Staying Ahead of Time

elasticsearch and time based data

Boaz Leskes
@bleskes

elasticsearch.

Basics

indices,types and other animals

A document

```
{
    "created_at": "Fri Jan 24 11:15:24 +0000 2014",
    "id": 426674590560305150,
    "text": "Prepping up for my #elasticsearch talk
        this afternoon at the UvA : http://t.co/rqhBI5zys0",
    "user": {
        "name": “Boaz Leskes",
        "screen_name": "bleskes",
    }
}
```

A type

= docs with similar data/structure


```
{
    "created_at": "Fri Jan 24 11:15:24 +0000 2014"
}
```


```
{
    "created_at": "Thu Jan 23 18:27:23 +0000 2014",
    "id": 426420915698544640,
    "text": "Elasticsearch es una maravilla !!!!",
    "user": {
        "name": "Abel Coronado",
        "screen_name": "abxda",
    }
}
```

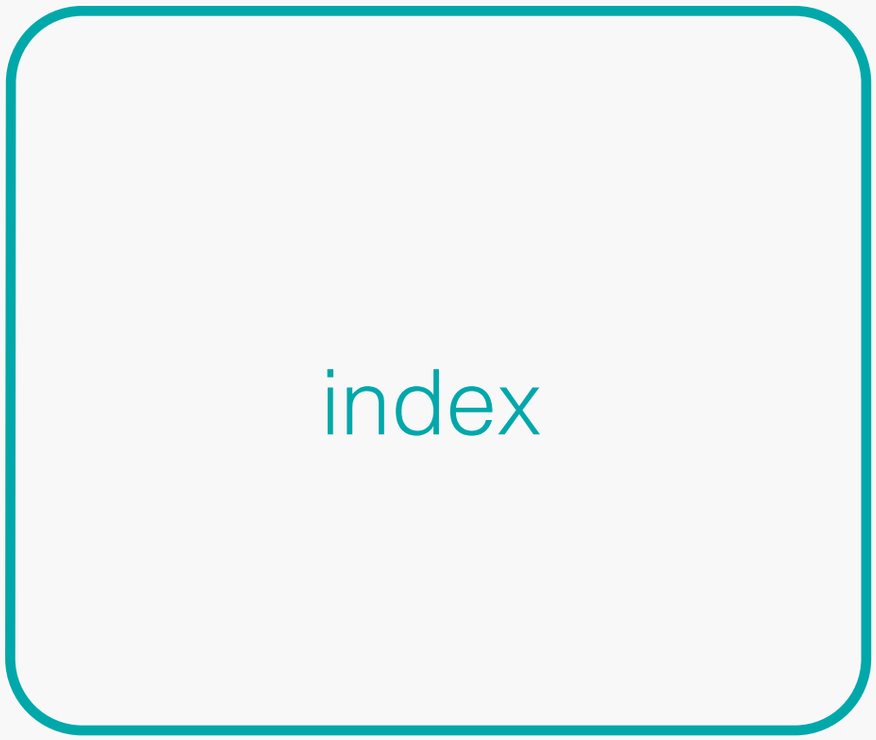
An index

= a collection of types

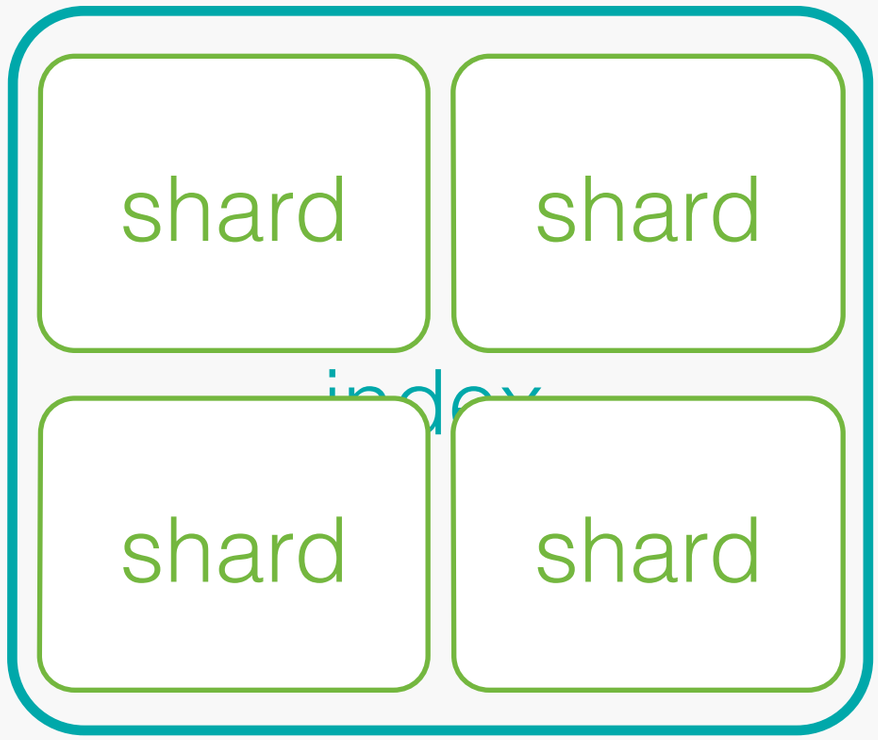
```
{
    "created_at": "Thu Jan 23
    "id": 426420915698544640,
    "text": "Elasticsearch Esc
    "user": {
        "name": "Abel Coronado
        "screen_name": "abxda"
    }
}
```

```
{
   "id": 19726002,
   "name": "Abel Coronado
   "screen_name": "abxda"
   "location":"Aguascalientes"
   "followers_count":871
   "friends_count":1794
   "listed_count":38
}
```

