XES Software Telemetry Extension

Citation for published version (APA):

Document status and date:
Published: 20/11/2017

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 12. May. 2019
1 Introduction and Background

During the execution of software, execution data can be recorded. With the development of process mining techniques on the one hand, and the growing availability of software execution data on the other hand, a new form of software analytics comes into reach. That is, applying process mining techniques to analyze software execution data. To enable process mining for software, event logs should be capable of capturing software-specific data.

When analyzing the performance of software applications, not only timing, but also resource utilization is of importance. The Software Telemetry extension supports the recording of basic performance profile related resource utilizations commonly used in many software profiler tools. The resource utilizations covered in this extension are: CPU usage, thread usage and memory usage.

Note that detailed timing information is already captured via the Software Event extension. Advanced resource utilization, like database requests, cache usage, network and socket usage and file I/O, are considered out of scope for this extension.

Extension definition

<table>
<thead>
<tr>
<th>Name</th>
<th>Software Telemetry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefix</td>
<td>swtelemetry</td>
</tr>
<tr>
<td>URI</td>
<td><a href="http://www.xes-standard.org/swtelemetry.xesext">http://www.xes-standard.org/swtelemetry.xesext</a></td>
</tr>
</tbody>
</table>

The remainder of this extension is organized as follows. In Section 2 we explain some basic terminology. In Sections 3, 4 and 5 we detail how various telemetry stats are recorded. Section 6 provides an example, and the XES Extension definition is given in Section 7. Finally, in Section 8 provides a reference glossary.

2 Terminology

In this extension, we will use some software-specific terminology. We provided a reference glossary in Section 8. As explained in the introduction, this extension supports the recording of basic performance profile related resource utilizations commonly used in many software profiler tools: **CPU usage, thread usage and memory usage.**

**CPU usage** When a program is executing, it consumes processor / CPU time. Most of this time is consumed in what is known as user space. When a program loops through an array, it is accumulating user CPU time. System time is the amount of time the CPU was busy executing code in kernel space. When a program executes a system call such as exec or fork, it is accumulating system CPU time. The CPU usage telemetry indicates how much of the CPU time was used. This gives a load indication for the CPU resource.

**Thread usage** A thread is a program’s path of execution. A multi-threaded program can execute multiple code paths in parallel, and possibly concurrently. A daemon thread is a background thread, typically used for providing a particular service for other threads. The thread count telemetry gives an indication of how many execution paths are concurrently active.

**Memory usage** As a program is executing, various objects and arrays are instantiated and destroyed. These objects are stored in what is known as the heap memory. The heap memory is used to allocate blocks of RAM memory on demand for use by the program, and can grow during the execution of a program. The memory usage telemetry indicates how much of the RAM memory was used. This gives a load indication for the RAM memory resource. Typically, this telemetry refers to the used heap memory.

3 CPU Usage

The CPU usage telemetry indicates how much of the CPU time was used. This gives a load indication for the CPU resource. We define the following CPU
usage attributes.

<table>
<thead>
<tr>
<th>Level</th>
<th>Key</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>event</td>
<td>cpuTotalUser</td>
<td>int</td>
<td>The total time in milliseconds that the CPUs spent in user space.</td>
</tr>
<tr>
<td>event</td>
<td>cpuTotalKernel</td>
<td>int</td>
<td>The total time in milliseconds that the CPUs spent in kernel space.</td>
</tr>
<tr>
<td>event</td>
<td>cpuTotalIdle</td>
<td>int</td>
<td>The total time in milliseconds that the CPUs spent idle.</td>
</tr>
<tr>
<td>event</td>
<td>cpuLoadUser</td>
<td>float</td>
<td>The fraction of time that the CPUs spent in user space. 1 represents 100% usage, 0 represents 0% usage.</td>
</tr>
<tr>
<td>event</td>
<td>cpuLoadKernel</td>
<td>float</td>
<td>The fraction of time that the CPUs spent in kernel space. 1 represents 100% usage, 0 represents 0% usage.</td>
</tr>
</tbody>
</table>

4 Thread Count

The thread count telemetry gives an indication of how many execution paths are concurrently active. We define the following thread count attributes.

<table>
<thead>
<tr>
<th>Level</th>
<th>Key</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>event</td>
<td>threadTotal</td>
<td>int</td>
<td>The total number of active threads.</td>
</tr>
<tr>
<td>event</td>
<td>threadDaemon</td>
<td>int</td>
<td>The number of active daemon threads.</td>
</tr>
</tbody>
</table>

5 Memory Usage

The Memory usage telemetry indicates how much of the RAM memory was used. This gives a load indication for the RAM memory resource. Typically, this telemetry refers to the used heap memory. We define the following memory usage attributes.

