



**keyrus**  
make data matter

# Cinq étapes pour bâtir une architecture data robuste

Construire une architecture data de confiance

[www.keyrus.com](http://www.keyrus.com)

# Cinq étapes pour bâtir une architecture data robuste

## Construire une architecture data de confiance

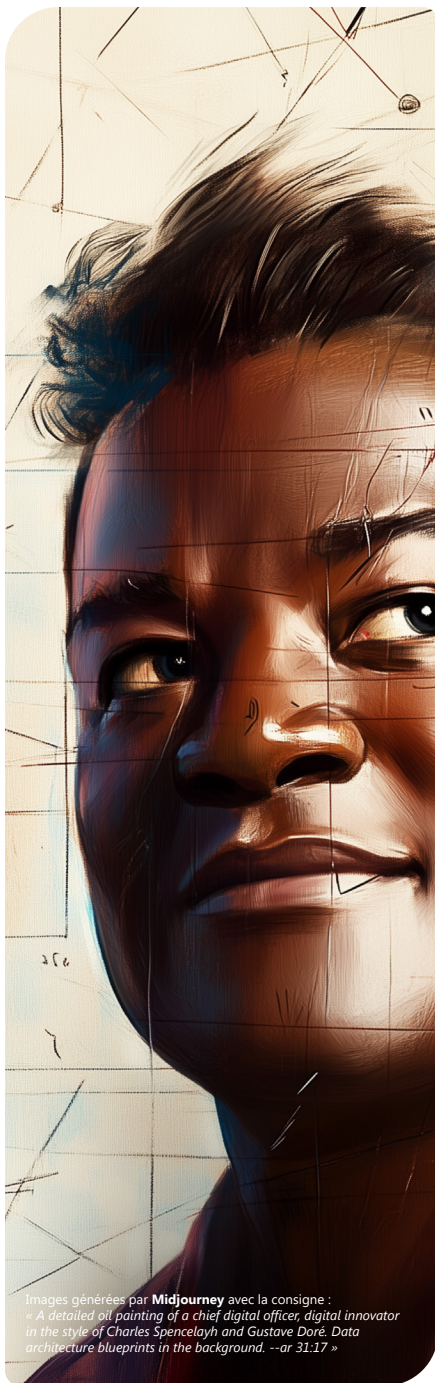
La transformation numérique des organisations repose de plus en plus sur leur capacité à exploiter efficacement les données. En lien avec cette contrainte, l'architecture data ne se résume plus à une simple infrastructure technique : elle devient le socle de la confiance indispensable à tout usage de la data, et à toute prise de décision éclairée.

Garantir cette confiance implique de maîtriser un ensemble de paramètres fondamentaux : l'alignement de l'architecture sur les ambitions stratégiques, la rigueur de la gouvernance, la qualité et la traçabilité des données, la robustesse des chaînes de traitement, et la capacité à évoluer dans le temps. Et – bien évidemment – la complexité croissante des environnements numériques combinée à l'irruption massive de l'intelligence artificielle renforcent ces exigences.

Ainsi, la diversité des sources de données, la variété des formats, les impératifs réglementaires et la pression économique viennent complexifier l'équation et imposent une vision globale et structurée de l'architecture data. En effet, cette dernière doit non seulement garantir l'intégrité des flux de données, mais aussi permettre d'en extraire de la valeur en toute sécurité, tout en préservant l'agilité nécessaire aux évolutions futures.

Construire une telle architecture requiert une approche méthodique, capable d'intégrer des considérations techniques, fonctionnelles et stratégiques. C'est à ce prix qu'une organisation peut réellement faire de ses données un levier de performance et de compétitivité durable.

Bref, c'est une magnifique feuille de route qui se trace devant nous. On vous explique tout !



Images générées par **Midjourney** avec la consigne :  
« A detailed oil painting of a chief digital officer, digital innovator  
in the style of Charles Spencelayh and Gustave Doré. Data  
architecture blueprints in the background. --ar 31:17 »

## Étape 1 | Définir la vision et les objectifs stratégiques

On ne le répètera jamais assez, toute architecture data de confiance se construit d'abord sur une intention claire : répondre aux ambitions stratégiques de l'organisation. Avant d'engager des choix techniques ou de déployer des infrastructures, il est indispensable de définir précisément ce que l'architecture doit permettre d'atteindre.

