University of HUDDERSFIELD

High-Performance Computing and Big Data Challenge

Dr Violeta Holmes

Matthew Newall

The University of Huddersfield

Inspiring tomorrow's professionals

THE AWARDS AWARD WINNER UNIVERSITY OF THE YEAR

theguardian UNIVERSITY AWARDS Winner 2013

2012 THE AWARDS WINNER Entrepreneurial University of the Year

THE QUEEN'S AWARDS FOR ENTERPRISE

Outline

University of HUDDERSFIELD

- High-Performance Computing
  - E-Infrastructure
  - Top500 -Tianhe-II
  - UoH experience: HPC resources and applications
- Big Data
  - Hadoop
  - Case study: Close Call
- Summary

High-Performance Computing

University of HUDDERSFIELD

- The HPC and Big Data are new technologies vital for the advancement in science, business and industry.
- ”High Performance Computing (HPC) and e-infrastructure in general are drivers of economic growth and societal well-being. They are also vital for maintaining international competitiveness in the generation of knowledge and its application.”(Kenway, Coveney, McGuire, Parchment, & Parsons, 2012)

Where we were

University of HUDDERSFIELD

50s-early 80s

- Centralised computing facilities
- Dummy Terminals
- Punch cards
- Batch Processing

µ Processors

University of HUDDERSFIELD

- Allowed users to have “intelligent” machines under their table
  - Seen as an expensive typewriter
  - Few organisations and educational institutions had them
- These machines were quickly overtaking Mainframes in price-performance factors, and later overtaking them in performance.

90s – early 2000

- Information on Demand - For the first time sharing and communicating became easy and quick.

Where we are now

University of HUDDERSFIELD

- Efficient Multitasking Desktops
- Capable Workstations
- Mobile Devices – less capable

Through the use of middlewares:

- Clusters
- Grids

The 5th Utility

Utility Computing/Cloud Computing/On-Demand Computing/Service Computing

University of Huddersfield

- The concept of just plugging in your problem into a computer system and getting results.

- Defined as:

"A large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted virtualised, dynamically-scalable, managed computing power, storage, platforms and services are delivered on demand to external customers over the internet."

(Foster et al. 2008)

HPC systems

University of HUDDERSFIELD

1969:	CDC 6600	1st system for scientific computing
1975:	CDC 7600	1st supercomputer
1985:	Cray X-MP / 4 8	1st vector supercomputer
1989:	Cray Y-MP / 4 64	
1993:	Cray C-90 / 2 128	
1994:	Cray T3D 64	1st parallel supercomputer
1995:	Cray T3D 128	
1998:	Cray T3E 256	1st MPP supercomputer
2002:	IBM SP4 512	1 Teraflops
2005:	IBM SP5 512	
2006:	IBM BCX	10 Teraflops
2009:	IBM SP6	100 Teraflops
2012:	IBM BG/Q	2 Petaflops

Inspiring tomorrow's professionals

THE AWARDS AWARD WINNER UNIVERSITY OF THE YEAR

theguardian UNIVERSITY AWARDS Winner 2013

2012 THE AWARDS WINNER Entrepreneurial University of the Year

THE QUEEN'S AWARDS FOR ENTERPRISE

Top 500 list

University of HUDDERSFIELD

http://www.top500.org/lists/2014/11/

- Top500 List of top supercomputers published every year in June and November.

Top 500 November 2014 (www.top500.org)

University of HUDDERSFIELD

List Highlights

- Total combined performance of all 500 systems has grown to 309 Pflop/s, compared to 274 Pflop/s in June and 250 Pflop/s one year ago.
- There are 50 systems with performance greater than 1 petaflop/s on the list, up from 37 six months ago.
- The No. 1 system, Tianhe-2, and the No. 7 system, Stampede, use Intel Xeon Phi processors to speed up their computational rate. The No. 2 system, Titan, and the No. 6 system, Piz Daint, use NVIDIA GPUs to accelerate computation.

