STORING AND PROCESSING ACCOUNTING INFORMATION BASED ON COLLECT/REPORT PARADIGM

Krassimira Ivanova

***Abstract**: The possibility to store and process accounting information using the Collect/Report Paradigm (CRP) is outlined in this paper. The main idea consists of using CRP to distribute accounting information in multi-dimensional information spaces and stored and processed it in parallel in cloud. Every account may be presented by a separated layer which contains two named sets – Dr and Cr. Storing and reporting may be provided simultaneously without recompilation of the information base.*

***Keywords**: Accounting, Collect/Report Paradigm, Big Data, Cloud computing, BigArM.*

***ACM Keywords:** E.1 Data Structures; Distributed data structures.*

Introduction

The milestone of Collect/Report Paradigm (CRP) [Markov & Ivanova, 2015] is the simple idea that we may use a special kind of organization of the information and this way to develop easy to use information bases and with very high speed for response which enables ***the real-time analytical processing*** (RTAP) [Markov, 2005]. (The RTAP multithreaded processing engine needs to support extremely large volumes of data in real time. The analytics performed are composed of combinations of algorithmic, statistical and logical functions [B-Jensen, 2002]).

RTAP is convenient approach for information service of business activities. Many business concerns have several hundred or even several thousand *business transactions* each day. This is reason to investigate using of Collect/Report Paradigm for storing and processing business information.

In this paper we will discuss the theoretical and practical aspects of implementing CRP in real accounting information service systems. Firstly we will remember the mathematical structures (Named Sets). Then we will outline main characteristics of structures for storing business information and theirs possibilities to be used in the frame of CRP.

Named sets

The concept “Named set” was defined by Mark Burgin. Here we will follow [Burgin, 2011].

Named set **X** is a triple **X** = (*X*, *μ*, *I*) where:

X is the support of X and is denoted by S(X);

I is the component of names (also called set of names or reflector) of X and is denoted by N(X);

μ: X → I is the naming map or naming correspondence (also called reflection) of the named set X and is denoted by n(X).

The most popular type of named sets is a named set **X** = (*X*, *μ*, *I*) in which *X* and *I* are sets and *μ* consists of connections between their elements. When these connections are set theoretical, i.e., each connection is represented by a pair (*x*, *a*) where *x* is an element from *X* and *a* is its name from *I*, we have a *set theoretical named set*, which is binary relation.

Named sets as special cases include:

Usual sets;

Fuzzy sets;

Multisets;

Enumerations;

Sequences (countable as well as uncountable)

etc.

A lot of examples of named sets we may find in linguistics studying semantical aspects that are connected with applying different elements of language (words, phrases, texts) to their meaning [Burgin & Gladun, 1989; Burgin, 2011].

Accounting cycle

Accounting documents contain a reflection of the financial activities of all proprietary information and are a good basis for analysis, and hence - forecasting and planning the future activities of the company [Samuelson & Nordhaus, 1989].

The sequence of accounting procedures used to record, classify, and summarize accounting information is often termed *accounting cycle* [Markov et al, 1993b; Meigs & Meigs, 1989].

The accounting cycle begins with the initial recording of business transactions and concludes with the preparation of formal financial statements summarizing the effects of these transactions upon the assets, liabilities, and owners' equity of the business. The term "cycle" indicates that these procedures must be repeated continuously to enable the business to prepare new, up-to-date financial statements at reasonable intervals.

Many business concerns have several hundred or even several thousand *business transactions* each day. It would not be practicable to prepare a balance sheet after each transaction and it is quite unnecessary to do so. Instead, the many individual transactions are recorded in the accounting records and, at the end of the month or other accounting period, a balance sheet is prepared from these records.

An accounting system includes a separate record for each item that appears in the balance sheet. For example, a separate record is kept for the asset cash, showing all the increases and decreases in cash which result from the many transactions in which cash is received or paid. A similar record is kept for every other asset, for every liability, and for owners' equity.

