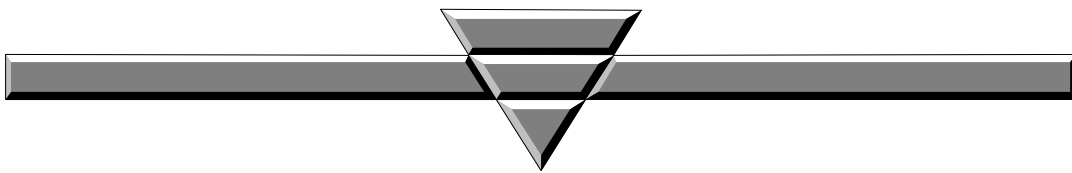


MARCHES PUBLICS DE FOURNITURES COURANTES ET SERVICES

Ministère de l'Education Nationale
UNIVERSITE JOSEPH FOURIER
BUREAU DES ACHATS ET MARCHES
BP 53
38041 GRENOBLE CEDEX 9



**ACHAT ET INSTALLATION D'UNE SOLUTION DE CALCUL HPC
À BASE DE PROCESSEURS 64 BITS
AVEC EXTENSION DE GARANTIE À 3 ANS**

UNIVERSITE JOSEPH FOURIER (GRENOBLE I)

Cahier des Clauses Techniques Particulières

Cahier des Clauses Techniques Particulières

Programme fonctionnel

pour la fourniture d'une solution de calcul HPC à base de processeurs 64 bits avec extension de garantie à 3 ans

1.	<i>Objet du dossier</i>	4
2.	<i>Plateforme matérielle</i>	4
2.1.	Serveur frontal	5
2.2.	Nœuds de calcul	5
2.3.	Les processeurs	6
2.4.	La mémoire centrale	6
2.5.	Systèmes d'entrées/sorties	6
2.6.	Blade	7
2.7.	Connectivité réseau	7
2.8.	Disponibilité et fiabilité	8
2.9.	Arrêt et démarrage	8
3.	<i>Système d'exploitation</i>	8
3.1.	Outils d'administration et de contrôle	8
3.2.	Sécurité	9
3.3.	Travaux interactifs	9
3.4.	Traitement par lots	9
3.5.	Aide à l'exploitation	9
4.	<i>Environnement de programmation</i>	9
4.1.	Langages de programmation	10
4.2.	Outils d'analyse des performances	10
4.3.	Outils de mise au point	10
4.4.	Bibliothèques numériques	10
4.5.	Bibliothèques de gestion de parallélisme	11
4.6.	Logiciels applicatifs	11
5.	<i>Maintenance, support technique et formation</i>	11
5.1.	Garantie	11
5.2.	Maintenance préventive et corrective des matériels	13
5.3.	Assistance technique et maintenance des logiciels	13
5.4.	Formation	13
5.5.	Documentation	14
6.	<i>Environnement physique</i>	14
7.	<i>Installation, modalités de la livraison</i>	15

8.	<i>Jeu de tests</i>	15
9.	<i>Les propositions</i>	16
10.	<i>Détail des tests</i>	16
10.1.	Codes Publics	16
10.2.	Codes spécifiques - «recherche»	17
11.	<i>Contacts</i>	18

1. Objet du dossier

Le présent marché a pour objet l'achat et l'installation d'une solution de calcul HPC à base de processeurs 64 bits avec extension de garantie à 3 ans pour le pôle PHYNUM (NANOSStar) du projet CIMENT (Université Joseph Fourier, Grenoble).

Contexte

Le pôle PHYNUM de CIMENT a, en particulier, pour objectif de proposer à ses utilisateurs des solutions de calculs optimales du point de vue des performances, de la stabilité des composants matériels/logiciels et le respect des contraintes environnementales.

Dans le contexte présent, le panel d'applications est relativement ciblé : il s'agit d'applications relatives aux nanosciences avec l'utilisation de codes DFT principalement (codes de type VASP, Abinit, Siesta, ...). Ce calculateur sera en outre utilisé dans le cadre de la grille de calcul CIGRI.

2. Plateforme matérielle

Terminologie : nous appellerons par la suite *nœud de calcul* une unité de calcul constituée de plusieurs processeurs fortement couplés à une mémoire partagée, et autorisant l'exécution d'un code OpenMP.

