



Talan[★]

DATAOPS

**Au service de la performance des
stratégies data des organisations**

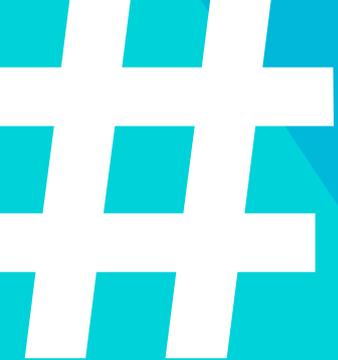
EDITO

Gartner définit le DataOps comme « une pratique collaborative de gestion des données axée sur l'amélioration de la communication, l'intégration et l'automatisation des flux de données entre les métiers de données (*Data Engineers, Data Architects, Data Stewards*) et les consommateurs de données (*Data Scientists, Business Analysts*, les métiers ou autres services) ».

Le DataOps utilise la technologie pour automatiser le déploiement et la gestion de la livraison des données. C'est un orchestrateur technologique au service de votre projet.

La promesse du DataOps est donc d'améliorer la qualité et d'optimiser le cycle de livraison des projets Data pour réduire le *time- to-market* des analyses, des prédictions.

Découvrez en quoi l'approche Talan de cette discipline est en mesure de soutenir les initiatives Data de l'entreprise.



Une petite idée de DataOps

Une petite idée de DataOps

Aujourd'hui, toutes les organisations, quels que soient leur domaine d'activité et leur taille, ont parfaitement conscience qu'il est **vital pour elles d'industrialiser et d'optimiser l'exploitation de leurs données**. Il convient alors de distinguer plusieurs « types » de données :

- Les données qui appartiennent déjà à l'organisation. Celles-ci constituent un capital souvent très sous-exploité qu'il faut apprendre à mieux connaître pour le valoriser.
- Les données produites par leur activité commerciale, industrielle ou de service,
- Les données échangées avec leurs partenaires et leurs clients.
- Toutes les données qui sont à leur portée sur Internet.

La capacité des organisations à tirer parti de cette formidable manne d'informations en la transformant en valeur est décisive dans une stratégie de transformation numérique. **La maîtrise de la data devient alors un avantage concurrentiel**. Elle permet de développer de nouveaux services, produits, méthodes, des axes d'amélioration, de développement, de diversification ou d'innovation. Un vaste programme !

“99,5% des données collectées ne sont jamais utilisées”

Les données en 2020

L'accélération de la création de données dans le monde a atteint un rythme sans précédent. Imaginez que le volume des données générées au cours des deux dernières années est plus important que celui produit dans toute l'histoire de l'humanité auparavant. En 2020, chaque être humain sur la planète va créer, tous les deux jours, suffisamment de données

pour remplir un disque dur de 256 Go.

Étonnamment, alors que depuis 2012, le Big Data a créé près de 15 millions d'emplois dans le monde, **99,5% des données collectées ne sont jamais utilisées ni analysées**.

Pourtant, le potentiel de ces données est tel que les analystes estiment que les entreprises du Fortune 1000 pourraient gagner plus de 65 millions de dollars de revenus nets supplémentaires, uniquement en augmentant leur accessibilité aux données de 10%.

Du DevOps au DataOps

Ne nous y trompons pas. DataOps ne se réduit pas à la simple transposition de DevOps (que nous connaissons dans le domainedu développement informatique) au domaine des data. Pour autant, il s'agit bien, d'une part, d'**améliorer la qualité et les cycles d'analyse et d'exploitation des données** en mettant en œuvre de nouveaux outils et de nouvelles méthodologies et, d'autre part, de **repenser l'optimisation, l'organisation du travail et de la communication des équipes**. C'est ici que s'établit le parallèle avec DevOps.

Depuis l'avènement de la Data Analytics, les experts data conçoivent des modèles pour en extraire des informations exploitables et actionnables à partir de très grands volumes de données hétérogènes. Mais cela met les opérations informatiques sous pression car **l'entreprise dépend de plus en plus de ces données analytiques qui doivent être mises en production rapidement**.