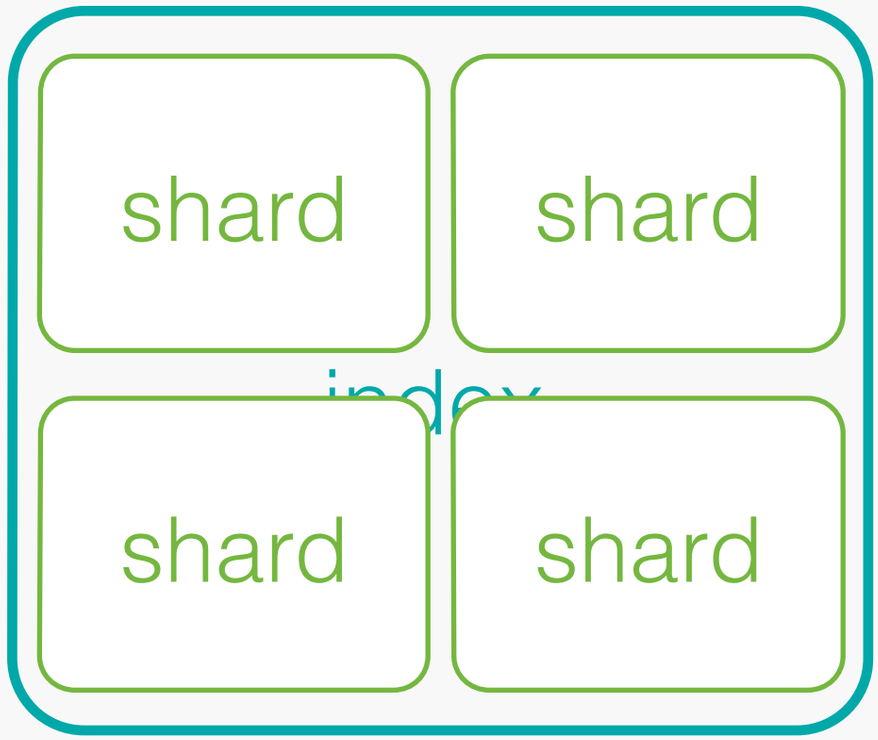
Sharding

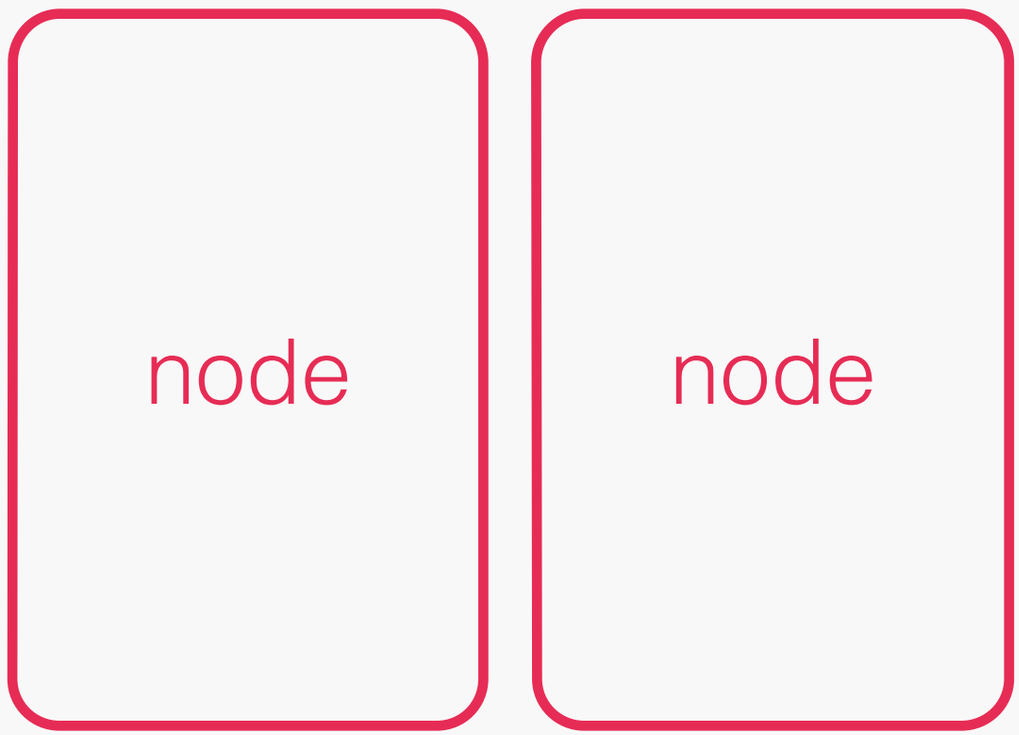


