

Zuzanna Jarosz

*Instytut Uprawy Nawożenia i Gleboznawstwa – Państwowy Instytut Badawczy
w Puławach*

JAKOŚĆ I INTEGRALNOŚĆ DANYCH W SYSTEMACH WSPOMAGANIA DECYZJI*

Słowa kluczowe: jakość danych, hurtownia danych, system wspomaganie decyzji

Wstęp

Systemy wspomaganie podejmowania decyzji stanowią intensywnie rozwijającą się dziedzinę informatyki. Systemy Wspomaganie Decyzji (SWD) są stosowane w obszarach, gdzie rozpatrywane problemy decyzyjne są szczególnie złożone i do ich rozwiązania potrzebne jest pozyskanie danych z różnych źródeł oraz skorzystanie z doświadczenia i wiedzy ekspertów z analizowanej dziedziny lub z innych dziedzin wpływających na rozwiązanie problemów decyzyjnych.

Zdobywanie przewagi konkurencyjnej wymaga przyspieszenia procesu podejmowania decyzji. Kluczowe w tym procesie jest posiadanie właściwej i łatwo dostępnej informacji. Obecnie systemy zarządzania zawierają dane, które mogą dostarczyć informacji niezbędnej do podjęcia decyzji. Jednak najczęściej dane te są rozrzucone po wielu różnych systemach, platformach i miejscach, co sprawia, że zapewnienie integralności danych oraz dostępu do nich w krótkim czasie jest niemożliwe. Celem ujednoczenia danych tworzy się hurtownie danych. Natomiast do analizowania danych zgromadzonych w hurtowni służą narzędzia SWD.

Rozwój Systemów Wspomaganie Decyzji zaowocował szeroką gamą dostępnych rozwiązań oraz szerokim spektrum zastosowań. Celem opracowania jest przedstawienie zgodnej z zaleceniami inżynierii oprogramowania podstawowej struktury SWD, podział na moduły i powiązania występujące między modułami. Szczególną uwagę zwrócono na jakość danych pochodzących z różnych źródeł i wykorzystywanych w systemie. Przedstawiono znaczenie procesu gromadzenia, ekstrakcji, trans-

* Opracowanie wykonano w ramach zadania 4.1 w programie wieloletnim IUNG-PIB.

formacji i integralności danych oraz ich wpływu na jakość informacji uzyskiwanych z systemu.

Systemy Wspomagania Decyzji (SWD)

SWD są zorganizowanym zbiorem procedur, baz danych i urządzeń wykorzystywanych w celu wspomagania podejmowania decyzji na wszystkich etapach tego procesu, poczynając od rozpoznania, czyli zdefiniowania problemu i zaklasyfikowania go do określonej grupy standardowej, następnie poprzez wybór odpowiednich danych stworzenie i analizę modelu informacyjnego opisującego rzeczywistość, a kończąc na pomocy w generowaniu wariantów dopuszczalnych rozwiązań oraz w wyborze najlepszego rozwiązania.

Systemy Wspomagania Decyzji stosowane są w sytuacji, gdy podjęcie decyzji jest zadaniem skomplikowanym, kiedy mamy do czynienia z problemami słabo strukturalizowanymi (2). Systemy te dotyczą wspomagania, a nie automatyzacji decyzji. Tym samym ich celem jest podnoszenie skuteczności, a nie sprawności zarządzania.

SWD pozwalają na realizację następujących zadań:

- wyszukiwanie danych jednostkowych, czyli wyodrębnionych ze zbiorów danych,
- swobodny dostęp do danych oraz ich analizę przyczynowo-skutkową,
- dostarczanie danych zbiorczych wcześniej zdefiniowanych,
- przygotowywanie projektów możliwych decyzji,
- przedstawianie konsekwencji (ocenę) proponowanych decyzji przy wykorzystaniu modeli obliczeniowych i symulacyjnych: „co - jeżeli?”,
- określanie danych problemu niezbędnych do realizacji określonego celu: wykonywanie analiz sterowanych celami,
- wybranie wariantu decyzji na podstawie zadanych kryteriów.

Biorąc pod uwagę powyższe cele, można wyodrębnić kilka poniższych cech SWD.

