

Dynamic Load Balancing

[for task-based parallel applications in distributed memory systems]

proactive load balancing
reactive load balancing
work stealing

Extending the performance model

- (1) Task graph with detailed weight of dependencies
- (2) Task-based runtime with heterogenous architectures and data movement overhead
- (3) Efficiency bound of load balancing with probability estimation

Extending the proactive load balancing approach

- (4) Generative model for online load prediction
- (5) Re-feeding or re-training prediction models at runtime
- (6) Proactive task offloading + topology-aware information
- (7) Portability protocol for co-scheduling tasks across multiple applications