Годишник на секция "Информатика"
Съюз на учените в България
Том 1, 2008, 81-92

Annual of "Informatics" Section
Union of Scientists in Bulgaria
Volume 1, 2008, 81-92

# GENERALIZED NET MODELS REPRESENTING DATA WAREHOUSE OPERATIONS

**Panagiotis Chountas[1], Violeta Tasseva[2], Boyan Kolev[2] and Krassimir Atanassov[2]**

[1]Data & Knowledge Management Group-HSCS, University of Westminster Watford Road, Northwick Park, London, HA1 3TP, UK

[2] CLBME – Bulgarian Academy of Sciences, Bl. 105, Sofia-1113, BULGARIA

## 1 Introduction

A Data Warehouse is a subject oriented, [1], [2] integrated, non-volatile and time-variant collection of data [3] in support of management's decisions".

In operational systems, data is organized to support specific business processes. A business typically employs many different operational systems, each optimised for a special business process, and each with its own data store. In the DW, data from all these systems is integrated, both by definition, i.e., the same data has the same type, and by content, i.e., the value sets of an attribute are the same, wherever they occur.

In the typical operational system, data is often kept only for a short period of time. In a data analysis situation, however, the need to discover trends in the way business is doing and compare them with those of previous periods. Data-warehouses keep data for at least a couple of years, and many intend to keep it much longer.

Operational-Transactional data does not always have an explicit temporal dimension. Operational systems often only store the current state of data. In Data Warehouses when analysing data for trends, it is almost always important to know "the time of the data," so that all data in a DW can be related to a specific time point or interval. Also, not only the current value of data is stored, but often either snapshots of data at specific points in time, or a complete history of changes of the data.

Overall it can be said that Data warehouse is an amalgamated view on the data within an enterprise and a first step in integrating enterprise systems. Typically time is one of the dimensions we find in data warehouses allowing comparisons of different periods.

However the instances of dimensions change over time, organisations unite and separate, and organisational structures emerge and vanish, or evolve. In current data warehouse architectures these changes cannot be represented adequately since all dimensions are considered as orthogonal, putting restrictions on the validity of queries defined over several eras. In this paper we propose architecture for data warehouse systems, which also allows the querying of data belonged to evolving hierarchical structures.

This paper concentrates on the fundamental aspects of data warehouses (DWH) and their effects on On-Line Analytical Processing OLAP tools. Based on a new multidimensional model as defined in [4], [5] that is able to treat with imprecision over conceptual hierarchies and facts using Intuitionistic fuzzy logic, we define GN models for multidimensional DWH-OLAP operators, motivated by the observation that ignoring versions and flexible hierarchies leads to questionable expressive power and query semantics in many real life scenarios. Our suggested models allow the expression of OLAP queries in an elegant and intuitive fashion.

## 2 GN-models

In the present paper consists of Generalized Net (GN, see [6]) models of different operations in a Data Warehouse. Their description is divided [7], [8], [9] into four logical entities depending on the number of the operations' arguments and the access rights to the data warehouse.

### 2.1. GN-model of unary and binary read-only operations in a Data Warehouse

On Fig 1 is shown a GN-model of unary and binary operations. The description of this model is the following:

$$Z_1 = \langle\{l_1, l_7, l_{10}\}, \{l_2, l_3, l_4, l_5, l_6, l_7\}, r_1\rangle,$$

where:

$r_1 =$

| | $l_2$ | $l_3$ | $l_4$ | $l_5$ | $l_6$ | $l_7$ |
|---|---|---|---|---|---|---|
| $l_1$ | $W_{1\_2}$ | *false* | $W_{1\_4}$ | *false* | *False* | *false* |
| $l_7$ | *false* | $W_{7\_3}$ | *false* | $W_{7\_5}$ | $W_{7\_6}$ | $W_{7\_7}$ |
| $l_{10}$ | $W_{10\_2}$ | *false* | $W_{10\_4}$ | *false* | *False* | *false* |

,

and:

$W_{1\_2}$ = "An unary operation has to be applied over chosen OLAP cube",

$W_{1\_4}$ = "A binary operation has to be applied over chosen couple of OLAP cubes",

$W_{7\_3} = W_{10\_2} = W_{1\_2}$,

$W_{7\_5} = W_{7\_6} = W_{10\_4} = W_{1\_4}$,

$W_{7\_7} = W_{1\_2} \vee W_{1\_4}$,

$E_1$ and $E_2$ are the GN-models from [1] and [2].