The form of record used to record increases and decreases in a single balance sheet item is called *an account.* The account is a means of accumulating in one place all the information, about changes in a specific asset, a liability, or owners' equity.

All separate accounts are usually kept in a loose-leaf binder, and the entire group of accounts is called a *ledger.* The electronic variant of the ledger is assumed in this paper.

Structure of the ledger account

➢ Basic types of information items of the transaction

In every transaction there are five main types of information items:

- Integer number;
- Real number;
- String;
- Date;
- Nomenclature.

The first four types are identical with those in programming languages like Pascal. The fifth, the nomenclature Nmcl, is specific named set:

$$Nmcl: (Nom\_code, \mu, Value),$$

where

- Nom_code is a set of nomenclature codes which usually are integer numbers or strings;
- μ - is biunique correspondence between Nom_code and Value;
- Value is a set of any of:
  - Integers;
  - Real numbers;
  - Strings;
  - Dates;
  - Complex records built of these types.

- **The ledger account and transactions**

The ledger account ACC is a specific named set [Markov et al, 1994] of two elements which are named Debit and Credit and which are sets Dr and Cr; i.e.

ACC : ({Debit, Credit}, μ, {Dr, Cr}).

Every element of Dr and Cr is named transaction and is connected with corresponded time point (usually this is any day of the year). This way Dr and Cr are sorted by the time of the transactions.

Every transaction may be described by the relation:

Tr : (<Doc No><Date><Type><Summ><Analitic information>)

where:

- <Doc No> is number of the primary document;
- <Date> is date of the primary document;
- <Type> is type of the primary document;
- <Sum> is the amount of the primary document;
- <Analytic information> is a special record of information items for additional analysis, sorting and searching the transaction record.

- **Correspondence between analytical information items**

The analytic information is the main tool for accounting analysis and preparing secondary documents for the leaders and the managers of the company and theirs decision making. Several additional items are usually connected with every account transaction. This information is needed for more complex analysis of the account or of the ledger as a whole. These items are known as an "analytical features" of the account.

There are three main goals of using the analytical features:

- To analyze the set of the account transactions in Dr or Cr for finding any mistakes or for obtaining any internal correspondences between transactions of the Dr or Cr of the ledger account;
- To analyze together the sets of the account transactions in Dr and Cr for finding any mistakes or for obtaining any internal correspondences between transactions of the ledger account which are included in sets of Dr and Cr;
- To analyze two or more accounts together for obtaining any correspondences between theirs transactions.

- **Time characteristics of the ledger account**

As we said in [Burgin & Gladun, 1989] the time dimension of the account transactions is very important one for the business analysis. The time-period principle is one of the generally accepted accounting principles that guide the interpretation of financial events and the preparation of financial statements. It

tells us that the life of a business entity must be divided for accounting purposes into time periods of equal length, so that decision-makers will be informed on current trends within business [Markov et al, 1993b; Meigs & Meigs, 1989].

Time is recorded when:

- The primary document is created;
- The account transaction correspond to the given primary document is entered in the account.

The business analysis may be made:

- For given date;
- For time interval, including or excluding the boundaries of the interval.

➢ **The account balance**

There are several accounting intervals which are used for analyze the set of transactions of the account. Usually these intervals are:

- A day;
- A month;
- A quarter;
- An interval of days;
- An interval of mounts;
- An interval of quarters;
- A year.

The account balance is prepared for every accounting interval which is important for the business. The account balance is amount between total sums of Dr decremented with total sum of Cr. The total sum of Dr or Cr is made only for the transactions belong to the given accounting time interval.

➢ **Subaccounts and batches**

Usually the account contains hundreds of transactions. It is very difficult to analyze so many transactions and prepare report for any analytic condition. Because of this the account may be divided into sub accounts. In this case the account is named set of subaccounts.

There is another kind of grouping the transactions of the account. Using any analytical feature we can group transactions in batches corresponded to every value of given analytical sign. The reports in this case are made for the account sorted and "virtually" divided in batches in accordance with given condition for one or more analytical signs.

