Mutualisation de machines HPC singulières dans Grid'5000

Pierre Neyron, Lucas Nussbaum

Mutualisation de machines HPC singulières dans Grid'5000

Pierre Neyron, Lucas Nussbaum

2016-03

1 Contexte

Plusieurs équipes de recherche travaillant sur la thématique HPC possèdent des machines expérimentales singulières : des machines hautement multicoeurs (par exemple : plus de 64 coeurs), des machines hybrides disposant d'une configuration matérielle atypique (par exemple : 8 GPU), ou encore des machines équipées de processeurs de nouvelle génération voire de nouvelle architecture (par exemple : ARM64).

En parallèle, Grid'5000 (`http://www.grid5000.fr`) propose une plate-forme nationale qui met à disposition de la communauté scientifique des infrastructures de type grappe de serveurs pour l'expérimentation scientifique en informatique distribuée, en particulier pour les domaines du cloud computing, du big data, du réseau et du HPC. Les points forts de Grid'5000 sont:

- une gestion administrative (politique d'accès, comptes, etc) unifiée au niveau national
- un envergure de plate-forme permettant des expérimentions sur des infrastructures dimensionnantes (permettant la validation du passage à l'échelle)
- une mécanique de reconfiguration (pile logicielle, réseau, stockage) et de contrôle/monitoring (API) de l'environnement expérimental très avancée
- une mutualisation des efforts d'ingénierie au niveau national

Dans ce contexte, un rapprochement de machines expérimentales d'équipes des recherche et de Grid'5000 à été entrepris en 2012 sous forme de prototype opérationnel à l'intérieur du site Grid'5000 de Grenoble. Il accueille les machines singulières des équipes LIG/Inria GRA DataMove et Polaris (ex Mescal et Moais). D'un point de vue local, ce projet s'intègre dans la plate-forme Digitalis (`http://digitalis.imag.fr`).

Ce document recense les besoins constatés et propose des pistes d'évolution pour aller au-delà de la situation actuelle.

Il s'inscrit dans le contexte grenoblois, mais vise à explorer la possibilité d'un modèle pouvant s'appliquer à d'autres sites.

Ce document est un document interne pour la gouvernance Grid'5000, mais pourra être une base pour un document diffusé plus largement.

2 Rappels

2.1 Les machines "singulières" de Digitalis, à Grenoble

Digitalis héberge aujourd'hui 7 machines singulières en plus des clusters locaux Grimages, Kinovis et Ppol

- idfreeze: machine quad-CPU AMD, 48 coeurs (2011, ~20K€)
- idgraf: machine bi-CPU Intel, équipée de 8 GPU (2011, ~25K€)
- idphix: machine équipée d'un accélerateur Intel Xeon Phi (2013, ~10K€)
- idbool: machine 12-CPU AMD interconnectés par la technologie Numascale de 192 coeurs (2014, ~35K€)
- idarm-1&2: 2 machines équipé de CPU ARM64 bits, ARM Juno development boards (2015, 2x ~5K€)
- idkat: machines quad-CPU Intel 48 coeurs, pouvant être équipée de 4 accélérateurs (GPU, Xeon Phi, etc) (2015, ~20K€ + coût des accélérateurs -> ~35K€)
- idcin-1&2: 2 machines bi-CPU, 28 coeurs equipés de 3 GPU Nvidia chacune (2015, 2x ~12K€ + coût des accélérateurs)

Voir la page `http://digitalis.imag.fr/index.php/Hardware` pour plus de détails.

2.2 Définition

Outre sa caractérisation par sa spécificité technologique, une machine HPC singulière est une machine de qui n'a pas vocation (au temps t) à être achetée en nombre conséquent. Elle se suffit à elle même en terme de support expérimental : pas de nécessité sur le plan de l'expérimentation scientifique d'une interconnexion particulière avec un ensemble d'autres machines (sauf cas particulier, ponctuel). Une machine singulière peut être prise en considération dans le cadre de cette étude dès lors que son coût (d'achat et/ou d'opération) ou sa rareté justifie à encourager une ouverture large de son accès à la communauté scientifique du domaine. En ce qui concerne le choix d'une mutualisation *dans Grid'5000*, la faisabilité technique et la pertinence vis-à-vis des objectifs scientifiques et/ou politique doit également être prise en compte.

