
This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: http://openaccess.city.ac.uk/4346/

Link to published version: /10.1109/SOSE.2014.19

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.
Customized Infrastructures for Monitoring Business Processes

Marco Comuzzi
City University London
London, United Kingdom
Email: marco.comuzzi.1@city.ac.uk

Ruben Ivan Rafael Martinez
Eindhoven University of Technology
Eindhoven, The Netherlands
Email: r.i.rafael.martinez@student.tue.nl

Abstract—Process enactment technology provides native tools and add-ons for monitoring, such as APIs and monitoring consoles, which are usually highly entangled with the underlying process enactment logic and not customizable by process users. In such a case, all users access the same set of monitoring data and functions and process management resources may be allocated for monitoring concerns not of interest for users. In this context, we present a model and a tool for customized process monitoring infrastructures executing on top of existing process enactment technology. The model classifies the options about monitoring over which the preferences of process users may diverge. The tool implements the proposed model, generating customized process monitoring infrastructures embedding the business logic of the monitoring options chosen by process users.

I. INTRODUCTION

Modern Workflow Management Systems (WFMS) provide native tools and add-ons for the monitoring of deployed business processes. WFMS log information about processes and executing process instances in a local database. The monitoring tools and add-ons are then built as a higher level interface of the database [1]. Monitoring data can vary from standard information, such as the status of instances and activities of active instances, to any process-specific variable logged by the WFMS, e.g. values of local and global process variables, or specific events, e.g. process variable updates or expiration of timeouts [2].

This approach to process monitoring presents limitations along the flexibility and scalability dimensions. About flexibility, it prevents monitoring customization, as all users get access to a common set of monitoring information tools that they may not need for their specific monitoring purposes. Different users, in fact, may have different monitoring requirements about the same process.

WFMS may also provide monitoring APIs that users can exploit for building their own monitoring infrastructure. Monitoring APIs are normally highly entangled with the underlying process technology [4], which makes them non reusable and hard to understand to a process developer unfamiliar with the underlying process technology. Moreover, APIs can support only the development of ad-hoc monitoring interfaces, which should evolve as the monitoring APIs evolve over time.

About scalability, the monitoring tools and APIs provided by WFMS clearly constitute a bottleneck of the monitoring activity, since all different monitoring requests have to be served by a single point of access to the data logged by the WFMS. In this regard, a non-customizable monitoring infrastructure is likely to lead to a waste of resources. By eliciting the users’ customer requirements before the actual monitoring customization occurs, the business process provider can determine and instantiate only the resources strictly required for monitoring and create synergies among similar monitoring requests, e.g. reuse indexes for common queries on the process logging database. Such a dynamic use of resources is made also possible by modern cloud technology, which allows the provisioning of computing resources on-demand.

Note that increased flexibility of monitoring infrastructures is a pre-requisite for improved scalability. Before it would be possible to efficiently allocate resources to the monitoring concerns requiring it most, it is necessary to disentangle monitoring concerns from the underlying process technology. This will allow encapsulating monitoring concerns into process technology-agnostic services, which can be then deployed efficiently on an elastic computing platform.

This paper focuses on the flexibility issue, leaving scalability to be addressed by future work.

In particular, we devise a solution enabling the business process provider, i.e. the business entity running the WFMS, to provide customized monitoring tools to process users, satisfying their individual monitoring requirements. Monitoring requirements may vary for a variety of reasons, such as the need for monitoring established contracts [5], or predicting their violation [6], [7], or synchronizing internal and external processes of the customer organizations [4], [8]. In this paper we consider the cross-instance monitoring problem. In cross-instance monitoring, customer organizations (users) require monitoring across a set of instances, e.g. monitoring the average execution time of a given activity for all instances of a given process started in the past week. As such, cross-instance monitoring represents the basis of several management control activities, such as Business Activity Monitoring (BAM) [10].

The paper is structured as follows. We first provide a multi-dimensional model for cross-instance business process monitoring to structure the customization solution space, i.e. the space of offered monitoring options over which users’ preferences may diverge (Section II). Then, we provide an implementation of the model in the form of a Web-based application that automatically synthesizes customized monitoring tools for the users (customer organizations) of the WFMS (Section III). Related work is discussed in Section IV, while we draw our conclusions and discuss future work in Section V.
II. A MODEL FOR CROSS-INSTANCE BUSINESS PROCESS MONITORING

The aim of our model is to capture the options for cross-instance process monitoring available to users in a multidimensional space. Such a space will serve as the blueprint for the implementation of the customized monitoring infrastructures specific to each user request. In Section II-A we discuss the rationale behind the design choices leading to our monitoring customization space, while Section II-B presents a formal characterization of our model.

