



BIG DATA

TOO BIG TO IGNORE

SÜMEYYE KAYNAK

SAKARY

OUTLINE

ElasticSearch

ELASTICSEARCH

- It is the Nosql technology used in text search operations in big data.
- Elasticsearch infrastructure includes Apache Lucene and Solr libraries.
- Apache Lucene and Solr libraries are open source.
- In the Elasticsearch module, data is indexed while it is being saved.

ELASTICSEARCH

■ In the word search, not all data is searched. Results can be found quickly through the created index list.



Implementing Elastic as a fundamental core to a drug-manufacturing data cloud



Customizing search results across multiple domains to make their students' lives easier



Making search smarter with machine learning at scale

COMPANIES USING ELASTICSEARCH



Monitoring application infrastructure across a major financial institution



Reducing system
downtime with
Elasticsearch at the basis
of Cisco's Cloud Native
platform



Creating a custom
Elasticsearch as a Service
platform



Creating a real-time search solution across over 4 million customer records



Delivering a better help experience for over a billion users



Driving better research, analysis, and journalism

1	The old n	ight <mark>kee</mark>	per the	keep in	the town
•		Sile itee	PCI CITE	ixeep iii	CITC COTTI

- In the big old house in the big old gown.
- The house in the town had the big old keep.
- 4 Where the old night keeper never did sleep.
- 5 The night keeper keeps the keep in the night.

And keeps in the dark and sleeps in the light.

6 document to index

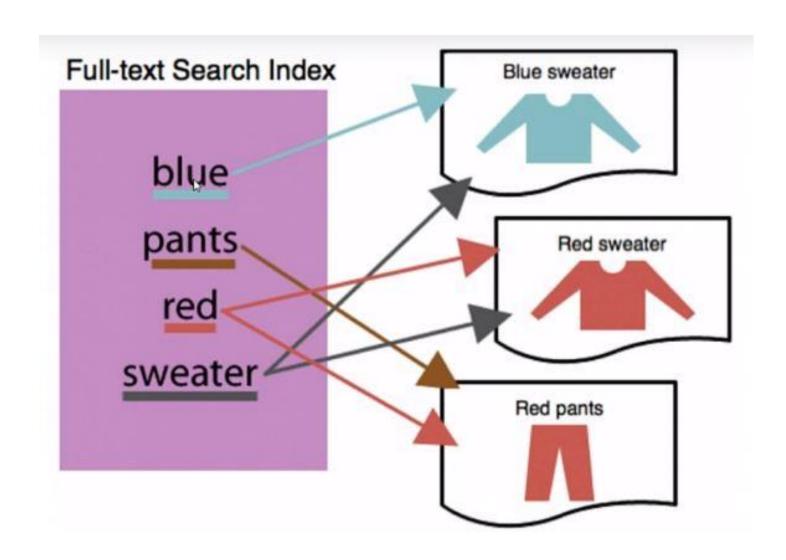
Sleeps	<6>
The	<1> <2> <3> <4> <5> <6>

Documents

Where <4>

Term

Term	Documents
And	<6>
Big	<2> <3>
Dark	<6>
Did	<4>
Gown	<2>
Had	<3>
House	<2> <3>
In	<1> <2> <3> <5> <6>
Keep	< > <3> <5>
Keeper	<1> <4> <5>
Keeps	<1> <5> <6>
Light	<6>
Never	<4>
Night	<1> <4> <5>
Old	<1> <2> <3> <4>
Sleep	<4>



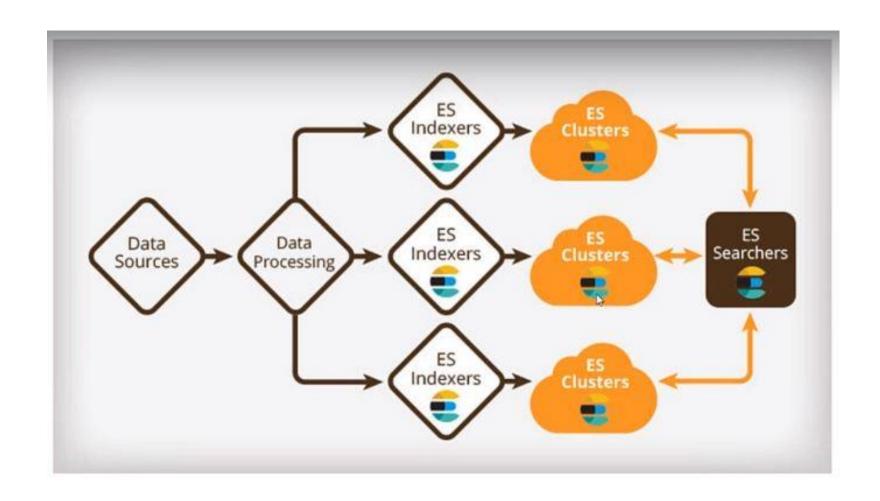
TERMINOLOGY

Relation Databases

- Database
- Table
- Row
- Column
- Schema

Elasticsearch

- Index
- Type
- Document
- Fields
- Mapping



DOCUMENT AS JSON

```
"id" : "abc123",
"title" : "A JSON Document",
"body" : "A JSON document is a ...",
"published on" : "2013/06/27 10:00:00",
"featured" : true,
"tags" : ["search", "json"],
"author" : {
    "first_name" : "Clara",
    "last_name" : "Rice",
    "email" : "clara@rice.org"
}
```

Since Elasticsearch is a document-type Nosql database, documents are stored in json format.

