Investigation report on Commvault for Manged Backup of Private Cloud

Revision history

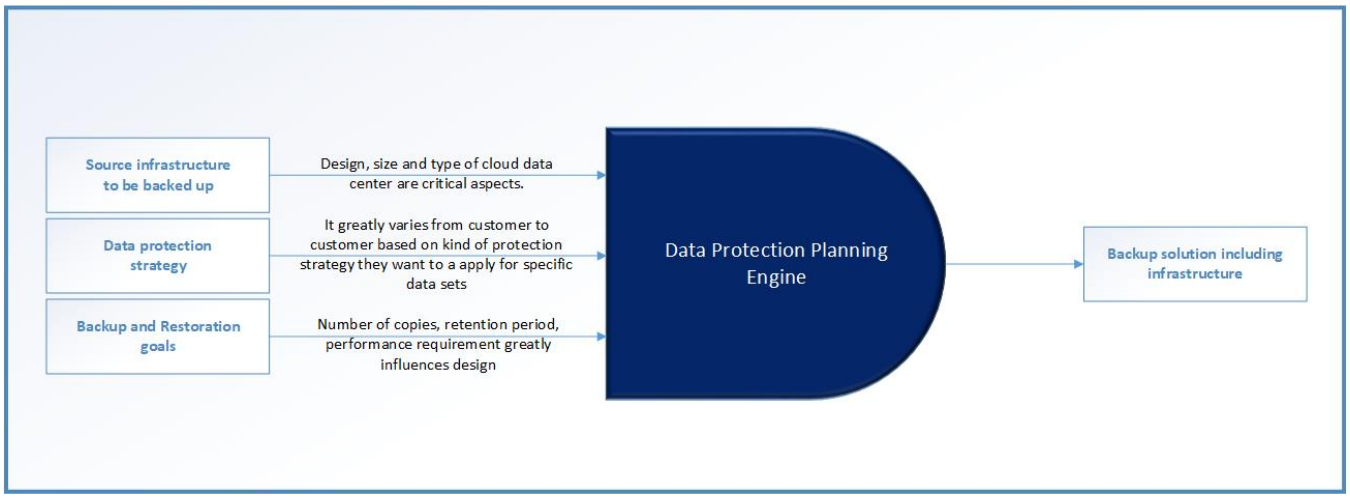
Revision	Author	Date	Description
0.1	Jyoti Ranjan	04-September-2019	Conceptualized the structure of document and drafted generic data protection strategy aspects from learning during CommVault backup investigation.
0.2	Jyoti Ranjan	10-September-2019	Populated the architecture and deployment steps.
0.5	Jyoti Ranjan	12-September-2019	Drafted logical architecture and use cases.
0.75	Jyoti Ranjan	13-September-2019	Drafted deployment architecture along with detailing network interface
0.9	Jyoti Ranjan	16-September-2019	Drafted intellisnap feature, backup and restore workflow.
1.0	Jyoti Ranjan	19-September-2019	Reviewed the content and made many minor changes.

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 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 scope for more 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 CommVault
- Defining backup and restoration solution for Agena cloud using CommVault
- 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 CommVault product suite for backup solution • Backup storage media = Customer selected storage array

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 CommVault

CommVault is one of the popular backup solution. The core of CommVault environment is CommCell which is comprised of following 3 core components:

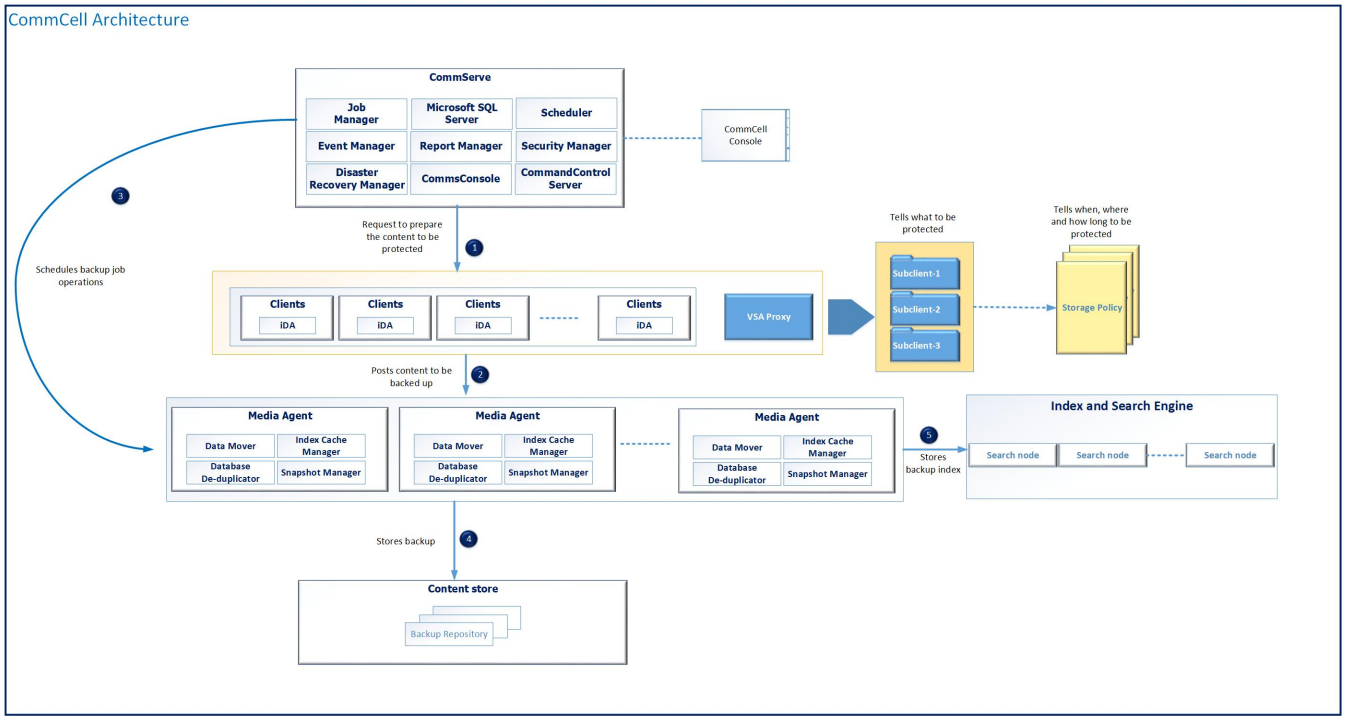
CommServe	It is brain of CommCell architecture and acts as command central server. Its core functionalities are:
	• Administration
	  • Administrative and configuration tasks
	  • Licensing
	  • Metadata management
	• Disaster Recovery
	• Event Orchestration. Uses Job manager
	• Reporting.
	  • Overall health
	  • Alerts
	  • Event logs
	• Security and Authentication. It can integrate with Active Directory or local user security groups

