Evaluating an Event-Based Approach to Workflow Services

Christina Webb
Academic Affiliation, fall 2003: Graduate Student, Michigan State University

SOARS® Summer 2003

Science Research Mentor: Aaron Andersen
Science Writing Mentor: Gayl Gray

ABSTRACT

Workflow management is crucial in monitoring and controlling business processes. Advancements have been made in computer-based workflow management, which allowed for the partial or complete automation of these processes. However, existing computer-based systems have not been successful in highly distributed domains. Hence, these domains have continued conducting their workflow management processes in an adhoc manner. Conducting workflow management in this way has three flaws: human dependency, confined knowledge, and inconsistency in tasks. Metis, an event-based workflow service, was developed to alleviate these flaws by bringing automation to workflow management in highly distributed domains, primarily digital libraries. This project conducted an evaluation of Metis using two data retrieval processes. Each workflow carried out via Metis soundness using a petri net analysis; in addition performance analysis was also conducted. Soundness tested Metis’ ability to handle exceptions and/or errors that may occur. The performance analysis is a qualitative measurement based on two major metrics: quality and efficiency. This evaluation found that while Metis’ approach is valid it needs further modification in areas requiring the automation of repetitive tasks such as multiple file transfers. Furthermore, the qualitative analysis showed an improvement in conducting workflow management using Metis versus in an adhoc manner.

This work was done under the auspices of the Significant Opportunity in Atmospheric Research and Science program of the University Corporation for Atmospheric Research, with funding from the National Science Foundation the U.S. Department of Energy, the National Oceanic and Atmospheric Administration, the Cooperative Institute for Research in Environmental sciences, and the National Aeronautics and Space Administration.
INTRODUCTION

When conducting any assignment a workflow plan is designed, either mentally or visually, to serve as a list of separate tasks to be carried out resulting in the completion of the original assignment. Technology advancements have allowed for workflow plans to be either partially or completely automated, reducing human dependency for the completion of the assignment. However, automated workflow technologies have not been successful in highly distributed domains. One reason for this is the emphasis placed on the task to be completed in a set order with no flexibility. Another reason is the assumption was made that all stakeholders belong to the same organization, which implies compliance to the same policies and procedures [Anderson et al., 2003]. For these two reasons existing workflow systems are not flexible and consequently restrict their use within highly distributed domains.

An example of a highly distributed domain in which current automated workflow systems cannot be utilized is the digital library domain. Within this domain, consider the case when professors from different universities (for our purposes we will call them University A and University B) are reviewing a paper for publishing on a digital library. In this case there are several steps that have to be completed in order to reach the final result, a published article. However, University A and University B are not on the same computing system and possibly do not work with the same software packages. Here we see that assuming all stakeholders are members of the same organization prohibits the use of current workflow technology in this situation. Consequently, within these highly domains all workflow management tasks have been conducted in an adhoc manner. By “adhoc”, we are referring to carrying out tasks in a non-uniform sequence of scripted and manual techniques. Conducting workflow management in this way has three inherent flaws: possibility of error due to human dependency, knowledge confined to a sole person, and inconsistency in tasks.

In an effort to alleviate the aforementioned flaws, Metis workflow services has developed a workflow technology that can be used in the digital library domain. Metis has the ability to extend automated workflow management processes by centering its model on events rather than tasks and avoiding the restrictions present in current automated workflow systems. Greater detail on an event-based approach will be presented in a later section. With the technology that Metis offers its developers hope to bring uniformity in executing tasks as well as shared knowledge to the community surrounding each particular workflow. Metis has the potential of improving workflow management by reducing time spent on each workflow, reducing human error, and bringing uniformity to each workflow [Anderson et al., 2003].

While the benefits of having an event-based workflow service system have been made evident, this project has conducted an evaluation to assist in validating the usefulness of Metis. We carried out the evaluation by completing two separate data retrieval processes, using both Metis and the existing adhoc procedure. The first process being evaluated is that of receiving, verifying, and uploading weather data from the National Center for Environmental Prediction (NCEP). The second process is similar to the previous process, except it collects sea surface temperature (SST) data obtained from the National Oceanic and Atmospheric Administration (NOAA). Both of these processes are carried out by the Data Support Service (DSS) section of the National Center for Atmospheric Research (NCAR). At the completion of each workflow we compared the Metis procedure to its respective adhoc procedure using a Process Improvement Matrix (PIM) [Altizer et al., 1995]. To further strengthen this evaluation we verified the structure of the two workflows as
constructed in Metis by testing for soundness (which will be defined in the Evaluation Methods section) [van der Aalst, 1998].

