

大数据技术-第七章：HBase数据库 HBase组件设置



CONTENTS

01. HBase体系架构

02. HBase组件设置

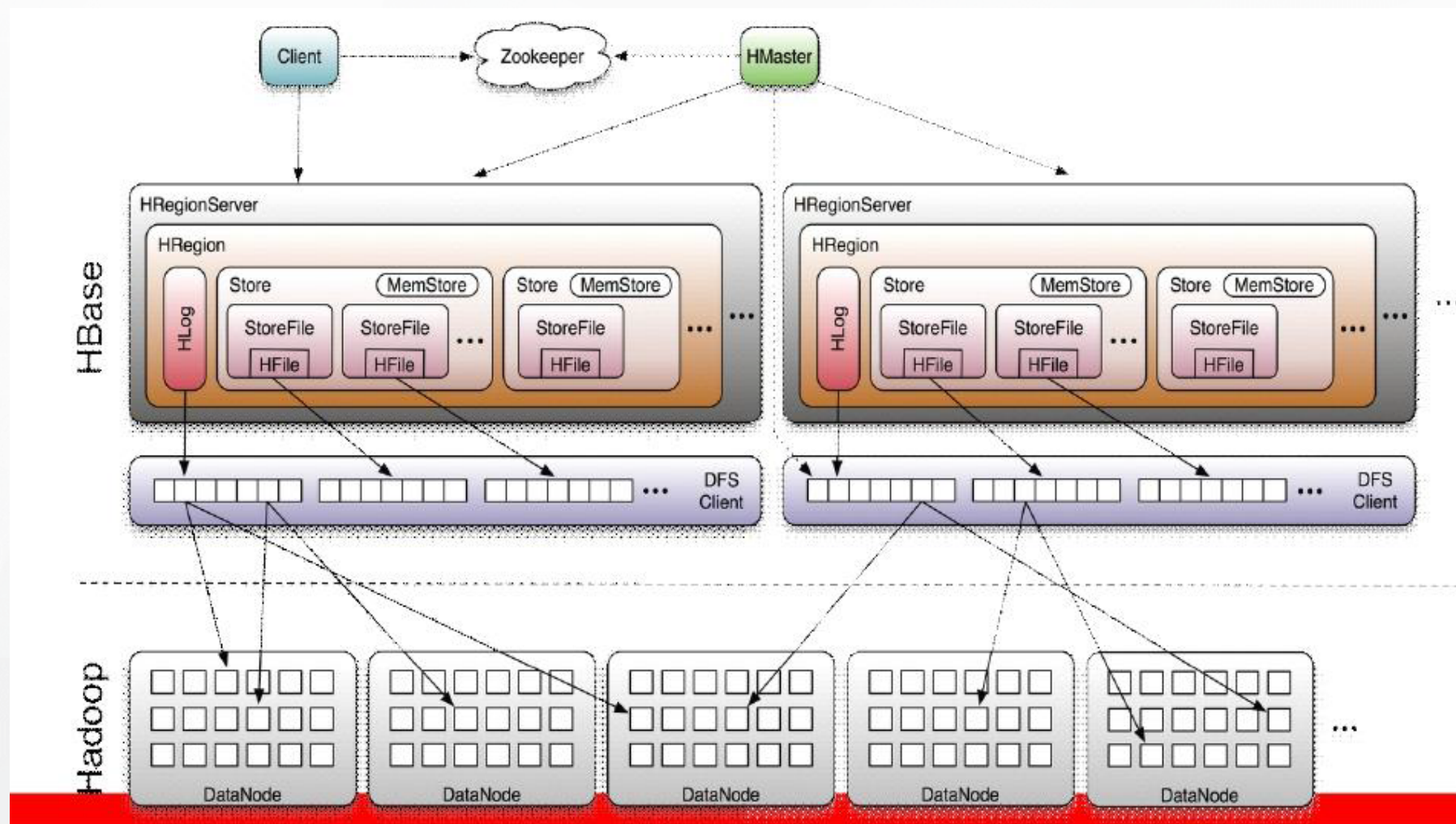


01

HBase体系架构



» HBase体系架构



- Zookeeper: 存储元数据
- HMaster: 不存储数据
- HBase: 是Master/Slave架构
- Hmaster: 类似HDFS中的NameNode
- HRegionServer: 类似HDFS中的DataNode

02

HBase组件设置



下面将介绍HRegion、HStore、MemStore、HFile、WAL等组件的协调操作过程。

(1) HRegion

每个HRegionServer内部管理了一系列HRegion，他们可以分别属于不同的逻辑表，每个HRegion对应了逻辑表中的一个连续数据段。HRegionServer只是管理表格，实现读写操作。Client直接连接到HRegionServer，并通信获取HBase中的数据。

(2) HStore

每个HRegion由多个HStore组成，每个HStore对应逻辑表在这个HRegion集合中的一个Column Family，建议把具有相近IO特性的Column存储在同一个Column Family中，以实现高效读取。

· (3) MemStore

· MemStore是一个缓存，当所有数据完成WAL日志写后，就会写入MemStore中，由MemStore根据一定的算法将数据Flush到底层HDFS文件中(HFile)，每个HRegion中的每个Column Family有一个自己的MemStore。

· (4) HFile

· HFile是最终保存HBase数据行的文件，一个HFile文件属于一张表中的某个列簇，当中的数据是按RowKey、Column Family、Column升序排序，对相同的Cell，则按timestamp倒序排列。

· (5) WAL

· WAL (Write-Ahead-Log, 又名HLog) 是HRegionServer中的日志记录的工具，当系统发生故障时，可以通过WAL恢复数据。在每次用户操作将数据写入MemStore的时候，也会写一份数据到WAL当中，WAL 当中包含了部分还没有写入HFile的文件。

Turing AI 万维图灵 | 大数据系列课程

大数据

BIG
DATA

智 / 能 / 科 / 技 放 / 眼 / 未 / 来

