

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:

Amazon OpenSearch Service > Domains > indexador6

indexador6 [Info](#) Delete Actions ▼

General information

| | | | |
|---|---|--|---|
| Name indexador6 | Domain status ☑ Active | Version Info Elasticsearch 7.8 Upgrade available | Kibana URL https://search-indexador6-4tauuhp7pwk5nyn4ivxm5r7w6y.us-east-1.es.amazonaws.com/_plugin/kibana/ ↗ |
| Domain ARN arn:aws:es:us-east-1:264727559380:domain/indexador6 | Cluster health Info Yellow | Service software version Info R20221114-P2 (latest) | |

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.

Sign in with your email and password

Email

Password

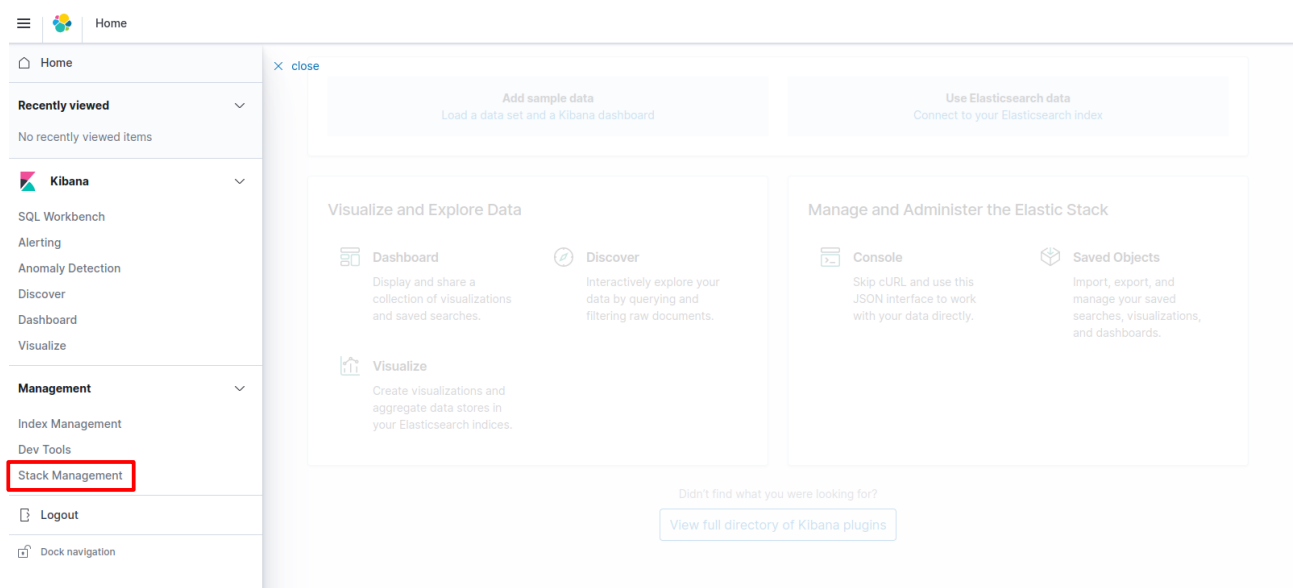
[Forgot your password?](#)

[Sign in](#)

Need an account? [Sign up](#)

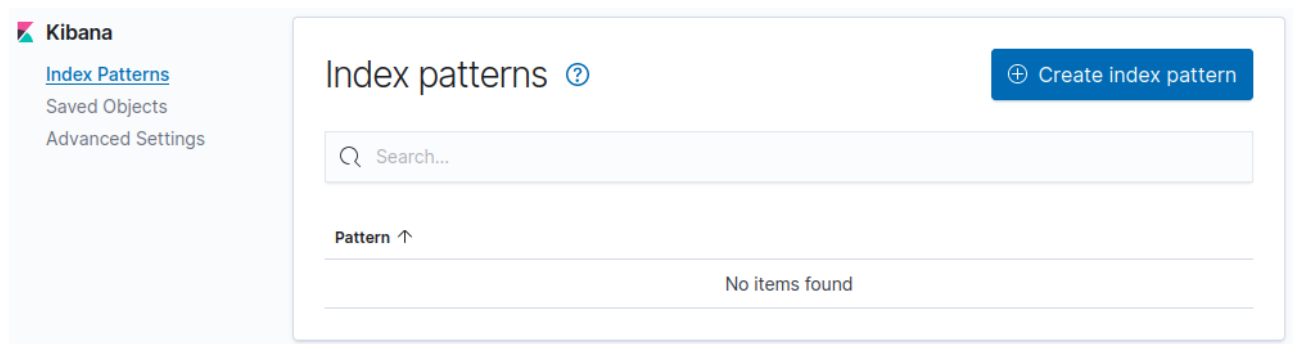
Criando um índice.