- **Zakres zastosowań.** SWD koncentrują się na wspomaganiu rozwiązywania rzeczywiste powstających problemów decyzyjnych. Są wyspecjalizowane w kierunku i tylko w kierunku podejmowania ściśle określonych decyzji. Skupiają się więc na konkretnym problemie i wspierają przede wszystkim pojedynczych decydentów i niewielkie grupy, a dopiero w dalszej kolejności całą organizację.
- **Konieczność wykonywania skomplikowanych, wyrafinowanych analiz i porównań, z wykorzystaniem zaawansowanych pakietów oprogramowania.** Zadania stawiane SWD wymagają zwykle zastosowania znacznie bardziej skomplikowanych algorytmów działania niż w przypadku tradycyjnie rozumianego przetwarzania danych. W związku z tym tworzone są one często w specjalistycznych środowiskach programistycznych dostarczających odpowiednich procedur

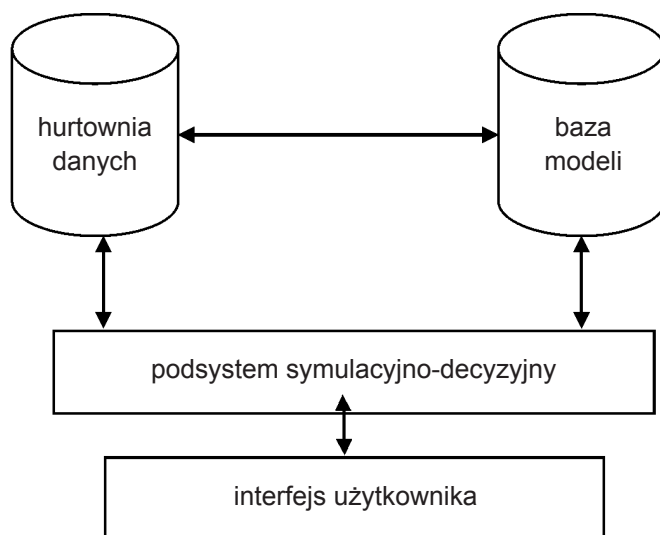
obliczeniowych (takich jak np. SAS, Matlab). SWD może również stanowić platformę integrującą dla samodzielnych zewnętrznych programów analitycznych.

- **Możliwość przetwarzania dużych ilości danych.** SWD wyposażone są w możliwości przeszukiwania obszernych baz danych, co pozwala na integrację tradycyjnego przetwarzania danych z metodami badań operacyjnych, ułatwiając kadrze menedżerskiej stosowanie ilościowych technik zarządzania.
- **Pobieranie i przetwarzanie danych z różnych źródeł.** Niektóre dane mogą rezydować w bazach danych na komputerach osobistych, kolejne mogą być ulokowane w innych systemach operacyjnych lub sieciowych. Dane mogą być przechowywane również w arkuszach kalkulacyjnych lub innych specyficznych formatach, związanych z np. oprogramowaniem analitycznym. SWD powinien mieć możliwości pozwalające na integrację danych z jak największej liczby źródeł.
- **Tworzenie raportów i elastyczność prezentacyjna.** Jedną z przyczyn powstania SWD był fakt, że systemy transakcyjne i systemy informatyczne zarządzania nie były dostatecznie elastyczne do zaspokojenia wszystkich potrzeb informacyjnych i problemów decydentów. Wyjściem z tego typu systemów są zwykle drukowane raporty o ustalonej strukturze i formacie. W przypadku Systemów Wspomaganie Decyzji otrzymuje się informację w formacie dostosowanym do wymagań użytkowników. SWD opierają się przede wszystkim na interakcyjnych raportach na ekranie komputera pozwalających na nawiązanie dialogu użytkownika z systemem i współpracę z nim w trybie on-line. W zależności od preferencji osoby rozwiązującej problem wyjście może być prezentowane oczywiście również w formie drukowanej.
- **Orientacja tekstowa i graficzna.** Sposób prezentacji informacji wyjściowych zwykle jest bardzo elastyczny. Obejmuje on zarówno format tekstowy, jak i graficzny (a coraz częściej również multimedialny). Współczesne SWD mogą tworzyć informacje tekstowe, tabelaryczne, graficzne. Stosowanie elastycznej techniki prezentacyjnej pozwala na lepsze zrozumienie sytuacji i łatwiejszą interpretację wyników działania systemu.
- **Elastyczność i adaptacyjność.** Systemy Wspomaganie Decyzji dostosowują się do zmian, jakie zachodzą w otoczeniu decydenta, umożliwiając indywidualne podejście do problemu decyzyjnego. Użytkownik często ma możliwość ingerencji w strukturę wewnętrzną SWD i konfigurowania jej w zależności od swoich indywidualnych potrzeb i uwarunkowań sytuacji decyzyjnej.

