Addressing World-Scale Challenges

Computation as a powerful change agent in areas such as Energy, Environment, Healthcare, Education

Collaboration and Community
Massive amounts of data collected and aggregated from the internet, satellites, sensors, and other sources.

We need to move from data to knowledge.

Computing technologies are enabling new approaches applied to world-scale challenges in disciplines such as medicine and healthcare, energy and the environment, educational and social progress.

Astronomy has been one of the first disciplines to embrace data-intensive science with the Virtual Observatory (VO), enabling highly efficient access to data and analysis tools at a centralized site. The image shows the Pleiades star cluster from the Digitized Sky Survey combined with an image of the moon, synthesized within the WorldWide Telescope service.
Division within Microsoft Research focused on partnerships between academia, industry and government to advance computer science, education, and research in fields that rely heavily upon advanced computing

Supporting groundbreaking research to help advance human potential and the wellbeing of our planet

Developing advanced technologies and services to support every stage of the research process

Microsoft External Research is committed to interoperability and to providing open access, open tools, and open technology
Judith Bishop was in the first group to study computer science in South Africa in 1970 and has stayed at the front of her field of programming languages for distributed systems ever since.

She wrote the first BASIC compiler for ICL computers in 1972 and was involved in the first Pascal compiler for the in 1976. Her doctorate investigated the relationship between the new languages of the 1970s (such as Ada and occam) and the stack and descriptor based mainframes of the time. She wrote the first Java textbook to become widely used in 1997 and one of the first C# textbooks in 2004. After having contributed to the field of configuration description languages in the 1990s, she now works on the principles of adaptive software in a multi-lingual and mobile environment, in collaboration with Microsoft Research, local companies and collaborators in Germany and Italy. Professor Bishop is the top NRF rated woman computer scientist in South Africa and has published over 70 journal and conference papers. Her 14 books are available in six languages and read worldwide.
Our goal is to accelerate research by collaborating with academic communities to create open tools and services based on Microsoft platforms and productivity software.

By building open software solutions in collaboration with the research community, we help scientists spend more time on their research and less time on IT issues.
announcing

Project Trident: A Scientific Workflow Workbench

Dryad and DryadLINQ
Accelerating the pace of discovery

- Makes it easier for scientists to ingest and make sense of data
- Get answers to questions at a rate not previously possible
- Capture provenance

Scientists in data-intensive fields such as oceanography, astronomy, environmental science and medical research can use these tools to manage, integrate and visualize volumes of information.

- The tools are available as no-cost downloads to academic researchers and scientists

What once required weeks or months of custom coding, now takes just hours
Project Trident for Researchers

• Visually program workflows
• Libraries of versioned activities and workflows
• Social annotations and search, export entire workflow libraries to share their methodology.
• Automatically schedules workflows over HPCS
• Support for administering and monitoring workflows
• Automatic provenance capture, for both workflows and results
• Cost model, including elapsed time, CPU, memory, data transfer
• Integrated data storage and access, from SQL to S3 and SDS
• Integrated visualization tools
• Fault tolerance, also used to facilitate smart reruns and what-if analysis
• Supports reproducible research

Project Trident is implemented on top of Microsoft’s Windows Workflow Foundation, using the existing functionality of a commercial workflow engine based on SQL Server and Windows HPC cluster technologies.
Project Trident: Scientific Workflow Workbench
University of Washington and Monterey Bay Aquarium Research Institute

Scientific workflow workbench to automate the data processing pipelines of the world’s first plate-scale undersea observatory

- From raw data to useable data products (visualizations)
- Focusing on cleaning, analysis, re-gridding, interpolation
- Support real time, on-demand visualizations
- Custom activities and workflow libraries for authoring
- Visual programming accessible via a browser
Allows users to connect to a Trident server;

To insert the output of a workflow (pipeline) into a document;

Each pipeline in a Word document is associated with an image or text;

A person reading the document can click on such an image or icon to view the associated pipelines and its input files, and rerun the pipelines on a Trident server while remaining in the Word application.
• Deployed at MIT BROAD, Gene Pattern Analysis Server, Jill Meserov PI.
• Trident Scientific Workflow Workbench
Turning a cluster into an easy-to-use tool:
- Dryad was designed to simplify the task of implementing distributed applications on clusters of Windows computers
- DryadLINQ is an abstraction layer, which simplifies the process of implementing Dryad-based applications

The Academic Release includes:
- Installation guide, programming samples, tutorials.
- Client SDK Installer – installs DryadLINQ, docs and code samples.
- Dryad & Dryad Management Tools installer (cluster-side installation)

The Pre-release was installed at Indiana University and the University of Washington
- Successfully developed bioinformatics application (pair-wise alignment of genetic sequences) with virtually no support
- Successfully developed queries for LSST data

Small community of internal DryadLINQ developers tested on a shared infrastructure (k18 cluster, 70 nodes)
Where to download the tools
research.microsoft.com/en-us/collaboration/tools

Other resources:

- Tools to Access Petabytes of Data
- Beyond Search with Data Driven Intelligence (11 AM, Cascade)
  Harry Shum, Corporate Vice President, Microsoft
  “The future of search focusing on data-driven research to help advance the state-of-the-art in the online world”

- DemoFest Booth 4
  Tools and Services for Data Intensive Research
Agenda
- On-line
- Printed

Please let us know if you have any questions or need any help