

1. DESCRIÇÃO

Este MVP descrito abaixo refere-se a análise de dados climatológicos das estações automáticas do INMET¹ instaladas em todo o território nacional.

2. OBJETIVO

Este MVP tem como objetivo a análise dos dados climatológicos referente as estações meteorológicas automáticas do INMET presentes em diversas regiões do país visando responder as seguintes questões:

- Qual a quantidade total de estações meteorológicas?
- Qual a distribuição das estações meteorológicas por estado?
- Qual o estado com o maior número de estações meteorológicas?
- Qual o estado com o menor número de estações meteorológicas?
- Qual o status das estações meteorológicas?
- Qual a estação operante a mais tempo?

3. COLETA DOS DADOS

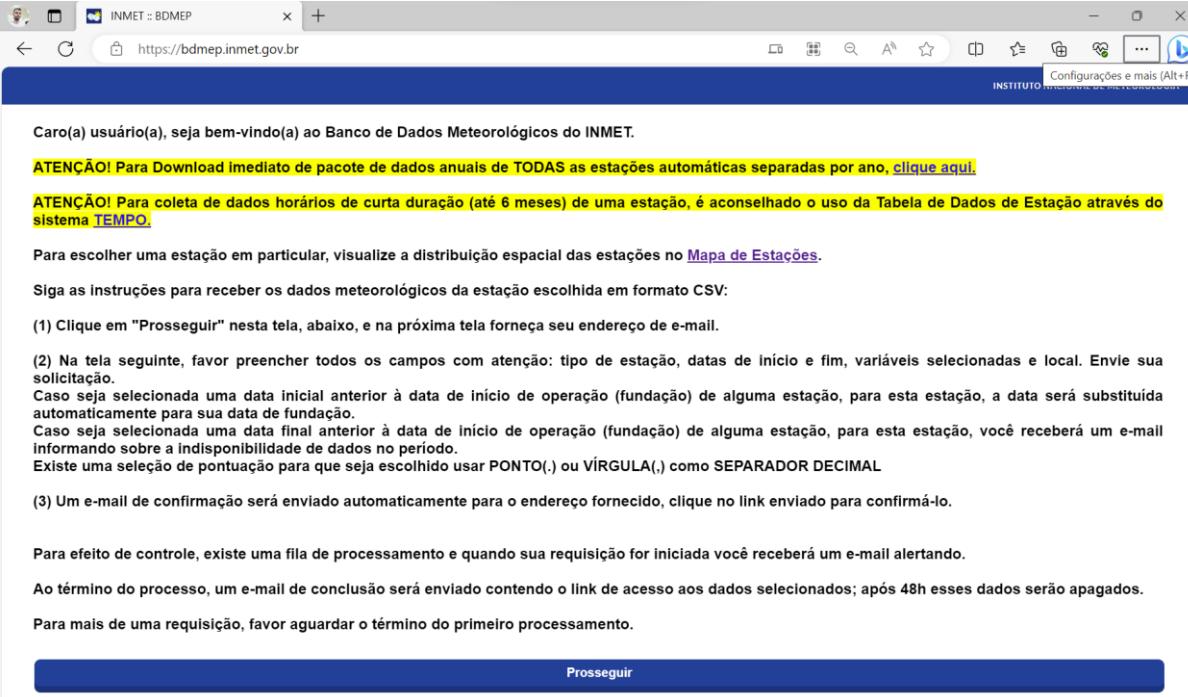
3.1. EXTRAÇÃO DAS INFORMAÇÕES

Foram utilizadas duas fontes de informações para extração da base de dados deste MVP:

- FONTE 01: <https://bdmep.inmet.gov.br/> (base de dados das estações meteorológicas)
- FONTE 02: <https://portal.inmet.gov.br/paginas/catalogoaut> (catálogo das estações meteorológicas)

3.1.1. FONTE 01 - BASE DE DADOS ESTAÇÕES METEOROLÓGICAS

¹ Instituto Nacional de Meteorologia



Caro(a) usuário(a), seja bem-vindo(a) ao Banco de Dados Meteorológicos do INMET.

ATENÇÃO! Para Download imediato de pacote de dados anuais de TODAS as estações automáticas separadas por ano, [clique aqui](#).

ATENÇÃO! Para coleta de dados horários de curta duração (até 6 meses) de uma estação, é aconselhado o uso da Tabela de Dados de Estação através do sistema [TEMPO](#).

Para escolher uma estação em particular, visualize a distribuição espacial das estações no [Mapa de Estações](#).

Siga as instruções para receber os dados meteorológicos da estação escolhida em formato CSV:

- (1) Clique em "Prosseguir" nesta tela, abaixo, e na próxima tela forneça seu endereço de e-mail.
- (2) Na tela seguinte, favor preencher todos os campos com atenção: tipo de estação, datas de início e fim, variáveis selecionadas e local. Envie sua solicitação.
- Caso seja selecionada uma data inicial anterior à data de início de operação (fundação) de alguma estação, para esta estação, a data será substituída automaticamente para sua data de fundação.
- Caso seja selecionada uma data final anterior à data de início de operação (fundação) de alguma estação, para esta estação, você receberá um e-mail informando sobre a indisponibilidade de dados no período.
- Existe uma seleção de pontuação para que seja escolhido usar PONTO(.) ou VÍRGULA(,) como SEPARADOR DECIMAL
- (3) Um e-mail de confirmação será enviado automaticamente para o endereço fornecido, clique no link enviado para confirmá-lo.

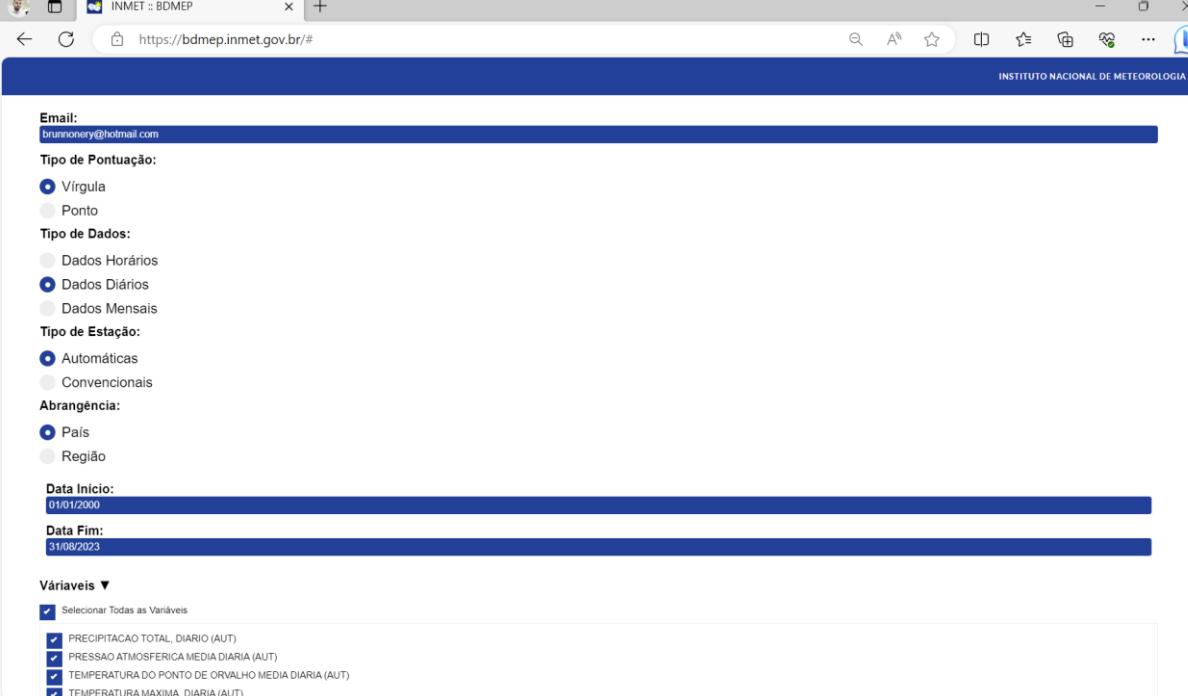
Para efeito de controle, existe uma fila de processamento e quando sua requisição for iniciada você receberá um e-mail alertando.

Ao término do processo, um e-mail de conclusão será enviado contendo o link de acesso aos dados selecionados; após 48h esses dados serão apagados.

Para mais de uma requisição, favor aguardar o término do primeiro processamento.

[Prosseguir](#)

Figura 1 – tela inicial para solicitação da base de dados das estações meteorológicas.



Email:

Tipo de Pontuação: Vírgula Ponto

Tipo de Dados: Dados Horários Dados Diários Dados Mensais

Tipo de Estação: Automáticas Convencionais

Abrangência: País Região

Data Início:

Data Fim:

Variáveis ▼

Selecionar Todas as Variáveis

PRECIPITACAO TOTAL, DIARIO (AUT)

PRESSAO ATMOSFERICA MEDIA DIARIA (AUT)

TEMPERATURA DO PONTO DE ORVALHO MEDIA DIARIA (AUT)

TEMPERATURA MAXIMA, DIARIA (AUT)

Figura 2 – definição e detalhamento do conjunto de dados das estações para extração das informações.

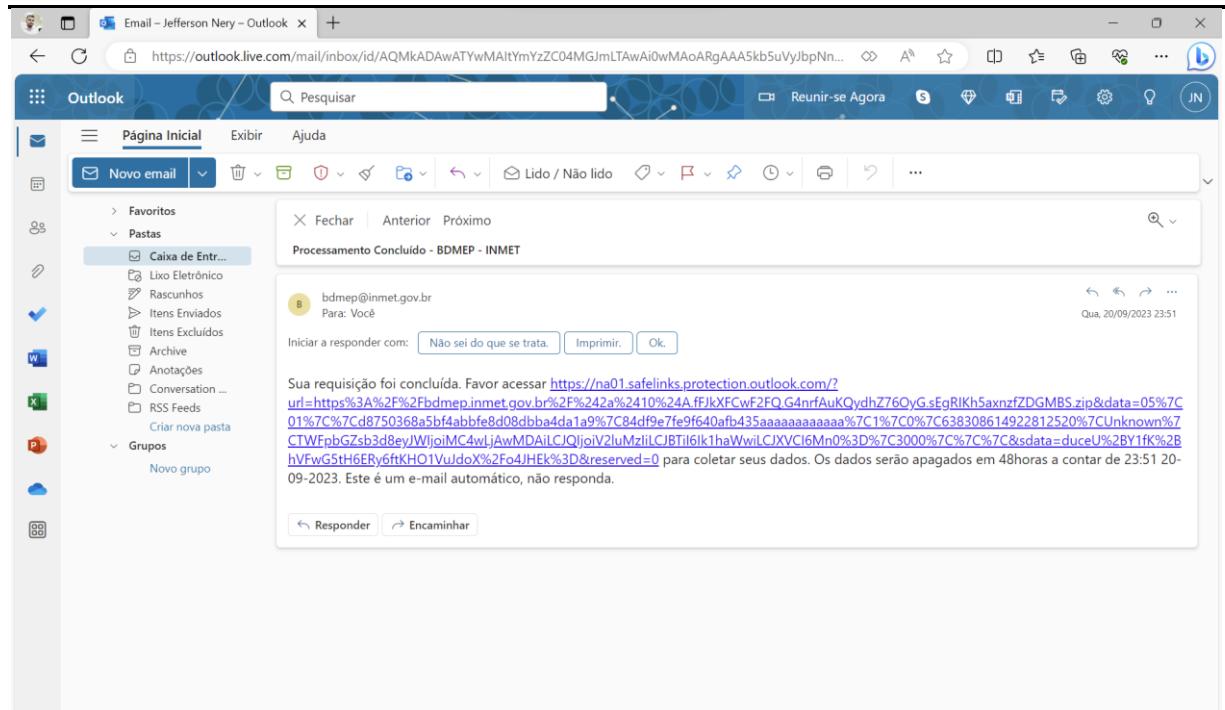


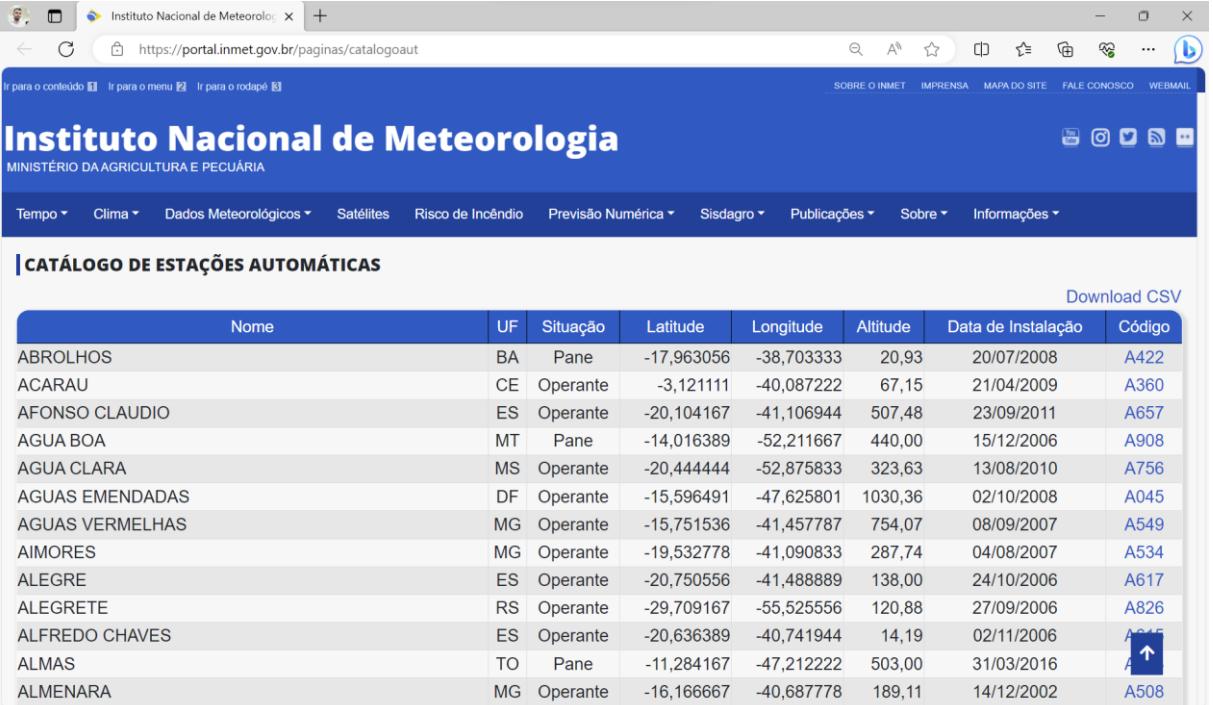
Figura 3 – recebimento do link para o download da base de dados.

Jefferson - Pessoal > PÓS-GRADUAÇÃO > SPRINT - Engenharia de dados > MVP > Base de dados > Dados - Estações Meteorológicas					Pesquisar em Dad...
Nome	Status	Data de modificação	Tipo	Tamanho	
dados_A001_D_2000-05-06_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	629 KB	
dados_A002_D_2001-05-28_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	596 KB	
dados_A003_D_2001-05-24_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	584 KB	
dados_A004_D_2001-05-30_2018-06-20.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	451 KB	
dados_A005_D_2001-06-03_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	529 KB	
dados_A006_D_2001-06-07_2006-09-24.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	128 KB	
dados_A007_D_2003-02-19_2005-08-15.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	64 KB	
dados_A008_D_2003-06-12_2013-02-19.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	263 KB	
dados_A009_D_2004-12-16_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	502 KB	
dados_A010_D_2005-03-03_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	497 KB	
dados_A011_D_2006-07-14_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	451 KB	
dados_A012_D_2006-10-20_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	456 KB	
dados_A013_D_2007-07-12_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	423 KB	
dados_A014_D_2007-07-17_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	433 KB	
dados_A015_D_2007-02-02_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	443 KB	
dados_A016_D_2007-05-22_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	437 KB	
dados_A017_D_2007-04-17_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	435 KB	
dados_A018_D_2006-11-30_2023-08-31.csv	Cloud	21/09/2023 11:46	Arquivo de Valores S...	438 KB	

Figura 4 – base de dados extraída em formato CSV².

