

Jerzy Korczak*, Wojciech Merchelski, Błażej Oleszkiewicz****

Uniwersytet Ekonomiczny we Wrocławiu*, TETA SA Wrocław**

WYBRANE PROBLEMY TECHNOLOGICZNE HURTOWNI DANYCH I OLAP W WYKRYWANIU OPERACJI PRANIA PIENIĘDZY. STUDIUM PRAKTYCZNE

Streszczenie: W artykule przedstawiono technologię hurtowni danych OLAP użytą do budowy systemu AML (Anti-MoneyLaundering) oraz pokazano konkretne jej zastosowania do przeciwdziałania praniu pieniędzy. W prezentacji wykorzystano System Analizy i Rejestracji Transakcji (w skrócie SART) wykorzystujący Analityczny Serwer SQL (w skrócie ASQL). SART stanowi kompletne narzędzie do przeciwdziałania praniu pieniędzy, spełniające europejskie wymagania legislacyjne. Wśród wybranych problemów omówiono: model hurtowni danych, wspomaganie OLAP, integrację warstwy aplikacyjnej i hurtowni danych oraz analiz OLAP, hurtownię danych łańcuchów transakcji ze wspomaganie OLAP.

Słowa kluczowe: ASQL, AML, hurtownie danych, łańcuchy transakcji, OLAP, ROLAP, SART, Księga Główna, Plan Kont.

1. Wstęp

Artykuł stanowi kontynuację prac badawczych związanych z problematyką technologii wykorzystywanych w systemach przeciwdziałania praniu pieniędzy (zob. nasze wcześniejsze publikacje, m.in. [Korczak, Merchelski, Oleszkiewicz 2008], w której opisano również aspekty legislacyjne, metodologiczne oraz technologiczne związane z problematyką AML).

Celem artykułu jest pokazanie nowych rozwiązań technologicznych związanych z tematyką hurtowni danych zawartych w ASQL i wykorzystanych w systemie AML wdrożonym w jednym z banków w Polsce. Omawiane w artykule aspekty wybrane zostały ze względu na specyficzny charakter modelowania danych w hurtowni oraz po to, aby pokazać, w jaki sposób poszczególne trudności zostały pokonane za pomocą ASQL. Po pierwsze, zwrócimy uwagę na problem modelowania hurtowni danych z użyciem tabel hierarchicznych i opartych na nich wymiarów jednorodnych i niejednorodnych. Następnie pokażemy, w jaki sposób utworzona została wielowymiarowa kostka OLAP, korzystająca z wymiarów niejednorodnych hurtowni danych. W kolejnej części artykułu przedstawimy sposób połączenia warstwy aplikacyjnej

i OLAP oraz opiszemy strukturę OLAP łańcuchów transakcji. W podsumowaniu wskażemy dalsze kierunki prac – wykorzystujące technologie ASQL stosowane w SART.

2. Opis struktury hurtowni danych w banku komercyjnym

W celu zilustrowania modelu i sposobu działania hurtowni danych OLAP zawartej w systemie SART pokażemy kilka istotnych rozwiązań technologicznych, korzystając z doświadczeń nabytych podczas wdrażania systemu w jednym z polskich banków. Do studium przypadku wybrane zostały nietypowe i trudne do zrealizowania w klasycznych rozwiązaniach zadania, związane bezpośrednio z modelowaniem struktur hurtowni danych oraz analiz OLAP, mianowicie:

- projektowanie modelu hurtowni danych SART wykorzystującego tabele hierarchiczne i oparte na nich wymiary jednorodne i niejednorodne,
- konstrukcja modelu kostki OLAP hurtowni danych SART,
- połączenie warstwy aplikacyjnej z analityczną na przykładzie procesu scalania transakcji,
- implementacja modelu struktury OLAP łańcuchów transakcji.

Wybór powyższych zagadnień był podyktowany tym, że z jednej strony są to problemy istotne technologicznie dla sprawnego działania hurtowni danych oraz oryginalnie rozwiązane w działającym produkcyjnie systemie SART, a z drugiej – ich realizacja w klasycznych modelach hurtowni jest niezwykle trudna lub wręcz niemożliwa. W przedstawionym studium hurtownia jest zaopatrywana z sześciu źródeł danych, pochodzących z trzech różnych systemów produkcyjnych. Z Księgi Głównej banku pozyskiwane są następujące informacje: dane o klientach banku, Plan Kont Księgi Głównej banku, konta NRB oraz raporty Dziennika Dokumentów (zawierające dekrety operacji księgowych). Z systemu rozliczeń instrumentów płatniczych dokonywanych pomiędzy uczestnikami rynku finansowego ELIXIR® pobierane są komunikaty systemu ELIXIR realizowane przez bank. Z serwisu internetowego Narodowego Banku Polskiego (www.nbp.gov.pl) wczytywane są notowania kursów wymiany walut banku centralnego.

Dane z poszczególnych źródeł są wczytywane do systemu SART za pośrednictwem tabel importu, a następnie integrowane do jednego spójnego repozytorium danych zawierającego wymiary hurtowni (klienci, plan kont, kursy walut), tabelę faktów (dekrety) oraz tabele pomocnicze (numery NRB, komunikaty ELIXIR). Warto tu zwrócić uwagę na fakt, że model ładowania danych realizowany jest zgodnie z konwencją ELT (Extract-Load-Transform), dzięki czemu uzyskano dużą elastyczność systemu pod kątem identyfikacji źródeł danych oraz pełną kontrolę nad procesem ładowania hurtowni danych. Dodatkowo w hurtowni danych wprowadzony jest wymiar czasu w formie hierarchii z następującymi podwymiarami: rok, kwartał, miesiąc, dzień.

