



Royaume du Maroc
Université Cadi Ayyad



Ecole Nationale des Sciences Appliquées de Safi

Département Informatique, Réseaux et Télécommunications (IRT)

Génie Réseaux et Télécommunications (GTR)

Niveau : 5ème Année

Rapport d'avancement

Etude et Mise en place d'un système de management proactive pour le cloud d'INWI

Réalisé par :

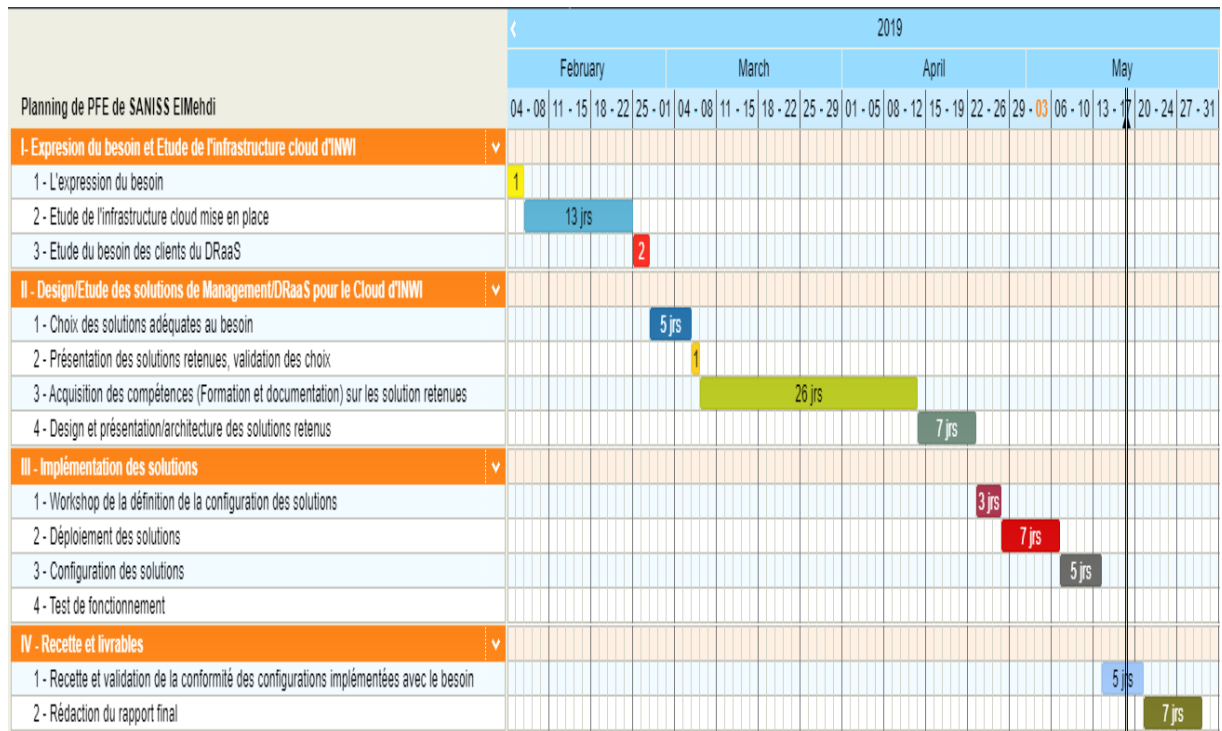
SANISS ElMehdi

Encadré par :

Mr. ZYANE Abdellah

Année universitaire : 2018/2019

Planning du Projet



Liste des figures

Figure 1 : Architecture du Cloud ICT INWI	2
Figure 2 : Architecture de la solution vRealize Operations Manager.....	12
Figure 3 : Architecture de la solution DRaaS avec Veeam Cloud Connect.....	14
Figure 4 : Configurations réseaux avec Veeam Cloud Connect.....	15
Figure 5 : Modèle de Management d'un IaaS	18

Liste des Tableaux

Tableau 1 : Les packs de gestion à utilisé sur vR Ops.....	13
Tableau 2 : Design vSphere pour les SLAs définie	20
Tableau 3 : Les SLAs de l'IaaS	20
Tableau 4 : Définition de la disponibilité pour les SLAs définie	21
Tableau 5 : Définition de la disponibilité pour les SLAs définie (suite)	21

Table des matières

Liste des figures	ii
Liste des Tableaux.....	ii
I. Expression du besoin et Étude de l'infrastructure Cloud d'INWI.....	1
1. L'expression du besoin	1
2. Etude de l'infrastructure Cloud mise en place	1
2.1. Contexte	1
2.2. Besoin Client.....	2
2.3. Etude de l'infrastructure cloud mise en place.....	2
2.4. Etude des besoins des clients du DRaaS.....	8
II. Etude / Design des solutions de Management/DRaaS pour le Cloud d'INWI	10
1. Choix des solutions adéquates au besoin	10
1.1. La solution de management proactive.....	10
1.2. Le DRaaS	10
2. Design et présentation/architecture des solutions retenues	10
2.1. Conception de l'architecture vRealize Operations Manager	10
2.2. Conception de l'architecture du DRaaS avec Veeam Cloud Connect.....	14
III. Implémentation des solutions.....	17
1. Définition de la configuration.....	17
2. Définition du management du cloud	17
2.1. La couche consommateur	18
2.2. La couche fournisseur.....	19
2.3. SLA de la performance.....	19
2.4. La disponibilité.....	20
2.5. Supervision du niveau 1 de service	21
2.6. Supervision de niveau 2 de service	24

I. Expression du besoin et Étude de l'infrastructure Cloud d'INWI

1. L'expression du besoin

Après l'implémentation d'un cloud public VMware chez INWI qui assure le IaaS (Infrastructure as a Service) et le BaaS (Backup as a Service) par l'intégrateur de solutions INEOS SA, des clients ont demandé le DRaaS et ce dernier n'existe pas encore. Les technologies de haute disponibilité inventées par VMware (DRS et HA) sont implémentées sur ce cloud mais il n'y a pas de système de management proactif qui va assurer la remédiation automatique des pannes et l'assurance des qualités de services promises. Pour cela, mon PFE vient pour répondre à ces deux besoins :

- Un système de management proactif,
- Et le DRaaS.

2. Etude de l'infrastructure Cloud mise en place

2.1. Contexte

INWI ambitionne de devenir un acteur ICT majeur, proposant des produits cloud innovants, sous forme de services de cloud publique, aux PME et aux grandes entreprises.

L'approche consiste à commencer avec deux offres différentes définies comme BaaS et IaaS :

- **BaaS** : un service de sauvegarde par les deux offres
- **BIC (Backup In the Cloud)** : permet aux clients de l'offre IAAS de sauvegarder les données de leurs infrastructures dans le Cloud.
- **BTC (Backup To the Cloud)** : permet aux clients externes, qui disposent de leurs propres infrastructures de sauvegarder les données de leurs serveurs/postes dans le Cloud
- **IaaS** : De l'infrastructure dans le cloud sous deux déclinaisons :
- **Offre VPS (Virtual Private Server)** : permet aux clients de louer une seule VM

- **Offre VDC (Virtual DataCenter)** : permet aux clients de louer un espace sous forme d'un DataCenter virtuel dans l'environnement multi-locataires dans le cloud.