Dans le même temps, il faut **garantir et maintenir un haut niveau de confiance dans la qualité et la pertinence de ces données**, tout en se pliant aux contraintes et aux **exigences réglementaires** dans un cadre législatif et éthique en perpétuelle évolution.

La versatilité du marché, l'imprévisibilité de notre monde – qui aurait pu prédire qu'un virus bouleverserait l'économie mondiale en 2020 ? -, la réactivité des concurrents et l'exigence des clients : autant de facteurs qui impliquent une adaptation à marche forcée des entreprises.

Il est donc indispensable que les entreprises adoptent l'ensemble des pratiques qui forment le DataOps pour améliorer la coordination entre les équipes data et les opérations. Le défi est simple dans son expression : **réduire le temps, de l'identification de la donnée à son analyse, jusqu'à la production de modèles actionnables qui vont permettre de créer de la valeur et de prendre des décisions.**

“La question n'est pas seulement de produire la prochaine analyse de données. Il s'agit de construire un système qui permette de produire plus d'analyses, plus rapidement.”

DataOps vient donc de l'idée que pour créer de la valeur rapidement, il faut **détecter, intégrer, nettoyer, dédoubler, classifier, agréger, indexer, référencer, modéliser, partager, gouverner les données de façon rapide et continue** pour que toutes les équipes de l'entreprise puissent en déduire des informations de

valeur. Ces informations permettront par la suite de prendre des décisions ou de mener des actions.

Trois grands principes

Bien entendu, DataOps n'est pas une science exacte, aussi n'y a-t-il pas de méthode unique. Plusieurs approches permettent effectivement de conduire des initiatives DataOps qui supporteront la stratégie de l'entreprise. Le socle commun à toutes les approches est la **vigilance qui doit être portée à la gouvernance**. En effet, nombreuses sont les entreprises qui conçoivent et adoptent des approches en mode “Centre d'Excellence” (CoE) dans leur stratégie Data où se mêlent les exigences métiers, stratégiques, technologiques et éthiques.

Cela étant dit, des grands principes incontournables doivent être appliqués afin de produire des résultats mesurables pour l'entreprise.

Principe 1 : Démocratisation

Toutes les récentes études tendent à démontrer qu'au sein de l'entreprise, les utilisateurs métiers expriment le besoin d'accéder aux panels de données les plus vastes de manière plus autonome. Selon Experian Data Quality¹, la quasi-totalité (96%) des Chief Data Officer (CDOs) le confirment. Pourtant, malgré le poids de ces demandes, ils admettent pour 50% d'entre eux, que l'exploitation que font les entreprises des données ne leur permet pas de bénéficier ensuite pleinement du

¹ [The Chief Data Officer, Powering business opportunities with data, Experian](#)

potentiel que recèlent ces informations. Les CDOs affirment que leur priorité est de **libérer l'accès aux données en garantissant la qualité et la gouvernance via une infrastructure sécurisée qui tende vers un accès en libre-service**. Car déverrouiller l'accès aux données ne suffit pas, encore faut-il que les processus de gouvernance soient maintenus tout au long de leur cycle de vie.

Ce sont ces conditions qui permettront aux métiers d'être confiants quant à la valeur des données mises à leur disposition, et d'engager avec responsabilité et efficacité leurs programmes de transformation numérique.

Tous les freins à l'accès aux données, au-delà des aspects de confidentialité et plus généralement de respect des règles RGPD, seront de nature à limiter les initiatives de transformation numérique.

