Cloud Patterns for mOSAIC-Enabled Scientific Applications

Teodor-Florin Fortiş[1,2], Gorka Esnal Lopez[3], Imanol Padillo Cruz[3], Gábor Ferschl[4], and Tamás Máhr[4]

[1] Institute e-Austria, Timişoara, Romania
[2] West University of Timişoara, Romania
fortis@info.uvt.ro
[3] Tecnalia (Industrial Systems Unit), San Sebastián, Spain
{gorka.esnal,imanol.padillo}@tecnalia.com
[4] AITIA International, Inc., Budapest, Hungary
{gferschl,tmahr}@aitia.ai

Abstract. Cloud computing has a huge potential to change the way data- and computing-intensive applications are performing computations. These specific categories of applications raise different concerns and issues that can be bypassed by identifying relevant reusable cloud computing patterns, on the top of specific cloud computing use cases. Development of new cloud patterns will help offering a better support for the development and deployment of scientific distributed application over a cloud infrastructure.

Keywords: Cloud computing, cloud use cases, cloud patterns.

1 Introduction

Cloud computing, with its different deployment and service models, has a huge potential to change the way we are performing computations, by moving it from a traditional model to a service-based model – utility computing –, as a next generation of commodity computing.

The National Institute of Standards and Technology (NIST) [13] offered a very clear and simple definition of cloud computing , as "*a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources [...] that can be rapidly provisioned and released with minimal management effort or service provider interaction.*" A similar point of view was expressed in a white paper from the UC Berkeley RAD Labs [1] : "*cloud computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services.*"

However, in Berkley RAD Lab's white paper a separation between utility computing and private clouds was identified. As a consequence, provided that cloud computing can be identified as "*the sum of SaaS and utility computing*" [1], the same kind of separation exists between cloud computing and private clouds.

M. Alexander et al. (Eds.): Euro-Par 2011 Workshops, Part I, LNCS 7155, pp. 83–93, 2012.

With a slightly different point of view, G. Perry [14] specified that "*cloud computing is a broader concept than utility computing and relates to the underlying architecture in which the services are designed*", that can be applied both to utility computing and internal corporate data centers.

With a growing interest in cloud computing, there are a series of research challenges, as identified in [19], that still require specific attention from the research communities. These challenges include *automated service provisioning*, *data security*, *software frameworks*, *storage technologies and data management*, or *novel cloud architectures*. In addition, having different approaches for the various setting of cloud computing, several important issues, like *open standards interfaces*, *opportunity of service level agreements* (SLA) for service delivery, *security in different service delivery models* [18], or *commonly accepted programming models* [12], still need to be addressed.

The mOSAIC project[1], developed in the framework of the FP7-ICT programme, is addressing a series of the issues and research challenges related with cloud computing, as specified above, to deliver an open-source platform that enables applications to negotiate cloud services as they are requested by their respective users, and to develop an open-source cloud API, by which applications will be able to specify and communicate their requirements to the platform.

Special attention is paid to the identification of a set o reusable patterns and use cases relevant to the selected set of mOSAIC-enabled applications [15], [11], that will be used both for validating the core of the mOSAIC platform [6], and the different mOSAIC specific components. The set of mOSAIC specific components include a semantic engine and a cloud agency [2], that will help applications to have a better response in relation with their requirements in terms of cloud resources, service level agreements (SLA) or quality of service.

2 Cloud Use Cases and Cloud Patterns

Different efforts were performed in order to identify cloud computing patterns. In [10] several sets of patterns were described, closely related with the capabilities offered by the Azure platform. Different categories of patterns were identified, including computing (on-demand application instance, worker patterns), storage (blob storage, structured storage patterns), communication (service interface, service-oriented integration, and messaging patterns), and administration (cloud deployment, design for operations, service instance management, managements alerts, and service level management patterns). A different approach was offered by a Sun presentation [17], where a set of software and infrastructure patterns were identified, in relation with cloud computing[2].

The different approaches for data- and computing-intensive applications (see [12], [16], [4], [9], [7]) identified different concerns and issues related with cloud computing. On the top of these issues and concerns, a common point of view

[1] Detailed information available on project's web site: `http://www.mosaic-cloud.eu`
[2] Also available on SUN wiki: `http://wikis.sun.com/display/cloud/Patterns`

can be developed by identifying and specifying a common set of reusable cloud computing patterns, and the set of complementary cloud computing use cases.

