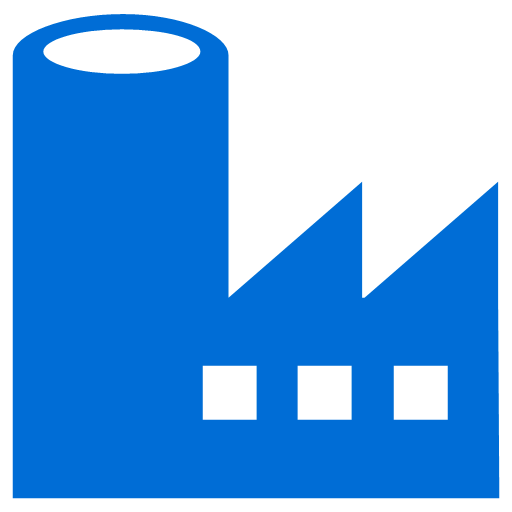
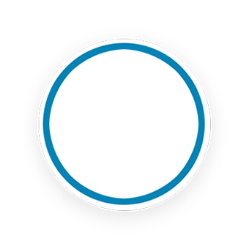


**Tutorial**

**Criação de um Data Factory usando a interface do usuário**



**Azure Data Factory**

[**1.** **Links Úteis** 3](#_Toc5401006)

[**2.** **Descrição** 3](#_Toc5401007)

[**2.1** **Pré-requisitos** 3](#_Toc5401008)

[**3.** **Criando os arquivos e pastas de entrada** 3](#_Toc5401009)

[**4.** **Criar um data factory** 5](#_Toc5401010)

[**5.** **Criar um serviço vinculado** 7](#_Toc5401011)

[**6.** **Criar conjuntos de dados** 9](#_Toc5401012)

[**7.** **Criar um pipeline** 11](#_Toc5401013)

[**8.** **Depurar o pipeline** 13](#_Toc5401014)

[**9.** **Disparar o pipeline manualmente** 14](#_Toc5401015)

[**10.** **Monitorar o Pipeline** 14](#_Toc5401016)

[**11.** **Disparar o pipeline em um cronograma** 15](#_Toc5401017)

[**12.** **Considerações finais** 17](#_Toc5401018)

# **Links Úteis**

* [Documentação](https://docs.microsoft.com/pt-br/azure/data-factory/)

# **Descrição**

É um serviço de integração de dados com baseado em nuvem que permite que você crie fluxos de trabalho orientados a dados na nuvem para orquestrar e automatizar a movimentação de dados e a transformação de dados. Usando o Azure Data Factory, é possível criar e agendar fluxos de trabalho orientados a dados (chamados de pipelines) que podem ingerir dados de diferentes repositórios de dados. Ele pode processar e transformar dados usando serviços de computação como o Azure HDInsight Hadoop, Spark, Azure Data Lake Analytics e Azure Machine Learning.

## **Pré-requisitos**

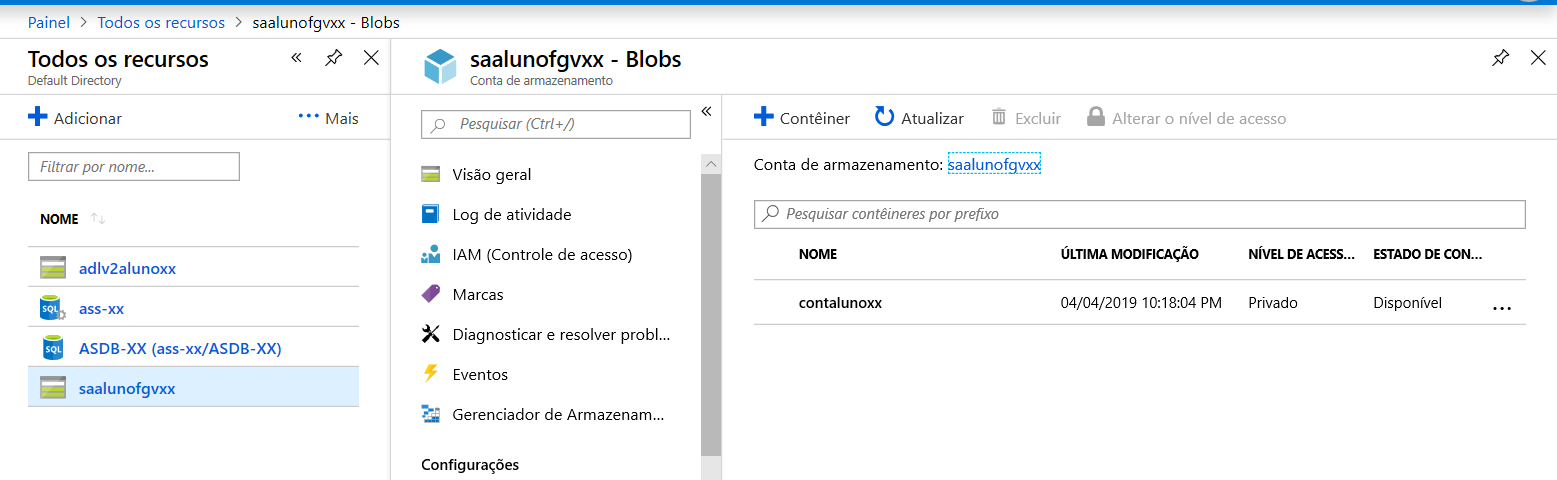
Para concluir este tutorial, verifique se está instalado:

* Possui uma conta no portal do Azure (<http://portal.azure.com>)

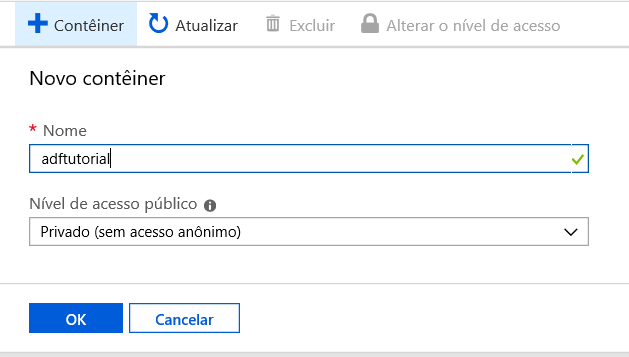
# **Criando os arquivos e pastas de entrada**

Nesta seção, você cria um contêiner de blobs chamado **adftutorial** no armazenamento de Blobs do Azure. Você cria uma pasta chamada **entrada** no contêiner e, em seguida, carrega um arquivo de exemplo na pasta de entrada.

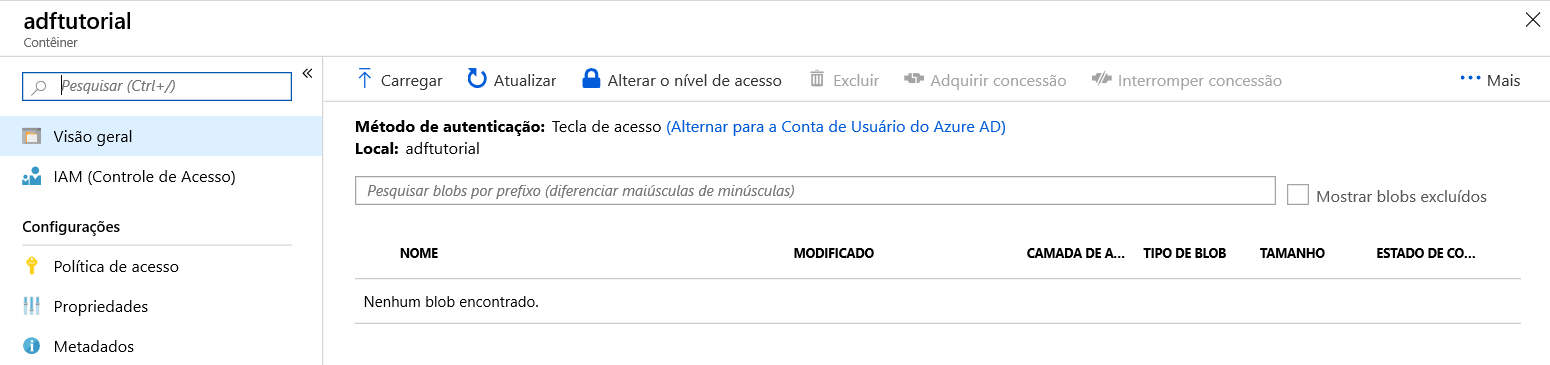
1. Na janela **Conta de armazenamento**, alterne para **Visão geral** e depois selecione **Blobs**.
2. Na página **Serviço Blob**, selecione **+ Contêiner** na barra de ferramentas.



