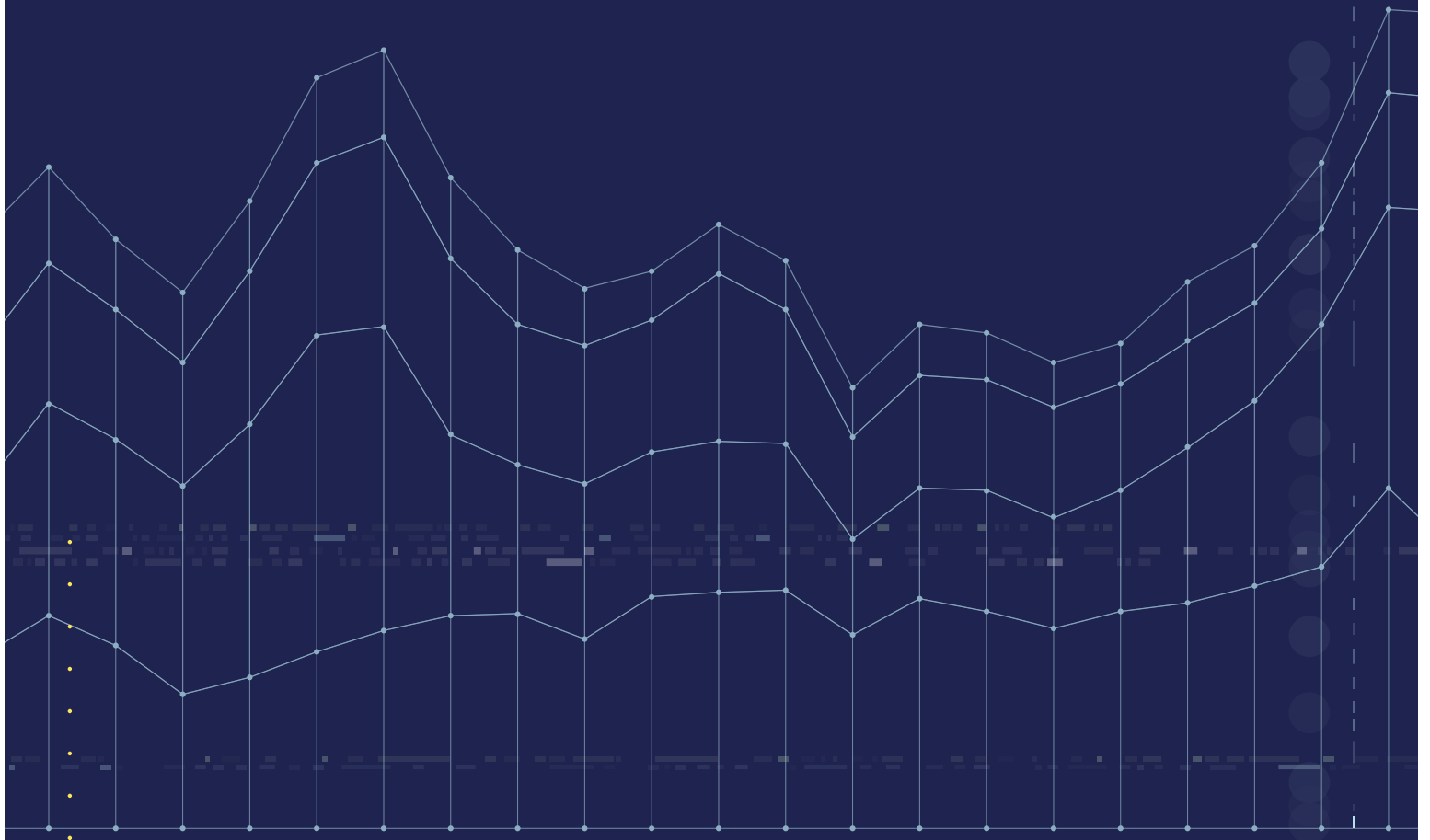




# Repensez la préparation des données

Découvrez Tableau Prep Conductor :  
la préparation de données en libre-service  
sécurisée et flexible



Ce livre blanc est destiné aux professionnels et aux décideurs du secteur IT. Il vous aidera à mieux comprendre ce qu'implique la mise en place généralisée de la préparation de données en libre-service. Vous découvrirez également comment Tableau Prep Conductor peut vous aider à gérer un environnement de préparation des données sous gouvernance. Nous partons du principe que vous connaissez déjà un peu nos solutions, notamment Tableau Prep Builder, Tableau Server, Tableau Online et Tableau Desktop.

## Sommaire

### Introduction : pourquoi démocratiser la préparation des données ?

Les défis de la préparation en libre-service dans un paysage changeant .....	3
Préparez comme un pro : comprendre tout ce qu'implique la préparation des données.....	3
Une démocratisation nécessaire .....	4
Découvrez Tableau Prep Conductor .....	5

### La préparation des données dans votre entreprise

Qui s'en charge ? Qui devrait s'en charger ? .....	7
Préparez comme un pro : attribuez les rôles essentiels.....	8
D'où viennent les données ? Où se trouvent-elles à présent ? .....	9
Comment les résultats profitent-ils à votre entreprise ? .....	9

### Repensez la préparation des données avec Tableau Prep Conductor

Comment mettre en place des flux de préparation des données ? .....	10
Préparez comme un pro : automatisez avec l'API REST .....	11
Comment surveiller vos flux de préparation des données ? .....	11
Préparez comme un pro : créez des rapports personnalisés avec le référentiel Tableau Server .....	11
Comment les problèmes sont-ils détectés ? .....	12
Comment utiliser les contrôles d'accès et assurer la sécurité des données à grande échelle ? .....	13

### Découvrez le potentiel de vos données

Comment faciliter la découverte des sources de données préparées ? .....	14
Comment aider l'entreprise à comprendre les données qui sont à sa disposition ? .....	14

La gouvernance à grande échelle .....	16
---------------------------------------	----

À propos de Tableau .....	17
---------------------------	----

Autres ressources.....	17
------------------------	----

# Introduction : pourquoi démocratiser la préparation des données ?

## Les défis de la préparation en libre-service dans un paysage changeant

Aujourd'hui, l'analytique en libre-service et la prise de décisions basées sur les données sont de mise dans les grandes entreprises partout dans le monde. La préparation des données a longtemps été dévolue à l'IT. En effet, peu de personnes avaient les compétences requises pour préparer et ajouter de nouvelles sources de données à l'entrepôt de données centralisé de leur entreprise. Mais la situation a évolué : le volume et la variété des données collectées ont explosé. Par ailleurs, les données ne sont pas toujours centralisées. Au contraire, elles se déplacent dans des pipelines, et leur ingestion, leur stockage, leur transformation et leur distribution sont déterminés au cas par cas. Enfin, les plates-formes analytiques deviennent accessibles à plus d'utilisateurs pour appuyer la prise de décision.

