Investigation report on Veeam for Manged Backup of Private Cloud

Revision history

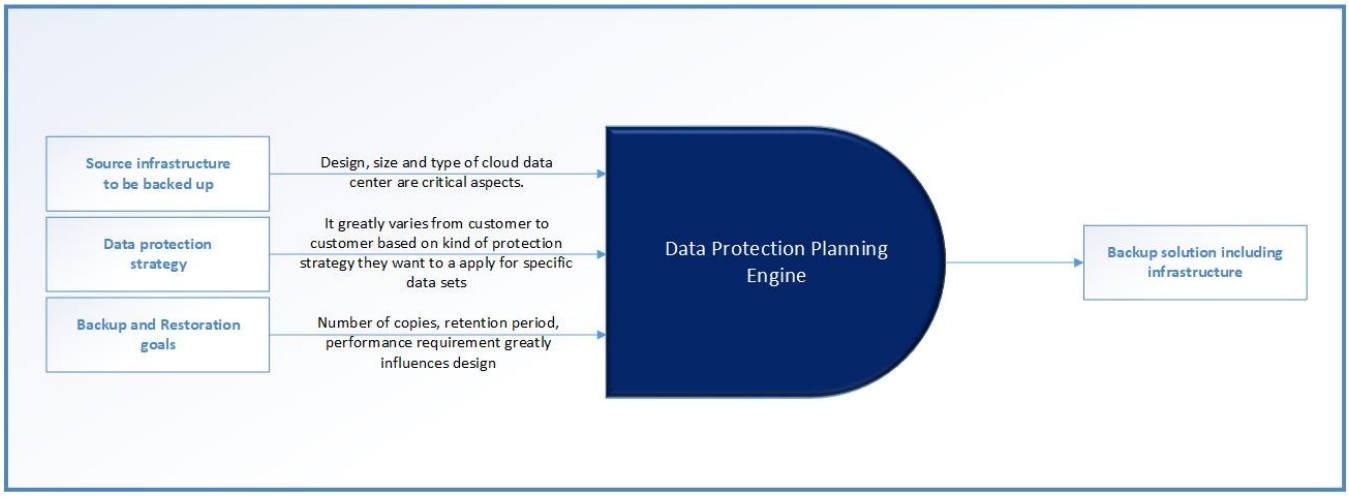
Revision	Author	Date	Description
0.1	Jyoti Ranjan	09-August-2019	Conceptualized the structure of document.
0.5	Jyoti Ranjan	09-August-2019	Copied the content which was lying in my laptop but was not able to maintain it because of JIRA not being created
0.6	Jyoti Ranjan	09-August-2019	Updating integration aspects of Agena's with Veeam.
0.75	Jyoti Ranjan	09-August-2019	Added more diagrams.
0.85	Jyoti Ranjan	09-August-2019	Added presentation for US421
1.0	Jyoti Ranjan	12-August-2019	Re-factored the diagram to improve visibility. Also, detailed the deployment steps.
1.05	Jyoti Ranjan	27-August-2019	Re-factoring document to be more specific to Agena as asked by AGE-275 - Getting issue details... STATUS (new story)
1.1	Jyoti Ranjan	29-August-2019	Drafted diagram to illustrate workflows instead of text for specific use cases. Also, added minute access details as asked in story AGE-275 - Getting issue details... STATUS (new story)

Introduction

Agena VMaaS cloud platform aims to provide Infrastructure as a Service (IaaS) which has all benefits of cloud consumption with embedded HPE value add. It provides end user to provision virtual infrastructure like a cloud in a way which has tenets like robustness, agility, pay-as-you-drink, scalable on demand. For every cloud data center, the user workload is utmost important and so is ensuring its availability in case of data corruption or site failure. It is important to provide security of data in motion or at rest. To achieve this, it requires comprehensive understanding and design thinking on data protection mechanism before concluding any backup, DR or data management solution. The backup, restore and replication is only one aspect of data protection system. In this document, we are not going to primarily focus on backup aspects and hence the reader is expected to ingest the content keeping this perspective unless it is stated explicitly.

The backup solution is highly influenced by customer's source infrastructure and backup requirements to be achieved. The following points greatly influence the design of Backup solution which is pictorially depicted as well:

1. Data protection strategy.
2. Capacity, consumption and design of source infrastructure to be backed up.
3. Backup and restoration goals



Considering the heterogeneity and diversity of data protection system as well as the need to focus on backup solution, it is very important to define what is covered and what is not covered as explicitly. See for details on it, The focus of this document is not to define complete data protection solution but to define a backup and restoration solution which can be used at the time of fail-over or disaster recovery.

Scope

The data protection strategy is bigger story than backup, replication and restore. In crude terms, the focus of backup is just to keep a offline cope of data which can be used at the time of data corruption or failure for restoration purposes. Many companies (like Veeam, Commvault) are really focusing on building products that offer “Resiliency” and “Business Continuity” which include both traditional backup and disaster recovery features. Also, many products may carry the same features but some will do it better than others. On the other hand, any data protection system is very much tied to customer requirement as well as source infrastructure, it is not possible to cover all permutation and combination of backup solution architecture without specific details with great degree of accuracy. So, it is very important to specifically detailed out what is covered and what is not in this document.

What is not in scope?

- No complete data protection strategy.
- No DR plan.

What is in scope?

- Understand backup and restoration solution provided by Veeam
- Defining backup and restoration solution for Agena cloud using Veeam
- As there is NO customer specific details are available, some assumptions have been made on various aspects. Those assumption greatly influence the architecture, design and configuration of backup and restoration solution. See below.
- Focus on following tenets (implicitly if not stated explicitly):
    - Robustness of backup infrastructure to avoid losing data in case of failure

- Scalable along with source infrastructure
- Adherence to 3-2-1 rule (if applicable)
- Minimal impact on user workload during backup and restore operation.
- Low RPO and RTO time
- Designed to strive for low cost

Assumption

Source infrastructure	• Hypervisor = ESXi • Maximum number of ESXi hosts = 24 • Maximum number of VMs = 2000 • Storage media used for VM = Nimble • Maximum amount of raw storage = 100 TB
Backup infrastructure	• Use Veeam product suite for backup solution • Backup storage media = Nimble

Design aspect of 'Data Protection System' for user workload(s)

The steps listed are based on author's understanding of data management and protection. The author has tried to present information as succinct as possible. It is implicitly assumed that below steps might need refinement for specific customer based on their data protection strategy for different types of data sets.

- List source infrastructure details
  - Type of hypervisor
  - Composition (number of hosts)
  - Number of VMs
  - Storage media used for VMs
  - etc
- Define data protection strategy (source: Tom Petrocelli)
  - **Backup and recovery.** Goal is to safeguard data by taking backup of data to be used in case of data corruption or failure.
  - **Remote data movement.** The real-time or near-real-time moving of data to a location outside the primary storage system or to another facility to protect against physical damage to systems and buildings.
  - **Storage system security.** Security data in rest as well as in-motion.
  - **Data Lifecycle Management (DLM).** Tiring and accessibility of data based on its age.
  - **Information Lifecycle Management (ILM).** A comprehensive strategy for valuing, cataloging and protecting information assets. It is tied to regulatory compliance as well.
- Define backup and restoration goals
  - How much data loss can the business afford (RPO – Recovery Point Objective)? How many backups we do want to maintain?
  - How quickly do the applications need to be up and running (RTO- Recovery Time Objective)?
  - How long does the data need to be retained?
  - How do we want to simplify backup and restoration mechanism? For e.g. do we need 1-click restoration?
- Understand specific backup solution and its fitment to source infrastructure
  - Feature
  - PoC
  - Licensing strategy
  - Direction in which product is headed
- Customize above steps for specific customer requirement (optional)

Evaluation of Veeam

Veeam is Backup, Replication and Recovery software. Veeam does not install any agent on VM being backed up.

