

# GRID Application Portal

Martin Matusiak<sup>1</sup>   Jonas Lindemann<sup>2</sup>

<sup>1</sup>The NTNU High Performance Computing Project  
Norwegian University of Science and Technology

<sup>2</sup>Lunarc, Center for Scientific and Technical Computing  
Lund University

1st Nordic Grid Neighbourhood Conference  
University of Oslo, Norway, 15-17 August 2005

# Outline

- 1 Motivation
  - The command line interface
  - A proposed solution
  - The portal interface
- 2 Design
  - Relation to Nordugrid/ARC
  - Authentication mechanism
- 3 Applications on a GRID portal - BLAST
  - BLAST - an introduction
  - BLAST at [norgrid.ntnu.no](http://norgrid.ntnu.no)

# Outline

- 1 **Motivation**
  - The command line interface
    - A proposed solution
    - The portal interface
- 2 Design
  - Relation to Nordugrid/ARC
  - Authentication mechanism
- 3 Applications on a GRID portal - BLAST
  - BLAST - an introduction
  - BLAST at [norgrid.ntnu.no](http://norgrid.ntnu.no)

# Introducing the command line interface to Nordugrid

```
alex@frasier:~  
marmat@norgrid blast $ grid-proxy-init  
Your identity: /O=Grid/O=NorduGrid/OU=ntnu.no/CN=Martin Matusiak  
Enter GRID pass phrase for this identity:  
Creating proxy ..... Done  
Your proxy is valid until: Sat Aug 13 07:08:39 2005  
marmat@norgrid blast $ ngsync -f job.xrsl -c norgrid.ntnu.no -d0  
run.sh      0 s:      0.2 kB      0.0 kB/s      0.0 kB/s      . . .  
martin_testset.txt  0 s:      0.9 kB      0.0 kB/s      0.0 kB/s      . . .  
  
Job submitted with jobid gsift://norgrid.ntnu.no:2811/jobs/2100011238665261699770589  
marmat@norgrid blast $ ngsync -fa -c norgrid.ntnu.no  
marmat@norgrid blast $ ngstat -a  
Job gsift://norgrid.ntnu.no:2811/jobs/2100011238665261699770589  
  Jobname: blastpJob4  
  Status: FINISHED  
marmat@norgrid blast $ ngget -a  
ngget: downloading files to /home/marmat/blast/2100011238665261699770589  
ngget: download successful - deleting job from gatekeeper.  
marmat@norgrid blast $ ls 2100011238665261699770589/  
blast_out.QUERY  gmlog/  martin_testset.txt  run.sh  stderr.txt  stdout.txt  
marmat@norgrid blast $
```

# Assessing the command line interface

## Advantages:

- Flexible
- Efficient
- Suitable for large data sets

Conclusion: **Ideal for the "power user"**

## Drawbacks:

- Intimidating at first sight
- Commands require memorizing
- Not everyone is comfortable with Unix

Conclusion: **Sub-par for the casual user**

# Outline

1

## Motivation

- The command line interface
- **A proposed solution**
- The portal interface

2

## Design

- Relation to Nordugrid/ARC
- Authentication mechanism

3

## Applications on a GRID portal - BLAST

- BLAST - an introduction
- BLAST at [norgrid.ntnu.no](http://norgrid.ntnu.no)

# A GRID Application Portal

A solution proposed by Jonas –  
the LUNARC Application Portal,

- offering a web-based interface for simplicity,
- revolving around a work flow model (create job, submit job, monitor job, get job),
- providing a unified interface to applications (adding support for new applications is straightforward),
- without compromising the security model.

# A portal in two flavors

## LUNARC Application Portal

- the original codebase
- developed at Lund University by Jonas

## GRIDportal

- a fork off LUNARC Application Portal
- developed at NTNU by Martin to suit NTNU needs

In spite of the split, both are moving toward an eventual merge.



# Outline

- 1 **Motivation**
  - The command line interface
  - A proposed solution
  - **The portal interface**
- 2 **Design**
  - Relation to Nordugrid/ARC
  - Authentication mechanism
- 3 **Applications on a GRID portal - BLAST**
  - BLAST - an introduction
  - BLAST at [norgrid.ntnu.no](http://norgrid.ntnu.no)

# Aims of the portal interface

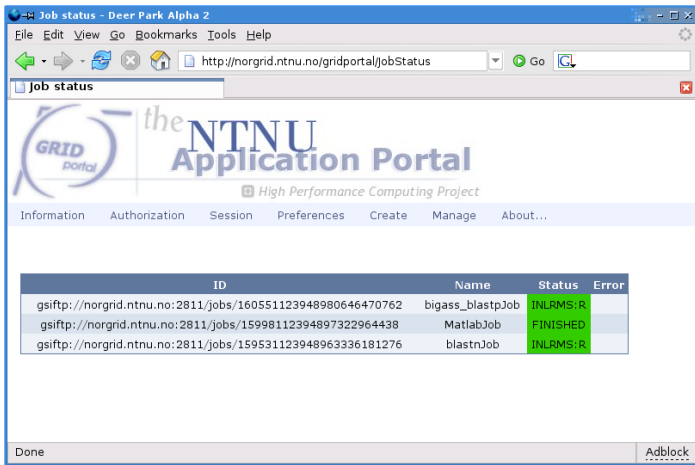
The portal aims to:

- make GRID computing easy to the "uninitiated" with a minimum of schooling
- conceal the intricate details of GRID computing
- offer a pluggable interface to applications

# Introducing the portal interface to Nordugrid (1/4)

The screenshot shows a web browser window titled "ManageJobPage - Deer Park Alpha 2". The address bar contains the URL "http://norgrid.ntnu.no/gridportal/ManageJobPage". The page header features the "GRID portal" logo and the text "the NTNU Application Portal" and "High Performance Computing Project". A navigation menu includes "Information", "Authorization", "Session", "Preferences", "Create", "Manage", and "About...". The main content area is titled "Manage jobs" and lists four job types: "MatlabJob", "blastnJob", "StorblastpJob", and "bigass\_blastpJob". Below the list are buttons for "Edit", "View files", "Submit", "Delete", and "Reset". The status bar at the bottom shows "Done" and "Adblock".

# Introducing the portal interface to Nordugrid (2/4)



The screenshot shows a web browser window titled "Job status - Deer Park Alpha 2". The address bar contains the URL "http://norgrid.ntnu.no/gridportal/jobStatus". The page header features the "GRID portal" logo and "the NTNU Application Portal" text, with a sub-header "High Performance Computing Project". A navigation menu includes "Information", "Authorization", "Session", "Preferences", "Create", "Manage", and "About...". The main content area displays a table of job status information.

ID	Name	Status	Error
gsiftp://norgrid.ntnu.no:2811/jobs/160551123948980646470762	bigass_blastpJob	INLRMS:R	
gsiftp://norgrid.ntnu.no:2811/jobs/15998112394897322964438	MatlabJob	FINISHED	
gsiftp://norgrid.ntnu.no:2811/jobs/159531123948963336181276	blastnJob	INLRMS:R	

The browser status bar at the bottom shows "Done" and "Adblock".

# Introducing the portal interface to Nordugrid (3/4)

ManageGridJobPage - Deer Park Alpha 2

File Edit View Go Bookmarks Tools Help

http://norgrid.ntnu.no/gridportal/ManageGridJobPag

ManageGridJobPage

**the NTNU Application Portal**  
High Performance Computing Project

Information Authorization Session Preferences Create Manage About...

**Manage GRID jobs**

JobID	JobName	Status
gsiftp://norgrid.ntnu.no:2811/jobs/160551123948980646470762	bigass_blastpJob	INLRMS:R
gsiftp://norgrid.ntnu.no:2811/jobs/15998112394897322964438	MatlabJob	FINISHED
gsiftp://norgrid.ntnu.no:2811/jobs/159531123948963336181276	blastnJob	INLRMS:R

Get Kill Clean Reset

Done Adblock

# Introducing the portal interface to Nordugrid (4/4)

File view - Deer Park Alpha 2

File Edit View Go Bookmarks Tools Help

http://nordgrid.ntnu.no/gridportal/ViewFilesPage

File view

the **NTNU**  
Application Portal

High Performance Computing Project

Information Authorization Session Preferences Create Manage About...

**Downloaded job files**

Type	File	Size	Last accessed	Last modified
	gmlog	4096	Sat Aug 13 18:22:23 2005	Sat Aug 13 18:22:32 2005
	f.m	68	Sat Aug 13 18:22:13 2005	Sat Aug 13 18:22:14 2005
	run.sh	67	Sat Aug 13 18:22:15 2005	Sat Aug 13 18:22:15 2005
	stdout.txt	10659030	Sat Aug 13 18:22:16 2005	Sat Aug 13 18:22:18 2005
	stderr.txt	0	Sat Aug 13 18:22:19 2005	Sat Aug 13 18:22:19 2005
	f.log	10658674	Sat Aug 13 18:22:20 2005	Sat Aug 13 18:22:22 2005

View Download Download all (tar.gz) Back Reset

Done Adblock

# Assessing the portal interface

## Advantages:

- Intuitive, easy-to-understand interface
- No memorizing necessary, all options are displayed
- Not restricted to Unix, easier for Windows users

Conclusion: **Ideal for the casual user?**

## Drawbacks:

- Inflexible (web interface does not provide the full array of command line switches)
- Inefficient with extensive use
- Unsuitable for large data sets (more on this later)

