

# DATA WAREHOUSE BASED DECISION SUPPORT SYSTEM IN NUCLEAR POWER PLANTS

**Berislav Nadinić**  
HTMobile  
Kružna 83 Zagreb  
berislav.nadinic@htmobile.hr

## ABSTRACT

Safety is an important element in business decision making processes in nuclear power plants. Information about component reliability, structures and systems, data recorded during the nuclear power plant's operation and outage periods, as well as experiences from other power plants are located in different database systems throughout the power plant. It would be possible to create a decision support system which would collect data, transform it into a standardized form and store it in a single location in a format more suitable for analyses and knowledge discovery. This single location where the data would be stored would be a data warehouse. Such data warehouse based decision support system could help make decision making processes more efficient by providing more information about business processes and predicting possible consequences of different decisions. Two main functionalities in this decision support system would be an OLAP (On Line Analytical Processing) and a data mining system. An OLAP system would enable the users to perform fast, simple and efficient multidimensional analysis of existing data and identify trends. Data mining techniques and algorithms would help discover new, previously unknown information from the data as well as hidden dependencies between various parameters. Data mining would also enable analysts to create relevant prediction models that could predict behaviour of different systems during operation and inspection results during outages. The basic characteristics and theoretical foundations of such decision support system are described and the reasons for choosing a data warehouse as the underlying structure are explained. The article analyzes obvious business benefits of such system as well as potential uses of OLAP and data mining technologies. Possible implementation methodologies and problems that may arise, especially in the field of data integration, are discussed and analyzed.

## 1 INTRODUCTION

Nuclear power plants, judging by their complexity, represent together with space exploration and military technology the pinnacle of technological development. In nuclear power plants technical achievements from many scientific fields are applied, notably from mechanical and electrical engineering and computer science. Principles of quality assurance and management based on deterministic and probabilistic principles are also used. Primary success criteria for nuclear power plants are safety and protection of environment from hazardous radiation, while classical economic criteria are of secondary importance.

It is usually assumed that managing such a system in regard to maintenance and operation requires accurate and up-to-date information. The problem is that information required for making optimal decisions about the operation of a nuclear power plant is usually stored in different database system scattered throughout different locations in the plant. Except being physically separated, such database systems usually contain mutually separated and inconsistent data. On top of that, the data is commonly in various formats.

So the following problem arises: how to extract any meaningful information (or knowledge) about the processes in the company from this multitude of different systems that could help decision-making processes in the nuclear power plant.

The solution to this problem is to create a data warehouse based decision support system which could integrate data from different source database systems and store it in a single location in a format more suitable for queries and analyses. The purpose of this article is to discuss possible applications of this concept to nuclear power plants taking into consideration all specific characteristics of nuclear technology.

## **2 OVERVIEW OF THE MOST IMPORTANT DATABASE SYSTEMS WHICH ARE USUALLY FOUND IN NUCLEAR POWER PLANTS**

From a technological and economical point of view, the most important database systems in nuclear power plants are the following ones:

- Database systems containing all relevant technical data about specific components and systems of the nuclear power plant. Such systems contain i.e. data about weights, dimensions, working parameters and characteristics, used materials and their mechanical, chemical and other properties that are relevant for their maintenance and operation.
- Database systems containing data about spare parts and consumables related to nuclear power plant's regular operation
- Database systems containing operational data based on records from various measurement devices (pressures, temperatures, mass flows, neutron fluxes,  $\alpha$   $\beta$  and  $\gamma$  radiation, steam humidity, currents, voltages, vibrations, quality parameters for the heat transmission media, levels in storage tanks etc.) This database (or databases) is usually extremely large, because it stores data from the start of nuclear power plant's operation and the frequency of readings (depending on the variable measured) can range from several readings per day up to several dozen times per second.
- Database systems containing data about inspection of components and systems using non-destructive methods during regular and forced outages which can influence the operation of a certain component or a system. Typical inspections of this type are pipeline tests, steam generator heat exchange and condenser tube inspections etc. Methods of non-destructive testing which are being used are mostly visual, ultrasound, eddy current, penetrants, radiography, magnetic particles, acoustic emission, etc.
- Database systems containing results of functional system tests during planned and forced outages. Such tests are used to check the functioning of the system with known inputs in order to check their behavior in different conditions. These tests are very important from a system regulation perspective, as well as functioning of its components including all electrical-electronic components and applied software.
- Database systems containing financial data. These systems include data about all existing expenses related to operation of nuclear power plant such as acquisition of spare parts, consumables, labor costs etc. All other financial data which could potentially affect the decision making processes are also included in this systems such as costs of produced electrical energy and credit debts.
- Database systems containing data about system and component reliability. This system should contain data resulted from Probabilistic Safety Assessment (PSA) of systems and components in nuclear power plant, as well as such data from other similar plants all over the world in the scope in accordance with its availability.
- Database system containing national as well as international (IAEA) regulatory requirements, codes, standards and norms. It would be useful to obtain access to data from countries with a developed nuclear industry (USA, France, Germany, Russia)

