

Atividade 4 – Ingestão e ETL com linguagem de programação (Python + Spark + SQL)



EDB-011 - Ingestão de Dados

Grupo 4

Daniel de Moraes Alves

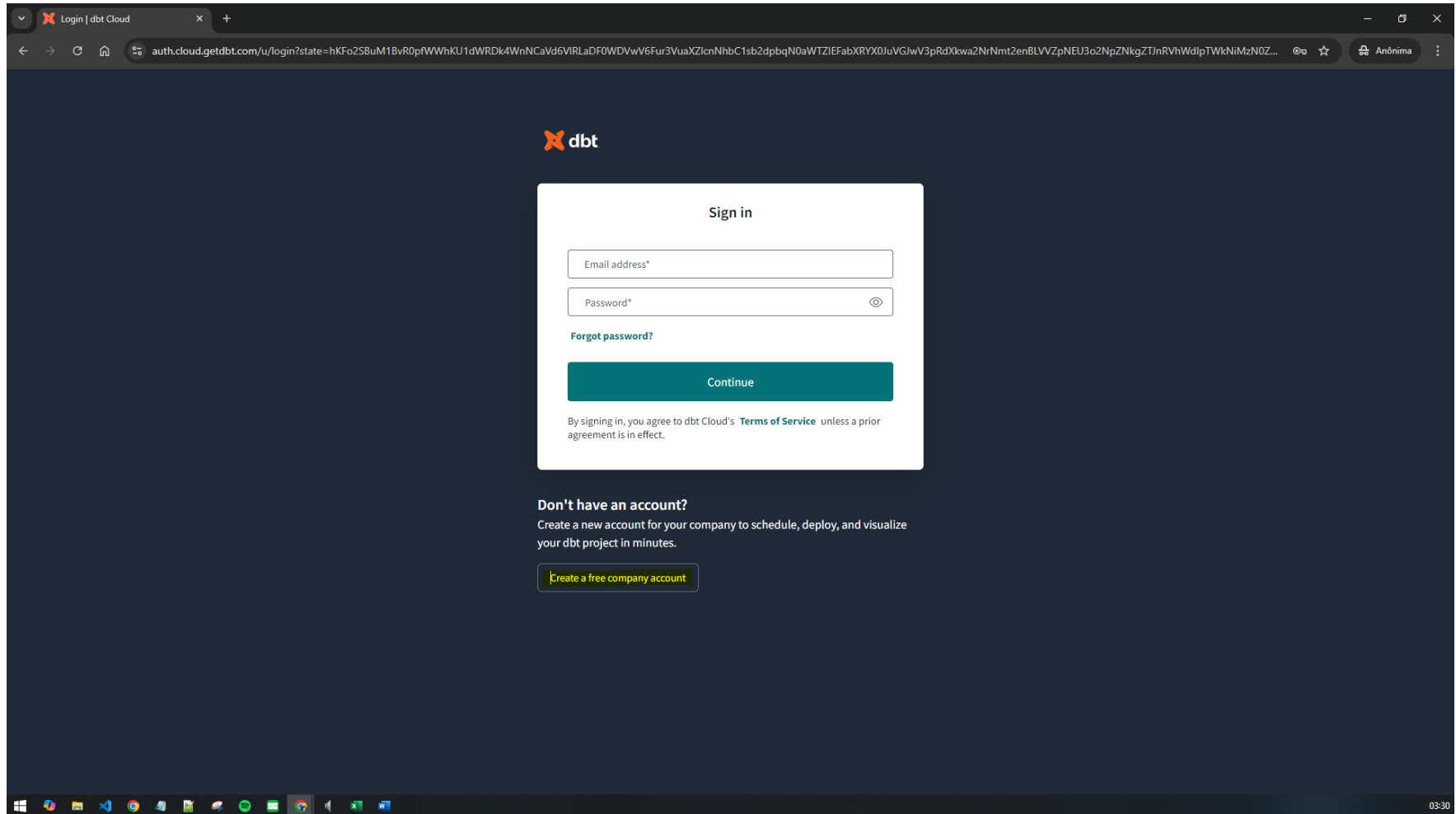
Gabriel Alencar Jeronimo Antonio

Julio Marques Honorio

Julia Muniz de Oliveira

Monalisa Rodrigues da Costa

1. Criar uma conta free no dbt



2. Criar uma nova account

dbt Cloud | Project Home

nd484.us1.dbt.com/accounts/new

Anônima

Navigation preview Data ingestion / Analytics

Create Account

Add a new account

Your dbt Cloud account is a shared workspace for developing and deploying dbt projects. Only you and invited teammates will have access to this account.

