Managing Continuous Data Integration Flows

Josef Schiefer¹, Jun-Jang Jeng¹, Robert M. Bruckner²

¹IBM Watson Research Center
19 Skyline Drive, Hawthorne, NY 10532
{josef.schiefer, jjjeng}@us.ibm.com

²Institute of Software Technology
Vienna University of Technology
 Favoritenstr. 9 / 188, A-1040 Vienna, Austria
bruckner@ifs.tuwien.ac.at

Abstract. Understanding the performance and dynamic behavior of workflow is crucial in being able to modify, maintain, and improve it. A particularly difficult aspect of measuring the performance of workflows is the dissemination of event data and its transformation into business metrics. In this paper we introduce an architecture that supports a continuous integration of event data from various source systems in near real-time into a data warehouse environment. The proposed architecture takes full advantage of existing J2EE (Java 2 Platform, Enterprise Edition) technology and uses an ETL container for the event data processing. We discuss the challenges of managing flows for continuously integrated data and show how a container environment can provide services that facilitate the flow management.