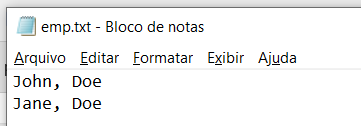
1. Na caixa de diálogo **Novo contêiner**, insira **adftutorial** como o nome e selecione **OK**.



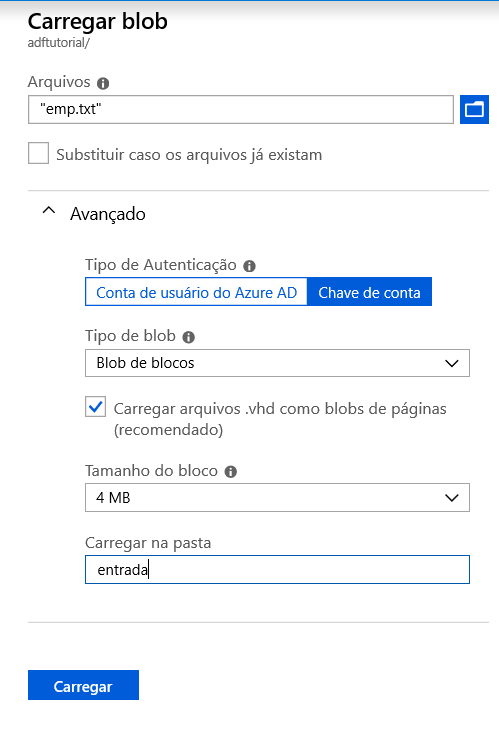
1. Selecione **adftutorial** na lista de contêineres.
2. Na página **Contêiner**, selecione **Carregar** na barra de ferramentas.



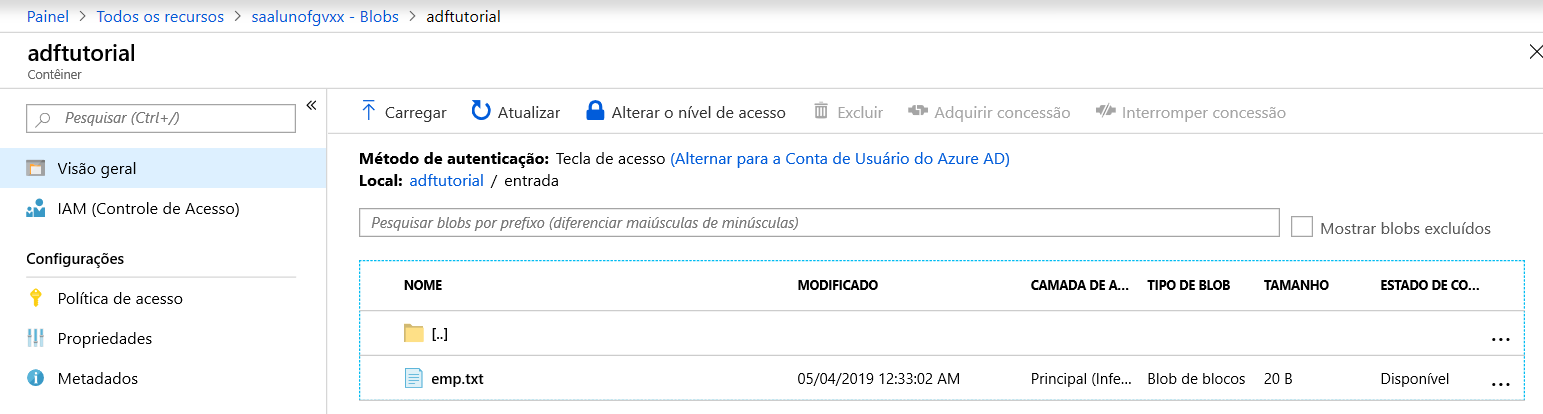
1. Na página **Carregar blob**, selecione **Avançado**.
2. Inicie o **Bloco de notas** e crie um arquivo chamado **emp.txt** com o seguinte conteúdo. Salve-o na pasta desejada.



1. No Portal do Azure, na página **Carregar blob**, procure e selecione o arquivo **emp.txt** para a caixa **Arquivos**.
2. Insira **entrada** como um valor da caixa **Carregar para a pasta**.
3. Confirme que a pasta é **entrada** e o arquivo é **emp.txt** e selecione **Carregar**.



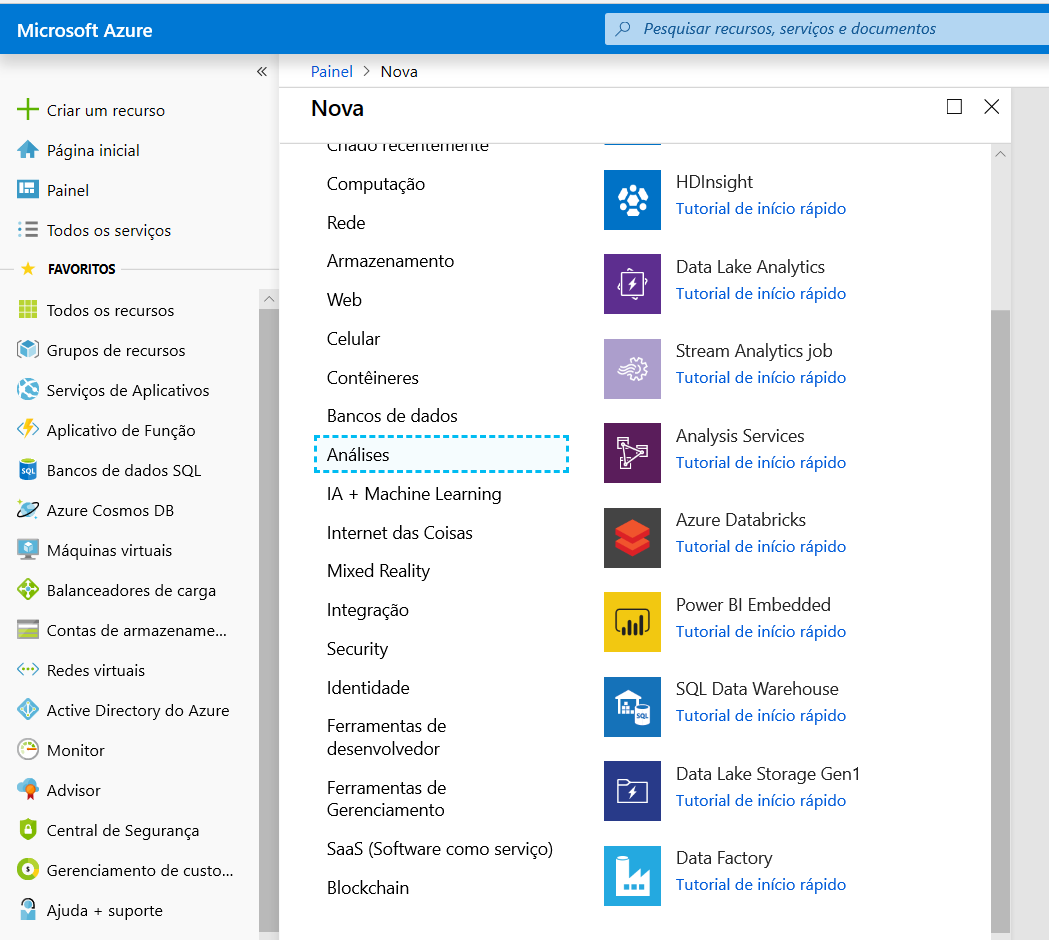
1. O arquivo **emp.txt** e o status do carregamento devem estar na lista.