A. Identifying monitoring dimensions for customization

A model for process monitoring customization aims at capturing the possible requirements of users of the process about monitoring, that is, understanding what could be monitored and how monitoring should occur. A first distinction within our model has thus to be made between the monitoring object and the monitoring lifecycle, which constitute the overarching dimensions of our monitoring model. The former identifies the monitoring information that users require from the process engine, whereas the latter identifies the modality according to which such information is acquired by or provided to the user over time.

For understanding the possible sub-dimensions and user options about the monitoring object, we rely on the integration of previous work in the area of Business Process Intelligence [11], BPMN extension for business activity monitoring [12], and on a model for business process analytics provided by the Workflow Management Coalition (WfMC) [1].

The work in [11] set the basis for identifying relevant process monitoring metrics for business intelligence. In the aim of building a data warehouse for business process monitoring, the paper identifies three types of monitoring data, i.e. Time, Status, and Resource data, which can be aggregated along the process granularity dimension. Granularity is defined at the level of process, service, and node. Processes invoke a set of services during their execution, and services are executed on physical nodes. The work in [12] extends BPMN for monitoring purposes identifying status and duration as main monitoring data and refining the granularity dimension considering also the process instance level and aggregation operators across the granularity levels, e.g. average, maximum, minimum, or frequency of occurrence (e.g. for status information, which is intrinsically not ordinal). The work in [1] mainly focuses on the status and time dimensions, providing a more detailed list of time-related variables relevant for monitoring and refining the list of possible statuses characterizing activities, instances, and processes.

To integrate the three proposal discussed above, in our monitoring model we consider time and status related monitoring variables. About granularity, in particular, we consider a simple process management model comprising processes, which are constituted by multiple activities and associated to multiple instances started by users. When an instance is executing, the activities belonging to the corresponding process are instantiated into assigned activities.

As far as status monitoring variables are concerned, our objective is to devise a generic process monitoring model, which can suit the case of both traditional workflow management systems and service-based business process enactment technology. Hence, we rely on the intersection between specifications proposed by the Workflow Management Coalition (WfMC) \(^1\) and in the BPEL language \(^2\), i.e. the de facto standard for service-based business processes.

Figure 1 shows finite state machines for instance and activity status that we consider in our monitoring model. Note that there are there are two final states for an instance: completed, which is reached when an instance is completed successfully, or failed, when the completion has not been successful, i.e. because of an error or because the user aborted the execution. Note that suspended instances can only terminate with a failure, whereas a running instance may either terminate successfully or fail. In particular, we excluded from our model the states defined for instance compensation in BPEL, as these do not find a counterpart in the WfMC specification and, consequently, in traditional workflow technology. Similarly, we excluded the initiated status defined by the WfMC as this is not defined in the BPEL state machine.

There is only one final state for activity execution, because the WfMC specification does not distinguish between states for successful and unsuccessful termination (as an unsuccessful termination should push the corresponding instance in the failed status).

As far as time-related monitoring variables are concerned, the time-related information in our monitoring model can be defined using status changes timestamps. As we will discuss later, in fact, process enactment technology usually logs the timestamps related to status changes of instances and activities. In particular, it has to be noted that, during their lifecycle, instances and activities may assume some states only once, whereas the state running and suspended can be assumed more than once.

For an instance \(i_{s_j}\) we define the individual timestamps \(t_{s_j}, t_{c_j}\), and \(t_{f_j}\) representing the instant in time in which the instance has reached the status terminated, failed, or completed, respectively. Note that, depending on the status in which an instance terminates, \(t_{c_j}\) coincides with either \(t_{c_j}\) or \(t_{f_j}\). Similarly for an activity \(a_{c_j}\) we define the individual timestamps \(t_{i_j}\) and \(t_{c_j}\) representing the instants in time when an activity becomes active and completes, respectively.

The statuses running and suspended, for both activities and instances, we can define a set of timestamps \(T R_j\) and \(T S_j\) representing the time instants in which an instance or activity becomes running or suspended, respectively:

\(^1\)http://www.wfmc.org/reference-model.html

\(^2\)http://pic.dhe.ibm.com/infocenter/dmndhelp/v8r0m1/index.jsp
Table I shows the time-related monitoring variables that we consider in our model. These are a restriction over the set of time-related process control variables defined in [1] on the basis of the states and related timestamps defined for instances and activities in our model.

Table I. Time-related Monitoring Variables.