Media Agent	It does participating in routing all data traffic except one specialized dump cases. It is high performance data mover server. It acts as a gateway between client and storage device. It does the actual copy job from source (or production) to target data device. It can have many sub-clients with its own storage policy which tells the aspects like what data is stored, where data is stored, how long data is stored etc. Its core functionalities are: • Acts as data mover. All data must move through media agent to reach destination except in the case of NDMP dumps. • Maintains index cache. It does the job of granular indexing of data management operation • Hosts de-duplication databases • Provides capabilities to execute array function and access to snapshots on the host in IntelliSnap operations
Client	Every client is the entity which host data to be protected. And every client hosts a software component called agents which directly interacts with application or file system for data protection. Examples of IDAs are: File System iDA and Application iDA.

Other than above mentioned pieces, these are following components which participates in various activities during management, backup and restore activities.

CommCell Console	Provides primary administrative interface
CommCell Web based Console	It is IIS based web application and provides browser based interface
ContentStore	It is not a physical entity. It is a virtualized content repository. It also provides catalog services keeping following information: • Where does data resides? • Where was data originated from? • How data got there? It is inherently scalable. One can add storage as demand is needed.
Search engine nodes	It provides content searching and indexing service. It facilitates user to search data at very granular level. it can be deployed like farms if scalability is needed. Every search node receives a cope of data set post-backup or archive operations. It is compute intensive operation and can be real CPU hogger. So, one must take care of resource when deploying it.

Architecture

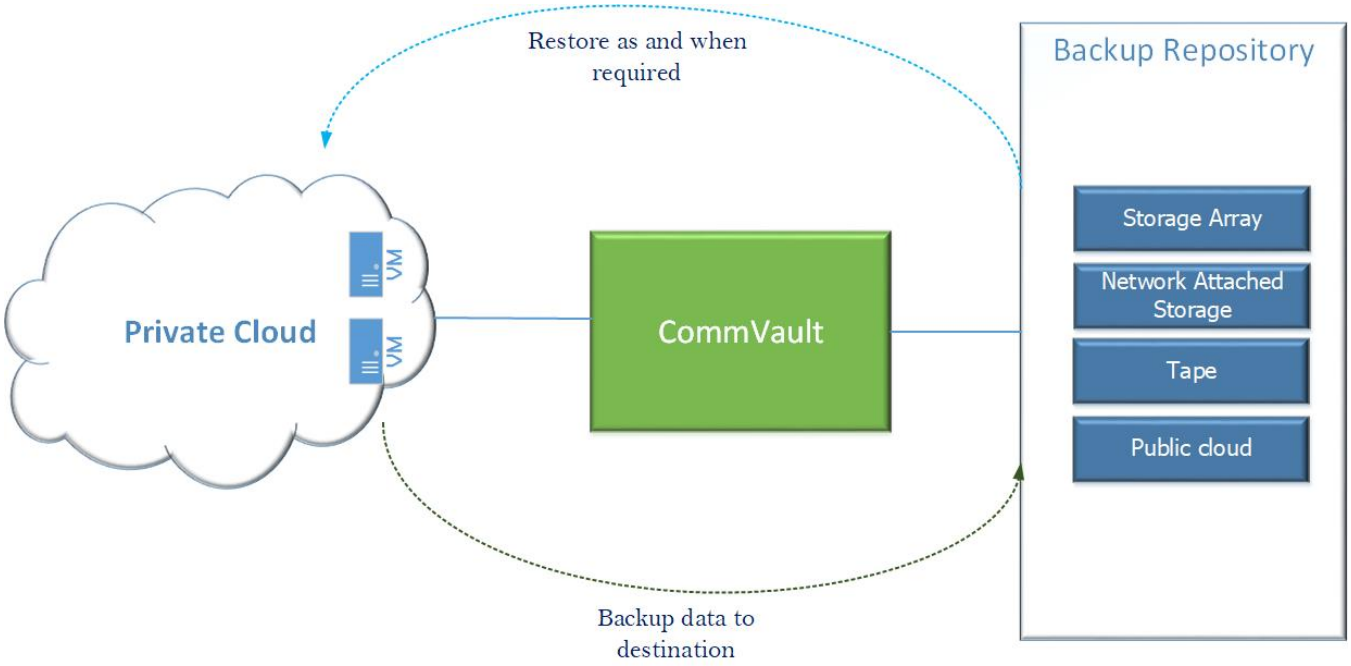


Integration guide for CommVault for Backup and restore

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.

GreenLake will support first approach. Our goal is to leverage customer deployed backup infrastructure. The focus of section is on Managed Service mode but being managed by customer. In this case, we will not be exposing Commvault API or Commvault UI. All integration aspects of backup and restoration activities will be performed by HPE but bringing up and management of backup infrastructure will be done by HPE customers. Customer can configure same infrastructure among tenants and have unified jobs for backup scheduling. Or, customer can choose application category based backup infrastructure and different jobs for backup scheduling. In other words, 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.

