

Krok 10: Projekt repozytorium metadanych

Tutaj omówiono następujące tematy:

- * Rzeczy do rozważenia podczas projektowania repozytorium metadanych lub podczas oceny produktów dostawcy repozytorium metadanych
- * W jaki sposób niedociągnięcia wczesnych inicjatyw dotyczących metadanych przyczyniły się do niepowodzeń w skutecznym zarządzaniu metadanymi
- * Mnogość źródeł metadanych, którymi musimy teraz zarządzać w środowisku wspomaganie decyzji BI
- * Zalety i wady trzech różnych rodzajów strategii wdrażania: scentralizowane repozytorium metadanych, zbudowane lub licencjonowane (kupione); zdecentralizowane repozytorium metadanych; oraz rozproszone rozwiązanie metadanych, które jest możliwe dzięki użyciu tagów Extensible Markup Language (XML)
- * Zalety i wady dwóch różnych typów projektów repozytoriów metadanych: projektowania relacji encji (ER) i projektowania obiektowego (OO)
- * Szczegółowe przykłady procesu oceny produktu i dostawcy
- * Krótkie opisy działań związanych z projektowaniem repozytorium metadanych, rezultatów wynikających z tych działań oraz zaangażowanych ról
- * Ryzyko niewykonania kroku 10

Rzeczy do rozważenia

Istniejące repozytorium metadanych

- * Czy mamy już repozytorium metadanych?
- * Czy musimy go rozbudowywać? Czy musimy dodawać więcej komponentów metadanych? Lub rozszerzyć funkcjonalność?
- * Kto aktualizuje repozytorium metadanych?
- * Kto go używa? Jak oni z tego korzystają? Z jakich części repozytorium metadanych korzystają?
- * Czy im się to podoba? Czy są jakieś skargi?
- * Jeśli nie mamy repozytorium metadanych, jak sobie radzimy bez niego?
- * Dlaczego go nie mamy? Brak budżetu? Brak środków? Brak zrozumienia?

Produkty z repozytorium danych meta

- * Czy istnieją produkty z repozytorium danych meta, które zadowolą naszą meta?
- wymagania dotyczące danych? A może musimy zbudować repozytorium metadanych?
- od zera?
- * Ile z naszych wymagań dotyczących metadanych nie może być spełnionych przez
- produktów repozytorium metadanych na rynku? Jak ważne są te wymagania dotyczące metadanych?

* Czy produkty repozytorium metadanych można ulepszyć, aby spełnić te specyficzne wymagania dotyczące metadanych?

* Które produkty repozytorium metadanych mają import i eksport możliwości?

Interfejsy

* W jaki sposób zautomatyzujemy interfejsy z repozytorium metadanych do innych narzędzi, które mają własne słowniki metadanych, na przykład inżynierii oprogramowania wspomaganego komputerowo (CASE), ekstrakcji/transformacji/ladowania (ETL) i przetwarzania analitycznego online (OLAP) narzędzia? Czy będziemy musieli kupić dodatkowe oprogramowanie pośredniczące?

* Czy inne narzędzia, z których musimy wyodrębnić metadane, mają?

możliwości importu i eksportu? Czy te narzędzia obsługują XML?

* Jak dostarczymy metadane ludziom biznesu? Poprzez

raporty? Przez funkcję pomocy? Przez interfejs WWW?

* Czy ludziom biznesu będzie trudno nauczyć się korzystać z meta?

interfejsy repozytorium danych? Jakie szkolenia musimy rozwijać?

Rekrutacja

* Czy będziemy potrzebować więcej pracowników do zainstalowania, ulepszenia i utrzymania licencjonowanego produktu repozytorium metadanych?

* Czy będziemy potrzebować więcej pracowników do projektowania, budowania i utrzymywania własnego, niestandardowego repozytorium metadanych?

Termin metadane nie jest nowy, podobnie jak próby zarządzania metadanymi. Nowością jest zwiększona świadomość, że metadane są ważnym rozszerzeniem informacji biznesowych i dlatego zarządzanie metadanymi jest obowiązkowe. Innym ważnym uznaniem jest to, że potrzebne są nowe narzędzia i techniki zarządzania metadanymi – i stają się one dostępne.

Silosy metadanych

Administratorzy danych próbowali inwentaryzować, definiować i organizować metadane od wczesnych lat 80-tych. Większość administratorów danych korzystała ze słowników danych generycznych (repozytoria metadanych nazywano kiedyś słownikami danych); tylko nieliczni próbowali zaprojektować i zbudować własne. Niektóre produkty ze słowników danych generycznych były raczej wyrafinowane i rozszerzalne oraz mogą przechowywać większość wymaganych składników metadanych. Jednak z tymi wczesnymi wysiłkami wiązało się mnóstwo problemów.

* Wypełnianie tych wczesnych słowników danych wymagało ręcznego wysiłku, który był czasochłonny i żmudny, podobnie jak wszystkie ręczne wysiłki.

* Brak umiejętności technicznych większości administratorów danych uniemożliwił im rozszerzenie produktów słownika danych o niestandardowe funkcje, aby były bardziej przydatne.

* Możliwości raportowania produktów wczesnych słowników danych były mniej niż pożądane. Niektóre produkty nie miały nawet funkcji interfejsu programowania aplikacji (API), które umożliwiałyby administratorom danych generowanie niestandardowych raportów.

* Niedojrzałe technologie używane w większości wczesnych słowników danych (które były produktami mainframe) nie zapewniały zautomatyzowanych interfejsów, łatwych w użyciu graficznych interfejsów użytkownika (GUI) ani funkcji pomocy kontekstowej.

* Brak standardów (lub brak egzekwowania standardów) stworzył nie do pokonania obciążenie dla administratorów danych, którzy musieli rozwiązać sprzeczne i niespójne nazwy danych, definicje danych i dziedziny danych.

