Scaling policy

Terms

- A/I/O/T : num of allocated/idle/other/total nodes
- R: num of reserved nodes
- J: job in the queue
- P:job is running
- Min: the minimum number of nodes reserved

Assumption

- All jobs will take a reasonable LONG time
- Jobs won’t come and go very frequently(slow change)

Test 1 simple increase and decrease

- Sleep 5 10
- Sleep 5 15

Test 2 make a way(on the queue)

- T=21 M=10
- Sleep 5 20
- Sleep 5 15
- Sleep 10 25(pending resource)
- Sleep 6 10(pending priority)

- -> R=11

Test 2 make a way(on the queue)

- T=21 M=10
- Sleep 5 20
- Sleep 5 15(killed)
- Sleep 10 25(pending resource)

- R-=1 I=6
- Sleep 6 10(backfilled)

Test 3 do not make a way

- T=21 M=10
- Sleep 5 20
- Sleep 5 15
- Sleep 10 25(pending resource)
- Sleep 6 20(pending priority)
- ->R=11

Test 3 do not make a way

- T=21 M=10
- Sleep 5 20
- Sleep 5 15(killed)
- Sleep 10 25(pending resource)
- Sleep 6 20(pending priority)
- ->R=11+5