2014… The Tianhe-2 (Milky Way-2)

University of HUDDERSFIELD

- Ranked 1st in the top500 list of the most “powerful” (computing intensive) computers (June 2013)
- Ranked 6th in the graph500 list of the most “powerful” (data intensive processing) computers (June 2013)
- Ranked 32nd in the green500 list of the most energy efficient computer (June 2013)
- China (National University of Defense Technology)

HPC - UK

University of HUDDERSFIELD

- HPC is of considerable importance to the UK economy
- In 2012 and 2013 the UK government invested extensively into the HPC systems infrastructure
- Created Hartree Centre, STFC HPC facilities, and
- Funded number of HPC centres across UK

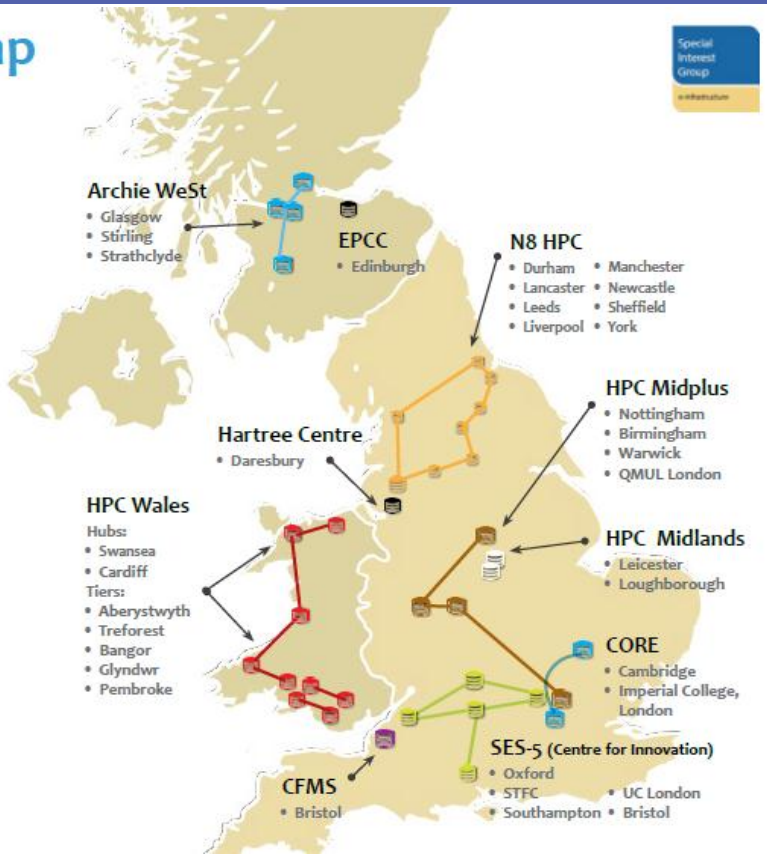
HPC Capability Map

https://sbri.innovateuk.org/documents/3108763/0/UK+HPC+capability+report/87d65dd0-82da-42da-9e7b-18a97ea06003

HPC Capability Map

HPC Facilties

- HPC facilities available for academic and commercial use.
- Facilities with similar capacity have been omitted shown if they are not available to the business community.



Inspiring tomorrow's professionals

HPC Resources at the University of Huddersfield

University of HUDDERSFIELD

- The University of Huddersfield Queensgate Grid enables access to Local and External National resources.
- Local HPC resources - Campus grid Queensgate Grid (QGG) compute clusters and Condor pool
- A share in an IBM iDataPlex system as part of the STFC enCore cloud service at Daresbury Laboratories, Hartree centre – IBM Blue Gene From 2011-2014

The Queensgate Grid – Local HPC resources

University of HUDDERSFIELD

- HPC Systems – Clusters
  - Intel based cluster – Eridani
  - AMD/Nvidia based Cluster with GPUs – Vega
  - SUN Cluster – SOL
- A HTC High Throughput System – 2000+ slot Condor Pool
- Cycle Stalling Render Farm – 120 Machine Render Farm

Local Resource: Sol Cluster

University of HUDDERSFIELD

- **Name:** Sol
- **Type:** Cluster
- **Cores:** 260
- **Memory:** 600GB
- **Interconnect:** Infiniband

- **R-max:** 2457.6 GFlops
- **R-peak:** 1004.8 GFlops
- **R-av:** 749.53 GFlops

Inspiring tomorrow's professionals

THE AWARDS AWARD WINNER UNIVERSITY OF THE YEAR

the guardian UNIVERSITY AWARDS Winner 2013

THE AWARDS 2012 WINNER Entrepreneurial University of the Year

THE QUEEN'S AWARDS FOR ENTERPRISE

Local Resource: Vega Cluster

University of HUDDERSFIELD

- The Head Node - 1* Quad Core Intel Xeon E5630 2.53Ghz, 32GB RAM, Windows Server 2012 R2
- The Compute Nodes - 2* Quad Core Intel Xeon E5620 2.40Ghz, 24GB RAM, 2 * NVidia TESLA M2050, Windows Server 2012 R2
- Netgear Gigabit Switch, internal network.

