South Ural State University | National Research University

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

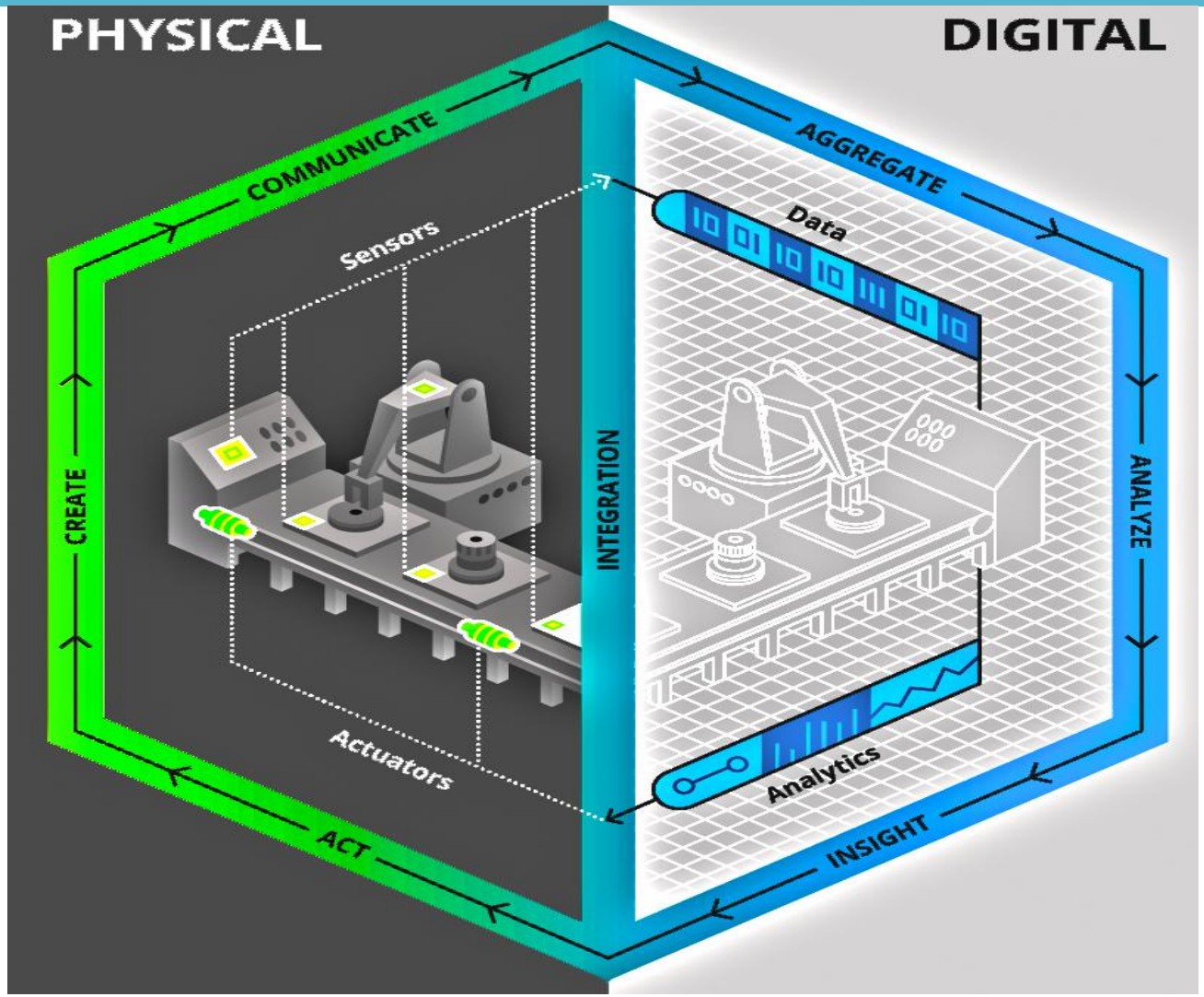
Alexey Podkorytov, School of electrical engineering and computer science, South Ural State University

Digital Twins (DT)