Principe 2 : Automatisation

Pour valoriser les données de façon plus rapide et efficace dans le cadre de projets qui requièrent de grands volumes de données, il est impératif **d'automatiser les étapes manuelles chronophages, fastidieuses et répétitives** qui monopolisent inutilement les équipes. Les professionnels des data du monde entier passent par exemple entre 60% et 80% de leur temps à formater, nettoyer et parfois à requalifier les données. En bout de chaîne, les **données doivent répondre à des critères de validité, de cohérence, d'intégrité, de précision, de pertinence, d'accessibilité, de cycles de vie et de niveau de mise à jour**. Cela doit également s'opérer conformément aux

critères de formats ou de standardisation de l'entreprise et aux règles de propriété et de confidentialité des données.

Les **technologies de Robotic Process Automation (RPA)**, de **Virtual Assistants** et, plus généralement, celles qui relèvent de **l'Intelligence Artificielle**, sont à même de permettre aux équipes de se concentrer sur les tâches à forte valeur ajoutée.

Principe 3 : Data Platform

Le marché impose aux métiers de mener des projets qui dépendent des données (*Data Driven*). Ces entreprises recrutent des CDOs qui participent à la transformation numérique de leur entreprise en s'efforçant de mener notamment une vraie stratégie de données pour atteindre les objectifs de l'entreprise.

L'approche DataOps, dans son implémentation de la chaîne d'intégration et de déploiement continue (CI/CD), requiert une **plateforme technique ouverte à l'environnement data** avec des outils tels que GitHub, Docker, Jenkins. Cette plateforme doit garantir la qualité et la sécurité tout en favorisant un accès en mode libre-service. Elle doit offrir la possibilité d'exécuter des traitements ou des calculs qui requièrent de très larges volumes de données ainsi que des processus conteneurisés.

La *Data Platform* ne doit en aucun cas être l'endroit où les données seront réservées uniquement à un petit groupe de data scientists ou data engineers. **Les données constituent un actif de l'entreprise à fort potentiel**. Aussi, la plateforme de données

se doit plutôt d'être un **lieu de partage accessible à tous les utilisateurs**, pour le plus grand bénéfice de l'entreprise.

“L’objectif n’est plus seulement de répondre à la question de votre client mais de pouvoir répondre rapidement à toutes les questions qui vont suivre.”

DataOps et efficacité

Les entreprises n'ont pas vocation à mettre en place de la technologie “pour la technologie” mais bien pour répondre à un objectif précis... Cela relègue alors la technologie à un outil au service de sa stratégie. L'implémentation de DataOps doit s'inscrire dans une démarche visant à servir ce même but. Mais comment en mesurer l'efficacité ?

Cette mesure se fera selon des **indicateurs concrets** liés notamment à la collaboration des équipes, aux taux d'erreur, à l'augmentation de la productivité, à la vitesse des déploiements, au bon déroulement des tests et au respect des délais.

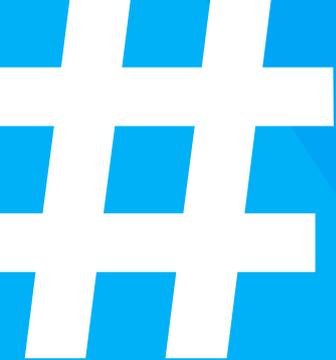
Le travail d'équipe apparaît au sein des espaces de travail virtuels collaboratifs. Leur monitoring permet de visualiser le niveau de collaboration. Le suivi des erreurs de données en production doit décrire une diminution constante jusqu'à atteindre un plancher minimum. **Ce suivi s'appuie sur des systèmes de data quality, des outils de visualisation et d'analyse.** Inversement, le monitoring de la productivité des équipes doit témoigner de l'optimisation des déploiements réussis en production. Tous ces facteurs

contribuent au respect des engagements de services (SLAs).

La représentation graphique la plus employée de DataOps décrit une boucle infinie que nous détaillerons plus loin : Vision, Définir, Planifier, Concevoir, Développer, Tester, Déployer, Monitorer, ... En effet, les **principes d'agilité et d'amélioration continue** ainsi que l'essence même des projets Data sous-tendent l'idée d'itérations.

La méthodologie et la plateforme data qui supporteront votre initiative DataOps révéleront toute leur pertinence grâce à l'expertise des équipes qui les mettront en œuvre.

