



PUC Minas

DIRETORIA DE EDUCAÇÃO CONTINUADA - IEC

Disciplina: **GOVERNANÇA DE DADOS**

Nome do curso:

Professor: Ricardo Brito Alves

Trabalho Prático: Iniciando a Governança de Dados no Apache
Atlas

Aluno:

-

-

FRAMEWORK APACHE ATLAS

Objetivos

Exercitar os conhecimentos adquiridos para implementação de um catálogo de dados utilizando o Apache Atlas, um framework de governança de dados open-source. Abordaremos a instalação, configuração e utilização do Apache Atlas, relacionando algumas etapas com os objetivos abaixo.

- Familiarizar o aluno com o conceito de catálogo de dados e sua importância na governança de dados.
- Demonstrar como utilizar o Apache Atlas para criar um catálogo de dados simples.
- Aplicar os conhecimentos do DAMA-DMBOK, especificamente nas áreas de Metadados, Governança de Dados e Arquitetura de Dados.

Pré-requisitos:

- Docker instalado e em funcionamento.
- Conhecimento básico de linha de comando.
- Entender o DAMA-DMBOK.
- Ter um editor de texto (ex: Visual Studio Code, Sublime Text).

Enunciado

Vamos analisar o seguinte cenário:

1. Subindo o Apache Atlas no Docker:

Você precisa ter o Docker rodando na sua máquina.

Temos disponível 3 imagens para o Apache Atlas. Vamos utilizar a primeira.

- Imagem sburn/apache-atlas no Docker Hub
- Imagem sansarip/apache-atlas no Docker Hub
- Imagem spydernaz/apache-atlas no Docker Hub

No terminal, vamos baixar a imagem do Docker:

```
docker pull sburn/apache-atlas
```

Rodar o container

```
docker run -d -p 21000:21000 --name apache-atlas sburn/apache-atlas
```

Explicação do comando:

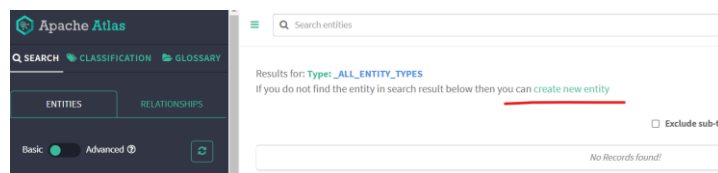
- -d: Executa o container em modo detached (em segundo plano).
- -p 21000:21000: Mapeia a porta 21000 do container para a porta 21000 do host, permitindo o acesso ao Apache Atlas através do navegador.
- --name atlas-server: Atribui um nome ao container para facilitar a sua identificação.

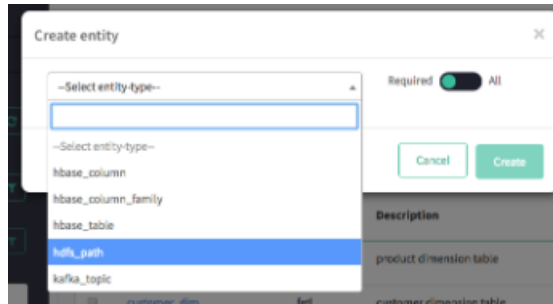
2. Acessar o Apache Atlas:

- Abrir um navegador e acessar a URL: <http://localhost:21000/>
- Criar um usuário e senha (tentar com admin/admin).
 - No primeiro acesso, você precisará criar um usuário administrador. Siga as instruções na tela para definir um nome de usuário e senha.
 - Este usuário terá permissões máximas para gerenciar todos os aspectos do Apache Atlas.

3. Criando um Modelo de Dados:

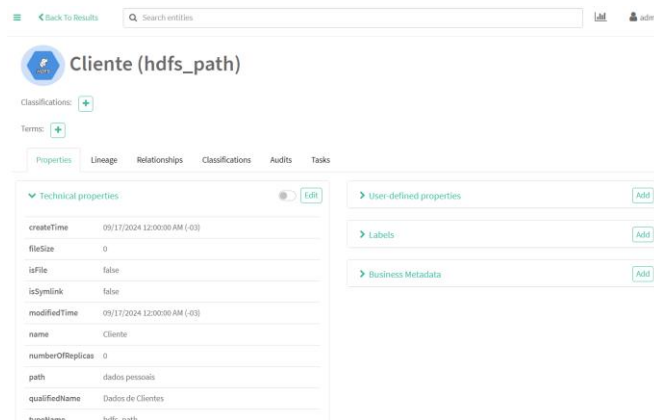
- **Acesso ao console:** Após fazer login, você terá acesso ao console do Apache Atlas.
- **Criar um tipo de entidade:** Representa uma tabela ou objeto em um banco de dados. Por exemplo, crie um tipo de entidade chamado "Cliente".
 - Em "Search by Type" digite Table e clique em Search.
 - Não vai achar e vai habilitar a possibilidade de se criar uma nova entidade.

