TeraGrid/XD Support For Knowledge Extraction

Sergiu Sanielevici, Ph.D.

Director, Scientific Applications and User Support, Pittsburgh Supercomputing Center

Area Director for User Support Coordination, TeraGrid

User Advocate, TeraGrid/XD Technology Insertion Service

sergiu@psc.edu

What is the TeraGrid

- World's largest distributed cyberinfrastructure for open scientific research, supported by NSF for the US academic research and education community
- Integrated high performance computers (>2 PF HPC & >27000 HTC CPUs), data resources (>2 PB disk, >60 PB tape, data collections), visualization, experimental facilities (VMs, GPUs, FPGAs), network at 11 Resource Provider sites
- Allocated to US researchers and their collaborators through national peer-review process
- Integration: Single {portal, sign-on, help desk, allocations process, advanced user support, EOT, campus champions}
- Current resources: https://portal.teragrid.org/systems-monitor
- Any questions? help@teragrid.org with "UST" in the subject line

http://www.teragrid.org/

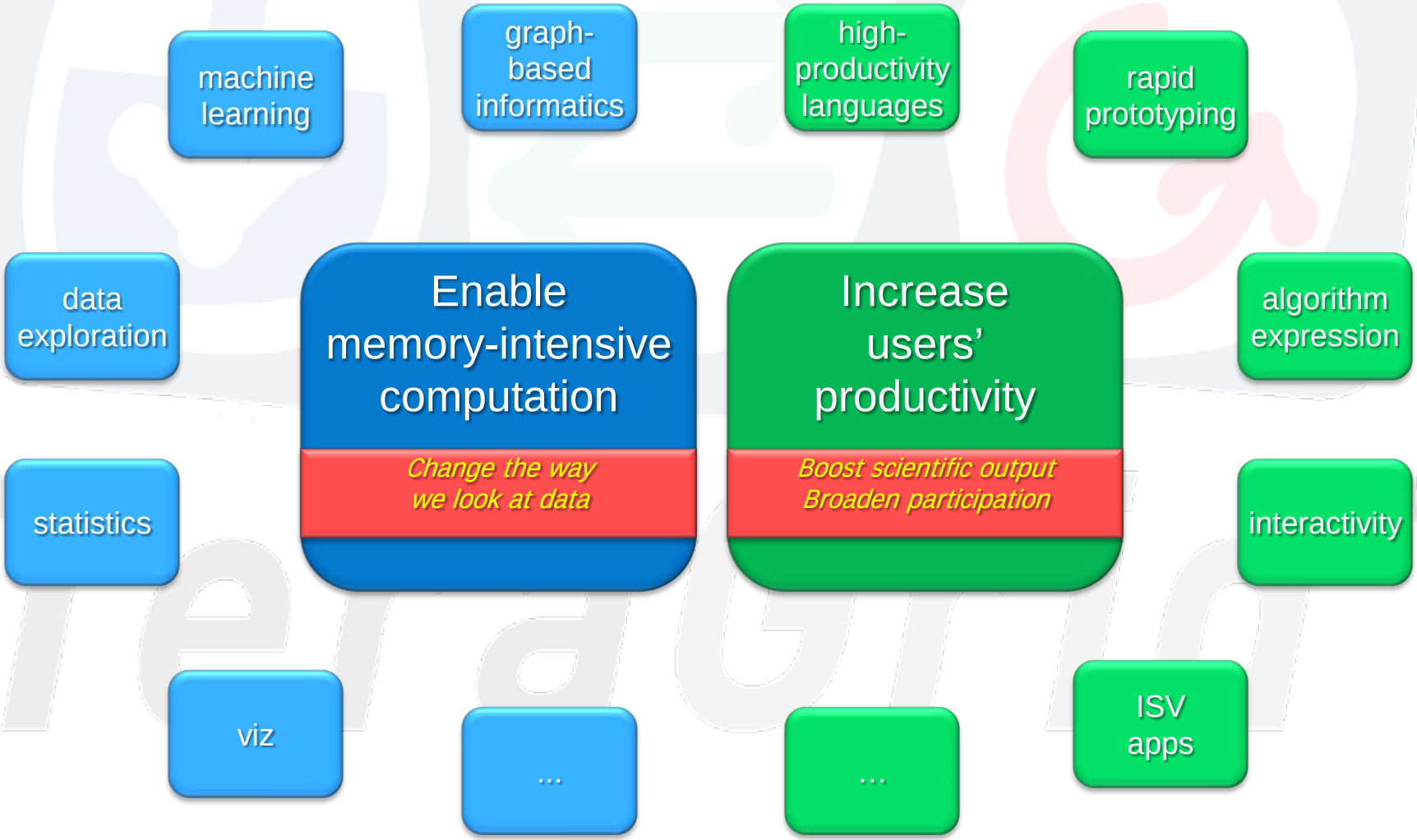
Who can use TeraGrid resources?

- Using TeraGrid is based on allocations
  - An allocation is a set of resources that are available and a quantity of each
    - Including advanced user support
  - Requests for allocations are peer-reviewed
  - Allocations PI must be from a US institution (faculty, staff, postdoc, students who are NSF fellows)
  - Allocations PI determines who can use the allocation (accounts)
  - Community allocations are available (e.g. for Science Gateways)
- TeraGrid allocations are free to US researchers and their collaborators

New Data Oriented Systems on TG

- PSC SGI UV Blacklight (now):
  - 256 blades configured in two Single System Images (SSIs) with 16 TB of coherent shared memory each are connected by a NumaLink5 Shared Memory Interconnect
  - Each blade has 2 Intel “Nehalem” eight-core processors and offers 144 gigaflops and 128 GB of RAM
  - 150 TB of disk storage, connected to TG-wide disk and archival storage systems
- SDSC Appro/ScaleMP Gordon (mid-2011):
  - 32 supernodes, each with:
    - 32 compute nodes at 240 gigaflops per node
    - 2 I/O nodes with 4 TB of flash memory per node
    - 10 TB of memory (2 TB of DRAM and 8 TB of flash memory)
    - 64 TB of DRAM
  - 256 TB of flash memory and 4 PB of disk storage

Benefits of Coherent Shared Memory



Plans for TeraGrid/XD

- Simplified, accelerated allocations and account creation process
- Campus bridging: standards based “grid/cloud” architecture that can be implemented and maintained by everyone (with advice and training from our experts) enabling work and data flows spanning lab and supercomputing resources
- Technologies and policies to better support data intensive, interactive, collaborative work
- Advanced User Support emphasizing proactive development of novel and innovative projects, and opportunities for community members to help each other
- Help to develop university curricula that include computational methods in various disciplines

Recommended reading: http://www.psc.edu/data-analytics/proceedings/proceedings.php