Sharding

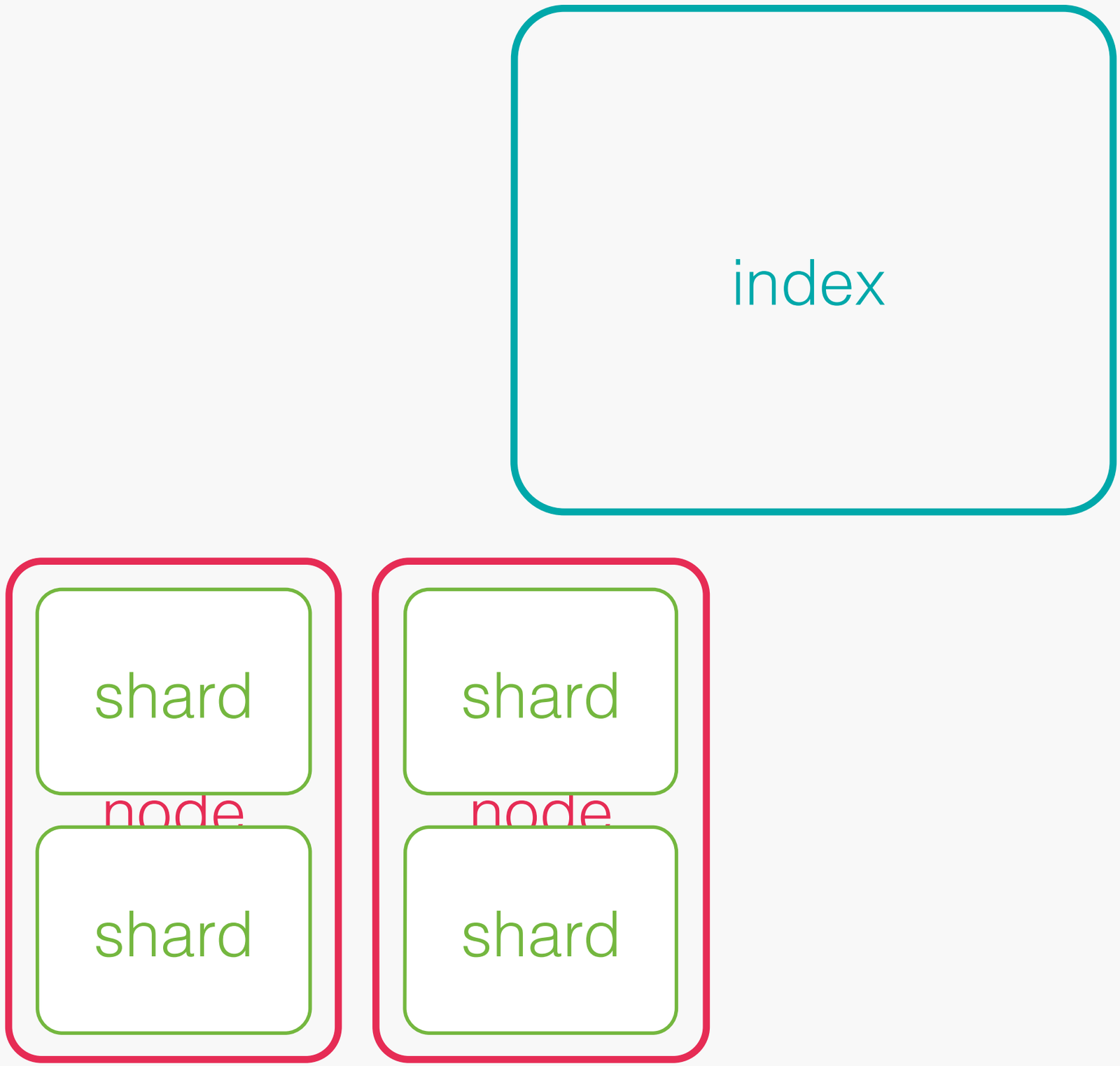


Sharding

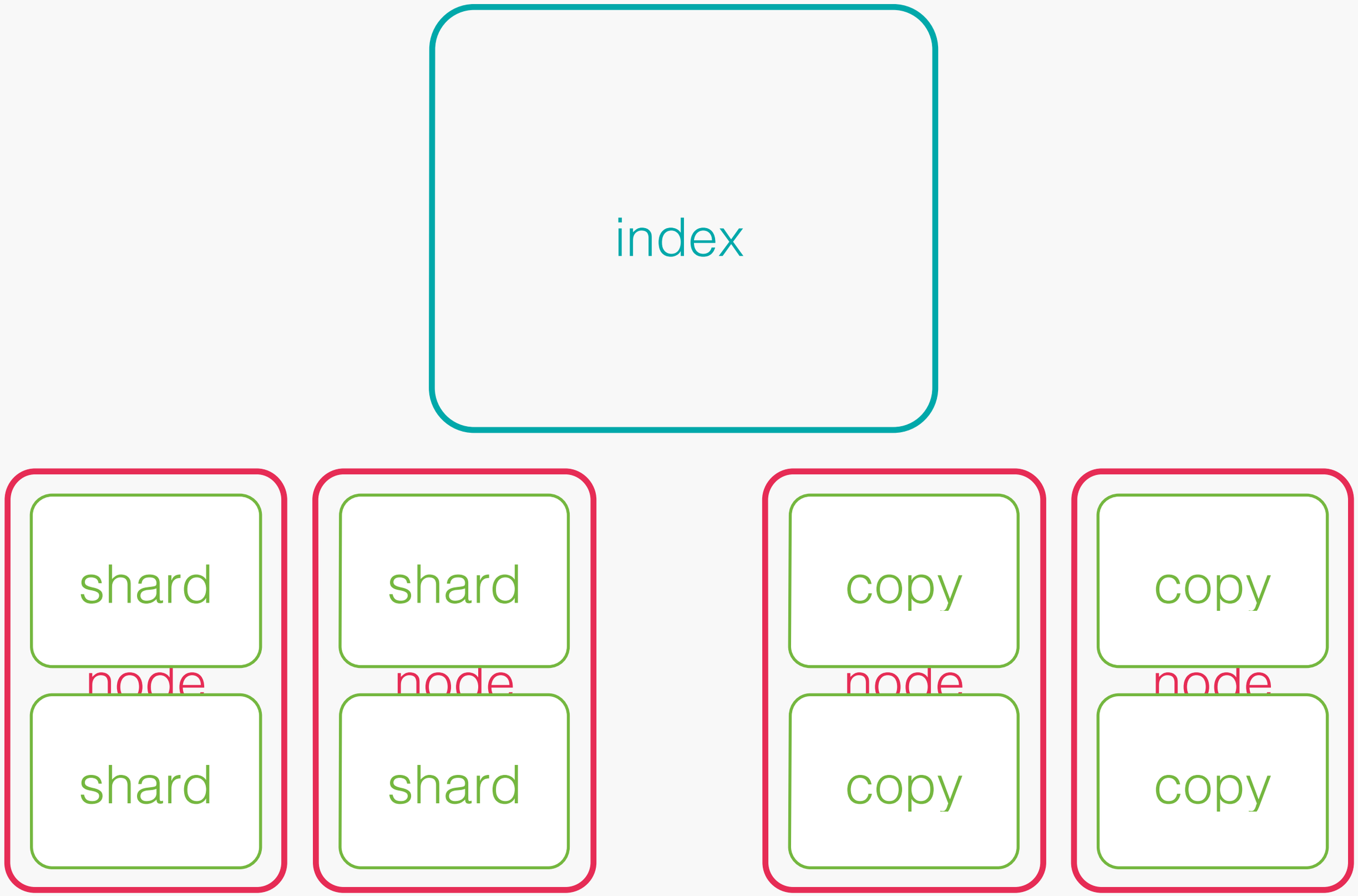




