DSS documentation for users

- 1. How to get access to DSS
- 2. Storing and accessing data in DSS
  - 2.1. Using DSS via SuperMUC and LinuxCluster
  - 2.2. Using DSS via Compute Cloud and VMware
    - 2.2.1. Prerequisites
    - 2.2.2. Preparing users and groups on the VM
    - 2.2.3. Mounting a DSS Container on a VM
  - 2.3. Using DSS world wide via GridFTP and Globus
    - 2.3.1. First time registration of your Grid certificate at LRZ
    - 2.3.2. Log in to the Globus Research data management portal
    - 2.3.3. Using Globus to transfer files between your workstation and DSS
    - 2.3.4. Using Globus to transfer files between your Servers and DSS
    - 2.3.5. Using Globus to transfer and share data world-wide
    - 2.3.6. Using Grid command line utilities to transfer data
- 3. Hints and possible pitfalls
  - 3.1. Known Limitations
  - 3.2. Do's and Don'ts

1. How to get access to DSS

Access to a DSS container is only possible on invitation by the data curators of the specific DSS container. If a data curator invites you for accessing a DSS container, an invitation Mail is sent to the EMail address, associated with your user in the LRZ IdPortal. If you are in doubt that this address is correct, please log in to the LRZ IdPortal and check by viewing your account settings (Account View) and contact your master user, if the address is wrong.

When you receive an invitation Mail, you have to confirm the invitation by clicking on the link in the Mail and follow the instructions on the website, in order to activate the access. Please note that there may be a latency of up to an hour before the updated access rights have been propagated to all attached systems.

2. Storing and accessing data in DSS

In the following we describe the versatile ways in which you may access the data stored in a DSS container.

2.1. Using DSS via SuperMUC and LinuxCluster

On LRZ's SuperMUC and the LinuxCluster, the DSS file systems are directly mounted via a high performance file system client on the following system parts:

<table>
<thead>
<tr>
<th>System</th>
<th>Subcomponent</th>
<th>Accumulated BW</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>SuperMUC</td>
<td>Phase I Login Nodes</td>
<td>20 Gbit/s</td>
<td>sb.supermuc.lrz.de sb-tsm.supermuc.lrz.de</td>
</tr>
<tr>
<td>SuperMUC</td>
<td>Phase II Login Nodes</td>
<td>20 Gbit/s</td>
<td>hw.supermuc.lrz.de hw-tsm.supermuc.lrz.de</td>
</tr>
<tr>
<td>LinuxCluster</td>
<td>CoolMUC II Login Nodes</td>
<td>60 Gbit/s</td>
<td>lxlogin[567].lrz.de</td>
</tr>
<tr>
<td>LinuxCluster</td>
<td>CoolMUC II Compute Nodes</td>
<td>200 Gbit/s</td>
<td></td>
</tr>
<tr>
<td>LinuxCluster</td>
<td>CoolMUC III Login Nodes</td>
<td>80 Gbit/s</td>
<td>lxlogin8.lrz.de</td>
</tr>
<tr>
<td>LinuxCluster</td>
<td>CoolMUC III Compute Nodes</td>
<td>20 Gbit/s</td>
<td></td>
</tr>
<tr>
<td>LinuxCluster</td>
<td>Teramem</td>
<td>200 Gbit/s</td>
<td></td>
</tr>
<tr>
<td>LinuxCluster</td>
<td>Hugemem</td>
<td>10 Gbit/s</td>
<td></td>
</tr>
</tbody>
</table>

On these systems, DSS containers can be accessed via the file system path

/dss/<data project>/<data container>/
2.2. Using DSS via Compute Cloud and VMware

2.2.1. Prerequisites

In order to access a DSS container from a LRZ Compute Cloud or VMware virtual machine, you must ask the data curator of the data project, to which the desired container belongs, to export the container to the IP address used by your VM.

Though technically not forbidden, you should only export DSS containers to IPs that are statically assigned to and trusted by you. NFS exports follow a "host based trust" semantic, which means the DSS NFS server will trust any IP/system to which a DSS container is exported. There is no additional user authentication between NFS server and client enforced. This is especially important if you want to export DSS containers to cloud machines, as these - by default - use a dynamically allocated IP, which may be reused by other machines as soon as you shut down your VM.

If you want to use NFS v4 instead of NFS v3 to mount the data container inside your VM, please make sure to configure the ID mapping daemon accordingly. The easiest thing to do is to set the `DOMAIN` parameter in `/etc/idmapd.conf` to LRZ.DE and follow the user and group setup steps described below. You may also be able to successfully setup some kind of user mapping between the local users on your VM and the users, used by DSS. However this is not covered by LRZ's support for DSS.

2.2.2. Preparing users and groups on the VM

As you may be aware, NFS user permissions are based on user ID (UID). UIDs of any users on the client must match those on the server in order for the user to have access. The typical ways of doing this are either through some kind of manual synchronisation or the use of some kind of directory service like LDAP for example.

Currently, LRZ does not allow access to its LDAP servers from customer VMs because of data privacy reasons. However, as DSS supports arbitrary users from the central TUM and LMU Identity Management Systems, you may be able to connect your VMs to the user directory of your organisation. For more information on how to do that, please contact the Servicedesk of your organisation, as this is out of scope of LRZ's support for DSS.

In order to create the users manually on your VM, you must determine the username and UIDs of the users, which are invited to the particular container and should be able to access the data via your VM. You can ask the data curators of your container, to provide you a list, mapping the usernames to UIDs.

Usually, you also have to create the particular container access groups on your VM. As normally the groups a user belongs to are also provided by the NFS client. However to work around a known limitation of the NFS protocol in handling groups (see this for more info), group membership is managed by the DSS NFS servers. So technically it is not mandatory to create the container groups on your side.

