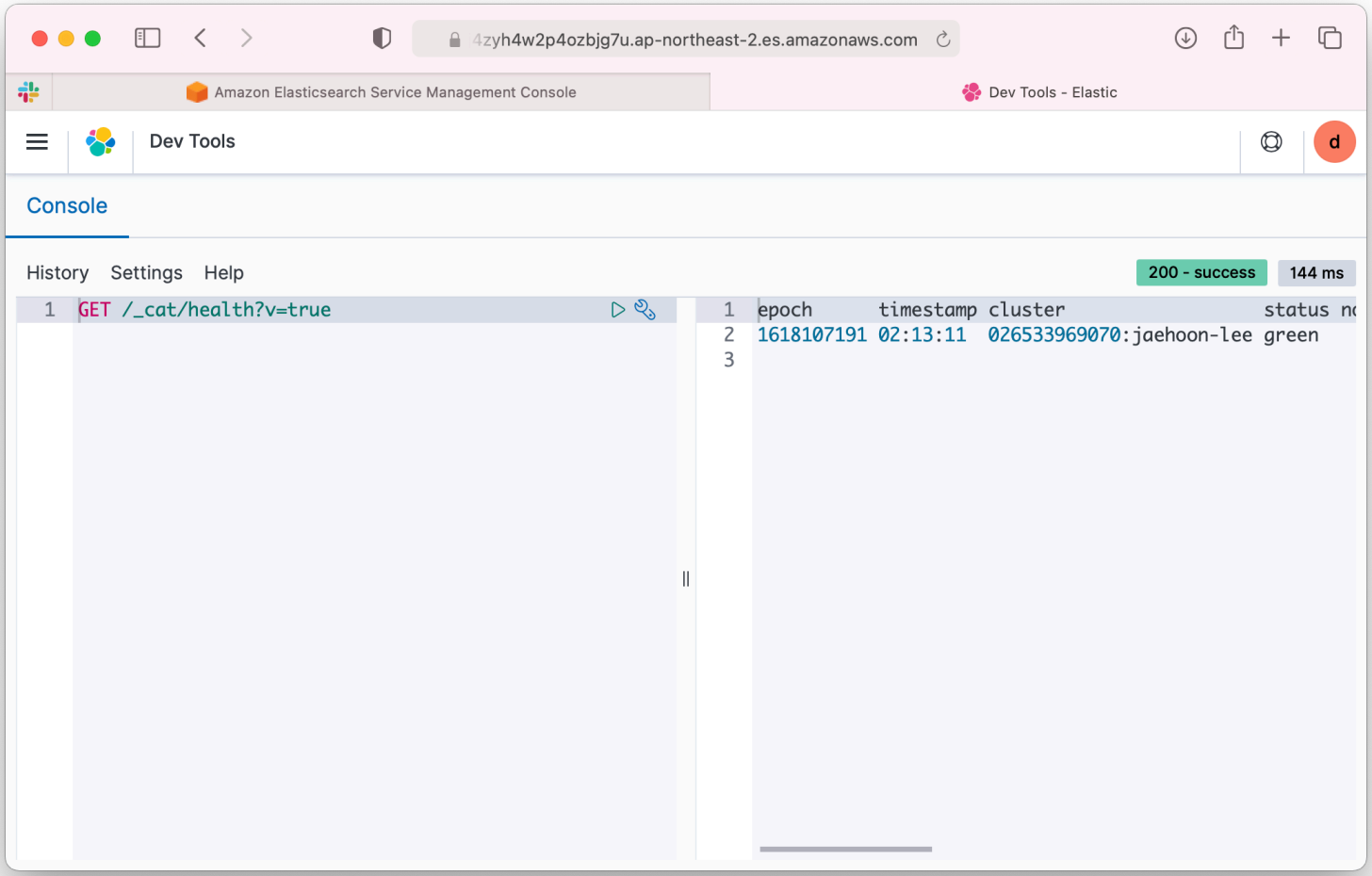
ElasticSearch Query

BAF675 금융 빅데이터 분석

이재훈, Week 2

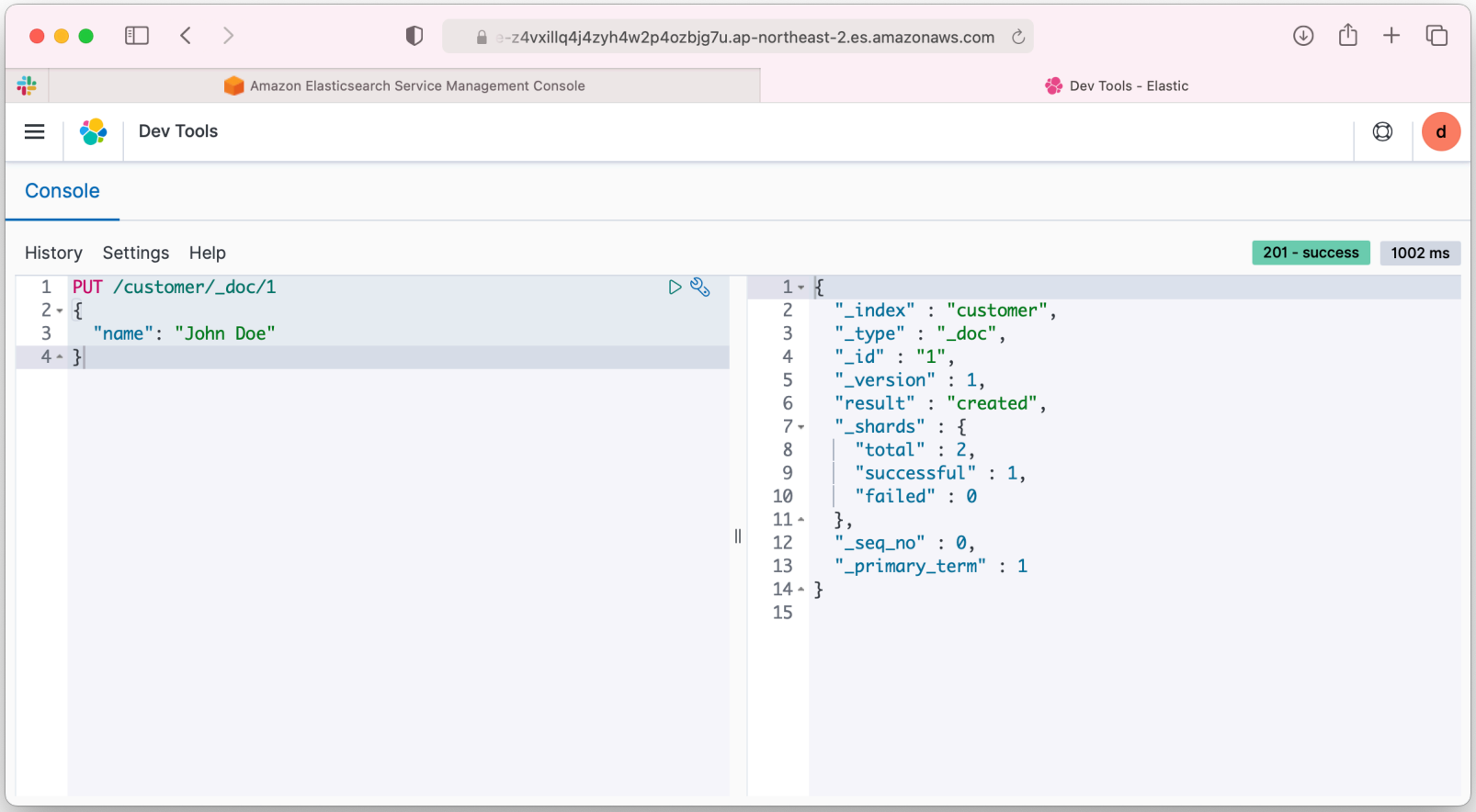
Kibana console 접속



HTTP request methods

- GET
- PUT / POST
- DELETE

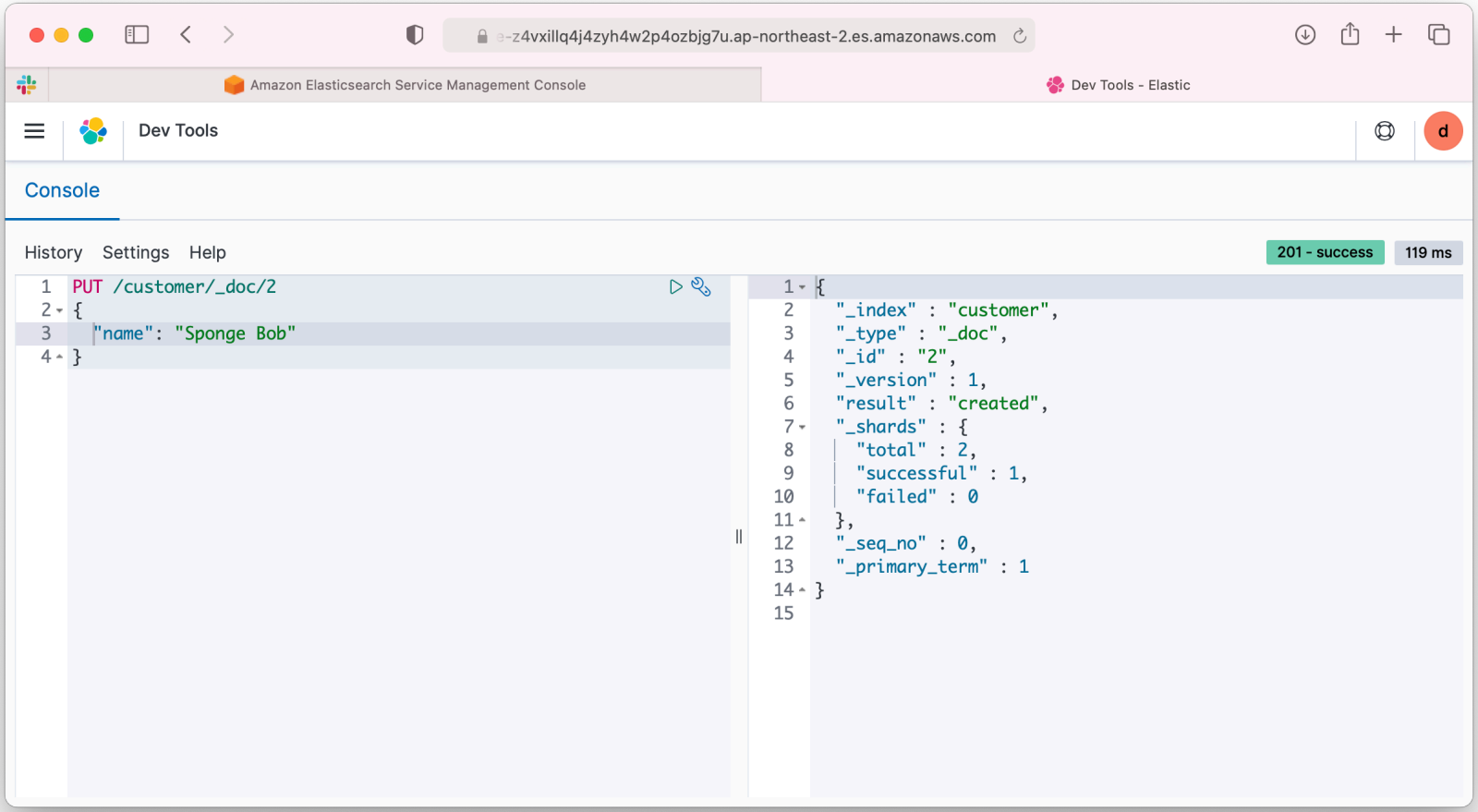
새로운 doc 입력


e-z4vxillq4j4zyh4w2p4ozbjg7u.ap-northeast-2.es.amazonaws.com
Amazon Elasticsearch Service Management Console
Dev Tools - Elastic
Dev Tools
d
Console
History
Settings
Help
201 - success
1002 ms
PUT /customer/_doc/1
{
"name": "John Doe"
}
{
"_index" : "customer",
"_type" : "_doc",
"_id" : "1",
