|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| hM | hM110 | hS1 | hS2 | hS3 | hS4 | hS5 |
| NameNode | DataNode | NameNode | DataNode | DataNode | DataNode | DataNode |
| DataNode | NodeManager | DataNode | NodeManager | HMaster | NodeManager | NodeManager |
| ResourceManager | JobHistoryServer | NodeManager | ResourceManager | HRegionServer | HMaster |  |
| NodeManager |  | JournalNode | JournalNode | Hive | Hive |  |
| JournalNode |  | QuorumPeerMain | JobHistoryServer |  |  |  |
| QuorumPeerMain |  | DFSZKFailoverController | QuorumPeerMain |  |  |  |
| DFSZKFailoverController |  | Kafka | Kafka |  |  |  |
| Kafka |  |  |  |  |  |  |

组件分布：

**zookeeper：**

需要zoo.cfg文件中server.1 server.2 server.3 + dataDir目录下的myid来搭建集群环境

**zoo.cfg**>hadoopMaster、hadoopSlave1、hadoopSlave2

**kafka**：

需要server.properties文件配置zk和broker.id搭建集群环境

**server.properties**>hadoopMaster、hadoopSlave1、hadoopSlave2

**hive**：

需要hdfs-site.xml、yarn-site.xml（这两个会自动在环境变量中找）

hadoopSlave3、hadoopSlave4 （启动hiveserver2进程，ps -ef | grep ‘HiveServer2’）

**\*hive不是集群，hive就是单点的入口，可以通过hiveserver2进程对外提供服务**

**hbase**：

需要hdfs-site.xml、core-site.xml文件+backup-master文件+regionservers文件搭建集群环境

**start-hbase.sh>active-master**：hadoopSlave3

**backup-masters文件>standby-master**：hadoopSlave4

**regionservers文件>regionserver**：hadoopSlave3、hadoopSlave4

**hadoop：**

**hdfs**：

**hdfs-site.xml>nameNode**：hadoopMaster、hadoopSlave1

**hdfs-site.xml>journalNode**：hadoopMaster、hadoopSlave1、hadoopSlave2

**配置加自定义启动进程>ZKFC**：hadoopMaster、hadoopSlave1

**slaves文件>dataNode**：全部

**yarn**：

**yarn-site.xml>resourceManager**：hadoopMaster、hadoopSlave2

**slaves>nodeManager**：全部

**mapreduce**：

**mapred-site.xml>jobHistoryServer**：hadoopSlave2

主机名：hadoopMaster

IP地址：10.83.198.130

|  |  |  |  |
| --- | --- | --- | --- |
| NameNode |  |  |  |
| DataNode |  |  |  |
| ResourceManager |  |  |  |
| NodeManager |  |  |  |
| JournalNode |  |  |  |
| JobHistoryServer |  |  |  |
| QuorumPeerMain |  |  |  |
| DFSZKFailoverController |  |  |  |
| Kafka |  |  |  |

主机名：hadoopMaster110

IP地址：10.82.193.223

|  |  |  |  |
| --- | --- | --- | --- |
| DataNode |  |  |  |
| NodeManager |  |  |  |
| JobHistoryServer |  |  |  |

主机名：hadoopSlave1

IP地址：10.83.198.131

|  |  |  |
| --- | --- | --- |
| NameNode |  |  |
| DataNode |  |  |
| ResourceManager |  |  |
| NodeManager |  |  |
| JournalNode |  |  |
| JobHistoryServer |  |  |
| QuorumPeerMain |  |  |
| DFSZKFailoverController |  |  |
| Kafka |  |  |

主机名：hadoopSlave2

IP地址：10.83.198.132

|  |  |  |  |
| --- | --- | --- | --- |
| DataNode |  |  |  |
| NodeManager |  |  |  |
| JournalNode |  |  |  |
| JobHistoryServer |  |  |  |
| QuorumPeerMain |  |  |  |
| Kafka |  |  |  |

主机名：hadoopSlave3

IP地址：10.93.202.77

|  |  |  |  |
| --- | --- | --- | --- |
| DataNode |  |  |  |
| HMaster |  |  |  |
| HRegionServer |  |  |  |

主机名：hadoopSlave4

IP地址：10.83.198.3

|  |  |  |  |
| --- | --- | --- | --- |
| DataNode |  |  |  |
| NodeManager |  |  |  |
| JobHistoryServer |  |  |  |
| HMaster |  |  |  |
| JobHistoryServer |  |  |  |

主机名：hadoopSlave5

IP地址：10.82.204.70

|  |  |  |  |
| --- | --- | --- | --- |
| DataNode |  |  |  |
| NodeManager |  |  |  |
| JobHistoryServer |  |  |  |

CREATE EXTERNAL TABLE `ad\_statistic\_data`(

`statistic\_type` int COMMENT 'from deserializer',

`project\_code` string COMMENT 'from deserializer',

`medium` string COMMENT 'from deserializer',

`material\_id` int COMMENT 'from deserializer',

`ad\_type` int COMMENT 'from deserializer',

`real\_ip` string COMMENT 'from deserializer',

`upload\_hour` int COMMENT 'from deserializer',

`banner\_location` string COMMENT 'from deserializer',

`country` string COMMENT 'from deserializer',

`province` string COMMENT 'from deserializer',

`city` string COMMENT 'from deserializer',

`region` string COMMENT 'from deserializer')

PARTITIONED BY (

`time\_part` string)

ROW FORMAT SERDE

'org.openx.data.jsonserde.JsonSerDe'

STORED AS INPUTFORMAT

'com.hadoop.mapred.DeprecatedLzoTextInputFormat'

OUTPUTFORMAT

'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'

LOCATION

'hdfs://ns/hive\_table/AdStatisticData