



**Prestation d'expertise et d'assistance Calcul Intensif  
(HPC/IA) de l'ONERA**

**CAHIER DES CLAUSES TECHNIQUES PARTICULIERES**

**« CCTP-HPC-IA-ONERA »**

**V 0.3 – 17/07/2025**

# SOMMAIRE

## Table des matières

1. Contexte général.....	4
1.1. L'ONERA .....	4
1.2. La Direction des Systèmes d'Information (DSI) de l'ONERA.....	5
1.3. Objet de la prestation .....	5
2. Environnement du Calcul Intensif à l'ONERA.....	7
2.1. Ressources de calcul .....	7
2.2. Ressources de visualisation.....	8
2.3. Stockage centralisé du CSI .....	8
2.4. Architecture-type des calculateurs CSI et criticités.....	9
2.5. Taille du parc de calculateurs et pondérations associées.....	11
2.6. Projets d'évolution du SI ONERA.....	11
2.5.1 Projet de regroupement immobilier en Île-de-France (PRISME) .....	11
2.5.2 RPO.....	12
3. Conditions d'exécution des prestations .....	13
3.1. Contexte existant .....	13
3.2. Périmètre des prestations .....	13
3.3. Localisation des prestations.....	13
3.4. Plage de service de la prestation .....	14
3.5. Organisation du MCO/MCS sur le périmètre Calcul Scientifique Intensif.....	14
3.5.1 Les acteurs actuels du périmètre .....	14
3.5.2 Le comité de suivi d'exploitation du CSI.....	15
3.5.3 Organisation cible du support .....	15
3.5.4 Répartition cible des rôles et responsabilités .....	16
3.5.5 Accès « administrateur » au périmètre.....	17
3.6. Compétences requises .....	18
3.6.1 Compétences communes .....	18
3.6.2 Profil « Ingénieur système HPC » .....	18
3.6.3 Profil « Architecte HPC et/ou IA » .....	18
3.6.4 Profil « Expert code HPC et/ou IA » .....	18
4. Déroulement du contrat.....	19
4.1. Poste 1 : phase d'initialisation .....	19

4.1.1	Prise de connaissance de l'existant .....	19
4.1.2	Livrables .....	19
4.2.	Poste 2 : MCO/MCS/assistance utilisateur avancée .....	20
4.2.1	Description détaillée des prestations techniques attendues .....	21
4.2.2	Éléments de volumétrie .....	25
4.2.3	Qualité et indicateurs contractuels .....	26
4.2.4	Comité de pilotage .....	27
4.3.	Poste 3 : prestations d'expertise ponctuelles sur demande.....	28
4.4.	Poste 4 : phase de réversibilité .....	30
5.	Contraintes spécifiques liées à l'administration de systèmes d'informations sensibles.	31
5.1.	Contraintes minimales d'accès aux sites et au SI de l'ONERA .....	31
5.2.	Habilitation des intervenants des prestation .....	31
5.3.	Accès à distance au SI général interne de l'ONERA .....	31

## 1. Contexte général

### 1.1. L'ONERA

L'ONERA (Office National d'Études et de Recherches Aérospatiales) est un établissement public scientifique et technique à caractère industriel et commercial, doté de l'autonomie financière, placé sous la tutelle de la Délégation Générale de l'Armement.

Premier acteur français dans le domaine de la recherche aéronautique, spatiale et de défense, l'ONERA compte 2200 salariés (hors stagiaires), dont 1500 ingénieurs et cadres et en moyenne 360 doctorants. Par ailleurs, l'ONERA accueille en moyenne 200 à 220 stagiaires en cours d'études, principalement entre le mois de mars et le mois de juillet. Le budget de l'ONERA est de l'ordre de 330 M€ dont plus de la moitié proviennent des prestations de recherches contractuelles au profit de l'industrie et d'agences de l'Etat.

Les missions principales de l'ONERA sont les suivantes :

- Anticiper les ruptures technologiques pour préparer l'avenir
- Favoriser les transferts vers l'industrie
- Réaliser et mettre en œuvre des moyens d'expérimentation et de simulation
- Fournir à l'industrie des expertises de haut niveau
- Expertiser pour l'État les grands choix technologiques de demain
- Former des ingénieurs et des chercheurs

L'ONERA est implanté sur 8 sites en France comme indiqué sur la figure ci-dessous :



Les personnels scientifiques de l'ONERA sont répartis dans 7 départements de recherche et 3 départements liés aux moyens d'essais et d'ingénierie pour les maquettes utilisées en soufflerie. La plupart de ces départements sont multisites.

Le Calcul Intensif représente un enjeu important pour diverses activités de l'ONERA. La tendance des besoins est en forte croissance, que ce soit pour les simulations numériques dans les domaines liés à l'aéronautique et l'espace (aérodynamique, turbomachines, électromagnétisme, matériaux, ...), mais aussi avec l'appropriation des dernières méthodologies d'intelligence artificielle par les équipes scientifiques.

Dans ce cadre, l'ONERA met en œuvre des moyens de calcul intensif au profit des ingénieurs de recherche et doctorants, composés actuellement d'un cluster de production, un cluster de développement et de production spécifique et un cluster dédié aux calculs classifiés de défense (niveau « Secret »).

### **1.2. La Direction des Systèmes d'Information (DSI) de l'ONERA**

La Direction des Systèmes d'Information fournit les infrastructures et services informatiques nécessaires aux métiers des départements de recherche et aux directions fonctionnelles.

La DSI doit réaliser ces missions en veillant à la qualité de service, à la maîtrise des coûts et au respect des contraintes, en particulier celles liées à la sécurité.

Les domaines d'intervention de la DSI sont les suivants :

- Calculateurs scientifiques de production et de développement centraux et moyens de pré et post traitements associés ;
- Postes de travail scientifiques et bureautiques et moyens d'impression (Windows et Linux principalement) ;
- Infrastructures de stockage et de sauvegardes dédiées ou mutualisées ;
- Infrastructure réseau, sécurisation et infrastructure de communication ;
- Serveurs d'infrastructure techniques et scientifiques ;
- Applicatifs de communication et outils collaboratifs divers ;
- Informatique de gestion et système de restitution de données associé.

Une description plus détaillée de l'organisation de la DSI de l'ONERA figure dans l'annexe « ax1-orga-DSI-ONERA.docx » joint à ce document.

### **1.3. Objet de la prestation**

La DSI de l'ONERA est de plus en plus sollicitée pour la mise en œuvre, l'administration et le maintien en conditions opérationnelles et de sécurité des calculateurs, sans compter la gestion de l'assistance utilisateurs ou les projets visant à couvrir les nouveaux besoins. En plus de l'augmentation du volume d'activité, les contraintes en termes de cybersécurité sont aussi en croissance significative.

Les ressources internes de l'ONERA dédiées à ces activités étant limitées, une partie des prestations fournies par la DSI sont externalisées. À ce jour, cette partie ne concerne que des actes courants d'exploitation et d'administration ainsi que de l'assistance de niveau 1, qui sont confiés à un prestataire d'infogérance généraliste. Compte tenu de la croissance des besoins

et de la limitation des activités pouvant être fournies en interne, l'ONERA a décidé de mettre en place une prestation complémentaire sur ce périmètre.

Cette prestation, objet de la présente consultation, concerne donc des activités d'expertise et d'assistance spécifiques aux domaines Calcul Haute-Performance (HPC) et Intelligence Artificielle (IA) dont le but est de :

- Consolider le circuit de maintien en conditions opérationnelles (MCO) / maintien en conditions de sécurité (MCS) des infrastructures de Calcul Intensif,
- Prendre en charge les demandes complexes d'assistance de la part des utilisateurs concernant les infrastructures de Calcul Intensif et l'usage des codes de calcul ONERA dans ces environnements,
- Concevoir des solutions innovantes répondant à des problématiques précises de ces domaines (nouvelle fonctionnalité, sécurisation d'un périmètre ...).

Les modalités de cette prestation seront définies dans un accord cadre décomposé en 4 postes distincts correspondant aux différentes phases rencontrées pendant le déroulement du contrat :

- Poste 1 : forfaitaire, phase d'appropriation et d'initialisation de la prestation d'un mois, ce poste consiste à prendre connaissance de l'environnement et à mettre en place la prestation.
- Poste 2 : forfaitaire, ce poste couvre le niveau N2/N3 du circuit de MCO / MCS des infrastructures de Calcul Intensif ainsi que l'assistance utilisateurs avancée.
- Poste 3 : poste à marchés subséquents couvrant des prestations ponctuelles d'expertise pour répondre à une problématique donnée du domaine.
- Poste 4 (tranche optionnelle) : forfaitaire, phase de réversibilité de la prestation d'un mois, ce poste consiste à assurer le transfert de savoir-faire acquis lors de la prestation vers l'ONERA ou vers un nouveau prestataire.