L'architecture data n'est pas une finalité en soi. Elle constitue un levier au service d'objectifs qui varient selon les contextes : renforcer un leadership sur un marché, réduire le time-to-market, améliorer l'excellence opérationnelle ou encore accélérer l'innovation. Chacun de ces objectifs stratégiques impose des arbitrages techniques et budgétaires qu'il convient d'assumer dès l'amont du projet.

Dit autrement, cette étape fondatrice consiste à formaliser un **contrat d'architecture** qui fixe la cible à atteindre et les moyens à mobiliser. Ce contrat engage l'ensemble des parties prenantes sur ce que l'architecture devra garantir, mais aussi sur ce qui pourra être différé ou ajusté en fonction des contraintes rencontrées. La notion de **trajectoire** prend alors tout son sens : il est légitime d'accepter des désalignements temporaires si la cible est connue et partagée.

La définition de la vision stratégique permet également d'adopter une approche proportionnée. Toute ambition suppose un engagement cohérent en matière de moyens techniques, humains et financiers. À titre d'exemple, viser une architecture capable de supporter des traitements en temps réel pour des usages critiques implique des investissements significatifs dans les infrastructures et la supervision des flux.

L'architecture data devient ainsi le reflet de la stratégie de l'entreprise. Elle doit évoluer au même rythme que les priorités business et s'adapter aux évolutions du contexte.



Sans cette discipline d'alignement permanent entre la vision et la réalisation technique, la robustesse de l'architecture reste illusoire.

Ainsi, définir la cible, clarifier les ambitions et assumer les arbitrages temporaires constituent les premières pierres d'une architecture de confiance. À ce stade, la responsabilité des décideurs consiste à exprimer cette ambition sans ambiguïté afin de guider les choix techniques qui suivront et d'assurer la cohérence de l'ensemble de la trajectoire data.

## Étape 2 | Instaurer une gouvernance des données étendue et structurée

À la suite du contrat d'architecture, la gouvernance des données constitue le premier levier opérationnel pour instaurer la confiance dans l'architecture data. Elle ne se réduit pas à la gestion des accès ou au respect des obligations réglementaires. Elle englobe l'ensemble des mécanismes permettant de garantir l'intégrité, la qualité, la traçabilité et la maîtrise économique des données tout au long de leur cycle de vie.



Image générée par Lexica Apture v5 avec la consigne « Vintage poster of a person in control the circulation of sensitive data. »

La première exigence de la gouvernance repose sur la capacité à contrôler la circulation des données sensibles. Chaque utilisateur ne doit accéder qu'aux informations pour lesquelles il est habilité. La moindre dérive dans cette gestion des droits entame immédiatement la confiance, expose l'organisation à des risques juridiques et affaiblit la solidité de l'architecture. La gouvernance doit ainsi s'assurer que toute donnée diffusée l'est auprès d'un public légitime, en respectant les règles de confidentialité et de conformité propres à chaque organisation.

La gouvernance étendue dépasse ces seuls aspects réglementaires. Elle intègre des dimensions techniques essentielles à la robustesse de l'ensemble. Les pratiques de **DataOps**, de **MLOps** ou de **LLMOps** participent pleinement à cette gouvernance élargie. Elles garantissent la reproductibilité des traitements, la maîtrise des cycles de développement et de déploiement, ainsi que la qualité des résultats produits. Cette mécanique bien huilée assure qu'à chaque évolution ou mise à jour, l'intégrité de la chaîne de traitement est préservée.

- Pour approfondir la manière dont les **DataOps** sécurisent vos chaînes de traitement, consultez notre eBook *Data Fabricadabra*.
- Pour explorer les bonnes pratiques et la mise en œuvre d'une démarche **MLOps** robuste, découvrez notre guide dédié *Standardisez vos workflows Machine Learning — Le guide du MLOps*.
- Pour comprendre les enjeux d'orchestration et de supervision des grands modèles de langage, nous vous invitons à lire *Chat-Geppetto Façonner la Conscience Artificielle* — Bien qu'axé sur la conscience artificielle, cet eBook traite des défis opérationnels de la gestion des modèles de grande taille, et amorce des notions proches du **LLMOps**.