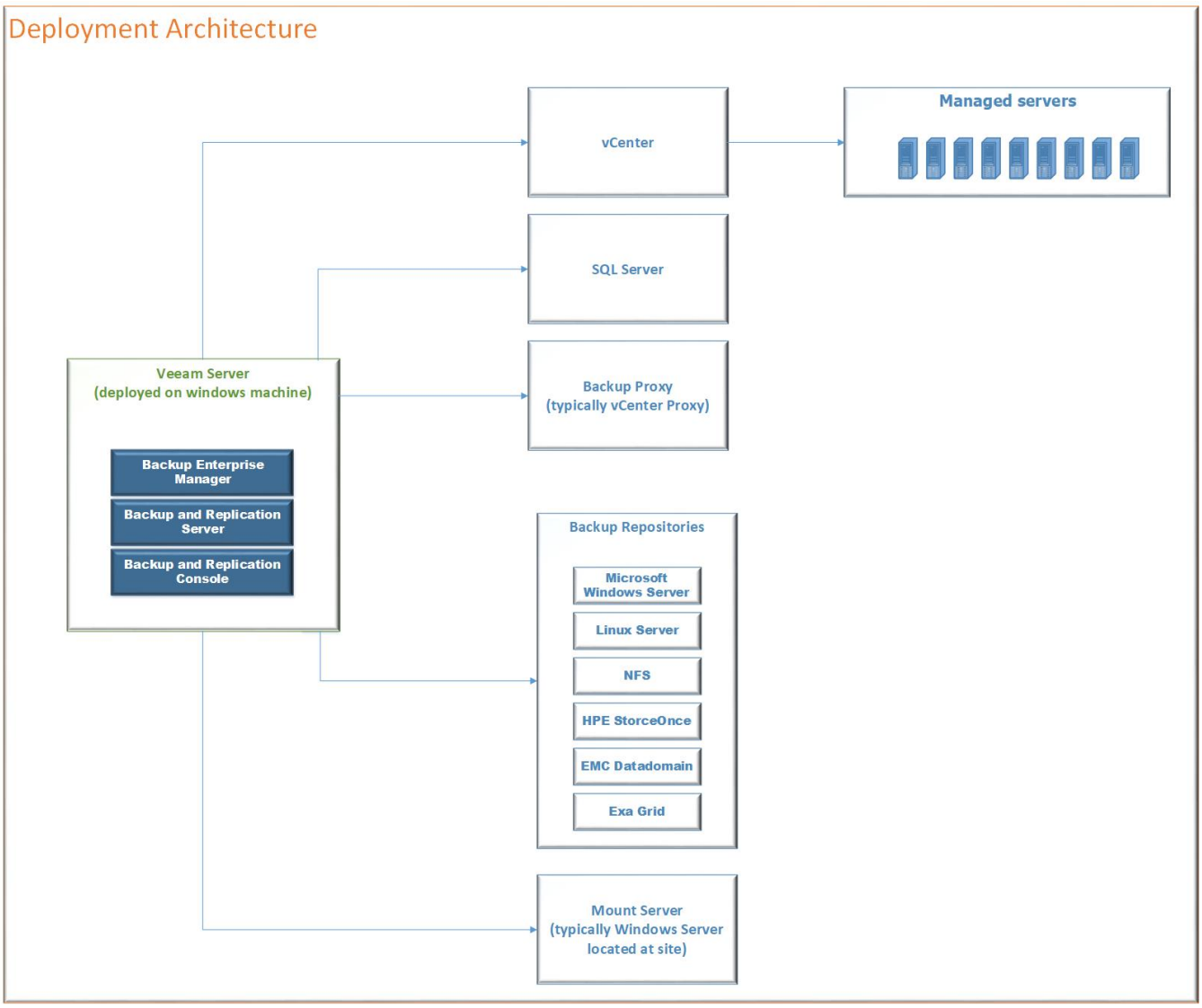
Veeam Backup & Replication performs backups at the image-level using APIs available from the underlying hypervisor. It has no direct visibility of the file structure after backup is finished. It is possible to Use File Level Recovery (FLR) wizard or Enterprise Manager to mount VMs from within a backup file and access/restore VM guest files. It provides self-service portal which can be used to configure backup jobs and remotely manage it. It is highly recommend to design Backup Infrastructure keeping in mind the number of source infrastructure being backed up so that there is no scarcity of resources affecting running VM workloads at the time of backup or restore is being carried out.

Use cases

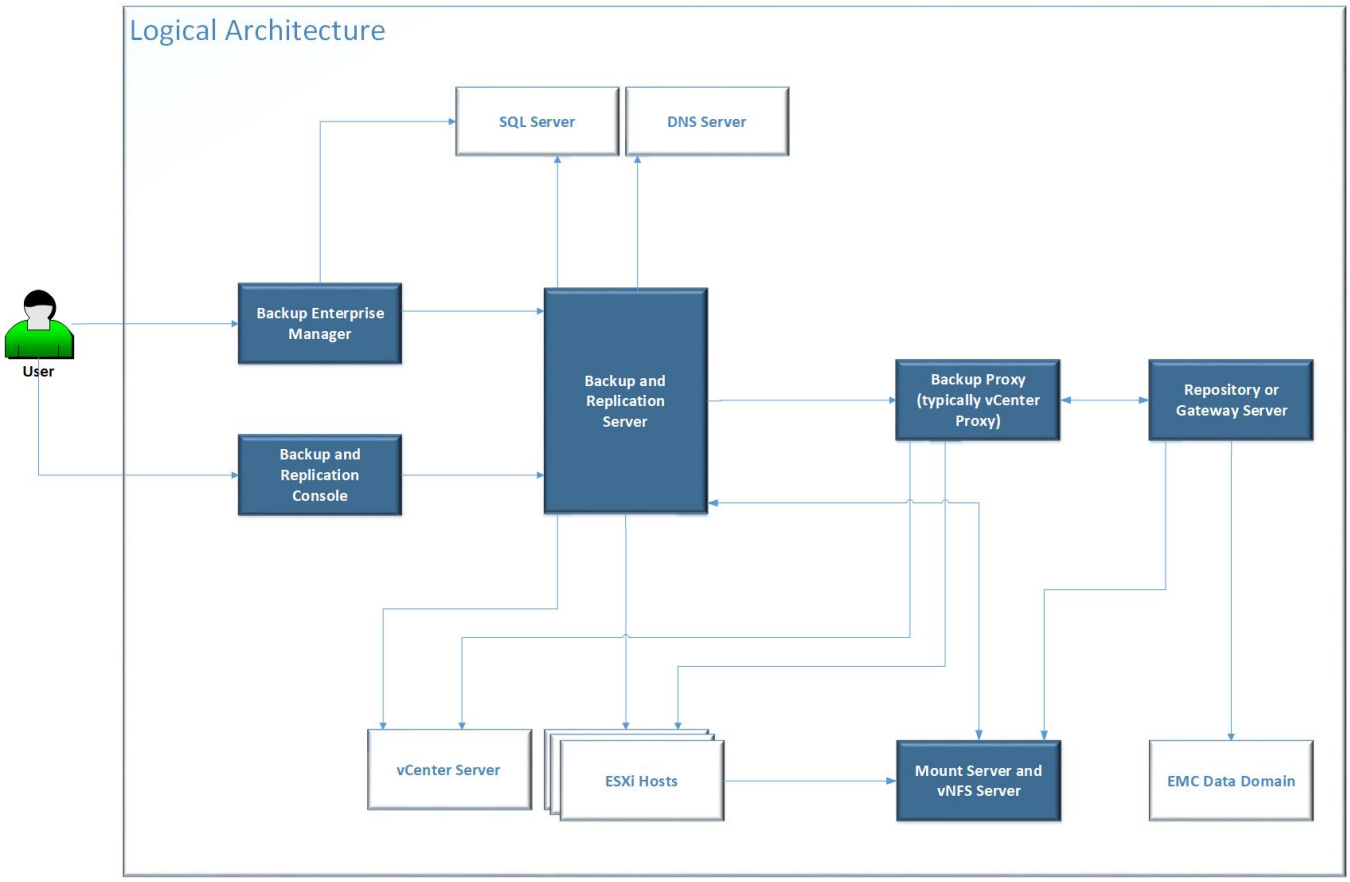
Veeam is a comprehensive backup and restore suite which allows integration with various third party component. The following use cases can be addressed.

