

CCTP

Appel d'offre passé selon la procédure de dialogue compétitif
d'après l'article 67 du code des marchés publics

Version Finale

Université de Franche-Comté

1 Objet du marché

L'université de Franche-Comté veut se doter d'un mésocentre de calcul, c'est-à-dire un centre offrant des ressources de calcul scientifique et de stockage avec une puissance intermédiaire entre les centres de calcul nationaux (CINES, IDRIS) et les ressources locales aux laboratoires. Ce mésocentre servira les composantes universitaires de la région, UFC, UTBM et ENSMM, et des industriels. Le but de ce centre est également de développer au niveau régional des compétences en calcul hautes performances pour aider les utilisateurs à tirer profit du matériel mis à disposition. Actuellement il n'existe aucune structure commune de calcul intensif. Les besoins de calcul sont donc satisfaits soit grâce à des clusters de taille réduite, dans les laboratoires, soit en accédant aux centres de calcul nationaux.

Pour équiper ce mésocentre, l'université souhaite acquérir un cluster d'ordinateurs qui sera mis à disposition de ses personnels et de ses partenaires. L'utilisation de ce cluster sera principalement orientée vers le calcul hautes performances. Les variantes et les besoins ont été affinés au cours de la procédure de dialogue compétitif que le pouvoir adjudicateur a lancé conformément à l'article 67 du code des marchés publics.

Le cahier des charges précise les caractéristiques matérielles et fonctionnelles attendues. Il laisse latitude aux entreprises de proposer des variantes quant à la puissance de calcul, la capacité de mémoire et de stockage.

2 Cahier des charges

Dans la suite nous donnons une description des différents éléments composant le cluster.

2.1 Caractéristiques générales

Les éléments clés de la configuration attendue sont la puissance de calcul, les ressources mémoire et de stockage, la performance des outils de développement et d'analyse des programmes et les bibliothèques de calcul scientifique et parallèle. Pour permettre une mise en exploitation rapide les solutions attendues doivent proposer une offre complète intégrant la climatisation, le matériel et les principaux logiciels qui seront utilisés.

Le domaine d'utilisation du cluster sera très varié (physique moléculaire, astronomie, mécanique, nano sciences, sciences humaines, ...). Pour répondre à ces besoins, l'offre devra donc tenir compte de la diversité des domaines et se montrer robuste dans l'ensemble de ces domaines. Une solution avec des nœuds ayant des caractéristiques matérielles différentes pourrait s'avérer nécessaire. L'offre inclura donc une version de base avec du matériel homogène et des propositions, sous la forme d'options, de variations sur les caractéristiques matérielles.

La configuration proposée s'appuiera soit sur une technologie de serveurs en lame soit en rack. Elle proposera un minimum de 64 nœuds et disposera d'un réseau d'interconnexion rapide permettant des échanges entre nœuds avec un débit minimal de 20 Gb/s et une latence faible. Elle disposera d'un espace de stockage d'un minimum de 20 To utiles, géré par un nœud spécifique, sera servie par un nœud d'administration et accédée par un ou plusieurs nœuds de login.

2.2 Nœud d'administration

Le nœud d'administration supporte les logiciels permettant la gestion du cluster. La configuration minimale requise pour ce nœud est la suivante :

- au moins un processeur de type identique à celui des nœuds de calcul,
- mémoire centrale d'une capacité d'au moins 6 Go,
- une capacité disque supérieure à 200 Go utiles, gérée en RAID,
- deux cartes réseaux 1000base/T connectique RJ45 pour la connexion vers le cluster et vers le réseau externe,
- châssis rackable.

2.3 Nœud(s) de login

Le nœud de login permet l'accès, la compilation et le test pour les utilisateurs. Si la configuration du cluster est hétérogène en terme de processeur, le nœud de login doit être doublé pour permettre d'avoir, sur ces nœuds, un environnement de compilation et de test équivalent à celui des différents types de processeurs de calcul. La configuration minimale requise pour chacun des nœuds est la suivante :

- deux processeurs 4 cœurs capables d'exécuter les programmes de tests, identiques aux processeurs des nœuds de calcul,
- au moins 24 Go de mémoire centrale,
- une capacité disque supérieure à 200 Go utiles, gérée en RAID,
- un accès de réseau rapide d'une bande passante minimale de 20 Gb/s et d'une latence faible,
- deux cartes réseaux 1000base/T connectique RJ45 pour la connexion vers le cluster et vers le réseau externe,

- contrôleur graphique,
- châssis rackable.

