ABOUT THE EXAM

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

The exam is for experienced IT professionals who can design, build, deploy, and manage any Elasticsearch solution.

A CLOUD GURU

Exam Topics

Data Management

Define indices and index templates. Create data streams and lifecycle management polices.

Searching Data

Write search queries and aggregations.

Cluster Management

Secure, diagnose, backup, and restore a cluster. Configure cross-cluster search and replication.

Developing Search Applications

Highlight, sort, and paginate search results. Define index aliases. Create search templates.

Data Processing

Define mappings and custom analyzers. Reindex and update documents. Create and use an ingest pipeline.

A CLOUD GURU

Exam Format

Hands-On

Operate in a live environment to complete tasks in real time.

3 Hours

You will have 3 hours to complete the exam, with a 10-minute break.

Remote

Take the exam remotely to avoid having to travel to a testing center.

A CLOUD GURU

Preparing for the Exam

Playground

Use the Cloud Playground to follow along with each lesson.

Hands-On Labs

Be able to complete the hands-on labs with only the Elastic documentation.

Quizzes

Take the available quizzes to reinforce important concepts and terminology.

Practice Exam

Be able to complete the comprehensive practice exam within the time limit.

Documentation

Be familiar with Elastic's documentation. This is your only resource during the exam!

Rinse and Repeat

Revisit any content for topics that you feel weak or unsure about.

ELASTIC STACK OVERVIEW

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

A CLOUD GURU

Beats

1 **Lightweight Data Shippers**

Collect and ship data from the source. There are Beats for just about any kind or source of data you wish to collect.

2 **Built-In Modules**

Pre-built configuration pipelines for well-known use cases.

3 **Extensible with libbeat**

Build your own Beat using the same foundation that all official Elastic Beats are built from.

```
#################### Filebeat Configuration Example #########################

# This file is an example configuration file highlighting only the most common
# options. The filebeat.reference.yml file from the same directory contains all the
# supported options with more comments. You can use it as a reference.
#
# You can find the full configuration reference here:
# https://www.elastic.co/guide/en/beats/filebeat/index.html

# For more available modules and options, please see the filebeat.reference.yml sample
# configuration file.

#============================ Filebeat inputs ===============================

filebeat.inputs:

# Each - is an input. Most options can be set at the input level, so
# you can use different inputs for various configurations.
# Below are the input specific configurations.

- type: log

  # Change to true to enable this input configuration.
  enabled: false

  # Paths that should be crawled and fetched. Glob based paths.
  paths:
    - /var/log/*.log
    #- c:\programdata\elasticsearch\logs\*

  # Exclude lines. A list of regular expressions to match. It drops the lines that are
  # matching any regular expression from the list.
  #exclude_lines: ['^DBG']

  # Include lines. A list of regular expressions to match. It exports the lines that are
  # matching any regular expression from the list.
  #include_lines: ['^ERR', '^WARN']

  # Exclude files. A list of regular expressions to match. Filebeat drops the files that
  # are matching any regular expression from the list. By default, no files are dropped.
  #exclude_files: ['.gz$']

  # Optional additional fields. These fields can be freely picked
  # to add additional information to the crawled log files for filtering
  #fields:
  #  level: debug
  #  review: 1

  ### Multiline options

  # Multiline can be used for log messages spanning multiple lines. This is common
  # for Java Stack Traces or C-Line Continuation
```

A CLOUD GURU

Logstash

Input, parse, enrich, and output live data streams. Logstash is an optional component of the Elastic Stack that provides advanced data processing capabilities.

It is worth noting that Logstash is powerful even outside of the Elastic Stack as a capable data processing tool.

```
input {
  beats {
    id => "filebeat"
    port => 5001
  }
}
filter {
  if [type] == "syslog" {
    grok {
      id => "parse_syslog_event"
      pattern_definitions => { "GREEDY
      match => { "message" => ["%{SYSL
      add_field => { "[@metadata][inde
      remove_field => "message"
    }
    date {
      id => "parse_syslog_date"
      match => [ "[@metadata][timestam
    }
  }
}
output {
  elasticsearch {
    id => "elasticsearch"
    hosts => "localhost:9200"
    index => "%{[@metadata][index]}"
    user => "logstash"
    password => "super_secret_password
  }
}
```

Input data from virtually any source.

Filter, parse, and enrich the data.

Output the processed data to virtually any destination.

Elasticsearch

Most data solutions only perform one part of the data pipeline. Not Elasticsearch.

Process

With Elasticsearch **ingest pipelines**, you can **parse** and **enrich** data as it comes into Elasticsearch.

Index

Elasticsearch's storage is **fast**, **highly scalable**, and **fault tolerant**.

Analyze

Elasticsearch can quickly **search** and **aggregate** data.

A CLOUD GURU

Kibana

Discover

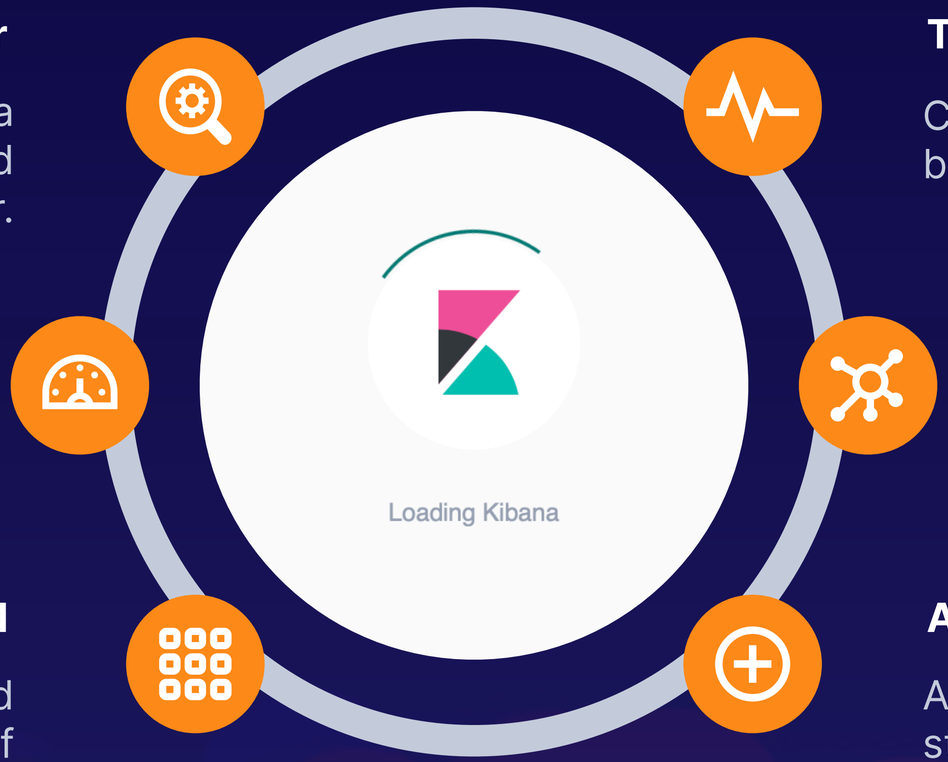
Search, filter, and view data stored in the attached Elasticsearch cluster.

Visualize

Craft informative visualizations from Elasticsearch data.

Dashboard

Combine visualizations and searches into a single pane of glass.



Time Series Visual Builder (TSVB)

Create purpose-built time series-based visualizations.

Machine Learning

Discover and visualize hidden insights and anomalies within Elasticsearch data.

And Much More

APM, uptime monitoring, SIEM, stack monitoring, and more are also possible with Kibana.

