

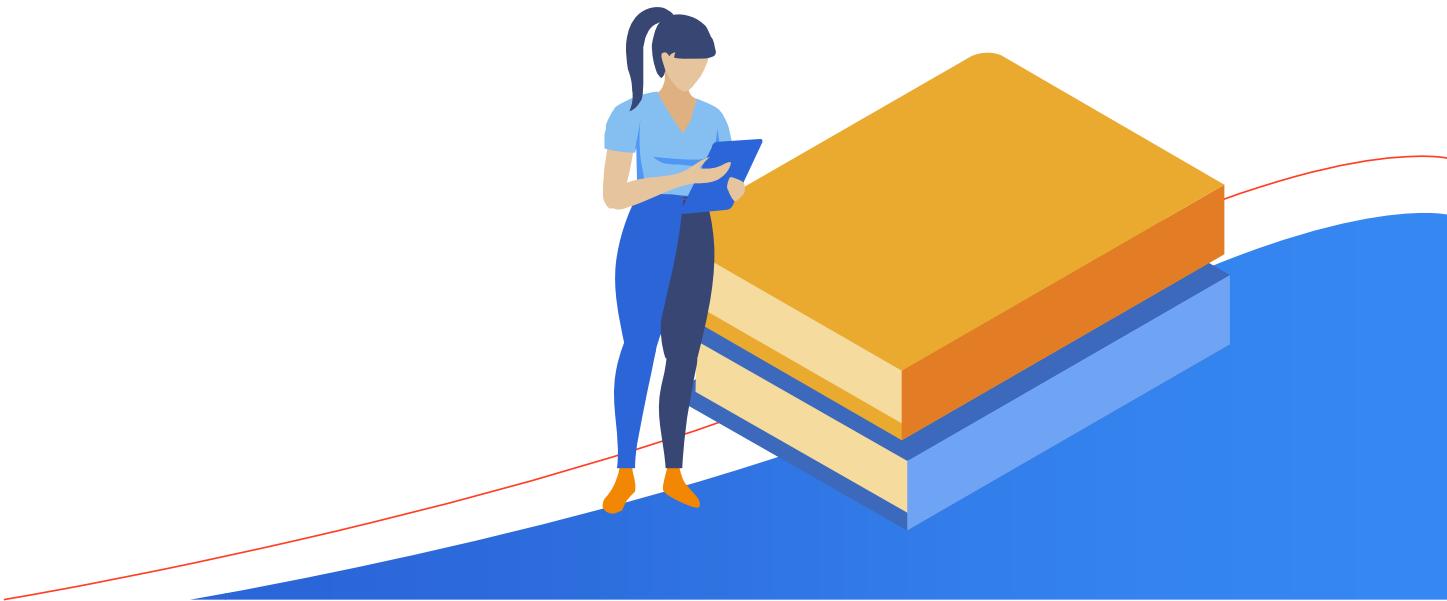
| 大数据技术-第七章：HBase数据库
HBase功能应用



CONTENTS

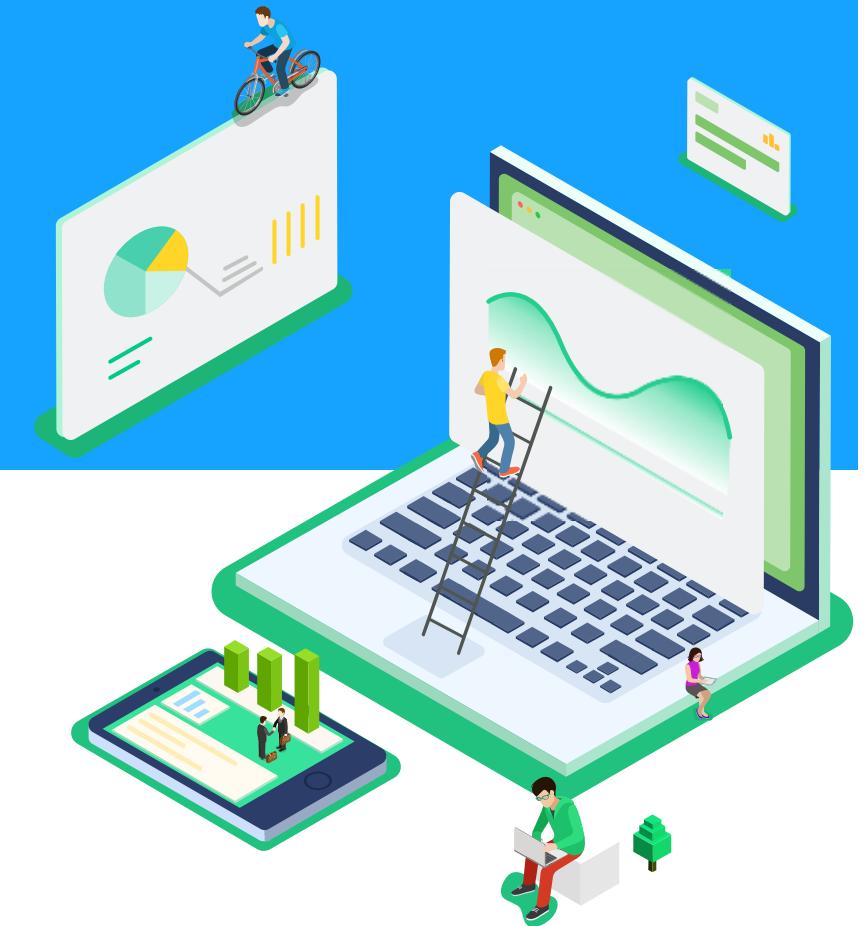
01. HBase适用场景

02. HBase功能应用



01

HBase适用场景

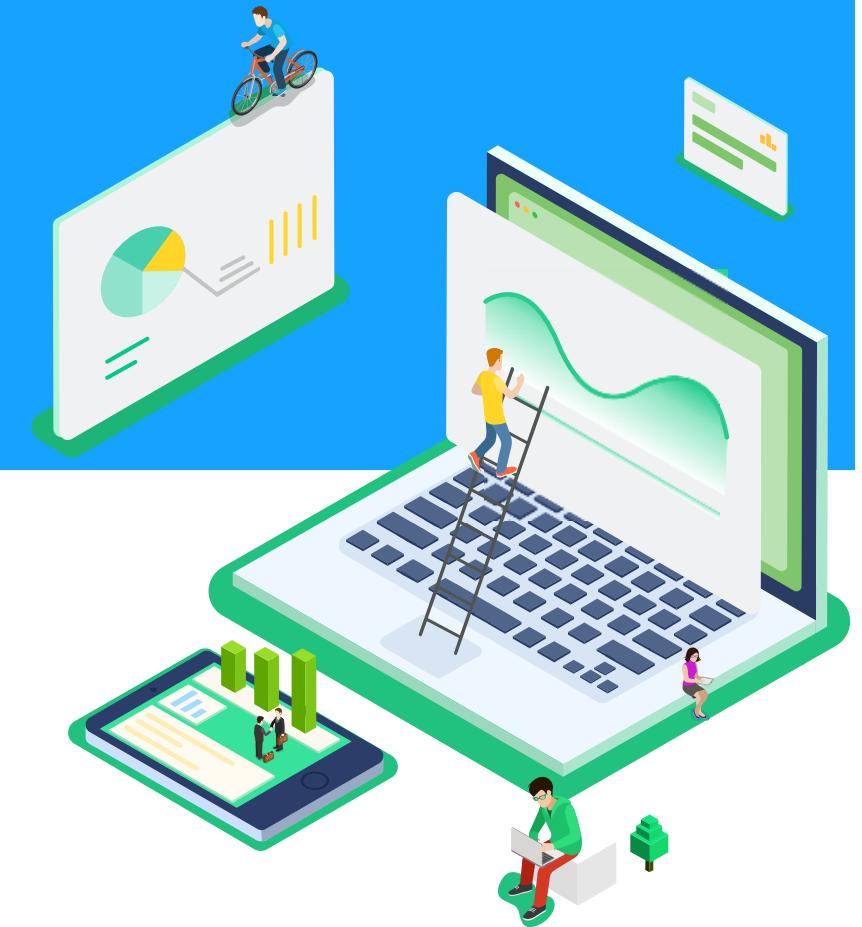


» HBase适用场景

- 数据模型比较简单；
- 对数据库性能要求较高；
- 不需要高度的数据一致性；
- 对于给定key，比较容易映射复杂值的环境。

02

HBase功能应用



HBase的架构是依托于Hadoop的HDFS作为最基本存储基础单元，在HBase的集群中由一个Master主节点管理多个RegionServer，而ZooKeeper进行协调操作。