The tokens enter place $l_1$ with initial characteristics:

"*demanded OLAP cube(s) and operation(s)*".

They pass through transition $Z_1$ and enter places $l_2$ or $l_4$ without obtaining any new characteristics. The choice of output place depends on the operation, which have to be applied over the selected cube(s). Initially there is one token in place $l_7$, which corresponds to the whole Data Warehouse. In case of existence of token in place $l_1$ the token in place $l_7$ splits into two or more tokens, one of which enters place $l_7$ and the others – places $l_3$, $l_5$ or $l_6$. The obtained in place $l_3$ characteristic is:

"*demanded OLAP cube*".

Tokens obtain as characteristics in places $l_5$ and $l_6$ respectively:

"*first demanded OLAP cube from the token in place* $l_1$",

"*second demanded OLAP cube from the token in place* $l_1$".

The tokens do not obtain any new characteristics in place $l_7$.

$$Z_2 = \langle\{l_8, l_9\}, \{l_{10}, l_{11}\}, r_2\rangle,$$

where:

$r_2 =$

| | $l_{10}$ | $l_{11}$ |
|---|---|---|

| | | |
|---|---|---|
| $l_8$ | $W_{8\_10}$ | $W_{8\_11}$ |
| $l_9$ | $W_{9\_10}$ | $W_{9\_11}$ |

,

and:
$W_{8\_10}$ = “There is a necessity of additional operation(s)”,
$W_{8\_11} = \neg W_{8\_10}$,
$W_{9\_10} = W_{8\_10}$,
$W_{9\_11} = W_{8\_11}$.

Tokens obtain as characteristics in places $l_8$ and $l_9$:
*“results of the work of GN $E_1$ ($E_2$) and necessity or not of additional operation(s)”*. They do not change these characteristics in places $l_{10}$ and $l_{11}$.

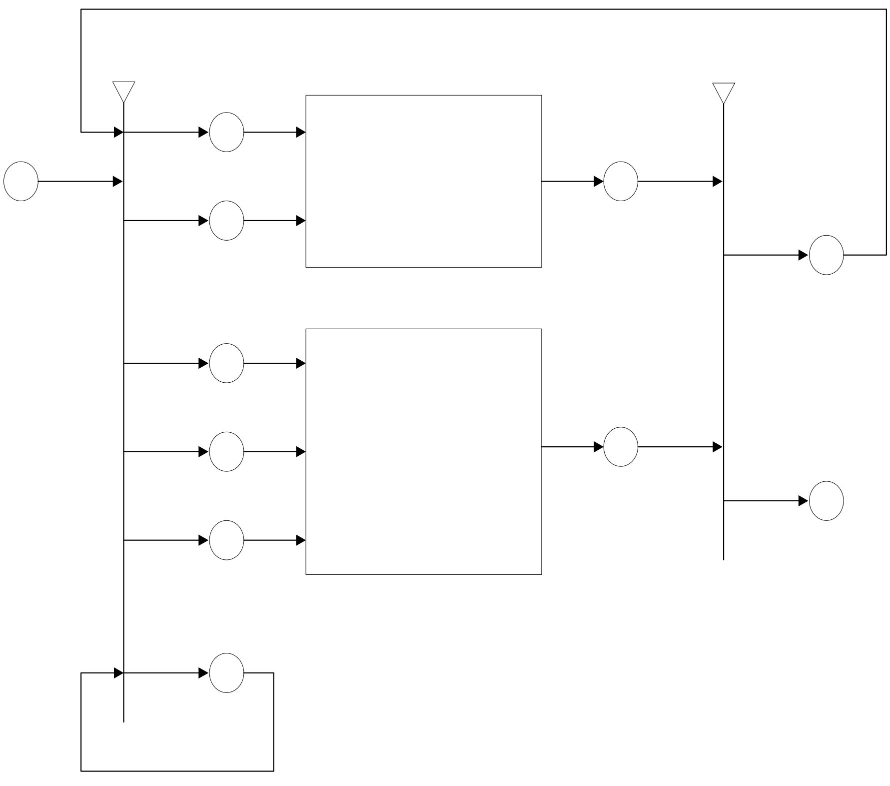

**Fig 1.** GN-model of unary and binary read-only operations in a Data Warehouse

$Z_1$



$l_2$

$l_1$

E

$l_3$

### 2.2. GN-model of Common Read-Only Operations in a Data Warehouse

On Fig 2 is represented a GN-model of common read-only operations in a Data Warehouse. Transition $Z_1$ has the following form:

$$Z_1 = <\{l_1, l_{2n+2}, l_{3n+3}\}, \{l_2, l_3, l_4, l_5, ..., l_{2n}, l_{2n+1}, l_{2n+2}\}, r_1>,$$

where:

$r_1 =$

| | $l_2$ | $l_3$ | $l_4$ | $l_5$ | … | $l_{2n}$ | $l_{2n+1}$ | $l_{2n+2}$ |
|---|---|---|---|---|---|---|---|---|
| $l_1$ | $W_{1\_2}$ | *false* | $W_{1\_4}$ | *false* | … | $W_{1\_2n}$ | *false* | *false* |
| $l_{2n+2}$ | *false* | $W_{2n+2\_3}$ | false | $W_{2n+2\_5}$ | … | *false* | $W_{2n+2\_2n+1}$ | $W_{2n+2\_2n+2}$ |
| $l_{3n+3}$ | $W_{3n+3\_2}$ | *false* | $W_{3n+3\_4}$ | *false* | … | $W_{3n+3\_2n}$ | *false* | *false* |

and:

$W_{1\_2}$ = “An unary operation has to be applied over chosen OLAP cube”,

$W_{1\_4}$ = “A binary operation has to be applied over chosen OLAP cubes”,

$W_{1\_2i}$ = “A more complex operation has to be applied over chosen OLAP cubes”,

$W_{2n+2\_2i+1} = W_{3n+3\_2i} = W_{1\_2i}$,

$W_{2n+2\_2n+2} = W_{1\_2} \vee W_{1\_4} \vee ... \vee W_{1\_2i}$,

$i \in \{1, n\}$, $n \in N$ and $E_1, E_2, ..., E_n$ are GN-models of respectively unary, binary and more complex operations over OLAP cubes.

As described above, the tokens enter place $l_1$ with initial characteristics:

*“demanded OLAP cube(s) and operation(s)”.*

They pass through transition $Z_1$ and each of them enters to one of places $l_2$, $l_4$, …, $l_{2i}$, …, $l_{2n}$ without obtaining any new characteristics. The choice of the output place depends on the operation, which have to be applied over the given cube. The token in place $l_{2n+2}$ corresponds to the Data Warehouse. Like in the previous model, this token splits into two or more tokens when given operation has to be applied. One of the resulting tokens enters again place $l_{2n+2}$, without obtaining new characteristic, the others enter respectively one or more of places $l_{2i+1}$, where $i \in \{1, n\}$. The tokens obtain as characteristics in places $l_{2i+1}$:

*“demanded OLAP-cube(s)”.*

Transition $Z_2$ has the following form:

$$Z_2 = <\{l_{2n+3}, l_{2n+4}, ..., l_{3n+2}\}, \{l_{3n+3}, l_{3n+4}\}, r_2>,$$

where:

$$r_2 = \begin{array}{c|cc} & l_{3n+3} & l_{3n+4} \\ \hline l_{2n+3} & W_{2n+3\_3n+3} & W_{2n+3\_3n+4} \\ l_{2n+4} & W_{2n+4\_3n+3} & W_{2n+4\_3n+4} \\ \dots & \dots & \dots \\ l_{3n+2} & W_{3n+2\_3n+3} & W_{3n+2\_3n+4} \end{array},$$