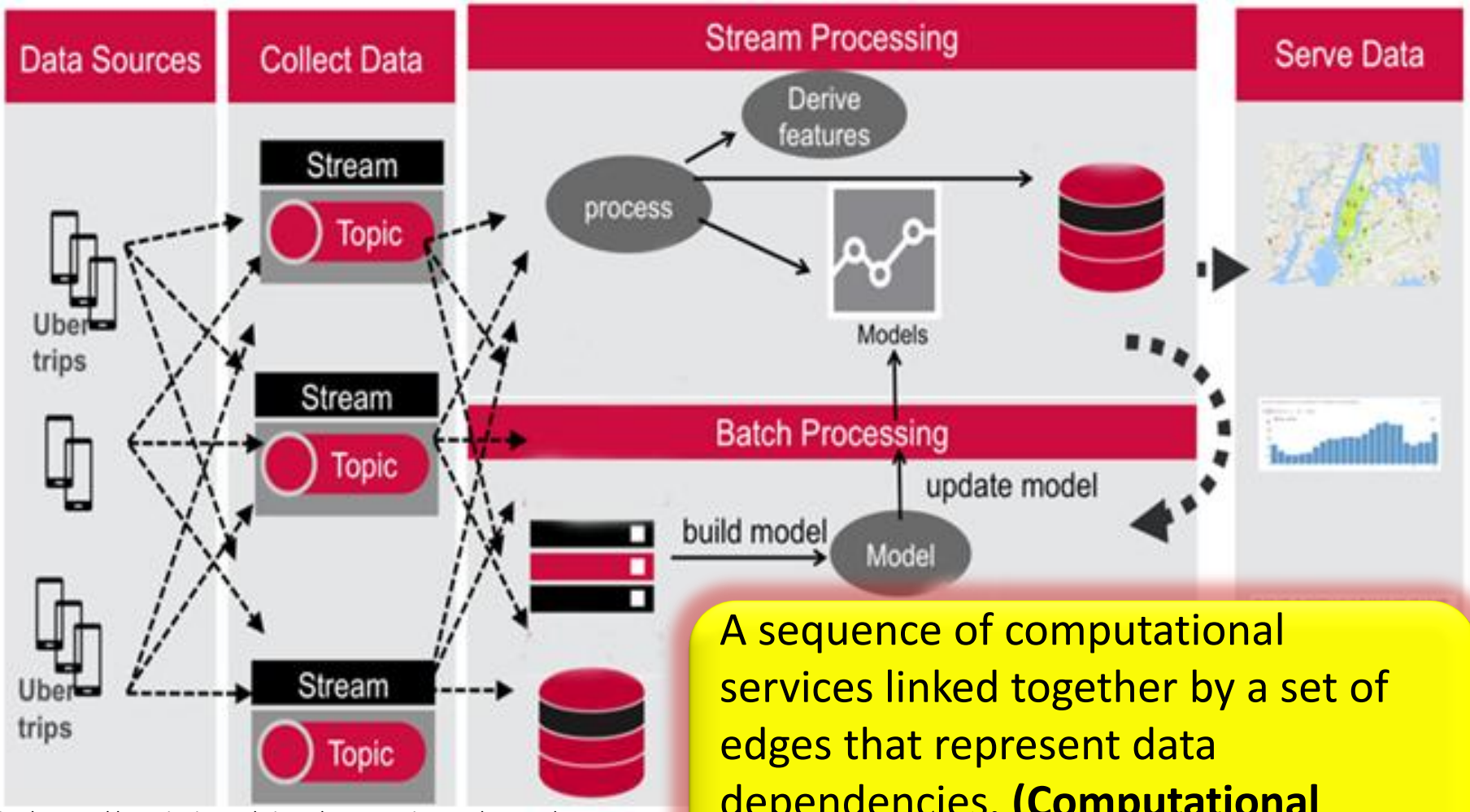
Fundamentally, can be defined, as an evolving digital profile of the historical and current behavior of a physical object or process that helps optimize business performance.

