

Analytics for Performance Optimization of BPMN2.0 Business Processes

**Robert M. Shapiro, Global 360, USA
Hartmann Genrich, GMD (retired), Germany**

INTRODUCTION

We describe a new approach to process improvement based on the combined use of statistics and simulation to study the structural aspects of process models. Past efforts to use simulation focused on resource optimization have led to some significant successes when coupled with Workforce Management scheduling technology¹, but that approach has not been particularly successful in making structural improvements in the actual processes. The difficulty of preparing satisfactorily detailed schedules, combined with the structural complexities introduced in particular by the event and looping structures in BPMN, requires a fresh look at the problem.

We have developed a tool, the Process Analytica Performance Analyzer, which can do basic timeline analysis of a BPMN model. It is aligned with Bruce Silver's modeling methodology². Not a resource schedule simulation engine, but an analysis engine that is much simpler to use but gives good results for a broad class of processes. Most simulation tools are too focused on the resource bottleneck problem but not so useful (and too hard to use) for the much more common process structure problem, in which resources spend a small fraction of their time on process tasks.

Extensive use is made of spread-sheets both to prepare simulation inputs and work with the analysis. Our major target is business analysts who will be comfortable with the spread-sheet approach. This is a much easier and more efficient way of setting up simulation runs and doing analysis.

Intermediate and Boundary Event simulation properties include probability of occurrence and time of occurrence. Like all simulation parameters, they can be expressions, not just constants. Timer Event is determined by the duration parameter (in combination with the Activity's duration). Error Event on a Subprocess boundary is triggered by an Error End Event in the Subprocess.

Inputs and outputs for Tasks are similar to what is done in typical Lean process improvement/redesign projects. Inputs for Tasks include: PT (processing time) is the active time per instance, with no interruptions; LT (lead time) is the additional time, after starting to work on it, for interruptions. Also important, is the batch time (BT), since work is often batched and forwarded N times a day instead of immediately. The Wait time (WT) is the time before the performer begins work on it. In BPMN terms, the Task has started but the performer has not started work on it yet.

¹ Integration of Workforce Management with a Business Process Management Suite, Robert Shapiro, 2008 BPM and Workflow Handbook.

² BPMN Method & Style, Bruce Silver, 2009

REQUIRED INPUTS TO THE ANALYSIS

There are three inputs to the analysis:

- A BPMN business model, serialized in either the XPD format or the BPMN2.0 format.
- An EXCEL Workbook containing the statistical performance characteristics of all the processes in the model.
- An EXCEL Workbook containing the arrival patterns for the work (work items) to be processed.

We discuss the details of these inputs in what follows.

BPMN Business Model

There are a number of modeling tools now available for creating and editing BPMN models. Prior to the completion of BPMN2.0, the only standardized serialization of BPMN models was the xml format defined in XPD2.1³. With the completion of BPMN2.0 there is now a second standardized serialization. The analysis tool accepts either format.

The BPMN2.0 specification⁴ defines four types of conformance, namely *Process Modeling Conformance*, *Process Execution Conformance*, *BPEL Process Execution Conformance* and *Choreography Modeling Conformance*. Here we focus on *Process Modeling*. Within *Process Modeling* there are subclasses. Our focus is the *Analytics* subclass which we believe is most appropriate for business analysts studying the structural characteristics of business processes. For a complete definition of the *Analytics* subclass refer to the BPMN2.0 specification.

In addition to the ability to load BPMN models in either serialization, the analysis tool contains an editor which allows visualization of the model and the ability to modify it if need be in the course of studying performance.

The behavioral interpretation of the BPMN model uses certain special features described in **BPMN METHOD & STYLE**.

- Exclusive Data-based Decision Gateways can be linked to End Events in an immediately preceding Sub-Process by the Names used. The exit path in the Sub-Process then determines which Gate is chosen for Sequence Flow out of the Gateway.
- Error Events on a Sub-Process boundary are linked to Error Events in the Sub-Process.

