

Master Data Management

– fundamentem zarządzania jakością danych

Jedną z podstawowych zasad, które poznajemy na początku informatycznej edukacji jest „garbage in – garbage out”. Ta fundamentalna prawda mająca zastosowanie nie tylko dla świata IT warta jest wspomnienia w kontekście jakości danych w złożonych systemach informatycznych.

W dyskusjach ze znajomymi, czy też obserwując naszą rzeczywistość polityczną, bardzo często mamy gotowe rozwiązania na lepsze zorganizowanie otaczającego nas świata. Każdy z nas musiał z konieczności doświadczyć w tym roku zmian w sposobie zarządzania odpadami. Skala chaosu oraz szkody dla środowiska zmusiły rząd do wprowadzenia centralnych, wymuszonych odgórnie zasad zbierania śmieci. Będąc CIO możemy zrobić to (przynajmniej na „naszym podwórku”) znacznie lepiej – możemy wyleczyć przyczynę a nie skutki.

Gwałtowny rozwój skali wykorzystania i złożoności systemów informatycznych, bardzo często przypomina „dziki” rozwój gospodarki rynkowej. Nasze „ustawy”: SOA, EIA, itp. nie są w stanie nadążyć za coraz to nowszymi pomysłami dotyczącymi zastosowania IT w biznesie. Bardzo często nie ma możliwości na dobre dopasowanie kolejnego elementu do aktualnej architektury. W rezultacie rośnie ilość wyzwań związanych z ich integracją. Rozwiązania „post factum” w typie centralnych hurtowni danych dają dobre rezultaty końcowe,

ale zwiększają ilość problemów związanych z integracją powstałych wcześniej danych.

Zamiast walczyć z lawinowym wzrostem ilości „śmieci” w naszym informatycznym krajobrazie, warto skupić się na metodach ograniczania możliwości ich powstawania. Należy stworzyć mechanizmy, które niezależnie od zmian w środowisku IT będą ograniczały „produkcję” błędnych lub niezintegrowanych danych. Jest to najważniejsze dla kluczowych danych wykorzystywanych wspólnie przez większość naszych systemów. Dane te określane są jako Master Data, a rozwiązaniem jest koncepcja Master Data Management (MDM).

MDM to fundament inicjatyw zarządzania jakością danych. W dużym uproszczeniu można traktować MDM jako kręgosłup spinający całość procesów powstawania i wykorzystywania danych we wszystkich systemach IT – od ich powstawania do końcowej konsumpcji w systemach zarządczych. Ważne jest, aby nie zaszufladkować go jako kolejnego rozwiązania jedynie informatycznego. MDM to określenie procesów i odpowiedzialności w odniesieniu do

danych, wpływających na rozumienie całości biznesu odzwierciedlonego w systemach IT.

Tradycyjny podział na role, w których IT jest traktowane jako dostawca informacji dla biznesu konsumującego te informacje, prowadzi wprost do kumulacji opisywanych wyżej problemów. Kwestii jakości danych nie można rozwiązać bez zaangażowania użytkowników biznesowych w zorganizowany proces zarządzania kluczowymi danymi. Wynika to głównie z faktu, iż IT ma świadomość jak powinien wyglądać proces utrzymywania jakości danych, a biznes wie w jaki sposób zarządzać danymi, które są niezbędne do efektywnego działania. Wdrożenie MDM musi być powiązane ze zmianami organizacyjnymi pozwalającymi na powstanie struktur wspólnie odpowiedzialnych za radzenie sobie z tym problemem. Dobrą praktyką jest ewolucyjna metoda stopniowych zmian.

Od czego zacząć?

Oczywiście, nie można twierdzić, że jakość danych jest sprawą zupełnie nową lub powszechnie zaniedbaną. Bardzo wiele organizacji przykładą dużą wagę do tych problemów



i stosuje szereg metod zapobiegawczych. Najczęściej są to rozwiązania rozwijane wraz z systemami integrującymi dane z wielu źródeł, czyli Hurtowniami Danych i Business Intelligence. Po wdrożeniu HD i BI organizacja sporo już wie na temat jakości danych i problemów z niejednoznacznością ich interpretacją. Na tym etapie muszą (a przynajmniej powinny) być zbudowane centralnie słowniki pozwalające na uniwersalną kategoryzację i interpretację pojęć biznesowych. Jest to dobra podstawa do budowy pierwszego, łatwiejszego kroku: budowy **analitycznego MDM**.

Mając powyższe „hurtowniane” doświadczenia, wdrożenie analitycznego MDM jest zadaniem osiągalnym dla każdej organizacji. Należy dodać do bieżących rozwiązań narzędzia oraz procesy łączące słowniki z procedurami zarządzania jakością i spójnością danych (Data Governance). Bardzo ważne jest umiej-

scowienie odpowiedzialności za te procesy i ich rezultaty. Aby móc mówić o rozwoju MDM muszą być one realizowane bezpośrednio przez użytkowników biznesowych lub dedykowane, stworzone w tym celu jednostki. Może to być sugerowana przez Gartnera Business Intelligence Competence Centre (BICC) lub inna, nawet wirtualna jednostka. Jeżeli kwestie te pozostaną w domenie IT – będzie to jedynie dalszy rozwój tradycyjnego systemu, bez szans na skuteczne wdrożenie MDM.

