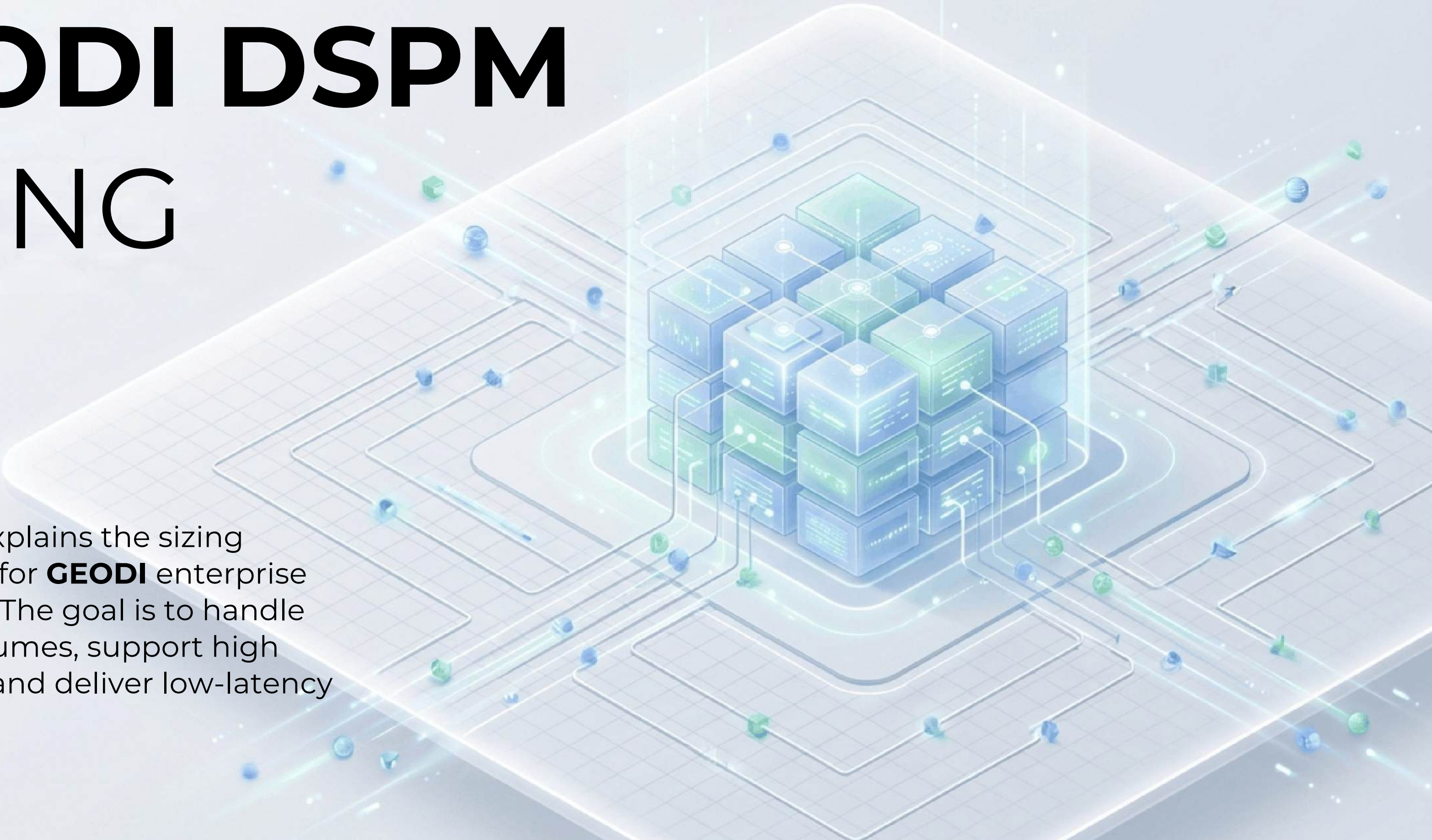


GEODI DSPM SIZING



This section explains the sizing methodology for **GEODI** enterprise deployments. The goal is to handle large data volumes, support high concurrency, and deliver low-latency performance.

Discovery vs Classification Sizing

GEODI DSPM supports both Discovery and Classification, but they require different sizing methods.



Discovery

Discovery is sized mainly according to data size, data type, and scan time expectations.