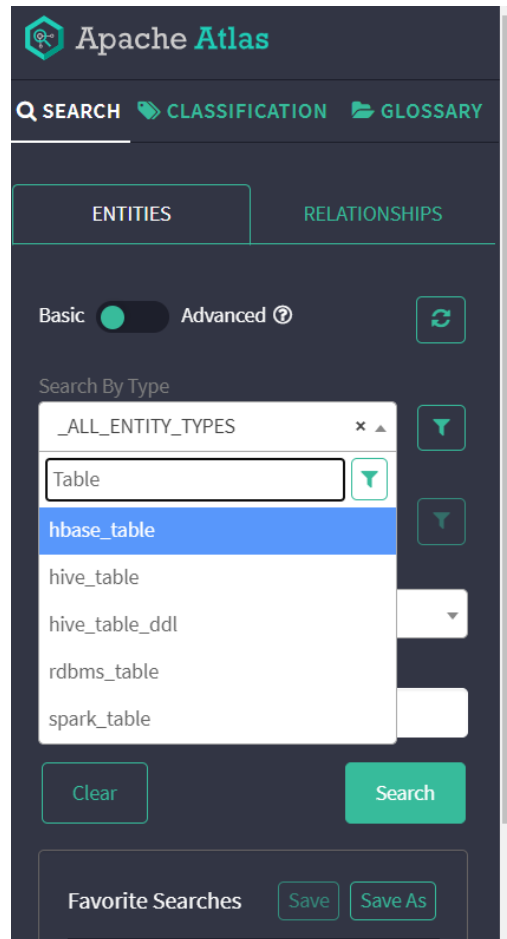




Insira as informações necessárias para a nova entidade. Clique em **All** para exibir as informações obrigatórias e não obrigatórias. Clique em **Create** para criar a nova entidade.

A entidade é criada e retornada nos resultados da pesquisa para o tipo de entidade aplicável. Agora você pode associar tags à nova entidade e controlar o acesso à entidade com políticas baseadas em tags.





- **Definir atributos:** Defina os atributos do tipo de entidade, como nome, sobrenome, email, etc. Associe cada atributo a um tipo de dado (string, inteiro, data, etc.).
- **Criar um tipo classificado:** Crie um tipo classificado para classificar os dados, por exemplo, "Dados Pessoais", "Dados Financeiros".
- **Associar tipos:** Associe o tipo de entidade "Cliente" ao tipo classificado "Dados Pessoais".

4. Importando Dados:

- **Utilizar a API REST:** O Apache Atlas possui uma API REST que permite a criação e atualização de entidades e atributos.
- **Criar um script:** Utilize uma linguagem de programação como Python ou Java para criar um script que faça chamadas à API REST, inserindo os dados no catálogo.
- **Exemplo em Python (utilizando a biblioteca requests):**

```
Python
import requests
```

```
url = "http://localhost:21000/api/atlas/v2/entity"
headers = {'content-type': 'application/json'}
```

```
data = {
    "typeName": "Cliente",
    "attributes": {
        "nome": "João da Silva",
        "email": "joao@example.com"
    }
}
response = requests.post(url, headers=headers, json=data)
print(response.text)
```

5. Criando Políticas de Governança:

- **Definir políticas de acesso:** Utilize as funcionalidades de segurança do Apache Atlas para definir regras de acesso aos dados, como quem pode visualizar, editar e deletar os dados.
- **Criar classificações:** Crie classificações para os dados, associando-as a políticas de acesso específicas.

6. Gerenciando Metadados:

- **Adicionar metadados:** Adicione metadados aos tipos de entidades, atributos e instâncias de entidades. Por exemplo, adicione um atributo "data_criacao" aos clientes para registrar a data de criação do registro.
- **Utilizar o catálogo de termos:** Crie um catálogo de termos para padronizar a terminologia utilizada na descrição dos dados.

7. Relatórios e Análises:

- **Utilizar os recursos de busca:** Utilize a interface de busca do Apache Atlas para encontrar entidades e atributos específicos.
- **Gerar relatórios:** Utilize os recursos de relatório para gerar informações sobre a qualidade dos dados, o uso dos dados e outras métricas relevantes.

Considerações

Relacionando com o DAMA-DMBOK:

- **Metadados:** O Apache Atlas permite criar e gerenciar metadados, que são essenciais para a descoberta e compreensão dos dados, como definido no DAMA-DMBOK.
- **Governança de Dados:** As políticas de acesso e classificação definidas no Apache Atlas são fundamentais para a governança de dados, garantindo a segurança e a qualidade dos dados.

- **Arquitetura de Dados:** O modelo de dados criado no Apache Atlas representa a arquitetura lógica dos dados, como definido no DAMA-DMBOK.

Próximos Passos

- Explorar outras funcionalidades: O Apache Atlas oferece diversas outras funcionalidades, como a integração com outras ferramentas de big data e a criação de fluxos de trabalho complexos.
- Personalizar o catálogo: Adapte o catálogo de dados às necessidades específicas da sua organização.
- Integrar com outras ferramentas: Integre o Apache Atlas com outras ferramentas de sua empresa, como ferramentas de BI e ETL.
- Este exercício fornece uma base para a utilização do Apache Atlas na implementação de um catálogo de dados. Ao seguir estes passos e explorar as funcionalidades do Apache Atlas, você poderá criar um catálogo de dados completo e eficaz para sua organização.
- Observação: Este exercício apresenta uma visão inicial do Apache Atlas. Para obter informações mais detalhadas, consulte a documentação oficial do Apache Atlas.

Considerações Finais:

- Este exercício é uma introdução ao uso do Apache Atlas para implementar um catálogo de dados.
- O DAMA-DMBOK fornece um framework abrangente para a gestão de dados, e o Apache Atlas é uma ferramenta poderosa para implementar algumas das práticas do DAMA-DMBOK.
- Existem muitas outras funcionalidades do Apache Atlas que podem ser exploradas, como a integração com outras ferramentas de big data e a criação de fluxos de trabalho complexos.