

HBase:

应用场景及特点

概念定义

架构模型

安装部署

HBase shell使用

1. 应用场景

100ms

上百亿行. 上百万列. 实时数据查询.

交通.

金融.

电商.

移动.

2. 特点

传统 500万行. 分表. 分库.

容量大:

高性能

写

LSM数据结构

Row key有序列列

读

region分

主键索引

缓存.

面向列: 列可扩展. 查询只需要少数几个字段. 能大大减少读取列.

多版本: 每一个列的数据存储多个 version.

稀疏性: 为空的列不占用存储空间.

扩展性: 基于HDFS.

高可用:

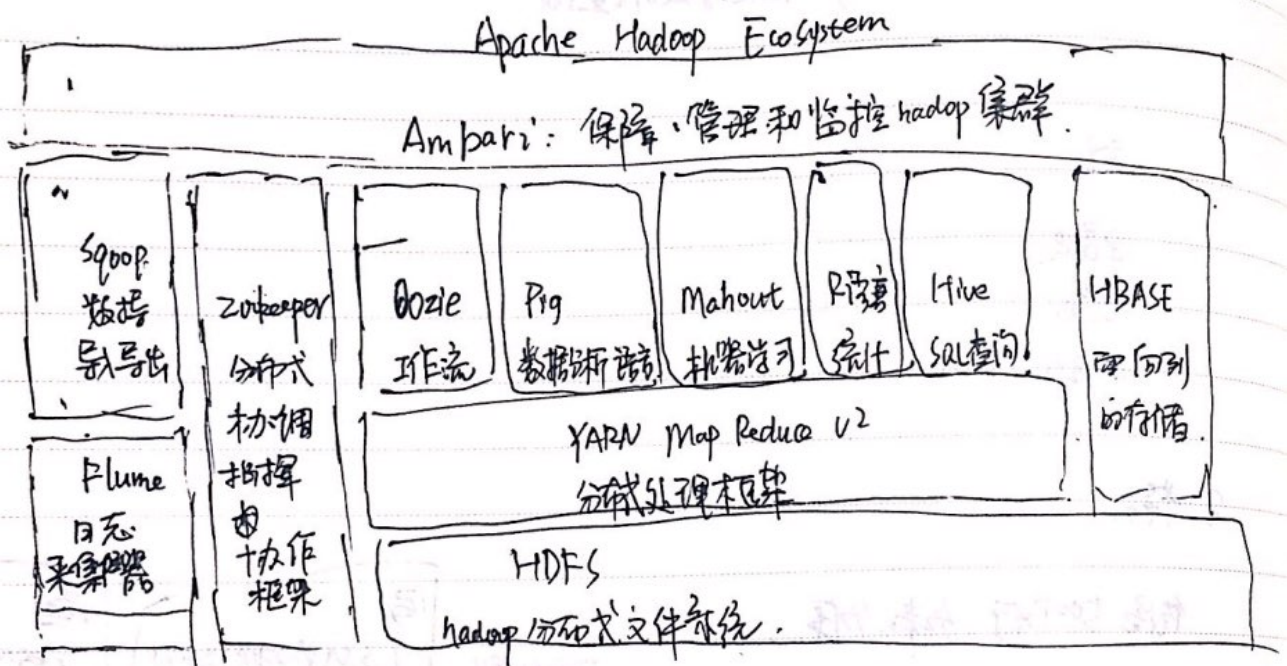


二. HBase 概念.

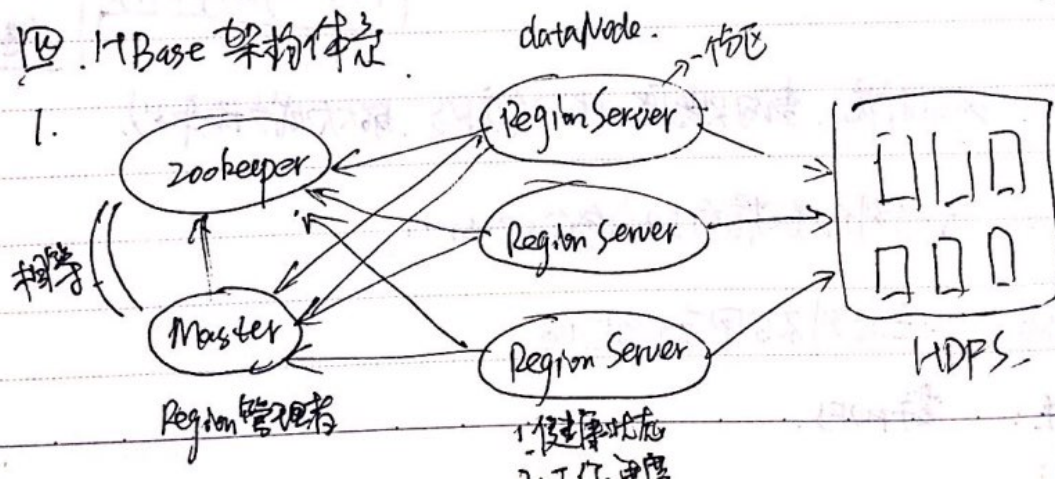
1. hadoop 数据库. 分布式. 高可用. 高性能

2. $\left\{ \begin{array}{l} \text{官网版本} \\ \text{CDH 版本: HBase + CDH 版本号} \end{array} \right.$ $\xrightarrow{\text{CDH}}$ hadoop 的版本保持一致.

三. HBase 在 Hadoop 2.x 生态系统中的定位.



四. HBase 架构体系



2. 数据模型.

个人信息

教育信息

工作经历

姓名 年龄 电话

大学 研究生 博士

工作1 工作2 工作3

- 列族:
- 1. 一张表列族不超过 5.
 - 2. 列族中的列无限制.
 - 3. 列只有插入数据后存在.
 - 4. 列在列族中是有效的序.