Offrant une évolutivité rapide en fonction des demandes des clients et de la croissance de vos charges de travail pour atteindre une capacité de 1 000 machines virtuelles par an.

2.2. Besoin Client

Pour devenir plus agile, INWI construit son propre cloud public. INWI utilisera la solution vCloud Director qui introduit des abstractions de ressources et des contrôles pour créer des pools dynamiques de ressources pouvant être utilisées par les clients hébergés d'INWI. Chaque client hébergé (Tenant) ne voit que les ressources qui ont été explicitement définies au cours du processus d'intégration ou via une demande ultérieure aux opérations du client hébergé.

INWI souhaite également donner à leurs clients, à leurs opérations et à leurs équipes de déploiement les moyens d'étendre les fonctionnalités prêtes à l'emploi pour répondre à leurs besoins actuels et futurs, tout en restant suffisamment souple pour répondre aux besoins des fonctions métiers, afin de répondre rapidement aux besoins du marché et de s'adapter à l'évolutivité options avec la possibilité de contrôler les coûts.

2.3. Etude de l'infrastructure cloud mise en place

En respectant le VMware Validated Design, l'architecture de projet **Cloud ICT INWI** est basée sur un certain nombre de couches et de modules permettant aux composants interchangeables de faire partie de la solution finale.

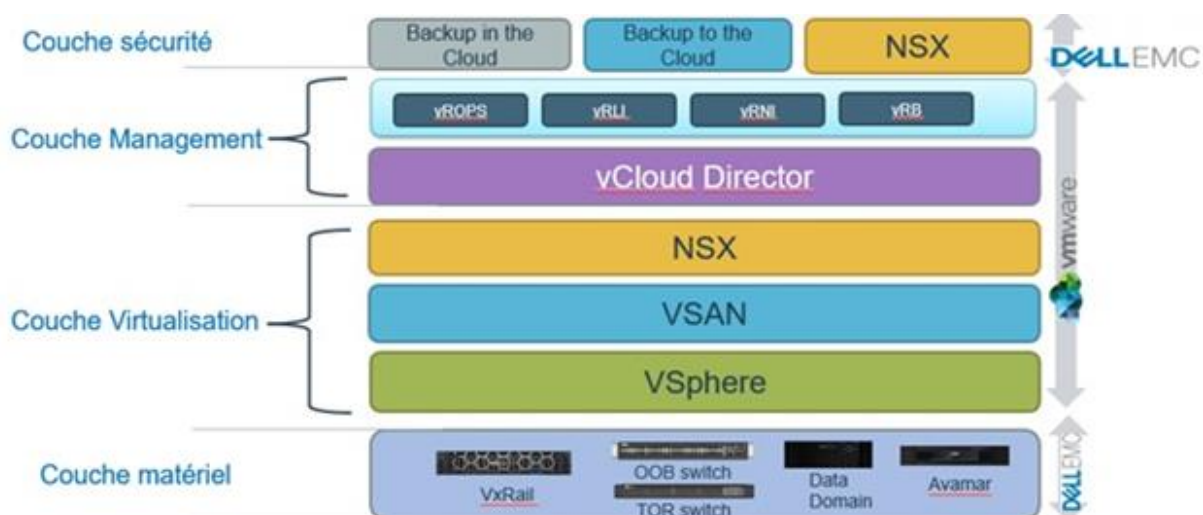


Figure 1 : Architecture du Cloud ICT INWI

Les composants de la solution pour le projet Cloud ICT INWI sont assurés par une collaboration entre les deux éditeurs mondiaux DELL EMC et VMware ; Étant cité, tous les équipements physiques sont de Dell EMC, alors que la partie Software est assurée par VMware.

2.3.1. Couche Hardware

Analogue à chaque infrastructure, la couche 1 réfère ici à l'infrastructure physique qui réside dans le Data Center pour fournir la base pour le reste des couches, Elle spécifie les entités physiques qui fonctionnent sur cette couche, telles que les systèmes de calcul, les périphériques réseau et les périphériques de stockage.

Une fonction clé de cette couche consiste à exécuter la requête générée à partir de la couche de virtualisation ou de la couche de contrôle. Des exemples de demandes émanant des couches, figurent le stockage de données sur les périphériques de stockage, la communication entre les systèmes de calcul, l'exécution de programmes sur des systèmes de calcul, la création d'une copie de sauvegarde des données ou l'exécution d'une politique de sécurité pour bloquer une activité non autorisée.

Compute/Storage:

- Dell EMC VXRAIL 4.5 E460 HYBRID (Management Cluster)
- Dell EMC VXRAIL 4.5 E460F ALL FLASH (Tenant Cluster)

Network:

- Dell EMC TOR (Top Of Rack) Switch S4048
- Dell EMC OOB (Out Of Band) Switch S3048

Backup/Restore:

- Dell EMC Data Domain DD6300
- Avamar M1200

2.3.2. Couche virtualisation

Dans le contexte du Cloud Computing, la virtualisation permet d'exécuter plusieurs instances de serveur, de réseau et de stockage sur le même matériel, chaque instance s'exécutant comme s'il s'agissait de la seule instance, occupant et consommant l'intégralité du matériel sous-jacent. Le Cloud Computing inclut la virtualisation dans trois catégories :

- **La virtualisation de serveur**
- **La virtualisation de réseau**

- **La virtualisation de stockage.**

Les serveurs représentent les capacités informatiques, les capacités de planification des processeurs et d'exécution des instructions de calcul. Le réseau doit diriger, séparer ou agréger les trafics de communication entre les ressources, tandis que le stockage est le lieu où les applications, les données et les ressources sont stockées.

2.3.2.1. La virtualisation de serveur (vSphere)

Par conséquent, la virtualisation de serveur signifie l'exécution de plusieurs instances de serveur sur le même matériel informatique, tandis que chaque instance de serveur (une VM) s'exécute de manière isolée, ce qui signifie que chaque instance de serveur s'exécute comme si elle était la seule à occuper et à utiliser tout le matériel.

2.3.2.2. La virtualisation de réseau (NSX)

La virtualisation du réseau représente la possibilité d'établir plusieurs instances de réseau « un schéma d'adressage IP », une résolution de noms, des sous-réseaux, etc. sur la même couche de réseau, chaque instance de réseau s'exécutant de manière isolée.

NSX introduit la virtualisation du réseau dans l'architecture de Data Center définie par logiciel, il fournit pour la mise en réseau ce que VMware a déjà fourni pour le calcul et le stockage. La virtualisation de serveur permet de créer, de supprimer et de restaurer des machines virtuelles. De même, NSX permet aux réseaux virtuels d'être créés, sauvegardés, supprimés et restaurés à la demande sans reconfiguration du réseau physique.

Les points forts de NSX

- Gestion centralisée des pare-feu distribués basés sur l'hyperviseur et des pare-feu NSX Edge
- Micro-segmentation qui garantit que chaque carte réseau virtuelle est soumise au traitement de la politique en entrée et en sortie
- Stratégies basées sur le réseau, les objets vCenter Server et l'état de sécurité de charge de travail
- Le pare-feu distribué fournit une micro-segmentation qui répond aux défis de sécurité.

