Migráció felhőbe Snowflake alapokon

Your specialist **data engineering** partner

www.infinitelambda.com

Zoltan Csonka

data architect

Passionate and experienced data architect

I have many years experience with modelling, designing, building, migrating data warehouse and BI solutions. I have worked for multinational enterprises and small companies too in the field of insurance, banking, e-commerce, marketing, education.

Certified Snowflake professional

I am currently at **Infinite Lambda,** we are building could based data warehouse solutions to help our clients to make transform to data driven organisation.

INFINITE LAMBDA

aws · snowflake · looker · Google Cloud Platform · tableau · Apache Airflow · SNOWPLOW · Fivetran · databricks · Microsoft Azure · HashiCorp · dbt_

We're hiring!

A feladat hátteréről

Az ügyfélről

Sikeres mobil alkalmazás fejlesztő angol cég.

A jelenlegi projekt egy gyerekeknek szóló, hang alapú, mobilos alkalmazásuk adatait dolgozza fel.

Az alkalmazás célja, hogy segítsen a gyerekeknek feszültséget oldani, segítse az elalvást.

- Világsztárok mondanak történeteket, meséket
- Altató, meditációs zenék, nyugtató hangok

A projekt célja

Segíteni az ügyfelet, hogy hatékonyabb, adatvezérelt vállalattá váljon.
Cél, hogy automatizált, modern adat platformot, riporting és elemző rendszert építsünk.

A mobil alkalmazással kapcsolatos

- használati,
- marketing,
- értékesítés segítő

riportokat és elemzéseket támogassuk.

A kihívás, problémák

Ami megvolt

- excel alapú kalkulációk, riportok
- külső alkalmazások adatait eltérő formában és úton importálták excelbe
- kézi folyamat

Ami nem jó

- ott kell lenni a fejlesztőnek
- nehezen átlátható, módosítható
- kézi feldolgozás miatt sok a hiba

Ami nem volt

- módszertani alapjuk
- excelen kívül más technológia
- nem volt történetiség
- nyers adat

INFINITE LAMBDA

Követelmények - 1. fázis

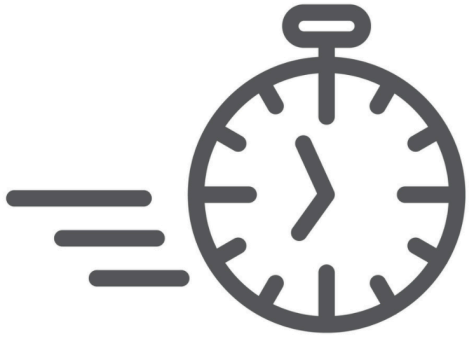
- adjon end-to-end **keretet a továbbfejlesztéshez**
- napi jelentés **automatizálás**a,
- a 31 KPI emigrálása (főként előfizetés és kampány adatok),
- BI eszközben elérhető napi jelentés **riport**

- **Könnyen módosítható,** gazdaságos legyen
- Oktatás, hogy a fejlesztést később belső kollégákkal folytathassák

Tulajdonosi elvárás a közel valós idejű riportok kiszolgálására alkalmas rendszer!

INFINITE LAMBDA

A megvalósítás határideje



4 hét

(+1 felmérés, előkészítés)

