从Hadoop 2.6开始支持异构存储，主要涉及两个概念：存储类型和存储策略。

## 存储类型

存储类型表示数据节点支持的多种物理存储介质的类型，有如下几类：

* RAM\_DISK(transient)
* SSD
* DISK(default)
* ARCHIVE

列表中存储类型分别对应内存、SSD、机械磁盘和归档存储。按照RAM\_DISK->SSD->DISK->ARCHIVE的顺序，存取速度由快到慢，单bit存储成本由高到低。通过datanode的配置项dfs.datanode.data.dir可以配置存储介质的位置和类型。

<property>

<name>dfs.datanode.data.dir</name>

<value>

[RAM\_DISK]/ram\_disk, [SSD]file:/ssd/dn, [DISK]/disk/dn, [ARCHIVE]/archive/dn

</value>

</property>

## 存储策略

某个文件的存储策略用来决定它的数据块将会被存放在什么类型的存储介质上。HDFS目前支持6种存储策略：

Lazy\_Persist： 一个副本存储在内存，其它副本存储机械硬盘。

All\_SSD： 所有副本存储在SSD

One\_SSD： 一个副本存储在SDD，其它副本存储机械硬盘

Hot ： 所有副本存储机械硬盘

Warm： 一个副本存储机械硬盘，其它副本归档存储

Cold： 所有副本归档存储

按照All\_SSD->One\_SSD->Hot->Warm->Cold的顺序，面向的数据“越来越冷”，可以根据数据的冷热程度选择合理的策略。Lazy\_Persist比较特殊，如果一个文件的存储策略被指定为Lazy\_Persist，在写入时会先写入内存，再异步地写入磁盘，主要用来降低小数据量的写入延迟，代价是在某些情况下会有数据丢失。

## HDFS Mover