Nowym podejściem w modelowaniu hurtowni danych jest zastosowanie tabel hierarchicznych oraz stworzenie za ich pomocą wymiarów jednorodnych i niejednorodnych hurtowni. Wymiar jednorodny jest ekwiwalentem klasycznego wymiaru hurtowni, mającym dodatkową własność ziarnistości we wszystkich swoich podwymiarach (klasyczne wymiary oparte na tabelach SQL posiadają ziarnistość tylko na najniższym poziomie). Wymiary niejednorodne ASQL nie mają swojego odpowiednika w klasycznych rozwiązaniach hurtowni danych i OLAP.

Rysunek 1 obrazuje fragment raportu przedstawiającego wymiar planu kont hurtowni danych systemu SART, będący przykładem wymiaru niejednorodnego hurtowni danych. W wymiarze tym zastosowano tabelę hierarchiczną z następującymi oznaczeniami klasy węzłów: Plan Kont (w skrócie PK) – węzeł nadrzędny hierarchii, Konto Syntetyczne (w skrócie KS) – pusta pozycja w kolumnie „KNT_CODE” raportu, Konto Analityczne (w skrócie KA) – znacznik „A” kolumny „KNT_CODE” raportu.

KNT_NAME	KNT_ID	KNT_KNT	KNT_CODE	KNT_WAL
Plan Kont	0	brak		
Aktywa trwałe	1	0		
Operacje z udziałem środków pieniężnych i operacje międzybankowe	336	1		
Kasa	337	10		
Gotówka/banknoty oraz monety/	338	100		
1 Gotówka w kasie PLN	339	100-1	A	PLN
Gotówka w walucie obcej USD	340	100-1-787	A	USD
Gotówka w walucie obcej - GBP	341	100-1-789	A	GBP
Gotówka w walucie obcej - CHF	342	100-1-797	A	CHF
Gotówka w walucie obcej - SEK	343	100-1-798	A	SEK
Gotówka w walucie obcej EURO	344	100-1-978	A	EUR
Gotówka w złotych w bankomatach	345	100-2	A	PLN
Czeki podróżnicze	346	101		
Czeki podróżnicze bez gwarancji wymiany	347	101-1	A	PLN
Czeki podróżnicze z gwarancją wymiany	348	101-2	A	PLN
Złoto niemonetarne i inne metale szlachetne	349	102		
Znaki wartościowe	352	103		
Czeki podróżnicze własnej emisji	358	104		
Czeki bankierskie	359	105		
Czeki bankierskie obce	360	105-1		
Czeki bankierskie obce w PLN	361	105-11	A	PLN
2 Czeki bankierskie obce w walucie	362	105-12		
Czeki bankierskie obce w USD	363	105-12-787	A	USD
Czeki bankierskie obce w EURO	364	105-12-978	A	EUR

Rys. 1. Niejednorodny wymiar hurtowni danych OLAP

Na rysunku zaznaczono dwa węzły hierarchii ilustrujące niejednorodność wymiaru wraz ze ścieżkami hierarchii klas węzłów dla:

- pozycji z numerem 1: [PK, KS, KS, KS, KA] dla konta KA „Gotówka w kasie PLN” (ID: 339),
- pozycji z numerem 2: [PK, KS, KS, KS, KS, KA] dla konta KA „Czeki bankierskie obce w USD” (ID: 363).

Łatwo zauważyć, że obie struktury hierarchiczne różnią się pod względem struktury: pozycja 1 ma trzy węzły pośrednie kont syntetycznych, w pozycji 2 węzłów pośrednich jest 5.

3. Wspomaganie OLAP hurtowni danych w systemie SART

We wdrażanym rozwiązaniu wielowymiarowa kostka OLAP miała następujące charakterystyki:

- model kostki OLAP: niepełna, wielowymiarowa, inkrementalna kostka OLAP ASQL,
- wymiary kostki: czas, klienci, plan kont,
- miary kostki oparte na pozycjach WINIEN i MA dekretów (w nawiasach podano identyfikatory kolumn raportów OLAP systemu SART występujące na rysunkach):
 - suma pozycji WINIEN (DDO_WN),
 - suma pozycji MA (DDO_MA),
 - liczba pozycji WINIEN (DDO_OCWN),
 - liczba pozycji MA (DDO_OCMA),
 - liczba pozycji WINIEN i MA (DDO_OC).

Od strony funkcjonalnej kostkę OLAP charakteryzuje:

- wielowymiarowość, tzn. przeliczane są tylko zaistniałe fakty z tabeli faktów (niepełna kostka OLAP) wraz ze wszystkimi nadrzędnymi podwymiarami danego wymiaru,
- inkrementalność, co oznacza, że przeliczane są tylko nowe fakty.

Zastosowanie w rozwiązaniu takiego modelu kostki istotnie wpłynęło na zmniejszenie przestrzeni dyskowej zajmowanej przez tę strukturę danych. Wielowymiarowa kostka OLAP przechowuje zagregowane dane dla iloczynów kartezyjskich wszystkich podwymiarów wymiarów OLAP. Na rysunku 2 pokazano raport z ustawionym widokiem Planu Kont; wszystkie wartości miar raportu są odczytywane z kostki OLAP i nie ma tu potrzeby agregacji danych na wyższych poziomach wymiarów, tak jak to jest w klasycznych systemach OLAP. Dodatkowo na rysunku pokazano operację rozszerzenia niepełnej kostki do postaci pełnej – zaznaczony na rysunku wiersz nie ma swojej reprezentacji fizycznej w niepełnej kostce i został w raporcie dodany dynamicznie z wartościami miar równymi zeru.