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:

Stack Management / Index patterns / Create index pattern

Kibana

Index Patterns
Saved Objects
Advanced Settings

Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

☐ Include system indices

Step 1 of 2: Define index pattern

Index pattern

index-name-*

You can use a * as a wildcard in your index pattern.
You can't use spaces or the characters \, /, ?, ", <, >, |.

> Next step

Your index pattern can match any of your **2 indices**, below.

- entities
- keyphrases

Rows per page: 10 ▾

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

Stack Management / Index patterns / Create index pattern

Kibana

Index Patterns
Saved Objects
Advanced Settings

Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

☐ Include system indices

Step 1 of 2: Define index pattern

Index pattern

entities

You can use a * as a wildcard in your index pattern.
You can't use spaces or the characters \, /, ?, ", <, >, |.

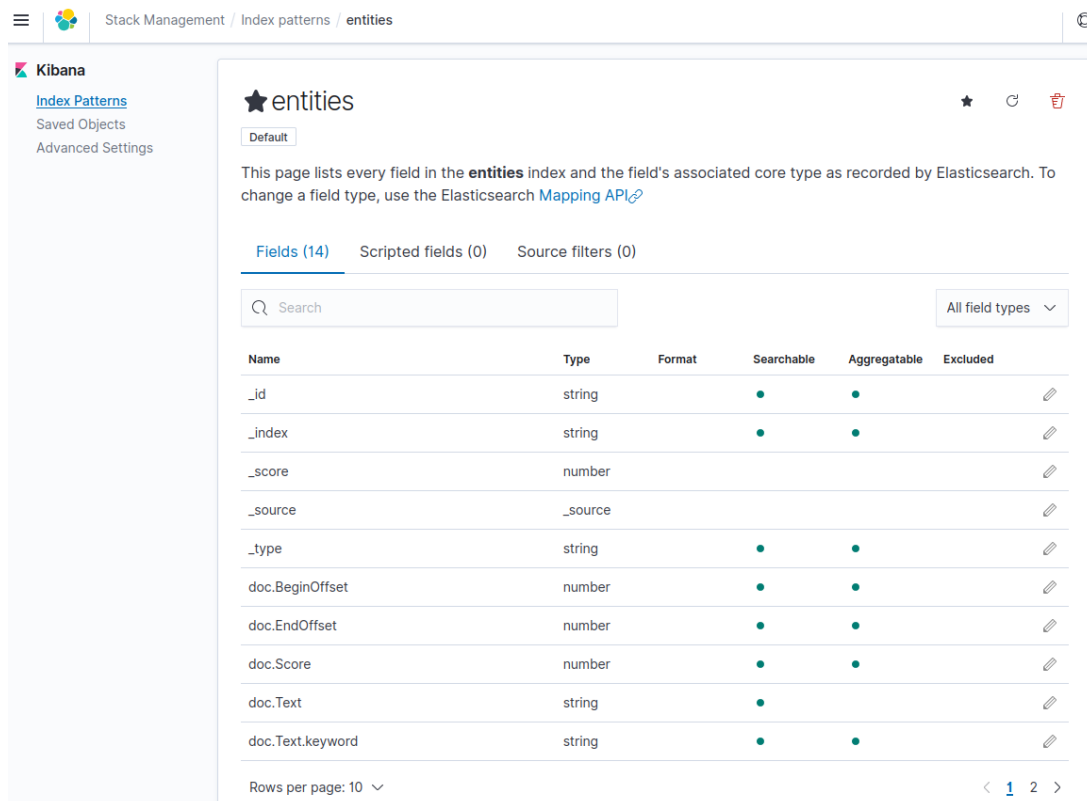
> Next step

✓ **Success!** Your index pattern matches **1 index**.

