Contribution ID: **4** Type: **Presentation (15' + 5' for questions)**

An architecture for the building and management of computing continuum systems using serverless blocks

Monday, 28 October 2024 17:50 (20 minutes)

Recently, the computing continuum has emerged as an extension of existing cloud services. Existing cloud services enable users to access almost unlimited resources for the processing, storage, and sharing of data. Nevertheless, the increasing volume of data, as well as their constant production, have motivated the distribution of both data and computation through existing computational nodes. In the computing continuum, organizations organize similar computational nodes in the form of vertical layers. The more typical layers are the edge, the fog, and the cloud. The goal is to take advantage of the characteristics of the infrastructure in each layer, like the proximity of sensors with edge devices and the highest computational specifications of fog and cloud machines. Nevertheless, the construction of these systems is a complex task as it involves the coordination of multiple actors. From sensors and Internet-of-things (IoT) devices collecting the data to visualization services, passing through intelligence artificial and machine learning applications to process the data to produce information, as well as storage silos to preserve and share both raw data and information. The movement of data through these environments is another issue to consider as actors are distributed through different infrastructures. This produces delays in the delivery of data, as well as security and reliability concerns as the data is moved through uncontrolled networks. Here we introduce an architecture to manage the deployment and execution of computing continuum systems as well as the movement of data through the different infrastructures considered. Using a high-level self-similar scheme, organizations can create the abstract representation of a system by chaining different applications and choosing the infrastructure where they will be deployed. These applications are managed as serverless containers, which contain all the requirements to enable these applications to be loosely coupled and deployed on the selected infrastructure. Furthermore, these containers include an API Rest to enable their remote invocation. Data channels through the multiple applications and infrastructures in a system are created using abstractions called data containers, which orchestrate the movement of data through different sites, creating a content delivery network. To meet security and reliability requirements when moving data through data containers, we designed a configurable non-functional scheme that allows organizations to couple computing continuum systems with applications to fulfil non-functional requirements. Moreover, to mitigate bottlenecks in a system, we include an adaptive mechanism that includes a monitoring scheme and an autoscaling mechanism based on parallel patterns. First, the monitoring scheme measures the performance of the applications by collecting metrics such as their service times, average waiting time in queue of objects, the volume of data processed, and throughput of the applications. Next, the autoscaling mechanism identifies the bottlenecks in the system and creates a parallelism strategy to mitigate them. The goal is to produce a steady data flow through the system without modifying the code of applications and without knowing the characteristics of the input workload and the hardware where applications are deployed.

Primary authors: SANCHEZ GALLEGOS, Dante Domizzi (Universidad Carlos III de Madrid); Ms CARRIZALES-ESPINOZA, Diana (Cinvestav Tamaulipas); Dr CARRETERO, Jesus (Universidad Carlos III de Madrid); Dr GONZALEZ-COMPEAN, Jose Luis (Cinvestav Tamaulipas)

Presenters: SANCHEZ GALLEGOS, Dante Domizzi (Universidad Carlos III de Madrid); Dr GONZALEZ-COMPEAN, Jose Luis (Cinvestav Tamaulipas)

Session Classification: IBERGRID

Track Classification: Developments oriented to foster the Compute Continuum