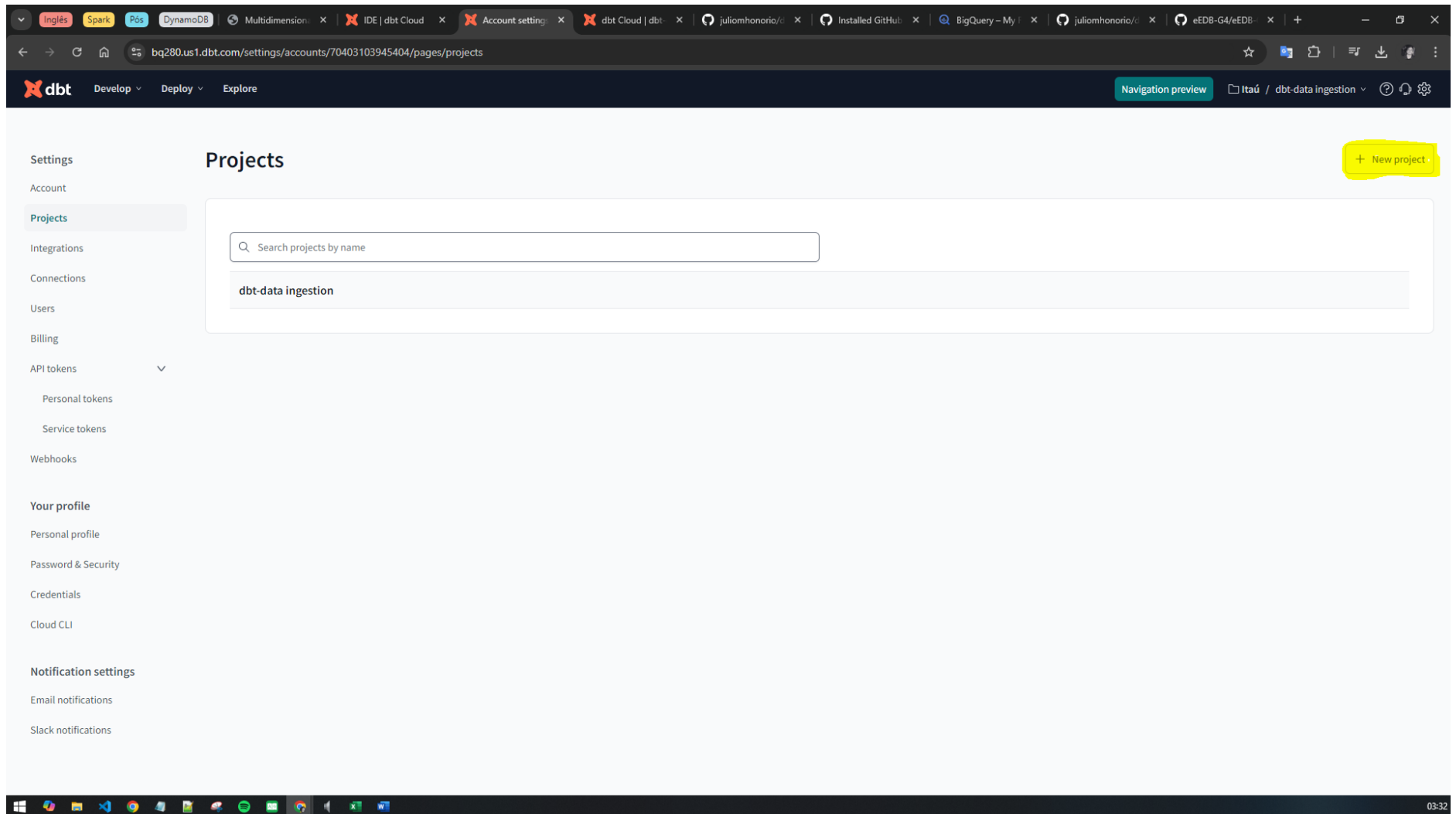
Account Name

dbt-data ingestion

Create

03:35

3. Criar um novo projeto



4. Criar uma nova conexão:

dbt Cloud | Project Home

sm023.us1.dbt.com/70403103946433/projects/70403103954634/setup

dbt Develop Deploy Explore

Navigation preview dbt-data ingestion / Analytics

Connections have been moved!
Connections are now an account-level resource (see them here). You are now able to share connections across projects and, within each project, use different connections for each environment. To learn more, check out our docs.

Complete project setup

Name your project Analytics ✓

Configure your development environment ✓

Connection
Select the warehouse connection to use for this environment.

Connection
Select...

Skip Save

Setup a repository ✓

03:36

5. Escolher qual o tipo de banco de dados que irá se conectar:

The screenshot displays the dbt Cloud interface for creating a new connection. The browser address bar shows the URL `sm023.us1.dbt.com/settings/accounts/70403103946433/pages/connections/new`. The dbt logo and navigation menu are at the top. A notification banner at the top states: "Connections have been moved! Connections are now an account-level resource (see them here). You are now able to share connections across projects and, within each project, use different connections for each environment. To learn more, check out our docs." The left sidebar lists various settings categories, with "Connections" currently selected. The main panel, titled "Add new connection", contains a "Connection" section with a link to a "connection setup guide". Below this, there is a grid of database provider buttons: Snowflake, PostgreSQL, BigQuery (highlighted in yellow), Redshift, Databricks, Apache Spark, Starburst (Trino compatible), Synapse, Fabric, and Athena. At the top right of the main panel, there are "Cancel" and "Save" buttons. The Windows taskbar at the bottom shows the time as 03:37.

6. Criar uma conta gratuita no BigQuery

Google Cloud | Visão Geral | Soluções | Produtos | Preços | Recursos | Entre Em Contato Com Nossa Equipe

O Google foi reconhecido como líder no relatório The Forrester Wave™: Data Lakehouses do segundo trimestre de 2024.

BigQuery

Do data warehouse a uma plataforma de dados unificada e pronta para a IA

O BigQuery é uma plataforma de análise de dados totalmente gerenciada e pronta para IA que ajuda a impulsionar o valor dos dados, funcionando com vários mecanismos, formatos e nuvens.

Armazene 10 GiB de dados e execute até 1 TiB de consultas gratuitamente por mês. **Novos clientes ganham US\$ 300 em créditos** para testar o BigQuery e outros produtos do Google Cloud.

[Comece a usar gratuitamente](#) [Entre em contato com a equipe de vendas](#)

RECURSOS

Capacite seus agentes de dados com o Gemini no BigQuery

O [Gemini no BigQuery](#) oferece recursos assistivos e de colaboração com tecnologia de IA, incluindo [assistência de código](#), preparação visual de dados e recomendações inteligentes que ajudam a aumentar a produtividade e otimizar custos. O BigQuery oferece um [único espaço de trabalho](#) que inclui SQL, um [notebook](#) e uma [interface de tela](#) baseada em NL para profissionais de dados com várias habilidades de programação para simplificar os fluxos de trabalho de análise, desde a ingestão e preparação de dados até a exploração e visualização de dados, além da criação e

Características do produto

- Plataforma de dados unificada para conectar todos os dados e ativar com IA
- Machine learning integrado para criar e executar modelos usando SQL simples
- Análise em tempo real com streaming e BI integrados

Inscreva-se no workshop de descoberta sem custos financeiros

vídeo

Saiba como criar agentes de dados com o

03:39

BigQuery - My First Project - C

console.cloud.google.com/bigquery?referrer=search&project=steel-binder-433800-p9&ws=11m511m414m311steel-binder-433800-p912sschema13s2021_tri_01

Status da avaliação gratuita: R\$ 1.928,04 de crédito e 91 dias restantes. Ative sua conta completa para ter acesso ilimitado a todos os recursos do Google Cloud. Use os créditos restantes e pague apenas pelo que usar.