Systemy Wspomaganie Decyzji tworzone są z myślą o rozwiązywaniu konkretnych, indywidualnych problemów. W związku z tym trudno jest mówić o jakimś uniwersalnym, jednolitym wzorcu ich budowy. SWD mają silnie zróżnicowaną strukturę wewnętrzną, w zależności od problemu, dla którego zostały one stworzone oraz preferencji użytkownika odnośnie działania systemu. Niemniej jednak możemy

wyróżnić pewien ramowy schemat ich budowy (rys. 1) obejmujący zwykle kilka standardowych podsystemów:

- **hurtownia danych** zawiera dane dotyczące działalności organizacji i jej otoczenia. Z tego powodu możemy czasami mówić o bazie danych wewnętrznych i zewnętrznych. Dane wewnętrzne pochodzą przede wszystkim z baz danych transakcyjnych oraz innych systemów informatycznych. Dane zewnętrzne pochodzą zwykle od otoczenia gospodarczo-politycznego organizacji. Działanie SWD w sposób kluczowy zależy od jakości danych, tak więc powinny być one uważnie kontrolowane.
- **baza modeli** składa się z wielu modułów, z których każdy zawiera opis odpowiednich zachowań związanych z daną sytuacją decyzyjną. Modele te wspierają podejmowanie decyzji na różnych poziomach zarządzania w zakresie różnych funkcji kierowniczych i w różnych dziedzinach działalności obiektu. Z tego powodu ważne jest zapewnienie szczególnie dla tego elementu SWD możliwości ciągłej modyfikacji i rozbudowy. Ta część systemu decyduje bowiem o rzeczywistych możliwościach całego systemu.
- **podsystem symulacyjno-decyzyjny** na podstawie żądań użytkownika oraz istniejących danych dokonuje wyboru kombinacji modeli niezbędnych do rozwiązania zadania, wyboru danych wejściowych dla tych modeli oraz wykonuje za ich pomocą niezbędne obliczenia. Zauważmy, że przepływy informacyjne tego podsystemu z bazą danych mają charakter dwukierunkowy. Może on nie tylko pobierać dane wejściowe, ale również zapisywać w bazie danych informacje będące wynikiem działania SWD. Podobnie w przypadku bazy modeli podsystem symulacyjno-decyzyjny może również modyfikować modele pod kątem konkretnego problemu oraz istniejących danych (np. poprzez reestymację ich parametrów), zapisując zmiany w bazie modeli.
- **interfejs użytkownika** (menedżer dialogu) ma zapewnić wysoki komfort obsługi. Pozwala decydentowi na łatwy dostęp i manipulowanie SWD. Steruje pracą modułu symulacyjno-decyzyjnego oraz przekazuje użytkownikowi otrzymane od niego wyniki obliczeń. Użytkownik musi otrzymywać to, czego zażąda w możliwie różnorodnej formie, a system musi być przygotowany na żądania niestandardowe. Nowoczesne interfejsy mają za zadanie jak największe uproszczenie sposobu komunikacji z systemem poprzez komunikację graficzną, multimedialną, z wykorzystaniem powszechnie stosowanej terminologii biznesowej, a nawet z użyciem języka naturalnego.



Rys. 1. Struktura Systemu Wspomagania Decyzji

Źródło: opracowanie własne

W celu dostarczenia niezbędnych informacji wypracowujących decyzje projektuje się hurtownie danych.