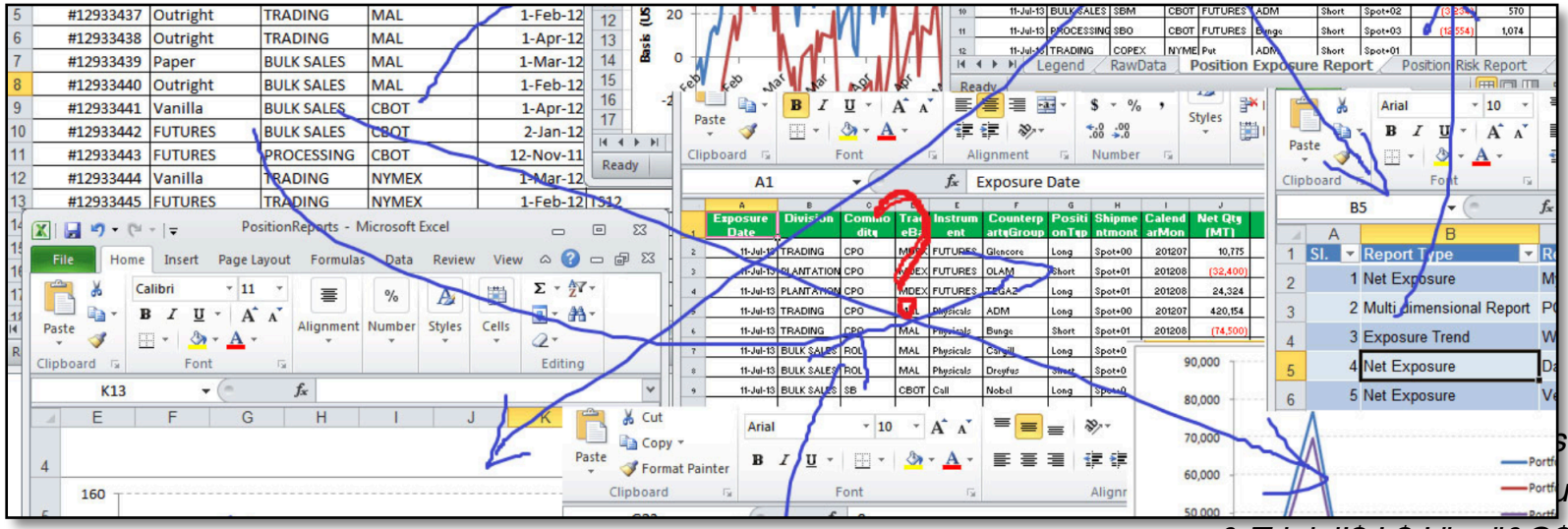
INFINITE LAMBDA

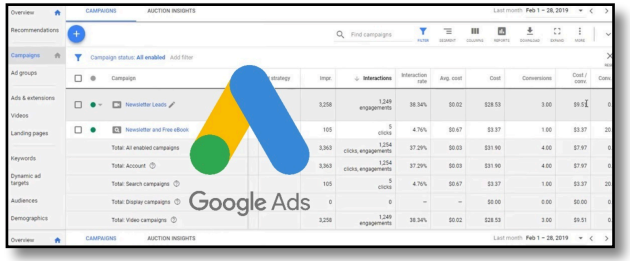
A projekt módszertana

Agilis, scrum, ticketing, minden hét végére leszállítandó és demo

Hol kezdjük?

A meglévő folyamatok és excel kalkulációk feltérképezése





s (Subs)'!$M:$M,'App Figures (Subs)'!$K:
ust Installs & Trials'!$S:$S,'Adjust Installs
& Trials'!$J:$J,"<="&G$8,'Adjust Installs & Trials'!$J:
$J,">"&G$8-7)+SUMIFS('Google & BT Acqs'!$M:$M,'Google &
BT Acqs'!$J:$J,G$8)

