

# Cinco maneiras de melhorar o Power BI com o Azure Synapse Analytics



# Cinco maneiras de melhorar o Power BI com o Azure Synapse Analytics

3 /

Introdução

21 /

#3: Explore o data lake usando um ponto de extremidade sem servidor totalmente gerenciado

36 /

Resumo

6 /

#1: Crie e analise painéis do Power BI diretamente do Azure Synapse

30 /

#4: Crie pipelines de dados sem código para integrar mais fontes de dados e enriquecer insights

37 /

Próximas etapas

15 /

#2: Gerencie o data lake e crie data marts para relatórios de BI

33 /

#5: Crie um data warehouse seguro e conecte-se aos painéis do Power BI

# Introdução

Hoje, todas as empresas buscam fazer mais com seus dados. As organizações estão procurando migrar para além dos relatórios, para fornecer aos usuários maiores insights de dados.

Usando o Azure Synapse Analytics, as organizações podem utilizar a nuvem como o repositório central para todos os dados. As informações não se limitam a apenas colunas de dados; com o Azure Synapse, as empresas podem mover e analisar outras informações, incluindo imagens, som ou caches de documentos PDF. Essas informações podem ser organizadas para maximizar a capacidade da empresa de acessar os dados e fornecer informações reveladoras com o Power BI. Os dados usados no Power BI podem ser obtidos de todo o armazenamento de dados criado pelo Azure Synapse. O Azure Synapse é usado como um repositório de dados central, pois contém as ferramentas necessárias para carregar dados em um data lake, que pode ser usado para o aprendizado de máquina ou um banco de dados virtual do Power BI. Para grandes volumes de dados, o Azure Synapse pode criar data warehouses e data marts com recursos otimizados usados para geração de relatórios do Power BI.

Depois que o Azure Synapse processar os dados necessários em uma organização, a próxima etapa será implementar um ambiente de visualização de dados, como o Power BI. O Power BI capacita os usuários a obter mais insights sobre dados, criando um ambiente em que as pessoas podem interagir com dados para encontrar respostas para perguntas de negócios. O desenvolvimento dos modelos de dados usados no Power BI é o primeiro passo para fornecer acesso a dados de todas as partes da organização. Os dados podem ser obtidos de dispositivos IoT e mídias sociais, bem como uma série de aplicações que a organização usa no ambiente diário. Os dados que precisam ser relatados podem ser transmitidos, talvez com Hubs de Eventos do Azure, um data warehouse, banco de dados ou data lake. Os dados precisam estar em um local acessível para que possam ser usados não apenas para relatórios, mas também para processos de IA/ML do Advanced Analytics. Se os dados forem armazenados em um único local, eles estarão acessíveis não apenas para relatórios do Power BI, mas também para modelagem de ciência de dados.

Como muitos dados são necessários para essa tarefa, os dados precisam ser armazenados em um local seguro e escalável e, para atender a esses requisitos, os dados devem ser armazenados na nuvem. Centralizar o armazenamento de dados no Microsoft Azure fornece um local seguro, escalável e tolerante a falhas. Hoje, muitas empresas procuram armazenar e analisar dados sem texto. Vídeo, imagens e arquivos de som costumam ser analisados por meio de bibliotecas de IA dedicadas. Para essa grande variedade de dados, um data lake é muitas vezes a solução que melhor se adapta a essas necessidades.

Um data lake é o local lógico para armazenar dados de transações diárias, streaming de dados de localização GPS e arquivos de imagem, junto com outros dados que a organização deseja manter. A forma como os dados são armazenados determina a capacidade de análise. O Azure fornece uma estrutura de arquivos hierárquica que pode ser usada para análise, bem como o Power BI. Os cientistas de dados podem usar esses dados para criar modelos para fornecer insights adicionais, que você pode usar no Power BI. Os cientistas de dados também podem usar o Power BI para ajudá-los a analisar rapidamente dados para determinar quais elementos são necessários em um modelo. O Power BI oferece muitos recursos de business intelligence diferentes, muitos dos quais recém-adicionados, como metas, influenciadores importantes, árvores de decomposição e narrativa inteligente, que podem fornecer insights adicionais sobre os fatores que contribuem para o sucesso organizacional.

A tarefa de gerenciar dados usados para análise pode ser um desafio, pois os dados podem ser muito grandes e incluir dados estruturados armazenados em um data warehouse ou dados não estruturados, como imagens, e algo intermediário. O gerenciamento dos dados precisará incluir um método para adicionar mais dados e organizá-los em um formato necessário para tarefas de business intelligence. Os métodos de extração, carga e transferência podem ser usados para coletar dados de diferentes partes da organização ou aplicações de terceiros. Com base nas habilidades em sua organização, a transferência de dados pode acontecer com soluções Low Code, fluxos de dados ou código Python. Não importa onde a origem dos dados está, pois você precisa para coletá-los para conseguir fornecer insights operacionais. Para fornecer respostas a perguntas de negócios, sua organização precisa desenvolver soluções para fornecer todas as informações necessárias para a tomada de decisão precisa. Os elementos de aprendizado de máquina podem aprimorar ainda mais o processo de tomada de decisão, fornecendo análises preditivas para determinar o estado futuro da empresa ou apontar anomalias na performance que podem não ter sido descobertas.

A tarefa de fornecer esses dados para que o Power BI possa acessá-los e os insights que outros possam fornecer com os dados está disponível em uma ferramenta, o [Azure Synapse Analytics](#), soluções analíticas unificadas. O Power BI pode ser usado para relatar dados de um data lake e o Azure Synapse fornece a capacidade de ingerir dados de várias soluções, criar grandes data warehouses, gerenciar data lakes, criar data marts direcionados para públicos de relatórios de BI, desenvolver soluções de aprendizado de máquina e fornecer ao Power BI as informações necessárias para migrar do fornecimento de relatórios para soluções de análise de negócios.

Neste artigo, vamos examinar cinco maneiras diferentes de usar o Azure Synapse e o Power BI:

- #1: Crie e analise painéis do Power BI diretamente do Azure Synapse
- #2: Gerencie o data lake e crie data marts para relatórios de BI
- #3: Explore o data lake usando um ponto de extremidade sem servidor totalmente gerenciado
- #4: Crie pipelines de dados sem código para integrar mais fontes de dados e enriquecer insights
- #5: Crie um data warehouse seguro e conecte-se aos painéis do Power BI

Essas são cinco maneiras diferentes em que você pode usar o Azure Synapse com o Power BI para fornecer mais insights sobre como o Azure Synapse pode ser usado para fornecer dados à sua organização. Vamos começar usando o Power BI para analisar dados no Azure Synapse.



# #1: Crie e analise painéis do Power BI diretamente do Azure Synapse

O Power BI fornece uma série de ferramentas para ajudar a analisar dados que podem já estar no Azure Synapse. O uso de uma ferramenta visual para examinar o estado dos dados atuais pode fornecer uma maneira rápida de determinar o status atual dos dados armazenados em diferentes áreas do Azure Synapse, incluindo pools sem servidor, pools dedicados ou para examinar os dados criados em um processo de carregamento provisório. Para adicionar a capacidade de usar o Power BI no Azure Synapse, você precisará de um espaço de trabalho do Azure Synapse e uma instância do Power BI no mesmo locatário do Azure. A criação de um espaço de trabalho do Azure Synapse leva alguns minutos e as etapas desse processo podem ser encontradas [aqui](#). Talvez os dados precisem ser examinados para revisar a qualidade dos dados recebidos para determinar o tipo de análise possível. Talvez você tenha recebido alguns dados de um cliente e queira fazer um relatório rápido para determinar como incorporar os dados em outro modelo de dados.

---

Saiba como criar um espaço de trabalho do Azure Synapse com [este tutorial em vídeo](#).

---

O Azure Synapse permite conectividade com outras aplicações por meio de serviços vinculados. Você pode incluir uma variedade de aplicações, incluindo o Azure Key Vault, o Azure Machine Learning e, claro, o Power BI. Os serviços vinculados fornecem as informações de conexão que permitem acessar a funcionalidade do Power BI no Azure Synapse. O serviço vinculado permite que você faça uma conexão com um espaço de trabalho no Power BI. Em versões anteriores, foi possível se conectar a um espaço de trabalho do Power BI. Atualmente, é possível se conectar a vários espaços de trabalho do Power BI.

No Azure Synapse, você pode ter vários outros serviços vinculados para elementos como uma conta adicional do ADLS ou um conjunto de dados publicamente disponível. Para criar uma conexão de serviço vinculado ao Power BI, selecione o ícone **Gerenciar** no lado esquerdo do portal do Azure Synapse e, depois, em **Conexões externas**, selecione **Serviços vinculados**, conforme mostrado na *Figura 1*. O serviço vinculado do Power BI tem um quadrado roxo em torno dele e é chamado de **PowerbiAcctDev**:

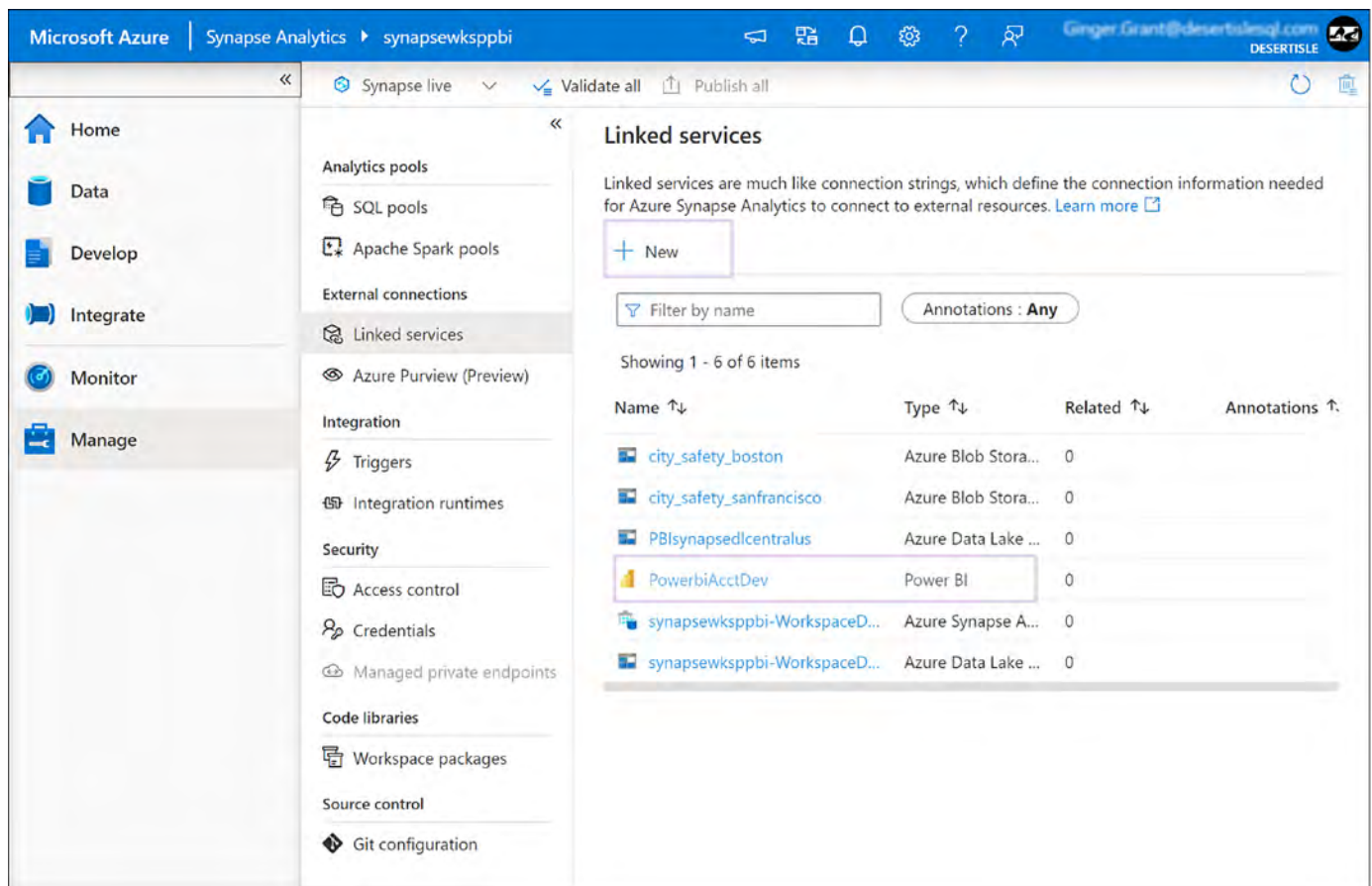


Figura 1: conexões de serviços vinculados criadas nesta conta do Azure Synapse Analytics

Observe que o serviço vinculado do Power BI é chamado de **PowerbiAcctDev**, que é um espaço de trabalho no mesmo locatário que a conta do Azure usada neste exemplo. A conta do Azure usada deve estar no mesmo locatário que a conta do Power BI. Por exemplo, se tivermos uma conta do Azure vinculada a uma conta Outlook.com, não poderemos vincular essa conta ao nosso locatário desertislesql.com.

