ATLASSIAN

Przegląd po incydencie

Awaria z kwietnia 2022 r.

Niniejsze tłumaczenie zostało udostępnione wyłącznie dla wygody. W przypadku jakichkolwiek niejasności lub konfliktów między tłumaczeniami pierwszeństwo ma oryginalna wersja angielska.

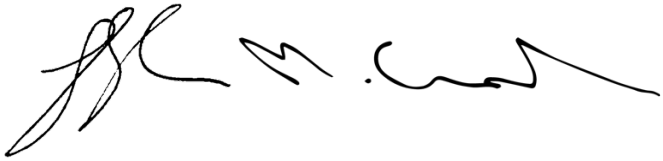
List od naszych współzałożycieli i dyrektorów generalnych

Pragniemy potwierdzić awarię, która spowodowała przerwę w dostawie usług dla klientów na początku tego miesiąca. Zdajemy sobie sprawę, że nasze produkty mają kluczowe znaczenie dla Twojej firmy i traktujemy zaistniałą sytuację z pełną powagą. To my ponosimy za nią odpowiedzialność. Kropka. Klientom, których dotknęła ta sytuacja, chcemy przekazać, że dokładamy starań, aby odzyskać Wasze zaufanie.

> Jedną z kluczowych wartości w Atlassian można opisać hasłem „Otwarta firma, bez nonsensów". Wcielamy tę wartość w życie częściowo poprzez otwarte omawianie incydentów i wykorzystywanie ich jako okazji do wyciągania wniosków. Niniejszy przegląd po incydencie publikujemy dla naszych klientów, członków społeczności Atlassian, a także szerszej społeczności technicznej. Firma Atlassian jest dumna ze swojego procesu zarządzania incydentami, w którym podkreśla się, że kultura eliminująca szukanie winnych i skupienie się na identyfikowaniu sposobów doskonalenia naszych systemów i procesów technicznych są niezbędne do świadczenia godnych zaufania usług na dużą skalę. Dokładamy wszelkich starań, aby uniknąć incydentów każdego rodzaju, ale jednocześnie uważamy, że stanowią one ważny sposób wprowadzania ulepszeń.

Zapewniamy, że platforma Atlassian Cloud pozwala nam sprostać różnorodnym potrzebom ponad 200 000 klientów każdej wielkości i z każdej branży, którzy korzystają z produktów chmurowych. Przed tym incydentem nasza platforma Cloud niezmiennie zapewniała 99,9% czasu dostępności, a nawet przekraczała poziomy nieprzerwanego działania określone w umowach SLA. Dokonaliśmy długoterminowych inwestycji w naszą platformę i szereg jej scentralizowanych możliwości, zapewniając skalowalną infrastrukturę i regularne dostarczanie ulepszeń w zakresie bezpieczeństwa.

Naszym klientom i partnerom dziękujemy za nieustające zaufanie i współpracę. Mamy nadzieję, że szczegółowe informacje i działania przedstawione w niniejszym dokumencie dowodzą, że Atlassian nadal będzie dostarczać światowej klasy platformę chmurową i oferować bogate portfolio produktów spełniające potrzeby każdego zespołu.



Scott i Mike

Streszczenie dla kierownictwa

We wtorek, 5 kwietnia 2022 r., począwszy od 7:38 czasu UTC 775 klientów Atlassian straciło dostęp do swoich produktów Atlassian. Dla części tych klientów awaria trwała nawet 14 dni, przy czym witryny pierwszych klientów zostały przywrócone 8 kwietnia, a wszystkich klientów stopniowo do 18 kwietnia.

Zaistniała sytuacja nie była wynikiem cyberataku i nie doszło do nieautoryzowanego dostępu do danych klientów. Atlassian prowadzi kompleksowy program zarządzania danymi z publikowanymi umowami SLA i historią przekraczania celów określonych w tych umowach.

Choć incydent był poważny, żaden klient nie utracił danych z okresu dłuższego niż pięć minut. Ponadto podczas działań związanych z przywracaniem danych ponad 99,6% naszych klientów i użytkowników nadal mogło korzystać z produktów chmurowych bez żadnych zakłóceń.

> W niniejszym dokumencie klientów, których witryny zostały usunięte w związku z incydentem, będziemy nazywali „poszkodowanymi". Ten przegląd po incydencie zawiera szczegółowe informacje na temat incydentu oraz opis kroków podjętych w celu odzyskania sprawności i sposobu zapobiegania wystąpieniu podobnych sytuacji w przyszłości. W tej sekcji prezentujemy ogólne podsumowanie incydentu. Dodatkowe szczegóły na jego temat można znaleźć w dalszej części dokumentu.

Opis incydentu

W 2021 roku zakończyliśmy przejęcie i integrację samodzielnej aplikacji Atlassian dla Jira Service Management i Jira Software o nazwie „Insight – Asset Management". Funkcje tej samodzielnej aplikacji stały się wówczas natywne w Jira Service Management i przestały być dostępne w Jira Software. W związku z tym musieliśmy usunąć starszą, samodzielną wersję aplikacji w witrynach klientów, w których była ona zainstalowana. Nasze zespoły inżynierskie użyły istniejącego skryptu i procesu w celu usunięcia instancji tej samodzielnej aplikacji, jednak wystąpiły dwa problemy:

- **Luka komunikacyjna**. Powstała luka komunikacyjna między zespołem, który złożył wniosek o usunięcie, a zespołem, który to usunięcie przeprowadził. Zamiast podania identyfikatorów wspomnianej aplikacji, *która miała zostać usunięta,*

zespół podał identyfikatory całej witryny Cloud*, z której aplikacje miały zostać usunięte.*

- **Niewystarczające ostrzeżenia systemowe.** Interfejs API używany do przeprowadzenia usunięcia akceptował zarówno identyfikatory witryn, jak i aplikacji, i zakładał, że dane wejściowe były poprawne. Oznaczało to, że podanie identyfikatora witryny spowoduje usunięcie witryny, a podanie identyfikatora aplikacji – usunięcie aplikacji. Nie było sygnału ostrzegawczego z wymogiem potwierdzenia żądanego typu usunięcia (witryny lub aplikacji).

Wykonany skrypt był zgodny z naszym standardowym procesem recenzowania przez współpracowników, który skupia się na identyfikacji wywoływanego punktu końcowego i sposobie jego wywoływania. Nie przeprowadzono krzyżowej weryfikacji podanych identyfikatorów witryn Cloud w celu potwierdzenia, że odnoszą się one do aplikacji Insight, a nie do całej witryny, a problem polegał na tym, że skrypt zawierał identyfikator całej witryny klienta. W rezultacie 5 kwietnia 2022 r. między godziną 7:38 a 8:01 czasu UTC usunięto bezpośrednio 883 witryny (należące do 775 klientów). *Patrz sekcja „Opis incydentu”*

Nasza reakcja

Po potwierdzeniu incydentu 5 kwietnia o 8:17 czasu UTC uruchomiliśmy nasz proces zarządzania poważnymi incydentami i powołaliśmy interdyscyplinarny zespół ds. zarządzania incydentem. Globalny zespół ds. reakcji na incydenty pracował 24 godziny na dobę, 7 dni w tygodniu przez cały czas trwania incydentu, dopóki wszystkie witryny nie zostały przywrócone, zatwierdzone i zwrócone klientom. Ponadto liderzy ds. zarządzania incydentem spotykali się co trzy godziny, aby koordynować strumienie prac.

Już na początku zdaliśmy sobie sprawę z licznych wyzwań związanych z równoczesnym przywróceniem wielu produktów setkom klientów.

Na początku incydentu wiedzieliśmy dokładnie, których witryn dotyczył problem, i naszym priorytetem było nawiązanie komunikacji z zatwierdzonym właścicielem każdej z usuniętych witryn w celu poinformowania o awarii.

Jednak dane kontaktowe niektórych klientów zostały usunięte. Oznaczało to, że klienci nie mogli utworzyć zgłoszeń do działu wsparcia w standardowym trybie. Oznaczało to również, że nie mieliśmy bezpośredniego dostępu do kluczowych osób kontaktowych po stronie klientów. *Więcej informacji zawiera sekcja „Ogólne omówienie strumieni roboczych w procesie odzyskiwania”*

Co robimy, aby zapobiec takim sytuacjom w przyszłości?

Podjęliśmy szereg natychmiastowych działań i zobowiązujemy się do wprowadzania zmian, które pozwolą nam uniknąć podobnej sytuacji w przyszłości. Poniżej przedstawiamy cztery konkretne obszary, w których wprowadziliśmy lub wprowadzimy istotne zmiany:

1. **Wprowadzenie uniwersalnej opcji „nietrwałego usuwania" we wszystkich systemach**. Zasadniczo usuwanie tego rodzaju powinno być zabronione lub zabezpieczone wielowarstwowo w celu uniknięcia błędów, z uwzględnieniem wdrażania etapowego i przetestowanego planu wycofania „nietrwałego usuwania". Globalnie uniemożliwimy usunięcie danych i metadanych klientów, które nie zostaną wcześnie poddane procesowi „nietrwałego usuwania.

2. **Przyspieszenie procesów w ramach naszego programu odzyskiwania awaryjnego w celu zautomatyzowania przywracania po zdarzeniach usunięcia wielu witryn i produktów dla większego zbioru klientów.** Wykorzystamy automatyzację oraz wnioski z zaistniałego incydentu w celu przyspieszenia kroków programu odzyskiwania awaryjnego, aby osiągnąć docelowy czas odzyskiwania (RTO) zgodny z naszymi zasadami w przypadku incydentów na taką skalę. Regularnie będziemy przeprowadzać ćwiczenia odzyskiwania awaryjnego z uwzględnieniem przywracania wszystkich produktów dla dużych zbiorów witryn.

3. **Modyfikacja procesu zarządzania incydentami w przypadku incydentów o dużej skali**. Udoskonalimy naszą standardową procedurę postępowania w przypadku incydentów o dużej skali i przećwiczymy ją w trakcie symulowanych incydentów o podobnej skali. Zaktualizujemy nasze szkolenia oraz narzędzia pod kątem obsługi dużej liczby zespołów pracujących równolegle.

4. **Opracowanie porad strategicznych dotyczących komunikacji w przypadku incydentów o dużej skali**. Będziemy potwierdzali wystąpienie incydentów na wczesnym etapie za pośrednictwem wielu kanałów. W ciągu kilku godzin będziemy ogłaszać publiczne komunikaty o incydentach. Aby ułatwić dotarcie do poszkodowanych klientów, ulepszymy tworzenie kopii zapasowej kluczowych kontaktów i unowocześnimy narzędzia wsparcia, aby umożliwić klientom nieposiadającym prawidłowego adresu URL lub identyfikatora Atlassian bezpośredni kontakt z naszym zespołem wsparcia technicznego.

