

Cluster-Benutzerverwaltung: erste Schritte mit Shibboleth

Steffen Hau, Reinhard Mayer, Sabine Richling, Tobias Kienzle,
Heinz Kredel

Rechenzentrum Universität Mannheim
Universitätsrechenzentrum Heidelberg

ZKI AK Supercomputing, Düsseldorf
20./21. September 2012



- 1 bwGRiD
 - in Baden-Württemberg
 - in Mannheim/Heidelberg
- 2 Cluster-Benutzerverwaltung Mannheim/Heidelberg
 - Administration
 - Benutzerverwaltung bisher
 - Föderative Authentifizierung und Autorisierung
 - Zugang über Shibboleth
- 3 Zusammenfassung und Ausblick
- 4 Demo

Was ist bwGRiD?

- Virtuelle Organisation (VO) bwGRiD
 - D-Grid Projekt der Universitäten in Baden-Württemberg
 - VO bwGRiD ist offen für alle Mitglieder der beteiligten Universitäten und deren Partner in wissenschaftlichen Projekten
- bwGRiD Ressourcen
 - Rechencluster an 8 Standorten
 - Zentraler Speicher in Karlsruhe
- bwGRiD Ziele
 - Nachweis der Funktionalität und des Nutzens von Grid-Konzepten für das Hochleistungsrechnen in Baden-Württemberg
 - Überwindung von Organisations-, Sicherheits- und Lizenz-Problemen
 - Entwicklung von neuen Cluster- und Grid-Anwendungen



Bundesministerium
für Bildung
und Forschung



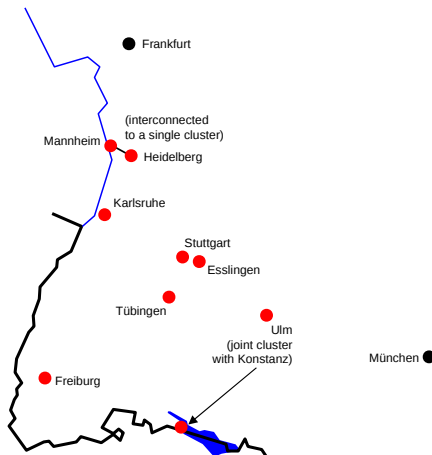
Baden-Württemberg

Rechencluster:

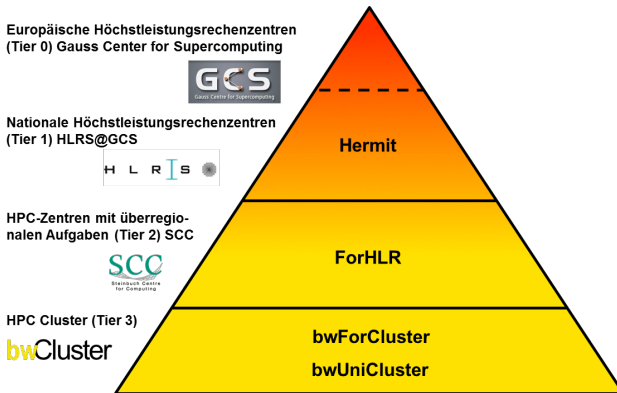
- Mannheim/Heidelberg:
280 Knoten
Direkte Kopplung
- Karlsruhe: 140 Knoten
- Stuttgart: 420 Knoten
- Tübingen: 140 Knoten
- Ulm (Konstanz): 280 Knoten
Hardware in Ulm
- Freiburg: 140 Knoten
- Esslingen: 180 Knoten
neuere Hardware

Zentraler Speicher:

- Karlsruhe:
128 TB (mit Backup)
256 TB (ohne Backup)



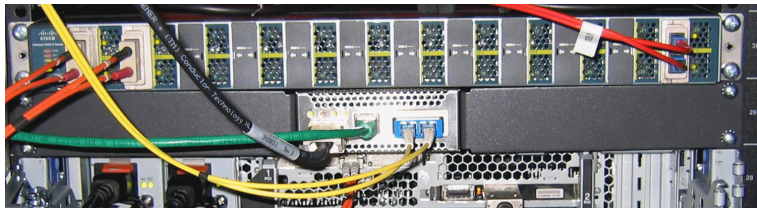
HPC Konzept von Baden-Württemberg:



- Ende 2012 ist die bwGRiD Hardware 5 Jahre alt
- Ersatzbeschaffungen angestrebt in 2012 bis 2015

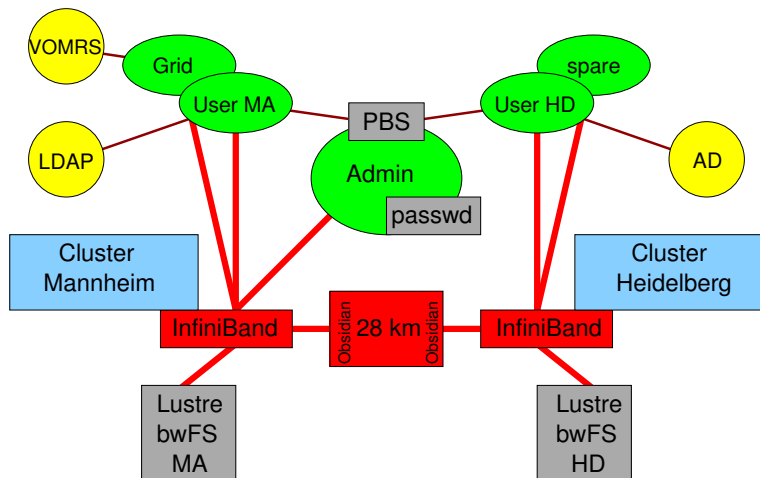
Kopplung der bwGRiD Cluster MA/HD

- Antrag in 2008
- Beschaffung und Aufbau bis Mai 2009
- in Betrieb seit Juli 2009
- Infiniband über Glasfaser: Obsidian Longbow



- Latenz ist hoch:
 $145 \mu\text{sec} = \text{Lichtlaufzeit (28 km)} + 2 \mu\text{sec lokale Latenz}$
- Bandbreite wie wartet: etwa 930 MB/sec
Lokale Bandbreite 1200-1400 MB/sec

bwGRiD Cluster Mannheim/Heidelberg



140 nodes cluster

InfiniBand Network

Login/Admin Server

Directory service

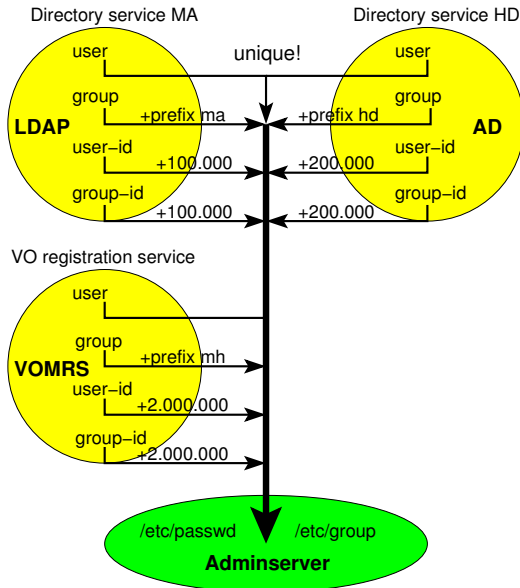
Storage

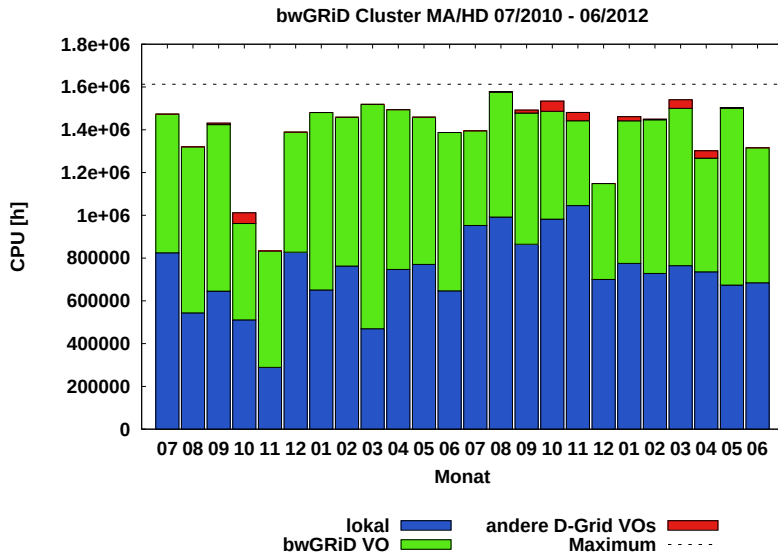
- Aufgaben des Admin-Servers:
 - DHCP+PXE Service für die Knoten
 - NFS Export des Root-Filesystems für die Knoten
 - NFS Verzeichnis für Software-Pakete
 - Batch System und Job Scheduler
- Administration der Knoten:
 - über Shell-Skripte
 - mit IBM Management-Modul (CLI und Web-GUI)
- Jobmanagement:
 - Kopplung (hohe Latenz, begrenzte Bandbreite)
 - bietet genug Bandbreite für Zugriff auf Storage-systeme
 - reicht nicht für MPI Jobs mit hoher Kommunikation
 - Jobs laufen auf Knoten in HD **oder** in MA
 - Vor der Kopplung:
 - In Mannheim: hauptsächlich 1-Knoten-Jobs → freie Knoten
 - In Heidelberg: viele MPI-Jobs → lange Wartezeiten
 - Nach der Kopplung: Bessere Auslastung

