

# A Visual Approach for Data-Intensive Workflow Validation<sup>☆</sup>

Minjae Park<sup>1</sup>      Hyun Ahn<sup>2</sup>      Kwanghoon Pio Kim<sup>2</sup>

## ABSTRACT

This paper presents a workflow validation method for data-intensive graphical workflow models using real-time workflow tracing mode on data-intensive workflow designer. In order to model and validate workflows, we try to divide as modes have editable mode and tracing mode on data-intensive workflow designer. We could design data-intensive workflow using drag and drop in editable-mode, otherwise we could not design but view and trace workflow model in tracing mode. We would like to focus on tracing-mode for workflow validation, and describe how to use workflow tracing on data-intensive workflow model designer. Especially, it is support data centered operation about control logics and exchange variables on workflow runtime for workflow tracing.

☞ keyword : Data-Intensive Workflow Validation, Workflow, BPM

## 1. Introduction

Traditionally, workflow management concept was business process automation and control[1]. Nowadays, workflow concept and application are going to diffuse not only business process but also several area and industry, so that a workflow has focused on execute system and data resources without human resources. [2, 3, 4] are researches about related data-intensive workflow or scientific workflow. [5] has formal representations of data-intensive workflow model, and [6] has implemented tool about data-intensive or scientific workflow model design.

This paper approach data-intensive workflows that are used automated service by data or systems. Particularly, we are focused on how to validate data-intensive workflow models with

graphical notations.

In organizing the paper, we start from describing the two modes concepts on data-intensive workflow designer. The next section describes the operations for data-intensive workflow validation are consist of workflow runtime control functions, applying colored workflow model and checking variables check with describing an implementation of data-intensive workflow model execution. And describing related works of this research. And we finalize the paper with conclusions.

## 2. Two Modes of Data-Intensive Workflow Designer

Data-intensive workflows are special in perspective of does not have properties about human resources. It support just data-processing with data-processing activity and workflow logics. So, it is most important validated data processing, because of data focused.

In general, well-known workflow management systems[7, 8, 9] are operated by the concept for build-time and run-time. In order to validate workflow with visual approach, it should has build-time concept and run-time concept for data-intensive workflow designer. So, the data-intensive workflow designer has two modes support both design and run workflow. We named it such as: one is editable mode and the other is tracing mode.

<sup>1</sup> Global Engineering Innovation Center, BISTel, Inc., Seoul, 06754, South Korea

<sup>2</sup> Dept. of Computer Science, Kyonggi University, Suwon, 16227, South Korea

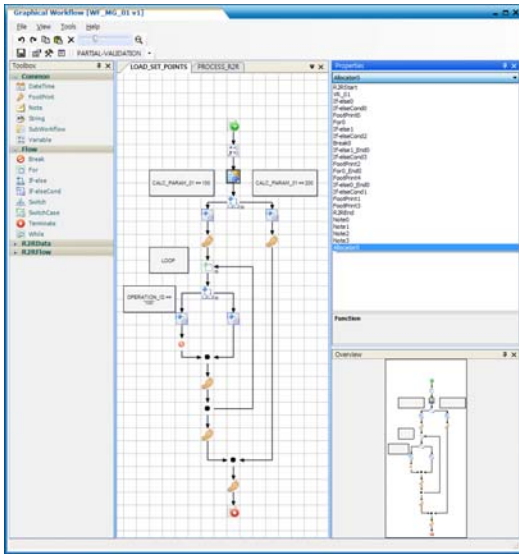
\* Corresponding author (kwang@kyonggi.ac.kr)

[Received 11 May 2016, Reviewed 26 May 2016, Accepted 4 July 2016]

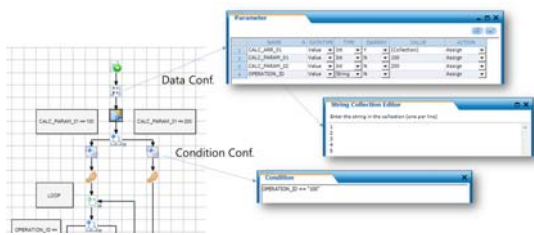
☆ A preliminary version of this paper was presented at ICONI 2015 and was selected as an outstanding paper. This research was supported by the Technology Innovation Program (10045913, Development of Big Data based Analysis and Control Platform for Semiconductor Manufacturing Plants) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea). Also this work was partially supported by Kyonggi University's Graduate Research Assitanship 2016.