DISPENSAR ATIVAR

Google Cloud My First Project

Pesquise (/) recursos, documentos, produtos e muito mais Pesquisa

Explorer + ADICIONAR

Pesquisar recursos do BigQuery

Você está visualizando os recursos. MOSTRAR APENAS COM ESTRELA

- steel-binder-433800-p9
 - Consultas
 - Notebooks
 - Telas de dados
 - Preparações de dados
 - Conexões externas
 - dbt_jmarquesonorio
 - schema
 - 2021_tri_01
 - 2021_tri_02
 - 2021_tri_03
 - 2021_tri_04
 - 2022_tri_01
 - 2022_tri_02
 - 2022_tri_03
 - 2022_tri_04
 - EnquadramentoInicia_v2
 - glassdoor_consolidado...
 - glassdoor_consolidado...

RESUMO

2021_tri_01

steel-binder-433800-p9.schema

Última modificação 27 de ago. de 2024, 03:41 UTC-3

2021_tri_01

CONSULTA COMPARTILHAR COPIAR SNAPSHOT EXCLUIR EXPORTAR ATUALIZAR

ESQUEMA DETALHES VISUALIZAÇÃO BUSCADOR DE TABELAS PRÉ-VISUALIZAÇÃO INSIGHTS LINHAGEM PERFIL DE DADOS QUALIDADE DOS DADOS

Filtro Insira o nome ou o valor da propriedade

<input type="checkbox"/>	Nome do campo	Tipo	Modo	Chave	Compilação	Valor padrão	Tags de políticas	Descrição
<input type="checkbox"/>	Ano	INTEGER	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	Trimestre	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	Categoria	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	Tipo	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	CNPJ IF	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	Instituicao financeira	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	Índice	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	Quantidade de reclamacoes reguladas procedentes	INTEGER	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	Quantidade de reclamacoes reguladas - outras	INTEGER	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	Quantidade de reclamacoes nao reguladas	INTEGER	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	Quantidade total de reclamacoes	INTEGER	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	Quantidade total de clientes - CCS e SCR	INTEGER	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	Quantidade de clientes - CCS	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	Quantidade de clientes - SCR	STRING	NULLABLE	-	-	-	-	-

EDITAR ESQUEMA VER POLÍTICAS DE ACESSO DE LINHA

Histórico de jobs

ATUALIZAR

8. Entrar no menu a esquerda do console do BigQuery > API e Serviços > Biblioteca e por fim pesquisar BigQuery API e ativa-la.

The screenshot shows the Google Cloud console interface. At the top, the browser address bar displays the URL: `console.cloud.google.com/apis/library/bigquery.googleapis.com?project=steel-binder-433800-p9`. Below the browser, a status bar indicates a free trial of R\$ 1,928.04 with 91 days remaining. The main header shows the Google Cloud logo and the project name "My First Project".

The page title is "Detalhes do produto". The main content area features the BigQuery API logo and the text "BigQuery API" and "Google Enterprise API". Below this, it states: "A data platform for customers to create, manage, share and query data." There are three buttons: "GERENCIAR", "TESTAR ESTA API", and a green button indicating "API ativada".

Below the buttons, there are three tabs: "VISÃO GERAL" (selected), "DOCUMENTAÇÃO", and "PRODUTOS RELACIONADOS".

The "Visão geral" section contains the same description: "A data platform for customers to create, manage, share and query data." To the right, under "Mais detalhes", it lists: "Tipo: SaaS & APIs", "Última atualização do produto: 21/07/2022", "Categoria: Big data, Google Enterprise APIs", and "Nome do serviço: bigquery.googleapis.com".

Below this, there is a section for "Tutoriais e documentação" with a link "Learn more". At the bottom, there is a "Termos de Serviço" section with the text: "Ao usar este produto, você concorda com os Termos e Condições da licença a seguir: Google Cloud Platform."

The Windows taskbar is visible at the bottom of the screen, showing various application icons and the system clock at 03:45.

9. Entrar no menu a esquerda do console do BigQuery > API e Serviços > Credentials > e criar uma service account para o dbt

API

APIs e serviços

APIs e serviços ativados

Biblioteca

Credenciais

Tela de permissão OAuth

Contratos de uso de página

Credenciais

+ CRIAR CREDENCIAIS

EXCLUIR

RESTAURAR CREDENCIAIS EXCLUÍDAS

Crie credenciais para acessar suas APIs ativas. [Saiba mais](#)

Lembre-se de configurar a tela de permissão OAuth com informações sobre seu aplicativo. [CONFIGURAR TELA DE CONSENTIMENTO](#)

Chaves de API

<input type="checkbox"/>	Nome	Data da criação	Restrições	Ações
Nenhuma chave de API a ser exibida				

IDs do cliente OAuth 2.0

<input type="checkbox"/>	Nome	Data da criação	Tipo	ID do cliente	Ações
Nenhum cliente do OAuth a ser exibido					

Contas de serviço

[Gerenciar contas de serviço](#)

<input type="checkbox"/>	E-mail	Nome	Ações
<input type="checkbox"/>	[redacted]@iam.gserviceaccount.com	dbt_data_ingestion	

Status da avaliação gratuita: R\$ 1.928,04 de crédito e 91 dias restantes. Ative sua conta completa para ter acesso ilimitado a todos os recursos do Google Cloud. Use os créditos restantes e pague apenas pelo que usar.

DISPENSAR

ATIVAR

Pesquise (/) recursos, documentos, produtos e muito mais

Pesquisa

1

?

<|

03:48

10. Durante a criação da service account devemos associar a role de bigQuery admin ou Bigquery Data Editor e posteriormente extrair as credenciais no formato json.