En général, les équipes et les employés ne relevant pas de l'IT (ou n'ayant pas accès aux processus formels de préparation des données) doivent soit attendre qu'une autre équipe prépare leurs données, soit essayer de résoudre leurs problématiques eux-mêmes. Souvent, les utilisateurs extraient les données depuis un système et les préparent dans une feuille de calcul. Cela produit des ensembles de données restructurés à usage unique. D'innombrables silos de données en résultent et plusieurs services peuvent faire le même travail sans le savoir. Ces solutions individuelles ne sont ni efficaces, ni évolutives, ni contrôlées.

L'analyse en libre-service est maintenant une pratique courante dans les entreprises qui s'appuient sur leurs données. Aussi, beaucoup de collaborateurs préparent ces dernières du mieux qu'ils peuvent avec les outils dont ils disposent, que ce soit pour les couper et coller ou pour créer des calculs gourmands qui sont loin d'être optimaux pour le serveur. Les analystes eux-mêmes indiquent que ce n'est pas l'analyse qui leur prend le plus de temps, mais le nettoyage et la réorganisation des données. Pour ces tâches, ils utilisent un processus ETL, des outils de préparation des données en libre-service ou des tableurs comme Microsoft Excel.

### **Préparez comme un pro : comprendre tout ce qu'implique la préparation des données**

Des facteurs tels que les erreurs humaines, les systèmes disparates et l'évolution des besoins des métiers peuvent contribuer à générer des données « sales ». Souvent, des étapes simples de nettoyage ne suffisent pas à assurer une bonne préparation des données. Les utilisateurs devront peut-être ajuster la granularité des données ou les transformer pour les harmoniser, puis les combiner à d'autres données avec une union ou une jointure. Une fois prêtes pour l'analyse, les données ont souvent l'air très différentes de ce qu'elles étaient dans la source originelle. Les étapes ci-dessous, et bien d'autres, peuvent intervenir dans le nettoyage, l'organisation et l'enrichissement des données :

**Permutation** : déplacer un champ des colonnes vers les lignes ou vice-versa.

**Jointure** : ajouter des champs à la source de données pour augmenter le nombre de champs à analyser.

**Union** : réunir deux ensembles de données, en gardant la même structure de table et en ajoutant des lignes.

**Filtrage ou exclusion** : exclure des valeurs ou des champs de l'analyse.

**Attribution de rôles** : confirmer les champs qui représentent une adresse e-mail, une URL ou des données géographiques.

**Modification des valeurs** : modifier manuellement une valeur ou utiliser une opération de nettoyage rapide pour changer la casse du texte ou supprimer des lettres, des chiffres, des signes de ponctuation, des espaces, etc.

**Regroupement et remplacement** : nettoyer les valeurs ayant une prononciation, une orthographe ou des caractères communs différents.

**Scission des valeurs** : scinder en plusieurs champs un champ qui contient plusieurs éléments d'informations.

**Création d'un champ calculé** : créer un champ à analyser en utilisant des calculs avec d'autres valeurs.

**Agrégation des données** : renvoyer une seule valeur à partir de plusieurs valeurs, par exemple une somme, une moyenne, un total ou un minimum.

Pour en savoir plus, lisez le livre blanc sur [les solutions aux problèmes courants de la préparation des données](#).

## Une démocratisation nécessaire

La collaboration avec les données devenant la norme, l'approche compartimentée est progressivement abandonnée. Ainsi, les employés dégagent plus facilement de nouvelles informations exploitables. Parallèlement, plus de données, au moment de leur collecte, ne sont pas prêtes ou faites pour être analysées. Il y a souvent une grande différence entre l'état optimal des données pour une collecte efficace et pour une analyse efficace. C'est le cas par exemple pour les données de streaming caractérisées par une grande vélocité ou pour les données transactionnelles. Qu'il s'agisse d'un problème de structure, de format ou de contexte, ces données ont besoin d'un nettoyage, parfois d'une curation, avant l'analyse. Il faut, par exemple, définir des règles métier et des types de demandes pour les données médicales.

De nombreuses structures adoptent des solutions de préparation des données en libre-service pour explorer et tester de nouvelles sources de données et façons d'utiliser l'analytique. Les outils de préparation en libre-service donnent le pouvoir à ceux qui connaissent le mieux les données, ce qui réduit la charge de travail de l'IT. Mais la préparation des données en libre-service est encore une compétence nouvelle qui doit être développée et déployée. Elle permettra alors aux utilisateurs de comprendre et d'exécuter des fonctions de préparation de manière efficace, d'établir des processus reproductibles et de les automatiser pour une efficacité maximale. Enfin, elle permettra de développer la confiance envers les données pour généraliser leur utilisation.

**Est-ce que tout ce travail de préparation en vaut la peine ? Selon une étude récente, elle présente de nombreux avantages qui ont déjà largement dépassé les attentes des entreprises : une vue unifiée et complète des données pertinentes à tous les niveaux de l'entreprise, réduction des silos d'analyses et amélioration de la prise de décisions basées sur les données.**

**Lisez le rapport gratuit du Business Application Research Center (BARC) sur la valorisation des données brutes lors de la préparation (en anglais)**

## Découvrez Tableau Prep Conductor

Nous allons révolutionner la préparation des données comme nous avons révolutionné l'analyse visuelle. Disponible depuis le printemps 2018, Tableau Prep Builder facilite l'accès aux données grâce à une préparation visuelle, directe et intelligente. Entièrement intégré à Tableau Desktop, il permet aux analystes et aux utilisateurs métier de préparer leurs propres données et de mieux s'immerger dans leur analyse. Avec Tableau Prep Conductor, nous étendons les fonctionnalités de préparation de données de la plate-forme Tableau : vous pouvez désormais automatiser vos flux de préparation pour qu'ils s'exécutent sans nécessiter de mise à jour manuelle lorsque vos données évoluent. Par ailleurs, vos données préparées sont plus faciles à découvrir.