Dans la deuxième partie de ce document, nous allons vous présenter l'approche Talan au travers des projets DataOps que nous avons menés auprès de nos clients.



Le DataOps au prisme de l'approche Talan

Le DataOps au prisme de l'approche Talan

Chez Talan, une équipe de notre pôle Data Intelligence Solutions a développé l'expertise de la conduite de projets DataOps. Un des enseignements que nous tirons de la conduite et de la réalisation de ces projets est sans doute que la majorité de nos clients a acquis un bon niveau de maturité quant à la **nécessité d'assurer une cohérence entre l'expression des besoins métiers et sa transposition en spécifications fonctionnelles**. Ils comprennent bien que l'adoption et l'implémentation de méthodologies agiles permettent de favoriser et de maintenir cette cohérence.

Aujourd'hui, nous avons identifié des **points d'attention de nature à compliquer l'adoption de leurs solutions** au sein de l'entreprise s'ils étaient négligés par les équipes clients. Parmi ces points, il y a notamment :

- Les régressions d'un périmètre considéré comme stable
- La complétude et la traçabilité des tests

- La sécurisation des livraisons techniques
- Le respect des normes et méthodes de développement et d'architecture

La conviction de Talan est que le propos même de DataOps est d'apporter notamment une solution opérationnelle et concrète à ces différents points.

Comme nous l'avons vu dans la première partie, l'enjeu de DataOps est de rapprocher les équipes de Data de façon générale, et plus précisément des Data Scientists et Data Engineers des équipes systèmes et d'administration.

Il s'agit de réduire le "*time-to-market*" des solutions pour **répondre le plus rapidement possible aux demandes métiers** et surtout couvrir efficacement à des points projets critiques comme :

- L'implémentation de cycles de développements courts
- La garantie d'un processus d'intégration continue
- La valorisation du projet dans un déploiement continu

"Sur le terrain, je vois de plus en plus de demandes pour intégrer de la méthodologie agile dans mes projets, parfois à très grande échelle, avec un focus sur l'optimisation des cycles de delivery data. Le DataOps devient incontournable pour rester compétitif, délivrer rapidement, de manière robuste, avec un haut niveau de qualité, en s'appuyant sur de l'automatisation couplé à de l'intégration, du développement et de l'exploration continue. Talan s'appuie sur un pool d'experts dans ces domaines pour répondre à ces enjeux, apporter de l'innovation et partager leurs riches retours d'expérience."

David Jayaseelan, Directeur Visual Analytics, Talan

La boucle DataOps

Depuis plusieurs années, les équipes Data Intelligence Solutions de Talan interviennent chez un leader français du secteur de l'énergie. La mission de Talan était de **construire une Data Factory ayant pour but de consolider, transformer, mettre en qualité et restituer les données provenant des différentes entités business.**

