

# Globus Toolkit Installation

QCDGrid Sysadmin course

23 March 2006

---

Jeremy Nowell  
EPCC

[jeremy@epcc.ed.ac.uk](mailto:jeremy@epcc.ed.ac.uk)

- Introduction to Globus
  - Globus Components
  - Security
- Globus Installation
  - From source
  - Virtual Data Toolkit (VDT)
- Post Installation Configuration
- Authorisation
- Optional components – RLS, MySQL
- Future Developments
- Exercise

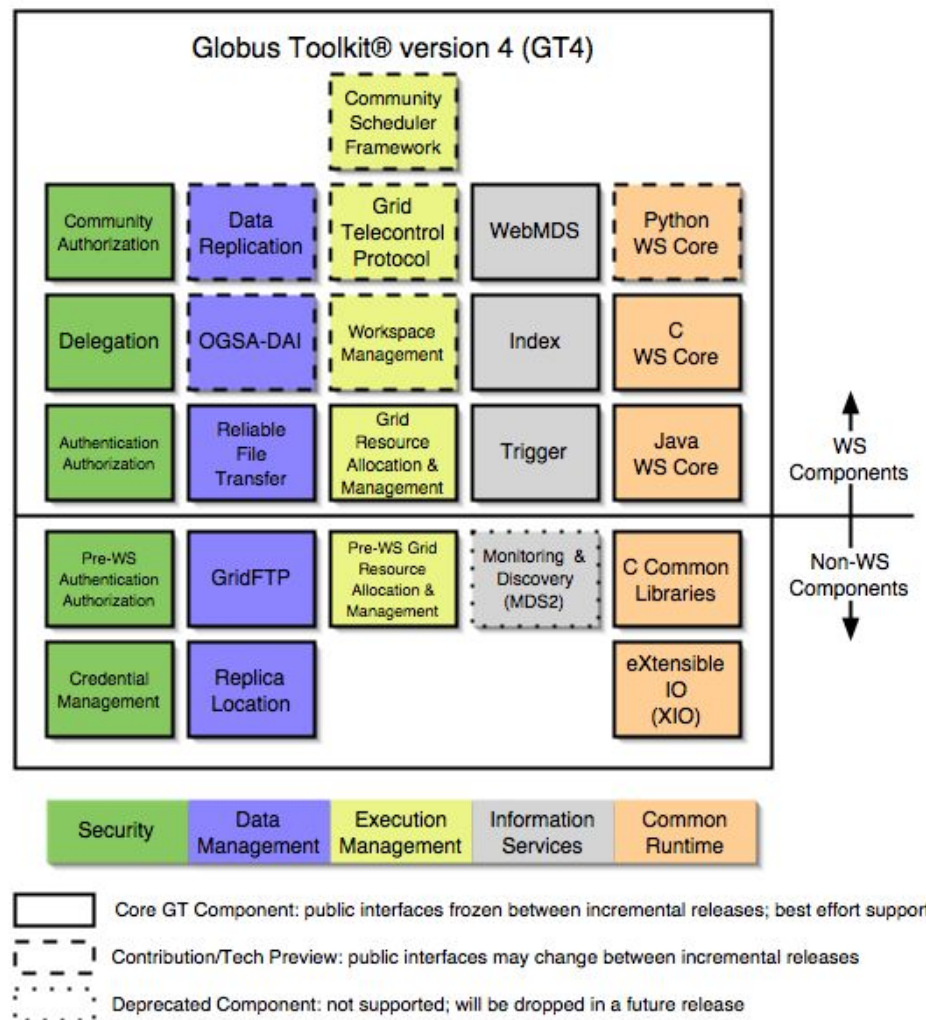
- The Globus Toolkit provides a set of services to enable the secure sharing of resources across corporate, institutional and geographic boundaries

- Data Management
- Execution Management
- Information Services
- Security

- “PreWS” (GT2) and “WS” (GT4)

