Collaborative and Distributed Architectures for Analysis of Economic Clusters

Cătălin BOJA, Cristian CIUREA
Department of Economic Informatics and Cybernetics
Bucharest University of Economic Studies
catalin.boja@ie.ase.ro, cristian.ciurea@ie.ase.ro

The paper describes the objectives, architecture of a collaborative distributed system used to analyze several economic indicators, as financial flows, goods and labor force movements, recorded at global and nationwide level. The results are used to validate the existence of economic clusters in some geographical areas or to indicate natural initiatives of clusters that have been sustained naturally by the existence of required economic conditions and factors without the direct intervention of public agencies. The paper describes the methodology for implementing such a project. The paper extends the idea of a single data processing and acquisition node to a distributed, collaborative system.
Keywords: *Cluster, Economy, Collaborative, Distributed, Information, System, Knowledge*

1 Economic clusters

Starting with Marshal [4],[17], many economists have proposed theoretical models that described the relation and the behavior of economic agents in different geographical and social environments. Many of these models have been defined based on observations of real environments that have proved to be successful in certain conditions. The objective of the model is to clearly define the factors, relations and components of the natural representation. One of these models is the economic cluster which has evolved from the *urban agglomeration* defined by Marshal, [17]. Economic clusters or geographic proximities of companies have many economic advantages, both for nations and private companies. These main advantages and economic benefits have been highlighted by many renowned economists [3], [4], [5], [8], [13], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]:

- reduced financial, time and transport costs;
- a larger labor pool of specialized workforce;better access to skilled employees;
- easier transfer of information and technology, directly or through spillovers;
- increase number of local factors that motivates competition between local agents;
- a greater local market for products and services; a high concentration of generate an increased market and hence more opportunities for reaching to more customers;
- facilitated access to resources;
- an attraction for new companies that see new opportunities to extend in this environment;
- provides a higher degree of specialization in products and services.

The concept of cluster is different from the classical approach of the competition policy at organization or government level as it involves analysis of "geographic concentrations of interconnected companies and institutions in the particular field" [20] that work together in such a competitive environment that increase productivity and efficiency of the cluster and thus of each member.

As the economic environment is slowly recovering from the biggest economic crisis that has hit global markets in the last 60 years, more economic analysts are trying to find the model that provides the most advantages and benefits to a local economy. In the first phase of the recovery plan world financial institutions as International

Monetary Fund, World Bank or European Union have backed up the international banking system and governments by massive infusion of funds in order to keep the public trust in these systems and to allow them to function and to implement their own contingency plans. In the current phase, governments and companies must implement plans that will use incoming financial support to rethink processes and develop own capabilities needed for future development.

Economic principles underlying the European Union (EU) common market and facilities generated by the free movement of goods and persons, partial or total removal of trade and financial barriers in the EU space have allowed the opening and the extension of economic centers, clusters, beyond a the geographical and administrative limits given by a single region or country. Thus, the single market concept is extended to the production centers which are placed over the entire European Union space.

The importance of economic clusters analysis in the EU is highlighted by the inclusion of this objective inthe European Commission policies for innovation[10], which are intended to help companies to function with better results in a competitive environment and to contribute to the overall objectives of the European Union, such as growth, employment and sustainability.

The paper describes the objectives, architecture of a collaborative distributed system used to analyze several economic indicators, as financial flows, goods and labor force movements, recorded at global and nationwide level. The results are used to validate the existence of economic clusters in some geographical areas or to indicate natural initiatives of clusters that have been sustained naturally by the existence of required economic conditions and factors without the direct intervention of public agencies. The paper describes the methodology for implementing such a project. The paper extends the idea of a single data processing and acquisition node to a distributed, collaborative system described in [6].

2 Different Architectures for the Information System

The project objective is to develop an information system for automatic analysis of clusters at Romanian level in the EU context. The analysis also aims to research the connections between clusters in EU and Romania and to identify areas in which Romania is dependent on external resources or areas in which the cluster is dependent entirely on local resources.