2.2.3. Mounting a DSS Container on a VM

In order to mount the container on your VM, you have to ask the data curators of your project for the IP address and path of your NFS export. Once you have this information, you should be able to mount the container on your VM. We suggest to use the following command:

```
your-vm:># mount -t nfs -o rsize=32768,wsize=32768,hard,tcp,bg,
    timeo=600,vers=3 <IP>:<Path>
```

2.3. Using DSS world wide via GridFTP and Globus

In order to access the data stored in DSS containers from outside of the LRZ, we provide a GridFTP service, which integrates into the Globus Research data management portal. With this setup, you can easily transfer and share DSS container data world wide, using a protocol which is optimised for high speed transfer via wide area networks (WAN).

Please note that it may take up to several hours after you have accepted your DSS invitation and performed first time registration (see step 2.3.1) before you can successfully access the DSS container via Globus. This is because there still has to trickle-down some information through various LRZ system components via regularly running cron jobs.

If you cannot access you DSS container after 12 hours after invitation acceptance and first time registration (see step 2.3.1) please raise a ticket via the LRZ Servicedesk.

2.3.1. First time registration of your Grid certificate at LRZ
The GridFTP infrastructure, on which the Globus research data management portal relies, works by using X.509 certificates for user authentication. However, as DSS is based on a POSIX file system, the GridFTP server needs to know how to map X.509 to your LRZ/TUM/LMU username. If the last two sentences sound complicated to you or don’t make sense at all, don’t worry. We worked hard in order to hide most of the complexity from you. All you really need to know is that in order to access your DSS containers via Globus, you first have to register here for each of your invited accounts.

If you are really curious and want to know all the nitty gritty details, we recommend you check out this.

2.3.2. Log in to the Globus Research data management portal

You can login to Globus by clicking on the Log In button on the upper right of the page.

You then will be directed to a page, where you have to select your identity provider. Depending on if you are using a LRZ, TUM or LMU user, select the appropriate institution from the dropdown list.

Then click on the continue button in order to start the login workflow.
The login workflow will redirect you to the Shibboleth Single Sign on Provider of your chosen institution. Use the username and password provided by your institution (LRZ/TUM/LMU).

After that you should be successfully logged in and see the following transfer view.

2.3.3. Using Globus to transfer files between your workstation and DSS

In order to transfer files between a DSS container and your workstation, you currently have to install the Globus Connect Personal software on your workstation and setup a so called Personal Endpoint. To do so, just click on the install link that is appropriate for your operating system on the Globus Connect Personal site and follow the instructions.

After that, you can go back to the transfer view and select the Personal Endpoint for your workstation, you have just created by clicking on the Start by selecting an endpoint link in one of the two windows. This will open a search window in which you can search and select your workstations endpoint.
After selecting your endpoint, you will see the content of the directories on your workstation, you have exported via Globus (usually the HOME directory)
After that, you can now select the DSS Endpoint on the other side of the page. Just again click on Start by selecting an endpoint link and this time search for LRZ DSS. Select the Endpoint Leibniz Supercomputing Centre’s DSS - CIlogon endpoint.

Now you should see the content of your workstation on one side and the base directory of DSS on the other side.
Now you can navigate/browse through the directories and start a transfer by just navigating to the destination folder on the destination, selecting the source folder or files on the source side and click the big blue button that points from the source to the destination.

Please note that you can also adjust some Transfer Settings on the bottom of the page. For example you can choose to encrypt the data transfer if you require an extra level of privacy, or you can even tell Globus to "Sync" your directories like you may be used to do by tools like rsync.
Basically as soon as you have started the transfer you are done. That means you can now navigate away from the page and do other stuff, while Globus is doing the heavy lifting in the background. Once the transfer has finished (or failed permanently), you will receive an email from Globus.

However, if you are curious and want to watch Globus while it does its magic you can click on the Activity tab and get an overview of your recent transfers.

When you click on one of those transfers, you can also get some more details about it.
2.3.4. Using Globus to transfer files between your Servers and DSS

If you want to transfer files regularly and with more performance than the Globus Personal Endpoint can deliver, you can also setup a Globus Endpoint Server on the servers on your institutions. Running a basic Globus Endpoint Server is free of charge. Just check out the Glo bus Connect Server page for download and installation/configuration instructions. Getting this up and running for the first time should be pretty doable in a day or two for medium to advanced Linux Sysadmin. However, if you struggle, please don’t hesitate to call out on us via the LRZ Servicedesk.

2.3.5. Using Globus to transfer and share data world-wide

Like the name Globus suggest, its mission is to connect data islands around the world and enable easy, fast and reliable data transfers between these islands. Therefore, many science institutions run their own Globus Endpoints. So if you need to transfer your data between LRZ and some other site, chances are good that there is already an Endpoint setup at your partner site. However, if this is not the case, you can ask them to install a Globus Connect Server or at least run a Globus Personal Endpoint for you on their systems. Both options are free of charge.

For DSS we even have signed a premium subscription for Globus which enables us to provide you another great feature, which basically is very unique. You can share your data in DSS world-wide with arbitrary people as easy as you may be already used to by tools like LRZ Sync+Share or Dropbox. Just provide Globus the Email address of the person you want to invite together with the permissions, you want to give the person (read or write access for example), and you are done. However, while tools like LRZ Sync+Share or Dropbox or Google Drive come to their limitations when we talk about data sizes of more than 1 TB, this is where Globus Sharing just begins...

2.3.6. Using Grid command line utilities to transfer data

If you don't want to use the graphical user interface for transferring data using Globus, you can also check out their Command Line Interface or RESTful API.

3. Hints and possible pitfalls

3.1. Known Limitations
3.2. Do's and Don'ts