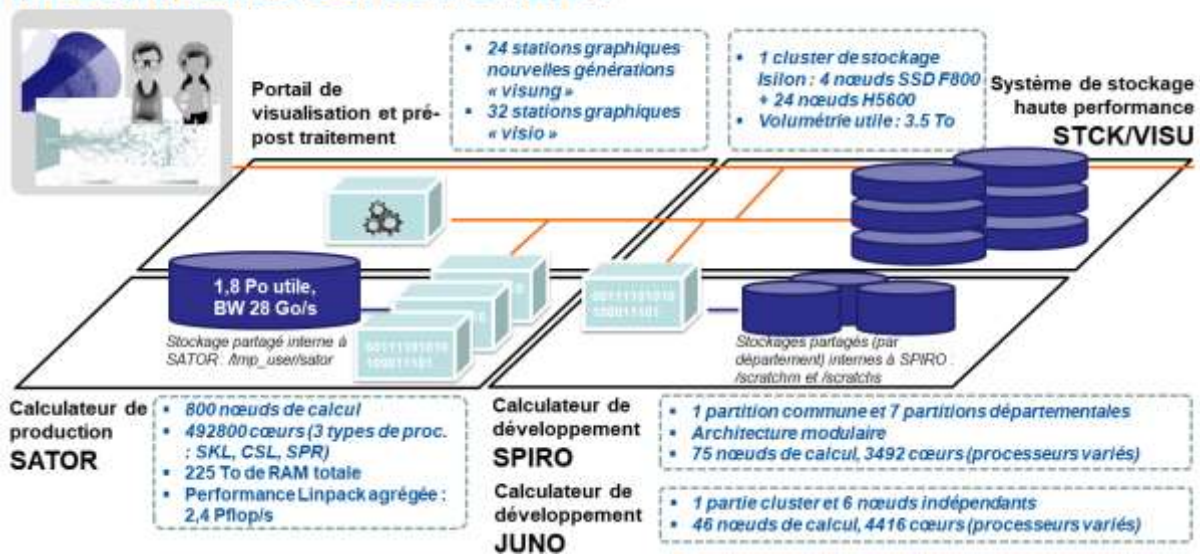
## 2. Environnement du Calcul Intensif à l'ONERA

Le périmètre du Calcul Intensif à l'ONERA repose principalement sur trois types de moyens centraux :

- Des ressources de calcul mutualisées, mises à disposition via des calculateurs ;
- Des ressources pour la visualisation et le pré-post traitement des données ;
- Une infrastructure de stockage mutualisée.

Pour accéder à ce périmètre, les utilisateurs doivent posséder un compte dit « CSI » pour Calcul Scientifique Intensif.

### Infrastructures CSI de l'ONERA



### 2.1. Ressources de calcul

L'ONERA dispose de 4 calculateurs centraux dont un en fin de vie. Les 3 calculateurs hébergés sur le réseau général (SI-E) sont situés actuellement à Châtillon et le calculateur classifié est situé à Palaiseau et connecté au nouveau réseau classifié (RSO).

Les calculateurs sont les suivants :

- Un calculateur de production, **SATOR**, constitué de plusieurs générations de matériel. Depuis le 2<sup>ème</sup> semestre 2024, SATOR est composé de 49 280 cœurs pour 800 nœuds de calcul, ce qui représente une puissance crête d'environ 2,4 Pflop/s.
- Un calculateur de développement en fin de vie, **SPIRO**, disposant d'une architecture modulaire : il s'agit d'un assemblage de partitions de calcul, composées soit de nœuds CPU, de nœuds GPU, de nœuds à grande mémoire. L'arrêt de ce calculateur de développement est prévu à la mi 2026.

- Un calculateur de développement, **JUNO**, remplaçant de la machine SPIRO, il reprend les concepts d'architecture modulaire. A partir de l'été 2025, il sera composé d'environ 60 nœuds de calcul de 96 cœurs.
- Un calculateur de production pour les travaux classifiés Secret, **SIROCCO**, en zone protégée à Palaiseau (donc non accessible depuis le réseau général mais accessible depuis le réseau RSO), dont l'architecture est identique à SATOR mais dans un format plus réduit : 108 nœuds de calcul, pour un total de 3984 cœurs.

## **2.2. Ressources de visualisation**

L'ONERA dispose également d'une infrastructure dédiée à la visualisation et au pré et post-traitement de données et constituée de différents types de stations à cartes graphiques. Ces machines graphiques sont centralisées dans le centre de calcul de Châtillon, au plus près des calculateurs et du stockage centralisé du CSI.

Leur accès s'effectue au travers du logiciel d'affichage déporté VirtualGL couplé avec TurboVNC, dont le client est installé sur les postes de travail sous la dénomination "Visualisation Distante CSI". Par rapport à un logiciel d'affichage déporté classique, cet outil permet d'accéder à l'accélération matérielle OpenGL fournie par les cartes graphiques présentes dans le centre informatique, sur les stations elles-mêmes.

## **2.3. Stockage centralisé du CSI**

En plus des espaces de stockage locaux à chaque calculateur, l'infrastructure de calcul intensif dispose également d'un espace de stockage de type NAS d'une capacité d'environ 3,5 Po (environ 1 Po utilisés), accessible via protocole NFS sur :

- L'ensemble des nœuds des calculateurs de développement,
- Les nœuds de connexion du calculateur de production,
- Les stations graphiques de pré/post-traitement,
- Des postes de travail scientifique (Linux),
- Les serveurs de fichiers « bureautiques » des départements (uniquement pour la mise à disposition de cet espace sur des postes Windows).



## **2.4. Architecture-type des calculateurs CSI et criticités**

Certains composants physiques d'un cluster ont une criticité plus élevée que d'autres. Même si ces composants critiques sont habituellement redondés, les délais d'intervention demandés en cas de dysfonctionnement sont plus courts.

A chaque composant du cluster est associée une criticité. Trois niveaux de criticité sont considérés :

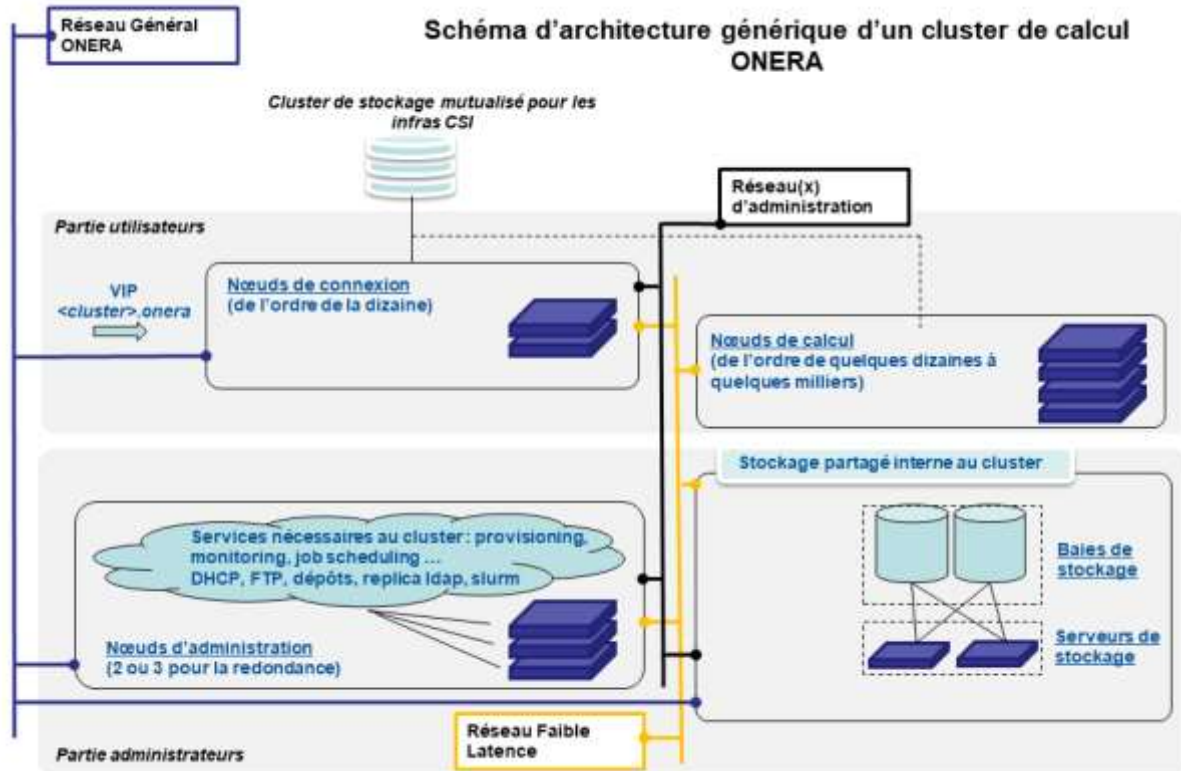
- Criticité 1 : équipement à faible criticité. La perte de plusieurs équipements de criticité 1 impacte uniquement la puissance de calcul du cluster.
- Criticité 2 : équipement à criticité moyenne. La perte d'un équipement unique n'est pas problématique mais si plusieurs équipements identiques sont en panne certains services critiques du cluster peuvent devenir instables ou inopérants.
- Criticité 3 : équipement à criticité élevée, la perte d'un de ces équipements doit être prise en charge sans attendre et résolue au plus vite.

