MONICA WACHOWICZ
Professor, Data Science
Cisco Innovation in Big Data
University of New Brunswick
Email: monicaw@unb.ca

Monica Wachowicz is full professor in Data Science, and the Cisco Innovation Chair in Big Data and the NSERC/Cisco Industrial Research Chair in Mobility Analytics at the University of New Brunswick, Canada. She is also the Director of the People in Motion Laboratory, a center of expertise in the application of Internet of Things (IoT) to smart cities. Her research interests include fog/edge computing, machine learning on graphs, mobility analytics, and IoT applications.

Wachowicz' research program focuses on the synergy of (1) streaming analytics for analyzing massive data streams that are being generated today in search of geospatial intelligence in real-time; and (2) cartography for designing maps for a world in which "intelligence" will be embedded in virtually everything around us.

Founding member of the Technical Committee on Big Data (TCBD), IEEE Communications Society and the International Journal of Big Data Intelligence, she has over 250 publications. Her pioneering work in multidisciplinary teams from government, industry and research organizations is fostering the next generation of data scientists for innovation.

Perspective

Society has a very ambitious vision of building smart interconnected cities through the Internet of Things (IoT). Research on Spatial Data Science must address a variety of new challenges that relate to the reproducibility of its results since they are often generated by complex analytical workflows. Over the past year, we have worked on developing and evaluating analytical workflows using our Analytics Everywhere framework, which encompasses the edge-fog-cloud continuum to support streaming analytics for maximizing the potential new insights from IoT data streams.

Currently, our framework consists of three components that can be considered as the main criteria to take into account in order to determine whether an edge-fog-cloud environment is required by an IoT application in smart cities. They can be described as follows:

- **Resource capability**: This element consists of organizing distributed computational nodes (i.e., cloud, fog and edge nodes) that will provide a message broker, data link, IoT device connector, data flow editor, parser, Machine Learning (ML) libraries , in-memory data storage and power for the execution of streaming tasks. Geographically adjacent compute nodes deployed at the edge, fog and cloud will be usually connected through a plethora of communication networks.

- **Analytical capability**: This element selects the best practice methods/algorithms for the orchestrated execution of analytical tasks that are vital to meet the requirements of IoT applications. The compute nodes are needed to perform a priori known analytical tasks to collect, contextualize, process and analyze data from IoT devices.
- **Data life-cycle**: This component describes the changes that data streams go through during the execution of analytical tasks. We expect many types of data life-cycles depending on the types of analytical tasks and compute nodes needed by an IoT application.

But the questions still remain: Are these components enough to build a streaming learning platform for supporting Spatial Data Science? Would it be different with AI in the near future?

Streaming analytics for IoT data is still in its infancy and IoT applications usually require algorithms that must work within limited resources (time and memory). Some open-source frameworks for IoT data stream analytics are being developed including MOA, SAMOA, and skit-multiflow using only streaming processors. Our proposed learning architecture is a step forward in finding a unique solution that combines the advantages of different computational resources into an integrated edge-fog-cloud fabric that is capable of capturing, managing, processing, analyzing, and visualizing IoT data streams. This fabric of computational resources is designed to work towards an asynchronous approach for supporting our Analytics Everywhere ecosystem making the development, deployment, and maintenance more pragmatic and scalable. By breaking down the processing and analytical capabilities into a network of streaming tasks and distributing them into an edge-fog-cloud computing environment, our proposed architecture can support streaming descriptive, diagnostic, and predictive analytics. A single computational resource (e.g., cloudGIS) is not sufficient to support all analytical capabilities that are needed for IoT applications, considering computing power, data stream management, storage and communication networks. We also do not expect that one IoT architecture will fit all IoT applications. More research is need to evaluate our understanding of the existing data lifecycles in smart cities.

From a conceptual perspective, an Analytics Everywhere framework can play an important role in exploring data streams in time and space. Time is an important dimension of this framework, and different approaches have been proposed in the literature to handle unbounded data streams, including landmark windows, sliding windows, and tilted windows. In contrast, the space dimension has been neglected so far, even though data streams are being generated over large geographical areas with fine spatial granularity. We are just unfolding how streaming descriptive analytics actually works. More research efforts are needed to support streaming diagnostic and predictive analytics.

From an implementation perspective, an Analytics Everywhere platform will require (1) a pre- built connector that supports data connectivity to communicate with several devices, (2) a low-latency database for storing data streams, and (3) high-performance processing for supporting the automated tasks. The technological challenge is to design an architecture that can perform analytical tasks without human intervention (e.g., an event from an IoT device triggers an analytical task), and at the same time, cope with the transportation of unbounded data streams where the data rate may overwhelm the processing power of this platform.

It is indisputable that IoT in smart cities will produce a large amount of high-speed streamed and heterogeneous data that poses challenges to performing management, processing, analysis, and visualization tasks within an acceptable time. The Spatial Data Science Symposium will be a discussion forum to advance our community conversation.

This position paper contains information from previous published papers:

Cao H. and Wachowicz, M. (2019). The design of an IoT-GIS platform for performing automated analytical tasks. Computers, *Environment and Urban Systems* 74: 23–40.

Cao, H., Wachowicz, M., Renso, C. and Carlini, E. (2019). Analytics Everywhere: generating insights from the Internet of Things. *IEEE Access Journal* 7: 71749–71769.

Cao, H. and Wachowicz, M. (2019). An edge-fog-cloud architecture of streaming analytics for Internet of Things applications. Sensors, *Special Issue on Issue Edge/Fog/Cloud Computing in the Internet of Things*; Velasco, L and Ruiz, M. (Eds), 19: 3594.