2.3.2.3. La virtualisation de stockage (vSAN)

Un concept similaire s'applique à la virtualisation du stockage, qui permet d'agréger différents types de périphériques de stockage tout en les présentant comme un seul périphérique logique avec un espace de stockage continu.

VSAN fournit un stockage sécurisé optimisé pour Flash à toutes les charges de travail vSphere stratégiques. VSAN repose sur des serveurs et des composants x86 aux normes du secteur, qui permettent de réduire le coût total de possession (TCO) de 50 % par rapport aux solutions de stockage traditionnelles. Son agilité facilite l'évolution des services informatiques, tout en offrant le premier système de chiffrement HCI natif du secteur.

En tant que seule plate-forme Software-Defined Storage native pour vSphere, vSAN permet aux clients d'évoluer sans risque vers une infrastructure hyperconvergente (HCI).

2.3.3. Couche de sécurité

La sécurité de la plateforme du projet Cloud ICT INWI est assurée au niveau réseau par la solution NSX. La micro-segmentation du NSX permet d'affecter des stratégies de sécurité à granularité fine aux applications du Data Center, jusqu'au niveau applicatif.

NSX intègre la sécurité directement dans une charge de travail virtualisé sans nécessité de pare-feu matériel. Cela dit, les stratégies de sécurité peuvent être synchronisées avec un réseau virtuel, une machine virtuelle (VM), un système d'exploitation (OS) ou d'autres cibles de sécurité virtuelles. La sécurité est affectée au niveau d'une interface réseau et les stratégies de sécurité peuvent être déplacées avec la VM ou la charge de travail, en cas de migration ou de reconfiguration du réseau.

Le système doit contenir aussi des éléments qui permettent d'assurer la continuité des activités dans les domaines de la sauvegarde, de la restauration et de la reprise après sinistre, afin qu'il soit garanti que c'est un système adapté à la solution choisi par le client **INWI**. Les domaines de la sauvegarde et de restauration garantissent que, en cas de perte de données, les éléments appropriés sont en place pour éviter toute perte permanente pour l'entreprise.

La solution Backup/Restore de l'infrastructure du projet Cloud ICT INWI est réalisé par :

Orchestration Backup/Restore:

- AVAMAR M1200
- AVE (Avamar Virtual Edition)

Emplacement Backup/Restore:

- DATA DOMAIN DD6300 (BIC)

- DATA DOMAIN DD6300 (BTC)

2.3.4. Couche de Management

Les machines virtuelles gérant l'intégralité de la plateforme du projet Cloud ICT INWI sont installées au niveau de la couche Management. La couche management est principalement présentée par le vCloud Director et les produits VMware de Monitoring et d'aide à l'extension (vROps, vRLi, vRNi et vRB).

2.3.4.1. vCloud Director

VMware vCD permet à INWI d'orchestrer les ressources du Data Center virtualisé par vSphere, pour répondre aux offres **Infrastructure As A Service (IAAS)**.

INWI utilise **vCloud Director v9.1** pour introduire les abstractions de ressources et des contrôles et créer des pools dynamiques de ressources afin d'être utilisées par les clients hébergés d'INWI (Tenants).

Le provisionnement de toutes les offres du catalogue sera de type « **Allocation** », Chaque client hébergé (Tenant) ne voit que les ressources qui ont été explicitement définies au cours du processus d'intégration ou via une demande ultérieure aux opérations du client hébergé.

2.3.4.2. VMware vRealize Suite

Le monitoring de la plateforme du projet Cloud ICT INWI est assuré par les produits VMware de la suite vRealize, Cette surcouche s'oriente autour de 4 axes : l'automatisation, le contrôle des opérations, l'extensibilité et la vision « business oriented ».

Les produits composant de la suite vRealize :

- vROps: vRealize Operation Manager
- vRLi: vRealize Log Insight
- vRNi: vRealize Network Insight
- vRB: vRealize Business

2.3.4.3. vRealize Operation Manager (vROps)

vRealize Operation Manager est l'outil de gestion intelligente des opérations de la plateforme Cloud ICT INWI. Il permet via un système de badge de créer des Dashboards personnalisés reflétant en temps réel l'état de l'infrastructure virtuelle. Il permet ainsi de maximiser les

performances et la disponibilité des applications hébergées au sein de l'infrastructure. Il offre aussi des fonctionnalités de Capacity Planning ainsi que d'aide au diagnostic des problèmes des Workloads basé sur des recommandations.

À l'aide des données collectées auprès des ressources du système (objets), vRealize Operations Manager identifie les problèmes liés aux composants du système surveillé, souvent avant que le client s'en rende compte. En outre, vRealize Operations Manager conseille souvent des actions correctives que vous pouvez entreprendre pour corriger le problème immédiatement. Pour les problèmes plus complexes, vRealize Operations Manager propose des outils d'analyse puissants qui vous permettent d'examiner et de manipuler les données des objets pour identifier les problèmes cachés, de rechercher la cause des problèmes techniques complexes, d'identifier des tendances ou d'effectuer des analyses pour évaluer l'état d'un objet.

2.3.4.4. vRealize Log Insight (vRLi)

Plus qu'un simple collecteur de logs, Log Insight permet une analyse en temps réel des logs des composants de l'infrastructure pour fournir des réponses aux problèmes liés aux systèmes, aux services et aux applications. Véritable aide au troubleshooting, il permet de résumer les différents événements remontés par les composants sous forme de diagrammes et Dashboards en supportant une très forte volumétrie en termes d'événements.

2.3.4.5. vRealize Network Insight (vRNI)

VMware vRealize Network Insight aide à créer une infrastructure réseau optimisée, hautement disponible et sécurisée dans l'environnement Cloud ICT INWI, il offre une analyse riche de trafic, permettant de déterminer les faiblesses de la sécurité et de la segmentation du réseau au sein des infrastructures. Sur la base de ces faiblesses, Network Insight fournit des recommandations utilisant les meilleures pratiques de réseau NSX en matière de micro-segmentation. Network Insight collecte les données de trafic auprès des hôtes vSphere.

Network Insight indique la quantité de trafic est-ouest, de trafic commuté et de trafic acheminé. Ces informations sont très utiles lors de l'évaluation de l'utilisation du routage réparti logique NSX. Si la quantité de trafic routé est élevée, Cela peut être optimisé à l'aide du routage distribué logique NSX.

2.3.4.6. vRealize Business for Cloud (vRB)

Ce produit permet de donner de la visibilité à INWI sur la partie business de l'infrastructure. En effet, il permet de gérer les coûts engendrés de manière fine et s'intègre avec vSphere pour permettre notamment la refacturation aux métiers des ressources consommées dans le Cloud. Il permet également de planifier des scénarios et d'évaluer leur impact en termes de coût. Il offre enfin des fonctionnalités de Reporting.

2.4. Etude des besoins des clients du DRaaS

2.4.1. Contexte

La probabilité qu'une entreprise survive à une catastrophe est faible. Parmi les entreprises qui ont connu une perte importante de données, 43% ne rouvrent jamais, 51% ferment leurs portes d'ici deux ans et seulement 6% survivront à long terme. (Cummings, Haag & McCubbrey 2005). En ce qui concerne les incendies, 44% des entreprises ne rouvrent pas et 33% de celles-ci ne survivent pas au-delà de trois ans. (IWS 2004.).