Hurtownia danych

Rozwój technologii informatycznych, a w szczególności baz danych, spowodował, że w każdej firmie czy instytucji gromadzone są różne dane na różnych etapach działalności. Powstały liczne systemy ukierunkowane na konkretne działania (kadrowe, księgowo, magazynowe). Często systemy te nie są ze sobą powiązane. Gromadzone dane wykorzystywane są do wspomaganie tylko bieżących działań. W różnych formatach przechowywane są te same dane, które po pewnym czasie stają się niepotrzebne. Tradycyjne systemy bazodanowe ukierunkowane na realizację wielu małych i prostych zapytań, mające zapewnić realizację bieżących działań nazwano OLTP (On – Line Transaction Processing) (1). Konieczność pozyskiwania dużej ilości danych analitycznych przyczyniła się do powstania systemów informatycznych typu OLAP (On – Line Analytical Processing). Podstawowe cech tych systemów to:

- przechowywane dane są zorientowane tematycznie,
- bardzo duże ilości gromadzonych danych,
- realizowane są bardzo złożone zapytania operujące na dużej ilości danych,
- przechowywane są dane bieżące i historyczne,
- przechowywane są dane elementarne i zagregowane,
- wykonywane są głównie operacje dopisywania danych, brak ich modyfikacji.

Elementem łączącym te dwa typy systemów są wyspecjalizowane bazy danych gromadzące w specjalnie zaprojektowanych strukturach dane historyczne zwane hurtowniami danych. „Hurtownia danych jest tematycznie zorientowaną, zintegrowaną, zmienną w czasie, nieulotną (tylko do odczytu) kolekcją danych służącą wsparciu podejmowania decyzji kierownictwa” (3).

Zorientowanie na temat oznacza, że zbierane dane dotyczą tematu (np. sprzedaży), a nie działań (np. zbierania zamówień).

Nieulotność oznacza, że dane raz umieszczone w hurtowni zazwyczaj pozostają niezmienione. Każdy użytkownik bazy danych ma pewność, że zapytanie zawsze zwróci taki sam wynik, niezależnie od tego, jak często jest wykonywane.

Zintegrowanie oznacza, że dane są jednolite, czyli na przykład daty przechowywane są zawsze w tym samym formacie, znaki kodowane są w ten sam sposób, pola zawierające tę samą informację mają tę samą postać.

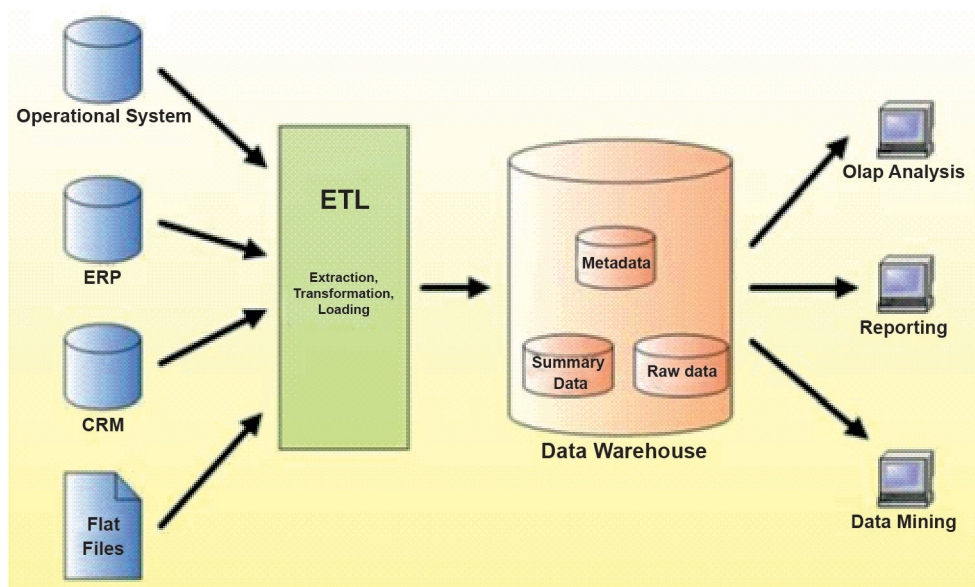
Zróżnicowanie czasowe oznacza, że gromadzone są dane historyczne. Prawie wszystkie zapytania kierowane do hurtowni danych wymagają prześledzenia jakiegoś odcinka czasu.