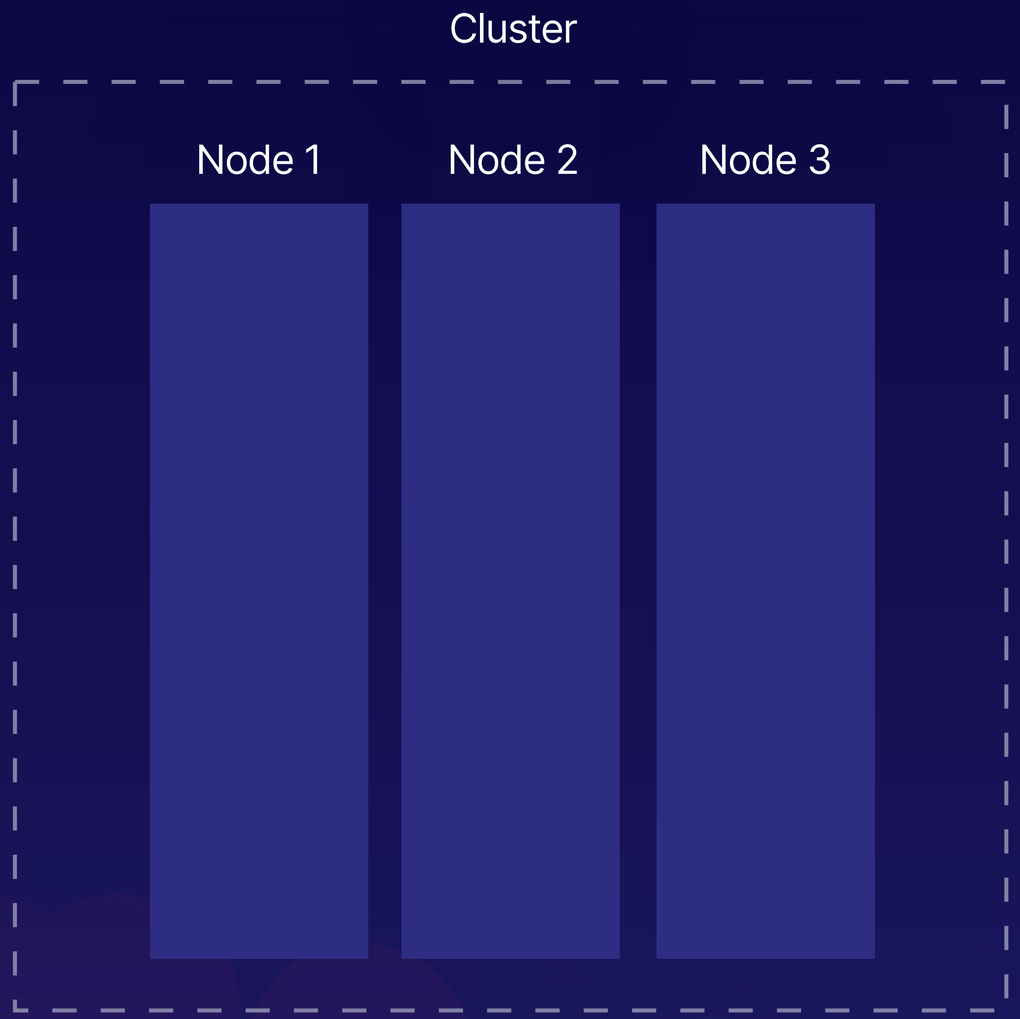
ELASTICSEARCH OVERVIEW

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Cluster

A collection of one or more nodes.

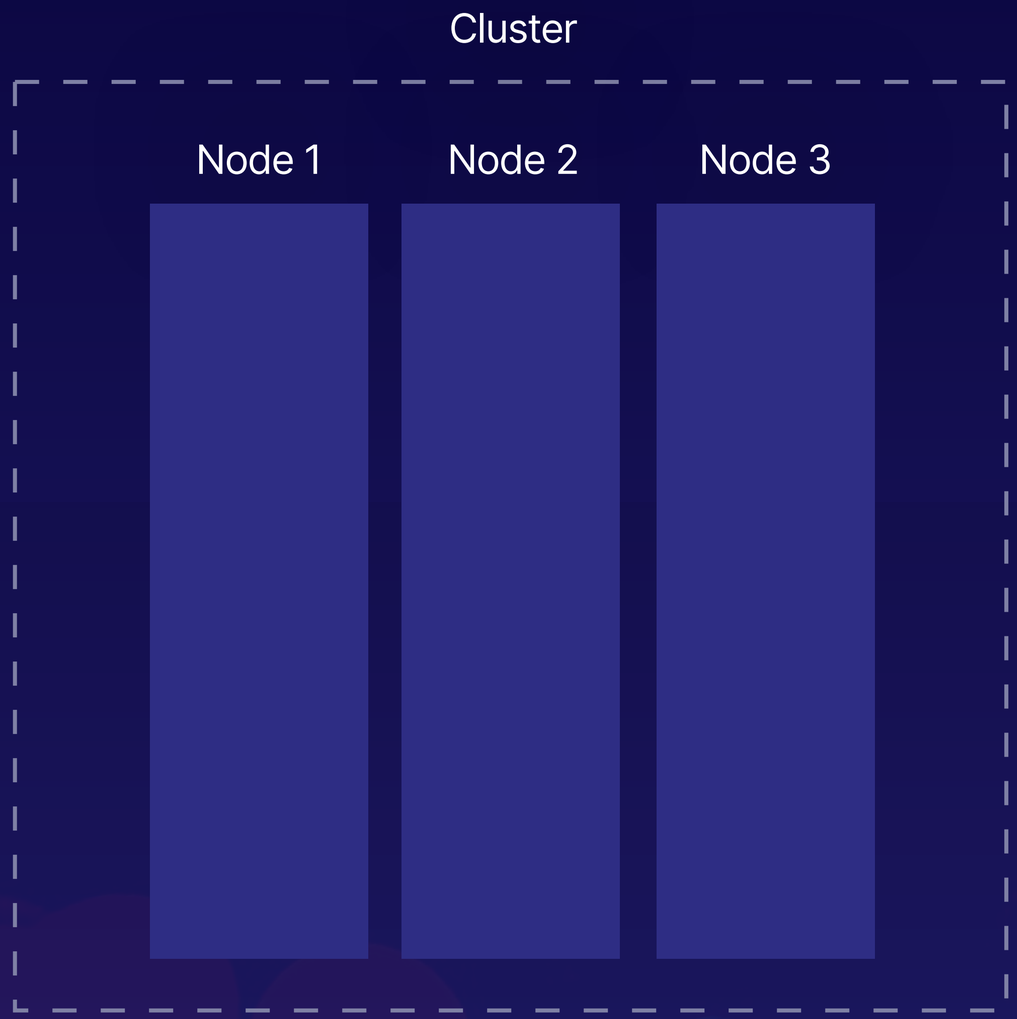
A **cluster** is a distributed entity which allows for horizontal scaling and redundancy.



Nodes

Each node has a set of defined roles.

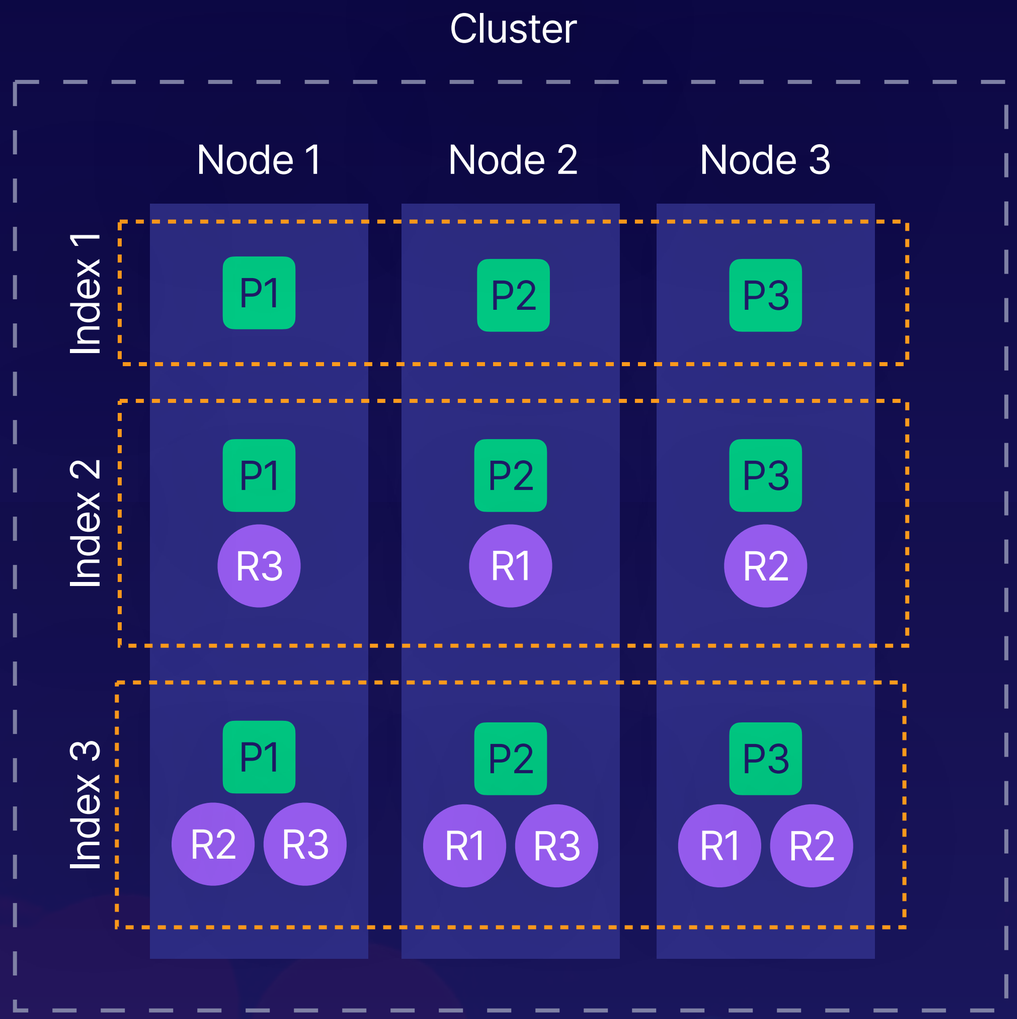
1. **Master-eligible** nodes can be elected as the **master** node which coordinates the cluster.
2. **Data** nodes index, store, and analyze data.
3. **Ingest** nodes can collect and process data.