Gemeinsame Benutzerverwaltung

- Benutzer bekommen exklusiven Zugang zu den Knoten
 - Kennungen und User-IDs müssen eindeutig sein
 - Direkte Verbindung zu PBS und Autorisierung über ein PAM-Modul
- Authentifizierung am Zugangsrechner
 - gegen LDAP (MA) und AD (HD)
 - mit D-Grid Zertifikat
- Aktivierung von Kennungen
 - Setzen eines Attributs im LDAP bzw. AD (Authentifizierung)
 - Update der Benutzerdaten auf dem Cluster (Autorisierung)
- Erstellung der Dateien für die Benutzerverwaltung
 - Prefix "ma", "hd" oder "mh" für Gruppen
 - Offsets für Group-IDs
 - Offsets für User-IDs
 - Kennungen von MA und HD müssen verschieden sein!

Erstellung der Dateien für die Benutzerverwaltung





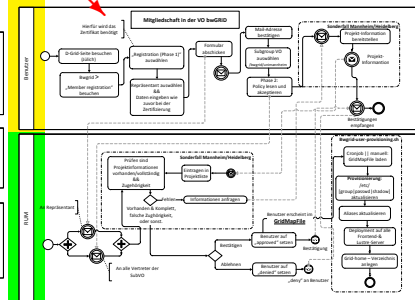
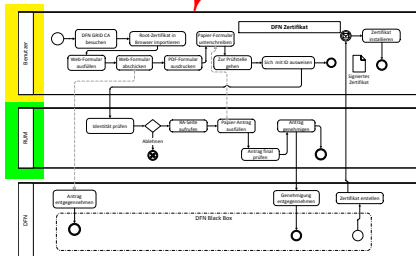
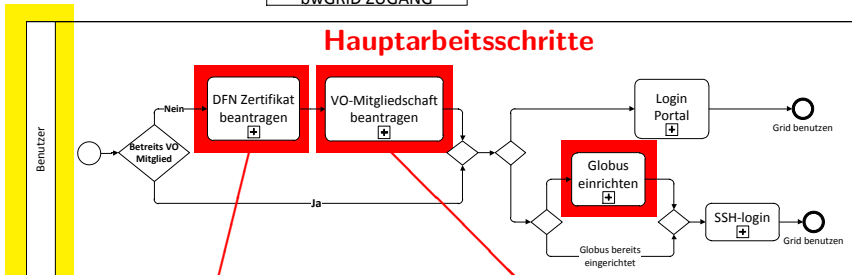
Warum neuer föderativer Zugang?

- D-Grid läuft eventuell Ende des Jahres 2012 aus (keine VO Dienste mehr)
- Grid-Zugang sehr komplex (siehe Modellierung)
 - Grid-Zertifikat und VO Registrierung
 - Viele Arbeitsschritte für Nutzer und Rechenzentrum
 - Einarbeitung in die Nutzung einer Middleware
- Zugang weiterhin Baden-Württemberg weit gewünscht
- Neubeschaffungen im Rahmen des HPC Konzeptes
- Föderativer Zugang als einfachere Alternative für
 - ssh-Zugang
 - bwGRiD-Portal
- keine zentrale Benutzerverwaltung notwendig
- keine Eingriffe in die Organisation der Benutzerverwaltung der Universitäten
- die Föderation der Benutzerverwaltungen stellt nur die Identitäten sicher