2.3 Objectifs

- Fédérer largement l'accès aux machines HPC expérimentales singulières, pour permettre facilement à des utilisateurs d'affiliations (et de localisations) diverses d'accéder aux machines
- Rationaliser les coûts d'opération des machines (passer d'une administration manuelle et dupliquée pour chaque machine/site, très coûteuse en temps ingénieur, à une administration unifiée et mutualisée)

- Améliorer la qualité du support expérimental (contrôle des systèmes, traçabilité, reproductibilité, etc)
- Ne pas trop contrevenir à la facilité d'utilisation (réservation, accès aux machines, etc)
- Ne pas trop réduire le champ d'expérimentation (adaptation facile de l'environnement matériel ou logiciel aux besoins des chercheurs)
- Ne pas trop alourdir l'interface administrative proposée aux utilisateurs (gestion des comptes, emails)

3 Synthèse des besoins remontés depuis le lancement du projet (2012)

Attention : les points remontés ci-dessous ne sont pas nécessairement tous cohérents entre eux. Certains peuvent ne pas être réalisables ou pertinents dans le contexte d'une mutualisation effective, suivant le niveau de mutualisation / de compromis visé. L'énumération aura cependant le mérite de permettre au lecteur d'entrevoir un certain nombre des questions soulevées pour une mutualisation dans Grid'5000.

3.1 Gestion de compte allégée et adaptée aux usages locaux

- Être capable de gérer un accès et un usage spécifique pour un sous-ensemble de machines (typiquement des machines singulières) : définition de groupes d'utilisateurs et possibilité de coupler ceux-ci avec la gestion de resources et de jobs
- Alléger les procédures administratives : définir la notion de responsable de groupe d'utilisateurs qui peut intervenir de manière autonome pour gérer ses utilisateurs (pas d'interaction avec un tiers responsable à un niveau supérieur pour les opérations courantes)
- Supporter une gestion de compte groupée : inscriptions groupées pour gérer facilement une classe d'étudiants (pour un TP ou pour l'année scolaire), notamment pour accéder aux machines singulières locales
- Rapport utilisateur optionnel en fonction du groupe d'utilisateurs
- Possibilité d'avoir des comptes utilisateurs liés à des expériences pour des tests automatisés (Jenkins)

3.2 Amélioration de la pertinence des médiums de communication

- Permettre une réception des emails Grid'5000 pertinente avec l'usage de la plate-forme : typiquement uniquement les emails concernant les services de base de la plate-forme et l'usage des machines singulières

- Adapter le site web Grid'5000 pour permettre un aiguillage facile vers les informations pertinentes pour les machines singulières, et leur documentation spécifique

3.3 Gestion des ressources adaptée pour les particularités des machines singulières

- Gestion spécifique pour les machines singulières
  - Différencier nettement clusters et machines singulières dans la gestion de ressources : mélanger les deux peut aboutir à des usages non pertinents ou non justifiés en plus de nuire à la lisibilité de la plateforme
  - Pouvoir garder certaines machines hors gestionnaire de ressources, suivant les desiderata du propriétaire
  - Pouvoir utiliser les machines sans passer par des jobs, au moins en journée
  - Avoir un accès interactif / sans attente en journée, multi-utilisateurs (job en temps partagé)
- Prise en comptes de 2 usages pour les machines :
  - Un usage pour le *développement* en journée (voire la nuit) avec un fonctionnement très proche d'une station de travail : accès immédiat en utilisant simplement ssh, édition de code directement sur la machine et compilation locale très fréquente, etc.
  - Un usage pour l'*expérimentation* la nuit, avec accès exclusif (et éventuellement kadeploy), via des réservations faites à l'avance
- Gestion de la priorité d'accès pour les propriétaires/utilisateurs locaux des machines
  - Avoir un accès prioritaire sans restriction pour les propriétaires/utilisateurs locaux
  - Permettre aux propriétaires/utilisateurs locaux de prendre la main arbitrairement sur les machines en cas de deadlines proches, et donc éventuellement de supprimer les jobs d'utilisateurs extérieurs
  - Pouvoir réserver des ressources pour un groupe d'utilisateurs (typiquement les utilisateurs locaux)
  - Limiter la durée des jobs pour les utilisateurs extérieurs
- Déploiement non activé sur certaines machines
  - Prendre en compte que certaines machines ne supportent techniquement pas Kadeploy
  - Permettre de n'activer le déploiement que la nuit par exemple
- Gestion dynamique de la durée des jobs
  - Pouvoir créer un job avec un calcul automatique de sa durée, par exemple pour terminer à la fin de la journée de travail
  - Pouvoir prolonger un job (sans interruption d'exécution pour les processus courants)

