Benchmarking and Configuration
of Workflow Management Systems

Michael Gillmann, Ralf Mindermann, Gerhard Weikum
University of the Saarland
P.O.Box 151150, D-66041 Saarbruecken
e-mail: {gillmann,minderma,weikum}@cs.uni-sb.de
WWW: http://www-dbs.cs.uni-sb.de

Abstract

Workflow management systems are a cornerstone of mission-criticial, possibly cross-organizational business processes. For large-scale applications both their performance and availability are crucial factors, and the system needs to be properly configured to meet the application demands. Despite ample work on scalable system architectures for workflow management, the literature has neglected the important issues of how to systematically measure the performance of a given system configuration and how to determine viable configurations without resorting to expensive trial-and-error or guesswork.

This paper proposes a synthetic benchmark for workflow management systems; based on the TPC-C order-entry benchmark, a complete e-commerce workflow is specified in a system-independent form. This workflow benchmark, which stresses all major components of a workflow system and is parameterized in a flexible manner, has been applied to two operational systems, the commercial system Staffware97 and our own prototype system Mentor-lite. The paper reports performance results from our measurements and discusses lessons learned. In particular, the results demonstrate the scalability of the Mentor-lite architecture. The measurements also underline the need for configuring systems intelligently, and the paper outlines an auto-configuration tool that we have been building to this end.

1   Introduction

1.1   Problem Statement

Workflow technology has penetrated into mission-critical, enterprise-wide or even cross-organizational, business applications. Typical examples are insurance claim processing, cargo shipping, or healt-care tracking and planning, and workflow technology is also embedded in many e-commerce services. Following the terminology of [WfMC] (see also [DKO+98, JB96, LR99, GHS95, CHR+98]), a workflow is a set of activities that belong together in order to achieve a certain business goal. Activities can be completely automated or based on interaction with a human user and intellectual decision-making. In particular, an activity can spawn requests to an arbitrary "invoked application" that is provided by some server independently of the current workflow. Workflow management systems (WFMS) orchestrate the control and data flow between a workflow's activities, based on a high-level specification of the intended behavior (e.g., using Petri-net variants, state charts, or some script language) with some leeway for exception handling and run-time improvisation (as needed, e.g., in medical applications).

Despite their business success, most WFMS products exhibit specific idiosyncracies and, by and large, significant deficiencies and limitations in terms of their performance. The current situation is probably comparable to that of relational database systems in the eighties. Also and similarly to database technology, configuring and tuning a WFMS for satisfactory performance falls more in the realm of black art (i.e., guesswork or expensive trial-and-error experimentation) and sorely lacks scientific foundations. Even such mundane basics such as systematic benchmarks are still missing.

1.2 Contributions

This paper aims to improve the state of the art on the systematic assessment and judicious configuration of WFMSs by defining a simple yet powerful benchmark workflow, based on an e-commerce scenario. The benchmark specification takes the well-known TPC-C order-entry application as a starting point, combines the three major transactions of TPC-C into a workflow, and extends this setting to systematically stress all components of a WMFS architecture. The benchmark is parameterized for adaptation to different scales and application scopes. It is specified using the statechart formalism [HG97], and can be easily converted into the specification languages of most WFMSs.

To demonstrate the platform-independence and usefulness of the benchmark, we present performance measurements of two systems, the commercial product Staffware97 and our own prototype system Mentor-lite. These experiments have revealed limitations and system-specific bottlenecks with regard to scalability, thus underlining the need for such benchmarking. An additional, important lesson is that proper system configuration is a key issue, and that performance depends quite sensitively on subtle aspects of the configuration.

Out of the benchmarking efforts, we have started work towards an "intelligent" auto-configuration tool for WFMS architectures, with specific emphasis on the Mentor-lite architecture. In our prior work [GWW+00] we have focused on the analytical underpinnings for such a tool; in the current paper we present the architecture of the tool itself and its interaction with the various components of the WFMS environment. The implementation of the tool is close to completion and we plan on evaluating the quality of its recommendations using our e-commerce benchmark.

1.3 Related Work

Although the literature includes much work on scalable WFMS architectures [AAA+97, CHR+98, DKO+98, GHS95, JB96, Moh99], there are only few research projects that have looked into the quantitative assessment of WFMS configurations. The work reported in [BD97, BD99] presents several types of distributed WFMS architectures and discusses the influence of different load distribution methods on the network and workflow-server load, mostly using simulations. [SNS99] presents heuristics for the allocation of workflow-type and workflow-instance data onto servers. Mechanisms for enhanced WFMS availability by replicating state data on a standby backup server have been studied in [HA98, KAG+96]. [KWA99] characterizes the workload of cross-organizational workflows by means of Markov models. None of this prior work has addressed the issue of how to systematically benchmark a WFMS and how to configure a WFMS for given performance goals.

Benchmarking is well established, with benchmark specifications gradually evolving, as a form of systematic performance assessment for processors and compilers [Jain91], and also in the area of database and transaction processing systems [Gr93, ONe97]. In particular, the suite of benchmarks published by the Transaction Performance Council [TPC] has proven extremely useful for driving performance enhancements of commercial database systems over the last decade. On the other hand, the work of the TPC also shows that defining a systematic benchmark involves many subtle pitfalls and presents a quite challenge. In particular, finding a careful balance between making the benchmark realistic and functionally comprehensive and ensuring that it can be installed and run on many different platforms with affordable effort is all but trivial.

For the WFMS area and even for modern middleware in general, benchmarking is in its infancy. [MP96] presents a benchmark for a specific system, namely, Lotus Notes, but this effort solely focuses on Notes's use as a message/mail engine. [BSR96] is even more specialized in its performance study of scientific lab processes in genome research. SAP has its product-specific benchmark suite [SAP] that stresses also the R/3-internal workflow engine, but this benchmark is completely tailored to the SAP environment, involves many issues of an ERP system that are not related to workflow technology, and would be difficult to port to another WFMS. The very recent TPC-W benchmark [TPC-W] considers

Web-based e-commerce, but emphasizes the routing, multiplexing, load balancing, caching, and security capabilities of Web application servers (of the category such as Apache, IIS, etc.) and pays no attention to the workflow aspects of e-services. Finally, [GBL+98] has made a laudable effort to define a general benchmark for active database systems, but has indeed restricted itself to the core functions of an active rule engine and cannot be generalized to the richer spectrum of WFMS services.

1.4   Paper Outline

The rest of the paper is organized as follows. Section 2 discusses the general rationale for a systematic benchmark of advanced WFMS architectures. Section 3 presents the specification of the e-commerce benchmark. Section 4 briefly reviews the architecture of the two systems under test. Section 5 presents the setup of the experimental testbed. Section 6 shows the results of our measurements. Section 7 discusses major lessons learned. Section 8 presents the architecture of the auto-configuration tool that we are building.