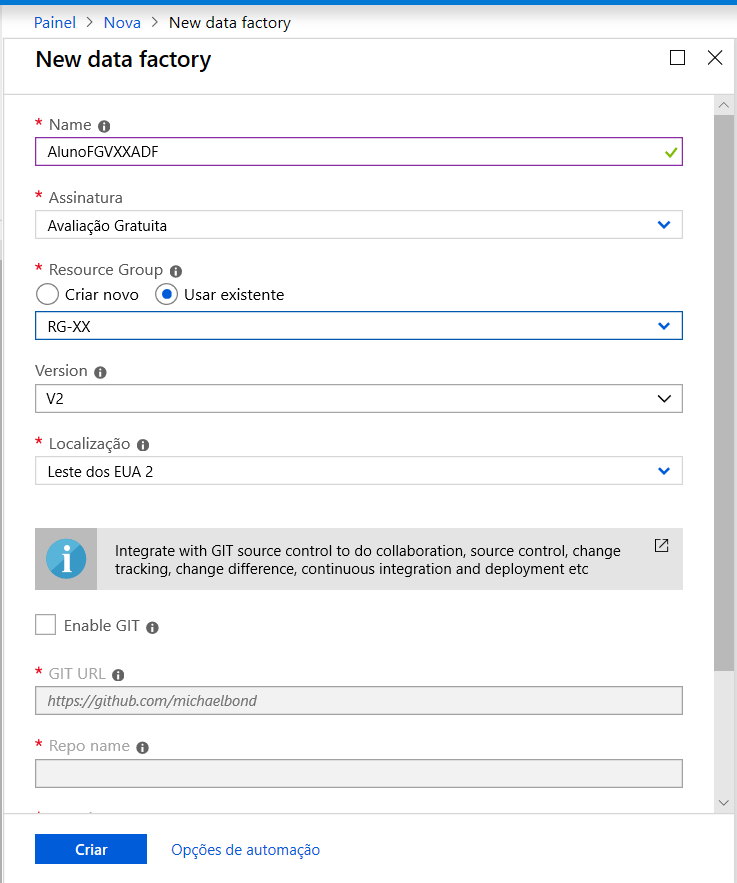
1. Feche a página **Carregar blob** clicando no **X** no canto superior.
2. Mantenha a página **Contêiner** aberta. Você a usa para verificar a saída no final do guia de início rápido.

# **Criar um data factory**

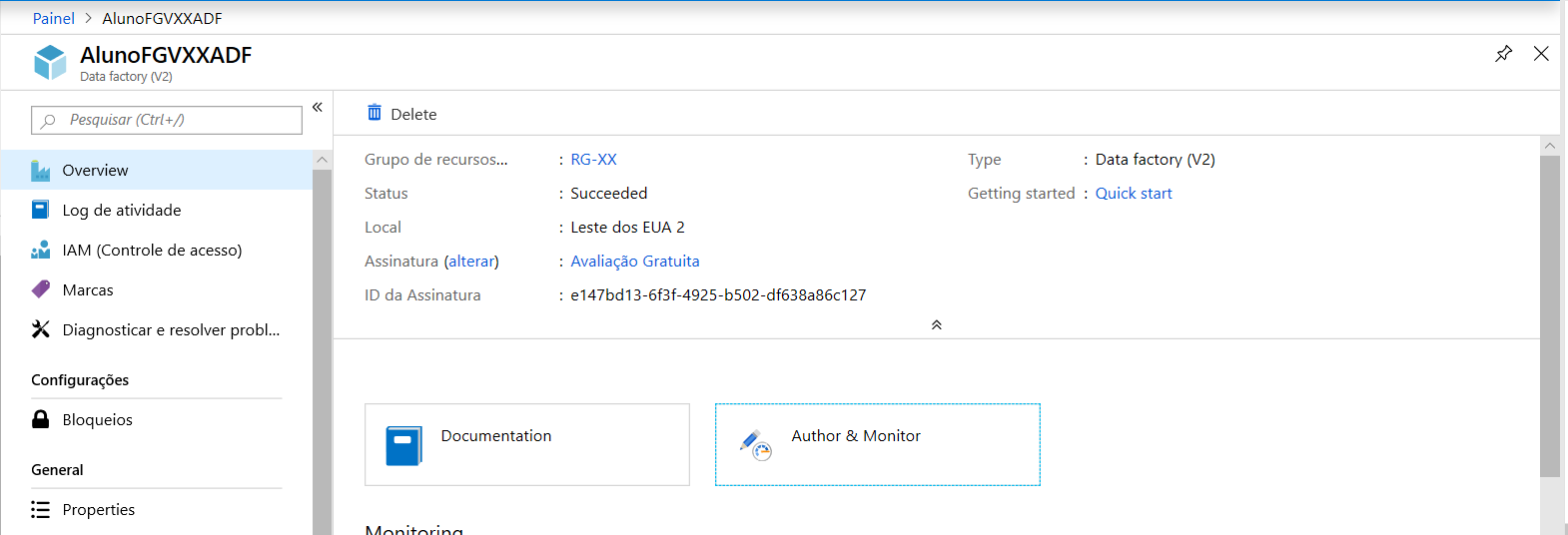
1. Iniciar o navegador da Web **Microsoft Edge** ou **Google Chrome**. Atualmente, a interface do usuário do Data Factory tem suporte apenas nos navegadores da Web Microsoft Edge e Google Chrome.
2. Vá para o [Portal do Azure](https://portal.azure.com/).
3. Selecione **Criar recurso** no menu à esquerda, depois **Análise** e, por fim, **Data Factory**.



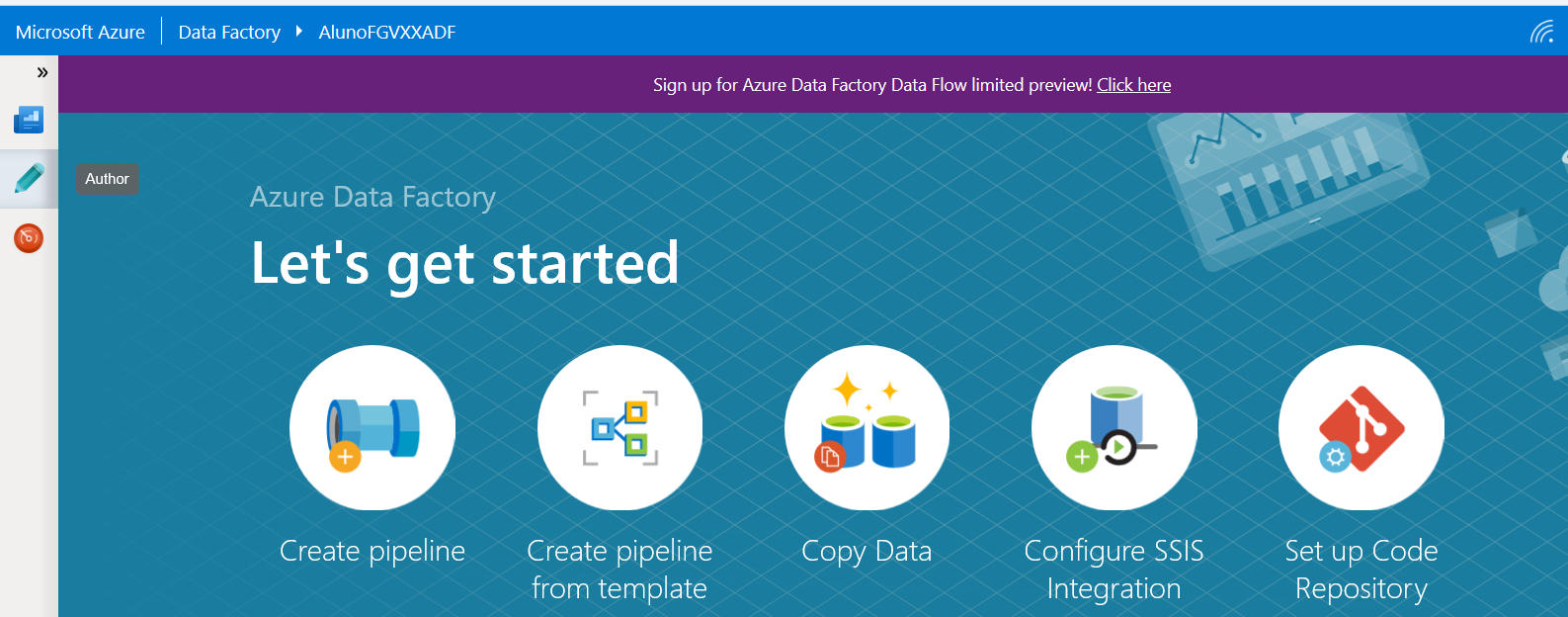
1. Na página **Novo data factory**, insira **ADF** no campo **Nome**.
2. O nome do Azure Data Factory deve ser globalmente exclusivo. Se você visualizar o seguinte erro, altere o nome de data factory (por exemplo, **AlunoFGVXXADF**) e tente criar novamente. Para ver as regras de nomenclatura para artefatos do Data Factory consulte o artigo [Data Factory - regras de nomenclatura](https://docs.microsoft.com/pt-br/azure/data-factory/naming-rules).
3. Para **Assinatura**, selecione a assinatura do Azure na qual você deseja criar o data factory.
4. Para o **Grupo de Recursos**, use uma das seguintes etapas:
   1. Selecione **Usar existente** e selecione um grupo de recursos existente na lista.
5. Para **Versão**, selecione **V2**.
6. Em **Local**, selecione uma localização para o data factory.
7. A lista mostra somente os locais aos quais o Data Factory dá suporte e em que os metadados do Azure Data Factory serão armazenados. Observe que os armazenamentos de dados (como o Armazenamento do Azure e o Banco de Dados SQL do Azure) e serviços de computação (como o Azure HDInsight) usados pelo Data Factory podem ser executados em outras regiões.
8. Selecione **Criar**.