**Tableau Prep Conductor** vous donne la possibilité de programmer vos flux pour qu'ils s'exécutent dans un environnement serveur centralisé, évolutif et fiable. Ainsi, vos données sont toujours à jour. De leur côté, les administrateurs ont de la visibilité sur la préparation des données en libre-service dans l'entreprise. Tableau Prep Conductor vous permet de gérer, surveiller et sécuriser les flux avec votre environnement serveur Tableau.

Intégré à Tableau Server et Tableau Online, Tableau Prep Conductor exploite leurs fonctionnalités de programmation, de surveillance et de sécurité. Tout comme les actualisations d'extraits, les tâches de flux planifiées et les exécutions de flux à la demande sont placées dans la file d'attente comme tâches d'arrière-plan. Vous pouvez publier des flux très simplement de Tableau Prep Builder vers Tableau Server ou Tableau Online. C'est similaire à la publication des sources de données et des classeurs dans Tableau Desktop.

### Des données toujours à jour, automatiquement

Programmez l'exécution de vos flux au moment qui vous convient le mieux. Automatisez les tâches d'exécution et créez un processus reproductible afin d'assurer l'uniformité de la préparation des données.

### Alertes et historique d'exécution

La vue historique vous permet de voir l'historique d'exécution de vos flux, avec les échecs et les réussites. Surveillez la qualité de vos flux de préparation grâce aux alertes intégrées qui vous informent en cas d'échec.

### Environnement de préparation sous gouvernance

Créez des règles et donnez des autorisations pour le partage et l'actualisation des données. Tirez parti de l'infrastructure et des autorisations existantes de Tableau Server ou Tableau Online pour indiquer qui peut publier, voir et exécuter des flux.

### Meilleure découvrabilité des données

Utilisez des fonctions de gestion simples, par exemple le balisage avec des mots-clés, le déplacement des flux d'un projet à l'autre ou encore la mise en place d'autorisations utilisateur, pour aider vos collègues à trouver des données préparées et pertinentes.

Home / Default / Superstore Beta Flow

## Superstore Beta Flow

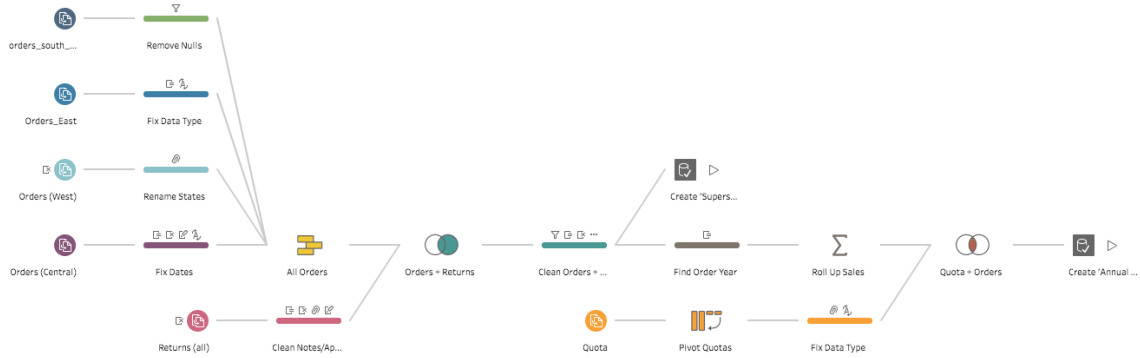
FLOW · By Youssef Shoukry · ☆ 0

**Overview** | Connections | Scheduled Tasks | Run History

Description  
Published to Prep Conductor

Tags No tags set on this flow.

Output Step	Output Name	Status	Schedule	Errors
Create 'Annual Regional Performance.tde'	Annual Regional Performance Beta	✓ Succeeded: Nov 5, 2018, 9:54 AM	Weekday 2:00AM	
Create 'Superstore Sales.tde'	Superstore Sales Beta (not yet published)	Never run	Weekday 2:00AM	



**Figure 1** Visualisez les flux de préparation des données publiés sur Tableau Server grâce à Tableau Prep Conductor.

# La préparation des données dans votre entreprise

Pour avoir tout le contexte de vos données, vous devez non seulement comprendre qui les utilise, mais aussi qui les a préparées, d'où elles proviennent, où elles sont disponibles pour l'analyse et quels bénéfices elles apportent à l'utilisateur final. Ces éléments sont fondamentaux pour la mise en place d'une préparation des données évolutive en libre-service.

## Qui s'en charge ? Qui devrait s'en charger ?

Dans de nombreuses entreprises, les analystes préparent les données pour les autres utilisateurs autant que pour eux-mêmes. La complexité des outils de préparation avancés limite leur usage à un nombre restreint d'utilisateurs confirmés. Mais les analystes et les utilisateurs métier peuvent utiliser d'autres applications ou programmes pour effectuer des tâches de préparation.

Les outils BI en libre-service ont ouvert la voie à de nouvelles possibilités pour des utilisateurs de tout niveau. Cependant, pour accéder aux informations exploitables que recèlent les données, ces mêmes utilisateurs doivent encore solliciter l'IT. Lorsque vous envisagez de déployer la préparation des données en libre-service à grande échelle, pensez aux rôles qui existent déjà dans votre entreprise et déterminez quelles responsabilités doivent être partagées ou ajoutées.

Le rôle de gestionnaire des données (ou Data Steward) peut s'ajouter à celui, plus traditionnel, d'administrateur de base de données. Aujourd'hui, le gestionnaire des données doit s'assurer que la distribution des données dans l'entreprise est fiable. Ainsi, les utilisateurs, qu'il s'agisse d'analystes, d'utilisateurs métier ou autres, peuvent vraiment exploiter tout le potentiel des données. Généralement, pour les administrateurs de base de données et les ingénieurs Data, le stockage et l'accessibilité sont les critères les plus importants. Des colonnes peuvent alors être ajoutées à l'usage exclusif de la base de données et non des utilisateurs. Lorsqu'un ingénieur crée un entrepôt de données destiné uniquement à l'analyse, il donne la priorité aux indicateurs qui permettront de répondre au plus grand nombre de questions. Si les informations recherchées ne sont pas déjà disponibles dans l'ensemble de données, les analystes peuvent être amenés à ajuster les agrégations ou à ajouter des sources externes. C'est ainsi qu'ils se retrouvent avec des silos ou des inexactitudes.