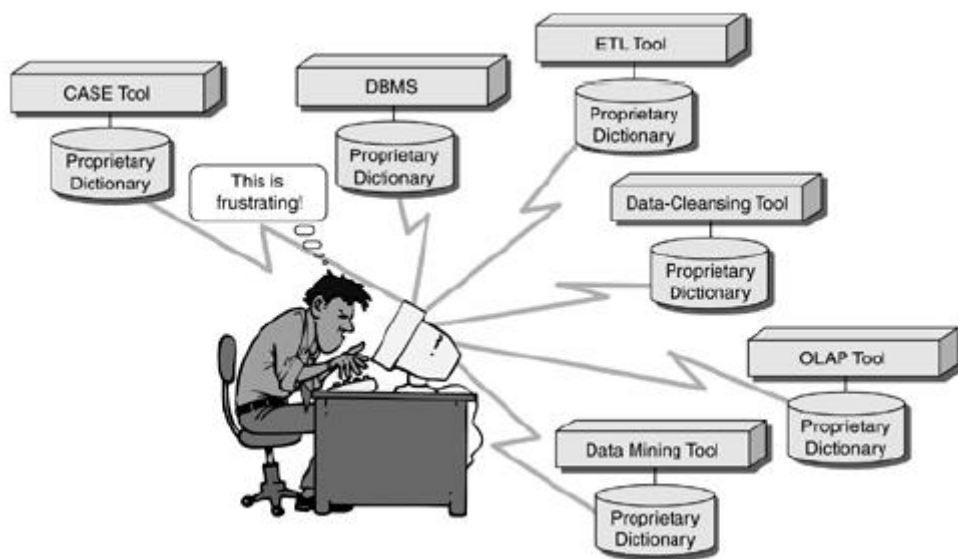
* Brak uznania przez kierownictwo wartości metadanych sprawił, że w większości organizacji metadane mają niski priorytet. Menedżerowie i dyrektorzy biznesowi, a także niektórzy menedżerowie ds. technologii informatycznych (IT), postrzegali metadane jako dokumentację systemową, którą uważali za ważną, ale bez której mogliby się obejść.

* W organizacjach nie istniały żadne inicjatywy międzyorganizacyjne, z wyjątkiem inicjatyw wydziałowych zwykle prowadzonych przez administratorów danych w IT. Dlatego wielu menedżerów i dyrektorów firm nie rozumiało wartości tego wysiłku i nie kupowało go. Popularność inicjatyw związanych z hurtowniami danych w latach 90. pomogła zwiększyć zrozumienie wartości inicjatyw międzyorganizacyjnych.

Z powodu tych problemów wysiłki administracji danymi mające na celu zarządzanie metadanymi były w przeszłości tylko marginalnie skuteczne. W wielu projektach wysiłki te były nawet uważane za przeszkody projektowe ze względu na dodatkowy czas potrzebny na zdefiniowanie i przechwycenie metadanych, gdy technicy zapragnęli pospieszyć się z kodowaniem. Menedżerowie IT i menedżerowie biznesowi często pytali: „Dlaczego jeszcze nie kodujemy?” — oczywiście postrzegali pisanie programów jako jedyną produktywną działalność związaną z opracowywaniem projektów.

Źródła metadanych

Dopiero wraz z pojawieniem się międzyorganizacyjnych inicjatyw BI i związanego z nimi mnóstwa narzędzi BI metadane zaczęły być właściwie rozpoznawane. Ludzie zaczęli zdawać sobie sprawę, że te narzędzia BI, z własnymi zestawami metadanych we własnych zastrzeżonych bazach danych słownikowych, stwarzają aż nazbyt znane problemy związane z nadmiarowością i niespójnością, tyle że tym razem z metadanymi. Pracownicy wiedzy, analitycy biznesowi, menedżerowie i technicy byli bardzo sfrustrowani splątaną siecią silosów metadanych



Nie można uniknąć metadanych, zwłaszcza metadanych technicznych, ponieważ systemy zarządzania bazami danych (DBMS) i większość narzędzi nie działają bez nich. To ich „język”. Na przykład metadane instruują system DBMS, jaki typ struktur bazy danych należy utworzyć, informują narzędzie ETL, jakie dane należy przekształcić, a narzędzie OLAP wie, jak agregować i podsumowywać dane. Różne komponenty metadanych są przechowywane w różnych narzędziach, a żadne z narzędzi (z wyjątkiem repozytorium metadanych) nie jest przeznaczone do przechowywania wszystkich innych komponentów metadanych ze wszystkich innych narzędzi. Na przykład:

- * Narzędzia CASE przechowują metadane biznesowe dla komponentów logicznego modelu danych oraz techniczne metadane dla komponentów fizycznego modelu danych (logiczny projekt bazy danych).

- * Słowniki DBMS przechowują techniczne metadane dla struktury bazy danych, takie jak bazy danych, tabele, kolumny, indeksy i tak dalej.

- * Narzędzia ETL przechowują metadane techniczne dotyczące mapowania danych źródłowych do docelowych oraz specyfikacji transformacji, które są używane przez te narzędzia do wykonywania procesów ETL.

- * Narzędzia do czyszczenia danych przechowują metadane biznesowe dla domen danych i reguł biznesowych, które pozwalają tym narzędziom identyfikować problemy z jakością danych. Przechowują również specyfikacje czyszczenia, które są używane przez te narzędzia do wykonywania funkcji czyszczenia danych.

- * Narzędzia OLAP przechowują techniczne metadane tabel i kolumn w docelowych bazach danych BI, definicje raportów i algorytmy do wyprowadzania, agregowania, podsumowywania i w inny sposób manipulowania danymi BI.

- * Narzędzia do eksploracji danych przechowują metadane techniczne dotyczące różnych modeli analitycznych i algorytmów operacji eksploracji danych.

