Batch Scheduling and Resource Management

Luke Tierney

Department of Statistics & Actuarial Science
University of Iowa

October 18, 2007

Background

- The earliest computers could run only one program at a time.
- Programs and data were written on punched cards.
- Decks of cards were submitted in batches.
- These batches were placed in a queue and run one at a time.
- Later, time-sharing allowed multiple simultaneous interactive users.
- Batch processing, or batch queueing, is still useful:
  - Two programs running simultaneously can slow each other down.
  - On a single core the slowdown will be at least a factor of two.
  - It can be *much* more with heavy memory or I/O use.
  - Programs may fail due to insufficient memory.

Batch System Features

- Basic batch systems provide:
  - A means of placing jobs on a queue.
  - Some means of examining the queue.
  - A way to remove jobs.
  - Jobs are run when resources are available.
- Additional features provided by some systems:
  - Load balancing across multiple processors
  - Management of resource usage (memory in particular).
  - Protection against runaway jobs.
  - A priority system.
  - Ways of managing parallel jobs.
- More sophisticated features:
  - Checkpointing, suspending, resuming, moving running programs.
  - Integration with grid computing frameworks.