



SOUL CODE

Relatório Telecomunicações

Fixo - Móvel | 2019 - 2020 - 2021

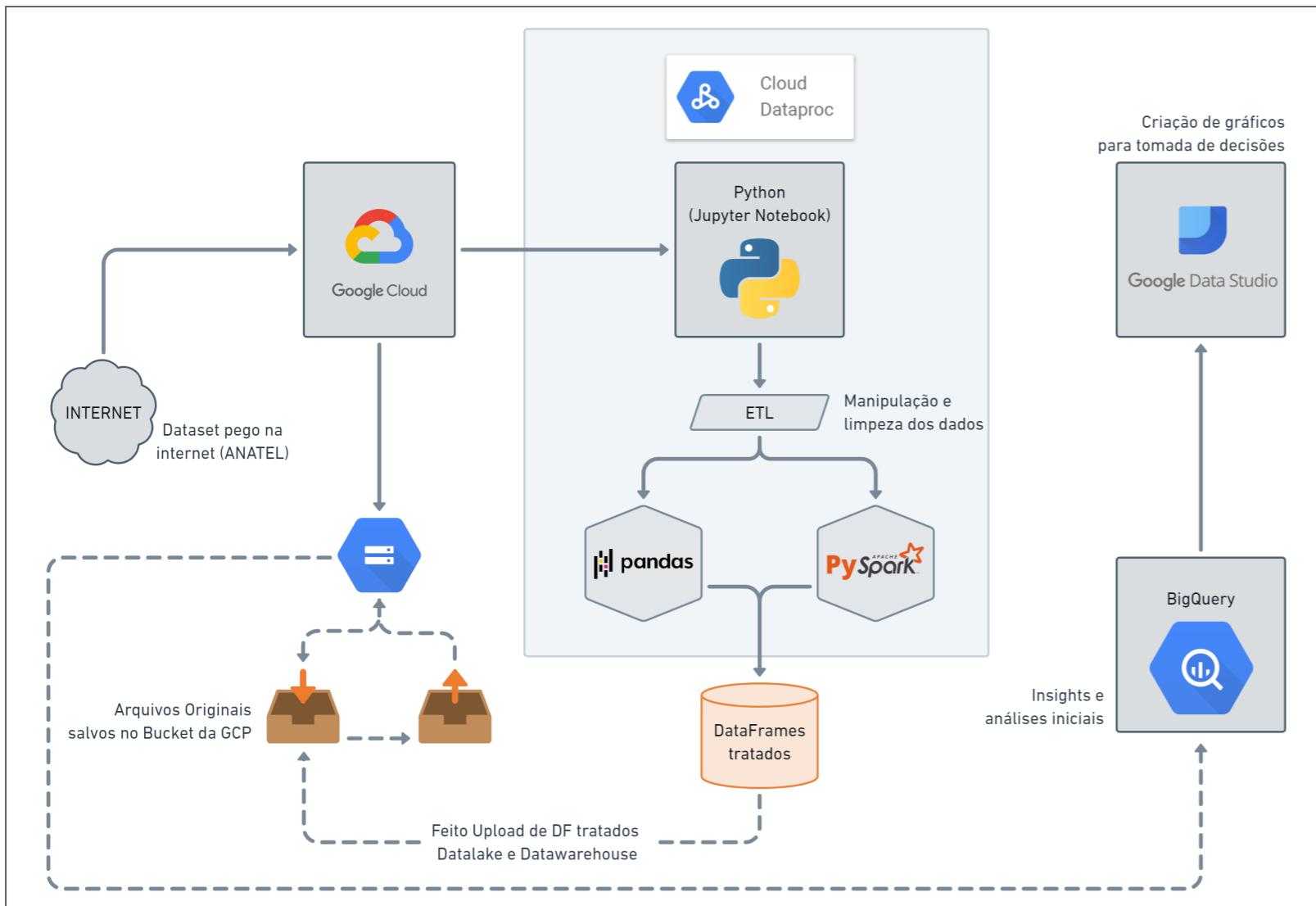


PROJETO TELECOM



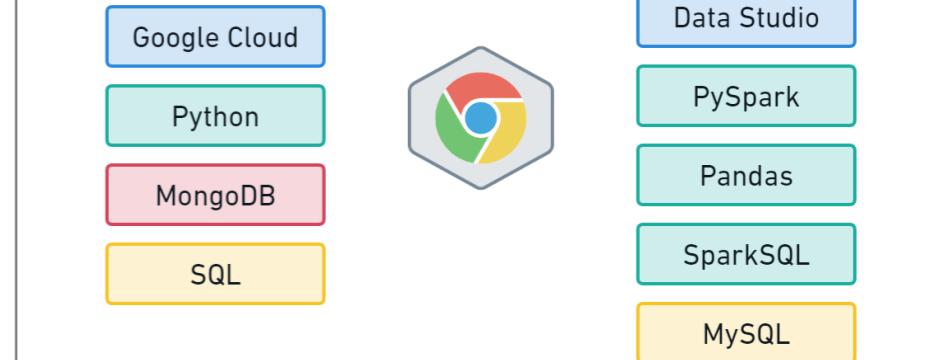
Requisitos do projeto

- Operações em Pandas, PySpark e SparkSQL
- Armazenamento dos Datasets no GCP - Google Cloud Platform e MySQL
- Dados tratados armazenados em Datalake e Data Warehouse do GCP
- Armazenar Data frames em coleções no Mongo DB Atlas
- Realizar análises no BigQuery em linguagem padrão SQL
- Dashboard trazendo dados tratados utilizando Datastudio
- Workflow de ETL com suas respectivas ferramentas



