# SRT's Areas of Research & Past RAMS Project

*Systems Research Team*
Computer Science and Mathematics Division
Oak Ridge National Laboratory
http://www.csm.ornl.gov/srt

# ORNL Systems Research Team (SRT)

- Christian Engelmann
- Thomas Naughton
- Hong Ong
- Stephen Scott
- Anand Tikotekar
- Geoffroy Vallée

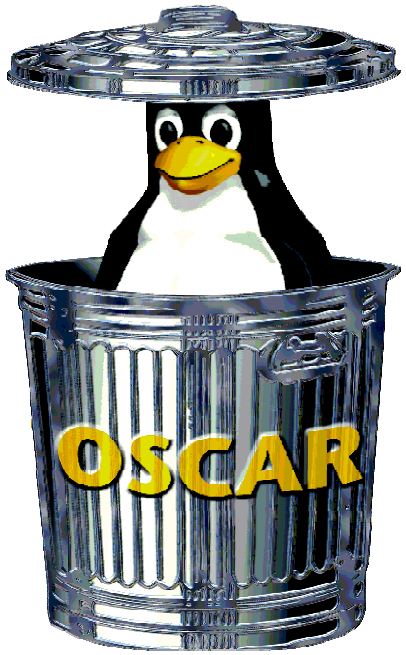
OAK RIDGE National Laboratory

# Outline

- **System Management Tools**
- System-level Virtualization
- Fault Tolerance / Resilience
- RAMS'08 Student Project

# Open Source Cluster Application Resources

- Snapshot of best known methods for building, programming, and using clusters
- International consortium of academic, research, and industry members




INRIA

OAK RIDGE National Laboratory

Bald Guy SOFTWARE

KERLABS

intel

LOUISIANA TECH UNIVERSITY

pervasivetechnologylabs AT INDIANA UNIVERSITY

NEC

OAK RIDGE National Laboratory

# OSCAR: C3 power tools

- Command-line interface for cluster system administration and parallel-user tools
- Parallel execution ***cexec***
  - Execute across a single cluster or multiple clusters at same time
- Scatter/gather operations ***cpush*** / ***cget***
  - Distribute or fetch files for all node(s)/cluster(s)
- Used throughout OSCAR
  - Mechanism for cluster-wide operations

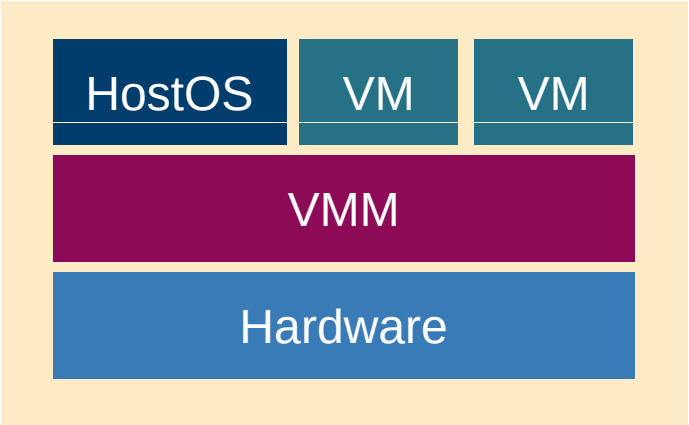
OAK RIDGE National Laboratory

# Outline

- System Management Tools
- **System-level Virtualization**
- Fault Tolerance / Resilience
- RAMS'08 Student Project

# System-level virtualization

- What is virtualization?
  - The capability of isolating an operating system from the bare hardware
- Many virtualization solutions:
  - Xen (Citrix Systems),
  - VMware,
  - VirtualBox (Sun Microsystems)
  - Etc.
- None of those solutions are “designed” for high-performance computing
  - Collaboration with SNL, Northwestern and UNM to develop new tools and virtualization solutions



# Why virtualization in HPC?

- Improved utilization
  - Users with differing OS requirements can be easily satisfied,
    - e.g., Linux, Catamount, others in future
  - Enable early access to petascale software environment on existing smaller systems
- Improved manageability
  - OS upgrades can be staged across VMs and thus minimize downtime
  - OS/RTE can be reconfigured and deployed on demand
- Improved reliability
  - Application-level software failures can be isolated to the VMs in which they occur
- Improved workload isolation, consolidation, and migration
  - Seamless transition between application development and deployment using petascale software environment on development systems
  - Proactive fault tolerance transparent to OS, runtime, and application

# Virtualization for HPC - Our Contributions

- Concept of “Virtual System Environment (VSE)”
  - Powerful abstraction concept that encapsulates OS, application runtime, & application
- OSCAR-V
  - Implements the VSE concept
  - Enhance the support of virtual clusters
  - Abstract the different virtualization solutions
- Development of a “module” mechanism for Xen
  - Similar to Linux modules
  - Enable the extension of Xen
  - “*Get the capabilities that perfectly fit your needs*”

