

Towards Enabling Scientific Workflows for the Future Internet of Things

Attila Kertesz^{1,2} and Tamas Pflanzner¹

¹ University of Szeged, H-6720 Szeged, Dugonics ter 13, Hungary,
{keratt,tampfla}@inf.u-szeged.hu

² MTA SZTAKI, P.O. Box 63, Hungary

Abstract. Cloud computing offers on-demand access to computational, infrastructure and data resources operated from a remote source. This novel technology has opened new ways of flexible resource provisions for businesses to manage applications and data responding to new demands from customers. In the current web application scenario a rapidly growing number of powerful devices join the Internet, significantly impacting on the global traffic volume and foreshadowing a world of smart devices, or things in the Internet of Things (IoT) perspective. This trend calls for an ecosystem that provides means to interconnect and control these devices. In this position paper we envision the integration of IoT into Cloud-enabled scientific workflows to support the proliferation of IoT with the help of cloud technologies. These enhanced workflows will enable the creation and management of user applications that bring clouds and IoT closer to users by hiding the complexity and cumbersome utilization of virtualized resources, data sources and things. The goal of this approach is to ease the lives of users and foster scientific work by engaging the Internet of Things.

Key words: cloud computing, scientific workflows, internet of things

1 Introduction

Cloud computing is a diverse research area that encompasses many aspects of sharing software and hardware solutions, including computing and storage resources, application runtimes or complex application functionalities. Cloud computing offers on-demand access to computational, infrastructure and data resources operated from a remote source. This novel technology has opened new ways of flexible resource provisions for businesses to manage Information technology (IT) applications and data responding to new demands from customers. The concept of cloud computing has been pioneered by commercial companies with the promise to allow elastic construction of virtual infrastructures, which attracted users early on. Its technical motivation has been introduced in [1][4]. Cloud solutions enable businesses with the option to outsource the operation and management of IT infrastructure and services, allowing the business and its employees to concentrate on their core competencies. As new products and technologies are offered in the near future, Gartner estimated that till 2015 \$112

billion would be spent by businesses and individuals on cloud computing offerings from service providers such as Amazon, IBM and Microsoft [2].

In the current worldwide Information and Communications Technology (ICT) scenario a growing number of powerful devices (smartphones, household appliances, etc.) join the Internet, significantly impacting on the global traffic volume (e.g. by data sharing, voice, multimedia) and foreshadowing a world of smart devices, or things in the Internet of Things (IoT) perspective. The Cluster of European Research Projects on the Internet of Things considers the Internet of Things as a vital part of Future Internet and they defined it as a dynamic global network infrastructure with self configuring capabilities based on standard and interoperable communication protocols. Things in this network interact and communicate among themselves and with the environment by exchanging data and information sensed, and react autonomously to events and influence them by triggering actions with or without direct human intervention [8]. According to another Gartner report there will be 30 billion devices always online and more than 200 billion devices discontinuously online by 2020 [3]. These trends and estimations call for an ecosystem that provides means to interconnect and control these devices.

Nowadays cloud computing has reached a maturity state and high level of popularity that various cloud services have become a part of our lives. These services are offered at different cloud deployment models ranging from the lowest infrastructure level to the highest software or application level. Within Infrastructure as a Service (IaaS) solutions we can differentiate public, private, hybrid and community clouds according to recent reports of standardization bodies. The previous two types may utilize more than one cloud system, which is also called as a cloud federation [7]. Such federations can be good candidates to serve as a base for the envisioned ecosystem. With the help of cloud solutions, user data can be stored in a remote location, and can be accessed from anywhere. Therefore mobile devices can also benefit from these cloud services: the enormous data users produce with these devices are continuously posted to online services, which may require the use of several Cloud providers at the same time to efficiently store and retrieve these data.

Gubbi et al. [6] have identified that to support the IoT vision, the current computing paradigm need to go beyond traditional mobile computing scenarios and cloud computing has the potential to address these needs, and it is able to hide data generation, processing and visualization tasks. M. D. Assuncao et al. [5] also highlighted that there are many open challenges in applying clouds for Big Data management. By addressing some of these challenges, the goal of our proposed approach is to support the proliferation of IoT with the help of cloud technologies, thus to integrate IoT into a cloud ecosystem, a complex system of interdependent components that work together to enable the creation and management of user applications in the form of heterogeneous service mashups or workflows. In this paper we present a vision to create such an integrated ecosystem, and discuss the main requirements for enabling this vision.

