HBase

# 一、概念

## 特点

|  |
| --- |
| 数据的最终持久化存储是基于HDFS--->俱容量可以 随时在线扩容，  HBASE的数据 增删查改功能模块是：分布式系统。它是一个分布式数据库系统 |

## 结构

|  |
| --- |
| 表名  行键（行唯一id）  列族（多个列族）：列族中存key-value 值。每个key可以有多个value，只是有版本号来区分。查询默认查最新的版本 |

# 二、安装

## 1、修改hbase-env.sh

|  |
| --- |
| vim $HBASE\_HOME/conf/hbase-env.sh  (原文件中中这两句是注释了的)  export JAVA\_HOME=/opt/module/jdk1.8  export HBASE\_MANAGES\_ZK=false #hbase是否启用一个自己的zookper |

## 2、修改hbase-site.xml

|  |
| --- |
| vim $HBASE\_HOME/conf/hbase-site.xml  <configuration>  <!—指定hbse在HDFS上的存储的路径 -->  <property>  <name>hbase.rootdir</name>  <!—哪台服务器的HDFS/数据的存储目录（可任意修改）-->  <value>hdfs://hadoop102:8020/hbase</value>  </ property>  <!—指定hbse是分布式的 -->  <property>  <name>hbase.cluster.distributed</name>  <value>true</value>  </ property>  <!—指定zookeeper,多个用“，”号分割-->  <property>  <name>hbase.zookeeper.quorum</name>  <value>hadoop102:2181,hadoop103:2181,hadoop103:2181</value>  </ property>  </configuration> |
|  |

## 3、修改regionservers

|  |
| --- |
| 修改此文件，填入要启动集群的主机。可方便启动集群  vim $HBASE\_HOME/conf/regionservers  hadoop102  hadoop103  hadoop104 |

## 4、启动服务

|  |
| --- |
| 启动前，应保证hdfs，和zookeeper启动完成  start-hbase.sh  可在<http://192.168.111.102:16010>查看 |

## 5、启动命令行客户端

|  |
| --- |
| bin/hbase shell |

# 三、操作

## 1、建表

|  |
| --- |
| create 'user\_info','base\_info','extra\_info'  create 表名,列簇1，列簇2......  表所有目录 hbase/库目录/表目录/region0/ |

## 2、增

|  |
| --- |
| put 'user\_info','001','base\_info:username','zhagnsan'  put 表名,行键,列簇 : key, value |

## 3、查

|  |
| --- |
| 1、get 'user\_info','001','base\_info:username'  get 表名,行键,列簇 : key  列族和key，可省。  2、scan 表名 查询所有数据  3、hbase的一个特性 |

## 4、删

|  |
| --- |
| 1、delete ’user\_info’,’001’,’base\_info:age’  delete 表名，行键，列族名：key  delete 只能删除一个key，不能删除行 |
| 2、deleteall ‘user\_info’,’001’  delete 表名，行键  3、drop ‘user\_info’ 删表  删除前，需要先停用 |

# 四、特性

## 1、排序

|  |
| --- |
| 排序规则：行键，列族名，列名；按字典顺序 |

## 2、二进制数据存储

|  |
| --- |
| 所有数据都是二进制进行存储。  因此没有数据类型  只有Bytes数组 |

## 3、存储

|  |
| --- |
| 较小的数据，先存储在一个hRegionServer中，同时也会在hdfs中记录日志。若这个 hRegionServer挂掉，其它hRegioinServer会读取日志，将数据恢复 |

# 五、JAVA API

|  |
| --- |
| 代码见hbase  需要在运行本机的host文件中添加zookeeper服务器的映射 |

# 六、批量数据快速导入

# 七、HBASE 性能调优

# 八、MAPREDUCE分析HAVSE中的数据

# 九、HBASE中的数据 建索引

## 9.1.过滤查询

## 9.2 二级索引

# 十、作业

日志查询web系统