Une gouvernance efficace repose enfin sur la traçabilité de l'ensemble des opérations. La possibilité d'auditer l'historique complet des données et des traitements sécurise l'organisation face aux contrôles et renforce la maîtrise de son patrimoine informationnel. Cette transparence garantit la capacité à démontrer à tout moment la conformité des actions menées.

Instaurer cette gouvernance étendue permet de transformer l'architecture data en un système fiable, pilotable et pérenne. La confiance se construit alors sur des bases solides : la maîtrise des usages, la rigueur des processus et la capacité à concilier performance technique et rentabilité économique.

“ Les pratiques de DataOps, de MLOps ou de LLMOps garantissent la **reproductibilité des traitements**, la **maîtrise des cycles** de développement et de déploiement, ainsi que la **qualité des résultats** produits. ”

## Étape 3 | Assurer la qualité et la traçabilité des données tout au long de leur cycle de vie

La qualité des données constitue, elle aussi, un pilier fondamental de la confiance dans l'architecture data. Quelle que soit la performance des infrastructures ou la sophistication des traitements, aucune valeur ne peut émerger d'un système alimenté par des données erronées, incomplètes ou incohérentes. L'adage « **garbage in, garbage out** » reste plus que jamais valable à l'ère de l'intelligence artificielle et de l'automatisation.

Garantir la qualité des données implique d'intervenir à chaque étape de leur cycle de vie : dès l'ingestion, lors des phases de transformation et d'enrichissement, jusqu'au stockage et à la restitution finale. Ce processus nécessite des mécanismes rigoureux de contrôle, de nettoyage et de validation systématique des flux entrants comme des traitements internes.

La mise en place d'un **Data Lineage** et d'un **Data Catalog** s'impose comme un prérequis. Ces outils permettent de tracer avec précision l'origine de chaque donnée, de reconstituer son parcours à travers l'architecture et de comprendre les transformations subies. Cette transparence garantit la capacité à répondre aux exigences des audits, à identifier rapidement l'origine d'un dysfonctionnement et à fiabiliser les analyses produites. Pour en savoir plus sur le rôle stratégique du **Data Lineage** et du **Data Catalog** dans la maîtrise et la gouvernance des flux de données, là encore, nous renvoyons vers notre eBook [Data Fabricadabra](#).

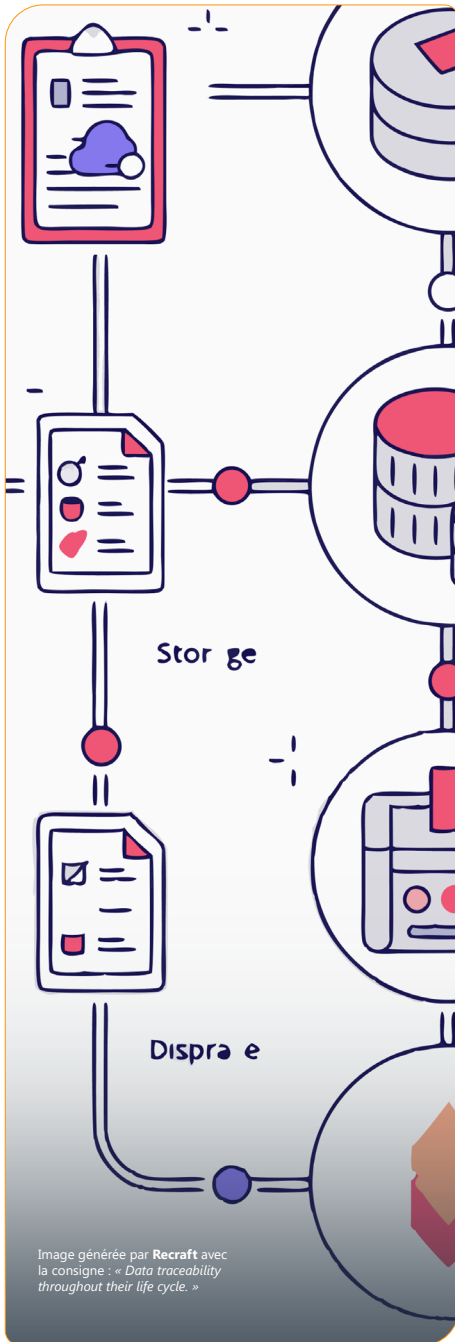


Image générée par Recraft avec la consigne : « Data traceability throughout their life cycle. »



Image générée par Midjourney avec la consigne : « Unstructured data — documents, text content, multimedia data — or complex datasets such as, for example, geospatial data or, to take just one, example from the medical field, omics data. »

La maîtrise de la qualité ne s'arrête pas aux données structurées. Les données non structurées — documents, contenus textuels, données multimédias — ou les jeux de données complexes tels que, par exemple, les données géospatiales ou, pour ne prendre qu'un seul exemple dans le domaine médical, les données omiques, doivent bénéficier de la même rigueur de traitement. La préparation et la mise en qualité de ces données conditionnent directement la pertinence des résultats des modèles d'intelligence artificielle ou des moteurs de recherche augmentés (RAG).