Modellierung Grid-Zugang

bwGRiD ZUGANG

Hauptarbeitsschritte



Benutzer

Rechenzentrum

bwGRiD “Ergänzende Maßnahmen”

Verbesserung der Nutzbarkeit und Weiterentwicklung des Angebots von bwGRiD in verschiedenen Arbeitspaketen in Zusammenarbeit mit bwIDM und bwLSDF

Ein Ziel

Vereinfachung der Anmelde-Prozeduren

Aufgaben

- Machbarkeit und Konzept
- Prototyp in HD-MA lauffähig
- allgemeine Verfügbarkeit
- Koordination mit bwIDM und DFN
- Lösungen: Shibboleth und eduroam

Untersuchung und Evaluation im Rahmen von bwIDM in 2011
(Nussbaumer, Simon aus KA, Grandpre, Becker aus KO, MA):

- Eduroam (Radius): hat keine / zu wenige Attribute über die Benutzer, zum Beispiel ist Mitarbeiter und Student nicht unterscheidbar
- Moonshot: zur Zeit noch nicht produktiv einsetzbar, große Eingriffe in die Infrastruktur notwendig (spezielles ssh)
- Shibboleth: Attribute vorhanden, kleine Eingriffe in die Infrastruktur notwendig (spezielles PAM_ECP, NSSWITCH), evtl. Abstriche beim Datenschutz wegen des SSHD

Ergebnis

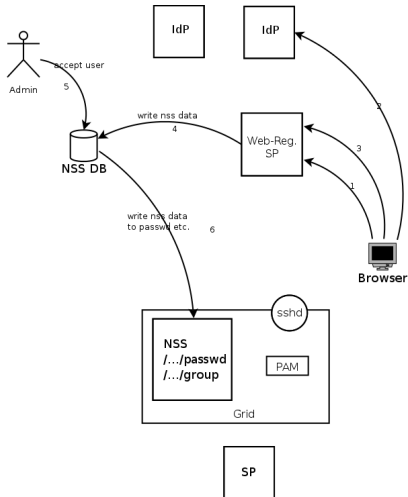
Shibboleth ist zur Zeit am besten geeignet.

Konzept

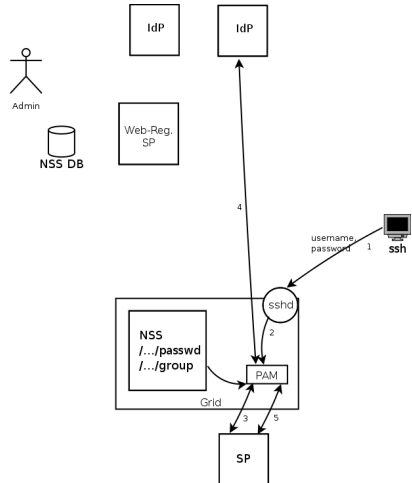
- Anforderung: es soll alles Datenschutz-konform sein
- Service Provider (SP), Diensteanbieter wie zum Beispiel bwForCluster, bwUniCluster, bwGRiD
- Identity Provider (IDP), Identitäten garantiert durch Benutzerverwaltungen der Universitäten
- Benutzer muss dem IDP der Heimateinrichtung vertrauen
 - auch dem SSHD muss vertraut werden
 - Einmal-Passwörter oder SSH-Keys als Alternative
- Beantragung eines Shibboleth Zugangs zu einem Dienst, wie z.B. bwGRiD, über ein Web-Formular
- keine weiteren Formulare und Unterschriften
- Dienst kann Benutzer ablehnen oder Bedingungen stellen
- Benutzer kann die Bedingungen ablehnen oder einwilligen

