

Veeam Backup for Azure Calculations

Veeam Worker Instances

Let's assume we're running a number of W_{veeam} worker instances, each providing a maximum throughput of ψ_{worker} MB/s, so the total throughput Ψ_{total} can be calculated as

$$\Psi_{total} = W_{veeam} \cdot \psi_{worker} \quad (1)$$

To process a given volume V of incremental backup data (in TB), these workers (all running in parallel) require a duration of

$$T_{parallel} = \frac{V \cdot 1024^2}{\Psi_{total}} = \frac{V \cdot 1024^2}{W_{veeam} \cdot \psi_{worker}} \quad \text{seconds, or}$$
$$T_{parallel} = \frac{V \cdot 1024^2}{W_{veeam} \cdot \psi_{worker} \cdot 3600} \quad \text{hours.} \quad (2)$$

Per default, Veeam Backup for Azure would start a dedicated worker instance for each VM to be backed up, i.e. W_{veeam} would be equal to the total number N_{vm} of VMs. But as we are limiting the number of worker instances to a value $\leq N_{vm}$, we have to consider running the chosen amount of workers multiple times in sequence, as each worker instance can process only a single workload at a time. The required amount $\rho_{workers}$ of these sequential "worker runs" is determined by

$$\rho_{workers} = \frac{N_{vm}}{W_{veeam}} \quad (3)$$

Hence, we need to multiply the duration $T_{parallel}$ by the number $\rho_{workers}$ of runs to obtain the duration T_{seq} (in hours) that is required to process our set of N_{vm} VMs (with $N_{vm} \geq W_{veeam}$):

$$T_{seq} = T_{parallel} \cdot \rho_{workers} \quad (4)$$

To be clear: T_{seq} defines the backup window T_{backup} required to process the given workload, i.e.

$$T_{backup} = T_{seq} = T_{parallel} \cdot \rho_{workers}$$

By substituting $T_{parallel}$ with (2) and $\rho_{workers}$ with (3), we get

$$\Leftrightarrow T_{backup} = \frac{V_{incr} \cdot 1024^2}{W_{veeam} \cdot \psi_{worker} \cdot 3600} \cdot \frac{N_{vm}}{W_{veeam}}$$

Simplified

$$\Leftrightarrow T_{backup} = \frac{1024^2}{3600} \cdot \frac{V_{incr} \cdot N_{vm}}{(W_{veeam})^2 \cdot \psi_{worker}} \quad (5)$$

By solving this equation for W_{veeam} , we are now able to calculate the required number of workers W_{veeam} as a function of their throughput ψ_{worker} (in MB/s), the number N_{vm} and (incremental) backup volume V_{incr} (in TB) of protected VMs as well as the limiting backup window T_{backup} (in hours):

$$\Leftrightarrow W_{veeam} = \sqrt{\frac{1024^2}{3600} \cdot \frac{V_{incr} \cdot N_{vm}}{T_{backup} \cdot \psi_{worker}}}$$

Simplified

$$\Leftrightarrow W_{veeam} = \frac{256}{15} \cdot \sqrt{\frac{V_{incr} \cdot N_{vm}}{T_{backup} \cdot \psi_{worker}}} \quad (6)$$

Storage Accounts

Ingress Limit

We must not forget that our target storage accounts have an ingress limit ψ_{stacc} (in MB/s). We have to divide the maximum throughput provided by the workers running in parallel by this limit to determine the minimum number of target storage accounts capable of ingesting the data at the required speed:

$$N_{stacc} \geq \frac{\Psi_{total}}{\psi_{stacc}}$$

Substituting Ψ_{total} with (1) results in

$$\Leftrightarrow N_{stacc} \geq \frac{W_{veeam} \cdot \psi_{worker}}{\psi_{stacc}} \quad (7)$$

By using (6) to substitute W_{veeam} in (7), the number of required storage accounts can be calculated as a function of well known input parameters N_{vm} , V_{incr} [TB], ψ_{worker} [MB/s], ψ_{stacc} [MB/s], and T_{backup} [hours]:

$$N_{stacc} \geq \frac{256}{15} \cdot \frac{\psi_{worker}}{\psi_{stacc}} \cdot \sqrt{\frac{V_{incr} \cdot N_{vm}}{T_{backup} \cdot \psi_{worker}}}$$

Simplified

$$\Leftrightarrow N_{stacc} \geq \frac{256}{15} \cdot \frac{1}{\psi_{stacc}} \cdot \sqrt{\frac{V_{incr} \cdot N_{vm} \cdot \psi_{worker}}{T_{backup}}} \quad (8)$$

IOPS Limit

There is another storage account limitation that needs to be considered: Each storage account has a maximum "request rate" of $r_{stacc/iops}$ (= 20.000 IOPS) as described in the [User Guide](#). The request rate $r_{veeam/iops}$ created by Veeam Backup for Azure during incremental backup processing can be calculated based data volume V_{incr} (total amount of data written during incremental backup in TB), Veeam's block size value β_{veeam} (= 1024 kB at source) and the backup time T_{backup} (in hours) as

$$r_{veeam/iops} = \frac{1024^3}{3600} \cdot \frac{V_{incr}}{\beta_{veeam} \cdot T_{backup}}$$

This request rate has to be lower than or equal to $r_{stacc/iops}$ multiplied with the number of storage accounts N_{stacc} we're targeting in parallel

$$r_{veeam/iops} \leq N_{stacc} \cdot r_{stacc/iops}$$

This leads to

$$r_{stacc/iops} \geq \frac{1024^3}{3600} \cdot \frac{V_{incr}}{N_{stacc} \cdot \beta_{veeam} \cdot T_{backup}}$$

