

# ANALISANDO ÍNDICES ENTRE OS BANCOS TRADICIONAIS E OS DIGITAIS

**Andre Ventura**

**Ewerton Matias**

**Helder Moura**

**Monnicy Tonon**

**Renata Santana**

- Proposta do projeto;
- Objetivo do Projeto;
- Datasets;
- Instituições financeiras;
- Escopo do Projeto;
- Desenvolvimento do Projeto;
- Análises dos dados.

# Proposta:

Criar uma réplica do banco de produção do cliente e tratá-los utilizando os processos de ELT através de procedures, gerando um Data Mart em nuvem.

Popular um Data Warehouse Cassandra utilizando dados brutos, tratados no processo de ETL via PySpark.

# Objetivo:

O objetivo do projeto é extrair dados provenientes do banco de produção do cliente, acerca das instituições financeiras selecionadas; tratar esses dados para a criação de um Data Mart com informações devidamente filtradas permitindo que os dados sejam analisados de forma eficiente e trazer para o cliente insights significativos que permitam a tomada de decisão com maior segurança.

# Datasets

---



Dados brutos totais:

**+ 5.5 milhões**

[HTTPS://WWW3.BCB.GOV.BR/IFDATA/](https://www3.bcb.gov.br/IFDATA/)

# Datasets

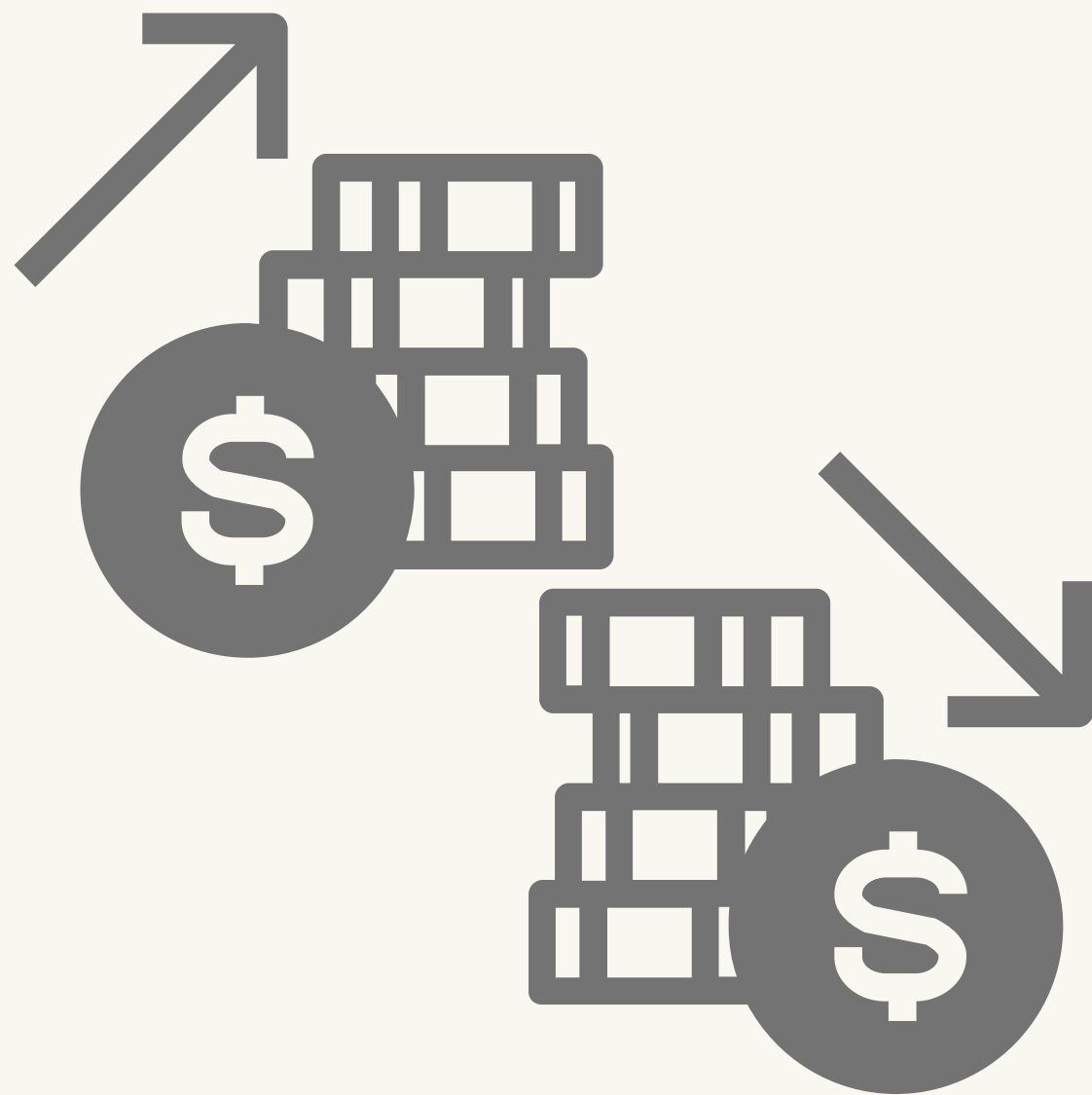
---



Buscamos nossos dados em sites que apresentaram o nível de satisfação dos consumidores com os serviços bancários.

**+ 3.4 milhões**

[HTTPS://WWW.CONSUMIDOR.GOV.BR/PAGES/DADOSABERTOS/EXTERNO](https://www.consumidor.gov.br/pages/dadosabertos/externo)



## Datasets

---

Além da satisfação de usuários com os serviços, também analisamos as ações desses bancos, as oscilações, pré pandemia até os dias atuais (2019 - 2021).

**+ 34 mil**

YAHOO FINANÇAS



# Datasets

---



Também optamos por verificar dados que envolvem as solicitações de financiamentos e os bancos envolvidos.

**+ 2.1 milhões**

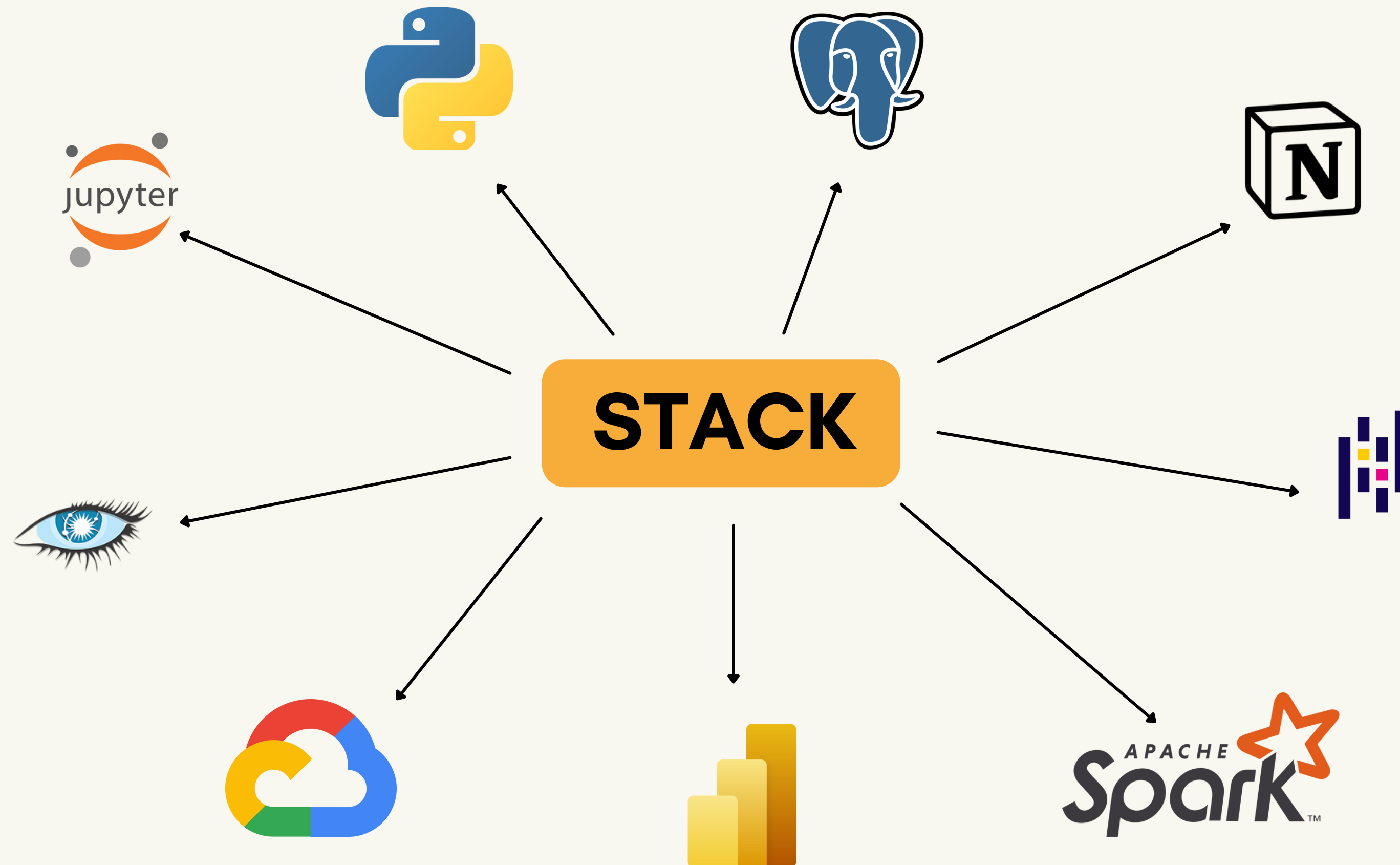
[HTTPS://WWW3.BCB.GOV.BR/IFDATA/](https://www3.bcb.gov.br/IFDATA/)