To achieve this objective the system will analyze the economic flows between different entities: European Union Member States, geographic regions, economic agents.

Important ongoing research projects in this area are at Harvard Business School - Institute for Strategy and Competitiveness, [11]. Based on the research have been identified profiles for more than 800 clusters in 52 countries. Each profile contains up to 120 variables, which include:

- basic descriptive data: cluster name, location and employment degree of labour;
- statistical indicators on the competitiveness of the cluster, the export growth, innovation;
- qualitative indicators regarding reasons behind the cluster and growth / decline of its competitiveness.

These profiles were collected by analyzing a vast literature devoted to the analysis and description of clusters, literature that has been created in recent years by practitioners and university researchers. The project has a predominant orientation to the American continent and Asia, mostly Japan and China.

Another important survey of economic clusters and their factors has been done by Brenner and Muhlig [7]. They have analyzed a vast literature on this topic and they have identified 35 factors classified in:

- local factors that represent the prerequisites and are present before the industrial cluster emerges. Some of these factors are represented by qualified labor, networks, universities and public research centers, tradition and culture, industrial structure, geographical location, local

demand, infrastructure, quality of life, wages.

- factors that are involved in a self-augmenting process during emergence like buyer-supplier relations, cooperation among firms, interaction with educational and research institutions, accumulation of local human capital, inter and intra industrial spillovers, spin-offs and start-up firms.
- specific events that trigger the emergence like promoting events, chance, historical events, specific innovations or leading firms.

In the European Union, it was launched in 2006, the European Cluster Observatory project [9] which is managed by the Centre for Strategy and Competitiveness (CSC) at the Stockholm School of Economics. This project is financed by the European Commission, Enterprise and Industry Directorate, through the Europe INNOVA initiative [12], [14] and the Competitiveness and Innovation Framework Programme - CIP, which aims to encourage competitiveness of European enterprises. During this project, till 2008 have been identified around 2000 clusters. At EU level, the cluster is considered an instrument of industrial policy, research and a competition and cooperation generator.

These results and researches in the field of clusters are showing that this form of economic relationship is beneficial to all parties and despite business processes are carried out at global level, the geographical proximity of partners is a real advantage.

