RESEARCH PAPERS OF THE WROCŁAW UNIVERSITY OF ECONOMICS
No. 35 — 2008
Advanced Information Technologies for Management – AITM 2008

Jerzy Korczak(1), Wojciech Merchelski(2), Błażej Oleszkiewicz

(1)Wrocław University of Economics, (2)TETA SA Wrocław
jerzy.korczak@ue.wroc.pl
w.merchelski@teta.com.pl
blajan@poczta.fm

A NEW TECHNOLOGICAL APPROACH TO MONEY LAUNDERING DISCOVERY USING ANALYTICAL SQL SERVER

Abstract: In the paper a new idea of system architecture for money laundering discovery is presented. The considerations relate to the software platform, called SART Analytical SQL Server, that was used to build complete anti-money laundering (AML) applications. This software platform is based on the data warehouse technology and contains advanced analytical functions, OLAP and data mining methods. Architecture and main functionalities of the SART system have been demonstrated using data coming from a bank information system. Our study will demonstrate how to build data warehouse structures to identify suspicious transactions and objects, based on the concepts of multidimensional, 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.

1. Introduction

New regulations and directives in the EU countries that are currently under development engage all financial institutions to implement efficient methods and tools in order to prevent money laundering. The problem remains open in spite of large number of research work [Alexander 2007; Unger 2007; Zdanowicz 2004] and software solutions (SART[1], DiMon[2], SPERT[3]). According to the Financial Action

[1] SART – System for Analysis and Registration of Transactions (in Polish: System Analiz i Rejestracji Transakcji – SART), AML software based on Analytical SQL Server developed by TETA SA Company.

[2] DiMon – Polish AML software developed by Signity Company.

[3] SPERT – Polish AML software developed by BSB Company.

Task Force (FATF) money launderers use increasingly sophisticated techniques to provide advice and assistance in laundering criminal funds [Schott 2006]. From economic point of view, money laundering is now one of the world's largest trades (Financial Crimes Enforcement Network). Therefore the research on new computational and intelligent techniques to discover money laundering transactions becomes critical [Kingdon 2004; Zdanowicz 2004].

The important factor of the efficiency of money laundering detection are techniques and methods applied in bank information systems [The Forty Recommendations 2004]. However, existing solutions concerning discovering money laundering transactions show a number of weaknesses, notably:

– frequently, such systems are not built on data warehouse technology;
– systems do not support advanced tools for analysing transaction chains;
– lack of support for efficient identification of suspected objects, particularly in the layer of data warehouse structures;
– finally, there is no uniform model for defined and archived analysis as well as for constructing scenarios of analyses.

Analysis of legislation in this area as well as technical and organizational situation in the banks indicate the necessity to develop new technological solutions for financial information systems. Before we introduce our view on this matter let us recall the concept of "money laundering". There is a large number of definitions [Muller 2007; Palermo Convention, http://www.undcp.org/adhoc/palermo/convmain.html]. According to the Directive 2005/60/EC of the European Parliament [Directive 2005] "money laundering" shall be understood as:

– "the conversion or transfer of property, knowing that such property is derived from criminal activity or from an act of participation in such activity, for the purpose of concealing or disguising the illicit origin of the property or of assisting any person who is involved in the commission of such activity to evade the legal consequences of his action;
– the concealment or disguise of the true nature, source, location, disposition, movement, rights with respect to, or ownership of property, knowing that such property is derived from criminal activity or from an act of participation in such activity;
– the acquisition, possession or use of property, knowing, at the time of receipt, that such property was derived from criminal activity or from an act of participation in such activity;
– participation in, association to commit, attempts to commit and aiding, abetting, facilitating and counselling the commission of any of the actions mentioned in the foregoing points."

The above Directive concerns credit and financial institutions, auditors, tax advisers, notaries, real estate advisers, casinos, and other natural and legal persons engaged in goods trading that involves payments made in cash to the amount of EUR 15000 or more.

This paper is aimed at presenting new approach to solving the above mentioned weaknesses of commercially available AML software technology. Firstly, we will demonstrate the idea of integrating application layer of AML system with OLAP data warehouse that in sequence is used for advanced analysis, including chains of transactions. Secondly, we will also point out the aspect of identifying analysed subjects including structures of OLAP data warehouse. Thirdly, we will describe the aspect of creating and archiving analyses, and building scenarios of analyses. We will discuss only financial institutions, but the proposed solution can easily be adapted to other types of activities. In conclusion of the paper, we will quote major characteristics of the system, showing how SART can be useful in discovering money laundering transactions.

This paper is divided into five sections. Section 2 will describe several areas that relate to AML system design, notably legislative aspects, methodological and functional issues. In section 3 technology of new Analytical Server SQL will be detailed. Server is supposed to be a foundation for AML Systems[4]. Functional specifications of AML SART system will be provided in section 4. The last section is a summary of ongoing project and describes future research on AML SART system.

2. Fundamental aspects of AML system design

Advanced information system for financial institutions dedicated to discovering money laundering operations should consider three main areas:

– legislative area, related to money laundering legislations,
– methodological area that concerns specificity of the problem solving and heterogeneity of data and information, and
– functional area, related to quality and performance of applied tools.

Legislative area

As the European Parliament has approved new legal regulations later published in [Alexander 2007; Directive 2005], Poland, being a Member State of the European Union, has introduced laws prohibiting introducing into financial turnover property derived from illegal sources, and counter-financing of terrorism [The Minister of Finance Regulation 2001; Directive 2005; Draft 2008]. New regulations support previous legislation in preventing and fighting money laundering which is a result of new requirements for registering transactions that exceed 15000 Euros. Regulations introduce range of new duties for financial institutions. When viewed from the perspective of bank information system, the risk of money laundering is currently one of the most important challenges [Draft 2008; Muller et al. (Eds.) 2007].

Methodological area

Main goal of information system that would prevent money laundering is to implement a method of managing risk evaluation of money laundering. Countermeasures against such practices are carried out in financial institutions on many levels in respective processes as well as in areas regarding heterogeneity of data

[4] AML Systems – software dedicated to Anti-Money Laundering (AML).

model [Alexander 2007; Unger, Busuioc 2007]. It is also advisable to pay attention to those issues in the context of analytical models of risk applied in banks.