The screenshot displays the Google Cloud IAM and Admin console interface. The browser address bar shows the URL: `console.cloud.google.com/iam-admin/serviceaccounts/details/100269581913568410585?project=steel-binder-433800-p9`. The page title is "dbt_data_ingestion - IAM e administrador". The left sidebar contains a navigation menu with options like IAM, PAM, Limite de acesso principal, Identidade e organização, Solucionador de problemas, Analisador de políticas, Políticas da organização, Contas de serviço (highlighted), Federação de identidade, Rótulos, Tags, Configurações, Privacidade e segurança, Identity-Aware Proxy, Papéis, Registro de auditoria, Gerenciar recursos, and Notas de lançamento. The main content area is titled "dbt_data_ingestion" and has tabs for DETALHES, PERMISSÕES, CHAVES, MÉTRICA, and REGISTROS. The "DETALHES" tab is active, showing the "Detalhes da conta de serviço" section. This section includes fields for "Nome" (dbt_data_ingestion), "Descrição", "E-mail" (a redacted email address ending in @p9.iam.gserviceaccount.com), and "ID exclusivo" (a redacted ID). Below this is the "Status da conta de serviço" section, which indicates the account is "Ativada" (Active) with a green checkmark and a "DESATIVAR CONTA DE SERVIÇO" button. At the bottom, there is a "Configurações avançadas" section with a dropdown arrow. The top right of the console shows a status message about a free trial and buttons for "DISPENSAR" and "ATIVAR". The bottom of the image shows a Windows taskbar with various application icons and a system clock showing 03:55.

11. Entramos nas connections e adicionamos o arquivo json pra preencher a conexão

The screenshot shows the dbt Cloud interface for setting up a BigQuery connection. The browser address bar shows the URL: `bq280.us1.dbt.com/settings/accounts/70403103945404/pages/connections/70403103940750`. The dbt logo is in the top left, and navigation links for 'Develop', 'Deploy', and 'Explore' are next to it. A 'Navigation preview' button and a breadcrumb 'Itaú / dbt-data ingestion' are in the top right.

The left sidebar contains the following menu items:

- Settings
 - Account
 - Projects
 - Integrations
 - Connections**
 - Users
 - Billing
 - API tokens
 - Personal tokens
 - Service tokens
 - Webhooks
- Your profile
 - Personal profile
 - Password & Security
 - Credentials
 - Cloud CLI
- Notification settings
 - Email notifications
 - Slack notifications

The main content area is titled 'Connections / BigQuery' and 'BigQuery'. It includes a 'Cancel' and 'Save' button in the top right. The 'Connection' section has a 'BigQuery' icon and a 'Connection name' field with the value 'BigQuery'. The 'Settings' section includes instructions on IP access, an 'Upload from file' button, and fields for 'Project ID', 'Job Execution Timeout Seconds', and 'Private Key ID'.

The 'Upload from file' button is highlighted with a yellow box and labeled 'Upload a Service Account JSON file'. The 'Project ID' field contains the value 'steel-bi'. The 'Job Execution Timeout Seconds' field contains the value '300'. The 'Private Key ID' field contains the value '08eb1'.

12. Configurar a conta git

The screenshot shows the 'Your profile' settings page in the dbt Cloud interface. The browser address bar shows the URL `bq280.us1.dbt.com/settings/profile`. The left sidebar contains a navigation menu with sections: 'Settings' (Account, Projects, Integrations, Connections, Users, Billing, API tokens), 'Your profile' (Personal profile, Password & Security, Credentials, Cloud CLI), and 'Notification settings' (Email notifications, Slack notifications). The main content area is titled 'Personal information' and includes an 'Edit' button. It features input fields for 'First name' (Julio), 'Last name' (Marques Honorio), and 'Email' (julio...), with a note about email verification. Below this is the 'Linked accounts' section, which shows a connected GitHub account 'juliomhonorio' with a 'Configure integration in GitHub' button and a warning about deprecation. A GitLab account is also listed with a 'Link' button. At the bottom, there is an 'API' section with a 'View docs' link and a yellow warning banner stating: 'WARNING: User API Keys are being deprecated on Sept 18. User API Keys are being replaced by the more secure, account-scoped Personal Access Tokens (PATs). To maintain uninterrupted access and enhanced security, we encourage you to...'

dbt Cloud interface showing the 'Your profile' settings page. The page is titled 'Personal information' and includes an 'Edit' button. The profile information is as follows:

- First name: Julio
- Last name: Marques Honorio
- Email: julio... (redacted)

Below the profile information, there is a section for 'Linked accounts'. It shows a connected GitHub account named 'juliomhonorio' with a 'Configure integration in GitHub' button. A note states: 'You have connected the dbt Cloud GitHub app to the following accounts:'. Below this, there is a GitLab account listed with a 'Link' button.

At the bottom of the page, there is a warning message: 'WARNING: User API Keys are being deprecated on Sept 18. User API Keys are being replaced by the more secure, account-scoped Personal Access Tokens (PATs). To maintain uninterrupted access and enhanced security, we encourage you to...'

13. Configurar conta git e criar um repositório dedicado para o projeto com dbt + Bigquery

The screenshot shows the GitHub web interface in a browser. The address bar displays `github.com/settings/installations/54193499`. The page is for a repository named **dbt Cloud**, installed 5 hours ago, developed by [dbt-labs](#), with a link to <https://www.getdbt.com>.

