

Offre d'emploi : « Architecte de solutions de traitement de données dans le cadre du projet européen PHIDIAS »

Type de contrat : CDD ;

Durée de contrat : 3 ans ;

Rémunération : de 2200€ à 3500€ bruts mensuels selon expérience ;

Niveau d'études souhaité : Bac +5 ;

Candidature : merci d'envoyer CV et lettre de motivation à recrutement@cines.fr

Plus d'informations : <https://www.cines.fr>

Le cadre du poste

Le poste est à pourvoir au Centre Informatique National de l'Enseignement Supérieur (CINES) situé 950 rue de St Priest à Montpellier. Le CINES est un établissement public national, placé sous la tutelle du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation, et qui a pour principales missions :

- Le calcul numérique intensif : mise à disposition d'un supercalculateur de niveau mondial avec une équipe d'ingénieurs support dédiée ;
- L'archivage pérenne : conservation à long terme de données électroniques sur disques et bandes magnétiques ;
- L'hébergement d'environnements informatiques : hébergement de baies informatiques appartenant à des établissements nationaux et régionaux.

Le CINES propose ainsi à 1700 utilisateurs scientifiques des moyens informatiques importants pour mener à bien leurs recherches. Il dispose pour cela d'un datacenter de 1500 m² constitué de 5 salles machines et doté d'équipements performants et redondés (onduleurs, groupes électrogènes, groupes froids, etc...), ainsi que de liaisons à très haut débit sur le nœud RENATER.

Contexte de la mission

Le projet européen PHIDIAS¹, d'une durée de 36 mois, est piloté par le CINES au nom de ses 13 partenaires (5 pays européens impliqués). Il a pour objectif principal de développer un ensemble de services et d'outils interdisciplinaires basés sur le HPC et le traitement des données massives, afin de permettre aux sciences du domaine « Système Terre » d'exploiter de grands ensembles de données d'intérêt européen public produit de l'observation de la Terre par satellite. Ils fourniront un accès FAIR (*Findable, Accessible, Interoperable, Re-usable*) à ces jeux de données traités ainsi que des services à valeur ajoutée : du traitement de données « standard » appliqué aux jeux de données hétérogènes Big-Data, aux services plus avancés tels que l'IA ou le HPC à la demande appelé « informatique urgente ». Cela sera rendu possible grâce à une grande capacité de stockage de données et à un réseau à large bande passante à travers l'Europe (réseau GEANT).

¹ <https://www.phidias-hpc.eu>

PHIDIAS développera et proposera un catalogue permettant aux utilisateurs de découvrir et d'accéder aux données, mais également aux logiciels open source pertinents, aux API publiques ainsi qu'aux services de traitement interactif. Ce catalogue implémentera des services interopérables pour la découverte, l'accès et le traitement des données, et sera connecté à d'autres référentiels de données importants tels que le Portail Européen de Données², GEOSS³, NextGEOSS et EOSC⁴. PHIDIAS mettra également en place un service de traitement interactif Web commun pour les utilisateurs finaux basé sur les technologies de cahier de notes et de cubes de données.

Description de la mission

Rattaché(e) au Département Calcul Intensif (DCI), vous travaillerez avec une équipe technique dédiée au calcul intensif sur machine massivement parallèle.

Le projet PHIDIAS est découpé en 7 Work Packages. Le CINES, qui pilote le projet, gère aussi les work packages 1 et 2, qui sont subdivisés en sous-tâches. Vous devrez ainsi piloter la sous-tâche 2.1 dédiée au HPC&HPDA et qui est adossée aux « data sciences ». Vous serez aussi amené(e) à participer aux tâches 2.2 (*Stockage et archivage pérenne*) et 2.3 (*On the fly computing*).

En tant qu'architecte de solutions pour le traitement de données, vos principales missions seront les suivantes :

- Etablir des liens avec les partenaires du projet afin de comprendre et formaliser les besoins des communautés ;
- Déterminer la meilleure manière d'y répondre et travailler avec le work package WP3 autour des données nécessaires à leur mise en œuvre ;
- Analyser et implémenter des stratégies et des outils d'échanges de données ;
- Définir et participer à la mise en place une infrastructure dédiée aux traitements des données ;
- Proposer et adapter des modèles et outils d'analyse de données : analyses sur-mesure, visualisation, etc... ;
- Partager les recherches et résultats avec le reste de l'équipe ;
- Participer à « l'évangélisation » des différentes communautés scientifiques partenaires de PHIDIAS pour l'adoption des solutions déployées ;
- Réaliser une veille sur les méthodes d'échange et de gestion des flux de données.

Profil

- Diplômé(e) d'un Bac +5 en école d'ingénieur ou équivalent universitaire, avec une bonne compétence en administration système et une spécialisation forte dans le traitement de données ;

² <https://www.europeandataportal.eu/fr/homepage>

³ <https://www.earthobservations.org/geoss.php>

⁴ <https://www.eosc-portal.eu/>

- Vous justifiez d'une expérience de 2 ans minimum en data science, idéalement dans des problématiques directement liées à des sujets autour des données satellitaires et de l'Intelligence Artificielle ;
- Vous justifiez d'une première expérience de développement ;
- Vous savez interpréter les questions métiers des communautés scientifiques et vous démontrez des qualités de communication pour partager vos idées et résultats de façon efficace ;
- Vous maîtrisez l'anglais pour pouvoir participer efficacement aux échanges avec les partenaires de ce projet européen.

Compétences techniques nécessaires

- Vous avez de fortes compétences en outils de gestion de flux de données ;
- Vous êtes capable de proposer et participer à la mise en œuvre d'une infrastructure de données complète depuis l'entrée jusqu'à la sortie définitive vers un service d'archivage pérenne ;
- Vous avez des compétences en design d'architecture d'ingestion, transformation & visualisation de données ;
- Vous savez mettre en place des tests de charge, du monitoring et utiliser des outils de *troubleshooting* ;
- Une expérience dans le domaine des images satellites serait un plus ;
- Vous connaissez les techniques suivantes :
 - Cassandra, Elasticsearch, PostgreSQL, MongoDB ;
 - Hadoop, Spark ;
 - iRODS, ZeroMQ ;
 - Docker, Kubernetes ;
 - Bash, Python...
 - Ansible, Packer, Chef ;
 - Kibana, Grafana, Prometheus.
- Vous connaissez l'environnement du HPC ;
- Toutes expériences en Data Mining, Text Mining, utilisation de NLP et technologies sémantiques sont également les bienvenues ;
- Vous possédez une bonne aptitude à la rédaction de rapports, l'organisation de réunions avec vos correspondants locaux et étrangers.

Qualités relationnelles

- Capacité à communiquer ;
- Capacité à travailler en équipe ;
- Capacité d'analyse et de synthèse ;
- Capacité d'adaptation.