Novel Energy Efficient Computer Systems - example

University of HUDDERSFIELD

- Iceotope innovative liquid cooled servers for HPC
- Use liquid to convect heat from all electronics to cold plate
- Use integrated water system to harvest heat from cold plate to heat exchangers
- The Iceotope solution is scalable, energy efficient and scilent

HPC research and resources for industry – 3M Buckley Business and Innovation Centre

University of HUDDERSFIELD

- 3M BIC HPC system – Iceotope solution http://www.iceotope.com/
- The Iceotope solution overcomes standard IT challenges:
  - lack of space,
  - need for more computing power
  - cooling problems,
  - energy challenges
- It uses 3M Novec liquid cooling
- It integrates compute and cooling systems that use less energy

Inspiring tomorrow's professionals

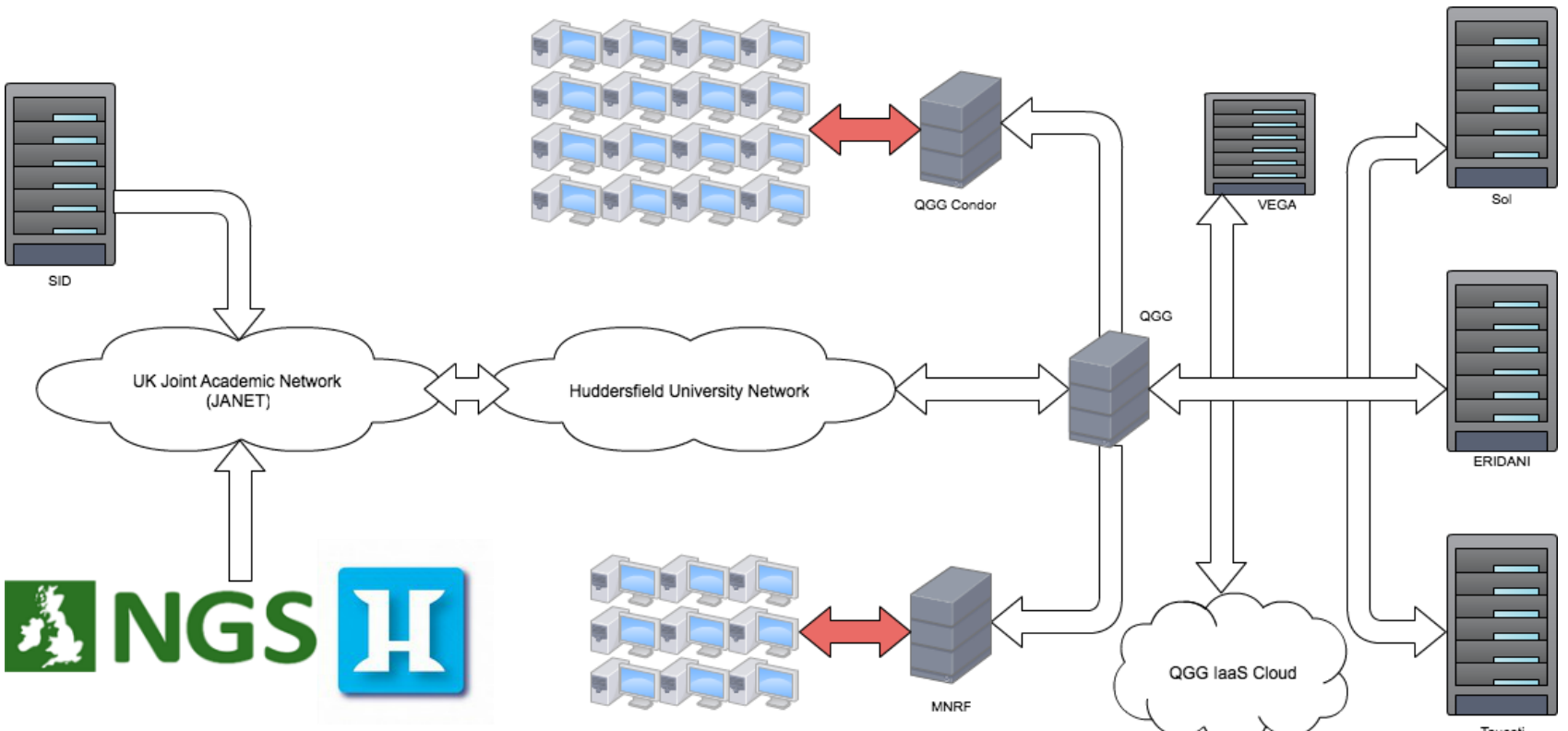
THE AWARDS AWARD WINNER UNIVERSITY OF THE YEAR