Kompletna lista czynności do wykonania została opisana szczegółowo w pełnym przeglądzie po incydencie poniżej. *Patrz sekcja „Wprowadzane ulepszenia"*

Spis treści

Omówienie architektury chmury Atlassian Strona 7

- Architektura hostingu w chmurze Atlassian
- Architektura usług rozproszonych
- Architektura z wielodostępem
- Aprowizacja i cykl życia dzierżaw
- Program odzyskiwania awaryjnego
    - Niezawodność
    - Możliwość przywrócenia magazynu usług
    - Zautomatyzowane przywracanie wielu produktów i witryn

Opis incydentu, oś czasu i odzyskiwanie Strona 13

- Co się stało
- Koordynacja działań
- Oś czasu incydentu
- Ogólne omówienie strumieni roboczych w procesie odzyskiwania
    - Strumień prac 1: Wykrycie, przystąpienie do odzyskiwania i identyfikacja podejścia
    - Strumień prac 2: Wczesne odzyskiwanie i podejście Przywracanie 1
    - Strumień prac 3: Przyspieszone odzyskiwanie i podejście Przywracanie 2
    - Minimalna utrata danych po przywróceniu usuniętych witryn

Komunikacja podczas incydentu Strona 21

- Co się stało

Wsparcie i kontakt z klientem Strona 23

- Wpływ na wsparcie udzielane naszym klientom
- Nasza reakcja

Wprowadzane ulepszenia Strona 25

- Wniosek 1: We wszystkich systemach stosować uniwersalnie „usuwanie nietrwałe”
- Wniosek 2: W ramach programu odzyskiwania awaryjnego zautomatyzować przywracanie po zdarzeniach usunięcia wielu witryn i produktów dla większych grup klientów
- Wniosek 3: Poprawić proces zarządzania incydentami w przypadku zdarzeń o dużej skali
- Wniosek 4: Poprawić nasze procesy komunikacji

Uwagi końcowe Strona 31

Omówienie architektury chmury Atlassian

Aby zrozumieć omówione w niniejszym dokumencie czynniki, które przyczyniły się do incydentu, warto najpierw zapoznać się z architekturą wdrażania produktów, usług i infrastruktury Atlassian.

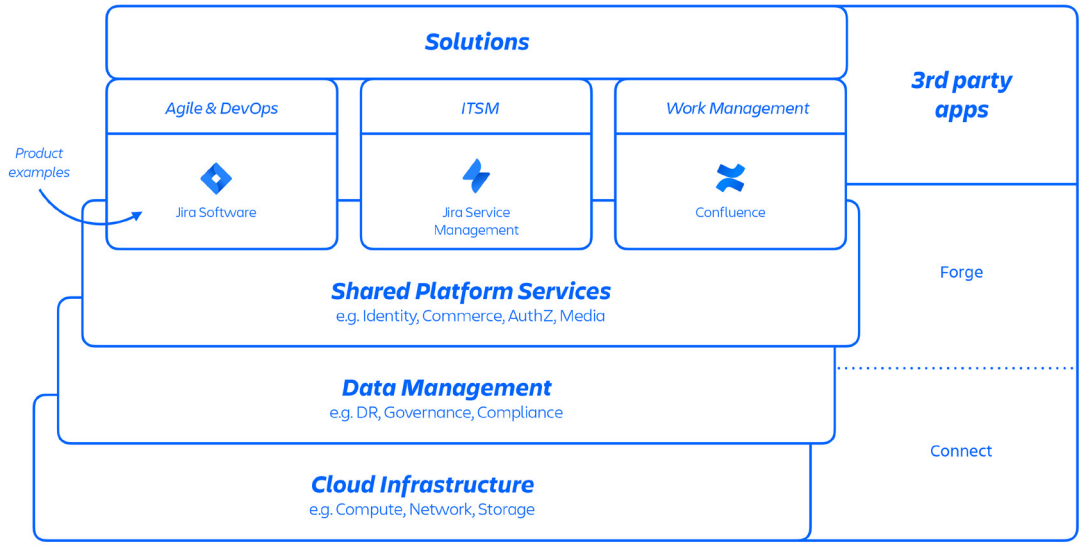
Architektura hostingu w chmurze Atlassian

Atlassian korzysta z Amazon Web Services (AWS) jako dostawcy usług w chmurze i ich centrów danych o wysokiej dostępności zlokalizowanych w wielu regionach na całym świecie. Każdy region AWS to oddzielna lokalizacja geograficzna z wieloma, odizolowanymi i fizycznie odrębnymi grupami centrów danych nazywanymi strefami dostępności (Availability Zone, AZ).

Podczas opracowywania komponentów naszych produktów i platform korzystamy z usług danych, obliczeniowych, magazynowania, i sieciowych AWS oraz możliwości nadmiarowości, które oferuje AWS, takich jak strefy dostępności i regiony.

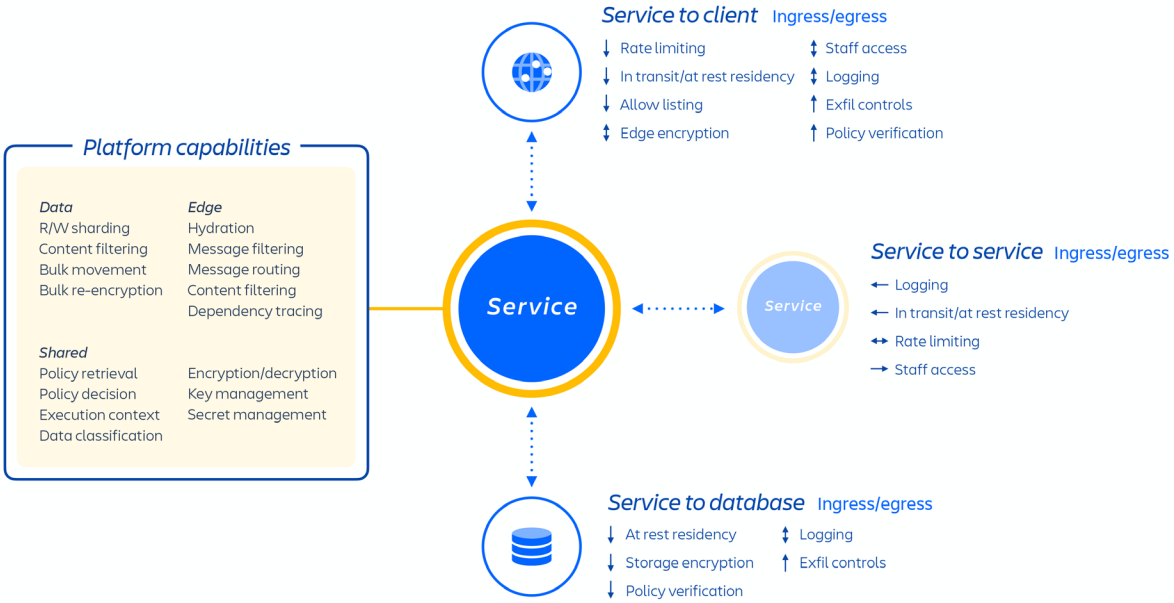
Architektura usług rozproszonych

Wykorzystujemy tę architekturę AWS do hostowania wielu usług platformowych i produktowych stosowanych w naszych rozwiązaniach. Dotyczy to funkcji platformy, które są współdzielone i stosowane w wielu produktach Atlassian, takich jak Media, Identity, Commerce, środowisk, takich jak nasz Editor, a także możliwości specyficznych dla konkretnego produktu, takich jak usługa zgłoszeń Jira i analizy Confluence.



Rysunek 1: Architektura platformy Atlassian.

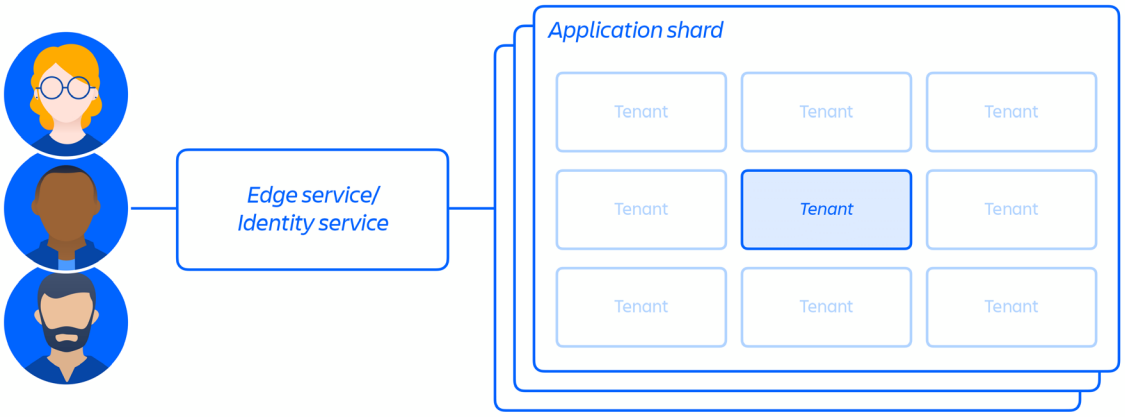
Programiści Atlassian aprowizują te usługi za pośrednictwem opracowanej wewnętrznie platformy jako usługi (PaaS) o nazwie Micros, która automatycznie koordynuje wdrażanie usług współdzielonych, infrastruktury, magazynów danych oraz możliwości zarządzania nimi, z uwzględnieniem wymagań w zakresie bezpieczeństwa i kontroli zgodności z przepisami (patrz *Rysunek 1* powyżej). Zazwyczaj produkt Atlassian składa się z wielu „konteneryzowanych" usług wdrażanych w środowisku AWS przy użyciu platformy Micros. Produkty Atlassian korzystają z możliwości platformy podstawowej (patrz *Rysunek 2* poniżej) – od przekierowywania wniosków po magazyny obiektów binarnych, uwierzytelnianie/autoryzację, zawartość transakcyjną generowaną przez użytkowników (UGC) i magazyny modeli encja-relacja, jeziora danych, wspólne logowanie, śledzenie wniosków, możliwość wglądu czy usługi analityczne. Te mikrousługi są tworzone przy użyciu zatwierdzonych pakietów rozwiązań technicznych znormalizowanych na poziomie platformy:



Rysunek 2: Przegląd mikrousług Atlassian.

Architektura z wielodostępem

Wykorzystując naszą infrastrukturę chmurową, zbudowaliśmy i utrzymujemy wielodostępną architekturę mikrousług ze współdzieloną platformą do obsługi naszych produktów. W architekturze z wielodostępem pojedyncza usługa obsługuje wielu klientów, w tym bazy danych i instancje obliczeniowe wymagane do uruchamiania naszych produktów chmurowych. Każdy shard (zasadniczo kontener – patrz *Rysunek 3* poniżej) zawiera dane wielu dzierżaw, jednak dane każdej dzierżawy są odizolowane i niedostępne dla innych dzierżaw.



Rysunek 3: Sposób przechowywania danych w architekturze z wielodostępem.

Aprowizacja i cykl życia dzierżaw

Po aprowizacji nowego klienta seria zdarzeń wyzwala orkiestrację usług rozproszonych i aprowizację magazynów danych. Te zdarzenia można zasadniczo przypisać do jednego z siedmiu etapów cyklu życia:

1. Systemy handlowe są niezwłocznie aktualizowane o najnowsze metadane i informacje o kontroli dostępu dla danego klienta, a następnie system orkiestracji aprowizacji dopasowuje „stan aprowizowanych zasobów” do stanu licencji poprzez serię zdarzeń związanych z dzierżawą i produktami.

