

Semana do Desenvolvedor – Dia 2

Aula 2: Ingestão de Arquivos via S3, Rastreamento e Integração com o Fluxo Principal de Pedidos

Documentação do Laboratório: Arquitetura de Processamento de Pedidos e Arquivos na AWS

(Duração Estimada: 160 minutos)

1. Visão Geral e Objetivos da Aula

Na aula anterior, estabelecemos o fluxo de ingestão de pedidos via API. Nesta segunda aula, vamos implementar um canal alternativo para a entrada de pedidos: o processamento de arquivos JSON enviados para um bucket S3. Esta funcionalidade é crucial para cenários onde pedidos são recebidos em lote ou de sistemas externos que geram arquivos.

O objetivo é configurar um bucket S3. Quando um novo arquivo JSON for carregado, uma notificação do S3 enviará um evento para uma fila SQS Standard. Uma função Lambda será acionada por esta fila, fará o download do arquivo, validará seu conteúdo e, crucialmente, **se o arquivo contiver dados de pedidos válidos, esses pedidos serão transformados e enviados para a mesma Amazon SQS FIFO Pedidos que é utilizada pelo fluxo da API (criada na Aula 1)**. Além disso, o resultado da validação do *arquivo* será registrado em uma tabela DynamoDB e, em caso de erro na validação do *arquivo*, uma notificação será enviada via SNS.

Ao final desta aula, você terá:

- Um bucket S3 configurado para receber arquivos JSON de pedidos.
- Uma fila SQS Standard para desacoplar o processamento inicial dos arquivos.
- Uma função Lambda para validar arquivos JSON do S3, **extrair pedidos e enviá-los para o pipeline principal**.
- Uma tabela DynamoDB para rastrear o histórico de validação de *arquivos*.
- Um tópico SNS para notificações de erro na validação de *arquivos*.
- **A integração efetiva de uma segunda fonte de pedidos no sistema.**

Recursos a Serem Criados/Modificados Nesta Aula:

- IAM Role: lambda-s3-validation-role-seu-nome (criação e adição de permissões)
- Amazon S3 Bucket: datalake-arquivos-seu-nome
- Amazon SQS DLQ (Standard): s3-arquivos-json-dlq-seu-nome
- Amazon SQS Queue (Standard): s3-arquivos-json-queue-seu-nome
- AWS Lambda Function: validacao-s3-arquivos-lambda-seu-nome
- Amazon DynamoDB Table: controle-arquivos-historico-seu-nome
- Amazon SNS Topic: notificacao-erro-arquivos-seu-nome

Lembrete Importante: Substitua seu-nome em todos os nomes de recursos. Trabalhe consistentemente na mesma região AWS.

2. Configuração de Permissões (IAM Role)

Criaremos uma IAM Role para a Lambda que processará os arquivos do S3. Esta role precisará de permissões para ler do S3, ler da SQS (Arquivos JSON), escrever no DynamoDB (Controle Arquivo Histórico), publicar no SNS e, **importante, enviar mensagens para a Amazon SQS FIFO Pedidos (criada na Aula 1)**.

2.1. Criar Role para a Lambda de Validação de Arquivos S3 (lambda-s3-validation-role-seu-nome)

1. Acesse o serviço **IAM** no Console AWS.
2. No menu à esquerda, clique em **"Roles"** e depois em **"Create role"**.
3. **Trusted entity type:** Selecione **"AWS service"**.
4. **Use case:** Selecione **"Lambda"** e clique em **"Next"**.
5. Na página **"Add permissions"**:
 - Procure pela política **AWSLambdaBasicExecutionRole** e marque-a.
 - Clique em **"Next"**.
6. **Role details:**
 - **Role name:** lambda-s3-validation-role-seu-nome
 - **Description:** (Opcional) Ex: "Role para Lambda de validacao de arquivos S3 e envio para SQS FIFO Pedidos".

7. Clique em **"Create role"**.

- *Nota: Após criar os outros recursos, voltaremos a esta role para adicionar as permissões específicas.*

3. Configuração do Enfileiramento de Mensagens (Amazon SQS Standard - para arquivos)

Esta fila receberá notificações do S3 sobre novos arquivos.

3.1. Criar a Fila de Mensagens Mortas (DLQ) (s3-arquivos-json-dlq-seu-nome)

1. Acesse o serviço **Amazon SQS**.
2. Clique em **"Create queue"**.
3. **Type:** Selecione **"Standard"**.
4. **Name:** s3-arquivos-json-dlq-seu-nome.
5. Clique em **"Create queue"**.
6. **Anote o ARN** desta DLQ.

