CHU de Toulouse

Intitulé de l'offre de stage

Conception et développement d'un pipeline ETL entre la base de données d'anesthésie-réanimation et l'entrepôt de données de santé du CHU de Toulouse

Stagiaire en	Data-ingéniérie
Affectation	CHU Toulouse
Durée	5 à 6 mois
Rémunération	environ 670 € net mensuel
Date de la publication	15/11/2026
Date d'embauche prévue	mars 2027 suivant disponibilités
Lieu	CHU, 2 rue Charles Viguerie, 31300 Toulouse, France

Le Centre Hospitalier Universitaire de Toulouse

Le Centre Hospitalier Universitaire (CHU) de Toulouse est constitué de plusieurs sites (les principaux étant les sites de Rangueil, Larrey et Purpan, ainsi que celui de l'oncopole en commun avec l'institut Claudius Régaud); il comprend 4 000 médecins et 12 000 personnels hospitaliers. La mission de recherche et d'innovation du CHU fait partie intégrante à la fois de son activité quotidienne et de sa stratégie pour l'avenir. Elle est menée en collaboration étroite avec les facultés et les organismes de recherche que sont notamment l'Institut national de la santé et de la recherche médicale (Inserm) et le Centre national de la recherche scientifique (CNRS).

L'équipe d'accueil

L'informatique, le traitement des données et l'intelligence artificielle sont appelés à prendre une place croissante dans le monde de la recherche médicale. Dans ce cadre, le CHU de Toulouse a créé le Centre de Données pour la Santé et la Recherche (CDSR) dont le but est le recueil, l'analyse et le traitement de la donnée médicale, en support aux activités de recherche médicales du CHU.

Le département anesthésie-réanimation du CHU participera également à l'encadrement du stage.

Objet du stage

Le stage s'inscrit dans la réutilisation des données de santé (recherche, BI, modélisation), avec pour fil rouge clinique la prédiction de la sévérité en réanimation, au sein d'une équipe réunissant médecins, data engineers et data scientists.

En réanimation, chaque patient génère en continu des milliers de mesures : constantes vitales, paramètres du respirateur, traitements, examens biologiques. Ces données sont stockées dans le dossier patient informatisé d'anesthésie-réanimation, dont la structure est pensée pour le soin et non pour l'analyse.

Le CHU de Toulouse construit un entrepôt de données de santé pour rendre ces données réutilisables, dans un cadre réglementaire strict. Il manque un maillon : un pipeline ETL fiable, testé et documenté entre la base de réanimation et l'entrepôt. Le ou la stagiaire contribue, au sein de l'équipe, à la construction de ce maillon.

Le fil rouge clinique : prédire la sévérité en réanimation. Les données mises en forme dans le cadre du stage servent des usages variés : BI, analyses, modélisation. La prédiction de la sévérité sert de cas d'usage de référence. Vous saurez toujours pourquoi vous transformez telle ou telle donnée, et l'usage qu'en feront les cliniciens.

Au sein de l'équipe, vous contribuez au pipeline de bout en bout, de la compréhension de la source jusqu'à la mise à disposition des données pour leurs réutilisations.

> Explorer et documenter la base source : comprendre la structure des données de réanimation, repérer les pièges (doublons, valeurs manquantes, unités, séries temporelles irrégulières) et les écarts entre la documentation et la pratique.
> Participer à la conception des règles de correspondance (mapping) entre la base source et le modèle cible déjà défini (OMOP-CDM), en lien avec les médecins qui connaissent la réalité clinique derrière chaque champ.
> Contribuer au développement du pipeline ETL en SQL, Python et dbt, avec des transformations modulaires, versionnées et reproductibles.
> Contribuer au contrôle qualité et aux bonnes pratiques d'ETL : participation aux étapes de contrôle qualité (tests automatisés, contrôles de cohérence, suivi des anomalies) et application des bonnes pratiques d'ETL dont le CHU de Toulouse est moteur.
> Valider avec l'équipe : vérifier avec les cliniciens que les données transformées reflètent bien la réalité, et permettent leur exploitation par des data scientists.
> Documenter votre travail pour que l'équipe puisse le maintenir et le faire évoluer après votre départ.

Ce que vous apprendrez sur place :

> Les formats FHIR et OMOP, dbt et la modern data stack
> Le vocabulaire médical et le fonctionnement d'un service de soin via l'exemple de la réanimation : vous serez accompagné(e) par les médecins de l'équipe

Ce que nous attendons de vous :

> de la curiosité pour la santé,
> de la rigueur (les données manipulées sont des données de patients)
> l'envie de travailler avec des médecins autant qu'avec des profils data.

Ce que le stage apporte :

- Des données réelles. Vous travaillez sur des données de santé issues de la vie réelle : volumineuses, bruitées, incomplètes, comme dans la pratique d'un data engineer.
- Une stack moderne et des standards de référence. Vous apprendrez à utiliser dbt, l'outil de référence des modern data stack, sur un cas d'usage complet. Le pipeline utilise des formats incontournables de la donnée de santé (FHIR, OMOP), dont la maîtrise sera indispensable dans le cadre de l'Espace européen des données de santé (EEDS).
- Une équipe pluridisciplinaire. Vous travaillez au quotidien avec des médecins, des data engineers et des data scientists. Vous apprenez à traduire un besoin clinique en pipeline de données.
- Un projet qui a du sens. Le pipeline alimente des usages variés (recherche, BI, modélisation), dont la prédiction de la sévérité en réanimation, avec un impact clinique à la clé.
- Un livrable qui reste. Votre travail est utilisé après votre départ. Vous pouvez le présenter en entretien, et il peut nourrir votre mémoire.

Formation

École d'ingénieur, de préférence avec spécialisation en informatique ou mathématiques (en année de césure ou stage long). Master 2 informatique ou mathématiques appliquées.

Compétences attendues

- Maitrise de **Python** (pandas, NumPy, scikit-learn, matplotlib)
- Connaissances de base en **SQL**.

Modalité de candidature

- CV à envoyer à: anes4@stages-medecine-numerique.fr
- Date limite de candidature : 01/01/2026
- Encadrants: Dr. Michael Poette (Département d'anesthésie-réanimation), Laure Feuillet (Equipe Data)

Une première sélection sera effectuée sur la base des CVs reçus. Les candidats seront immédiatement informés du résultat, et ceux qui seront retenus à l'issue de la première sélection seront invités à un ou deux entretiens individuels en téléconférence avant sélection définitive.

L'ensemble des offres de stage est disponible sur http://www.stages-medecine-numerique.fr.