



INFINITE
LAMBDA

Adattranszformáció DBT-vel a gyakorlatban

Infinite Lambda

www.infinitelambda.com

Bemutakozás

Introducing Infinite Lambda

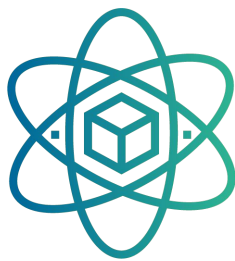
- ▶ 13 different industries
- ▶ from startups to Enterprise
- ▶ small teams to big data departments
- ▶ over 80 consultants and engineers
- ▶ first MVP delivery around 4 - 6 weeks



Introducing Infinite Lambda



**Fullstack
Data
Development**



**Business
Intelligence &
Data Science**



**Cloud
Application
Development**



**Flexible
DataOps
Services**

Introduction





Google Cloud



Apache
Airflow



+ a b | e a u®



PREFECT



Azure

looker



dbt



databricks



Fivetran



snowflake®



Terraform



docker



kubernetes

Introduction



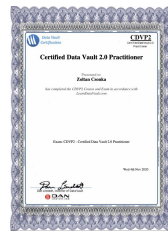
INFINITE
LAMBDA

Zoltan Csonka

Zoltan is Data Architect at Infinite Lambda, specialized in Data Warehousing and Business Intelligence. He has 18 years of experience working on data warehouse design and implementation with on-prem and cloud-based warehouses.

He is focusing on working on modern data stack with data modelling, Data Vault, Snowflake and DBT.

Zoltan is Certified Data Vault 2.0 Practitioner and Snowflake Snowpro certified.



A DBT filozófiája

Data Build Tool



ELT - adat transzformációs eszköz

DBT filozófia



Mindenki számára elérhető
adattranzformációs eszköz



Adatok tesztelése, ellenőrzését
minden lépésnél el kell végezni



Az SQL a közös nyelv
az adatok világában



Adattranzformációt, lineaget
és dokumentációt együtt kell kezelni



Adatos környezetben az együttműködésnek úgy kell működnie mint egy
szoftverfejlesztésnek (egy kódbázis, git, pull, etc.)

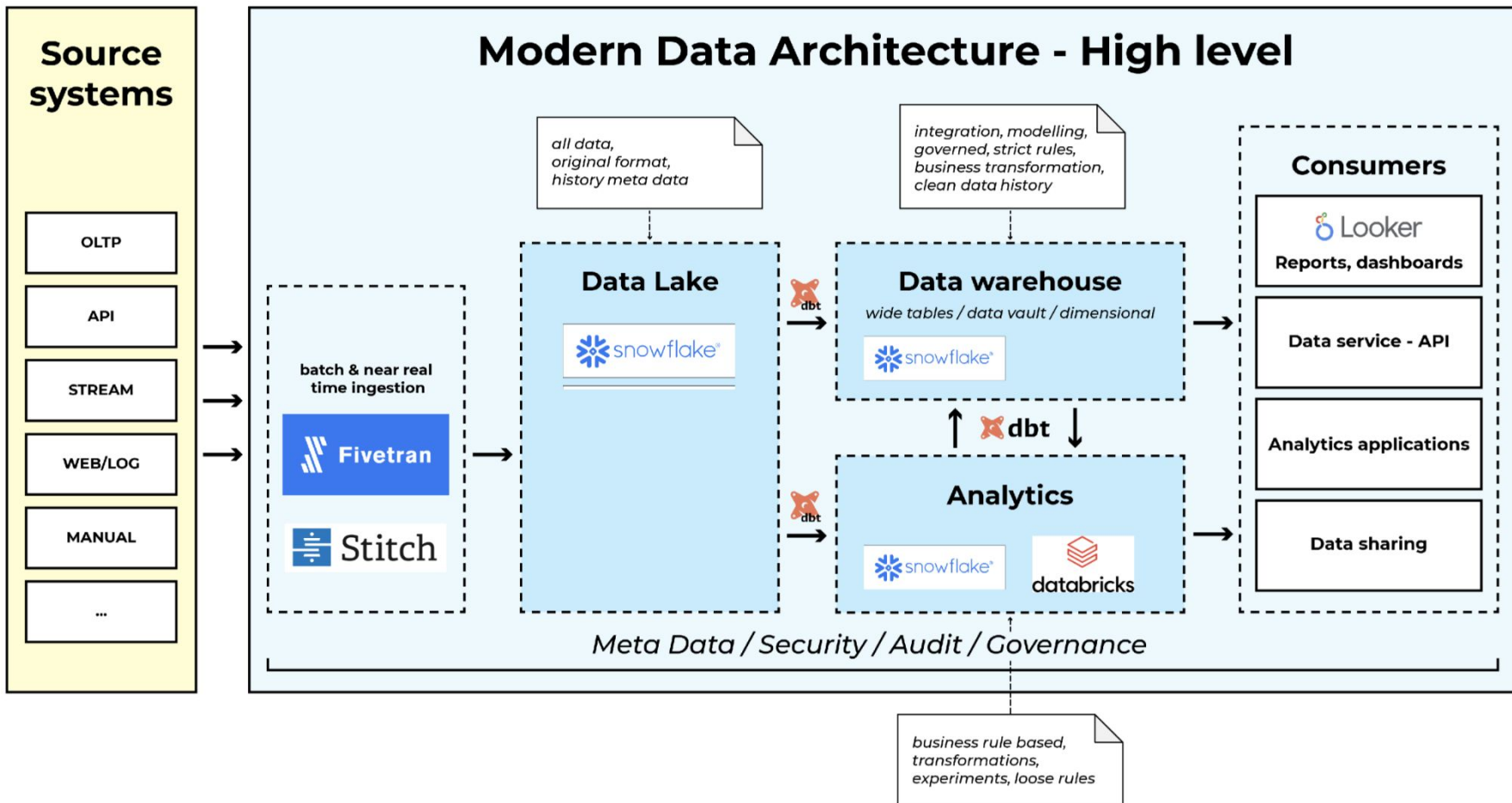
Kinek az eszköze?



