

# Large-scale image processing on the Grid

Edward Schofield



# Overview

---

Background

What it does

How it works

How to make it scalable

# Background

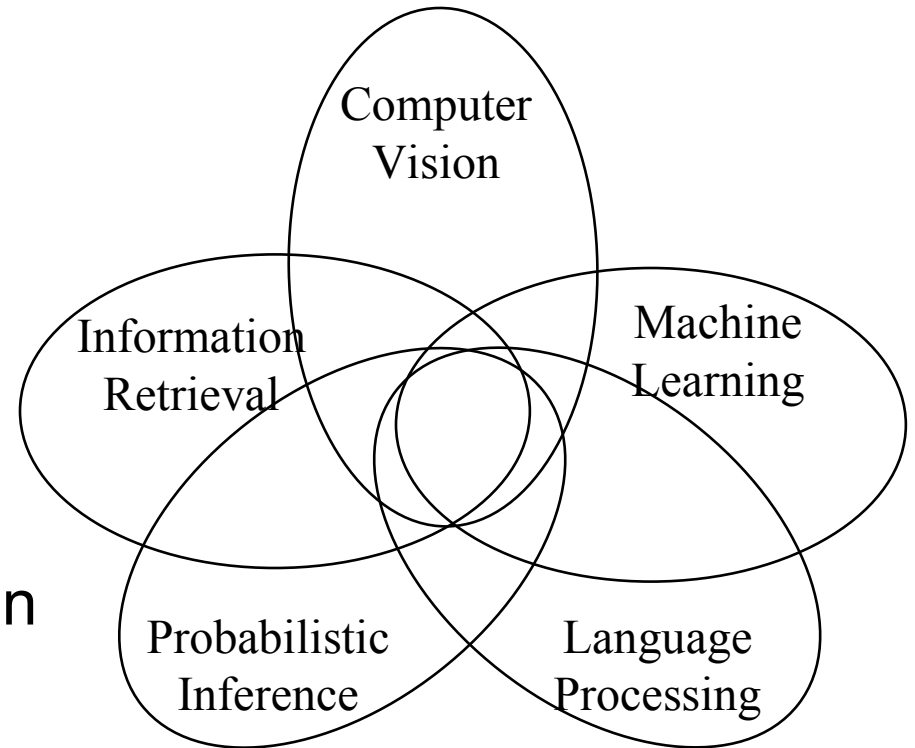
---

## Academic

- Undergrad maths at Cambridge
- PhD Computer Science at Imperial: machine learning

## Commercial

- Imense Ltd founded in 2004
- Beta website launched at BAPLA, May 2008
- New product launches coming soon



# “purple flowers with green leaves”

purple flowers with green leaves - Google Image Search - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://images.google.lk/images?svnum=10&um=1&hl=en&q=purple+flowers+with+green+leaves&btnG=Search+Images

Google

purple flowers with green leaves Search Images Search the Web Advanced Image Search Preferences

Moderate SafeSearch is on

Images Showing: All image sizes Results 1 - 18 of about 304,000 for [purple flowers with green leaves](#). (0.10 seconds)

View all web results for [purple flowers with green leaves](#)

... purple flowers, and green leaves ...  
517 x 768 - 131k - jpg  
[www.dkimages.com](#)

... purple flowers and green leaves ...  
311 x 424 - 21k - jpg  
[www.dkimages.com](#)  
[ [More from www.dkimages.com](#) ]

Purple flower in the middle of green ...  
380 x 253 - 85k - jpg  
[www.istockphoto.com](#)

Large Purple Rhododendron Flower ...  
380 x 253 - 46k - jpg  
[www.istockphoto.com](#)  
[ [More from www1.istockphoto.com](#) ]

... PURPLE FLOWERS AND GREEN LEAVES  
432 x 432 - 9k - jpg  
[www.funkandjunk.com](#)

... tree with purple flowers and ...  
768 x 1024 - 138k  
[www.upscalesale.com](#)

