Commercial e-Science Applications on UK Light

Terence Harmer, P Donachy and R Perrott
Belfast e-Science Centre, The Queen's University of Belfast, Belfast BT71NN
{t.harmer, p.donachy, r.perrott}@qub.ac.uk
http://www.qub.ac.uk/escience

1. Background and Network Infrastructure

The Belfast e-Science Centre is focused on commercial, inter-enterprise applications of grid technology with commercial partners in biotechnology, broadcast media, data mining and military media services. The UK Light experiment aim to expand and increase the capabilities of the BeSC application test beds that are prototyping service grids in these areas. In particular, UK Light connectivity will enhance the existing grid test bed that links BeSC, BBC NI, BBC R&D and JANET that has been running GT3 services since October 2003; the current BeSC infrastructure is depicted below.

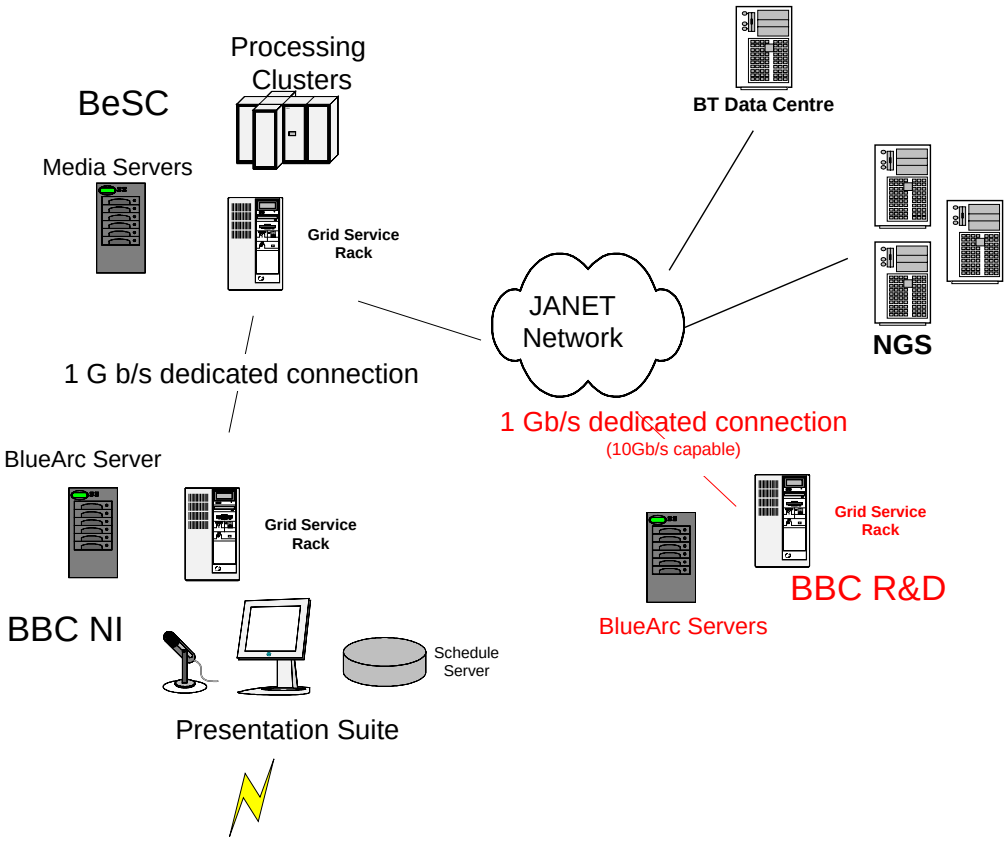

Figure 1: BeSC Application Testbed

The current infrastructure has a BeSC—JANET 1Gb/s dedicated link, BBCNI—BeSC—JANET 1Gb/ dedicated link and BBC R&D Tadworth—TeleHouse North, London—JANET 1Gb/s Link. The BeSC-BBCNI link is capable of 2G/s. The BBC R&D to Telehouse North connection is by project owned and managed dark fibre that is provisioned for 10Gb/s connectivity. The Gridcast project has a 40+-node blade server in BeSC, a 4-node server in BBCNI and a 4-node server in BBC R&D. The Gridcast project has a 4 Tb BlueArc data server in BBCNI and two 8 Tb BlueArc server in BBC R&D. Each of these data servers is capable of 2Gb/s sustained read or write transfer rate or 1Gb/s sustained simultaneous read *and* write transfer rate. This network configuration is primarily designed to support the Gridcast project but is used by all of the projects in the centre portfolio as a test bed infrastructure. UK Light will be used to provide a backbone network infrastructure to replace that currently provided by JANET, and enable high-speed remote access linking BeSC—BBCNI—BBC R&D and to other BeSC collaborating sites such as SDSC.

2. Project Portfolio and Exemplar UK Light Experiments

2.1 **Gridcast. Partners:** BeSC, BBC NI (Belfast), BBC R&D (London) and BT plc.

The Gridcast project aims to prototype the next generation of broadcast media infrastructure and enable the distributed and secure management of broadcast content (audio and video) and secure provision of media processing services. The project is scenario driven and manages large-scale video files (25 Gb+ per hour of programming) and live video streams that are part of a BBC broadcast schedule. A project aim is to model the management and distribution of the media content in parallel with the current day-to-day BBC infrastructure. The UK Light experiments will model the large-scale movement of video files between the test bed sites to model the day-to-day operation of the BBC infrastructure that supports 10 television channels. Of particular interest in using UK Light, is the prototyping of scheduling, transfer and processing services for high-definition video where 150 Gb+ per hour of broadcast content will be the norm; i.e supporting 10 TV Channels, each with 15+ hours of programming per day and each using high-definition video. These experiments are timely because the BBC is planning a Lambda-based network infrastructure for the next generation of its production and broadcasting infrastructure and is interested in the quality of service issues in such a network infrastructure.