Na tym etapie biznes będzie w stanie przejąć odpowiedzialność za własne dane na poziomie analitycznym, czyli będzie

mógł zidentyfikować problemy a następnie stworzyć reguły i procesy korygujące. W analitycznym MDM będzie to wykonalne w obszarze zasilania systemów raportowo-analitycznych (HD/BI). Analityczny MDM nie da jeszcze pełnych możliwości zapobiegania problemom u źródła – w miejscach powstawania danych. Częściowo będzie to jednak możliwe, tam gdzie analiza jakości danych umożliwi diagnozę przyczyn i łatwe ich usunięcie. Na przykład błędem wynikającym ze źle wprowadzanych danych można zapobiegać zamieniając interfejs użytkownika lub poprzez dodatkowe szkolenia dla obsługujących systemy źródłowe.

” MDM to określenie procesów i odpowiedzialności w odniesieniu do danych, wpływających na rozumienie całości biznesu odzwierciedlonego w systemach IT

Operacyjny MDM

Wyniki i obserwacje z wykorzystania analitycznego MDM powinny posłużyć jako podstawa do wdrożenia **operacyjnego MDM**. W operacyjnym wydaniu, MDM pełni już znacznie bardziej aktywną rolę. Poza samą obserwacją i analizą jakości danych, jest głównym systemem, w którym realizowane są procesy zarządzania danymi podstawowymi. W MDM określone są m.in. hierarchie, atrybuty, parametry źródłowe, itp. konieczne do zapewnienia spójności i wysokiej jakości danych. System umożliwi też łączenie wykorzystywanych we wszystkich systemach danych ze źródłami ich powstawania.

Ważne dla operacyjnego podejścia do MDM jest określenie i realizacja procesów zarządzania danymi podstawowymi. Powinny one bazować na wynikach zebranych przez komórkę MDM. Analogicznie od pierwszego kroku kluczowe jest, aby zaangażować właściwych użytkowników biznesowych. Za jednoznaczną i właściwą interpretację danych musi odpowiadać związany z nimi biznesowy właściciel obszaru danych. Zazwyczaj bywa on określany

mianem *data steward* lub *data officer*. Niezależnie od nazwy, ważne jest, aby mógł on na podstawie realnych obserwacji biznesowych proponować obszary zmian procesowych tworzących operacyjny MDM. Im bardziej biznes będzie uczestniczył w procesie MDM tym większe będzie miał zaufanie do danych, zaś koszty IT będą się obniżać – zwłaszcza te obszary transformacji i zasilania danych, które bazują dziś na milionach linii kodu SQL.

Pełny zestaw referencyjnych danych podstawowych musi być następnie udostępniany innym systemom i użytkownikom (subskrybentom MDM). W tym wydaniu MDM jest centralnym źródłem wszystkich słowników danych (choć zależność od przyjętego stylu implementacji nie musi być miejscem ich powstawania) i ich właściwej interpretacji biznesowej. Trzeba przy tym pamiętać o zapewnieniu odpowiednich reguł bezpieczeństwa. W dalszej perspektywie rozwoju procesy MDM mogą być rozszerzone o elementy automatycznej detekcji i obsługi zmian w systemach źródłowych dla danych podstawowych.

” Oczekiwany efektem rozwoju operacyjnego MDM powinien być wzrost jakości danych, automatyzacja procesów zarządzania danymi podstawowymi i tym samym stopniowy zanik znaczenia analitycznego MDM

Podsumowanie

Oczekiwany efektem rozwoju operacyjnego MDM powinien być wzrost jakości danych, automatyzacja procesów zarządzania danymi podstawowymi i tym samym stopniowy zanik znaczenia analitycznego MDM. Docelowo, wdrożenie MDM powinno wygenerować zmiany w organizacji, które wyodrębnią jednostki biznesowe odpowiedzialne za jakość danych podstawowych. Jednostka MDM będzie mieć do dyspozycji zintegrowany system IT umożliwiający:

- Wykrywanie zmian w systemach źródłowych;
- Zapewnienie odpowiedniej jakości danych podstawowych;
- Umożliwienie odpowiednich transformacji i parametryzacji danych;
- Dostarczenie danych w odpowiedniej formie do wszystkich wykorzystujących je systemów.

Przy takim podejściu możliwe będzie również łatwe spełnienie jakichkolwiek zewnętrznych wymogów dotyczących jakości danych, np. takich jak wymagania nowej rekomendacji D KNF. Ostatecznym celem wdrożenia MDM jest takie umiejscowienie procesów zarządzania jakością, aby stały się częścią codzienności, niezauważalną dla organizacji – to długa droga, najlepiej zatem niezwłocznie zacząć od pierwszego małego kroku w kierunku MDM. □

Radosław Kozieja
wiceprezes zarządu
Codec Systems Sp. z o.o.
rkoziejac@codec.pl

Innowacyjność w sektorze ICT

Przy okazji myślenia o innowacyjności, zwłaszcza w obszarze ICT, jednym z pierwszych skojarzeń jest Dolina Krzemowa i marzenie aby jej odpowiednik mógł powstać w Polsce. I chociaż szanse na powtórzenie sukcesu nie są duże, jako że Silicon Valley stała się samonapędzającym mechanizmem, to zapoznawanie się i inteligentne naśladowanie jej praktyk ma sens i tą drogą idzie inicjatywa Top 500 Innovators. Z drugiej strony, jeśli porówna się potencjał innowacyjności polskiej branży ICT i jej realne osiągnięcia można zobaczyć ile jest jeszcze do zrobienia.

