

Título:	IT_73 Disponibilização de Conjuntos de Dados do Azure Databricks Unity no Power BI	
Diretoria:	Gerência Responsável:	
DSC/STI	Gerência de Tecnologia do Negócio	Data publicação: 12/07/2024



1. Controle de Versão

Histórico (para uso da Gerência de Governança):

Versão	Data Publicação	Elaborador	Motivo da Versão
1.0	12/07/2024	Thiago Silva	Estruturação de documento para orientar as equipes de negócio e da STI a elaborar conjuntos de dados do Power BI a partir do Azure Databricks Unity

Aprovação da Versão Atual:

Dono do Processo	Cargo
Leonardo Mafra Silva	Coordenador de Data & Analytics

2. Objetivo/Introdução

Auxiliar a equipe da Coordenação de Data & Analytics da Gerência de Tecnologia do Negócio/STI e das áreas de negócio a criar e disponibilizar conjuntos de dados (*datasets*) baseados no *Unity Catalog* do *Azure Databricks*.

3. Referências

N/A

4. Pré-requisitos para Execução das Atividades

As atividades desse documento serão apresentadas usando como exemplo uma computação do tipo *SQL Warehouse* do *Azure Databricks* e, portanto, é esperado que ela já tenha sido criada e que os desenvolvedores do Power BI da área e o usuário de integração (inteligencia@cni.com.br) tenham permissão de usá-la. Além disso, esses mesmos usuários precisam ter acesso aos ativos via *Unity*.

O colaborador que vai realizar essa atividade deve possuir licença Pro do Power BI, deve ter acesso a um workspace do tipo STAGE no serviço do Power BI, deve ter acesso ao SQL Warehouse do Databricks e aos ativos de interesse via *Unity Catalog*.

5. Descrição das Atividades

Os processos apresentados aqui serão divididos em 2 grandes etapas: uma para configurar a conexão para um conjunto de dados novo e outra para reapontar uma conexão de um conjunto de dados já existente.

Título:	IT_73 Disponibilização de Conjuntos de Dados do Azure Databricks Unity no Power BI	
Diretoria:	DSC/STI	Gerência Responsável: Gerência de Tecnologia do Negócio Data publicação: 12/07/2024



5.1. Configurar a conexão para um conjunto de dados novo

Existem dois métodos para estabelecer a conexão do Unity com o Power BI: uma delas vai fornecer um arquivo do Power BI com a conexão já criada e uma outra em que será preciso configurar essa conexão de ponta a ponta. Os dois métodos serão apresentados em detalhe.

Inicialmente, acesse o Azure Databricks e clique na seção SQL Warehouse. Para esse exemplo, será usado o Warehouse SQL-STI, mas selecione aquele que tem acesso. É a partir dele que a conexão será realizada.

Status	Name	Created by	Size	Active / Max	Type
○	SQL-SDH	victor.ribeiro@cni.com.br	2X-Small	1 / 1	Serverless
○	SQL-STI	t-thomaz.moreira@cni.com.br	2X-Small	1 / 1	Serverless
○	SQL-UNINDUSTRIA	victor.ribeiro@cni.com.br	2X-Small	1 / 1	Serverless
○	oni-lançaini-gpt	victor.ribeiro@cni.com.br	2X-Small	0 / 1	Pro
○	serverless	victor.ribeiro@cni.com.br	2X-Small	0 / 1	Serverless
○	SQL-ARRECADACAO	victor.ribeiro@cni.com.br	2X-Small	0 / 1	Serverless
○	SQL-COAL	thiago.silva@senaiaci.com.br	2X-Small	0 / 1	Serverless
○	SQL-DIRCOM	victor.ribeiro@cni.com.br	2X-Small	0 / 1	Serverless
○	SQL-DTI	victor.ribeiro@cni.com.br	2X-Small	0 / 1	Serverless
○	SQL-GCH	victor.ribeiro@cni.com.br	2X-Small	0 / 1	Serverless
○	SQL-ONI	victor.ribeiro@cni.com.br	2X-Small	0 / 1	Serverless
○	SQL-PowerBI-Refresh	victor.ribeiro@cni.com.br	2X-Small	0 / 1	Serverless
○	SQL-SCG	victor.ribeiro@cni.com.br	2X-Small	0 / 1	Serverless
○	SQL-SSI	victor.ribeiro@cni.com.br	2X-Small	0 / 1	Serverless
○	SQL-UNIMERCADO	victor.ribeiro@cni.com.br	2X-Small	0 / 1	Serverless

Figura 1 - Seleção do SQL Warehouse

Na tela seguinte, navegue até a aba “Connection details”.