2.1 Cloud Computing Use Cases

The Cloud Computing Use Cases Group defined a set of use case scenarios for cloud computing [5], by using both the experience of cloud vendors, and cloud consumers. The document offers a consumer-oriented view of cloud usage, and highlights several concerns related with interoperability, standards and standardization, security and SLA. Seven core cloud usage scenarios were identified, and the specific interaction concerns related with these usage scenarios were exposed: *end user to cloud*, *enterprise to cloud to end user*, *enterprise to cloud*, *enterprise to cloud to enterprise*, *private cloud*, *changing cloud vendors*, and *hybrid cloud* [5, pp. 18-19]. Also, Cloud Computing Use Cases Group provided an uniform view of the relationships between cloud requirements and specific use cases, as well as a set of developer requirements for cloud-based applications.

NIST, with the Standards Acceleration to Jumpstart Adoption of Cloud Computing (SAJACC) initiative, developed a set of cloud use cases [3], by combining the experience from industry, government agencies and academia. The efforts of the SAJACC initiative were oriented to "*facilitate Standards Development Organizations in their efforts to develop high-quality standards*", addressing the important needs identified, on the basis of a pool of use case scenarios that were developed to better support portability, interoperability and security.

With the document [7], the Distributed Management Task Force (DMTF), via its Open Cloud Standards Incubator, offered the basis for a reference architecture for managing clouds, and identified specific requirements oriented to management, security and resource models. Different interaction patterns were identified along with this reference architecture, covering four larger categories. Based on the set of interaction patterns, a distinct document was offered for a uniform view of cloud management use cases [8]. These use cases were aligned with NIST definitions for cloud computing [13], offering a distinctive view on cloud management, security and cloud government requirements and concerns.

3 Usage Scenario: A mOSAIC Perspective

Different data- and computing-intensive applications were considered for mOSAIC implementation, as proof-of-the-concept applications. Cloud usage patterns and cloud use cases were identified after a detailed analysis of these applications, starting from typical usage scenarios, as in [15]. Considered data-intensive applications, include applications used for process optimization and maintenance in different industrial sectors, storage and data distribution, fast data access for crisis situations, or running simulations on the cloud, with different configured simulation scenario. The area of usage cases covered by the minimal set of mOSAIC applications range from virtual storage, data distribution, peak in processing resource demand, fast data access for crisis situations,

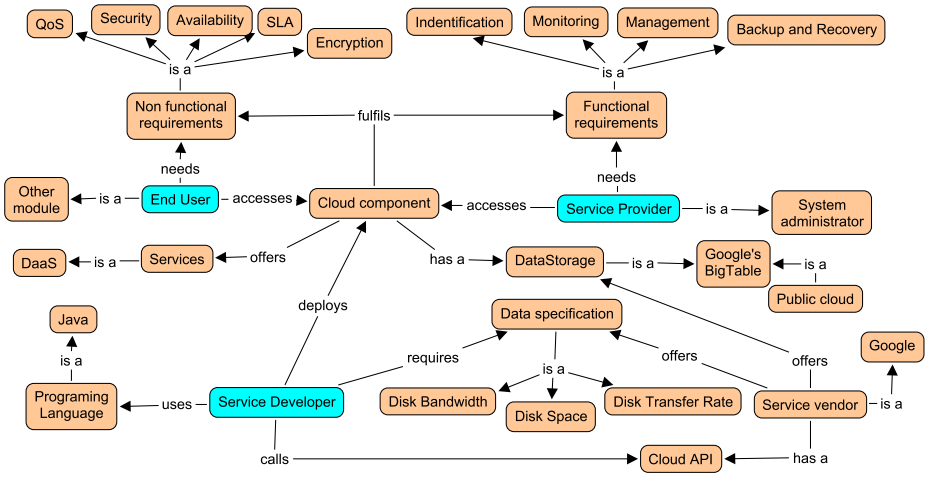


Fig. 1. Ontological view: data storage and data access (data intensive, [11])

data-intensive telemetry scenario, on-demand and distributed storage allocation, balanced computing, to computing-intensive scenario.

Data Intensive Applications. The Intelligent Maintenance System (IMS) allows the maintenance of devices from different industrial scenarios (wind turbines, airplanes, machine tools, etc) through early diagnosis of faults in critical components and real-time monitoring of key variables. It uses specific Artificial Intelligence techniques that permit modelling the experience of expert technical people and reason about it to support IMS in making decisions about maintenance actions and advance warning of critical situations.

