príjem z položky alebo produktu.
- **transaction_id** – Unikátny identifikátor transakcie.
- **item_views** – Počet zobrazení položky alebo produktu.
- **item_add_to_carts** – Počet pridaní položky alebo produktu do košíka.
- **item_purchases** – Počet nákupov položky alebo produktu.

Iné, špecifické dimenzie či atribúty sú:

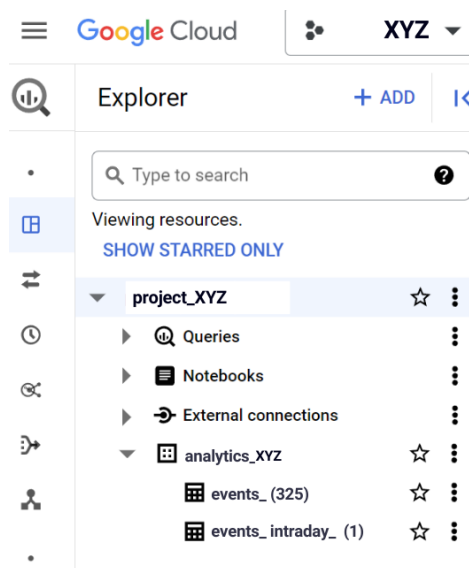
- **session_engaged** – Tento údaj informuje, či bol používateľ počas svojej návštevy webu alebo aplikácie aktívny alebo neaktívny. Je kľúčový hlavne pre výpočet miery okamžitého odchodu z webu, v angličtine známeho ako „bounce rate”. Hodnoty tohto atribútu sú typu boolean (pravdivostné).

- **engagement_time_sec** – Doba zapojenia používateľa na webe alebo v aplikácii. Označuje celkový čas, ktorý používateľ strávil na webe počas jedného sedenia (sčítaný za celé toto obdobie) alebo čas strávený na jednej podstránke webu alebo v sekcii aplikácie. V angličtine sa tento čas často označuje ako „*engagement time*”.
- **first_traffic_source** – Uvádza názov prvého zdroja návštevy, ktorým sa používateľ dostal na webové stránky či do aplikácie v rámci prvej návštevy alebo prvého otvorenia aplikácie.
- **first_traffic_medium** – Jedná sa o dodatočný atribút ku `first_traffic_source`, prináša informáciu o spôsobe, akým sa používateľ dostal na webové stránky či do aplikácie v rámci prvej návštevy alebo prvého otvorenia aplikácie. Príkladom je napríklad platená návštevnosť (reklama) alebo referencia na inej webovej stránke.
- **first_traffic_campaign** – Rozširuje atribút `traffic_source` o názov kampane, ktorou sa používateľ dostal na webové stránky či do aplikácie v rámci prvej návštevy alebo prvého otvorenia aplikácie.

6 IMPLEMENTÁCIA NÁVRHU V BIGQUERY

6.1 Východiskový stav

Nová štruktúra dátového skladu bude implementovaná použitím BigQuery, v ktorej sa nachádzalo aj pôvodné riešenie, ktoré nebolo nijakým spôsobom modifikované. Pôvodná štruktúra dátového skladu bola dopredu stanovená Google Analytics 4 exportom. Dáta, ktoré sú na dennej báze exportované zo služby Google Analytics 4 vo vybranom formáte exportu sú uložené v projekte, ktorý je pre anonymitu firmy nazvaný *project_XYZ*. Tento projekt bol vytvorený pod účtom spoločnosti ešte pri základnom prepojení služby Google BigQuery a nástroja Google Analytics 4. V projekte možno nájsť dataset nazvaný *analytics_XYZ*, XYZ reprezentuje ID vlastníctva danej spoločnosti v Google Analytics 4. Tento dataset obsahuje dátové tabuľky, ktoré tvoria dátový sklad a dočasné dátové úložisko. Tabuľky začínajúce slovom *events_*, obsahujú dáta z vždy príslušného dňa, ktorý je špecifikovaný za daným výrazom.

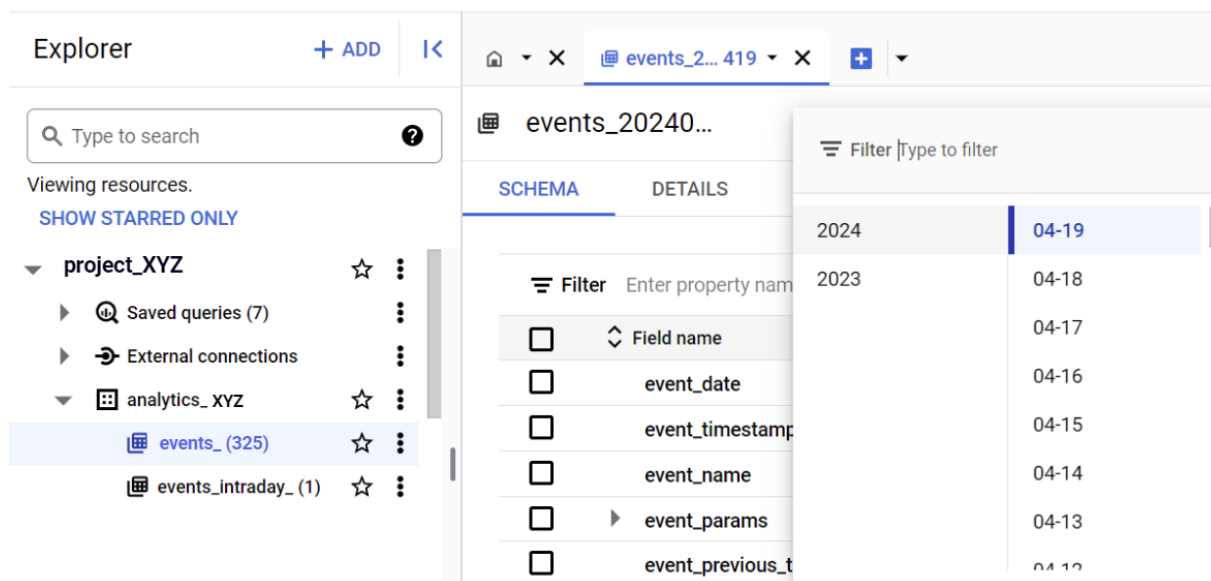


Obrázok 22 – Východiskový stav dátového skladu v BigQuery

(Zdroj: Vlastné spracovanie)

Do tabuliek s názvom *events_intraday_*, ktoré tvoria dočasné dátové úložisko, sa exportujú dáta na kontinuálnej báze, ich názov funguje na rovnakom princípe, avšak tieto tabuľky sú vždy vymazané po konci daného dňa, kedy sa dáta pretransformujú do tabuliek s názvom *events_*. Preto je tabuľka s predponou *events_intraday_* len jedna, kým tabuliek *events_* je 325, čo tvorí 325 denných exportov dát.

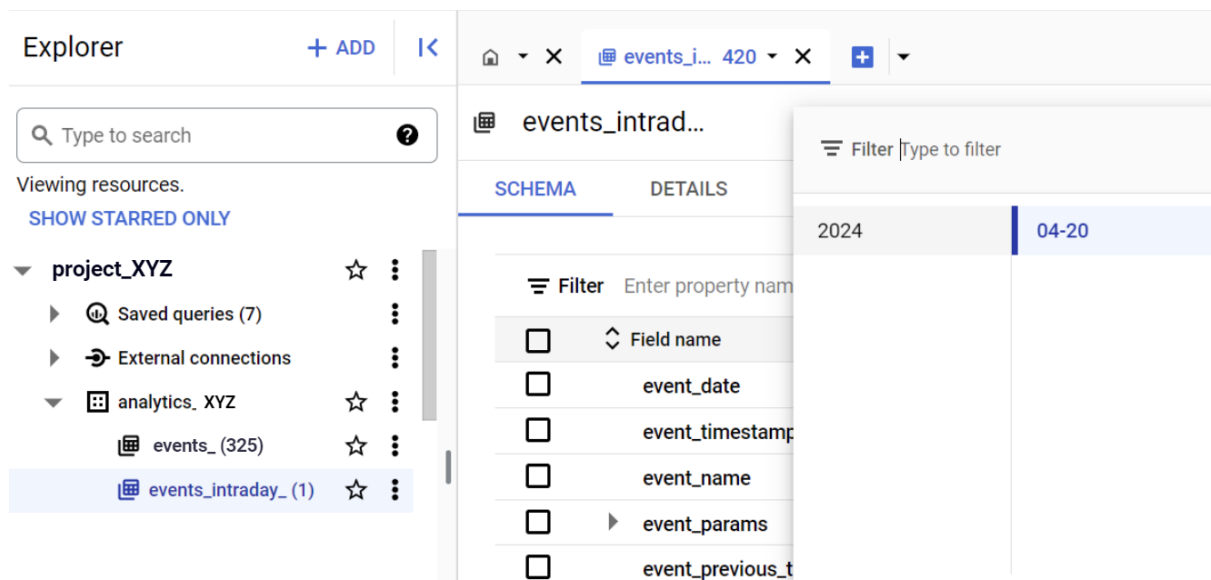
Jednotlivých 325 denných exportov možno ľahko prehľadávať cez filter. Následne vie konzola poskytnúť náhľad ich štruktúry v tabuľke a dodatočné informácie, ako je možno vidieť na obrázku 23.



Obrázok 23 – Dátový sklad v prostredí BigQuery

(Zdroj: Vlastné spracovanie)

Na obrázku 24 možno vidieť časový filter pre dočasné dátové úložisko *events_intraday_*. V tomto dočasnom úložisku sa nachádzajú dáta za aktuálny deň a po polnoci GMT+1 času sa presunú do dátového skladu a dočasné dátové úložisko bude premazané.



Obrázok 24 – Dočasné dátové úložisko v prostredí BigQuery

(Zdroj: Vlastné spracovanie)

6.2 Postup a plnenie dátového trhoviska

Pre implementáciu návrhu bolo potrebné vytvoriť nový dataset s názvom *datamart_XYZ*, ktorý bude obsahovať dátové tabuľky z návrhu nachádzajúceho sa v návrhovej časti diplomovej práce.

Tento dataset bol následne naplnený historickými dátami a boli nastavené plánované dotazy (dátové pumpy), ktoré zabezpečujú automatizovaný proces extrakcie, transformácie a načítania dát (ETL) z dočasného úložiska *events_intraday_* a dátového skladu do dátového trhoviska, ktoré slúži k plneniu dátového trhoviska novými dátami.

Plánované dotazy, ktoré budú vytvorené slúžia na pravidelné dopĺňanie nových dát do dátového trhoviska prostredníctvom procesu ETL, sa spúšťajú každé 4 hodiny. Tieto dotazy extrahujú dáta nielen z dátového skladu, to sú tabuľky *events_*, ale aj z dočasného úložiska, z tabuliek *events_intraday_*. Dotazy boli nastavené pre získavanie dát za posledné dva dni a aktuálny deň. Ako preventívne opatrenie proti duplicitným záznamom sa pred každým spustením týchto dotazov vykoná príkaz, ktorý odstraňuje dáta za rovnaké časové obdobie z nového dátového trhoviska.

Pre čo najlepšiu úspešnosť ETL procesov bolo implementované monitorovanie mechanizmov a posielanie upozornení, ktoré upozorňujú na akýkoľvek problémy alebo chyby, ktoré by mohli ovplyvniť kvalitu alebo aktuálnosť dát. Ak sa vyskytne chyba pri vykonávaní plánovaného dotazu, BigQuery automaticky odošle e-mail na adresu, ktorá je spojená s účtom Google Cloud. Tento e-mail obsahuje podrobnosti o chybe a odkaz na stránku s detailmi plánovaného dotazu. Postup implementácie bol rozdelený do niekoľkých krokov, ktoré logicky oddelia proces napĺňania datasetu.

6.2.1 Naplnenie dátových tabuliek historickými dátami

Pre naplnenie jednotlivých dátových tabuliek historickými dátami bol pripravený SQL dotaz, ktorý konvertoval a upravil dáta do požadovanej formy, ktorá bola opísaná v konceptuálnom a logickom návrhu, za určené historické obdobie. Dáta boli získané podmienkou spätne za posledných niekoľko dní a to od začiatku roka 2024. Následne bol tento pripravený dotaz spustený. Výstupy sú po tomto kroku zobrazené v dočasnom okne, kde sa nastaví ich presun do konkrétneho datasetu a tabuľky, prostredníctvom kopírovania tejto dočasnej dátovej tabuľky (výstupu).