La solution sera constituée :

- d'un serveur frontal, compatible au niveau binaire avec les autres nœuds, dédié au service batch et à l'administration,
- d'un certain nombre de serveurs de calcul à base de processeurs 64 bits quadri-cœurs,
- d'une infrastructure réseau suffisante pour couvrir les besoins en terme de calcul parallèle et d'administration,
- d'une solution de stockage de capacité supérieure ou égale à 3,5 To (débit minimal de 150 Mo/s dont le niveau de performance sera testé par IOZONE),
- d'un système de contrôle pour l'administration à distance.

Les besoins en calcul vont soumettre la solution fournie à un usage intensif de la solution avec une utilisation 24h sur 24 des performances maximales. La solution proposée devra donc non seulement répondre à des besoins de haute performance, mais encore à des besoins de robustesse et de haute disponibilité du matériel et des logiciels. Enfin, l'aspect énergétique est un élément important de ce dossier (minimisation de la consommation électrique et de la dissipation thermique).

2.1. Serveur frontal

Le serveur frontal sera dédié à l'administration, au service batch et à la compilation des codes, il devra présenter les caractéristiques suivantes :

- compatible binaire avec les nœuds de calcul,
- mono ou bi processeur x86-64, quadri-cœur, de fréquence ≥ 2.5 GHz, avec au moins 6 Mo cache L2 et un bus ≥ 1333 MHz,
- 8 Go de mémoire, 667 MHz ou plus,
- 2 disques internes, de capacité individuelle au minimum 300 Go, qui seront exploités en RAID1,
- 1 disque interne en secours (identique aux deux précédents), pour un système de secours ou tester de nouvelles versions.

2.2. Nœuds de calcul

Chaque nœud de calcul doit être constitué de deux processeurs (quadri-cœur) x86-64, de fréquence ≥ 2.5 GHz, avec au moins 12 Mo (2x6 Mo) de cache L2, et d'un bus de fréquence ≥ 1300 MHz.

La capacité mémoire de chaque nœud de calcul devra être de 16 Go (fréquence de la mémoire ≥ 667 MHz).

Chaque nœud de calcul devra accéder à un espace de stockage dont la capacité totale sera d'au moins 3,5 To, avec un débit soutenu d'au moins 150 Mo/s.

Les spécifications complémentaires et les renseignements à détailler sont les suivants :

Le soumissionnaire présentera l'architecture générale du système :

- 1) noeuds de calcul ;
- 2) système de base ;
- 3) système d'entrées/sorties.

Les architectures mémoire et caches au sein d'un processeur et au sein d'un nœud devront être décrites précisément. Les fréquences d'horloge et les latences devront être indiquées.

La bande passante mémoire et cache par processeur doit être spécifiée, en indiquant les contentions en multi-processing.

Dans tous les cas, il faudra préciser :

- 1) le partage d'entrées/sorties disques ;
- 2) l'administration globale des ressources.

2.3. Les processeurs

Ce seront des processeurs superscalaires performants, 64 bits, homogènes sur l'ensemble de la solution.

Le soumissionnaire décrira de manière détaillée leur architecture sur les aspects suivants :

1. fréquence ;
2. caches mémoires (hiérarchie, taille, latences et débits etc...) ;
3. registres ;
4. unités de calcul ;
5. pipelines ;
6. débits d'entrées/sorties ;
7. puissance crête ;
8. puissance soutenue ;
9. parallélisme au niveau des instructions : branchements spéculatifs, exécution dans un ordre modifié... ;
10. le type d'instrumentation hardware (présence de sondes en vue de mesures de performances des programmes).

2.4. La mémoire centrale

Une partie des utilisateurs soumet des travaux très consommateurs de ressources en mémoire centrale.

Il est souhaité, pour ces applications, que les ressources en mémoire centrale puissent être partagées entre plusieurs processeurs, sans qu'il en résulte de conséquences sur la tenue en charge.

La configuration mémoire, s'appuie sur la base suivante :

- au moins 2 Go de mémoire (de fréquence au moins égale à 667 MHz) par cœur sur l'ensemble de la plateforme.