IMS is associated with a SLA template customized to its needs. The system automatically acquires data from distributed sensors and stores them into a cloud-based datastore. Once data is acquired, a knowledge extraction process over raw data is executed, using a cloud-based infrastructure, in order to get enough information that will assist in the detection process of potentially critical situations. These diagnoses are also stored in a cloud-based datastore.

After the data has been stored properly in a cloud storage system, registered users have the ability to query these data. For this purpose, users must be properly registered in the system and they must have sufficient permissions. Users who meet these requirements can access stored data (as raw data storage and diagnoses) in an asynchronous way, i.e. while the system continues storing new data and generating new diagnoses.

For this specific application type several usage scenarios are possible, on the basis of which one can identify the set of specific requirements and relationships, as depicted in Figure 1, and described in [11].

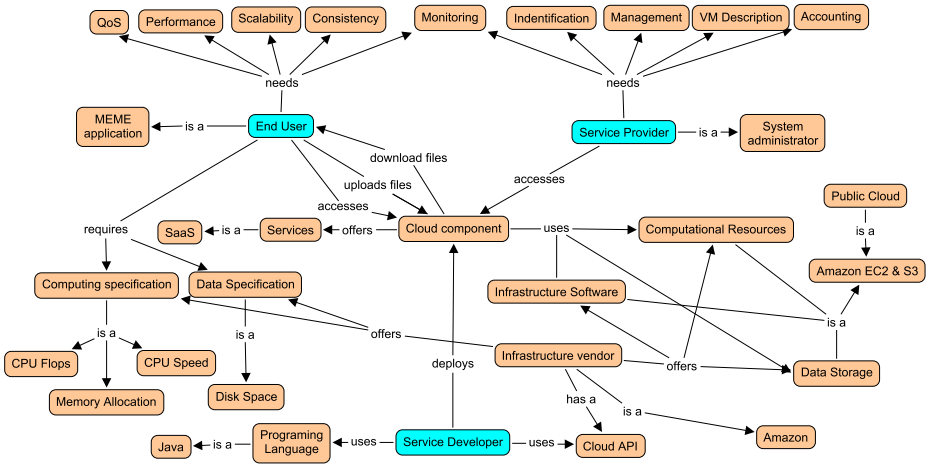


Fig. 2. Ontological view: distributed parameter sweep (computing intensive, [11])

Computing Intensive Applications. The Model Exploration Portal (MEP) is a service and website that enables researchers to run social simulation models in the cloud. Various user needs can be satisfied when renting computational resources for the users. They can run agent-based simulations in an 'as-fast-as-possible' mode by acquiring as many resources as needed. Alternatively, in a so called 'best-effort' mode, it is possible to use the spare resources in the system which become available when a simulation finishes without using its resources to the full time. Finally, users may set a deadline for the simulations, and the system will automatically determine the proper operational mode. Users can monitor the progress of their simulations and modify the SLA online. The development of agent-based models requires repeated experimentation with the model, which in turn means that a large amount of simulations have to be run.

One of the typical usage is to run *parameter-sweep* experiments, where the parameter space of the model is explored to get an overview of the general behavior. This particular usage scenario was fully described in [11], and the set of identifiable requirements and relationships was specified as in Figure 2.

4 Execution Scenarios for mOSAIC-Enabled Applications

Based on the initial analysis of the data- and computing-intensive applications (implementing simulation scenarios and intelligent maintenance), a series of requirements were identified for further implementation of SLA, and for implementing specific cloud policies and constraints. These requirements are reflecting both the functional and non-functional requirements that are coming from the analyzed applications, see [11], and express a wide range of cloud-specific concerns.

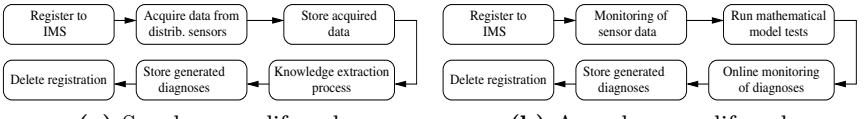
Functional requirements that were identified in [11], and depicted in Figure 1 and Figure 2, include the following: (1) consistency, (2) identifications, (3) monitoring, (3) backup and recovery, (4) replication, (5) management, (6) accounting, (7) virtual machine descriptions, (8) encryption. Non-functional properties include (1) availability, (2) QoS, (3) communication (non-functional part), (4) computing (non-functional part), (5) security, (6) scalability, (7) performance, (8) autonomous, (9) data (non-functional part), (10) reliability.