Para adicionar um novo serviço vinculado, clique no texto **+ Novo** no painel **Serviços vinculados**. Se você ainda não criou uma conexão com um espaço de trabalho do Power BI, seu menu será semelhante a este:

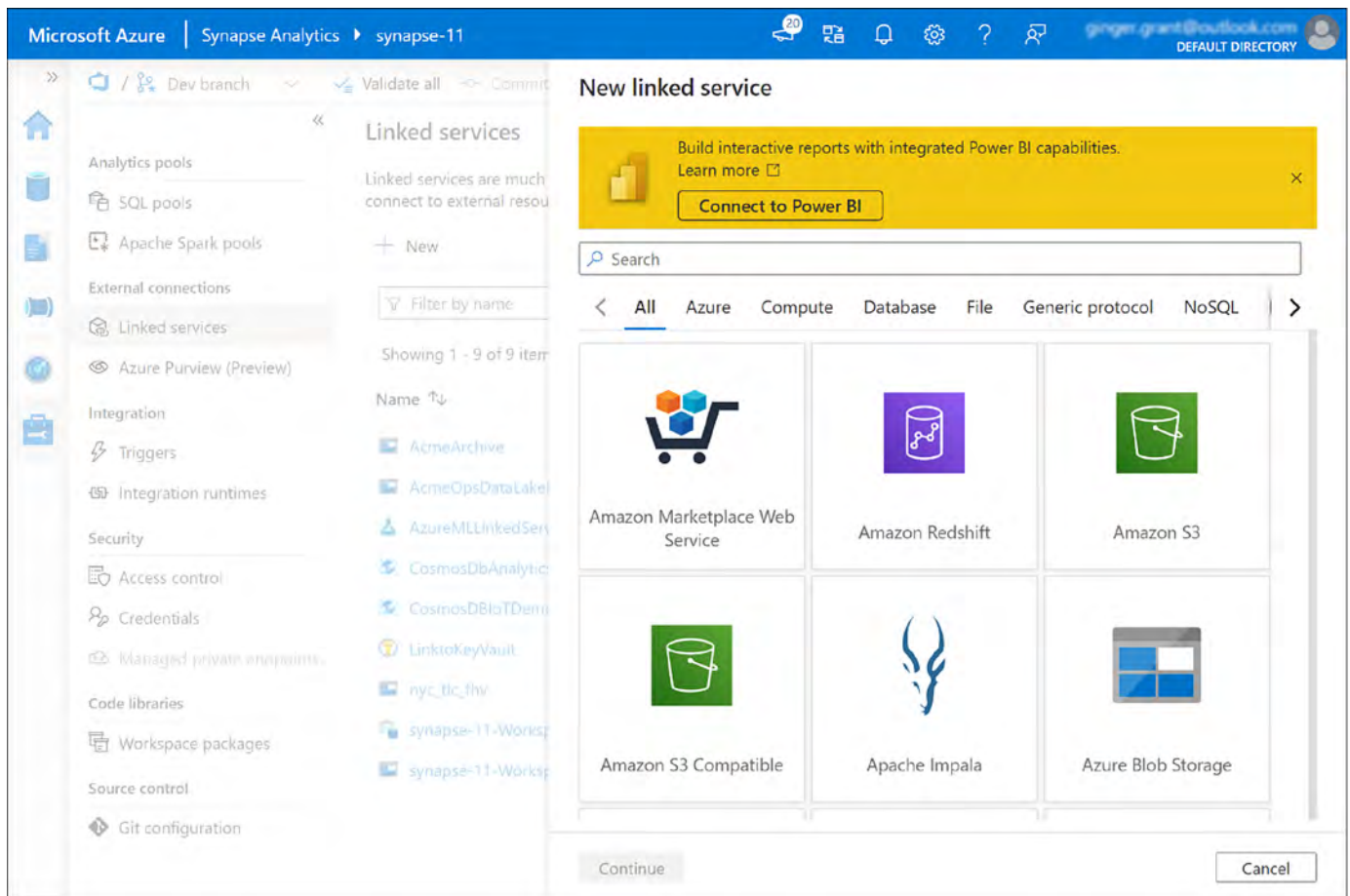


Figura 2: menu Novo serviço vinculado do Azure Synapse Analytics



Clique no botão **Conectar ao Power BI**. O Power BI também está listado nos menus; basta clicar na seta no lado direito do painel para acessar a opção de menu. Ambas as opções levam você para a mesma tela, mostrada na *Figura 3*, que você precisa concluir para conectar o Azure Synapse a um espaço de trabalho do Power BI. Você pode nomear o serviço vinculado no Azure Synapse da forma desejada, mas saiba que o nome não poderá ser alterado posteriormente:

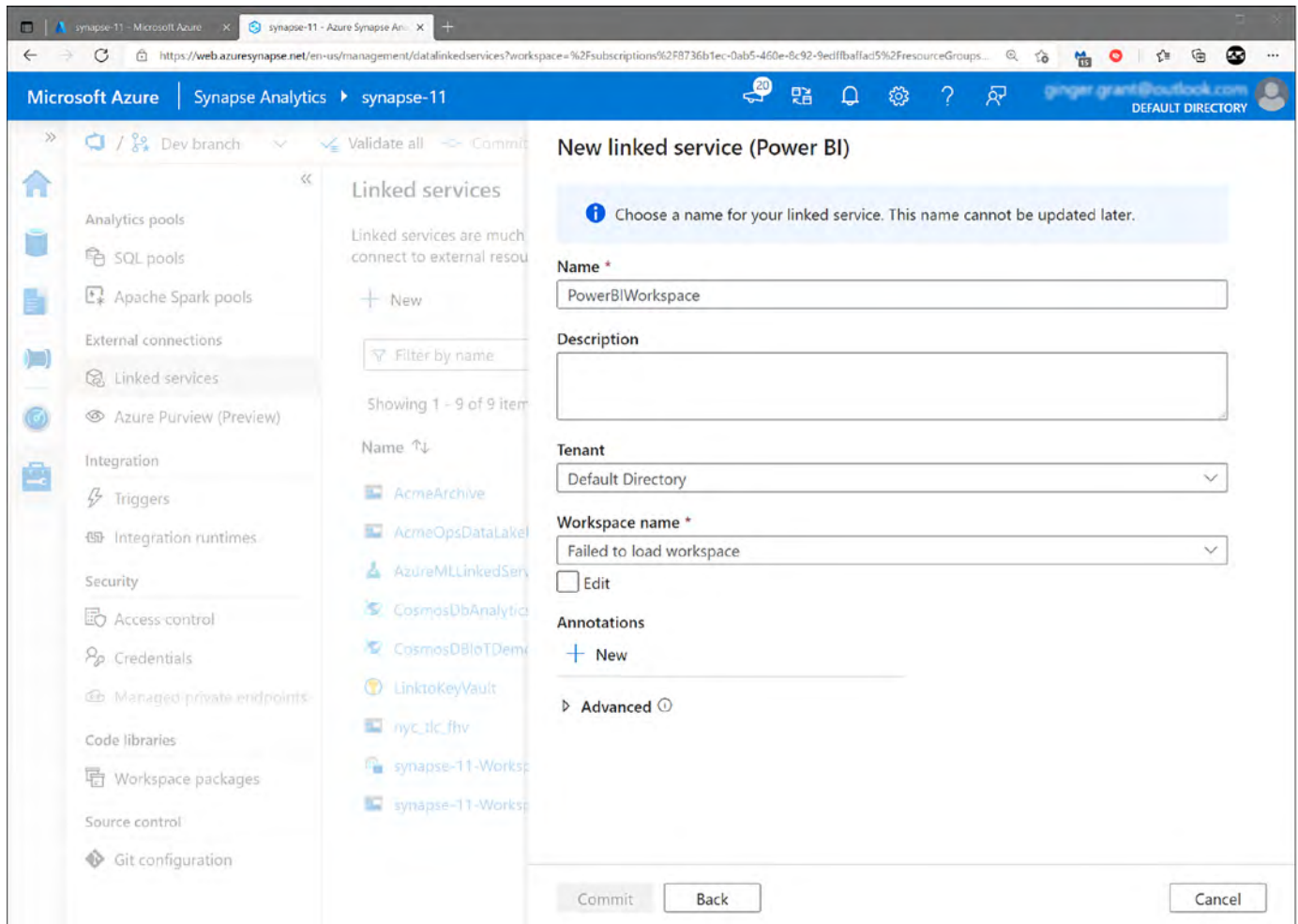


Figura 3: conectar um espaço de trabalho do Power BI ao Azure Synapse usando serviços vinculados

Você precisará selecionar o **Locatário** da lista suspensa e, depois, a lista suspensa do **Nome do espaço de trabalho** será preenchida com todos os espaços de trabalho aos quais você tem acesso. Os dados em um espaço de trabalho podem ser obtidos de vários locais diferentes, incluindo ADLS, e também de locais como um arquivo do Excel armazenado no OneDrive ou um espaço de trabalho SQL do Azure. Ao concluir este formulário, clique no botão **Confirmar**. Você publicará o novo serviço vinculado para concluir o processo de adicioná-lo.

Após a adição do Power BI ao Azure Synapse, você poderá usá-lo no painel **Desenvolver**. Você notará que uma seção **Power BI** foi adicionada. Nela, você encontra os conjuntos de dados e relatórios contidos em seu espaço de trabalho. Você pode usar o Power BI no Azure Synapse para modificar um relatório existente e, ao fazê-lo, certifique-se de salvá-lo para garantir que seja disponibilizado para todos que têm acesso ao espaço de trabalho. A *Figura 4* mostra como modificar um conjunto de dados existente no Power BI:

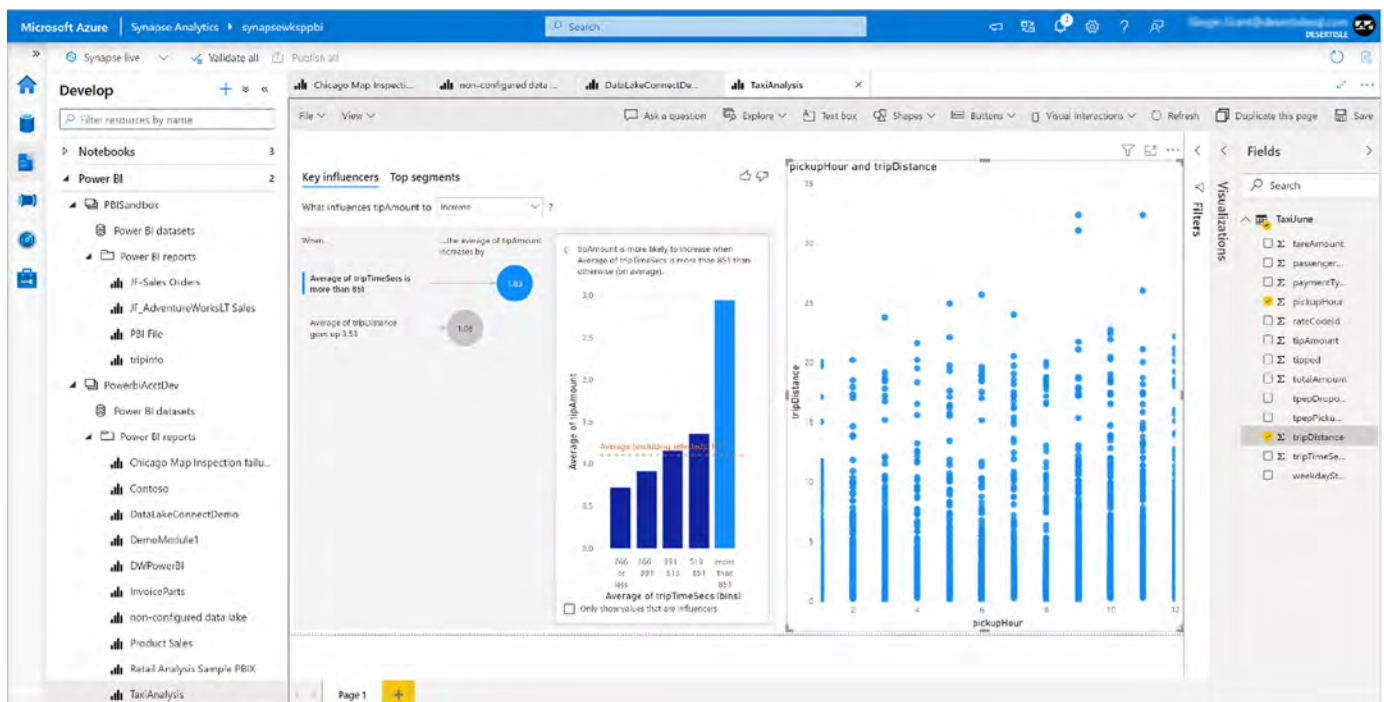


Figura 4: desenvolvimento do relatório de serviços vinculados do Power BI no Azure Synapse

Os conjuntos de dados são listados em sua própria janela, que também é o local para a criação de novos relatórios no Azure Synapse para Power BI. Se você passar o cursor sobre um dos nomes de conjunto de dados na lista, verá dois ícones, que são mostrados na *Figura 5*:

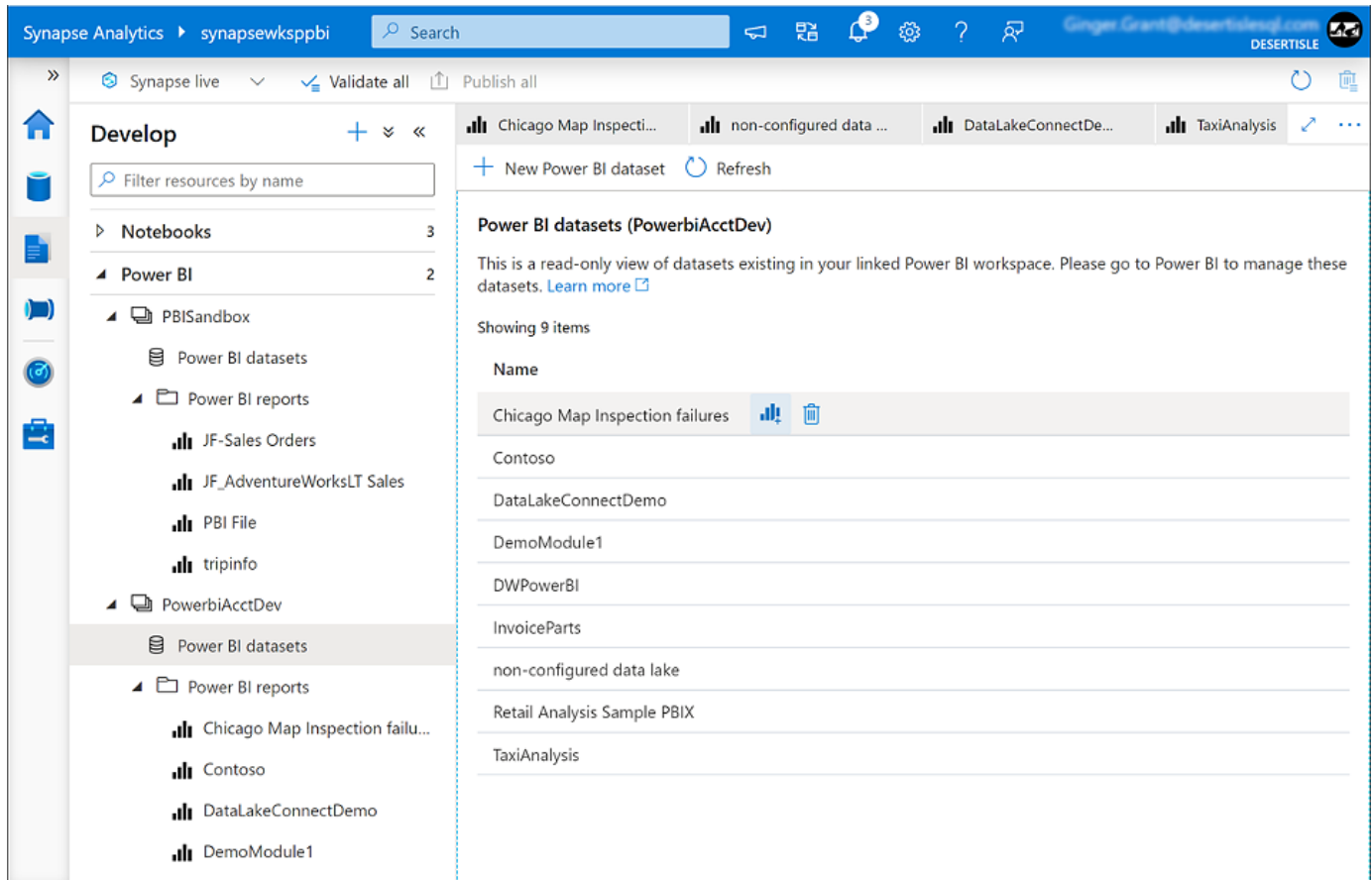


Figura 5: exibição do conjunto de dados de espaços de trabalho do Power BI no Azure Synapse usado para criar novos relatórios

O primeiro permitirá que você adicione um novo relatório com base no conjunto de dados destacado ou pode excluir o conjunto de dados. O uso do Power BI para analisar dados no Azure Synapse pode acelerar o tempo que leva para examinar dados em modelos de dados e usar qualquer modelo de dados combinado a ser criado no Power BI. No Azure Synapse, todos os modelos no espaço de trabalho podem ser usados. Quando novos modelos são carregados no espaço de trabalho, eles estão disponíveis no Azure Synapse. Talvez seja necessário atualizar o conjunto de dados usando o botão **Atualizar** na parte superior da tela mostrada na *Figura 5* para que os novos modelos sejam exibidos.

## Uso de um exemplo de conjunto de dados do Centro de conhecimento

Se você ainda não tiver um conjunto de dados a ser usado com o Azure Synapse, poderá usar um dos conjuntos de dados de exemplo oferecido no Centro de conhecimento, que pode ser encontrado clicando no botão **Aprender** na página inicial do espaço de trabalho, conforme mostrado na *Figura 6*, que é destacada em roxo:

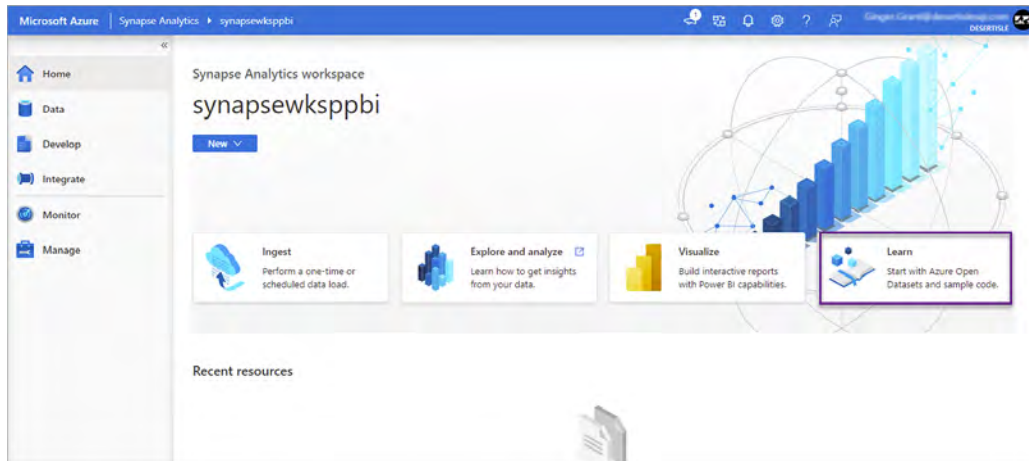


Figura 6: o Azure Synapse Learn contém o Centro de conhecimento, onde você pode saber mais sobre o Azure Synapse

O Azure Synapse fornece vários exemplos de codificação e conjuntos de dados a serem usados para se familiarizar com o Azure Synapse. Para acessá-los, após selecionar **Aprender**, a página Centro de conhecimento será carregada. No Centro de conhecimento, selecionaremos **Procurar na galeria** e carregaremos um banco de dados de exemplo em um novo banco de dados lógico do Azure Synapse. Você verá **Procurar na galeria** destacado em roxo na *Figura 7*:

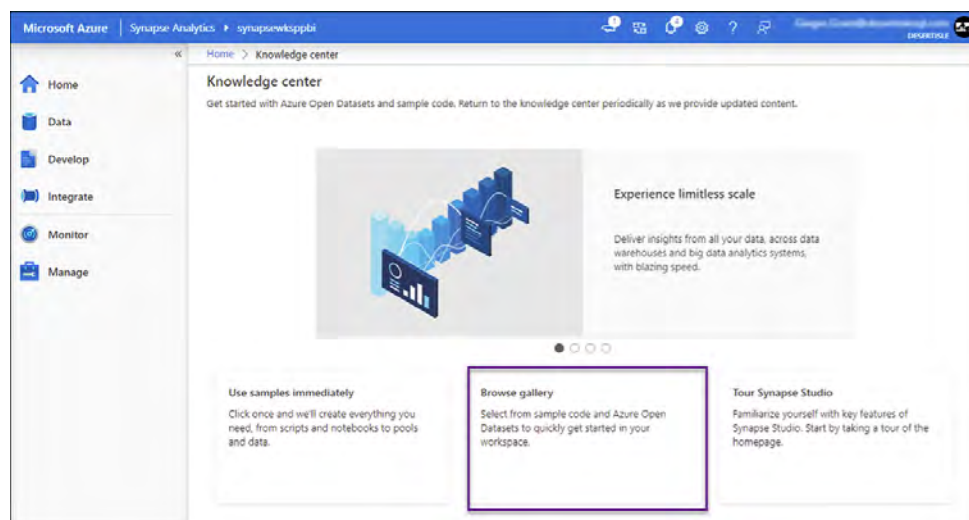


Figura 7: a seção Procurar na galeria do Centro de conhecimento

No Centro de conhecimento, selecionaremos o exemplo de **horas e ganhos de emprego de estado nos EUA** e, depois, clicaremos no botão azul **Continuar**. Uma breve descrição aparecerá e um botão azul **Adicionar conjunto de dados** aparecerá no canto superior esquerdo da janela, que você deve clicar. O conjunto de dados aparecerá em **Armazenamento de Blobs do Azure**. Clique nas reticências (os três pontos) ao lado da pasta **us-employment-hour-earnings** e selecione **Novo script SQL** seguido de **Criar tabela externa**, conforme mostrado na *Figura 8*:

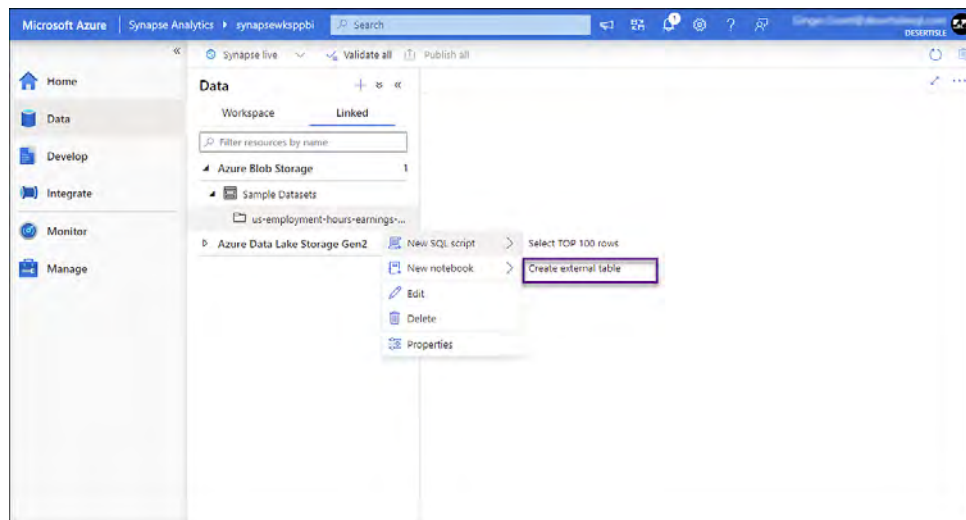


Figura 8: crie tabela externa a partir do exemplo de conjunto de dados Horas e ganhos de emprego de estado nos EUA

Uma janela pop-up aparecerá fornecendo as solicitações necessárias para criar uma tabela externa. Crie um novo banco de dados chamado **USEmp** e para **Nome de tabela externa**, use **HoursData**, como mostrado na *Figura 9*:

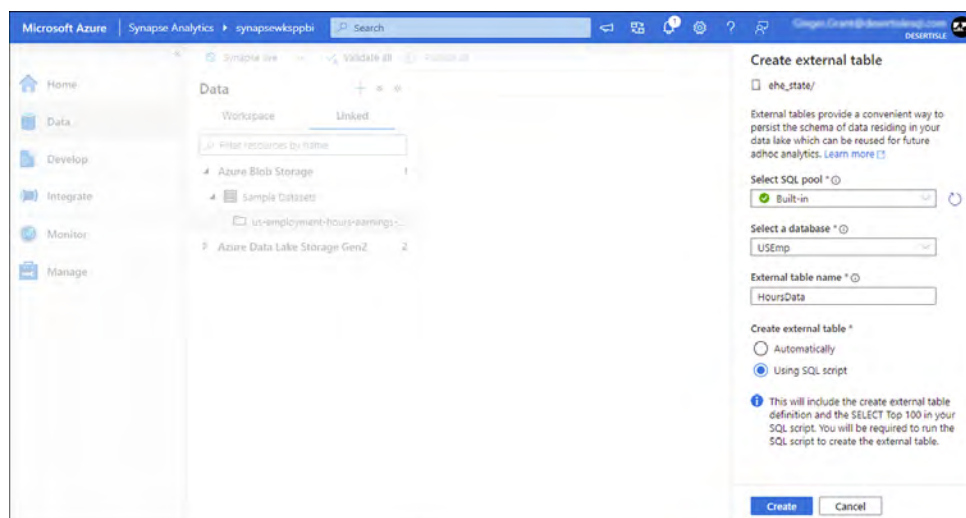
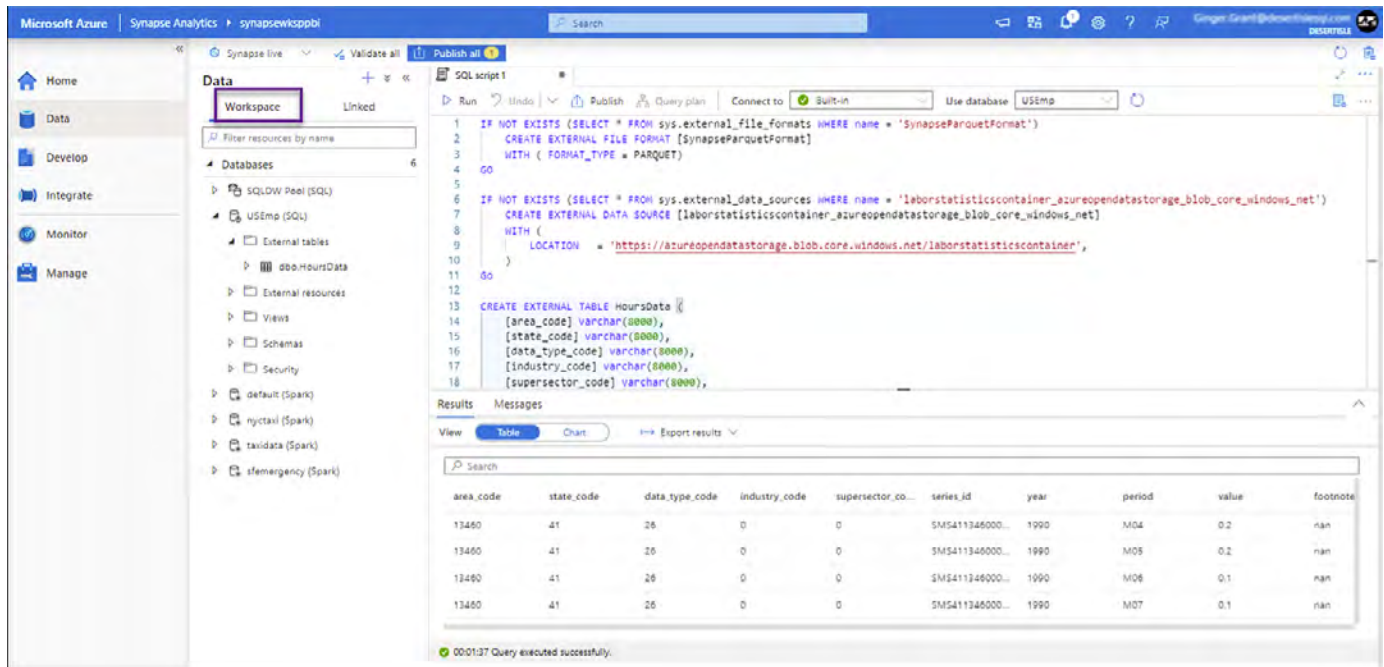


