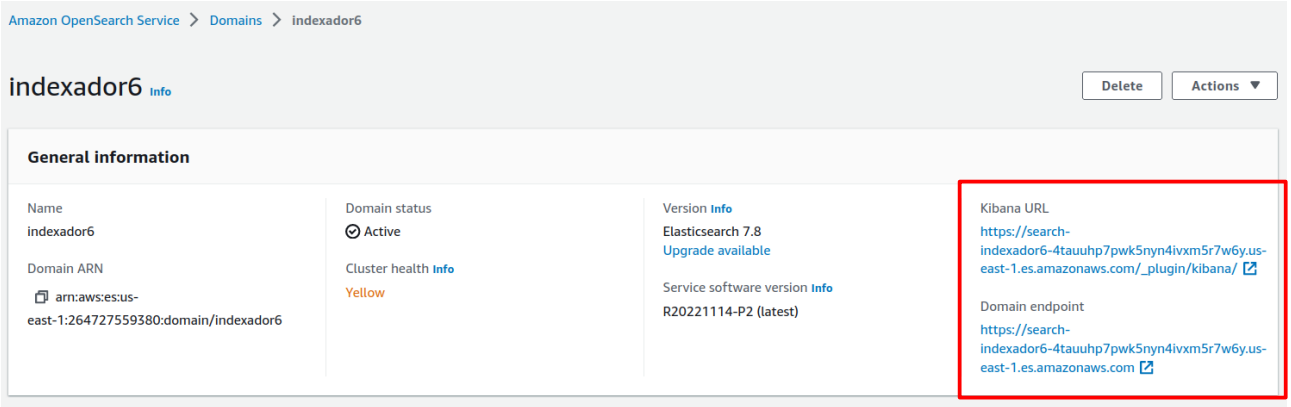
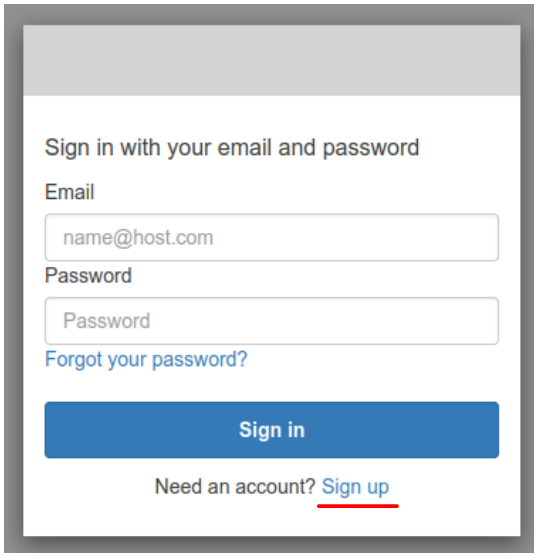
UTILIZANDO O KIBANA PARA ANALISAR A BASE DE DOCUMENTOS

Acessando o Kibana.

No console da AWS, pesquise pelo serviço Amazon OpenSearch Service. Uma vez que seu domínio do ElasticSearch esteja criado, selecione-o. Na janela aberta será mostrado a URL do endpoint e o caminho para acessar o Kibana:

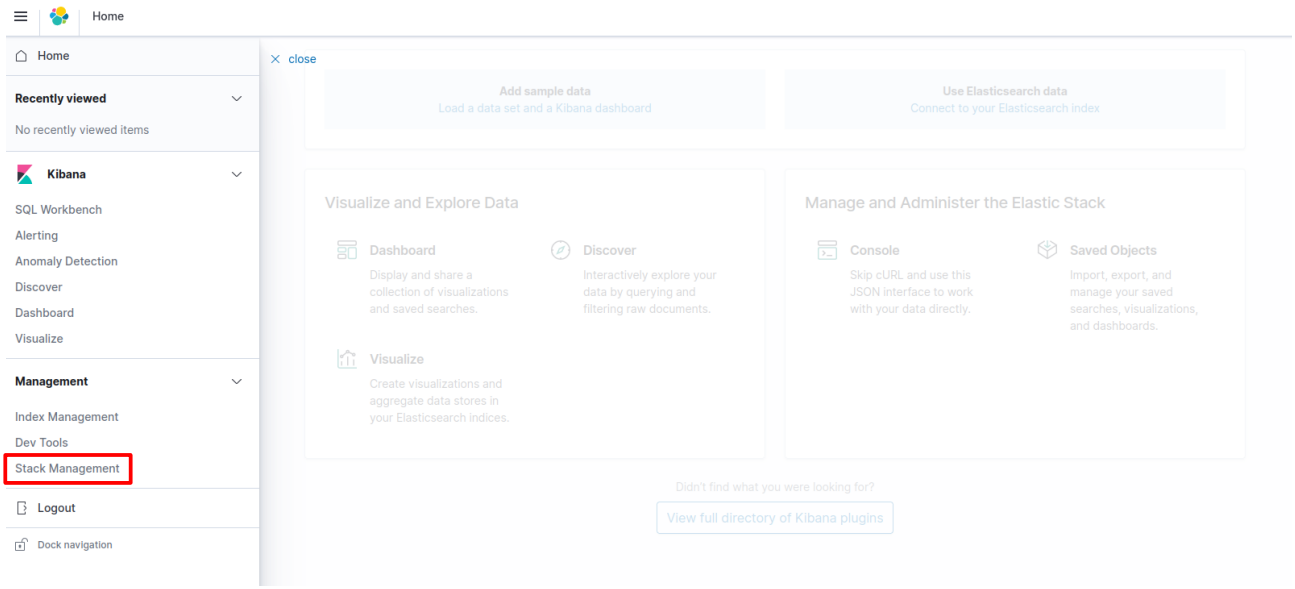


O Kibana requer o cadastro de um login específico. Clique em “Sign up” e cadastre um email e senha; um código será enviado para o email informado para validar o acesso.



