

Projeto Final

Turma : Cientista de Dados - T04
Aluna : Luciana Lanzoni Menges

Projeto Final

→ Meu primeiro projeto de Ciência de Dados:

◆ Objetivo : Analisar dados brutos do Comex Stat utilizando serviço em nuvem, AWS.



Comex Stat é o portal para acesso gratuito às estatísticas de comércio exterior do Brasil.

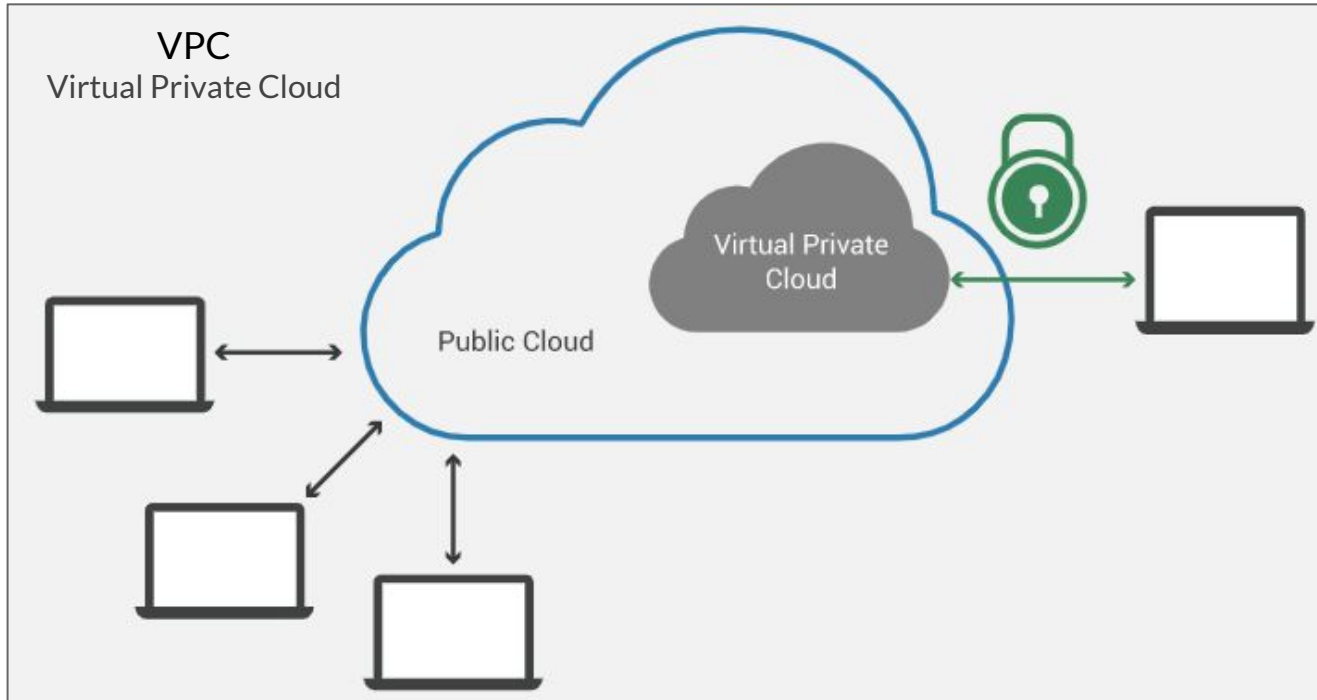


Amazon Web Services (ou AWS) existe desde 2006 e oferece serviços de infraestrutura de TI para empresas por meio de serviços web – a chamada computação na nuvem.

Etapas do Projeto

1. Criar conta na Amazon Web Services - AWS
2. Criar Database na Amazon RDS – Relational Database Service
3. Conectar-se a uma instância de banco de dados PostgreSQL
4. Quais arquivos serão analisados?
5. PgAdmin 4 - “Importar arquivo”
6. Análise dos dados das tabelas
7. Visualização com Power BI

Visão Geral do ambiente virtual



1. Criar conta na AWS

As contas da AWS incluem
12 meses de acesso ao nível
gratuito

Incluindo o uso do Amazon EC2, do Amazon S3 e do Amazon
DynamoDB

Acesse aws.amazon.com/free para ver os termos completos da
oferta

Criar uma conta da AWS

Endereço de e-mail

Senha

Confirmar senha

Nome da conta da AWS ⓘ

Continuar

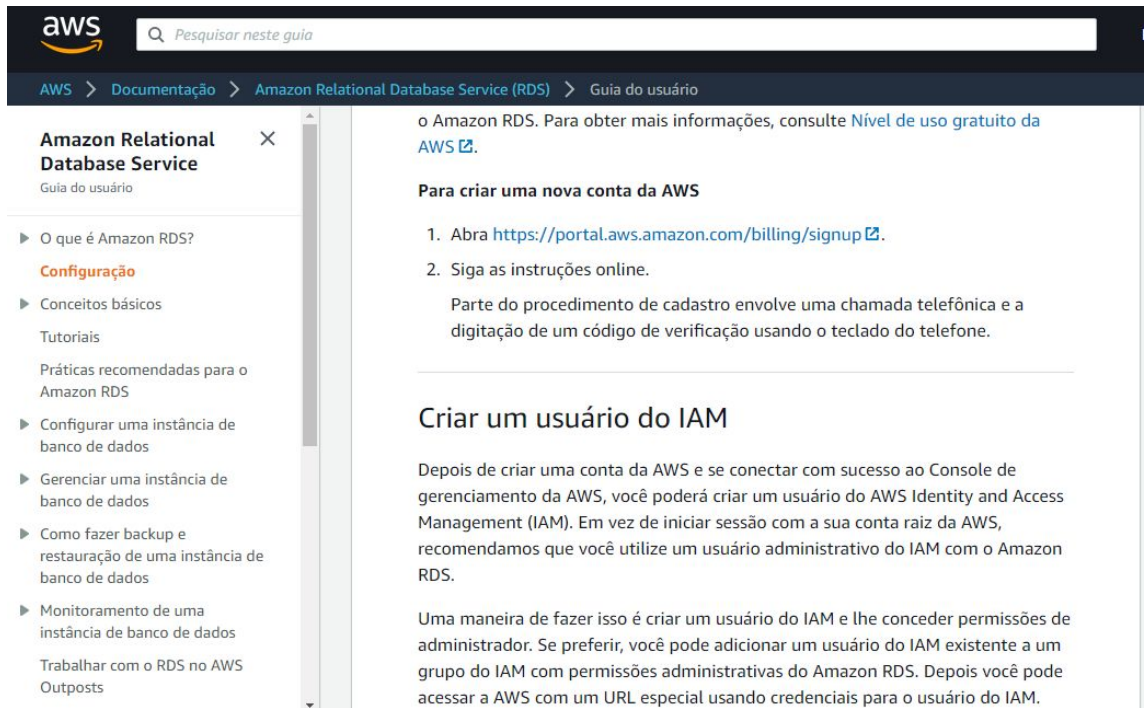
[Fazer login com uma conta existente da AWS](#)

© 2020 Amazon Web Services, Inc. ou suas afiliadas.
Todos os direitos reservados.