3.2. Criar a Fila Principal de Arquivos JSON (s3-arquivos-json-queue-seu-nome)

1. Ainda no **SQS**, clique em **Queues**, no menu lateral.
2. Clique em **"Create queue"**.
3. **Type:** Selecione **"Standard"**.
4. **Name:** s3-arquivos-json-queue-seu-nome.
5. Na seção **"Configuration"**:
 - **Visibility timeout:** Altere o valor padrão de 30 segundos para 70 Seconds.
 - **Nota:** Este valor é definido para ser maior que o timeout da função Lambda que processará mensagens desta fila, que configuraremos posteriormente para 60 segundos. Isso é importante para evitar o reprocessamento acidental de mensagens.)
6. Role para baixo até a seção **"Dead-letter queue (DLQ)"**:
 - Marque **"Enabled"**.

- Em "**Amazon Resource Name (ARN) of the dead-letter queue**" ;, cole o ARN da DLQ s3-arquivos-json-dlq-seu-nome (criada no Passo 3.1).
 - **Maximum receives:** 3.
7. Mantenha as demais configurações padrão e clique em "**Create queue**".
 8. **Anote o ARN e a URL** desta fila, pois você precisará deles futuramente.

4. Configuração do Armazenamento de Histórico de Arquivos (Amazon DynamoDB)

Esta tabela rastreará o processamento de cada *arquivo*.

4.1. Criar a Tabela DynamoDB (controle-arquivos-historico-seu-nome)

1. Busque pelo serviço do **Amazon DynamoDB** na barra de pesquisa.
2. No menu à esquerda, clique em **Tables**.
3. Clique em "**Create table**".
4. **Table name:** controle-arquivos-historico-seu-nome
5. **Partition key:** nomeArquivo (String).
6. Clique em "**Create table**".
7. Em **Tables**, clique na tabela que você acabou de criar.
8. Na seção "**General information**", **anote o ARN** desta tabela no seu bloco de notas.

5. Configuração do Sistema de Notificação de Erros de Arquivo (Amazon SNS)

Notificações para erros no processamento do *arquivo* em si.

5.1. Criar o Tópico SNS (notificacao-erro-arquivos-seu-nome)

1. Busque pelo serviço do **Amazon SNS** na barra de pesquisa.
2. No menu à esquerda, clique em "**Topics**".
3. Clique em "**Create topic**".
4. **Type:** Selecione "**Standard**".
5. **Name:** notificacao-erro-arquivos-seu-nome.

6. Clique em **"Create topic"**.
7. **Anote o ARN** deste tópico no seu bloco de notas.

5.2. Criar uma Subscrição (Ex: E-mail)

1. Selecione o tópico criado, vá em **"Subscriptions"** > **"Create subscription"**.
2. **Protocol:** Selecione **"Email"**.
3. **Endpoint:** Insira seu endereço de e-mail (que você tenha acesso imediato).
4. Clique em **"Create subscription"**.
5. **Acesse o seu Email e confirme o e-mail** que você receberá da AWS:

Subscription confirmed!

You have successfully subscribed.

Your subscription's id is:

arn:aws:sns:us-east-1:237482015336:notificacao-erro-arquivos-seu-nome:09c1736f-2a61-43b4-86f7-b61f872c26dc

If it was not your intention to subscribe, [click here to unsubscribe](#).

6. Implementação da Lógica de Validação e Extração de Pedidos (AWS Lambda)

Esta é a Lambda central desta aula. Ela será acionada por mensagens da s3-arquivos-json-queue-seu-nome.

6.1. Criar a Função Lambda (validacao-s3-arquivos-lambda-seu-nome)

1. Busque pelo serviço do **AWS Lambda** na barra de pesquisa.
2. Acesse o serviço **AWS Lambda**.
3. Clique em **"Create function"**
4. Selecione **"Author from scratch"**.
5. **Function name:** validacao-s3-arquivos-lambda-seu-nome.
6. **Runtime:** Python 3.12.
7. Expanda o menu **Change default execution role**.
8. Selecione **Use an existing role**.
9. Em **Existing role** selecione a role: lambda-s3-validation-role-seu-nome.
10. Clique em **"Create function"**.

