

# Intégration automatique

**une nouvelle approche  
pour les données**



**Silicon** INSIGHTS FOR IT PROFESSIONALS

 **Fivetran**

# Sommaire

**03** Édito par Liliane Roux,  
Directrice commerciale EMEA

**04** Intégrer, maintenir  
et faire évoluer ses connexions

**06** Comment fonctionne  
l'intégration de données ?

**07** Spendesk bâtit un guichet  
de données métiers

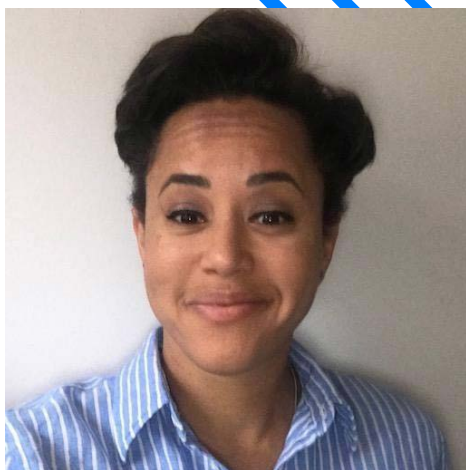
**09** Conclusion

**10** Glossaire

**11** À propos de Fivetran

ÉDITO

# « Avec l'automatisation de l'intégration, la Business Intelligence devient plus accessible »



**Liliane Roux**

**Directrice commerciale de la zone EMEA chez Fivetran**

**Longtemps réservée aux structures les plus grandes, l'exploitation à grande échelle des données devient de plus en plus accessible. À cela, deux raisons principales : les coûts de stockage et de bande-passante se sont réduits et une majorité d'applications cloud proposent des API.**

Ces évolutions du marché contribuent à l'émergence de la dernière pièce du puzzle : les pipelines de données intégrant automatiquement les données sources variées vers son data warehouse. Dans les faits, ce schéma reliant des outils métiers à une plateforme d'analyse existait auparavant. Mais la révolution de l'intégration automatique se situe dans la possibilité de se concentrer sur la valorisation des données plutôt que sur les outils techniques de consolidation des données sources. Il s'agit de libérer les data analysts, les ingénieurs affectés à l'interprétation des données, de tâches comme la création et de la maintenance de connecteurs développés sur-mesure, ou encore la synchronisation des données... Ces équipes ont besoin de connexions et d'automatismes fiables.

Nous voyons quotidiennement de plus en plus de métiers adopter une culture orientée données grâce à l'accessibilité renforcée de ces outils : à la fois en termes de coût, mais aussi d'ergonomie. Ils ouvrent la voie, pour les équipes marketing, à l'optimisation du "funnel" d'acquisition, et pour les équipes finances, à un pilotage affiné des budgets, entre autres exemples. Nous sommes ainsi fiers de compter des clients actifs aussi bien dans l'e-commerce, les solutions B2B SaaS que les FinTech comme Spendesk. Dans le contexte économique actuel, les données sont plus que jamais essentielles au pilotage de la performance et à la conduite du changement, et elles le demeureront sans nul doute. C'est parce que nous croyons au potentiel de cette technologie et à ses applications que nous continuons à investir pour l'améliorer. En juin 2020, nous avons ainsi levé 100 millions de dollars pour investir dans des produits complémentaires comme dbt (Fishtown Analytics), désormais intégré à Fivetran pour la transformation, les tests et la documentation des jeux de données des clients. ●

# 1. Intégrer, maintenir et faire évoluer ses connexions

Les usages décisionnels se heurtent souvent aux difficultés d'intégration, de maintenance et d'évolution des accès aux données. L'intégration automatique des données rend cependant les utilisateurs métiers plus autonomes.



Dans un pipeline de données, celles venant de plusieurs sources peuvent être déversées vers un entrepôt centralisé où elles seront analysées. « Les bénéfices de notre approche ETL (extract, load, transform) ? Le gain de temps et de flexibilité, la rapidité d'accès aux données à jour et les tableaux de bord de qualité, pour une prise de décision en toute confiance », résume Aymen Benazouz, ingénieur solutions de Fivetran.