- International database systems. These systems partially include systems from all the categories listed above, the only difference being that this data comes from similar power plants all over the world. Such data is usually acquired through IAEA, EPRI, WANO etc.

All types of database systems listed above contain operational data (operational data is usually considered data that is used for performing business processes in an organization). Usually the majority of this data is collected and stored via transactional systems. The name "transactional" is derived from the fact that access to this data is usually performed using transactions (each read, write or update operation is a transaction) resulting in a large number of transactions containing small data quantities.

When it comes to extracting useful information or knowledge that could assist in decision making, several problems appear:

- physical separation – data is stored in physically separated database systems
- different formats – data can be stored in a variety of database types (ORACLE, MS Access, Informix...) in different formats (character, string, text...)
- lack of concurrency – different database system may contain inaccurate information often as a result of different computation algorithms

The following example illustrates the problem: suppose that in a certain part of the plant a circulation pump broke down. This would usually mean that a long and painstaking process of manual data analysis would begin. The analyst would have to manually inspect database tables containing thousands of rows that have record certain measurements for the pump (operating data, results of nondestructive examinations performed in previous years...). All calculations would have to be done manually in end user tools like MS Excel. The analyst would also have to browse through many other separate systems in order to look for any relevant stored information that could be related to the circulation pump that broke down. He would have problems with joining data from different systems (i.e. *mass flow* can be a numeric value in an ORACLE database and a string value in MS Access) and interpreting stored transactional data (i.e. what do table fields *pu\_id* and *pu\_fx* actually store?). The analyst would require weeks to browse and filter out the data that could provide him with some information about the causes of the pump's breakdown. However, he would still lack the "big picture" that would provide him with insights. Also, all hidden dependencies, relations and trends would be virtually unavailable to him.

Suppose now that the plant in question has an operational data warehouse based decision support system. The data warehouse would contain data that is historic, integrated, cleansed concurrent and stored in a single format. In the case of the circulation pump, data warehouse would contain among other, data from the following systems:

- System that keeps track of all the parameters during plant's operation. From this data it would be seen whether the pump in question (or even other pump of the same type) had caused any disturbances during operation and how did this reflect itself onto the operation of the system and the entire plant
- System that contains records about all repairs in the plant. From this data it would be seen if the pump in question (and other pumps of this type) was repaired, what were the reasons of those previous breakdowns, and the amount of time and spare parts used for repair
- system that tracks all spare parts and consumables expenditure with corresponding delivery times for each part
- System that records all workforce allocations. In this case it would be the workforce necessary to perform regular maintenance for the mentioned pump
- System containing data about all other pumps that could be used as a replacement with all corresponding data about their performance
- Systems that provide reliability factors for the specific components of the system. In this case it would be the reliability factor for the relevant pump and for the potential types of replacement pumps. This kind of information is essential, because it is closely linked with safety criteria which is of the utmost importance when it comes to decision making in nuclear power plants

- Systems that provide information about the influence of a certain component on the general safety of the nuclear power plant. In the case of the circulation pump, this influence may vary from small (i.e. a pump in the secondary circuit) to crucial (i.e. a pump in the safety system)

As mentioned before, if such data is stored in the data warehouse cleansed, integrated, concurrent and in a format more suitable for analyses it is possible to using analyzing tools and techniques to extract useful information from the data warehouse that could help the decision making process. In this case this means that management of the plant (or the analyst) could analyze data on different levels of summarization to get a clear picture of the pump's impact on the operation of other systems, search for hidden dependencies between various parameters, analyze possible scenarios of operation and trends, predict breakdowns and create optimal maintenance strategy etc.