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



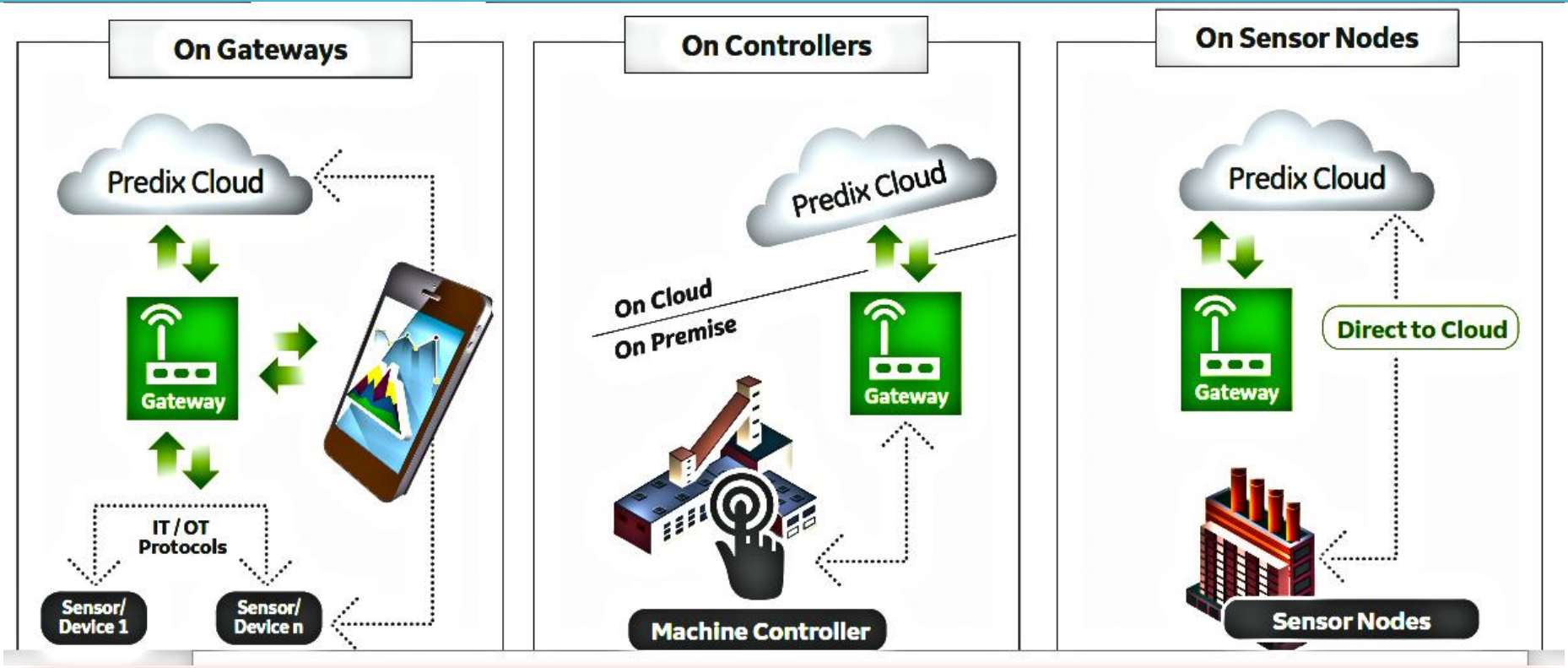
Real example, Uber



Carol McDonald, Monitoring Real-Time Uber Data Using Apache APIs, dzone.com, Jul. 03, 2017 · AI Zone

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

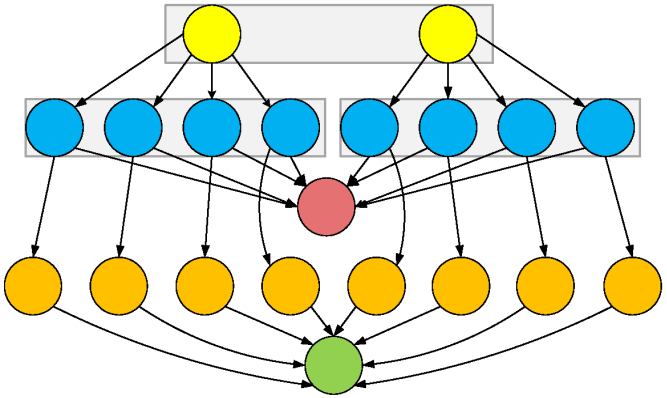
Commercial DT Platform example: (Predix), General Electric Co.

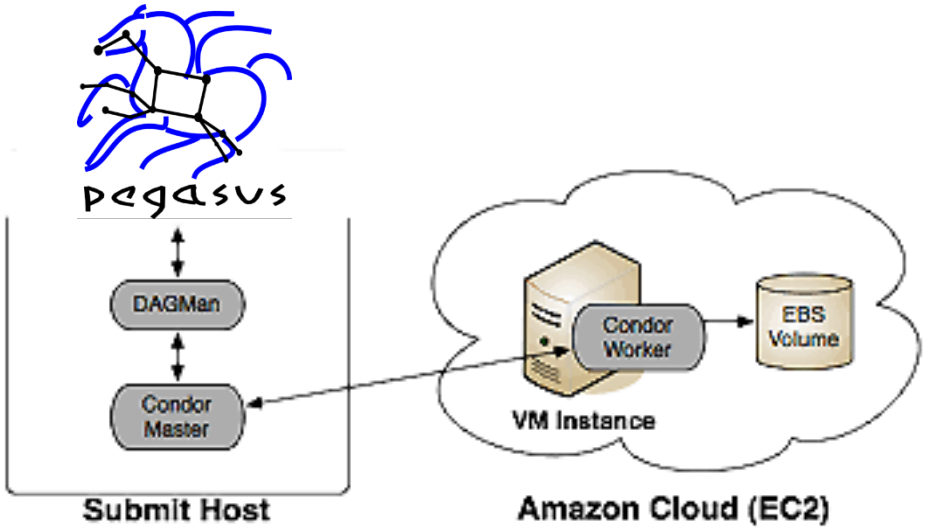


Different Computational Workflows on different resources for each deployment models. Vendor lock?, Changing?, Resource limit?, Ability or right to modify?, The scientists with no technological experience?..Etc (MONEY!) OR (NO RIGHT TO MODIFY!)