theguardian UNIVERSITY AWARDS Winner 2013

THE AWARDS 2012 WINNER Entrepreneurial University of the Year

THE QUEEN'S AWARDS FOR ENTERPRISE

University of HUDDERSFIELD



Inspiring tomorrow's professionals

THE AWARDS AWARD WINNER UNIVERSITY OF THE YEAR

theguardian UNIVERSITY AWARDS Winner 2013

2012 THE AWARDS WINNER Entrepreneurial University of the Year

THE QUEEN'S AWARDS FOR ENTERPRISE

HPC systems users at the University of Huddersfield

University of HUDDERSFIELD

- In 2010 there was a handful of HPC researchers using modest departmental HPC resource
- In 2015 there are 200 active users from:
  - Engineering 23%
  - Chemistry 25%
  - Physics 20%
  - Informatics 12%
  - Biology 5%, and others
- This dramatic increase of HPC systems users is due to the university's investment in the HPC resources and support for research from 2010-2014

Supported Applications

University of HUDDERSFIELD

- 3d Studio Max
- Abaqus
- Amber
- Blender
- Castep
- Comsol
- DL_POLY
- Fluent
- GAMESS-UK
- Gulp
- HMMER
- LAMMPS
- Metadise
- NWChem
- OpenFoam
- Opera 3D
- Matlab
- Mental Ray
- Octave

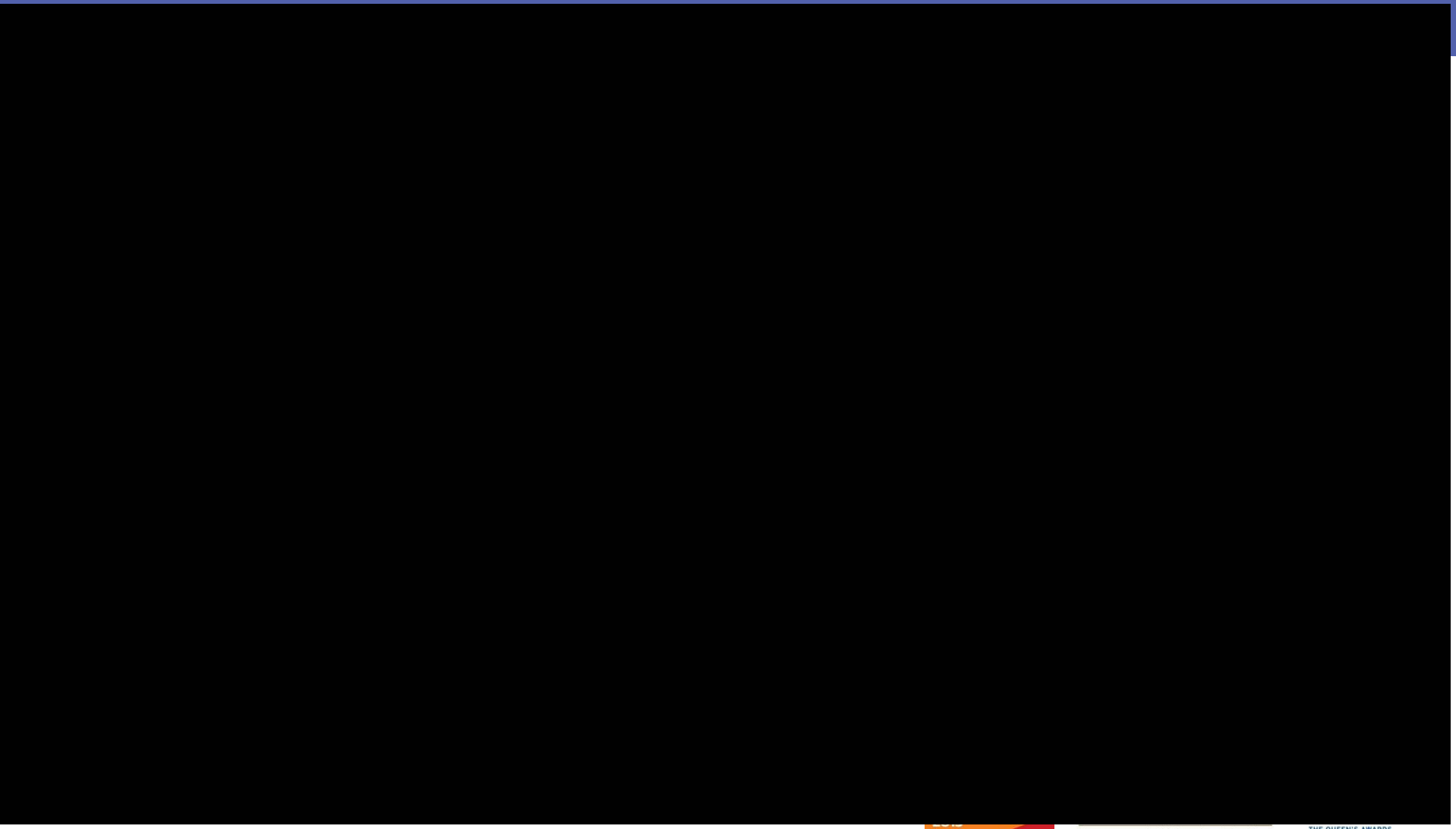
Quicker and more insightful research outcomes - examples