"_version" : 1,
"result" : "created",
"_shards" : {
"total" : 2,
"successful" : 1,
"failed" : 0
},
"_seq_no" : 0,
"_primary_term" : 1
}

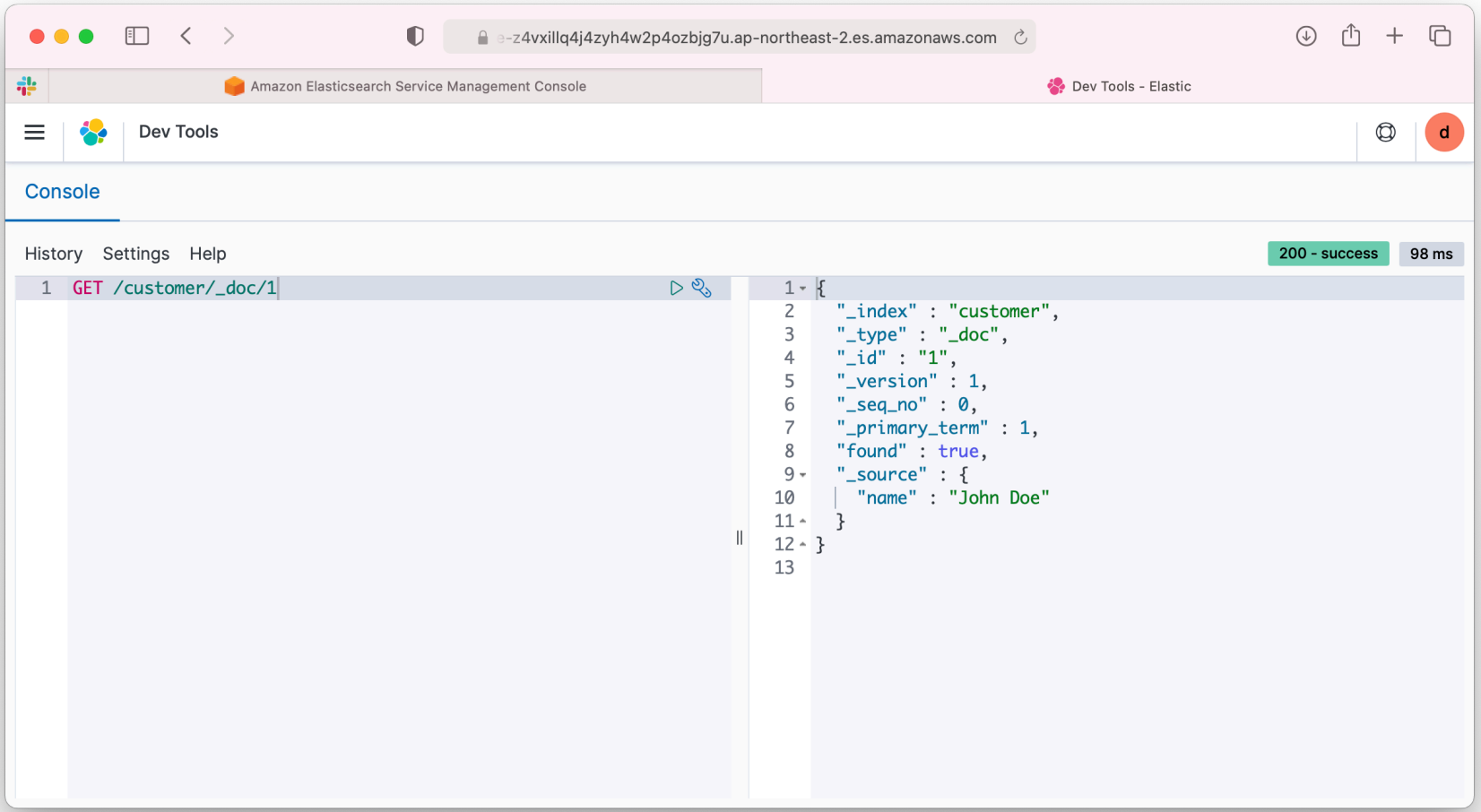
쿼리 해석

- SQL 이 row 단위로 데이터를 관리한다면, ElasticSearch 는 JSON document (doc) 단위로 기록 관리
- customer 라는 이름의 새로운 index 생성
  - ElasticSearch 의 index 는 SQL 의 table 같은 개념
- Mapping concepts across SQL and Elasticsearch
  - https://www.elastic.co/guide/en/elasticsearch/reference/current/_mapping_concepts_across_sql_and_elasticsearch.html
- customer 라는 index 내 id = 1 로 John Doe 라는 고객 정보 doc 생성