[Política de privacidade](#) | [Termos de uso](#)

- Criar a conta na AWS é fácil!
- Preencha os dados iniciais nesta página e depois
- Siga as instruções online.
- Será necessário indicar o número de um cartão de crédito internacional.
- Para autenticação dos seus dados será debitado o valor de US\$1,00 no cartão,
- Além do envio de um código de verificação via SMS.
- Etapa Concluída, criada conta RAÍZ!!
- **Ao se cadastrar, você tem acesso automático a todos os serviços da AWS.**

1. Criar conta na AWS



The screenshot shows the AWS documentation interface. At the top, there's a search bar with the text 'Pesquisar neste guia'. Below it, a breadcrumb trail reads: 'AWS > Documentação > Amazon Relational Database Service (RDS) > Guia do usuário'. On the left, a sidebar titled 'Amazon Relational Database Service' lists various topics, with 'Configuração' (Configuration) highlighted. The main content area is titled 'Guia do usuário' and contains the following text:

o Amazon RDS. Para obter mais informações, consulte [Nível de uso gratuito da AWS](#).

Para criar uma nova conta da AWS

1. Abra <https://portal.aws.amazon.com/billing/signup>.
2. Siga as instruções online.

Parte do procedimento de cadastro envolve uma chamada telefônica e a digitação de um código de verificação usando o teclado do telefone.

Criar um usuário do IAM

Depois de criar uma conta da AWS e se conectar com sucesso ao Console de gerenciamento da AWS, você poderá criar um usuário do AWS Identity and Access Management (IAM). Em vez de iniciar sessão com a sua conta raiz da AWS, recomendamos que você utilize um usuário administrativo do IAM com o Amazon RDS.

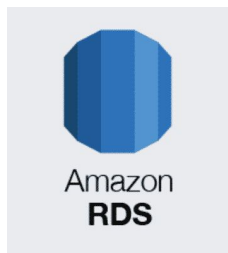
Uma maneira de fazer isso é criar um usuário do IAM e lhe conceder permissões de administrador. Se preferir, você pode adicionar um usuário do IAM existente a um grupo do IAM com permissões administrativas do Amazon RDS. Depois você pode acessar a AWS com um URL especial usando credenciais para o usuário do IAM.

- A AWS tem uma documentação bem ampla, explica o passo a passo de cada serviço apresentado.
- Além de sugestões para boas práticas para utilização
- Exemplo: a criação do Usuário Administrador, no IAM - Identity and Access Management
- Depois você adiciona o usuário a um grupo de administradores
- Podendo criar grupos de usuários, com permissões específicas

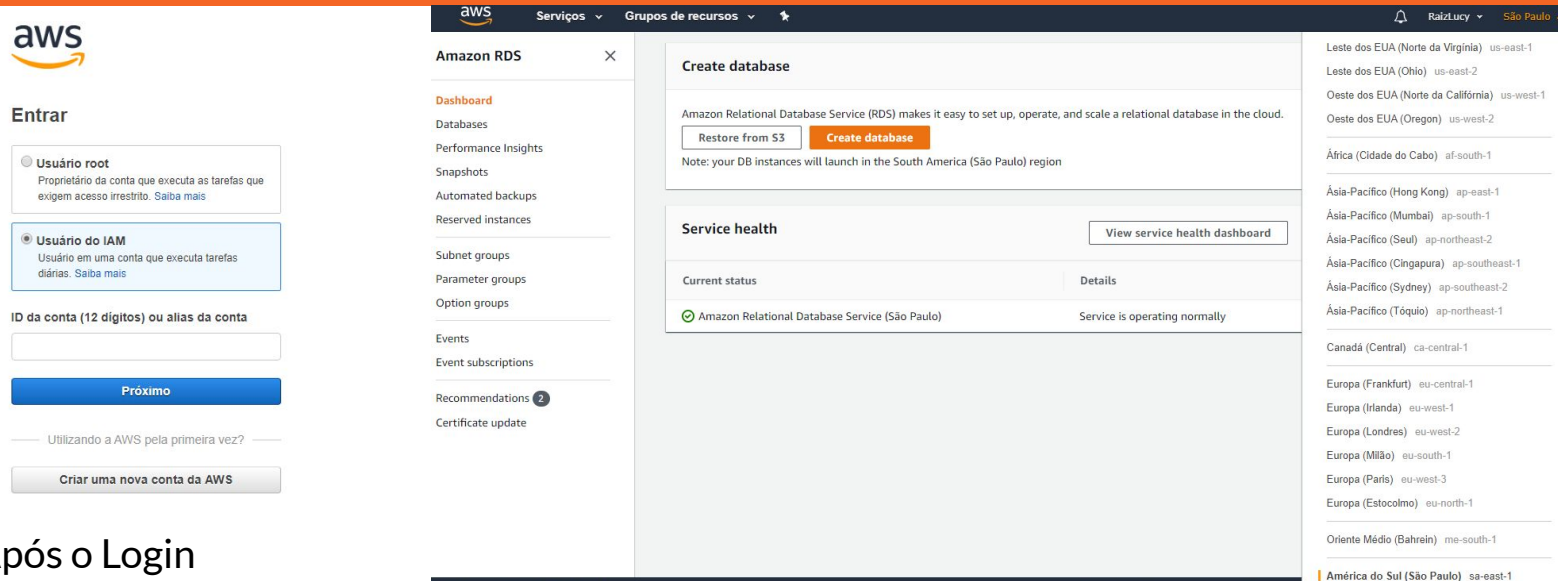
2. Criar Database na Amazon RDS

- Amazon RDS - **Relational Database Service**, permite a fácil configuração, operação e escalabilidade de bancos de dados relacionais em cloud computing.
- O serviço está disponível em variados modelos de instâncias de banco de dados com desempenho e otimização para memória.
- Oferece seis diferentes tipos banco de dados (versões ou sabores):

- ◆ **PostgreSQL;**
- ◆ Amazon Aurora;
- ◆ MySQL;
- ◆ Microsoft SQL Server;
- ◆ Oracle;
- ◆ MariaD.



2. Criar Database na Amazon RDS



The screenshot displays the AWS Management Console interface for creating a new Amazon RDS database instance. The left-hand navigation pane shows the 'Amazon RDS' section expanded, listing various management tools. The main panel is titled 'Create database' and provides introductory text about the service, along with buttons to 'Restore from S3' or 'Create database'. A note specifies that instances will be launched in the South America (São Paulo) region. Below this, a 'Service health' section indicates that the Amazon RDS service is currently operating normally. On the right side of the console, a scrollable list of AWS regions is visible, with 'América do Sul (São Paulo)' highlighted at the bottom.

- Após o Login
- Primeiro Passo é escolher a região da AWS na qual você deseja criar a instância de banco de dados. (EC2 disponibiliza 77 zonas em 24 regiões)
- Depois ir em Criar banco de dados

2. Amazon RDS – Relational Database Service

Create database

Choose a database creation method [Info](#)

☐ Standard Create

You set all of the configuration options, including ones for availability, security, backups, and maintenance.

☒ Easy Create

Use recommended best-practice configurations. Some configuration options can be changed after the database is created.

Configuration

Engine type [Info](#)

☐ Amazon Aurora



☐ MySQL



☐ MariaDB



☒ PostgreSQL



☐ Oracle

ORACLE®

☐ Microsoft SQL Server



DB instance size

☐ Production

db.r4.large
2 vCPUs
15.25 GiB RAM
500 GiB
1.346 USD/hour

☐ Dev/Test

db.m4.large
2 vCPUs
8 GiB RAM
100 GiB
0.269 USD/hour

☒ Free tier

db.t2.micro
1 vCPUs
1 GiB RAM
20 GiB
0.044 USD/hour

2. Amazon RDS – Relational Database Service

DB instance identifier

Type a name for your DB instance. The name must be unique across all DB instances owned by your AWS account in the current AWS Region.