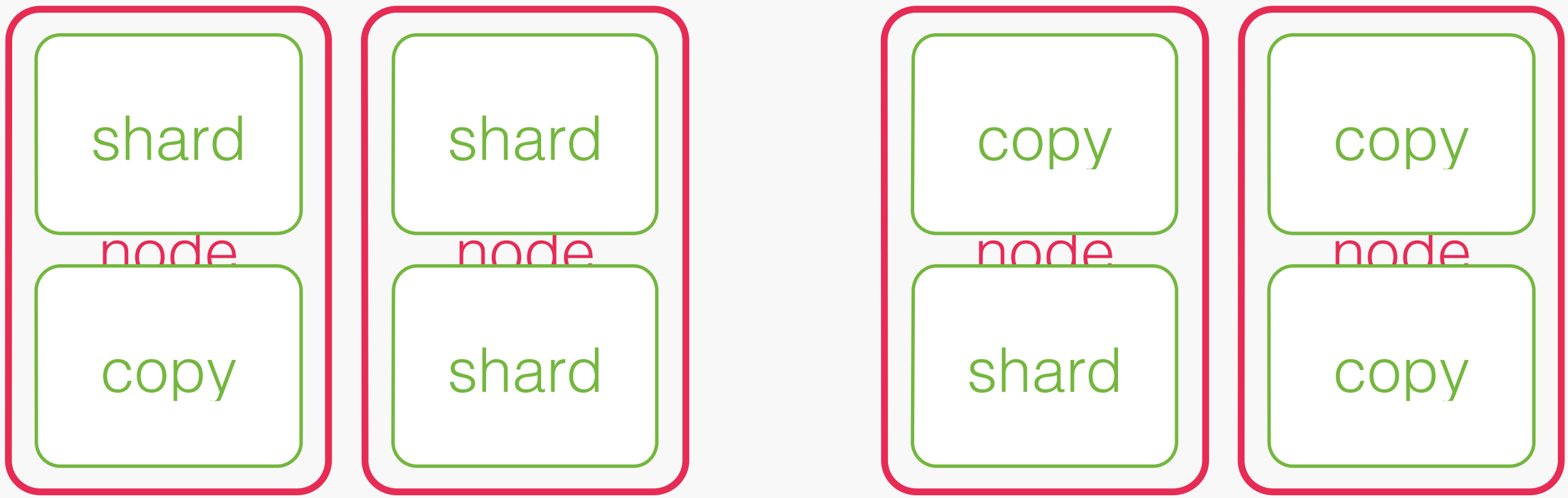
Sharding



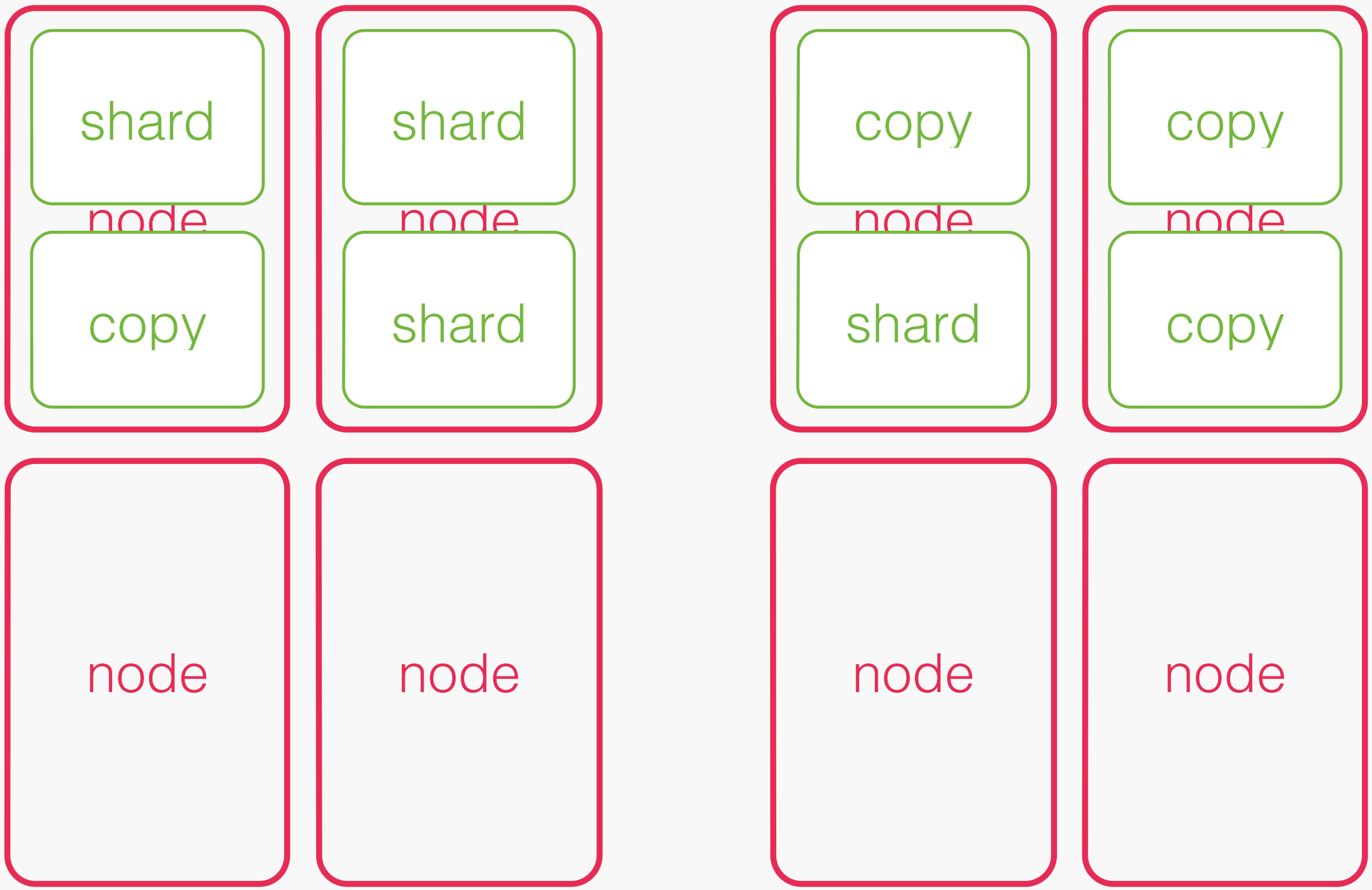