GENERAL FEATURES

- Elasticsearch is scalable up to petabytes of structured and unstructured data.
- Elasticsearch uses denormalization to improve the search performance.
- Elasticsearch is one of the popular enterprise search engines, and is currently being used by many big organizations like Wikipedia, The Guardian, StackOverflow, GitHub etc.
- Elasticsearch is an open source and available under the Apache license version 2.0

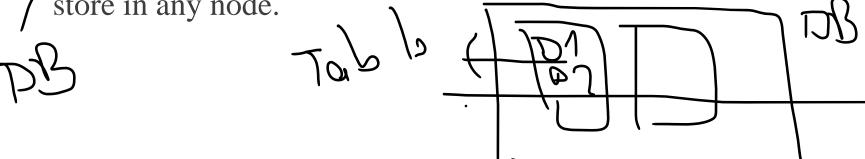
KEY CONCEPTS

- **Node:** It refers to a single running instance of Elasticsearch.
- Cluster: It is a collection of one or more nodes. Cluster provides collective indexing and search capabilities across all the nodes for entire data.
- Index: It is a collection of different type of documents and their properties. Index also uses the concept of shards to improve the performance.

KEY CONCEPTS

- **Document**: It is a collection of fields in a specific manner defined in JSON format. Every document belongs to a type and resides inside an index. Every document is associated with a unique identifier called the UID.
- **Shard**: Indexes are horizontally subdivided into shards. This means each shard contains all the properties of document but contains less number of JSON objects than index. The horizontal separation makes shard an independent node, which can be

store in any node.



KEY CONCEPTS

■ **Replicas**: Elasticsearch allows a user to create replicas of their indexes and shards. Replication not only helps in increasing the availability of data in case of failure, but also improves the performance of searching by carrying out a parallel search operation in these replicas.

ADVANTAGES

- Elasticsearch is developed on Java, which makes it compatible on almost every platform.
- Elasticsearch is real time, in other words after one second the added document is searchable in this engine.
- Elasticsearch is distributed, which makes it easy to scale and integrate in any big organization.
- Creating full backups are easy by using the concept of gateway, which is present in Elasticsearch.

ADVANTAGES

- Handling multi-tenancy is very easy in Elasticsearch when compared to Apache Solr.
- Elasticsearch uses JSON objects as responses, which makes it possible to invoke the Elasticsearch server with a large number of different programming languages.

DISADVANTAGES

■ Elasticsearch does not have multi-language support in terms of handling request and response data (only possible in JSON) unlike in Apache Solr, where it is possible in CSV, XML and JSON formats.

ELASTIC SEARCH-INSTALLATION

For Windows OS;

- Download Elasticsearch from https://www.elastic.co/downloads/elasticsearch
- The downloaded file is copied to the C:

ELASTIC SEARCH SERVER

To run the Elastic Search server;

```
Microsoft Windows [Version 10.0.18363.1556]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\Sumeyye>cd ..

C:\Users>cd ..

C:\>cd elasticsearch-7.15.0

C:\elasticsearch-7.15.0>cd bin

C:\elasticsearch-7.15.0\bin>elasticsearch.bat
```

CRUD OPERATION

- In client/server architecture, a request-response communication is performed between client-server.
- Http methods determine the purpose of the request made to the server side.

HTTP METHODS

Common request types are:

- GET
- POST
- HEAD
- OPTIONS
- TRACE
- PUT
- DELETE
- CONNECT

Method	Description
GET	Request to read a Web page
HEAD	Request to read a Web page's header
PUT	Request to store a Web page
POST	Append to a named resource (e.g. a Web page)
DELETE	Remove the web page
TRACE	Echo the incoming request
CONNECT	Reserved for future use
OPTIONS	Query certain options

GET AND POST METHODS

- **Get method:** GET is used to request data from a specified resource.
- Post method: POST is used to send data to a server to create/update a resource.

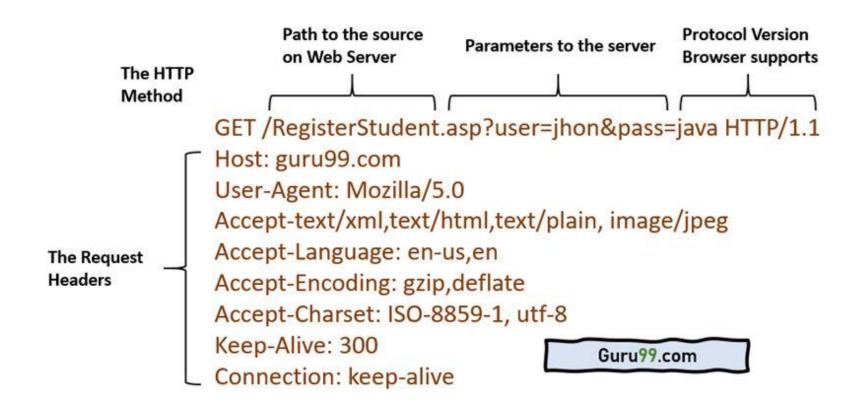
GET METHOD

- Get requests can be cached.
- In GET method, values are visible in the URL.
- Get requests remain in the browser history.
- Get requests can be bookmarked.
- Get requests should never be used when dealing with sensitive data
- GET requests have length restrictions
- GET requests are only used to request data (not modify)

