

CYKL ŻYCIA HURTOWNI DANYCH W PRZEDSIĘBIORSTWIE PRODUKCYJNYM SPRZĘTU PRZECIWOŻAROWEGO

Marian KOPCZEWSKI, Marek TOBOLSKI

Streszczenie: hurtownia danych to „architektura informatyczna organizująca proces pozyskiwania i porządkowania danych potrzebnych do wspomaganie podejmowania decyzji w przedsiębiorstwie”. W tym szerokim ujęciu, hurtownię danych definiuje się więc jako bazę danych wraz z mechanizmami pozyskiwania danych z różnych źródeł danych oraz procesami ich przetwarzania do postaci odpowiedniej dla analityków i decydentów. Dla potrzeb niniejszej pracy, hurtownia danych traktowana jest *sensu largo*, tj. w szerokim ujęciu, jako system wspomaganie decyzji, który jest oparty na hurtowni danych *sensu stricte*, a jej wykorzystanie zależy od jej cyklu życia w określonej organizacji, którą jest przedsiębiorstwo produkujące sprzęt przeciwpożarowy.

Słowa kluczowe: systemy informacyjne, decyzje, hurtownia danych, cykl życia hurtowni.

1. Elementy cyklu życia hurtowni danych

Zasoby informacyjne każdej organizacji, stanowią kapitał niematerialny, którego użyteczność zależy od umiejętności ich przetworzenia i dostarczenia użytkownikom w formie najlepiej zaspokajającej ich potrzeby. Zasoby informacyjne organizacji funkcjonują w ich systemach informacyjnych w postaci baz danych, hurtowni danych, jak też w bardziej zaawansowanych systemach informacyjnych jako bazy procedur, bazy modeli, bazy wiedzy, czy bazy tekstów. Zasoby danych utrzymywane są w systemach informacyjnych organizacji dzięki odpowiedniemu oprogramowaniu i za jego pośrednictwem są udostępniane użytkownikom.

Pojęcie „hurtownia danych” pojawiło się w języku polskim jako odpowiednik angielskiego terminu *data warehouse* i mimo, że budzi kontrowersje w środowiskach nie związanych blisko z technologiami informacyjnymi, to powszechnie funkcjonuje w środowisku nauk technicznych. Nie mniej jednak, celowym wydaje się jego bliższe w tym miejscu scharakteryzowanie. Pojęcie *data warehouse* wprowadzili w 1988 roku Barry Devlin i Paul Murphy z irlandzkiego oddziału firmy IBM, którzy zajmowali się problemem integracji organizacji. Określili tym terminem środowisko do wspierania użytkowników informacji biznesowej, zapewniające informatykom możliwość dbania o jakość danych. Współcześnie jednak termin ten jest interpretowany nieco odmiennie. W wąskim zakresie oznacza bazę danych, która zawiera dane dobrane i zorganizowane pod kątem zapewnienia możliwości łatwego i szybkiego ich wykorzystania w procesie podejmowania decyzji. W szerszym zaś ujęciu, hurtownię danych traktuje się jako cały system, który pobiera dane z baz źródłowych, umieszcza je w nowej bazie, a następnie udostępnia je użytkownikom podejmującym decyzje biznesowe, utożsamiając tym samym definitywnie hurtownię danych z systemem BI i potwierdzając kluczową rolę jej I warstwy – informacji i składowania.

Hurtownia danych nie może jednak istnieć w oderwaniu od systemu zasilania jej danymi. I tak, powołując się na A. Wiecką, dyrektora zarządzającego polskiego oddziału SAS Institute, hurtownia danych to „architektura informatyczna organizująca proces

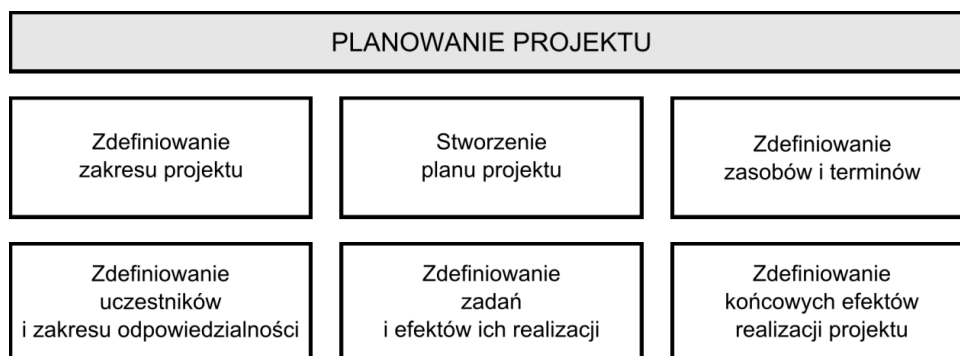
pozyskiwania i porządkowania danych potrzebnych do wspomaganie podejmowania decyzji w przedsiębiorstwie”. W tym szerokim ujęciu, hurtownię danych definiuje się więc jako bazę danych wraz z mechanizmami pozyskiwania danych z różnych źródeł danych oraz procesami ich przetwarzania do postaci odpowiedniej dla analityków i decydentów. Dla potrzeb niniejszej pracy, hurtownia danych traktowana jest *sensu largo*, tj. w szerokim ujęciu, jako system wspomaganie decyzji, który jest oparty na hurtowni danych *sensu stricte*, a jej wykorzystanie zależy od cyklu życia. Syntezując wyniki przeprowadzonej analizy literatury, etapy cyklu życia hurtowni danych zebrać można jako, następujące po sobie:

1. Planowanie;
2. Zbieranie wymagań na dane;
3. Tworzenie fizycznego projektu i rozwój bazy danych;
4. Pozyskiwanie i integrację danych;
5. Zapewnianie hurtowni danych;
6. Automatyzowanie procesu zarządzania danymi;
7. Tworzenie początkowego zbioru predefiniowanych raportów;
8. Kontrolowanie poprawności i testowanie danych;
9. Szkolenie użytkowników;
10. Uruchomienie hurtowni danych.

1. 1. Planowanie

Zgodnie z powszechną definicją planowania, obejmuje ono „*definiowanie celów organizacji, określenie ogólnej strategii ich osiągnięcia oraz opracowanie zwartej hierarchii planów służących do integracji i koordynacji działań*”. Z powyższej, ogólnej definicji wynika więc, że planowanie dotyczy zarówno celów (co ma być zrobione), jak i środków (jak to ma być zrobione). Na tym tle, pierwszy etap planowania hurtowni danych nie różni się od analogicznych etapów innych projektów z zakresu zarządzania organizacją, z tą jedynie uwagą, że silniej wiąże się z zadaniami, które wynikają z kolejnych etapów cyklu życia wspomaganie podejmowania decyzji. W związku z tym, przed stworzeniem planu projektu hurtowni danych istotne jest dokonanie analizy poszczególnych, dalszych etapów cyklu życia hurtowni danych i wzięcie ich pod uwagę już na etapie planowania. Dla przykładu, przed rozpoczęciem przygotowania projektu hurtowni danych należy bezwarunkowo określić architekturę danych i infrastruktury techniczne, które *de facto* będą realizowane w etapie trzecim, t.j. tworzenia fizycznego projektu i rozwoju bazy danych. Do etapu planowania hurtowni danych zalicza się więc działania zebrane na rysunku 1.

Dodatkowo, na etapie planowania uwzględnić należy ewentualne problemy techniczne, które mogą pojawić się w dalszych etapach. Mowa tu chociażby o efektach realizacji zadań, które wymagać będą środków technicznych i czasu. Zaliczyć do nich można m.in. zagadnienia planowania pojemności, strategię archiwizowania danych, procedury dostępu użytkowników różnicowanych pod względem uprawnień dostępu do danych. Ponadto, jeżeli w dziedzinie infrastruktury technicznej hurtownia danych projektowana jest niejako „od podstaw” i nie istnieją żadne „zastane” infrastruktury techniczne, to ich implementacja musi zostać uwzględniona w planie projektu, a obejmie m.in. wybór technologii sieci LAN/WAN, wybór platformy serwerów i stacji roboczych oraz ich konfigurację, czy wybór narzędzi dostępu do danych.



Rys. 1. Działania na etapie planowania projektu w cyklu życia hurtowni danych

1. 2. Zbieranie wymagań na dane

Etap ten dotyczy zrozumienia potrzeb informacyjnych przyszłych użytkowników hurtowni oraz wstępnego modelowania tych wymagań. Zbieranie wymagań polega na wyjaśnieniu, w jaki sposób użytkownik prowadzi działania, co nim kieruje, jakich atrybutów (miar) potrzebuje.

Podczas rozmów i wywiadów, jakie na tym etapie prowadzi się z użytkownikami należy określić:

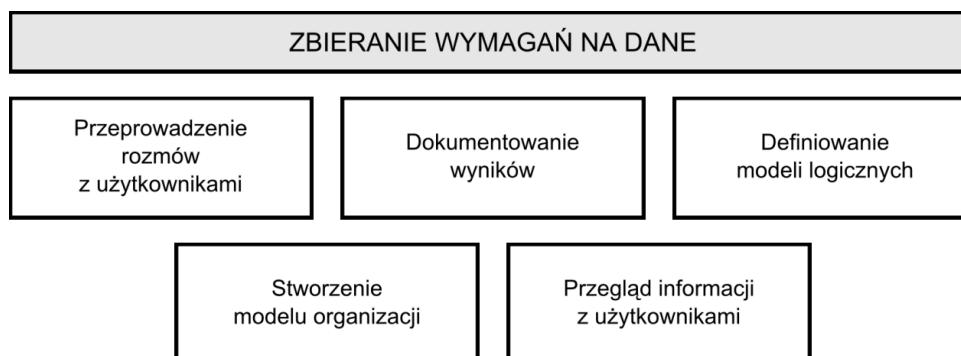
- Jakie dane są obecnie przez nich wykorzystywane ?
- Jakie dane chcieliby posiadać ?
- Jakie dane są użytkownikom absolutnie niezbędne, a jakie mają charakter wyłącznie życzeniowy ?
- Jakiego poziomu szczegółowości i jakich podsumowań potrzebują ?
- W jakiej postaci chcą otrzymywać wyniki odpowiedzi na swoje zapytania ?
- Z jakiego rodzaju narzędzi dostępu do danych chcieliby korzystać ?