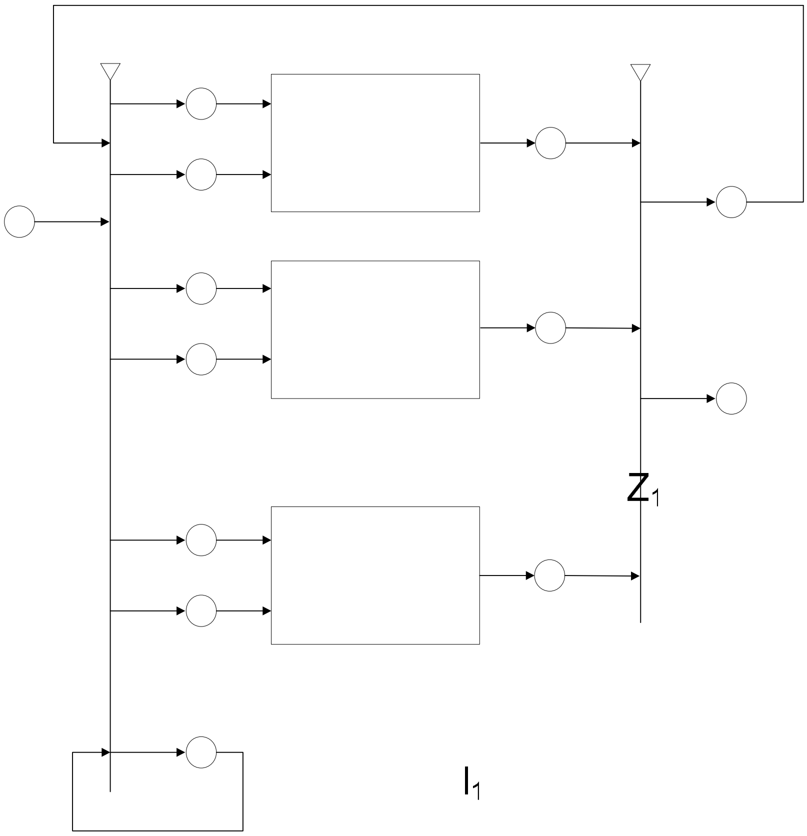


**Fig 2.** GN-model of common read-only operations in a Data Warehouse

and:

$W_{2n+i\_3n+3}$ = "There is a necessity of additional operation(s)",

$W_{2n+i\_3n+4} = \neg W_{2n+i\_3n+3}$,

$i \in \{1, n+2\}$.

Tokens obtain as characteristics in places $l_{2n+3}$, $l_{2n+4}$, ..., $l_{3n+2}$:

"*results of the work of the corresponding GN ($E_1$, $E_2$, …, $E_n$) and necessity or not of additional operation(s)*".

They do not change these characteristics in places $l_{3n+3}$ and $l_{3n+4}$.

**2.3. GN-model of Unary and Binary Read/Write Operations in a Versioned Data Warehouse**

The descriptions of the models from Figs 3 and 4 are analogous to the shown in Figs 1 and 2. The differences are in the possibility to write in the Data Warehouse. That is why the data could be refreshed by constructing of versions of the base.

$$Z_1 = <\{l_1, l_7, l_8, l_{11}, l_{12}\}, \{l_2, l_3, l_4, l_5, l_6, l_7\}, r_1>,$$

where:

| | | $l_2$ | $l_3$ | $l_4$ | $l_5$ | $l_6$ | $l_7$ |
|---|---|---|---|---|---|---|---|
| | $l_1$ | $W_{1\_2}$ | *false* | $W_{1\_4}$ | *false* | *false* | *false* |
| $r_1 =$ | $l_7$ | *false* | $W_{7\_3}$ | *False* | $W_{7\_5}$ | $W_{7\_6}$ | $W_{7\_7}$ |
| | $l_8$ | *false* | *false* | *False* | *false* | *false* | $W_{8\_7}$ |
| | $l_{11}$ | *false* | *false* | *False* | *false* | *false* | $W_{11\_7}$ |
| | $l_{12}$ | $W_{12\_2}$ | *false* | $W_{12\_4}$ | *false* | *false* | *false* |

,

and:
$W_{1\_2}$ = "An unary operation has to be applied over chosen OLAP cube",
$W_{1\_4}$ = "A binary operation has to be applied over chosen couple of OLAP cubes",
$W_{7\_3} = W_{12\_2} = W_{1\_2}$,
$W_{7\_5} = W_{7\_6} = W_{12\_4} = W_{1\_4}$,
$W_{7\_7} = W_{1\_2} \vee W_{1\_4}$,
$W_{8\_7}$ = "Data Warehouse has to be refreshed",
$W_{11\_7} = W_{8\_7}$,
$E_1$ and $E_2$ are the GN-models from [1] and [2].