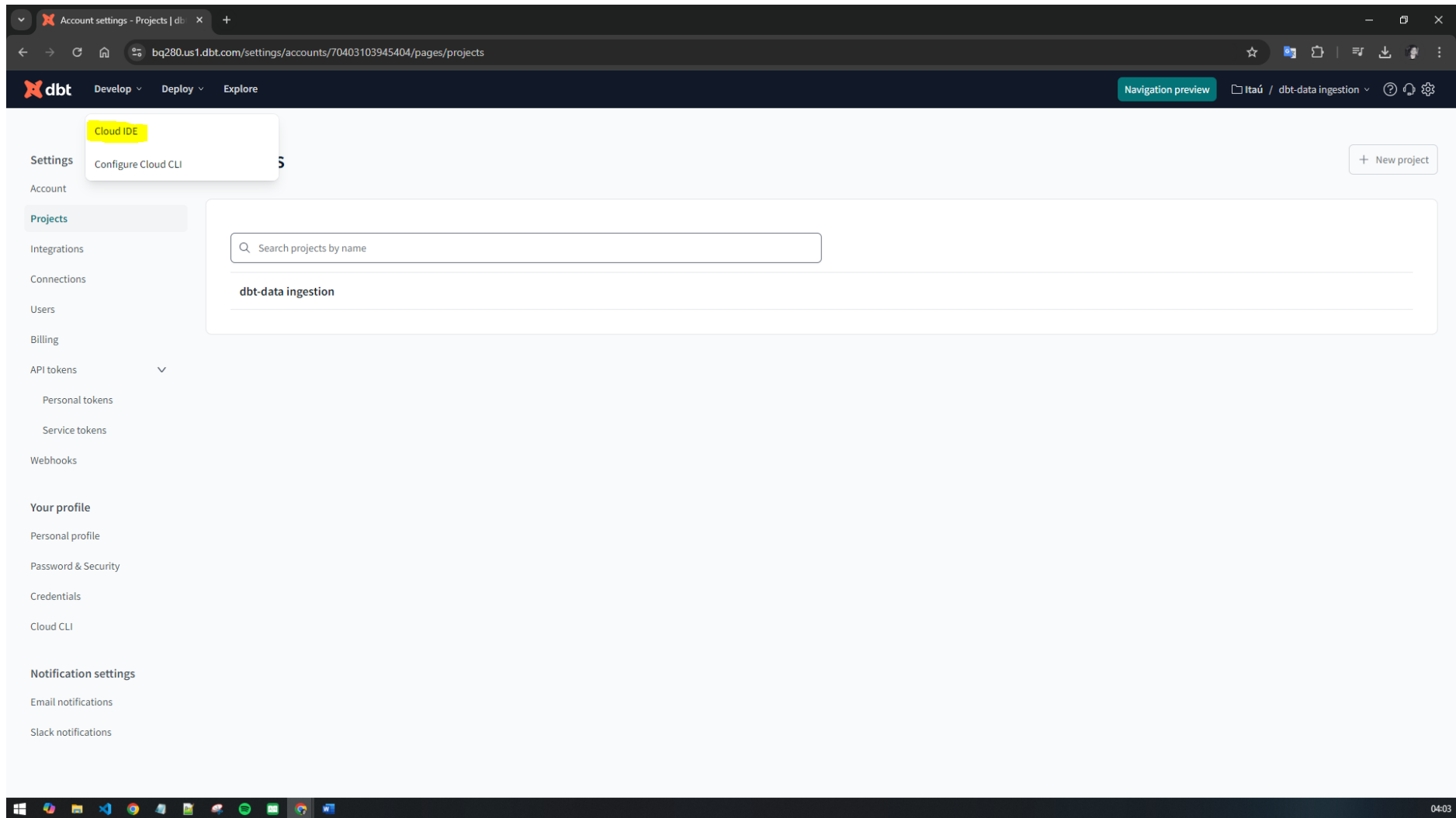
On the left, the user profile for **Julio Marques Honorio (juliomhonorio)** is shown, with a sidebar menu containing options like Public profile, Account, Appearance, Accessibility, Notifications, Access, Billing and plans, Emails, Password and authentication, Sessions, SSH and GPG keys, Organizations, Enterprises, Moderation, Code, planning, and automation, Repositories, Codespaces, Packages, Copilot, Pages, Saved replies, Security, and Code security and analysis.

The main content area is divided into sections:

- Permissions:** Shows two permissions with green checkmarks:
 - Read access to metadata
 - Read and write access to checks, code, commit statuses, pull requests, repository hooks, and workflows
- Repository access:** Contains two radio button options:
 - All repositories:** This applies to all current and future repositories owned by the resource owner. Also includes public repositories (read-only).
 - Only select repositories:** Select at least one repository. Also includes public repositories (read-only). This option is selected.A button labeled **Select repositories** is present. Below it, a list shows **Selected 1 repository:**
 - `juliomhonorio/dbt_data` (with a lock icon and a close button)

At the bottom of the repository access section, there are **Save** and **Cancel** buttons. The Windows taskbar is visible at the very bottom of the screen.

14. Após as configurações realizadas, acessar o Cloud IDE via dbt.



15. Entrar no Cloud IDE e dar o primeiro commit para que os arquivos possam ser criados no repositório

The screenshot displays the dbt Cloud IDE interface in a web browser. The browser's address bar shows the URL `bq280.us1.dbt.com/develop/70403103945404/projects/70403103954283`. The interface includes a top navigation bar with the dbt logo and tabs for 'Develop', 'Deploy', and 'Explore'. On the left side, there is a sidebar with a 'Version control' section. This section contains a 'Commit and sync' button and a list of files to be committed, each with a status icon (A for added, M for modified). The files listed are:

- `.gitkeep` analyses (A)
- `.gitkeep` macros (A)
- `my_first_dbt_model.sql` (A)
- `my_second_dbt_model.sql` (A)
- `schema.yml` (A)
- `.gitkeep` seeds (A)

Below the 'Version control' section is a 'File explorer' showing the project structure:

- `dbt` (M)
- `analyses` (M)
- `dbt_packages` (M)
- `macros` (M)
- `models` (M)
- `example` (M)
- `my_first_dbt_model.sql` (A)
- `my_second_dbt_model.sql` (A)
- `schema.yml` (A)
- `seeds` (M)
- `snapshots` (M)
- `target` (M)
- `tests` (M)
- `.gitignore` (A)
- `README.md` (A)
- `dbt_project.yml` (A)

The main area of the IDE is currently empty, displaying a message: 'No Open Files. Get started by opening a file, or create a new Scratchpad.' with a '+ Create new file' button. At the bottom of the interface, there is a status bar showing the command `dbt build --select <model_name>` and a 'Defer to staging/production' toggle set to 'Ready'.

16. Configurar o arquivo `dbt_project.yml` com as variáveis corretas e dessa forma conseguir criar os modelos com o lineage e code quality.

The screenshot displays the dbt Cloud IDE interface. The top navigation bar includes the dbt logo and tabs for 'Develop', 'Deploy', and 'Explore'. The main workspace is divided into three panels: a left sidebar, a central editor, and a bottom status bar.