Podobnie jak w przypadku dostawców innych stosunkowo nowych linii produktów oprogramowania i oprogramowania pośredniczącego, dostawcy produktów repozytorium metadanych konkurują o dominację, co spowalnia standaryzację linii produktów. W rezultacie organizacje kończą z splątana

siecią rozproszonych i rozproszonych metadanych rozproszonych po zastrzeżonych słownikach ich narzędzi. Aby poradzić sobie z tą sytuacją, muszą teraz wyodrębnić, scalać i dokładnie integrować metadane z tych słowników narzędzi, co może być tak samo wyzwaniem, jak wyodrębnianie, scalanie i dokładna integracja danych biznesowych z systemów operacyjnych.

Rozwiązania dotyczące repozytoriów metadanych

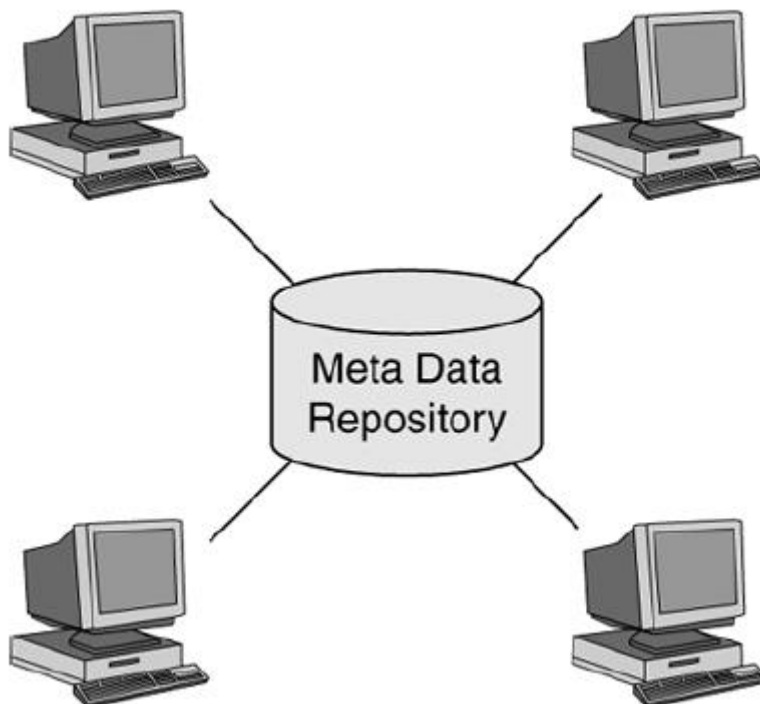
Rozwiązaniem dla silosów metadanych jest oczywiście podejście do repozytorium metadanych obejmujące całe przedsiębiorstwo. Można to osiągnąć na trzy sposoby:

1. Jedno scentralizowane rozwiązanie bazy danych repozytorium metadanych.
2. Zdecentralizowane rozwiązanie metadanych, które wykorzystuje jeden zintegrowany model meta, ale fizycznie dystrybuje metadane w wielu bazach danych
3. Rozproszone rozwiązanie obsługujące XML, w którym metadane są znakowane XML i przechowywane (przechowywane) w różnych typach zastrzeżonych słowników narzędziowych na różnych platformach.

Niezależnie od wybranego rozwiązania repozytorium metadanych, projekty i podprojekty dotyczące repozytorium metadanych są duże i kosztowne. Dlatego każde rozwiązanie repozytorium metadanych powinno być budowane w iteracjach.

Scentralizowane repozytorium metadanych

Scentralizowane repozytorium metadanych jest najpopularniejszym rozwiązaniem i najłatwiejszym do wdrożenia, ponieważ istnieje tylko jedna baza danych, relacyjna lub obiektowa, i tylko jedna aplikacja do utrzymania



Aktualizacja repozytorium metadanych nie musi być koordynowana między bazami danych, a pobieranie metadanych z repozytorium można łatwo wykonać za pomocą prostego interfejsu GUI lub aplikacji sieci Web. Scentralizowane repozytorium metadanych może być zbudowane na zamówienie lub licencjonowane przez dostawcę.

Niestandardowe repozytorium: Zaprojektowanie i zbudowanie dostosowanego scentralizowanego repozytorium metadanych jest alternatywą, którą należy rozważyć. Ponieważ rozwiązanie repozytorium metadanych powinno obejmować całe przedsiębiorstwo, metamodels (zarówno logiczne, jak i fizyczne) będą uogólnione, a nie specyficzne dla aplikacji. Oznacza to, że nie będzie repozytorium metadanych dla aplikacji marketingowej i innego dla aplikacji sprzedażowej; zamiast tego będzie jedno scentralizowane repozytorium metadanych dla wszystkich aplikacji. Zalety i wady tworzenia dostosowanego scentralizowanego repozytorium metadanych są podobne do tych związanych z tworzeniem niestandardowej aplikacji biznesowej.

Zalety:

- * Dostosowany projekt bazy danych zawiera wszystkie wymagania dotyczące metadanych.
- * Interfejs dostępu i interfejsy do narzędzi (ETL, OLAP itd.) są zaprojektowane na zamówienie, aby spełnić wszystkie wymagania.
- * Raporty oraz funkcje pomocy są projektowane dokładnie według potrzeb.
- * Technicy mają pełną kontrolę nad projektem i funkcjonalnością repozytorium metadanych.

Niedogodności:

- * Do utrzymania bazy danych repozytorium metadanych i raportów metadanych potrzebny jest pełnoetatowy personel.
- * Przedni koniec dostępu i interfejsy do narzędzi muszą być zaprogramowane i utrzymywane, a oba te procesy są czasochłonne.
- * Repozytorium metadanych musiałyby być okresowo ulepszone (czasem przeprojektowywane), ponieważ nie można go zbudować od samego początku ze wszystkimi funkcjami.
- * Treść może nie być zsynchronizowana z treścią zastrzeżoną słowniki narzędzi i DBMS.