Sharding



Sharding

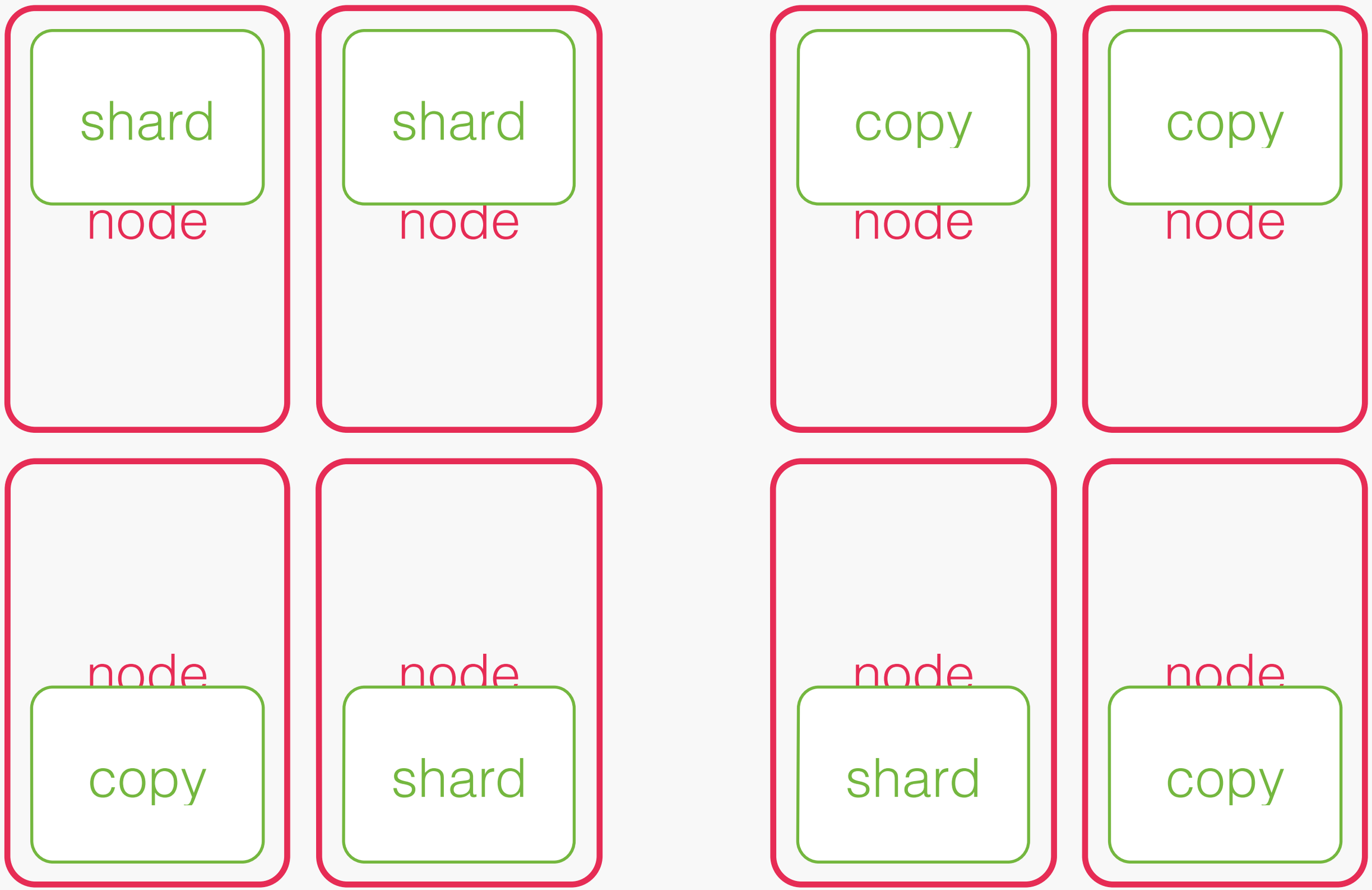


Sharding



elasticsearch.