Zdarzenia dotyczące dzierżawy
Zdarzenia, które wpływają na dzierżawę jako całość, i mogą obejmować:

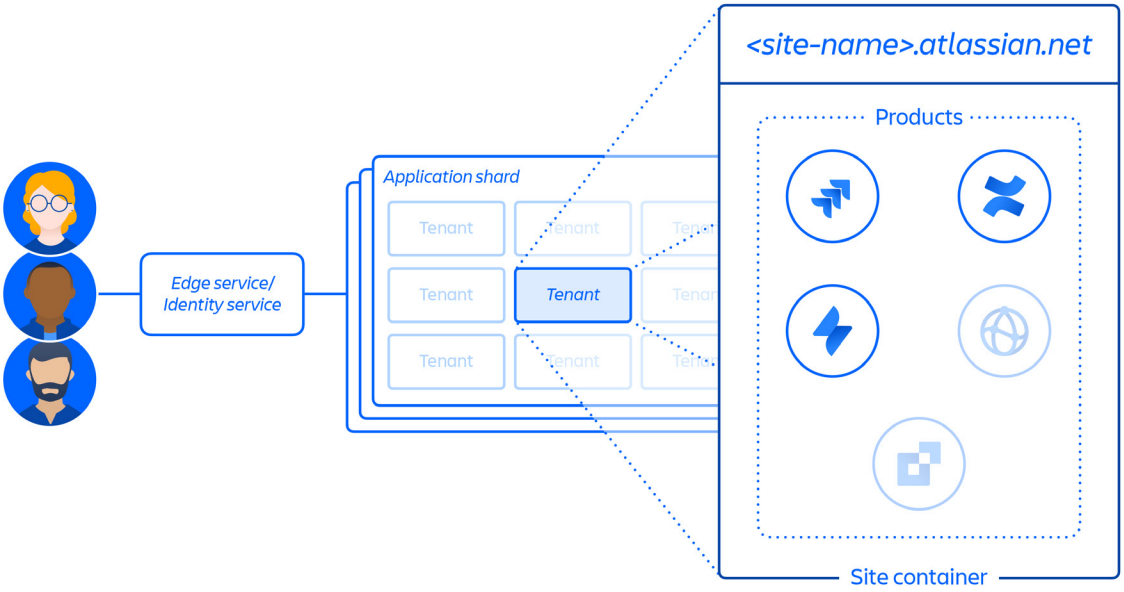
- Tworzenie: dzierżawa jest tworzona i stosowana do zupełnie nowych witryn.
- Likwidacja: cała dzierżawa jest usuwana.

Zdarzenia dotyczące produktów

- Aktywacja: po aktywacji licencjonowanych produktów lub aplikacji innych firm.
- Dezaktywacja: po dezaktywacji niektórych produktów lub aplikacji.
- Zawieszenie: po zawieszeniu konkretnego istniejącego produktu i tym samym zablokowaniu dostępu do posiadanej witryny.
- Wycofanie zawieszenia: po cofnięciu zawieszenia konkretnego istniejącego produktu i tym samym zapewnieniu dostępu do posiadanej witryny.

Aktualizacja licencji: zawiera informacje dotyczące liczby stanowisk licencyjnych danego produktu oraz jego statusu (aktywny/nieaktywny).

2. Utworzenie witryny klienta i aktywacja poprawnego zestawu produktów dla klienta. Zgodnie z koncepcją witryna jest kontenerem wielu produktów licencjonowanych przez konkretnego klienta (np. Confluence i Jira Software w przypadku witryny <site-name>.atlassian.net). To (patrz *Rysunek 4* poniżej) ważna kwestia w kontekście tego raportu, ponieważ w trakcie tego incydentu to właśnie kontenery witryn zostały usunięte, a koncepcja witryny jest przywoływana w całym dokumencie.



Rysunek 4: Przegląd kontenera witryny.

3. Aprowizacja produktów w obrębie witryny klienta w wyznaczonym regionie.

   Podczas aprowizacji produktu większość związanej z nim zawartości jest hostowana w pobliżu lokalizacji, w której użytkownicy uzyskują dostęp do produktu. Aby zoptymalizować wydajność produktu, nie ograniczamy przepływu danych, gdy są one hostowane globalnie, a w razie potrzeby możemy przenosić dane między regionami.

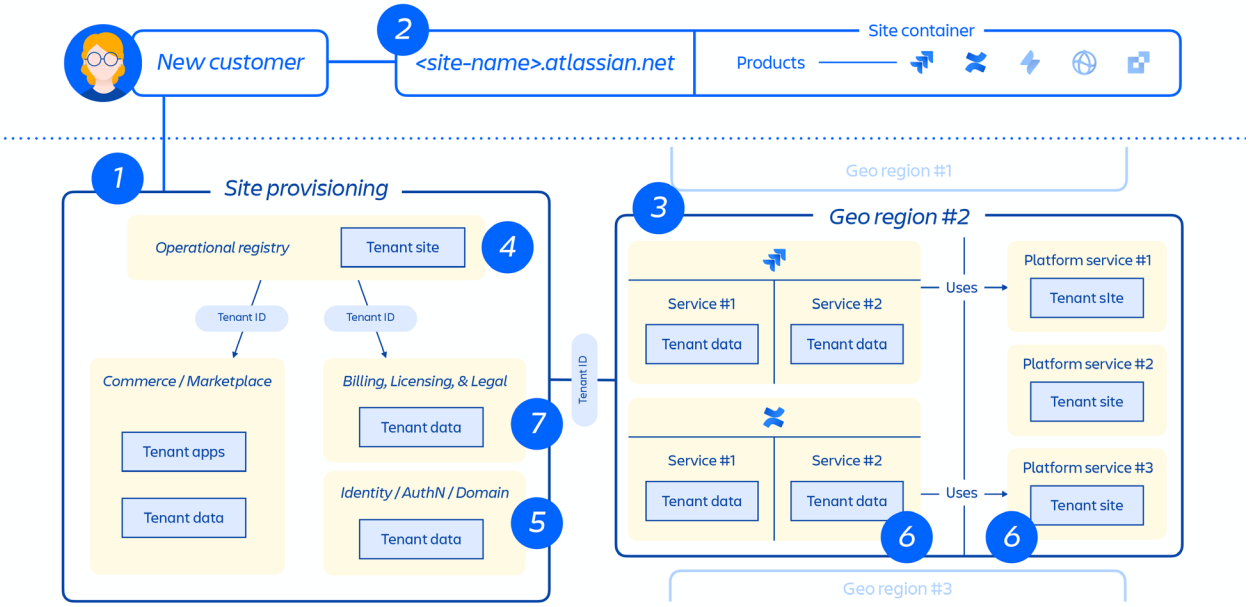
   W przypadku niektórych naszych produktów oferujemy również funkcję jurysdykcji danych. Pozwala ona klientom wybrać, czy dane produktów mają być rozproszone globalnie, czy przechowywane w jednej z naszych zdefiniowanych lokalizacji geograficznych.

4. Tworzenie i przechowywanie głównych metadanych i konfiguracji witryny oraz produktów klienta.

5. Tworzenie i przechowywanie danych tożsamości witryny i produktów, takich jak użytkownicy, grupy, uprawnienia itp.

6. Aprowizacja baz danych produktów w obrębie witryny, np. rodziny produktów Jira, Confluence, Compass, Atlas.

7. Aprowizacja licencjonowanych aplikacji do produktów.



Rysunek 5: Przegląd sposobu aprowizacji witryny klienta w naszej architekturze rozproszonej.

Na *Rysunku 5* powyżej przedstawiono sposób wdrażania witryny klienta w całej naszej architekturze rozproszonej, a nie tylko w pojedynczej bazie danych lub w pojedynczym magazynie. W procesie uczestniczy wiele lokalizacji fizycznych i logicznych, w których są przechowywane metadane, dane konfiguracji, dane produktu, dane platformy oraz inne powiązane informacje o witrynie.

Program odzyskiwania awaryjnego

Nasz program odzyskiwania awaryjnego obejmuje wszystkie nasze wysiłki na rzecz zapewnienia odporności na awarie infrastruktury i przywrócenia magazynu usług z kopii zapasowych. Z programami odzyskiwania awaryjnego wiążą się dwa ważne pojęcia:

- **Docelowy czas odzyskiwania (RTO):** określa, jak szybko można odzyskać i zwrócić klientowi dane po awarii.
- **Docelowy punkt odzyskiwania (RPO):** określa, na ile aktualne będą dane odzyskane w razie awarii z kopii zapasowej, i jaka ilość danych zostanie utracona od momentu wykonania ostatniej kopii zapasowej?

W trakcie tego incydentu nie udało nam się zmieścić w docelowym czasie odzyskiwania, ale osiągnęliśmy przewidziany docelowy punkt odzyskiwania.

Niezawodność

Przygotowujemy się na awarie na poziomie infrastruktury, na przykład na utratę całej bazy danych, usługi lub stref dostępności AWS. Takie przygotowanie obejmuje replikację danych i usług w wielu strefach dostępności oraz regularne testowanie przełączania na inną strefę w razie awarii.

Możliwość przywrócenia magazynu usług

Przygotowujemy się również na przywracanie sprawności po uszkodzeniu danych magazynu usług z powodu zagrożeń, takich jak ransomware, osoby działające w złej wierze, wady oprogramowania i błędy operacyjne. To przygotowanie obejmuje tworzenie niezmiennych kopii zapasowych i testowanie przywracania kopii zapasowych magazynu usług. Jesteśmy w stanie przywrócić dowolny indywidualny magazyn danych do poprzedniego punktu w czasie.

Zautomatyzowane przywracanie wielu produktów i witryn

W trakcie incydentu nie mieliśmy możliwości wyboru dużego zbioru witryn klientów i przywrócenia wszystkich znajdujących się w nich wzajemnie powiązanych produktów z kopii zapasowych do poprzedniego punktu w czasie.

Nasze możliwości były skoncentrowane na problemach infrastrukturalnych, uszkodzeniu danych, zdarzeniach związanych z pojedynczą usługą lub usunięciami pojedynczych witryn. W przeszłości mieliśmy do czynienia z tego rodzaju awariami i testowaliśmy sposoby postępowania w przypadku ich wystąpienia. W przypadku usunięcia na poziomie witryny nie mieliśmy wykazów procedur, które dałoby się szybko zautomatyzować, biorąc pod uwagę skalę tego zdarzenia. Wymagało ono skoordynowanego zastosowania narzędzi i automatyzacji obejmujących wszystkie produkty i usługi.

W kolejnych sekcjach omówimy bardziej szczegółowo tę złożoność oraz działania, jakie podejmujemy w Atlassian w celu rozszerzenia i zoptymalizowania naszych zdolności do utrzymywania tej architektury na dużą skalę.

Opis incydentu, oś czasu i odzyskiwanie

Co się stało

W 2021 roku zakończyliśmy integrację samodzielnej aplikacji Atlassian dla Jira Service Management i Jira Software o nazwie „Insight – Asset Management". Funkcje tej samodzielnej aplikacji stały się wówczas natywne w Jira Service Management i przestały być dostępne w Jira Software. W związku z tym musieliśmy usunąć starszą, samodzielną wersję aplikacji w witrynach klientów, w których była ona zainstalowana. Nasze zespoły inżynierskie użyły istniejącego skryptu i procesu w celu usunięcia instancji tej samodzielnej aplikacji.

Wystąpiły jednak dwa krytyczne problemy:

- **Luka komunikacyjna.** Powstała luka komunikacyjna między zespołem, który złożył wniosek o usunięcie, a zespołem, który je przeprowadził. Zamiast podania identyfikatorów wspomnianej aplikacji, która miała zostać usunięta, zespół podał identyfikatory całej witryny Cloud, z której aplikacje miały zostać usunięte.
- **Niewystarczające ostrzeżenia systemowe.** Interfejs API używany do przeprowadzenia usunięcia akceptował zarówno identyfikatory witryn, jak i aplikacji, i zakładał, że dane wejściowe były poprawne. Oznaczało to, że podanie identyfikatora witryny spowoduje usunięcie witryny, a podanie identyfikatora aplikacji – usunięcie aplikacji. Nie było sygnału ostrzegawczego z wymogiem potwierdzenia żądanego typu usunięcia (witryny lub aplikacji).