La solution proposée (en terme de fréquence mémoire / FSB) devra limiter au maximum les problèmes de contention.

La configuration mémoire de chaque nœud devra respecter les règles de placement optimales pour les performances (les tailles des barrettes mémoire seront éventuellement choisies en conséquence).

2.5. Systèmes d'entrées/sorties

Les spécifications globales du sous-système disque sont :

- 1) **un espace disque cumulé d'au moins 3,5 To** pour la globalité de l'offre matérielle ;
- 2) un débit effectif de transfert entre le sous-système disques et un nœud de la plate-forme de calcul au moins égal à 150 Mo/s.

Le soumissionnaire pourra proposer :

- soit un système d'entrées/sorties distribué.
- soit un système d'entrées/sorties centralisé.

Dans le cas d'un système distribué, une capacité minimale de 140 Go par processeur est requise (pour le système et les besoins temporaires de stockage). Un espace minimum de 1 To devra être visible de tous les nœuds. Les données seront alors partagées via NFS, ou via un système de fichiers parallèles plus performant. Le système de fichiers doit être natif 64 bits. Il doit être possible de stocker des fichiers de grande taille (supérieure à 100 Go).

Dans le cas d'un système de fichiers distribué, il doit être partitionné en conséquence.

Dans le cas d'un système centralisé, nous laissons libre cours au soumissionnaire pour proposer une solution performante, adaptée à la configuration et assurant un débit effectif d'au moins 150 Mo/s. Dans ce cas, la perte simultanée de 2 disques ne doit pas entraîner la perte de données.

Un lecteur de DVD (de dernière technologie) sera installé sur le nœud d'administration et un deuxième nœud de secours.

La configuration et le tuning du système seront réalisés avec l'aide du constructeur.

Seront indiqués les débits d'écriture et de lecture.

Le soumissionnaire détaillera la configuration des entrées/sorties et en particulier indiquera le type de connectique externe (fibre...), la taille maximale d'un fichier et la possibilité d'extension dynamique. Les différents niveaux de configuration Raid disponibles seront indiqués.

2.6. Blade

Si la solution proposée est constituée de blades, leur descriptif détaillé est demandé :

- nombre de nœuds par blade ;
- consommation électrique en charge (linpack), efficacité des boîtiers d'alimentation, communication, redondance, ventilateurs, connectique, etc...

2.7. Connectivité réseau

La configuration proposée doit à la fois assurer des entrées/sorties disque rapides et un débit réseau performant pour les applications parallèles MPI.

Le soumissionnaire proposera l'architecture réseau qui lui semblera la plus performante globalement au regard des benches demandés.

La latence mpi maximum effective au moins niveau infiniband 4 X DDR (<4 micro-secondes). Cependant ces spécificités devront être adaptées à la taille de nœuds dans le but d'obtenir globalement une architecture équilibrée.

Un ou plusieurs commutateurs devront être fournis pour relier l'ensemble des ports.

Le cas échéant, le ou les commutateurs ethernet devront être administrables et supporter les VLANS.

Dans tous les cas, il devra être possible de connecter l'ensemble de la solution en 10Gb ethernet (1 seul port suffit).

Les principaux critères pris en compte pour juger de la performance des commutateurs seront la bande passante en bisection sur la configuration proposée et les performances obtenues sur les benches MPI.

2.8. Disponibilité et fiabilité

L'utilisation intensive du supercalculateur nécessite une disponibilité maximale de la totalité de la configuration. Dans ce cadre-là, le soumissionnaire détaillera les points suivants :

- fiabilité des sources d'alimentations électriques ;
- éléments pouvant être remplacés à chaud ;
- redondance des éléments critiques ;
- fonctionnement en mode dégradé.

2.9. Arrêt et démarrage

Le «reset» matériel doit pouvoir être assuré à distance en toute fiabilité, et ce quelque soit l'état du système.

Indiquer la durée du démarrage de la plateforme proposée :

- lors d'un démarrage après coupure totale d'électricité ;
- lors d'un redémarrage, tous les composants étant sous tension.