Statistical Performance Characteristics

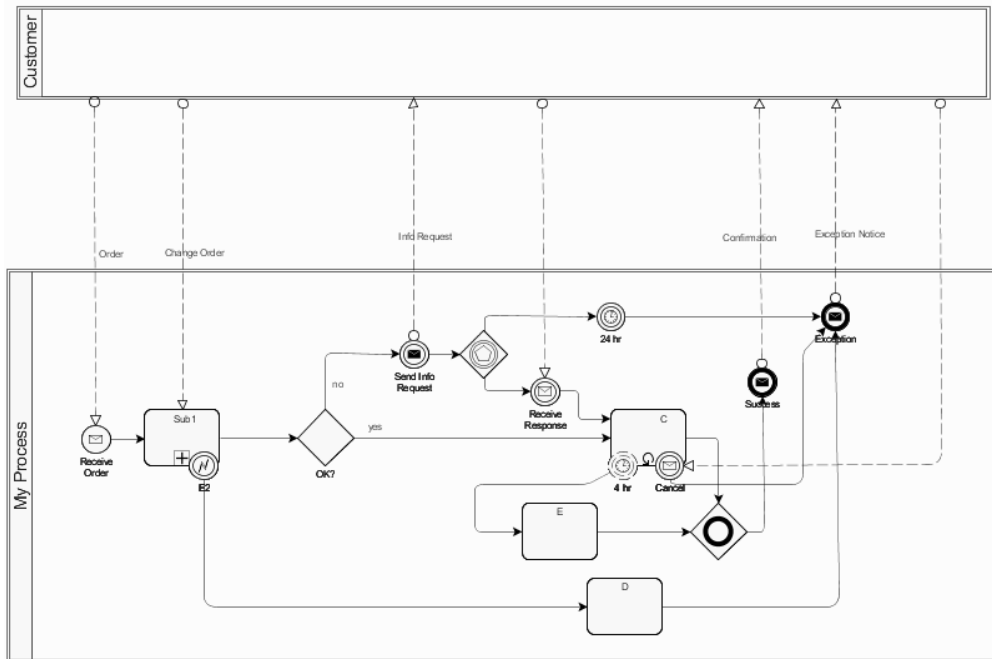
The performance characteristics for a model are contained in a single Excel Workbook. A template for the Workbook is generated by the analysis tool based on the model. For each Process or Subprocess in the model there is a set of Sheets. The set consists of three types of Sheets:

- Flow Nodes
- Routes
- Properties

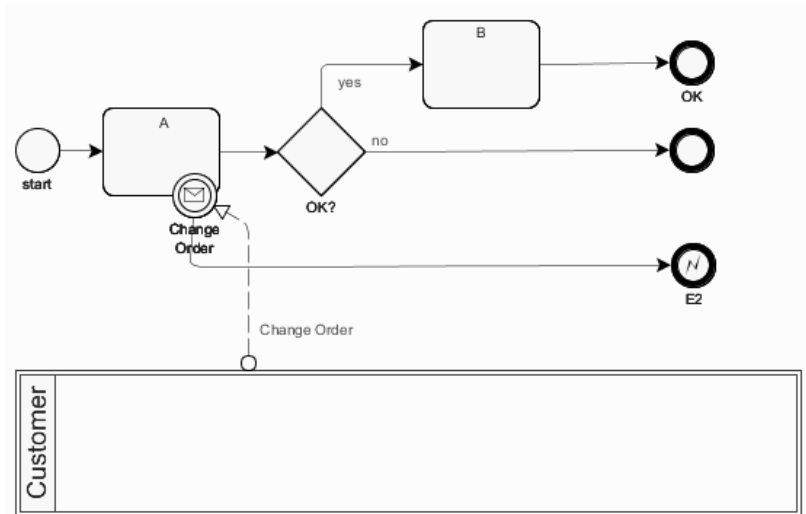
In the discussion that follows we will refer to one sample model which we depict here.