HBase的体系结构是一个主从式的结构，主节点HMaster在整个集群当中只有一个在运行，从节点HRegionServer有很多个在运行，主节点HMaster与从节点HRegionServer实际上指的是不同的物理服务器，即有一个服务器上面跑的进程是HMaster，很多服务器上面跑的进程是HRegionServer，HMaster没有单点问题，HBase集群当中可以启动多个HMaster，但是通过ZooKeeper的事件处理机制保证整个集群当中只有一个HMaster在运行。既然HBase是数据库，那么数据库从根本上来说就是存储表Table的，但是必须注意HBase并非是传统的关系型数据库(例如：MySQL、Oracle)，而是非关系型数据库，因为HBase是一个面向列的分布式存储系统。

(1) HMaster

HMaster用于启动任务管理多个HRegionServer，侦测各个HRegionServer之间的状态，当一个新的HRegionServer登录到HMaster时，HMaster会告诉它等待分配数据。而当某个HRegionServer死机时，HMaster会把它负责的所有HRegion标记为未分配，并恢复HRegionServer的故障。

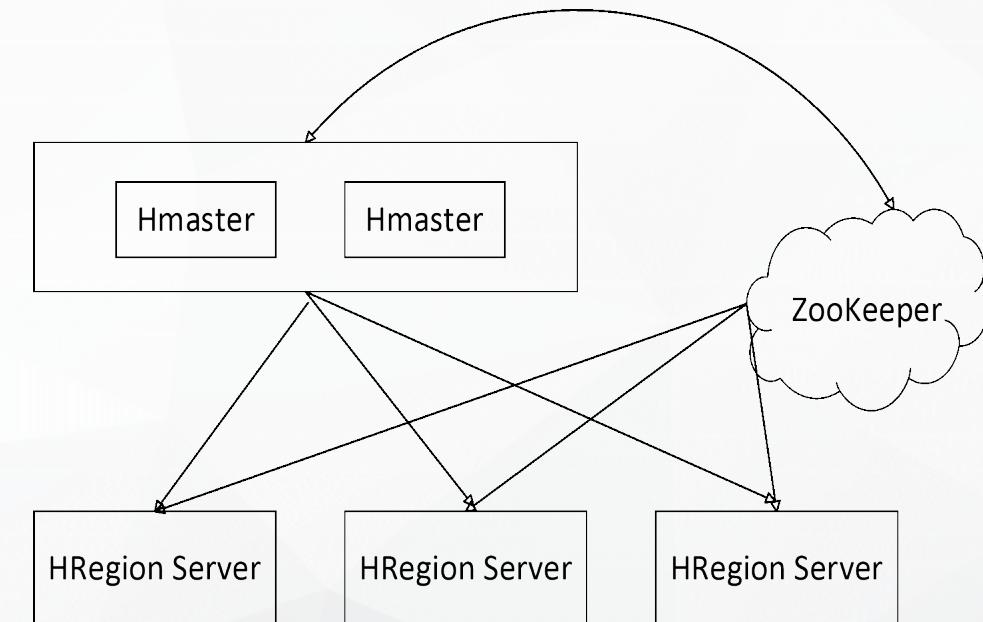
(2) HRegionServer

HBase中的所有数据从底层来说一般都是保存在HDFS中的，用户通过一系列HRegionServer获取这些数据。

HRegionServer主要负责响应用户I/O请求，向HDFS文件系统读写数据，是HBase中最核心的模块。

(3) ZooKeeper

Apache ZooKeeper起源于Hadoop的分布式协同服务，是负责协调集群中的分布式组件，经过多年的发展ZooKeeper已经成为了分布式大数据框架中容错性的标准框架，被多个分布式开源框架所应用。



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