... Deep purple flowers above green ...  
300 x 300 - 25k - jpg  
[www.peoriagardens.com](#)

... green leaves. Purple flowers ...  
300 x 300 - 18k - jpg  
[www.peoriagardens.com](#)  
[ [More from www.peoriagardens.com](#) ]

Three complete flowers on green and ...  
864 x 385 - 477k - jpg  
[www.ucalgary.ca](#)

Three flowers: purple flowers with ...  
864 x 560 - 709k - jpg  
[www.ucalgary.ca](#)

... leaves and pink-purple flowers, ...  
500 x 333 - 56k - jpg  
[www.nativeplants.org](#)

... dark green leaves and bright ...  
540 x 360 - 60k - jpg  
[www.nativeplants.org](#)

Find: Next Previous Highlight all

Done

# “purple flowers with green leaves”

## Image Search Demo

Results per page



Content



Text



Combined

Displaying 1 - 21 of 3565 images that matched (in 35 ms) query 'purple flowers with green leaves'

[Next >>](#)



1. <http://www.photo.net/phot..>



2. <http://www.pbase.com/imag..>



3. <http://public.fotki.com/v..>



4. <http://www.pbase.com/imag..>



5. <http://www.flickr.com/pho..>



6. <http://www.flickr.com/pho..>



7. <http://www.pbase.com/imag..>



8. <http://www.photo.net/phot..>



9. <http://www.flickr.com/pho..>



10. <http://www.flickr.com/pho..>



11. <http://www.photo.net/phot..>



12. <http://commons.wikimedia...>



13. <http://www.flickr.com/pho..>



14. <http://www.pbase.com/imag..>



15. <http://www.photo.net/phot..>



16. <http://commons.wikimedia...>



17. <http://www.freefoto.com/p..>



18. <http://www.pbase.com/imag..>



19. <http://www.pbase.com/imag..>



20. <http://www.flickr.com/pho..>



21. <http://www.flickr.com/pho..>



# “children indoors”

children indoors - Google Image Search - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://images.google.lk/images?svnum=10&um=1&hl=en&q=children+indoors&btnG=Search+Images

Google children indoors Search Images Search the Web Advanced Image Search Preferences

Moderate SafeSearch is on

Images Showing: All image sizes Results 1 - 18 of about 234,000 for **children indoors**. (0.09 seconds)

View all web results for **children indoors**

... to keep their **children indoors**, ... that 21st century **children** are ...  
360 x 288 - 63k  
[www.ilovemybaby.org](http://www.ilovemybaby.org)

PlayStation: Keeps **children indoors**  
228 x 310 - 7k - jpg  
[www.dailymail.co.uk](http://www.dailymail.co.uk)

Child Size Teak Garden Bench - ...  
490 x 440 - 31k - jpg  
[www.thepierceco.com](http://www.thepierceco.com)

... have kept **children indoors**, ...  
600 x 450 - 173k - jpg  
[www.tomorrowproject.net](http://www.tomorrowproject.net)

Keep the **Children Indoors!**  
490 x 495 - 75k - jpg  
[theoddsandsods.com](http://theoddsandsods.com)

Keep the **Children Indoors!**  
321 x 258 - 60k - jpg  
[theoddsandsods.com](http://theoddsandsods.com)

... keeping their **children indoors**.  
450 x 250 - 17k - jpg  
[angelingo.usc.edu](http://angelingo.usc.edu)

... to keep **children indoors**.  
300 x 239 - 12k - jpeg  
[mothergaia.blog.ca](http://mothergaia.blog.ca)

... keep their **children indoors** to ...  
212 x 334 - 34k - jpg  
[www.100blackmenba.org](http://www.100blackmenba.org)

... you and the **children indoors**, ...  
200 x 236 - 13k - jpg  
[kcsos.kern.org](http://kcsos.kern.org)