A contrario, les outils ETL (extract, transform, load) traditionnels génèrent une complexité et des coûts élevés réservant les applications analytiques aux seules grandes entreprises. En effet, il faut organiser l'équipe data, embaucher des ingénieurs pour gérer cette complexité qui croît avec les changements d'API. Or, plus les intervenants sont nombreux, plus les risques d'erreurs augmentent. D'autres problèmes découlent d'une moindre qualité des données ou de la lenteur de leur collecte. « S'il faut plusieurs jours ou semaines pour construire

**ELT**  
Extract, Load,  
Transform,  
l'intégration  
des données.

*un nouveau tableau de bord, la réalité a changé entretemps. Faute de données fiables, on ne pourra pas prendre les bonnes décisions », poursuit-il.*

## Cloud et automatisation démocratisent les usages

*« Nous avons des partenariats avec les principaux data warehouses et gestionnaires de données, y compris dans le cloud où s'exécutent de plus en plus d'applications », observe l'ingénieur. Les ressources du cloud sont réservées puis payées à la consommation par l'entité utilisatrice. Grâce à cela,*



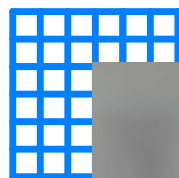
●● de nombreuses PME découvrent l'univers cloud analytique, sans avoir à maintenir leur propre infrastructure dédiée.

La simplicité du pipeline de données Fivetran leur apporte une flexibilité nouvelle et l'assurance d'analyses menées sur des données récentes. Du coup, les utilisateurs peuvent mieux planifier leurs activités et prendre des décisions "data-driven".

« Nous observons un mélange croissant parmi nos utilisateurs issus des métiers, en particulier de la finance et du marketing. Ces utilisateurs souhaitent, avant tout, une solution facile à utiliser, via une interface graphique ergonomique pour ajouter simplement de nouvelles sources de données. Ce sont rarement eux-mêmes des experts de la data. Avec Fivetran, la mise en place d'une nouvelle source s'effectue en 5 minutes, après authentification », confirme-t-il.

### Des utilisateurs autonomes, en confiance

La solution Fivetran permet de régler la fréquence de synchronisation des données de 5 minutes à 24 heures. En outre, on peut programmer une mise à jour, chaque jour, à l'heure souhaitée. La qualité des données



## « Grâce au Modern Data Stack, les responsables métiers bénéficient de tableaux de bord avec des données actualisées. »

**Aymen Benazouz, ingénieur solutions de Fivetran**

analysées dépend en partie de cette synchronisation fine des données.

« Les responsables métiers n'ont plus besoin de faire appel systématiquement aux analystes ou à l'équipe IT pour préparer les schémas de données et définir leur tableau de bord. Fivetran met à leur disposition des données récentes, dans un schéma qu'ils connaissent bien, et gère les transformations nécessaires. Les utilisateurs peuvent vérifier rapidement que leurs tableaux de bord reflètent bien la réalité du jour et que la collecte configurée est bien digne de confiance.

Pour rester conformes aux règles de l'entreprise comme au RGPD (règlement général sur la protection des données), ils peuvent masquer des colonnes ou appliquer des filtres. Des diagrammes de causes à effets, qu'on ne voyait pas auparavant, contribuent aussi à améliorer

la qualité et les performances de leurs activités. »

L'utilisateur métier doit pouvoir agréger simplement les sources de données dont il a besoin pour générer son tableau de bord de pilotage d'activités. « C'est possible grâce à l'interface utilisateur ergonomique, accessible aux non-experts ainsi qu'au nombre et à la diversité des connecteurs proposés par Fivetran. Durant la période d'essai de 14 jours, il devient vite autonome et capable de trouver,

en quelques clics, la solution recherchée. »

La sélection des données nécessaires reste assez simple pour qu'un utilisateur sans expertise IT puisse déclencher lui-même l'opération. Une fois collectées, les données sont consolidées dans le data warehouse puis, ensuite seulement, transformées. Leur agrégation et leur transformation dépendent des analyses et tableaux de bord à générer. Avec Fivetran, l'utilisateur peut intégrer une transformation programmée en SQL qui sera exécutée à chaque mise à jour. Mieux, grâce aux packages de connecteurs et à l'outillage dbt, il peut orchestrer des transformations complexes pour des tableaux de bord personnalisés. Une fois les sources identifiées et configurées, il ne reste qu'à déclencher la lecture des données. Tout le reste est automatisé ! ●

