

Kurzfassung

Verteilte Systeme können heutzutage große Mengen von Daten verarbeiten. Der Transport dieser Daten auf ein verarbeitendes System wird Data Staging genannt. Die UNICORE-Middleware ermöglicht das Data Staging mithilfe verschiedener Datentransfer-Protokolle, unter anderem auch mit dem UNICORE File Transfer Protocol, kurz UFTP. Im Rahmen dieser Arbeit wurde die Funktionsweise von UFTP untersucht und dessen Anwendbarkeit, Performance und Skalierbarkeit im Hinblick auf Data Staging Anwendungsfälle am ZIH der TU Dresden evaluiert. Dazu wurde eine Reihe von Tests mit synthetischen und realen Daten durchgeführt. Es hat sich dabei herausgestellt, dass das Data Staging am ZIH durch die Nutzung von UFTP durchaus optimiert werden kann. Es hat sich jedoch auch gezeigt das UFTP nicht in allen Anwendungsfällen das zweckdienlichste Data Staging Protokoll ist. So ist der Transport kleiner Dateien (< 1 MiB) mit UFTP bspw. sehr ineffizient. Um in diesem und in anderen Fällen die optimale Integration des Protokolls in bestehende Arbeitsabläufe zu ermöglichen, wurde deshalb eine Richtlinie entwickelt, die für alle relevanten Anwendungsfälle von UFTP die optimalen UFTP Parameter empfiehlt.

Abstract

Distributed Systems are nowadays capable of processing vast amounts of data. The process of transferring that data onto the processing system is called data staging. The UNICORE-middleware allows data staging through several data transfer protocols, including the UNICORE File Transfer Protocol, or short UFTP. This thesis examined how UFTP functions and evaluated its applicability, performance and scalability with regards to data staging use cases at the ZIH of the TU Dresden. Multiple tests with synthetic and real data were conducted on this basis. It became clear that data staging at the ZIH can be optimized by using UFTP. It has however also become evident, that UFTP is not the most appropriate data staging protocol in every situation. The transfer of small files (< 1 MiB) for example is very inefficient in UFTP. To optimally integrate the protocol in existing workflows for all data staging use cases, a guideline was developed that recommends the best UFTP parameters for all relevant use cases.