

Krok 7: Analiza repozytorium metadanych

Ta część obejmuje następujące tematy:

- * Rzeczy do rozważenia podczas analizy, czy licencjonować (kupić) lub zbudować repozytorium metadanych
- * Dlaczego ważne jest dostarczanie metadanych z każdym projektem BI
- * Różnice między dwiema kategoriami metadanych: metadanymi biznesowymi i metadanymi technicznymi
- * W jaki sposób repozytorium metadanych może pomóc ludziom biznesowym znaleźć i wykorzystać ich dane biznesowe po ich ustandaryzowaniu pod kątem środowiska wspomagania decyzji BI
- * Cztery grupy komponentów metadanych: własność, cechy opisowe, reguły i polityki oraz cechy fizyczne
- * Jak priorytetyzować metadane do celów wdrożeniowych?
- * Pięć typowych trudności napotykanym w przypadku inicjatyw dotyczących repozytoriów metadanych: wyzwania techniczne, kadrowe, budżetowe, użyteczności i polityczne
- * Meta model relacji encji (E-R) używany do dokumentowania wymagań dotyczących metadanych
- * Definicja i przykłady meta-metadanych
- * Krótkie opisy działań związanych z analizą repozytorium metadanych, rezultatów wynikających z tych działań oraz zaangażowanych ról Ryzyko niewykonania Kroku 7

Rzeczy do rozważenia

Wykorzystanie repozytorium metadanych

- * Kto będzie używał metadanych w repozytorium metadanych?
- * Jakie mamy standardy dotyczące jego użytkowania? Jakie standardy musimy opracować?
- * Czy mamy już repozytorium metadanych? Jeśli nie, czy udzielimy licencji? (kupić) czy zbudować?
- * Czy to repozytorium metadanych obsługuje tylko wsparcie decyzji BI środowiska, czy będzie stosowany we wszystkich systemach w całej organizacji?
- * Skąd będziemy wiedzieć, czy skutecznie korzystamy z repozytorium metadanych?

Wymagania dotyczące metamodelu

- * Czy musimy stworzyć meta model dla repozytorium meta danych, czy już go mamy?
- * Jeśli taki mamy, jakie obiekty metadanych musimy do niego dodać?
- * Które produkty z repozytorium danych meta obsługują nasz model meta?
- * Czy te produkty repozytorium metadanych można rozszerzać?

Bezpieczeństwo repozytorium danych meta

- * Jakiego rodzaju zabezpieczenia będziemy potrzebować dla repozytorium metadanych?
- * Kto będzie upoważniony do wprowadzania i utrzymywania metadanych?

* Czy każdy będzie miał w dowolnym momencie dostęp do metadanych?

Przechwytywanie metadanych

* Jakie typy metadanych biznesowych musimy przechwycić?

* Czy będziemy używać narzędzia do wspomagania komputerowego inżynierii oprogramowania (CASE) do przechwytywania metadanych biznesowych?

* Jaki rodzaj metadanych technicznych musimy przechwycić?

* Czy będziemy przechwytywać techniczne metadane w ekstrakcji/transformacji/ladowaniu (ETL), przetwarzaniu analitycznym online (OLAP) oraz innych narzędziach dostępu i analizy?

* Jak wyodrębnimy metadane z tych narzędzi i zmigrujemy je do repozytorium metadanych? Kto jest odpowiedzialny za jego migrację?

* Kto połączy metadane biznesowe z narzędzia CASE z metadanymi technicznymi z ETL, OLAP i innych narzędzi?

Dostarczanie metadanych

* Jak będą wyświetlane metadane? Jak będzie dostępny? Czy będziemy mieć interfejs WWW do repozytorium metadanych?

* Czy będziemy musieli tworzyć raporty metadanych? Jakie rodzaje raportów?

* Jak będą dystrybuowane te raporty?

* Czy będzie funkcja pomocy (samouczek online)? Czy funkcja pomocy będzie zależna od kontekstu?

Rekrutacja

* Czy mamy już administratora metadanych? Jeśli nie, to czy mamy administratora danych z umiejętnościami technicznymi, który może pełnić funkcje administratora metadanych?

* Czy będziemy musieli zatrudnić więcej administratorów metadanych?

* Jak zostaną podzielone obowiązki dotyczące metadanych pomiędzy administratorów danych i administratorów metadanych?

Repozytorium metadanych to baza danych. Ale w przeciwieństwie do zwykłych baz danych, repozytorium metadanych nie jest przeznaczone do przechowywania danych biznesowych dla aplikacji biznesowej. Zamiast tego jest przeznaczony do przechowywania informacji kontekstowych o danych biznesowych. Przykłady informacji kontekstowych o danych biznesowych obejmują:

* Znaczenie i treść danych biznesowych

* Zasady regulujące dane biznesowe

* Atrybuty techniczne danych biznesowych

* Specyfikacje, które przekształcają dane biznesowe

* Programy manipulujące danymi biznesowymi

Informacje kontekstowe dotyczące danych biznesowych z natury istnieją w każdej organizacji, niezależnie od tego, czy są udokumentowane, czy nie. Gdy informacje kontekstowe są

udokumentowane, nazywane są metadanymi. Gdy informacje nie są udokumentowane, zwykle nie są znane wszystkim w organizacji. W rezultacie ludzie biznesu często wymyślają własne reguły biznesowe i tworzą własne nadmiarowe dane wraz z nadmiarowymi procesami, nie zdając sobie sprawy (a czasami nie dbając o to, że to, czego potrzebują, może już istnieć). Czterdzieści lat takich praktyk doprowadziło obecnie większość organizacji na skraj utonięcia w danych i odwodnienia z powodu braku informacji.