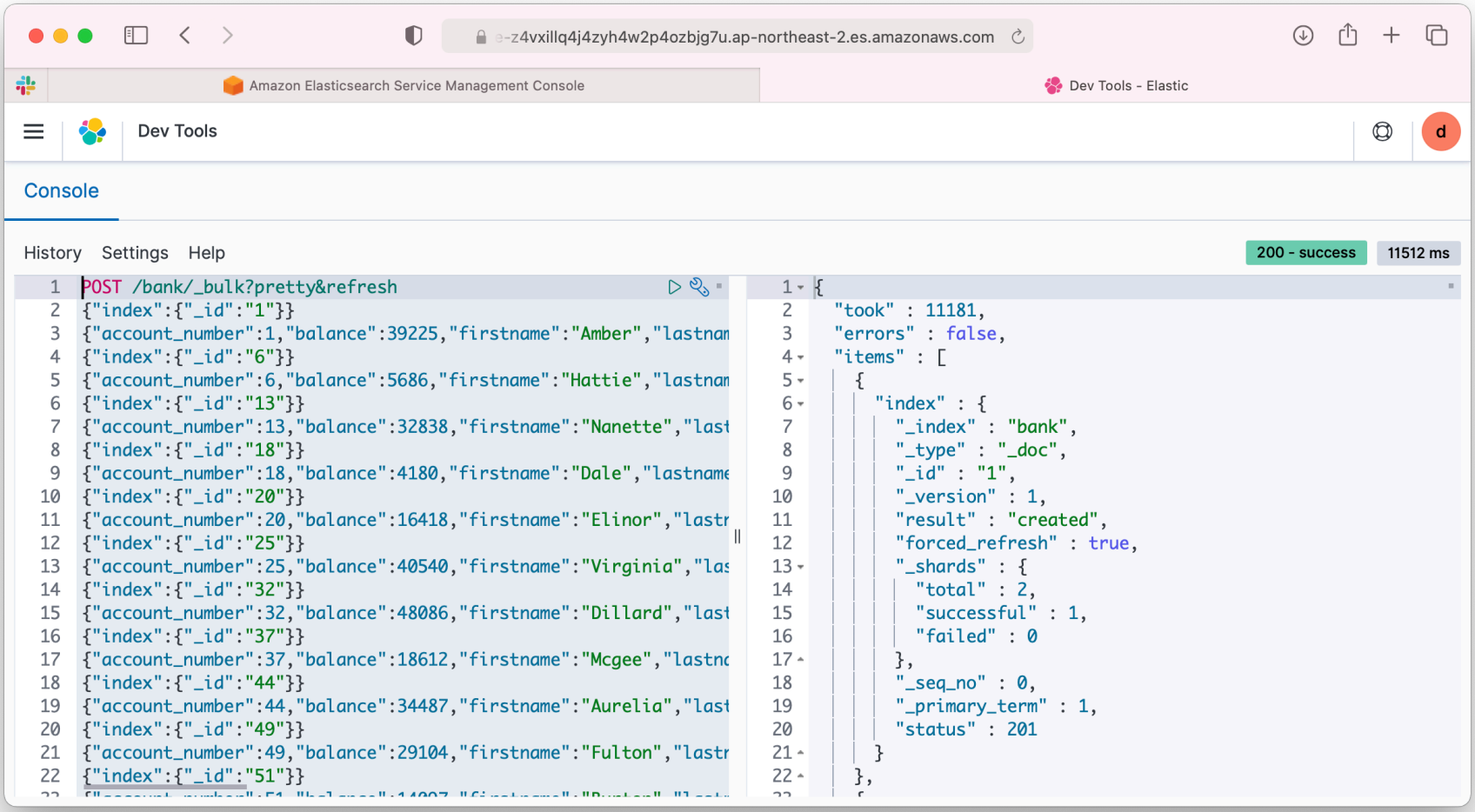
Let's add one more



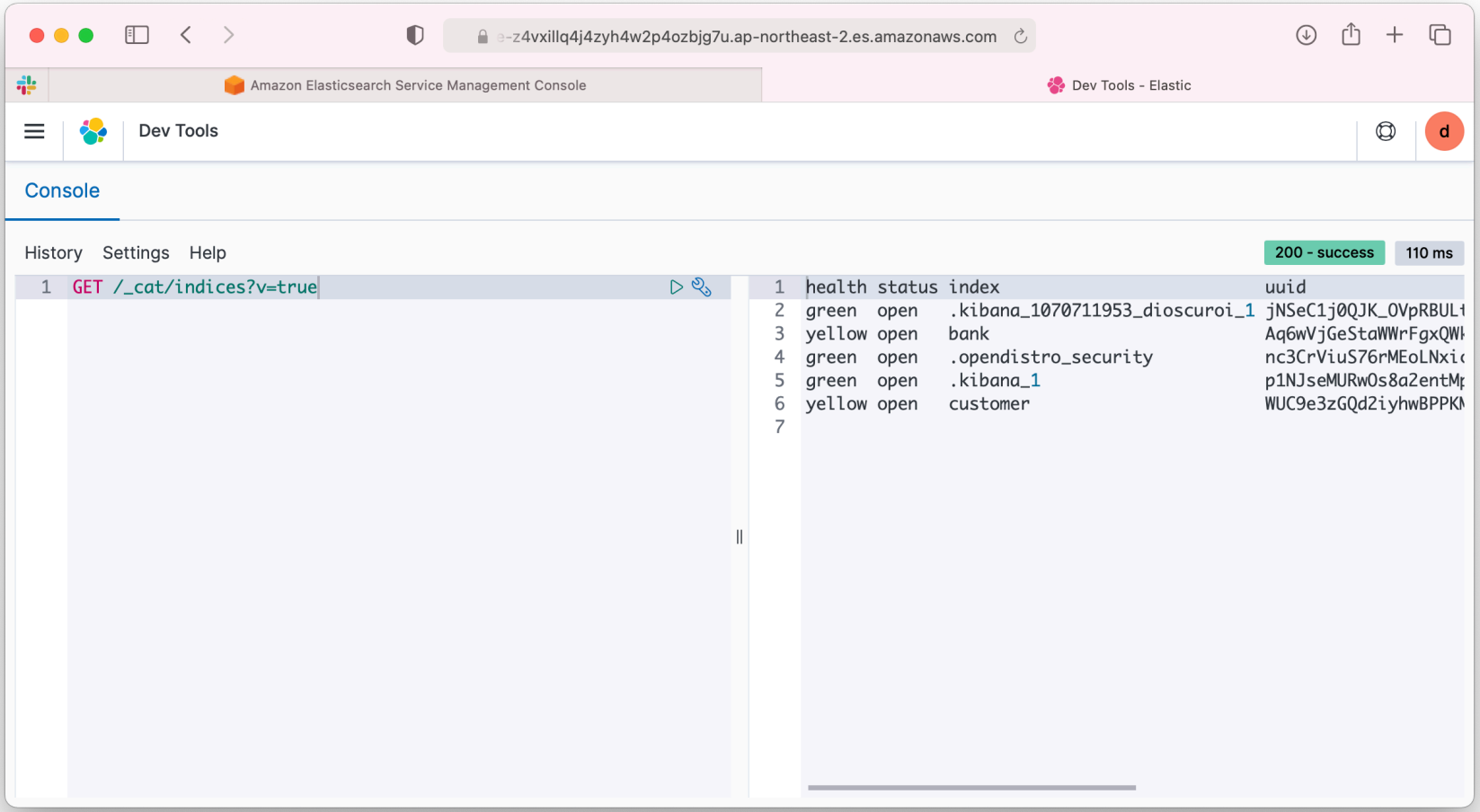
방금 입력한 데이터 조회


e-z4vxillq4j4zyh4w2p4ozbjg7u.ap-northeast-2.es.amazonaws.com
Amazon Elasticsearch Service Management Console
Dev Tools - Elastic
Dev Tools
d
Console
History
Settings
Help
200 - success
98 ms
1 GET /customer/_doc/1
1 {
2 "_index" : "customer",
3 "_type" : "_doc",
4 "_id" : "1",
5 "_version" : 1,
6 "_seq_no" : 0,
7 "_primary_term" : 1,
8 "found" : true,
9 "_source" : {
10 "name" : "John Doe"
11 }
12 }
13