1. VM use cases
    - Recover entire VM to the Original or Different host
    - Quickly restore to user by starting a VM directly from a backup file on regular backup storage
    - Recover individual VM files (such as VMX) and virtual disks (vhd, vhdx or vmdk)
    - Full VM recovery
    - Instant VM recovery
    - VM file and virtual disk recovery
    - Restore to cloud
2. File level recovery
    - Recovery files from 19 common file systems used by Windows, Linux, Unix, MacOS, Novell, Solaris
3. Application aware backup (for specific enterprise application)
    - Microsoft Active Directory
    - Microsoft Exchange
    - Microsoft SQL
    - Oracle
    - Microsoft SharePoint
4. Replication
    - Replicate image
    - Use case: high availability or off-site for disaster recovery
    - Move production site to disaster recovery site
    - Mostly used for DR testing
    - Create backup from repositories without affecting workload
    - Facilitate data-center migration with zero-data loss
    - Get replicas offsite up to 50x faster and save bandwidth
    - Enterprise edition supports built-in WAN acceleration to Veeam Cloud Connect targets only
    - Enterprise plus edition supports built-in WAN acceleration to any target
    - Image based replication
    - Fail-over and fail back
    - Replication from a backup
    - Planned fail-over
    - 1-click fail-over orchestration
    - Built-in WAN acceleration
5. Search and restore
    - User should be able to perform advanced search and restore

Deployment architecture (simplified view)



Logical Architecture

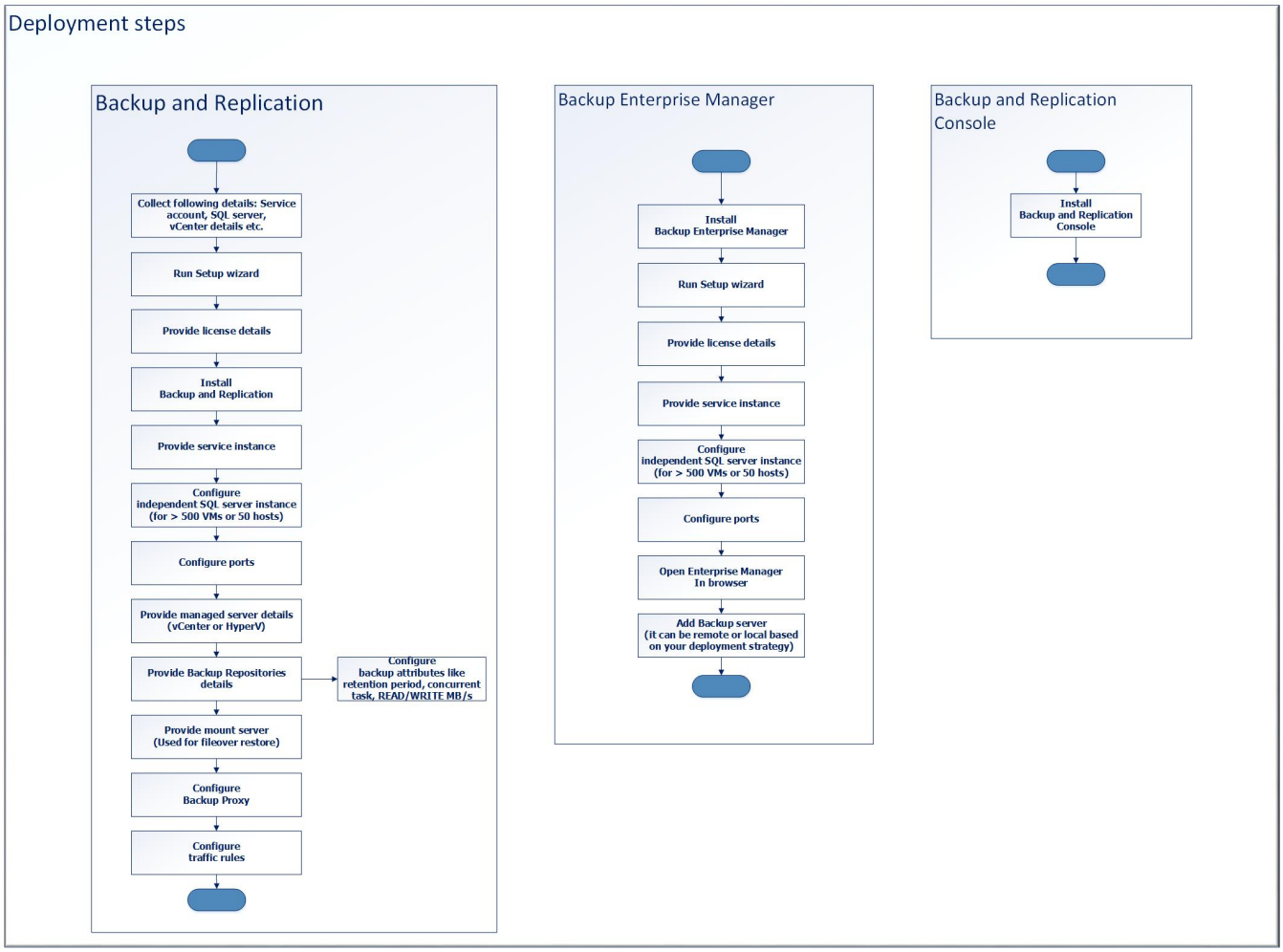


Deployment sequences

The below steps lists how we can deploy and configure Veeam solution for backup of user workloads running for source virtual infrastructure. It does not detail each and every steps very minutely as those can be found in Veeam documentation. The purpose is to give coarse steps which can be exercised with the assistance of Veeam product suite which is self-explanatory up to great extent.

1. Preparation phase
    - Windows machine (VM or baremetal?)
    - Windows service account
    - SQL server instance
    - vCenter details
    - Backup solution details (like Microsoft Windows Server or HP StoreOnce)
    - Mount server details
2. Deployment and configuration phase for Veeam
    - SQL server instance
    - Port configuration for Catalog service port, Backup service port and Secure connection port
    - Provide managed servers details: ESXi or HyperV
    - Configure Backup repositories, You can choose one of the following:
        - Microsoft Windows Sever
        - Linux server which can Direct or internal attached server or NFC mount point
        - Shared folder (common for mediocre use case)
        - De-duplicating storage appliance like EMC Data Domain, Exa Grid, HP StoreOnce etc
    - Provide backup configuration attributes
        - Concurrent task settings (a critical parameter based on your system configuration in terms of resources, number of disks)
        - Read and Write in MB/s
    - Provide mount server details
    - Remove default backup repository which is local folder in Veeam server
    - Configure backup proxy
3. Configure jobs for VM backups as per the customer workload.

The same is pictorially represented below!



Pros and cons

The below section captures author review of Veeam subjected to limited time spent on this.