Majority of models for risk analysis in finance (e.g. risk of rate of currency, credit risk, risk of liquidity, etc.) involve evaluating only a part of measurable values (e.g. rate of currency, value of credit, etc.) with respect to the scope of analysis. When risk of money laundering is considered not only definite value, but rather a specific feature called risk of money laundering, usually associated with different areas of bank operations, should be taken into account. Therefore, new solutions for money laundering risk have to be defined for various financial areas.

In terms of system scope, new solutions should fulfil the following conditions:

- identification of money laundering risks that take into consideration multidimensional structure of problems and heterogeneous character of data,
- identification and clustering of business areas where money laundering might occur,
- establishing risk weighting within analysed business areas,
- uniform way of evaluating money laundering risk within analysed business areas (e.g. client, groups of clients, accounts, groups of accounts, products, etc.).

Functional area

Tools applied for data storage and analysis must be designed suitably to the scale of the problem. In particular, designers have to pay attention to the fact that all transactions should be subjected by bank in long term time horizon, that can be translated directly to data volume designed for storage. Therefore, use of OLAP data warehouse with appropriate software oriented towards Business Intelligence and Computational Intelligence [Turban et al. 2005] becomes necessary [Inmon 2005; Zdanowicz 2004; Wrembel, Koncilia 2006].

Choosing hardware and software solutions, designers can consider capability of mirroring in the data warehouse and software tools that have to apply specific to financial institutions' OLAP data structures [Wrembel, Koncilia 2006], such as the structure of bank organization; General Ledger, risk profile, etc. Above mentioned structures exist in bank activities in the form of non-uniform hierarchy of objects. Hence, it is difficult to determine exact representation of these objects in the databases, in data warehouses and tools, and sometimes it might even be impossible [Inmon 2005; Turban et al. 2005; Wrembel, Koncilia 2006].

Another technological limitation is related to lack of support for processing chains of objects, in other words sequences of financial transactions. Analysis of chains of transactions is one of the fundamental functions in preventing money laundering. Proposed solution concerns an extension of data warehouse model that allows to analyse effectively chains of bank transactions.

3. Description of Analytical SQL Server

The proposed Analytical SQL Server is one of the first integrated systems on the market. Analytical SQL Server's main applications include systems belonging to the class of Business Intelligence, such as financial applications (e.g. controlling, budgeting, forecasting activity, MiFID), banking trade (BASEL II, AML), insurances (Solvency II), industrial applications, medicine and health services. Extensions of Business Intelligence concern relational OLAP (ROLAP) with full support for multidimensional data processing (MOLAP). Integration of these two technologies (ROLAP and MOLAP) is directly transposed into facility of modelling and processing multidimensional data using standard SQL interface. This is rarely available in other solutions. System presented in the paper, called SART, is an example of solution where OLAP model has been built on non-uniform dimensions, and more specifically on Chart of Accounts.

Server's functional architecture has been schematically shown on Figure 1.

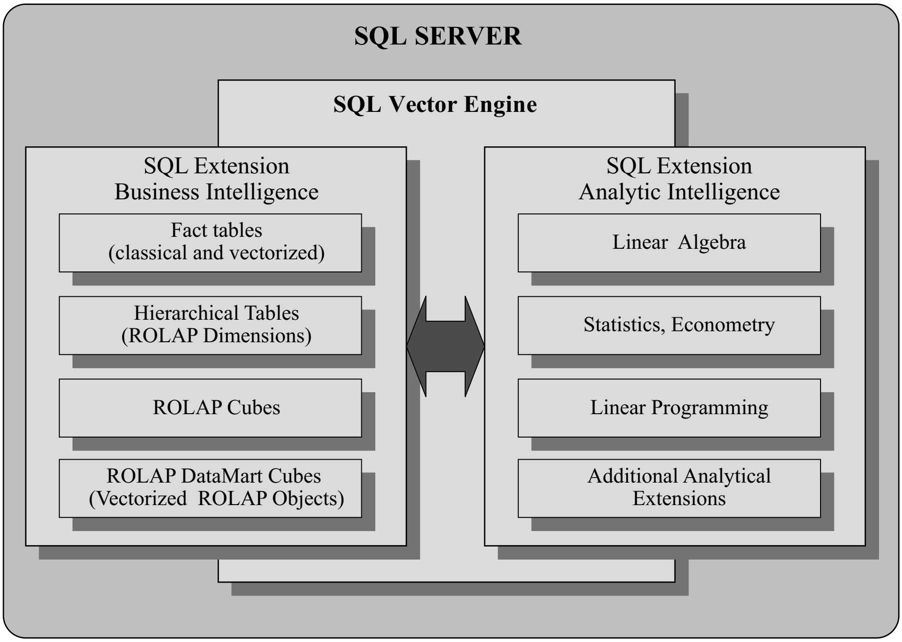


Figure 1. Architecture of Analytical SQL Server

Extensions of Analytical SQL Server allow also to use advanced analysis based on mathematical and statistical models. Similar to Business Intelligence features, the extensions of Computational Intelligence are built-in in the database management system and are available through standard SQL interface. This feature is also not found in currently commercialized systems.

During designing Analytical SQL Server a lot of attention has been directed to the system performance which resulted in very high efficiency of data processing. In one of the largest data warehouses (imported data concerned period of 14 months) all multidimensional analyses were performed within duration of time that did not exceed 10 seconds, where 98% analyses were carried out in less than 5 seconds. Above mentioned technological properties of Analytical SQL Server make implementation of data warehouses with advanced data analysis a low risk task, and allow to achieve very easily full user requirements.

Homogeneity of analytical system offered by Analytical SQL has also considerable influence on reducing license fees and implementation costs, because it does not require payment for third-party licenses as well as reducing integration cost related to different tools and systems (e.g. DBMS and statistical programmes).

4. SART as implementation of AML software based on Analytical SQL Server

In order to meet banking sector's requirements TETA SA has developed *System of Analysis and Registration of Transaction*, called SART. This is a complete and integrated analytical suite, fully compliant with the Draft Act requirements [Draft 2008]. SART is a modular system composed of twelve modules, schematically shown in Figure 2.