- entities

Rows per page: 10 ▾

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

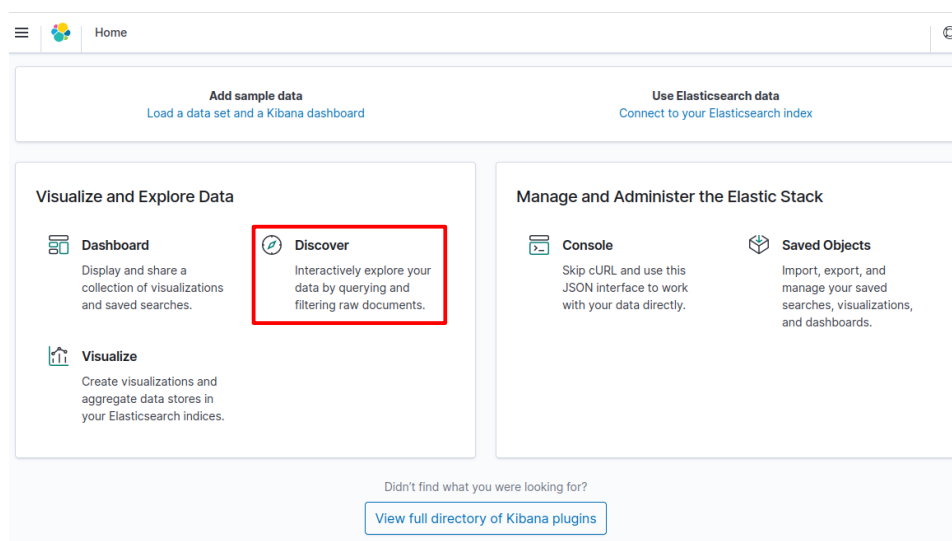


The screenshot shows the Kibana interface with the 'entities' page selected. The page title is 'entities' with a star icon. Below the title is a 'Default' button. A descriptive text states: 'This page lists every field in the **entities** index and the field's associated core type as recorded by Elasticsearch. To change a field type, use the Elasticsearch [Mapping API](#).' There are three tabs: 'Fields (14)', 'Scripted fields (0)', and 'Source filters (0)'. The 'Fields (14)' tab is active, showing a search bar and a dropdown for 'All field types'. Below is a table with columns: Name, Type, Format, Searchable, Aggregatable, and Excluded. The table lists 14 fields, including '_id', '_index', '_score', '_source', '_type', 'doc.BeginOffset', 'doc.EndOffset', 'doc.Score', 'doc.Text', and 'doc.Text.keyword'. Each row has an edit icon. At the bottom, it says 'Rows per page: 10' and has pagination links '< 1 2 >'.

| Name | Type | Format | Searchable | Aggregatable | Excluded |
|------------------|---------|--------|------------|--------------|----------|
| _id | string | | • | • | |
| _index | string | | • | • | |
| _score | number | | | | |
| _source | _source | | | | |
| _type | string | | • | • | |
| doc.BeginOffset | number | | • | • | |
| doc.EndOffset | number | | • | • | |
| doc.Score | number | | • | • | |
| doc.Text | string | | • | | |
| doc.Text.keyword | string | | • | • | |

Analizando a base de dados.

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



The screenshot shows the Kibana Home page. At the top, there's a 'Home' link. Below it, two main sections are visible: 'Add sample data' (with a link 'Load a data set and a Kibana dashboard') and 'Use Elasticsearch data' (with a link 'Connect to your Elasticsearch index'). The main content area is divided into two columns. The left column is titled 'Visualize and Explore Data' and contains three options: 'Dashboard' (with a description 'Display and share a collection of visualizations and saved searches.'), 'Discover' (highlighted with a red box, with a description 'Interactively explore your data by querying and filtering raw documents.'), and 'Visualize' (with a description 'Create visualizations and aggregate data stores in your Elasticsearch indices.'). The right column is titled 'Manage and Administer the Elastic Stack' and contains two options: 'Console' (with a description 'Skip cURL and use this JSON interface to work with your data directly.') and 'Saved Objects' (with a description 'Import, export, and manage your saved searches, visualizations, and dashboards.'). At the bottom, there's a link 'View full directory of Kibana plugins'.