Figura 9: crie avisos de tabela externos



Certifique-se de que o botão de opção, **Usando o script SQL**, esteja selecionado, como mostrado aqui, antes de selecionar o botão **Criar**. Verifique o código gerado na janela **Script SQL 1** pois isso pode ser um guia para criar tabelas externas. Selecione a guia **Espaço de trabalho**, que é destacada em roxo na *Figura 10*, e atualize os bancos de dados selecionando as reticências e você verá o novo banco de dados **USEmp**:



```
1 IF NOT EXISTS (SELECT * FROM sys.external_file_formats WHERE name = 'SynapseParquetFormat')
2   CREATE EXTERNAL FILE FORMAT [SynapseParquetFormat]
3   WITH ( FORMAT_TYPE = PARQUET)
4   GO
5
6 IF NOT EXISTS (SELECT * FROM sys.external_data_sources WHERE name = 'laborstatisticscontainer_azureopendatastorage_blob_core_windows_net')
7   CREATE EXTERNAL DATA SOURCE [laborstatisticscontainer_azureopendatastorage_blob_core_windows_net]
8   WITH (
9     LOCATION = 'https://azureopendatastorage.blob.core.windows.net/laborstatisticscontainer',
10  )
11  GO
12
13 CREATE EXTERNAL TABLE HoursData (
14   [area_code] varchar(8000),
15   [state_code] varchar(8000),
16   [data_type_code] varchar(8000),
17   [industry_code] varchar(8000),
18   [supersector_code] varchar(8000),
19 )
```

area_code	state_code	data_type_code	industry_code	supersector_code	series_id	year	period	value	footnote
13450	41	26	0	0	SM5411346000...	1990	M04	0.2	nan
13450	41	26	0	0	SM5411346000...	1990	M05	0.2	nan
13450	41	26	0	0	SM5411346000...	1990	M06	0.1	nan
13450	41	26	0	0	SM5411346000...	1990	M07	0.1	nan

Figura 10: banco de dados lógico do Azure Synapse recém-criado com a tabela HoursData mostrada

O código selecionado gerará um novo banco de dados e uma tabela que você pode usar no Power BI. A tabela externa contém o conjunto de dados que você pode conectar ao Power BI e usá-lo em uma visualização.

Em seguida, mostraremos como o Azure Synapse gerencia o data lake e trabalha com o Power BI para obter relatórios mais significativos.

## #2: Gerencie o data lake e crie data marts para relatórios de BI

Muitas empresas hoje trabalham com data lakes para o armazenamento a longo prazo de todos os dados existentes em seu ambiente. O Azure Synapse pode ser uma ótima solução, pois fornece a capacidade de criar pipelines de integração que podem organizar dados em um data lake em diferentes áreas funcionais para diferentes casos de uso.

A empresa Acme é uma empresa de vendas pela Internet que tem vários itens armazenados em depósitos que precisarão ser embalados e enviados depois que uma ordem for lançada. Embora deseje apenas criar relatórios diários dos últimos dois anos, ela precisa armazenar o restante dos dados de forma segura e organizada para que os dados estejam disponíveis para análise ad hoc ou aprendizado de máquina. Os dados podem ser recebidos de dispositivos IoT e podem precisar ser relatados de forma resumida e armazenados no formato bruto para posterior análise de aprendizado de máquina.

O Azure Synapse contém todas as ferramentas necessárias para organizar dados em um data lake para que os dados possam ser disponibilizados para muitos usos diferentes. Os dados podem ser obtidos de diferentes fontes para o data lake. Usando um pipeline de integração, os dados podem ser transformados de dados brutos armazenados na área de bronze do data lake, adicionados em alguns metadados na área prateada do data lake e selecionados para um conjunto de arquivos modelados em um esquema de estrela do data warehouse. O Azure Synapse pode ajudar a organizar dados brutos para que possam ser melhor utilizados por organizações, fornecendo a estrutura para revisar e organizar data lakes. Depois que o Azure Synapse coletar os dados, eles estarão prontos para serem usados pelo Power BI.

Os dados armazenados no data lake costumam ser armazenados em formato parquet, pois esse formato fornece compactação de dados em colunas e pode ser consultado mais rapidamente usando o ponto de extremidade sem servidor padrão e lido mais rapidamente no Azure Synapse. Isso não é um impedimento para acessar os dados no Power BI. Pelo contrário, ele fornece um método para garantir que você possa carregar talvez 24 meses ininterruptos de dados, carregando apenas esses dados no modelo. Vamos examinar como podemos nos conectar ao data lake que foi coletado com o Azure Synapse.

O acesso ao data lake criado pelo Azure Synapse é configurado primeiro no portal de administração do Power BI. No momento em que este documento foi escrito, esta era uma configuração de recurso de visualização no portal de administrador do Power BI. Para acessar os dados, um usuário com acesso administrativo do Power BI precisará usar o portal de administração e selecionar a opção **Conexões do Azure**, como mostrado na *Figura 11*. Ao se conectar a uma conta do Azure Data Lake Storage Gen 2, o método mais econômico para usar os dados no Power BI é garantir que o data lake e o locatário do Power BI estejam no mesmo data center. Você também precisará do nome da **Assinatura** do Azure, do **Grupo de recursos** e do nome da sua **Conta de armazenamento** do Azure Data Lake para preencher as informações na seção **Conexões do Azure** do portal de administração, conforme mostrado na *Figura 11*:

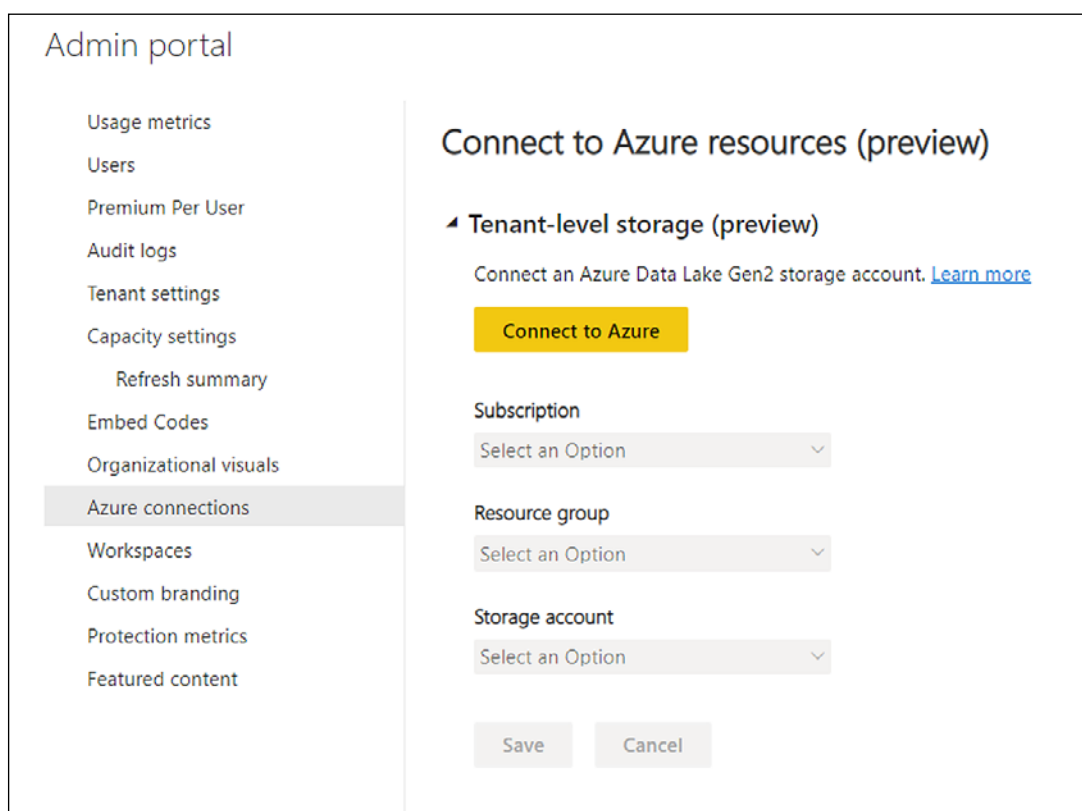


Figura 11: portal de administração do Power BI para conexões do Azure e Azure Data Lake Gen2

Após essa conexão, você poderá acessar os dados no Power BI. Usamos o nome do grupo de recursos **SynapseRGPBI** e a conta do Azure Data Lake Storage Gen 2 é **synapsedlpbi**. Depois que a conexão do Azure for criada e salva, os dados poderão ser atualizados no serviço do Power BI.

---

**Observação:** quando este documento foi escrito, você só podia se conectar a uma conta do Azure Data Lake Gen2 Storage (ADLS) e não a outros tipos de conta de armazenamento.

---



Na área de trabalho do Power BI, as contas do Azure Data Lake Gen 2 podem ser acessadas selecionando-as a partir das opções do Azure no menu **Obter Dados**. Você precisará acessar os dados com a URL com o local dos dados do ADLS usando este padrão:

```
https://<ADLS_accountname>.dfs.core.windows.net/<filesystemname>/<subfolder>
```

Isso extrairá todos os dados da pasta. Por exemplo, temos uma conta do ADLS chamada **PBISynapseCentralUS**. Ela tem uma pasta chamada **powerbi** e há uma subpasta chamada **Bronze**. Para acessar todos os arquivos desta pasta, utilizaríamos a seguinte URL no Power BI:

```
https://synapsedlcentralus.dfs.core.windows.net/powerbi/Bronze/  
Chicagofoodinspections/2020FoodInspectionsparquet
```

A URL seria inserida como mostrado na *Figura 12*:



Azure Data Lake Storage Gen2

URL  
[.windows.net/powerbi/Bronze/Chicagofoodinspections/2020FoodInspectionsparquet]

Data View

☒ File System View

☐ CDM Folder View (Beta)

OK Cancel

Figura 12: obter dados do Azure Data Lake Storage Gen2 usando uma URL

Ao especificar o local, certifique-se de que exista uma correspondência exata das letras maiúsculas e minúsculas. Se não coincidirem, você obterá um erro "Não encontrado". Como não especificamos um nome de arquivo, todos os arquivos nesse diretório foram devolvidos. Nosso subdiretório contém três arquivos Parquet. Os arquivos Parquet são comumente usados com o Azure Synapse, pois agilizarão a performance ao criar tabelas em pools sem servidor e diminuirão o tamanho do armazenamento. Esses arquivos são binários e ilegíveis por humanos, mas podem ser lidos pelo

Power BI. Há três arquivos Parquet no subdiretório listado. Eles devem ser lidos como um elemento de dados, os dados de inspeção 2020 Chicago Food. Vejamos isso no Power BI, ilustrado na *Figura 13*:

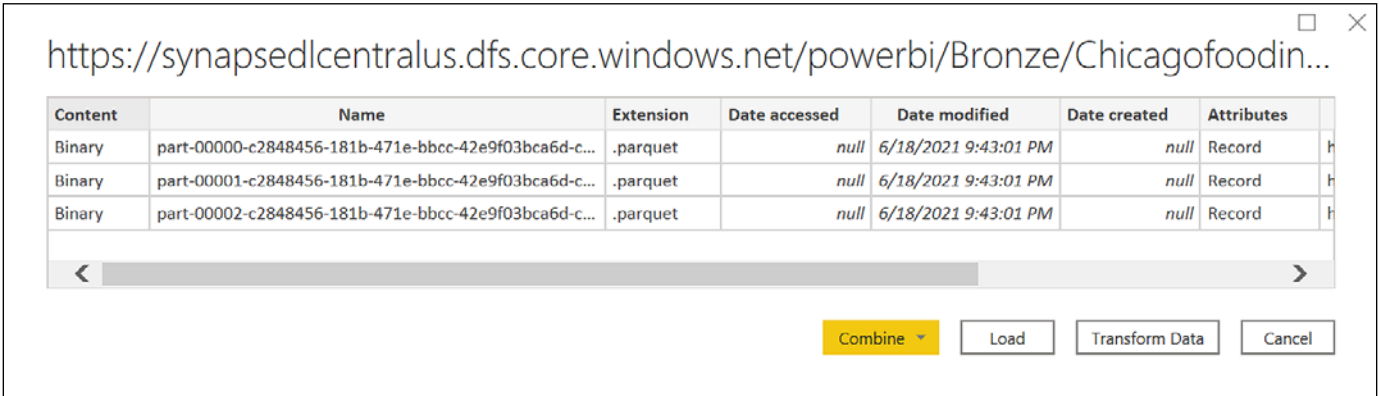


Figura 13: arquivos Parquet a serem carregados do ADLS no Power BI

O botão amarelo **Combinar** nos permitirá juntar os arquivos e incorporar o conteúdo em nosso modelo de dados. Vamos selecionar **Combinar** e **Transformar Dados** para coletar os dados e determinar se precisamos fazer algo antes de carregá-lo no Power BI. Esta etapa carregará todos os dados e os transformará de Parquet em um formato legível por humanos, mostrado na *Figura 14*:

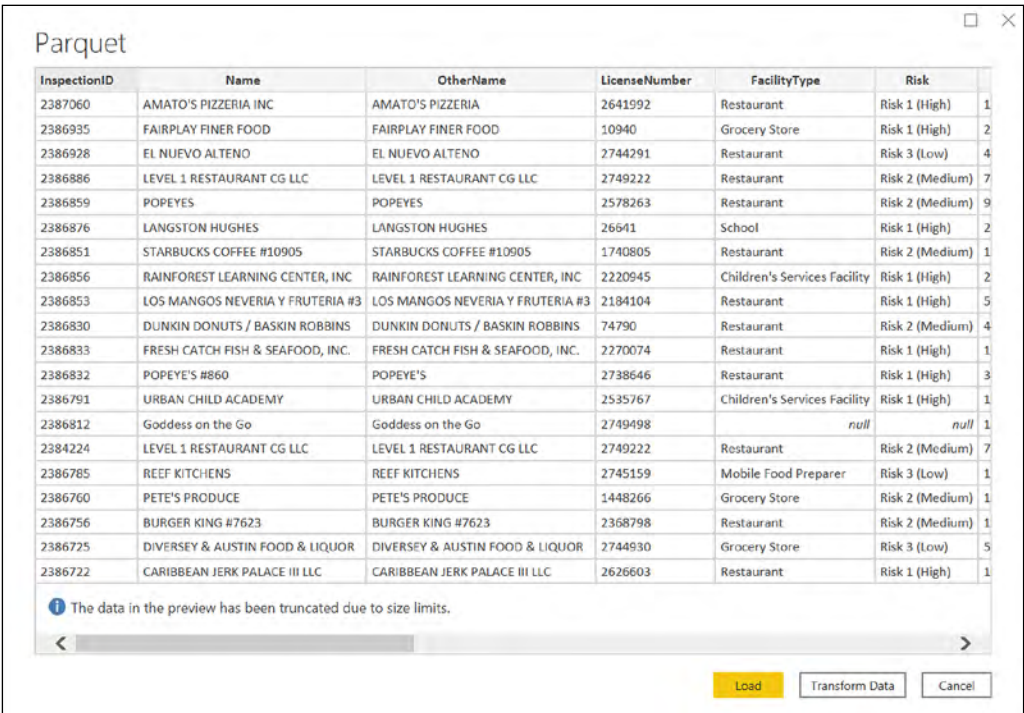


Figura 14: amostra de arquivos Parquet do ADLS transformados em Power BI

Note que é fácil ler os dados; então, vamos transformá-los. A *Figura 15* mostra que, no lado esquerdo da tela, o Power BI executou automaticamente várias etapas de transformação diferentes nos dados:

The screenshot displays the Power Query Editor window. The main area shows a table with the following data:

	Source.Name	InspectionID	Name	OtherName
1	part-00000-c2848456-181b-4...	2387060	AMATO'S PIZZERIA INC	AMATO'S PIZZERIA
2	part-00000-c2848456-181b-4...	2386935	FAIRPLAY FINER FOOD	FAIRPLAY FINER FOOD
3	part-00000-c2848456-181b-4...	2386928	EL NUEVO ALTENO	EL NUEVO ALTENO
4	part-00000-c2848456-181b-4...	2386886	LEVEL 1 RESTAURANT CG LLC	LEVEL 1 RESTAURANT CG LLC
5	part-00000-c2848456-181b-4...	2386859	POPEYES	POPEYES
6	part-00000-c2848456-181b-4...	2386876	LANGSTON HUGHES	LANGSTON HUGHES
7	part-00000-c2848456-181b-4...	2386851	STARBUCKS COFFEE #10905	STARBUCKS COFFEE #10905
8	part-00000-c2848456-181b-4...	2386856	RAINFOREST LEARNING CENTER, INC	RAINFOREST LEARNING CENTE
9	part-00000-c2848456-181b-4...	2386853	LOS MANGOS NEVERIA Y FRUTERIA #3	LOS MANGOS NEVERIA Y FRUT
10	part-00000-c2848456-181b-4...	2386830	DUNKIN DONUTS / BASKIN ROBBINS	DUNKIN DONUTS / BASKIN RO
11	part-00000-c2848456-181b-4...	2386833	FRESH CATCH FISH & SEAFOOD, INC.	FRESH CATCH FISH & SEAFOOD
12	part-00000-c2848456-181b-4...	2386832	POPEYE'S #860	POPEYE'S
13	part-00000-c2848456-181b-4...	2386791	URBAN CHILD ACADEMY	URBAN CHILD ACADEMY
14	part-00000-c2848456-181b-4...	2386812	Goddess on the Go	Goddess on the Go
15	part-00000-c2848456-181b-4...	2384224	LEVEL 1 RESTAURANT CG LLC	LEVEL 1 RESTAURANT CG LLC
16	part-00000-c2848456-181b-4...	2386785	REEF KITCHENS	REEF KITCHENS
17	part-00000-c2848456-181b-4...	2386760	PETE'S PRODUCE	PETE'S PRODUCE
18	part-00000-c2848456-181b-4...	2386756	BURGER KING #7623	BURGER KING #7623
19	part-00000-c2848456-181b-4...	2386725	DIVERSEY & AUSTIN FOOD & LIQUOR	DIVERSEY & AUSTIN FOOD & LI
20	part-00000-c2848456-181b-4...	2386722	CARIBBEAN JERK PALACE III LLC	CARIBBEAN JERK PALACE III LLC
21	part-00000-c2848456-181b-4...	2386695	WALGREENS # 09000	WALGREEN'S
22	part-00000-c2848456-181b-4...	2386687	HAGEN'S FISH MARKET, INC	HAGEN'S FISH MARKET
23	part-00000-c2848456-181b-4...	2386631	LAS ISLAS MARIAS	LAS ISLAS MARIAS
24	part-00000-c2848456-181b-4...	2386636	CAFFE UMBRIA, INC	CAFFE UMBRIA
25				

The interface includes a ribbon with various transformation options (Close & Apply, New Source, Recent Sources, Enter Data, Data source settings, Manage Parameters, Refresh Preview, Advanced Editor, Choose Columns, Remove Columns, Keep Rows, Remove Rows, Sort, Split Column, Group By, Data Type: Text, Use First Row as Headers, Replace Values, Merge Queries, Append Queries, Combine Files, Text Analytics, Vision, Azure Machine Learning, AI Insights). The sidebar on the left shows the Queries list with 'Query1' selected. The right sidebar shows the Query Settings for 'Query1', including the Properties (Name: Query1) and the Applied Steps (Source, Filtered Hidden Files1, Invoke Custom Function1, Renamed Columns1, Removed Other Columns1, Expanded Table Column1, Changed Type).

Figure 15: etapas transformacionais criadas no Power Query

As consultas de auxiliares obtêm os dados da conta do ADLS e os transformam em formato legível por humanos. A **Consulta1** capta as alterações feitas. Aqui, removeremos as colunas **Source.Name** e **InspectionID**, pois elas não serão usadas no modelo de dados. Renomearemos **Consulta1** como **Inspeção**. Usando o Power BI, podemos usar os dados armazenados no formato ideal para o Azure Synapse e usar o Power BI para relatar os dados, conforme mostrado na *Figura 16*:

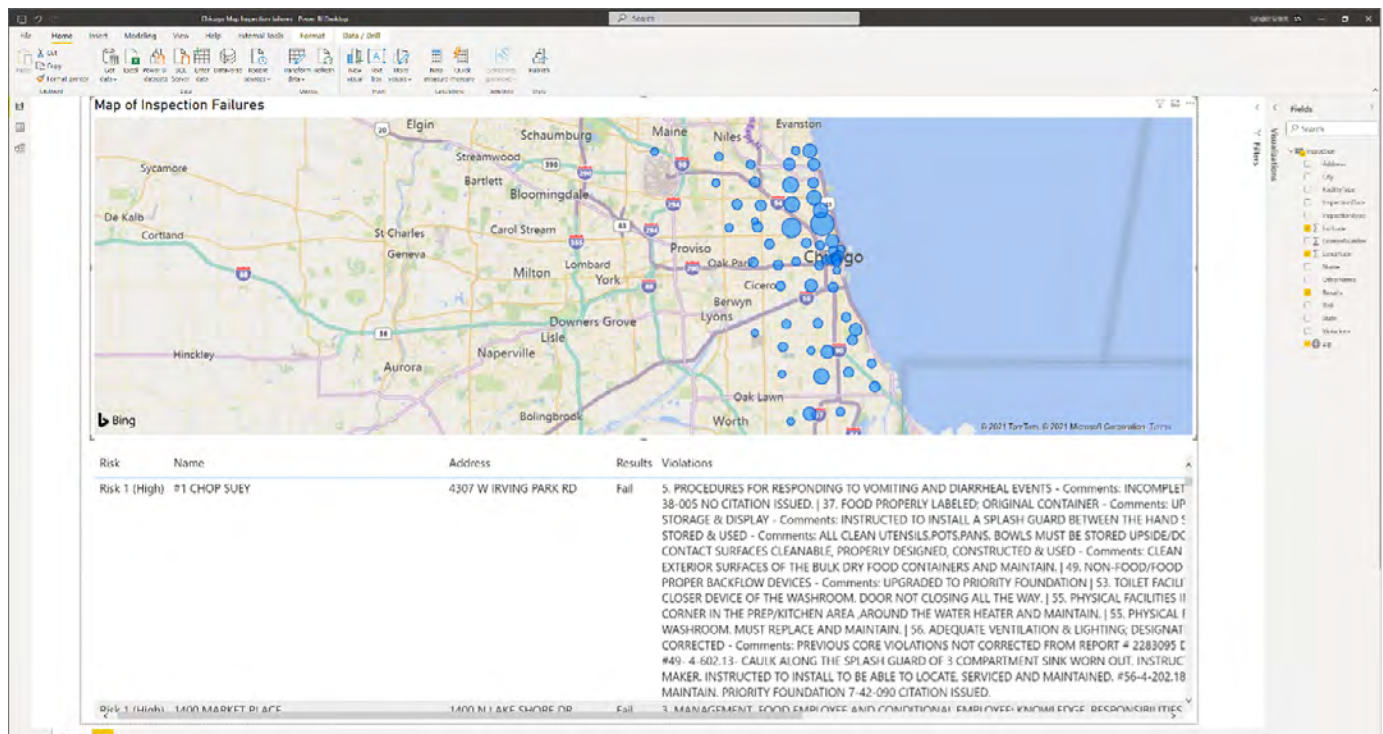


Figura 16: visualização do Power BI realizada com dados de Parquet

## #3: Explore o data lake usando um ponto de extremidade sem servidor totalmente gerenciado

Muitas empresas optam por armazenar grande parte de seus dados em um data lake. Uma conta do ADLS Gen2 pode armazenar anos de dados a baixo custo, especialmente quando os dados são armazenados no formato Parquet, que oferece compactação de dados, reduzindo a necessidade de armazenamento dos dados. Quando você cria um espaço de trabalho do Azure Synapse, uma conta do Azure Data Lake Gen2 (ADLS) é provisionada, fornecendo o armazenamento necessário para criar um data lake. O Azure Synapse Analytics pode gerenciar os conjuntos de dados e fornecer um banco de dados lógico com base na conta do ADLS, em que os dados podem ser consultados como se estivessem em um banco de dados usando o ponto de extremidade sem servidor incluído. Não é necessário provisionar itens, nem interromper e iniciar tarefas, pois ele usa um modelo de pagamento por consulta.

Bancos de dados lógicos podem ser criados usando o armazenamento do ADLS para que os usuários possam explorar os dados contidos no data lake usando o T-SQL. O banco de dados virtual também inclui a capacidade de criar um row-level security no banco de dados porque esse recurso é compatível. Naturalmente, você pode criar seu modelo de dados seguindo as práticas recomendadas e criando exibições classificadas para acessar os dados inseridos no modelo de dados do Power BI. Ao usar exibições, você pode garantir que as colunas estejam em formato legível por humanos e classificadas pelas colunas com o menor número de valores para melhorar a compactação do mecanismo de VertiPaq e a performance do relatório do Power BI.

Usando o ponto de extremidade sem servidor no Azure Synapse, você só é cobrado pelas consultas executadas. Os dados podem ser acessados criando bancos de dados lógicos no Azure Synapse com instruções T-SQL. Você pode definir tabelas além de dados de vários diretórios e data lakes diferentes e combiná-los em um banco de dados. Essa tarefa é executada por meio do comando T-SQL **CETAS, Create External Table AS SELECT**. Esse comando é usado para criar tabelas externas que podem ser acessadas pelo Power BI ou por **SSMS**.