Overview	Connection details	Monitoring
Status	Running	
Name	SQL-STI (ID: e2113fc5c25f0bbb)	
Type	Serverless	

Figura 2 - SQL Warehouse

As opções de conexão são apresentadas em seguida, com diversas possibilidades de conexão, mas o foco aqui é mostrar como isso ocorre no Power BI.

Título:	IT_73 Disponibilização de Conjuntos de Dados do Azure Databricks Unity no Power BI	
Diretoria:	DSC/STI	Gerência Responsável: Gerência de Tecnologia do Negócio Data publicação: 12/07/2024

SQL Warehouses >
SQL-STI

Overview Connection details Monitoring

Server hostname
adb-6523536500265509.9.azuredatabricks.net

Port
443

Protocol
https

HTTP path
/sql/1.0/warehouses/e2113fc5c25f0bbb

JDBC URL 2.6.25 or later

Databricks supports drivers released within the last two years. Download drivers here
jdbc:databricks://adb-6523536500265509.9.azuredatabricks.net:443/default;transportMode=http;ssl=1;AuthMech=3;httpPath=/sql/1.0/warehouses/e2113fc5c25f0bbb;

Use these details to connect to this warehouse

Tableau Power BI dbt Python Java Node.js Go More tools

Figura 3 - Detalhes de conexão com o SQL Warehouse

Clicando no ícone do Power BI, será apresentada a tela com a opção para o download do arquivo de conexão do Unity com o Power BI. Faça o download desse arquivo. Ele será criado com o nome databricks-SQL-STI.pbids.

Microsoft Azure databricks

SQL Warehouses > **SQL-STI**

Search data, notebooks, recent, and more... CTRL + P

Connect to partner Microsoft Power BI

You can use Partner Connect to connect Power BI to a Azure Databricks SQL Warehouse. Download and open the connection file (.pbids) to start Power BI. You must have Power BI version 2.99.562.0 or above installed.

Learn more Compute SQL-STI

Download connection file Cancel

Figura 4 - Download do arquivo pbids para conexão do Unity com o Power BI

Título:

IT_73 Disponibilização de Conjuntos de Dados do Azure Databricks Unity no Power BI



Diretoria:

DSC/STI

Gerência Responsável:

Gerência de Tecnologia do Negócio

Data publicação:

12/07/2024

A extensão “.pbids”, conhecida como *Power BI Data Source*, é utilizada para facilitar o compartilhamento e a reutilização das configurações de conexão com as fontes de dados, eliminando a necessidade de configurar esses parâmetros manualmente. Contudo, ela difere do arquivo de painel do Power BI, cuja extensão é “.pbix”. Apenas após abrir o arquivo “.pbids” e selecionar os ativos do Unity desejados, é que o dataset e/ou o painel poderá ser criado, salvando-o com o nome e a extensão “.pbix”. Observe que ao fazer isso o arquivo de configuração “.pbids” não é alterado, ele apenas facilita o processo de conexão para gerar um “.pbix”. Tal funcionalidade é especialmente vantajosa, dado que o arquivo “.pbids” baixado pode ser reutilizado em diferentes contextos. Isso permite a conexão com outros ativos do Unity que utilizem o mesmo SQL Warehouse, facilitando a criação de novos painéis sem a necessidade de baixar o “.pbids” novamente.

Para tanto, ao abrir esse arquivo pbids baixado, o Power BI abrirá já na tela que permite a seleção dos ativos do Unity para os quais o usuário tem acesso.

