

# MVP de Engenharia de Dados

Thiago Correa do Quinto

## 1. Definição do Problema

O conjunto de dados utilizado neste projeto foi o **World GDP by Country, Region, and Income Group**. Este conjunto de dados compreende vários indicadores-chave relacionados ao PIB, cobrindo de 1960 até 2022 e foi obtido do Grupo Banco Mundial, uma instituição renomada em dados econômicos globais. O Conjunto de Dados do PIB (Produto Interno Bruto) Mundial foi elaborado para fornecer insights abrangentes sobre o desempenho econômico de países em todo o mundo, categorizados por região e grupo de renda. Em uma economia global cada vez mais interconectada, compreender o panorama econômico de diferentes países, regiões e grupos de renda é vital para formuladores de políticas, pesquisadores e empresas.

Fonte: <https://www.kaggle.com/datasets/sazidthe1/world-gdp-data/data>

O conteúdo do projeto é formado por dois arquivos CSV. Esses conjuntos de dados oferecem uma visão abrangente da economia global, centrada no Produto Interno Bruto (PIB). No conjunto de dados principal (`gdp_data.csv`), você encontrará valores específicos do PIB por país, anos e códigos de país exclusivos. O conjunto de dados suplementar (`country_codes.csv`) complementa-o fornecendo regiões de países e classificações de grupos de renda, facilitando uma análise multifacetada do panorama econômico global.

O conjunto de dados principal (`gdp_data.csv`) inclui as seguintes colunas:

**country\_name:** O nome do país.

**country\_code:** Um identificador único para cada país.

**year:** O ano ao qual os dados correspondem.

**value:** O valor do PIB em US\$ para o ano e país correspondentes.

O conjunto de dados suplementar (`country_codes.csv`) inclui as seguintes colunas:

**country\_code:** Um identificador único que se relaciona com o conjunto de dados principal.

**region:** A região à qual cada país pertence.

**income\_group:** A classificação do grupo de renda de cada país.

O objetivo deste conjunto de dados é fornecer uma visão abrangente do Produto Interno Bruto (PIB) global por país, região e grupo de renda ao longo do tempo. Ele serve como uma ferramenta valiosa para análise e pesquisa econômica, permitindo que formuladores de políticas, pesquisadores e empresas entendam melhor a dinâmica econômica mundial e tomem decisões informadas com base nessas informações.

Com base neste projeto vamos responder algumas questões baseadas nos dados obtidos:

1. Quais são os países com os maiores PIBs globais em 2022?
2. Qual é a tendência de crescimento do PIB global ao longo das décadas, com base nos dados desde 1960?

3. Quais regiões do mundo estão experimentando um rápido crescimento econômico e quais estão ficando para trás?
4. Como as classificações de grupos de renda se relacionam com o PIB de diferentes países e regiões?
5. Quais países ou regiões tiveram flutuações significativas em seu PIB ao longo do tempo e quais foram as principais causas dessas flutuações?
6. Como a crise econômica global de 2008 afetou o PIB de diferentes países e regiões?
7. Quais são os países ou regiões que mais dependem de setores específicos (por exemplo, agricultura, indústria, serviços) em sua economia?
8. Qual é a distribuição regional do PIB global e como isso mudou ao longo das décadas?
9. Quais países ou regiões têm visto um declínio consistente em seu PIB e quais medidas podem ser tomadas para reverter essa tendência?
10. Como a pandemia de COVID-19 afetou o PIB global?

Essas perguntas e insights podem ajudar a orientar a análise e a tomada de decisões relacionadas à economia global, contribuindo para um melhor entendimento das tendências econômicas e das disparidades entre países e regiões.

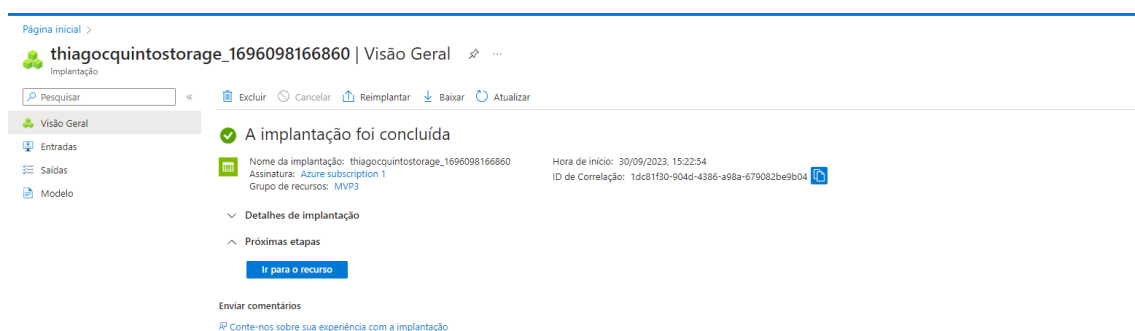
Neste trabalho será construído um pipeline de dados utilizando tecnologias na nuvem no Microsoft Azure. O pipeline irá envolver a busca, coleta, modelagem, carga e análise dos dados que será consumido pelo Power BI.

## 2. Coleta

Uma vez já definida o conjunto de dados, seguiremos para a etapa de coleta dos dados que serão hospedados em um serviço de nuvem do Microsoft Azure. A seguir estão descritas as etapas para a utilização do serviço de nuvem.

### Criação de Conta de Armazenagem

Após a inscrição e login no portal do Microsoft Azure foi adicionado um recurso de armazenagem thiagocquintostorage.



thiagocquintostorage

Conteúdo de armazenamento

Carregar blob

Arraste e solte os arquivos aqui ou Procure arquivos

Selecionar um contêiner existente

Carregar

Enviar comentários

Propriedades

Monitoramento

Funcionalidades (7)

Recomendações (0)

Tutoriais

Ferramentas + SDKs

Serviço Blob

Nome de espaço de nomes

Desabilitado

Camada de acesso padrão

Hot

Acesso anônimo ao blob

Desabilitado

Exclusão reversível do blob

Habilitado (7 dias)

Exclusão reversível do contêiner

Habilitado (7 dias)

Controle de versão

Desabilitado

Feed de alterações

Desabilitado

NFS v3

Desabilitado

Permitir replicação entre locais

Desabilitado

Serviço de arquivo

Compartilhamento de arquivo grande

Desabilitado

Active Directory

Não configurado

Definir permissões padrão de nível de compartilhamento

Desabilitado

Exclusão reversível

Habilitado (7 dias)

Compartilhar capacidade

5 TiB

Segurança

Exigir transferência segura para o cliente

da API REST

Acesso à chave de conta de armazenamento

Versão mínima do TLS

Criptografia de infraestrutura

Rede

Permitir o acesso de

Número de conexões de ponto de extremidade privado

Roteamento de rede

Acesso para serviços Microsoft

Tipo de ponto de extremidade

Em seguida foram carregados os arquivos para o Contêiner thiagocquintocontainer:

Microsoft Azure

thiagocquintocontainer

Contêiner

Método de autenticação: Chave de acesso (Alternar para uma Conta de Usuário do Azure AD)

Local: thiagocquintocontainer

Buscar blobs por prefixo (diferenciar maiúsculas de minúsculas)

Mostrar blobs excluídos

Nome	Modificado	Camada de acesso	Status do arquivo	Tipo de blob	Tamanho	Estado de concessão
country_codes.csv	30/09/2023, 15:28:53	Principal (inferido)		Blob de blocos	9,07 KiB	Disponível ***
gdp_data.csv	30/09/2023, 15:28:54	Principal (inferido)		Blob de blocos	445,94 KiB	Disponível ***

A seguir foi criado um SQL Database thiagocquintosql.

## Criar Servidor do Banco de Dados SQL ...



Microsoft



### Detalhes do servidor

Insira as configurações necessárias para este servidor, incluindo o fornecimento de um nome e localização. Esse servidor será criado na mesma assinatura e grupo de recursos que seu banco de dados.

Nome do servidor *	<input type="text" value="thiagocquintosql"/> ✓
	.database.windows.net
Localização *	<input type="text" value="(US) East US"/> ▼

### Autenticação

 Azure Active Directory (Azure AD) is now Microsoft Entra ID. [Saiba mais](#) 

Selecione seus métodos de autenticação preferidos para acessar este servidor. Crie um logon de administrador servidor e uma senha para acessar seu servidor com autenticação SQL, selecione apenas a autenticação Microsoft Azure Active Directory [Saiba mais](#)  usando um usuário, grupo ou aplicativo existente do Microsoft Azure Active Directory como administrador do Microsoft Azure Active Directory [Saiba mais](#)  ou selecione a autenticação do SQL e do Microsoft Azure Active Directory.

Método de autenticação	<input type="radio"/> Usar somente a autenticação do Azure Active Directory (Azure AD)
	<input type="radio"/> Usar a autenticação do SQL e do Microsoft Azure Active Directory
	<input checked="" type="radio"/> Usar autenticação SQL
Logon do administrador do servidor *	<input type="text" value="serveradmin"/> ✓
Senha *	<input type="password" value="....."/> ✓
Confirmar senha *	<input type="password" value="....."/> ✓

OK

E todos os recursos associados foram implantados.



Página inicial >

## Microsoft.DataFactory-20230930154023 | Visão Geral

Implantação

Pesquisar

Excluir Cancelar Reimplantar Baixar Atualizar

Visão Geral

Entradas

Saídas

Modelo

**A implantação foi concluída**

Nome da implantação : Microsoft.DataFactory-20230930154023 Hora de início : 30/09/2023, 15:41:32

Assinatura : Azure subscription 1 ID de correlação : ce9e161e-34ab-474f-927d-82f2ff813a24

Grupo de recursos : MVP3

**Detalhes de implantação**

Recurso	Tipo	Status	Detalhes da operação
thiagocquintofactory	Data factory (V2)	OK	<a href="#">Detalhes da operação</a>

**Próximas etapas**

[Ir para o recurso](#)

Enviar comentários

[Conte-nos sobre sua experiência com a implantação](#)

Em seguida, o Factory criado inicializamos o Data Factory Studio.

Página inicial > Recente >

## thiagocquintofactory

Data factory (V2)

Pesquisar

Excluir

Visão geral

Log de atividade

IAM (Controle de acesso)

Marcações

Diagnosticar e resolver problemas

Configurações

Rede

Identidades gerenciadas

Propriedades

Bloqueios

Introdução

Início rápido

Monitoramento

Alertas

Métrica

Configurações de diagnóstico

Logs

Fundamentos

Grupo de recursos [\(move\)](#) : MVP3

Status : Succeeded

Local : East US

Assinatura [\(move\)](#) : Azure subscription 1

ID da Assinatura : e466163b-00bf-460f-b633-af037c1d686d

Tipo : Data factory (V2)

Introdução : [início rápido](#)

### Estúdio do Azure Data Factory

[Iniciar estúdio](#)

Inícios Rápidos

Tutoriais

Galeria de Modelos

Módulos de Treinamento

Monitoramento

Microsoft Azure | Data Factory | thiagocquintofactory

Pesquisar atalhos e documentação

Seus insights são importantes! Participe do nosso [breve pesquisa](#) sobre nosso recurso de nível superior do CDC e ajude-nos a aprimorar sua experiência.

O Azure Data Factory permite que você configure um repositório Git com o Azure DevOps ou o GitHub. O Git é um sistema de controle de versão que permite o controle de alterações e a colaboração mais fidedigna. [Saiba mais](#)

[Configurar repositório de código](#)

## Data factory thiagocquintofactory

[Novo](#)

Ingerir  
Copie dados em escala uma vez ou mediante agendamento.

Orquestrar  
Pipelines de dados sem código.

Transformar dados  
Transforme dados usando fluxos de dados.

Configurar SSIS  
Gerencie e execute os pacotes SSIS na nuvem.

Recursos recentes

**Não há itens para mostrar**

Seus recursos abertos recentemente serão mostrados aqui.