**METIS BACKGROUND**

Metis has four features, which make it suitable for highly distributed domains. Metis is web-based, lightweight, flexible, and event-based. Each of these features aid in correcting the flaws of conducting workflow management in an adhoc manner, as well as, alleviating the organizational assumptions made by current automated systems [Anderson, *et al.*, 2003]. Below is an explanation of each feature:

**Web-based:** Since Metis is Web-based it is accessible by most organizations. With over 35 million Web servers detected in the December 2002 Netcraft Survey, it is reasonable to assume that most organizations have access to Web browsers if not Web servers [Anderson, *et al.*, 2003]. By utilizing the web, Metis allows for the execution of workflow events across multiple organizations.

**Lightweight:** Metis is referred to as lightweight because it has a low entry barrier. In other words Metis does not have extensive software requirements, which require installing, configuring, and deploying multiple complex packages.

**Flexible:** Workflow services need to be flexible and customizable in the sense that the workflow technology will conform to the organizations and not the other way around. Metis also offers flexibility to support the highly distributed domains by basing each workflow on events rather than tasks.

**Event-based:** Since Metis is event-based, it specifies the events which must occur for a workflow to progress, but does not require that the completion of the events be carried out in any particular order [Anderson, *et al.*, 2003].

**METIS FRAMEWORK**

The framework of Metis caters to the set goals of being web-based, lightweight, flexible, and event-based. When discussing the framework of Metis, we first must note that Metis is hosted by an application server, Tomcat, which is layered on top of an event notification system, Siena. Metis is composed of six Java servlets, which are composed in a three-layer structure. These servlets are the user registry, event editor, workflow editor, action editor, execution engine, and a workflow monitor [Anderson, *et al.*, 2003].

Figure 1 is a high level architectural view of the Metis system. Each of the previously mentioned servlets can be found on one of the three layers. Other components of Metis can be found in this illustration such as the event announcer, workflow status, etc. These are components, which are controlled by one of the six Java servlets and help the user monitor and control the workflow process. The lowest layer of the framework is where the event system is found. This system receives event notification and sends them to the workflow management layer, the second layer. The workflow management layer handles the interpretation of received events, the execution, interaction and archival of workflows and processes actions associated with events. The tool layer is responsible for supporting the editing and creation of events, workflows, actions, users, groups, and roles [Andersen, 2003].
Each of these Java servlets holds certain responsibilities throughout each process as shown here:

**User profiles:** User profiles are created in the user registry. Each profile consists of first and last name and organization affiliation; additionally groups and roles are specified for each individual. Each user is given different levels of access to Metis according to his/her group or role.

**Event editor:** The event editor is a user interface in which the definition of events can be created, modified and deleted. There is a characteristic within the event editor, which allows the user to set a default attribute for events if an attribute is not supplied.

**Workflow editor:** A user can create, edit, and delete workflow definitions within the workflow editor. The workflow editor represents the process being carried out as a tree structure. The workflow editor presents the tree structure and enforces preset constraints built into Metis.

**Action editor:** The action editor is used after the workflow has been defined. This editor is used to define actions that are associated with each of the Event nodes. Some actions include sending email, creation of directories, and/or uploading (downloading) files.

**Execution engine:** The execution engine is responsible for managing instances of workflows while the process is being carried out. Some specific roles of the execution engine include activating and deactivating workflows and retrieving the status of active workflows. The execution engine is also

---

**Figure 1. A high level architectural view of Metis system.**
responsible for invoking an event’s actions once the event has been delivered using an event notification.

**Workflow monitor:** Users are able to determine the status of workflows using the workflow monitor and specified users can activate and deactivate workflows. This monitor allows users to publish events to active workflows [Anderson, *et al.*, 2003].

Figure 2 displays the architecture of Metis within the complete executing environment (i.e. the world wide web). The user begins the process via the web, which enters the Tomcat server to initiate Metis. Once in Metis the workflow is created and particulars to each workflow (i.e. events, actions, etc.) are edited and maintained. Defined workflows are then instantiated. The event server allows for events to be published or posted via the web. The Internet icon is the portion of the architecture, which connects users across multiple organizations. If this diagram were duplicated to represent additional organizations working on the same workflow, they would connect at the Internet icon [Anderson, *et al.*, 2003].

