



Agilis adattárházak startup környezetben

Projekt tapasztalatok, kihívások és megoldások

Zoltan Csonka

Data Architect at Infinite Lambda



Zoltan Csonka

Data Architect
Infinite Lambda

This is me

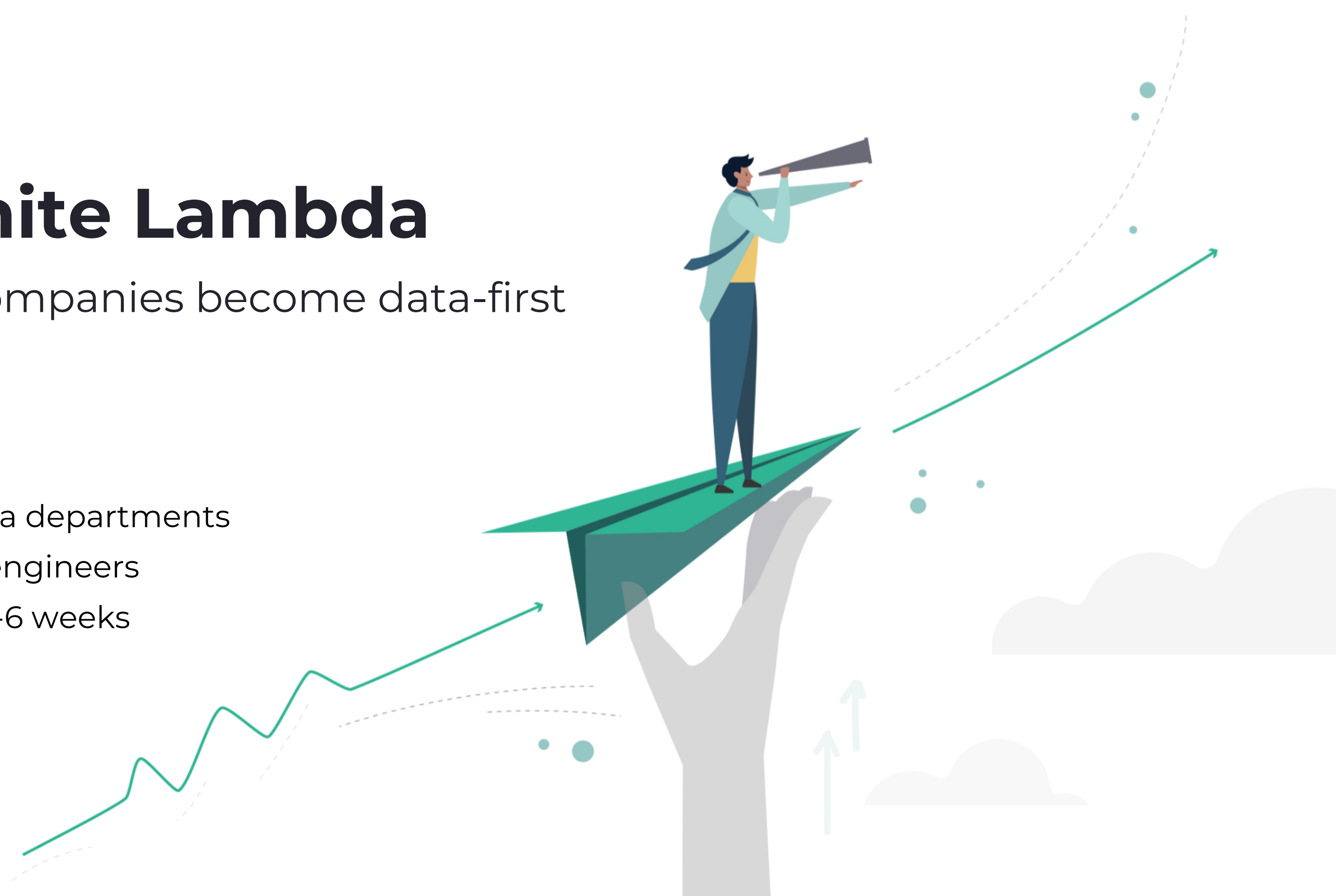
- Specialised in data warehousing and BI
- 18 years of experience in cloud-based and on-prem data warehouse design and implementation
- Focused on the modern data stack, data modelling, Data Vault, Snowflake and dbt
- Certified Data Vault 2.0 Practitioner
- Snowflake SnowPro certified
- Proud father of 3