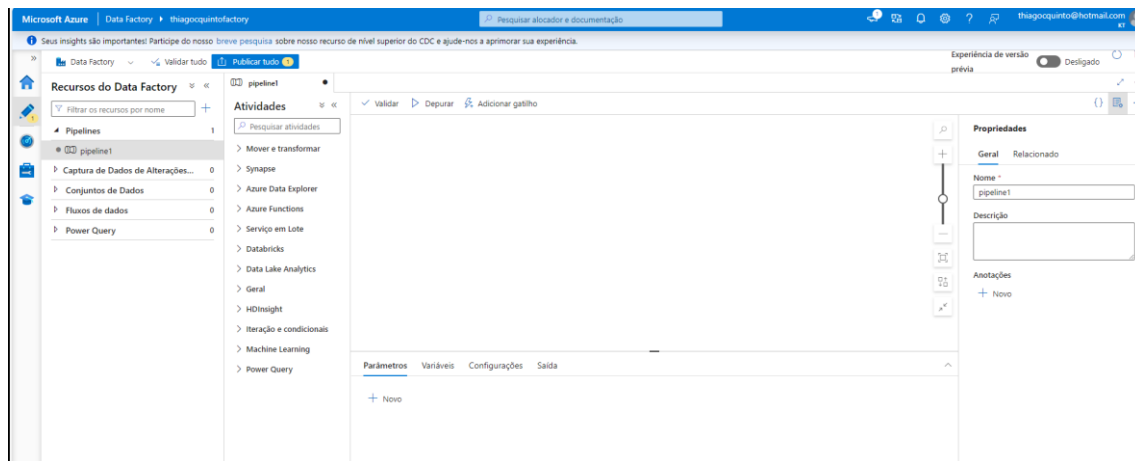
Descobrir mais

Procurar parceiros (prévia)

