***Abstract*—**Resource management in cloud infrastructures is one of the most challenging problems due to the heterogeneity of resources, variability of the workload and scale of data centers. Efficient management of physical and virtual resources can be achieved considering performance requirements of hosted applications and infrastructure costs. In this paper, we present a self-adaptive resource management system based on a hierarchical multi-agent based architecture. The system uses novel adaptive utilization threshold mechanism and benefits from reinforcement learning technique to dynamically adjust CPU and memory thresholds for each Physical Machine (PM). It periodically runs a Virtual Machine (VM) placement optimization algorithm to keep the total resource utilization of each PM within given thresholds for improving Service Level Agreement (SLA) compliance. Moreover, the algorithm consolidates VMs into the minimum number of active PMs in order to reduce the energy consumption. Experimental results on real workload traces show that our recourse management system provides substantial improvement over other approaches in terms of performance requirements, energy consumption and the number of VM migrations.