

Documentação Operacional: Processo de Extração de Dados

1. Introdução

Este documento serve como um guia para a execução do processo de extração de dados. O objetivo é buscar informações de um sistema de origem e salvá-las em arquivos CSV locais. Esses arquivos serão a matéria-prima para as análises subsequentes.

O processo é conduzido por um script inteligente (`query_executor.py`) que funciona como um "robô". Nós o instruímos sobre quais dados buscar através de um arquivo de configuração simples, o `queryConfig.csv`.

2. O Que Você Precisa Para Começar

Antes de iniciar, garanta que seu ambiente esteja preparado com os seguintes itens:

- **Arquivos Essenciais:** Na mesma pasta, você deve ter os seguintes arquivos:
 - `query_executor.py`: O script principal que executa a extração.
 - `queryConfig.csv`: O "cardápio" de extrações, onde você escolhe o que buscar.
 - `private.pem`: Sua chave de segurança para acessar o sistema.
 - `.env`: Um arquivo de texto simples para armazenar as credenciais de acesso de forma segura.
- **Instalação Inicial (Apenas uma vez):** Se for a primeira vez que você executa este processo, é necessário instalar alguns assistentes para o Python. Abra o terminal de comando do seu computador e execute o comando abaixo:

None

```
pip install requests pyjwt cryptography python-dotenv tqdm
```

3. Configurando o Acesso: O Arquivo `.env`

O arquivo `.env` funciona como um cofre para as suas credenciais. É aqui que o script busca as informações de login sem que você precise editar o código principal.

Abra o arquivo `.env` com um editor de texto e preencha as seguintes informações:

Variável	Descrição	Exemplo
----------	-----------	---------

SF_LOGIN_URL	A URL de login do ambiente (Ex: Produção ou Sandbox).	https://login.salesforce.com
SF_CLIENT_ID	O identificador único da aplicação conectada.	3MVG9...
SF_USERNAME	O nome de usuário para a autenticação.	seu.email@empresa.com
SF_PRIVATE_KEY_FILE	O nome exato do arquivo da sua chave privada.	private.pem

Importante: Verifique se as informações estão corretas, pois um erro aqui impedirá o acesso do script ao sistema.

4. O Coração do Processo: O Arquivo **queryConfig.csv**

Este arquivo é o seu painel de controle. É aqui que você decide **quais dados serão extraídos**. Cada linha representa uma tarefa de extração, com duas colunas principais.

- **queryType:** Define o **método** de extração.
- **query:** A **instrução** exata do que buscar.

Como Ativar e Desativar Extrações

A forma mais simples de controlar o que será executado é usando o caractere **#**.

- **Para desativar uma extração:** Coloque um **#** no início da linha. O script irá ignorá-la.
- **Para reativar uma extração:** Simplesmente remova o **#** do início da linha.

Exemplo:

Code snippet

None

- **# Esta linha está desativada e não será executada**
- **#Tooling,"SELECT Id, CreatedById, CreatedDate FROM MktDataModelObject --DataModelObjectsDetails"**
-
- **# Esta linha está ativa e será executada**
- **Standard,"SELECT Id, Name FROM User --Users"**

Entendendo a Estrutura da Consulta (**query**)

Cada instrução na coluna **query** tem uma estrutura lógica:

SELECT [Colunas] FROM [Objeto] --[NomeDoArquivoFinal]

- **SELECT Id, Name:** Informa quais colunas (campos) você deseja extrair do objeto.
- **FROM User:** Indica de qual objeto (ou "tabela") os dados devem ser retirados.
- **--Users: (Muito Importante)** Este comentário no final define o nome do arquivo que será gerado. No exemplo, o resultado será salvo como **Users.csv**.

Os Tipos de Extração (queryType)

Cada tipo de extração tem uma finalidade específica. Pense neles como diferentes ferramentas para diferentes trabalhos:

Tipo	Quando Usar	Analogia
Bulk	Para extrair um volume muito grande de dados (milhões de registros).	É como encomendar um contêiner de produtos. Demora um pouco para preparar, mas é o método mais eficiente para grandes volumes.
Standard	Para consultas de rotina com volume de dados de pequeno a médio.	É como pedir um item diretamente no balcão da loja. Rápido e direto para a maioria das necessidades.
Tooling	Para buscar metadados, ou seja, informações sobre a estrutura dos dados (configurações, objetos, etc.).	É como pedir a planta da loja em vez dos produtos. Útil para entender como as coisas estão organizadas.
ssot	Para extrair dados de fontes complexas e aninhadas que os outros métodos não alcançam.	É como pedir um relatório customizado que busca informações de diferentes departamentos e as organiza para você.
localSQL e localParse	Para processar arquivos que já foram extraídos. Elas não se conectam ao sistema; em vez disso, leem um CSV da pasta dataExtract e extraem novas informações dele.	É como pegar um relatório já pronto e usar um marca-texto para destacar e reorganizar apenas as partes mais importantes.

5. Executando a Extração

Com tudo configurado, o processo de execução é muito simples:

1. **Abra o Terminal:** Inicie o terminal de comando (Prompt de Comando, PowerShell ou Terminal).
2. **Navegue até a Pasta:** Use o comando **cd** para navegar até o diretório onde você salvou os arquivos do script.
 - Exemplo: **cd C:\Users\SeuUsuario\Documentos\ExtracaoSalesforce**
3. **Execute o Script:** Digite o comando abaixo e pressione Enter.

None

```
python query_executor.py
```

Agora, o script começará a trabalhar. Ele exibirá o progresso na tela, mostrando qual consulta está executando e quantos registros já processou.

Atenção: Dependendo do volume de dados e do tipo de consulta (especialmente **Bulk**), o processo pode levar de alguns minutos a várias horas. Você pode acompanhar o progresso diretamente na janela do terminal.

6. Verificando os Resultados

Após a conclusão do script, uma nova pasta chamada **dataExtract** será criada no mesmo local.

Dentro desta pasta, você encontrará todos os arquivos **.csv** gerados, cada um nomeado de acordo com o que foi definido no comentário **--NomeDoArquivo** no arquivo **queryConfig.csv**.

Ocasionalmente, você também poderá ver arquivos com a extensão **.skipped**. Estes arquivos são registros de itens que, por algum motivo específico (como um erro de API temporário), não puderam ser processados. Eles são úteis para auditoria, mas na maioria dos casos podem ser ignorados.

7. Perguntas Frequentes (FAQ)

- **Apareceu um erro de "Authentication error". O que fazer?**
 - Este erro quase sempre está relacionado às credenciais. Verifique se os dados no seu arquivo **.env** estão corretos e se o arquivo **private.pem** está na mesma pasta.
- **O script executou, mas um dos arquivos que eu esperava não foi criado.**
 - Verifique o arquivo **queryConfig.csv** e certifique-se de que a linha correspondente à extração **não** está comentada com um **#** no início.
- **Vi uma mensagem de "Token expired" na tela. É um problema?**
 - Não se preocupe. Isso é normal. O token de acesso tem uma validade. O script foi projetado para detectar isso e solicitar um novo token automaticamente, continuando o processo de onde parou.
- **Posso executar apenas uma extração específica em vez de todas?**
 - Sim! Essa é a principal vantagem do **queryConfig.csv**. Comente todas as linhas adicionando um **#** no início, exceto a linha da extração que você deseja realizar.