

10 - Agregace

Elasticsearch není jen vyhledávací engine, je také analytický engine. Pomocí agregací lze získat přehled o datech uložených v Elasticsearch.

Aggregations syntax

Agregace se zapisují do těla requestu pod klíč `aggregations`, případně zkráceně `aggs`:

```
GET _search
{
  "query": {...},
  "aggs": {
    ...
  }
}
```

Agregací je možné počítat více zároveň:

```
GET _search
{
  "query": {...},
  "aggs": {
    "aggregations" : {
      "<aggregation_name_1>" : {
        "<aggregation_type_1>" : {
          ...
        }
      },
      "<aggregation_name_2>" : {
        ...
      }
    }
  }
}
```

Nebo je možné agregace zanořovat:

```
GET _search
{
  "query": {...},
  "aggs": {
    "aggregations" : {
```

```
"<aggregation_name_1>" : {
  "<aggregation_type_1>" : {
    ...
  },
  "<aggregation_name_2>" : {
    ...
  }
}
}
```

Název agregace si lze zvolit libovolný. Pokud jsou agregace zanořeny, počítají se vnořené agregace pro jednotlivé skupiny agregací na vyšší úrovni. Pokud budeme chtít spočítat průměrnou cenu objednávky každý den, napíšeme nejprve agregaci, která data rozdělí po dnech (bucket - např. `date_histogram`) a následně použijeme agregaci vnořenou (metric - např. `avg`).

S agregacemi se budeme setkávat také v Kibaně - veškeré vizualizace získávají data právě pomocí vizualizací.

Agregací existuje několik druhů, které se liší způsobem, kterým pracují s daty:

Bucket aggregations

Prvním typem agregací jsou tzv. bucket agregace. Ty seskupí hodnoty ve vyhledaných dokumentech podle daných pravidel do skupin. Jde o alternativu k GROUP BY relačních databází.

- `terms`: seskupení podle hodnoty tokenu
- `date_histogram`: seskupení podle data/času do časových úseků
- `filter`: seskupení podle uživatelsky definovaných filtrů
- `range`: seskupení podle uživatelsky definovaných rozsahů (číselných nebo časových)

Metric aggregations

Druhým typem agregací jsou ty, které provádějí statistické výpočty nad nalezenými produkty - minima, maxima nebo průměry.

- `avg`: průměr
- `max`: maximum
- `min`: minimum
- `stats`: více statistických výpočtů zároveň

Pipeline aggregations

Speciální druh agregací, které namísto práce s dokumenty pracují s výsledky agregací. Obecně existují dva typy takových agregací:

Parent agregace pracují s výsledky nadřazených agregací a přidávají do nich další výsledky výpočtů. Zpravidla je třeba uvést cestu, s kterou nadřazenou agregací pracují jako `buckets_path`.

Sibling agregace pracují s výsledky agregací na shodné úrovni.

Příkladem pipeline agregace může být `cumulative_sum`, která přičítává hodnoty jednotlivých výsledků agregace. Lze ji použít například pro výpočet stavu účtu na základě jednotlivých transakcí:

```
GET /_search
{
  "aggs": {
    "my_agg_days": {
      "date_histogram": {
        "field": "date",
        "calendar_interval": "day"
      },
      "aggs": {
        "my_agg_sum": {
          "sum": {"field": "amount"}
        },
        "my_agg_balance": {
          "cumulative_sum": {"buckets_path": "my_agg_sum"}
        }
      }
    }
  }
}
```

Dalším příkladem může být `bucket_sort`. Pokud bychom chtěli seřadit měsíce podle největších zůstatků na účtu, bylo by to možné následujícím způsobem:

```
GET /_search
{
  "aggs": {
    "my_agg_days": {
      "date_histogram": {
        "field": "date",
        "calendar_interval": "day"
      },
      "aggs": {
        "my_agg_sum": {
          "sum": {"field": "amount"}
        },
        "my_agg_top_3_days": {
          "bucket_sort": {
            "sort": [
              {"my_agg_sum": {"order": "desc"}}
            ],
            "size": 3
          }
        }
      }
    }
  }
}
```

```
    }
  }
}
}
```