Figure 3: **Children** experience high ...  
545 x 372 - 29k - jpg  
[www.hedon.info](http://www.hedon.info)

Find: Next Previous Highlight all

Done

# “children indoors”

## Image Search Demo

Results per page



Content



Text



Combined

Displaying 1 - 21 of 30 images that matched (in 14 ms) query 'children indoors'

[Next >>](#)

1. <http://www.pbase.com/imag..>



2. <http://www.pbase.com/imag..>



3. <http://www.pbase.com/imag..>



4. <http://www.pbase.com/imag..>



5. <http://www.pbase.com/imag..>



6. <http://www.pbase.com/imag..>



7. <http://www.pbase.com/imag..>



8. <http://www.pbase.com/imag..>



9. <http://www.pbase.com/imag..>



10. <http://www.pbase.com/imag..>



11. <http://www.pbase.com/imag..>



12. <http://www.pbase.com/imag..>



13. <http://www.pbase.com/imag..>



14. <http://www.pbase.com/imag..>



15. <http://www.pbase.com/imag..>



18. <http://www.pbase.com/imag..>



21. <http://www.pbase.com/imag..>



# “beach without people”

beach without people - Google Image Search - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://images.google.co.uk/images?hl=en&q=beach+without+people&btnG=Search+Images&gbv=2

Google

beach without people



















Search Images Search the Web

Advanced Image Search Preferences

Moderate SafeSearch is on

Images Showing: All image sizes

Results 1 - 18 of about 10,500,000 for **beach without people**. (0.15 seconds)

 <p>... too much time <b>without people</b> ... 2048 x 1536 - 432k - jpg darylcook.com</p>	 <p>... <b>Beach</b> lots <b>without people</b> and ... 640 x 512 - 54k - jpg www.simtropolis.com</p>	 <p>We started from Railay <b>beach without</b> ... 568 x 401 - 65k - jpg www.the-andaman-eye.net</p>	 <p>The new 'Pink <b>Beach</b>' is an exclusive ... 1108 x 1424 - 263k - jpg img.hexus.net</p>	 <p>Wreck <b>Beach</b>, British Columbia 640 x 480 - 47k - jpg www.cbc.ca</p>	 <p>... <b>without people</b> in it; the <b>beach</b> ... 700 x 525 - 113k - jpg plasticdaffodils.wordpress.com</p>
 <p>... <b>without people</b> in it; the <b>beach</b> ... 800 x 600 - 85k - jpg plasticdaffodils.wordpress.com</p>	 <p>... the <b>beach without</b> worrying about ... 289 x 225 - 14k - jpg www.time.com</p>	 <p>... the <b>beach without</b> looking back. 640 x 480 - 17k - jpg www.worldriders2.com</p>	 <p>... found nice <b>beach without people</b> ... 1024 x 696 - 13k - jpg picasaweb.google.com</p>	 <p>A pretty <b>beach without</b> a lot of ... 1600 x 1200 - 43k - jpg picasaweb.google.com</p>	 <p>... and the <b>beach without</b> including ... 480 x 360 - 45k - jpg blog.planetc1.com</p>
					

http://images.google.co.uk/imgres?imgurl=http://img.hexus.net/v2/gaming/screenshots\_pc/lara1\_large.jpg&imgrefurl=http://www.hexus.net/content/item.php?item=9179&h=1424&w=1108&sz=...



# “beach without people”

## Image Search Demo

Results per page

☐ Content☐ Text☒ Combined

Displaying 1 - 21 of 25345 images that matched (in 41 ms) query 'beach without people'.