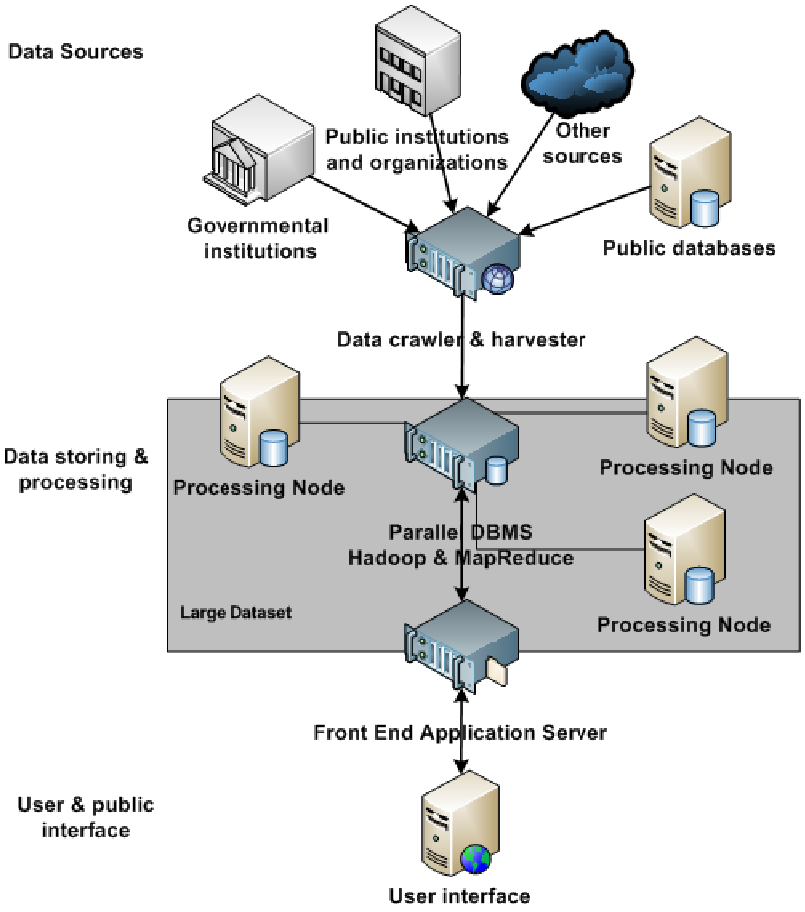

Fig. 1. Proposed architecture for a centralized system with a single data storing and processing node

The analysis is based on processing large data set of national economic reports, ministerial reports, centralized statistical data - statistical yearbooks, reports of manufacturers in various fields, reports of the Financial Administration and those from the

European Union, like Eurostat, in order to search for dependency relationships used to describe clusters. These large data volumes, requires an automated processing system that will implement effective automated data acquisition, processing and representation of these clusters.

All proposed architectures are based on three levels: data source input, data storing and processing and user interface. They differ in the technical solution used for each layer.

Figure 1 describes a proposed architecture, [28], for a centralized system with a single data storing and processing node. In order to reduce response times, the system will implement a parallel architecture for large data processing based on Hadoop framework [26], [27] and MapReducealgorithm.

The layers of the Figure 1 architecture are [28]:

- the input layer gets data from different sources, reports, data repositories and archives which are managed by governmental and public structures, economic agencies and institutions, NGO projects. The input layer is a single point of failure. Because in many cases sources use independent data schemes, the input layer may normalize them to a common format.This is an intensive data processing stage and can become a very complex one if we consider as input national data sources or crawling the Web for free data;
- the data layer stores and process large datasets of economic and financial records. This layer implements distributed, parallel processing based on Hadoop framework and Parallel Distributed Systems. The ETL intermediary layer collects data from the data crawler and harvester component, converts it in a new form and loads it in the parallel DBMS data store. The ETL normalize data, transforms it based on a predefined structure and discards not needed or inconsistent information.The objective of the layer is to normalize data and bring it to a common format, requested by the parallel DBMS.
- the user layer provides access to data and manage requests for analysis and reports.

Figure 2 describes a proposed architecture for a collaborative system which use agents, placed at the data source, to collect, filter and locally process needed data.

The proposed architecture for the information system is a collaborative one, because there are involved multiple processing nodes, institutions and organizations that are working together in order to achieve a common goal.

Collaborative information systems will represent the new generation of intelligent and auto-adaptive systems, which are encountered in many economic fields, [1], [2]. The progress of technologies leads to the development of this new generation of collaborative systems.

The information system for automatic analysis of clusters is a collaborative system that accedes at the 3C rule, meaning communication, coordination and cooperation, which cannot be achieved without the interaction between agents or system components.

The architecture of the proposed information system will be refined according with the agreements established with the institutions and organization involved in this project. If some institutions will offer incomplete data sets, the data storing and processing module must not be affected and the data crawler agent must extract data that are missing from other sources. This situation can be considered an example of an intelligent and auto-adaptive system.

