

Cloud Computing and Grid Computing 360-Degree Compared

Ioan Raicu

**Computer Science Department
Illinois Institute of Technology**

CS 595

Hot Topics in Distributed Systems: Data-Intensive Computing

September 27th, 2010

Resource Management

- Compute Model
 - batch-scheduled vs. time-shared
- Data Model
 - Data Locality
 - Combining compute and data management
- Virtualization
 - Slow adoption vs. central component
- **Monitoring**
- **Provenance**

Programming and Application Model

- Grids:
 - Tightly coupled
 - High Performance Computing (MPI-based)
 - Loosely Coupled
 - High Throughput Computing
 - Workflows
 - Data Intensive
 - Map/Reduce
- Clouds:
 - Loosely Coupled, transactional oriented

Programming Model Issues

- **Multicore** processors
- Massive **task parallelism**
- Massive **data parallelism**
- Integrating **black box applications**
- Complex **task dependencies** (task graphs)
- **Failure**, and other execution management issues
- **Dynamic task graphs**
- Documenting **provenance** of data products
- **Data management**: input, intermediate, output
- **Dynamic data access** involving large amounts of data

Gateways

- Aimed to simplify usage of complex resources
- Grids
 - Front-ends to many different applications
 - Emerging technologies for Grids
- Clouds
 - Standard interface to Clouds

An Example of an Application in the Grid



Security Model

- Grids
 - Grid Security Infrastructure (GSI)
 - Stronger, but steeper learning curve and wait time
 - Personal verification: phone, manager, etc
- Clouds
 - Weaker, can use credit card to gain access, can reset password over plain text email, etc

Conclusion

- Move towards a mix of micro-production and large utilities, with load being distributed among them dynamically
 - Increasing numbers of small-scale producers (local clusters and embedded processors—in shoes and walls)
 - Large-scale regional producers
- Need to define protocols
 - Allow users and service providers to discover, monitor and manage their reservations and payments
 - Interoperability

Conclusion (cont)

- Need to combine the centralized scale of today's Cloud utilities, and the distribution and interoperability of today's Grid facilities
- Need support for on-demand provisioning
- Need tools for managing both the underlying resources and the resulting distributed computations
- Security and trust will be a major obstacle for commercial Clouds by large companies that have in-house IT resources to host their own data centers

Questions