<table>
<thead>
<tr>
<th>Granularity</th>
<th>Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>instance</td>
<td>Duration</td>
<td>$DU_R(ins_j) = tt_j - tr_j,i$</td>
</tr>
<tr>
<td>instance</td>
<td>Processing Time</td>
<td>$PIT(ins_j) = \sum_{i=0}^{I}(ts_{j,i} - tr_{j,i}) + \begin{cases} tt_j - tr_j,i, &amp; \text{if } \exists tr_j,i, \ 0, &amp; \text{otherwise} \end{cases}$</td>
</tr>
<tr>
<td>instance</td>
<td>Suspended Time</td>
<td>$SIT(ins_j) = \sum_{i=0}^{I}(tr_{j,i+1} - ts_{j,i}) + \begin{cases} tf_j - ts_j,i, &amp; \text{if } \exists tf_j, \ 0, &amp; \text{otherwise} \end{cases}$</td>
</tr>
<tr>
<td>activity</td>
<td>Duration</td>
<td>$DU_R(act_j) = tc_j - tr_j$</td>
</tr>
<tr>
<td>activity</td>
<td>Waiting Time</td>
<td>$WAT(act_j) = tr_j - ti_j$</td>
</tr>
<tr>
<td>activity</td>
<td>Turnaround time</td>
<td>$TUT(act_j) = tc_j - tr_j$</td>
</tr>
<tr>
<td>activity</td>
<td>Processing Time</td>
<td>$PAT(act_j) = \sum_{i=0}^{I}(ts_{j,i} - tr_{j,i}) + (tc_j - tr_j,I)$</td>
</tr>
<tr>
<td>activity</td>
<td>Suspended Time</td>
<td>$SAT(act_j) = \sum_{i=0}^{I}(tr_{j,i+1} - ts_{j,i}) + \begin{cases} tt_j - ts_j,i, &amp; \text{if } tt_j &gt; tr_j,i, \ tr_j,i - ts_j,i, &amp; \text{otherwise} \end{cases}$</td>
</tr>
</tbody>
</table>

$TR_j = \{tr_{j,i}\}_{i=1,\ldots,I}$, with $tr_{j,i+1} > tr_{j,i}, \forall i$ \hspace{1cm} (1)

$TS_j = \{ts_{j,k}\}_{k=1,\ldots,K}$, with $ts_{j,k+1} > ts_{k,j}, \forall j$ \hspace{1cm} (2)

For identifying the possible user options about the monitoring lifecycle, we refer to our previous work on customizable single-instance process monitoring [9]. In that work, we identified a standard lifecycle for business process monitoring, leading to the definition of two dimensions over which users can specify their options about monitoring, i.e. Management and Notification.

**Management.** Users may express options about the way monitoring data will be managed by the monitoring infrastructure before being provided. For instance, monitoring information can be stored in a batch or provided as soon as a new value is captured. The data may be stored, for instance to provide historical series to the user, or destroyed in the monitoring infrastructure once communicated to the user. Specifically, new values acquired can overwrite existing values (rewrite - rw) or being persisted (persist - pe) by the monitoring infrastructure. When notified to the user, monitoring values currently stored by the monitoring infrastructure may be consumed (co), i.e. they will be no longer available in the future, or they may be only read (re) and remain persisted by the monitoring infrastructure, e.g. to build historical time series. The options available to the user for the management monitoring design are derived from the combination of the modalities discussed before, that is, we identify a total of four possible options, i.e. rw-co, rw-re, pe-co, pe-re.

**Notification.** According to a well-recognized paradigm in distributed computing [13], there are two main ways monitoring data can be notified to the user. Users may pull information as they wish, or they may be pushed information by the monitoring infrastructure according to a pre-specified policy, e.g. periodically.

### B. Formal model

The monitoring model discussed in the previous section can be formally characterized as presented in Table II. The formalization of Table II gives us a model to specify succinctly the monitoring customization options of users. Such a representation of users’ monitoring requirements is exploited by the implementation of our tool (see Section III).

Specifically, a monitoring customization $c_{i,j}$ is a request to the process engine of the provider organization to monitor a certain (set of) variable(s) $var$ over a possible restriction of the processes currently deployed within the engine. Such a restriction is defined on features characterizing the elements of the process management model introduced before, i.e. processes ($pro$), instances ($ins$), and activities ($act$).

