**MVP de Engenharia de Dados**

**Aluno: Bruno Soares Rodrigues da Silva**

Construção de um pipeline de dados utilizando tecnologias na nuvem

**1. Definição do problema**

O dataset objeto deste estudo refere-se a uma base de spend de uma empresa do segmento mineração, com resultados de 2018 a 2020. Entende-se como spend a relação de invoices (notas fiscais) cadastradas no ERP. A base lista as invoices de diversos países, entretanto para este exercício será premissado a análise do resultado do Brasil (que possui o spend mais representativo). O objetivo é analisar o spend das categorias de compras e a expectativa é de uma queda no resultado de 2020 devido ao impacto da pandemia de Covid-19. Estima-se que este impacto será maior nas categorias de serviço.

Contextualizando... a Covid-19 foi caracterizada como pandemia pela Organização Mundial da Saúde (OMS) em março de 2020. Muitas empresas impuseram o regime de trabalho remoto visando minimizar a propagação do vírus e, com isso, alguns contratos de fornecedores, especialmente de prestadores de serviços, tiveram que ser renegociados e/ou distratados.

Os atributos que compõem a base de dados e auxiliarão na análise estão descritos a seguir:

|  |  |
| --- | --- |
| ATRIBUTOS | DESCRIÇÃO |
| YEAR | Ano de registro do documento no ERP |
| MONTH | Mês de registro do documento no ERP |
| VENDOR\_CODE | ID do Fornecedor no ERP |
| VENDOR\_NAME | Nome do Fornecedor |
| CONTRACT | Número do contrato no ERP |
| PURCHASE\_ORDER | Número do Pedido de compra no ERP |
| CATEGORY\_CODE | Código do grupo de mercadoria no ERP, que segue a metodologia UNSPSC |
| CATEGORY\_NAME | Descrição do grupo de mercadoria |
| COMPANY\_CODE | Código da empresa demandante |
| COST\_TYPE | Código do tipo de custo, que definirá se trata-se de OPEX ou CAPEX |
| COUNTRY | País do Pedido de Compra |
| CURRENCY | Moeda do Pedido de compra |
| INCOTERM | Código da condição de frete |
| INVOICE\_NUMBER | Número da Invoice registrada no ERP |
| PAYMENT\_CONDITION | Código da condição de pagamento do Pedido de Compra |
| PLANT | Planta/Armazém do Pedido de Compra |
| PLANT\_REGION | Estado da Planta/Armazém |
| PURCHASING\_GROUP | Grupo de compra do gestor de categoria / comprador |
| PURCHASING\_ORGANIZATION | Organização de compras onde foi gerado o pedido de compras |
| QUANTITY | Quantidade total do Pedido de Compra |
| SPEND | Spend Total em USD |
| SUPPLIER\_COUNTRY | País do Fornecedor |
| UNIT\_OF\_MEASURE | Unidade de Medida do Pedido de compras |
| VENDOR\_REGION | Estado do Fornecedor |
| PO\_DOCUMENT\_TYPE | Código no ERP que define se é uma compra de contrato ou spot |
| COMMODITY\_GROUP | Definição da categoria de compra em Material ou Serviço |

**2. Carga e Análise dos Dados**

Utilizou-se os recursos da Azure para upload das bases e criação do banco de dados, conforme as etapas demonstradas a seguir:

2.1. Criação da conta de armazenamento

Foram criados dois contêineres: Na pasta *insumos* temos as bases com a descrição das categorias de compras e Nome dos fornecedores

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Na pasta *raw* temos as bases spend separadas por ano.