Indiquer la durée de la séquence d'arrêt complet de la machine et de re-démarrage. Si cette durée est proportionnelle au nombre de processeurs ou à la taille du sous-système disque, indiquer la méthode de calcul.

Le pic de consommation électrique à la mise sous tension sera communiqué ainsi que le mode d'allumage (tous les nœuds simultanément, en décalage...).

3. Système d'exploitation

Le système d'exploitation sera basé sur un système Linux. Il devra prendre pleinement en compte l'aspect multiprocesseurs de la machine et devra pouvoir être utilisé concurremment par plusieurs utilisateurs.

Le soumissionnaire décrira en outre :

- les extensions aux fonctionnalités de base, s'il y en a ;
- les outils d'administration propres fournis dans la solution ;
- la possibilité d'installer une nouvelle version du système pendant une exploitation normale, s'il y a lieu ;
- la possibilité d'avoir plusieurs versions différentes du système et de pouvoir basculer de l'une à l'autre, s'il y a lieu.

3.1. Outils d'administration et de contrôle

L'administration, le contrôle et la répartition des ressources devront se faire à partir d'un nœud d'administration. Le soumissionnaire devra intégrer à sa proposition les

logiciels d'administration et de contrôle qu'il préconise pour son matériel. Ces outils doivent être utilisables via une interface graphique mais aussi en ligne de commande. Les outils développés par le constructeur devront être utilisés le plus souvent possible. Leurs fonctionnalités seront présentées et résumées. Nous souhaitons une offre packagée maximale, supportée à 100% par le constructeur **de préférence basé sur des briques open source.**

Ce logiciel d'administration devra entre autres permettre de :

- déployer les images des nœuds,
- afficher l'état des nœuds (informations telles que la Température par exemple, l'espace disque, etc...),
- permettre de déclencher des alarmes par mail.

Le système d'exploitation proposé devra permettre l'utilisation des ressources en séquentiel ou en parallèle, en mode interactif ou en traitement par lots, et ceci de façon simultanée.

3.2. Sécurité

Les projets traités par le calculateur demandent un niveau de sécurité élevé, tant pour l'accès au système que pour l'accès aux données.

Le système d'exploitation devra permettre d'effectuer les mises à jour de sécurité de façon régulière, sans contrainte et sans limite de durée.

3.3. Travaux interactifs

Il devra être possible, utilisateur par utilisateur, de définir les ressources maximales utilisables en interactif : temps 'cpu', taille mémoire, taille disque, nombre de 'cpu'..., et ceci quel que soit le nombre de sessions interactives ouvertes par un utilisateur donné.

3.4. Traitement par lots

Nous utiliserons le système de traitement par lot OAR. Le soumissionnaire devra s'assurer que ce logiciel opensource est compatible avec la solution proposée.

3.5. Aide à l'exploitation

Un outil avec une interface graphique doit permettre de surveiller l'utilisation des ressources de la machine : processeurs, mémoire par nœuds et globale, disques, occupation du swap, interfaces réseaux, etc. Le décrire en détail.

4. Environnement de programmation

Le système de développement doit être le plus complet possible : compilateurs fortran, C/C++, bibliothèques scientifiques scalaires et parallèles, débogueurs (séquentiel), outils d'analyse et de profils de performance, éditeurs de texte (vim, emacs ou Xemacs), outils de génie logiciel (lex, yacc, make scalaire/parallèle). Il doit être possible de mélanger les langages dans une même application.

Les codes exécutables obtenus doivent permettre d'accéder à la totalité des ressources en mémoire centrale et en capacité disque.

Les programmeurs doivent avoir accès à des outils de mesures statistiques d'utilisation des ressources de la machine (mesure de Mflops, des accès aux mémoires cache, aux disques, etc...).

Le soumissionnaire devra fournir les renseignements techniques sur les points suivants :

4.1. Langages de programmation

Les **compilateurs** pour les langages suivants

- 1) fortran 77/90/95/2003 (fournir le calendrier de support du fortran 2003 en cas de non disponibilité immédiate) ;
- 2) C/C++ ;

☞ doivent fournir des **codes exécutables en mode 64 bits** ;

☞ doivent être livrés avec le nombre de licences nécessaires pour exécuter au moins deux compilations simultanées sur le noeud «frontal» et des exécutions sur les autres nœuds (les codes parallèles peuvent potentiellement être exécutés sur l'ensemble de la plateforme).