Workflow

Tecnologias e Linguagens

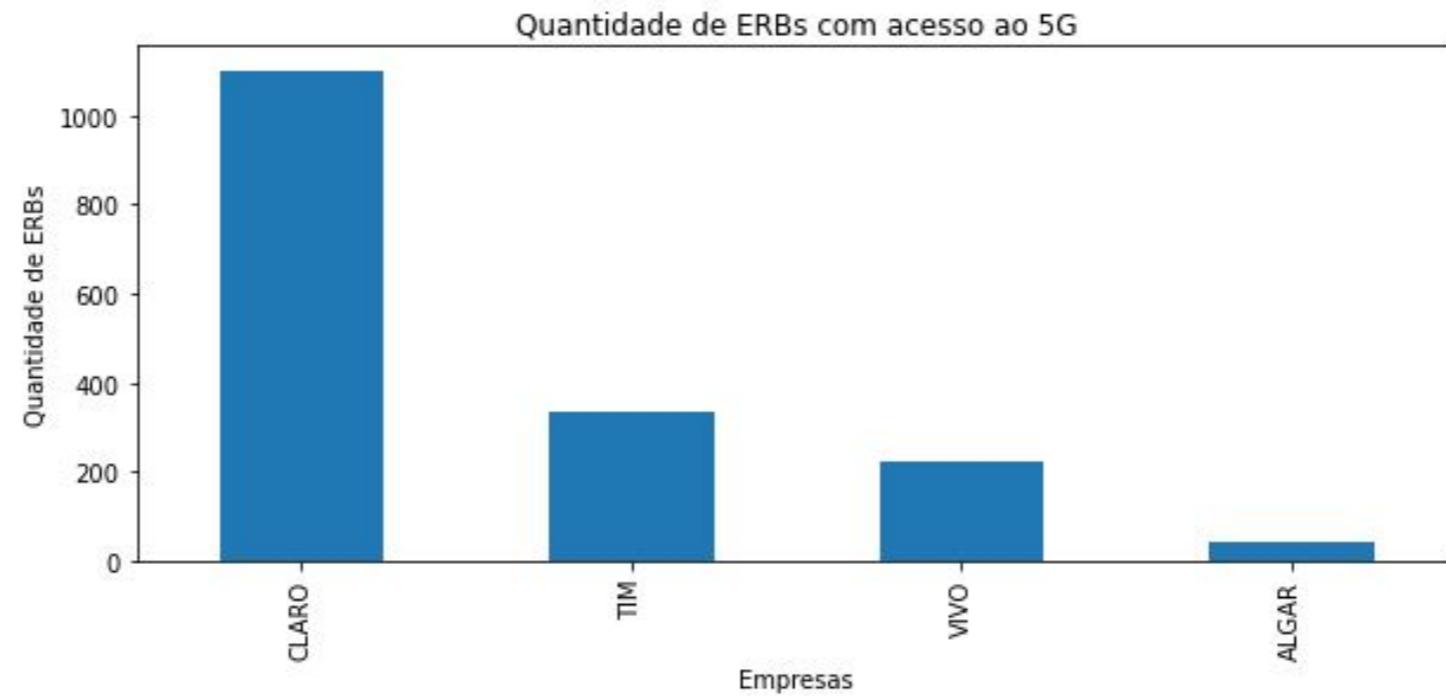




Procedimento de análise dos dados: Banda Larga Móvel

```
In [25]: print(dfa[dfa['5G'] == 1]['Empresas'].value_counts().plot.bar\\\n    (xlabel='Empresas', ylabel='Quantidade de ERBs',\\\n     title='Quantidade de ERBs com acesso ao 5G', figsize=(10,4)))
```

AxesSubplot(0.125,0.125;0.775x0.755)



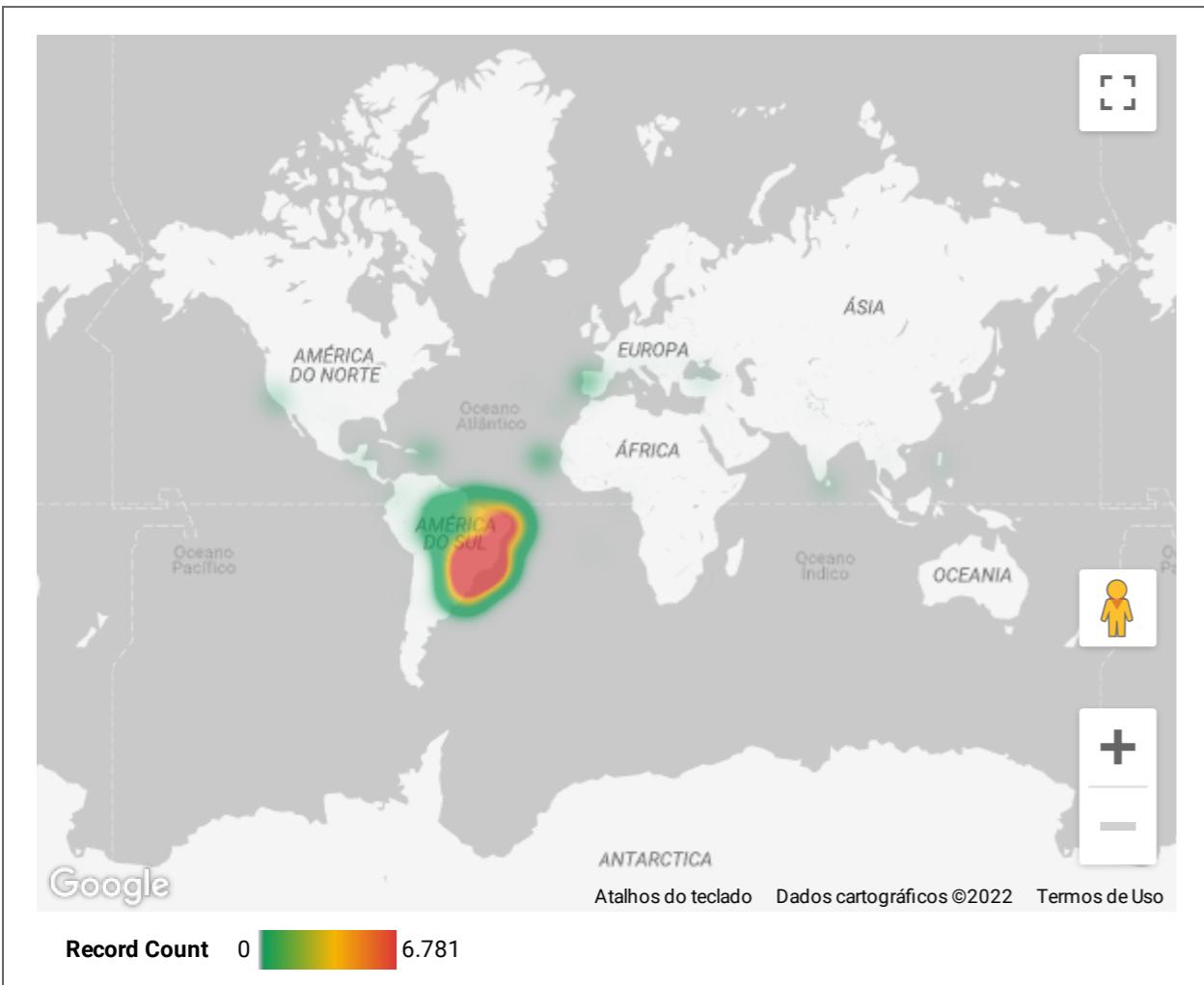


POR QUE A OI NÃO OFERECE TECNOLOGIA 5G?