Licencjonowane repozytorium: Licencjonowanie (kupowanie) produktu scentralizowanego repozytorium metadanych jest atrakcyjną alternatywą dla jego budowania. Utrata niektórych korzyści związanych z niestandardowym rozwiązaniem jest równoważona przez zyskanie korzyści związanych z licencjonowaniem produktu dostawcy

Zalety:

- * Oszczędza się czas, ponieważ nie trzeba projektować i budować bazy danych repozytorium metadanych, interfejsów, interfejsu i raportów.
- * Większość licencjonowanych produktów z repozytoriami metadanych zawiera interfejsy, a większość zawiera pełny zestaw interfejsów API.
- * Jeśli produkt repozytorium metadanych jest certyfikowany dla narzędzia, w których znajdują się metadane, zapewni interfejsy narzędzi.

Wady:

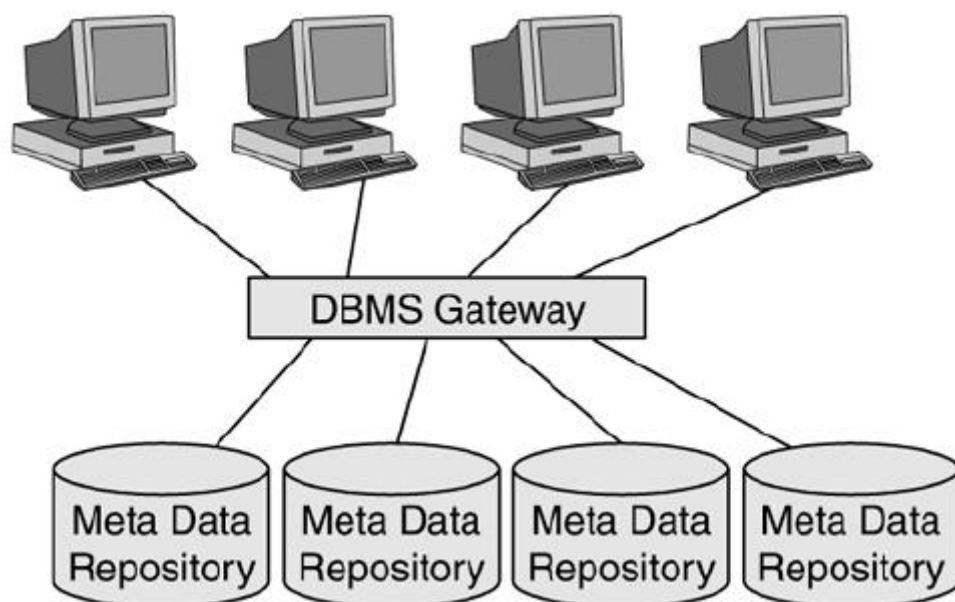
* Wersja „plain vanilla” licencjonowanego produktu prawdopodobnie nie spełni wszystkich wymagań dotyczących metadanych. Dlatego do obsługi i ulepszania licencjonowanego produktu potrzebny jest pełnoetatowy administrator.

* Będzie krzywa uczenia się, aby zapoznać się z architekturą produktu, interfejsami i interfejsami API.

* Im bardziej zaawansowany produkt repozytorium metadanych, tym jest droższy i tym więcej umiejętności potrzebują technicy aby go utrzymać.

Zdecentralizowane repozytorium metadanych

Jak wskazuje termin, zdecentralizowane rozwiązanie repozytorium metadanych przechowuje metadane w wielu bazach danych w wielu lokalizacjach



Powszechnie używane komponenty metadanych mogą być replikowane w wielu bazach danych, ale należy zachować szczególną ostrożność, aby zachować spójność tych komponentów. Brama kieruje wywołania dostępu do metadanych do odpowiedniej bazy danych w celu pobrania żądanych metadanych. To rozwiązanie ma kilka wyraźnych zalet i wad, niezależnie od tego, czy jest ono zbudowane we własnym zakresie, czy na licencji dostawcy

Zalety:

* Różni właściciele mogą oddzielnie zarządzać własnymi zestawami metadanych.

* Bazy danych repozytoriów meta są mniejsze i łatwiejsze w użyciu

ponieważ każda baza danych zawiera tylko te metadane

komponenty, które są interesujące dla określonej grupy biznesowej

ludzie.

- * Każde repozytorium danych meta może mieć własny model meta, czyli własny, dostosowany projekt.
- * Raporty można dostosować do każdego repozytorium danych meta.
- * Brama tworzy nazwę i lokalizację metadanych

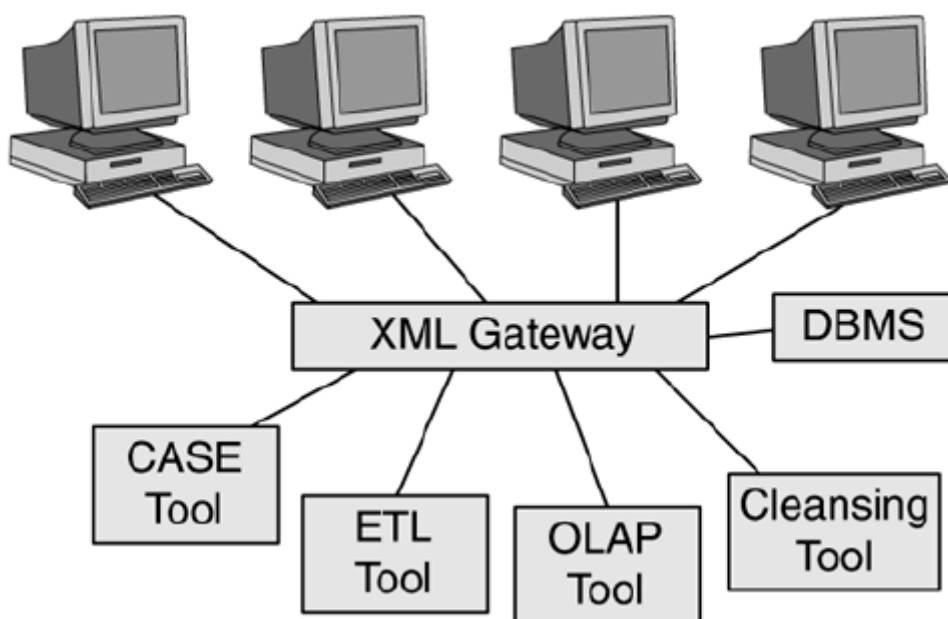
repozytorium przejrzyste dla osoby uzyskującej do niego dostęp.