## 2.1 The Editable Mode

For workflow process design and activity property configure, it is opened editable mode for workflow model designer. We could design workflow model and configure property of activity using drag and drop method.



(Figure 1) The Editable Mode of Designer



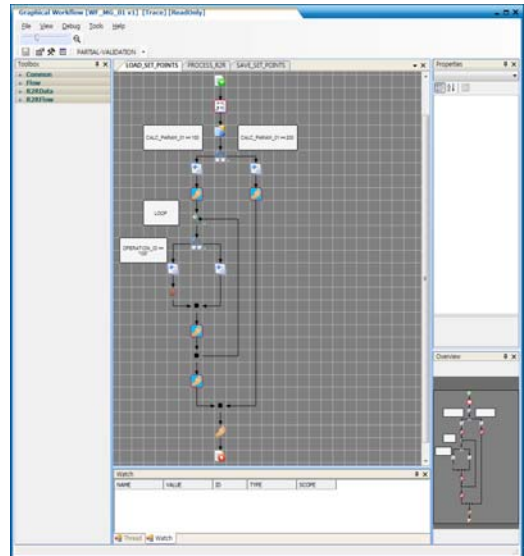
(Figure 2) Activity Properties of Data-Intensive Workflow

## 2.2 The Tracing Mode

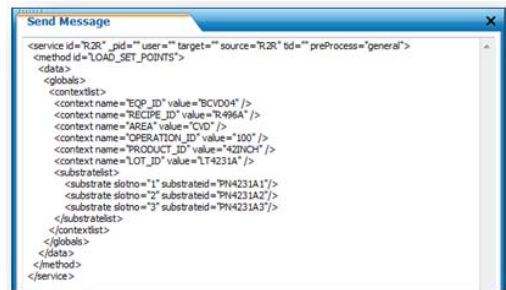
Contrary to editable mode, the tracing mode are support semi-automated execute data-intensive workflow model for validation of workflow which includes control and data-processing with real-time concept. In tracing mode, it can support colored workflow concept like program code debugging for data-intensive workflow designer. So, that has docking area

like canvas for control, data grid for variable expression and thread for workflow instance management. In figure 3, centered area is tabs for canvas and below area are tabs about variables grid and threads.

A data-intensive workflow is started by triggering message. This message is a trigger whenever a user starts workflow. It is following the specified XML data format as application type. Note that, this workflow framework should operate many applications. So, messages are defined different as application specification. In other words, data-intensive workflow models are executed by triggering messages which is depended by each application. Therefore, we try to make and send the message on a general application has data-intensive workflow functions.



(Figure 3) The Tracing Mode on Designer



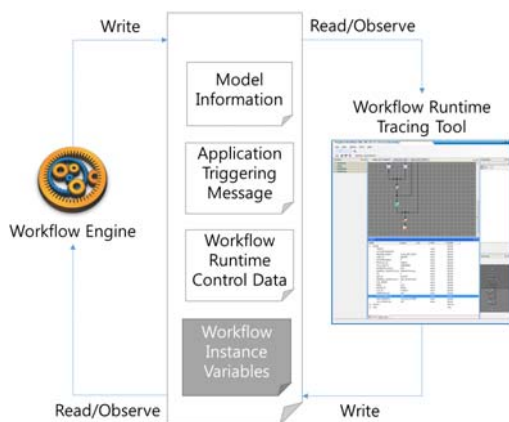
(Figure 4) A Message for Workflow Triggering