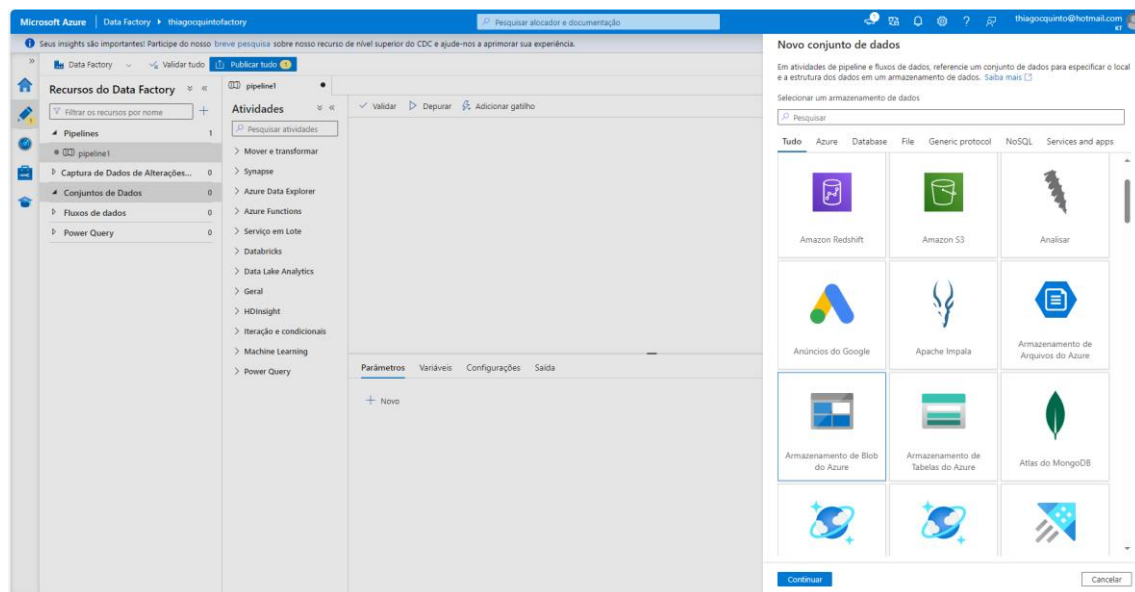
Modelos de pipeline

Modelos de pipeline SAP

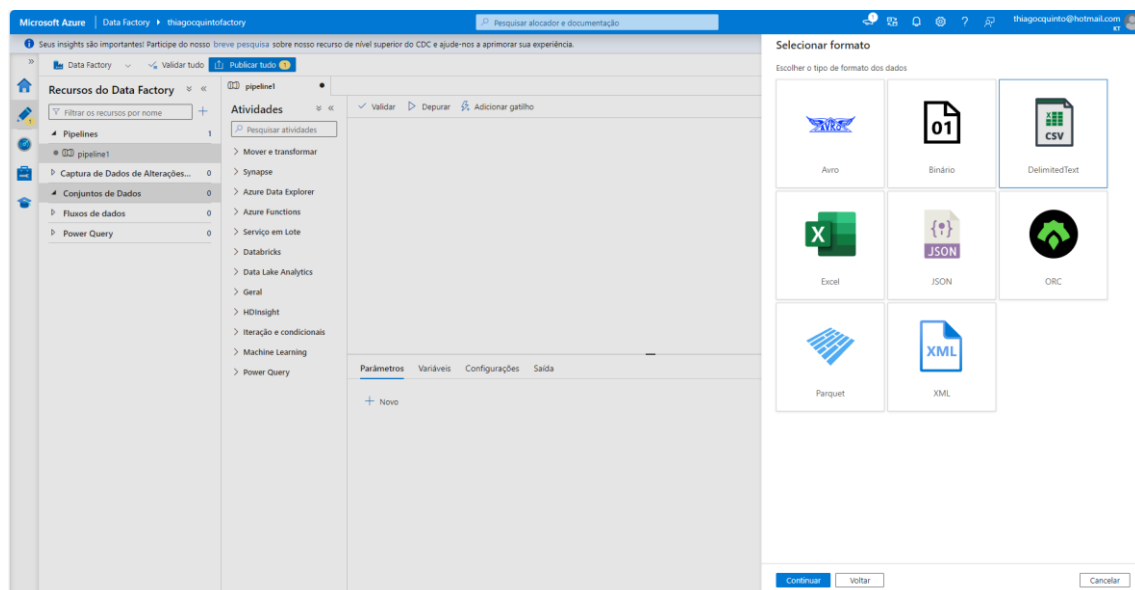
Criado um pipeline

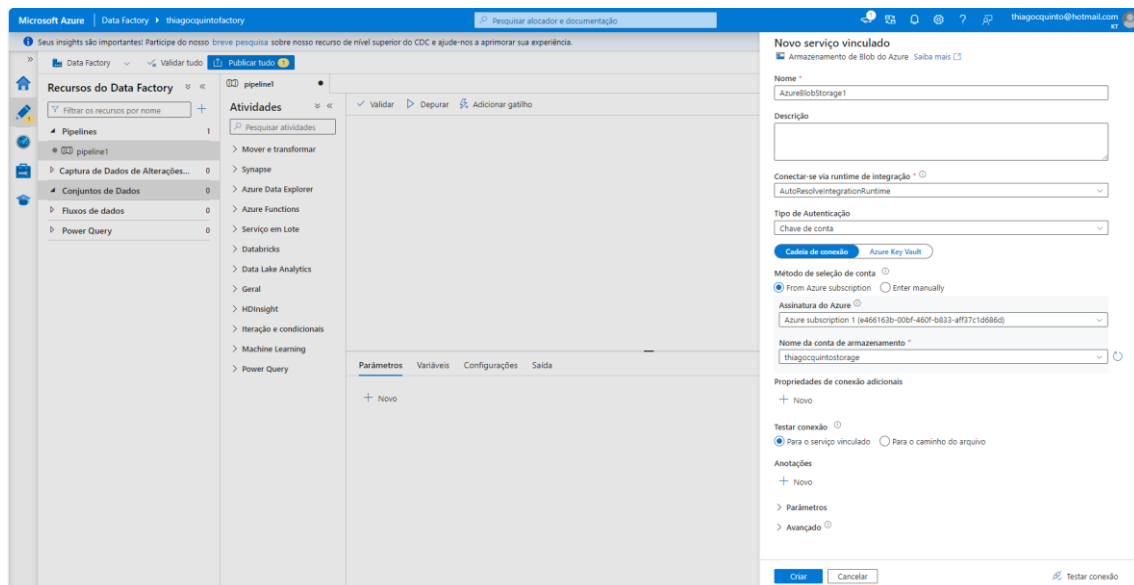


Em seguida criado Conjuntos de Dados do Blob Storage

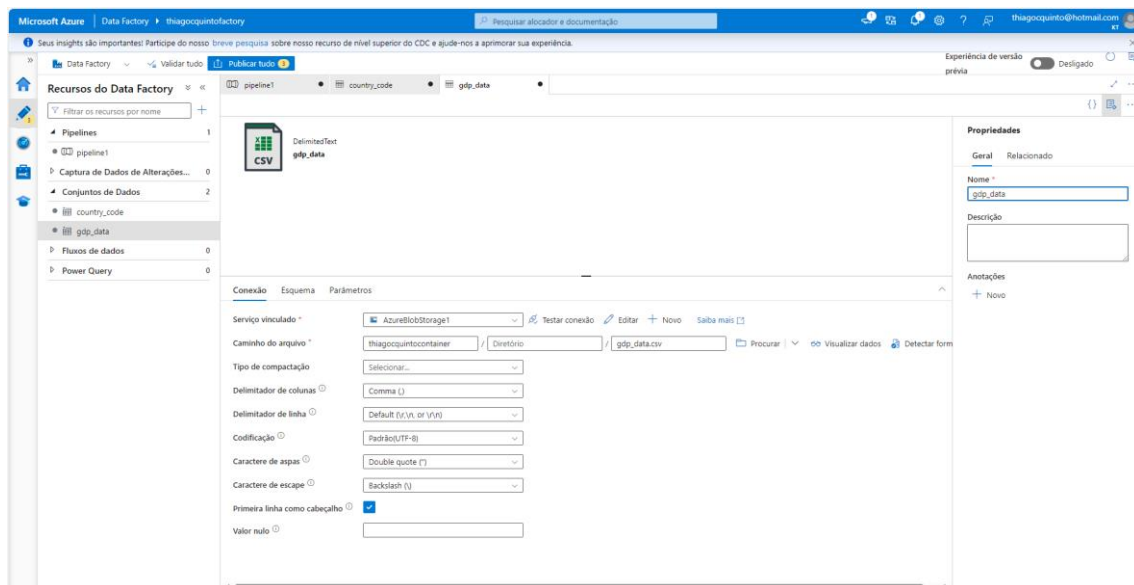


Carregados os arquivos de formato csv





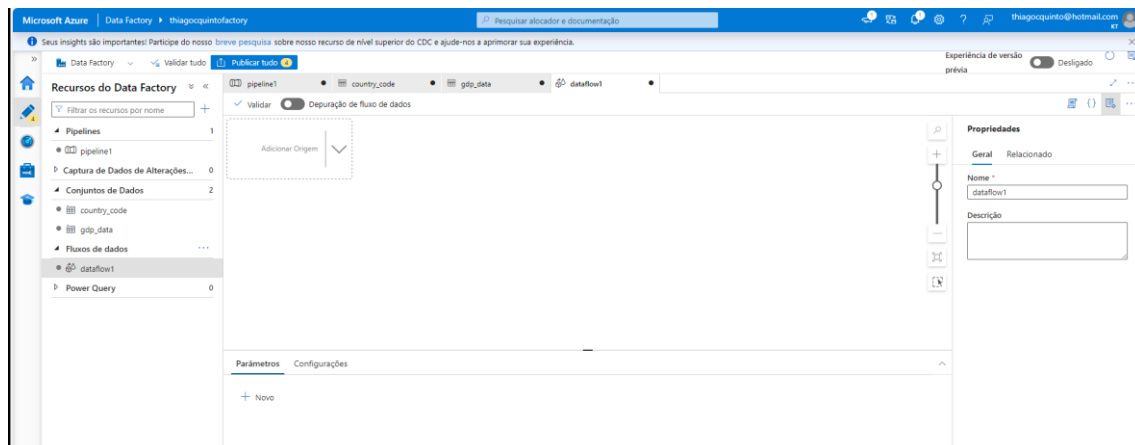
Carregados os arquivos



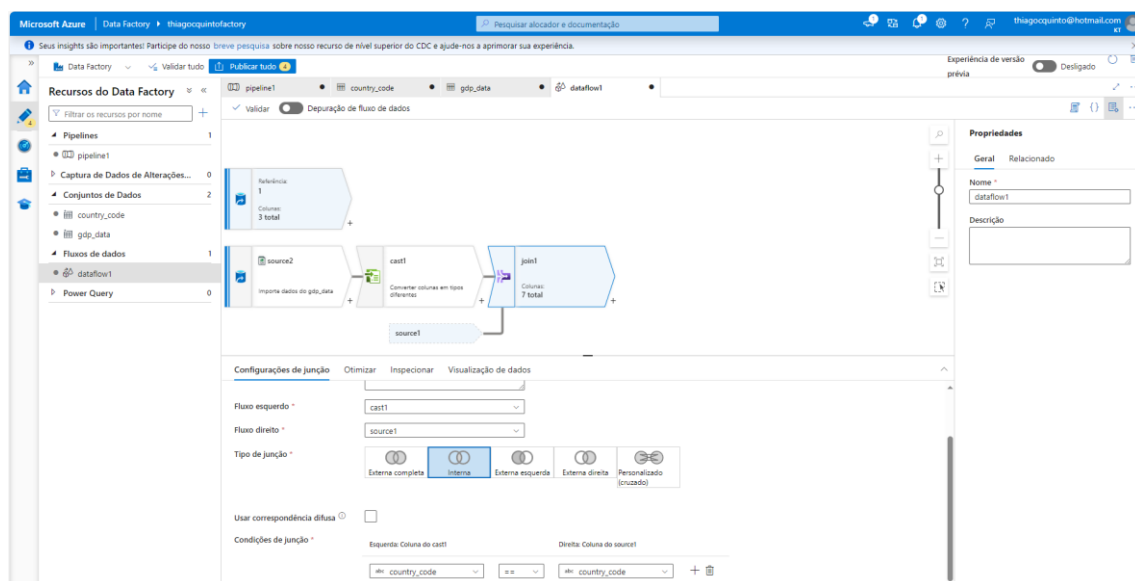
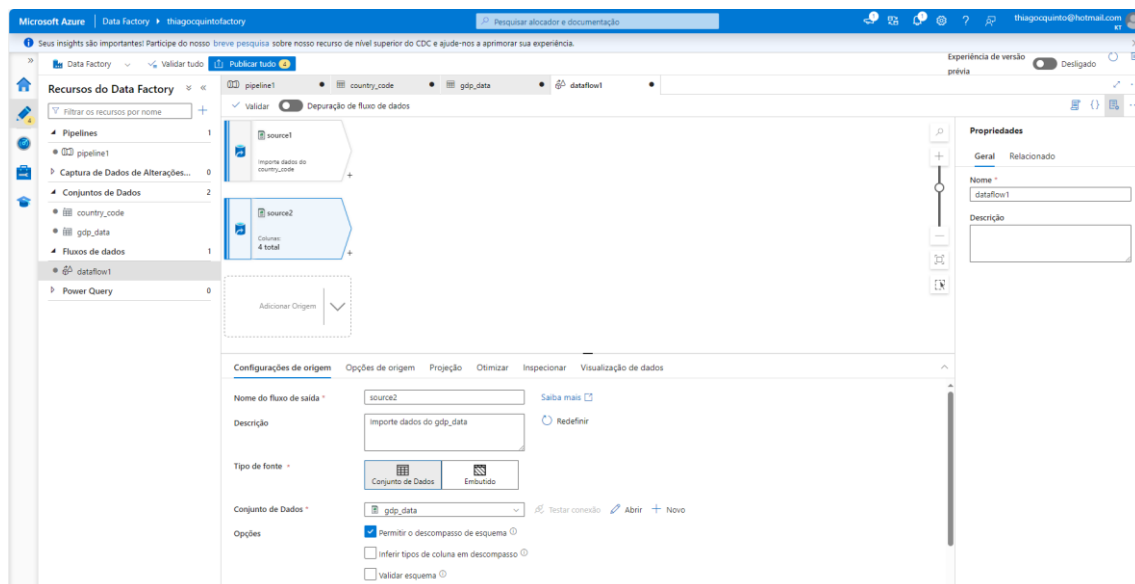
## Modelagem

Segue o passo para criação de um Dataflow





Nesta etapa, são realizados alguns ajustes nos dados via DataFlow (ETL). Os dados contidos no gdp\_data.csv além de conter os dados dos países, também tinham as agregações por região. Realizando o Inner Join com os dados do country\_code.csv, essas agregações foram excluídas evitando que os dados ficassem duplicados.



Microsoft Azure

Data Factory

thiago@continfactory.com

Seus insights são importantes! Participe do nosso breve questionário sobre nosso recurso de nível superior do CDC e ajude-nos a aprimorar sua experiência.

Data Factory

Validar tudo

Publicar tudo

Recursos do Data Factory

Filterar os recursos por nome

Pipelines

1

pipeline1

0

Captura de Dados de Alterações...

0

Conjuntos de Dados

2

country\_code

0

gdp\_data

0

Fluxos de dados

1

dataflow1

0

Power Query

0

pipeline1

country\_code

gdp\_data

dataflow1

Validar

Depuração de fluxo de dados

source1

Importe dados do country\_code

source2

Importe dados do gdp\_data

cast1

Converter colunas em tipos diferentes

join1

Junção de interna em 'cast1' e 'source1'

sink1

Colocar: 6 total

Coletor

Configurações

Erros

Mapeamento

Otimizar

Inspeccionar

Visualização de dados

Nome do fluxo de saída \*

sink1

Saiba mais

Descrição

Adicionar conjunto de dados do coletor

Redefinir

Fluxo de entrada \*

join1

Tipo de coletor \*

Conjunto de Dados

Embutido

Cache

Conjunto de Dados \*

Selecione...

+ Novo

Opções

☒ Permitir o descompasso de esquema

☐ Validar esquema

Novo conjunto de dados

Em atividades de pipeline e fluxos de dados, referência um conjunto de dados para especificar o local e a estrutura dos dados em um armazenamento de dados. Saiba mais

Selecione um armazenamento de dados

Pesquisar

Tudo

Azure

Database

File

Generic protocol

NoSQL

Services and apps

Amazon S3

Armazenamento de Blob do Azure

Azure Cosmos DB for NoSQL

Azure Data Explorer (Kusto)

Azure Data Lake Storage Gen1

Azure Data Lake Storage Gen2

Azure Synapse Analytics

Banco de Dados SQL do Azure

Banco de Dados do Azure para MySQL

Continuar

Cancelar

Microsoft Azure

Data Factory

thiagocunha@hotmail.com

Seus insights são importantes! Participe do nosso breve pesquisa sobre nosso recurso de nível superior do CDC e ajude-nos a aprimorar sua experiência.

Data Factory

Validar tudo

Publicar tudo

Recursos do Data Factory

Filter os recursos por nome

Pipelines

pipeline1

Captura de Dados de Alterações...

Conjuntos de Dados

country\_code

gdp\_data

Fluxos de dados

dataflow1

Power Query

pipeline1

country\_code

gdp\_data

dataflow1

Validar

Depuração de fluxo de dados

source1

Importe dados do country\_code

source2

Importe dados do gdp\_data

cast1

Converter colunas em tipos diferentes

join1

Junção de interna em 'cast1' e 'source1'

sink1

Colunas: 6 total

Coletor

Configurações

Erros

Mapeamento

Otimizar

Inspeccionar

Visualização de dados

Nome do fluxo de saída \*

sink1

Salva mais

Descrição

Adicionar conjunto de dados do coletor

Redefinir

Fluxo de entrada \*

join1

Tipo de coletor \*

Conjunto de Dados

Embutido

Cache

Conjunto de Dados \*

Selecionar...

+ Novo

Opções

Permitir o descompasso de esquema

Validar esquema

Novo serviço vinculado

Banco de Dados SQL do Azure

Salva mais

Nome \*

AzureSqlDatabase1

Descrição

Conectar-se via runtime de integração \*

AutoResolveIntegrationRuntime

Credenciais de conexão

Azure Key Vault

Método de seleção de conta

From Azure subscription

Enter manually

Assinatura do Azure

Azure subscription 1 (a466163b-020f-460f-b833-af37c16586d)

Nome do servidor \*

thiagocunhaosql

Nome do Banco de Dados \*

thiagocunhaosql

Tipo de Autenticação \*

SQL authentication

Nome do usuário \*

serveradmin

Senha

Azure Key Vault

Senha \*

\*\*\*\*\*

Sempre criptografado

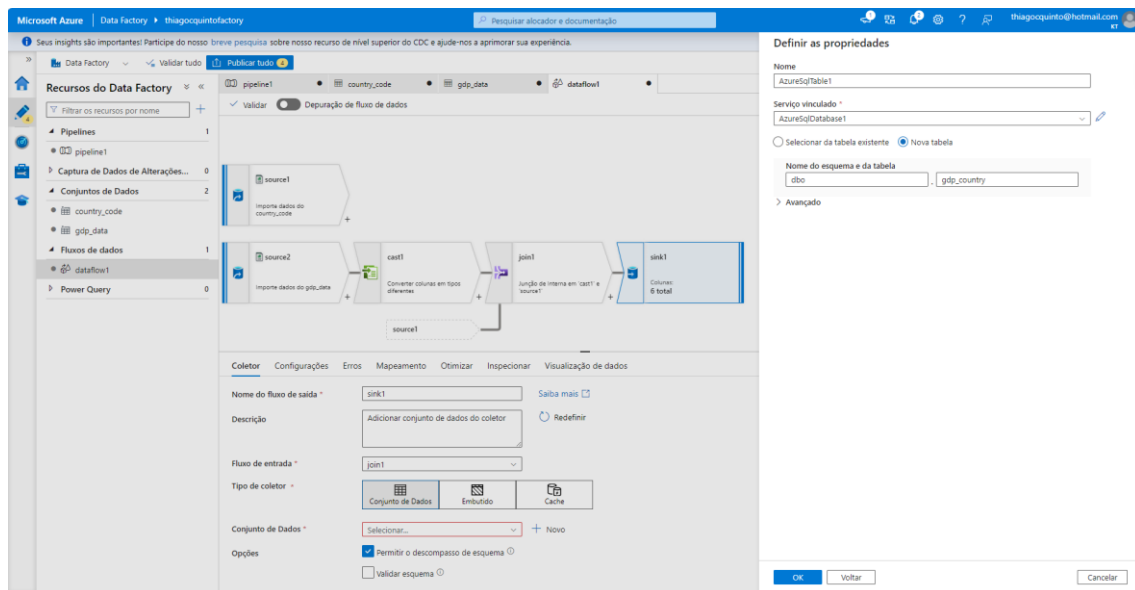
Propriedades de conexão adicionais

Criar

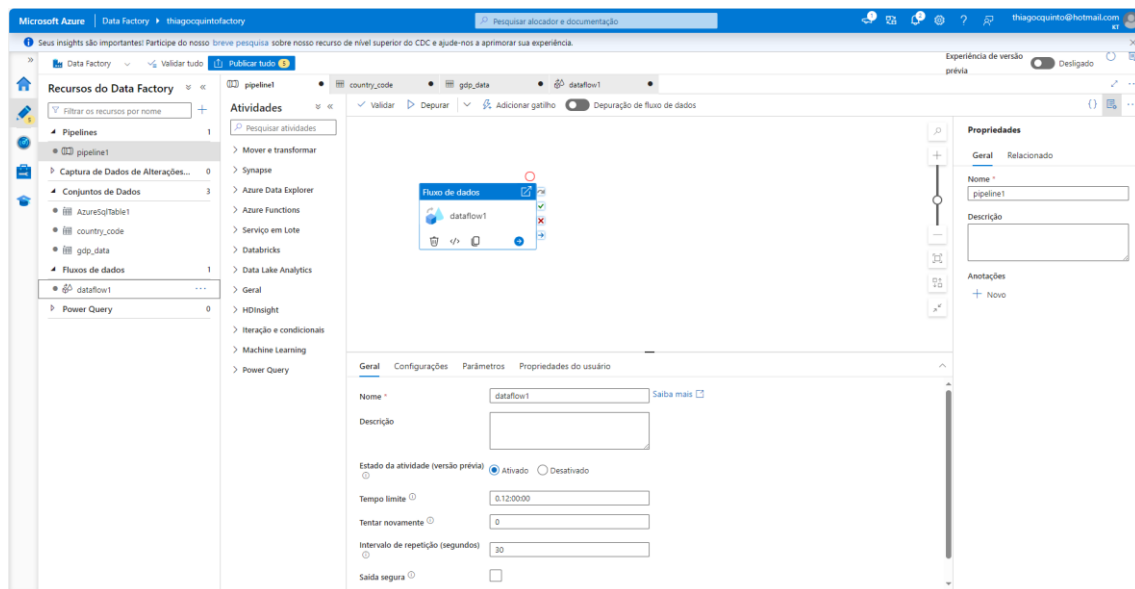
Cancelar

Testar conexão

Os dados são coletados para a Tabela `gdp_country` que estará vinculado ao Azure SQL Database.



