**1       获取默认配置**

配置hadoop，主要是配置core-site.xml,hdfs-site.xml,mapred-site.xml三个配置文件，默认下来，这些配置文件都是空的，所以很难知道这些配置文件有哪些配置可以生效，上网找的配置可能因为各个hadoop版本不同，导致无法生效。浏览更多的配置，有两个方法:

1.选择相应版本的hadoop,下载解压后，搜索\*.xml,找到core-default.xml,hdfs-default.xml,mapred-default.xml,这些就是默认配置,可以参考这些配置的说明和key，配置hadoop集群。

2.浏览apache官网,三个配置文件链接如下:

   http://hadoop.apache.org/common/docs/current/core-default.html

   http://hadoop.apache.org/common/docs/current/hdfs-default.html

   http://hadoop.apache.org/common/docs/current/mapred-default.html

   这里是浏览hadoop当前版本号的默认配置文件，其他版本号，要另外去官网找。其中第一个方法找到默认的配置是最好的，因为每个属性都有说明，可以直接使用。另外，core-site.xml是全局配置,hdfs-site.xml和mapred-site.xml分别是hdfs和mapred的局部配置。

**2       常用的端口配置**

***2.1  HDFS端口***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 参数 | 描述 | 默认 | 配置文件 | 例子值 |
| fs.default.name namenode | namenode RPC交互端口 | 8020 | core-site.xml | hdfs://master:8020/ |
| dfs.http.address | NameNode web管理端口 | 50070 | hdfs- site.xml | 0.0.0.0:50070 |
| dfs.datanode.address | datanode　控制端口 | 50010 | hdfs -site.xml | 0.0.0.0:50010 |
| dfs.datanode.ipc.address | datanode的RPC服务器地址和端口 | 50020 | hdfs-site.xml | 0.0.0.0:50020 |
| dfs.datanode.http.address | datanode的HTTP服务器和端口 | 50075 | hdfs-site.xml | 0.0.0.0:50075 |

***2.2  MR端口***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 参数 | 描述 | 默认 | 配置文件 | 例子值 |
| mapred.job.tracker | job-tracker交互端口 | 8021 | mapred-site.xml | hdfs://master:8021/ |
| job | tracker的web管理端口 | 50030 | mapred-site.xml | 0.0.0.0:50030 |
| mapred.task.tracker.http.address | task-tracker的HTTP端口 | 50060 | mapred-site.xml | 0.0.0.0:50060 |

***2.3  其它端口***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 参数 | 描述 | 默认 | 配置文件 | 例子值 |
| dfs.secondary.http.address | secondary NameNode web管理端口 | 50090 | hdfs-site.xml | 0.0.0.0:50090 |
|  |  |  |  |  |
|  |  |  |  |  |

