Condor Implementation Strategies
Preston Smith <psmith@purdue.edu>

Rosen Center
for Advanced
Computing

- # Condor Grid a nexus for CI at Purdue
  - Vehicle for federating resources across campus and state
  - Resource also available to regional (NWICG) and national grids (OSG and TeraGrid)
- RCAC Clusters are primarily scheduled with PBS
  - Condor runs on same clusters, configured to run when PBS isn't using a node
  - ~4000 cores in PBS clusters available to Condor
- Some dedicated Condor-only resources
  - Sizeable chunk of CMS Tier-2
  - 48-node GPU rendering cluster in Envision Center
    - http://teradre.rcac.purdue.edu/

Rosen Center
for Advanced
Computing
RCAC

- # Condor on a PBS Cluster?

  - A node is in the OWNER state when PBS is using it

  - How does it work?

    - START = $(START) && $(PBSRunning) == FALSE

    - PBS Prologue

      - condor_config_val -rset -startd PBSRunning = TRUE

      - If jobs are running, run condor_vacate

    - PBS Epilogue

      - condor_config_val -rset -startd PBSRunning = FALSE

Rosen Center
for Advanced
Computing

- RCAC runs central manager services
  - 3 pools for clusters, 1 pool for campuswide participants.
- Sister unit in ITaP runs instructional computing labs for students
  - ~2500 machines - available for Condor 90% of the time!
- Other departments
  - Libraries (catalog terminals)
  - Physics (LSST, VERITAS)
  - Structural Biology (**major** users)
- Now other campuses!
  - Notre Dame, Purdue Calumet today
  - Soon: IU-Purdue Fort Wayne, Indiana State

Rosen Center
for Advanced
Computing

- # Science highlights:
  - – "Database of Hypothetical Zeolite Structures"
    - • TeraGrid project, used 3M hours in last year
  - – Cryo-EM image processing (structural biology)
    - • Campus user - consumed 2.76M hours over last year.
    - • Now capable of using all idle Windows cycles.. Look out..
  - – Protein folding
    - • On campus - 1.9M hours
  - – "Football pool problem"
    - • TeraGrid project, 280K hours used in one month
  - – Plus many others
    - • Distributed rendering, astrophysics, nanotechnology, hydrology, network simulation, cytometry, management