Bulk 데이터 입력


z4vxillq4j4zyh4w2p4ozbjg7u.ap-northeast-2.es.amazonaws.com
Amazon Elasticsearch Service Management Console
Dev Tools - Elastic
Dev Tools
d
Console
History
Settings
Help
200 - success
11512 ms
POST /bank/_bulk?pretty&refresh
{"index":{"_id":"1"}}
{"account_number":1,"balance":39225,"firstname":"Amber","lastnam
{"index":{"_id":"6"}}
{"account_number":6,"balance":5686,"firstname":"Hattie","lastnam
{"index":{"_id":"13"}}
{"account_number":13,"balance":32838,"firstname":"Nanette","last
{"index":{"_id":"18"}}
{"account_number":18,"balance":4180,"firstname":"Dale","lastname
{"index":{"_id":"20"}}
{"account_number":20,"balance":16418,"firstname":"Elinor","lastr
{"index":{"_id":"25"}}
{"account_number":25,"balance":40540,"firstname":"Virginia","las
{"index":{"_id":"32"}}
{"account_number":32,"balance":48086,"firstname":"Dillard","last
{"index":{"_id":"37"}}
{"account_number":37,"balance":18612,"firstname":"Mcgee","lastna
{"index":{"_id":"44"}}
{"account_number":44,"balance":34487,"firstname":"Aurelia","last
{"index":{"_id":"49"}}
{"account_number":49,"balance":29104,"firstname":"Fulton","lastr
{"index":{"_id":"51"}}
{
"took" : 11181,
"errors" : false,
"items" : [
{
"index" : {
"_index" : "bank",
"_type" : "_doc",
"_id" : "1",
"_version" : 1,
"result" : "created",
"forced_refresh" : true,
"_shards" : {
"total" : 2,
"successful" : 1,
"failed" : 0
},
"_seq_no" : 0,
"_primary_term" : 1,
"status" : 201
}
},

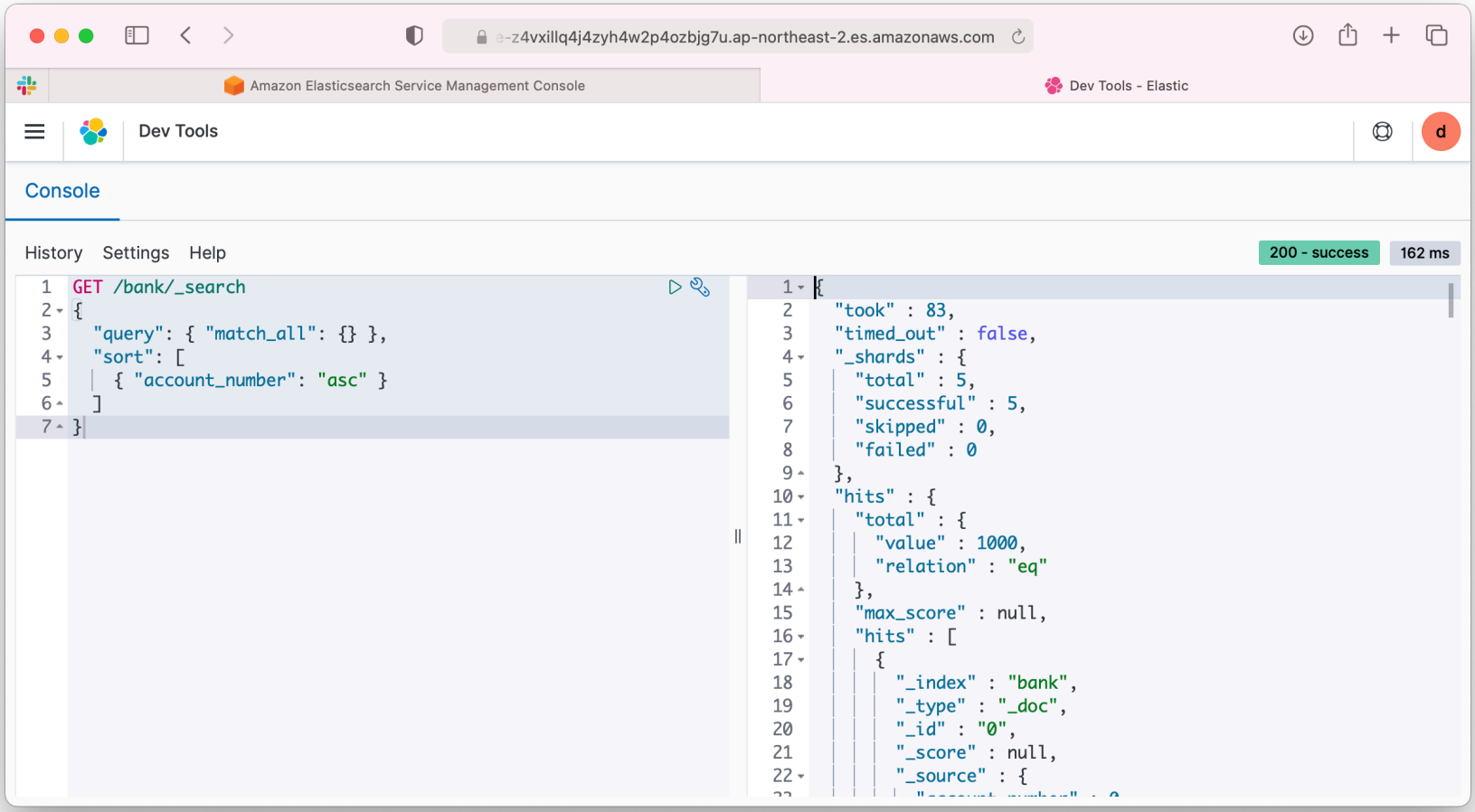
데이터 입력 상황 확인


Amazon Elasticsearch Service Management Console
Dev Tools - Elastic
Dev Tools
Console
History
Settings
Help
200 - success
110 ms
GET /_cat/indices?v=true
health status index uuid
green open .kibana_1070711953_dioscuroi_1 jNSeC1j0QJK_OVpRBULt
yellow open bank Aq6wVjGeStaWWrFgxQWh
green open .opendistro_security nc3CrViuS76rMEoLNxi
green open .kibana_1 p1NJseMURwOs8a2entMp
yellow open customer WUC9e3zGQd2iyhwBPPKN

Index status

- Health
  - Green: all good
  - Yellow: primary shards are allocated but replicas are not ready
  - Red: something has gone wrong.. some primary shards are not ready