Left Sidebar:

- Version control:** Shows a 'Commit and sync' button and a list of changes, including 'reviews_complaints_delivery.sql'.
- File explorer:** Displays a tree view of the project files, including 'dbt', 'analyses', 'dbt_packages', 'macros', 'models', 'example', 'seeds', 'snapshots', 'target', 'tests', '.gitignore', 'README.md', and 'dbt_project.yml' (highlighted).

Central Editor:

The editor shows the `dbt_project.yml` file with the following content:

```
23 clean-targets: # directories to be removed by `dbt clean`
26
27
28 # Configuring models
29 # Full documentation: https://docs.getdbt.com/docs/configuring-models
30
31 # In dbt, the default materialization for a model is a view. This means, when you run
32 # dbt run or dbt build, all of your models will be built as a view in your data platform.
33 # The configuration below will override this setting for models in the example folder to
34 # instead be materialized as tables. Any models you add to the root of the models folder will
35 # continue to be built as views. These settings can be overridden in the individual model files
36 # using the `{{ config(...) }}` macro.
37
38 models:
39   my_new_project:
40     # Applies to all files under models/example/
41     example:
42       +materialized: table
43
```

Bottom Status Bar:

The status bar shows the command `dbt build --select <model_name>` and a toggle for 'Defer to staging/production' (currently off). A 'Ready' status indicator is visible.

17. Agora devemos criar as tabelas na camada RAW

The screenshot displays the dbt Cloud IDE interface. The top navigation bar includes the dbt logo and tabs for 'develop', 'Deploy', and 'Explore'. The left sidebar contains a 'Version control' section with a 'Create a pull request on GitHub' button and a 'File explorer' section showing the project structure. The main editor area shows the 'raw_banks.sql' file with the following SQL query:

```
1 SELECT
2
3 *
4
5 FROM schema.EnquadramentoInicia_v2
6
```

Below the editor, there are buttons for 'Preview', 'Compile', 'Build', and 'Format'. The 'Results' tab is active, showing a table with 1.1s execution time and 500 rows. The table has three columns: 'Segmento', 'CNPJ', and 'Nome'.

Segmento	CNPJ	Nome
S1	0	BANCO DO BRASIL - PRUDENCIAL
S1	60746948	BRADERCO - PRUDENCIAL
S1	30306294	BTG PACTUAL - PRUDENCIAL
S1	360305	CAIXA ECONOMICA FEDERAL - PRUDENCIAL
S1	60872504	ITAU - PRUDENCIAL
S1	90400888	SANTANDER - PRUDENCIAL
S1	0	BANCO DO BRASIL S/A
S1	33058660	BANCO ITAU S/A

The bottom status bar shows the command 'dbt build --select <model_name>' and a 'Defer to staging/production' toggle.

18. Posteriormente criar as tabelas na camada Trusted, aqui podemos identificar que o lineage já está sendo aplicado.

The screenshot displays the dbt Cloud IDE interface. The top navigation bar includes tabs for 'Inglês', 'Spark', 'Pós', 'DynamoDB', 'Multidimension...', 'IDE | dbt Cloud', 'dbt Cloud | dbt...', 'juliomhonorio/c...', 'Installed GitHub', 'juliomhonorio/c...', 'dbt_data_ingest...', 'eEDB-G4/eEDB...', and 'Nova guia'. The address bar shows the URL 'bq280.us1.dbt.com/develop/70403103945404/projects/70403103954283'. The main interface is divided into three sections: a left sidebar, a central editor, and a bottom lineage graph.

Left Sidebar:

- Version control:** Includes a button 'Create a pull request on GitHub'.
- File explorer:** Shows a tree view of the project structure:
 - dbt
 - analyses
 - dbt_packages
 - macros
 - models
 - example
 - my_first_dbt_model.sql
 - my_second_dbt_model.sql
 - raw_banks.sql
 - raw_complaints.sql
 - raw_employees.sql
 - reviews_complaints_delivery.sql
 - schema.yml
 - trusted_banks.sql
 - trusted_complaints.sql** (selected)
 - trusted_employees.sql
 - seeds
 - snapshots
 - target
 - tests
 - .gitignore
 - README.md
 - dbt_project.yml

Central Editor:

The editor shows the SQL code for the model 'trusted_complaints.sql'. The code is as follows:

```
1 WITH raw_complaints AS (  
2   SELECT * FROM {{ref('raw_complaints')}}  
3 )  
4  
5 SELECT  
6  
7   CAST('Ano' AS INTEGER) AS ano ,  
8   CAST('Trimestre' AS STRING) AS trimestre ,  
9   CAST('Categoria' AS STRING) AS categoria ,  
10  CAST('Tipo' AS STRING) AS tipo ,  
11  CAST('CNPJ IF' AS STRING) AS cnpj_if ,  
12  CAST('Instituicao financeira' AS STRING) AS instituicao_financeira ,  
13  CAST('Indice' AS STRING) AS indice ,  
14  CAST('Quantidade de reclamacoes reguladas - outras' AS INTEGER) AS quantidade_de_reclamacoes_reguladas_outras ,  
15  CAST('Quantidade de reclamacoes nao reguladas' AS INTEGER) AS quantidade_de_reclamacoes_nao_reguladas ,  
16  CAST('Quantidade total de reclamacoes' AS INTEGER) AS quantidade_total_de_reclamacoes ,  
17  CAST('Quantidade total de clientes - CCS e SCR' AS STRING) AS quantidade_total_de_clientes_ccs_scr ,  
18  CAST('Quantidade de clientes - CCS' AS STRING) AS quantidade_de_clientes_ccs ,  
19  CAST('Quantidade de clientes - SCR' AS STRING) AS quantidade_de_clientes_scr  
20  
21 FROM schema.2021_tri_01
```

Bottom Lineage Graph:

The lineage graph shows the data flow from the 'raw_complaints' model to the 'trusted_complaints' model, and then to the 'reviews_complaints_delivery' model. The 'trusted_complaints' model is highlighted in yellow. The graph includes a button '2+trusted_complaints+2' and an 'Update Graph' button.