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

Discover

New

Save

Open

Share

Inspect

PERSON

KQL

Refresh

+ Add filter

entities

Search field names

Filter by type

Selected fields

Available fields

Popular

Top 5 values in 74 / 74 records

74 hits

>

doc.Type: PERSON

guid: 5e27cef6-cf3b-42fa-9674-46bdcacc3807

doc.BeginOffset: 641

doc.EndOffset: 651

doc.Score: 0.699

doc.Text: presidente

_id: n6HDv4cBsy7GfXeHkeX0

_type: _doc

_index: entities

_score: 0

>

doc.Type: PERSON

guid: 5e27cef6-cf3b-42fa-9674-46bdcacc3807

doc.BeginOffset: 2,386

doc.EndOffset: 2,396

doc.Score: 0.484

doc.Text: Costureira

_id: qqHDv4cBsy7GfXeHkeX0

_type: _doc

_index: entities

_score: 0

>

doc.Type: PERSON

guid: 5e27cef6-cf3b-42fa-9674-46bdcacc3807

doc.BeginOffset: 3,183

doc.EndOffset: 3,187

doc.Score: 0.997

doc.Text: Dana

_id: saHDv4cBsy7GfXeHkeX0

_type: _doc

_index: entities

_score: 0

>

doc.Type: PERSON

guid: 6aa5844b-b16c-4c75-ba6f-fa1e92e38773

doc.BeginOffset: 357

doc.EndOffset: 372

doc.Score: 0.643

doc.Text: José de Alencar

_id: JaHRv4cBsy7GfXeH-On-

_type: _doc

_index: entities

_score: 0

>

doc.Type: PERSON

guid: 6aa5844b-b16c-4c75-ba6f-fa1e92e38773

doc.BeginOffset: 637

doc.EndOffset: 643

doc.Score: 0.502

doc.Text: Camões

_id: MKHRv4cBsy7GfXeH-On-

_type: _doc

_index: entities

_score: 0

>

doc.Type: PERSON

guid: 6aa5844b-b16c-4c75-ba6f-fa1e92e38773

doc.BeginOffset: 876

doc.EndOffset: 881

doc.Score: 0.974

doc.Text: Haiti

_id: NqHRv4cBsy7GfXeH-On-

_type: _doc

_index: entities

_score: 0

>

doc.Type: PERSON

guid: 6aa5844b-b16c-4c75-ba6f-fa1e92e38773

doc.BeginOffset: 2,783

doc.EndOffset: 2,790

doc.Score: 0.999

doc.Text: Cecilia

_id: V6HRv4cBsy7GfXeH-On-

_type: _doc

_index: entities

_score: 0

>

doc.Type: PERSON

guid: 6aa5844b-b16c-4c75-ba6f-fa1e92e38773

doc.BeginOffset: 3,797

doc.EndOffset: 3,803

doc.Score: 0.873

doc.Text: Senhor

_id: ZqHRv4cBsy7GfXeH-On-

_type: _doc

_index: entities

_score: 0

>

doc.Type: PERSON

guid: 6aa5844b-b16c-4c75-ba6f-fa1e92e38773

doc.BeginOffset: 4,130

doc.EndOffset: 4,139

doc.Score: 0.982

doc.Text: Tamandaré

_id: a6HRv4cBsy7GfXeH-On-

_type: _doc

_index: entities

_score: 0

Iracema

8.1%

Eugênio Verneck

5.4%

Martin

5.4%

Peri

5.4%

Tamandaré

5.4%

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

Ocorrência do termo “Alencar” no áudios indexados:

The screenshot shows the 'Discover' interface with a search for 'Alencar'. The results are displayed in a table with 4 hits. The left sidebar shows the 'entities' section with a search for 'Alencar' and a list of available fields. The 'Selected fields' section shows the '_source' field. The 'Available fields' section lists various fields including '_id', '_index', '_score', '_type', 'doc.BeginOffset', 'doc.EndOffset', 'doc.Score', and 'doc.Text'. A red box highlights the 'Top 5 values in 4 / 4 records' section, which shows the following data:

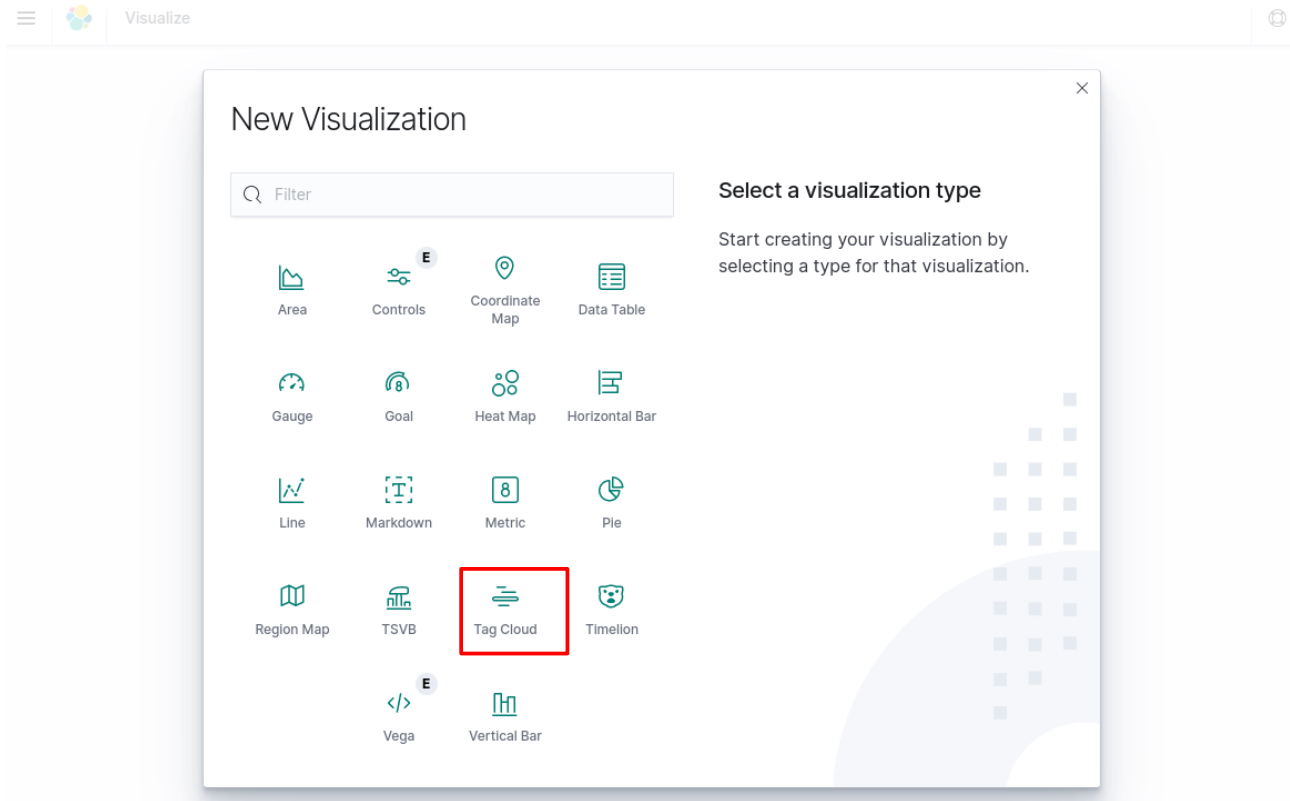
| Value | Count | Percentage |
|----------------------|-------|------------|
| José de Alencar | 2 | 50.0% |
| Alencar | 1 | 25.0% |
| José Martiniano d... | 1 | 25.0% |

Criando visualizações das análises.