Em seguida, foi feita a ligação do Dataflow ao Pipeline



E publicado todo pipeline

Microsoft Azure | Data Factory | thiagocquintofactory

Seus insights são importantes! Participe do nosso breve pesquisa sobre nosso recurso de nível superior do CDC e ajude-nos a aprimorar sua experiência.

Recursos do Data Factory

- Pipelines
- Conjuntos de Dados de Alterações...
- Conjuntos de Dados
- Fluxos de dados
- Power Query

Atividades

- Mover e transformar
- Synapse
- Azure Data Explorer
- Azure Functions
- Serviço em Lote
- Databricks
- Data Lake Analytics
- Gerar
- HDInsight
- Iteração e condicionais
- Machine Learning
- Power Query

Fluxo de dados

dataflow1

Propriedades

Nome: dataflow1

Descrição:

Estado da atividade (versão prévia): ☒ Ativo ☐ Desativado

Tempo limite: 0:12:00:00

Tentar novamente: 0

Intervalo de repetição (segundos): 30

Salida segura:

Publicar tudo

Você está prestes a publicar todas as alterações pendentes para o ambiente ativo. [Saiba mais](#)

Alterações pendentes (5)

NOME	ALTERAR	EXISTENTE
Pipelines		
pipeline1	(Novo)	-
Conjuntos de Dados		
country_code	(Novo)	-
gdp_data	(Novo)	-
AzureSqlTable1	(Novo)	-
Fluxos de dados		
dataflow1	(Novo)	-

Publicar Cancelar

Microsoft Azure | Data Factory | thiagocquintofactory

Seus insights são importantes! Participe do nosso breve pesquisa sobre nosso recurso de nível superior do CDC e ajude-nos a aprimorar sua experiência.

Recursos do Data Factory

- Pipelines
- Conjuntos de Dados de Alterações...
- Conjuntos de Dados
- Fluxos de dados
- Power Query

Atividades

- Mover e transformar
- Synapse
- Azure Data Explorer
- Azure Functions
- Serviço em Lote
- Databricks
- Data Lake Analytics
- Gerar
- HDInsight
- Iteração e condicionais
- Machine Learning
- Power Query

Fluxo de dados

dataflow1

Propriedades

Nome: dataflow1

Descrição:

Estado da atividade (versão prévia): ☒ Ativo ☐ Desativado

Tempo limite: 0:12:00:00

Tentar novamente: 0

Intervalo de repetição (segundos): 30

Salida segura:

Validar Depurar Depuração de fluxo de dados

Experiência de versão: ☒ Prévia ☐ Desativado

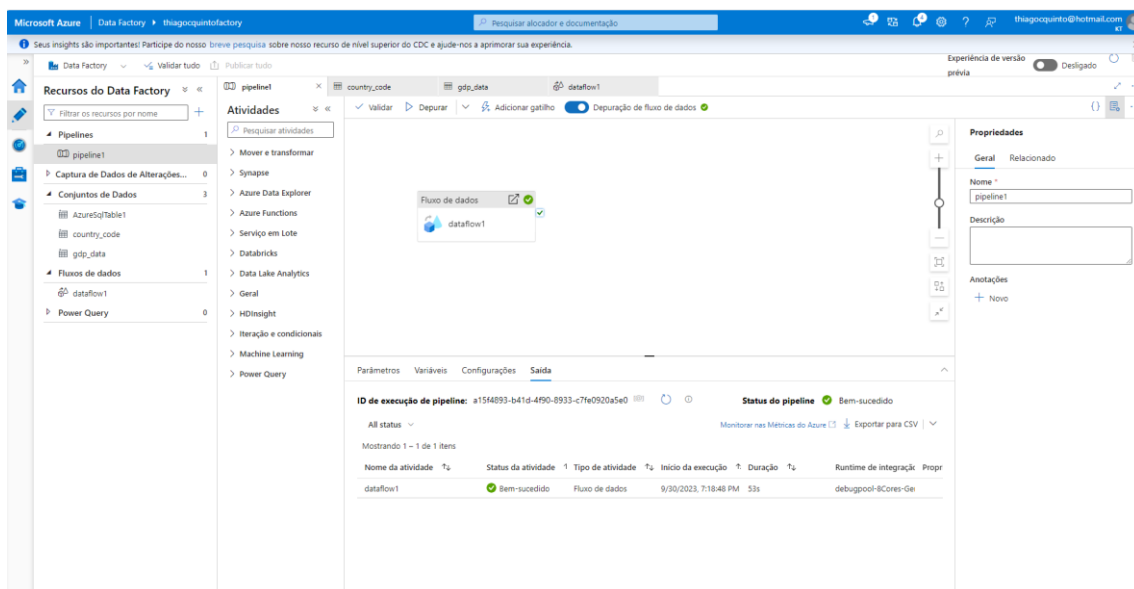
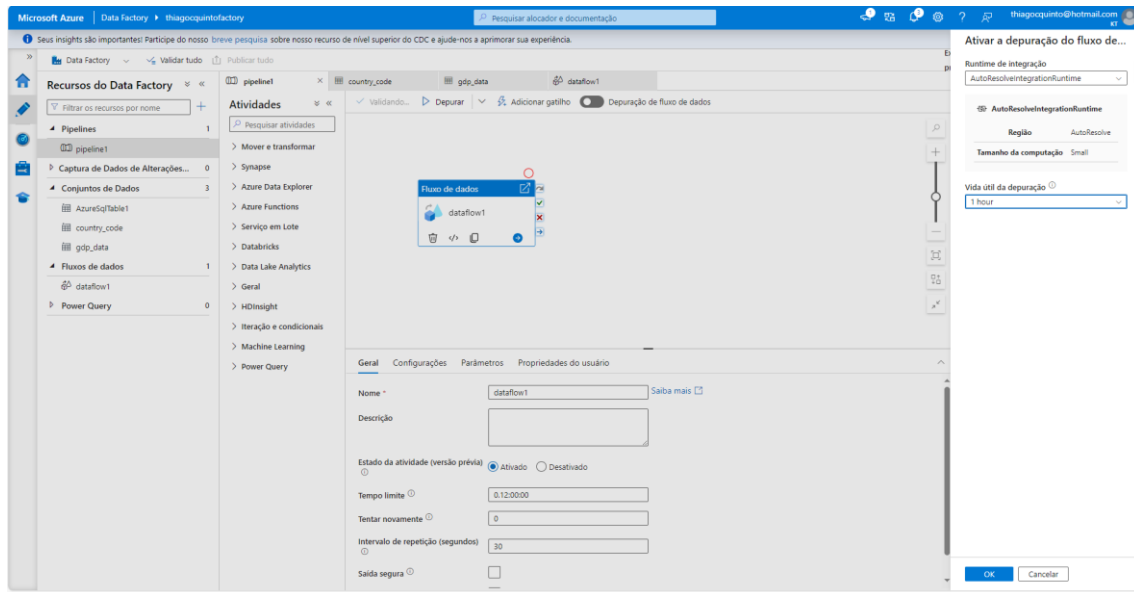
Validar tudo

Validação de pipeline

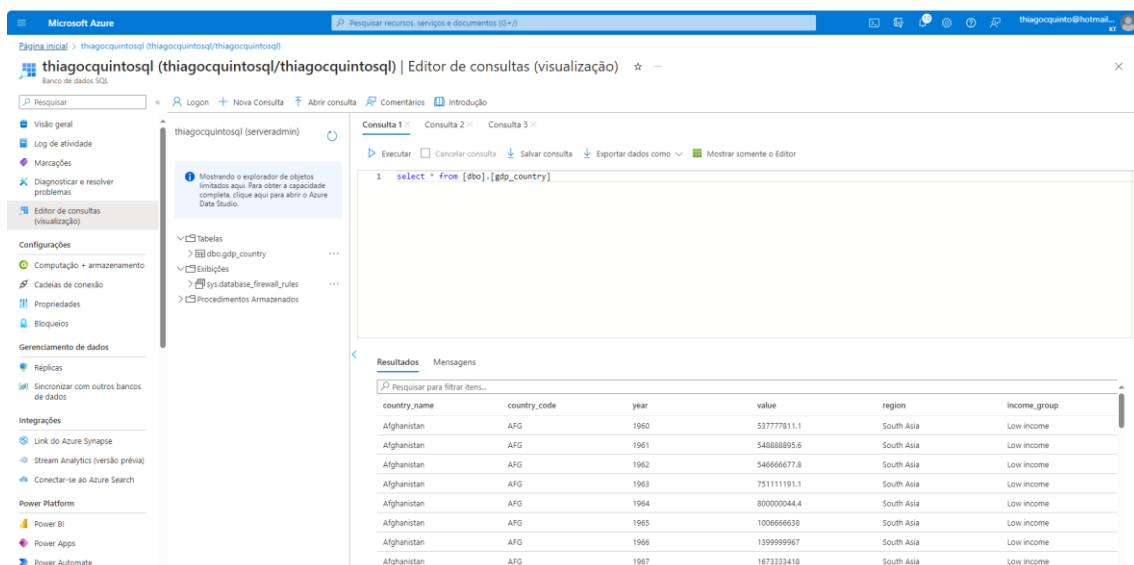
O pipeline foi validado. Nenhum erro foi encontrado.

