

Desafio AeC Automação

Descrição

Esse documento visa dar um overview sobre o projeto de **web scraping do site Alura**, na parte de cursos.

Foi utilizado o tipo de projeto Console por se tratar de um projeto mais rápido de ser implementado, requerendo menos codificação.

Arquitetura

A solução apresentada segue padrões de arquitetura DDD, com as seguintes camadas (projetos):

- **Console** – Onde reside o Program Main e de onde parte as chamadas dos serviços de scraping. **Faz referência aos projetos: Common e Domain**
- **Domain** – Onde temos as definições dos contratos/interfaces a serem implementadas pelas demais camadas, e onde se tem a única entidade “**Course**” que vai representa a unidade de um documento/linha/registro associado aos dados dos cursos recuperados do site Alura. **Não faz referência a nenhum projeto.**
- **Service** – Onde temos as implementações das operações crud a serem realizadas no modelo de armazenamento nosql do projeto. **Faz referência ao projeto Domain.**
- **Common** – Onde temos a parte responsável pela **injeção de depedência** do projeto. **Faz referência aos projetos Domain e Service.**

O projeto faz o armazenamento de dados em um banco **NoSql MongoDB**, rodando em um container Docker. **Para fins de teste, o banco foi levantado sem usuário e senha.**

O projeto acompanha um arquivo ***docker-compose.yml*** para ser executado antes dos teste, para que o container do banco seja levantado.

Vale observar que os testes foram realizados localmente numa máquina rodando Windows 11, com Docker for Windows.

Desafio AeC Automação

Decisões técnicas

Para o processo de scraping no site Alura, foi criada uma classe específica para o serviço, deixando toda a lógica dentro dela. Com isso, outras classes poderão ser anexadas com outros scrapings, deixando a lógica de cada serviço isolado e de fácil manutenção.

Foi usado um modelo **nosql de banco para facilitar os testes** e por ser um banco mais rápido de ser usado e para os fins desse teste.

Foi baixando do **Docker Hub a versão mais atual da imagem do banco**, assim como o **package NuGet para uso do MongoDB**, foi a mais recente.

Nas camadas **Domain e Service**, buscou-se aplicar o conceito de **classes genéricas** para implementar as operações de banco, facilitando assim a chamada declarativa das operações, diminuindo acoplamentos, e facilitando a manutenibilidade do código, e claro, fazendo uso da injeção de dependência, para menor acoplamento.

Dentro da solution, se encontra anexado o **binário do navegador Chrome**, que é acionado durante a execução do projeto.

Para poder realizar os testes do sistema, primeiro abra um terminal na pasta onde se encontra o projeto Console, e execute o comando ***docker-compose up -d***, para que seja baixada a imagem do MongoDB e levantado o container.

Sugere-se que utilize uma IDE gráfica que facilite a visualização dos dados capturados. Sugiro o **MongoDB Compass**.

Desafio AeC Automação

Lógica

O sistema faz a captura dos seguintes dados:

- **Título do Curso**
- **Descrição do Curso**
- **Link do Curso**
- **Nomes dos Professores**
- **Carga Horária**

Esses dados estão representados na classe **Course** localizada no projeto **Domain**.

Primeiro passo é percorrer todas as páginas de curso do site, recuperando o título, a descrição e o link do curso.

Num segundo passo, por meio do link salvo em cada curso, é acessada a página dos cursos e recuperado o nome dos professores e a carga horária. São ao todo **820 cursos** até a data de 31/07/2024.

Foram aplicados try/catches em pontos críticos do projeto para garantir que o mesmo rode sem quebras.

Um ponto fundamental para execução do projeto, caso venha a ser feito via linha de código, é que o mesmo **requer a passagem de um parâmetro de pesquisa**.

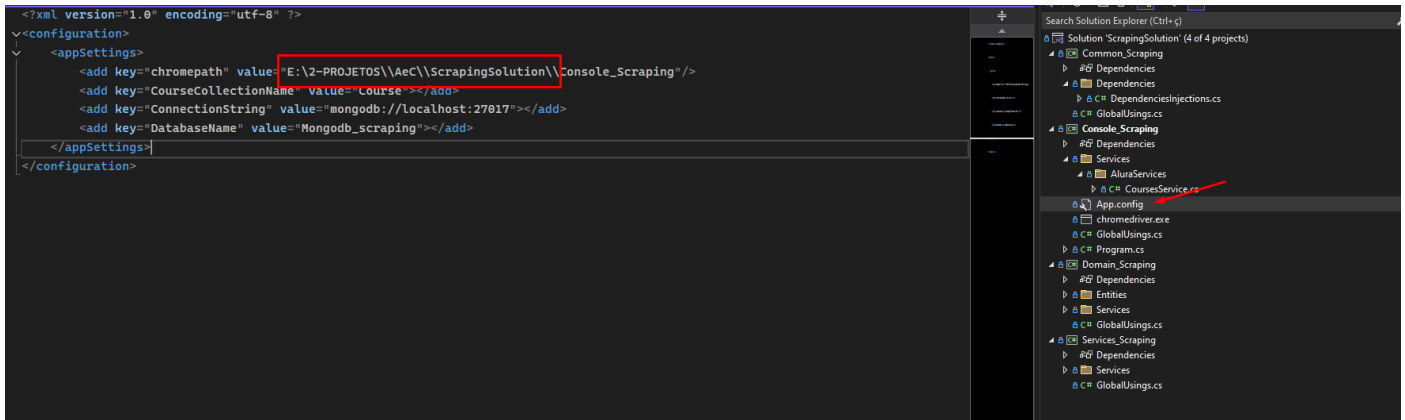
Para fins apenas de teste, foram apenas considerados termos simples, sem espaços em branco. Portanto, no exemplo usado no desenvolvimento, usou-se o termo “**IA**”, mas podem ser usados outros, desde que apenas 1 único sem espaços em branco.

