

Impact case study (REF3b)

<p>Institution: University of Cambridge</p>
<p>Unit of Assessment: UoA9</p>
<p>Title of case study: Using grid technology to develop image processing software</p>
<p>1. Summary of the impact (indicative maximum 100 words) Grid computing research conducted by the High Energy Physics (HEP) Group at the University of Cambridge, Department of Physics has enabled software company IMENSE to develop and commercialise a range of content based image recognition products. The research gained substantial media interest and was featured at the BA Festival of Science 2008.</p>
<p>2. Underpinning research (indicative maximum 500 words) Research conducted by Professor Andy Parker at the University of Cambridge Department of Physics (Lecturer from 1992, now Professor) involves the use of large scale grid computing for the analysis of high energy physics data. Based within the High Energy Physics (HEP) Group, Professor Parker has been involved in distributed computing since the late 1980's when he was responsible for computing and data handling for the CERN flagship UA2 experiment, and produced a series of groundbreaking physics results, for example [1]. He is a member of GridPP, a collaboration of particle physicists and computer scientists from the UK and CERN which developed the software to build a distributed computing Grid across the UK for particle physicists. Professor Parker was one of the authors of the original GridPP proposal in 2000 [2]. Funded by the Science and Technology Facilities Council (STFC), GridPP was developed to handle and analyse the UK's share of the petabytes (one petabyte is one quadrillion bytes) of data being generated by the LHC annually, requiring huge data storage and processing capabilities.</p> <p>Within the project, Professor Parker developed the Ganga distributed computing management system between 2002 and 2004 as a unique contribution from Cambridge [3,4]. The scale of data analysis task means that datasets are sent from the "Tier 0" centre at CERN to 10 international "Tier 1" centres, including one in the UK. Each national centre is linked to "Tier 2" centres at Universities which supply computing and storage resources. Data processing tasks submitted to the Grid need to be broken into a series of jobs, often numbering in the thousands, which are distributed to the centres housing the relevant data. The results then need to be collated and sent back the originating user. Ganga provides a framework for managing thousands of such processing applications in a coherent way, logging progress and dealing with failures. Parker also designed, developed and managed the Tier 2 Grid facilities in Cambridge, which he further developed to provide a campus-wide Grid computing facility (CamGrid) to other researchers, using custom software solutions developed by the HEP Group. The grid system in Cambridge was unique in connecting to both the outside grid and local resources. These facilities and the software systems developed by the HEP group are now part of the UK Grid, equivalent to 40,000 PCs, and were used for the recent discovery of the Higgs boson. Since 2010 Professor Parker has exploited the system to search for new physics outside of the current standard model of particle physics [5,6]. These studies included placing limits on the existence of new particles, such as supersymmetric partners to the standard model quarks and leptons, and on constraining models which predict extra space dimensions. The physics research outputs relied on the success of Parker's computing research and development.</p> <p>The University set up the eScience Centre (now part of the Centre for Scientific Computing) in 2002 in order to exploit this research in academia and industry. As Director of eScience at the Centre for Scientific Computing Centre from 2001 to 2007, Professor Parker led the development of grid computing initiatives across the University, including projects in life sciences, earth sciences, medicine, physics, chemistry and engineering, all requiring the sharing of large computational and data resources.</p> <p>Software company IMENSE Ltd made contact with Prof. Parker in 2006 following an eScience industry forum where grid research was presented, requesting his help to address its needs in high throughput computing. Prof. Parker won funding from the STFC research council Innovations Partnership Scheme for two research projects with IMENSE. The first project (2006-2007) used an</p>

Impact case study (REF3b)

adapted version of the Ganga system to process millions of images. The HEP research group modified Ganga to manage a search for images across the internet, and to process them using the IMENSE algorithms. The Cambridge Tier 2 site and collaborating sites were used as the platform. During the second project (2007-2009) Professor Parker's research centred on escalating his previous work to provide the level of technology required for industrial use.

3. References to the research (indicative maximum of six references)

[1] A search for new intermediate vector bosons and excited quarks decaying to 2-jets at the CERN $\bar{p}p$ collider, Alitti J et al, NUCL PHYS B 400(1-3):3-22 1993, DOI: 10.1016/0550-3213(93)90395-6 (61 citations on WoS)

*[2] GridPP: Development of the UK computing Grid for particle physics
Faulkner PJW, Lowe LS, Tan CLA, Watkins PM, Bailey DS, Barrass TA, Brook NH, Croft RJH, Kelly MP, Mackay CK, et al.
Journal of Physics G: Nuclear and Particle Physics **32**(1):N1-N20 2006, DOI: 10.1088/0954-3899/32/1/N01

*[3] K. Harrison et al., "Ganga: a user-Grid interface for ATLAS and LHCb", in Proceedings of "e-Science All Hands Meeting 2004", Nottingham, 31st August - 3rd September 2004, URL: http://ganga.web.cern.ch/ganga/documents/pdf/ganga_allhands03.pdf