6.2. Configurar o Código e Variáveis de Ambiente

1. Na aba "**Code**", substitua `lambda_function.py` pelo código em python que você encontrará [clikando aqui](#).
2. Clique em "**Deploy**".
3. Vá para "**Configuration**" > "**Environment variables**" > "**Edit**".
4. Adicione as seguintes variáveis. Clique em "**Add environment variable**":
 - **Key 1:** DYNAMODB_TABLE_NAME
 - **Value 1:** controle-arquivos-historico-seu-nome

Atenção: Se você não anotou o nome exato da sua tabela no DynamoDB, volte lá e confirme antes de prosseguir. Se o nome estiver diferente, essa validação pode falhar.

5. Clique novamente em "**Add environment variable**":
 - **Key 2:** SNS_TOPIC_ARN
 - **Value 2:** Cole o ARN do seu tópico SNS notificacao-erro-arquivos-seu-nome (você anotou anteriormente no bloco de notas).
6. Clique novamente em "**Add environment variable**":
 - **Key 3:** SQS_FIFO_PEDIDOS_URL
 - **Value 3:** Cole a **URL** da fila pedidos-fifo-queue-seu-nome.fifo (criada na Aula 1).
7. Clique em "**Save**".
8. Vá para "**Configuration**" > "**General configuration**" > "**Edit**".
 - a. Aumente o **Timeout** para 1 minuto (ou mais, dependendo do tamanho/complexidade dos arquivos):

Timeout

| | | | |
|--------------------------------|-----|--------------------------------|-----|
| <input type="text" value="1"/> | min | <input type="text" value="0"/> | sec |
|--------------------------------|-----|--------------------------------|-----|

- b. Clique em "**Save**".

6.3. Passo Complementar: Adicionar Permissão Inicial de SQS à Role da Lambda

Antes de adicionarmos o trigger SQS à nossa função Lambda `validacao-s3-arquivos-lambda-seu-nome`, precisamos garantir que sua IAM Role (`lambda-s3-validation-role-seu-nome`) tenha as permissões mínimas para interagir com a fila `s3-arquivos-json-queue-seu-nome`. Sem isso, a criação do trigger falhará.

1. Acesse o serviço **IAM** no Console AWS.

2. No menu à esquerda, clique em **"Roles"** e encontre a role **lambda-s3-validation-role-seu-nome** (criada no Passo 2.1). Clique nela.
3. Na aba **"Permissions"**, sob "Permissions policies", clique em **"Add permissions"** e selecione **"Create inline policy"**.
4. Selecione a aba **JSON**.
5. Apague o esboço existente.
6. Cole o seguinte JSON, substituindo:D
 - **COLE AQUI O ARN DA SUA FILA SQS s3-arquivos-json-queue-seu-nome** Pelo ARN da sua Fila SQS s3-arquivos-json-queue, **criada no passo 3.2**.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "sqs:ReceiveMessage",
        "sqs:DeleteMessage",
        "sqs:GetQueueAttributes"
      ],
      "Resource": "COLE AQUI O ARN DA SUA FILA SQS s3-arquivos-
json-queue-seu-nome"
    }
  ]
}
```

7. **Next**
8. **Name:** SQSReadFromArquivosJsonQueueTrigger-seu-nome (um nome que indique que é para o trigger).
9. **Clique em** "Create policy".

Com esta permissão em vigor, podemos prosseguir para adicionar o trigger.

6.4. Adicionar o Trigger SQS à Lambda

1. Busque pelo serviço do **AWS Lambda** na barra de pesquisa.
 2. Acesse o serviço **AWS Lambda**.
 3. **Selecione a Função Lambda** “validacao-s3-arquivos-lambda-seu-nome” criada no passo 6.1.
 4. Role para baixo.
10. Vá para "**Configuration**" > "**Triggers**" > "**Add trigger**".
1. Fonte: **SQS**.
 2. SQS queue: s3-arquivos-json-queue-seu-nome.
 3. Batch size: 1.
 4. Clique em "**Add**".
-

7. Configuração do "Data Lake" (Amazon S3) e Notificações

7.1. Criar o Bucket S3 (datalake-arquivos-seu-nome)

1. Busque pelo serviço do **Amazon S3** na barra de pesquisa.
2. Acesse o serviço **Amazon S3**.
3. Clique em "**Create function**".
4. **Bucket name:** datalake-arquivos-seu-nome (deve ser globalmente único).
5. Mantenha "**Block all public access**" marcado.
6. Clique em "**Create bucket**".
7. **Anote o ARN e o nome do Bucket.**

7.2. Configurar Notificação de Eventos do S3 para SQS

1. Navegue até seu bucket > "**Properties**" > "**Event notifications**" > "**Create event notification**".
2. **Event name:** s3-json-upload-to-sqs-seu-nome.
3. **Suffix:** .json.
4. **Event types:** "All object create events".

5. **Destination:**

- **"SQS queue".**
- SQS queue: s3-arquivos-json-queue-seu-nome.

6. Clique em **"Save changes"**.

7.3. Atenção para uma possível erro aqui!

Caso você se depare com este erro **"Unable to validate the following destination configurations"** ao tentar configurar a notificação de eventos do S3 para SQS geralmente indica um problema de permissão **do serviço S3 para a fila SQS**.

Quando você configura o S3 para enviar eventos para uma fila SQS, o S3 precisa de permissão para executar a ação `sqs:SendMessage` *nessa fila específica*. Normalmente, o console do S3 tenta adicionar automaticamente uma política à fila SQS para permitir isso. No entanto, às vezes esse processo automático pode falhar ou encontrar problemas.

Causas Comuns e Soluções:

1. **Política da Fila SQS Restritiva (Principal Causa):**

- Se a sua fila SQS `s3-arquivos-json-queue-seu-nome` já tiver uma política de acesso (resource-based policy) que seja muito restritiva, ela pode impedir que o console do S3 (ou o serviço S3) adicione a permissão necessária.
- **Solução:** Verifique a política da sua fila SQS.
 1. Vá para o serviço **SQS**.
 2. Selecione a fila `s3-arquivos-json-queue-seu-nome`.
 3. Clique na aba **"Queue policies"** e na seção **"Access policy"**, clique em **"Edit"**.
 4. Desça até **Access policy**.
 5. **Ação Recomendada:** Remova a política existente e cole a política abaixo, modificando as seguintes informações:

```
{
  "Version": "2012-10-17",
  "Id": "PolicyForS3ToSQS",
  "Statement": [
    {
      "Sid": "AllowS3ToSendMessage",
      "Effect": "Allow",
      "Principal": {
        "Service": "s3.amazonaws.com"
      },
      "Action": "SQS:SendMessage",
      "Resource": "COLE AQUI O ARN DA SUA FILA s3-arquivos-json-queue-seu-nome",
      "Condition": {
        "ArnLike": {
          "aws:SourceArn": "COLE AQUI O ARN DO SEU BUCKET datalake-arquivos-seu-nome"
        },
        "StringEquals": {
          "aws:SourceAccount": "COLE AQUI O ID DA SUA CONTA"
        }
      }
    }
  ]
}
```

6. Clique em **Save**.

Volte ao passo **7.2. Configurar Notificação de Eventos do S3 para SQS**. Desta vez, você conseguirá finalizar o processo.

8. Atualizar Permissões da Role da Lambda de Validação S3

Conceder à lambda-s3-validation-role-seu-nome as permissões específicas.

1. Volte ao **IAM** > **"Roles"** > lambda-s3-validation-role-seu-nome.
2. Na aba **"Permissions"** > **"Add permissions"** > **"Create inline policy"**.
3. Selecione a aba **JSON**.
4. Cole o seguinte JSON (substitua REGION, ACCOUNT_ID e nomes dos recursos):

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "s3:GetObject",
      "Resource": "arn:aws:s3:::datalake-arquivos-seu-nome/*"
    },
    {
      "Effect": "Allow",
      "Action": "dynamodb:PutItem",
      "Resource": "COLE AQUI O ARN DA SUA table/controle-arquivos-historico-seu-nome"
    },
    {
      "Effect": "Allow",
      "Action": "sns:Publish",
      "Resource": "COLE AQUI O ARN DA SUA SNS notificacao-erro-arquivos-seu-nome"
    },
    {
      "Effect": "Allow",
      "Action": "sqs:SendMessage",
      "Resource": "COLE AQUI O ARN DA SUA FILA pedidos-fifo-queue-seu-nome.fifo"
    }
  ]
}
```

5. **Next.**

6. Em "**Review policy**" > **Name:**

S3ReadSQSPProcessDynamoSNSSQSFIFOPermissions-seu-nome.

7. Clique em "**Create policy**".

9. Teste do Fluxo de Ingestão de Arquivos S3 e Integração

9.1. Preparar Arquivos de Teste

1. Baixe o arquivo: [arquivo_com_pedidos.json](#)

○ Abra e edite o arquivo colocando o seu nome:

- S3P001-seu-nome
- S3P002-seu-nome
- Salve.