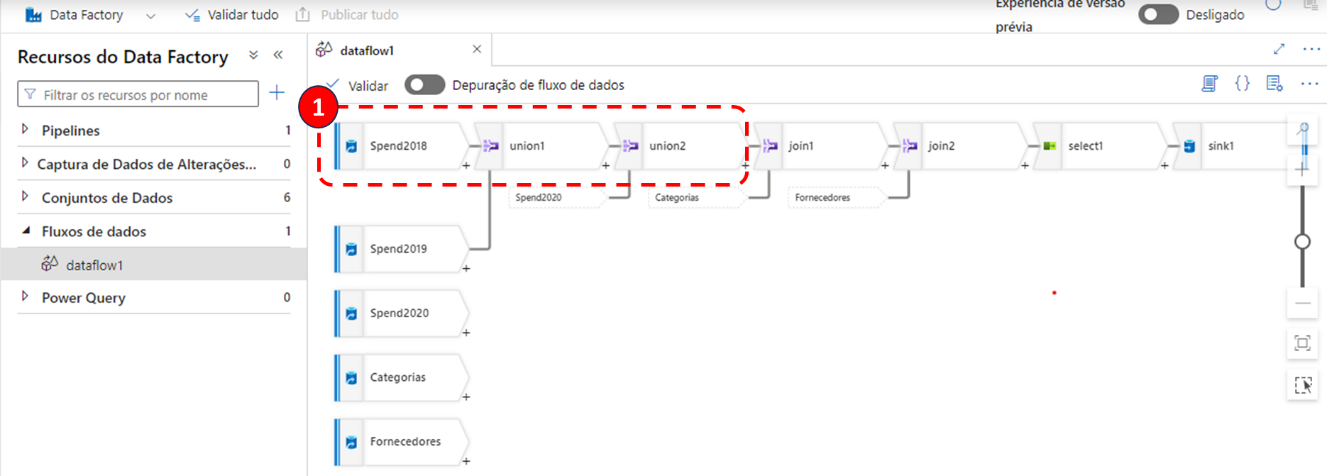
Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

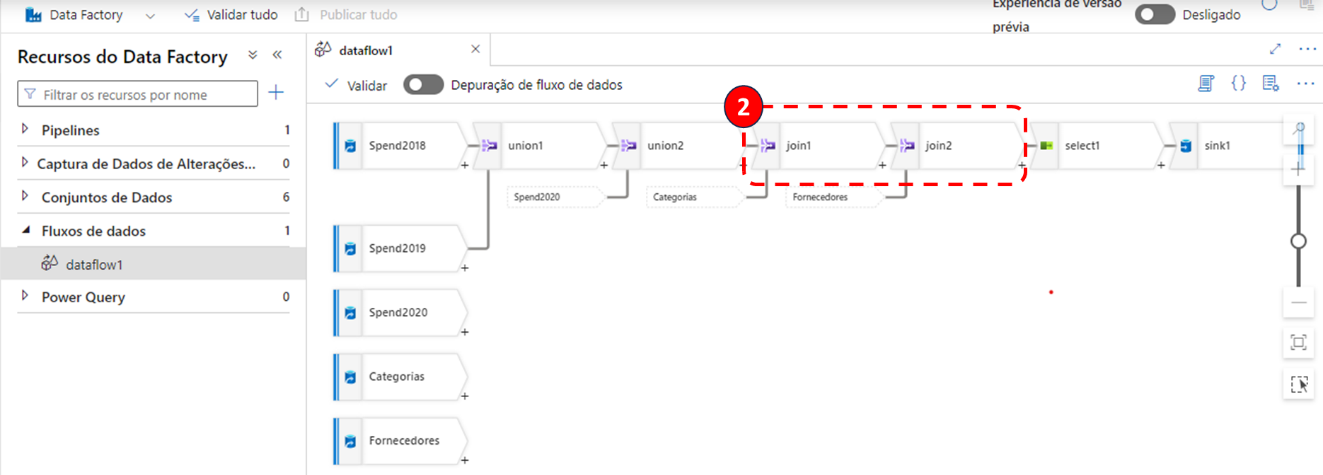
2.2. Criação do fluxo de dados e pipeline no Azure Data Factory

Utilizou-se o *dataflow* para alguns tratamentos, conforme descrito nas etapas a seguir:

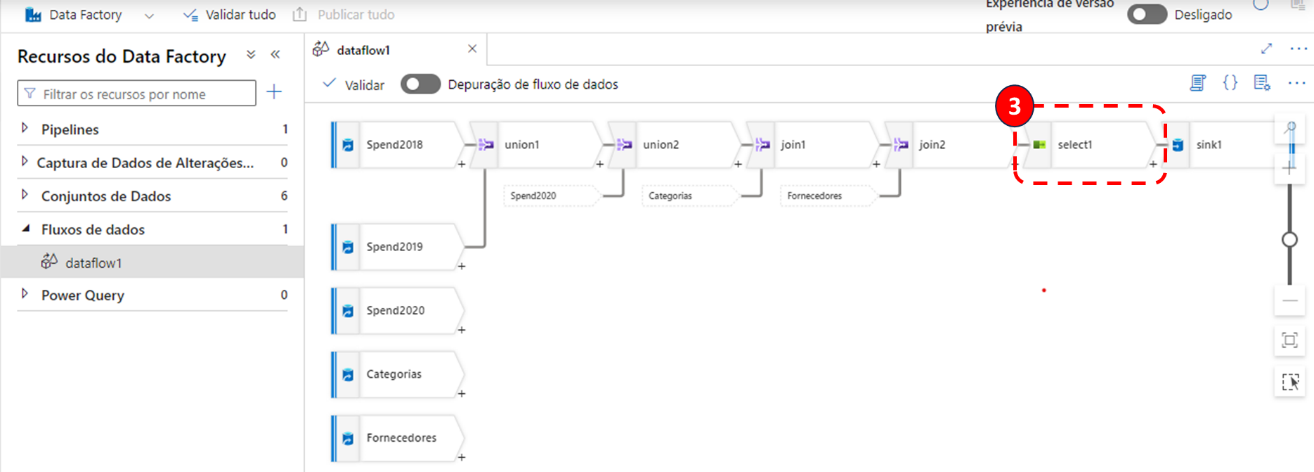
1. União dos datasets 2018. 2019 e 2020



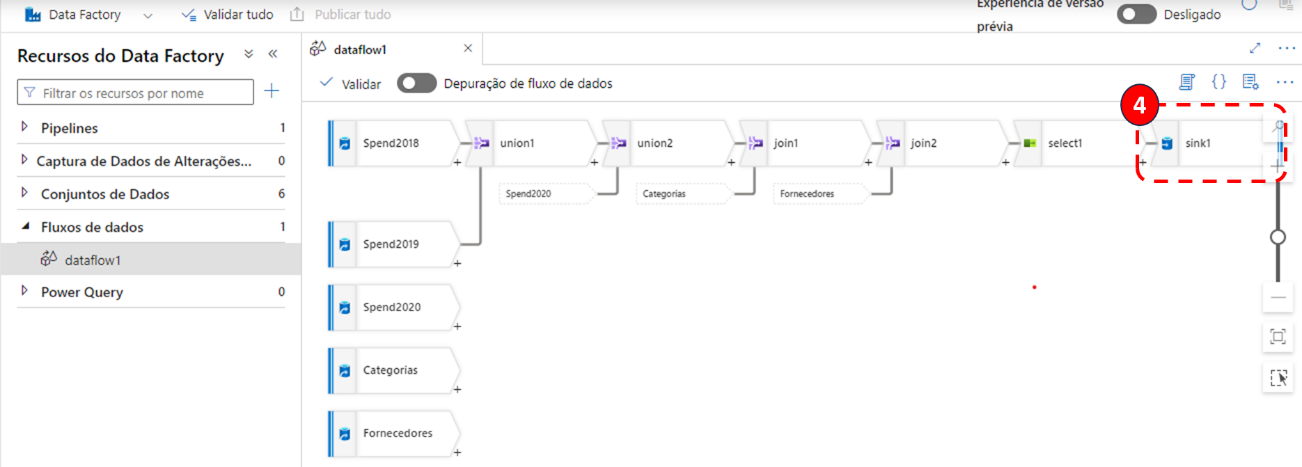
2. *Join* com as bases Categorias, para trazer a descrição e Fornecedores para trazer o nome dos mesmos