For the $pro$, $ins$, and $act$ sub-dimensions, we consider the following restriction options:

- **ALL**: it refers to all the processes/instances/activities currently associated to the user $u_i$ by the process engine;
- **[USER_SELECTION]**: it restricts the set of processes/instances/activities to a list specified by the user, e.g. a set of process IDs;
- **[TIME_FRAME]**: it restricts the set of processes/instances/activities to the ones deployed/started/allocated within a specific time frame, i.e. before and/or after a specific date;

For the elements $ins$ and $act$ we consider also the following restriction options:

- **[STATUS_BASED]**: it restricts the set of instances/activities to the ones in a certain status at the moment a monitoring request will be issued to the process engine;

Note that the monitoring options are not mutually exclusive. For example, a user can specify options on both **TIME_FRAME** and **STATUS_BASED** to restrict monitoring of a certain variable to all instances (and/or activities) started in a certain timeframe and currently in a certain status, e.g. running or terminated.
As previously discussed, the monitoring variables can relate to the status of instances, \( i_{\text{status}} \), and activities, \( a_{\text{status}} \), and to the time-related variables of Table I.

Eventually, the \( agg \) element in \( MO \) specifies the possible aggregation of monitoring data required over the restrictions applied. We adopt in this work the aggregation operators of the SQL language, i.e. \( agg \in \{\text{sum, avg, max, min, count}\} \). Note that not all combinations of options are feasible. For instance, it is not possible to apply aggregation operators such as \( \text{sum} \) or \( \text{avg} \) to non-numeric values, e.g. activity or instance status. Similarly, operators \( \text{max} \) and \( \text{min} \) can be applied only to monitoring variables for which a total order relationship is defined, i.e. all the time related monitoring variables.

As far as the monitoring lifecycle \( ML \) is concerned, the Notification dimension \( not \) involves two possible possible options. The user can in fact \( \text{pull} \) data from the monitoring infrastructure when needed, or the monitoring infrastructure can \( \text{push} \) monitoring data to the user according to a certain policy. In this work we consider only periodic push of monitoring data. Hence, the \( policy \) element represents the notification period, expressed in time units. For the Management dimension \( man \) we consider the four options discussed in the previous section.

As a sample, Eq. 3 captures the monitoring requirements of user \( u_1 \) interested in counting the number of activities in the instances of process \( p_1 \) started before January 1st and terminated in the status \( \text{failed} \); \( u_1 \) wants to pull the information when required from the monitoring infrastructure and requires the monitoring information to be persisted by the monitoring infrastructure as long the customization is active (therefore, \( man = \{\text{pe}, \text{re}\} \)).

\[
e_{1,1} = (MO, ML) \\
MO = \{a_{\text{status}}; p_1; \} \\
< 2013 - 01 - 01T00: 00: 00 + 00: 00; \quad \text{(3)} \\
ALL; \text{failed}; \text{count} \\
ML = \{\text{pull}; \text{pe}, \text{re}\}
\]

Eq. 4 captures the requirements of users \( u_2 \) interested in monitoring the average processing time of instances of all instances of the process \( p_2 \) terminated in the status \( \text{completed} \). The user wants to be notified this information every six hours. Each time, the user requires only the average processing time of instances completed in the last six hours (that is, the user is not interested in building time series of monitoring data; \( man = \{\text{re}, \text{co}\} \)).

\[
e_{2,1} = (MO, ML) \\
MO = \{\text{pIT}; p_2; \text{ALL; -; completed}; \text{avg}\} \\
ML = \{\text{push(6hr)}; \text{re}, \text{co}\}
\]

### III. A TOOL FOR CUSTOMIZED CROSS-INSTANCE PROCESS MONITORING

In this section we describe the tool we implemented. Our tool generates customized process monitoring infrastructures exploiting the monitoring model described in the previous section. We first provide a conceptual outline of our tool, in Section III-A, and then discuss its technical implementation in Section III-B.

#### A. Conceptual Architecture

The conceptual architecture of the tool is shown in Figure 2(a). The design aims at isolating components for each of the monitoring dimensions in the monitoring model of Table II. Conceptually, user interaction with the system can occur at customization time, i.e. when monitoring customizations are requested, and at run time, i.e. when the monitoring starts [14].

At customization time, users access the Monitoring Customization Console (MCC). This comprises the Customization Interface (CI), through which users specify their customization options. Customization requests are received by the Monitor Generator (MG), which instantiates a Custom Monitoring Infrastructure (CMI) for each monitoring customization request. The CMI contains the Monitoring Console (MC), through which users get access to the monitoring data according to the monitoring options specified in the customization. It also comprises three components in charge of realizing the business logic of the monitoring options identified in our model:

- **Query Engine (QE):** It instructs the BPE Handler to capture the appropriate monitoring variable(s) requested by the user in the customization;
the user can interact with our tool. The MG represents the **controller** in our architecture, connecting the interfaces to the CMI, which implements the **model** driving the cross-instance process monitoring, i.e. the monitoring options described in Section II.

