INTEGRATION PROCESSES QUALITY IN HETEROGENOUS ENVIRONMENTS

Ion Ivan[1]
Leonard Săcuiu[2]

Abstract

It presents heterogeneous distributed software applications concept. It describes integration techniques. It defines quality of integration processes regarding heterogeneous environments. It defines quality metrics for heterogeneous e-commerce applications.

Keywords*:* integration, distributed environments, *software process*.

1. Heterogeneous Distributed Software Applications

Economic entities struggling with increasingly higher competitiveness are extending its operating markets. Develop strategies to minimize costs and maximize profits by setting up working points in the country with lower cost of human resources for similar training and qualification level. They devise appropriate marketing strategies that focus on target market characteristic's depending on the local economic environment. Organizations are expanding their influence through acquisitions of other companies or through partnerships by initiating joint projects on short or long term. Companies feel the need to integrate other business partner's information flows with its own information flows.

From this perspective geared IT resources must meet the requirements of this emphasis dynamic. The traditional approach in which software companies promote their current standards is no longer useful because they no longer meet the needs of data collection and information from multiple data sources focusing on data and process integration.

IT developments brought major changes to the applications with acquisition costs often prohibitive for small and medium organizations. Financially strong organizations wishing decrease software purchase costs, because of this economic environment has shown restraint in buying computer products to keep pace with dynamic software technology. Another issue concerns the failure to waive the software application or data sources within the organization, the result of their information management departments. Since practically all companies are faced with the need to extract data from multiple data sources, analyzing, reporting and formation of new deposits of refined information to meet current operational needs.

This situation generates the existence of a set of software products used by economic entities, data sources and technologies that mark the evolution and dynamics of information technology and communications. Being economic entities E1, E2, ..., En.

[1] Ion Ivan, ASE, Bucharest, Romania, email: ionivan@ase.ro

[2] Leonard Săcuiu, ASE, Bucharest, Romania, email: leo.sacuiu@gmail.com

Each entity has a number of operational software applications or systems designed and implemented at some point in time using:

- technologies available at that time or permitted by company budget restrictions formed in accordance with the concrete conditions of the economic environment;
- economic entities existing expertise or attracted by contracts and short or long time cooperation with other economic entities to produce software application;
- appropriate restrictions to specific strategic objectives developed by management in accordance with the existing technological level.

This approach led to the creation of an information environment that meets the operational requirements set by top management of economic organization-wide information technology resulted in a working framework with pronounced features of heterogeneity. At one point, the flow of information in the enterprise is or will be affected by the influence of heterogeneous IT implementations. Setting up automated data flows is impossible or is achieved by converting data and manually conducted transactions affected by human error which requires high processing time.
For economic entity E_i with i=1…n, we define following sets:

- software applications sets: $AI = \{AI_{i1}, AI_{i2}, ..., AI_{in}\}$;
- data sources sets: $S = \{S_{i1}, S_{i2}, ..., S_{im}\}$.

AI set is formed by all software applications designed and implemented at a certain moment of time belonging to time sets $T = \{T_{i1}, T_{i2}, ..., T_{in}\}$. Computer applications update and extract data sets the S set. The data sources set, S consists of data sources of unstructured data, structured data, hierarchical or network databases, relational databases or object oriented. Unstructured data category may be text files, HTML files format or spreadsheet type. Structured data sources include binary of data aggregated or XML. Database systems is the approach that offers higher quality transfer mechanisms and converting data to information flow with flexible mechanisms for validation and data integrity verification.

A schematic representation of typical IT environment of economical entity E_i is given in figure 1.1.

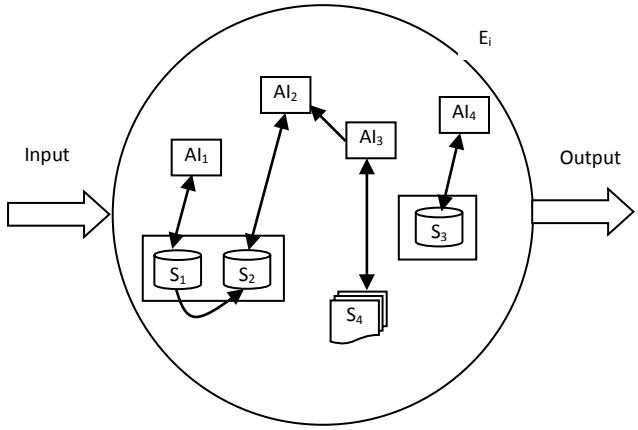

Fig. 1.1. Informational flow for E_i

Inputs to entity informational environment are processed through AI_1, AI_2, AI_3, AI_4 software applications accessing data sources S_1, S_2, S_3, S_4. Due to heterogeneous structure of computing environment difficulties appear in establishing data flows for processing input data and transactions, to provide financial values and economic indicators of the economic entity. Data is stored and processed in the data sources through independent software applications and data conversions are done by manual work procedures due to the lack of integration functionality of these applications. To minimize the degree of heterogeneity, companies must design and implement long-term strategies for their business intelligence.

