

Challenge 2020

Processo de Seleção para DATA ENGINEER na BOA

O desafio

Empresa

A TUBES & CO é uma indústria familiar que produz tubos flexíveis. Estes tubos são muito utilizados em instalações residenciais e comerciais.

Contexto

A TUBES & CO está passando por um processo de transformação digital.

Descrição do problema

Levar os dados da TUBES & CO para a nuvem. Usando os produtos e serviços do Google.

Sobre os dados do desafio

Price Quote

Esse arquivo contém informações sobre price quotes dos fornecedores. Estes preços podem ser:

- Bracket pricing
- No bracket

Bill of Materials

Esse arquivo contém a lista de componentes e as quantidades que foram utilizadas em cada conjunto de tubos.

Comp Boss

Esse arquivo contém as informações sobre cada componente.

Solução

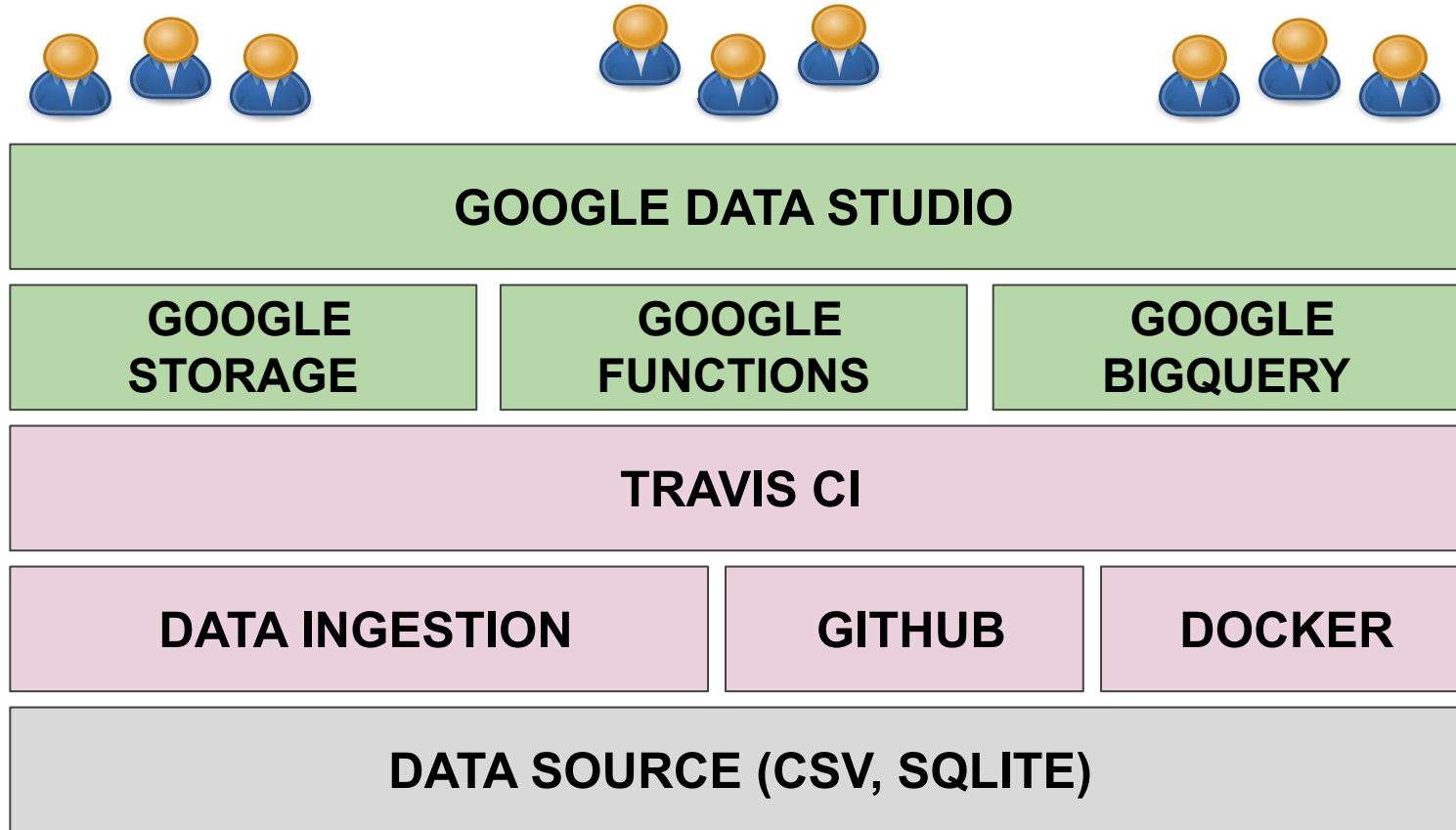
Usando o Google Cloud

GCS - Cloud Storage

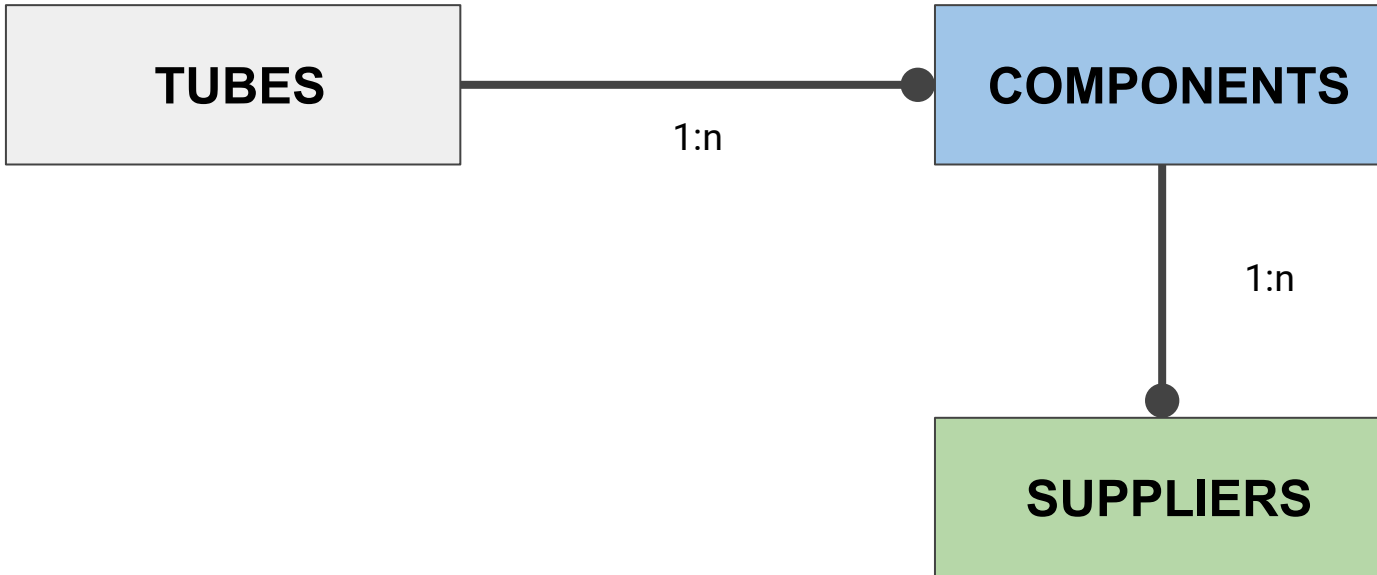
GCF - Cloud Functions (Python)