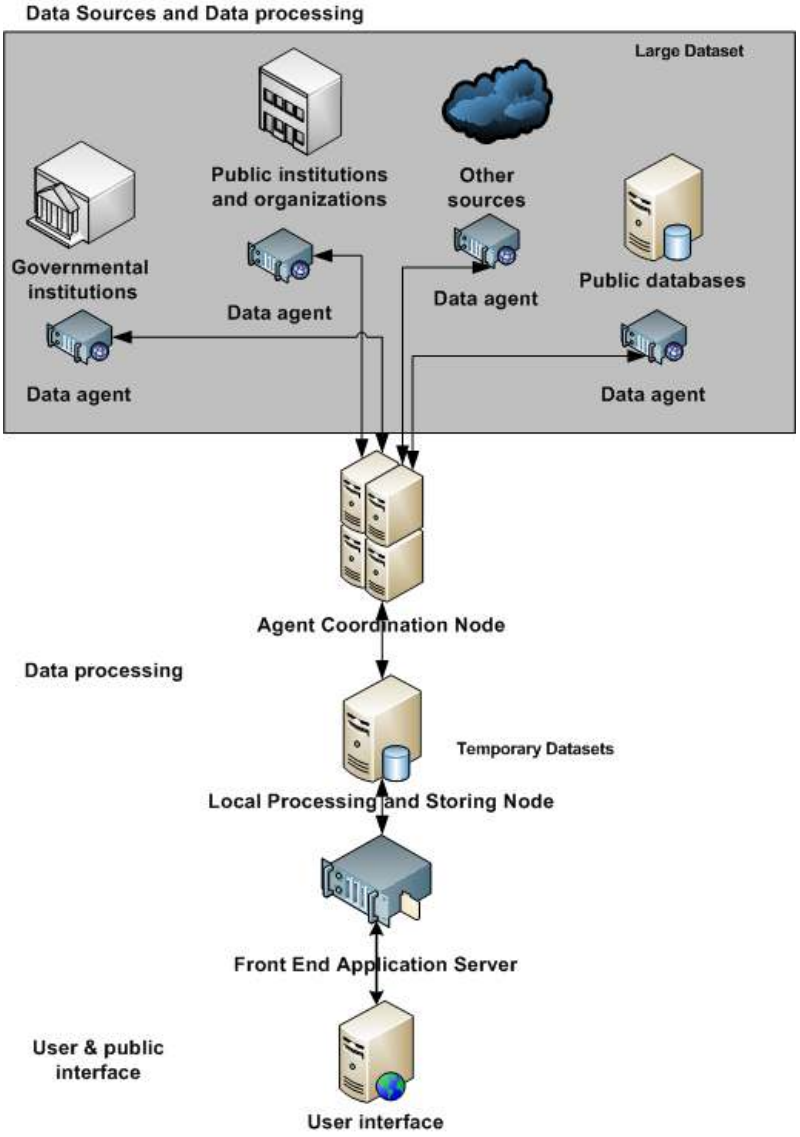

Fig. 2. Proposed architecture for a collaborative system

Figure 3 presents the main components from the architecture of the proposed collaborative auto-adaptive system.

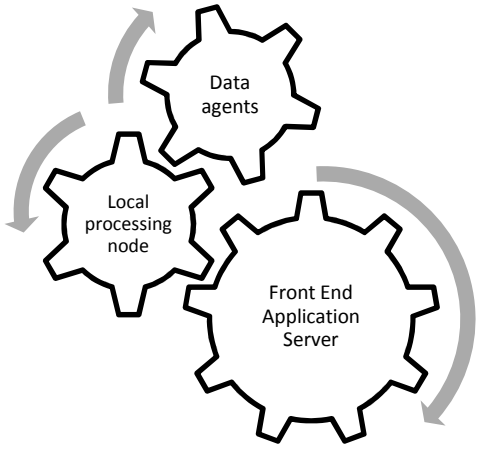

Fig. 3. Components of the collaborative auto-adaptive system

An analysis of the two proposed architectures highlights in Table 1 the advantages and difficulties generated by implementing a solution versus the other.

