#ASLI ENGINEERING

nrtSearch - Yelp's in-house Search Engine

SWIPE

BY

ARPIT BHAYANI

nrt Search : Yelp's search engine on top of lucene

Yelp is a website where
businesses list themselves
people engage with them for services
and post reviews about them

To make discovery better, Yelp used ElasticSearch, but they soon built their own, ... here's why...

Why replace Elasticsearch ?

- Document based replication — Document is indexed independently on every replica and hence scaling out also requires scaling CPUs on replicas.

- uneven load distribution

shard distribution is managed by ES and hence can make load uneven. hence some nodes become hot

Manual rebalancing required for even distribution

- autoscaling is challenging

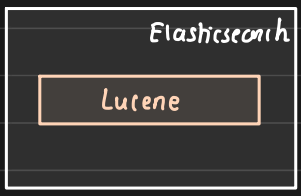
We provision for peak load. Adding new nodes or removing some requires shard migration/movement. It is expensive and non-trivial.

ARPIT BHAYANI

Elasticsearch and lucene

Elasticsearch is built on top of lucene

It is an HTTP server running on top of lucene



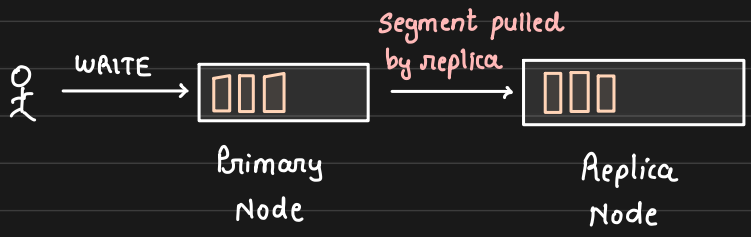
Elasticsearch makes lucene "simple" and "distributed"

↳ replication, sharding, LTR, custom fields and analytics

Two key features of lucene

1. Near-realtime segment replication

lucene has segments → immutable

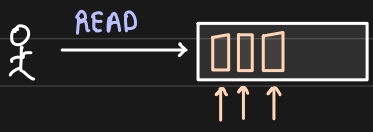


No need to redo the operation on replica.

[no reindexing]

2. Concurrent Search

lucene can search over multiple segments in parallel



leveraging multiple cores.

ARPIT BHAYANI

Implementation

Because they threw out Elasticsearch.

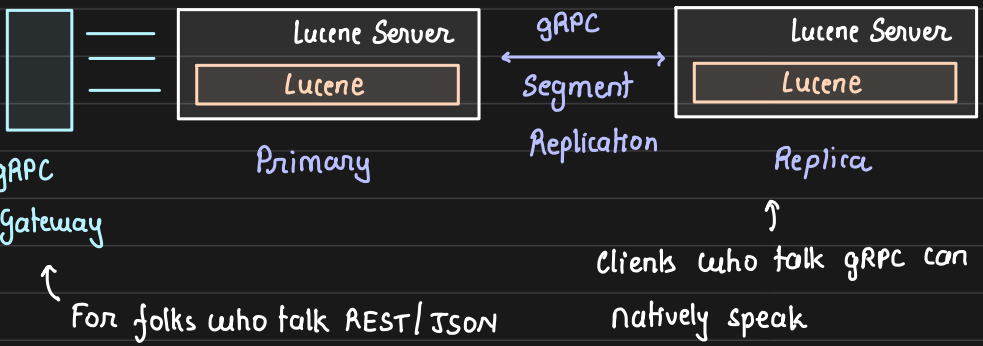
They needed HTTP server over lucene

Lucene Server project

gRPC and protobuf

Replaced REST/JSON based API

with gRPC/protobuf to improve serialization/deserialization performance



Quick Failovers

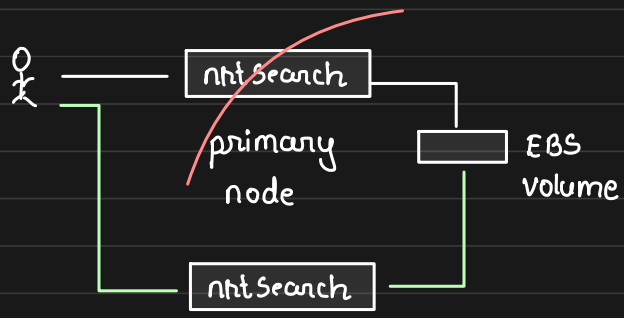
Requirement: When a node fails, time for takeover should be less

Approach 1:

- primary writes to disk
- periodically uploads to s3
- other node on boot, downloads from s3

ARPIT BHAYANI

Approach 2: Attached storage

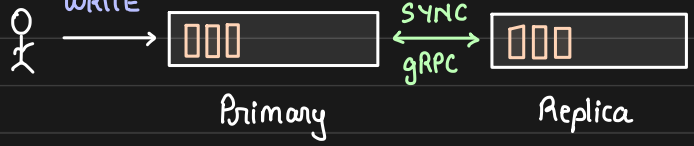


if primary instance crashes or moves to other node, no need to download the index to the new disk, just plug the same EBS.

↳ Very fast recovery / boot

When new replicas start, they can download the latest backup from S3 and start serving.

Primary Replica Sync



downloads latest segments from primary node

(notifies replicas about the new writes)

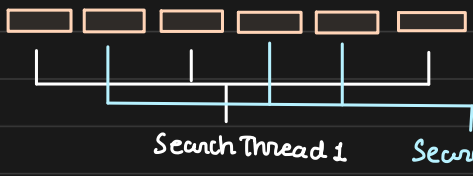
Apart from these, Yelp also added features that would make nrtSearch closer to Elasticsearch.

ARPIT BHAYANI

Performance Improvements

1. Virtual Sharding

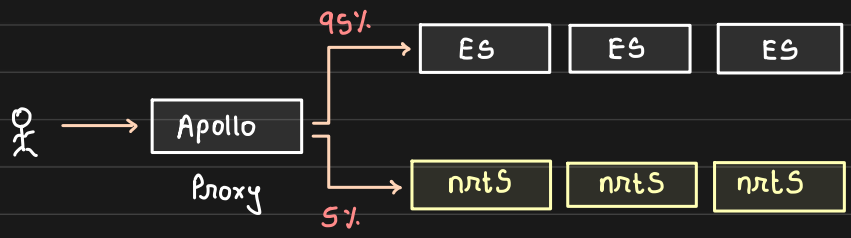
Search request is sent to multiple segments (called slices) in parallel



More even allocation of search threads on these slices (consistent work)

2. Parallel fetching of document fields
3. Segment level search timeout (consistent SLA)

Migration from existing to nrt Search



Dark launch: Send request to both
Send 5% to new as well
} Compare the results for correctness

Once confident then do a phased rollout with exclusive traffic moved to nrt Search → 100%.

ARPIT BHAYANI