This is Infinite Lambda

We help leading companies become data-first

- 15+ industries
- Startups to enterprise
- Small teams to big data departments
- 100+ consultants and engineers
- First MVP delivery in 4-6 weeks

www.infinitelambda.com





Full Stack Data Development

All your real-time or batch data into a single integrated platform



BI and Data Science

Automated reports and drill-down dashboards for data-driven decisions



Cloud-Native Applications

Cloud-native development of robust, scalable apps



Flexible DevOps

Building and maintaining core cloud infrastructure





INFINITE
LAMBDA

Ügyfél elvárások

startup környezetben

www.infinitelambda.com

1. Üzleti érték szállítása gyorsan

2. Együttműködés, megosztás és átláthatóság

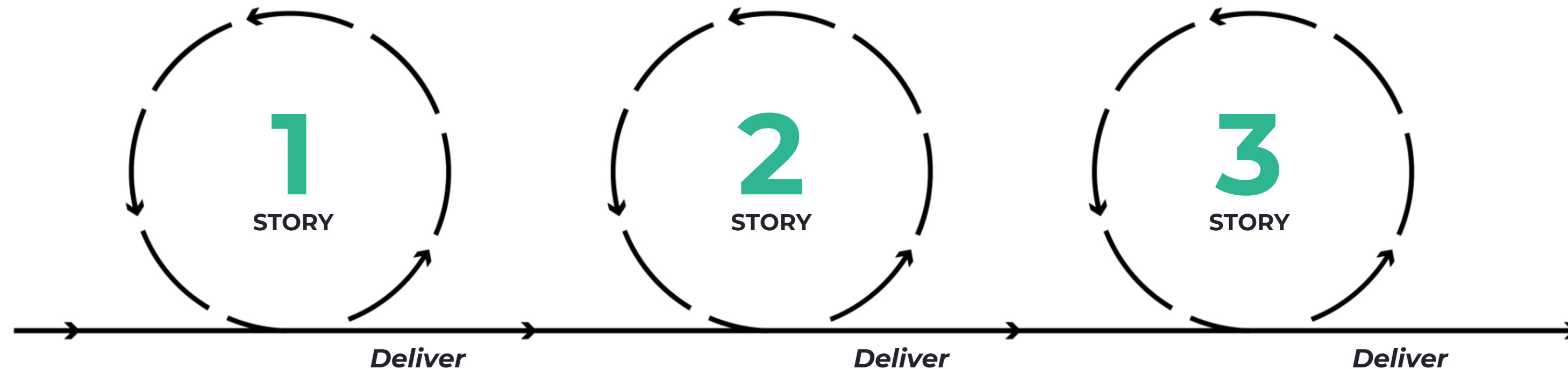
3. Gyors változáskezelés

4. Technológiai függetlenség biztosítása

5. Modern data platform és metodológiák

Agilis modellezés és fejlesztés - data vault

Hozzuk létre vagy bővítsük a modellünket **egy történetre, egy folyamatra** egy időben



Üzleti érték szállítása gyorsan

Automatizáció

- Metaadat- és template-alapú kód generálás
- Teszt automatizáció
- Élesítés
- Reverse engineering és forrás profilozás