2. Integration Techniques

A large number of applications developed over time running in organizations using various programming techniques and technologies have lead to the concept of reusing applications using integration processes.

Benefits of integration applications:

- introduction in the organization of new applications with high efficiency;
- changing business processes according to current requirements of the organization;
- operations and processes triggered automatically to replace the manual ones.

Whether distributed applications A_1, A_2, ..., A_n, as their data transaction data belonging to disjunctive sets. Applications run on independent servers S_1, S_2, ..., S_n. Direct integration requires building a set of applications A_1, A_2, ..., A_m selected having regard to the following:

- homogeneity of their functional structure;
- security and confidentiality of data at a similar level of confidence;
- the level of quality reliability of software is equally sensitive;
- the user interface is simple easy to use;

Integrator has an important mission, to verify that the necessary conditions are reached to achieve effective integration with high degree of utility and quality.

We have to thoroughly test the following:

- quality indicators of integrating software;
- generating sets of test data;
- clear evidence of differences in the test.

The result tests determine the selection of applications that recorded high values of characteristics. Direct integration involves:

- defining the objective of integrating applications must fulfill;
- building the interface that allows access to all software entities of distributed applications integrated;
- implementation of functional connections with these n applications;

Figure 2.1. presents a simple direct integrating scheme. The user accesses the application integrating distributed applications A_1, A_2 and A_3 through reference software interfaces made for them.
Another way to approach is the integrating by concatenation. There are k numbers of

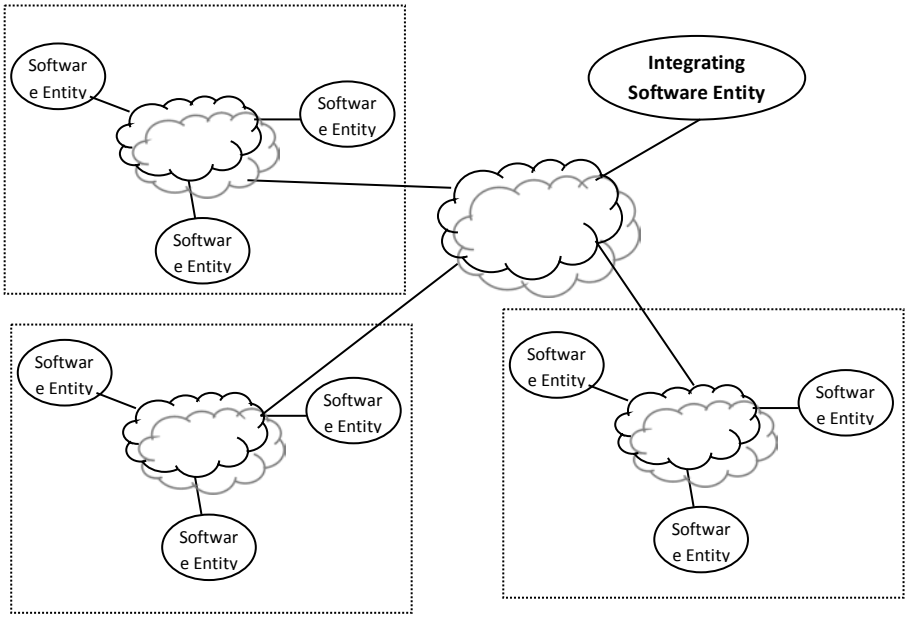

Fig. 2.1. Direct integration

applications which process the same kind of data.

Applications A_i i=1...k, process data D_{i1} , D_{i2}, D_{i3}, D_{i4}, ..., D_{in}.. Although applications of the same class each one has its specific that leads to lots of intersections of processed data to empty set of data. Integration by concatenation involves development of a software integration that incorporates data processing applications for the k applications, concatenates them, resulting in a new set of data.