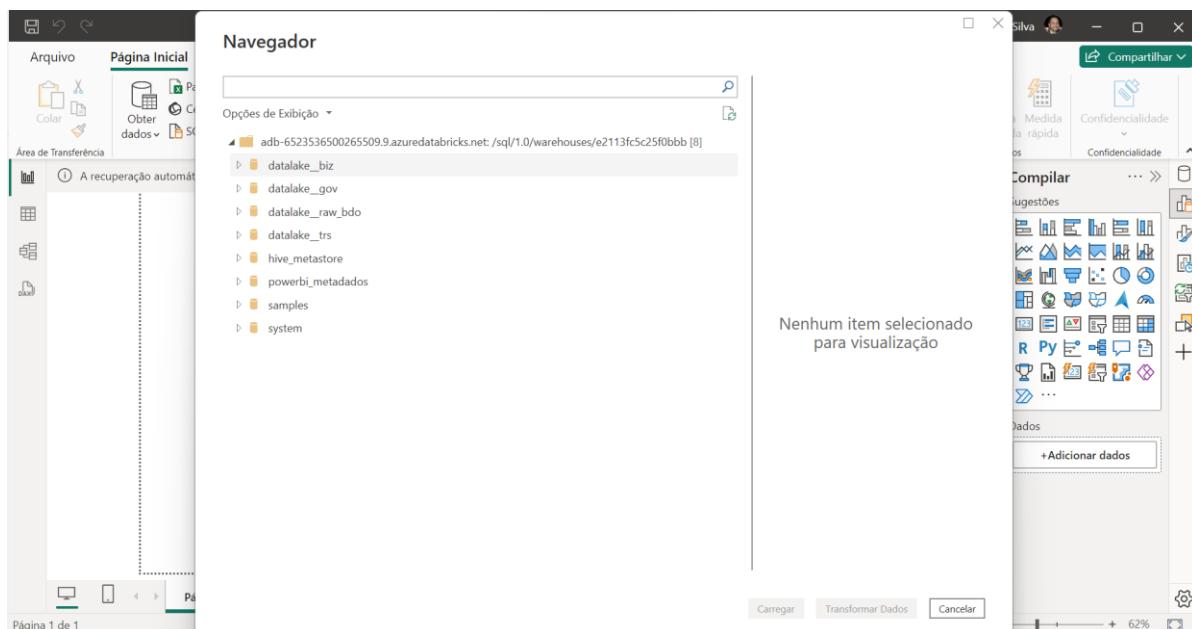


Figura 5 - Conexão com ativos do Unity via arquivo “.pbids”

Nesse ponto, se percebe uma das grandes vantagens no uso dessa conexão, pois podem ser escolhidos quantos ativos forem necessários para o dataset/painel em construção, bastando apenas selecioná-los para que a conexão seja estabelecida com eles de uma vez só, inclusive de catálogos e schemas distintos. Ao final da seleção, clique em “Transformar dados” se precisar realizar ajustes pelo Power Query nos dados carregados ou clique em “Carregar” para já levá-los direto à área de desenvolvimento do painel. O exemplo seguirá com a escolha de quatro origens, seguindo diretamente pelo botão “Carregar”.

Observação: caso seja necessário ter um *preview* dos dados, pode se clicar no ativo e uma amostra dos dados será apresentada para visualização nessa mesma tela.

Título:

IT_73 Disponibilização de Conjuntos de Dados do Azure Databricks Unity no Power BI



Diretoria:

DSC/STI

Gerência Responsável:

Gerência de Tecnologia do Negócio

Data publicação:

12/07/2024

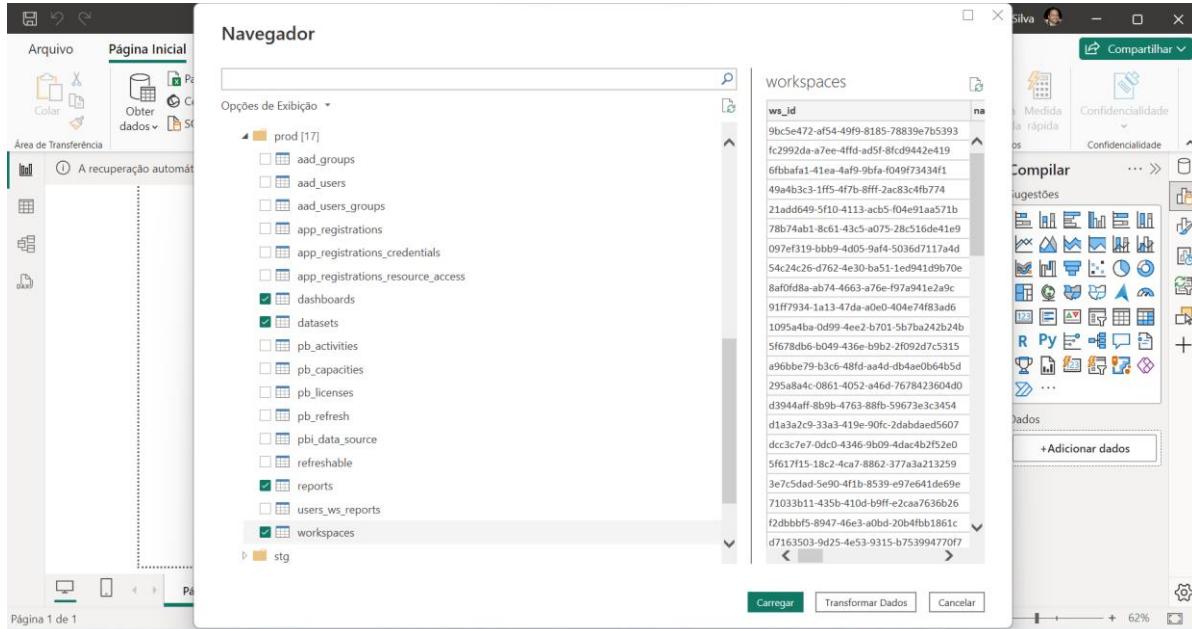


Figura 6 - seleção dos ativos do Unity para conexão com o Power BI

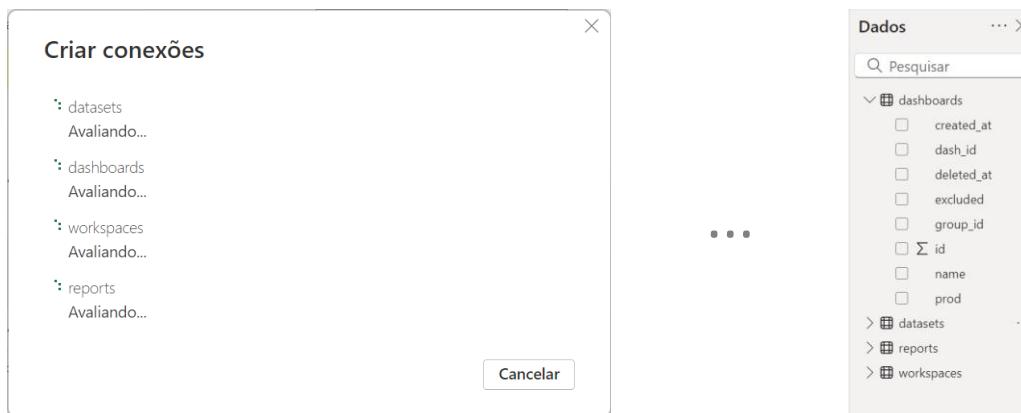


Figura 7 - Carregamento dos dados selecionados

Com isso, pode ser realizada a modelagem e desenvolvimento de visuais como em qualquer outro contexto de elaboração de painéis. Ao final, salve o arquivo “.pbix” e publique o painel no workspace ao qual tem acesso.

Título:	IT_73 Disponibilização de Conjuntos de Dados do Azure Databricks Unity no Power BI	
Diretoria:	Gerência Responsável:	Data publicação:
DSC/STI	Gerência de Tecnologia do Negócio	12/07/2024

5.2. Configurar a conexão para um conjunto de dados já existente

Nessa etapa, será feita a alteração da origem dos dados de um dataset que já existe e que precisa ser reapontado para o Unity.

Inicialmente, acesse o Azure Databricks e clique na seção SQL Warehouse. Para esse exemplo, será usado o Warehouse SQL-STI, mas selecione aquele que tem acesso. É a partir dele que a conexão será realizada.

The screenshot shows the Microsoft Azure Databricks interface. On the left sidebar, under the 'Compute' category, the 'SQL Warehouses' option is selected and highlighted with a red box. The main area displays a table of SQL warehouses. One row, 'SQL-STI', has a red arrow pointing to it. The table columns include Status, Name, Created by, Size, Active / Max, and Type. The 'Type' column for 'SQL-STI' shows 'Serverless'.