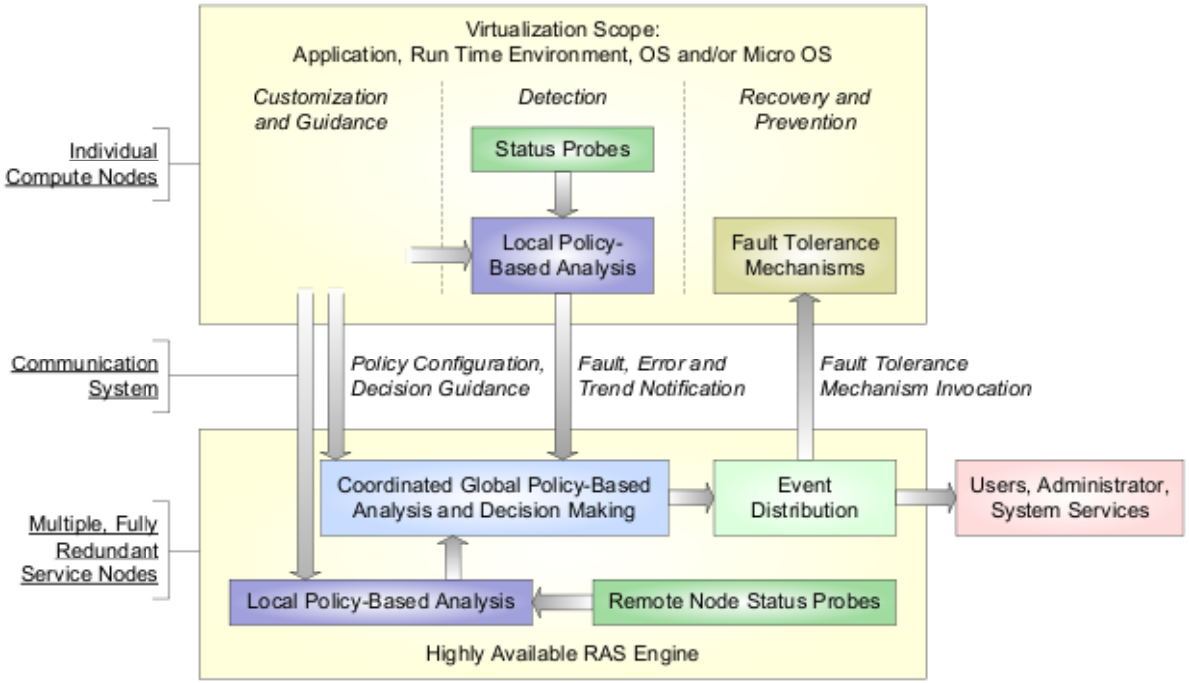
# Outline

- System Management Tools
- System-level Virtualization
- **Fault Tolerance / Resilience**
- RAMS'08 Student Project

# System and application resilience

- Goal: "*Application should continue to run even in case of failures*"

- Modular framework
  - Support virtualization
  - Designed to support process-level checkpoint/restart and migration
  - Proactive fault tolerance adaptation: Possible to implement new policies using our SDK




- Policy simulator
  - Ease the initial phase of study of new policies
  - Results from simulator match experimental virtualization results

# Outline

- System Management Tools
- System-level Virtualization
- Fault Tolerance / Resilience
- **RAMS'08 Student Project**

# RAMS'08 Student Project

*"Automating the Run-time Infrastructure for the Performance Evaluation Process"*, Tara McQueen (Delaware State)

- Student investigated
  - Performance evaluation framework (Cbench)
  - Cluster configuration & management system (OSCAR)
- Extend OSCAR to support performance evaluation package
  - Separation of Cbench into core components & benchmarks
  - Create packages (RPMs) for core & HPCC benchmark

# Results

- Integration of new benchmark (LAMMPS) into Cbench
  - Write parser module to process benchmark data
- Created HowTo guide for new benchmark integration
- Provided benchmarking framework for OSCAR
  - Source code already available for the community
  - To be included in future OSCAR release!
- Student take away…
  - Learn to read/write Makefiles
  - Learn to read/write Perl module
  - Learn about binary package/distribution files (RPMS)
  - Exposure to HPC benchmarks & system management details

# Questions?

***Systems Research Team***
**Computer Science and Mathematics Division**
**Oak Ridge National Laboratory**
**http://www.csm.ornl.gov/srt**

Supported by:

This work was supported by the U.S. Department of Energy, under Contract DE-AC05-00OR22725.