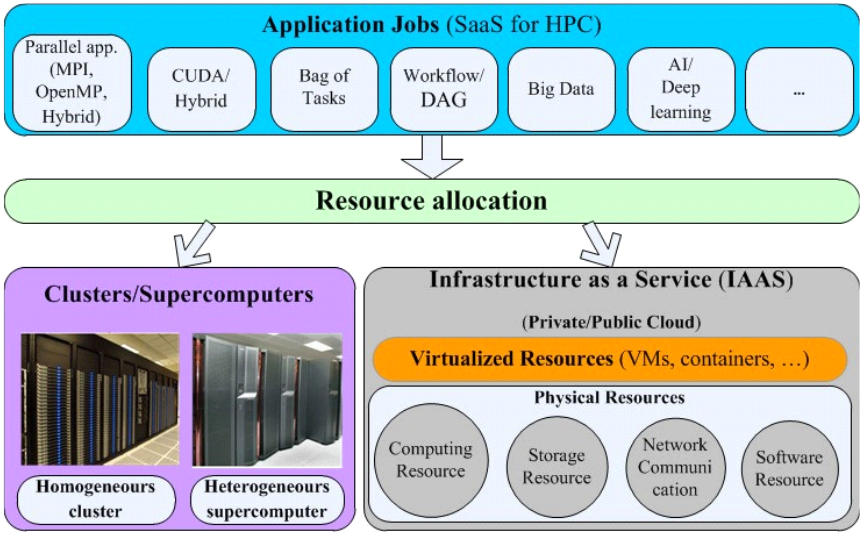
# Prediction of Job Characteristics for Intelligent Resource Allocation in HPC Systems: A Survey and Future Directions

**Zhengxiong HOU, Hong SHEN, Xingshe ZHOU, Jianhua GU, Yunlan WANG, Tianhai ZHAO**

Frontiers of Computer Science, DOI: 10.1007/s11704-022-0625-8

# Problems & Ideas

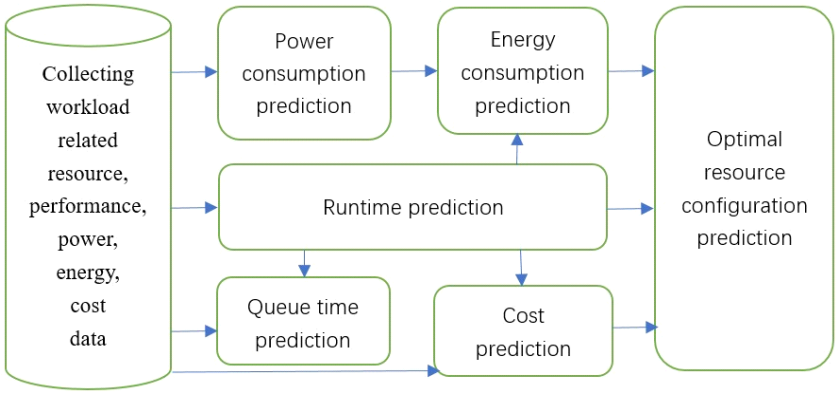
- Problems of conventional resource allocation in HPC systems:
  - For executing applications, both HPC end-users and cloud users need to request specific resources for different workloads by themselves. It is hard for them to estimate runtime and select optimal resource configurations in terms of performance, cost, and energy efficiency
  - how to provide on-demand HPC services with intelligent resource allocation



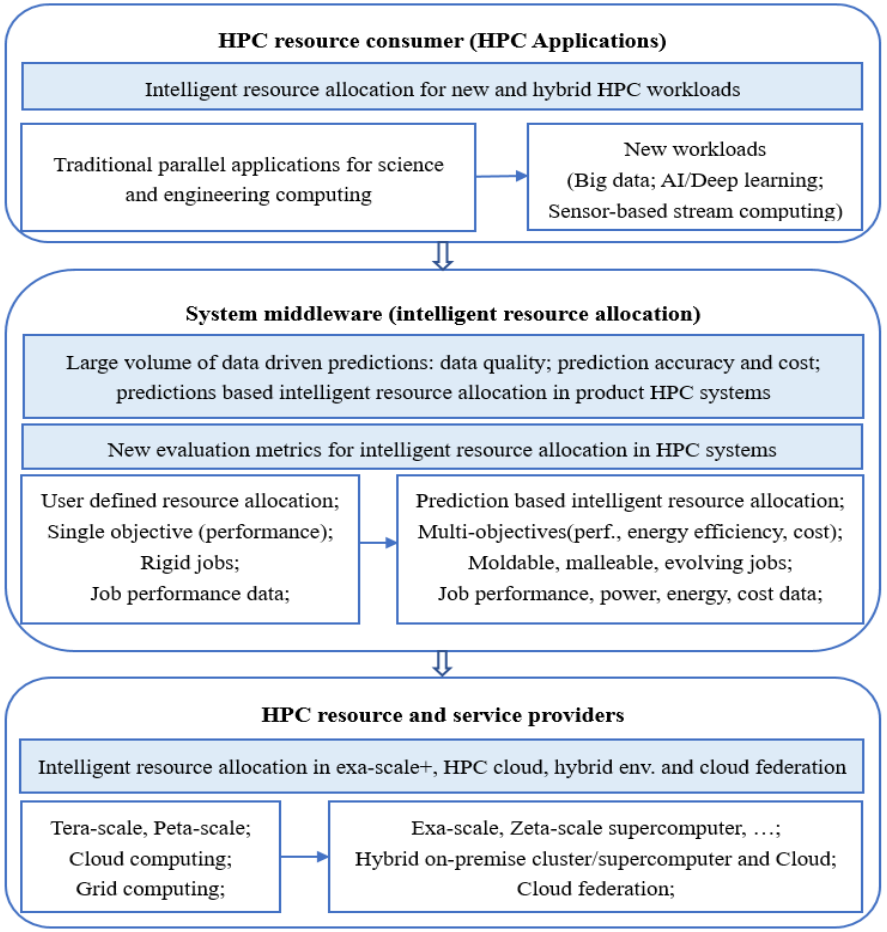

- Ideas: Prediction of job characteristics plays a key role for intelligent resource allocation. We present a survey of the existing work and future directions for prediction of job characteristics for intelligent resource allocation in HPC systems

# Main Contributions

- Contributions:
  - Review the existing techniques in obtaining fine-grain performance and energy consumption data of jobs;
  - We survey the techniques for single-objective oriented predictions on runtime, queue time, power and energy consumption, cost and optimal resource configuration for input jobs, as well as multi-objective oriented predictions;
  - We discuss future trends, research challenges and possible solutions towards intelligent resource allocation in HPC systems.




Prediction of job characteristics for intelligent resource allocation in HPC systems




Future research directions