L'architecture-type des clusters de l'ONERA est détaillée pour pouvoir faire correspondre une criticité à chaque composant. Cette architecture-type pourra évoluer, à la marge, au cours du temps selon les solutions proposées par les fournisseurs d'infrastructures HPC.

Elle se décompose comme suit (cf. schéma de référence ci-dessous) :

- Nœuds d'administration (généralement 2 ou 3) : ils portent la solution de déploiement des nœuds de calcul et l'ensemble des services nécessaires au bon fonctionnement du cluster. Selon la solution du fournisseur ces services s'exécuteront directement sur les nœuds d'administration ou de manière virtualisée ou conteneurisée sur ces nœuds. Les nœuds d'administration sont configurés en haute-disponibilité pour se prémunir de pannes ou d'incidents. Ces serveurs doivent tout de même être considérés comme critiques.
- Les serveurs de stockage (de 2 à 6 actuellement) : ils portent la solution de stockage parallèle et haute-performance qui est partagée et montée sur tous les nœuds du cluster. Le fonctionnement du cluster est résilient à la perte d'un de ces serveurs mais la performance du système de stockage s'en trouve significativement impactée. Par conséquent ces serveurs doivent être considérés comme critiques.
- Les baies de stockage (1 à 4) : elles sont résilientes à la perte de plusieurs disques, leur criticité est moyenne.
- Les commutateurs et liens du réseau faible latence (de quelques commutateurs à quelques dizaines) : la topologie du réseau faible latence utilisée est actuellement généralement du « fat tree » sur 2 niveaux ou plus. La criticité des commutateurs et liens dépend du niveau dans la topologie. Les composants du niveau edge ont une criticité faible alors que pour les autres niveaux la criticité est moyenne.
- Les nœuds de connexion (de l'ordre d'une dizaine) : ils servent de points d'entrée au calculateur pour les utilisateurs et sont configurés pour avoir de la redondance avec répartition de charge. La criticité des nœuds de connexion est moyenne.
- Les nœuds de calcul (de quelques dizaines à quelques milliers) : les clusters sont constitués d'un nombre élevé de nœuds de calcul, l'impact de la perte d'un nœud de calcul est négligeable. Ils sont de criticité faible.

Certains clusters ont des nœuds de visualisation légère, on considère qu'ils ont le même niveau de criticité que les nœuds de connexion. On entend par nœuds de visualisation légère, des nœuds de connexion munis de cartes graphiques destinés à faire du déport d'affichage vers le poste de travail de l'utilisateur. La solution utilisée actuellement à l'ONERA est développée en interne et se base sur TurboVNC et VirtualGL.



Les niveaux de criticités sont synthétisés ci-après pour les équipements d'un cluster de calcul, en utilisant un niveau de 1 à 3 :

Type d'équipement	Criticité
Nœud d'administration	3
Serveurs de stockage	3
Volumes de la baie de stockage	2
Disques de la baie de stockage	2
Commutateurs Réseau Faible Latence Core	2
Commutateurs Réseau Faible Latence Edge	1
Liens entre 2 commutateurs Core	2
Liens entre commutateurs Edge et Core	1
Nœuds de connexion	2
Nœuds de calcul	1 (*)

Les délais de résolution demandés en cas de panne ou d'incident diffèrent selon le niveau de criticité de l'équipement.

(\*) : L'indisponibilité de plus 10% de nœuds de calculs d'un cluster requalifie les incidents concernés en criticité 2, une valeur de 25 % les requalifie au niveau de criticité 3.

## **2.5. Taille du parc de calculateurs et pondérations associées**

Pour quantifier le parc de calculateurs et les charges associées, il est proposé de prendre en compte des critères qui sont représentatifs de la charge relative à l'exploitation, l'administration et l'assistance utilisateur sur chaque ordinateur. Deux critères viennent naturellement à l'esprit :

- Nombre d'utilisateurs, comptés en « utilisateurs actifs » mensuellement. En effet, plus le nombre d'utilisateurs qui travaillent sur un cluster est élevé, plus le volume de tickets d'assistance est susceptible d'augmenter. Est compté comme « utilisateur actif » un utilisateur s'étant connecté et ayant soumis au moins une tâche durant le mois dans le questionnaire de travaux du ordinateur concerné.
- La « taille » du cluster en nombre de cœurs de calcul. Le nombre de cœurs de calcul va traduire la taille globale du cluster et surtout le nombre d'équipements physiques qui le compose. Ce paramètre va avoir en particulier un impact sur le volume de tickets.

Il paraît assez réaliste de considérer des charges relatives via une formule du type suivant :

$$\text{poids ordinateur} = K_1 * (K_2 * \text{poids utilisateurs} + K_3 * \text{poids cœurs})$$

Cependant d'autres paramètres peuvent être pris en compte par le titulaire. Une description plus détaillée des calculateurs de production (SATOR) et de développement (JUNO) figure dans l'annexe « ax2-fiche-detaillée-infras-CSI.xlsx » joint à ce document.

Le titulaire doit proposer une formule de tarification pour le poste 2 dans un document de réponse dédié (type excel par exemple) où il explique en détail les paramètres utilisés. La formule applicable sera bien entendue celle proposée par le titulaire et validée.

Une évolution prévisionnelle des infrastructures est indiquée en annexe 3 du CCTP. Par ailleurs, un scénario d'évolution du parc de calculateurs pendant les 3 années fermes du contrat est proposé. Il est demandé d'estimer les tarifs du poste 2 pour ces 3 années en prenant en compte le scénario d'évolution décrit.

## **2.6. Projets d'évolution du SI ONERA**

Pendant la durée du présent accord-cadre, l'ONERA et son système d'information vont connaître des transformations majeures qui sont susceptibles d'impacter les prestations.

### **2.6.1 Projet de regroupement immobilier en Île-de-France (PRISME)**

L'ONERA prépare plusieurs années un projet de regroupement immobilier de ses 3 sites d'Île-de-France afin de simplifier son organisation, réduire ses dépenses de fonctionnement immobilier et bien entendu permettre aux équipes de travailler sur le même site.

Ainsi des bâtiments vont être construits à Palaiseau pour accueillir les équipes de Châtillon et de Meudon, les installations scientifiques associées et les infrastructures informatiques situées majoritairement à Châtillon (environ 400 serveurs, le cœur de réseau des liaisons intersites ONERA, les calculateurs intensifs de production et de développement).

Il est en particulier prévu de construire :

- Le bâtiment N2 : bâtiment tertiaire pour les fonctions support, en particulier la DSI.
- Le bâtiment N4 : à proximité immédiate du bâtiment N2, ce bâtiment sera le futur centre informatique central de l'ONERA (datacenter). Il comprendra, au rez-de-chaussée, toutes les servitudes électriques et froid (à l'exception de groupes situés en toiture), et au premier étage, les salles informatiques : calcul intensif, infrastructures générales, calcul classifié de défense et un local télécom.

Les bâtiments N2 et N4 seront livrés en principe fin 2026 et seront ensuite intégrés au SI de l'ONERA.

Il est prévu d'installer le futur calculateur de production, successeur de SATOR, dans la salle dédiée du bâtiment 4 à partir du 2<sup>ème</sup> trimestre 2027

### 2.6.2 RPO

L'ONERA héberge des données sensibles directement sur son SI-Entreprise, notamment des données à mention de protection « Diffusion Restreinte » (DR) ou « Diffusion Restreinte Spécial France » (DR-SF), ou encore des informations relevant du PSTN (Potentiel Scientifique et Technique de la Nation). Même si ces données sont la plupart du temps stockées de manière chiffrée, cette situation n'est pas satisfaisante car elle n'est pas conforme aux dispositions de l'instruction interministérielle 901 du 28/01/2015 relative à la protection des systèmes d'informations sensibles.

C'est d'ailleurs une situation qui impose à ce jour que tous les administrateurs du SI-Entreprise de l'ONERA soient ressortissants français et habilités au moins au niveau « Secret ».

Pour se mettre en conformité avec la réglementation et augmenter le niveau de confiance de ses clients, l'ONERA devait soit mettre à niveau son SI-Entreprise pour pouvoir l'homologuer au niveau DR, soit construire un nouveau SI protégé qui serait, par construction, homologable au niveau DR. Compte tenu des enjeux et des ressources à engager (pour rappel, le SI-E ONERA comprend plus de 5000 PC et plusieurs centaines de serveurs sur 8 sites), la décision a été discutée pendant assez longtemps, y compris avec la tutelle DGA et les instances de sécurité associées. Toutefois, il a été décidé en 2024 de financer et construire un nouveau SI protégé au niveau DR, baptisé « Réseau Protégé ONERA » (RPO).