The below picture depicts very simplified view of connecting Private Cloud, CommVault Backup solution and storage entities used to host backup data.



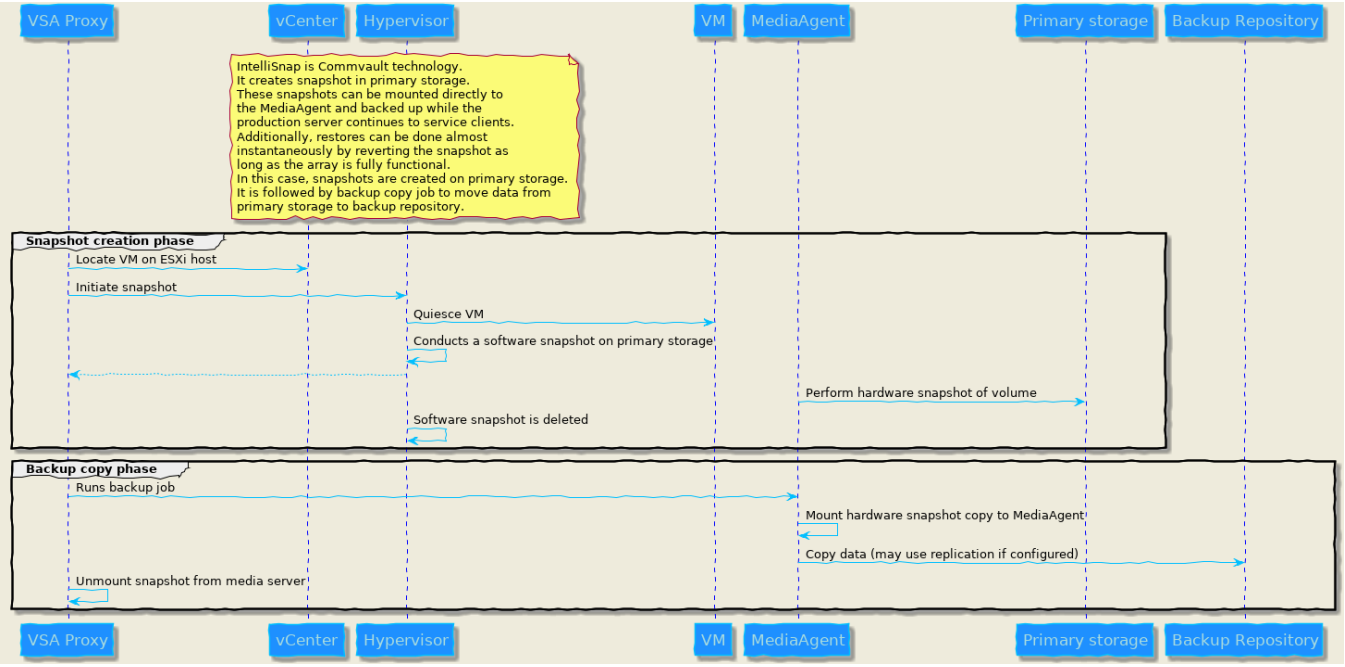
Use cases

CommVault provides a comprehensive backup solution. It supports multiple hypervisors, multiple backup media and public cloud. 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 intended to be supported:

- Backup of VMware objects
  - VM
  - Virtual disk
  - Guest files
- Live VM recovery
- Live Mount
- Live Browse or Live File Recovery
- VM recovery
  - Full VM
  - Virtual disk only
  - Guest files

Using IntelliSnap feature

IntelliSnap backup enables to create a point-in-time snapshot of the data used for backups. An effective way to back up live data is to quiesce it temporarily, take a snapshot, and then resume live operations. **IntelliSnap** backup works in conjunction with storage arrays to provide snapshot **functionality** for backup. It is very common method used for backup if storage array is used in VMware eco-system. IntelliSnap feature does require creation of hardware snapshot in primary storage. So, there must be provision for extra capacity to host snapshot. The IntelliSnap workflow looks like as depicted below.



Deployment architecture

CommVault 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, CommVault backup infrastructure can be possibly deployed in following ways:

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

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

There are plenty of permutation and combination in the way CommVault 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 CommVault documentation.

Assumptions

1. Customer is responsible for bringing up backup infrastructure.
2. The environment is configured to use only primary backup. No secondary backup is configured.
3. 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.