2   Benchmark Rationale

2.1   Metrics for Performance Assessment

From a business perspective, the benefit of a WFMS should ideally be measured in terms of how smooth and cost-effective the system runs a company's business processes. This involves issues like how much the WFMS contributes to customer satisfaction, effectivity of office workers (e.g., call-center agents), meeting deadlines, reducing operational cost, and ultimately the company's success in the market. Obviously, despite some recent business research along these lines (e.g., [Sch96]), such benefits are hard to quantify. Our goal is much more limited in scope, focusing on metrics that can be directly attributed to the computer-support infrastructure of business processes.

Similar to transaction-processing benchmarks, a key metric is obviously the **throughput** of the entire WFMS. Unlike OLTP, our notion of throughput has a much larger granularity and extension in time: we measure throughput in terms of completed workflows (i.e., instances of a given workflow type), and we have to collect these numbers separately for different workflow types. Note that it does not make sense to amalgate throughput figures for a complex workflow type that involves week-long activities (e.g., processing a credit request in a bank) and a simple, short workflow type that usually completes within a few minutes (e.g., opening a bank account) into a single, seemingly unified metric, which could, however, no longer be meaningfully interpreted.

When we know the number of clients that interact with the WFMS, the rates at which these initiate new workflows, and the relative frequencies of the different workflow types, then the throughput that the system can sustain must be at least as high as the aggregate load induced from all clients. In addition, it is equally crucial that the WFMS is sufficiently responsive. In the context of long-lived business processes, this entails two subgoals:

- The **turnaround time** for an entire workflow must be acceptable. Obviously this depends also on the efficiency of the human users that process intellectual activities, but it is important that the WFMS does not incur any additional bottlenecks. Also, and importantly for benchmarking, the turnaround time from the initiation of a workflow to its completion, as perceived by the initiating user, is the key metric for fully automated workflows that contain no interactive activities.

- For workflows that include interactive activities and could thus have an extended lifetime of several days or weeks, the critical units are the individual interactions between a user and the WFMS that occur during activities and, especially, at the start and end of an activity when role resolution and other worklist-handling tasks are performed. For the purpose of our

benchmark definition we are thus interested in the **step response time**, where a step could be a specific type of user interaction or an entire activity.

2.2   Components under Test

A WFMS is a complex system that consists of a build-time component for specifying workflows, a run-time component with the actual **workflow engine** for the proper interpretation of control and data flow in between activities, and a suite of administration tools for monitoring etc. For our purpose, the run-time engine is the component that matters. In most WFMSs, the workflow engine is run as a multithreaded server process, or sometimes as a collection of processes on the same or different computers. This workflow server may have its own persistent storage system for tracking the states of long-lived workflows, or more typically relies on a **database server** for this purpose. In addition, it may interact with **application servers** for invoked applications that result from automated activities. Finally, the communication among these various servers may be based on other, lower-level, middleware that is itself implemented in the form of dedicated **communication servers**. The latter serve as a reliable request brokers, with TP monitors (MQ Series, MTS, etc.) or ORBs (e.g., Orbix) being prevalent examples.

With the workflow engine really being the heart of a WFMS, it is crucial for a benchmark to stress the entire functional spectrum of the engine. Therefore, one requirement is to exercise the different types of **control-flow constructs**, notably, conditional branching, fork-join parallelism, and also loops. These should be supported by all industrial-strength WFMSs, but it is widely open to what extent they are handled efficiently.

An equally important, orthogonal, aspect to be examined by the benchmark is the support for different types of activities. Fully **automated activities**, which may invoke an **external application** on an application server, and **interactive activities** that run on client machines pose fairly different requirements on the workflow engine, and some commercial WFMS are known to be particularly geared for one type of activities.

Our benchmark will include all major types of control-flow constructs and both types of activites. However, the benchmark will be designed such that different, restricted, levels of specialization can be derived when the interest is on specific application requirements or special-purpose WFMSs (e.g., for call-center applications or canonical, simple types of e-commerce). The components under test will include the workflow server as well as any underlying database or storage servers and also external application servers and communication servers if these are present in a given WFMS architecture. The actual external applications (e.g., the program for testing the authenticity of a hand-written or digital signature) will, however, only be emulated by a stub (i.e., a "dummy" program) inside the application server. Similar to the functional specializations, we will allow different scoping levels of the benchmark to either include or exclude such additional components. So, in the simplest case the benchmark solely examines the workflow engine, and in the most advanced case it takes the full suite of surrounding software and the corresponding servers into account.

In none of these cases, however, we consider client machines and the client software as part of the system under test. Excluding clients is justified as they are almost always uncritical with regard to performance (or can be easily configured to this end). The benchmark will include only stubs that emulate client behavior.

3   Benchmark Specification

In this section we describe our benchmark specification. The benchmark reflects all previously described metrics and tests all interesting system components. Our proposal captures an e-commerce scenario. It is similar to the TPC-C benchmark for transaction systems [TPC], with the key difference that we combine multiple transaction types into a workflow and further enhance the functionality by control and data flow handling. Furthermore, we explicitly take application invocations into account.

3.1 Control Flow as State Chart

The control flow specification is given in the form of state charts [Har87, HG97]. This specification formalism has been adopted for the behavioral dimension of the UML industry standard, and it has been used for our own prototype system Mentor-lite [WW97, MWW+98a].

State charts specify the control flow between activities. A state chart is essentially a finite state machine with a distinguished initial state and transitions driven by event-condition-action rules (ECA rules). A transition from state s_i to state s_j, annotated with an ECA rule of the form *E[C]/A*, fires when event *E* occurs and if condition *C* holds. The effect is that state s_i is left, state s_j is entered, and action *A* is executed. Conditions and actions are expressed in terms of variables that are relevant for the control and data flow among activities. In addition, an action *A* can explicitly start an activity, expressed by *st!(activity)*, and can generate an event *E* or modify a condition variable *C* (e.g., *fs!(C)* sets the condition *C* to false). Each of the three components of an *E[C]/A* triple may be empty.

Important additional features of state charts are nested states and orthogonal components. Nesting of states means that a state can itself contains an entire state chart. The semantics is that upon entering the higher-level state, the initial state of the embedded lower-level state chart is automatically entered, and upon leaving the higher-level state all embedded lower-level state charts are left. The capability for nesting states is especially useful for the refinement of specifications during the design process and for incorporating subworkflows. Orthogonal components denote the parallel execution of two state charts that are embedded in the same higher-level state (where the entire state chart can be viewed as a single top-level state). Both components enter their initial states simultaneously, and the transitions in the two components proceed in parallel, subject to the preconditions for a transition to fire.

Figure 1 shows the top-level state chart for our e-commerce (EC) benchmark workflow. Each state corresponds to an activity or one (or multiple, parallel) subworkflow(s). We assume that for every activity *act* the condition *act*_DONE is set to true when *act* is finished. So, we are able to synchronize the control flow so that a state of the state chart is left when the corresponding activity terminates. For parallel subworkflows, the final states of the corresponding orthogonal components serve to synchronize the termination (i.e., join in the control flow).