Znaczenie metadanych

Metadane opisują organizację pod względem jej działań biznesowych oraz obiektów biznesowych, na których wykonywane są działania biznesowe. Rozważmy na przykład sprzedaż produktu klientowi przez pracownika. Sprzedaż jest czynnością biznesową, a produkt, klient i pracownik są obiektami biznesowymi, na których dokonywana jest czynność sprzedaży. Działania biznesowe i obiekty biznesowe, zarówno ręczne, jak i automatyczne, działają zgodnie z zestawem relacji i reguł zdefiniowanych przez firmę. Te działania i obiekty oraz rządzące nimi relacje i reguły zapewniają kontekst, w którym pracownicy biznesowi codziennie korzystają z danych biznesowych. Metadane są tak ważne dla środowiska wspomagania decyzji BI, ponieważ pomagają w metamorfozie danych biznesowych w informacje. Różnica między danymi a informacjami polega na tym, że informacje są surowymi danymi w kontekście biznesowym. Metadane zapewniają kontekst biznesowy; oznacza to, że metadane zapewniają prawidłową interpretację (na podstawie działań, obiektów, relacji i reguł) tego, co faktycznie oznaczają dane biznesowe. Na przykład, czym jest zysk? Czy jest to kwota pozostała po sprzedaży produktu i zapłaceniu wszystkim, którzy byli zaangażowani w ten produkt? A może jest to bardziej skomplikowana kalkulacja, taka jak „całkowity roczny dochód minus suma średniego kosztu bazowego na produkt minus rzeczywiste koszty ogólne personelu minus skumulowane roczne premie produkcyjne minus rabaty hurtowe minus kupony podzielone przez dwanaście?” Czy każdy człowiek biznesu ma takie samo rozumienie zysku? Czy istnieje jedna i tylko jedna kalkulacja zysku? Jeśli istnieją różne interpretacje zysku, czy wszystkie interpretacje są uzasadnione? Jeśli istnieje wiele legalnych wersji do obliczania zysków, należy utworzyć wiele elementów danych, każdy z własną unikatową nazwą, definicją, treścią, regułami, relacjami itd. Wszystkie te kontekstowe informacje o zysku to metadane. Ponieważ metadane zapewniają kontekst biznesowy, w którym używane są dane biznesowe, metadane mogą być postrzegane jako semantyczna (interpretacyjna) warstwa środowiska wspomagania decyzji BI. Ta warstwa semantyczna pomaga ludziom biznesowym nawigować po docelowych bazach danych BI, w których znajdują się dane biznesowe. Pomaga także technikom zarządzać docelowymi bazami danych BI oraz aplikacjami BI. Poniżej wymieniono niektóre ważne cechy metadanych i repozytoriów metadanych.

* Repozytorium metadanych jest wypełnione metadanymi z wielu różnych narzędzi, takich jak narzędzia CASE, narzędzia ETL, narzędzia OLAP i narzędzia do eksploracji danych.

* Metadane dokumentują transformację i czyszczenie danych źródłowych oraz zapewniają ścieżkę audytu okresowych wczytywania danych.

* Metadane pomagają śledzić wymagania bezpieczeństwa BI, miary jakości danych i metryki wzrostu (wielkość danych, sprzęt itd.).

* Metadane zapewniają spis wszystkich danych źródłowych, które wypełniają aplikacje BI.

* Metadane mogą być zarządzane centralnie lub mogą być dystrybuowane. Tak czy inaczej, każde wystąpienie komponentu metadanych powinno być unikalne, niezależnie od jego fizycznej lokalizacji.

Kategorie metadanych

Istnieją dwie kategorie metadanych: metadane biznesowe i metadane techniczne.

1. Metadane biznesowe zapewniają ludziom biznesu plan dostępu do danych biznesowych w środowisku wspomagania decyzji BI. Ponieważ wielu ludzi biznesu jest stosunkowo nietechnicznych, powinni mieć dostęp do metadanych, które definiują środowisko wspomagania decyzji BI w zrozumiałych dla nich terminach biznesowych.

2. Techniczne metadane wspierają techników i "zaawansowanych użytkowników" poprzez dostarczanie im informacji o ich aplikacjach i bazach danych, których potrzebują do utrzymania aplikacji BI.

W tabeli przedstawiono pewne różnice między metadanymi biznesowymi a metadanymi technicznymi.