/test/demo_form.php?name1=value1&name2=value2

GET METHOD

GET/RegisterStudent.asp?user=value1&pass=value2



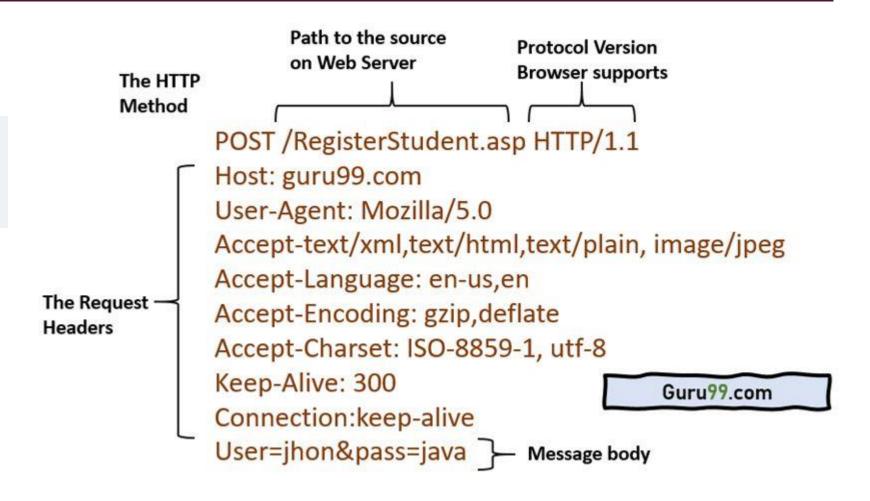
POST METHOD

- POST requests are never cached
- In POST method, values are not visible in the URL.
- POST requests do not remain in the browser history
- POST requests cannot be bookmarked
- POST requests have no restrictions on data length

/test/demo_form.php

POST METHOD

POST/RegisterStudent.asp HTTP/1.1 Host: www.guru99.com user=value1&pass=value2



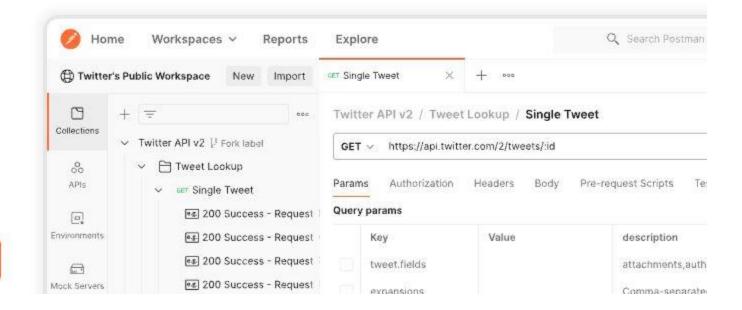
Download Postman

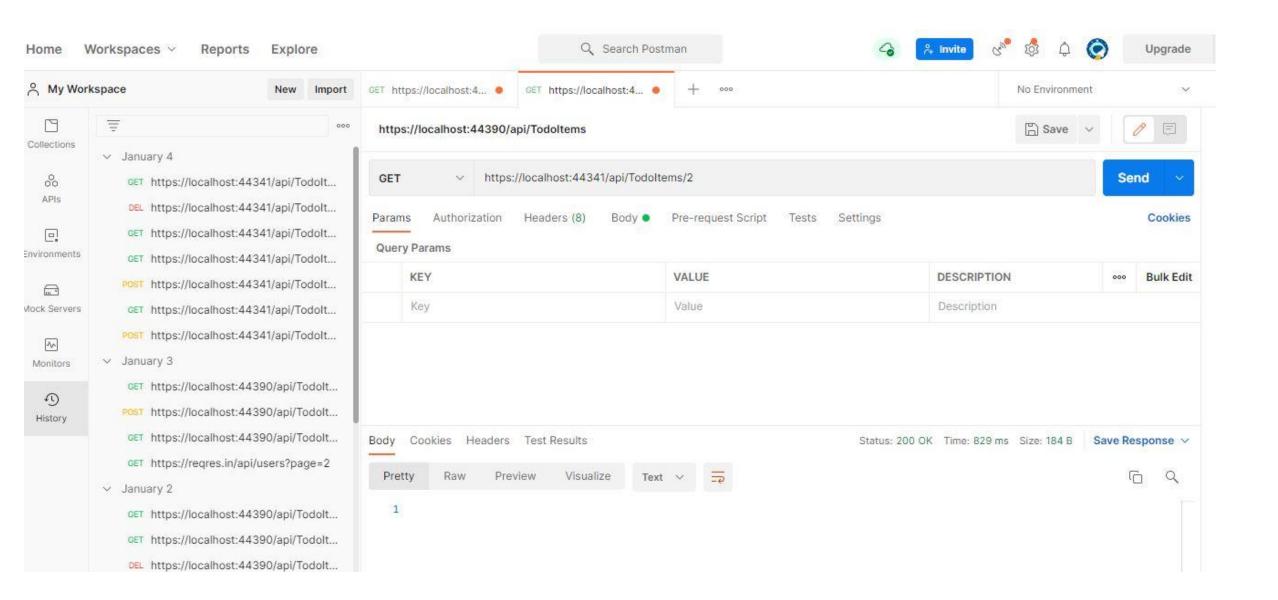
Download the app to quickly get started using the Postman API Platform. Or, if you prefer a browser experience, you can try the new web version of Postman.