### 3. Operation for Data-Intensive Workflow Validation

#### 3.1 Overall Concept of Operation

In order to validate workflow model using graphical method, we start workflow designer by workflow tracing mode. On tracing mode, we could operate and control real-time workflow. There are data for managing and controlling the data-intensive workflow like belows:

- Model information: Workflow model that could be defined by modeling tools and executed by workflow engines.
- Application triggering message: Triggering message could execute that is defined workflow model as workflow application.
- Workflow runtime control data: Control data has break point and current state information. If control data changed by workflow runtime tracing tool, workflow engine should be observed.
- Workflow instance variables: workflow instance variables are created and modified by workflow engine that run workflow step by step. Therefore, workflow runtime tracing tool observe the file during workflow running time.



(Figure 5) Expression of Break Activities with Red Colored Icons

The workflow engine and a workflow tracing mode (workflow runtime tracing tool) communicate via shared workflow files. In particular, there are meaningful ‘workflow instance variables’ file has information about workflow runtime data should be checked.

#### 3.2 Colored Workflow Models

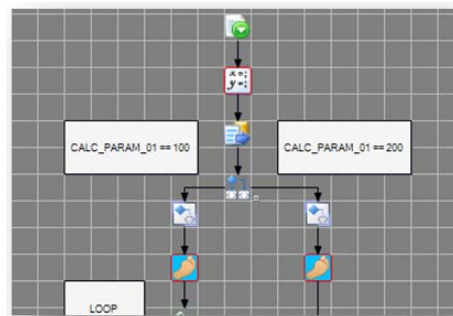
This data-intensive workflow designer could support a special type of workflow model, which is called a colored workflow model[18]. So, we would define activity states as colored icons.

- Red colored icon(s): Expression of the break activity on run-time workflow.
- Yellow colored icon(s): Expression of current activity on run-time workflow.

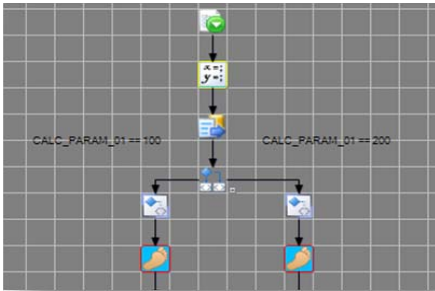
In order to operate workflow trace, it needs to define commands such as start, stop, step over and continue.

- Start: Starting workflow trace with a triggering message.
- Stop: Stopping workflow trace on break activity.
- Step over: Moving the next step from current step.
- Continue: Moving the next break activity from current step.

After workflow real-time tracing start with the message, workflow run until break-point. Figure 6 is expression about break activities of a workflow model with red colored icons (xy: variables icon and foot-print icons), and figure 7 is expression about current holed activity with yellow colored icon (xy: variables icon) for the workflow real-time tracing.



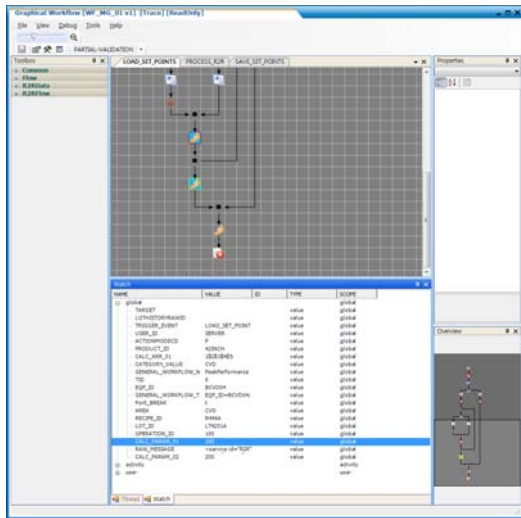
(Figure 6) Expression of Break Activities with Red Colored Icons