Le soumissionnaire précisera, langage par langage, les niveaux de version. Il indiquera les niveaux de support des directives **OpenMP** et des bibliothèques **Pthreads** dans les langages fortran et C. Il signalera également les directives propres de compilation pour le système proposé. S'il existe plusieurs offres de compilateurs pour la plate-forme, il devra justifier le choix qui figurera dans son offre.

4.2. Outils d'analyse des performances

Le soumissionnaire décrira les outils inclus dans son offre permettant de mesurer les performances des codes exécutables scalaires et parallèles en 64 bits.

4.3. Outils de mise au point

Le soumissionnaire décrira les outils de mise au point pour les compilateurs. Ceux-ci doivent permettre d'analyser les programmes scalaires et parallèles.

4.4. Bibliothèques numériques

Les bibliothèques doivent pouvoir être appelées depuis fortran, C, C++.

Le soumissionnaire indiquera les **bibliothèques numériques disponibles en 64 bits**, pour la plate-forme proposée.

Il indiquera en particulier la présence de :

- 1) Blas 1, 2 et 3 et Blas parallélisé pour 64 bits ;
- 2) Lapack pour 64 bits ;
- 3) ScaLAPACK pour 64 bits ;
- 4) bibliothèques spécifiques optimisées avec FFT en 64 bits ;

5) toute bibliothèque mathématique permettant de traiter en scalaire et en parallèle des problèmes d'algèbre linéaire de grands systèmes creux (résolution de systèmes linéaires, recherche d'extremum, calculs de valeurs propres et vecteurs propres) ;

6) GSL (Gnu Scientific library).

4.5. Bibliothèques de gestion de parallélisme

Les bibliothèques d'échange de messages MPI (*Message Passing Interface*) doivent être disponibles pour la plate-forme proposée en mode 64 bits.

La bibliothèque MPI doit être au niveau 2.

La bibliothèque d'échange de messages MPI (*Message Passing Interface*) doit être fournie en version optimisée pour la plate-forme proposée.

Le soumissionnaire indiquera :

- la version de la bibliothèque MPI fournie (le détail des fonctionnalités de MPI-2 offertes) ;
- la valeur de la latence effective MPI (sera vérifiée) ;
- la bande-passante effective MPI.

Les compilateurs C et fortran doivent supporter **OpenMP version 1.1**.

Le soumissionnaire indiquera le détail des fonctionnalités de OpenMP 2 fournies.

Le soumissionnaire indiquera quel est le statut des licences correspondant aux différents composants logiciels (compilateurs, bibliothèques, batch...) proposés, et son niveau d'engagement pour leur support et leur évolution.

4.6. Logiciels applicatifs

Les logiciels VASP, ABINIT et SIESTA doivent être utilisables dans leur version parallèle.

5. Maintenance, support technique et formation

5.1. Garantie

Une extension de garantie incluant la maintenance et le support technique pour la solution globale fournie devra être incluse pour une durée de trois ans à compter de l'admission du matériel, frais de déplacement de personnel inclus. Sans facturation supplémentaire, le titulaire assurera le dépannage en 24 heures ouvrables maximum qui suivent l'appel demandant l'intervention d'un technicien sur site. Toute intervention s'effectuera entre 8h30 et 17h, du lundi au vendredi (non compris les jours fériés).

A titre indicatif, le soumissionnaire chiffrera le coût annuel des prestations contenues dans l'extension de garantie/maintenance au-delà des 3 années prévues.

Dans le cas où les composants d'un ensemble installé ne proviendraient pas d'un même constructeur, le titulaire sera tenu pour seul responsable d'un mauvais fonctionnement ou de toute défectuosité qui pourrait résulter d'un assemblage de pièces ou d'accessoires mal adaptés.

Dans le cas où un composant de l'équipement nécessiterait un retour atelier ou si le temps nécessaire au dépannage ou à la fourniture de pièces détachées dépasse une semaine (ou 5 jours ouvrés), le titulaire mettra gratuitement à la disposition de l'Université Joseph FOURIER un composant dans les 48 heures ouvrées.

