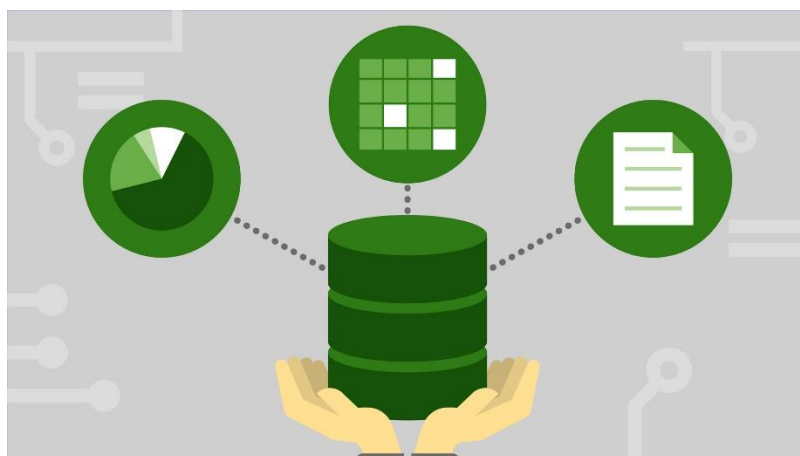


به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



آزمایشگاه پایگاه داده

دستورکار شماره ۵

مهلت تحویل :

۱۴۰۱/۰۹/۱۵

مجتبی بنائی

دستور کار شماره ۵ - کار با الستیک سرچ

یکی از دیتابیس های معروف در حوزه جستجوی متن با سرعت بالا در ذخیره انواع داده های متنی و پاسخگویی به انواع کوئری های کاربر بر روی آنها، الستیک سرچ است که در اکوسیستم استارتاپی ایران هم بسیار پرطرفدار است.

در این دستور کار هم مشابه با دستورکارهای اخیر، هدف اصلی، آشنایی اولیه با این دیتابیس و نحوه کار با آن است که حداکثر با دوساعت صرف زمان، میتوانید به راحتی آنرا انجام دهید.

گام اول: کافی است به آدرس زیر مراجعه کرده و تمامی مراحل آنرا انجام دهید :

<http://yun.ir/fi9loe>

تمام دستورات آنرا از ابتدا تا انتها در محیط کیبانا که محیط گرافیکی کار با الستیک سرچ است انجام داده، با گرفتن اسکرین شات از خروجی ده دستور آخر از این راهنمای آموزشی، گزارش خود را آپلود کنید. داده ها را طوری وارد کنید که هر کوئری حداقل دوجواب در خروجی برگرداند.

گام دوم: یک پروژه پایتون برای دریافت توئیت های کاربران سایت سهام یاب به کمک وب سوکت این سایت که در صفحه اصلی آن قابل مشاهده است (البته در بخش کنسول های توسعه وب مرورگر)، ضمیمه این دستور کار شده است.

ابتدا فایل [sahamyab.py](#) را اجرا کنید تا خروجی توئیت ها را در کنسول مشاهده کنید. خروجی ها در فایل [pipe_line_connector.py](#) تولید شده اند و شما هم تنها با همین فایل سروکار خواهید داشت.

به کمک فایل [insert2elasticsearch.py](#)، با نحوه درج اطلاعات در الستیک سرچ آشنا شده و تابع [on_item_update](#) را به گونه ای تغییر دهید که با دریافت هر توئیت از وب سوکت، آنرا در الستیک سرچ درج کند.