Ce SI devrait être mis en place progressivement entre 2025 et 2028 pour une part des besoins métier, sur la base d'une infrastructure physique séparée du SI-Entreprise et multisites.

L'année 2024 a été consacrée à l'étude de besoins et la définition de l'architecture, notamment des SI de services communs nécessaires, du SI d'administration, et l'architecture intersites. Les premiers équipements devraient arriver en 2025 pour les services de base, puis il est prévu en 2026 de basculer le SI de contrôle d'accès existant sur l'infrastructure physique du RPO, et enfin, les premiers besoins métier devraient pouvoir s'appuyer sur le RPO à partir de 2027.

Il est prévu d'intégrer le calculateur de production et le calculateur de développement au RPO à horizon 2028 dans le cadre d'un projet spécifique.

### **3. Conditions d'exécution des prestations**

#### **3.1. Contexte existant**

Les infrastructures de Calcul Intensif de l'ONERA sont localisées sur deux de ses sites en Île-de-France : Châtillon et Palaiseau. Les équipes de la DSI de l'ONERA intervenant sur ces infrastructures opèrent actuellement depuis le site de Châtillon. Un déménagement de ces équipes et de toutes les infrastructures de Calcul Intensif est prévu sur le site de Palaiseau en 2027 et 2028.

#### **3.2. Périmètre des prestations**

Les prestations concernent tous les calculateurs de l'ONERA à l'exception du calculateur dédié aux études classifiées Secret et du calculateur de développement en fin de vie SPIRO.

Dans le périmètre des prestations, il y a au moins un calculateur de production et un calculateur de développement, avec parfois des périodes de biseau en cas de renouvellement.

Les calculateurs du périmètre sont actuellement connectés au réseau général de l'ONERA mais à partir de 2028, comme indiqué au paragraphe 2.6.2, ils seront transférés dans un système d'information en cours de construction (RPO : Réseau Protégé de l'ONERA) prévu pour héberger les données sensibles de l'ONERA et qui fera l'objet d'une homologation au niveau « Diffusion Restreinte ». L'administration de ces infrastructures comporte des contraintes spécifiques de sécurité qui s'appliqueront au titulaire dès à présent dans la mesure où les calculateurs existants sont susceptibles d'héberger des données sensibles à mention de protection « Diffusion Restreinte » ou « Diffusion Restreinte Spécial France ».

#### **3.3. Localisation des prestations**

Pour le poste 1, en phase d'initialisation, un transfert de compétences sera organisé entre la DSI de l'ONERA et le titulaire. Cette phase aura lieu sur le site de l'ONERA à Châtillon.

Les prestations pourront être exécutées dans les locaux du titulaire (équipe mutualisée sur plateforme) et/ou sur le site ONERA de référence pour le poste 2 (MCO / MCS / assistance utilisateur avancée) et le poste 3 (prestations sur demande). Le site ONERA de référence est initialement Châtillon, puis, lorsque les équipes de la DSI de l'ONERA auront déménagé à Palaiseau, le site de référence pour le titulaire deviendra Palaiseau.

Si les prestations sont conduites dans les locaux du titulaire, la connexion au SI de l'ONERA doit respecter les conditions spécifiées au paragraphe 5 du présent document.

Certaines activités précises du poste 2, comme par exemple les maintenances avec arrêt de production, pourront nécessiter une présence obligatoire sur le site ONERA de référence.

Pour le poste 4, la phase de réversibilité, si elle est effectuée, se déroulera sur le site ONERA de référence.

### **3.4. Plage de service de la prestation**

La prestation est demandée :

- Les jours ouvrés, c'est-à-dire habituellement 5 jours sur 7 ;
- De 9h à 18h.

L'ONERA ferme une semaine tous les mois d'août et une semaine fin décembre.

Il est demandé d'avoir la possibilité de réaliser quelques opérations planifiées en dehors de ces plages, typiquement entre 18h et 21h ou durant les week-ends, pour un maximum de 15 h par an.

### **3.5. Organisation du MCO/MCS sur le périmètre Calcul Scientifique Intensif**

#### **3.5.1 Les acteurs actuels du périmètre**

##### Infogérance généraliste

Un certain nombre d'activités est externalisé dans le cadre d'une prestation d'infogérance généraliste pilotée par la DSI de l'ONERA.

Cette prestation est organisée en différentes entités :

- Helpdesk : point d'entrée des demandes, anomalies ou incidents rapportés au support
- Intervenants de proximité : ils sont en charge de réaliser les actes techniques en salle machine. Pour le périmètre CSI, cela peut être changer du matériel défaillant, effectuer des contrôles visuels sur le matériel, vérifier l'état d'un câblage... Cette partie de la prestation d'infogérance est actuellement réalisée par la société Akkodis.
- Intervenants support N1 : il s'agit des intervenants de l'équipe « infrastructure » de l'infogérance généraliste, actuellement réalisée par la société Capgemini. Ils sont en charge de l'exploitation courante, des demandes procédurées et de la gestion des incidents niveau 1.

##### Le pôle Calcul Scientifique Intensif (Service ISR de la DSI de l'ONERA)

Les intervenants du pôle CSI sont en charge du socle système (matériel, pilotes, système d'exploitation, logiciels spécifiques au HPC : Slurm, IB/OPA, GPFS, MPI ...)

Jusqu'à présent les intervenants du pôle CSI traitent les demandes complexes, les incidents N2/N3, les problèmes et la plupart des changements du périmètre. C'est cette partie de l'activité qui doit être transférée dans le cadre de cette prestation.

##### Le pôle Outils et Composants Linux (Service OST de la DSI de l'ONERA)

Les intervenants du pôle sont en charge de configurer, compiler, packager les composants de l'environnement logiciel scientifique et de les déposer dans les dépôts adéquats.

Le déploiement des paquets scientifiques est automatisé pour les calculateurs de développement et il est effectué de manière manuelle par les intervenants du pôle CSI sur le calculateur de production et le calculateur classifié sur demande du pôle Outils et Composants Linux.



### 3.5.2 Le comité de suivi d'exploitation du CSI

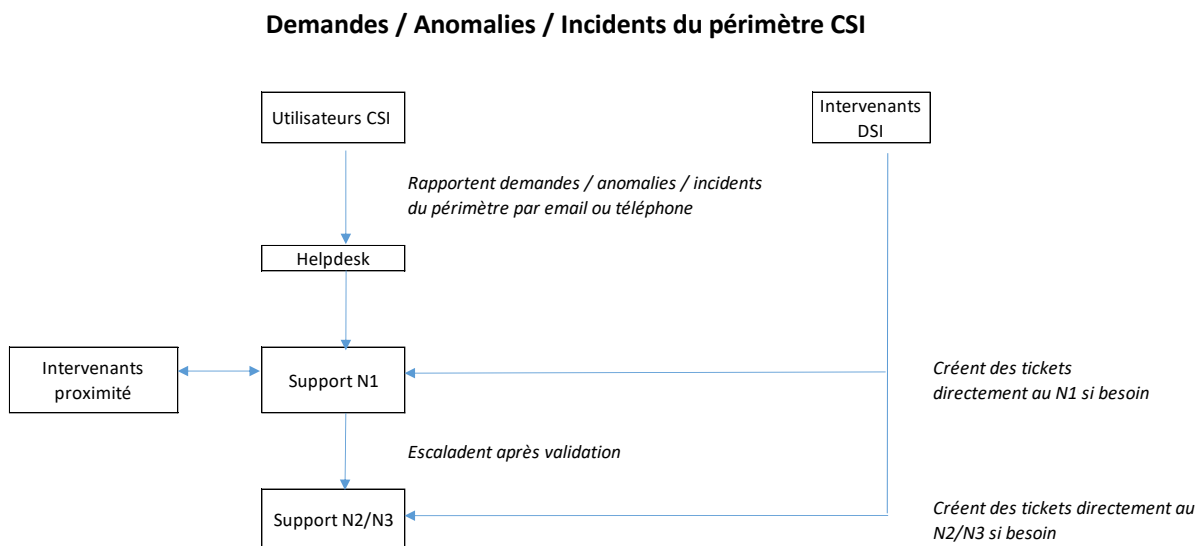
#### Comité hebdomadaire CSI :

Un point hebdomadaire d'environ 1h est réalisé entre les différents intervenants du domaine afin :

- D'identifier d'éventuels tickets bloqués ou points de blocage,
- D'analyser les incidents récurrents pour mise en place d'un plan d'action (problèmes, changements) ;
- De lancer les actions techniques spécifiques le cas échéant ;
- De discuter des évolutions de modes opératoires ou de consignes le cas échéant ;
- D'établir des plans d'action et des délais associés aux tâches en cours ;
- D'arbitrer la requalification des tickets d'assistance et éventuellement les problèmes de fond ou problématique sur le domaine.

