Scaling stuck due to management operations inside the same virtual cluster

Last updated by | Vitor Tomaz | Nov 16, 2022 at 12:59 PM PST

Contents

Issue

Customers complain that scaling their Managed Instance is taking too long.

Investigation/Analysis

Management operations on a managed instance can affect other management operations of the instances placed inside the same virtual cluster:

- Long-running restore operations in a virtual cluster will put other instance creation or scaling operations in the same subnet on hold. **Example**: If there is a long-running restore operation and there is a create or scale request in the same subnet, this request will take longer to complete as it waits for the restore operation to complete before it continues.
- A subsequent instance creation or scaling operation is put on hold by a previously initiated instance creation or instance scale that initiated a resize of the virtual cluster. **Example**: If there are multiple create and/or scale requests in the same subnet under the same virtual cluster, and one of them initiates a virtual cluster resize, all requests that were submitted 5+ minutes after the initial operation request will last longer than expected, as these requests will have to wait for the resize to complete before resuming.
- Create/scale operations submitted in a 5-minute window will be batched and executed in parallel. **Example**: Only one virtual cluster resize will be performed for all operations submitted in a 5-minute window (measuring from the moment of executing the first operation request). If another request is submitted more than 5 minutes after the first one is submitted, it will wait for the virtual cluster resize to complete before execution starts.

To investigate stuck management operations:

1. Extract all Managed Instances that belong in same Virtual cluster using below Kusto:

```
MonManagedServers
| where TIMESTAMP >= {startTime} and TIMESTAMP <= {endTime}
| where private_cluster_tenant_ring_name == {tenantringName} //Tenant Ring where MI resides - can be obtained
| summarize arg_max(TIMESTAMP,*) by managed_server_id
| where customer_subscription_id =~ {customerSubscriptionId}
| project TIMESTAMP, state, name, managed_server_id, private_cluster_tenant_ring_name, customer_subscription_i
```

2. Run ASC report for the timeframe to check for ongoing long running restore requests on the MI in the same Virtual cluster that are/were running at the time of scaling. Alternatively, below kusto can also help identify if there are/were ongoing restore operations:

```
MonRestoreRequests
| where subscription_id =~ {customerSubscriptionId}
| where target_managed_server_name in ({MINames}) //Managed Instances that belong in the same virtual cluster
| project restore_request_id, target_managed_server_name, restore_start_time = start_time, restore_end_time =
```

Mitigation

This behavior is by design. Management operations that are put on hold because of another operation that is in progress will automatically be resumed once conditions to proceed are met. No user action is necessary to resume the temporarily paused management operations.

In such scenarios, customer can either wait until the restore operation is completed or opt to cancel the restore to resume the scaling operation. In the event that restore appears to be long running or stuck, investigate long running restores using Backup and Restore TSGs.

Public Doc Reference

Management operations cross-impact

Internal Reference

ICM 214949237

ICM 266979036

How good have you found this content?