

Self hosted IR about disaster recovery

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

Contents

- [Summary](#)
- [Info](#)
- [More detail](#)

Summary

Customer is using Self hosted IR in their Azure data factory and would like to understand the disaster recovery for it. They have created backup VMs for SHIR VMs, once they goes down then it will switch to backup VMs. So if the VMs on which SHIR is hosted from East US region goes down and it gets Switched to backup VMs then what action needed to be taken care by customer for DR? Customer would like to know the disaster recovery for SHIR in above scenario.

Info

SHIR needs to talk with ADF service and all SHIR metadata are persisted in one region. So there are two levels of disaster recovery here.

1. ADF service outage: ADF will take action, like failover to backup region to guarantee its service SLA. This is not customer facing and awareness.
2. SHIR client outage (the VMs hosting SHIR are outage): Customer should setup SHIR high availability with some VMs in primary region and other VMs in backup region. Once primary region VMs are outage, resulting in those nodes are offline, as long as backup region VMs are online (automatically spin up), SHIR can automatically use these backup nodes to execute tasks. Pipeline in running state on the failed nodes will be reset back to queue and repicked up by the backup nodes.

More detail

1. Customer should register SHIR on backup VMs and make sure it is in the same SHIR HA cluster as primary VMs because linked service can be only connected via one IR.
2. If primary VMs crash, running pipelines are reset back to queue. Once backup VMs is automatically spin up/manually started by customer (so the backup nodes can be online), the backup nodes will repick up these pipelines to execute. It is automatic and customer doesn't need to re-run the pipeline. We guarantee no task is lost in SHIR failover.
3. One SHIR HA cluster can only have 4 nodes at most. Reserving 1 or 2 nodes for disaster recovery might be a waste. Keeping backup VMs always online can increase the SHIR capacity (it can also execute pipeline but there will be VM cost and cross region billing/performance concern). This is all based on customer's business requirement.