La reprise après sinistre est le processus qui consiste à ramener les systèmes informatiques critiques d'une entreprise à un état fonctionnel. Un plan de reprise après sinistre bien testé permet à une entreprise de continuer à servir ses clients, ses partenaires et ses employés pendant les temps d'arrêt imprévus, et peut ainsi réduire considérablement les pertes de données.

La récupération après sinistre en tant que service (DRaaS) est une option de plus en plus utilisée pour gérer la reprise après sinistre. DRaaS utilise les technologies d'orchestration pour automatiser la réplication et la récupération afin d'améliorer la protection et la facilité de gestion.

2.4.2. Besoin des clients

La majorité des systèmes d'information des sociétés marocaines sont virtualisés avec les technologies VMware et Hyper-V, reste une petite partie qui utilise la virtualisation de Citrix et les technologies de virtualisation open source. Ils ont besoin de répliquer partiellement ou complètement leurs machines virtuelles d'une façon sécurisée, fiable et avec un coût si réduit comparativement avec la construction d'un nouveau site de secours. L'assurance de disposer instantanément de toutes les ressources nécessaires le jour où elles souhaiteront passer à un

niveau supérieur de ressource. De plus, tous les utilisateurs qui manipulent les données doivent être authentifiés de manière appropriée.

2.4.3. Scénarios possibles

1. **DRaaS en libre-service** — Ici, le fournisseur met à la disposition de l'entreprise tous les outils nécessaires pour assembler, surveiller et réaliser les répliques de son environnement de production, pour être prêt en cas de sinistre. L'équipe informatique de l'entreprise intervient au moment de la panne pour exécuter les procédures de reprise.
2. **DRaaS assistée** — Même si cette option n'est que très rarement proposée, ici le fournisseur met à la disposition de l'entreprise les ressources nécessaires, et l'aide à mettre en œuvre, tester et gérer son plan de reprise d'activité. En cas de sinistre, le fournisseur l'aide à compléter les ressources manquantes pour qu'elle atteigne les objectifs de reprise définis.
3. **DRaaS managée** — Avec ce modèle, le fournisseur est responsable de tout. C'est la forme d'externalisation des options de reprise d'activité la plus courante. Le prestataire gère tous les aspects de la reprise — la mise en application de la stratégie de récupération de l'entreprise étant sa priorité absolue.

II. Etude / Design des solutions de Management/DRaaS pour le Cloud d'INWI

1. Choix des solutions adéquates au besoin

1.1. La solution de management proactive

vRealize Operations Manager est la solution de gestion intelligente des opérations dans un cloud VMware. Elle peut gérer toutes ces composants y compris les OS invités et ses applications ainsi que d'autre plateforme non-vSphere et les dispositifs réseaux physique ce qui la rend la solution qui répond parfaitement à notre besoin.

1.2. Le DRaaS

Pour effectuer le DRaaS, les clients répliquent ses machines virtuelles partiellement ou intégralement dans un conteneur vApp. Dans notre cas de cloud VMware, ceci peut être effectué en répliquant d'une infrastructure VMware vSphere du client vers l'un des dépôts de ressources VMware vCloud Director utilisant **VMware vCloud Availability for vCloud Director**, ou d'une infrastructure Hyper-V vers l'un des dépôts de ressources VMware vCloud Director utilisant **Veeam Cloud Connect** (Il supporte VMware vSphere aussi).

2. Design et présentation/architecture des solutions retenues

2.1. Conception de l'architecture vRealize Operations Manager

2.1.1. Analyse de l'état actuel

- INWI a 1 centre de données physique
- Il existe deux (2) instances de serveur vCenter. Un serveur vCenter gère les charges de travail de gestion et un autre serveur vCenter gère les charges de travail de calcul.
- INWI a 1 instance de gestionnaire NSX pour les charges de travail de calcul intégrée au serveur de calcul vCenter.

- Les charges de travail de gestion et de calcul d'INWI sont hébergées dans des banques de données (datastores) VSAN.
- INWI dispose d'une instance de vCloud director permettant aux utilisateurs publics d'accéder au Datacenter virtuel.

2.1.2. Architecture de l'appliance virtuelle vRealize Operations Manager

vRealize Operations Manager collecte et analyse les informations provenant de plusieurs sources de données au sein de l'entreprise. Il utilise des algorithmes d'analyse avancés pour apprendre et reconnaître le comportement «normal» de chaque objet surveillé. Ces informations sont présentées aux utilisateurs via des vues, des rapports et des tableaux de bord. L'interface utilisateur permet aux utilisateurs d'accéder aux résultats des analyses sous forme de badges, d'alertes, de graphiques et de rapports.

Afin de répondre à l'exigence INWI, une instance de vROps est déployée en mode haute disponibilité sur le site de production.

