

Creating indexes for multi-tenant solutions

Patryk Orwat

Software Architect - Bluestone

Who am I

Humanity is OK, but 99% of people are boring idiots – Slavoj Žižek

I'm a boring idiot, write code for living, deliver solutions using AWS services



Challenge



Let's create a fancy
solution with
Elasticsearch



And make it for different
clients



Customers have different
requirements/can
interfere with each other



We want to optimize
costs

Plan

Multi tenancy modes

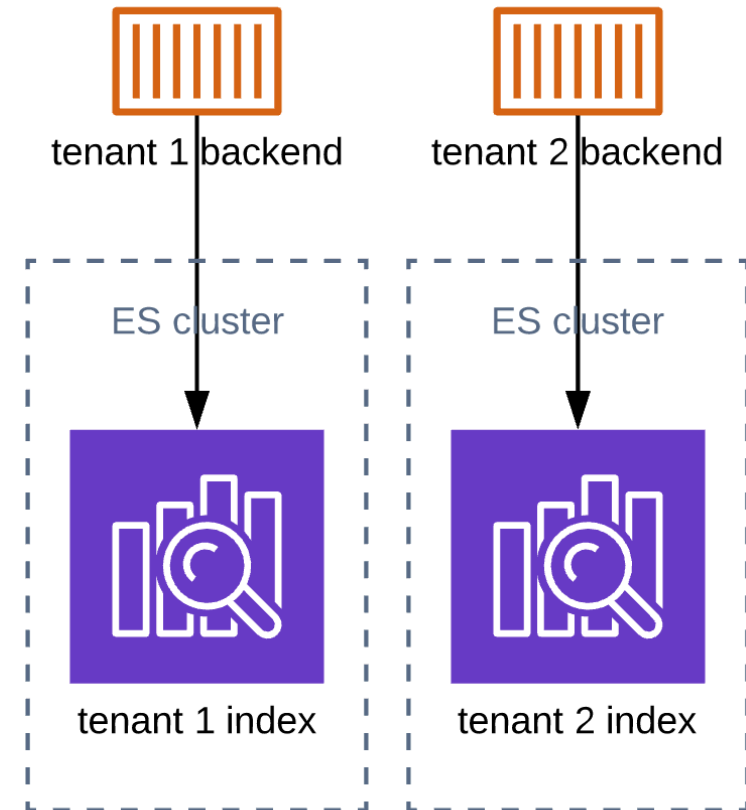
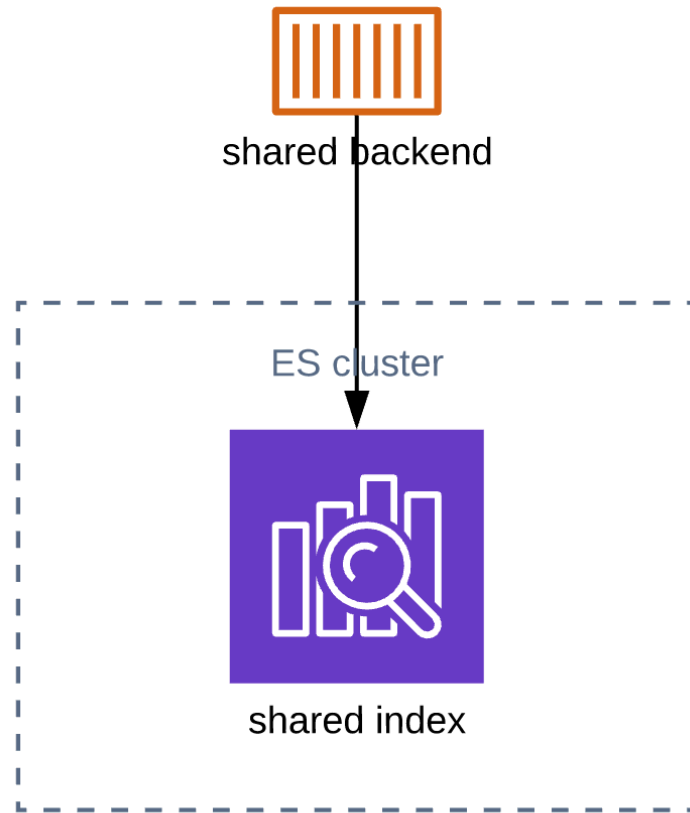
Multitenant indexes

Tenant index mapping – index templates

Rolling updates – aliases

I assume you know ES (indexes, documents, clustering)

Tenancy modes (2 extremes) + example



Tenancy modes (2 extremes)

Shared single index

- Shared functionalities ✓
- Quick initial setup ✓
- High complexity ✗
- Little tenant separation ✗

Separate everything

- High agility ✓
- Best tenant separation ✓
- High maintenance cost ✗
- Unused memory ✗

Best solution

Shared functionalities

High agility

Moderate maintenance costs

Moderate tenant separation

Multitenant index



Let's have tenant on single shard



Indexes up to ES 7.0 have 5 shards, later have 1



Shard allocation can be done using `_routing`

_routing field + example

A document is routed to a particular shard in an index using the following formula:

```
shard_num = hash(_routing) % num_primary_shards
```

The default value used for `_routing` is the document's `_id`.

Custom routing patterns can be implemented by specifying a custom `routing` value per document. For instance:

```
PUT my_index/_doc/1?routing=user1&refresh=true ❶
{
  "title": "This is a document"
}

GET my_index/_doc/1?routing=user1 ❷
```

[COPY AS CURL](#) [VIEW IN CONSOLE](#) 

- ❶ This document uses `user1` as its routing value, instead of its ID.
- ❷ The same `routing` value needs to be provided when [getting](#), [deleting](#), or [updating](#) the document.