Table 1. Comparative analysis of the proposed architectures

Distributed system with a centralized processing node	Collaborative system with distributed processing agents
Data is obtained in a centralized manner from a single point.	Data is gathered by distributed agents placed at the data source
Data is obtained by crawling or by accessing public services or data repositories.	Data is obtained from internal sources. The agents are integrated in the source data infrastructure.
The system does not have to comply to data sources regulations as it uses public interfaces	The agents must comply with the data sources regulations regarding data access policies
Data has proprietary formats and must be normalized	The agents extract data and transform it in a predefined format
Input data is stored locally for future use. The system needs storage facilities for large volumes of data. Data is stored indefinitely.	Input data is stored locally temporarily only for the current processing session. The data is stored at the source and is transferred by the agents only when is needed.
The input layer gathers a wide range of data because can be used in further processing sessions.	The agents collect and transfer data for the current processing task.
Data is processed only at the second layer	Data can be processed at the input layer by agents
The processing node implements a parallel distributed system for managing large volumes of data	The processing node processes small amounts of data because it is filtered and delivered by agents
The system uses an Extract, Transform and Load (ETL) component	The ETL tasks are done at the data source by the agents

As a side effect of the system objectives, it can be used to refine existing economic models or to define new ones that allow a better qualitative analysis ofeconomic clusters. The development and implementation of the information system will generate a tool used to assist cluster analysis by automatic processing large volumes of data and by providing more suggestive visual representationsfor the results.

Testing methodologies and analysis of existing models, such as Harvard model, can lead to its refinement. Automation of the analysis may lead to its expansion by including a larger number of factors. The system Web platform will provide public access to economic flows and clusters in different regions.

Although in Romania there are many initiatives to generate clusters in various economic fields, like the FP7 program, in the economic literature they are underrepresented. The objectives of this project may become an initial interface to these local economic models.

3 Methodology

The project is implemented in two major phases, described by Figure 4.

The first phase of the methodology is centered on the economic approach of clusters:

- it is studied the economic literature in order to indentify qualitative and quantitative characteristics that describe a cluster;
- the characteristics set is used to define a model or to complete an existing one in order to use it to analyze existing clusters;
- the economic model is tested to help identify premises of a cluster initiative based on financial data;
- the project requires an analysis of available economic data sources; these data will be used to validate the model and to provide a quantitative image of the cluster;
- there are identified important economic influence factors.

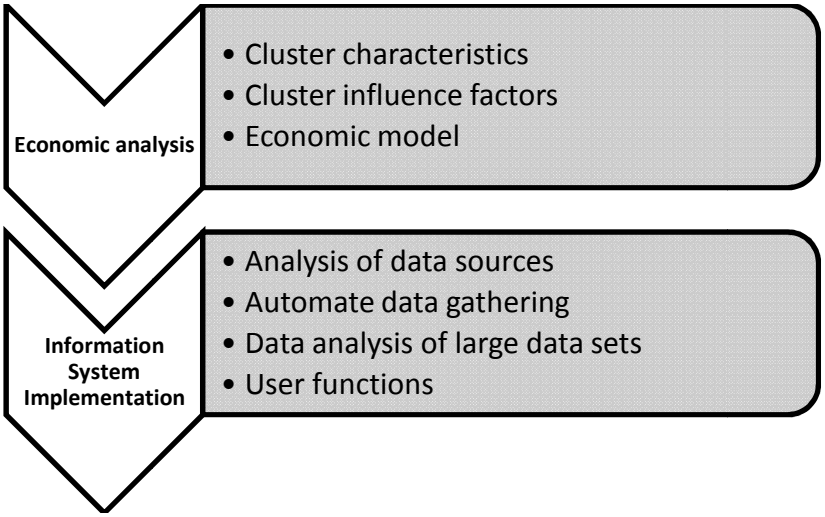

Fig. 4. Stages in defining and implementing the information system

The second phase is the development of the information system – the software solution.

- the objective is to automate the analysis stage and to provide software tools for visual representation of clusters using GIS, Adobe Flex or Google/Bing maps API.
- the system will also managed and process large data sets acquired from EuroStat and other financial institutions.

Another important phase in project implementation is the testing of the information system. This is done using test automation tools in order to calculate some metrics and determine some measurements. The testing phase of the project is a part of the quality assurance process and is realized in the followings steps:

- test planning and strategy;
- analysis, design and test case development;
- test preparation and execution;
- results analysis;
- test summary reports.

