## **Elasticsearch Merging Test Results**

We did our 'Merge' testing with 2 different sets of settings. This documents lists the steps to do for merging using these and the pros and cons associated with both.

## Settings 1:

1. Create an index with settings as shown below.

```
PUT /test/
{
    "settings": {
        "index": {
            "store": {
                  "type": "default"
            },
            "number_of_shards": 3,[Ideal Value = Number of Team Members]
            "number_of_replicas": 0
        }
}
```

- 2. Merge the documents.
- 3. Once all the documents have been merged, update the index settings to have replicas.

```
PUT /test/_settings
{
    "number_of_replicas":2 [Ideal Value = Number of team members - 1]
}
```

4. Monitor your shards' status. The newly created replicas will be in 'INITIALIZING' status. IMPORTANT: Wait until the status of replica shards change to 'STARTED' state.

```
un _cuej muccesi i
                                                                                         shard prirep state
                                                                                                                 docs store ip
                                                                        .marvel-2015.06.17 0
                                                                                                                 1559 6.4mb 127.0.1.1 Arul
                                                                                                     STARTED
                                                                                             Γ
GET _cluster/health
                                                                        .marvel-2015.06.17 0 p
                                                                                                     STARTED
                                                                                                                 1559 6.1mb 127.0.1.1 Arul_2
GET _cat/nodes?v
                                                                                         0 p
                                                                                                     STARTED
                                                                                                                 3927 619.4mb 127.0.1.1 Arul
                                                                        test
GET _cat/shards?v
                                                                                            ٢
                                                                                         0
                                                                                                     INITIALIZING
                                                                                                                             127.0.1.1 Arul 2
                                                                        test
                                                                                                     INITIALIZING
                                                                                             Γ
                                                                                                                             127.0.1.1 Arul
                                                                        test
                                                                                         1
GET cat/health
                                                                                             P
                                                                                                     STARTED 4136 680.9mb 127.0.1.1 Arul_2
                                                                        test
DELETE /merge_test
                                                                        test
                                                                                                     STARTED
                                                                                                                 3937 646.4mb 127.0.1.1 Arul
                                                                    9
                                                                        test
                                                                                                     INITIALIZING
                                                                                                                             127.0.1.1 Arul_2
GET _cluster/settings
                                                                   10
GET /vsa/ stats
```

## Settings 2:

1. Create an index with number of shards equal to the number of team members and number of replicas equal to the team number count - 1.

```
PUT /test/
{
    "settings": {
        "index": {
            "type": "default"
        },
        "number_of_shards": 3,[Ideal Value = Number of Team Members]
        "number_of_replicas": 2 [Ideal Value = Number of Team Members - 1]
    }
}
```

2. Start your merging program and merge the results.

Settings 1	Settings 2
Pro: Less time to merge.	Con: More time to merge.
Test Results:	Test Results:
<ol> <li>Takes about 20-25 mins to merge 35000 documents.</li> <li>Takes about 20 mins to create replica shards.</li> </ol>	1. Takes about 3-4 hours to merge 25000 documents.
Con:  1.No way to check the status of replica shards' state transformation from 'INTIALIZING' to 'STARTED'.  2.While the state transformation always succeeds, it takes some time over which we have no control or monitoring ability.  Test Results: State Transformation from 'INITIALIZING' to 'STARTED' takes about 6-10 mins for shards of size 700MB.	Pro: No need to create any additional replicas

## **Index Mapping:**

```
PUT /test/document/_mapping
{
 "document": {
  "properties": {
    "docno": {
     "type": "string",
     "store": true,
     "index": "analyzed",
     "term_vector": "with_positions_offsets_payloads"
   },
    "HTTPheader": {
     "type": "string",
     "store": true,
     "index": "not_analyzed"
   "title":{
     "type": "string",
     "store": true,
     "index": "analyzed",
     "term_vector": "with_positions_offsets_payloads"
   },
    "text": {
     "type": "string",
     "store": true,
     "index": "analyzed",
     "term_vector": "with_positions_offsets_payloads"
   },
    "html_Source": {
     "type":"string",
     "store": true,
     "index": "no"
    "in_links":{
     "type": "string",
     "store": true,
     "index": "no"
    "out_links":{
     "type": "string",
     "store": true,
     "index": "no"
   },
```

```
"author":{
    "type": "string",
    "store": true,
    "index": "analyzed"
    }
}
```