Fechar

Validando o pipeline e depurando



A seguir foram feitos testes de execução de queries no SQL Database



Microsoft Azure

thiagocqintosoql (thiagocqintosoql/thiagocqintosoql) | Editor de consultas (visualização)

thiagocqintosoql (serveradmin)

Consultas

Consultar 1

Consultar 2

Consultar 3

Executar

Cancelar consulta

Salvar consulta

Exportar dados como

Mostrar somente o Editor

```
1 select distinct(region) from [dbo].[gdp_country]
```

Resultados

Mensagens

region

North America
Latin America & Caribbean
East Asia & Pacific
Europe & Central Asia
Middle East & North Africa
South Asia
Sub-Saharan Africa

Êxito na consulta | 0s

Microsoft Azure

thiagocqintosoql (thiagocqintosoql/thiagocqintosoql) | Editor de consultas (visualização)

thiagocqintosoql (serveradmin)

Consultas

Consultar 2

Executar

Cancelar consulta

Salvar consulta

Exportar dados como

Mostrar somente o Editor

```
1 select year,region,count(*) as quant_paises,sum(value) as total_income
2 from [dbo].[gdp_country]
3 group by region,year
4 having year=2020
5 order by sum(value) desc
```

Resultados

Mensagens

year	region	quant_paises	total_income
2020	East Asia & Pacific	36	2646478520277.1
2020	North America	3	22756867147000
2020	Europe & Central Asia	57	22149009427263
2020	Latin America & Caribbean	39	4552123889331.8
2020	South Asia	6	3488089161507
2020	Middle East & North Africa	20	3138086103377
2020	Sub-Saharan Africa	46	1707444708747.6

Êxito na consulta | 0s

Microsoft Azure

thiagocqintosoql (thiagocqintosoql/thiagocqintosoql) | Editor de consultas (visualização)

thiagocqintosoql (serveradmin)

Consultas

Consultar 1

Executar

Cancelar consulta

Salvar consulta

Exportar dados como

Mostrar somente o Editor

```
1 select distinct top 15 *
2 from [dbo].[gdp_country]
3 where year=2022
4 order by value desc
```

Resultados

Mensagens

country_name	country_code	year	value	region	income_group
United States	USA	2022	25500000000000	North America	High income
China	CHN	2022	18000000000000	East Asia & Pacific	Upper middle income
Japan	JPN	2022	42300000000000	East Asia & Pacific	High income
Germany	DEU	2022	40700000000000	Europe & Central Asia	High income
India	IND	2022	33900000000000	South Asia	Lower middle income
United Kingdom	GBR	2022	30700000000000	Europe & Central Asia	High income
France	FRA	2022	27800000000000	Europe & Central Asia	High income
Russian Federation	RUS	2022	22400000000000	Europe & Central Asia	Upper middle income

Êxito na consulta | 0s

A seguir, iniciou-se a criação do Catálogo de Dados no Microsoft Purview, adicionado este recurso.

Microsoft Azure

Página inicial > Criar um recurso > Marketplace > Microsoft Purview

### Criar conta Microsoft Purview

Forneça informações da conta Microsoft Purview

🔗 Lançamos o novo portal do Microsoft Purview. Você está criando uma conta Microsoft Purview em nível de locatário para obter segurança e governança de dados em toda a empresa. [Saiba mais sobre contas em nível de locatário](#)

\* **Notões básicas** | Configuração | Marcas | Examinar + Criar

Crie uma conta Microsoft Purview para desenvolver uma solução de governança de dados com apenas alguns cliques. Uma conta de armazenamento e um eventub serão criados em um grupo de recursos gerenciados em sua assinatura para cenários de ingestão de catálogo. [Saiba mais](#)

**Detalhes do projeto**

Assinatura \*

Grupo de recursos \*   
[Criar novo](#)

**Detalhes da instância**

Nome da conta Microsoft Purview \*

Localização \*

📘 1 unidade de capacidade (CU) = 25 operações/s e 10 GB de armazenamento de metadados. Qualquer nova conta do Microsoft Purview será provisionada com 1 CU com recursos de dimensionamento automático. [Saiba mais](#)

**Recursos gerenciados**

Um grupo de recursos, uma conta de armazenamento será criada na assinatura selecionada para cenários de ingestão de catálogo. Os provedores de recursos Microsoft.Storage e Microsoft.EventHub serão registrados. [Saiba mais](#)

Nome do grupo de recursos gerenciados

Nome da conta de armazenamento

O nome será gerado automaticamente durante a criação da conta.

[Examinar + Criar](#) [Anterior](#) [Próximo: Configuração >](#)

Microsoft Azure

Página inicial > **Microsoft.PurviewAccount-thiagocquintopurview-1696152635612** | Visão Geral

Implantação

[Excluir](#) [Cancelar](#) [Reimplantar](#) [Baixar](#) [Atualizar](#)

**Visão Geral**

Entradas

Saídas

Modelo

✅ **A implantação foi concluída**

Nome da implantação : Microsoft.PurviewAccount-thiagocquintopurview-1696152635612

Assinatura : Azure subscription 1

Grupo de recursos : MVP3

Hora de início : 01/10/2023, 06:30:46

ID de correlação : a1a4e67b-b35e-451b-a91c-14c88548d786

**Detalhes de implantação**

Recurso	Tipo	Status	Detalhes da operação
thiagocquintopurview	Conta do Microsoft Purview	OK	<a href="#">Detalhes da operação</a>

**Próximas etapas**

[Ir para o recurso](#)

Enviar comentários

[Conte-nos sobre sua experiência com a implantação](#)

## Configurando o purview

Microsoft 365

Experimente o novo Microsoft Purview

**Coleções**

Coleções são grupos de itens, como fontes de dados e ativos, que são organizados juntos no Mapa de Dados. [Saiba mais sobre coleções](#)

**thiagocquintopurview**

[+ Nova coleção](#) [Atualizar](#)

**Visão Geral** | **Atribuições de função** | Atualizado em 1 de outubro de 2023, 06:30 por Thiago Corrêa Quinto

[Editar atribuições de função](#)

**Coleção de administradores**

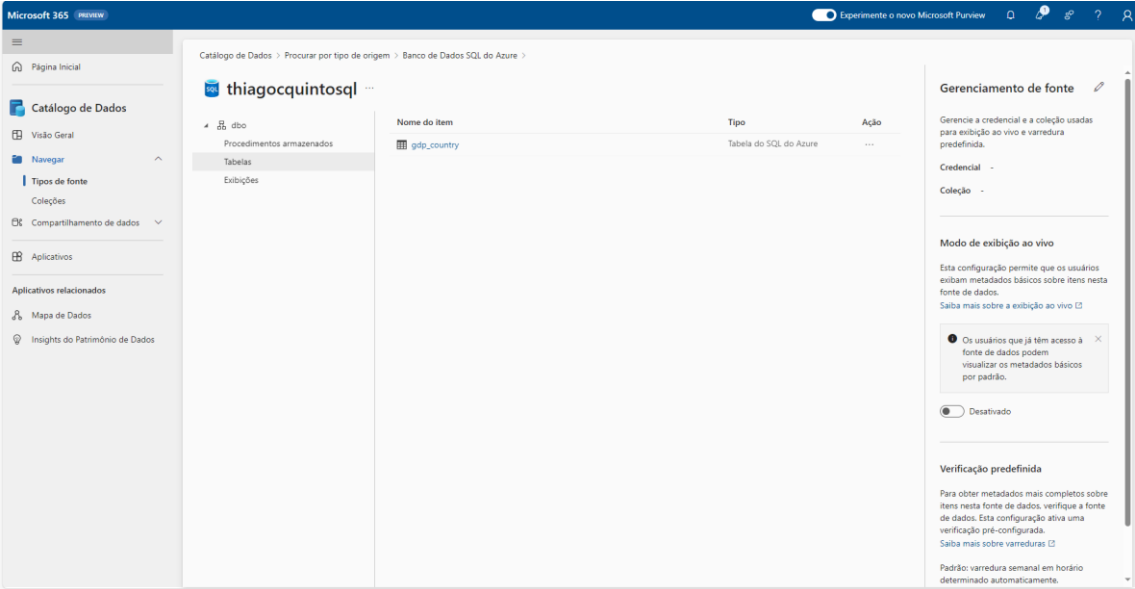
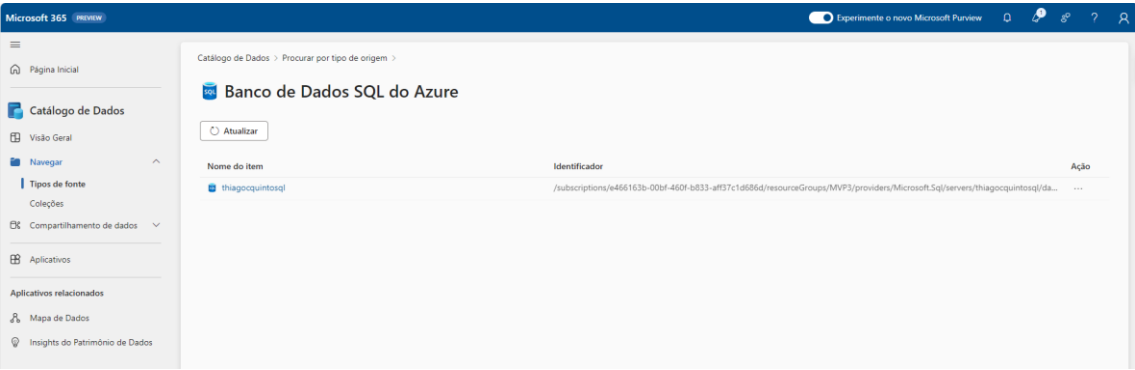
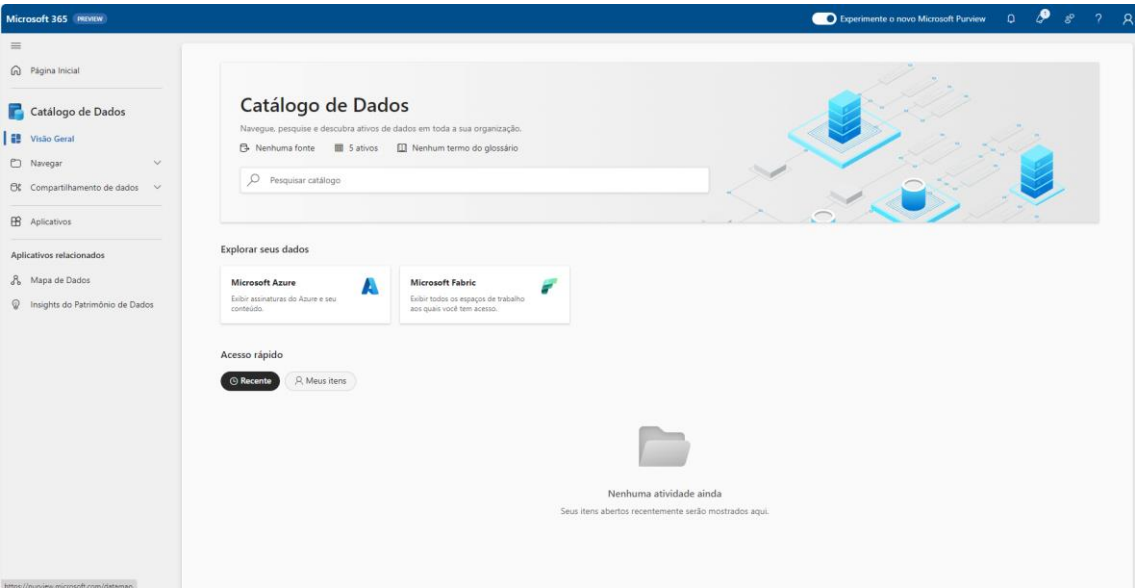
Os administradores da coleção podem editar a coleção, seus detalhes e adicionar subcoleções. Eles também podem adicionar curadores de dados, leitores de dados e outras funções do Microsoft Purview a um escopo de coleção. Os administradores de coleção herdados automaticamente de uma coleção pai não podem ser removidos.