Figura 9 - Seleção do SQL Warehouse

Na tela seguinte, navegue até a aba “Connection details”.

The screenshot shows the 'SQL-STI' page. At the top, there are three tabs: 'Overview' (selected), 'Connection details' (highlighted with a red box), and 'Monitoring'. Below the tabs, there are sections for 'Status' (Running), 'Name' (SQL-STI (ID: e2113fc5c25f0bbb)), and 'Type' (Serverless). The 'Overview' tab contains some placeholder text about the warehouse's purpose.

Figura 10 - SQL Warehouse

As opções de conexão são apresentadas em seguida, com diversas possibilidades de conexão, mas o foco aqui é mostrar como isso ocorre no Power BI, por meio do **Server hostname** e o **HTTP Path**.

Título:	IT_73 Disponibilização de Conjuntos de Dados do Azure Databricks Unity no Power BI	
Diretoria:	DSC/STI	Gerência Responsável: Gerência de Tecnologia do Negócio Data publicação: 12/07/2024

SQL Warehouses >
SQL-STI

Overview Connection details Monitoring

Server hostname
adb-6523536500265509.9.azuredatabricks.net

Port
443

Protocol
https

HTTP path
/sql/1.0/warehouses/e2113fc5c25f0bbb

JDBC URL 2.6.25 or later

Databricks supports drivers released within the last two years. Download drivers here
jdbc:databricks://adb-6523536500265509.9.azuredatabricks.net:443/default;transportMode=http;ssl=1;AuthMech=3;httpPath=/sql/1.0/warehouses/e2113fc5c25f0bbb;

Use these details to connect to this warehouse

Tableau Power BI dbt Python Java Node.js Go More tools

Figura 11 - Detalhes de conexão com o SQL Warehouse

Nessa etapa, abra o Power BI Desktop e procure pelo conector “Azure Databricks” e clique em Conectar (figuras 12 e 13).

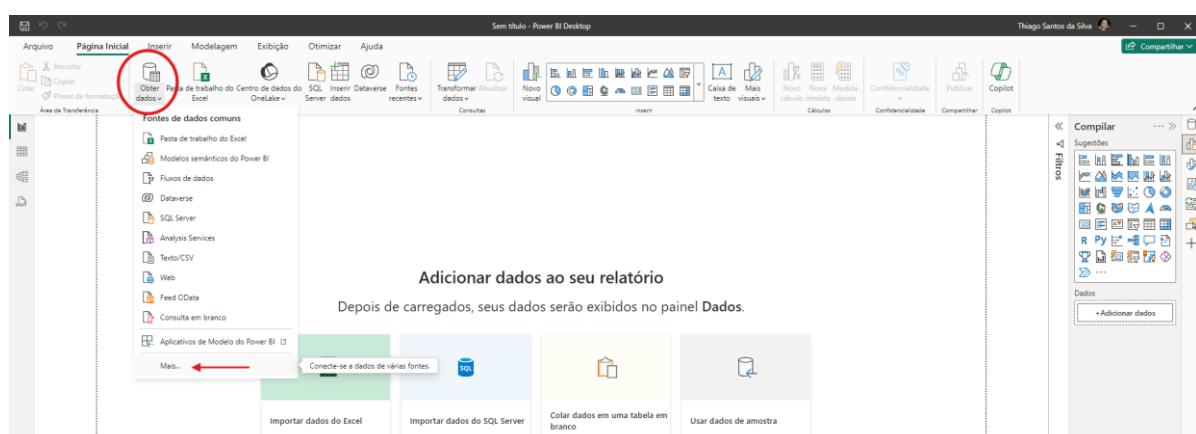


Figura 12 - Seleção do conector Azure Databricks

Título:	IT_73 Disponibilização de Conjuntos de Dados do Azure Databricks Unity no Power BI	
Diretoria:	Gerência Responsável:	Data publicação:
DSC/STI	Gerência de Tecnologia do Negócio	12/07/2024

