**# Based On**[**Apache Hadoop 2.9.2**](http://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/index.html)

# 配置hadoop进程的环境变量

etc/hadoop/hadoop-env.sh

etc/hadoop/mapred-env.sh

etc/hadoop/yarn-env.sh

管理员使用以上脚本完成hadoop进程环境的site-specific定制化配置。至少，应该配置JAVA\_HOME环境变量。

|  |  |  |
| --- | --- | --- |
| **运行脚本** | **运行进程** | **环境变量** |
| hadoop-env.sh | NameNode | HADOOP\_NAMENODE\_OPTS |
| DataNode | HADOOP\_DATANODE\_OPTS |
| Secondary NameNode | HADOOP\_SECONDARYNAMENODE\_OPTS |
|  | HADOOP\_PID\_DIR |
|  | HADOOP\_LOG\_DIR |
|  | HADOOP\_HEAPSIZE |
| mapred-env.sh | Map Reduce Job History Server  Map Reduce Job History Server | HADOOP\_JOB\_HISTORYSERVER\_OPTS |
| HADOOP\_JOB\_HISTORYSERVER\_HEAPSIZE |
| yarn-env.sh | ResourceManager  ResourceManager | YARN\_RESOURCEMANAGER\_OPTS |
| YARN\_RESOURCEMANAGER\_HEAPSIZE |
| NodeManager  NodeManager | YARN\_NODEMANAGER\_OPTS |
| YARN\_NODEMANAGER\_HEAPSIZE |
| WebAppProxy  WebAppProxy | YARN\_PROXYSERVER\_OPTS |
| YARN\_PROXYSERVER\_HEAPSIZE |
|  | YARN\_HEAPSIZE |

# 配置hadoop进程

## etc/hadoop/core-site.xml

| Parameter | Value | description |
| --- | --- | --- |
| fs.defaultFS | [dfs://host:port/](hdfs://host:port/) | NameNode URI.  The name of the default file system. |
| hadoop.tmp.dir | /tmp/hadoop-${user.name} | A base for other temporary directories. |
| io.file.buffer.size | 131072 |  |

## etc/hadoop/hdfs-site.xml

### 作用于NameNode的配置

| Parameter | Value | description |
| --- | --- | --- |
| dfs.namenode.name.dir | file://${hadoop.tmp.dir}/dfs/name | Determines where on the local filesystem the DFS name node should store the name table(fsimage). |
| dfs.namenode.http-address | 0.0.0.0:50070 | The address and the base port where the dfs namenode web ui will listen on. |
| dfs.hosts  dfs.hosts.exclude | List of permitted/excluded DataNodes. |  |
| dfs.blocksize | 268435456 | The default block size for new files, in bytes. |
| dfs.namenode.handler.count | 10 | The number of Namenode RPC server threads that listen to requests from clients. |

### 作用于SecondaryNameNode的配置

| Parameter | Value | description |
| --- | --- | --- |
| dfs.namenode.checkpoint.dir | file://${hadoop.tmp.dir}/dfs/namesecondary | Determines where on the local filesystem the DFS secondary name node should store the temporary images to merge. |
| dfs.namenode.checkpoint.edits.dir | ${dfs.namenode.checkpoint.dir} | Determines where on the local filesystem the DFS secondary name node should store the temporary edits to merge. |
| dfs.namenode.checkpoint.period | 3600 | The number of seconds between two periodic checkpoints. |
| dfs.namenode.checkpoint.max-retries | 3 | The SecondaryNameNode retries failed checkpointing. |
| dfs.image.compress | false | Should the dfs image be compressed? |
| dfs.image.compression.codec | org.apache.hadoop.io.compress.DefaultCodec | If the dfs image is compressed, how should they be compressed? |
| dfs.image.transfer.timeout | 60000 | Socket timeout for the HttpURLConnection instance used in the image transfer. |
| dfs.namenode.secondary.http-address | 0.0.0.0:50090 | The secondary namenode http server address and port. |
| dfs.namenode.secondary.https-address | 0.0.0.0:50091 | The secondary namenode HTTPS server address and port. |

### 作用于DataNode的配置

| Parameter | Value |  |
| --- | --- | --- |
| dfs.datanode.data.dir | file://${hadoop.tmp.dir}/dfs/data | Determines where on the local filesystem an DFS data node should store its blocks. |
| dfs.datanode.handler.count | 10 | The number of server threads for the datanode. |
| dfs.datanode.address | 0.0.0.0:50010 | The datanode server address and port for data transfer. |
| dfs.datanode.http.address | 0.0.0.0:50075 | The datanode http server address and port. |
| dfs.datanode.ipc.address | 0.0.0.0:50020 | The datanode ipc server address and port. |

## etc/hadoop/yarn-site.xml

### 作用于ResourceManager和NodeManager的配置

| Parameter | Value |
| --- | --- |
| yarn.acl.enable | true / false |
| yarn.admin.acl | Admin ACL |
| yarn.log-aggregation-enable | *false* |