**3       三个缺省配置参考文件说明**

***3.1  core-default.html***

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 参数名 | 参数值 | 参数说明 |
| 1 | hadoop.tmp.dir | /tmp/hadoop-${user.name} | 临时目录设定 |
| 2 | hadoop.native.lib | true | 使用本地hadoop库标识。 |
| 3 | hadoop.http.filter.initializers |  | http服务器过滤链设置 |
| 4 | hadoop.security.group.mapping | org.apache.hadoop.security.ShellBasedUnixGroupsMapping | 组内用户的列表的类设定 |
| 5 | hadoop.security.authorization | false | 服务端认证开启 |
| 6 | hadoop.security.authentication | simple | 无认证或认证设置 |
| 7 | hadoop.security.token.service.use\_ip | true | 是否开启使用IP地址作为连接的开关 |
| 8 | hadoop.logfile.size | 10000000 | 日志文件最大为10M |
| 9 | hadoop.logfile.count | 10 | 日志文件数量为10个 |
| 10 | io.file.buffer.size | 4096 | 流文件的缓冲区为4K |
| 11 | io.bytes.per.checksum | 512 | 校验位数为512字节 |
| 12 | io.skip.checksum.errors | false | 校验出错后是抛出异常还是略过标识。True则略过。 |
| 13 | io.compression.codecs | org.apache.hadoop.io.compress.DefaultCodec,  org.apache.hadoop.io.compress.GzipCodec,  org.apache.hadoop.io.compress.BZip2Codec,  org.apache.hadoop.io.compress.SnappyCodec | 压缩和解压的方式设置 |
| 14 | io.serializations | org.apache.hadoop.io.serializer.WritableSerialization | 序例化和反序列化的类设定 |
| 15 | fs.default.name | file:/// | 缺省的文件URI标识设定。 |
| 16 | fs.trash.interval | 0 | 文件废弃标识设定，0为禁止此功能 |
| 17 | fs.file.impl | org.apache.hadoop.fs.LocalFileSystem | 本地文件操作类设置 |
| 18 | fs.hdfs.impl | org.apache.hadoop.hdfs.DistributedFileSystem | HDFS文件操作类设置 |
| 19 | fs.s3.impl | org.apache.hadoop.fs.s3.S3FileSystem | S3文件操作类设置 |
| 20 | fs.s3n.impl | org.apache.hadoop.fs.s3native.NativeS3FileSystem | S3文件本地操作类设置 |
| 21 | fs.kfs.impl | org.apache.hadoop.fs.kfs.KosmosFileSystem | KFS文件操作类设置. |
| 22 | fs.hftp.impl | org.apache.hadoop.hdfs.HftpFileSystem | HTTP方式操作文件设置 |
| 23 | fs.hsftp.impl | org.apache.hadoop.hdfs.HsftpFileSystem | HTTPS方式操作文件设置 |
| 24 | fs.webhdfs.impl | org.apache.hadoop.hdfs.web.WebHdfsFileSystem | WEB方式操作文件类设置 |
| 25 | fs.ftp.impl | org.apache.hadoop.fs.ftp.FTPFileSystem | FTP文件操作类设置 |
| 26 | fs.ramfs.impl | org.apache.hadoop.fs.InMemoryFileSystem | 内存文件操作类设置 |