E-commerce distributed applications A_i i=1...k, manages product list:

A_1: P_{11} , P_{12}, P_{13}, P_{14},...,P_{1n}, A_2:P_{21} ,P_{22},P_{23},P_{24}, ...,P_{2n}, A_k:P_{k1} , P_{k2}, P_{k3}, P_{k4}, ..., P_{kn}.

Integrating application presents his clients concatenated product list:

P_{11} , P_{12}, P_{13}, P_{14}, ..., P_{1n} || P_{21} , P_{22}, P_{23}, P_{24}, ..., P_{2n} ||… || P_{k1} , P_{k2}, P_{k3}, P_{k4}, ..., P_{kn} .

M1 : books: a, b, c, d, M2 : books : m, n, o, p, M3: books : x, y,;

Through concatenation integrated e-shop offers following books: a, b, c, d, m, n, o, p, x, y,.

Uniform access to resources is a way to integrate distributed applications at data level. It is noted that in order to create uniform access to resources transformation are to be made from a bunch of heterogeneous e-commerce shops integration is achieved through a virtual shop. E-commerce shop will have a virtual database containing virtual databases of e-commerce shops subject to integration. Extraction and data processing to build virtual

database do not access mechanisms in order to enforce restrictions of integrity of data used.

3. Integration Processes Quality

Data quality and data integration software applications tend to be combined in one because economic entities need accurate data and complex sets of data. Another reason is a software distributor must address different needs in order to compete this market.

Usually when referring to data quality, software may offer following functionality:

- data profiling, the process of analyzing data from data sources like databases, structured or unstructured files in order to gather statistics about it. These statistics may refer to search capabilities by using keywords, text descriptors, create a taxonomy or data quality metrics, measure difficulties with data integration procedure. One example of data integration is using data from different databases by joining them by useful criteria. Metadata used to describe actual data may be verified in order to see if accurate.
- data standardization refer to an environment that ensures one or a set of rules that must be enforced in a company;
- matching or linking provide mechanisms to compare similar data, identify duplicates and other links between data;
- monitoring useful for logging data quality parameters defined and send reports when margin restriction are reached, autocorrect using rules loaded into running environment;
- data cleansing, to identify and correct corrupt data from tables of databases. This is not data validation and involves correction against a list of predefined values, writing errors;

Software applications may offer all of these functions or a limited set. Data profiling tools offer functionality to define content, structure, and quality of highly complex data structures. These tools allows business users and data management staff to perform a large variety of analyses using a set of indicators, patterns and rules for each data element being analyzed or monitored. It analyzes data on an ongoing basis, and analyzes changes to source data over time to help improve data quality.
Data quality metrics may offer simple or advanced statistics applied to text string analysis, including summary data and statistical distributions of records. The patterns are preset or customized expressions that define the expected form of data analyzed and the data quality rules help define custom business thresholds and value ranges.

Strategic goals of organization is achieved through data quality level can achieve at every level starting from collecting, validation, security and integrity mechanism applied and management on transaction process. To achieve required level of quality mechanism to implement continuous monitoring of data must be in place. These mechanisms must define moment in time, the periodicity, what to data to monitor, a set of metrics to evaluate changes and a set of procedures to evaluate and asses the problem.
Integration techniques must ensure software quality system implementation to reach quality levels planned by the mechanisms for selecting applications. Selecting

applications involve defining useful quality metrics to achieve the integration objective of complex information.

Quality characteristics of distributed applications with integration needs in mind become more facets:

- integrating distributed application maintainability should ensure easy modification to architecture to allow future disposal of premises which led to approach of software functionality integrating solution, the formation of complex data processing flows to meet the needs of the company. Also implemented to ensure easy inclusion in terms of efficiency of other distributed applications to make appropriate processing streams. The basic idea is that at any stage of life cycle changes are possible to application data flows, processed data in the same economic entity or between economic entities that share data and functionality. For example processing flow of collaboration between suppliers and customers changes occur frequently. This refers to changing the terms of collaboration, inclusion of new suppliers or customers, initiating new partnerships and cooperation agreements, legislative changes. From this perspective the low values of metrics indicators has a major impact on operational status of integrating software;
- reliability is a quality characteristic affecting the data and transactions quality. Metrics for evaluating the reliability should be assessed continuously during the life cycle to identify implementation errors or dependent on external factors. Mostly decrease of reliability is due to low levels of data quality. The human factor is a cause of errors in the raw data. Legislative changes that are not implemented in time also affect the raw data. High maintainability can increase reliability through permanent monitoring;
- efficiency is the feature that ensures success of integration approach. High cost in implementing integration solutions determines whether the project is continued or not. Requiring a high level of quality leads to increased costs but must be considered also the benefits integration brings, reduced processing time, multicriterial and complex analysis of stored data are possible in real time, creating a virtual business environment possible.
- usability to enable easy learning mode of operation. Operators coming from different fields of activity, not IT experts, must be able to quickly learn how integrating environment works. Mechanisms for validation of data entry should provide clear information to operators about occurring errors and implement mechanisms for negotiating state error between component applications;
- portability under a high degree of integrated distributed applications heterogeneity of hardware and software architectures requires a comprehensive approach based on abstract interfaces that abstract proprietary implementations. Further integration with other distributed applications must be possible without affecting the operational status of the application. Scalability issues to be considered by performing connectivity solutions;
- functionality is the aim of integrating distributed applications. Integration allows you to combine the functionality offered by software applications. Integrating