Zugang über Shibboleth

Registrierung



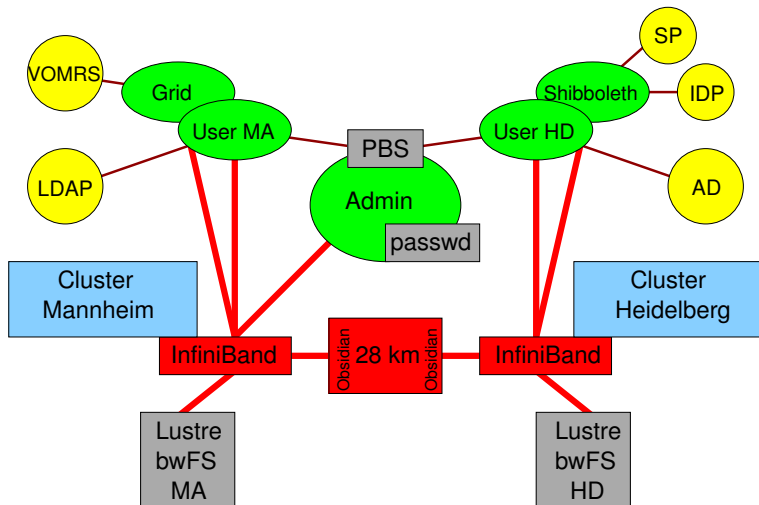
ssh-Zugang



Aufgaben bei der Realisierung

- Ersatz Login-Server wird Shibboleth Zugangsrechner
- ssh-Zugang mit `kennung@einrichtung` unter Verwendung und Anpassung des PAM_ECP Moduls aus Karlsruhe
- Aufbau einer Datenbank, die mit dem Web-Portal verbunden ist und die, die für den Betrieb notwendigen Infos enthält
- Erzeugung von `passwd.shib`, `group.shib` und `aliases.shib` aus der Datenbank
- Überarbeitung des Deployments damit diese Dateien entsprechend integriert werden
- Kann `kennung@einrichtung` auch als UID verwendet werden?
 - wenn dies nicht möglich ist, muss ein Mapping zu 'normalen' Kennungen entwickelt werden

bwGRiD Cluster Mannheim/Heidelberg mit Shibboleth



140 nodes cluster

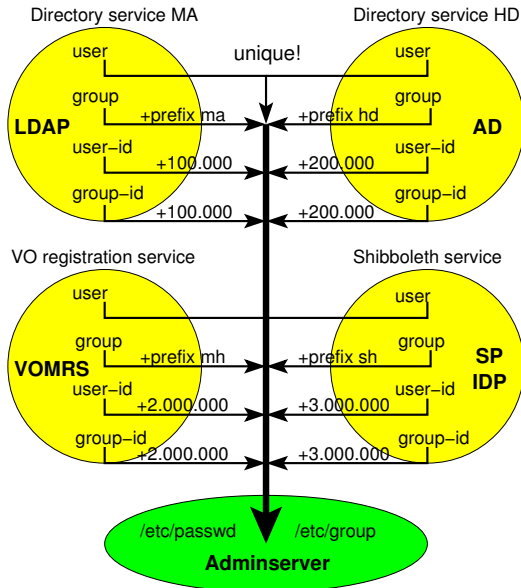
InfiniBand Network

Login/Admin Server

Directory service

Storage

Erstellung der Dateien für die Benutzerverwaltung



Tests

- Aufsetzen des Shibboleth bwGRiD SP
- Anpassen des PAM_ECP Moduls
- Aufsetzen der Shibboleth IDPs der Universitäten

Probleme

- @ in UID `kennung@uni-heidelberg.de` funktioniert nicht, wenn das @ eine Bedeutung für die Anwendung hat, wie zum Beispiel bei PBS `kennung@host`
- _ in UID `kennung_uni-heidelberg.de` würde funktionieren, aber evtl. Probleme mit Datenschutz

⇒ NSS Modul muss entwickelt werden

Erste Ergebnisse

- Generierung und Deployment der neuen Kennungen auf die Zugangsrechner und die Lustre Server funktioniert
- Zugang funktioniert mit `kennung@uni-heidelberg.de` und `kennung@uni-mannheim.de`
 - PAM_ECP Modul authentifiziert gegen Shibboleth IDP der jeweiligen Einrichtung
 - NSS Modul bildet `kennung@einrichtung` auf neue Kennungen `sbwxxxxxx` ab
 - Abbildung der Kennungen `sbwxxxxxx` auf reale Email-Adressen in `/etc/aliases`
 - Verwendung von Lustre funktioniert
 - Verwendung von PBS einschließlich Email-Benachrichtigung funktioniert ebenfalls