GBQ - Cloud BigQuery

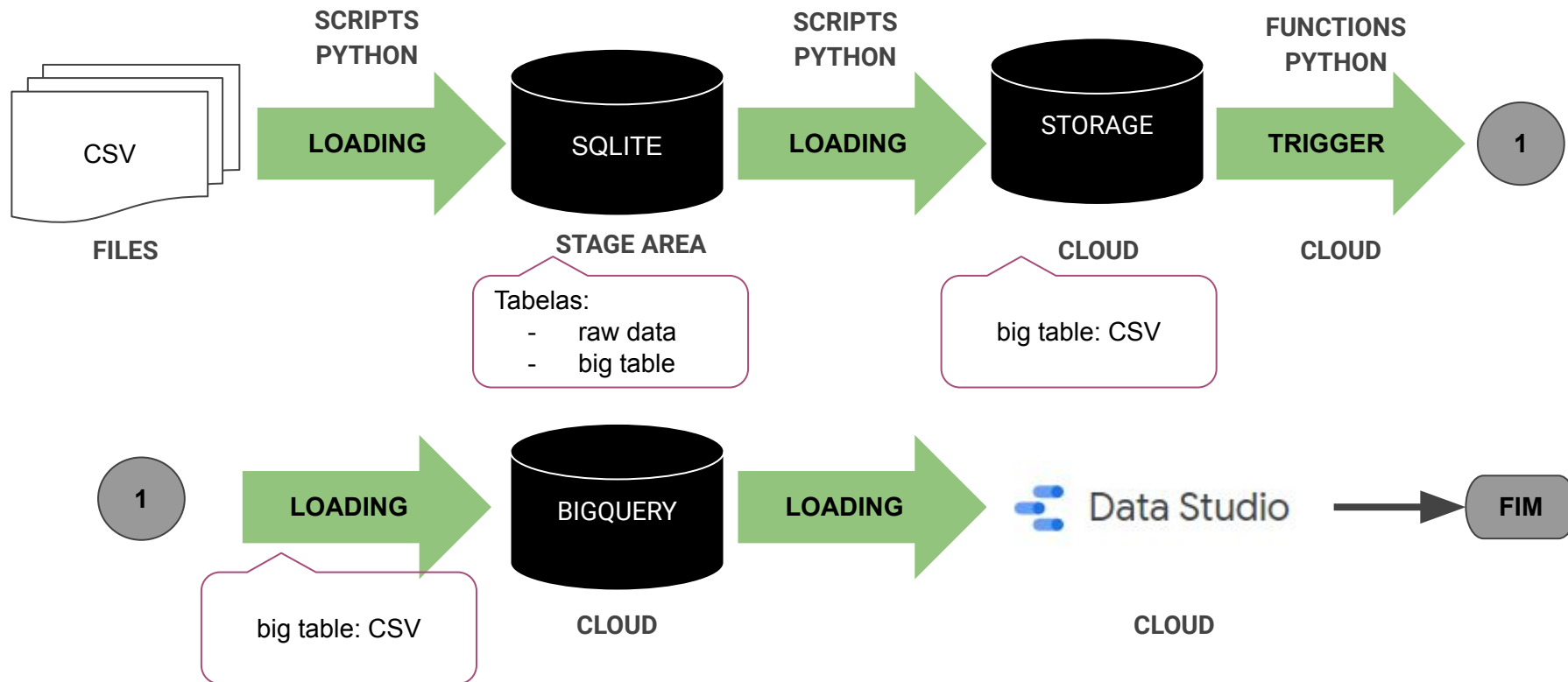
Arquitetura



Modelo de dados



Fluxo de dados

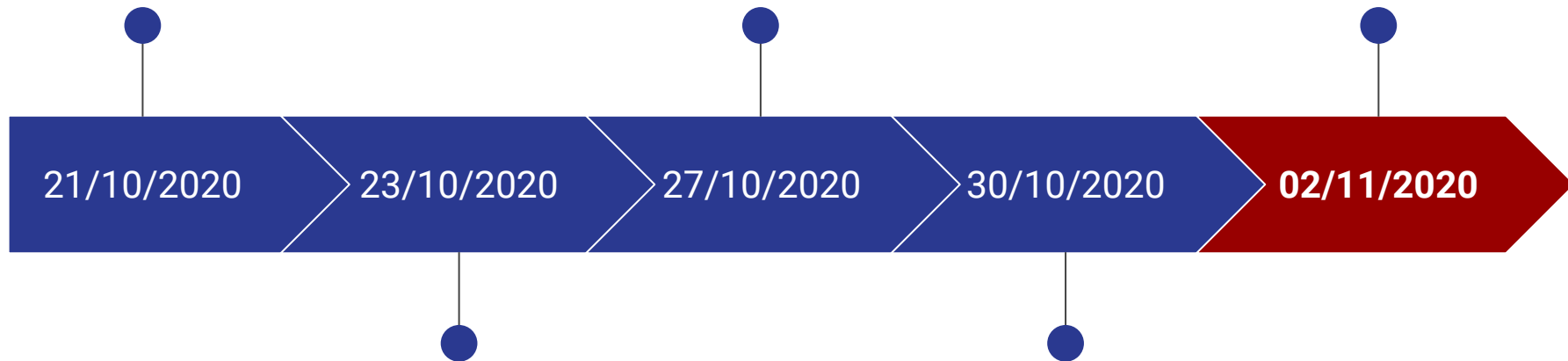


Timeline

Entrevista com o RH e recebimento do material do teste após a entrevista com o RH

Entrevista técnica e alinhamento de expectativas

Entrega do teste ao RH para análise técnica



21/10/2020

23/10/2020

27/10/2020

30/10/2020

02/11/2020

Análise do material do teste, planejamento das atividades e feedback ao RH

Retomada das atividades do teste depois dos problemas técnicos

Considerações Finais

O processo

Estou gostando muito de conhecer a empresa, os desafios e as oportunidades que existem.