**B. Technical Realization**

The tool has been implemented as a Java Web application, using Struts 2 as a framework supporting the MVC design pattern. We experimented with the Apache ODE BPEL engine, which is freely available and open source, that is, a specific BPE Handler has been developed for Apache ODE.

Figure 2(b) shows the technical architecture of our tool. The MG component in the conceptual architecture is implemented by the controller component. It comprises the struts.xml file, which is used to configure the appropriate actions in the model component to implement the monitoring options requested by the user. The controller implements the business logic of the customized monitoring infrastructure. Specifically, the BPE Handler issues query over the Apache ODE database, whereas the Management Engine component controls a local database where data acquired from Apache ODE may be stored in case the user requests monitoring data to be persisted.

The view part of our tool comprises a set of dynamic Web pages, which are configured by the data requested through the actions implemented in the controller. The tool comprises a page for the login of the user and pages implementing the Customization Interface and the Monitoring Console.

As introduced before, the main objective of the BPE Handler is to translate a monitoring customization expressed according to the formal model of Section II-B into a query for the Apache ODE database logging monitoring information. In particular, the translation process concerns the monitoring object \( MO \) element of a customization \( c_{i,j} \).

A typical strategy adopted by current process management technology to log relevant process monitoring information is **event-based**. Relevant monitoring information in this case is captured in events logged by the BPE. Events refer to entities of the process management model, have a timestamp, and log different types of occurrences in the process engine, e.g. status changes, variable value changes, faults. An event-based logging strategy is followed by most process engines available on the market. As far as open source and freely available engines are concerned, this strategy is for instance adopted by the Apache ODE BPEL engine and YAWL. The event-based strategy is opposed to the **entity-based** strategy, adopted for instance by the OpenESB BPEL engine. The limitation of this strategy is that monitoring information is stored at the level of individual entities of the process management model, i.e. processes, instances, and activities. This prevents logging multiple status changes timestamps (in OpenESB, for instance, only the last status change is persisted at a given moment in time).

**IV. RELATED WORK**

Data warehousing and OLAP (On Line Analytical Processing) [11] exploit the concept of cross-instance monitoring, but
taking a rather static perspective about monitoring, supporting the more static use case of executive decision making. In this work, we focus on a more dynamic perspective about monitoring, supporting scenarios such as cross-organizational business process synchronization [15] or online Business Activity Monitoring (BAM) [10].

The Monita [18] approach to process monitoring highlights the issue of entanglement between monitoring concerns and workflow specification and it proposes to implement the former as aspects blended within workflow specifications. The definition of monitoring aspects has to be performed by monitoring experts and cannot be delegated to process users, since the set of possible monitoring concerns is not abstracted into a high-level model understandable by non-expert users. A framework for ECA rules to support advanced process monitoring is presented in [19]. Also in this case, the expertise required for the specification of ECA monitoring rules prevents the delegation of the monitoring customization activity to process users.

Several approaches [2], [20], [21] to business process monitoring have been developed in the context of service-based business processes and, more specifically, processes specified using BPEL. An architecture for self-supervision of BPEL processes has been proposed in [2]. It involves the definition of aspects for the analysis and recovery of running BPEL processes, which are blended at runtime within the Active BPEL engine execution logic. This approach represents a typical example in which the resulting monitoring logic is highly entangled with the chosen process engine technology. The language defined for the definition of monitoring concerns, i.e. wsCol, is more expressive than our model of monitoring customization. Hence, it could be exploited to extend the range and complexity of user-defined monitoring variables and concerns of this work. A visual approach to specify monitoring queries on running BPEL processes is presented in [20]. Visual queries are then translated into one or more XPath and SQL queries on the BPEL process specification and/or data logged by the BPEL engine. Although, also in this case, the range and complexity of monitoring concerns is higher when compared to our formal model, the approach is specific only to BPEL processes and does not specifically address the need for customizing the resulting monitoring infrastructure.

V. CONCLUSIONS AND OUTLOOK

This paper considers the issue of customized business process monitoring infrastructures in the context of cross-instance process monitoring. The design of our tool decouples the monitoring concern from the underlying process engine, making our approach generic, i.e. replicable using alternative process enactment technology. In this work we focused on the flexibility of current process monitoring technology. As stated in the Introduction, our future work will also concentrate on the scalability issue. In particular, we plan to work in two different directions. A first optimization will concern the access to the BPEL database, such as indexes for highly requested monitoring configurations. A second optimization will concern the deployment of our Customized Monitoring Facilities on a cloud infrastructure. Individual CMIs can be deployed on virtualized appliances using only the resources that are strictly necessary, improving the performance of our tool.

REFERENCES