³ XPD2.2: Incorporating BPMN2.0 Modeling Extensions, 2010 BPM & Workflow Handbook and also XPD2.1 – Integrating Process Interchange & BPMN <http://www.wfmc.org/>

⁴ Business Process Model and Notation Version 2.0, June 2010, OMG Document Number: dtc/2010-06-05 Standard document URL: <http://www.omg.org/spec/BPMN/2.0>



Analysis Example



Sub Process

Flow Nodes Sheet

The Activities, Gateways and Events for a Process or Sub-Process are listed as rows on this sheet. Some of the information (columns) are determined by the model and cannot be changed. Other information characterizes the performance. The specifics of the behavior depend on model details.

- Activity: All activities can have Looping or Multi-Instance specifications
- Task: Tasks can have Wait Time, Processing Time, Batch Time and Lead Time specifications.

- Event: All Intermediate and Boundary Events can have a Time-of-Occurrence specification
- Events other than Timer can have a probability specification.
- Exclusive Event-Based Gateways can have a routing rule specification.

Since the intended user of the tool is a Business Analyst rather than an IT professional, these specifications are all done without requiring knowledge of a scripting language (such as JavaScript).

The specification is done by filling in a preformatted Sheet. This is the most efficient way of supplying the simulation specific information. A section of the Sheet for the Sub-Process is depicted in the following diagram:

| Name | Id | Node Type | Trig/Result | Duration |
|--------------|----|------------------------|-------------|-----------------------------|
| <78> | 78 | end event/None | None | |
| A | 70 | task | | 10.0 (min) with spread:10 % |
| B | 73 | task | | 20.0 (min) with spread:10 % |
| Change Order | | boundary event/Message | Message | |
| E2 | 80 | end event/Error | Error | |
| OK | 76 | end event/None | None | |
| OK? | 72 | exclusiveData gateway | | |
| start | 69 | start event/None | None | |

Flow Node Sheet (partial)

The left four columns are determined by the BPMN model and cannot be changed.

In addition, the tool also supports modification of any specification by user-invoked dialogs. We illustrate with diagrams that show an abbreviated version of the sheet and a dialog for editing the Task specification:

Flow Nodes

| Name | Id | Node Type | Duration | Wait Time | Probability | When |
|--------------|----|------------------------|----------------------------|-----------|-------------|-----------|
| <78> | 3 | end event/None | | | | |
| A | 3 | task | 10.0 (min) with spread:... | 4.0 (min) | | |
| B | 3 | task | 20.0 (min) with spread:... | 4.0 (min) | | |
| Change Order | 3 | boundary event/Message | | | 5.0% | 0.0 (min) |
| E2 | 3 | end event/Error | | | | |
| OK | 3 | end event/None | | | | |
| OK? | 3 | exclusiveData gateway | | | | |
| start | 3 | start event/None | | | | |

Abbreviated Sheet Panel

Expression Builder for: A:Duration

Current Expression
 10.0 (min) + quantity * 5.0 with spread:10 %

The top line in this view is the current formula for determining the processing time for a task. You cannot edit the formula. You use this dialog to edit the formula fragments and test them. For specifying a duration, wait time or the time of occurrence of an event there are two different expression formats available: ARITHMETIC or WITH CONDITION. You can switch between them using the radio buttons on the left side of the dialog.

Expression Fragments

ARITHMETIC 10.0 MINUTES + quantity * 5 + * 0.0 +/-

WITH CONDITION 10 %

Change any entry and press Apply. Property names selected by scroll button.

Preview Apply Cancel

Dialog for Specifying Task Processing Duration

Routes Sheet

The Routes specification determines the rule to use in choosing the exit gate for Sequence Flow out of an Exclusive Data-Based Decision Gateway. The following Sheet fragment shows the specification for the Gateway in the Sub-Process:

| Gateway Name | Id | Routing Type | Route Name |
|--------------|----|--------------|------------|
| OK? | 72 | defaults | no |
| | | | yes |

Route Sheet (partial)

In this case the default logic reverts to choosing the exit gate at random. The user could have designated the use of Weights to bias the choice, or the use of Conditional Expressions to select the Gate based on testing properties of the Work Item. If the gateway were preceded by a sub-process, then if an End Event of that subprocess was labeled (in the BPMN modeler) 'OK' it would control the Gate selection, choosing the 'Yes' Gate only if the subprocess terminated with that End Event.

