Externalisation ou mutualisation : Quels choix pour les infrastructures ?

Mathieu Molinéris
DOSI / Aix-Marseille Université
Avenue Escadrille Normandie Niemen
13397 Marseille

Yves Azamberti
DOSI / Aix-Marseille Université
Avenue Escadrille Normandie Niemen
13397 Marseille

Résumé

Externaliser une partie ou l'intégralité de son infrastructure peut sembler être une solution efficiente pour répondre à des besoins en constante évolution et garantir une qualité de service maitrisée.

L'autre voie consiste à mobiliser l'ensemble des ressources en interne afin de proposer idéalement un service équivalent en misant principalement sur une rationalisation des infrastructures.

En effet, si nous souhaitons mutualiser nos ressources en interne, nous devons à minima être en capacité de pouvoir proposer une infrastructure hautement tolérante à la panne, fortement et facilement extensible et ceci avec une forte maitrise technique, organisationnelle et financière.

Dans ce cas, ce sont les compétences internes et la volonté de tout un chacun qui pourraient s'avérer déterminantes.

C'est ce choix éminemment stratégique que nous vous proposons de présenter à travers l'exemple de la construction de notre infrastructure Datacenter menée lors des quelques mois qui ont précédé la fusion d'Aix-Marseille Université. De plus, nous présenterons un inventaire des résultats obtenus deux ans après, ainsi que des futures pistes de travail afin d'améliorer et d'étoffer l'offre de service proposée.

Enfin, nous conclurons sur les avantages et les inconvénients d'un tel choix.

Mots-clefs

Infrastructures, Datacenter, virtualisation, mutualisation, PCA, PRA

1 Introduction

L'université d'Aix-Marseille (AMU) a été créée le 1er janvier 2012. Dès la fin de l'année 2010, la fusion de nos trois établissements nous a été annoncée avec une lettre de mission assez floue et somme toute relativement limitée : « Etre en capacité de payer les agents et les fournisseurs du nouvel établissement ».

A cette date, seuls trois éléments étaient parfaitement clairs :

- Il paraissait parfaitement illusoire que les services nécessaires aux besoins de l'établissement soient aussi limités ;
- Les infrastructures informatiques existantes étaient soit en fin de vie, soit en non capacité de répondre aux seuls besoins du SI du futur établissement ;
- Il nous restait moins de 11 mois afin de trouver une solution permettant de s'adapter à des besoins qui s'annonçaient pléthoriques vu la taille de la future université.

2 Un groupe de travail dédié

La volonté des trois directions (DSII, DOSI, DSI) des trois établissements (Université de Provence, de la Méditerranée, Paul Cézanne) était de s'assurer qu'un groupe de travail composé de personnels issus de leur service puisse prendre en charge l'étude de la construction du futur Datacenter.

Ce groupe de travail s'est composé de cinq personnes volontaires dont parmi elles un animateur.

Chacun exposa ses propres architectures, ses points forts, ses points faibles ainsi que les services qu'il aurait apprécié d'y voir développer si le temps l'avait permis.

Nous nous sommes rapidement aperçus que les compétences et les savoir-faire au sein du groupe étaient réunis et que la volonté de chacun était de construire une infrastructure maitrisée.

Et que dans ce cas précis, le nombre de personnes impliqué était suffisant pour prendre en charge ce travail.

Le mot « maitrisé » dans notre cas, devait être une ligne de conduite commune sur l'ensemble des éléments que nous allions mettre en œuvre : l'infrastructure, les architectures, les technologies, les investissements financiers et le calendrier serré.

C'est dans ce contexte que nous nous sommes mis au travail

2.1 Inventaire

Dans un souci d'économie, le but était de recenser ce qui pouvait être réutilisé sans toutefois mettre en péril l'existant et sans pour autant hypothéquer la future architecture.

2.1.1 Quelle salle machine avec quels services ?

Le point sur lequel nous étions tous unanimes, est que nous devions être en mesure de redémarrer l'intégralité de la production sur un autre site en cas de sinistre grave.

Dans ce cas, nous étions dans l'obligation d'avoir à minima deux salles machines dont les infrastructures devaient être robustes et maitrisées.