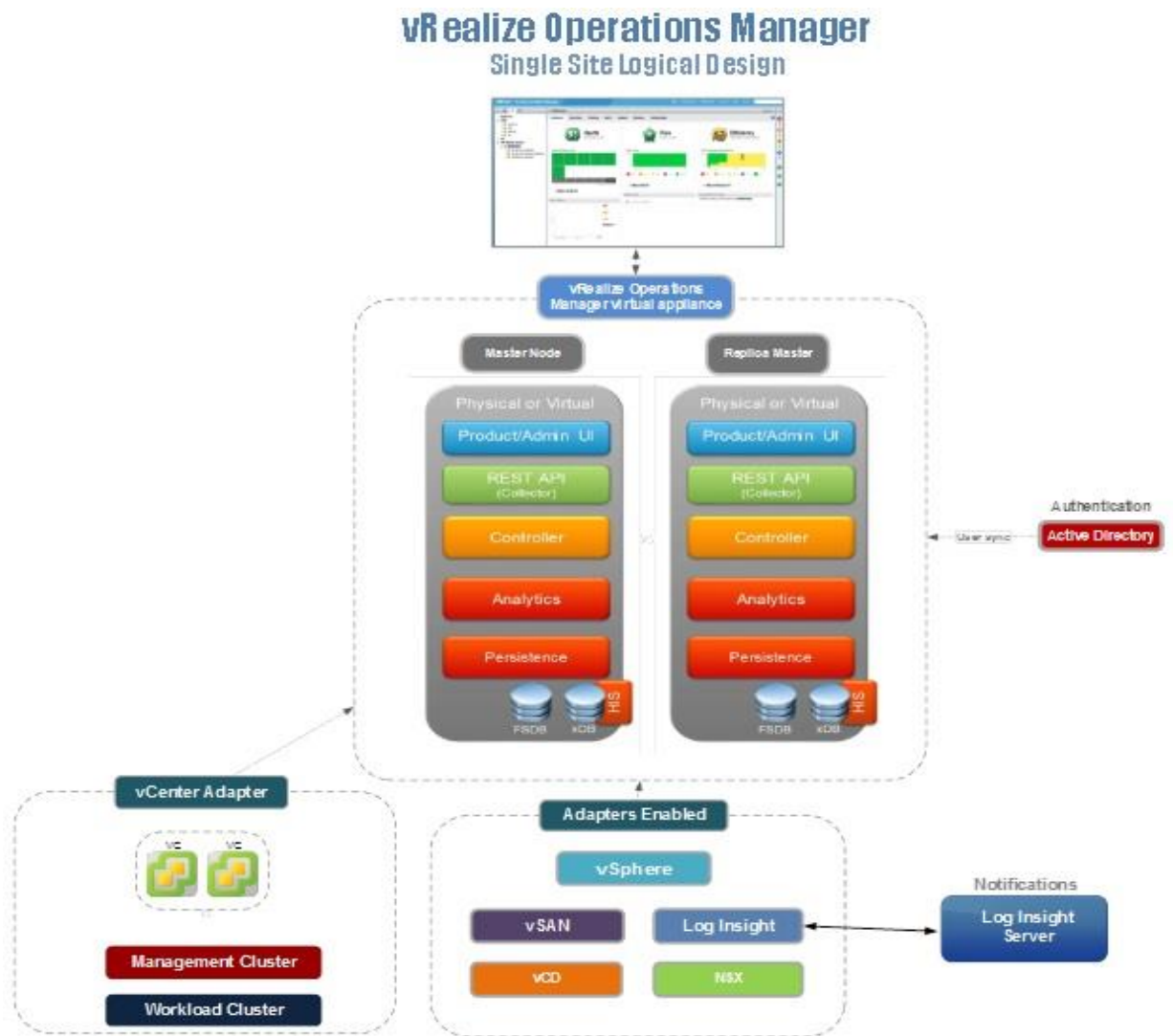


Figure 2 : Architecture de la solution vRealize Operations Manager

vRealize Operations Manager prend en charge l'utilisation de nombreux adaptateurs pour collecter des informations à partir de diverses sources. Ces sources peuvent être situées sur différents sites à partir desquels vRealize Operations Manager est déployé.

Le cluster de nœuds vROps Master & Replica est déployé dans le cluster Management de site de production et surveille les composants suivants :

- Production vCenter 6.5
- VMware NSX pour vSphere
- VMware vSAN
- vRealize Log Insight

- vCloud Director
- Les nœuds Master & Replica sont déployés en une taille Medium
- Le HA est activé pour les nœuds vROps
- Thick provisioning du disque est utilisé pour les appliances vROps pour une meilleure performance.

La connectivité de vCenter Server via vRealize Operations Manager permet d'améliorer la visibilité et la facilité de gestion des données opérationnelles via une interface utilisateur unique.

vRealize Operations Manager est connecté au vCenter du cluster Management et au vCenter du Cluster Tenant (**vc01 et vc02**).

vRealize Operations Manager Management Packs

Le SDDC contient des produits VMware pour la gestion du réseau, du stockage et du cloud. Nous pouvons surveiller et effectuer des diagnostics sur chacun d'entre eux dans vRealize Operations Manager à l'aide de packs de gestion. Fournit une surveillance granulaire supplémentaire pour toutes les applications d'infrastructure virtuelle et de gestion du cloud. Les packs d'administration suivants sont disponibles et doivent être déployés et configurés dans vROps :

Tableau 1 : Les packs de gestion à utilisé sur vR Ops

Management Pack	Version	Constructeur
vSphere (default)	6.0.818	VMware
vSAN	1.0.1	VMware
vRealize Log Insight (default)	6.0.8183621	VMware
NSX	3.2.1	VMware
vCD	1.1	VMware

2.2. Conception de l'architecture du DRaaS avec Veeam Cloud Connect

Veeam Cloud Connect est une fonctionnalité ajoutée à la solution de base de backup et de réplication de Veeam nommé Veeam Backup and Replication. Comme cité avant, cette solution répond au besoin DRaaS des clients possédant une infrastructure vSphere ou hyper-V.

L'autre solution de DRaaS est délégué un autre membre de mon équipe, pour cela on ne va pas la voir dans ce document.

Les composants utilisés dans Veeam Cloud Connect sont celle de toute solution de backup et/ou de réplication :

- Un Backup Server : pour gérer les opérations de backup/réplication.
- Un Backup Repository : C'est le Tenant Cloud Host dans ce contexte cloud, qui représente un ensemble de ressources alloué dans une organisation de VMware vCloud Director.
- Et un Cloud Gateway : Pour router les répliquions vers ces propres Tenant Cloud Host.

L'architecture peut être présentée comme suit :

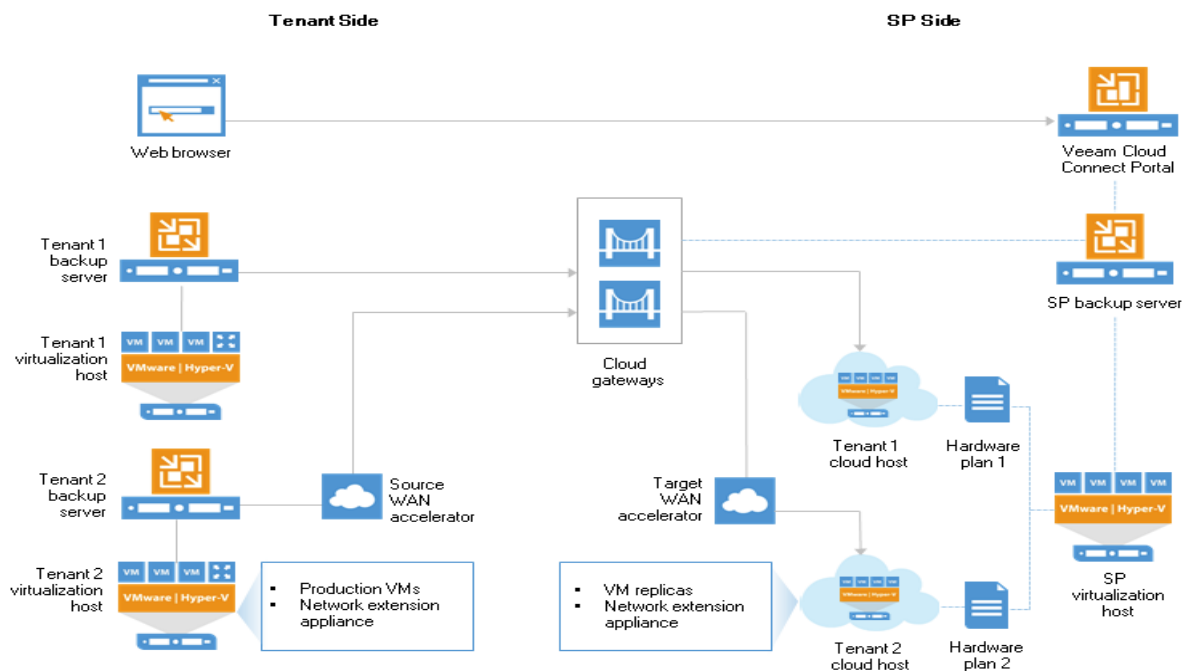


Figure 3 : Architecture de la solution DRaaS avec Veeam Cloud Connect

Dans chaque Tenant Cloud Host une Network extension appliance est créé automatiquement dans les deux sites (client et fournisseur de services) pour faire un tunnel entre eux à travers le Cloud Gateway, et un Plan Hardware pour planifier les opérations de répliquions.

