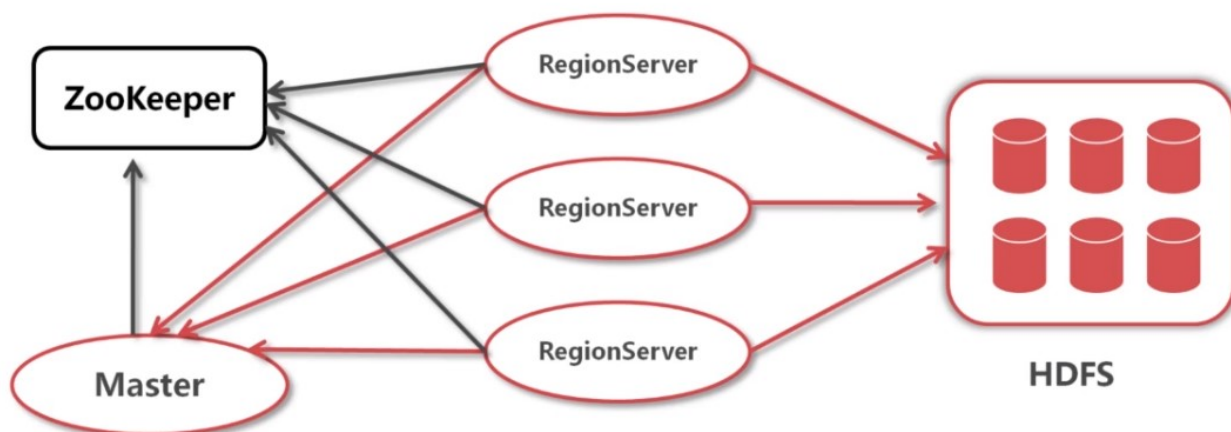


## HBase架构体系



## 资源网站

- <http://hbase.apache.org/>
- <http://archive.apache.org/dist/hbase/>
- <http://archive.cloudera.com/cdh5/>
- <http://blog.csdn.net/liuygvip/article/details/78610797>

## 版本

- CentOS-6.5
- JDK-1.8.0\_101
- Hadoop-2.5.0 (cdh5.2.0)
- Zookeeper-3.4.5 (cdh5.2.0)
- Hbase-0.98.6 (cdh5.2.0)

## 机器信息

机器	应用	端口
192.168.243.18	Hadoop、Hbase	50070、60010
192.168.243.19	Zookeeper	2181

## 特点

- 容量大——HBase单表可以有百亿行、百万列
- 面向列——数据写入之前不需要设定列，列是动态增加的
- 高扩展性——基于HDFS，支持弹性水平伸缩

- 高性能——海量数据的查询能够达到毫秒级别

优点：满足宽表的需求

缺点：不支持复杂查询

## Filter：过滤器，筛选数据

[http://blog.csdn.net/horizon\\_zy/article/details/78362031](http://blog.csdn.net/horizon_zy/article/details/78362031)

[http://blog.csdn.net/horizon\\_zy/article/details/78335975](http://blog.csdn.net/horizon_zy/article/details/78335975)

## Hadoop 单节点安装配置

解压完成后，删掉 `share/doc` 文件夹，和 `etc/hadoop` 下的 `*.cmd` 文件。

配置 `hadoop-env.sh`，修改Java环境变量：

```
1 export JAVA_HOME=/usr/lib/jdk1.8.0_111
```

配置 `core-site.xml`：

```
1 <configuration>
2     <!-- 默认的HDFS路径 -->
3     <property>
4         <name>fs.defaultFS</name>
5         <value>hdfs://centos-1.cipher.com:9000</value>
6     </property>
7     <!-- 临时文件目录 -->
8     <property>
9         <name>hadoop.tmp.dir</name>
10        <value>/opt/modules/hadoop-2.5.0/data/tmp</value>
11    </property>
12 </configuration>
```

配置 `hdfs-site.xml`：

```
1 <configuration>
2     <!-- 副本数量 -->
3     <property>
4         <name>dfs.replication</name>
5         <value>1</value>
6     </property>
7     <!-- 权限 -->
```

```
8     <property>
9         <name>dfs.permissions.enabled</name>
10        <value>>false</value>
11    </property>
12 </configuration>
```

配置 `slaves` ，修改为主机名：

```
1 centos-1.cipher.com
```

启动：

```
1 bin/hdfs namenode -format
```

```
1 # namenode
2 sbin/hadoop-daemon.sh start namenode
3
4 # datanode
5 sbin/hadoop-daemon.sh start datanode
```

开放50070端口：

```
1 # 写入修改
2 /sbin/iptables -I INPUT -p tcp --dport 50070 -j ACCEPT
3 # 保存修改
4 /etc/init.d/iptables save
5 # 重启防火墙，修改生效
6 service iptables restart
```

验证：

<http://10.2.144.118:50070>

## Zookeeper 集群安装配置

在 `/opt/modules/zookeeper-3.4.5/conf` 下，修改 `zoo_sample.cfg` 为 `zoo.cfg` ，编辑内容：

```
1 dataDir=/opt/modules/zookeeper-3.4.5/zkData
2
```

```
3 # 集群信息
4 server.1=centos-1.cipher.com:2888:3888
5 server.2=centos-2.cipher.com:2888:3888
6 server.3=centos-3.cipher.com:2888:3888
```

在 `/opt/modules/zookeeper-3.4.5/zkData` 下创建 `myid` 文件，内容为 `1`

分发文件到其他机器：

```
1 scp -r zookeeper-3.4.5/ centos-2.cipher.com:/opt/modules/zookeeper-3.4.5/
2 scp -r zookeeper-3.4.5/ centos-3.cipher.com:/opt/modules/zookeeper-3.4.5/
```

修改机器的 `myid` 文件，内容为对应的集群号。

启动服务：

```
1 bin/zkServer.sh start
```

## HBase 安装配置

在 `/opt/modules/hbase-0.98.6/conf` 下，修改 `hbase-env.sh`：

```
1 export JAVA_HOME=/usr/lib/jdk1.8.0_111
2 export HBASE_MANAGES_ZK=false
```

修改 `hbase-site.xml`：

```
1 <configuration>
2   <property>
3     <name>hbase.tmp.dir</name>
4     <value>/opt/modules/hbase-0.98.6/data/tmp</value>
5   </property>
6
7   <property>
8     <name>hbase.rootdir</name>
9     <value>hdfs://centos-1.cipher.com:9000/hbase</value>
10  </property>
11
12  <property>
13    <name>hbase.cluster.distributed</name>
14    <value>true</value>
```

```
15     </property>
16
17     <property>
18         <name>hbase.zookeeper.quorum</name>
19         <value>centos-2.cipher.com,centos-3.cipher.com</value>
20     </property>
21 </configuration>
```

修改 `regionservers` :

```
1 centos-1.cipher.com
```

启动 :

```
1 bin/start-hbase.sh
```

## HBase shell 使用

```
1 bin/hbase shell
```

```
1 # 创建表: create 表名,列簇名
2 create 'test', 'info'
3
4 # 扫描表
5 scan 'test'
6
7 # 插入一行数据: put 表名,rowKey,列簇名:列名,列值
8 put 'test', '0001', 'info:username', 'cipher'
9
10 # 禁用表
11 disable 'test'
12
13 # 删除表
14 drop 'test'
```