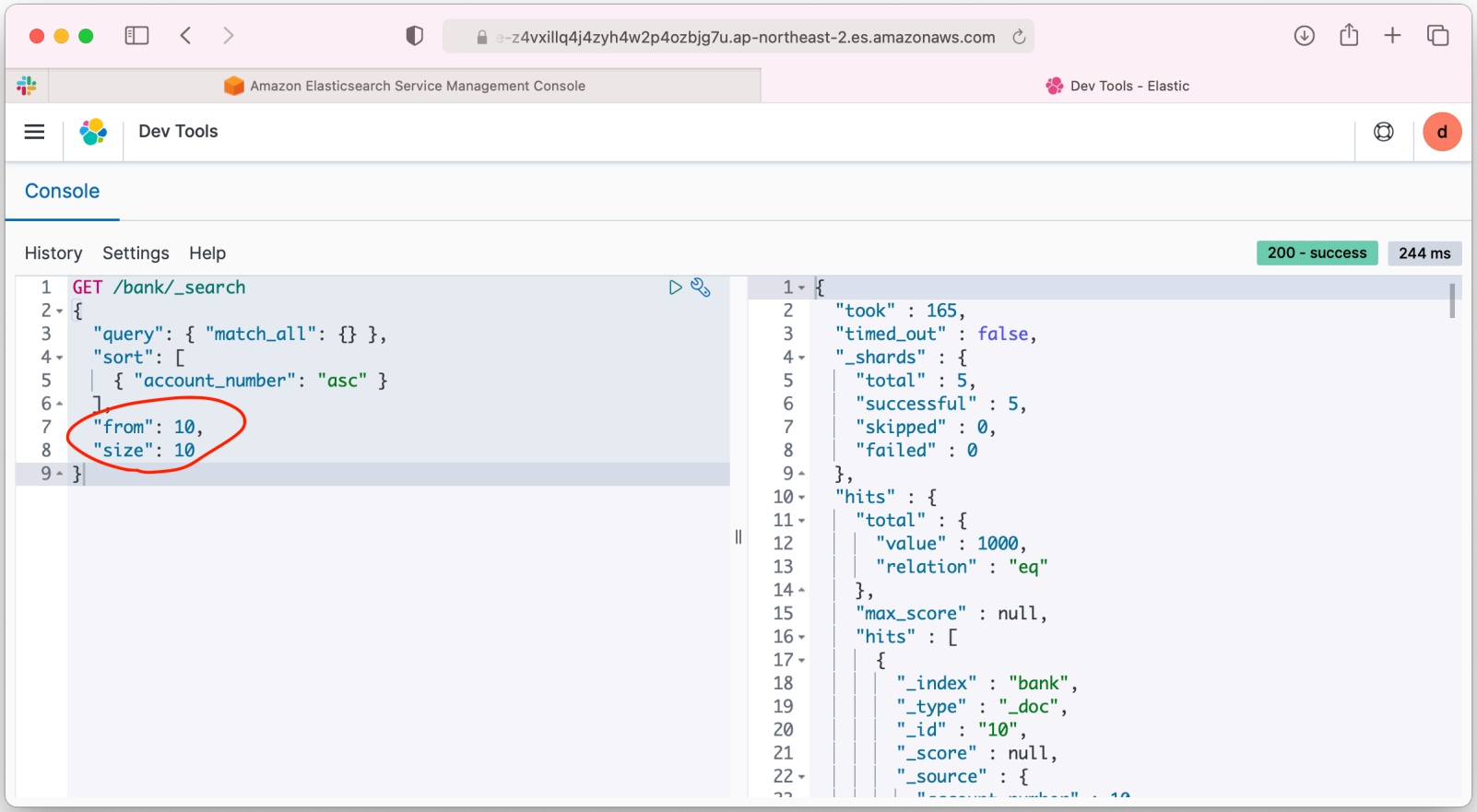
검색하기


e-z4vxillq4j4zyh4w2p4ozbjg7u.ap-northeast-2.es.amazonaws.com
Amazon Elasticsearch Service Management Console
Dev Tools - Elastic
Dev Tools
d
Console
History
Settings
Help
200 - success
162 ms
GET /bank/_search
{
"query": { "match_all": {} },
"sort": [
{ "account_number": "asc" }
]
}
{
"took" : 83,
"timed_out" : false,
"_shards" : {
"total" : 5,
"successful" : 5,
"skipped" : 0,
"failed" : 0
},
"hits" : {
"total" : {
"value" : 1000,
"relation" : "eq"
},
"max_score" : null,
"hits" : [
{
"_index" : "bank",
"_type" : "_doc",
"_id" : "0",
"_score" : null,
"_source" : {