Do realizacji tego etapu, przyjmuje się zazwyczaj metodę iteracyjną, która zakłada, że użytkownicy mogą zapoznać się z wynikami pierwszej iteracji i nanosić swoje uwagi i spostrzeżenia w iteracjach kolejnych.

Wynikiem takiego działania jest stworzenie logicznego modelu danych oraz wielowymiarowego modelu organizacji, który odzwierciedla przepływ danych w organizacji i wynikające z niego zależności informacyjne.

Ostatnim przedsięwzięciem, a zarazem zwieńczeniem tego etapu jest przegląd uzyskanych informacji wraz z użytkownikami.

Podsumowując, do etapu zbierania wymagań na dane zalicza się działania zebrane na rysunku 2.



Rys. 2. Działania na etapie zbierania wymagań na dane w cyklu życia hurtowni danych

1. 3. Tworzenie fizycznego projektu i rozwój bazy danych

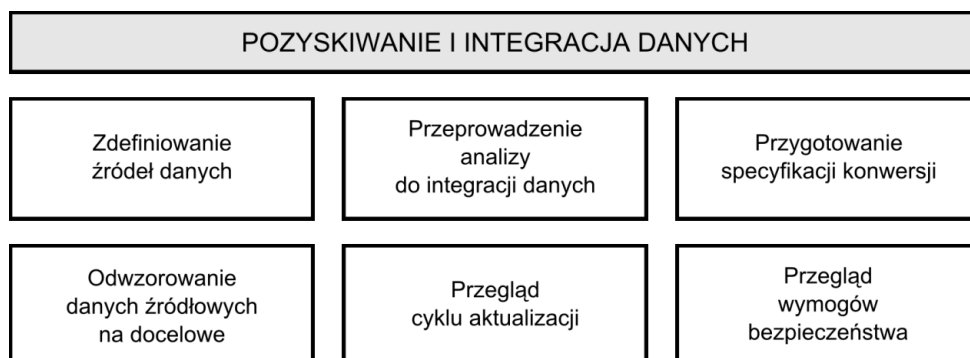
Etap trzeci cyklu życia hurtowni danych obejmuje projekt i denormalizację bazy danych. Zawiera zadania kluczowe dla przetwarzania danych wspomagającego podejmowanie decyzji oraz rozwoju tego wspomaganie. Jako, że etap ten jest silnie związany z fizycznym projektem bazy danych, wymaga rozumienia pojęć z zakresu systemów baz danych m.in. ogólnych pojęć hierarchii, wymiarów i faktów, ogólnych zasad projektowania baz danych o schemacie gwiazdowym oraz ogólnych pojęć z zakresu wspomaganie decyzji.

Działania realizowane na tym etapie powinny koncentrować się na: zaprojektowaniu bazy danych łącznie z tabelami faktów, wymiarów i tabelami zależności (relacji), stworzeniu właściwych obiektów (encji) w bazie danych, denormalizacji danych, stworzeniu strategii indeksowania danych, przygotowaniu strategii agregacji danych, uściśleniu pojemności baz danych.

1. 4. Pozyskiwanie i integracja danych

Działania przewidziane na tym etapie mają na celu bieżące weryfikowanie opracowanego na etapie poprzednim fizycznego projektu bazy danych hurtowni. Dlatego też, etap pozyskiwania i integracji danych jest wykonywany razem z etapem tworzenia fizycznego projektu bazy danych, gdyż projekt ten jest niezbędny do odwzorowywania danych źródłowych na dane docelowe, co stanowi sedno niniejszego etapu.

Działania realizowane w ramach niniejszego etapu, zebrane na rysunku 3, obejmują w szczególności: zdefiniowanie możliwych systemów źródłowych i ustalenie wykorzystywanych formatów plików, wykonanie analizy danych w celu określenia najlepszego źródła danych, wykonanie analizy danych pod kątem możliwości ich integracji, przygotowanie specyfikacji konwersji (transformacji) danych dla każdego pola źródłowego i docelowego, opisanie (dokumentowanie) procesu odwzorowania danych źródłowych na dane docelowe.