For this reasons a data warehouse based decision support system would be extremely useful in nuclear power plants because it would enable different analyses and predictions to be performed on integrated historical data which could discover unexpected relationships and identify trends. Figure 1 clearly illustrates the types of data that would be loaded in such a system:

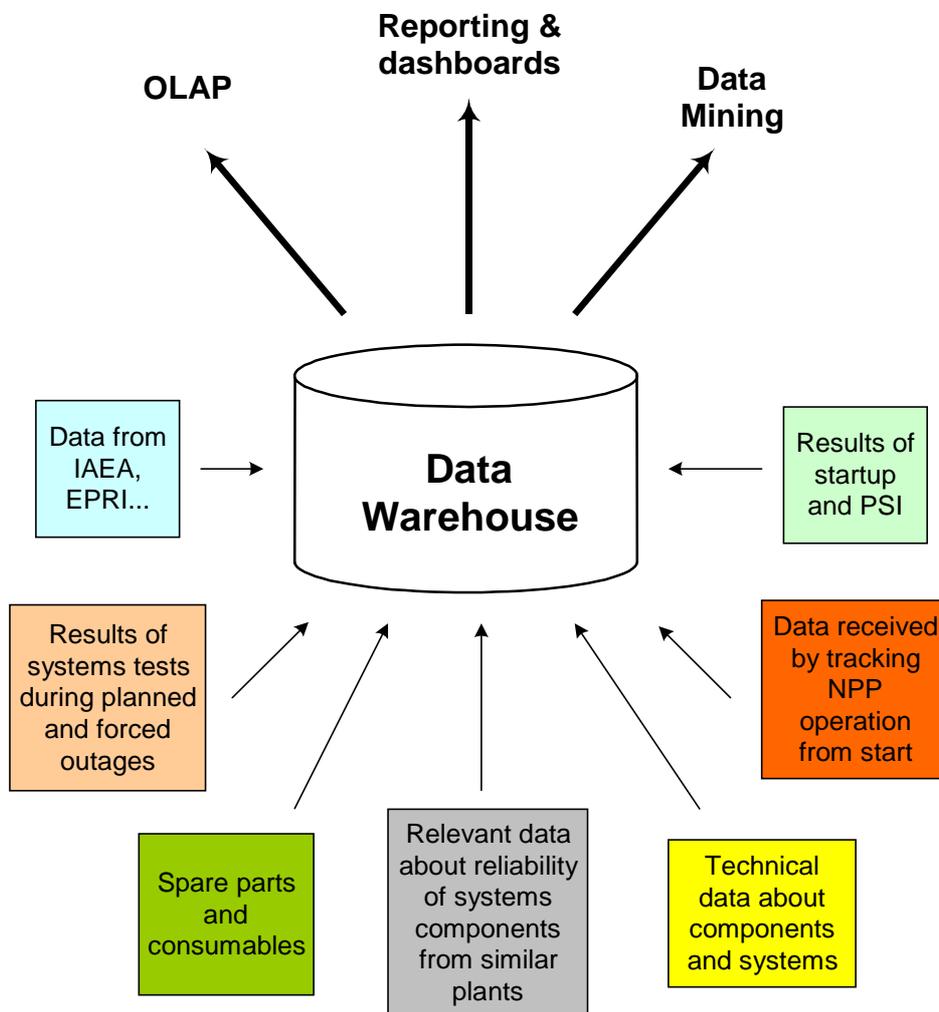


Figure 1. Types of data loaded into the data warehouse

### 3 DATA WAREHOUSE

Firstly, it is important to stress that a data warehouse is not a substitute for standard operation systems, but in fact it only stores copies of transactional data transformed and specially structured for queries and analyses.

The main element of a data warehouse based decision support system is naturally the data warehouse itself. W. H. Inmon, who developed the information architecture, which became the foundation on which data warehouses are designed, offers the following definition. A data warehouse is a subject-oriented, integrated, time-variant, non-volatile collection of data in support management's decision making process. A subject-oriented view allows analysts to easily locate and use the data they need. Because the warehouse is nonvolatile and time variant, data are never deleted. This allows users to perform historical or trend analysis. The warehouse data are integrated and users can therefore examine and join data generated from many systems.