Metadane biznesowe: Metadane techniczne

Dostarczone przez ludzi biznesu: Dostarczone przez techników lub narzędzia

Udokumentowane pod względem biznesowym na modelach danych i w słownikach danych :
Udokumentowane pod względem technicznym w bazach danych, plikach, programach i narzędziach

Używany przez ludzi biznesu : używany przez techników, „zaawansowanych użytkowników”, bazy danych, programy i narzędzia (np. ETL, OLAP)

Nazwy w pełni napisane w języku biznesowym : skrócone nazwy ze znakami specjalnymi, takimi jak „_” (podkreślenie) lub „-” (myślnik), używane w bazach danych, plikach i programach

Repozytorium metadanych jako narzędzie nawigacyjne

Metadane nie są nowe; zawsze była częścią systemów operacyjnych. Można go znaleźć w dokumentacji systemów, układach rekordów, katalogach baz danych i sekcjach deklaracji danych w programach. Rola metadanych w środowisku operacyjnym zawsze była postrzegana jako dokumentacja systemowa, z której korzystali głównie technicy zajmujący się obsługą systemów operacyjnych. Gdy część dokumentacji systemowej (metadane) stała się nieaktualna, personel techniczny miał wystarczające umiejętności, aby przeczytać rzeczywisty kod programistyczny, aby wyodrębnić informacje, których szukali, takie jak znaczenie i zawartość elementu danych. Dlatego najczęściej metadane były traktowane jako refleksja. W środowisku wspomagania decyzji BI metadane nabierają nowego poziomu ważności. Trzeba obsłużyć nową publiczność, a mianowicie ludzi biznesu. Metadane pomagają im lokalizować, zarządzać, rozumieć i wykorzystywać dane w docelowych bazach danych BI. Metadane mają nową rolę: nawigację, a nie tylko dokumentację. Ludzie biznesu zwykle nie mają umiejętności technicznych, ani czasu ani chęci, aby rozszyfrować kod programistyczny. Nie chcą też pozostawać w zależności od działu IT, który interpretuje znaczenie i zawartość danych po tym, jak zostały one zmanipulowane przez programy. Zamiast dzwonić do programisty, przedsiębiorca powinien mieć dostęp do metadanych, które pomogłyby mu efektywnie poruszać się po środowisku wspomagania decyzji BI i interpretować dane BI.

Standaryzacja danych

Gdyby dane biznesowe były przez cały czas przechowywane i wykorzystywane w spójny, zatwierdzony sposób, problemy z nadmiarowością i niespójnością danych, które obecnie nękają wiele systemów operacyjnych, nie istniałyby w takim stopniu, w jakim występują obecnie. Niestety, złe nawyki ciężko umierają. Deweloperzy i ludzie biznesu nadal jawnie lub niejawnie ponownie wykorzystują dane biznesowe w systemach operacyjnych do różnych celów. Na przykład, programiści nadal jawnie przeddefiniują elementy danych w swoich programach, a ludzie biznesu nadal niejawnie przeddefiniują (wynajdują nowe kody dla) istniejących elementów danych w celu przechwytywania niepowiązanych informacji. Dokumentacja tych redefinicji również pozostaje słaba lub nie istnieje. Jeśli

jakakolwiek dokumentacja istnieje, rzadko jest przekazywana wszystkim w organizacji, którzy jej potrzebują, i bardzo rzadko jest aktualizowana. Dlatego ludzie biznesu nadal wymyślają własne reguły biznesowe i tworzą własne nadmiarowe dane wraz z nadmiarowymi procesami. Każdy zespół projektowy BI musi zająć się tym istniejącym chaosem danych i musi dołożyć wszelkich starań, aby promować standaryzację danych. Podczas standaryzacji danych biznesowych dla środowiska wspomagania decyzji BI, zespół projektowy BI powinien dokumentować wszystkie zmiany wprowadzone do danych, tak aby każdy mógł być ich świadomy. Ta dokumentacja ma postać metadanych w repozytorium metadanych. Na przykład można zmienić nazwę źródłowego elementu danych, aby była zgodna z nowymi standardami nazewnictwa, lub wartości danych można filtrować, dodawać lub przekształcać w celu wymuszenia reguły biznesowej. W obu przypadkach dane BI w docelowej bazie danych BI nie będą już pasować do danych źródłowych w pliku źródłowym lub źródłowej bazie danych. Metadane zapewniłyby nawigację między nimi. Korzystanie z aplikacji BI bez wiedzy, że dane biznesowe zostały zmienione i jak zostały zmienione, może być frustrującym doświadczeniem, które może ostatecznie zakończyć się tym, że ludzie biznesu w ogóle nie będą chcieli już korzystać z aplikacji BI. Byłoby to katastrofalne, ponieważ jednym z najważniejszych aspektów inicjatywy wspierającej podejmowanie decyzji BI jest zapewnienie łatwego w użyciu, intuicyjnego sposobu, w jaki ludzie biznesu mogą uzyskać dostęp do danych i przeszukiwać je. Łatwa w obsłudze aplikacja oznacza, że ludzie biznesu:

- * Nie musisz być ekspertem w dziedzinie technologii relacyjnych
- * Nie musisz znać strukturalnego języka zapytań (SQL)
- * Nie musisz znać fizycznej struktury baz danych
- * Nie musisz znać lokalizacji swoich danych
- * Nie musisz zgadywać znaczenia danych
- * Nie musisz szukać wymaganych informacji

Klasyfikacje metadanych

Ponieważ projekty BI mogą generować dużą liczbę komponentów metadanych, warto sklasyfikować te komponenty i nadać im priorytety do stopniowej implementacji.