Post filter

V případě kombinace vyhledávání a agregací je třeba někdy vypočítat agregace před samotnou filtrací. V takovém případě je možné tyto filtrace přesunout do sekce `post_filter` zapsané na nejvyšší úrovni request body. Příkladem může být e-shop, kde nejprve spočítám data pro filtry pomocí agregací a následně pomocí `post_filter` vyfiltruji data podle zaškrtnutých hodnot v uživatelském filtru.

```
GET _search
{
  "query": { // <= query spuštěná před agregacemi
    "term": {
      "field": "value"
    }
  },
  "aggs": {
    ...
  },
  "post_filter": { // <= documents filtrovány po výpočtu agregací
    "term": {
      "field": "value"
    }
  }
}
```

Úkol: agregace

V indexu `kibana_sample_data_ecommerce`:

1. Vypište `email` adresy, které vytvořily nejvíc objednávek
2. Pro každý `day_of_week` zjistěte průměrnou cenu objednávky (s použitím pole `taxful_total_price`)
3. Vyfiltrujte data podle pole `geoip.continent_name` s hodnotou `North America` (můžete použít `term` query); poté pro každé pohlaví (`customer_gender`) spočítejte průměrný počet produktů (použijte `avg` agregaci a pole `total_unique_products`)

Kibana visualizations

Kibana využívá agregace pro tvorbu vizualizací. Ty je možné vytvořit prostřednictvím **Visualize Library**:

Visualize Library

Building a dashboard? Create and add your visualizations right from the [Dashboard application](#).

Search...

Tags

Title	Type	Description	Tags	Actions
<input type="checkbox"/> Failed Logons [Windows System Security]	Metric			
<input type="checkbox"/> % of target revenue (\$10k)	Lens			
<input type="checkbox"/> Admin Logons Simple [Windows System Security]	Metric			
<input type="checkbox"/> Administrator Logons [Windows System Security]	TSVB			
<input type="checkbox"/> Administrator Users [Windows System Security]	Pie			
<input type="checkbox"/> Avg. items sold	Lens			
<input type="checkbox"/> Blocked Accounts [Windows System Security]	Metric			
<input type="checkbox"/> Blocked Accounts TSVB [Windows System Security]	TSVB			
<input type="checkbox"/> Blocked Accounts Tag [Windows System Security]	Tag cloud			

Zvolte **Aggregation based** visualizations:

New visualization

Lens
Create visualizations with our drag and drop editor. Switch between visualization types at any time. *Recommended for most users.*

Maps
Create and style maps with multiple layers and indices.

TSVB
Perform advanced analysis of your time series data.

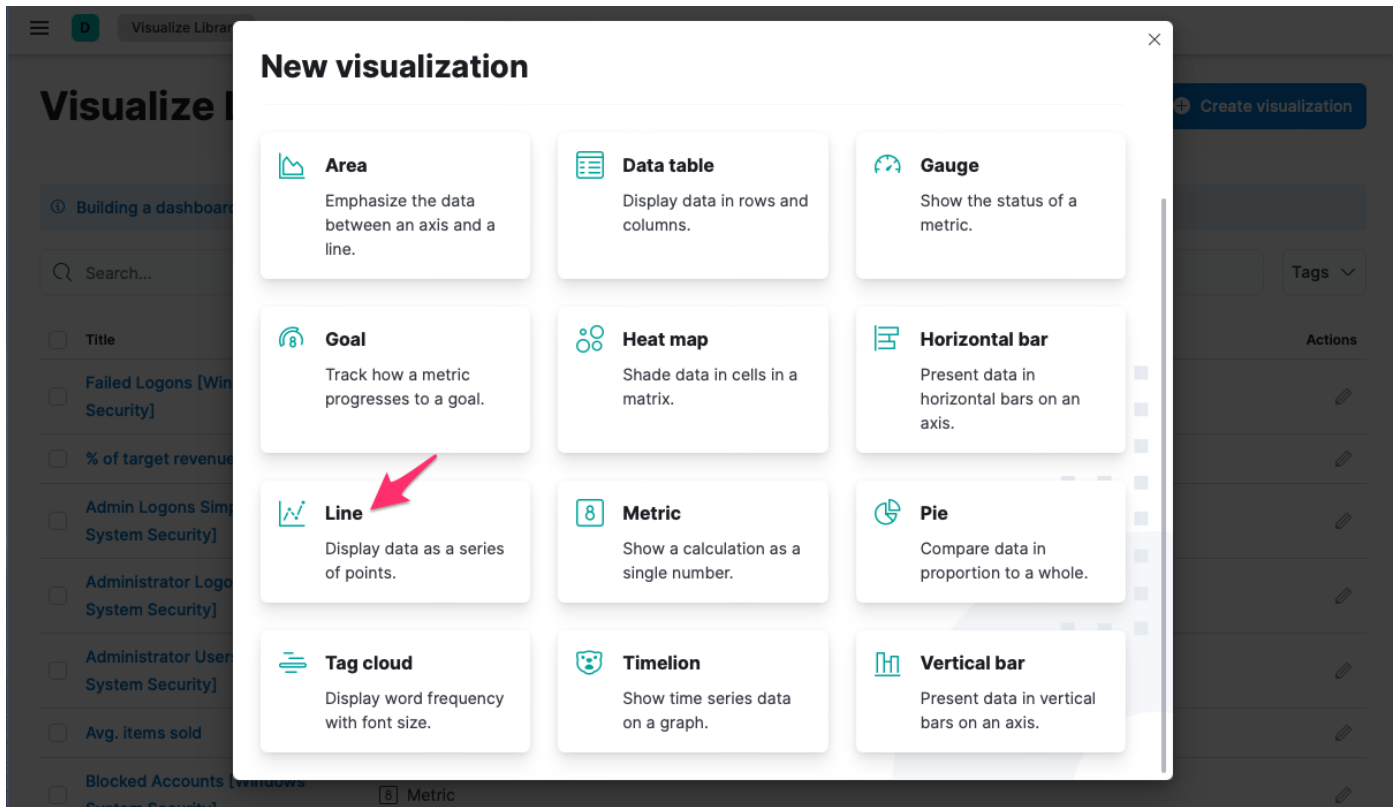
Custom visualization
Use Vega to create new types of visualizations. *Requires knowledge of Vega syntax.*