3.4 Adaptation du système pour l'expérimentation

- Mise à disposition rapide des dernières versions des logiciels dans l'environnement par défaut :
  - gcc/openmp, llvm, librairies HPC, cuda, driver Nvidia, Intel mpss, icc, vtune, ...
  - les versions souhaitées peuvent devoir être installées à partir des sources car pas encore packagées, voire encore en béta
- Possibilité pour l'utilisateur d'exécuter des actions avec les privilèges root dans l'environnement par défaut (sans déployer) :
  - Paramétrage du système (schedtool, récupération de compteurs matériels avec liwkid, perf, ...)
  - Paramétrage du matériel (nvidia-smi, (dés)activation de l'hyperthreading/c-states/p-states. turboboost, ...)
- Offrir un système de nettoyage systématique de l'environnement expérimental pour éviter des effets de bord dus à des reliques d'expériences précédentes
- Permettre à l'utilisateur d'avoir un traçabilité des évolutions de son environnement d'expérimentation, et de pouvoir reproduire des expériences ultérieurement tout en permettant une évolution du système d'exploitation (installation de nouvelles versions de logiciels et effets de bord associés)
- Possibilité pour l'utilisateur de prendre la main complètement sur la machine avec Kadeploy ou équivalent :
  - pour ajouter/modifier des logiciels sans contrainte vis-à-vis des autres utilisateurs
  - pour modifier l'OS librement (voire à l'extrême : le *casser*) sans risque et sans crainte de compromettre l'usage ultérieur de la machine
- Alléger le système de reconfiguration/reinstallation de machine
  - Rendre plus facile de la mise au point et la persistance du système pour une machine unique (kadeploy est moins pertinent pour une machine unique que pour un cluster)
- Permettre de changer arbitrairement la configuration hardware au cours de la vie des machines (ex: changement de carte PCI-E, GPU) pour les besoins locaux (propriétaires)
- Permettre le fonctionnement de certaines machines avec un OS spécifique (RedHat pour Xeon Phi/stack Intel, Ubuntu/noyau patché pour Numascale, OpenEmbedded pour des plates-formes SoC)
- Supporter des architectures matérielles hétéroclites, par exemple : ARM64, MPPA, OpenPower, FPGA, ...
- Permettre à l'administrateur local d'installer des packages dans l'environnement par défaut très rapidement (moins de 5 minutes)

- Faciliter l'accès pour les machines depuis et vers l'extérieur
  - Pas de proxy, accès direct à Internet
  - Limitation du nombre des rebonds SSH nécessaires pour se connecter sur les machines
  - Accès réseau *de proximité* pour privilégier un meilleur débit et une meilleure latence (pour X11 par exemple)
- Offrir une interface utilisateur de qualité sur la base de ligne de commandes

4 Propositions

Ce paragraphe propose deux modèles pour rapprocher Digitalis de Grid5000. Pour faciliter l'étude des intérêts et inconvénients de ces propositions, nous considérons les différents acteurs suivants :

- la structure d'hébergement (Inria, université, laboratoire, etc) prenant en charge les coûts d'opération (personnels, fluides), et éventuellement mais pas toujours le financement de l'achat de machines
- l'équipe technique locale (ingénieur(s) système local(locaux) en charge des opérations sur les machines)
- les propriétaires (acheteurs) de machines (pour une même structure d'hébergement, il est possible d'avoir différents propriétaires ne partageant pas les mêmes avis sur le mode d'opération des machines)
- les utilisateurs locaux des machines (qui ne sont pas nécessairement des utilisateurs de Grid'5000)
- la gouvernance de Grid'5000
- l'équipe technique Grid'5000
- les utilisateurs Grid'5000

Il faut noter que chaque propriétaire (acheteur) de machines tient un rôle prédominant sur le choix du mode d'opération des machines qu'il achète. Bien qu'un financement vienne éventuellement d'une structure (Inria, Université, etc), les machines sont généralement achetées dans le contexte de projets (ou contrats) pour lesquels le propriétaire est le décideur. De ce fait, la volonté de mutualisation est moins évidente que lorsque la structures est effectivement prédominante sur les choix (dans le cadre d'achats d'envergure nationale par exemple). Par ailleurs, le propriétaire est également souvent l'utilisateur local prédominant.