W systemie produkcyjnym wydajność rozwiązania dla pojedynczego importu danych zawierającego 57 952 operacji (dekretów) jest następująca: przeliczenie tabeli OLAP (jeden proces) trwało 250 sek. (4 min. 10 sek.), liczba wykonanych operacji OLAP wyniosła 5 275 770, natomiast liczba utworzonych wierszy w kostce OLAP wyniosła 991 675. Wydajność systemu charakteryzują odpowiednio: średnia liczba przeliczanych dekretów tabeli faktów 231,8 dekretów na sekundę, średnia liczba operacji OLAP 21103,1 operacji na sekundę oraz średnia liczba przeliczanych iloczynów kartezyjskich na dekret 91¹.

¹ Specyfikacja systemu: Serwer (Part No.: 798454G): IBM x3455, 2xAMD Opteron Dual Core 2218 95W 2.6GHz/2x1MB L2, 4x512MB, 160GB SATA; System Operacyjny: SUSE LINUX Enterprise Server 10; RAID (Part No.: 18147VH): IBM Express DS4200 Model 7V, 16xHDD Express DS4200 500GB 7.2K SATA EV-DDM (Part No.: 42D0389); QLogic 4Gb FC Dual-Port PCIe HBA dla IBM System x (Part No.: 39R6527).

KNT_NAME	KNT_ID	KNT_KNT	DDO_WN	DDO_MA	DDO_OCWN	DDO_OCMA	DDO_OC
Plan Kont	0	brak	147 447 771 263,953	147 447 771 263,274	863 673	920 587	1 784 260
Aktywa trwałe	1		30 882 589,99	18 505 210,44	42	46	88
Operacje z udziałem środków pieniężnych i operacje międzybankowe	336	1	97 104 905 375,5607	98 234 569 174,6214	285 717	155 419	441 136
Kasa	337	10	364 873 519,1601	361 976 479,37	113 821	13 187	127 008
Gotówka/banknoty oraz monety/	338	100	364 853 749,6601	361 972 151,37	113 810	13 180	126 990
Gotówka w kasie PLN	339	100-1	339 523 748,9001	339 543 253,16	113 270	12 325	125 595
Gotówka w walucie obcej USD	340	100-1-787	2 694 775,77	2 297 742,27	191	211	402
Gotówka w walucie obcej - GBP	341	100-1-789	2 216 411,44	1 901 212,99	85	65	150
Gotówka w walucie obcej - CHF	342	100-1-797	32 355,46	22 978,12	16	8	24
Gotówka w walucie obcej - SEK	343	100-1-798	0,00	0,00	0	0	0
Gotówka w walucie obcej EURO	344	100-1-978	9 617 488,09	9 045 444,83	181	198	379
Gotówka w złotych w bankomatach	345	100-2	10 768 970,00	9 161 520,00	67	373	440
Cześć podręcznicze	346	101	0,00	0,00	0	0	0
Złoto niemonetarne i inne metale szlachetne	349	102	0,00	0,00	0	0	0
Znaki wartościowe	352	103	19 769,50	4 328,00	11	7	18
Cześć podręcznicze własnej emisji	358	104	0,00	0,00	0	0	0
Cześć bankierskie	359	105	0,00	0,00	0	0	0
Należności i zobowiązania wobec banku centralnego	366	11	17 541 926 675,84	17 636 036 840,97	963	857	1 820
Należności normalne i rezerwy celowe od podmiotów finansowych	434	12	50 139 004 174,5402	49 258 114 635,7101	105 769	98 856	204 625
Należności pod obserwacją od podmiotów finansowych i rezerwy cel	4 027	13	0,00	0,00	0	0	0
Należności poniżej standardu i rezerwy celowe od podmiotów finans	6 420	14	0,00	0,00	0	0	0
Należności wątpliwe i rezerwy celowe od podmiotów finansowych	8 818	15	0,00	0,00	0	0	0
Należności stracone i rezerwy celowe od podmiotów finansowych	10 449	16	0,00	0,00	0	0	0
Zobowiązania, należności i rezerwy celowe od podmiotów finansowy	11 403	17	28 028 745 209,41	29 949 641 288,6499	24 769	32 923	57 692
konto wolne	12 477	18	0,00	0,00	0	0	0
Inne operacje z podmiotami finansowymi	12 478	19	1 030 355 796,58	1 028 799 929,92	40 395	9 596	49 991
Operacje z podmiotami niefinansowymi	12 586	2	3 334 997 874,36	2 859 554 151,5802	188 590	165 828	354 418
Operacje z instytucjami rządowymi i samorządowymi	38 785	3	203 710 745,03	182 664 748,10	29 649	4 071	33 720
Papiery wartościowe	43 236	4	2 272 334 339,58	1 600 416 756,74	233	128	361
Operacje różne	44 898	5	44 172 093 572,7999	44 089 485 437,9016	243 499	376 601	620 100
Fundusze własne	48 827	6	142 523 740,65	280 310 123,16	1 009	2 130	3 139
Koszty i straty	48 877	7	28 615 331,23	131 109 311,37	63 714	55 564	119 278