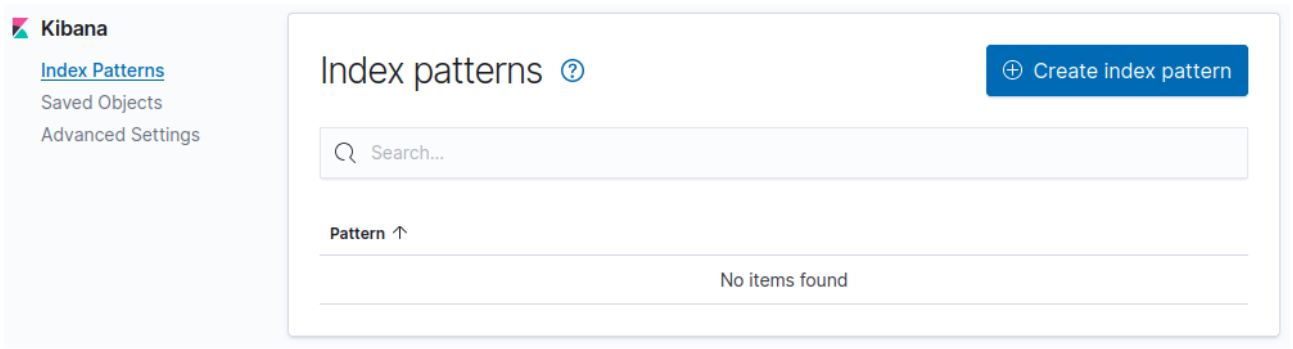
Criando um índice.

Uma vez logado no Kibana, no menu esquerdo, clique em “Stack Management”:

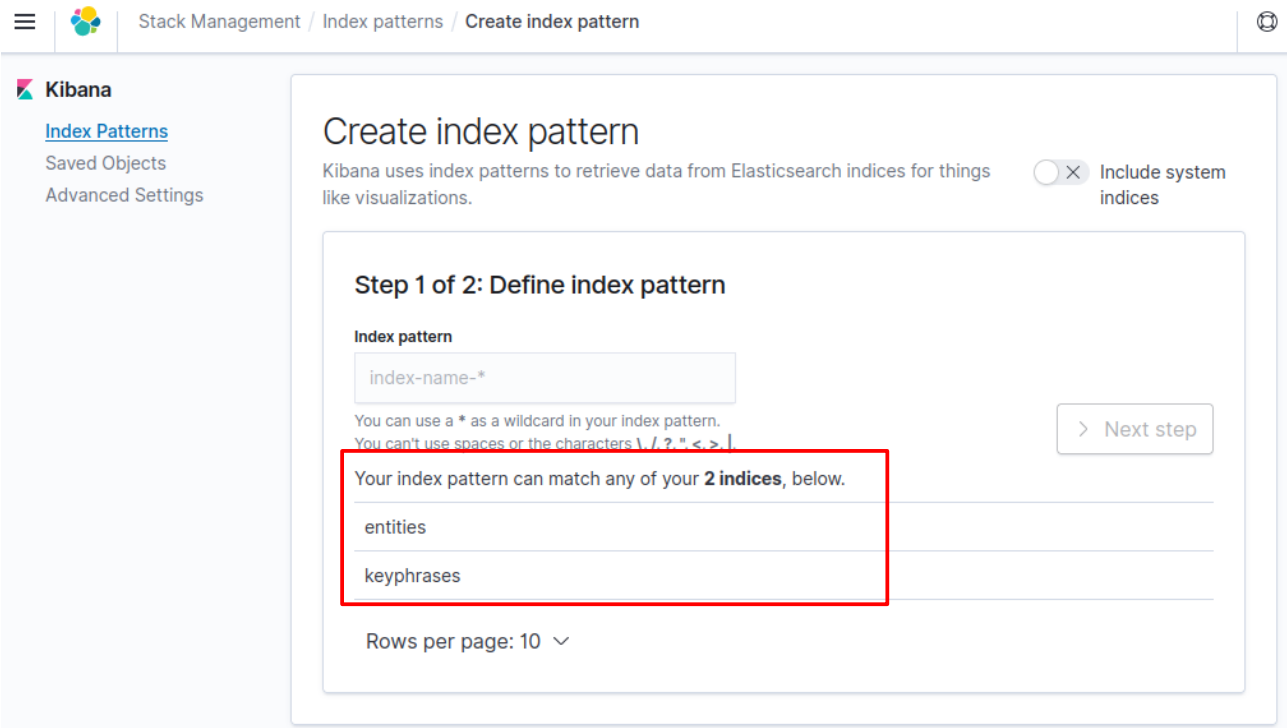


Crie um padrão de índice:

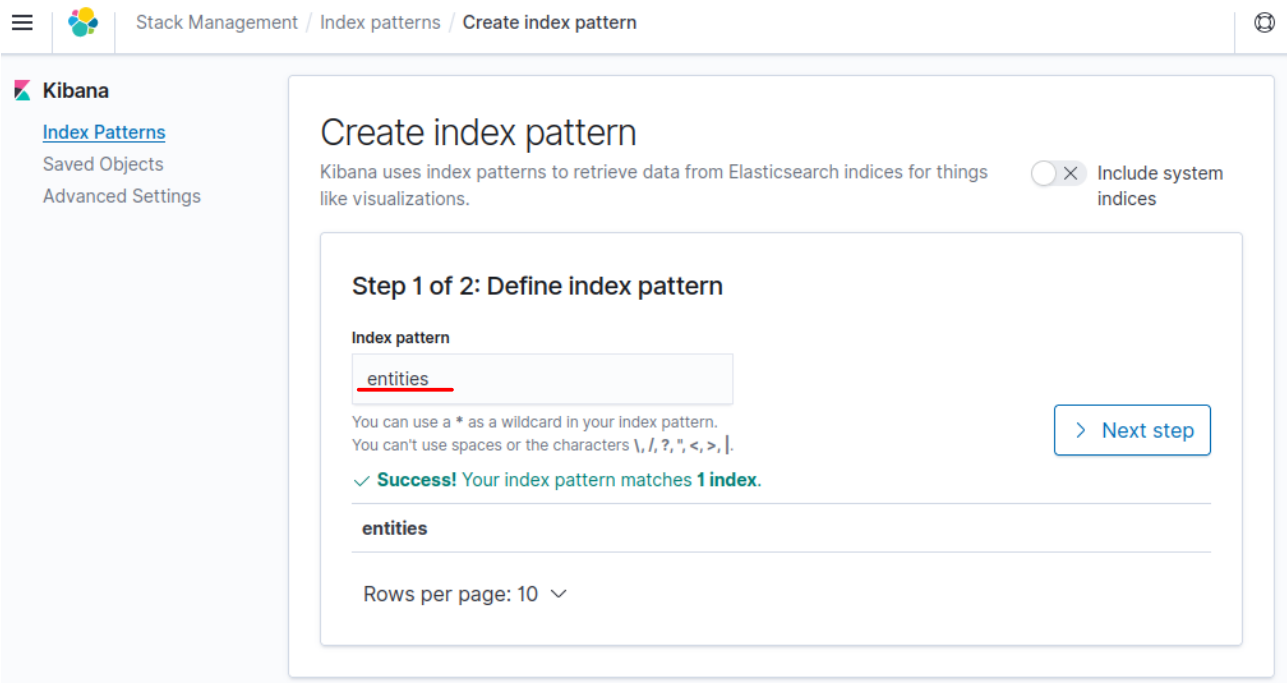
Index patterns



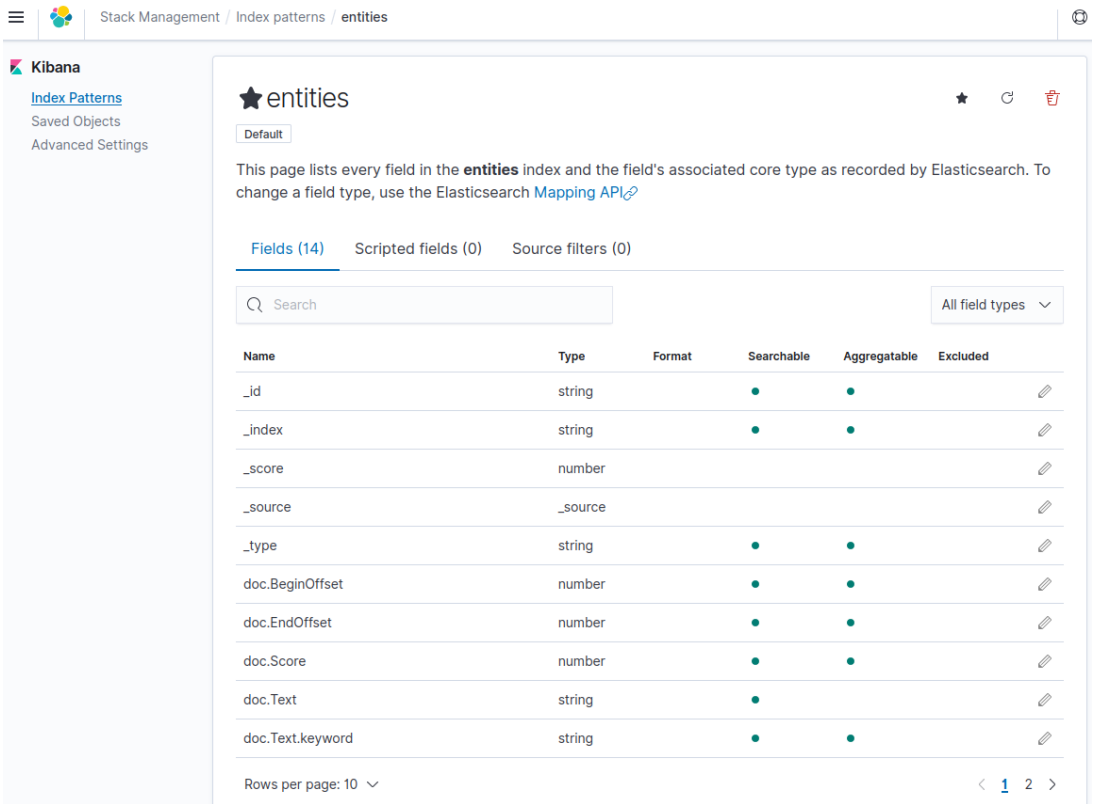
A partir dos seus dados já indexados (a transcrição dos áudios já processados), o Kibana irá indicar os índices disponíveis:



Digite o nome de um dos índices (**entities ou entidades**) e clique em “Next step”, e depois em “Create index pattern”:

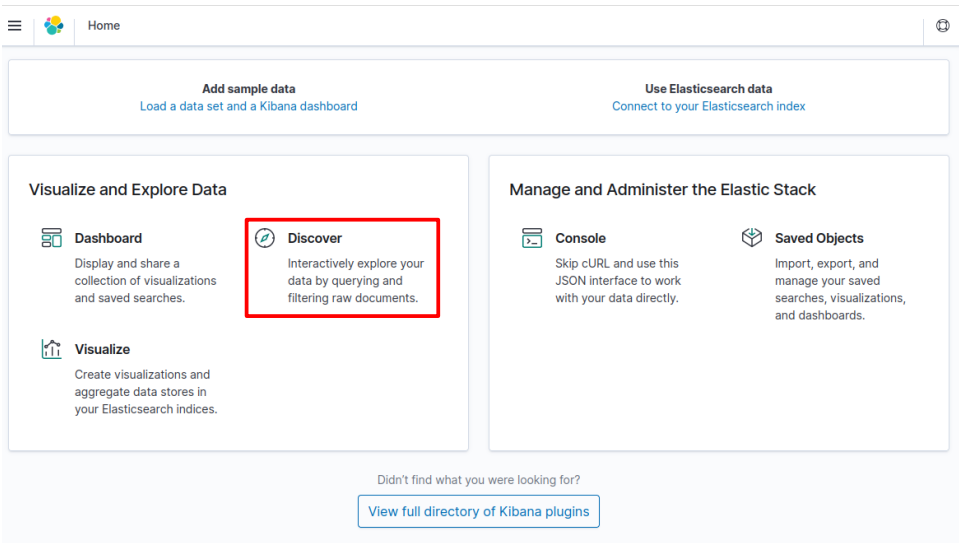


A próxima página irá listar o campos do índice (incluindo campos de metadados), com informaçõe como tipo e operações possíveis:

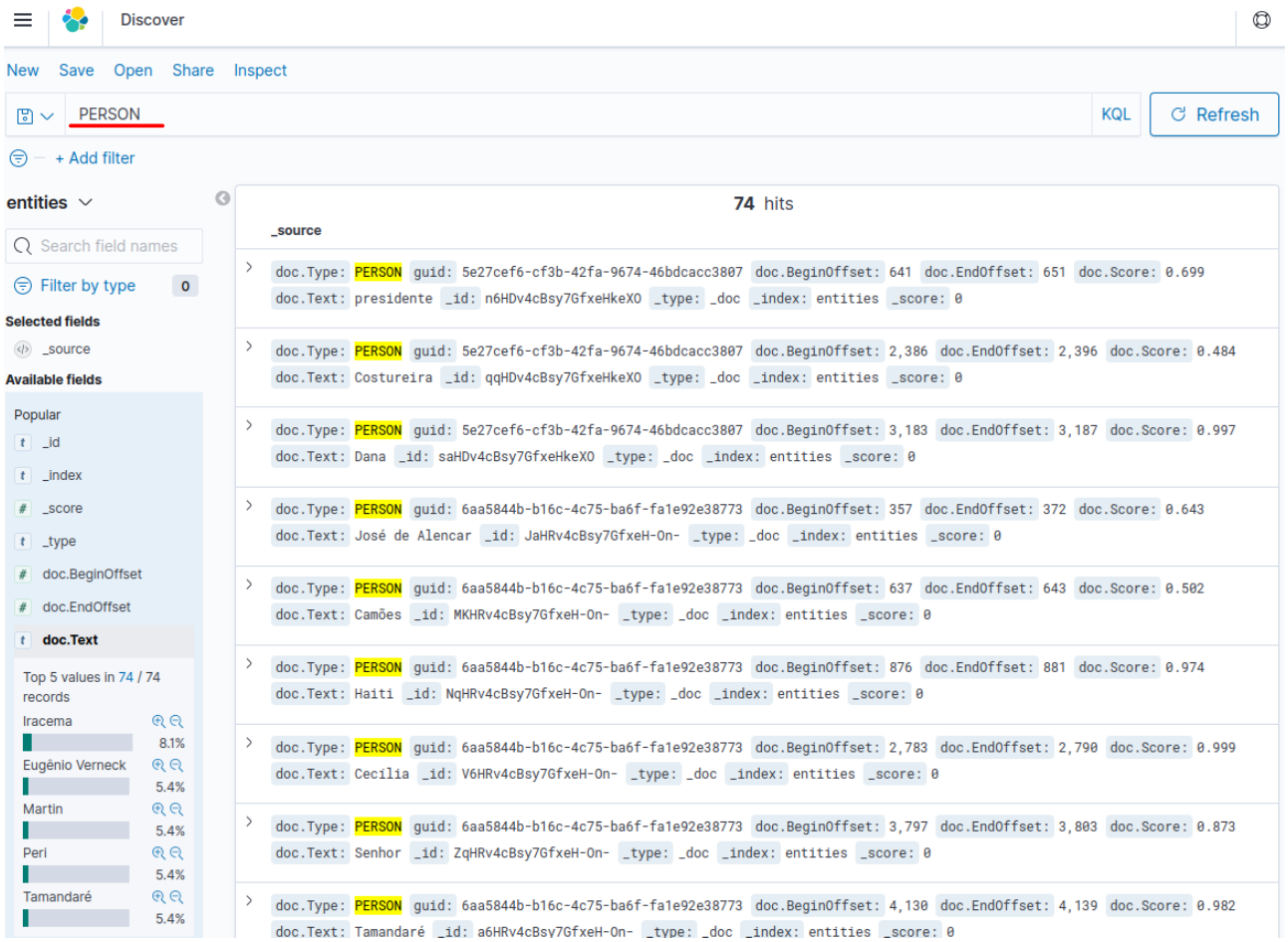


Analisando a base de dados.

Uma vez criado o índice, retorne ao menu principal e clique em “Discover”:

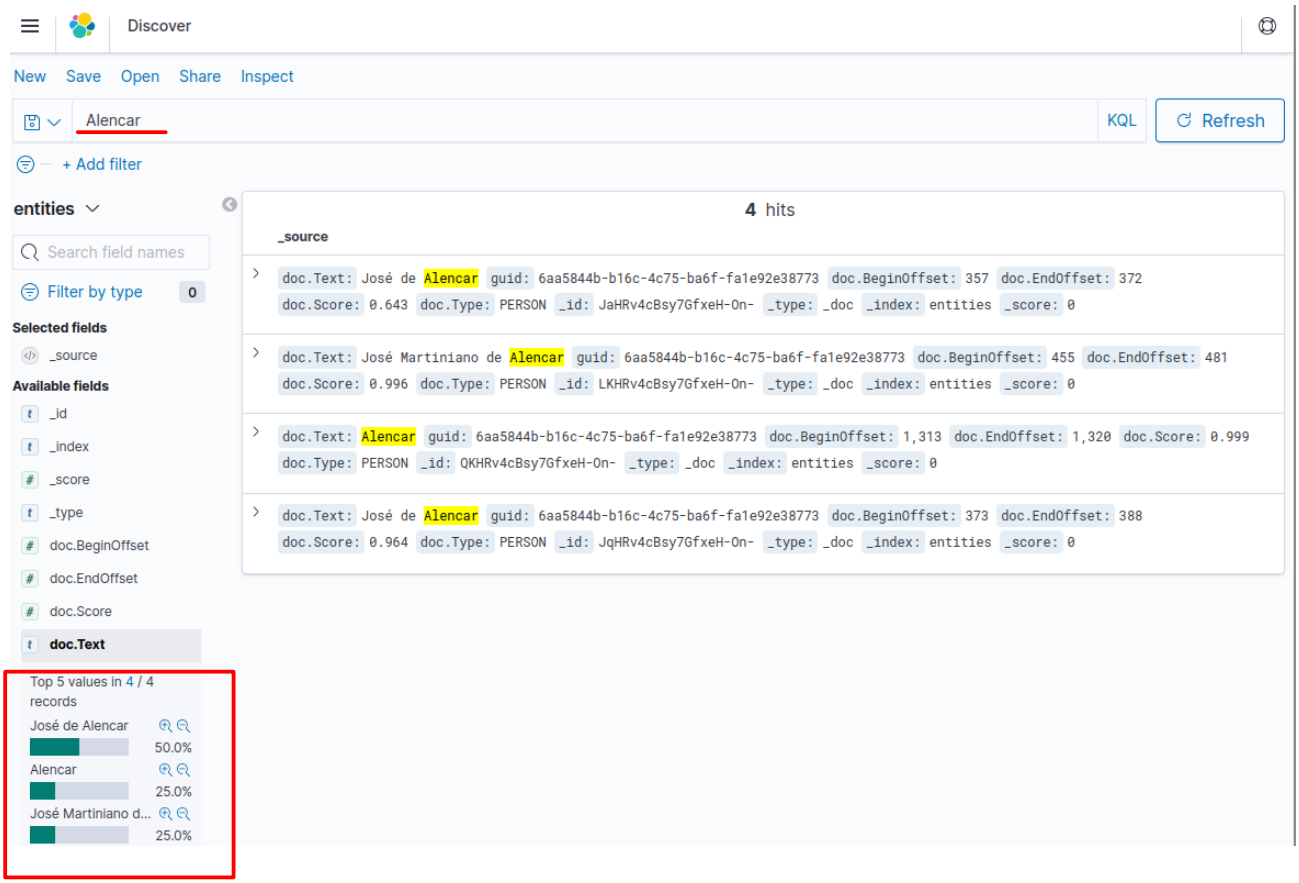


Na janela aberta, é possível realizar pesquisas sobre base de dados, como por exemplo, as entidades nomeadas do tipo pessoa (PERSON):