The workflow behaves as follows. After the initialization of the workflow instance, the *NewOrder* activity is started. After the termination of *NewOrder*, the control flow is split. If the customer wants to pay by credit card, the condition *PayByCreditCard* is set and the *CreditCardCheck* activity checks the validity of the credit card. If there are any problems with the credit card, the workflow is terminated. Otherwise the shipment, represented by the nested top-level state *Shipment_S*, is initiated spawning two orthogonal/parallel subworkflows, specified in the state charts *Notify_SC* (Figure 2) and *Delivery_SC* (Figure 3), respectively. The first subworkflow *Notify* has only one activity that sends an acknowledgment mail. The delivery of goods is done by one or several external, eventual autonomous stores. So, the second subworkflow, *Delivery*, (sequentially) invokes for each ordered item an activity that identifies a store from which the item could be shipped. Then, a second activity instructs the store to deliver the item and waits for an acknowledgement. The both activities *FindStore* and *CheckStore* are repeated within a loop over all ordered items. If the customer wants to pay by bill, a reminder counter and the due day for the payment have to be initialized. After the termination of both subworkflows, the control flow is synchronized, and split again depending on the mode of payment. If the customer wants to pay by credit card, the credit card is now charged and the workflow terminates. If the customer wants to pay by bill, a activity is invoked that waits for the settlement of the invoice. If the payment is confirmed within the running period, the workflow also terminates. If the payment is not confirmed after two weeks, within a loop an activity is invoked that sends a reminder to the customer. Moreover, the reminder counter is increased and the due day initialized again. This is repeated at most three times. If the payment is not receipted within the period after the third reminder, an activity is invoked that inform the legal department and the workflow terminates.

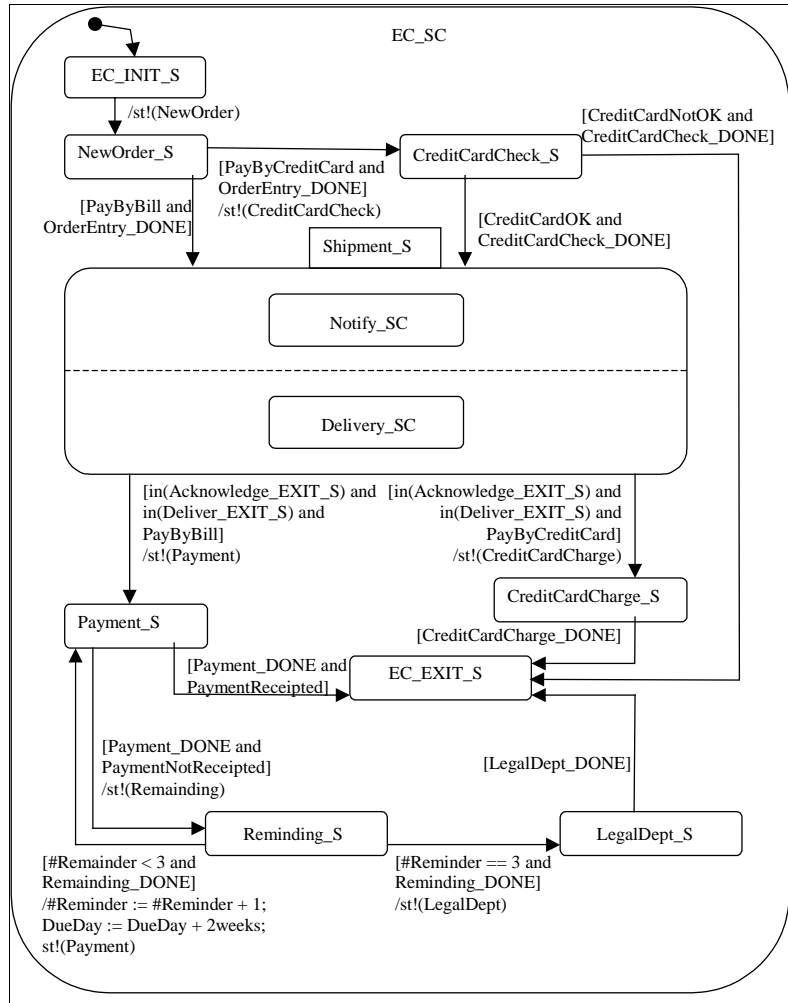Note, that we have so far neglected the exception handling (e.g. if an item is not deliverable).

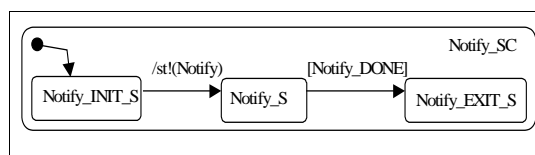**Figure 1 :** State chart of the *electronic commerce (EC)* workflow example

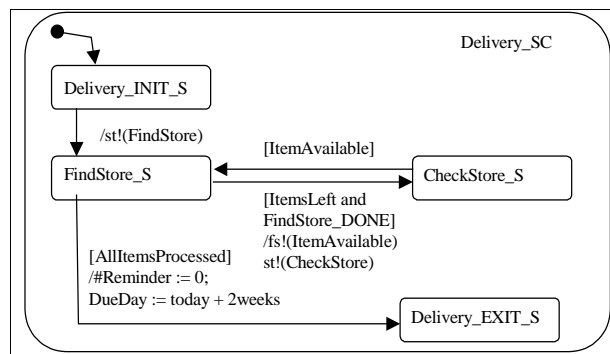**Figure 2 :** State chart of the *Notify* subworkflow

Figure 3 : State chart of the *Delivery* subworkflow

3.2 Data Flow as Activity Chart

The data flow specification is given in the form of an activity chart [Har87, HG97]. Activities reflect the functional decomposition of a system and denote the "active" components of a specification; they correspond directly to the activities of a workflow. An activity chart specifies the data flow among activities, in the form of a directed graph with data items as arc annotations.