En cas de panne persistante ou répétitive, le titulaire s'engage sur demande écrite de l'Université Joseph FOURIER à remplacer le matériel défectueux dans un délai d'une semaine (ou 5 jours ouvrés) à compter de la réception du courrier par le titulaire.

Que le soumissionnaire assure lui-même la garantie ou qu'il la sous-traite à un tiers, le soumissionnaire s'engage à désigner un seul interlocuteur (numéro d'appel et/ou e-mail) propre à son organisation pour toute demande d'intervention.

Le soumissionnaire doit préciser séparément pour chaque composant garanti :

- le mode opératoire ;
- le support technique proposé ;
- les délais d'intervention ;
- les délais de réparation ;
- les moyens qu'il entend mettre en œuvre pour assurer le support matériel, en particulier le matériel de remplacement restant à demeure dans les locaux de l'acheteur ;
- les autres prestations incluses ;
- le cas de sous-traitance de la prestation de garantie à un tiers, le (ou les) nom(s) et coordonnées du (ou des) sous-traitant(s) auquel il fera appel.

Pour les logiciels inclus dans la proposition du soumissionnaire, celui-ci s'engage à fournir les mises à jour de ces logiciels durant la période de garantie sans facturation supplémentaire.

Le support technique des différents composants des nœuds devra intégrer la mise à jour des BIOS, drivers, firmware, etc... de ces derniers. À défaut, le soumissionnaire s'engage à mettre à disposition des outils permettant d'automatiser ces tâches et leur reconfiguration.

Le titulaire s'engage à fournir :

- les mises à jour des pilotes de périphériques qui ne sont pas supportés dans le noyau linux officiel (si linux) ;
- les mises à jour des logiciels et les correctifs de bogues pour les équipements réseau.

Le contrat de garantie devra rester valable en cas d'extensions ajoutées aux nœuds (mémoire par exemple), cependant les extensions ajoutées après l'admission du matériel ne sont pas concernées par la garantie décrite dans le présent marché.

5.2. Maintenance préventive et corrective des matériels

Le soumissionnaire précisera au moyen de l'annexe 3 :

- la périodicité de la maintenance préventive et le temps d'indisponibilité généré, s'il y a lieu ;
- les périodes de prise en compte des demandes d'intervention ;
- les périodes d'intervention de la maintenance corrective qui s'étendra au minimum de 08h30 à 17h00 du lundi au vendredi inclus, non compris les jours fériés ;
- la localisation des personnels faisant la maintenance corrective ;
- les délais de disponibilité des composants critiques (mémoire, processeur, disque, contrôleur disque, carte réseau...).

5.3. Assistance technique et maintenance des logiciels

Comme indiqué auparavant, l'extension de garantie doit fournir toutes les mises à niveau des logiciels livrés. Au moins une fois par an une mise à niveau globale de l'ensemble de ces logiciels devra être effectuée par le soumissionnaire.

Le soumissionnaire précisera au moyen de l'annexe 3 :

- la périodicité prévue de sortie des nouvelles versions des logiciels fournis (système, compilateurs, etc...) ;
- le nombre de journées sur site d'intervention du personnel du soumissionnaire prévu dans la proposition pour la période de garantie de trois ans (ainsi que le contenu des travaux de maintenance prévus pendant ces journées).

5.4. Formation

Un transfert de compétence devra être prévu sur site pour permettre l'administration et l'exploitation de la solution proposée. Un programme pour la formation de 3 personnes devra être proposé par le soumissionnaire.

Les points suivants seront notamment abordés (**liste non exhaustive**) :

- Arrêts / redémarrages (à distance) ;
- Logiciel d'administration ;
- Programmation du gestionnaire de batch ;
- Processus de mise à jour ;
- Déploiement systèmes et logiciels sur les nœuds ;
- Simulation de panne (réinstaller un nœud par exemple) ;
- Outils d'analyse de performance.

Le coût de la formation est compris dans le prix indiqué à l'article 2 de l'acte d'engagement.

5.5. Documentation

Toute la documentation se rapportant aux matériels, aux systèmes d'exploitation, aux compilateurs, aux bibliothèques doit être accessible en ligne, au moyen d'un navigateur Internet.