**Figure 2.** The architecture of Metis within its complete computing environment.
EVALUATION METHODS

Verification

In order to conduct an evaluation of Metis we had to choose processes to be carried out using Metis. After choosing two particular processes (which are discussed in the next section) we had to verify that they were complete. To do this we chose to depict each process as a petri net, which is a mathematical model of any process. Petri nets are used because of their clear, precise depiction of a workflow. Furthermore, there has been extensive research on petri net analysis, which this study was able to utilize. Figure 3 is a simple example of a petri net. In this particular example, note the construction of the petri net. The large hollow circles represent a given state (also referred to as a place) in the workflow. In order to go from one state to the next certain preset conditions must be met and a transition occurs. The broad black bars denote transitions. The solid black circles are referred to as tokens. Tokens denote the progression of the workflow by acting as a “place holder”. A token is placed in a particular place if the preset conditions are met.

Figure 3. an example of a general process depicted using petri nets.

As mentioned earlier, extensive studies have been conducted on petri net analysis. For this study we used soundness (Definition 1) to test each workflow for completeness. Soundness is a minimal requirement that should be met in order to ensure the proper execution and completion of workflows [van der Aalst, 1998].

Definition 1. A procedure (workflow) is considered sound if and only if

1. For every state, \( m \), reachable from the state, \( i \), there exists a firing sequence leading from state \( m \) to the terminating state, \( o \)
2. State \( o \) is the only state reachable from state \( i \) with at least one token in place \( o \)
3. There are no dead transitions, i.e. state \( m \) leads to some state \( m' \) until \( m' = o \) [van der Aalst, 1998].

The soundness property explores whether all possible states are considered. If all parts of definition 1 are covered this implies that the workflow is in fact complete and once the workflow has begun there is some terminating state, regardless of what path is taken.
Performance Analysis

The second aspect of this evaluation is a performance analysis, which is primarily a qualitative evaluation of Metis’ ability to meet certain requirements. The two major qualitative metrics are quality and efficiency. Each of these metrics was assessed using a Process Improvement Matrix (PIM). PIM is a tool used to continuously monitor and evaluate a process as qualitative improvements are made [Altizer et al., 1995]. For this project the PIM was constructed to show any qualitative comparison discovered in carrying out each data retrieval process in an adhoc manner versus using Metis (Table 1).

Table 1. An example of a Process Improvement Matrix (PIM), for Data Support Services (DSS) Sea Surface Temperature (SST) retrieval, which was used in the performance analysis.

<table>
<thead>
<tr>
<th>PIM For DSS/SST Retrieval</th>
<th>Adhoc Process</th>
<th>Metis Process</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error handling</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Efficiency</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turn-around time</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When evaluating each process, overall quality will be determined by Metis’ ability to maintain or improve the perceived value of workflow management received using the current adhoc process. For further analysis the metric quality has been divided into more specific sub-metrics: satisfaction, quality of output, and error handling. These subcategories were rated so that any problematic (or exceptional) areas can be pinpointed.

Efficiency was based on the rate of completion of each process. This metric has subcategories of total time and turnaround time. Total time was rated according to the total time of completion once the first step of the process begins. Turnaround time was measured from the time the data is made available for retrieval to the time the data is uploaded onto the NCAR website.

Both metrics were rated based on interviews conducted with Joey Comeaux of the DSS section of NCAR. Comeaux oversees the processes being used in this evaluation; hence he is very familiar with the adhoc procedures currently used in these two processes. Comeaux was consulted on Metis and involved with the execution of Metis’ workflow procedures, which makes him an eligible candidate for interviewing. To keep consistency in the rating of both the adhoc procedures and the Metis procedures, Comeaux’s interview will be the basis for the color-coding measurements of each procedure.

The measuring technique for the PIM is a color-coding based on how much attention should be given to each aspect. For all purposes of this project the codes will be r-red, y-yellow, and g-green; where red represents those areas needing immediate attention, yellow corresponds to areas of concern, and a green code will be given to areas that do not need much attention, if any at all. Since the yellow rating corresponds to a wide range of levels of concern, any metric given a yellow
coding was further measured by a given weight. The weight of the yellow coding was indicated by percentage of fulfillment of that particular metric. For example a metric given a rating of y-8 indicates a yellow rating with 80 percent fulfillment, which would correspond to an area that is closer to a green rating as opposed to a red rating [Altizer et al., 1995].

PROCESS DESCRIPTIONS

Weather Data Retrieval

The process for weather data retrieval is conducted by DSS on an interrupt basis. The data is received from the National Center for Environmental Prediction (NCEP). NCEP has two active satellites obtaining two readings daily. DSS attempts to collect all the daily data on a monthly basis (the 15th of each month). For the remainder of this paper this process will be noted as DSS/NCEP.