Figure 4 shows the activity chart for our benchmark workflow. First of all, the activity chart contains a pointer to the corresponding state chart (@EC_SC) as well as all involved activities. All customer specific input data is collected by the *NewOrder* activity and distributed to the other activities as needed. Moreover, the activity *FindStore* sents the ID of the selected store and its specific item list to the activity *CheckStore*. Finally, *CheckStore* returns the acknowledgment to *FindStore*.
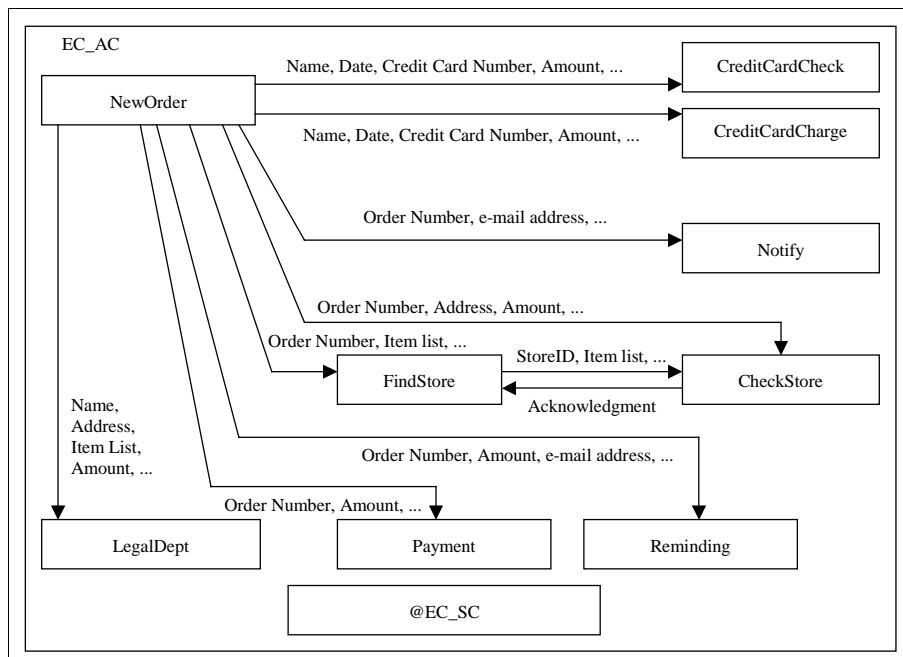

Figure 4 : Activity chart of the *electronic commerce (EC)* workflow example

3.3 Applicability to other WFMS

The EC workflow specification is straightforward to implement when the WFMS that is to be benchmarked uses state charts. In this case only the stubs for the simulated application programs have to be implemented, but this can be done with one generic stub with different parameter settings.

Installing the benchmark on other WFMSs is still simple when the workflow specification language of the target WFMS is a well-founded "high-level" formalism such as Petri nets, etc. In fact, automatic convertion of state charts into other formalisms are largely feasible [NFZ98, MWW+98b].

4 Systems under Test

In this section, we shortly describe the architectures of the systems under test and point out the respective features that have a major impact on the performance measurements. The systems under test are our own research prototype Mentor-lite [MWG+99] and the commercial product Staffware97 [Staff].

4.1 Mentor-lite

The core of Mentor-lite is an *interpreter* for state chart specifications. The interpreter performs a stepwise execution of the workflow specification according to its formal semantics [WW97]. For each step, the activities to be performed by the step are determined and started. Two additional components, the *communication manager* and the *log manager*, are closely integrated with the workflow interpreter. All three components together form the *workflow engine*. The execution of a workflow instance can be distributed over several workflow engines at different sites. A separate *workflow log* is used at each site where a Mentor-lite workflow engine is running. The communication manager is responsible for sending and receiving synchronization messages between the engines. These messages contain information about locally raised events, updates of state chart variables and state information of the local engine [MWW+98b]. When a synchronization message is received, the corresponding updates at the receiving site are performed. In order to guarantee a consistent global state even in the presence of site or network failures, Mentor-lite uses the CORBA Object Transaction Services to implement reliable message queues. The CORBA implementation Orbix provides the basic communication infrastructure for distributed execution. The workflow engine, comprising the three components interpreter, communication manager, and log manager, is implemented as an Orbix server. Its IDL interface provides a method to start a workflow instance and a method to set variables and conditions within the workflow instance.

Databases like the *workflow repository* (i.e., a repository of workflow specifications) or the *worklist database* can be shared by Mentor-lite workflow engines at different sites. In the current setup, the underlying DBMS is Oracle 7.

Communication interfaces to application programs are implemented by wrappers using the distributed computing environment CORBA. On top of these interfaces, protocols for complex interactions with application programs are specified in terms of state and activity charts. The workflow engine starts the wrappers asynchronously and uses the methods of the wrapper objects to read or set variables. The application wrappers can in turn use the workflow engine's method to set control flow variables.

In this paper, we consider two different versions of the Mentor-lite implementation. The difference between the two releases is in the handling of application invocations. The first version, referred to as "*ml-proc*", starts a new process for each external application on the workflow-server site for the data exchange with the external application. The advantage of this approach is the increased availability of the system as only one workflow is affected when the communication with the application fails. However, the required main memory on the workflow server site increases significantly. The second version of Mentor-lite, referred to as "*ml-thr*", uses threads within a single process, which is much more memory-efficient.

4.2 Staffware97

Staffware97 has a client-server architecture with a monolithic server. All components like log manager, worklist handler, etc. are implemented inside the *workflow engine*. The workflow engine can be run on multiple workflow servers, but each workflow instance is bound to one server and exclusively handled by this server throughout its lifetime. So Staffware97 does not support a partitioned and distributed workflow execution. The workflow engine manages several *work queues* that start application programs and can be polled by the user's clients.

A work queue schedules the work items of one or several users in a FIFO manner with priorities, i.e., FIFO among all processes with the same priority. The number of users per work queue as well as the number of parallel processes per user are system parameters. This is a critical issue especially for mostly automated workflows, i.e., workflows with mostly automated, non-interactive activities, because all such activities are scheduled in the work queue of a single user (i.e., dummy user "auto").

Staffware97 provides an interface to application programs based on Dynamic Data Exchange (DDE). External application programs are called via scripts that use DDE commands. Automated, non-interactive activities can also be started without the use of any scripts but only under the restriction that they run on the workflow engine's server machine.

The exchange of data between the workflow engine and the application programs is handled via the file system. Input data to the application as well as results from the application are stored in and read from temporary files. In an asynchronous application invocation, the calling script has to raise an *event* when the workflow engine is allowed to read the result data. Automated, non-interactive programs running on the server machine are able to communicate directly with the workflow engine also via *pipes*. In our measurements, we used both options for comparison. We refer to the file- and event-based as "*sw-ev*", and to the pipe-based alternative as "*sw-pi*".

Note that Staffware97 is no longer the current release of Staffware's workflow product, but the newer version Staffware2000 [Staff] became available only very recently.

5   Experimental Testbed

5.1   Setup and Instrumentation

For every WFMS under test we must create its full-fledged system environment with all necessary servers, including middleware components (e.g. Corba or a TP-monitor), since even for simulated external applications the benchmark includes the invocation and completion steps. So the testbed is significantly more complex than for TPC-C-like benchmarks.

Our testbed consists of the following modules:

- A synthetic load generator starts new workflow instances with a given interarrival time distribution. In our case, we used a Poison arrival process with a given mean as a parameter.

- A monitoring component observes and logs the start and stop times of the activities and entire workflows.