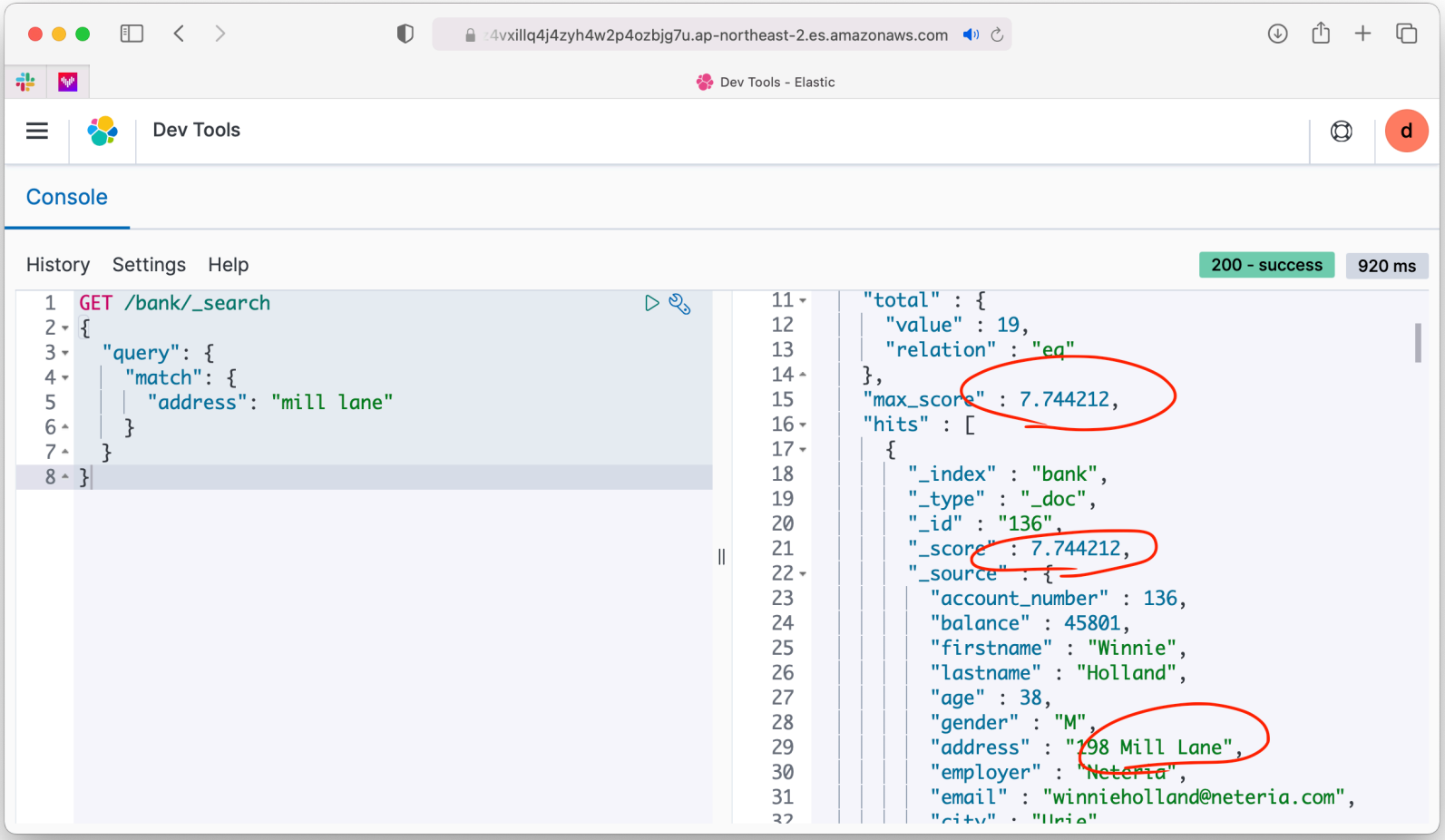
검색 결과

- took: 검색에 걸린 시간
- hits: 검색 결과 포함
  - 검색 결과 전체 갯수는 1,000 개이지만 그 중 상위 10개만 리턴

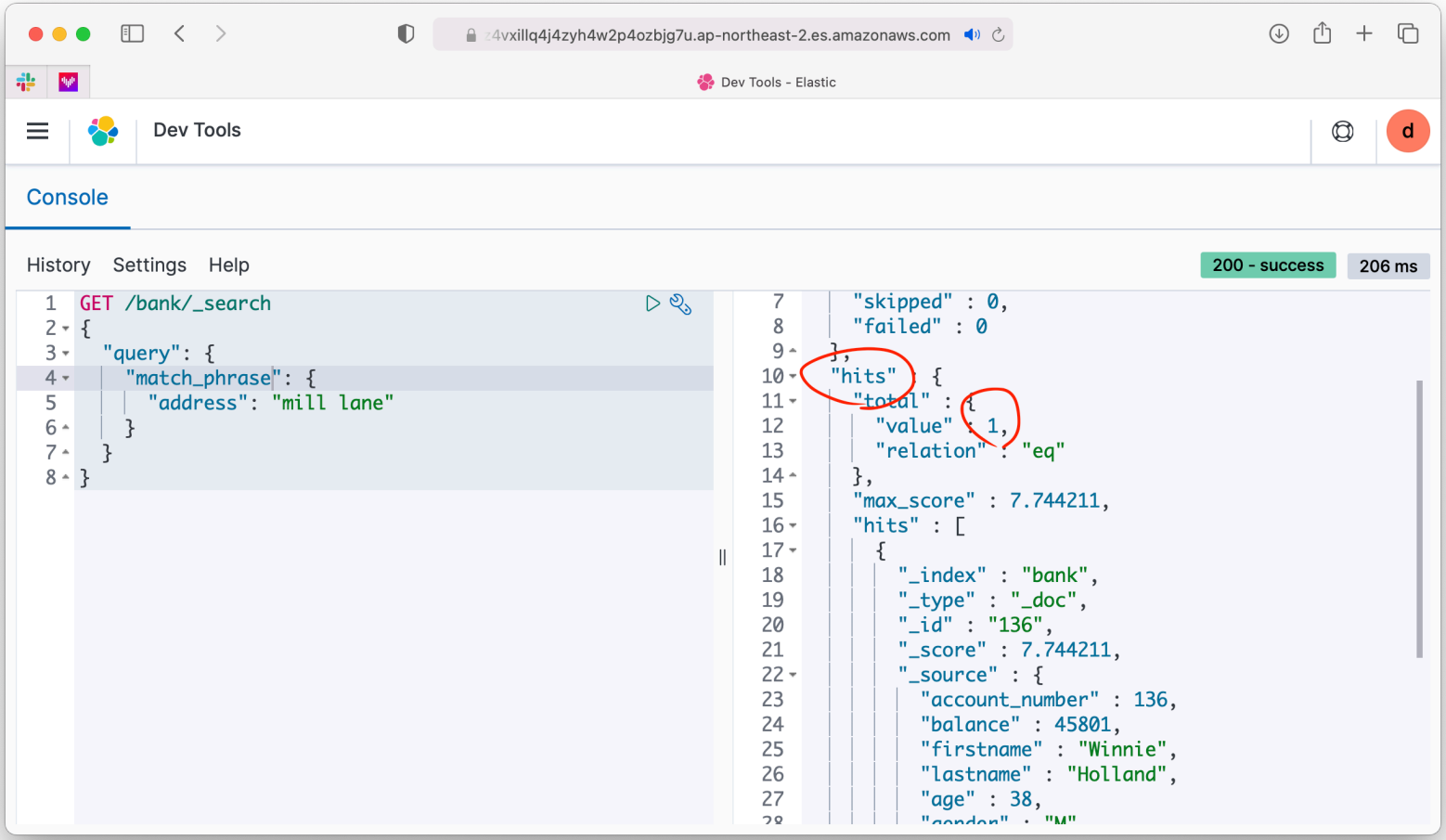
다음 10개 요청하기


z4vxillq4j4zyh4w2p4ozbjg7u.ap-northeast-2.es.amazonaws.com
Amazon Elasticsearch Service Management Console
Dev Tools - Elastic
Dev Tools
Console
History
Settings
Help
200 - success
244 ms
GET /bank/_search
{
"query": { "match_all": {} },
"sort": [
{ "account_number": "asc" }
],
"from": 10,
"size": 10
}
{
"took" : 165,
"timed_out" : false,
"_shards" : {
"total" : 5,
"successful" : 5,
"skipped" : 0,
"failed" : 0
},
"hits" : {
"total" : {
"value" : 1000,
"relation" : "eq"
},
"max_score" : null,
"hits" : [
{
"_index" : "bank",
"_type" : "_doc",
"_id" : "10",
"_score" : null,
"_source" : {

검색: 필드 지정

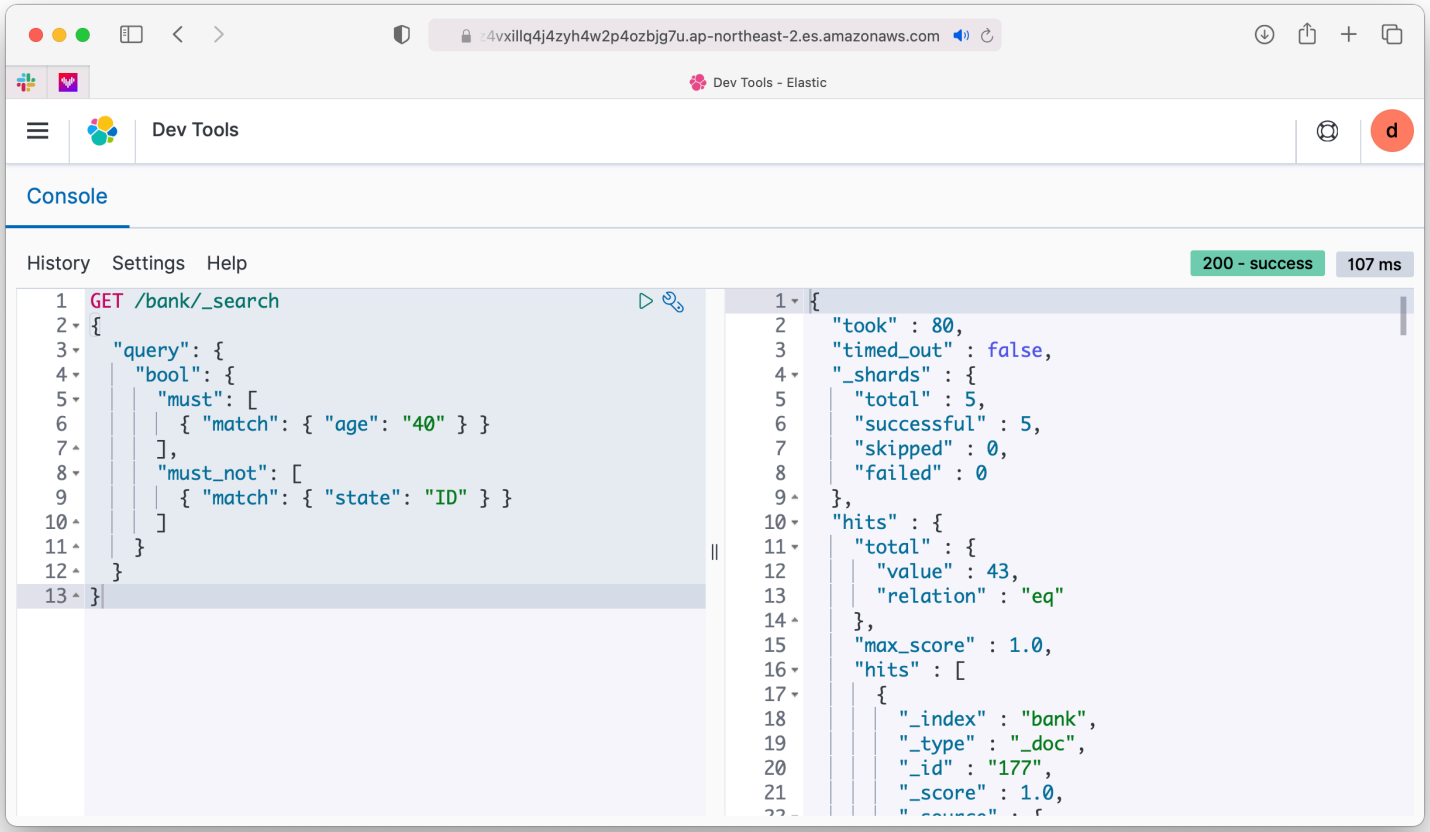

Dev Tools - Elastic
Dev Tools
Console
History
Settings
Help
200 - success
920 ms
GET /bank/_search
{
"query": {
"match": {
"address": "mill lane"
}
}
}
"total" : {
"value" : 19,
"relation" : "eq"
},
"max_score" : 7.744212,
"hits" : [
{
"_index" : "bank",
"_type" : "_doc",
"_id" : "136",
"_score" : 7.744212,
"_source" : {
"account_number" : 136,
"balance" : 45801,
"firstname" : "Winnie",
"lastname" : "Holland",
"age" : 38,
"gender" : "M",
"address" : "198 Mill Lane",
"employer" : "Neteria",
"email" : "winnieholland@neteria.com",