2.2 **GEDDM** (Grid EnableD Data Mining). Partners: BeSC and Datactics Ltd.

The GEDDM project aims to implement high performance, on-demand data-mining grid services for the commercial Datactics data-mining engine. Datactics aim to move from being a company with in-house high performance computing equipment to a high-performance data mining service provider using utility computing. The project aims to prototype remote data mining services that require high speed, multi-format data transfers and processing. The UK Light experiments will prototype high-performance data conversion and mining services for legal data mining on multi source and multi format files. The experiments will be based upon the commercial use-case of mining 25 Tb of litigation material to provide a test bed for remote high-performance data mining services. For the purposes of such experiments, the project will use the Gridcast infrastructure outlined above in addition to co-locating data mining services at UK Light end-points.

2.3 **GeneGrid.** Partners: BeSC, Amtec Medical Ltd (Belfast), Fusion Antibodies Ltd, SDSC

The GeneGrid project aims to create a virtual bioinformatics laboratory using grid-based services. The laboratory will link the BeSC, the commercial partners, SDSC and publicly available services and data sources. (e.g. EBI) The UK Light experiments will allow GeneGrid to truly demonstrate inter-enterprise collaboration within the life science sector. Within the UK, UK Light will allow GeneGrid to experiment with data, workflow and application services across various UK light end-points. This will include collaborations with other UK projects including MyGrid and Bridges. Such datasets are multi Gb files in multiple locations. On the international arena GeneGrid has a working relationship with the SDSC Encyclopedia of Life project and is interested in deploying services across both sites utilising the potential UKLight/StarLight backbone. This will allow the project to experiment with deploying complex services across truly geographically independent locations. This novel architecture and bandwidth will bring a new level of functionality to bench Biologists and Bioinformaticians in both projects.

2.4 **OpenRiskGrid.** Partners: BeSC, First Derivatives plc (NI), BT plc and KX Systems

The project is implementing high-performance, on-demand financial risk management services for stock portfolios. The financial sector depends heavily on calculations as part of financial regulation and also to improve financial competitiveness. Risk management applications are characterised by user demand for high performance and the high volume of data involvement and the need to use multiple remote data sources. The UK Light tests will model the provision of real-time stock market risk analysis on a market cycle basis using multi-source trading information, stock market portfolio data and historical trading information. The infrastructure will use that defined for Gridcast with the addition of co-located equipment at UK Light endpoints.