2.4 Nœuds de calcul

La configuration minimale requise pour chacun des nœuds de calcul est la suivante :

- deux processeurs 4 cœurs, capables d'exécuter les programmes de test,
- au moins 16 nœuds possédant 24 Go de mémoire centrale, 12 Go de mémoire centrale minimum pour les autres, des slots disponibles.
- un accès réseau rapide d'une bande passante minimale de 20 Gb/s et d'une latence faible,
- un connecteur réseau 1000base/T supportant le protocole PXE (Preboot eXecution Environment),

Tout le matériel pour mettre la machine – les nœuds de service et nœuds de calcul – sous forme de rack (armoires et kit de mise en rack) ou de lame ainsi que l'alimentation devront être inclus dans l'offre.

2.5 Réseau

Certains des calculs sont parallèles et nécessitent de ce fait un débit de communication important. L'offre devra comprendre deux réseaux d'interconnexion :

- un réseau rapide pour relier les nœuds de calcul, de login, d'administration et de stockage d'un débit minimal de 20 Gb/s et une latence faible (de l'ordre de la micro seconde).
- Un réseau gigabit ethernet permettant de relier les nœuds de calcul et le nœud d'administration.
- Le câblage complet pour l'interconnexion nœuds et maîtres pour les deux solutions réseau.

2.6 Stockage des données

Les calculs utilisent et génèrent un grand nombre de données. Il est donc nécessaire de proposer une solution de stockage de l'ordre de 20 To utiles.

- Baies de disques pour une capacité de stockage de 20 To utiles avec une garantie par rapport aux pannes (système RAID 6). L'offre doit donner le débit d'accès aux fichiers.
- Un nœud spécifique servira le stockage de données. Il disposera d'un accès réseau rapide d'une bande passante minimale de 20 Gb/s et d'une latence faible.
- Système de sauvegarde simple d'une capacité de 10 To, sur la base de disques qui peuvent être arrêtés et démarrés de manière logicielle, et de synchronisations cycliques depuis l'espace principal.

2.7 Administration et exploitation

L'offre devra intégrer des possibilités physiques d'administration et d'exploitation de la machine (écran pour le nœud d'administration, câblage spécifique, KVM, etc.). Elle devra en outre donner des possibilités d'administration à distance (IPMI, KVM, etc), d'arrêter et de redémarrer individuellement chaque nœud ou l'ensemble, de manière logicielle à distance, d'en couper et redémarrer l'alimentation et de réaliser un démarrage progressif. Le matériel d'administration devra être intégré dans les armoires rack ou lame.

Du point de vue logiciel :

- système d'exploitation installé : Linux. La configuration devra être supportée par linux (<http://www.kernel.org>), si les pilotes de périphériques ne font pas partie de ce noyau Linux officiel, le candidat s'engage à les fournir et à maintenir des versions à jour avec le noyau. La distribution utilisée ou les caractéristiques du système par rapport aux distributions classiques doivent être fournies.
- Le candidat est invité à intégrer à sa proposition les logiciels d'administration et de contrôle qu'il préconise pour son matériel. Pour chacun de ces logiciels, le candidat précisera les modalités d'accès pour les logiciels d'usage gratuit, et les conditions d'accès (nombre de licences, accès au code source...) pour les logiciels normalement payants, fournis à titre gratuit par le candidat.

2.8 Logiciel

Du point de vue logiciel la solution devra intégrer un environnement de développement optimisé pour le processeur choisi qui comprendra :

- un compilateur commercial fortran optimisé pour le processeur proposé, compatible fortran 77, 95 et openMP pour 5 utilisateurs simultanés,
- un compilateur commercial pour les langages C et C++, optimisé pour le processeur proposé, compatible ANSI C89 et C99, pour 5 utilisateurs simultanés,
- la fourniture de bibliothèques scientifiques d'algèbre linéaire, de transformée de Fourier, voire plus, utilisable facilement depuis un programme fortran ou C et optimisées pour le processeur et le compilateur proposés,
- une version optimisée et adaptée à la configuration matérielle de la bibliothèque de communication MPI.

Le système devra impérativement supporter l'exécution des logiciels et bibliothèques suivants :

- Matlab
- Modélisation multi-physique : Comsol
- Compilateurs : icc (intel), Intel ifort Linux, Portland Group
- Bibliothèques informatiques : Blas, Lapack, Scalapack, IMSL et NAG, Libgoto, Phonon
- Modélisation moléculaire : Gaussian, Namd, Amber, Vasp, Siesta, VMD
- communication MPI : OpenMPI, MPI-CH, LAM - MPI

2.9 Climatisation

L'offre devra comprendre une solution de climatisation adaptée et performante, dimensionnée en fonction du cluster. La salle destinée à recevoir la machine est déjà équipée d'une climatisation mais ce système est insuffisant pour recevoir un cluster de la taille envisagée.