Rys. 2. Przykład raportu OLAP z widokiem Planu Kont

Testy wydajnościowe systemu pokazały, że w ASQL możliwe jest projektowanie systemów OLAP przetwarzających ogromne zbiory danych. Przy założeniu liniowości procesu, przy niezmiennych uwarunkowaniach sprzętowych można dziennie przetwarzać miliony transakcji, skracając proces ładowania systemu do akceptowalnych wartości czasowych na poziomie dwóch, trzech godzin. Zaznaczmy, że możliwe jest przetwarzanie współbieżne, wykorzystujące własności dynamiczne kostek OLAP w ASQL (przetwarzamy tylko nowe fakty), oraz wykorzystanie metody akceleracji operacji ładowania kostek za pośrednictwem tabel przejściowych umieszczonych w pamięci RAM systemu. Metody te pozwalają na ponaddziesięciokrotne przyspieszenie procesu – w prezentowanym rozwiązaniu nie było potrzeby zastosowania tej metody.

4. Scalanie transakcji – integracja warstwy aplikacyjnej SART z hurtownią danych

Moduł scalania transakcji w SART polega na transformacji danych z modelu systemu księgowego (dekrety Księgi Głównej banku; rys. 3A) na model danych transakcji bankowych (rys. 3B). We wdrożonym rozwiązaniu proces scalania transakcji identyfikuje konta stron transakcji, korzystając z wymiarów Planu Kont hurtowni danych jako tabeli słownika.

Na rysunku 3 przedstawione zostały dwa modele logiczne wykorzystywane w działalności operacyjnej banku:

A) model księgowy, w którym zapis polega na dekretowaniu pojedynczych operacji na odpowiednie pozycje WN lub MA,

B) model transakcyjny, w którym jeden rekord danych zawiera informacje o kontach (pozycja 1): źródłowym (ID dekretu: 281 187), docelowym (ID dekretu: 281 188) oraz kwocie operacji (pozycja 2, zaznaczona kwota operacji 2 000 000); model ten jest zgodny z wymaganiami ustawy [Ustawa... 2001].

A

TDD_ID	TDD_KOD1	TDD_KOD3	TDD_KOD4	TDD_IDMD	TDD_IDKT	TDD_UWAG	TDD_WN	TDD_MA
281 188	2	1022	03	21 687	460		0,00	2 000 000,00
281 187	2	2142	03	21 687	11 625	ST-07-36993	2 000 000,00	0,00
281 190				1 301	45 003	004	2 004 076,80	0,00
281 189				4	45 003	001	0,00	2 004 076,80
281 186	2	1023	03	21 687	460		0,00	4 076,71
281 185	2	2192	03	21 687	11 625	ST-07-36993	4 076,71	0,00
281 184	2	2120	03	21 687	11 625	ST-07-36993	0,00	4 076,71
281 183	2	2192	03	21 687	12 426	ST-07-36993	4 076,71	0,00
281 182	2	2191	09	21 687	12 426	ST-07-36993	0,00	271,78
281 181	2	2191	09	21 687	48 893	ST-07-36993	271,78	0,00
281 180	2	2193	03	21 687	12 426	ST-07-36993	0,00	3 804,93
281 179	2	2193	03	21 687	12 432	ST-07-36993	3 804,93	0,00

B

TTD_ID	TTD_IDDS	TTD_IDKS	TTD_IDMS	TTD_IDDD	TTD_IDKD	TTD_IDMD	TTD_AM	TTD_I...	TTD...
1	281187	40	21687	281188	41	21687	2000000	1	2
2	281109	40	21865	281200	41	21865	1000000	1	1
3	281211	40	21713	281212	41	21713	2500000	1	2
4	281223	40	21832	281224	41	21832	1500000	1	1
5	281235	40	21687	281236	41	21687	2000000	1	2
6	281247	40	21735	281248	41	21735	1000000	1	1
7	281259	42	21686	281260	41	21686	4700000	1	4

Rys. 3. Przykład reprezentacji danych dwóch modeli logicznych: księgowego (A) i transakcyjnego (B) w systemie SART

Scalenie dwóch różnych modeli logicznych możliwe było dzięki identyfikacji obiektów hurtowni danych (tabela faktów dekretów) i wykorzystaniu wymiaru Planu Kont jako metadanych w warstwie aplikacyjnej systemu SART.

Na rysunku 4 przedstawiony został formularz definiowania reguł scalania transakcji na podstawie dekretów księgowych. Zaznaczono nań definicje kont: źródłowego (ID: 11 625) i docelowego (ID: 460). Na podstawie tej definicji wszystkie dekrety oznaczone jednym identyfikatorem operacji, w których występują powyższe konta, scalone zostaną do transakcji: „Wypłata kapitału z lokaty (przelew wewnętrzny)”. Stworzony pełny rejestr definicji transakcji pozwala na automatyzację

procesu transformacji dekretów do postaci transakcji bankowych wymaganych przez regulator.

Definicje transakcji dekretów Dziennika Dokumentów ...