Le respect des exigences réglementaires en matière de conservation des données renforce cette nécessité. Certaines données doivent être stockées pendant des durées précises et pouvoir être restituées dans leur intégralité. Toute purge prématurée ou perte d'historique expose l'organisation à des risques majeurs.

Enfin, la qualité des données contribue à la cohérence des décisions. Des données fiables assurent la solidité des analyses et des indicateurs stratégiques. À l'inverse, la moindre faille altère la confiance des utilisateurs et dégrade la performance globale de l'architecture.

Faire de la qualité et de la traçabilité des données une priorité absolue permet de sécuriser l'ensemble de la chaîne de valeur de l'architecture data. C'est la condition indispensable pour que la donnée devienne un véritable levier d'action et de productivité et non un facteur de risque.

“ Des données fiables assurent la solidité des analyses et des indicateurs stratégiques. À l’inverse, **la moindre faille altère la confiance des utilisateurs** et dégrade la performance globale de l’architecture. ”

## Étape 4 | Adopter une architecture technique modulaire et évolutive

La robustesse d'une architecture data ne repose pas uniquement sur la puissance des infrastructures déployées. Elle dépend avant tout de la capacité de l'architecture à absorber la complexité croissante des flux de données, à s'adapter aux évolutions technologiques et à intégrer de nouveaux usages sans remettre en cause ses fondations.

Une architecture de confiance doit permettre la coexistence et l'orchestration de flux hétérogènes. La distinction entre **Cold Path**, **Warm Path** et **Hot Path** répond à cette exigence. Les traitements en mode batch restent pertinents pour les historiques ou les flux à faible fréquence. Les traitements en streaming s'imposent pour les flux opérationnels nécessitant une actualisation régulière. Les flux en (pseudo-)temps réel, quant à eux, deviennent incontournables dès lors qu'une réactivité immédiate est exigée.

La capacité à gérer simultanément ces différents régimes de traitement garantit la performance de l'architecture sans surcoût inutile. Elle évite également de multiplier les silos technologiques, source fréquente d'inefficacité et de perte de contrôle.

La mise en place d'une **Single Source of Truth (SSoT)** représente un autre pivot essentiel. Cette source unique de référence permet de centraliser les données consolidées et validées, offrant ainsi un socle fiable pour l'ensemble des traitements et des analyses. La SSoT ne prétend pas détenir une vérité absolue, mais elle constitue le point de départ indispensable pour produire des data products de qualité, sur lesquels les métiers peuvent s'appuyer en toute confiance.



Image générée par **Lexica Aperture v5** avec la consigne : « *Minimalist painting of data architect design in the style of Olvia De Berardinis* »

L'architecture doit également être conçue pour accueillir la diversité des typologies de données. Les données documentaires ou issues d'applications métier complexes, par exemple, nécessitent des traitements et des modèles adaptés. L'évolution vers des usages basés sur l'intelligence artificielle ou des systèmes de **Retrieval Augmented Generation (RAG)** renforce cette exigence d'adaptabilité.

Enfin, la modularité de l'architecture conditionne sa capacité à évoluer dans le temps. L'intégration de nouveaux services, la montée en charge ou l'évolution des besoins métiers doivent pouvoir se faire sans refondre l'ensemble du dispositif. Cette agilité technologique permet de sécuriser les investissements dans la durée et de maintenir le niveau de confiance requis.

Concevoir une architecture technique modulaire et évolutive, capable d'absorber la complexité des flux et des usages, constitue la seule garantie d'une performance durable. C'est à cette condition que l'architecture data devient un véritable levier d'innovation et de création de valeur.



