

A WORKFLOW MANAGEMENT SYSTEM FOR SCALABLE DATA MINING ON CLOUDS

Abstract :

The extraction of useful information from data is often a complex process that can be conveniently modeled as a data analysis workflow. When very large data sets must be analyzed and/or complex data mining algorithms must be executed, data analysis workflows may take very long times to complete their execution. Therefore, efficient systems are required for the scalable execution of data analysis workflows, by exploiting the computing services of the Cloud platforms where data is increasingly being stored. The objective of the paper is to demonstrate how Cloud software technologies can be integrated to implement an effective environment for designing and executing scalable data analysis workflows. We describe the design and implementation of the Data Mining Cloud Framework (DMCF), a data analysis system that integrates a visual workflow language and a parallel runtime with the Software-as-a-Service (SaaS) model. DMCF was designed taking into account the needs of real data mining applications, with the goal of simplifying the development of data mining applications compared to generic workflow management systems that are not specifically designed for this domain.

SHIELD TECHNOLOGIES,

2232, 3RD FLOOR, 16TH B CROSS, YELAHANKA NEW TOWN, BANGALORE-64

Mail us: shieldtechnobl@gmail.com / manager@shieldtechno.com

Contact: 9972364704 / 8073744810