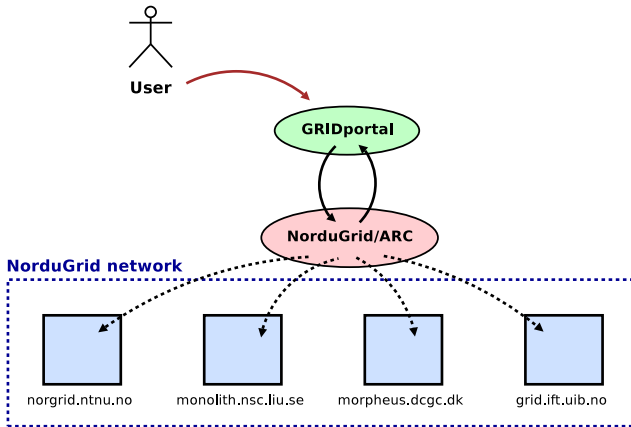
Conclusion: **Sub-par for the "power user"**

# Outline

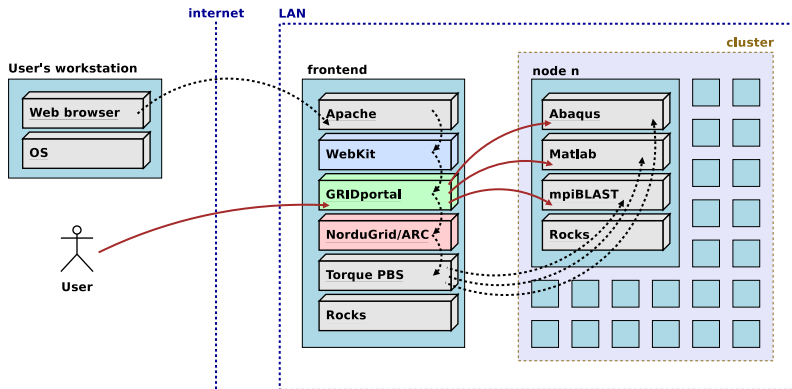
- 1 Motivation
  - The command line interface
  - A proposed solution
  - The portal interface
- 2 **Design**
  - **Relation to Nordugrid/ARC**
  - Authentication mechanism
- 3 Applications on a GRID portal - BLAST
  - BLAST - an introduction
  - BLAST at [norgrid.ntnu.no](http://norgrid.ntnu.no)



# The top level perspective



# A real world example – norgrid.ntnu.no



# Outline

- 1 Motivation
  - The command line interface
  - A proposed solution
  - The portal interface
- 2 Design
  - Relation to Nordugrid/ARC
  - **Authentication mechanism**
- 3 Applications on a GRID portal - BLAST
  - BLAST - an introduction
  - BLAST at [norgrid.ntnu.no](http://norgrid.ntnu.no)

# Problem description

Nordugrid requires the following steps to be completed before a user can gain access to the network:

- 1 The user must create a user certificate
- 2 The certificate must be signed by a Certificate Authority
- 3 The user must be accepted into a Virtual Organization
- 4 The user must generate a user proxy for every session

So how do we combine this with a web portal?

# Proposed solution – myProxy to the rescue

We deploy a client application for download to:

- 1 Create a certificate
- 2 Mail certificate for signing
- 3 Register certificate with a myProxy server (a certificate store)

For every session:

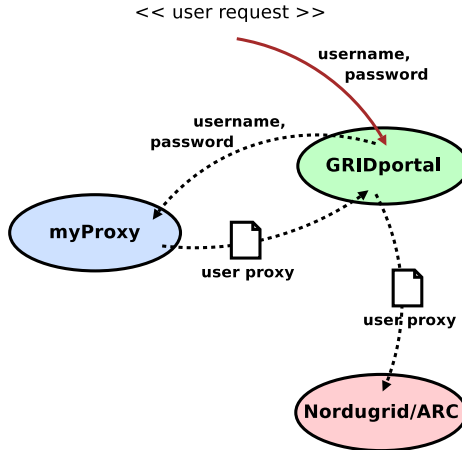
- 1 The user logs in with a username/password, which is passed to the myProxy server
- 2 The portal receives a user proxy and passes it onto ARC

# myProxy: a short description

**Q.** So what is this myProxy thing?

**A.** myProxy is a certificate store, which can store user certificates in a "safe place". Since we wish to relieve the user of the burden of creating a user proxy for every session (as is the case with the command line interface), we transfer the responsibility of storing the certificate onto myProxy. The portal can then query myProxy for a user proxy whenever needed.

# Authentication at a glance



# Outline

- 1 Motivation
  - The command line interface
  - A proposed solution
  - The portal interface
- 2 Design
  - Relation to Nordugrid/ARC
  - Authentication mechanism
- 3 Applications on a GRID portal - BLAST
  - **BLAST - an introduction**
  - BLAST at norgrid.ntnu.no



# BLAST demystification – a short description

## BLAST

- 1 compares biological sequences (written as text strings),
- 2 and yields results which describe the alignment between the sequences (the strings).