**5 min.**  
C'est le temps  
de mise en place  
d'un connecteur.

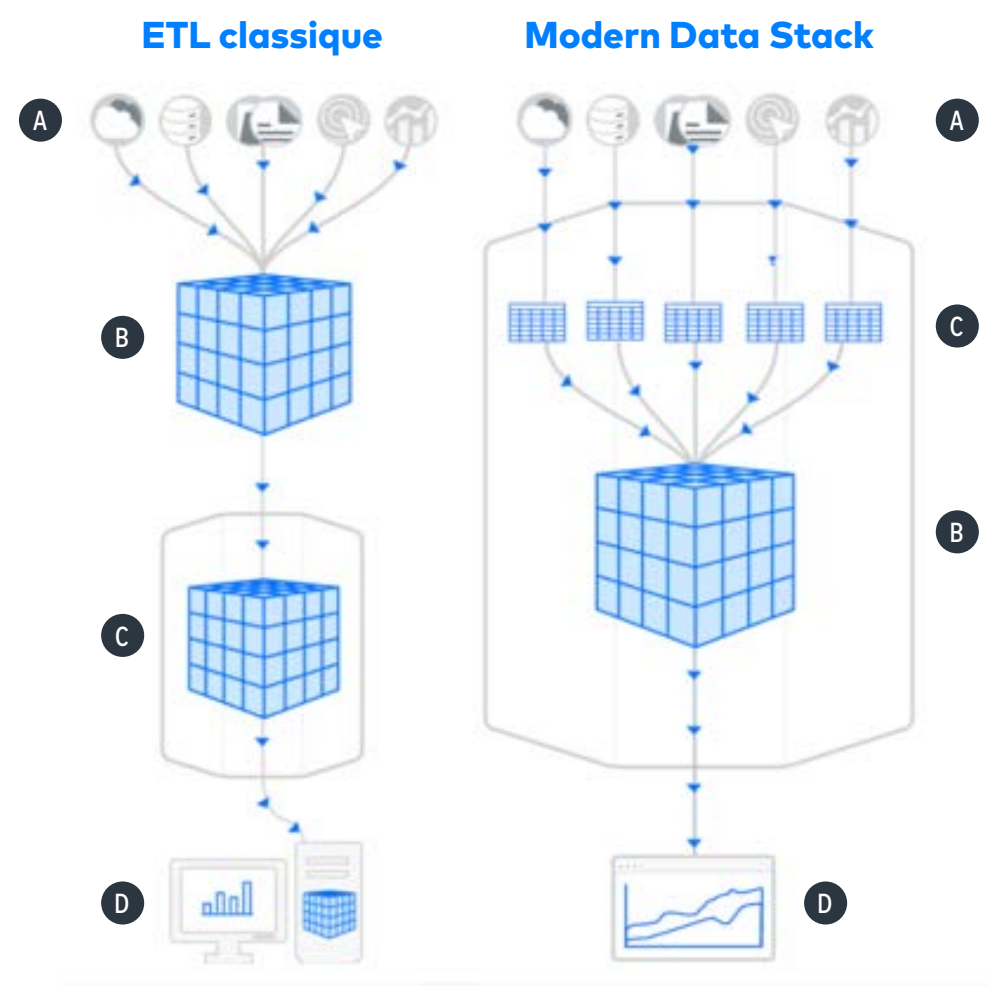
# 2. Comment fonctionne l'intégration de données ?

Intégrer automatiquement ses données, c'est mettre en place un pipeline de données automatisé, évolutif et sécurisé, soutenu par une réplication sans perte de données.

Les automatismes interviennent dès l'intégration de données. Ils libèrent les ingénieurs cherchant à intégrer de nouvelles sources d'information de nombreuses manipulations fastidieuses. Mieux encore, ils accélèrent les phases d'extraction, de chargement et de transformation, préalables à l'analyse de données. Contrairement aux outils ETL traditionnels, l'approche Modern Data Stack, aussi connue sous le terme ELT, choisie par Fivetran, est automatisée, sans configuration des pipelines de données. Elle permet aux schémas et aux modèles de données d'évoluer librement, sans exiger de reconfiguration.