| 27 | fs.har.impl | org.apache.hadoop.fs.HarFileSystem | 压缩文件操作类设置. |
| 28 | fs.har.impl.disable.cache | true | 是否缓存har文件的标识设定 |
| 29 | fs.checkpoint.dir | ${hadoop.tmp.dir}/dfs/namesecondary | 备份名称节点的存放目前录设置 |
| 30 | fs.checkpoint.edits.dir | ${fs.checkpoint.dir} | 备份名称节点日志文件的存放目前录设置 |
| 31 | fs.checkpoint.period | 3600 | 动态检查的间隔时间设置 |
| 32 | fs.checkpoint.size | 67108864 | 日志文件大小为64M |
| 33 | fs.s3.block.size | 67108864 | 写S3文件系统的块的大小为64M |
| 34 | fs.s3.buffer.dir | ${hadoop.tmp.dir}/s3 | S3文件数据的本地存放目录 |
| 35 | fs.s3.maxRetries | 4 | S3文件数据的偿试读写次数 |
| 36 | fs.s3.sleepTimeSeconds | 10 | S3文件偿试的间隔 |
| 37 | local.cache.size | 10737418240 | 缓存大小设置为10GB |
| 38 | io.seqfile.compress.blocksize | 1000000 | 压缩流式文件中的最小块数为100万 |
| 39 | io.seqfile.lazydecompress | true | 块是否需要压缩标识设定 |
| 40 | io.seqfile.sorter.recordlimit | 1000000 | 内存中排序记录块类最小为100万 |
| 41 | io.mapfile.bloom.size | 1048576 | BloomMapFiler过滤量为1M |
| 42 | io.mapfile.bloom.error.rate | 0.005 |  |
| 43 | hadoop.util.hash.type | murmur | 缺少hash方法为murmur |
| 44 | ipc.client.idlethreshold | 4000 | 连接数据最小阀值为4000 |
| 45 | ipc.client.kill.max | 10 | 一个客户端连接数最大值为10 |
| 46 | ipc.client.connection.maxidletime | 10000 | 断开与服务器连接的时间最大为10秒 |
| 47 | ipc.client.connect.max.retries | 10 | 建立与服务器连接的重试次数为10次 |
| 48 | ipc.server.listen.queue.size | 128 | 接收客户连接的监听队例的长度为128 |
| 49 | ipc.server.tcpnodelay | false | 开启或关闭服务器端TCP连接算法 |
| 50 | ipc.client.tcpnodelay | false | 开启或关闭客户端TCP连接算法 |
| 51 | webinterface.private.actions | false | Web交互的行为设定 |
| 52 | hadoop.rpc.socket.factory.class.default | org.apache.hadoop.net.StandardSocketFactory | 缺省的socket工厂类设置 |
| 53 | hadoop.rpc.socket.factory.class.ClientProtocol |  | 与dfs连接时的缺省socket工厂类 |
| 54 | hadoop.socks.server |  | 服务端的工厂类缺省设置为SocksSocketFactory. |
| 55 | topology.node.switch.mapping.impl | org.apache.hadoop.net.ScriptBasedMapping |  |
| 56 | topology.script.file.name |  |  |
| 57 | topology.script.number.args | 100 | 参数数量最多为100 |
| 58 | hadoop.security.uid.cache.secs | 14400 |  |