Üzleti érték szállítása gyorsan



CI/CD

dbt és gitlab / github integráció

- Automatizáció
- Sebesség
- Együttműködés



Üzleti érték szállítása gyorsan



1. Üzleti érték szállítása gyorsan

2. Együttműködés, megosztás és átláthatóság

3. Gyors változáskezelés

4. Technológiai függetlenség biztosítása

5. Modern data platform és metodológiák

A kommunikáció kulcsfontosságú



Együttműködés, megosztás és átláthatóság

Adatmodellezés & elemzés



Együttműködés, megosztás és átláthatóság

Adatmodellezés & elemzés

data profiling example

File

Edit

View

Insert

Format

Data

Tools

Extensions

Help

Last edit was seconds ago

100%

\$

%

.0

.00

123

Calibri

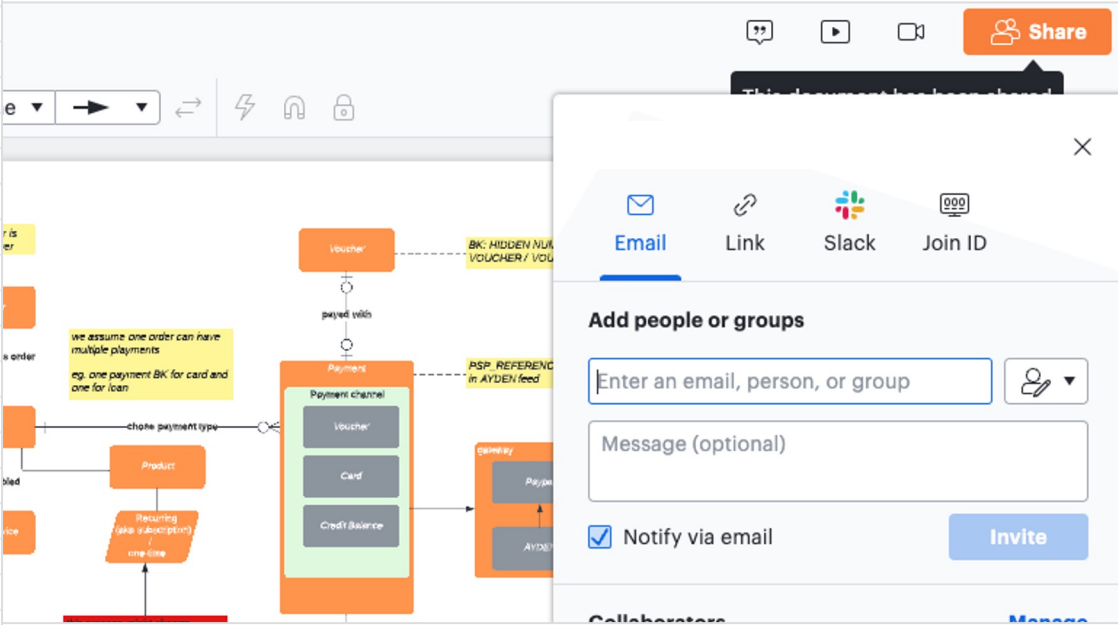
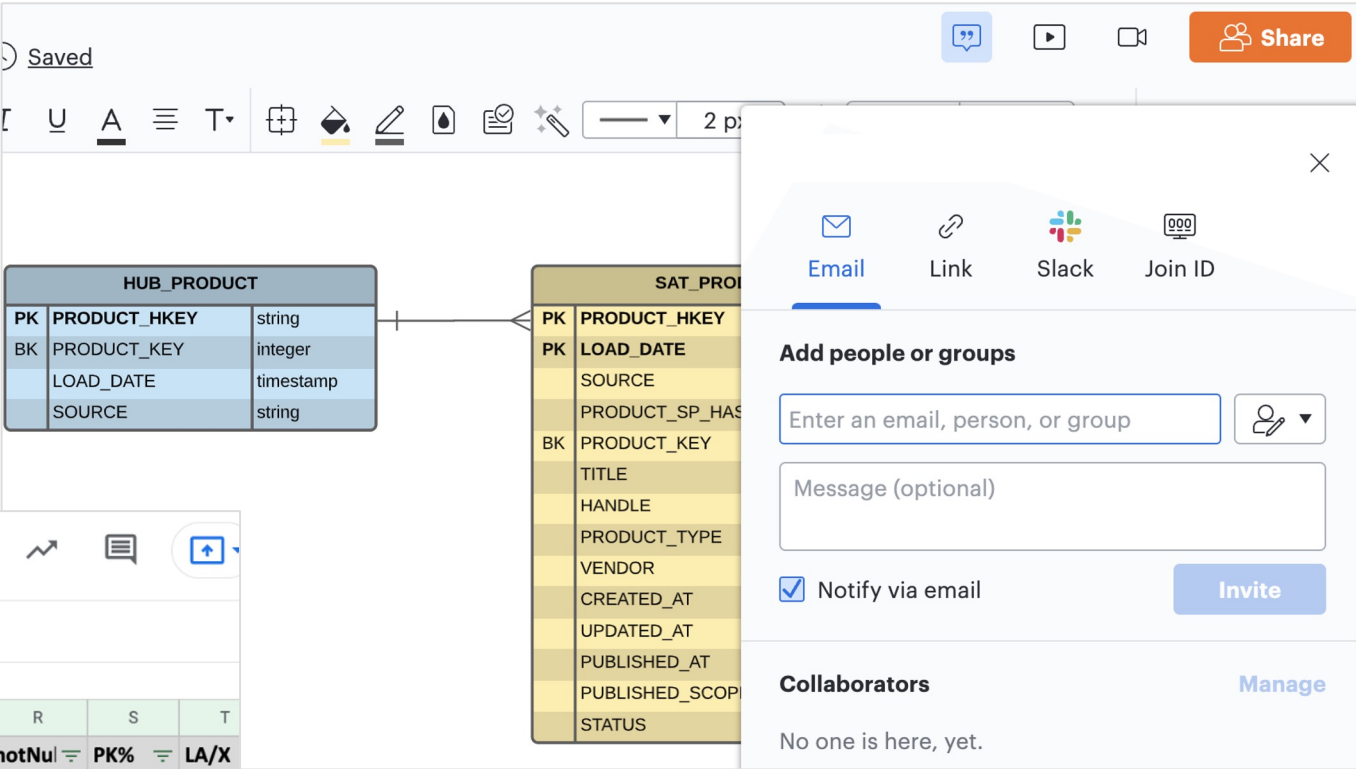
11