<table>
<thead>
<tr>
<th>Level</th>
<th>Key</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>event</td>
<td>memoryUsed</td>
<td>int</td>
<td>The amount of memory used, measured in bytes.</td>
</tr>
<tr>
<td>event</td>
<td>memoryTotal</td>
<td>int</td>
<td>The amount of memory available, measured in bytes.</td>
</tr>
<tr>
<td>event</td>
<td>memoryLoad</td>
<td>float</td>
<td>The fraction of memory in use. 1 represents 100% usage, 0 represents 0% usage.</td>
</tr>
</tbody>
</table>

6 Example

In Listing 1 an example XES trace is given for a simple application with telemetry data. We have annotated the load attributes with how the load values
were computed based on the differences between events. For example, the `cpuLoadUser` value in the second event is calculated by taking the difference in `cpuTotalUser` divided by the total difference, i.e., user plus kernel plus idle. In Figure 1, example performance charts are given, based on the telemetry data recorded in the example XES log in Listing 1.

Figure 1: Example performance charts for the telemetry data recorded in the example XES trace in Listing 1.

Listing 1: Example XES log for a simple application with telemetry data.

```xml
<log>
  <extension name="Concept" prefix="concept"
  <extension name="Software Telemetry" prefix="swtelemetry"
  <trace>
    <event> -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- Event 1
      <string key="concept:name" value="readSettings" />
      <int key="swtelemetry:cpuTotalUser" value="20" />
      <int key="swtelemetry:cpuTotalKernel" value="0" />
      <int key="swtelemetry:cpuTotalIdle" value="60" />
      <float key="swtelemetry:cpuLoadUser" value="0.25" />
      <int key="swtelemetry:threadTotal" value="1" />
      <int key="swtelemetry:threadDaemon" value="0" />
      <int key="swtelemetry:memoryUsed" value="2097152" />
      <int key="swtelemetry:memoryTotal" value="1073741824" />
      <float key="swtelemetry:memoryLoad" value="0.00" />
    </event>
    <event> -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- Event 2
      <string key="concept:name" value="loadFile" />
      <int key="swtelemetry:cpuTotalUser" value="37" />
      <int key="swtelemetry:cpuTotalKernel" value="34" />
      <int key="swtelemetry:cpuTotalIdle" value="159" />
      <float key="swtelemetry:cpuLoadUser" value="0.1" />
      <float key="swtelemetry:cpuLoadKernel" value="0.2" />
      <int key="swtelemetry:threadTotal" value="1" />
      <int key="swtelemetry:threadDaemon" value="0" />
      <int key="swtelemetry:memoryUsed" value="6291456" />
      <int key="swtelemetry:memoryTotal" value="1073741824" />
      <float key="swtelemetry:memoryLoad" value="0.01" />
    </event>
    <event> -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- Event 3
      <string key="concept:name" value="writeSettings" />
      <int key="swtelemetry:cpuTotalUser" value="10" />
      <int key="swtelemetry:cpuTotalKernel" value="15" />
      <int key="swtelemetry:cpuTotalIdle" value="75" />
      <float key="swtelemetry:cpuLoadUser" value="0.15" />
      <int key="swtelemetry:threadTotal" value="1" />
      <int key="swtelemetry:threadDaemon" value="0" />
      <int key="swtelemetry:memoryUsed" value="1073741824" />
      <int key="swtelemetry:memoryTotal" value="1073741824" />
      <float key="swtelemetry:memoryLoad" value="0.00" />
    </event>
  </trace>
</log>
```
7 XES Extension

Listing 2: XES Extension - Software Telemetry.

```xml
<extension name="Software Telemetry" prefix="swtelemetry" uri="http://www.xes-standard.org/swtelemetry.xesext"/>
<event>
<int key="cpuTotalUser" value="129"/>
<int key="cpuTotalKernel" value="57"/>
<int key="cpuTotalIdle" value="203"/>
<float key="cpuLoadUser" value="0.4"/>
<float key="cpuLoadKernel" value="0.1"/>
<int key="threadTotal" value="2"/>
<int key="threadDaemon" value="1"/>
<int key="memoryUsed" value="15728640"/>
<int key="memoryTotal" value="1073741824"/>
<float key="memoryLoad" value="0.02"/>
</event>
<event>
<int key="cpuTotalUser" value="231"/>
<int key="cpuTotalKernel" value="57"/>
<int key="cpuTotalIdle" value="255"/>
<float key="cpuLoadUser" value="0.3"/>
<float key="cpuLoadKernel" value="0.0"/>
<int key="threadTotal" value="2"/>
<int key="threadDaemon" value="1"/>
<int key="memoryUsed" value="16777216"/>
<int key="memoryTotal" value="1073741824"/>
<float key="memoryLoad" value="0.02"/>
</event>
</trace>
</log>
```
8 Glossary

**user CPU time** User time is the amount of time the CPU was busy executing code in user space. When a program loops through an array, it is accumulating user CPU time.

**system CPU time** System time is the amount of time the CPU was busy executing code in kernel space. When a program executes a system call such as `exec` or `fork`, it is accumulating system CPU time.

**kernel CPU time** See System CPU time.

**heap memory** The heap memory is used to allocate blocks of RAM memory on demand for use by the program, and can grow during the execution of a program.

**thread** A thread is a program’s path of execution. A multi-threaded program can execute multiple code paths in parallel, and possibly concurrently.

**daemon thread** A background thread, typically used for providing a particular service for other threads.