



R&D

Biométrie / Data management

leem  
les entreprises  
du médicament

# Ingénieur(e) des données

L'ingénieur(e) de données, ou ingénieur(e) informatique, crée la structure des bases de données et les algorithmes d'intelligence artificielle pour traiter ces données. Il/elle répond à un besoin défini au préalable. Il/elle développe les bases de données, les teste et réalise les architectures data. Il lui revient également d'organiser l'échange de flux d'informations entre les sources et la base de stockage des données. Il/elle intervient en amont de l'analyse du/de la data scientist.



## Profil de recrutement :

Métier accessible aux jeunes diplômé(e)s ou après une première expérience. L'entreprise peut souhaiter un profil plus confirmé en fonction du besoin et de l'équipe existante. Ce poste tend à être internalisé, notamment pour les profils disposant d'une double compétences informatique /IT et scientifique /médicales, capables de bien comprendre et intervenir en appui des besoins métiers (particulièrement la R&D et le marketing).



## Formations

### Parcours recommandés :

- Formation de niveau Bac +5 de type Master Big Data en école d'ingénieur ou école spécialisée en informatique

Pour aller plus loin : <https://imfis.fr>



## Passerelles métier :

- Data Architect
- Data Scientist
- Technical Leader Big Data

Pour aller plus loin : [www.macarrieredanslapharma.org](http://www.macarrieredanslapharma.org)



## Autres appellations :

- Data Ingénieur(e)
- Data Engineer
- Ingénieur(e) Big Data

## ACTIVITÉS

### Développement, construction et tests des architectures bases de données

- Recueil des besoins métiers pour la collecte et le stockage des données et rédaction des spécifications techniques
- Intégration de l'ensemble des éléments juridiques, de conformité et de sécurité relatifs à la RGPD et à la bonne utilisation/exploitation des données
- Construction et développement des solutions avec les outils et logiciels adaptés pour construire la solution, que ce soit pour le Big Data les bases de données ou les infrastructures
- Réalisation des tests unitaires et des tests d'intégration
- Mise en place et maintenance des « batchs » (automatisation d'une série de traitements)

### Traitement des données, optimisation de la qualité

- Optimisation et gestion de la base de données pour garantir la fluidité de la collecte, du stockage et de l'exploitation d'une grande masse de données.
- Création de pipelines de données fiables et combinaison des sources de données
- Animation des processus de maintenance digitale, tant préventive que corrective et anticipation de l'obsolescence des solutions dans une logique de maintenance prédictive

### Adaptation de l'architecture aux besoins des utilisateurs

- Sécurisation des solutions développées
- Garantie de la lisibilité pour les utilisateurs/trices (notamment data analysts et data scientists) et de l'adaptation aux besoins métiers
- Implémentation du suivi de la validité des modèles statistiques et/ou de machine learning (apprentissage machine grâce à l'analyse de flux de données) et développement de leur industrialisation

### Animation et coordination d'équipe dans le cadre de gestion de projets

- Pilotage de la collecte de données, management et accompagnement de son équipe dans l'apprentissage d'une méthodologie, contribution à l'amélioration continue des process
- Garantie de la qualification des demandes des utilisateurs par son équipe et accompagnement de son équipe dans la compréhension et l'analyse des demandes
- Collaboration avec les data scientists sur les projets intégrant le machine learning et/ou l'intelligence artificielle
- Collaboration avec le(s) data analyst(s) et le(s) data scientist(s), qui analysent et exploitent les données à différents degrés de complexité

# R

## COMPÉTENCES CLÉS

### ←|→ Transverses

- Savoir travailler en équipe et comprendre l'impact de son travail dans l'utilisation faite par les autres services/unités
- Mettre son sens de l'organisation et de la rigueur au service de la qualité des bases de données
- Être force de proposition dans l'architecture de la base de données et les algorithmes d'intelligence artificielle à mettre en place
- Faire preuve de réactivité et d'un esprit analytique pour décoder des situations informatiques complexes
- Maîtriser le vocabulaire anglais technique et informatique des outils de la fonction



### Métier

- Disposer des fondamentaux métiers sur les domaines d'activité de l'entreprise
- Gérer les sources de données médicales (EEG, ECG, etc.)
- Maîtriser le développement informatique et les langages de code (Java, Python, Scala, etc.)
- Fiabiliser et combiner les données
- Maîtriser les technologies de gestion de bases de données (SQL, NoSQL, NewSQL, DB2, etc.), les outils framework Big Data (Hadoop, Spark, Kafka, Hbase, Hive, etc.)
- Modéliser des données
- Utiliser les fondamentaux du machine learning et de l'intelligence artificielle
- Bien connaître les outils et méthodes d'amélioration continue des process
- Bien connaître les méthodes de développement agile
- Collecter des données en utilisant l'ETL et les outils associés
- Savoir qualifier les demandes et faire preuve de pédagogie pour en favoriser l'intégration par son équipe
- Maîtriser la réglementation liée aux données

## Evolution du métier

Les compétences de l'ingénieur(e) de données sont de plus en plus internalisées par les entreprises, du fait de leur aspect stratégique dans le soutien à la prise de décision.

L'augmentation de la quantité et de la complexité des données entraîne un besoin accru en profils spécialisés dans leur traitement et leur mise à disposition pour les autres services. Ce métier est amené à prendre de l'ampleur et à voir ses champs de responsabilités se développer, notamment dans les petites entreprises via la cartographie et la documentation des données et la contribution à la mise en place d'une culture « data driven ».



Ingénieur(e)  
des données