Grupowanie komponentów metadanych

Składniki metadanych można podzielić na cztery grupy lub klasyfikacje metadanych: własność, cechy opisowe, reguły i zasady oraz cechy fizyczne. Repozytorium metadanych powinno być w stanie przechowywać składniki metadanych wszystkich czterech klasyfikacji, jak wymieniono poniżej.

Własność

* Właściciel danych: dane są własnością organizacji. Ponieważ jednak organizacja jest osobą prawną, a nie osobą, ktoś w organizacji musi przejąć uprawnienia i odpowiedzialność za ustalanie polityki, określanie reguł i ustanawianie standardów danych organizacyjnych. Te uprawnienia i odpowiedzialność można rozdzielić między menedżerów linii biznesowych lub przypisać do komitetu własności danych (którego członkami najprawdopodobniej będą niektórzy lub wszyscy menedżerowie linii biznesowej). Przykładem własności danych rozproszonych jest kierownik działu zasobów ludzkich, który ma uprawnienia i odpowiedzialność za ustalanie zasad, reguł i standardów dotyczących danych płacowych, ale nie danych produktów. W przypadku posiadania danych przez komitet, komitet ustala

zasady, zasady i standardy dla wszystkich danych w drodze konsensusu, delegacji na członka komitetu lub na podstawie innej zasady komitetu.

* Właściciel aplikacji: Tradycyjnie własność przypisywana jest do systemu jako całości. Ponieważ system zazwyczaj składa się z aplikacji i jej danych, „własność systemu” oznacza, że ta sama osoba ma uprawnienia do ustalania zasad, określania reguł i ustanawiania standardów zarówno dla danych, jak i funkcjonalności (aplikacji). Może to być prawidłowy warunek dla systemów operacyjnych, z których pochodzą dane, ale nie dotyczy aplikacji BI, ponieważ większość ludzi biznesu korzystających z aplikacji BI to nie te same osoby, które są źródłem danych operacyjnych. Dlatego konsumenci informacji BI mogą być właścicielami aplikacji BI, ale większość z nich nie będzie właścicielem danych.

Charakterystyka opisowa

* Nazwa: Każdy obiekt danych, element danych i proces biznesowy powinien mieć unikalną nazwę.

* Definicja: Każdy obiekt danych, element danych i proces biznesowy powinien mieć krótką definicję wyjaśniającą, co to jest.

* Typ i długość: Każdy element danych powinien mieć zadeklarowany oficjalny typ i długość, nawet jeśli elementy danych w systemach źródłowych lub kolumny lub komórki w docelowych bazach danych mogą od niego odbiegać. To odchylenie byłoby również zdefiniowane jako metadane w elemencie danych, kolumnie lub komórce, w której wystąpiło.

* Domena: każdy element danych powinien mieć zadeklarowany zestaw dopuszczalnych wartości, nawet jeśli jest to zestaw obejmujący wszystkie elementy, taki jak „dowolny znak, liczba lub znak”.

* Uwagi: Należy uwzględnić dodatkowe fakty dotyczące danych lub procesów. Jest to przechwytywanie dowolnych komentarzy, takich jak „Spór między inżynierią a marketingiem dotyczący znaczenia kodu typu podkomponentu produktu został przekazany do rozstrzygnięcia komitetowi sterującemu BI”.

Zasady i reguły

* Relacje: Obiekty danych są ze sobą powiązane poprzez działania biznesowe. Repozytorium metadanych powinno być w stanie przechowywać informacje o tych relacjach.

* Reguły biznesowe i polityki biznesowe: te komponenty mogą dotyczyć zarówno danych, jak i procesów. Mogą to być reguły konwersji danych technicznych, reguły domeny danych biznesowych, reguły integralności danych biznesowych lub reguły przetwarzania.

* Bezpieczeństwo: Wymagania dotyczące bezpieczeństwa mogą dotyczyć danych, metadanych, procesów, baz danych, aplikacji (programów i ekranów), narzędzi i witryn sieci Web.

* Czystość: Należy przechowywać metryki dotyczące sum uzgadniania ETL i jakości danych BI. Metryki mogą być wyrażone jako procenty wiarygodności ładunku danych (np. 89 procent kodu typu klienta jest prawidłowe) lub jako liczniki rekordów określające liczbę rekordów przefiltrowanych (odrzuconych) oraz liczbę rekordów przepuszczonych podczas procesu ETL.