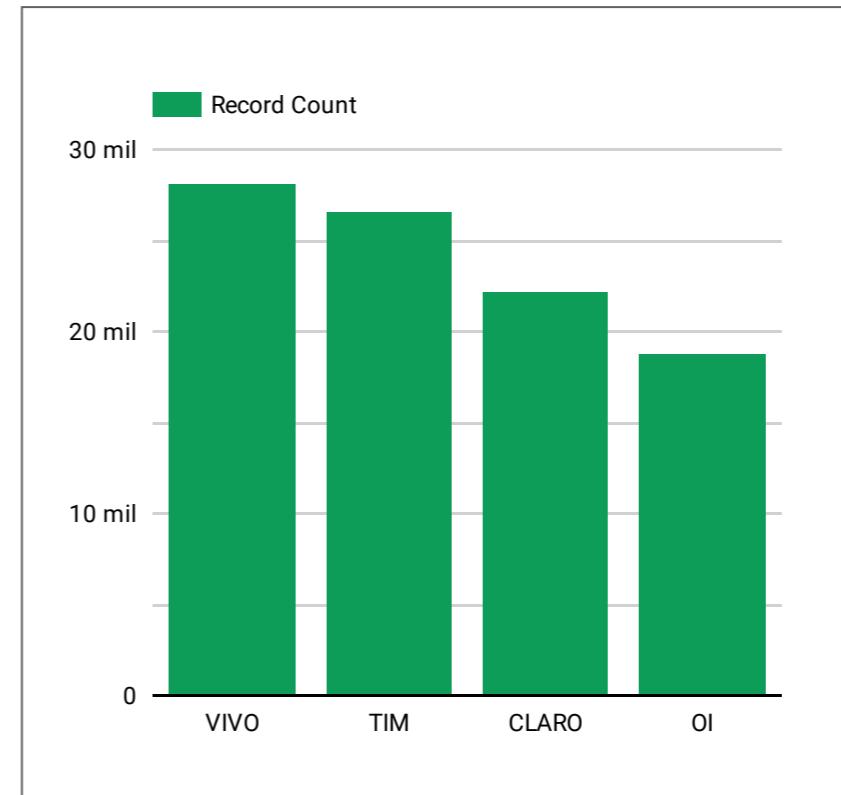




Predominância das ERBs no Brasil



ERBs por Operadoras



*ERB - Estação de Rádio Base

ERBs Prioritária

96.490

Tecnologia 2G

58.982

Tecnologia 3G

74.518

Tecnologia 4G

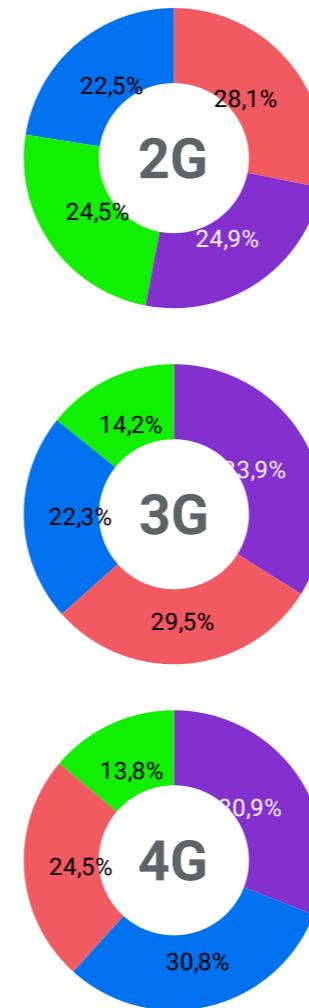
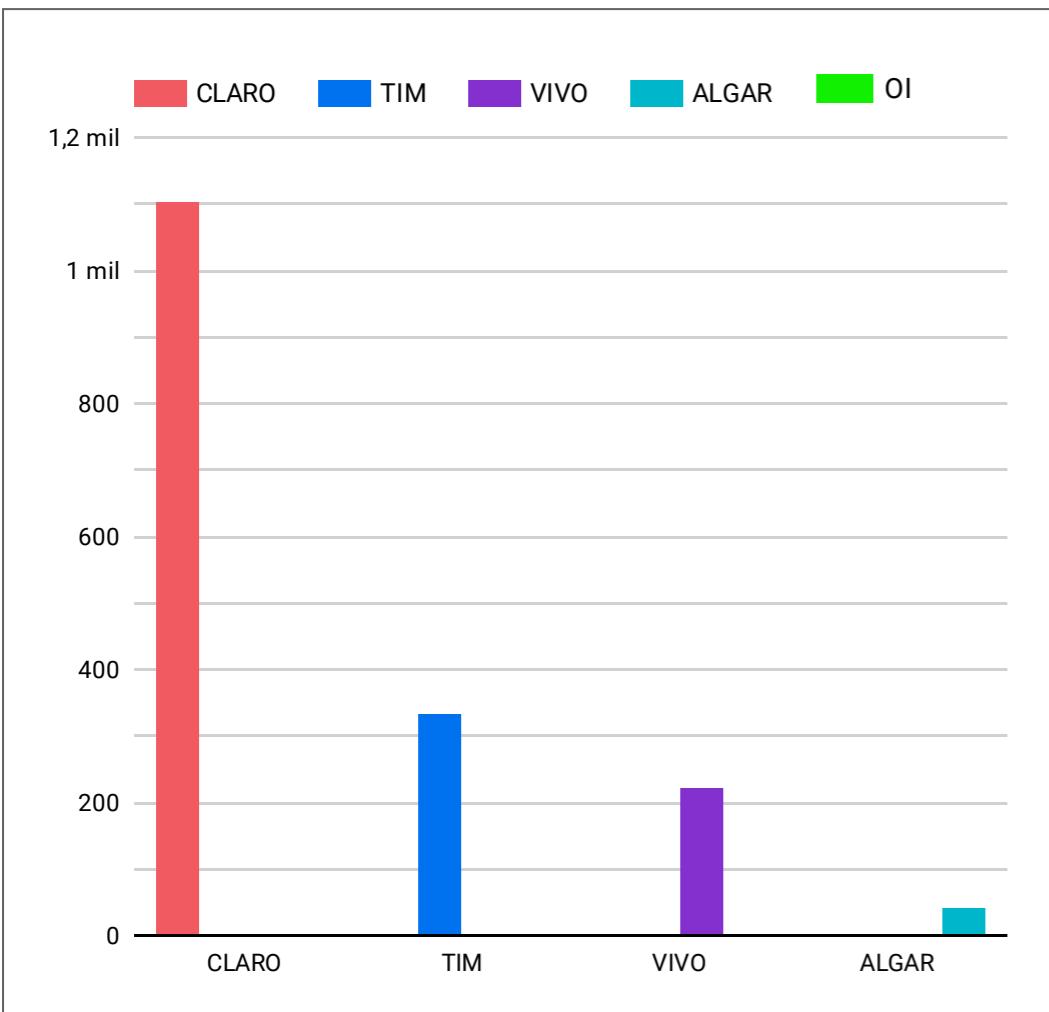
81.169

Tecnologia 5G

1.702



Empresas com Estações 5G



Recuperação Judicial OI

Em setembro de 2016, com uma dívida de R\$ 65 bilhões, a Oi teve aprovado o seu plano de recuperação judicial, o maior já visto no Brasil até hoje. Na assembleia, ficou definido que a companhia seria dividida em quatro partes: torres, ativos móveis, data center e fibra ótica.