University of HUDDERSFIELD

- University's HPC system was used in designing a component for truck trailers to reduce drag using Computational Fluid Dynamics.
- On a single work station each simulation required 28.5 days to complete.
- It took just under 5000 simulations over 2 years to find the best working model for the product.
- This was only possible because of HPC was used to run simulations. Every simulation took between 12-18 hours to complete and about 10-15 simulations could run at the same time.
- HPC was able to do 97.5 years of computing in 2 years.
- The end product, when prototyped and tested under controlled conditions at the Mira test track, resulted in an improved fuel efficiency of 3%.

Examples of applications in Engineering

University of HUDDERSFIELD



Case Study – Sonic SETI

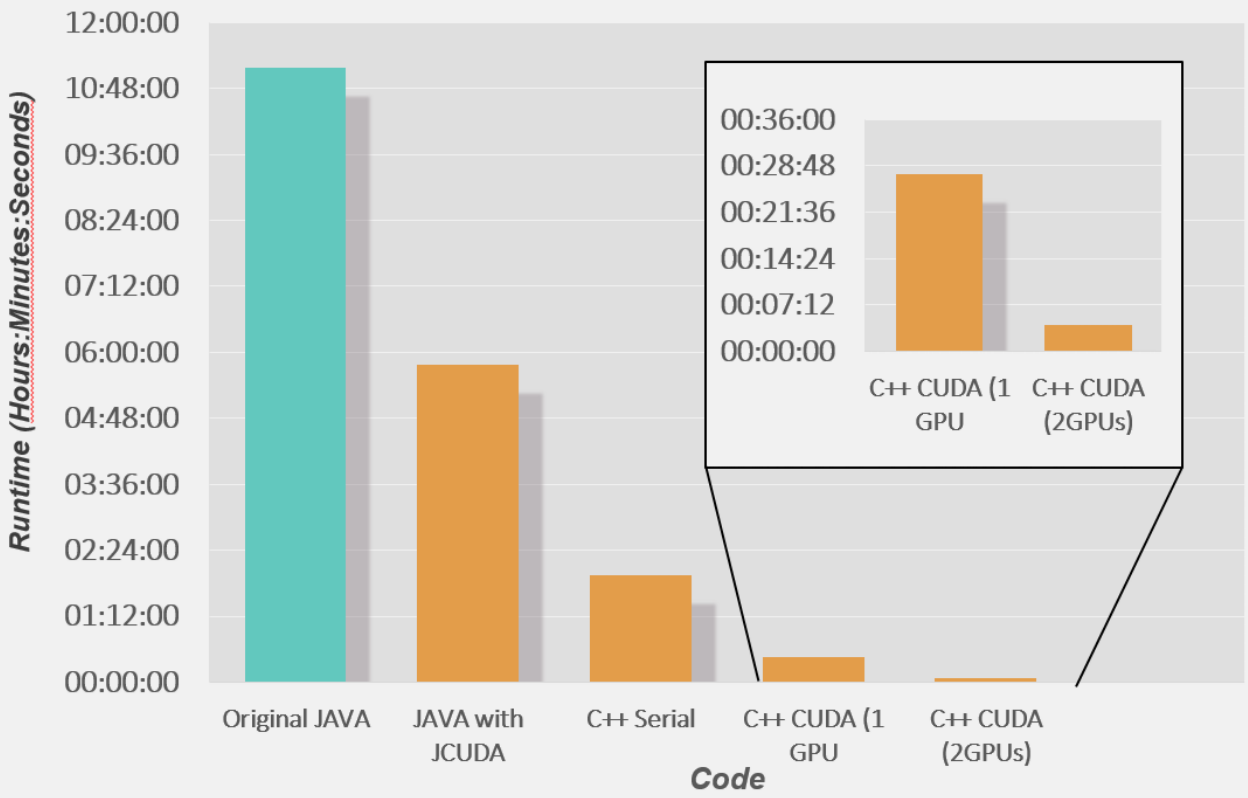
- Original software was written in JAVA.
- Runtime was deemed unacceptable (11 hours 10 minutes for 3 2GB datasets)
- Rewritten in C++
- Parallelized, using a GPU aware FFT algorithm
- Accelerated further using MPI to allow multiple GPUs to be used simultaneously