Rearranged

$$\Leftrightarrow N_{stacc} \geq \frac{1024^3}{3600} \cdot \frac{V_{incr}}{r_{stacc/iops} \cdot \beta_{veeam} \cdot T_{backup}} \quad (9)$$

Combining Ingress and IOPS Limit

While (8) provides a result based on storage account ingress speed limit (in MB/s), equation (9) is based on the storage accounts' ingress rate limit (in IOPS). For a *real world* calculation, we have to use both formulas and pick the higher result, because both limits must not be exceeded. We can write this down as

$$N_{stacc} = \mathbf{max\ of} \begin{cases} \frac{256}{15} \cdot \frac{1}{\psi_{stacc}} \cdot \sqrt{\frac{V_{incr} \cdot N_{vm} \cdot \psi_{worker}}{T_{backup}}} & \text{(based on ingress limit } \psi_{stacc} \text{)} \\ \frac{1024^3}{3600} \cdot \frac{V_{incr}}{r_{stacc/iops} \cdot \beta_{veeam} \cdot T_{backup}} & \text{(based on request rate limit } r_{stacc/iops} \text{)} \end{cases} \quad (10)$$

Policies

A single policy within Veeam Backup for Azure can only target a single storage account (as a primary backup target). So, the number of storage accounts N_{stacc} we just derived also defines the *minimum* number N_{pol} of required policies, in other words

$$N_{pol} \geq N_{stacc}$$

Assuming we do not want to create more policies than required (i.e., we're maxing out the policies as much as possible), the number N_{pol} of policies is given by

$$N_{pol} = N_{stacc} \quad (11)$$

In an ideal world, we can also assume an equal distribution of all N_{vm} VMs across all existing policies N_{pol} , with each policy containing the same amount $N_{vm/pol}$ of VMs, defined as

$$N_{vm/pol} = \frac{N_{vm}}{N_{pol}}$$

Substituting N_{pol} with (11) results in

$$\Leftrightarrow N_{vm/pol} = \frac{N_{vm}}{N_{stacc}} \quad (12)$$

It's as simple as that and it's all we need because N_{stacc} is already known from (10) and N_{vm} should be a well known input parameter.

Appliance Requirements

According to the sizing and scalability guidelines provided in the [User Guide](#), the amount M_{pol} of RAM required by running a number of N_{pol} policies (each processing $N_{vm/pol}$ VMs) in parallel on a single Veeam Backup for Azure appliance can be calculated as

$$M_{pol} = N_{pol} \cdot (10 \text{ MB} + (N_{vm/pol} \cdot 1 \text{ MB})) \quad (13)$$

Additionally, some *overhead* RAM $M_{appl/oh}$ required by the appliance OS and the WebUI and REST API service needs to be taken into account. This value depends on the amount $M_{appl/total}$ of RAM provisioned to the appliance VM.

$$M_{appl/oh} = 1.5 \text{ GB} + \frac{M_{appl/total}}{20}$$

The available RAM $M_{appl/avail}$ for processing policies is therefore given as

$$\begin{aligned} M_{appl/avail} &= M_{appl/total} - M_{appl/oh} \\ \Leftrightarrow M_{appl/avail} &= M_{appl/total} - \left(1.5 \text{ GB} + \frac{M_{appl/total}}{20}\right) \\ \Leftrightarrow M_{appl/avail} &= M_{appl/total} - 1.5 \text{ GB} - \frac{M_{appl/total}}{20} \\ \Leftrightarrow M_{appl/avail} &= \left(\frac{19}{20} M_{appl/total}\right) - 1.5 \text{ GB} \end{aligned} \quad (14)$$

Rearranging (14) to get $M_{appl/total}$

$$\begin{aligned} \Leftrightarrow M_{appl/total} &= \frac{20}{19} \cdot (1.5 \text{ GB} + M_{app/avail}) \\ \Leftrightarrow M_{appl/total} &= \frac{30}{19} \text{ GB} + \frac{20}{19} M_{app/avail} \end{aligned} \quad (15)$$

To enable processing of N_{pol} policies with each working on $N_{vm/pol}$ VMs, the appliance needs an available amount of RAM given as

$$M_{appl/avail} \geq M_{pol}$$

The absolute minimum requirement being

$$M_{appl/avail} = M_{pol}$$

Inserting this identity into (15)

$$\Leftrightarrow M_{appl/total} = \frac{30}{19} \text{ GB} + \frac{20}{19} M_{pol}$$

We can substitute M_{pol} with the knowledge of (13)

$$\Leftrightarrow M_{appl/total} = \frac{30}{19} \text{ GB} + \left(\frac{20}{19} N_{pol} \cdot (10 \text{ MB} + (N_{vm/pol} \cdot 1 \text{ MB})) \right) \quad (16)$$

This provides us with a *theroretical* value of the amount of required appliance RAM based on our assumptions and input parameters. The [User Guide](#) recommends allocating an additional margin of 20% of RAM for production environments. In addition, I would recommend also increasing the number of policies N_{pol} used in (16), since our assumptions from above (all policies are maximally utilized, and that all VMs are equally distributed across all policies) are unlikely to be achieved in reality. To be on the safe side, I'd start with doubling the value of N_{pol} calculated by (11) while keeping (not reducing) the value of VMs per policy $N_{vm/pol}$.

Putting these two additional thoughts (doubling N_{pol} and adding 20% overall) into (16) results in

$$M_{appl/total} = \frac{36}{19} \text{ GB} + \left(\frac{48}{19} N_{pol} \cdot (10 \text{ MB} + (N_{vm/pol} \cdot 1 \text{ MB})) \right) \quad (17)$$