Stream Processing in Production-to-Business Software

Michael Cammert, Christoph Heinz, Jürgen Krämer, Tobias Riemenschneider,
Maxim Schwarzkopf, Bernhard Seeger, Alexander Zeiss
Department of Mathematics and Computer Science
University of Marburg, Germany
{cammert, heinzch, kraemerj, tobys, maxim, seeger, zeissa}@mathematik.uni-marburg.de

Abstract

In order to support continuous queries over data streams, a plethora of suitable techniques as well as prototypes have been developed and evaluated in recent years. In particular, it is of utmost importance to confirm their necessity and feasibility in real-world applications. For that reason, we have successfully coupled our infrastructure for data stream processing (PIPES) with an industrial Production-to-Business software (i-Plant) dedicated to highly automated manufacturing processes.

1. PIPES: Stream Processing

PIPES (*Public Infrastructure for Processing and Exploring Streams*) [3] is a powerful toolkit providing the essential components to build and execute continuous queries over data streams. The library character of PIPES substantially facilitates the construction of a data stream management system (DSMS) tailored to the particular requirements of an application domain. Our demonstration validates this approach by showing up the intuitive and fruitful composition of PIPES and i-Plant.

PIPES relies on a sound semantics [4] which is compatible with the Continuous Query Language (CQL). Its core is a push-based, time-interval operator algebra. Continuous queries are implemented as a directed acyclic operator graph which enables PIPES to gain from subquery sharing.

In addition, PIPES provides flexible frameworks for the runtime components, e. g., scheduler, memory manager, and query optimizer. Each component is parameterized by a strategy in order to guarantee runtime adaptivity. Appropriate strategies are usually based on a cost model [1] which incorporates continuously gathered runtime metadata such as stream rates, memory usage, and operator costs. As more complex metadata, PIPES offers online, resource-adaptive estimations of data distributions based on complex statistical methods, like wavelet-based density estimators [2].

2. i-Plant: Production-to-Business Software

Unlike common business processes, highly automated manufacturing processes produce significantly more data. Typically, such a manufacturing environment autonomously generates continuous streams of process and machine data respectively. Hence, factory automation is a suitable target application for stream processing approaches.

Production-to-Business (P2B) software aims to close the gap between the production and the business layer. Therefore, it is essential to process, monitor, and analyze production data in real-time in order to provide adequate support for economic decision making. A commercial P2B product is *i-Plant* [5], a software platform developed by *Langner Communications AG*[1]. At the production layer, i-Plant allows to integrate heterogeneous data sources, especially low-level factory devices, by standardizing their underlying formats. At the business layer, it provides tailored web-based software applications for managing factory automation and production processes.

3. Composition of PIPES and i-Plant

Figure 1 illustrates the main features of PIPES and i-Plant as well as their composition. Both infrastructures nicely fit together, since (i) they are implemented in Java, (ii) they provide powerful and flexible connectivity facilities, and most importantly, (iii) both rely on push-based data processing concepts.

With regard to i-Plant, so-called *factory automation connectors* establish the connection to factory automation hardware and transform proprietary industrial protocols or *OLE for Process Control* (OPC) into *FactoryXML*, a XML-based communication technology designed for factory automation. These connectors act as data sources. They are managed together with other i-Plant components by the *di-*

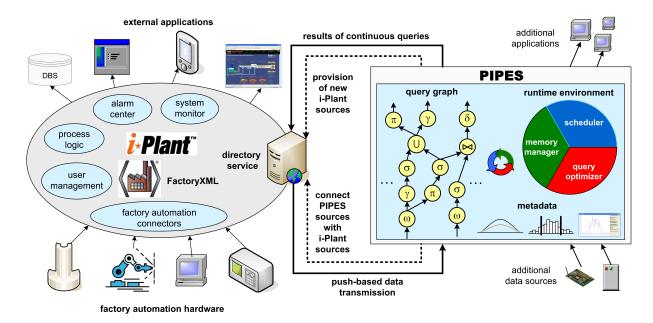[1]i-Plant, P2B, and FactoryXML are registered trademarks of Langner Communications AG.

Figure 1. Composition of PIPES and i-Plant

rectory service which also represents the bridge between i-Plant and PIPES. Whenever a continuous query refers to an i-Plant data source, PIPES establishes a connection via the directory service and receives new values henceforth. Installing a new data source in the directory service makes the results of a continuous query available to i-Plant.

The synergies resulting from the composition of PIPES and i-Plant are manifold. On the one hand, coupling PIPES to i-Plant exploits continuous query processing in factory automation. Thus, sophisticated techniques from research like windowed queries, subquery sharing, adaptive query processing, and complex data analysis become applicable. On the other hand, processing real-world streams proves the feasibility of PIPES. In addition, PIPES profits from the professional web-frontends of i-Plant and its connections to other external applications such as SCM and EAI systems or e-mail and PDA services. Moreover, PIPES gains from user management, backup services, and fault tolerance when accessing i-Plant data sources.

4. Demonstration Content

In our demonstration, we will first give a brief overview of PIPES and i-Plant and explain their core functionality as well as application domains. Second, the connection and communication between PIPES and i-Plant will be illustrated. Third, we will point out the advantages of this seamless composition and emphasize the feasibility of stream processing for this industrial setting.

The entire demonstration will be based on a real-world application scenario of i-Plant with data streams being ob-

tained from sensors of an industrial production line. We will build and execute valuable continuous queries with PIPES, while giving insight into the underlying semantics, the operator implementation, and the metadata generation for the adaptive runtime environment.

The synergies of coupling i-Plant with PIPES confirmed us in our library approach [3], as it often suffices to integrate new functionality into existing systems, instead of building complete DSMS from scratch.

Acknowledgements

References

[1] M. Cammert, J. Krämer, B. Seeger, and S. Vaupel. An Approach to Adaptive Memory Management in Data Stream Systems. In *Proc. of the IEEE Conference on Data Engineering (ICDE)*, 2006.

[2] C. Heinz and B. Seeger. Wavelet Density Estimators over Data Streams. In *Proc. of the ACM Symposium on Applied Computing (SAC)*, pages 578–579, 2005.

[3] J. Krämer and B. Seeger. PIPES - A Public Infrastructure for Processing and Exploring Data Streams. In *Proc. of the ACM SIGMOD*, pages 925–926, 2004.

[4] J. Krämer and B. Seeger. A Temporal Foundation for Continuous Queries over Data Streams. In *Proc. of the Int. Conf. on Management of Data (COMAD)*, pages 70–82, 2005.

[5] Langner Communications AG. *Executive Briefing on i-Plant.* http://www.factoryxml.com/pdfs/executivebriefing_e.pdf.