Motivation

scaling in parallel computing

fixed-workload (Amdahl’s law) limited by resources

fixed-time (Gustafson’s law)

pay-per-use charging

application resource demand scale differently

On Cloud, application scalability is limited only by the cost budget!

Approach

measurement-driven analytical modeling approach

cloud resources have different cost and performance e.g. Amazon EC2

Given an application, time deadline and a cost budget
- largest size of the application executable?
- execution cost and time?
- cloud configuration?

Observation 1: Multiple Pareto-optimal problem sizes.

Observation 2: Increasing the cost budget and relaxing the time deadline does not always result in obtaining a larger problem size.

Observation 3: Resource demand is allocated in order of highest PCR.

Observation 4: Among Pareto-optimal problem sizes, tightening the time deadline results in relatively smaller reduction of problem size.

Observation 5: Among Pareto-optimal problem sizes, increasing cost results in relatively smaller increase of problem size because PCR is non-linear across resource types.

Evaluation

- Applications
  - n-body (linear, quadratic)
  - sand (linear, logarithmic)
- Cloud Resources
  - Amazon EC2 cloud
  - 9 resource types