À mesure que vous étendez la préparation des données en libre-service, posez-vous ces questions : qui devrait être impliqué dans la mise en place et la mise en œuvre des pratiques de gouvernance ? Qui utilise des outils de préparation, et lesquels ? Quelles sont les formations requises ? Comment mesurer la réussite dans l'entreprise ?

## Préparez comme un pro : attribuez les rôles essentiels

Des facteurs tels que les erreurs humaines, les systèmes disparates et l'évolution des besoins des métiers peuvent contribuer à générer des données « sales ». Souvent, des étapes simples de nettoyage ne suffisent pas à assurer une bonne préparation des données. Les utilisateurs devront peut-être ajuster la granularité des données ou les transformer pour les harmoniser, puis les combiner à d'autres données avec une union ou une jointure. Une fois prêtes pour l'analyse, les données ont souvent l'air très différentes de ce qu'elles étaient dans la source originelle. Les étapes ci-dessous, et bien d'autres, peuvent intervenir dans le nettoyage, l'organisation et l'enrichissement des données :

### Rôles Professionnels IT/BI

L'**administrateur de base de données** est responsable de l'administration, de la surveillance, de la maintenance et de la sécurité des bases de données de l'entreprise. En coordination avec les ingénieurs Data et les gestionnaires des données, l'administrateur de base de données assure l'accès aux données et aide à la modélisation, la structuration et l'optimisation des sources de données accessibles dans Tableau.

L'**administrateur Système** installe, configure, gère et assure la maintenance du matériel et du système d'exploitation sur lequel est installé Tableau Server dans le data center ou le cloud. Il s'assure également que les politiques de l'entreprise sont respectées et conformes aux stratégies technologiques et métier.

L'**administrateur Client** configure les logiciels clients, en installant notamment les pilotes de base de données et les produits Tableau et en **activant Tableau Prep Conductor** dans Tableau Server ou Tableau Online.

### Rôles d'administrateurs Tableau

L'**administrateur Serveur** dispose d'un accès total aux paramètres de Tableau Server, aux sites sur le serveur, aux utilisateurs et groupes, et à tous les éléments de contenu tels que les sources de données et les classeurs, ce qui lui permet de surveiller et de préserver l'état global du serveur.

**Les administrateurs de site Tableau** créent et gèrent les utilisateurs et les groupes du site. Ils créent également des projets pour organiser le contenu sur le site et attribuent des autorisations pour permettre aux utilisateurs (groupes) d'accéder au contenu. Ils promeuvent et certifient également le contenu. Par ailleurs, ils peuvent mesurer l'utilisation de l'analytique dans leur site.

### Rôles de créateur de contenu

Un **gestionnaire de données** (avec une licence Tableau Creator) comprend le domaine d'activité et les interactions des processus métier avec l'analytique. Un gestionnaire de données s'assure que des procédures et consignes documentées ont été mises en place pour l'accès aux données et leur utilisation. Par ailleurs, il collabore avec l'administrateur de base de données ou les ingénieurs Data pour planifier et exécuter une politique de gouvernance data et de conformité à l'échelle de l'entreprise. Un gestionnaire de données peut publier des flux de préparation ou des sources de données.

Les **auteurs de contenus** (avec une licence Tableau Creator) créent et publient des tableaux de bord, des flux de préparation ou des sources de données.



## **D'où viennent les données ? Où se trouvent-elles à présent ?**

Aujourd'hui, étonnamment, de nombreuses entreprises ont très peu de visibilité sur la façon dont leurs données sont préparées. Répétition des mêmes tâches, interventions manuelles sans processus reproductibles ou encore incohérences dans les sources de données : sans approche contrôlée et normalisée, le travail de préparation et les analyses ad hoc peuvent générer des problèmes. Pour résoudre ces problèmes, il faut notamment déterminer d'où viennent les données et, une fois qu'elles sont nettoyées, où elles seront accessibles. En somme, il s'agit de déterminer le lien entre la personne qui prépare les données et celle qui les analyse.

- Comment les données sont-elles sécurisées et qui a besoin des bonnes autorisations pour y accéder et les organiser ?
- Quels utilisateurs accèdent aux sources de données brutes ? Aux données nettoyées ?
- Les utilisateurs auront-ils besoin de sources de données combinées (ou peut-être de données externes) pour approfondir leurs questions ou effectuer des analyses plus robustes ?
- Comment les sources de données préparées sont-elles partagées avec d'autres utilisateurs pour les analyses ?

Il n'est pas rare d'exporter des données dans un fichier CSV ou un autre type de feuille de calcul pour les nettoyer et effectuer une analyse ad hoc. Mais cela peut soulever des inquiétudes quant à la sécurité, car le partage de ces fichiers n'est pas toujours sécurisé. Les entreprises qui utilisent Tableau Server ou Tableau Online comme référentiel pour les sources de données et les classeurs auront une meilleure visibilité sur leurs processus de préparation des données. En effet, Tableau Prep Conductor publiera les flux sur Tableau Server ou Tableau Online. Ainsi, les utilisateurs trouveront leurs flux de préparation des données au même endroit et pourront également voir l'intégrité des flux, ce qui leur permettra d'apprendre les meilleures pratiques de la préparation des données en libre-service.

## **Comment les résultats profitent-ils à votre entreprise ?**

Aujourd'hui, étonnamment, de nombreuses entreprises ont très peu de visibilité sur la façon dont leurs données sont préparées. Répétition des mêmes tâches, interventions manuelles sans processus reproductibles ou encore incohérences dans les sources de données : sans approche contrôlée et normalisée, le travail de préparation et les analyses ad hoc peuvent générer des problèmes. L'un des facteurs majeurs pour résoudre ces problèmes consiste à identifier l'origine des données et, une fois qu'elles sont nettoyées, où elles seront accessibles. En somme, il s'agit de déterminer le lien entre la personne qui prépare les données et celle qui les analyse.

