

Guest Editorial Preface

Advances Services for Big Data and Cloud Computing (Part 1)

Ugo Fiore, University Federico II, Naples, Italy

Antonio J. Jara, Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland

The size of datasets produced by large-scale scientific experiments, sensor networks, and the Internet of Things (Singh, Tripathi, & Jara, 2014) is increasing, with the exabyte range to be reached in the near term. Such large datasets need special techniques, approaches, and tools, to allow researchers to gain valuable insights from them, this emerging techniques to deal with huge quantities of data is the so-called “Big Data”. In the same way, Cloud Computing has gained momentum, as it combines flexibility, interoperability, and control over expenditure. It has created the opportunity for a new paradigm, with advanced services and changes also in the way people and organizations work and collaborate. All the computing infrastructure resources are provided as services over the Internet, giving raise to business models called “X as a service (XaaS)” where X is a generic resource that could be, for example, software, hardware, data storage, or network.

Interdisciplinary research is required to ensure a coordinated development of multifaceted solutions that will allow the combination of the advantages of both these innovative scenarios. The reader will, thus, find an articulated discussion of different aspects that provide insights for a deeper understanding of research efforts in this fascinating area.

An important respect where the benefits of this coordinated development are readily visible is the Web of Data. Large amounts of interconnected data are made publicly available through different domains in an open environment, taking advantage of structured data sets freely accessible all over the world, related to many kinds of information. Concerns regarding integration, interoperability, security, and distribution should be addresses by application and services that have to manage this massive knowledge base. In the context of Cloud computing, several solutions have been proposed to address some of the security issues, in particular availability and access control, as well as efficient mapping techniques that organize efficient connection and delivery of data, leading to the Data as a Service (DaaS) model.

RamodeAlcarria, Tomom Robles, Diego Martin, and Álvaro Sánchez, in *Enabling Efficient Service Distribution using Process Model Transformations*, explore service distribution for application deployment with data sets distributed over multiple cloud zones. By means of process model transformation techniques, they develop a cost model and an algorithm to distribute workflow activities efficiently.

Extracting usable information from data sets is a multifaceted task that involves disparate techniques, between which semantic analysis for textual data and structural similarity for graph-based structure, e.g., those representing relationships among entities or events. In *Measuring Semantic-based Structural Similarity in Multi-relational Networks*, Yunchuan Sun, Rongfang Bie, and Junsheng Zhang make the two methods converge, developing a metric for quantitative assessment of the semantic-based structural similarity between two networks.

Federated infrastructures for digital identity management are an enabler for cloud services involving different organizations. In *A blind signature-based approach for cross-domain authentication in the cloud environment*, Aniello Castiglione, Francesco Palmieri, Chin-Ling Chen, and Yao-Chung Chang deal with inter-cloud authentication, proposing a cross-domain scheme that allows users to access transparently to multiple cloud services using a single authentication.

Hadoop, an open-source implementation of the MapReduce programming model for parallel processing of massive datasets on non-dedicated distributed computing hardware is commonly used in Big Data Analytics. *Recent Developments on Security and Reliability in Large-scale Data Processing with MapReduce*, by Christian Esposito and Massimo Ficco, reviews and illustrates the available solutions to security and reliability concerns within large-scale data processing infrastructures, highlighting open issues and subjects for future research.

Ugo Fiore
Antonio J. Jara
Guest Editors
IJDWM

REFERENCES

Fiore, U., Palmieri, F., Castiglione, A., & De Santis, A. (2014). A cluster-based data-centric model for network-aware task scheduling in distributed systems. *International Journal of Parallel Programming*, 42(5), 755–775. doi:10.1007/s10766-013-0289-y

Singh, D., Tripathi, G., & Jara, A. J. (2014). A survey of Internet-of-things: Future vision, architecture, challenges and services. In *2014 IEEE World Forum on Internet of Things (WF-IoT)* (pp. 287–292).