En combinant l'élasticité des plateformes cloud, l'isolation de processus et le chiffrement de données sensibles, la plateforme garantit le passage à l'échelle et la fiabilité d'applications prévisionnelles et décisionnelles. ●

- A** EXTRACTION
- B** TRANSFORMATION
- C** CHARGEMENT
- D** VISUALISATION



# 3. Cas d'usage : Spendesk bâtit un guichet de données réservé aux métiers

Le spécialiste français de l'externalisation des dépenses professionnelles compte 200 salariés, dont une équipe de 10 analystes et experts data focalisée sur les prévisions utiles aux métiers. Le Modern Data Stack de Fivetran a permis à Spendesk de créer un guichet unique des données réservé aux responsables d'activités.



Toute l'organisation obtient des réponses rapides à ses questions grâce à cette nouvelle ressource accessible en libre-service après authentification. Dernièrement, un analyste a partagé un tableau de bord déterminant l'impact du confinement Covid-19 sur les ventes grâce à l'accès fiable aux dernières données disponibles.

## Des analyses simplifiées par l'intégration automatisée

À l'origine, Spendesk devait fournir un meilleur accès aux données pour mesurer les performances de ses campagnes marketing, prévoir ses ventes et segmenter ses clients. Après avoir essayé un pipeline trop exigeant en termes de configuration manuelle, l'entreprise s'est tournée vers Fivetran, une solution jugée efficace, rapide et rentable par Damien Maillard, ingénieur en chef de Spendesk : « Nous pensons que Fivetran est l'outil le plus robuste du marché ; il est performant pour répliquer rapidement nos bases

**40 heures**  
c'est le temps  
d'ingénierie  
économisé  
par semaine.

de données. Il offre une vision claire de l'historique des synchronisations et un suivi de l'évolution des structures, incluant les nouvelles tables et nouveaux schémas de données. »

Sans intégration automatisée, Spendesk aurait besoin d'un ingénieur à plein temps pour la seule phase de préparation ELT (extractions, chargements, transformations).

L'intégration automatisée des données permet aux analystes de se focaliser sur les calculs statistiques requis par chaque nouvelle expression de besoin métier. « À présent, l'équipe data peut travailler sur des



●● *analyses par anticipation, plus approfondies, pour n'importe quel service.* »

Fivetran est associé au data lake Snowflake et à l'outil dbt pour transformer les données en un datamart interrogé par Looker qui assure la visualisation des données.

Le Modern Data Stack forme un élément de réponse essentiel à l'environnement économique actuel. « *Nous devons utiliser les données pour déterminer rapidement la direction à prendre sur le marché actuel* », conclut Damien Maillard. ●

### Le Modern Data Stack de Spendesk

- **Sources :** HubSpot, Intercom, MongoDB, MySQL RDS, PostgreSQL, Salesforce, Zendesk
- **Pipeline :** Fivetran
- **Destination :** Snowflake
- **Transformations :** dbt (outil de construction de données)
- **Outil de Business Intelligence :** Looker



## Quatre bénéfices appréciés

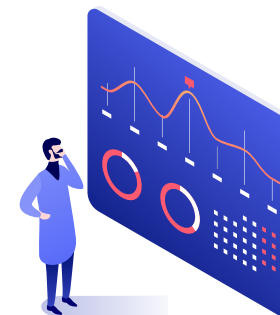


**La culture data est propagée auprès de tous les salariés au travers du guichet unique de données accessible en libre-service.**

**L'automatisation libère la moitié du temps consacré jusqu'ici aux rapports ad hoc et à l'extraction de données, soit 40 heures d'ingénierie libérées par semaine.**



**Des performances notables sont révélées, notamment au niveau de la réplique rapide des bases de données MySQL, PostgreSQL et MongoDB.**