Aggregation based
Use our classic visualize library to create charts based on aggregations.
[Explore options →](#)

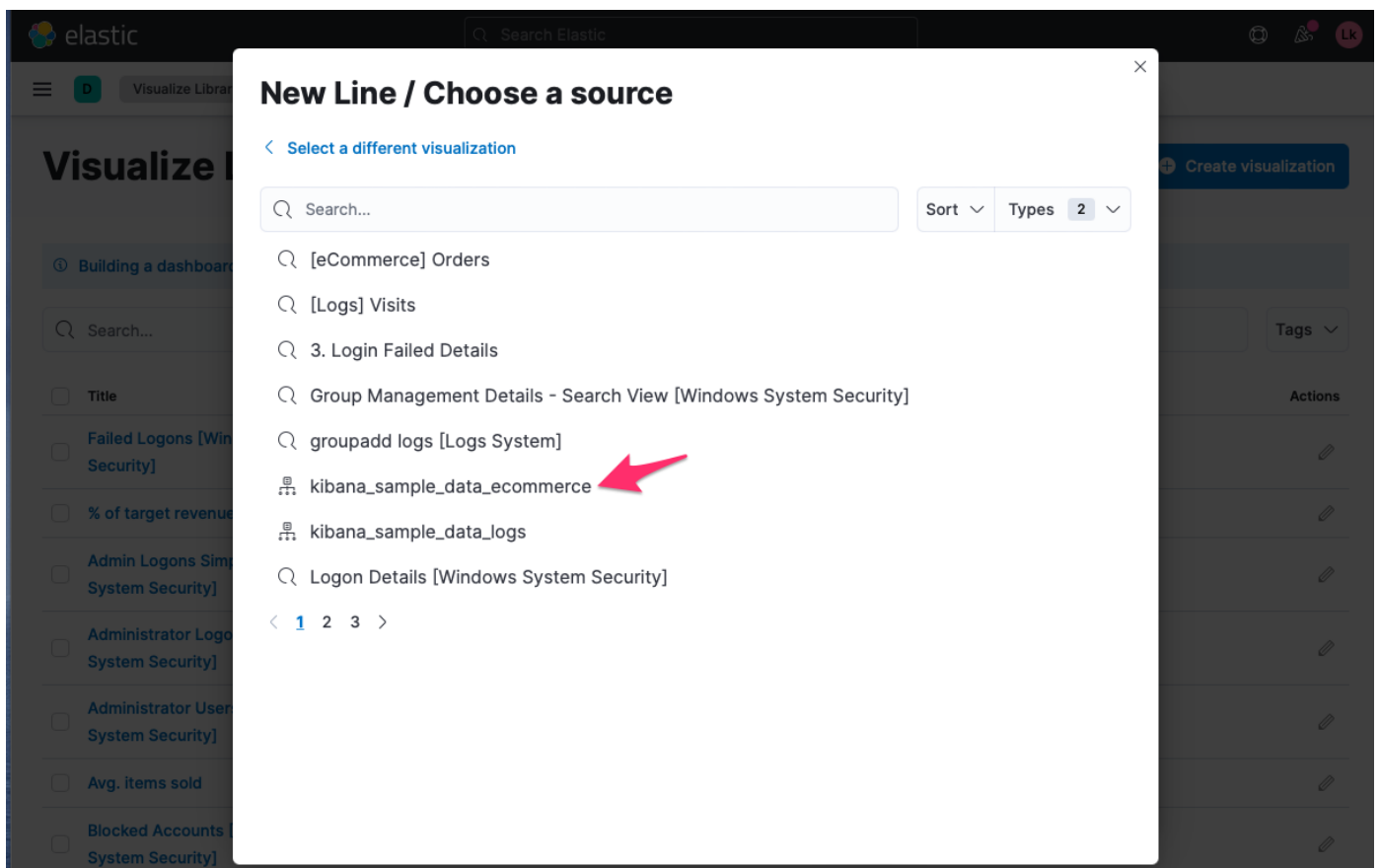
Tools
Text
Add text and images to your dashboard.
Controls
Add dropdown menus and range sliders to your dashboard.