Podstawowym celem hurtowni danych jest zcentralizowanie oraz ujednoczenie danych pochodzących z różnych systemów źródłowych wspierających zadania operacyjne przedsiębiorstwa. Dodatkowo dane przetworzone są z postaci transakcyjnej na formę analityczną, znacznie lepiej dostosowaną do przekrojowego badania trendów oraz wszelkiego rodzaju analiz.

Typowa hurtownia danych składa się z następujących komponentów (9):

- źródła danych,
- obszar wstępnego przetwarzania i kontroli jakości danych (staging),
- obszar łączenia danych oraz transformacji do postaci analitycznej (data warehouse),
- obszar eksploatacji danych do użytkownika końcowego (data mart).

Na rysunku 2 w sposób schematyczny została przedstawiona architektura hurtowni danych. Oczywiście konkretne rozwiązania mogą się różnić od poniższego, ale ten schemat ma za zadanie przybliżyć ogólną ideę.



Rys. 2. Ogólna architektura hurtowni danych

Źródło: <http://datawarehouse4u.info> (13)

Jak widzimy, do hurtowni danych dane pobierane są z zewnętrznych źródeł danych. Zanim dane znajdują się w hurtowni muszą zostać pobrane (*extraction* – ekstrakcja), podane transformacji (*transformation* – transformacja), czyli zwalidowane (*validation* – walidacja), zintegrowane (*integration* – integracja) i odwzorowane (*mapping* – mapowanie, odwzorowanie), a w końcu załadowane do hurtowni (*loading* – ładowanie).

Proces **ekstrakcji** polega na wybraniu informacji, które mają trafić do hurtowni, a następnie pozyskaniu tych informacji z baz źródłowych. Pierwsze pytania, jakie należy sobie zadać: **co i skąd** pobrać. Odpowiedzi na te pytania nie udzieli żadne narzędzie, musi to zrobić projektant, analizując dostępne źródła. Następne pytanie, to **kiedy** pobierać dane. Kolejny kluczowy problem, to **jak** pobrać dane. Skala trudności pobierania danych zależy od rodzaju źródła. Niektóre źródła są odpytywalne (np. przez SQL), inne wymagają specjalnych programów do ekstrakcji z danych surowych (wrappers).

Na **transformację** danych składają się wszelkie operacje dostosowujące treść i format danych do potrzeb hurtowni. Są to między innymi:

- wypełnianie pustych wartości,
- zmiana formatu (daty, liczby),
- zmiana wartości (np. przeliczanie jednostek),
- ujednolicanie wartości (np. na podstawie słowników),
- utrzymanie integralności danych (więzy).

Zanim dane trafią do hurtowni muszą zostać oczyszczone, zintegrowane i odwzorowane. Przykładowo operacjami czyszczenia danych są:

- wykrywanie i (w miarę możliwości) poprawianie błędów literowych, słownikowych,
- wykrywanie i usuwanie niezgodności między nazwą atrybutu a zawartością,
- normalizacja wartości, np. zamiana pustych ciągów znaków „ ” czy „spacji” na NULL,
- uzupełnienie danych na podstawie zewnętrznych źródeł informacji.