- Comment sont collectés les besoins pour les sources de données et les rapports ?
- À quelles questions les données doivent-elles permettre de répondre ?
- Quelles sont les priorités stratégiques de l'utilisateur qui accède aux données ?
- Comment permettez-vous à la fois de trouver des réponses rapides et d'explorer dans les détails ?
- Quels processus sont déjà en place pour assurer la qualité des flux de données et des sources de données publiées (par exemple, assurance qualité, certification) ?

# Repensez la préparation des données avec Tableau Prep Conductor

Pour une mise en place réussie de la préparation des données en libre-service à grande échelle, il faut réunir les personnes et les technologies dans un cadre de gouvernance qui concilie équitablement les besoins de l'IT et ceux des métiers. L'IT peut se concentrer sur les moyens donnés aux utilisateurs en automatisant les flux et sur la surveillance de l'utilisation, les performances et l'accès pour s'assurer que les pratiques de préparation des données se déploient efficacement dans l'entreprise.

## Comment mettre en place des flux de préparation des données ?

Dans certains cas, les fonctions basiques de préparation de données de Tableau Desktop suffisent. Pour l'exploration ad hoc d'ensembles de données peu volumineux, simples ou déjà nettoyés, faire pivoter une colonne ou la masquer peut permettre d'effectuer une analyse pertinente. En revanche, ce n'est pas aussi simple avec les ensembles de données plus volumineux et plus complexes ou avec les flux qui alimentent les tableaux de bord essentiels à l'entreprise. Il faut alors avoir des sources de données à jour pour une prise de décision fiable. Selon les cas, vous aurez peut-être besoin d'aller au-delà des actualisations d'extraits programmées dans Tableau Server. Plutôt que d'actualiser votre extrait, automatisez l'exécution de vos flux de préparation. Ainsi, les données passeront par les étapes de nettoyage nécessaires et vous disposerez d'une source de données prête à être analysée.

Si des collègues utilisent Tableau Prep Builder pour nettoyer leurs données, Tableau Prep Conductor vous permet de profiter automatiquement des fruits de leur travail. Programmez l'exécution des flux à un moment précis ou à intervalles réguliers grâce aux tâches planifiées de Tableau Server, ou créez des tâches de flux qui s'exécuteront selon un calendrier défini dans Tableau Online. La réexécution des flux peut s'effectuer en dehors des heures de travail sans nécessiter d'intervention humaine, ce qui permet de gagner du temps et de réduire les coûts fixes. Ainsi, les utilisateurs bénéficient d'un environnement serveur stable, au lieu d'utiliser les ressources de leur propre ordinateur pour exécuter des flux.

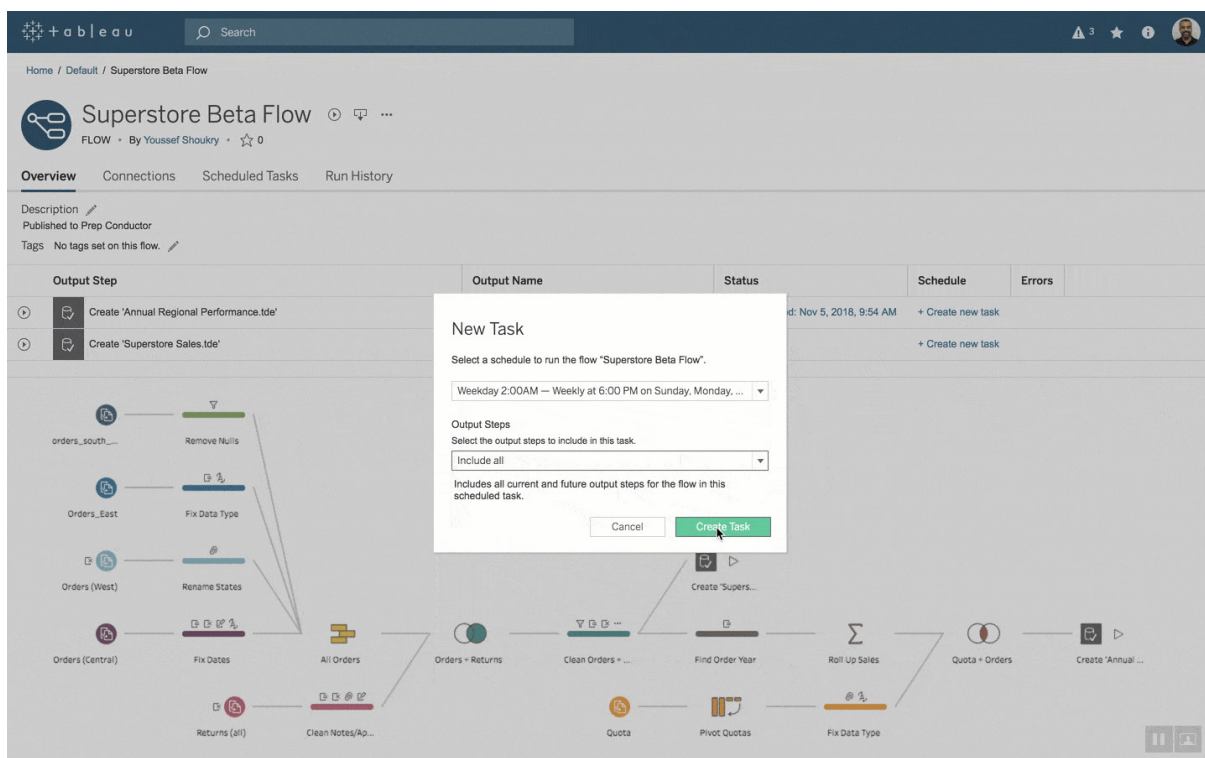


Figure 2 Créer une tâche pour exécuter régulièrement les flux de préparation selon un calendrier défini.

## Préparez comme un pro : automatisez avec l'API REST

Utilisez l'API REST pour créer des workflows avec des systèmes tiers et pour optimiser le potentiel votre pipeline de données. Cela vous permet de publier, programmer, télécharger et interroger des flux. Vous pouvez également mettre à jour vos connexions de flux, exécuter des flux ou des tâches de flux à la demande, gérer les autorisations, etc.