Indices

A logical structure composed of shards.

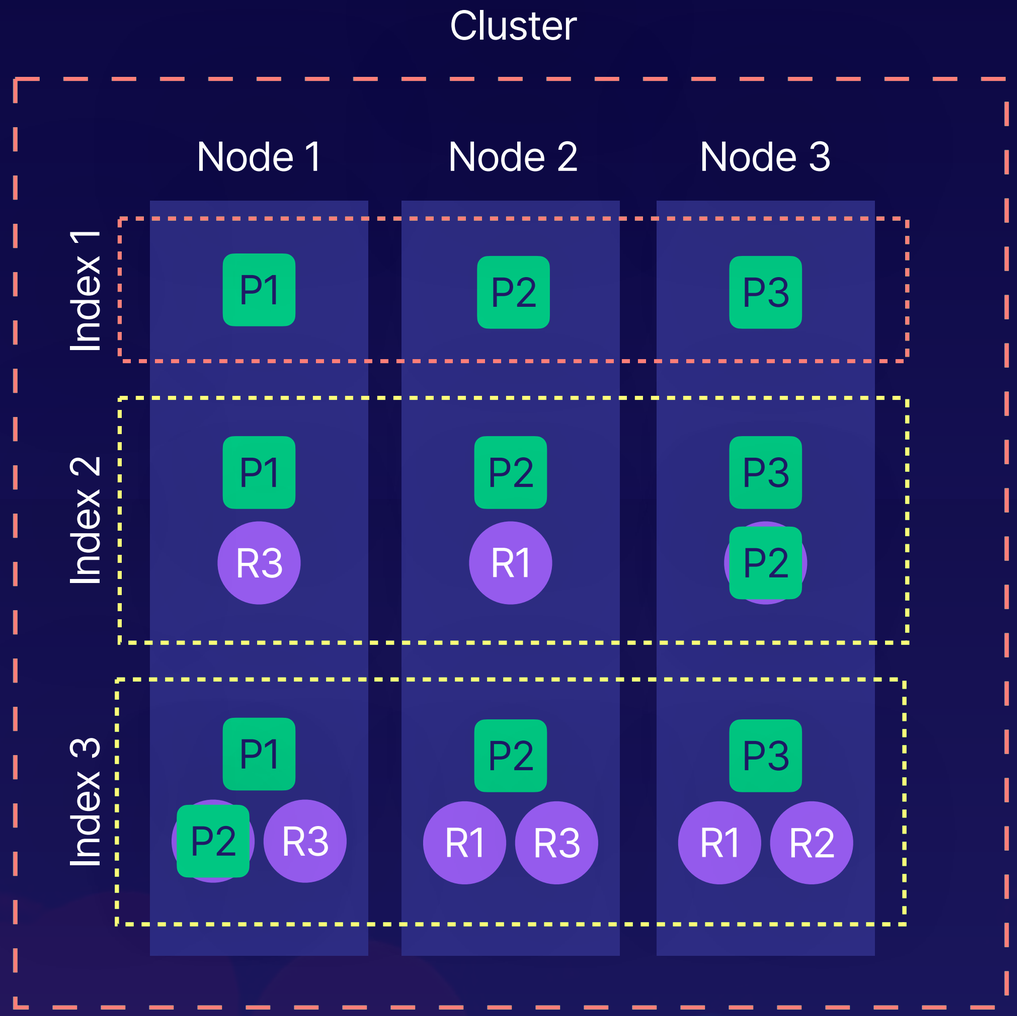
1. **Primary** shards contain a portion of the index.
2. **Replica** shards are copies of a primary shard.
3. Replica shards **cannot** allocate to the same node of their primary shard.
4. Replica shards provide redundancy and increased search throughput.



Cluster States

Based on shard allocation.

1. **Green** state means all primary and replica shards are allocated.
2. **Yellow** state means all primary shards, but **not** all replica shards, are allocated.
3. **Red** state means not all primary shards are allocated.



SETTING UP YOUR OWN ENVIRONMENT

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Setting Up Your Own Environment

A CLOUD GURU

Pre-Built Elastic Certified Engineer Exam Environment

1

Create the Environment

Go to **Cloud Playground** and under **Cloud Servers**, create a new server using the **Elastic Certified Engineer** distribution. The first startup will take some time to build out the environment.

2

Log In to Kibana

Once your cloud server has finished starting up, navigate to its **public IP address** in your local web browser. Then, log in with the username **`elastic`** and password **`elastic_acg`**.

3

Follow Along

We will be using the same environment for all the lessons, so you can follow along each step of the way.

INTRODUCTION TO SEARCHING DATA

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

LESSON BREAKDOWN

Introduction to Managing Data

Defining Indices

Defining Index Templates

Using the Data Visualizer to Upload Data

Establishing an Index Lifecycle Management (ILM) Policy

Creating Data Streams

Myles Young
Training Architect

Setting Up Your Own Environment

A CLOUD GURU

Pre-Built Elastic Certified Engineer Exam Environment

1

Create the Environment

Go to **Cloud Playground** and under **Cloud Servers**, create a new server using the **Elastic Certified Engineer** distribution. The first startup will take some time to build out the environment.

2

Log In to Kibana

Once your cloud server has finished starting up, navigate to its **public IP address** in your local web browser. Then, log in with the username **`elastic`** and password **`elastic_acg`**.

3

Follow Along

We will be using the same environment for all the lessons, so you can follow along each step of the way.

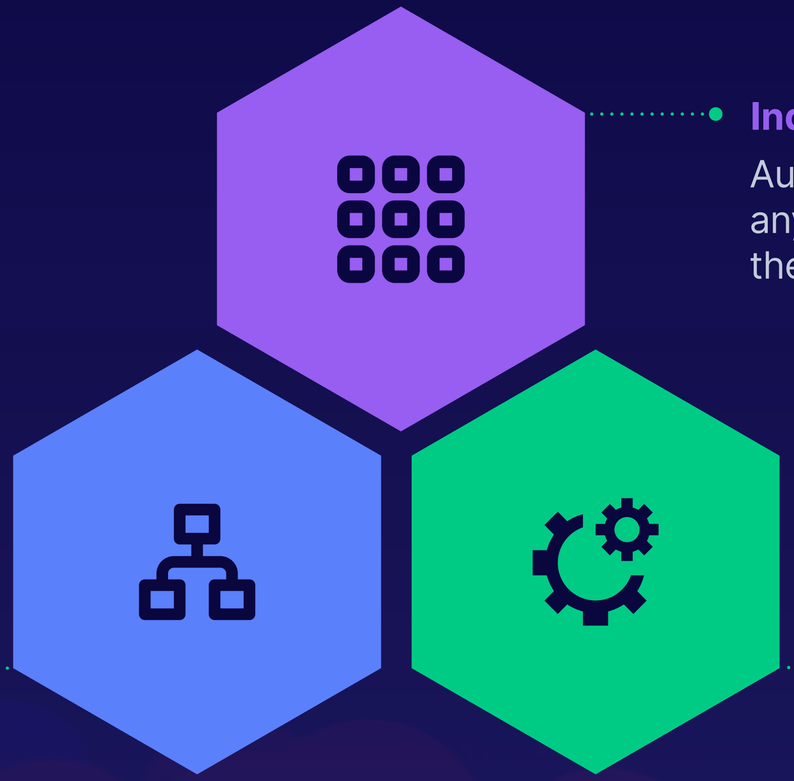
DEFINING INDEX TEMPLATES

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

A CLOUD GURU

Index Templates

Intelligently manage your Elasticsearch indices.



Index Patterns

Automatically manage any index that matches the index pattern.

Component Templates

Reusable building blocks for constructing index templates.

Advanced Features

Data streams, index lifecycle management, snapshotting, etc.

USING THE DATA VISUALIZER TO UPLOAD DATA

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Data Visualizer

Understand the fields of a dataset by quickly uploading and analyzing it in order to determine whether or not to ingest it into Elasticsearch.

Great for ad-hoc data analysis.