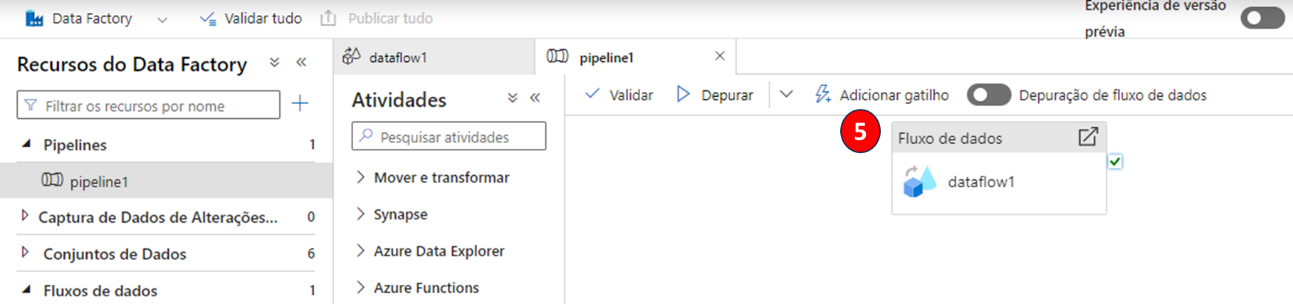
3. Exclusão de algumas colunas



4. Subir os dados para o SQL Database



5. Validação e depuração do fluxo de dados



2.3. Executando consultas no *database* através do Azure Data Studio.

Interface gráfica do usuário, Aplicativo, Tabela

Descrição gerada automaticamente

Iniciaremos a análise do spend do Brasil pelo número de Pedidos de Compra das categorias de material e serviço emitidas por ano:

Interface gráfica do usuário, Aplicativo, Teams

Descrição gerada automaticamente

Pode-se verificar um aumento do número de pedidos de compra de serviço e diminuição do número de pedidos de compra de material. Vamos fazer a mesma análise com valor do spend para verificar se isso se mantem.

Interface gráfica do usuário, Aplicativo, Teams

Descrição gerada automaticamente

Nesta consulta foi possível identificar que, ao contrário da expectativa inicial de impacto no spend das categorias de serviço, tivemos uma queda no spend das categorias de material em 2020 (de USD 4,5 Bi em 2019 para USD 4,0 Bi em 2020). Isso nos leva a entender que os contratos de serviço que foram distratados não tinham grande representativa no spend da categoria. Outro ponto seria uma renegociação da condição de homem/hora trabalhada.

E um dos cenários possíveis que justifiquem a queda do spend de material seria uma renegociação de prazo de entrega de material, considerando que as operações estavam operando em regime de escala para diminuição do efetivo e, consequentemente, minimizar os riscos de contaminação... ou até mesmo cancelamento de pedidos que não seriam essenciais.

2.4. Análise das categorias de compras através do PowerBI

Usamos o powerbi como recurso para visualizarmos o comportamento das top 20 categorias de material no período 2018 – 2020 e, desta forma, colhermos algum insight.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

O spend das categorias de material entre 2018 e 2019 teve variação entre as top 20, mas tudo dentro do esperado, com as categorias de combustível mantendo-se no topo.

Vamos analisar o resultado de 2020.

Uma imagem contendo Tabela

Descrição gerada automaticamente

Podemos observar uma grande queda do spend de Diesel, indo de USD 8,9Mi em 2019 para USD 6,1Mi em 2020... entretanto o que chamou a atenção foi o spend da categoria *Personal safety and protection* que teve um spend de USD 7Mi, sendo a 6° categoria de maior spend neste ano, considerando que nos anos anteriores a mesma não configurou no top 20. Isso faz total sentido, considerando o cenário de pandemia em 2020.

**3. Considerações Finais**

Este trabalho buscou analisar o spend das categorias de compras do Brasil no período de 2018 a 2020 e verificar um possível impacto da pandemia de Covid-19 no resultado de 2020. A expectativa era de uma diminuição do spend das categorias de serviço visto que, devido ao cenário de imposição do trabalho remoto nos escritórios para minimizar o risco de propagação do vírus, assim como o estabelecimento de turnos nas operações com a mesma finalidade... muitos contratos foram renegociados ou distratados. Entretanto o que se observou foi uma queda de USD 500Mi no spend das categorias de materiais. O diesel, que é a categoria mais representativa, teve uma queda de USD 2,8Mi de 2019 para 2020... e a categoria de Segurança e Proteção pessoal que, não figurava entre as 20 categorias de maior spend em 2018 e 2019, teve grande representatividade em 2020, com um spend de USD 7Mi.

Como próximos passos para este trabalho, fica a sugestão de análise das categorias nos anos pós pandemia para avaliação do comportamento... assim como estender este mesmo exercício para outros países, a fim de comparação.