Wykonany skrypt był zgodny z naszym standardowym procesem recenzowania przez współpracowników, który skupia się na identyfikacji wywoływanego punktu końcowego i sposobie jego wywoływania. Nie przeprowadzono krzyżowej weryfikacji podanych identyfikatorów witryn Cloud w celu potwierdzenia, że odnoszą się one do aplikacji Insight, a nie do całej witryny. Skrypt został przetestowany w środowisku przejściowym zgodnie z naszymi standardowymi procesami zarządzania zmianami, jednak test nie dawał możliwości wykrycia braku poprawności wprowadzonych identyfikatorów, ponieważ identyfikatory nie występowały w środowisku przejściowym.

W środowisku produkcyjnym skrypt zastosowano najpierw do 30 witryn. Pierwsze uruchomienie produkcyjne zakończyło się powodzeniem i aplikacja Insight została pomyślnie usunięta w tych 30 witrynach bez żadnych skutków ubocznych. Jednak identyfikatory tych 30 witryn pozyskano przed wystąpieniem błędu w komunikacji i dlatego były to poprawne identyfikatory aplikacji Insight.

Skrypt dla kolejnego uruchomienia produkcyjnego zawierał identyfikatory witryn zamiast identyfikatorów aplikacji Insight i został wykonany w odniesieniu do 883 witryn. Wykonywanie skryptu rozpoczęto 5 kwietnia o 7:38 czasu UTC, a zakończono o 8:01 czasu UTC. Skrypt usuwał kolejne witryny zgodnie z listą danych wejściowych, zatem pierwsze witryny klientów zostały usunięte krótko po rozpoczęciu wykonywania skryptu o 7:38 czasu UTC. Efektem było bezpośrednie usunięcie 883 witryn bez żadnego sygnału ostrzegawczego dla naszych zespołów inżynierskich.

Poszkodowani klienci utracili dostęp do następujących produktów Atlassian: produkty z rodziny Jira, Confluence, Atlassian Access, Opsgenie i Statuspage.

Gdy tylko dowiedzieliśmy się o incydencie, nasze zespoły skoncentrowały się na przywróceniu danych wszystkich poszkodowanych klientów. Na tamten moment oszacowaliśmy, że liczba witryn dotkniętych problemem wynosi w przybliżeniu 700 (łącznie były to 883 witryny, jednak odjęliśmy witryny należące do Atlassian). Znaczną część spośród tych 700 witryn stanowiły nieaktywne, darmowe lub niewielkie konta o małej liczbie aktywnych użytkowników. Na tej podstawie początkowo oszacowaliśmy przybliżoną liczbę poszkodowanych klientów na około 400.

Obecnie dysponujemy znacznie dokładniejszym obrazem. Dla pełnej przejrzystości, awaria dotknęła 775 klientów odpowiadających oficjalnej definicji klienta obowiązującej w Atlassian. Jednak większość użytkowników należała do oszacowanej początkowo liczby 400 klientów. W przypadku podzbioru tych klientów awaria trwała maksymalnie 14 dni, przy czym działanie produktów dla pierwszego zbioru klientów przywrócono 8 kwietnia, natomiast 18 kwietnia przywrócono sprawność w przypadku wszystkich klientów.

Koordynacja działań

Pierwsze zgłoszenie do działu wsparcia zostało utworzone przez poszkodowanego klienta 5 kwietnia o 7:46 czasu UTC. Nasz monitoring wewnętrzny nie wykrył problemu, ponieważ witryny zostały usunięte zgodnie ze standardowym przepływem pracy. O 8:17 czasu UTC uruchomiliśmy nasz proces zarządzania poważnymi incydentami, powołując interdyscyplinarny zespół ds. zarządzania incydentem, i w ciągu siedmiu minut, o 8:24 czasu UTC incydent został eskalowany do stanu krytycznego. O 8:53 czasu UTC nasz zespół potwierdził, że zgłoszenie do działu wsparcia, które napłynęło od klienta, było powiązane z uruchomieniem skryptu. Gdy zdaliśmy sobie sprawę, jak złożony będzie proces przywracania, o 12:38 czasu UTC nadaliśmy incydentowi najwyższy poziom istotności.

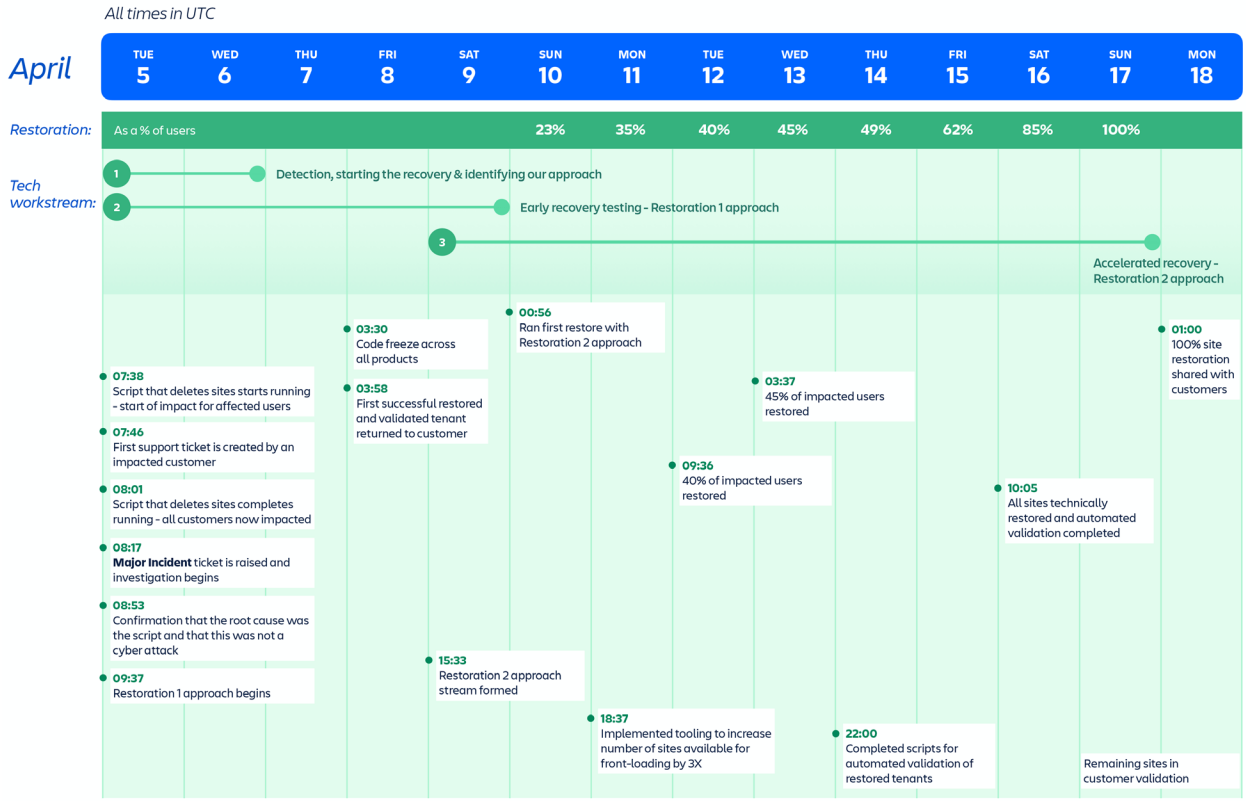
Zespół ds. zarządzania incydentem składał się z osób reprezentujących różne zespoły w Atlassian, w tym zespół inżynierski, obsługi klienta, zarządzania programami, komunikacji i wiele innych. Główny zespół spotykał się co trzy godziny przez cały czas

trwania incydentu, dopóki wszystkie witryny nie zostały przywrócone, zatwierdzone i zwrócone klientom.

W celu zarządzania procesem przywracania utworzyliśmy nowy projekt Jira o nazwie SITE oraz przepływ pracy pozwalający monitorować przebieg przywracania na poziomie poszczególnych witryn w wielu zespołach (inżynierskim, zarządzania programami, wsparcia itp.). Takie podejście umożliwiło wszystkim zespołom łatwe identyfikowanie i śledzenie zgłoszeń związanych z przywracaniem poszczególnych witryn.

8 kwietnia o 3:30 czasu UTC wdrożyliśmy również zamrożenie kodu we wszystkich zespołach inżynierskich na czas trwania incydentu. Pozwoliło nam to skupić się na przywracaniu danych klientów, wyeliminować ryzyko zmian powodujących niespójności w danych klientów, zminimalizować ryzyko innych awarii i zmniejszyć prawdopodobieństwo wystąpienia niepowiązanych zmian odwracających uwagę zespołu od procesu odzyskiwania.

Oś czasu incydentu



Rysunek 6: Oś czasu incydentu i kluczowe kamienie milowe przywracania.

Ogólne omówienie strumieni roboczych w procesie odzyskiwania

Odzyskiwanie przeprowadzono w oparciu o trzy podstawowe strumienie pracy – wykrywanie, wczesne odzyskiwanie i przyspieszenie. Choć poniżej każdy z tych strumieni pracy opisaliśmy oddzielnie, podczas odzyskiwania praca odbywała się równolegle we wszystkich strumieniach.

Strumień prac 1: Wykrycie, przystąpienie do odzyskiwania i identyfikacja podejścia

Znacznik czasu: dni 1–2 (5–6 kwietnia)

5 kwietnia o 8:53 czasu UTC ustaliliśmy, że skrypt przeznaczony do usunięcia aplikacji Insight spowodował usunięcie witryn. Potwierdziliśmy, że nie było to wynikiem wewnętrznego złośliwego działania ani cyberataku. Odpowiednie zespoły ds. produktów i infrastruktury platformy zostały przywołane i włączone do incydentu.

Na początku incydentu stwierdziliśmy, że:

- Przywracanie setek usuniętych witryn jest złożonym, wieloetapowym procesem (szczegółowo opisanym w sekcji architektury powyżej), którego pomyślne przeprowadzenie potrwa wiele dni i będzie wymagać udziału wielu zespołów.
- Mamy możliwość odzyskania pojedynczej witryny, ale nie opracowaliśmy funkcji ani procesów odzyskiwania dużej partii witryn.

W rezultacie musieliśmy w znacznym stopniu postawić na działania równoległe oraz automatyzację procesu przywracania, aby ułatwić poszkodowanym klientom jak najszybsze odzyskanie dostępu do ich produktów Atlassian.

W ramach strumienia prac 1 wiele zespołów programistycznych uczestniczyło w następujących działaniach:

- Identyfikacja i wykonanie kroków przywracania partii witryn w pipelinie.
- Utworzenie i udoskonalenie automatyzacji w celu umożliwienia zespołom wykonywania kroków przywracania większej liczby witryn partiami.