Delimited text files (CSV, TSV)

Newline-delimited JSON (NDJSON)

Log files

ESTABLISHING AN INDEX LIFECYCLE MANAGEMENT (ILM) POLICY

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Index Lifecycle Management (ILM)

Automatically manage your indices based on your usage requirements.

Phases

Prioritize your data based on how you use it with hot, warm, cold, and delete phases.

Phase Transitions

Move indices through the lifecycle based on age, size, or document count.

Phase Executions

Perform actions on indices at each phase, like rollover, force merge, migrate, shrink, freeze, and delete.

Index Lifecycle

A CLOUD GURU

Write Optimized

The index is actively being written to and frequently queried.

Storage Optimized

The index is not being written to and is infrequently queried.

Read Optimized

The index is no longer being written to but is still queried.

Retired

The index is no longer needed and can be deleted from the cluster.

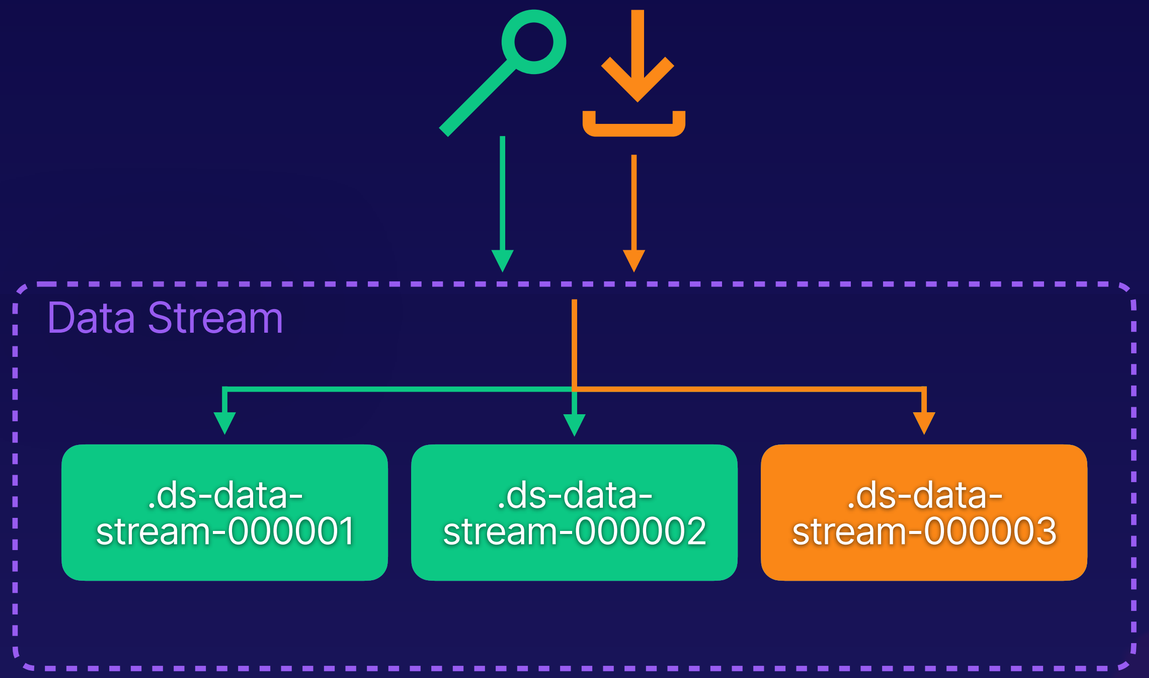
CREATING DATA STREAMS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Data Streaming

Store and search time series data spread across multiple indices with a single resource.

1. Backing indices store the actual data, but the data stream abstracts them.
2. Search requests made to a data stream are directed to all backing indices.
3. Indexing to a data stream gets routed to the latest index in the stream.



SECTION SUMMARY

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Section Summary

Defining Indices

Defining Index Templates

Using the Data Visualizer to Upload Data

Establishing an Index Lifecycle Management (ILM) Policy

Creating Data Streams

Myles Young
Training Architect

INTRODUCTION TO SEARCHING DATA

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

LESSON BREAKDOWN

Introduction to Searching Data

Understanding the Elasticsearch Query DSL

Writing Term-Level Search Queries

Writing Full-Text Search Queries

Writing Compound Search Queries

Executing Asynchronous Search Queries

Executing Cross-Cluster Search Queries

Myles Young
Training Architect

A CLOUD GURU

Setting Up Your Own Environment

A CLOUD GURU

Pre-Built Elastic Certified Engineer Exam Environment

1

Create the Environment

Go to **Cloud Playground** and under **Cloud Servers**, create a new server using the **Elastic Certified Engineer** distribution. The first startup will take some time to build out the environment.

2

Log In to Kibana

Once your cloud server has finished starting up, navigate to its **public IP address** in your local web browser. Then, log in with the username **`elastic`** and password **`elastic_acg`**.

3

Follow Along

We will be using the same environment for all the lessons, so you can follow along each step of the way.

UNDERSTANDING THE ELASTICSEARCH QUERY DSL

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

A CLOUD GURU

Storing Analyzed Fields

Analyzed

Original Text

The students learned a NEW concept.

↓

Tokenized Text

students | learned | NEW | concept

↓

Normalized Tokens

student | learn | new | concept

VS

Non-Analyzed

Original Text

The students learned a NEW concept.

↓

Tokenized Text

The students learned a NEW concept.

⇣

Normalized Token

the students learned a new concept.

A CLOUD GURU

Searching Analyzed Fields

Analyzed

Stored Field

student | learn | new | concept

Matched Searches

Students.

New concepts.

How to learn?

Learning new ideas.

Never stop learning!

Newer concepts.

Training new students.

Students learning new concepts.

Learning.

Understanding fresh ideas.

VS

Non-Analyzed

Stored Field

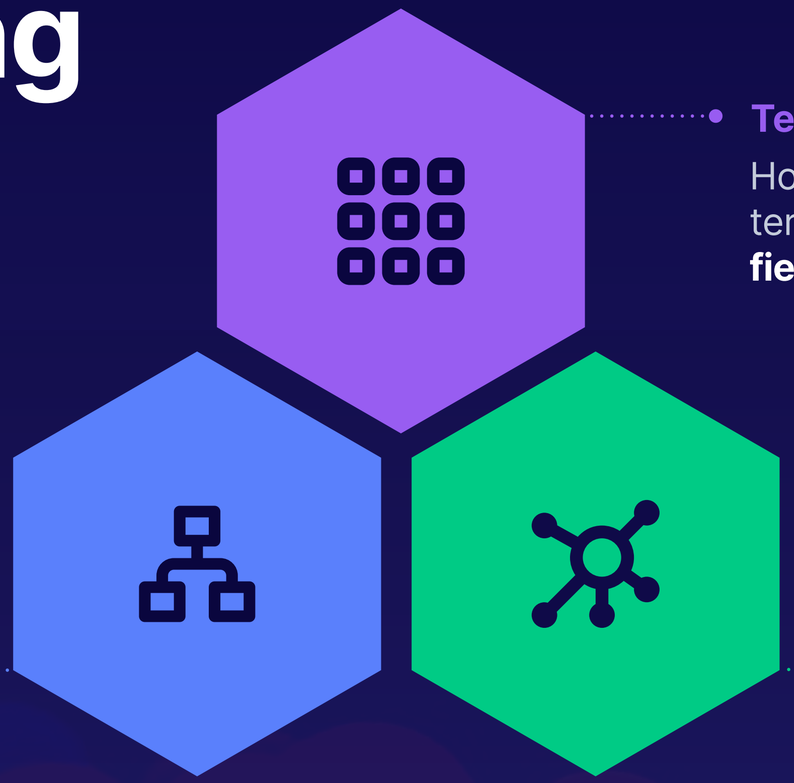
The students learned a NEW concept.

Matched Searches

The students learned a NEW concept.

Relevancy Scoring

How well a search matches a document.



Term Frequency

How often does the term appear in the **field**?

Field-Length Normalization

How many terms are in the field?

Inverse Document Frequency

How often does the term appear in the **index**?

Query and Filter Context

QUERY CONTEXT

Q: How well does the search clause match the document?

A: `<relevancy_score>`

VS

FILTER CONTEXT

Q: Does the search clause match the document?

A: `true` or `false`

WRITING TERM-LEVEL SEARCH QUERIES

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Term-Level Queries

Search for documents based on exact values.

1. Term-level queries do not analyze search terms.
2. Search terms can be normalized for term-level queries.
3. Avoid using term-level queries on analyzed fields.

```
# No results
GET shakespeare/_search
{
  "query": {
    "term": {
      "text_entry.keyword": {
        "value": "yes"
      }
    }
  }
}

# 8 results
GET shakespeare/_search
{
  "query": {
    "term": {
      "text_entry.keyword": {
        "value": "Yes."
      }
    }
  }
}
```

WRITING FULL-TEXT SEARCH QUERIES

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Full-Text Queries

Search for documents based on analyzed values.