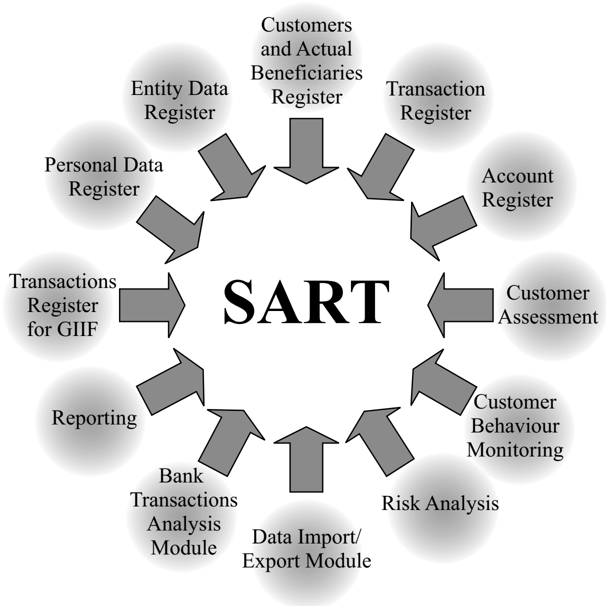


Figure 2. Major modules in SART system

Modules support essential functionalities of the system, more specifically:

- Data Import and Export;
- Data Registers;

– Object Group Management Module;
– Reporting Module;
– Analysis Module;
– Transactions Register for GIIF (General Inspector for Financial Information) Module.

Data Import and Export Modules provide an interface for uploading or saving data from different sources (SQL database, MS Office/Open Office files, text files, XML, WebService, etc.). This functionality provides also immediate information at the General Inspector's request. It should be noted that SART fully supports data import from the DefBank system in relation to banking transactions (transaction details), General Ledger and customers.

Data Registers assure two main functions:

– they support identification of properties required by the Draft Act and connected to the Responsible Institution's business (customer's properties, actual beneficiary's properties, business entity's properties, etc.);
– they enable categorisation of transaction-related events and are used for monitoring money flow, customer and actual beneficiary behaviour analysis, money laundering, financing terrorism and risk analysis.

Basic set of registers contains:

Personal Data Register	collects data concerning individuals involved in financial transactions.
Entity Data Register	contains data relating to entities involved in financial transactions.
Customers Register and Actual Beneficiaries Register	identifies customers and actual beneficiaries in the context of transaction register and banking transactions analysis. The two registers are closely connected with the *Personal Data Register* and *the Entity Data Register.*
Bank Account Register	contains bank accounts connected with the *Customers Register* and the *Actual Beneficiaries Register.*
Transaction Register	contains all transactions processed by the institution, together with FX rates.

Object Group Management Module provides comprehensive model for creating and managing groups of objects, in particular:

– black and white list of customers and individual customer groups;
– black and white list of bank accounts and individual account groups;
– customers, actual beneficiaries and high-profiled individuals.

Depending on client's requirements the group model can be modified according to individual needs. The system can be upgraded with new group models (e.g. business customer groups, bank product groups, etc.).

Reporting Module offers extensive reporting facilities allowing the user to present data according to user's preferences. This functionality has been achieved through ASQL Business Classes – an innovative solution enabling to convert intricate SQL database attribute classes to user-oriented classes (e.g. customer, account, beneficiary) and analytical rules (appropriately to carry out analyses with intuitive graphical user interfaces). SART reporting module also provides unique ability to save generated reports in a repository of the system, and then use the reports in analysis as an extra data source. Thanks to ASQL Business Classes all reports can be automatically labelled on the basis of the context they are used in. Thus, when looking through archived reports we have an access to complete information on every single report.

Analysis Module. Apart from the existing requirements regarding bank transactions analysis, the Draft Act introduces new requirements concerning monitoring of customer behaviour and customer assessment based on risk analysis. Currently system provides the following analysis:

Bank Transactions Analysis Module	enables detection of suspicious and linked transactions as well as money flow analysis.
Customer Behaviour Monitoring	contains set of customer behaviour analyses. The module can also be used in "Know Your Client" (KYC) programmes.
Risk Analysis	contains risk analyses attributed to customers, actual beneficiaries and bank transactions.
Customer Assessment	provides multi-level and multi-context customer assessment on the basis of scoring models.

As in its previous version, Bank Transactions Analysis Module in SART is based on analytical rules grouped into four categories: selection, aggregation, clustering and transactions chains. It is worthwhile to mention that SART, as the only system on the market, offers an advanced analysis of linked transactions (regulatory requirement). Bank Transactions Analysis Module may also save analysis results and then use them as data sources in analyses scenarios.

Customer Behaviour Monitoring analyses have been based on the data warehouse of bank transactions stored in the Transactions Register. The analysis produces qualitative and quantitative data on bank transactions in the context of the customer, actual beneficiary and linked transactions (in relation to accounts, customers and actual beneficiaries). Bank transactions analyses are performed in the context of the bank's General Ledger, hence there is no need to categorise transactions in the SART system again (the categorisation is imposed by the General Ledger).

Risk Analysis Module provides tools for modelling risks connected with suspicious and linked transactions in relation to customers and current beneficiaries.

Customer Assessment Module is designed to facilitate decisions regarding customer assessment in the context of various bank operations and on the basis of risk analysis. The module operates on the basis of scoring sheet model.

Transactions Register for GIFI (General Inspector for Financial Information) Module is compliant with the Minister of Finance regulation of the 21st September, 2001, concerning data transfer to the General Inspector for Financial Information.

To resume SART system, in relation to current legislations and business requirements, it meets all the basic criteria for modern AML systems, notably:
- modularity and scalability of the system functions;
- system integrity at the conceptual and logical levels (the full integration of the various data models);
- simple rules of system implementation and ease of use;
- high system performance that allows to carry out data analyses almost in real time.