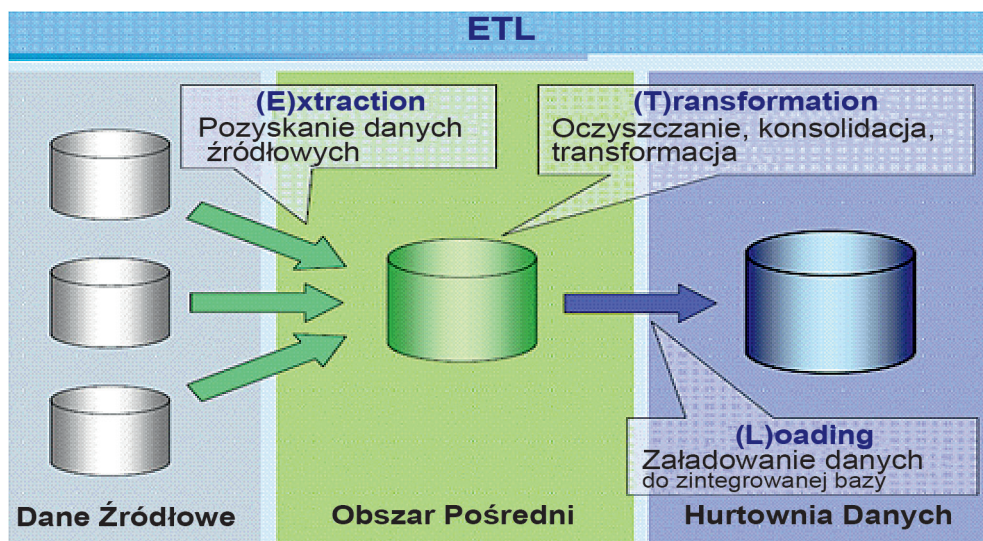
Kolejnym krokiem procesu transformacji jest integracja, która obejmuje integrację formatów oraz integrację semantyczną. Zintegrowanie danych jest konieczne przede wszystkim dlatego, że dane pochodzą z wielu źródeł. Ostatnim krokiem procesu transformacji jest odwzorowywanie (mapowanie), czyli przekształcanie danych do formatu docelowego. Zadaniem tego procesu jest ułatwienie wprowadzania zmian, np. przy zmianie jednego systemu źródłowego na inny.

Ładowanie danych do hurtowni to problem przede wszystkim techniczny. Wyodróżniamy trzy podstawowe metody ładowania danych:

- przekształcanie i ładowanie rekord po rekordzie – jest to najprostsza metoda ładowania danych, jednak mało wydajna w warunkach dużych hurtowni danych,
- zewnętrzna obróbka (scalanie, sortowanie) i ładowanie gotowych danych, często z wykorzystaniem ODS-ów (magazynów danych operacyjnych stanowiących warstwę pośrednią, w której dane są już zintegrowane),
- wykorzystanie dedykowanych, wydajnych mechanizmów docelowej bazy danych (ładowanie binarne gotowych rekordów we własnym formacie silnika hurtowni).

Poza ładowaniem początkowym, wykonywanym jednokrotnie na początku, dokonywane są ładowania przyrostowe w celu odświeżania hurtowni. Do ekstrakcji, transformacji i ładowania służy narzędzie ETL (Extract – Transformation – Load), które może być częścią hurtowni bądź też narzędziem zewnętrznym (rys. 3). Bardzo często proces ETL nazywany jest także procesem integracji danych, natomiast narzędzie ETL znane jest jako platforma integracyjna.

Dane zgromadzone w hurtowni powinny być wysokiej jakości. W praktyce hurtownie są bazami danych integrującymi informacje ze wszystkich pozostałych systemów bazodanowych w przedsiębiorstwie i na każdym etapie należy zapewnić jakość danych.



Rys. 3. Schemat procesu ETL

Źródło: opracowanie własne na podstawie Vassiliadis i in., 2002 (12)

Pojęcie danych i jakość danych

Dane definiuje się jako ciągi symboli uporządkowane zgodnie z przyjętymi regułami i zapisane na nośniku danych (kartka papieru, mowa, dźwięk, dysk twardy komputera) (8). Według innej definicji dane to fakty, zdarzenia, transakcje, które zostały zapisane. Stanowią surowy materiał wejściowy, z którego produkowane są informacje (7). Dane są także definiowane jako postać informacji, którą możemy przetwarzać z użyciem sprzętu komputerowego (6). Natomiast fizyczny zapis wartości danych jest nazywany bazą danych. W takim ujęciu baza jest to zorganizowany zestaw danych, w którym można je przetwarzać, aktualizować, dodawać nowe bądź wyszukiwać według odpowiednich kryteriów (5).

Istnieje podstawowy zestaw wymagań, jakie powinien spełniać zestaw danych:

- konieczna jest odpowiednia liczba danych o poprawnej strukturze i właściwie reprezentujących modelowane zjawisko,
- niezbędny jest dostęp do danych we właściwym czasie i miejscu,
- musi istnieć możliwość weryfikacji i aktualizacji danych oraz ich ochrony.