* Zastosowanie: Dane nie są wieczne. Czasami nowe dane są wymyślane i przechwytywane, a stare dane są wycofywane i nie są już używane. Ponieważ docelowe bazy danych BI przechowują wiele lat historii, niektóre kolumny lub komórki nie będą miały wartości dla wszystkich okresów, ponieważ dane nie miały zastosowania lub nie istniały w pewnych okresach czasu. Jeśli na wykresach analizy trendów pojawiają się skoki, należy skorzystać z repozytorium metadanych w celu określenia możliwości zastosowania tego konkretnego fragmentu danych.

* Terminowość: Ludzie biznesu będą chcieli wiedzieć, kiedy dane źródłowe zostały ostatnio zaktualizowane i które wersje systemów operacyjnych zostały użyte do aktualizacji. Nie wszystkie systemy operacyjne działają codziennie lub tego samego dnia miesiąca. Jeden system operacyjny może działać w ostatnim dniu kalendarzowym miesiąca, podczas gdy inny może działać w ostatnim dniu roboczym miesiąca. Niektóre systemy operacyjne nie „zamykają miesiąca”, dopóki nie zakończą cyklu korekty cztery do dziesięciu dni po ostatnim dniu kalendarzowym miesiąca.

Charakterystyka fizyczna

* Pochodzenie (źródło): Ponieważ docelowe bazy danych BI przechowują tylko istniejące dane operacyjne (wygenerowane wewnętrznie i zakupione zewnętrznie), pochodzenie lub źródło każdego elementu danych powinno być udokumentowane. Jedna kolumna w docelowej bazie danych BI może być wypełniona elementami danych z wielu źródeł. Na przykład, kolumna Saldo rachunku w tabeli Konta może być wypełniona z elementu danych Saldo rachunku lokaty na żądanie w źródłowej bazie danych Rachunek bieżący oraz z elementu danych Saldo dzienne rachunku lokaty terminowej w pliku Transakcja rachunku oszczędnościowego. I odwrotnie, jeden element danych źródłowych może zasilać wiele kolumn w docelowej bazie danych BI. Na przykład element danych Type Code może być wykorzystywany w systemie operacyjnym do dwóch celów. Wartości danych „A”, „B” i „C” kodu typu mogą być użyte do wypełnienia kolumny Kod typu klienta w tabeli Klient, a wartości danych „N”, „O” i „P” ten sam kod typu może być użyty do wypełnienia kolumny Kod typu produktu w tabeli Produkt.

* Fizyczna lokalizacja: kilka komponentów metadanych (np. tabele, kolumny, nazwy zestawów danych) powinno opisywać, gdzie dane znajdują się w środowisku wspomaganie decyzji BI.

* Transformacja: bardzo niewiele elementów danych można przenieść ze źródła do celu bez jakiegokolwiek transformacji. Może zajść konieczność zmiany typu i długości danych, a kody jednoznakowe muszą zostać przetłumaczone na wieloznakowe mnemoniki. W najgorszym przypadku długie reguły biznesowe mogą wymagać bardziej skomplikowanych przekształceń obejmujących edycję, filtrowanie, łączenie, rozdzielanie lub tłumaczenie wartości danych.

* Pochodne: Ten składnik przechowuje obliczenia dla kolumn pochodnych. Chociaż kolumny pochodne zwykle nie są przechowywane w systemach operacyjnych, normą jest przechowywanie ich w docelowych bazach danych BI.

* Agregacja i podsumowanie: Podobnie jak w przypadku wyprowadzania, reguły agregacji i podsumowania powinny być przechowywane jako metadane.

* Wolumen i wzrost: Rozmiar i wzrost docelowych baz danych BI są często ogromne. Dlatego zarówno prognozowane, jak i rzeczywiste wolumeny powinny być dokumentowane jako metadane pod względem liczby wierszy i procentu oczekiwanego wzrostu. Ludzie biznesu najczęściej uzyskują dostęp do komponentów metadanych w klasyfikacji cech opisowych oraz klasyfikacji reguł i polityk. Technicy zazwyczaj uzyskują dostęp do komponentów metadanych w klasyfikacji cech fizycznych

Priorytetyzacja komponentów metadanych

Przechwytywanie wszystkich składników metadanych może nie być konieczne lub praktyczne w przypadku wszystkich projektów BI. Jednak schwytywanie żadnego jest niedopuszczalne. Z reguły metadane powinny być dostarczane z każdym projektem BI. Dzięki niej przedsiębiorcy będą mogli rozpoznać swoje stare dane, prześledzić, co się z nimi stało (transformacja), zlokalizować je w nowych docelowych bazach danych BI i określić, jak je właściwie wykorzystać. Innymi słowy, ludzie biznesu odniosą ogromne korzyści z udostępnienia metadanych, które pomogą im poruszać się po środowisku wspomaganie decyzji BI. Nie wszystkie składniki metadanych mają tę samą wartość dla wszystkich ludzi

biznesu lub wszystkich aplikacji BI. Przydatne może być uszeregowanie według priorytetów komponentów metadanych w trzech grupach: obowiązkowe, ważne (korzystne, ale nie obowiązkowe) i opcjonalne.