Nome	Tipo
Administrador global Função do Microsoft Entra	Função do Mi...
Thiago Corrêa Quinto thiagocquinto@hotmail.com	Usuário

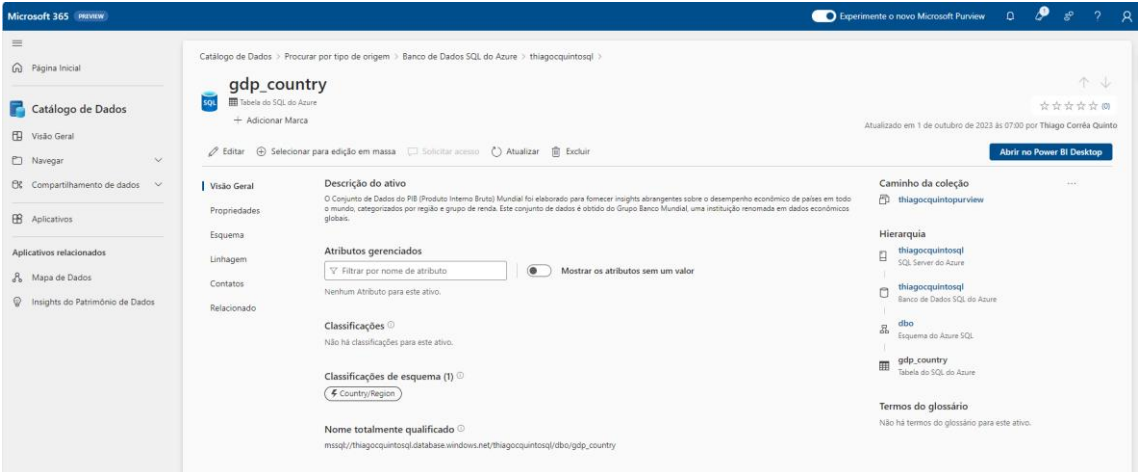
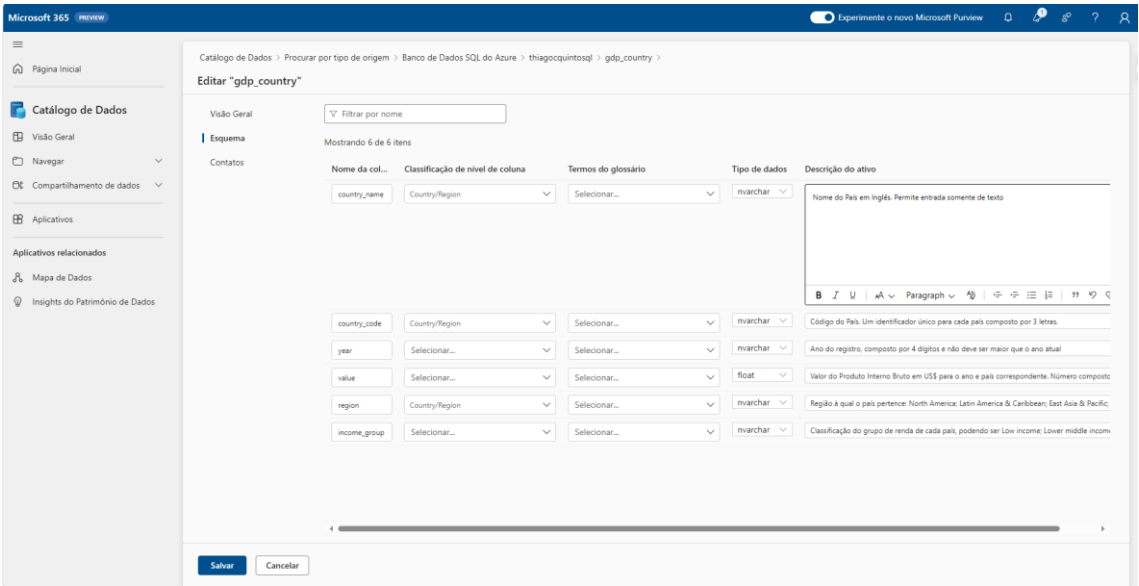
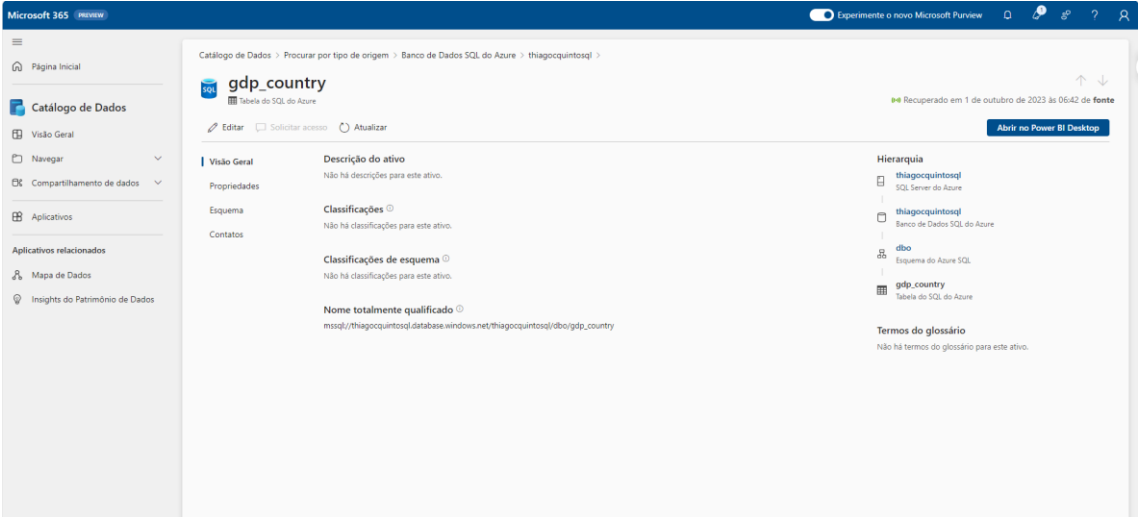
**Administradores das fontes de dados**

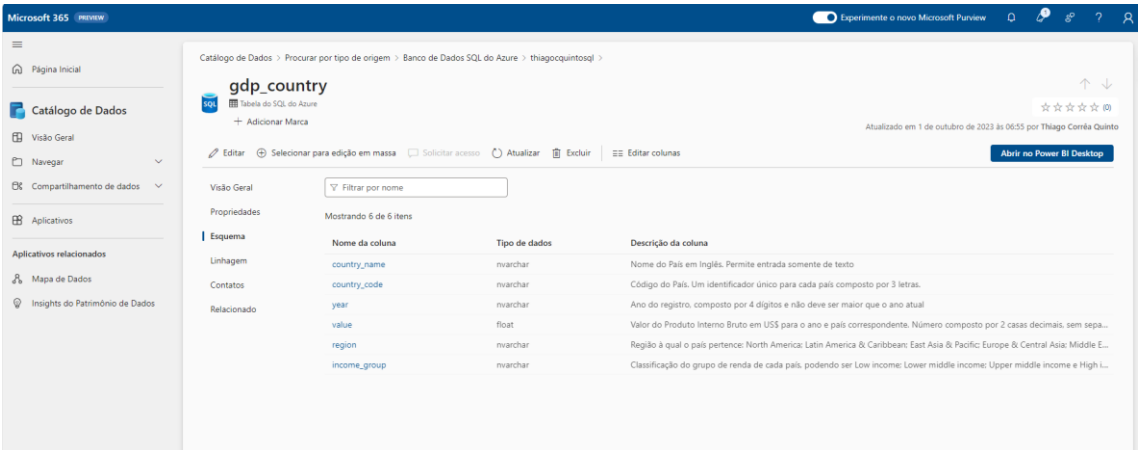
Os administradores da fonte de dados podem gerenciar fontes de dados e verificações de dados.

Nome	Tipo
Thiago Corrêa Quinto thiagocquinto@hotmail.com	Usuário





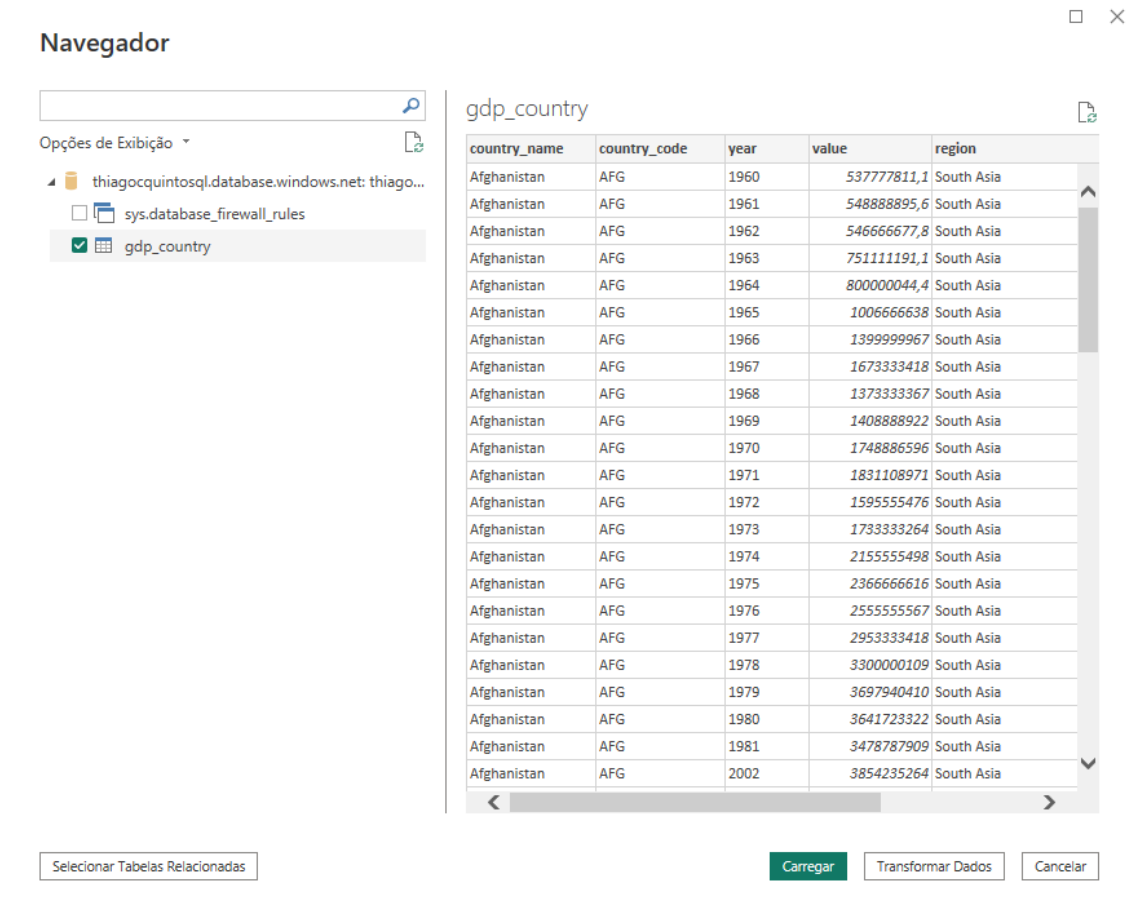




Após a conclusão destas etapas o conjunto de dados está pronto para ser consumido em qualquer serviço de análise de dados. Neste projeto foi escolhido o Microsoft Power BI para usufruir do ecossistema da Microsoft.