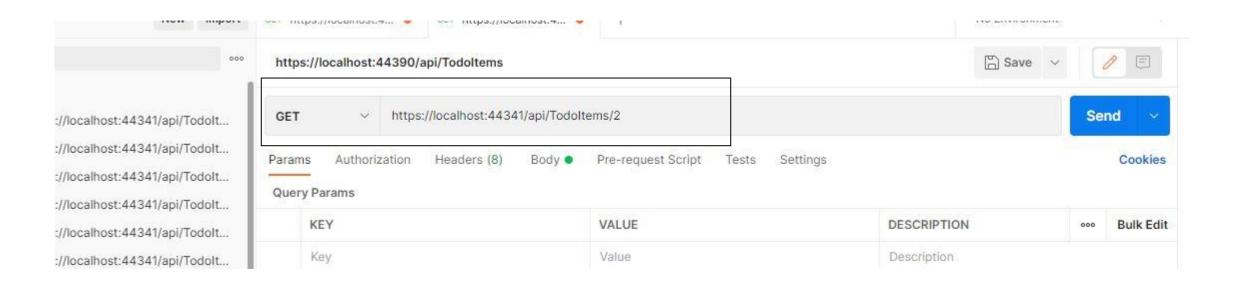
The Postman app

The ever-improving Postman app (a new release every two weeks) gives you a full-featured Postman experience.

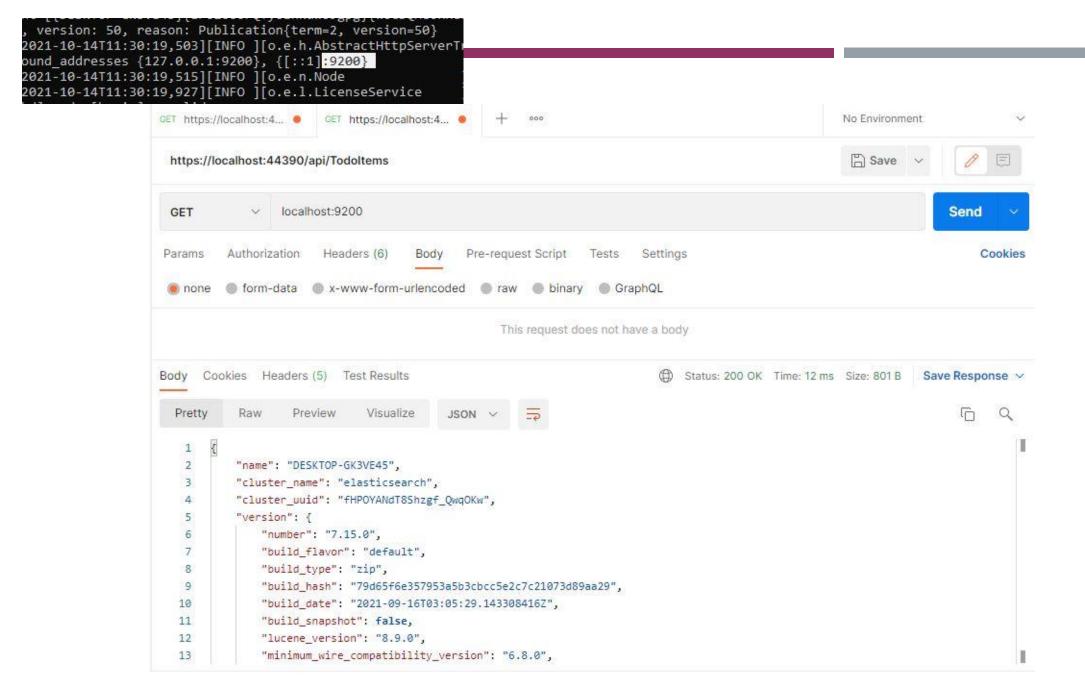
■ Download the App







```
Command Prompt - elasticsearch.bat
Microsoft Windows [Version 10.0.18363.1556]
(c) 2019 Microsoft Corporation. All rights reserved.
C:\Users\Sumeyye>cd ..
C:\Users>cd ...
C:\>cd elasticsearch-7.15.0
C:\elasticsearch-7.15.0>cd bin
C:\elasticsearch-7.15.0\bin>elasticsearch.bat
[2021-10-14T11:29:36,238][INFO ][o.e.n.Node
                                                         [ DESKTOP-GK3VE45] version[7.15.0], pid[18184], build[defau
/zip/79d65f6e357953a5b3cbcc5e2c7c21073d89aa29/2021-09-16T03:05:29.143308416Z], OS[Windows 10/10.0/amd64], JVM[Eclipse
undation/OpenJDK 64-Bit Server VM/16.0.2/16.0.2+7]
[2021-10-14T11:29:36,335][INFO ][o.e.n.Node
                                                         [DESKTOP-GK3VE45] JVM home [C:\elasticsearch-7.15.0\jdk],
ing bundled JDK [true]
[2021-10-14T11:29:36,350][INFO ][o.e.n.Node
                                                         [ DESKTOP-GK3VE45] JVM arguments [-Des.networkaddress.cache
tl=60, -Des.networkaddress.cache.negative.ttl=10, -XX:+AlwaysPreTouch, -Xss1m, -Djava.awt.headless=true, -Dfile.encodi
=UTF-8, -Djna.nosys=true, -XX:-OmitStackTraceInFastThrow, -XX:+ShowCodeDetailsInExceptionMessages, -Dio.netty.noUnsafe
rue, -Dio.netty.noKeySetOptimization=true, -Dio.netty.recycler.maxCapacityPerThread=0, -Dio.netty.allocator.numDirectA
nas=0, -Dlog4j.shutdownHookEnabled=false, -Dlog4j2.disable.jmx=true, -Djava.locale.providers=SPI,COMPAT, --add-opens=j
a.base/java.io=ALL-UNNAMED, -XX:+UseG1GC, -Djava.io.tmpdir=C:\Users\Sumeyye\AppData\Local\Temp\elasticsearch, -XX:+Hea
umpOnOutOfMemoryError, -XX:HeapDumpPath=data, -XX:ErrorFile=logs/hs err pid%p.log, -Xlog:gc*,gc+age=trace,safepoint:fi
```