Wady:

- * Kontrolowanie nadmiarowości w wielu repozytoriach metadanych i utrzymanie spójności metadanych jest trudne.
- * Utrzymanie i zarządzanie wieloma bazami danych na wielu platformach zajmie więcej czasu. Mogą również wystąpić problemy z synchronizacją z nowymi wydaniem DBMS.
- * Komunikacja między opiekunami różnych repozytoriów metadanych będzie musiała wzrosnąć. Ponadto będzie to wymagało utrzymania meta-modelu meta, który jest zintegrowaną (scaloną) ogólną architekturą wielu meta-modeli.
- * Powiązanie metadanych w różnych bazach danych może być trudne. Na przykład metadane biznesowe nie są automatycznie łączone z metadanymi technicznymi, jeśli znajdują się w różnych bazach danych.
- * Architektura tego rozwiązania jest bardziej skomplikowana, a nauczanie się korzystania z wielu baz danych o potencjalnie różnych projektach może być wysokie.

Rozproszone rozwiązanie metadanych z obsługą XML

Chociaż najbardziej obiecującą odpowiedzią na temat repozytorium metadanych jest rozproszone rozwiązanie metadanych z obsługą XML, jest ono również najtrudniejsze do wdrożenia, ponieważ przenosi koncepcję zdecentralizowanego repozytorium metadanych na wyższy poziom. Zamiast przechowywać metadane w wielu bazach danych, w rozwiązaniu obsługującym XML metadane pozostają w swojej pierwotnej lokalizacji, czyli w różnych słownikach narzędzi



Brama działa jak katalog do różnych lokalizacji, w których przechowywane są komponenty metadanych (np. tabele katalogu systemu DBMS lub słownik narzędzi ETL). Sprzedawcy energicznie badają to rozwiązanie „krwawiące”, ponieważ zmniejsza ono konieczność podwójnego utrzymywania metadanych. Podwójna konserwacja odnosi się do utrzymywania metadanych w źródłach źródłowych (DBMS i słowniki narzędzi) oraz utrzymywania ich w oddzielnej bazie danych repozytorium metadanych.

Zalety :

- * Tagi XML umożliwiają dostęp do metadanych w dowolnym typie przechowywania danych poprzez ustandaryzowaną kategoryzację i tagowanie składników metadanych.
- * Metadane nigdy nie muszą być duplikowane ani przenoszone z oryginalnego źródła (z wyjątkiem celów raportowania).
- * Brama sprawia, że lokalizacja metadanych jest przejrzysta dla osoba uzyskująca do niej dostęp.
- * Standardowe wyszukiwarki internetowe powinny być w stanie zlokalizować dowolne metadane w dowolnym miejscu.
- * Metadane i dane biznesowe mogą być łączone i przesyłane jednocześnie.

Wady:

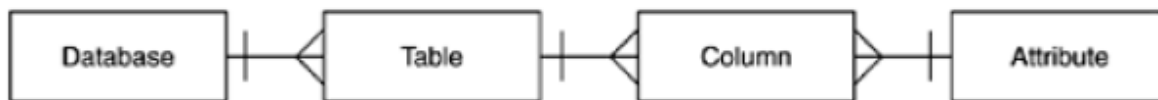
- * Wstępne tagowanie wszystkich metadanych za pomocą tagów XML jest procesem ręcznym i pracochłonnym. Ponadto tagowanie XML nie może być używane dla wszystkich metadanych.
- * Tagi XML zwiększają wymagania dotyczące przechowywania słownikowych baz danych, które przechowują metadane (DBMS i słowniki narzędzi).
- * Meta-metamodel musi być utworzony jako mapa wszystkich różnych typów przechowywania metadanych, z których każdy jest zaprojektowany zgodnie z własnym unikalnym metamodelem.
- * DBMS i dostawcy narzędzi muszą przestrzegać standardów branżowych dotyczących znaczników XML metadanych, aby umożliwić bezproblemowy dostęp do metadanych we wszystkich produktach. Należy obsługiwać wiele standardów.
- * Nie wszystkie DBMS i narzędzia obsługują XML. To najnowocześniejsza i niesprawdzona technologia.

Projektowanie repozytorium metadanych

Jeśli zostanie podjęta decyzja o stworzeniu własnego repozytorium metadanych we własnym zakresie, musisz wybrać między projektem E-R a projektem OO.

Projektowanie relacji encji

Ponieważ projekt E-R jawnie reprezentuje obiekty metadanych i ich relacje, a także ponieważ projekty E-R są intuicyjne i łatwe do zrozumienia, wiele organizacji wybiera ten typ projektu bazy danych dla swoich repozytoriów metadanych. Aby zilustrować intuicyjność projektu E-R, założmy, że fizyczny model meta zawiera cztery obiekty (baza danych, tabela, kolumna i atrybut) i że te obiekty są powiązane w kardynalności jeden-do-wielu, jak pokazano na rysunku



Ten rodzaj struktury bazy danych jest wystarczająco łatwy, aby zrozumieć, że ludzie biznesu obeznani z technologią mogą pisać własne zapytania ad hoc w języku SQL (Structured Query Language). Jeśli jednak te zapytania ad hoc są wykonywane w dużym scentralizowanym repozytorium metadanych, wydajność może stanowić problem. Napisanie słabo wydajnych zapytań SQL jest stosunkowo łatwe. Ponadto, ponieważ każdy obiekt metadanych jest zaimplementowany jako oddzielna tabela, będą dziesiątki tabel, a niektóre zapytania będą zawierać bardzo skomplikowane JOIN w wielu z tych tabel. Może to również wpłynąć na wydajność.