1. Após a criação, a página do **Data Factory** será exibida. Clique no bloco **Criar e Monitorar** para iniciar o aplicativo de interface do usuário (IU) do Azure Data Factory em uma guia separada.



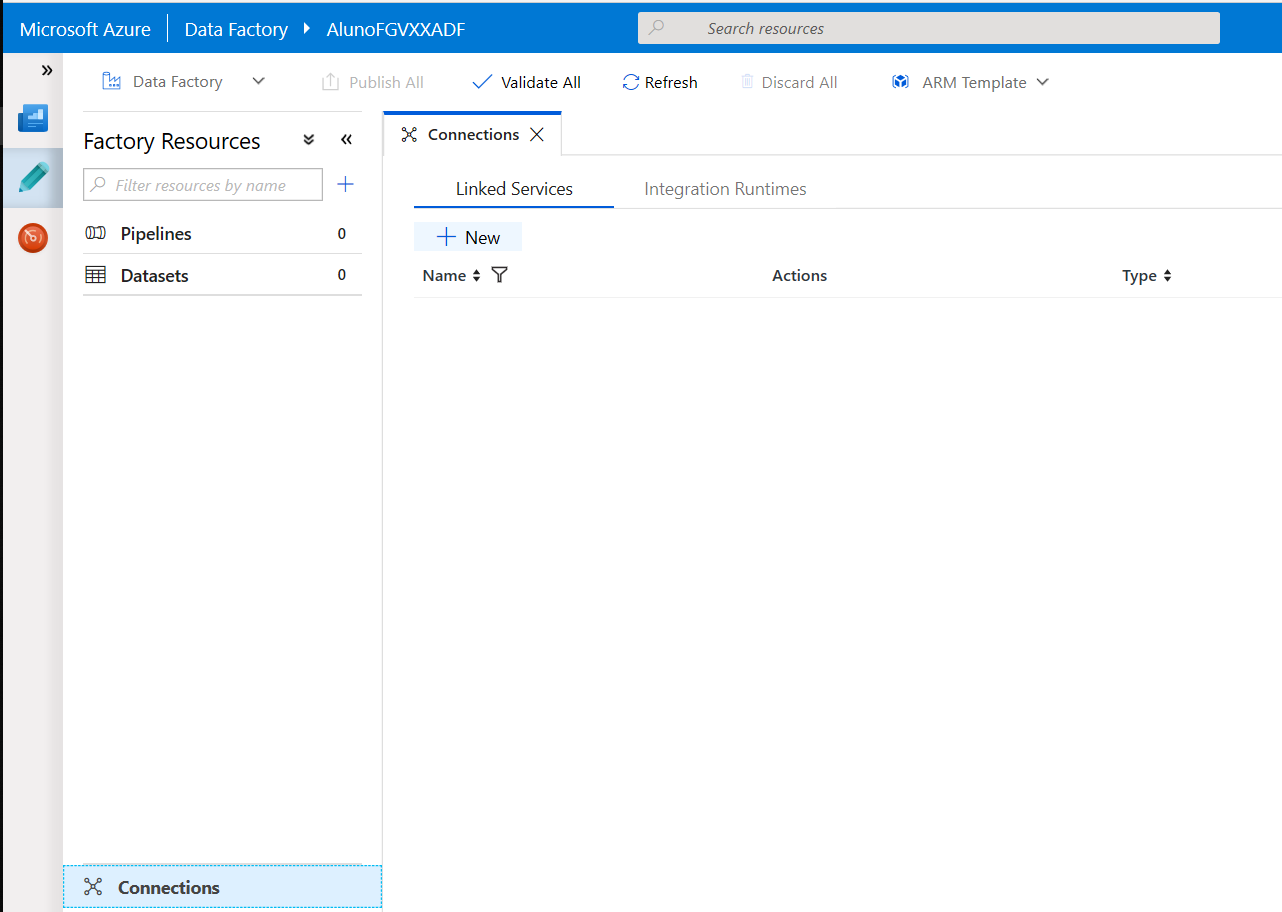
1. Na página **Introdução**, acesse a guia **Autor** no painel esquerdo.



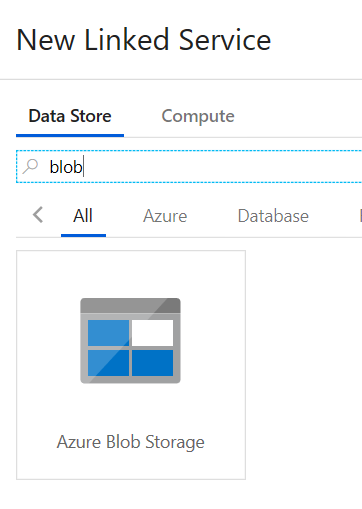
# **Criar um serviço vinculado**

Nesta etapa, você criará um serviço vinculado para vincular sua Conta de Armazenamento do Azure ao Data Factory. O serviço vinculado tem as informações de conexão que o serviço do Data Factory usa no tempo de execução para se conectar a ele.

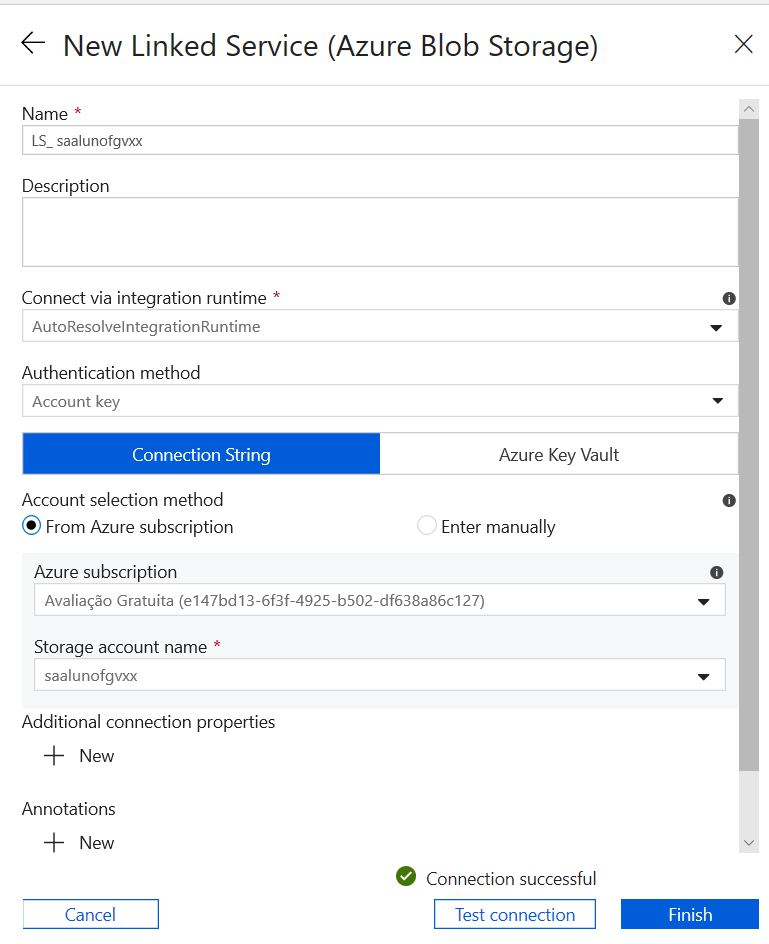
1. Clique em **Conexões** e, em seguida, selecione o botão **Novo** na barra de ferramentas.



