Properties of Distributed
High Throughput Computing
Applications

Mats Rynge <rynge@isi.edu>

OSG User Support

Why High Throughput Computing?

For many experimental scientists, scientific progress and quality of research are strongly linked to computing throughput. In other words, they are less concerned about instantaneous computing power. Instead, what matters to them is the amount of computing they can harness over a month or a year - they measure computing power in units of scenarios per day, wind patterns per week, instructions sets per month, or crystal configurations per year.

Slide credit: Miron Livny

High Throughput Computing

Sustained computing over long periods of time. Usually serial codes, or low number of cores threaded/MPI.

High Performance Computing

Great performance over relative short periods of time. Large scale MPI.

High Throughput Computing
is a
24 - 7 -365
activity

FLOPY ≠ (60*60*24*7*52)*FLOPS

•Slide credit: Miron Livny

Distributed?

No shared filesystem

(except for the OASIS readonly filesystem used for software distribution)

How to access your application

- You can ship it with your jobs…
  - if your application can be built self-contained, this solution provides easy software updates and management by the user
  - for example: static linking, python scripts

- … or have it installed on OASIS
  - currently, only admins can install software

- More about this in later sessions

Application Ideal Case

- ## Single threaded
  - Limited support for multithreaded and small core count MPI

- ## Use less than 2 GB RAM
  - Limited support for larger memory requirements

- ## Have a runtime of 4 – 12 hours
  - Short jobs leads to inefficient scheduling
  - Long jobs – preemption can be a problem

Preemption

Preemption is when your job gets killed on the remote compute resource because something with a higher priority came in

- Remember, the OSG resources are owned and operated by someone else
- These resources prefer jobs from the owner, and provides left-over cycles to you

Your application has to support being restarted

Support?

connect-support@opensciencegrid.org