

The Virtual Data Toolkit

Miron Livny

(Alain Roy)

Computer Sciences Department
University of Wisconsin-Madison

VDT

VDT's Mission

- **GriPhyN:** "...a primary GriPhyN deliverable will be a suite of virtual data services and virtual data tools designed to support a wide range of applications. The development of this Virtual Data Toolkit (VDT) will enable the real-life experimentation needed to evaluate GriPhyN technologies. The VDT will also serve as a primary technology transfer mechanism to the four physics experiments and to the broader scientific community".
- **iVDGL:** "... VDT Robustification and Troubleshooting team (VRT) will not develop new software but instead will work to enhance the releases of the GriPhyN VDT software, as follows:
 - (1) develop and maintain a VDT test suite;
 - (2) define and implement a unified and fully integrated error reporting framework across all VDT components;
 - (3) equip the VDT with dynamically configurable event logging capabilities;
 - (4) extend the VDT with new components required for specific application purposes;
 - (5) help users to maintain and trouble shoot the VDT software,
 - (6) provide documentation, and
 - (7) create procedures and tools for reporting, tracking, and fixing bugs."

VDT

Core VDT Working Group



- Miron Livny (lead)
- UW-Madison:
 - Alain Roy (technical lead)
 - Carey Kireyev
 - Parag Mhashilkar
- UW-Milwaukee:
 - Scott Koranda (Alum)
 - Brian Moe
- Univ. of Chicago:
 - Xin Zhao
- Caltech:
 - Erik Aslakson
- CMS:
 - Rick Cavanaugh
 - Ruth Pordes
- Atlas:
 - Kaushik De
 - Saul Youssef
- And a host of supporting people...

What is the VDT?

- A collection of software
 - Grid software
 - Virtual data software
 - Utilities
- An easy installation mechanism
 - Goal: Push a button, everything just works
 - Two methods:
 - Pacman: installs and configures it all
 - RPM: installs some of the software, no config
- A support infrastructure
 - Coordinate bug fixing
 - Help desk

Software in the VDT

- Condor Group
 - Condor/Condor-G
 - DAGMan
 - Fault Tolerant Shell
 - ClassAds
- Globus Alliance
 - Job submission (GRAM)
 - Information service (MDS)
 - Data transfer (GridFTP)
 - Replica Location (RLS)
- EDG & LCG
 - Make Gridmap
 - Certificate Revocation List Updater
 - Glue Schema/Info prov.
- ISI & UC
 - Chimera & Pegasus
- NCSA
 - MyProxy
 - GSI OpenSSH
- LBL
 - PyGlobus
 - Netlogger
- Caltech
 - MonaLisa
- VDT
 - VDT System Profiler
 - Configuration software
- Others
 - KX509 (U. Mich.)

Pacman

- The VDT uses Pacman to install software
 - Pacman easily installs software of all types
 - Pacman is very easy for users to use
 - Pacman does not require root access (unlike RPM)
- Pacman is developed by:
 - Saul Youssef from Boston University (Primary)
 - VDT has contributed features/patches
- Pacman has helped make the VDT to be successful.

Pacman Installation

- Goal:
 - Type a single command
 - Everything downloads
 - Everything installs
 - Everything is configured
 - No questions asked
- We're close:
 - A few questions if you're root
 - Basic configuration, may need changing

