

Proposta de provisionamento de recursos para plataformas de cloud continuum baseado no framework Globus Compute

Mateus Tenorio dos Santos¹, Guilherme Galante¹

¹PPGComp – Universidade Estadual do Oeste do Paraná - Unioeste
R. Universitária, 2069 - Bloco B - Universitário, Cascavel - PR, 85819-110

{mateus.santos28, guilherme.galante}@unioeste.br

Resumo. *Este trabalho tem como objetivo investigar e implementar uma arquitetura híbrida de computação utilizando Kubernetes e a plataforma Globus Compute, visando otimizar a execução de funções distribuídas em ambientes de nuvem e borda (edge). A pesquisa aborda o desafio da integração entre sistemas de computação em nuvem e na borda, considerando as limitações de latência, escalabilidade e eficiência energética. A pesquisa pretende preencher a lacuna que existe da viabilização da utilização de plataformas FaaS, como o Globus Compute, em arquiteturas hierárquicas, para o Cloud Continuum.*

1. Informações gerais

A evolução da Internet das Coisas (IoT) tem sido um dos marcos mais transformadores da era digital, caracterizada pela interconexão de bilhões de dispositivos que interagem entre si e com os ambientes onde estão inseridos. Essa rede globalizada de sensores, atuadores e sistemas computacionais gerou um volume massivo de dados que desafia os paradigmas convencionais de armazenamento e processamento.

Nesse contexto, o conceito de *IoT Edge-Cloud Continuum* surge como uma solução inovadora para superar as limitações de abordagens isoladas, como computação de borda ou nuvem. Essa arquitetura integra diferentes camadas de infraestrutura computacional, desde dispositivos na borda, que processam dados próximos à sua origem, até sistemas de nuvem que oferecem capacidade de processamento e armazenamento praticamente ilimitadas. Essa continuidade entre borda, nevoeiro (*fog computing*) e nuvem minimiza a latência, otimiza o consumo de banda e oferece escalabilidade para diversas aplicações, incluindo cidades inteligentes, agricultura de precisão, transporte autônomo e saúde conectada.

A integração de plataformas avançadas de computação distribuída, como o Globus Compute (<https://docs.globus.org/compute/>), uma plataforma distribuída de FaaS que permite execução remota de funções, ao IoT Edge-Cloud Continuum representa uma oportunidade única de avançar o estado da arte em arquiteturas computacionais. Além de preencher lacunas na literatura, a pesquisa possui relevância prática, abordando demandas crescentes por soluções escaláveis em setores críticos. A capacidade de criar clusters hierárquicos interoperáveis é especialmente relevante em aplicações de alta complexidade, onde a eficiência e a resiliência do sistema são cruciais.

Desta forma, este trabalho exploratório será desenvolvido utilizando do Globus Compute juntamente com outras tecnologias, principalmente *Kubernetes* (<https://kubernetes.io/>), para controlar os clusters. Inicialmente, serão realizados testes

exploratórios das capacidades e limitações de uma arquitetura mais simples, onde apenas um cluster é montado utilizando o *Kubernetes*, e no decorrer deste trabalho uma arquitetura mais robusta será implementada. Deste trabalho, espera-se que seja possível alcançar um modelo de arquitetura de referência para implementação de soluções de alto desempenho, escaláveis e resilientes, facilitando a adoção em aplicações práticas e industriais. Espera-se também, aumentar o leque de possibilidades da utilização do Globus Compute, tecnologia que atualmente é utilizada extensamente para aplicações científicas.

2. Arquitetura Conceitual

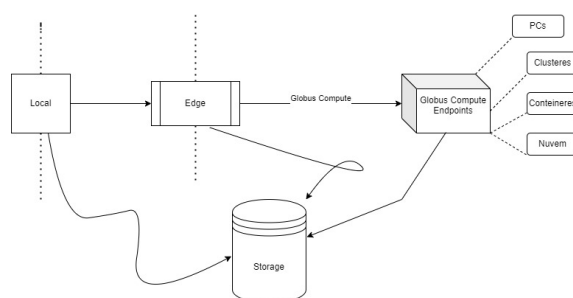


Figura 1. Arquitetura conceitual

A arquitetura conceitual será composta por 3 partes. Como temos como objetivo explorar o provisionamento de recursos no Cloud-continuum, cada parte dessa arquitetura será capaz de realizar algum processamento. Na camada local, a mais simples, será composta por nós que coletarão dados. São nós mais simples em nível de processamento, mas que são mais eficazes em questão energética e possuem baixa latência para qualquer processamento.

A camada de borda ficará responsável por qualquer processamento que possa ser mais complexo que os realizados na camada dos nós de captura de dados, mas também é responsável por processar qualquer requisição ou chamada aos *endpoints* da camada em nuvem. Por fim, temos a camada de nuvem, que é o ponto onde serão alocados todos os *endpoints* do *Globus Compute*. Como principal vantagem do *Globus Compute*, nela, os componentes podem ser heterogêneos [Li et al. 2022]. Portanto, podemos ter endpoints que são computadores, endpoints que são clusters de computadores, ou clusters em nuvem, contêineres gerenciados por *kubernetes* ou até mesmo instâncias em algum outro provedor de nuvem. Isso tem suma importância, pois irá garantir a escalabilidade necessária para essa solução.

Essa arquitetura necessita de uma solução de armazenamento, para que seja possível obter o conhecimento dos nós de *endpoint*, bem como armazenar funções e *inputs* ou *outputs* que extrapolem o limite imposto pela utilização do *Globus compute*. Essa solução de armazenamento pode ser um servidor dedicado ou algum serviço na nuvem.

Referências

Li, Z., Chard, R., Babuji, Y., Galewsky, B., Skluzacek, T. J., Nagaitsev, K., Woodard, A., Blaiszik, B., Bryan, J., Katz, D. S., Foster, I., and Chard, K. (2022). funcx: Federated function as a service for science. *IEEE Transactions on Parallel and Distributed Systems*, 33(12):4948–4963.