Fonte: <https://www.infomoney.com.br/mercados/oi-oibr3-oibr4-super-tele-a-maior-recuperacao-judicial-do-brasil-entenda-o-que-aconteceu/>

InfoMoney

Venda Telefonia OI Móvel

Para quem não sabia, na verdade, a Oi Móvel foi vendida para a Vivo, Claro e TIM em 14/12/2020, pelo valor do R\$ 16,5 bilhões. Contudo, a venda dependia de aprovação pelos órgãos reguladores e antitruste competentes no negócio mais importante das telecomunicações no Brasil. A Anatel, após analisar as condições societária e regulatória, aprovou a venda em 31/01/2022. E o Cade, após analisar a condição concorrencial, concluiu o processo com a devida aprovação na data de hoje, 9/02/2022.

Fonte:<https://www.convergenciadigital.com.br/Opiniao/O-fim-da-Oi-Movel-59405.html?UserActiveTemplate=mobile>

**Convergência
DIGITAL**

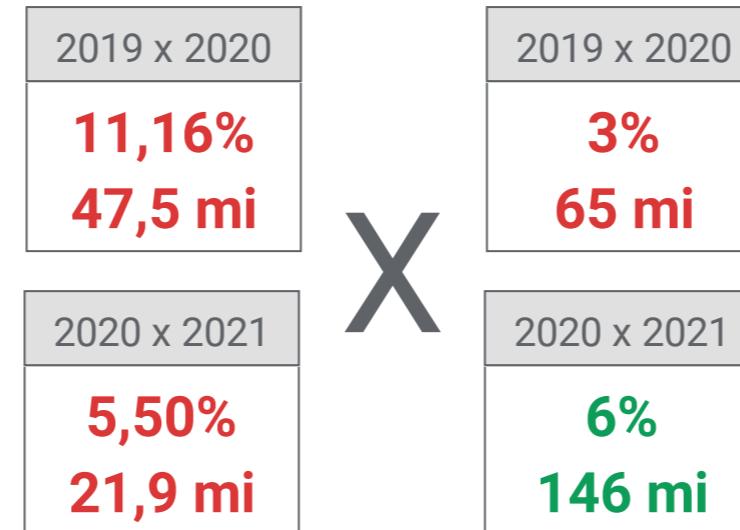
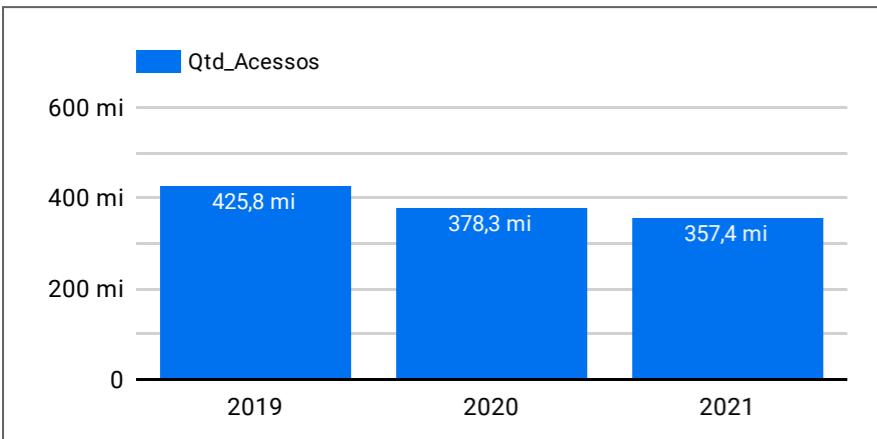