Computational (Scientific) Workflows Systems

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



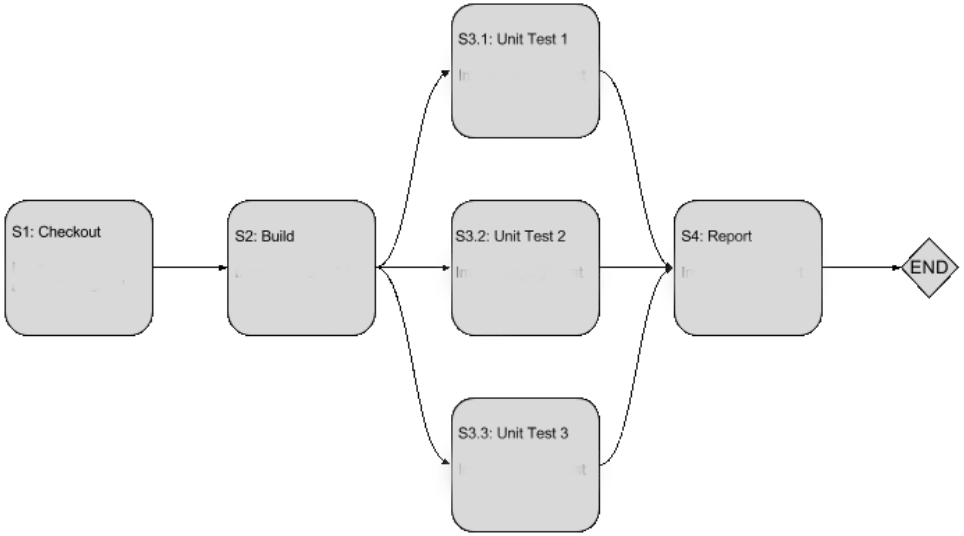


Kepler

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**.

- N. Dragoni *et al.*, “Microservices: Yesterday, Today, and Tomorrow,” *Present Ulterior Softw. Eng.*, pp. 195–216, 2017.

VS !?

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**.

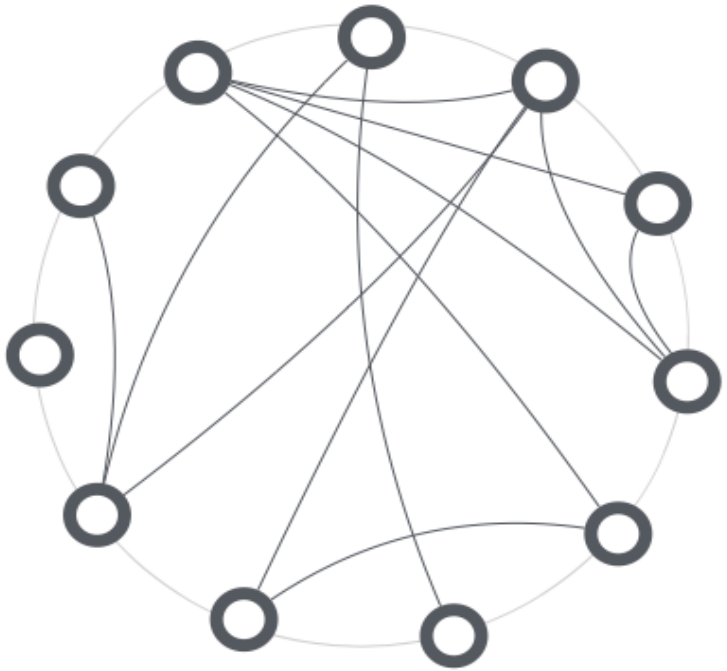
- C. Zheng, B. Tovar, and D. Thain, “Deploying high throughput scientific workflows on container schedulers with makeflow and mesos,” *Proc. - 2017 17th IEEE/ACM Int. Symp. Clust. Cloud Grid Comput. CCGRID 2017*

Microservices

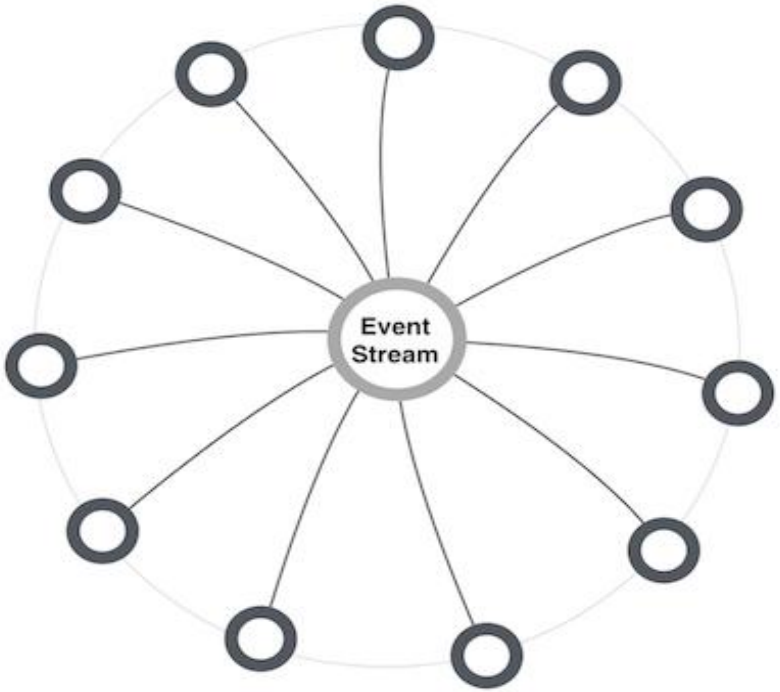
A microservice is a cohesive, independent process interacting **via messages**.