Analytics engineer / Data Engineer / BI engineer / Data analyst

Az dolgozik DBT-vel, aki érti is az adatot és az üzletet, van technológiai és modellezési ismerete.

Architektúra, komponensek és szerep az adatkörnyezetben



Miből főzünk?

Verziókezelő



Adatbázis



DBT vagy - és DBT Cloud



Modell

Egy select utasítás
text file, sql kiterjesztésel,
model neve a tábal neve

```
1  with source as (  
2      select * from {{ ref('raw_customers') }}  
3  )  
4  ),  
5  
6  renamed as (  
7  
8      select  
9          id as customer_id,  
10         first_name,  
11         last_name  
12  
13         from source  
14  
15     )  
16  
17  select * from renamed
```

Materializáció

A cél adathalmaz töltési stratégiáját vezérli

Nézet / tábla,
Teljes / inkrementális töltés...

```
1  {{
2    config(
3      materialized='view'
4    )
5  }}
6
7  with source as (
8    select * from {{ ref('raw_customers') }}
9
10 ),
11
12 renamed as (
13
14   select
15     id as customer_id,
16     first_name,
17     last_name
18
19   from source
20
21 )
```

Függőség

Egy futtatási sorrend meghatározására

```
1  {{
2      config(
3          materialized='view'
4      )
5  }}
6
7  with source as (
8      select * from {{ ref('raw_customers') }}
9  )
10 ),
11
12 renamed as (
13
14     select
15         id as customer_id,
16         first_name,
17         last_name
18
19     from source
20
21 )
```

Mi történik futtatáskor

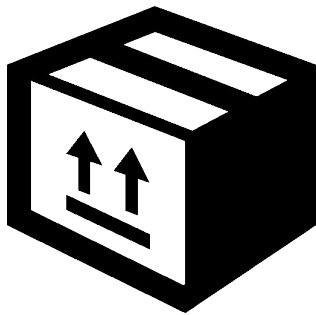
Compiler (dbt compile)



Runner (dbt run)



Csomagkezelő

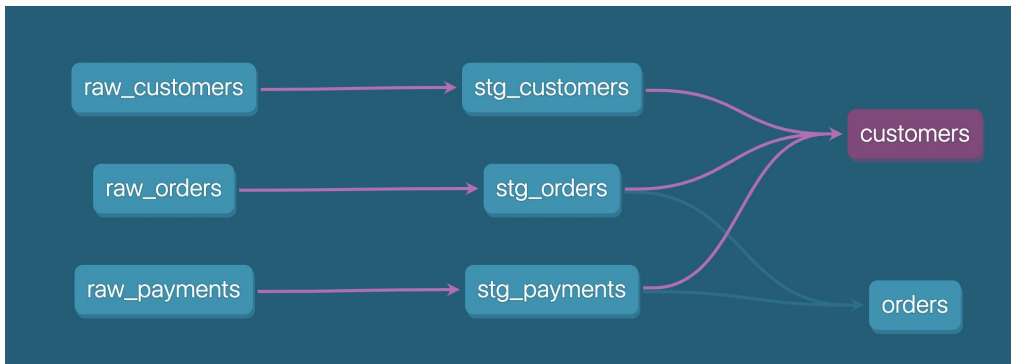


GIT url + DBT DEPS

Sok sok hasznos csomag élvezzük az open source eszköz elonyeit.