B

I

S

A



Együttműködés, megosztás és átláthatóság



Kód, repo és dokumentáció



Conversation 6

Commits 6

Checks 0

Files changed 43

+843 -0

commented 15 days ago

No description provided.

added 4 commits 16 days ago

CDI-152/create guuid map pre-process-raw-vault ... c20bbec

CDI-152/create load_sal macro and first test versions d5cb14a

CDI-152/add more tests to stg_sal 343b879

CDI-152/add data integrity tests 09accf0

reviewed 15 days ago

dbt/docs/macros/pre_process_raw_vault_pii/guuid_map/get__stg__pii__member_id_imei.md

Outdated Show resolved

reviewed 15 days ago

dbt/macros/utils/get_cte_stg_guuid_sal.sql

... @@ -0,0 +1,32 @@

1 + {%-- macro get_cte_stg_guuid_sal(

15 days ago

We should add a specific test to make sure this macro is doing what we want

14 days ago

I have added a unit test on the get_cte_stg_guuid_sal macro

Reply...

Reviewers

Assignees

Labels

Projects

Milestone

Development

dim_customer_entity table

Details Description Columns Referenced By Depends On SQL

Description

This table shows customer entity attributes. A customer entity is a customer legal entity with which one or more Railsbank (or Payrnet) legal entities has signed one or more commercial contracts to deliver some products or services.

Note - work is currently being carried out on customer account hierachies which might change this definition.

The following are conditions to being classed as a customer legal entity

- At least one opportunity with current status of Closed Won
- The Closed Won opportunities must be against the New Business and Upsell record types

Columns

COLUMN	TYPE	DESCRIPTION	TESTS	MORE?
id	TEXT	The primary key f...	+	>
name	TEXT	The name of the ...	+	>
status	TEXT	Salesforce field u...	+	>
customer_success_manager	TEXT	Representative fr...	+	>
offboarding_notification_date	DATE	The date that eith...	+	>

Együttműködés, megosztás és átláthatóság



1. Üzleti érték szállítása gyorsan
2. Együttműködés, megosztás és átláthatóság
- 3. Gyors változáskezelés**
4. Technológiai függetlenség biztosítása
5. Modern data platform és metodológiák