***3.2  hdfs-default.html***

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 参数名 | 参数值 | 参数说明 |
| 1 | dfs.namenode.logging.level | info | 输出日志类型 |
| 2 | dfs.secondary.http.address | 0.0.0.0:50090 | 备份名称节点的http协议访问地址与端口 |
| 3 | dfs.datanode.address | 0.0.0.0:50010 | 数据节点的TCP管理服务地址和端口 |
| 4 | dfs.datanode.http.address | 0.0.0.0:50075 | 数据节点的HTTP协议访问地址和端口 |
| 5 | dfs.datanode.ipc.address | 0.0.0.0:50020 | 数据节点的IPC服务访问地址和端口 |
| 6 | dfs.datanode.handler.count | 3 | 数据节点的服务连接处理线程数 |
| 7 | dfs.http.address | 0.0.0.0:50070 | 名称节点的http协议访问地址与端口 |
| 8 | dfs.https.enable | false | 支持https访问方式标识 |
| 9 | dfs.https.need.client.auth | false | 客户端指定https访问标识 |
| 10 | dfs.https.server.keystore.resource | ssl-server.xml | Ssl密钥服务端的配置文件 |
| 11 | dfs.https.client.keystore.resource | ssl-client.xml | Ssl密钥客户端的配置文件 |
| 12 | dfs.datanode.https.address | 0.0.0.0:50475 | 数据节点的HTTPS协议访问地址和端口 |
| 13 | dfs.https.address | 0.0.0.0:50470 | 名称节点的HTTPS协议访问地址和端口 |
| 14 | dfs.datanode.dns.interface | default | 数据节点采用IP地址标识 |
| 15 | dfs.datanode.dns.nameserver | default | 指定DNS的IP地址 |
| 16 | dfs.replication.considerLoad | true | 加载目标或不加载的标识 |
| 17 | dfs.default.chunk.view.size | 32768 | 浏览时的文件块大小设置为32K |
| 18 | dfs.datanode.du.reserved | 0 | 每个卷预留的空闲空间数量 |
| 19 | dfs.name.dir | ${hadoop.tmp.dir}/dfs/name | 存贮在本地的名字节点数据镜象的目录,作为名字节点的冗余备份 |
| 20 | dfs.name.edits.dir | ${dfs.name.dir} | 存贮文件操作过程信息的存贮目录 |
| 21 | dfs.web.ugi | webuser,webgroup | Web接口访问的用户名和组的帐户设定 |
| 22 | dfs.permissions | true | 文件操作时的权限检查标识。 |
| 23 | dfs.permissions.supergroup | supergroup | 超级用户的组名定义 |
| 24 | dfs.block.access.token.enable | false | 数据节点访问令牌标识 |
| 25 | dfs.block.access.key.update.interval | 600 | 升级访问钥时的间隔时间 |
| 26 | dfs.block.access.token.lifetime | 600 | 访问令牌的有效时间 |
| 27 | dfs.data.dir | ${hadoop.tmp.dir}/dfs/data | 数据节点的块本地存放目录 |
| 28 | dfs.datanode.data.dir.perm | 755 | 数据节点的存贮块的目录访问权限设置 |
| 29 | dfs.replication | 3 | 缺省的块复制数量 |
| 30 | dfs.replication.max | 512 | 块复制的最大数量 |
| 31 | dfs.replication.min | 1 | 块复制的最小数量 |
| 32 | dfs.block.size | 67108864 | 缺省的文件块大小为64M |
| 33 | dfs.df.interval | 60000 | 磁盘空间统计间隔为6秒 |
| 34 | dfs.client.block.write.retries | 3 | 块写入出错时的重试次数 |
| 35 | dfs.blockreport.intervalMsec | 3600000 | 块的报告间隔时为1小时 |
| 36 | dfs.blockreport.initialDelay | 0 | 块顺序报告的间隔时间 |
| 37 | dfs.heartbeat.interval | 3 | 数据节点的心跳检测间隔时间 |
| 38 | dfs.namenode.handler.count | 10 | 名称节点的连接处理的线程数量 |
| 39 | dfs.safemode.threshold.pct | 0.999f | 启动安全模式的阀值设定 |
| 40 | dfs.safemode.extension | 30000 | 当阀值达到量值后扩展的时限 |
| 41 | dfs.balance.bandwidthPerSec | 1048576 | 启动负载均衡的数据节点可利用带宽最大值为1M |
| 42 | dfs.hosts |  | 可与名称节点连接的主机地址文件指定。 |
| 43 | dfs.hosts.exclude |  | 不充计与名称节点连接的主机地址文件设定 |
| 44 | dfs.max.objects | 0 | 文件数、目录数、块数的最大数量 |
| 45 | dfs.namenode.decommission.interval | 30 | 名称节点解除命令执行时的监测时间周期 |
| 46 | dfs.namenode.decommission.nodes.per.interval | 5 | 名称节点解除命令执行是否完检测次数 |
| 47 | dfs.replication.interval | 3 | 名称节点计算数据节点的复制工作的周期数. |
| 48 | dfs.access.time.precision | 3600000 | 充许访问文件的时间精确到1小时 |
| 49 | dfs.support.append | false | 是否充许链接文件指定 |
| 50 | dfs.namenode.delegation.key.update-interval | 86400000 | 名称节点上的代理令牌的主key的更新间隔时间为24小时 |
| 51 | dfs.namenode.delegation.token.max-lifetime | 604800000 | 代理令牌的有效时间最大值为7天 |
| 52 | dfs.namenode.delegation.token.renew-interval | 86400000 | 代理令牌的更新时间为24小时 |
| 53 | dfs.datanode.failed.volumes.tolerated | 0 | 决定停止数据节点提供服务充许卷的出错次数。0次则任何卷出错都要停止数据节点 |

