Personalized healthcare: workflow orchestration for cyberinfrastructure platforms

Rosa M Badia, Miguel Vázquez, Sergi Girona
Barcelona Supercomputing Center

1. What innovative capabilities/functionalities will the proposed candidate platform demonstrate (e.g. transcontinuum workflow, edge computing, data logistics, distributed reduction engine, etc.)?

This demonstrator will focus on the orchestration of workflows across the platform. The demonstrator will be based on COMPSs, which while it is a task-based programming model it also provides means for the development and orchestration of edge-to-cloud/HPC workflows. The COMPSs runtime has been recently modified to become a decentralized engine to fit the requirements of such infrastructures. For such scenario, COMPSs is combined with dataClay which offers an object-oriented library that is able to federate in-memory data from multiple stores.

These are the main innovations on the demonstrator:

- Orchestration of dynamic end-to-end workflows in an edge-to-cloud/HPC platform, that take into account the volatility of the edge devices. By dynamic, we mean that the actual workflow depends on the actual inputs of the workflow and also that can change at execution time due to the occurrence of specific events.
- Integration of machine learning, data analytics and computational processes in a single workflow.
- Multiple aspects of data management in such infrastructures: sharing, privacy, locality, security, management of devices' failures or disappearance, etc.

2. What applications/communities would/could be addressed?

For the BDEC demonstrator we want to propose a use case based on **personalized healthcare**, with wearable devices collecting continuous information on multiple parameters of the patient. The parameters will be processed in the edge and used to train AI models in the cloud/HPC to provide personalized notifications, alerts, and recommendations for prevention. To avoid personal data to leave the controlled environment of edge computing, data reaching the cloud will be encrypted or encoded.

However, since COMPSs is a general-purpose programming model, there is a wide range of applications where what we are proposing can be applied. We are contributing to multiple European funded projects involving fog-to-cloud scenarios where the COMPSs programming models is used. Find here some examples:

- mF2C: smart airport fog hub, enriched boat navigation service, emergency management in smart cities
- CLASS: smart city use, with heavy sensorized urban area and connected cars

- ELASTIC: smart tramway

3. What is the "platform vision," i.e. what kind of shared cyberinfrastructure (CI) for science would the further research/design/development of this platform lead to?

We are considering a hierarchical platform with devices in the edge (wearable devices) that are the sources of data, mobile devices in the edge connected to the wearable devices that can be used for prediction and cloud/HPC devices used to train the models.

4. How available/ready/complete is the set of software components to be used to build the demonstrator?

COMPSs runtime is already functional in fog-to-cloud scenarios. The use of HPC instead of cloud has not been tried, but should be almost immediate.

The actual personalized healthcare use case is at proposal level, therefore needs to be developed.

5. As far as one can tell at this early date, to what extent can this be done with existing and/or otherwise available hardware/software/human resources?

The wearable devices are not available right now, and some of the development resources are not allocated. The cloud/HPC devices could be provided by BSC.

6. What is the potential international footprint of the demonstrator?

As we have said before, the COMPSs runtime is already involved in multiple scenarios that involve several European countries. Given the nature of the use case that we are considering, the demonstrator has a potential for becoming very attractive and we expect contributors from worldwide level.