Új feature/érték, a meglévő folyamatok megtörése nélkül

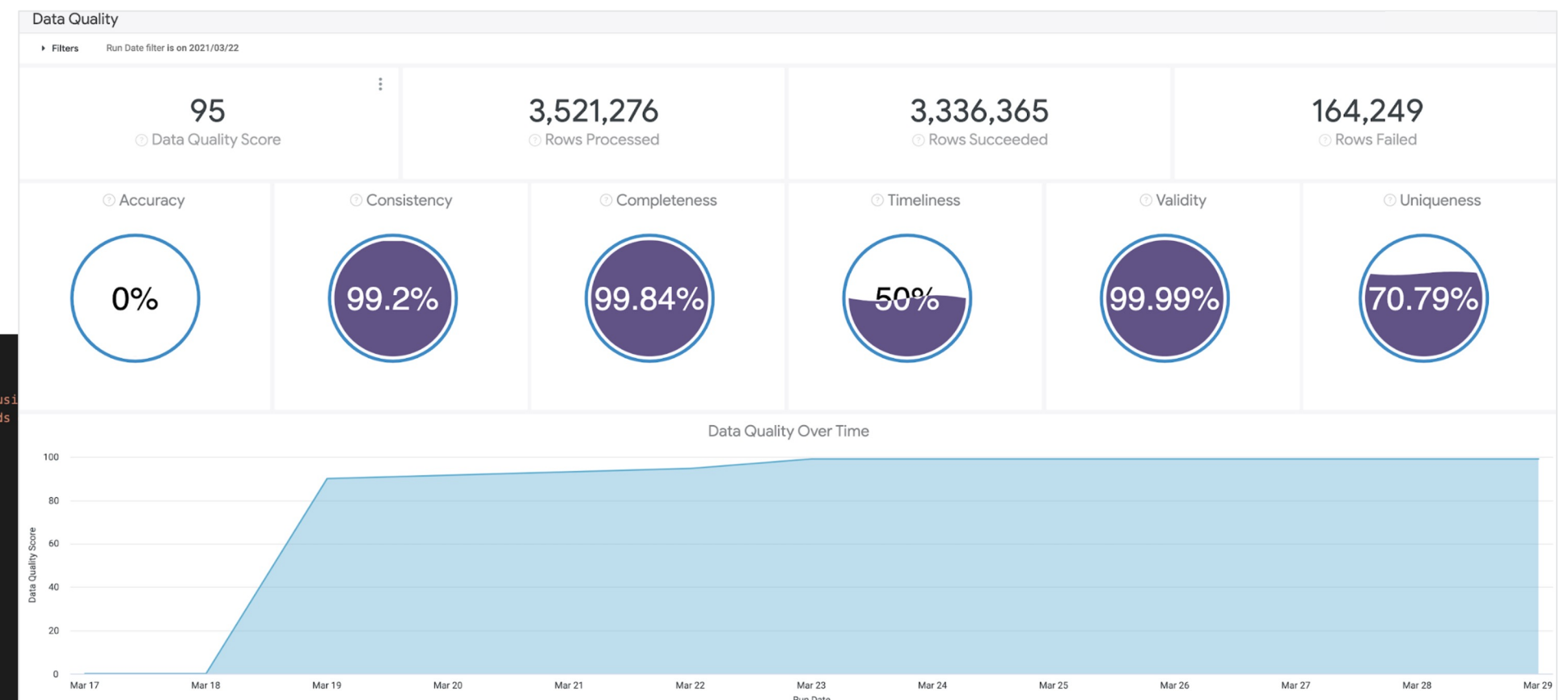
- Új verzió meglévő objektumok változtatása helyett
- Változások követése metaadatok használatával

SAT_CUSTOMER_V2_1		
PK	CUSTOMER_HKEY	binary
PK	LOAD_DATE	timestamp
	SOURCE	varchar
	CUSTOMER_HASHDIFF	binary
	FIRST_NAME	varchar
	LAST_NAME	varchar
	ACCEPTS_MARKETING	boolean
	EMAIL	varchar
	NOTE	varchar
	PHONE	varchar
	STATE	varchar
	VERIFIED_EMAIL	boolean