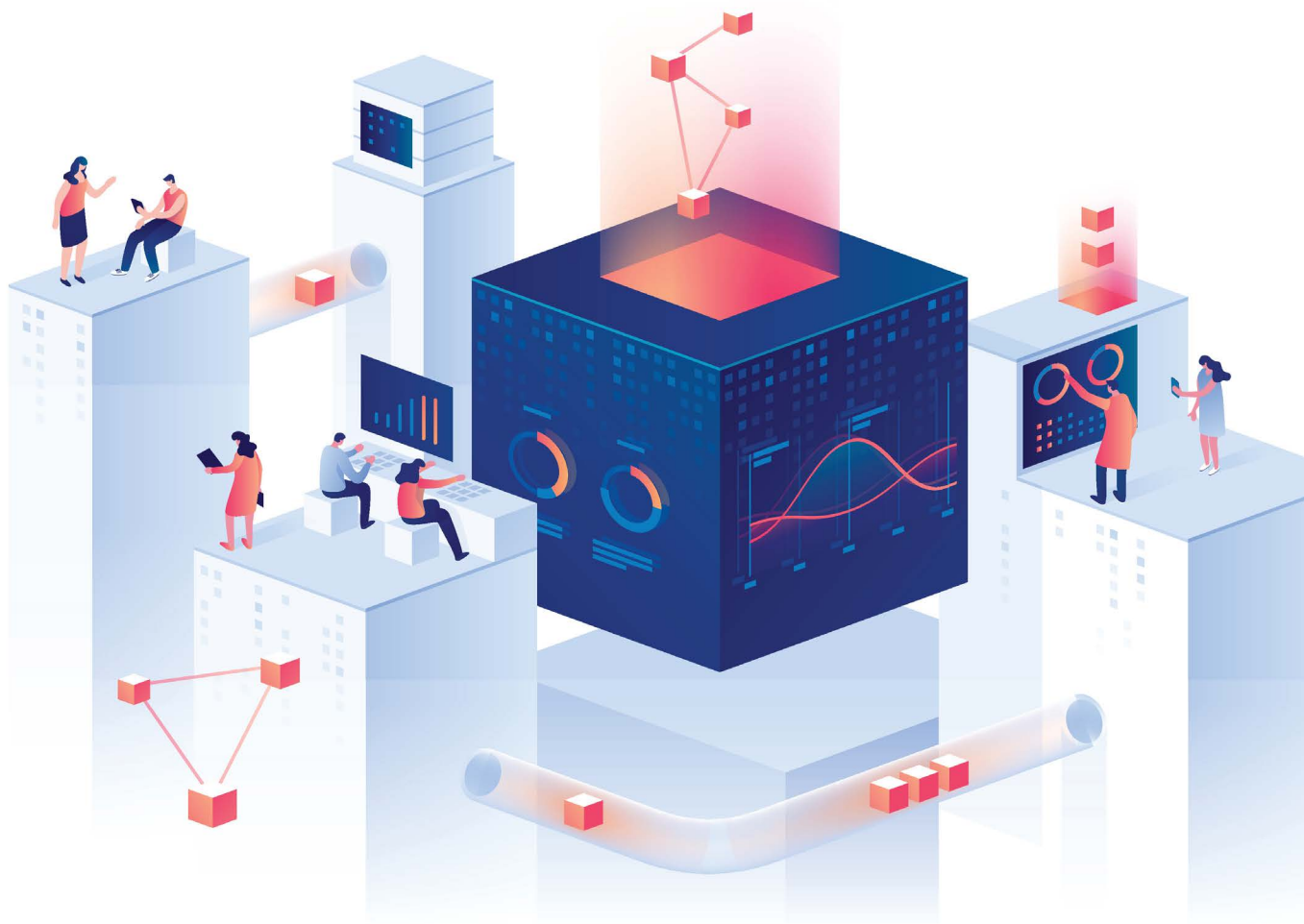


**La relance est permise au travers d'un nouveau tableau de bord qui évalue l'impact commercial du confinement.**



# Conclusion

D'ici à 2025, 93 % des données mondiales ne seront plus stockées mais collectées, traitées et analysées en mémoire et en temps réel, estime IDC. Toute organisation a besoin d'un pipeline de données automatisé et évolutif pour rester compétitive. Avec un Modern Data Stack, les décideurs, soucieux de faire les meilleurs choix stratégiques et guidés par la donnée, peuvent compter sur une solution fiable, performante et évolutive.



# Glossaire

## BI (Business Intelligence) :

la Business Intelligence – ou l'informatique décisionnelle – désigne l'ensemble des applications et bonnes pratiques délivrant l'accès aux données et permettant leur analyse, pour faciliter la prise de décision ou améliorer les performances professionnelles.