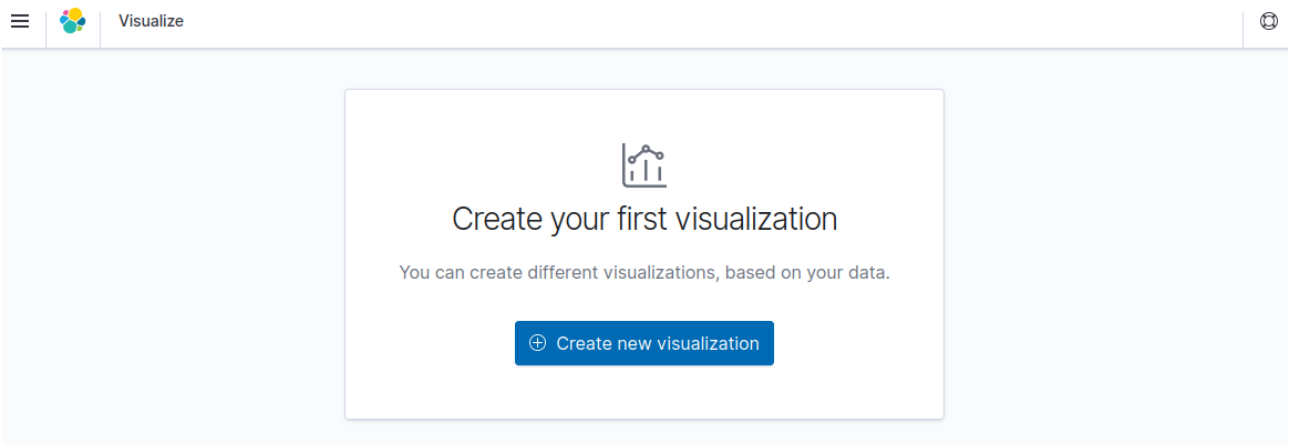
Também é possível construir e aplicar filtros por campos, ou filtrar por documento.

Ocorrência do termo “Alencar” no áudios indexados:

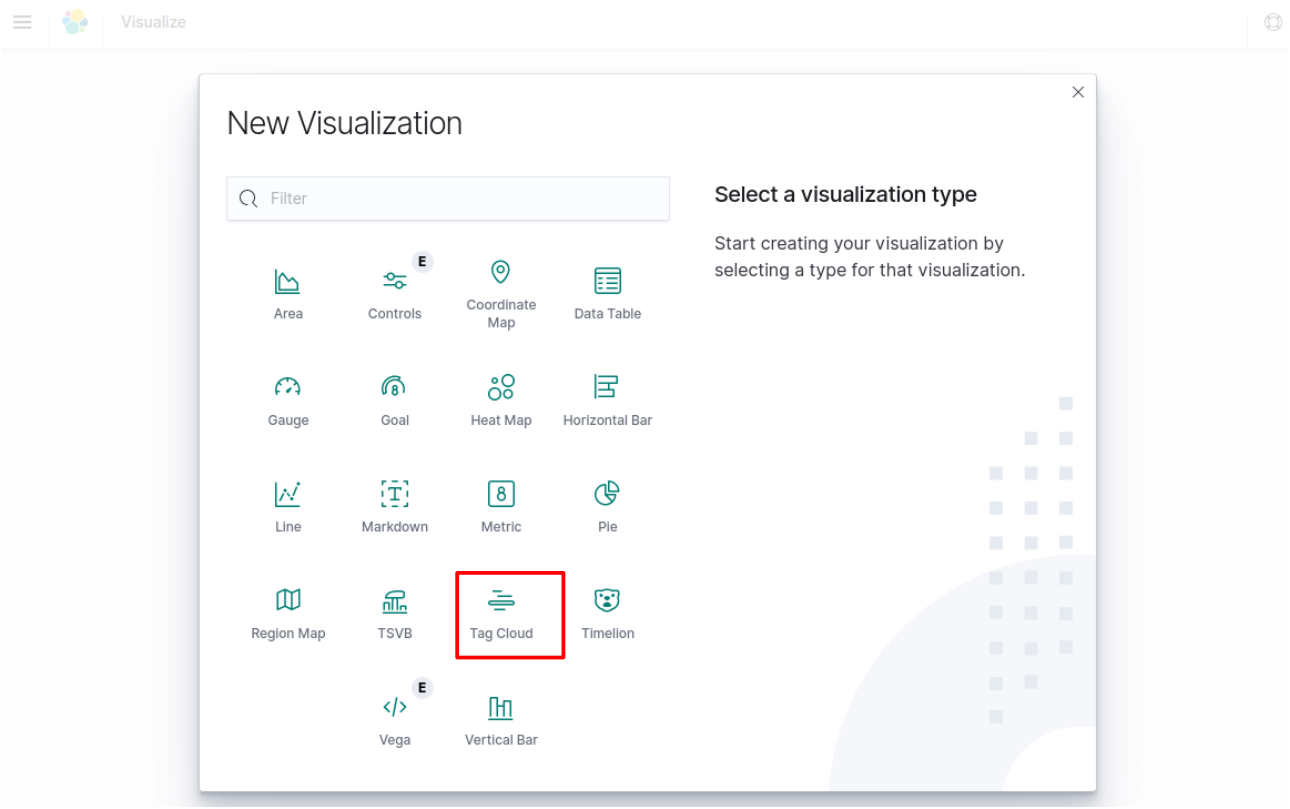


Criando visualizações das análises.