## Components

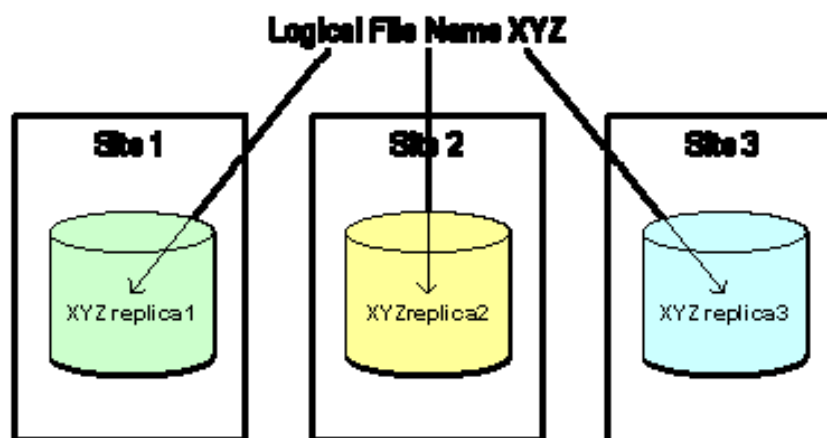
- QCDGrid currently based around GT2 (below line in diagram)



- Services to allow secure execution of processes on remote resources
  - Job submission, management, cancellation
- At its most basic can be seen as providing a common interface to heterogeneous batch schedulers
  - Batch scheduler may mean here “fork a process”
- GRAM – Grid Resource Allocation and Management is main component

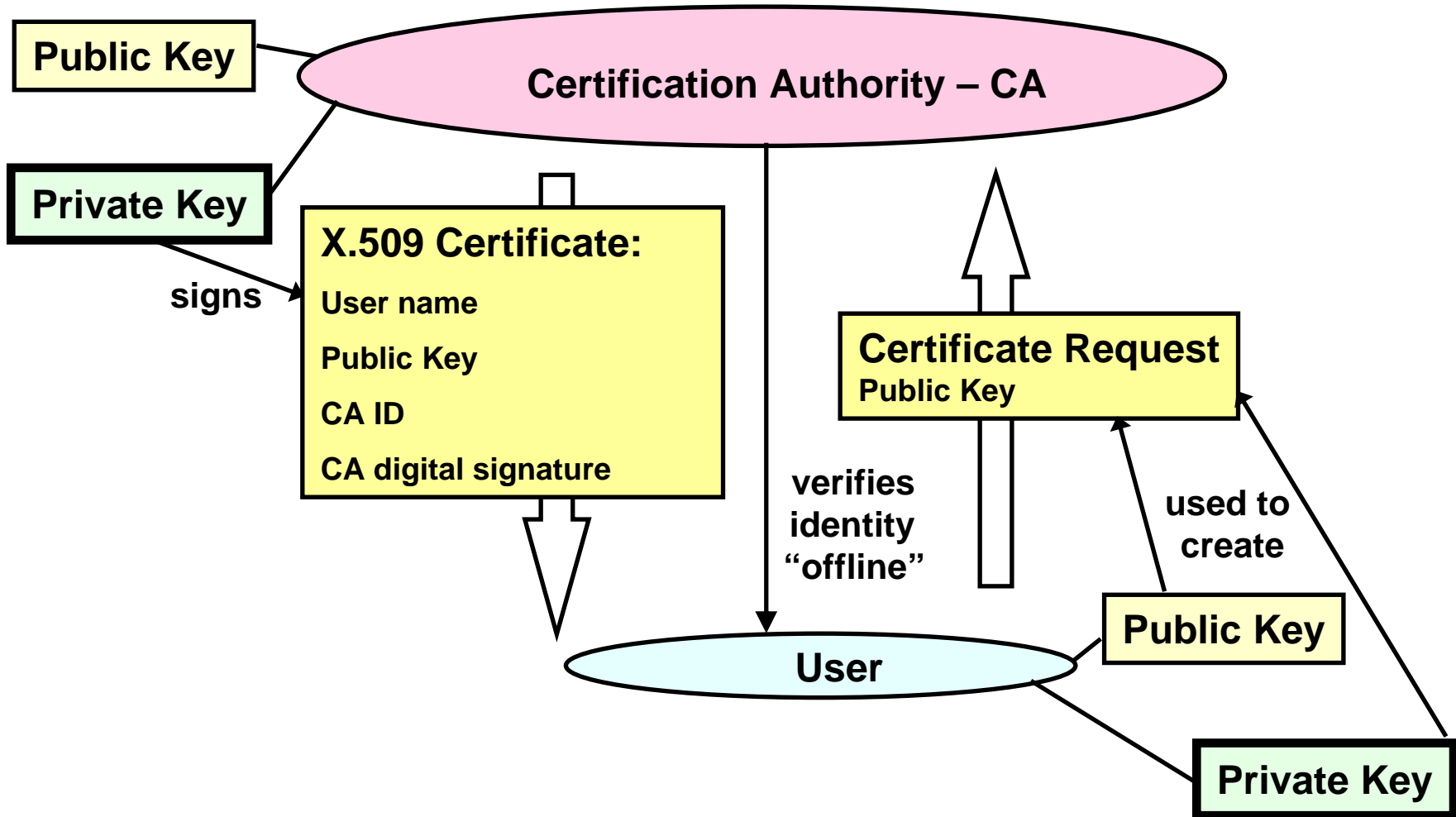
- GridFTP
  - GSI compliant implementation of FTP protocol
  - Optimised for high-bandwidth WANs
  - High performance
    - Multiple TCP streams
    - Striping
    - Large files
  - Third party transfer support
  - Interoperable between GT2 and GT4

