Resource Management in the Greenplum Parallel Database

Siva Narayanan (siva@greenplum.com)
Consultant Software Engineer, Query Processing, EMC Greenplum
Why, you ask?
Problem

- Finite resources - CPU/memory/IO/network
- Concurrent activity
  - Different business value (Loads/Reports/Analytics)
  - Different system impact (Simple/Complex queries)
- How can a DBA manage the system and keep everyone happy?
Solution

- Determine business value of a query upon arrival
  - Translate that to fair share of CPU and Memory
- Resource reservation / Admission control
  - Are the resources available?
- Run-time resource allocation
  - Ensure that reservations are honored
  - Adjust behavior as necessary
CPU vs Memory
CPU Sharing

- Every query operator in an execution plan
  - Continually measures its actual CPU usage and compares it with fair share
  - If it uses too much, it sleeps for a short while
  - Rinse, repeat
- I/O and network bandwidths are similar
Every query operator in a execution plan
  - Gets a portion of memory reserved for the entire query
  - Memory intensive operators vs not
  - Re-use memory between blocking operators
  - If data is too large, they spill
Net effect, every query uses up to its fair share
Resource management is a big problem with big data
Align resource allocation with business value
Greenplum Parallel Database has mechanisms for CPU and Memory
Questions?

We’re hiring!
siva@greenplum.com