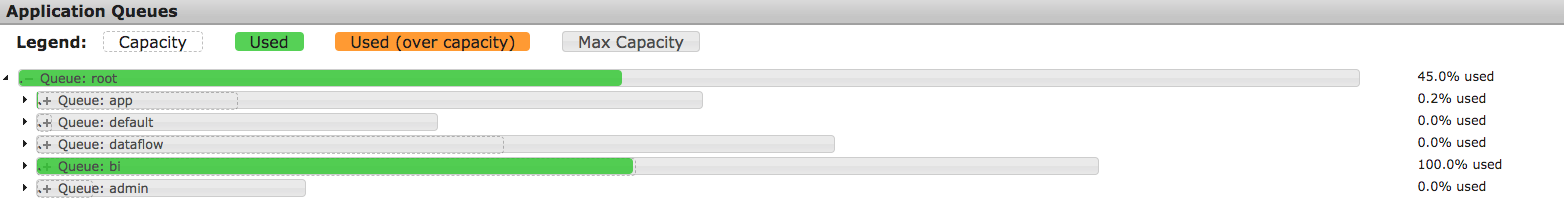
在做Tez性能测试时候，发现Tez并没有非常高效地使用相应yarn queue的资源，需要提高队列资源的利用率。

基于以上考虑，YARN允许用户配置每个节点上可用的物理内存资源，因为一个节点上的内存会被若干个服务共享，比如一部分给YARN，一部分给HDFS，一部分给HBase等。

Capacity Scheduler是YARN中默认的资源调度器，但是在默认情况下只有root.default 一个queue。目前，我们的root队列下又分成app、default、dataflow、bi和admin五个队列。



而当不同用户提交任务时，任务都会在这个队列里面按优先级先进先出，大大影响了多用户的资源使用率。

对于单个队列而言，只能有一个任务在运行，其他任务在等待。虽然设置了队列的资源容量capacity，但是由于资源是共享式的，所以如果有其他queue的资源处理空闲，那么该queue就会从空闲的queue借资源，但是该queue资源最多不会超过 maximum-capacity。如果有队列被其他队列借用是，该队队列接收到了新任务，dailyTask就会等待借出去的资源被回收再分配。

Yarn配置参数如下：

1. yarn.scheduler.capacity.root.capacity

root队列的资源容量（百分比）。 当系统非常繁忙时，应保证每个队列的容量得到满足，而如果每个队列应用程序较少，可将剩余资源共享给其他队列，所有队列的容量之和应小于100。目前设置为100.

1. yarn.scheduler.capacity.maximum-am-resource-percent

集群中用于运行应用程序ApplicationMaster的资源比例上限，该参数通常用于限制处于活动状态的应用程序数目。默认是0.1。目前设置0.2.

1. yarn.scheduler.capacity.root.app.minimum-user-limit-percent

app队列每个任务占用的最少资源数。默认100，目前设置4.

1. yarn.scheduler.capacity.root.default.user-limit-factor=1

default队列对单一用户的最大容量控制，默认1，目前设置1。即单个用户可以申请到的最大资源为min(系数\*capacity, maximum-capacity)

1. yarn.scheduler.capacity.root.default.maximum-capacity

root队列下的default队列最大可使用的资源率，目前设置为30

Tez配置参数如下：

1. tez.am.resource.memory.mb

application master container的内存大小。建议与yarn.scheduler.minimum-allocation-mb一致。目前设置14GB

1. hive.tez.container.size

hive在tez框架下的容器大小。建议设置为yarn.scheduler.minimum-allocation-mb的1～2倍，但不要超过yarn.scheduler.maximum-allocation-mb。目前设置14GB。

1. hive.tez.java.ops

默认设置80%\*hive.tez.container.size，目前设置“-server -Djava.net.preferIPv4Stack=true -XX:NewRatio=8 -XX:+UseNUMA -XX:+UseG1GC -XX:+ResizeTLAB -XX:+PrintGCDetails -verbose:gc -XX:+PrintGCTimeStamps -Xmx3276m”建议设置10GB

1. tez.runtime.io.sort.mb

输出排序需要的内存大小。建议设置40%\*hive.tez.container.size，一般不超过2G；默认值100，目前设置3276MB

1. hive.auto.convert.join.noconditionaltask.size

多个mapjoin转换为1个时，所有小表的文件大小总和的最大值，这个值只是限制输入的表文件的大小，并不代表实际mapjoin时hashtable的大小。 建议值：1/3\* hive.tez.container.size。目前设置1145044992B。

1. tez.runtime.unordered.output.buffer.size-mb

如果不直接写入磁盘，将使用缓冲区的大小。 建议设置10%\* hive.tez.container.size。默认100，目前设置245MB

1. hive.exec.parallel

在同一个sql中的不同的job是否可以同时运行,默认为false.建议true。

LLAP配置参数如下：

1. dfs.client.mmap.enable

是否尝试执行内存映射读取。目前设置为false，建议设置为true。

1. hive.llap.auto.allow.uber

是否允许规划程序在Application Master中运行。目前设置为false，建议设置为true。