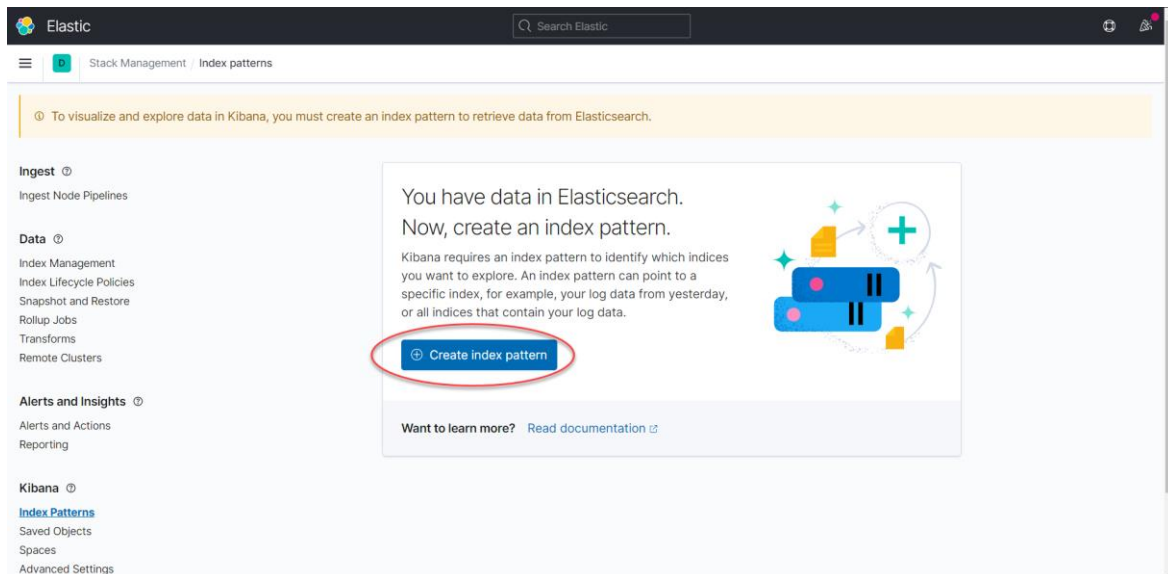
حتما هشتگ های هر توئیت را استخراج کنید و فیلدهای اصلی هر توئیت مانند زمان درج، هشتگ، محتوای توئیت، نوع توئیت (فیلد [type](#)) نام ارسال کننده را در الستیک درج کنید.

برنامه را تا دریافت حداقل ۳۰۰ توئیت اجرا کنید.

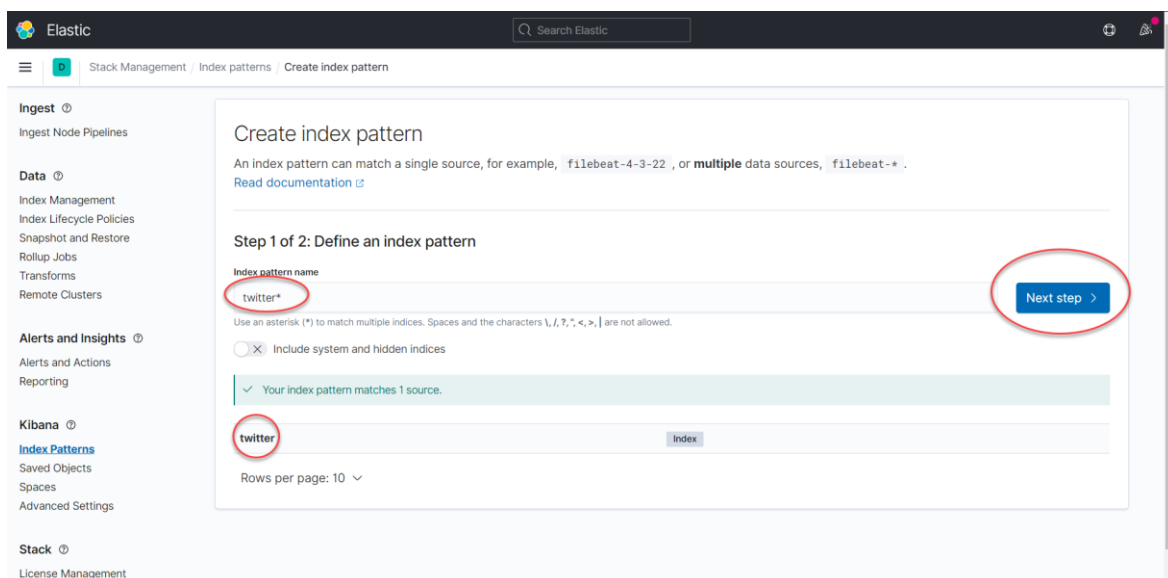
گام سوم: سپس یک داشبورد با چهار پنل ایجاد کنید. این چهار پنل می تواند مواردی مثل نمودار خطی یا ستونی تعداد توئیت ها در بازه های مشخص مثلا یکساعته، ابرکلمات برای هشتگ ها، فعالترین کاربران، توزیع انواع توئیت ها (یک نمودار ستونی یا پای چارت شامل مجموع تعداد هر نوع)، تعداد توئیت های دارای عکس در هر یک ساعت و مانند آن.