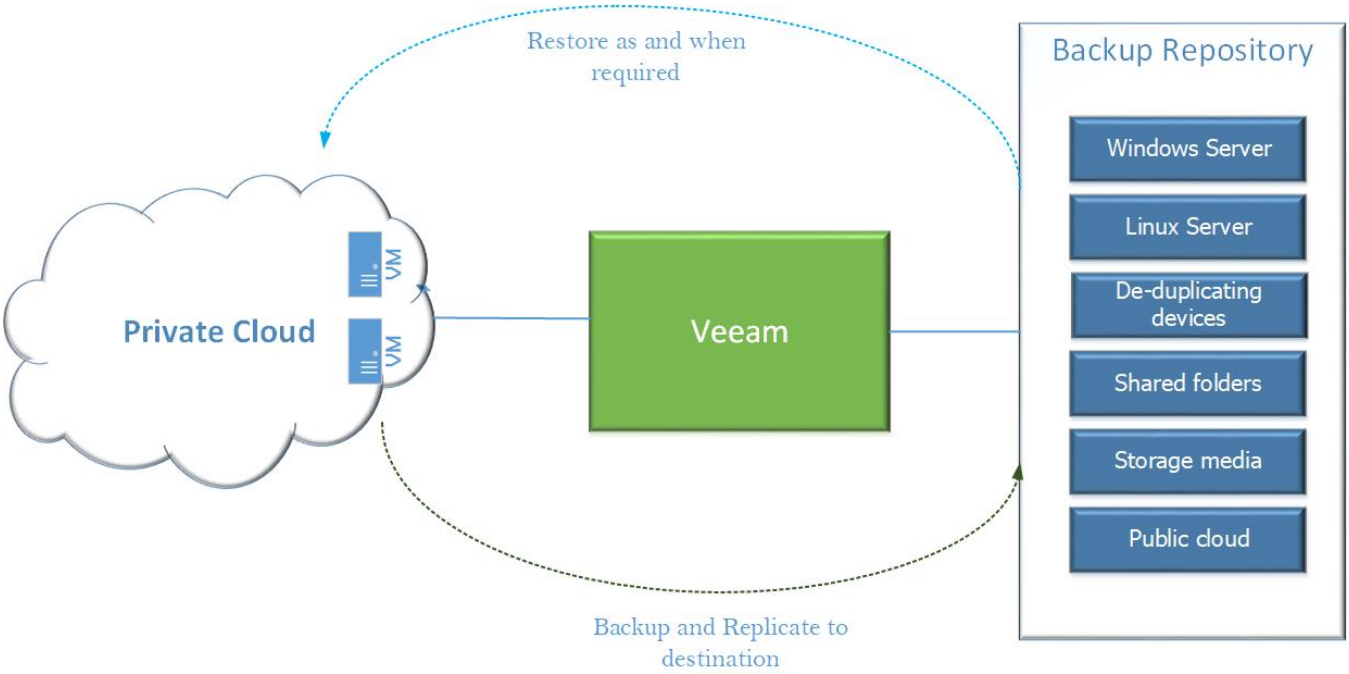
Pros	• Intuitive and self-service portal • Rich feature set • Very good integration with ESXi and HyperV • Supports public cloud integration • Good supportability for HPE Storage System like HPE Store Once, HPE Nimble.
Cons	• Does not support KVM virtual infrastructure • Scale and performance can be eyebrow raising if system has greater than 2000 VMs (to be discovered more) • Data optimization during backup and restoration is average. I feel that HPE SimpliVity can lead here significantly.

Integration guide for Veeam for Backup and Restore

GreenLake Hybrid Cloud (GLHC) cloud platform is a flagship managed cloud offering by HPE. Gemini is one of the product line up which aims to support **VMaaS** (VM as a Service) in private cloud space with the robustness, simplicity, scalability and pay-as-you go model. In other words, the goal of the “VMaaS” offering is to reduce friction between on-prem infrastructure and customers need of public cloud like experience using on-prem hosted service. The solution will allow customers to deploy various “services” (virtual machines, networks, storage etc.) through a self-service portal. The services will run on infrastructure residing on the customer premise or in a co-location facility. In order to simplify overall user experience, HPE will manage and operate certain aspects of the solution. Customers will no longer have to deal with the complexity of managing and integrating cloud services.

As part of managing end user's virtual infrastructure, it is paramount to have a data availability strategy for cloud as things fail in production environment. Backup is of the critical aspect of data protection strategy. it is important to perform backup of user's virtual infrastructure and ability to restore it to point in time when last backup was taken. Also, it needs to be ensured that application workload gets minimal impact when backup is being carried our and restoration of failed entities (VM or virtual disk or VM files) is carried out as soon as possible. One of the way to achieve is to integrate Agena cloud with third party backup solution like Veeam, CommVault etc. In this section, we go deeper into aspects needed to integrate Veeam Backup & Replication with Agena cloud without too much focus on specific of Veeam feature. If you are looking for Veeam features in more detail, please see Veeam documentation.

The below picture depicts very simplified view of connecting Private Cloud, Veeam Backup solution and Backup devices ranging from simple Windows server to storage array to deduplication devices like HPE StoreOnce.



Approach for backup solution belong to primarily two groups: Manged Service or Self-managed Service approach. In former, the control is with service provider i.e. HPE GreenLake team. In later, the control is with end user. Considering Agena VMaaS cloud being a managed cloud and our goal is to leverage customer Veeam infrastructure, the focus of section is on Managed Service mode. In this case, we will not be exposing Veeam API or Veeam UI. All integration aspects, backup and restoration activities will be performed by HPE. HPE will use customer deployed infrastructure. It can be configured to use same infrastructure among tenants and unified jobs for backup scheduling. Or, customer can choose application category based backup infrastructure and different jobs for backup scheduling. Henceforth, customer is responsible for setting the backup policies according to the agreed service level agreement (SLA), and for performing all restore operations after receiving a request from the their end user or on failure of system. Rest of document provides more insight keeping this perspective in mind.

Use cases

Veeam is comprehensive backup, replication and restoration solution. It supports multiple hypervisors, multiple backup media and public cloud including reasonably good DR through replication of data across multiple sites. This document focuses on backup and restoration of VM deployed in private cloud only and does not address remote site backup, disaster recovery or backup to public cloud. As of now, the following use cases are validated and supported:

1. Recover entire VM to the Original or Different host
2. Quickly restore to user by starting a VM directly from a backup file on regular backup storage
3. Recover individual VM files (such as VMX) and virtual disks (vhd, vhdx or vmdk)
4. Full VM recovery
5. Instant VM recovery
6. VM file and virtual disk recovery
7. User should be able to perform advanced search and restore

Deployment architecture

Veeam requires its backup infrastructure (services, network, port) to be configured so that it can connect to vCenter to perform backup of VMs over host management network and store it to destined backup repository. In context of Agena, Veeam backup infrastructure can be possibly deployed in following ways:

1. **Organic deployment.** Customer deploys Veeam on infrastructure used to user workloads or VMs.
2. **Inorganic deployment.** Customer has pre-existing deployment of Veeam and he or she wants to integrate with Agena cloud.

In this document, the scenario (2) is explored, where customer is responsible for hosting Veeam components.

There are plenty of permutation and combination in the way Veeam and Backup devices can be integrated which can vary based on:

- Source storage media used for VMware datastore
- Destination backup media
- Organization of VMware resources like VMs, virtual disk etc
- Backup policy
- Backup restoration need and restoration mechanism applied at the time of recovery
- Data protection strategy
- etc.

It is not possible to explore and illustrate all permutation and combination in this document. This document aim to illustrate how the most common use cases can be realized by following steps detailed below. For others, refer Veeam documentation.

