Data Transfer Options on SuperMUC-NG

Overview
SuperMUC-NG provides a high performance network connection to the outside world. In order to perform easy and fast data staging to/from SuperMUC-NG, we provide various data transfer methods. However, because of limited support resources, there is a distinction on what is available to transfer data to/from arbitrary external sites and to/from our GCS and PRACE partner sites. In the following document, we want to give you the necessary information on using the available data transfer options.

SuperMUC-NG DTN Setup
The diagram below shows the SuperMUC-NG DTN topology. As you can see, we operate currently four Login nodes, two Globus Online and two Special Purpose Data Transfer Nodes (DTNs), each one connected via a 100GE network link to our Data Center Network. The Data Center Network is connected via four 100GE links to the German Research Network (DFN), that are operated in a pair wise active/passive mode. The DFN then provides high bandwidth connections to the internet and via the European Research Network (Geant) to other Research Networks all over the world.

- Overview
- SuperMUC-NG DTN Setup
- Data Transfer to/from arbitrary external sites
  - Globus Online
  - Secure Copy
- Data Transfer to/from GCS sites
  - Unicore FTP
  - Grid Community Toolkit (GridFTP)
- Data Transfer to/from PRACE sites
  - Grid Community Toolkit (GridFTP)
- Beyond Data Transfer: Sharing and Public Access

Please note that the Data Transfer System for SuperMUC-NG is still under development. So this document, as well as the available data transfer options will increase over the next few months. So check back from time to time to see what's new.

Note that the Data Transfer Nodes (DTNs) only provide server side functionality. Initiating Transfers to/from the DTNs is always done by the user from the Login Nodes or via the Globus Online Webinterface. User Login to the DTNs is not possible.
Data Transfer to/from arbitrary external sites

In the following we will outline the two data transfer methods, we currently support from/to arbitrary external endpoints.

⚠️ In order to be able to transfer data to/from an arbitrary external site, the IP(s) of the external endpoint(s) have to be registered as trusted IP(s) by your project in the SuperMUC-NG firewall. If not already done so, please ask the Master User of your project to submit a Service Request to register your IP(s) in the SuperMUC-NG firewall.

**Globus Online**

The preferred data transfer method on SuperMUC-NG to/from arbitrary external sites is to use Globus Online. LRZ operates an dedicated Globus Endpoint for SuperMUC NG called LRZ SuperMUC-NG Globus DTN. In order to get started with Globus Online check out this short tutorial. Note that you can use your SuperMUC-NG credentials to log in to Globus Online and must use it to authenticate against the LRZ SuperMUC-NG Globus DTN endpoint. Just search for LRZ or Leibniz in the list of organisations on the login page(s).

✅ In general Globus Online performs best if you transfer multiple (>8) large (>10GB) files so that it can efficiently parallelise data transfers. This is especially important over long distance links.
If you already used Globus on SuperMUC, you may ask yourself now one or more of the following questions:

> Is it really that simple? => Yes. It is!
> What about getting certificates and linking them to my user account? => No more need to hassle with certificates. Your SuperMUC-NG login credentials will be all you need.
> What about that complicated globus-url-copy command? => Gone. Check out the new Globus Online CLI.

Secure Copy

For smaller amounts of data you also can use secure copy (`scp`), secure ftp (`sftp`) or RSYNC (`rsync`) using `ssh` as remote shell from your external endpoint to the SuperMUC-NG login nodes. Note that the direction from external to SNG is important as the firewall of SuperMUC-NG does not permit outgoing ssh connections.

When choosing this data transfer method, please bear in mind that `ssh`, `scp` and `sftp` are very limited in the achievable performance, especially over high latency WAN links. See also https://fasterdata.es.net/data-transfer-tools/say-no-to-scp/ on this topic.

Data Transfer to/from GCS sites

The three GCS Sites (HLRS in Stuttgart, JSC in Jülich and LRZ in Garching) operate a network of Data Transfer Nodes (DTNs), that have been optimised to enable you to transfer data with up to 10GB/s between the three German National Flagship Supercomputers.

Unicore FTP

Grid Community Toolkit (GridFTP)

Data Transfer to/from PRACE sites

Coming Soon.

Grid Community Toolkit (GridFTP)

Beyond Data Transfer: Sharing and Public Access

If you want to share data generated on SuperMUC-NG with external parties or even make this data publicly available, you can have a look at our LRZ Data Science Storage (DSS). GCS has funded 20PB of DSS storage for SuperMUC-NG. For details on how SuperMUC-NG projects can apply for storage space on DSS see: Data Science Storage for SuperMUC.