- Stub applications simulate the external applications. These stubs simply read their input data, sleep for a specified time period, and return control-flow-relevant result data. The mean turnaround time of these stubs is a parameter of the experiments.

- A dedicated SUN Sparc5 is used as an application server connected to the workflow-server machine with an Ethernet LAN. The use of a separate machine for the application server, as opposed to running applications on the workflow-server machine, was optional and varied in the experiments.

- The workflow server itself runs on a dedicated SUN Sparc10 with all, WFMS-specific, additional components as described in following. In all experiments reported here, we limited ourselves to a single workflow server (i.e., did not make use of multiple workflow engines or even partitioned and distributed workflow execution).

Mentor-lite

Mentor-lite additionally needed the Oracle7 database system for its logging and recovery component and as a repository containing workflow specifications etc. The database was not dedicated for the WFMS and ran on a remote database server machine. For the communication interfaces of the workflow engine and the application wrappers we used Orbix 2.3.

Staffware97

Staffware97 offers a choice between logging in a database or to the file system. We chose the file system variant. In the experiments that involved a separate application server, the application invocation was done by a remote-shell call from a script on the workflow-engine machine.

5.2 Parameter Settings

In this paper, we present three series of measurements with the arrival rate of new workflow instances systematically increased from a lightly loaded level to system saturation.

In our baseline experiment, the system configuration consists of exactly one server machine. So, all components of the WFMS including the the external application programs were run on the dedicated SUN Sparc10. The turnaround time of the activities was normally distributed with a mean of 10 seconds and a standard deviation of 4 seconds.

In the second series of measurements, we study the sensitivity with regarad to the activity turnaround time. We increased the mean turnaround time of the external applications to 60 seconds and its standard deviation to 15 seconds.

In the third series, we measured the impact of the system configuration. Specifically, we added a dedicated computer for the application server and the external applications.

5.3 Test Levels

As mentioned before, our e-commerce benchmark supports different test levels with regard to functionality and the scope of the benchmarking. For example, we can easily derive specialized, simplified variants that contain only automated activities or no loops.

In the measurements presented here, we specialized the benchmark by using only the control flow path for the case of credit card payment, disregarding the reminder loops, and limiting the delivery loop to exactly one iteration. As a consequence, the workflow contained only automated activities, no interactive ones.

6   Experimental Results

This section presents the results of the three series of measurements. For each experiment, we give the system throughput in terms of completed workflow instances per hour, the mean turnaround time of the workflow instances in seconds, and the step response time of a single control flow step, i.e., the time between the completion of an activity and the start of the next activity.

As mentioned in Section 4, we benchmarked two versions of Mentor-lite, ml-proc and ml-thr, and two versions of Staffware97, sw-ev and sw-pi. In all measurements, we used an observation window of 8 hours.

6.1 Baseline Experiment

The baseline experiment used one dedicated computer for all servers. The mean turnaround time of the activities was set to 10 seconds each.

Figure 5 shows the system throughput as a function of the workflow arrival rate. As expected, the throughput increases linearly with the arrival rate until the system saturates. The maximum sustainable throughput of ml-thr is about 10% higher than that of sw-pi. The other two candidates under test performed poorly. sw-ev exhibited very high delays because of waiting for events to read result data from the file system. The low throughput of ml-proc resulted from hardware and operating-system bottlenecks, especially with regard to memory, for too many active processes (one per activity) had to be handled.
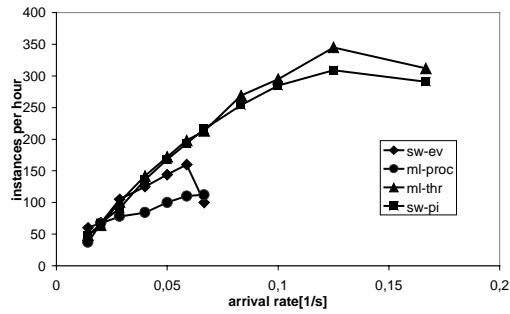
Figure 5 : Throughput of baseline experiment

Figure 6 shows the mean turnaround time of workflow instances. The turnaround times of Mentor-lite stays almost constant for all given arrival rates. Staffware97 is bounded by the number of workflow instances running in parallel. As the execution of invoked applications is carried out on behalf of a single, artificial "dummy" user (called the "auto" user in Staffware terminology), the work queue handling for this user caused congestion, which resulted in long waiting times. The major drawback of sw-ev is again its inefficient treatment of data that is returned from the application to the workflow engine.
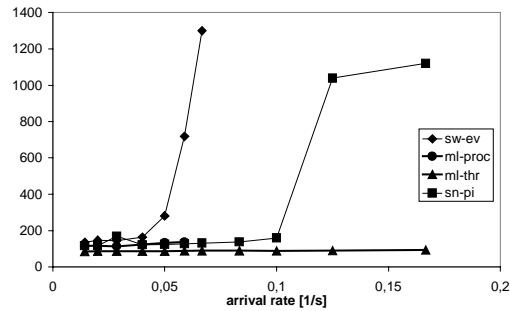
Figure 6 : Turnaround time of baseline experiment

Figure 7 shows the mean of the step response times. The step response times showed patterns that were similar to the workflow turnaround times. For Mentor-lite, the response times could be kept almost constant with increasing arrival rate.
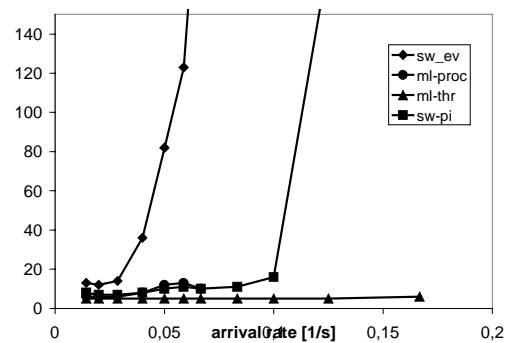

Figure 7 : Step response times of baseline experiment

As the ml-proc version of Mentor-lite performed much worse than ml-thr in all experiments, this architectural variant is no longer considered in the following sensitivity studies.