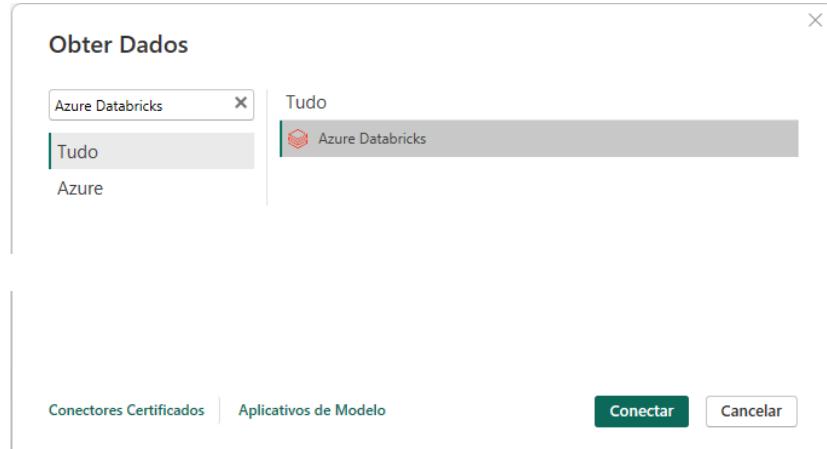


Figura 13 - Seleção do conector Azure Databricks

Nesse momento será preciso configurar a conexão com base nos valores da figura 11. Em “Nome do Host do Servidor” coloque o valor de “Server hostname” e em “Caminho HTTP” coloque o “HTTP Path”. As demais opções podem se manter como estão, sobretudo o “Modo de Conectividade de Dados” que precisa ficar como “Importar”. Ao final, clique em OK.

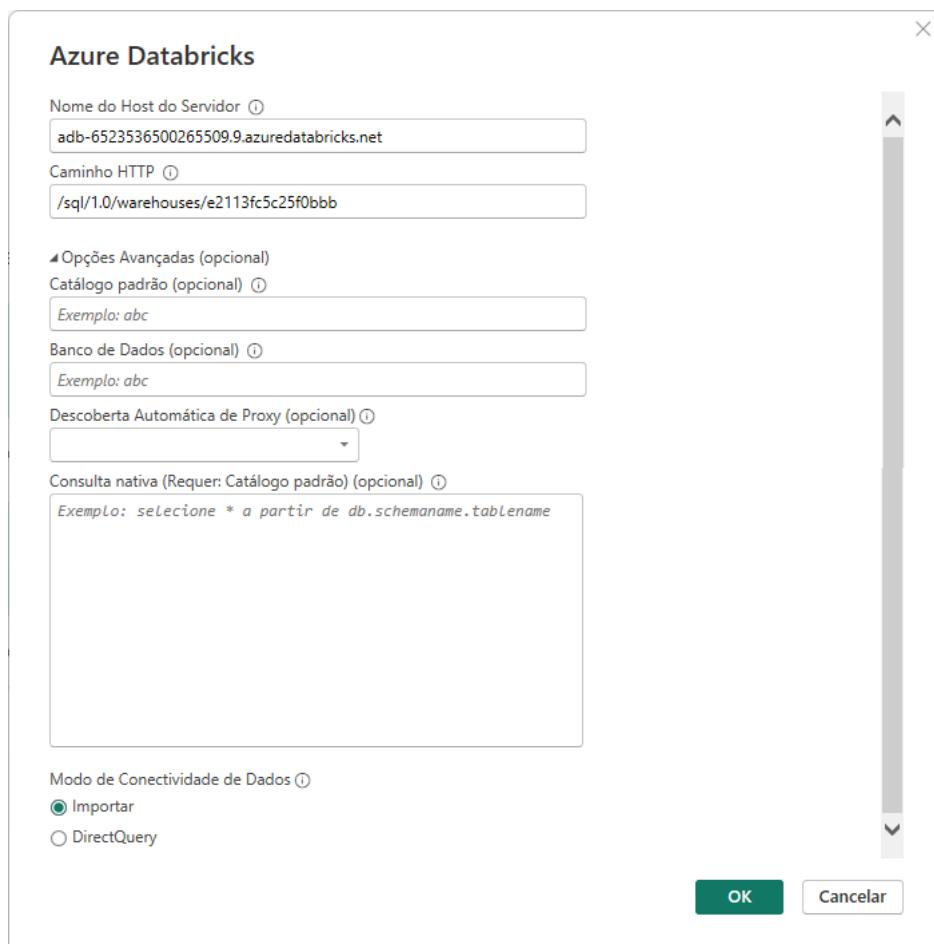


Figura 14 - Configuração do conector “Azure Databricks”