Também estou gostando bastante da forma como o processo está sendo conduzido.

O desafio

Era pra ser um desafio DE BOA. Mas, acabei encarando alguns imprevistos no meio do caminho que me tiraram do foco. Mas, com perseverança e resiliência segui em frente!

O futuro

O futuro é incerto.

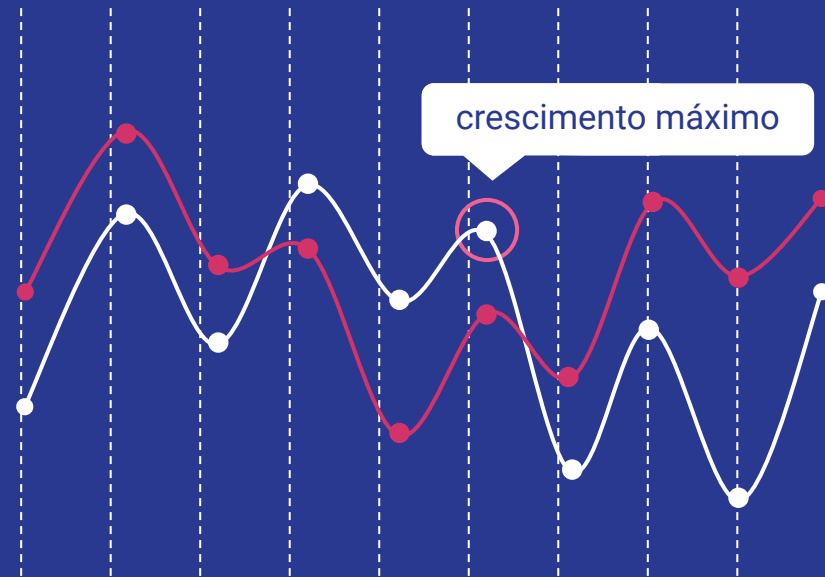
Mas hoje a única certeza que eu tenho, é o meu desejo de fazer parte desta equipe e me tornar um DE BOA !

(Data Engineer na BOA).

Thank you!

Laercio Serra

laercio.serra@gmail.com



Gerando Relatórios EDA

WELCOME TO BOA CHALLENGE FOR DATA ENGINEER!