Image générée par Midjourney avec la consigne : « Trust in data. »

## Étape 5 | Piloter la trajectoire, la maintenance et l'évolution continue de l'architecture

La confiance dans une architecture data se mesure dans le temps à la capacité de l'organisation à piloter la trajectoire de son architecture, à en assurer la maintenance et à anticiper les évolutions technologiques, métiers ou réglementaires.

Une architecture data figée devient rapidement un facteur de risque. Les technologies évoluent, les usages se transforment, les exigences de sécurité et de conformité se renforcent. Ce qui était fiable hier peut devenir obsolète demain. Les algorithmes de chiffrement en sont un exemple concret : robustes à leur création, ils finissent par devenir vulnérables sous l'effet des avancées technologiques. La confiance est donc, par nature, **relative**, **transitive** et **contextuelle**.

Piloter la trajectoire de l'architecture implique de prévoir les moyens nécessaires pour maintenir la performance et la sécurité du système sur le long terme. Ce pilotage s'appuie sur une gouvernance étendue, qui dépasse la simple gestion des flux de données. Il s'agit d'intégrer la maintenance préventive, les mises à jour régulières, le suivi des performances et le contrôle des coûts dans un cadre maîtrisé.

La dimension **FinOps** prend ici toute sa place. Une architecture techniquement parfaite mais dont le coût devient disproportionné fragilise la confiance qu'on lui accorde. Le pilotage économique permet d'ajuster les moyens mobilisés en fonction de la valeur réellement créée par l'architecture et de garantir la soutenabilité du modèle.

La capacité d'adaptation de l'architecture aux nouveaux besoins métier et aux évolutions du contexte réglementaire constitue un autre facteur de robustesse. L'introduction de nouveaux cas d'usage, le déploiement de modèles d'intelligence artificielle ou la prise en compte de nouveaux formats de données doivent pouvoir s'intégrer sans déstabiliser l'ensemble de l'écosystème.

Enfin, l'architecture doit rester lisible et pilotable dans la durée. Les mécanismes de traçabilité dont nous avons parlé, les audits réguliers et la documentation des évolutions garantissent que la confiance acquise ne se dégrade pas avec le temps. La vigilance doit être permanente pour éviter l'accumulation de dettes techniques ou de dérives fonctionnelles.

Assurer cette gouvernance dans la durée, piloter les arbitrages et maintenir la cohérence globale du système permet de préserver la confiance dans l'architecture data. La robustesse devient alors une qualité évolutive, portée par la capacité de l'organisation à anticiper, corriger et s'adapter.

## La *robustesse de confiance* pour produire de la valeur durable

En résumé, La robustesse d'une architecture data se mesure à sa capacité à produire de la valeur de manière durable, tout en maîtrisant les risques et les coûts. Elle repose sur un équilibre subtil entre ambitions stratégiques, contraintes opérationnelles et exigences réglementaires. Plus encore, elle s'incarne dans la capacité de l'organisation à piloter dans la durée la trajectoire de son architecture, en garantissant la cohérence des choix et l'adaptabilité aux évolutions des usages et des technologies.

Face à ces défis, **Keyrus propose à ses clients un accompagnement global, combinant conseil stratégique et expertise opérationnelle.**

Cet accompagnement couvre l'ensemble des étapes clés : de la définition des ambitions à la conception de l'architecture, de la mise en œuvre technique à la gouvernance continue et au pilotage de la performance.

La force de Keyrus réside dans sa capacité à s'adapter aux spécificités de chaque organisation, en proposant des approches modulaires et sur-mesure. Qu'il s'agisse d'architectures *Data Fabric* ou *Data Mesh*, d'intégration de l'intelligence artificielle ou de traitement de données complexes, les équipes Keyrus apportent à la fois la vision, la méthode et la capacité d'exécution pour faire de la donnée un véritable levier de performance durable.

L'enjeu n'est plus seulement d'exploiter les données, mais de construire une relation de confiance entre l'organisation, ses données et ses utilisateurs. C'est cette ambition que Keyrus porte aux côtés de ses clients, en les accompagnant à chaque étape vers des architectures data à la fois robustes, évolutives et créatrices de valeur.