Tento postup rozdelíme do jednotlivých krokov:

- A. **Vytvorenie SQL dotazu a jeho spustenie** – výsledná tabuľka je v podobe, ktorá je opísaná konceptuálnym a logickým modelom a obsahuje historické dáta od začiatku roku 2024.
- B. **Uloženie výsledkov do novej tabuľky** – dáta, ktoré sú výsledkom dotazu budú následne uložené do požadovanej tabuľky, čo sú tabuľky, ktoré tvoria dátové trhovisko.

SQL dotazy pre jednotlivé dátové tabuľky

V rámci nadchádzajúceho časti bude ukázaných 7 SQL dotazov alebo predpisov (pre dátové tabuľky webových stránok), zvyšných 7 dotazov (pre dátové tabuľky aplikácie) alebo celé dlhšie dotazy, možno nájsť v prílohe II-XII.

1. **web_first_visits** dátová tabuľka a jej SQL dotaz pre získanie historických dát z dátového skladu je znázornený na obrázku 25.

```
SELECT
  CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  user_pseudo_id,
  user_id,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_referrer,
  REGEXP_EXTRACT(
    (SELECT value.string_value FROM UNNEST(event_params)
     WHERE key = 'page_location'), r'([^\?]*\?)') AS page_location_stripped,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
  END AS first_traffic_source,
  traffic_source.medium AS first_traffic_medium,
  traffic_source.name AS first_traffic_campaign
FROM
  `project_XYZ.analytics_XYZ.events_*`
WHERE
  event_name = "first_visit"
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
```

Obrázok 25 – SQL dotaz pre získanie dát do tabuľky web_first_visits

(Zdroj: Vlastné spracovanie)

2. **web_traffic** dátová tabuľka a jej SQL dotaz pre získanie historických dát z dátového skladu je znázornený na obrázku 26.

```
SELECT
  CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  'session_aggregated' AS event_name,
  user_pseudo_id,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = 'ga_session_id') AS session_id,
  MAX(user_id) AS user_id,
  MAX(CASE WHEN event_name = 'user_logged' THEN TRUE ELSE FALSE END) AS user_logged,
  MAX((SELECT value.string_value FROM UNNEST(event_params)
  | WHERE KEY = 'session_engaged')) AS session_engaged,
  SUM((SELECT value.int_value FROM UNNEST(event_params)
  | WHERE KEY = 'engagement_time_msec'))/1000 AS engagement_time_sec,
  CASE
    WHEN (SELECT value.int_value FROM UNNEST(event_params)
    | WHERE event_name = 'page_view' AND KEY = 'entrances') = 1
    THEN (SELECT value.string_value FROM UNNEST(event_params)
    | WHERE event_name = 'page_view' AND KEY = 'page_location')
  END AS landing_page,
  CASE
    WHEN (SELECT value.int_value FROM UNNEST(event_params)
    | WHERE event_name = 'page_view' AND KEY = 'entrances') = 1
    THEN REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
    | WHERE event_name = 'page_view' AND key = 'page_location'), r'([?]*)')
  END AS landing_page_stripped,
  CASE
    WHEN (SELECT value.int_value FROM UNNEST(event_params)
    | WHERE event_name = 'page_view' AND KEY = 'entrances') = 1
    THEN (SELECT value.string_value FROM UNNEST(event_params)
    | WHERE event_name = 'page_view' AND KEY = 'page_referrer')
  END AS page_referrer,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
  END AS traffic_source,
  traffic_source.medium AS traffic_medium,
  traffic_source.name AS traffic_campaign,
  COUNTIF (event_name = 'page_view') as page_views
FROM
  `project_XYZ.analytics_XYZ.events_*`
WHERE
  platform = "web"
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
GROUP BY
  event_date, user_pseudo_id, session_id, device, os,
  country, city, traffic_source, traffic_medium, traffic_campaign, platform, landing_page,
  landing_page_stripped, page_referrer
```

Obrázok 26 – SQL dotaz pre získanie dát do tabuľky web_traffic

(Zdroj: Vlastné spracovanie)

3. **web_page_visits** dátová tabuľka a jej SQL dotaz pre získanie historických dát z dátového skladu je znázornený na obrázku 27.

```
SELECT
  CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  user_pseudo_id,
  user_id,
  (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
  REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
  | WHERE KEY = 'page_location'), r'([^\?]*)') AS page_location_stripped,
  SUM((SELECT value.int_value FROM UNNEST(event_params)
  | WHERE KEY ='engagement_time_msec'))/1000 AS engagement_time_sec,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
  | WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  | WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  | ELSE traffic_source.source
  END AS traffic_source,
  traffic_source.medium AS traffic_medium,
  traffic_source.name AS traffic_campaign
FROM
  `project_XYZ.analytics_XYZ.events_*`
WHERE
  event_name = "page_view"
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
```

Obrázok 27 – SQL dotaz pre získanie dát do tabuľky **web_page_visits**

(Zdroj: Vlastné spracovanie)

4. **web_ecommerce_funnel** dátová tabuľka a jej SQL dotaz pre získanie historických dát z dátového skladu sa skladá zo 8 častí zodpovedajúcich 8 udalostiam, ktoré boli znázornené na obrázku 17 v predchádzajúcej kapitole. Každá časť dotazu extrahuje dáta týkajúce sa jednej konkrétnej udalosti, ktoré sú nastavené a merané v Google Analytics 4.

Univerzálny prepis časti dotazu sa nachádza na obrázku 28 a celý SQL dotaz možno nájsť v prílohe V.

```

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step n" as step_number, -- číslo kroku vo web_ecommerce_funnel (n nahradit' číslom)
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY = "isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
| WHERE KEY = 'page_location'), r'([^\?]*)') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
| WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
| WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
| ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "" -- názov udalosti vo web_ecommerce_funnel
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL -- pridané len v prípade ak nasleduje SELECT príkaz
-- zvyšok dotazu pre ostatné udalosti vo web_register_funnel

```

Obrázok 28 – Vzor SQL dotazu pre získanie dát do tabuľky web_ecommerce_funnel

(Zdroj: Vlastné spracovanie)

5. **web_register_funnel** dátová tabuľka a jej SQL dotaz sa skladá z 3 častí zodpovedajúcich 3 udalostiam, ktoré boli znázornené na obrázku 18 v predchádzajúcej kapitole.

Každá časť dotazu extrahuje dáta týkajúce sa jednej konkrétnej udalosti. Celý dotaz možno nájsť v prílohe VII. Univerzálny prepis časti dotazu sa nachádza na obrázku 29.


```

SELECT
  CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  event_name,
  "step n" as step_number, -- číslo kroku vo web_register_funnel (n nahradiť číslom)
  user_pseudo_id,
  user_id,
  (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
  REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
  | WHERE KEY = 'page_location'), r'([^\?]*)') AS page_location_stripped,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
  | WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  | WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  | ELSE traffic_source.source
  END AS traffic_source,
  traffic_source.medium AS traffic_medium,
  traffic_source.name AS traffic_campaign
FROM
  `project_XYZ.analytics_XYZ.events_*`
WHERE
  platform = "web"
  AND event_name = "" -- názov udalosti vo web_register_funnel
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL -- pridané len v prípade ak nasleduje SELECT príkaz
-- zvyšok dotazu pre ostatné udalosti vo web_register_funnel

```

Obrázok 29 – Vzor SQL dotazu pre získanie dát do tabuľky web_register_funnel

(Zdroj: Vlastné spracovanie)

6. **web_promo_funnel** dátová tabuľka a jej SQL dotaz sa skladá zo 3 častí zodpovedajúcich 3 udalostiam, ktoré boli znázornené na obrázku 19 v predchádzajúcej kapitole. Každá časť dotazu extrahuje dáta týkajúce sa jednej konkrétnej udalosti. Celý dotaz možno nájsť v prílohe IX. Univerzálny prepis časti dotazu sa nachádza na obrázku 30.


```

SELECT
  CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  event_name,
  "step n" as step_number, -- číslo kroku vo web_promo_funnel (n nahradiť číslom)
  user_pseudo_id,
  user_id,
  (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
  REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
  | WHERE KEY = 'page_location'), r'([^\?]*)') AS page_location_stripped,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
  | WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  | WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  | ELSE traffic_source.source
  END AS traffic_source,
  traffic_source.medium AS traffic_medium,
  traffic_source.name AS traffic_campaign
FROM
  `project_XYZ.analytics_XYZ.events_*`
WHERE
  platform = "web"
  AND event_name = "" -- názov udalosti vo web_promo_funnel
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL -- pridané len v prípade ak nasleduje SELECT príkaz
-- zvyšok dotazu pre ostatné udalosti vo web_items_funnel

```

Obrázok 30 – Vzor SQL dotazu pre získanie dát do tabuľky web_promo_funnel

(Zdroj: Vlastné spracovanie)

7. **web_items_funnel** dátová tabuľka a jej SQL dotaz sa skladá zo 3 častí zodpovedajúcich 3 udalostiam, ktoré boli znázornené na obrázku 20 v predchádzajúcej kapitole.

Každá časť dotazu extrahuje dáta týkajúce sa jednej konkrétnej udalosti. Celý dotaz možno nájsť v prílohe XI. Univerzálny prepis časti dotazu sa nachádza na obrázku 31.

```

SELECT
  CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  event_name,
  "step n" as step_number, -- číslo kroku vo web_items_funnel (n nahradíť číslom)
  item_id,
  item_name,
  item_brand,
  item_category,
  item_category2,
  item_category3,
  price,
  quantity,
  item_revenue,
  ecommerce.transaction_id AS transaction_id,
  user_pseudo_id,
  user_id,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") AS session_id,
  (SELECT value.string_value FROM UNNEST(user_properties) WHERE KEY = "isLoggedIn") AS user_logged,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_title") AS page_title,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_location,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
  REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = 'page_location'), r'([?]*)*') AS page_location_stripped,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
  END AS traffic_source,
  traffic_source.medium AS traffic_medium,
  traffic_source.name AS traffic_campaign,
  COUNT(CASE WHEN event_name = "view_items" THEN CONCAT(user_pseudo_id,
    (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = 'ga_session_id'))
  END) AS item_views,
  COUNT(CASE WHEN event_name = "add_to_cart" THEN CONCAT(user_pseudo_id,
    (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = 'ga_session_id'))
  END) AS item_add_to_carts,
  COUNT(CASE WHEN event_name = "purchase" THEN CONCAT(user_pseudo_id,
    (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = 'ga_session_id'))
  END) AS item_purchases
FROM
  `project_XYZ.analytics_XYZ.events_*`,
  UNNEST(items) AS items
WHERE
  platform = "web"
  AND event_name = "" -- názov udalosti vo web_items_funnel
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", DATE '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL -- pridané len v prípade ak nasleduje SELECT príkaz
-- zvyšok dotazu pre ostatné udalosti vo web_items_funnel

```

Obrázok 31 – Vzor SQL dotazu pre získanie dát do tabuľky web_items_funnel

(Zdroj: Vlastné spracovanie)

Uloženie výsledkov do príslušných tabuliek

Po spustení každého dotazu sú výsledky uložené a to prostredníctvom vytvorenia novej dátovej tabuľky, a to možnosťou *Save Results* a *BigQuery Table*. Zakaždým je vybraný *project_XYZ*, dataset *datamart_XYZ* a názov tabuľky je zvolený podľa toho, pre ktorú tabuľku boli tieto výsledky SQL dotazom vygenerované. Možnosť tohto ukladania sa teda opakuje 14-krát, t.j. 1-krát pre každú tabuľku. Ostatné rozšírené možnosti ostávajú na predvolenom nastavení.

Save to BigQuery Table

Destination

Project *
project_XYZ BROWSE

Dataset *
datamart_XYZ

Table *
názov_tabuľky

Maximum name size is 1,024 UTF-8 bytes. Unicode letters, marks, numbers, connectors, dashes, and spaces are allowed. The job will create the specified destination table if needed.