Adatminőség tesztelése

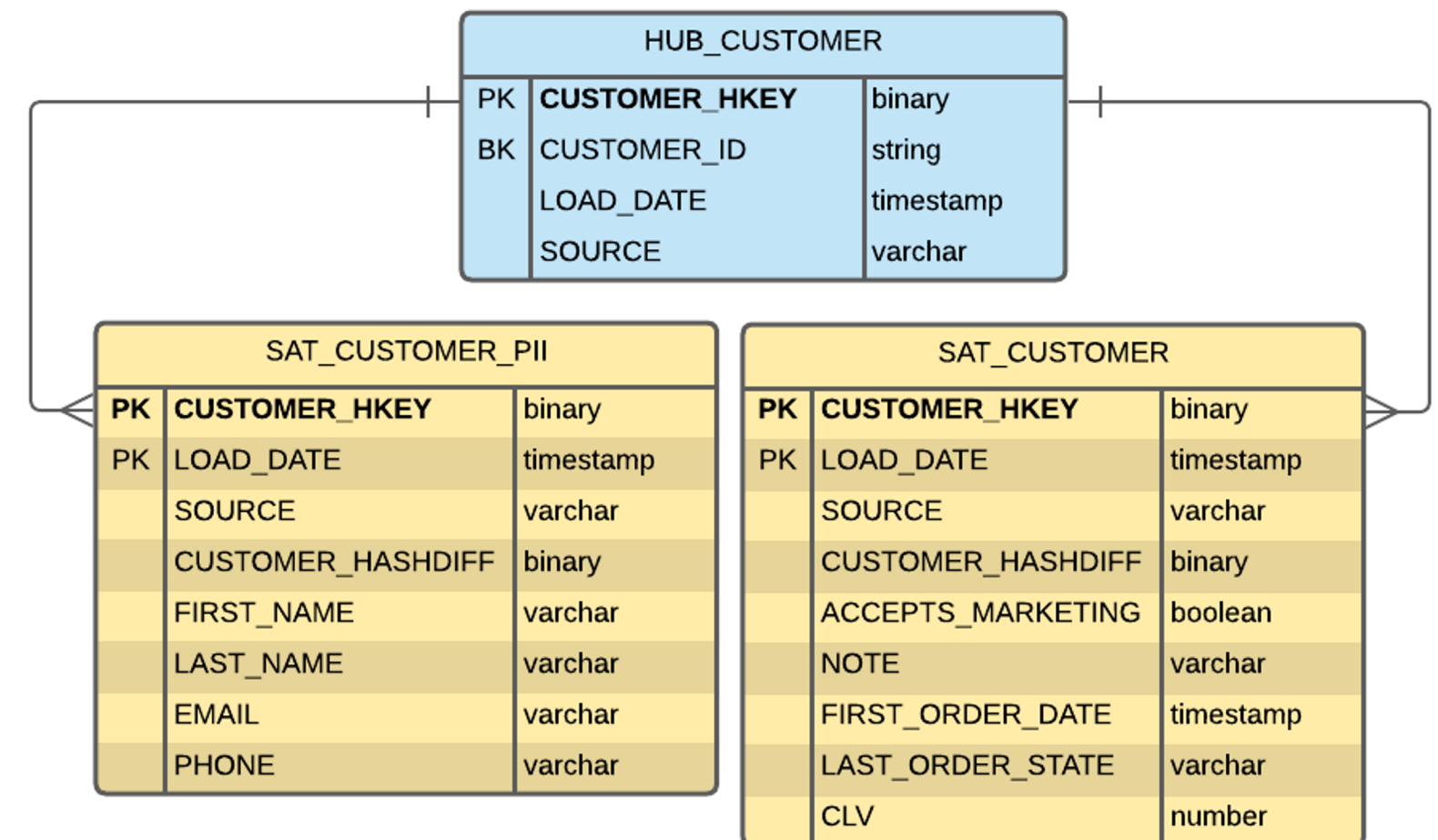
- Template-alapú
- Automatizált
- Adatteszt (null, pk, fk..)
- Üzleti logika teszt

```
models:
  - name: fct_transaction
    description: >
      Card transaction data is financial data collected through the transfer of funds between a card holder's account and a busi...
      It consists of the use of either a debit card or a credit card to generate data on the transfer for the purchase of goods...
      Transaction data describes an action composed of events in which master data participates.
      Transaction focuses on the price, discount and method of payment interaction between the customer and the organization.
    columns:
      - name: id
        description: "The primary key for this table"
        tests:
          - dq_tools.not_null_where_db:
              kpi_category: Completeness
          - dq_tools.unique_where_db:
              kpi_category: Uniqueness
      - name: source
        description: "Name and version of the data source"
        tests:
          - dq_tools.not_null_where_db:
              kpi_category: Completeness
      - name: loaded_at_utc
        description: "System generated date and time when the data was loaded into the table"
        tests:
          - dq_tools.not_null_where_db:
              kpi_category: Completeness
```



Személyes adatok védelme

- Szabályok szerint meghatározott (GDPR/CCPA) és azonosított (profiling)
- Data Vault best practices szerint modellezett
- Azonosítás Snowflake classification, tagging
- Védelem Snowflake dynamic data masking
- Törlés (RTBF) szabályok alapján



Semi structured adat

Snowflake variant típus adattárolásra.

