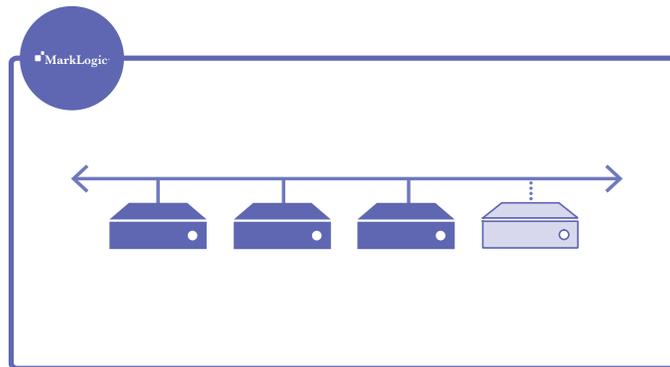


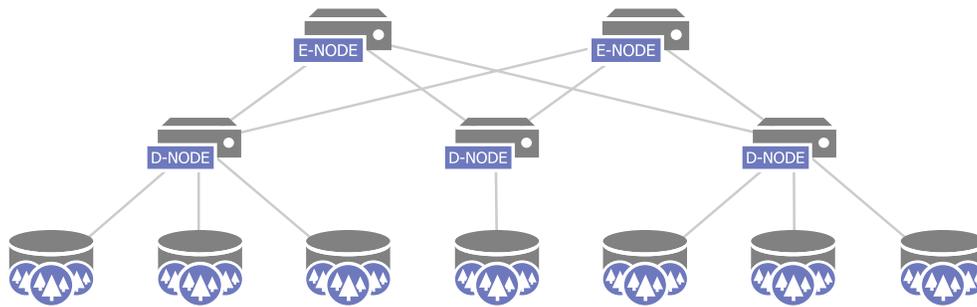
Scalability & Elasticity

With today's massive volumes of data, it is imperative to have a database that scales quickly, easily, and at low cost. But, it is also important to have elasticity—to be able to scale down based upon fluctuating demand. MarkLogic® is a massively scalable Enterprise NoSQL database that scales horizontally in clusters on commodity hardware to hundreds of nodes, petabytes of data, and billions of documents—and still processes tens of thousands of transactions per second. And, when demand dissipates, MarkLogic can scale back down without having to worry about complex sharding. With these features, organizations can handle incredible volumes of data and run large scale web applications—all without breaking the bank.



Start Small, Go Big

- **Scalability** – From three nodes to hundreds of nodes, or 10,000 documents to 1 billion documents—MarkLogic clusters scale horizontally as your data or access demand grows and shrinks
- **Elasticity** – Add or remove nodes in minutes and take advantage of automatic cluster rebalancing, helping you keep the database in line with performance needs without over-provisioning
- **Run on Commodity Hardware** – MarkLogic doesn't need "big iron." You can run MarkLogic on cost-effective commodity hardware in any environment—in the cloud, virtualized, on-premises, or a combination
- **Shared Nothing Architecture** – MarkLogic uses a shared-nothing architecture with no master-slave relationships, which means there is no risk of data loss if a node fails. If one node fails, another node automatically picks up the workload
- **No Performance Degradation** – MarkLogic was designed from the start to run large enterprise applications, and does not reach a limit where there are large performance cliffs while scaling
- **Fewer Nodes and Licenses** – MarkLogic datasets and indexes do not have to fit in-memory, which means you can scale without the expense of dozens of boxes and licenses
- **Tiered Storage** – As your data volumes grow, you can triage your data to less expensive storage tiers based on cost and performance trade-offs, making scaling even easier and more cost effective



Clustering in MarkLogic

MarkLogic is designed for extremely large data volumes, and scales to clusters of hundreds of machines, each of which runs MarkLogic. Each machine in a MarkLogic cluster is called a *host*, or *node*. Some hosts (Data Managers, or *D-nodes*) manage a subset of data in what are called *forests* (also known as *shards*). Other hosts (Evaluators, or *E-nodes*) handle incoming user queries and internally federate queries across D-nodes to access the data. As you load more data, you add more D-nodes. As the user load increases, you add more E-nodes.

High Availability

Clustering enables high availability. In the event that an E-node should fail, there is no host-specific state to lose—just the in-process requests (which can be retried)—and a load balancer can route traffic to the remaining E-nodes. Should a D-node fail, that subset of the data can be brought online by another D-node.

Commodity Hardware

MarkLogic clusters across commodity hardware connected on a LAN. A commodity server can be anything from a laptop, to a simple virtualized instance, all the way up to a high-end box with two CPUs—each with 12 cores, 512 gigabytes of RAM, and either a large local disk array or access to a SAN. A high-end box like this can store terabytes of data.

More Information

- **Documentation** – Scalability, Failover, and High Availability (<http://docs.marklogic.com/guide/cluster>)
- **Technical White Paper** – Fundamentals of Resource Consumption (<http://developer.marklogic.com/media/fundamentals-of-resource-consumption.pdf>)

About MarkLogic

MarkLogic is the world's best database for integrating data from silos, providing an operational and transactional Enterprise NoSQL database platform that integrates data better, faster, with less cost. Visit www.marklogic.com for more information.