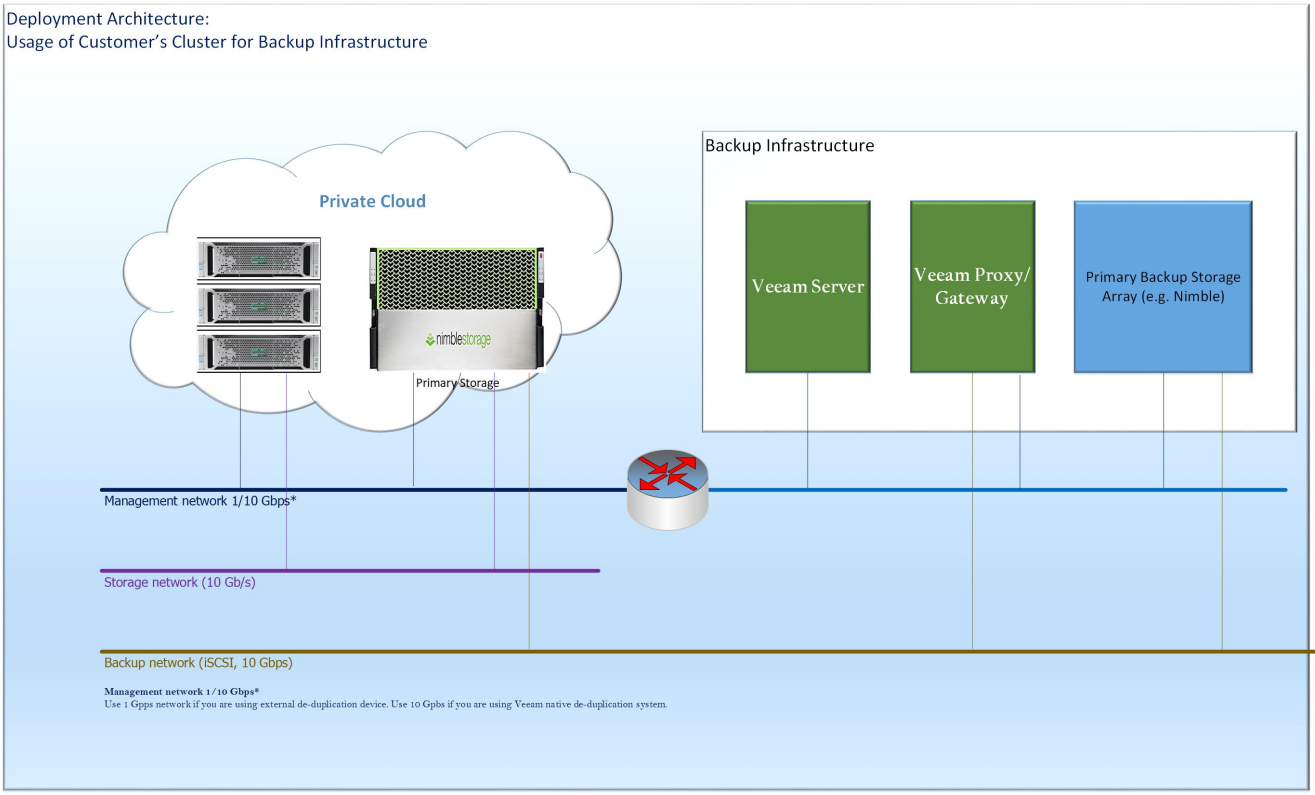
Assumptions

1. Customer is responsible for backup job configuration instead of usage of backup service by direct tenant.
2. The environment is configured to use only primary backup. No secondary backup is configured.
3. Veeam services are deployed as an appliance and hence iSCSI network is used as backup network. It helps to offload the burden on management network.
4. Nimble has been used as primary storage media as well primary backup. Usage of storage instead of tape provides foster faster restoration process and is typically used for secondary backup now a days.
5. Storage-level snapshots is used as VMware hypervisors are relieved of the resource usage due to long-lasting VM snapshots and their consolidation (delete) that occurs at the end of the backup operations.
6. Veeam Backup Proxy is configured to use 'Direct Storage Access' transport mode.
7. iSCSI protocol is used to connect ESXi host with primary storage media.

Deployment architecture

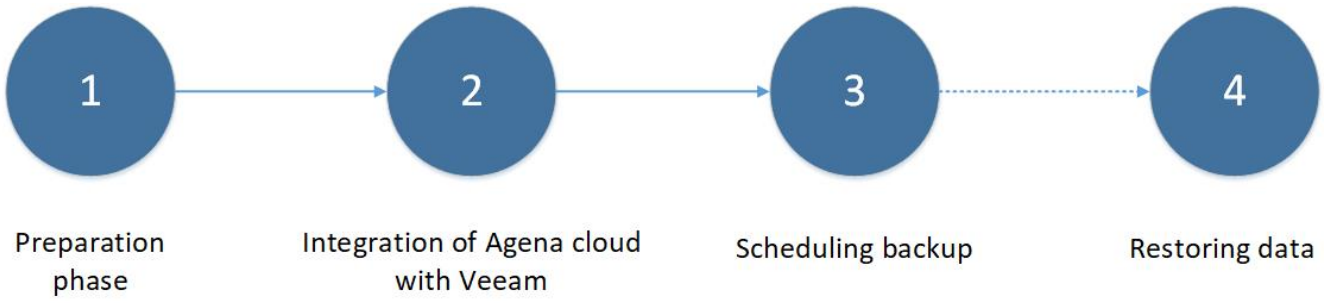
It is not possible to explore and illustrate all permutation and combination in this document. This document aim to illustrate how the most common use cases can be realized by following steps detailed below. For others, refer Veeam documentation. However, it is worth to have a better understanding of networks as mentioned below.

Network	Purpose	Integration specific information
Management network	It is the data center management network used to connect vCenter with ESXi hosts.	As the expectation is to use pre-deployed backup infrastructure used by customer, a routing mechanism needs to be devised where components of backup infrastructure is connected to datacenter management network. It is owned by customer. **Note:** We have not been able to find reference architecture where vCenter management network is not shared with Veeam backup infrastructure. As Veeam uses VADP and hence a router connecting two network should work fine. This piece needs to be tested in lab.
Storage network	It is used to connect ESXi host with primary storage.	It is owned by HPE.
Backup network	It is used to Direct storage access during backup. It helps Proxy servers to read data directly primary storage volumes and hardware snapshots via FC or iSCSI. It greatly reduces the workload on the production hypervisor because the data path is Storage->Proxy instead of Storage->Hyper visorProxy. It is strongly recommended configuration.	Private cloud should facilitate an independent network termed as backup network) which facilitated FC or iSCSI connection among Primary Storage, Veeam Proxu /Gateway and Primary Backup Storage. It is owned by customer.



Deployment and configuration steps for backup operation

The integration activity can be categorized in four groups primarily as mentioned below.



Preparation phase

In this phase, we primarily collect following sets of information which will be needed during integration phase.

Backup repository	Identify backup repository or backup volume in destination storage array or file store based on your backup infrastructure.
DNS information	FQDN or IP address of DNS server begin used in our vCenter environment
Source and target permission	1. Root permissions on the source ESX(i) server. 2. Write permission on the target folder and share. 3. vCenter server permission (more specific detailed below)

vCenter Server permission

1. It is preferred to have admin user with full rights unless and until there is specific reason to do that. It is so because configuring elemental operation might get interrupted because of access reason if not configured properly or changed post configuration. Considering that backup and restore is a managed service and is not exposed to end user, it is reasonably acceptable to proved admin operation.
2. Admin user name and password
   a. Full rights or
   b. Specific rights for activities: Backup operation, Replication, Instance VM recovery, Quick recovery, SureBackup, Entire VM recovery, Replica failover, Replica failback, File-level store and vSphere web client plug-in for Veeam backup and replication. The cumulative details for backup and replication is depicted below.