2. Baixe o arquivo: [arquivo_schema_invalido.json](#)

9.2. Testar com arquivo_com_pedidos.json

1. Faça upload de arquivo_com_pedidos.json para o bucket S3.

2. **Verificações:**

- **s3-arquivos-json-queue-seu-nome:** Mensagem deve passar rapidamente.
- **Logs da validacao-s3-arquivos-lambda-seu-nome:**
 - Verifique o processamento do arquivo.
 - Verifique as mensagens "Pedido ... enviado para SQS FIFO Pedidos." para S3P001 e S3P002.
 - Verifique a mensagem de erro para o pedido inválido S3P003.
- **DynamoDB (controle-arquivos-historico-seu-nome):** Deve haver um item para arquivo_com_pedidos.json com statusValidacao: "ARQUIVO_VALIDADO".
- **pedidos-fifo-queue-seu-nome.fifo (Fila da Aula 1):**
 - Vá para esta fila no console SQS > "Send and receive messages" > "Poll for messages".
 - Você deverá ver **duas novas mensagens**, correspondentes a S3P001 e S3P002. Examine seus corpos.

- **Importante:** Se essas mensagens não estiverem mais visíveis, isso indica que já foram processadas pela função validacao-pedidos-lambda-seu-nome (Lambda da Aula 1), que publica eventos no EventBridge. Verifique também os logs dessa função no CloudWatch para confirmar o processamento dos pedidos originados do S3.
- **Logs da validacao-pedidos-lambda-seu-nome:** Verifique os logs no CloudWatch para esta Lambda. Você deve ver:
 1. "Evento SQS recebido".
 2. "Processando pedido do SQS".
 3. "Pedido X validado com sucesso".
 4. "Evento publicado no EventBridge".
- **SNS/E-mail:** Nenhuma notificação de erro de *arquivo* deve ser enviada.

9.3. Testar com arquivo_schema_invalido.json

1. Faça upload de arquivo_schema_invalido.json para o bucket S3.
2. **Verificações:**
 - **Logs da validacao-s3-arquivos-lambda-seu-nome:** Verifique o erro de schema do arquivo.
 1. **OBS:** A partir de agora, você já sabe onde encontrar essas informações. Isso faz parte do seu crescimento como profissional. É importante estar atento e aprender a identificar, com poucas palavras, onde buscar o que precisa. Eu confio no seu potencial, mostre do que é capaz e vamos juntos nessa jornada.
 - **DynamoDB:** Item para arquivo_schema_invalido.json com statusValidacao: "ERRO_VALIDACAO_ARQUIVO". **Caminho:** DynamoDB > Explore items > controle-arquivos-historico-seu-nome

Table: controle-arquivos-historico-seu-nome - Items returned (2)

Scan started on May 10, 2025, 11:33:11

< 1 > ⚙

| <input type="checkbox"/> | nomeArquivo (String) ▾ | statusValidacao ▾ | time... |
|--------------------------|--|------------------------|---------|
| <input type="checkbox"/> | arquivo_com_pedidos.json | ARQUIVO_VALIDADO | 2025 |
| <input type="checkbox"/> | arquivo_schema_invalido... | ERRO_VALIDACAO_ARQUIVO | 2025 |

- **SNS/E-mail:** Você deve receber uma notificação de erro de *arquivo*.
- **pedidos-fifo-queue-seu-nome.fifo:** Nenhuma nova mensagem de pedido deve ser enviada para esta fila.

Conclusão da Aula 2

Parabéns! Você não apenas implementou um pipeline de ingestão de arquivos, mas também o **integrou com sucesso ao fluxo principal de processamento de pedidos**. Isso demonstra um padrão poderoso para lidar com múltiplas fontes de dados. Você aprendeu a:

- Configurar S3 para disparar eventos para SQS.
- Desenvolver uma Lambda para processar arquivos, validar seu conteúdo, **extrair dados relevantes (pedidos), transformá-los e enviá-los para outra fila SQS (FIFO)**.
- Rastrear o processamento de *arquivos* em DynamoDB e notificar erros de *arquivo* via SNS.
- Gerenciar permissões IAM para interações complexas entre serviços.

Na próxima aula, focaremos no processamento central dos pedidos que agora chegam ao EventBridge (originados tanto da API quanto do S3) e em sua persistência.

Prepare-se para a Aula 3!