

**CHENNAI – PONDICHERRY**

**Supporting Data-intensive Workflows in Software-defined Federated Multi-Clouds**

**Abstract:**

Cloud computing is emerging as a viable platform for scientific exploration. Elastic and on-demand access to resources (and other services), the abstraction of “unlimited” resources, and attractive pricing models provide incentives for scientists to move their workflows into clouds. Generalizing these concepts beyond a single virtualized datacenter, it is possible to create federated marketplaces where different types of resources (e.g., clouds, HPC grids, supercomputers) that may be geographically distributed, are collectively exposed as a single elastic infrastructure. This presents opportunities for optimizing the execution of application workflows with heterogeneous and dynamic requirements, and tackling larger scale problems. In this paper, we introduce a framework to manage the end-to-end execution of data-intensive application workflows in dynamic soft ware-defined resource federation. This framework enables the autonomic execution of workflows by elastically provisioning an appropriate set of resources that meet application requirements, and by adapting this set of resources at runtime as the requirements change. It also allows users to customize scheduling policies that drive the way resources federated and used. To demonstrate the benefits of our approach, we study the execution of two different data-intensive scientific workflows in a multi-cloud federation using different policies and objective functions.

**Existing System:**

However, enabling users to efficiently use these marketplaces presents a new set of challenges. For example, these marketplaces provide a wide variety of capabilities and capacities as well as pricing models, which makes provisioning the appropriate blend of resources/services to meet application and user requirements, non-trivial. Moreover, application and resource dynamics requires runtimes adaptations of the provisioned resources as well as the application execution.

**Proposed System:**

We present an end-to-end framework that enables the autonomic execution of data-intensive scientific workflows on software-defined multi-cloud environments. This framework is built on top of our federation model, which enables the dynamic creation of federated “Cloud-of-Clouds”. The resulting solution is a platform that takes an infrastructure-independent workflow description from the user, and autonomously orchestrates the execution of such a workflow by elastically composing appropriate resources and services to ensure that appli cation requirements and user’s objectives. Furthermore, runtime monitoring, failure handling, and adaptations enable the system to respond to application and resource dynamics in an appropriate and timely manner.

A customizable autonomic manager that supports multiple scheduling policies and allows the definition of new ones. These policies can be defined based on different optimization criteria, such as cost, performance, etc. These policies are used to identify relevant resources and evaluate possible scheduling strategies, and to choose the one that best meets the objectives.