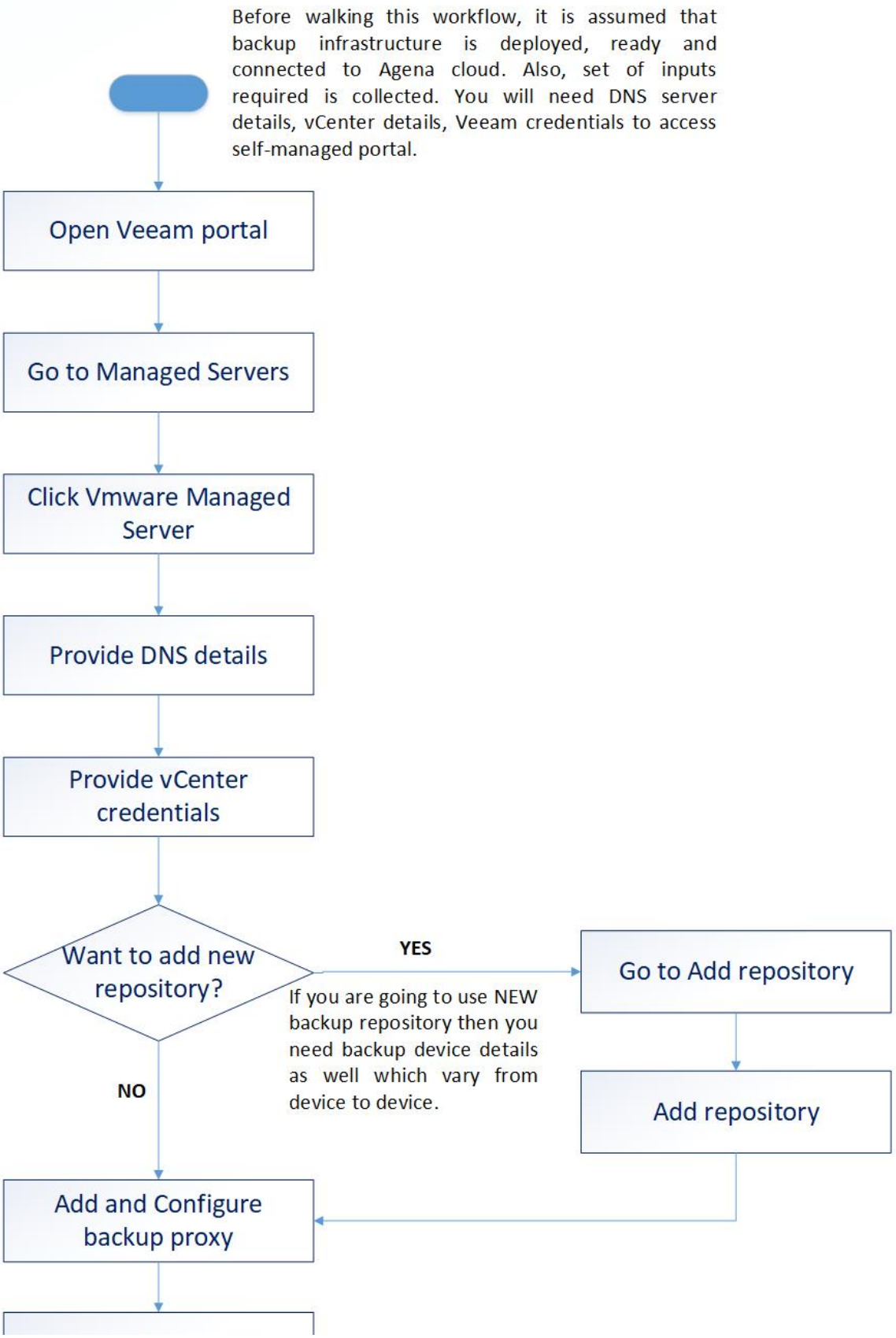
Cryptographic operations	• Add disk • Direct Access • Encrypt • Encrypt new • Migrate
Datastore	• Allocate space • Browse datastore • Low-level file operations • Remove file
Datastore cluster	• Configure a datastore cluster
Extension	• Register extension • Unregister extension
Folder	• Create folder Delete folder
Global	• Disable methods • Enable methods • Licenses • Log event • Manage custom attributes • Set custom attribute • Settings
Host	• Network configuration • Storage partition configuration
Network	• Assign network • Configure
Resource	• Assign virtual machine to resource pool • Create resource pool • Migrate powered off virtual machine • Migrate powered on virtual machine • Remove resource pool
Storage Profiles	• Profile-driven storage update • Profile-driven storage view

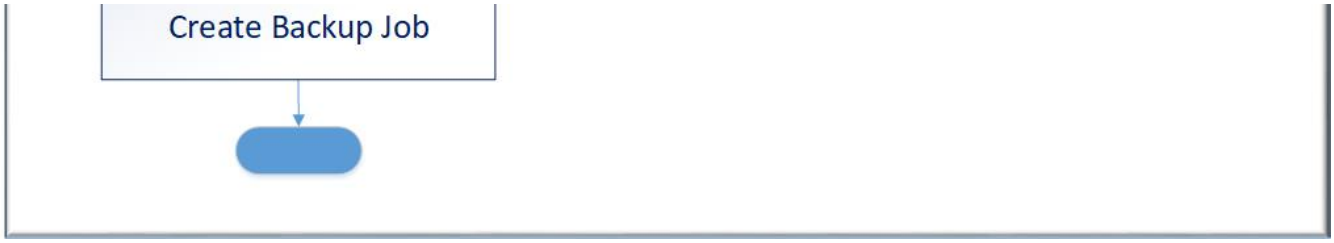
<table>
<tr><td></td><td>
<table>
<tr><td>Virtual Machine</td><td>
<table>
<tr><td>Change Configuration</td><td>• Acquire disk lease
• Add existing disk
• Add new disk
• Add or remove device
• Advanced configuration
• Change Settings
• Change resource
• Extend virtual disk
• Modify device settings
• Remove disk
• Rename
• Set annotation
• Toggle disk change tracking</td></tr>
<tr><td>Guest operations</td><td>• Guest operation modifications
• Guest operation program execution
• Guest operation queries</td></tr>
<tr><td>Interaction</td><td>• Console interaction
• Connect devices
• Guest operating system management by VIX API
• Power Off
• Power On
• Suspend</td></tr>
<tr><td>Edit Inventory</td><td>• Register
• Remove
• Unregister</td></tr>
<tr><td>Provisioning</td><td>• Allow disk access
• Allow read-only disk access
• Allow virtual machine download
• Allow virtual machine files upload</td></tr>
<tr><td>Snapshot Management</td><td>• Create snapshot
• Remove snapshot
• Revert to snapshot</td></tr>
</table>
</td></tr>
<tr><td>vSphere Tagging</td><td>• Assign or Unassign vSphere Tag</td></tr>
<tr><td>dvPort Group</td><td>• Create
• Delete</td></tr>
<tr><td>vApp</td><td>• Add virtual machine
• Assign resource pool
• Unregister</td></tr>
</table>
</td></tr>
<tr><td>Credential for Veeam services</td><td>1. Host name or the IP address of the Veeam Backup Enterprise Manager.
2. HTTP(S) port of the Veeam Backup Enterprise Manager API
3. Username and password to authenticate with the Veeam Backup Enterprise Manager.</td></tr>
</table>

Integration of Agena cloud with Veeam

The below workflow how one can integrate Veeam with private cloud for backup and restoration activity of VM, virtual disks and VM files.