Strumień prac 2: Wczesne odzyskiwanie i podejście Przywracanie 1

Znacznik czasu: dni 1–4 (5–9 kwietnia)

W ciągu godziny od zakończenia wykonywania skryptu wiedzieliśmy już, co spowodowało usunięcie witryn 5 kwietnia o 8:53 czasu UTC. Zidentyfikowaliśmy również proces przywracania stosowany wcześniej w przypadku odzyskiwania niewielkiej liczby

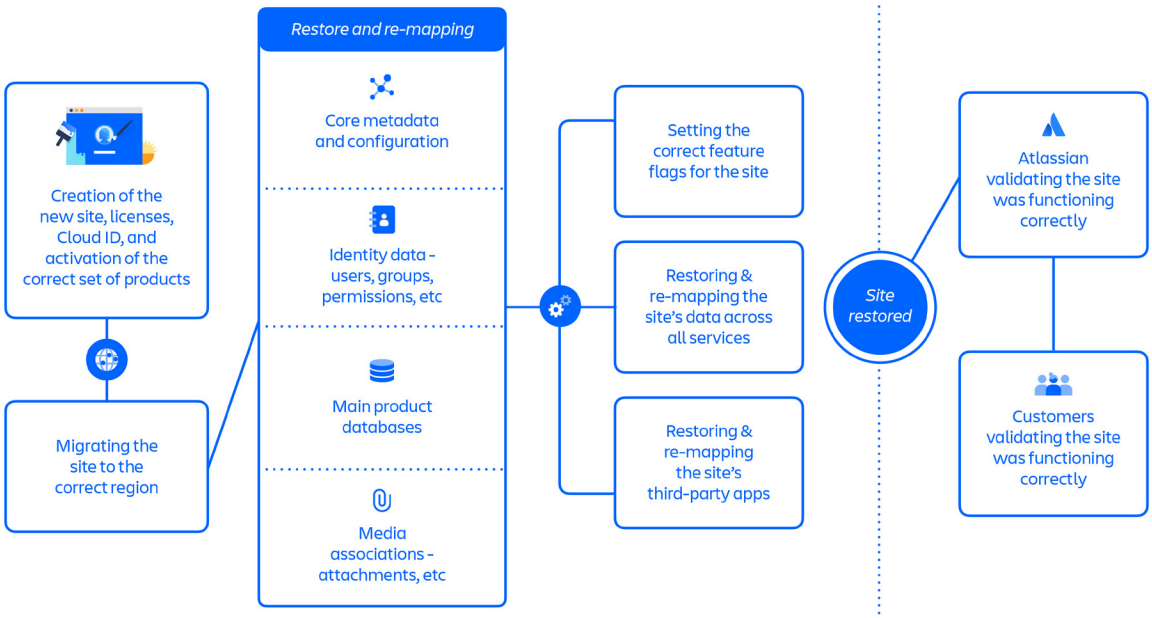
witryn do środowiska produkcyjnego. Jednak proces odzyskiwania obejmujący przywrócenie usuniętych witryn na taką skalę nie był dobrze zdefiniowany.

Aby szybko przystąpić do działania, na wczesnych etapach incydentu powołano dwie grupy robocze:

- Grupa robocza odpowiedzialna za procesy ręczne weryfikowała poprawność wymaganych kroków i ręcznie wykonywała proces przywracania na niewielkiej liczbie witryn.
- Grupa robocza odpowiedzialna za automatyzację przyjrzała się istniejącemu procesowi przywracania i opracowała reguły automatyzacji pozwalające bezpiecznie wykonywać kroki na większych partiach witryn.

Przegląd podejścia Przywracanie 1 (patrz *Rysunek 7* poniżej):

- Konieczne było utworzenie nowej witryny dla każdej usuniętej, a następnie przywrócenie danych każdego należącego do witryny produktu, magazynu danych i każdej usługi.
- Nowa witryna zawierała nowe identyfikatory, takie jak cloudId. Identyfikatory te są uważane za niezmienne, co oznacza, że wiele systemów osadza te identyfikatory w rekordach danych. W rezultacie zmiana tych identyfikatorów powoduje konieczność zaktualizowania dużej liczby danych, co jest szczególnie trudne w przypadku aplikacji innych firm wchodzących w skład ekosystemu.
- Zmodyfikowanie nowej witryny w celu odtworzenia stanu usuniętej witryny powodowało powstawanie złożonych i często nieprzewidzianych zależności między poszczególnymi krokami.



Rysunek 7: Kluczowe kroki podejścia Przywracanie 1.

Podejście Przywracanie 1 obejmowało około 70 pojedynczych kroków, które po zagregowaniu na poziomie ogólnym tworzyły zasadniczo następujący sekwencyjny przepływ czynności:

- Utworzenie nowej witryny, nowych licencji, identyfikatora Cloud i aktywowanie właściwego zestawu produktów.
- Migracja witryny do właściwego regionu.
- Przywrócenie i ponowne zmapowanie podstawowych metadanych i konfiguracji witryny.
- Przywrócenie i ponowne zmapowanie danych tożsamości witryny – użytkowników, grup, uprawnień itp.
- Przywrócenie baz danych głównych produktów witryny.
- Przywrócenie i ponowne zmapowanie powiązań multimediów w witrynie – załączników itp.
- Ustawienie prawidłowych flag funkcji dla witryny.
- Przywrócenie i ponowne zmapowanie danych witryny we wszystkich usługach.
- Przywrócenie i ponowne zmapowanie aplikacji innych firm używanych w witrynie.
- Zweryfikowanie przez Atlassian poprawności działania witryny.
- Zweryfikowanie przez klientów poprawności działania witryny.

Po zoptymalizowaniu przywrócenie partii witryn z zastosowaniem podejścia Przywracanie 1 zajmowało około 48 godzin i ten sposób odzyskiwana użyto w przypadku 53% poszkodowanych użytkowników i 112 witryn w okresie od 5 do 14 kwietnia.

Strumień prac 3: Przyspieszone odzyskiwanie i podejście Przywracanie 2

Znacznik czasu: dni 4–13 (9–17 kwietnia)

Przywrócenie danych wszystkich klientów przy zastosowaniu podejścia Przywracanie 1 zajęłoby nam do trzech tygodni. W związku z tym 9 kwietnia zaproponowaliśmy nowe podejście, Przywracanie 2, aby przyspieszyć przywracanie wszystkich witryn (patrz *Rysunek 8* poniżej).

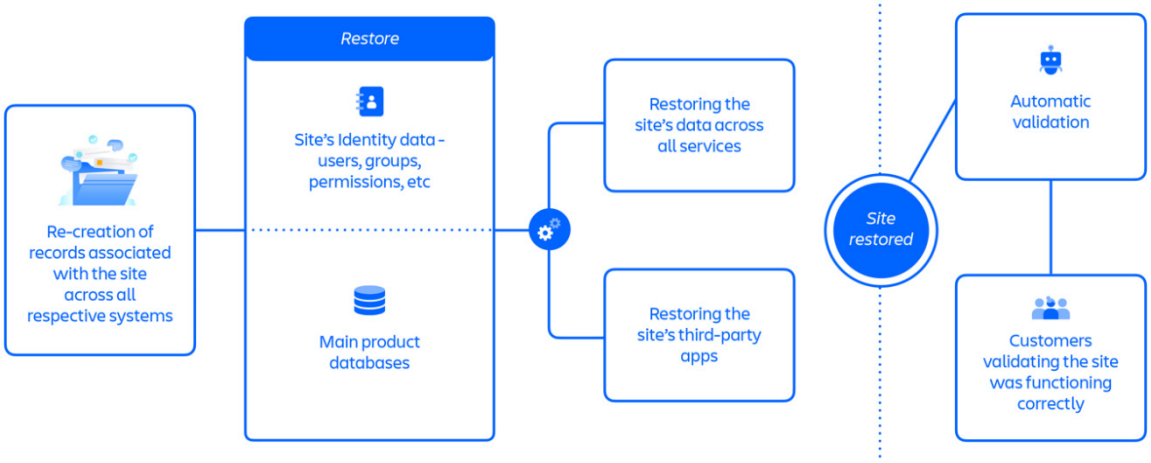
W przypadku podejścia Przywracanie 2 poszczególne kroki przebiegały w większym stopniu równolegle dzięki ograniczeniu złożoności oraz liczby zależności, które występowały w podejściu Przywracanie 1.

Podejście Przywracanie 2 polegało na ponownym utworzeniu (lub cofnięciu usunięcia) rekordów powiązanych z witryną we wszystkich odpowiednich systemach, począwszy od rekordu usługi katalogu. Kluczowym elementem tego nowego podejścia było *ponowne wykorzystanie wszystkich starych identyfikatorów witryny*. Dzięki temu usunięto z poprzedniego procesu ponad połowę kroków związanych z mapowaniem

starych identyfikatorów na nowe identyfikatory, eliminując między innymi konieczność koordynowania działań z poszczególnymi dostawcami aplikacji zewnętrznych w przypadku każdej witryny.

Jednak przejście od podejścia Przywracanie 1 do Przywracanie 2 spowodowało znaczne zwiększenie nakładu pracy w zakresie reagowania na incydent:

- Wiele skryptów automatyzacji oraz procesów opracowanych w ramach podejścia Przywracanie 1 wymagało zmodyfikowania w podejściu Przywracanie 2.
- Zespoły odpowiedzialne za przeprowadzanie przywracania (w tym koordynatorzy ds. incydentu) musiały zarządzać równoległymi partiami zadań przywracania w obu podejściach, podczas gdy my przeprowadzaliśmy testy i weryfikację procesu w podejściu Przywracanie 2.
- Zastosowanie nowego podejścia oznaczało, że musieliśmy przetestować i zweryfikować proces w podejściu Przywracanie 2 przed zastosowaniem go na szerszą skalę, co wymagało zdublowania prac weryfikacyjnych przeprowadzonych wcześniej w ramach podejścia Przywracanie 1.



Rysunek 8: Kluczowe kroki podejścia Przywracanie 2.

Przedstawiony powyżej wykres przedstawia podejście Przywracanie 2 obejmujące ponad 30 kroków realizowanych w ramach następującego przepływu prac odbywających się w dużej mierze równolegle:

- Odtworzenie rekordów powiązanych z witryną we wszystkich odpowiednich systemach.
- Przywrócenie danych tożsamości witryny – użytkowników, grup, uprawnień itp.
- Przywrócenie baz danych głównych produktów witryny.
- Przywrócenie danych witryny we wszystkich usługach.
- Przywrócenie aplikacji innych firm używanych w witrynie.
- Automatyczna weryfikacja.
- Zweryfikowanie przez klientów poprawności działania witryny.

W ramach przyspieszonego odzyskiwania podjęliśmy również kroki w celu przeniesienia na wcześniejsze etapy i zautomatyzowania procesu przywracania witryny, ponieważ skalowanie przywracania ręcznego nie sprawdziłoby się w przypadku dużych partii. Sekwencyjny charakter procesu odzyskiwania oznaczał wolniejszy przebieg przywracania witryn w przypadku dużych baz danych i baz/uprawnień użytkowników. Wdrożyliśmy następujące optymalizacje:

- Opracowaliśmy narzędzia i „bariery ochronne" potrzebne do *przesunięcia na początek* i przeprowadzenia długotrwałych kroków, takich jak przywracanie baz danych i synchronizacja tożsamości, aby zostały ukończone przed innymi etapami przywracania.
- Zespoły inżynierskie opracowały reguły automatyzacji dla poszczególnych kroków, umożliwiając bezpieczne przywracanie większych partii danych.
- Opracowano reguły automatyzacji w celu weryfikacji działania witryn po wykonaniu wszystkich korków przywracania.

Przywrócenie witryny z zastosowaniem przyspieszonego podejścia Przywracanie 2 trwało około 12 godzin i ten sposób odzyskiwana użyto w przypadku około 47% poszkodowanych użytkowników i 771 witryn w okresie od 14 do 17 kwietnia.