Results

University of HUDDERSFIELD

Original Sonification Code

Code produced for the case study



Big Data

University of HUDDERSFIELD

- “Digital universe” is doubling in size every two years
- By 2020 it will reach 44 trillion Gigabytes
- This is driven by the growth in connected sensors and devices – from smart phones to electricity meters, traffic systems and cars, and ubiquitous sensors - Internet of Things (IoT)
- To process this Big Data, HPC and new programming paradigms are needed

Hadoop

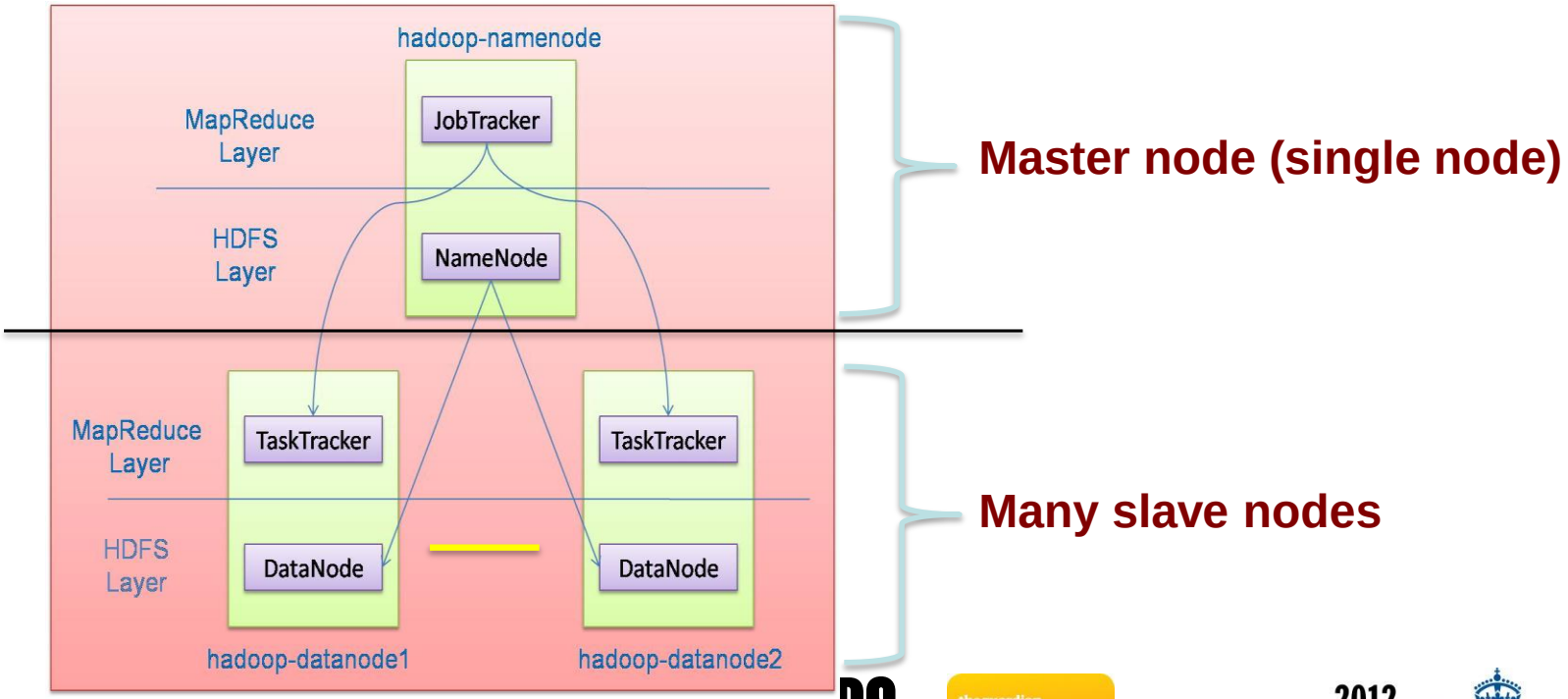
University of HUDDERSFIELD

- Hadoop as an open source framework for the storing and processing of internet-scale data in a distributed manner
- Hadoop tackles the problem of 'Big Data' by distributing both the storage and processing of data to numerous machines.