Título:

IT_73 Disponibilização de Conjuntos de Dados do Azure Databricks Unity no Power BI



Diretoria:

DSC/STI

Gerência Responsável:

Gerência de Tecnologia do Negócio

Data publicação:

12/07/2024

A tela seguinte será usada para a autenticação com o Azure Databricks, via Azure Active Directory. Nessa opção, clique em “Entrar” e siga com a autenticação da Microsoft. Ao final do processo de autenticação, clique em “Conectar”.

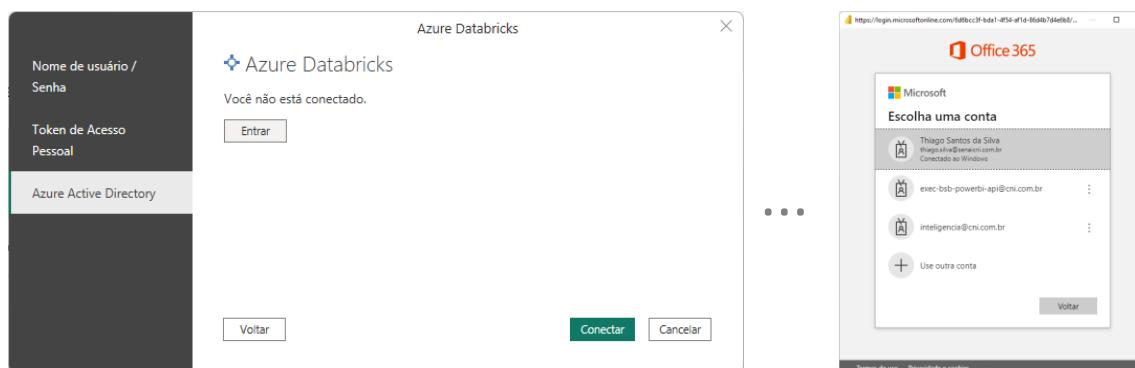


Figura 15 - Autenticação com o “Azure Databricks”

Na tela seguinte, poderá ser escolhido os ativos de interesse para que a conexão seja criada, e de forma análoga ao que foi mostrado nas figuras 6 e 7, concluir a etapa de conexão com os dados. Com isso, pode ser realizada a modelagem e desenvolvimento de visuais como em qualquer outro contexto de elaboração de painéis. Ao final, salve o arquivo “.pbix” e publique o painel no workspace ao qual tem acesso.