The objective of the testing phase is to increase the quality of the information system by test automation using programming languages as Ruby and specific testing tools.

4 Conclusions

The fast integration of information systems in all economic and social fields and the current state of technology infrastructure for large data storage, processing and network transfer has allowed organizations to record everything. This advantage has become in short time the current obstacle in getting consistent information regarding the economic phenomenon which it describes. The two proposed architectures describe solutions used to gather information from different sources and to process large amount

of data either by using a high capacity single processing node or by a large collaborative and distributed system.

The data generated by this systemis used by:

- government institutions to provide legal and economic support to private companies to promote further development of the cluster;
- companies to define /change competitive policies in order to join existing clusters or to identify areas where future clusters can be formed;
- effect analysis of clusters on the labor market, human capital and on policies for development and competitiveness.

Integration of the results obtained by the collaborative information system help to automate current operations carried out by the system, but also to provide strategic, tactical and operational information required in the decision-making process.

Acknowledgement

This work was supported from the European Social Fund through Sectored Operational Programmer Human Resources Development 2007-2013, project number POSDRU/89/1.5/S/59184, "Performance and excellence in postdoctoral research in Romanian economics science domain".

References

[1] L. Bătăgan, "Smart Cities and Sustainability Models," *Informatica Economica Journal* , Vol.15, No.3, 2011, ISSN 1453-1305, B+, Available at: http://www.revistaie.ase.ro/content/59/07%20-%20Batagan.pdf

[2] L. Bătăgan, "Intelligent Solution for Future Development,"*Journal of Applied Quantitative Methods*, Vol. 6, Issue 1, March 30, 2011, ISSN 1842-4562, B+, Available at: http://www.jaqm.ro/issues/volume-6,issue-1/pdfs/8_batagan.pdf

[3] R. Baptista and P. Swann, "Do firms in clusters innovate more?", *Research Policy*, Vol. 27, No. 5, 1998, pp. 525–540.

[4] G. Becattini, "From Marshall's to the ''Italian industrial districts''. A brief critical reconstruction,"2001,Available at: http://www.tci-network.org/media/asset_publics/resources/000/000/685/original/becattini_marshall.pdf.

[5] W. Bönte, "Innovation and Employment Growth inIndustrial Clusters: Evidence fromAeronautical Firms in Germany", ERSA conference papers, ERSA03, European Regional Science Association, 2003, Available at: http://www-sre.wu-wien.ac.at/ersa/ersaconfs/ersa03/cdrom/papers/284.pdf

[6] C. Boja, "Information System for Economic Cluster Analysis", Proceedings of the 11th International Conference on INFORMATICS in ECONOMY (IE 2012), "Education, Research & Business Technologies", Bucharest, Romania, May 10-11, 2012, ISSN 2284-7472, ISSN-L = 2247-1480, pp. 534 – 538.

[7] T. Brenner, A. Mühlig, "Factors and Mechanisms Causing the Emergence of Local Industrial Clusters - A Meta-Study of 159 Cases,"*Papers on Economics and Evolution*, 2004, ISSN 1430-4716.

[8] H. Etzkowitz, "The Triple Helix of University-Industry-Government: Implications for Policy and Evaluation", *Working Paper 2002: 11*, Science Policy Institute, 2002.

[9] European Cluster Observatory, 2006, Available at: www.clusterobservatory.eu, (Accessed 23/02/12)

[10] European Comission, 2012, "Industrial innovation, Innovation Policy", Available at: http://ec.europa.eu/enterprise/policies/innovation/policy/index_en.htm(Accessed 23/02/12)

[11] Harvard Business School - Institute for Strategy and Competitiveness, 2011, "The International Cluster Competitiveness Project," Available at: http://www.isc.hbs.edu/econ-clusters.htm, (Accessed 20/01/12)

[12] INNOVA Project, 2008 - European Cluster Organisation Directory, Available at: www.europe-innova.eu, (Accessed 5/02/12)