Aqui está uma amostra do código que usei para criar um banco de dados lógico no Azure Synapse e para usar o agrupamento UTF8, necessário para tratar alguns dos textos em nossos arquivos Parquet:

```
CREATE DATABASE EMR

GO

--Change the internal database to EMR to issue comments

USE    EMR

GO

ALTER DATABASE EMR COLLATE Latin1_General_100_BIN2_UTF8;
```

A próxima etapa é definir a localização dos dados a serem lidos e definir o local para recuperar os dados. Adicionamos instruções **Not Exists** para permitir executar o script várias vezes. Se os dados forem armazenados em vários locais diferentes, você poderá criar várias fontes de dados externas para fazer referência a elas:

```
IF NOT EXISTS (SELECT * FROM sys.external_file_
formats WHERE name = 'SynapseParquetFormat')

    CREATE EXTERNAL FILE FORMAT [SynapseParquetFormat]

    WITH ( FORMAT_TYPE = PARQUET )

GO

IF NOT EXISTS (SELECT * FROM sys.external_data_sources WHERE name = 'EMRParquetData')

    CREATE EXTERNAL DATA SOURCE [EMRParquetdata]

    WITH (

        LOCATION 'https://synapsedatalake.dfs.core.windows.net/root/Curated/EMR',
        CREDENTIAL = [ManagedIdentityCredential]

    )

GO
```



Depois que o formato e o local tiverem sido especificados, você poderá criar as tabelas além dos dados. Aqui, estamos criando duas tabelas e formatando a consulta:

```
IF EXISTS (SELECT * FROM sys.external_tables WHERE name = 'MailingAddress')

DROP EXTERNAL TABLE MailingAddress

GO

CREATE EXTERNAL TABLE MailingAddress (

    [MailingAddressID] varchar(100) ,

    [StreetAddress] varchar(300),

    [City] varchar(100),

    [StateAbbr] char(2),

    [Zipcode] varchar(10),

    [ModifiedDate] varchar(13)

)

WITH (

    LOCATION = '/2021MailingAddress/Address.snappy.parquet',

    DATA_SOURCE = [RawParquetData],

    FILE_FORMAT = [SynapseParquetFormat]

)

GO
```

```
IF EXISTS (SELECT * FROM sys.external_tables WHERE name = 'Patient')

DROP EXTERNAL TABLE Patient

GO

CREATE EXTERNAL TABLE Patient (

PatientID varchar(100) ,

SSN varchar(12) ,

FirstName varchar(60) ,

MiddleName varchar(60) ,

LastName varchar(60) ,

Gender char(1) ,

MailingAddressID varchar(100) ,

Email varchar(256) ,

DateOfBirth varchar(13) ,

ModifiedDate varchar(13)

)

WITH (

LOCATION = '/2021Patient/Patient.snappy.parquet',

DATA_SOURCE = [RawParquetData],

FILE_FORMAT = [SynapseParquetFormat]

)

GO
```



Aqui estão as exibições que criamos com esses dados, em que vamos nos conectar com o Power BI:

```
CREATE VIEW vPatient

AS

SELECT TOP 999999999 [PatientID]

    ,FirstName AS [First Name]

    ,LastName AS [Last Name]

    ,[Gender]

    ,CAST([AddressID] AS INT) AS AddressID

    ,[Email]

    ,DATEDIFF(YYYY, CAST([DateOfBirth] AS DATE), GETDATE() ) as Age

FROM [EMR].[dbo].[Patient]

ORDER BY Gender, Age

GO

CREATE VIEW vAddress

AS

SELECT TOP 999999999 cast([MailingAddressID] as int) as AddressID

    ,[StreetAddress] as [Address]

    ,[City]

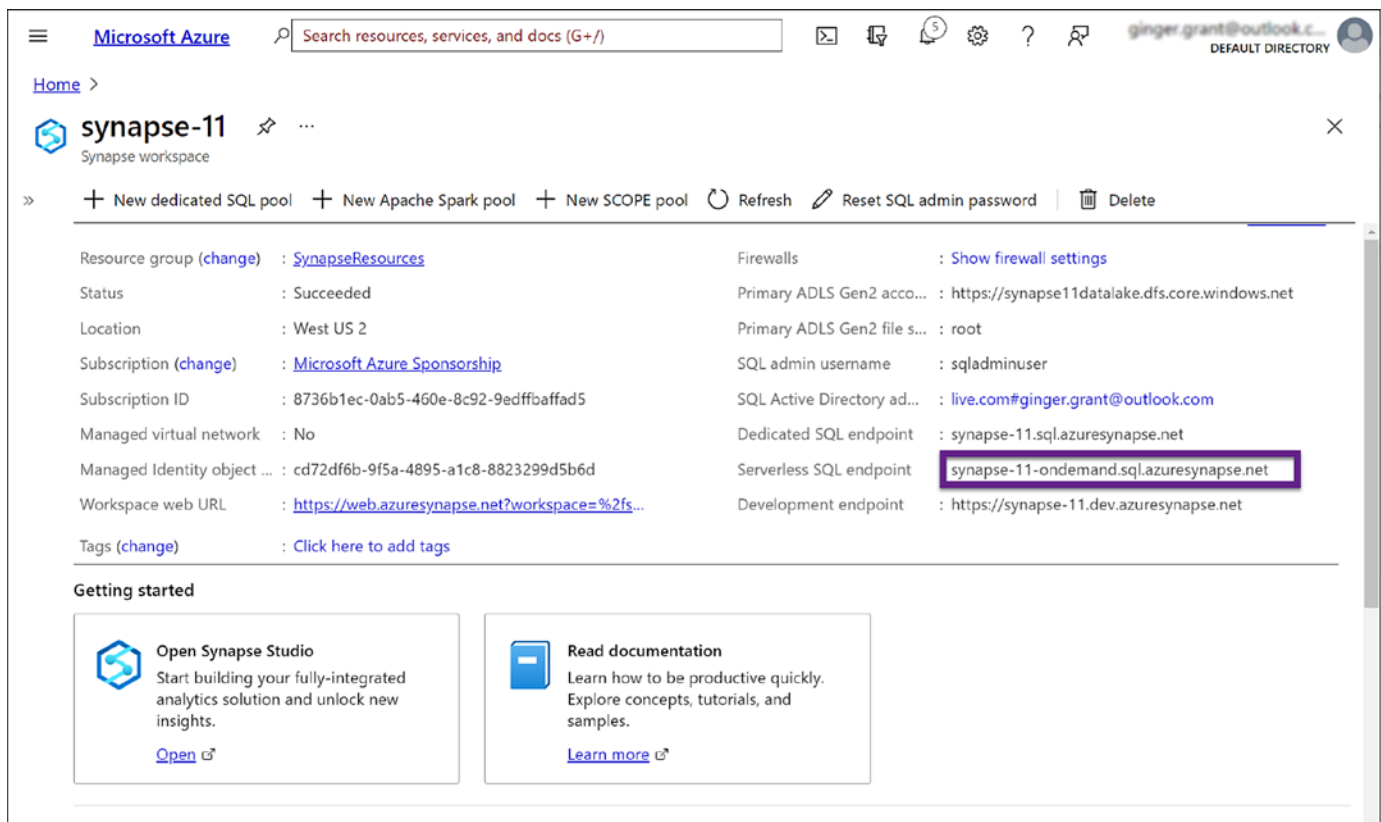
    ,[StateAbbr] as [State or Province]

    ,[Zipcode] as [Postal Code]

FROM [EMR].[dbo].[MailingAddress]

ORDER BY [StateAbbr] , [City]
```

Quando isso for concluído, poderemos nos conectar ao banco de dados EMR no Power BI e outras ferramentas de banco de dados, como o SQL Server Management Studio. Quando nos conectarmos a esses dados, será exatamente como estamos conectados ao SQL Server. O nome do banco de dados usado para o pool sem servidor se encontra na página do Azure Synapse no portal do Azure, como mostrado na *Figura 17*, onde há uma caixa roxa em torno do nome do ponto de extremidade sem servidor que usaremos para conexão ao Power BI:



The screenshot displays the Azure Synapse workspace 'synapse-11' in the Azure portal. The page shows various configuration details for the workspace, including resource group, status, location, subscription, and endpoints. The 'Serverless SQL endpoint' is highlighted with a red box, showing the URL 'synapse-11-ondemand.sql.azuresynapse.net'.

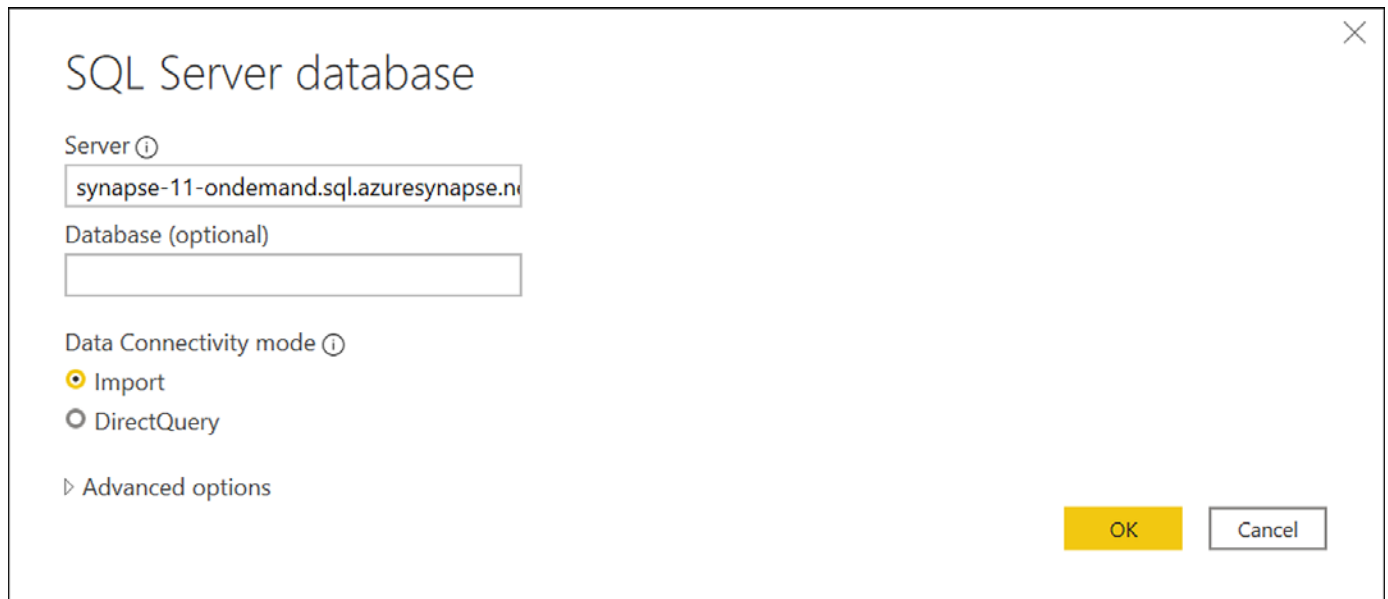
Property	Value
Resource group (change)	<a href="#">SynapseResources</a>
Status	Succeeded
Location	West US 2
Subscription (change)	<a href="#">Microsoft Azure Sponsorship</a>
Subscription ID	8736b1ec-0ab5-460e-8c92-9edffbaaffad5
Managed virtual network	No
Managed Identity object ...	cd72df6b-9f5a-4895-a1c8-8823299d5b6d
Workspace web URL	<a href="https://web.azuresynapse.net?workspace=%2fs...">https://web.azuresynapse.net?workspace=%2fs...</a>
Tags (change)	<a href="#">Click here to add tags</a>
Firewalls	<a href="#">Show firewall settings</a>
Primary ADLS Gen2 acco...	<a href="https://synapse11datalake.dfs.core.windows.net">https://synapse11datalake.dfs.core.windows.net</a>
Primary ADLS Gen2 file s...	root
SQL admin username	sqladminuser
SQL Active Directory ad...	<a href="#">live.com#ginger.grant@outlook.com</a>
Dedicated SQL endpoint	synapse-11.sql.azuresynapse.net
Serverless SQL endpoint	<b>synapse-11-ondemand.sql.azuresynapse.net</b>
Development endpoint	<a href="https://synapse-11.dev.azuresynapse.net">https://synapse-11.dev.azuresynapse.net</a>

**Getting started**

- Open Synapse Studio**  
Start building your fully-integrated analytics solution and unlock new insights.  
[Open](#)
- Read documentation**  
Learn how to be productive quickly. Explore concepts, tutorials, and samples.  
[Learn more](#)

Figura 17: Azure Synapse do portal do Azure que mostra o ponto de extremidade sem servidor

A próxima etapa é conectar o Power BI a essa fonte de dados a partir de uma conexão do SQL Server, conforme mostrado na *Figura 18*:



SQL Server database

Server ⓘ  
synapse-11-ondemand.sql.azuresynapse.n

Database (optional)

Data Connectivity mode ⓘ  
☒ Import  
☐ DirectQuery

▸ Advanced options

OK Cancel

Figura 18: acessando um pool do Azure Synapse sem servidor usando uma conexão de banco de dados do SQL Server

Como você pode ver, optamos por importar dados e estamos nos conectando ao ponto de extremidade sem servidor do Azure Synapse. Vamos usar as exibições que criamos no script anterior, mas note que estamos autorizados a selecionar dados de qualquer banco de dados. Ao autenticar o banco de dados, você desejará usar sua conexão do Azure AD.

The screenshot shows the Power BI Navigator interface. On the left, the 'Display Options' pane shows a tree view of the data source 'synapse-11-ondemand.sql.azuresynapse.net'. Under the 'EMR [4]' folder, the 'vPatient' table is selected. Below the tree view, there are checkboxes for 'vAddress', 'MailingAddress', and 'Patient', all of which are currently unchecked. At the bottom left of the Navigator, there is a button labeled 'Select Related Tables'.