La Network extension appliance présente dans options de basculement :

- Basculement complet : Fournit à la fois un accès à Internet pour les répliques et un accès à des répliques à partir d'Internet.
- Basculement partiel : Étend le réseau client au site du fournisseur de service, de sorte que les ordinateurs virtuels de production puissent communiquer avec les répliques.

Elle est utilisé pour séparer les réseaux, les VLAN locataires ne doivent pas être routés au niveau de la couche de commutateur physique, ils peuvent uniquement communiquer entre eux à l'aide des fonctionnalités de routage de l'appliance réseau. Pendant le basculement, le trafic va directement de Network Appliance à Internet.

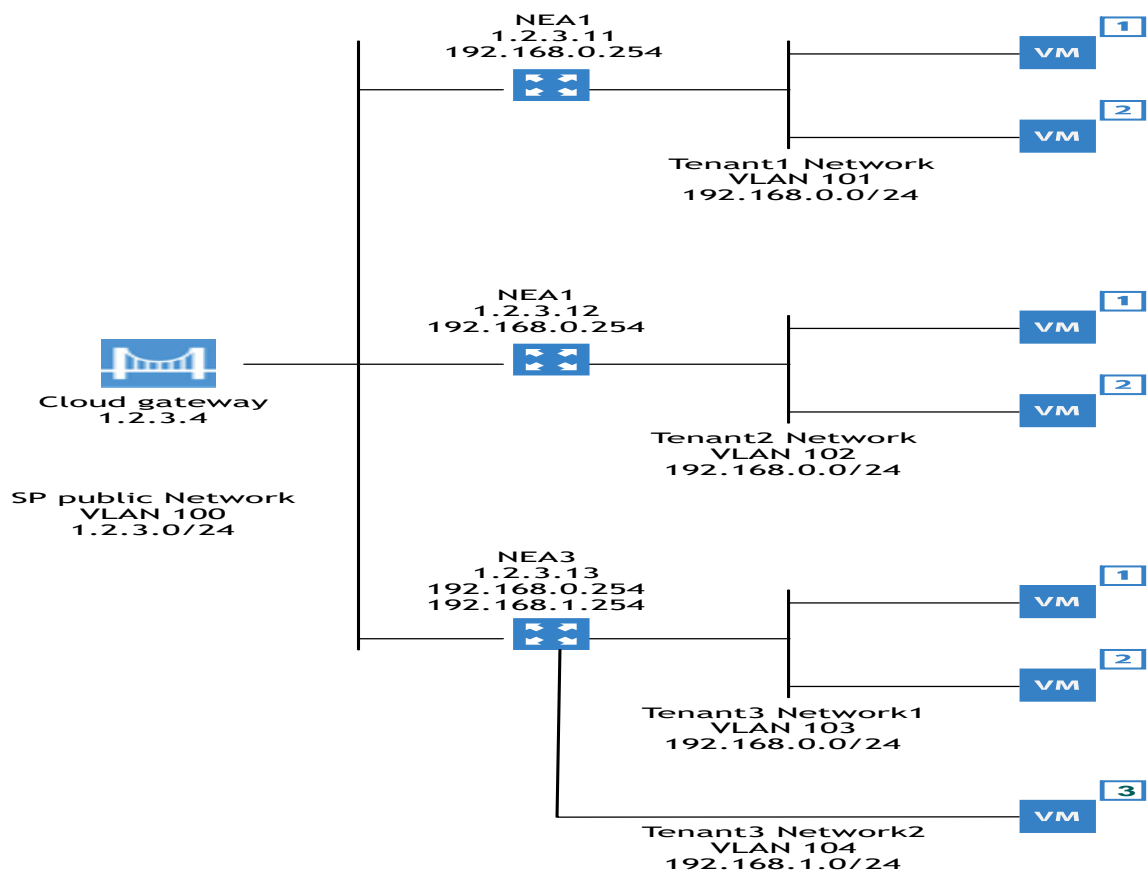


Figure 4 : Configurations réseaux avec Veeam Cloud Connect

Les clients peuvent gérer leurs plans de basculement à travers le portail web de Veeam Cloud Connect Portal installé et configurer avec le manager Veeam « Enterprise Manager ». Les locataires peuvent se connecter avec leurs identifiants d'abonnement créé par le fournisseur de service et démarrer leurs plans de basculement en cas de sinistre.

III. Implémentation des solutions

1. Définition de la configuration

Après avoir passé par l'étape de l'acquisition des compétences j'ai trouvé que le management de cloud n'est pas une simple configuration à faire, l'infrastructure doit être redéfinie selon les qualités de services demandés par les clients, des SLAs doivent être définies et une supervision proactive forte doit être implémentée pour garantir que les contrats de niveau de service ne sont pas violés. La supervision proactive intègre les machines virtuelles des clients et les composants de l'infrastructure IaaS répartis sur toutes les couches de virtualisations.

Quant à la solution de DRaaS, la préparation d'un environnement de test et la licence Cloud Connect sont encore retardés pour faire le POC.

2. Définition du management du cloud

Le mot "management" dans un SDDC (Software Defined Data Center, e.g. ce nouveau cloud VMware d'INWI) couvre tout le spectre d'activité nécessaire au bon fonctionnement de ce dernier. Dans certaines situations, nous devons utiliser un mot plus étroit, tel que surveillance, planification ou dépannage.

La gestion de la performance comporte deux sous-activités :

- Surveillance
- Dépannage

La gestion de la capacité comporte deux sous-activités :

- Planification
- Surveillance

La gestion de la configuration et de la disponibilité :

- Surveillance
- Dépannage

On peut les voir dans les deux couches suivantes :

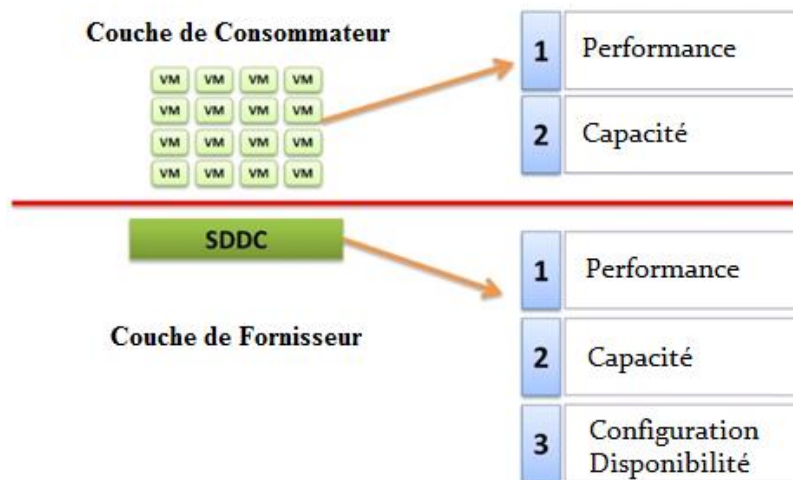


Figure 5 : Modèle de Management d'un IaaS

2.1. La couche consommateur

Voici les propriétés de la couche consommateur :