VDT

1. Download Pacman

- `http://physics.bu.edu/~youssef/pacman/`

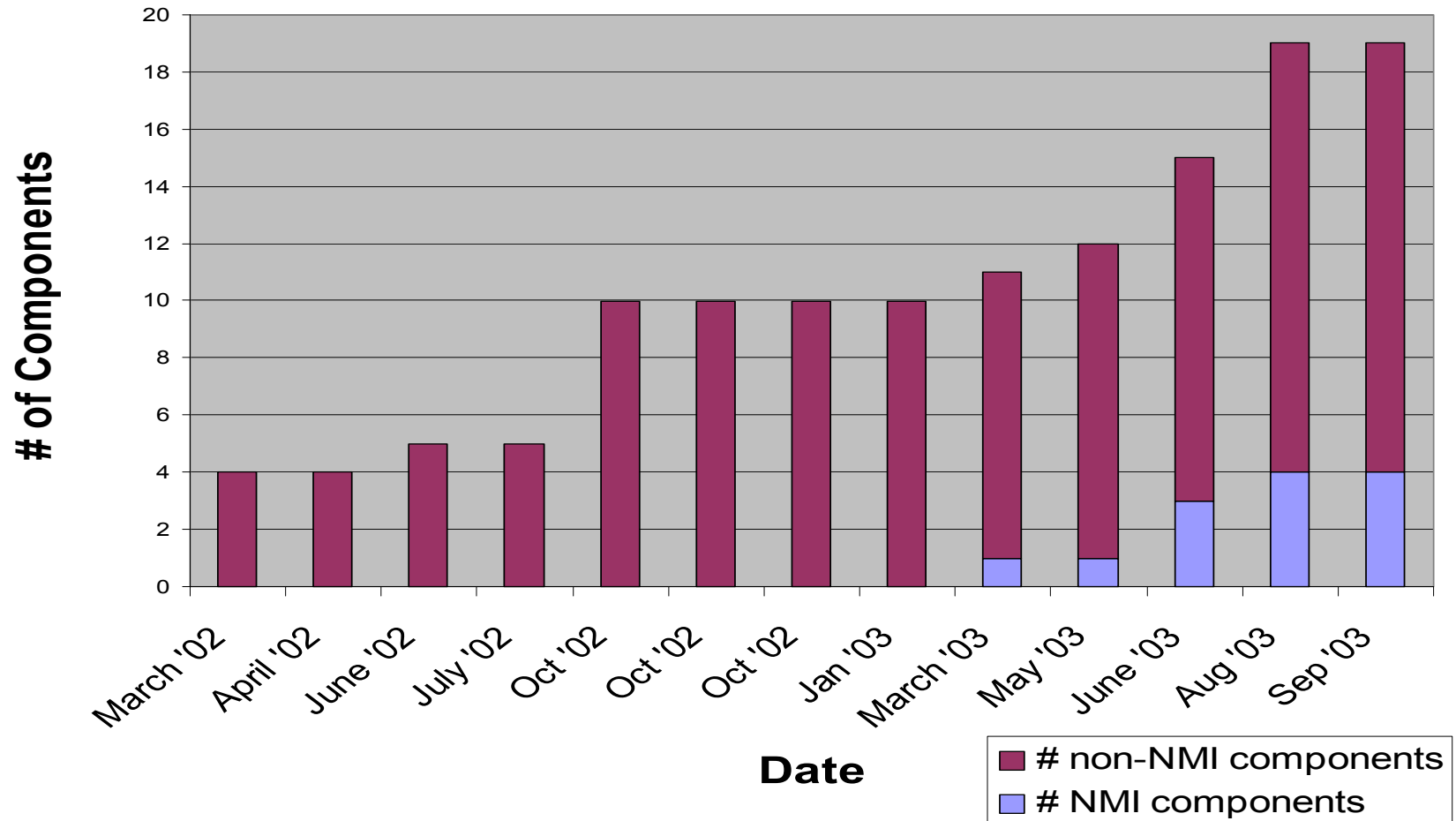
2. Install VDT

- `cd <install-directory>`
 - `pacman -get VDT-Server`
 - `pacman -get VDT-Client`
 - `ls`
- | | | | |
|----------------------|----------------------|----------------------------|------------------------------|
| <code>condor/</code> | <code>globus/</code> | <code>post-install/</code> | <code>setup.sh</code> |
| <code>edg/</code> | <code>gpt/</code> | <code>replica/</code> | <code>vdt/</code> |
| <code>ftsh/</code> | <code>perl/</code> | <code>setup.csh</code> | <code>vdt-install.log</code> |

3. Use

- *Fall 2001:* VDT started by GriPhyN and iVDGL
 - Supported USCMS testbed in 2002
- *Early 2003:* Adopted by EDG & LCG
- *Fall 2003:* Supporting Grid3
 - More than 20 sites
 - Six different applications
- *Fall 2003:* Adopted by PPDG

VDT Evolution



VDT

VDT Users

- GriPhyN collaborators
 - USCMS
 - USAtlas
 - LIGO
 - SDSS
- European Data Grid
- LCG
- PPDG collaborators (just starting)

Testing

- Thorough testing is essential for high-quality releases
- We do automated nightly testing:
 - Installation procedure
 - Functionality of software
- We work with National Middleware Initiative of the NSF to leverage their testing capabilities
- All releases tested by iVDGL VDT Testers Group before distribution

Challenges

- How do we help users transition between versions of core grid software? (like Globus 2 to Globus 3)
- How do we deal with evolution of the infrastructure (OS, HW, utilities, ...)
- Improving testing
 - Stress testing
 - Real applications
- More assistance in deploying and configuring a complete cluster/grid site
- How do we guarantee longevity.
- How do we scale the support infrastructure.



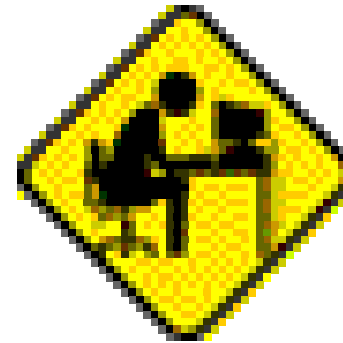
Where do you learn more?



- <http://www.griphyn.org/vdt>
- Support:
 - Alain Roy: roy@cs.wisc.edu
 - Miron Livny: miron@cs.wisc.edu
 - Ticket system: vdt-support@ivdgl.org



*Typical Grid Software
Installation Experience...*



**VDT Installation
Experience!**

VDT