

Test results of YARN-4390

Wangda Tan

Test configurations

Using SLS mock a 100 nodes cluster, each node has 128G memory.

- SLS test framework is patched YARN-4778/YARN-4779/YARN-4899/YARN-4900 (bug fixes for SLS)

Queue configuration:

- Using CapacityScheduler
- 4 queues under root: a.capacity=5%, b.capacity=10%, c.capacity=25%, d.capacity=60%
- Maximum-capacity of all queues are 100
- User-limit-factor is set to 100 / queue's capacity. (So single job can take all resources of a cluster)
- Other configurations are using default value

Preemption configuration:

- yarn.resourcemanager.monitor.capacity.preemption.monitoring_interval=1000
- yarn.resourcemanager.monitor.capacity.preemption.max_wait_before_kill=30000
- yarn.resourcemanager.monitor.capacity.preemption.total_preemption_per_round=0.15
- yarn.resourcemanager.monitor.capacity.preemption.natural_termination_factor=1

Test Results

Case description

Submit 11 jobs to the cluster.

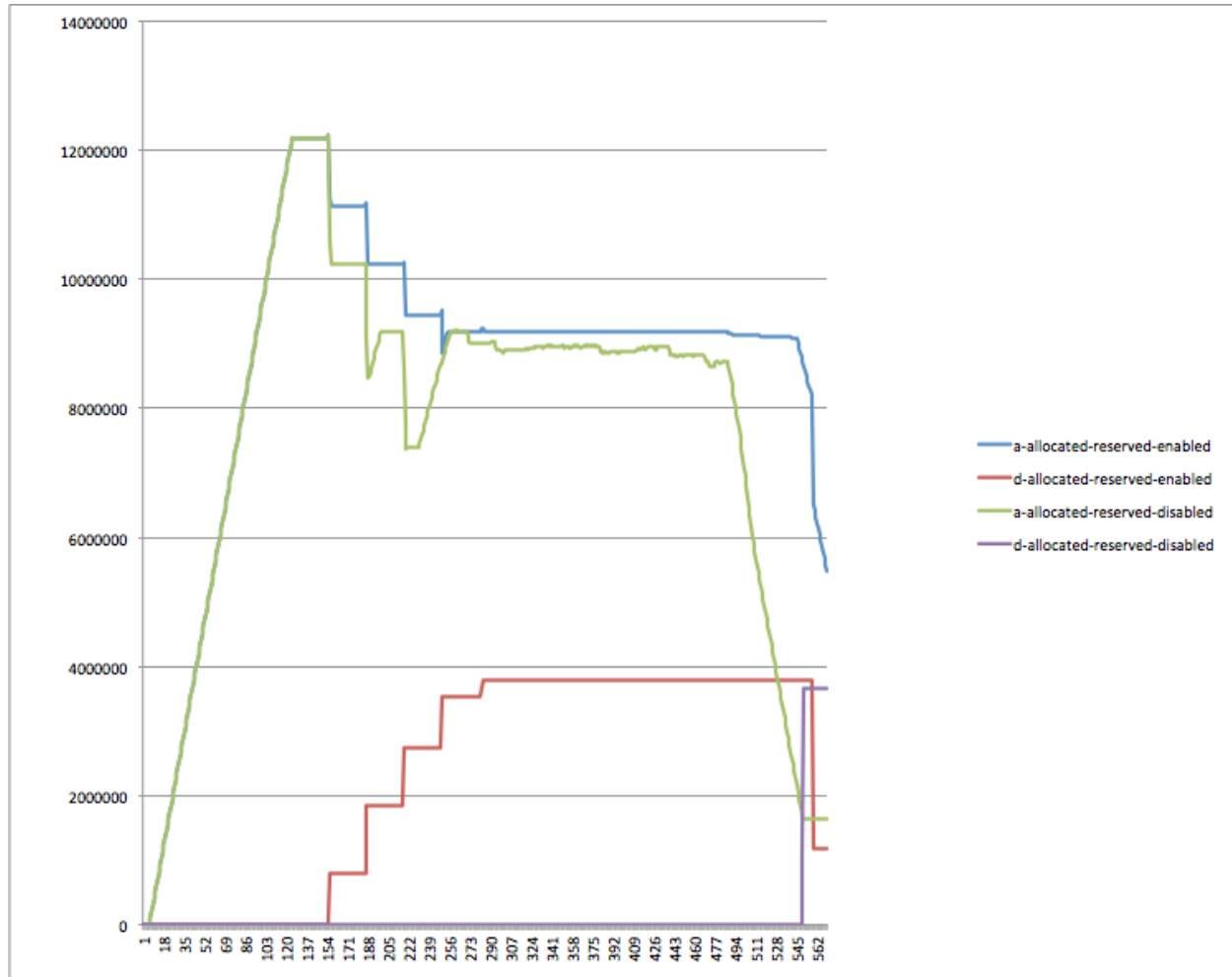
- Job#1 - Job#10 running at queue-A, runs many 1G containers, lifetime of each container is 8 mins.
- Job#11 running at queue-D, submitted to RM after job#1 runs for 1mins, asks for 10 * 128G containers, lifetime of each container is 8 mins.