The tokens enter place $l_1$ with initial characteristics:

*"demanded version of the Data Warehouse, OLAP cube(s) and operation(s)"*.

They pass through transition $Z_1$ and enter places $l_2$ or $l_4$ without obtaining any new characteristics. The choice of output place depends on the operation, which have to be applied over the selected cube(s). Initially there is one token in place $l_7$, which corresponds to all versions of the Data Warehouse. In case of existence of token in place $l_1$ the token in place $l_7$ splits into two or more tokens, one of which enters place $l_7$ and the others – places $l_3$, $l_5$ or $l_6$. The obtained in place $l_3$ characteristic is:

*"demanded OLAP cube from the needed version of the Data Warehouse"*.

Tokens obtain as characteristics in places $l_5$ and $l_6$ respectively:

*"first demanded OLAP cube from the token in place $l_1$"*,

*"second demanded OLAP cube from the token in place $l_1$"*.

The tokens obtain or do not obtain new characteristics in place $l_7$, depending on the existence or non existence of token in one of places $l_8$ or $l_{11}$. If the operation updates the cube, a new version of the cube is created and the token exits the subnet $E_1$ or $E_2$ through $l_8$ or $l_{11}$ and afterwards merges with the token in $l_7$. If the operation is read-only, the token exits the subnet $E_1$ or $E_2$ through $l_9$ or $l_{10}$ and then moves either to $l_{12}$, if the result cube needs to participate in another operation, or otherwise exits the GN through $l_{13}$. The possible new characteristic in place $l_7$ is:

*"refreshed Data Warehouse"*.

$$Z_2 = <\{l_9, l_{10}\}, \{l_{12}, l_{13}\}, r_2>,$$

where:

$$r_2 = \begin{array}{c|cc} & l_{12} & l_{13} \\ \hline L_9 & W_{9\_12} & W_{9\_13} \\ L_{10} & W_{10\_12} & W_{10\_13} \end{array},$$

$W_{9\_12}$ = “There is a necessity of additional operation(s)”,
$W_{9\_13} = \neg W_{9\_12}$,
$W_{10\_12} = W_{9\_12}$,
$W_{10\_13} = W_{9\_13}$.

Tokens obtain as characteristics in places $l_8$ and $l_9$:
*“results of the work of GN $E_1$ ($E_2$) and necessity of refreshment of the Data Warehouse”*.
In places $l_9$ and $l_{10}$ tokens obtain as characteristics:
*“results of the work of GN $E_1$ ($E_2$) and necessity or not of additional operation(s)”*.
They do not change these characteristics in places $l_{12}$ and $l_{13}$.