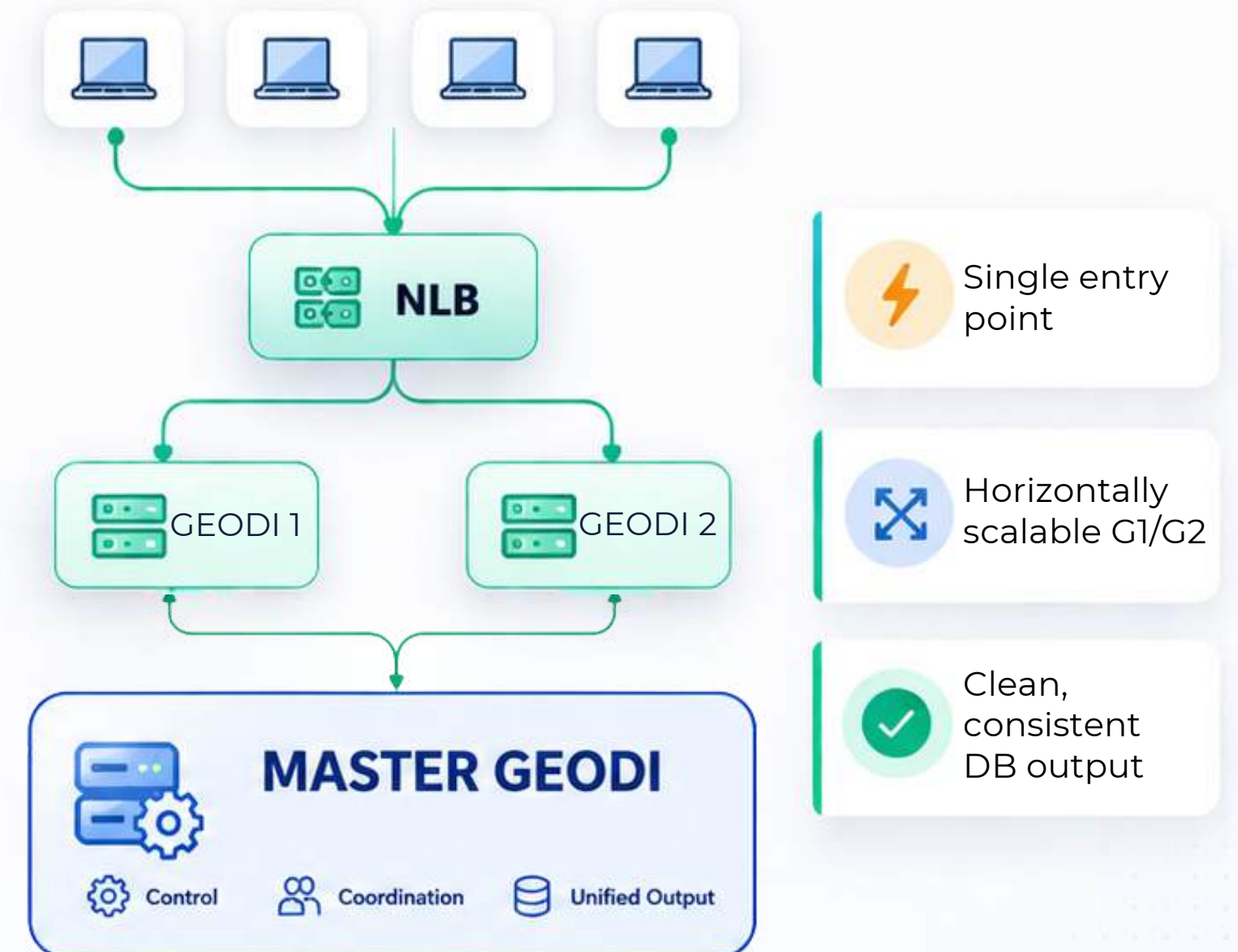
Multiple Data Sources



Classification

Classification requires continuity and performance, so beyond a certain point it should use active-active or NLB.

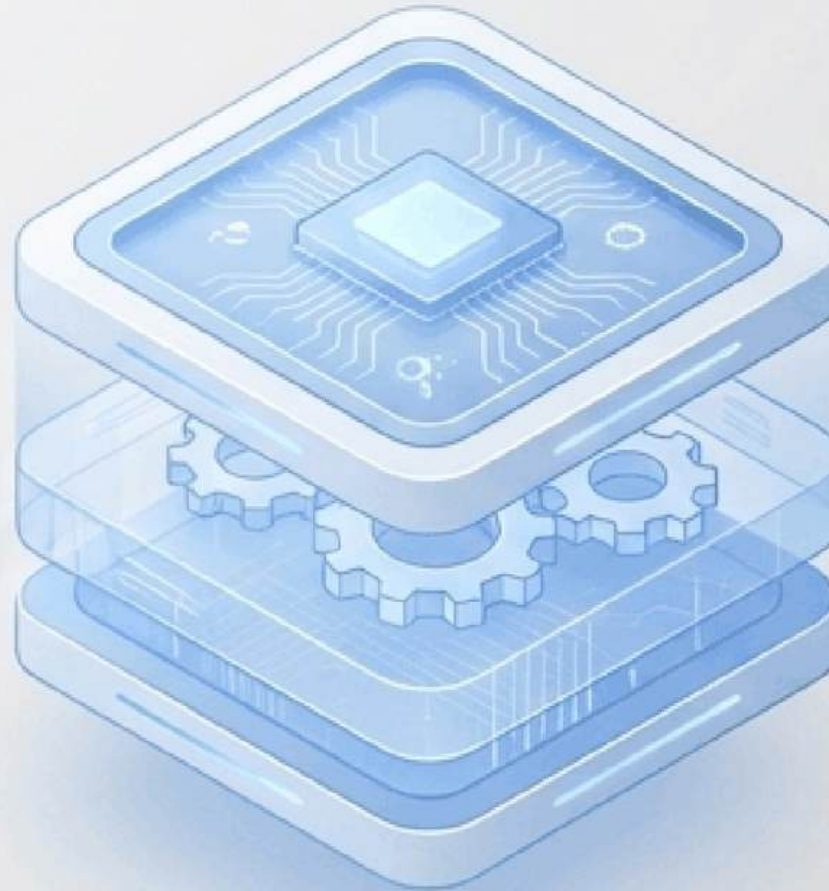
Many EndPoints



Main Factors Affecting Sizing

Several factors directly affect performance and storage needs:

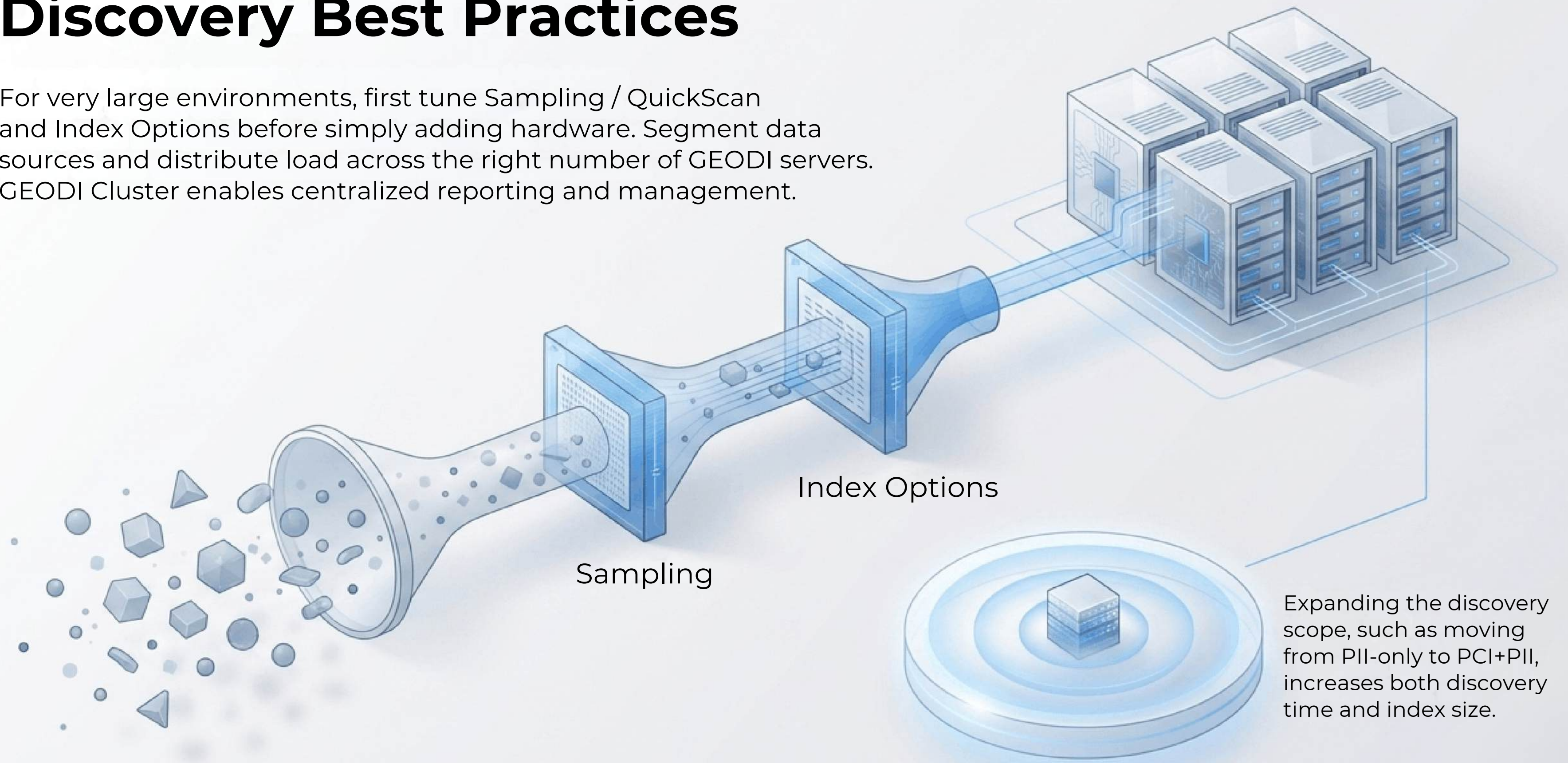
- Sampling / Quick Scan modes
- Index strategy
- Number of GEODI nodes
- Type of data source



These influence throughput, index size, I/O load, latency, and resilience.

Discovery Best Practices

For very large environments, first tune Sampling / QuickScan and Index Options before simply adding hardware. Segment data sources and distribute load across the right number of GEODI servers. GEODI Cluster enables centralized reporting and management.