The remainder of this paper is as follows: Section 2 presents a survey of related works in the corresponding categories, and Section 3 presents the vision and requirements of our proposed approach to engage things in scientific workflows. Finally, the contributions are summarized in Section 5.

2 Related works

Most of current cloud computing offerings still lock customers into a single cloud infrastructure, platform or application, preventing the portability of data or software created by them. Even if portability is supported, it is barely used by customers due to its complexity and high switching costs. The European Network and Information Security Agency (ENISA) has also recognized the lock-in problem as a high risk that cloud infrastructures entail [9]. The increasing competition between the leading vendors in the cloud market, such as Amazon, Microsoft, Google and Salesforce, each of which promotes its own, incompatible cloud standards and formats [10], prevents them from agreeing on a widely accepted, standardized way to utilize cloud details and specifications. However, an interoperable cloud environment would benefit customers, as they could migrate their virtual machines, data and applications between cloud providers without setting data at risk. Nevertheless there are promising approaches that work for enabling interoperability: DeltaCloud [11] is an open source software that moves towards standard public/hybrid cloud interaction APIs, with an emphasis on compute and VM-based IaaS, and has been submitted to DTMF. OCCI [12] is a family of specifications and standards for cloud interfaces, geared towards interoperability and extensibility, by means of a definition of a Core model, and a suite of already developed documents, falling under two categories, Renderings and Extensions, the former aimed at definition and description of interaction methods and APIs for extending the model with new resource types, attributes, and available actions. The need for intermediary components (coordinators, brokers, exchange) is explained in [13], where the authors outline an architecture for a federated network of clouds. Federation issues in cloud environments have been considered in some research projects, such as mOSAIC [26] and OPTIMIS [27].

In the field of resource abstraction for IoT, good efforts have been made towards the description and implementation of languages and frameworks for efficient representation, annotation and processing of sensed data. There are several standardization bodies, such as the OGC Sensor Web Enablement [14] that develops languages and semantic annotations for abstracting sensors and sensor networks. It has taken important steps towards enabling the web-based discovery, exchange, and processing of sensor observations. Web service interface specifications such as the Sensor Observation Service [15] facilitate the discovery, access and search over the sensor data, and it provides means to integrate data from heterogeneous sources in a standard format accessible to internet users. The W3C Semantic Sensor Network Incubator Group [16] has also been established, aiming at extending this syntactic level interoperability to a semantic

level, through the investigation of two separate but closely related tasks: the development of an ontology for describing sensors and sensor data, and the development of an annotation framework for using semantic metadata.

Concerning virtualization for IoT, Alam et al. [18] propose an Internet of Things (IoT) virtualization framework to support connected objects sensor event processing and reasoning by providing a semantic overlay of underlying IoT cloud. The framework uses the sensor-as-a-service notion to expose IoT clouds connected objects functional aspects in the form of web services. The framework uses an adapter oriented approach to address the issue of connectivity with various types of sensor nodes. Virtual sensor networks have also been proposed towards virtualization in order to offer a new and dynamic collaboration paradigm that requires accommodating multiple logical network instances over a single physical network infrastructure with the ultimate goal to support applications with different requirements and to utilize in an efficient manner the available network resources.

The integration of IoT and clouds has been envisioned by Botta et al. [23] by summarizing their main properties, features, underlying technologies, and open issues. A solution for merging IoT and clouds is proposed by Nastic et al. [24]. They argue that system designers and operations managers face numerous challenges to realize IoT cloud systems in practice, due to the complexity and diversity of their requirements in terms of IoT resources consumption, customization and runtime governance. They propose a novel approach to IoT cloud that encapsulates fine-grained IoT resources and capabilities in well-defined APIs in order to provide a unified view on accessing, configuring and operating IoT cloud systems, and demonstrate the framework for managing electric fleet vehicles.