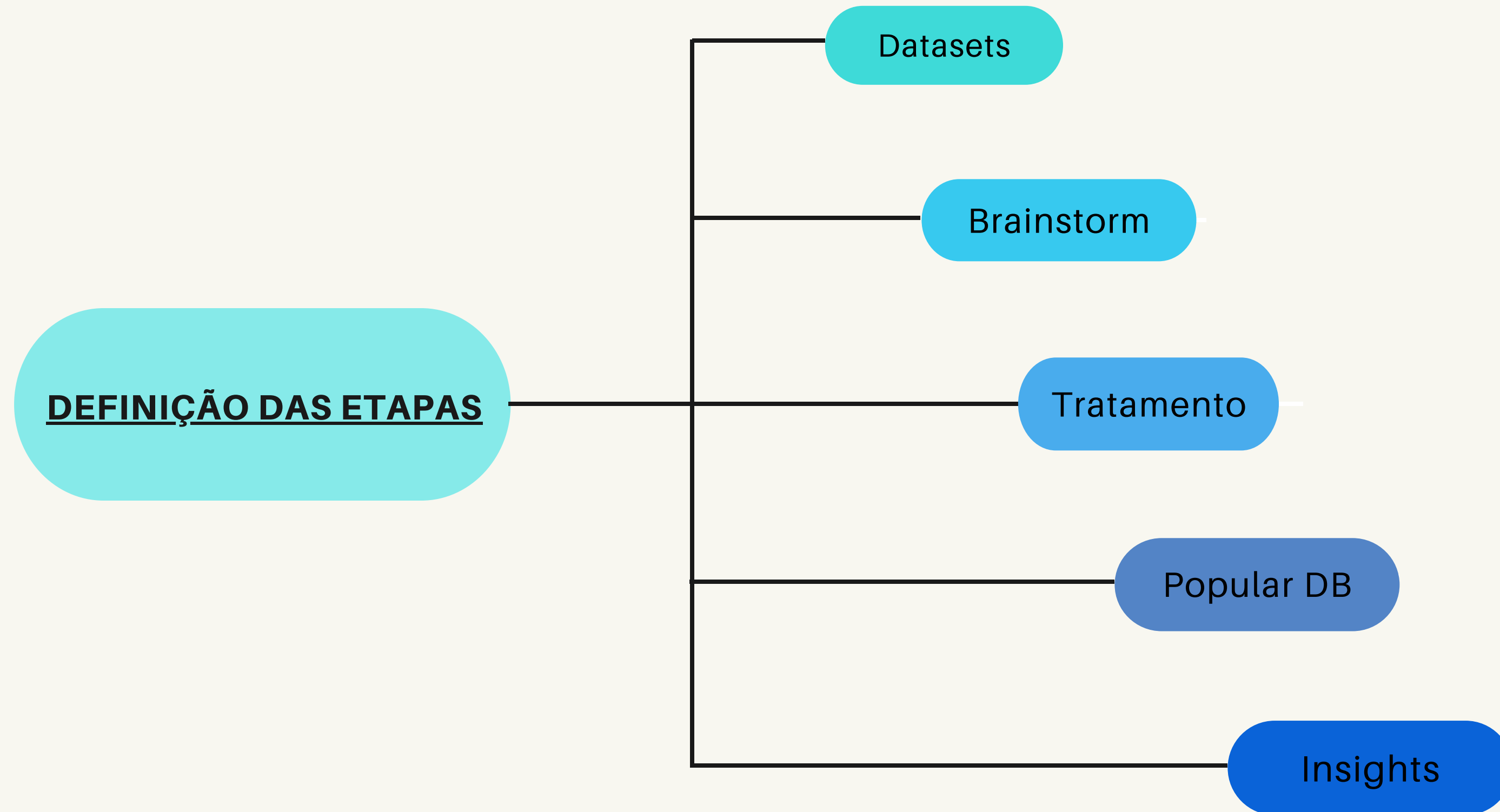
# Bancos digitais:

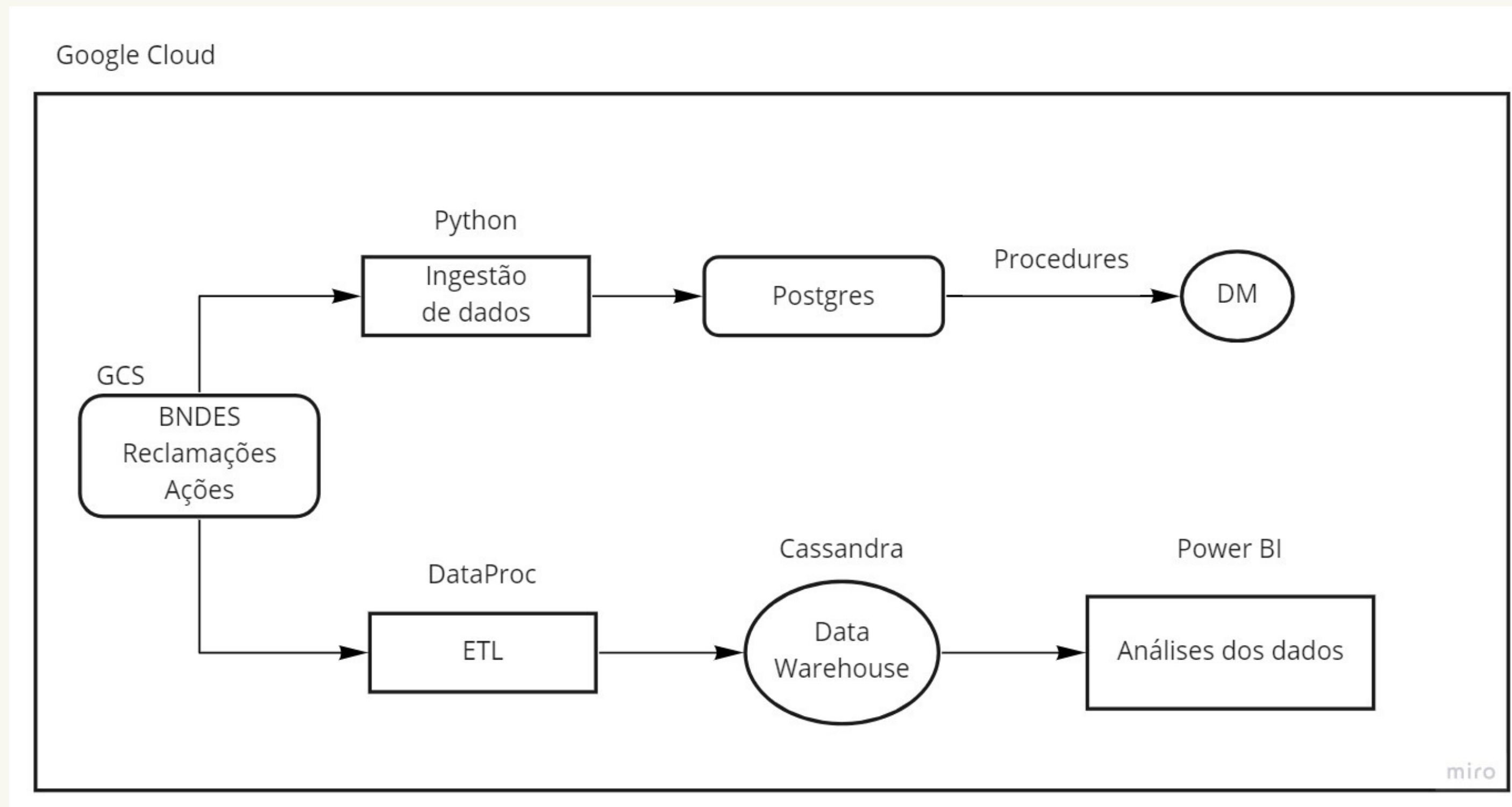
- PAN
- BMG
- INTER
- MODAL
- NUBANK

## Bancos tradicionais (físicos):

- BRADESCO
- SANTANDER
- ITAÚ UNIBANCO
- BANCO DO BRASIL
- CAIXA ECONÔMICA FEDERAL







# Preparação da Cloud

SQL	Instâncias	+ CR
Filtro Insira o nome ou o valor da propriedade		
<input type="checkbox"/>	ID da instância ? ↑	Tipo
<input type="checkbox"/>	<input checked="" type="radio"/> trabalho-final	PostgreSQL 13

Conexões	
Redes autorizadas	
Especifique intervalos CIDR para permitir que endereços IP nesses intervalos acessem a instância. <a href="#">Saiba mais</a>	
worker-1 (34.151.240.47)	▼
worker-0 (34.151.213.102)	▼
helder (200.124.166.217)	▼
andre (131.0.48.23)	▼
cluster-master (34.151.208.44)	▼

