Scientific Micro-Workflows: Where Event-Driven Approach Meets Workflows to Support Digital Twins

Ameer B. A. Alaasam, PhD Student, Department of System programming, School of electrical engineering and computer science, South Ural State University (alaasamab@susu.ru),

Gleb Radchenko, Associate Professor, Department of System programming, School of electrical engineering and computer science, South Ural State University

Andrey Tchernykh, Professor CICESE Research Center (Ensenada, Mexico)

Kirill Borodulin, School of electrical engineering and computer science, South Ural State University

Alexey Podkorytov, School of electrical engineering and computer science, South Ural State University
Digital Twins (DT)

**Fundamentally**, can be defined, as an evolving digital profile of the historical and current behavior of a physical object or process that helps optimize business performance.

**Technically**, The DT concept contains three main parts: *physical products* in a real space. *virtual products* in a virtual space. *connections* of data and information that ties the virtual and real products together.
Digital Twins, little bit deeper
Real example, Uber

A sequence of computational services linked together by a set of edges that represent data dependencies. (**Computational Workflow**)

Carol McDonald, Monitoring Real-Time Uber Data Using Apache APIs, dzone.com, Jul. 03, 2017 · AI Zone
Different Computational Workflows on different resources for each deployment models. Vendor lock?, Changing?, Resource limit?, Ability or right to modify?, The scientists with no technological experience?..Etc (MONEY!) OR (NO RIGHT TO MODIFY!)
Computational (Scientific) Workflows Systems

A sequence of computational services linked together by a set of edges that represent data dependencies. (Computational Workflow), (Scientific Workflow)

The main goal of the Kepler SWfMS is to support different execution scenarios.
SWF Challenges

- Tightly coupled dependencies.
- Sequential execution.
- A job which may succeed on one host but it may fail in another host.
- Lack of fully stream-processing support.

- Running one large-scale SWF faces a series of obstacles. For example, the limitations appear, when we deal with big data problems, including data scale and computation complexity, resource provisioning, and collaboration in heterogeneous environments.
Microservice vs SWf, is it applicable?

**Microservices**
- By nature, the microservice architecture is *loosely coupled*, meaning that there is a small number of links between services and services themselves being independent.


**SWf**
- In scientific workflow each task is *tightly coupled*, often having intricate dependencies on other tasks for example, the input of one task may be produced by other tasks, thus, usually large workflow divide into several phases and executed sequentially.

A microservice is a cohesive, independent process interacting via messages.

The 2 Microservice Architecture:

- **Orchestration**

- **Choreography**
Kepler SWF example

The rising of our Micro-Workflow Idea
Docker is used for running software packages called "containers". All containers are run by a single OS kernel and are thus more lightweight than virtual machines. All containers are isolated from each other and bundle their own tools, libraries, and configuration files; they can communicate with each other through well-defined channels.
Micro-Workflow

Data Flow
Micro-Workflow
Micro-Workflow first result

We used sensors data from DEBS Grand Challenge: Manufacturing equipment. It includes a set of queries to process the data. The delay between two consecutive source data points is about 10 ms.

$\text{Av}_\text{SM}$ (average interval between source messages)
$\text{Av}_\text{RM}$ (average interval between result messages)
$\text{Av}_\text{TAT}$ (average turnaround time)

Table 1. Testing results

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Test time</td>
<td>1 hour</td>
</tr>
<tr>
<td>Number of messages</td>
<td>472279</td>
</tr>
<tr>
<td>$\text{Av}_\text{SM}$ (millisecond)</td>
<td>7.62</td>
</tr>
<tr>
<td>$\text{Av}_\text{RM}$ (millisecond)</td>
<td>7.62</td>
</tr>
<tr>
<td>$\text{Av}_\text{TAT}$ (millisecond)</td>
<td>1.38</td>
</tr>
</tbody>
</table>
Thank you