- Replica Location Service (RLS)
  - Distributed registry and discovery service
  - Maps *logical file name* to *physical file name*
  - Based on relational database technology (MySQL)
  - Allows hierarchical catalogues to be created



- Authentication:
  - Is the user who they say they are?
- Authorisation:
  - Is the user allowed to do what they are requesting?
- Message-level security:
  - Integrity
  - Privacy
- Single sign-on:
  - Minimise security overheads in an ongoing client-service dialog
- Delegation:
  - Delegate privileges/rights to (other) services
- Based around PKI and X.509 certificates

- A key is a very large number!
- Key pairs:
  - Public key – available to the world
  - Private key – your-eyes-only
    - Stored in an encrypted file protected with a password
- Encryption and decryption:
  - Approach 1 – Sender Encrypts with Receiver’s Public Key:
    - ReceiverPublicKey(Message) => EncryptedMessage
    - ReceiverPrivateKey(EncryptedMessage) => Message
  - Approach 2 – Sender Encrypts with Sender’s Private Key:
    - SenderPrivateKey(Message) => EncryptedMessage
    - SenderPublicKey(EncryptedMessage) => Message



- Used for identification and authentication
- Guarantees that:
  - Your public key is indeed yours
  - You are who you claim to be
- Host certificates:
  - AKA server certificates
  - Required if a server must perform operations on behalf of a user
  - For example, Globus Toolkit job submission (GRAM) and file transfer (GridFTP)
  - User must authenticate the host
- Certification Authorities (CA):
  - Globus – accessible via Globus Toolkit tools (only for testing)
  - UK e-Science: <http://ca.grid-support.ac.uk/>

- Several choices available:
  - Build from source using globus provided packages
    - Can be a very long and tedious process, depending on machine
    - Hard to debug problems
    - My recommendation for anything other than Linux/x86 (until GT4 release)
  - Install Globus built binaries
    - Selected platforms only
    - Need to install updates manually
  - Install binaries using third party tools
    - Added layer of packaging tools
    - Can be easier and quicker
    - Help with configuration