Properties Sheet

The Properties specification allows the user to define Work Item properties specifically for the analysis. BPMN2.0 uses Data Objects and associated constructs for this purpose. XPD L provides Data Fields. These, if present in the model, provide initial definitions for the Properties specification. The user may extend or override these.

The properties are used in other specifications to allow the handling and behavior of work items to depend upon the values of their properties. This is illustrated in the Dialog for Specifying Task Processing Duration (above) where the value of the 'quantity' property affects the duration calculation. Properties are also commonly used in the condition expressions for routing logic and the time-of-occurrence of Events. The Arrival Patterns supply the property values for Work Items. The following is a snippet from the Properties Sheet:

| Property Name | Data Type | Value |
|---------------|-----------|-------|
| quantity | INTEGER | 5 |

Route Sheet (partial)

Arrival Patterns

The Arrival Patterns for executing a model are contained in a single Excel Workbook. A template sheet for the Workbook is generated by the analysis tool based on the model. The template sheet is also influenced by the Properties Sheet just discussed.

| Name | ProcessId | Arrival Time:yyyy-MM-dd HH:mm:ss | How Many | quantity |
|------------|-----------|----------------------------------|----------|----------|
| Monday Q20 | 2 | 2010-10-25 09:00:00 | 20 | 10 |
| Monday Q10 | 2 | 2010-10-25 09:00:00 | 10 | 5 |

Arrival Sheet (partial)

ANALYTICS GENERATED BY THE SIMULATION

Applying the statistical specifications to the Arrival Patterns generates a collection of analysis tables and charts. These are all available for interactive use within the tool and for export to EXCEL spreadsheets for independent analysis, depiction

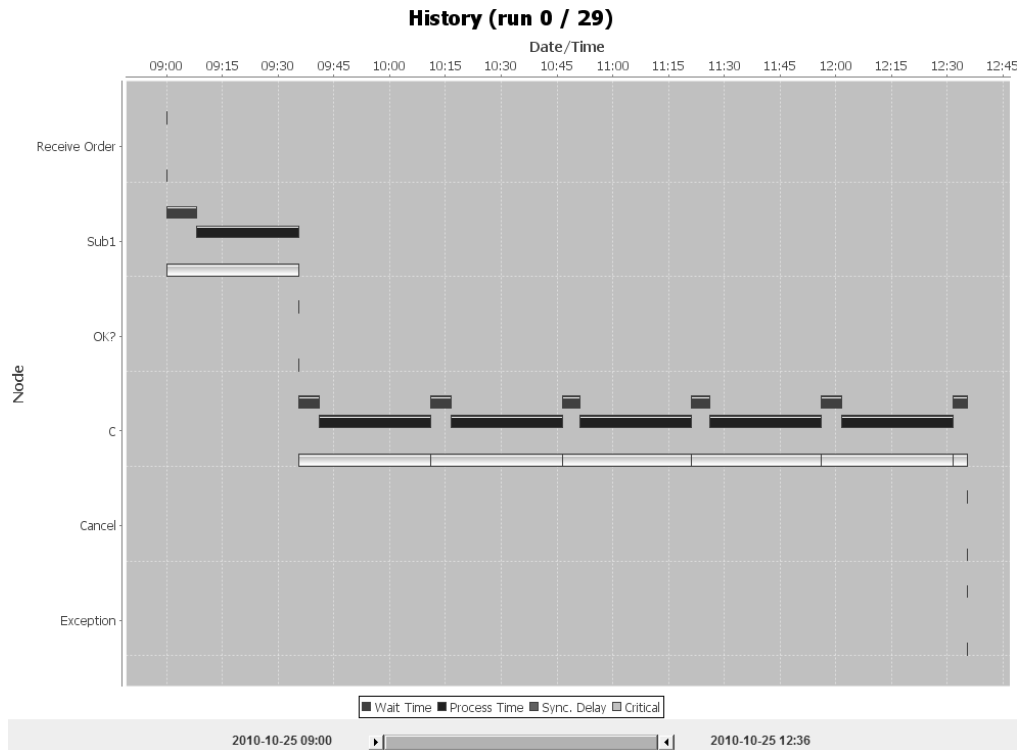
and transfer to other tools used by Business Analysts. Here we mention some of the most useful Charts.