Basic characteristics of data warehouse systems are the following:

- 1) Subject oriented - operational database systems, such as ones storing measured parameters in components (pressure, temperature...) or payroll databases, are organized around processes or functional areas. Thus, the data was relative to the measurement process or the payroll application. Data on a particular subject, such as production, repairs or employees, is maintained separately (and usually inconsistently) in a number of different databases. In contrast, a data warehouse is organized around subjects. This subject orientation presents the data in a much easier-to-understand format for analysts and management.
- 2) Integrated - integration of data within a warehouse is accomplished by making the data consistent in format, naming, and other aspects. Operational systems, for historic reasons, often have major inconsistencies in data representations. For example, a set of operational database systems may represent status "working" and "broken down" by using codes such as "w" and "b", by "1" and "2", or by "yes" and "no". Often, the inconsistencies are more complex and subtle. In a data warehouse, on the other hand, data is always maintained in a consistent fashion.
- 3) Time variant - data warehouses are time variant in the sense that they maintain both historical and (nearly) current data. Operational database systems, in contrast, contain only the most current, up-to-date data values. Furthermore, they generally maintain this information for no more than a year (and often much less). In contrast, data warehouses contain data that is generally loaded from the operational databases daily, weekly, or monthly which is then typically maintained for a period of 3 to 10 years. This is a major difference between the two types of environments.
- 4) Historical - historical information is of high importance to decision making, because management often wants to understand trends and relationships between data. For example, the management of a nuclear power plant wants to see the correlation between water chemistry regime in the secondary side in the previous years and corrosion processes on steam generator tubes. This is information that is almost impossible - and certainly in most cases not cost effective - to determine with an operational system.
- 5) Non-volatile - non-volatility, the final primary aspect of data warehouses, means that after the data warehouse is loaded there are no changes, inserts, or deletes performed against the informational database. The data warehouse is, of course, first loaded with transformed data that originated in the operational database system. The data warehouse is subsequently reloaded or, more likely, appended on a periodic basis (usually nightly, weekly, or monthly) with new transformed data from the operational databases. Outside of this loading process, the data warehouse generally stays static. Due to non-volatility, the data warehouse can be heavily optimized for query processing.

For better reference, basic characteristics of operational systems and data warehouses are listed in table 1:

Table 1. Basic characteristics of operational systems and data warehouses

Operational systems	Data warehouses
Operational systems are generally designed to support high-volume <i>transaction processing</i> with minimal back-end reporting.	Data warehousing systems are generally designed to support high-volume <i>analytical processing</i> (i.e. <i>OLAP</i> ) and subsequent, often elaborate <i>report generation</i> .
Operational systems are generally <i>process-oriented</i> or <i>process-driven</i> , meaning that they are focused on specific business processes or tasks. Example tasks include measurement of a certain component, financial data processing etc.	Data warehousing systems are generally <i>subject-oriented</i> , organized around functional areas that the organization needs information about. Such subject areas are usually populated with data from one or more operational systems. As an example, revenue may be a subject area of a data warehouse that incorporates data from operational systems that contain data about electrical energy production, running costs, consumables and spare parts cost, payroll data...
Operational systems are generally concerned with <i>current data</i> .	Data warehousing systems are generally concerned with <i>historical data</i> .
Data within operational systems are generally <i>updated regularly</i> according to need.	Data within a data warehouse is generally <i>non-volatile</i> , meaning that new data may be added regularly, but once loaded, the data is <i>rarely changed</i> , thus preserving an ever-growing <i>history of information</i> . In short, data within a data warehouse is generally <i>read-only</i> .
Operational systems are generally optimized to perform <i>fast inserts and updates</i> of relatively <i>small volumes of data</i> .	Data warehousing systems are generally optimized to perform <i>fast but complex retrievals</i> of relatively <i>large volumes of data</i> .
Operational systems are generally <i>application-specific</i> , resulting in a multitude of partially or non-integrated systems and <i>redundant data</i> (e.g. expenses data is not integrated with payroll data).	Data warehousing systems are generally <i>integrated</i> at a layer above the application layer, avoiding data redundancy problems.
Operational systems generally are used by a large number of users	Data warehouse systems are usually used by a small number of analysts and management

### 3.1 Design of a data warehouse based decision support system

How to create a decision support system? Firstly, it needs to be understood that such a system is not a "black box" which can be bought from a vendor and used in many different organizations. A decision support system is developed inside an organization from the beginning and is inseparable from processes taking place in the organization. The design of the system is usually performed in the following steps:

- Identification of source systems – the first step. It is necessary to analyze all data in database systems within the organization (or even outside) and determine which of them could be the source systems (systems from which the data will be loaded in the data warehouse). For source system identification it is necessary to be thoroughly familiarized with all processes within the organization. It is also necessary to foresee possible analyses and queries that will be performed in the system. This means that if the management of the nuclear power plants wishes to analyze electrical power production, all relevant data from different systems that could be related to power production must be loaded in the data warehouse.
- Definition of architecture and creation of a data model - it is necessary to define system's architecture and create a data model that includes all relational tables in the data warehouse formed in special *star* and *snowflake* schemes
- Creation of ETL processes – the process of loading the data into the data warehouse is performed in three distinct phases – extraction, transformation and loading (hence the name ETL processes). Extraction phase usually encompasses all tasks related to extraction of data from source systems and their loading into the "staging area", a special pre-warehouse repository where the transformation processes are taking place. In the transformation phase data is cleansed, joined and transformed in a single format. During the loading phase such transformed data is loaded into the data warehouse itself. Since the data from the source system will be regularly loaded into the data warehouse, it is necessary to pay a lot of attention to creation of such automated processes. Three key issues that deserve special attention during the process creation are the following:
  - a) Data integration – integration of data related to different functional entities (i.e. production, spare parts, employees...) must be achieved. This means that data from different source systems must be integrated into single entity notwithstanding possible differences in the database type and data format. Data integration should also discard all table columns or rows that are not needed for storage in the data warehouse
  - b) Data cleansing – before the loading phase, data should be cleansed. Cleansing usually resolve all inconsistencies that may arise from different systems. Also, all errors that may be present in the data must be addressed and corrected before the loading phase.
  - c) Summarization and aggregation – in the source systems data is usually stored at the highest detail level. This means that measurements of a certain component are recorded and stored i.e. each second. It is necessary to determine whether it analyses and queries that will be performed from the data warehouse actually need such high detail level or perhaps storing average values for each hour or even day in the data warehouse may be sufficient. In this case data is aggregated to a higher level in the data warehouse.
- Implementation of tools and interfaces for analysts – it is necessary to implement tools that will enable end users of the decision support system (management or trained analysts) to perform analyses and queries identify trends and predict behaviors.

Figure 2 shows a high level architecture of a data warehouse based decision support system. The data is loaded from source systems into the staging area, where data transformation occurs. An Operational Data Store (ODS) is a type of database often used as an interim area for a data warehouse. Unlike a data warehouse, which contains static data, the contents of the ODS are updated through the course of business operations. An ODS is designed to quickly perform relatively simple queries on small amounts of data (such as finding the status of a certain part of the system), rather than the complex queries on large amounts of data typical of the data warehouse. An ODS is similar to your short term memory in that it stores only very recent information; in comparison, the data warehouse is more like long term memory in that it stores relatively permanent information. From the ODS, data is being loaded into the data warehouse. Information and knowledge can be extracted from data warehouse in various ways. Analysts use reporting and query tools to perform direct queries into the data warehouse and obtain reports. OLAP system enables them to get the complete picture of the data from different perspectives and different levels of aggregation. Data Mining

system searches for hidden dependencies in the data and can be used to predict future behaviour of components and systems based on historical data in the data warehouse.

