

[ABSTRACT]

The present invention provides a method for provisioning and managing a multi-cluster on a cloud platform. When provisioning of a plurality of clusters capable of operating a container-based application in a multi-cloud environment is requested, the method allows a cloud platform system to enable following the steps of: inputting type information of each of the clusters; generating configuration information of each of the clusters when the type information of the cluster is input; confirming whether to generate or modify each of the clusters when public cloud or system access information is registered; performing remote simultaneous provisioning of the cluster by requesting for generation of instance, network, and storage and configuring same, installing container runtime software, and setting configuration information of the cluster when the cluster is generated; modifying the cluster from a remote place by confirming configuration history information of the cluster and updating the configuration of the cluster when the cluster is modified; adding a node of the cluster or replacing a failed node of the cluster, and performing backup of the node of the cluster; and automatically scaling the cluster, wherein the configuration information of the cluster includes at least one of the number of instances, an instance specification (GPU, a memory type), network configuration information, and storage configuration information.