Steps to add Veeam for Private cloud's backup





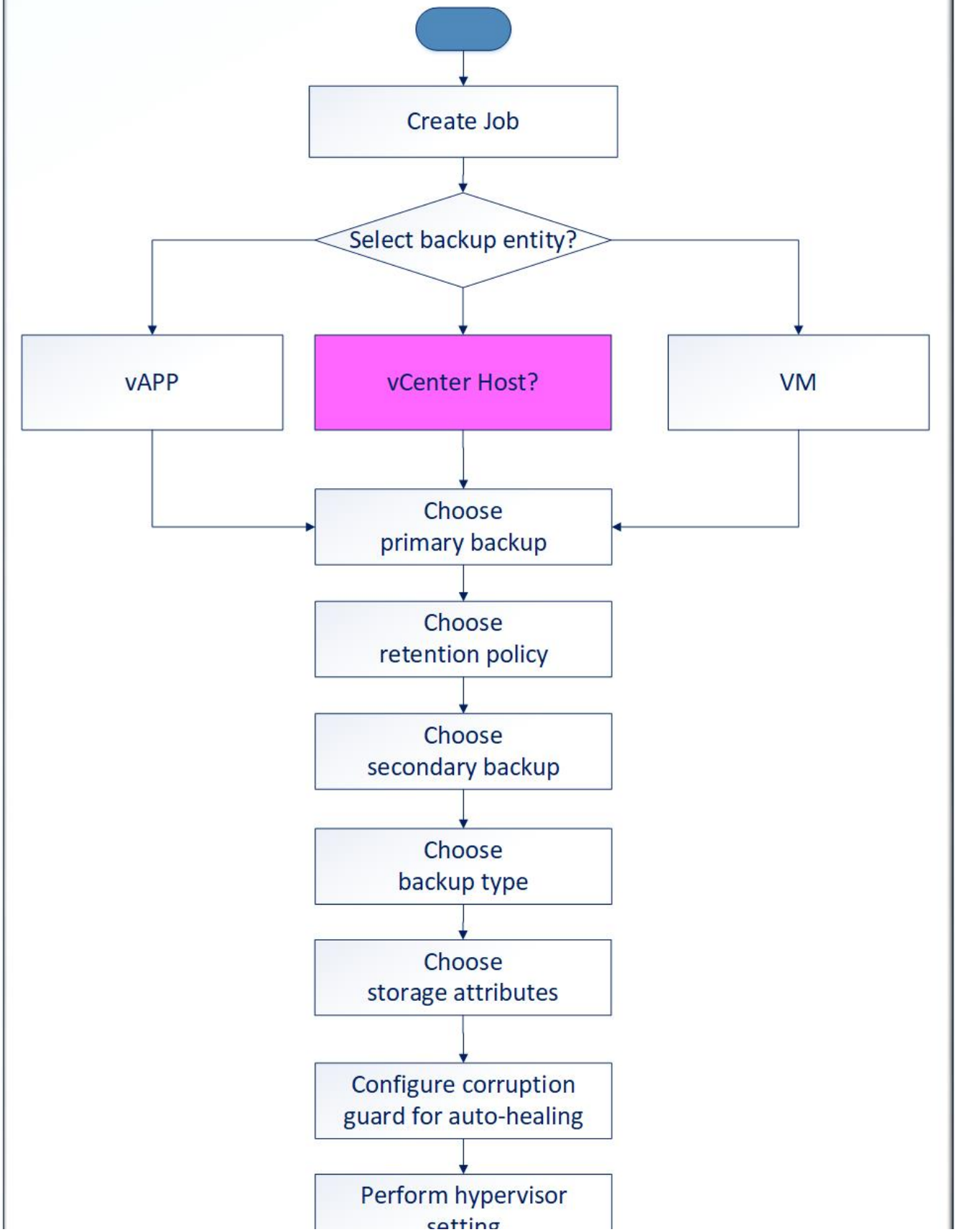
Scheduling backup

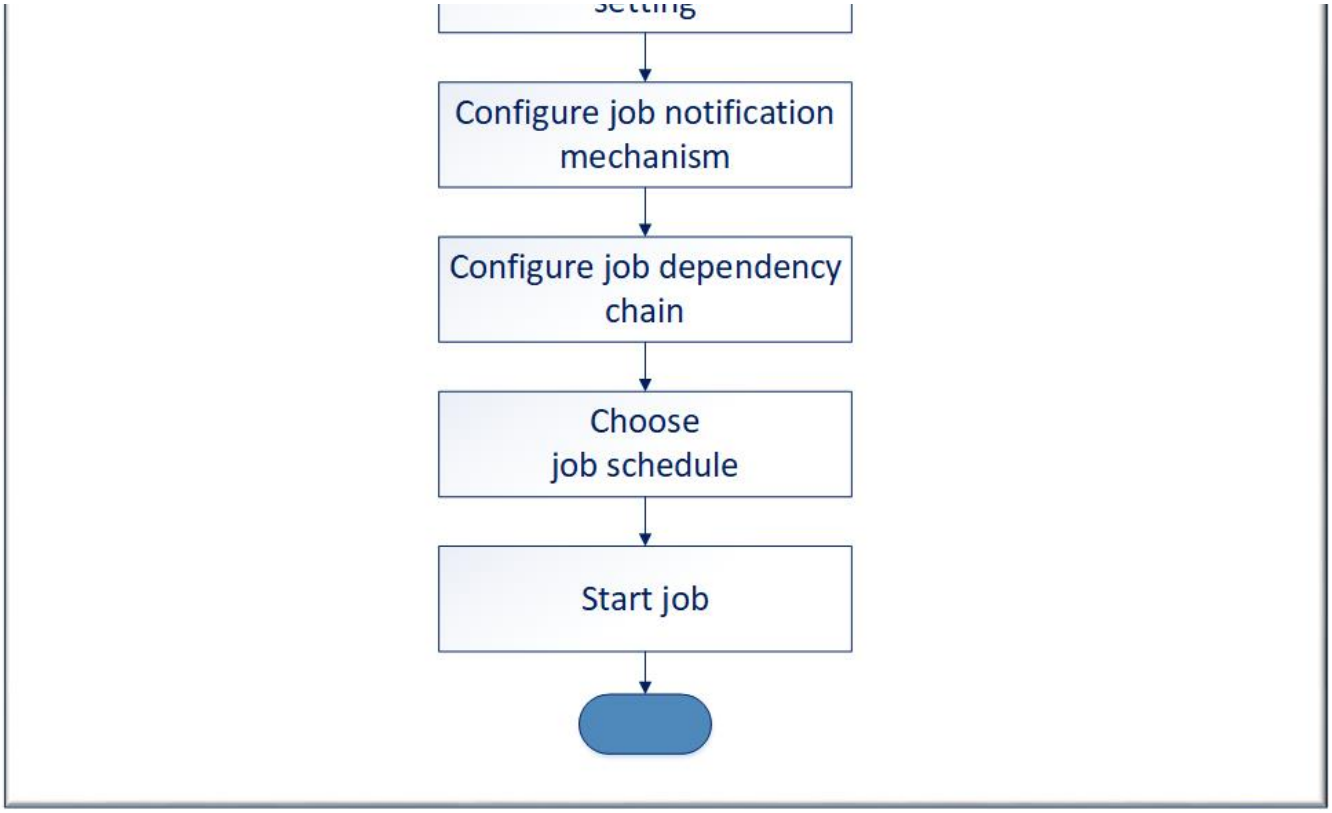
As mentioned above, job is the one which provides scheduling of backup. Configuration of job is very much tied to customer requirement and workload characteristics. We do recommend following standard values, if there is no specific requirement.

Retention period	7 days
Number of simultaneous backup	10
Number of job / VMs	30
Backup type	'Forever Forward Incremental Backup'

The steps to create backup job is depicted below.