INFINITE LAMBDA

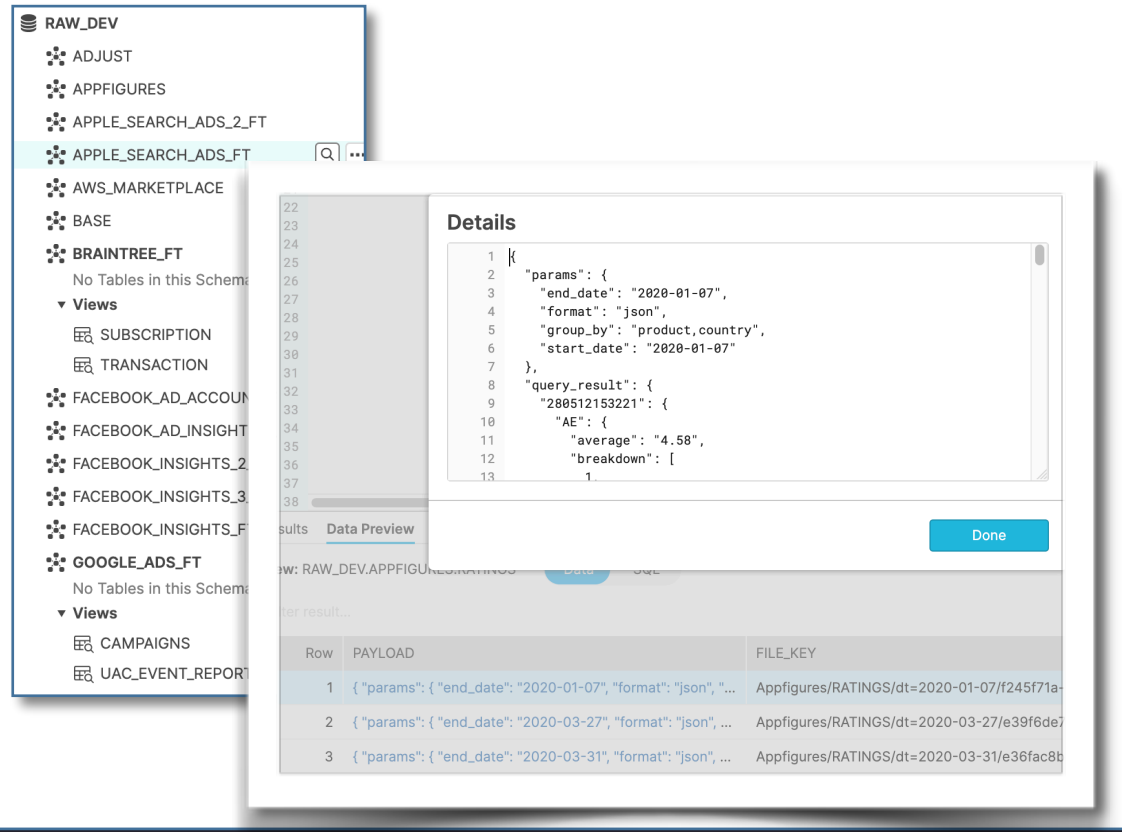
Az architektúra és a folyamat

Architecture - Data Lake

DATA LAKE

- Historikusan tárolt nyers adatok
- Tisztított, kiterített adatok
- Üzleti logika nélkül

NEM VESZÍTHETÜNK ADATOT,
MINDEN ADATOT LÁTNI AKARUNK!