The DB instance identifier is case-insensitive, but is stored as all lowercase (as in "mydbinstance"). Constraints: 1 to 60 alphanumeric characters or hyphens (1 to 15 for SQL Server). First character must be a letter. Can't contain two consecutive hyphens. Can't end with a hyphen.

Master username [Info](#)

Type a login ID for the master user of your DB instance.


1 to 16 alphanumeric characters. First character must be a letter

☒ Auto generate a password

Amazon RDS can generate a password for you, or you can specify your own password

► View default settings for Easy create

Easy create sets the following configurations to their default values, some of which can be changed later. If you want to change any of these settings now, use [Standard Create](#).

 You are responsible for ensuring that you have all of the necessary rights for any third-party products or services that you use with AWS services.

Create database

2. Amazon RDS – Relational Database Service

★ Banco de dados criado

RDS > Databases > database-1

database-1 Modify

Summary

DB identifier database-1	CPU <div><div></div></div> 2.83%	Info ✓ Available	Class db.t2.micro
Role Instance	Current activity <div><div></div></div> 0 Connections	Engine PostgreSQL	Region & AZ sa-east-1a

América
do Sul -
São Paulo

2. Amazon RDS – Relational Database Service

Connectivity & security

Endpoint & port

Endpoint

database-1[REDACTED].sa-east-1.rds.amazonaws.com

Port

5432

Networking

Availability zone

sa-east-1a

VPC

vpc-22baa745

Subnet group

default-vpc-22baa745

Subnets

subnet-0c69e76a

subnet-5af66201

subnet-f967ebb0

Security

VPC security groups

rds-launch-wizard (sg-073b8eecd3310e69c)
(active)