History Chart

This chart is an interactive window into the paths taken by all the arriving work items through the BPMN model. Controls allow the user to determine the level of detail, the focus in terms of processes and flow nodes selected, the time frame etc. The work item paths can be grouped in a variety of ways and a path pattern analyzer uses heuristics to identify the set of unique patterns generated by the complete arrival set.

The representation is based on Gantt charts⁵. The selected set of flow nodes is depicted on the left edge and each flow node that a work item traverses has one or more horizontal bars depicting the start and end times for each occurrence of that node. Parallel paths, critical paths and slack time (synchronization delays) are all depicted. Multiple occurrences of the same activity (e.g. from use of the BPMN Multi-Instance attribute) are folded into a single bar, but can be unfolded as well.

The following chart shows a bird’s eye view of the example model, depicting detail only for the top level process, with single bars for each sub-process. The particular path is for a work item that encounters a boundary interrupt during the 6th looping repetition of Task ‘C’. (In this example there is no parallelism: parallel paths are associated with parallel gateways and non-interrupting boundary events).

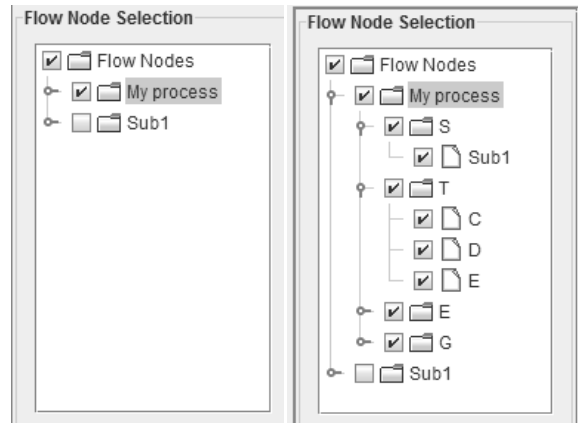


Bird’s Eye View (grayscale depiction⁶)

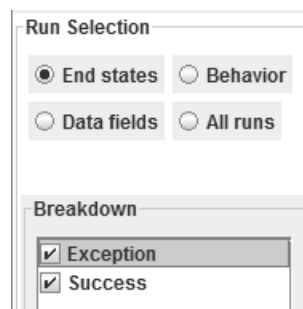
⁵ http://en.wikipedia.org/wiki/Gantt_chart

⁶ The Analytics tool uses colors to distinguish the different kinds of bars.

History Filters



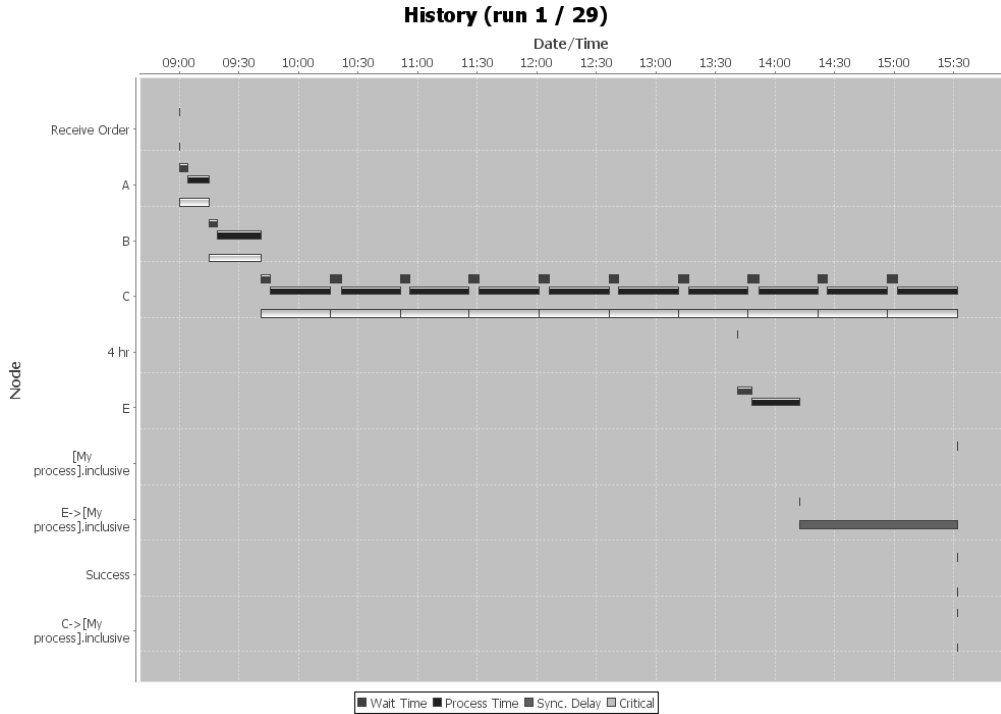
On the left is the setting for the Bird's Eye View. On the right is an Expansion to show scope of selection. Flow node types are S=Sub-Process, T=Task, E=Event, G=Gateway. The node names appear after the page icon.