نحوه ایجاد داشبورد هم به صورت زیر است :

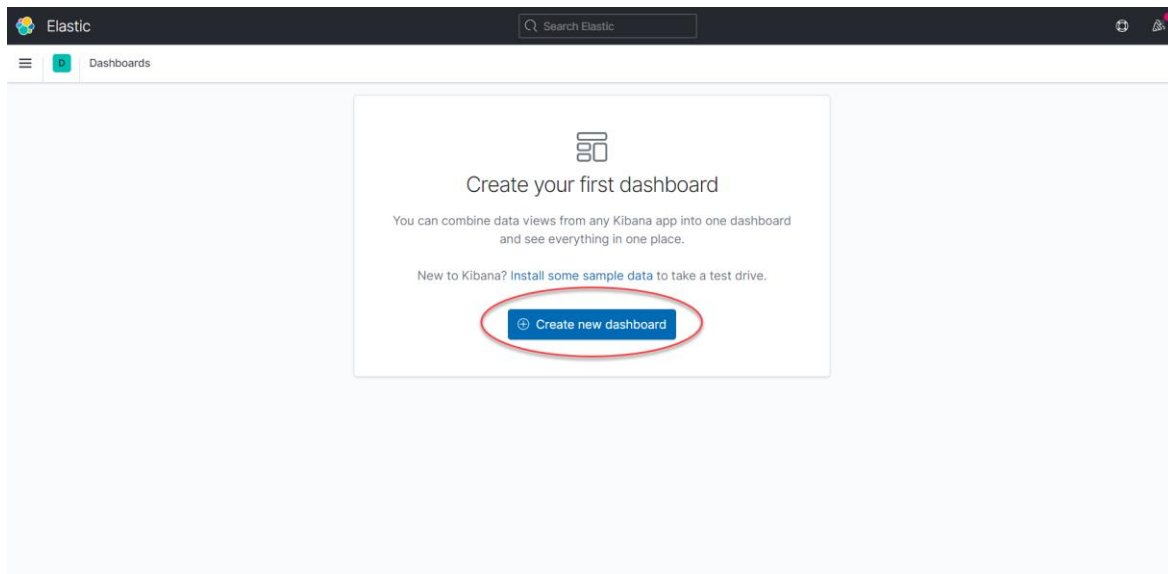
ابتدا یک [index pattern](#) می سازیم.



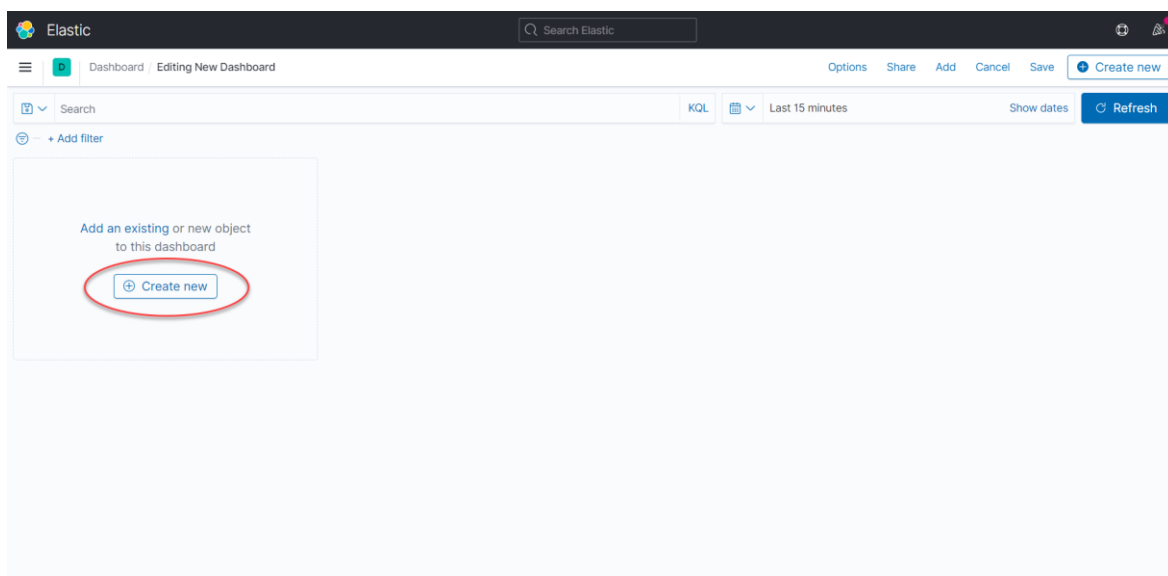
این دکس با نام **twitter** را انتخاب کرده و **next** را می زنیم.