Public accessibility

No

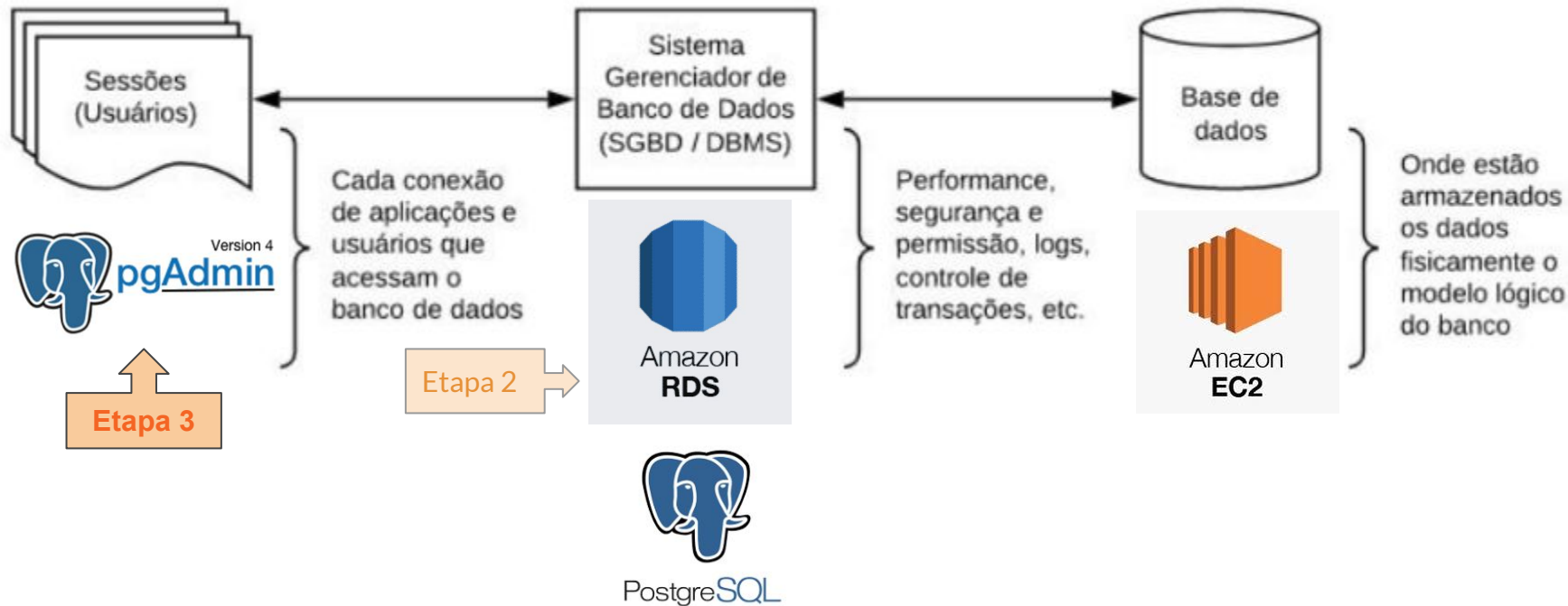
Certificate authority

rds-ca-2019

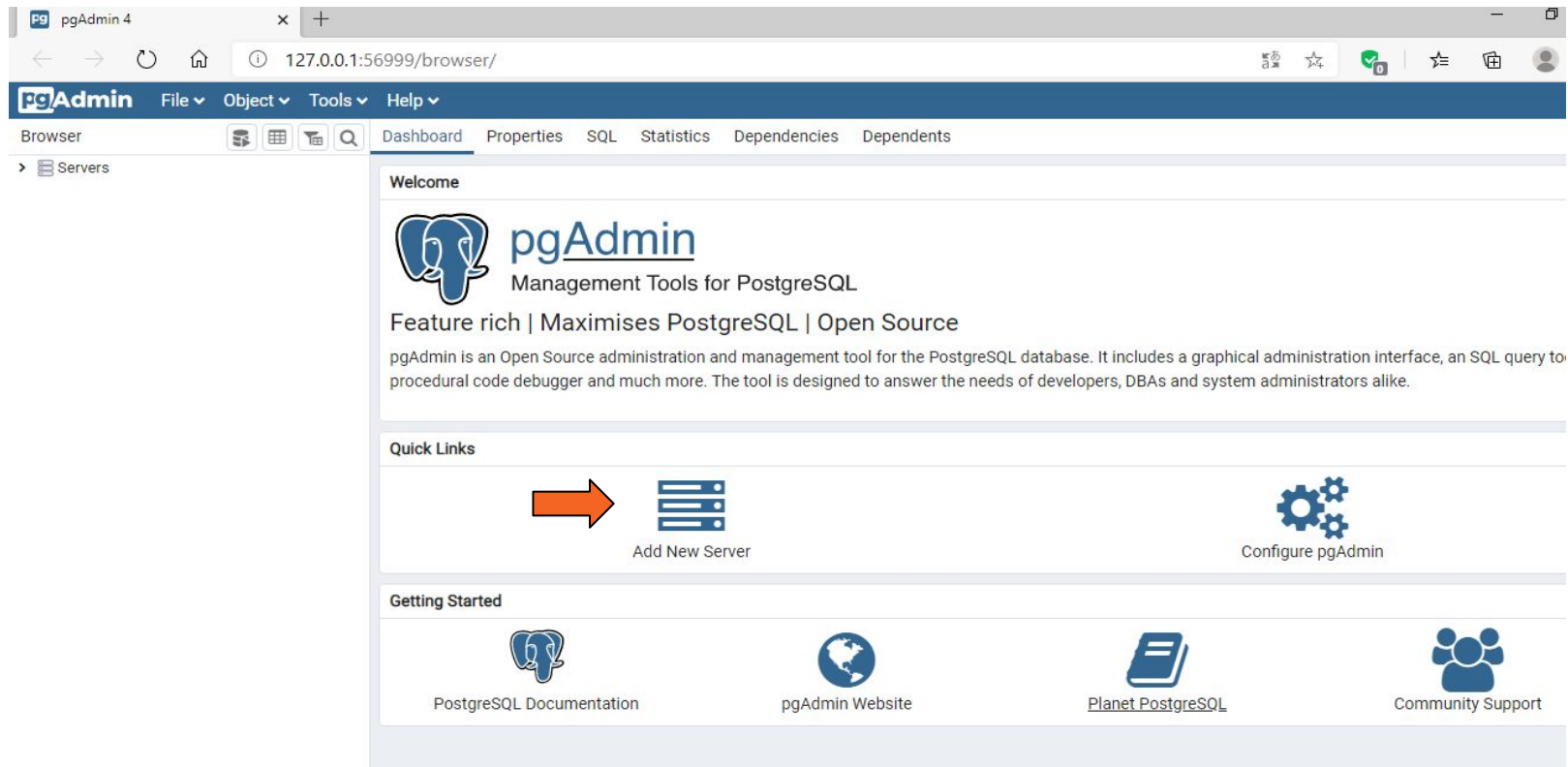
Certificate authority date

Aug 22nd, 2024

3. Conectar-se a instância de banco dados PostgreSQL



3. Conectar-se a instância de banco dados PostgreSQL



3. Conectar-se a instância de banco dados PostgreSQL

Create - Server

General Connection SSL SSH Tunnel Advanced

Name **rds_lucy**

Server group Servers

Background ☒

Foreground ☒

Connect now? ☒

Comments

Name must be specified.

Cancel Reset Save

Create - Server

General Connection SSL SSH Tunnel Advanced

Host **Endpoint RDS**

Port 5432

Maintenance database postgres

Username postgres

Password **Senha criada na RDS**

Save password? ☐

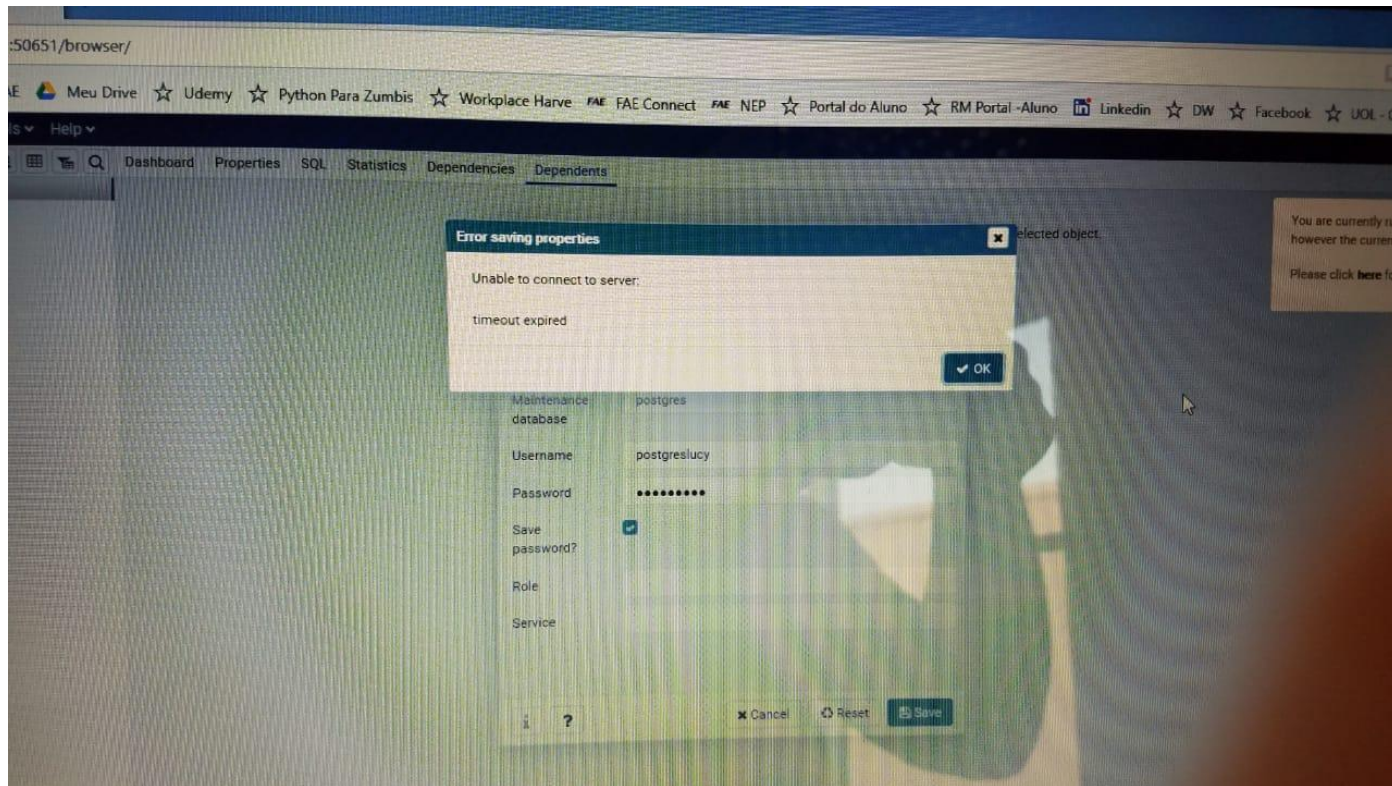
Role

Service

Name must be specified.

Cancel Reset **Save**

3. Conectar-se a instância de banco dados PostgreSQL



Erro:
Incapaz de se conectar
com o servidor

Tempo limite expirado

2. Amazon RDS – Relational Database Service

Connectivity & security

Endpoint & port

Endpoint

database-1.([REDACTED]).sa-east-1.rds.amazonaws.com

Port

5432

Networking

Availability zone

sa-east-1a

VPC

[vpc-22baa745](#)

Subnet group

default-vpc-22baa745

Subnets

[subnet-0c69e76a](#)

[subnet-5af66201](#)

[subnet-f967ebb0](#)

Security

VPC security groups

[rds-launch-wizard \(sg-073b8eecd3310e69c\)](#)
(active)

Public accessibility

No

Certificate authority

rds-ca-2019

Certificate authority date

Aug 22nd, 2024

2. Amazon RDS – Relational Database Service

Grupos de segurança (1/2) [Informações](#)

[<](#) **1** [>](#)

<input type="checkbox"/>	Name	ID do grupo de segur...	Nome do grupo de ...	ID da VPC	Descrição
<input type="checkbox"/>	–	sg-073b8eecd3310e69c	rds-launch-wizard	vpc-22baa745 ↗	Created from the P
<input checked="" type="checkbox"/>	–	sg-4a6ac63b	default	vpc-22baa745 ↗	default VPC securi

sg-4a6ac63b - default

[Detalhes](#) **[Regras de entrada](#)** [Regras de saída](#) [Tags](#)

Regras de entrada [Editar regras de entrada](#)

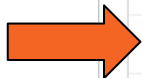
Tipo	Protocolo	Intervalo de portas	Origem	Descrição - opcional
PostgreSQL	TCP	5432	0.0.0.0/0	–
PostgreSQL	TCP	5432	:::/0	–
Todo o tráfego	Tudo	Tudo	sg-4a6ac63b (default)	–

2. Amazon RDS – Relational Database Service

▼ View default settings for **Easy create**

Easy create sets the following configurations to their default values, some of which can be changed later. If you want to change any of these settings now, use [Standard Create](#).