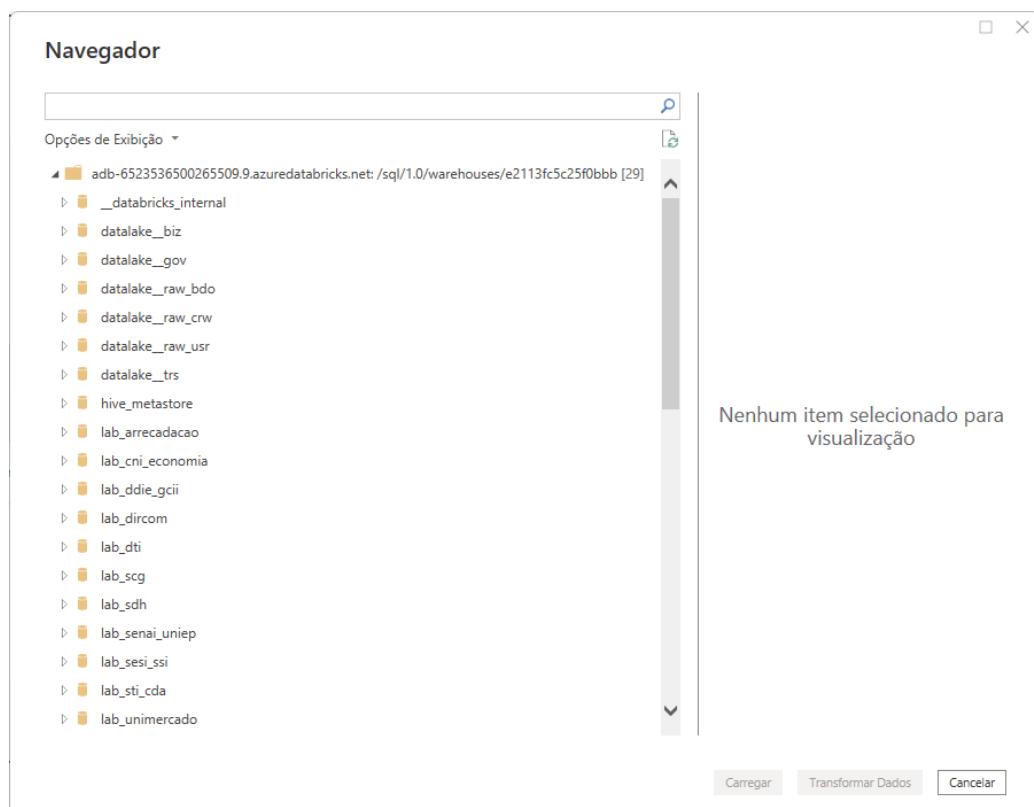


Figura 16 - seleção dos ativos do Unity para conexão com o Power BI

Título:

IT_73 Disponibilização de Conjuntos de Dados do Azure

Databricks Unity no Power BI



Diretoria:

DSC/STI

Gerência Responsável:

Gerência de Tecnologia do Negócio

Data publicação:

12/07/2024

Os demais parâmetros de “Opções Avançadas” da figura 14 são opcionais, mas podem ser úteis em alguns contextos, principalmente quando se precisa usar um script SQL para conectar aos dados. Isso pode ser feito para, por exemplo, fazer coisas como um SELECT apenas de algumas colunas ou a inclusão de uma cláusula WHERE para filtrar registros ou até mesmo um JOIN com outro ativo. Nesse contexto a configuração deve ser feita de maneira diferente, como mostrado na figura 16. Observe que ao usar uma Consulta nativa para incluir um script SQL, é obrigatório informar o Catálogo padrão (“database”). Além disso o nome do “schema” precisa estar no script.

The screenshot shows the 'Advanced Options' section of a Power BI dataset configuration. It includes fields for 'Default Catalog' (set to 'datalake_raw_bdo'), 'Native Query' (containing a SELECT statement), and a 'Catalog' pane on the right showing database and schema structures.

Catalog

- SQL-COAL Serverless 2XS
- Type to filter
- _databricks_internal
- datalake_biz
- datalake_gov
- datalake_raw_bdo** (highlighted)
- coal (highlighted)
- coal_integracao_os_vw_rdr_ds_com
- coal_integracao_os_vw_rdr_ds_grau
- coal_integracao_os_vw_rdr_ds_perfil
- coal_integracao_os_vw_rdr_ds_vota
- coal_integracao_os_vw_rdr_grupo
- coal_integracao_os_vw_rdr_perfilin
- coal_legisdata_cni_robo_legisdata_i
- coal_legisdata_cni_robo_legisdata_o
- coal_legisdata_cni_robo_legisdata_r

Figura 17 - Configuração das Opções Avançadas

Observação: a estratégia usada aqui é particularmente necessária quando um dataset já foi criado, já conta com outras tabelas do Unity ou até mesmo de outras fontes e que precisa ser incrementado com novas origens do Unity. Nesse caso, não será feita a substituição, mas o processo de conexão com os dados explicado acima (via **Server hostname** e **HTTP path**) será o mesmo, podendo ser usado o Power Query e o editor avançado de consultas, assim como demonstrado na [IT_54](#).

6. Utilização de Dados Pessoais (LGPD)

Embora a gestão dos dados que servem aos conjuntos de dados do Power BI não seja de alçada da STI, vale salientar que esses *datasets* podem conter dados sensíveis e a área de negócio deve usá-los apenas para o fim ao qual se propõe e zelar pelo seu compartilhamento em conformidade com o seu uso.

7. Conclusão

N/A