Integrating sensor network approaches to clouds has been investigated by Hassan et al. [17]. They proposed a framework of sensor-cloud integration, in which they introduced a publish/subscribe based model to simplify the integration of sensor networks with cloud-based community-centric applications. The core component to manage subscriptions is the publish/subscribe Broker, which is responsible for monitoring, processing and delivering events to registered users through SaaS applications. In [19] an infrastructure called Sensor-Cloud infrastructure is proposed, that can manage physical sensors on IT infrastructure. The Sensor-Cloud infrastructure virtualizes a physical sensor as a virtual sensor on the cloud computing and dynamic grouped virtual sensors on cloud computing can be automatic provisioned when the users need them. Combining sensors with cloud computing has been addressed by Mitton et al. [25]. They state that the constantly growing number of powerful devices join the Internet significantly impacting data traffic. Heterogeneous resources can be aggregated and abstracted according to tailored thing-like semantics, thus enabling a cloud of Things. They say that in the Future Internet initiatives, sensor networks will assume even more of a crucial role, especially for making smarter cities. Smart sensors are very heterogeneous in terms of communication technologies, sensing features and elaboration capabilities. They also propose an architecture based on standard specifications, and define an approach by the phrase Sensing and

Actuation as a Service (SAaaS). It envisages new scenarios and innovative, ubiquitous, value-added applications, disclosing the sensing and actuation world to any user, a customer and at the same time a potential provider as well, thus enabling an open marketplace of sensors and actuators. In this paper we also follow and build on this definition.

Gesing et al. [20] states that in the last decades many mature workflow engines and workflow editors have been developed to support primarily scientific communities in managing workflows. While there is a trend followed by these workflow managers to ease the creation of workflows tailored to their specific workflow system, the available tools still require much understanding of the workflow concepts and languages. They propose an approach targeting various workflow systems and building a single user interface for editing and monitoring workflows under consideration of aspects such as optimization and provenance of data. They envision a workflow dashboard offered in a web browser and connecting seamlessly to available workflow systems and external resources like cloud infrastructures. In this paper we plan to follow a similar vision by enabling the combination of various cloud, social networking and IoT services to workflows in a simple way, similarly to the approach followed by ITTT [21] that enables interconnecting web-based services by so-called, predefined channels. DashMash [22] also follows a web service composition approach that is easy to grasp by users. They propose a web platform that allows end users to develop their own mashups making use of an intelligible paradigm that abstracts from technical variables. They use also-called recommendation mechanisms that take into account quality variables to help end users select data sources, components and composition patterns specifying also non-functional user requirements.

Apart from these research approaches our vision is to combine cloud computational and data services, and IoT capabilities into an ecosystem. This vision brings these technologies closer to users, and provides a simple way to form general purpose mashups or workflows of these services in order to ease their everyday lives, and scientific works. In this way their own mobile devices can be used to participate in this ecosystem, and they can be interlinked with their data of interest either coming from social media sites, other public sources or personal cloud storages, and their required data manipulation or processing can be computed dynamically in infrastructure clouds, or fed to traditional web services.

3 An approach for enabling workflows for IoT

3.1 A vision for integrating clouds and IoT services in an ecosystem

An overview of our approach is shown in Figure 1, representing an ecosystem of compute, data, networking and sensing resources (computers, disks, mobile or other sensing devices (i.e. things)) in the cloud belonging to separate administrative domains, managed according to the policies by the local systems. In order to create and manage such a cloud ecosystem, we need to use or enable the

abstraction and virtualization these heterogeneous resources and provide customization features for encapsulating them into services. We use the following naming conventions for these resource groups:

- Infrastructure as a Service (IaaS): represents cloud compute and data services in Virtual Machines (VMs) offered or managed by public or private clouds. In the ecosystem it is possible to use heterogeneous providers both from industrial and academic providers.
- Storage as a Service (STaaS): represents online data storage services such as Personal Clouds (e.g. DropBox, Google Drive).
- Data as a Service (DaaS): this category represents data sources for user communities covering big and small data. Social networking (e.g. Twitter or Facebook) and social media sites (e.g. YouTube) are one of the main sources, but traces from workload or experiment archives/provenances can also be made available through these services.
- Sensing and Actuation as a Service (SAaaS): this category brings the Internet of Things to the ecosystem and provides access to various devices and their sensors (e.g. mobile phones, tablets, smart televisions as well as their sensors such as thermometer, GPS, microphone or camera).
- Web Service (WS): represents the traditional web services covering various areas, e.g. business processes, compute services, travel planners, data analyzers or search engines.