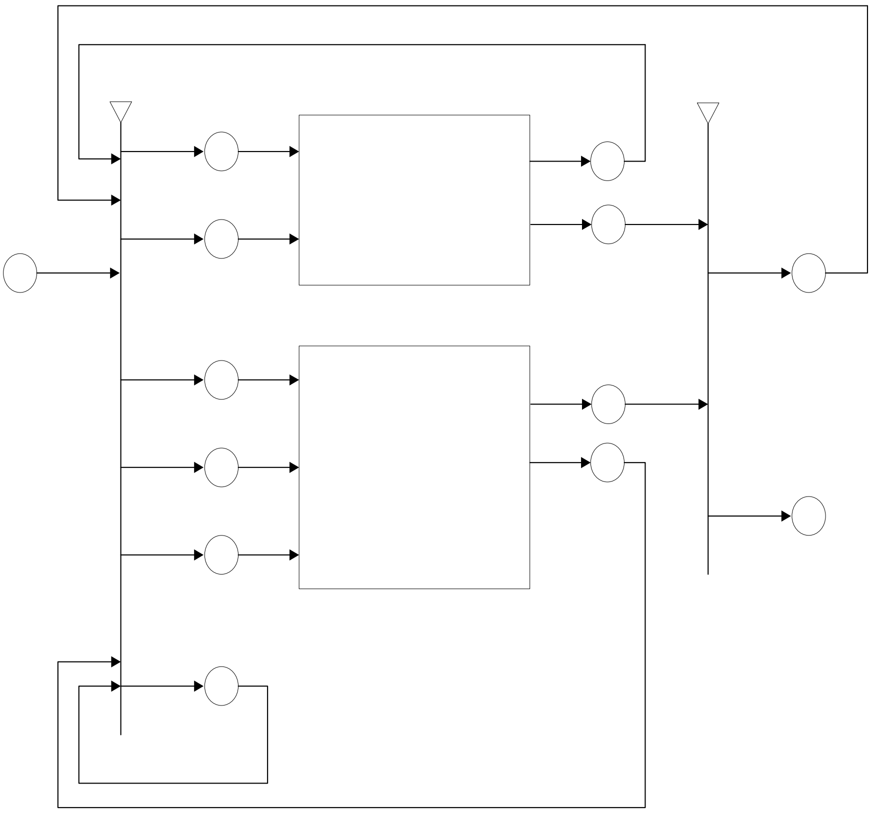

**Fig 3.** GN-model of unary and binary read/write operations in a versioned Data Warehouse



$Z_1$

$l_2$

E

$l_3$

$l_1$

**2.4. GN-model of Common Read/Write Operations in a Versioned Data Warehouse**

On Fig 4 is represented GN-model of common read/write operations in a versioned Data Warehouse. The description of the model is the following

$$Z_1 = <\{l_1, l_{2n+2}, l_{2n+3}, ..., l_{4n+2}, l_{4n+3}\}, \{l_2, l_3, l_4, l_5, ..., l_{2n}, l_{2n+1}, l_{2n+2}\}, r_1>,$$

where:

$r_1$ =

| | $l_2$ | $l_3$ | $l_4$ | $l_5$ | … | $l_{2n}$ | $l_{2n+1}$ | $l_{2n+2}$ |
|---|---|---|---|---|---|---|---|---|
| $l_1$ | $W_{1\_2}$ | *false* | $W_{1\_4}$ | *false* | … | $W_{1\_2n}$ | *false* | *false* |
| $l_{2n+2}$ | *false* | $W_{2n+2\_3}$ | *false* | $W_{2n+2\_5}$ | … | *false* | $W_{2n+2\_2n+1}$ | $W_{2n+2\_2n+2}$ |
| $l_{2n+3}$ | *false* | *false* | *false* | *false* | … | *false* | *False* | $W_{2n+3\_2n+2}$ |
| … | … | … | … | … | … | … | … | … |
| $l_{4n+2}$ | *false* | *false* | *false* | *false* | … | *false* | *false* | $W_{4n+2\_2n+2}$ |
| $l_{4n+3}$ | $W_{4n+3\_2}$ | *false* | $W_{4n+3\_4}$ | *false* | … | $W_{4n+3\_2n}$ | *false* | *false* |

where:

$W_{1\_2}$ = "An unary operation has to be applied over chosen OLAP cube",

$W_{1\_4}$ = "A binary operation has to be applied over chosen OLAP cubes",

$W_{1\_2j}$ = "A more complex operation has to be applied over chosen OLAP cubes", for $j \in \{3, n\}$

$W_{2n+2\_2i+1} = W_{4n+3\_2i} = W_{1\_2i}$,

$W_{2n+2\_2n+2} = W_{1\_2} \lor W_{1\_4} \lor ... \lor W_{1\_2i}$,

$W_{2n+2i+2\_2n+2}$ = "Data Warehouse has to be refreshed",
$i \in \{1, n\}$, $n \in N$ and $E_1$, $E_2$, ..., $E_n$ are GN-models of respectively unary, binary and more complex operations over OLAP cubes.

As described above, the tokens enter place $l_1$ with initial characteristics:

*"demanded version of the Data Warehouse, OLAP cube(s) and operation(s)"*.