Zalety :

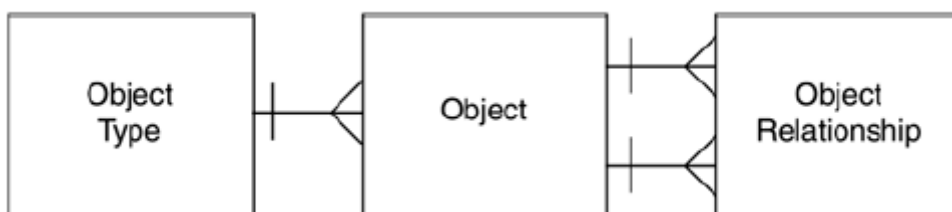
- * Projekty E-R są łatwe do odczytania i zrozumienia.
- * Ze względu na intuicyjny i jednoznaczny charakter projektu, zapytania można pisać za pomocą stosunkowo prostych instrukcji SQL.
- * Projekty E-R są łatwe do wdrożenia jako struktury relacyjnych baz danych.

Wady :

- * Zmiany i ulepszenia mogą wymagać przeprojektowania bazy danych, a także usunięcia i ponownego załadowania repozytorium metadanych.
- * Fizyczny model meta jest dość duży, z wieloma obiektami i wieloma relacjami, co sprawia, że architektura jest nieco złożona.
- * Obiekty metadanych i ich relacje muszą być bardzo dobrze zdefiniowane i rozumiane, aby fizyczny model meta był dokładny.

Projektowanie zorientowane obiektowo

Choć projekty E-R mogą być popularne w przypadku baz danych repozytoriów metadanych, projekty obiektowe są bardziej wydajne. Ponieważ są bardziej abstrakcyjne, skutkują mniejszą liczbą tabel, wydajniej uruchamiają zapytania i są znacznie łatwiejsze do rozwinięcia. Korzystając z tego samego przykładu, co powyżej, model OO zawierałby tylko trzy obiekty, ale obiekty te byłyby bardziej ogólne, jak pokazano na rysunku



Tego typu struktura bazy danych nie jest łatwa do zrozumienia, a ludzie biznesu prawdopodobnie nie będą w stanie napisać własnych zapytań SQL ad hoc w stosunku do niej. Nie jest intuicyjnie oczywiste, że obiekt o nazwie Object zawiera instancje (wiersze) dla wszystkich obiektów metadanych, takich jak instancje bazy danych, instancje tabel, instancje kolumn, instancje atrybutów i tak dalej. Nie jest

również oczywiste, że obiekt o nazwie Typ obiektu rozróżnia różne instancje obiektów metadanych przez przypisanie odpowiedniej etykiety bazy danych, tabeli, kolumny, atrybutu i tak dalej. A niewprawne oko miałoby jeszcze większe trudności z rozpoznaniem, że wszystkie relacje między tymi instancjami obiektów znajdują odzwierciedlenie w trzecim obiekcie o nazwie Relacja obiektu. Jednak łatwo zauważyć, że rozszerzenie tego typu ogólnego projektu jest tak proste, jak dodanie nowych wystąpień (wierszy) do tych trzech obiektów (tabel).

Zalety :

- * Projekty OO są niezwykle elastyczne; nie będą potrzebować żadnych przeprojektowań bazy danych, gdy konieczne będą zmiany.
- * Projekty OO są uproszczone, a zatem łatwe w utrzymaniu i ulepszaniu.
- * Projekty obiektowe są łatwe do zaimplementowania jako zorientowane obiektowo struktury baz danych.

Wady :

- * Ponieważ obiekt o nazwie Object zawiera wszystkie wystąpienia (wiersze) metadanych, tabela ta stanie się bardzo duża. Może to wpłynąć na wydajność dostępu.
- * Zapytania są znacznie trudniejsze do napisania i będą wymagały wielu rekurencyjnych sprzężeń. Wymagana jest zaawansowana znajomość SQL.
- * Projekty OO wymagają dużej krzywej uczenia się. Bardzo abstrakcyjny fizyczny model danych jest trudny do zrozumienia, a zrozumienie obszernych reguł wymaga czasu.

Projektowanie i budowanie własnego repozytorium metadanych może nie mieścić się w zakresie Twojego projektu BI. Może nie być budżetu ani personelu dla oddzielnego projektu repozytorium metadanych. Być może Twoja organizacja woląaby licencjonować (kupić) produkt z repozytorium metadanych. Podobnie jak w przypadku wszystkich gotowych produktów, licencjonowane repozytorium metadanych prawdopodobnie nie będzie idealnym rozwiązaniem, ale może być najbardziej opłacalne. Z pewnością byłoby to lepsze niż całkowite ignorowanie metadanych.

Licencjonowanie (kupowanie) repozytorium metadanych

Wybierając produkt z repozytorium metadanych (lub dowolny rodzaj produktu), nigdy nie należy zaczynać od pytania „Jaki jest najlepszy produkt tego typu na rynku?” Zamiast tego zawsze zaczynaj od następujących pytań:

- * Jakie są nasze wymagania?
- * Jakie wymagania to:
 - Obowiązkowe (musi mieć)
 - Ważne (korzystne mieć)
 - Opcjonalnie (miło mieć)
- * Które produkty spełniają nasze obowiązkowe wymagania?

* Które produkty spełniają nasze ważne wymagania?