In our vision resources from these categories are encapsulated to services (some are already available), and these services are composed into a dynamic workflow forming a unique application to be executed within the ecosystem. These services can be put together just like pieces of a puzzle, and this composition should be done in a straightforward and user friendly web-based graphical environment.

The main goals of this vision are:

- to create a straightforward, easy to use web-based graphical environment for users to use and manage this ecosystem;
- to bring clouds and IoT closer to users by hiding the complexity and cumbersome utilization of virtualized resources and data sources;
- and to open the world made available by this cloud ecosystem to non-expert users as well as to the scientific community – to enable the combination of social networking, cloud computing and the Internet of Things.

3.2 Requirements to enable the proposed vision

In order to realize this vision, the following issues should be addressed by bringing innovation to current state of the art:

- Sensor abstraction: sensing and actuation resources and devices have to be abstracted and encapsulated to services, providing a homogeneous view of heterogeneous sensors and actuators hosted by both mobiles and sensor networks, also providing adequate hardware contextualization and isolation features.

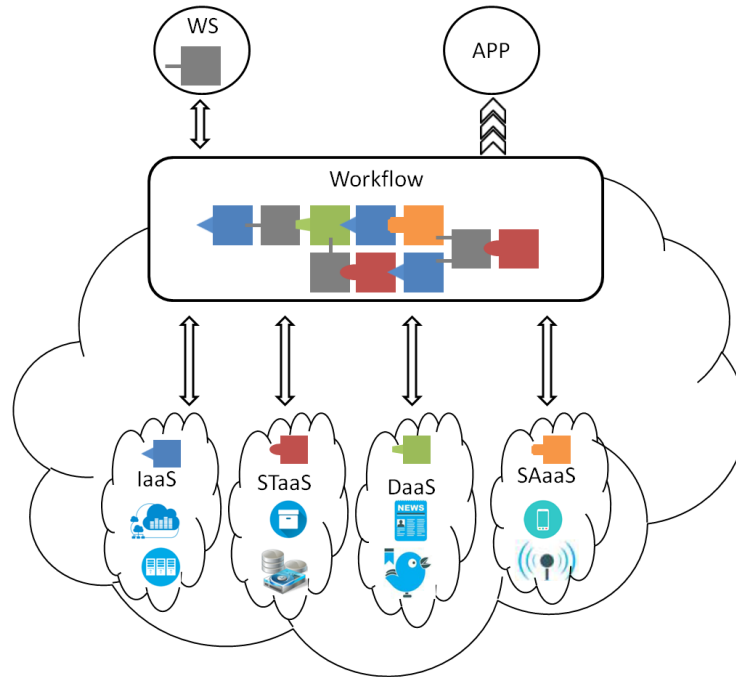


Fig. 1. Workflow applications on top of a cloud ecosystem

- Cloud service management: adequate mechanisms and tools have to be provided in order to manage a cloud provider subscription, and implementing and enforcing policies merging device owner and cloud provider objectives.
- Workflow management and orchestration: facilities for enabling higher level features for integrating cloud, IoT and social networking services into mashups, and enhanced services/APIs have to be provided, also in terms of methods and processes for software development and management applying software engineering principles. All these necessary mechanisms and tools have to implement basic functionalities that are easily customizable and extensible, also adaptive to changes such as load fluctuations.
- Straightforward, easy-to-use application development: There is a need to bring cloud technologies closer to users, and provide a simple way to form general purpose mashups of these services in order to ease their lives, everyday works. In our vision user devices can be interlinked with their data of interest either coming from social media sites, other public sources or personal cloud storages, and their required data manipulation or processing can be computed dynamically in infrastructure clouds.
- Security: effective mechanisms are required to provide identity management, privacy, trustiness, resources and metadata protection and integrity. We plan

to provide specific policies that preserve user privacy and conformation to national legislation for data processing.