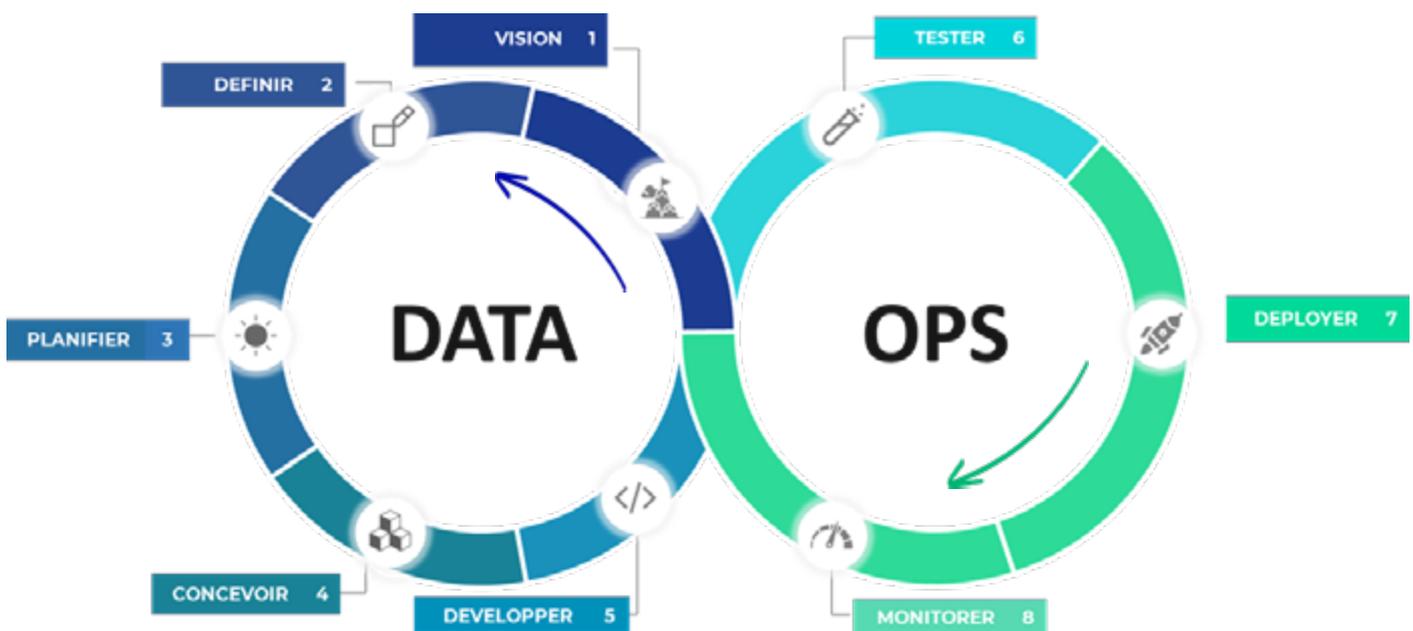
Nos équipes interviennent dans toutes les phases de la donnée : de la récolte et mûrissement des besoins à la restitution de la donnée en passant par la modélisation du Data Warehouse d'entreprise,

l'intégration, la transformation et la mise en qualité de la donnée (MDM).

Dans ce cadre, nos experts Data ont **mis en place une méthodologie DataOps décomposant successivement le flux de traitement des données en 8 phases spécifiques.** On y distingue deux principaux cycles :

- un premier **cycle orienté données** dit "DATA" ;
- un **cycle orienté administration/exploitation** dit "OPS"

Ces cycles sont respectivement identifiés en bleu et vert dans le graphique ci-dessous :



Les cinq premières étapes sont inspirées de la **méthode Scrumban**.

Pour référence, cette méthode permet dans un premier temps **d'optimiser la visualisation des workflows et leur localisation** dans l'ensemble des processus. Ensuite, elle conduit à **réduire le travail en cours** en attribuant la limite des tâches simultanées à chaque état d'un flux de travail. Enfin, elle permet de **mesurer**

et d'optimiser les délais d'exécution des processus afin de les rendre aussi courts et prévisibles que possible.

L'enjeu est notamment de permettre de **prendre des décisions au moment opportun**, d'établir un **processus d'amélioration continue** et de **minimiser tout ce qui n'apportera pas de valeur** au client.

Les trois dernières étapes, celles du cycle vert, concentrent une grande partie de nos propositions. Elles visent à **améliorer la qualité des projets en automatisant au maximum ces tâches** et en surveillant les environnements et les flux.

Voici ce à quoi correspondent ces huit étapes qu'il conviendrait de détailler dans un article spécifique.