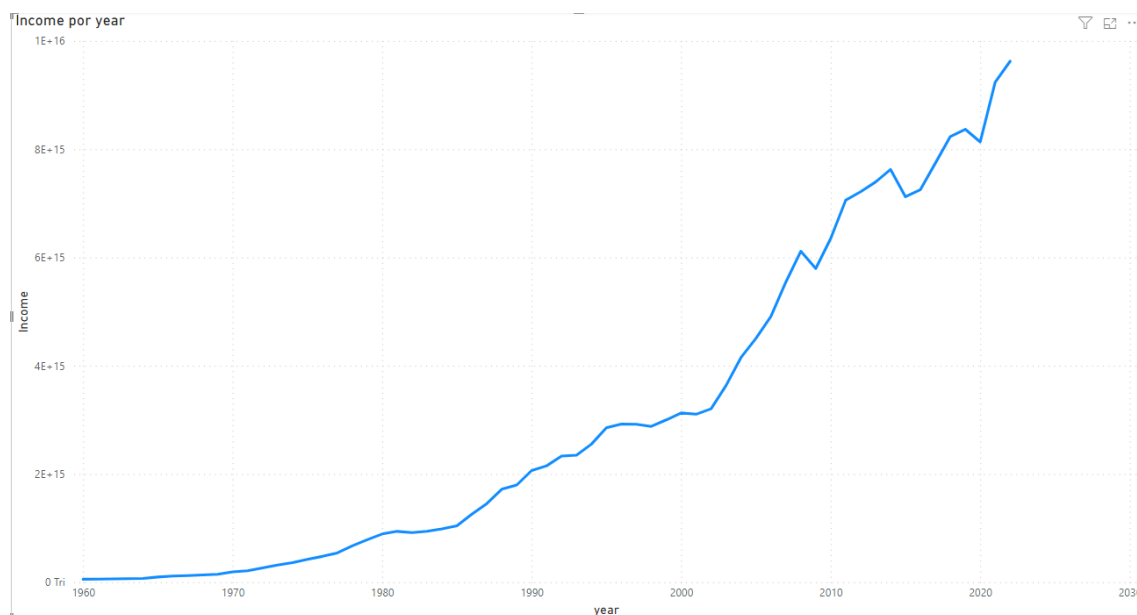
# Análise dos Dados

Foi feita a conexão dos dados no Power BI via Direct Query, conforme o screen abaixo:

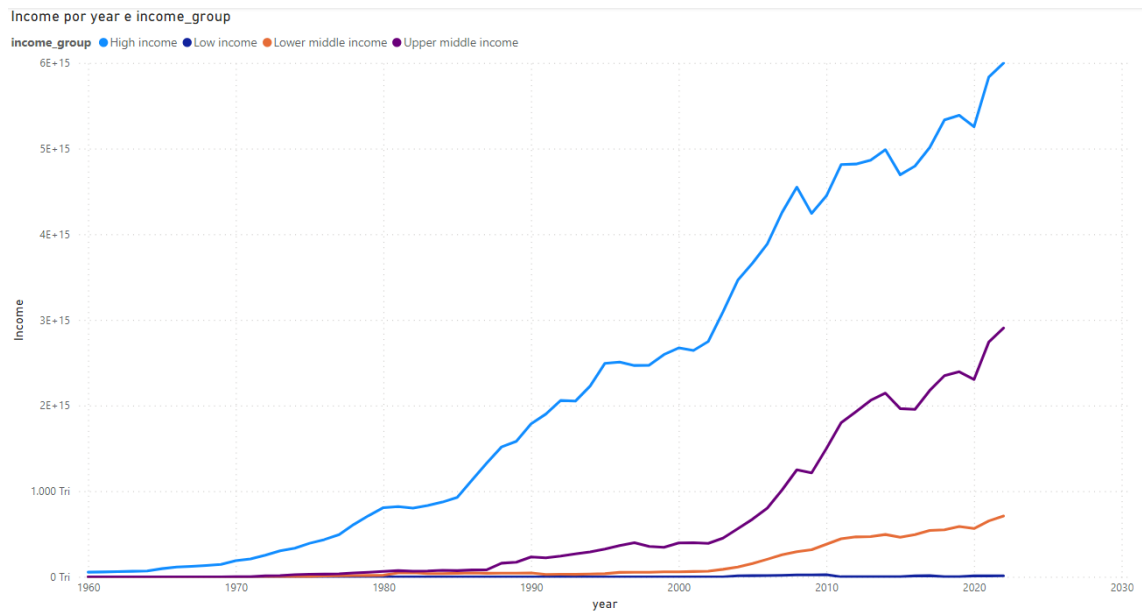
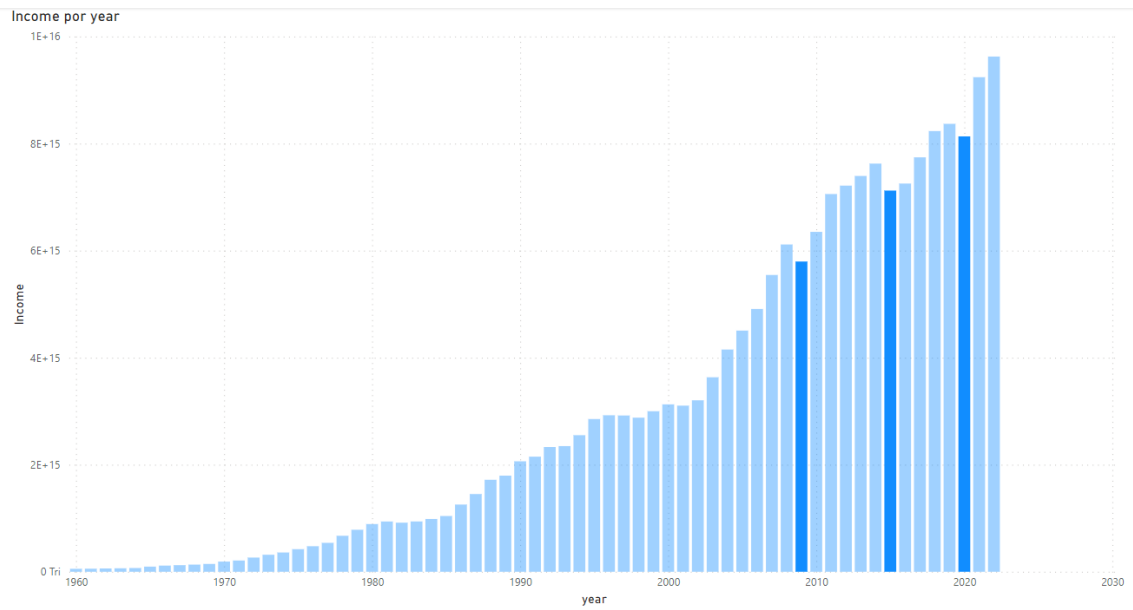


= Sql.Database("thiagocquintosql.database.windows.net", "thiagocquintosql")[[Schema = "dbo", Item = "gdp_country"]][Data]						
	A <sub>C</sub> country_name	A <sub>C</sub> country_code	A <sub>C</sub> year	1.2 value	A <sub>C</sub> region	A <sub>C</sub> income_group
1	Afghanistan	AFG	1960	537777811,1	South Asia	Low income
2	Afghanistan	AFG	1961	548888895,6	South Asia	Low income
3	Afghanistan	AFG	1962	546666677,8	South Asia	Low income
4	Afghanistan	AFG	1963	751111191,1	South Asia	Low income
5	Afghanistan	AFG	1964	800000044,4	South Asia	Low income
6	Afghanistan	AFG	1965	1006666638	South Asia	Low income
7	Afghanistan	AFG	1966	1399999967	South Asia	Low income
8	Afghanistan	AFG	1967	1673333418	South Asia	Low income
9	Afghanistan	AFG	1968	1373333367	South Asia	Low income
10	Afghanistan	AFG	1969	1408888922	South Asia	Low income
11	Afghanistan	AFG	1970	1748886596	South Asia	Low income
12	Afghanistan	AFG	1971	1831108971	South Asia	Low income
13	Afghanistan	AFG	1972	1595555476	South Asia	Low income
14	Afghanistan	AFG	1973	1733333264	South Asia	Low income
15	Afghanistan	AFG	1974	2155555498	South Asia	Low income
16	Afghanistan	AFG	1975	2366666616	South Asia	Low income

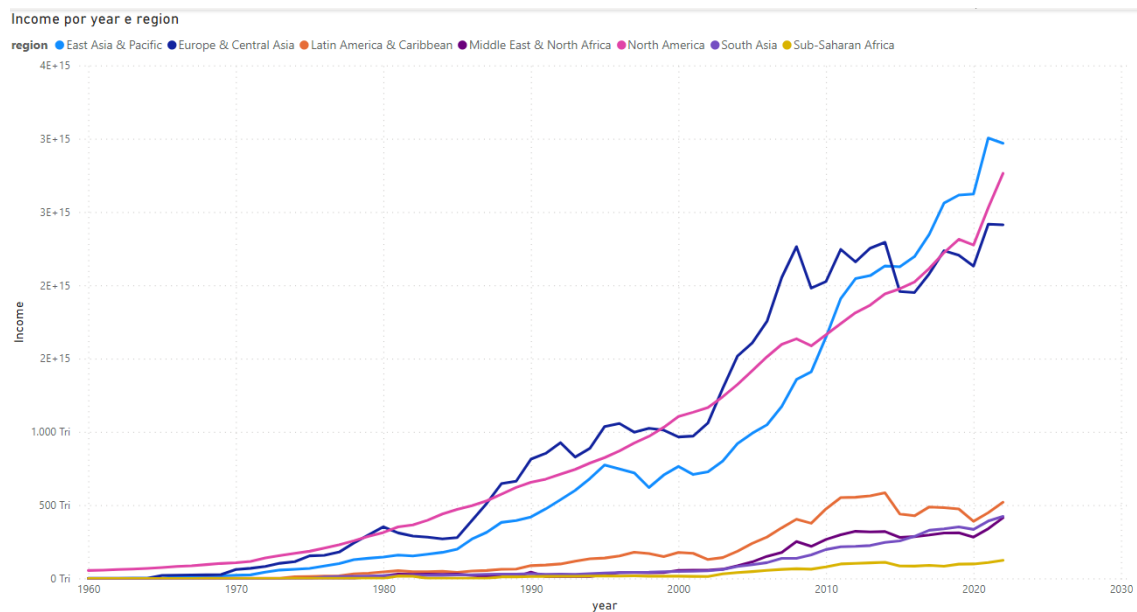
Como a principal parte da ETL foi realizada no Azure Data Factory, os dados não apresentaram nenhuma inconsistência bastando apenas carregar para ambiente do Power BI.



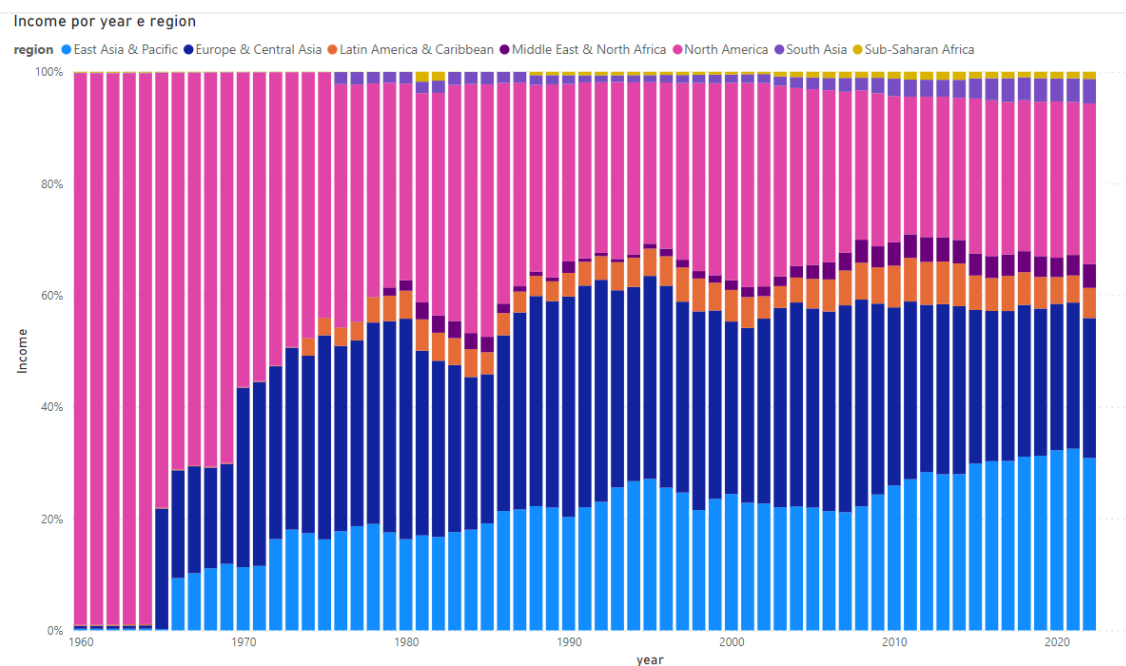
PIB por ano: Podemos perceber que desde o início da série histórica, o PIB Mundial vem aumentando exponencialmente, principalmente após o ano 2000. Destaque para 3 quedas do PIB Mundial, que ocorreram em 2009 (Crise Financeira Global e Recessão Global), 2015 (Crise da Dívida na zona do Euro e Baixos Preços do Petróleo) e 2020 (Pandemia COVID-19). Estes PIBs podem ser visualizados no próximo gráfico.