➢ **The Ledger**

The ledger LGR is a named set which elements are accounts, i.e.

$$LGR : (NACC, \mu, SACC),$$

where

- NACC is set of ledger accounts' names;
- μ - is biunique correspondence between NACC and SACC;
- SACC is set of ledger accounts.

Names of accounts are strings of digits. It is possible to use any letters but this form is not preferred in the everyday practice.

All accounts in the ledger describe all financial depended activities of the business. Every primary document must be registered in at least of two accounts. This way every account corresponds to some others - there are mappings between accounts. These mappings are very important for business analysis.

The specific of the mappings between accounts is the principle "the domain of mapping always is Dr of an account and co-domain is Cr of the same or other account". Mappings are registered both in corresponded transactions of domain and co-domain. This is an explicit registration of the relationship.

One account may correspond to every other one, but in practice there are clear rules for building accounting correspondence.

So, all financial business activities may be and must be reflected in the ledger by corresponded transactions and relationships between them.

Information operations with the ledger account

The main information operations with the ledger accounts are**:**

- Creating;
- Analyzing.

There are several service information operations - renaming, moving from one place of the ledger to another, deleting, sorting and preparing the batches and etc.

➢ Creating the ledger account

Creating the ledger account is a two steps' process. The first step is defining of the account and establishing its relations with balance sheet and income statement. The second is everyday entering new transactions in the Dr and Cr of the account and preparing work sheet and other financial statements.

➢ Defining the ledger account

The main procedure for the account is its defining. This is sequence of several decisions and activities:

- Naming the account and including it in the hierarchy of the ledger;
- Initializing of the control fields of the account;
- Set up the analytical parameters of the account and connection it with appropriate nomenclatures.

Naming the account is depended with national and international accounting standards. Usually it is clear what name is needed and where in the hierarchy is it.

More complicated is the initializing of the control fields of the account. There are two main types control fields:

- Relations with balance sheet and income statement;
- Initial values in Dr and Cr [Markov et al, 1993b] for the beginning of year and for all preceding months of current year if the account is initialized in the middle of the year.

Setting up the analytical parameters of the account and connection it with appropriate nomenclatures is very important for future analysis of the account and preparing reports for requests which in the stage of initializing could not be expected.

- **Memorial order**

The everyday accounting is based on two main technologies:

- Memorial orders oriented;
- Accounting journals oriented.

In computer based accounting the journals oriented technology is used when accounting cycle contains little number of transactions. In this case a spreadsheet may be used for information service of accounting cycle. So, all correspondences between accounts are made manually and analysis is very difficult.

Memorial orders oriented technology is simpler for working and has great power for computer analysis and preparing the secondary information - all standard financial statements and non standard but very useful reports. Our information model is connected with memorial orders' technology.

Memorial order (MO) contains header and body.

Header is standard information which describes the primary document:

- Type;
- Number;
- Date;
- Total sum;
- Number and name of business partner;
- Type of business operation;

- Income tax if it is included in the sum of the contract; etc.

Body of the MO is specific table used for entering the follow information about transactions:

- Description of the transaction;
- Corresponded accounts;
- Analytical information for each account of the transaction;
- Sum of the transaction.

➢ Information analysis of the MO

The information analysis of the MO is aimed:

- To make preliminary work sheet in the frame of the MO and to find and to correct any mistakes;
- To prepare two main journals of purchases and sales;
- To compute the income taxes;
- To make information retrieval and to prepare non-standard reports based on the MO description of the primary documents.

It is possible to classify MO in three sets:

- MO for purchases;
- MO for sales;
- MO for other business operations like money transfer between bank account and cash of the company and vice versa.

The information retrieval is needed for preparing reports based on primary documents. The main features of the information retrieval are described below.

➢ Entering new transactions in the Dr or Cr of the accounts

One row of the body of the MO describes one transaction. One primary document may be described by one or more transactions and by one or more MO.