² Comma-separated values (valores separados por vírgula)

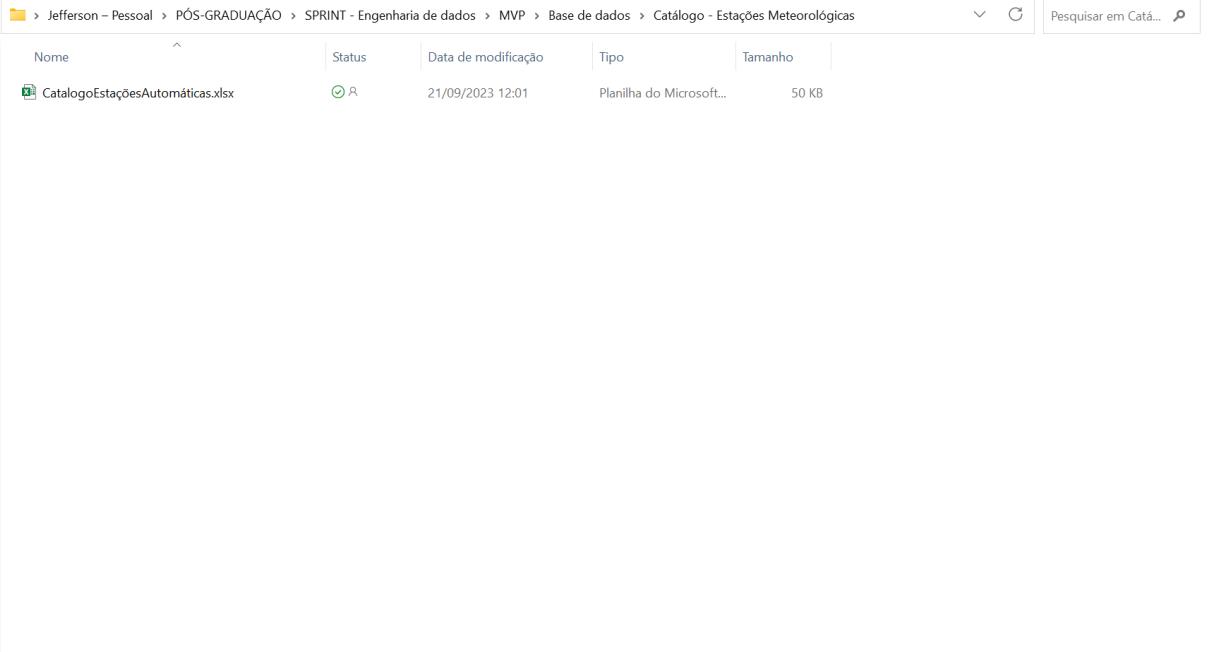
3.1.2. FONTE 02 – CATÁLOGO DAS ESTAÇÕES METEOROLÓGICAS



The screenshot shows a web browser window for the Instituto Nacional de Meteorologia (INMET) at <https://portal.inmet.gov.br/paginas/catalogoaut>. The page title is "Instituto Nacional de Meteorologia". Below it, a sub-header reads "CATÁLOGO DE ESTAÇÕES AUTOMÁTICAS". A table lists 18 automatic stations with their details: Nome (Name), UF (State/Region), Situação (Status), Latitude, Longitude, Altitude, Data de Instalação (Installation Date), and Código (Code). The table includes a "Download CSV" link in the top right corner.

Nome	UF	Situação	Latitude	Longitude	Altitude	Data de Instalação	Código
ABROLHOS	BA	Pane	-17,963056	-38,703333	20,93	20/07/2008	A422
ACARAU	CE	Operante	-3,121111	-40,087222	67,15	21/04/2009	A360
AFONSO CLAUDIO	ES	Operante	-20,104167	-41,106944	507,48	23/09/2011	A657
AGUA BOA	MT	Pane	-14,016389	-52,211667	440,00	15/12/2006	A908
AGUA CLARA	MS	Operante	-20,444444	-52,875833	323,63	13/08/2010	A756
AGUAS EMENDADAS	DF	Operante	-15,596491	-47,625801	1030,36	02/10/2008	A045
AGUAS VERMELHAS	MG	Operante	-15,751536	-41,457787	754,07	08/09/2007	A549
AIMORES	MG	Operante	-19,532778	-41,090833	287,74	04/08/2007	A534
ALEGRE	ES	Operante	-20,750556	-41,488889	138,00	24/10/2006	A617
ALEGRETE	RS	Operante	-29,709167	-55,525556	120,88	27/09/2006	A826
ALFREDO CHAVES	ES	Operante	-20,636389	-40,741944	14,19	02/11/2006	A645
ALMAS	TO	Pane	-11,284167	-47,212222	503,00	31/03/2016	A508
ALMENARA	MG	Operante	-16,166667	-40,687778	189,11	14/12/2002	A508

Figura 5 – tela inicial para extração das informações das estações meteorológicas.



The screenshot shows a Microsoft Excel spreadsheet titled "CatalogoEstaçõesAutomáticas.xlsx". The sheet displays a table of data with columns: Nome (Name), Status, Data de modificação (Modification Date), Tipo (Type), and Tamanho (Size). The data corresponds to the stations listed in Figure 5. The file is identified as a "Planilha do Microsoft..." (Microsoft spreadsheet) with a size of 50 KB.

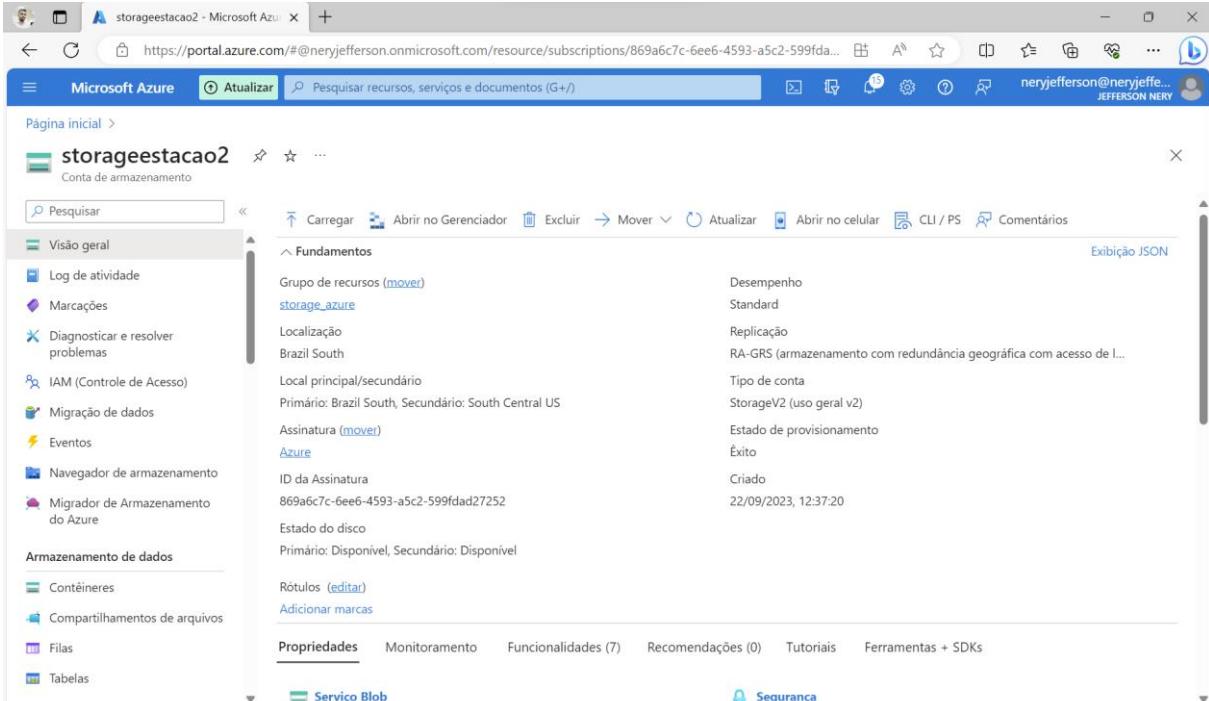
Nome	Status	Data de modificação	Tipo	Tamanho
CatalogoEstaçõesAutomáticas.xlsx	OK	21/09/2023 12:01	Planilha do Microsoft...	50 KB

Figura 6 – base de dados extraída em formato CSV.

4. CARGA DOS DADOS

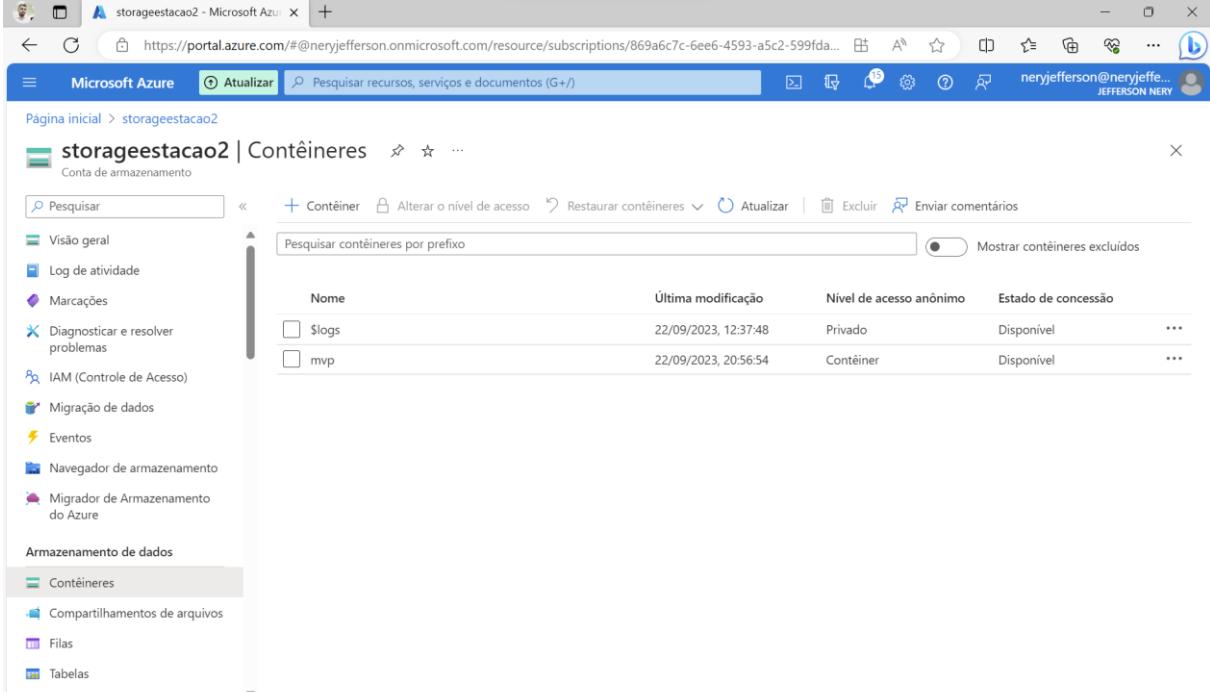
Foi criado uma conta de armazenamento blobs da Azure (storageestacao2) para armazenamento dos dados não estruturados nos contêineres na nuvem.

4.1. CRIAÇÃO DA CONTA DE ARMAZENAMENTO



The screenshot shows the Microsoft Azure portal interface for the storage account 'storageestacao2'. The left sidebar contains navigation links like 'Visão geral', 'Log de atividade', 'Marcções', 'Diagnosticar e resolver problemas', 'IAM (Controle de Acesso)', 'Migração de dados', 'Eventos', 'Navegador de armazenamento', and 'Migrador de Armazenamento do Azure'. The main content area displays detailed information about the storage account, including its name ('storageestacao2'), resource group ('storage_azure'), location ('Brazil South'), primary and secondary locations ('Brazil South, South Central US'), and access type ('StorageV2'). It also shows the creation date ('22/09/2023, 12:37:20'). The bottom navigation bar includes tabs for 'Propriedades', 'Monitoramento', 'Funcionalidades (7)', 'Recomendações (0)', 'Tutorial', and 'Ferramentas + SDKs'.

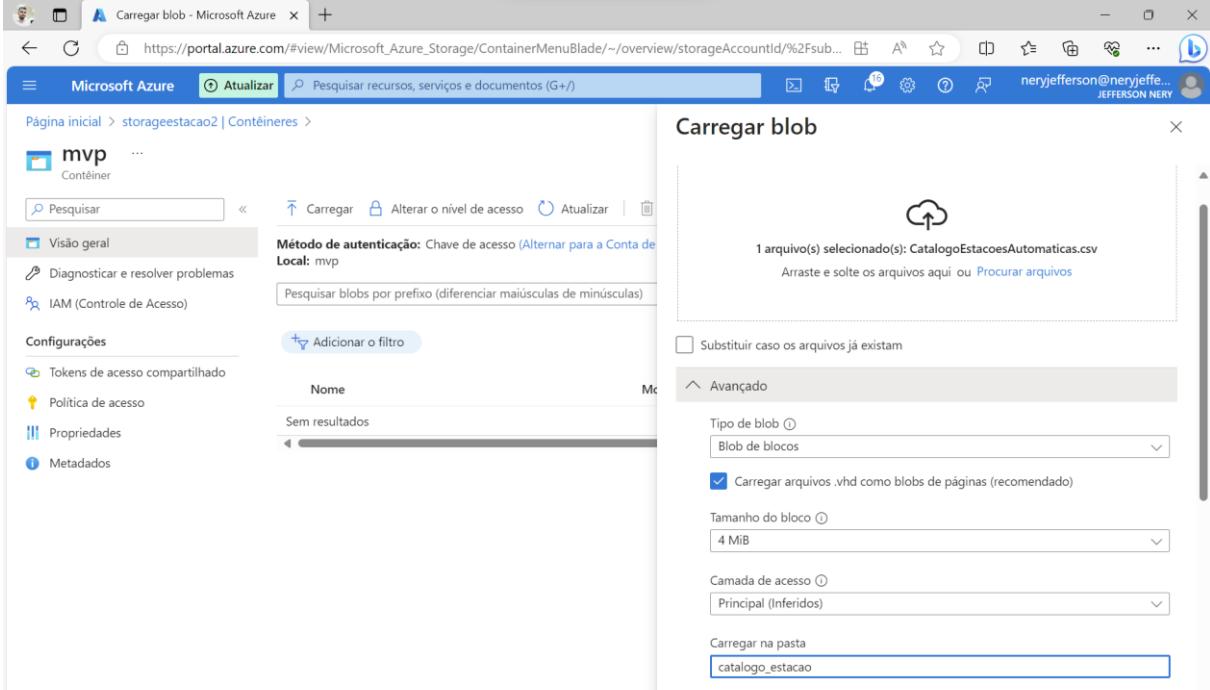
Figura 7 – criação da conta de armazenamento "storageestacao2".



The screenshot shows the Microsoft Azure Storage blade for the 'storageestacao2' account. On the left, there's a sidebar with options like Visão geral, Log de atividade, Marcações, etc. The main area shows a table of containers. One row is selected, showing details for the 'mvp' container: Nome (mvp), Última modificação (22/09/2023, 20:56:54), Nível de acesso anônimo (Contêiner), and Estado de concessão (Disponível). There are three more rows in the table: '\$logs' (Privado) and two others (Contêiner, Disponível).

Figura 8 – criação do contêiner “mvp” para carga das fontes de dados 1 e 2 referente as estações meteorológicas.

4.2. CARGA DOS ARQUIVOS PARA O BLOB



The screenshot shows the 'Carregar blob' (Upload blob) dialog in the Microsoft Azure Storage blade. It's set to upload 'CatalogoEstacoesAutomaticas.csv' to the 'mvp' container. The 'Método de autenticação' (Authentication method) is set to 'Chave de acesso' (Access key) and 'Local' (Location) is set to 'mvp'. The 'Carregar' (Upload) button is highlighted. The right side of the dialog shows advanced settings: 'Tipo de blob' (Block blob), 'Carregar arquivos .vhd como blobs de páginas (recomendado)' (Load .vhd files as page blobs (recommended)), 'Tamanho do bloco' (Block size) set to '4 MiB', 'Camada de acesso' (Access tier) set to 'Principal (Inferidos)' (Primary (Inferred)), and 'Carregar na pasta' (Upload to folder) set to 'catalogo_estacao'. There are also checkboxes for 'Substituir caso os arquivos já existam' (Replace if files already exist) and 'Avançado' (Advanced).

Figura 9 – carga do arquivo “CatalogoEstacoesAutomaticas.csv”.

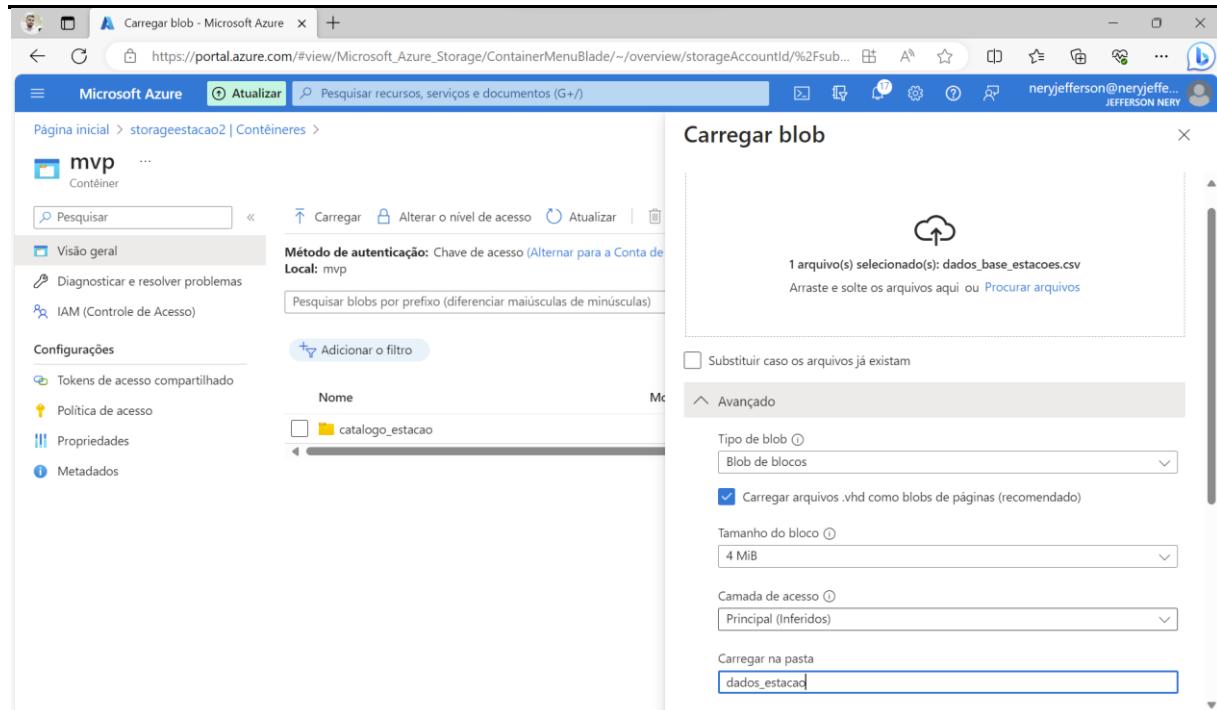
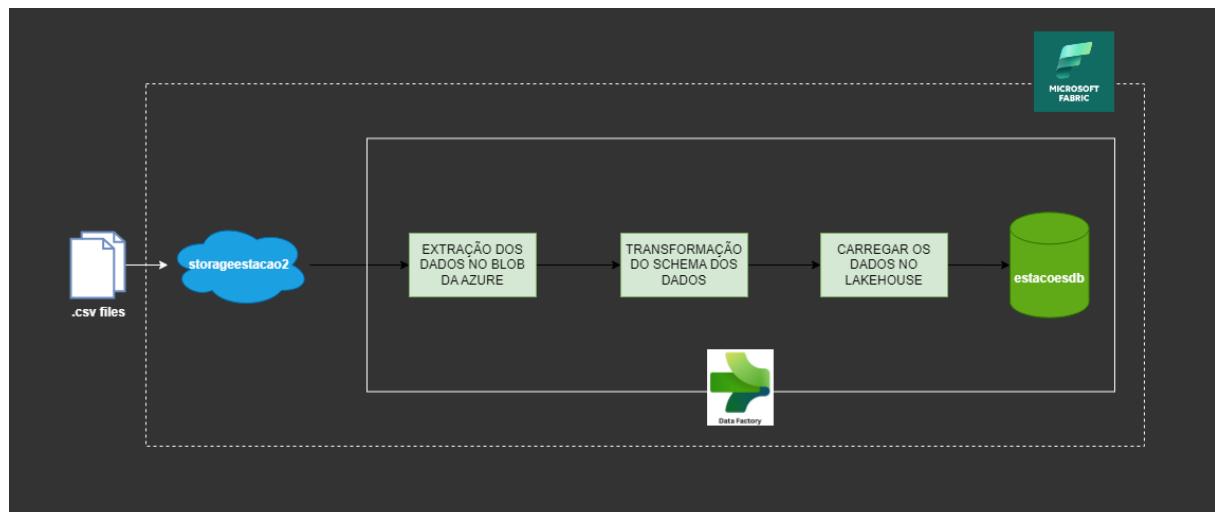


Figura 10 – carga do arquivo “dados_base_estacoes.csv”.