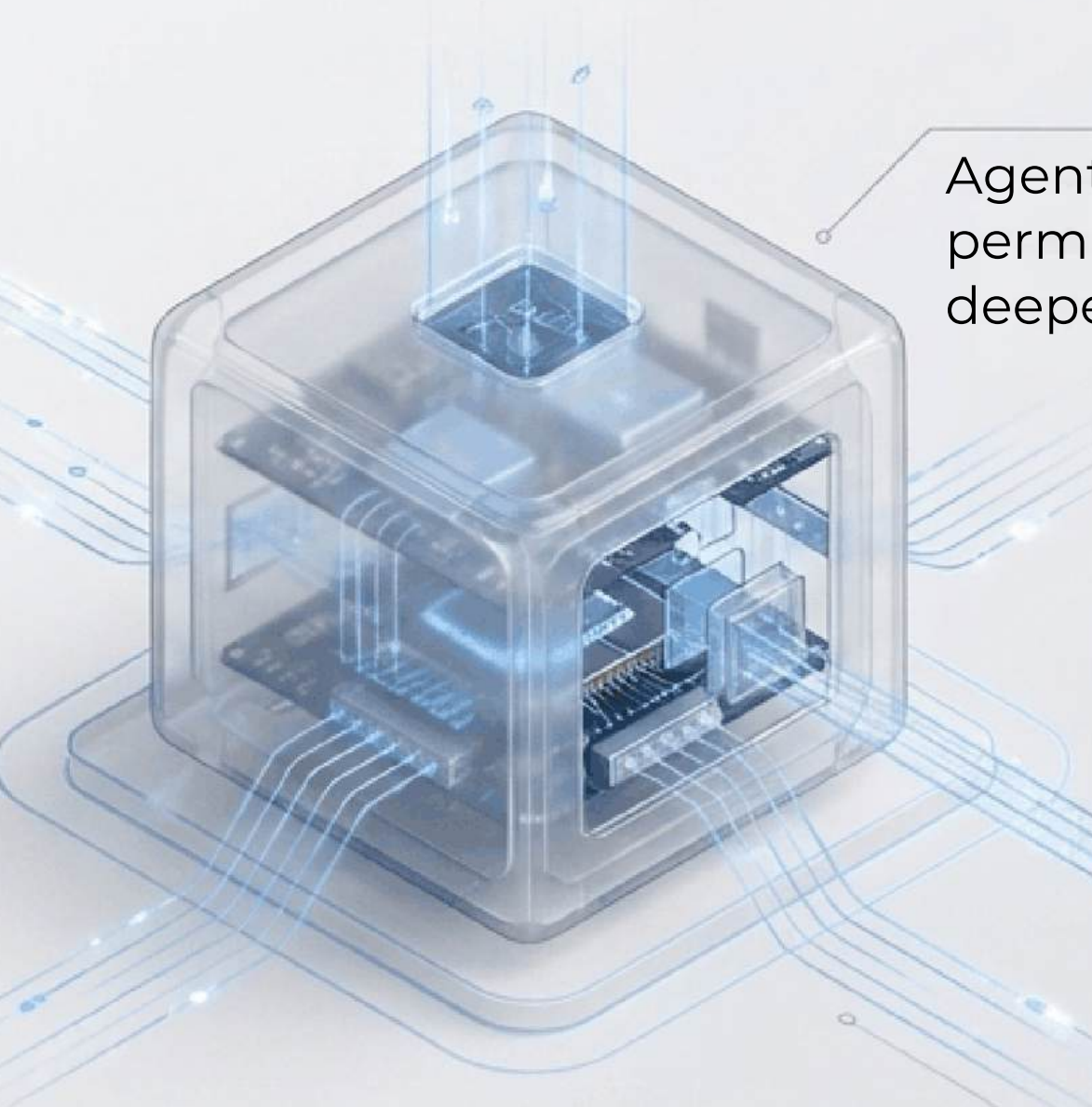
Discovery Options & Performance

	Full Text Index	Metadata Only	Sampling/QuickScan
	Typically 0.5-1.5 TB/day, full search capability, index size roughly 10-20% of a typical dataset.	About 40-50% faster, with 60-80% smaller index size, but search is limited to metadata and content names.	Discovery becomes almost proportionally faster and smaller according to the sampling rate, but results are limited to sampled content.
 Indexing Speed	0.5-1.5 TB/day	40-50% faster	Fastest
 Index Size	100% Approximately 10-20% of a typical dataset	60-80% reduction	Almost proportionally smaller by the sampling rate
 Search Capability	 Full	 Limited to metadata and content names	 Limited to sampled content



Discovery Architecture Notes

You may use GDE agents or agentless discovery.
Both provide similar search and remediation outcomes.



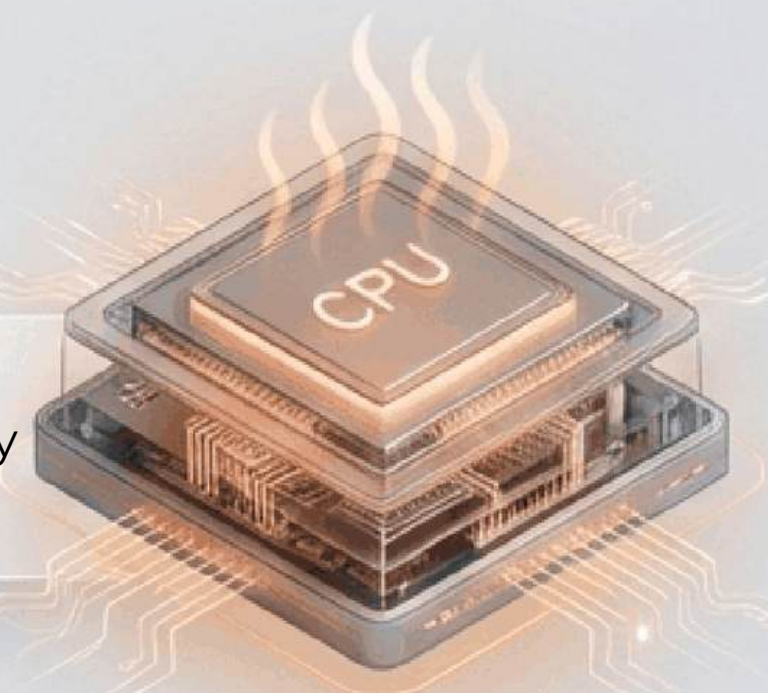
Agents may help with permissions and provide deeper endpoint access.



Agentless methods reduce deployment overhead.



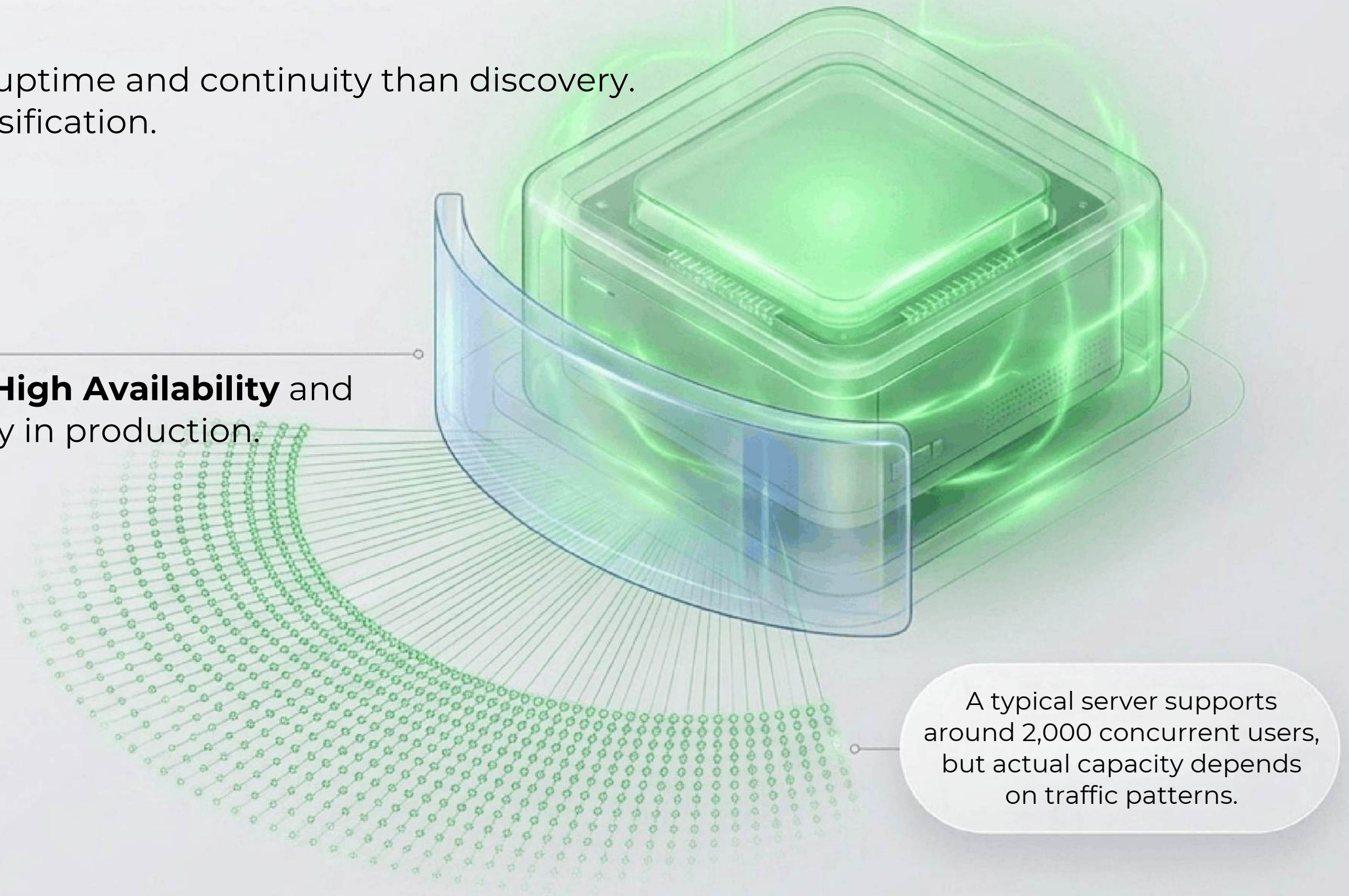
OCR is CPU-intensive.
Use Delayed-OCR and apply OCR only where needed.