- Globus packaged using GPT – Grid Packaging Tool
- Source or binary packages
  - Flavour describes compiler and build type, eg gcc32dbg, vendorcc64pthr
  - Need to decide on client, server or sdk packages

```
export GPT_LOCATION=/opt/local/globus/gpt3.0.1
./build_gpt
export GLOBUS_LOCATION=/opt/local/globus/globus2.4.3
$GPT_LOCATION/sbin/gpt-build globus-data-management-client-2.4.3-
    src_bundle.tar.gz vendorcc64
$GPT_LOCATION/sbin/gpt-postinstall
$GLOBUS_LOCATION/setup/globus/setup-gsi
```

- GT4 adds extra Makefile to hide GPT

- <http://vdt.cs.wisc.edu/>
- Collection of grid middleware for easy installation and configuration
  - Globus
  - Condor
  - MySQL
  - Useful utilities
- Originally created by iVDGL and GriPhyN, now used by LCG, PPDG projects
- Includes patches to Globus as contributed by above projects and others
- Binaries for various different Linux platforms

- Packaging in VDT is based on pacman
  - <http://physics.bu.edu/~youssef/pacman/>
  - Python based
  - Higher level than GPT/RPM
  - Helps with configuration using setup scripts that run when packages are installed
  - Uses “caches” as download locations
  - More features not used in a basic Globus install

- Simple installation of GT2.4.3
  - Can install either as root or normal user
  - Need to answer questions about licenses during install

```
# Create a directory for VDT
$ mkdir vdt
$ cd vdt
# Get pacman
$ wget http://physics.bu.edu/pacman/sample_cache/tarballs/pacman-2.129.tar.gz
# Install pacman
$ tar zxvf pacman-2.129.tar.gz
$ export PATH=`pwd`/pacman-2.129:$PATH
# Point Pacman at the correct VDT cache
$ pacman -cache:http://vdt.cs.wisc.edu/vdt_124_cache
# Install any one or more of these VDT packages with pacman
$ pacman -get VDT:Globus
# Setup the environment (use setup.sh / setup.csh depending on your shell)
$ source ./setup.sh
$ pacman -get VDT:Globus-2-Core
$ pacman -get VDT:Globus-RLS
```

- Setup globus services to run using xinetd/inetd
- (Configure globus job-manager if using batch system)
- Security
  - Install CA certificates as required
  - Obtain and install host certificates if running services
  - Configure Globus authorisation
  - Firewall configuration

- Globus services need adding to system configuration files
  - GridFTP, Globus Gatekeeper, (MDS)
- Add port numbers to `/etc/services`

```
gsigatekeeper 2119/tcp # Globus Gatekeeper
gsiftp        2811/tcp # GridFTP
```

- Add services to `xinetd/inetd`

```
service gsigatekeeper {
    socket_type = stream
    protocol = tcp
    wait = no
    user = root
    env = LD_LIBRARY_PATH=GLOBUS_LOCATION/lib
    env = GLOBUS_TCP_PORT_RANGE=50000,52000
    server = GLOBUS_LOCATION/sbin/globus-gatekeeper
    server_args = -conf GLOBUS_LOCATION/etc/globus-gatekeeper.conf
    disable = no
}
```

- **xinetd - GridFTP**

```
service gsift {
    instances = 1000
    socket_type = stream
    wait = no
    user = root
    env = LD_LIBRARY_PATH=GLOBUS_LOCATION/lib
    env = GLOBUS_TCP_PORT_RANGE=50000,52000
    server = GLOBUS_LOCATION/sbin/in.ftpd
    server_args = -l -a -G GLOBUS_LOCATION
    log_on_success += DURATION USERID
    log_on_failure += USERID
    nice = 10
    disable = no
}
```

- Done by VDT if installed as root user
- Different for GT4 services – see user doc