1. The query is processed using the same analyzer as the text field.
2. Text fields can be indexed and searched using custom analyzers.
3. Avoid using full-text queries on non-analyzed fields.

```
# 199 results
GET shakespeare/_search
{
  "query": {
    "match": {
      "text_entry": "yes"
    }
  }
}

# 199 results
GET shakespeare/_search
{
  "query": {
    "match": {
      "text_entry": "Yes."
    }
  }
}

```

WRITING COMPOUND SEARCH QUERIES

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Boolean Query

Search for documents matching a boolean combination of queries.

With boolean queries, the scores of **`must`** and **`should`** clauses will be added together to calculate the final relevancy score for each result.

Clause	Description
`must`	The search term **must** appear and is scored.
`should`	The search term **should** appear and is scored. You can configure how many `should` clauses must match.
`must_not`	The search term **must not** appear and is not scored.
`filter`	The search term **must** appear but is not scored.

EXECUTING ASYNCHRONOUS SEARCH QUERIES

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Async Search

A CLOUD GURU

Write the query

`_async_search` accepts the same parameters and body as the `_search` API.

Get search results

Search results can be fetched while `_async_search` executes and after it completes.

Submit → **Check** → **Results** → **Delete**

Check search status

`_async_search` requests return an ID that can be used to check the search status.

Delete the search

Search results can be deleted when no longer needed.

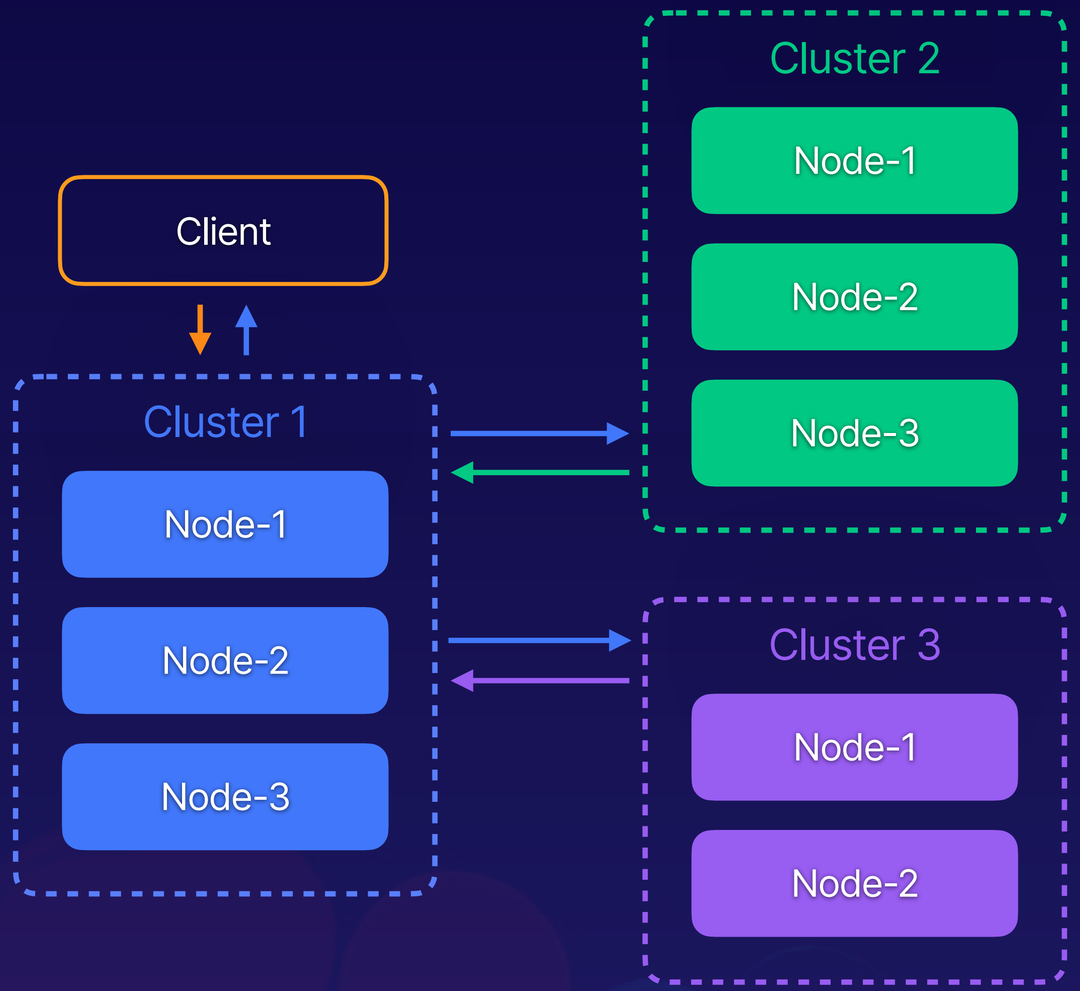
EXECUTING CROSS-CLUSTER SEARCH QUERIES

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Cross-Cluster Search

Run a search request against 1 or more remote clusters.

You can search remote clusters that are 1 major version behind or ahead of the coordinating node.



SECTION SUMMARY

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Section Summary

Understanding the Elasticsearch Query DSL

Writing Term-Level Search Queries

Writing Full-Text Search Queries

Writing Compound Search Queries

Executing Asynchronous Search Queries

Executing Cross-Cluster Search Queries

Myles Young
Training Architect

INTRODUCTION TO AGGREGATING DATA

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

LESSON BREAKDOWN

Introduction to Aggregating Data

Writing Metrics Aggregations

Writing Bucket Aggregations

Writing Sub-Aggregations

Writing Pipeline Aggregations

Myles Young
Training Architect

A CLOUD GURU

Setting Up Your Own Environment

A CLOUD GURU

Pre-Built Elastic Certified Engineer Exam Environment

1

Create the Environment

Go to **Cloud Playground** and under **Cloud Servers**, create a new server using the **Elastic Certified Engineer** distribution. The first startup will take some time to build out the environment.

2

Log In to Kibana

Once your cloud server has finished starting up, navigate to its **public IP address** in your local web browser. Then, log in with the username **`elastic`** and password **`elastic_acg`**.

3

Follow Along

We will be using the same environment for all the lessons, so you can follow along each step of the way.

WRITING METRICS AGGREGATIONS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Metrics Aggregations

Computes numeric values.

Metrics aggregations are either single or multi-value aggregations that can operate on a variety of **non-analyzed** fields to produce a numerical value.

```
# single-value metrics aggregation
GET ecommerce/_search
{
  "size": 0 # simplify the output
  "aggs": {
    "total_sales": { # any name
      "sum": { # metrics agg
        "field": "taxless_total_price"
      }
    }
  }
}
```

WRITING BUCKET AGGREGATIONS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Bucket Aggregations

Creates buckets of documents.

Establish a criterion to categorize documents into groups or buckets.

```
# bucket aggregation
GET ecommerce/_search
{
  "size": 0, # simplify the output
  "aggs": {
    "orders_per_day": { # any name
      "date_histogram": { # bucket agg
        "field": "order_date",
        "calendar_interval": "day"
      }
    }
  }
}
```

WRITING SUB-AGGREGATIONS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

```
GET ecommerce/_search
{
  "size": 0, # simplify the output
  "aggs": {
    "total_sales_per_day": { # any name
      "date_histogram": { # bucket agg
        "field": "order_date",
        "calendar_interval": "day"
      },
      "aggs": { # sub-agg
        "total_sales": { # any name
          "sum": { # metrics agg
            "field": "taxless_total_price"
          }
        }
      }
    }
  }
}
```

Sub-Aggregations

Aggregates per bucket.

Each bucket of a parent pipeline aggregation can have sub-aggregations performed on it.

WRITING PIPELINE AGGREGATIONS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

```
GET flights/_search
{
  "size": 0,
  "aggs": {
    "delay_minutes_per_day": {
      "date_histogram": { # parent agg
        "field": "timestamp",
        "calendar_interval": "day"
      },
      "aggs": {
        “delay_minutes”: {
          "sum": { # sub-agg
            "field": "FlightDelayMin"
          }
        },
        “cumulative_delay_minutes": {
          "cumulative_sum": { # pipeline agg
            "buckets_path": "delay_minutes"
          }
        }
      }
    }
  }
}
```

Parent Pipeline Aggregations

Takes the output of a parent aggregation.