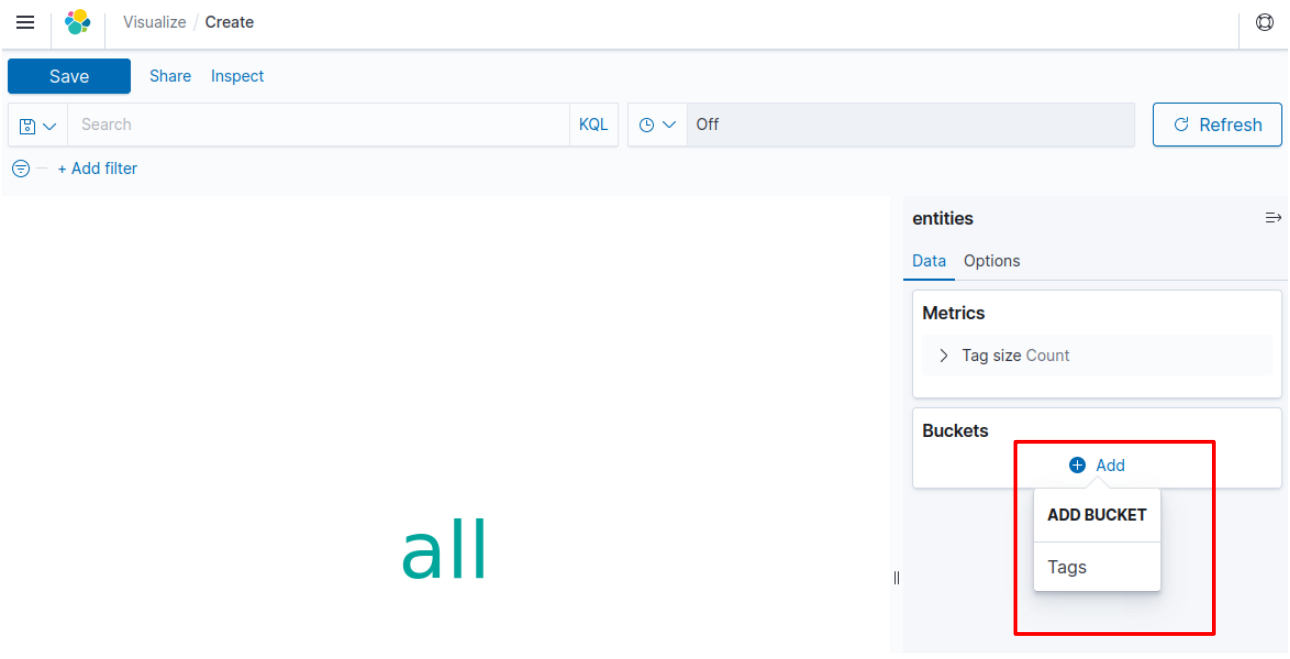
Na página inicial, selecione "Visualize" (Visualização).



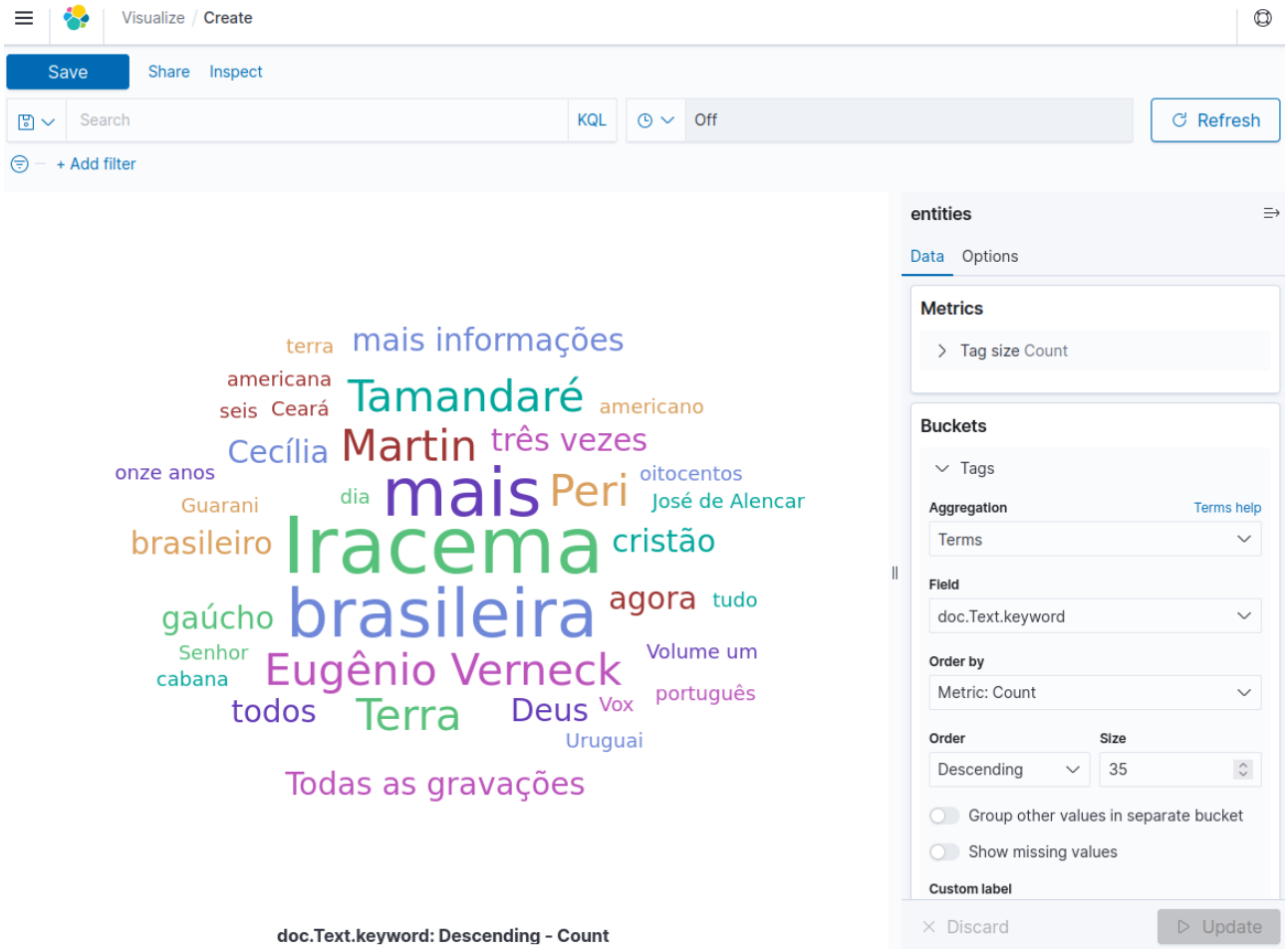
Há diversos tipos de visualizações disponíveis. Por exemplo, selecione “Tag Cloud” (nuvem de palavras):