6.2 Impact of Activity Turnaround Time

In these experiments the mean turnaround time of the activities was increased to 60 seconds. Figure 8 shows the impact of this significantly higher activity turnaround time on the throughput. For all systems, the maximum sustainable throughput was reduced by a factor of 2 or higher. Mentor-lite suffered the highest drop of throughput and became even inferior to Staffware.
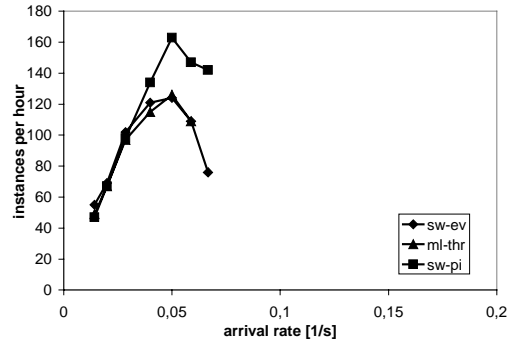

Figure 8 : Activity turnaround time impact on throughput

Figure 9 shows the impact on the workflow turnaround times. For all systems, the mean workflow turnaround time increases because its dominated by the activity tournaround times. Again, the effects of queueing within the work queues increase the turnaround time for Staffware97 dramatically as the arrival rate and thus the number of active workflows increases.
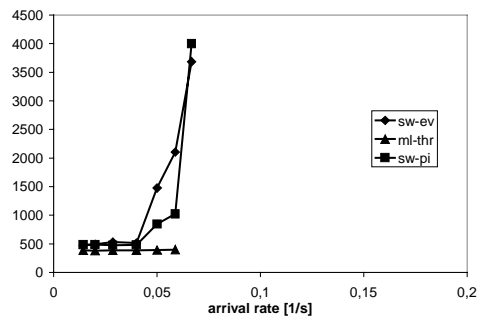

Figure 9 : Activity turnaround time impact on workflow turnaround time

Figure 10 shows the impact on the step response times. With regard to this metric nothing significant changed compared to the baseline experiment.
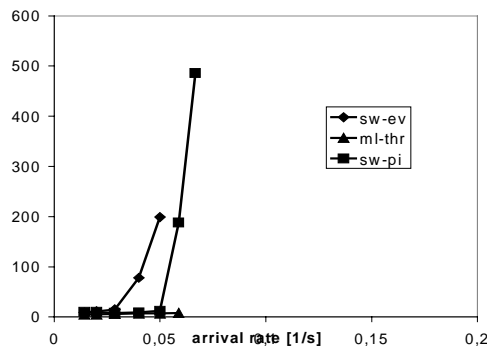

Figure 10 : Activity turnaround time impact on step response time

6.3 Impact of System Configuration

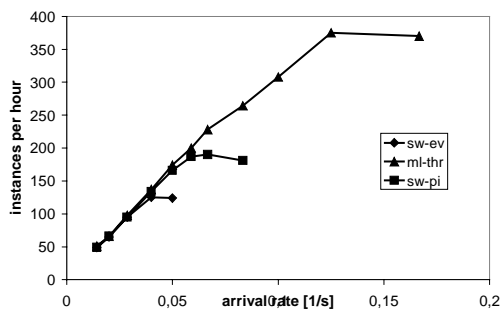The impact of the system configuration was evaluated by adding a dedicated computer for the application server.

Figure 11 :   System configuration impact on throughput

Figure 11 shows the impact on the system throughput for the case of short activities with a mean duration of 10 seconds. Mentor-lite is able to fully exploit the additional hardware. The sustainable throughput increases even more with longer activities, i.e., the case of a mean duration of 60 seconds, as shown in Figure 12. Surprisingly, the throughput of Staffware97 dropped compared to the baseline experiment. So for Staffware97 the additional hardware turned out to be a penalty rather than an opportunity. The reason for this seemingly counterintuitive phenomenon lies in the pecularity of Staffware97 that it needs remote-shell calls and thus dynamically spawns remote processes for applications that are invoked on a machine other than the workflow server.
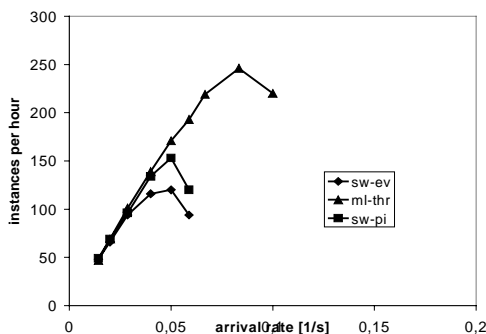

Figure 12 :   System configuration impact on throughput with longer activities

Figure 13 shows the impact on the turnaround time. The turnaround time for Mentor-lite increases only marginally due to the remote application invocation.
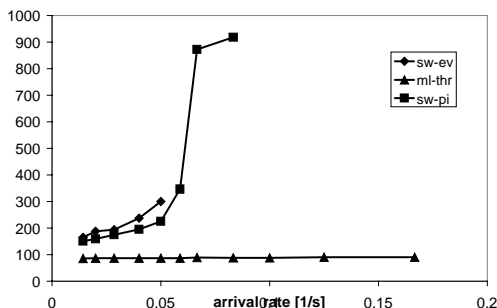

Figure 13 :   System configuration impact on turnaround time
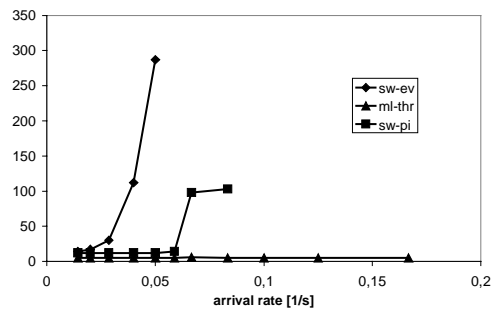
Figure 14 shows the impact on the step response times.

13

Figure 14 : System configuration impact on step response time

7 Lessons Learned

In addition to demonstrating the viability of the proposed benchmark, the experiments provided a number of important general insights:

- The Corba-based architecture of Mentor-lite proved to be reasonably scalable. In particular and unlike Staffware97, it was able to fully exploit the additional hardware resources of a dedicated application server machine.

- Multi-threading inside the workflow engine turned out to be crucial for bounding the memory consumption of the Mentor-lite workflow server. The version with a separate process for each activity performed very poorly and became unstable for higher load.

- Staffware97 did not scale as well as we expected. First, the file- and event-based version for the communication with invoked applications incurred high delays and resulted in long step response times, critically adding to the workflow turnaround times. Second and even more importantly, the architecture turned out to handle automated activities in a rather inefficient way. Remote invocations on a dedicated application server machine even resulted in decreasing throughput. So Staffware97 obviously is geared mostly for workflows with interactive activities running on client machines, with considerably less stress on the server side.

- Generally, the responsiveness of Staffware97 seems to be critically depending on the configuration of its work queues. For automated activities, all "work items" were assigned to a single queue, for the artificial user "auto", forming a response-time bottleneck. An alternative could be to add more artificial users with some intelligent assignment of activity types to these "auto"-type users, but this would entail a sophisticated load distribution problem and was beyond the scope of our experiments.

8 Architecture of an Auto-Configuraion Tool for Mentor-lite

A distributed configuration of Mentor-lite consists of different workflow servers (i.e., instances of the workflow engine), application servers, and one communication server (i.e., ORB). Each server of the first two categories can be dedicated to a specified set of workflow activities and external applications, resp., on a per type basis. Each of these dedicated servers and also the communication server can be replicated across multiple computers for enhanced performance and availability. Given this flexibility (which is provided in similar ways also by some commercial WFMSs), it is a difficult problem to choose an appropriate configuration for the entire WFMS that meets all requirements with regard to throughput, interaction response time, and availability. Moreover, it may be necessary to adapt an initial configuration over time due to changes of the workflow load, e.g., upon adding new workflow types.