Nous avons donc établi une liste de critères pertinents pondérés par leur niveau de criticité pour guider notre choix :

Critère	**Coef**
Sécurité incendie de la salle	3
Réseau	
Sécurité de la pénétrante	2
Redondance physique	1
Disponibilité fibres	2
Refroidissement du local serveur	3
Courant secouru	3
Accès salle machine (manutention)	1
Sécurité d'accès à la salle machine (Badge, alarme intrusion, accès limité, intervention)	3
Sécurité sanitaire	2
Hébergement du personnel (bureau, clim, chauffage, etc…)	1
Sans possibilité d'évolution	
Qualité courant EDF	2
Sécurité environnementale :	
Incendie	3
Inondation	3
Perturbations (malveillance, grève, etc. ..)	3
Surface au sol	1

Figure 1 - Critères pour le choix des salles machines

Comme on peut le voir, nous avons initialement privilégié dans notre réflexion les infrastructures physiques comme le refroidissement ou le courant secouru, vis à vis du réseau que nous étions en capacité de faire évoluer.
En effet, en cas de coupure réseau, seul l'accès aux services est interrompu. Lorsque le réseau est rétabli, l'accès au service est immédiat. En cas de problème électrique ou de température, les conséquences sont tout autres, que ce soit en termes de reprise d'activité ou d'intégrité des données.

Nous avons utilisé cette méthode pour effectuer un audit interne des salles machines de nos établissements respectifs.
Au total, 7 salles ont été passées en revue et notées selon nos critères de choix.
Au final, cette étude à démontrée que deux salles répondaient au niveau de service optimal attendu. Leurs caractéristiques spécifiques étaient que les groupes froids, les onduleurs et les groupes électrogènes étaient tous opérés par les personnels des services informatiques.
Nous avons donc concentré tous nos efforts sur la consolidation de ces deux salles, respectivement de 300 et 50 m2, qui sont devenues le socle de notre infrastructure Datacenter.
Bilan de l'opération : zéro euro d'investissement. Nous pouvions désormais compter sur une infrastructure totalement maitrisée par la future DOSI, condition indispensable à la construction du Datacenter.

Figure 2 - Une partie de l'infrastructure DOSI AMU

On peut visualiser sur la photo de gauche (site de St-Jérôme) deux groupes électrogènes en arrière plan. Le jaune (250 Kva) secoure l'intégralité des baies informatiques, le bleu (165 Kva) l'intégralité des cinq groupes froids. Au premier plan, un équipement de production d'eau glacée. Sur la photo de droite (site du Pharo), on y observe le groupe électrogène (165 kva) qui secoure l'intégralité de l'infrastructure de la salle plus modeste en superficie.

2.1.2 La grande aventure de l'achat dans un contexte de pré-fusion

La seconde partie de ce processus d'inventaire était naturellement de savoir comment nous allions pouvoir acquérir les matériels nécessaires au déploiement de notre architecture système et réseau ainsi que les moyens financiers mis à notre disposition.
La complexité annoncée des modalités d'achats comme les conventions de financement et le groupement d'achat interuniversitaire, fut toutefois partiellement gommée par la volonté des trois directions et des trois gouvernances de faciliter la mise en œuvre de la nouvelle infrastructure.
Nous avons utilisé les marchés existants et opérationnels des trois établissements lorsqu'ils convenaient à nos besoins. Toutefois chaque commande devait être montée avec une convention de financement avec l'accord des trois présidents. Cette méthode a eu pour conséquence l'allongement significatif des temps d'acquisition. Il est toutefois remarquable que malgré ce contexte, les différents acteurs ont largement contribué à réduire ce temps de traitement.
Seul subsistait la problématique du choix et de l'acquisition des matériels de stockage qui s'annonçaient stratégiques.
Cette situation fut l'occasion d'engager les évaluations des solutions proposées par les différents acteurs du marché. Nous avons mis au banc d'essai l'ensemble de ces matériels avec nos propres jeux d'essai. Nos critères nous ont permis de mettre en adéquation nos besoins avec leur capacité d'extensibilité aussi bien en termes de volumétrie que de performance (Bande passante, Iops[1] / temps de réponse).
Suite à cette évaluation de deux mois, Nous avons mis en œuvre un appel d'offre en groupement d'achat porté par les trois établissements.
La solution de stockage retenue après ce long processus nous permet de pouvoir évoluer, avec les mêmes contrôleurs jusqu'à une volumétrie de 1,5 Po répliqués et 200 000 Iops. A titre comparatif, la messagerie de l'établissement (plus de 100 000 comptes à ce jour) génère à elle seule des pics de prêt de 2000 Iops.
Cette solution efficiente intègre nativement, en plus des fonctionnalités classiques, les toutes dernières technologies comme: le Tiering, la virtualisation de blocs, la migration à chaud inter-baie des volumes...
La technologie de Tiering est un procédé adaptatif capable d'organiser les données de manière pertinente afin de maximiser les performances et d'optimiser l'utilisation des unités de stockage. Les données (au niveau bloc) sont écrites ou modifiées sur les disques les plus performants et elles transitent sur des disques capacitifs beaucoup moins onéreux si elles ne sont plus ou faiblement accédées. Auparavant, une augmentation de quota généralisée (1Go à 3 Go) de la messagerie était synonyme d'achat couteux. Désormais, on procède uniquement à l'achat de disques capacitifs à des tarifs très abordables.