1. Na página **Novo Serviço Vinculado**, selecione **Armazenamento de Blobs do Azure** e selecione **Continuar**.



1. Conclua as seguintes etapas:
   1. Para o campo **Nome**, insira **LS\_ saalunofgvxx**.
   2. Insira o nome da sua Conta de Armazenamento do Azure em **Nome da conta de armazenamento**.
   3. Selecione **Testar conectividade** para confirmar se o serviço do Data Factory pode se conectar à conta de armazenamento.
   4. Para salvar o serviço vinculado, selecione **Concluir**.



# **Criar conjuntos de dados**

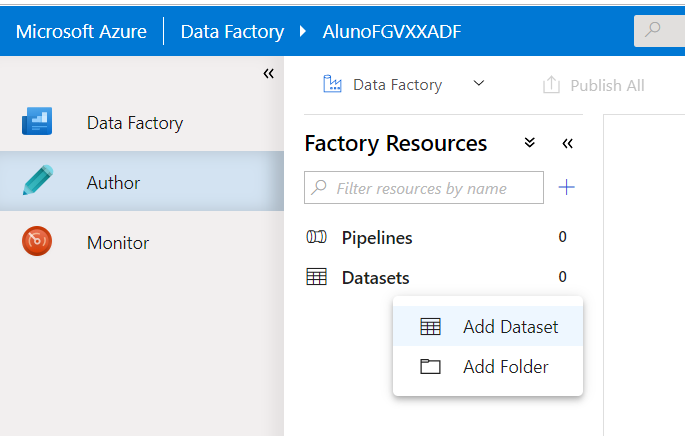
Neste procedimento, você criará dois conjuntos de dados: **InputDataset** e **OutputDataset**. Esses conjuntos de dados são do tipo **AzureBlob**. Eles se referem ao Serviço vinculado do Armazenamento do Azure que você criou na seção anterior.

O conjunto de dados de entrada representa os dados de origem na pasta de entrada. Na definição de conjunto de dados de entrada, especifique o contêiner de blob (**adftutorial**), a pasta (**entrada**) e o arquivo (**emp.txt**) que contém os dados de origem.

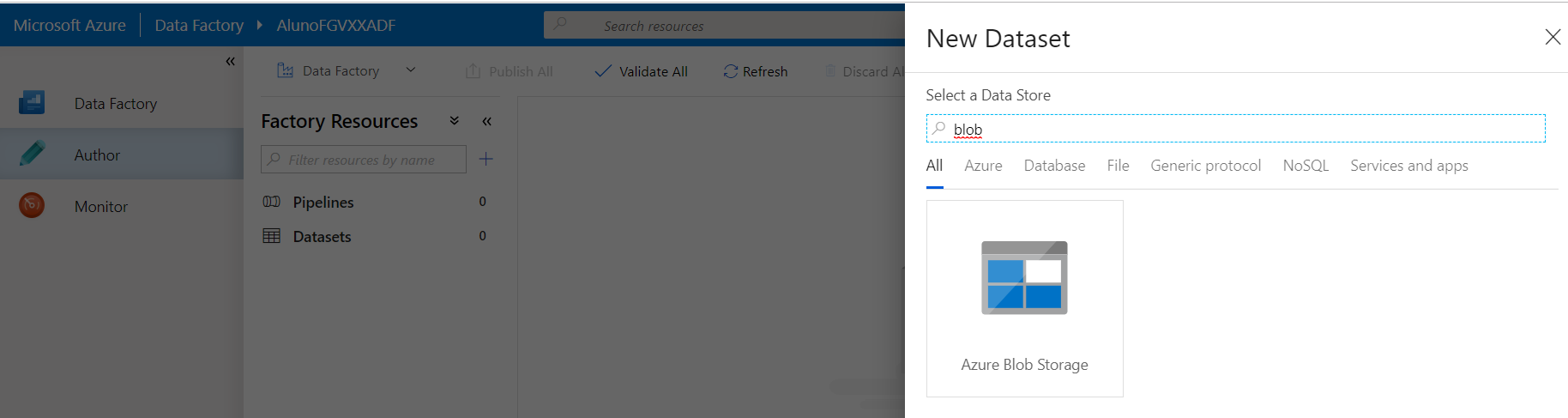
Esse conjunto de dados de saída representa os dados que são copiados para o destino. Na definição de conjunto de dados de saída, especifique o contêiner de blob (**adftutorial**), a pasta (**saída**) e o arquivo para o qual os dados são copiados. Cada execução de um pipeline tem uma ID exclusiva associada a ele. Você pode acessar essa ID, usando a variável do sistema **RunId**. O nome do arquivo de saída é avaliado dinamicamente com base na ID de execução do pipeline.

Nas configurações de serviço vinculado, você especificou a conta de armazenamento do Azure que contém os dados de origem. Nas configurações do conjunto de dados de origem, especifique onde exatamente os dados de origem residem (contêiner de blob, pasta e arquivo). Nas configurações de conjunto de dados do coletor, especifique para onde os dados são copiados (contêiner de blob, pasta e arquivo).

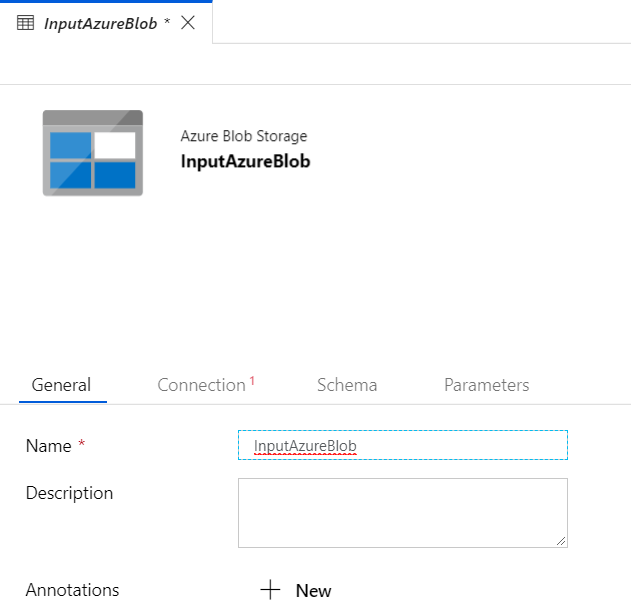
1. Selecione o botão **+** (mais) adição e, em seguida, selecione **Conjunto de Dados**.



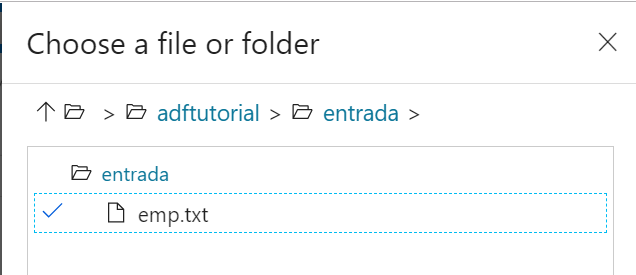
1. Na página **Novo Conjunto de Dados**, selecione **Armazenamento de Blobs do Azure** e selecione **Concluir**.



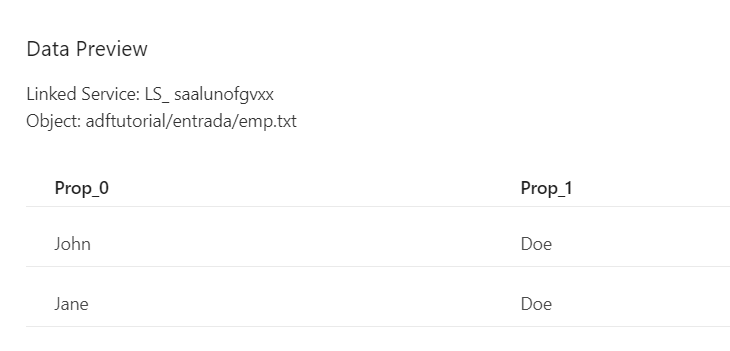
1. Na janela **Geral** do conjunto de dados, insira **InputDataset** para o **Nome**.