Edytor definicji transakcji Dziennika Dokumentów

Nazwa definicji transakcji:

Typ transakcji

Definicja typu transakcji:

Kategoria transakcji: Kod kategorii transakcji:

Typ transakcji: Kod typu transakcji:

Sposób wydania dyspozycji: Kod sposobu wyd. dysp.:

Konto źródłowe transakcji

Definicja konta źródłowego transakcji:

ID Konta źródłowego:

Numer konta źródłowego:

Nazwa konta źródłowego:

Wartość pola "Kod3":

Konto docelowe transakcji

Definicja konta docelowego transakcji:

ID konta docelowego:

Numer konta docelowego:

Nazwa konta docelowego:

Wartość pola "Kod3":

Zarejestrowane definicje transakcji Dziennika Dokumentów DeBank

Strona raportu: 1/1; Liczba elementów w raporcie: 39

TRD_ID	TRD_NAME	TRD_I...	KNT_KNT	KNT_NAME	ATD_K...	TRD_I...	KNT_KNT	KNT_NAME
30	Wypłata z kapitału lokat...	38	27103-2403-1	Depozyty 3 dniowe w PLN	2142	39	27003-10-1	Rachunki bieżące w
31	Wypłata z kapitału lokat...	40	17102-2414-1	Depozyty 14 dniów w PLN	2142	41	12012-10-1	Rachunek bieżący w
32	Wypłata z kapitału lokat...	42	17102-1401-1	Depozyty 1 dniowe w PLN	2142	41	12012-10-1	Rachunek bieżący w
33	Wypłata z kapitału lokat...	43	27102-2403-1	Depozyty 3 dniowe w PLN	2142	44	27002-10-1	Rachunki bieżące w

Rys. 4. Przykład formularza definicji transakcji, w którym to formularzu odbywa się definiowanie stron transakcji bankowej

Rysunek 5 przedstawia fragment raportu OLAP w kontekście Planu Kont z wyszczególnionym kontem „Rachunek bieżący w PLN” użytym we wcześniejszych przykładach (rysunki 3 i 4) definicji scalania transakcji. Na rysunku zaznaczono pozycję zagregowanych obrotów na koncie „Rachunek bieżący w PLN” (ID konta: 460), które zostało wyszczególnione na rys. 3A (pozycja pierwsza, ID konta kolumna „TDD_IDKT”).

Metryka Formularza

Nazwa raportu: Plan Kont DeBank OLAP (niepełny)

KNT_NAME	KNT_ID	KNT_KNT	DDO_WN	DDO_MA	DDO_OCWN	DDO_OCMA	DDO_OC
Plan Kont	0	brak	147 447 771 263,953	147 447 771 263,274	863 673	920 587	1 784 260
Aktywa trwałe	1	0	30 882 589,99	18 505 210,44	42	46	88
Operacje z udziałem środków pieniężnych i operacje międzybankowe	336	1	97 104 905 375,5607	98 234 569 174,6214	285 717	155 419	441 136
Kasa	337	10	364 873 519,1601	361 976 479,37	113 821	13 187	127 008
Należności i zobowiązania wobec banku centralnego	366	11	17 541 926 675,84	17 636 036 940,97	963	857	1 620
Należności normalne i rezerwy celowe od podmiotów finansowych	434	12	50 139 004 174,5402	49 259 114 635,7101	105 769	98 856	204 625
Rachunki bieżące	435	120	34 198 738 224,3899	34 295 739 129,3497	102 106	95 559	197 665
Konto bieżące nostro	436	1200	579 376 708,88	574 652 079,71	1 661	2 181	3 842
Konto bieżące loro	455	1201	33 619 361 515,5099	33 721 087 049,6397	100 445	93 378	193 823
Rachunki bieżące loro pozostałych monetarnych instytucji finansowych	456	12012	33 619 361 515,5099	33 721 087 049,6397	100 445	93 378	193 823
Rachunek bieżący	459	12012-10	33 377 235 047,9199	33 375 504 970,7496	100 236	92 968	193 204
Rachunek bieżący w PLN	460	12012-10-1	32 705 257 522,3199	32 703 493 750,1997	95 623	89 053	184 876
Rachunki bieżące w USD	461	12012-10-787	116 879 929,25	116 905 366,35	1 685	1 551	3 236
Rachunki bieżące w GBP	462	12012-10-789	21 403 394,96	21 406 657,63	302	289	591
Rachunki bieżące w EURO	465	12012-10-978	533 694 201,39	533 694 156,57	2 426	2 075	4 501

Rys. 5. Fragment raportu OLAP

Dzięki opisanej możliwości identyfikacji obiektów łatwo można przejść od analiz transakcji bankowych do analiz OLAP dekretów opartych na Planie Kont oraz drugą stronę, czyli od raportów OLAP do transakcji bankowych, których dany raport dotyczy.

5. Hurtownia danych łańcuchów transakcji

Oryginalnym rozwiązaniem zawartym w systemie SART jest hurtownia danych łańcuchów transakcji oraz skojarzona z nim funkcjonalność OLAP, zwana analizą łańcuchów przepływu pieniądza (Cash Flow Chains Analysis, w skrócie CFCA). Zadaniem tego modułu jest ułatwienie analizy łańcuchów przepływu pieniądza pomiędzy rachunkami bankowymi.

Podstawowymi raportami w prezentowanym rozwiązaniu są:

- informacje o łańcuchach zaczynających się od zadanego konta, zawierających dane konto bądź kończących się na danym koncie,
- informacje o cyklach w łańcuchach (tzn. gdy konto początkowe i końcowe w łańcuchu są takie same),
- informacje o łańcuchach rozpoczynających się, zawierających bądź kończących się na zadanej sekwencji kont,
- informacji o zagregowanych przepływach pieniądza względem konta bądź transakcji,
- informacje o budowie wskaźników ilościowych dotyczących liczby łańcuchów zawierających daną transakcję bądź konto.