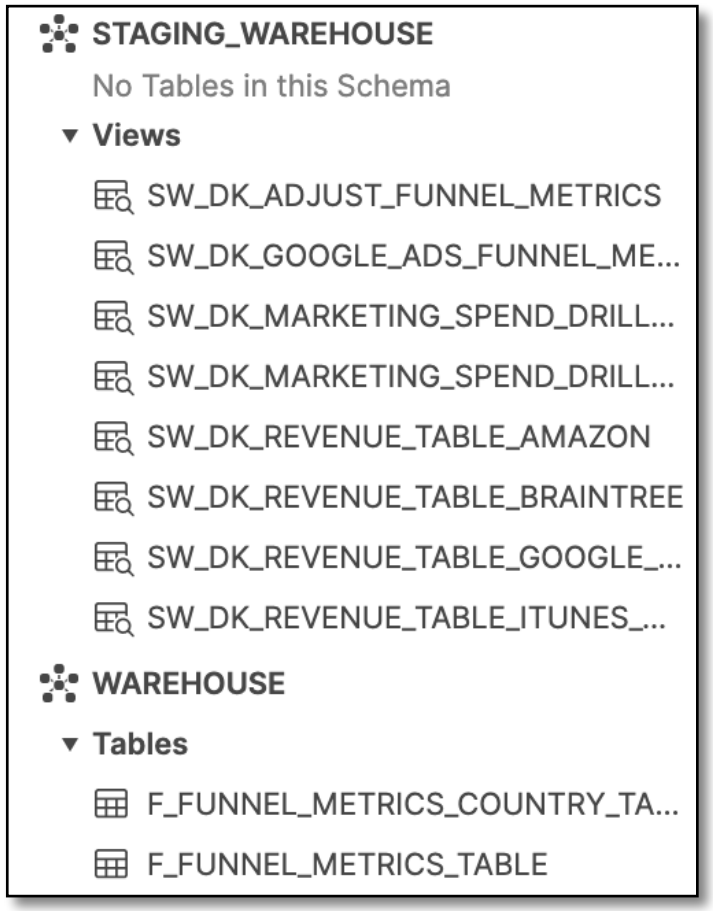


INFINITE LAMBDA

Architecture - Warehouse

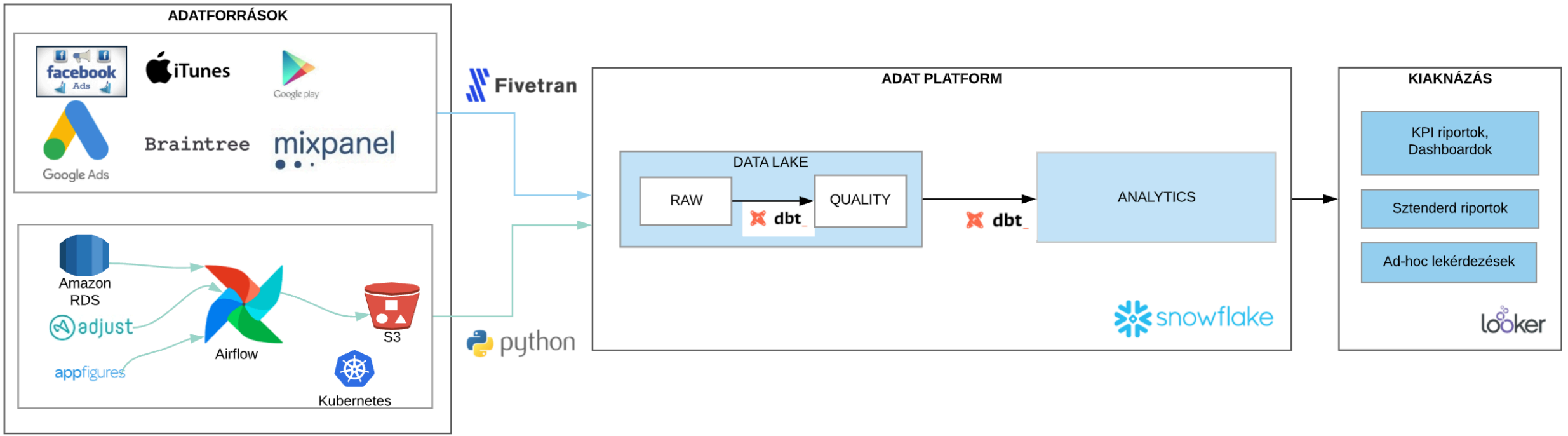
Data Warehouse

- Historikusan feldogozott adatok
- Üzleti KPI-ok
- Dimenzionális modell BI-hoz



INFINITE LAMBDA

Egyszerűsített architektúra



Fivetran python

Adatkinyerés, betöltés

dbt_

Transzformáció

snowflake

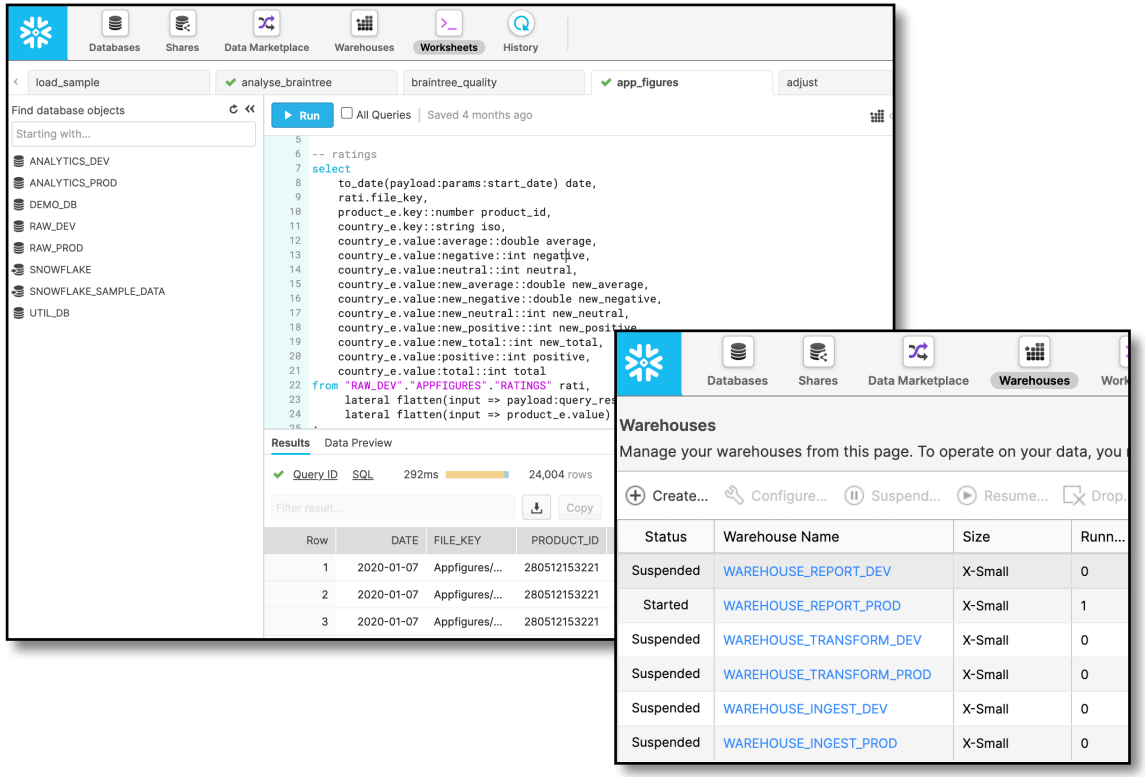
Adatbázis, platform

Apache Airflow

Folyamatvezérlés

Az eszközök

Data warehouse, data lake platform



Felhőre fejlesztett relációs adatbázis

SQL alapú adatbázismotor

Egy platformon data lake, data warehouse

Számítások és adattárolás szétválasztott

Fantasztikus dokumentáció, rengeteg esettanulmány, pezsgő közösség

INFINITE LAMBDA

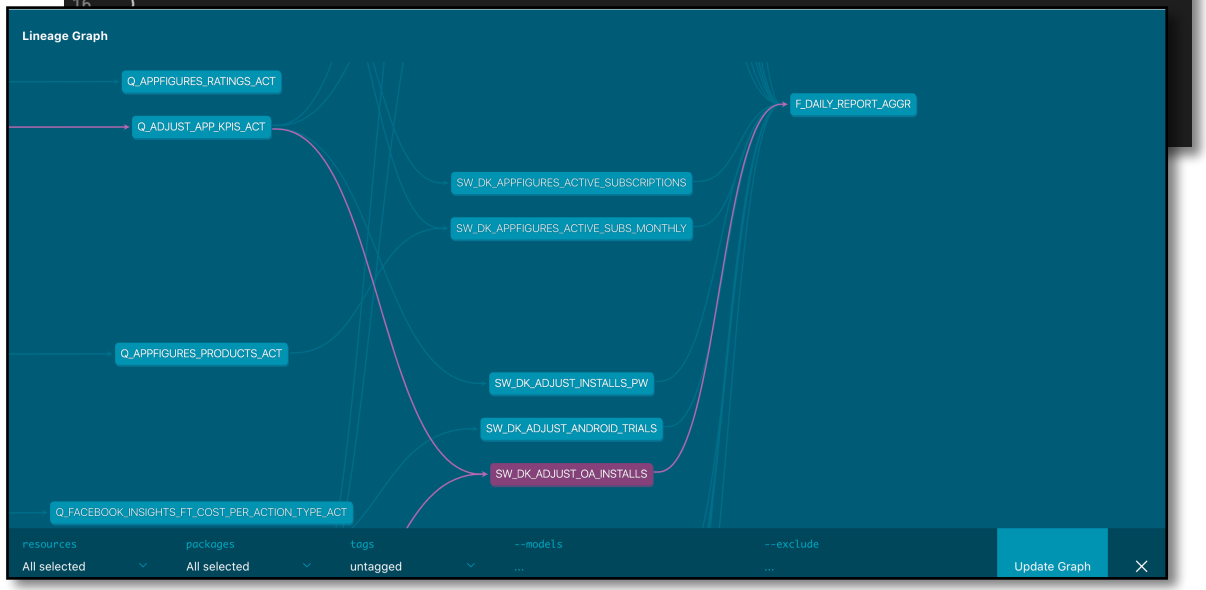
Transformation - ELT

dbt_ az adat transzformációk eszköze

- CTAS (SQL)
- Lineage (függőség kezelés)
- Dokumentációs segítség
- Open source

```sql
{{      You, 4 months ago • add KPI views
  config(
    materialized ='view',
    schema = generate_schema_name('STAGING_WAREHOUSE')
  )
}}

with source_adjust_events as (
    select
        events_kpis.date,
        SUM(case when network_groups.GROUP_TYPE = 'Group1' then events_kpis.SUBSCRIPTION_PAID_INIT_YHVRJP_E
        SUM(case when network_groups.GROUP_TYPE = 'Group2' then events_kpis.SUBSCRIPTION_PAID_INIT_YHVRJP_E
    from {{ ref('Q_ADJUST_EVENTS_KPIS_ACT') }} events_kpis
    left join {{ ref('Q_ADJUST_NETWORK_GROUPS') }} network_groups on events_kpis.NETWORK = network_groups.N
    group by 1
)
```

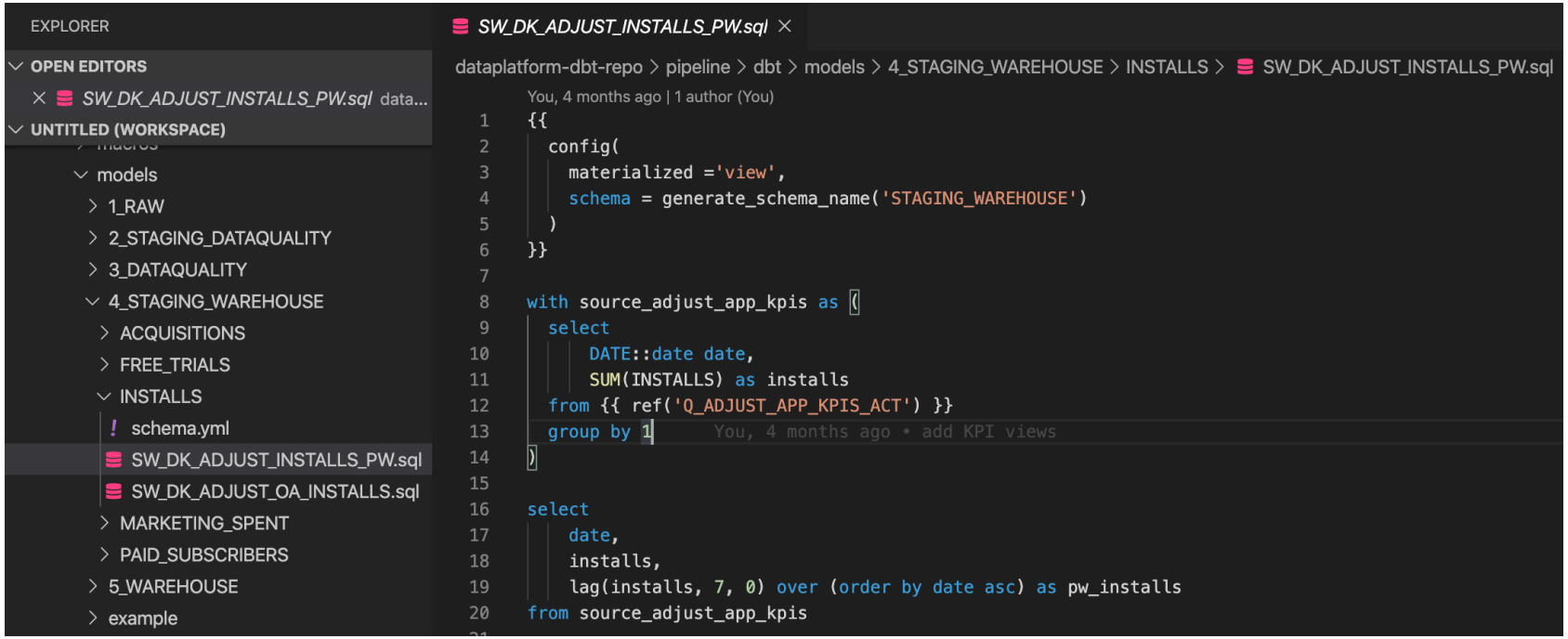


INFINITE LAMBDA

Transzformációk

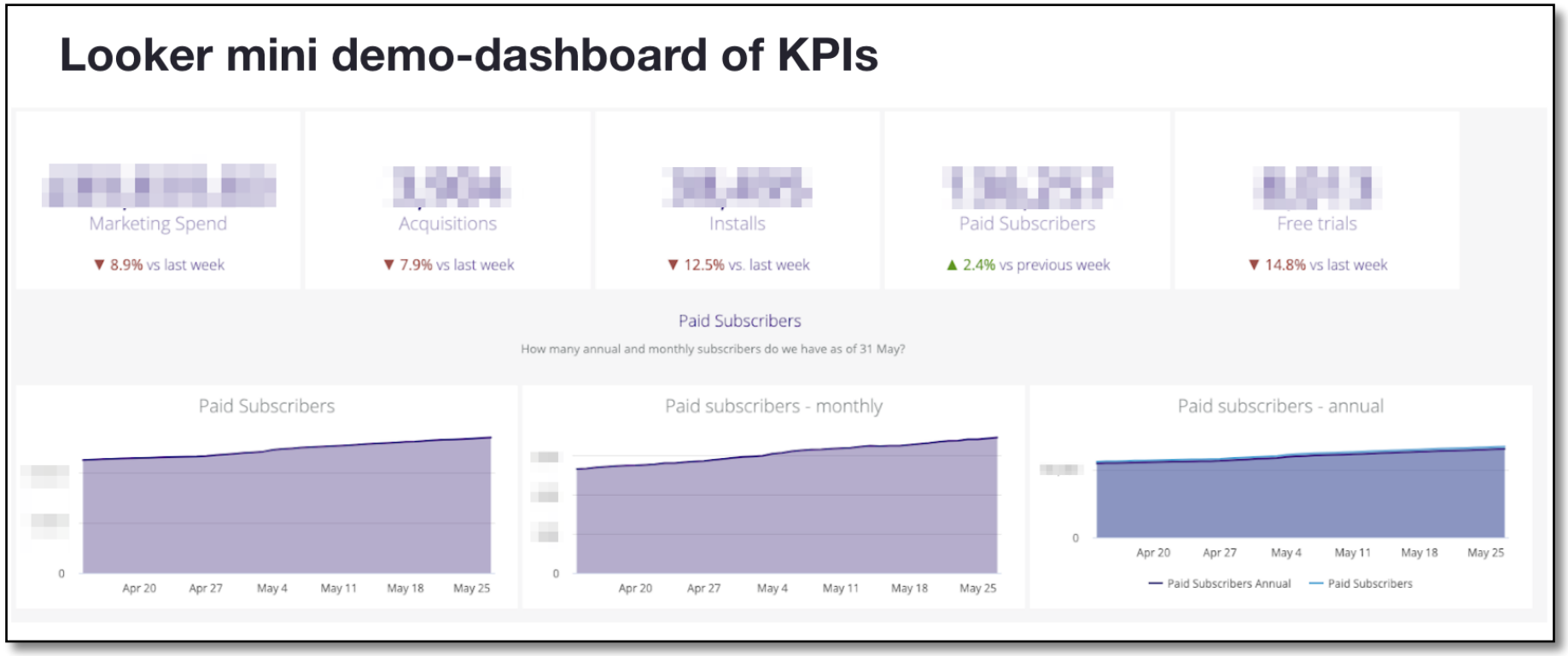
DBT struktúra

- Funkcionális rétegek