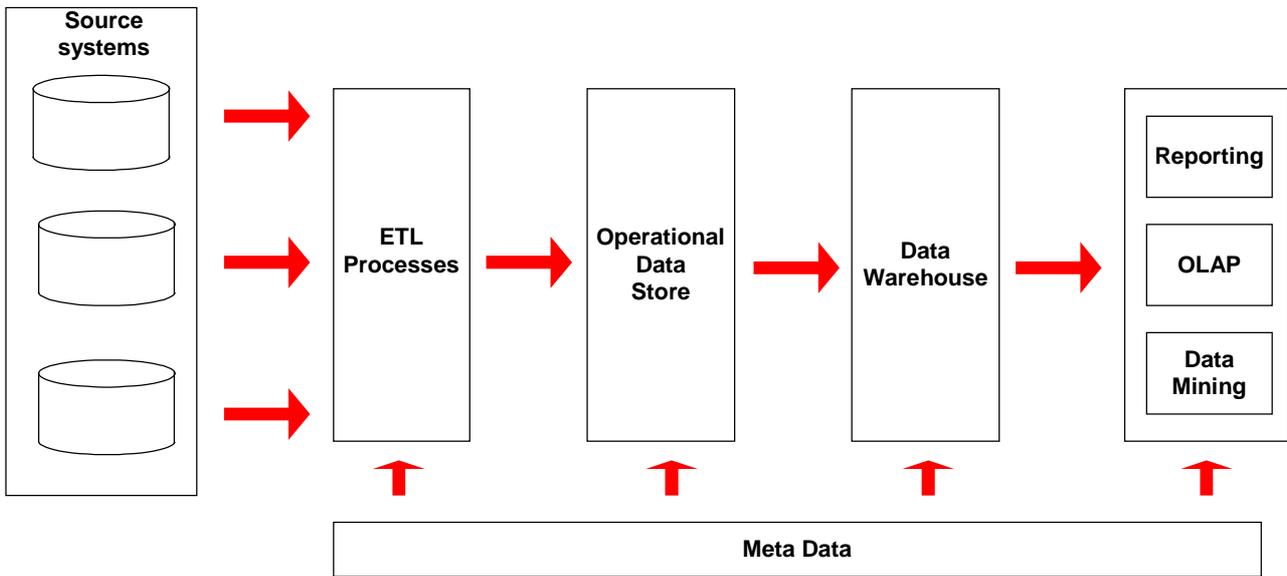


Figure 2. Decision support system architecture

### 3.2 Possible uses for the decision support system

Queries and analyses can be performed on the data stored in the data warehouse. Types of queries and analyses that can be performed on the data stored in the data warehouse range from simple to complex and depend on the end-user tools.

Analysts can get information from the decision support system by using the following methods and technologies:

- Information processing – this encompasses basic functionalities for creating reports and performing *ad hoc* queries for data retrieval from the warehouse. Analyses on this level are very simple and are usually consisting of sorting and summing of the data. Typical visualization tools for this method are MS Excel or MS Access.
- OLAP (On-Line Analytical Processing) –OLAP or FASMI (Fast Analysis of Shared Multidimensional Information) involves representing data (typically summary data) as user-defined dimensions, not tables. For example, data from production analysis may be organized by component, system, safety importance and other dimensions. Dimensions are almost always related in hierarchies, and a multidimensional database can have multiple hierarchies. System hierarchies can range from the nuclear power plant as a whole down to individual subsystems and components. Time is common to almost all analytical applications, so most multidimensional servers have built-in time hierarchies (weeks roll up into months). OLAP tools enable analysts to analyze trends and multidimensional data using different views and different levels of aggregation.
- Data Mining - an analytic process designed to explore large amounts of data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. The ultimate goal of data mining is prediction - and predictive data mining is the most common type of data mining. The term Predictive Data Mining is usually applied to identify data mining projects with the goal to identify a statistical or neural network model or set of models that can be used to predict some response of interest. Data mining presents the

"evolution" of systems based on knowledge discovery. The purpose of data mining is to automate the process of trend identification and discovery of patterns in order to minimize the end user involvement. Patterns do not have to be directly stored in warehouse tables, but they can span over dimensional and hierarchies. Such patterns are very hard to discover using OLAP techniques.

- Dashboards – visualization software can be used to create dashboards with graphic images that represent various reports, analyses and events. A personalized dashboard may be created which represents various key performance indicators such as electrical energy production, spare parts and consumables costs etc.

## 4 CONCLUSIONS

As a result of previous considerations some important conclusions can be reached. Firstly, decision making processes in nuclear power plants are very complex due to the fact that huge quantities of complex data are present in its operational systems. Also, economic and safety criteria must be taken into account while making decisions. It is almost impossible to extract meaningful information that could assist in the decision making processes because the data is spread out throughout separated systems in the plant. In order to extract such hidden information, a data warehouse based decision support system may be used. The decision support system could offer analysts and management analytical capabilities on integrated and archived historical data through querying, reporting, OLAP and data mining. However, it is important to stress the fact that in nuclear power plants the safety criteria is of primary concern, so the decision support system has to provide in depth view of relevant safety parameters associated with any dimension. The proposed concept is a novel way to increase safety, availability and profit in nuclear power plants.

## REFERENCES

William H. Inmon: Building the data warehouse, Wiley & Sons 2002.

Ralph Kimball: Two powerful ideas – the foundations for modern data warehousing, Intelligent Enterprise 2002.