[Next >>](#)

1. <http://www.freefoto.com/p..>



2. <http://www.freefoto.com/p..>



3. <http://www.freefoto.com/p..>



4. <http://www.freefoto.com/p..>



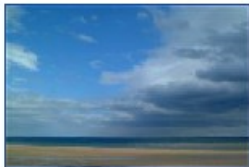
5. <http://www.freefoto.com/p..>



6. <http://www.freefoto.com/p..>



7. <http://www.freefoto.com/p..>



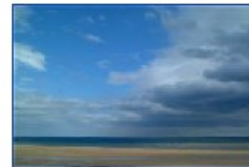
8. <http://www.freefoto.com/p..>



9. <http://www.freefoto.com/p..>



10. <http://www.freefoto.com/p..>



11. <http://www.freefoto.com/p..>



12. <http://www.freefoto.com/p..>



13. <http://www.flickr.com/pho..>



14. <http://www.freefoto.com/p..>



15. <http://www.freefoto.com/p..>



16. <http://www.freefoto.com/p..>



17. <http://www.freefoto.com/p..>



18. <http://www.freefoto.com/p..>



19. <http://www.freefoto.com/p..>



20. <http://www.freefoto.com/p..>



21. <http://www.freefoto.com/p..>

# What it does

---

Automatic image classification and recognition:

“search inside the picture”

Regions: sand, skin, sky, hair, glass

Scenes / concepts: beach, market, basketball, Christmas

Face recognition: age, sex, ethnicity

# Applications

---

automated analysis and retrieval of un-annotated images

improved retrieval of pre-annotated images

automated image annotation (“keywording”)

industry needs: [media](#), [advertising](#), [satellite imaging](#) ...



# Problems with metadata

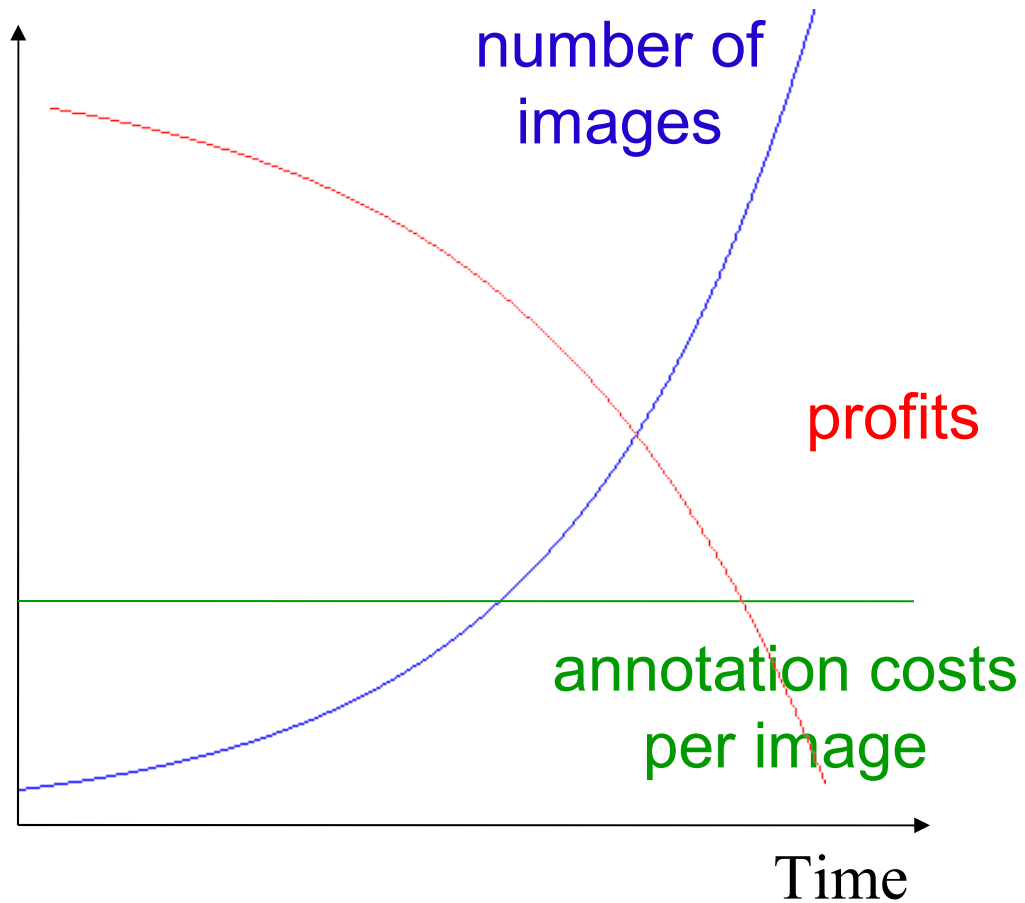
---

manual image annotation:

- expensive and tedious
- inconsistent and error prone
- language-specific
- homonymy: keywords can be ambiguous
- no spatial context
- not object-specific
- binary: no relevance assessment

# The future

---



**“Evolve or die”**

# The Imense approach

---

## Manual image annotation:

- expensive and tedious
- inconsistent and error prone
- language specific
- ambiguous keywords
- no spatial context
- not object-specific
- no relevance assessment

## Automated image analysis:

- automated, therefore cheaper
- consistent and probabilistic
- language independent (concepts)
- clear semantics
- spatial localisation
- object specific
- relevance is quantifiable



# How it works

---

- many thousands of labelled images as training data
- machine learning / statistical classifiers
- rich semantic and visual ontologies
- ontology-based query language

# Image analysis



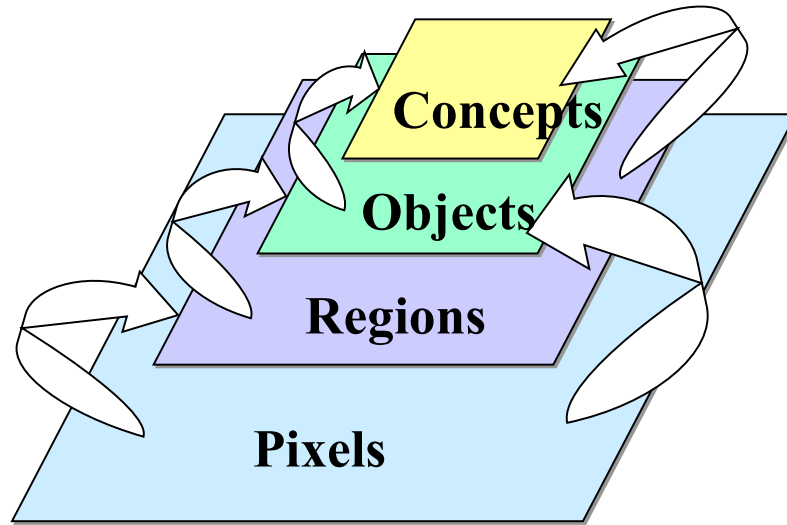
**Ontological Description  
Process**

Edges, pixel, colour,  
textures, shapes,  
shading etc  
(sun above mountain)

