

## 1 - Situation géographique :

INRA - Unité GenPhySE (Génétique, Physiologie et Systèmes d'Élevage)  
Auzerville Tolosane

## 2 - Contact :

M. Alexandre JOURNAUX, 05.61.28.51.73, [alexandre.journaux@inra.fr](mailto:alexandre.journaux@inra.fr)  
M. François LAPERRUQUE, 05.61.28.51.75, [francois.laperruque@inra.fr](mailto:francois.laperruque@inra.fr)

## 3 - Description du poste :

La personne recrutée sera en charge d'étudier puis de mettre en place la meilleure solution possible pour le stockage et l'interrogation de gros volume de données (Big Data).

### Environnement fonctionnel

GenPhySE est une unité de recherche basée sur le centre Inra de Toulouse dont les travaux vont de la gestion des populations animales (toutes espèces) à l'évaluation et la conception de systèmes d'élevage plus durables.

Le CTIG (Centre de Traitement de l'Information Génétique), basé sur le centre Inra de Jouy en Josas, est le centre informatique du Département de Génétique Animale (GA). Une de ses missions est la gestion des infrastructures de calcul et l'hébergement de services informatiques pour le Département GA.

La personne recrutée sera attachée administrativement au CTIG mais intégrera l'unité GenPhySE à Toulouse, au sein de l'équipe Informatique et Automatisation. Toutefois, en fonction du besoin, l'affectation géographique pourrait se faire au CTIG.

### Environnement technique

L'équipe Informatique et Automatismes met en place des outils pour intégrer des données de phénotypage animal depuis l'acquisition jusqu'au traitement statistique. Des nouveaux outils permettent désormais l'enregistrement automatique de mesures standards, répétées et fréquentes. C'est le cas, par exemple, des distributeurs automatisés qui enregistrent, de manière continue, des informations sur le comportement des animaux lorsqu'ils viennent se nourrir et s'abreuver.

Jusqu'à présent, ces données étaient stockées dans une base de données relationnelles (MySQL). Une première solution orientée « Big data » s'appuyant sur les technologies Spark, Solr, et Cassandra a été mise en place sur un environnement de test. La personne recrutée devra finaliser et valider cette solution pour l'implémenter dans l'environnement de production. Ensuite, il lui sera demandé d'utiliser et configurer Docker pour généraliser la solution retenue.

## 4 - Compétences attendues :

Techniques : MySQL, Cassandra (ou autre base NoSql), Solr, Spark, Scala, Docker

Environnement : Linux

Conception : UML

Travail en équipe, bonne capacité de rédaction et de communication

## 5 - Durée du contrat :

3 mois en 2018 + 7 mois en 2019