



Using MATLAB on the TeraGrid

Nate Woody, CAC

John Kotwicki, MathWorks

Susan Mehringer, CAC



MATLAB on the TeraGrid

- This is an effort to provide a large parallel MATLAB resource available to a national (and inter-national) community in a secure, useable manner.
- Several different components make up this system and integrate with the MATLAB client at different levels.
- All functions are provided by various “services”, meaning you never actually logon to any CAC systems. The client software simply makes requests to CAC systems.



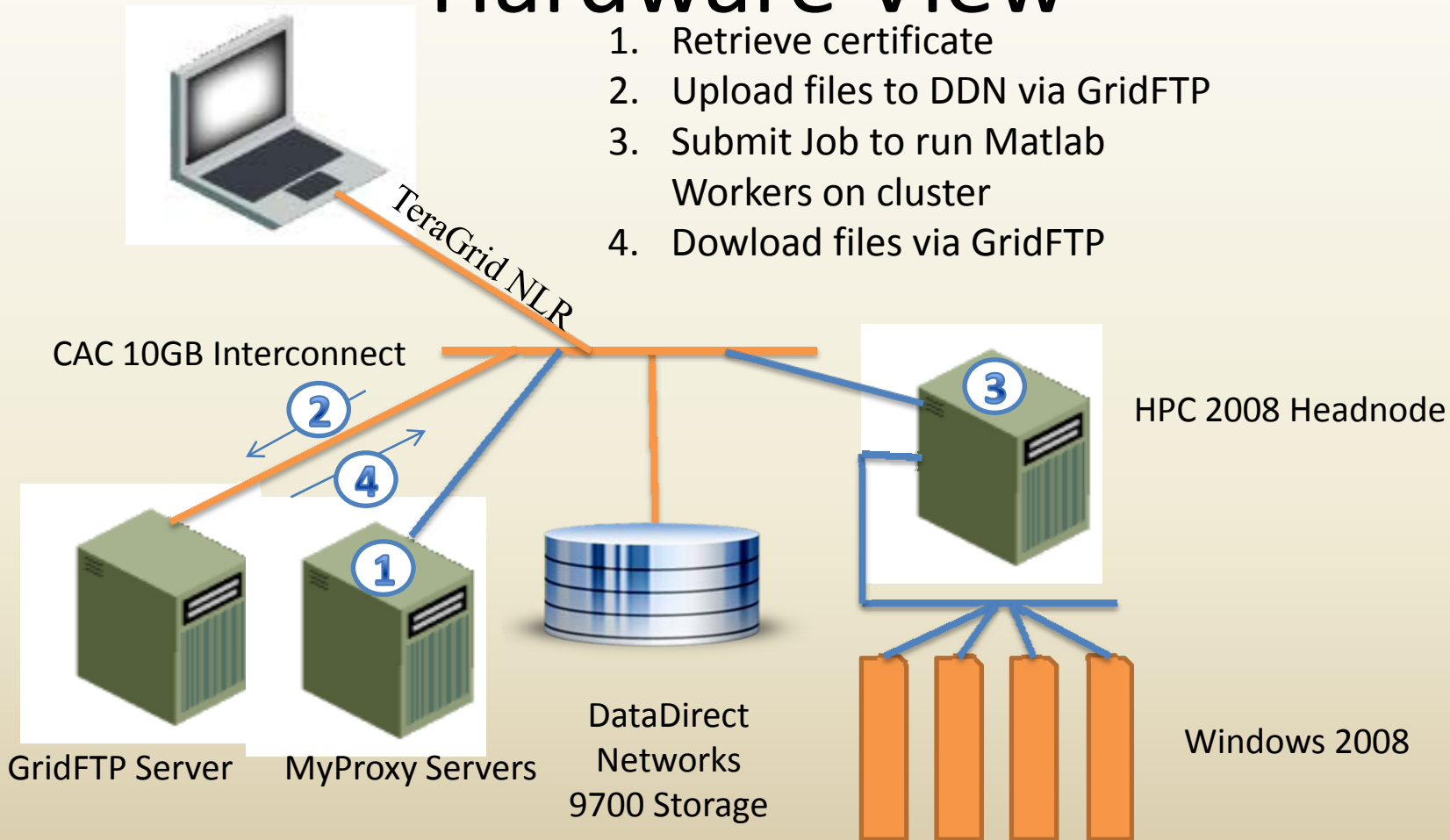
High-Level Process

- Security is managed via short-lived certificates. When you login to the system you are trading your username-password for a certificate that you will allow you to use the services.
- **File transfer service** – allows you to move files through a specialized FTP server to a file system that is mounted on all compute nodes.
- **Job submission service** – allows you to submit and query jobs on the cluster, these jobs are executed by MATLAB workers on the compute nodes.