The process will begin using CRON, a program that is used to execute a script at a future date. CRON will execute a script that will obtain the available data on the fifteenth of each month. As previously mentioned there are two active satellites, therefore when retrieving data we must then determine which satellite the data is coming from. We would then specify the year and month of the data we would like to retrieve. The next few steps include using FTP to get the information from NCEP to local NCAR directories. The data must then be verified for completeness. The final step is to send notification in a long format listing of the data to the supervisor of the process.

Below a flowchart of this workflow as depicted by Metis is shown (Figure 3). Figure 3 shows the individual tasks along with events and action triggers that will be fired throughout the process.

Figure 3. Flowchart of tasks that must occur for the weather data retrieval process.
As mentioned earlier Metis recognizes each workflow in a tree structure. This structure displays each event independently reemphasizing the event-based approach. Figure 4 depicts the DSS/NCEP process as configured in Metis.

![Tree Diagram](image)

**Figure 4.** tree diagram depicting Metis’ view of the weather data retrieval process

### Sea Surface Temperature Data Retrieval

The process for SST data retrieval is conducted by DSS upon request. The data is received from the National Oceanic and Atmospheric Administration (NOAA). NOAA uploads the data on a weekly and monthly basis into annual files. Currently DSS does not retrieve SST data regularly. For the remainder of this paper this process will be noted as DSS/SST.

The DSS/SST workflow is comparatively simpler than the DSS/NCEP process. When the DSS/SST workflow is executed it must receive the SST data from NOAA. Once the data is gathered we must determine what format was used for the data. With each retrieval of data the file is overwritten, therefore file size is used to ensure the data was received (versus retrieving an empty file). Once this is complete an email is sent to the supervisor of the DSS/SST process reporting the readings. Figure 5 shows the flow of tasks associated with this process. Again, Metis recognizes workflows in a tree structure, which is displayed in figure 6 for this particular workflow.
RESULTS AND DISCUSSION

Verification

When testing for soundness we were investigating the completeness of each workflow. In other words we wanted to make sure that no matter what path was taken to get to some point in the workflow, there is a sequence leading to the terminating state.

Figure 6 shows the petri net construction for the weather data retrieval process. This is a snapshot of the workflow after leaving the loop receiving 60 days of data. At this point the next
transition, E₁ is complete and we can transition to the next state. This workflow met all three conditions of soundness; hence we consider this workflow to be complete.

When testing the weather data retrieval process for soundness we found that complete automation of each workflow was not possible. In each process the data was sent to specific local directories according to date and, in the earlier case, satellite. The step that requires the numerous file transfers required human interaction to input the path name.

Below, figure 7 shows the petri net display of the SST data retrieval process. Again, we see this process is also sound according to the definition given before. The snapshot of the workflow shown in figure 7 is that of the initial state of the workflow, which is waiting for the data to be made available.

Figure 6. Petri Net representing weather data retrieval workflow. This is a snapshot of the workflow once 60 days of data are collected.

Figure 7. Petri Net representing sea surface temperature data retrieval process. This is a snapshot of the workflow in its initial state.
Performance Analysis

After conducting several interviews with Joey Comeaux of the Data Support Services (DSS) section we were able to give color-coded ratings to each metric for both the adhoc and Metis processes. As mentioned before g-green is assigned to those areas needing little or no attention, y-yellow is assigned areas of concern (with additional weighting), and r-red is reserved for areas needing immediate attention.

Table 2 is the Process Improvement Matrix (PIM) for the weather data retrieval process (DSS/NCEP). First let us explore the ratings of the current adhoc process. Overall quality was given a yellow coding with 50 percent compliance based on the ratings of the subcategories. Satisfaction was given a yellow rating with 20 percent compliance because Comeaux would be more satisfied if the process were automated which would free time for other activities. The final output was given a green coding because the final outputs are adequate. Error handling was given a yellow rating with a weight of 80 percent. The error handling of the adhoc process is an area of concern because when a problem occurs there is an extensive process that must take place for the correction. Efficiency for the adhoc process is given a yellow coding with 30 percent weight. The total time it takes to complete the process is an area of no concern, however the turn-around time (time from when data is first made available to the time it is uploaded) is an area needing immediate attention.