They pass through transition $Z_1$ and each of them enters to one of places $l_2$, $l_4$, …, $l_{2i}$, …, $l_{2n}$ without obtaining any new characteristics. The choice of the output place depends on the operation, which have to be applied over the given cube. The token in place $l_{2n+2}$ corresponds to the Data Warehouse with all its versions. Like in the previous model, this token splits into two or more tokens when given operation has to be applied. One of the resulting tokens enters again place $l_{2n+2}$, the others enter respectively one or more of places $l_{2i+1}$, where $i \in \{1, n\}$. If the operation updates the cube, a new version of the cube is created and the token exits the subnet $E_i$ through $l_{2n+2i+2}$ and afterwards merges with the token in $l_{2n+2}$. If the operation is read-only, the token exits the subnet $E_i$ through $l_{2n+2i+1}$ and then moves either to $l_{4n+3}$, if the result cube needs to participate in another operation, otherwise exits the GN through $l_{4n+4}$. Depending on the existence or non existence of token in one of places $l_{2n+2i+2}$, the tokens obtain or do not obtain new characteristics. In case of existence, the token merges with the token from place $l_{2n+2}$ and obtains as a characteristic:

*"refreshed Data Warehouse"*.

The tokens obtain as characteristics in places $l_{2i+1}$:

*"demanded OLAP cube(s)"*.

Transition $Z_2$ has the following form:

$$Z_2 = <\{l_{2n+4}, l_{2n+5}, ..., l_{4n+8}\}, \{l_{4n+10}, l_{4n+11}\}, r_2>,$$

where:

$r_2$ =

| | $l_{4n+10}$ | $l_{4n+11}$ |
|---|---|---|
| $l_{2n+4}$ | $W_{2n+4\_4n+10}$ | $W_{2n+4\_4n+11}$ |
| $l_{2n+5}$ | $W_{2n+5\_4n+10}$ | $W_{2n+5\_4n+11}$ |
| … | … | … |
| $l_{4n+8}$ | $W_{4n+8\_4n+10}$ | $W_{4n+8\_4n+11}$ |

,

where:
$W_{2n+4\_4n+10}$ = "There is a necessity of additional operation(s)",
$W_{2n+2i+1\_4n+10} = W_{2n+4\_4n+10}$,
$W_{2n+i\_3n+4} = \neg W_{2n+i\_3n+3}$,
$i \in \{1, n+2\}$.

Tokens obtain as characteristics in places $l_{2n+3}$, $l_{2n+6}$, ..., $l_{4n+2}$:

*"results of the work of the corresponding GN ($E_1$, $E_2$, …, $E_n$) and necessity of refreshment of the Data Warehouse"*.

In places $l_{2n+4}$, $l_{2n+5}$ and $l_{4n+1}$ tokens obtain as characteristics:

*"results of the work of the corresponding GN ($E_1$, $E_2$, …, $E_n$) and necessity or not of additional operation(s)"*.

They do not change these characteristics in places $l_{4n+3}$ and $l_{4n+4}$.

