CiTIUS

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA — **CAMPUS VIDA** CAMPUS DE EXCELENCIA INTERNACIONAL

PhD Defense: 'Big Data and Large-scale Analytics: Efficiency of Sustainable Scalability and Security of Centralized Clouds and Edge Deployment Architectures'

Data: venres, 28 febreiro, 2020 - 11:00 - 13:00

Lugar: CiTIUS Assembly Hall

Poñente(s): Feras Mahmoud Awaysheh (CiTIUS predoctoral researcher)

Idioma: Inglés

Streaming: Por confirmar

One of the significant shift of next-generation computing technologies will certainly be in the development of Big Data (BD) platforms. This trend results in a wide range of revolutionary and state-of-the-art enhancements within the data science. Apache Hadoop, the BD landmark, evolved as a widely deployed large-scale data operating system. In principle, Hadoop is designed to utilize clusters of commodity machines. Over the years, the model was ported to be compatible with different types of architecture and paradigms. Its new features include federation configuration to provide Hadoop with the maturity to serve different markets.

This dissertation focuses on the architectural elements of the BD processing frameworks and focusing on the scalability aspects to improve its performance and security. We propose a hybrid BD execution environment (called EME) that keeps spawning a containerized DataN- odes for the benefit of BD applications. A dynamic provisioning scheduler (called OPERA) is presented to take advantage of the underutilized on-premise and cloud resources. The results demonstrate that OPERA has immense potential as it significantly decreases the time of exe- cution up to 74% for CPU bound jobs (as in the PiEst benchmark), and up to 26% for HDFS bound jobs (as in the wordcount benchmark) compared with a native Hadoop cluster.

Also, the privacy and security of BD architectures are discussed as a significant concern among prac- titioners. A BD federation single-sign-on authentication module and a novel access broker framework are introduced. Experimental results demonstrate the efficiency of the proposed access broker with only 1% impact on the Hadoop performance compared with a non-secure one. Finally, a modern secure case of study regarding data streaming of edge nodes to the clouds in vehicular clouds is explained to validate the thesis findings.

Supervisors: José Carlos Cabaleiro Domínguez and Tomás Fernández Pena