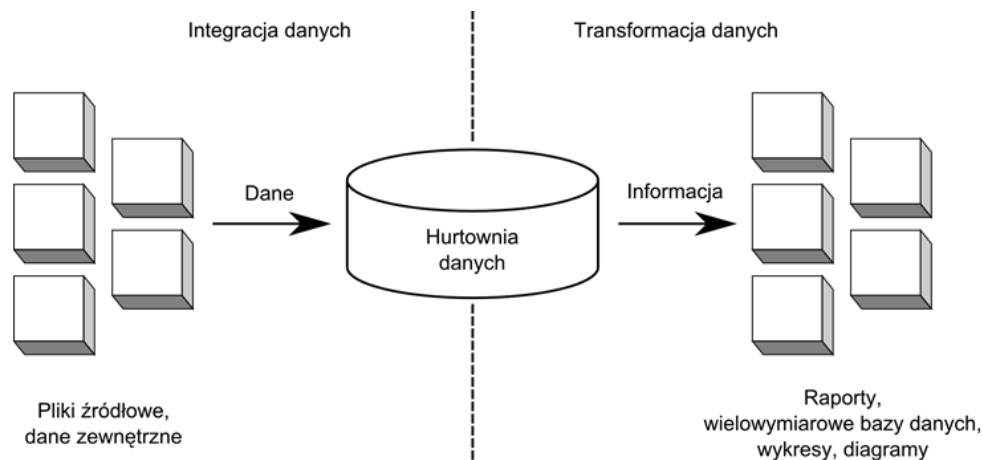


Rys. 3. Działania na etapie pozyskiwania i integracji danych w cyklu życia hurtowni danych

W toku realizacji powyższych działań, okazać może się, że dane które są faktycznie możliwe do pozyskania, różnią się od zapotrzebowań na dane zgłoszonych przez użytkownika, co z jednej strony nadaje tym zapotrzebowaniom charakter życzeniowy, z drugiej zaś zmusza do poszukiwania takich źródeł danych, które pozwolą na zaspokojenie zgłoszonych potrzeb informacyjnych. Może to spowodować konieczność modyfikacji wcześniejszych wymagań i fizycznego projektu bazy danych hurtowni, co po raz kolejny podkreśla nierozzerwalny związek niniejszego etapu z etapem poprzednim.

Użytkownicy hurtowni danych zazwyczaj mają świadomość, jakich informacji potrzebują. Nie zawsze jednak wiedzą, czy dostępne dane dostarczą im faktycznie informacji, której zażądali. W tym miejscu pojawia się więc konieczność uświadomienia potencjalnym użytkownikom hurtowni danych różnicy między danymi a informacją. W kontekście hurtowni danych terminu „dane” używa się dla oznaczenia dyskretnego elementu, jakim jest dla przykładu plik. Takie dane są zbierane i organizowane w określone struktury danych zaprojektowane tak, aby mogły być przetwarzane przez systemy komputerowe.

W tym miejscu po raz kolejny pojawia się różnica między bazodanowymi systemami transakcyjnymi OLTP, a analitycznymi OLAP. Bazodanowe systemy transakcyjne (OLTP) przetwarzają dane, wykorzystywane w codziennej działalności organizacji. Gdy zaś dane te zostaną scalone, zgrupowane, zsyntezowane z innych danych, zestrukturalizowane, uporządkowane i odpowiednio zaprezentowane, stają się użyteczną „informacją”. Dzięki takiemu właśnie podejściu tworzone są bazodanowe systemy analityczne OLAP, a wśród nich hurtownie danych, tj. systemy komputerowe, w których dane są zbierane, integrowane i przechowywane w celu takiego ich przetwarzania, by mogły przekształcić się w użyteczne informacje. Punktami decydującymi o pomyślnej realizacji niniejszego etapu są procesy integracji i konwersji (transformacji) danych. Ideę tych procesów, przedstawia schematycznie rysunku 4.



Rys. 4. Integracja i transformacja danych w cyklu życia hurtowni danych

Proces integracji danych uważany jest za wstępne przetwarzanie danych w celu: ustalenia standardów nazw i wartości danych źródłowych i docelowych; rozwiązania problemów niezgodności w reprezentacji danych źródłowych (np. występowanie synonimów, homonimów czy analogii); ustalenia różnych ścieżek dostępu do danych źródłowych poprzez identyfikację kluczy głównych, alternatywnych i niejednoznacznych.