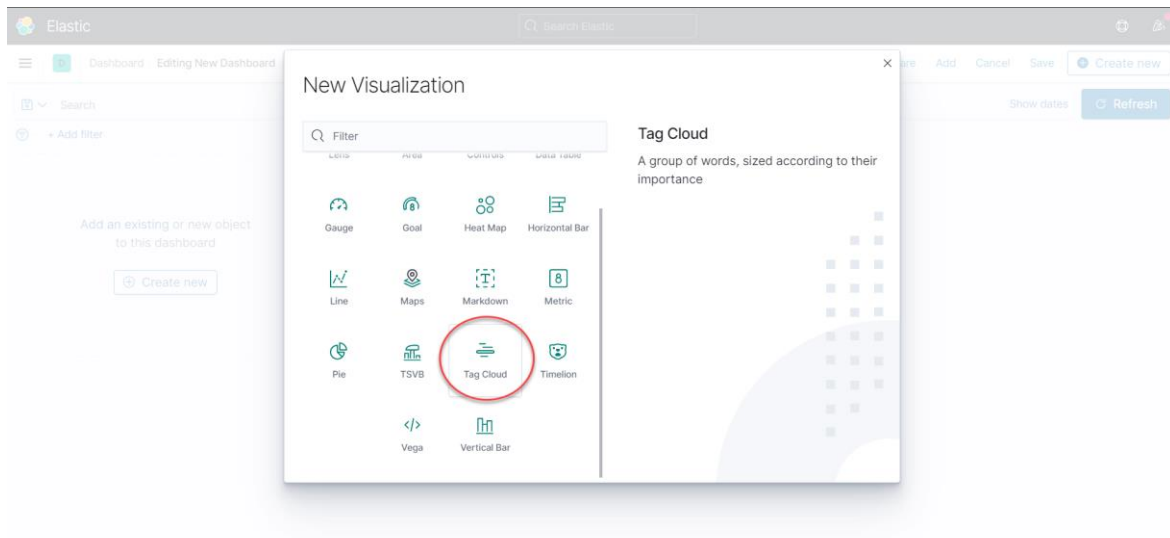
سپس یک داشبورد می سازیم.



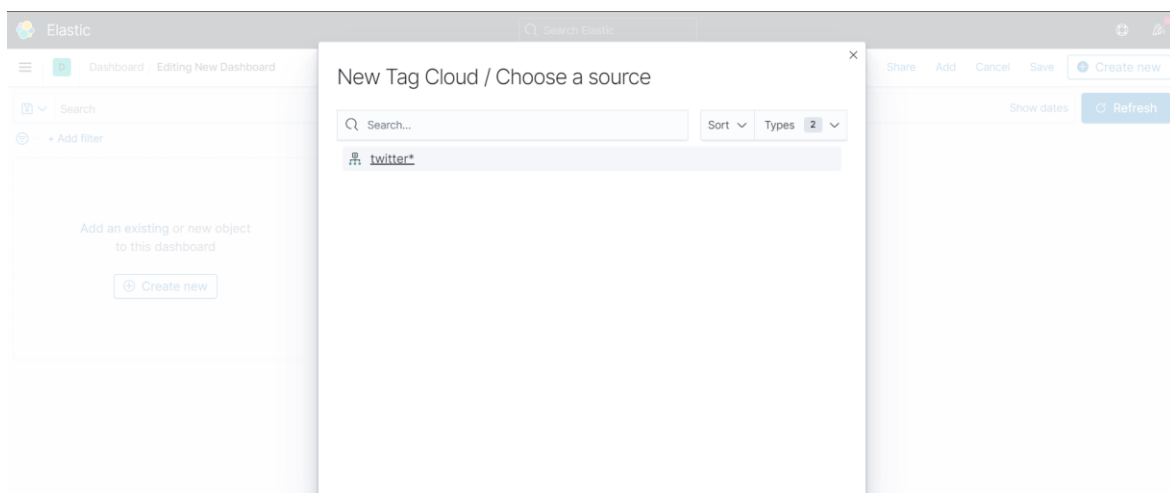
سپس **create new** را می‌زنیم تا پنل‌های مختلف را به این داشبورد اضافه کنیم.