Minimalna utrata danych po przywróceniu usuniętych witryn

Kopie zapasowe naszych baz danych tworzymy z wykorzystaniem kombinacji pełnych oraz przyrostowych kopii zapasowych, dzięki czemu możemy przywracać nasze magazyny danych do dowolnego konkretnego „punktu w czasie" przypadającego w okresie przechowywania kopii zapasowych (30 dni). W związku z tym incydentem w przypadku większości klientów ustaliliśmy główne magazyny danych produktów i zdecydowaliśmy się przyjąć za punkt przywrócenia okres pięciu minut poprzedzających usunięcie witryn jako bezpieczny punkt synchronizacji. Magazyny danych inne niż główne zostały przywrócone do tego samego punktu w czasie lub poprzez odtworzenie zarejestrowanych zdarzeń. Użycie stałego punktu przywracania dla głównych magazynów pozwoliło nam uzyskać spójność danych we wszystkich magazynach danych.

W przypadku 57 klientów przywróconych na wczesnym etapie naszego procesu reagowania na incydenty brak spójnych zasad oraz ręczne odzyskiwanie z migawek kopii zapasowych baz danych spowodowały przywrócenie niektórych baz danych Confluence i Insight do punktu przypadającego *wcześniej* niż pięć minut przed usunięciem witryny. Brak spójności został wykryty podczas audytu przeprowadzanego po przywróceniu. Od tamtego momentu odzyskaliśmy pozostałą część danych, skontaktowaliśmy się z poszkodowanymi klientami i pomagamy tym klientom zastosować zmiany w celu dalszego przywrócenia ich danych.

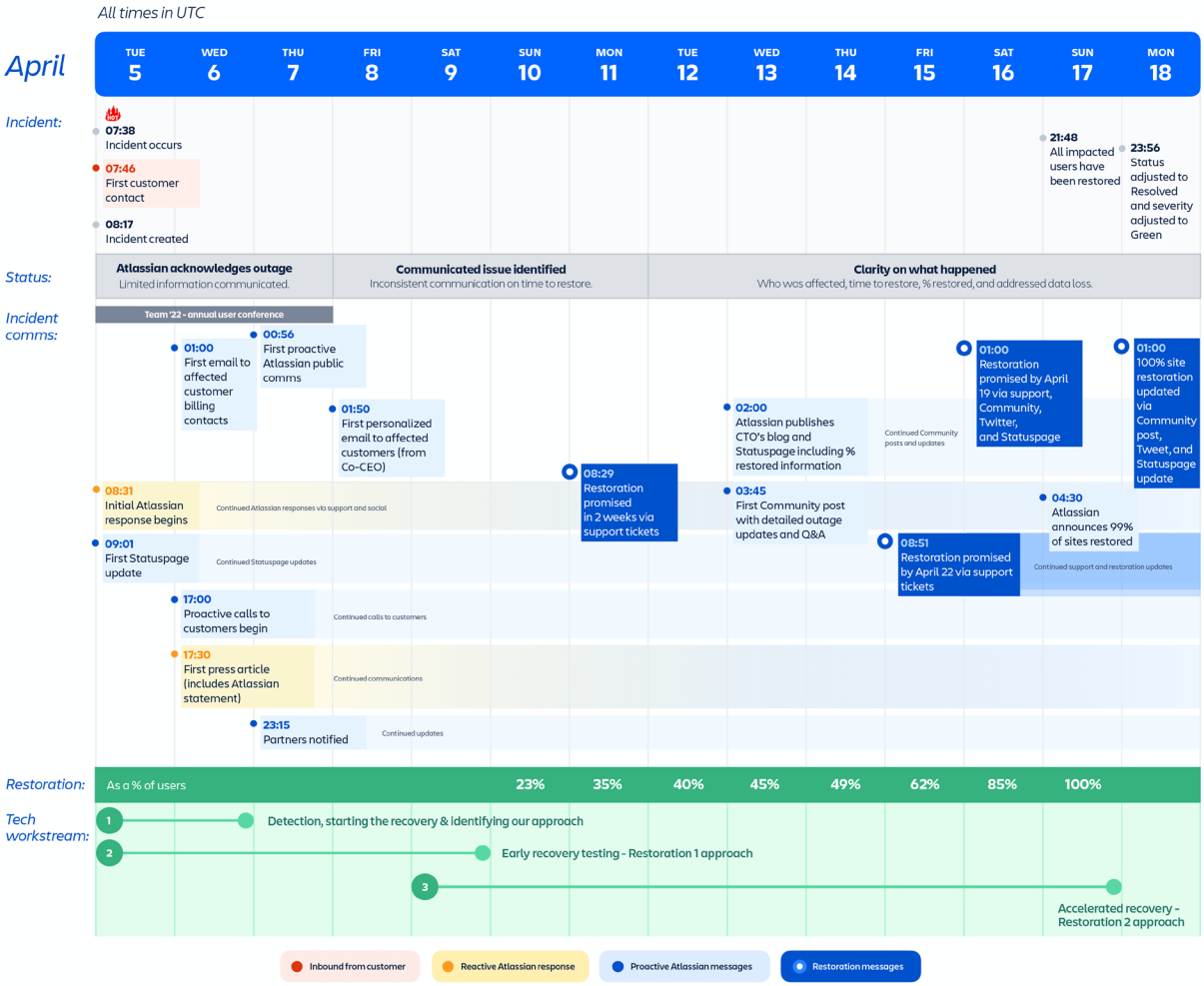
Podsumowanie:

- Podczas tego incydentu udało nam się osiągnąć założony docelowy punkt odzyskiwania (RPO) wynoszący jedną godzinę.
- Utrata danych w wyniku incydentu została ograniczona do pięciu minut poprzedzających usunięcie witryny.
- W przypadku niewielkiej liczby klientów bazy danych Confluence i Insight zostały przywrócone do stanu wcześniejszego niż pięć minut przed usunięciem, jednak udało nam się odzyskać dane i obecnie współpracujemy z klientami nad przywróceniem tych danych.

Komunikacja podczas incydentu

Mówiąc o komunikacji w sprawie incydentu, mamy na myśli kontakt z klientami, partnerami, mediami, analitykami branżowymi, inwestorami oraz szeroko rozumianą społecznością technologiczną.

Co się stało



Rysunek 9: Oś czasu kluczowych kamieni milowych komunikacji podczas incydentu.

Znacznik czasu: dni 1–3 (5–7 kwietnia)

Wczesna reakcja

Pierwsze zgłoszenie do działu wsparcia zostało utworzone 5 kwietnia o 7:46 czasu UTC, a dział wsparcia Atlassian odpowiedział, potwierdzając incydent do 8:31 czasu UTC. O 9:03 czasu UTC opublikowano pierwszą aktualizację w witrynie Statuspage informującą klientów o trwającej analizie incydentu. O 11:13 czasu UTC potwierdziliśmy w witrynie Statuspage, że zidentyfikowaliśmy przyczynę źródłową i pracujemy nad usunięciem problemu. Do godziny 1:00 czasu UTC 6 kwietnia przekazano klientom początkowy komunikat poprzez zgłoszenie z informacją, że awaria wynikała ze skryptu konserwacyjnego i spodziewamy się minimalnej utraty danych. W odpowiedzi na pytania mediów firma Atlassian opublikowała oświadczenie 6 kwietnia o 17:30 czasu UTC. 7 kwietnia o 0:56 czasu UTC firma Atlassian opublikowała na Twitterze pierwszy zewnętrzny komunikat do szerokiego grona odbiorców z informacją o potwierdzeniu incydentu.

Znacznik czasu: dni 4–7 (8–11 kwietnia)

Rozpoczęcie spersonalizowanej kampanii komunikacyjnej o szerszym zasięgu

8 kwietnia o 1:50 czasu UTC firma Atlassian wysłała e-mail do poszkodowanych klientów z przeprosinami od współzałożyciela i dyrektora generalnego, Scotta Farquhara. W następnych dniach pracowaliśmy nad przywróceniem usuniętych danych kontaktowych i utworzeniem zgłoszeń do działu wsparcia dla wszystkich dotkniętych incydentem witryn, w przypadku których takie zgłoszenie jeszcze nie powstało. Następnie za pośrednictwem zgłoszeń do działu wsparcia powiązanych z każdą dotkniętą witryną nasz zespół wsparcia wysyłał regularne aktualizacje dotyczące procesu przywracania.

Znacznik czasu: dni 8–14 (12–18 kwietnia)

Większa przejrzystość i ukończenie przywrócenia

12 kwietnia firma Atlassian opublikowała komunikat od dyrektora ds. technicznych, Sri Viswanatha, podając więcej szczegółów technicznych na temat zaistniałej sytuacji, poszkodowanych klientów, utraty danych oraz naszych postępów w przywracaniu, informując dodatkowo, że pełne przywrócenie wszystkich witryn może potrwać do dwóch tygodni. Oprócz wpisu na blogu pojawił się również komunikat prasowy od Sri. Dodaliśmy również odniesienie do bloga Sri w naszym pierwszym proaktywnym wpisie w społeczności Atlassian od kierownika działu inżynierskiego, Stephena Deasy'ego. Ten wpis stał się następnie specjalnym miejscem, w którym zamieszczane były dodatkowe aktualizacje oraz pytania i odpowiedzi dostępne dla szerszego grona odbiorców. 18 kwietnia wpis zaktualizowano o komunikat o pełnym przywróceniu wszystkich witryn poszkodowanych klientów.

Dlaczego nie udzieliliśmy odpowiedzi publicznej wcześniej?

1. Priorytetem była dla nas bezpośrednia komunikacja z poszkodowanymi klientami za pośrednictwem witryny Statuspage, poczty elektronicznej, zgłoszeń do działu wsparcia i bezpośrednich interakcji. Nie byliśmy jednak w stanie dotrzeć do wielu klientów, ponieważ utraciliśmy ich dane kontaktowe po usunięciu witryn. Powinniśmy postawić na szerszą komunikację znacznie wcześniej, aby poinformować poszkodowanych klientów i użytkowników końcowych o naszej reakcji na incydent oraz przedstawić oś czasu rozwiązywania.

2. Choć od razu wiedzieliśmy, co doprowadziło do incydentu, złożoność architektoniczna i wyjątkowe okoliczności tego incydentu spowolniły naszą zdolność do szybkiego zdefiniowania zakresu i precyzyjnego oszacowania czasu potrzebnego do rozwiązania. Zamiast czekać, aż zyskamy pełny obraz sytuacji, powinniśmy wyraźnie poinformować o tym, co wiemy, a czego jeszcze nie. Podanie przybliżonych czasów przywrócenia (nawet orientacyjnych) oraz wyraźne zakomunikowanie, kiedy spodziewamy się uzyskać bardziej kompletny obraz całej sytuacji, pozwoliłoby naszym klientom lepiej skorygować własne plany w związku z wystąpieniem incydentu. W szczególności dotyczy to administratorów systemu i osób kontaktowych ds. technicznych, które znajdują się na pierwszej linii zarządzania interesariuszami i użytkownikami w swoich organizacjach.

Wsparcie i kontakt z klientem

Jak już wcześniej wspomniano, skrypt, który spowodował usunięcie witryn klientów, doprowadził również do usunięcia kluczowych identyfikatorów i danych kontaktowych tych klientów (np. adresu URL witryny Cloud, danych kontaktowych administratora systemu w witrynie) z naszych środowisk produkcyjnych. Jest to niezwykle istotne, ponieważ nasze podstawowe systemy (np. wsparcia, licencjonowania, rozliczeń) wykorzystują istniejący adres URL witryny Cloud oraz dane kontaktowe administratora systemu w witrynie jako główne identyfikatory do celów związanych z bezpieczeństwem, przekierowywaniem i priorytetyzacją. Po utraceniu tych identyfikatorów początkowo straciliśmy możliwość systematycznego identyfikowania klientów i nawiązywania z nimi kontaktu.