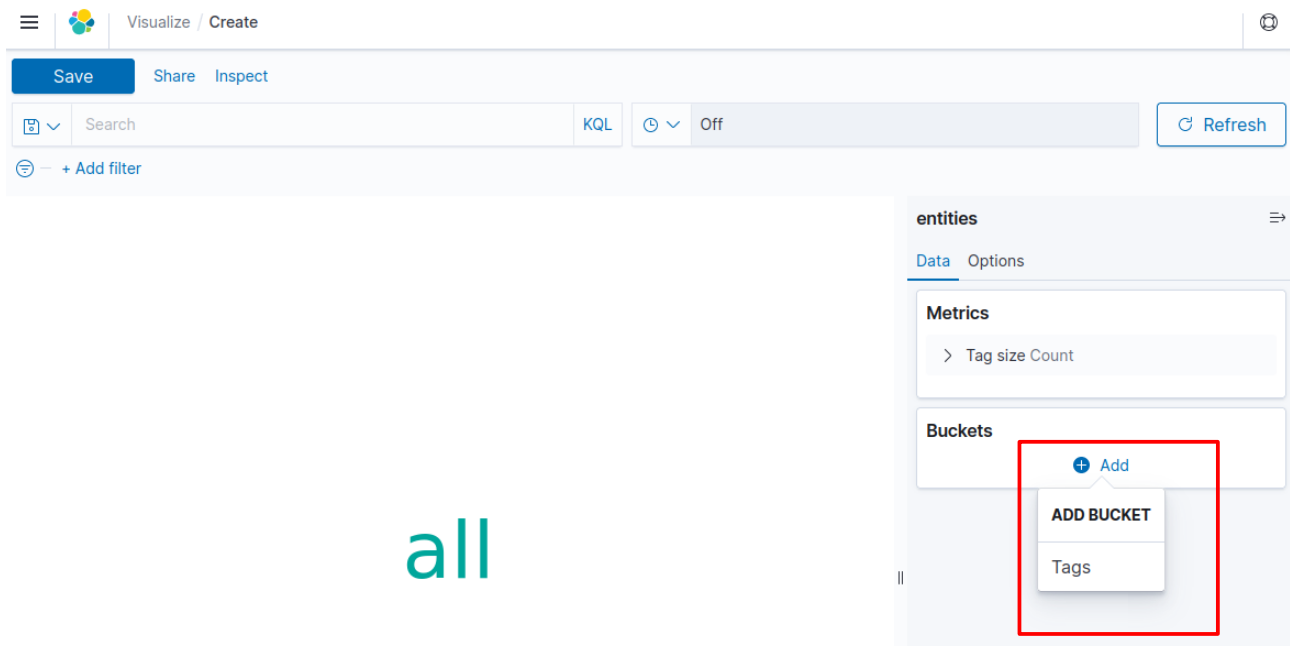
Na página inicial, selecione "Visualize" (Visualização).

The screenshot shows the 'Visualize' interface with a prompt to create a visualization. The prompt text is: 'Create your first visualization. You can create different visualizations, based on your data.' Below the prompt is a button labeled 'Create new visualization'.

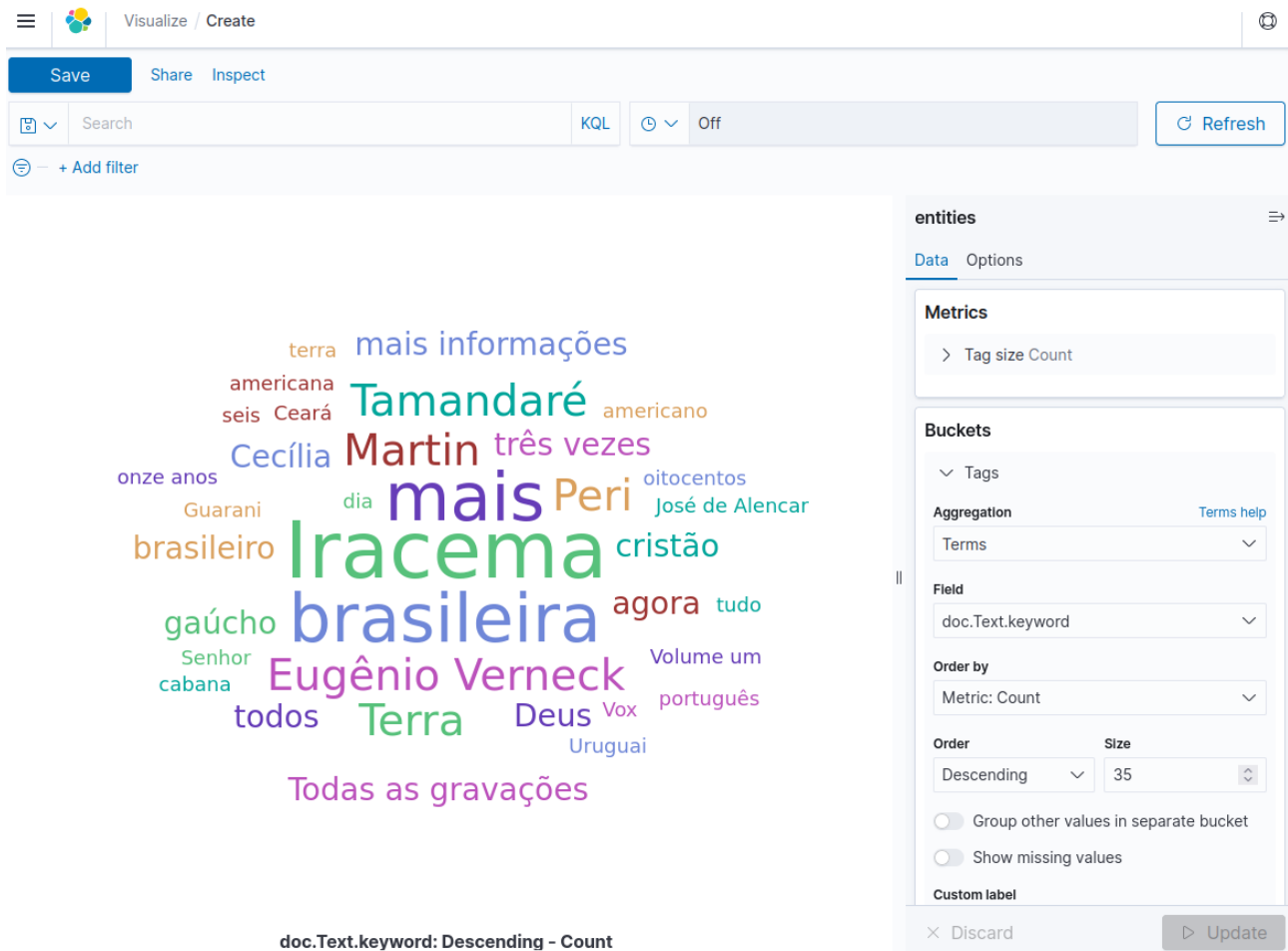
Há diversos tipos de visualizações disponíveis. Por exemplo, selecione “Tag Cloud” (nuvem de palavras):



