CASE STUDY

SQUARESHIFT

Migration Strategy (via Remote Re-indexing) and Index Template Consolidation for a Business Travel Startup based out of California

About the Client

The client is a business travel startup in California that evolved into an agile development and design studio. They build travel solutions used by some of the largest companies on the planet.

Project Context

In the on-premises Elasticsearch environment, the system contained 1,400 indices with 22 unique mappings for similar data across different indices. These inconsistencies in mappings, such as fields representing similar data types (e.g., locations) being defined as either text or keyword, led to inefficiencies and difficulties in managing and querying the data.

The goal of the migration was to address the mapping inconsistencies, improve index management, and resolve oversharding issues while migrating the data to a cloud-based Elasticsearch deployment.

Migration Strategy

Index Template Consolidation:

- The system had 22 unique mappings across indices that contained similar data, with variations in field types and settings. For example, fields representing geographical locations were defined as text in some indices and as keyword in others, which led to inconsistencies.
- A thorough review of these mappings was performed to identify common fields and resolve conflicts in field types.
- A unified schema was created to standardize the mappings for all similar data across indices. This schema ensured that all fields representing similar types of data were consistently defined.
- A new index template was developed to apply to any new indices following a generic pattern. This template ensured that future indices would follow the standardized structure, improving both data consistency and query performance.
- Existing data was restructured to align with the new template using remote reindexing, ensuring that all historical data followed the unified schema.

Remote Reindexing:

- Remote reindexing enabled the seamless migration of data from the on-premises cluster to the cloud-based deployment while aligning the data with the new index templates.
- This process preserved the integrity of the original indices and ensured a smooth transition to the new template with minimal disruption.
- No manual export or re-ingestion was required, and the migration was executed with minimal downtime.

CASE STUDY

SQUARESHIFT

Migration Strategy (via Remote Re-indexing) and Index Template Consolidation for a Business Travel Startup based out of California

Solutions and Recommendations

Delete Empty Indices:
Over 380 empty indices were identified and deleted to free up resources and reduce the shard count, alleviating strain on the cluster.

Optimize Shard Size:
The shard sizes were optimized, targeting a range of 10GB to 50GB for better resource utilization and performance.

Consolidation of Indices:
Indices that contained minimal data (e.g., 1 document) were consolidated into larger indices using the Reindex API, which reduced the overall shard count and improved cluster performance.

Index Lifecycle Management (ILM):
ILM policies were implemented to manage the lifecycle of indices, automate rollovers, and delete older or less useful indices, ensuring that shard counts remained manageable over time.

Shard Reduction:
For smaller indices that could not be consolidated, the number of primary shards was reduced to 1, and the number of replicas was lowered to 0 for indices that did not require redundancy.

Reducing Shards Using Index Aliases:
Index aliases were used to consolidate smaller indices into fewer, larger indices, while maintaining query and application continuity. This allowed the system to reduce the total number of shards without disrupting existing queries or applications.

Ongoing Maintenance:
The system was set up to regularly monitor shard sizes and use consolidation techniques as needed to prevent future oversharding issues.

Outcome

The migration successfully standardized the mappings across all indices, consolidating the 22 unique mappings into a single, unified schema. This ensured that all future data ingested into Elasticsearch followed a consistent structure, improving both performance and ease of management.

By addressing the oversharding issue through consolidation, shard reduction, and the deletion of empty indices, the cluster's resources were optimized, resulting in better query performance and lower operational overhead. The use of remote reindexing allowed for a seamless transition to the cloud deployment with minimal disruption, ensuring that both legacy and new data followed the standardized index structure