- KPI definíciók



INFINITE LAMBDA

KPI riportok



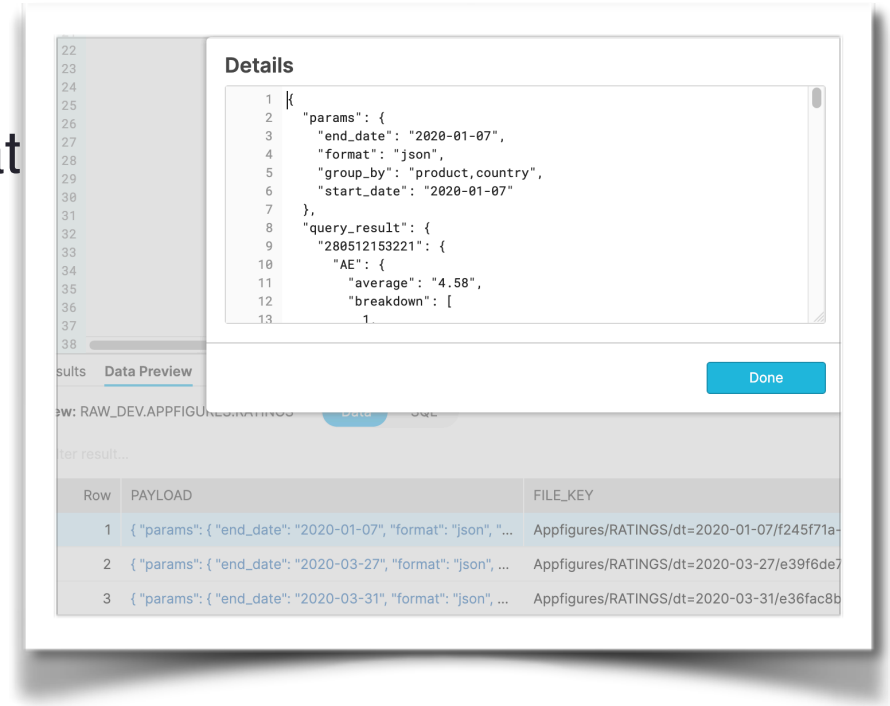
Néhány kiváló Snowflake funkció, amit a projektben használtunk

