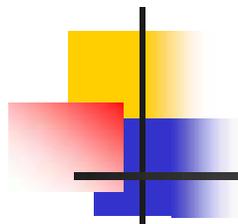


# GridFTP Project Report

---

Igor Mandrichenko

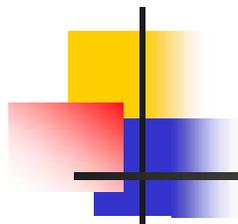
Grid, Data Movement and Storage Project  
Status



# History

---

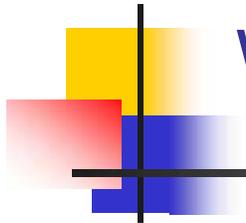
- During GGF11, February 2003 GridFTP Working Group was rechartered
- Goals
  - Finalize GridFTP v1 Protocol Document
  - Start working on GridFTP v2
    - Basis – so called Toronto list of issues/possible improvements



# History

---

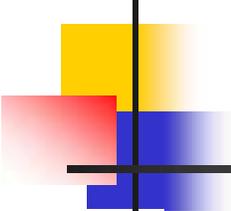
- In collaboration with ANL and working group members, published 2 GGF documents:
  - GridFTP v1 Definition
  - GridFTP Improvements
- Developed a prototype implementation of GridFTP v2 server and client
- Produced definition of GridFTP v2



# What is in GridFTP v2.0

---

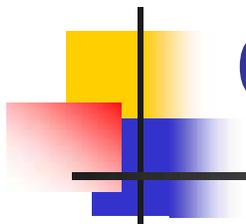
- Data Streaming
- eXtended block transfer mode (X-block mode)
- GET/PUT
- Explicit EOF communication in S mode
- Data integrity verification commands



## eXtended (X) block mode

---

- Parallel data transfer mode
- Improvement of Extended (E) block mode
- Data flow direction independent of data socket connection
  - Makes parallel transfers possible in both directions in presence of firewall, NAT
- Dynamic data channel (bandwidth) management
- Data integrity verification with an option to retransmit corrupt blocks during the transfer
- Data streaming option



# GET/PUT commands

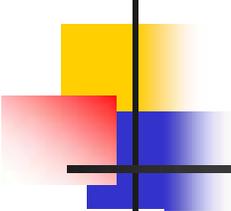
---

## ■ Example:

<u>Client</u>	<u>Server</u>
GET path=/tmp/file.dat;pasv;mode=e;	
	1xx wait
	1xx PORT=134,23,145,2,48,114
	1xx Data connection established
	2xx Transfer complete

## ■ Instead of:

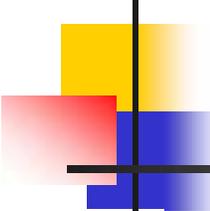
<u>Client</u>	<u>Server</u>
MODE E	
	200 OK
PASV	
	200 PORT=134,23,145,2,48,114
RETR /tmp/file.dat	
	1xx Data connection established
	2xx Transfer complete



# GET/PUT commands

---

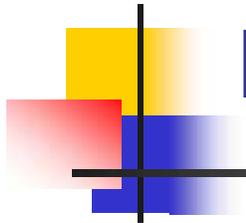
- Allow the server to create data channel after it knows what file to transfer and other transfer parameters
  - Allows scalable implementation of distributed servers such as dCache
- Pack all data transfer parameters into single atomic command
- Flexible extensions without the need to add new commands to the protocol
  - Data Streaming
- Web Services - friendly



## Explicit EOF communication in S mode

---

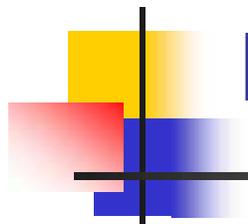
- New optional EOF command
- If client disconnects without BYE or EOF, consider last upload transfer failed
- Helps tape-based server such as dCache discard incomplete file instead of storing it on tape



# Data Integrity Verification Commands

---

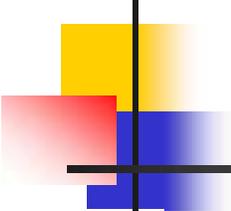
- CKSM – get checksum for a file or its portion
  - The file is already on server
- SCKS – send checksum ahead of file to be transferred
  - the server can verify integrity of the file
  - do not store corrupt file on tape



# Data Streaming

---

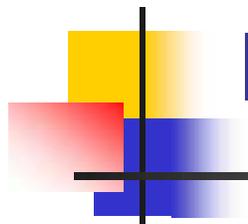
- Relatively new addition, after GGF10
- Allow concurrent transfers of multiple files in the same session
- Identify individual transfers with transaction IDs
- Use GET/PUT to initiate the transaction
- Send transaction ID with each X-mode block



# GridFTP WG Future

---

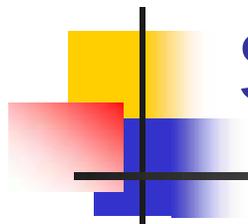
- GridFTP WG Charter:
  - Publish GridFTP v1.0 protocol document as an official GGF document - by GGF8
  - Collect and discuss list of improvements to GridFTP v1.0 - by GGF9
  - Produce GridFTP v2.0 protocol definition draft document - by GGF10
  - Implement GridFTP v2.0 client and server and document the implementation - by GGF10
  - Produce GridFTP v2.0 protocol document - by GGF11



# Future Plans

---

- Last minute modifications, polishing
  - Deadline - July 4
- Submit GridFTP v2 for Public Review
  - by July 16
- Once the document becomes official GGF Recommendation
  - congratulations, we are done !!



# Summary

---

- In cooperation with ANL significantly improved GridFTP protocol
  - Allow scalable implementation of distributed servers such as Disk Cache
  - Allow full parallel transfer functionality in presence of firewalls, NAT
  - Provisions for data integrity verification, correction
- Gained experience in using GGF as a venue for communication with grid community