Starting from the set of functional and non-functional properties, and complemented by the discussion about SLA from Cloud Computing Use Cases Group's document [5, p. 54], a set of basic SLA requirements was identified, including, but not limited to: (1) compliance, (2) transparency, (3) security, (4) metric, (5) privacy, (6) auditability, (7) data encryption, (8) monitoring, (9) certification.

4.1 Intelligent Maintenance System Execution Scenarios

On the top of the identified use cases for an Intelligent Maintenance System (IMS), several execution scenario are possible for an IMS-enabled application. These execution scenarios respond to specific requirements from application point of view.

IMS Synchronous Execution Scenario. In this scenario data acquired from distributed sources (sensors) by the IMS system is automatically stored. A knowledge extraction process over these acquired raw data is executed, in order to get information that will help to detect possible critical situations. Storage of these diagnoses is subsequently performed. The entire process is carried out in synchronous mode, without any interaction required from expert or specialized user to the system.



(a) Synchronous lifecycle **(b)** Asynchronous lifecycle

Fig. 3. IMS execution scenarios

The implementation of this scenario is based on the existence of several use cases, as specified in Figure 3a, including *data acquisition from distributed sources*; *storing acquired data from distributed sources*; *running knowledge extraction processes*; *storing generated diagnoses*.

IMS Asynchronous Execution Scenario. This scenario is applicable to a situation where expert or specialized users access generated data (e.g. reports, statistics) by using a specialized web interface. Moreover, expert users have the ability to access independent applications through which they have the possibility to configure (change, or add) the rules on which the intelligent system is based. Additionally, the ability to deploy algorithms is ensured, such that the

newly configured rules can be used. Different use cases, including *IMS registration*; *online monitoring of data from sources*; *running mathematical model tests*; *online monitoring of diagnoses*; *storing generated diagnoses*; *registration deletion*, are used for this execution scenario, as identified in Figure 3b.

4.2 Model Exploration and Service Level Agreement Execution Scenarios

With the model exploration series of execution scenarios, different usages are described for this system, including a model exploration portal (MEP) as an administrative cloud application, the model exploration service (MES), which enables distributed simulations and collection of simulation results, the model exploration management in cloud (MEMiC), which is responsible with provisioning of cloud resources, including negotiation of a service level agreement (SLA) with the cloud providers, resource acquisition and monitoring, and the model exploration module (MEME), a system used to run agent based simulations for building experiment's designs.

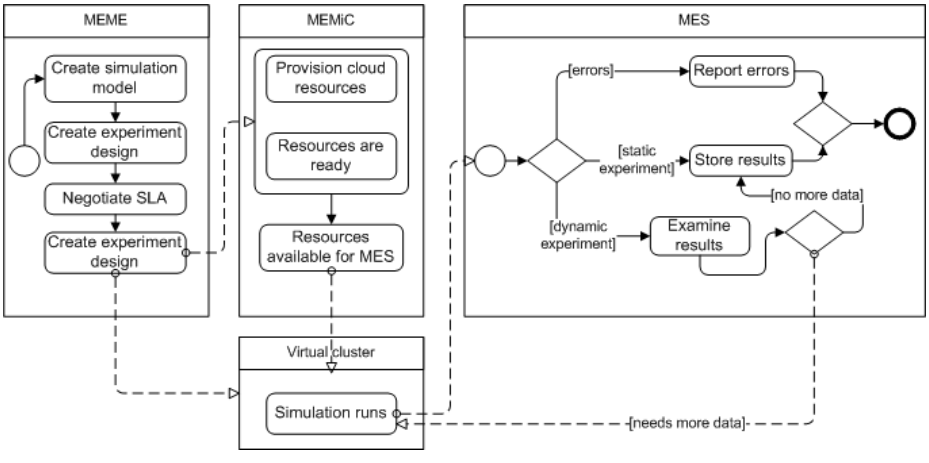


Fig. 4. Simulation execution scenario: activity workflow

Simulation Execution Scenario. The simulation execution scenario is used for conducting both static and dynamic simulation experiments. In both situation a simulation model was identified, and an analysis of this model is to be performed by running it with various settings, in order to follow its behavior. For a dynamic experiment design, an initial set of parameter settings will be run, and based on results analysis new sets of parameter settings will be identified and run, until a stopping condition was met. On the other hand, in the case of static experiment design, there is only one set of parameter settings that will be run, then the simulation is ended. Different use cases, including *creating a simulation model*;

setting up model exploration; *choosing an SLA template*; *contracting an SLA*; *provisioning cloud resources*; *running, monitoring and simulation maintenance,* are used for this execution scenario.