5. MODELAGEM DOS DADOS

Foi utilizada a recém-lançada plataforma da Microsoft Fabric que oferece um conjunto de serviços incluindo datalake, engenharia de dados e integração de dados, tudo em um só lugar. Dentro desta plataforma foi utilizado um conjunto de ferramentas do Data Factory para o ETL conforme demonstra o diagrama abaixo:



5.1. CRIAÇÃO DO PIPELINE

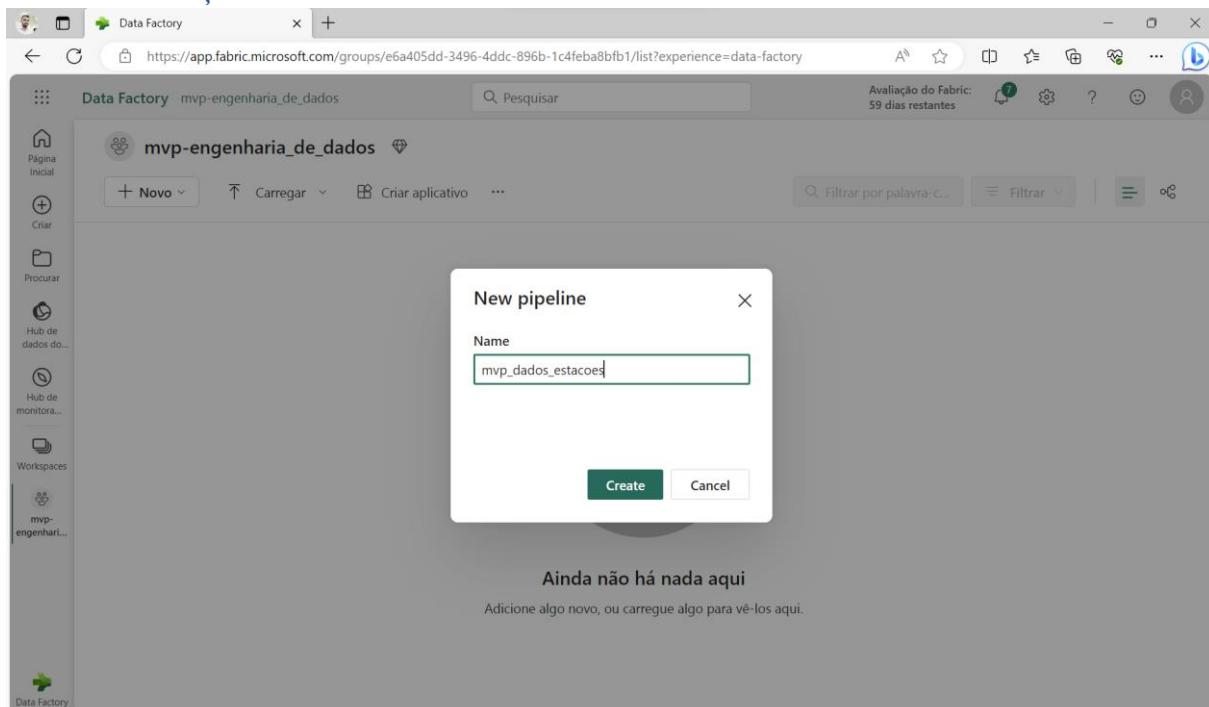


Figura 11 – criação do pipeline para realizar a extração dos dados no blob “storageestacoes2” da Azure.

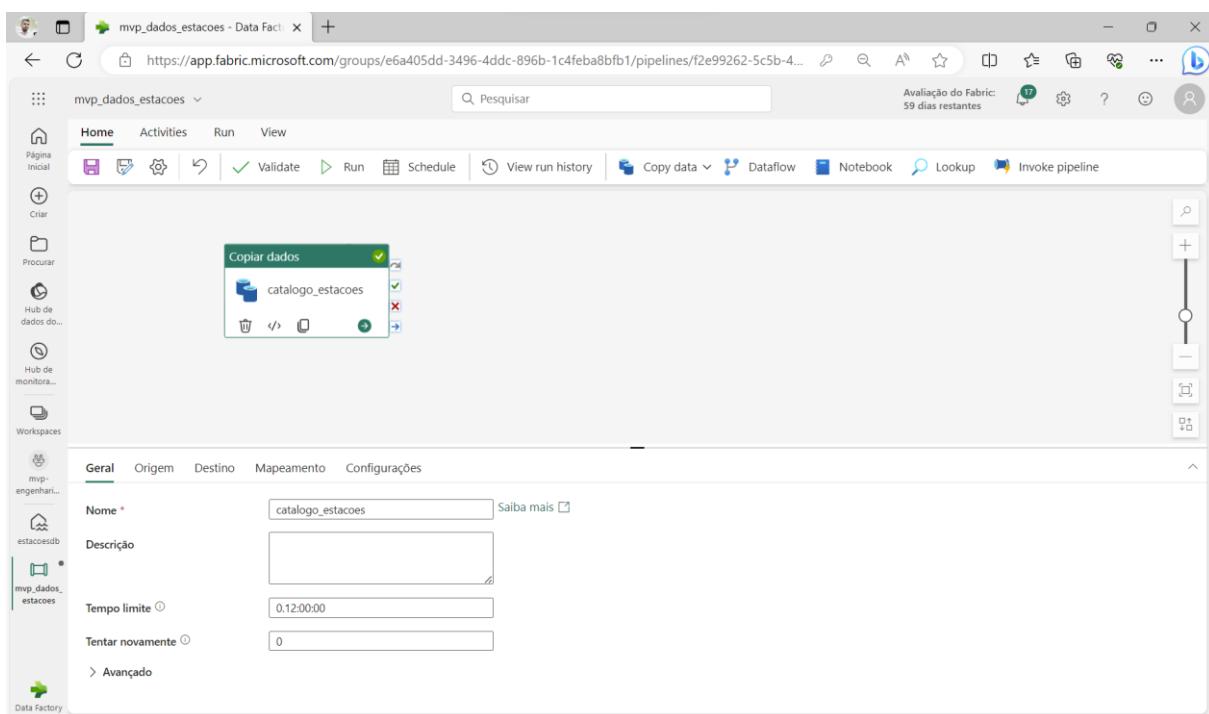
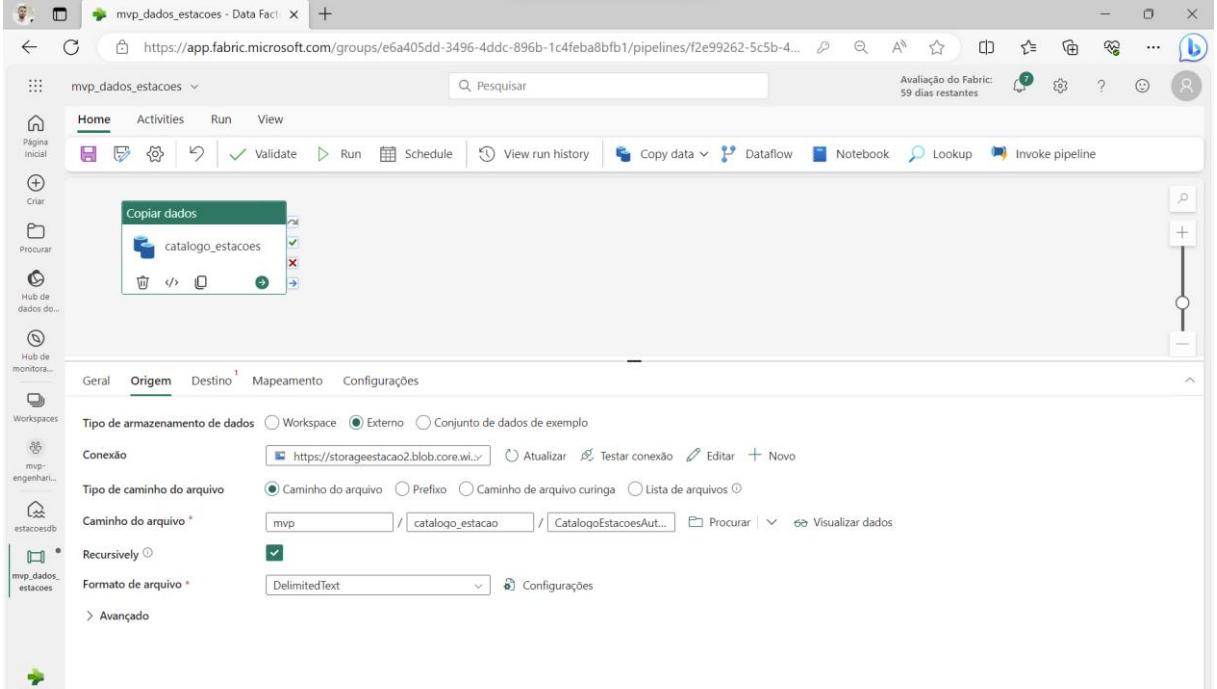


Figura 12 – definindo o nome da rotina acerca do catálogos das estações.



Copiar dados

catalogo_estacoes

Origem

Destino

Mapeamento

Configurações

Tipo de armazenamento de dados: Externo

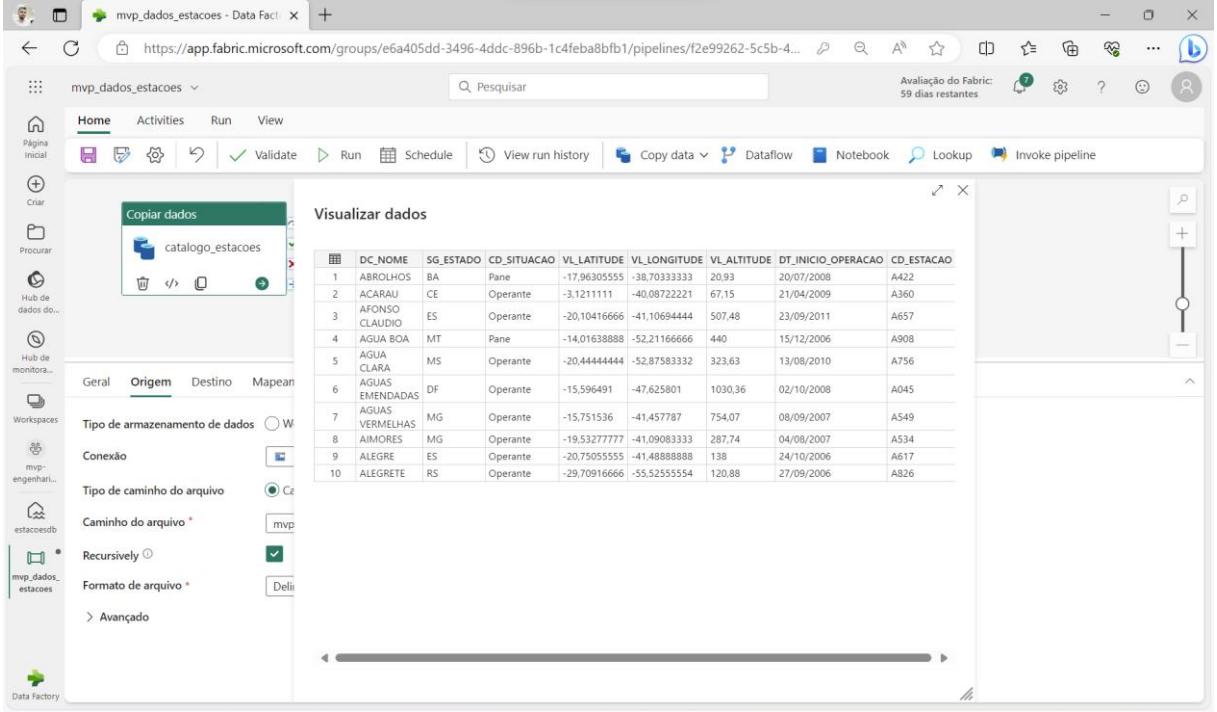
Conexão: https://storageestacao2.blob.core.windows.net

Tipo de caminho do arquivo: Caminho do arquivo (Recursive)

Caminho do arquivo: mvp / catalogo_estacao / CatalogoEstacoesAut...

Formato de arquivo: DelimitedText

Figura 13 – estabelecendo conexão com o armazenamento blob da Azure.



Visualizar dados

DC_NOME	SG_ESTADO	CD_SITUACAO	VL_LATITUDE	VL_LONGITUDE	VL_ALTITUDE	DT_INICIO_OPERACAO	CD_ESTACAO
ABROLHOS	BA	Pane	-17.9630555	-38.0333333	20.93	20/07/2008	A422
ACARAU	CE	Operante	-3.1211111	-40.08722221	67.15	21/04/2009	A360
AFONSO CLAUDIO	ES	Operante	-20.10416666	-41.10694444	507.48	23/09/2011	A657
AGUA BOA	MT	Pane	-14.01638888	-52.21166666	440	15/12/2006	A908
AGUA CLARA	MS	Operante	-20.44444444	-52.87583332	323.63	13/08/2010	A756
AGUAS EMENDADAS	DF	Operante	-15.596491	-47.625801	1030.36	02/10/2008	A045
AGUAS VERMELHAS	MG	Operante	-15.751536	-41.457787	754.07	08/09/2007	A549
AIMORES	ES	Operante	-19.53277777	-41.90903333	287.74	04/08/2007	A534
ALEGRETE	RS	Operante	-29.70916666	-55.52555554	138	24/10/2006	A617
					120.88	27/09/2006	A826

Figura 14 – pré-visualização dos dados.

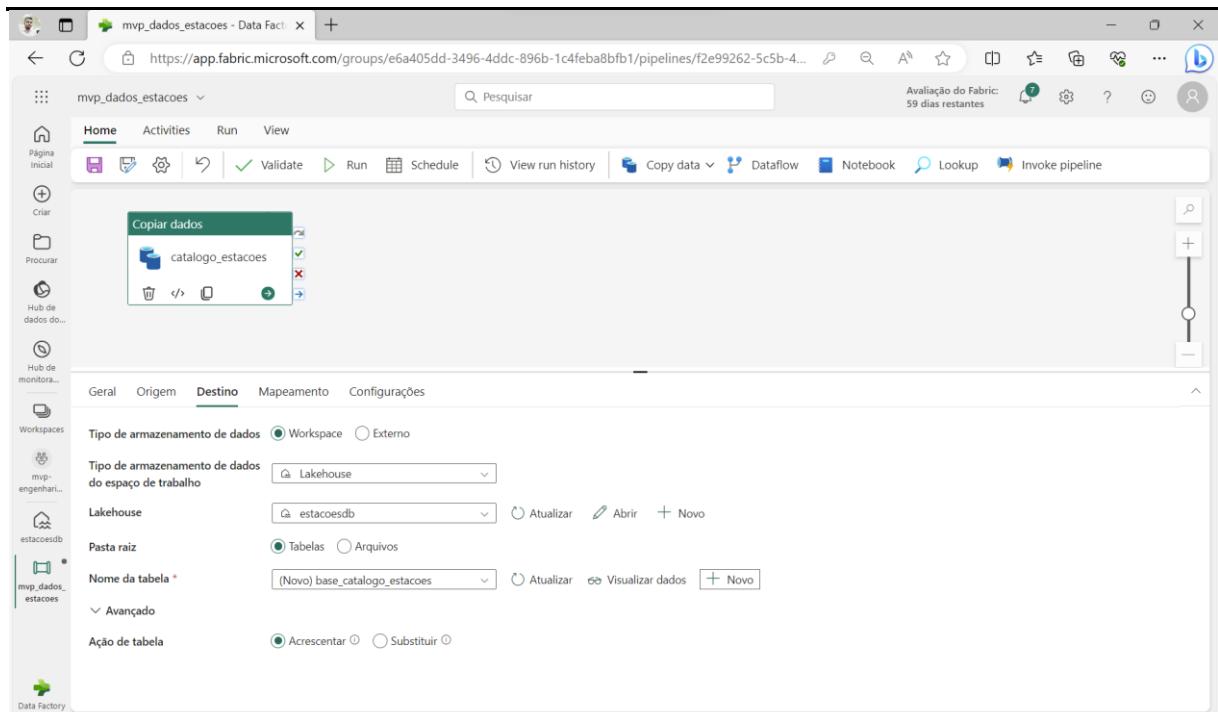


Figura 15 – definindo o destino dos dados, neste caso criamos um Lakehouse chamado “estacoesdb” da plataforma Microsoft Fabric e atribuímos o nome da tabela “base_catalogo_estacoes” que será salva.

