

„Service Level Agreement aware Resource Management “

Matthias Hovestadt

Zusammenfassung

Grid Systeme der nächsten Generation zielen auf die Einbindung kommerzieller Benutzer, so dass diese das Grid für geschäftskritische Berechnungen nutzen. Solche Kunden fordern jedoch vertraglich zugesicherte Dienstgütern, die die fristgemäße Verfügbarkeit von Rechenergebnissen sicherstellen. Hierbei ist das Service Level Agreement (SLAs) ein zentrales Instrument zur Festlegung eines umfassenden Anforderungsprofils.

Zahlreiche Forschungsprojekte weltweit arbeiten bereits an einer Integration von SLA-Mechanismen in die Bausteine der Grid Middleware. Allerdings ist dies allein noch nicht ausreichend. Dienste im Bereich der Grid Middleware akzeptieren zwar die Aufträge der Kunden, sind zur Realisierung der Jobs jedoch auf Dienste lokaler Resource Management Systeme (RMS) angewiesen, die ihre Ressourcen innerhalb des Grids zur Verfügung stellen. Derzeitige RMS bieten lediglich einen „Best-Effort“ Dienst, bei dem keinerlei Garantien hinsichtlich der Dienstgüte gegeben werden. Aus diesem Grund können derzeitige Grid Middleware Komponenten nur eingeschränkte Dienstgütegarantien geben.

In dieser Arbeit wird die Architektur und Funktionsweise eines SLA-fähigen Resource Management Systems vorgestellt, das es den Grid Middleware Komponenten erlaubt über die Inhalte der SLAs zu verhandeln. Das System nutzt interne Mechanismen, um applikationstransparente Fehlertoleranz zu bieten. Diese erlauben die Einhaltung von SLAs auch im Fall lokaler Ausfälle einzelner Ressourcen. Der Hauptteil dieser Arbeit beschreibt Aspekte des Scheduling sowie Strategien zur Einhaltung von SLAs, sowie hierzu erforderliche Aspekte des Systemdesigns und konkreter Implementation.

Nach der Vorstellung von Anforderungen der Grid Middleware hinsichtlich Dienstgüte folgt eine Beschreibung der Ausführungsphasen des SLA-fähigen RMS sowie clusterinterne Schedulingstrategien, die den Level der erbringbaren Dienstgüte bestimmen. Hierbei nutzt das System ausschließlich eigene Ressourcen und Mechanismen, um die Folgen eines Ausfalls zu kompensieren.

Zur Verbesserung der Fehlertoleranz werden hiernach Strategien zur grenzüberschreitenden Migration vorgestellt, bei der nicht nur Cluster innerhalb der gleichen Domäne, sondern auch beliebige Ressourcen innerhalb des Grids genutzt werden können. Um die erfolgreiche Fortführung von Berechnungen auf diesen Ressourcen sicherzustellen, wird ein Mechanismus zur Beschreibung eines Kompatibilitätsprofils vorgestellt.

Das Konzept des SLA-fähigen RMS wurde im Rahmen des EU-geförderten Projekts HPC4U implementiert. Es werden Aspekte dieser Implementation sowie Ergebnisse aus der praktischen Anwendung vorgestellt.

Keywords

Grid Computing, Grid Middleware, Grid, Resource Management, Service Level Agreement, SLA, Negotiation, Fault Tolerance, Scheduling, Process Checkpointing, Storage Snapshot, Network Fault Tolerance, Quality of Service, QoS