***3.3  mapred-default.html***

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 参数名 | 参数值 | 参数说明 |
| 1 | hadoop.job.history.location |  | 作业跟踪管理器的静态历史文件的存放目录。 |
| 2 | hadoop.job.history.user.location |  | 可以指定具体某个作业的跟踪管理器的历史文件存放目录 |
| 3 | mapred.job.tracker.history.completed.location |  | 已完成作业的历史文件的存放目录 |
| 4 | io.sort.factor | 10 | 排完序的文件的合并时的打开文件句柄数 |
| 5 | io.sort.mb | 100 | 排序文件的内存缓存大小为100M |
| 6 | io.sort.record.percent | 0.05 | 排序线程阻塞的内存缓存剩余比率 |
| 7 | io.sort.spill.percent | 0.80 | 当缓冲占用量为该值时，线程需要将内容先备份到磁盘中。 |
| 8 | io.map.index.skip | 0 | 索引条目的间隔设定 |
| 9 | mapred.job.tracker | local | 作业跟踪管理器是否和MR任务在一个进程中 |
| 10 | mapred.job.tracker.http.address | 0.0.0.0:50030 | 作业跟踪管理器的HTTP服务器访问端口和地址 |
| 11 | mapred.job.tracker.handler.count | 10 | 作业跟踪管理器的管理线程数,线程数比例是任务管理跟踪器数量的0.04 |
| 12 | mapred.task.tracker.report.address | 127.0.0.1:0 | 任务管理跟踪器的主机地址和端口地址 |
| 13 | mapred.local.dir | ${hadoop.tmp.dir}/mapred/local | MR的中介数据文件存放目录 |
| 14 | mapred.system.dir | ${hadoop.tmp.dir}/mapred/system | MR的控制文件存放目录 |
| 15 | mapreduce.jobtracker.staging.root.dir | ${hadoop.tmp.dir}/mapred/staging | 每个正在运行作业文件的存放区 |
| 16 | mapred.temp.dir | ${hadoop.tmp.dir}/mapred/temp | MR临时共享文件存放区 |
| 17 | mapred.local.dir.minspacestart | 0 | MR本地中介文件删除时，不充许有任务执行的数量值。 |
| 18 | mapred.local.dir.minspacekill | 0 | MR本地中介文件删除时，除非所有任务都已完成的数量值。 |
| 19 | mapred.tasktracker.expiry.interval | 600000 | 任务管理跟踪器不发送心跳的累计时间间隔超过600秒，则任务管理跟踪器失效 |
| 20 | mapred.tasktracker.resourcecalculatorplugin |  | 指定的一个用户访问资源信息的类实例 |
| 21 | mapred.tasktracker.taskmemorymanager.monitoring-interval | 5000 | 监控任务管理跟踪器任务内存使用率的时间间隔 |
| 22 | mapred.tasktracker.tasks.sleeptime-before-sigkill | 5000 | 发出进程终止后，间隔5秒后发出进程消亡信号 |
| 23 | mapred.map.tasks | 2 | 每个作业缺省的map任务数为2 |
| 24 | mapred.reduce.tasks | 1 | 每个作业缺省的reduce任务数为1 |
| 25 | mapreduce.tasktracker.outofband.heartbeat | false | 让在任务结束后发出一个额外的心跳信号 |
| 26 | mapreduce.tasktracker.outofband.heartbeat.damper | 1000000 | 当额外心跳信号发出量太多时，则适当阻止 |
| 27 | mapred.jobtracker.restart.recover | false | 充许任务管理器恢复时采用的方式 |
| 28 | mapred.jobtracker.job.history.block.size | 3145728 | 作业历史文件块的大小为3M |
| 29 | mapreduce.job.split.metainfo.maxsize | 10000000 | 分隔元信息文件的最大值是10M以下 |
| 30 | mapred.jobtracker.taskScheduler | org.apache.hadoop.mapred.JobQueueTaskScheduler | 设定任务的执行计划实现类 |
| 31 | mapred.jobtracker.taskScheduler.maxRunningTasksPerJob |  | 作业同时运行的任务数的最大值 |
| 32 | mapred.map.max.attempts | 4 | Map任务的重试次数 |
| 33 | mapred.reduce.max.attempts | 4 | Reduce任务的重试次数 |
| 34 | mapred.reduce.parallel.copies | 5 | 在复制阶段时reduce并行传送的值。 |
| 35 | mapreduce.reduce.shuffle.maxfetchfailures | 10 | 取map输出的最大重试次数 |
| 36 | mapreduce.reduce.shuffle.connect.timeout | 180000 | REDUCE任务连接任务管理器获得map输出时的总耗时是3分钟 |