Ce point hebdomadaire rassemble a minima un représentant opérationnel de chaque entité intervenant sur le domaine : infogérance généraliste, prestation d'expertise et DSI.

### 3.5.3 Organisation cible du support



Les demandes ou incidents rapportées par les utilisateurs du périmètre sont toutes centralisées par le helpdesk qui est en charge de les orienter vers le support N1.

Les demandes procédurées sont directement traitées par le N1.

Dans le cas d'une anomalie ou d'un incident, le support N1 effectue une première analyse avec qualification du niveau de criticité.

Dans le cas d'une panne matérielle, le support N1 pilote les interventions de remplacement de pièce si besoin.

Dans le cas d'une anomalie ou d'un incident concernant un des aspects suivants, le support N1 va escalader au support N2/N3, objet de cette consultation, après validation systématique auprès de la DSI :

- Slurm, aspects logiciels du réseau faible latence ou de la solution de stockage partagée,
- Exécution d'un code,
- Problème de performance sans problème matériel identifié.

Le fait de valider systématiquement les escalades du N1 vers le N2/N3 par la DSI pourra être réévalué une fois que les périmètres d'intervention seront bien intégrés par l'ensemble des équipes du domaine.

### 3.5.4 Répartition cible des rôles et responsabilités

Le tableau RACI ci-dessous montre la répartition des rôles et responsabilités entre les différents types d'intervenants.

La mention « R » indique que l'intervenant est responsable de l'action. « A » désigne l'approbateur si cela est nécessaire. « C » veut dire consulté et « I » informé.

La mention « N/A » signifie « non applicable ».

	Activité	Proxi / N1	N2 – N3	DSI
Transverse	Gestion comptes CSI	R	N/A	N/A
	Gestion quotas CSI	R	N/A	A
Logistique	Gestion des stocks	R	I	A
	Gestes de proximité	R	I	A
	Accompagnement interventions	R	I	A
Pilotage	Météo des calculateurs 2 fois / j	R	I	I
	Suivi d'indisponibilité	R	I	I
	Vérification sauvegarde	R	N/A	I
Gestion d'incident	Premier diagnostic anomalie / incident	R	I	I
	Traitement incident N1	R	I	A
	<b>Traitement incident N2/N3</b>	<b>I</b>	<b>R</b>	<b>A</b>
Gestion de problème	<b>Analyse d'un problème</b>	<b>I</b>	<b>R</b>	<b>C</b>
Demandes	<b>Mise à jour logiciel sur nœud d'administration</b>	<b>I</b>	<b>R</b>	<b>A</b>



	Mise à jour logiciel nœud utilisateur	I	R	A
	Configuration gestionnaire de batch	I	R	A
	Mise à jour gestionnaire de batch	I	R	A
	Configuration solution de stockage partagé	I	R	A
	Mise à jour solution de stockage partagé	I	R	A
	Configuration réseau faible latence	I	R	A
	Mise à jour solution de réseau faible latence	I	R	A
Maintenance	Arrêts / démarrages complets	R	A	C
	Vérification arrêts / démarrages complets	I	R	C
	Préparations maintenance N1	R	I	A
	Préparations maintenance N2/N3	I	R	A
	Maintenance N2/N3	I	R	A
Outillage	Amélioration outillage interne / scripts	I	R	A
Gestion documentaire	Rédaction / mise à jour DEX	R	C	A
	Rédaction / mise à jour DAT	I	R	A
	Rédaction / mise à jour documentation utilisateurs	I	C	R
	Rédaction / mise à jour guide de sécurisation	I	C	R
Assistance	Assistance utilisateur de base	R	I	I
	Assistance utilisateur approfondie	I	R	I
Support applicatif	MCO Applicatif	I	R	I
	Gestion Compilation	I	R	I
	Gestion Optimisation	I	R	I

### 3.5.5 Accès « administrateur » au périmètre

Les accès « administrateurs » sont partagés entre les différentes entités intervenant dans l'organisation du support du périmètre CSI :

- Support N1
- Support N2/N3
- Intervenants DSI : ISR/CSI et OST/Linux.

Chaque intervenant possède un compte « administrateur » individuel.

### **3.6. Compétences requises**

Le titulaire propose différents profils d'intervenants lui permettant de répondre aux besoins exprimés dans ce CCTP.

Le titulaire tient à jour une matrice de compétences pour l'ensemble de ses intervenants. Un format de matrice de compétences est fourni, le titulaire peut éventuellement ajouter des compétences qui ne sont pas listées, il ne doit pas en enlever.

Le titulaire peut proposer des profils supplémentaires s'il le juge utile pour son activité.

#### **3.6.1 Compétences communes**

Les compétences générales attendues pour l'ensemble des profils sont :

- bonne capacité de communication écrite et orale,
- bonne capacité rédactionnelle,
- capacité d'analyse méthodique et sens relationnel,
- capacité à travailler en français (oral et écrit) et en anglais (écrit a minima).

#### **3.6.2 Profil « Ingénieur système HPC »**

Ce profil possède des bases d'administrateur système Linux, tout en ayant quelques années d'expérience d'exploitation d'environnements de type cluster HPC en production.

#### **3.6.3 Profil « Architecte HPC et/ou IA »**

Ce profil possède a minima toutes les compétences techniques d'un ingénieur système HPC. Il a plus de 5 ans d'expérience dans le domaine.

Il est capable d'analyser les besoins scientifiques et de concevoir des solutions complexes prenant en compte ces besoins et les contraintes de l'ONERA, de les mettre en œuvre en respectant les délais et de les documenter.

#### **3.6.4 Profil « Expert code HPC et/ou IA »**

Ce profil possède des bases de développeur informatique avec des connaissances approfondies sur les techniques de parallélisation et/ou l'apprentissage autonome.

## 4. Déroulement du contrat

### 4.1. Poste 1 : phase d'initialisation

La phase d'initialisation a pour objectif de faire découvrir l'environnement technique et fonctionnel de la prestation au titulaire. Elle s'effectue sur le site ONERA Châtillon pendant un mois.

#### 4.1.1 Prise de connaissance de l'existant

L'objectif premier de cette phase est la prise de connaissance de l'existant. L'ONERA s'assure de fournir la documentation existante, les accès informatiques aux infrastructures concernées et un accompagnement lors de cette période.

Cette prise de connaissance porte en particulier sur :

- l'architecture des calculateurs du périmètre,
- les environnements logiciel de production et de développement,
- l'outillage transverse au périmètre et les procédures d'exploitation en place,
- les principaux codes et benchmarks de référence utilisés par l'ONERA sur ses infrastructures.

#### 4.1.2 Livrables

A l'issue de cette phase, la documentation technique (dossiers d'architecture, dossiers d'exploitation, documentation utilisateurs) sera annotée par le titulaire de propositions d'amélioration.

Il fournit un bilan de la phase d'initialisation : document formalisant l'état de départ de la prestation et recensant les préconisations du titulaire.

Par ailleurs, il propose les versions initiales des livrables suivants à la fin de cette phase :

- le plan qualité de service (PQS),
- la convention de service (CS),
- le plan de réversibilité (PR).

Les versions finalisées du plan qualité de service (PQS), de la convention de service (CS) et du plan de réversibilité (PR) sont attendues 4 mois après le démarrage du poste 1.

#### Plan qualité de service (PQS)

Le PQS devra notamment détailler :

- l'organisation de la relation entre le titulaire et l'ONERA,
- les procédures de prise en compte d'incidents, de problèmes, de demandes de changement ou d'assistance utilisateurs,
- la gestion de la documentation,

- la périodicité des réunions prévues,
- le processus d'intégration des nouveaux arrivants sur le projet,
- le processus de maintien des compétences et contrôles associés.

Le titulaire s'engage à effectuer toutes les mises à jour nécessaires au cours du contrat, le cas échéant, afin que le PQS soit le reflet de la prestation réalisée. Toute modification du PQS est soumise à validation en comité de pilotage.

#### Convention de service (CS)

Le titulaire précise dans la CS les engagements pris pour chaque prestation due au titre de cet accord-cadre et en définit les niveaux de service contractuels attendus.

#### Plan de réversibilité (PR)

Même si la phase de réversibilité est optionnelle, un document récapitulant les modalités de cette phase (tâches principales, planning, référent en charge de la réversibilité) sera fourni. A ce stade il ne s'agit que de la structure du document qui sera complété au fur et à mesure.

### **4.2. Poste 2 : MCO/MCS/assistance utilisateur avancée**

Actuellement le support sur les infrastructures de Calcul Intensif se base sur l'intervention d'une infogérance généraliste, non spécialisée, pour les activités récurrentes et procédurées. Les analyses approfondies ainsi qu'une grande partie des MCO/MCS sont réalisées par les équipes internes de la DSI de l'ONERA.