Classification Best Practices

Classification requires higher uptime and continuity than discovery.
Use dedicated servers for Classification.

Use **NLB** both for **High Availability** and for scaling capacity in production.

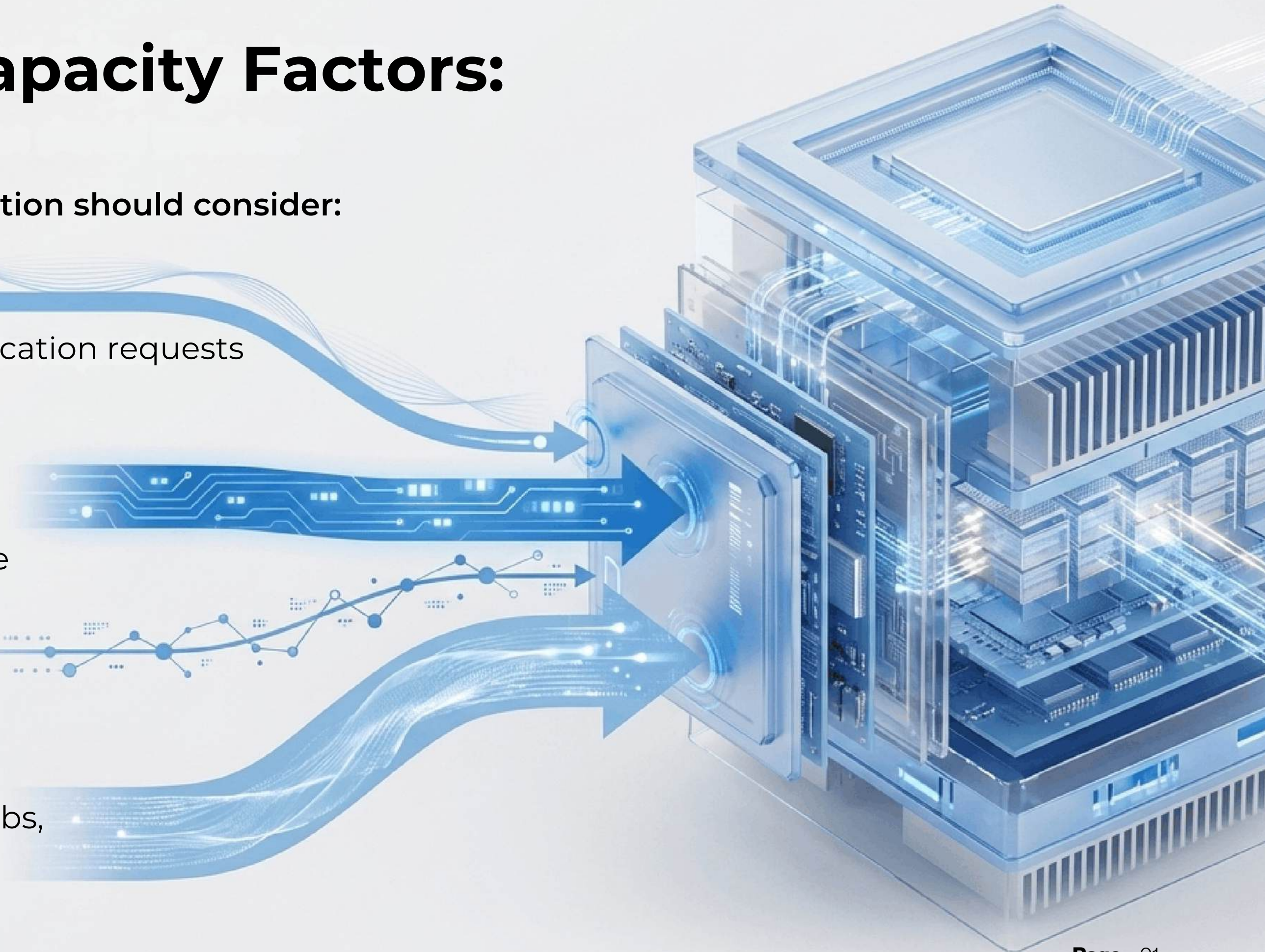


A typical server supports around 2,000 concurrent users, but actual capacity depends on traffic patterns.