5. Case study: Selected aspects of SART system implementation in commercial bank

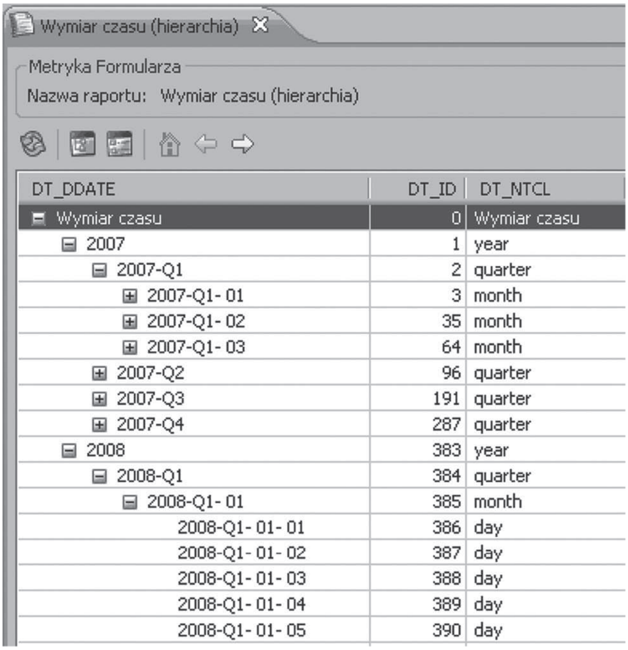
Defining the problem

In this section we will demonstrate several aspects of AML SART system implementation in one of the Polish banks. In the project, four main business requirements have been specified by the client:
- creating coherent data repository formed as multidimensional data warehouse, fed by bank's information systems and external data sources, enabling to start analyses preventing money laundering;
- automatic merging bank decrees into forms of transactions in accordance with the definition included in the Act;
- parametrisation of SART analytical modules that include set of parameters that enable to select bank transactions in accordance with the Act requirements taking linked transactions into consideration;
- creation of Transaction Register GIIF in accordance with the Act.

It might seem that it is a typical task related to data warehouse design using OLAP and functions that operate on multidimensional data structures. However, preliminary analysis carried out in the bank showed that the problem is more complex than initially thought. Preliminary works have demonstrated that:
- data concerning transactions carried in the bank are stored as defined in General Ledger, but this model is not adequate with the specification of transactions included in the Act;
- data warehouse model must include heterogeneous dimension of Bank's General Ledger and allow to identify each single analysed object;

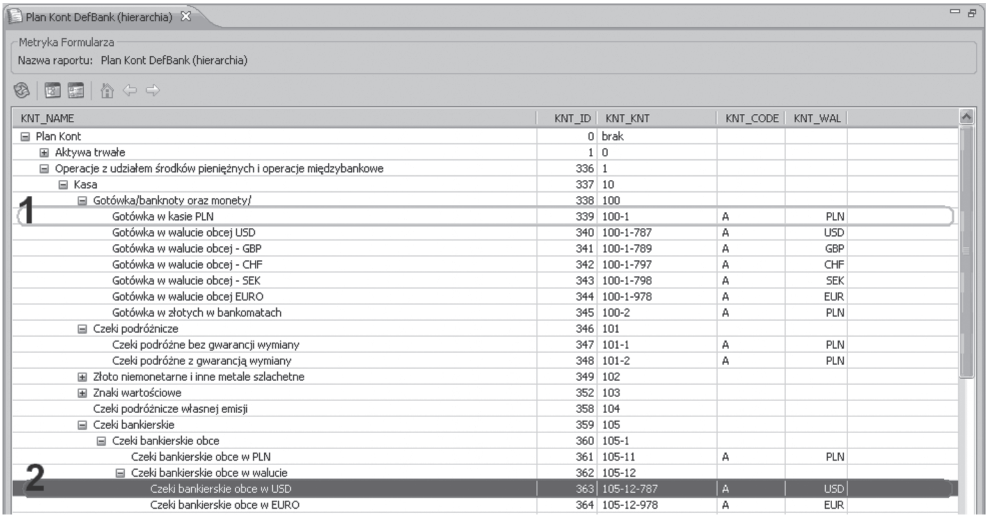
- OLAP cubes must aggregate data including heterogeneous dimension of General Ledger; it means that granularity of aggregation should be defined on each of the dimension levels, and each of the Cartesian product defined by the OLAP cube measurements should be able to be identified;
- analyses included in SART must take into consideration all operations carried in bank's accounting system;
- system must take into account volume of data, up to 400 thousand records relating to bank operations (accounting decrees and ELIXIR announcements) can be processed on daily basis;
- system must be ready to implement new functionalities included in the Draft Act, without the necessity to introduce important changes in data warehouse model.

Undoubtedly the biggest challenge that SART system had to face was the definition of data warehouse model and OLAP cubes that would take into account heterogeneous dimension of General Ledger. It is worthwhile mentioning that presently used data warehouse models do not take into account heterogeneous dimensions in their specification. The example of differences between homogeneous and heterogeneous dimensions is illustrated in Figures 3A and 3B.



DT_DDATE	DT_ID	DT_NTCL
Wymiar czasu	0	Wymiar czasu
2007	1	year
2007-Q1	2	quarter
2007-Q1- 01	3	month
2007-Q1- 02	35	month
2007-Q1- 03	64	month
2007-Q2	96	quarter
2007-Q3	191	quarter
2007-Q4	287	quarter
2008	383	year
2008-Q1	384	quarter
2008-Q1- 01	385	month
2008-Q1- 01- 01	386	day
2008-Q1- 01- 02	387	day
2008-Q1- 01- 03	388	day
2008-Q1- 01- 04	389	day
2008-Q1- 01- 05	390	day

Figure 3A. A homogeneous dimension of data warehouse (consistent with the classical model for data warehouse) where each class of dimension: Time, Year, Quarter, Month, Day has a homogeneous structure of the hierarchy



Plan Kont DefBank (hierarchia)

Metryka Formularza
Nazwa raportu: Plan Kont DefBank (hierarchia)

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		
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 podróżne bez gwarancji wymiany	347	101-1	A	PLN
Czeki podróżne 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
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

Figure 3B. Heterogeneous data warehouse dimensions where each class of dimension: General Ledger (PK – in Polish Plan Kont), Synthetical Account (KS – in Polish Konto Syntetyczne), Analytical Account (KA – in Polish Konto Analityczne) may have a heterogeneous structure of the hierarchy

In Figure 3B the heterogeneity can be recognised in two encircled dimensions that differ in terms of the structure of the hierarchy. For example, Figure 3B marked heterogeneity of the hierarchy:

Green colour (marked as 1):

- PK "General Ledger",
- KS "operations involving cash and interbank operations",
- KS "Kasa",
- KS "Cash (banknotes and coins)",
- KA "Cash in hand GBP".

Red colour (marked as 2):

- PK "Chart of Account",
- KS "operations involving cash and interbank operations",
- KS "Kasa",
- KS "bankers' cheques",
- KS "foreign bankers' cheques",
- KS "bankers' cheques, foreign currency", KA "bankers' cheques in USD"

Hence, we can see that both hierarchical structures:

- [PK, KS, KS, KS, KA] for a KA "Cash in hand GBP (ID: 339 in column KNT_ID);
- [PK, KS, KS, KS, KS, KS, KA] for a KA "foreign bankers' cheques in U.S. dollars" (ID: 363 in column KNT_ID)

are different from each other in terms of the structure of the hierarchy.