[1] Input/Outpout Operation per seconds

2.2 Quelles solutions pour quels services dans notre Datacenter ?

Dès lors que nous nous étions assurés de la qualité de nos infrastructures ainsi que leur capacité à évoluer à qualité constante, nous pouvions commencer à concevoir l'architecture réseau et système qui nous permettrait de répondre aux futurs besoins.
Idéalement, nous souhaitions que cette architecture cible nous apporte les services suivants :

- Déploiement rapide des ressources système (serveurs et stockage) ;
- Evolution rapide tant en terme de capacité que de performances ;
- Sécurité et redondance maximale des éléments physiques et logiques (réseau et système) ;
- PRA[2] ;
- Un PCA[3] serait un plus.

Afin de répondre à ces spécificités, le choix de la généralisation de la virtualisation s'est naturellement imposé.
Cette compétence indispensable et présente au sein du groupe de travail a permis d'élaborer plusieurs concepts d'architecture intégrant les contraintes système et réseau indissociables.
Dans le détail, l'ensemble de ces compétences a aussi induit certains de nos choix techniques. Par exemple, le déploiement du stockage généralisé en iSCSI[4]. L'expertise réseau présente au sein du groupe de travail nous permettait d'entrevoir tout le potentiel du protocole iSCSI déployé au travers de nos architectures réseau.
La convergence de nos réseaux de données était alors un choix évident en cette fin d'année 2010.
De plus, notre maitrise du réseau métropolitain RAIMU[5] allait grandement nous simplifier cette mise en œuvre.
Afin de répartir la charge au plus juste et de permettre un plan de reprise d'activité en cas de sinistre plus efficace, nous avons choisi de déployer notre Datacenter en mode actif-actif réparti sur les deux sites précédemment choisis.