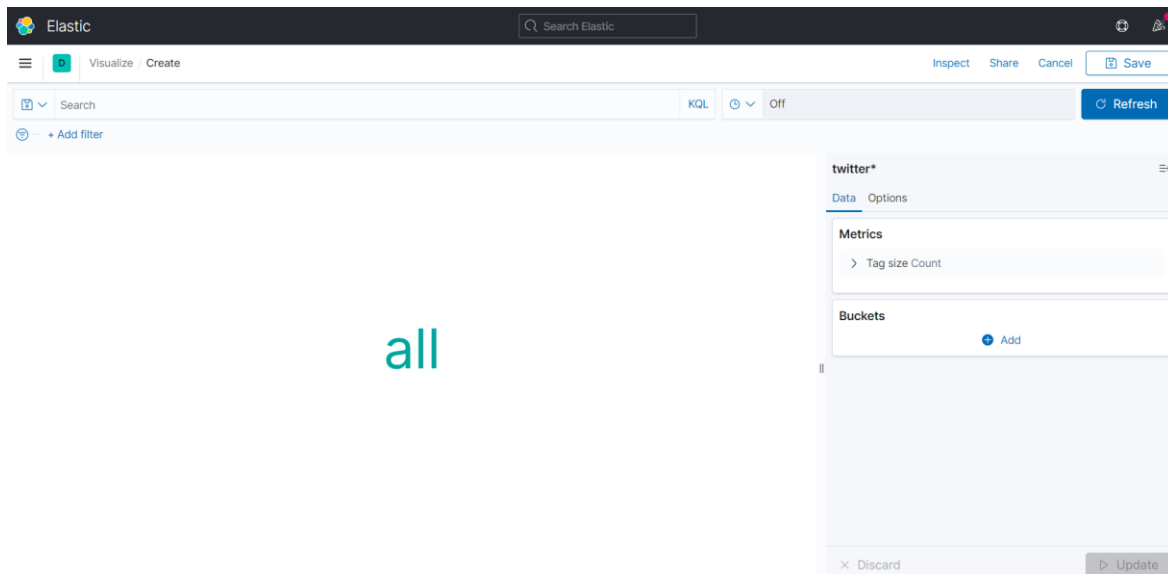
برای **visualization** اول **tag cloud** را انتخاب می‌کنیم که بسته به تعداد تکرار کلمه، سایز آن ست می‌کند.



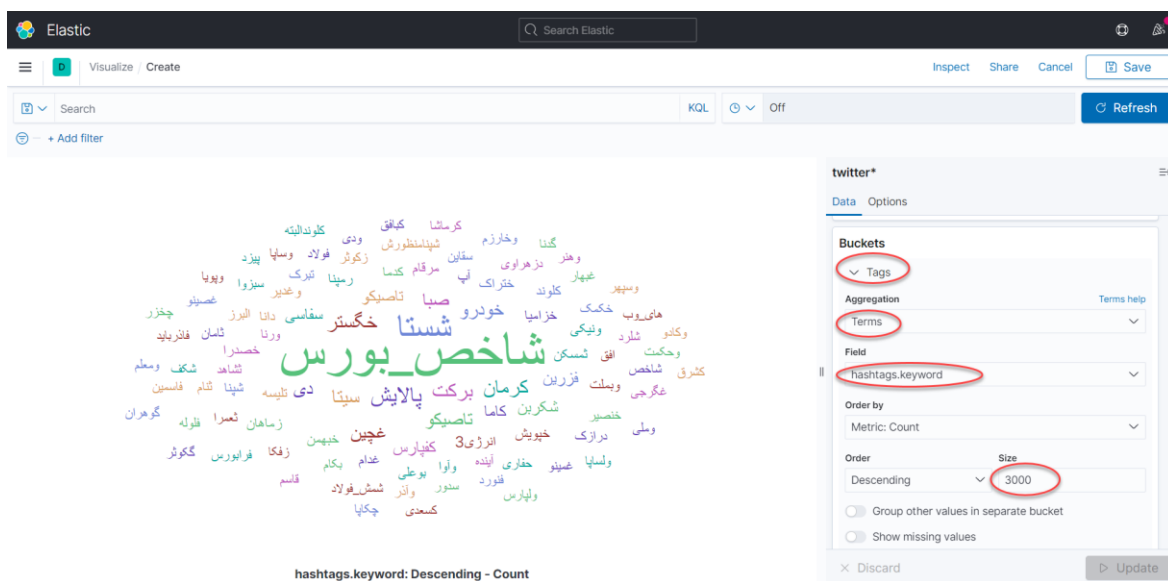
سپس **twitter** را انتخاب می کنیم.