Another issue that had to be included in SART concerned the requirements to identify all analysed objects (including warehouse measurements and Cartesian products of OLAP cubes). Technically speaking, it is rather OLAP implementation problem that is currently solved in two ways: by description of data included in data warehouse and OLAP *via* metadata (e.g. XML) or by direct identification of objects included in data warehouse and OLAP. The problem can be described using evaluation of OLAP cube as an example. On the basis of information obtained from bank's production systems, the following basic measurements of data warehouse dimensions have been assumed: bank's General Ledger 60 000 entries, bank customers 500 000 entries, the duration of operations 3600 entries (10 years), number of measures in OLAP cube is 5.

Approximate calculation of OLAP cube's[5] size indicated that it would not be possible to apply a solution that would store OLAP data without compression – approximate size of OLAP cube of 230.2 PB exceeds considerably size of disc matrixes available on the market and which could be applied (up to 10 TB) considering project's budget.

Another specific task that must have been completed in SART system was the delivery of functionality that would allow to merge the decrees into bank. This process was to enable transformation of data saved as accounting decrees (list of endorsement operations along with DEBIT and CREDIT) [Figure 7A] into a form of bank transaction (source and destination accounts, and transaction amount) [Figure 7B]. It is worthwhile emphasising that such functionality should integrate information coming from General Ledger dimension (identifier of source and destination account of bank transaction) in order to maintain the consistency of data between bank transactions table and facts table (decrees) of data warehouse.

SART Data Warehouse

In case described here, data warehouse has been supplied by three different IT systems that make six sources of data available. From bank's General Ledger the following data are acquired: data relating to the bank's customers, bank's General Ledger, NRB accounts and Documents' Journal reports (accounting operations endorsement). ELIXIR system announcements are acquired from the system of financial instruments settlement made between the participants of financial market ELIXIR® and completed by the bank. WEB Service of Polish National Bank (www.nbp.gov.pl) provides central bank exchange rates that are then loaded. Additionally in data warehouse there is Time dimension implemented in the form of hierarchy including the following time sub-dimensions: Year, Quarter, Month, Day.

Data coming from the individual sources are loaded into SART system *via* import tables and then integrated into one coherent repository of data that includes data warehouse measurements (clients, General Ledger, exchange rate), facts table

---

[5] Approximate number of entries in OLAP cube was: $60\,000 \times 500\,000 \times 3600 \times 5 = 0.54*10^{15}$. Considering the minimum size of data stored in OLAP cube (4 bytes dimension identifier, 8 bytes measure's value) this value should increase by $3\times4\times5\times8 = 480$ times that is $259.2*10^{15}$ data bytes (23 5741TB, 230.2 PB). Above presented calculations do not take into account the size of OLAP cube data table indexes.

(decrees) and auxiliary tables (NRB numbers, ELIXIR announcements). It is important to notice that model of data loading is completed in accordance with ELT convention (Extract-Load-Transform), thus allowing for considerable flexibility of system in the aspect of data source identification and full control over the process of data warehouse loading.

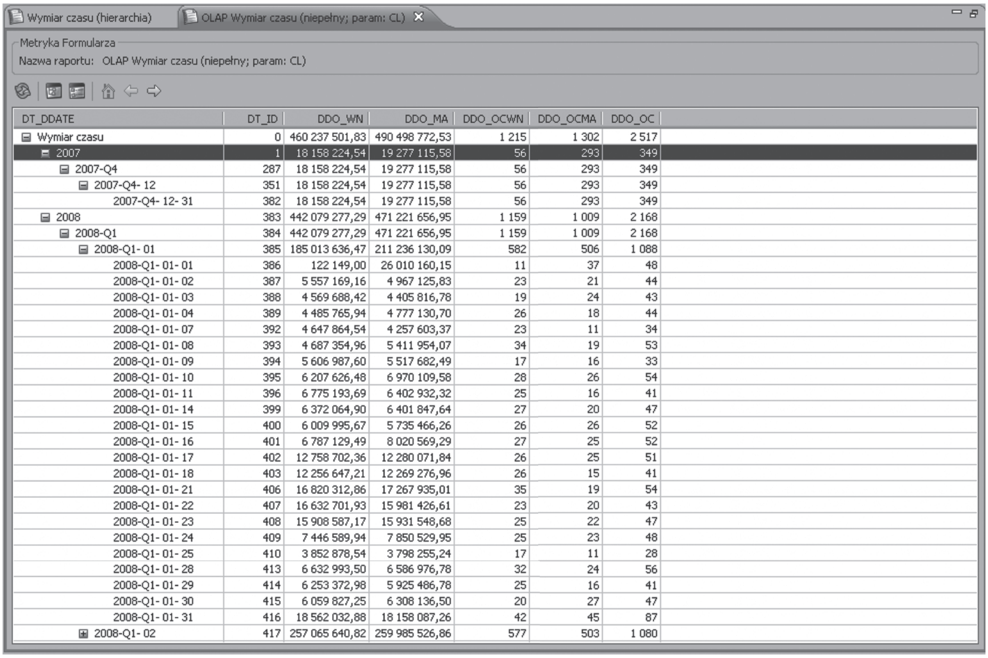
OLAP support in SART system

The solution that has been implemented links the data warehouse scheme with multidimensional OLAP cube with the following characteristics:

– OLAP cube model: partial, multidimensional, incremental OLAP ASQL cube;
– OLAP cube's dimensions: time dimension, clients, General Ledger;
– 5 OLAP cube's measures.

Partial, multidimensional, incremental OLAP ASQL cube defines model of OLAP cube in which only ensuing facts from the facts table (incomplete) along with all overriding sub-dimensions of given dimension (multidimensional) are calculated in incremental manner (only new facts are calculated). The concept of "multidimensional OLAP cube" in this project requires additional comment: the model of OLAP cube in ASQL server stores aggregated data for the Cartesian products of all sub-dimensions of OLAP cube (Figures 4 and 5).