Uogólniając, etap pozyskiwania i integracji danych składa się z następujących podetapów:

1. Pozyskiwanie danych źródłowych – polega na określeniu znaczenia danych źródłowych i zidentyfikowaniu użytecznych pól w plikach źródłowych, spełniających wymagania systemu docelowego. Wynikiem tego podetapu są wyselekcjonowane pliki i pola źródłowe, spełniające wymagania systemu docelowego, tj. hurtowni danych.
2. Konsolidacja danych – polega na scaleniu źródeł danych w jedną, zintegrowaną, docelową strukturę bazodanową. Wynikiem tego podetapu jest zintegrowany model danych lub zweryfikowany projekt fizycznej bazy danych.
3. Konwersja danych – polega na określeniu sposobów przekształcania danych źródłowych do zintegrowanej struktury danych docelowych. Wynikiem tego podetapu jest specyfikacja konwersji, tj. algorytmiczny opis, na podstawie którego programiści przygotują programy ładujące dane źródłowe do docelowej bazy danych hurtowni.
4. Przenoszenie danych – polega na zdefiniowaniu kroków, które muszą zostać wykonane przed faktycznym umieszczeniem danych źródłowych w docelowej strukturze danych. Wynikiem tego podetapu są programy konwersji używane do umieszczenia danych źródłowych w docelowej bazie danych hurtowni, realizujące również funkcje testowania poprawności danych, obsługi zdarzeń wyjątkowych czy zbierania statystyk.

1. 5. Zapelnianie hurtowni danych

Realizując ten etap, opracowuje się programy, które pobierają dane z systemów źródłowych zdefiniowanych w etapie poprzednim, konwertują je według opracowanych specyfikacji oraz zapisują w bazie danych hurtowni. Zasadniczym celem tego etapu jest opracowanie programów lub wykorzystanie istniejących narzędzi do pobierania, konwersji i ładowania danych (narzędzi ETL).

W sferze organizacyjnej wymagane jest określenie strategii zapelniania hurtowni, tj. wskazanie częstości i kolejności pobierania danych z systemów źródłowych, zaś w sferze infrastrukturalnej określenie zakresu wiedzy i umiejętności zarówno użytkowników, jak i administratorów systemu na temat m.in. administrowania bazą danych, wykorzystania narzędzi programistycznych, wykorzystania systemów i platform czy zapewnienia jakości. Powyższe zagadnienia sfery organizacyjnej będą rzutowały na treści, jakie będą musiały być uwzględnione w jednym z kolejnych etapów cyklu życia hurtowni danych, tj. etapie szkolenia użytkowników.

1. 6. Automatyzowanie procesu zarządzania danymi

Etap niniejszy dotyczy opracowania procedur (programów, skryptów), które zautomatyzują procesy ETL, tj. pobierania, konwersji i ładowania danych, realizując je w ustalonych wcześniej odstępach czasu. Ponadto na etapie tym tworzy się procedury automatycznego wykonywania kopii zapasowych i ewentualnego odzyskiwania danych. Wszystkie, przygotowane na tym etapie, automatyczne procedury muszą zostać poddane szczegółowym testom, dla wykrycia potencjalnych błędów w ich działaniu. Działania realizowane na tym etapie zostały przedstawione schematycznie na rysunku 5.



Rys. 5. Działania na etapie automatyzowania procesu zarządzania danymi w cyklu życia hurtowni danych

1. 7. Tworzenie początkowego zbioru predefiniowanych raportów

Wynik realizacji niniejszego etapu stanowi najbardziej dla użytkowników widoczny efekt rozwoju systemu hurtowni danych. Dokonuje się w głównej mierze dzięki przygotowaniu kilku początkowych, predefiniowanych raportów.

Predefiniowane raporty stanowią bezpośrednią odpowiedź na potrzeby informacyjne użytkowników, ujawnione na etapie zbierania wymagań na dane. Stosownie do potrzeb

użytkowników, raporty standardowe tworzone mogą być w z góry przyjętych odstępach czasu, bądź jako odpowiedź na ich zapytania *ad hoc*. W obu jednak przypadkach wymagają udokumentowania tego procesu.

Jak już powiedziano, podstawą przygotowania zbioru początkowych raportów są odpowiedzi użytkowników na pytania, zadane na etapie zbierania wymagań na dane, a wśród nich:

- W jakiej postaci chcą otrzymywać wyniki odpowiedzi na swoje zapytania ?
- Jakiego poziomu szczegółowości i jakich podsumowań potrzebują ?

Biorąc pod uwagę zarówno postać, jak i poziom szczegółowości raportów stwierdzić można, że użytkownik ma w tym względzie pełne spektrum możliwości. Najprostszą z nich (niestety jedynie z punktu widzenia administratora hurtowni) jest własnoręczne napisanie instrukcji w języku zapytań SQL. Możliwość taka, choć istnieje, wymaga od użytkownika szczególnej znajomości nie tylko zawartości hurtowni danych, listy jej tabel i kolumn, ale i samego języka zapytań SQL. Jest więc, z punktu widzenia użytkownika, który nie jest, i nie musi być, specjalistą z zakresu informatyki, niezwykle wymagająca.