| 37 | mapreduce.reduce.shuffle.read.timeout | 180000 | REDUCE任务等待map输出数据的总耗时是3分钟 |
| 38 | mapred.task.timeout | 600000 | 如果任务无读无写时的时间耗时为10分钟，将被终止 |
| 39 | mapred.tasktracker.map.tasks.maximum | 2 | 任管管理器可同时运行map任务数为2 |
| 40 | mapred.tasktracker.reduce.tasks.maximum | 2 | 任管管理器可同时运行reduce任务数为2 |
| 41 | mapred.jobtracker.completeuserjobs.maximum | 100 | 当用户的完成作业数达100个后，将其放入作业历史文件中 |
| 42 | mapreduce.reduce.input.limit | -1 | Reduce输入量的限制。 |
| 43 | mapred.job.tracker.retiredjobs.cache.size | 1000 | 作业状态为已不在执行的保留在内存中的量为1000 |
| 44 | mapred.job.tracker.jobhistory.lru.cache.size | 5 | 作业历史文件装载到内存的数量 |
| 45 | mapred.child.java.opts | -Xmx200m | 启动task管理的子进程时的内存设置 |
| 46 | mapred.child.env |  | 子进程的参数设置 |
| 47 | mapred.child.ulimit |  | 虚拟机所需内存的设定。 |
| 48 | mapred.cluster.map.memory.mb | -1 |  |
| 49 | mapred.cluster.reduce.memory.mb | -1 |  |
| 50 | mapred.cluster.max.map.memory.mb | -1 |  |
| 51 | mapred.cluster.max.reduce.memory.mb | -1 |  |
| 52 | mapred.job.map.memory.mb | -1 |  |
| 53 | mapred.job.reduce.memory.mb | -1 |  |
| 54 | mapred.child.tmp | /tmp | Mr任务信息的存放目录 |
| 55 | mapred.inmem.merge.threshold | 1000 | 内存中的合并文件数设置 |
| 56 | mapred.job.shuffle.merge.percent | 0.66 |  |
| 57 | mapred.job.shuffle.input.buffer.percent | 0.70 |  |
| 58 | mapred.job.reduce.input.buffer.percent | 0.0 |  |
| 59 | mapred.map.tasks.speculative.execution | true | Map任务的多实例并行运行标识 |
| 60 | mapred.reduce.tasks.speculative.execution | true | Reduce任务的多实例并行运行标识 |
| 61 | mapred.job.reuse.jvm.num.tasks | 1 | 每虚拟机运行的任务数 |
| 62 | mapred.min.split.size | 0 | Map的输入数据被分解的块数设置 |
| 63 | mapred.jobtracker.maxtasks.per.job | -1 | 一个单独作业的任务数设置 |
| 64 | mapred.submit.replication | 10 | 提交作业文件的复制级别 |
| 65 | mapred.tasktracker.dns.interface | default | 任务管理跟踪器是否报告IP地址名的开关 |
| 66 | mapred.tasktracker.dns.nameserver | default | 作业和任务管理跟踪器之间通讯方式采用的DNS服务的主机名或IP地址 |
| 67 | tasktracker.http.threads | 40 | http服务器的工作线程数量 |
| 68 | mapred.task.tracker.http.address | 0.0.0.0:50060 | 任务管理跟踪器的http服务器的地址和端口 |
| 69 | keep.failed.task.files | false | 失败任务是否保存到文件中 |
| 70 | mapred.output.compress | false | 作业的输出是否压缩 |
| 71 | mapred.output.compression.type | RECORD | 作业输出采用NONE, RECORD or BLOCK三种方式中一种压缩的写入到流式文件 |
| 72 | mapred.output.compression.codec | org.apache.hadoop.io.compress.DefaultCodec | 压缩类的设置 |
| 73 | mapred.compress.map.output | false | Map的输出是否压缩 |
| 74 | mapred.map.output.compression.codec | org.apache.hadoop.io.compress.DefaultCodec | Map的输出压缩的实现类指定 |
| 75 | map.sort.class | org.apache.hadoop.util.QuickSort | 排序键的排序类指定 |
| 76 | mapred.userlog.limit.kb | 0 | 每个任务的用户日志文件大小 |
| 77 | mapred.userlog.retain.hours | 24 | 作业完成后的用户日志留存时间为24小时 |
| 78 | mapred.user.jobconf.limit | 5242880 | Jobconf的大小为5M |
| 79 | mapred.hosts |  | 可与作业管理跟踪器连接的主机名 |