To estimate the system performance for data import containing 57 952 operations (decrees) the following measures have been considered: computing time of

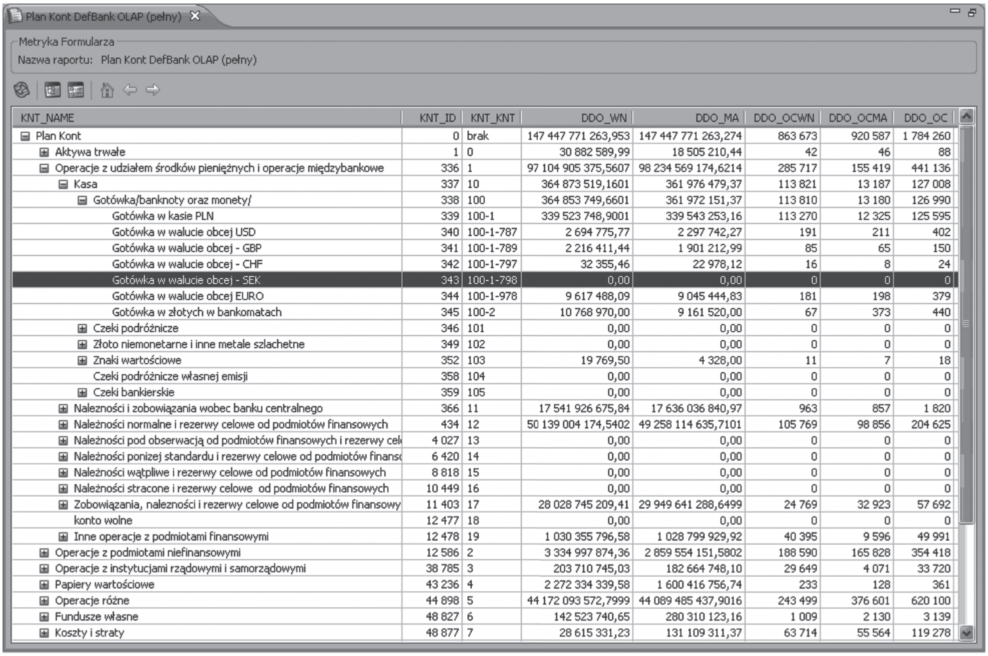


Wymiar czasu (hierarchia) | OLAP Wymiar czasu (niepełny; param: CL)

Metryka Formularza
Nazwa raportu: OLAP Wymiar czasu (niepełny; param: CL)

DT_DDATE	DT_ID	DDO_WN	DDO_MA	DDO_OCWN	DDO_OCMA	DDO_OC
Wymiar czasu	0	460 237 501,83	490 498 772,53	1 215	1 302	2 517
2007	1	18 158 224,54	19 277 115,58	56	293	349
2007-Q4	287	18 158 224,54	19 277 115,58	56	293	349
2007-Q4- 12	351	18 158 224,54	19 277 115,58	56	293	349
2007-Q4- 12- 31	382	18 158 224,54	19 277 115,58	56	293	349
2008	383	442 079 277,29	471 221 656,95	1 159	1 009	2 168
2008-Q1	384	442 079 277,29	471 221 656,95	1 159	1 009	2 168
2008-Q1- 01	385	185 013 636,47	211 236 130,09	582	506	1 088
2008-Q1- 01- 01	386	122 149,00	26 010 160,15	11	37	48
2008-Q1- 01- 02	387	5 557 169,16	4 967 125,83	23	21	44
2008-Q1- 01- 03	388	4 569 688,42	4 405 816,78	19	24	43
2008-Q1- 01- 04	389	4 485 765,94	4 777 130,70	26	18	44
2008-Q1- 01- 07	392	4 647 864,54	4 257 603,37	23	11	34
2008-Q1- 01- 08	393	4 687 354,96	5 411 954,07	34	19	53
2008-Q1- 01- 09	394	5 606 987,60	5 517 682,49	17	16	33
2008-Q1- 01- 10	395	6 207 626,48	6 970 109,58	28	26	54
2008-Q1- 01- 11	396	6 775 193,69	6 402 932,32	25	16	41
2008-Q1- 01- 14	399	6 372 064,90	6 401 847,64	27	20	47
2008-Q1- 01- 15	400	6 009 995,67	5 735 466,26	26	26	52
2008-Q1- 01- 16	401	6 787 129,49	8 020 569,29	27	25	52
2008-Q1- 01- 17	402	12 758 702,36	12 280 071,84	26	25	51
2008-Q1- 01- 18	403	12 256 647,21	12 269 276,96	26	15	41
2008-Q1- 01- 21	406	16 820 312,86	17 267 935,01	35	19	54
2008-Q1- 01- 22	407	16 632 701,93	15 981 426,61	23	20	43
2008-Q1- 01- 23	408	15 908 587,17	15 931 548,68	25	22	47
2008-Q1- 01- 24	409	7 446 589,94	7 850 529,95	25	23	48
2008-Q1- 01- 25	410	3 852 878,54	3 798 255,24	17	11	28
2008-Q1- 01- 28	413	6 632 993,50	6 586 976,78	32	24	56
2008-Q1- 01- 29	414	6 253 372,98	5 925 486,78	25	16	41
2008-Q1- 01- 30	415	6 059 827,25	6 308 136,50	20	27	47
2008-Q1- 01- 31	416	18 562 032,88	18 158 087,26	42	45	87
2008-Q1- 02	417	257 065 640,82	259 985 526,86	577	503	1 080

Figure 4. Example of OLAP report for homogenous dimension of Time based on incomplete dice OLAP



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
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
Czeki podróżnicze	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
Czeki podróżnicze własnej emisji	358	104	0,00	0,00	0	0	0
Czeki bankierskie	359	105	0,00	0,00	0	0	0
Nalezności 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 ponizej 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, naleznośći 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

Figure 5. Example of OLAP report with a view of General Ledger as heterogeneous dimension with enabled filling of partial OLAP cube (marked row)

OLAP table (one process) takes 250 seconds; number of OLAP operations carried out: 5 275 770; and number of lines created in OLAP cube: 991 675, average number of calculated decrees in facts table was 231.8 decrees/second, average number of OLAP operation was 21 103.1 operations/second and average number of calculated Cartesian products per decree was 91. According to the functional requirements the system performances were considered acceptable.

Transaction merging in SART

Module of transaction merging in SART delivers functionality that includes transformation of data from accounting system (decrees of bank's General Ledger) (Figure 6) into bank transactions data model . In the implemented solution the process of transactions merging uses table of dimensions of data warehouse General Ledger as a dictionary table for identifying accounts of transactions parties. Figure 6 illustrates the definitions of accounts: source (ID: 11 625) and target (ID: 460), that correspond to the accounts in Figure 7 that shows the connections between data warehouse objects and OLAP application layer of system SART.