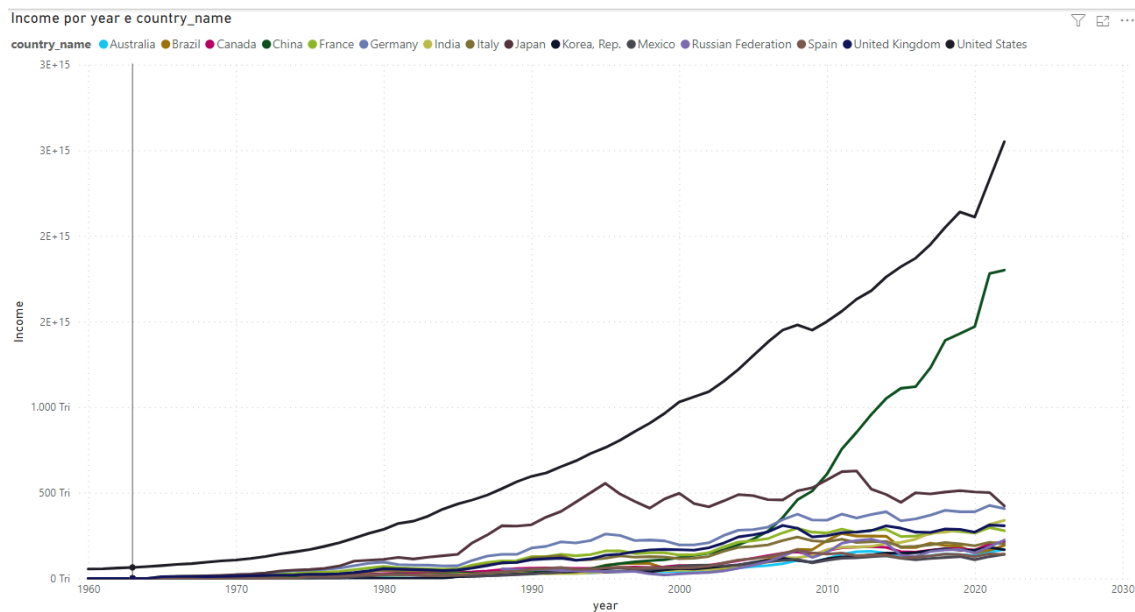
PIB por ano e por Grupo. Podemos perceber pelo gráfico que os Países com os Maiores PIBs foram os que mais cresceram na série histórica.



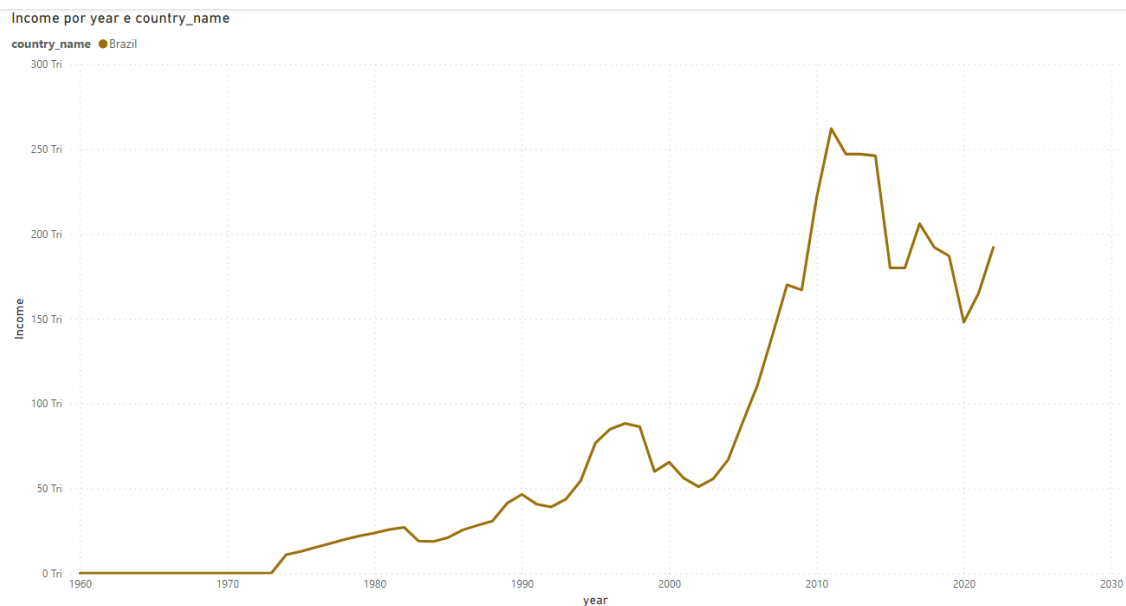
PIB por ano região: Podemos ver mais claramente que as Regiões da Ásia, Europe e América do Norte foram as que mais concentraram o aumento do PIB. A discrepância é elevada quando comparada com o restante do globo.



Participação do PIB por região e ano. Podemos Perceber que no início da série histórica a América do Norte tinha quase que toda a participação do PIB Mundial, em virtude do pós-guerra e provavelmente falta de dados no início da série. Esse cenário foi se alterando ao longo dos anos com o aumento da participação da Ásia e Europa.

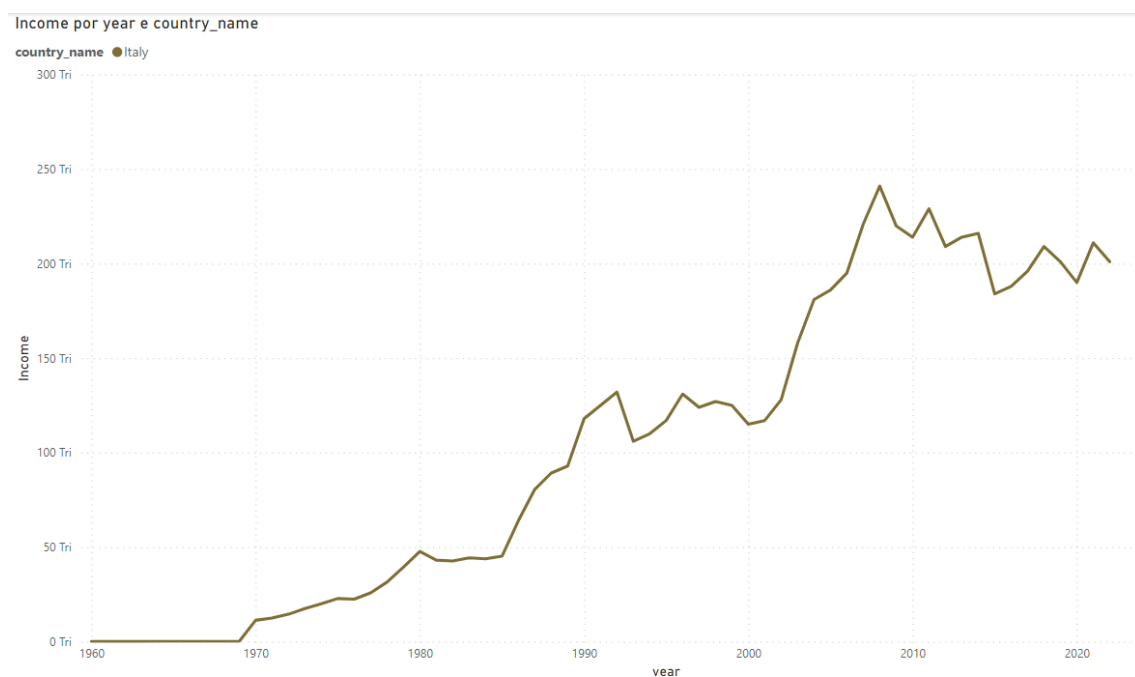
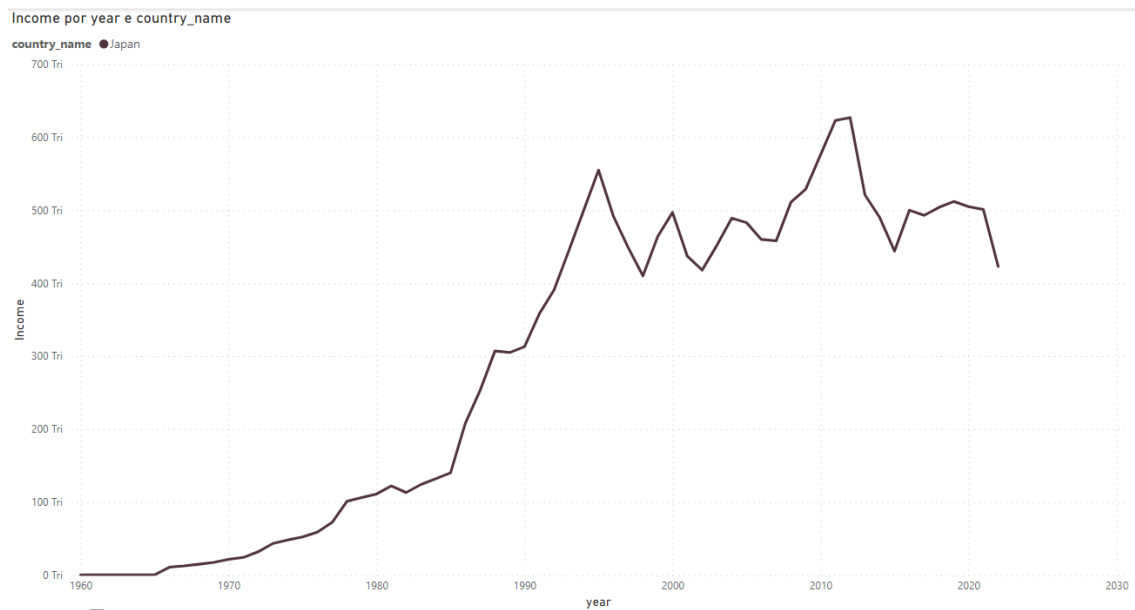


PIB por ano e País: Percebemos que a vanguarda do PIB mundial fica concentradas nos Estados Unidos e China. Destaque para duas informações: Os Estados Unidos sempre lideraram na série histórica e a China foi o país que apresentou o maior crescimento mundial.



Ainda no PIB por ano e país, destaque especial para o Brasil, cujo PIB teve um grande crescimento de 2000 a 2011, e em seguida uma tendência de queda.

Resultados similares ocorreram com a Itália e Japão destacados a seguir.



## Conclusão

A implementação de um Data Warehouse no Azure é desafiadora pois envolve vários componentes e considerações complexas. A utilização do serviço de Nuvem da Azure foi implementada com sucesso e com os dados obtidos foi possível fazer uma análise sobre o PIB Mundial. Praticamente todas as questões levantadas no início do projeto puderam ser respondidas com o conjunto de dados.

A implementação possibilita um automatismo na nuvem além de uma facilidade maior na integração de dados, permitindo sejam ingeridos dados de diversas fontes e formatos sem a necessidade de transformá-los imediatamente. Isso simplifica a ingestão de dados e a eliminação de problemas de incompatibilidade entre fontes. O Azure permitiu ainda a utilização de recursos avançados de governança e segurança para proteger dados confidenciais e garantir conformidade regulatória através do Catálogo de Dados. O data warehouse ainda

pode permitir que sejam processados dados em tempo real ou em lote, dependendo das necessidades da sua organização. Isso é importante para análises em tempo real e também para a análise histórica.