(Figure 7) Expression of a Current Activity with Yellow Colored Icons

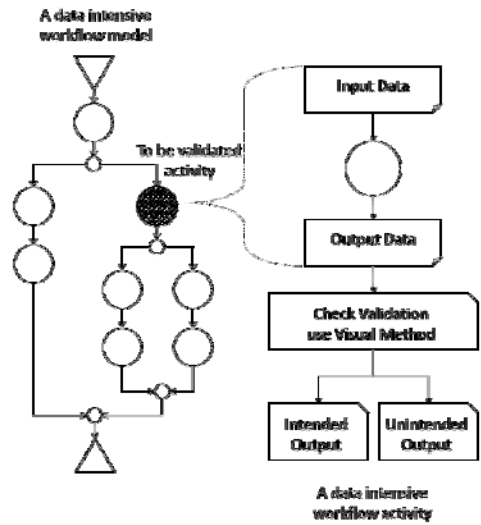
### 3.3 Variables Check

In real-time tracing mode on workflow designer, workflow instances are able to express by a workflow thread, which could watch workflow information related with workflow-data and variables. In figure 8, we can watch workflow-data and variables of the workflow.



(Figure 8) Workflow Data and Variables of Real-time workflow

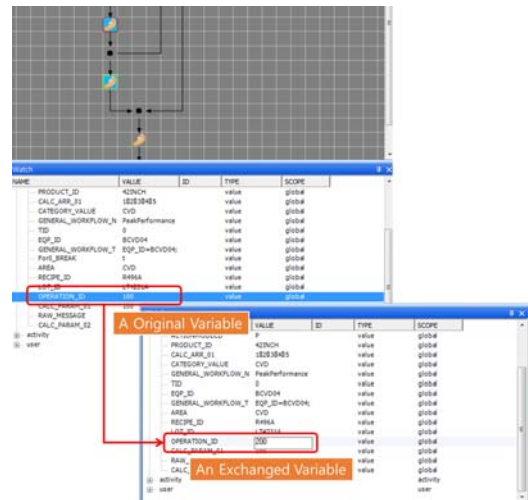
In order to validate an activity of workflow models, we would like to check variables that are input and output data by visual approach. If output data are intended from input data for the activity, the activity are validated. We could check with visual approach on real-time workflow tracing of workflow model designer.



(Figure 9) Workflow Validation with Activity Data

### 3.4 Variables Exchanges

In order check workflow logics correctly, occasionally, it should be changed a variable of workflow. When hold time, it would be possible by changed the file which manage workflow variable data. This function is very useful for validation runtime without workflow control logic exchange or triggering message exchange.



(Figure 10) A Variable Exchange of Data-Intensive Workflow

## 4. Related Works

[10] describes data flow and validation on business process. And [11] approach workflow validation with workflow net theory. And, [12, 13] describes verification and validation of grid workflow management system for modelling, redesign and execution of large-scale sophisticated scientific and business processes in many complex e-science and e-business applications. [14] is focused on rule-based workflow validation like business-to-business workflow interoperability across virtual organizations. [15] was researched validation framework for e-science business processes. There were research for validation and verification on business process and data process workflow management. And, [16, 17] has each workflow simulation tools information but, that describes business process area, mainly. In visualization of workflow informations, [18, 19, 20] are focused on workflow information analysis like social resources, human resources or system resources. Therefore, data-intensive workflow validation with visualization method is rare. So, this research and implemented results are worth.

## 5. Conclusions

So far, this paper has explained about a validation method with the tracing mode of workflow designer and its descriptions that include configurations and operations of the data-intensive workflow. Based upon the tracing mode on workflow model designer, we are able to create or modify workflow models as validated. We so believe that this work might be one of good contributions for improving and advancing the data-intensive workflow management technology.

In future work, we will attempt to advance and improve functions of data-intensive workflow validation.