BUCKET

POSTGRES

CLUSTER

PERMISSÕES

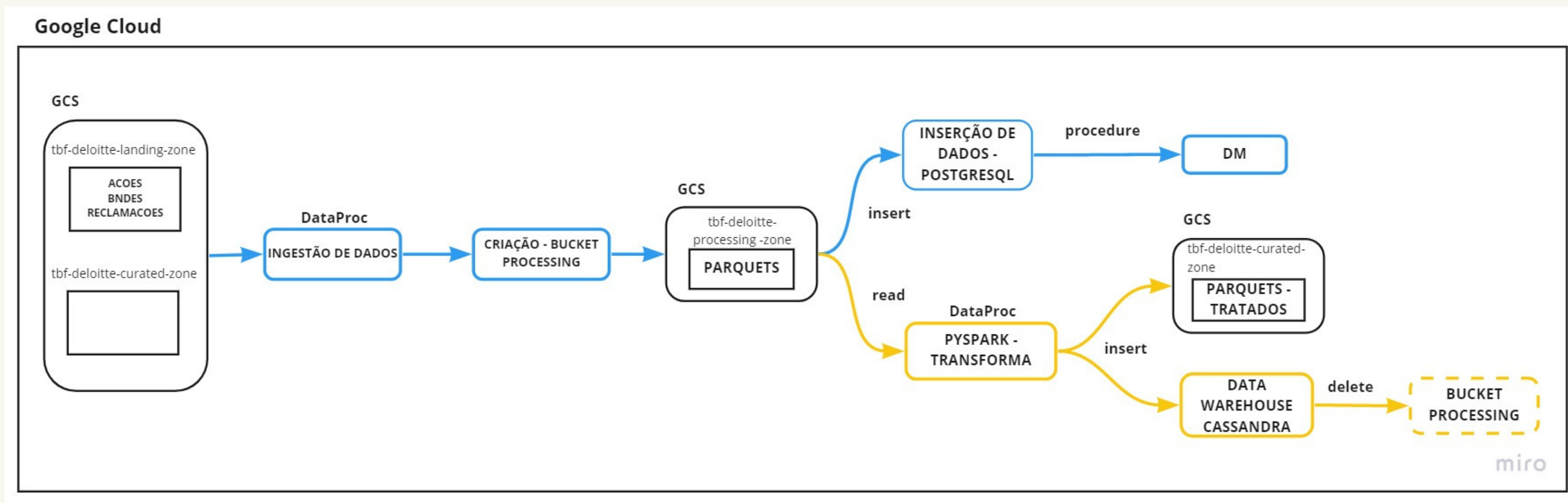
CASSANDRA

<input type="checkbox"/>	Nome ↑
<input type="checkbox"/>	dataproc-staging-us-central1-534305...
<input type="checkbox"/>	dataproc-temp-us-central1-53430531...
<input type="checkbox"/>	tbf-deloitte-codigos
<input type="checkbox"/>	tbf-deloitte-curated-zone
<input type="checkbox"/>	tbf-deloitte-landing-zone

Nome	cluster-610d
UUID do cluster	5590401b-dcbd-486e-a
Tipo	Cluster do Dataproc
Status	✓ Em execução