COMPARATIVO ENTRE TELEFONIA FIXA E MÓVEL

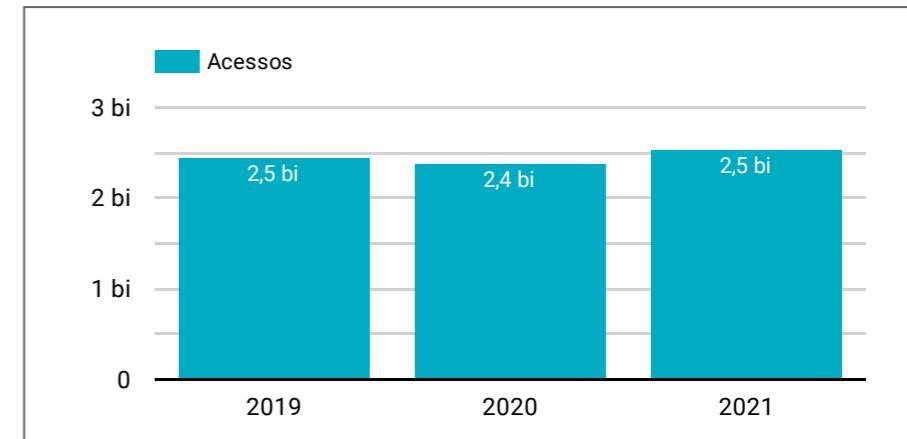




Telefonia Fixa



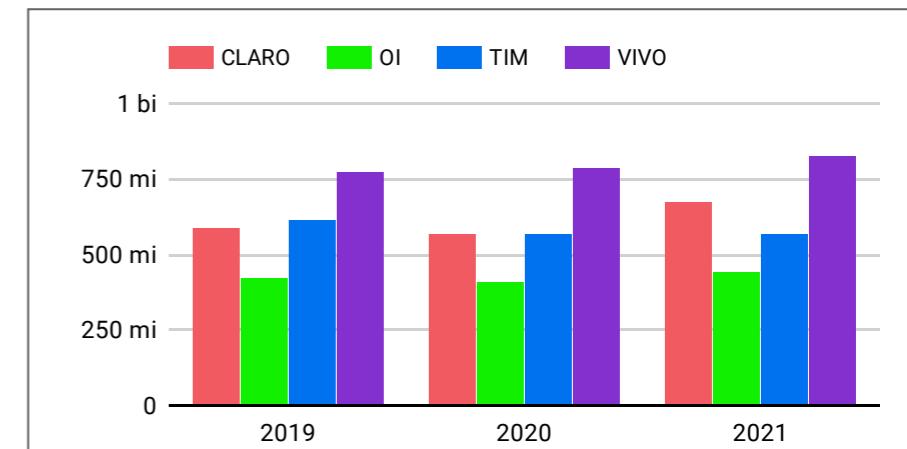
Telefonia Móvel



Comparação Fixo x Móvel | Ano

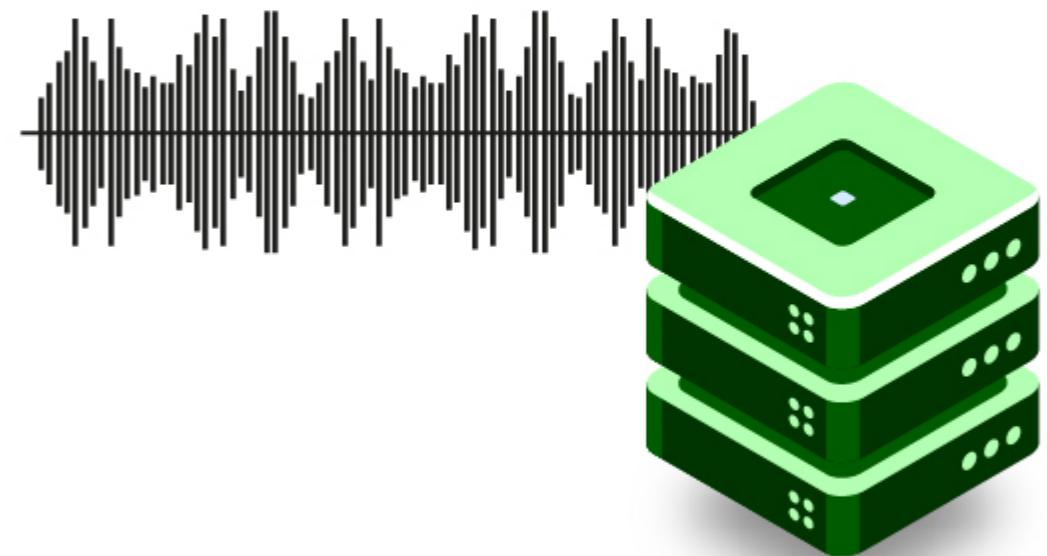
Ano	Fixo	Móvel	Crescimento Móvel	Redução Fixo
2019	425.796.967	2.457.808.389	577,23%	17,32%
2020	378.261.139	2.392.771.207	632,57%	15,81%
2021	357.439.263	2.538.754.020	710,26%	14,08%

Serviço Telefonia Móvel | Ano



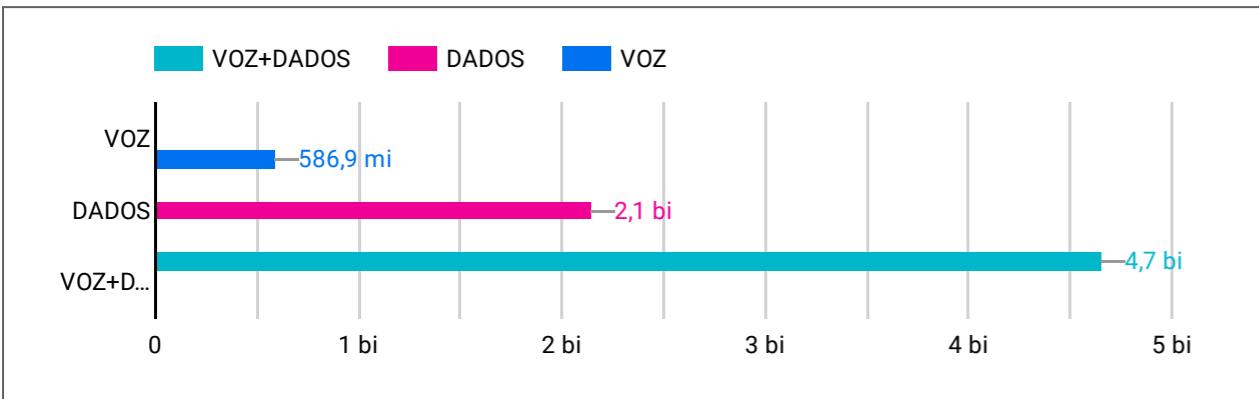


QUAL O DOMÍNIO DOS TIPOS DE SERVIÇO POR OPERADORA?