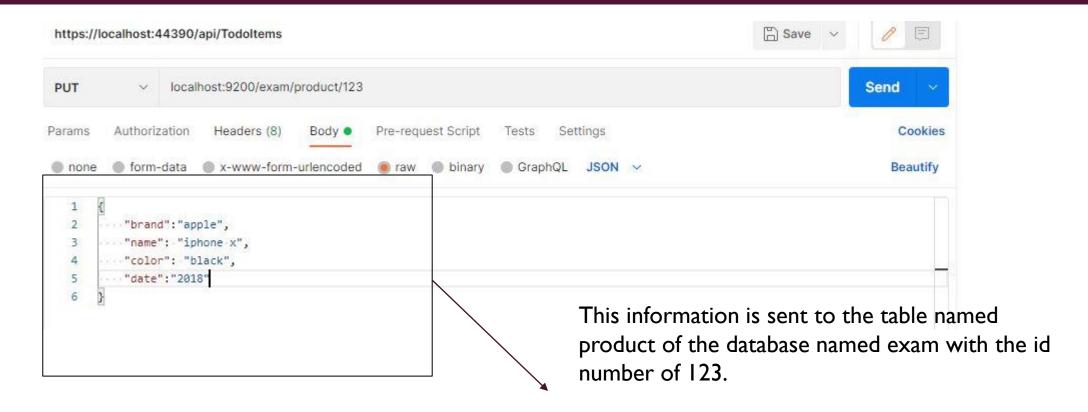


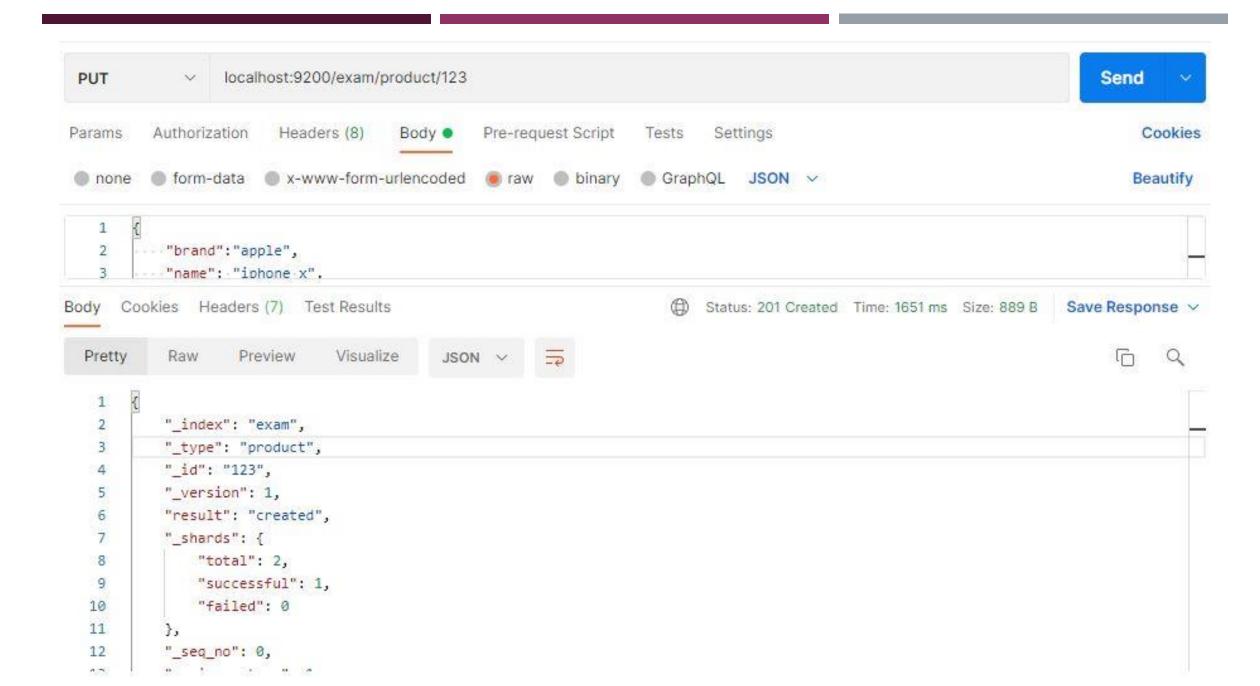
MAPPING CONCEPTS ACROSS SQL AND ELASTIC SEARCH

SQL	ELASTIC SEARCH
column	field
row	document
table	index
schema	implicit

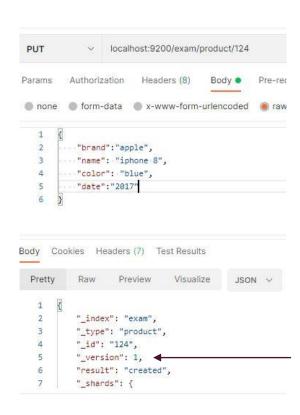
```
(i) localhost:9200
Uygulamalar G Google 🛊 Bookmarks PQ Arama Sonuçları - P...
 "name": "DESKTOP-GK3VE45",
 "cluster name" : "elasticsearch",
 "cluster uuid" : "fHPOYANdT8Shzgf QwqOKw",
 "version" : {
   "number" : "7.15.0",
   "build_flavor" : "default",
   "build type" : "zip",
   "build hash": "79d65f6e357953a5b3cbcc5e2c7c21073d89aa29",
   "build date" : "2021-09-16T03:05:29.143308416Z",
   "build snapshot" : false,
   "lucene version": "8.9.0",
   "minimum wire compatibility version" : "6.8.0",
   "minimum_index_compatibility_version" : "6.0.0-beta1"
 "tagline" : "You Know, for Search"
```