Using the output of a parent aggregation, parent pipeline aggregations create new buckets or new values for existing buckets.

```
GET flights/_search
{
  "size": 0,
  "aggs": {
    "miles_per_day": {
      "date_histogram": { # parent agg
        "field": "timestamp",
        "calendar_interval": "day"
      },
      "aggs": {
        “total_miles": {
          "sum": { # sub-agg
            "field": "DistanceMiles"
          }
        }
      }
    },
    “most_miles_per_day": {
      "max_bucket": { # pipeline agg
        "buckets_path": “miles_per_day>total_miles"
      }
    }
  }
}
```

Sibling Pipeline Aggregations

Takes the output of a sibling aggregation.

Using the output of a sibling aggregation, sibling pipeline aggregations create new outputs at the same level as the sibling aggregation.

SECTION SUMMARY

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Section Summary

Writing Metrics Aggregations

Writing Bucket Aggregations

Writing Sub-Aggregations

Writing Pipeline Aggregations

Myles Young
Training Architect

A CLOUD GURU

INTRODUCTION TO DEVELOPING SEARCH APPLICATIONS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

LESSON BREAKDOWN

Introduction to Developing Search Applications

Highlighting Search Terms

Sorting Search Results

Paginating Search Results

Defining Index Aliases

Defining Search Templates

Myles Young
Training Architect

A CLOUD GURU

Setting Up Your Own Environment

A CLOUD GURU

Pre-Built Elastic Certified Engineer Exam Environment

1

Create the Environment

Go to **Cloud Playground** and under **Cloud Servers**, create a new server using the **Elastic Certified Engineer** distribution. The first startup will take some time to build out the environment.

2

Log In to Kibana

Once your cloud server has finished starting up, navigate to its **public IP address** in your local web browser. Then, log in with the username **`elastic`** and password **`elastic_acg`**.

3

Follow Along

We will be using the same environment for all the lessons, so you can follow along each step of the way.

HIGHLIGHTING SEARCH TERMS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Highlight

Emphasize the search term(s).

Customize the way Elasticsearch highlights matched search term(s).

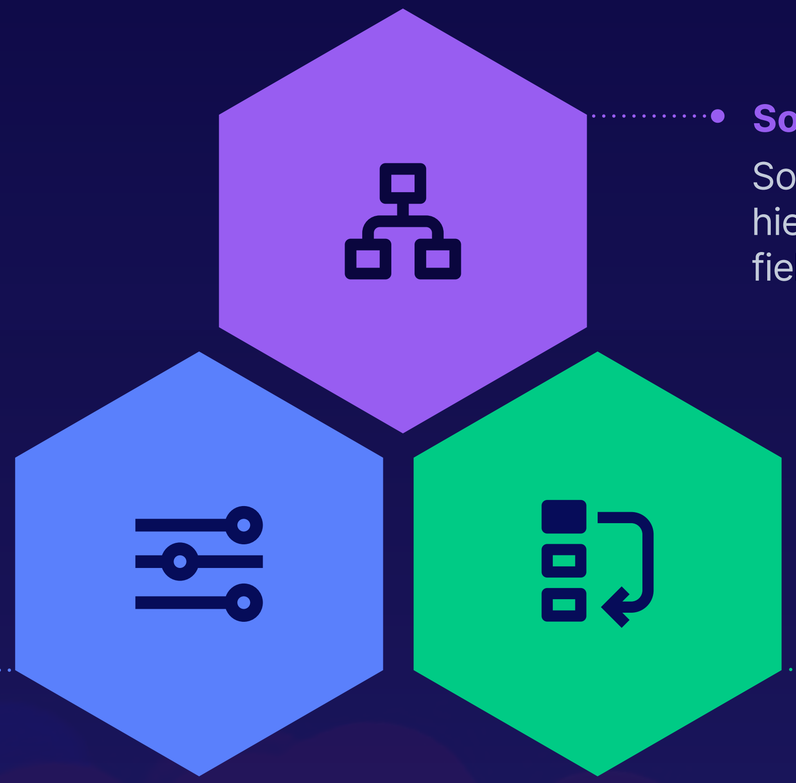
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris quis metus lacus. Praesent interdum aliquam libero at pretium. Suspendisse nisl nisl, porta quis porta et, sagittis vel nunc. Praesent laoreet arcu in odio lobortis, vitae efficitur ipsum tincidunt. In at erat sit amet velit feugiat suscipit vitae in erat. Aliquam ut odio maximus, laoreet lorem ac, luctus libero. Morbi dignissim ullamcorper pharetra. Fusce tempus eros id lobortis vehicula. Pellentesque nisi metus, venenatis quis ligula eget, posuere posuere mauris. Fusce mattis elit quis ex malesuada, eget feugiat felis tempor. Nunc congue auctor lacus, at rhoncus neque faucibus sit amet. Aenean mollis diam sed feugiat viverra. Mauris pellentesque justo quis gravida tempor.

SORTING SEARCH RESULTS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Sort

Meaningfully organize search results.



Sort Values

Sort on the values of a hierarchy of specified fields.

Sort Mode

For multi-value fields, sort by `min`, `max`, `sum`, `average`, or `median`.

Sort Order

Determine whether it should be ascending or descending.

PAGINATING SEARCH RESULTS

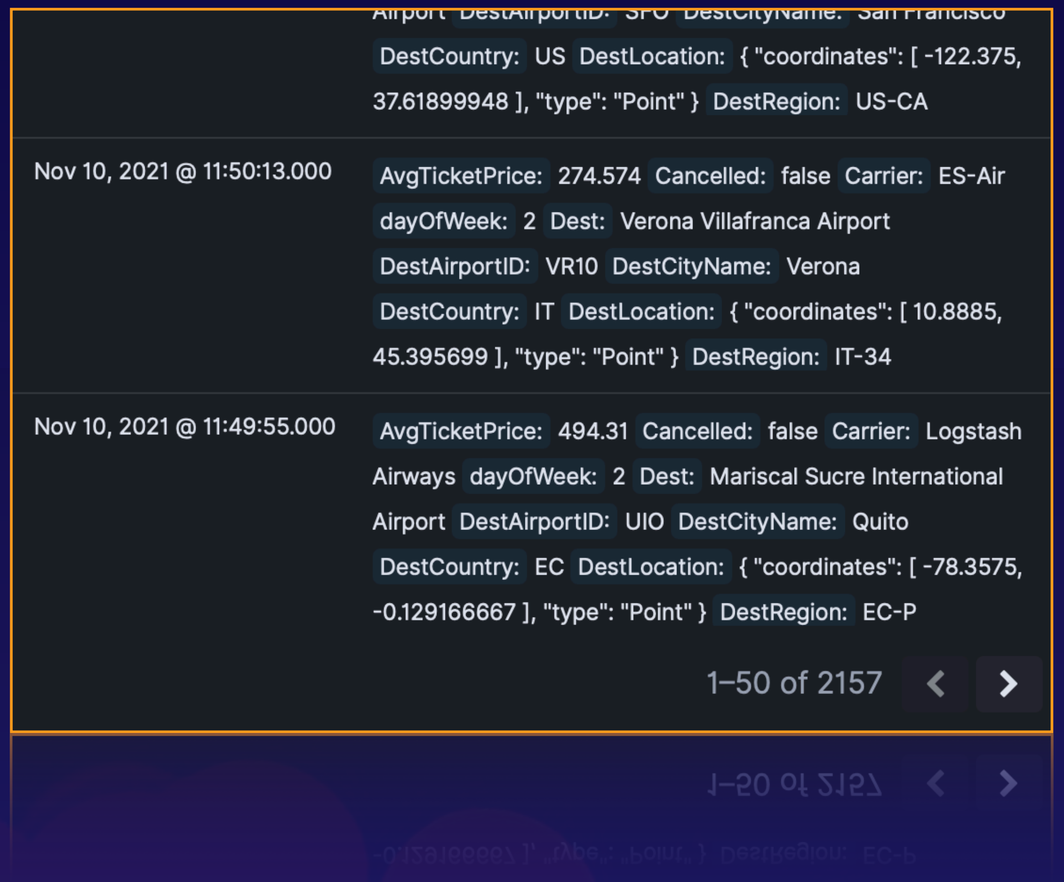
A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

A CLOUD GURU

Paginate

Split search results into discrete pages.

Having too many search results displayed at once can be overwhelming. More often than not, what you actually care about is on the first page.