## References

- [1] Hollingsworth, David, and U. K. Hampshire. "Workflow management coalition the workflow reference model." *Workflow Management Coalition* 68, 26, 1993.  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.198.5206&rep=rep1&type=pdf>
- [2] Barker, Adam, and Jano Van Hemert. "Scientific workflow: a survey and research directions." *Parallel Processing and Applied Mathematics*. Springer Berlin Heidelberg, 746-753, 2008.  
[http://dx.doi.org/10.1007/978-3-540-68111-3\\_78](http://dx.doi.org/10.1007/978-3-540-68111-3_78)
- [3] Yildiz, Ustun, Adnene Guabtini, and Anne HH Ngu. "Business versus scientific workflows: A comparative study." *Services-I, 2009 World Conference on*. IEEE, 2009.  
<http://dx.doi.org/10.1109/SERVICES-I.2009.60>
- [4] Qin, Jun, and Thomas Fahringer. "Scientific Workflows: Programming, Optimization, and Synthesis with ASKALON and AWDL." Springer Science & Business Media, 2012.  
<http://dx.doi.org/10.1007/978-3-642-30715-7>
- [5] Ahn, Hyun, Minjae Park, and Kwanghoon Pio Kim. "sciCN: Scientific information control nets." *Advanced Communication Technology (ICACT), 2014 16th International Conference on*. IEEE, 2014.  
<http://dx.doi.org/10.1109/ICACT.2014.6779140>
- [6] Minjae Park, Hyukjun Na, Ahn, Hyun and Kwanghoon Pio Kim. "A Scientific Workflow Model Designer based on Scientific Information Control Nets." *Advanced Communication Technology (ICACT), 2015 17th International Conference on*. IEEE, 2015.  
<http://dx.doi.org/10.1109/ICACT.2015.7224897>
- [7] Hyung-Jin Ahn and Kwang -Hoon Kim, "Design and Implementation of Very Large-Scale Workflow Management System", *Journal of Internet Computing and Services*, Vol.10, No.6, pp. 205-217, 2009.  
[http://society.kisti.re.kr/sv/SV\\_svpsbs03V.do?method=view&art\\_no=1&id=0&iSet\\_id=22652&totalCount=3&menuid=1&subid=11](http://society.kisti.re.kr/sv/SV_svpsbs03V.do?method=view&art_no=1&id=0&iSet_id=22652&totalCount=3&menuid=1&subid=11)
- [8] Jablonski, Stefan, and Christoph Bussler. "Workflow management: modeling concepts, architecture and implementation," 1996.  
<http://www.citeulike.org/group/6987/article/3401448>
- [9] Taylor, Ian J., et al. *Workflows for e-Science: scientific workflows for grids*. Springer Publishing Company, Incorporated, 2014.  
<http://dl.acm.org/citation.cfm?id=1196459>
- [10] Sadiq, S., Orłowska, M., Sadiq, W., & Foulger, C.. Data flow and validation in workflow modelling. In *Proceedings of the 15th Australasian database conference*, Volume 27,