Od strony algorytmicznej tego typu struktury danych przedstawiamy w postaci grafu, jednak w przypadku łańcucha transakcji bankowych należy pamiętać, że operacje na grafie są silnie zdeterminowane budową pojedynczych ścieżek (łańcuchów). Uwarunkowanie to powoduje, że autonomiczne operacje na grafie muszą być zawsze zdefiniowane w kontekście całego diagramu zdefiniowanego na grafie.

Od strony technicznej hurtownia danych, jak również zawarte w niej struktury OLAP stanowią rozszerzenie standardu SQL i pozwalają operować na nowym typie danych SQL, jakimi są „łańcuchy obiektów”, podobnie jak ma to miejsce w przypadku typów prostych SQL. Niezwykle istotne jest tu przyspieszenie operacji analitycznych polegających na indeksowaniu łańcuchów, a także operacji na zbiorach zawartych w łańcuchach oraz możliwości tworzenia reguł. Istnieje również możliwość rozszerzenia funkcjonalności CFCA o proces ekstrakcji reguł z hurtowni łańcuchów, wykorzystując do tego metody eksploracji danych (algorytmy grupowania, asocjacji czy drzew decyzyjnych).

Raport CFCA, przedstawiony na rys. 6, pokazuje wszystkie łańcuchy transakcji zaczynających się od konta o ID:22921, a kończących się na koncie o ID:14037 (bez względu na długość łańcucha), tym samym pozwalając na wizualizację graficzną wyników łańcuchów transakcji w postaci drzewiastej, która odpowiada grafowi przepływu pieniądza pomiędzy kontami. Każdy rekord raportu reprezentuje transak-

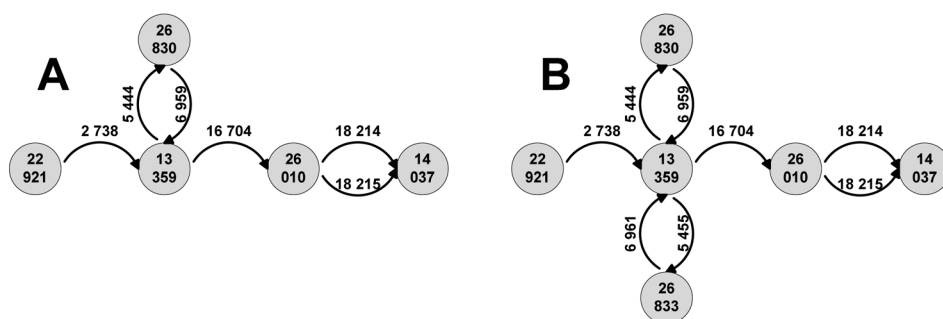
cję zawartą w łańcuchu, przy czym hierarchia drzewa determinuje zależność przepływu pieniądza pomiędzy kontami – konto docelowe TEX_IDND elementu nadrzędnego odpowiada kontu źródłowemu TEX_IDNS elementu podrzędnego. Dzięki zastosowaniu tej metody wizualizacji możemy pokazać w sposób przejrzysty i łatwy duże zbiory danych. Zaznaczmy, że stosowanie klasycznej metody wizualizacji w postaci diagramu przejść przy dużej liczbie wierzchołków i krawędzi powoduje że analiza diagramu z powodu jego wielkości i nieczytelności staje się niemożliwa.

Na rysunku 7 przedstawiono diagram transakcji stworzony na podstawie dwóch pierwszych poddrzew rys. 6 – węzły grafu przedstawiają konta (numery identyfikatorów kont kolumn TEX_IDNS, TEX_IDND), krawędzie reprezentują transakcje (identyfikatory transakcji zawarte w drugiej kolumnie TEX_ID). Należy zwrócić uwagę, że analiza przepływu pieniądza, wykorzystująca reprezentację diagramów, wymaga dużo większego nakładu pracy niż w przypadku wizualizacji w postaci lasu drzew pokazanego na rys. 6.