The bottom status bar shows the command 'dbt build --select <model_name>' and a toggle for 'Defer to staging/production'.

19. Criar o modelo da camada delivery, aqui temos o lineage completo

IDE | dbt Cloud

dbt Cloud | dbt-dat... | juliomhonorio/dbt... | Installed GitHub Ap... | juliomhonorio/dbt... | dbt_data_ingestion... | eEDB-G4/eEDB-011

bq280.us1.dbt.com/develop/70403103945404/projects/70403103954283

dbt Develop Deploy Explore Navigation preview Itaú / dbt-data ingestion

develop Change branch

Version control

Create a pull request on GitHub

File explorer

- dbt
- analyses
- dbt_packages
- macros
- models
 - example
 - my_first_dbt_model.sql
 - my_second_dbt_model.sql
 - raw_banks.sql
 - raw_complaints.sql
 - raw_employees.sql
 - reviews_complaints_delivery.sql
- schema.yml
- trusted_banks.sql
- trusted_complaints.sql
- trusted_employees.sql

seeds

snapshots

target

tests

.gitignore

README.md

dbt_project.yml

reviews_complaints_deliv... raw_banks.sql trusted_complaints.sql

models / example / reviews_complaints_delivery.sql

```
1 with
2   trusted_banks as (select * from {{ ref("trusted_banks") }}),
3
4   trusted_complaints as (select * from {{ ref("trusted_complaints") }}),
5
6   trusted_employees as (select * from {{ ref("trusted_employees") }}),
7
8
9 df_complaints AS (
10  SELECT
11    YEAR,
12    QUARTER,
13    INSTITUTION_CATEGORY,
14    INSTITUTION_TYPE,
15    CNPJ,
16    INSTITUTION_NAME,
17    QTY_JUSTIFIED_REGULATED_COMPLAINTS,
18    QTY_OTHER_REGULATED_COMPLAINTS,
19    QTY_NOT_REGULATED_COMPLAINTS,
20    QTY_COMPLAINTS
21  FROM
```

Preview </> Compile Build Format Results Code quality Compiled code Lineage

2+reviews_complaints_deliv Update Graph

```
graph LR
    raw_banks[raw_banks] --> trusted_banks[trusted_banks]
    raw_complaints[raw_complaints] --> trusted_complaints[trusted_complaints]
    raw_employees[raw_employees] --> trusted_employees[trusted_employees]
    trusted_banks --> reviews_complaints_delivery[reviews_complaints_delivery]
    trusted_complaints --> reviews_complaints_delivery
    trusted_employees --> reviews_complaints_delivery
```

dbt build --select <model_name>

Defer to staging/production Error

04:10

20. Rodar o comando “dbt run” e efetuar o run de todos os modelos.

The screenshot displays the dbt Cloud web interface. At the top, the browser address bar shows the URL `bq280.us1.dbt.com/develop/70403103945404/projects/70403103954283`. The interface is divided into several sections:

- Command Input:** A dropdown menu shows `dbt build --select <model_name>`. Below it, a text input field contains `dbt run`.
- Recent Runs Table:** A table listing recent executions of `dbt run` on the `develop` branch. The table has columns for the command, status, and duration.
- Run Details Panel:** A panel on the right provides details for the selected `dbt run` command, including its status (success), branch, and a list of system logs.

Recent Runs Table:

Command	Status	Duration
<code>dbt run</code>	success	11.4s
<code>dbt run</code>	success	10.6s
<code>dbt run</code>	failed	10.4s
<code>dbt run --select trusted_banks</code>	success	3.1s
<code>dbt run</code>	success	5.2s
<code>dbt run</code>	failed	0.0s
<code>dbt run</code>	failed	0.0s
<code>dbt run</code>	failed	0.0s
<code>dbt run</code>	failed	0.0s
<code>dbt run</code>	failed	0.0s
<code>dbt run</code>	failed	0.0s
<code>dbt build --select raw_banks</code>	failed	0.0s

Run Details Panel:

dbt run success

develop
less than a minute ago

> System Logs

All 9 Pass 9 Warn 0 Error 0 Skip 0 Running 0

Model	Duration
my_first_dbt_model	2.92s
raw_banks	2.92s
raw_complaints	2.93s
raw_employees	2.79s
trusted_employees	3.09s
my_second_dbt_model	2.48s
trusted_banks	3.93s
trusted_complaints	4.88s
reviews_complaints_delivery	2.59s

21. Rodar o comando dbt run --models example.trusted_bank para adicionar as tabelas tratadas ao BigQuery

The screenshot displays the dbt Cloud IDE interface. At the top, the browser address bar shows the URL `bq280.us1.dbt.com/develop/70403103945404/projects/70403103954283`. The IDE workspace includes a file explorer on the left with tabs for `reviews_complaints_deliv...`, `raw_banks.sql`, `trusted_complaints.sql`, `trusted_employees.sql`, `dbt_project.yml`, and `trusted_banks.sql`. The main editor shows a SQL file `models / example / reviews_complaints_delivery.sql` with the following content:

```
1 with
2 trusted_banks as (select * from {{ ref("trusted_banks") }}),
3
```

Below the editor, a command input field contains `dbt run --models example.trusted_bank`. To the right of this field are buttons for `Defer to staging/production` (disabled), `Ready` (active), and a menu icon. Below the command input, a list of recent runs is shown:

Command	Environment	Status	Duration
dbt run --models example.trusted_bank	develop	success	0.0s
dbt run	develop	success	11.4s
dbt run	develop	success	10.6s
dbt run	develop	failed	10.4s
dbt run --select trusted_banks	develop	success	3.1s
dbt run	develop	success	5.2s
dbt run	develop	failed	0.0s
dbt run	develop	failed	0.0s
dbt run	develop	failed	0.0s
dbt run	develop	failed	0.0s
dbt run	develop	failed	0.0s
dbt run	develop	failed	0.0s

On the right side of the interface, a detailed view of the selected run is shown:

dbt run --models example.trusted_bank success

develop
1 minute ago

> System Logs

At the bottom right, there are `Cancel` and `Rerun` buttons. The system clock at the bottom right indicates 04:22.