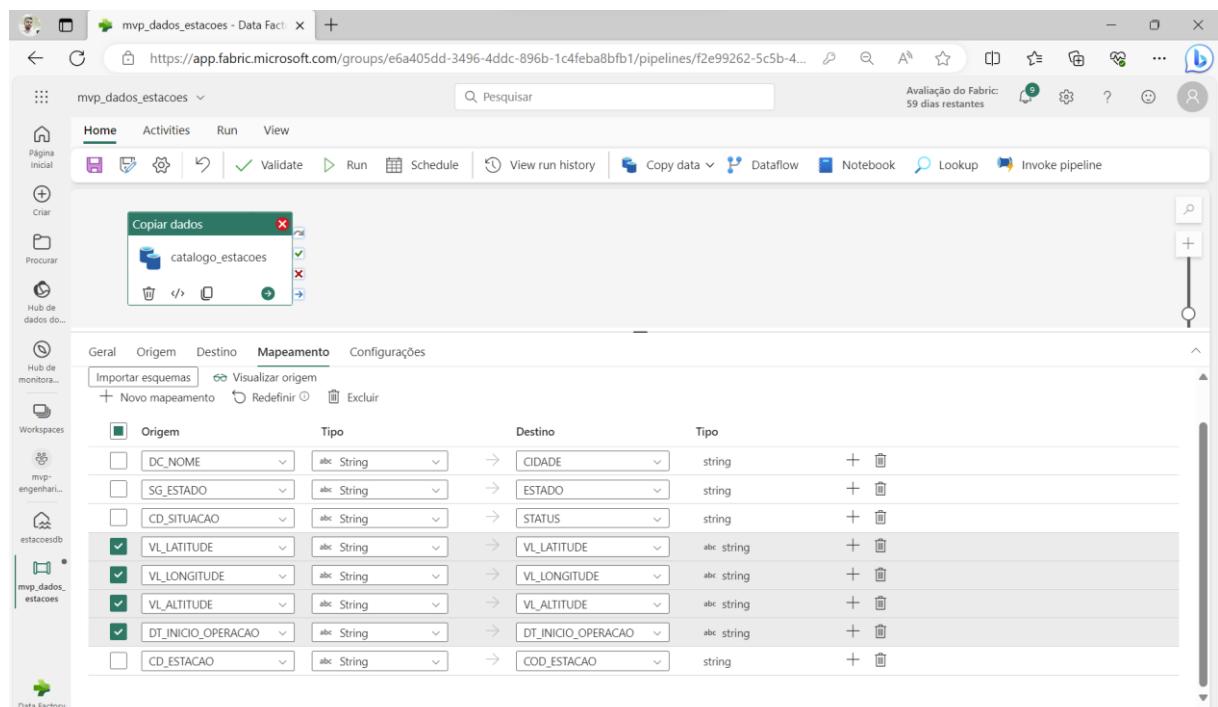


Figura 16 – realizado a importação do schemas, fizemos a alteração do nome das colunas e os tipos de valores esperados (string, int e outros) que as mesmas devem ter, algumas colunas foram selecionadas para exclusão pois não serão necessárias para análise das informações.

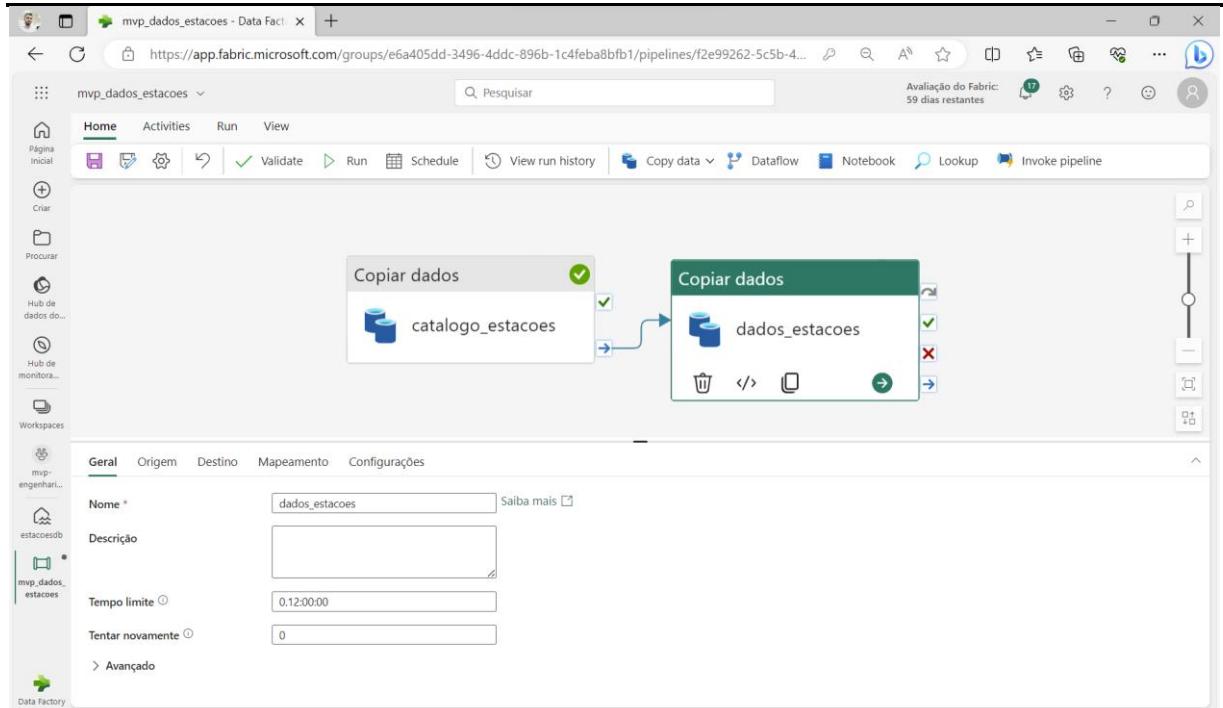


Figura 17 – foi inserido uma segunda atividade em nosso pipeline que se refere a nossa base de dados das estações.

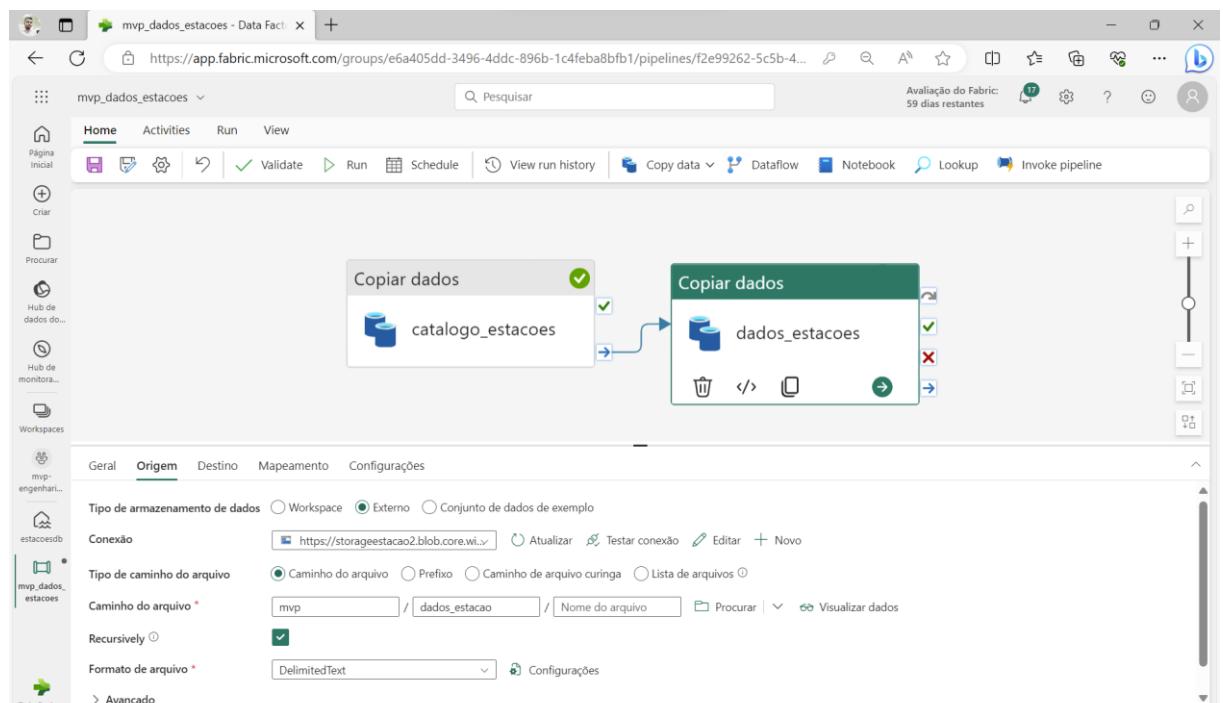
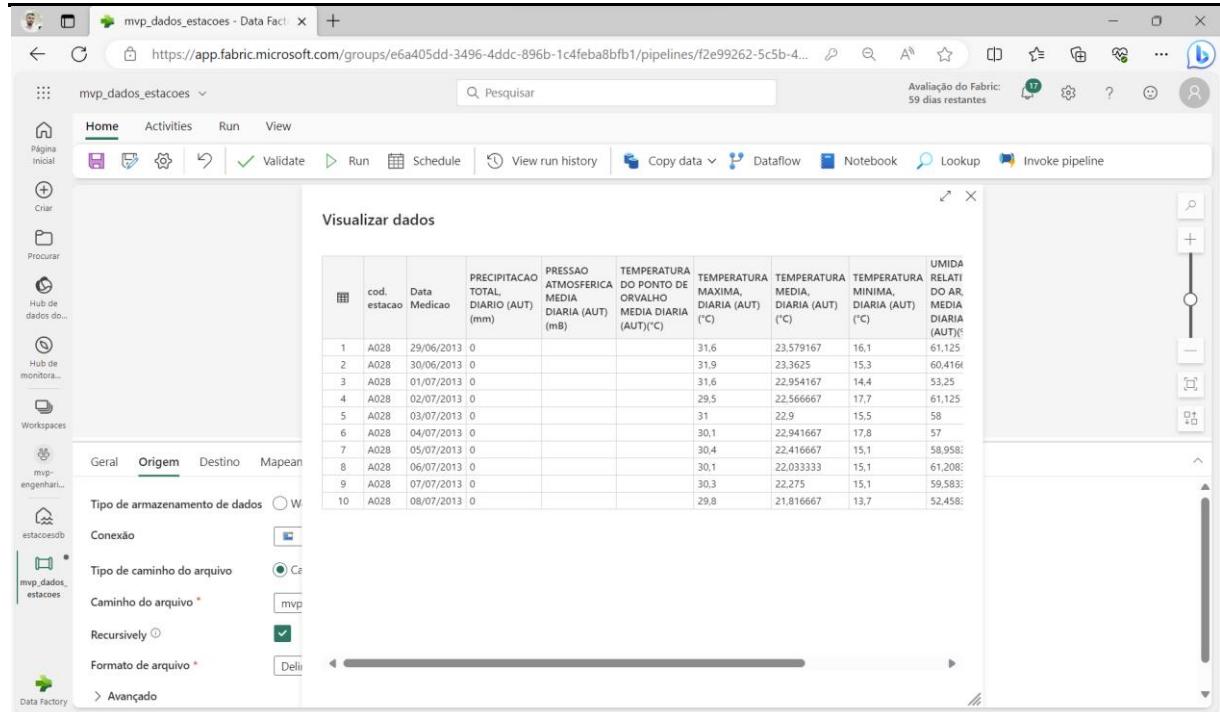


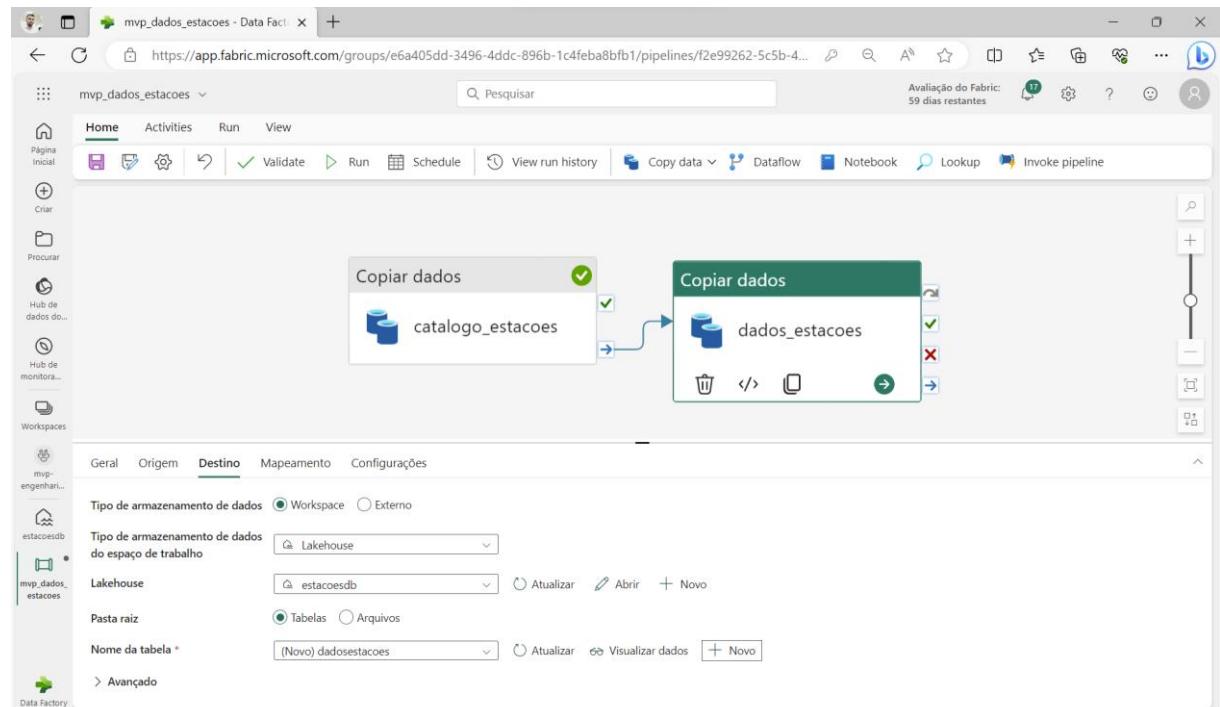
Figura 18 – estabelecendo conexão com o armazenamento blob da Azure.



The screenshot shows the Microsoft Data Factory interface. On the left, there's a sidebar with various workspace options like 'Página Inicial', 'Criar', 'Procurar', 'Hub de dados do...', 'Hub de monitora...', 'Workspaces', and 'Data Factory'. The main area is titled 'mvp_dados_estacoes' and shows a 'Visualizar dados' (Preview data) window. This window contains a table with 10 rows of data. The columns are labeled: cod. estacao, Data Medicao, PRECIPITACAO TOTAL DIARIO (AUT) (mm), PRESSAO ATMOSFERICA MEDIA DIARIA (mB), TEMPERATURA DO PONTO DE ORVALHO MEDIA DIARIA (AUT)(°C), TEMPERATURA MAXIMA, DIARIA (AUT) (°C), TEMPERATURA MEDIA, DIARIA (AUT) (°C), TEMPERATURA MINIMA, DIARIA (AUT) (°C), and UMIDA RELATI DO AR, MEDIA DIARIA (%). The data represents weather measurements for station A028 from June 29 to July 8, 2013.