1. Alterne para a guia **Conexão** e siga estas etapas:
   1. Selecione **AzureStorageLinkedService** para **Serviço vinculado**.
   2. Para **Caminho do arquivo**, selecione o botão **Procurar**.
   3. Na janela **Escolher um arquivo ou uma pasta**, navegue até a pasta **entrada** no contêiner **adftutorial**, selecione o arquivo **emp.txt** e clique em **Concluir**.



* 1. (opcional) Selecione **Visualizar dados** para visualizar os dados no arquivo emp.txt.

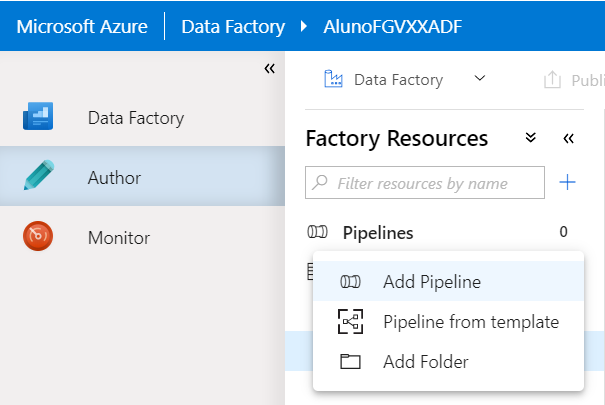


1. Repita as etapas para criar o conjunto de dados de saída:
   1. Selecione o botão **+** (mais) adição e, em seguida, selecione **Conjunto de Dados**.
   2. Na página **Novo Conjunto de Dados**, selecione **Armazenamento de Blobs do Azure** e selecione **Concluir**.
   3. Na tabela **Geral**, especifique **OutputDataset** para o nome.
   4. Na guia **Conexão**, selecione **AzureStorageLinkedService** como serviço vinculado e, no campo de diretório, insira **adftutorial/output** para a pasta. Se a pasta de **saída** não existir, a atividade de cópia a cria em tempo de execução.

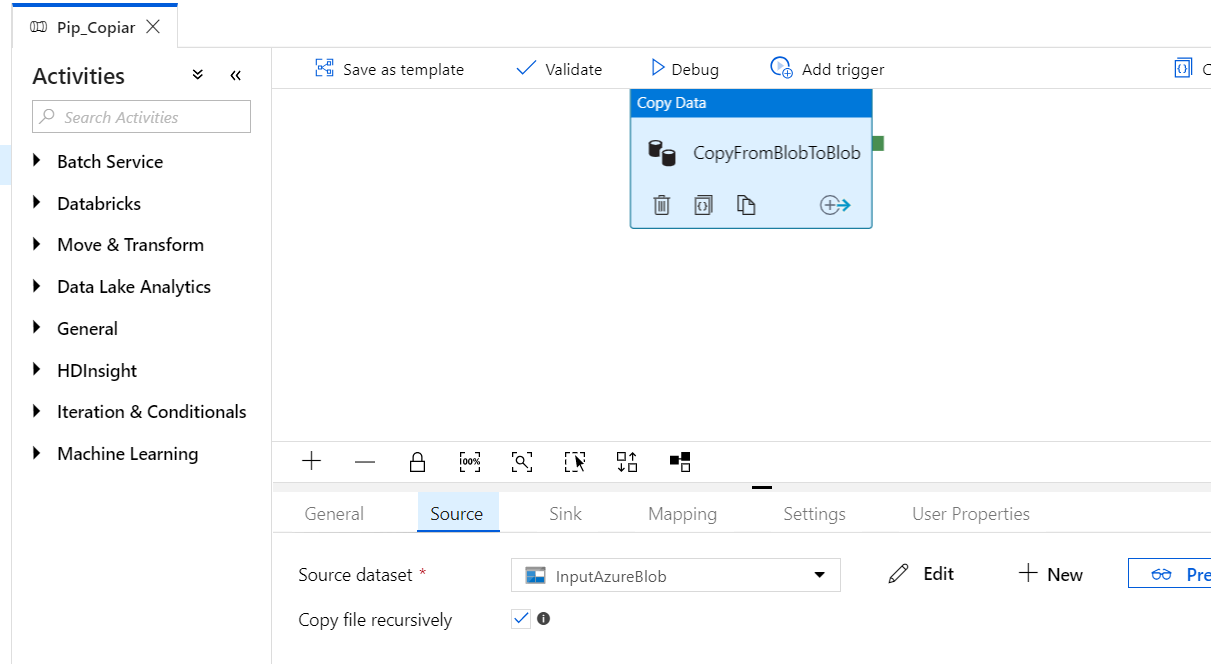
# **Criar um pipeline**

Neste procedimento, você criará e validará um pipeline com uma atividade Copiar que usa o conjunto de dados de entrada e saída. A Atividade de cópia copia os dados do arquivo especificado por você nas configurações do conjunto de dados de entrada para o arquivo especificado por você nas configurações do conjunto de dados de saída. Se o conjunto de dados de entrada especifica apenas uma pasta (não o nome de arquivo), a Atividade de cópia copia todos os arquivos da pasta de origem para o destino.

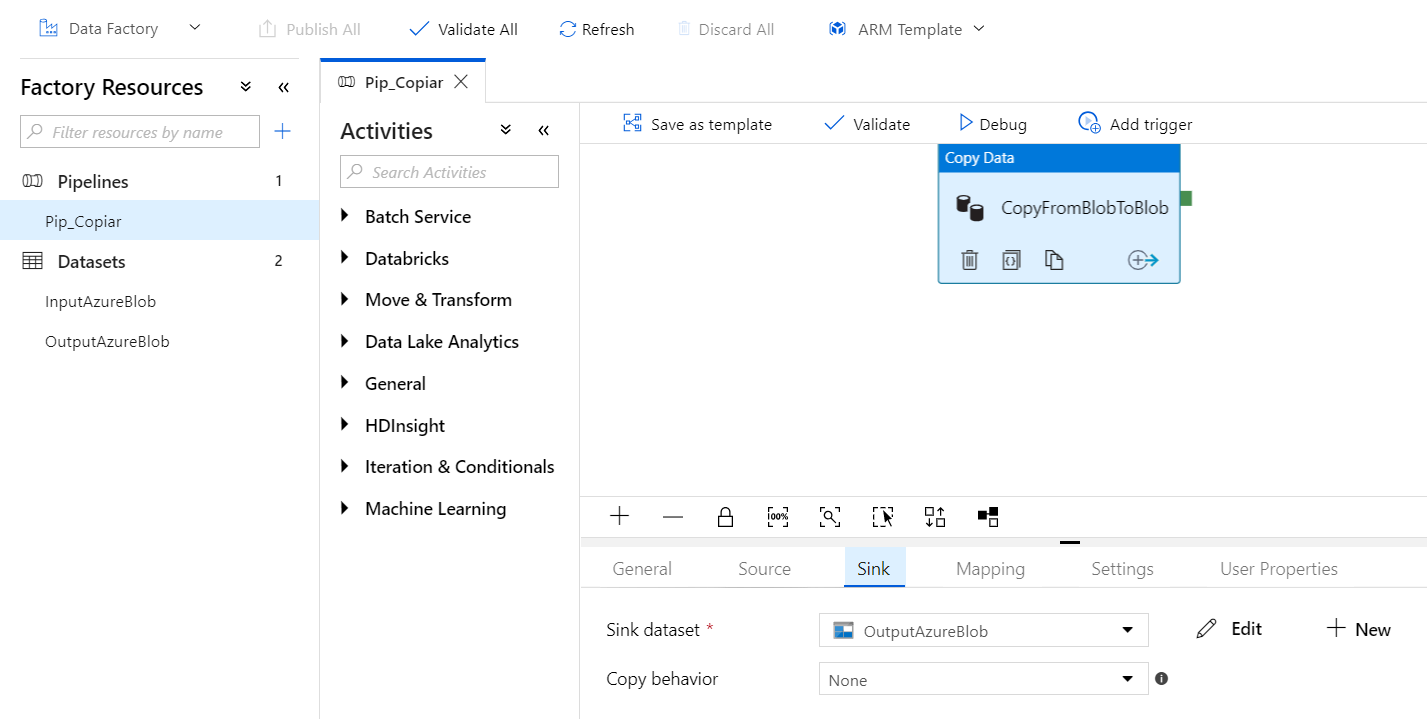
1. Selecione o botão **+** (mais) adição e, em seguida, selecione **Pipeline**.