Adicione o bucket com os dados:

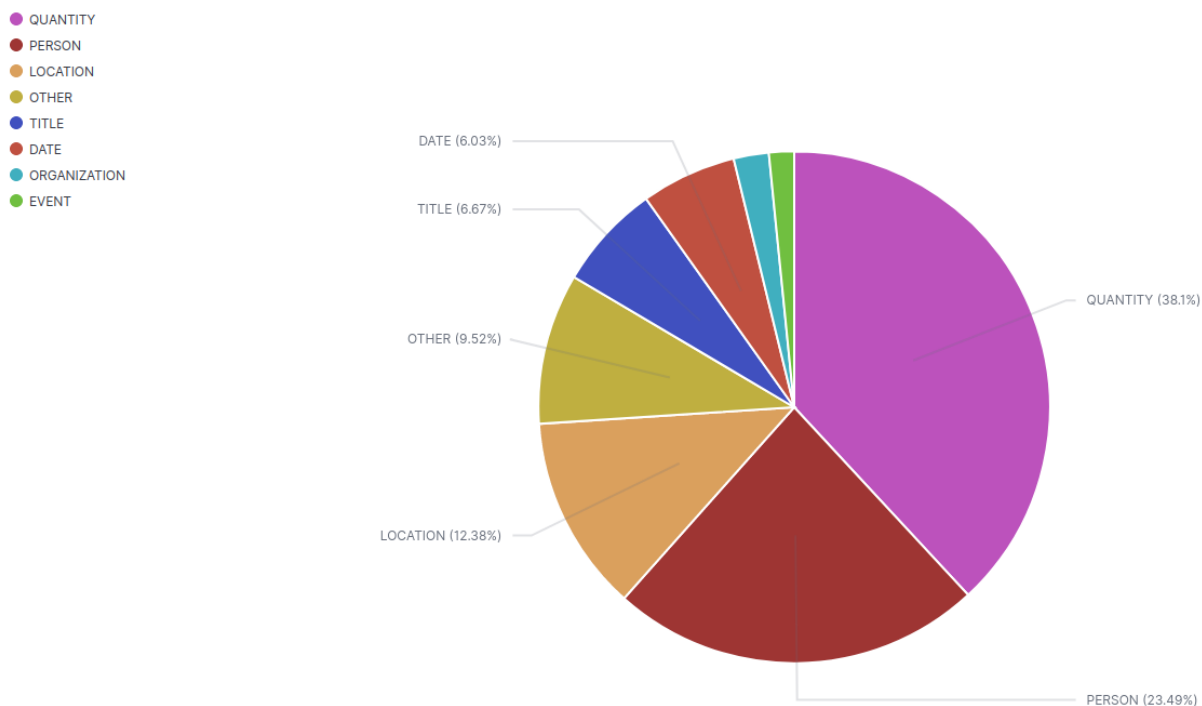


Personalize a seleção e clique em “Update”:



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.