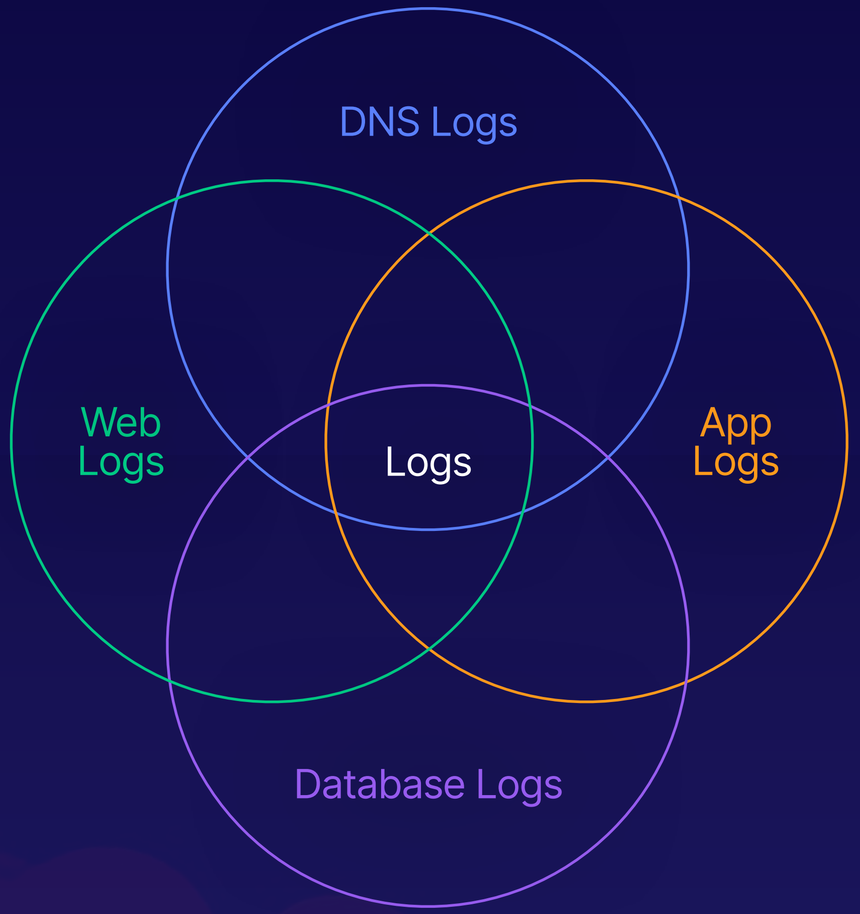
DEFINING INDEX ALIASES

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Aliases

Simplify data referencing.

Search across multiple indices with ease by assigning meaningful aliases to indices manually or automatically using index or component templates.



DEFINING SEARCH TEMPLATES

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Search Templates

Reusable parameterized queries.

When using Elasticsearch as a backend for a search application, search templates enable you to pass user inputs to your search without exposing the whole query.

First Name	John
Last Name	Doe
Age	34
Email	john.doe@gmail.com
Phone Number	1+(234)-567-8900

SECTION SUMMARY

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Section Summary

Highlighting Search Terms

Sorting Search Results

Paginating Search Results

Defining Index Aliases

Defining Search Templates

Myles Young
Training Architect

A CLOUD GURU

INTRODUCTION TO DATA PROCESSING

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

LESSON BREAKDOWN

Introduction to Data Processing

- Explicitly Mapping Fields
- Dynamically Mapping Fields
- Defining a Custom Analyzer
- Defining Multi-Fields
- Reindexing Documents
- Updating Documents
- Defining Ingest Pipelines
- Handling Nested Arrays of Objects

Myles Young
Training Architect

A CLOUD GURU

Setting Up Your Own Environment

A CLOUD GURU

Pre-Built Elastic Certified Engineer Exam Environment

1

Create the Environment

Go to **Cloud Playground** and under **Cloud Servers**, create a new server using the **Elastic Certified Engineer** distribution. The first startup will take some time to build out the environment.

2

Log In to Kibana

Once your cloud server has finished starting up, navigate to its **public IP address** in your local web browser. Then, log in with the username **`elastic`** and password **`elastic_acg`**.

3

Follow Along

We will be using the same environment for all the lessons, so you can follow along each step of the way.

EXPLICITLY MAPPING FIELDS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

A CLOUD GURU

Explicit Mapping

Manually define the fields and data types.

1. Determine which string fields are analyzed.
2. Define which numbers should be integers, floats, percents, etc.
3. Customize the date format for date fields.

```
{
  "kibana_sample_data_flights" : {
    "mappings" : {
      "properties" : {
        "AvgTicketPrice" : {
          "type" : "float"
        },
        "Cancelled" : {
          "type" : "boolean"
        },
        "Carrier" : {
          "type" : "keyword"
        },
        "Dest" : {
          "type" : "keyword"
        },
        "DestAirportID" : {
          "type" : "keyword"
        },
        "DestCityName" : {
          "type" : "keyword"
        },
        "DestCountry" : {
          "type" : "keyword"
        },
        "DestLocation" : {
          "type" : "geo_point"
        },
        "DestRegion" : {
          "type" : "keyword"
        },
        "DestWeather" : {
          "type" : "keyword"
        },
        "DistanceKilometers" : {
          "type" : "float"
        },
```

DYNAMICALLY MAPPING FIELDS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Dynamic Mapping

Automatically add new fields and data types.

1. New fields not already in the index will be added automatically.
2. Data detectors will automatically determine data types for new fields.
3. Dynamic templates allow for the customization of dynamic mapping behavior.

```
{
  "component_templates" : [
    {
      "name" : "strings_as_keywords",
      "component_template" : {
        "template" : {
          "mappings" : {
            "dynamic_templates" : [
              {
                "strings_as_keywords" : {
                  "mapping" : {
                    "ignore_above" : 256,
                    "type" : "keyword"
                  },
                  "match_mapping_type" : "string"
                }
              }
            ]
          }
        }
      }
    }
  ]
}
```

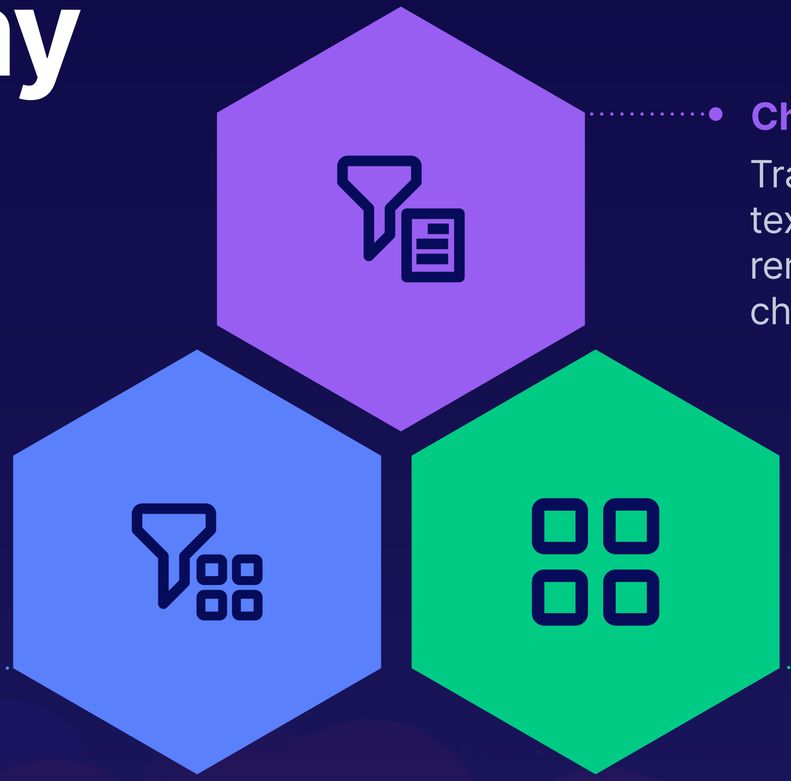
DEFINING A CUSTOM ANALYZER

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

A CLOUD GURU

Analyzer Anatomy

Analyzers allow Elasticsearch to return relevant results rather than exclusively exact matches.



Character Filter

Transforms a string of text by adding, removing, or changing characters.

Token Filters

Transforms a token by adding, removing, or changing it.

Tokenizer

Converts a string of text into an array of individual tokens.

DEFINING MULTI-FIELDS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Multi-Fields

Index the same field in multiple ways.

Multi-fields allow you to index fields multiple times. This way, you can have one field mapped in different ways depending on the potential use cases.

Map a string as both an analyzed `text` field and non analyzed `keyword` field.

Map an analyzed `text` field with multiple different analyzers.

REINDEXING DOCUMENTS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Reindexing Documents

1

Source

Specify the local or remote source index and filter the source with a query.

2

Destination

Specify the local index to reindex into.

3

Pipeline

Specify an ingest pipeline to modify the data in flight before it's written to the destination index.

UPDATING DOCUMENTS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Updating Documents

Update by document ID or query.

The `update` and `update_by_query` APIs enable the modification of documents after they have already been indexed.

Update documents to pick up mapping changes.

Update documents with a script to change their source values.