Advanced options ▼

SAVE CANCEL

Obrázok 32 – Ukladanie získaných dát SQL dotazov do novej tabuľky

(Zdroj: Vlastné spracovanie)

Výsledkom tohto procesu je 14 dátových tabuliek, obrázok 33, ktoré majú schému, ktorá bola navrhnutá v logickom modeli návrhu dátového trhoviska, tabuľka **web_first_visits** je znázornená na obrázku 34.

The screenshot shows the Google Cloud BigQuery Explorer interface. On the left, the 'Explorer' pane displays a tree view of resources under the 'datamart_XYZ' dataset, listing 14 tables including 'web_first_visits'. The main pane shows the 'web_first_visits' table selected, displaying its schema with columns and their data types and modes.

Field name	Type	Mode
event_date	STRING	NULLABLE
user_pseudo_id	STRING	NULLABLE
user_id	STRING	NULLABLE
session_id	INTEGER	NULLABLE
page_title	STRING	NULLABLE
page_location	STRING	NULLABLE
page_location_stripped	STRING	NULLABLE
page_referrer	STRING	NULLABLE
platform	STRING	NULLABLE
device	STRING	NULLABLE
os	STRING	NULLABLE
country	STRING	NULLABLE
city	STRING	NULLABLE
first_traffic_source	STRING	NULLABLE
first_traffic_medium	STRING	NULLABLE
first_traffic_campaign	STRING	NULLABLE

Obrázok 33 – Dátové trhovisko s vytvorenými tabuľkami

(Zdroj: Vlastné spracovanie)

The screenshot shows the BigQuery interface for a table named 'web_first_visits'. At the top, there are navigation options: 'web_first_visits', 'QUERY', 'SHARE', 'COPY', and 'SNAPSHOT'. Below these are tabs for 'SCHEMA', 'DETAILS', 'PREVIEW', 'LINEAGE', 'DATA PROFILE', and 'DATA QU'. A filter input field is present with the text 'Filter Enter property name or value'. The main content is a table listing the columns of the schema.

<input type="checkbox"/>	Field name	Type	Mode	Key	Collat
<input type="checkbox"/>	event_date	STRING	NULLABLE	-	-
<input type="checkbox"/>	user_pseudo_id	STRING	NULLABLE	-	-
<input type="checkbox"/>	user_id	STRING	NULLABLE	-	-
<input type="checkbox"/>	session_id	INTEGER	NULLABLE	-	-
<input type="checkbox"/>	page_title	STRING	NULLABLE	-	-
<input type="checkbox"/>	page_location	STRING	NULLABLE	-	-
<input type="checkbox"/>	page_location_stripped	STRING	NULLABLE	-	-
<input type="checkbox"/>	page_referrer	STRING	NULLABLE	-	-
<input type="checkbox"/>	platform	STRING	NULLABLE	-	-
<input type="checkbox"/>	device	STRING	NULLABLE	-	-
<input type="checkbox"/>	os	STRING	NULLABLE	-	-
<input type="checkbox"/>	country	STRING	NULLABLE	-	-
<input type="checkbox"/>	city	STRING	NULLABLE	-	-
<input type="checkbox"/>	first_traffic_source	STRING	NULLABLE	-	-
<input type="checkbox"/>	first_traffic_medium	STRING	NULLABLE	-	-
<input type="checkbox"/>	first_traffic_campaign	STRING	NULLABLE	-	-

Obrázok 34 – Vytvorená tabuľka web_first_visits a jej schéma v dátovom trhovisku

(Zdroj: Vlastné spracovanie)

6.2.2 Tvorba ETL procesu a plnenie tabuliek novými dátami

Ďalším postupom je tvorba ETL procesu, dátovej pumpy, ktorá doplní dáta za posledné dva dni. Pre tento proces budú využité tie isté SQL dotazy ako pri naplňaní tabuliek historickými dátami avšak zmení sa časť podmienky a to konkrétne časový filter. Táto podmienka, znamená, že tabuľky budú plnené dátami za posledné dva dni (vrátane prebiehajúceho dňa).

```
WHERE
  (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", DATE_SUB(CURRENT_DATE(), INTERVAL 2 DAY))
  | OR _TABLE_SUFFIX LIKE "%intraday%")
```

Obrázok 35 – Časová úprava podmienky v SQL dotazoch pre ETL proces

(Zdroj: Vlastné spracovanie)

Táto podmienka bude upravená pri všetkých 14 dotazoch a to v každej časti, ktorá ju obsahuje. Následne sa vytvorí 14 plánovaných dotazov v prostredí BigQuery, ktoré budú slúžiť ako spomínaný ETL proces či dátová pumpa. Tento proces musí byť nastavený tak, aby obsahoval príkaz pre odstránenie časti starých dát pre možné duplicity.

Nový plánovaný dotaz je možné konfigurovať zadaním kódu do konzoly a následným kliknutím na možnosť *Create a new scheduled query*, kde sa vyplnia informácie o intervale opakovania, mene samotného plánovaného dotazu (po vytvorení sa nedá modifikovať) a časovania. Rozšírené možnosti ostávajú na predvolenom nastavení.

New scheduled query
✔ Settings valid.

Details and schedule

INSERT web_first_visits

Schedule options

Repeat frequency *
 Custom

Custom schedule *
 4

Start now
 Start at set time

Start date and run time *
 5/10/24, 2:00PM CEST 📅

End never
 Schedule end time

End date CET 📅

🔔 Scheduled to run on the custom schedule *4* starting on May 10, 2024 at 2:00:00 PM CEST.

Destination for query results

Set a destination table for query results

Dataset *
✔ datamart_XYZ

Table Id *
 web_first_visits

Destination table partitioning field

Destination table write preference

Append to table
 Overwrite table

Automatic location selection ?

Location type ?

Region
Specify a region to colocate your datasets with other Google Cloud services.

Multi-region
Allow BigQuery to select a region within a group to achieve higher quota limits.

Region

Notification options

Send email notifications ?

Cloud Pub/Sub topic ?
projects/<project-name>/topics/<topic-name>

SAVE
CANCEL

Obrázok 36 – Konfigurácia ETL procesu pre tabuľku web_first_visits

(Zdroj: Vlastné spracovanie)

Po uložení detailov a rozvrhu tohto dotazu je potrebné vybrať finálnu destináciu nových dát. Tu sa vyberie požadovaný dataset *datamart_XYZ*, projekt *project_XYZ* a výsledná tabuľka (zakaždým jedna zo 14 navrhovaných). Rozšírené možnosti ostávajú opäť na predvolenom nastavením, no je odčiarknutá emailová notifikácia, ktorá poskytne informáciu v prípade zlyhania procesu. Rovnakým postupom sa vytvorenie príkaz na prečistenie dát, avšak pri tomto procese sa nevyplňujú parametre destinácie, obrázok 38, keďže tieto informácie sú už

zahrnuté v samotnom dotaze. Konfigurácia dátových púmp a mazacích príkazov je rovnaká pre všetky tabuľky v dátovom trhovisku.

```
DELETE
FROM
`project_XYZ.datamart_XYZ.web_first_visits`
WHERE
event_date = FORMAT_DATE("%Y%m%d", DATE_SUB(CURRENT_DATE(), INTERVAL 0 DAY))
OR event_date = FORMAT_DATE("%Y%m%d", DATE_SUB(CURRENT_DATE(), INTERVAL 1 DAY))
OR event_date = FORMAT_DATE("%Y%m%d", DATE_SUB(CURRENT_DATE(), INTERVAL 2 DAY))
```

Obrázok 37 – SQL príkaz pre vymazanie potencionálnych duplicitných dát web_first_visits

(Zdroj: Vlastné spracovanie)

New scheduled query Settings valid.

Details and schedule

Name for scheduled query *
DELETE web_first_visits

Schedule options

Repeat frequency *
Hours

Repeats every *
4 hours

Start now Start at set time

Start date and run time *
5/10/24, 1:50PM CEST

End never Schedule end time

End date CEST

i Scheduled to run every 4 hours starting on May 10, 2024 at 1:50:00 PM CEST.

Destination for query results

Set a destination table for query results

Dataset

Table Id

Destination table partitioning field

Destination table write preference

Append to table
 Overwrite table

Automatic location selection

Location type

Region
Specify a region to colocate your datasets with other Google Cloud services.

Multi-region
Allow BigQuery to select a region within a group to achieve higher quota limits.

Region

Notification options

Send email notifications

Cloud Pub/Sub topic
projects/<project-name>/topics/<topic-name>

SAVE **CANCEL**

Obrázok 38 – Konfigurácia plánovaného dotazu pre zmazanie dát pre tabuľku web_first_visits

(Zdroj: Vlastné spracovanie)

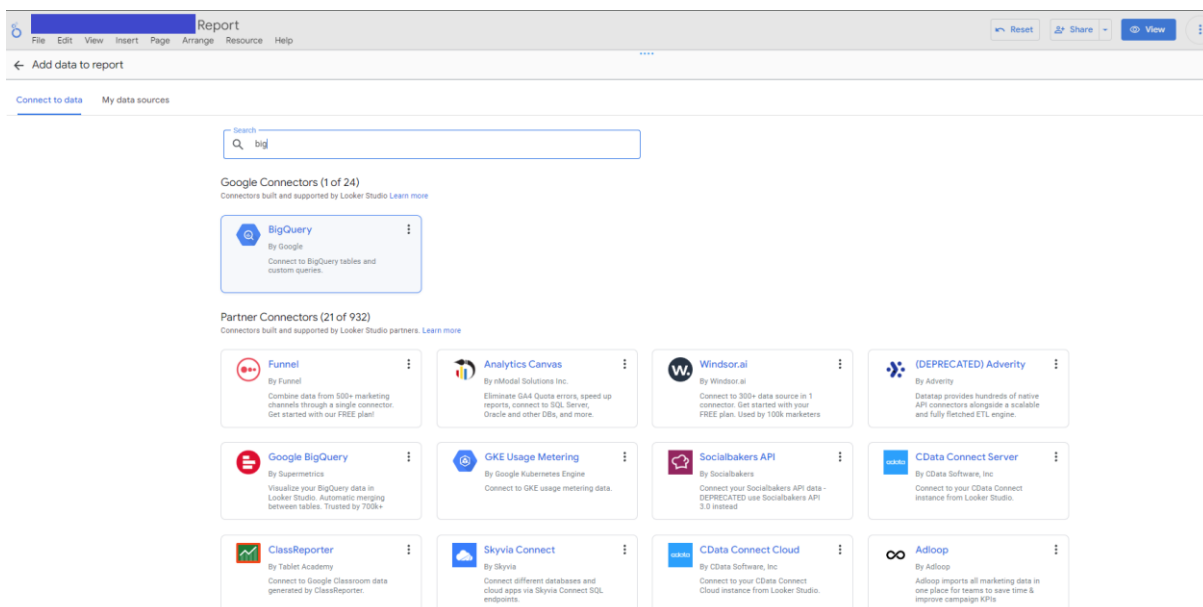
7 REPORTOVACÍ DASHBOARD

Novovytvorený dataset, ktorý sa nachádza v BigQuery je následne nutné vizualizovať. Pre vytvorenie reportovacieho dashboardu bol vybraný nástroj Looker Studio.

Looker Studio je nástroj na vizualizáciu dát, ktorý dokáže transformovať surové dáta na intuitívne a zrozumiteľné vizualizácie. Navyše, integrácia Looker Studio s BigQuery je možná neplateným konektorom, pretože obidva produkty sú produkty spoločnosti Google, LLC., čo zjednodušuje proces extrakcie a manipulácie s dátami. Prostredníctvom tohto nástroja môžu byť dáta efektívne analyzované a prezentované v jednoduchých podobách vyššiemu managementu, ale aj špecializovaným zamestnancom marketingu.