Steps to create and schedule backup job





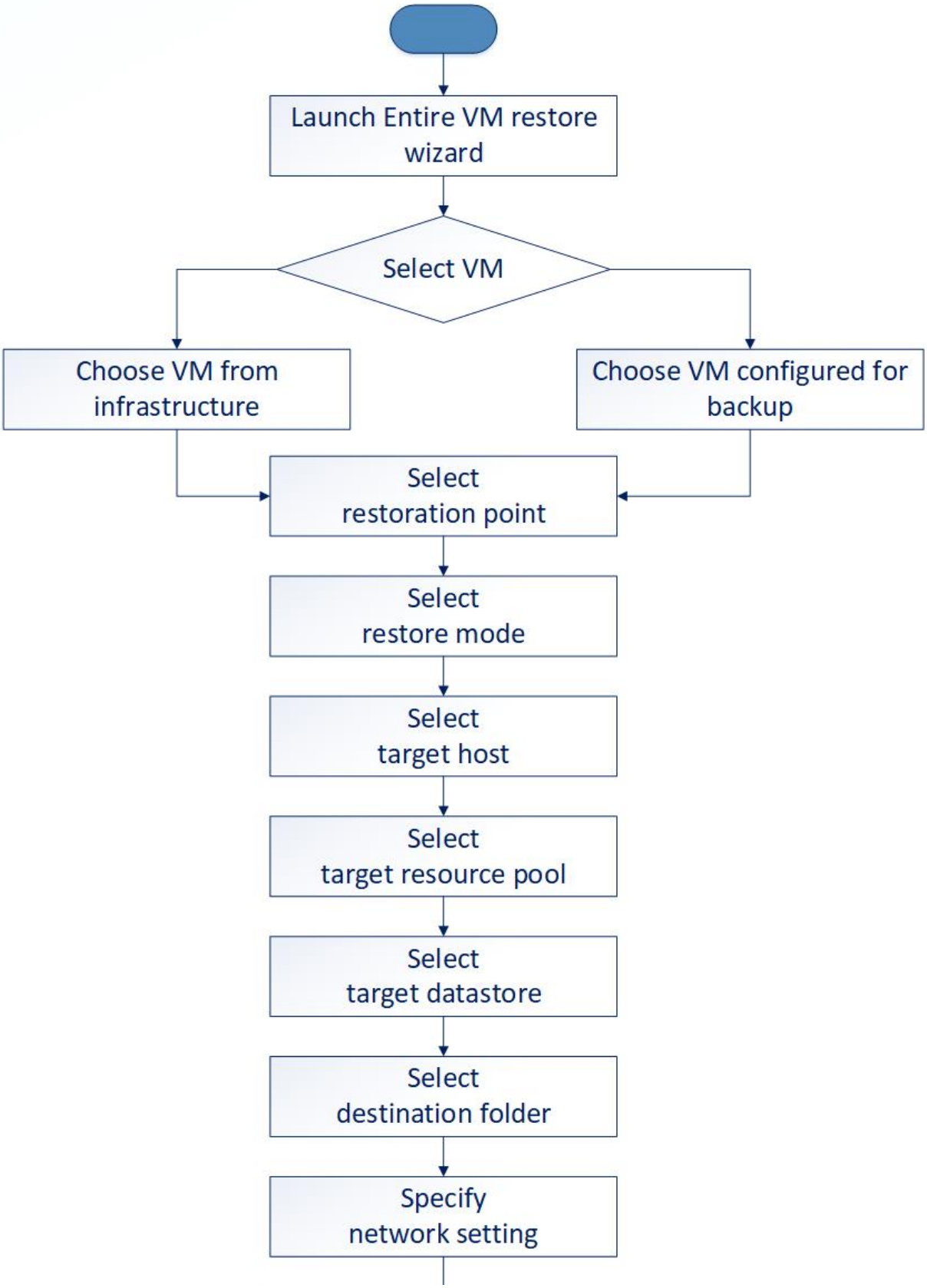
Restoring data

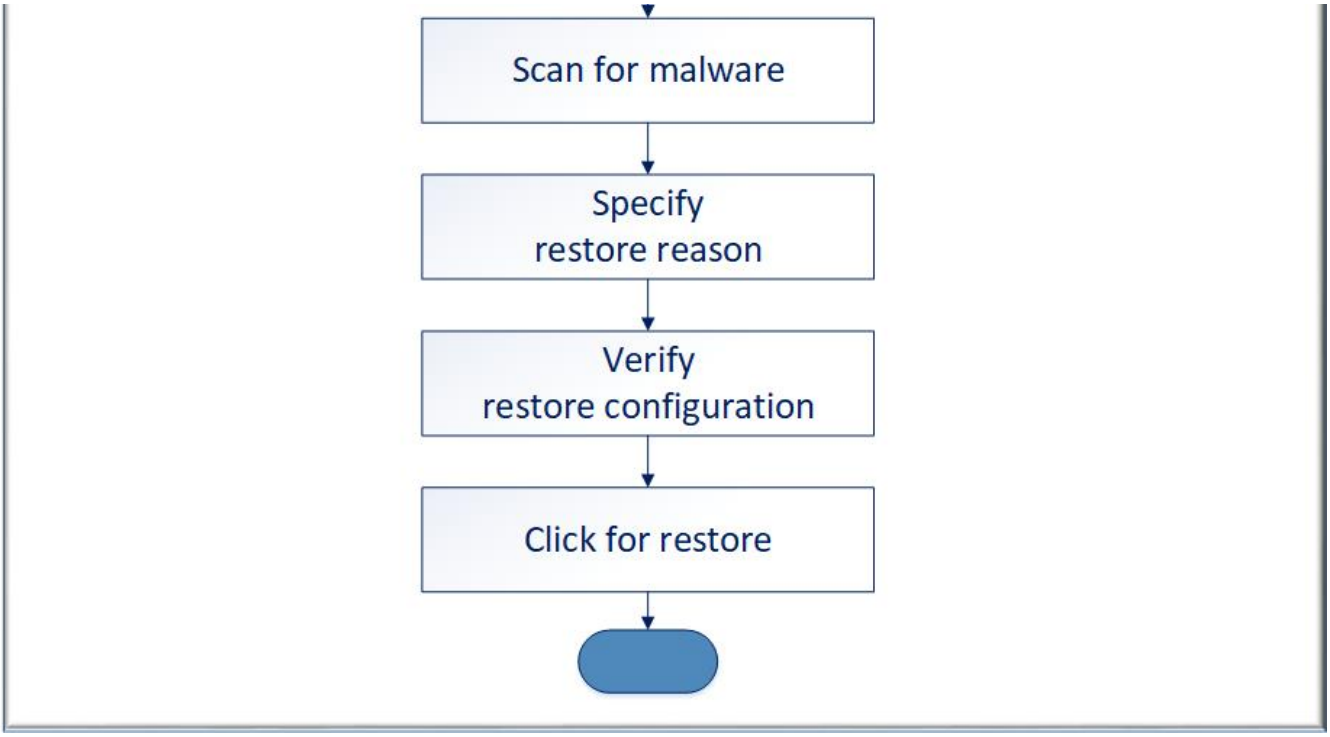
The restoration process depends upon what entities are backed up and which one you want to restore at given point of time. Veeam provides many eco-system. In the context of arena, our focus will be primarily on restoring following entities:

- Restore VM using instant recovery - should be used only for ultra production restoration need
- Restore entire VM
- Restore virtual disk
- Restore VM files

The 'Restore entire VM' use case is commonly used. The workflow for same is depicted below. For other use case please refer Veeam documentation.

Steps to restore entire VM





Best practices (general recommendation)

Backup strategy	• Define your data protection strategy • Category workloads in the terms of medium, high and critical workload • Adhere to 3-2-1 rule

Backup infrastructure	• Use 10 Gbps if you are relying on Veeam Backup Proxy for de-duplication.
	• Do not deploy backup infrastructure in source cloud infrastructure being backed up. Keep source and backup destination different.
	• Veeam deployment
	  • Deploy multiple Backup and Replication server if numbers of VMs are greater than 500.
	  • Use independent instance of Microsoft SQL Server
	• Backup repository
	  • Use Nimble as primary backup storage to meet better RTO instead of tape.
	  • Use secondary backup at remote site

Job configuration	• Number of backups = 10 • Retention period = 7 days • Use 'Forever Forward Incremental Backup' as backup algorithm • Number of VMs per backup job = 30 • [Optional] For backup of secondary backup, use backup copy job in the chain of backup operation.

More activities

The story US421 is necessary for but not sufficient for covering all aspects of providing advisory document for GLHC customer for backup and restoration process. There are other activities like PoC, scale and sizing, integration with public cloud (if desired) etc needs to be covered. Effectively, these can be translated into user stories for further work on it if we intend to provide refined document to customer. At this point of time, the following activities have been done:

- Evaluation of Veeam product suite
  - Feature
  - Logical architecture
  - Deployment architecture
- Define backup and replication architecture for Agena Cloud
- Document procedure to integrate Veeam with GLHC

It will be good to carry out more steps like PoC, Detailing scale and sizing aspect, Integration with HPE de-duplication solution HPE StoreOnce, Integration with public cloud (if desired) etc.

Summary

In a nutshell, Veeam provides rich feature suite for backup, replication and restoration of VM workloads. Also, it can be integrated with different backup repositories like Windows Server, Linux Server, Nimble, EMC Data Domain, Exabyte, HPE StoreOnce etc. The self-service portal really makes admin life simplified. The only aspect which needs to be double clicked is of scale and performance. As per various source of documents, the number of VM counts greater than 2000 seems to be very large infrastructure to be managed for Veeam. And they strongly hint for laying out backup infrastructure and configure it to distribute the load optimally. Otherwise, there is very like hood of performance bottleneck which might affect running workload significantly. It is strongly recommend to carry out scale and performance test in an environment which is closer to user production deployment in terms of size as well as performance expectation.

Presentation