PUT METHOD

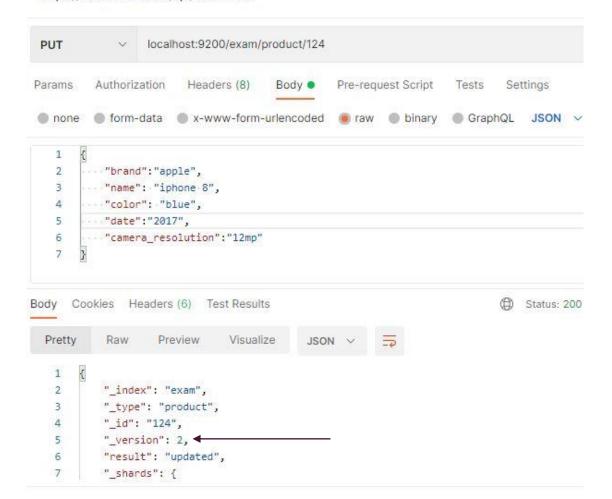




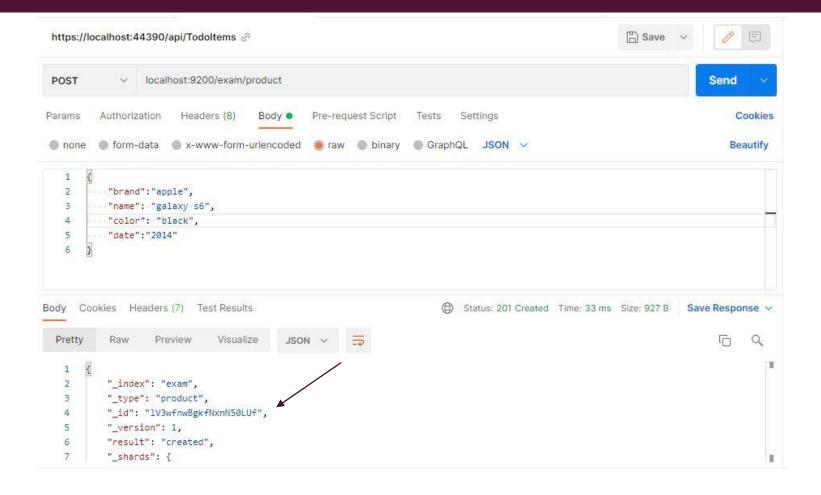
UPDATE WITH POSTMAN



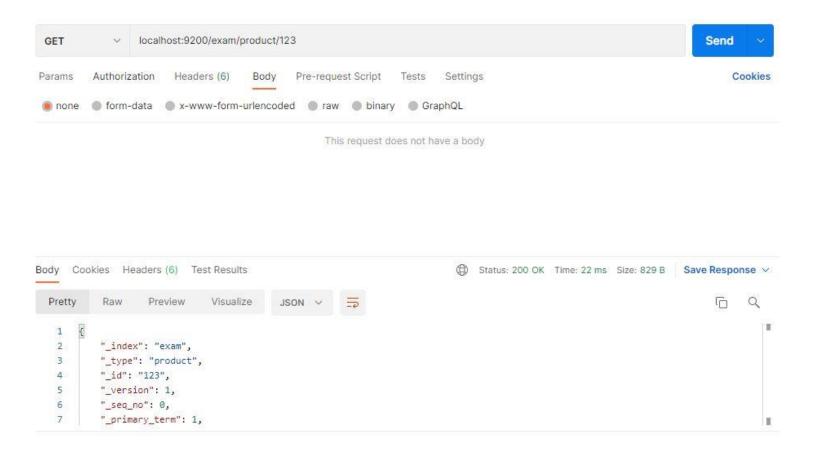
https://localhost:44390/api/TodoItems



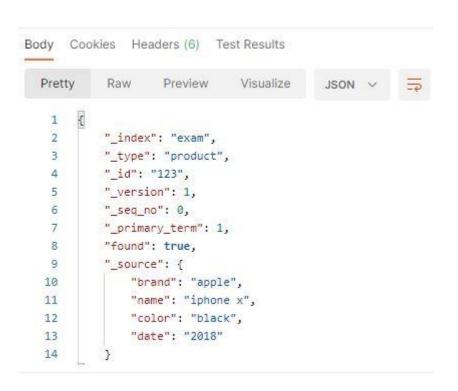
POST METHOD

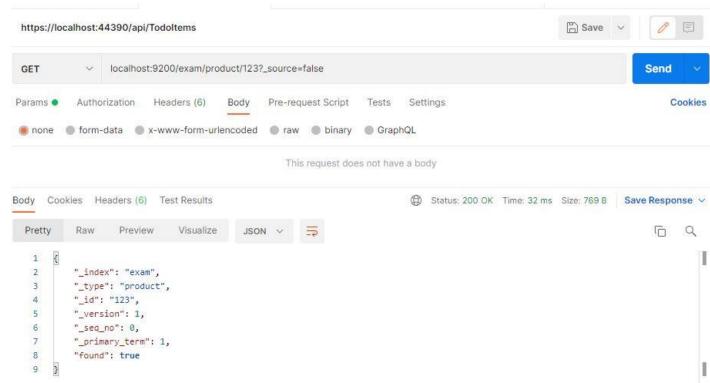


QUERY WITH POSTMAN

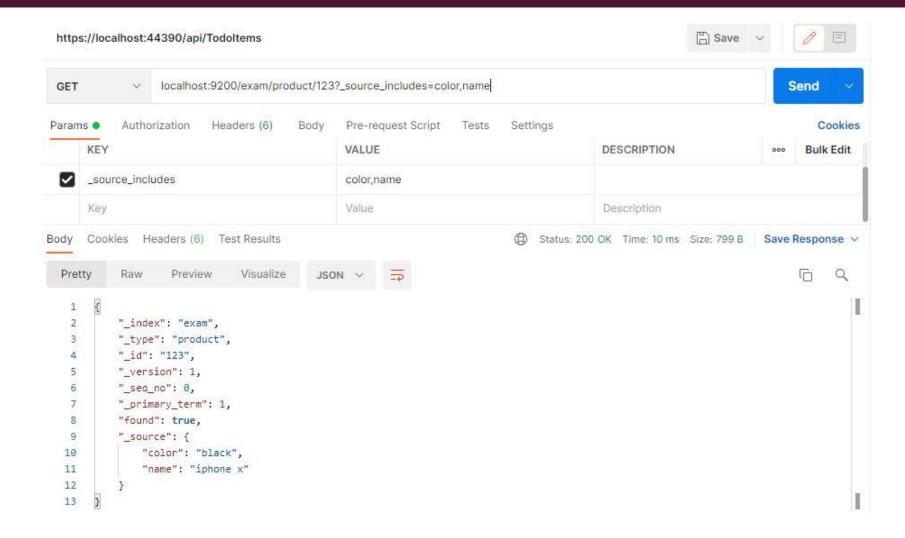


QUERY WITH POSTMAN

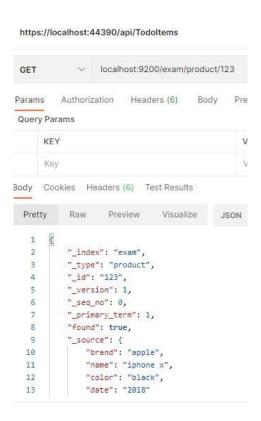


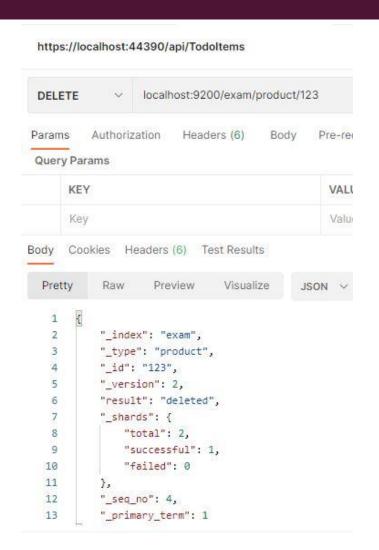


QUERY WITH POSTMAN



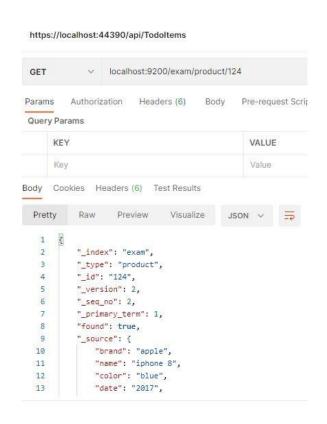
DELETE COMMAND WITH POSTMAN

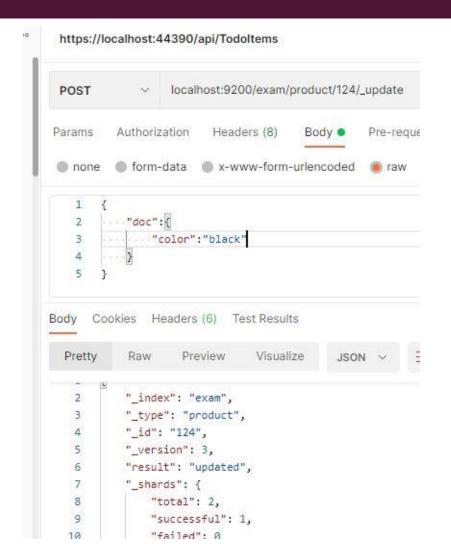




https://localhost:44390/api/Todoltems localhost:9200/exam/product/123 GET Params Authorization Headers (6) Body Pre-re Query Params KEY VAL Valu Key Body Cookies Headers (6) Test Results Pretty Raw Preview Visualize JSON V " index": "exam", "_type": "product", "_id": "123", "found": false 6

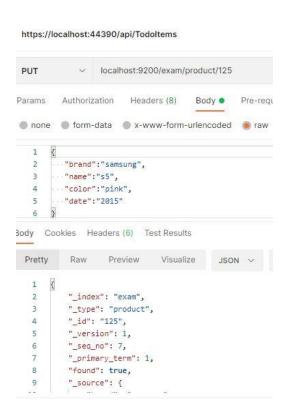
UPDATE COMMAND WITH POSTMAN





https://localhost:44390/api/Todoltems localhost:9200/exam/product/124 GET Params Authorization Headers (6) Body x-www-form-urlencoded form-data Body Headers (6) Test Results Pretty Raw Preview Visualize JSOI "_seq_no": 6, "_primary_term": 1, "found": true, " source": { 9 10 "brand": "apple", 11 "name": "iphone 8", 12 "color": "black", 13 "date": "2017", 14 "camera_resolution": "12mp" 15