As mentioned previously, Metis did not have some key automation attributes, which negatively affected the quality ratings. Satisfaction received a yellow rating with 10 percent fulfillment because all parts of the workflow were not automated. Error handling has improved, increasing to a 90 percent weight, because Metis will download with exceptions. Efficiency has received a green rating due to Metis’ ability to fire the process regularly and on time each month (which was a major concern in the adhoc process).

Table 2. Process Improvement Matrix for weather data retrieval process.

<table>
<thead>
<tr>
<th>PIM For DSS/NCEP Retrieval</th>
<th>Adhoc Process</th>
<th>Metis Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>y-5</td>
<td>y-5</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>y-2</td>
<td>y-1</td>
</tr>
<tr>
<td>Output</td>
<td>g</td>
<td>g</td>
</tr>
<tr>
<td>Error handling</td>
<td>y-8</td>
<td>y-9</td>
</tr>
<tr>
<td>Efficiency</td>
<td>y-5</td>
<td>g</td>
</tr>
<tr>
<td>Total time</td>
<td>g</td>
<td>g</td>
</tr>
<tr>
<td>Turn-around time</td>
<td>r</td>
<td>g</td>
</tr>
</tbody>
</table>

The PIM for the sea surface temperature data retrieval process is found in Table 3. For the adhoc process in this workflow satisfaction is given a yellow rating with a weigh of 20 percent, while output and error handling are both given green ratings. Error handling in this case is given a green rating because if an error in the data occurs it is simple to contact the overseer of the data. Consequently, overall quality was given a yellow rating with a weight of 80 percent. Again, in this
workflow the overall total time has a green rating, but the turn-around time has a red rating. Efficiency has a yellow rating with a 50 percent weight.

For the Metis process the satisfaction measure has increased, which results in improved quality. All features in this process were automated because each month the files are transferred to the same location. Recall, each month the same file is overwritten which makes this process respectively simpler than the previous process. Turn-around time improved using Metis, making its efficiency rating higher than that of the adhoc process.

<table>
<thead>
<tr>
<th>PIM For DSS/SST Retrieval</th>
<th>Adhoc Process</th>
<th>Metis Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>y-8</td>
<td>g</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>y-2</td>
<td>g</td>
</tr>
<tr>
<td>Output</td>
<td>g</td>
<td>g</td>
</tr>
<tr>
<td>Error handling</td>
<td>g</td>
<td>g</td>
</tr>
<tr>
<td>Efficiency</td>
<td>y-5</td>
<td>g</td>
</tr>
<tr>
<td>Total time</td>
<td>g</td>
<td>g</td>
</tr>
<tr>
<td>Turn-around time</td>
<td>r</td>
<td>g</td>
</tr>
</tbody>
</table>

**Table 3. Process Improvement Matrix for sea surface temperature retrieval process.**

**CONCLUSION**

This project conducted an evaluation of Metis using qualitative analysis techniques as well as validating the processes. During the validation analysis we found the two workflow processes tested were sound as defined previously in this paper. However, we found that some modifications were necessary for the complete automation of the two processes. This has prompted the developers to implement these modifications, which are currently under way.

While carrying out the qualitative performance analysis, this study discovered the overall quality of Metis did not improve in the weather data retrieval process due to the lack of some key automation features necessary for this process. The SST data retrieval process was not affected by the lack of this feature because there is only one file being transferred to the same location each month. Metis improved the efficiency in both processes due to the increased rating for the submetric turn-around time. Within Metis, the retrieval processes will be fired on a monthly basis. Whereas, when the processes are carried out in the adhoc manner they tend to be carried out when someone requests the new data.

In addition to the metrics evaluated during the performance analysis, this study found an additional qualitative improvement obtained by using Metis for workflow management. Metis’ ability to monitor the progress of a workflow offers additional insurance in case of some unpredictable circumstance. For instance, take the case that a supervisor of a particular process unexpectedly can no longer continue his duties. If he oversees a particular process conducted in an adhoc manner it will be difficult, if not impossible, for someone to follow up the last completed step of the workflow. However, by using Metis, equipped with a workflow monitor, the progression of the process can be seen and the appropriate steps to follow are clear.
FUTURE WORK

Due to time constraints some qualitative metrics were not evaluated. In further investigations of Metis the author recommends the evaluation of additional metrics such as scalability. The metric scalability would rate Metis’ ability to carry out numerous workflows simultaneously.

Recall, modifications are being made to some of Metis’ functionality. After those modifications are made a follow up evaluation should be conducted to point out the improvements made as compared to the current version of Metis.
REFERENCE