Configuration ▼	Value	Editable after database is created ▲
Encryption	Enabled	No
VPC	Default VPC (vpc-22baa745)	No
Option Group	default:postgres-11	No
Subnet Group	default-vpc-22baa745	Yes
Automatic Backups	Enabled	Yes
VPC Security Group	sg-4a6ac63b	Yes
Publicly Accessible	No	Yes
Database Port	5432	Yes
DB Instance Identifier	database-1	Yes



3. Conectar-se a instância de banco dados PostgreSQL



Create - Server [X]

General | Connection | SSL | SSH Tunnel | Advanced

Name:

Server group:

Background: ☐

Foreground: ☐

Connect now?: ☒

Comments:

Name must be specified. [X]

[i] [?] [Cancel] [Reset] [Save]

Create - Server [X]

General | **Connection** | SSL | SSH Tunnel | Advanced

Host name/address: [!]

Port:

Maintenance database:

Username:

Password:

Save password?: ☐

Role:

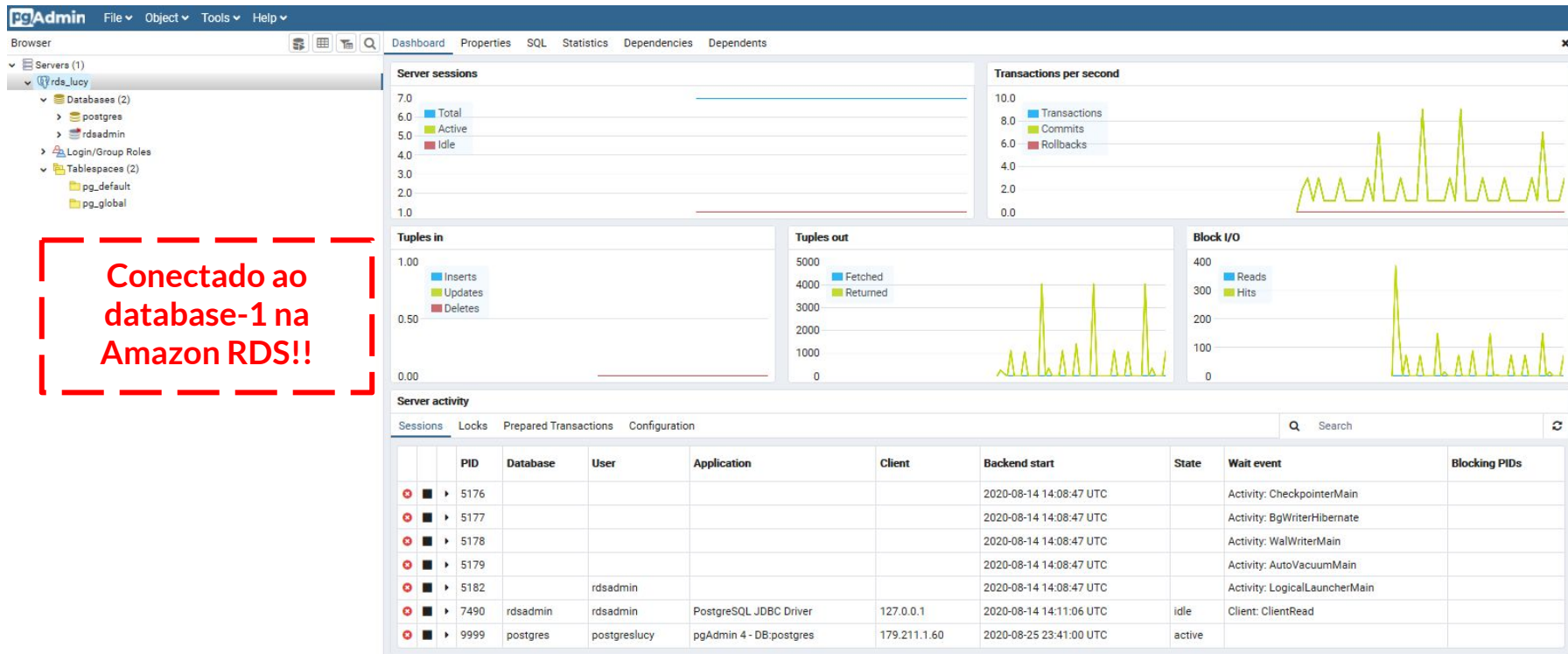
Service:

Name must be specified. [X]

[i] [?] [Cancel] [Reset] [Save]

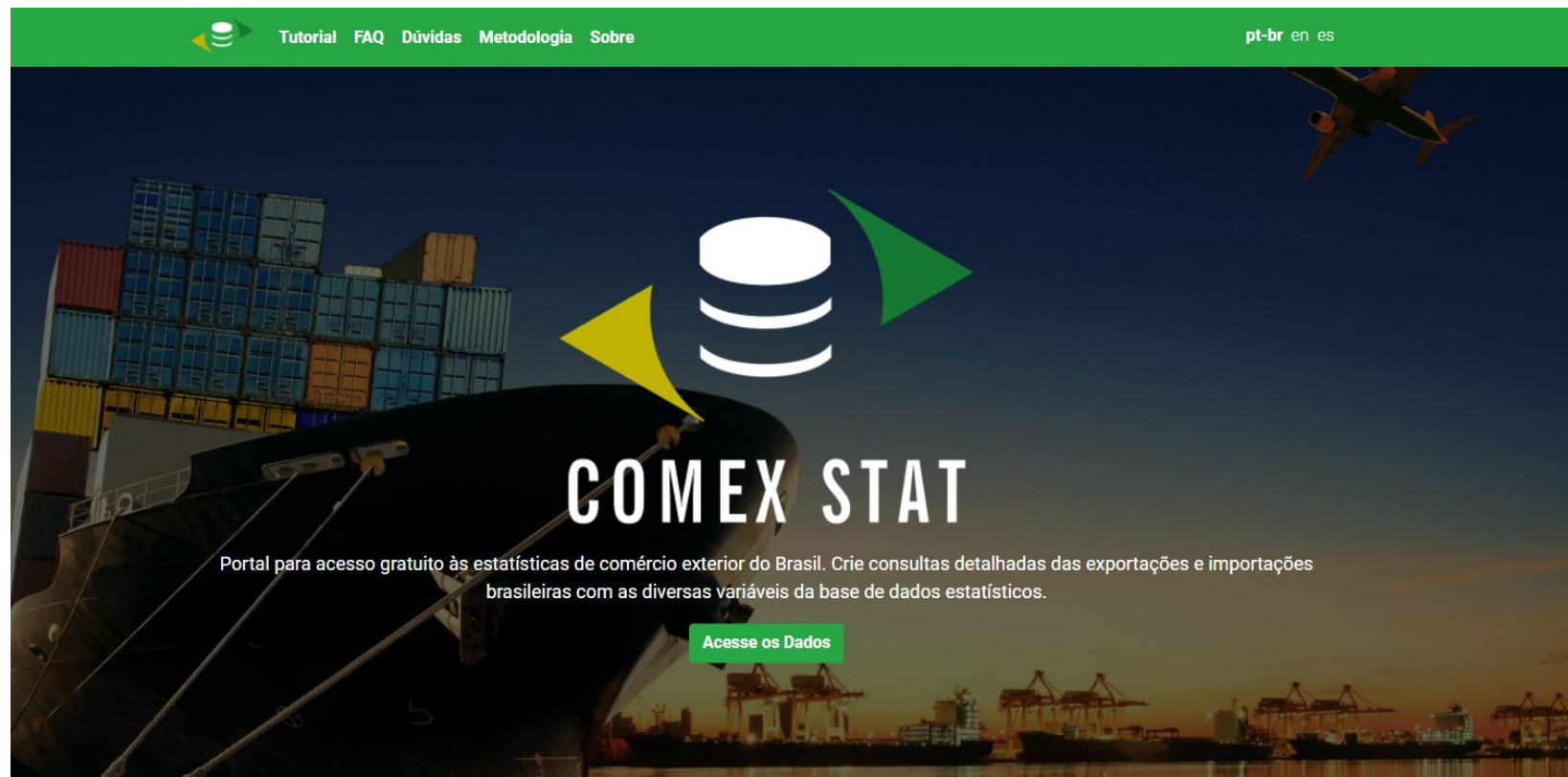


3. Conectar-se a instância de banco dados PostgreSQL



Conectado ao
database-1 na
Amazon RDS!!