TEX_ID	DT_SDATE	TEX_BNOS	TEX_IDNS	TEX_BNOD	TEX_IDND	TEX_AMNT	TEX_DTYP	T	TE...
2 743	2 738	2007-12-28	13 200 019	22 921	20 601 054	13 359	247,50	ELIXR	0 1
9 323	5 444	2008-01-04	20 600 002	13 359	12 401 789	26 830	5 000,00	ELIXR	0 3
16 763	6 959	2008-01-04	12 401 789	26 830	20 601 054	13 359	47,46	ELIXR	0 3
45 033	16 704	2008-01-08	20 600 002	13 359	10 101 010	26 010	30,00	ELIXR	0 5
57 792	18 214	2008-01-08	10 101 010	26 010	20 601 054	14 037	610,00	ELIXR	0 5
56 288	18 215	2008-01-08	10 101 010	26 010	20 601 054	14 037	540,00	ELIXR	0 5
10 654	5 455	2008-01-04	20 600 002	13 359	12 401 789	26 833	10 000,00	ELIXR	0 3
16 886	6 961	2008-01-04	12 401 789	26 833	20 601 054	13 359	24,28	ELIXR	0 3
45 156	16 704	2008-01-08	20 600 002	13 359	10 101 010	26 010	30,00	ELIXR	0 5
55 915	18 214	2008-01-08	10 101 010	26 010	20 601 054	14 037	610,00	ELIXR	0 5
56 411	18 215	2008-01-08	10 101 010	26 010	20 601 054	14 037	540,00	ELIXR	0 5
44 847	16 704	2008-01-08	20 600 002	13 359	10 101 010	26 010	30,00	ELIXR	0 5
55 606	18 214	2008-01-08	10 101 010	26 010	20 601 054	14 037	610,00	ELIXR	0 5
56 102	18 215	2008-01-08	10 101 010	26 010	20 601 054	14 037	540,00	ELIXR	0 5
2 745	2 739	2007-12-28	13 200 019	22 921	20 601 054	13 359	247,50	ELIXR	0 1
9 325	5 444	2008-01-04	20 600 002	13 359	12 401 789	26 830	5 000,00	ELIXR	0 3
16 765	6 959	2008-01-04	12 401 789	26 830	20 601 054	13 359	47,46	ELIXR	0 3
45 035	16 704	2008-01-08	20 600 002	13 359	10 101 010	26 010	30,00	ELIXR	0 5
55 794	18 214	2008-01-08	10 101 010	26 010	20 601 054	14 037	610,00	ELIXR	0 5
56 290	18 215	2008-01-08	10 101 010	26 010	20 601 054	14 037	540,00	ELIXR	0 5
10 656	5 455	2008-01-04	20 600 002	13 359	12 401 789	26 833	10 000,00	ELIXR	0 3
44 849	16 704	2008-01-08	20 600 002	13 359	10 101 010	26 010	30,00	ELIXR	0 5
2 747	2 740	2007-12-28	13 200 019	22 921	20 601 054	13 359	114,00	ELIXR	0 1
2 749	2 741	2007-12-28	13 200 019	22 921	20 601 054	13 359	70,50	ELIXR	0 1
2 751	2 742	2007-12-28	13 200 019	22 921	20 601 054	13 359	70,00	ELIXR	0 1
2 753	2 743	2007-12-28	13 200 019	22 921	20 601 054	13 359	70,00	ELIXR	0 1
2 755	2 744	2007-12-28	13 200 019	22 921	20 601 054	13 359	71,00	ELIXR	0 1
9 335	5 444	2008-01-04	20 600 002	13 359	12 401 789	26 830	5 000,00	ELIXR	0 3
16 775	6 959	2008-01-04	12 401 789	26 830	20 601 054	13 359	47,46	ELIXR	0 3
45 045	16 704	2008-01-08	20 600 002	13 359	10 101 010	26 010	30,00	ELIXR	0 5
55 804	18 214	2008-01-08	10 101 010	26 010	20 601 054	14 037	610,00	ELIXR	0 5

Poszczególne kolumny na rysunku oznaczają: TEX_ID – identyfikator łańcucha, TEX_ID – identyfikator transakcji, DT_SDATE – data transakcji, TEX_BNOS – bank zleceniodawcy, TEX_IDNS – identyfikator konta zleceniodawcy, TEX_BNOD – bank odbiorcy, TEX_IDND – identyfikator konta odbiorcy, TEX_AMNT – kwota transakcji.

Rys. 6. Fragment raportu analizy CFCA



Rys. 7. Diagramy transakcji pokazanej na rys. 6 (diagram A – pierwsze poddrzewo, diagram B – scalone dwa pierwsze poddrzewa)

Zauważmy, że diagram B na rys. 7, mimo iż składa się tylko z 6 wierzchołków i 8 krawędzi, nie wskazuje bezpośrednio sposobu ekstrakcji do postaci z diagramu A w ujęciu przepływu pieniądza pomiędzy rachunkami. W przypadku dużych zbiorów danych, złożonych z setek czy tysięcy wierzchołków lub krawędzi, analiza struktur diagramów pod kątem przepływu pieniądza praktycznie staje się niemożliwa.

6. Podsumowanie i plan przyszłych prac

W artykule przedstawiono konkretne problemy technologiczne, które zostały rozwiązane za pomocą ASQL i zaimplementowane w produkcyjnym systemie AML SART. Wyróżniono kilka krytycznych problemów związanych z modelowaniem struktur danych hurtowni danych OLAP oraz integracją warstwy aplikacyjnej i analitycznej (transakcje bankowe, hurtownia i struktury OLAP). Pokazano też budowę łańcuchów transakcji, które stanowią innowacyjną koncepcję rozszerzenia istniejących modeli hurtowni danych. Przeprowadzone eksperymenty pokazały, że zaproponowany model pozwala na pełną integrację danych operacyjnych prezentowanych w dekretach oraz transakcjach ze strukturami hurtowni danych OLAP poprzez identyfikatory poszczególnych obiektów. Zaprezentowana została spójna architektura hurtowni danych, oparta na niejednorodnym wymiarze Planu Kont, oraz struktury OLAP, bazujące na hurtowni danych. Architektura ta ma zasadnicze znaczenie dla przyszłych prac badawczo-rozwojowych, związanych z przygotowaniem konkretnych modeli analiz i scenariuszy analitycznych służących do typowania transakcji powiązanych i podejrzanych, zgodnie z wymaganiami zawartymi zarówno w bieżących, jak i w planowanych regulacjach prawnych.