Hardware View

1. Retrieve certificate
2. Upload files to DDN via GridFTP
3. Submit Job to run Matlab
Workers on cluster
4. Download files via GridFTP





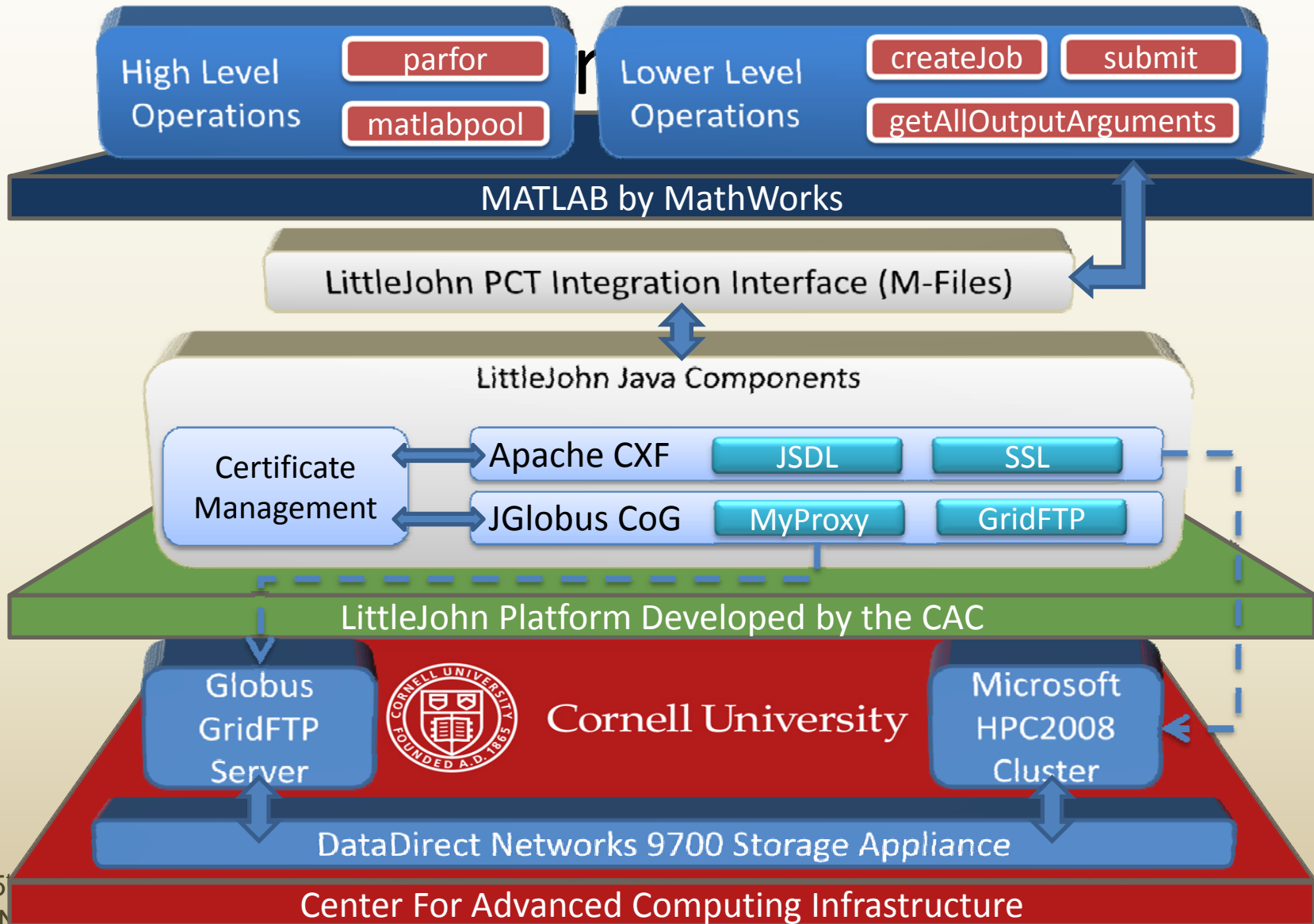
Software View

- Much of the file movement and job submission interactions is hidden by software integrated with MATLAB.
- The software package we call LittleJohn is a mix of Java and .m files that interfaces through the PCT 'generic scheduler' interface to allow access to TUC from your MATLAB client.
 - LittleJohn – software package installed on the client machine that allows you to use our parallel resource.
 - TUC – the 512 core cluster where MATLAB jobs are run.



Client Software

Ithaca, NY





A Word on Security

- Logging on to MyProxy returns a short-lived X.509 certificate that is used to authenticate to services.
- Most users will use the CAC MyProxy server, performing security this way allows any TeraGrid user to access the system using their TeraGrid username/password on a TeraGrid MyProxy server.
- The job submission and status information is accessed via a web service call that is secured by a client-certificate SSL (or TLS) connection. Your data and job requests are transferred over secure channels.



GridFTP

- GridFTP is an extension of the standard File Transfer Protocol that has been developed as part of the Globus Toolkit.
- GridFTP provides two key extensions that LittleJohn makes use of to make this project work.
 - GSI Security – The Grid Security Infrastructure provides authentication and encryption of the file transfer level and interoperates with MyProxy X.509 certificates.
 - Parallel Transfers (Extended Mode) – A higher percentage of available bandwidth can be used by making use of multiple simultaneous connections.



A note on the platform

- The compute nodes that run the MATLAB jobs are running Windows HPC 2008 (64 bit).
 - Since a minority of people are running a Win64 platform, files likely need to be compiled on TUC (mex and otherwise).
 - MATLAB is relatively resilient to paths with the wrong direction of slashes, but this is a Windows path:
 - C:\Users\naw47\myfiles\this.dat
 - This is a Mac and linux path:
 - /home/naw47/myfiles/this.dat



Support

- As a funded project, the system is free to use for research applications
 - we will ask for information on your project so that we can learn who we are supporting and how to best address problems.
- We also provide support for the system
 - Troubleshooting
 - Guidance on optimizing your application
 - General help with parallel MATLAB