1. **VISION** : Identifier et challenger les différentes Epics Agile
2. **DÉFINIR** : Détailler les besoins - Lotissement du projet - Détecter et définir les sources
3. **PLANIFIER** : Finaliser l'étude du besoin - Définir les phases de développement - Prioriser les user stories
4. **CONCEVOIR** : Construire le modèle de données - Transformer la maquette crayonnée en maquette graphique
5. **DÉVELOPPER** : Développer les User Stories - Ecriture de tous les cas de test - Préparer le déploiement - Documenter
6. **TESTER** : Vérifier l'exhaustivité des cas de test - Intégrer les nouveaux tests et vérifier leurs impacts dans le référentiel de test du projet
7. **DÉPLOYER** : Création d'une nouvelle release - Déployer la nouvelle version du produit
8. **MONITORER** : Exécuter le référentiel de test - Analyser l'impact sur les serveurs

Qualité et communauté : deux piliers fondamentaux des projets DataOps

Cette boucle complète repose sur deux fondamentaux que Talan valorise

particulièrement dans les projets DataOps : la qualité et la communauté (ou promotion) autour des solutions.

1. L'expérience utilisateur au service de la qualité

En se penchant sur ce qui amène de la qualité, il apparaît que l'expérience utilisateur produite par les livrables n'est absolument pas négligeable.

Ainsi, pour contribuer à l'amélioration de la qualité des projets Data, il est primordial de se livrer à un **exercice de sensibilisation globale des équipes aux bonnes pratiques des techniques de data visualisation**. Il faut également instaurer les deux étapes qui deviendront incontournables lors des phases de conception d'une *User Story* de restitution : **le crayonnage et le maquettage**. Ces deux étapes sont essentielles lors de la création d'un nouveau rapport ou d'un nouveau tableau, mais aussi lors de la refonte d'une application.

En effet, le **crayonnage sert de base lors des phases de prototypage** et permet de valider notamment l'organisation de la donnée, le cheminement de l'analyse voulue, la pertinence des graphiques par rapport au besoin et de réfléchir à la navigation au sein de l'application.

De même, le **maquettage** (sur Powerpoint ou sur un outil web) **permet d'intégrer la dimension interactive**. Par exemple, des liens peuvent être établis afin de matérialiser la navigation possible entre différentes pages.

Cela permet de présenter aux utilisateurs des éléments sur lesquels ils vont pouvoir réagir.

Aussi l'objectif est de **faciliter l'évolution de l'interface proposée au départ, afin de corriger ou valider des choix** (techniques ou fonctionnels) avant de commencer le développement du projet.

2. La promotion au service de l'adoption

Nous avons systématiquement pu le constater chez nos clients : la **promotion des outils et des réussites de l'équipe DataOps permet l'adoption des solutions**. En effet, la confiance des métiers dans le projet et leur appétence pour la solution dépend aussi des efforts de communication et de partage de la valeur de la part des équipes DataOps. Il est crucial d'**identifier et d'alimenter les sponsors du projet pour maintenir leur niveau d'engagement** et les inciter à promouvoir à leur tour la solution.

Adopter une démarche d'intégration des projets Data

Avant de conclure, voici quelques mots sur notre vision de la partie "Ops" de la boucle. **La recommandation de Talan est de ne pas s'affranchir de la mise en place d'une Testing Factory au sein des projets.**

Le but est de répondre aux besoins d'amélioration des cycles de la démarche DataOps en introduisant un framework de test industriel qui contiendra aussi bien les cas de tests purement techniques que les tests métiers et les retours de recette/anomalie. Cela amène à augmenter et à **maintenir la fiabilité de la donnée** au cours du temps, d'être **proactif dans la détection d'éventuelles erreurs**, de **réduire le coût d'exécution** grâce à l'automatisation et surtout, de **renforcer et de pérenniser la valeur des flux de données**.

Un autre volet important du DataOps est de **procéder au déploiement sous la forme d'une démarche d'intégration continue**. Cette étape est cruciale car elle crée un cadre qui favorise, pour les équipes, l'exécution de tests de façon quotidienne, sans que l'intégration des nouvelles fonctionnalités n'ait d'impact négatif sur l'existant. **L'intégration continue prend sa source dans la Testing Factory et se poursuit lors des tests de**

non-régressions en restitution.