- Comprises two main components:
  - the Hadoop Distributed File System (HDFS), used for storing data across a Hadoop cluster and
  - the Map/Reduce programming framework, used to process the data

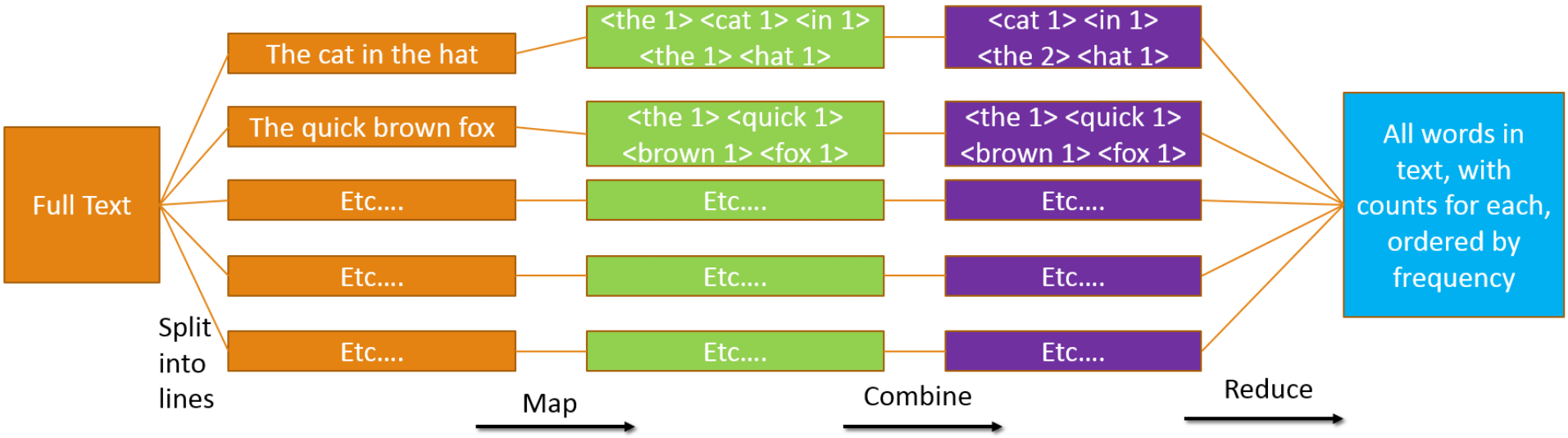
Hadoop Master/Slave Architecture

University of HUDDERSFIELD

- Hadoop is designed as a *master-slave* *shared-nothing* architecture



Example Hadoop Task - Wordcount



Hadoop Cluster at UoH

University of HUDDERSFIELD

- Hadoop cluster specification

Component	Head Node	Data Node
CPU	Intel Q8400	Intel Q8400
RAM	4GB DDR2	8GB DDR2
HDD	250GB (7200RPM)	250GB (7200RPM)

- Head node and 8 Data nodes
- Hadoop allows for a system to assess the quality of high volume of data as it scales the workload across a computer cluster.