The main pane displays the 'vPatient' table with the following columns: PatientID, First Name, Last Name, Gender, AddressID, and Email. The table contains 25 rows of data. At the bottom right of the main pane, there are three buttons: 'Load' (highlighted in yellow), 'Transform Data', and 'Cancel'.

PatientID	First Name	Last Name	Gender	AddressID	Email
99	Kelly	Phillips	F	98	valandsh
82	Teresa	Nugent	F	82	rracher@
81	Melissa	Drye	F	81	mhem@:
74	Dawn	Begay	F	75	philrieste
96	Joan	Barba	F	95	tnugen@
35	Stacy	Prickett	F	40	Irene.mir
14	Luz	Mares	F	21	debheal@
44	Kathryn	Ottosen	F	49	lbreaux@
31	Barbara	Olson	F	37	iam300i
7	Merrie	Heath	F	15	celticlady
3	Connie	Gover	F	11	benitab5
5	Susan	Schulte	F	13	bromgn@
95	Susan	Maestas	F	94	tindle358
20	Sheri	Williams	F	27	dprince0
34	Cynthia	Knutson	F	4	acdpc50(
11	Deborah	Jackson	F	19	Comp@h
53	Julee	Gilson	F	56	mboro2@
41	Bonnie	Russell	F	46	kle@ne
77	Debra	Sawyer	F	78	rayban92
40	Sue	Budman	F	45	KNUTSOI
80	Irene	Russell	F	80	rmerrick
100	Rosalind	Racher Kirk	F	99	WDHOU5
25	Lynne	Myers	F	31	GregFlyH

Figure 19: selecionar dados de um banco de dados lógico no Azure Synapse no Power BI

Usando dados do Azure Synapse Analytics, agora podemos criar relatórios do Power BI por meio de bancos de dados sem servidor, como mostrado na *Figura 19*.

O uso do banco de dados lógico permite acessar dados armazenados em diferentes formatos, como Parquet, e explorar os dados armazenados em um data lake para fornecer resultados significativos sem precisar criar serviços adicionais, pois toda essa funcionalidade é proporcionada no Azure Synapse.

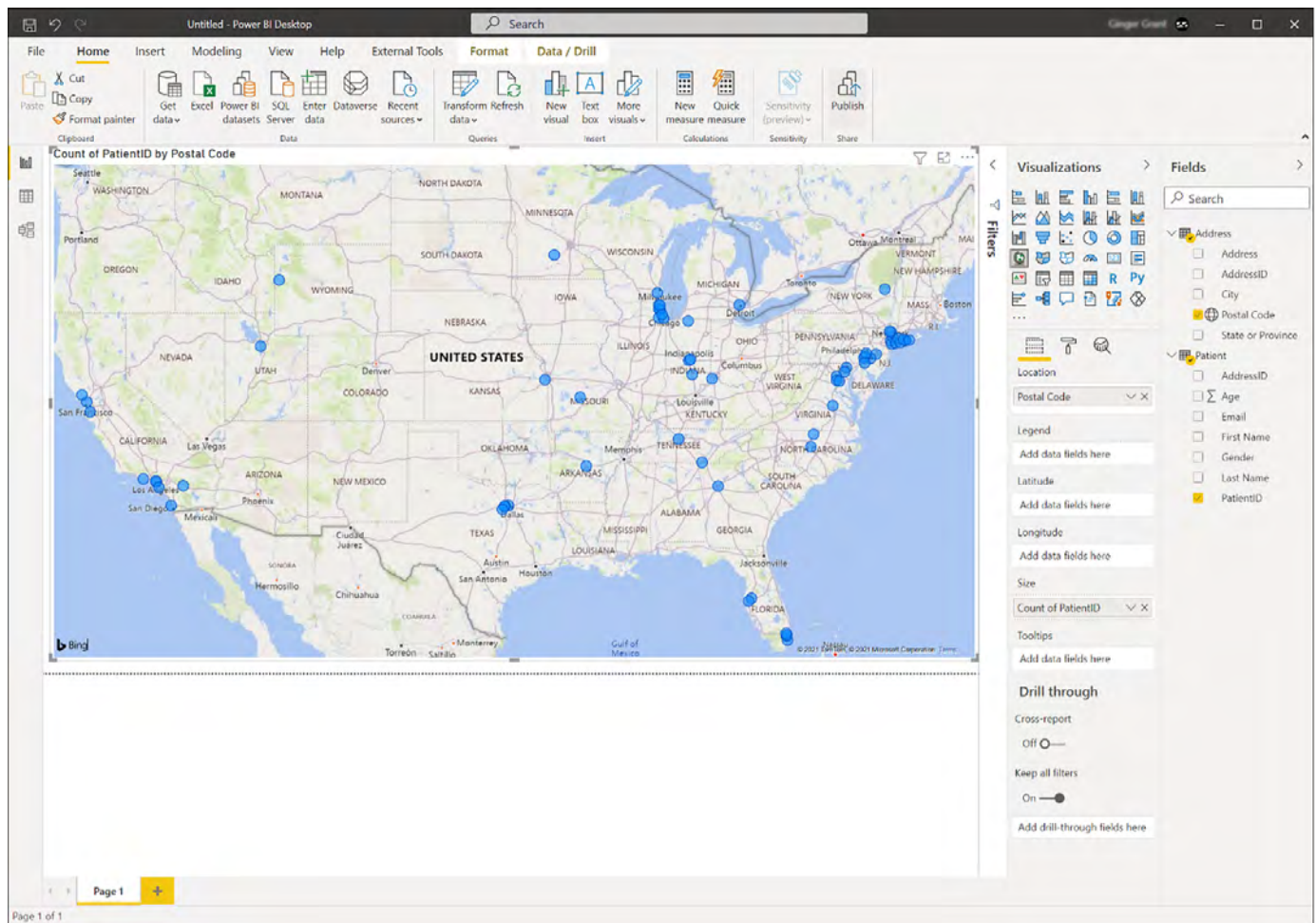


Figura 20: relatório do Power BI com dados de pool sem servidor do Azure Synapse

## #4: Crie pipelines de dados sem código para integrar mais fontes de dados e enriquecer insights

O Azure Synapse pode coletar dados de várias fontes de dados diferentes e armazená-los centralmente em uma conta de armazenamento do Azure Data Lake Gen2, que atua como um único repositório para todas as outras aplicações que precisam de acesso a esses dados. Os cientistas de dados podem querer combinar dados de aplicações Web, como jornada de trabalho, juntamente com dados da empresa usando fluxos de dados para coletar dados de diferentes fontes. Esses dados também podem ser transformados em um data warehouse antes de serem acessados. Como parte do processo de coleta de dados, os dados também podem ser analisados usando modelos de aprendizado de máquina, com o Power BI usado para visualizar os resultados. Por exemplo, como parte de um processo noturno, os clientes podem ser analisados para determinar a probabilidade de rotatividade do cliente usando o aprendizado de máquina. Os dados analisados podem ser de comentários não estruturados com o pessoal do serviço de atendimento ao consumidor associado ao histórico de transações armazenado no data warehouse, que é armazenado em um pool dedicado no Azure Synapse. Essa análise pode ser feita como parte de um processo em lote para alimentar dados no Power BI e, depois, a análise de rotatividade do cliente pode ser incluída no Power BI usando pipelines de integração no Azure Synapse.

As integrações do Azure Synapse permitem criar pipelines de dados da mesma forma que o Azure Data Factory, além de fornecer a capacidade de incluir blocos de anotações PySpark para gravar dados em pools sem servidor que podem ser incluídos em um modelo do Power BI. Vejamos como fazer isso com o Azure Synapse.

Primeiro, vamos criar um novo pipeline para processar os dados brutos do serviço de atendimento ao consumidor. A *Figura 21* mostra um pipeline de exemplo que criamos para analisar e pontuar os dados extraídos do serviço de atendimento ao consumidor, que vamos produzir como parte de um pipeline do Azure Synapse:

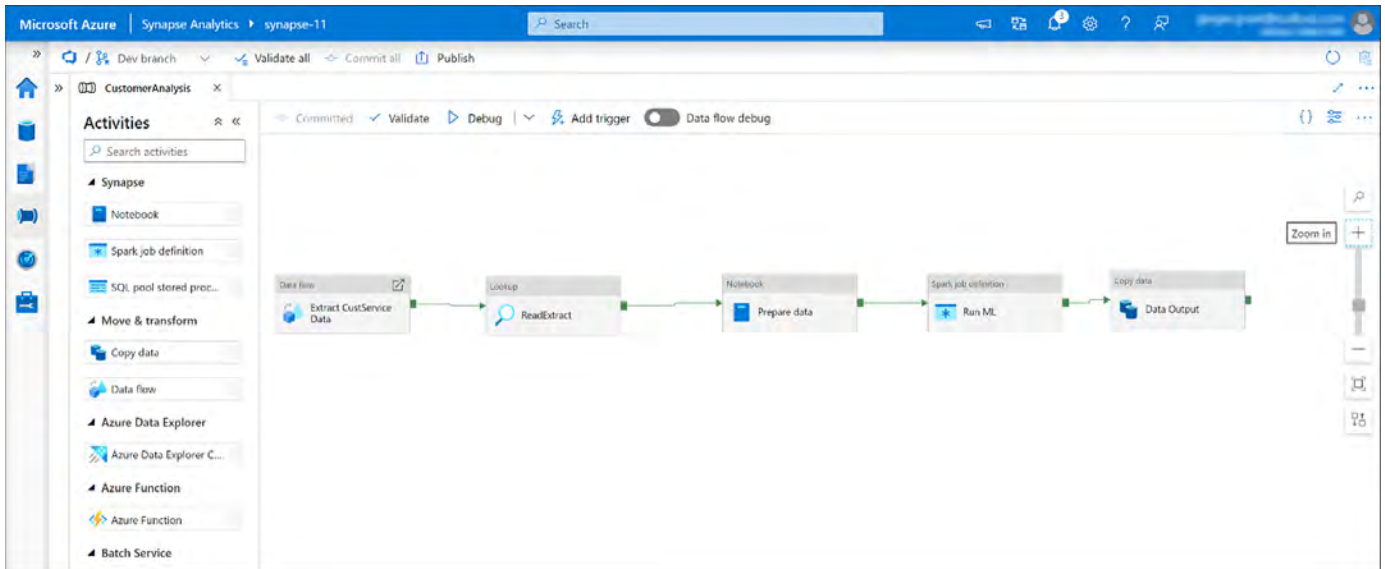


Figura 21: Azure Pipelines para extrair dados do serviço de atendimento ao consumidor e analisá-los com o Spark

O resultado da análise de aprendizado de máquina é criar uma transação a partir dos dados brutos do serviço de atendimento ao consumidor, processar os dados usando o PySpark e, depois, usar os dados processados para alimentar um objeto de aprendizado de máquina para pontuar a transação do cliente e gravar os dados em um pool sem servidor no Azure Synapse. Vejamos nossos novos elementos de dados no Power BI, pois os dados foram gravados no LogicalDW.

Vamos nos conectar usando uma conexão do SQL Server com o pool sem servidor da mesma forma que fizemos no exemplo anterior, como mostrado na *Figura 21*. Poderemos usar as informações coletadas da análise do sentimento do serviço de atendimento ao consumidor e outros dados do cliente para mostrar a probabilidade média de rotatividade, junto com uma pontuação de satisfação do cliente com base em uma análise dos dados no Azure Synapse e fornecer essas informações como parte de um relatório do Power BI, conforme mostrado na *Figura 22*:

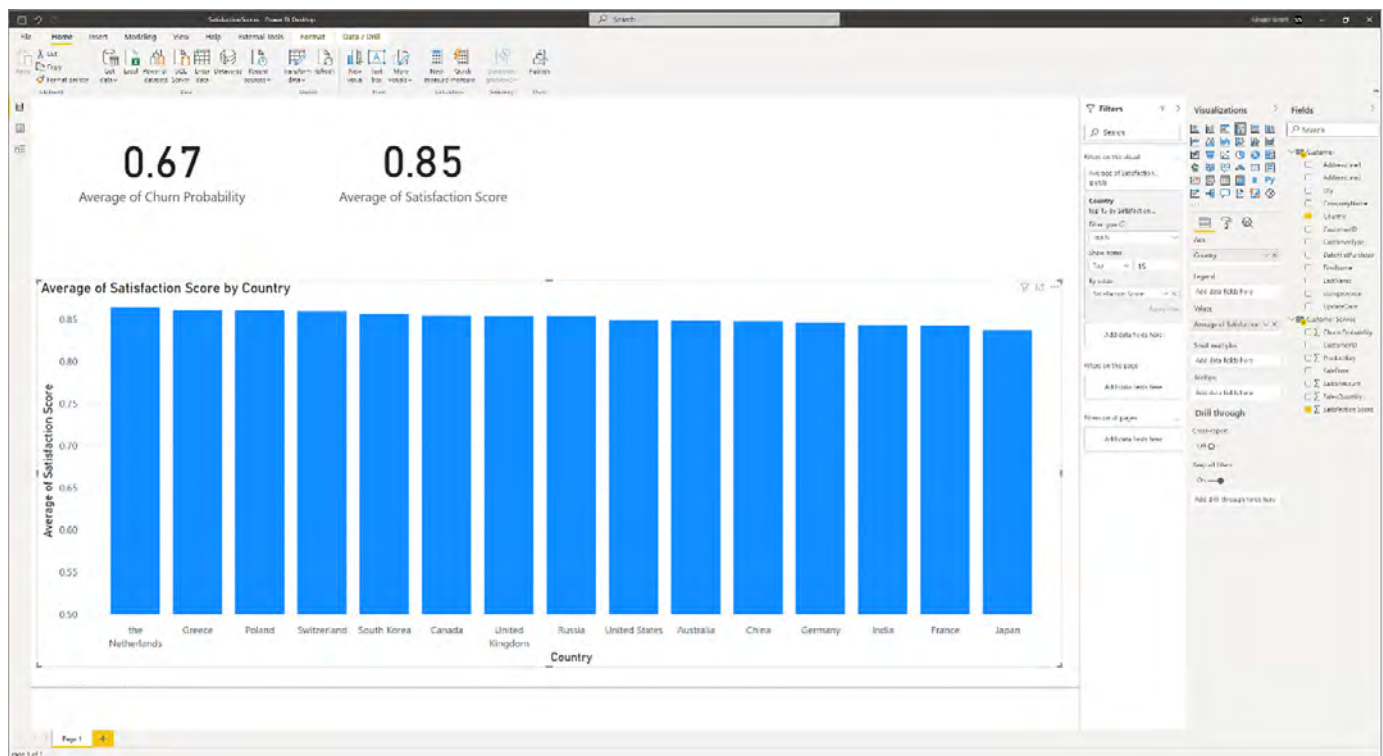
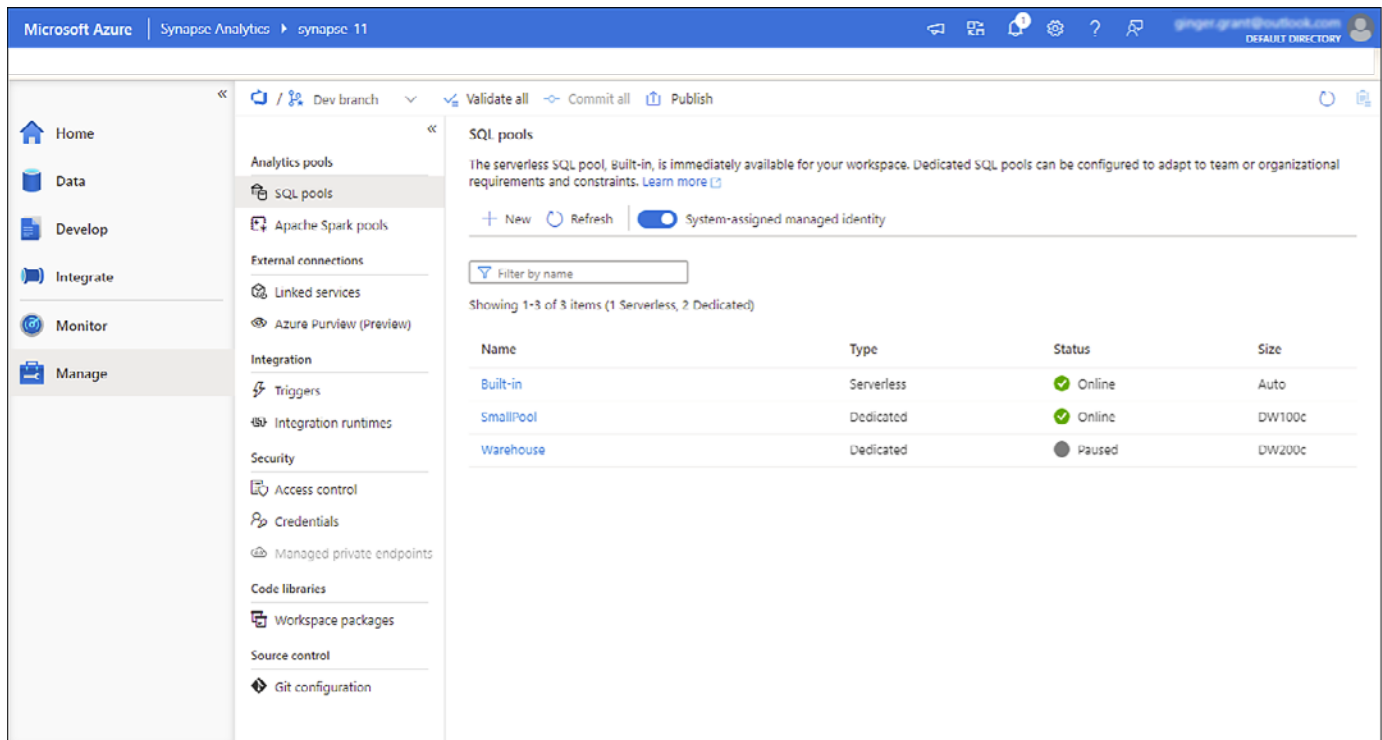


Figura 22: o resultado do processamento de dados no Azure Synapse para fazer análises de rotatividade no conjunto de dados do serviço de atendimento ao consumidor e a pontuação de satisfação do cliente



## #5: Crie um data warehouse seguro e conecte-se aos painéis do Power BI

O Power BI pode se conectar a fontes de dados, incluindo conjuntos de dados muito grandes — como os encontrados em um pool dedicado no Azure Synapse. Pode haver desafios para relatar um grande volume de dados, com que os praticantes do Power BI já estão familiarizados. No Power BI, existem diferentes métodos para ajustar a performance de dados, incluindo tabelas de agregação e dobramento de consulta para citar apenas dois. Os grandes conjuntos de dados geralmente usam consultas diretas ou conexões duplas com os dados, pois os modelos de dados costumam ser grandes para importação. O Azure Synapse também contém métodos para ajudar a melhorar a performance do relatório por meio da classificação de gerenciamento de workload, que permite priorizar recursos para tarefas de relatório do Power BI. Os pools dedicados são muitas vezes pausados para poupar recursos do Azure. Para certificar-se de que o pool dedicado esteja em execução, acesse a seção **Gerenciar** no Azure Synapse e analise os pools SQL. Na *Figura 23*, note que temos dois pools dedicados, **SmallPool** e **Warehouse**. O status mostra que **SmallPool** está em execução:



The screenshot displays the 'SQL pools' management interface in the Microsoft Azure Synapse Analytics portal. The left sidebar shows the navigation menu with 'Manage' selected. The main content area shows a list of SQL pools. The 'SmallPool' is highlighted as being online.

Name	Type	Status	Size
Built-in	Serverless	Online	Auto
SmallPool	Dedicated	Online	DW100c
Warehouse	Dedicated	Paused	DW200c

Figura 23: gerenciar pools SQL

Se ele não estivesse em execução, eu clicaria no botão **Retomar**.

A primeira etapa para fornecer mais recursos para o relatório do Power BI é criar um grupo de workloads no Azure Synapse. O grupo de workloads especificará a quantidade de isolamento de workload associada a um grupo. Precisaremos criar um novo script SQL para este código e garantir que estamos nos conectando ao **SmallPool** ao executá-lo. A *Figura 24* mostra a seção **Desenvolver**, onde um novo script SQL foi criado, que está associado ao pool dedicado em execução, **SmallPool**:

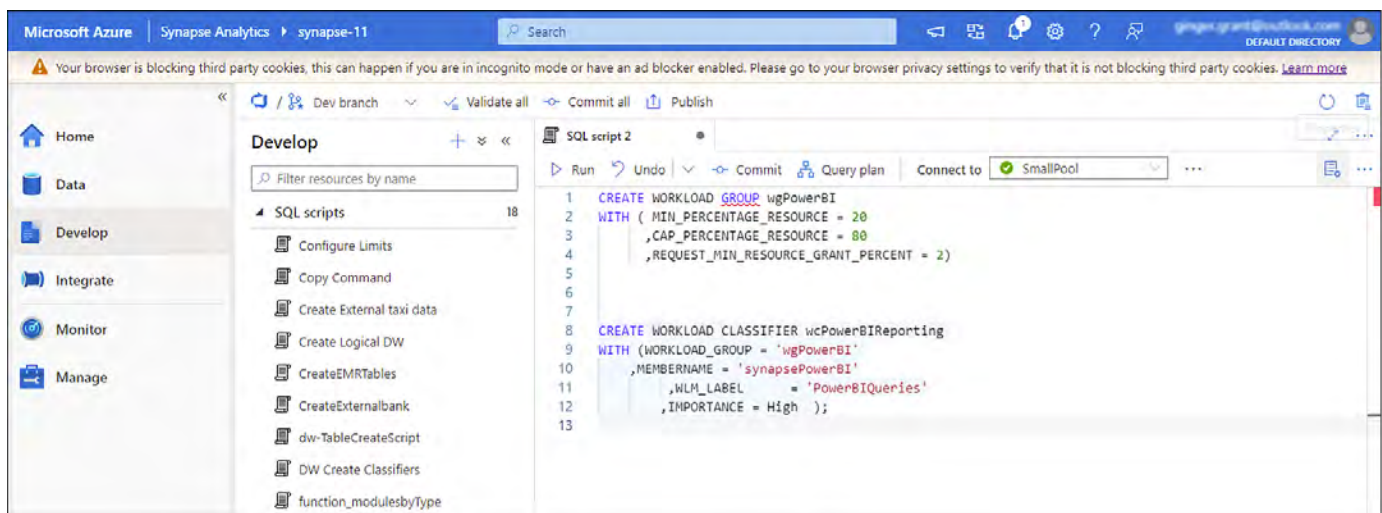


Figura 24: desenvolver um script SQL para criar workloads

Para executar a consulta SQL em um bloco de anotações do SQL no Azure Synapse, conforme mostrado na seguinte instrução, certifique-se de que o pool dedicado seja iniciado, pois você não pode executar consultas nele até que ele seja executado:

```
CREATE WORKLOAD GROUP wgPowerBI

WITH ( MIN_PERCENTAGE_RESOURCE = 20

      ,CAP_PERCENTAGE_RESOURCE = 80

      ,REQUEST_MIN_RESOURCE_GRANT_PERCENT = 2)
```

O grupo de workloads **wgPowerBI** fornecerá no máximo 80% dos recursos do Azure Synapse Analytics para elementos nesse workload. Cada membro será concedido um mínimo de 2% dos recursos na conexão inicial. Garanta que, ao criar grupos de recursos, o total de todos os recursos de porcentagem mínima adicione até 100%, pois a alocação de recursos em excesso pode criar problemas de performance. **REQUEST\_MIN\_RESOURCE\_GRANT\_PERCENT** é o número de blocos de recursos que podem ser alocados de uma só vez. O isolamento do workload permite dedicar recursos a grupos de workloads. Aqui vemos que o grupo de workloads ELT recebe pelo menos 20% dos recursos disponíveis usando **MIN\_PERCENTAGE\_RESOURCE**. Na ausência de isolamento de workload, as solicitações operam no pool compartilhado de recursos. O acesso a recursos no pool compartilhado não é garantido e é atribuído com base na importância.

A classificação de workload permite classificar consultas de Power BI para receber mais prioridade geral e recursos em solicitações menos importantes. Há cinco pesos diferentes que você pode atribuir a um classificador de workload: *usuário*, *função*, *rótulo*, *contexto* e *tempo*. Esses elementos diferentes são listados aqui na ordem em que são ponderados. O elemento com a maior prioridade é o *usuário*. Se você incluir *usuário* em um classificador, qualquer item executado por esse usuário terá maior prioridade. O próximo maior valor é a *função*. Os usuários de uma função específica têm menos prioridade do que um usuário, mas são mais fáceis de gerenciar. Como os usuários podem ter várias linhas, a maior prioridade é usada em classificadores de workload:

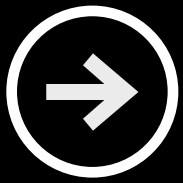
```
CREATE WORKLOAD CLASSIFIER wcPowerBIReporting  
  
WITH (WORKLOAD_GROUP = 'wgPowerBI'  
  
      ,MEMBERNAME = 'synapsePowerBI'  
  
      ,WLM_LABEL = 'PowerBIQueries'  
  
      ,IMPORTANCE = High );
```

Esse classificador usará os recursos dedicados definidos no grupo de workloads com membros incluídos no grupo de diretórios do Azure chamado **synapsePowerBI** para melhorar a performance dos relatórios que se conectam a dados em um pool dedicado no Azure Synapse.



## Resumo

Neste documento, analisamos cinco maneiras diferentes de aproveitar os diferentes recursos do Azure Synapse e usá-los com o Power BI. O Azure Synapse permite criar data marts além de data lakes para uso no Power BI. Você pode usar os recursos de exploração de dados do Power BI para analisar dados no Azure Synapse para melhorar os recursos analíticos. No Azure Synapse, você pode criar bancos de dados virtuais para uso no Power BI. Você também pode adicionar a capacidade de incluir funções de aprendizado de máquina criadas no Azure Synapse e relatar os resultados. Analisamos os métodos no Azure Synapse para ajuste de performance do Power BI, adicionando gerenciamento de workload para garantir que o Power BI tenha todos os recursos necessários para coletar dados. Essas são algumas formas de aproveitar os recursos do Azure Synapse e, depois de começar a usá-lo com o Power BI, você poderá aproveitar todos esses elementos. Incorporar os elementos discutidos neste documento fornecerá as habilidades necessárias para aproveitar o Power BI com o Azure Synapse.



## Próximas etapas

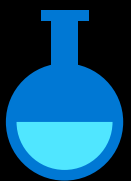
3 etapas para levar o BI e a análise para o próximo patamar:



Crie um [espaço de trabalho gratuito do Azure Synapse](#)



Assista à [série de treinamentos práticos gratuita](#)



Use as [quantidades gratuitas no Azure Synapse](#)  
para continuar experimentando