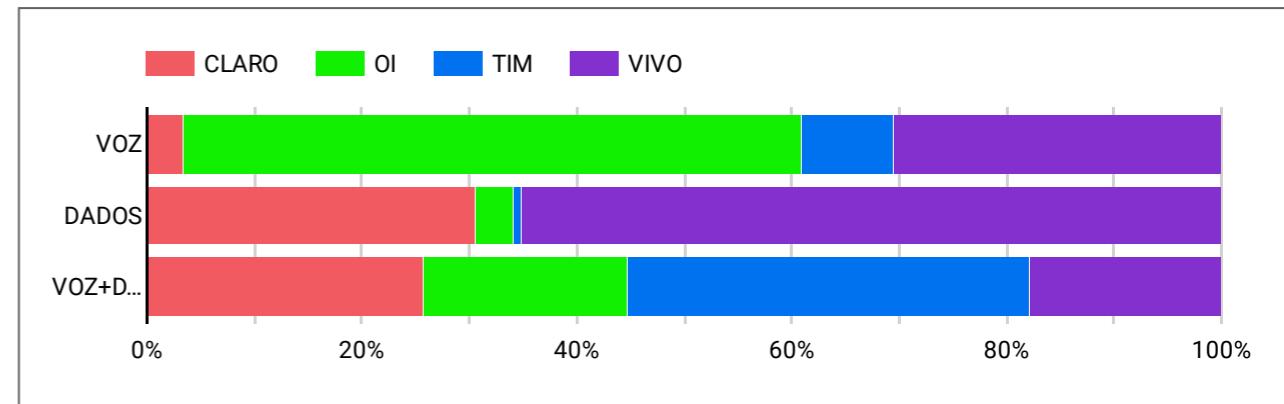




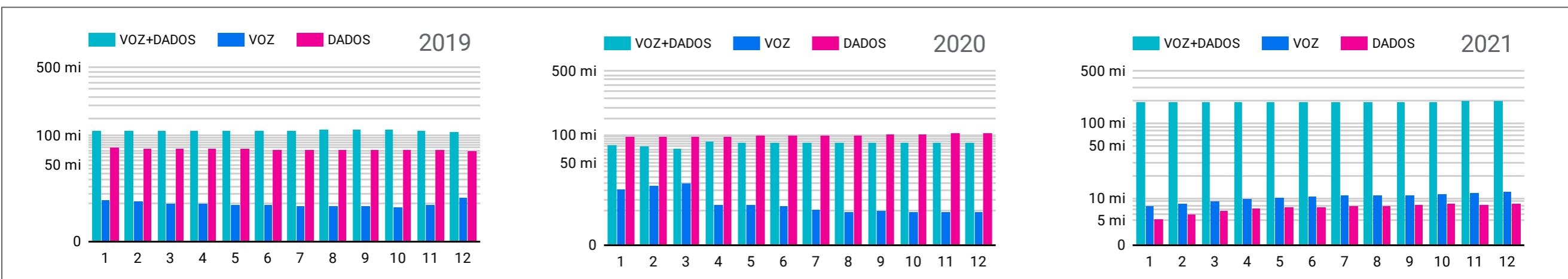
Tipos de Serviços de Telefonia Móvel



Porcentagem de Serviços por Operadora



Tipos de Serviço Telefonia Móvel | Classificação mensal entre 2019 e 2021



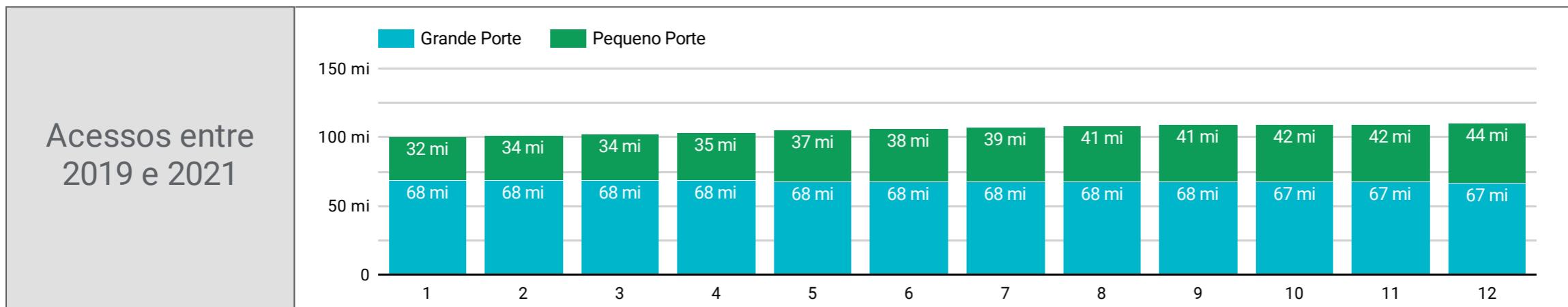
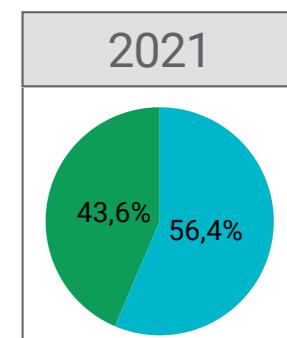
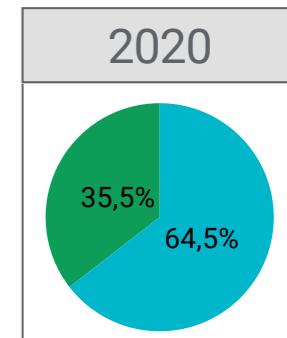
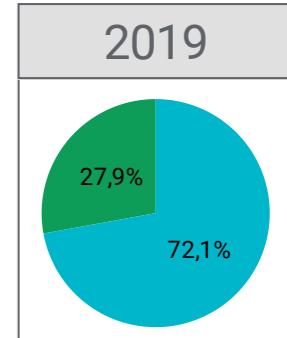
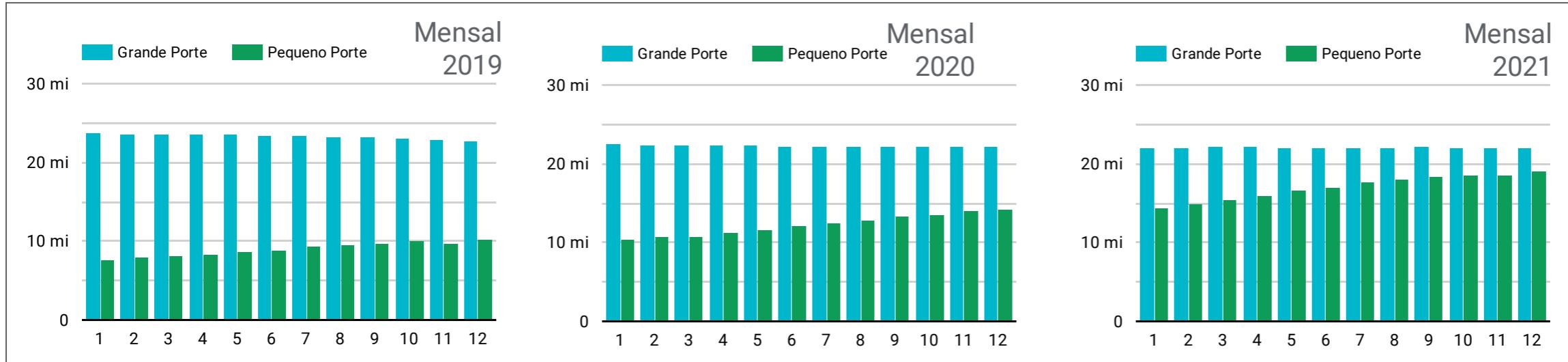


QUAL O CRESCIMENTO DE ACESSO NAS PEQUENAS PRESTADORAS DE SERVIÇO BANDA LARGA?





Crescimento de acessos nas Pequenas Prestadoras de Serviço Banda Larga





Procedimentos de ETL mais complexos - Criação de DF com base em diversos datasets

DataFrame Telefonia Móvel - PySpark

```
In [45]: # Criando Dataframe Telefonia Móvel e concatenando todos os Datasets do diretório do Datalake
```

```
bucket = storage.Client().get_bucket('projetofinal')
movel_names = [file.name for file in bucket.list_blobs(prefix='Originais/Movel/A')]

if movel_names == []:
    print('Não encontrado nenhum arquivo para ser inserido no DataFrame.')
else:
    print('Iniciando inserção de DataSets...')
    dfm = spark.read.csv('gs://projetofinal/' + movel_names[0], sep=';', header=True, inferSchema=True)
    dfm = dfm.drop('Código IBGE Município', 'CNPJ', 'Código Nacional (Chip)', 'Porte da Prestadora', 'Grupo Econômico', 'Código N
    print(f"DataSet '{movel_names[0][16:]}' incluído com sucesso.")
    if len(movel_names) > 1:
        movel = movel_names[1:]
        for i in movel:
            dfnovo = spark.read.csv('gs://projetofinal/' + i, sep=';', header=True, inferSchema=True)
            dfnovo = dfnovo.drop('Código IBGE Município', 'CNPJ', 'Código Nacional (Chip)', 'Porte da Prestadora', 'Grupo Econômico
            dfm = dfnovo.union(dfm)
            print(f"DataSet '{i[16:]}' incluído com sucesso.")

# Dropando colunas desnecessárias / indesejadas antes de concatenar para não onerar o processo.
```



Iniciando inserção de arquivos ao DataFrame...

DataFrame 'Acessos_Telefonia_Movel_201901-201906.csv' incluído com sucesso.



Procedimentos de ETL mais complexos - Conexão Mongo DB Atlas

```
In [ ]: # Criado conexão com MongoDB
client = pymongo.MongoClient('mongodb+srv://USUARIO:SENHA@projetofinal.356e0.mongodb.net/myFirstDatabase?retryWrites=true&w=major
client

In [ ]: antenas = client['Projeto_Final'].Antenas
antenas.insert_many(dfa.to_dict('records'))

# Criado Banco de Dados "Projeto_Final" e coleção "Antenas"
# Fazendo inserção do DataFrame na coleção Antenas

Out[25]: <pymongo.results.InsertManyResult at 0x7fe6bd8a7fd0>

In [ ]: antenas.count_documents({})

Out[26]: 96490
```

Robo 3T - 1.4

File View Options Window Help

New Connection (3) Welcome db.getCollection('Antenas').find...

