**Elasticsearch Capacity Planning Formula with Example**

Capacity planning in Elasticsearch depends on factors like data size, indexing rate, query load, retention period, and replication. Below is a general approach to estimate storage and nodes required.

**Step 1: Estimate Raw Data Size**

Assume the raw log size is **X GB/day**.

**Step 2: Calculate Indexed Storage Size**

Elasticsearch typically indexes data at **1.2x to 1.5x** the raw data size. If you enable replication, storage needs increase accordingly.

**Formula:**

Indexed Data per Day=Raw Data per Day×Indexing Overhead\text{Indexed Data per Day} = \text{Raw Data per Day} \times \text{Indexing Overhead}Indexed Data per Day=Raw Data per Day×Indexing Overhead Total Storage Needed=Indexed Data per Day×Retention Days×(1+Number of Replicas)\text{Total Storage Needed} = \text{Indexed Data per Day} \times \text{Retention Days} \times (1 + \text{Number of Replicas})Total Storage Needed=Indexed Data per Day×Retention Days×(1+Number of Replicas)

**Step 3: Choose Number of Nodes**

* **Storage per Node:** Assume each node can store **Y GB**.
* **Number of Data Nodes Needed:**

Nodes Required=Total Storage NeededStorage per Node\text{Nodes Required} = \frac{\text{Total Storage Needed}}{\text{Storage per Node}}Nodes Required=Storage per NodeTotal Storage Needed​

**Example Calculation**

**Scenario**

* Raw log size = **100GB/day**
* Indexing overhead = **1.4x**
* Retention period = **30 days**
* Replication factor = **1** (1 primary + 1 replica)
* Storage per node = **2TB**

**Step 1: Indexed Data per Day**

100GB×1.4=140GB100GB \times 1.4 = 140GB100GB×1.4=140GB

**Step 2: Total Storage Needed**

140GB×30×(1+1)=140GB×30×2=8.4TB140GB \times 30 \times (1 + 1) = 140GB \times 30 \times 2 = 8.4TB140GB×30×(1+1)=140GB×30×2=8.4TB

**Step 3: Number of Nodes**

8.4TB2TB=4.2≈5 data nodes\frac{8.4TB}{2TB} = 4.2 \approx 5 \text{ data nodes}2TB8.4TB​=4.2≈5 data nodes

**Final Recommendation**

* **5 Data Nodes** (each with 2TB storage)
* **3 Master Nodes** (for cluster stability)
* **1 Coordinator Node** (optional for load balancing)