Wszystkie obowiązkowe komponenty metadanych i jak najwięcej ważnych komponentów metadanych powinny być przechwytywane i przechowywane w repozytorium metadanych. Opcjonalne komponenty metadanych mogą zostać odłożone w przyszłych wydaniach aplikacji BI.

Wyzwania związane z repozytorium metadanych

Dobre pomysły są często trudne do wdrożenia. Udostępnienie repozytorium metadanych jest dobrym pomysłem, ale też dość trudnym, niezależnie od tego, czy podjęta zostanie decyzja o licencjonowaniu (zakupie) dostępnego na rynku produktu, czy też o zbudowaniu repozytorium od podstaw. W tej sekcji pokrótce opisano wyzwania stojące przed implementacją repozytorium metadanych.

Wyzwania techniczne

Budowanie repozytorium metadanych nie jest trywialnym zadaniem. Jest to projekt sam w sobie, z własnym planem projektu, własnymi etapami rozwoju i własnym personelem. Wszystkie wyzwania technologiczne, które dotyczą baz danych i aplikacji, mogą pojawić się w projektach repozytorium metadanych. Licencjonowanie produktu repozytorium metadanych jest alternatywą dla jego budowania, ale „zwykłe waniliowe” wersje komercyjnie dostępnych produktów repozytorium metadanych często nie spełniają wszystkich wymagań dotyczących metadanych środowiska wspierającego decyzje BI. Dlatego licencjonowanie produktu będącego repozytorium metadanych nadal wymaga szczegółowej analizy wymagań w celu wybrania odpowiedniego produktu, a także znacznego wysiłku wdrożeniowego w celu jego ulepszenia. Ulepszanie licencjonowanego oprogramowania wiąże się z własnymi wyzwaniami. Kod źródłowy produktu może być niedostępny. Sprzedawca może nalegać na włączenie żądanych ulepszeń za odpowiednią cenę i we własnym tempie. Zwiększa się czas i wysiłek wymagany do konserwacji produktu, ponieważ ulepszenia muszą zostać ponownie zastosowane w nowych wydaniach i wersjach licencjonowanego produktu repozytorium metadanych.

Wyzwania kadrowe

Metadane powinny być „żywą” dokumentacją przechowywaną w bazie danych, czyli w repozytorium metadanych. Przechowywanie metadanych jako dokumentów papierowych gwarantuje, że zmienią się one w „półkę” w ciągu miesiący, jeśli nie tygodni. Oznacza to, że co najmniej jeden administrator metadanych musi poświęcić cały czas na zarządzanie zawartością repozytorium metadanych i oprogramowaniem. Jeśli w ramach projektu BI budowane jest repozytorium metadanych, jeden personel nie wystarczy. Wysiłek związany z repozytorium metadanych będzie wymagał analityka, projektanta danych, projektanta baz danych i co najmniej jednego programisty.

Wyzwania budżetowe

Chociaż wielu ekspertów BI uważa metadane za „klej” środowiska wspomaganie decyzji BI, większość organizacji przeznaczają niewiele lub wcale na tworzenie i utrzymywanie repozytorium metadanych. Nadal traktują metadane jako dokumentację systemową dla techników, a nie narzędzie nawigacyjne dla ludzi biznesu. Ból związany z frustracją związaną z dostępem i zamieszaniem w danych musi często osiągnąć niedopuszczalny poziom, zanim organizacje uwzględnią metadane jako obowiązkowy i standardowy element ich projektów BI.

Brak metadanych był często wymieniany jako jedna z przyczyn niepowodzenia aplikacji BI.

Wyzwania użyteczności

Korzystanie z repozytorium metadanych powinno być całkowicie intuicyjne. Ludzie biznesu powinni mieć możliwość kliknięcia ikony i natychmiastowego uzyskania żądanych informacji o tabeli lub kolumnie, wykresie lub raporcie, a nawet zapytaniu biznesowym. Bardziej złożone zapytania dotyczące repozytorium metadanych powinny być obsługiwane za pomocą wbudowanych lub dostosowanych makr. Jednak najbardziej dopracowanym sposobem prezentacji metadanych jest uwzględnienie ich w zapytaniach BI. Niestety, wiele produktów repozytoriów metadanych jest nadal projektowanych przez techników dla techników, a nie dla ludzi biznesu. Niektóre z tych produktów nadal mają zagadkowy język metadanych, brak wyrafinowanych możliwości raportowania, brak kontekstu i wymaga zrozumienia metamodelu opisującego obiekty metadanych i ich relacje.

Wyzwania polityczne

Stworzenie rozwiązania metadanych obejmującego całe przedsiębiorstwo jest trudne, ponieważ należy pogodzić różnice między departamentami i rozwiązać politykę międzywydziałową. Spory te, choć całkowicie przewidywalne, rzadko są brane pod uwagę podczas tworzenia planu projektu. W rezultacie projekty są opóźniane, a kwestie te są adresowane lub przekazywane do kadry kierowniczej i komitetów sterujących. Sprawia to wrażenie, że projekty BI są trudną, kontrowersyjną, męczącą, wyczerpującą, powolną i ogólnie niepożądaną pracą.