Source: <https://www.elastic.co/guide/en/elasticsearch/reference/current/mapping-routing-field.html>

Index template

- How to define mappings (mapping types are obsolete)?
- Let's create many indexes – *data_en*, *data_th*, *data_en_big_tenant*
- *data_en_big_tenant* can have custom schema
- How to manage mappings?
- Create index templates for patterns:
 - **data_**
 - **data_en_big_tenant*

Index template + example

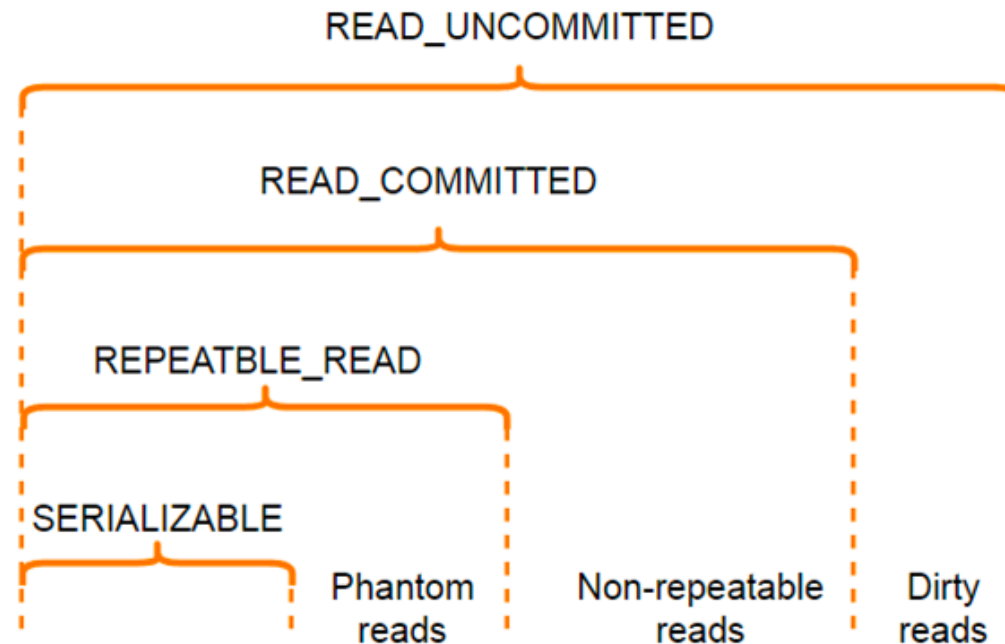
```
PUT _template/template_1
{
  "index_patterns": ["te*", "bar*"],
  "settings": {
    "number_of_shards": 1
  },
  "mappings": {
    "_source": {
      "enabled": false
    },
    "properties": {
      "host_name": {
        "type": "keyword"
      },
      "created_at": {
        "type": "date",
        "format": "EEE MMM dd HH:mm:ss Z yyyy"
      }
    }
  }
}
```

[COPY AS CURL](#) [VIEW IN CONSOLE](#) 

Source: <https://www.elastic.co/guide/en/elasticsearch/reference/current/indices-templates.html>

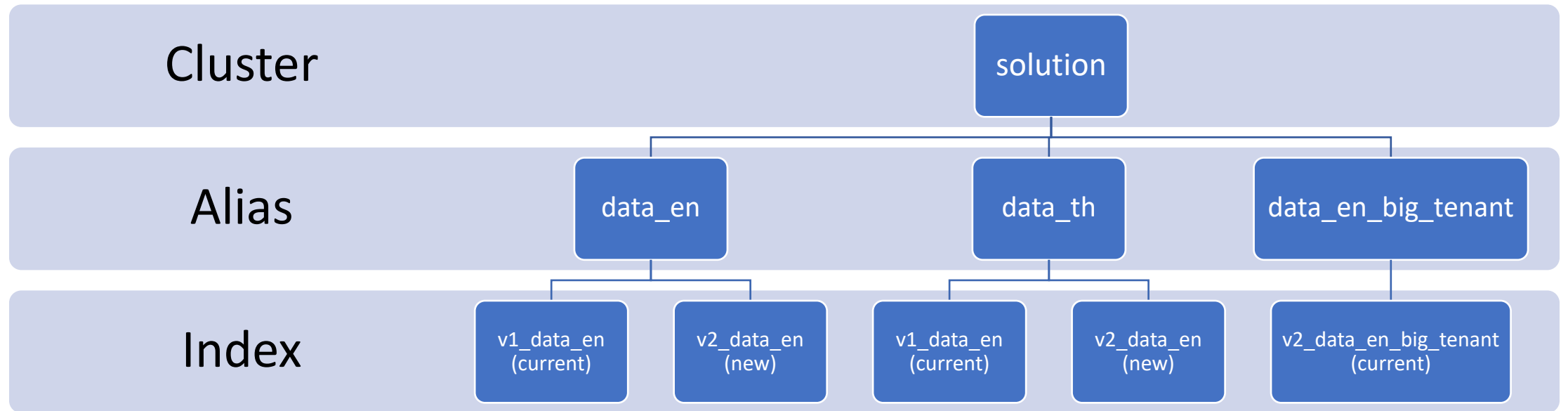
How to manage index updates?

- What is ES isolation level?



Source: <https://www.codeproject.com/Articles/1190420/Database-Transactions-with-Spring-Framework>

Index aliases



Links

- <https://www.elastic.co/blog/found-multi-tenancy>
- https://github.com/meshuga/es_meetup_2019