Want to learn more? [Read documentation](#)

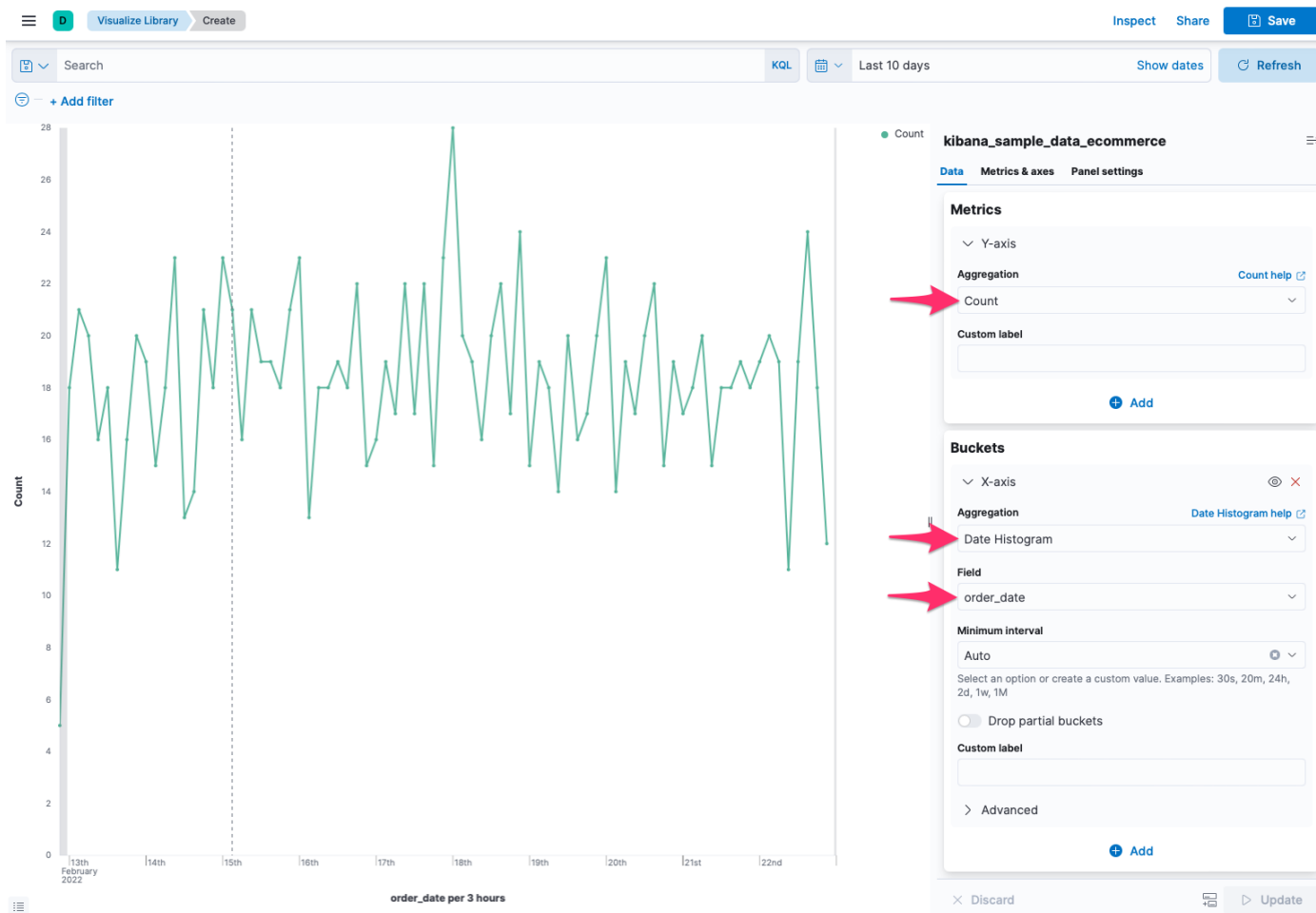
Poté zvolte libovolný typ agregace, např. **Line**:



