



**Título:** Alocação preditiva de recursos em nuvem utilizando redes neurais de memória de curto longo prazo empilhadas

**Data:** 29/11/2022

**Horário:** 14h00

**Local:** Videoconferência

**Resumo:**

A computação em nuvem tem sido apontada como uma solução para o uso racional de

recursos da Tecnologia da Informação. Os provedores de serviços em nuvem oferecem ambientes compartilhados que podem ser dimensionados para atender aos requisitos flutuantes de seus clientes. O desafio imposto é a utilização de mecanismos capazes de otimizar o uso dos recursos e, simultaneamente, garantir que o desempenho desses serviços continue atendendo às métricas Quality of Experience (QoE), Quality of Service (QoS), bem como aos Service Level Indicators (SLI) e aos respectivos Service Level Agreements (SLA) estabelecidos. Os provedores necessitam oferecer mecanismos autônomicos para promover a escalabilidade dos recursos em tempo hábil, ao mesmo tempo, em que os clientes precisam confiar no desempenho e nos custos envolvidos nas negociações. No processo de Gerenciamento de Recursos em Nuvem, diversas abordagens preditivas de escalonamento de recursos já foram propostas para superar as limitações das abordagens reativas convencionais. Entretanto, tais abordagens ainda não demonstraram resultados satisfatórios, em termos de custo, desempenho e autonomia. Esta pesquisa propõe um Modelo Autônomico de Gerenciamento de Recursos em Nuvem que combina características reativas e preditivas para o escalonamento de recursos. Para suportar o provisionamento preditivo de recursos, foi utilizado Recurrent Neural Networks (RNNs) na arquitetura Stacked Long Short-Term Memory, buscando suplantar os resultados já alcançados. Na abordagem autônômica, adotou-se o modelo MAPE-K, fazendo uso dos princípios da Autonomic Cloud Computing (ACC). Visando demonstrar a viabilidade da proposta foi elaborado um estudo de caso utilizando traces experimentais. Na avaliação da acurácia do modelo utilizou-se um comparativo entre a rede LSTM clássica e diferentes configurações da rede Stacked LSTM. Para análise da viabilidade operacional foi implementado um protótipo utilizando componentes de simuladores de infraestrutura em nuvem. Os resultados obtidos demonstraram a viabilidade da proposta, trazendo como benefício a utilização das Stacked LSTMs na predição do provisionamento de recursos em nuvem. Como trabalhos futuros, pretende-se evoluir o protótipo para uma ferramenta operacional, em código livre, para apoiar provedores de serviços de pequeno e médio porte e permitir o planejamento de capacidade no processo de migração para nuvem.

### **Banca examinadora:**

- Prof. Dr. José Neuman de Souza (MDCC/UFC - Orientador)
- Prof. Dr. Alberto Sampaio Lima (UFC)
- Prof. Dr. Lincoln Souza Rocha (UFC)
- Prof. Dr. José Maria da Silva Monteiro Filho (UFC)
- Prof. Dr. Joaquim Celestino Junior (UECE)