La mise à jour de cette documentation devra être fournie pendant toute la période de maintenance.

6. Environnement physique

Le soumissionnaire devra préciser toutes les conditions d'environnement physique nécessaires à l'installation et au bon fonctionnement de la machine proposée.

Il précisera impérativement :

1. **côtes d'encombrement pour l'installation au sol ;**
2. **poids, dimensions ;**
3. **puissance électrique totale consommée (VA / kW) ;**
4. **efficacité des boîtiers d'alimentation utilisés (en %) ;**
5. tension d'alimentation monophasé ;
6. intensité nominale ;
7. courbe du disjoncteur (ou surintensité au démarrage en A) ;
8. protection différentielle (30, 200mA ou 1A) ;
9. type de raccord au réseau électrique du laboratoire (référence des éléments de connectique) ;
10. **la fourchette de consommation électrique globale pour la machine proposée ; cette fourchette doit notamment tenir compte du pic à l'allumage du calculateur ; la consommation électrique globale en charge (Linpack) sera fournie ;**
11. type de refroidissement ;
12. **niveau sonore en Db pour le système complet ;**
13. **la puissance thermique dégagée par l'ensemble du calculateur en BTU/h ;**
14. la circulation du flux d'air de refroidissement ;
15. recommandations éventuelles ;
16. accessoires à prévoir ;
17. un schéma électrique de raccordement

En outre, le soumissionnaire précisera dans sa réponse les actions qui ont été engagées d'une façon générale (toute la chaîne de fabrication du matériel) dans le but de diminuer les émissions de gaz à effet de serre et dans le but de diminuer la consommation énergétique tout en maintenant un bon niveau de performance.

7. Installation, modalités de la livraison

Avant d'être installé dans les locaux de la DSI-GU, le matériel aura été préalablement monté et testé dans les locaux du titulaire. Le ordinateur sera également de préférence configuré en usine d'un point de vue logiciel (nœud d'administration, déploiement sur les nœuds de calcul, installation des derniers drivers et firmwares, compilateurs, installation de logiciels et applications demandées -notamment ceux du jeu de tests-, etc...). Le matériel sera ainsi livré «clef en main». Sur place, l'accent sera ainsi mis sur l'aspect formation à l'administration et à l'exploitation de la machine.

La configuration des éléments réseaux de la machine est à la charge du soumissionnaire ainsi que l'étiquetage de l'ensemble des câbles aux deux extrémités.

Le titulaire à qui le marché aura été notifié s'engage à visiter les locaux destinés à recevoir le cluster au moins 30 jours avant la date de livraison du matériel, il pourra proposer une répartition des plateformes, favorisant une circulation correcte de l'air climatisé dans la salle.

Le matériel sera installé à l'adresse suivante :

Direction des Systèmes d'Information
351, avenue de la bibliothèque
Domaine Universitaire de Saint Martin d'Hères
38041 Grenoble CEDEX

Le matériel devra être livré au plus tard le 15 décembre 2008.

8. Jeu de tests

Il existe un jeu de tests, accessible uniquement par réseau (prendre contact avec Françoise BERTHOUD et/ou Bruno BZEZNIK - cf. point 11 - contacts), qui devra être exécuté sur une plate-forme correspondant à une version réduite, mais identique en tous points à la configuration proposée par le soumissionnaire. Cette plate-forme disposera d'un minimum de 24 cœurs (cf point 10, détails des tests).

Ce jeu comporte un test public et des tests spécifiques aux utilisateurs du pôle PHYNUM. Pour ces derniers, le soumissionnaire s'engage à ne pas les divulguer et à en effacer toute trace (code source, données, résultats...) sur ses machines de test.

Ces tests devront être reproduits à l'identique (durées et résultats) sur la machine livrée.

Les résultats de ce jeu de tests serviront à évaluer l'architecture proposée, mais ne serviront en aucun cas à décider, à leur seule vue, de la solution la plus performante.