**Article co-écrit par Keyrus, ChatGPT, DeepSeek, Mistral, Claude, Copilot, Perplexity et Gemini**



Image générée par Midjourney avec la consigne : « Hyperrealistic portrait of a data, ai, and business consultant.. »

## Les unes auxquelles vous avez échappé...

Lors de la création de nos ebooks, chaque image naît d'un processus itératif avec le moteur d'IA. Des allers-retours répétés permettent d'affiner chaque visuel, jusqu'à atteindre l'expression la plus fidèle de l'idée souhaitée. Mais le travail ne s'arrête pas là : parmi les images générées, il faut ensuite opérer des choix. Ce processus de sélection repose sur un effort d'objectivation, où il s'agit de mettre de côté sa propre subjectivité pour respecter au mieux le « cahier des charges ». Une démarche délicate, tant les images peuvent inspirer, surprendre, voire fasciner, mais essentielle pour rester fidèle à l'intention initiale.

## L'atelier du prompteur...

Images générées par  
**Midjourney** avec la

**consigne** : *« A detailed oil painting of a chief digital officer, digital innovator in the style of Charles Spencelayh and Gustave Doré. Data architecture blueprints in the background. --ar 31:17 »*



## Étape 1 | Vous avez trouvé cette lecture utile ?

Vous aimerez sûrement aussi :

### MLOps Standardisez vos workflows de Machine Learning

Guide pratique de démarrage

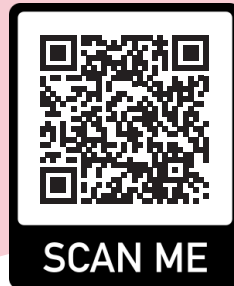
Cet ebook est un guide pratique pour la mise en œuvre du MLOps, visant à standardiser et optimiser les workflows de machine learning. Il aborde les défis liés à la gestion complète du cycle de vie des modèles, de leur conception à leur déploiement, en passant par la surveillance continue et la gouvernance. Le MLOps permet d'automatiser ces processus pour assurer la performance, la fiabilité et l'évolutivité des modèles, tout en réduisant les interventions manuelles. Il met également l'accent sur l'importance de la reproductibilité, de l'auditabilité et de la conformité réglementaire, notamment dans un contexte d'IA responsable et éthique.

Quelles sont les 3 principales idées ?

**1. Standardisation et automatisation des workflows :** le MLOps optimise le cycle de vie des modèles de machine learning en automatisant la conception, le déploiement et la surveillance pour une gestion efficace et évolutive.

**2. Reproductibilité et auditabilité :** assurer la traçabilité et la capacité à reproduire les expériences pour faciliter le débogage, l'amélioration continue et la conformité réglementaire.

**3. IA responsable et éthique :** intégrer des principes de transparence, de sécurité et de gouvernance pour minimiser les biais et garantir un usage éthique des modèles d'IA.



# keyrus

make data matter

Du conseil en management à l'intégration des technologies digitales, Keyrus met depuis 28 ans les données au cœur de chaque transformation pour accompagner ses clients dans l'amélioration continue et durable de leur performance au travers de 5 domaines d'expertise :

- **IA & Automatisation** : accompagner les entreprises et organisations publiques dans l'optimisation de leurs processus, l'augmentation de leur productivité et de leur performance opérationnelle pour leur permettre ainsi de se concentrer sur des activités à forte valeur ajoutée.
- **Expérience digitale** : aider les entreprises à imaginer et à créer des expériences digitales multicanales inspirantes et engageantes pour atteindre leurs objectifs commerciaux.
- **Data & Analytics** : permettre aux organisations de développer et de déployer les capacités nécessaires pour donner du sens et de la valeur aux données.
- **Cloud & Sécurité** : offrir des solutions Cloud robustes, flexibles et sécurisées, garantissant la confidentialité et l'intégrité des données dans un environnement en pleine transformation.
- **Transformation & Innovation** : aider les organisations à accélérer leur transformation métier et digitale et renforcer leur agilité, résilience et compétitivité dans un contexte en perpétuelle évolution.

Avec une présence dans 28 pays et comptant plus de 3 300 experts, Keyrus est l'acteur incontournable et inspirant dans les domaines du conseil en management, de la data, du digital et bien sûr de l'IA en France et à l'international.

Pour en savoir plus : [www.keyrus.fr](http://www.keyrus.fr)

#HumanizingTheFuture

**Jean-Philippe CLAIR**

Directeur Marketing, Communication & Expérience client

[jean-philippe.clair@keyrus.com](mailto:jean-philippe.clair@keyrus.com)