[4] K. Harrison et al., "Ganga: a user-Grid interface for ATLAS and LHCb", in Proceedings of the 2003 Conference for Computing in High Energy and Nuclear Physics, La Jolla, California, 24th-28th March 2003, URL: http://ganga.web.cern.ch/ganga/documents/pdf/ganga_chep03.pdf

*[5] Search for squarks and gluinos using final states with jets and missing transverse momentum with the Atlas detector in $s=7$ TeV proton-proton collisions, Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics **710**(1):67-85 2012, DOI: 10.1016/j.physletb.2011.05.061

[6] Using gamma plus jets production to calibrate the Standard Model $Z \rightarrow \nu(\bar{\nu}) + \text{jets}$ background to new physics processes at the LHC, Ask S, Parker MA, Sandoval T, Shea ME, Stirling WJ, J HIGH ENERGY PHYS Article number 058 2011

* References which best reflect the quality of the underpinning research.

4. Details of the impact (indicative maximum 750 words)

This first STFC research project described above was fundamental to IMENSE being able to finalise its image processing software, leading directly to its Series A investment of £0.5M in 2007 and subsequently thereafter to IMENSE producing its commercial products. Following the second STFC project, ending in 2009, the company was able to offer a variety of products for sale, these being based largely on the understanding of raw images gained in Parker's STFC Innovations Partnership Scheme (IPS) projects. The company is still in the market, using large scale computing to analyse images. Its success can be directly attributed to its direct use of the infrastructure and software developed by Professor Parker's HEP research group in the Cavendish. Until engaging closely in Professor Parker's research projects, the company had no knowledge or capability in grid computing.

IMENSE Ltd (they actually operate as 'imense') has thrived to the extent that it is now (2013) building the next generation of image search, developing innovative solutions that make retrieval of images even easier and more powerful than any existing search for images on the Internet. The applications all rely on the IMENSE proprietary core image recognition algorithms. Their effectiveness results from the very large samples of images used to train them, which in turn are made possible by the successful outputs of Professor Parker's research. This corpus of data had been well beyond the reach of the company alone during its start-up phase. A recent product, which also relies on Parker's image processing algorithms, is a number plate image-based app for iPhones.

Dr David Sinclair, CEO of IMENSE Ltd, states

“Imense's current success and promising future business prospects are thanks in no small part to the grid computing research conducted by the HEP group which enabled the development and commercialisation of our products. We have also benefited from the continuing engagement and support of the HEP group.

We have diversified our product portfolio from recognising scene content to reading textual information in scenes. This enhancement in our position has been facilitated by machine learning infrastructure developed as part of our IPS project and executed in the case of some of the larger training jobs on hardware resident in the HEP machine room. Much of the training data for this new market was amassed while working with the HEP.

As part of our diversification strategy we now supply B2B software solutions for reading ID card, number plates (for about 12 different countries), legacy utility meters and general text. As a marketing strategy we have approximately 30 apps available from iTunes and Google Play. Ready University based expertise has enabled us to build our own in house secure licensing server so we can license Android software without paying Google its distribution fees.”[7]

“This is an excellent example of what happens when the possible wider applications of new research and technologies are considered” Dr Liz Towns-Andrews, Director of Knowledge Exchange, STFC.[8]

The work with IMENSE offered a number of public engagement opportunities. When the LHC was first switched on in 2008, STFC invited a select handful of companies, including IMENSE, to be part of the “Big Bang Breakfast” event at Westminster, attended by Ministers and VIPs. Numerous online technology news sites reported on the high-profile event.

Project activities were also presented to an enthusiastic general public at the BA Festival of Science, Liverpool 5-10 September 2008, in the session “Latest research and today’s technology – Grid Computing in the UK”. Over 30,000 people attended the Festival, 87% of were outside academia and nearly 90% rated the Festival as excellent or good.[15]

The work attracted other positive media coverage, for example an article in the Telegraph in 2008 and a 2008 interview with David Sinclair shown on the BBC website).

5. Sources to corroborate the impact (indicative maximum of 10 references)

[7] Statement from CEO of Imense Ltd

[8] STFC press release, held on file

[9] Press coverage <http://www.theengineer.co.uk/news/internet-search-party/308683.article>

[10] itunes IMENSE number plate app: <https://itunes.apple.com/gb/app/uk-anpr/id584397416?mt=8>

[11] Camtology: Intelligent Information Access for Science. Proc of the NAACL HLT 2010, Los Angeles, June 2010.

[12] Example press coverage of IMENSE’s involvement “Big Bang Breakfast”
<http://news.zdnet.co.uk/emergingtech/0,1000000183,39486541-5,00.htm>

[13] Telegraph article as example of positive media coverage:
<http://www.telegraph.co.uk/finance/yourbusiness/brightideas/3174154/Search-engines-tap-into-world-of-sound-and-vision.html>

[14] 2008 interview with David Sinclair shown on the BBC website
<http://news.bbc.co.uk/1/hi/technology/7621089.stm>

[15] <http://www.britishtscienceassociation.org/sites/default/files/root/Liverpool%202008%20-%2020Evaluation%20report.pdf>