4. Quais arquivos serão analisados ?



Tutorial FAQ Dúvidas Metodologia Sobre

pt-br en es

COMEX STAT

Portal para acesso gratuito às estatísticas de comércio exterior do Brasil. Crie consultas detalhadas das exportações e importações brasileiras com as diversas variáveis da base de dados estatísticos.

[Acesse os Dados](#)

4. Quais arquivos serão estudados ?



Exportações e Importações Geral

Consultas com dados mensais a partir de 1997 ao ano atual. Filtros e detalhamentos de países, blocos, UF do produto, NCM e sistema harmonizado (SH6, SH4, Capítulo e Seção), classificação por grandes categorias econômicas e Classificação Uniforme para o Comércio Internacional.

[Acessar ↗](#)

Dados disponíveis até 07 /2020



Exportações e Importações Municípios

Consultas com dados mensais de 1997 ao ano atual. Filtros e detalhamentos de países, blocos, sistema harmonizado (apenas SH4, Capítulo e Seção), município de domicílio fiscal do exportador/importador e UF do município.

[Acessar ↗](#)

Dados disponíveis até 07 /2020



Comex Vis

Acompanhe representações gráficas e interativas de dados do comércio exterior brasileiro, destacando as principais informações de forma simples e intuitiva.

[Acessar ↗](#)



Índice de Preço & Quantum

As séries dos Índices de Preços e Quantum decompõem os valores das exportações e importações em seus aspectos de preços e volumes, viabilizando análises mais precisas dos fluxos de comércio.

[Acessar ↗](#)



Base de Dados

A base de dados completa do sistema está disponível em formato de dados brutos (csv) na página da Secretaria de Comércio Exterior, com periodicidade mensal de publicação. Clique no link para redirecionar.

[Acessar ↗](#)



Dados Históricos

Consultas com dados históricos mensais entre 1989 a 1996. Filtros e detalhamentos de países, UF e NCM

[Acessar ↗](#)



Tabelas Auxiliares

Consultas às tabelas auxiliares com descrições de produtos, classificações e agregações de NCM, Sistema Harmonizado, Classificação por Grandes Categorias Econômicas, Classificação Uniforme para o Comércio Internacional, países, blocos, UF, UR e via.

4. Quais arquivos serão estudados ?

Ambiente em

Migração

MINISTÉRIO DA ECONOMIA

Buscar no portal

Ouvidoria | Área de imprensa | Acesso a Sistemas

PÁGINA INICIAL > COMÉRCIO EXTERIOR > ESTATÍSTICAS DE COMÉRCIO EXTERIOR > BASE DE DADOS DO COMÉRCIO EXTERIOR BRASILEIRO

Balança Comercial

Comex Vis

Siscomex

ASSUNTOS

Camex

4 Comércio Exterior

4 Estatísticas de Comércio Exterior

Balança comercial brasileira: Semanal

Balança comercial

Base de dados do Comex Stat, tabelas de correlações de códigos e calendário brasileiro com indicação dos dias úteis

Tweetar

Curtir 2,9 mil

Os arquivos a seguir apresentam todos os dados utilizados para construção da balança comercial brasileira, bem como a fonte de dados do sistema **Comex Stat**, detalhados por **NCM** ou por **Municípios do exportador/importador (e SH4)**. Os dados estão disponíveis em CSV com os nomes das colunas na primeira linha. Para aproveitamento completo dos arquivos de dados será necessário utilizar as **Tabelas de Correlações de Códigos e Classificações**. *Obs: Devido ao tamanho dos arquivos, recomenda-se a utilização de ferramentas apropriadas para manejo de massas de dados (Bancos de Dados, Access, softwares estatísticos, etc). Softwares de planilhas podem não abrir os arquivos adequadamente.*

4. Quais arquivos serão estudados ?

Balança comercial brasileira: Acumulado do ano
Séries Históricas
Comex Vis: Visualizações de Comércio Exterior
Base de dados do Comércio Exterior Brasileiro
Empresas brasileiras exportadoras e importadoras
Outras estatísticas de Comércio Exterior
Metodologia de Produção de Estatísticas de Comércio Exterior
Despachos de Operações de Comércio Exterior
Seminários de Comércio Exterior
Negociações Internacionais
Tarifa Externa Comum - TEC (NCM)

Últimos dados disponíveis: janeiro - julho de 2020

1. Base de dados detalhada por NCM: Arquivos CSV com separador ponto e vírgula (;) detalhado por ano, mês, código NCM, código da unidade estatística, código de país de destino/origem do produto, código da UF de origem/destino do produto, código da via de transporte, código da URF de embarque/desembarque, quantidade estatística, quilograma líquido, valor dólar FOB (US\$).

Layout: CO_ANO;CO_MES;CO_NCM;CO_UNID;CO_PAIS;SG_UF_NCM(sigla UF origem/destino da NCM);CO_VIA;CO_URF(Unidade da Receita Federal);QT_STAT;KG_LIQUIDO;VL_FOB

OS ARQUIVOS NÃO DEVEM SER UTILIZADOS EM SOFTWARES DE PLANILHA, CORRENDO O RISCO DE PERDA DE LINHAS E INFORMAÇÕES.

▪ Exportação:

1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
2017	2018	2019	2020						
Arquivo único 1997-2020					Totais para validação				

▪ Importação:

1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
2017	2018	2019	2020						
Arquivo único 1997-2020					Totais para validação				

4. Quais arquivos serão estudados ?

ACESSO À INFORMAÇÃO

Institucional

Agendas

Legislação

Licitações

Contratos

Auditorias

Convênios

Atos Internacionais

Gestão Estratégica

Dados Abertos

Servidores

Informações
Classificadas

Perguntas
Frequentes

Serviço de
Informação ao
Poder Judiciário

4. Tabelas de Correlações de Códigos e Classificações: Arquivos em CSV para correspondência dos códigos utilizados nos dados acima, apresentando as descrições, classificações e agrupamentos utilizados nas divulgações e publicações. Todas as tabelas também estão disponíveis em um único arquivo Excel (.xlsx)

Produtos	Países e Blocos	Municípios e Estados	Via e URF
<ul style="list-style-type: none">▪ <u>NCM</u>▪ <u>SH - Sistema Harmonizado</u>▪ CUCI - Classificação Uniforme do Comércio Internacional▪ ISIC - Classificação Internacional Padrão por Atividade Econômica▪ ISICxCUCI- Seção da ISIC e nível Grupo da CUCI (classificações usadas nas divulgações semanais e mensais preliminares; e coletiva da balança comercial).▪ CGCE - Classificação por Grandes Categorias Econômicas▪ Fator Agregado▪ PPE - Pauta de Produtos Exportados▪ PPI - Pauta de Produtos Importados▪ Unidades Estatísticas da NCM▪ NBM (1989-1996) - Nomenclatura Brasileira de Mercadorias▪ NBMxNCM - Tabela de conversão	<ul style="list-style-type: none">▪ <u>Países</u>▪ Blocos	<ul style="list-style-type: none">▪ Municípios▪ Estados	<ul style="list-style-type: none">▪ <u>Via</u>▪ URF (Unidade da RFB)
Todas tabelas em planilhas Excel (.xlsx)			

A Nomenclatura Comum do Mercosul (NCM) é um código composto por 8 dígitos que oferece informações ao governo, para controle e identificação de produtos sujeitos a tributação nas operações comerciais.