The 2 Microservice Architecture :

- Orchestration



- Choreography



Kepler SWF example

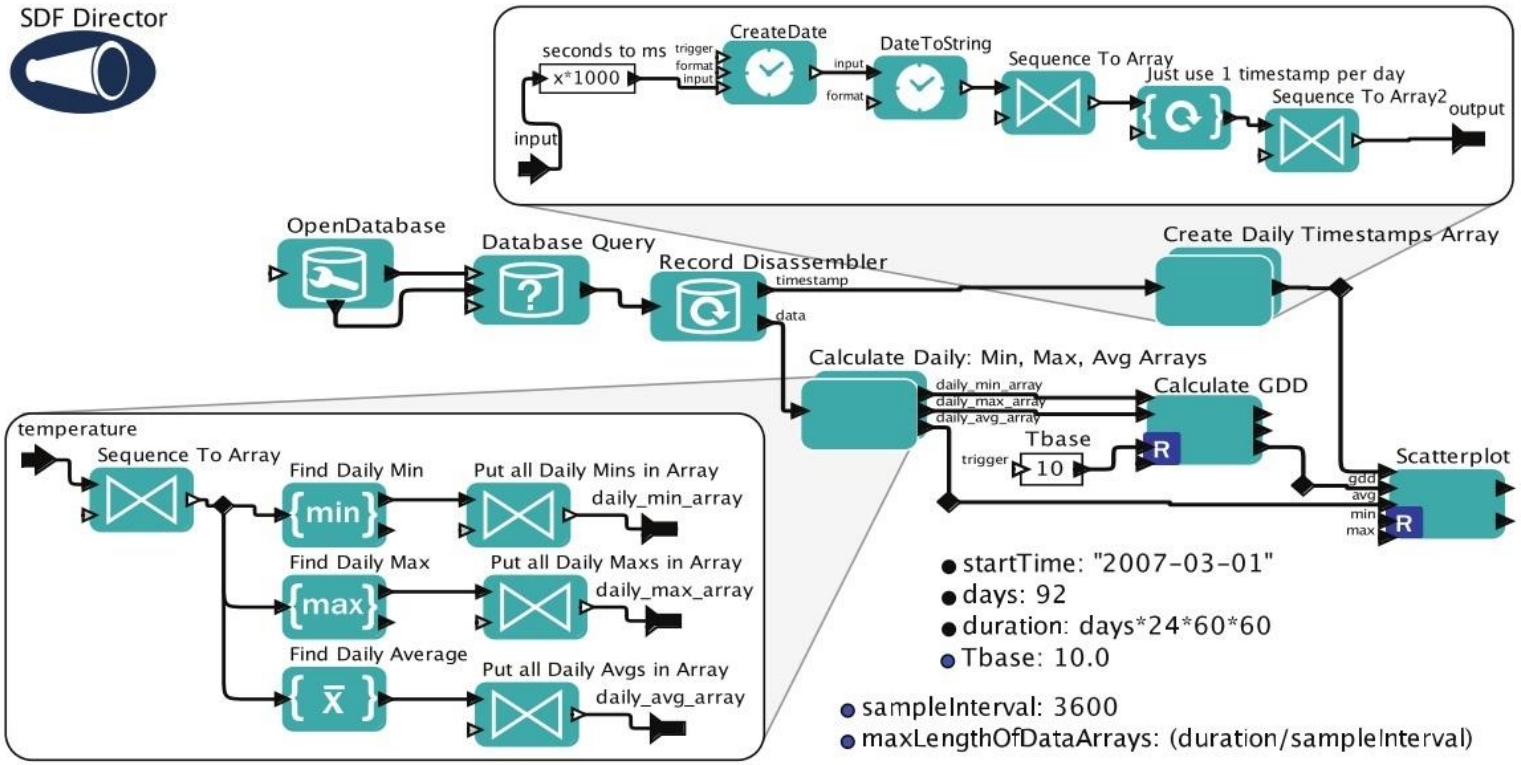
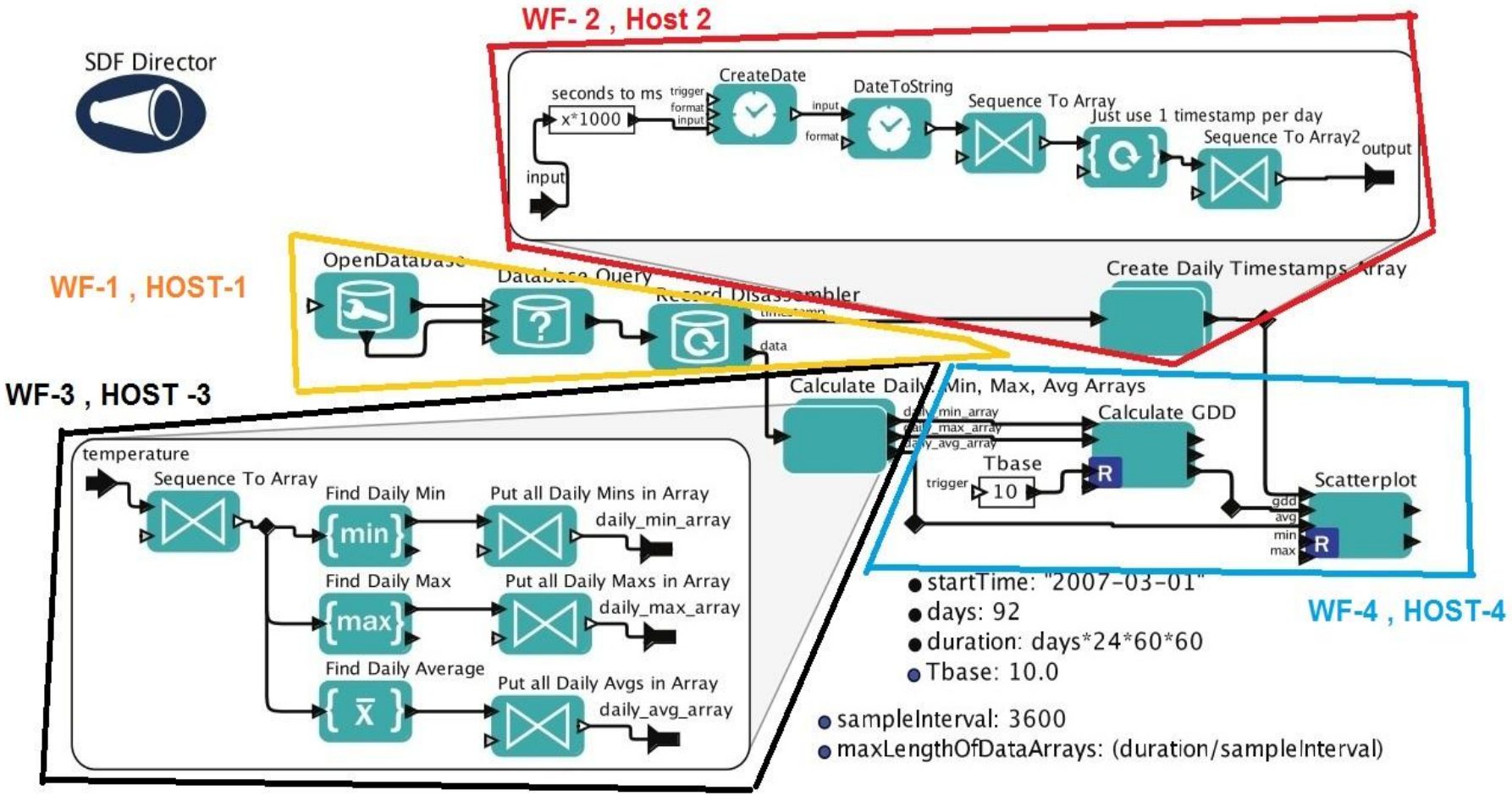