Classification Capacity Factors:

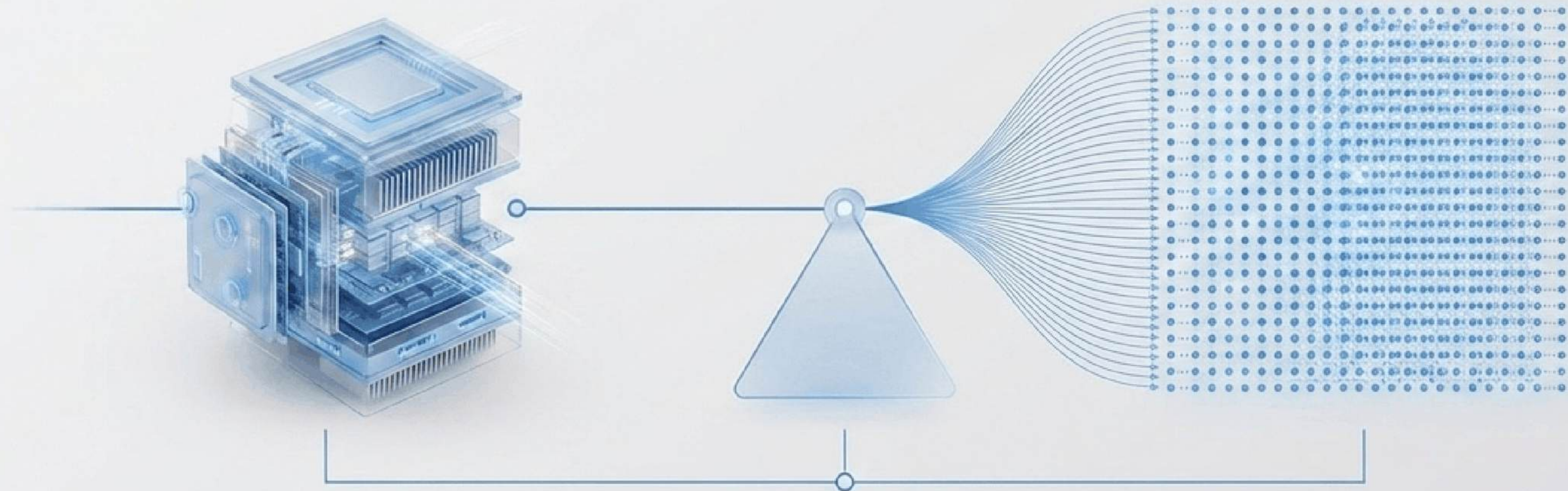
Capacity planning for Classification should consider:

- Manual and automatic classification requests
- AI-based classification workloads, which may be CPU/GPU intensive
- Number of concurrent users
- Traffic volume, such as bulk jobs, email flows, and Office add-ins



Endpoint Discovery Sizing

GDE (**GEODI Desktop Explorer**) can crawl thousands of endpoints and send results to GEODI servers.

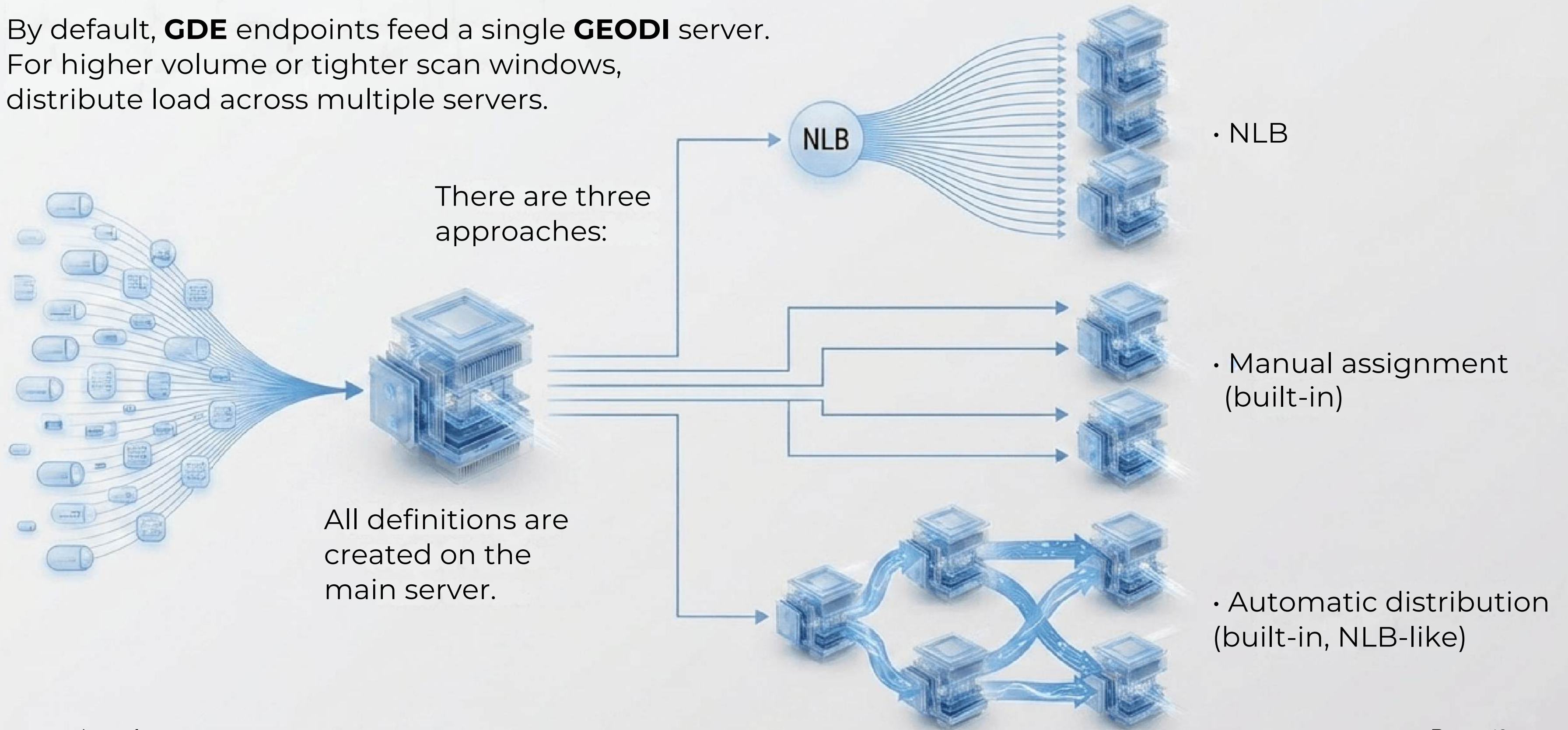


As a practical guideline, allocate a new **GEODI** server for roughly **10 TB** of data or about **1,000 endpoints** averaging **10 GB** each.