Update documents with an ingest pipeline.

DEFINING INGEST PIPELINES

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Ingest Pipelines

Process and enrich your data.

1. Pipelines use processors to perform some action on a document.
2. Processors are executed in order.
3. Ingest pipelines can be used in the `update_by_query` and `reindex` APIs.

```
PUT _ingest/pipeline/migrate_accounts
{
  "description": "refactoring accounts dataset and adding a bonus five percent",
  "processors": [
    {
      "remove": {
        "field": "account_number"
      }
    },
    {
      "set": {
        "field": "fullname",
        "value": "{{firstname}} {{lastname}}"
      }
    },
    {
      "script": {
        "description": "adds a bonus and increments the bonus counter",
        "source": """
          ctx.balance += ctx.balance * 0.05;
          if (ctx.bonus_pct == null) {
            ctx.bonus_pct = 5;
          } else {
            ctx.bonus_pct += 5;
          }
        """
      }
    }
  ]
}
```

HANDLING NESTED ARRAYS OF OBJECTS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Nested Arrays of Objects

Maintain the independence of objects in an array.

1. Arrays of objects are flattened, thereby losing the relationships of which values belong to which object.

2. The `nested` data type maintains the independence of each object in an array.

3. The `nested` query searches object arrays as if each object were a separate document.

```
# nested array of objects
PUT courses/_doc/1
{
  "name" : "Elasticsearch for Beginners",
  "student" : [
    { "first" : "Myles", "last" :  "Young" },
    { "first" : "Michael", "last" :  "Bender" }
  ]
}

# matches when not using a nested datatype
GET my-index-000001/_search
{
  "query": {
    "bool": {
      "must": [
        { "match": { "student.first": "Myles" }},
        { "match": { "student.last":  "Bender" }}
      ]
    }
  }
}
```

SECTION SUMMARY

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Section Summary

- Explicitly Mapping Fields
- Dynamically Mapping Fields
- Defining a Custom Analyzer
- Defining Multi-Fields
- Reindexing Documents
- Updating Documents
- Defining Ingest Pipelines
- Handling Nested Arrays of Objects

Myles Young
Training Architect

INTRODUCTION TO CLUSTER MANAGEMENT

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

LESSON BREAKDOWN

Introduction to Cluster Management

Diagnosing Shard Issues

Snapshotting Data

Configuring Searchable Snapshots

Restoring Data

Setting Up Cross-Cluster Replication

Defining Access Control

Myles Young
Training Architect

A CLOUD GURU

Setting Up Your Own Environment

A CLOUD GURU

Pre-Built Elastic Certified Engineer Exam Environment

1

Create the Environment

Go to **Cloud Playground** and under **Cloud Servers**, create a new server using the **Elastic Certified Engineer** distribution. The first startup will take some time to build out the environment.

2

Log In to Kibana

Once your cloud server has finished starting up, navigate to its **public IP address** in your local web browser. Then, log in with the username **`elastic`** and password **`elastic_acg`**.

3

Follow Along

We will be using the same environment for all the lessons, so you can follow along each step of the way.

DIAGNOSING SHARD ISSUES

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Cluster Allocation Explain API

Make the cluster explain itself.

This API will provide a detailed explanation as to why something is allocated or not allocated.

Providing no parameters will explain the first **unassigned** shard.

Providing an index and a shard will explain the allocation of that shard.

SNAPSHOTTING DATA

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

A CLOUD GURU

Snapshotting Data

1

Register

Register a snapshot repository.

2

Snapshot

Create a snapshot of 1 or more indices.

3

Automate

Use snapshot lifecycle management (SLM).

CONFIGURING SEARCHABLE SNAPSHOTS

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Searchable Snapshots

Offload data while still being able to search it.

1. No replica shards by default. The snapshot repository is responsible for redundancy.

2. Replica shards can be configured in order to increase search throughput.

3. Can be configured through index lifecycle management (ILM) automatically or the `_mount` API manually.

```
# automatically with ilm
PUT _ilm/policy/my_policy
{
  "policy": {
    "phases": {
      "cold": {
        "actions": {
          "searchable_snapshot" : {
            "snapshot_repository" : "my_repo"
          }
        }
      }
    }
  }
}

# manually with _mount api
POST _snapshot/my_repo/my_snapshot
{
  "index": "my_index",
  "renamed_index": "my_index_backup"
}
```

RESTORING DATA

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Restoring Data

1

Restore

Restore 1 or more indices from a snapshot.

2

Rename

Restored indices can be renamed during the restore process.

3

Cross-Cluster

Restore a snapshot from a different cluster.

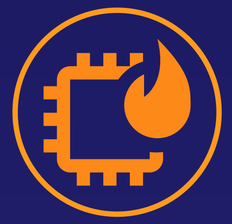
SETTING UP CROSS-CLUSTER REPLICATION

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Cross-Cluster Replication

A leader indexes and a follower replicates.

Configure a remote cluster and follow 1 or more indices from it. Follower indices will replicate all actions performed on the remote leader index.



Disaster recovery

Data locality

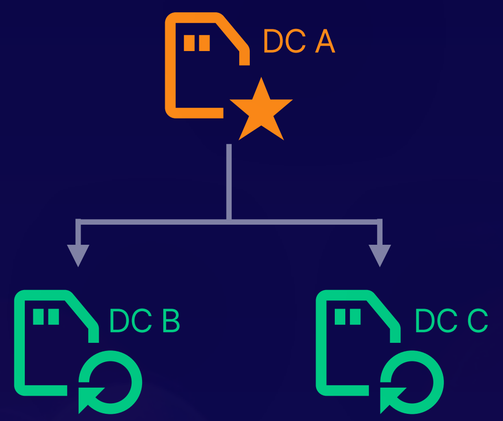
Centralized reporting

A CLOUD GURU

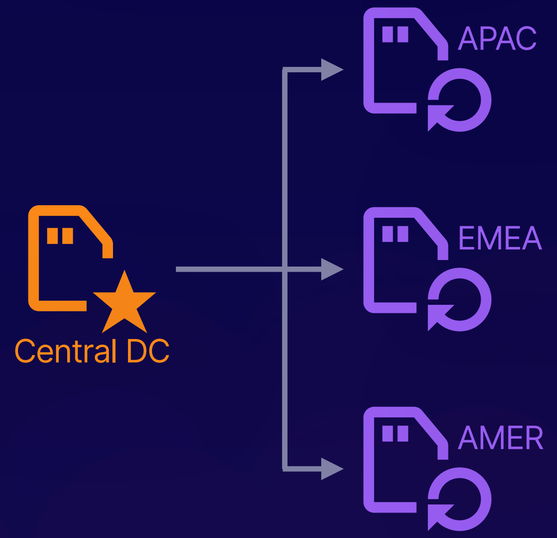
Cross-Cluster Replication

Basic Architectures

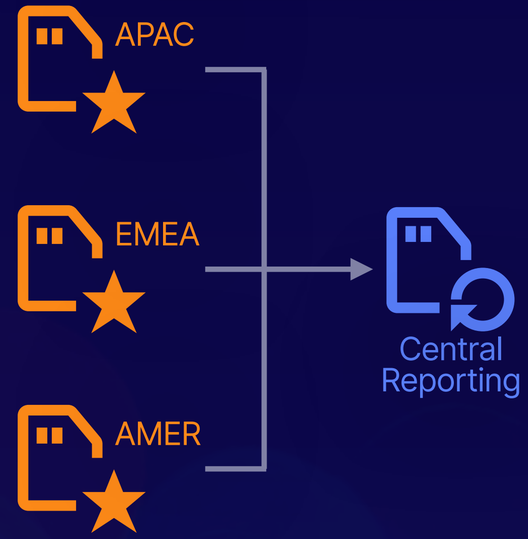
Disaster Recovery



Data Locality



Centralized Reporting



DEFINING ACCESS CONTROL

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

User Access Control

Authenticate and authorize.

Prevent access to your data with a variety of authentication realms. Limit access to your data with roles.

Authenticate

Built-In Realms

- `native`
- `ldap`
- `active_directory`
- `pki`
- `file`
- `saml`
- `oidc`

Custom Realm

Using the token and API key services, you can build a custom realm.

Authorize

Cluster

Define **cluster-level** privileges. Cluster privileges are typically reserved for administrators.

Index

Define the **index-level** privileges for one or more indices. This can optionally include **document-level** and **field-level** privileges.

SECTION SUMMARY

A Cloud Guru's Elastic Certified Engineer Exam Preparation Course

Section Summary

- Diagnosing Shard Issues
- Snapshotting Data
- Configuring Searchable Snapshots
- Restoring Data
- Setting Up Cross-Cluster Replication
- Defining Access Control

Myles Young
Training Architect

A CLOUD GURU