검색: 개별 단어 대신 phrase 비교



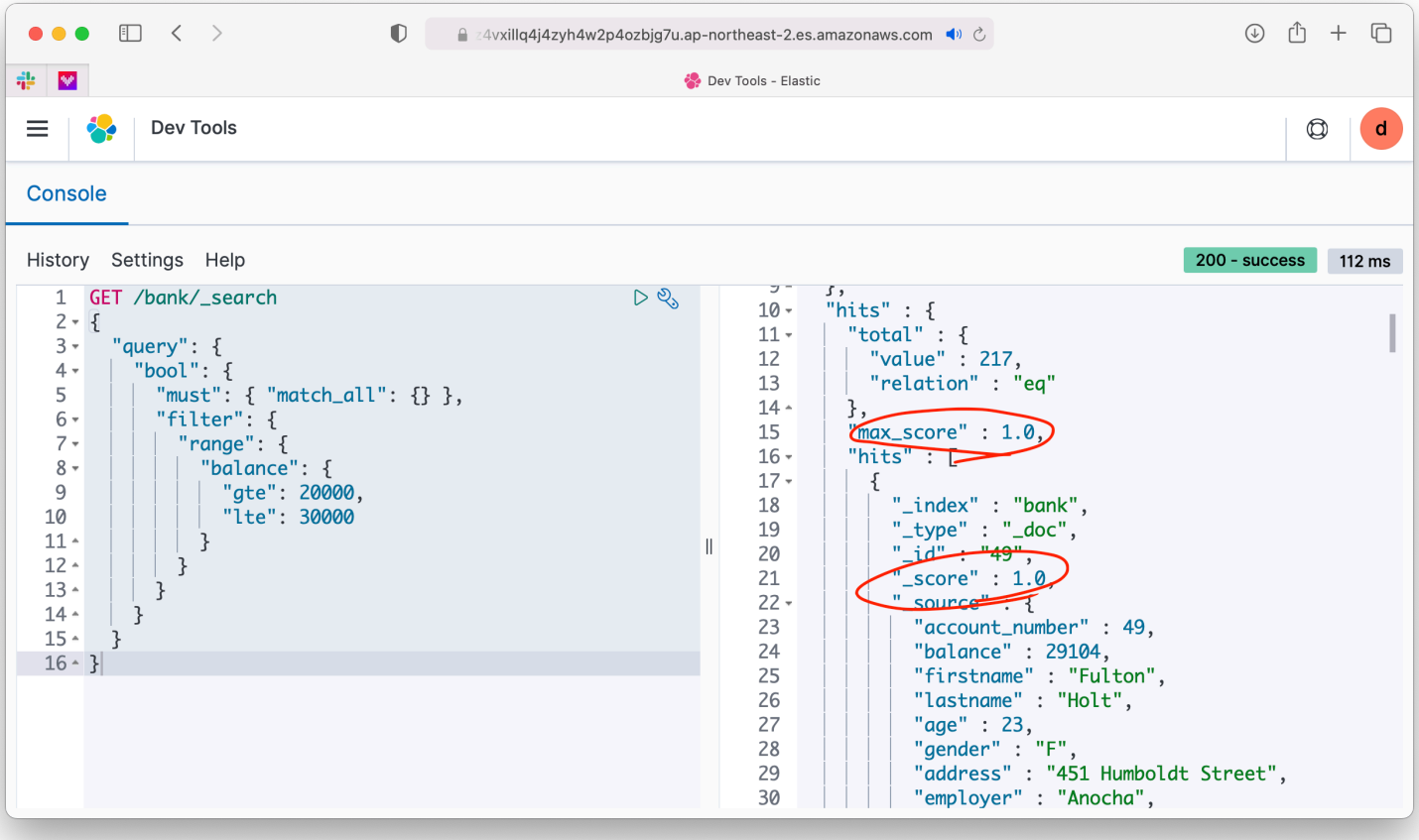
검색: bool 로 여러 조건 조합하기

- must, should, must_not 으로 조합



검색: filter 로 검색 결과 제한하기

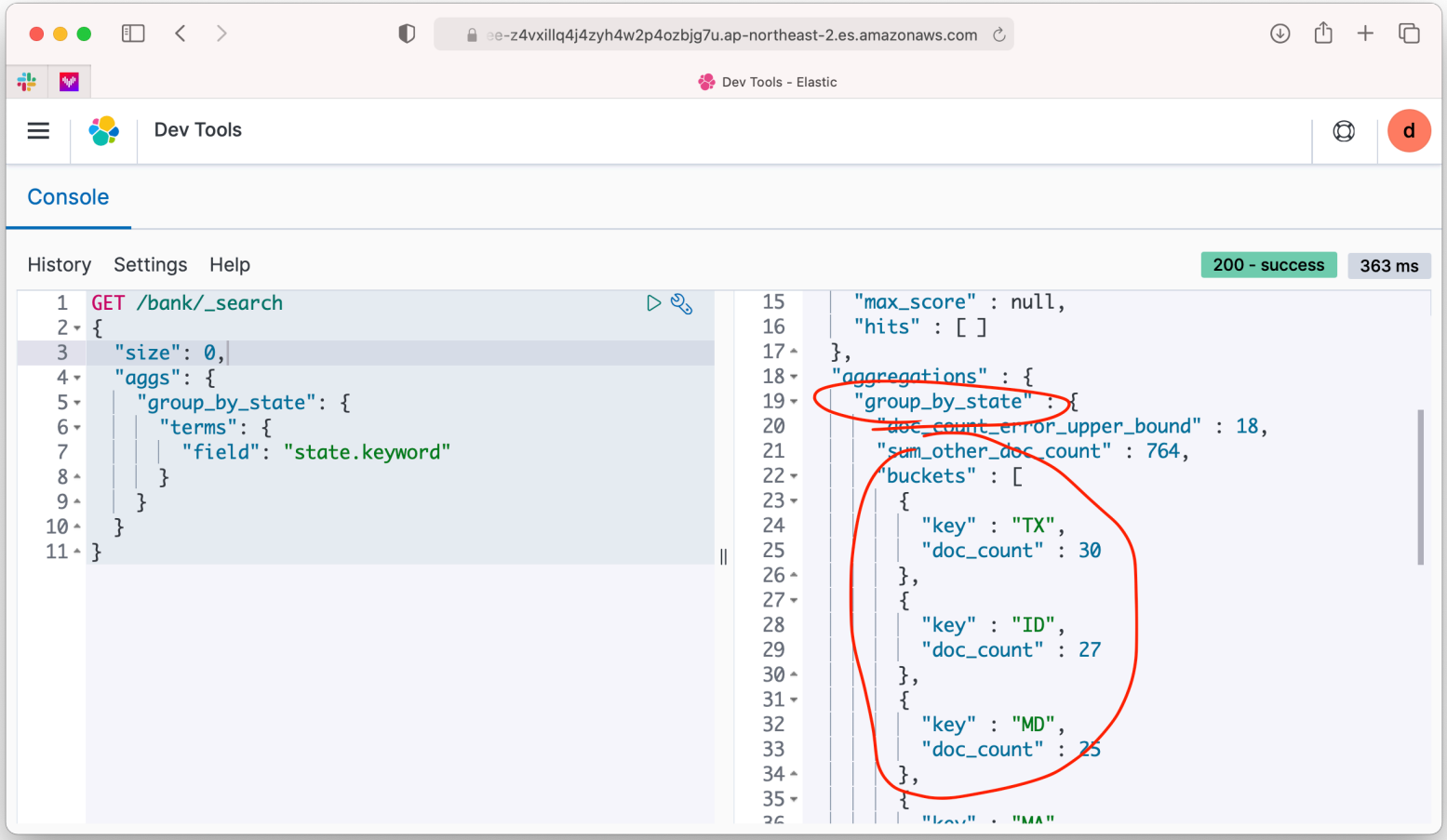
- 다른 query 와의 차이점: filter 는 score 에 영향을 주지 않는다