Sharding



Important fact for later

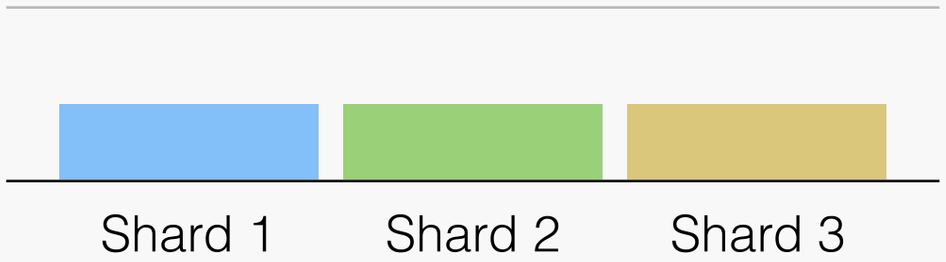
indexing & searching is done on shards, not indices

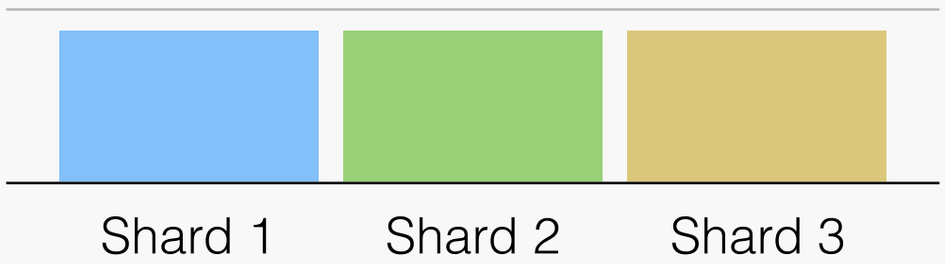
To the subject at hand

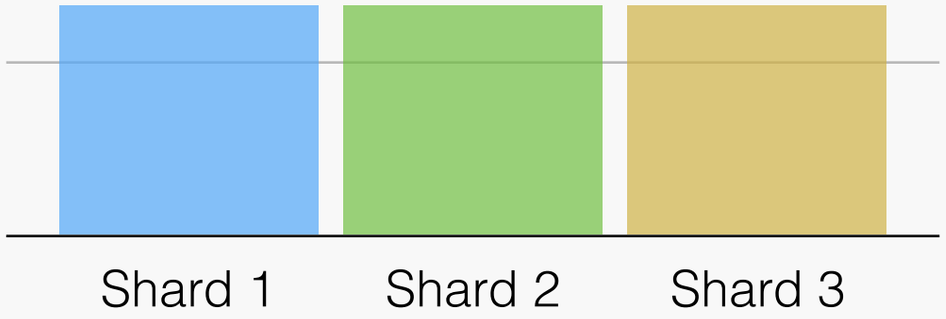
time based data

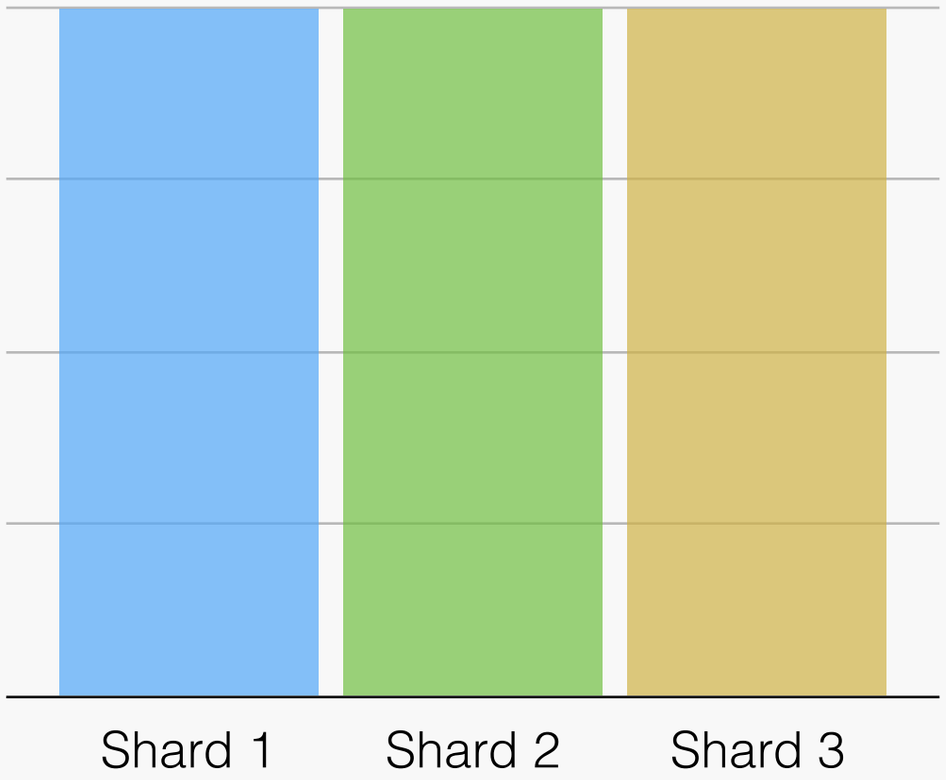
easy to get, easy to index

```
# curl -XPUT localhost:9200/tweets/tweet/426674590560305150 -d '{
    "created_at": "Fri Jan 24 11:15:24 +0000 2014",
    "id": 426674590560305150,
    "text": "Prepping up for my #elasticsearch talk
        this afternoon at the UvA : http://t.co/rqhBI5zys0",
    "user": {
        "name": “Boaz Leskes",
        "screen_name": "bleskes",
    }
}'
```