Une solution à base d'eau froide est à envisager - solutions type "armoires climatisées" ou "porte réfrigérante" - dans la mesure où un groupe d'eau froide de capacité suffisante existe dans le bâtiment et qu'il est possible de s'y raccorder.

Pour plus d'information sur le système actuel et l'intégration possible de la proposition de climatisation, il est possible de contacter le responsable technique de l'offre.

2.10 Alimentation

L'offre doit comprendre une solution d'alimentation adaptée, à partir d'une arrivée basée sur 4 câbles 32A triphasés nus et de prises murles connectées à un système de courant secouru, qui seront disponibles dans la salle où sera installé le cluster.

Un système de courant secouru existe déjà dans le bâtiment destiné à recevoir le cluster. Néanmoins, le secours des nœuds n'est pas envisagé. Seuls les nœuds de services et le stockage seront branchés sur l'onduleur existant. Ceci suppose l'adjonction de batteries pour maintenir sa capacité actuelle.

Pour plus d'information sur le système actuel et l'intégration possible de la proposition d'alimentation, il est possible de contacter le responsable technique de l'offre.

2.11 Garantie, maintenance

3ans , J+1 sur site.

2.12 Variantes

Dans le but de prendre en compte les besoins spécifiques de certains utilisateurs, toutes les offres doivent donner le coût des variantes suivantes :

Nœuds : le coût d'un nœud supplémentaire, sur la base d'un nœud de calcul standard.

Processeurs : le coût de passage à une version de processeur plus rapide.

Mémoire centrale : certains codes de calcul ayant des besoins de mémoire centrale plus importants, il faut préciser le coût de passage, pour un nœud, à une capacité de 24 Go et de 96 Go.

Processeurs graphiques : le coût d'ajout à un nœud d'une unité graphique permettant les calculs en double précision, dans le seul but de faire du calcul sur GPU.

Garantie : préciser une extension de garantie à 4 ans J+1 sur site.

Le constructeur est encouragé à proposer des variantes si il pense que celles-ci peuvent apporter un gain d'utilisation de la machine.

2.13 Installation /transfert de compétences

La machine sera installée à l'université de Franche-Comté, dans les locaux du CRI de Besançon. La prestation doit comprendre l'installation du cluster, le transfert de compétences et la validation des fonctionnalités attendues. L'installation devra avoir lieu dans la seconde quinzaine de septembre.

- le raccordement électrique (prises) devra être fourni, les candidats sont encouragés à entrer en contact avec le référent technique pour élaborer une solution d'alimentation en rapport avec les possibilités du site.
- le raccordement de la solution de climatisation au système actuel, la prise en charge des condensas.
- installation matérielle sur site : l'ensemble du cluster (nœuds, serveur de fichiers, commutateurs et KVM si présent) seront installés.
- Installation du logiciel d'administration et transfert de compétences (déploiement d'un nœud, paramétrage,...)
- Fourniture d'une documentation technique (en version informatique) explicitant la configuration matérielle et logicielle, ainsi que les principes de base de l'administration et de l'utilisation de la machine.
- Formation 3 jours sur site des administrateurs et utilisateurs lors de l'installation de la machine plus une formation annuelle pour les mises à jour.
- La mise en ordre de marche portera sur la conformité de l'installation à l'offre technique du candidat et se déroulera de la façon suivante :
 1. installation de tous les équipements ;
 2. installation d'une distribution linux 64 bits sur tous les nœuds, dont au moins une installation utilisant la méthode PXE ;
 3. validation de la reconnaissance complète des caractéristiques physiques des machines : nombre de processeurs, taille de la mémoire, (en particulier pour les nœuds à grande capacité), cartes graphiques si présentes, interfaces réseau, ipmi, ... ;
 4. validation du système de contrôle : reboot à distance, mise hors tension, mise sous tension, vérification du déport clavier/écran ;
 5. installation et configuration des bibliothèques implémentant le protocole MPI ;
 6. exécution des programmes de tests fourni par le mésocentre à travers une queue d'exécution ;
 7. validation du système de sauvegarde.

- En cas de mise à jour majeure de l’environnement logiciel et/ou matériel, le fournisseur s’engage à intervenir sur site pour réaliser cette mise à jour, en collaboration avec l’ingénieur du mésocentre.
- Évacuation des emballages à l’issue de l’installation.