**Datalake** : l'emplacement de stockage centralisé des données massives, issues d'un grand nombre de sources distinctes. Il stocke en vrac des données structurées, semi-structurées ou non structurées.

**Datamart** : sous-ensemble du datawarehouse limité à quelques applications, à des données moins variées et moins volumineuses.

**Data warehouse** : l'entrepôt – ou gisement de données – facilite la prise de décision dans l'entreprise. Ce concept a été développé pour passer de données brutes à l'alimentation de systèmes décisionnels.

## ELT (Extract, Load, Transform) :

ce sigle désigne la suite d'outils d'intégration de données pour l'analyse de données massives. Le chargement de données brutes s'effectue directement sur la cible, permettant les transformations ultérieures.

**ETL (Extract, Transform, Load)** : ce sigle désigne la suite d'outils logiciels pour capter des données brutes, en extraire l'information utile à l'analyse, et les transformer pour les charger dans un datawarehouse.

**Pipeline de données (Data Pipeline)** : il définit l'ensemble des traitements consécutifs de données permettant à l'entreprise d'intégrer plusieurs sources d'informations pour prendre de meilleures décisions.

**Modern Data Stack** : système incluant l'extraction de données source, le workflow de transport et de transformation en données exploitables pour l'analyse, jusqu'à la synchronisation des données.

Pour aller plus loin, vous pouvez consulter nos autres ressources en français :