7.1 Napojenie dátového zdroja

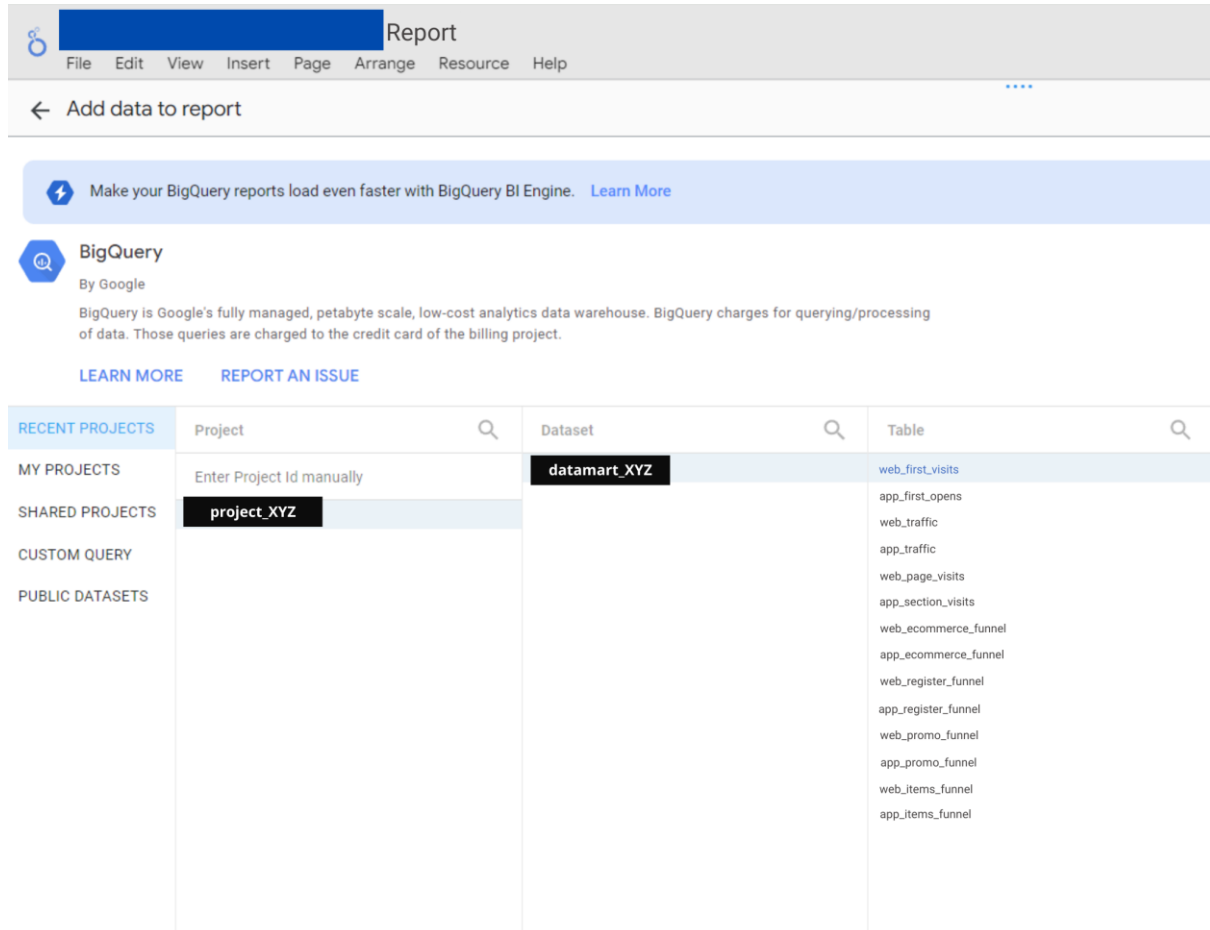
V Google Looker Studio bol vytvorený nový dashboard s názvom podľa potreby firmy. Pre vizualizáciu akýkoľvek dát je nutné pridať aspoň jeden dátový zdroj, pričom je možné ich pridávať aj dodatočne počas práce s reportom. Taktiež je nutný prístup pod rovnakým účtom, pod ktorým je Looker Studio report vytvorený. Na paneli nástrojov po zvolení *Resource* a následne možnosť *Manage added data sources* bol pridaný dátový zdroj prostredníctvom možnosti *Add a data source*. Pre účely tejto práce bol vybraný konektor pre dátový zdroj BigQuery, ktorý je jedným z 24 zabudovaných konektorov, ktorými nástroj disponuje.



Obrázok 39 – Napojenie cloudovej služby BigQuery do Looker Studia

(Zdroj: Vlastné spracovanie)

Následne je potrebné zvoliť konkrétny projekt, dataset a konečnú tabuľku s dátami, ktorá definuje štruktúru dát. Pre účely tejto práce bol vybraný projekt *project_XYZ*, dataset s názvom *datamart_XYZ* a jednotlivé dátové tabuľky, ktoré boli vytvorené v implementačnej časti.



Obrázok 40 – Napojenie dátových tabuliek dátového trhoviska do Looker Studia

(Zdroj: Vlastné spracovanie)

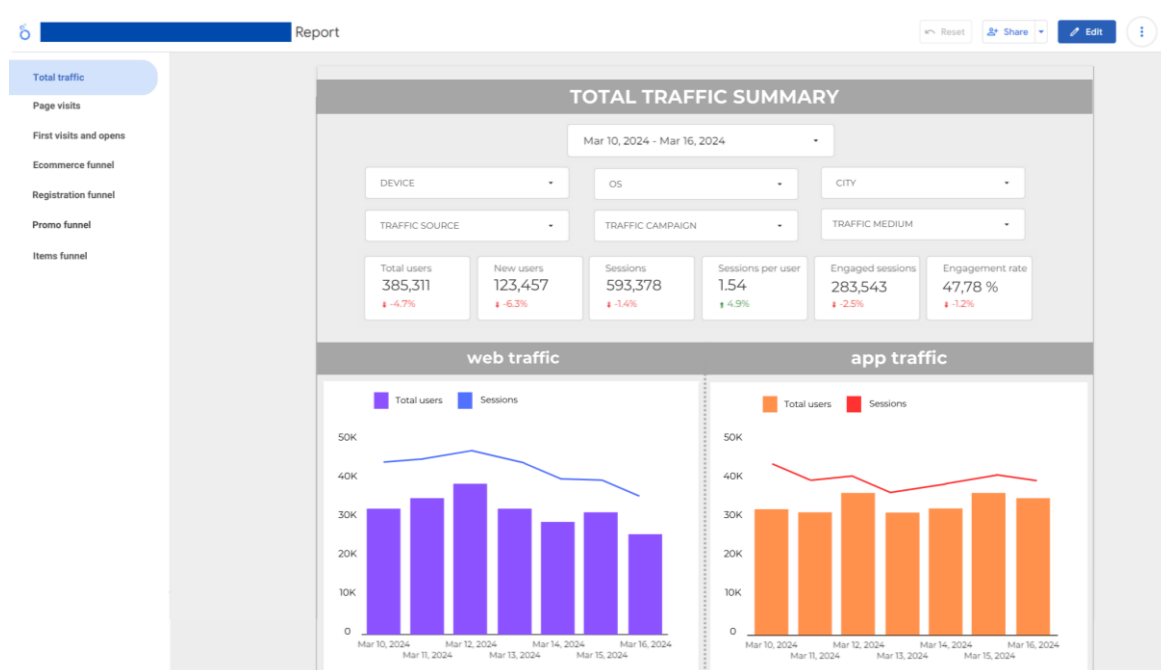
Správa jednotlivých zdrojov je možná pod tlačidlami *Resource, Manage added data sources*, kde možno každý zo zdrojov upravovať alebo zmazať aj počas aktívnej tvorby reportu, nachádza sa tu aj prehľad o tom v koľkých tabuľkách či grafoch bol zdroj použitý či jeho druh a status.

Pre kombinovanie rôznych zdrojov možno použiť *Blended data sources*. Tie sú užitočné, keď potrebné analyzovať dáta z rôznych zdrojov v jednej vizualizácii alebo tabuľke (napríklad dáta z BigQuery a Google Ads) alebo keď sú dáta uložené vo viacerých dátových tabuľkách.

7.2 Tvorba dashboardu

Vytvorenie reportingového riešenia, ktorý oddeľuje dáta pre aplikáciu a web, ale zároveň ich aj kombinuje, bolo dosiahnuté vytvorením viacerých sekcií alebo „kariet“ v rámci jedného reportu. Každá karta alebo sekcia je zameraná na určitú oblasť analýzy. Dashboard je štruktúrovaný na karty:

- **Všeobecná návštevnosť:** Táto karta obsahuje metriky z tabuliek web_traffic a app_traffic. Zobrazuje počet prvých návštev a otvorení, celkovú návštevnosť a porovnanie medzi návštevnosťou webovej stránky a aplikácie. Taktiež používa prepojenie s inými dátovými tabuľkami.



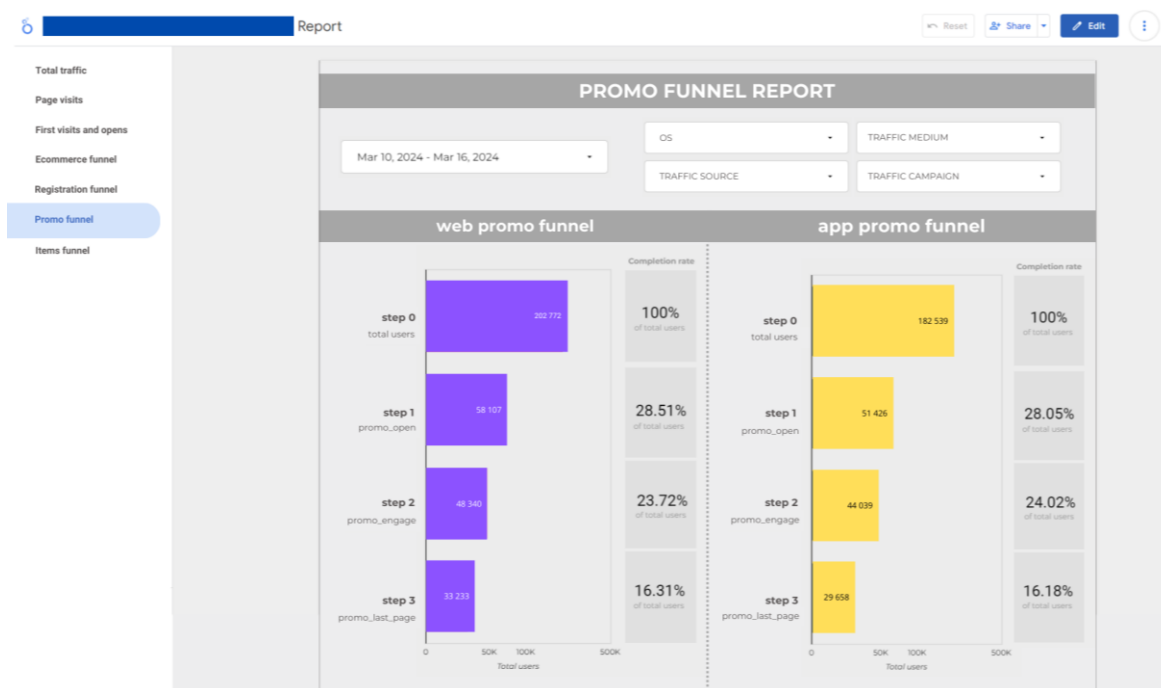
Obrázok 41 – Vizualizácia dát všeobecnej návštevnosti v Looker Studiu

(Zdroj: Vlastné spracovanie)

- **Podrobnosti o návštevnosti stránok/sekcii:** Táto karta používa dáta z tabuliek web_page_visits a app_section_visits na zobrazenie informácií o tom, ktoré stránky alebo sekcie sú najnavštevovanejšie.
- **Prvé návštevy a otvorenia aplikácie:** Táto karta obsahuje dáta z tabuliek web_first_visits a app_first_opens a to najmä z pohľadu marketingových kampaní a zdroja týchto prvých návštev.
- **Nákupný lievrik:** Táto karta kombinuje dáta z tabuliek web_ecommerce_funnel, app_ecommerce_funnel, web_items_funnel a app_items_funnel na vizualizáciu

nákupného lievika pre web a aplikáciu. Zobrazuje, koľko používateľov vytvorilo objednávku a aké produkty boli najpopulárnejšie. Taktiež používa prepojenie s inými dátovými tabuľkami.