| 80 | mapred.hosts.exclude |  | 不可与作业管理跟踪器连接的主机名 |
| 81 | mapred.heartbeats.in.second | 100 | 作业管理跟踪器的每秒中到达的心跳数量为100 |
| 82 | mapred.max.tracker.blacklists | 4 | 任务管理跟踪器的黑名单列表的数量 |
| 83 | mapred.jobtracker.blacklist.fault-timeout-window | 180 | 任务管理跟踪器超时180分钟则訪任务将被重启 |
| 84 | mapred.jobtracker.blacklist.fault-bucket-width | 15 |  |
| 85 | mapred.max.tracker.failures | 4 | 任务管理跟踪器的失败任务数设定 |
| 86 | jobclient.output.filter | FAILED | 控制任务的用户日志输出到作业端时的过滤方式 |
| 87 | mapred.job.tracker.persist.jobstatus.active | false | 是否持久化作业管理跟踪器的信息 |
| 88 | mapred.job.tracker.persist.jobstatus.hours | 0 | 持久化作业管理跟踪器的信息的保存时间 |
| 89 | mapred.job.tracker.persist.jobstatus.dir | /jobtracker/jobsInfo | 作业管理跟踪器的信息存放目录 |
| 90 | mapreduce.job.complete.cancel.delegation.tokens | true | 恢复时是否变更领牌 |
| 91 | mapred.task.profile | false | 任务分析信息是否建设标志 |
| 92 | mapred.task.profile.maps | 0-2 | 设置map任务的分析范围 |
| 93 | mapred.task.profile.reduces | 0-2 | 设置reduce任务的分析范围 |
| 94 | mapred.line.input.format.linespermap | 1 | 每次切分的行数设置 |
| 95 | mapred.skip.attempts.to.start.skipping | 2 | 在跳转模式未被设定的情况下任务的重试次数 |
| 96 | mapred.skip.map.auto.incr.proc.count | true | MapRunner在调用map功能后的增量处理方式设置 |
| 97 | mapred.skip.reduce.auto.incr.proc.count | true | 在调用reduce功能后的增量处理方式设置 |
| 98 | mapred.skip.out.dir |  | 跳过记录的输出目录 |
| 99 | mapred.skip.map.max.skip.records | 0 |  |
| 100 | mapred.skip.reduce.max.skip.groups | 0 |  |
| 101 | job.end.retry.attempts | 0 | Hadoop偿试连接通知器的次数 |
| 102 | job.end.retry.interval | 30000 | 通知偿试回应的间隔操作为30秒 |
| 103 | hadoop.rpc.socket.factory.class.JobSubmissionProtocol |  | 指定与作业跟踪管理器的通讯方式，缺省是采用rpc方式 |
| 104 | mapred.task.cache.levels | 2 | 任务缓存级别设置 |
| 105 | mapred.queue.names | default | 分隔作业队例的分隔符设定 |
| 106 | mapred.acls.enabled | false | 指定ACL访问控制列表 |
| 107 | mapred.queue.default.state | RUNNING | 定义队列的状态 |
| 108 | mapred.job.queue.name | default | 已提交作业的队列设定 |
| 109 | mapreduce.job.acl-modify-job |  | 指定可修改作业的ACL列表 |
| 110 | mapreduce.job.acl-view-job |  | 指定可浏临作业的ACL列表 |
| 111 | mapred.tasktracker.indexcache.mb | 10 | 任务管理跟踪器的索引内存的最大容器 |
| 112 | mapred.combine.recordsBeforeProgress | 10000 | 在聚合处理时的记录块数 |
| 113 | mapred.merge.recordsBeforeProgress | 10000 | 在汇总处理时的记录块数 |
| 114 | mapred.reduce.slowstart.completed.maps | 0.05 |  |
| 115 | mapred.task.tracker.task-controller | org.apache.hadoop.mapred.DefaultTaskController | 任务管理器的设定 |
| 116 | mapreduce.tasktracker.group |  | 任务管理器的组成员设定 |
| 117 | mapred.healthChecker.script.path |  | 脚本的绝对路径指定，这些脚本是心跳服务的 |
| 118 | mapred.healthChecker.interval | 60000 | 节点心跳信息的间隔 |
| 119 | mapred.healthChecker.script.timeout | 600000 |  |
| 120 | mapred.healthChecker.script.args |  | 参数列表 |
| 121 | mapreduce.job.counters.limit | 120 | 作业计数器的最小值 |
|  |  |  |  |