Data Vault model

- Kiterített struktúra
- Tárolás eredeti formában



83
84 `select *`
85 `from "PERSISTENT_STAGING_AREA_PII"."DBT_DATA_FEEDS".PSA__NON_PII__PAYMENT_FAILURE__V1_0`
86 `limit 10;`
87
88
89
90
91
92
93
94
95
96
97

results Data Preview

Query ID SQL

Filter result...

EVENTTYPE

payment-fail... 2022-04-30... 2022-04-30... 2022-05-01... { "body": {}, "header": { "correlationId": "payment_20220501_003103.psv", "eventTimestamp": "2022-04-30T10:20:24.000Z", "eventType": "payment-failure", "fileName": "payment_20220501_003103.psv", "id": "20220430102024F2118CB0857C928864D9E6E67CC44D5AE0FBB0CD92A2971519", "processedTimestamp": "2022-04-30T23:53:51.756Z", "source": "portal", "version": "1.0" } } 7 Both posta... null

payment-fail... 2022-04-30... 2022-04-30... 2022-05-01... { "body": {}, "header": { "correlationId": "payment_20220501_003103.psv", "eventTimestamp": "2022-04-30T10:20:24.000Z", "eventType": "payment-failure", "fileName": "payment_20220501_003103.psv", "id": "20220430102024F2118CB0857C928864D9E6E67CC44D5AE0FBB0CD92A2971519", "processedTimestamp": "2022-04-30T23:53:51.756Z", "source": "portal", "version": "1.0" } } 7 Both posta... null

payment-fail... 2022-04-30... 2022-04-30... 2022-05-01... { "body": {}, "header": { "correlationId": "payment_20220501_003103.psv", "eventTimestamp": "2022-04-30T10:20:24.000Z", "eventType": "payment-failure", "fileName": "payment_20220501_003103.psv", "id": "20220430102024F2118CB0857C928864D9E6E67CC44D5AE0FBB0CD92A2971519", "processedTimestamp": "2022-04-30T23:53:51.756Z", "source": "portal", "version": "1.0" } } 7 Both posta... null

Details

```
1 {
2   "body": {},
3   "header": {
4     "correlationId": "payment_20220501_003103.psv",
5     "eventTimestamp": "2022-04-30T10:20:24.000Z",
6     "eventType": "payment-failure",
7     "fileName": "payment_20220501_003103.psv",
8     "id":
9       "20220430102024F2118CB0857C928864D9E6E67CC44D5AE0FBB0CD92A2971519",
10    "processedTimestamp": "2022-04-30T23:53:51.756Z",
11    "source": "portal",
12    "version": "1.0"
13  }
14 }
```

Done

1. Üzleti érték szállítása gyorsan
2. Együttműködés, megosztás és átláthatóság
3. Gyors változáskezelés
- 4. Technológiai függetlenség biztosítása**
5. Modern data platform és metodológiák

Modellezés függőségek nélkül

- Data Vault modell platform függetlenül
- Könnyű integrálhatóság
- Lazán kapcsolt komponensek (HUB / LINK / SAT)

Rugalmas és lazán kapcsolt technológiák

- Nyelv: **SQL** / python
- Adattranzformáció és automatizáció: **dbt**
- Verziókezelés: **gitlab** / github
- Modellező eszköz: **lucidchart** / **sqldbm**
- Adat platform: **Snowflake** / BigQuery
- Orchestrator: **Airflow** / Prefect
- Cloud: **AWS** / AZURE / GCP

1. Üzleti érték szállítása gyorsan
2. Együttműködés, megosztás és átláthatóság
3. Gyors változáskezelés
4. Technológiai függetlenség biztosítása
- 5. Modern data platform és metodológiák**