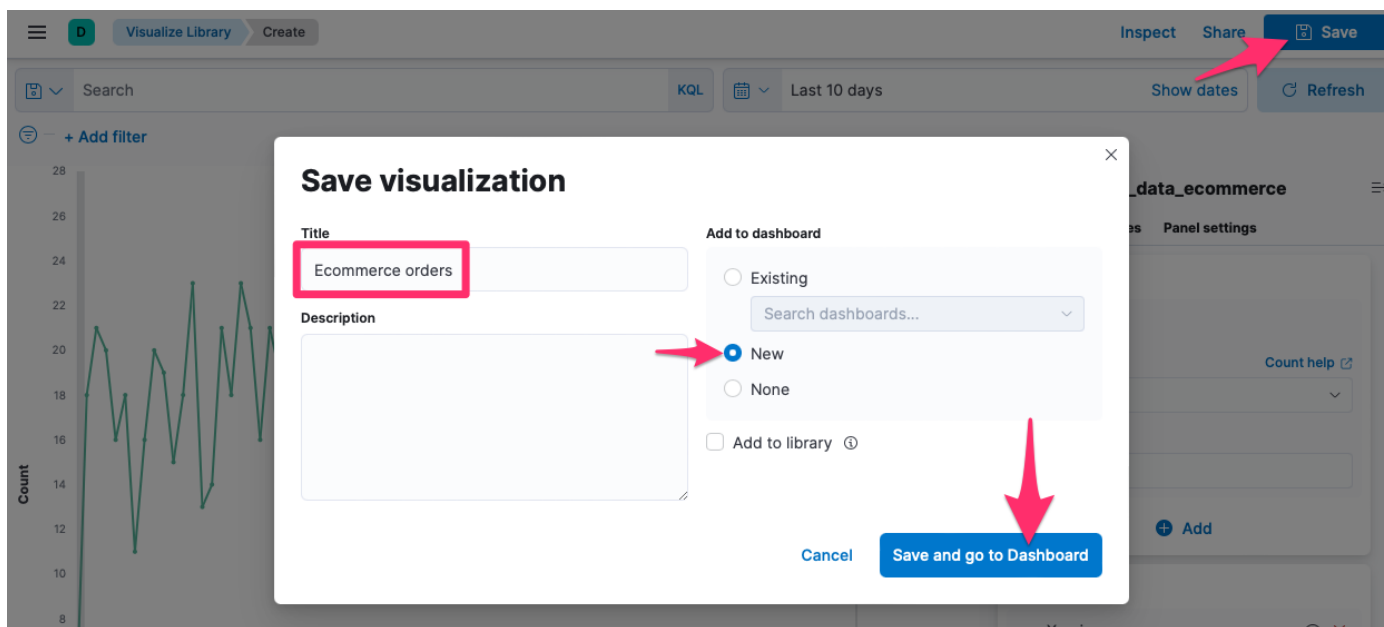
V posledním kroku zvolte data, která mají být pomocí vizualizace zobrazena - Data view, nebo uložená tabulka Discover:



V nově vytvořené vizualizaci lze zvolit metrickou agregaci, která bude reprezentována na ose Y a bucket agregaci, která se promítne na osu X.



Kompletní vizualizaci je možné uložit a případně přidat na dashboard:



Dashboard je stránka v Kibaně sestávající z více vizualizací:

