

## Krok 15: Wdrożenie

Tu omówiono następujące tematy:

- \* Rzeczy do rozważenia przy wdrażaniu aplikacji BI
- \* Dwa rodzaje wdrożeń zabezpieczeń: bezpieczeństwo scentralizowane i bezpieczeństwo zdecentralizowane
- \* Jak przeprowadzić analizę luk w zabezpieczeniach?
- \* Trzy rodzaje procedur tworzenia kopii zapasowych: przyrostowa kopia zapasowa, szybka kopia zapasowa na komputerze mainframe i częściowa kopia zapasowa
- \* Monitorowanie wykorzystania zasobów komputerowych, sieciowych i kadrowych
- \* Zarządzanie wzrostem ilości danych, wzrostem użycia i wzrostem sprzętu
- \* Krótkie opisy działań związanych z wdrożeniem, rezultatów wynikających z tych działań oraz zaangażowanych ról
- \* Ryzyko niewykonania kroku 15

Rzeczy do rozważenia

Przygotowanie do produkcji

- \* Czy zdefiniowaliśmy wszystkie biblioteki produkcyjne i produkcyjne bazy danych?
- \* Czy dzienne, tygodniowe i miesięczne procesy ekstrakcji/transformacji/ładowania (ETL) są w harmonogramie zadań?
- \* Czy regularnie zaplanowane programy raportów aplikacji są w harmonogramie zadań?
- \* Czy regularnie zaplanowane programy repozytoriów metadanych są na?

planista pracy?

- \* Czy personel operacyjny jest gotowy do przejęcia?
- \* Czy zatwierdzili wyniki testów zapewnienia jakości (QA)? Robić mają jakieś obawy?
- \* Czy musimy napisać procedury operacyjne dla personelu operacyjnego?
- \* Dla wszystkich komponentów aplikacji BI? A może tylko do procesu ETL?
- \* Kiedy skopiujemy wszystkie programy do bibliotek produkcyjnych?
- \* Kiedy załadujemy produkcyjne bazy danych?

Względy bezpieczeństwa

- \* Jakiego rodzaju zabezpieczenia są nam potrzebne? Co zabezpieczamy?
- \* Jak zabezpieczamy dane? Aplikacje? Narzędzia? Interfejsy?
- \* Czy środki bezpieczeństwa muszą obejmować szyfrowanie i deszyfrowanie, zwłaszcza w przypadku dostępu do Internetu i części analitycznej aplikacji BI?

\* Czy usługi uwierzytelniania pojedynczego użytkownika portalu informacyjnego przedsiębiorstwa są częścią tej aplikacji BI?

Utrzymanie bazy danych

\* Jaka jest nasza procedura tworzenia kopii zapasowych i odzyskiwania?

\* Jaka jest nasza procedura odzyskiwania po awarii?

\* Jak będzie przebiegać monitorowanie wydajności bazy danych? Jakie narzędzia będą użyte? Kto odpowiada za monitorowanie baz danych?

\* Skąd będziemy wiedzieć, czy spełniliśmy warunki umowy dotyczącej poziomu usług (SLA) na wydajność?

\* Jak będziemy monitorować wzrost wykorzystania i wzrost ilości danych?

Szkolenia i wsparcie

\* Czy przedsiębiorcy przeszli szkolenie z obsługi aplikacji BI?

\* Czy przedsiębiorcy zostali przeszkoleni w zakresie korzystania z repozytorium metadanych?

\* Czy „zaawansowani użytkownicy” przeszli szkolenie w zakresie pisania wydajnych zapytań w języku SQL (Structured Query Language)?

\* Czy pracownicy działu pomocy są przygotowani do mentorowania ludzi biznesu w korzystaniu z aplikacji BI, w tym repozytorium metadanych?

\* Czy personel help desku został odpowiednio przeszkolony?

\* Jeśli nie ma help desku, kto wesprze ludzi biznesu?

Teraz, gdy aplikacja BI jest zbudowana i przetestowana, jest gotowa do wdrożenia w środowisku produkcyjnym. Nową aplikację BI można wdrożyć na dwa sposoby, wszystkie naraz, tak jak ma to miejsce tradycyjnie, lub w przyrostach, jak krótko opisano w poniższej sekcji.

### **Wdrażanie przyrostowe**

Planując wdrożenie, zastosuj to samo iteracyjne podejście, które zastosowano przy tworzeniu aplikacji BI i repozytorium metadanych. Podejście iteracyjne, czyli incremental rollout, sprawdza się dobrze, ponieważ zmniejsza ryzyko ujawnienia potencjalnych defektów aplikacji BI na całą organizację. Ponadto daje możliwość nieformalnego zademonstrowania koncepcji BI i funkcji narzędzia BI ludziom biznesowym, którzy nie byli bezpośrednio zaangażowani w projekt BI. Oto parę sugestii.

\* Zaczynaj od małej grupy ludzi biznesu. Ta niewielka grupa powinna składać się nie tylko z „zaawansowanych użytkowników”, ale także z pracowników wiedzy i analityków biznesowych, którzy nie znają się na technologii, a także z głównego przedstawiciela biznesowego, który był zaangażowany w prace rozwojowe jako członek głównego zespołu.

\* Traktuj ludzi biznesu jak klientów – pamiętaj o dbałości o klienta. Bezproblemowe wdrożenie, interaktywne szkolenie i stałe wsparcie pomogą Ci uzyskać ich poparcie. Zawsze zadawaj sobie pytanie: „Co to ma dla klientów?”

\* Skorzystaj z okazji, aby przetestować swoje podejście do wdrażania. Możesz rozważyć dostosowanie swojego podejścia do implementacji lub zmodyfikowanie aplikacji BI przed pełnym wdrożeniem (np. zmiana uciążliwych procedur logowania).

\* Może być konieczne powielenie działań wdrożeniowych w wielu lokalizacjach. Powolne dodawanie tych witryn z czasem jest łatwiejsze niż uruchamianie ich wszystkich jednocześnie.

Ten przykład ilustruje, że wymagane środki bezpieczeństwa muszą być dobrze przemyślane, a funkcje bezpieczeństwa DBMS oraz narzędzi dostępu i analizy muszą być dobrze zrozumiane i przetestowane. Całkowite poleganie na jednym kompleksowym pakiecie zabezpieczeń, który ma możliwość wdrożenia wszystkich rodzajów zabezpieczeń, nie jest rozwiązaniem zabezpieczającym, ponieważ taki pakiet zabezpieczeń nie istnieje. Aby uzyskać potrzebne zabezpieczenia, najprawdopodobniej będziesz musiał wdrożyć szereg różnych środków bezpieczeństwa, w tym zakupione pakiety zabezpieczeń. Pamiętaj jednak, aby zminimalizować liczbę wdrażanych pakietów bezpieczeństwa, ponieważ może się zdarzyć jedna z dwóch rzeczy.

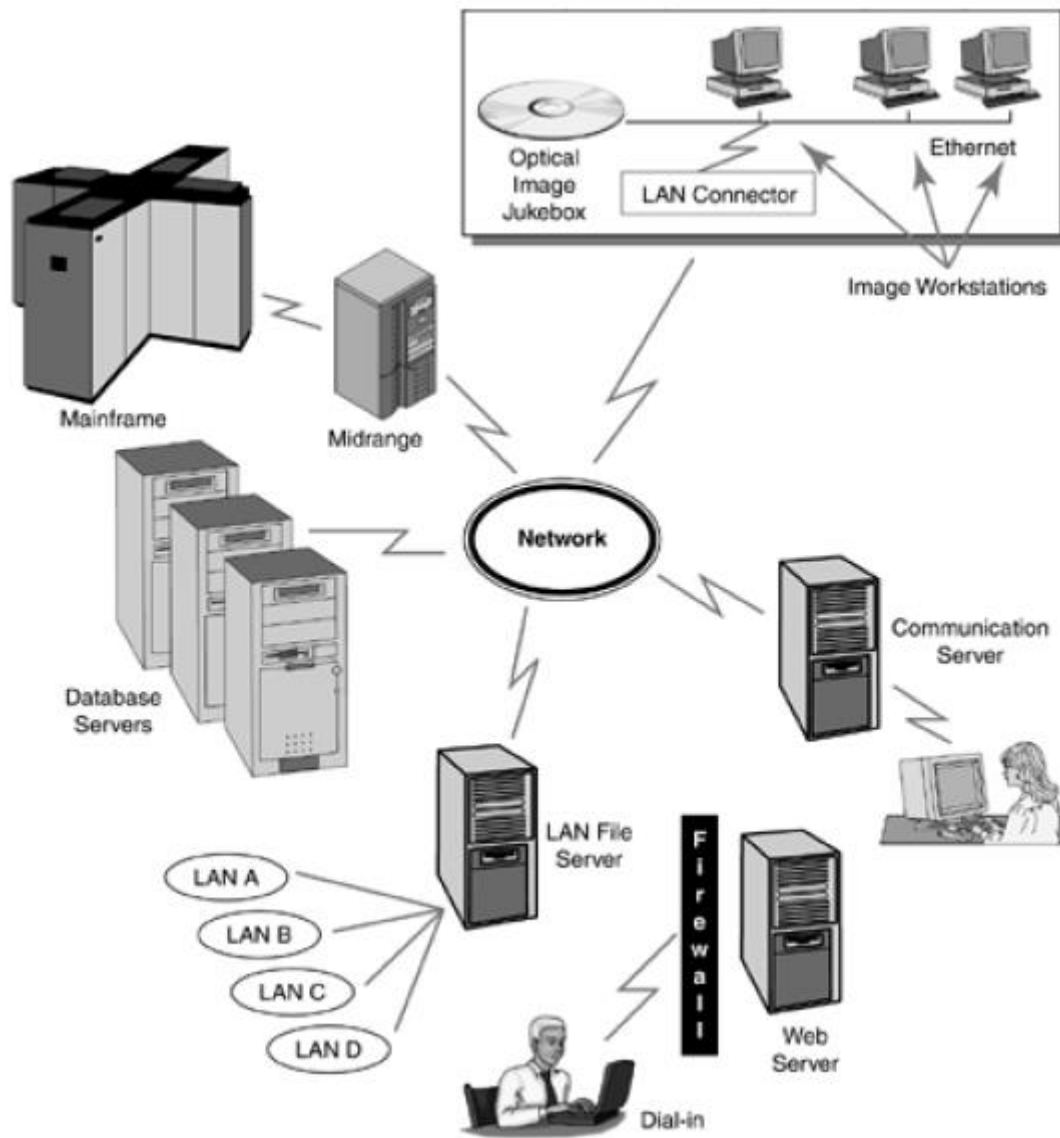
1. Przedsiębiorcy będą logować się za pomocą wielu pakietów bezpieczeństwa, używając wielu identyfikatorów logowania (ID) i wielu haseł, które wygasają w różnym czasie. Będą sfrustrowani bardzo szybko, jeśli będą musieli przejść przez różne procedury logowania i zapamiętać różne identyfikatory i hasła dla każdej procedury. Narastają skargi.

2. Ludzie biznesu przestaną całkowicie korzystać ze środowiska wspomagania decyzji BI, ponieważ jest ono zbyt uciążliwe. Nie chcesz, żeby tak się stało.

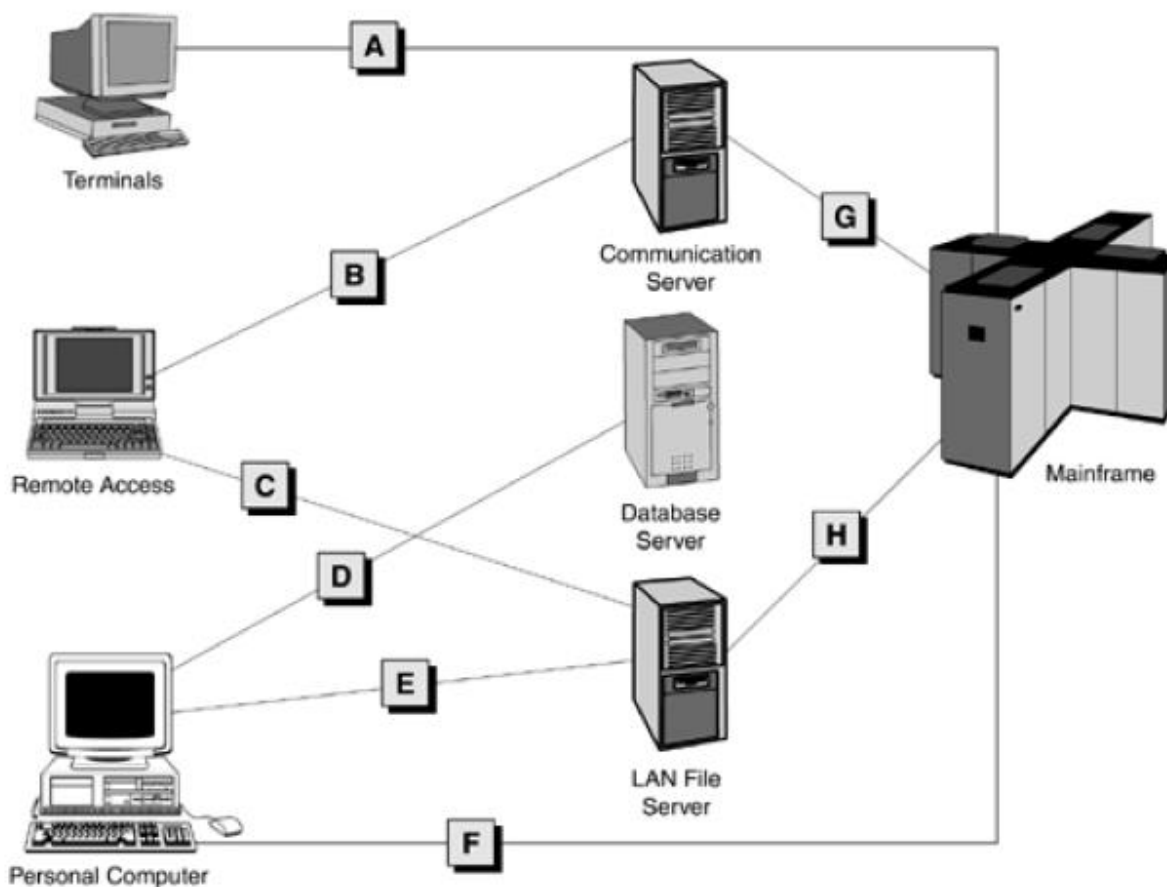
Wiele organizacji unika tego problemu, stosując schemat jednokrotnego logowania, który ogranicza poziom frustracji do minimum, ale nadal umożliwia śledzenie wszelkich naruszeń bezpieczeństwa, aczkolwiek w mniej wyrafinowany sposób.

### **Bezpieczeństwo w środowisku wielopoziomowym**

Wdrażanie środków bezpieczeństwa w scentralizowanym środowisku jest mniej skomplikowane niż w środowisku wielowarstwowym. W scentralizowanym środowisku wszystkie środki bezpieczeństwa można wdrożyć w jednym miejscu, ponieważ wszystkie dane znajdują się w jednym miejscu. Celem scentralizowanego bezpieczeństwa jest „jeden punkt wejścia, jeden strażnik”. O wiele łatwiej jest strzec pojedynczych drzwi niż wielu. W środowisku wspierającym podejmowanie decyzji BI przechowywanie wszystkich danych w jednym centralnym miejscu nie zawsze jest wykonalne lub pożądane. Jeśli dane muszą być przechowywane w sposób rozproszony w środowisku wielowarstwowym, wdrażanie środków bezpieczeństwa staje się znacznie bardziej skomplikowane. Poniższa lista zawiera krótki opis odpowiednich kroków. Zidentyfikuj punkty końcowe w architekturze sieci i ścieżki łączące punkty końcowe. Narysuj diagram swojej architektury fizycznej, podobny do rysunku



2. Określ ścieżki łączności (od punktów wejścia) używane do uzyskania dostępu do danych. Narysuj diagram z linkami i etykietami dla ścieżek łączności



3. Porównaj ścieżki z istniejącymi środkami bezpieczeństwa. Być może masz już zainstalowane pewne pakiety bezpieczeństwa, a niektóre z nich mogą wystarczyć do ochrony podzbioru danych. Narysuj macierz do celów analizy luk bezpieczeństwa

Connectivity Path	Mainframe Security Package	LAN Security Package	PC Security Package	Password Security	Encryption Function	DBMS Security	Generic Security Package
A							
B							
C							
D							
E							
F							
G							
H							

■ Security exists

□ No security

Macierz analizy luk w zabezpieczeniach pomoże Ci określić, gdzie zabezpieczenia są nadal potrzebne i jaki rodzaj zabezpieczeń jest potrzebny. Pamiętaj o następujących punktach:

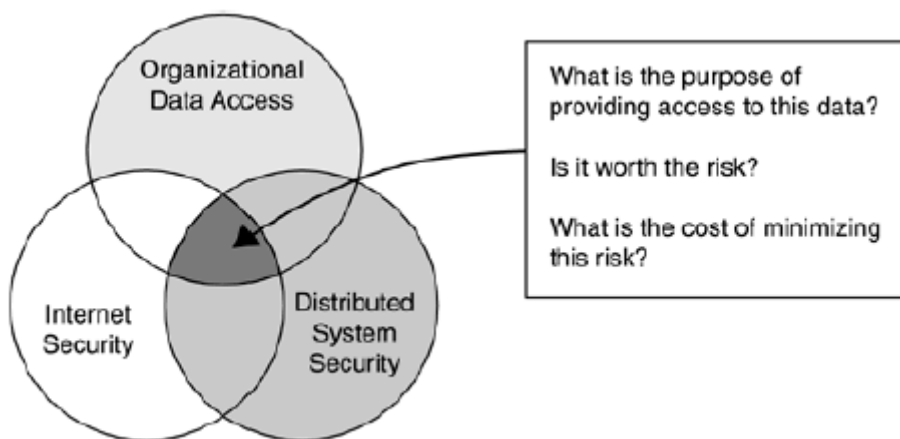
\* Zabezpieczenie hasłem może być najtańsze we wdrożeniu, ale można je łatwo złamać.

\* Bezpieczeństwo DBMS jest najważniejszym elementem rozwiązania zabezpieczającego i powinno mieć pierwszeństwo przed wszystkimi innymi środkami bezpieczeństwa, które mogą być sprzeczne z uprawnieniami dostępu do danych przyznawanymi przez DBMS.

\* Szyfrowanie nie jest tak rozpowszechnione w środowiskach wspierających decyzje BI ze względu na skomplikowane algorytmy szyfrowania i deszyfrowania. Procesy szyfrowania i deszyfrowania również znacznie obniżają wydajność. Jednak przy częstym wykorzystywaniu Internetu jako mechanizmu dostępu i dostarczania należy poważnie rozważyć szyfrowanie, aby chronić organizację przed kosztownymi naruszeniami bezpieczeństwa.

### Bezpieczeństwo dostępu do Internetu

Internet umożliwia dystrybucję informacji na całym świecie, a środowisko wspomaganie decyzji BI zapewnia łatwy dostęp do danych organizacyjnych. Połączenie tych dwóch możliwości wydaje się być ogromnym krokiem naprzód w angażowaniu się w e-commerce. Jednak dokładnie rozważ konsekwencje połączenia tych technologii, zanim zdecydujesz się podjąć ryzyko potencjalnego ujawnienia wrażliwych danych organizacyjnych.



Wielu dostawców produktów umożliwia ogólny dostęp do baz danych w sieci Web, a niektórzy dostawcy w szczególności umożliwiają dostęp do docelowych baz danych BI. To komplikuje obawy dotyczące:

\* Ogólne bezpieczeństwo środowiska wspomaganie decyzji BI

\* Kwestie bezpieczeństwa związane z umożliwieniem dostępu przez Internet do danych organizacji

Najważniejsze w kwestii bezpieczeństwa jest to, że musisz wcześniej zdefiniować swoje wymagania dotyczące bezpieczeństwa, aby mieć czas na rozważenie i rozważenie wszystkich czynników. Jeśli zdecydujesz się na wyświetlanie danych w sieci, poświęć dodatkowy czas i pieniądze na uwierzytelnianie i autoryzację personelu wewnętrznego i klientów zewnętrznych. Jeśli przesyłasz poufne dane do i od klientów zewnętrznych, rozważ zainwestowanie w oprogramowanie do szyfrowania i deszyfrowania.

\* Uwierzytelnianie to proces identyfikacji osoby, zwykle na podstawie identyfikatora logowania i hasła. Ten proces ma na celu upewnienie się, że osoba jest tym, za kogo się podaje.

\* Autoryzacja to proces przyznawania lub odmawiania osobie dostępu do zasobu, takiego jak aplikacja lub strona internetowa. W oprogramowaniu zabezpieczającym uwierzytelnianie różni się od

autoryzacji, a większość pakietów bezpieczeństwa implementuje dwuetapowy proces uwierzytelniania i autoryzacji.

\* Szyfrowanie to „tłumaczenie” danych na tajny kod. Jest to najskuteczniejszy sposób na osiągnięcie bezpieczeństwa danych. Aby odczytać zaszyfrowany plik, musisz mieć dostęp do tajnego klucza lub hasła, które umożliwiają jego odszyfrowanie. Dane niezaszyfrowane są zwykle określane jako zwykły tekst, podczas gdy dane zaszyfrowane są zwykle określane jako tekst zaszyfrowany.

### **Tworzenie kopii zapasowych i odzyskiwanie danych**

Po wydaniu kilku milionów dolarów na środowisko wspomagające podejmowanie decyzji BI, chcesz mieć pewność, że nigdy nie stracisz zawartości docelowych baz danych BI i nigdy nie zostaniesz pozbawiony możliwości analitycznych aplikacji BI przez długi okres czasu. Istnieje szkoła myślenia, która mówi: „Nie martw się o tworzenie kopii zapasowych docelowych baz danych BI, ponieważ dane te pochodzą z innych systemów — jeśli dane zostaną zniszczone, po prostu je odbuduj”. Jest to nieostrożne i kosztowne podejście w przypadku bardzo dużej bazy danych (VLDB). Chociaż tworzenie kopii zapasowej bazy danych jest czasochłonne i powoduje wyłączenie bazy danych na kilka godzin, alternatywa polegająca na ponownym załadowaniu danych z lat do VLDB zajmie znacznie więcej czasu — jeśli w ogóle można to zrobić. Nie każda organizacja decyduje się na przechowywanie wszystkich plików źródłowych przez lata, na wypadek konieczności ich ponownego przetworzenia. Regularne tworzenie kopii zapasowych docelowych baz danych BI jest obowiązkowe, ale sam rozmiar baz VLDB sprawia, że jest to wyzwanie technologiczne. Wiele platform sprzętowych, na których znajdują się aplikacje BI, często ma ograniczenia dotyczące ilości danych, które mogą być regularnie archiwizowane. Ograniczenia te wynikają z niskiej szybkości przesyłania danych między serwerem a urządzeniem do tworzenia kopii zapasowych. Dostępnych jest kilka strategii tworzenia kopii zapasowych, aby złagodzić ten problem.

\* Przyrostowa kopia zapasowa: Jedną ze strategii jest wykorzystanie aspektu tylko wzrostu docelowych baz danych BI (bez aktualizacji wierszy) poprzez tworzenie kopii zapasowych tylko rzeczywistych zmian w bazie danych (nowych wierszy) od ostatniej aktualizacji, a nie całej bazy danych. Ta przyrostowa („zmiana netto”) strategia tworzenia kopii zapasowych jest możliwa nawet w przypadku większości codziennych kopii zapasowych. Ponieważ jednak w środowisku wspomagania decyzji BI istnieje zwykle wiele baz danych, a dane podsumowane muszą pozostać zsynchronizowane z danymi szczegółowymi, żadne ładowanie ani odświeżanie nie może nastąpić do żadnej z tych baz danych, dopóki tworzenie kopii zapasowych wszystkich baz danych nie zostanie pomyślnie zakończone.

\* Szybka kopia zapasowa na komputerze mainframe: Inną możliwością jest użycie narzędzi do przesyłania danych na komputerach mainframe w celu przekazania danych BI z powrotem do komputera typu mainframe w celu utworzenia szybkiej kopii zapasowej, która jest obsługiwana tylko na komputerze mainframe. Łącza kanałowe na komputerze mainframe umożliwiają prędkości, do których nie można jeszcze zbliżyć się na większości serwerów klasy średniej. To drogie rozwiązanie, ale solidne, które zwykle działa.

\* Częściowa kopia zapasowa: Inna strategia polega na partycjonowaniu tabel bazy danych według daty w celu obsługi częściowych kopii zapasowych. Podczas tworzenia kopii zapasowej jednej partycji pozostałe partycje mogą pozostać dostępne. Rozważania dotyczące tej strategii są wymienione poniżej.

- Bazy danych, które obsługują równoległe tworzenie kopii zapasowych, mają dużą zaletę w tej strategii, ponieważ można tworzyć kopie zapasowe wielu partycji w tym samym czasie.

- Jeśli docelowe bazy danych BI są ładowane codziennie, zgrupuj wiele dni w jedną partycję, zamiast tworzyć nową partycję na każdy dzień. Podczas tworzenia kopii zapasowej dane ze wszystkich dni na partycji, której kopia zapasowa jest tworzona, nie byłyby dostępne.

- Dużą wadą tej strategii jest to, że jeśli tabela jest podzielona na partycje według kolumny daty do celów tworzenia kopii zapasowej (co oznacza, że jest pogrupowana według kolumny daty), nie można jej grupować w żaden inny sposób w celu uzyskania dostępu. Może to mieć wpływ na wydajność podczas uruchamiania raportów i zapytań, chyba że używany jest równoległość bazy danych.

### **Monitorowanie wykorzystania zasobów**

Należy stale monitorować wykorzystanie różnych zasobów w środowisku wspomagania decyzji BI, w szczególności wykorzystanie komputerów, sieci i personelu. Jeśli którykolwiek z tych zasobów zostanie zaniedbany, może potencjalnie stać się wąskim gardłem dla aplikacji BI.

### **Wykorzystanie komputera**

Wykorzystanie komputera obejmuje jednostkę centralną (CPU), kanały wejścia/wyjścia (I/O), pamięć o dostępie swobodnym (RAM), urządzenia pamięci masowej o dostępie bezpośrednim (DASD) lub napędy dysków oraz inny powiązany sprzęt. Urządzenia te powinny być dedykowane dla środowiska wspomagania decyzji BI i nie powinny być współdzielone przez inne aplikacje. Ponieważ wykorzystanie tych urządzeń wzrasta z czasem, należy je monitorować.

Niezwykle ważne jest, aby używać odpowiednich narzędzi monitorujących i alarmujących, które wykryją wszelkie problemy z zasobami i uruchomią alarm w przypadku rzeczywistej lub oczekującej awarii systemu lub niedoboru zasobów. Wybór odpowiedniego narzędzia do monitorowania jest szczególnie istotny w środowisku rozproszonym.

### **Wykorzystanie sieci**

Wykorzystanie sieci może być stosunkowo niewielkim problemem w codziennym wykonywaniu aplikacji BI, ale może być dużym problemem dla procesu ETL. Pewna duża organizacja telekomunikacyjna odkryła, że nawet mając dostęp do najnowszych i najlepszych technologii komunikacyjnych, jej przepustowość jest niewystarczająca do przesyłania danych źródłowych z komputera mainframe na serwer bazy danych w celu terminowego zapełnienia docelowych baz danych BI. Monitorując wykorzystanie sieci, weź pod uwagę również następujące punkty:

\* Przepustowość może stanowić poważny problem dla aplikacji BI, które często wymagają kilku poziomów dostępu do szczegółowych danych. Zapytania po zapytaniach mogą być wykonywane przez wiele osób z wielu lokalizacji, z których każda potencjalnie zwraca ogromne ilości danych w całej sieci.

\* Jeśli przepustowość jest ciągłym problemem, jednym z potencjalnych rozwiązań jest przejście w kierunku rozproszonej implementacji BI i odejście od scentralizowanej strategii. Jeśli jednak wymagana jest duża komunikacja i integracja pomiędzy rozproszonymi docelowymi bazami danych BI, rozwiązanie to może pogorszyć sytuację.

### **Wykorzystanie personelu**

Środowisko wspomagające podejmowanie decyzji BI to środowisko wymagające dużej konserwacji. Jest to szczególnie ważne podczas początkowego opracowywania i wdrażania kilku pierwszych aplikacji BI. Odnosi się to również do zakresu, w którym procesy ręczne są używane podczas lub po cyklach ETL do sprawdzania i analizowania statystyk obciążenia generowanych przez przebiegi ETL. Środowisko



wspomagające podejmowanie decyzji BI wymaga dedykowanego wsparcia szeregu pracowników IT, takich jak:

- \* Twórcy aplikacji
- \* Administratorzy danych
- \* Administratorzy baz danych
- \* Specjaliści od sprzętu i systemów operacyjnych
- \* Specjaliści od oprogramowania pośredniego
- \* Administratorzy sieci

Często trudno jest znaleźć dobrych techników. Wielu techników chce pracować nad aplikacjami BI, zwłaszcza wielowymiarowymi, ponieważ te wymagające aplikacje obejmują nowe technologie (a przez to dobrze wyglądają w CV). Jednakże, ponieważ systemy operacyjne nadal wymagają wsparcia – ponieważ są one bardziej krytyczne niż aplikacje BI – często wynikiem tego jest „przeciąganie liny” o dostępny personel. Dotyczy to zwłaszcza administratorów baz danych, którzy posiadają już pełną płytę obsługującą systemy operacyjne. Niektórzy starsi administratorzy baz danych muszą zostać zwolnieni z tych obowiązków, aby zostać głównymi projektantami i menedżerami docelowych baz danych BI. Ci administratorzy baz danych są odpowiedzialni za projektowanie, budowanie, monitorowanie, dostrajanie i utrzymywanie docelowych baz danych BI oraz, do pewnego stopnia, wywołania dostępu do bazy danych z programów aplikacji BI oraz narzędzi dostępu i analizy (np. pisanie zapytań przekazujących). Ponadto muszą zajmować się monitorowaniem i zarządzaniem rosnącymi ilościami danych oraz rosnącym wykorzystaniem bazy danych, co będzie wymagało jeszcze więcej czasu. Redystrybucja obciążenia administratorów baz danych to zmiana kulturowa i problem w wielu organizacjach, które postrzegają funkcję administratora baz danych jako nic innego jak uruchamianie procedur konserwacji DBMS na produkcyjnych bazach danych.

### **Zarządzanie wzrostem**

Według ostrożnych szacunków ilość danych w środowisku wspomagania decyzji BI podwaja się co dwa lata. Dobrą wiadomością jest to, że koszt zapytania w przypadku większości środowisk wspomagania decyzji BI spada wraz z odpowiednio zarządzanym wzrostem. Zła wiadomość jest taka, że ogólny koszt rośnie, przy założeniu, że coraz więcej ludzi biznesu korzysta ze środowiska wspomagania decyzji BI w miarę upływu czasu. Trzy kluczowe obszary rozwoju, które należy obserwować, to dane, wykorzystanie i sprzęt.

### **Wzrost danych**

Przyrost danych to nie tylko dodawanie nowych wierszy do tabel, ale także rozszerzanie docelowych baz danych BI o dodatkowe kolumny i nowe tabele. Dodawanie nowych kolumn do tabeli wymiarów nie jest tak skomplikowane, jak dodawanie nowych tabel wymiarów do istniejącego schematu gwiazdy lub płatka śniegu, co zwykle wymaga wykonania następujących czynności:

- \* Rozładowanie tabeli faktów
- \* Dodanie kolejnego klucza obcego do tabeli faktów w celu odniesienia się do nowej tabeli wymiarów
- \* Przeliczanie faktów na mniejszą szczegółowość (ze względu na nowy wymiar)
- \* Ponowne ładowanie tabeli faktów

Docelowe bazy danych BI wymagają dużej ilości miejsca na dysku, przy czym obszar roboczy i indeksy zajmują od 25 do 40 procent tej przestrzeni. W świecie relacyjnym dane stanowią tylko ułamek całkowitego rozmiaru bazy danych; główną jego częścią jest przestrzeń indeksu. Indeksowanie jest wymagane, aby zapewnić lepszy czas odpowiedzi, gdy odczytywane są ogromne ilości danych.

Podczas obliczania zapotrzebowania na miejsce rozsądnym może być zastosowanie standardowej maksymy inżynierskiej: oblicz, jak duże będą docelowe bazy danych BI (w tym indeksy), a następnie potroj te liczby.

Wraz ze wzrostem ilości danych musi istnieć plan agregowania i podsumowywania danych w miarę ich starzenia. Analitycy biznesowi rzadko wymagają takiego samego poziomu szczegółowości w przypadku bardzo starych danych, jak w przypadku danych najnowszych. Dlatego poziom szczegółowości powinien spadać wraz z ruchomym kalendarzem. Załóżmy na przykład, że ludzie biznesu chcą przechowywać dane historyczne z dziesięciu lat. Wymagają miesięcznych zestawień według departamentów za dwa lata, ale zadowolają się zestawieniami miesięcznymi według regionu za pozostałe osiem lat. Przed załadowaniem nowego miesiąca do docelowej bazy danych BI dane na poziomie działu z 24. miesiąca są podsumowywane w podsumowaniach regionalnych i przenoszone do innej tabeli faktów, tak że 23. miesiąc staje się 24., a 22. miesiąc staje się 23. miesiącem, i tak dalej. Poniższa lista zawiera niektóre z nowych technologii dostępnych do obsługi ogromnych ilości danych i możliwości analizy tych ogromnych baz danych:

- \* Technologie równoległe
- \* Wielowymiarowe bazy danych
- \* Nowe technologie indeksowania
- \* Narzędzia do relacyjnego przetwarzania analitycznego online (ROLAP)
- \* Rozproszone narzędzia i narzędzia do konserwacji baz danych

### **Wzrost użytkowania**

Kolejnym kluczowym obszarem wzrostu jest wykorzystanie. Organizacje, które zbudowały skuteczne aplikacje BI, często odkrywały stłumione zapotrzebowanie na informacje w całej organizacji. Ta potrzeba przekłada się na to, że coraz więcej ludzi biznesu korzysta z istniejących aplikacji BI i pyta o nowe. Liczba ludzi biznesu uzyskujących dostęp do docelowych baz danych BI może z łatwością podwoić się lub potroić każdego roku, co prowadzi do wykładniczego wzrostu użycia. Ponieważ różni ludzie biznesu chcą widzieć różne dane i patrzeć na nie na różne sposoby, chcą podzielić dane według nowych wymiarów biznesowych, co zwiększa ich objętość. Chociaż ilość danych jest o wiele bardziej krytycznym czynnikiem przy określaniu wymagań procesora, równie ważna jest liczba osób uzyskujących dostęp do docelowych baz danych BI. Technicy postrzegają wzrost użycia jako coś negatywnego. Menedżerowie uważają jednak wzrost użytkowania za coś pozytywnego, o ile istnieje zwrot z inwestycji (ROI). Zakup nowego sprzętu lub aktualizacja istniejącego sprzętu w celu obsługi wzrostu mogą nie stanowić problemu, jeśli organizacja osiąga znaczne zyski dzięki lepszym możliwościom podejmowania decyzji za pomocą aplikacji BI. Dlatego wzrost użycia może oznaczać, że strategia BI działa.

Docelowe bazy danych BI są z natury tylko do odczytu i tylko do wzrostu. Dlatego kluczem jest zaprzestanie prób oszczędzania miejsca na dysku, jeśli aplikacje BI i dane BI pomagają organizacji osiągnąć zysk.

### **Rozwój sprzętu**

Biorąc pod uwagę informacje o wzroście danych i wzroście użycia, powinno być oczywiste, że skalowalność architektury sprzętowej BI jest kluczowa. Ale zanim planujesz pięć lat naprzód, pamiętaj, że koszt sprzętu to tylko jedna część całkowitego kosztu BI. Spójrz na horyzont planowania od 12 do 24 miesięcy; najlepiej zacząć od małych, ale też plan długoterminowego wzrostu. Rozważ następujące czynniki.

\* Pamiętaj o progu wydajności Twojej platformy BI. Jeśli przekroczysz tę pojemność, musisz dodać więcej procesorów, kanałów we/wy, niezależnych kontrolerów dysków i innych szybkich komponentów, aby utrzymać środowisko wspomagania decyzji BI na akceptowalnym poziomie wydajności.

\* Ze wszystkich urządzeń BI najważniejsza jest platforma serwerowa BI. Przy zamawianiu nowego sprzętu dowolnego rodzaju, musi być wystarczająco dużo czasu na dostarczenie sprzętu, przetestowanie go i przygotowanie do środowiska programistycznego i produkcyjnego.

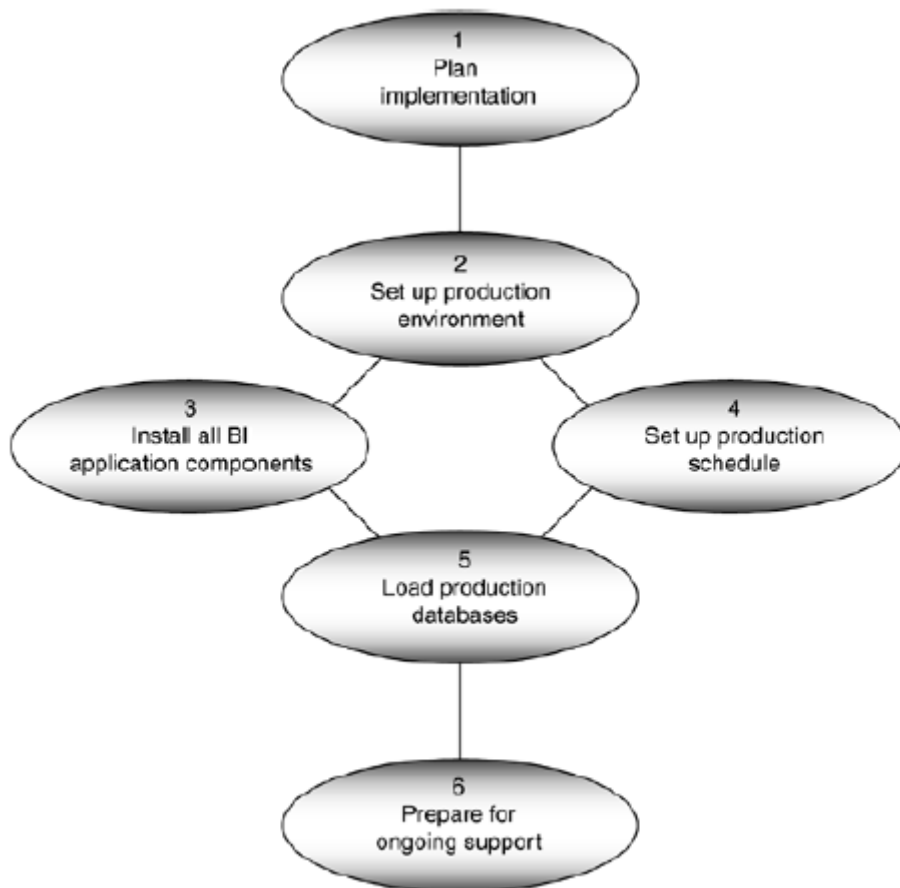
\* Technologia równoległa jest absolutną koniecznością dla VLDB. Możliwość przechowywania danych na dyskach rozłożonych oraz możliwość posiadania wielu niezależnych kontrolerów dysków odgrywają niezwykle ważną rolę w wydajności procesów działających na docelowych bazach danych BI.

\* Protokół kontroli transmisji/protokół internetowy (TCP/IP) jest odpowiedni dla większości platform sprzętowych. TCP/IP szybko staje się standardem w zakresie skalowalności, rozważań dotyczących rozwoju i środowisk wieloplatformowych.

\* Rozważ zalety podejścia do zbiorczej bazy danych z oddzielnymi docelowymi bazami danych BI. Takie podejście umożliwia skalowalność w mniejszych i tańszych przyrostach.

### **Działania wdrożeniowe**

Czynności do wdrożenia nie muszą być wykonywane liniowo. Rysunek 15.5 wskazuje, jakie czynności można wykonywać jednocześnie. Poniższa lista zawiera krótki opis działań związanych z Krokiem 15, Wdrażanie.



1. Zaplanuj wdrożenie. Ustaw datę wdrożenia i upewnij się, że wszystkie zasoby potrzebne do wdrożenia będą dostępne. W zależności od poczynionych postępów, wyciągniętych wniosków i napotkanych trudności, możesz chcieć wdrażać aplikację BI w społeczności biznesowej etapami. Zaczynij od małej grupy ludzi biznesu, ucz się z doświadczenia i w razie potrzeby zmodyfikuj swoje podejście (np. zwiększ czas szkolenia lub zmień zabezpieczenia) przed udostępnieniem aplikacji BI większej liczbie osób. Jeśli aplikacja BI ma jakikolwiek wpływ organizacyjny, przygotuj się do wprowadzenia tych zmian organizacyjnych (np. zmian usprawniających procesy biznesowe lub przesuniętych ról i obowiązków).

2. Skonfiguruj środowisko produkcyjne. W większości dużych organizacji przygotowanie środowiska produkcyjnego wymaga przestrzegania ścisłych procedur.

1. – Skonfiguruj biblioteki programu produkcyjnego (ETL, aplikacja, repozytorium metadanych).

2. – Utwórz produkcyjne bazy danych (docelowe bazy danych BI, baza repozytorium metadanych).

3. – Nadaj odpowiednie uprawnienia dostępu do produkcyjnych baz danych.

4. – Przyznaj programistom, personelowi operacyjnemu i ludziom biznesowym odpowiednie uprawnienia dostępu do wykonywania programów z bibliotek programów produkcyjnych.

5. – Napisz procedury operacyjne dla personelu operacyjnego wraz z instrukcjami uruchamiania procesu ETL, a także regularnie zaplanowane zadania raportowania aplikacji.

6. – Przygotuj poradnik dla pracowników help desku i ludzi biznesu z instrukcjami korzystania z aplikacji BI.

7. – Określ poziomy bezpieczeństwa produkcji dla wszystkich komponentów aplikacji BI.

3. Zainstaluj wszystkie składniki aplikacji BI. Przenieś wszystkie programy ETL, aplikacje i programy repozytorium metadanych do odpowiednich bibliotek produkcyjnych.
4. Skonfiguruj harmonogram produkcji. Wszystkie programy ETL, programy do raportowania aplikacji i programy repozytoriów metadanych, które będą uruchamiane regularnie, muszą być skonfigurowane w harmonogramie zadań. Harmonogram zadań ETL musi zawierać programy metadanych, które są częścią procesu ETL (np. przechwytywanie statystyk obciążenia, sum uzgadniania, współczynniki niezawodności danych).
5. Załaduj produkcyjne bazy danych. Załaduj docelowe bazy danych BI, uruchamiając proces wstępnego ładowania, a następnie proces ładowania historycznego. Załaduj również repozytorium metadanych metadanymi z różnych źródeł metadanych, takich jak arkusze kalkulacyjne, narzędzie do wspomaganego komputerowo inżynierii oprogramowania (CASE), narzędzie ETL i narzędzie do przetwarzania analitycznego online (OLAP).
6. Przygotuj się na stałe wsparcie. Ustal harmonogram pomocy w nagłych wypadkach na wezwanie. Zaplanuj regularne kopie zapasowe, a także okazjonalne reorganizacje baz danych dla wszystkich baz danych produkcyjnych. Zaplanuj użycie narzędzi dostarczonych przez system DBMS do tych czynności związanych z konserwacją bazy danych. Ponadto zaplanuj monitorowanie wydajności, wzrostu, użycia i jakości w ramach bieżących działań konserwacyjnych bazy danych. Okresowo przeglądaj i koryguj plany pojemności procesorów, pamięci dyskowej, sieci i przepustowości.

#### **Rezultaty wynikające z tych działań**

1. Produkcyjna biblioteka programów ETL . Wszystkie w pełni działające programy i skrypty ETL powinny znajdować się w produkcyjnej bibliotece programów ETL.
2. Produkcyjna biblioteka programów aplikacyjnych. Wszystkie w pełni działające programy dostępu i analizy oraz skrypty powinny znajdować się w produkcyjnej bibliotece programów aplikacji.
3. Biblioteka programowa repozytorium metadanych produkcyjnych. Wszystkie w pełni działające programy i skrypty repozytorium metadanych powinny znajdować się w bibliotece programu repozytorium metadanych.
4. Produkcyjne bazy danych docelowych BI. Instrukcje SQL języka definicji danych (DDL) i języka sterowania danymi (DCL) są uruchamiane w środowisku produkcyjnym w celu tworzenia produkcyjnych docelowych baz danych BI. Proces ładowania początkowego ETL i proces ładowania historycznego ETL są uruchamiane w celu zapełnienia tych docelowych baz danych produkcyjnych BI.
5. Baza danych repozytorium metadanych produkcyjnych. Instrukcje DDL i DCL dla repozytorium metadanych są uruchamiane w środowisku produkcyjnym w celu zbudowania produkcyjnej bazy danych repozytorium metadanych.

Programy do migracji metadanych i programy interfejsu narzędzi są uruchamiane w celu zapełnienia produkcyjnej bazy danych repozytorium metadanych metadanymi biznesowymi, metadanymi technicznymi i metadanymi ETL (statystyki obciążenia, sumy uzgadniania, metryki jakości danych) z ładowania początkowego i obciążenia historycznego procesy.

6. Dokumentacja produkcyjna. Dokumentacja produkcyjna aplikacji BI obejmuje:

- Procedury operacyjne dla personelu operacyjnego obejmujące proces ETL i wszystkie zaplanowane zadania raportowania aplikacji
- Poradnik dla pracowników help desku i ludzi biznesu z instrukcjami korzystania z aplikacji BI

## **Role zaangażowane w te działania**

\* Twórcy aplikacji. Twórcy aplikacji współpracują z personelem operacyjnym, aby przenieść programy raportów, skrypty zapytań, programy interfejsu i programy funkcji pomocy online do biblioteki programów aplikacji produkcyjnych.

\* Główny programista aplikacji. Główny programista aplikacji nadzoruje działania wdrożeniowe części dostępowej i analitycznej aplikacji BI. Jest odpowiedzialny za konfigurację biblioteki programów aplikacji produkcyjnych oraz pisanie części dotyczących dostępu i analizy procedur operacyjnych oraz przewodnika referencyjnego. Jest również odpowiedzialny za konfigurowanie programów raportów aplikacji w harmonogramie zadań.

\* Ekspert od eksploracji danych. Ekspert eksploracji danych współpracuje z administratorem bazy danych w celu tworzenia, korygowania i utrzymywania baz danych eksploracji danych dla planowanych działań eksploracji danych. Eksploracja danych to przedsięwzięcie iteracyjne i ad hoc, które wymaga ciągłych zmian w bazach danych eksploracji danych, analitycznych modelach danych i operacjach eksploracji danych.

\* Administrator bazy danych. Administrator bazy danych tworzy produkcyjne docelowe bazy danych BI oraz produkcyjną bazę danych repozytorium metadanych i przyznaje odpowiednie uprawnienia dostępu do tych produkcyjnych baz danych. Musi uruchomić proces ładowania początkowego i proces ładowania historycznego, aby załadować docelowe bazy danych BI. Planuje działania związane z utrzymaniem bazy danych (kopie zapasowe, reorganizacje) oraz bieżące monitorowanie bazy danych (pod kątem wydajności, wzrostu, wykorzystania). Przegląda również plany pojemności procesorów, pamięci dyskowej i przepustowości sieci.

\* Deweloperzy ETL. Programiści ETL współpracują z personelem operacyjnym, aby przenieść programy ETL do produkcyjnej biblioteki programów ETL.

\* Główny programista ETL. Główny programista ETL nadzoruje działania wdrożeniowe części ETL aplikacji BI. Współpracuje z personelem operacyjnym w celu przygotowania środowiska produkcyjnego i konfiguracji produkcyjnej biblioteki programów ETL. Powinien napisać część ETL procedur operacyjnych i przewodnik referencyjny. Jest również odpowiedzialny za konfigurację procesu ETL w harmonogramie zadań.

\* Administrator metadanych. Administrator metadanych jest odpowiedzialny za przeniesienie wszystkich programów repozytorium metadanych do biblioteki programów repozytorium metadanych produkcyjnych. Musi również uruchomić proces migracji (ładowania) metadanych i zaplanować bieżące działania monitorujące jakość danych, takie jak zbieranie metryk metadanych i wykonywanie wyrywkowych kontroli jakości danych.

\* Twórcy repozytorium danych meta. Twórcy repozytorium metadanych pomagają administratorowi metadanych w przenoszeniu wszystkich programów repozytorium metadanych do produkcyjnej biblioteki programów repozytorium metadanych.

\* Autorzy strony. Deweloperzy sieci Web są odpowiedzialni za przenoszenie swoich stron sieci Web i skryptów ze swoich serwerów lokalnych na produkcyjny serwer sieci Web.

\* Mistrz sieci. Webmaster odpowiada za skonfigurowanie produkcyjnego serwera WWW. Musi również współpracować z personelem służb bezpieczeństwa i usług sieciowych w celu zainstalowania i przetestowania zapory i innych wymaganych funkcji bezpieczeństwa.

### **Ryzyko niewykonania kroku 15**

Jeśli dotarłeś tak daleko, oczywiście wdrożysz swoją aplikację BI. Jeśli wykonasz ten krok z należytą starannością, możesz mieć rozsądną pewność, że środowisko wspomagania decyzji BI będzie tak stabilne, niezawodne i bezpieczne, jak każde inne środowisko produkcyjne. Jeśli wykonasz ten krok w pośpiechu, ryzykujesz, że środowisko wspierające podejmowanie decyzji BI nie będzie tak solidnym i bezpiecznym środowiskiem, jakiego oczekują ludzie biznesu.