Zusammenfassung und Ausblick

- InfiniBand Cluster-Kopplung über 28 km ist stabil im Betrieb
- bessere Auslastung durch Kopplung sowie vereinfachte Administration und Benutzerverwaltung
- Ziel: vereinfachter Zugang für die Benutzer
 - Entwicklung von PAM_ECP im Rahmen von bwIDM von M. Simon, Karlsruhe
 - Aufsetzen der Shibboleth Infrastruktur: IDP und SP
 - Entwicklung eines NSS Moduls zum Mapping auf generische Kennungen
 - Proof of Concept für bwGRiD Cluster Heidelberg-Mannheim
 - Freigabe für bwGRiD Nutzer aus Heidelberg-Mannheim geplant für Ende 2012
- Produktiver Einsatz im bwUniCluster geplant für Anfang 2013

- Nussbaumer, Simon (KIT, Karlsruhe)
- Grandpre, Becker (Universität Konstanz)
- Mosch (Universität Ulm)
- Bogus, Skowronek, Maylein (Universität Heidelberg)
- Hänger, Wagner, Pfeffer (Universität Mannheim)

- 1 Select IDP for registration
- 2 Registration login
- 3 Registration request
- 4 Approve request pending
- 5 Approve request processing
- 6 Approval status overview
- 7 User provisioning in the cluster
- 8 SSH login via Shibboleth
- 9 Job submission with PBS

Select IDP for registration



Home Organisation auswählen



Um auf einen Dienst auf dem Server 'bwgrid-auth-test3.rz.uni-

[Über AAI](#) | [FAQ](#) | [Hilfe](#) | [Datenschutz](#)

'mannheim.de' zuzugreifen, wählen oder suchen Sie bitte die Organisation, der Sie angehören.

Uni Mannheim - TestIDP



Auswählen

☐ Auswahl für die laufende Webbrowser Sitzung speichern.

[SWITCH](#) erbringt innovative, einzigartige Internet-Dienstleistungen für die Schweizer Hochschulen und Internetbenutzer.



Example Login Page

This login page is an example and should be customized. Refer to the [documentation](#).

The web site described to the right has asked you to log in and you have chosen <FILL IN YOUR SITE> as your home institution.

Username:

Password:

bwgrid-auth-test3.rz.uni-mannheim.de

You have asked to login to bwgrid-auth-test3.rz.uni-mannheim.de

Registration request



PAM/ECP Proof of Concept

Registration for bwGRiD

[Home](#)

UEbersicht

Feld	Wert
BW Username	shau@uni-mannheim.de
E-Mail-Adresse	shau@mail.uni-mannheim.de
Heimatinrichtung	uni-mannheim.de

☒ you must accept our [conditions](#)

Register for	
ES	☐
FR	☐
KA	☐
KO	☐
MA/HD	☒
ST	☐
Tue	☐
Ulm	(no request required)
Uni-Cluster	(no request required)

Please enter an abstract of your work here

Ganz ganz tolles Projekt :)

Zugang anlegen: [Register](#)

Approve request pending



PAM/ECP Proof of Concept

Registration for bwGrID

[Home](#)

UEbersicht

Feld	Wert
BW Username	shau@uni-mannheim.de
E-Mail-Adresse	shau@mail.uni-mannheim.de
Heimateinrichtung	uni-mannheim.de

You are registered with the following entries

local username	uid	gid	homedir	scoped username
sbw00005	3000005	3030012	/bwfs/home/shib/sbw00005	shau@uni-mannheim.de

You can connect to the following clusters

Cluster	login permission	request pending
ES	no	no
FR	no	no
KA	no	no
KO	no	no
MA/HD	no	yes
ST	no	no
Tue	no	no
Ulm	yes	no
Uni-Cluster	yes	no

Approve request processing

delete	username	scoped username	abstract	cluster	requested	activate	deactivate
☐	sbw10002	rmayer2@uni-heidelberg.de	DGII-9VuGOP..				
				KA		☐	
				ES		☐	
				Tue		☐	
				Ulm			☐
				FR		☐	
				Uni-Cluster			☐
				KO	yes	☒	
				ST	yes	☒	
				MA/HD	yes		☐
☐	sbw00004	reimayer@uni-mannheim.de	lalal bla..				
				KA		☐	
				ST		☐	
				Tue		☐	
				Ulm			☐
				FR		☐	
				KO		☐	
				Uni-Cluster			☐
				ES	yes	☒	
				MA/HD	yes		☐
☐	sbw00006	shau@uni-mannheim.de	Ganz tolle..				
				KA		☐	
				ST		☐	
				ES		☐	
				Tue		☐	
				Ulm			☐
				FR		☐	
				KO		☐	
				Uni-Cluster			☐
				MA/HD	yes	☒	

save

Approval status overview



PAM/ECP Proof of Concept

Registration for bwGRiD

[Home](#)