Run Selection establishes the criteria for grouping paths.

- End States: Groups by the End Events in the top level process.
- Behavior: Groups by the path patterns heuristics which the user can adjust.
- Data fields: Groups by property values of the Work Items.

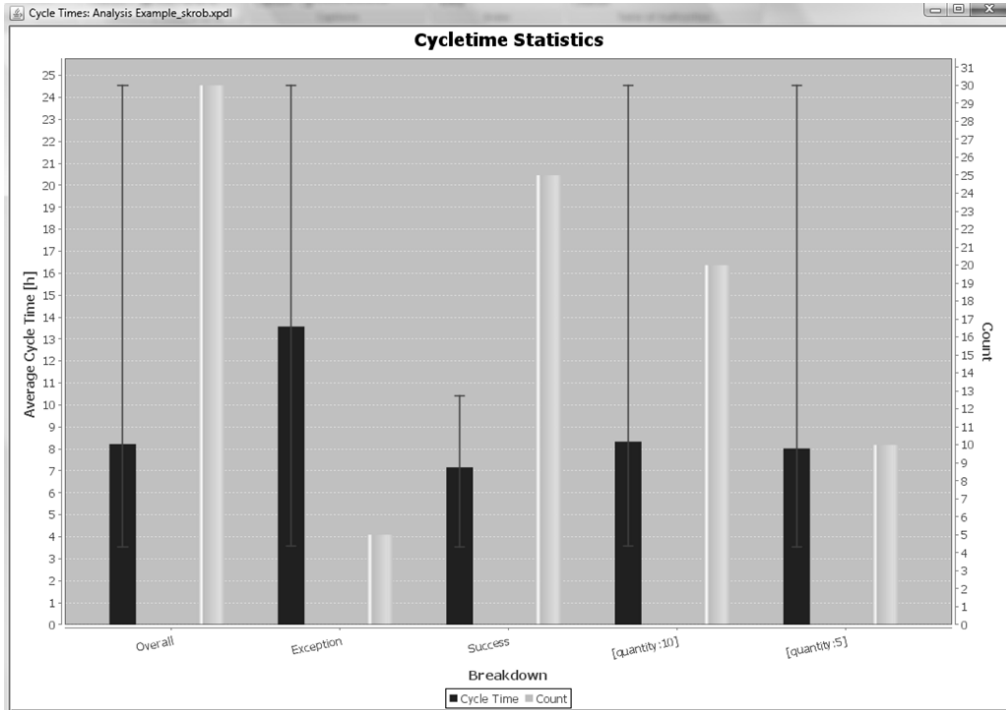
This ends a brief description of the History controls. We expect that real use by Business Analysts will aid in fine tuning our approach and adding other views that can aid in understanding the behavior of a model. In concluding this section we show one more history chart with some parallelism due to a non-interrupting boundary event.