4.1 Proposition 1 : développer la notion de site satellite dans lequel les machines sont gérées localement et sous l'autorité des propriétaires

4.1.1 Principes de fonctionnement :

- On reste très proche du fonctionnement actuel de Digitalis (voir `http://digitalis.imag.fr/index.php/usage`)

- Le site satellite n'utilise qu'un sous-ensemble des services et ne propose qu'un sous-ensemble des fonctionnalités de Grid'5000
  - Gestion de comptes
  - Mediums de communication
  - Infrastructure réseau nationale
  - Interconnexion au réseau rapide du site si pertinent
  - Services réseaux (accès à la plate-forme, DNS, ...)
  - Stockage (home directories et autres hébergements de données scientifiques)
  - Hébergement de machines de services
  - Infrastructure de monitoring
  - Souches d'environnement kadeploy
- Pas d'ouverture large à l'ensemble de la communauté d'utilisateur de Grid'5000, mais au cas par cas par une demande aux propriétaires
- Pas d'ingérence de la politique globale Grid'5000 sur les choix fait localement pour le site satellite, hormis pour les services de base utilisés
  - Définition d'une gouvernance locale avec des acteurs/rôles bien identifiés pour l'interaction avec Grid'5000
- Aspects locaux du site satellite opérés par l'équipe technique locale
  - le choix de fournir des fonctionnalités proches de Grid'5000 (gestion de ressources évoluée, redéploiement, etc) est laissé aux propriétaires et à équipe technique locale

4.1.2 Intérêts par rapport à une gestion locale non mutualisée :

- Pour la structures d'hébergement et l'équipe technique locale :
  - Mutualisation des coûts pour les services de base, s'appuyant sur une infrastructure et des services existants
  - L'équipe technique peut s'investir sur des problématiques plus avancées
- Pour les utilisateurs locaux :
  - Comptes utilisateurs communs avec Grid'5000 (accès aux machines Grid'5000 avec le même compte, partage du home directory) et d'autres sites satellites
  - Réseau informatique national facilitant l'interconnexion pour une expérience occasionnelle entre plusieurs sites
  - Environnement d'expérimentation se rapprochant éventuellement de Grid'5000 (capitalisation sur la connaissance des outils communs)
- Pour les propriétaires :
  - Ceux en temps qu'utilisateurs locaux principaux +

  - Ceux relatifs à l'hébergement +
  - Obtenir plus facilement les financements du fait qu'il mutualise au moins une partie des coûts associés à ses machines expérimentales

- Pour la gouvernance de Grid'5000 :
  - Élargissement de la communauté supportée par Grid'5000, et ouverture sur des usages plus larges et plus souples que ceux proposés dans le modèle de base de Grid'5000
  - Développement d'une collaboration plus proche avec l'équipe technique locale, bénéfique pour l'accueil de jeunes ingénieurs Grid'5000 sur le site, pour les tâches que l'équipe technique Grid'5000 ne peut pas exécuter à distance, ou sur des compétences spécifiques. Favoritation de projets de développement communs entre local et Grid'5000
  - Apport d'un regard externe plus proche sur la plate-forme Grid'5000, bénéfique pour son évolution. Meilleure connaissance de Grid'5000 sur les sites locaux

- Pour l'équipe technique Grid'5000 :
  - Idem gouvernance Grid'5000

- Pour les utilisateurs de Grid'5000 :
  - Réutilisation du compte Grid'5000 et des services de base pour l'accès à des machines singulières, suivant l'acceptation au cas par cas par le propriétaire (contact privilégié/partenariat)

4.1.3 Inconvénients par rapport à une gestion locale non mutualisée :

- Pour les utilisateurs locaux :
  - Complexité accrue pour les services de base (accès ssh avec rebond), ou lourdeur (par exemple : service homes NFS surchargé à cause de l'usage Grid'5000)
  - Incompatibilité éventuelle de la politique de sécurité de Grid'5000 avec des contraintes de confidentialité et d'accès pour des partenaires industriels
  - Impact des maintenances inhérente à une grosse plate-forme mutualisée sur l'utilisation des machines locales

