ATLASSIAN

# Revue post-incident

## Panne d’avril 2022

*Cette traduction n'est fournie que pour des raisons de commodité. En cas d'ambiguïté ou de conflit entre les traductions, la version originale anglaise prévaut.*

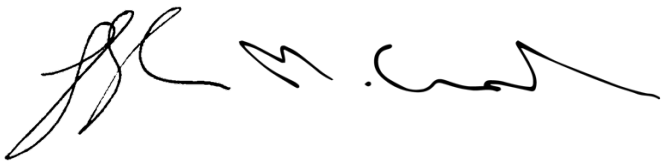
# Lettre de nos cofondateurs et codirecteurs généraux

Nous sommes pleinement conscients qu'une panne de services a affecté nos clients au début du mois d'avril. Nous avons conscience que nos produits sont essentiels pour votre activité, et nous ne prenons pas cette responsabilité à la légère. La responsabilité nous incombe. Un point c'est tout. Pour les clients touchés, nous nous employons à regagner votre confiance.

> Chez Atlassian, l'une de nos valeurs fondamentales est « oui à la transparence, non au baratin ». Nous incarnons cette valeur en partie en discutant ouvertement des incidents et en les utilisant comme des opportunités d'apprentissage. Nous publions cette revue post-incident pour nos clients, notre communauté Atlassian et la communauté technique au sens large. Atlassian est fière de son processus de gestion des incidents qui montre qu'une culture sans reproches combinée à la volonté d'identifier les moyens d'améliorer nos systèmes et processus techniques sont essentiels pour offrir un service de grande qualité et digne de confiance. Bien que nous fassions de notre mieux pour éviter tout type d'incident, nous partons également du principe que les incidents sont une manière efficace de nous améliorer.

Soyez assurés que la plateforme cloud d'Atlassian permet de répondre aux besoins variés de plus de 200 000 clients cloud de toutes tailles et de tous les secteurs. Avant cet incident, notre cloud a toujours assuré 99,9 % de temps d'activité et dépassait les accords de niveau de service (SLA) de disponibilité. Nous avons investi à long terme dans notre plateforme et dans un certain nombre de fonctionnalités centralisées de la plateforme, avec une infrastructure évolutive et une cadence régulière d'améliorations en matière de sécurité.

Nous tenons à remercier nos clients et nos partenaires pour leur confiance et leur partenariat continus. Nous espérons que les détails et actions décrits dans ce document démontrent qu'Atlassian continuera à fournir une plateforme cloud de classe mondiale et un puissant portefeuille de produits permettant de répondre aux besoins de chaque équipe.



-Scott et Mike

# Synthèse

Le mardi 5 avril 2022, à compter de 7 h 38 UTC, 775 clients Atlassian ont perdu l'accès à leurs produits Atlassian. La panne a duré jusqu'à 14 jours pour un sous-ensemble de ces clients, le premier groupe de clients ayant récupéré leur accès le 8 avril et le reste des sites client ayant été progressivement restauré au plus tard le 18 avril.

L'incident n'est pas le résultat d'une cyberattaque, et il n'y a pas eu d'accès non autorisé aux données des clients. Atlassian dispose d'un programme complet de gestion des données avec des contrats de niveau de service publiés et un historique de dépassement de ces SLA.

Bien qu'il s'agisse d'un incident majeur, aucun client n'a perdu plus de cinq minutes de données. En outre, plus de 99,6 % de nos clients et utilisateurs ont continué d'utiliser nos produits cloud sans aucune interruption pendant les activités de restauration.

> Dans ce document, nous appelons les clients dont les sites ont été supprimés dans le cadre de cet incident les clients « touchés » ou « impactés ». Cette revue post-incident fournit les détails exacts de l'incident, décrit les mesures que nous avons prises pour y remédier et indique comment nous allons éviter que de telles situations ne se reproduisent à l'avenir. Dans cette section, nous fournissons un résumé général de l'incident, avec des détails plus précis dans le reste du document.

## Que s'est-il passé ?

En 2021, nous avons procédé à l'acquisition et à l'intégration d'une application Atlassian autonome pour Jira Service Management et Jira Software appelée Insight – Asset Management. Cette application autonome a alors été intégrée en tant que fonctionnalité native dans Jira Service Management et n'était plus disponible pour Jira Software. De ce fait, nous avons dû supprimer l'ancienne application autonome sur les sites des clients qui l'avaient installée. Nos équipes d'ingénieurs ont utilisé un script et un processus existants pour supprimer les instances de cette application autonome, mais deux problèmes se sont posés :

- **Communication insuffisante.** Tout d'abord, la communication entre l'équipe qui a demandé la suppression et celle qui l'a exécutée était insuffisante. Au lieu de fournir les identifiants de *l'app prévue* pour suppression, l'équipe a fourni les identifiants de *l'ensemble du site cloud* sur lequel les apps devaient être supprimées.

- **Avertissements système insuffisants.** L'API utilisée pour effectuer la suppression acceptait les identifiants de site et d'application et supposait que la saisie était correcte. Cela signifiait que si un identifiant de site était transmis, un site serait supprimé ; si un identifiant d'application était transmis, une application serait supprimée. Aucun signal d'avertissement n'a été émis pour confirmer le type de suppression (site ou application) demandé.

Le script qui a été exécuté utilisait notre processus standard de revue par des pairs, qui se concentrait sur le point de terminaison appelé et de quelle manière il l'était. Il ne vérifiait pas si les identifiants des sites cloud fournis faisaient référence à Insight App ou au site complet. Il s'est avéré que le script contenait l'identifiant pour le site client complet. Le résultat a été la suppression immédiate de 883 sites (représentant 775 clients) entre 7 h 38 UTC et 8 h 01 UTC le mardi 5avril 2022. *Voir « Ce qui s'est passé »*

## Comment avons-nous réagi ?

Une fois l'incident confirmé le 5 avril à 8 h 17 UTC, nous avons déclenché notre processus de gestion des incidents majeurs et formé une équipe interfonctionnelle de gestion des incidents. L'équipe mondiale de réponse aux incidents a travaillé 24 h/24 et 7 j/7 pendant toute la durée de l'incident jusqu'à ce que tous les sites touchés aient été restaurés, validés et restitués aux clients. En outre, les responsables de la gestion des incidents se réunissaient toutes les trois heures pour coordonner les flux de travail.

Très vite, nous nous sommes rendu compte de l'ampleur de la tâche qui consistait à restaurer des centaines de clients avec plusieurs produits simultanément.

Au début de l'incident, nous savions exactement quels sites étaient touchés, et notre priorité était d'établir la communication avec le propriétaire agréé de chaque site concerné afin de l'informer de la panne.

Cependant, certaines informations de contact clients étaient supprimées. Par conséquent, les clients n'ont pas pu déposer de tickets de support comme ils l'auraient fait normalement. Cela signifie également que nous n'avons pas eu immédiatement accès aux contacts clés des clients. *Pour plus de détails, voir « Vue d'ensemble de haut niveau des flux de travail de récupération »*

## Que faisons-nous pour éviter ce type de situation à l'avenir ?

Nous avons pris un certain nombre de mesures immédiates et nous nous sommes engagés à apporter des changements pour éviter qu'une telle situation se reproduise à l'avenir. Voici quatre domaines spécifiques dans lesquels nous avons apporté ou apporterons des modifications importantes :

1. **Établir des « suppressions en douceur » universelles sur tous les systèmes**. Dans l'ensemble, une suppression du type concerné par l'incident doit être interdite ou comporter plusieurs niveaux de protection pour éviter les erreurs, y compris un déploiement échelonné et un plan de restauration testé pour les « suppressions en douceur ». Nous empêcherons globalement la suppression des données clients et des métadonnées qui n'ont pas fait l'objet d'un processus de suppression en douceur.

2. **Accélérer notre programme de reprise d'activité (DR) afin d'automatiser la restauration des événements de suppression multisites et multiproduits pour un plus grand nombre de clients.** Nous exploiterons l'automatisation et les leçons tirées de cet incident pour accélérer le programme de reprise d'activité afin d'atteindre l'objectif de temps de récupération (RTO) tel que défini dans notre politique pour cette ampleur d'incident. Nous organiserons régulièrement des exercices de reprise d'activité qui impliquent la restauration de tous les produits pour un grand nombre de sites.

3. **Réviser le processus de gestion des incidents pour les incidents de grande échelle**. Nous améliorerons notre procédure opérationnelle standard pour les incidents de grande envergure et la mettrons en pratique avec des simulations de cette ampleur d'incident. Nous mettrons à jour notre formation et nos outils pour prendre en charge un grand nombre d'équipes travaillant en parallèle.

4. **Créer un playbook de communication sur les incidents à grande échelle.** Nous reconnaîtrons les incidents rapidement, par le biais de plusieurs canaux. Nous publierons des communications publiques sur les incidents dans les heures qui suivent. Pour mieux atteindre les clients concernés, nous améliorerons la sauvegarde des contacts clés et adapterons les outils d'assistance afin de permettre aux clients ne disposant pas d'une URL valide ou d'un identifiant Atlassian d'entrer en contact direct avec notre équipe d'assistance technique.

Notre liste complète des éléments d'action est détaillée dans la revue complète post-incident ci-dessous. *Voir « Comment allons-nous nous améliorer ? »*

# Sommaire

# Présentation de l'architecture cloud d'Atlassian

Pour saisir les facteurs qui ont contribué à l'incident, tels qu'ils sont exposés tout au long de ce document, il est utile de comprendre tout d'abord l'architecture de déploiement pour les produits, les services et l'infrastructure d'Atlassian.

## Architecture de l'hébergement cloud d'Atlassian

Atlassian utilise Amazon Web Services (AWS) comme fournisseur de services cloud et ses datacenters hautement disponibles dans de nombreuses régions du monde. Chaque région AWS est un emplacement géographique distinct comprenant plusieurs groupes de datacenters isolés et physiquement séparés, appelés zones de disponibilité (AZ).