Należy podkreślić, że przedstawione rozwiązania i modele nie są obecnie stosowane ani w standardach modeli hurtowni danych i OLAP [Zdanowicz 2004; Inmon 2005; Wrembel, Koncilia 2006; Turban i in. 2007], ani też w rozwiązaniach komercyjnych (z wyjątkiem systemu SART). Zaznaczmy też, że zaproponowane rozwiązania mają charakter uniwersalny, ponieważ bazują na danych i strukturach danych

systemów informacyjnych banku (Księga Główna), stanowiących międzynarodowy standard księgowy wykorzystywany nie tylko w bankach, ale również w większości instytucji finansowych.

W artykule wskazano ponadto kilka możliwości związanych z analizą danych i typowaniem transakcji powiązanych i podejrzanych w ramach przeciwdziałania praniu pieniędzy. Tu przede wszystkim należy wyróżnić:

- analizy oparte na strukturze OLAP Planu Kont (zagregowana wartość operacji, ich liczba oraz kontekst wynikający z planu kont),
- analizy porównawcze wynikające z integracji kostki OLAP Planu Kont z transakcjami bankowymi,
- analizy związane z przepływem pieniądza w łańcuchach transakcji.

Budowa konkretnych modeli analitycznych opartych na zaprezentowanych strukturach danych będzie przedmiotem naszych dalszych prac. Podejmiemy m.in. prace nad algorytmami rozpoznawania wzorców, tworzeniem baz reguł decyzyjnych oraz analizą funkcjonalną i implementacją modułu raportującego zgodnie z projektem ustawy.

Literatura

- Alexander R.C.H., *Insider Dealing and Money Laundering in the EU: Law and Regulation*, Ashgate Publ. 2007.
- Dyrektywa 2005/60/WE Parlamentu Europejskiego i Rady z dnia 26 października 2005 r. w sprawie przeciwdziałania korzystaniu z systemu finansowego w celu prania pieniędzy oraz finansowania terroryzmu.
- Inmon W.H., *Building the Data Warehouse*, Wiley 2005.
- Kingdon J., *AI fights money laundering*, „IEEE Intelligent Systems”, May/June 2004.
- Korczak J., Merchelski W., Oleszkiewicz B., *A new technological approach to money laundering discovery using analytical SQL server*, [w:] *Advances Information Technologies for Management – AITM 2008*, eds. J. Korczak, H. Dudycz, M. Dyczkowski, UE, Wrocław 2008.
- Muller W.H., Kalin C.H., Goldsworth J.G. (eds.), *Anti-Money Laundering: International Law and Practice*, Wiley 2007.
- Philippsohn S., *Money laundering on the Internet*, „Computers & Security” 2001, no. 20.
- Rządowy projekt ustawy o zmianie ustawy o przeciwdziałaniu wprowadzaniu do obrotu finansowego wartości majątkowych pochodzących z nielegalnych lub nieujawnionych źródeł oraz o przeciwdziałaniu finansowaniu terroryzmu oraz o zmianie niektórych innych ustaw, <http://orka.sejm.gov.pl/proc6.nsf/opisy/1660.htm>.
- Rozporządzenie Ministra Finansów z dnia 21 września 2001 r. w sprawie określenia wzoru rejestru transakcji, sposobu jego prowadzenia oraz trybu dostarczania danych z rejestru Generalnemu Inspektorowi Informacji Finansowej, tekst opracowany na podstawie: DzU 2001 nr 113, poz. 1210, zm. DzU 02.32.308, zm. DzU 03.101.935.
- Schott P.A., *Reference Guide to Anti-money Laundering and Combating the Financing of Terrorism*, World Bank Publ. 2006.
- The Forty Recommendations, Report Financial Action Task Force on Money Laundering*, FATF-GAFI, <http://www.fatf-gafi.org/>.
- Turban E. i in., *Business Intelligence*, Prentice Hall 2007.

Unger B., Busuioc E.M., *The Scale and Impacts of Money Laundering*, E. Elgar Publ. 2007.

Ustawa z dnia 16 listopada 2000 r. o przeciwdziałaniu wprowadzaniu do obrotu finansowego wartości majątkowych pochodzących z nielegalnych lub nieujawnionych źródeł oraz o przeciwdziałaniu finansowaniu terroryzmu, tekst opracowany na podstawie: DzU 2000 nr 116, poz. 1216; DzU 2003 nr 153, poz. 1505, 2004 nr 62, poz. 577, nr 96, poz. 959, nr 116, poz. 1203, 2005 nr 183, poz. 1538, 2006 nr 104, poz. 708 i 711, nr 157, poz. 1119.

Wrembel R., Koncilia C., *Data Warehouses and OLAP: Concepts, Architectures and Solutions*, IRM Press 2006.

Zdanowicz J.S., *Detecting Money Laundering and Terrorist Financing via Data Mining*, Comm. of the ACM, 5/2004, vol. 47.

SELECTED ISSUES OF DATA WAREHOUSES AND OLAP TECHNOLOGY IN THE DISCOVERY OF MONEY LAUNDERING OPERATIONS – A CASE STUDY

Summary: In the paper a few important issues of data warehouse technology for Anti-Money Laundering (AML) applications are presented. The considerations relate to the software platform, called SART Analytical SQL Server, based on the data warehouse technology, OLAP and data mining methods. Architecture and main functionalities of the SART system are demonstrated using data coming from a bank of information system. The paper describes how to build data warehouse structures to identify suspicious transactions and objects, based on the concepts of multi-dimensional, heterogenic ROLAP cubes and transaction chains. In conclusion, the advantages of the proposed solution are summed up in the context of new trends in OLAP as well as in Business Intelligence research and development.