Deployment architecture

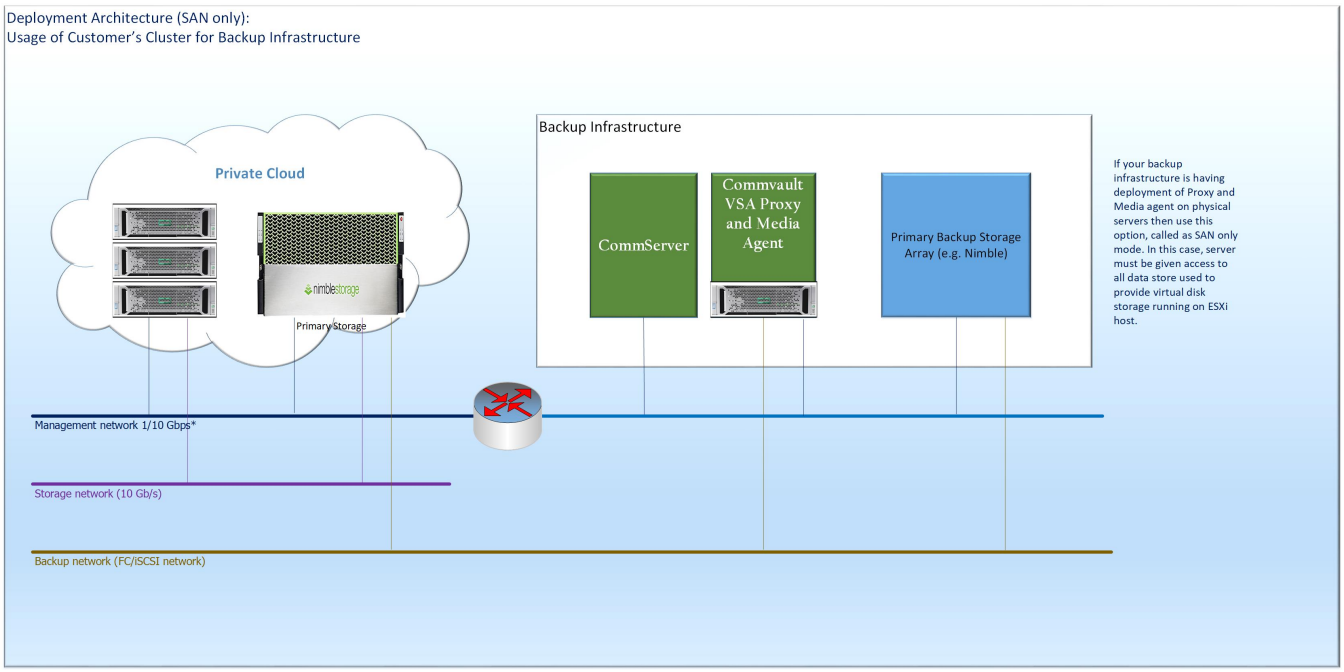
We strongly recommend to use Storage array for primary backup instead of tape etc though Agena cloud. Based on the awareness that Agena cloud uses Nimble storage array and we are going to adapt pre-deployed backup infrastructure by customer, these are following choices we have:

- SAN only mode
- HotAdd mode

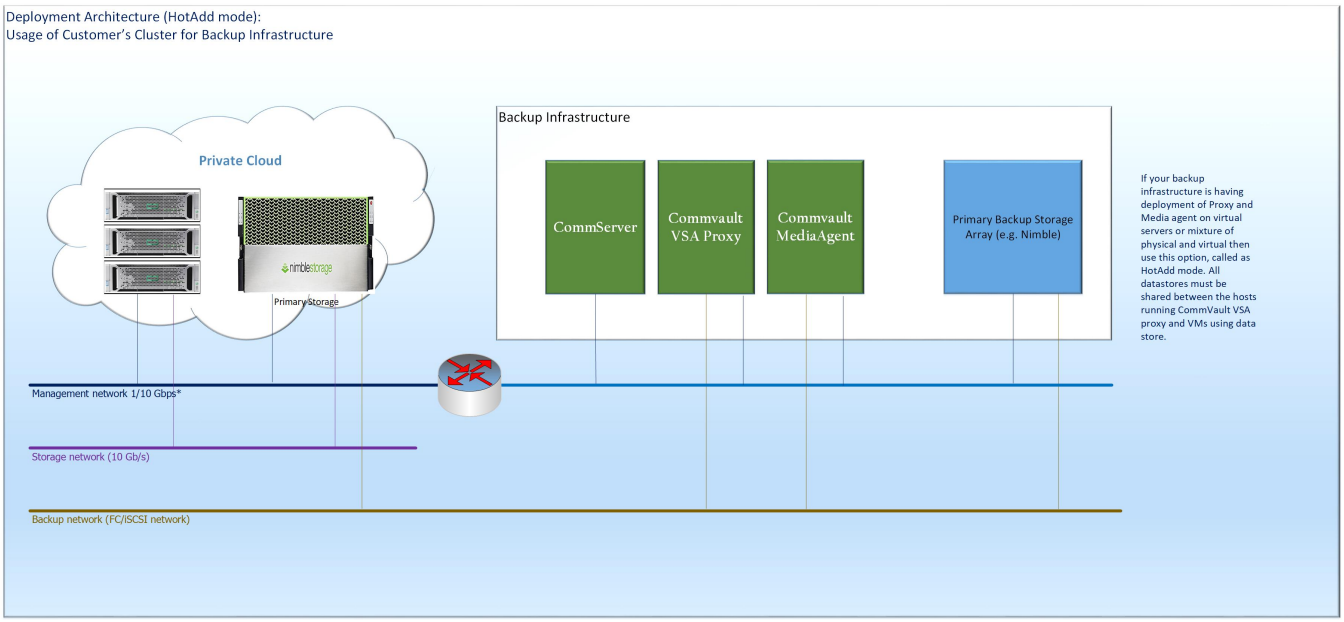
- Network mode (should be used only if tape is used as primary backup)

The network diagram for each use cases are given below. But it is worth to note that Agena cloud supports only SAN only mode and HotAdd mode because Network mode needs connectivity between ESXi hosts and VSA proxy.