Replica Set (3 nodes) New Connection projetofinal-shard-00-02.356e0.mongodb.net:27017 Projeto_Final

System

Projeto_Final

Collections (2)

Antenas

Telefonia_Fixa

Functions

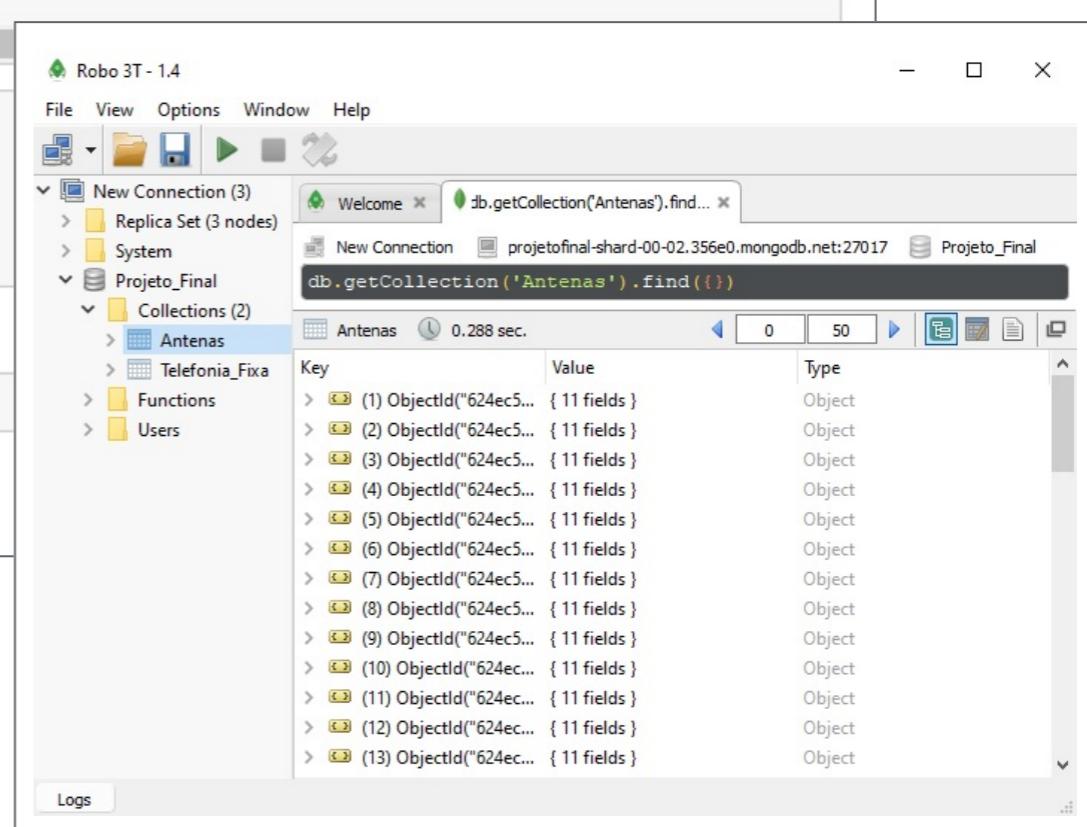
Users

db.getCollection('Antenas').find({})

Antenas 0.288 sec.

Key	Value	Type
> (1) ObjectId("624ec5...")	{ 11 fields }	Object
> (2) ObjectId("624ec5...")	{ 11 fields }	Object
> (3) ObjectId("624ec5...")	{ 11 fields }	Object
> (4) ObjectId("624ec5...")	{ 11 fields }	Object
> (5) ObjectId("624ec5...")	{ 11 fields }	Object
> (6) ObjectId("624ec5...")	{ 11 fields }	Object
> (7) ObjectId("624ec5...")	{ 11 fields }	Object
> (8) ObjectId("624ec5...")	{ 11 fields }	Object
> (9) ObjectId("624ec5...")	{ 11 fields }	Object
> (10) ObjectId("624ec5...")	{ 11 fields }	Object
> (11) ObjectId("624ec5...")	{ 11 fields }	Object
> (12) ObjectId("624ec5...")	{ 11 fields }	Object
> (13) ObjectId("624ec5...")	{ 11 fields }	Object

Logs





Procedimentos de ETL mais complexos - Envio de DataFrame original para MySQL via Pandas

Envio do DF ERBs - Antenas Original para MySQL via Pandas

```
from sqlalchemy import create_engine
```

```
In [33]: # CRIANDO CONEXÃO COM MYSQL
```

```
con = mysql.connector.connect(host='34.73.155.150', user='root', password='123456', db='telecom')
cur = con.cursor()

engine = create_engine("mysql+pymysql://root:123456@34.73.155.150/telecom")
```

```
In [34]: # Transformando DF em SQL e enviando para tabela ERBs_Antenas
```

```
dfa.to_sql('ERBs_Antenas', con = engine, if_exists='replace', index = False)
```

```
In [35]: # Conferindo se todas as linhas foram enviadas corretamente
```

```
cur.execute('SELECT count(*) FROM ERBs_Antenas')
cur.fetchall()
```

```
# FIM DO ENVIO PARA MySQL
```

```
Out[35]: [(96567,)]
```





Procedimentos de ETL mais complexos - Inserção manual de DataFrame no MySQL via Console

```
mysql> describe Acessos_Telefonia_Movel_2019_01;
```

Field	Type	Null	Key	Default	Extra
Ano	int(11)	YES		NULL	
Mes	int(11)	YES		NULL	
Grupo_Economico	varchar(255)	YES		NULL	
Empresa	varchar(255)	YES		NULL	
CNPJ	int(11)	YES		NULL	
Porte_da_Prestadora	varchar(255)	YES		NULL	
UF	varchar(255)	YES		NULL	
Municipio	varchar(255)	YES		NULL	
Codigo_IBGE_Municipio	int(11)	YES		NULL	
Codigo_Nacional	int(11)	YES		NULL	
Codigo_Nacional_Chip	int(11)	YES		NULL	
Modalidade_de_Cobranca	varchar(255)	YES		NULL	
Tecnologia	varchar(255)	YES		NULL	
Tecnologia_Geracao	varchar(255)	YES		NULL	
Tipo_de_Pessoa	varchar(255)	YES		NULL	
Tipo_de_Produto	varchar(255)	YES		NULL	
Acessos	int(11)	YES		NULL	

17 rows in set (0.00 sec)

Source

Choose a file to import from. Make sure you have read access first. [Learn more](#)

bucket-name/file-name *

projetofinal/Originais/Movel/Controle/Acessos_Telefonia_Movel_201 BROWSE

Browse for a Cloud Storage file or enter the path to one (bucket/folder/file)

File format

SQL

A plain text file with a sequence of SQL commands, like the output of mysqldump

CSV

If your Cloud Storage file is a CSV file, select CSV. The CSV file should be a plain text file with one line per row and comma-separated fields.