Run SLS with

yarn.resourcemanager.monitor.capacity.preemption.select_candidates_for_reserved_containers = true/false

And compare results

Allocated resources (excluded reserved resources) of queues



X-axis is time (in sec)

Y-axis is allocated resource of memory (in MB)

Red/Purple lines are queue-D's allocated resource when reservation-preemption enabled/disabled:

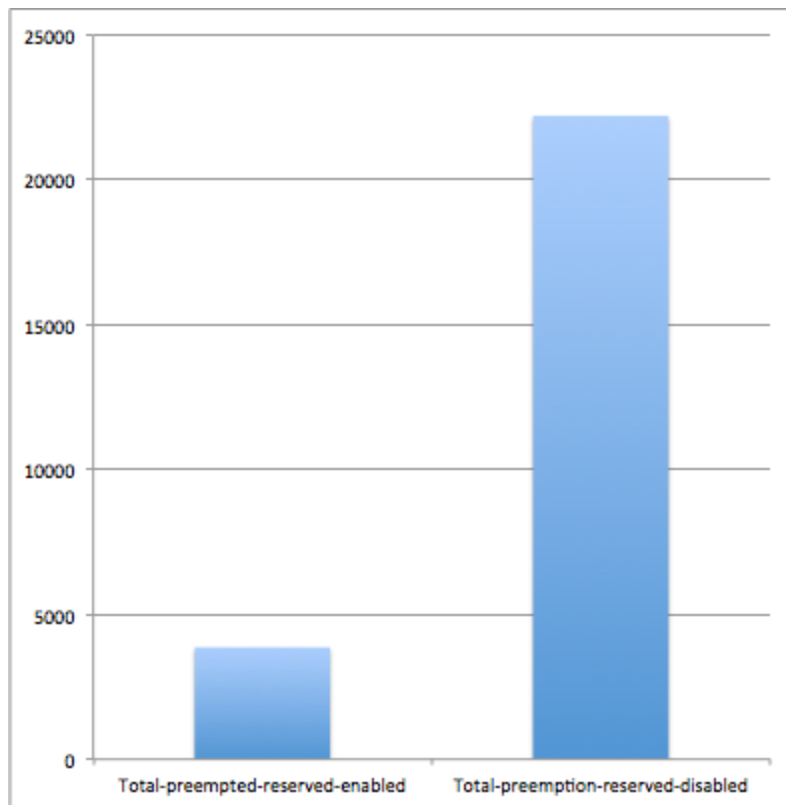
- When reservation preemption is enabled (Red), queue-D can preempt from queue-A in short time, and containers start to run.
- When reservation preemption is disabled (Purple), queue-D cannot precisely preempt correct resources from queue-A, so its allocated resource is 1G until queue-A releases most of its resources

Blue/Green lines are queue-A's allocated resources when reservation-preemption enabled/disabled:

- When reservation-preemption is disabled (green), queue-A's allocated resource is fluctuating. This is caused by excessive preemption happens. We can see this in next section.

Total number of preempted containers when reservation-preemption enabled/disabled

Total-preempted-reserved-enabled	Total-preemption-reserved-disabled
3865	22210



From the chart, when preemption for reserved container is disabled, it preempts almost 6X more containers than when preemption for reserved container is enabled.