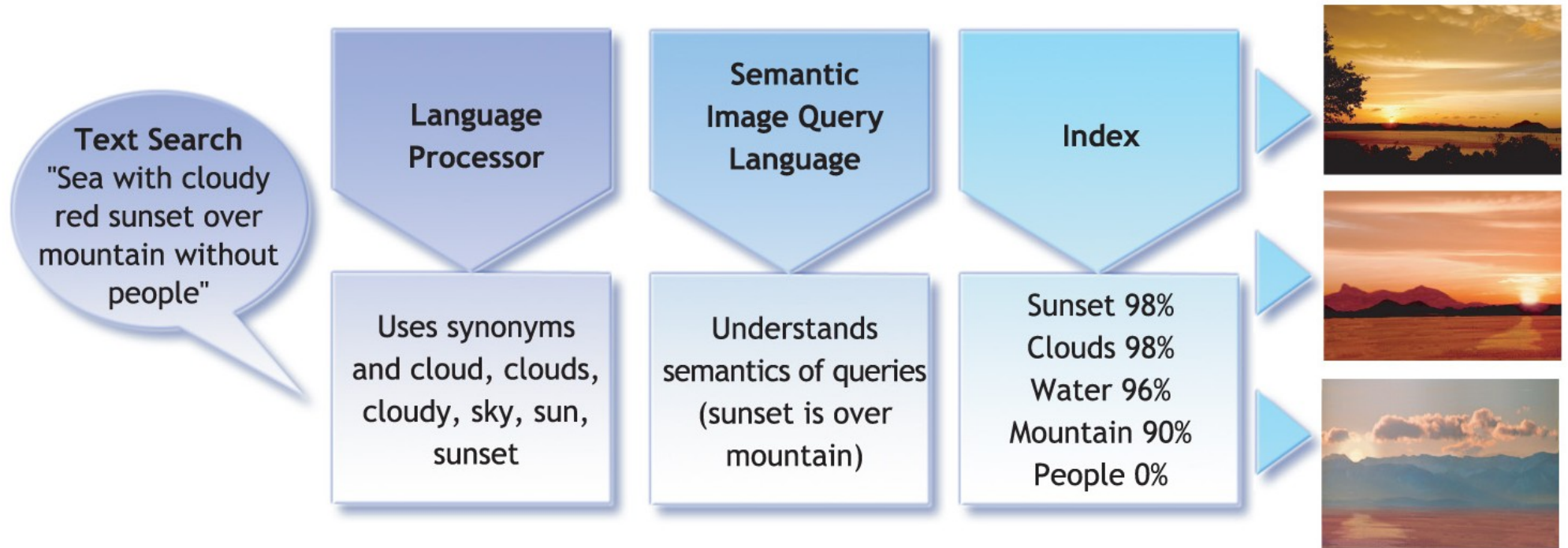
**Statistical  
Class Matching**

Sunset 98%  
Clouds 98%  
Water 96%  
Mountain 90%

**Index**



# Image retrieval





# Scalability

---

Image analysis and classification is

- computationally expensive
- easy to parallelize (at the image level):  
no interdependencies, no hard scheduling requirements

The **High-Energy Physics (HEP)** community has created a worldwide computing Grid to meet the requirements of the **Large Hadron Collider** at CERN.

This Grid provides access to tens of thousands of CPUs, with UK participation through the GridPP collaboration