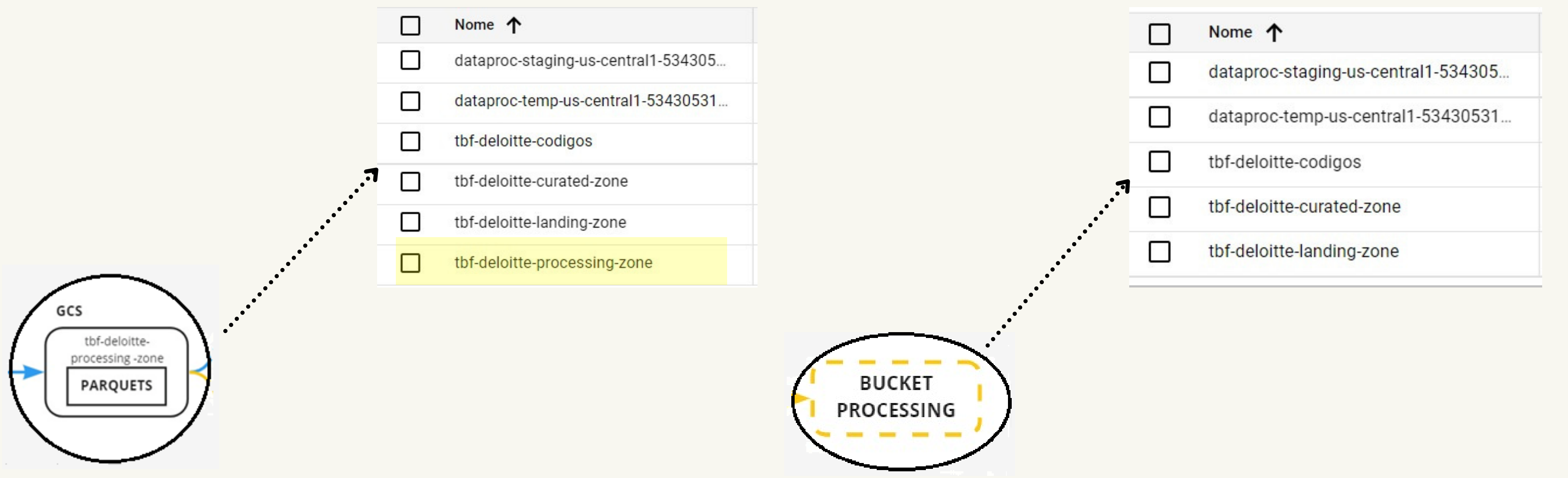
Inclusão do cluster

# Pipeline





# Pipeline - tbf-deloitte-processing-zone



# Ingestão de dados no PostgreSQL

Utilizamos um conector JDBC via PySpark para instanciar a conexão

CONEXÃO

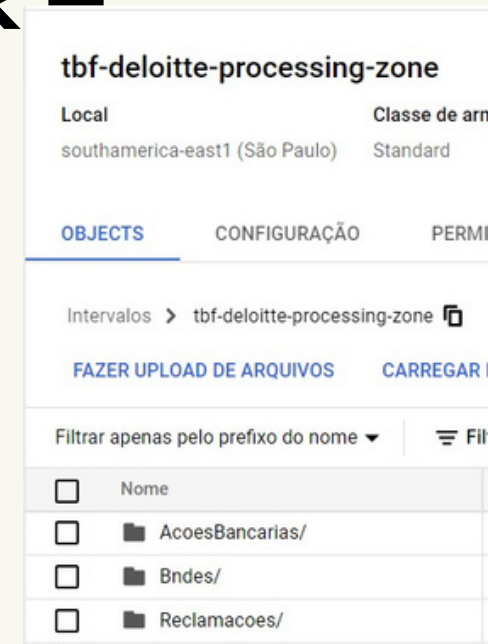
LEITURA

TRATAMENTO

PARQUET

INSERÇÃO

Usamos Pandas/PySpark para ler e renomear as colunas do DF evitando problemas de codificação de texto



```
Lendo os CSVs em tbfdeloitte-landing-zone
Concatenando os DataFrames de reclamacoes
Renomeando as colunas dos DataFrames
Criando o Bucket tbfdeloitte-processing-zone
Bucket tbfdeloitte-processing-zone created.
Escrevendo os parquets brutos em tbfdeloitte-processing-zone para ser utilizado pelo Pyspark e inserido no Cassandra
22/01/21 01:35:46 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageFileSystem: Successfully repaired
22/01/21 01:35:53 WARN org.apache.spark.sql.catalyst.util.package: Truncated the string representation of a plan since it was too large. This be
22/01/21 01:36:39 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageFileSystem: Successfully repaired
22/01/21 01:37:41 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageFileSystem: Successfully repaired
Inserindo os dados de acoes
insercao acoes - 0:00:32.127978

Inserindo os dados de bndes
insercao bndes - 0:06:45.188749

Inserindo os dados de reclamacoes
insercao reclamacoes - 0:12:59.688572
```

# Gerando o Data Mart

PROCEDURES

TRATAMENTO

TEMP. TABLE

DATA MART

Tratamento completo dos dados, gerando tabelas temporárias que serão consumidas pelo Data Mart