L'objectif du poste 2 de cet accord cadre est donc de créer un nouveau niveau de support externalisé avec expertise et permettant d'alléger la charge des équipes DSI sur l'exploitation des infrastructures de Calcul Intensif et l'assistance avancée de leurs utilisateurs.

Cette partie forfaitaire se compose principalement des activités suivantes :

- installation, configuration ou mise à jour d'un gestionnaire de batch (Slurm),
- installation, configuration ou mise à jour d'un réseau faible latence (InfiniBand, OPA...),
- installation, configuration ou mise à jour d'une solution de stockage partagé (GPFS, Lustre...),
- maintenance du calculateur : mise à jour des composants (firmware, logiciels), montée de version mineure du système d'exploitation,
- création et amélioration de scripts d'exploitation, d'outils internes,
- rédaction ou mise à jour de la documentation technique (dossier d'architecture, d'exploitation, documentation utilisateur, guide de sécurisation),
- traitement des incidents escaladés par l'infogérance,
- participation aux analyses pour la résolution des problèmes,
- assistance aux utilisateurs sur demande complexe.

#### 4.2.1 Description détaillée des prestations techniques attendues

Les actes d'exploitation sont tracés et comptabilisés via l'outil de ticketing de la DSI de l'ONERA, actuellement Qualiparc. Les actions des projets et des analyses plus longues, pour la gestion de problème par exemple, sont tracés dans un projet Redmine du périmètre CSI.

##### Traitement incident N2/N3

Le titulaire gère les incidents N2/N3 qu'ils soient d'origine matérielle ou logicielle.

Le niveau des incidents dépend de la complexité et des aspects concernés, ils sont escaladés au support N2/N3 après validation de la DSI.

Dans tous les cas, un incident doit être pris en charge dans les 2h après son affectation ou escalade.

Les délais de résolution d'incident demandés (à partir de l'arrivée du ticket dans le groupe de traitement du titulaire) dépendent de la criticité de la panne ; trois niveaux de criticité sont considérés comme indiqué ci-dessous.

Degré d'impact aux utilisateurs	Criticité	Délai de résolution (h. ouvrées)
Le service est indisponible ou significativement dégradé pour plus de 25% des utilisateurs.	3	12 h
Le service est dégradé pour un nombre d'utilisateurs restreints.	2	24 h
Une anomalie est constatée pour un utilisateur.	1	72 h

##### Analyse d'un problème

Le titulaire identifie les incidents et les dysfonctionnements récurrents pénalisant le service et les remonte à l'ONERA pour apprécier leur qualité de problème lors du comité hebdomadaire CSI.

La recherche de la cause racine d'un problème peut prendre du temps. A chaque comité hebdomadaire CSI les problèmes en cours sont parcourus et les actions à mener d'ici le comité suivant sont discutées entre le titulaire et la DSI de l'ONERA. Les délais des actions dépendent de la criticité du problème et sont donnés à titre indicatif.

##### Mise à jour logiciel sur nœud d'administration

Le titulaire met à jour un composant logiciel sur un nœud d'administration.

##### Mise à jour logiciel sur nœud utilisateur

Le titulaire met à jour un composant logiciel ou un méta-paquet dans les images utilisées sur les nœuds de calcul des clusters.

#### Configuration initiale ou modification de la configuration du gestionnaire de batch

Le titulaire définit la configuration initiale ou modifie la configuration existante du gestionnaire de batch en fonction du besoin exprimé par l'ONERA.

Le titulaire peut être amené à effectuer des recherches dans la documentation du gestionnaire de batch pour être force de proposition, choisir une implémentation et appliquer sa mise en production.

#### Mise à jour logiciel du gestionnaire de batch

Le titulaire prépare la mise à jour du gestionnaire de batch en consultant les *release notes* de la version cible. Il effectue une analyse d'impact et en communique une synthèse à l'ONERA.

Il planifie, organise et réalise la montée de version du gestionnaire de batch à chaud ou en arrêt de production sur l'ensemble des nœuds du cluster.

Le titulaire sauvegarde l'état de la base de données du gestionnaire de batch puis le met à jour. Il s'assure du bon fonctionnement du gestionnaire de batch une fois la mise à jour effectuée.

#### Configuration initiale ou modification de la solution de stockage partagé

Le titulaire apporte son expertise haut-niveau sur la solution de stockage partagée, il est force de proposition concernant l'optimisation de la configuration et produit des recommandations de paramétrage.

Il crée ou migre des systèmes de fichiers partagés. Il déploie de nouvelles fonctionnalités (ex : snapshots, quota, export NFS...)

#### Mise à jour logiciel de la solution de stockage partagé

Le titulaire prépare la mise à jour de la solution de stockage partagé en consultant les *release notes* de la version cible. Il effectue une analyse d'impact et en communique une synthèse à l'ONERA.

Il planifie, organise et réalise la montée de version du système de fichiers partagé en minimisant l'arrêt de production sur l'ensemble des nœuds du cluster (clients et serveurs).

#### Configuration initiale ou modification de la configuration du réseau faible latence

Le titulaire apporte son expertise haut-niveau sur le réseau faible-latence. Il optimise sa configuration et fait des recommandations concernant son paramétrage.

Il est amené à déployer de nouvelles fonctionnalités et à optimiser les paramétrages pour améliorer les performances du calculateur (MPI, RDMA, GPU Direct ...)

### Mise à jour logiciel de la solution de réseau faible latence

Le titulaire prépare la mise en jour des logiciels de la solution de réseau faible latence en vérifiant les release notes, analysant les impacts sur l'exploitation.

Il planifie, organise et réalise la montée de version de la solution utilisée pour le réseau faible latence (pilotes, logiciels de la solution) en minimisant l'arrêt de production sur l'ensemble des nœuds du cluster.

### Maintenance

Plusieurs maintenances planifiées peuvent avoir lieu chaque année, donnant lieu à la mise hors-production d'une partie ou de l'intégralité d'un calculateur.

C'est l'occasion de mettre à jour un maximum de composants du cluster : en allant des micrologiciels du matériel au système d'exploitation des nœuds d'administration et/ou de calcul (mise à jour mineure pour des raisons de sécurité).

Le titulaire participe à l'opération de maintenance et réalise les actions. Pour des raisons pratiques, il est demandé à l'ensemble des intervenants de la maintenance d'être présents sur le site ONERA de référence.

### Préparations de maintenance N2/N3

Le titulaire interagit avec la DSI de l'ONERA pour définir le plan d'action de maintenance.

Le titulaire liste les prérequis et établit le chronographe prévisionnel des tâches dont il est responsable.

### Vérification de l'arrêt / démarrage d'un cluster

Le titulaire supervise l'exécution de la procédure d'arrêt ou de démarrage d'un cluster par les intervenants du support N1. Des points de vérification entre le titulaire et les intervenants du N1 concernés doivent être planifiés en amont des phases critiques de ces procédures.

### Amélioration d'outils ou scripts internes

Le titulaire développe ou améliore les scripts d'exploitation existants. De par sa maîtrise des langages usuels (python, shell...), il met en place des outils qui permettent d'optimiser ou d'automatiser l'exploitation des clusters.

### Rédaction initiale ou mise à jour d'un document technique de référence

Les documents techniques de référence pour chaque cluster du périmètre CSI sont les suivants :

- Dossier d'architecture : il est rédigé sur la base de la documentation fournie par le constructeur à la livraison du cluster et décrit le fonctionnement du cluster ;
- Dossier d'exploitation : il contient l'ensemble des procédures d'exploitation et des informations nécessaires au maintien en condition opérationnelle. Ce document est enrichi tout au long de la durée d'exploitation du cluster ;

- Documentation utilisateur : ensemble de pages intranet décrivant un cluster et son utilisation ; les informations listées vont des spécifications techniques du matériel à l'environnement logiciel en passant par les modalités pour exécuter des calculs ;
- Guide de sécurisation : il recense les procédures à appliquer pour sécuriser un cluster.

Le titulaire rédige ou met à jour les dossiers d'architecture, notamment lors des montées de version ou d'évolution de l'outillage.

Le titulaire est consulté à propos du contenu du dossier d'exploitation dédié au support N1 et aux intervenants de proximité. Il complète le document de procédures ou mémos dédiés aux intervenants N2/N3. En cas d'évolution des procédures le concernant, il réalise les mises à jour.

Le titulaire est consulté concernant le contenu de la documentation utilisateur et du guide de sécurisation d'un cluster.

#### Assistance utilisateur approfondie

Tous les tickets relevant de l'assistance aux utilisateurs et du périmètre CSI reçus au helpdesk sont d'abord qualifiés d'assistance utilisateur initiale à destination du support N1. Les demandes simples couvrent des sujets variés allant de la gestion des espaces de stockage, la soumission des jobs dans SLURM, le mode opératoire à suivre pour le déport d'affichage...