### 作用于ResourceManager的配置

| Parameter | Value | description |
| --- | --- | --- |
| yarn.resourcemanager.hostname | 0.0.0.0 | The hostname of the RM. |
| yarn.resourcemanager.address | ${yarn.resourcemanager.hostname}:8032 | The address of the applications manager interface in the RM. |
| yarn.resourcemanager.scheduler.address | ${yarn.resourcemanager.hostname}:8030 |  |
| yarn.resourcemanager.resource-tracker.address | ${yarn.resourcemanager.hostname}:8031 |  |
| yarn.resourcemanager.admin.address | ${yarn.resourcemanager.hostname}:8033 | The address of the RM admin interface. |
| yarn.resourcemanager.webapp.address | ${yarn.resourcemanager.hostname}:8088 | The http address of the RM web application. |
| yarn.resourcemanager.webapp.https.address | ${yarn.resourcemanager.hostname}:8090 | The https address of the RM web application. |
| yarn.resourcemanager.scheduler.class | org.apache.xxx.CapacityScheduler | The class to use as the resource scheduler. |
| yarn.scheduler.minimum-allocation-mb | 1024 | Minimum limit of memory to allocate to each container request at the Resource Manager. |
| yarn.scheduler.maximum-allocation-mb | 8192 | Maximum limit of memory to allocate to each container request at the Resource Manager. |
| yarn.resourcemanager.nodes.include-path  yarn.resourcemanager.nodes.exclude-path |  | Path to file with nodes to include. |

### 作用于NodeManager的配置

| Parameter | Value |
| --- | --- |
| yarn.nodemanager.hostname | The hostname of the NM. |
| yarn.nodemanager.address | The address of the container manager in the NM. |
| yarn.nodemanager.resource.memory-mb | Resource i.e. available physical memory, in MB, for given NodeManager |
| yarn.nodemanager.vmem-pmem-ratio | Maximum ratio by which virtual memory usage of tasks may exceed physical memory |
| yarn.nodemanager.local-dirs | Comma-separated list of paths on the local filesystem where intermediate data is written. |
| yarn.nodemanager.log-dirs | Comma-separated list of paths on the local filesystem where logs are written. |
| yarn.nodemanager.log.retain-seconds | 10800 |
| yarn.nodemanager.remote-app-log-dir | */logs* |
| yarn.nodemanager.remote-app-log-dir-suffix | *logs* |
| yarn.nodemanager.aux-services | mapreduce\_shuffle |

**# Monitoring Health of NodeManagers**

| Parameter | Value |
| --- | --- |
| yarn.nodemanager.health-checker.script.path | Node health script |
| yarn.nodemanager.health-checker.script.opts | Node health script options |
| yarn.nodemanager.health-checker.interval-ms | Node health script interval |
| yarn.nodemanager.health-checker.script.timeout-ms | Node health script timeout interval |

### 作用于HistoryServer的配置

| Parameter | Value |
| --- | --- |
| yarn.log-aggregation.retain-seconds | *-1* |
| yarn.log-aggregation.retain-check-interval-seconds | *-1* |

## etc/hadoop/mapred-site.xml

### 作用于MapReduce应用的配置

| **Parameter** | **Value** |  |  |
| --- | --- | --- | --- |
| mapreduce.jobhistory.address | 0.0.0.0:10020 | MapReduce JobHistory Server IPC host:port |  |
| mapreduce.jobhistory.webapp.address | 0.0.0.0:19888 | MapReduce JobHistory Server Web UI host:port |  |
| mapreduce.jobhistory.webapp.https.address | 0.0.0.0:19890 | The https address the MapReduce JobHistory Server WebApp is on. |  |
| mapreduce.framework.name | yarn |  |  |
| mapreduce.map.memory.mb | 1536 |  |  |
| mapreduce.map.java.opts | -Xmx1024M |  |  |
| mapreduce.reduce.memory.mb | 3072 |  |  |
| mapreduce.reduce.java.opts | -Xmx2560M |  |  |
| mapreduce.task.io.sort.mb | 512 |  |  |
| mapreduce.task.io.sort.factor | 100 |  |  |
| mapreduce.reduce.shuffle.parallelcopies | 50 |  |  |

### 作用于MapReduce JobHistory Server的配置

| **Parameter** | **Value** |
| --- | --- |
| mapreduce.jobhistory.address | MapReduce JobHistory Server *host:port* |
| mapreduce.jobhistory.webapp.address | MapReduce JobHistory Server Web UI *host:port* |
| mapreduce.jobhistory.intermediate-done-dir | /mr-history/tmp |
| mapreduce.jobhistory.done-dir | /mr-history/done |

## etc/hadoop/slaves

List all slave hostnames or IP addresses in your etc/hadoop/slaves file, one per line.

## Web Interfaces

| **Daemon** | **Web Interface** | **Notes** |
| --- | --- | --- |
| NameNode | <http://nn_host:port/> | Default HTTP port is 50070. |
| ResourceManager | <http://rm_host:port/> | Default HTTP port is 8088. |
| MapReduce JobHistory Server | <http://jhs_host:port/> | Default HTTP port is 19888. |