Wpływ na wsparcie udzielane naszym klientom

Po pierwsze, większość poszkodowanych klientów nie mogła skontaktować się z naszym zespołem wsparcia za pośrednictwem zwykłego internetowego formularza kontaktowego. Ten formularz został zaprojektowany w taki sposób, aby wymagać od użytkownika zalogowania się przy użyciu identyfikatora Atlassian i podania prawidłowego adresu URL witryny Cloud. Bez podania prawidłowego adresu URL użytkownik nie może przesłać zgłoszenia do działu wsparcia. W trakcie normalnej działalności ta weryfikacja jest celowa w celu zapewnienia bezpieczeństwa w witrynie i umożliwienia klasyfikacji zgłoszeń. Jednak nieprzewidzianym skutkiem tego wymagania było uniemożliwienie klientom poszkodowanym w przypadku tej awarii przesłanie do działu wsparcia zgłoszenia dotyczącego witryny o wysokim priorytecie.

Po drugie, usunięcie danych administratora systemu w witrynie w wyniku tego incydentu utrudniło nam proaktywne nawiązanie kontaktu z poszkodowanymi klientami. W ciągu pierwszych kilku dni od momentu wystąpienia incydentu wysyłaliśmy proaktywne komunikaty do osób kontaktowych ds. rozliczeniowych i technicznych poszkodowanych klientów zarejestrowanych w Atlassian. Szybko jednak okazało się, że wiele danych osób kontaktowych ds. rozliczeniowych i technicznych poszkodowanych klientów było nieaktualne. Bez danych administratora systemu każdej witryny nie dysponowaliśmy kompletną listą aktywnych i potwierdzonych danych kontaktowych, które moglibyśmy wykorzystać do nawiązania kontaktu.

Nasza reakcja

Przed naszymi zespołami wsparcia postawiono trzy równie ważne priorytety w celu przyspieszenia przywracania witryn i wyeliminowania luki w naszych kanałach komunikacyjnych w ciągu pierwszych dni od wystąpienia incydentu.

Po pierwsze, uzyskanie wiarygodnej listy zweryfikowanych osób kontaktowych po stronie klientów. Nasze zespoły inżynierskie pracowały nad przywróceniem witryn klientów, więc zespoły odpowiedzialne za kontakt z klientem skoncentrowały się na odzyskaniu zweryfikowanych danych kontaktowych. Wykorzystaliśmy każdy mechanizm, jakim dysponujemy (systemy rozliczeniowe, wcześniejsze zgłoszenia do działu wsparcia, inne zabezpieczone kopie zapasowe użytkowników, bezpośredni kontakt z klientami itp.), aby odbudować naszą listę kontaktów. Naszym celem było utworzenie jednego zgłoszenia do działu wsparcia powiązanego z incydentem dla każdej dotkniętej witryny w celu usprawnienia bezpośredniego kontaktu z klientami i skrócenia czasów reakcji.

Po drugie, ponowne ustanowienie przepływów pracy, kolejek i umów SLA właściwych dla tego incydentu. Usunięcie identyfikatora Cloud i brak możliwości prawidłowego uwierzytelniania użytkowników wpłynęły również na naszą możliwość przetwarzania zgłoszeń do działu wsparcia powiązanych z incydentem za pośrednictwem naszych

standardowych systemów. Zgłoszenia nie pojawiały się poprawnie w kolejkach i na pulpitach odpowiednich priorytetów i eskalacji. Szybko powołaliśmy zespół interdyscyplinarny (z udziałem pracowników działu wsparcia, produktowego i IT), który zaprojektował i dodał nową logikę, umowy SLA, statusy przepływu pracy i pulpity. Te zmiany należało wprowadzić w naszym systemie produkcyjnym, dlatego ich pełne opracowanie, przetestowanie i wdrożenie zajęło kilka dni.

Po trzecie, masowe skalowanie ręcznych weryfikacji w celu przyspieszenia przywracania witryn. W miarę jak inżynierowie robili postępy w zakresie początkowego przywracania, stało się jasne, że do przyspieszenia procesu przywracania witryn konieczne będzie wykorzystanie potencjału wykonawczego naszych globalnych zespołów wsparcia, które zajęły się ręcznym testowaniem i sprawdzaniem poprawności. Ten proces weryfikacji nabrał krytycznego znaczenia w toku przywracania witryn naszym klientom, gdy zespół inżynierski zdołał przyspieszyć proces przywracania danych. Musieliśmy opracować niezależny strumień standardowych procedur operacyjnych (SOP), przepływów pracy, przekazywania zadań i grafików pracowników w celu zmobilizowania ponad 450 inżynierów wsparcia. Przeprowadzali oni sprawdzenia poprawności w trybie zmianowym przez całą dobę, siedem dni w tygodniu, aby przywrócone dane mogły szybciej trafić w ręce klientów.

Jednak mimo ustalenia tych kluczowych priorytetów już pod koniec pierwszego tygodnia, nasze możliwości przekazywania *znaczących* aktualizacji były ograniczone ze względu na niejasny przebieg osi czasu rozwiązywania incydentu, co było skutkiem złożoności procesów przywracania. Powinniśmy szybciej poinformować o braku pewności co do daty przywrócenia witryn i wcześniej umożliwić osobisty kontakt z nami, aby nasi klienci mogli odpowiednio zweryfikować swoje plany.

Wprowadzane ulepszenia

Natychmiast zablokowaliśmy zbiorcze usuwanie witryn do czasu wprowadzenia odpowiednich zmian.

W związku z tym incydentem i ponowną oceną naszych wewnętrznych procesów chcemy zauważyć, że to nie ludzie są przyczyną incydentów. To raczej systemy dopuszczają możliwości pomyłki. W tej części podsumowujemy czynników, które przyczyniły się do wystąpienia tego incydentu. Omawiamy również nasze plany przyspieszenia sposobów eliminowania tych słabości i problemów.

Wniosek 1: We wszystkich systemach stosować uniwersalnie „usuwanie nietrwałe”

Zasadniczo usuwanie tego rodzaju powinno być zabronione lub zabezpieczone wielowarstwowo w celu uniknięcia błędów. Podstawowym ulepszeniem, które

wprowadzamy, jest globalne uniemożliwienie usunięcia danych i metadanych klientów, które wcześniej nie zostały poddane procesowi nietrwałego usuwania.

a) Usuwanie danych powinno być wyłącznie usuwaniem nietrwałym

Usuwanie całej witryny powinno być zabronione, a usuwanie nietrwałe powinno wymagać wielopoziomowych zabezpieczeń, aby zapobiec błędom. Wdrożymy zasady „nietrwałego usuwania" uniemożliwiające zewnętrznym skryptom lub systemom usuwanie danych klientów w środowisku produkcyjnym. Nasze zasady „nietrwałego usuwania" będą pozwalać na wystarczająco długie przechowywanie danych, aby można było je szybko i bezpiecznie odzyskać. Dane zostaną usunięte ze środowiska produkcyjnego dopiero po upływie okresu przechowywania.

Działania:

- **Wdrożenie „nietrwałego usuwania" w przepływach pracy aprowizacji i wszystkich odpowiednich magazynach danych:** dodatkowo, zespół ds. platformy dzierżaw sprawdzi, czy usuwanie danych może nastąpić tylko po dezaktywacji, a także zapewni inne zabezpieczenia w tym zakresie. W dłuższej perspektywie zespół ds. platformy dzierżaw będzie odgrywał kluczową rolę w dalszym rozwoju zarządzania prawidłowym stanem danych dzierżaw.

b) Usuwanie nietrwałe powinno być objęte znormalizowanym i zweryfikowanym procesem przeglądu

Usuwanie nietrwałe należy do operacji wysokiego ryzyka. W związku z tym powinniśmy wprowadzić znormalizowane lub zautomatyzowane procesy przeglądu obejmujące zdefiniowane procedury wycofywania i testowania tych operacji.

Działania:

- **Wymuszone etapowe wprowadzanie wszelkich czynności nietrwałego usuwania:** wszystkie nowe operacje wymagające usuwania będą najpierw testowane w naszych własnych witrynach w celu zweryfikowania przyjętego podejścia oraz reguł automatyzacji. Po takim sprawdzeniu będziemy stopniowo wdrażać ten sam proces u klientów, kontynuując testy pod kątem nieprawidłowości, zanim zastosujemy automatyzację do całej bazy wybranych użytkowników.

- **Przetestowany plan wycofywania czynności nietrwałego usuwania:** każde działanie obejmujące nietrwałe usuwanie danych będzie wymagało uprzedniego przetestowania przywracania usuniętych danych przed wdrożeniem w środowisku produkcyjnym oraz przetestowanego planu wycofywania.

Wniosek 2: W ramach programu odzyskiwania awaryjnego zautomatyzować przywracanie po zdarzeniach usunięcia wielu witryn/produktów dla większych grup klientów

Dokument Zarządzanie danymi w Atlassian zawiera szczegółowy opis naszych procesów zarządzania danymi. W celu zapewnienia wysokiej dostępności aprowizujemy i utrzymujemy synchroniczną replikę rezerwową w wielu strefach dostępności (AZ) AWS. Przełączenie awaryjne strefy AZ odbywa się automatycznie i zazwyczaj trwa 60–120 sekund, a my zwykle radzimy sobie z awariami centrów danych oraz innymi typowymi zakłóceniami w sposób nieodczuwalny dla klienta.

Przechowujemy również niezmienialne kopie zapasowe zaprojektowane w taki sposób, aby były odporne na zdarzenia uszkodzenia danych, umożliwiając przywrócenie do poprzedniego punktu w czasie. Kopie zapasowe są przechowywane przez 30 dni, a Atlassian stale testuje i kontroluje kopie zapasowe pamięci masowej pod kątem możliwości przywrócenia. W razie potrzeby jesteśmy w stanie przywrócić wszystkich klientów w nowym środowisku.

Korzystając z tych kopii zapasowych, regularnie przywracamy przypadkowo usunięte dane pojedynczych klientów lub małych grup klientów. Jednak w przypadku usunięcia na poziomie witryny nie mieliśmy wykazów procedur, które dałoby się szybko zautomatyzować, biorąc pod uwagę skalę tego zdarzenia. Wymagało ono skoordynowanego zastosowania narzędzi i automatyzacji obejmujących wszystkie produkty i usługi.

Tym, czego (jeszcze) nie zautomatyzowaliśmy, jest przywracanie dużego podzbioru klientów w naszym istniejącym (i aktualnie używanym) środowisku bez wpływu na innych naszych klientów.

W naszym środowisku chmurowym każdy magazyn danych zawiera dane pochodzące od wielu klientów. Dane usunięte w trakcie tego incydentu stanowiły jedynie część magazynów danych, z których inni klienci nadal korzystali, dlatego musieliśmy ręcznie wyodrębniać i przywracać poszczególne elementy z naszych kopii zapasowych. Przywracanie danych każdego klienta jest procesem długotrwałym i złożonym, wymagającym wewnętrznego sprawdzania poprawności i ostatecznej weryfikacji przez klienta po przywróceniu witryny.