- vSphere est bien connu pour sa stabilité. Par conséquent, la disponibilité n'est pas un problème auquel nous devons faire face quotidiennement. En cas de défaillance, vSphere HA fera apparaître les machines virtuelles concernées en quelques minutes, probablement avant que le propriétaire de la machine virtuelle ne le remarque, si cela se produit en dehors des heures de pointe.
- La performance, d'autre part, doit être gérée quotidiennement et surveillée en quelques minutes. En fait, il s'agit généralement de la première source de désaccord entre l'équipe d'application et l'équipe Infrastructure. C'est pourquoi nous mettons un grand "numéro un" à côté de la performance. Comme nous le savons bien, un propriétaire de machine virtuelle ne se soucie que de sa machine virtuelle. Les autres ordinateurs virtuels de l'hôte ESXi ou du cluster vSphere ne sont pas pertinents. Lorsqu'un propriétaire de machine virtuelle se plaint de la lenteur des performances de sa machine virtuelle, pouvons-nous prouver en une minute à peine que notre système IaaS le sert bien ? Si nous ne pouvons pas, c'est un signe que nous n'avons pas de SLA de performance.
- Lorsque la performance est prise en compte, nous pouvons examiner la capacité. Il est inutile de discuter de la capacité lorsqu'il y a un problème de performance. C'est pourquoi on lui attribue le numéro deux.

- La capacité d'une machine virtuelle est une question de redimensionnement. En général, c'est la responsabilité de l'équipe d'application. En tant que fournisseur IaaS, nous pouvons ajouter de la valeur en recommandant la bonne taille.

2.2. La couche fournisseur

Voici les propriétés de la couche fournisseur :

- La couche IaaS est l'endroit où nous avons ou devrions avoir un contrôle et une visibilité complets. Si nous ne l'avons pas, nous devons le réparer, comme nos clients le supposent et l'attendent.
- À cette couche, nous nous soucions de tout le monde, pas seulement d'une machine virtuelle spécifique. Nous devons nous préoccuper à la fois de la machine virtuelle et de l'infrastructure, car nous voulons être assurés que l'infrastructure partagée n'est pas un goulot d'étranglement. La grande image est la clé ici. C'est à ce niveau que nous répondons à des questions telles que "Quel est l'état de santé de mon environnement VMware ?»
- Une bonne performance signifie que tout le monde est bien servi. La définition de "bien" est que le niveau de discordance pour une VM donnée est inférieur à ce que nous promettons officiellement.
- Une bonne capacité signifie que nous équilibrons nos achats de matériel. Ni une utilisation élevée ni une utilisation faible ne sont bonnes pour notre activité IaaS.
- La gestion de la configuration devient critique dans notre SDDC, car il est facile d'effectuer des modifications. Cependant, il est en deçà des performances et de la capacité en termes de priorité commerciale car il impacte moins directement nos clients.
- La gestion de la disponibilité est importante en raison d'un engagement excessif. Elle est moins critiques en termes de performances et de capacité, car nous avons généralement une redondance et sommes protégés par vSphere HA. Dans SDDC, la gestion de la disponibilité doit couvrir l'infrastructure VM. Par exemple, nous devons assurer la disponibilité de tous nos ordinateurs virtuels NSX Edge et Trend Micro, car ils se trouvent dans le chemin de données.

2.3. SLA de la performance

Nous utilisons deux niveaux de service car cela constitue un bon équilibre pour notre petit environnement.

Les contrats de niveau de service performants étant liés à la qualité de la conception, nous devons définir clairement les différents différenciateurs de conception.

Voici quelques points forts du design INWI :

Tableau 2 : Design vSphere pour les SLAs définie

Niveau de service	CPU	RAM	Stockage
1 (élevé)	Pas de sur-provisionnement	Pas de sur-provisionnement	- Provisionnement Tick - Disques Full Flash
2	3 fois sur-provisionnée	2 fois sur-provisionnée	- Provisionnement Thin. - Disques Hybride

Le niveau de service 2 est encore en phase d'implémentation.

Parlons maintenant des contrats de niveau de service relatifs aux performances que nous avons passés dans le contrat avec nos clients. Nous devons définir le service pour chacun des quatre composants de l'infrastructure (CPU, RAM, disque et réseau). Pour chacun, nous spécifions la valeur réelle et la métrique à suivre.

Tableau 3 : Les SLAs de l'IaaS

Niveau de service	CPU	RAM	Network	Stockage
1 (élevé)	<1% Contention de CPU	0% Contention de RAM	0 paquet perdu	10 ms de latence
2	<13% Contention de CPU	10% Contention de RAM	0 paquet perdu	30 ms de latence

Le taux de consolidation n'est pas pertinent car il ne prend pas en compte les conflits.

2.4. La disponibilité

Pour le nœud de calcul, le tableau du diagramme suivant fournit un exemple de conception de cluster pour les niveaux de machine virtuelle. L'essentiel est que nous ayons besoin d'une conception qui définisse clairement notre service :

Tableau 4 : Définition de la disponibilité pour les SLAs définie

Niveau de service	vSphere HA	Taille maximale du cluster	Nb Max des VM
1 (élevé)	2 Hôtes de HA (N+2)	8 nœuds	15 VM par hôte 100 VM par cluster
2	1 Hôte de HA (N+1)	16 nœuds	60 VM par hôte 750 VM par cluster

Regardons maintenant la partie de stockage :

Tableau 5 : Définition de la disponibilité pour les SLAs définie (suite)

Niveau de service	Nb de copies	Tolérance aux pannes de disque	Snapshot
1 (élevé)	2 copies	2 disques en panne	Toutes les 2 heures
2	1 copie	1 disque en panne	à la demande

2.5. Supervision du niveau 1 de service

Pour la partie calcul de la gestion de la capacité, la surveillance devient beaucoup plus simple, car il n'y a pas de sur-engagement.

Nous avons juste besoin de vérifier notre SLA de disponibilité.

Pour nous aider à surveiller, nous pouvons créer cinq graphiques linéaires, trois pour le calcul et deux pour le stockage. Voici les graphiques en courbes pour le calcul :

- Un graphique en courbes indiquant le nombre total de vCPU restantes dans le cluster.
- Un graphique en courbes indiquant le nombre total de vRAM restant dans le cluster.
- Un graphique en courbes indiquant le nombre total de machines virtuelles restantes dans le cluster.

Regardez les trois graphiques précédents en tant que groupe. Prenez celui avec le plus petit nombre.

Contrairement au calcul, le stockage dépend des performances et de l'utilisation de tous les niveaux de service. Voici les graphiques en courbes pour le stockage :

- Un graphique linéaire montrant la latence de stockage maximale et moyenne de toute machine virtuelle du cluster de banque de données.

- Un graphique linéaire montrant la capacité de disque restante dans le cluster de banque de données.

Le stockage 100% flash devient une option courante en 2016. Avec les disques SSD, même ceux de grande capacité, il est possible que nous soyons liés par l'espace disque et non par les performances. Le deuxième graphique linéaire nous aide à suivre l'espace disponible (en Go). Nous utilisons le stockage distribué VSAN, cela modifie la gestion de la capacité, car nous ne pouvons pas effectuer le calcul et le stockage différemment.

