Aggregations

> Official ElasticSearch Documentation can be found here

Example: Average Length of File Content

We had a case where we wanted to see what the average length of the ***fileContent*** property was. We were able to use aggregations to get statistics about the property including the ***count***, ***minimum length***, ***maximum length***, ***average length***, and ***entropy*** using the string_stats aggregation.

We used the API Console in Elastic Cloud to run the aggregation query.

> We used the _async_search API because the query took so long to run. During this time searches slowed to a crawl. We were running this at night so it wasn't a huge deal. If you have an async search that is running too long and / or causing performance issues you can stop the request by sending a DELETE call to _async_search/{async search id here}, and this is detailed further below.

Request

- Method: POST
- Path: node-index/_async_search?size=0
- Body (see below)

```
{
 "query": {
  "exists": {
   "field": "fileContent.wildcard"
  }
 },
 "aggs": {
  "message_stats": { "string_stats": { "field": "fileContent.wildcard" } }
 }
```

```
}
```

Response

Initial Response Body


```
{
  "is_partial": true,
  "is_running": true,
  "id": "FjZFS3lJTnd0UTNla29TcnFFRnR5NncfT0NycU9EcnFRYjZqRkdTRFQwdi1PZzo0NDY0MjA5NA==",
  "expiration_time_in_millis": 1716350043005,
  "response": {
    "hits": {
      "hits": [],
      "total": {
        "relation": "gte",
        "value": 0
      },
      "max_score": null
    },
    "_shards": {
      "successful": 0,
      "failed": 0,
      "skipped": 0,
      "total": 400
    },
    "took": 1031,
    "timed_out": false,
    "terminated_early": false,
    "num_reduce_phases": 0
  },
  "start_time_in_millis": 1715918043005
}
```


The request is running asynchronously, and you can monitor progress with the following request

- Method: GET
- Path: _async_search/FjZFS3lJTnd0UTNla29TcnFFRnR5NncfT0NycU9EcnFRYjZqRkdTRFQwdi1PZzo0NDY0MjA5NA==

```
{
  "is_partial": true,
  "is_running": true,
  "id": "FjZFS3lJTnd0UTNla29TcnFFRnR5NncfT0NycU9EcnFRYjZqRkdTRFQwdi1PZzo0NDY0MjA5NA==",
  "expiration_time_in_millis": 1716350043005,
  "response": {
    "hits": {
      "hits": [],
      "total": {
        "relation": "gte",
        "value": 10000
      },
      "max_score": null
    },
    "_shards": {
      "successful": 17,
      "failed": 0,
      "skipped": 0,
      "total": 400
    },
    "took": 215630,
    "timed_out": false,
    "terminated_early": false,
    "num_reduce_phases": 4,
    "aggregations": {
      "message_stats": {
        "count": 8880102,
        "min_length": 1,
        "max_length": 12175466,
        "entropy": 4.977113043941186,
        "avg_length": 12041.666635923777
      }
    }
  },
  "start_time_in_millis": 1715918043005
}
```


You can delete the search query (if it's running too long or causing other performance issues with the following request

- Method: DELETE
- Path: _async_search/FjZFS3lJTnd0UTNla29TcnFFRnR5NncfT0NycU9EcnFRYjZqRkdTRFQwdi1PZzo0NDY0MjA5NA==

Revision #1
Created 20 May 2024 22:13:38 by Quinn Godfrey
Updated 30 June 2024 13:44:18 by Quinn Godfrey