

Policy-Based Data Management

Reagan W. Moore, Arcot Rajasekar, Mike Wan, Wayne Schroeder, Chien-Yi Hou

Data grids organize distributed data into shareable collections through virtualization mechanisms. Logical name spaces are managed for storage resources, users, files and files. Protocol interoperability between clients and storage systems is handled by mapping client actions to standard operations, and mapping the standard operations to storage specific protocols. The next generation data grid technology represented by the integrated Rule-Oriented Data System automates policy enforcement. Policies are cast as computer actionable rules that apply processing workflows at the remote storage location. The workflows are composed from micro-services (modular C code). Resulting state information is managed in a central metadata catalog. The combination of data grid virtualization mechanisms with policy based data management makes it possible to automate administrative functions, enforce policies such as distribution, retention, disposition, replication, and periodically validate assessment criteria.