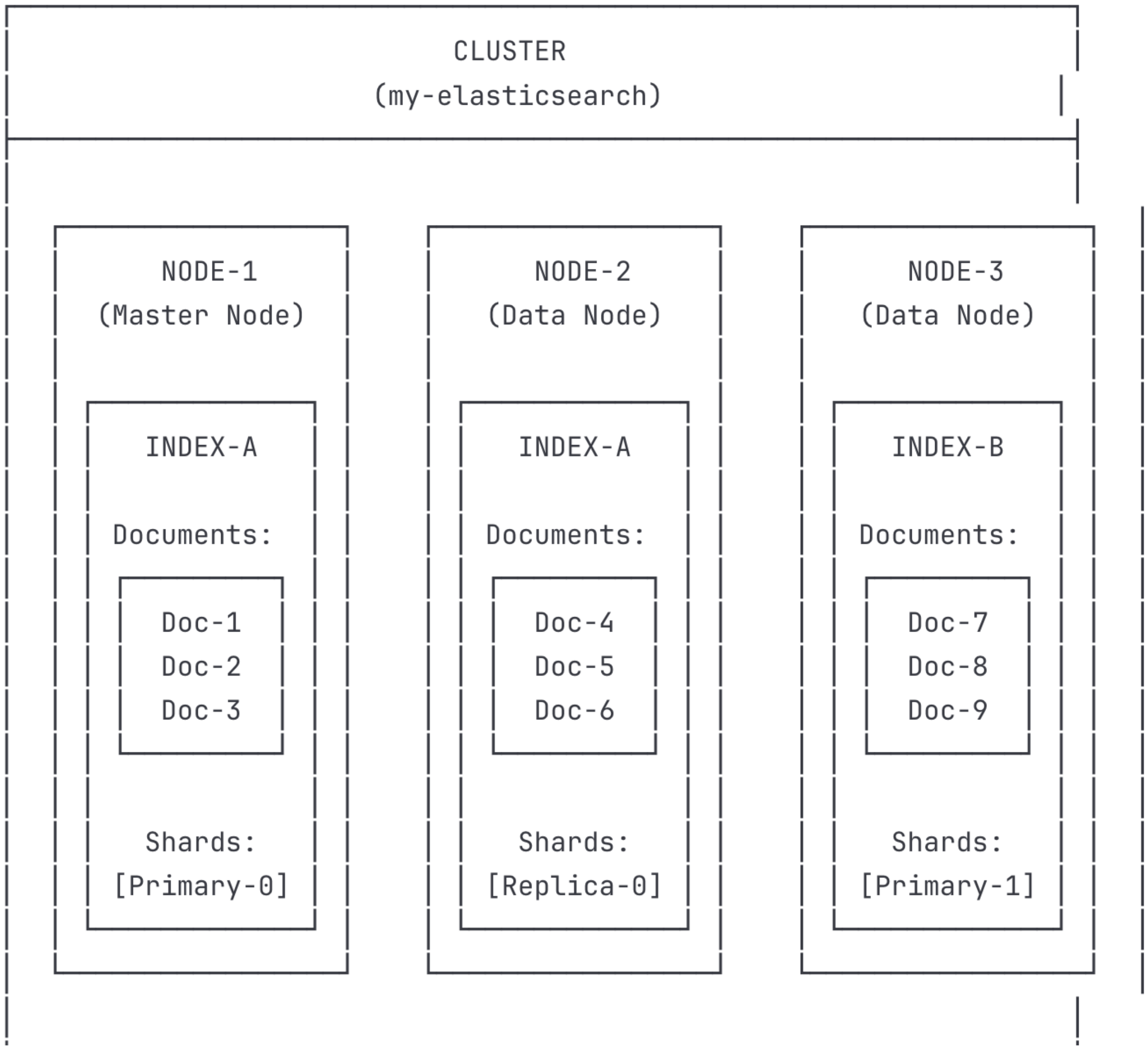
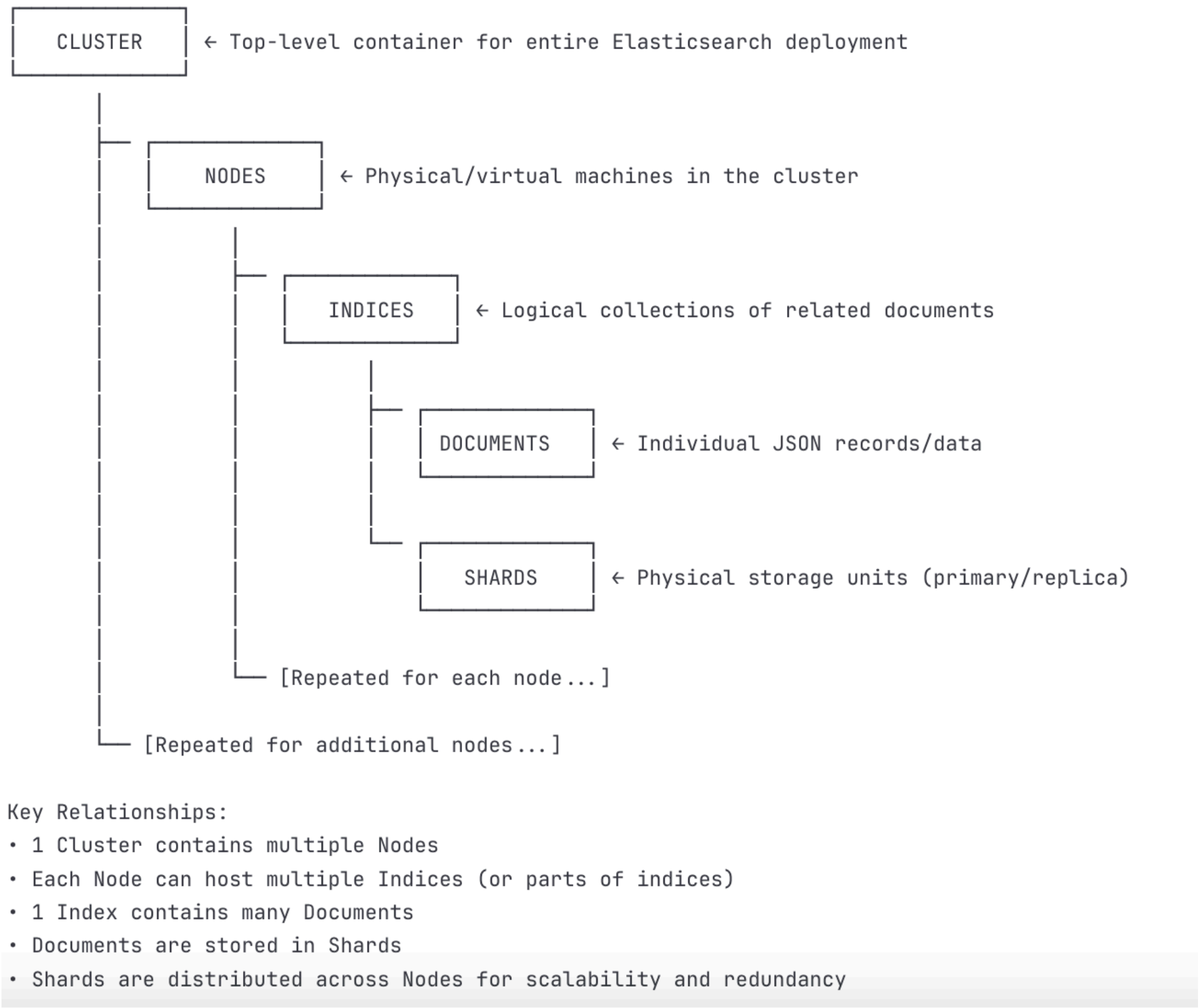
Can you explain the concepts of a node, cluster, index, document, and shard in Elasticsearch?



```
CLUSTER
(my-elasticsearch)

NODE-1              NODE-2              NODE-3
(Master Node)       (Data Node)         (Data Node)

INDEX-A             INDEX-A             INDEX-B

Documents:          Documents:          Documents:

Doc-1               Doc-4               Doc-7
Doc-2               Doc-5               Doc-8
Doc-3               Doc-6               Doc-9

Shards:             Shards:             Shards:
[Primary-0]         [Replica-0]         [Primary-1]
```



```
┌──────────────┐
│   CLUSTER    │ ← Top-level container for entire Elasticsearch deployment
└──────────────┘
       │
       ├──┌──────────────┐
       │  │    NODES     │ ← Physical/virtual machines in the cluster
       │  └──────────────┘
       │         │
       │         ├──┌──────────────┐
       │         │  │   INDICES    │ ← Logical collections of related documents
       │         │  └──────────────┘
       │         │         │
       │         │         ├──┌──────────────┐
       │         │         │  │  DOCUMENTS   │ ← Individual JSON records/data
       │         │         │  └──────────────┘
       │         │         │
       │         │         └──┌──────────────┐
       │         │            │    SHARDS    │ ← Physical storage units (primary/replica)
       │         │            └──────────────┘
       │         │
       │         └── [Repeated for each node...]
       │
       └── [Repeated for additional nodes...]

Key Relationships:
• 1 Cluster contains multiple Nodes
• Each Node can host multiple Indices (or parts of indices)
• 1 Index contains many Documents
• Documents are stored in Shards
• Shards are distributed across Nodes for scalability and redundancy
```