NAME: LAERCIO SERRA

=> [OK] - EXPLORATORY DATA ANALYSIS

```
>> Generating Profile Report: comp_boss.csv
```

```
>> Exporting Profile Report to HTML: comp_boss.csv
```

```
Summarize dataset: 100% | 29/29 [00:07<00:00, 3.76it/s, Completed]
```

```
Generate report structure: 100% | 1/1 [00:06<00:00, 6.28s/it]
```

Render HTML: 100% | 1/1 [00:00<00:00, 1.00it/s]

```
Export report to file: 100% | 1/1 [00:00<00:00, 259.47it/s]
```

```
>> Generating Profile Report: bill_of_materials.csv
```

```
>> Exporting Profile Report to HTML: bill_of_materials.csv
```

```
Summarize dataset: 100% | 31/31 [00:03<00:00, 7.94it/s, Completed]
```

```
Generate report structure: 100% | 1/1 [00:06<00:00, 6.88s/it]
```

Render HTML: 100% 1/1 [00:00<00:00, 1.59it/s]

```
Export report to file: 100% | 1/1 [00:00<00:00, 311.57it/s]
```

```
>> Generating Profile Report: price_quote.csv
```

```
>> Exporting Profile Report to HTML: price_quote.csv
```

Google Cloud Storage

Google Cloud Platform

MyFirstProject

Search products and resources

Bucket details

REFRESH

Updated 2 hours ago

my-bigdata-projects

OBJECTS

CONFIGURATION

PERMISSIONS

RETENTION

LIFECYCLE

Buckets > my-bigdata-projects > data > csv

UPLOAD FILES

UPLOAD FOLDER




CREATE FOLDER

MANAGE HOLDS

DELETE

Filter

Filter by object or folder name prefix

<input type="checkbox"/>	Name	Size	Type	Created time ?	Storage class	Last modified	Public access ?	Encryption ?	Retention expiration date ?	Holds ?	
<input type="checkbox"/>	 bt_challenge_bo	45.1 KB	text/csv	Nov 2, 2020, 8:...	Standard	Nov 2, 202...	Not public	Google-managed key	—	None	 

Google Cloud Functions

Google Cloud Platform

MyFirstProject

Search products and resources

Cloud Functions

Functions

CREATE FUNCTION

REFRESH

DELETE

COPY

SHOW INFO PANEL

Filter functions

		Name ↑	Region	Trigger	Runtime	Memory allocated	Executed function	Last deployed	Authentication ?	Actions
		CSV_LOADER_GCS_GBQ	us-central1	Bucket: my-bigdata-projects	Python 3.7	256 MiB	csv_loader	Nov 1, 2020, 9:51:39 PM		

Google Cloud Functions

Google Cloud Platform MyFirstProject Search products and resources

7:55:51 AM - 9:55:51 AM PAGE LAYOUT

Logs Explorer OPTIONS REFINE SCOPE Project

New features are available in the Logs Explorer. Dismiss Learn more

Query preview
resource.type="cloud_function" resource.labels.function_name="CSV_LOADER_GCS_GBQ" resource.labels.region="us-central1"