- **Registrácia do vernostného programu:** Táto karta používa dáta z tabuliek web_register_funnel a app_register_funnel na zobrazenie, koľko používateľov sa zaregistrovalo do vernostného programu cez web alebo aplikáciu. Taktiež používa prepojenie s inými dátovými tabuľkami.
- **Akčné ponuky:** Táto karta používa dáta z tabuliek web_promotion_funnel a app_promotion_funnel na zobrazenie, koľko používateľov otvorilo akčnú ponuku na webe alebo v aplikácii. Taktiež používa prepojenie s inými dátovými tabuľkami.



Obrázok 42 – Náhl'ad reportu v karte s promo_funnel dátami

(Zdroj: Vlastné spracovanie)

- **Produkty a ich lievik:** Táto karta používa dáta z tabuliek web_items_funnel a app_items_funnel, zobrazuje prehľad produktov a ich zakúpenia, prezretia, či pridania do košíka. Taktiež používa prepojenie s inými dátovými tabuľkami.

Každá karta obsahuje vizualizácie, ako sú grafy, tabuľky alebo heatmaps, ktoré pomáhajú lepšie pochopiť jednotlivé dáta. V hornej časti každej karty sa nachádzajú filtre, ktoré umožňujú užívateľom zmeniť časový rámec alebo iné parametre analýzy, tieto filtre zahŕňajú:

- **Filter na dimenzie časového rámca:** Tento filter umožňuje užívateľom vybrať konkrétny časový rámec, za ktorý chcú vidieť dáta. Napríklad, môžu si vybrať, či chcú vidieť dáta za posledný týždeň, mesiac, rok atď.
- **Filter kampane:** Tento filter umožňuje užívateľom zobraziť dáta pre konkrétnu marketingovú kampaň.
- **Filter zdroja:** Tento filter umožňuje užívateľom zobraziť dáta podľa zdroja, odkiaľ pochádzajú návštevníci.
- **Filter média:** Tento filter umožňuje užívateľom zobraziť dáta podľa typu média, ktoré návštevníci použili na prístup k webovej stránke alebo aplikácii.
- **Filter zariadenia alebo operačného systému:** Tento filter umožňuje užívateľom zobraziť dáta podľa typu zariadenia či operačného systému, ktoré návštevníci použili na prístup k webovej stránke alebo aplikácii.

Tieto filtre môžu byť použité samostatne alebo v kombinácii, užívatelia tak môžu získať potrebné dáta pre ich analýzy.

ZÁVER

Teoretická časť diplomovej práce sa venuje oblasti Business Intelligence (BI), sú tu definované a rozobraté rôzne metódy a prístupy v tvorbe BI riešení, pričom sú uvedené ich výhody a nevýhody. Ďalej sú opísané jednotlivé komponenty, ktoré sú typické pre štruktúru BI riešení. Táto časť práce sa tiež zaoberá návrhom BI riešenia a poskytuje pochopenie základov dimenzionálneho modelovania, ktoré je esenciálne pre návrh dátových skladov.

V tejto časti je taktiež predstavená webová analytika a jej úloha v kontexte Business Intelligence, pričom je zdôraznený jej význam a možnosť prispieť k lepšiemu pochopeniu a optimalizácii online aktivít používateľov.

V závere teoretickej časti sú predstavené konkrétne nástroje a technológie pre BI riešenia v cloudovom prostredí, ktoré sú súčasťou riešenia v diplomovej práci. Medzi tieto nástroje patria Google Analytics 4, BigQuery a Looker Studio. Tieto nástroje sú súčasťou ekosystému Google a boli vybrané na základe existujúceho riešenia firmy a plánovaného budúceho riešenia, pre ktoré bol tento návrh vypracovaný.

Praktická časť práce začína predstavením firmy, pre ktorú bolo riešenie dátového skladu navrhované, jej webovej stránky a webovej aplikácie. Kapitola obsahuje popis súčasného stavu dát z dvoch platforiem a opis požiadaviek, ktoré toto riešenie pokrýva. Na základe týchto požiadaviek je vytvorený návrh štruktúry dátového skladu, ktorý obsahuje aj pôvodné komponenty predchádzajúceho riešenia. Návrh, ktorý táto práca opisuje, sa primárne sústredil na tvorbu dátového modelu dátového trhoviska. Tento návrh bol opísaný konceptuálnym a logickým modelom a následne realizovaný v prostredí BigQuery.

V časti implementácie sa pojednáva o tvorbe dátových tabuliek, ktoré sú naplnené historickými dátami od začiatku roka 2024 pomocou SQL dotazov. Následne sú vytvorené ETL príkazy, ktoré naplňajú dátové tabuľky dátového trhoviska novými dennými dátami. V závere praktickej časti bola vytvorená reportingová vrstva tohto riešenia v nástroji Looker Studio, kde boli napojené nové dátové trhoviská a následne bol vytvorený základný analytický report aplikácie a webovej stránky, ktorý má automatický tok aktuálnych dát.

V rámci diplomovej práce boli splnené všetky ciele stanovené v úvode práce. Prehľad jednotlivých čiastkových cieľov, spôsob ich naplnenia a umiestnenie naplnenia cieľa v práci sú uvedené v nasledujúcej tabuľke.

Tabuľka 6 – Prehľad čiastkových cieľov a ich naplnenia

(Zdroj: Vlastné spracovanie)

Názov čiastkového cieľa	Spôsob naplnenia	Kapitola
Zozbieranie a definícia požiadaviek riešenia.	Požiadavky firmy na riešenie boli zozbierané a definované.	4.3
Návrh dátového modelu dátového skladu a dátového trhoviska.	Bola vytvorená schéma nového riešenia dátového skladu a bol navrhnutý model dátového trhoviska z konceptuálneho a logického pohľadu.	5
Implementácia dátového trhoviska a tvorba ETL procesov.	Navrhované dátové tabuľky boli implementované v prostredí BigQuery a boli nastavené ETL procesy pre automatizáciu naplňovania dátového trhoviska aktuálnymi dátami.	6
Vytvorenie reportingovej vrstvy a integrácia dátového trhoviska.	Bola vytvorená reportingová vrstva vo vizualizačnom nástroji a základný automatizovaný analytický report, kde bolo napojené dátové trhovisko.	7

Implementácia BI riešenia v tejto firme prináša množstvo strategických a finančných prínosov. Na strategickom fronte, BI riešenie umožňuje lepšie pochopenie a interpretáciu dát, ktoré firma denne generuje. To vedie k informovanejšiemu rozhodovaniu na všetkých úrovniach organizácie. Okrem toho, implementácia BI riešenia pomáha identifikovať oblasti pre zlepšenie - či už ide o interné procesy, efektivitu práce alebo využitie zdrojov. A ako hlavnú výhodu, umožňuje lepšie pochopiť a analyzovať správanie zákazníkov, čo môže byť cenné pri tvorbe cielenejších marketingových stratégií a kampaní a zlepšenie kľúčových ukazovateľov výkonnosti.

Na finančnej strane, vytvorenie špecificky navrhnutého dátového trhoviska s viacerými datasetmi, ktoré obsahujú len tie najdôležitejšie informácie, vedie k výraznému zníženiu množstva dát, ktoré je potrebné prechádzať pri tvorbe ad hoc reportov. To nielenže znižuje čas potrebný na generovanie reportov, ale aj náklady na použitie BigQuery, ktorý účtuje poplatky na základe množstva spracovaných dát. Teda, čím menej dát je potrebné spracovať, tým nižšie sú poplatky za dotazovanie.

V kontexte tejto diplomovej práce, využitie BigQuery a Looker Studio ako súčasť BI riešenia umožňuje firme efektívne využívať výhody cloudových technológií a pokročilých analytických nástrojov. Vďaka menším, špecificky navrhnutým datasetom, je možné v Looker Studiu generovať reporty rýchlejšie, čo zvyšuje efektivitu práce s dátami. Previazanosť nástrojov a technológií spoločnosti Google, LLC umožňuje jednoduchosť pri využívaní viacerých takýchto nástrojov.

Implementácia tohto BI riešenia tak prináša firme nielen lepšie pochopenie svojich dát a informácií, ale aj významné finančné úspory vďaka efektívnejšiemu využitiu cloudových technológií a pokročilých analytických nástrojov.

Zoznam použitej literatúry

AMARESAN, Swetha. *Everything You Need to Know about Customer Lifecycle Management*. Online. Dostupné z: <https://blog.hubspot.com/service/customer-lifecycle-management>. [cit. 2024-05-01].

BAPNA, Himan. *BigQuery Table Schema*. Online. Dostupné z: <https://medium.com/@himan.bapna/bigquery-table-schema-fd3c25eb8d39>. [cit. 2024-05-01].

BEKAVAC, Ivan a Daniela Garbin PRANICEVIC, 2015. *Web analytics tools and web metrics tools: An overview and comparative analysis*. Croatian Operational Research Review; Zagreb. Online. 6(2), 373–386. ISSN 18480225. Dostupné z: <https://hrcak.srce.hr/file/218174>

BRUCE CLAY EUROPE. *The Customer Life Cycle*. Online. Dostupné z: <https://www.bruceclay.com/eu/analytics/customerlifecycle/>. [cit. 2024-05-01].

BRUNEC, J. *Google analytics*. Průvodce (Grada). Praha: Grada Publishing, 2017. ISBN 978-80-271-0338-6.

BURTON, Colleen. *Understanding Data Streams In Google Analytics 4*. Online. Data Driven U. 2024. Dostupné z: <https://datadrivenu.com/understanding-data-streams-google-analytics-4/>. [cit. 2024-04-29].

COFFMAN, Liz. *What is the CRM process? 5 key steps*. Online. Dostupné z: <https://www.zendesk.com/in/blog/crm-process/>. [cit. 2024-05-01].

DENNIS, Amber. *The Data Warehouse: From the Past to the Present*. Online. Dostupné z: <https://www.dataversity.net/data-warehouse-past-present/>. [cit. 2024-04-30].

DIGITAL MARKETING INSTITUTE. *Lifecycle Marketing vs. Sales Funnel: Which Works Best?* Online. Dostupné z: <https://digitalmarketinginstitute.com/blog/lifecycle-marketing-sales-funnel-which-works-best>. [cit. 2024-05-01].

EDMONDSON, Mark. *Learning Google Analytics: Creating Business Impact and Driving Insights*. Oreilly Media, 2022. ISBN 9781098113087.

FARAZ, Muhammad. *Load Data into BigQuery: Easy Step-by-Step Guide*. Online. Dostupné z: <https://hevodata.com/learn/load-data-into-bigquery/>. [cit. 2024-05-01].

FARRUH, Kushnazarov. *Consumer Life Cycle and Profiling: A Data Mining Perspective*. Online. In: REYES, Matthew (ed.). *Consumer Behavior and Marketing*. IntechOpen, 2020. ISBN 978-1-78923-855-6. Dostupné z: <https://doi.org/10.5772/intechopen.85407>. [cit. 2024-05-05].

FOY, Peter. *15+ Most Important Ecommerce Metrics to Measure*. Online. Dostupné z: <https://agencyanalytics.com/blog/ecommerce-analytics-metrics-to-measure-success>. [cit. 2024-05-01].

GOOGLE ANALYTICS HELP. *[GA4] About Events*. Online. 2024. Dostupné z: <https://support.google.com/analytics/topic/13367566?hl=en&sjid=13502373747430386772-EU>. [cit. 2024-04-29].

GOOGLE ANALYTICS HELP. *[GA4] Automatically collected events*. Online. 2024. Dostupné z: <https://support.google.com/analytics/answer/9234069?hl=en>. [cit. 2024-04-29].

GOOGLE ANALYTICS HELP. *[GA4] Google Analytics 360 (Google Analytics 4 Properties)*. Online. 2024. Dostupné z: <https://support.google.com/analytics/answer/11202874?hl=en>. [cit. 2024-04-29].

GOOGLE CLOUD. *BigQuery APIs and libraries overview*. Online. 2024. Dostupné z: <https://cloud.google.com/bigquery/docs/reference/libraries-overview>. [cit. 2024-04-29].