Pomimo tych wszystkich wyzwań repozytorium metadanych jest obowiązkowym elementem każdego środowiska wspomagającego podejmowanie decyzji BI.

Logiczny metamodel

Niezależnie od tego, czy repozytorium metadanych jest licencjonowane, czy zbudowane oraz niezależnie od metody implementacji (scentralizowana, zdecentralizowana lub rozproszona, jak omówiono w Kroku 10, Projektowanie repozytorium metadanych), repozytorium metadanych powinno obsługiwać logiczny metamodel, który odzwierciedla wymagania dotyczące metadanych. Podobnie jak w przypadku danych biznesowych, każdy składnik metadanych jest z natury wyjątkowy. Ważne jest, aby zdefiniować te unikalne obiekty metadanych, ich zawartość, ich wzajemne powiązania i współzależności, niezależnie od tego, jak będą przechowywane lub dostępne. Technika tego działania jest logiczne modelowanie danych, tylko w tym przypadku wytworzy logiczny meta model. Logiczny meta model to model danych, który wskazuje obiekty, relacje między obiektami oraz licznosc i opcjonalność relacji. Różnica między logicznym modelem meta dla repozytorium danych a logicznym modelem danych dla aplikacji biznesowej tkwi w naturze obiektów. Obiekty w logicznym modelu meta reprezentują metadane, takie jak jednostka, atrybut, definicja, domena, tabela, kolumna i indeks. Obiekty w logicznym modelu danych reprezentują dane biznesowe, takie jak klient, produkt, pracownik, konto i lokalizacja.

Metamodel relacji encji

Logiczny meta model jest tworzony podczas pierwszej inicjatywy metadanych i jest rozszerzany przy każdej kolejnej inicjatywie. Logiczną reprezentację obiektów metadanych należy uchwycić jako diagram E-R ze względu na wyraźne definicje obiektów metadanych, relacje między nimi i zawartość obiektów. Metamodel E-R przede wszystkim pomaga ludziom zrozumieć, komunikować się i weryfikować wymagania dotyczące metadanych. Dlatego metamodel E-R powinien być postrzegany jako model wymagań, który ma być używany do oceny produktów repozytorium metadanych oraz do

ustalania punktu odniesienia podczas projektowania dostosowanego repozytorium metadanych, nawet jeśli jego fizyczny metamodel (projekt bazy danych) stanie się obiektem- zorientowany (OO).

Meta-Metadane

Ponieważ metadane to informacje kontekstowe o danych biznesowych, meta-dane to informacje kontekstowe o metadanych. Wiele składników meta-danych jest podobnych do tych z metadanych. Na przykład każdy obiekt metadanych powinien zawierać komponenty obejmujące nazwę, definicję, rozmiar i długość, zawartość, własność, relacje, reguły biznesowe, bezpieczeństwo, czystość, lokalizację fizyczną, zastosowanie, terminowość, objętość i uwagi. Meta-dane meta dla obiektu metadanych mogą wyglądać tak:

- * Nazwa: podmiot
- * Związek: związany z jedną lub wieloma tabelami
- * Bezpieczeństwo: odczytane przez wszystkich, aktualizowane przez administratora danych
- * Własność: administrator danych
- * Pochodzenie: narzędzie ERWIN CASE
- * Fizyczna lokalizacja: tabela MDRSYSENT
- * Czystość: 2 procent brakujących danych
- * Aktualność: ostatnia aktualizacja 1 listopada 2002 r.
- * Objętość i wzrost: 2391 rzędów, tempo wzrostu 1 procent rocznie

Czynności związane z analizą repozytorium metadanych

Czynności związane z analizą repozytorium metadanych nie muszą być wykonywane liniowo. Poniższa lista zawiera krótki opis czynności związanych z krokiem 7, Analiza repozytorium metadanych.

1. Przeanalizuj wymagania dotyczące repozytorium metadanych. Współpracuj z przedstawicielem biznesowym, aby określić i ustalić priorytety wymagań dotyczących metadanych dla konkretnego projektu BI. Wskaż, które składniki metadanych są obowiązkowe, ważne i opcjonalne. Jeśli repozytorium metadanych już istnieje, określ, które składniki metadanych należy dodać, jeśli w ogóle. Zaktualizuj najnowszą wersję dokumentu wymagań aplikacji (poprawionego podczas lub po prototypowaniu).
2. Przeanalizuj wymagania interfejsu dla repozytorium metadanych. Niezależnie od tego, czy repozytorium metadanych jest licencjonowane, czy zbudowane, musi akceptować metadane z różnych źródeł. Metadane biznesowe będą musiały zostać wyodrębnione z narzędzi CASE, dokumentów edytora tekstu lub arkuszy kalkulacyjnych. Techniczne metadane będą musiały zostać wyodrębnione i połączone ze słownikami systemu zarządzania bazami danych (DBMS), narzędziami ETL, narzędziami do czyszczenia danych, narzędziami OLAP, autorami raportów i narzędziami do eksploracji danych.
3. Przeanalizuj wymagania dotyczące dostępu do repozytorium metadanych i raportowania. Wypełnianie bazy danych jest bez znaczenia, chyba że można uzyskać dostęp do zawartości, przeszukać ją i zgłosić. Dotyczy to zarówno metadanych, jak i danych biznesowych. Zidentyfikuj wymagania dotyczące dostępu do metadanych, wymagania dotyczące bezpieczeństwa i wymagania funkcji pomocy. Oceń alternatywne tryby wyświetlania, takie jak Portable Document Format (PDF), Hypertext Markup Language (HTML), SQL, gotowe zapytania lub zastrzeżone oprogramowanie do raportowania