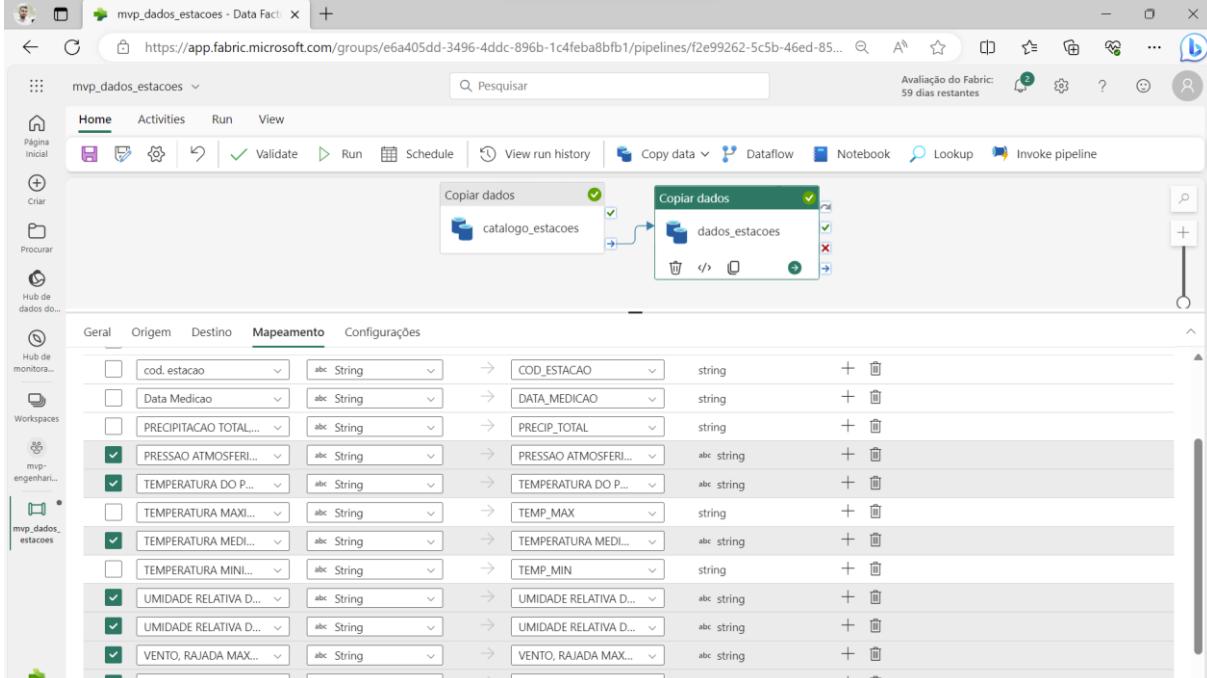
	cod. estacao	Data Medicao	PRECIPITACAO TOTAL DIARIO (AUT) (mm)	PRESSAO ATMOSFERICA MEDIA DIARIA (mB)	TEMPERATURA DO PONTO DE ORVALHO MEDIA DIARIA (AUT)(°C)	TEMPERATURA MAXIMA, DIARIA (AUT) (°C)	TEMPERATURA MEDIA, DIARIA (AUT) (°C)	TEMPERATURA MINIMA, DIARIA (AUT) (°C)	UMIDA RELATI DO AR, MEDIA DIARIA (%)
1	A028	29/06/2013	0		31,6	23,579167	16,1	61,125	
2	A028	30/06/2013	0		31,9	23,3625	15,3	60,416	
3	A028	01/07/2013	0		31,6	22,954167	14,4	53,25	
4	A028	02/07/2013	0		29,5	22,566667	17,7	61,125	
5	A028	03/07/2013	0		31	22,9	15,5	58	
6	A028	04/07/2013	0		30,1	22,941667	17,8	57	
7	A028	05/07/2013	0		30,4	22,416667	15,1	58,958	
8	A028	06/07/2013	0		30,1	22,033333	15,1	61,208	
9	A028	07/07/2013	0		30,3	22,275	15,1	59,583	
10	A028	08/07/2013	0		29,8	21,816667	13,7	52,458	

Figura 19 – pré-visualização dos dados.



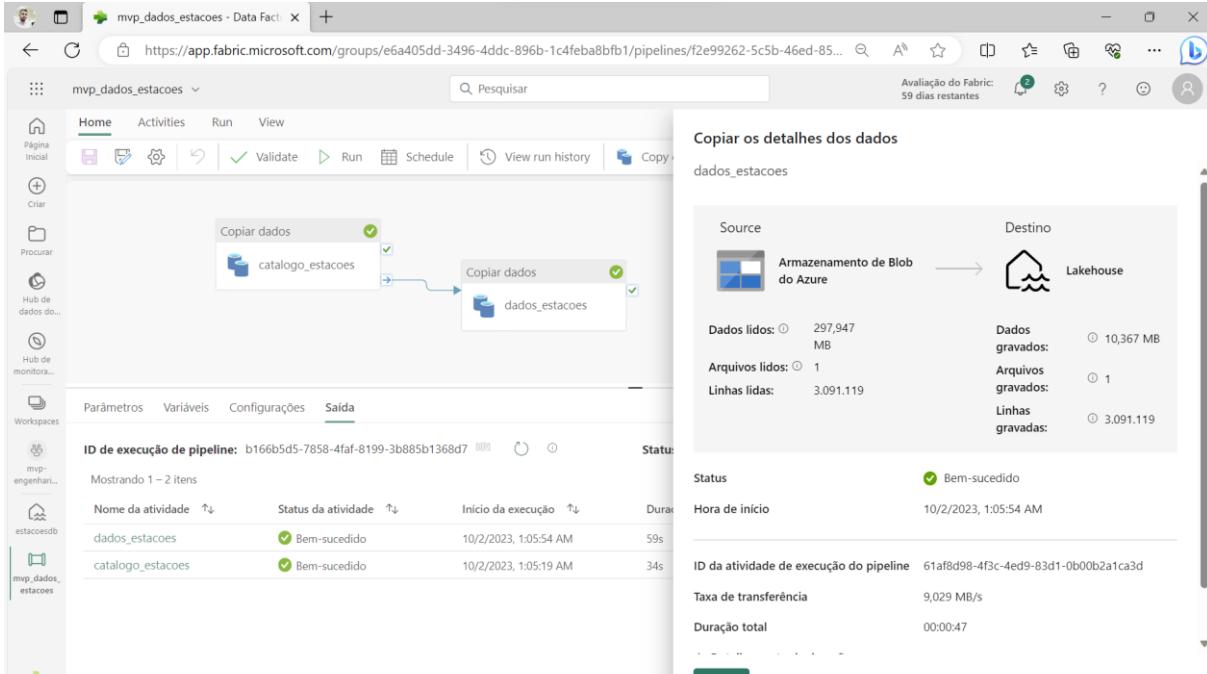
The screenshot shows the Microsoft Data Factory interface with a 'Copiar dados' (Copy data) step configuration window open. The 'Destino' tab is selected. The 'Tipo de armazenamento de dados' (Destination data storage type) is set to 'Workspace'. The 'Tipo de armazenamento de dados do espaço de trabalho' (Type of data storage in the workspace) is set to 'Lakehouse'. The 'Lakehouse' dropdown shows 'estacessdb'. The 'Nome da tabela' (Table name) is set to '(Novo) dadesestacoes'. There are two validation status boxes: one for 'catalogo_estacoes' which is green with a checkmark, and another for 'dados_estacoes' which is red with a cross.

Figura 20 – definindo o destino dos dados, neste caso utilizamos o Lakehouse criado anteriormente e atribuímos o nome da tabela “dadosestacoes” que será salva.



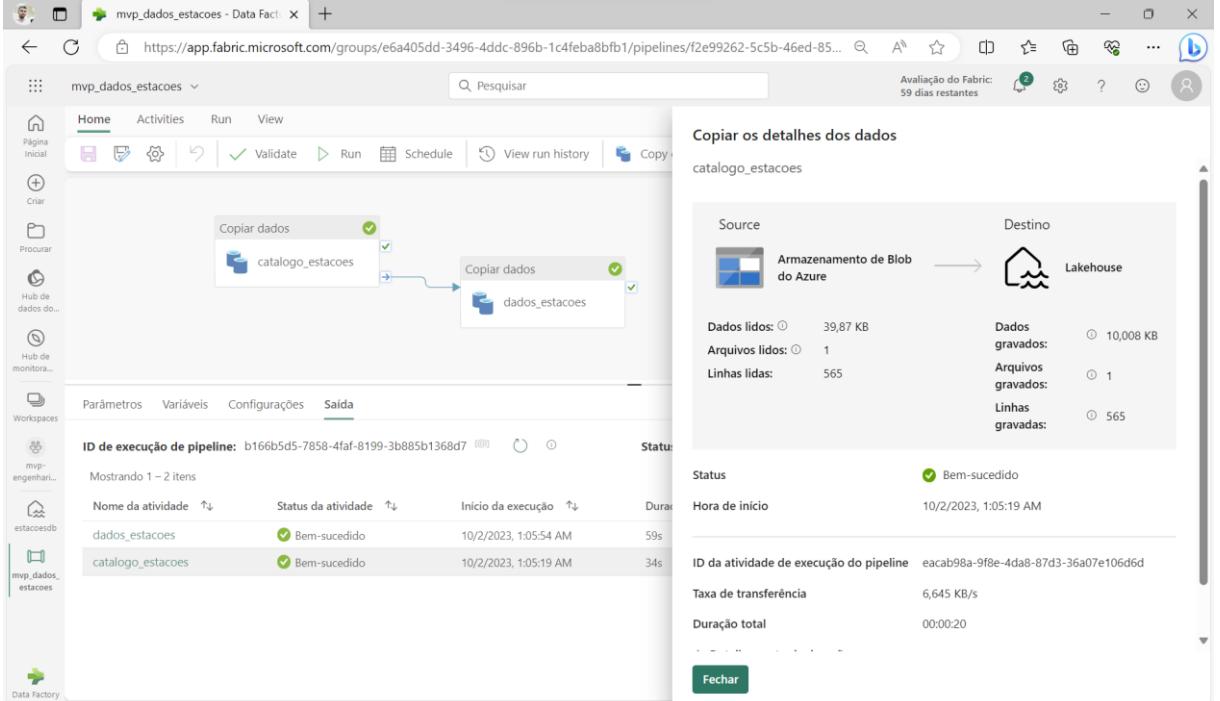
The screenshot shows the Microsoft Data Factory interface for a pipeline named 'mvp_dados_estacoes'. On the left sidebar, under 'mvp_dados_estacoes', the 'Data Factory' icon is selected. The main area displays a pipeline diagram with two 'Copy data' activities. The first activity copies from 'catalogo_estacoes' to a temporary location, and the second activity copies from that temporary location to 'dados_estacoes'. Below the diagram, the 'Mapeamento' (Mapping) tab is active, showing a list of columns from the source and their corresponding destination names and data types. Some columns have checkboxes next to them, indicating they are selected for mapping.

Figura 21 – realizamos a importação do schemas, fizemos a alteração do nome das colunas e definidos os tipos de valores esperados (string, int e outros) que as mesmas devem ter, algumas colunas foram selecionadas para exclusão pois não serão necessárias para análise das informações.



The screenshot shows the Microsoft Data Factory interface for the same pipeline. The left sidebar shows the 'mvp_dados_estacoes' workspace. The main area shows the pipeline diagram with the two 'Copy data' activities. The right side of the screen displays the execution details for the second 'Copy data' activity, titled 'Copiar os detalhes dos dados' for 'dados_estacoes'. It shows the 'Source' as 'Armazenamento de Blob do Azure' and the 'Destino' as 'Lakehouse'. Summary statistics are provided: Dados lidos: 297,947 MB, Arquivos lidos: 1, Linhas lidas: 3.091.119; and Dados gravados: 10,367 MB, Arquivos gravados: 1, Linhas gravadas: 3.091.119. The status is listed as 'Bem-sucedido' (Successful). The execution time is shown as 10/2/2023, 1:05:54 AM, and the total duration is 00:00:47.

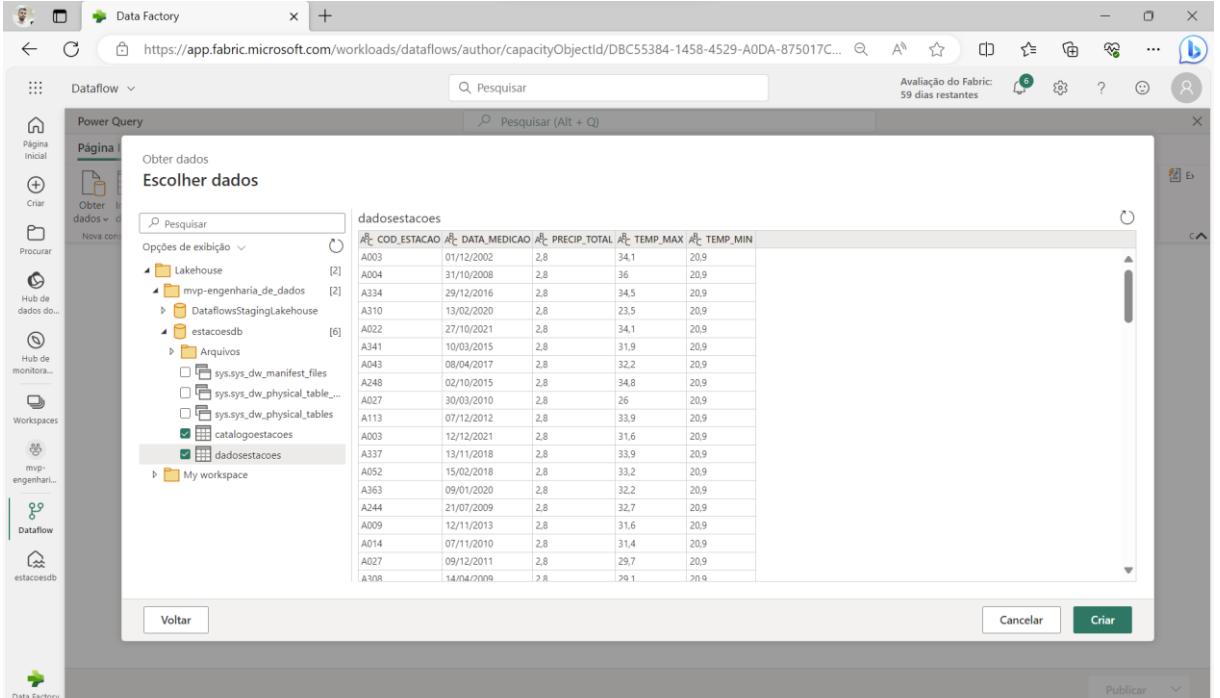
Figura 22 – pipeline com os dados das estações finalizado com sucesso, o número de linhas lidas foram gravadas no Lakehouse.



The screenshot shows the Microsoft Data Factory interface. On the left, there's a sidebar with various workspace options like 'Página Inicial', 'Criar', 'Procurar', 'Hub de dados do...', 'Hub de monitora...', 'Workspaces', and specific ones for 'mvp_engenharia...' and 'estacoesdb'. The main area shows a pipeline named 'mvp_dados_estacoes' with two activities: 'Copiar dados' (catalogo_estacoes to dados_estacoes) and 'Copiar dados' (dados_estacoes to Lakehouse). Both activities are marked as 'Bem-sucedido'. Below the activities, a table provides detailed execution statistics. On the right, a large panel titled 'Copiar os detalhes dos dados' displays the 'Source' (Armazenamento de Blob do Azure) and 'Destino' (Lakehouse) details, along with summary statistics for data read and written.

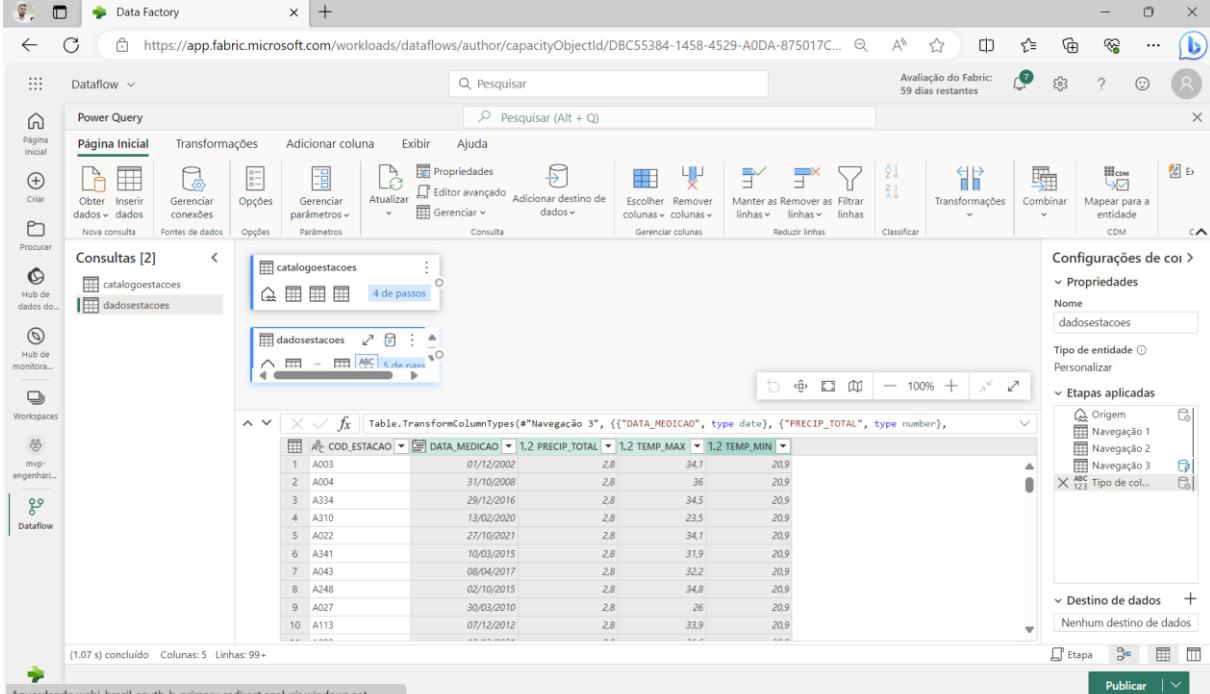
Figura 23 – pipeline com os catálogos das estações finalizado com sucesso, o número de linhas lidas foram gravadas no Lakehouse.