[13] P. Krugman, *Geography and trade*, London: MIT Press/Leuven UP, 1991, pg.142.
[14] C. Ketels, O. Solvell,„Innovation Clusters in The 10 New Member States of The European Union", *European Commission, Europe INNOVA paper N° 1*, 2006, ISBN 92-79-03196-1.
[15] A. Malmberg and P. Maskell, "The elusive concept of localization economies – Towards a Knowledge-based Theory of Spatial Clustering,"*AAG Annual Conference*, New York, 27 February – 3 March 2001.
[16] A. Malmberg, Ö. Sölvell and I. Zander,"Spatial clustering, local accumulation of knowledge and firm competitiveness," *Geografiskaannaler*, Series B, Human Geography, Vol. 78, No. 2, 1996, pp. 85-97.
[17] A. Marshall, *Principles of Economics, 8th edition*, London: Macmillan, 1920.
[18] P. Morosini, "Industrial Clusters, Knowledge Integration and Performance,"*World Development*, Vol. 32, No. 2, 2004, pp. 305-326.
[19] M. E. Porter, *The Competitive Advantage of Nations*, Macmillan, London, 1990.
[20] M. E. Porter, "Clusters and the new economics of competition,"*Harvard Business Review*, Vol. 76, No. 6, 1998, pp. 77–90.
[21] M. E. Porter, "Location, competition, and economic development: Local clusters in a global economy,"*Economic Development Quarterly*, Vol.14, No. 1, 2000, pp. 15-34.
[22] A. Saxenian, *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Harvard University Press, 1996, Available at: http://www.google.com/books
[23] F. Sforzi, "The industrial district and the 'new' Italian economic geography", *European Planning Studie*s, Vol. 10, No. 4, 2002.
[24] Ö. Sölvell, G. Lindqvist and C. Ketels, *The Cluster Initiative Greenbook*, Stockholm, 2003.
[25] Ö.Sölvell, *Clusters, Balancing Evolutionary and Constructive Forces*, Ivory Tower Publishers, Stockholm,2008, ISBN 978-91-974783-3-5.
[26] C. Toma, "Practical Results Using Apache Hadoop Platform forDistributed and Parallel Computing", *Proceedings of the 11th International Conference on INFORMATICS in ECONOMY (IE 2012), "Education, Research & Business Technologies"*, Bucharest, Romania, May 10-11, 2012, ISSN 2284-7472, ISSN-L = 2247-1480, pp. 30 – 35.
[27] T. White, *Hadoop: The Definitive Guide*, O'Reilly, 2009.
[28] A. Pocovnicu, C. Boja and L. Batagan, "Distributed Parallel Architecture for "Big Data"," *Informatica Economica Journal*, Vol. 16, No. 2, 2012, pp. 116 – 127, ISSN 1453-1305.

Cătălin BOJA is Lecturer at the Economic Informatics Department at the Academy of Economic Studies in Bucharest, Romania. In June 2004 he has graduated the Faculty of Cybernetics, Statistics and Economic Informatics at the Academy of Economic Studies in Bucharest. In March 2006 he has graduated the Informatics Project Management Master program organized by the Academy of Economic Studies of Bucharest. He is a team member in various undergoing university research projects where he applied most of his project management knowledge. Also he has received a type D IPMA certification in project management from Romanian Project Management Association which is partner of the IPMA organization. He is the author of more than 40 journal articles and scientific presentations at conferences. His work focuses on the analysis of data structures, assembler and high level programming languages. He is currently holding a PhD degree on software optimization and on improvement of software applications performance.

Cristian CIUREA has a background in computer science and is interested in collaborative systems related issues. He has graduated the Faculty of Economic Cybernetics, Statistics and Informatics from the Bucharest Academy of Economic Studies in 2007. He has a master in Informatics Project Management (2010) and a PhD in Economic Informatics (2011) from the Academy of Economic Studies. Other fields of interest include software metrics, data structures, object oriented programming in C++, windows applications programming in C# and mobile devices programming in Java.