Parallel Paths due to non-interrupting Boundary Event

Process Cycle Time

This chart depicts the minimum, maximum and average cycle times for groups of work items and the number of work items in each group. We anticipate broadening the ability to specify the grouping criteria to match the capabilities in the History filters described previously. We are also complementing the choice of presentations to include a Histogram format.

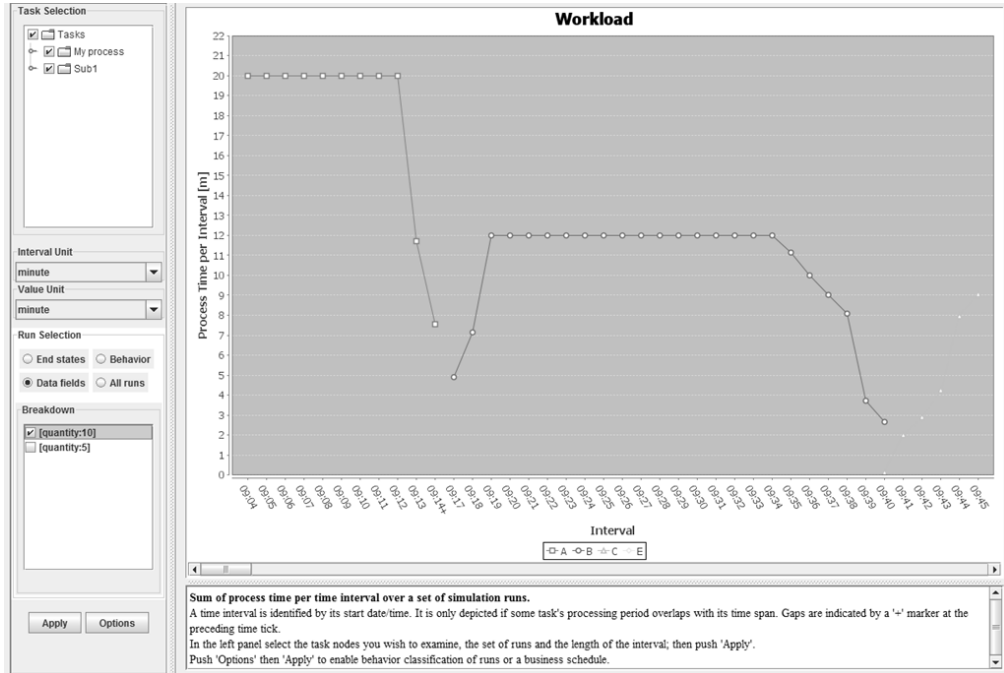
**(Grayscale Depiction)****Workload Wait Time and similar charts for Task Specific Durations**

These charts make extensive use of color so the depiction in this paper is curtailed.

The Workload chart presents the processing time associated with every Task as a collection of lines, one for each task. The Y axis represents the processing time in a time period and the X axis represents clock time.

The user can determine which tasks to include and which subset of the work items to include, using the same filtering techniques available in the History chart. The user can also choose the size of a time period and the scale for the Y axis.

These controls allow viewing the overall picture of task processing in a single screen, or detailed examination of any time segment. Here is a view generated by our example model:



The same viewing technology is used for Wait Time and other Task-related metrics. The data is all exportable to EXCEL.

It should be noted that the Workload statistics is important for determining the resource requirements for a BPM enactment system. It also feeds into the analysis of costs and their allocation. Workload and Wait Time data, combined with critical path information, play an important role in optimization.