Adicione o bucket com os dados:

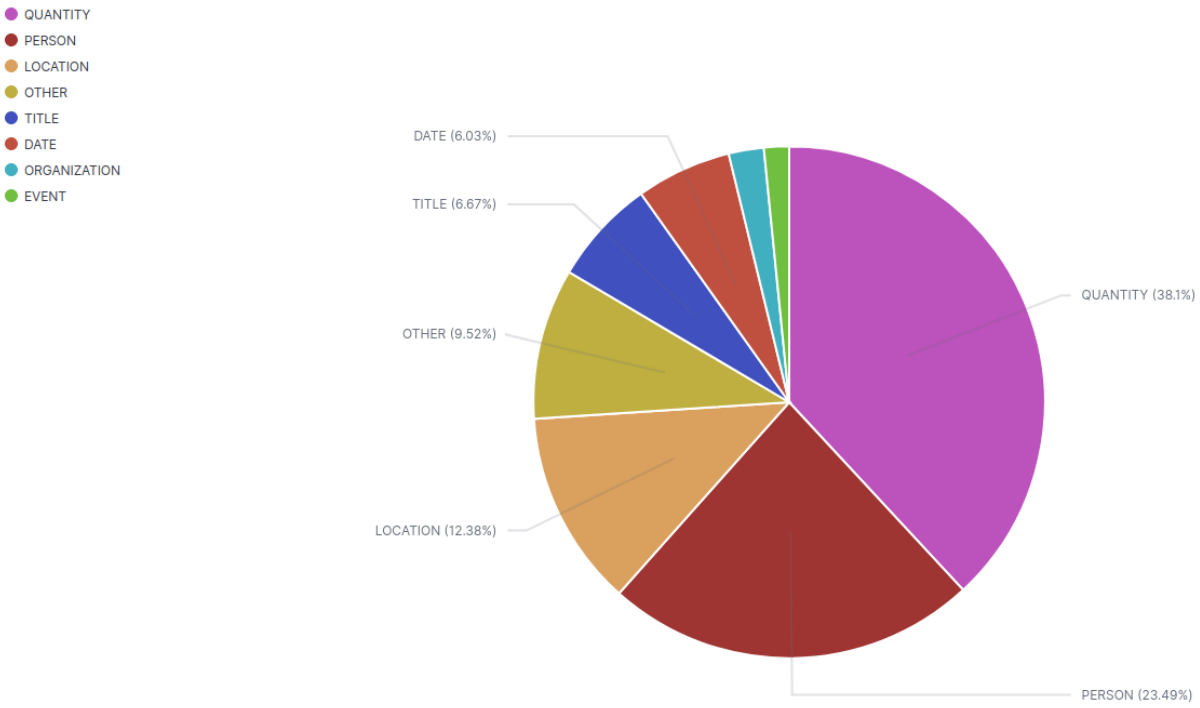


Personalize a seleção e clique em “Update”:



A nuvem de palavras criada mostra a ocorrência das palavras mais frequentes nos documentos indexados (stop words já foram tratadas e não são consideradas). A base de dados aqui utilizada consistia de transcrições de audiobooks de José de Alencar.

Outro exemplo de visualização, um gráfico de setores mostrando a proporção de tipos de entidades nomeadas indexadas na base:



As visualizações construídas são iterativas, e atualizadas conforme forem inseridos novos documentos, e a partir das visualizações, é possível compor dashboards que podem ser compartilhados para análises em tempo real da base de dados.