Summary

In this section we have discussed only crucial aspects of implementation that required non-standard processing in the context of currently used models and solutions for OLAP data warehouse. Special attention was paid to the fact of techni-

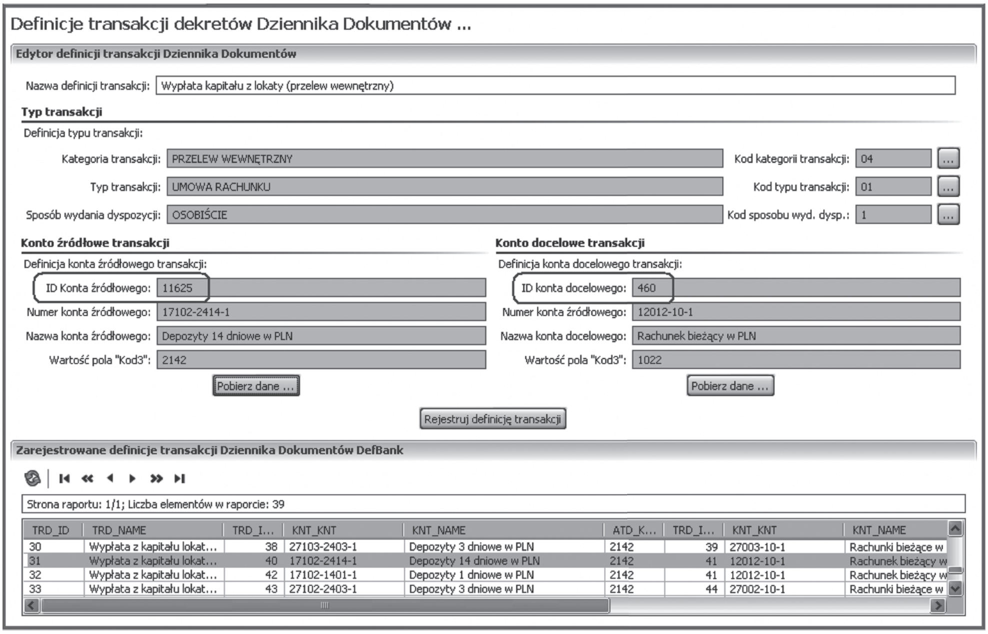


Figure 6. An example of transaction definition

cal and logical marketability of OLAP cube. Taking into account logical aspect we should point out the fact of implementing OLAP cube that deals with heterogeneous dimension of General Ledger which undoubtedly is important. Structures of such type have not yet been supported in currently available models of OLAP data warehouses. Functionality of incomplete OLAP AQL cubes coincides completely with functionality of complete OLAP cubes. Assuming the size of designed OLAP cube to be 230 PB the raid storage device of 2 TB has been installed.

Considering functional aspect of data warehouse, it was feasible to build a repository of data that fully illustrates structure of General Ledger's decrees. It should be noted that it is essential in data analysis because the primal categorisation of bank operations basing on General Ledger has been maintained. Therefore it was possible to obtain on the one hand automatic categorisation of bank transactions, and on the other – manual aggregation of data in accordance with General Ledger. It is important to emphasise the fact of integrating two different logical models (accounting and transactional) into one coherent application model.

Another essential aspect of this issue concerned the full exploitation of the functionality of ROLAP Analytic Server SQL. In the implemented solution that functionality has been used during definition process of bank transactions in respect to General Ledger data. In general, model of OLAP relational data warehouse included in Analytical Server SQL enables to use tables of dimensions as tables of relational database in the application layer. Thus, dimension of General Ledger data warehouse has been used as a dictionary table during defining transactions in the entries concerning source and target accounts. This let us to obtain full integration between analytical layer (data warehouse and OLAP) and application layer in SART system. As a consequence of analysing bank transactions we can easily move to OLAP analyses of General Ledger's decrees and back again. Interaction between those two logical models of data is essential during verification of analysed areas, for example: does number of cash transactions selected for reporting contain all operations carried out by the bank including appropriate limitations (it can be checked in the OLAP cube through comparing number of appropriate decrees).