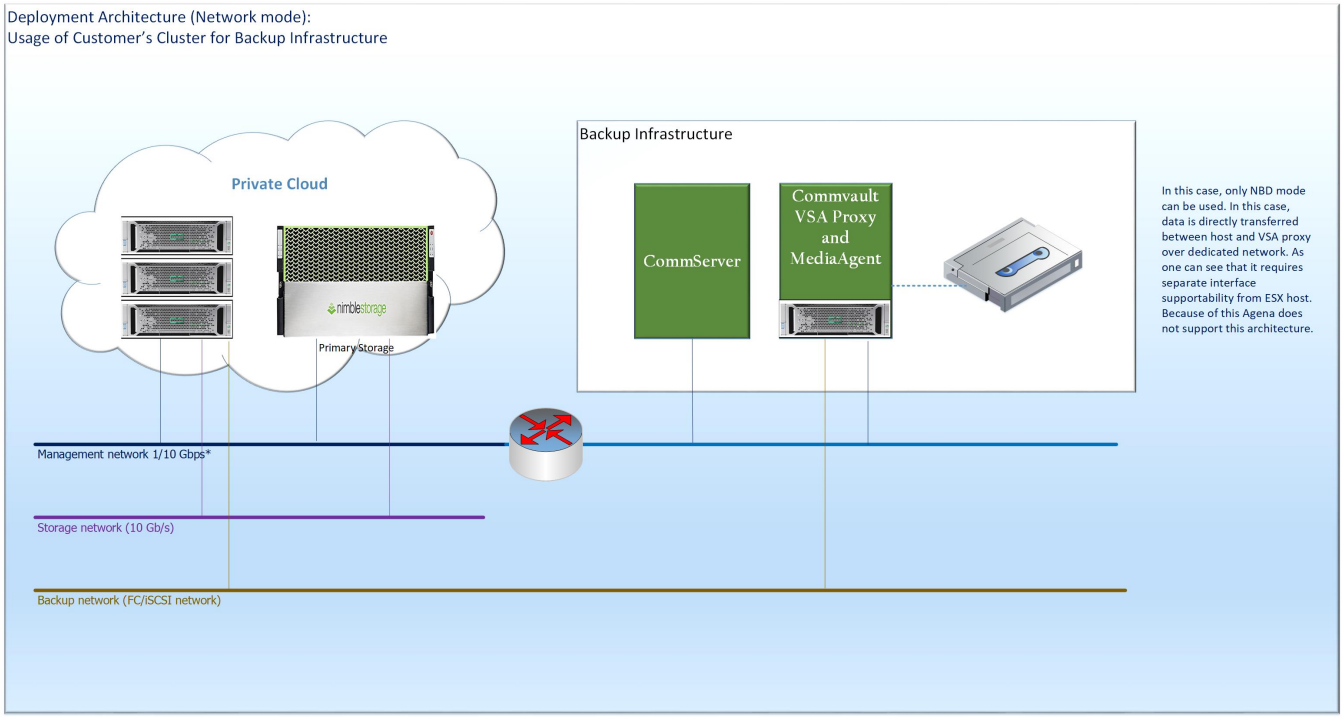
Deployment architecture using SAN only mode



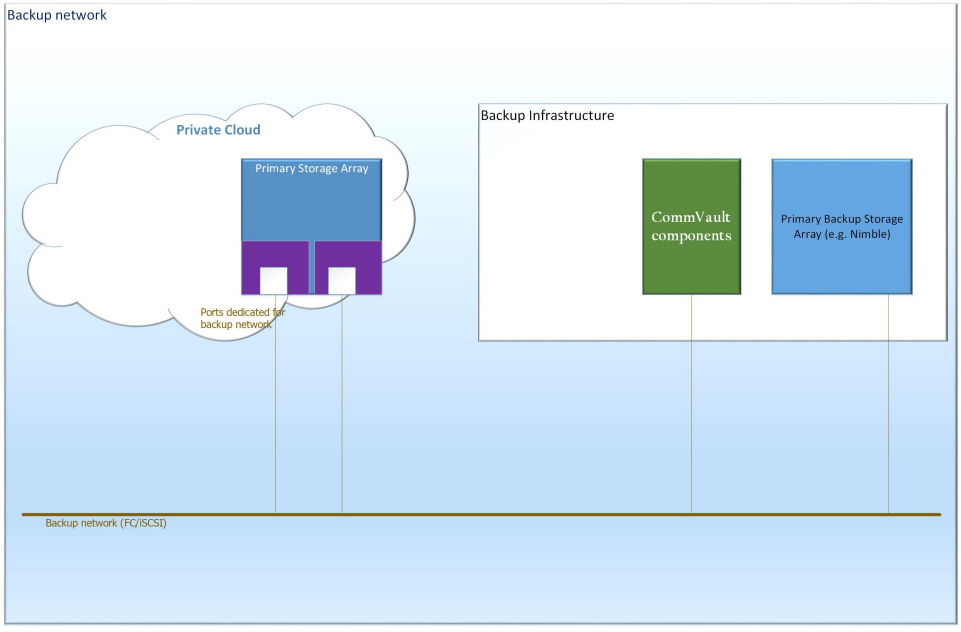
Deployment architecture using HotAdd mode



Deployment architecture using tape library



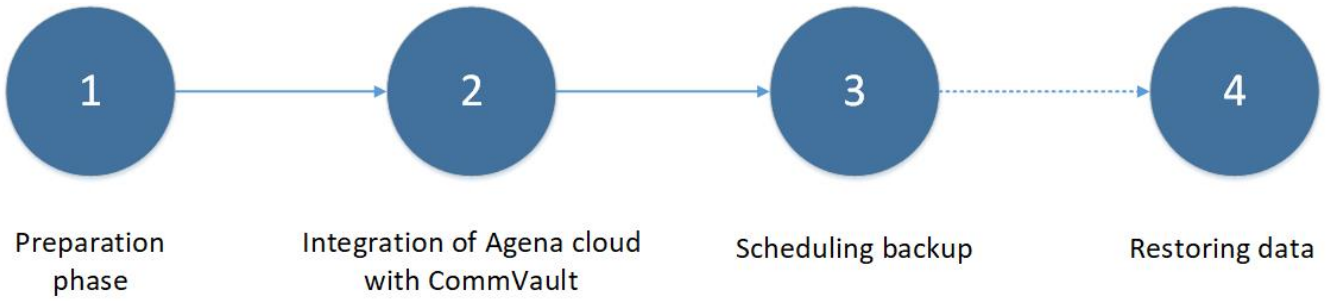
Network details

Network	Purpose	Integration specific information
Backup network	It is used to segregate data traffic from backup traffic.	Private cloud should facilitate an independent network termed as backup network) which facilitated FC or iSCSI connection am components and Backup Storage in case of SAN only and HotAdd mode. The Network mode is not supported. It is owned by c 

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 w 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 bac 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.