Porównaj logiczny meta model każdego dostawcy (jeśli taki istnieje) lub ich fizyczny meta model (projekt produktu) z logicznym metamodelem i ustal, czy model dostawcy obejmuje wszystkie wymagania dotyczące metadanych odzwierciedlone w „metamodelu wymagań”. Model dostawcy musi co najmniej obsługiwać wszystkie obowiązkowe wymagania dotyczące metadanych. Jeśli tak, sprawdź, czy meta model i oprogramowanie dostawcy można rozszerzyć, aby można było dodać własne funkcje, aby spełnić ważne wymagania dotyczące metadanych. Możliwości rozbudowy produktów repozytorium metadanych powinny obejmować:

- * Dodawanie obiektów metadanych
- * Dodawanie relacji
- * Zmiana niewłaściwych relacji
- * Dodawanie atrybutów meta-metadanych do obiektów metadanych
- * Zmiana rozmiaru i długości komponentów meta-metadanych
- * Dostosowywanie raportów dostarczonych przez dostawcę
- * Tworzenie i przechowywanie kodu dla dodatkowych raportów
- * Importowanie metadanych z innych narzędzi
- * Eksportowanie metadanych do innych narzędzi

Ocena produktu

Użyj standardowych technik oceny, aby wybrać produkt repozytorium metadanych. Na przykład przygotuj listę kryteriów oceny produktów dla wymagań dotyczących metadanych i przypisz każdemu kryterium wagę od 1 do 10 (1 oznacza najmniej ważne, a 10 oznacza najważniejsze).

Zauważ, że „Produkt może spełnić nasze obowiązkowe wymagania dotyczące metadanych” nie jest kryterium ważnym. Jeśli produkt nie spełnia obowiązkowych wymagań, nawet nie bierz tego pod uwagę.

Oceń każdy produkt według wszystkich kryteriów oceny produktu, przypisując mu ocenę w skali od 0 do 10 (0 oznacza, że produkt nie ma tej cechy, 10 oznacza, że cecha produktu jest wyjątkowa).

Pomnóż oceny produktów przez współczynniki ważenia kryteriów, aby uzyskać końcowe oceny dla każdego produktu. Zsumuj wszystkie wyniki i ułóż produkty w kolejności od najwyższego wyniku całkowitego do najniższego wyniku całkowitego.

Ocena dostawcy

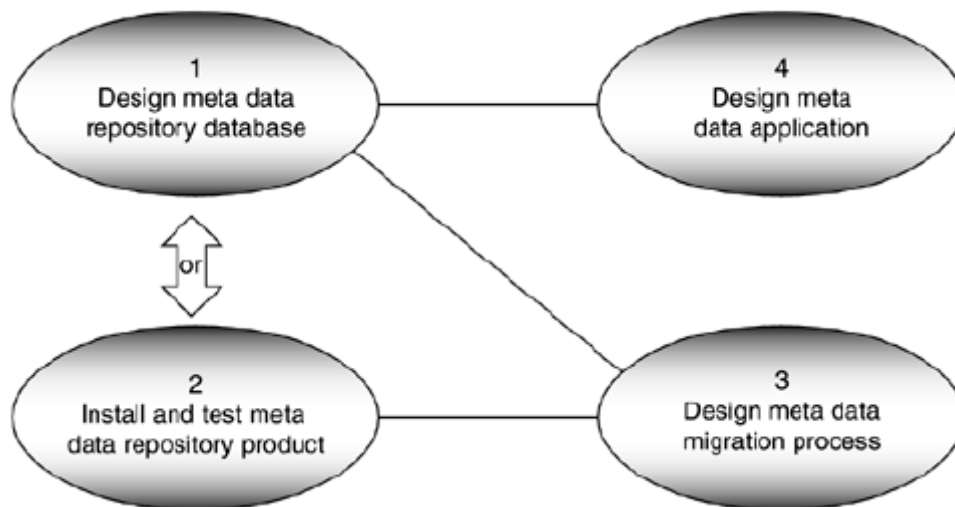
Większość organizacji nie poświęca wystarczającej ilości czasu, jeśli w ogóle, na ocenę dostawców oprócz oceny produktów. Ważne jest, aby zrozumieć stabilność firmy każdego dostawcy, jego zaangażowanie w produkt i poziom wsparcia. Utwórz listę kryteriów oceny dostawców i przypisz współczynnik wagowy do każdego kryterium w skali od 1 do 10 (1 oznacza najmniej ważne, a 10 najważniejsze). Inne kryteria, które należy wziąć pod uwagę, to reputacja w zakresie wsparcia, uczciwość dostawcy i wcześniejsze doświadczenie z dostawcą.

Oceń każdego dostawcę według wszystkich kryteriów oceny dostawców, przypisując ocenę w skali od 0 do 10 (0 oznacza, że dostawca nie może spełnić kryterium, 10 oznacza, że dostawca wyróżnia

się w tym kryterium). Pomnóż oceny dostawców przez współczynniki ważenia kryteriów, aby uzyskać końcowe oceny dla każdego dostawcy. Zsumuj wszystkie wyniki i wymień dostawców w kolejności od najwyższego wyniku całkowitego do najniższego wyniku całkowitego. Na koniec porównaj listy ocen produktów i dostawców i wybierz dwa najlepsze produkty. Sprawdź referencje dostawców, zaplanuj prezentacje produktów i umów się na 30-dniową instalację próbną przed dokonaniem ostatecznego wyboru produktu.

Działania związane z projektowaniem repozytorium metadanych

Czynności związane z projektowaniem repozytorium metadanych nie muszą być wykonywane liniowo. Rysunek wskazuje, jakie czynności można wykonywać jednocześnie.



Poniższa lista zawiera krótki opis czynności związanych z krokiem 10, projektowanie repozytorium metadanych.

1. Zaprojektuj bazę danych repozytorium metadanych. Jeśli zostanie podjęta decyzja o budowie repozytorium metadanych zamiast licencjonowania (kupowania), zaprojektuj bazę danych repozytorium metadanych. Wybierz między E-R

projekt i projekt obiektowy oraz stwórz (lub ulepsz) fizyczny model meta (projekt bazy danych). Wygeneruj DDL dla struktur bazy danych. Opracuj procedury konserwacji bazy danych repozytorium metadanych, takie jak tworzenie kopii zapasowych i odzyskiwanie, oraz utwórz plany wersjonowania i archiwizacji.

