

# WORKLOAD PREDICTION METHODS AND APPARATUSES FOR SERVICE IN SERVICE CLUSTER

## ABSTRACT

Embodiments of this specification provide a workload prediction method and apparatus for a service in a service cluster. The method includes: obtaining a load indicator sequence of each service in a service cluster corresponding to a workload indicator in a same historical time period; determining, based on the load indicator sequence corresponding to each service, a service representation corresponding to each service; performing clustering processing based on the service representation corresponding to each service, to obtain a target category cluster to which each service belongs in multiple category clusters; obtaining multiple sequence prediction models pre-trained for multiple tasks, and enabling the multiple category clusters to correspond to the multiple tasks; and inputting at least a load indicator sequence of any service into a target sequence prediction model corresponding to a target category cluster of the service in the multiple sequence prediction models.

15

FIG. 1