Notre projet est lié à l’activité de recherche de certains laboratoires de l’UFC, l’ENSMM et l’UTBM. Cette dynamique entraînera l’implication naturelle des étudiants de troisième cycle. Un tel projet sera largement facilité par un accompagnement et le partenariat du constructeur de la machine retenue. En particulier, l’organisation d’un séminaire d’une journée pour les utilisateurs, à Besançon, en vue du démarrage des moyens de calcul est souhaitée. Puis, une journée par an pourra être organisée pour présenter les nouveautés et répondre à des demandes plus techniques des chercheurs. Un accompagnement fort des chercheurs (en particulier des doctorants) est souhaité, par exemple, en ouvrant l’accès aux centres de compétences du constructeur retenu. Toute proposition en ce sens devra être clairement exprimée (type de partenariat, localisation des centres, durée de l’engagement, moyens complémentaires mis à disposition, etc.) dans la réponse du fournisseur.

3 Informations complémentaires

3.1 L’offre sera remise sous forme

Il conviendra de fournir une réponse détaillée au niveau des coûts. Les offres financières devront être accompagnées des fiches techniques des produits proposés. L’accompagnement dans la mise en œuvre de la solution sera détaillé.

Informations à fournir :

- descriptif des contraintes liées à l’accueil du cluster : consommation électrique, rapport performance/consommation par nœud, ventilation, nombre d’armoires, nombre d’unités par nœuds...
- délais de livraison ainsi que les délais d’installation et de mise en route du matériel dans les locaux de l’université. Un diagramme explicitant l’ordonnancement des différentes tâches devra être fourni,
- estimation de la fiabilité de l’offre (MTBF).

3.2 Discussions techniques

Dans le cadre de la procédure, des entretiens avec le responsable technique de l’appel d’offres seront organisés.

Pour la réalisation des mesures de performances deux solutions doivent être envisagées : la mise en place d’un accès depuis l’université de Franche-Comté vers les machines de test pour une personne de l’université et la mise à disposition des codes de test pour être testés

en interne par les entreprises répondant à l'offre.

3.3 Évolution de l'offre

Dans l'hypothèse d'un financement ultérieur, l'offre doit permettre des possibilités d'évolution, soit par l'ajout de nouveaux nœuds, soit par l'évolution des nœuds en place (augmentation mémoire, espace scratch, ajout d'unités graphiques pour calcul sur GPU), soit par l'évolution du réseau en place pour offrir plus de bande passante. L'offre devra développer les possibilités d'évolution (paliers possibles) sur ces trois points.

4 Critères d'attribution

Pour être valable l'offre devra impérativement :

- avoir validé l'exécution des codes de test qui seront fournis par le mésocentre de calcul,
- décrire de quelle manière les concepts de développement durable sont pris en compte aux différents stades de la construction, de l'exploitation et de la fin de vie du matériel,
- donner les performances de la machine sur les principaux benchmark CINT 2006, SpecIntRate 2006, CFP 2006, SpecFPRate 2006, bande passante crête et bande passante utile du réseau.

L'attribution se fera à l'offre économiquement la plus avantageuse appréciée en fonction des trois critères énoncés ci-dessous (par ordre de priorité décroissante) :

Qualité technique du matériel proposé (50%)

pour permettre l'évaluation de ce critère l'offre devra détailler :

- performances des codes de test fournis, performances sur les principaux benchmarks CINT 2006, SpecIntRate 2006, CFP 2006, SpecFPRate 2006, Linpack (HPL), support de bibliothèques,etc.
- caractéristiques techniques : nombre de cœurs et nœuds, mémoire par cœurs, capacité de stockage, débit du réseau, débit d'accès aux fichiers, alimentation, options proposées, etc.
- assistance technique, SAV, accompagnement dans le déploiement de la solution.
- possibilités d'évolution.

Prix (30 %) calculé à partir du coût d'achat et du coût de fonctionnement.

Développement durable (20 %)

pour permettre l'évaluation de ce critère l'offre devra inclure :

- informations sur les labels et normes garantis par l'offre,
- les caractéristiques environnementales des équipements (consommation des processeurs, dispositifs d'économie d'énergie, efficacité des alimentations électriques).

- politique mise en œuvre par rapport à la protection de l'environnement et en particulier par rapport au problème de l'émission des gaz à effet de serre et de la recyclabilité des produits.
- politique mise en œuvre pour le respect de la dignité humaine.