metadanych repozytorium. Samouczek pomocy kontekstowej byłby korzystną funkcją do uwzględnienia.

4. Utwórz logiczny meta model.

Narysuj logiczny model meta jako model E-R, aby jawnie pokazać relacje między obiektami metadanych, nawet jeśli planujesz zaimplementować repozytorium metadanych jako bazę danych OO. Innymi słowy, logiczny meta model powinien zawsze być modelem E-R, podczas gdy fizyczny meta model (projekt bazy danych repozytorium meta danych utworzony w kroku 10, Meta Data Repository Design) może być modelem E-R lub modelem OO.

5. Utwórz meta-dane meta.

Podczas gdy logiczny model meta pokazuje na pierwszy rzut oka wymagania dotyczące repozytorium metadanych, meta-dane opisują szczegółowo wymagane składniki metadanych.

Rezultaty wynikające z tych działań

1. Logiczny meta model

Ten model danych jest w pełni znormalizowanym diagramem E-R pokazującym jednostki jądra, jednostki asocjacyjne, jednostki charakterystyczne, relacje, kardynalność, opcjonalność, unikalne identyfikatory i wszystkie atrybuty obiektów metadanych repozytorium.

2. Meta-metadane

Encje i atrybuty metadanych z logicznego modelu meta muszą być opisane za pomocą metadanych. Komponenty metadanych specyficzne dla metadanych (metametadane) to nazwy metadanych, definicje metadanych, relacje metadanych, unikalne identyfikatory, typy, długości, domeny, reguły biznesowe, polityki i własność metadanych.

Role zaangażowane w te działania

* Administrator danych

Administrator danych gromadzi metadane biznesowe w narzędziu CASE podczas logicznych czynności modelowania danych. Te metadane będą jednym ze źródeł repozytorium metadanych. Administrator danych we współpracy z administratorem metadanych opracowuje i publikuje standardy danych. Może również asystować przy tworzeniu metamodelu i meta-danych.

Administrator metadanych

Na administratorze metadanych spoczywa główna odpowiedzialność za przechowywanie i udostępnianie metadanych oraz za utrzymywanie repozytorium metadanych. Musi przeanalizować wymagania dotyczące metadanych, zidentyfikować komponenty metadanych oraz stworzyć lub ulepszyć logiczny model meta i meta-dane.

Ekspert merytoryczny

Ekspert merytoryczny uczestniczy w tym kroku, reprezentując ludzi biznesu i ich wymagania dotyczące metadanych. Ekspert w danej dziedzinie identyfikuje wymagania bezpieczeństwa i własność danych oraz współpracuje z administratorem danych, administratorem metadanych, właścicielami danych i innymi osobami biznesowymi w celu standaryzacji nazw, definicji, treści i reguł biznesowych.

Ryzyko niewykonania kroku 7

Ponieważ jednym z celów wspomaganie decyzji BI jest eliminacja niespójności, dane źródłowe muszą być ustandaryzowane. Standaryzacja niezmiennie skutkuje zmianą wielu danych źródłowych. Zmiany mogą obejmować zmianę nazwy danych, podzielenie jednego elementu danych źródłowych na wiele kolumn docelowych lub wypełnienie jednej kolumny docelowej z wielu elementów danych źródłowych. Może to również oznaczać tłumaczenie kodów na mnemoniki, standaryzację (zmianę) wartości danych oraz filtrowanie nieodpowiednich lub nieprawidłowych danych. W końcu ludzie biznesu nie będą mogli uzgodnić swoich operacyjnych danych źródłowych z danymi docelowymi BI, chyba że będą mieli ślad tych zmian. Ten ślad nazywa się metadanymi, a ludzie biznesu potrzebują go do efektywnego poruszania się po środowisku wspomaganie decyzji BI.

Bez metadanych ludzie biznesu mieliby trudności ze zrozumieniem i wykorzystaniem przekształconych danych w docelowych bazach danych BI. Byłoby to równie frustrujące, jak bezcelowa jazda samochodem przez tydzień lub miesiąc bez mapy, zgadywanie drogi do celu. Gdy ludzie biznesu postrzegają aplikację BI jako trudną w użyciu lub uważają, że dane BI są niewiarygodne, ponieważ nie pasują już do danych źródłowych w systemach operacyjnych, mogą nazwać inicjatywę wspomaganie decyzji BI jako niepowodzenie.