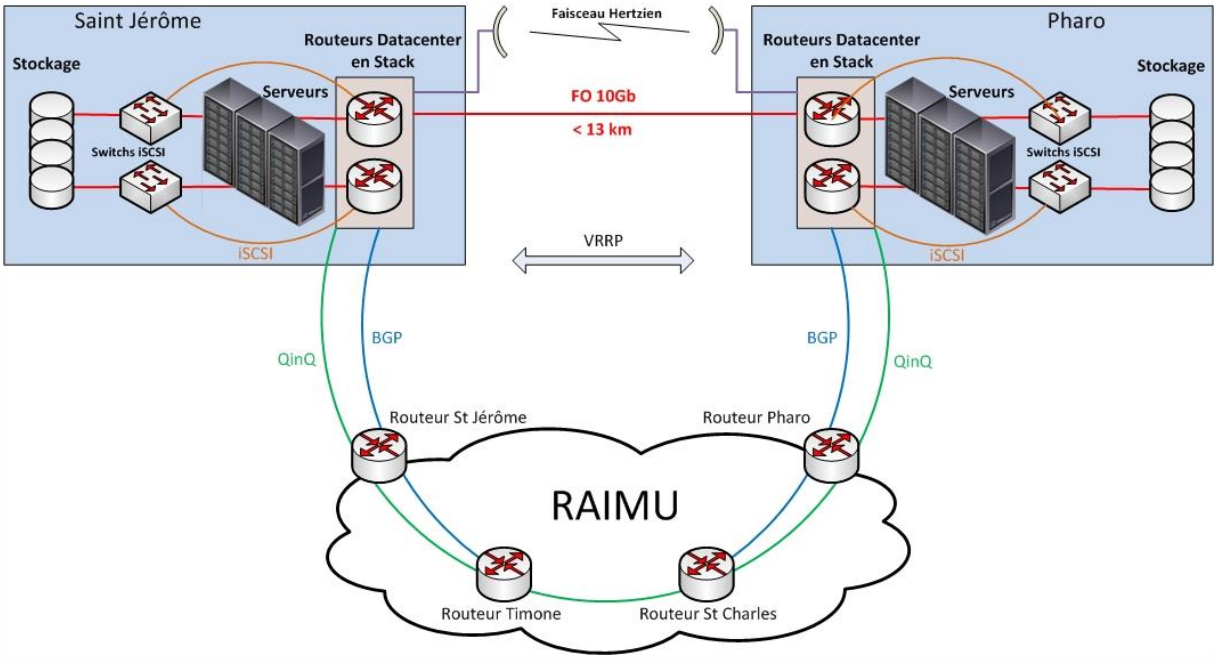

Figure 3 - Schéma de l'architecture Datacenter DOSI AMU

Pour ce faire, les réseaux sont annoncés des 2 cotés en BGP[6] sur RAIMU certains avec un poids fort sur le premier site et plus faible sur le second et vice-versa pour les autres en jouant sur l'attribut BGP MED[7].

[2] Plan de reprise d'activité
[3] Plan de continuité d'activité
[4] Internet Small Computer System Interface
[5] **R**éseau **AI**x **M**arseille **U**niversité
[6] Border Gateway Protocol

Pour renforcer la tolérance de panne, les peerings BGP se font sur deux routeurs différents en cœur de RAIMU.
Les passerelles sont redondées via le protocole VRRP[8]. La convergence de VRRP utilise BFD[9] comme mécanisme de détection de pannes. Afin de renforcer cette sécurité, l'ensemble des routeurs du backbone est doublé et "stacké".
Tous les réseaux sont propagés en niveau 2 d'un site à l'autre par une fibre noire que nous éclairons à 10Gb et sur laquelle les « jumbo frame » sont autorisés pour les besoins des vlans iSCSI.
La fibre noire est secourue par un lien « QinQ[10] » à 1Gb au travers de RAIMU.

Dans un souci de sécurisation de l'infrastructure fibre en cas de sinistre majeur sur l'ensemble des brins optiques, nous avons intégré un faisceau hertzien (sept kilomètres à vol d'oiseau) d'un débit théorique de 300Mb/s mais compte tenu de la distance et du milieu urbain très bruité, le débit mesuré est de 40Mb/s. Ce pont qui nous permet de conserver la maitrise des opérations d'exploitation de manière globale a été intégralement mis en œuvre par le groupe de travail : étude, configuration et pose.

3 De la conception à la réalisation

Si la conception de notre future infrastructure nous laissait à certains moments un peu de répit, sa réalisation a demandé de la part du groupe de travail un investissement total sur plusieurs mois.
En effet, l'ensemble des opérations de déploiement a été réalisé en interne.
Nous avons toutefois fait appel à un minimum de service externe : 1 journée pour la validation d'une partie spécifique de la configuration réseau et cinq jours pour la mise en œuvre de la solution de stockage qui était l'élément, à cet instant, le moins maitrisé.
Au mois de septembre de l'année 2011, l'ensemble de notre réseau et de notre solution de virtualisation étaient opérationnels. Nous avons donc pu nous concentrer sur l'intégration des solutions de stockage. Le travail effectué en amont nous a permis de tirer profit de cette prestation qui s'est déroulée comme un échange constructif afin d'obtenir le maximum que pouvait nous apporter la solution en terme de fonctionnalités. Dans ce contexte, nous avons pu mettre en œuvre un PCA qui n'était pas au départ une fonctionnalité considérée comme indispensable.
Afin d'accentuer notre maitrise, nous avions prévu dans le cadre de l'acquisition des matériels de stockage un transfert de compétence. De même, une formation complète sur notre solution de virtualisation a été dispensée à toute l'équipe.

Il est à noter, que d'autres groupes de travail DOSI œuvraient aussi durant l'année 2011, autour des thèmes de l'authentification, des applications financières et de scolarité.
Ceux-ci avaient un besoin immédiat de ressources système et ne pouvaient attendre le 1er janvier 2012.
Nous avons déployé des serveurs et du stockage associé pour le nouvel établissement sur une infrastructure existante provisoire au tout début de l'année 2011. La migration vers le nouveau Datacenter à la fin du mois de décembre 2011 a été grandement facilitée grâce à la somme des choix effectués (virtualisation, convergence des réseaux de données) et à leur totale maitrise.