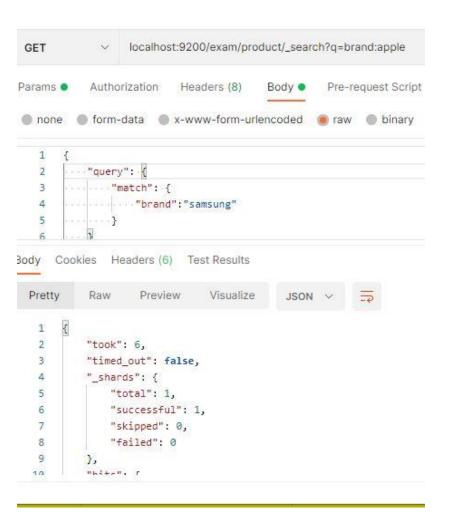
DELETE COMMAND WITH POSTMAN





```
localhost:9200/exam/product/_delete_by_query
POST
                                               Pre-request 5
Params:
         Authorization
                       Headers (8)
                                     Body .
                      x-www-form-urlencoded
         form-data
                                                raw
           "query": {
        --- "match": {
        "brand": "samsung"
        Cookies Headers (6) Test Results
 Pretty
                   Preview
                              Visualize
                                          JSON V
                                                      -2
           "took": 261,
          "timed out": false,
           "total": 1,
          "deleted": 1,
           "batches": 1,
          "version conflicts": 0,
           "noops": 0,
   9
           "retries": {
  10
              "hulle". O
```

SEARCH COMMAND WITH POSTMAN



```
10
         "hits": {
10
11
              "total": {
                 "value": 2,
12
13
                 "relation": "eg"
14
15
             "max score": 0.13353139,
             "hits": [
16
17
18
                     "_index": "exam",
Pretty
           Raw
                   Preview
                               Visualize
                                            JSON V
              "max score": 0.13353139,
 15
 16
              "hits": [
 17
 18
                      " index": "exam",
 19
                      "_type": "product",
 20
                      "_id": "1V3wfnwBgkfNxnN50LUf",
 21
                       "_score": 0.13353139,
 22
                      " source": {
                          "brand": "apple",
 23
```

```
22
                     " source": {
23
                        "brand": "apple",
24
                        "name": "galaxy s6",
25
                        "color": "black",
                        "date": "2014"
26
27
28
29
30
                     " index": "exam",
24
                     Raw
                 Preview
                             Visualize
Pretty
31
                     "_type": "product",
32
                     " id": "124",
33
                     " score": 0.13353139,
34
                     " source": {
35
                        "brand": "apple",
```

"name": "iphone 8",

"camera_resolution": "12mp"

"color": "black",

"date": "2017".

Visualize

JSON V

Preview

Pretty

36

37

38

39

Raw

GET ALL INFORMATION IN THE DATABASE



ELASTIC SEARCH ASP.NET INTEGRATION

You can install NEST from the package manager console.

PM> Install-Package NEST

• Alternatively, simply search for NEST in the package manager UI.

CONNECTION

Connecting to a single node

```
var node = new Uri("http://localhost:9200");
var settings = new ConnectionSettings(node);
var client = new ElasticClient(settings);
```

Connecting to multiple nodes using a connection pool

```
var nodes = new Uri[]
{
          new Uri("http://myserver1:9200"),
          new Uri("http://myserver2:9200"),
          new Uri("http://myserver3:9200")
};

var pool = new StaticConnectionPool(nodes);
var settings = new ConnectionSettings(pool);
var client = new ElasticClient(settings);
```

INDEXING

```
var tweet = new Tweet
{
    Id = 2,
    User = "kimchy",
    PostDate = new DateTime(2009, 11, 15),
    Message = "Trying out NEST, so far so good?"
};

var response1 = client.Index(tweet, idx => idx.Index("mytweetindex"));
```

All the calls have async variants:

```
var response = await client.IndexAsync(tweet, idx => idx.Index("mytweetindex"));
// awaits a Task<IndexResponse>
```

GETTING A DOCUMENT

```
var response2 = await client.GetAsync<Tweet>(2, idx => idx.Index("mytweetindex"));
var tweetData = response2.Source; // the original document
                    public async Task<IActionResult> Index()
                       var node = new Uri("http://localhost:9200");
                       var settings = new ConnectionSettings(node);
                       var client = new ElasticClient(settings);
                       var tweet = new Tweet
                           Id = 2
                           User = "kimchy",
                           PostDate = new DateTime(2009, 11, 15),
                           Message = "Trying out NEST, so far so good?"
                       };
                       var response1 = client.Index(tweet, idx => idx.Index("mytweetindex"));
                       var response2 = await client.GetAsync<Tweet>(2, idx => idx.Index("mytweetindex"));
                       var tweetData = response2.Source; // the original document
                        return View();
```

SEARCH

```
public IActionResult Search()
   var node = new Uri("http://localhost:9200");
    var settings = new ConnectionSettings(node);
    var client = new ElasticClient(settings);
   var response = client.Search<Tweet>(s => s
.Index("mytweetindex") //or specify index via settings.DefaultIndex("mytweetindex");
.From(0)
.Size(10)
.Query(q \Rightarrow q)
.Term(t => t.User, "kimchy") | q
.Match(mq => mq.Field(f => f.User).Query("nest"))
   return View();
```

UPDATE

```
public IActionResult Update()
{
    var node = new Uri("http://localhost:9200");
    var settings = new ConnectionSettings(node);
    var client = new ElasticClient(settings);

    var response = client.Update<Tweet>(2, x => x.Index("mytweetindex").Doc(new Tweet { Message = "güncellendi" }));
    return View();
}
```

DELETE

```
public IActionResult Delete()

var node = new Uri("http://localhost:9200");
var settings = new ConnectionSettings(node);
var client = new ElasticClient(settings);

var response = client.Delete<Tweet>(2, x => x.Index("mytweetindex"));
return View();
}
```