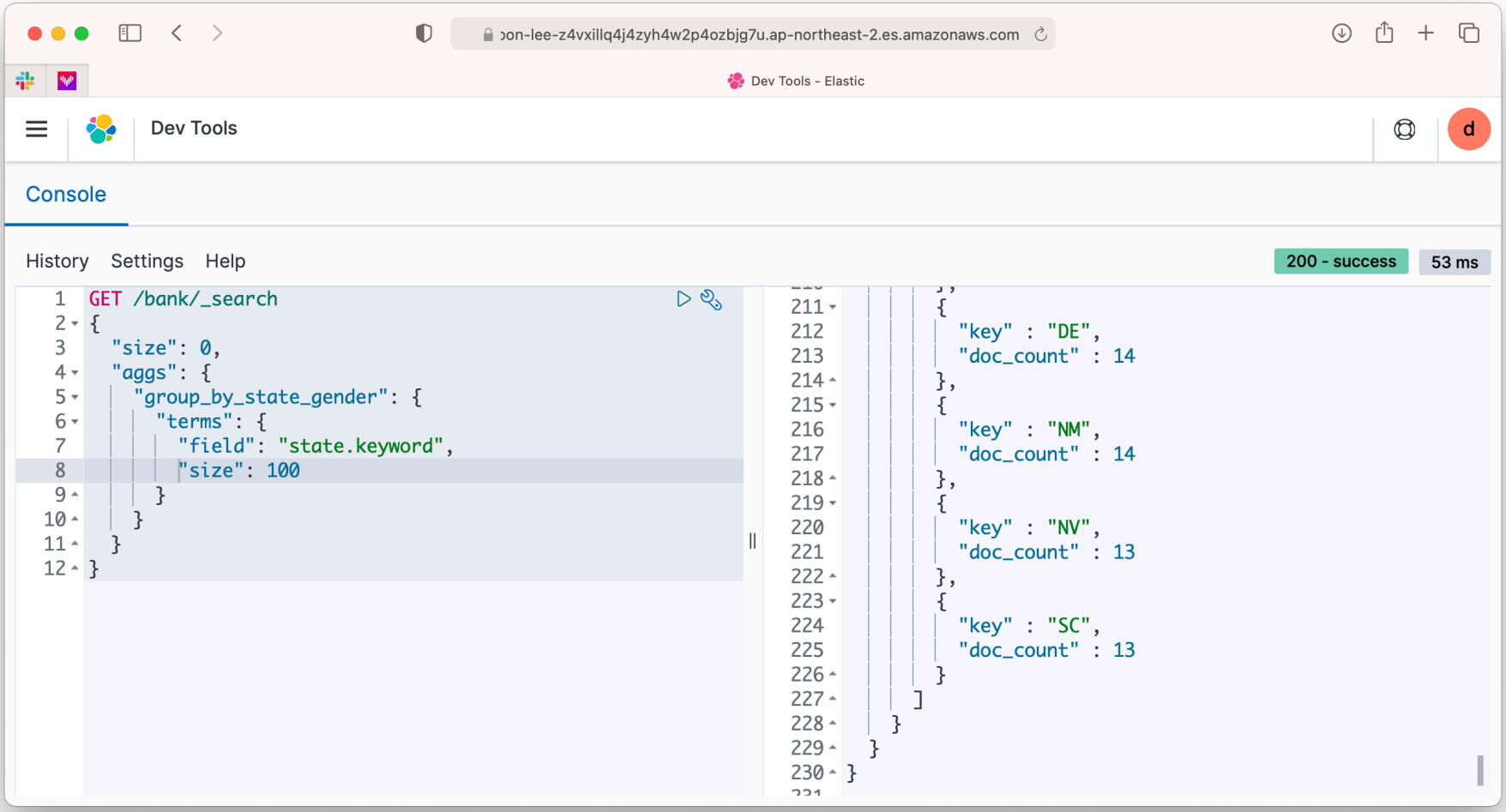
Aggregations

- pandas 에서 하던 group-by 기능
  - 분류별 갯수 세기, 합계, 평균 등

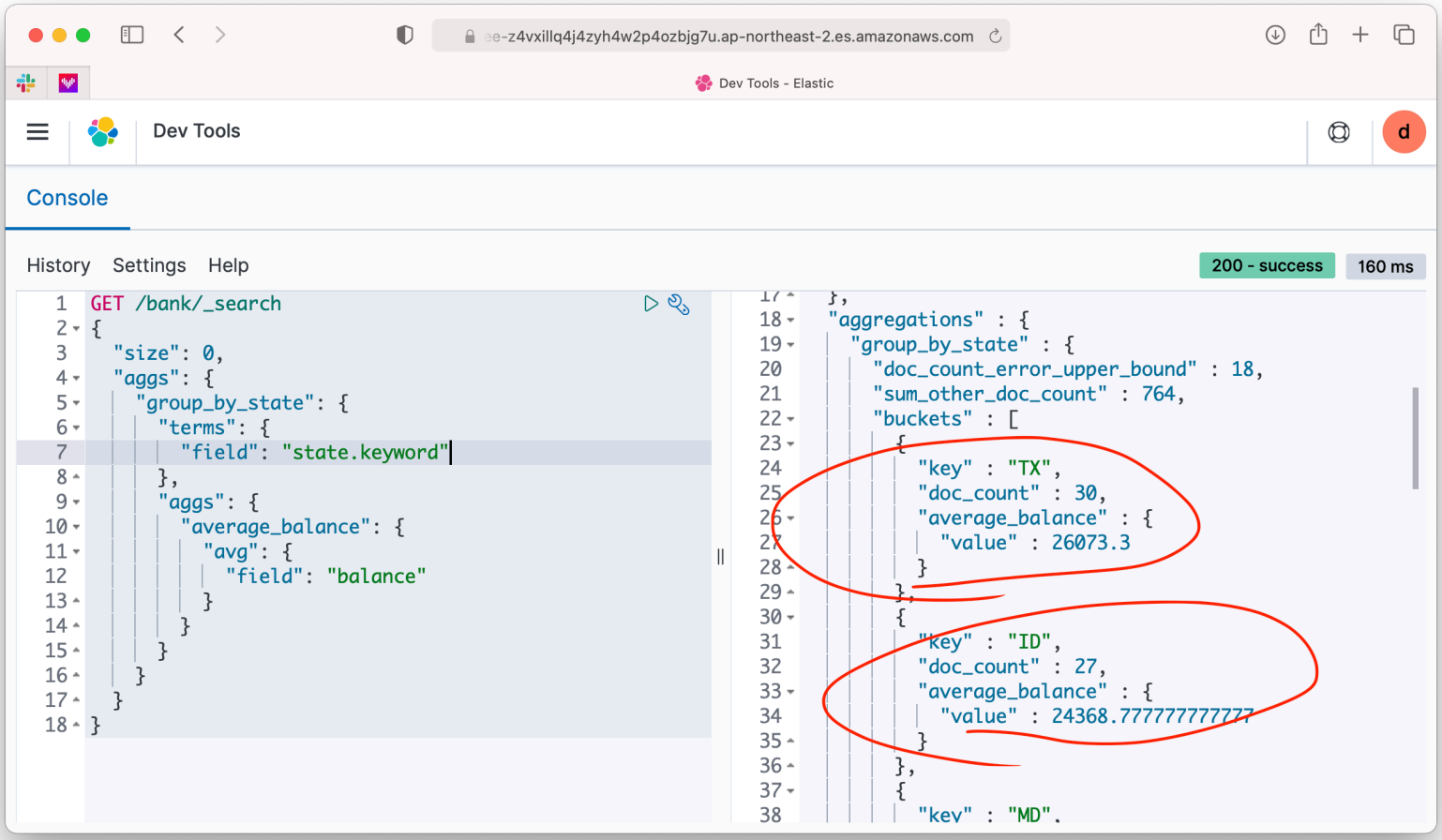
Aggregations: count



Aggregations: count - 52개 주 모두 요청하기



Aggregations: average balance



Aggregations: sort by average balance