- `/etc/grid-security/` is default location for everything certificate related in Globus
  - Can also be `$GLOBUS_LOCATION/etc/share/certificates/`
- Subdirectory `certificates/` contains root certificates of all the Certificate Authorities (CAs) that you want to trust – VDT will install a selection of these for you if installed as root
- UK e-Science CA certificate available as `01621954.0` from <http://www.grid-support.ac.uk>, also need signing policy file, `01621954.signing_policy`
- Need host certificate and private key, should have following permissions:

```
$ ls -l /etc/grid-security
drwxr-xr-x  2 root    other      512 Apr  7  2004 certificates
-r--r--r--  1 root    other     2653 May 26  2005 hostcert.pem
-r-----  1 root    other      891 May 26  2005 hostkey.pem
```

- Need to open ports in firewalls for Globus to talk to outside world
  - 2119 for Gatekeeper
  - 2811 for GridFTP
  - (2135 for MDS)
  - Port for RLS if appropriate
  - GLOBUS\_TCP\_PORT\_RANGE, also called the ephemeral port range
    - GridFTP data channels, job management processes
    - Required by both servers and clients
    - Can be anything you like, although 50000-52000 and 65000-65255 have become common
    - Need environment variable \$GLOBUS\_TCP\_PORT\_RANGE
- <http://www.globus.org/toolkit/security/firewalls/>

- grid-mapfile
  - Maps X.509 Certificate Distinguished Names to local user account

```
$ more /etc/grid-security/grid-mapfile  
"/C=UK/O=eScience/OU=Edinburgh/L=NeSC/CN=jeremy nowell" jeremy
```

- Quotes required when spaces are present
- Difficult to manage for more than a few users or machines

- Virtual Organisation Management Service
- QCDGrid version
  - LDAP based server contains certificate list of approved users
  - EDG mkgridmap command runs daily and retrieves list
  - Maps users to “pool” accounts, typically a list such as user001 etc
  - Need to configure mkgridmap to talk to QCDGrid VOMS server
  - See QCDGrid documentation for further details

- If you are setting up a control node then need to install RLS
  - Replica Location Service
    - Needs backend database, MySQL by default
    - Easiest is to install both MySQL and RLS using VDT

```
$ pacman -get MySQL
```

```
$ pacman -get Globus-RLS-Server-Setup-MySQL
```

- Answer y to all questions!
- Can also install manually, or use an already existing database if required – see VDT and Globus documentation for details

- GT4
  - Web services components
  - All “GT2” services still present
  - More stable and tested, better documentation
  - More binary builds available
  - Smoother process
  - May need extra port in firewall for Web Services
- Should be available in VDT

- QCDGrid recommendation for Globus installation is to use VDT
- For a server need to configure
  - Gatekeeper
  - GridFTP
  - Security
  - Authorisation
  - Firewall
  - Optionally RLS

- Simple client installation of GT2.4.3 into home directory

```
# Create a directory for VDT
$ mkdir vdt
$ cd vdt
# Get pacman
$ wget http://physics.bu.edu/pacman/sample_cache/tarballs/pacman-2.129.tar.gz
# Install pacman
$ tar zxvf pacman-2.129.tar.gz
$ export PATH=`pwd`/pacman-2.129:$PATH
# Point Pacman at the correct VDT cache
$ pacman -cache:http://vdt.cs.wisc.edu/vdt_124_cache
# Install any one or more of these VDT packages with pacman
$ pacman -get VDT:Globus
```

- Now go for coffee while install progresses!

- By now VDT Globus should have installed, along with CA certificates. Now need to
  - Install missing components
  - Install user X.509 certificate
  - Test everything is working

```
# Setup the environment (use setup.sh / setup.csh depending on your shell)
$ source ./setup.sh
$ pacman -get VDT:Globus-2-Core
$ pacman -get VDT:Globus-RLS
# Put certificate and key into expected place:
$ mkdir ~/.globus
$ cd ~/.globus
$ scp user@host:usercert.pem .
$ scp user@host:userkey.pem .
# Test certificate
$ grid-proxy-init -debug -verify
# Run testjob
$ globus-job-run qcdmachine /bin/date
```