Peak load cause long queue

Last updated by | Veena Pachauri | Mar 8, 2023 at 11:10 PM PST

Issue: Custer see activity delay

Query:

Please run query below. Replace the activity id, and IR effective region.

IntegrationRuntime_DiagnoseCopyPoolCapacity('4ceef452-8b28-4171-b688-9eb8e8f76b01', 'westus')

If you see heath column pop up 'Unhealthy' That means batch pool has capacity issue.

TIMESTAMP	PoolName	health	IdleNodeRatio	TotalVM	IdleVm	RunningVm	CurrentCoreCount	So
2022-12-20 01:45:00.0000000	transferpoola1	health	0.742424242424242	66	49	17	66	Ho
2022-12-20 01:45:00.0000000	transferpoola1	health	0.742424242424242	66	49	17	66	Ho
2022-12-20 01:50:00.0000000	transferpoola1	health	0.696969696969697	66	46	20	66	Ho
2022-12-20 01:50:00.0000000	transferpoola1	health	0.696969696969697	66	46	20	66	Ho
2022-12-20 01:55:00.0000000	transferpoola1	health	0.863636363636364	66	57	9	66	Ho
2022-12-20 01:55:00.0000000	transferpoola1	health	0.863636363636364	66	57	9	66	Ho
2022-12-20 02:00:00.0000000	transferpoola1	health	0.742424242424242	66	49	17	66	Ho
2022-12-20 02:00:00.0000000	transferpoola1	health	0.742424242424242	66	49	17	66	Ho
2022-12-20 02:05:00.0000000	transferpoola1	health	0.106060606060606	66	7	59	66	Ho
2022-12-20 02:05:00.0000000	transferpoola1	health	0.106060606060606	66	7	59	66	Ho
2022-12-20 02:10:00.0000000	transferpoola1	health	0.730509932535304119	66	48	18	66	Ho

Root Cause:

Usually caused by adhoc peak load coming and our public IR compute pool no enough resource.

Suggestion:

If adhoc see this issue, please help to manager customer expectation that it might happen.

We regular update our scale rule depends on history workload. So we won't expect customer repeat hit delay issue at same time range. If customer repeat see this issue, you can let us know, we will continue improve scale system.