Use **NLB** or the built-in distribution methods to balance **GDE traffic** across multiple servers.

GDE Multi-Server Distribution

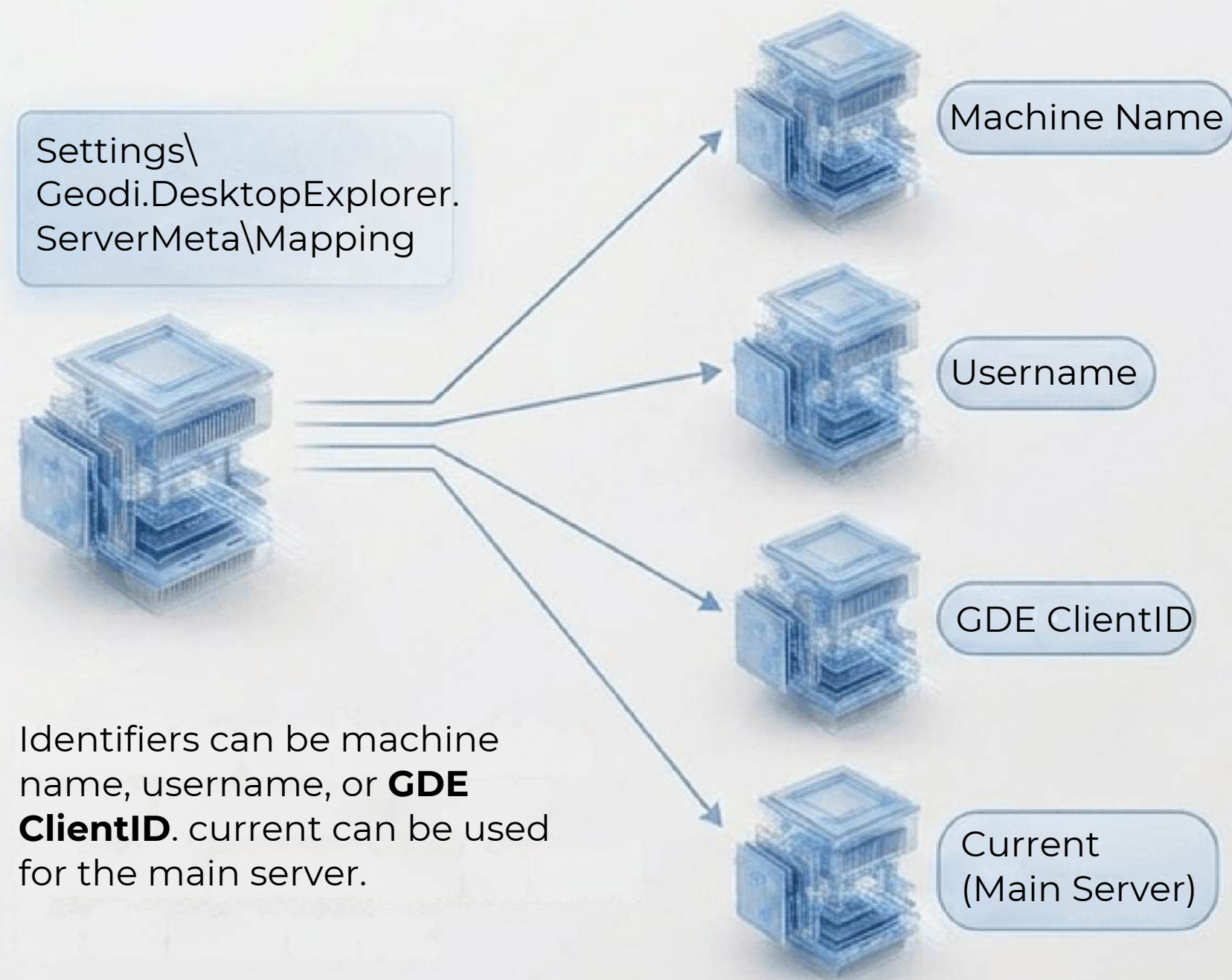
By default, **GDE** endpoints feed a single **GEODI** server. For higher volume or tighter scan windows, distribute load across multiple servers.



