Scouting Big Data Campaigns using TOREADOR Labs

Claudio A. Ardagna, Paolo Ceravolo
Università degli Studi di Milano
Computer Science Department
Crema, CR 26013, Italy
{claudio.ardagna,paolo.ceravolo}@unimi.it

Marcello Leida
Taiger
Madrid 28036, Spain
marcello.leida@taiger.com

Ernesto Damiani
Consorzio Interuniversitario
Nazionale per l'Informatica
Rome 00198, Italy
ernesto.damiani@unimi.it

ABSTRACT

TOREADOR Labs[1] offer a Big Data Analytics-as-a-Service environment for testing simplified but real-life Big Data analytics vertical scenarios. Users are challenged with requirements, described from a business perspective, and are requested to compare alternative options, investigating the consequences of their choices. This "trial and error" approach brings up the interconnections and interferences of the different design stages typically addressed in preparing a Big Data campaign.

Keywords
Parallel computing methodologies; Modeling and simulation.

1. INTRODUCTION
Today, the level of complexity of architectures supporting Big Data Analytics (BDA) and the lack of standardisation for them represents a huge barrier towards the adoption of Big Data technologies, especially for those organisations and SMEs not having the sufficient amount of competences and skills. Another major hindering factor is the so-called "regulatory barrier", that is, concerns about violating data access, sharing and custody regulations when using BDA, and the high cost of obtaining legal clearance for specific scenarios, which is discouraging companies, particularly SMEs, from taking over BDA.

Project TOREADOR aims to overcome some of these hurdles, by providing a platform that supports customers lacking Big Data expertise in the management of BDA and deployment of a full Big Data pipeline [2]. Users with different skills and expertise can benefit by using TOREADOR. Users lacking proper data science expertise (e.g., modeling, analysis, problem solving) can use TOREADOR for preparing the real analytics, reason on data to find out hidden patterns and information, and solve business problems. Users lacking expertise proper of data engineers (e.g., builds a robust and fault-tolerant data pipeline, install a Big Data system) can use TOREADOR to automatically identify and deploy the proper set of technologies that accomplish their requirements. Users lacking both type of expertise can use TOREADOR for a proper initiation in the Big Data realm.

2. BIG DATA-AS-A-SERVICE
Big Data Analytics-as-a-Service (BDAaaS) consists of a set of automatic tools and methodologies that allows customers to design BDA and deploy a full Big Data pipeline addressing their goals [1]. BDAaaS can be seen as a function that takes as input users' Big Data goals and preferences, and returns as output a ready-to-be executed Big Data pipeline.

While declarative goals underlying the use of Big Data services are usually industry-dependent, we argue that identifying a core set of standard indicators is an important step towards increasing transparency of the commitments taken by Big Data service providers, as well as the awareness of users adopting a Big Data solution. Indicators present a way for measuring or assessing a business goal, such as analytics tasks or regulatory constraints on personal data protection, and are accompanied by Big Data objectives representing the target to be achieved for fulfilling the goal.

3. TOREADOR LABS
The model driven approach adopted by TOREADOR supports the creation of a virtual environment particularly suited for training Big Data professionals using a "trial and error" approach. This environment supports users in understanding the interrelations and interferences of the different design options available when preparing a BDA.

In this context, the TOREADOR Labs provide a free-limited access to TOREADOR using a Platform-as-a-Service solution. It proposes a simplified version of real-life vertical scenarios and success stories organised in a set of challenges, where the trainees are requested to identify alternative options, and investigate the consequences of their choices. Note that this kind of experience is usually not available in the professional Big Data platforms today in the market, where the architectural and data complexity make it difficult to compare different runs of a composite BDA.

REFERENCES
[1] E. D. Claudio Ardagna, Paolo Ceravolo. Big data analytics as-a-service: Issues and challenges. In Proceedings of the 3rd International Workshop on Privacy and Security of Big Data (PSBD). IEEE, 2016.

[2] M. Leida, C. Ruiz, and P. Ceravolo. Facing big data variety in a model driven approach. In Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), 2016 IEEE 2nd International Forum on, pages 1–6. IEEE, 2016.