After analysis of the MO all its transactions may be entered in the ledger accounts. As a rule, one transaction is entered in two corresponded accounts. The main rule in this case is: the transaction is entered in Dr of the first account and in Cr of the second. Which account is the first and which is the second is shown in the MO.

✓ Analysis of Dr or Cr

The main goal of accounting is to distribute transactions in accounts for more convenient analysis of the financial information and this way to support decisions making and the control in the business.

✓ Making the work sheet

After every entering the transactions the content of the Dr and Cr of the accounts is changed. The analysis of the current content of a given account is the first step. Usual it is to prepare the total of all sums of the transactions belongs to Dr or Cr respectively.

The totals are very useful for analysis of the dynamic of changes in the accounts. This analysis is based on:

- The total sums from beginning of the given accounting period so cold trial balance (TB);
- The total sum of the transaction belong to the given accounting period or so cold adjustments (ADJ);
- The final result of the getting together of the trial balance and adjustments or so cold adjusted trial balance (ATB).

These three elements (TB, ADJ and ATB) are used for preparing the work sheet of the ledger accounts.

As a rule, the heading of the work sheet consist of three parts:

- The name of the business;
- The title Work Sheet;
- The period of time covered.

In the American and European accounting the body of work sheet contains five pairs of money columns. In Bulgarian accounting the body of the work sheet contains three pairs of money columns (for TB, ADJ and ATB), each pair consisting of a debit and a credit column. The rest two pairs are represented in separate income statement and balance sheet.

Making the work sheet is a standard procedure and may be done automatically.

- ✓ **Making the journals**

The journals are another form to show the relations between ledger accounts. The journal is a table which visualizes the mapping between given account and all others in the ledger. For every account there are two main mappings - for Dr and Cr respectively.

Making the journals is a standard procedure too and may be done automatically.

- ✓ **Preparing the batches**

Batch is a subset of the Dr or Cr which corresponds to given condition. The process of preparing the batches is based on additional analytical information in the Dr or Cr of the account. So, there are many variants of grouping the transactions of Dr or Cr in batches. This is main reason for the importance of this information operation for the accounting cycle - it permits the multi lateral analysis in the single account.

Preparing the batches is a standard procedure too and may be done automatically.

✓ **Complex requests and reports**

The Dr and Cr may be represented as relations in the Codd relation model. In this case it is possible to use relation algebra or relation calculus for extracting any subsets from Dr or Cr. The resulting subsets may be used for preparing the batches which will contain transactions only from the report of the given relational request. This way number of the elements of the set of batches for analyze may be limited.

Executing the complex requests and preparing the reports and batches based on them in the practice may be done only automatically.

➢ **Common analysis of Dr and Cr**

After separate analysis of Dr and Cr it is important to continue with the common view of the two parts of the ledger account. This common view permits to find relationships between elements of Dr and Cr. Only common analysis of Dr and Cr gives total map of the financial operations. It is possible to analyze both Dr and Cr of one account or of the different accounts. Common analysis of Dr's or Cr's of any accounts has little practical effect.

✓ **Grouping the transactions and establishing the correspondences between groups**

The main work in common analysis of the Dr and Cr is to reorganize the account in new subsets and establish any correspondences between these subsets. All analytical signs may be used as basis for grouping the transactions of Dr and Cr. After finishing grouping it is possible to integrate transactions in every group to make more general view on the financial operations and processes.

✓ **Requests and reports**

As a rule, accounts contain thousands of transactions. In this case grouping is very slow process and in many cases it is impossible to work with so much information. For this goal it is need to have special tool for selecting the appropriate transactions before starting the grouping process. The requests and reports of this kind are similar to the relation algebra or relation calculus but are executed on different relations (Dr and Cr) and the result is again different subsets of Dr and Cr.

Storing and processing accounting information based on CRP

We have used strong hierarchies of named sets to create a specialized mathematical model, for new kind of organization of information bases called Multi-domain Information Model" [Markov, 2004]. The "Information Spaces" defined in the model are kind of strong hierarchies of enumerations (named sets) [Ivanova, 2015]. This permits it to be implemented for storing and processing of accounting information. Let remember that the ledger accounts are kind of named sets.