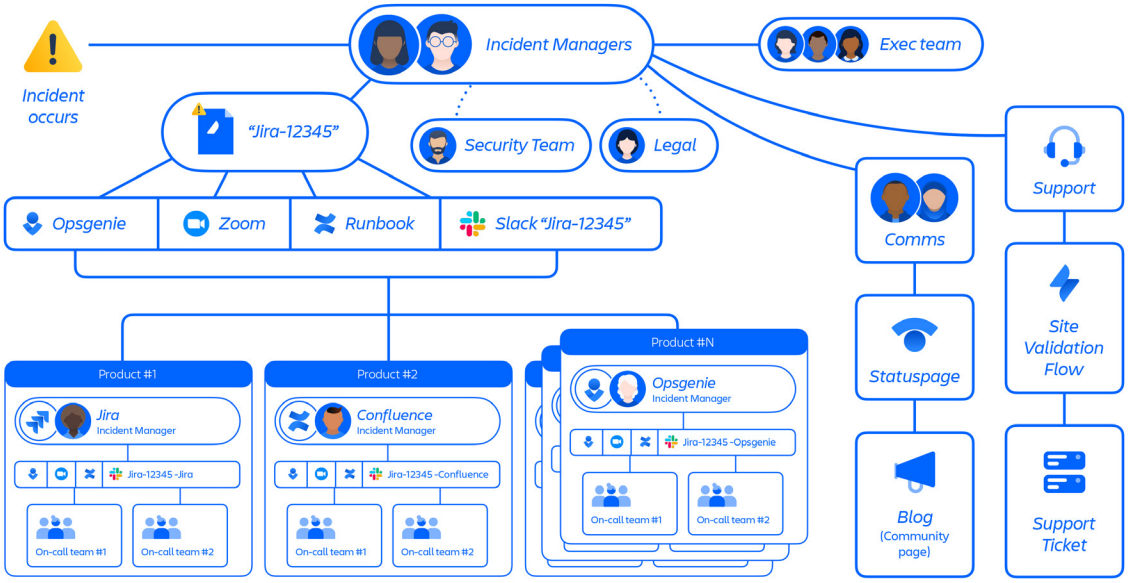
Działania:
Przyspieszenie przywracania wielu produktów i witryn dla większych grup klientów: program odzyskiwania awaryjnego spełnia nasze aktualne normy docelowego punktu odzyskiwania (RPO) wynoszące jedną godzinę. Wykorzystamy automatyzację oraz wnioski płynące z tego incydentu w celu przyspieszenia realizacji programu odzyskiwania awaryjnego, aby wartość docelowego czasu odzyskiwania (RTO) określona w naszych zasadach była osiągana również w przypadku incydentu na taką skalę.

Automatyzacja i dodanie weryfikacji tego przypadku w testach odzyskiwania awaryjnego: będziemy regularnie przeprowadzać ćwiczenia odzyskiwania awaryjnego z uwzględnieniem przywracania wszystkich produktów dla dużego zbioru witryn. Te testy odzyskiwania awaryjnego pozwolą zweryfikować, czy wykazy procedur pozostają aktualne w miarę rozwoju naszej architektury i pojawiania się nowych przypadków skrajnych. Będziemy stale doskonalili nasze podejście do przywracania w celu zautomatyzowania większej części procesu przywracania oraz skrócenia czasu odzyskiwania.

Wniosek 3: Poprawić proces zarządzania incydentami w przypadku zdarzeń o dużej skali

Nasz program zarządzania incydentami doskonale sprawdza się w zarządzaniu poważnymi i drobnymi incydentami, które miały miejsce na przestrzeni lat. Często symulujemy reakcje na incydenty na mniejszą skalę i krótszym czasie trwania, które zazwyczaj obejmują mniej osób i zespołów.

Jednak w okresie szczytowym wskutek tego incydentu nad przywróceniem witryn klientów równocześnie pracowały setki inżynierów i pracowników działu obsługi klienta. Opracowując nasz program zarządzania incydentami i struktury zespołów, nie byliśmy przygotowani na taką skalę oraz taki zakres i czas trwania incydentu, z jakimi mieliśmy do czynienia w tym przypadku (patrz *Rysunek 10* poniżej).



Rysunek 10: Przegląd procesu zarządzania incydentami o dużej skali.

Nasz proces zarządzania incydentami o dużej skali zostanie lepiej zdefiniowany i będzie często ćwiczony.

Dysponujemy poradami strategicznymi na wypadek incydentów na poziomie produktów, ale nie zdarzeń na taką skalę, nad którymi równocześnie pracują setki pracowników firmy. W narzędziach do zarządzania incydentami mamy reguły automatyzacji, które pozwalają tworzyć strumienie komunikacji obejmujące aplikacje, takie jak Slack i Zoom, oraz dokumenty Confluence. Nie uwzględniają one jednak tworzenia strumieni podrzędnych koniecznych w przypadku incydentów o dużej skali w celu wyizolowania strumieni dotyczących przywracania.

Działania:
Opracowanie porad strategicznych i narzędzi na wypadek incydentów o dużej skali oraz przeprowadzanie symulowanych ćwiczeń: zdefiniujemy i udokumentujemy typy incydentów, które można uznać za incydenty o dużej skali i które wymagają reakcji na takim właśnie poziomie. Nakreślimy kluczowe etapy koordynacji i opracujemy narzędzia, które pomogą menedżerom ds. zarządzania incydentami oraz innym pracownikom firmy usprawnić reagowanie i przystąpić do odzyskiwania. Menedżerowie ds. zarządzania incydentami wraz z zespołami będą regularnie przeprowadzać symulacje i szkolenia oraz ulepszać narzędzia i dokumenty w ramach ciągłego doskonalenia.

Wniosek 4: Poprawić nasze procesy komunikacji

a) Usunęliśmy krytyczne identyfikatory klientów, co wpłynęło na komunikację z osobami poszkodowanymi i podejmowanie działań w odniesieniu do nich
Skrypt, który spowodował usunięcie witryn klientów, doprowadził również do usunięcia kluczowych identyfikatorów klientów (np. adresu URL witryny, danych kontaktowych administratora systemu w witrynie) z naszych środowisk produkcyjnych. adres URL witryny, danych kontaktowych administratora systemu) z naszych środowisk produkcyjnych. W rezultacie (1) klienci nie mogli wysłać zgłoszenia do działu wsparcia przez standardowy kanał wsparcia; (2) uzyskanie wiarygodnej listy kluczowych osób kontaktowych po stronie klientów (np. administratorów systemu w witrynie) poszkodowanych w wyniku awarii w celu proaktywnego nawiązania z nimi kontaktu zajęło nam kilka dni; (3) początkowo przepływy pracy wsparcia, umowy SLA, pulpity i procedury eskalacji nie działały poprawnie ze względu na unikatowy charakter incydentu.

Podczas awarii eskalacje klientów napływały również wieloma kanałami (e-mailem, telefonicznie, poprzez zgłoszenia do dyrektora generalnego, za pośrednictwem LinkedIn i innych serwisów społecznościowych oraz poprzez zgłoszenia do działu wsparcia). Zróżnicowane narzędzia oraz procesy w naszych zespołach odpowiedzialnych za kontakt

z klientem spowolniły naszą reakcję i utrudniły śledzenie tych eskalacji oraz ich zgłaszanie.

b) Nie mieliśmy wystarczająco dokładnych porad strategicznych w zakresie komunikacji podczas incydentu, aby poradzić sobie z takim poziomem złożoności. Nie mieliśmy porad strategicznych w zakresie komunikacji podczas incydentu, które obejmowałyby zasady, a także role i obowiązki pozwalające zmobilizować wystarczająco szybko jednolity, interdyscyplinarny zespół ds. komunikacji podczas incydentu. Nie potwierdziliśmy incydentu wystarczająco szybko i spójnie za pośrednictwem wielu kanałów, zwłaszcza w mediach społecznościowych. Właściwym podejściem byłaby zakrojona na szerszą skalę, publiczna komunikacja dotycząca awarii, wraz z powtórzeniem kluczowego komunikatu, że nie doszło do utraty danych i nie był to wynik cyberataku.

Działania:

- **Ulepszenie tworzenia kopii zapasowej najważniejszych osób kontaktowych:** będziemy tworzyli kopię zapasową autoryzowanych osób kontaktowych przypisanych do poszczególnych kont poza instancją produktu.
- **Modernizacja narzędzi wsparcia:** utworzymy mechanizmy pozwalające klientom nieposiadającym adresu URL witryny ani identyfikatora Atlassian nawiązać bezpośredni kontakt z naszym zespołem wsparcia technicznego.
- **System i procesy eskalacji dla klientów:** zainwestujemy w ujednolicony system eskalacji oparty na kontach oraz powiązane z nim przepływy pracy, aby umożliwić przechowywanie wielu obiektów roboczych (zgłoszeń, zadań itp.) w ramach jednego obiektu konta klienta, co pozwoli usprawnić koordynację i poprawi widoczność we wszystkich zespołach mających kontakt z klientem.
- **Rozszerzenie zasięgu obsługi 24/7 działu zarządzania eskalacjami:** wdrożymy plany rozszerzenia globalnego zasięgu działu zarządzania eskalacjami, aby zapewnić stałą obsługę 24/7 z wyznaczonymi pracownikami w każdym z głównych regionów geograficznych. Ponadto w zespołach wsparcia wyznaczymy role w celu zapewnienia odpowiednich ekspertów w dziedzinie produktów i sprzedaży oraz przywództwa.
- **Aktualizacja porad strategicznych w zakresie komunikacji podczas incydentu z uwzględnieniem nowych wniosków oraz ich regularna weryfikacja:** przeprowadzimy weryfikację porad strategicznych w celu zdefiniowania wyraźnych wewnętrznych ról i kanałów komunikacyjnych. Wykorzystamy macierz DACI do obsługi incydentów i zapewnimy obsadę rezerwową 24/7 każdej roli na wypadek choroby, urlopu lub innych nieprzewidzianych zdarzeń. Co kwartał będziemy przeprowadzać audyt w celu potwierdzenia stałej gotowości.

Działania (cd.)
Będziemy stosować szablon komunikacji podczas incydentu we wszystkich komunikatach: będziemy informować o zaistniałym zdarzeniu, poszkodowanych podmiotach, osi czasu przywracania, procentowych postępach przywracania witryn, spodziewanej utracie danych, łącznie z powiązanymi poziomami ufności, oraz przedstawimy wyraźne wytyczne dotyczące kontaktowania się z działem wsparcia.

Uwagi końcowe

Choć awaria została usunięta, a dane klientów zostały w pełni przywrócone, kontynuujemy nasze działania. Na tym etapie wdrażamy opisane powyżej zmiany, aby usprawnić nasze procesy, zwiększyć odporność i nie dopuścić do ponownego wystąpienia podobnej sytuacji.

Atlassian jest organizacją nastawioną na naukę, a nasze zespoły z pewnością wyciągnęły z tej sytuacji wiele trudnych wniosków. Wprowadzamy te wnioski w życie, aby dokonać trwałych zmian w naszej działalności. Dzięki temu doświadczeniu staniemy się silniejsi i zapewnimy lepszą obsługę naszym klientom.

Mamy nadzieję, że wnioski z tego incydentu pomogą innym zespołom, które wkładają wiele wysiłku w zapewnienie swoim klientom dostępu do niezawodnych usług.

Na koniec chciałbym podziękować tym, którzy to czytają i uczą się razem z nami, a także tym, którzy są częścią naszej szerokiej społeczności i zespołu Atlassian.

– Sri Viswanath, dyrektor ds. technicznych

©2022 Atlassian. Wszelkie prawa zastrzeżone. ENTMIG-245 04/29/22