Hierarchical Management of the Dynamic Allocation of Resources in a Multi-Node System

Oracle Corporation: Feb 2005
Benny Souder, Debashish Chatterjee, Lakshminarayanan Chidambaran, David Brower, Carol Colrain, Daniel Semler, Christopher A. Kantarjiev, James W. Stamos
Oracle Systems Engineer and Consultant in Chicago
OCP, RAC since 2002,
Systems Admin and Developer in previous lives
Likes to Swing Dance
http://www.ardentperf.com
Summary

- This paper details a technique to implement grid computing concepts for databases. *(Related concepts: virtualization and cloud computing.)*
- Each particular type or category of work performed by clients is labeled as a “service”. *(Finance, payroll, etc.)*
- The server allocates and manages resources among services on multiple levels, arranged in a hierarchy.
  - Node, database, cluster and farm.
- At each level, a process called a “director” is responsible for doing the resource management work.
Grid Architecture

- A node is a single physical server. An instance is a set of database processes and cache on a node.
- A database consists of multiple instances.
- A cluster may contain multiple databases.
- A cluster farm may contain multiple clusters.

google.com/patents?id=jZiQAAAAEBAJ&pg=PA2
Workload Monitoring

- Policies are defined by administrators according to service-level agreements for performance metrics.
  - Example metrics include transaction execution time or percent CPU utilization.
- A workload monitor runs on cluster nodes and measures performance characteristics of each service.
- Performance data is sent to directors to evaluate whether service-level agreements are being met.
- Directors can take automatic action (reallocate resources) to remediate performance shortfalls.
Director Hierarchy

- Single director process can act in multiple roles.
- Database director assigns services in multinode DB
- Cluster director can allocate nodes among databases by starting/stopping instances.
- Farm director can move nodes between clusters.

google.com/patents?id=kpiQAAAAEBAJ&pg=PA4
Director Operation

- To remedy service-level performance problems, resource adjustments start at the lowest level, in the least disruptive and cheapest manner. Higher-level resource allocations are only adjusted if required.
- Directors communicate using a transactional, recoverable message queue.
- Every node has a director process.
  - Active director is elected through global lock, IPC or static assignment.
  - All others act as standby directors.
Critiques

• Major shortcoming: cannot share single node between multiple databases. (Oracle namespace design often prevents consolidation of many small databases which don't need a dedicated server.)

• Doesn't mention automatic rebalancing of existing connections after expanding service to new nodes. Capability is outlined in Transparent Session Migration paper [referenced] but not used here.
Three Questions

• What are the three proposed levels of directors?

• What is the smallest unit of granularity at which resources can automatically be reallocated within the grid?

• What could cause resources to be automatically reallocated?
Questions, comments, suggestions?