Rozwiązaniem jest zastosowanie narzędzi dostępu do danych, tj. programów, które pozwolą użytkownikowi sformułować żądanie w sposób intuicyjny, a następnie „przetłumaczyć” je na zapytanie w języku SQL kierowane do systemu zarządzania bazą danych hurtowni.

Dane, otrzymane w wyniku realizacji takich zapytań można przedstawić na wiele sposobów, których celem jest ułatwienie użytkownikowi poznania i zrozumienia tych danych, a w efekcie – przekształcenia ich w informacje pomocne w podejmowaniu decyzji.

Do podstawowych form prezentacji danych należą:

- raporty – prezentujące dane w dwóch podstawowych formatach tekstowych (kolumnowym i tabelarycznym skośnym),
- wykresy – prezentujące dane w formie graficznej, m.in. wykresów liniowych, słupkowych, stosowych, kołowych czy map obszaru,
- odwzorowania – umożliwiające przejście z prezentacją raportów i wykresów do wymiaru geograficznego, często angażując w ten proces narzędzia GIS – owe.

W zakresie zaś szczegółowości danych, użytkownik może żądać przedstawienia:

- wielkości prostych – przechowywanych bezpośrednio w kolumnach bazy danych hurtowni,
- wielkości wyliczanych – obliczanych z wielkości prostych za pomocą standardowych operacji matematycznych – np. wartość średnia, minimalna, maksymalna, mediana itp.,
- wielkości złożonych – obliczanych na podstawie danych z wielu poziomów hierarchii danych,
- wielkości przekształcalnych – wykorzystywane w obliczeniach, w których informacje zgodne z jednym, konkretnym zbiorem kryteriów porównuje się z informacjami zgodnymi z pokrewnym zbiorem kryteriów,
- agregacji: agregacji prostych – akumulacji informacji według hierarchii i atrybutów wewnątrz modelu danych; w hurtowniach danych agregacje proste są zwykle przygotowywane wcześniej i przechowywane w bazie danych hurtowni; agregacji *ad hoc* – akumulacji informacji wzdłuż niestandardowych ścieżek; obliczeń na wierszach – obliczeń na wartościach zebranych w wierszach tabel; danych nie podlegających agregacji – tj. danych, które z punktu widzenia logiki nie nadają się do agregacji.

1. 8. Kontrolowanie poprawności i testowanie danych

Realizując niniejszy etap należy przeprowadzić wieloaspektowe kontrole poprawności i testowanie danych. Testy realizowane na tym etapie stanowią niejako uzupełnienie standardowych procesów kontroli danych, realizowanych w fazie wydzielania, integrowania i wprowadzania danych do hurtowni. Zasadniczą ich częścią jest uruchomienie programów wydzielania, konwersji i wprowadzania w środowisku testowym i sprawdzenie, czy dane zostały umieszczone w hurtowni zgodnie z przyjętymi zasadami. Testy mają dodatkowo na celu wykrycie błędnych lub brakujących danych przy zastosowaniu technik takich jak chociażby liczenie rekordów czy analiza dziennika zdarzeń wyjątkowych. W ramach testowania danych analizuje się także zestaw początkowych predefiniowanych raportów.

1. 9. Szkolenie użytkowników

Najlepiej nawet zaplanowana, zaprojektowana, zintegrowana, zapełniona, automatycznie tworząca szereg predefiniowanych raportów i gotowa do odpowiedzi na nietypowe zapytania hurtownia danych nie spełni swojej roli, jeśli ani o jej istnieniu, ani o możliwościach i sposobach jej wykorzystania nie będą wiedzieli jej użytkownicy.

Dla uniknięcia powyższego, etap szkolenia koncentruje się na stworzeniu programów szkoleniowych dla użytkowników. Realizując niniejszy etap, należy mieć na uwadze, że przeszkoleni muszą zostać użytkownicy wszystkich szczebli.

W szczególności, użytkownicy hurtowni winni odbyć szkolenie w zakresie:

- przestrzeni danych hurtowni, tj. jej zawartości i stosowanej terminologii,
- posługiwania się wysuniętymi narzędziami (aplikacjami) dostępu do danych oraz zasad ich funkcjonowania,
- sposobów korzystania z metadanych w celu uzyskania informacji o danych zapisanych w hurtowni.

Administratorzy hurtowni danych powinni zaś zostać przeszkoleni w zakresie:

- zawartości hurtowni,
- sposobów zarządzania metadanymi,
- zasad i procedur zapełniania hurtowni,
- możliwości monitorowania eksploatacji hurtowni (np. śledzenia statystyk ładowania danych i korzystania z nich przez użytkowników),
- systemu zarządzania użytkownikami i ich uprawnieniami do dostępu do hurtowni,
- procedur sporządzania kopii zapasowych i odzyskiwania danych,
- kontroli poprawności działania wszelkich mechanizmów realizujących operacje na hurtowni danych w sposób zautomatyzowany.