Puisque notre stockage fait partie de nos calculs, nous devons les considérer comme une seule et même unité. Nous examinons donc les cinq graphiques linéaires ensemble.

Ajouter du compute signifie ajouter du stockage. Il est vrai que l'ajout de stockage ne signifie pas l'ajout du compute si nous avons des emplacements physiques disponibles. Cependant, cela nécessite une planification préalable et le choix intentionnel d'un modèle de serveur avec suffisamment d'emplacements de disque. Pour VSAN, il n'y a qu'un seul datastore, un cluster de datastores n'est donc pas nécessaire.

Si le nombre le plus bas approche un nombre peu élevé, il est temps d'ajouter de la capacité. À quel point ce nombre devrait être bas dépend de la rapidité avec laquelle nous pouvons nous procurer du matériel.

2.5.1. Niveau 1 de service - partie calcul – CPU

Regardons le graphique de première ligne. Le nombre de vCPU restantes est essentiellement l'offre et la demande. L'offre et la demande peuvent être définies comme suit :

- L'offre = (Le nombre total de cœurs physiques de tous les hôtes ESXi) - (mémoire tampon HA)
- La demande = nombre total de vCPU pour toutes les machines virtuelles

2.5.2. Niveau 1 de service - partie calcul - RAM

Nous avons couvert le processeur. La RAM suit une logique similaire. La quantité de vRAM restante est également basée sur l'offre et la demande, où :

- L'offre = (Mémoire RAM physique totale de tous les hôtes ESXi) - (mémoire tampon HA)
- La demande = La vRAM totale pour toutes les machines virtuelles

Il n'est pas nécessaire d'inclure la RAM ESXi vmkernel car elle est suffisamment petite pour être ignorée. En outre, pratiquement aucune machine virtuelle n'utilise sa RAM configurée et nous avons le partage de page transparent. Ces deux raisons expliquent également pourquoi nous ne voyons pas la nécessité d'ajouter un seuil.

Pour le vSAN, il faut laisser au moins 32 Go de RAM libre sur les hôtes pour sa bonne fonctionnement, car :

- VSAN réduit sa consommation de mémoire lorsque 32 hôtes de moins.
- VSAN consomme de la mémoire supplémentaire lorsque le nombre de nœuds du cluster est supérieur à 16.
- Tous les Flash VSAN consommation des ressources de mémoire supplémentaires par rapport aux configurations hybrides.

2.5.3. Calcul de niveau 1 de service – Les VMs

Le troisième graphique linéaire à afficher indique le nombre total de machines virtuelles restantes dans le cluster. Bien que les deux premiers graphiques linéaires proviennent de la stratégie de performance, ceux de la stratégie de disponibilité.

Normalement, la disponibilité est associée à l'hôte ESXi. Par exemple, nous conservons le nombre maximal de machines virtuelles dans un hôte ESXi donné. La logique est que si un ESXi tombe en panne, nous limitons le nombre de machines virtuelles affectées.

Nous appliquons la politique de disponibilité au niveau du cluster car cela a plus de sens. L'application au niveau de l'hôte ESXi est moins pertinente en raison de la haute disponibilité. Oui, la probabilité qu'un hôte tombe en panne est supérieure à celle de l'ensemble du cluster. Toutefois, HA va redémarrer les ordinateurs virtuels et les propriétaires d'ordinateurs virtuels peuvent ne pas le remarquer. D'un autre côté, si un cluster tombe en panne, c'est un problème majeur.

L'offre et la demande peuvent être définies comme suit :

- L'offre = Le nombre maximal de machines virtuelles autorisées dans un cluster (en tenant compte du tampon haute disponibilité).
- La demande = Le nombre total de VM dans le cluster

La formule suppose que la taille de notre cluster de niveau 1 ne varie pas. Cette hypothèse est juste car nous devrions garder les choses cohérentes dans le niveau 1. Si nous avons plusieurs tailles de cluster (par exemple, 8, 10 ou 12), nous avons alors un super métrique par taille.

2.5.4. Niveau 1 de service - résumé

Nous avons récapitulé que nous devons produire trois graphiques linéaires. Examinez les graphiques linéaires des trois sous-sections précédentes en tant que groupe. Prenez celui avec le nombre le plus bas comme facteur déterminant pour le manque de capacité.

Nous avons l'offre et la demande. Nous savons quand le nombre atteindra zéro.

La question est : quand achèterons-nous du nouveau matériel ?

Comme il s'agit du niveau 1 et que nous promettons des spécifications identiques, cela pourrait signifier un nouveau cluster. L'achat d'un cluster peut être un engagement important, il doit donc être planifié correctement.

2.6. Supervision de niveau 2 de service

Pour la partie de calcul, le niveau 2 est plus complexe que le niveau 1, car il existe maintenant une possibilité de conflit. Nous devons vérifier à la fois les contrats de niveau de service et de disponibilité.

Voici les graphiques en courbes pour le calcul :

- Un graphique linéaire montrant les contentions de CPU maximum et moyen rencontrés par n'importe quelle machine virtuelle du cluster.
- Un graphique en courbes illustrant les contentions de RAM maximum et moyen rencontrés par toute machine virtuelle du cluster.
- Un graphique en courbes indiquant le nombre total de machines virtuelles restantes dans le cluster.

Voici comment nous utilisons les lignes maximales et moyennes pour suivre la capacité :

- Le nombre maximum doit être inférieur au SLA promis à tout moment. Si le nombre maximal de latences atteint notre seuil, il est temps d'ajouter de la capacité. L'ajout d'une machine virtuelle va enfreindre notre contrat de niveau de performance.
- Si le maximum est inférieur à notre SLA mais que notre nombre minimal approche de notre seuil, la capacité est presque complète. L'ajout de machines virtuelles peut constituer une violation de notre contrat de niveau de performance, le nombre maximal pouvant le dépasser.
- Si le maximum est bien inférieur à notre SLA, cela signifie que nous avons surapprovisionné le matériel.

Regardez ces trois graphiques en tant que groupe. Prenez celui qui a la valeur la plus proche de notre SLA. Les deux premières lignes proviennent de notre contrat de niveau de service Performance et la troisième ligne de notre contrat de niveau de disponibilité.

Pour la partie stockage, c'est la même chose, car elle dépend des performances et de l'utilisation de tous les niveaux de service mentionnés précédemment.

Pour surveiller la partie réseau, nous pouvons créer les éléments suivants :

- Un graphique en courbes indiquant le nombre maximal de paquets rejetés par le réseau au niveau du centre de données physique. Il est appliqué au centre de données physique et non au cluster vSphere, car ils partagent éventuellement les mêmes commutateurs centraux.
- Un graphique en courbes montrant les valeurs maximales et moyenne d'ESXi vmnic au même niveau que dans le point précédent. Cela nous dira si l'un de nos vmnics est saturé.

Et maintenant je suis dans l'étape de la configuration de ces différents SLAs et niveaux de services sur la solution de virtualisation vSphere et sur la solution de management du cloud VMware vRealize Operations manager.