[7] Multi Exit Discriminator
[8] Virtual Router Redundancy Protocol
[9] Bidirectional Forwading Detection
[10] 802.1ad

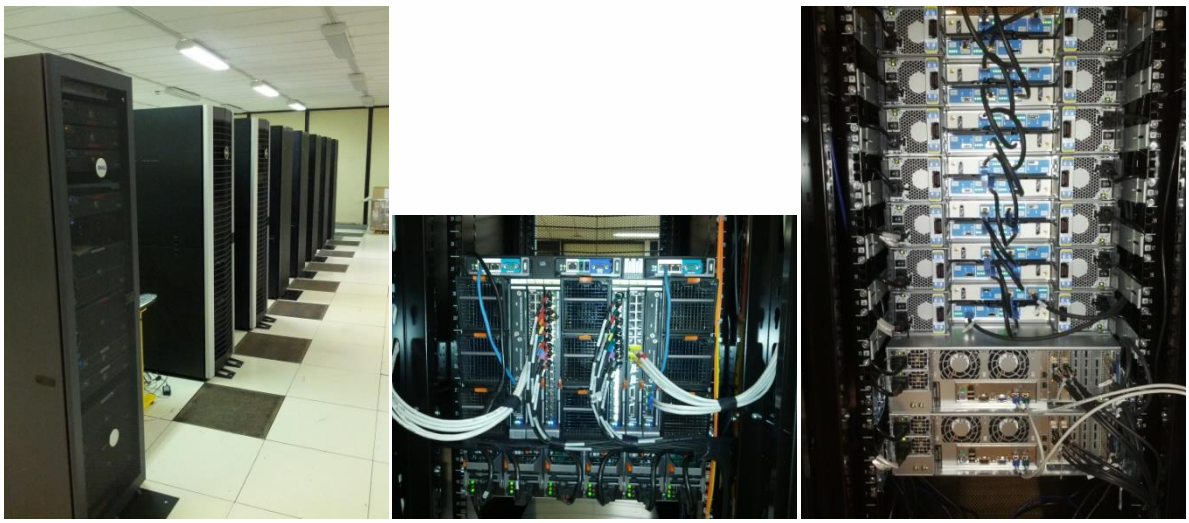

Figure 4 - Matériels mis en œuvre dans le Datacenter DOSI AMU

4 L'infrastructure mise au banc d'essai

Le groupe de travail a pleinement réussi son objectif de mettre à disposition au premier jour de l'existence d'AMU sa propre infrastructure. Mais le choix de la déployer totalement en interne et reposant sur les compétences mutualisées ne peut être validé définitivement que si on évalue ses capacités à répondre à la demande et aux éventuels incidents par la suite.

4.1 Scénarios d'incidents et résultats 2012/2013

Nous avions anticipé différents scénarios auxquels devait faire face l'infrastructure.
Nous vous présentons ici le résultat de l'exploitation :
En rouge le nombre d'événements produits en 2012 et en vert le nombre d'heures d'arrêt évitées.

- Scénario 1 : Coupure électrique réseau EDF: Pas d'arrêt (11 ; 49) ;
- Scénario 2 : Défaut courant ondulée : Pas d'arrêt (1 ; 4) ;
- Scénario 3 : Défaut sur annonces RAIMU : Bascule automatique en moins de 20 s. (5 ; 360) ;
- Scénario 4 : Défaut sur un équipement réseau (Ethernet ou iSCSI) de l'infrastructure : Pas d'arrêt (1 ; 4) ;
- Scénario 6 : Arrêt électrique prévu ou maintenance globale d'un des deux sites : bascule en production des serveurs notifiés PCA et des données associées sur l'autre site (la majorité). (2 ; 14).

Le scénario 6 est extrêmement intéressant du point de vue technique mais aussi en termes de continuité de service. A ce jour peu d'offres d'infrastructure externalisée sont capables de proposer un tel service quel que soit le type d'applicatif. Et si cela est disponible, le tarif reste prohibitif.

Il est à noter qu'à ce jour nous avons rencontré un seul problème ayant perturbé la production. Il a eu pour conséquence de créer des problèmes d'accès aux services de messagerie durant une période d'un mois. A l'origine, un effet de latence causé par une combinaison d'optimisations non effectuées en cours de production et induites par les multiples urgences qui ont diminué la vigilance des équipes en charge de l'infrastructure.