application functionality is the result of joining functionality sets belonging to applications included in the chain process flows.

4. Quality Metrics for Integration Processes

In order to build indicators one must verify the extend to which every one of them meet the following:

- sensitivity, small changes in factor values to record also small variations of the indicator;
- non-catastrophic, meaning at small variations of factors not to record very large variations;
- non-compensatory, there are no situations in which factors at different levels to obtain identical values of indicators;
- representativeness, which means that the indicator values or subintervals corresponding place in correspondence with qualitative statements reflect two-way phenomenon.

Given distributed software applications $A_n = \{A_1, A_2, ..., A_n\}$. To implement integrating solution the following set of them are proposed $A_i = \{A_1, A_2, ..., A_i\}$ each one having values ranged between [0,8-1] for created indicator set $I_k = \{I_1, I_2, ..., I_k\}$. Distributed applications having values under minimal value of accepted range will not be included in A_i. I_k set includes following $I_k = \{I_t, I_e, I_s\}$where,

I_t – sets of technical indicators;
I_e – sets of economical indicators;
I_s – sets of social indicators.

I_t set includes the following indicators accuracy, reliability and maintainability, portability, modularity, stability, testability.
For I_e set proposed indicators are efficiency, complexity.
I_s set is formed by :

- safety in use, to assess the extent to which data operations managed are subject to unauthorized alteration;
- adaptability, an important indicator that helps to streamline the application investment by increasing its life;
- flexibility, evaluates restrictions on users in terms of their tolerance for mistakes;
- interoperability, to assess how an application is coupling with other applications for further use.

For each application the aggregate quality indicator is calculated. A level for this aggregate quality indicator must be chosen to generate efficiency. Out of A_n set A_i set is formed with the condition the aggregate quality indicator value must register for each application higher values than the acceptable minimum, figure 4.1.

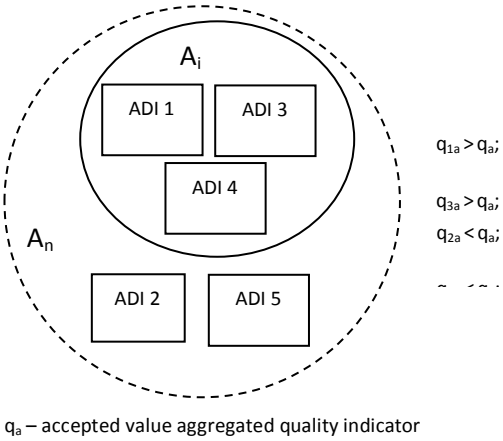

Fig. 4.1 Integrability applications candidates

Quality integration project is closely linked to the quality of software components. Low quality level of a single application will entail a low level of subjective assessment of the beneficiaries.

If modifications are done to integrating application structure to ensure comparability of the indicators:

- define the transformation coefficients;
- procedures are built to make possible the calculus of new indicators

Given customer satisfaction indicator, GS defined as the ratio between numbers of customers that bought certain products and total number of customers.

$$GS = \frac{NCF}{NTC}$$

where NCF – customers number that successfully completed a transaction from product order to receiving;

NTC – total number of customers.

Insatisfaction indicator , GI given by :

$$GI = 1 - GS - GInf$$

where GInf – represents information indicator of virtual shop visitors defined as:

$$GInf = \frac{NCInf}{NTC}$$

where NCInf – number of customers enters information pages of virtual shop.

To collect this primary data counters are implemented in the functionality of the virtual shop to record numbers of visitors entered certain pages.