🔷 Key Concepts in Elasticsearch

1. Node 🖥️

- A **single server** in Elasticsearch.
- Runs Elasticsearch software and stores data.
- Can be physical or virtual.

👉 **Analogy:** One **student** in a class.

---

2. Cluster 🏫

- A collection of **nodes** working together.
- They share the data and workload.
- Identified by a **unique cluster name**.

👉 **Analogy:** The **classroom** (all students together).

---

3. Document 📄

- The **basic unit of data** in Elasticsearch (stored in JSON format).
- Example: A single log entry, one order detail, one customer record.

👉 **Analogy:** A **single page in the notebook**.

---

4. Index 📁

- A collection of **documents** that share similar characteristics.
- Think of it as a **database** in SQL.
- Example: logs-2025, ecommerce-orders.

👉 **Analogy:** A **notebook** where students write notes on a specific subject.

---

5. Shard 🔪

- An index can be split into smaller pieces = **shards**.
- Each shard is a self-contained Lucene index.
- Two types:
  - **Primary shard** → actual data.
  - **Replica shard** → copy for fault tolerance.

👉 **Analogy:** If a notebook (index) is too big, you **tear it into chapters** (shards) so different students (nodes) can hold parts of it.
If one student loses their part, a **backup copy** (replica) is still safe.

🔷 How They Work Together

Imagine you're running an **online shopping site**:

- **Cluster** = your whole data system.
- **Nodes** = multiple servers storing data
- **Documents** = each order placed by a customer.
- **Index** = "orders" database.
- **Shards** = splitting the "orders" index into smaller pieces so they can be spread across servers for performance and reliability.

🔷 Short Interview Answer

In Elasticsearch:

- A **node** is one server.
- A **cluster** is a group of nodes.
- An **index** is like a database that stores related documents.
- A **document** is the smallest unit of data, stored in JSON.
- A **shard** is a partition of an index, with replicas for fault tolerance.

This design makes Elasticsearch scalable and fault-tolerant.

A node in Elasticsearch is best compared to:
A) A database in SQL
B) A single server running Elasticsearch
C) A partition of data
D) A JSON document
Answer: B) A single server running Elasticsearch

What does a cluster represent in Elasticsearch?
A) A single JSON document
B) A collection of shards

C) A group of nodes working together
D) A physical hard disk
Answer: C) A group of nodes working together

Which of the following best describes a document in Elasticsearch?
A) A single entry of data stored in JSON format
B) A group of servers
C) A collection of indices
D) A backup copy of data
Answer: A) A single entry of data stored in JSON format

What is an index in Elasticsearch similar to?
A) A class of students
B) A database in SQL that holds related documents
C) A single server
D) A JSON file
Answer: B) A database in SQL that holds related documents

Why does Elasticsearch use shards?
A) To remove duplicate documents
B) To split an index into smaller pieces for scalability and fault tolerance
C) To create multiple clusters
D) To reduce JSON size
Answer: B) To split an index into smaller pieces for scalability and fault tolerance

---