Integration steps

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

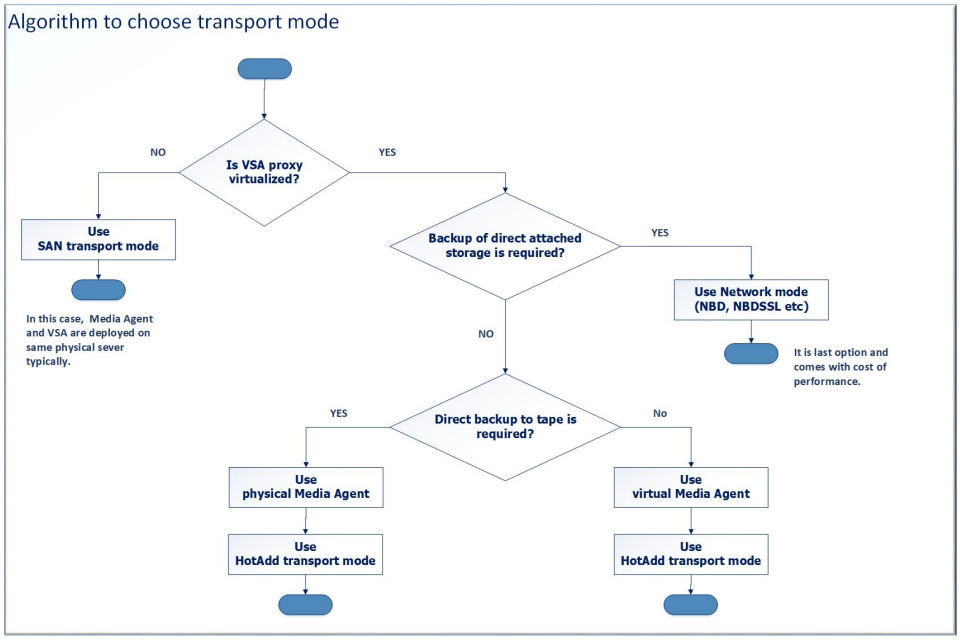


Preparation phase

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

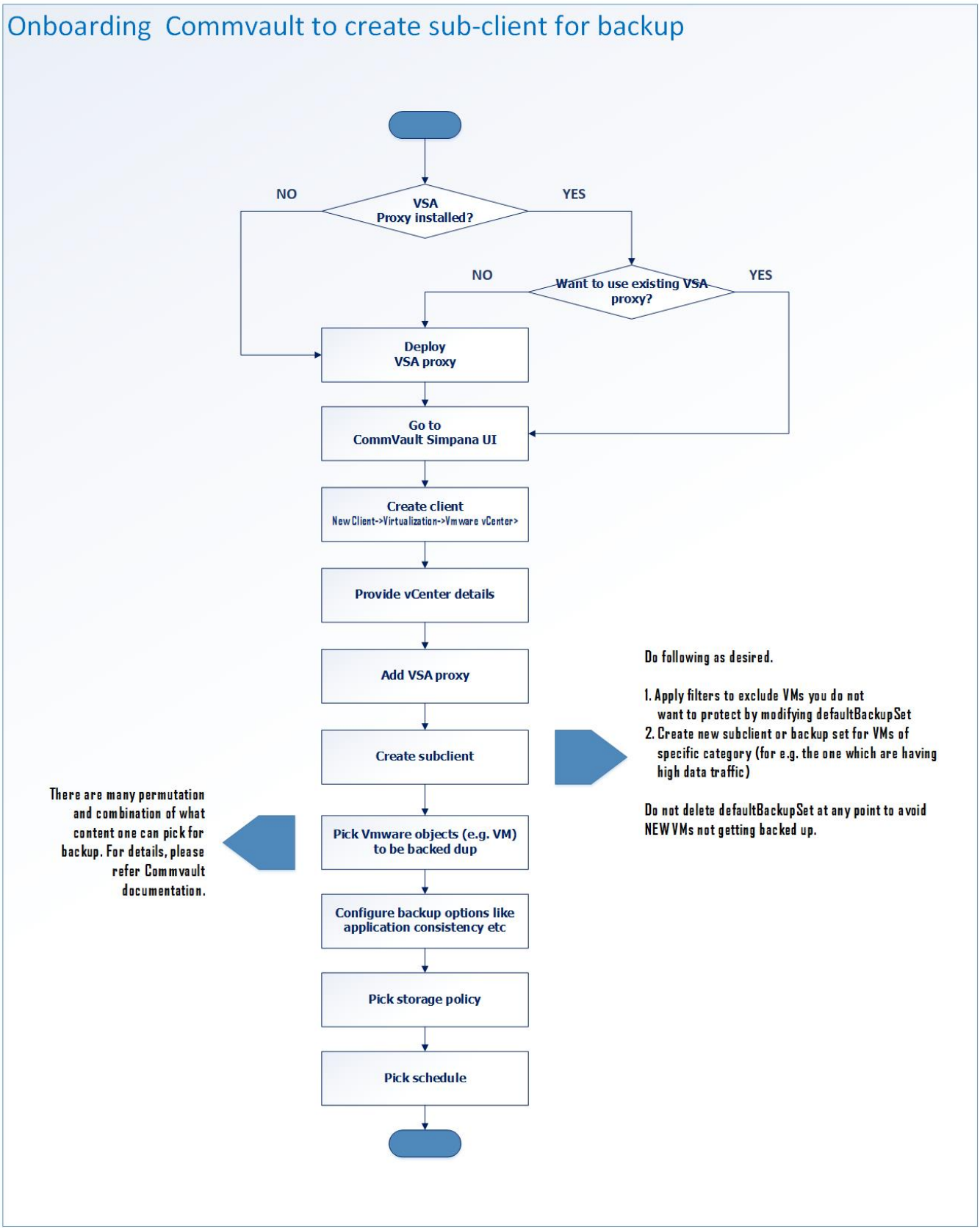
Nimble storage array	The following set of information is needed only if we are using IntelliSnap feature of Nimble storage array.
	