Nous utilisons les services de computing, de stockage, de réseau et de données d'AWS pour créer nos produits et les composants de notre plateforme, ce qui nous permet d'utiliser les capacités de redondance offertes par AWS, telles que les zones et les régions de disponibilité.

## Architecture de services distribuée

Grâce à cette architecture AWS, nous hébergeons un certain nombre de services de plateforme et de produits qui sont utilisés sur l'ensemble de nos solutions. Ces services incluent les fonctionnalités de la plateforme qui sont partagées et utilisées par plusieurs produits Atlassian, tels que Media, Identity, Commerce, des expériences comme notre éditeur, ainsi que des fonctionnalités spécifiques aux produits, telles que le service Jira Issue et Confluence Analytics.

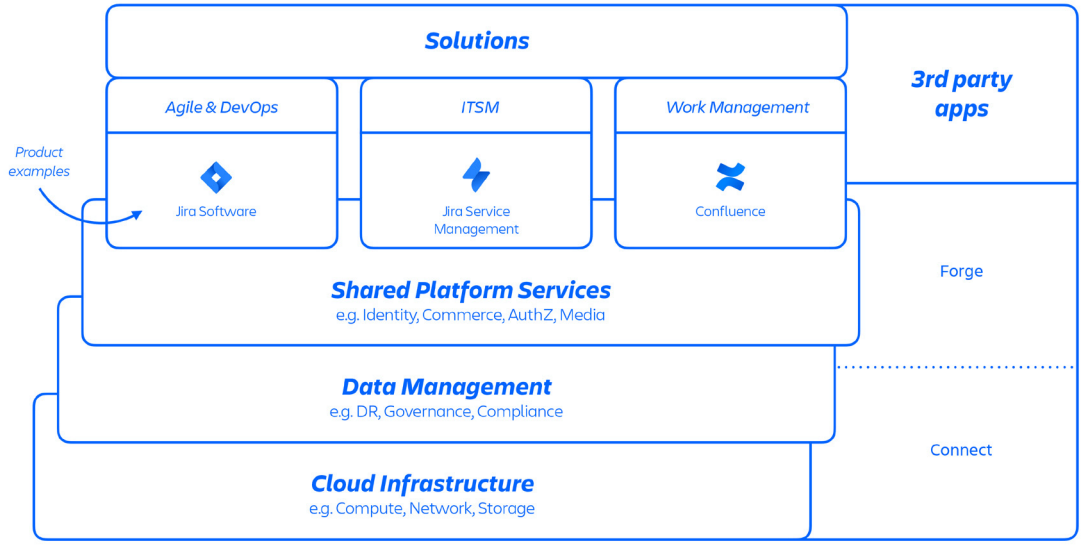



*Image 1 : architecture de la plateforme Atlassian.*

Les développeurs Atlassian fournissent ces services via une plateforme en tant que service (PaaS) développée en interne, appelée Micros, qui orchestre automatiquement le déploiement des services partagés, de l'infrastructure, des ensembles de données et de leurs capacités de gestion, y compris les exigences en matière de contrôle de la sécurité et de la conformité (voir *image 1* ci-dessus). En général, un produit Atlassian se compose de plusieurs services « conteneurisés » déployés sur AWS à l'aide de Micros. Les produits Atlassian utilisent les fonctionnalités de base de la plateforme (voir *image 2* ci-dessous), qui vont du routage des requêtes aux ensembles d'objets binaires, en passant par l'authentification/autorisation, le contenu transactionnel généré par l'utilisateur (UGC) et les ensembles de relations entre entités, les data lakes, la journalisation en commun, les requêtes de services de traçage, l'observabilité et l'analyse. Ces microservices sont établis à l'aide de piles techniques approuvées et normalisées au niveau de la plateforme :

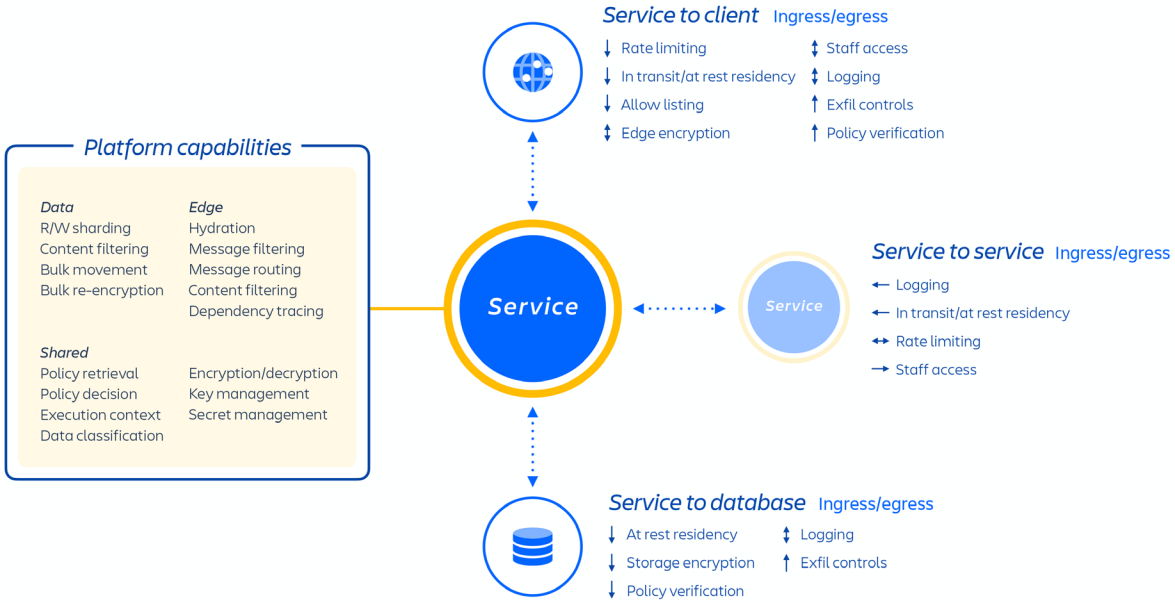



*Image 2 : vue d'ensemble des microservices Atlassian.*

## Architecture multilocataire

En plus de notre infrastructure cloud, nous avons créé et nous exploitons une architecture de microservices multi-locataires ainsi qu'une plateforme partagée qui prend en charge nos produits. Dans une architecture multi-locataires, un service unique dessert plusieurs clients, y compris les bases de données et les instances de calcul nécessaires à l'exécution de nos produits cloud. Chaque partition (dans l'essence, un conteneur : voir *image 3* ci-dessous) contient les données de plusieurs locataires, mais les données de chacun d'entre eux sont isolées et inaccessibles aux autres locataires.

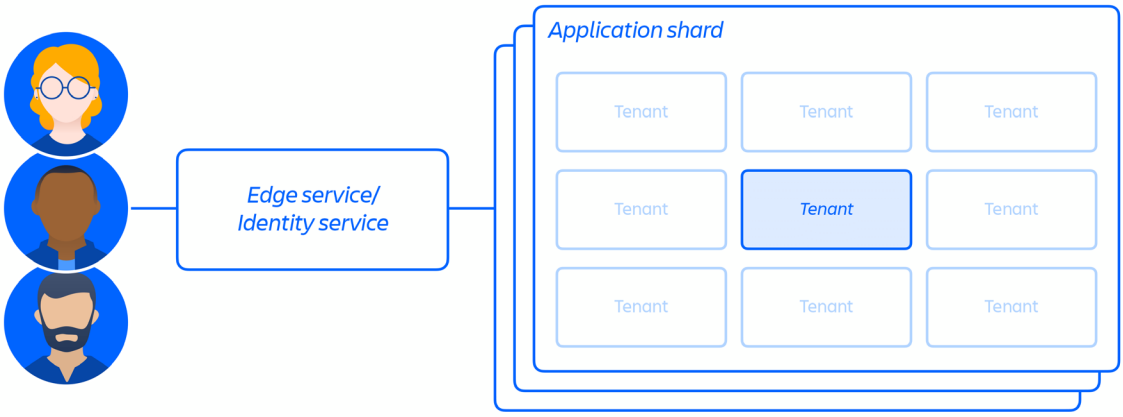



*Image 3 : notre manière de stocker les données dans une architecture mutualisée.*

## Provisionnement d'un locataire et cycle de vie

Lorsqu'un nouveau client est provisionné, une série d'événements déclenche l'orchestration des services distribués et le provisionnement des magasins de données. Ces événements peuvent généralement être mis en correspondance avec l'une des sept étapes du cycle de vie :

1. Les systèmes commerciaux sont immédiatement mis à jour avec les dernières métadonnées et informations de contrôle d'accès pour ce client, puis un système d'orchestration du provisionnement aligne l'« état des ressources provisionnées » avec l'état de la licence par le biais d'une série d'événements liés au locataire et aux produits.

**Événements liés au locataire**
Ces événements affectent le locataire dans son ensemble et peuvent prendre deux formes :

- Création : un locataire est créé et utilisé pour de nouveaux sites
- Destruction : un locataire entier est supprimé

**Événements liés aux produits**

- Activation : après l'activation de produits sous licence ou d'applications tierces
- Désactivation : après la désactivation de certains produits ou de certaines applications
- Suspension : après la suspension d'un produit existant donné, désactivant ainsi l'accès à un site donné dont il est propriétaire
- Dé-suspension : après la dé-suspension d'un produit existant donné, permettant ainsi l'accès à un site dont il est propriétaire

Mise à jour de licence : contient des informations concernant le nombre de postes de licence pour un produit donné ainsi que son statut (actif/inactif)

2. Création du site client et activation du bon ensemble de produits pour le client. Le concept d'un site est le conteneur de plusieurs produits concédés sous licence à un client spécifique (par exemple, Confluence et Jira Software pour `<nom du site>.atlassian.net`). Ceci (voir *image 4* ci-dessous) est un point qu'il est important de comprendre dans le contexte de ce rapport, car le conteneur de site représente ce qui a été supprimé lors de cet incident, et le concept de site est discuté tout au long de ce document.

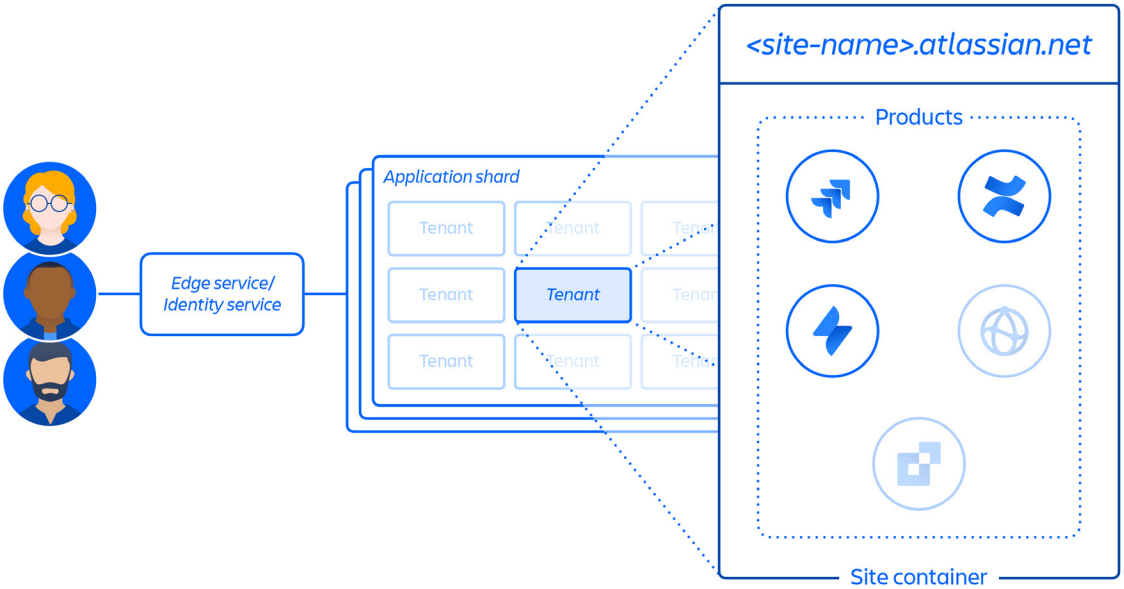



*Image 4 : vue d'ensemble du conteneur de site.*

3. Provisionnement de produits sur le site client dans la région désignée.

   Lorsqu'un produit est provisionné, la plus grande partie de son contenu est hébergée à proximité de l'endroit où les utilisateurs y accèdent. Pour optimiser les performances des produits, nous ne limitons pas le mouvement des données lorsqu'elles sont hébergées dans le monde entier et nous pouvons déplacer des données entre les régions selon les besoins.

   Pour certains de nos produits, nous offrons également la résidence des données. La résidence des données permet aux clients de choisir si les données des produits sont distribuées dans le monde entier ou conservées dans l'une de nos zones géographiques définies.

4. Création et stockage de la configuration et des métadonnées de base du site client et du ou des produits.

5 Création et stockage des données d'identité du site et du ou des produits, telles que les utilisateurs, les groupes, les autorisations, etc.

6 Provisionnement de bases de données de produits sur un site, par ex. famille de produits Jira, Confluence, Compass, Atlas.

7 Provisionnement des applications sous licence du ou des produits.

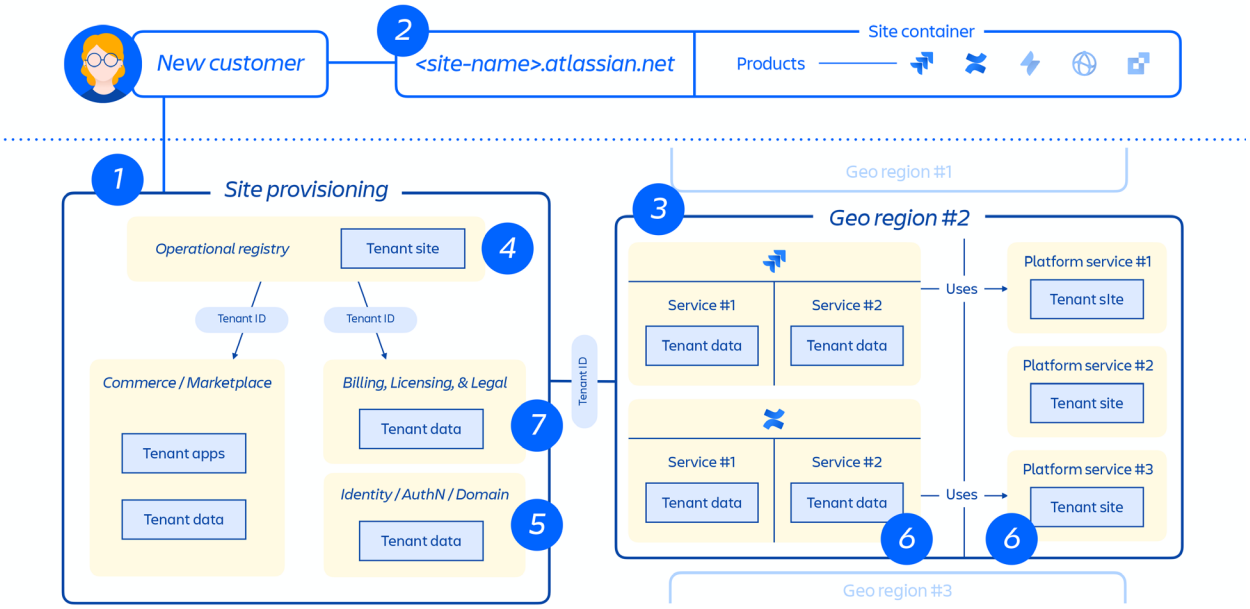



*Image 5 : vue d'ensemble de la façon dont le site client est provisionné sur l'ensemble de notre architecture distribuée.*

*L'image 5* ci-dessus montre comment le site d'un client est déployé dans notre architecture distribuée, et pas uniquement dans un ensemble ou base de données unique. Cela inclut plusieurs emplacements physiques et logiques qui stockent des métadonnées, des données de configuration, des données de produit, des données de plateforme et d'autres informations de site connexes.

## Programme de reprise d'activité

Notre programme de reprise d'activité (DR) englobe toutes les actions que nous menons pour assurer la résilience contre les pannes de l'infrastructure et la restaurabilité du stockage des services à partir des sauvegardes. Les deux concepts importants pour comprendre les programmes de reprise d'activité sont les suivants :

- **Objectif de temps de récupération (RTO) :** À quelle vitesse les données peuvent-elles être récupérées et restituées à un client en cas de sinistre ?
- **Objectif de point de récupération (RPO) :** Quelle est la fraîcheur des données récupérées après leur récupération à partir d'une sauvegarde ? Quelle quantité de données sera perdue depuis la dernière sauvegarde ?

Au cours de cet incident, nous avons dépassé notre RTO, mais avons respecté notre RPO.

## Résilience

Nous nous préparons aux défaillances au niveau de l'infrastructure, par exemple la perte totale d'une base de données, d'un service ou de zones de disponibilité AWS. Cette préparation comprend la réplication des données et des services sur plusieurs zones de disponibilité et des tests réguliers de basculement.

## Restauration du stockage de service

Nous nous préparons également à récupérer en cas de corruption des données du stockage de services due à des risques tels que les rançongiciels, les mauvais acteurs, les défauts logiciels et les erreurs opérationnelles. Cette préparation comprend des sauvegardes immuables et des tests de restauration des sauvegardes de stockage de services. Nous avons la possibilité de prendre n'importe quel magasin de données individuel et de le restaurer à un point antérieur dans le temps.

## Restauration automatisée multi-site et multi-produits

Au moment de l'incident, nous n'étions pas en mesure de sélectionner un nombre important de sites clients et de restaurer tous leurs produits interconnectés à partir de sauvegardes à un point antérieur dans le temps.

Nos capacités se sont concentrées sur l'infrastructure, la corruption des données, les événements de service uniques ou les suppressions de sites uniques. Dans le passé, nous avions dû faire face à ce type de défaillances et les tester. La suppression au niveau des sites ne disposait pas de runbooks pouvant être rapidement automatisés pour un événement de cette ampleur qui nécessitait des outils et une automatisation sur tous les produits et services pour se dérouler de manière coordonnée.

Les sections suivantes vont approfondir ces nuances et présenter ce que nous faisons chez Atlassian pour faire évoluer et optimiser notre capacité à maintenir cette architecture à grande échelle.

# Événements, chronologie et récupération

## Que s'est-il passé ?

En 2021, nous avons procédé à l’acquisition et à l’intégration d’une application Atlassian autonome pour Jira Service Management et Jira Software appelée Insight – Asset Management. Cette application autonome a alors été intégrée en tant que fonctionnalité native dans Jira Service Management et n’était plus disponible pour Jira Software. De ce fait, nous avons dû supprimer l’ancienne application autonome sur les sites des clients qui l’avaient installée. Nos équipes d’ingénieurs ont utilisé un script et un processus existants pour supprimer les instances de cette application autonome.

Cependant, deux problèmes critiques se sont posés :

- **Communication insuffisante.** Tout d'abord, la communication entre l'équipe qui a demandé la suppression et celle qui l'a exécutée était insuffisante. Au lieu de fournir les identifiants de l'app prévue pour suppression, l'équipe a fourni les identifiants de l'ensemble du site cloud sur lequel les apps devaient être supprimées.
- **Avertissements système insuffisants.** L'API utilisée pour effectuer la suppression accepte les identifiants de site et d'application et suppose que la saisie est correcte. Cela signifie que si un identifiant de site est transmis, un site est supprimé ; si un identifiant d'application est transmis, une application est supprimée. Aucun signal d'avertissement n'a été émis pour confirmer le type de suppression (site ou application) demandé.

Le script qui a été exécuté utilisait notre processus standard de revue par des pairs, qui se concentrait sur le point de terminaison appelé et de quelle manière il l’était. Il ne vérifiait pas si les identifiants des sites cloud fournis faisaient référence à Insight App ou au site complet. Le script a été testé en staging (environnement de test pré-production) conformément à nos processus standard de gestion des changements, mais il n’aurait pas détecté que les identifiants saisis étaient incorrects, car ces identifiants n’existaient pas dans l’environnement de staging.

Une fois exécuté en production, le script a été exécuté initialement sur 30 sites.
Le premier run du script en production a été correctement exécuté et a supprimé l’application Insight pour ces 30 sites sans autres effets secondaires. Cependant, les identifiants de ces 30 sites avaient été obtenus avant l’incident et comprenaient les identifiants corrects de l’application Insight.

Le script pour l'exécution de production suivante incluait des identifiants de site à la place des identifiants d'application Insight et s'exécutait sur un ensemble de 883 sites. Le script a commencé à s'exécuter le 5 avril à 07 h 38 UTC et s'est terminé à 8 h 01 UTC. Le script a supprimé séquentiellement des sites en fonction de la liste d'entrée, ce qui fait que le site du premier client a été supprimé peu de temps après le début de l'exécution du script à 07 h 38 UTC. Le résultat a été une suppression immédiate des 883 sites, sans aucun signal d'avertissement pour nos équipes d'ingénieurs.

Les produits Atlassian suivants n'étaient pas disponibles pour les clients concernés : la famille de produits Jira, Confluence, Atlassian Access, Opsgenie et Statuspage.

Dès que nous avons eu connaissance de l'incident, nos équipes se sont concentrées sur la restauration pour tous les clients impactés. À ce moment-là, nous avons estimé qu'environ 700 sites avaient été impactés (883 sites au total ont été impactés, mais nous avons retiré les sites appartenant à Atlassian). Un grand nombre de ces 700 sites comprenait des comptes inactifs, gratuits ou de petite taille avec un faible nombre d'utilisateurs actifs. À partir de ce constat, nous avons initialement estimé que 400 clients environ étaient impactés.

Nous disposons maintenant d'une vue beaucoup plus précise. Par souci de transparence totale, sur la base de la définition officielle des clients d'Atlassian, 775 clients ont été touchés par la panne, mais la majorité des utilisateurs étaient représentés dans l'estimation initiale de 400 clients. La panne a duré jusqu'à 14 jours pour un sous-ensemble de ces clients : le premier ensemble de clients a été restauré le 8 avril, et tous les clients ont été restaurés à compter du 18 avril.

## Comment nous nous sommes coordonnés

Le premier ticket d'assistance a été créé par un client concerné à 7 h 46 UTC, le 5 avril. Notre surveillance interne n'a pas détecté de problème car les sites ont été supprimés via un flux de travail standard. À 8 h 17 UTC, nous avons lancé notre processus de gestion des incidents majeurs, pour former une équipe interfonctionnelle de gestion des incidents, et en sept minutes, à 8 h 24 UTC, le statut de l'incident a été revu à la hausse, sur Critique. À 8 h 53 UTC, notre équipe a confirmé que le ticket d'assistance client et l'exécution du script étaient liés. Une fois que nous avons pris conscience de la complexité de la restauration, nous avons attribué notre plus haut niveau de gravité à l'incident à 12 h 38 UTC.

L'équipe de gestion des incidents était composée de personnes issues de plusieurs équipes d'Atlassian, notamment l'ingénierie, le support client, la gestion des programmes, les communications, et bien d'autres encore. L'équipe mondiale de réponse

aux incidents a travaillé 24 h/24 et 7 j/7 pendant toute la durée de l'incident jusqu'à ce que tous les sites touchés aient été restaurés, validés et restitués aux clients.

Pour gérer l'avancement de la restauration, nous avons créé un nouveau projet Jira, SITE, et un flux de travail pour suivre les restaurations site par site au sein de plusieurs équipes (ingénierie, gestion de programme, support, etc.). Cette approche a permis à toutes les équipes d'identifier et de suivre facilement les problèmes liés à la restauration de chaque site.

Nous avons également mis en place un gel du code sur l'ensemble de l'ingénierie pour la durée de l'incident le 8 avril à 03:30 UTC. Cela nous a permis de nous concentrer sur la restauration des clients, d'éliminer le risque de changement entraînant des incohérences dans les données clients, de minimiser le risque d'autres pannes et de réduire la probabilité que des changements sans rapport avec le problème ne viennent distraire l'attention de l'équipe occupée à la récupération.

## Chronologie de l'incident

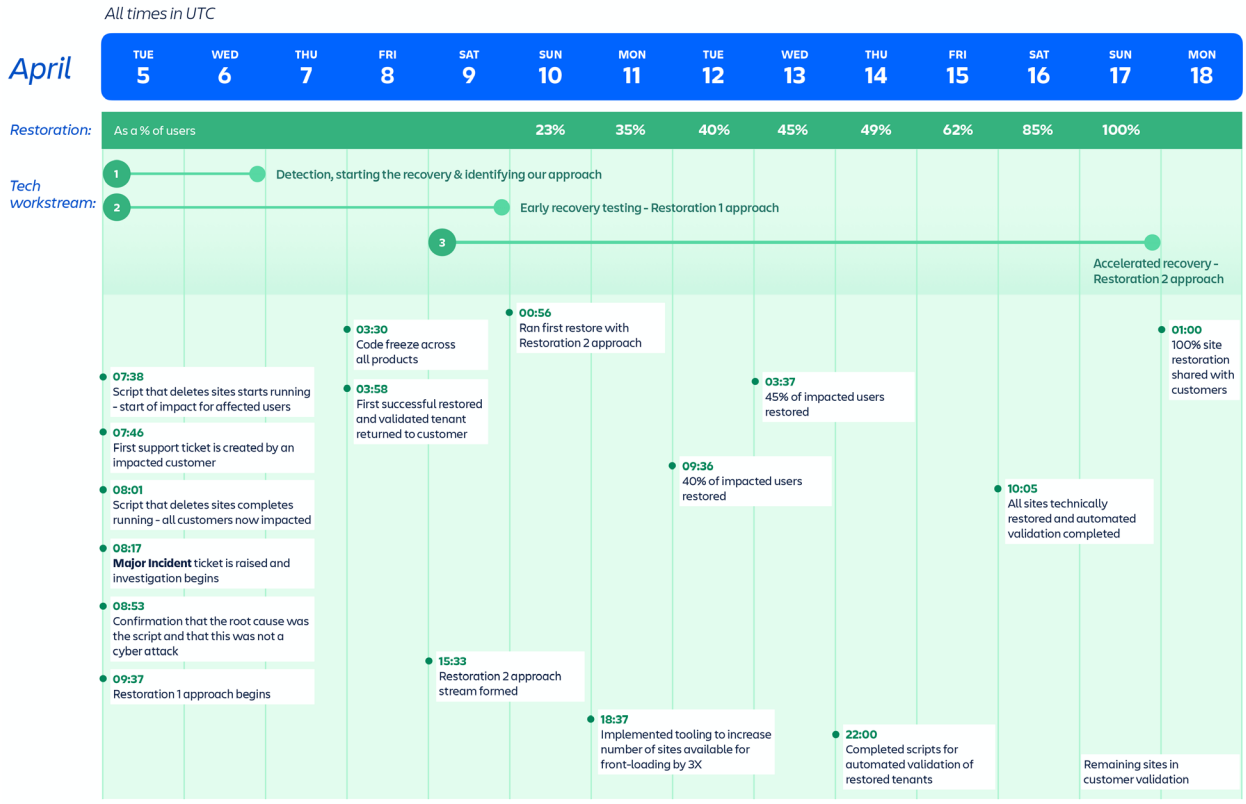



*Image 6 : chronologie de l'incident et des principales étapes de restauration.*

# Aperçu de haut niveau des flux de travail de récupération

La récupération s'est déroulée sous la forme de trois flux de travail principaux : détection, récupération anticipée et accélération. Bien que nous ayons décrit chaque domaine de travail séparément ci-dessous, pendant la récupération, les travaux se sont déroulés en parallèle dans tous les flux de travail.

## Flux de travail 1 : détection, démarrage de la récupération et identification de l'approche à suivre

***Dates : jours 1 à 2 (5 au 9 avril)***

À 8 h 53 UTC le 5 avril, nous avons constaté que le script de l'application Insight provoquait la suppression de sites. Nous avons confirmé que cela n'était pas dû à un acte malveillant interne ou à une cyberattaque. Les équipes chargées des produits et de l'infrastructure de la plateforme ont été appelées et mobilisées sur l'incident.

Au début de l'incident, nous avons fait le constat suivant :

- La restauration de centaines de sites supprimés est un processus complexe en plusieurs étapes (détaillé dans la section sur l'architecture ci-dessus), qui nécessite de nombreuses équipes et plusieurs jours pour être menée à bien.
- Nous avions la capacité nécessaire pour récupérer un site unique, mais nous n'avions pas mis en place de capacités et de processus permettant de récupérer un grand nombre de sites.

Par conséquent, nous avons dû considérablement paralléliser et automatiser le processus de restauration afin d'aider les clients concernés à retrouver l'accès à leurs produits Atlassian le plus rapidement possible.

Le flux de travail 1 a impliqué un grand nombre d'équipes de développement chargées de mener les activités suivantes :

- Identifier et exécuter les étapes de restauration des lots de sites dans le pipeline.
- Établir et améliorer l'automatisation pour permettre aux équipes de suivre les étapes de restauration pour un plus grand nombre de sites par lot.

## Flux de travail 2 : premières récupérations et approche Restauration 1

***Dates : jours 1 à 4 (5 au 9 avril)***

Nous avons repéré la cause de la suppression des sites le 5 avril à 8 h 53 UTC, moins d'une heure après la fin de l'exécution du script. Nous avons également identifié le processus de restauration qui avait déjà permis de récupérer quelques sites. Cependant,

le processus de récupération pour restaurer les sites supprimés à une telle échelle n'était pas bien défini.

Pour progresser rapidement, deux groupes de travail se sont formés pour gérer les premières étapes de l'incident :

- Le groupe de travail manuel a validé les étapes requises et a exécuté manuellement le processus de restauration pour un petit nombre de sites.
- Le groupe de travail automatisé a repris le processus de restauration existant et l'a automatisé pour en exécuter les étapes en toute sécurité sur un plus grand nombre de sites.

Aperçu de l'approche Restauration 1 (voir *image 7* ci-dessous) :

- Cette approche nécessitait de créer un nouveau site pour chaque site supprimé, puis de recréer chaque produit, service et ensemble de données en aval dont les données devaient être restaurées.
- Le nouveau site serait alors doté de nouveaux identifiants tels que CloudID. Ces identifiants sont tous considérés comme immuables, ce qui signifie que de nombreux systèmes intègrent ces identifiants dans les enregistrements de données. Par conséquent, toute modification de ces identifiants exigeait de mettre à jour de grandes quantités de données, ce qui est particulièrement problématique pour les applications de l'écosystème tiers.
- La modification d'un nouveau site pour répliquer le statut du site supprimé entraînait des dépendances complexes et souvent imprévues entre chaque étape.

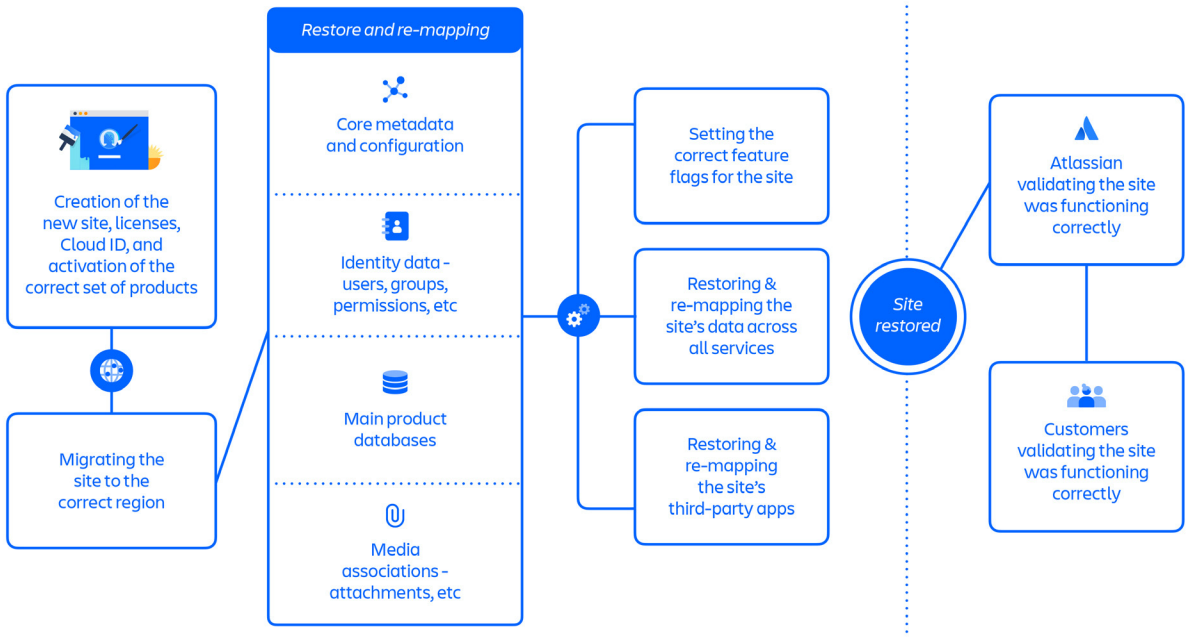



*Image 7 : étapes clés de l'approche Restauration 1.*

L'approche Restauration 1 comprenait environ 70 étapes individuelles qui, une fois regroupées à un niveau élevé, suivaient un flux principalement séquentiel des étapes suivantes :

- Création des nouveaux sites, licences et Cloud ID et activation de l'ensemble de produits approprié
- Migration du site vers la bonne région
- Restauration et remappage des principales métadonnées et de la configuration du site
- Restauration et remappage des données d'identité du site (utilisateurs, groupes, autorisations, etc.)
- Restauration des principales bases de données de produits du site
- Restauration et remappage des associations médias du site (pièces jointes, etc.)
- Configuration des indicateurs de fonctionnalité appropriés pour le site
- Restauration et remappage des données du site pour tous les services
- Restauration et remappage des applications tierces du site
- Validation par Atlassian du bon fonctionnement du site
- Validation par le client du bon fonctionnement du site

Une fois optimisée, l'approche Restauration 1 a pris environ 48 heures pour restaurer un lot de sites, et a été utilisée pour la récupération de 53 % des utilisateurs concernés à travers 112 sites entre le 5 et le 14 avril.

## Flux de travail 3 : Récupération accélérée et approche Restauration 2

***Dates : jours 4 à 13 (9 au 17 avril)***
Avec l'approche Restauration 1, il nous aurait fallu trois semaines pour restaurer les données de tous les clients. C'est pourquoi nous avons proposé une nouvelle approche le 9 avril, pour accélérer la restauration de tous les sites : Restauration 2 (voir *image 8* ci-dessous).

L'approche Restauration 2 a amélioré le parallélisme entre les étapes de restauration en réduisant la complexité et le nombre de dépendances dont souffrait l'approche Restauration 1.

Restauration 2 impliquait de recréer (ou d'annuler la suppression) des enregistrements associés aux sites dans tous les systèmes respectifs, en commençant par l'enregistrement de service de catalogue. Un élément clé de cette nouvelle approche était de *pouvoir conserver tous les anciens identifiants de site*. Cela a supprimé la moitié des étapes du processus précédent, qui visaient à mapper les anciens identifiants avec les nouveaux, y compris la nécessité de se coordonner avec chaque fournisseur d'applications tiers pour chaque site.

Cependant, le passage de l'approche Restauration 1 à l'approche Restauration 2 a ajouté des frais généraux importants à la réponse à l'incident :

- De nombreux scripts et processus d'automatisation établis selon l'approche Restauration 1 ont dû être modifiés pour Restauration 2.
- Les équipes effectuant des restaurations (y compris les coordinateurs d'incidents) ont dû gérer des lots parallèles de restaurations selon les deux approches, pendant que nous testions et validions le processus Restauration 2.
- L'utilisation d'une nouvelle approche signifiait que nous devions tester et valider le processus Restauration 2 avant de le mettre à l'échelle, ce qui a nécessité de dupliquer le travail de validation précédemment effectué sur Restauration 1.

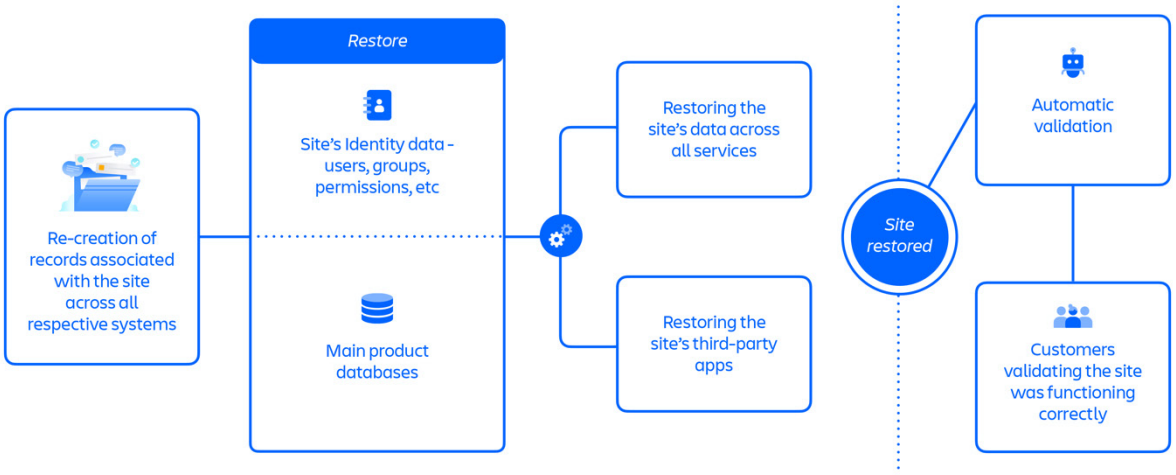



*Image 8 : étapes clés de l'approche Restauration 2.*

Le graphique ci-dessus représente l'approche Restauration 2, qui comprenait plus de 30 étapes qui suivaient un flux principalement parallélisé de :

- Recréation des enregistrements associés au site dans tous les systèmes respectifs
- Restauration des données d'identité du site (utilisateurs, groupes, autorisations, etc.)
- Restauration des principales bases de données de produits du site
- Restauration des données du site pour tous les services
- Restauration des applications tierces du site
- Validation automatique
- Validation par le client du bon fonctionnement du site

Dans le cadre de la restauration accélérée, nous avons également pris des mesures pour charger à l'avance et automatiser la restauration des sites, car le processus manuel ne pouvait pas être adapté aux grands lots. La nature séquentielle du processus de récupération signifiait que la restauration du site pouvait être plus lente pour celles des

bases de données et des bases utilisateur/autorisations volumineuses. Voici les optimisations que nous avons mises en œuvre :

- Nous avons développé les outils et les protections nécessaires pour les étapes de longue durée et *intenses en début de période*, telles que la restauration de bases de données et la synchronisation d'identité, afin de les terminer avant les autres étapes de restauration.
- Les équipes d'ingénierie ont mis en place une automatisation pour les étapes séparées, ce qui a permis d'exécuter de grands lots de restaurations en toute sécurité.
- L'automatisation a été conçue pour vérifier que les sites fonctionnaient correctement une fois toutes les étapes de restauration terminées.

L'approche Restauration 2 accélérée a pris environ 12 heures pour restaurer un site et a permis de récupérer environ 47 % des utilisateurs concernés sur 771 sites entre le 14 et le 17 avril.

## Perte de données minimale suite à la restauration des sites supprimés

Nos bases de données sont sauvegardées à l'aide d'une combinaison de sauvegardes complètes et de sauvegardes incrémentielles qui nous permettent de choisir un « point dans le temps » spécifique selon lequel restaurer nos ensembles de données pendant la période de conservation des sauvegardes (30 jours). Pour la plupart des clients lors de cet incident, nous avons identifié les principaux ensembles de données pour nos produits et avons décidé d'utiliser un point de restauration de cinq minutes avant la suppression des sites comme point de synchronisation sécurisé. Les ensembles de données secondaires ont été restaurés au même point ou en rejouant les événements enregistrés. L'utilisation d'un point de restauration fixe pour les ensembles principaux nous a permis d'obtenir une cohérence des données dans tous les ensembles de données.

Pour 57 clients restaurés au début de notre réponse aux incidents, l'absence de politiques cohérentes et la récupération manuelle des sauvegardes de base de données ont entraîné la restauration de certaines bases de données Confluence et Insight à leur statut *plus* de cinq minutes avant la suppression du site. Cette incohérence a été découverte lors d'un processus d'audit post-restauration. Depuis, nous avons récupéré le reste des données, contacté les clients concernés et nous les aidons à appliquer les modifications afin de restaurer davantage leurs données.

En résumé :

- Nous avons atteint notre objectif de perte de données maximale admissible (PDMA) d'une heure au cours de cet incident.

- La perte de données résultant de l'incident est limitée à cinq minutes avant la suppression du site.
- Un petit nombre de clients ont vu leurs bases de données Confluence ou Insight restaurées à un statut antérieur à la suppression du site de plus de cinq minutes. Cependant, nous sommes en mesure de récupérer les données et nous travaillons actuellement avec les clients pour restaurer ces données.

# Communications sur les incidents

Lorsque nous parlons de communications sur les incidents, cela englobe les points de contact avec les clients, les partenaires, les médias, les analystes du secteur, les investisseurs et la communauté technologique au sens large.

## Que s'est-il passé ?

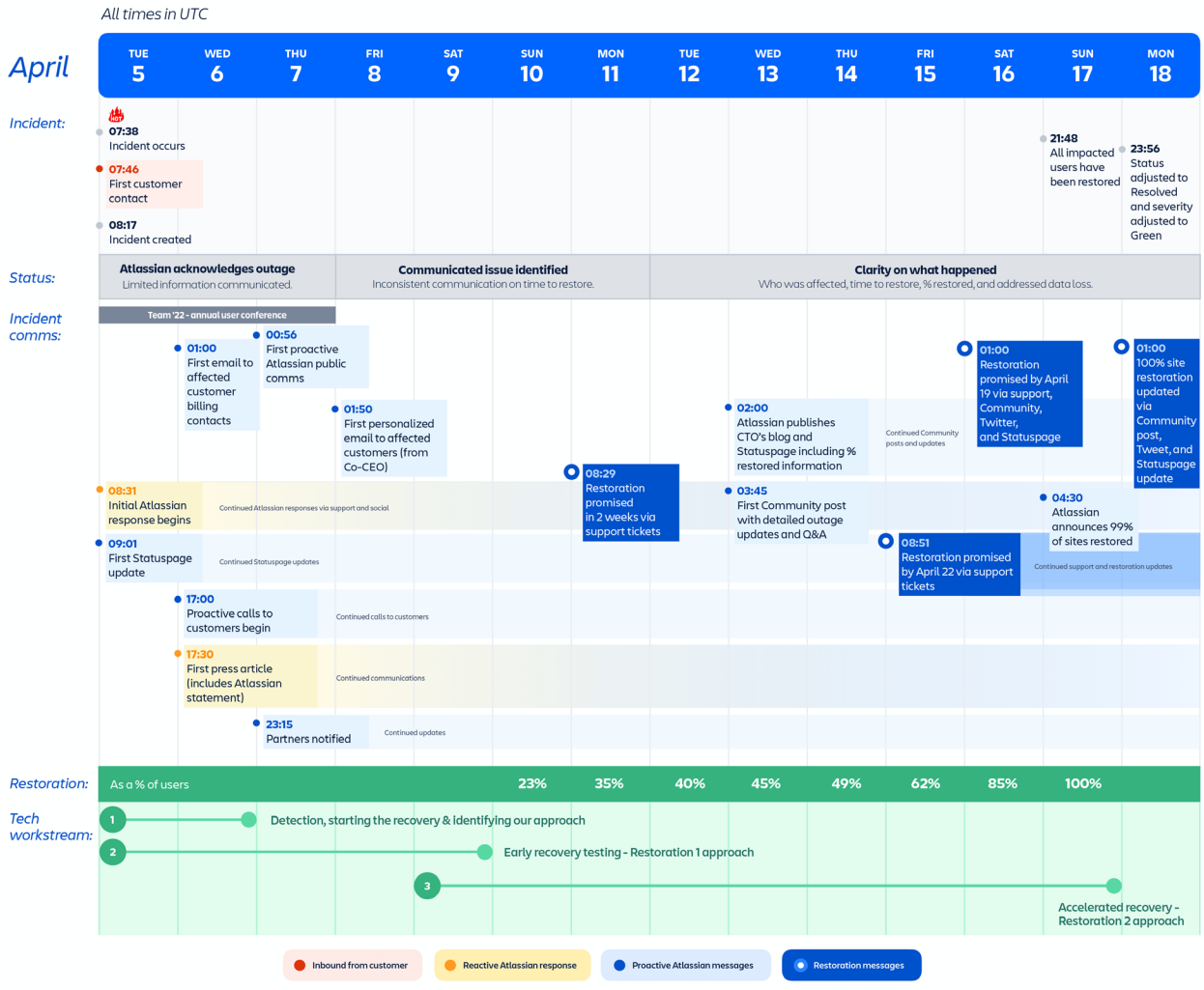



*Image 9 : chronologie des principales étapes de la communication des incidents.*

***Dates : Jours 1 à 3 (5 au 7 avril)***

## Première réponse

Le premier ticket d'assistance a été créé le 5 avril à 7 h 46 UTC et l'assistance Atlassian a répondu en accusant réception de l'incident à 8 h 31 UTC. À 9 h 03 UTC, la première mise à jour de Statuspage a été publiée pour informer les clients que nous enquêtions sur l'incident. À 11 h 13 UTC, nous avons confirmé via Statuspage que nous avions identifié la cause première et que nous travaillions sur un correctif. À 1 h UTC le 6 avril, les premières communications des tickets client indiquaient que la panne était due à un script de maintenance et que nous prévoyions une perte de données minimale. Atlassian a répondu aux demandes des médias par une déclaration le 6 avril à 17 h 30 UTC. Atlassian a tweeté son premier message externe général reconnaissant l'incident le 7 avril à 0 h 56 UTC.

***Dates : jours 4 à 7 (8 au 11 avril)***

## Début d'une sensibilisation plus large et personnalisée

Le 8 avril à 1 h 50 UTC, Atlassian a envoyé aux clients concernés des excuses de la part de son co-fondateur et co-PDG, Scott Farquhar. Dans les jours qui ont suivi, nous nous sommes efforcés de restaurer les informations de contact supprimées et de créer des tickets d'assistance pour tous les sites concernés qui n'en avaient pas encore déposé un. Notre équipe d'assistance a ensuite continué à envoyer régulièrement des mises à jour concernant nos efforts de restauration via les tickets d'assistance correspondant à chaque site concerné.

***Dates : Jours 8 à 14 (12 au 18 avril)***

## Plus de clarté et restauration complète

Le 12 avril, Atlassian a publié une mise à jour de la part du directeur technique, Sri Viswanath, afin de fournir davantage de détails techniques sur ce qui s'est passé, qui a été affecté, s'il y a eu perte de données, nos progrès en matière de restauration et le fait que jusqu'à deux semaines allaient être nécessaires pour restaurer complètement tous les sites. L'article était accompagné d'un autre communiqué de presse attribué à Sri. Nous avons également fait référence à l'article de Sri dans notre premier billet proactif de la communauté Atlassian, rédigé par Stephen Deasy, responsable de l'ingénierie. Ce billet est ensuite devenu le lieu dédié aux mises à jour supplémentaires et aux discussions avec le grand public. Une mise à jour du 18 avril de cet article a annoncé la restauration complète de tous les sites clients concernés.

**Pourquoi n'avons-nous pas répondu publiquement plus tôt ?**

1. Nous avons privilégié la communication directe avec les clients concernés via Statuspage, e-mail, tickets d'assistance et interactions individuelles. Cependant, nous n'avons pas pu joindre de nombreux clients car nous avons perdu leurs coordonnées lorsque leurs sites ont été supprimés. Nous aurions dû produire des communications plus globales beaucoup plus tôt, afin d'informer les clients et les utilisateurs finaux concernés de notre réponse aux incidents et de notre calendrier de résolution.

2. Bien que nous ayons immédiatement compris la cause de l'incident, la complexité architecturale et les circonstances uniques de cet incident ont ralenti notre capacité à déterminer rapidement la portée et à estimer avec précision le délai de résolution. Plutôt que d'attendre d'avoir une image complète, nous aurions dû faire preuve de transparence sur ce que nous savions et ce que nous ne savions pas. Fournir des estimations générales quant à la restauration des données (même si elles étaient directionnelles) et avoir une idée précise du moment où nous prévoyions d'avoir une image plus complète aurait permis à nos clients de mieux s'adapter à l'incident. Cela est particulièrement vrai pour les administrateurs système et les contacts techniques, qui sont en première ligne dans la gestion des parties prenantes et des utilisateurs au sein de leur organisation.

## Expérience d'assistance et sensibilisation des clients

Comme mentionné précédemment, le script qui a supprimé les sites clients a également supprimé les principaux identifiants clients et les informations de contact (par ex. URL du cloud, contacts de l'administrateur système du site) de nos environnements de production. Il est important de noter ceci car nos systèmes principaux (par ex. les services d'assistance, de licences et de facturation) utilisent tous l'URL du cloud et les contacts de l'administrateur système du site comme identifiants principaux, à des fins de sécurité, de routage et de priorisation. Lorsque nous avons perdu ces identifiants, nous avons perdu notre capacité à identifier systématiquement les clients et à interagir avec eux.

## Quel a été l'impact de l'assistance pour nos clients ?

Tout d'abord, la majorité des clients concernés n'ont pas pu joindre notre équipe d'assistance via le formulaire de contact en ligne habituel. Ce formulaire est conçu pour obliger l'utilisateur à se connecter avec son ID Atlassian et à fournir une URL de cloud valide. Sans URL valide, l'utilisateur ne peut pas envoyer de ticket d'assistance technique. Dans le cours normal des activités, cette vérification est intentionnelle, pour veiller à la sécurité du site et au tri des tickets. Cependant, cette exigence a eu un résultat imprévu sur les clients touchés par cette panne : ils ne pouvaient pas soumettre de ticket d'assistance de site de haute priorité.

Deuxièmement, la suppression des données de l'administrateur système du site à la suite de l'incident a créé une lacune dans notre capacité à interagir de manière proactive avec les clients concernés. Au cours des premiers jours de l'incident, nous avons envoyé des communications proactives aux contacts techniques et de facturation des clients concernés enregistrés auprès d'Atlassian. Cependant, nous avons rapidement constaté que beaucoup de ces contacts n'étaient pas à jour. Sans les informations relatives à l'administrateur système pour chaque site, nous ne disposions pas d'une liste complète des contacts actifs et approuvés, par l'intermédiaire desquels nous pouvions interagir.

## Comment avons-nous réagi ?

Nos équipes d'assistance avaient trois priorités tout aussi importantes pour accélérer la restauration des sites et combler les ruptures de nos canaux de communication dans les premiers jours de l'incident.

**Tout d'abord, obtenir une liste fiable de contacts clients validés.** Alors que nos équipes d'ingénierie travaillaient à restaurer les sites des clients, nos équipes en contact avec ces derniers se sont concentrées sur la restauration des informations de contact validées. Nous avons utilisé tous les moyens à notre disposition (systèmes de facturation, tickets d'assistance antérieurs, autres sauvegardes sécurisées des utilisateurs, contact direct avec les clients, etc.) pour reconstituer notre liste de contacts. Notre objectif était d'avoir un ticket d'assistance lié à l'incident pour chaque site touché, afin d'optimiser les contacts directs et les délais de réponse.

**Ensuite, rétablir les flux de travail, les files d'attente et les SLA spécifiques à cet incident.** La suppression de l'ID du cloud et l'impossibilité d'authentifier correctement les utilisateurs ont également eu un impact sur notre capacité à traiter les tickets d'assistance liés aux incidents par le biais de nos systèmes habituels. Les tickets n'apparaissaient pas correctement dans les files d'attente et les tableaux de bord de priorité et de remontées adéquats. Nous avons rapidement créé une équipe interfonctionnelle (assistance, produit, informatique) chargée de concevoir et ajouter une

logique, des SLA, des états de flux de travail et des tableaux de bord supplémentaires. Comme cela devait être fait au sein de notre système de production, il a fallu plusieurs jours pour tout développer, tester et déployer.

**Troisièmement, la mise à l'échelle massive des validations manuelles afin d'accélérer la restauration des sites.** Au fur et à mesure que nos ingénieurs progressaient dans les premières restaurations, il est devenu évident que nos équipes d'assistance mondiales seraient nécessaires pour accélérer la restauration des sites via des tests manuels et des contrôles de validation. Ce processus de validation est devenu un moyen essentiel de restaurer les sites de nos clients, dès lors que notre équipe d'ingénierie a accéléré la restauration des données. Nous avons dû créer un flux indépendant de procédures opérationnelles permanentes (POP), de flux de travail, de transferts et de listes de personnel afin de mobiliser plus de 450 ingénieurs d'assistance pour exécuter des contrôles de validation, avec des équipes offrant une couverture 24 heures sur 24, 7 jours sur 7, afin de pouvoir rendre aux clients leurs données au plus vite.

Même si ces priorités clés étaient bien établies à la fin de la première semaine, nous étions limités dans notre capacité à fournir des mises à jour *significatives*, en raison du manque de clarté dans les délais de résolution des incidents, du fait de la complexité des processus de restauration. Nous aurions dû reconnaître plus tôt notre incapacité à fournir une date de restauration des sites et nous aurions dû nous rendre disponibles plus tôt pour des discussions en personne, afin que nos clients puissent aviser en conséquence.

## Comment pouvons-nous nous améliorer ?

Nous avons immédiatement bloqué les suppressions en lot des sites en attendant que les modifications appropriées puissent être apportées.

En allant de l'avant après cet incident, en réévaluant nos processus internes, nous voulons déclarer que les individus ne sont pas à l'origine des incidents. À la place, ce sont les systèmes qui rendent possibles les erreurs. Cette section résume les facteurs qui ont contribué à cet incident. Nous discutons également de nos plans pour accélérer la façon dont nous allons corriger ces faiblesses et ces problèmes.

### Enseignement 1 : les « suppressions en douceur » devraient être universelles à tous les systèmes

Dans l'ensemble, toute suppression du type concerné par l'incident doit être interdite ou comporter plusieurs niveaux de protection pour éviter les erreurs. La principale amélioration que nous apportons est d'empêcher partout la suppression des données clients et des métadonnées qui n'ont pas fait l'objet d'un processus de suppression en douceur.

**a) La suppression des données ne doit avoir lieu que dans le cadre d'une suppression en douceur**

La suppression d'un site entier devrait être interdite ; et la suppression en douceur devrait nécessiter des protections à plusieurs niveaux pour éviter les erreurs. Nous allons mettre en œuvre une politique de « suppression en douceur », empêchant les scripts ou systèmes externes de supprimer les données des clients dans un environnement de production. Notre politique de « suppression en douceur » permettra une conservation suffisante des données afin que la récupération des données puisse être exécutée rapidement et en toute sécurité. Les données ne seront supprimées de l'environnement de production qu'après l'expiration d'une période de conservation.

**Actions :**

- **Implémenter une « suppression en douceur » dans les flux de travail de provisionnement et tous les ensembles de données pertinents :** en outre, l'équipe de plateforme locataire vérifiera que les suppressions de données ne peuvent avoir lieu qu'après des désactivations, ainsi que d'autres mesures de protection dans cet espace. À plus long terme, l'équipe de plateforme locataire jouera un rôle de premier plan dans le développement de la bonne gestion de l'état des données des locataires.

**b) La suppression en douceur doit bénéficier d'un processus de révision normalisé et vérifié**

Les actions de suppression en douceur sont des opérations à haut risque. Par conséquent, nous devons disposer de processus de révision standardisés ou automatisés qui incluent des procédures de restauration et de test définies pour traiter ces opérations.

**Actions :**

- **Déploiement progressif imposé de toutes les actions de suppression en douceur :** toutes les nouvelles opérations nécessitant une suppression seront d'abord testées sur nos propres sites afin de valider notre approche et de vérifier l'automatisation. Une fois cette validation terminée, nous appliquerons progressivement le même processus aux clients et continuerons à tester les irrégularités avant d'appliquer l'automatisation à l'ensemble de la base d'utilisateurs sélectionnée.

- **Les actions de suppression en douceur doivent avoir un plan de restauration testé :** toute activité de suppression en douceur de données doit passer un test de restauration des données supprimées avant d'être exécutée en production et doit disposer d'un plan de restauration testé.

## Enseignement 2 : dans le cadre du programme de reprise d'activité, automatiser la restauration pour les événements de suppression multisites et multiproduits concernant un plus grand nombre de clients

La gestion des données Atlassian décrit en détail nos processus de gestion de données. Pour garantir la haute disponibilité, nous provisionnons et tenons à jour une réplique de secours synchrone dans plusieurs zones de disponibilité (ZD) AWS. Le basculement entre les différentes ZD est automatisé et prend généralement de 60 à 120 secondes. Par ailleurs, nous gérons régulièrement des pannes de data center et d'autres interruptions courantes sans impact sur le client.

Nous tenons également à jour des sauvegardes immuables pensées pour résister aux corruptions de données, ce qui permet une récupération à un point antérieur. Les sauvegardes sont conservées pendant 30 jours, et Atlassian teste et audite en permanence les sauvegardes de stockage à des fins de restauration. Si nécessaire, nous pouvons restaurer les sites de tous les clients dans un nouvel environnement.

À l'aide de ces sauvegardes, nous annulons régulièrement les suppressions accidentelles de leurs propres données par des clients individuels ou un petit groupe de clients. Cependant, la suppression au niveau du site ne comportait pas de runbooks pouvant être rapidement automatisés selon l'ampleur de cet événement, ce qui nécessitait des outils et une automatisation de tous les produits et services de manière coordonnée.

Cependant, nous n'avons pas (encore) automatisé la restauration pour un large sous-ensemble de clients dans notre environnement existant (et actuellement utilisé) sans affecter aucun de nos autres clients.

Au sein de notre environnement cloud, chaque ensemble de données stocke des données de plusieurs clients. Étant donné que les données supprimées au cours de cet incident ne concernaient qu'une partie des ensembles de données qui continuent d'être utilisés par d'autres clients, nous devons extraire et restaurer manuellement des parties individuelles de nos sauvegardes. La récupération de chaque site client est un processus long et complexe, qui nécessite une validation interne et une vérification finale du client au terme de l'opération.

**Actions :**

- **Accélérer les restaurations multiproduits et multisites pour un plus grand nombre de clients :** le programme de reprise d'activité répond à nos normes RPO actuelles d'une heure. Nous exploiterons l'automatisation et les leçons tirées de cet incident pour accélérer le programme de reprise d'activité, afin de respecter le RTO tel que défini dans notre politique pour cette ampleur d'incident.

**Automatiser et ajouter la vérification de ce cas aux tests de reprise d'activité :** nous organiserons régulièrement des exercices de reprise d'activité qui impliquent la restauration de tous les produits pour un grand nombre de sites. Ces tests de reprise d'activité vérifieront que les runbooks sont à jour au fur et à mesure que notre architecture évolue et que de nouveaux cas périphériques sont rencontrés. Nous améliorerons continuellement notre approche de restauration, automatiserons davantage le processus de restauration et réduirons le temps de restauration.

## Enseignement 3 : améliorer le processus de gestion des incidents pour les événements à grande échelle

Notre programme de gestion des incidents est parfaitement adapté à la gestion des incidents majeurs et mineurs qui se sont produits au fil des ans. Nous procédons souvent à des simulations de réponse aux incidents de plus petite envergure et de plus courte durée, qui impliquent généralement moins de personnes et d'équipes.

Cependant, au plus fort de cet incident, des centaines d'ingénieurs et d'employés de l'assistance client ont travaillé simultanément pour restaurer les sites des clients. Notre programme et nos équipes de gestion des incidents n'ont pas été conçus pour gérer l'ampleur, l'étendue et la durée de ce type d'incident (voir *image 10* ci-dessous).

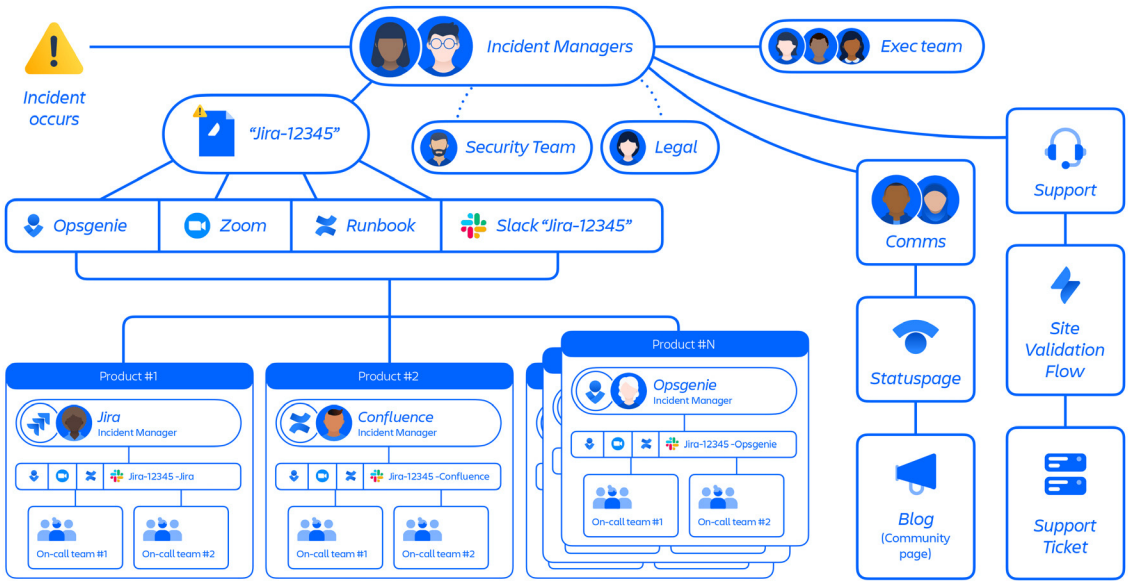



*Image 10 : vue d'ensemble du processus de gestion des incidents à grande échelle.*

**Notre processus de gestion des incidents à grande échelle sera mieux défini et souvent mis en pratique**

Nous disposons de mini-guides pratiques pour les équipes concernant les incidents au niveau des produits, mais pas pour les événements de cette ampleur, avec des centaines de personnes travaillant simultanément dans toute l'entreprise. Nos outils de gestion des incidents comportent une automatisation qui crée des flux de communication, tels que Slack, Zoom et Confluence, mais il leur manque la capacité de créer les sous-flux nécessaires pour les incidents à grande échelle afin d'isoler les flux de restauration.

**Actions :**
**Définir un playbook des outils pour les incidents de grande envergure et mener des exercices simulés :** définir et documenter les types d'incidents qui peuvent être considérés comme de grande échelle et qui nécessitent ce niveau de réponse. Décrire les principales étapes de coordination et créer des outils pour aider les gestionnaires d'incidents et les autres fonctions de l'entreprise à optimiser la réponse et à démarrer la récupération. Les gestionnaires d'incidents avec des équipes organiseront régulièrement des simulations, des formations et affineront les outils et les documents afin de les améliorer continuellement.

## Enseignement 4 : améliorer nos processus de communication

**a) Nous avons supprimé les identifiants clients majeurs, ce qui a eu un impact sur les communications et les actions des personnes concernées**
Le même script qui a supprimé les sites clients a également supprimé les principaux identifiants clients (par ex. l'URL du cloud, les contacts de l'administrateur système du site) de nos environnements de production. En conséquence, (1) les clients n'ont pas pu déposer de ticket d'assistance technique via notre canal d'assistance habituel ; (2) il nous a fallu des jours pour obtenir une liste fiable des principaux contacts clients (tels que les administrateurs système des sites) touchés par la panne pour permettre un engagement proactif ; et (3) les flux de travail d'assistance, les SLA, les tableaux de bord et les processus de remontée n'ont pas fonctionné correctement à l'origine, en raison de la nature unique de l'incident.

Pendant la panne, les remontées des clients sont également passées par plusieurs canaux (e-mails, appels téléphoniques, tickets du PDG, LinkedIn et autres réseaux sociaux, et tickets d'assistance). Des outils et des processus disparates au sein de nos équipes en contact avec les clients ont ralenti notre réponse et ont rendu plus difficiles le suivi et les rapports holistiques de ces remontées.

**b) Nous ne bénéficions pas d'un playbook sur la communication sur les incidents qui soit suffisamment complet pour faire face à ce niveau de complexité**

Nous n'avions pas de playbook de communication sur les incidents qui décrivait les principes ainsi que les rôles et les responsabilités nécessaires pour mobiliser suffisamment rapidement sur les incidents une équipe de communication unifiée et interfonctionnelle. Nous n'avons pas reconnu de manière rapide et cohérente l'existence de l'incident par le biais de multiples canaux, en particulier sur les réseaux sociaux. Plus largement, la bonne approche aurait été de faire des communications publiques concernant la panne, ainsi que de répéter le message critique selon lequel il n'y a pas eu de perte de données et que cela n'était pas le résultat d'une cyberattaque.

**Actions :**

- **Améliorer la sauvegarde des contacts clés :** sauvegarder les informations de contact de compte autorisé en dehors de l'instance du produit.
- **Outils d'assistance de mise à niveau :** créer des mécanismes permettant aux clients ne disposant pas d'une URL de site valide ou d'un identifiant Atlassian d'entrer en contact direct avec notre équipe d'assistance technique.
- **Système et processus de remontée client :** investir dans un système de remontée unifié et basé sur les comptes et dans des flux de travail qui permettent de stocker plusieurs objets de travail (tickets, tâches, etc.) sous un seul objet de compte client, afin d'améliorer la coordination et la visibilité au sein de toutes nos équipes en contact avec les clients.
- **Accélérer la couverture de la gestion des remontées 24 heures sur 24, 7 jours sur 7 :** exécuter les plans d'expansion de la présence mondiale de la fonction de gestion des remontées afin de permettre une couverture cohérente 24 heures sur 24, 7 jours sur 7, avec du personnel dédié basé dans chaque région géographique majeure, ainsi que des rôles d'assistance pour venir en aide aux experts et à la direction nécessaires en matière de produits et de ventes.
- **Mettre à jour notre playbook de communication sur les incidents avec les nouvelles leçons apprises et le consulter régulièrement :** consulter le playbook pour définir clairement les rôles et les lignes de communication en interne. Utiliser le framework DACI pour les incidents et disposer de sauvegardes 24 heures sur 24 et 7 jours sur 7 pour chaque rôle en cas de maladie, vacances ou autre événement imprévu. Procéder à un audit trimestriel pour vérifier à tout moment l'état de préparation.

*Actions (suite)*

Suivre le modèle de communication sur les incidents dans toutes les communications : expliquer ce qui s'est passé, qui a été touché, indiquer le calendrier de restauration, les pourcentages de restauration du site, les pertes de données attendues, les niveaux de confiance associés, ainsi que des conseils clairs sur la façon de contacter l'assistance.

# Remarques de clôture

Même si la panne est résolue et les clients sont entièrement restaurés, notre travail n'est pas terminé. À ce stade, nous mettons en œuvre les changements décrits ci-dessus pour améliorer nos processus, accroître notre résilience et empêcher qu'une telle situation se reproduise.

Atlassian est une organisation qui continue d'apprendre, et nos équipes ont tiré de nombreuses leçons difficiles de cette expérience. Nous mettons ces leçons à profit afin d'apporter des changements durables à notre activité. En fin de compte, nous en sortirons plus forts et vous fournirons un meilleur service grâce à cette expérience.

Nous espérons que les leçons tirées de cet incident seront utiles aux autres équipes qui travaillent avec diligence pour fournir des services fiables à leurs clients.

Enfin, je tiens à remercier ceux qui lisent ce document et apprennent avec nous, ainsi que ceux qui font partie de notre communauté et de notre équipe Atlassian.

-Sri Viswanath, directeur technique





ATLASSIAN