# Detailed Notes: Elasticsearch Architecture

## 1. Introduction to Elasticsearch Architecture

1. Elasticsearch was successfully installed along with Kibana.

2. This lecture focuses on the core components of Elasticsearch architecture, such as nodes, clusters, and data organization.

## 2. Nodes in Elasticsearch

1. A node is **an instance of Elasticsearch** that stores data.

2. Multiple nodes can be started to store large amounts of data, even terabytes.

3. Key facts about nodes:

* - Nodes can be distributed across multiple physical or virtual machines.
* - A single machine can host multiple nodes.
* - For production environments, nodes should run on dedicated machines, virtual machines, or containers.

## 3. Clusters in Elasticsearch

1. A cluster is a collection of nodes that together store and manage data (manage means that data is properly replicated etc.).

2. Features of clusters:

* - Data is distributed across nodes within a cluster.
* - A cluster can be automatically formed when a node starts.
* - Multiple clusters can be created for different purposes (e.g., e-commerce search, APM).

3. Clusters are independent by default but can support cross-cluster searches if needed.

4. Common reasons for using multiple clusters include logical separation and different configurations.

5. When the 1st node is started, it forms a cluster automatically having that only node itself.

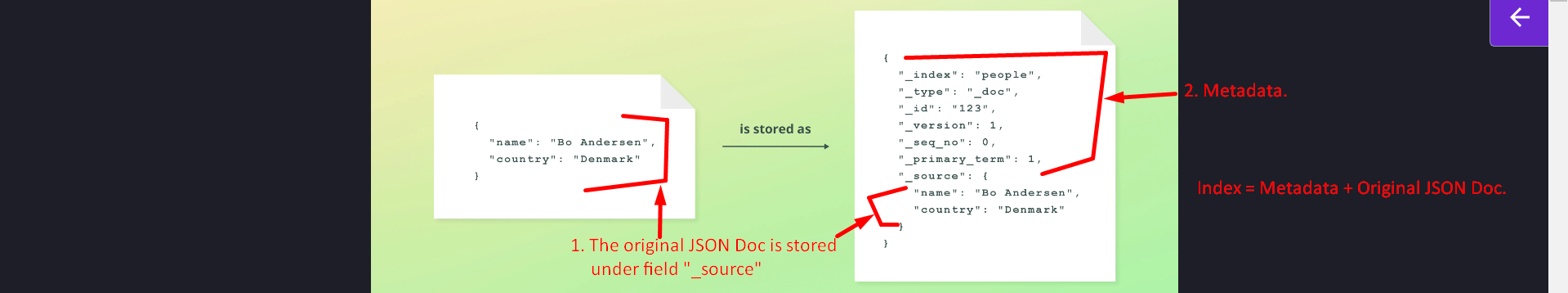
6. A new node will create either a new cluster or will become part of either existing cluster.

7. If one node in the cluster, then availability and scalability issues may occur.

Let’s look at how Data is organized and stored.

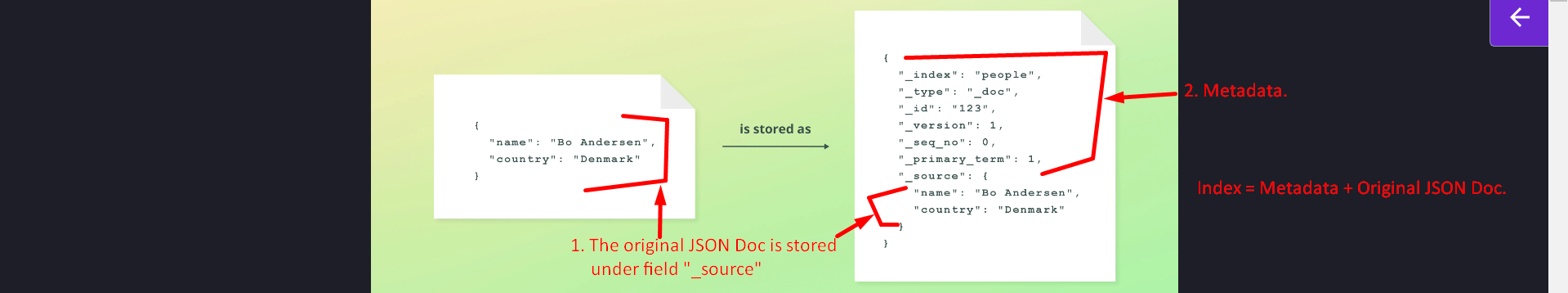
## 4. Documents in Elasticsearch

1. Data in Elasticsearch is stored as **documents which is smallest unit**.

2. Documents are JSON objects containing fields with any desired data.  


3. When a document is indexed:

* - The original JSON object is stored in the `\_source` field.
* - Elasticsearch adds metadata to the document for internal use.

4. Example of a document:  


* - Fields: `name`, `country`.
* - Stored within Elasticsearch along with metadata.

## 5. Indices in Elasticsearch : How are Docs are organized?

1. Documents are organized into indices.

2. An index is a logical grouping of documents with similar characteristics.

3. Features of indices:

* - Provide scalability and availability configurations.
* - Allow searches to be run against specific indices.

- Examples:

* - A `people` index for storing person documents.
* - A `departments` index for storing department documents.

4. Indices can contain an unlimited number of documents.  


## 6. Key Takeaways

1. An Elasticsearch cluster is a collection of nodes responsible for storing and managing data.

2. A node is an instance of Elasticsearch that can run on physical or virtual machines or containers.

3. Data is stored as JSON documents, which can represent any type of information.

4. Documents are grouped into indices, which are logical collections of related documents.

5. This structure ensures scalability and flexibility in managing data.

## 7. Next Steps

In the next lectures, we will explore how data is distributed across nodes, search functionality, and configurations for scalability and availability.