Cloudet:
A Cloud-Driven Visual Cognition of Large Streaming Data

George Baciu, GAMA Lab, Department of Computing, Hong Kong Polytechnic University, Hung Hom, Hong Kong

Chenhui Li, Hong Kong Polytechnic University, Hung Hom, Hong Kong

Yunzhe Wang, Hong Kong Polytechnic University, Hung Hom, Hong Kong

Xiujun Zhang, Shenzhen University, Shenzhen, China

ABSTRACT

Streaming data cognition has become a dominant problem in interactive visual analytics for event detection, meteorology, cosmology, security, and smart city applications. In order to interact with streaming data patterns in an elastic cloud environment, we present a new elastic framework for big data visual analytics in the cloud, the Cloudet. The Cloudet is a self-adaptive cloud-based platform that treats both data and compute nodes as elastic objects. The main objective is to readily achieve the scalability and elasticity of cloud computing platforms in order to process large streaming data and adapt to potential interactions between data stream features. Our main contributions include a robust cloud-based framework called the Cloudet. This is a cloud profile manager that attempts to optimize resource parameters in order to achieve expressivity, scalability, reliability, and the proper aggregation of the compute nodes and data streams into several density maps for the purpose of dynamic visualization.

INTRODUCTION

With the exponential increase in streaming data big data cognition has become a real challenge. Examples can be found in traffic monitoring and management, weather, astronomy, genomics, on-line financial transactions, and electronic tracking of large capital flows. For large scale streaming data, pattern cognition involves interactive pattern querying, filtering, smoothing, classification, rendering, and finally visualization (Patterson et al., 2014). The capacity to store and display even moderately sized data sets (e.g. in the Terabyte range) has become limited.

Currently, it is difficult to add cognitive analytics to big data sets (Tudoran et al., 2015). Big data sets are characterized by *Volume*, *Velocity*, and *Variety* (VVV). Processing big data has specific computational requirements both for both storage (i.e. volume) and speed (i.e. velocity). These requirements cannot be satisfied by simply allocating one fat server or a large number of thin client-server machines. However, within a cloud platform, access to a large number of compute nodes and leveraging on the availability of Petabyte-scale storage resources can often accelerate the processing time by allowing caching of large data streams, and distributing the data amongst store components of sizable VRAMs (i.e. in the Terabyte range). This gives the possibility to virtually adapt to any

type of data streaming, filtering, smoothing and rendering requirements. Furthermore, mobile client machines, such as GPU enhanced tablets, and virtual desktops, can access an order of magnitude more resources within the cloud system in parallel from arbitrary locations.

In this work, we focus on the real-time visualization of streaming data in the context of big data and cloud computing platforms. In this context, we propose an elastic platform that would allow thin clients to extend beyond mobile device limitations in both storage and speed and open the door to scalable big data cognitive analytics. In this paper, we present the *Cloudet*, a flexible SPARK (Zaharia et al., 2010) based framework, which can adapt to the big data VVV characteristics: Volume, Velocity and Variety by monitoring data streams and adjusting the internal resource parameters in order to maintain quality of service, in this case rendering and interactive visualization requirements. Large numbers of data features can be processed on several interacting high performance cloud compute nodes and the results can be dynamically adjusted for arbitrary number of displays with different sizes and form factors.

The main idea is to allow the system to intelligently adapt computations, storage and communication connectivity based on the characteristics of the data stream patterns. A quad-tree structure for the cloud data (Ding et al., 2011) is amongst the first examples of a discrete elastic cloud management system. However, one limitation is the manual presetting of the cluster resources. The other limitation is that the adjustment is often done in one direction, i.e. as the demand for resources grows.

We are concerned with a more fine grained elastic resource management of the cloud platform during the incoming data stream. That is, adapt to the resource requirements in terms of compute nodes, storage, and communication channels on live data streams. The streaming features of the data can then be selected and rendered at interactive rates for visual cognition. Spatial-temporal patterns can then be visualized interactively in order to understand the interaction between data features. This framework can also be extended to a variety of multi-cluster architectures that can be automatically generated in order to accommodate to the complex interactions between tasks in the process of visual cognition of data streams.

In addition, our framework is suitable for adaptively handling data streams with different peaks in one task. We first address the problem of setting up adaptive clusters according to different types of tasks. Second, a *stream mapp*ing operation is proposed to normalize the unstructured data. Third, the use of adaptive color-theme based rendering with filtering is used in order to enhance visual cognition of the rendered data features. Critical regions and periods can then be easier to detect in the process of interactive visualizations of data streams.

Our experiments show that the *Cloudet* framework can effectively transform large streaming data into structural data. In addition, our framework can also be applied to predict, prioritize, and tag the saliency regions in the visualizations of streaming data. In other words, our major contribution is a robust elastic framework for the cognitive feature enhancement and pattern detection in heterogeneous data streams.

The rest of this paper is organized as follows. In the *Related Work* section, we summarize the related research work. In the *Overview* section we present an overview of the *Cloudet* framework. We continue with the details of the framework in the *Cloudet* section. Details of the visual cognition are proposed in the *Visual Cognition* section. *Implementation* section describes the architecture of the proposed framework. We finalize the paper with *Applications* section showing our experiments and the discussion of the results. Finally, we conclude our work and lay down the plan for our future work.

Related Content

The Event Search Engine
Takeshi Okadome, Yasue Kishino, Takuya Maekawa, Koji Kamei, Yutaka Yanagisawa and Yasushi Sakurai (2012). *Developments in Natural Intelligence Research and Knowledge Engineering: Advancing Applications (pp. 70-82).*
www.igi-global.com/chapter/event-search-engine/66439?camid=4v1a

The Activity Circle: A Social Proxy Interface to Display the Perceived Distributed Viscosity about Workflow Technology
Marcello Sarini (2012). *Cognitively Informed Intelligent Interfaces: Systems Design and Development (pp. 201-217).*
www.igi-global.com/chapter/activity-circle-social-proxy-interface/66275?camid=4v1a

AiryLight: Ambient Environmental Data
Annelie Berner (2015). *Handbook of Research on Maximizing Cognitive Learning through Knowledge Visualization (pp. 487-492).*
www.igi-global.com/chapter/airylight/127493?camid=4v1a

The Operationalization of the Unconscious: Models of Subcognitive Informatics

www.igi-global.com/article/the-operationalization-of-the-unconscious/133294?camid=4v1a