Task, Event and Exclusive Data-Based Decision Statistics

These Tables present basic information about the Tasks, Events and Decision Gateways.

| History | Process Statistics | Task Statistics | Event Statistics | XOR Statistics | | | | | |
|---------|--------------------|-----------------|------------------|------------------|----------|-------------|--------|--------|-----------|
| row ^ | Name | Process | Frequency | Interrupt Fre... | Waittime | Processtime | PT min | PT max | Slacktime |
| 0 | C | My process | 7.43 | 0.01 | 300.722 | 1,775.785 | 0 | 1,800 | 0 |
| 1 | E | My process | 0.53 | 0.00 | 484.666 | 1,440 | 1,440 | 1,440 | 0 |
| 2 | <34> | My process | 1.60 | 0.00 | 0 | 0 | 0 | 0 | 1,556.774 |
| 3 | D | My process | | | | | | | |
| 4 | A | Sub1 | 1.00 | 0.00 | 240 | 598.698 | 540 | 660 | 0 |
| 5 | B | Sub1 | 0.50 | 0.00 | 240 | 1,229.268 | 1,080 | 1,320 | 0 |

| History | Process Statistics | Task Statistics | Event Statistics | XOR Statistics | |
|---------|--------------------|-----------------|------------------|----------------|-------------|
| row ^ | Name | Process | Type | Frequency | Processtime |
| 0 | start | Sub1 | S | 1.00 | 0 |
| 1 | <78> | Sub1 | E | 0.50 | 0 |
| 2 | Receive Re... | My process | I | 0.43 | 0 |
| 3 | 4 hr | My process | B | 0.53 | 0 |
| 4 | Success | My process | E | 0.83 | 0 |
| 5 | 24 hr | My process | I | 0.07 | 0 |
| 6 | Send Info ... | My process | I | 0.50 | 0 |
| 7 | Exception | My process | E | 0.17 | 0 |
| 8 | E2 | My process | B | | |
| 9 | Cancel | My process | B | 0.10 | 0 |
| 10 | E2 | Sub1 | E | | |
| 11 | Change Or... | Sub1 | B | | |
| 12 | Receive Or... | My process | S | 1.00 | 0 |
| 13 | OK | Sub1 | E | 0.50 | 0 |

| History | Process Statistics | Task Statistics | Event Statistics | XOR Statistics |
|---------|--------------------|-----------------|------------------|----------------------------------|
| row ▲ | Name | Process | Frequency | Branch Frequencies |
| 0 | OK? | Sub1 | 1.00 | [<78>:0.50, B:0.50] |
| 1 | OK? | My process | 1.00 | [C:0.50, Send Info Request:0.50] |

THE SIMULATION ENGINE AND PERFORMANCE ANALYZER

Engine Architecture

The engine incorporates a novel approach to reduce the amount of time performing a simulation for a large number of work items. Instead of proceeding in strict time order, as the typical discrete event simulator would, the engine follows pathways into the future and if subsequent events cancel a pathway, the log events generated on that pathway are erased. Future experimentation will determine whether this offers any real benefits.

It should be noted that analysis of the log events generated by simulation consumes more CPU power and memory than the simulation proper. That limits the benefits obtainable from a speedup in the engine.

Environment

The Performance Analyzer is currently embedded in the Sketchpad⁷ desktop application for creating and editing BPMN models. In principle, it could be divided into two independent web-based facilities. The simulation facility, fed with a BPMN model (using either XPD L or BPMN XML formats) and the two EXCEL Workbooks, would generate a stream of log events for the analytics facility. The analytics facility would deliver the results of analysis (charts, data) to the Business Analyst.

This is a long-term objective. In the meanwhile it is critical that Business Analysts use the tool and provide the feedback necessary to refine the approach and improve the user interface.

Future Extensions

Previous work by the authors focused on Automatic Optimization⁸, using a discrete event simulator⁹ and sophisticated analysis of historical data generated either by an enactment engine or by the simulator. We are incorporating these ideas so that our tool will be able to generate suggestions for process improvement.

SUMMARY

We have described an ongoing effort to provide performance analysis and business process optimization tools to Business Analysts who currently do this kind of work without building sophisticated process models or simulation of such models. We believe that a sub-class of BPMN2.0 and a tool like the Process Analytica Performance Analyzer could be of value to this community.

⁷ Sourceforge Open Source project: <http://sketchpadbpmn.sourceforge.net/>

⁸ The AutoOptimizer, Robert Shapiro and Hartmann Genrich, 2009 BPM & Workflow Handbook

⁹ The Cape Visions ‘Vision Forecaster’, used by FileNet/IBM, Fujitsu and Global 360.