Gyakorlati megoldások

Lineage & dokumentáció



customers view

[Details](#) [Description](#) [Columns](#) [Referenced By](#) [Depends On](#) [SQL](#)

Details

TAGS	OWNER	TYPE	PACKAGE	RELATION
untagged	DATA_ENGINEERING	table	jaffle_shop	DBT_DEMO.dbt_rossonka.customers

APPROXIMATE SIZE	LAST MODIFIED	ROW COUNT
4 KB	2021-06-23 21:48UTC	100

Description

This table has basic information about a customer, as well as some derived facts based on a customer's orders

Columns

COLUMN	TYPE	DESCRIPTION	TESTS	MORE
customer_id	NUMBER	This is a unique identifier for a customer	U N F	>
first_name	TEXT	Customer's first name. PII		>
last_name	TEXT	Customer's last name. PII		>

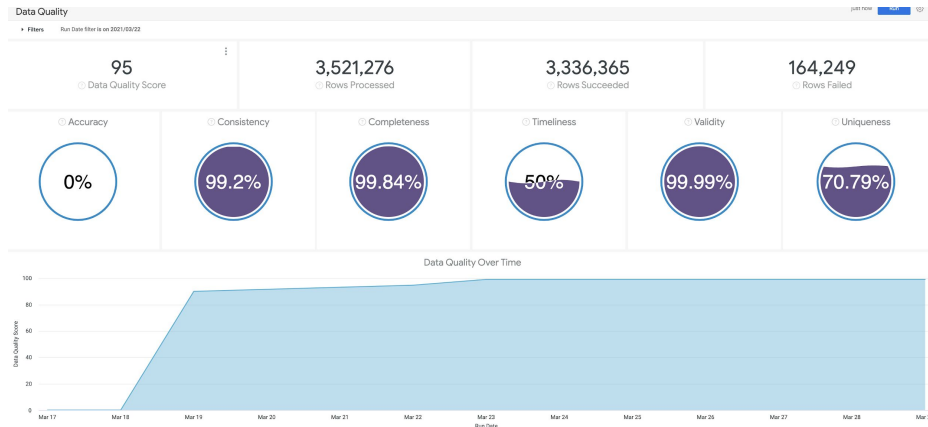
Tesztelés

Adatbeszt implementáció

schema.yml

```
columns:
  - name: order_id
    tests:
      - unique
      - not_null
    description: This is a unique identifier for an order
  - name: customer_id
    description: Foreign key to the customers table
    tests:
      - not_null
      - relationships:
          to: ref('customers')
          field: customer_id
```

Vizualizáció Lookerben



Makró

A Jijnja - SQL makrók segítségével rugalmas eljárásokat, templateket készíthetünk

union.sql

```
1 {% macro union_relations(relations, column_override=None, include=[], exclude=[], source_column_name='_dbt_source_relation') -%}  
2   {{ return(adapter.dispatch('union_relations', packages = dbt_utils.get_utils_namespaces()))(relations, column_override, include, exclude, sou  
3 {% endmacro %}  
4  
5 {% macro default__union_relations(relations, column_override=None, include=[], exclude=[], source_column_name='_dbt_source_relation') -%}  
6  
7   {% if exclude and include -%}  
8     {{ exceptions.raise_compiler_error("Both an exclude and include list were provided to the 'union' macro. Only one is allowed") }}  
9   {% endif -%}  
10  
11   {%#-- Prevent querying of db in parsing mode. This works because this macro does not create any new refs. --#}  
12   {% if not execute %}  
13     {{ return('') }}  
14   {% endif -%}  
15  
16   {% set column_override = column_override if column_override is not none else {} -%}  
17
```

```
{{ dbt_utils.union_relations(  
    relations=[ref('my_model'),  
    source('my_source', 'my_table')],  
    exclude=["_loaded_at"]  
) }}
```

Hol kezdjük el?



dbt

getdbt.com



getdbt.slack.com



infinitelambda.com

Demo

A DBT kiváló adattranszformációs eszköz

Open sources és menedzselt Cloud

SQL alapú

Lineage - dokumentáció beépített

Segíthetünk?

<https://www.linkedin.com/in/zoltan-csonka/>

www.infinitelambda.com

