

Veeam Backup for Azure Calculations

Veeam Worker Instances

Calculating the required number of worker instances W_{veeam} as a function of a predefined backup window T_{backup} :

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 theoretically 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

$$\Leftrightarrow W_{veeam} = \frac{1024}{60} \cdot \sqrt{\frac{V_{incr} \cdot N_{vm}}{T_{backup} \cdot \psi_{worker}}} \quad (6)$$

Storage Accounts

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 required number of target storage accounts capable of ingesting the data:

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

Substituting Ψ_{total} with (1) results in

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

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} , ψ_{worker} , ψ_{stacc} , and T_{backup} :

$$N_{stacc} = \frac{1024}{60} \cdot \frac{\psi_{worker}}{\psi_{stacc}} \cdot \sqrt{\frac{V_{incr} \cdot N_{vm}}{T_{backup} \cdot \psi_{worker}}}$$

Simplified

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

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

$$N_{pol} = N_{stacc} \quad (9)$$

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 (9) results in

$$\Leftrightarrow N_{vm/pol} = \frac{N_{vm}}{N_{stacc}} \quad (10)$$

We can replace N_{stacc} with (8)

$$\Leftrightarrow N_{vm/pol} = N_{vm} \div \left(\frac{256}{15} \cdot \frac{1}{\psi_{stacc}} \cdot \sqrt{\frac{V_{incr} \cdot N_{vm} \cdot \psi_{worker}}{T_{backup}}} \right)$$

Simplified

$$\begin{aligned}
\Leftrightarrow N_{vm/pol} &= \frac{15}{256} \cdot N_{vm} \cdot \psi_{stacc} \cdot \sqrt{\frac{T_{backup}}{V_{incr} \cdot N_{vm} \cdot \psi_{worker}}} \\
\Leftrightarrow N_{vm/pol} &= \frac{15}{256} \cdot \psi_{stacc} \cdot \sqrt{\frac{N_{vm} \cdot T_{backup}}{V_{incr} \cdot \psi_{worker}}}
\end{aligned} \tag{11}$$

Well, this formula looks nice but it doesn't add any real value as (10) already provides us with the desired result for $N_{vm/pol}$. It's also much easier to calculate using (10) because we've already gained N_{stacc} from our previous calculations.

Important note: Depending on input values, these formulas might give you a higher number of VMs per policy than there are VMs in total. This is no mistake, it just tells us that we don't need more than just **one** policy to cover all source VMs. It becomes more clear when realizing that all the calculated values that represent a number of "instances" of a certain "object" (like VMs, storage accounts, policies, etc.) need to be **rounded up to the next greater integer**, as there is no such thing as "2.4 storage accounts" - if the result is 2.4, you need "3 storage accounts".

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 (with each policy processing $N_{vm/pol}$ VMs) in parallel on a single Veeam Backup for Azure appliance can be calculated as

$$M_{pol} = N_{pol} \cdot (10 \text{ MB} + (N_{vm/pol} \cdot 1 \text{ MB})) \tag{12}$$

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

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

This means that the available RAM $M_{appl/avail}$ for processing policies is given as

$$\begin{aligned}
M_{appl/avail} &= M_{appl/total} - M_{appl/oh} \\
\Leftrightarrow M_{appl/avail} &= M_{appl/total} - (1.5 \text{ GB} + \frac{M_{appl/total}}{20}) \\
\Leftrightarrow M_{appl/avail} &= M_{appl/total} - 1.5 \text{ GB} - \frac{M_{appl/total}}{20} \\
\Leftrightarrow M_{appl/avail} &= \left(\frac{19}{20} M_{appl/total}\right) - 1.5 \text{ GB}
\end{aligned} \tag{13}$$

Rearranging (13) to get $M_{appl/total}$

$$\begin{aligned}
\Leftrightarrow M_{appl/total} &= \frac{20}{19} \cdot (1.5 \text{ GB} + M_{appl/avail}) \\
\Leftrightarrow M_{appl/total} &= \frac{30}{19} \text{ GB} + \frac{20}{19} M_{appl/avail}
\end{aligned} \tag{14}$$

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

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

So, the absolute minimum available RAM requirement would be

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

Inserting this identity into (14)

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

We can substitute M_{pol} with the knowledge of (12)

$$\Leftrightarrow M_{appl/total} = \frac{30}{19} \text{ GB} + \left(\frac{20}{19} N_{pol} \cdot \left(10 \text{ MB} + (N_{vm/pol} \cdot 1 \text{ MB}) \right) \right) \quad (15)$$

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 (15), since our assumption that all policies are maximally utilized (leading to equation (9) above) is unlikely to be achieved in reality. To be on the safe side, let's simply double the value of N_{pol} for this calculation.

Putting these two additional constraints (doubling N_{pol} and adding 20% overall) into (15) results in

$$M_{appl/total} = \frac{36}{19} \text{ GB} + \left(\frac{48}{19} N_{pol} \cdot \left(10 \text{ MB} + (N_{vm/pol} \cdot 1 \text{ MB}) \right) \right) \quad (16)$$

Main Results

The main results are in formulas no. 6, 8, 10 and 15/16.

Worker Instances

The amount of worker instances required to process a given amount of VMs and their incremental backup volume in a given backup window:

$$W_{veeam} = \frac{1024}{60} \cdot \sqrt{\frac{V_{incr} \cdot N_{vm}}{T_{backup} \cdot \psi_{worker}}} \quad (6)$$

Storage Accounts

The minimum amount of storage accounts required to ingest these incremental backups within the backup window:

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

Number of Policies

The minimum number of policies required to process these workloads is equal to the number of storage accounts (8), while each policy should not exceed the maximum amount of VMs per policy:

$$N_{vm/pol} = \frac{N_{vm}}{N_{stacc}} \quad (10)$$

or

$$N_{vm/pol} = \frac{15}{256} \cdot \psi_{stacc} \cdot \sqrt{\frac{N_{vm} \cdot T_{backup}}{V_{incr} \cdot \psi_{worker}}} \quad (11)$$

Appliance RAM Requirements

Theoretical value

$$M_{appl/total} = \frac{30}{19} \text{ GB} + \left(\frac{20}{19} N_{pol} \cdot \left(10 \text{ MB} + (N_{vm/pol} \cdot 1 \text{ MB}) \right) \right) \quad (15)$$

"Real life" value

$$M_{appl/total} = \frac{36}{19} \text{ GB} + \left(\frac{48}{19} N_{pol} \cdot \left(10 \text{ MB} + (N_{vm/pol} \cdot 1 \text{ MB}) \right) \right) \quad (16)$$

Thanks for reading this far! :-)