Save Run Query

Log fields

Search fields and values

LOG NAME
cloudfunctions.googleapis.com/cloud-functio... 7

FUNCTION_NAME
CSV_LOADER_GCS_GBQ Clear X

PROJECT_ID
sturdy-conduit-260900 7

REGION
us-central1 Clear X

RESOURCE TYPE
Cloud Function Clear X

SEVERITY
Info 5

Histogram

Nov 2, 7:54 AM 8:30 AM 9:00 AM 9:30 AM Nov 2, 9:56 AM

Query results Jump to Now Actions Configure

SEVERITY	TIMESTAMP	BRT	SUMMARY
> [i]	2020-11-02 08:54:00.439 BRT	CSV_LOADER_GCS_GBQ qq31491jbv92	Function execution started
> [i]	2020-11-02 08:54:01.324 BRT	CSV_LOADER_GCS_GBQ qq31491jbv92	>> Function: CSV LOADER GCS TO GBQ
> [i]	2020-11-02 08:54:01.325 BRT	CSV_LOADER_GCS_GBQ qq31491jbv92	>> Storage: gs://my-bigdata-projects/data/csv/bt_chall...
> [i]	2020-11-02 08:54:01.325 BRT	CSV_LOADER_GCS_GBQ qq31491jbv92	>> Starting Job: ef6e2ab0-c515-4415-bd0d-a520face7382
> [i]	2020-11-02 08:54:03.005 BRT	CSV_LOADER_GCS_GBQ qq31491jbv92	>> Total Rows: 342
> [i]	2020-11-02 08:54:03.005 BRT	CSV_LOADER_GCS_GBQ qq31491jbv92	>> Job finished successfully!

Google BigQuery

Query editor

1

 Run  Save query  Save view  Schedule query

bt_challenge_boa

Schema

Details

Preview

Field name	Type	Mode	Policy tags 
tube_assembly_id	STRING	NULLABLE	
quantity_component	FLOAT	NULLABLE	
component_id	STRING	NULLABLE	
component_type_id	STRING	NULLABLE	
component_type	STRING	NULLABLE	
connection_type_id	STRING	NULLABLE	
outside_shape	STRING	NULLABLE	
base_type	STRING	NULLABLE	

bt_challenge_boa

 QUERY TABLE

 SHARE TABLE

Schema

Details

Preview

Description 

Big Table Challenge BOA

Labels 

challenge : boa


Table info 

Table ID	sturdy-conduit-260900:de_boa.bt_challenge_boa
Table size	46.2 KB
Number of rows	342
Created	Nov 1, 2020, 6:19:54 PM
Table expiration	Nov 11, 2020, 6:19:54 PM
Last modified	Nov 2, 2020, 11:25:34 AM
Data location	US

Google BigQuery

Google Cloud Platform

MyFirstProject

Search products and resources

BigQuery

FEATURES & INFO

SHORTCUT

Query history

Saved queries

Job history

Transfers

Scheduled queries

Reservations

BI Engine

Resources

+ ADD DATA

Search for your tables and datasets

sturdy-conduit-260900

listenbrainz

Query editor

+ COMPOSE NEW QUERY

HIDE EDITOR

FULL SCREEN

1 SELECT * FROM `sturdy-conduit-260900.de_boa.bt_challenge_boa` LIMIT 1000

Run

Save query

Save view

Schedule query

More

This query will process 46.2 KB when run.

Query results

SAVE RESULTS

EXPLORE DATA

Query complete (0.4 sec elapsed, 46.2 KB processed)

Job information

Results

JSON

Execution details

Row	tube_assembly_id	quantity_component	component_id	component_type_id	component_type	connection_type_id	outside_shape	base_type	height_over_tube	bolt_pattern
1	TA-01947	1.0	C-0161	CP-020	null	9999	null	null	19.2	
2	TA-11878	1.0	C-0161	CP-020	null	9999	null	null	19.2	
3	TA-13310	4.0	C-0243	CP-019	null	B-012	null	null	36.0	
4	TA-14192	4.0	C-0243	CP-019	null	B-012	null	null	36.0	
5	TA-13359	1.0	C-0770	CP-020	null	9999	null	null	21.5	
6	TA-10009	1.0	C-0780	CP-020	null	9999	null	null	21.5	
7	TA-20386	2.0	C-1919	CP-019	null	B-012	null	null	35.0	