5. E-Commerce Software

The way of accessing through internet virtual shops resources determines:

- customer information through referring a large number units which means resource consumption and limitation due to learning ratio of virtual shops sets;
- optimization necessity of diversity through quality and price;
- accessing one web address in order to complete one transaction comprising different products from different shops in one session.

Given following virtual shops M_1, M_2, ..., M_n, referred by addresses www.mv_1.com, www.mv_2.com, ... , www.mv_n.com, having products from classes P_1, P_2, ..., respectively, P_n . Product classes are disjoint. Virtual shops are running on web servers S_1, S_2, ..., S_n.

Direct integrating impose forming the virtual shop set M_1, M_2, ..., M_n by analyzing following:

- homogeneity of functional structure, meaning all shops have modules for product management, of valid orders, for access on-line payment systems
- quality level of their products and services is equally sensitive, meaning same warranty and delivery conditions;
- support services after acquisition to insure an efficient and useful service;
- easy to learn user interface without elements to emphasize design to esthetics ratio;
- use of selection for entering delivery and payment.

One must proceed to thorough testing procedures:

- quality metrics indicator for virtual shops, usability, functionality, reliability, consistency, response time and high customer satisfaction;
- choose a community of people with different education and different experience;
- test data sets generation;
- referring to all nodes associated with the virtual shop distributed application;

Testing results determine virtual shops set that have aggregated indicator value ranged in an accepted domain.

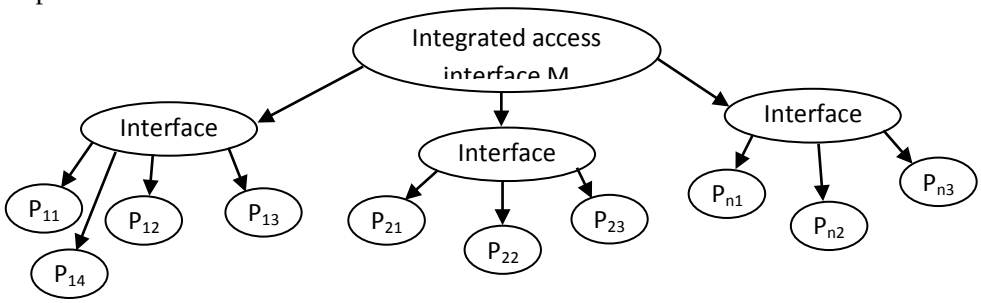

Fig.5.1. Direct integrating of 3 virtual shops

Figure 5.1. underline a simple direct integration scheme. Customer accesses the integrated shop page where you can find references to virtual stores integrated component virtual shop. The client chooses one of the shops on the existing descriptions and promotional offers available with the functionality exposed by it. Thus the buyer will see the product catalog, product basket will be based on preferences, will complete the delivery and payment dates.

6. Conclusions

Heterogeneous integrated distributed applications are a challenge of longer duration to IT specialists. The objective is to achieve integration of distributed applications in high-efficiency, cost and reduced time for implementation in conditions of high security, essential prerequisite in the virtual business environment.

Integrating e-shops becomes a necessity if one aims to close customers as wide range of products and services on a similar quality, product quality and ways to purchase and return them. Designers of technologies and applications must respond to requests recipients computer applications, users are actually those who require new trends, new approaches, all they are the ones who validate the proposed implementations.

Bibliography

[ROŞC04]	Ion Gh. ROŞCA, Cristina-Mihaela BUCUR, Carmen TIMOFTE-STANCIU – *Comerţul electronic,* Editura Economică, Bucureşti, 2004, ISBN 973-590-989-8
[IVAN09]	Ion IVAN, Leonard SĂCUIU, Daniel MILODIN - *Integrarea aplicaţiilor de comerţ electronic*, Revista Română de Informatică şi Automatică, vol. 19, nr. 2, 2009, pag. 23 - 42
[PETT08]	Christy Pettey, Holly Stevens - *Gartner Says as Number of Business Processes Using Open-Source Software Increases*, Companies Must Adopt and Enforce an OSS Policy, http://www.gartner.com/it/page.jsp?id=801412, STAMFORD, Conn., November 17, 2008
[IVAN08]	Ion IVAN, Mihai DOINEA**,** Leonard SĂCUIU – *Evaluating security of non-homogenous distributed applications*, 4th International Conference on Applied Statistics, Bucharest, November, 20 – 22, 2008, Bucharest University of Economics, National Institute of Statistics, Special Issue of Romanian Statistical Review (CNCSIS B+), ISBN 1018-046x