Pour en savoir plus, lisez [Capacités de l'API REST avec Tableau Prep Conductor](#)

## Comment surveiller vos flux de préparation des données ?

En se chargeant de la surveillance, l'IT peut vraiment assurer un bon déploiement à grande échelle des pratiques de préparation des données. Grâce au suivi des performances et de l'utilisation, à l'automatisation et aux alertes, les données sont toujours préparées à l'aide de flux efficaces. Ainsi, elles sont toujours à jour et sécurisées.

Avec Tableau Prep Conductor, les administrateurs peuvent utiliser les mêmes outils que sur Tableau Server pour surveiller les flux, notamment Tableau Services Manager (TSM), la page « État » et les vues administratives prêtes à l'emploi. Intégrées à Tableau Server et Tableau Online, ces vues peuvent vous aider à répondre à des questions essentielles sur les pratiques de préparation des données de votre entreprise. Remarque : certaines vues de Tableau Server ne sont pas disponibles ou pertinentes dans Tableau Online.

- **Performances des exécutions de flux** : vous aide à comprendre quelles tâches de flux sont actuellement programmées, quelles tâches sont en cours d'exécution, la durée des tâches de flux, lesquelles s'exécutent le plus souvent et la proportion de flux ad hoc par rapport aux flux programmés.
- Actions par **tous les utilisateurs**, par **certains utilisateurs** ou par **utilisateurs récents** : cette dernière option peut être très utile lorsque vous souhaitez effectuer une maintenance sur le serveur et que vous avez besoin de savoir quel type d'actions les utilisateurs effectuent et quel effet aura la maintenance sur leurs activités.
- **Statistiques d'utilisation de l'espace** : vous aide à identifier les sorties de flux qui prennent le plus d'espace disque sur le serveur.
- **Retard des tâches d'arrière-plan** : cette vue vous aide à déterminer comment améliorer les performances du serveur en optimisant les tâches et en répartissant les programmations.

## Préparez comme un pro : créez des rapports personnalisés avec le référentiel

### Tableau Server

En plus des vues administratives prêtes à l'emploi, vous pouvez utiliser Tableau Desktop pour analyser vous-même l'activité du serveur. Pour ce faire, vous pouvez vous connecter à vos vues ou envoyer des requêtes dans le référentiel Tableau Server, une base de données PostgreSQL.

Découvrez comment créer des vues personnalisées pour Tableau Server sur [Windows](#) ou [Linux](#).

## Comment les problèmes sont-ils détectés ?

Problèmes de connexion, erreurs dans un flux... Soyez attentif aux problèmes qui peuvent ralentir vos flux de préparation des données. Pour réduire le risque d'utilisation de données obsolètes, Tableau Prep Conductor vous informe dès qu'un problème survient pendant l'exécution de vos flux et vous suggère des corrections.

- **Historique d'exécution** : avec l'historique des actualisations d'un flux, les utilisateurs peuvent voir rapidement les exécutions qui ont réussi ou échoué. Ainsi, vous pouvez mieux estimer la qualité de tous vos flux et avoir davantage confiance en l'exactitude de vos données.
- **Alertes** : des alertes vous indiquent si vos flux fonctionnent correctement. En cas d'erreur, les utilisateurs reçoivent dans les meilleurs délais une notification par e-mail ainsi qu'une alerte dans l'interface de Tableau Server. Les liens fournis dans l'alerte permettent aux analystes et aux gestionnaires de données d'agir rapidement : ils peuvent examiner les erreurs, les corriger en suivant les suggestions, puis reprendre leur préparation ou leur analyse.
- **Pages État** : les pages État de Tableau Server et de Tableau Services Manager contiennent les processus **Tableau Server**, ainsi que des liens vers de la documentation de dépannage à consulter si un processus ne s'exécute pas comme prévu. Si vous passez le curseur sur un indicateur d'état, une infobulle apparaît, avec le nom du nœud et le port sur lequel le processus s'exécute.

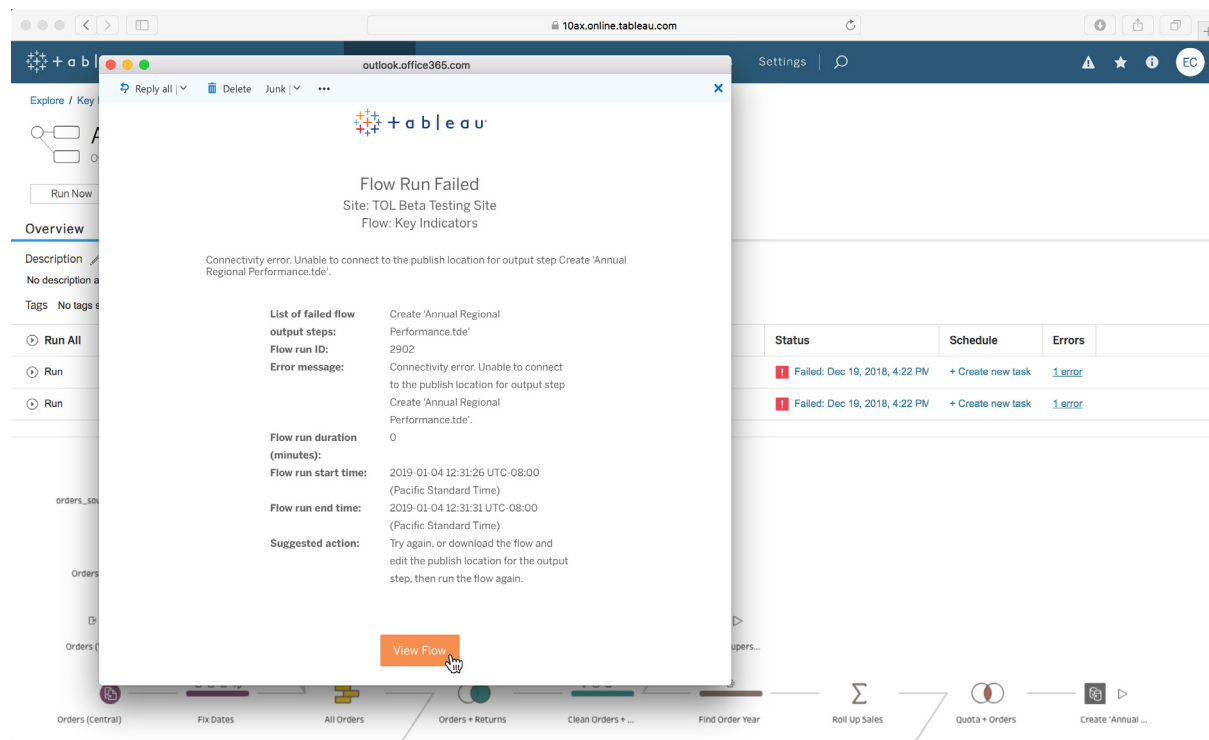


Figure 3 Recevez des alertes lorsqu'une exécution de flux échoue : par e-mail et sur Tableau Online.