BLAST: `<http://www.ncbi.nlm.nih.gov/BLAST/>`

# BLAST demystification – an example

The two sequences:

- 1 a gene sequence from a specimen from the laboratory
- 2 a set of gene sequences from a known bacteria disease

The specimen sequence is compared to every sequence in the bacteria and for every alignment match (above a given threshold), a match is returned, along with a match score.

Depending on the results, there is something to be said for the presence of a sequence known in a common bacteria disease, in a specimen we take from a patient's blood.

# BLAST vs speed – an N:M problem

A typical BLAST query involves comparing

- 1 many specimen sequences (anything from one sequence to millions of sequences)
- 2 to a sizeable database of sequences (e.g. 4GB)

The BLAST algorithm, comparing sequences one by one, is characterized as *embarrassingly linear*, so a speed boost could be possible through symmetric processing.

# The solution: mpiBLAST

mpiBLAST, built with the Message Passing Interface (MPI), is a parallelized flavor of BLAST, designed for use on a cluster. It

- 1 divides the database into equal segments,
- 2 distributes each segment onto a node,
- 3 performs BLAST search on each node in parallel,
- 4 and merges the results from each node into a common result set.

# Evaluating mpiBLAST

*"Database segmentation yields near linear speedup of BLAST in most cases and super-linear speedup in low memory conditions."*

The Design, Implementation, and Evaluation of mpiBLAST

A. Darling, L. Carey, and W. Feng

ClusterWorld Conference & Expo in conjunction with the 4th International Conference on Linux Clusters: The HPC Revolution 2003, San Jose, CA, June 2003.

# Outline

- 1 Motivation
  - The command line interface
  - A proposed solution
  - The portal interface
- 2 Design
  - Relation to Nordugrid/ARC
  - Authentication mechanism
- 3 Applications on a GRID portal - BLAST
  - BLAST - an introduction
  - **BLAST at norgrid.ntnu.no**

# Creating a BLAST job with GRIDportal

The screenshot shows a web browser window titled "BlastJobPageblastn - Deer Park Alpha 2" with the URL "http://norgrid.ntnu.no/gridportal/BlastJobPageblastn". The main content area is titled "Create a blastn job" and contains the following fields and options:

- \* CPU time (min): 30
- \* Job name: blastnJob
- Email notification: numerodix@gmail.com
- \* Enter query sequence(s): (empty text area)
- Sequences input file: input.seq (with a "Browse..." button)
- Location on query sequence [1,400]: 0,100
- Database: refseq\_genomic (dropdown menu)
- Filter query sequence: Yes (dropdown menu)
- Expectation value (E): 10.0 (dropdown menu)
- Word Size: 11 (dropdown menu)
- HTML format on output: No (dropdown menu)
- Number of alignments reported: 250 (dropdown menu)
- Number of scores/descriptions reported: 500 (dropdown menu)
- Restrict search of database to list of GI's: (empty text area)
- Advanced parameters: (empty text area)

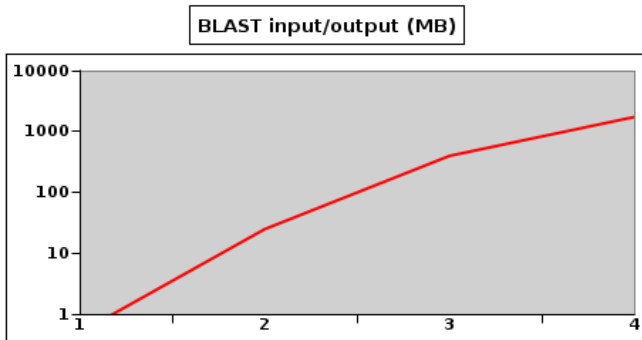
Fields marked with an asterisk (\*) must be filled in.  
Note that query sequences must be supplied either in the input box or in an input file.  
Defaults displayed on this page reflect the defaults of NCBI BLAST.

Buttons: Create, Reset

Status: Done (bottom left), Adblock (bottom right)

# BLAST with large data sets

Depending on the number of matches in a BLAST query, the result file may become rather large.





# GRIDportal vs large data sets

The portal is web-based, uploading/downloading of input/output files is over HTTP. On slow links, the transfer is likely to suffer from bad connectivity, network congestion etc. *And there is no resume function for interrupted transfers.*

Thus, heavy BLAST users are better off using the command line interface. **In general, the portal is well suited for jobs with heavy processing but small input & output files.**

# Links

- GRIDportal project website  
<<http://gridportal.dynalias.org/>>
- GRIDportal deployment site  
<<http://norgrid.ntnu.no/gridportal/>>

Thank you for your attention!