4.2 Des besoins en constante progression

Immédiatement après l'officialisation de création d'Aix-Marseille université, une explosion des besoins en hébergement est apparue notamment par la mise en place des outils communs nécessaires au fonctionnement dont les plus significatifs sont:

- La messagerie et ses services associés pour 110 000 utilisateurs ;
- La plateforme d'enseignement en ligne pour près de 100 000 utilisateurs ;
- Les serveurs de fichiers pour plus de 7000 utilisateurs potentiels ;
- L'ensemble des applications de gestion (Apogée, Sifac, …) ;
- La ferme RDS[11] centralisant en grande partie les accès aux applications métiers ;
- Les services WEB intégrés AMU et les services d'hébergement mutualisés ou dédiés ;
- L'ENT et son architecture répartie;
- Les référentiels d'authentification (LDAP[12], AD[13], Radius[14]) …

Les capacités d'évolutions du Datacenter sont telles que certaines entités de la communauté de l'enseignement supérieur et de la recherche ont souhaité intégrer nos infrastructures comme le Cancéropôle PACA (2 VM et 10 To) qui a fait ce choix après une étude comparative de notre offre avec des offres d'hébergement privées. Nous avons mis en place des conventions d'hébergement avec ces partenaires externes.

La capacité d'évolution des infrastructures physiques d'un des 2 sites sont telles que nous proposons aussi des conventions d'hébergement physique. Ainsi le Mésocentre d'Aix-Marseille Université (14 TFlops) et 3 autres clusters de calculs sont hébergés dans cette salle machines.

A ce jour, l'infrastructure DOSI AMU représente 200 To de volumétrie dynamique et 200 To secondaire (déporté sur un local tiers dédié à la sauvegarde). Plus de 200 serveurs virtuels ont été déployés. Les demandes continuent d'affluer et nous sommes capables d'y répondre sereinement sans hypothéquer l'existant et tout en maitrisant les coûts.

[11] Remote Desktop Services
[12] Lightweight Directory Access Protocol
[13] Active Directory
[14] Remote Authentication Dial-In User Service

5 Evolutions

A très court terme, nous allons proposer un catalogue de services afin d'afficher clairement l'offre d'hébergement, soit physique en nombre de U, soit de ressources système complètes et de faciliter leur mise à disposition. L'évolution de notre réseau métropolitain en cours apportera au Datacenter une connectivité à 10Gbits. Sa technologie MPLS[15] de bout en bout permettra de remplacer le lien de secours QinQ par du L2VPN[16] à 10 Gbits avec une MTU[17] à 9000 inclus. La prise en compte des communautés BGP apportera plus de souplesse dans la gestion des réseaux ip. Une amélioration du temps de convergence BGP sera obtenue par l'ajout d'un peering iBGP entre les 2 sites, l'utilisation de BFD et l'optimisation des timers BGP.

Le renforcement de la sécurité va s'opérer au travers de l'intégration de pare-feux applicatifs de dernière génération qui permettront d'identifier les applications voire même les prioriser, d'identifier les utilisateurs et d'inspecter le flux en temps réel.

Sur le long terme, nous aurons à mener une étude orientée cloud privé de type IAAS[18] autour de la solution OpenStack, ainsi qu'une étude de déploiement d'une solution VDI[19] en mutualisant les compétences de la DOSI permettant de virtualiser les postes de travail de certains sites.

6 Conclusion

Nous démontrons ici, que la mutualisation des compétences et des moyens de façon générale, permet d'obtenir des niveaux de service opérationnels efficaces, évolutifs et à coût maitrisé par l'économie effectuée sur les prestations de service. A ce jour, cinq personnes issues du pôle système et du pôle réseau travaillent conjointement sur le suivi et l'évolution de l'architecture du Datacenter. En condition d'exploitation standard ceci représente environ 1 ETP[20] et en mode projet (évolution, conception), on peut atteindre 2,5 ETP.

De plus, nous sommes dans un cercle vertueux qui garantit un niveau de motivation élevé du fait de la confiance établit vis à vis de la direction et de la gouvernance qui permet à chacun de mettre en œuvre l'ensemble de ses compétences et de ses savoir-faire tout en les renforçant. Les personnes se sentent valorisées, ce qui garantit leur implication.
L'ensemble des moyens de sécurisation et de redondance des infrastructures ainsi que la mise en place d'une organisation des personnels basée uniquement sur la simple continuité de service permet d'assurer un très haut niveau de disponibilité sans astreintes.

D'autre part, dans le contexte de la fusion, cette approche basée sur la maitrise de l'infrastructure et sur la compétence a permis l'intégration rapide d'autres personnels dans ce projet commun et dans la nouvelle organisation.
En définitive, nous sommes capables de proposer des services de manière plus efficaces tout en construisant notre équipe.

[15] MultiProtocol Label Switching
[16] Layer 2 Virtual Private Network
[17] Maximum Transmission Unit
[18] Infrastructure As A Service
[19] Virtual Desktop Infrastructure
[20] Equivalent Temps Plein