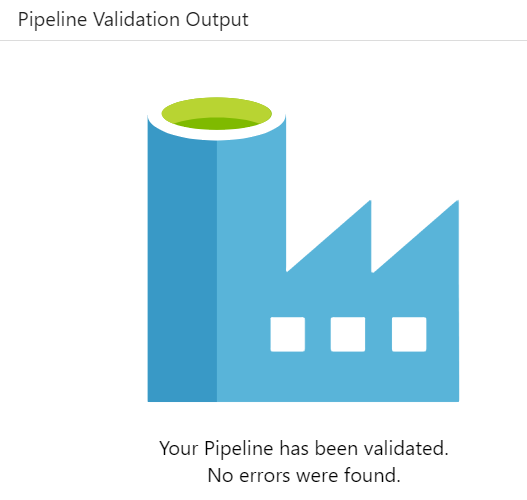
1. Na guia **Geral**, especifique **CopyPipeline** para o **nome**.
2. Na caixa de ferramentas **Atividades**, expanda **Mover e transformar**. Arraste e solte a atividade de **Cópia** da caixa de ferramentas **Atividades** para a superfície do designer do pipeline. Você também pode pesquisar atividades na caixa de ferramentas **Atividades**. Especifique **CopyFromBlobToBlob** para o **Nome**.
3. Alterne para a guia **Fonte** nas configurações da atividade de cópia e selecione **InputDataset** para o **Conjunto de dados de origem**.



1. Alterne para a guia **Coletor** nas configurações da atividade de cópia e selecione **OutputDataset** para o **Conjunto de dados do coletor**.



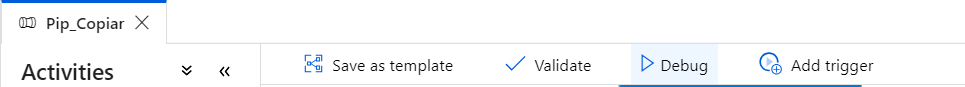
1. Clique em **Validar** na barra de ferramentas do pipeline sobre a tela para validar as configurações de pipeline. Confirme se esse pipeline foi validado com êxito. Para fechar a saída de validação, selecione o botão **>>** (seta para a direita).



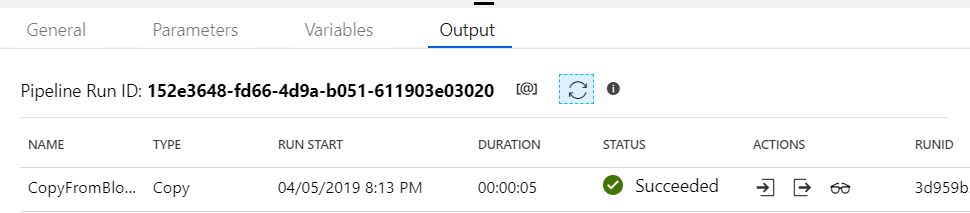
# **Depurar o pipeline**

Nesta etapa, você depura o pipeline antes de implantá-lo no Data Factory.In this step, you debug the pipeline before deploying it to Data Factory.

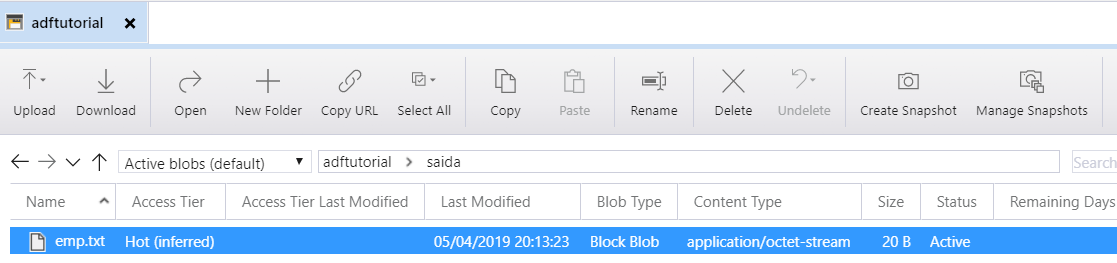
1. Na barra de ferramentas do pipeline acima da tela, clique em **Depurar** para disparar uma execução de teste.On the pipeline toolbar above the canvas, click **Debug** to trigger a test run.



1. Confirme se você vê o status da execução do pipeline na guia **Saída** das configurações do pipeline na parte inferior.Confirm that you see the status of the pipeline run on the **Output** tab of the pipeline settings at the bottom.



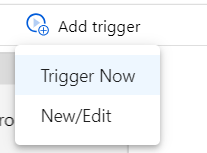
1. Confirme que você vê um arquivo de saída na pasta **saída** do contêiner **adftutorial**.Confirm that you see an output file in the **output** folder of the **adftutorial** container. Se a pasta de saída não existir, o serviço do Data Factory a cria automaticamente.If the output folder does not exist, the Data Factory service automatically creates it.

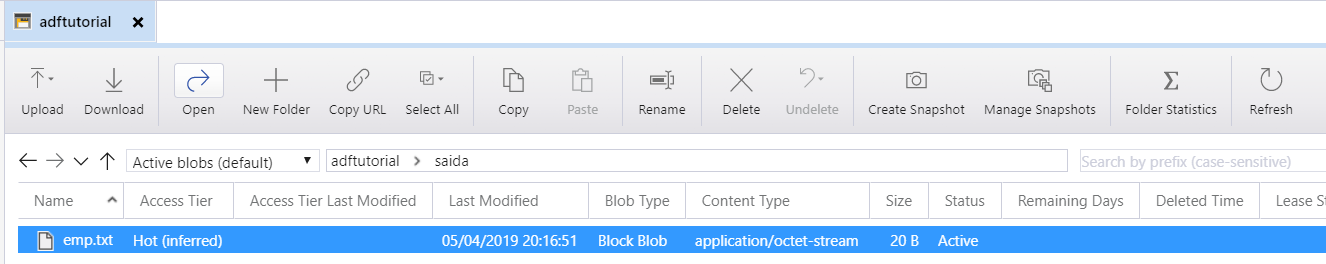


# **Disparar o pipeline manualmente**

Nesta procedimento, você implanta entidades (serviços vinculados, conjuntos de dados, pipelines) ao Azure Data Factory.In this procedure, you deploy entities (linked services, datasets, pipelines) to Azure Data Factory. Depois, dispare manualmente a execução do pipeline.Then, you manually trigger a pipeline run.

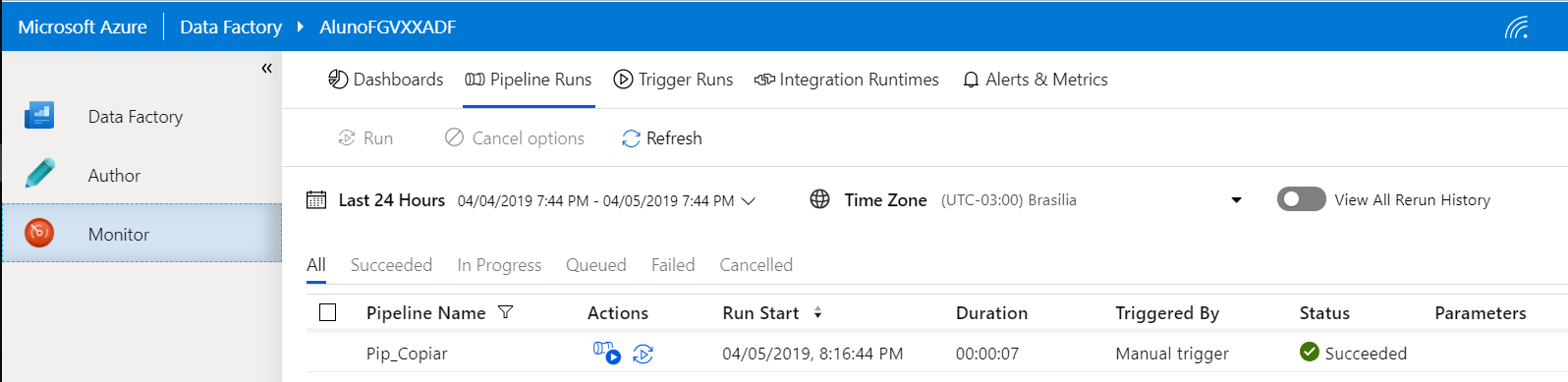
1. Antes de disparar um pipeline, você deve publicar as entidades no Data Factory.Before you trigger a pipeline, you must publish entities to Data Factory. Para publicar, selecione **Publicar Tudo** na parte superior.To publish, select **Publish All** on the top.
2. Para disparar o pipeline manualmente, selecione **Gatilho** na barra de ferramentas do pipeline e selecione **Disparar Agora**.To trigger the pipeline manually, select **Trigger** on the pipeline toolbar, and then select **Trigger Now**.