Destination

Choose the database and table in your instance for this file to import into. [Learn more](#)

Database *

telecom

Table *

Acessos_Telefonia_Movel_2019_01

Enter the name of an existing table in the database to house your CSV file



Análise realizada em Spark SQL

In [17]: `dfbl.createOrReplaceTempView("blsql")`In [23]: *# Comparativo de crescimento entre Prestadoras de Grande e Pequeno porte nos últimos 3 anos.*

```
spark.sql('WITH bl2019 as (SELECT Porte_da_Prestadora, Ano, SUM(Acessos) AS Soma_de_Acessos, round((sum(Acessos) / 388393110)*100) AS Porcentagem from blsql WHERE Ano = 2019 \
GROUP BY Porte_da_Prestadora, Ano), \
\
bl2020 as (SELECT Porte_da_Prestadora, Ano, SUM(Acessos) AS Soma_de_Acessos, round((sum(Acessos) / 414932279)*100) AS Porcentagem from blsql WHERE Ano = 2020 \
GROUP BY Porte_da_Prestadora, Ano), \
\
bl2021 as (SELECT Porte_da_Prestadora, Ano, SUM(Acessos) AS Soma_de_Acessos, round((sum(Acessos) / 469976308)*100) AS Porcentagem from blsql WHERE Ano = 2021 \
GROUP BY Porte_da_Prestadora, Ano) \
\
SELECT bl2019.Porte_da_Prestadora, bl2019.Ano, bl2019.Soma_de_Acessos, bl2019.Porcentagem, \
bl2020.Ano, bl2020.Soma_de_Acessos, bl2020.Porcentagem, bl2021.Ano, bl2021.Soma_de_Acessos, bl2021.Porcentagem \
FROM bl2019 LEFT JOIN bl2020 ON bl2019.Porte_da_Prestadora = bl2020.Porte_da_Prestadora LEFT JOIN bl2021 ON \
bl2020.Porte_da_Prestadora = bl2021.Porte_da_Prestadora').show()
```

[Stage 42::=> (4 + 1) / 6][Stage 45:> (0 + 0) / 1][Stage 48:> (0 + 0) / 1]1

Porte_da_Prestadora	Ano	Soma_de_Acessos	Porcentagem	Ano	Soma_de_Acessos	Porcentagem	Ano	Soma_de_Acessos	Porcentagem
Grande Porte	2019	280170257	72.0	2020	267709124	65.0	2021	265117279	56.0
Pequeno Porte	2019	108222853	28.0	2020	147223155	35.0	2021	204859029	44.0



BigQuery - Análises

EDITOR X *QTD_AC... X

RUN SAVE SHARE SCHEDULE

```
1 -- Quantidade de Acessos de Telefonia Móvel
2 -- por Região e Modalidade nos últimos 3 anos
3 SELECT
4 DISTINCT R.Regiao,
5 Modalidade_de_Cobran__a,
6 SUM(M.Acessos) AS Qtd_Acessos
7 FROM
8 telecom.regioes R
9 INNER JOIN
10 telecom.movel M
11 ON
12 R.UF = M.UF
13 GROUP BY
14 Regiao,
15 Modalidade_de_Cobran__a
16 ORDER BY
17 Regiao
```

Query results SAVE RESULTS RESULTS JSON EXECUTION D

Row	Regiao	Modalidade_de_Cobran__a	Qtd_Acessos
1	Centro-Oeste	Pré-pago	341194537
2	Centro-Oeste	Pós-pago	276098964
3	Nordeste	Pré-pago	1289369828
4	Nordeste	Pós-pago	509149503
5	Norte	Pré-pago	405005900
6	Norte	Pós-pago	160771132
7	Sudeste	Pré-pago	1621746772
8	Sudeste	Pós-pago	1722508621
9	Sul	Pré-pago	558366584
10	Sul	Pós-pago	505121775

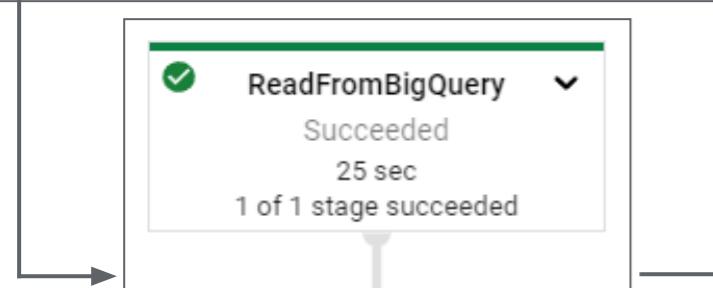


Google
Big Query



Dataflow - Exportando tabela resultado para um Datalake em formato Parquet

Name	Type	End time	Elapsed time	Start time	Status	SDK version	ID	Region
bigqueryparquet2	Batch	Apr 9, 2022, 11:55:40 AM	5 min 19 sec	Apr 9, 2022, 11:50:21 AM	Succeeded		2022-04-09_07_50_21- 18040831372762234582	us-east1



Job name	bigqueryparquet2
Job ID	2022-04-09_07_50_21- 18040831372762234582
Job type	Batch
Job status	✓ Succeeded
SDK version	Apache Beam SDK for Java 2.37.0
Job region	us-east1
Worker location	us-east1-d
Current workers	0
Latest worker status	Worker pool stopped.
Start time	April 9, 2022 at 11:50:21 AM GMT-3
Elapsed time	5 min 19 sec
Encryption type	Google-managed key

BEAM
SUMMIT

Armazenamento Cloud Storage

Buckets	→	projetofinal	→	Dataflow_bigquery	
UPLOAD FILES	UPLOAD FOLDER	CREATE FOLDER			
Filter by name prefix only ▼ Filter Filter objects and					
<input type="checkbox"/>	Name		Size		
<input type="checkbox"/>	output-00000-of-00004.parquet		10.3 MB		
<input type="checkbox"/>	output-00001-of-00004.parquet		10.4 MB		
<input type="checkbox"/>	output-00002-of-00004.parquet		10.3 MB		
<input type="checkbox"/>	output-00003-of-00004.parquet		10.4 MB		



Integrantes

Artur Venturoli

<https://br.linkedin.com/in/arturventuroli>

Maurício da Silva Barros

<https://www.linkedin.com/in/mauriciodasilvabarros/>

Renan Marques

www.linkedin.com/in/renan-marques-rodrigues

Renato Romão

www.linkedin.com/in/renato-cm-romao

Referências

Datasets

<https://conexis.org.br/numeros/mapa-de-antenas/>

<https://informacoes.anatel.gov.br/paineis/acessos/telefonia-fixa>

<https://informacoes.anatel.gov.br/paineis/acessos/telefonia-movel>

<https://informacoes.anatel.gov.br/paineis/acessos/banda-larga-fixa>

<https://www.todamateria.com.br/capitais-do-brasil/>

Notícias

<https://www.convergenciadigital.com.br/Opiniao/O-fim-da-Oi-Movel-59405.html?UserActiveTemplate=mobile>

<https://www.infomoney.com.br/mercados/oi-oibr3-oibr4-super-tele-a-maior-recuperacao-judicial-do-brasil-entenda-o-que-aconteceu/>

Imagens e vetores

<https://br.freepik.com>