	1. iSCSI Discovery IP of the subnet that carries the data traffic.
	2. User name and password (of administrator role or the power user role)
	3. Snapshot properties
	   • Snap Operation Retry Interval (in seconds)
	   • Snap Operation Retry Count
	   • Enable Diagnostic Logging
	   • CHAP Username
	   • CHAP Password

Backup repository	It is assumed that your backup infrastructure is pre-deployed and ready to be used. As we know that Nimble is being used as primary sto backup copy job to perform backup to your primary backup media in two ways: 1. Nimble storage replication, if your backup repository is also a Nimble storage array 2. Generic backup copy feature supported by Commvault For first case, you need to create the replication relationship between the local and the remote arrays, and set up replication parameters. Nimble Storage Snapshot Replication Feature for more details.
Firewall settings	In an environment with firewalls, the vCenter, ESX servers, Virtual Server Agent, and MediaAgent must be able to communicate with eac components can communicate through the firewall, ensure that the ports for web services (default: 443) and TCP/IP (default: 902) are op each of these machines. See Entering Required Firewall Settings for the VSA with VMware for more details.
vCenter version	One must ensure that appropriate version of vCenter is used before exercising specific feature as mentioned in vCenter Server Versions
vCenter credentials	You can create a separate user account in vSphere for backup and restore operations. The user must have access rights as mentioned i Custom User Accounts
vSphere license	Depending upon the version of vCenter edition, you can use HotAdd transport mode. Please check your vCenter edition. If not then your use different transport mode.

Choose transport mode	Based on your CommVault deployment architecture, you need to choose appropriate transport mode. It is recommended to use auto-sele explicit selection because of technical reason, adhere to following decision flow. As Agena uses SAN array for VMDK datastore, one is e HotAdd transport mode. The selection of transport mode affects backup, replication and restore operation greatly. Refer CommVault doc 
Port requirement	• vCenter Port for web service (default: 443) must be opened. If vCenter is configured to use non-default ports, the non-default ports r • ESXi host Ports for web service (default: 443) and TCP/IP (default: 902) must be opened for the vStorage APIs for Data Protection. • vCloud Director (if used) port for the vCloud REST API (default: 443) must be opened

Integration of Agena cloud with CommVaullt

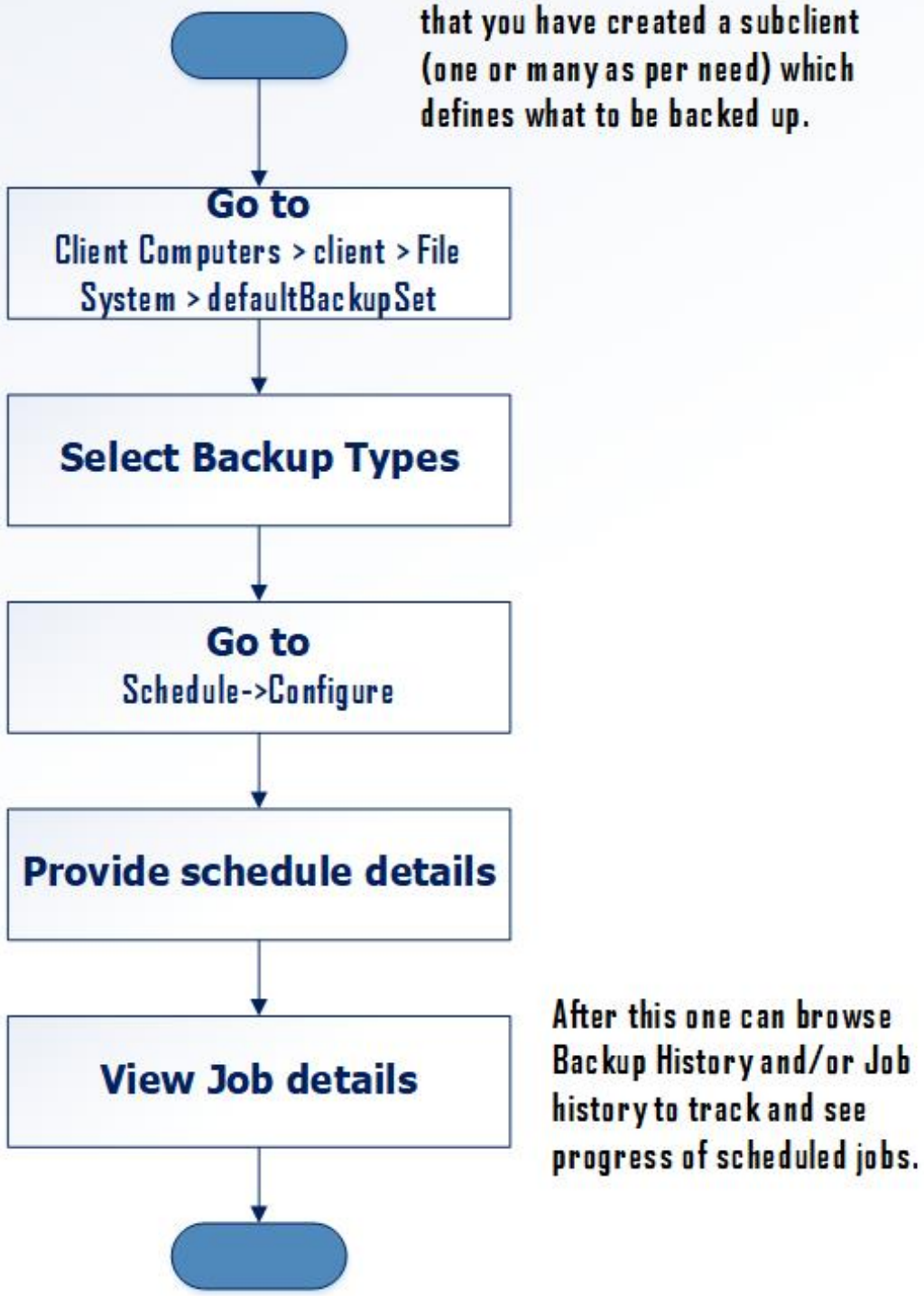
The below workflow tells how to create a sub-client, a term used to define what needs to be backed up. It is typically tied to storage policy. It is advised to have different sub-client for different category of workloads. By category, we mean by high traffic workload, critical workload, normal workload, application specific workload etc. For more details, see Commvault documentation.



