NorduGrid activities in Oslo

Resources available:
• 1 cluster (controlled by OpenPBS)
  2 computing nodes (uni– and dualprocessor machines).
  Gatekeeper with Globus Toolkit™ software.

Main activities:
• Testing upcoming new software.
• Trying to apply existing software for real task.

Conclusion:
- Not too much is done.
NorduGrid activities in Oslo

There are few job staging models, that can be adopted.

1. Job description contains information about necessary data.

   Implementations:
   staging software is responsible for providing data.
   job is staged to the place where data is available directly.

2. Small-sized job is staged. It is up to the job to download/access necessary data.

   Implementations:
   only wrapper, capable of downloading data and, maybe, job binary itself is staged.
   "grid aware" application capable of locating and obtaining data

3. Job is staged with all necessary data.

   Implementations:
   data is pre-staged by user using available protocols.
   job is self-containing.
NorduGrid activities in Oslo

**Chosen model:**
Staging self-contained job.

**Realisation:**
Wrapper scripts around Globus tools.

**Reasons:**
Implementation simplicity.
Compatibility with other models.
Expandable – new features can be added as system evolves.

**Disadvantages:**
Not comfortable for jobs with big amount of data (depending on software used it can be also unstable).
Ineffective for running many jobs with same/overlapping set of input data (this is the case for HEP applications).
NorduGrid activities in Oslo

*Is Globus Toolkit™ suitable for chosen model?*

Test staging of jobs up 0.5GB+ was performed successfully on local network (smaller jobs were submitted to geographically remote locations).

Staging speed \(\sim 7\text{MB/s}\) (close to the limit on 100Mbps networks). Only minimal amount of software in addition to standard Globus utilities is required.
NorduGrid activities in Oslo

Is chosen model (implemented using Globus Toolkit™) suitable for Grid?

Although can be used for jobs with **not too big amount** of data in **geographically compact** environments, this way is not suitable for Grid.

**Arguments:**

Used job submission through GASS does not support resuming data transfer and can’t be reliably/effectively used in geographically distributed environments (at least using standard Globus tools). Moving a lot of data through highly loaded networks can’t be recognized as good solution.