Case Study – Close Call Data

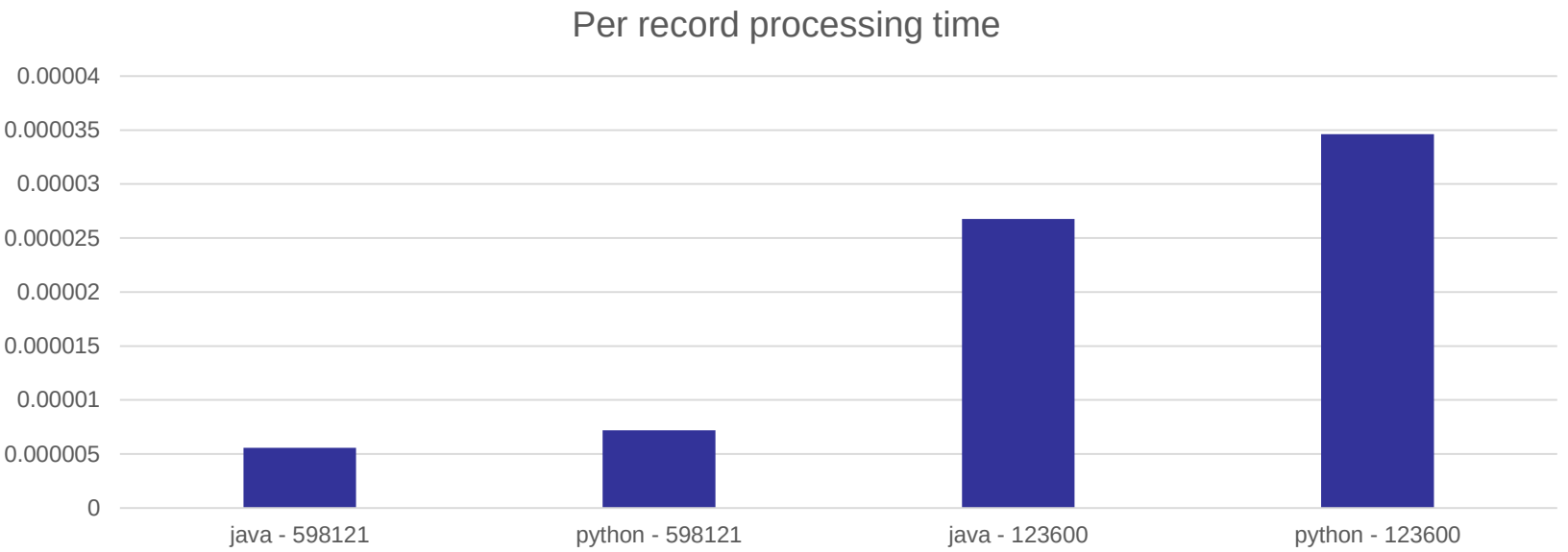
University of Huddersfield

- Unstructured free form text records pertaining to close call railway events
- Processed to identify and tag locations, dates, etc in the text
- Processing time is prohibitive (measured in days)

Close Call data processing on UoH Hadoop cluster

University of HUDDERSFIELD

- Python/Java versions were compared for efficiency



Performance evaluation

University of HUDDERSFIELD

- The program was re-implemented using java with Hadoop mapreduce
- Processing time was reduced to under 5 minutes
- There is potential for further performance improvements using GPUs

Summary

University of HUDDERSFIELD

- Scientific and design and manufacturing simulation will continue to require more powerful HPC systems
  - To increase the precision of the result
  - To get to an answer sooner
- Computer architectures will continue to get more complex, and achieving high performance will get harder.
- New Parallel Models & Algorithms are required for processing large amount of data and need to be
  - Scalable
  - Energy efficient
- New systems are required for Big data processing, storage, management and visualisation

Thank you

University of HUDDERSFIELD

Inspiring tomorrow's professionals

THE AWARDS AWARD WINNER UNIVERSITY OF THE YEAR

theguardian UNIVERSITY AWARDS Winner 2013

THE AWARDS 2012 WINNER Entrepreneurial University of the Year

THE QUEEN'S AWARDS FOR ENTERPRISE