Dane stanowią więc zasób, który należy pozyskać, przetwarzać i udostępniać gwarantując jakość zarówno danych, jak i procesu zarządzania danymi (8).

Gromadzone dane potrzebne są zarówno do codziennej pracy, jak i podejmowania decyzji. Istotne jest zatem, aby na każdym kroku zapewniać najwyższą jakość przetwarzanych danych. Według Stecyka (10) „jakość danych należy rozumieć jako kwalifikację poprawności danych, ale także ich przydatności”. Pojęcie to jest subiektywne

i zależy od punktu widzenia zainteresowanego. Inne znaczenie pojęcia „jakość danych” ma dla programisty, inne dla administratora, a jeszcze inne dla odbiorcy (decydenta).

Projektant systemu i programista są zainteresowani mierzaniem schematu hurtowni danych. Etap tworzenia systemu ma największy wpływ na jakość końcową danych systemu. Administrator zaś jest zainteresowany raportowaniem błędów, dostępem do metadanych oraz ich aktualnością. Na jakość w świetle projektowania i administracji składają się dwa główne wymiary: jakość schematu i uwzględnienie zmian metadanych (4).

Jakość schematu określa zdolność modelu do przechowywania informacji pozwalającej na dokładne opisanie istniejącej sytuacji rzeczywistej. Składa się ona z szeregu wymiarów odzwierciedlających jej pełną funkcjonalność (4, 11):

- poprawność (validation) – stanowi właściwe zrozumienie obiektów świata rzeczywistego, właściwe zrozumienie wymagań użytkowników, poprawne schematy źródeł oraz uzasadnione obliczenia i agregacje w systemie;
- kompletność (completeness) – stopień kompleksowego ujęcia niezbędnej wiedzy w strukturach hurtowni danych oraz systemach źródłowych, uzupełnienie brakujących rekordów i pól;
- spójność (consistency) – oznacza ujednoczenie i integrację przechowywanych w systemie informacji – zgodność czasową;
- redukcja redundancji (reduction of recurrence) – stopień usunięcia pokrywających się informacji gromadzonych w systemie;
- weryfikowalność (traceability) – oznacza, iż wszelkie wymagania użytkowników końcowych, projektantów i administratorów powinny być możliwe do wyśledzenia w schemacie hurtowni danych;
- interpretowalność (interpretability) – stanowi właściwe opisanie składników struktury, co pozytywnie wpływa na łatwość administrowania i użytkowania systemu.

Drugim elementem wpływającym na jakość projektowania i administrowania hurtownią danych jest **uwzględnienie zmian metadanych**. Zarówno źródła danych, jak i wymagania w stosunku do systemu ulegają ewolucji w czasie. Konieczne jest zatem odzwierciedlenie tych zmian w strukturze metadanych systemu (4).

Projektant systemu, a w szczególności hurtowni danych, musi pamiętać o zapewnieniu następujących aspektów jakości danych:

- jakość definicji danych,
- jakość architektury danych,
- jakość zawartości danych,
- jakość prezentacji danych.

Dla użytkownika jakość danych oznacza możliwość efektywnego pozyskania niezbędnych informacji. Użytkownik końcowy wymaga od systemu, aby informacje, jakie uzyskał, były jak najlepszej jakości, tj. użyteczne i zgodne czasowo, a wyniki analiz generowane szybko, w postaci czytelnej i zrozumiałej.

Jednym z głównych wymiarów jakości danych jest **dostępność** (accessibility). Oznacza ona możliwość dostępu za pomocą kierowanych do bazy zapytań. Następujące kryteria pozwolą dokładniej opisać ten składnik jakości (4):

- bezpieczeństwo (security) – ograniczenie dostępu do informacji (autoryzacja) z uwagi na rolę i wymagania użytkowników;
- dyspozycyjność transakcyjna (transactional availability) – czas, w jakim zgromadzone dane w systemie są dostępne dla użytkownika;
- dyspozycyjność systemu (system availability) – czas, w jakim system hurtowni danych jest dostępny dla użytkownika.