SLA Execution Scenario. This complex scenario is in relation with the model exploration module (MEME) and the model exploration management in cloud (MEMiC). The SLA is being negotiated with MEME, and is based on interactions with MEMiC for establishing SLA options. The scenario offers the necessary support for SLA preparation, enabling SLA, as well as modifying the SLA. However, this execution scenario depends on the capability of the MES provider to define and modify SLA templates in MEMiC, and to offer the necessary support for negotiations.

4.3 Cloud Patterns for mOSAIC-Enabled Applications

On the top of identified use cases and typical execution scenario, a set of cloud patterns can be extracted in order to support the development of mOSAIC-enabled applications. The identified cloud patterns could be categorized, following category description, as in [7] and [10]. The common identified patterns, together with additional pattern categories and applicable security federation pattern categories, as specified in [5], are as follows.

Establish Identity (identify [7, p.18]**).** Establish the identity of accessing user of a scientific application. Different security policies, including *trust*; *identity management*; *single sign-on/sign-off* [5, pp.46-47]; can be used in the implementation of this pattern.

Establish SLA (identify). A complementary pattern for establishing identity. An SLA contract was established for the accessing entity and the cloud provider for accessed services. The 'Establish SLA' pattern is responsible with enforcing the established SLA. Security policies and federation patterns, like *trust*; *access management*; *audit and compliance* [5, pp.46-47]; as well as cryptography-related security controls, are applicable for this pattern.

Browse Available Offerings (administer [7, p.18]**).** A generic pattern, usable for browsing offerings, related with available (e.g. sensor) data collections; available data storage facilities; available components for monitoring. The *configuration management* security federation pattern [5, p.47] can be used in the implementation of this pattern.

Establish Parameters (administer). Parameters, as metadata of accessed scientific services, can be specified for different aspects (e.g. processes) of sample mOSAIC applications: parameters for simulations, parameters for knowledge extraction, parameters for data mining, monitoring parameters. Usable security federation patterns include *trust* and *access management* [5, pp.46-47].

Update SLA (administer). Updating a previously generated SLA to better respond to requirements of the accessing user. The mOSAIC SLA Generation Tool (MSGT), used for the generation of the SLA, could be used for this pattern, too.

Same security federation patterns as for the 'Establish SLA' pattern could be used in the implementation of this pattern.

Obtaining Metadata of Provisioned Services (administer). Different metadata of provisioned services, like parameters, SLA information, or diagnoses results are the subject of this pattern. The security federation pattern usable for implementing this pattern is *configuration management* [5, p.47].

Online Adjustment and Negotiation of Resources (deploy and update [7, p.19]**).** Adjustment and negotiation of resources are necessary when cloud resources initially allocated to the execution of the scientific processes (e.g. simulation) may not fulfill anymore the accessing entity's requirements (e.g. established SLA). An automatic monitoring of the scientific process (e.g. simulation) and online adjustments of cloud resource allocation are performed. *Access management*; *audit and compliance*; and *configuration management* [5, pp.46-47] are the security policies applicable for this situation.

Provision Scientific Process (deploy and update [7, p.19]**).** Scientific processes require access to large quantities of data: sensor data, simulation data, satellite images, as well as large quantities of resources. This pattern is responsible with uploading of selected data for the execution of scientific processes, as well as provisioning of resources that satisfies established SLAs. A set of security policies similar with those specified for the 'Update SLA' and 'Online adjustment and negotiation of resources' patterns, are usable for this case.

Data Storage and Aggregation (steady state). A generic cloud pattern, for storing data outputs from scientific application, like results of knowledge extraction process, of simulation results. Aggregated and raw data are the subject of this identified pattern. While *access management* is the basic security federation pattern, it can come into effect accompanied by several security controls, like *data/storage security* [5, pp.44-46].

Monitoring and Notification of Data Models (steady state). This pattern is related with execution of the knowledge extraction process; execution of simulations for adjusting provisioned resources, or job distribution. *Audit and compliance*; *configuration management*; *access management* [5, pp.46-47] are the security policies usable for the implementation of this pattern.

5 Conclusions

The cloud offers an ideal environment for developing and deploying of data- and computing-intensive scientific applications. These applications have to use complex mechanisms for assuring SLA, or QoS, like negotiation, brokering or provisioning, in order to maximize the utilization of a cloud infrastructure.

Typical usage of these mechanisms could be captured by the specification of a set of cloud patterns, oriented to scientific applications, in addition to previously described patterns, as in [10] and [17], and offer the basis for the development of powerful reusable building blocks for cloud-enabled applications.