Si la demande de l'utilisateur nécessite des connaissances spécifiques ou une analyse plus approfondie, le support N1 se concentre d'abord, dans le cadre de cette assistance initiale, sur la collecte d'information, et s'assure que tous les éléments nécessaires à l'analyse ont été fournis par l'utilisateur.

Une fois ces précisions apportées, le ticket peut être requalifié si besoin en assistance approfondie à destination du support N2/N3 avec validation de l'ONERA lors du point hebdomadaire CSI.

#### Maintien en condition (MCO) des codes de calcul

Le titulaire a accès à un référentiel de cas de référence utilisés en interne ONERA. Ces cas sont régulièrement exécutés sur les clusters, notamment avant et après les opérations de maintenance. Leurs résultats sont analysés. Toute régression de plus de 5% donne lieu à une investigation sur la cause de l'écart.

#### Gestion de compilations

Le titulaire analyse les options de compilation disponibles pour un code de calcul sur une architecture donnée. Il fait des préconisations selon l'objectif spécifié (précision, performance ...) et illustre sur les cas précis qui lui sont fournis.

#### Gestion d'optimisations

Le titulaire réalise une analyse de profilage d'un cas donné et identifie les contentions et pistes d'optimisation en termes de performances.



#### 4.2.2 Eléments de volumétrie

	Activité	Volumétrie cluster PROD / an	Volumétrie cluster DEV / an
Gestion d'incident	Traitement incident N2/N3	3 tickets x 12 mois	3 tickets x 12 mois
Gestion de problème	Analyse d'un problème	60 h	60 h
Demandes	Mise à jour logiciel sur nœud d'administration	3 x 5 h	3 x 5 h
	Mise à jour logiciel nœud utilisateur	30 x 3 h	5 x 3h
	Configuration gestionnaire de batch	5 x 10 h	5 x 10 h
	Mise à jour gestionnaire de batch	1 x 20 h	1 x 20 h
	Configuration solution de stockage partagé	4 x 20 h	4 x 20 h
	Mise à jour solution de stockage partagé	1 x 40 h	1 x 40 h
	Configuration réseau faible latence	4 x 20 h	4 x 20 h
	Mise à jour solution de réseau faible latence	1 x 20 h	1 x 20 h
Maintenances	Préparations maintenance N2/N3	1 x 40 h	1 x 40 h
	Maintenances N2/N3	1 x 20 h	1 x 20 h
Outillage	Amélioration outillage interne / scripts	60 h	60 h
Gestion documentaire	Rédaction / mise à jour document	60 h	60 h
Assistance	Assistance utilisateur approfondie	5 tickets x 12 mois	10 tickets x 12 mois
Support applicatif	MCO Applicatif	60 h	60 h
	Gestion Compilation	60 h	60 h
	Gestion Optimisation	60 h	60 h

### 4.2.3 Qualité et indicateurs contractuels

#### Qualité de service

L'ONERA est attaché à la qualité de ses recherches et de ses fournitures techniques et la plupart des processus internes participant à son activité sont intégrés à un système de management de la qualité. Ainsi, l'ONERA a été l'un des premiers organismes de recherche français à être certifié ISO 9001 pour la conduite d'études et de recherches, preuve de son engagement pour une gestion de la recherche par projet selon les meilleurs standards industriels. Dans le cadre de cette certification, les processus de soutien informatique sont régulièrement audités.

Aussi la DSI est attentive à la qualité de service mesurable au travers d'indicateurs adaptés à ses principales attentes. Dans le cadre du poste 2 de ce contrat, les indicateurs contractuels sont :

- le taux mensuel de réouverture des tickets,
- et le taux mensuel de traitement en conformité avec les délais.

#### Notion de réouverture à la suite d'un traitement

Les intervenants techniques du titulaire ont à la responsabilité de placer les tickets directement dans un état « résolu » lorsqu'ils estiment être certains que la solution proposée fonctionne et qu'ils ont pu, si possible, la tester en situation.

L'analyse de l'intervenant peut cependant être contestée par l'utilisateur ou par la DSI, ce qui peut conduire à 2 états particuliers existant dans le workflow des tickets :

- le premier état s'appelle « réouverture demandée » ; lorsqu'un acteur (utilisateur, la DSI, le titulaire) signale qu'un ticket a été traité incorrectement, ce dernier est placé dans cet état,
- le second état s'appelle « rouvert » ; un ticket est placé dans cet état s'il est jugé que la demande de réouverture est justifiée, après arbitrage technique par un groupe de gouvernance commun titulaire-ONERA. Dans le cas contraire, l'utilisateur – ou le demandeur de la réouverture - est informé de la raison du refus de réouverture et le ticket est placé à nouveau dans l'état « résolu ».

Le nombre de tickets « rouverts » dans le mois en cours, rapporté au nombre de tickets résolus dans le mois, après correction des cas dans lesquels l'intervenant n'est pas en cause, est un élément significatif de la qualité de service de la DSI et de ses prestataires. Il est appelé « taux de réouverture » et calculé mensuellement.

Le titulaire s'engage à que ce taux de réouverture soit inférieur à 5%. Ce taux sera calculé mensuellement.

#### Délais de résolution

Cet indicateur ne concerne que le traitement des incidents N2/N3 pour lesquels le titulaire s'engage à respecter des délais conformément à ce qui est indiqué au paragraphe 4.2.1.

Le compteur de temps de traitement d'un ticket devient actif lorsque le ticket arrive dans le groupe de traitement du titulaire. Il est arrêté lorsque le ticket est passé à l'état « résolu ».

Un taux de traitement d'incidents en conformité avec les délais contractuels est calculé mensuellement. Il est obtenu en effectuant le rapport entre les tickets marqués « résolu » dans les délais dans le mois et le nombre total de tickets « résolu » dans le mois.

Le titulaire s'engage à ce que ce taux de traitement d'incidents dans les délais, calculé mensuellement, soit supérieur à 95%.

#### **4.2.4 Comité de pilotage**

Une réunion de pilotage du contrat se tient tous les 3 mois. Cette réunion recense les faits marquants de la période écoulée, et est l'occasion d'une revue des indicateurs de mesure de la qualité. Elle permet aussi d'anticiper toute évolution impactant le fonctionnement de la prestation.

Sa fréquence peut être revue à tous les 2 mois à la demande de l'ONERA selon l'activité ou le besoin.

Le comité de pilotage réunit au moins les personnes suivantes :

- Le chef du service « Infrastructure, Systèmes et Réseaux » et le responsable de pôle « Calcul Scientifique Intensif » de la DSI de l'ONERA
- le responsable contractuel du titulaire,
- le ou les responsables opérationnels des prestations.

Le directeur de la DSI ou son adjoint pourront être conviés ponctuellement si besoin.

Les points suivants doivent obligatoirement figurer à l'ordre du jour de chaque réunion, sur la base des documents fournis :

- présentation des états statistiques et des indicateurs de la période écoulée,
- revue des faits marquants et incidents majeurs de la période écoulée,
- revue des éléments de facturation.

Les éléments statistiques devront être fournis par le titulaire pour validation au plus tard 3 jours ouvrés avant le comité de pilotage.

Le comité de pilotage validera toutes les évolutions de périmètre ayant un impact sur la facturation.

Un compte rendu de réunion est rédigé et proposé par le titulaire, validé par les participants avant diffusion.

### **4.3. Poste 3 : prestations d'expertise ponctuelles sur demande**

En complément des prestations récurrentes, le besoin technique concerne également des prestations complémentaires ponctuelles sur demande pour un volume significativement inférieur.

Ces prestations sont toutes exécutées dans le contexte du Calcul Intensif ou de l'Intelligence Artificielle. Les sous-thèmes considérés pour ce type de prestations sont listés de manière non-exhaustive ci-dessous :

- sécurité d'infrastructures type ordinateur HPC ou de cluster de stockage,
- optimisation de l'architecture et de la configuration d'un ordinateur HPC,
- optimisation de l'architecture et de la configuration d'un cluster de stockage,
- optimisation de l'usage des architectures de processeurs ou coprocesseurs (CPU, GPU...) pour un code de calcul donné,
- mise en place d'outillage pour les usages d'Intelligence Artificielle...

Le mécanisme de déclenchement des prestations est décrit dans l'accord cadre. Il sous-entend, dans tous les cas, la description du besoin technique par l'ONERA et la fourniture d'une proposition par le titulaire, incluant un devis évaluant le volume d'activité associé à chaque profil technique nécessaire.

Les paragraphes suivants proposent des exemples de synoptiques de prestations envisagées. Le titulaire s'engage à pouvoir répondre a minima à des prestations proches de celles présentées.

#### **4.3.1 Prestation de mise en œuvre d'un portail d'accès au ordinateur de développement JUNO**

La solution Open OnDemand fournit un portail web pour que les utilisateurs puissent gérer leurs fichiers et travaux sur les ordinateurs.