# Gerando o Data Warehouse - Cassandra

Leitura dos arquivos  
Parquet dentro do  
bucket *tbj-deloitte-  
processing-zone*

Criação dos arquivos  
Parquet tratados em  
*tbj-deloitte-curated-  
zone*

CONEXÃO

LEITURA

TRATAMENTO

PARQUET

INSERÇÃO

Utilizado o driver  
Cassandra  
Pyspark


Filtragem de linhas  
e colunas; alteração  
de tipos e ajuste de  
pontuações.

Inserção do  
DataFrame tratado  
no Cassandra

# Resultados dos processos


DM

DW

trabalho-final on postgres@trabalho_final		Data Output	
1	<code>select count(*) from dm_reclamacoes;</code>		count bigint
		1	420867




▷ Run on active connection   ≡ Select block   ▶ Run SQL		count
	<code>select count(*) from reclamacoes;</code>	abc Filter...
		420867

trabalho-final on postgres@trabalho_final		Data Output	
1	<code>select count(*) from dm_bndes;</code>		count bigint
		1	1072713



▷ Run on active connection   ≡ Select block   ▶ Run SQL		count
	<code>select count(*) from bndes;</code>	abc Filter...
		1072713

trabalho-final on postgres@trabalho_final		Data Output	
1	<code>select count(*) from dm_acoes;</code>		count bigint
		1	34211



▷ Run on active connection   ≡ Select block   ▶ Run SQL		count
	<code>select count(*) from acoes;</code>	abc Filter...
		34211

# Análise dos dados

## CONEXÃO

Conexão externa  
do Power BI com  
o Cassandra

## ANÁLISE





# Obrigado!



**Andre Ventura**  
[/andreventuraa](#)



**Ewerton Matias**  
[/ewerton-matias](#)



**Helder Moura**  
[/heldermourabr](#)



**Monnicy Tonon**  
[/monnicy-tonon](#)



**Renata Santana**  
[/renatassantana](#)