


Cs comments service ES usage

 **TNL-9617** - [Discover] Find out extent of cs_comments_service ElasticSearch use
DONE

Discovery work as per the ticket

Questions:

Discover exactly what cs_comments_service with ES, and if that is the only place that ES is used for Infinity

Yes, cs_comments_service is the only place ES is used for Infinity

How much data to we load into ElasticSearch, how do we query it, what type of searches do we do?

Total data in prod-edx-forum = 48.8 GB

Total data in prod-edx-edge = 1.32 GB

Index #1: comment_threads

Index used to search all the comment threads. Queries search comment threads by "title" and "body"

Data in prod-edx-forum = 28.3 GB, Number of documents = 27321310

Data in prod-edx-edge = 714 MB, Number of documents = 526353

Index #2: comments

Index used to search all the comments. Queries search comments by "body"

Data in prod-edx-forum = 20.5 GB, Number of documents = 27882231

Data in prod-edx-edge = 637 MB, Number of documents = 721126

Conclusion:

Although there is a lot of data, the usage of Elastic search in forums is only Vanilla.

How much effort would it take to replace the existing ElasticSearch usage with a more native MySQL implementation?

The code is written in ruby and since I know very little ruby it is very difficult for me to estimate the time or effort required to replace this with a MySQL. Another question would be do we want to make an effort on ruby if the plan is to convert it to python eventually?

Raw Data:



prod-edx-forum-es.txt



prod-edge-forum-es.txt