O Sistema Harmonizado de Designação e de Codificação de Mercadorias, ou simplesmente Sistema Harmonizado (SH), é um **método internacional** de classificação de mercadorias, baseado em uma estrutura de códigos e respectivas descrições. São 6 dígitos

Composição do Código NCM e SH



01. - Animais Vivos - 2 primeiros dígitos do SH



01.02. - Animais Vivos da espécie bovina - 4 primeiros dígitos do SH

01.02.29. - Bovinos domésticos - 6 primeiros dígitos do SH

01.02.29.1. - Para reprodução - 7º dígito do NCM



01.02.29.1.0 - Outros - 8º dígito do NCM

5. PgAdmin 4 - “Importar arquivo”

- Primeiro deve-se criar um Table (pré-arquivo) com o mesmo número de colunas do arquivo a ser estudado.
- Existem duas maneiras para fazer isso:
 - Utilizando o assistente do PgAdmin ou
 - Criando manualmente com query



5. PgAdmin 4 - “Importar arquivo”

pg

File Object View Tools View Help

Browser

Servers (1)

rds_lucy

Databases (2)

postgres

Cast

Catalogs

Event Triggers

Extensions

Foreign Data Wrappers

Languages

Schemas (1)

public

Collations

Domains

FTS Configurations

FTS Dictionaries

Aa FTS Parsers

FTS Templates

Foreign Tables

Functions

Materialized Views

Procedures

Sequences

Tables (6)

impo

impo2

ncm

ncm_sh

pais

via

Trigger Functions

Types

Views

rdsadmin

Login/Group Roles

Tablespaces (2)

pg_default

pg_global

Dashboard

Properties

SQL

Statistics

Dependencies

Dependents

Projeto Final apresentação.sql

postgres/postgreslucy@rds_lucy

Query Editor Query History

```

1 select 'Hello happy world';
2
3 -- Projeto Final
4 -- PRIMEIRO PASSO É CRIAR UMA TABLE PARA O ARQUIVO A SER IMPORTADO
5
6 CREATE TABLE impo (
7     ano          VARCHAR(4) NOT NULL,    --"CO_ANO";
8     mes          VARCHAR(2) NOT NULL,    --"CO_MES";
9     ncm          VARCHAR(8) NOT NULL,    --"CO_NCM";
10    cod_unidade  VARCHAR(2) NOT NULL,    --"CO_UNID";
11    cod_pais_origem VARCHAR(3) NOT NULL, --"CO_PAIS";
12    cod_uf_destino VARCHAR(3) NOT NULL, --"SG_UF_NCM";
13    cod_via      VARCHAR(3) NOT NULL,    --"CO_VIA";
14    cod_urf      VARCHAR(7) NOT NULL,    --"CO_URF";
15    qtidade_estat VARCHAR(15) NOT NULL, --"QT_ESTAT";
16    kg_liquido   VARCHAR(20) NOT NULL,   --"KG_LIQUIDO";
17    vl_fob_dolar VARCHAR(25) NOT NULL    --"VL_FOB" valor dólar FOB (US$)
18 );
19
20 -- SEGUNDO PASSO É IMPORTAR PARA A TABLE O ARQUIVO .CSV
21
22 \COPY impo FROM 'D:\Harve\Projeto Final\teste\IMP_2020.csv' DELIMITER ';' QUOTE '"' CSV HEADER ENCODING 'UTF8'
23 --NÃO RECONHECEU ESTA QUERY, ENTÃO FIZ ATRAVES DO ASSISTENTE PARA IMPORTAR O ARQUIVO.
24

```

Data Output

Explain

Messages

Notifications

6. Análise dos dados das tabelas



7. Visualização - Power BI

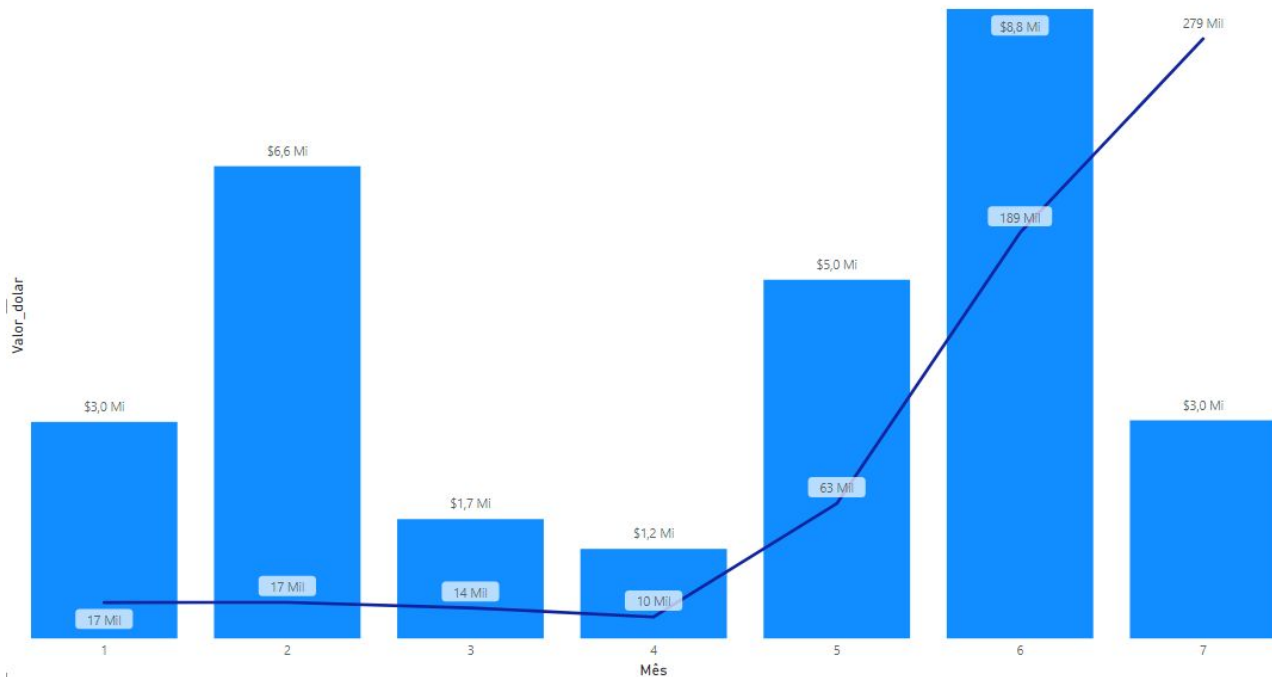
3002 - Sangue humano;

sangue animal preparado para usos terapêuticos, profiláticos ou de diagnóstico; anti-soros, outras frações do sangue, **produtos imunológicos modificados**, mesmo obtidos por via biotecnológica; vacinas, toxinas, culturas de microrganismos.

