

Síntesis de una Arquitectura de Procesamiento de Datos del RADAR Meteorológico de INTA Anguil

Mario Diván¹, Yanina Bellini², María de los Ángeles Martín¹, María Laura Belmonte², Guillermo Lafuente¹ & Juan Marcelo Caldera²

¹ Facultad de Ingeniería, UNLPam, General Pico, La Pampa, Argentina

² INTA, Estación Experimental Anguil, La Pampa, Argentina.

{mjdivan, martinma, lafuente}@ing.unlpam.edu.ar
{bellini.yanina, belmonte.maria, caldera.juan}@inta.gob.ar

1 Resumen

Actualmente, existen aplicaciones que procesan un conjunto de datos a medida, generados en forma continua, a los efectos de responder a consultas y/o adecuar su comportamiento en función del propio arribo de los datos[1], como es el caso de las aplicaciones para el monitoreo de signos vitales de pacientes; del comportamiento de los mercados financieros; entre otras. En dicho tipo de aplicaciones, se enmarca el Enfoque Integrado de Procesamiento de Datos centrado en Metadatos de Mediciones (*EIPFDcMM*)[2,3], el cual sustentado en el marco de medición y evaluación C-INCAMI (*Context-Information Need, Concept model, Attribute, Metric and Indicator*)[4,5], incorpora metadatos al proceso de medición, promoviendo la repetitividad, comparabilidad y consistencia del mismo. Desde el punto de vista del sustento semántico y formal para la medición y evaluación (*M&E*), el marco conceptual C-INCAMI establece una ontología que incluye los conceptos y relaciones necesarias para especificar los datos y metadatos de cualquier proyecto de M&E. Por otra parte, y a diferencia de otras estrategias de procesamiento de flujos de datos[6,7,8], gracias a la incorporación de metadatos, el EIPFDcMM es capaz de guiar el procesamiento de las medidas provenientes de fuentes de datos heterogéneas, analizando cada una de ellas dentro de su contexto de procedencia, como así también su significado dentro del proyecto de M&E en el que se definió. Así, el EIPFDcMM es un gestor de flujos de datos sustentado en un marco de medición y evaluación, el cual incorpora comportamiento detectivo y predictivo, mediante el empleo de las mediciones y metadatos asociados.

La Estación Experimental Agropecuaria (EEA) INTA Anguil (La Pampa) tiene instalado un radar meteorológico (RM) capaz de generar un flujo de datos de 17gb diarios aproximadamente, lo que representa un desafío para su almacenamiento, gestión y posterior servicio al público, principalmente considerando la importancia que los datos poseen para la región productiva de influencia.

Por otro lado, y si bien el EIFPDcMM incorporó en su esquema de procesamiento la posibilidad de manejar grandes volúmenes de datos como memoria organizacional para el entrenamiento de sus clasificadores[9], no estaba orientado al acopio de la totalidad de los datos procesados en línea y menos aún, al hecho de contar con la posibilidad de integrar la prestación de servicios de datos públicos a terceros.

Este trabajo propone una arquitectura de procesamiento que extiende el EIFPDcMM para incorporar el procesamiento de grandes datos almacenados. Ello permitiría dotar al RM de un comportamiento detectivo y predictivo sobre los datos en línea, como así también incluir una capa de servicios al público, que fomente el consumo de datos generados por el RM de INTA Anguil

Así, y como contribuciones específicas se plantea, *(i) relacionado con la configuración de las fuentes de datos*: la posibilidad de incorporar dentro de la arquitectura de procesamiento del EIFPDcMM como una fuente heterogénea al RM de INTA Anguil en forma transparente, lo que permite incorporar un comportamiento detectivo y predictivo, a partir del proyecto de M&E que los expertos del INTA definan sobre las medidas observadas; *(ii) relacionado con la captación de medidas*: la posibilidad de definir formalmente un proyecto de M&E que permite detectar eventuales desvíos del funcionamiento del RM y promover los ajustes meteorológicos pertinentes, basando los cursos de acción probables a partir de la información disponible en la memoria organizacional; *(iii) relacionado con la arquitectura de procesamiento*: ahora se tiene una arquitectura actualizada para el EIFPDcMM que permite incorporar la gestión y utilización de grandes volúmenes de datos persistentes junto con el procesamiento de flujos de datos; *(iv) relacionado con los servicios de datos*: la arquitectura incorpora en forma transparente la posibilidad de prestar servicios de datos a terceros interesados. De este modo, los datos generados por el RM de la EEA Anguil, podrían ser consumidos en forma directa y sin mediar solicitud o intervención humana, por Centros de Investigación, Servicios Meteorológicos u otros organismos o personas.

Estas contribuciones representan un importante avance con respecto al modelo de procesamiento presentado en [3,9], ya que ahora se incorpora la posibilidad de gestionar grandes volúmenes de datos persistentes en forma adicional a su procesamiento en línea, y contar con la posibilidad de establecer capas de servicios para el eventual consumo de los mismos.

Referencias

1. J. Gehrke, J. Balakrishnan, H. Namit, "Towards a Streaming SQL Standard," Proceedings of the VLDB Endowment, vol. 1, no. 2, pp. 1379-1390, August 2008.
2. M Diván, L Olsina, and S Gordillo, "Strategy for Data Stream Processing Based on Measurement Metadata: An Outpatient Monitoring Scenario," Journal of Software Engineering and Applications, vol. 4, no. 12, pp. 653-665, December 2011.
3. M Diván and L Olsina, "Process View for a Data Stream Processing Strategy based on Measurement Metadata," Electronic Journal of SADIO, vol. 13, no. 1, pp. 16-31, June 2014.
4. Olsina L. Molina H, "Towards the Support of Contextual Information to a Measurement and Evaluation Framework," in QUATIC, Lisboa, Portugal, 2007, pp. 154-163.

5. Papa F, Molina H. Olsina L, "How to Measure and Evaluate Web Applications in a Consistent Way," in Ch. 13 in *Web Engineering.*: Springer, 2007, pp. 385-420.
6. M. Aref, W. Bose, R. Elmagarmid, A. Helal, A. Kamel, I. Mokbel, M. Ali, "NILE-PDT: A Phenomenon Detection and Tracking Framework for Data Stream Management Systems," in *VLDB*, Trondheim, Norway, 2005, pp. 1295-1298.
7. S. Chandrasekaran, S. Cooper, O. Deshpande, A. Franklin, M. Hellerstein, J. Hong, W. Madden, s. Reiss, F. & Shah, M. Krishnamurthy, "TelegraphCQ: An Architectural Status Report," *IEEE Data Engineering Bulletin*, vol. 26, 2003.
8. D. Ahmad, Y. Balazinska, M. Cetintemel, U. Cherniack, M. Hwang, J. Lindner, W. Maskey, A. Rasin, A. Ryvkina, E. Tatbul, N. Xing, Y. & Zdonik, S. Abadi, "The Design of the Borealis Stream Processing Engine," in *Conference on Innovative Data Systems Research (CIDR)*, Asilomar, CA, 2005, pp. 277-289.
9. M Diván, M Martín, and L Olsina, "Towards the feedback of the Data Stream Processing based on Organizational Memory (in spanish)," in *Congreso Nacional de Ingeniería Informática/Sistemas de Información*, Córdoba, Argentina, 2013.