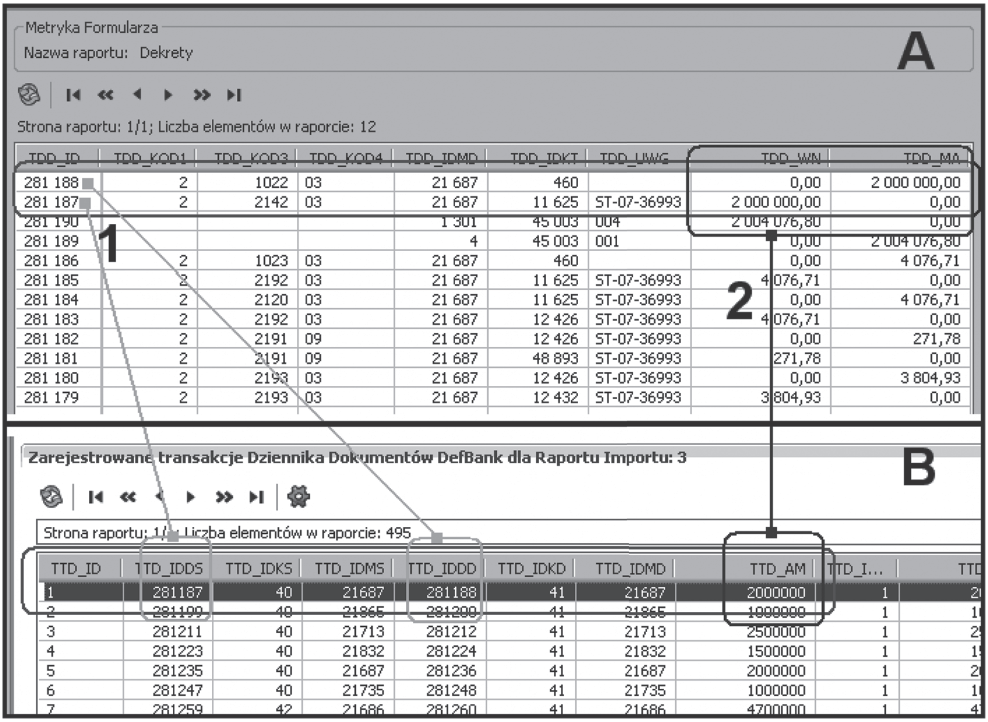


Figure 7. An example describing integration of the two models in SART system

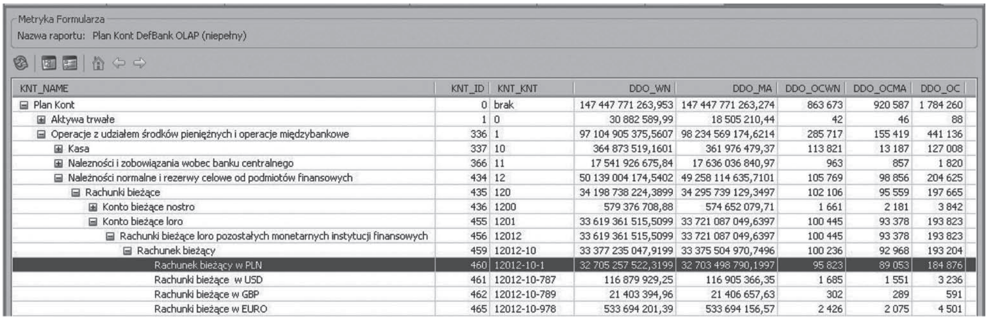
Figure 7 shows an example of integration of two different data models:

A. *Accounting model* that concerns individual operations on the positions CR and DB;

B. *Transaction model*, in which data record contains information about accounts (number 1 in Figure 7): source (ID decree: 281 187), target (ID decree: 281 188) and the amount of operation (number 2 on Figure 7).

Merging two different models is possible using OLAP object identification function and General Ledger as a dictionary in the application layer.

Figure 8 indicates the position of the aggregated turnover on the account "Current account in PLN" (account ID: 460), which was listed in Figure 7A (the first item, account ID column "TDD_IDKT"). Thanks to applying object identification function, one can easily move from analysis of banking transactions (Figure 8B) to OLAP analysis of decrees based on General Ledger.


Metryka Formularza
Nazwa raportu: Plan Kont DefBank 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
Nalezności 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
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	32 703 498 790,1997	95 823	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

Figure 8. Extract of OLAP report in the context of the General Ledger

As a result of applying functions described in this article, system SART is able to analyse and discover money laundering transactions, notably:
- through selection operation; due to extraction criteria indicating transactions of above 15 000 euro;
- through aggregation operation, enabling classical OLAP analysis;
- through clustering operation, that allows grouping predefined objects;
- through analysis of chains of transactions related to a particular chain of transactions.

6. Conclusions and future works

System presented in this paper is a complete and integrated analytical suite comprising of Analytical SQL Server (SQL database server with analytical extensions) and SART application (database and client layer). SART is one of scalable and extensible systems. In standard configuration SART system does not need any supplementary components, application licenses or dedicated server licenses (e.g. ORACLE, IBM DB2 or MS SQL Server or extra Windows server licenses etc.). It should also be noted that SART has been designed in accordance with guidelines on banking analysis determined in the EU Capital Requirements Directive (CRD) or BASEL II Accord. Hence, data warehouse and suite of data registers in SART are oriented towards advanced banking analysis and management of market, credit analysis and evaluation of operational risk.

Further research and development of the Analytical SQL Server will be carried out, paying particular attention to its use in the area of AML using the SART system. Focusing on the most important issues, the works will concern the following topics:

- standardisation of the data warehouse model OLAP with regard to compliance with the classical model of OLAP;
- standardisation of the ETL process with the SART model (ELT) and the application layer (the combination of Java and Eclipse, the script language Python);
- development of Business Intelligence and Analytical Intelligence extensions, in particular in the context of the analysis related to methodology Know Your Customer – KYC;
- analysis of mapping Object/Relational model in SART system;
- application of solutions built-in SART to design financial monitoring based on ASQL, General Ledger, including controlling, planning and budgeting;
- research and evaluation of selected data mining algorithms, in particular clustering, pattern recognition and interpretation of results.

To summarize, the presented technological approach offers new possibilities to build dynamic, adaptive and efficient systems capable to discover money laundering transactions.

References

Act of 16th November, 2000 on preventing introducing into financial turnover properties derived from illegal or undisclosed sources and on counter-financing of terrorism, DzU 2003 Nr 153, poz. 1505, with changes at a later date.

Alexander R.C.H. (2007), *Insider Dealing and Money Laundering in the EU: Law and Regulation*, Ashgate Publ., Aldershot.

Directive 2005/60/EC of the European Parliament and the Council of October 26, 2005. Commission Directive 2006/70/EC of 1st August 2006 laying down implementing measures for Directive 2005/60/EC of the European Parliament and the Council, Euro-Lex http://eur-lex.europa.eu.

Draft of 26th May, 2008 of the Act on altering the act on preventing introducing into financial turnover properties derived from illegal or undisclosed sources and counter-financing of terrorism and on altering several other acts, BIP Web site http://www.mf.gov.pl.

Financial Crimes Enforcement Network (FinCEN), http://www.fincen.gov – official Web site of the US Department of Treasury.

Inmon W.H. (2005), *Building the Data Warehouse*, Wiley, New York.

Kingdon J. (2004), AI Fights Money Laundering, *IEEE Intelligent Systems*, May/June, pp.87-89.

Muller W.H., Kalin C.H., Goldsworth J.G. (Eds.) (2007), *Anti-Money Laundering: International Law and Practice*, Wiley, New York.

Philippsohn S. (2001), Money Laundering on the Internet, *Computers & Security*, No. 20, pp. 485--490.

Regulation of the Minister of Finance on determining the model of transaction register, the mode of keeping and method of delivering data from register to the General Inspector of Financial Information, Warsaw, 21/09/2001.

Schott P.A. (2006), *Reference Guide to Anti-money Laundering and Combating the Financing of Terrorism*, World Bank Publ.

The Forty Recommendations (2004), Report Financial Action Task Force on Money Laundering (FATF-GAFI), http://www.fatf-gafi.org/

Turban E. et al. (2007), *Business Intelligence*, Prentice Hall, New Jersey.

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

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

Zdanowicz J.S. (2004), Detecting money laundering and terrorist financing via data mining, *Comm. of the ACM*, Vol. 47, No. 5, pp. 53-55.