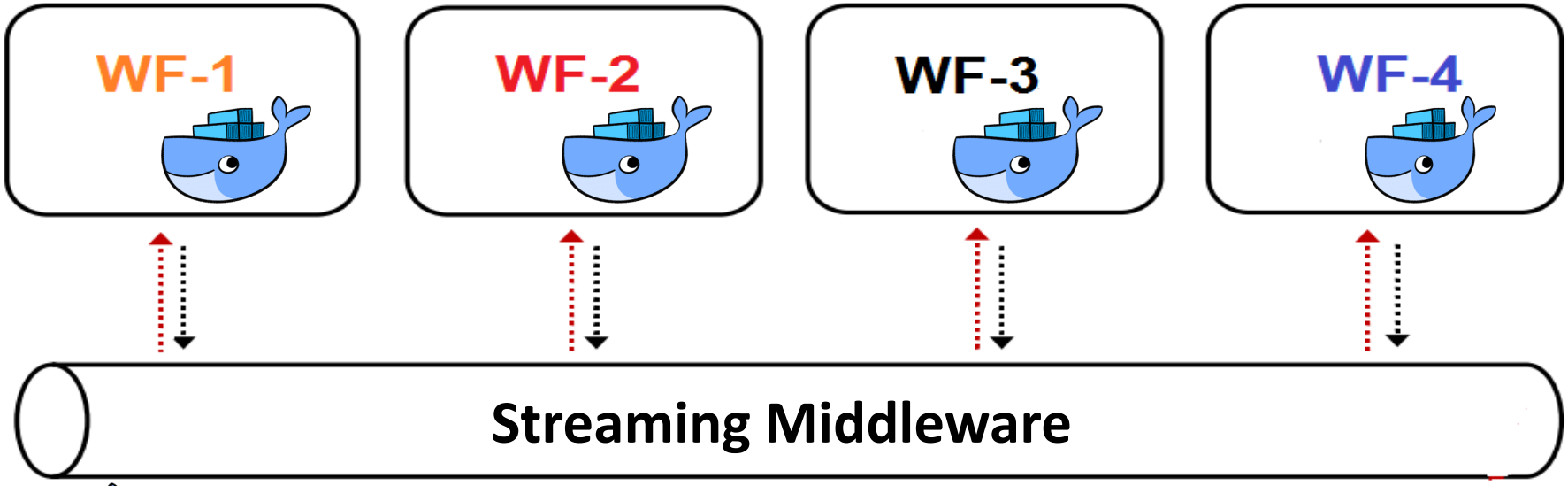
Figure 2: Conventional "Token Counting" GDD workflow: The data stream is split into non-overlapping windows of fixed size by counting data tokens using `Sequence_to_Array` with a statically defined data token count. Here, the maximum length of data arrays is calculated from constant parameters provided in the lower right of the workflow graph.