Przygotowując programy szkolenia użytkowników, należy wziąć pod uwagę fakt, że najwięcej pytań pojawi się nie podczas samego szkolenia, ale już w czasie faktycznego użytkowania hurtowni. W związku z tym, konieczne jest stworzenie i rozwój systemu pomocy technicznej dla użytkowników.

Prócz powyższych przedsięwzięć, etap szkolenia musi „zderzyć się” z problemem natury socjologicznej. Użytkownicy, którzy dotychczas swoje zadania realizowali w oparciu o istniejący system wspomagania podejmowania decyzji, niechętnie będą wykorzystywać nowe narzędzie. W związku z tym wymaga się, by na etapie szkoleń użytkowników poświęcić czas na promocje hurtowni wewnątrz organizacji, ukierunkowując ją na dwie grupy użytkowników: zaawansowanych technologicznie oraz

najbardziej wpływowych. Pierwsza grupa będzie zainteresowana technologią i szybko nabyte umiejętności wymagane do efektywnego posługiwania się narzędziami dostępu do danych, zachęcając innych do pójścia w ich ślady. Druga grupa, składająca się z liderów w swoich zespołach będzie zachęcała swoich podwładnych do korzystania z hurtowni poprzez osobisty przykład.

1. 10. Uruchomienie hurtowni danych

Ostatni etap cyklu życia hurtowni danych obejmuje:

- zainstalowanie fizycznych infrastruktur dla wszystkich użytkowników (ustanowienie połączeń sieciowych między bazami danych, skonfigurowanie stacji roboczych, zainstalowanie wysuniętych narzędzi dostępu do danych),
- zainstalowanie aplikacji wspomagających podejmowanie decyzji,
- stworzenie struktur pomocy dla użytkowników,
- stworzenie procedur dodawania nowych raportów i rozszerzenia możliwości aplikacji wspomagających podejmowanie decyzji,
- przygotowanie procedur wykonywania kopii zapasowych aplikacji, a nie tylko kopii danych zebranych w hurtowni,
- stworzenie procedur wykrywania i rozwiązywania problemów związanych z integralnością danych,
- stworzenie procedur zarządzania metadanymi,
- stworzenie procedur zarządzania zmianami.

Przedstawione etapy cyklu życia hurtowni danych dają obraz zakresu zadań, jakie należy wykonać w czasie jej tworzenia. Każdy z powyższych etapów zwieńczony musi być produktem końcowym w postaci dokumentu opisującego np. projekt bazy danych, przebieg i efekty szkolenia użytkowników czy konfiguracje stacji roboczej. Prócz dokumentacji, za produkt końcowy uznać można faktyczny efekt, osiągnięty po zakończeniu danego etapu, np. przeszkolony użytkownik, wypełniona danymi struktura bazy danych hurtowni czy działający program transformacji danych.

Wykaz kluczowych produktów opisanych 10 etapów cyklu życia hurtowni danych zebrano w tabeli poniżej.

Tab. 1. Kluczowe produkty poszczególnych etapów w cyklu życia hurtowni danych

Etap cyklu życia hurtowni danych	Kluczowe produkty etapu
Planowanie	- Harmonogram projektu (zadania, terminy, osoby odpowiedzialne) - Słownik terminów - Efekty etapów i efekt końcowy - Strategie zarządzania hurtownią
Zbieranie wymagań na dane	- Specyfikacja wymagań - Wielowymiarowy model logiczny hurtowni
Tworzenie fizycznego projektu i rozwój bazy danych	- Projekt struktury i fizyczna struktura bazy danych - Strategie gromadzenia danych
Pozyskiwanie i integracja danych	- Opis źródeł danych - Specyfikacja konwersji dla każdego pola bazy źródłowej

	- Specyfikacja odwzorowania danych źródłowych w dane docelowe
Zapełnianie hurtowni danych	- Programy ETL do wydzielania danych ze źródeł, ich konwersji i zapisywania do hurtowni, - Procedury użycia programów lub narzędzi do zapełniania hurtowni
Automatyzowanie procesu zarządzania danymi	- Procedury (programy) automatyzujące zapełnianie hurtowni - Plany cyklicznego zapełniania hurtowni - Procedury automatycznego wykonywania kopii zapasowych i odzyskiwania danych
Tworzenie początkowego zbioru predefiniowanych raportów	- Szablony, wzorce predefiniowanych raportów, które będą generowane na podstawie danych zapisanych w hurtowni
Kontrolowanie poprawności i testowanie danych	- Wyniki kontroli poprawności danych wprowadzonych do hurtowni
Szkolenie użytkowników	- Programy szkolenia, - Przeszkoleni użytkownicy i administratorzy hurtowni, - Dokumentacja ze szkolenia
Uruchomienie hurtowni danych	- Zainstalowana i skonfigurowana infrastruktura techniczna dla wszystkich użytkowników, - Procedury zarządzania metadanymi, - Procedury zarządzania zmianami w hurtowni.