22.Checar se as o novo db e as tabelas foram criadas como parquet no BigQuery

Google Cloud

My First Project

Pesquise (/) recursos, documentos, produtos e muito mais

Pesquisa

Explorer

+ ADICIONAR

Consultas... ulo

raw_emp... ees

+

-

Pesquisar recursos do BigQuery

Você está visualizando os recursos.

MOstrar APENAS COM ESTRELA

Consultas

Notebooks

Telas de dados

Preparações de dados

Conexões externas

dbt_marquesonorio

raw_banks

raw_complaints

raw_employees

reviews_complaints...

trusted_banks

trusted_complaints

trusted_employees

schema

RESUMO

raw_employees

steel-binder-433800-p9.dbt_marquesonorio

Última modificação

27 de ago. de 2024, 04:15:37 UTC-3

Local dos dados

US

Descrição

Rótulos

Tipo de tabela

table

raw_employees

CONSULTA

COMPARTILHAR

COPIAR

SNAPSHOT

EXCLUIR

EXPORTAR

ATUALIZAR

ESQUEMA

DETALHES

VISUALIZAÇÃO

BUSCADOR DE TABELAS

PRÉ-VISUALIZAÇÃO

INSIGHTS

LINHAGEM

PERFIL DE DADOS

QUALIDADE DOS DADOS

Linha	employer_name	reviews_count	culture_count	salaries_count	benefits_count	employer-website	employer-headquarters	employer-found	employer-industry	employer-revenue
1	China Construction Bank	871	194	720	240	http://www.ccb.com	Pequim, China	null	/Explorar/melhores-serviços-bancários-e-de-crédito-empresas_II.9.40_IIND200048.htm	Mais de US\$ 10 bilhões
2	Mercedes-Benz Mobility	282	90	513	118	https://www.mercedes-benz-m...	Estugarda, Alemanha	1990.0	/Explorar/melhores-serviços-bancários-e-de-crédito-empresas_II.9.40_IIND200048.htm	Mais de US\$ 10 bilhões
3	Santander	17000	5900	29000	5100	http://www.santandercareers.c...	Madrid, Espanha	1856.0	/Explorar/melhores-serviços-bancários-e-de-crédito-empresas_II.9.40_IIND200048.htm	Mais de US\$ 10 bilhões
4	BNP Paribas	13000	4100	20000	3600	http://www.group.bnpparibas	Paris, França	2000.0	/Explorar/melhores-serviços-bancários-e-de-crédito-empresas_II.9.40_IIND200048.htm	Mais de US\$ 10 bilhões
5	J.P. Morgan	859	322	1600	227	http://www.jpmorgan.com	New York, Estados Unidos	1799.0	/Explorar/melhores-gestão-de-ativos-e-investimentos-empresas_II.9.41_IIND200146.htm	Mais de US\$ 10 bilhões
6	JPMorgan Chase & Co	22000	8800	65000	9600	https://www.jpmorganchase.c...	New York, Estados Unidos	1799.0	/Explorar/melhores-serviços-bancários-e-de-crédito-empresas_II.9.40_IIND200048...	Mais de US\$ 10 bilhões

Resultados por página: 50 1 - 34 de 34

Histórico de jobs

ATUALIZAR

04:25

23.Fazer o commit final para a branch de develop e posteriormente para a main

The screenshot displays the dbt Cloud IDE interface. The top navigation bar includes tabs for 'Inglês', 'Spark', 'Pós', 'DynamoDB', and 'Multidimensional'. The main workspace shows a SQL editor with a query for 'trusted_complaints.sql'. A 'Commit Changes' dialog box is overlaid in the center, prompting the user to enter a commit message. The message 'feat: Finalização do pipeline de ingestão de dados' is entered in the text field. The dialog has 'Cancel' and 'Commit Changes' buttons. The background shows the file explorer on the left, the SQL editor in the center, and a DAG (Directed Acyclic Graph) on the right. The bottom status bar indicates the command 'dbt run --models example.trusted_bank' and a 'Ready' status.

dbt Cloud IDE interface showing a commit dialog box. The dialog box prompts the user to enter a commit message for changes to 1 file. The message entered is: `feat: Finalização do pipeline de ingestão de dados`. The background shows the SQL editor with a query for `trusted_complaints.sql` and a DAG (Directed Acyclic Graph) view.

SQL Query (trusted_complaints.sql):

```
1 WITH raw_complaints AS (  
2   SELECT * FROM {{ref('raw_complaints')}}  
3 )  
4  
5 SELECT  
6  
7   CAST('Ano' AS INTEGER) AS ano,  
8   CAST('Trimestre' AS STRING) AS trimestre,  
9   CAST('Categoria' AS STRING) AS categoria,  
10  CAST('Tipo' AS STRING) AS tipo,  
11  CAST('CNPJ IF' AS STRING) AS cnpj_if,  
12  CAST('Instituicao financeira' AS STRING) AS instituicao_financieira,  
13  CAST('Indice' AS STRING) AS indice,  
14  CAST('Quantidade de reclamacoes' AS INTEGER) AS quantidade_reclamacoes,  
15  CAST('Quantidade de reclamacoes n' AS INTEGER) AS quantidade_reclamacoes_n,  
16  CAST('Quantidade total de reclama' AS INTEGER) AS quantidade_total_reclamacoes,  
17  CAST('Quantidade total de cliente' AS INTEGER) AS quantidade_total_clientes,  
18  CAST('Quantidade de clientes - CC' AS INTEGER) AS quantidade_clientes_ccs_scr,  
19  CAST('Quantidade de clientes - SC' AS INTEGER) AS quantidade_clientes_scs_scr,  
20  
21 FROM schema.2021_tri_01
```

DAG (Directed Acyclic Graph) view:

```
graph LR  
    raw_complaints[raw_complaints] --> trusted_complaints[trusted_complaints]  
    trusted_complaints --> reviews_complaints[reviews_complaints]  
    reviews_complaints -- "2+trusted_complaints+2" --> UpdateGraph[Update Graph]
```