5.2. DATAFLOW



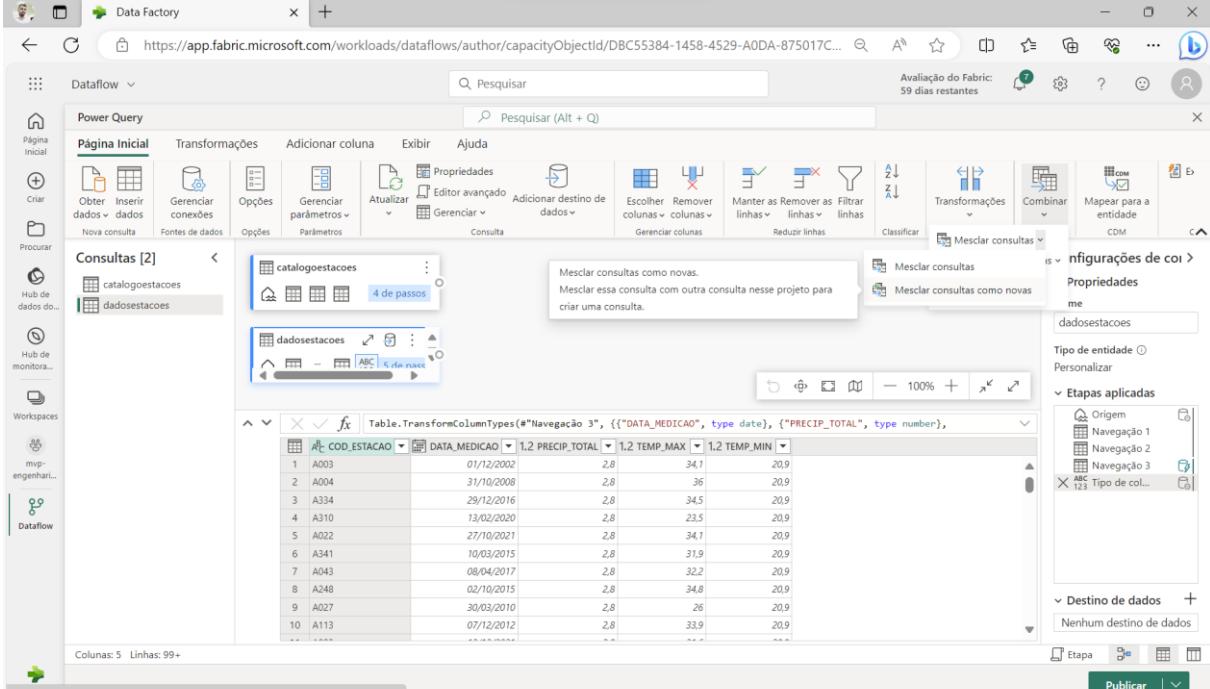
The screenshot shows the Microsoft Power Query 'Escolher dados' (Select Data) dialog. On the left, the 'Power Query' ribbon tab is selected. The main area shows a tree view of data sources under 'Obter dados' (Get Data). Under 'Lakehouse', there are two entries: 'mvp-engenharia_de_dados' and 'estacoesdb'. The 'estacoesdb' entry has several sub-folders: 'Arquivos', 'sys.sys_dw_manifest_files', 'sys.sys_dw_physical_files', 'catalogoestacoes', and 'dadosestacoes'. The 'dadosestacoes' folder is currently selected, and its contents are displayed in a grid table on the right. The table has columns: COD_ESTACAO, DATA_MEDICAO, PRECIP_TOTAL, TEMP_MAX, and TEMP_MIN. It contains approximately 30 rows of data, with the first few rows being: A003 (01/12/2002, 2.8, 34.1, 20.9), A004 (31/10/2008, 2.8, 36, 20.9), A334 (29/12/2016, 2.8, 34.5, 20.9), etc.

Figura 24 – para a criação do dataflow ao acessar o Lakehouse “estacoesdb” nota-se que as duas rotinas do pipeline funcionaram e as tabelas foram criadas e ambas possuem as informações extraídas do storage da Azure.



The screenshot shows the Microsoft Power Query Editor interface. On the left, there's a navigation pane with options like 'Dataflow', 'Power Query', 'Consultas [2]', and 'dadosestacoes'. The main area displays two tables: 'catalogoestacoes' and 'dadosestacoes'. The 'dadosestacoes' table has 99 rows and 5 columns. The 'DATA_MEDICAO' column is currently set to 'type date'. A tooltip above the column header indicates that the data type is being changed from string to date. The 'Transformações' ribbon tab is selected, and the 'Table.TransformColumnTypes' function is being used to change the data types of multiple columns at once. The right side of the screen shows configuration panels for 'Configurações de col' (column settings) and 'Etapas aplicadas' (applied steps), which include 'Origem' (origin) and 'Navegação 1' through 'Navegação 3'. The status bar at the bottom shows 'Colunas: 5 Linhas: 99+'.

Figura 25 – após a importação da tabela “dadosestacoes” vamos alterar o tipo de dados da coluna “DATA_MEDICAO” do tipo string para data, “PRECIP_TOTAL”, “TEMP_MAX” e “TEMP_MIN” de string para número decimal.



This screenshot is from the same session as Figure 25, showing the results of the transformation. The 'dadosestacoes' table now has the 'DATA_MEDICAO' column correctly typed as 'date'. The rest of the table structure remains the same, with 99 rows and 5 columns. The 'Transformações' ribbon tab is still selected, and the 'Table.TransformColumnTypes' function is visible in the formula bar. The right-hand configuration panels show the completed transformation steps and the unchanged destination settings.

Figura 26 – as tabelas foram mescladas para criar uma nova consulta unindo as informações da tabela de “catalogoestacoes” com a tabela “dadosestacoes”.

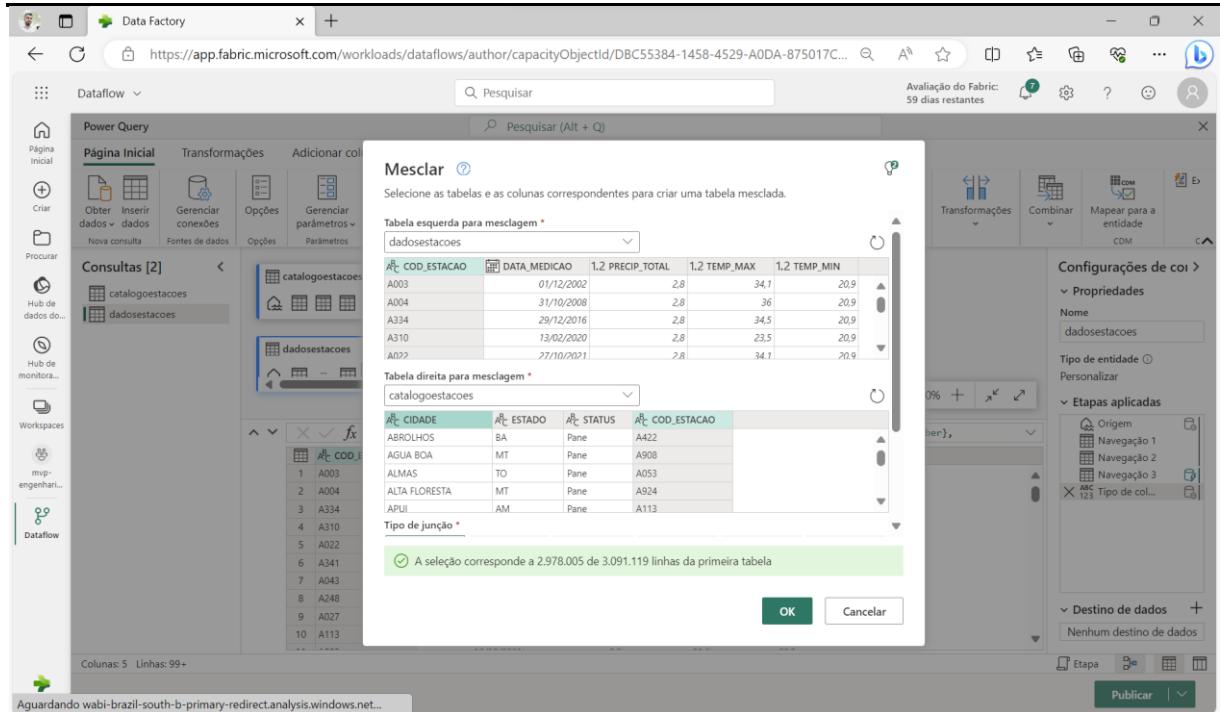


Figura 27 – foi utilizado a coluna “COD_ESTACAO” que possuem correspondências em ambas as tabelas para mesclá-las.

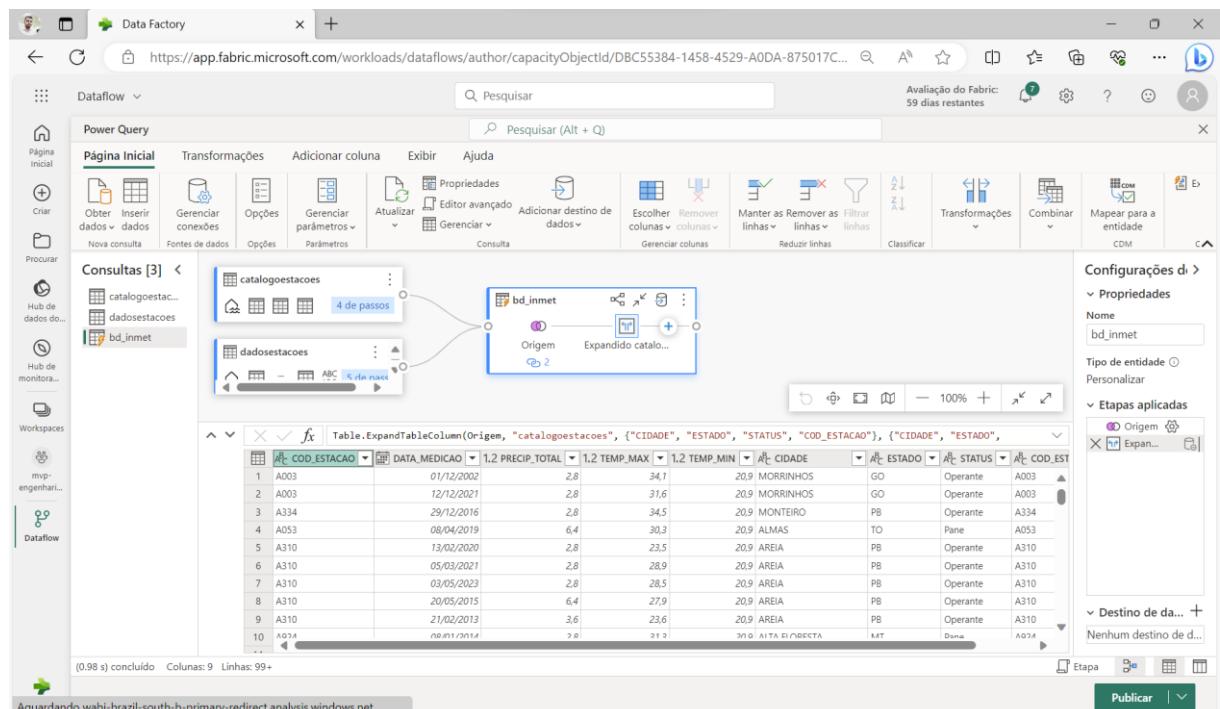


Figura 28 – após a junção das tabelas foi criada uma nova tabela nomeada de “bd_inmet”. Nesta nova tabela será excluída a última coluna “COD_ESTACAO” pois a mesma está duplicada.

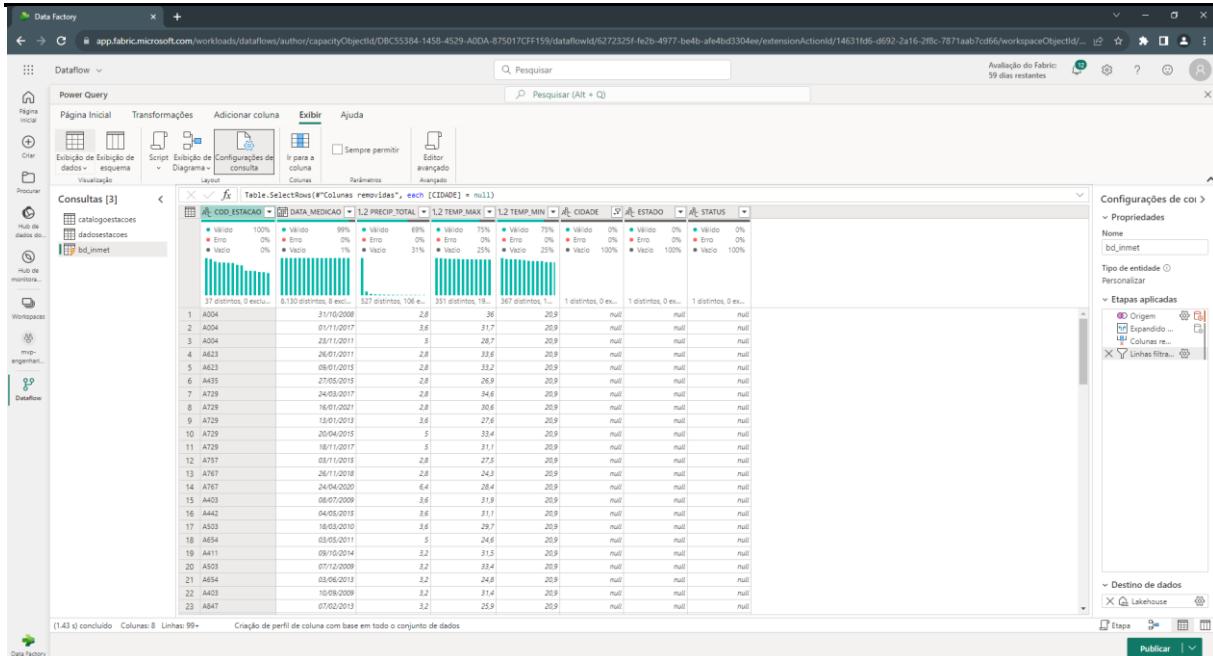
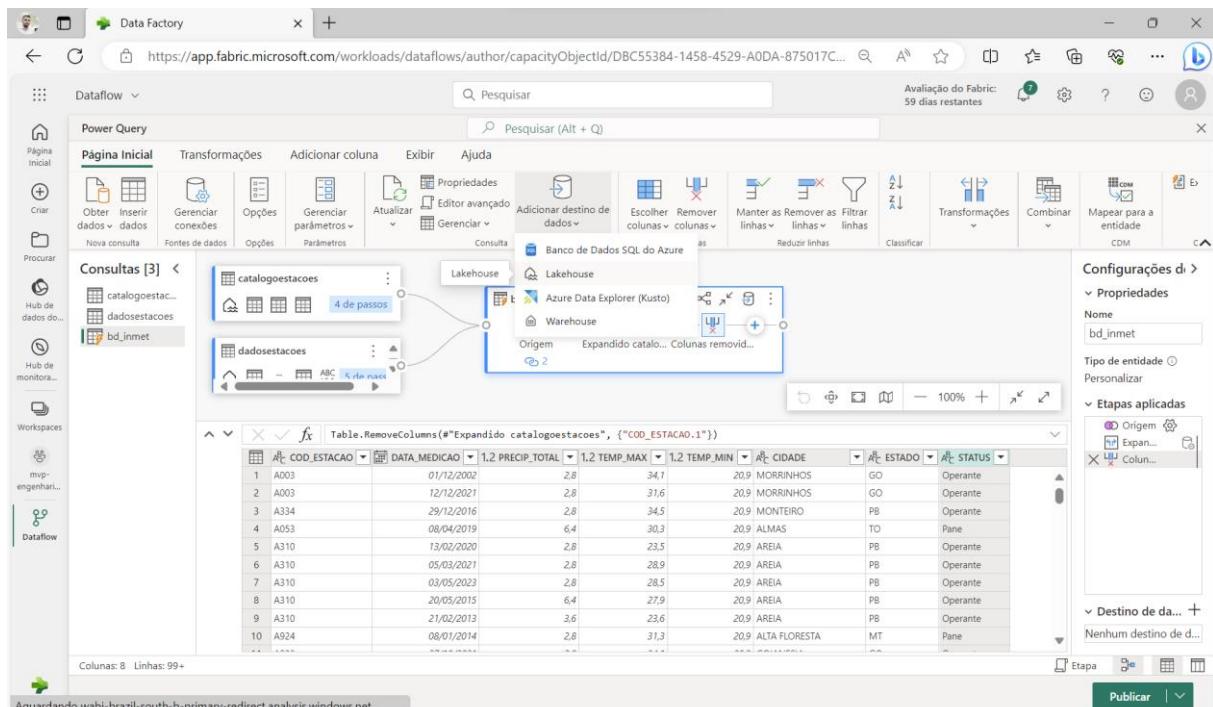


Table structure from Power Query Editor:

COD_ESTACAO	DATA_MEDICAO	PRECIP_TOTAL	TEMP_MAX	TEMP_MIN	CIDADE	ESTADO	STATUS
A004	01/10/2008	2,8	36	20,9			
A004	01/11/2017	3,6	31,7	20,9	null	null	
A004	23/11/2011	5	28,7	20,9	null	null	
A623	26/01/2011	2,8	33,6	20,9	null	null	
A623	09/01/2015	2,8	33,2	20,9	null	null	
A435	27/05/2015	2,8	26,9	20,9	null	null	
A729	24/03/2017	2,8	34,8	20,9	null	null	
A729	16/01/2017	2,8	30,9	20,9	null	null	
A729	13/01/2013	3,6	27,6	20,9	null	null	
A729	29/04/2015	5	33,4	20,9	null	null	
A729	18/11/2017	5	31,1	20,9	null	null	
A757	09/11/2018	2,8	27,5	20,9	null	null	
A767	26/11/2018	2,8	24,3	20,9	null	null	
A767	24/04/2020	6,4	28,4	20,9	null	null	
A403	08/07/2009	3,6	31,9	20,9	null	null	
A442	04/05/2013	3,6	31,1	20,9	null	null	
A503	18/03/2010	3,6	29,7	20,9	null	null	
A654	03/05/2011	5	24,6	20,9	null	null	
A411	09/10/2014	3,2	31,5	20,9	null	null	
A503	07/12/2009	3,2	33,4	20,9	null	null	
A654	03/06/2013	3,2	24,8	20,9	null	null	
A403	10/09/2009	3,2	31,4	20,9	null	null	
A647	07/02/2013	3,2	25,9	20,9	null	null	

Figura 29 – foi identificado colunas com valores null, realizando uma busca detalhada foi encontrado um total de 37 estações que não constam informações na tabela “catalogoestacoes”, verificando no site INMET foi verificado que essas estações foram desativadas.



