

Promotionsthema

Verteilte Datenstromverarbeitung in privaten P2P-Netzwerken

von Timo Michelsen

In vielen Bereichen der IT-gerichteten Informationsverarbeitung existieren keine klassischen persistenten Datenbanken, sondern flüchtige, theoretisch unbegrenzte Datenströme. Datenstrommanagementsysteme (DSMS) bieten allgemeine Lösungen für die kontinuierliche Verarbeitung solcher Daten. Verteilte DSMS koordinieren die kontinuierliche Datenstromverarbeitung über mehrere Maschinen in einem Netzwerk. Dadurch erreichen sie im Vergleich zu monolithischen DSMS eine erhöhte Leistungsfähigkeit, Zuverlässigkeit und Skalierbarkeit. Sie nutzen häufig klassische Client/Server-Architekturen, Grids oder Computercluster, welche jedoch nicht in jedem Anwendungsszenario zur Verfügung stehen, etwa aus organisatorischen, wirtschaftlichen oder technischen Gründen.

Die Kernidee dieser Dissertation ist, bereits vorhandene Systemressourcen von privaten Maschinen der Nutzer zu einem Peer-to-Peer (P2P) Netzwerk zu verbinden und diese zur verteilten Datenstromverarbeitung zu nutzen. Jeder sogenannte Peer bietet dem Netzwerk verschiedene Systemressourcen an (Heterogenität). Zudem nehmen sie freiwillig an der verteilten Datenstromverarbeitung teil und können zu keiner Tätigkeit gezwungen werden (Autonomie). Dementsprechend kann ein Peer jederzeit das Netzwerk betreten oder verlassen, was zu einem dynamischen und dezentral organisierten P2P Netzwerk führt.

In der Dissertation wird ein Ansatz konzipiert, um ein verteiltes DSMS auf Basis eines solchen P2P-Netzwerks prototypisch um- und einzusetzen. Ausgangspunkt ist ein existierendes, monolithisches DSMS sowie ein bereits aufgebautes P2P-Netzwerk. Es wird untersucht, wie in einer solchen Umgebung kontinuierlichen Anfragen verteilt ausgeführt und die Vor- und Nachteile des P2P Netzwerks berücksichtigt werden können. Dazu befasst sich die Dissertation mit Replikation, Fragmentierung, Lastbalancierung und Recovery im Datenstromkontext. Der resultierende Prototyp soll im Rahmen zweier Anwendungsfälle – Smart Home und Basketballspielanalyse in Echtzeit – evaluiert und mit einem monolithischen DSMS verglichen werden.