2. Wnioski

Punktem wyjścia do przeprowadzonych w analiz była potrzeba dokonanie analizy stanu teorii i praktyki wykorzystania hurtowni danych do wspomaganie podejmowania decyzji w organizacjach oraz określenie funkcjonalnych i organizacyjnych możliwości zastosowania hurtowni danych do w zarządzaniu.

Syntezyując wyniki przeprowadzonych analiz, stwierdzić jednoznacznie należy, że umiejscowienie hurtowni danych w klasie systemów Business Intelligence potwierdza, iż hurtownia danych jest zaawansowanym narzędziem wspomaganie podejmowanie decyzji, którego dotychczasowe, główne obszary zastosowania wynikają przede wszystkim z potrzeby: konsolidacji dużych zbiorów danych, wielowymiarowej analizy tych danych, jak najlepszego dostosowania zakresu analiz do żądań użytkownika, oraz jak najlepszego dostosowania formy prezentacji wyników analiz do jego żądań.

Zasadniczym wnioskiem z analizy stanu teorii i praktyki wykorzystania hurtowni danych jest stwierdzenie, że hurtownie danych stały się architektoniczną odpowiedzią na chaos różnorodnych systemów informatycznych i przechowywanych w nich, niezgodnych często ze sobą, danych. Stąd, głównym celem stosowania hurtowni danych jest uzyskanie integracji i integralności danych w skali całej organizacji lub wybranego obszaru tematycznego jej działalności. By taką integralność osiągnąć, przy wdrażaniu hurtowni danych dążyć należy do osiągnięcia celu, jakim jest zdolność do dostarczenia właściwych informacji, właściwym odbiorcom, we właściwym czasie. Rozwiązaniem problemu braku właściwych informacji jawi się w literaturze hurtownia danych, gromadząca dane pochodzące z wszystkich systemów źródłowych dostępnych w organizacji.

Z kolei, rozwiązaniem dostępności informacji we właściwym czasie, a zarazem zasadniczym czynnikiem warunkującym zastosowanie hurtowni danych w rozpatrywanym obszarze, jest przyjęcie określonej architektury hurtowni, tj. określonych rozwiązań technologicznych dotyczących systemów zarządzania bazami danych hurtowni oraz wynikających z nich typów struktur i schematów tych baz danych. W tym zakresie, do obsługi baz danych hurtowni można wybrać system zarządzania relacyjną bazą danych (RDBMS) lub system zarządzania bazą wielowymiarową (MDBMS). Przyjęcie określonego rozwiązania determinuje bazę, w której będą przechowywane agregacje danych, stanowiące bezpośrednią podstawę wspomaganie podejmowania decyzji. Chcąc więc zastosować hurtownię danych do wspomaganie procesu zarządzania w organizacji, należy m.in. określić, jaką architekturę i jaką infrastrukturę techniczną będzie miała taka hurtownia (tj. zdecydować np., jakie zastosować rozwiązania w zakresie systemu zarządzania bazą danych hurtowni) oraz podjąć szereg decyzji, wynikających z kolejnych kroków jej wdrażania. Efektem wdrożenia hurtowni danych o środowisku organizacji ma być system, który powinien być w stanie odpowiedzieć natychmiastowo na praktycznie każde pytanie związane z obszarem zarządzania organizacją.

Litertura

1. Ficoń K.: Systemy Informatyczne Przedsiębiorstw, BEL Studio, Warszawa 2005.
2. Inmon W. H.: Building the Data Warehouse, 3rd edition. Nowy Jork 2002.
3. Januszewski A.: Funkcjonalność informatycznych systemów zarządzania, Warszawa 2008.
4. Kopczewski M., Pączek B.: Data Warehouse as a Part of Defence System, Olsztyn 2008.
5. Kisielnicki J.: MIS Systemy Informatyczne Zarządzania, Warszawa 2008.
6. Kwiatkowska A. M.: Systemy wspomaganie decyzji, Warszawa 2007.
7. Poe V., Klauer P., Brobst S.: Tworzenie hurtowni danych, Warszawa 2000.
8. Turban E.: Decision Support and Expert Systems, Macmillan Publishing Company, New York 1993.

Prof. dr hab. inż. Marian KOPCZEWSKI
Kierownik Katedry Nauk o Bezpieczeństwie
Wyższa Szkoła Bezpieczeństwa w Poznaniu
60-782 Poznań, ul. E. Orzeszkowej 1
Tel. 61 8510518

Mgr inż. Marek TOBOLSKI
Delegatura Wojskowej Ochrony Przeciwpożarowej w Gdyni
81-301 Gdynia ul. Jana z Kolna 8A
Tel. 58 6266178