GOOGLE CLOUD. *Introduction to Looker*. Online. 2024. Dostupné z: <https://cloud.google.com/looker/docs/intro>. [cit. 2024-04-29].

GOOGLE CLOUD. *Introduction to LookML*. Online. 2024. Dostupné z: <https://cloud.google.com/looker/docs/what-is-lookml>. [cit. 2024-04-30].

GOOGLE CLOUD. *Scheduling queries*. Online. 2024. Dostupné z: <https://cloud.google.com/bigquery/docs/scheduling-queries>. [cit. 2024-04-29].

INMON, William H. *Building the data warehouse*. 3rd ed. New York: J. Wiley, c2002. ISBN 978-0471081302.

JANSEN, Bernard J.; JUNG, Soon-gyo; SALMINEN, Joni a SULEMAN, Hussein. *Measuring user interactions with websites: A comparison of two industry standard analytics approaches using data of 86 websites*. Online. PLOS ONE. 2022, roč. 17, č. 5. ISSN 1932-6203. Dostupné z: <https://doi.org/10.1371/journal.pone.0268212>. [cit. 2024-05-12].

JOSHI, Amar. *Understanding the Differences Between Inmon, Kimball, and Data Vault Data Models*. Online. Dostupné z: Medium, <https://medium.com/@amarrjoshi/understanding-the-differences-between-inmon-kimball-and-data-vault-data-models-79868aa99525>. [cit. 2024-05-05].

KAUSHIK, A. *Web Analytics: 2.0 The Art of Online Accountanility and Science of Customer Centricity*. 2nd ed. Indianapolis: Sybex, 2009. 475 p. ISBN 978-0-470-52939-3.

KIMBALL, Ralph., CASERTA, Joe. *The data warehouse ETL toolkit practical techniques for extracting, cleaning, conforming, and delivering data*. Indianapolis, IN: Wiley, 2004. ISBN 0764579231.

KIRSCH, Katrina. *How Conversion Funnels Create a Better Customer Journey + How to Optimize Yours*. Online. Dostupné z: <https://blog.hubspot.com/marketing/conversion-funnel>. [cit. 2024-05-01].

LABERGE, R. *Datové sklady - Agilní metody a Business Intelligence*. 1. vyd. Praha: Computer Press, 2012. 352 s. ISBN 978-80-251-3729-1.

LOKKEN, Bob. *Business Intelligence: An Intelligent Move or Not?* Online. ProClarity Corporation. Dostupné z: <http://bi.ittoolbox.com/browse.asp?c=BIPeerPublishing&r=/pub/AO031202.pdf>. [cit. 2024-05-12].

MANGOLD, Benjamin. *Creating Visually Engaging Reports with Looker Studio: A Step-by-Step Guide*. Online. Loves Data. Dostupné z: <https://www.lovesdata.com/blog/creating-visually-engaging-reports-with-looker-studio-a-step-by-step-guide>. [cit. 2024-04-29].

MANGOLD, Benjamin. *Google Analytics 4 Event Tracking: Your Complete Guide*. Online. Loves Data. Dostupné z: <https://www.lovesdata.com/blog/google-analytics-4-events>. [cit. 2024-05-01].

MATOMO. *What is Funnel Analysis? A Complete Guide for Quick Results*. Online. Dostupné z: <https://matomo.org/blog/2024/01/funnel-analysis/>. [cit. 2024-04-29].mat

MCDANIEL, Zachary. *The Ins and Outs of Customer Lifecycle Marketing*. Online. Dostupné z: <https://encharge.io/customer-lifecycle-marketing/>. [cit. 2024-05-01].

MORGAN, Edward. *How To Optimize Your Lifecycle Marketing Strategy*. Online. Dostupné z: <https://www.constantcontact.com/blog/lifecycle-marketing/>. [cit. 2024-05-01].

MOSCOSO-ZEA, Oswaldo; PAREDES-GUALTOR, Joel a LUJAN-MORA, Sergio. *A Holistic View of Data Warehousing in Education*. Online. IEEE Access. 2018, roč. 6, s. 64659-64673. ISSN 2169-3536. Dostupné z: <https://doi.org/10.1109/ACCESS.2018.2876753>. [cit. 2024-05-05].

MUKHERJEE, Subhabrata. *Inmon & Kimball: Datawarehousing & their views!*. Online. Dostupné z: <https://medium.com/analytics-vidhya/inmon-kimball-datawarehousing-their-views-579291acdf08>. [cit. 2024-05-01].

NOVOTNÝ, O.; POUR, J. a SLÁNSKÝ, D. *Business Intelligence: Jak využít bohatství ve vašich datech*. 1. vyd. Praha: Grada, 2005. 256 s. ISBN 80-247-1094-3.

NOVOTNÝ, O.; POUR, J.; MARYŠKA, M. a BASL, J. *Řízení výkonnosti podnikové informatiky*. 1. vyd. Praha: Professional Publishing, 2010. 275 s. ISBN 978-80-7431-040-9.

NYBERG, Tommi. *Google Looker Simplifies Business Intelligence in the Cloud*. Online. The Cloud People. 2022. Dostupné z: <https://www.thecloudpeople.com/sv/blog/google-looker-simplifies-business-intelligence-in-the-cloud>. [cit. 2024-04-29].

Online marketing. Brno: Computer Press, 2014. ISBN 978-80-251-4155-7.

PEARCE, Phil. *Using Google Analytics 4 with BigQuery: Definitive Guide*. Online. Measure Minds. 2021. Dostupné z: <https://measuremindsgroup.com/bigquery-with-google-analytics-4>. [cit. 2024-04-29].

PITTMANN, Christopher. *Advanced Guide to Google Analytics 4: Improve your online sales*. SMP Publishing, 2022. ISBN 9798471742178.

POUR, Jan, Miloš MARYŠKA, Iva STANOVSKÁ a Zuzana ŠEDIVÁ. *Self service Business Intelligence: Jak si vytvořit vlastní analytické, plánovací a reportingové aplikace*. Praha: Grada Publishing, 2018. Management v informační společnosti. ISBN 978-80-271-0616-5.

SAHA, Dipan. *Harnessing the Power of Scheduled Queries in BigQuery*. Online. Dostupné z: <https://medium.com/@dipan.saha/google-bigquery-schedule-your-queries-7ceabb50592e>. [cit. 2024-04-29].

SHARMA, Himanshu. *events_ & events_intraday_ tables in BigQuery for GA4*. Online. Dostupné z: https://www.optimizesmart.com/event-events_intraday-data-tables-bigquery-for-ga4-google-analytics-4/. [cit. 2024-05-01].

SHARMA, Himanshu. *GA4 User Properties (User Scoped Dimensions) – Tutorial*. Online. OptimizeSmart. Dostupné z: <https://www.optimizesmart.com/how-to-create-and-use-user-properties-in-ga4/>. [cit. 2024-05-01].

SCHEPS, Swain, 2008. *Business intelligence for dummies*. Hoboken: Wiley. ISBN 978-0470127230.

SCHMIEDICKER, Nick. *Customer lifecycle*. Online. Dostupné z: <https://encharge.io/customer-lifecycle-marketing/>. [cit. 2024-05-01].

SILVA, Mariano. *Harnessing the Power of Scheduled Queries in BigQuery*. Online. Dostupné z: <https://medium.com/@mhsilvav/harnessing-the-power-of-scheduled-queries-in-bigquery-a2680720d599>. [cit. 2024-04-29].

SLÁNSKÝ D. *Data a analytika pro 21. století*. 1. vyd. Praha: Professional Publishing, 2018. 5 svazků. ISBN 978-80-88260-25-7.

SOPA, Faton. *Google Analytics 4 for Higher Education: A Comprehensive 2023 Guide*. Online. Manaferra. Dostupné z: <https://www.manaferra.com/google-analytics-4-for-universities/>. [cit. 2024-05-01].

THALLAM, Rajesh. *BigQuery Explained: Data Ingestion*. Online. Dostupné z: <https://medium.com/google-cloud/bigquery-explained-data-ingestion-cdc26a588d0>. [cit. 2024-05-01].

THALLAM, Rajesh. *BigQuery Explained: Storage Overview*. Online. Dostupné z: <https://medium.com/google-cloud/bigquery-explained-storage-overview-70cac32251fa>. [cit. 2024-05-01].

TVRDÍKOVÁ, Milena. *Aplikace moderních informačních technologií v řízení firmy: nástroje ke zvyšování kvality informačních systémů*. Management v informační společnosti. Praha: Grada, 2008. ISBN 978-80-247-2728-8.

YESSAD, Lamia a LABIOD, Aissa. *Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault*. Online. In: 2016 International Conference on System Reliability and Science (ICSRS). IEEE, 2016, s. 95-99. ISBN 978-1-5090-3277-8. Dostupné z: <https://doi.org/10.1109/ICSRS.2016.7815845>. [cit. 2024-05-05].

Zoznam použitých obrázkov

Obrázok 1 – Obecná koncepcia architektúry BI	14
Obrázok 2 – Diagram ETL procesu	16
Obrázok 3 – Schéma tvorby dátového skladu podľa Kimballovej metodológie.....	21
Obrázok 4 – Schéma tvorby dátového skladu podľa Inmonovej metodológie	23
Obrázok 5 – Fázy zákaznickeho životného cyklu	32
Obrázok 6 – Hierarchia entít v dátovom modeli GA4	35
Obrázok 7 – Diagram udalostí a ich parametrov	37
Obrázok 8 – Schéma štruktúry BigQuery	38
Obrázok 9 – Diagram ukladania dát v relačných databázach a BigQuery	40
Obrázok 10 – Plánované dotazy a ich tvorba v BigQuery	41
Obrázok 11 – Prehľad zabudovaných konektorov v Looker Studiu	42
Obrázok 12 – Wireframe webovej stránky firmy.....	47
Obrázok 13 – Wireframe aplikácie firmy.....	49
Obrázok 14 – Návrh dátového skladu	54
Obrázok 15 – Náhľad štruktúry exportu dát do BigQuery.....	56
Obrázok 16 – Schéma konceptuálneho dátového modelu dátového trhoviska.....	60
Obrázok 17 – Schéma nákupného lievika a jeho jednotlivé fázy.....	61
Obrázok 18 – Schéma nákupného lievika v rámci produktov.....	62
Obrázok 19 – Schéma registračného lievika	62
Obrázok 20 – Schéma lievika prezerania akčnej ponuky.....	63
Obrázok 21 – Schéma čiastočného logického dátového modelu dátového trhoviska pre webové dáta (1/2).....	64
Obrázok 22 – Východiskový stav dátového skladu v BigQuery	68
Obrázok 23 – Dátový sklad v prostredí BigQuery	69
Obrázok 24 – Dočasné dátové úložisko v prostredí BigQuery	69
Obrázok 25 – SQL dotaz pre získanie dát do tabuľky web_first_visits.....	71
Obrázok 26 – SQL dotaz pre získanie dát do tabuľky web_traffic.....	72
Obrázok 27 – SQL dotaz pre získanie dát do tabuľky web_page_visits.....	73
Obrázok 28 – Vzor SQL dotazu pre získanie dát do tabuľky web_ecommerce_funnel	74
Obrázok 29 – Vzor SQL dotazu pre získanie dát do tabuľky web_register_funnel	75
Obrázok 30 – Vzor SQL dotazu pre získanie dát do tabuľky web_promo_funnel	76

Obrázok 31 – Vzor SQL dotazu pre získanie dát do tabuľky web_items_funnel.....	77
Obrázok 32 – Ukladanie získaných dát SQL dotazov do novej tabuľky	78
Obrázok 33 – Dátové trhovisko s vytvorenými tabuľkami	78
Obrázok 34 – Vytvorená tabuľka web_first_visits a jej schéma v dátovom trhovisku.....	79
Obrázok 35 – Časová úprava podmienky v SQL dotazoch pre ETL proces.....	79
Obrázok 36 – Konfigurácia ETL procesu pre tabuľku web_first_visits	80
Obrázok 37 – SQL príkaz pre vymazanie potencionálnych duplicitných dát web_first_visits	81
Obrázok 38 – Konfigurácia plánovaného dotazu pre zmazanie dát pre tabuľku web_first_visits	81
Obrázok 39 – Napojenie cloudovej služby BigQuery do Looker Studia.....	82
Obrázok 40 – Napojenie dátových tabuliek dátového trhoviska do Looker Studia.....	83
Obrázok 41 – Vizualizácia dát všeobecnej návštevnosti v Looker Studiu	84
Obrázok 42 – Náhľad reportu v karte s promo_funnel dátami	85