Transformation steps from Power Query Editor:

- Step 1: Catalogoestacoes (4 de passos) → Lakehouse
- Step 2: dadosestacoes → Origen
- Step 3: Table.RemoveColumns("Expandido_catalogoestacoes", {"COD_ESTACAO_1"})

Table structure after transformation:

COD_ESTACAO	DATA_MEDICAO	PRECIP_TOTAL	TEMP_MAX	TEMP_MIN	CIDADE	ESTADO	STATUS
A003	01/12/2002	2,8	34,1	20,9	MORRINHOS	GO	Operante
A003	12/12/2021	2,8	31,6	20,9	MORRINHOS	GO	Operante
A334	29/12/2016	2,8	34,5	20,9	MONTEIRO	PB	Operante
A503	08/04/2019	6,4	30,3	20,9	ALMAS	TO	Pane
A310	13/02/2020	2,8	23,5	20,9	AREIA	PB	Operante
A310	05/03/2021	2,8	28,9	20,9	AREIA	PB	Operante
A310	03/05/2023	2,8	28,5	20,9	AREIA	PB	Operante
A310	20/05/2015	6,4	27,9	20,9	AREIA	PB	Operante
A310	21/02/2013	3,6	23,6	20,9	AREIA	PB	Operante
A924	08/01/2014	2,8	31,3	20,9	ALTA FLORESTA	MT	Pane

Figura 30 – após o tratamento das informações e criação de uma nova tabela foi definido o destino dos dados desta nova tabela, no qual será dentro do Lakehouse já utilizado.

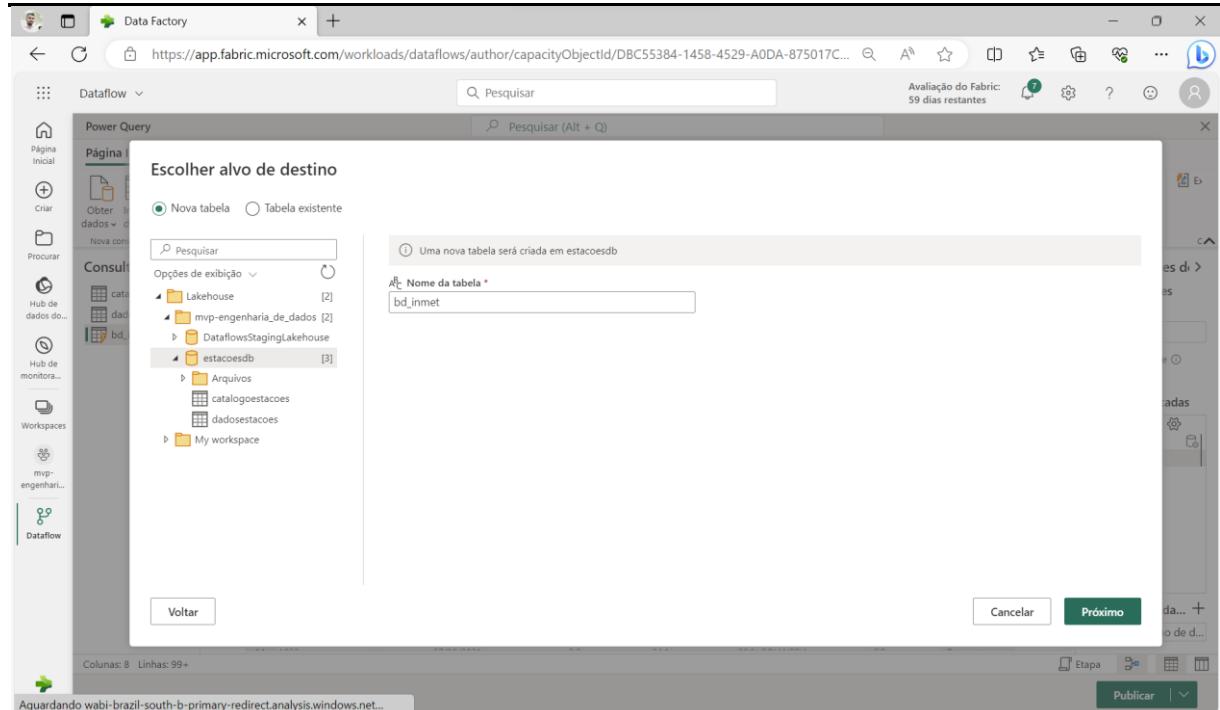


Figura 31 – a nova tabela será salva dentro do Lakehouse com o nome “bd_inmet”.

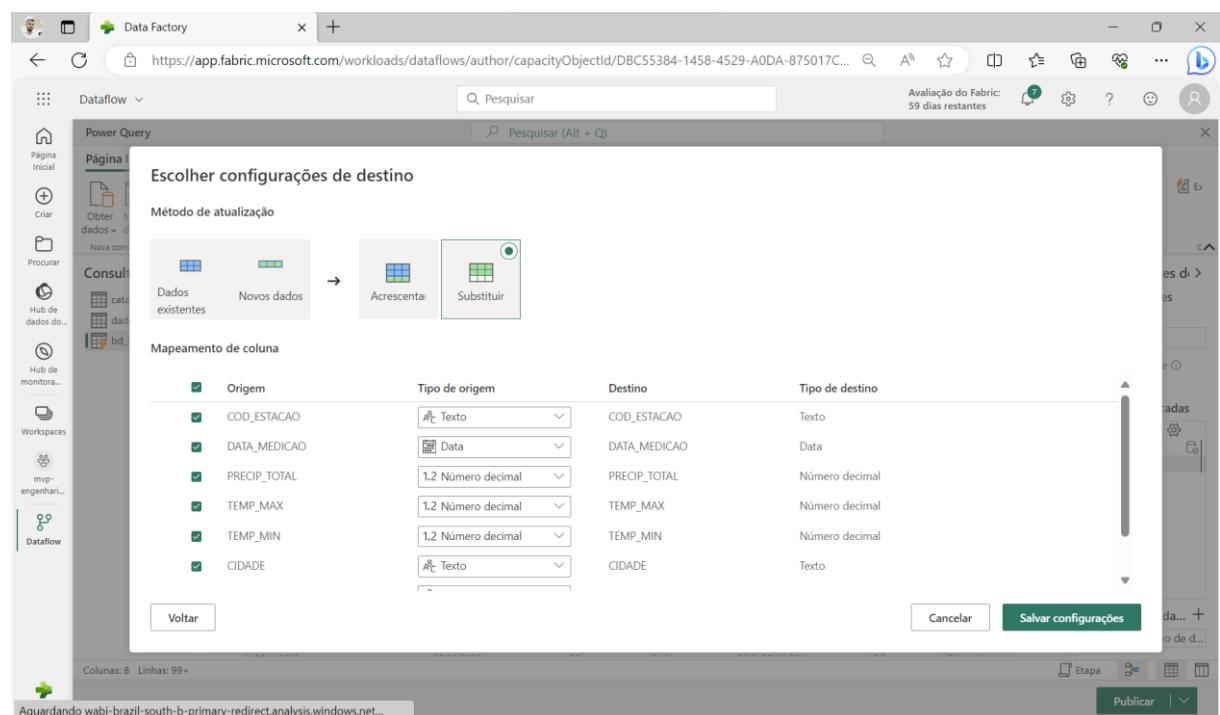


Figura 32 – o método de atualização no banco de dados utilizado foi o de substituição das informações.



PÓS-GRADUAÇÃO: Ciência de Dados e Analytics
SPRINT: Engenharia de Dados
DISCIPLINA: MVP – Minimum Viable Product
ORIENTADOR: Vitor Almeida
ALUNO: Jefferson Brunno Leandro Nery

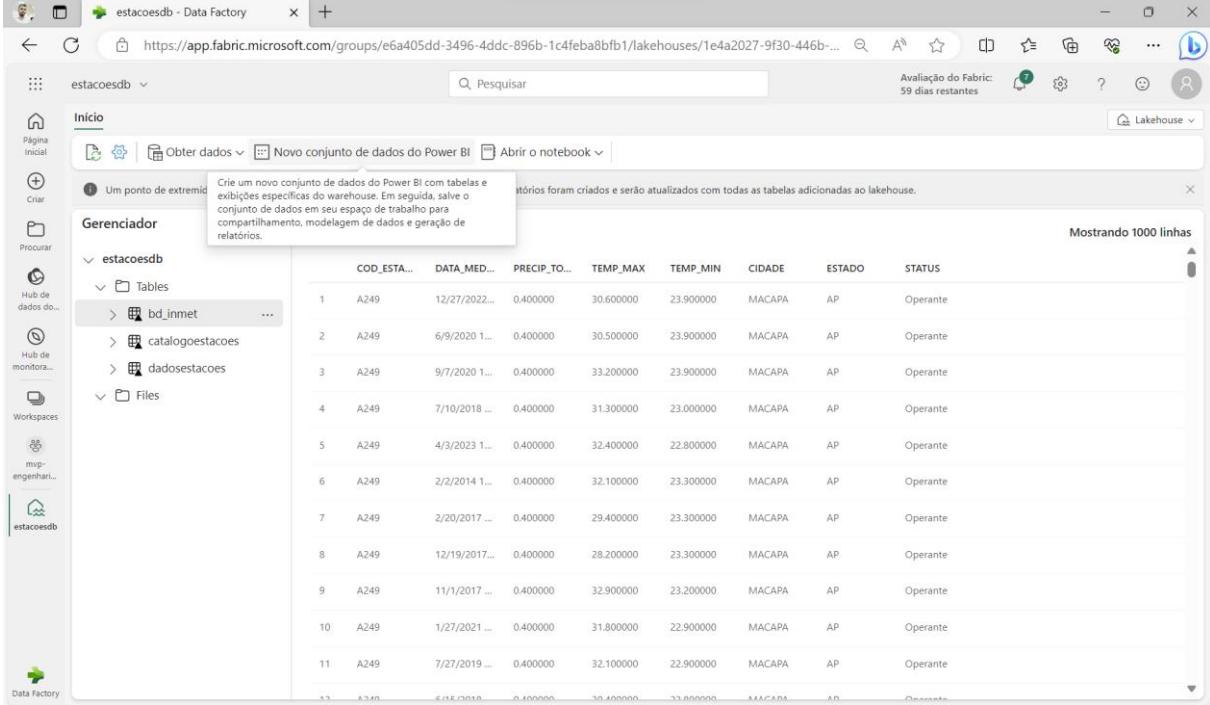
The screenshot shows the Microsoft Power Query Editor interface. On the left, there's a navigation bar with links like 'Dataflow', 'Página inicial', 'Criar', 'Procurar', 'Hub de dados...', 'Hub de monitora...', 'Workspaces', 'mvp-engenhar...', and 'Dataflow'. The main area has a toolbar with buttons for 'Power Query' (Página Inicial, Transformações, Adicionar coluna, Exibir, Ajuda), 'Consultas' (3), and various data manipulation tools (Propriedades, Gerenciar conexões, Opções, Gerenciar parâmetros, Atualizar, Editor avançado, Adicionar destino de dados, Consulta, Escolher colunas, Remover colunas, Manter as linhas, Remover linhas, Filtrar linhas, Reduzir linhas, Classificar, Transformações, Combinar, Mapear para a entidade CDM). A central workspace displays a data flow with three tables: 'catalogoestacoes', 'bd_inmet', and 'dadosestacoes'. The 'catalogoestacoes' table has 4 steps. The 'bd_inmet' table is the source ('Origem') for an expansion step ('Expandido catalogoestacoes...'). The 'dadosestacoes' table is also connected to the expansion step. Below the tables is a preview of a transformed table with columns: COD_ESTACAO, DATA_MEDICAO, 1.2_PRECIP_TOTAL, 1.2_TEMP_MAX, 1.2_TEMP_MIN, CIDADE, ESTADO, and STATUS. The preview shows 10 rows of data. On the right, there are configuration panels for 'Propriedades' (Nome: bd_inmet, Tipo de entidade: Personalizar) and 'Etapas aplicadas' (Origem, Expand..., Colun...). At the bottom, there are buttons for 'Destino de dados' (Lakehouse), 'Etapas', 'Publicar', and a save button.

Figura 33 – após a definição do destino dos dados o dataflow foi publicado.

6. ANÁLISE DOS DADOS

Após a conclusão das etapas anteriores vamos utilizar as informações armazenadas no Lakehouse para realizar a análise dos dados para visando responder as questões mencionadas no objetivo deste MVP. Nesta etapa utilizaremos uma ferramenta de BI³ disponível no Microsoft Fabric conhecida como PowerBI.

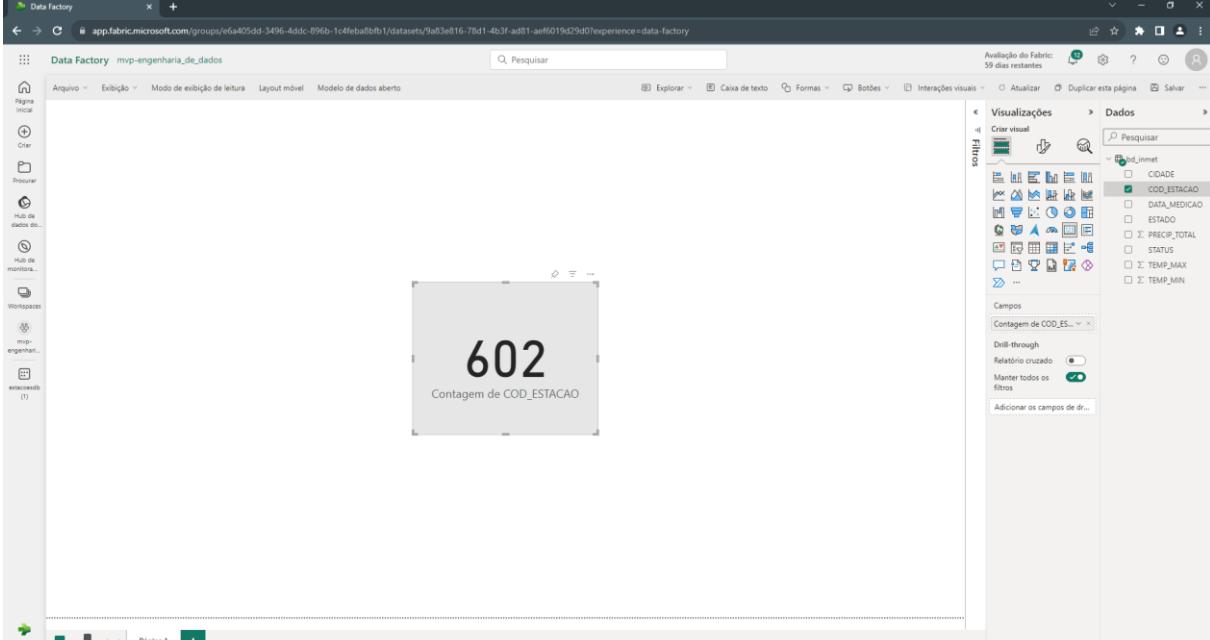
3 business intelligence



The screenshot shows the Microsoft Power BI Data Factory interface. On the left, there's a sidebar with various options like 'Página inicial', 'Criar', 'Procurar', 'Hub de dados do...', 'Hub de monitora...', 'Workspaces', and 'mvp-engenharia...'. The main area is titled 'estacoedb' and shows a 'Gerenciador' section with a 'Tables' folder containing 'bd_inmet', 'catalogoestacoes', and 'dadosestacoes'. A 'Files' folder is also present. A modal window titled 'Um ponto de extremidade' is open, providing instructions on how to create a new Power BI dataset from the tables in the warehouse. To the right, a table view displays 1000 rows of data from the 'bd_inmet' table, showing columns like COD_ESTACAO, DATA_MEDICAO, PRECIP_TOTAL, TEMP_MAX, TEMP_MIN, CIDADE, ESTADO, and STATUS. The data shows 11 entries for station A249 across different dates and locations.

Figura 34 – criando um novo conjunto de dados para análise dos dados no PowerBI.

- Qual a quantidade total de estações meteorológicas?



The screenshot shows the Microsoft Power BI Data Factory interface with a visual representation of the data. A large digital gauge or counter visualization displays the number '602' in a prominent font. Below the visualization, it says 'Contagem de COD_ESTACAO'. The interface includes a navigation bar at the top with 'Data Factory', 'mvp-engenharia_de_dados', and a search bar. On the left, there's a sidebar with 'Arquivo', 'Exibição', 'Modo de exibição de leitura', 'Layout móvel', and 'Modelo de dados aberto'. The main workspace shows a 'Visualizações' pane on the left with various chart icons and a 'Dados' pane on the right listing fields like CIDADE, COD_ESTACAO, DATA_MEDICAO, ESTADO, PRECIP_TOTAL, STATUS, TEMP_MAX, and TEMP_MIN. A filter for 'Contagem de COD_ESTACAO' is applied to the visualization.