S. Köhler, S. Gulati, G. Cao, Q. Hart, and B. Ludäscher, "Sliding window calculations on streaming data using the Kepler scientific workflow system," Procedia Comput. Sci., vol. 9, pp. 1639–1646, 2012.

The rising of our **Micro-Workflow Idea**

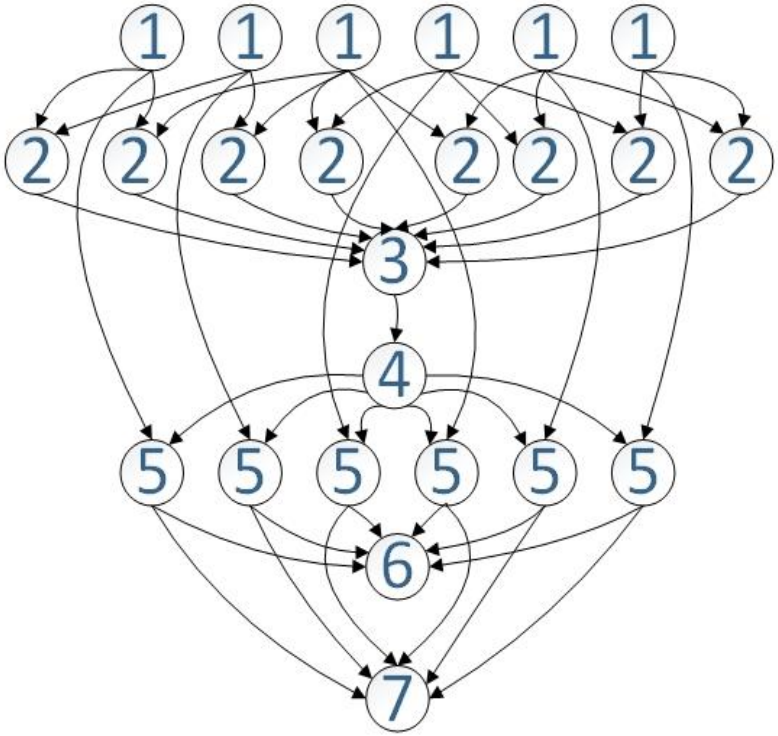


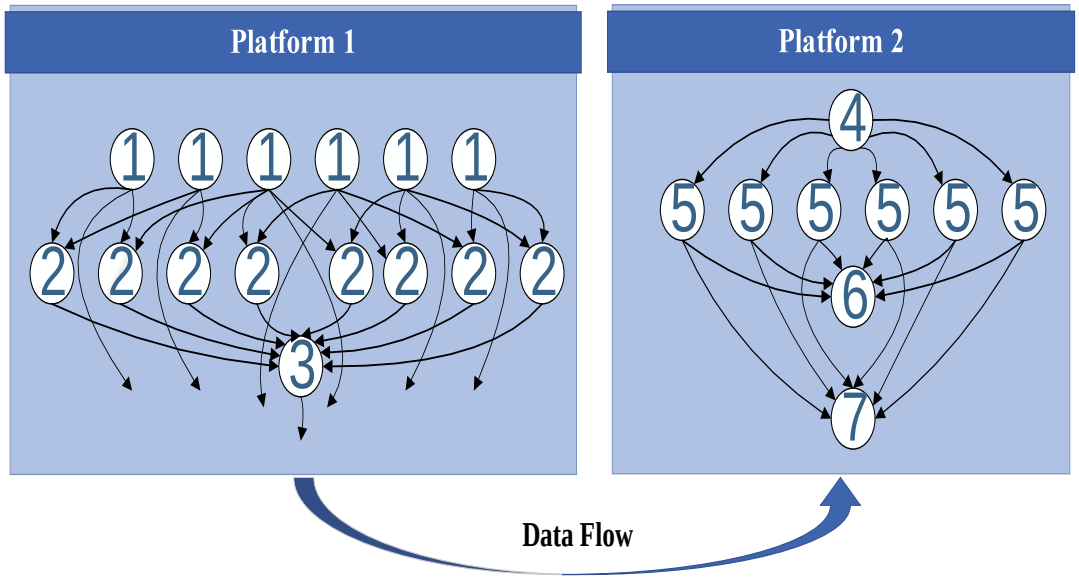
Micro-Workflow



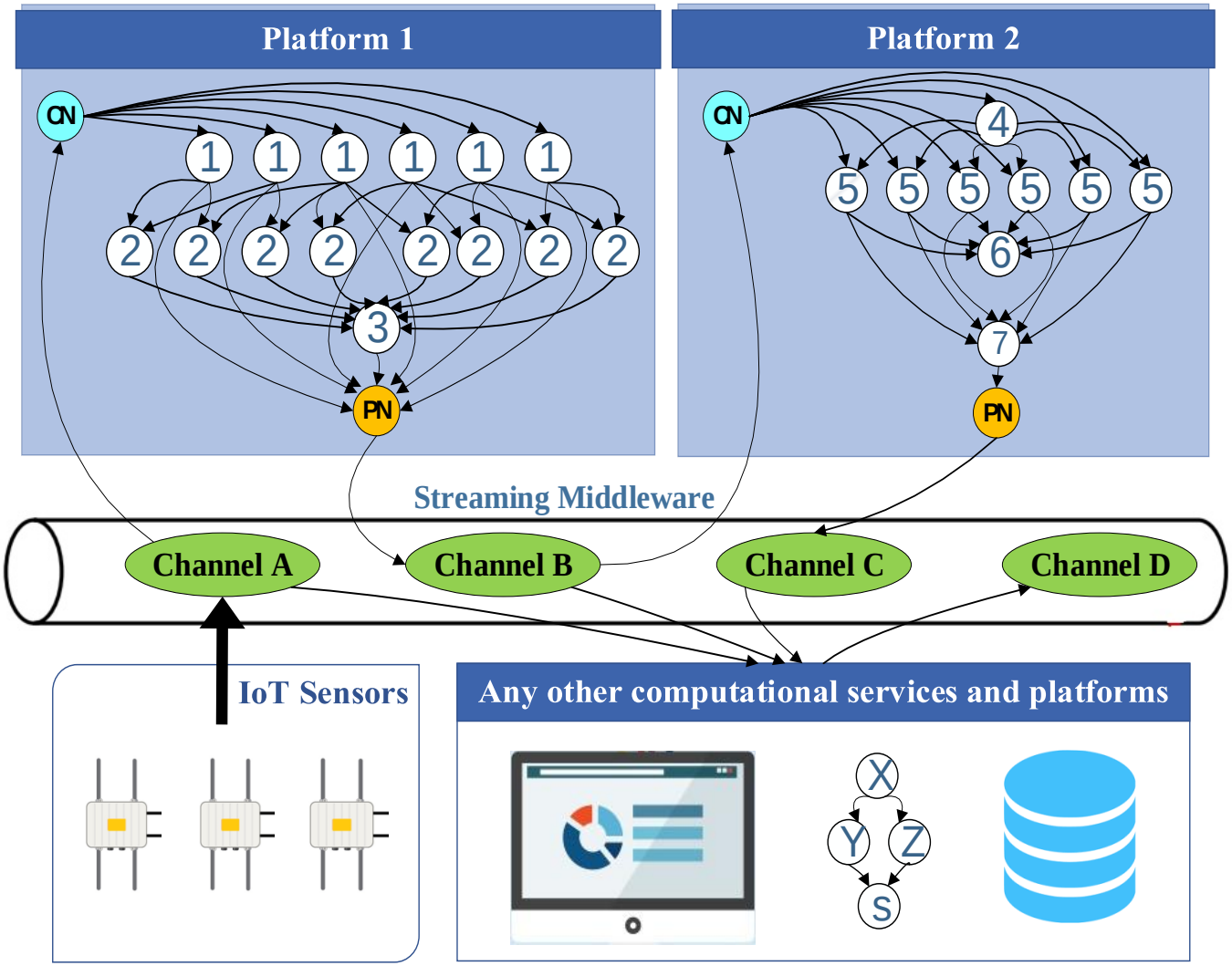
Docker is used for running software packages called "containers". All containers are run by a single OS kernel and are thus more lightweight than virtual machines. All container are isolated from each other and bundle their own tools, libraries and configuration files; they can communicate with each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

Micro-Workflow





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.

Av_SM (average interval between source messages
Av_RM (average interval between result messages)
Av_TAT (average turnaround time)

Table 1. Testing results

Test time	**1 hour**
Number of messages	**472279**
Av_SM (millisecond)	**7.62**
Av_RM (millisecond)	**7.62**
Av_TAT (millisecond)	**1.38**


Thank you