Hadoop作业调优参数

第一部分：core-site.xml
•core-site.xml为Hadoop的核心属性文件，参数为Hadoop的核心功能，独立于HDFS与MapReduce。
参数列表
•fs.default.name
•默认值 file:///
•说明：设置Hadoop namenode的hostname及port，预设是Standalone mode,如果是伪分布式文件系统要设置成 hdfs://localhost:9000，如果使用集群模式则配置为 hdfs://hostname:9000
•hadoop.tmp.dir
•默认值/tmp/hadoop-${user.name}
•会在tmp下根据username生成不同的目录

•fs.checkpoint.dir
•默认值${hadoop.tmp.dir}/dfs/namesecondary
•Sencondary NameNode 镜像存储目录
•fs.checkpoint.period
•默认值: 3600(秒)
• 控制 secondary namenode 的 checkpoint 时间间隔。如果距离上次 checkpoint 的时间大于这个参数的设定，就会触发 checkpoint。secondary namenode 會把 namenode 的 fsimage 和 editlog 做 snapshot。如果存取 Hadoop 的次数频繁或为了减少重起 namenode 的 downtime，可以把這个值设小一点。

•fs.checkpoint.size
•默认值67108864(byte)
•如果 Hadoop 非常的忙碌，editlog 可能会在短时间内变的很大，fs.checkpoint.period 的设定不见得可以完全预测这个状况，所以保险的做法会多设定这个值，以保证当数据大到超过 fs.checkpoint.size 的值也会触发 checkpoint。
•io.file.buffer.size
•默认值 4096
•這是读写 sequence file 的 buffer size, 可减少 I/O 次数。在大型的 Hadoop cluster，建议可设定为 65536 到 131072。

•ipc.client.connection.maxidletime
•默认值 10000(毫秒)
•设定 Hadoop client 连接時最大的闲置，默认是是 10 秒。如果 Hadoop cluster 的网络联系不稳，可以把这个值设到 60000(60秒)
•ipc.server.tcpnodelay
•默认值 false
•在 Hadoop server 是否启动 Nagle’s 算法。设 true 会 disable 这个演算法，关掉会减少延迟，但是会增加小数据包的传输。server site 不太需要这定这个值。

•hadoop.security.authorization
•默认值 false
•是不是要开启 账号验证机制，开启之后 Hadoop 在执行任何动作之前都会先确认是否有权限。详細的权限设定会放在 hadoop-policy.xml 裡。例如要让 fenriswolf 这个 account 及 mapreduce group 可以 submit M/R jobs，要设定 security.job.submission.protocol.acl
•hadoop.security.authentication
•默认值 simple
• simple 表示沒有 authentication，Hadoop 會用 system account 及 group 來控管q权限。另外可以指定為 kerberos，这部分相对比較复杂，要有一個 kerberos server 并产生 account keytab，在执行任何操作前 client 要先用 kinit 指令對 kerberos server 認證，之後的任何操作都是以 kerberos account 來執行。

•fs.trash.interval
•默认值 0（分）
•清掉垃圾筒的时间。预设是不清, 所以在删除文件时要自己执行
•hadoop.native.lib
•默认值 true
• 默认 Hadoop 会去找所有可用的 native libraries 并自动 load 进來使用，例如压缩类的 libraries 像 GZIP, LZO 等等。

第二部分：hdfs-site.xml
参数列表
•dfs.block.size
•默认值67108864（字节）
•默认每個 block 是 64MB。如果確定存取的文件块都很大可以改為 134217728(128MB)。Client 也可自行決定要使用的 block size 而不需要更改整個 cluster 的设定。
•dfs.safemode.threshold.pct
•默认值 0.999f
• Hadoop 启动时会进入 safe mode，也就是安全模式，這时是不能写入数据的。只有当99.9% 的 blocks 达到最小的 dfs.replication.min 数量(默认是3)才会离开safe mode。在 dfs.replication.min 设的比较大或 data nodes 数量比较多时会等比较久。

•dfs.namenode.handler.count
•默认值 10
•设定 namenode server threads 的数量，这些 threads 会用 RPC 跟其他的 datanodes 沟通。当 datanodes 数量太多时会发现很容易出现RPC timeout，解决方法是提升网络速度或提高这个值，但要注意的是 thread 数量多也表示 namenode 消耗的内存也随着增加
•dfs.datanode.handler.count
•默认值 3
• 指定 data node 上用的 thread 数量。

•dfs.datanode.max.xcievers
•默认值 256
•这个值是指定 datanode 可同時处理的最大文件数量、
•dfs.datanode.du.reserved
•默认值 0
•默认值表示 data nodes 会使用整个 磁盘，写满之后会导致无法再写入 M/R jobs。如果还有其他程式共用这些目录也会受到影响。建议保留至少 1073741824(1G) 的空间。

第三部分：mapred-site.xml
参数列表

•io.sort.mb
•默认值100
•缓存map中间结果的buffer大小(in MB)
•io.sort.record.percent
•默认值 0.05
•io.sort.mb中用来保存map output记录边界的百分比，其他缓存用来保存数据
•io.sort.spill.percent
•默认值0.80
•map开始做spill操作的阈值
•io.sort.factor
•默认值 10
•做merge操作时同时操作的stream数上限。

•min.num.spill.for.combine
•默认值3
•combiner函数运行的最小spill数
•mapred.compress.map.output
•默认值 false
•map中间结果是否采用压缩
•mapred.map.output.compression.codec
•org.apache.hadoop.io.compress.DefaultCodec
•min.num.spill.for.combine
•默认值3

•combiner函数运行的最小spill数
•mapred.compress.map.output
•默认值 false
•map中间结果是否采用压缩
•mapred.map.output.compression.codec
•org.apache.hadoop.io.compress.DefaultCodec
•mapred.reduce.parallel.copies
•默认值5
•每个reduce并行下载map结果的最大线程数
•mapred.reduce.copy.backoff
•默认值 300
•reduce下载线程最大等待时间（in sec）
•io.sort.factor
•默认值10
•org.apache.hadoop.io.compress.DefaultCodec
•mapred.job.shuffle.input.buffer.percent
•默认值0.7
•用来缓存shuffle数据的reduce task heap百分比
•mapred.job.shuffle.merge.percent
•默认值 0.66
•缓存的内存中多少百分比后开始做merge操作
•mapred.job.reduce.input.buffer.percent
•默认值0.0
•sort完成后reduce计算阶段用来缓存数据的百分比