Desafio AeC Automação

Observação

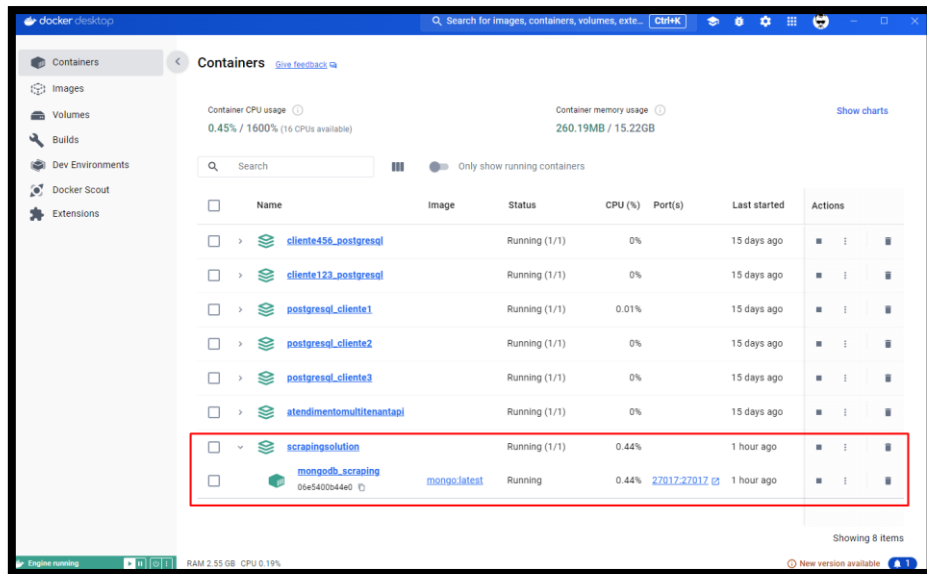
Após baixar o projeto para a máquina local, abrir o arquivo App.config, e alterar o caminho indicado no box vermelho, apontando para o caminho onde você instalou a solution, mantendo apenas o trecho final “\Console_Scraping”.

Esse caminho se refere ao local onde o sistema vai buscar o executável Chromedriver.exe.

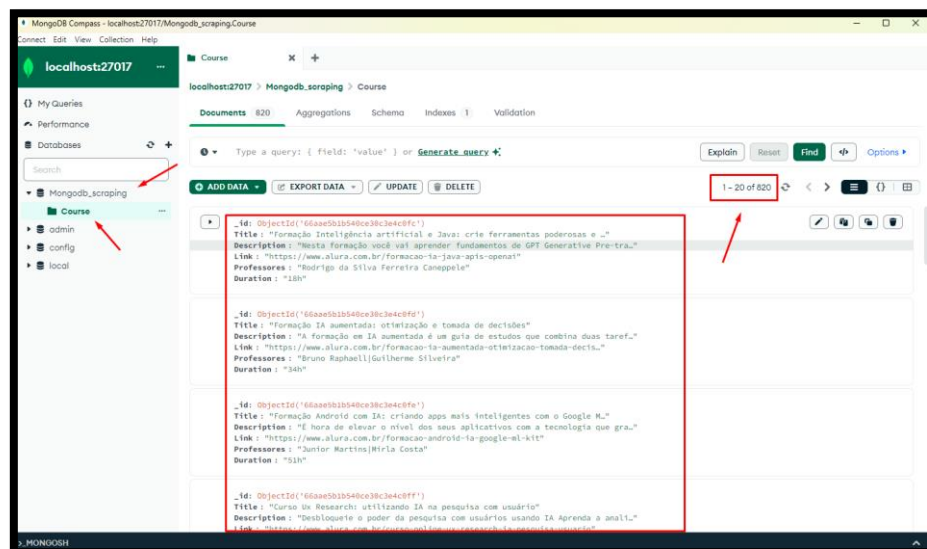


Desafio AeC Automação

Prints



Container do banco MongoDB



Exemplo com alguns dados do banco MongoDB visto pela IDE MongoDB Compass