

DataTAG

Research and Technological Development for a Transatlantic GRID

Abstract

The main idea behind the DataTAG project was to strengthen the collaboration between, Globus middleware based, Grid projects in Europe and in the USA. We provide a common dedicated test-bed, in order to facilitate the deployment of a “transparent” Grid infrastructure. It gives access to the massively distributed computing infrastructure that is needed to meet the challenges of modern high energy physics experiments such as the data intensive LHC (Large Hadron Collider at CERN) applications. Consequently, the DataTAG project has addressed the issues which arise in the sectors of high performance inter-Grid and interoperability between the Grid middleware layers such as information and security services. The advances made have been disseminated into each of the associated Grid projects.

Objectives

The goal is to create a large-scale multi-vendor transatlantic test-bed for Data Intensive Grids. Together with the European DataGrid (EDG) project and other related Grid projects in the USA such as GriPhyN, iVDGL and PPDG, this will allow to explore advanced networking technologies as well as interoperability issues between different Grid domains.

Technical Approach

The Grid interoperability issues have been addressed by INFN in collaboration with DataGrid and the Globus team through the definition of a new interoperable schema for describing Grid resources such as computing, storage and network elements. Whereas, the networking issues have been addressed by procuring a dedicated 2.5 Gbps (phase 1) then 10Gbps (phase 2) circuit between CERN in Geneva and the StarLight Internet Exchange located at Northwestern University in Chicago.



Simplified view of the DataTAG project showing the EC project partners as well as the main co-operating networks on both sides of the Atlantic.

Testbed

The transatlantic DataTAG testbed is based on equipment and circuit funded jointly by the European Commission, the US National Science Foundation (NSF) through the Electronic Visualisation Lab (EVL) at University of Illinois in Chicago (UIC) and the US Department of Energy (DoE) through the California Institute of Technology (Caltech).

The DataTAG multi-vendor layer 2 (Ethernet) and layer 3 (Internet) testbed with equipment from mainly Alcatel, Cisco and Juniper connected to high end Linux servers at 1 Gigabit as well as 10 Gigabit Ethernet is believed to be unique. The main use of this testbed has been for advanced network research and demonstrations focusing, in particular, on very high speed transport protocols, QoS and advance bandwidth reservation techniques. Access to the testbed for experimental purposes is made possible in Europe through GEANT/DANTE and the National Research and Education Networks (NREN) involved using transparent VPN layer 2 tunnels, as well as through a 10Gbps circuit from Amsterdam contributed by Surfnets and a 2.5Gbps circuit to Lyon contributed by the French VTHD pilot network. The DataTAG test-bed is open to other EC Grid projects as well as to other network research projects in the field of high performance transport protocols, on a case by case basis.

Experiments

Together with Surfnets that also operates a similar transatlantic circuit, the project pioneered Ethernet over SONET/SDH transport networks, i.e. next generation G.709 networks (digital wrapper).

Applications

Together with DataGrid and the LHC Computing Grid (LCG) project, the software of the CERN LHC experiments ALICE, ATLAS CMS and LHCb has been adapted to the Grid environment.

Innovation

The DataTAG project has many innovative components in the area of high performance transport, Quality of Service (QoS), advance bandwidth reservation, EU-US Grid interoperability and new tools for easing the management of Virtual Organizations such as the Virtual Organization Membership Server (VOMS) and Grid monitoring (GridICE).

Results

Thanks to the excellent cooperative spirit between the European and American teams involved in DataTAG, remarkable results have been achieved in a very short while. Regarding high performance transport, new Internet2 landspeed records for both IPv4 and IPv6 have been established and TCP/IP stacks variants such as FAST, Scalable TCP and High Speed TCP have been extensively tested.

Regarding EU-US Grid interoperability, the Grid Laboratory Universal Environment (GLUE) activity led to the definition of common/interoperable Information specifications to be used for Resource Discovery.

The "GLUE" schema has been subsequently incorporated in the Globus and DataGrid middleware.

The WorldGrid demos made at IST2002 and SC2002 conferences and featuring cross job submission between the US and EU with "GLUE" enhanced EDG and Globus middleware on both sides of the Atlantic marked a very significant milestone towards the interoperability of EU & US Grid projects such as DataGrid, GriPhyN, iVDGL and PPDG.

Contribution to Standards

DataTAG is heavily involved in GGF activities with presence in the steering group (GFSG) and in several research groups (RG) and working groups (WG) such as Authorization Frameworks and Mechanisms, Grid Resource Allocation Agreement Protocol, High Performance Networking, Data Transport, OGSA, OGSF. The Generic AAA work co-authored by the University of Amsterdam

(UvA) is described in IETF RFCs 2903/2904/2906 and IRTF Draft draft-irtf-aaaarch-generic-policy-03.txt

Success Stories

Following the upgrade of the DataTAG circuit to 10Gb/s in September, the first ever transatlantic testbed with 10Gigabit Ethernet access capabilities became available, just in time for live demonstrations at Telecom World 2003 where CERN had a booth on the Lake Geneva Region stand. Thanks to this, a new 5.44Gb/s single stream TCP/IP data transfer record between CERN and Starlight was established by Caltech and CERN on October 1, and considerable publicity was given to this very significant technical feat by numerous media around the world. Indeed, ultra high speed data transfer will enable to download CDs, DVDs, complex images, etc, in seconds rather than minutes or hours which, in turn, will pave the way to new innovative applications.

Project name:
DataTAG

Contract no.:
IST-2001-32459

Project type:
RTD

Start date:
01/01/2001

Duration:
24 months

Total budget:
€ 4,105,196

Funding from the EC:
€3,980,826

Total effort in person-months:
525

Website:
<http://www.datatag.org>

Contact person:
Dr. Olivier H. Martin
email: Olivier.Martin@cern.ch
tel.: +41-22-767-4894
fax.: +41-22-767-7155

Project participants:

CERN	Int.Org
INFN	IT
INRIA	FR
PPARC	UK
UVA-NWI	NL

Keywords:
GRID
Networks
Scientific research
Telecommunications

Collaboration with other EC funded projects:
CrossGrid
DataGRID
GÉANT
Gridstart

IST - Research Networking - Research on Networks - Grids