Heterogeneous Storage

Hadoop got its start targeting a very homogenous storage landscape. In fact, part of the success of HDFS was due to how effective it was in pooling massive amounts of commodity hard drives across the cluster, without requiring any expensive hardware or software solutions such as RAIDs. The assumptions HDFS made were:

- DataNode is a single storage unit
- Storage medium is uniform
- Any other storage medium is not exposed to HDFS clients.

Homogenous Storage Landscape (by Hortonworks, Inc., retrieved from LinkedIn SlideShare)

In a modern data center, however, a storage deployment strategy needs to take into account the fact that on each physical server there may be a different storage medium available, and a filesystem needs

to adapt to pick the best one for the given job, and ultimately allow the client to fine-tune those choices. The kinds of attached storage could be:

- Disk
- SSDs
- Memory.

And then, there are also remote storage options: EBS, SAN, etc.

Collection of tiered storages

Tiered Storages (by Hortonworks, Inc., retrieved from LinkedIn SlideShare)

Learn About Verified Certificates