State-of-the-art scientific workflow development and execution environments or gateways support many of these features, so they are good candidates to realize this vision. The most wanting, missing feature is the support for IoT integration. In the next section we introduce how we envision to perform this extension in the near future.

4 Towards integrating the approach to a scientific gateway

Researchers of various disciplines ranging from life sciences and astronomy to computational chemistry, create and use scientific applications producing large amount of complex data relying heavily on compute-intensive modeling, simulation and analysis. The ever growing number of such computation-intensive applications calls for the interoperation of distributed infrastructures including private and public clouds, grids and clusters. Scientific workflows have become a key paradigm for managing complex tasks and have emerged as a unifying mechanism for handling scientific data. Workflow applications capture the essence of the scientific process, providing means to describe it via logical data- or control-flows. During the execution of a workflow, its jobs are mapped onto resources of concrete Distributed Computing Infrastructures (DCIs) to perform large-scale experiments.

One of the popular workflow execution environments and scientific gateways is the WS-PGRADE/gUSE portal environment [29] that is capable of executing scientific workflows in an interoperability way, through multiple DCIs. Recently a new feature has been introduced in this gateway by enabling a new type of workflow called infrastructure-aware workflow [28]. These are scientific workflows extended with new node types that enable the on-the-fly creation and destruction of the required infrastructures in the clouds. The paper also describes the semantics of these new types of nodes, how these new type of workflows can be implemented by a new service called as One Click Cloud Orchestrator, and how this service can be integrated with the WS-PGRADE/gUSE portal to provide the required functionalities.

By following this approach we plan to develop an IoT-aware workflow, which is a scientific workflow extended with new resource adaptor node types. These special nodes will be used to encapsulate sensors and things by realizing a SAaaS call in the workflow to send or receive data to/from entities of the IoT world.

5 Conclusions

In this paper we envisioned a future ecosystem integrating IoT, web and cloud-based services. The goal of our proposed approach is to support the proliferation

of IoT with the help of cloud technologies. In this complex system interdependent components can work together to enable the creation and management of user applications in the form of heterogeneous service mashups or workflows. These enhanced workflows will enable the creation and management of user applications that bring clouds and IoT closer to users by hiding the complexity and cumbersome utilization of virtualized resources, data sources and things. Our future work will address the development of this approach in a real-world workflow execution environment.

6 Acknowledgment

The research leading to these results has received funding from the European COST programme under Action identifier IC1304 (ACROSS).

References

1. Buyya B, Yeo C S, Venugopal S, Broberg J, and Brandic I, Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility, *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599-616, June 2009.
2. Pring B et. al., Forecast: Public Cloud Services, Worldwide and Regions, Industry Sectors, 2009-2014. Gartner report. Online: <http://www.gartner.com/Display-Document?ref=clientFriendly-Url&id=1378513>, June 2010.
3. J. Mahoney and H. LeHong, The Internet of Things is coming, Gartner report. Online: <https://www.gartner.com/doc/1799626/internet-things-coming>, September 2011.
4. Vaquero L M, Rodero-Merino L, Caceres J, and Lindner M, A break in the clouds: towards a cloud definition, *SIGCOMM Comput. Commun. Rev.* 39, 1, pp. 50-55, 2008.
5. M.D. Assuncao, R.N. Calheiros, S. Bianchi, M. A. S. Netto, R. Buyya. Big Data Computing and Clouds: Challenges, Solutions, and Future Directions. arXiv:1312.4722, Dec. 2013.
6. J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, Volume 29, Issue 7, pp. 1645-1660, September 2013.
7. A. Kertesz. Characterizing Cloud Federation Approaches. In book: *Cloud Computing - Challenges, Limitations and R&D Solutions*, Z. Mahmood (Ed.), Springer Series on Computer Communications and Networks, pp 277-296. Oct 2014.
8. H. Sundmaeker, P. Guillemin, P. Friess, S. Woelffle. Vision and challenges for realising the Internet of Things. CERP IoT - Cluster of European Research Projects on the Internet of Things, CN: KK-31-10-323-EN-C, March 2010.
9. D. Catteddu and G. Hogben, Cloud Computing-Benefits, risks and recommendations for information security, ENISA report, 2009.
10. G. Sperb Machado, D. Hausheer, and B. Stiller, Considerations on the Interoperability of and between Cloud Computing Standards. In 27th Open Grid Forum (OGF27), G2CNet Workshop: From Grid to Cloud Networks, Banff, Canada, 2009.
11. Apache Deltacloud website. Online: <http://deltacloud.apache.org/developers.html>. Accessed in December 2014.