- Pour les propriétaires :
  - Ceux en temps qu'utilisateurs locaux principaux +
  - Alourdissement des procédures administratives (gestion de compte, relation avec la gouvernance Grid'5000)

- Pour l'équipe technique locale :
  - Besoin de négocier, de se conformer aux choix et d'attendre l'implémentation par l'équipe technique Grid'5000 pour toute demande de changement dans les services de base

- Pour l'équipe technique de Grid'5000 :
  - Besoin de gérer dans la configuration des services de la plate-forme des machines qui ne sont pas complètement intégrées dans Grid'5000, et de prendre en compte ces machines lors des maintenances
  - Besoin de gérer des utilisateurs qui ne sont pas des utilisateurs Grid'5000 (qui peuvent reporter des incidents ou demander de l'aide sans clarifier que leur problème ne concerne que les machines du site satellite)
  - Besoin d'interagir avec des ingénieurs externes à l'équipe

4.1.4 Évolutions/développements induits pour Grid'5000 :

1. Évolutions requises :
   - Adapter les procédures d'administration de Grid'5000 :
     - Pour prendre en compte les machines du sites satellites qui n'utilisent pas tous les services et qui ne sont pas gérées par l'équipe technique Grid'5000
     - Pour interagir avec une équipe technique locale et la gouvernance associée
   - Améliorer la gestion des comptes :
     - Pour permettre la gestion de groupes d'utilisateurs pour l'accès à certaines machines sous l'autorité d'un responsable
       * L'adhésion au groupe est contrôlée par le responsable (il peut valider ou refuser l'adhésion)
       * Le responsable peut effectuer les actions sur le groupe de manière autonome (sans devoir faire intervenir un tiers)
       * Les informations d'appartenance d'un utilisateur à un groupe sont disponibles pour un couplage avec les systèmes (OAR, PAM, etc)
     - Pour permettre une gestion plus fine des mailing lists Grid'5000 (par exemple : une mailing list pour les services commums, une mailing list pour le coeur de Grid'5000, une mailing list par site satellite)
2. Évolutions fortement souhaitables :
   - Adaptation du site web pour permettre un aiguillage facile vers les informations pertinentes pour le site satellite
   - Par ailleurs, pour que cette proposition 1 soit équilibrée, il semble indispensable qu'une publicité large soit faite sur les machines du site satellite (pas juste via du bouche à oreille), par exemple en ajoutant à la fin du tutoriel Getting Started :

     > Grid'5000 also hosts a few "satellite machines", that share some aspects with the rest of Grid'5000, but are not entirely part of Grid'5000. Those machines are generally large multi-core/manycore nodes, or otherwise experimental hardware.

Usage policy and access conditions are specific to each machine, but you are welcomed to ask for access if needed for your research. For more information, see Grenoble:Digitalis, Bordeaux:Dalton, etc.

- Simplifier l'accès aux machines et depuis les machines : limiter le besoin de rebonds ssh, accès facile à Internet (direct)

3. Évolutions souhaitables :
   - Il faudrait quand même travailler:
     - à une vue synthétique des machines disponibles, idéalement en les intégrant dans la Reference API (mais sans la vérification avec g5k-checks)
     - à une vue synthétique de leur niveau d'intégration dans Grid'5000, et des points bloquants (pour les documenter), par exemple sur le modèle de la page `https://www.grid5000.fr/mediawiki/index.php/Tasks-January-2013#Services_.26_features_deployment_status`

4.2 Proposition 2 : offrir le support pour une intégration complète dans Grid'5000 de machines HPC singulières sélectionnées

4.2.1 Principes de fonctionnement

- Cette proposition repose sur les idées suivantes :
  - L'intégration effective dans Grid'5000 nécessite et vise la capacité de reconfiguration et contrôle de la machine
  - Il ne faut pas confondre *prise en charge par Grid'5000* et *mutualisation* : il peut y avoir d'autres plates-formes d'accueil pour mutualiser des machines, par exemple les mésocentres comme CIMENT à Grenoble (`https://ciment.ujf-grenoble.fr`) ou le Centre Blaise Pascal à Lyon (`http://www.cbp.ens-lyon.fr/doku.php?id=developpement:productions:plateaux`)
  - Toute machine expérimentale HPC singulière n'a pas vocation à entrer dans le modèle Grid'5000 (reconfiguration et contrôle bas niveau, etc) : d'autres modèles de fonctionnement, mutualisés ou non, peuvent être pertinents pour des utilisateurs, typiquement pour des expériences sur les couches logicielles hautes uniquement
  - Certaines machines sont techniquement incompatibles avec le modèle de fonctionnement Grid'5000
- La prise en charge par Grid'5000 se fait sous conditions :
  - Toutes les machines fonctionnent avec les mêmes règles: pas d'exception au cas par cas
  - Signature d'un accord explicite avec les propriétaire/utilisateurs locaux sur le mode de fonctionnement