Identification of the new cloud patterns was based on data- and computing-intensive applications, that are considered as proof-of-the-concept mOSAIC applications. Different execution scenarios were considered for the identification of specific cloud use cases, and specification of corresponding cloud patterns. By putting the newly identified patterns in relation with already specified pattern categories, and linking them with interaction patterns, security federation patterns, as well as with security controls, a high level of reusability for the set of considered cloud patterns is achieved.

Acknowledgments. This research was partially supported by the grant FP7-ICT-2009-5-256910 (mOSAIC).

References

1. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R.H., Konwinski, A., Lee, G., Patterson, D.A., Rabkin, A., Stoica, I., Zaharia, M.: Above the clouds: A Berkeley view of cloud computing (February 2009), http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html
2. Aversa, R., di Martino, B., Rak, M., Venticinque, S.: Cloud agency: A mobile agent based cloud system. In: Barolli, L., Xhafa, F., Vitabile, S., Hsu, H.H. (eds.) CISIS, pp. 132–137. IEEE Computer Society (2010)
3. Badger, L., Bohn, R., Chandramouli, R., Grance, T., Karygiannis, T., Patt-Corner, R., Voas, J.: Cloud computing use cases (October 2010), http://www.nist.gov/itl/cloud/use-cases.cfm
4. Beloglazov, A., Buyya, R., Lee, Y.C., Zomaya, A.: A taxonomy and survey of energy-efficient data centers and cloud computing systems. Advances in Computers 82, 47–111 (2011)
5. Cloud Computing Use Cases Group: Cloud computing use cases white paper (July 2010), http://opencloudmanifesto.org/Cloud_Computing_Use_Cases_Whitepaper-4_0.pdf
6. Di Martino, B., Petcu, D., Cossu, R., Goncalves, P., Máhr, T., Loichate, M.: Building a Mosaic of Clouds. In: Guarracino, M.R., Vivien, F., Träff, J.L., Cannatoro, M., Danelutto, M., Hast, A., Perla, F., Knüpfer, A., Di Martino, B., Alexander, M. (eds.) Euro-Par-Workshop 2010. LNCS, vol. 6586, pp. 571–578. Springer, Heidelberg (2011)
7. DMTF: Architecture for managing clouds (June 2010), http://dmtf.org/sites/default/files/standards/documents/DSP-IS0102-1.0.0.pdf
8. DMTF: Use cases and interactions for managing clouds (June 2010), http://www.dmtf.org/sites/default/files/standards/documents/DSP-IS0103_1.0.0.pdf
9. Guo, W., Gong, J., Jiang, W., Liu, Y., She, B.: OpenRS-cloud: A remote sensing image processing platform based on cloud computing environment. SCIENCE CHINA Technological Sciences 53, 221–230 (2010)
10. Joseph, J.: Patterns for high availability, scalability, and computing power with Windows Azure. MSDN Magazine (May 2009)
11. Lazkanotegi, I., Esnal, G.: Cloud usage patterns (March 2011), http://www.mosaiccloud.eu/dissemination/deliverables/FP7-256910-D3.1-1.0.pdf

12. Malawski, M., Meizner, J., Bubak, M., Gepner, P.: Component approach to computational applications on clouds. Procedia Computer Science 4, 432–441 (2011)
13. Mell, P., Granc, T.: The NIST definition of cloud computing (January 2011), http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf
14. Perry, G.: How cloud and utility computing are different (February 2008), http://gigaom.com/2008/02/28/how-cloud-utility-computing-are-different/
15. Petcu, D.: Identifying cloud computing usage patterns. In: Proceedings of Cluster 2010, pp. 1–4. IEEE Computer Society (September 2010)
16. Rafique, M.M., Butt, A.R., Nikolopoulos, D.S.: A capabilities-aware framework for using computational accelerators in data-intensive computing. Journal of Parallel and Distributed Computing 71(2), 185–197 (2011)
17. Stanford, J., Mattoon, S., Pepple, K.: Practical cloud computing patterns (2009), http://wikis.sun.com/download/attachments/116065636/Practical-Cloud-Patterns-S311528.pdf
18. Subashini, S., Kavitha, V.: A survey on security issues in service delivery models of cloud computing. Journal of Network and Computer Applications 34(1), 1–11 (2011)
19. Zhang, Q., Cheng, L., Boutaba, R.: Cloud computing: state-of-the-art and research challenges. Journal of Internet Services and Applications 1, 7–18 (2010)