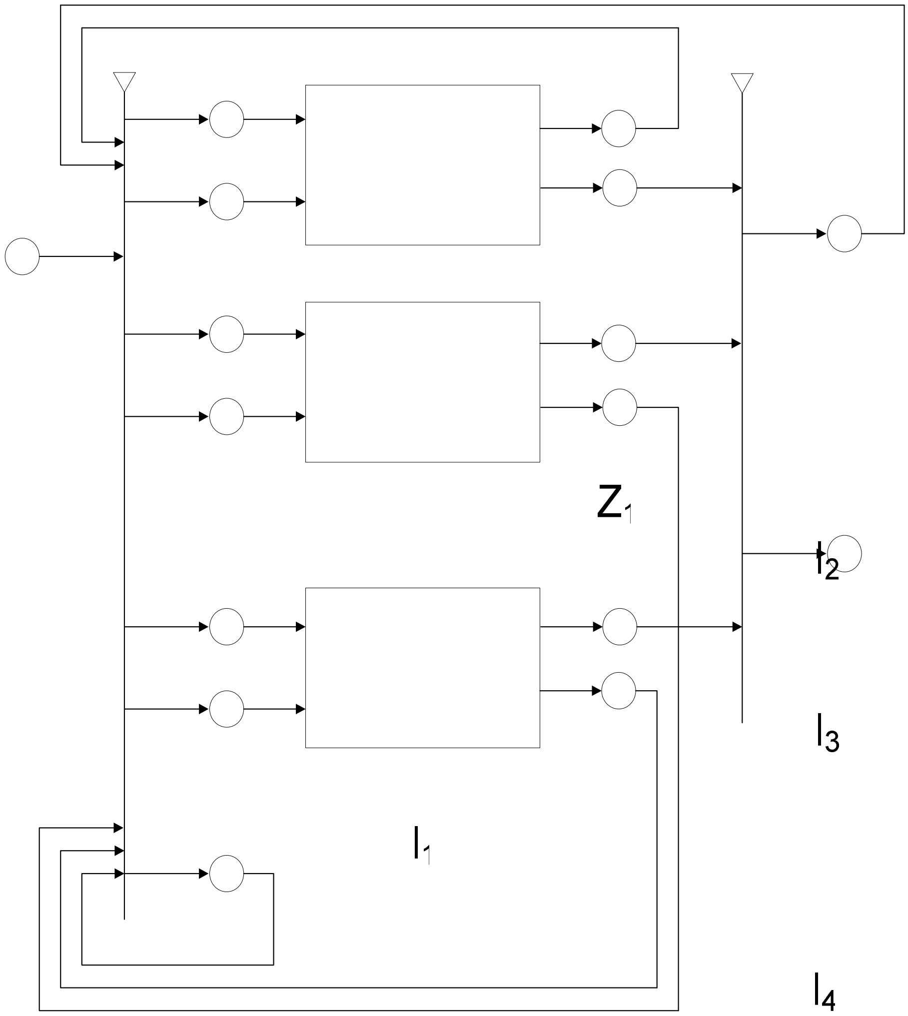


**Fig 4.** GN-model of common read/write operations in a versioned Data Warehouse

## 3 Conclusion

Traditionally, there is no real-time connection between a DWH and its data sources. This is mainly because the write-once read-many decision support characteristics would conflict with the continuous update workload of operational systems resulting in poor response times. Consequently, up until recently, timeliness requirements were restricted to mid-term or long-term time windows. We reviewed the issue of time in data warehouses and meaning of time as part of a conventional OLAP architecture. We proposed an extension of the conventional OLAP architecture in order to handle flexible and evolving hierarchical structures defined as well over the time using versions. We introduced the automatic recommendation of analysis according to the context of users' explorations in order to guide the decision making with the aid of GN & IFS and the knowledge represented through them.

## References


[1] KIMBALL, R. "The Data Warehouse Toolkit" New York: John Wiley & Sons, 2001.

[2] JARKE M., et al. "Fundamentals of data warehouses", Springer-Verlag, London, 2002.

[3] PEDERSEN T., JENSEN C., DYRESON C., "A foundation for capturing and querying complex multidimensional data", Information Systems, Vol. 26, 2001, 383-483.

[4] ROGOVA E., CHOUNTAS P., ATANASSOV, K., "Flexible Hierarchies and Fuzzy Knowledge-based OLAP" FSKD 2007, published by IEEE Computer Society Press.

[5] ROGOVA, E., CHOUNTAS, P., KOLEV, B., "Intuitionistic Fuzzy Knowledge- based OLAP", Notes Intuitionistic Fuzzy Sets, Vol. 13, No. 2, ISSN 1310-4926, 2007, 88-100.

[6] ATANASSOV, K., "Generalized Nets", World Scientific, Singapore, 1991.

[7] CHOUNTAS, P., ATANASSOV, K., TASSEVA, V., KOLEV, B., "GN Model of binary operations over IF cubes". – In: Prof. of Eighth Int. Workshop on GNs, Sofia, 26 June 2007, 66-72

[8] CHOUNTAS, P., ATANASSOV, K., TASSEVA, V., KOLEV, B., "Generalized net model of selection operation over Intuitionistic fuzzy cube" – Issues in Intuitionistic Fuzzy Sets and Generalized nets, 2007, Vol. 6, Wydawnictwo, WSISIZ, in press.

[9] CHOUNTAS, P., ATANASSOV, K., TASSEVA, V., KOLEV, B., "Generalized net model for querying operators over multiple Intuitionistic fuzzy cubes". – Issues in Intuitionistic Fuzzy Sets and Generalized nets, 2007, Vol. 7. Wydawnictwo, WSISIZ, in press.