- pp. 207-214, January 2004. Australian Computer Society, Inc.  
<http://dl.acm.org/citation.cfm?id=1012317>
- [11] Van der Aalst, Wil MP., "Verification of workflow nets," *Application and Theory of Petri Nets 1997*, Springer Berlin Heidelberg, 407-426, 1997.  
[http://dx.doi.org/10.1007/3-540-63139-9\\_48](http://dx.doi.org/10.1007/3-540-63139-9_48)
- [12] Chen, Jinjun, and Yun Yang. "A taxonomy of grid workflow verification and validation." *Concurrency and Computation: Practice and Experience 20.4*, pp. 347-360, 2008.  
<http://dx.doi.org/10.1002/cpe.1220>
- [13] Chen, Jinjun, and Yun Yang. "Key research issues in grid workflow verification and validation." *Proceedings of the 2006 Australasian workshops on Grid computing and e-research-Volume 54*. Australian Computer Society, Inc., 2006.  
<http://dl.acm.org/citation.cfm?id=1151841>
- [14] Haq, I. U., Paschke, A., Schikuta, E., & Boley, H. (2009, May). Rule-based workflow validation of hierarchical service level agreements. In *Grid and Pervasive Computing Conference, GPC'09. Workshops at the IEEE*, pp. 96-103, 2009.  
<http://dx.doi.org/10.1109/GPC.2009.25>
- [15] Wong, Sylvia C., et al. "Provenance-based validation of e-science experiments," *The Semantic Web - ISWC 2005*. Springer Berlin Heidelberg, 801-815, 2005.  
[http://dx.doi.org/10.1007/11574620\\_57](http://dx.doi.org/10.1007/11574620_57)
- [16] Jansen-Vullers, Monique, and Mariska Netjes. "Business process simulation - a tool survey," *Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools*, Aarhus, Denmark. Vol. 38. 2006.  
<http://dx.doi.org/10.1.1.87.8291>
- [17] Van Der Aalst, Wil MP. "Woflan: a Petri-net-based workflow analyzer," *Systems Analysis Modelling Simulation 35.3*, pp. 345-358, 1993.  
<http://dl.acm.org/citation.cfm?id=316234>
- [18] Woo-Jin Jeong and Kwang-Hoon Kim, "A Colored Workflow Model for Business Process Analysis", *Journal of Internet Computing and Services*, Vol. 10, No. 3, pp. 113-129, 2009.  
<http://www.dbpia.co.kr/Article/NODE01210913>
- [19] Hyun Ahn and Kwanghoon Kim, "An Activity-Performer Bipartite Matrix Generation Algorithm for Analyzing Workflow-supported Human-Resource Affiliations", *Journal of Internet Computing and Services*, Vol.14, No.2, pp.25-34, 2013.  
<http://www.dbpia.co.kr/Article/NODE02162420>
- [20] Sungjoo Park and Kwanghoon Pio Kim, "A Closeness Centrality Analysis Algorithm for Workflow-supported Social Networks", *Journal of Internet Computing and Services*, Vol.14, No.5, pp.77-85, 2013.  
<http://www.dbpia.co.kr/Article/NODE02277043>

## ● Authors ●



### Minjae Park

is a senior member of research staff at the solution R&D research center of BISTel, Inc., South Korea. He received B.S., M.S., and Ph.D. degrees in computer science from Kyonggi University in 2004, 2006, and 2009, respectively. His research interests include groupware, workflow systems, BPM, CSCW, collaboration theory, process warehousing and mining, workflow-supported social networks discovery and analysis, and process-aware factory automation systems



### Hyun Ahn

is a full-time Ph.D. student of computer science department and a graduate member of the collaboration technology research laboratory at Kyonggi University, South Korea. He received B.S. and M.S. degrees in computer science from Kyonggi University in 2011 and 2013, respectively. His research interests include workflow systems, BPM, scientific workflow systems, workflow-supported social and affiliation networks discovery, analysis, and visualization.



### Kwanghoon Pio Kim

is a full professor of computer science department and the founder and supervisor of the collaboration technology research laboratory at Kyonggi University, South Korea. He received B.S. degree in computer science from Kyonggi University in 1984. And he received M.S. degree in computer science from Chungang University in 1986. He also received his M.S. and Ph.D. degrees from the computer science department at University of Colorado Boulder, in 1994 and 1998, respectively. He had worked as researcher and developer at Aztek Engineering, American Educational Products Inc., and IBM in USA, as well as at Electronics and Telecommunications Research Institute (ETRI) in South Korea. In present, he is a vice-chair of the BPM Korea Forum. He has been in charge of a country-chair (Korea) and ERC vice-chair of the Workflow Management Coalition. He has also been on the editorial board of the journal of KSII, and the committee member of the several conferences and workshops. His research interests include groupware, workflow systems, BPM, CSCW, collaboration theory, Grid/P2P distributed systems, process warehousing and mining, workflow-supported social networks discovery and analysis, process-aware information systems, data intensive workflows, and process-driven Internet of Things.