Un outil de préparation des données doit permettre aux analystes de répondre aux questions ponctuelles et être itératif.

**GORDON STRODEL**

CONSULTANT GESTION DE L'INFORMATION ET ANALYTIQUE, SLALOM

### Comment utiliser les contrôles d'accès et assurer la sécurité des données à grande échelle ?

Pour des flux suivis de près et bien encadrés, il est essentiel de s'assurer que les utilisateurs concernés ont accès aux flux de préparation des données. Les administrateurs IT peuvent gagner du temps et s'alléger la tâche avec les commandes d'autorisations consolidées dans Tableau Server ou Tableau Online. Vous pouvez définir les autorisations des flux de préparation gérés par Tableau Prep Conductor lorsque vous les publiez. Vous pouvez par exemple choisir qui peut les voir, les modifier, les exécuter, etc. Si un flux se connecte à des bases de données, vous pouvez spécifier le type d'authentification et définir des identifiants permettant d'accéder aux données.

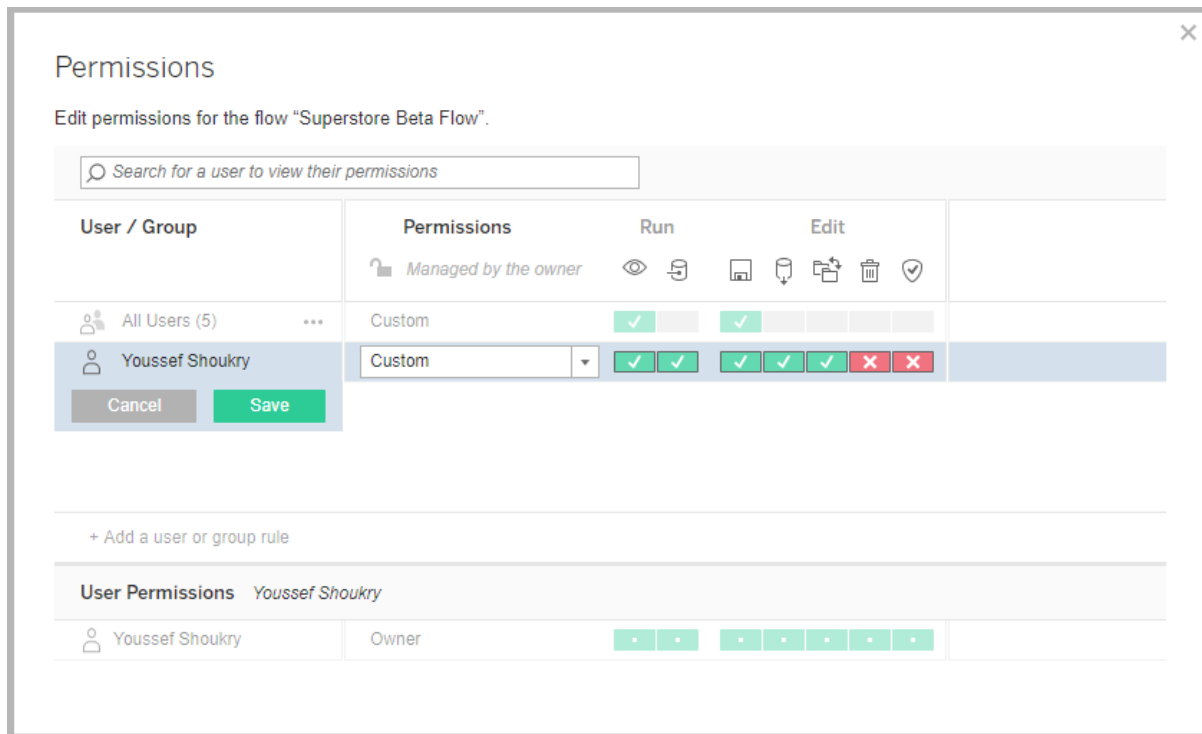


Figure 4 Gérez et personnalisez vos autorisations de flux dans une interface simple et familière.

De plus, les vues administratives fournissent des rapports avec les autorisations et les actions consultables pour un utilisateur spécifique ou pour tous les utilisateurs. Cela vous permet de mieux comprendre qui fait quoi pour chaque flux.

# Découvrez le potentiel de vos données

Pour mieux tirer profit des données préparées, votre entreprise doit pouvoir les découvrir, comprendre leur pertinence et leur faire confiance. C'est d'autant plus crucial à mesure que l'utilisation des données se démocratise. C'est dans ce contexte que la préparation des données en libre-service peut alimenter et faire progresser l'analyse en libre-service dans un environnement sous gouvernance, comme Tableau Server ou Tableau Online.

## Comment faciliter la découverte des sources de données préparées ?

Tableau Prep Conductor dispose de nombreuses fonctionnalités de gestion qui aident les utilisateurs à découvrir plus facilement des données pertinentes et fiables lorsqu'ils recherchent, filtrent et trouvent des flux et des sources de données publiées sur Tableau Server ou Tableau Online.

**Balisage** : appliquez des mots-clés à des flux pour aider les utilisateurs à trouver, filtrer et catégoriser le contenu, comme dans un classeur. Vous pouvez ajouter une balise à un flux individuel ou à plusieurs flux en même temps.

**Organisation** : déplacez les flux d'un projet à l'autre pour qu'ils restent bien organisés avec les bonnes sources de données, les bons classeurs et tout autre contenu pertinent. Par défaut les propriétaires des flux peuvent effectuer cette action, mais d'autres utilisateurs peuvent demander l'autorisation Déplacer aux administrateurs concernés.