  - Absence de notion de priorité arbitraire pour les propriétaires/utilisateurs locaux : l'acheteur "donne" la machine à Grid'5000 (à la communauté)

- Grid'5000 définit et propose les éléments suivants :
  - Des critères d'accueil de machine : toutes les machines ne sont pas acceptées, elles doivent être compatibles avec le mode d'opération Grid'5000
    * En amont, ces critères sont utilisés par l'acheteur dans le cahier des charges pour l'acquisition de machine
    * En aval, la question de l'acceptation de la machine est soumise à la gouvernance de Grid'5000, après étude du cas par l'équipe technique
  - Des règles d'usage et une charte faisant autorité pour l'utilisation des machine, quel que soit l'utilisateur (acheteur ou autre)

- La gestion de ressources proposée par Grid'5000 est adaptée pour des machines singulière (notamment accès concurrent multi-utilisateur / en temps partagé, etc)

- Le spectre d'intervention de l'équipe technique Grid'5000 est élargi pour supporter les machines singulières (uniques et diverses).

- La structure locale/les propriétaires contribuent à l'effort national sur Grid'5000, en allouant ou en finançant du temps d'ingénieur pour Grid'5000 sur la base de l'équivalent du temps d'ingénieur qui serait nécessaire pour opérer la machine avec un même niveau de service dans une plate-forme locale.

4.2.2 Intérêts par rapport à une gestion locale non mutualisée :

- Pour les utilisateurs locaux :
  - Utilisation des machines bien intégrée dans Grid'5000, uniformisation, facilité de passage d'une machine à une autre au sein de Grid'5000
  - Bénéfice de services typiques de Grid'5000 permettant contrôle et reconfiguration, qui sont souvent trop chers à mettre en place sur une plate-forme locale

- Pour la structure d'hébergement :
  - Augmentation du niveau de mutualisation et rationalisation des coûts au niveau national
  - Amélioration du service en beneficiant de l'expertise Grid'5000
  - Augmentation de la pertinence du travail de l'équipe technique locale : bénéficiant des services communs de la plate-forme nationale, l'équipe technique locale contribue au projet national et peut viser des objectifs de plus grande envergure

- Pour les propriétaires :

  - Ceux en temps qu'utilisateurs locaux principaux +
  - Prise en charge complète de la gestion des machine par Grid'5000 (dès la procédure d'achat ?), bénéficiant de services rodés et de qualité

- Pour la gouvernance de Grid'5000 :
  - Support d'une communauté scientifique plus importante, notamment en offrant une plate-forme expérimentale plus intéressante pour le domaine HPC ou pour les travaux nécessitant des machines atypique reconfigurables
  - Renforcement de l'équipe technique avec les ingénieurs missionés par la structures locales
  - Augmentation des compétences de l'équipe technique grâce à l'intégration d'ingénieurs possédant des expertises "locales"

- Pour l'équipe technique Grid'5000:
  - Idem gouvernance Grid'5000

- Pour les utilisateurs Grid'5000 :
  - Accès facile à des machines singulières et reconfigurables

4.2.3 Inconvénients par rapport à une gestion locale non mutualisée :

- Pour les utilisateurs locaux:
  - Perte de flexibilité (sur les évolutions, qui suivent le rythme des évolutions Grid'5000 ; sur la configuration matérielle des machines)
  - Accès en concurrence avec beaucoup plus d'utilisateurs, notamment des utilisateurs exterieurs : il ne suffit plus de traverser un couloir pour régler un conflit d'usage
  - Charge plus forte sur les machines (besoin d'une règulation plus stricte)
  - Non uniformité du parc des machines locales si du coup certaines machines ne peuvent pas être dans Grid'5000