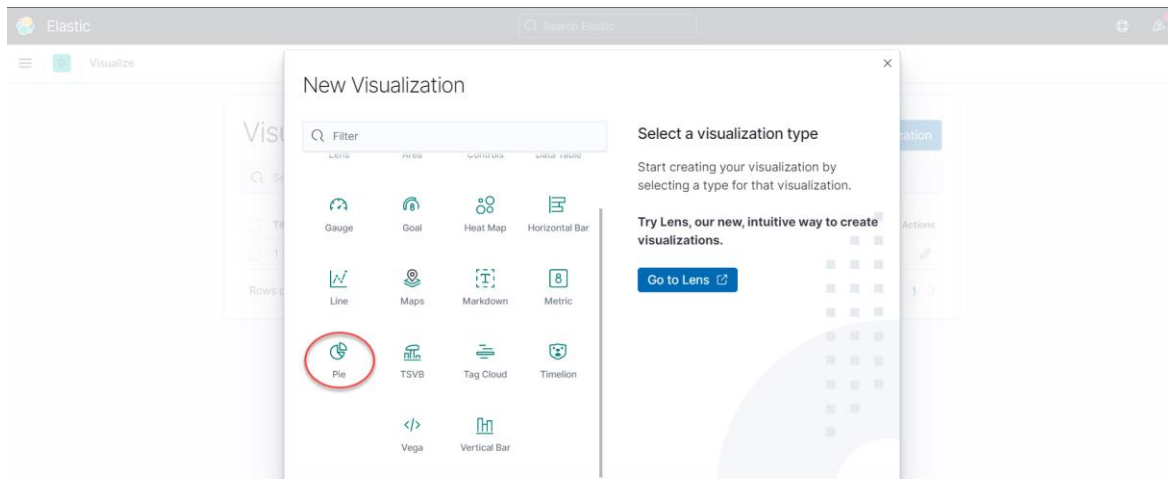
در ابتدا تنها **ali** نشان داده می شود.



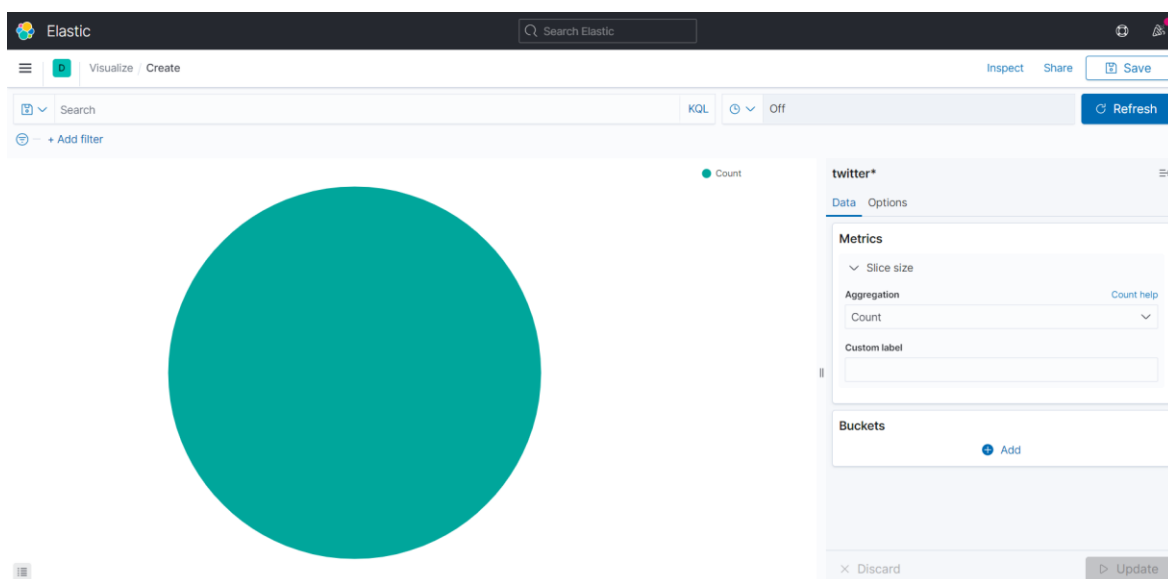
تنظیمات رو در سمت راست تغییر می دهیم تا برای hashtag ها این visualization انجام شود که خروجی به صورت زیر می شود.



برای visualization دوم، Pie را انتخاب می کنیم که سایز slice دایره بر اساس تعداد تکرار کلمه می شود.



خروجی اولیه به این صورت می شود.



در تنظیمات **bucket** مانند قبل **hashtags** رو انتخاب می کنیم و خروجی نهایی به این صورت می شود.