Scheduling backup

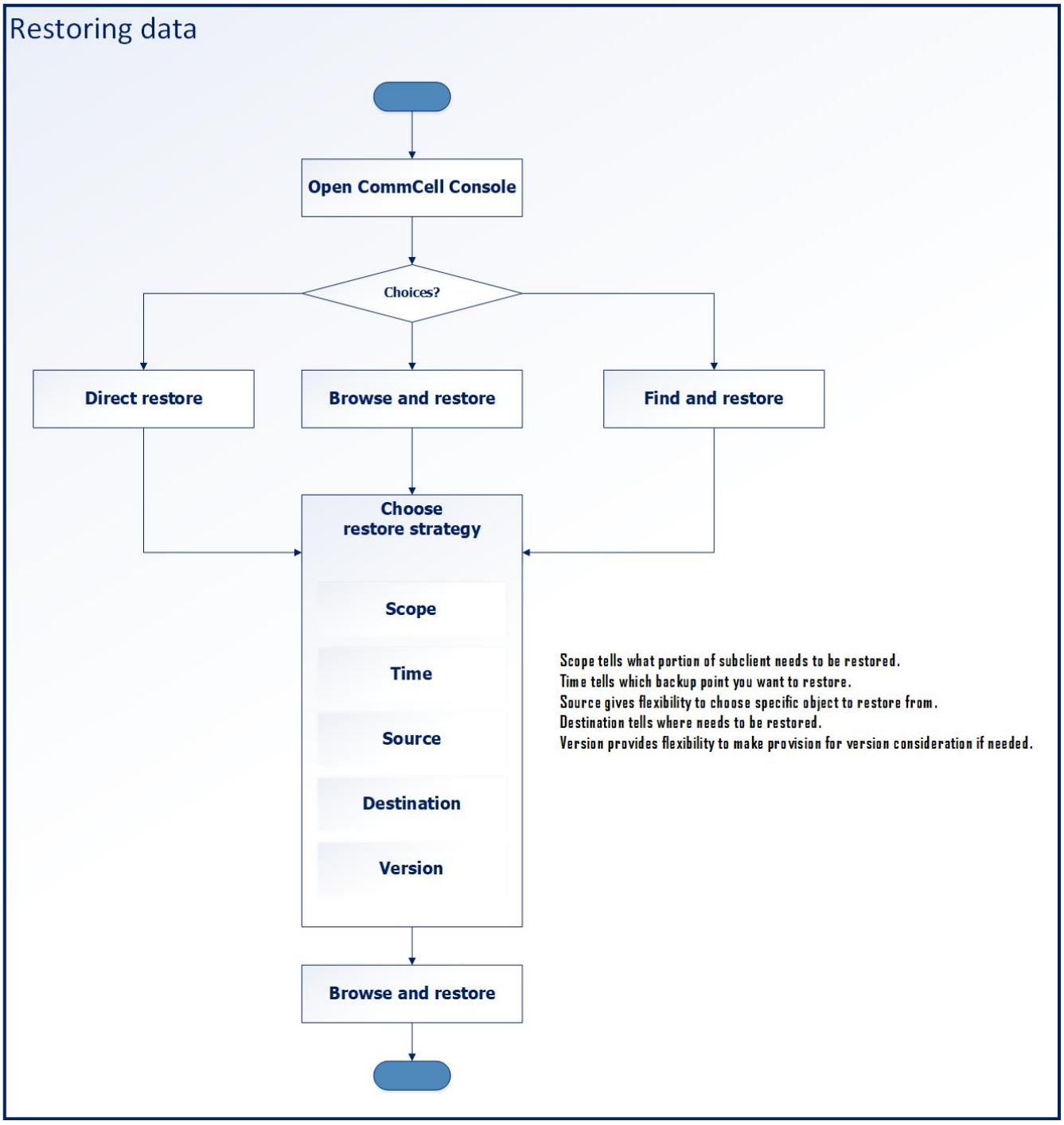
There are many ways to schedule a backup but the most commonly used workflow is depicted below. For others, see Commvault documentation.

Scheduling backup



Restoring data

There are many ways to take a backup but the most commonly used workflow is depicted below. Also, exact steps might vary based on kind of restoration we perform. For e.g. steps for instant recovery will different from normal steps to restore data. For others , see Commvault documentation.



Summary

Comvault Backup, Replication and Restoration provides a rich feature set to solve data availability and protection strategy for Private cloud of varying size. Based on size of private cloud, backup infrastructure for Veeam should be scaled for specific component to avoid sluggish performance. Integration of Veeam involves primarily following steps:

- Collect datasets needed to add and configure Veeam
- Adding Veeam
- Configuring backup job
- On demand, restoring data from backup repository

It is strongly recommended to use storage array as primary backup device and use IntelliSnap feature of CommVault for snapshot management as well as backup operation. Also, the selection of transport mode should be made based on deployment of backup infrastructure at customer site in following decreasing order of preference: SAN copy mode, HotAdd mode and Network mode.