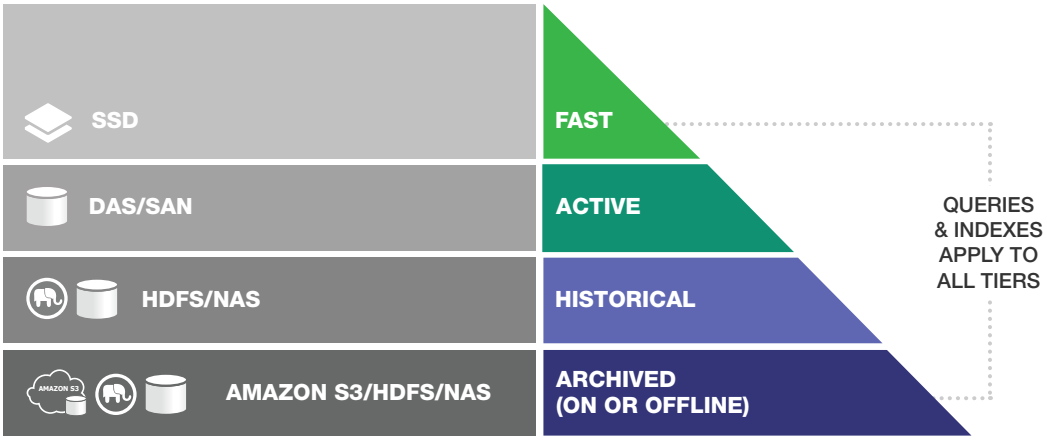
MarkLogic®

DATASHEET

Tiered Storage

Today organizations are storing more data than ever and are now challenged with deciding how to keep it available and manage it. MarkLogic® tiered storage provides the ability to store and manage data in different tiers based on cost and performance trade-offs—whether it's flash storage, traditional local or shared disk storage, HDFS, or Amazon* cloud storage. With tiered storage, data is easily migrated between these tiers without any ETL, additional software, or expensive infrastructure changes. Organizations can easily balance performance and capacity through the information lifecycle—meeting performance SLAs and making data governance easy.



Store and Manage Data More Efficiently

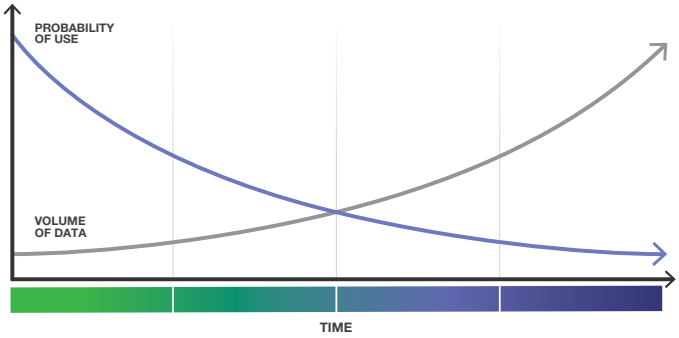
As we venture further into the Exabyte era, in which 90% of the world's data was created in just the past two years, organizations are facing increasingly difficult choices around storing and managing data to fit their business needs and budget. Unfortunately, traditional databases are not designed to efficiently move data between storage tiers, and data older than 90 days is often stored on tape archives never to be seen again. MarkLogic tiered storage solves this problem by efficiently storing and managing data across the information lifecycle—from fast down to archived, and then back again.

- **Optimize data availability** – Move data across tiers without having to take the data offline, perform any ETL, or even re-index it—allowing you to move without regret, knowing that it is always available when and where you need it
- **Reduce the cost of storage** – Storage costs can vary widely from around $1 to $25 per Gigabyte. With MarkLogic, you can avoid over-provisioning expensive storage for data that can be easily stored on a cheaper tier
- **Save time managing storage** – Partition data to different storage tiers using a set policy in database administration. For example, a policy can be created to automatically archive data if it is older than a certain date
- **Leverage the economics of Amazon S3 or HDFS** – Use Amazon S3 or HDFS as distributed file systems for cheaply storing large volumes of archival data, without losing the ability to bring that data back into an active, operational storage tier quickly, and without any ETL or re-indexing

MarkLogic®

Match Storage With The Information Lifecycle

Organizations are storing larger and larger volumes of historical data, often for compliance purposes. However, over time, the probability of needing that data decreases. Tiered storage manages data over this information lifecycle, helping the business meet performance SLAs while not breaking the bank, and efficiently handling the unpredictable nature of when archived data will need to be active again.



Use Multiple Storage Options

The below list includes storage options that work with MarkLogic tiered storage. For each option, we include an example for a configuration designed to store a few hundred terabytes of data. But, you should consult with MarkLogic professionals to discuss the storage options that would work best for your unique use case.

- **SSD** – Solid state drives can be used for MarkLogic Fast Data Directories. A configuration might include a few SSDs to handle a few gigabytes of active data. When the limit of capacity is reached, slower data directories pick up the workload
- **Direct Attached Storage (DAS)** – Local disk storage can be used for active, operational data. A configuration might include local 10K Serial Attached SCSI (SAS) RAID10 hard drives for a small number of hosts and a few dozen terabytes
- **Shared Disk (SAN, NAS)** – Storage Area Networks and Network Attached Storage can be used for active data, but is more commonly used for older historical or archived data that is rarely updated. A configuration might include a few dozen hosts
- **HDFS** – The Hadoop Distributed File System is well-designed as an inexpensive tier for historical or archived data. A configuration might include a large cluster of dozens, or even hundreds of hosts to handle hundreds of terabytes of data
- **Amazon Cloud** – MarkLogic has pre-configured AMIs to quickly get going on Amazon Web Services with Amazon EBS Storage Volumes or Amazon S3 buckets. Amazon S3 is similar to HDFS, providing a cheap storage mechanism for older, non-transactional data

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.

MARKLOGIC CORPORATION
999 Skyway Road, Suite 200 San Carlos, CA 94070
+1 650 655 2300 | +1 877 992 8885 | www.marklogic.com | sales@marklogic.com