

Big Data Experience: a framework for management strategy evaluation

Luis G. C. Russi, Carlos R. Senna, and Edmundo R. M. Madeira
Institute of Computing - University of Campinas (UNICAMP)
luisrussi@lrc.ic.unicamp.br, {crsenna, edmundo}@ic.unicamp.br

Abstract—Nowadays using big data capacities is an emergent topic where data is generated from a lot of sources, in a real information explosion. Big Data is naturally big and complex, and overloads capacity from the current data management tools and traditional applications. Cloud computing with its ubiquitous characteristic, on demand and dynamic resource provisioning at low cost has potential to be the environment to treat this data explosion. In this context, this work proposes an architecture that allows running experiments involving Big Data.

I. INTRODUCTION

The term Big Data is new and has an interesting aspect, it's not always directly related to big amounts of data, also being applied to its data characteristic and velocity they are created [1]. The data are multi-structured in various formats like social media, videos, images and sensors that move rapidly between source and destinations, where they can load multiple informations. Big Data is generated continuously where can be extracted information windows that content can be in a real time form. We propose the creation of a hybrid cloud architecture composed of a private cloud environment provided by OpenStack software and the grid computing supported by Globus Toolkit, an Open Grid Services Architecture implementation. This joint will allow us evaluate strategies based on services providing technical studies of distributed systems to access data.

II. RESEARCH SCENARIOS ON BIG DATA EXPERIENCE

There are a lot of scenarios that can be tried on Big Data Experience (BDE), network simulations, scientific workflows, generation and image treatment, and others. Here we'll discuss the scenario to analyze data from sensor networks. This one based on real data [2], proposes a solution that combines web services and software components to process and visualize data obtained through network sensors. The architecture receives data through the Internet from a sensor pool and sort them applying filters as the visualization needs. Furthermore would be possible analyze a data burst traffic (big quantities of data sent in a short while) and try the cloud scalability in a unique way.

III. BIG DATA EXPERIENCE ARCHITECTURE

To evaluate the Big Data treatment strategies on hybrid environments we propose a service infrastructure pointed to big data research on the cloud. The data base is formed by collect informations through a sensor network simulator service (SNSS). To collect data we use the collector data

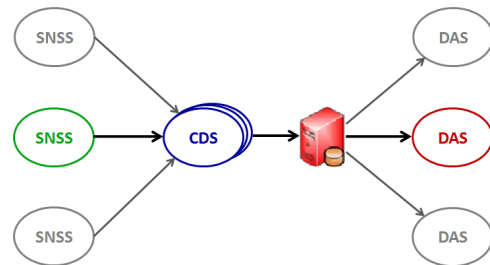


Fig. 1. Big Data Experience Infrastructure

service (CDS), that allows us evaluate various strategies to receive data. The big mass of data is available through the data access service (DAS), giving flexibility to our testbed granting the architecture scalability. Using a service access make easy to compare data access strategies, and evaluate virtual machine utilization, grid sensors or traditional applications. Big Data Experience architecture is shown in Figure 1.

Its infrastructure is composed of an IaaS middleware, a computational grid, and a set of support services. To provide our basic cloud infrastructure we've chosen the open source platform called OpenStack, composed of a computational infrastructure, storage system, virtual machines image management and its software defined network provider, the OpenStack has the necessary components to deploy cloud computing and it allows SDN to evaluate management strategies that fit on the Big Data needs. Through the computational grid Globus Toolkit we create a distributed environment to support concurrent tasks developed on the infrastructure. And finally, to do the workflow management and support parallel executions, we will use our middleware named CEO, developed at the Computer Network Laboratory (LRC).

At our proposed architecture we wish to evaluate strategies based on services, to compare them using virtual machines used by applications and combine different ways to study them. Moreover, the Big Data capacity stored in one place inside the cloud can perform a huge data moving, overloading the environment and generating slowness, which allow study new distributed systems techniques to data access.

REFERENCES

- [1] S. B. Siewert, "Big data in the cloud data velocity, volume, variety, and veracity," IBM Developerworks, Tech. Rep., 2013.
- [2] I. K. Koga, C. B. Medeiros, and O. Branquinho, "Handling and publishing wireless sensor network data: a hands-on experiment," *Journal of Computational Interdisciplinary Science*, pp. 13–22, 2011.