In the same time, the RDF structures are kind of named sets. This means that the accounting information can be directly represented by RDF triples or quadruples and stored and processed in the frame of CRP.

The advantage of this is the possibility to pass the boundaries of limitations of classical systems and to have freedom of cloud processing.

Conclusion

In the beginning of this paper we have remembered the main structure of the Collect/Report Paradigm. Further we included a short presentation of named sets. After that, the structure and operations with accounting information have been presented in details. In the great companies it may be very large and memory for storing and time processing pass the limitations of classical data bases.

Finally, we outlined the possibility to store and process accounting information using the Collect/Report Paradigm. Following it, the accounting information may be distributed in the cloud multidimensional information spaces and stored and processed in parallel. Every account is presented by a separated layer which contains two named sets – Dr and Cr. Storing and reporting may be provided simultaneously without recompilation of the information base.

More concretely, the advantages of such approach are:

Collecting accounting information may be done for all ledger accounts independently in parallel;

Reporting accounting information may be provided only by the ledger accounts which really contain information related to the request;

Input data as well as results may be in RDF-triple or RDF-quadruple format.

Bibliography

[B-Jensen, 2002] M.T. B-Jensen, “High Tower Software's Tower View is the Odds-On Favorite of International Game Technology for Real-Time Data Management”, Product Review published in DM Review Magazine July 2002 Issue. http://www.dmreview.com/article_sub.cfm?articleId=5403

[Burgin & Gladun, 1989] M. Burgin, V. Gladun, “Mathematical Foundations of Semantic Networks Theory”, Springer verlag: Lecture Notes in Computer Science, 1989, v.364, pp. 117 - 135.

[Burgin, 2011] Mark Burgin, “Theory of Named Sets”, Nova Science Publishers Inc (United States), 2011, ISBN-13: 9781611227888, 681 p.

[Ivanova, 2015] Krassimira Ivanova, “A Survey of Mathematical and Informational Foundations of the BigArM Access Method”, International Journal "Information Technologies & Knowledge" Vol. 9, Number 3, 2015, ISSN 1313-0455 (printed), 1313-048X (online), pp. 272 – 299.

[Markov & Ivanova, 2015] Krassimir Markov, Krassimira Ivanova, “General Structure of Collect/Report Paradigm for Storing and Accessing Big Data”, International Journal “Information Theories and Applications”, Vol. 22, Number 3, 2015, ISSN 1310-0513 (printed), 1313-0463 (online), pp. 266 - 276.

[Markov et al, 1993b] Kr. Markov, Kr. Ivanova, I. Mitov, J. Ikonomov, "Information Model of the Ledger Account", IJITA Vol. 2, Num. 7, 1993, pp. 3 - 12.

[Markov et al, 1994] Kr. Markov, Kr. Ivanova, I. Mitov, J. Ikonomov, "Base Conception and Main Information Structures of Complex FOI", FOI-COMMERCE, Sofia, 1994 (in Bulgarian)

[Markov, 2004] Markov, K., "Multi-domain information model", Int. J. Information Theories and Applications, 11/4, 2004, pp. 303 – 308

[Markov, 2005] Markov, K., "Building data warehouses using numbered multidimensional information spaces", International Journal of Information Theories and Applications, 12(2), 2005, pp. 193 – 199

[Meigs & Meigs, 1989] R. Meigs, W. Meigs, "Financial Accounting", McGraw-Hill, 1989, 736 p.

[Samuelson & Nordhaus, 1989] P. Samuelson, W. Nordhaus, "Microeconomics", McGraw-Hill, 1989, 658 p.

Authors' Information

Ivanova Krassimira *– University of National and World Economy, Sofia, Bulgaria;*

e-mail: krasy78@mail.bg

Major Fields of Scientific Research: Software Engineering, Business Informatics, Data Mining, Multidimensional multi-layer data structures in self-structured systems