Manual & Automatic GDE Assignment

Manual assignment maps specific endpoints to specific servers using files under

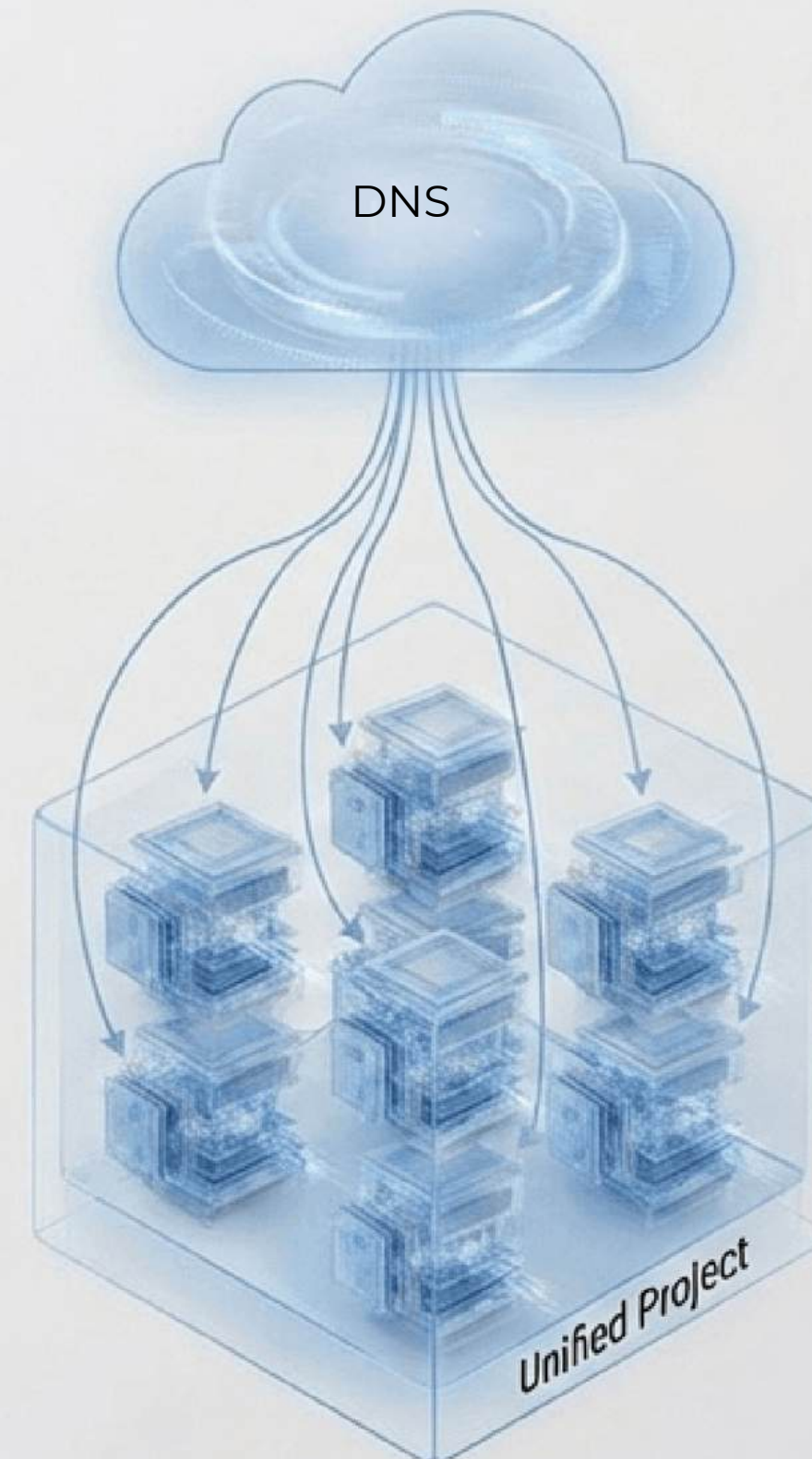
Settings\Geodi.Desktop Explorer.Server Meta\Mapping.



Identifiers can be machine name, username, or **GDE ClientID**. current can be used for the main server.

Automatic assignment uses Settings

`\Geodi.Desktop Explorer.Server Meta\Servers,`
with one server per line.



All defined servers must have the same project and be reachable with the same GDE token.

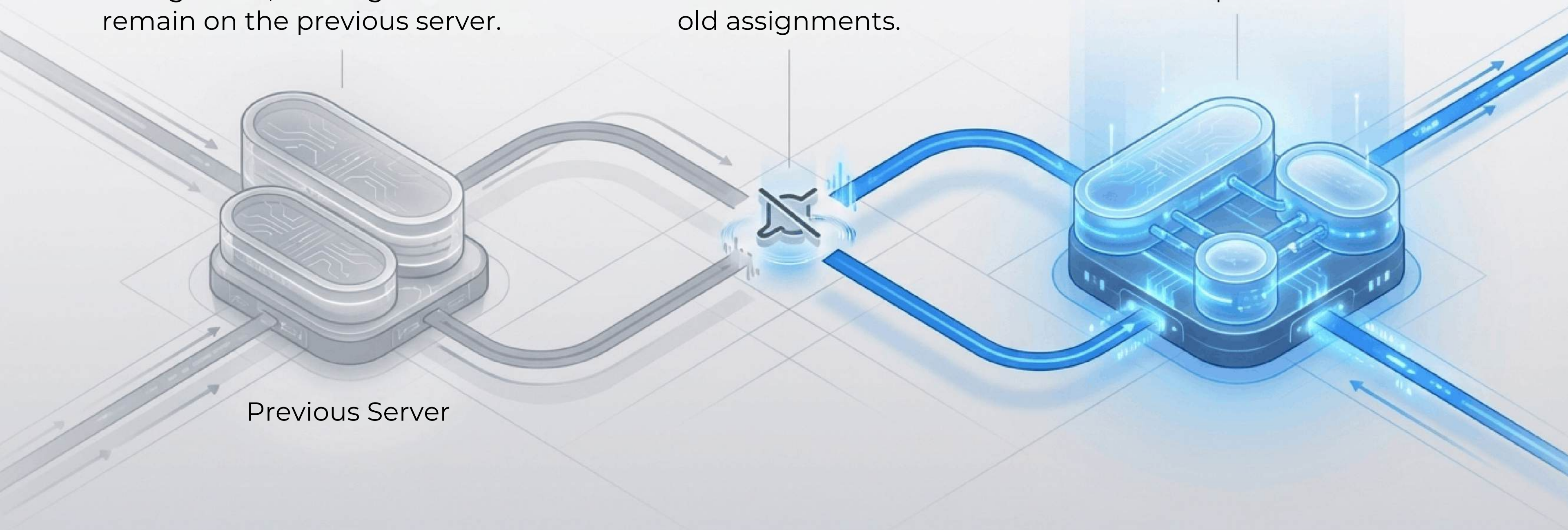
Endpoints must be able to reach all listed servers directly. DNS is recommended.

GDE Operational Notes

If mappings or server assignments change later, existing feeds remain on the previous server.

Use Forget from the Clients Panel to clean old assignments.

Only new feeds follow the updated rules.



Key Takeaways