**Gestion du cycle de vie** : des fonctionnalités de gestion supplémentaires, comme l'autorisation d'enregistrer, de renommer ou de supprimer des flux si nécessaire, vous aideront à garantir la pertinence et la bonne organisation des flux de préparation des données. Les administrateurs, les propriétaires des flux et les propriétaires des projets peuvent également réattribuer la propriété des flux.

La possibilité de rechercher des flux de données déjà existants dans Tableau Server ou Tableau Online vous permettra également d'éviter que deux utilisateurs effectuent la même tâche de préparation sans le savoir. En effet, si un utilisateur trouve une source de données dont le flux de préparation convient à ses besoins, il n'a pas besoin de passer du temps à reproduire les étapes de préparation ou à exécuter le flux. Il peut également trouver un flux existant qu'il peut télécharger et modifier pour une utilisation similaire, sans avoir besoin de créer le flux de préparation à partir de zéro.

## Comment aider l'entreprise à comprendre les données qui sont à sa disposition ?

Toute entreprise qui souhaite déployer la préparation des données et l'analyse en libre-service à grande échelle se doit de développer la connaissance des données. Beaucoup mettent en place un centre d'excellence dans lequel l'IT et les champions de l'analytique fournissent des ressources et de l'aide aux autres utilisateurs. Composé notamment d'un groupe d'utilisateurs internes, ce centre investit dans des formations pour le développement des compétences analytiques. Pour évaluer les besoins, il faut prendre en compte le type et le nombre d'utilisateurs qui utiliseront des outils tels que Tableau Prep Builder et Tableau Prep Conductor pour la préparation des données en libre-service.

S'ils savent comment une source de données Tableau a été créée, les analystes et les utilisateurs métier lui feront davantage confiance. Avec Tableau Prep Conductor, tous les utilisateurs peuvent voir l'origine d'une source de données créée par un flux et y accéder directement pour voir comment elle a été construite. Avec un niveau de compréhension basique des données et la possibilité de « lire » les étapes d'un flux, l'utilisateur devrait être en mesure de déterminer si une source de données lui sera utile ou non.

Les types de licences Tableau Creator et Tableau Explorer leur permettent de se connecter aux sources de données ou encore de modifier, personnaliser ou créer du contenu. Pour cela, l'IT et les gestionnaires de données doivent avoir mis en place un processus qui certifie les sources de données exploitant des flux de préparation exécutés régulièrement. Avec des sources de données certifiées, l'entreprise sait qu'elle dispose de données fiables et prêtes à être analysées. Les sources de données certifiées seront également mises en avant dans les résultats des recherches et dans les contenus filtrés sur Tableau Server ou Tableau Online.



La préparation visuelle donne une visibilité de bout en bout et permet de déceler les erreurs potentielles dès le départ, qu'il s'agisse de fautes d'orthographe, d'espaces en double ou de clauses de jointure incorrectes. Cela renforce également la confiance dans l'analyse finale.

**JASON HARMER** CONSULTANT, NATIONWIDE INSURANCE

# La gouvernance à grande échelle

Chaque entreprise a des besoins spécifiques : il n'existe pas de solution universelle en matière de préparation des données. Toutefois, au moment de choisir un outil en libre-service, il faut tenir compte de sa capacité à évoluer au fil du temps pour permettre une approche itérative et agile. Les utilisateurs auront davantage envie de préparer et comprendre leurs données s'ils peuvent connaître les effets de chaque étape de la préparation.

C'est pour cela que la collaboration et la gouvernance sont essentielles : un utilisateur peut essayer de régler ses problèmes de données lui-même, mais l'IT joue un rôle majeur dans la résolution des problèmes au niveau de l'entreprise. Les pratiques de gouvernance permettent aux bonnes personnes d'accéder aux bonnes données, assurent que les données utilisées pour la prise de décision sont fiables, et garantissent aussi bien le respect des politiques internes que des réglementations externes à l'entreprise.

Il ne s'agit pas de demander à l'IT de renoncer au contrôle, mais de permettre à tous les utilisateurs d'être plus autonomes dans un environnement fiable et centralisé. Les analystes et les utilisateurs métier sont les premiers à identifier les problèmes ou irrégularités liés aux données dans un modèle de gouvernance défini conjointement par l'IT et les métiers.

Tout comme le passage à l'analytique en libre-service, la participation des utilisateurs métier dans la gouvernance pour démocratiser la préparation des données amène son lot de défis : conduite du changement pour les processus et les technologies, limitation des risques liés à la sécurité, compétences insuffisantes des utilisateurs, etc. Toutefois, gardez bien à l'esprit qu'avec une approche agile et itérative du déploiement et une approche collaborative de la gouvernance, étendre l'accès à la préparation des données peut se révéler plus bénéfique que prévu.



Avant Tableau Prep, notre équipe passait des heures et des heures à vérifier que les sources de données étaient nettoyées et organisées, simplement pour s'assurer de la fiabilité de l'analyse. Tableau Prep nous a permis de gagner énormément de temps, en changeant totalement notre façon de regarder nos données et en réduisant considérablement le délai entre la collecte et les informations exploitables.

**GESSICA BRIGGS-SULLIVAN** ADMINISTRATRICE TABLEAU, CHARLES SCHWAB, INC.



# À propos de Tableau

Tableau est une plate-forme BI visuelle, exhaustive et facile à utiliser, conçue pour l'entreprise, qui vous permet de voir et comprendre vos données grâce à une analytique en libre-service rapide et évolutive. Sur site ou dans le cloud, sous Windows ou Linux, Tableau s'appuie sur vos investissements technologiques et évolue en fonction de vos besoins en s'adaptant à votre environnement de données. Libérez la puissance de vos ressources les plus précieuses : vos données et vos collaborateurs.

## Autres ressources

[En savoir plus : La préparation des données avec Tableau](#)

[En savoir plus : La gestion des données avec Tableau](#)

[Aide en ligne : Tableau Prep Conductor](#)

[Livre blanc : Les données mal préparées ont un coût : 4 façons d'aborder les problèmes courants de la préparation de données](#)

[Livre blanc : Meilleures pratiques pour nettoyer les données « sales »](#)

[Étude du BARC : Préparation des données : valorisez vos données brutes \(en anglais\)](#)

[Tableau pour l'entreprise : L'IT rend l'analytique accessible à tous](#)