9. Les propositions

Le soumissionnaire rédigera une proposition décrivant les caractéristiques de la solution technique proposée **sous forme de réponses à tous les points précisés dans le présent document. Avec la remise de ce rapport, une réponse sur support électronique sera fournie pour les fichiers de sortie.**

On demande aux soumissionnaires qui adresseront leur proposition définitive :

- d'avoir une expérience significative dans les solutions matérielles et logicielles des besoins du calcul scientifique et de fournir les références les plus significatives dans ce domaine et de spécifier les taux de disponibilité de solutions équivalentes à celle(s) proposée(s) ;
- de faire un engagement fort du constructeur du matériel proposé pour la pérennité et la validité de ce matériel, vis à vis des demandes et des tests de ce cahier des clauses techniques particulières.

Les candidats sont autorisés à proposer des variantes dans le respect des prescriptions techniques minimales indiquées dans le CCTP y compris les jeux de test (cf. point 10). La solution de base fera l'objet d'un rapport et chacune des variantes fera l'objet d'un rapport complémentaire et distinct.

10. Détail des tests

Nous vous demandons de bien vouloir respecter nos demandes (nombre de procs/cœurs, taille des problèmes à traiter, unité des résultats) et de justifier les absences de résultats le cas échéant.

10.1. Codes Publics

Code fftw, V3.1.2 FFTW (<http://www.fftw.org>)

Si N est le nombre de cœurs et TxTxT est la dimension de la matrice, L'exécution du code (`mpirun -np N -s1 rfftw_mpi_test -s TxTxT`) génère un fichier de résultats. Les valeurs suivantes nous intéressent.

- SPEED TEST: 768x768x768, FFTW_FORWARD, in place, generic
 - **time for one fft (uniprocessor): XXX s**
- et
- TRANSP.,w/WORK: **time for one fft (N cpus): XXX s** (XXX ns/point)
 - TRANSP.,w/WORK: "mflops" = $5/2 (N \log_2 N) / (t \text{ in microseconds}) = \mathbf{XXX}$
 - TRANSP.,w/WORK: **parallel speedup: XX**

Pour les configurations suivantes :

Nombre de coeurs (N)	Taille matrice (T)	Time (s) for 1 fft	Time (s) for N fft	Mflops MF	speedup	Mflops pour l'ensemble de la proposition (Nb_total_coeurs*MF/N)
N=8	T=192					
N=8	T=716					
N=16	T=716					
N=24	T=716					

10.2. Codes spécifiques - «recherche»

Les codes source ne doivent en aucun cas subir de modification.

Site de téléchargement : <ftp.grenoble.cnrs.fr>

Login et password sur demande auprès de l'un des deux contacts (cf point 11)

VASP (contact Alain.Pasturel@grenoble.cnrs.fr)

Les tests seront effectués à partir de la version la plus récente fournie par l'utilisateur.

Attention : Il faudra supprimer l'option DNGZhalf dans le makefile (option du CPP dans la version parallèle)

Pour ce code fournir :

- le fichier Makefile utilisé
- les fichiers de résultats

Pour les configurations suivantes :

Code VASP	précisions	Nb de coeurs (N)	Total CPU time used (sec) (CPU)	User time (sec)	Elapsed time (sec)	Indice de perf globale : Nb total de coeurs / (CPU*N)
Version séquentielle	N/A	1				
4 coeurs (1 procs quadri-core)		4				
4 coeurs (2 coeurs par processeur))		4				
8 coeurs (2 procs quadri core)		8				
24 coeurs (6 procs)		24				

SIESTA (contact : thierry.deutsch@cea.fr)

La documentation et les instructions se trouvent dans le répertoire SIESTA du site de téléchargement.

AB INIT (contact : valerio.olevano@grenoble.cnrs.fr)

Utiliser la version téléchargeable : http://www.abinit.org/package/?text=5_4_4

Compiler le code avec les options adéquates.

Les fichiers d'entrée sont disponibles à partir du site de téléchargement des benches.

11. Contacts

Les personnes à contacter (hors codes spécifiques) sont :

Françoise BERTHOUD

Responsable technique du projet PHYNUM (CIMENT)

Email : Francoise.Berthoud@grenoble.cnrs.fr

Bruno BZEZNIK

Responsable technique projet CIMENT

Email : Bruno.Bzeznik@imag.fr