Figura 35 – utilizando o conjunto de dados identificamos que temos um total de 602 estações meteorológicas.

- Como estão distribuídas as estações meteorológicas por estado?

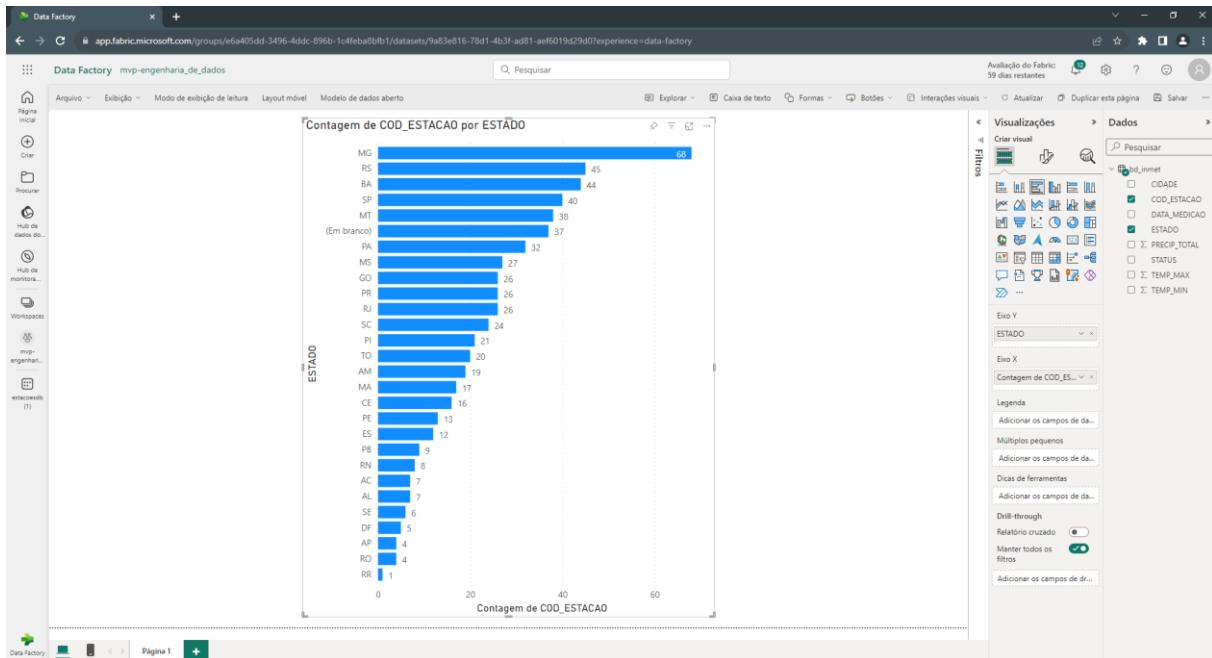


Figura 36 – distribuição das estações meteorológicas por estado e 37 estações que estão desativadas.

- Qual o estado com o maior número de estações meteorológicas?

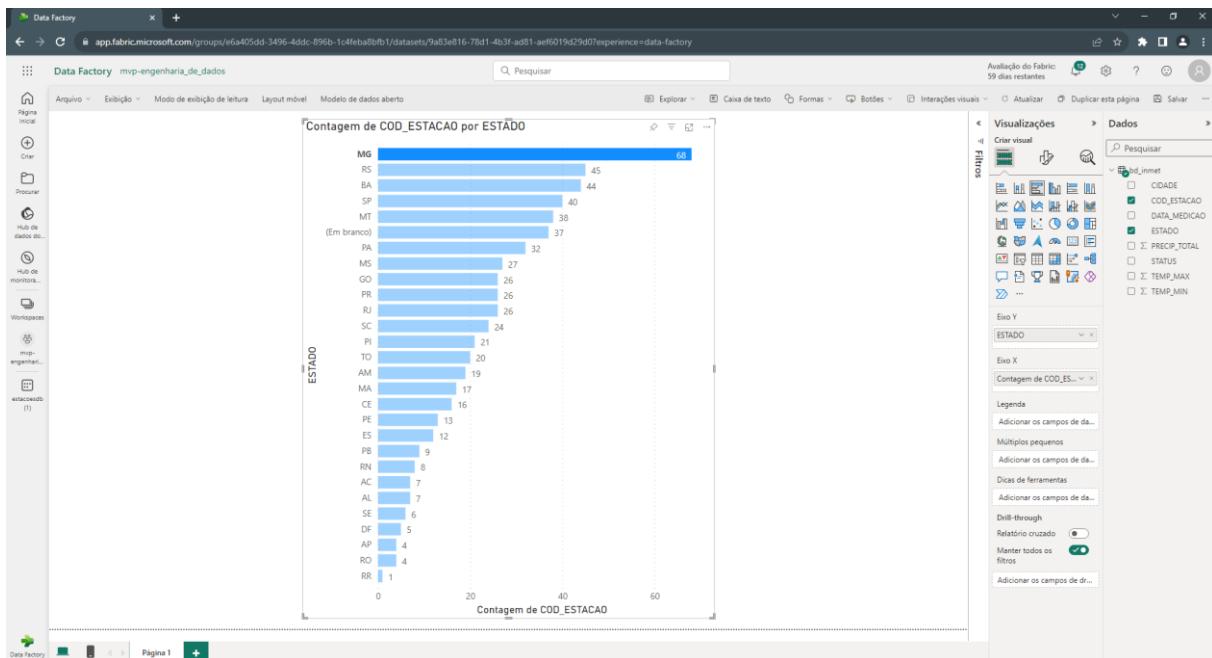


Figura 37 – o estado com o maior número de estações meteorológicas é Minas Gerais com um total de 68 estações.

- Qual o estado com o menor número de estações meteorológicas?

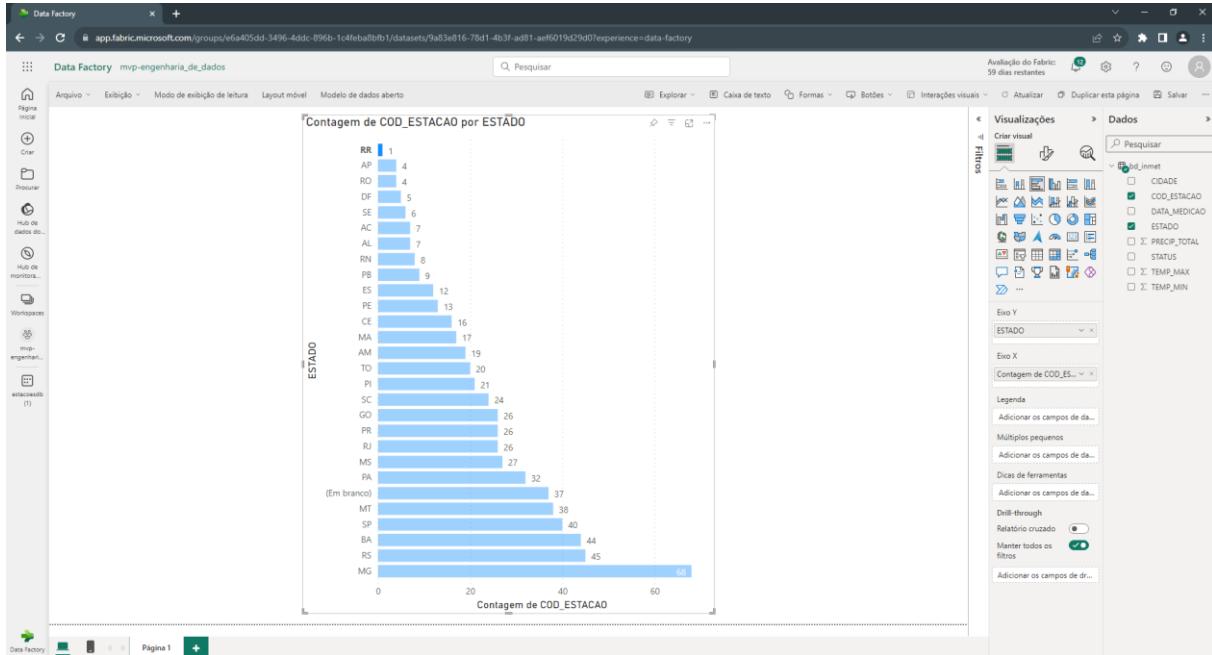


Figura 38 – o estado com o menor número de estações meteorológicas é Roraima com apenas 1 (uma) estação em todo o estado.

- Qual o status das estações meteorológicas?

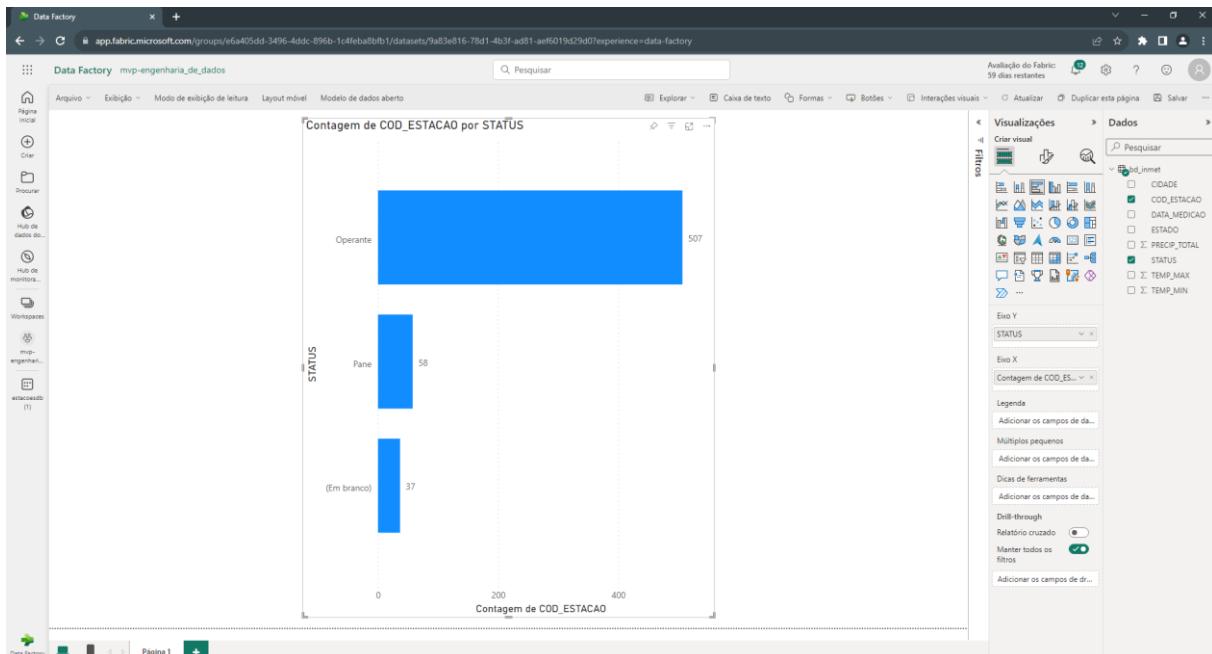
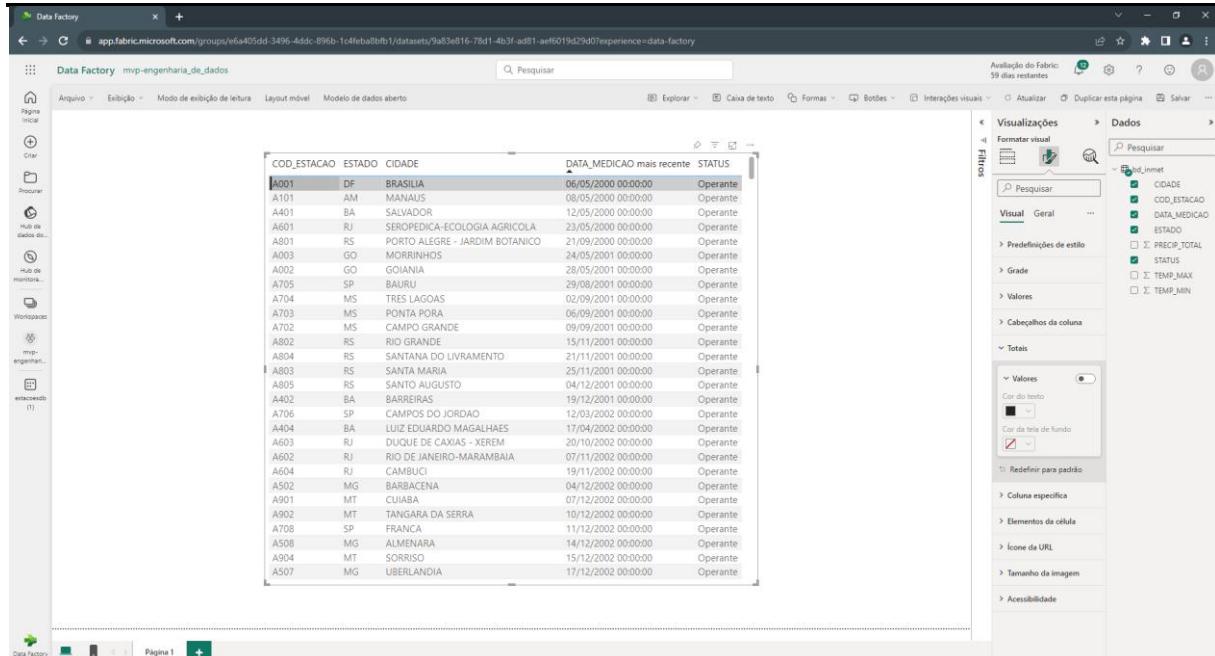


Figura 39 – temos um total de 507 estações operantes, 58 com pane e 37 desativadas.



COD_ESTACAO	ESTADO	CIDADE	DATA_MEDICAO mais recente	STATUS
A001	DF	BRASILIA	06/05/2000 00:00:00	Operante
A101	AM	MANAUS	08/05/2000 00:00:00	Operante
A401	BA	SALVADOR	12/05/2000 00:00:00	Operante
A601	RJ	SEROPEDICA-ECOLOGIA AGRICOLA	23/05/2000 00:00:00	Operante
A801	RS	PORTO ALEGRE - JARDIM BOTANICO	21/09/2000 00:00:00	Operante
A003	GO	MORRINHOS	24/05/2001 00:00:00	Operante
A002	GO	GOIANIA	28/05/2001 00:00:00	Operante
A705	SP	BAURO	29/08/2001 00:00:00	Operante
A704	MS	TRES LAGOAS	02/09/2001 00:00:00	Operante
A703	MS	PONTA PORA	06/09/2001 00:00:00	Operante
A702	MS	CAMPOM GRANDE	09/09/2001 00:00:00	Operante
A802	RS	RIO GRANDE	15/11/2001 00:00:00	Operante
A804	RS	SANTANA DO LIVRAMENTO	21/11/2001 00:00:00	Operante
A803	RS	SANTA MARIA	25/11/2001 00:00:00	Operante
A805	RS	SANTO AUGUSTO	04/12/2001 00:00:00	Operante
A402	BA	BARREIRAS	19/12/2001 00:00:00	Operante
A706	SP	CAMPOS DO JORDAO	12/03/2002 00:00:00	Operante
A404	BA	LUIZ EDUARDO MAGALHAES	17/04/2002 00:00:00	Operante
A603	RJ	DUQUE DE CAXIAS - XEREM	20/10/2002 00:00:00	Operante
A602	RJ	RIO DE JANEIRO-MARAMBAIA	07/11/2002 00:00:00	Operante
A604	RJ	CAMBUCI	19/11/2002 00:00:00	Operante
A502	MG	BARBACENA	04/12/2002 00:00:00	Operante
A901	MT	CUIABA	07/12/2002 00:00:00	Operante
A902	MT	TANGARA DA SERRA	10/12/2002 00:00:00	Operante
A708	SP	FRANCA	11/12/2002 00:00:00	Operante
A508	MG	ALMENARA	14/12/2002 00:00:00	Operante
A904	MT	SORRISO	15/12/2002 00:00:00	Operante
A507	MG	UBERLANDIA	17/12/2002 00:00:00	Operante

Figura 40 - a estação A001 localizada em Brasília - DF é a estação meteorológica a mais tempo em operação (23 anos).

7. CONCLUSÃO

Os dados extraídos do site INMET acerca das estações meteorológicas disponíveis no território nacional demonstram uma qualidade “aceitável” das informações. Pode-se colocar um ponto de melhoria dos dados na base de catálogos das estações inserir as informações das 37 estações que foram desativadas com o intuído de manter um histórico para outros usuários que venham a realizar um trabalho com essas informações. Foi realizado a extração dos dados em formato de arquivo do tipo “csv” e a carga destes dados foram realizados na plataforma da Azure em blob de armazenamento “nuvem”, dentro da plataforma Microsoft Fabric foi utilizado a ferramenta Data Factory para a criação do pipeline e dataflow das informações para o Lakehouse, foi realizado o tratamento dos dados e utilizou-se uma ferramenta de BI “PowerBI” para conectar-se ao conjunto de dados criado para a análise das informações visando responder os questionamentos deste MVP. Conclui-se que os objetivos deste MVP foram atingidos com a utilização das ferramentas da Microsoft Fabric e Azure para a obtenção dos resultados e análises finais.