14

To this end, we have developed a suite of analytic models, using stochastic methods like continuous-time Markov chains and Markov reward models, to predict the performance, availability, and performability under a given load. The performance model estimates the maximum sustainable throughput in terms of workflow instances per time unit and the mean waiting time for service requests such as interactions upon starting an activity on the basis of a Markov chain model for the statistical behavior of the various workflow types. The availability model estimates the mean downtime of the entire system for given failure and restart rates for the various components. Finally, the performability model takes into account the performance degradation during transient failures and estimates the effective mean waiting time for service requests with explicit consideration of periods during which only a subset of a server type's replicas are running. These models, which are described in detail in [GWW+00], form the underpinnings of an auto-configuration tool for distributed WFMSs.

The auto-configuration tool is primarily driven by statistics on the workload from the monitoring tool of Mentor-lite. It can feed this information into its analytic models to a hypothetical configuration in a what-if analysis. By systematic variation of the parameters for such hypothetical configurations the tool is also able to derive the (analytically) best configuration, i.e., the minimum degree of replication of each of the involved server types to meet given availability and performance or performability goals, and recommend appropriate reconfigurations. The tool is largely independent of a specific WFMS, using product-specific stubs for its various components that need to interact with the WFMS.
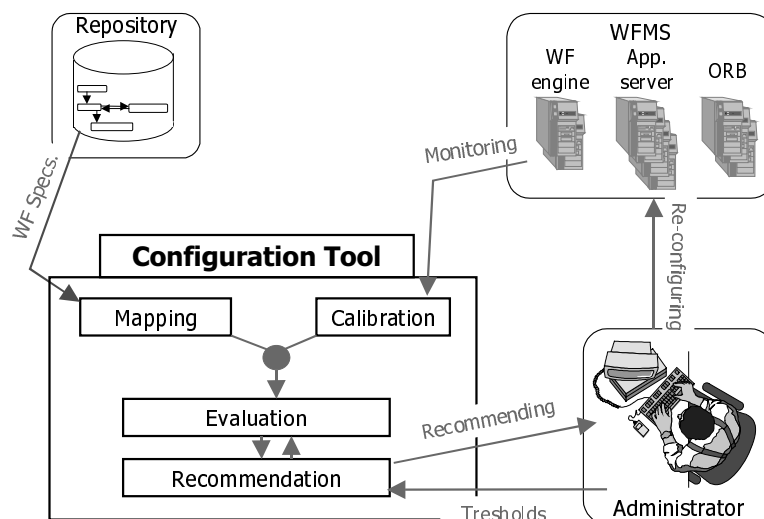

Figure 15 : Integration of the auto-configuration tool

The components of the configuration tool and its embedding into the overall system environment are illustrated in Figure 15. The tool consists of four main components:

- the *mapping* of workflow specifications onto the tool's internal models,
- the *calibration* of the internal models by means of statistics from monitoring the system,
- the *evaluation* of the models for given input parameters, and
- the computation of *recommendations* to system administrators and architects, with regard to specified goals.

For the mapping the tool interacts with a workflow repository where the specifications of the various workflow types are stored. In addition, statistics from online monitoring are used as a second source (e.g., to estimate typical control flow behavior etc.). The configuration tool translates the workflow specifications into corresponding continuous-time Markov chain models. For the evaluation

of the models, additional parameters may have to be calibrated; for example, the first two moments of server-type-specific service times for various elementary service requests (e.g., starting an activity) have to be fed into the models. This calibration is again based on appropriate online monitoring. So both the mapping and calibration components exploit online statistics about the running system. However, when the tool is to be used for configuring a completely new workflow environment, the parameters can be estimated by results of the above presented benchmark. Later, after the system has been operational for a while, these parameters can be automatically adjusted, and the tool can then make appropriate recommendations for reconfiguring the system.

The evaluation of the tool's internal models is primarily driven by specified performance goals with consideration of transient component downtimes, so-called *performability goals*. System administrators or architects can specify goals of the following kinds:

- the minimum throughput in terms of completed workflow instances per time unit that the entire WFMS must be able to sustain,

- a tolerance threshold for the mean waiting time of service requests that would still be acceptable to the end-users,

- a tolerance threshold for the unavailability of the entire WFMS, or in other words, a minimum availability level.

The first two goals requires evaluating the performance or the more comprehensive performability model, whereas the third one merely needs the availability model. The tool can invoke these evaluations either for a given system, or it can search for the minimum-cost configuration that satisfies all goals. The cost of a configuration is assumed to be proportional to the total number of servers that constitute the entire WFMS, but this could be further refined with respect to different server types. Also, the goals can be refined into workflow-type-specific goals, by requiring, for example, different maximum waiting times or availability levels for specific types.

The tool uses the results of the model evaluations to generate recommendations to the system administrators or architects. Such recommendations may be asked for regarding specific aspects only (e.g., focusing on performance and disregarding availability), and they can take into account specific constraints such as limiting or fixing the degree of replication of particular server types (e.g., for cost reasons).

The most far-reaching use of the configuration tool is to ask it for the minimum-cost configuration that meets specified performability and availability goals. Computing this configuration requires searching the space of possible configurations, and evaluating the tool's internal models for each candidate configuration. While this may eventually entail full-fledged algorithms for mathematical optimization such as branch-and-bound or simulated annealing, our first version of the tool uses a simple greedy heuristics. The algorithm iterates over candidate configurations by increasing the number of replicas of the most critical server type until both the performability and the availability goals are satisfied. Since either of the two criteria may be the critical one and because an additional server replica improves both metrics at the same time, the two criteria are considered in an interleaved manner. Thus, each iteration of the loop over candidate configurations evaluates the performability and the availability, but adds servers to two different server types only after re-evaluating whether the goals are still not met. This way the algorithm avoids "oversizing" the system configuration.

To summarize, the functionality of the configuration tool comprises an entire spectrum ranging from the mere analysis and assessment of an operational system all the way to providing assistance in designing a reasonable initial system configuration, and, as the ultimate step, automatically recommending a reconfiguration of a running WFMS.

9   Conclusion and Outlook

In this paper we have proposed the first systematic benchmark for WFMS architectures that we are aware of. We have demonstrated the viability and usefulness of the benchmark with measurements of two different WFMSs (each in two different versions). In particular, porting the benchmark, which is initially specified in terms of state and activity charts, to the specification language of Staffware was fairly easy and the installation of the entire setup for the measurements was relatively straightforward (albeit more time-consuming than we expected).