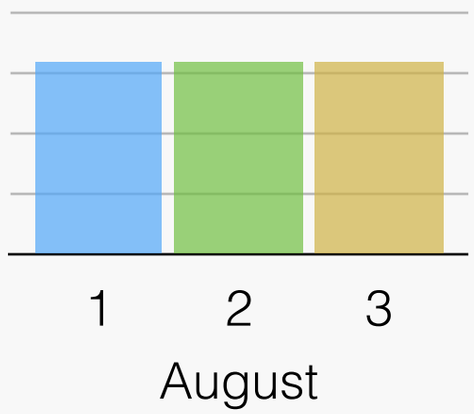


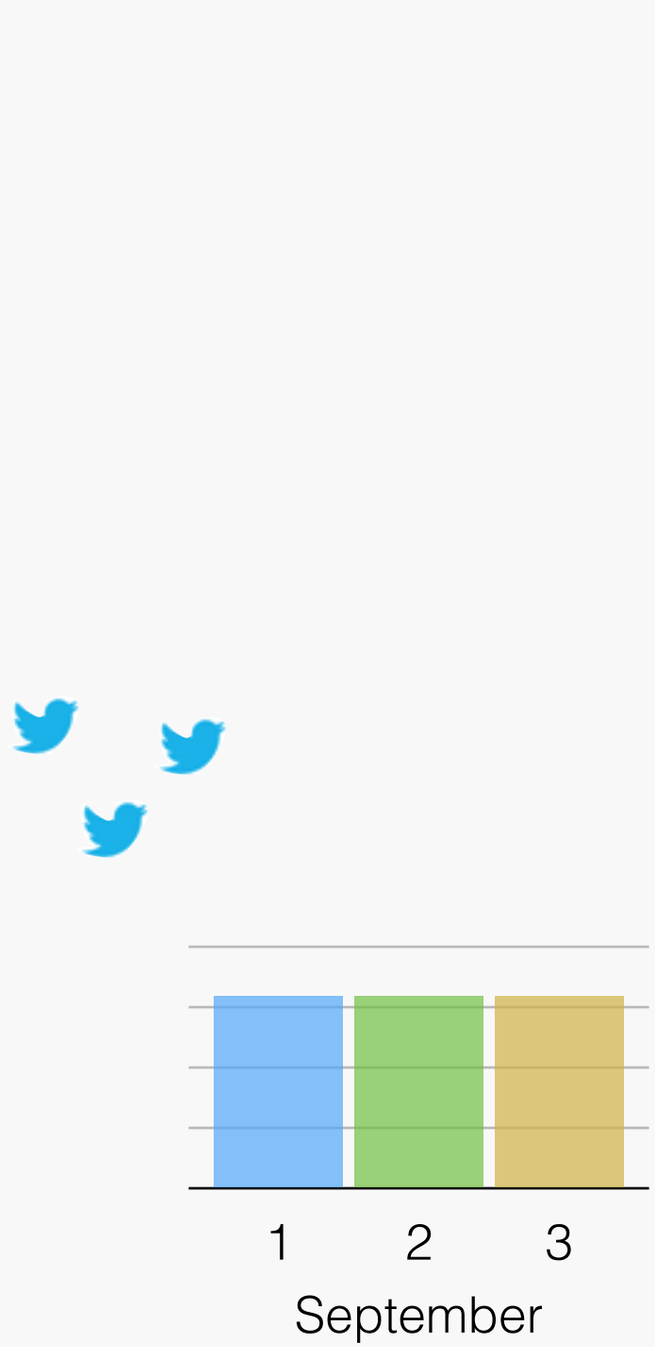


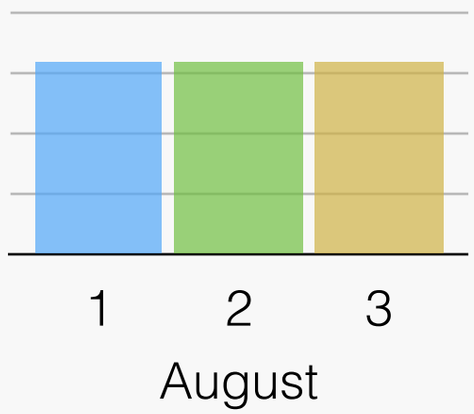


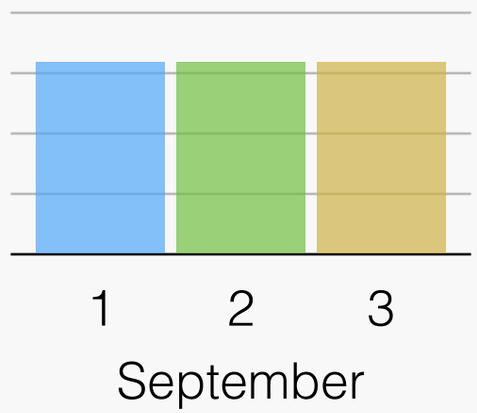
Reminds of a tile at my aunt's house

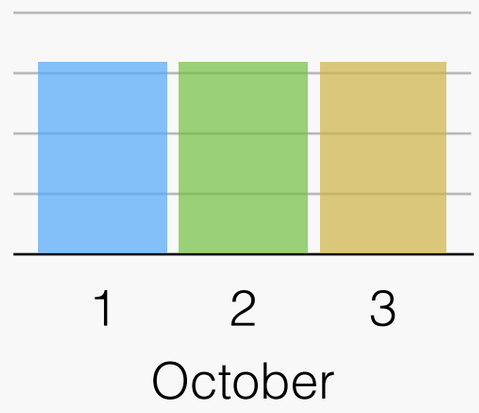
Today is the tomorrow we were all afraid of yesterday....