L'objectif de cette prestation est de définir une architecture permettant l'hébergement de la plateforme Open OnDemand pour un accès au ordinateur de développement junos, ainsi que sa configuration en fonction des besoins scientifiques. Le titulaire doit ainsi lister les cas d'usage et codes principaux envisagés et de s'assurer de l'intégration de ces usages dans le portail.

L'architecture préconisée sera décrite dans un dossier d'architecture puis validée par la DSI de l'ONERA.

Une fois l'architecture validée, l'installation de la solution pourra être commencée. A l'issue de l'installation et de la configuration, une phase de tests sera effectuée conjointement avec la DSI pour vérifier l'aptitude de la solution.

Une fois la phase de tests finie, les procédures d'exploitation ainsi que le guide d'utilisation du portail, répertoriant les cas d'usage, doivent être livrés.

#### **4.3.2 Analyse de l'usage d'un système de fichiers, préconisations de configuration et de solution nouvelle**

Le cluster de production SATOR est composé d'une solution de stockage interne au calculateur qui est parallèle et distribué sur l'ensemble des nœuds, basé sur la solution Spectrum Scale d'IBM et de baies de stockage composées principalement de disques rotatifs.

L'objectif de cette prestation est d'analyser les usages de ce système de stockage : suivi du débit et des IOPS par serveur dans le temps, répartition des fichiers par taille, répartition des fichiers par date de modification.

Dans un premier temps, le titulaire fournit une synthèse des usages en faisant ressortir les indicateurs qui lui semblent pertinents.

Ensuite, le titulaire fait des préconisations concernant la configuration de la solution IBM Spectrum Scale existante.

Pour finir, le titulaire se base sur l'analyse des usages existants pour préconiser une nouvelle architecture, pouvant être envisagée dans un nouveau cluster, intégrant une partition constituée de disques SSD et définir des politiques de migrations automatiques des données de la partition SSD vers celle à disques rotatifs et inversement.

#### **4.3.3 Prestation de mise en place d'une solution type chatbot basée sur modèle LLM avec utilisation de ressources GPU du cluster de développement**

L'ONERA dispose d'une solution de chatbot interactif utilisable par portail web et utilisant actuellement comme ressources en arrière-plan des serveurs dédiés à cartes GPU.

L'objectif de cette prestation est de faire évoluer la solution globale pour utiliser aussi en débord des ressources GPU du calculateur de développement.

Le titulaire devra proposer un mode opératoire permettant de réserver des ressources dans le gestionnaire de batch du cluster de développement, sur des périodes bien délimitées et de déporter les traitements sur ces ressources.

Il devra analyser les transferts de données nécessaires du portail web vers les ressources d'arrière-plan, tout en s'assurant du respect de leur confidentialité.

#### **4.3.4 Prestation de mise en place d'un puit de logs, d'outils d'analyse de logs avec modèle d'IA pour maintenance prédictive**

L'objectif de cette prestation est de concevoir une solution d'analyse de logs basée sur la suite ELK (Elasticsearch, Kibana et Logstash) permettant à la fois de rassembler les traces des différents clusters dans un même répertoire de stockage et de les analyser, en particulier pour détecter certains motifs prédisant des pannes matérielles.

Le titulaire propose une maquette d'architecture temporaire pour illustrer le principe de la solution. Des tests de détection de motif de pannes sont réalisés pour mettre en évidence la faisabilité de la solution.

Il propose ensuite une architecture cible, qui sera décrite dans un dossier d'architecture puis validée par la DSI de l'ONERA.

Une fois l'architecture validée, les phases d'installation et de configuration sont réalisées, puis une phase de tests sera effectuée conjointement avec la DSI pour vérifier l'aptitude de la solution.

Une fois la phase de tests finie, les procédures d'exploitation de la solution sont livrées à la DSI.

#### **4.3.5 Audit de sécurité d'un cluster HPC et rédaction d'un plan de durcissement**

Le titulaire étudie la documentation et audite l'architecture d'un calculateur existant, il liste les failles de sécurité et pistes d'amélioration.

Il rédige un plan d'action permettant de corriger les failles et de mettre en œuvre les améliorations possibles.

Ces suggestions sont ajoutées au guide de sécurisation du cluster.

#### **4.4. Poste 4 : phase de réversibilité**

La phase de réversibilité est optionnelle et doit être affirmée en fin d'accord-cadre. Sa durée est d'un mois, elle se déroule sur le site ONERA de référence et son contenu est défini dans le « Plan de Réversibilité ». Le titulaire assure un transfert de connaissance complet vers l'ONERA et le cas échéant vers le titulaire du marché suivant. Il leur remet notamment l'ensemble de la documentation technique à jour et des outils d'exploitation, toute la production dans le cadre de la prestation étant propriété de l'ONERA. Des sessions de travail sont organisées entre les différentes parties selon le planning figurant dans le « Plan de Réversibilité ».

## **5. Contraintes spécifiques liées à l'administration de systèmes d'informations sensibles**

### **5.1. Contraintes minimales d'accès aux sites et au SI de l'ONERA**

Dans le cadre du dispositif de Protection du Potentiel Scientifique de la Nation (PPST, instruction interministérielle 3415/SGDSN/AIST/PST du 07/11/2012, article 413-7 du code pénal notamment), les centres ONERA sont des zones à régime restrictif (ZRR). Dans ces conditions, l'accès au réseau interne de l'ONERA, tout comme l'accès physique à ses sites, nécessite un agrément ministériel de tutelle (délivré en pratique par la DGA) pour chaque personne concernée.

Dans ces conditions, chaque personnel du titulaire travaillant sur le SI de l'ONERA ou dans ses locaux doit disposer d'une autorisation ministérielle d'accès aux ZRR concernées (AZR). Il doit donc soumettre au préalable un dossier d'agrément à l'administration de tutelle de l'ONERA (DGA) via l'officier de sécurité du site ONERA de référence. Cette procédure de demande d'AZR peut durer jusqu'à deux mois (les délais actuels constatés sont de l'ordre de 1 à 2 semaines). **Cette contrainte doit être prise en compte pour la gestion des ressources pendant toute la durée du contrat.**

### **5.2. Habilitation des intervenants des prestations**

Les prestations concernent des SI contenant des informations avec une mention de protection « Diffusion Restreinte » (DR) ou « Diffusion Restreinte Spécial France » (DR SF) sans cloisonnement à ce jour.

De ce fait, chaque intervenant du titulaire doit être ressortissant français.

En phase d'initialisation et en début de contrat, les intervenants de prestations devront au moins être titulaires d'une autorisation ministérielle à travailler dans une ZRR (cf. paragraphe 5.1).

Ensuite et dès que possible, l'ensemble des intervenants ayant des droits d'administration sur tout ou partie du périmètre devra être habilité au minimum au niveau « Secret ».

### **5.3. Accès à distance au SI général interne de l'ONERA**

Une partie des prestations peut être réalisée en dehors des locaux de l'ONERA, sous réserve de l'application des règles suivantes :

- En application des dispositions déjà indiquées aux paragraphes ci-dessus 5.1 et 5.2, la connexion depuis l'extérieur au SI de l'ONERA n'est possible que pour des intervenants ressortissants français, habilités au niveau « Secret » (ou en début de prestation titulaires d'une autorisation ministérielle à travailler dans une ZRR).
- Les postes de travail utilisés sont des postes informatiques portables fournis et administrés par l'ONERA et à usage exclusif des personnes déclarées et autorisées. Ils ne doivent en aucune manière être utilisés par d'autres personnes.
- Pendant l'exécution de travaux à distance sur le système d'information de l'ONERA, ils ne doivent pas être accessibles visuellement (vue sur les moniteurs) par des personnes non autorisées par l'ONERA. Le titulaire doit pouvoir présenter à tout moment le dispositif ou la formalisation des règles internes utilisés pour respecter cette contrainte.

- Il n'est pas permis d'établir de VPN global vers un centre de supervision puis de le router vers certains postes via de l'interconnexion de réseau ; en effet, chaque connexion doit se faire par un utilisateur authentifié et agréé en mode point à point depuis un poste autorisé du prestataire vers un poste ONERA, et non depuis un sous réseau entier du prestataire.
- Le nombre d'autorisations d'accès (nombre d'intervenants en centre de service) doit être en rapport avec les charges réelles et la variété des compétences nécessaires afin de ne pas diluer ces accès au sein d'un groupe de personnes trop important <sup>(1)</sup>. Outre une meilleure maîtrise de la sécurité, cette disposition rend plus réaliste le fait que les personnes concernées soient bien formées au contexte ONERA (connaissance des outils utilisés par la DSI et ses prestataires, connaissance de l'architecture technique et des règles de supervision, connaissance des règles de traitement des tickets).

---

<sup>(1)</sup> Un exemple admissible est de configurer l'accès au réseau de l'ONERA pour trois administrateurs système pour traiter une charge équivalente moyenne correspondant à celle d'un administrateur. Au contraire, configurer un accès pour vingt personnes pour une charge équivalente à « un administrateur système + un DBA » n'est pas satisfaisant.