Enfin, la dernière étape de la boucle dans la partie OPS est le monitoring. Les systèmes, une fois industrialisés comportent de nombreuses applications en RUN. Ces dernières sont soumises aux contraintes du self-service, de disponibilité, de performances et du respect de règles RGPD par exemple. **Il est crucial d'être en mesure d'identifier d'éventuelles alertes en avance de phase**. Cela peut concerner la consommation de ressources (disque, mémoire, CPU), les temps de réponse ou le cycle de vie des données par exemple.

C'est cette phase de **surveillance qui permet de maintenir le niveau de satisfaction** des utilisateurs de l'application.

Conclusion

Si vous en doutiez encore, la mise en œuvre des pratiques DataOps constitue un véritable avantage lorsque vous menez un projet Data.

Ces pratiques aident les entreprises à **améliorer la qualité de leurs projets données dont un des premiers bénéfices constatés est un impact important sur la réduction des coûts**. L'expérience montre qu'elles conduisent aussi bien à améliorer la productivité des équipes Data que l'efficacité des équipes administration/opérations de l'entreprise. **L'optimisation des initiatives Data constitue aujourd'hui une étape cruciale dans la valorisation des projets d'entreprise** qui visent à développer l'activité et prendre de l'avance sur les concurrents.

À PROPOS

Depuis plus de 15 ans, Talan conseille les entreprises et les administrations, les accompagne et met en œuvre leurs **projets de transformation en France et à l'international**.

Pour répondre aux défis d'un monde de plus en plus complexe, nous aidons nos clients à s'inscrire dans des cycles courts de transformation, de leur conception à leur mise en œuvre. Pour ce faire, nous nous appuyons à la fois sur le levier technologique et sur la force de notre ADN basé sur l'**intelligence collective, l'agilité et le goût d'entreprendre**.

L'intelligence collective consiste à conjuguer la diversité des expertises des femmes et des hommes du groupe avec celles de nos clients et de nos partenaires. L'agilité nous permet de capitaliser sur nos expériences et de nous réinventer à chaque mission avec pragmatisme et humilité.

Chez Talan, nous avons des convictions mais pas de certitudes. Nous avons le goût d'entreprendre, d'oser, d'être audacieux et n'avons pas peur d'expérimenter.

Notre expertise Data Intelligence

Les entreprises sont confrontées à la multiplication des canaux d'information, à la complexité de leur SI et la difficulté d'exploiter leurs données. Pourtant, la maîtrise de ces trois enjeux est nécessaire pour parvenir à :

- être plus **réactif face à la concurrence**
- **personnaliser l'expérience clients** pour la rendre unique
- **ouvrir et développer les usages de la donnée** dans l'entreprise pour faire communiquer les métiers entre eux
- avoir une **vision globale de l'entreprise**

Cette transformation numérique fondamentale place la **donnée au cœur de tous les modèles d'affaires et des opérations à orchestrer**. Il s'agit de pouvoir **assurer une gouvernance de ces données efficace et adaptative**.

Sur le volet technique on parle de **visualisation des données**, de **machine/reinforcement learning**, de **data science**, de **data warehousing as-a-service**, d'**intégration des données**, de **sécurité des données**, de **plateformes hybrides cloud (big/smart) data**, de **Master Data Management**, de **Robotic Process Automation (RPA)**.

Talan accompagne les organisations pour réussir concrètement cette mutation en fournissant les moyens humains, les savoirs et les expertises technologiques requis.

Pour en savoir plus, rendez-vous sur [notre site web](#).

CONTACTS

Raphaël GONCALVES

Solution Manager, Talan Solutions
raphael.goncalves@talan.com

Frederic JACQUET

Directeur de l'Innovation, Talan Solutions
frederic.jacquet@talan.com

David JAYASEELAN

Directeur Visual Analytics, Talan Solutions
david.jayaseelan@talan.com

Suivez toute l'actualité de Talan
sur [notre site web](#) et nos réseaux sociaux :



Talan[★]