2. Zainstaluj i przetestuj produkt repozytorium metadanych. Jeśli zostanie podjęta decyzja o licencjonowaniu repozytorium metadanych, a nie o jego budowie, oceń produkty repozytorium metadanych oraz ich dostawców. Produkty i dostawcy z najwyższymi ocenami na karcie wyników powinni znaleźć się na krótkiej liście (dwa najlepsze wybory), z której zostanie wybrany produkt końcowy. Zainstaluj i przetestuj produkt repozytorium metadanych.

3. Zaprojektuj proces migracji metadanych. Zidentyfikuj wszystkie narzędzia i DBMS, z których trzeba będzie wyodrębnić metadane biznesowe i metadane techniczne. Określ możliwości importu, eksportu i interfejsu API tych narzędzi i systemów DBMS, a także produktu z repozytorium metadanych, jeśli posiadasz licencję. Zaprojektuj programy do migracji metadanych, w tym interfejsy narzędzi, i napisz specyfikacje programistyczne.

4. Zaprojektuj aplikację metadanych. Jeśli nie masz licencji na produkt z repozytorium metadanych, zaprojektuj aplikację metadanych, która obejmuje interfejsy dostępu, funkcje internetowe, raporty i funkcję pomocy online. Po wybraniu nośnika raportu (np. Portable Document Format [PDF], Hypertext Markup Language [HTML]) przygotuj specyfikacje programowe dla różnych typów aplikacji.

Rezultaty wynikające z tych działań

Jeśli licencjonujesz produkt z repozytorium metadanych, twoim rezultatem jest zainstalowany i przetestowany produkt. Jeśli projektujesz własne repozytorium metadanych, powinieneś stworzyć następujące elementy projektu:

1. Fizyczny model meta. Fizyczny model meta to diagram fizycznych struktur bazy danych dla repozytorium metadanych. W zależności od wybranego schematu projektu bazy danych ten diagram może być modelem E-R lub modelem OO. Pokazuje tabele, kolumny, klucze podstawowe, klucze obce, licznosc i reguły integralności referencyjnej.
2. Język definicji danych dla repozytorium metadanych. Język definicji danych (DDL) to zestaw instrukcji SQL, które informują DBMS, jakie typy fizycznych struktur baz danych należy utworzyć dla repozytorium metadanych, takich jak bazy danych, obszary tabel, tabele, kolumny i indeksy.
3. Język kontroli danych dla repozytorium metadanych. Język kontroli danych (DCL) to zestaw instrukcji SQL, które informują SZBD, jaki typ dostępu do odczytu/zapisu do repozytorium metadanych ma udzielić. Dostęp można przyznać osobie, grupie osób, programowi lub narzędziu.
4. Specyfikacje programowania meta repozytorium danych. W przypadku procesu migracji metadanych te specyfikacje programistyczne powinny określać logikę programowania dla wyodrębniania, przekształcania i ładowania programów metadanych, a także interfejsów narzędzi. W przypadku aplikacji metadanych te specyfikacje programu powinny określać logikę programowania dla raportów i zapytań dotyczących metadanych, interfejsów dostępu oraz funkcji pomocy online.

Role zaangażowane w te działania

* Architekt infrastruktury BI. Ponieważ architekt infrastruktury BI ponosi ostateczną odpowiedzialność architektoniczną za całe środowisko wspomagania decyzji BI, musi on/ona przejrzeć wszystkie działania projektowe. Jeśli repozytorium metadanych jest licencjonowane, architekt infrastruktury BI może uczestniczyć w przygotowaniu kryteriów oceny i decydowaniu o współczynnikach ważenia kryteriów. Będzie również zaangażowany w ostateczny wybór produktu.

* Administrator danych. Administrator danych będzie współpracował z administratorem metadanych w zakresie wymagań dotyczących metadanych i pomoże administratorowi metadanych w działaniach związanych z modelowaniem danych. Administrator danych może być również skutecznym łącznikiem między administratorem metadanych, ekspertem w danej dziedzinie i przedstawicielem biznesowym, gdy wymagane jest zweryfikowanie wymagań dotyczących metadanych lub przejrzanie lub zakomunikowanie decyzji projektowych.

* Administrator metadanych. Administrator metadanych ponosi główną odpowiedzialność za metadane i repozytorium metadanych. Jeśli repozytorium metadanych jest budowane we własnym zakresie, odpowiada za jego zaprojektowanie i rozwój. Jeśli repozytorium metadanych jest licencjonowane, odpowiada za jego instalację, testowanie, ulepszanie i utrzymanie. Administrator metadanych jest również odpowiedzialny za projektowanie interfejsów dostępowych oraz programów migracji do DBMS, narzędzia CASE, narzędzia ETL, narzędzia OLAP i innych narzędzi.

Ryzyko niewykonania kroku 10

Dostarczenie rozwiązania repozytorium metadanych nie jest przypadkowym przedsięwzięciem. Te same dyscypliny i rygor, które dotyczą tworzenia aplikacji BI, dotyczą również tworzenia repozytorium metadanych. „Wyrzucenie” bazy danych i kilku gotowych zapytań SQL nie jest równoznaczne z trwałym rozwiązaniem repozytorium metadanych. Jak każdy inny system, musi być zaprojektowany z dużą dozą przemyśleń i przewidywania, aby zapewnić pożądaną poziom funkcjonalności, wydajności, skalowalności i łatwości konserwacji w czasie. Jeśli podjęto decyzję o licencjonowaniu produktu repozytorium metadanych, należy zadbać o proces oceny z taką samą dbałością, jak przy zakupie krytycznego pakietu systemów operacyjnych. Jeśli nie poświęcisz czasu na zaprojektowanie solidnego i trwałego rozwiązania do repozytorium metadanych, będziesz musiał przerobić swoje rozwiązanie lub skończyć z gorszym środowiskiem wspomagania decyzji BI. Jak mówi przysłowie: „Zapłać mi teraz albo zapłać później, ale zapłacisz”.