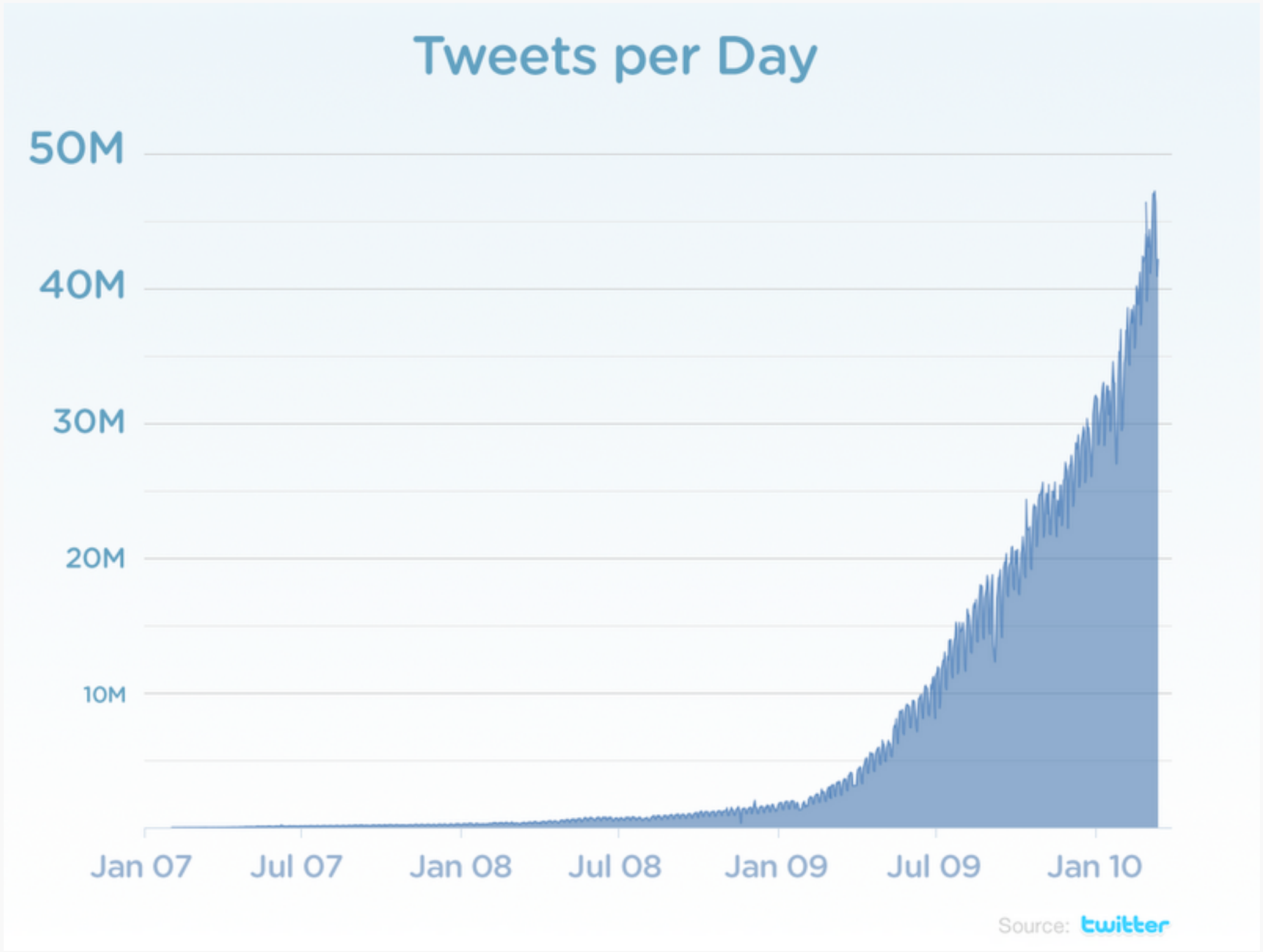






one little tweak…

```
# curl -XPUT localhost:9200/tweets_201310/tweet/426674590560305150
-d '{
    "created_at": "Fri Jan 24 11:15:24 +0000 2014",
    "id": 426674590560305150,
    "text": "Prepping up for my #elasticsearch talk
       this afternoon at the UvA : http://t.co/rqhBI5zys0",
    "user": {
       "name": “Boaz Leskes",
       "screen_name": "bleskes",
    }
}'
```



elasticsearch.

Another fact

index is the basic unit of configuration

index templates

```
curl -XPUT localhost:9200/_template/twitter -d '
{
    "template" : “twitter_*",
    "settings" : {
        "number_of_shards" : 4,
        "number_of_replicas" : 1
    }
}'
```

older data

```
# elasticsearch.yml

node.disk: spinning_disks
```

```
curl –XPUT localhost:9200/twitter_2012*/_settings –d '{
    "index.routing.allocation.include.disk" : “spinning_disks”,
    “index.routing.allocation.exclude.disk" : "ssd"
}'
```

older data

```
curl -XPOST localhost:9200/twitter_2012*/_optimize

curl -XDELETE localhost:9200/twitter_201201/

curl -XPOST localhost:9200/twitter_201201/_close
```

aliases

```
curl -XPUT localhost:9200/_aliases -d '{
      "add": {
        "index": "twitter_201311", "alias": "last_2_months"
    },
    "remove": {
        "index": "twitter_201309", "alias": "last_2_months"
    }
}'
```

Implications

- Use indices to manage data as it scales
- Use aliases to efficiently point your searches at the relevant shards

One More Thing..

Time is just a (strictly) monotonic function
Primary keys are just as good

thank you!

@elasticsearch , @bleskes

http://elasticsearch.org/guide

http://elasticsearch.com/support

elasticsearch.