# **Monitorar o Pipeline**

1. Alterne para a guia **Monitorar** à esquerda.Switch to the **Monitor** tab on the left. Use o botão **Atualizar** para atualizar a lista.Use the **Refresh** button to refresh the list.

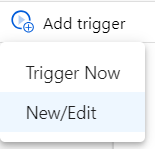


1. Selecione **Exibir Execuções de Atividade** em **Ações**.Select the **View Activity Runs** link under **Actions**. Você vê o status de execução da atividade de cópia nesta página.You see the status of the copy activity run on this page.
2. Para exibir detalhes sobre a operação de cópia, selecione **Detalhes** (imagem de óculos) na coluna **Ações**.To view details about the copy operation, select the **Details** (eyeglasses image) link in the **Actions** column. Para obter detalhes sobre as propriedades, confira [Visão geral da atividade de cópia](https://docs.microsoft.com/pt-br/azure/data-factory/copy-activity-overview).For details about the properties, see [Copy Activity overview](https://docs.microsoft.com/pt-br/azure/data-factory/copy-activity-overview).
3. Confirme que você vê um arquivo novo na pasta de **saída**.Confirm that you see a new file in the **output** folder.
4. Você pode alternar de volta para o modo de exibição **Execuções do pipeline** no modo de exibição **Execuções de Atividade** selecionando **Pipelines**.You can switch back to the **Pipeline Runs** view from the **Activity Runs** view by selecting the **Pipelines** link.

# **Disparar o pipeline em um cronograma**

Este procedimento é opcional neste tutorial.This procedure is optional in this tutorial. Você pode criar um *agendador de gatilho* para agendar a execução periódica do pipeline (por hora, diariamente, e assim por diante).You can create a *scheduler trigger* to schedule the pipeline to run periodically (hourly, daily, and so on). Nesta procedimento, você cria um gatilho para ser executado a cada minuto até a data e hora de término especificadas.In this procedure, you create a trigger to run every minute until the end date and time that you specify.

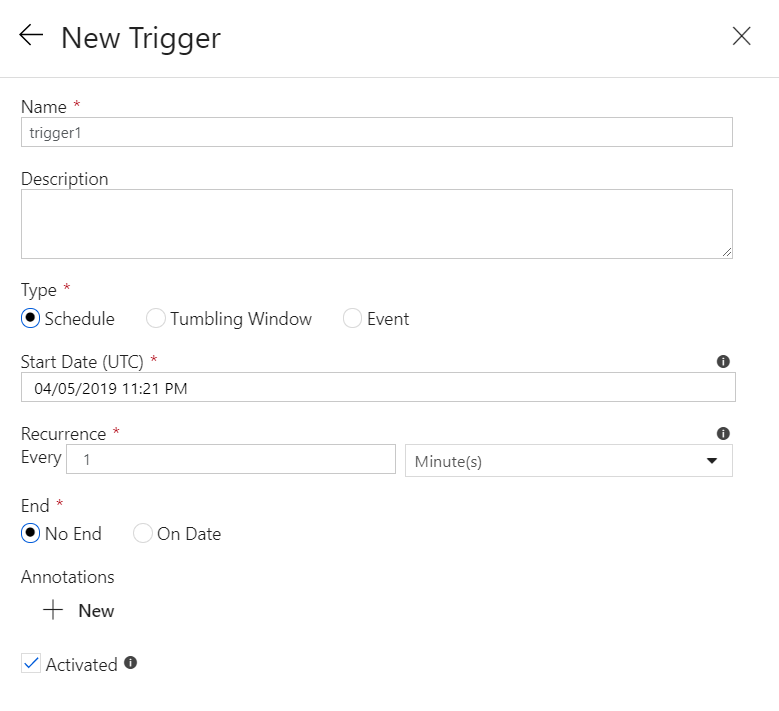
1. Alterne para a guia **Autor**.Switch to the **Author** tab.
2. Vá até o pipeline, selecione **Gatilho** na barra de ferramentas do pipeline e depois selecione **Novo/Editar**.Go to your pipeline, select **Trigger** on the pipeline toolbar, and then select **New/Edit**.



1. Na página **Adicionar gatilhos**, selecione **Escolher gatilho** e, em seguida, selecione **Novo**.On the **Add Triggers** page, select **Choose trigger**, and then select **New**.
2. Na página **Novo gatilho**, no campo **Final**, selecione **Na Data**, especifique como hora de término alguns minutos após a hora atual e selecione **Aplicar**.On the **New Trigger** page, under **End**, select **On Date**, specify an end time a few minutes after the current time, and then select **Apply**.

Um custo associado a cada execução de pipeline, então, especifique o a hora de término como apenas alguns minutos após a hora de início.A cost is associated with each pipeline run, so specify the end time only minutes apart from the start time. Verifique se está como o mesmo dia.Ensure that it's the same day. No entanto, verifique se há tempo suficiente para a execução do pipeline entre a hora da publicação e a hora de término.However, ensure that there is enough time for the pipeline to run between the publish time and the end time. O gatilho só entra em vigor depois de você publicar a solução no Data Factory, e não ao salvar o gatilho na interface do usuário.The trigger comes into effect only after you publish the solution to Data Factory, not when you save the trigger in the UI.

1. Na página **Novo gatilho**, selecione a caixa de seleção **Ativado** e, em seguida, selecione **Avançar**.On the **New Trigger** page, select the **Activated** check box, and then select **Next**.
2. Examine a mensagem de aviso e selecione **Concluir**.Review the warning message, and select **Finish**.



1. Clique em **Publicar Tudo** para publicar as alterações no Data Factory.Select **Publish All** to publish changes to Data Factory.
2. Alterne para a guia **Monitorar** à esquerda.Switch to the **Monitor** tab on the left. Selecione **Atualizar** para atualizar a lista.Select **Refresh** to refresh the list. Você verá que o pipeline é executado uma vez por minuto desde o momento da publicação até hora de término.You see that the pipeline runs once every minute from the publish time to the end time.
3. Observe os valores na coluna **Disparado Por**.Notice the values in the **Triggered By** column. A execução do gatilho manual foi feita em uma etapa anterior (**Disparar agora**).The manual trigger run was from the step (**Trigger Now**) that you did earlier.
4. Alterne para o modo de exibição **Execuções de gatilho**.Switch to the **Trigger Runs** view.
5. Confirme que um arquivo de saída é criado para cada execução de pipeline até a data e hora de término especificadas na pasta **saída**.

# **Considerações finais**

Neste tutorial, foi apresentado as seguintes tarefas:You learned how to:

* Criando os arquivos e pastas de entrada
* Criar um data factory
* Criar um serviço vinculado
* Criar conjuntos de dados
* Criar um pipeline
* Depurar o pipeline
* Disparar o pipeline manualmente
* Monitorar o Pipeline
* Disparar o pipeline em um cronogramaConnect to the database with [**SQL Server Management Studio**](https://docs.microsoft.com/sql/ssms/sql-server-management-studio-ssms) (SSMS)

Esse material foi adaptado a partir do link original abaixo:

* <https://docs.microsoft.com/en-us/azure/sql-database/sql-database-design-first-database>