Zoznam použitých tabuliek

Tabuľka 1 – Výhody a nevýhody Kimballovej metodológie.....	22
Tabuľka 2 – Výhody a nevýhody Inmonovej metodológie.....	24
Tabuľka 3 – Prehľad základných udalostí a typov v GA4.....	36
Tabuľka 4 – Prehľad významu kontrolných celkov v BigQuery	39
Tabuľka 5 – Prehľad navrhovaných dátových tabuliek dátového trhoviska.....	58
Tabuľka 6 – Prehľad čiastkových cieľov a ich naplnenia	88

Zoznam príloh

Príloha I: Schéma čiastočného logického dátového modelu dátového trhoviska pre aplikačné dáta (2/2)	100
Príloha II: Snímka obrazovky SQL dotazu pre app_first_opens.....	101
Príloha III: Snímka obrazovky SQL dotazu pre app_traffic	101
Príloha IV: Snímka obrazovky SQL dotazu pre app_section_visits	102
Príloha V: Snímky obrazovky SQL dotazu pre web_ecommerce_funnel	102
Príloha VI: Snímky obrazovky SQL dotazu pre app_ecommerce_funnel	107
Príloha VII: Snímky obrazovky SQL dotazu pre web_register_funnel	111
Príloha VIII: Snímky obrazovky SQL dotazu pre app_register_funnel.....	113
Príloha IX: Snímky obrazovky SQL dotazu pre web_promo_funnel	114
Príloha X: Snímky obrazovky SQL dotazu pre app_promo_funnel	115
Príloha XI: Snímky obrazovky SQL dotazu pre web_items_funnel.....	117
Príloha XII: Snímky obrazovky SQL dotazu pre app_items_funnel	121

Prílohy

Príloha I: Schéma čiastočného logického dátového modelu dátového trhoviska pre aplikáciu dáta (2/2)

app_first_opens		app_ecommerce_funnel		app_items_funnel	
event_date	DÁTUM	event_date	DÁTUM	event_date	DÁTUM
user_pseudo_id	TEXT	event_name	TEXT	event_name	TEXT
user_id	TEXT	step_number	TEXT	step_number	TEXT
session_id	ČÍSLO	user_pseudo_id	TEXT	item_id	TEXT
screen_name	TEXT	user_id	TEXT	item_name	TEXT
screen_class	TEXT	user_logged	BOOLEAN	item_brand	TEXT
platform	TEXT	session_id	ČÍSLO	item_category	TEXT
device	TEXT	screen_name	TEXT	item_category2	TEXT
os	TEXT	screen_class	TEXT	item_category3	TEXT
country	TEXT	previous_screen	TEXT	price	ČÍSLO
city	TEXT	platform	TEXT	quantity	ČÍSLO
first_traffic_source	TEXT	device	TEXT	item_revenue	ČÍSLO
first_traffic_medium	TEXT	os	TEXT	transaction_id	TEXT
first_traffic_campaign	TEXT	country	TEXT	user_pseudo_id	TEXT
		city	TEXT	user_id	TEXT
		traffic_source	TEXT	session_id	ČÍSLO
		traffic_medium	TEXT	user_logged	BOOLEAN
		traffic_campaign	TEXT	screen_name	TEXT
		transaction_id	TEXT	screen_class	TEXT
		purchase_revenue	ČÍSLO	previous_screen	TEXT
		total_item_quantity	ČÍSLO	platform	TEXT
		shipping_value	ČÍSLO	device	TEXT
		unique_items	ČÍSLO	country	TEXT
				city	TEXT
				traffic_source	TEXT
				traffic_medium	TEXT
				traffic_campaign	TEXT
				item_views	ČÍSLO
				item_add_to_carts	ČÍSLO
				item_purchases	ČÍSLO
app_section_visits		app_register_funnel		app_promo_funnel	
event_date	DÁTUM	event_date	DÁTUM	event_date	DÁTUM
user_pseudo_id	TEXT	event_name	TEXT	event_name	TEXT
user_id	TEXT	step_number	TEXT	step_number	STRING
user_logged	BOOLEAN	user_pseudo_id	TEXT	user_pseudo_id	TEXT
session_id	ČÍSLO	user_id	TEXT	user_id	STRING
screen_name	TEXT	user_logged	BOOLEAN	user_logged	BOOLEAN
screen_class	TEXT	session_id	ČÍSLO	session_id	ČÍSLO
previous_screen	TEXT	screen_name	TEXT	screen_name	TEXT
engagement_time_sec	ČÍSLO	screen_class	TEXT	screen_class	TEXT
platform	TEXT	previous_screen	TEXT	previous_screen	TEXT
device	TEXT	platform	TEXT	platform	TEXT
os	TEXT	device	TEXT	device	TEXT
country	TEXT	os	TEXT	os	TEXT
city	TEXT	country	TEXT	country	TEXT
traffic_source	TEXT	city	TEXT	city	TEXT
traffic_medium	TEXT	traffic_source	TEXT	traffic_source	TEXT
traffic_campaign	TEXT	traffic_medium	TEXT	traffic_medium	TEXT
		traffic_campaign	TEXT	traffic_campaign	TEXT
app_traffic					
event_date	DÁTUM				
event_name	TEXT				
user_pseudo_id	TEXT				
user_id	TEXT				
session_id	ČÍSLO				
user_logged	BOOLEAN				
session_engaged	TEXT				
engagement_time_sec	ČÍSLO				
landing_screen_class	TEXT				
landing_screen_name	TEXT				
platform	TEXT				
device	TEXT				
os	TEXT				
country	TEXT				
city	TEXT				
traffic_source	TEXT				
traffic_medium	TEXT				
traffic_campaign	TEXT				

Príloha II: Snímka obrazovky SQL dotazu pre app_first_opens

```
SELECT
  CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  user_pseudo_id,
  user_id,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
  END AS first_traffic_source,
  traffic_source.medium AS first_traffic_medium,
  traffic_source.name AS first_traffic_campaign
FROM
  `project_XYZ.analytics_XYZ.events_*`
WHERE
  event_name = "first_open"
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
```

Príloha III: Snímka obrazovky SQL dotazu pre app_traffic

```
SELECT
  CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  'session_aggregated' AS event_name,
  user_pseudo_id,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = 'ga_session_id') AS session_id,
  MAX(user_id) AS user_id,
  MAX(CASE WHEN event_name = 'user_logged' THEN TRUE ELSE FALSE END) AS user_logged,
  MAX((SELECT value.string_value FROM UNNEST(event_params)
  | WHERE KEY = 'session_engaged')) AS session_engaged,
  SUM((SELECT value.int_value FROM UNNEST(event_params)
  | WHERE KEY = 'engagement_time_msec'))/1000 AS engagement_time_sec,
  CASE
    WHEN (SELECT value.int_value FROM UNNEST(event_params)
    | WHERE event_name = 'screen_view' AND KEY = 'entrances') = 1
    THEN (SELECT value.string_value FROM UNNEST(event_params)
    | WHERE event_name = 'screen_view' AND KEY = 'firebase_screen')
  END AS landing_screen_name,
  CASE
    WHEN (SELECT value.int_value FROM UNNEST(event_params)
    | WHERE event_name = 'screen_view' AND KEY = 'entrances') = 1
    THEN (SELECT value.string_value FROM UNNEST(event_params)
    | WHERE event_name = 'screen_view' AND KEY = 'firebase_screen_class')
  END AS landing_screen_class,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
  END AS traffic_source,
  traffic_source.medium AS traffic_medium,
  traffic_source.name AS traffic_campaign,
  COUNTIF (event_name = 'screen_view') as screen_views
FROM
  `project_XYZ.analytics_XYZ.events_*`
WHERE
  platform = "web"
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
GROUP BY
  event_date, user_pseudo_id, session_id, device, os,
  country, city, traffic_source, traffic_medium, traffic_campaign, platform, landing_screen_class,
  landing_screen_name
```

Príloha IV: Snímka obrazovky SQL dotazu pre app_section_visits

```
SELECT
  CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  user_pseudo_id,
  user_id,
  (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
  SUM((SELECT value.int_value FROM UNNEST(event_params)
  | WHERE KEY = 'engagement_time_msec')/1000 AS engagement_time_sec,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
  END AS traffic_source,
  traffic_source.medium AS traffic_medium,
  traffic_source.name AS traffic_campaign
FROM
  `project_XYZ.analytics_XYZ.events_*`
WHERE
  event_name = "screen_view"
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
```

Príloha V: Snímky obrazovky SQL dotazu pre web_ecommerce_funnel

```
SELECT
  CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  event_name,
  "step 1" as step_number,
  user_pseudo_id,
  user_id,
  (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
  REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
  | WHERE KEY = 'page_location'), r'([^\?]*)) AS page_location_stripped,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
  END AS traffic_source,
  traffic_source.medium AS traffic_medium,
  traffic_source.name AS traffic_campaign,
  ecommerce.transaction_id AS transaction_id,
  ecommerce.purchase_revenue AS purchase_revenue,
  ecommerce.total_item_quantity AS total_item_quantity,
  ecommerce.shipping_value AS shipping_value,
  ecommerce.unique_items AS unique_items
FROM
  `project_XYZ.analytics_XYZ.events_*`
```

```

WHERE
    platform = "web"
    AND event_name = "view_item_list"
    AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
    event_name,
    "step 2" as step_number,
    user_pseudo_id,
    user_id,
    (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
    (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
    REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
    | WHERE KEY = 'page_location'), r'([^\?]*)) AS page_location_stripped,
    platform,
    device.category AS device,
    device.operating_system AS os,
    geo.country AS country,
    geo.city AS city,
    CASE
        WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
        WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
        ELSE traffic_source.source
    END AS traffic_source,
    traffic_source.medium AS traffic_medium,
    traffic_source.name AS traffic_campaign,
    ecommerce.transaction_id AS transaction_id,
    ecommerce.purchase_revenue AS purchase_revenue,
    ecommerce.total_item_quantity AS total_item_quantity,
    ecommerce.shipping_value AS shipping_value,
    ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
    platform = "web"
    AND event_name = "view_item"
    AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
    event_name,
    "step 3" as step_number,
    user_pseudo_id,
    user_id,
    (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
    (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,

```



```

REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
| WHERE KEY = 'page_location'), r'([^\?]*') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
| WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
| WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
| ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "add_to_cart"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 4" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
| WHERE KEY = 'page_location'), r'([^\?]*') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
| WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
| WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
| ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "view_cart"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,

```

```

event_name,
"step 5" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
| WHERE KEY = 'page_location'), r'([?]*)') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
| WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
| WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
| ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items

```

FROM

```
`project_XYZ.analytics_XYZ.events_*
```

WHERE

```

platform = "web"
AND event_name = "begin_checkout"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

```

UNION ALL

SELECT

```

CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 6" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
| WHERE KEY = 'page_location'), r'([?]*)') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
| WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
| WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
| ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items