- Pour les propriétaires :
  - Ceux en temps qu'utilisateurs locaux principaux +
  - Perte du prestige/de l'exclusivité de posséder une machine singulière (pour la publication, ou pour la négociation de collaborations)
  - Doit se contraindre aux décisions de la gouvernance Grid'5000
  - Limitation du choix de machines expérimentales possibles (ou nécessité de trouver un autre fonctionnement pour des machines non prises en charge par Grid'5000)

- Pour l'équipe technique de Grid'5000 :
  - Diversification du parc des machines à administrer (qui se complexifie)

- Besoin de supporter un mode de fonctionnement non-cluster (accès partagé en journée, etc)

- Pour la structure d'hébergement :
  - Nécessité de maintenir un effort local s'il y a des machines non compatibles/acceptées pour être prises en charge par Grid'5000

4.2.4 Évolutions/développements induits pour Grid'5000

1. Évolutions requises :
   - Trouver une séparation claire entre les machines singulières et les clusters pour éviter des utilisations non pertinentes (pour ref. cluster *borderline* à Bordeaux en 2007-2008)
   - Mettre en place l'accès en temps partagé (dans OAR et dans tous les outils associés)
   - Fournir un document donnant les conditions d'acceptation (à posteriori) d'une machine dans Grid'5000 (au niveau matériel).
   - Fournir la charte d'usage des machines singulières pour que l'acheteur local puisse accepter en connaissance de cause de *donner* sa machine à Grid'5000, et ne soit pas insatisfait par la suite
   - Adapter le site web Grid'5000 pour permettre un aiguillage facile vers les informations sur les machines singulières/non cluster
   - Veiller au bon fonctionnement de l'équipe technique pour la coopération avec de nouveaux ingénieurs, typiquement ingénieurs permanent travaillant à temps partiel sur la plate-forme
2. Évolutions fortement souhaitables :
   - Fournir une canevas pour la rédaction d'un cahier des charges par un acheteur local visant une prise en charge par Grid'5000
   - Support d'architectures non x86-64 (boot, environnements, etc)
   - Permettre aux utilisateurs de changer la configuration bas niveau mais être capable de garantir une configuration matérielle stable (cas des changements résistants au reboot/redeploiement, par exemple l'activation ECC pour la mémoire d'un GPU)
   - Mise en place d'autorisations pour `sudo` dans l'environnement par défaut pour les commandes comme `schedtool`, la récupération de compteurs matériels avec `liwkid` ou `perf`, `nvidia-smi`, l'activation de l'hyperthreading/c-states (éventuellement en mode exclusif uniquement, suivant la commande)
   - Amélioration des environnements utilisateur pour faciliter l'expérimentation HPC (éventuellement compilation à la main pour fournir des versions spécifiques; Modules d'environnement, Nix)
3. Évolutions souhaitables :
   - Permettre dans la charte d'usage et la gestion de ressources du supporter une priorité d'accès pour des groupes d'utilisateurs financeurs d'une machine

- Développer un système de déploiement plus léger que kadeploy pour les machines singulières

5 Conclusion

Suite à l'expérience au sein de la plate-forme Digitalis à Grenoble pour la mutualisation dans Grid'5000 de machines HPC singulières, ce document fait 2 propositions pour une évolution de Grid'5000 permettant d'accueillir plus formellement de telles machines. Ces 2 propositions ont pour point commun un rapprochement entre Grid'5000 et des plates-formes expérimentales existantes ou futures, mais avec 2 idées orthogonales : la première proposition repose en effet sur une fédération large autour d'une plate-forme Grid'5000 fournissant des services de base, alors que la seconde vise à renforcer les mécanismes intrinsèques de Grid'5000 pour le contrôle et la reconfiguration de système, pour supporter un nouveau type de machines. S'il serait bien sûr souhaitable de trouver l'intersection de ces 2 propositions, il ne faut cependant par oublier que la complexité de la mutualisation tient autant à l'humain qu'à la technique, rendant le problème très sensible. En particulier, il ne faut pas négliger l'impact de l'augmentation de la distance entre les utilisateurs sur leur coopération, en plus de l'augmentation de leur nombre. Typiquement il faut garder à l'esprit qu'on ne peut pas appliquer simplement une recette fonctionnant pour un partage de ressources au sein d'un petit laboratoire par exemple, pour une mutualisation dans une plate-forme nationale.