Rows per page: 100 1 - 100 of 342 First page < < > > Last page

Google Data Studio

Data Studio

Pesquisar no Data Studio



Criar

Recentes

Relatórios

Origens de dados

Explorador



Recentes



Compartilhados
comigo



De minha propriedade



Lixeira

Começar com um modelo

Galeria de modelos



Relatório em branco
Data Studio



Relatório de tutorial
Data Studio



Marketing da Acme
Google Analytics



Search Console Report
Search Console

Nome

Pertencem a qualquer pessoa

Última vez que abri



Executive Dashboard

Laércio Serra

10:16

Google Data Studio

 Data Studio

 Pesquisar no Data Studio



Criar

Recentes

Relatórios

Origens de dados

Explorador



Recentes



Compartilhados
comigo



De minha propriedade



Lixeira



As fontes de dados incorporadas são um novo tipo de fonte gerenciada com um relatório. Como elas são associadas a um único relatório, não aparecem nessa lista e, por padrão, não é possível reutilizá-las. [Saiba mais](#)

[Dispensar](#)

Nome

Pertencem a qualquer pessoa ▼

Última vez que abri ▼




BigQuery - 01/11/2020 23:36

Laércio Serra

00:49

Google Data Studio

 Data Studio

 Pesquisar no Data Studio



Criar



Recentes



Compartilhados
comigo



De minha propriedade



Lixeira

Recentes

Relatórios

Origens de dados

Explorador

Nome

Última modificação






Data Explorer

10:16




Google Data Studio

Executive Dashboard

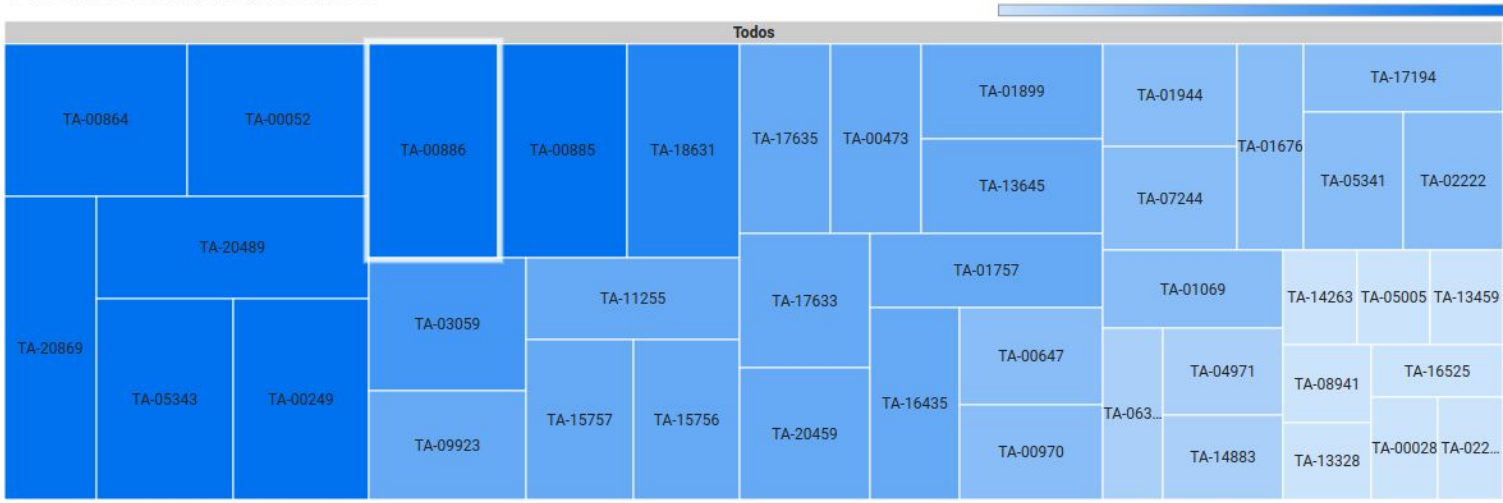


Compartilhar

Edit



Produtos Mais Fabricados



Custo do Produto por Componente

Comp ID / Custo						
Tube ID	C-0699	C-0199	C-0817	C-1536	C-1231	C-1725

Top 10 Suppliers