```

FROM

```
`project_XYZ.analytics_XYZ.events_*
```

```

WHERE
    platform = "web"
    AND event_name = "add_shipping_info"
    AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
    CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
    event_name,
    "step 7" as step_number,
    user_pseudo_id,
    user_id,
    (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
    (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
    REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
        | WHERE KEY = 'page_location'), r'([?]*)') AS page_location_stripped,
    platform,
    device.category AS device,
    device.operating_system AS os,
    geo.country AS country,
    geo.city AS city,
    CASE
        WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
        WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
        ELSE traffic_source.source
    END AS traffic_source,
    traffic_source.medium AS traffic_medium,
    traffic_source.name AS traffic_campaign,
    ecommerce.transaction_id AS transaction_id,
    ecommerce.purchase_revenue AS purchase_revenue,
    ecommerce.total_item_quantity AS total_item_quantity,
    ecommerce.shipping_value AS shipping_value,
    ecommerce.unique_items AS unique_items

FROM
    `project_XYZ.analytics_XYZ.events_*`
WHERE
    platform = "web"
    AND event_name = "add_payment_info"
    AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
    CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
    event_name,
    "step 8" as step_number,
    user_pseudo_id,
    user_id,
    (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
    (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
    REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
        | WHERE KEY = 'page_location'), r'([?]*)') AS page_location_stripped,
    platform,
    device.category AS device,
    device.operating_system AS os,
    geo.country AS country,
    geo.city AS city,
    CASE
        WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
        WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
        ELSE traffic_source.source
    END AS traffic_source,
    traffic_source.medium AS traffic_medium,
    traffic_source.name AS traffic_campaign,
    ecommerce.transaction_id AS transaction_id,
    ecommerce.purchase_revenue AS purchase_revenue,
    ecommerce.total_item_quantity AS total_item_quantity,
    ecommerce.shipping_value AS shipping_value,
    ecommerce.unique_items AS unique_items

FROM
    `project_XYZ.analytics_XYZ.events_*`
WHERE
    platform = "web"
    AND event_name = "purchase"
    AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

```


Príloha VI: Snímky obrazovky SQL dotazu pre app_ecommerce_funnel

```
SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  event_name,
  "step 1" as step_number,
  user_pseudo_id,
  user_id,
  (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
  END AS traffic_source,
  traffic_source.medium AS traffic_medium,
  traffic_source.name AS traffic_campaign,
  ecommerce.transaction_id AS transaction_id,
  ecommerce.purchase_revenue AS purchase_revenue,
  ecommerce.total_item_quantity AS total_item_quantity,
  ecommerce.shipping_value AS shipping_value,
  ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
  platform != "web"
  AND event_name = "view_item_list"
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  event_name,
  "step 2" as step_number,
  user_pseudo_id,
  user_id,
  (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
  END AS traffic_source,
  traffic_source.medium AS traffic_medium,
  traffic_source.name AS traffic_campaign,
  ecommerce.transaction_id AS transaction_id,
  ecommerce.purchase_revenue AS purchase_revenue,
  ecommerce.total_item_quantity AS total_item_quantity,
  ecommerce.shipping_value AS shipping_value,
  ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
```



```

WHERE
    platform != "web"
    AND event_name = "view_item"
    AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
    event_name,
    "step 3" as step_number,
    user_pseudo_id,
    user_id,
    (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
    (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
    platform,
    device.category AS device,
    device.operating_system AS os,
    geo.country AS country,
    geo.city AS city,
    CASE
        WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
        WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
        ELSE traffic_source.source
    END AS traffic_source,
    traffic_source.medium AS traffic_medium,
    traffic_source.name AS traffic_campaign,
    ecommerce.transaction_id AS transaction_id,
    ecommerce.purchase_revenue AS purchase_revenue,
    ecommerce.total_item_quantity AS total_item_quantity,
    ecommerce.shipping_value AS shipping_value,
    ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
    platform != "web"
    AND event_name = "add_to_cart"
    AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
    event_name,
    "step 4" as step_number,
    user_pseudo_id,
    user_id,
    (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
    (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
    platform,
    device.category AS device,
    device.operating_system AS os,
    geo.country AS country,
    geo.city AS city,
    CASE
        WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
        WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
        ELSE traffic_source.source
    END AS traffic_source,

```

```

traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "view_cart"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL
SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 5" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY = "isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,|
CASE
WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "begin_checkout"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL
SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 6" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY = "isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
platform,

```

```

device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
  `project_XYZ.analytics_XYZ.events_*`
WHERE
  platform != "web"
  AND event_name = "add_shipping_info"
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL
SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  event_name,
  "step 7" as step_number,
  user_pseudo_id,
  user_id,
  (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
  `project_XYZ.analytics_XYZ.events_*`
WHERE
  platform != "web"
  AND event_name = "add_payment_info"
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL
SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  event_name,
  "step 8" as step_number,
  user_pseudo_id,
  user_id,

```



```

(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "purchase"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

```

Príloha VII: Snímky obrazovky SQL dotazu pre web_register_funnel

```

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 1" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
  WHERE KEY = 'page_location'), r'([?]*)) AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "register_start"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 2" as step_number,
user_pseudo_id,
user_id,

```

```

(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
| WHERE KEY = 'page_location'), r'([^\?]*)') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
| WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
| WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
| ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "register_add_info"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 3" as step_number,
user_pseudo_id,
user_id,

(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
| WHERE KEY = 'page_location'), r'([^\?]*)') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
| WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
| WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
| ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "register_success"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

```

Príloha VIII: Snímky obrazovky SQL dotazu pre app_register_funnel

```
SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 1" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "register_start"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 2" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "register_add_info"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 3" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
```

```

platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "register_success"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

```

Príloha IX: Snímky obrazovky SQL dotazu pre web_promo_funnel

```

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 1" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY = "isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
  WHERE KEY = 'page_location'), r'([^\?]*)') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "promo_open"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL
SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 2" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY = "isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
  WHERE KEY = 'page_location'), r'([^\?]*)') AS page_location_stripped,

```



```

platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "promo_engage"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 3" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
  WHERE KEY = 'page_location'), r'([?]*)') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "promo_last_page"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

```

Príloha X: Snímky obrazovky SQL dotazu pre app_promo_funnel

```

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 2" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,

```



```

CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "promo_open"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL
SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 3" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "promo_engage"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL
SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 4" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "promo_last_page"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

```

Príloha XI: Snímky obrazovky SQL dotazu pre web_items_funnel

```
SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  event_name,
  "step 1" as step_number,
  user_pseudo_id,
  user_id,
  (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
  REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
  | WHERE KEY = 'page_location'), r'([?]*)') AS page_location_stripped,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
  | WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  | WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  | ELSE traffic_source.source
  END AS traffic_source,
  traffic_source.medium AS traffic_medium,
  traffic_source.name AS traffic_campaign,
  ecommerce.transaction_id AS transaction_id,
  ecommerce.purchase_revenue AS purchase_revenue,
  ecommerce.total_item_quantity AS total_item_quantity,
  ecommerce.shipping_value AS shipping_value,
  ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
  platform = "web"
  AND event_name = "view_item_list"
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL
SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  event_name,
  "step 2" as step_number,
  user_pseudo_id,
  user_id,
  (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
  REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
  | WHERE KEY = 'page_location'), r'([?]*)') AS page_location_stripped,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
  | WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  | WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  | ELSE traffic_source.source
```

```

END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "view_item"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 3" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
WHERE KEY = 'page_location'), r'([?]*)') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "add_to_cart"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 4" as step_number,
user_pseudo_id,
user_id,

```



```

(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
| WHERE KEY = 'page_location'), r'([?]*)') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
| WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
| WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
| ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
| `project_XYZ.analytics_XYZ.events_*`
WHERE
| platform = "web"
| AND event_name = "view_cart"
| AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 5" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
| WHERE KEY = 'page_location'), r'([?]*)') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
| WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
| WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
| ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
| `project_XYZ.analytics_XYZ.events_*`
WHERE
| platform = "web"
| AND event_name = "begin_checkout"
| AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

```

```

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 6" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
| WHERE KEY = 'page_location'), r'([^\?]*')) AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
| WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
| WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
| ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "add_shipping_info"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 7" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
| WHERE KEY = 'page_location'), r'([^\?]*')) AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
| WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
| WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
| ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,

```

```

ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "add_payment_info"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 8" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_title") AS page_title,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "page_location") AS page_location,
(SELECT value.string_value FROM UNNEST(event_params) WHERE KEY = "page_location") AS page_referrer,
REGEXP_EXTRACT((SELECT value.string_value FROM UNNEST(event_params)
| WHERE KEY = 'page_location'), r'([?]*)*') AS page_location_stripped,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
| WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
| WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
| ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform = "web"
AND event_name = "purchase"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

```

Príloha XII: Snímky obrazovky SQL dotazu pre app_items_funnel

```

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 1" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,

```



```

CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
    `project_XYZ.analytics_XYZ.events_*`
WHERE
    platform != "web"
    AND event_name = "view_item_list"
    AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL
SELECT
    CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
    event_name,
    "step 2" as step_number,
    user_pseudo_id,
    user_id,
    (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
    (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
    platform,
    device.category AS device,
    device.operating_system AS os,
    geo.country AS country,
    geo.city AS city,
    CASE
        WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
        WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
        ELSE traffic_source.source
    END AS traffic_source,
    traffic_source.medium AS traffic_medium,
    traffic_source.name AS traffic_campaign,
    ecommerce.transaction_id AS transaction_id,
    ecommerce.purchase_revenue AS purchase_revenue,
    ecommerce.total_item_quantity AS total_item_quantity,
    ecommerce.shipping_value AS shipping_value,
    ecommerce.unique_items AS unique_items
FROM
    `project_XYZ.analytics_XYZ.events_*`
WHERE
    platform != "web"
    AND event_name = "view_item"
    AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL
SELECT
    CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
    event_name,
    "step 3" as step_number,
    user_pseudo_id,
    user_id,
    (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
    (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
    (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
    platform,
    device.category AS device,
    device.operating_system AS os,
    geo.country AS country,
    geo.city AS city,

```

```

CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
  `project_XYZ.analytics_XYZ.events_*`
WHERE
  platform != "web"
  AND event_name = "add_to_cart"
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
  CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  event_name,
  "step 4" as step_number,
  user_pseudo_id,
  user_id,
  (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY = "isLoggedIn") AS user_logged,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,
  CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
  END AS traffic_source,
  traffic_source.medium AS traffic_medium,
  traffic_source.name AS traffic_campaign,
  ecommerce.transaction_id AS transaction_id,
  ecommerce.purchase_revenue AS purchase_revenue,
  ecommerce.total_item_quantity AS total_item_quantity,
  ecommerce.shipping_value AS shipping_value,
  ecommerce.unique_items AS unique_items
FROM
  `project_XYZ.analytics_XYZ.events_*`
WHERE
  platform != "web"
  AND event_name = "view_cart"
  AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
  CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
  event_name,
  "step 5" as step_number,
  user_pseudo_id,
  user_id,
  (SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY = "isLoggedIn") AS user_logged,
  (SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
  (SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
  platform,
  device.category AS device,
  device.operating_system AS os,
  geo.country AS country,
  geo.city AS city,

```



```

CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items

FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "begin_checkout"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 6" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
platform,

device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
  WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
  WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
  ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items

FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "add_shipping_info"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

UNION ALL

SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 7" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,

```

```

platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "add_payment_info"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")
UNION ALL
SELECT
CAST(TIMESTAMP_MICROS(event_timestamp) AS DATE) AS event_date,
event_name,
"step 8" as step_number,
user_pseudo_id,
user_id,
(SELECT value.string_value FROM UNNEST (user_properties) WHERE KEY ="isLoggedIn") AS user_logged,
(SELECT value.int_value FROM UNNEST(event_params) WHERE KEY = "ga_session_id") as session_id,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen") AS screen_name,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_screen_class") AS screen_class,
(SELECT value.string_value FROM UNNEST(event_params) WHERE key = "firebase_previous_screen") AS previous_screen,
platform,
device.category AS device,
device.operating_system AS os,
geo.country AS country,
geo.city AS city,
CASE
    WHEN traffic_source.source LIKE "%facebook%" THEN "facebook"
    WHEN traffic_source.source LIKE "%instagram%" THEN "instagram"
    ELSE traffic_source.source
END AS traffic_source,
traffic_source.medium AS traffic_medium,
traffic_source.name AS traffic_campaign,
ecommerce.transaction_id AS transaction_id,
ecommerce.purchase_revenue AS purchase_revenue,
ecommerce.total_item_quantity AS total_item_quantity,
ecommerce.shipping_value AS shipping_value,
ecommerce.unique_items AS unique_items
FROM
`project_XYZ.analytics_XYZ.events_*`
WHERE
platform != "web"
AND event_name = "purchase"
AND (_TABLE_SUFFIX >= FORMAT_DATE("%Y%m%d", '2024-01-01') OR _TABLE_SUFFIX LIKE "%intraday%")

```