The measured results clearly show that the configuration of a WFMS architecture has a crucial impact on the achieved performance. For example, running an application server on a dedicated machine, as opposed to running it on the same machine as the workflow engine's server, can make a big difference in terms of throughput and turnaround time. To aid system administrators in finding a proper configuration, we have started working on an auto-configuration tool, which we have sketched in the paper. One important issue among the options of the tool is Mentor-lite's capability of having multiple workflow servers on different machines and running even single workflows in a distributed manner. Presumably, high-end commercial workflow systems have similar capabilities or will add such support in the near future. Our future work will include also additional performance measurements for this advanced type of system configuration.

References

[AAA+97]   G. Alonso, D. Agrawal, A. Abbadi, C. Mohan: Functionality and Limitations of Current Workflow Management Systems, IEEE Expert Vol. 12, No. 5, 1997

[BD97]   T. Bauer, P. Dadam: A Distributed Execution Environment for Large-Scale Workflow Management Systems with Subnets and Server Migration, IFCIS Int'l Conf. on Cooperative Information Systems (CoopIS), Kiawah Island, South Carolina, 1997

[BD99]   T. Bauer, P. Dadam, Distribution Models for Workflow Management Systems - Classification and Simulation (in German), Technical Report, University of Ulm, Germany, 1999

[BSR96]   A. Bonner, A. Shrufi, S. Rozen: LabFlow-1: a Database Benchmark for High-Throughput Workflow Management, Int'l Conf. on Extending Database Technology (EDBT), Avignon, France, 1996

[CHR+98]   A. Cichoki, A. Helal, M. Rusinkiewicz, D. Woelk: Workflow and Process Automation, Concepts and Technology, Kluwer, 1998

[DKO+98]   A. Dogac, L. Kalinichenko, M. Tamer Ozsu, A. Sheth (Eds.), Workflow Management Systems and Interoperability, NATO Advanced Study Institute, Springer-Verlag, 1998

[GBL+98]   A. Geppert, M. Berndtsson, D. Lieuwen, C. Roncancio: Performance Evaluation of Object-Oriented Active Database Management Systems Using the BEAST Benchmark, Theory and Practice of Object Systems (TAPOS), Vol. 4, No. 4, 1998

[GHS95]   D. Georgakopoulos, M. Hornick, A. Sheth: An Overview of Workflow Management: From Process Modeling to Workflow Automation Infrastructure, Distributed and Parallel Databases, Vol. 3, No. 2, 1995

[GWW+00]   M. Gillmann, J. Weissenfels, G. Weikum, A. Kraiss: Performance and Availability Assessment for the Configuration of Distributed Workflow Management Systems, Int'l Conf. on Extending Database Technology (EDBT), Konstanz, Germany, 2000

[Gr93]   J Gray (ed.): The Benchmark Handbook, 2nd Edition, Morgan Kaufmann, 1993

[HA98]   C. Hagen, G. Alonso: Flexible Exception Handling in the OPERA Process Support System, Int'l Conf. on Distributed Computing Systems (ICDCS), Amsterdam, The Netherlands, 1998

[Har87]   D. Harel, State Charts: A Visual Formalism for Complex Systems, Science of Computer Programming, Vol. 8, 1987

[HG97]   D. Harel, E. Gery: Executable Object Modeling with Statecharts, IEEE Computer, Vol. 30, No. 7, 1997

[Jain91]   R. Jain: The Art of Computer Systems Performance Analysis, John Wiley & Sons, 1991

[JB96]   S. Jablonski, C. Bussler: Workflow-Management, Modeling Concepts, Architecture and Implementation, International Thomson Computer Press, 1996

[KWA99]   J. Klingemann, J. Waesch, K. Aberer, Deriving Service Models in Cross-Organizational Workflows, 9th Int'l Workshop on Reasearch Issues in Data Engineering (RIDE), Sydney, Australia, 1999

[KAG+96]   M. Kamath, G. Alonso, R. Günthör, C. Mohan, Providing High Availability in Very Large Workflow Management Systems, 5th Int'l Conf. on Extending Database Technology (EDBT), Avignon, France, 1996

[LR99]   F. Leymann, D. Roller, Production Workflow: Concepts and Techniques, Prentice Hall, 1999

[Moh99]   C. Mohan, Workflow Management in the Internet Age, Tutorial, http://www-rodin.inria.fr/~mohan

[MP96]   K. Moore, M. Peterson: A Groupware Benchmark Based on Lotus Notes, Int'l Conf. on Data Engineering (ICDE), New Orleans, Louisiana, 1996

[MWG+99]   P. Muth, J. Weissenfels, M. Gillmann, G. Weikum: Integrating Light-Weight Workflow Management Systems within Existing Business Environments, Int'l Conf. on Data Engineering (ICDE), Sydney, Australia, 1999

[MWW+98a]   P. Muth, D. Wodtke, J. Weissenfels, G. Weikum, A. Kotz Dittrich, Enterprise-wide Workflow Management based on State and Activity Charts, in [DKO+98]

[MWW+98b]   P. Muth, D. Wodtke, J. Weissenfels, A. Kotz Dittrich, G. Weikum: From Centralized Workflow Specification to Distributed Workflow Execution, Intelligent Information Systems, Special Issue on Workflow Management, Vol. 10, No. 2, 1998

[NFZ98]   M. Nüttgens, T. Feld, V. Zimmermann: Business Process Modeling with EPC and UML: Transformation or Integration, in: M. Schader, A. Korthaus (eds.): The Unified Modeling Language - Technical Aspects and Applications, Workshop des Arbeitskreises "Grundlagen objektorientierter Modellierung" (GROOM), Heidelberg Germany, 1998

[ONe97]   P. O'Neil: Database Performance Measurement, in: A.B. Tucker (ed.): The Computer Science and Engineering Handbook, CRC Press, 1997

[SAP]   SAP AG: SAP E-Business Solutions, http://www.sap-ag.de/solutions/technology/index.htm

[Sch96]   A.W. Scheer: Benchmarking Business Process, in: Okino, N.; Tamura, H.; Fujii, S. (eds.): Advances in Production Management Systems, IFIP TC5/WG5.7 Int'l Conf. on Production Management Systems (APMS), Kyoto, Japan, 1996

[SNS99]   H. Schuster, J. Neeb, R. Schamburger, A Configuration Management Approach for Large Workflow Management Systems, Int'l Joint Conf. on Work Activities Coordination and Collaboration (WACC), San Francisco, California, 1999

[Staff]   Staffware, http://www.staffware.com

[TPC]   Transaction Processing Performance Council, http://www.tpc.org

[TPC-W]   TPC-W Benchmark Specification, http://www.tpc.org/wspec.html

[WfMC]   Workflow Management Coalition, http://www.wfmc.org

[WW97]   D. Wodtke, G. Weikum, A Formal Foundation For Distributed Workflow Execution Based on State Charts, Int'l Conf. on Database Theory (ICDT), Delphi, Greece, 1997