12. OCCI-Open Cloud Computing Interface, <http://occi-wg.org/>. Accessed in Dec. 2014.
13. R. Buyya, R. Ranjan, R. Calheiros, Intercloud: Utility-oriented federation of cloud computing environments for scaling of application services, *Algorithms and Architectures for Parallel Processing*, pp. 1331, 2010.
14. C. Reed, M. Botts, J. Davidson, G. Percivall, OGC Sensor Web Enablement: Overview and High Level Architecture, *IEEE Autotestcon*, pp. 372-380, 2007.
15. Open Geospatial Consortium Inc., Sensor Observation Service. Document: OGC 06-009r6, Version: 1.0, Category: OpenGIS Implementation Standard, 2007.
16. H. Neuhaus, and M. Compton, The Semantic Sensor Network Ontology: A Generic Language to Describe Sensor Assets. *AGILE Workshop Challenges in Geospatial Data Harmonisation*, 2009.
17. M. M. Hassan, B. Song, and E. Huh. A framework of sensor-cloud integration opportunities and challenges. In *proc. of the 3rd International Conference on Ubiquitous Information Management and Communication*. ACM, New York, NY, USA, 618-626, 2009.
18. S. Alam, M. M. R. Chowdhury, and J. Noll, SenaaS: An Event-driven Sensor Virtualization Approach for Internet of Things Cloud. In *Proc. of the 1st IEEE International Conference on Networked Embedded Systems for Enterprise Applications*, Nov. 2010.
19. M. Yuriyama, and T. Kushida, Sensor-Cloud Infrastructure: Physical Sensor Management with Virtualized Sensors on Cloud Computing. In *Proc. of the 13th International Conference on Network-Based Information Systems*, Gifou, Japan, September 2010.
20. S. Gesing, M. Atkinson, R. Filgueira, I. Taylor, A. Jones, V. Stankovski, C. S. Liew, A. Spinuso, G. Terstyanszky, and P. Kacsuk, Workflows in a dashboard: a new generation of usability. In *Proceedings of the 9th Workshop on Workflows in Support of Large-Scale Science (WORKS '14)*. IEEE Press, Piscataway, NJ, USA, 82-93, 2014.
21. ITTT website. Online: <https://ifttt.com/wtf>. Accessed in December 2014.
22. C. Cappiello, M. Matera, M. Picozzi, G. Sprega, D. Barbagallo, C. Francalanci, DashMash: A Mashup Environment for End User Development. *Web Engineering, Lecture Notes in Computer Science*, Volume 6757, pp 152-166, 2011.
23. A. Botta, W. de Donato, V. Persico, A. Pescapè, On the Integration of Cloud Computing and Internet of Things. *The 2nd International Conference on Future Internet of Things and Cloud (FiCloud-2014)*, August 2014.
24. S. Nastic, S. Sehic, D. Le, H. Truong, and S. Dustdar, Provisioning Software-defined IoT Cloud Systems. *The 2nd International Conference on Future Internet of Things and Cloud (FiCloud-2014)*, August 2014.
25. N. Mitton, S. Papavassiliou, A. Puliafito and K. S. Trivedi, Combining Cloud and sensors in a smart city environment, *EURASIP Journal on Wireless Communications and Networking*, 2012:247, 2012.
26. mOSAIC website. Open Source API and Platform for multiple Clouds. Online: <http://www.mosaic-cloud.eu>. Accessed in December 2014.
27. OPTIMIS website. Online: <http://www.optimis-project.eu>. Accessed in Dec. 2014.
28. P. Kacsuk, G. Kecskemeti, A. Kertesz, Zs. Nemeth, A. Visegradi and M. Gergely, Infrastructure aware scientific workflows and their support by a science gateway. *7th International Workshop on Science Gateways*, Budapest, Hungary, 2015.
29. gUSE science gateways. Online: <http://www.guse.hu/portals/sg>, 2015.