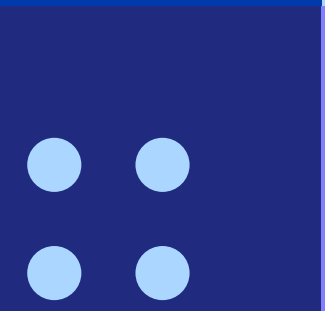
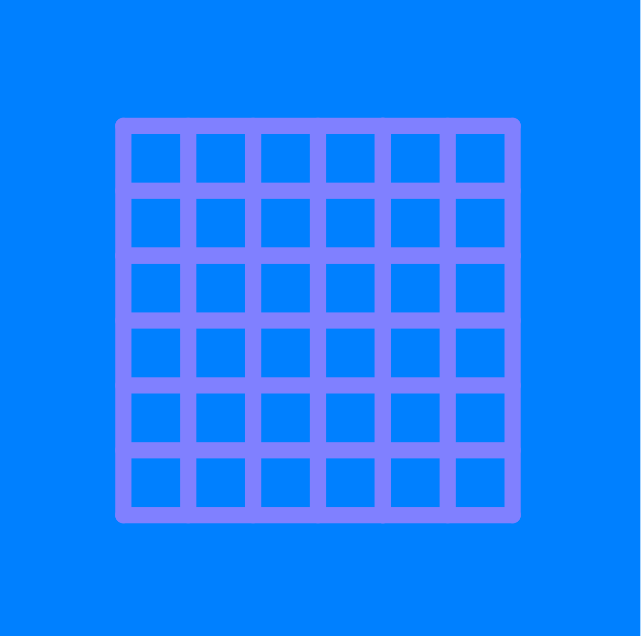
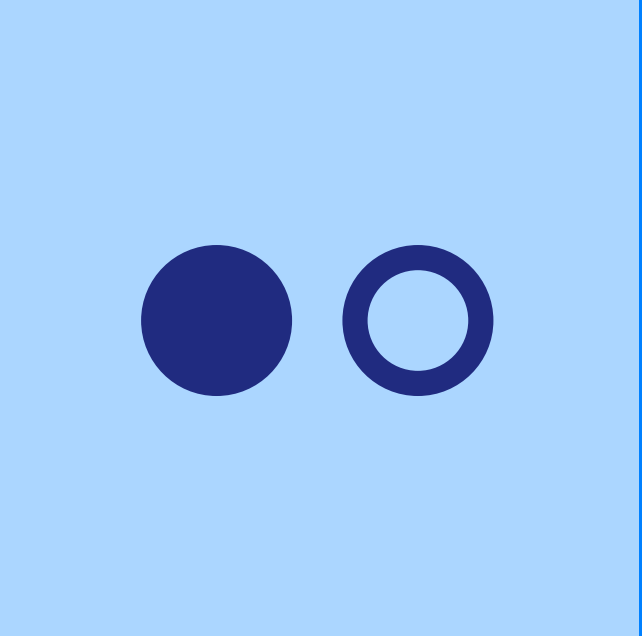
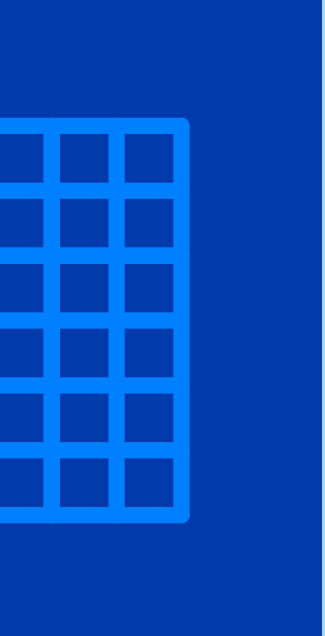
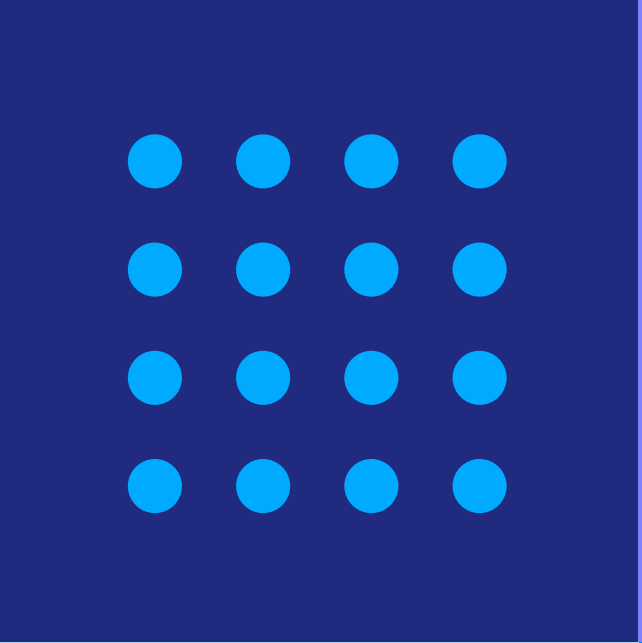
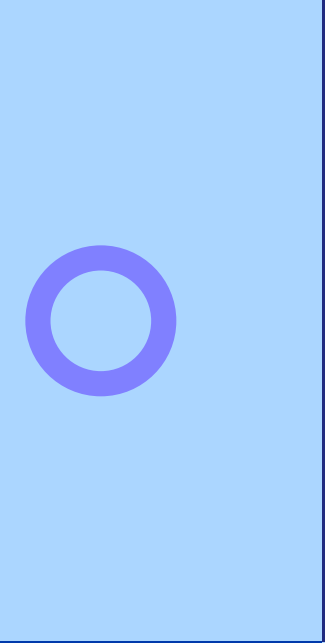
- ***Benchmark Data Warehouses Cloud 2020***
- ***Le guide essentiel de l'intégration de données***

De plus, découvrez ***les missions Fivetran*** présentées par Nathaniel Spohn, General Manager, EMEA.

# À propos de Fivetran

Leader de l'intégration automatisée des données, Fivetran fournit des connecteurs prêts à l'emploi qui s'adaptent automatiquement à l'évolution des schémas et des APIs, garantissant ainsi un accès cohérent et fiable aux données. La société améliore l'efficacité de la prise de décision grâce à la synchronisation continue des données provenant d'applications sources vers toutes les destinations. Les analystes peuvent ainsi travailler avec les données actualisées les plus récentes. Pour accélérer les facultés d'analyse, Fivetran s'adapte aux transformations au sein des data warehouses et fournit des modèles spécifiques pour chaque source de données. Comptant plus de 1 300 clients, Fivetran a pour siège Oakland en Californie, et possède des bureaux dans le monde entier.

Découvrez-en davantage sur l'intégration automatique de données capable de suivre les changements de modèles et de schémas de données sur [www.fivetran.com](http://www.fivetran.com) ou débutez un essai gratuit sur [www.fivetran.com/signup](http://www.fivetran.com/signup)



**Silicon** INSIGHTS FOR IT PROFESSIONALS

 **Fivetran**