INFINITE LAMBDA

Snowflake feature - semi-structured adat kezelése

VARIANT általános adattípus

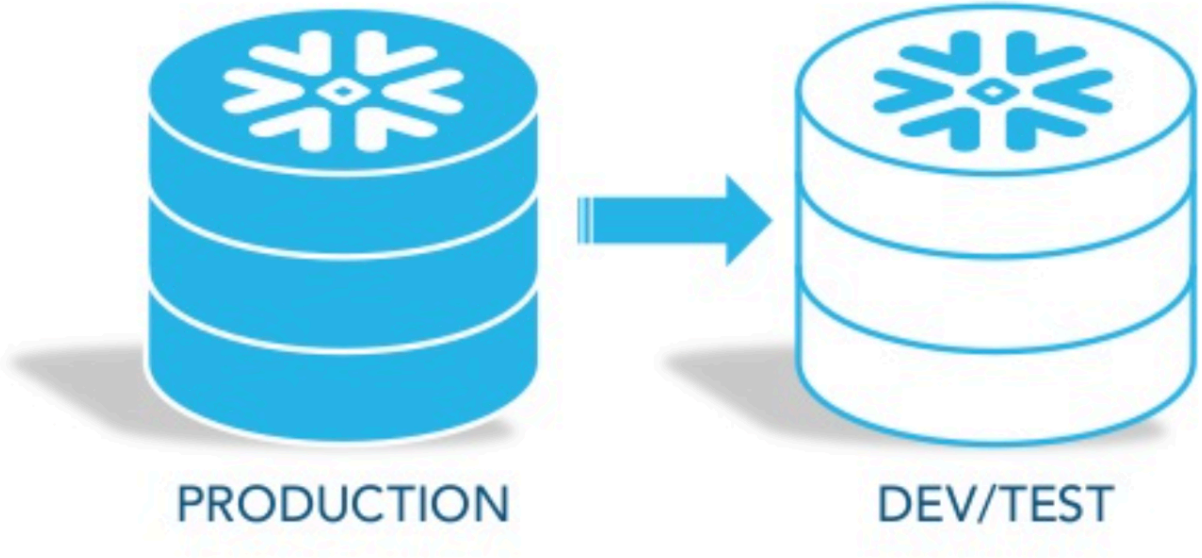
- Variant type 16 MB bármilyen adattípust tartalmazhat
  (pl. JSON, AVRO, ORC, Parquet, XML forrásokból)
- DATA LAKE építéshez
  Séma változások kezelése végtelenül egyszerű
- Data science feladatokhoz
- SQL-ben egyszerűen lekérdezhető



INFINITE LAMBDA

Snowflake feature - Zero copy clone

Teszt környezet létrehozása ingyen és azonnal



- táblákat, sémákat vagy egész adatbázist klónózik azonnal
- Nincs tárterület igény
- Az éles környezetet nem terheli
- A jogosultság örökölt
- Bármennyi készíthető

INFINITE LAMBDA

Snowflake feature - tuning és optimalizálás

- Többnyire nincs rá szükség
  nem kell és nem is lehet indexelni vagy hintelni, (és ezért én nagyon hálás vagyok)
  partícionálni is csak ritkán szükséges (1TB táblaméret fölött)

- Separate WH, cache, auto scale out, scale up

Lehet rossz SQL-t írni, hibás folyamatokat lefejleszteni itt is :)

INFINITE LAMBDA

Snowflake feature - Időutazás

Törölt és módosított adatok lekérdezése, visszaállítása

- táblákat, sémákat vagy egész adatbázist visszaállíthatunk
- 1-90 nap között
- van tárolási költsége

```
undrop table subscritptions;
```

```
select * from campaigns
at(timestamp => 'Mon, 01 May 2020 16:20:00 -0700'::timestamp);
```

INFINITE LAMBDA

Összefoglalás

Nehézségek a projekt során

- Google, apple, amazon (is) eltérő definíciókat használ, eltérően számolja a mutatókat. pl. a letöltés darabszámot
- Fivetran replikáció nem tudja milyen adatot kap, a sorrendiség kezelése nem egyértelmű *pl. apple tranzakciók nem megfelelő sorrendben érkeznek*
- Némelyik forrásrendszer csak aggregált adatokat ad
- Looker riportokat nehéz volt Excellé varázsolni (a Looker sem Excel)

INFINITE LAMBDA

Leghasznosabb előnyök a projektből az ügyfél számára

- Teljesen automatikus folyamatok
- End-to-end keret, rugalmas módszertan és eszközök a továbbfejlesztéshez
- Nem kell adminisztrálni az eszközöket!
- Népszerű és könnyen használható eszközök. Snowflake, DBT, Airflow
- Könnyű a közös munkavégzés, megoszthatóak a folyamatok, adatok.

Nincs varázslat, működik!

INFINITE LAMBDA

Felhő migrációs gondolatok

- Felhő vagy sem, az a fontos, hogy ne az eszközzel kelljen küzdeni
- Nem kérdés, hogy Felhőben dolgozunk
- Hogyan kerül az adat a Felhőbe? *(ahol nincs delta az fáj itt is)*
- Felhő migrációnál a legnagyobb félelem még mindig az adatbiztonság
- Felhőben használjunk felhőre tervezett eszközöket

Még mindig jó megérteni mi kell az üzletnek, mit csinálunk.

Olyan rendszert építsünk,
ami rugalmas és
könnyen alkalmazkodik a változáshoz,
mert holnap már más lesz az igény!

Köszönöm a figyelmet!

Van kérdésed?

Contact us

info@infinitelambda.com

www.infinitelambda.com

Agile, Cloud Based Data Solutions

We're hiring!