HBase

关系型数据库

列存储结构

~~行存储结构~~ ~~需要预定义~~

数据自动切分.

高并发读写.

不支持条件查询

五. 安装.

1. 需求 - ① 1.7 ↑

② Hadoop 2.5.0 ↑

③ zookeeper - 3.4.5

Noodepad++

FTP 插件.



No.

Date:

hdfs 1台机器

2. ① 官网 → document → 2.5.2 → 配置

②. Configuration. $\begin{cases} \text{core-default.xml} \\ \text{hdfs} \\ \text{mr} \\ \text{Yarn} \end{cases}$

1. core-site.xml

$\begin{cases} \text{fs.defaultFS} \\ \text{hdfs://主机名:9000} \end{cases}$

hadoop.tmp.dir

/opt/modules/hadoop-2.5.0/data/tmp

2. hdfs-site.xml

$\begin{cases} \text{dfs.replication} \\ 1 \end{cases}$

dfs.permissions

false

3. slaves

主机名

4. 只放存储, 无需配置 Yarn.

3. zookeeper 分布式安装. 3台机器

① tar xzf zookeeper.tar -C /opt/modules/

②. zoo 2k.cfg 配置, 参考官网

③. 创建 myid, 1. 2. 3.
host1 host2 host3



4. HBase 安装配置

① Hbase-env.sh. {
 1. export JAVA_HOME
 2. export HBASE_MANAGES_ZK=false

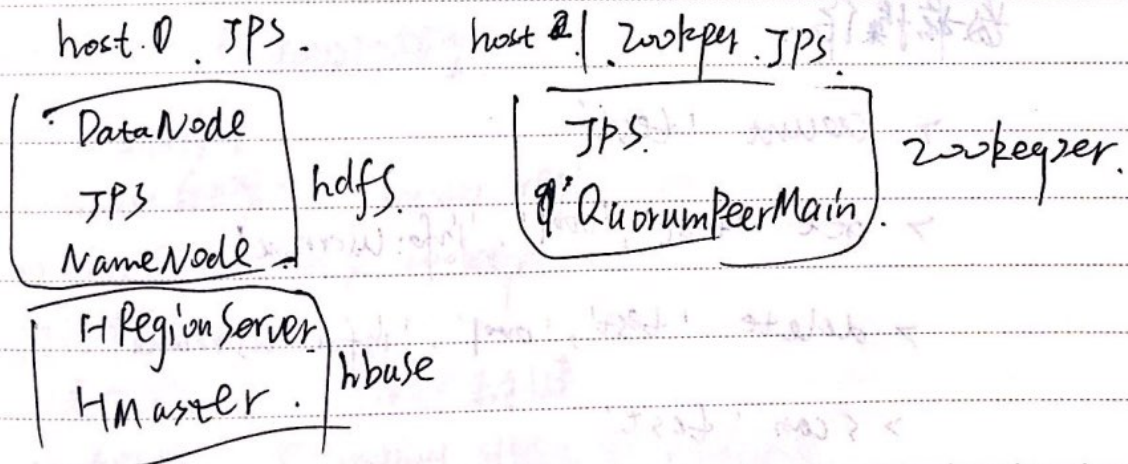
② HBase-site.xml. {
 1. hbase.tmp.dir = /opt/test/hbase/data/temp.
 2. hbase.root = hdfs://~~域名~~:9001/hbase.
 3. hbase.cluster.distributed = true.
 4. hbase.zookeeper.quorum = ~~域名~~, ~~端口~~2, ~~域名~~
 ↓
 都能通过局域网.

③ 配置 register server, 主节点.

5. hbase 目录结构

6. hbase 启动/停止 {
 master
 RegionServer
 1. hbase-daemon.sh
 2. hbase-daemon.sh
 3. start-hbase.sh
 4. stop-hbase.sh
 HBase shell

7. hbase.web
 port: 60010



2 Hbase 表操作.

①. bin/hbase shell

ctrl + space

② creat :test, 'info'

hbase> list

hbase> put 'test', '001', 'info:username', 'henry'

hbase> scan 'test'

> put 'test', '001', 'info:age', '20'

> describe 'test'

> disable 'test'

> is-enabled 'test'

> drop 'test'

数据操作:

> count 'test'

> get 'test', '001', 'info:username'

> delete 'test', '001', 'info:username'

> scan 'test'



> truncate 'test'

disable

drop

creat .

总结:

HBase是一个分布式的数据库

主要作用: 海量数据存储和实时查询

一. 应用场景. 特点.

>> 交通. 金融. 电商. 移动.

某列新行体.

>> 容量大, 可扩展, 高性能, 可扩展, 自动增加, 多副本, 稀疏性.

↓
dataNode

二. HBase的定义和定位

>> 官方文档. hbase的概述描述

>> . hadoop生态中 hbase定位.

三. HBase的架构体系和设计模型

>> 服务架构体系

1) hbase的主进程: master, regionserver

2) hbase所依赖的服务: zookeeper, hdfs.

>> 设计模型

1) 表结构. 表名, 列簇名, 表创建.

2) 表数据. 行. rowkey, 列簇名, 各个列的时间.



四. 安装部署

→ 条件:

- 1) JDK 1.7 ↑
- 2) ^{hdfs} ~~hadoop~~ 2.5.0 ↑
- 3) zookeeper 3.4.x ↑ 集群 3台

→ 部署

- 1) hbase-env.sh jdk, zookeeper
- 2) hbase-site.xml → 官网
- 3) regionserver 主机名. 在哪些台机器上

五. Hbase shell

→ DDL 表: create describe disable enable drop

→ DML 数据: put, delete, get, word scan