Drugim wymiarem jakości danych jest **użyteczność**, która jest charakteryzowana przez następujące kryteria:

- interaktywność (interactivity) – możliwość nawiązania łatwej komunikacji między systemem a użytkownikiem, uzyskiwanie dodatkowych informacji;
- interpretowalność (interpretability) – stopień zrozumienia informacji będących wynikami oraz możliwość ich analizy pod kątem odzwierciedlenia rzeczywistości;
- wiarygodność (credibility) – przekonanie o prawidłowości otrzymanych wyników;
- dokładność (accuracy) – zgodność wartości przechowywanych z wartościami rzeczywistymi.

Kolejnym wymiarem jakości w świetle użytkownika danych jest **zgodność czasowa** (timelines). Wymiar czasu jest podstawowym elementem w hurtowni danych. Dzięki niemu możliwe jest uzyskanie historii podmiotu oraz analiz z uwzględnieniem różnych przedziałów czasowych. Aby otrzymywane dane były wiarygodne, niezbędne jest restrykcyjne przestrzeganie aktualności gromadzonych w systemie danych. Zapewnienie przez system, w sposób spójny i konsekwentny, jakości wszystkich wymiarów wpływa na końcową jakość danych.

Podsumowanie

Niska jakość danych – ich niekompletność, niepoprawność, może być przyczyną niechęci użytkowników do korzystania z nich jako źródła informacji. Istotnym problemem jest także podejmowanie decyzji na podstawie dużej ilości danych, często niespójnych i rozproszonych (pochodzących z wielu systemów operacyjnych). Rozwiązaniem jest wprowadzenie systemu wspomagającego podejmowanie decyzji. Dobrze zaprojektowana hurtownia danych to solidna podstawa dla systemów analitycznych i źródło informacji do podejmowania decyzji w przedsiębiorstwie.

Należy jednak podkreślić, iż jakość danych powinna być postrzegana całościowo, jako zespół wszystkich wymiarów jakości informacji. Natomiast mechanizmy kontroli jakości winny obejmować kompleksowe działania zarówno w sferze projektowania hurtowni danych, pierwszego zasilania, jak i każdorazowej aktualizacji. Należy

również wskazać na konieczność kontroli jakości danych gromadzonych w bazach transakcyjnych i innych zewnętrznych źródłach danych wykorzystywanych w SWD.

Literatura

1. Codd E.F., Codd S.B., Salley C.T.: Providing OLAP to user-analysts: An IT mandate. Technical report, E.F. Codd & Associates, 1993.
2. Holsapple C.W.: Decision Support Systems. W: Encyclopedia of Information Systems, Elsevier Inc., 2004.
3. Inmon W.H.: Building the Data Warehouse. 4st Ed. Wiley and Sons, New York, 2005.
4. Jarke M., Lenzerini M., Vassiliou Y., Vassiliadis P.: Hurtownie danych. Podstawy organizacji i funkcjonowania. WSiP, Warszawa 2003, 272.
5. Kisielnicki J., Gwiazda T.: Wstęp do informatyki w zarządzaniu. Wyd. Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego, 2007, 171.
6. Kisielnicki J.: Systemy informatyczne zarządzania. Placet, Warszawa 2008.
7. Lucey T.: Management Information Systems. 6th edition, DP Publications Ltd., London, 1991, 376.
8. Ostrowska T.: Relacyjne systemy bazodanowe. Politechnika Warszawska, 2002, 237.
9. Poe V., Kluwer P., Brobst S.: Tworzenie hurtowni danych. WNT, Warszawa 2000, 286.
10. Stecyk A.: Jakość i integralność informacji w hurtowniach danych. Gazeta IT, 2005, **9(39)**.
11. Wojtachnik R.: Problemy we wdrażaniu hurtowni danych. Gazeta IT, 2005, **9(39)**.
12. Vassiliadis P., Simitis A., Skiaadopoulos S.: Conceptual modeling for ETL processes. In Proceedings of the ACM 5th International Workshop on Data Warehousing and OLAP (DOLAP '02), 2002, 14-21.
13. <http://datawarehouse4u.info>

Adres do korespondencji:

dr Zuzanna Jarosz
Zakład Agrometeorologii i Zastosowań Informatyki
IUNG-PIB
ul. Czartoryskich 8
24-100 Puławy
tel. (81) 886 34 21 w. 210
e-mail: zjarosz@iung.pulawy.pl