UEbersicht

Feld	Wert
BW Username	shau@uni-mannheim.de
E-Mail-Adresse	shau@mail.uni-mannheim.de
Heimateinrichtung	uni-mannheim.de

You are registered with the following entries

local username	uid	gid	homedir	scoped username
sbw00006	3000006	3030012	/bwfs/home/shib/sbw00006	shau@uni-mannheim.de

You can connect to the following clusters

Cluster	login permission	request pending
ES	no	no
FR	no	no
KA	no	no
KO	no	no
MA/HD	yes	yes
ST	no	no
Tue	no	no
Ulm	yes	no
Uni-Cluster	yes	no

User provisioning in the cluster

```
root@intern1:~ - bwGRiD
[root@intern1 ~]# time bwgrid-user-provisioning.sh
Passwd:
Shibboleth added: sbw00006
Aliases:
Shibboleth added: sbw00006
Deploying passwd files (Frontends, Lustre)...
Creating missing home directories for grid users...
Creating missing home directories for shibb users...
sbw00004
sbw00006
sbw10002
Setting Lustre Quotas for new users...
Deploying aliases files (Frontends)...

real    0m19.955s
user    0m5.432s
sys     0m12.799s
[root@intern1 ~]#
```


SSH login via Shibboleth

```
root@intern1:~ - bwGRID root@frbw4:~ - bwGRID
steffen@haurumtop /tmp $ ssh -l shau@uni-mannheim.de zug-hd-2.urz.uni-heidelberg.de
shau@uni-mannheim.de@zug-hd-2.urz.uni-heidelberg.de's password:
***      bwGRID Cluster Mannheim/Heidelberg (Spare Access Node)      ***
* General Information:      http://www.bw-grid.de
* Local Documentation:      http://www.uni-mannheim.de/rum/zs/hpc/bwgrid_cluster
* D-Grid User Support:      https://helpdesk.ngi-de.eu
* E-mail Support MA/HD:      dgrid-support@mailman.uni-mannheim.de
*****

/usr/bin/xauth:  creating new authority file /bwfs/home/shib/sbw00006/.Xauthority

Setting up your ssh key...

...setting BASIC BWGRID environment...
...setting MACHINESPECIFIC environment...

Quota information for your home directory
You have      0 GB in      17 files and directories
[sbw00006@frbw4 ~]$
```

Job submission with PBS

```
root@intern1:~ - bwGRiD                                X root@frbw4:~ - bwGRiD
steffen@haurumtop /tmp $ ssh -l shau@uni-mannheim.de zug-hd-2.urz.uni-heidelberg.de
shau@uni-mannheim.de@zug-hd-2.urz.uni-heidelberg.de's password:
Last login: Wed Sep 19 18:04:37 2012 from haurumtop.rz.uni-mannheim.de
***          bwGRiD Cluster Mannheim/Heidelberg (Spare Access Node)          ***
* General Information:      http://www.bw-grid.de
* Local Documentation:     http://www.uni-mannheim.de/rum/zs/hpc/bwgrid_cluster
* D-Grid User Support:     https://helpdesk.ngi-de.eu
* E-mail Support MA/HD:    dgrid-support@mailman.uni-mannheim.de
-----
...setting BASIC BWGRID environment...
...setting MACHINESPECIFIC environment...

Quota information for your home directory
You have      0 GB in      18 files and directories
[sbw00006@frbw4 ~]$ qsub -I -lnodes=2,walltime=05:00
qsub: waiting for job 325649.intern1 to start
qsub: job 325649.intern1 ready

...setting BASIC BWGRID environment...
...setting MACHINESPECIFIC environment...

Quota information for your home directory
You have      0 GB in      18 files and directories
[sbw00006@n060212 ~]$ cat $PBS_NODEFILE
n060212
n060205
[sbw00006@n060212 ~]$ █
```