(*)Inclui Kits de teste para Covid-19, baseados em reações imunológicas

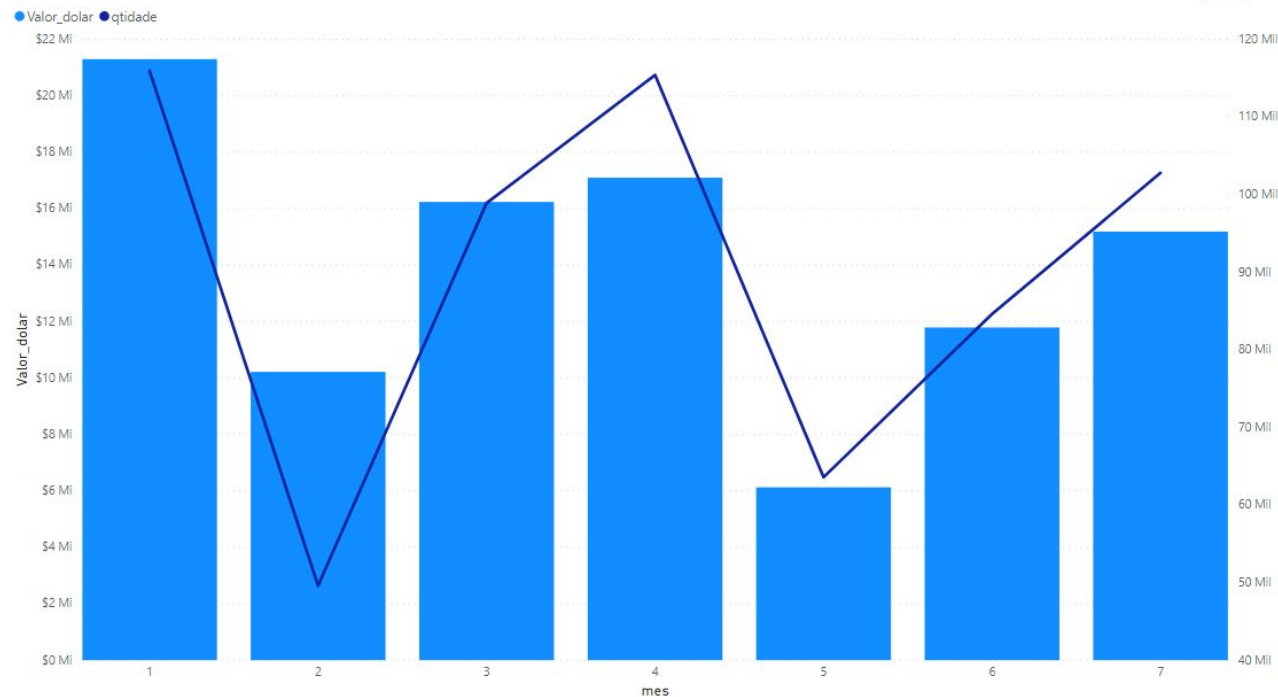
Importações do Estado Paraná da NCM 3002

Valor_dolar qtidade



7. Visualização - Power BI

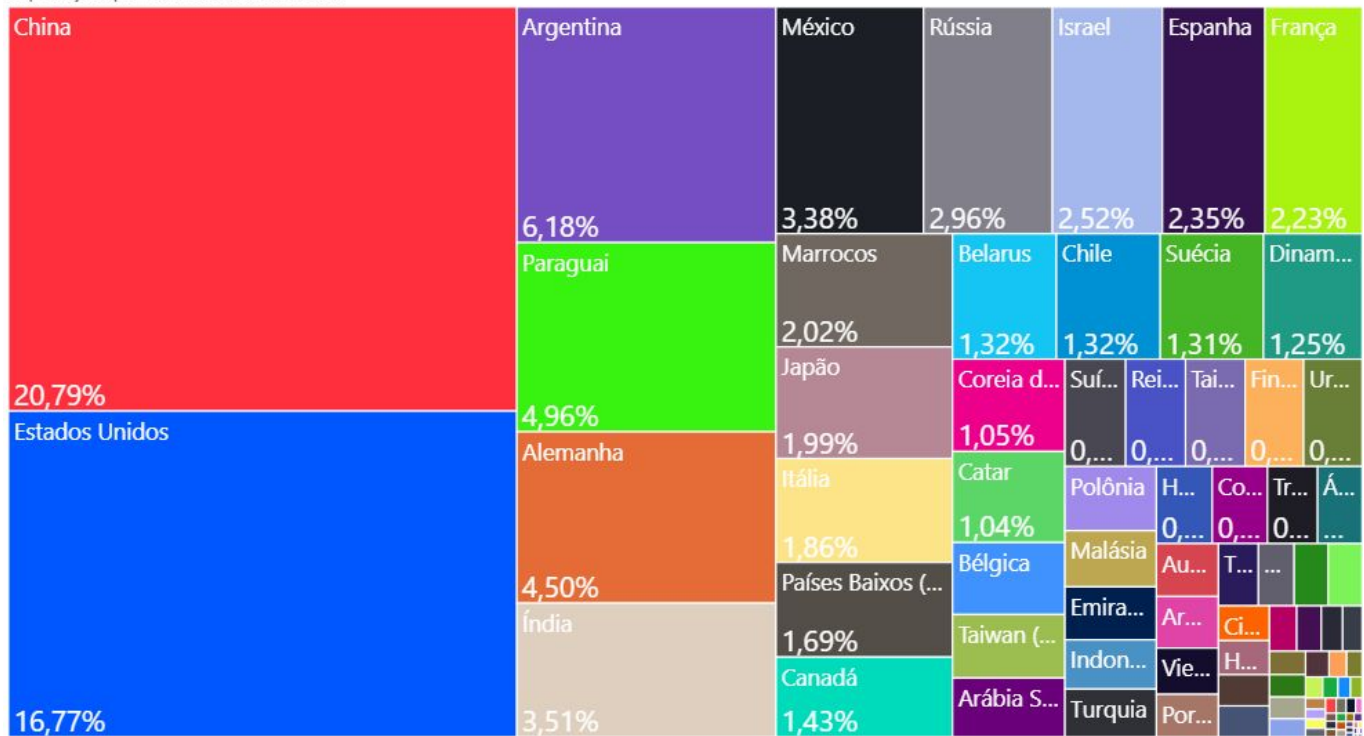
Importações do Estado Paraná da NCM 3004



3004 - Medicamentos
(exceto os produtos das posições 3002, 3005 ou 3006) constituídos por produtos misturados ou não misturados, preparados para fins terapêuticos ou profiláticos, apresentados em doses (incluindo os destinados a serem administrados por via sub

7. Visualização - Power BI

Importações por Países do Estado Paraná



(*)Valor FOB em
porcentagem
total

Referências

- ❖ <http://comexstat.mdic.gov.br/pt/home>
- ❖ <http://www.mdic.gov.br/comercio-exterior/estatisticas-de-comercio-exterior/>
- ❖ <https://www.escolalinux.com.br/blog/o-que-e-aws-pra-que-serve-e-por-que-devo-domina-la>
- ❖ <https://www.opservices.com.br/principais-servicos-da-aws-amazon-web-services/>
- ❖ https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/CHAP_GettingStarted.CreatingConnecting.PostgreSQL.html
- ❖ <https://www.postgresql.org/>
- ❖ <https://signin.aws.amazon.com/>

Projeto Final

Muito obrigada!

Turma : Cientista de Dados - T04
Aluna : Luciana Lanzoni Menges
