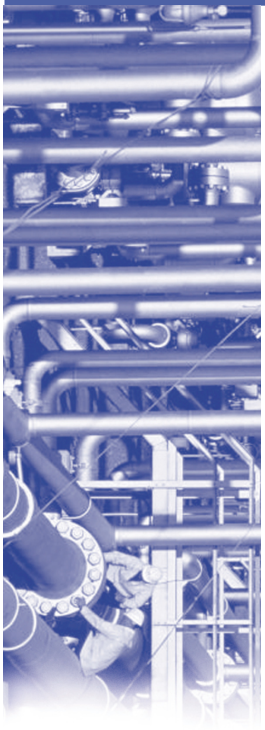
NEWS RELEASE



DILIGENT Data Challenge on EGEE Infrastructure

Geneva, 6 December 2007 – The success of DILIGENT's recent data challenge on image feature extraction, executed on the EGEE infrastructure, has enabled one of the world's largest collections of multimedia metadata to be made publicly available for research purposes.

The DILIGENT team used the EGEE computing Grid to process 37 million images from the online Flickr database in just 16 weeks. This computation generated approximately 112 million text and image objects—nearly 5 TB of data—containing more than 150 million extracted features. This is equivalent to an average processing capacity of over 300,000 images per day.

This unique collection will be used by the SAPIR project to develop new large-scale content-based data retrieval and automatic data classification techniques that combine both text and image content, expanding the limits of conventional search engines, which can only search text associated to images and audio-visual content.

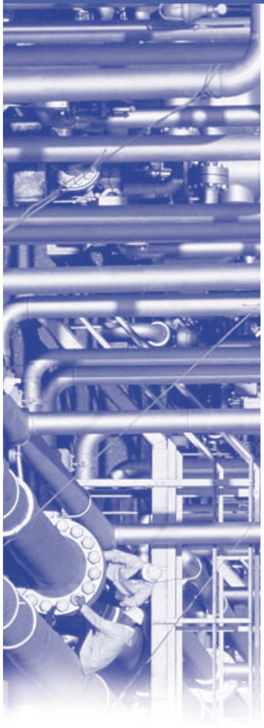
The computational load required to generate this massive data collection was outsourced to DILIGENT, and then delegated to the EGEE Pre-Production Service (PPS) Grid infrastructure via the gLite middleware. A total of 44,333 gLite jobs were successfully executed by the EGEE PPS infrastructure resource broker. Each job processed approximately 1000 images.

The data challenge lasted for 116 days, from 16 June to 9 October 2007, and was organized in three different phases. During the initial preparation phase experimental jobs were submitted to some EGEE PPS sites to test the feature extraction application and optimize the number of images to process per day. The next two phases involved actual execution of the data challenge, exploiting ten EGEE PPS sites that contributed their computational resources: University of Athens, Scuola Normale Superiore, ISTI-CNR, LIP, ESA-ESRIN, CERN, CESGA, University of Macedonia, Ben Gurion University, and CYFRONET. Four of these sites are maintained by DILIGENT partners.

###

Updated 06/12/2007

EGEE is a project funded by the European Commission - contract number INFSO-RI-031688



Notes for editors:

1. For more information about the DILIGENT data challenges, including detailed results, please visit https://twiki.cern.ch/twiki/bin/view/DILIGENT/DiligentFlickrDC.

2. The DILIGENT project is co-funded by the European Commission. The project objective is to create an advanced test-bed that will allow virtual e-Science communities to share knowledge and collaborate in a secure, coordinated, dynamic and cost-effective way. For more information see http://www.diligentproject.org and http://www.gcube-system.org or contact info@diligentproject.org

3. The SAPIR project is co-funded by the European Commission. SAPIR's goal is to provide searches over huge quantities of multimedia objects. The searches are based on both text and multimedia features and exploit similarity and search-by-examples queries. For more information see http://www.sapir.eu/

3. The Enabling Grids for E-sciencE (EGEE) project is co-funded by the European Commission. The project aims to provide researchers in both academia and industry with access to major computing resources, independent of their geographic location. For more information see http://www.eu-egee.org/ or contact Sarah Purcell, EGEE Dissemination, Outreach and Communications Manager, on + 41 22 767 41 76 or email sarah.purcell@cern.ch

4. The EGEE Pre-Production Service runs in parallel with the production service, but in a way that is as much like a production service as possible. In general this is where the next version of the middleware is first deployed, where applications test that it is useable, and that the site operators verify that it meets their needs too. More information about EGEE PPS at https://cern.ch/egee-pre-production-service/

5. Flickr is a well known photo management and sharing web application. It helps people make their photos available to others. Thousands of new photos are uploaded every minute and million of photos are geo-tagged. For additional information see http://www.flickr.com/tour/