# Grid Computing – CERN data centre

---

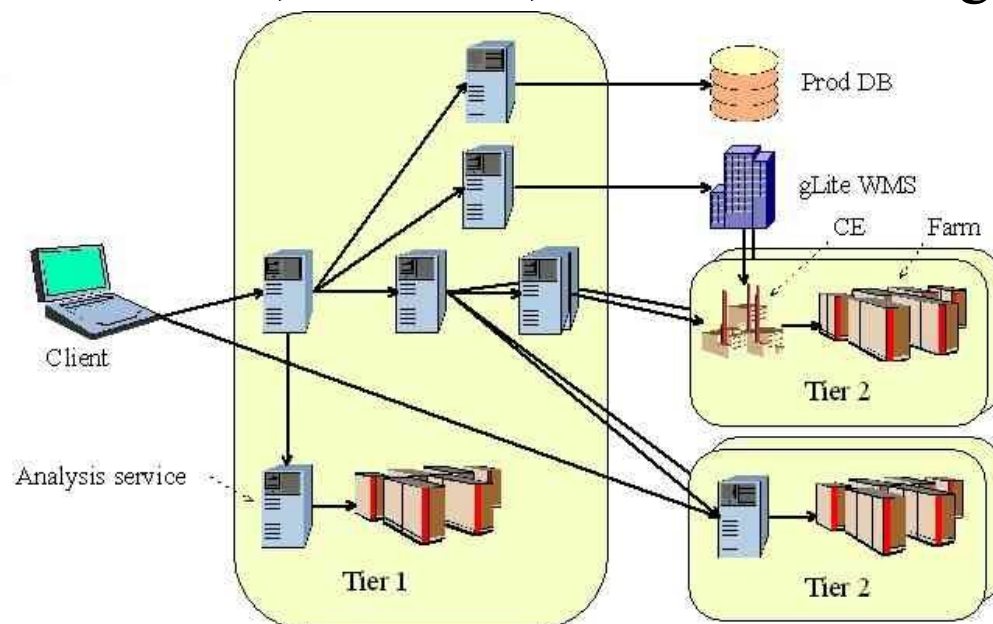
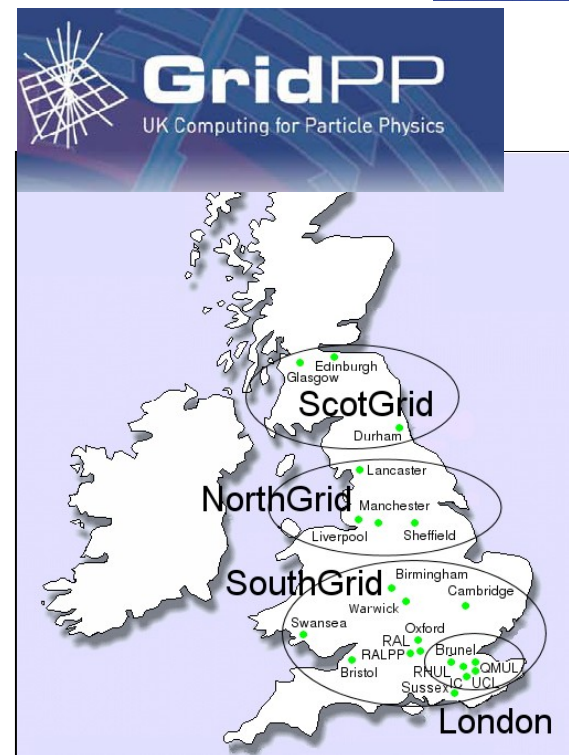


# Grid computing: overview

Collaboartion with the Cambridge eScience Centre and the HEP group at the Cavendish (Prof Andy Parker)

Have access to CamGrid, GridPP (UK),  
and EGEE (Europe) grids

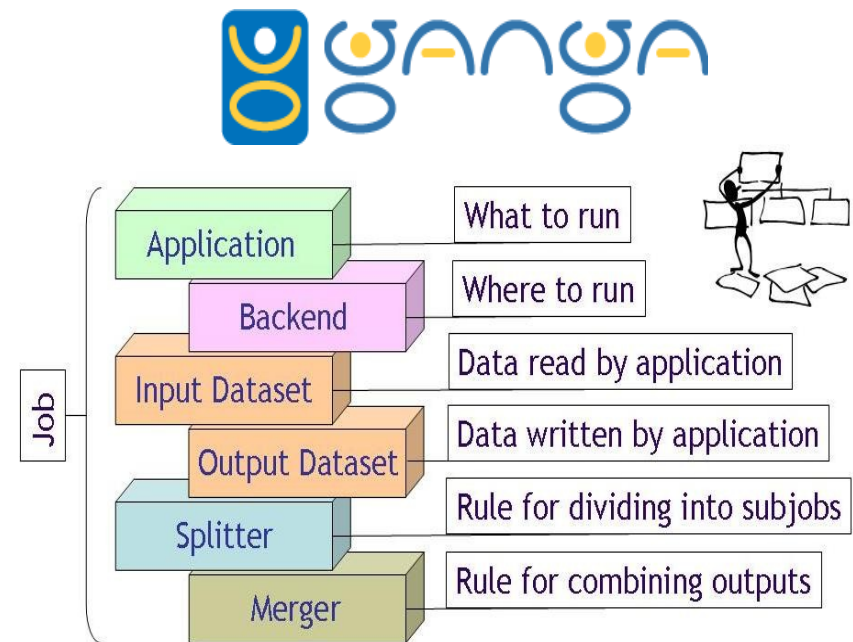
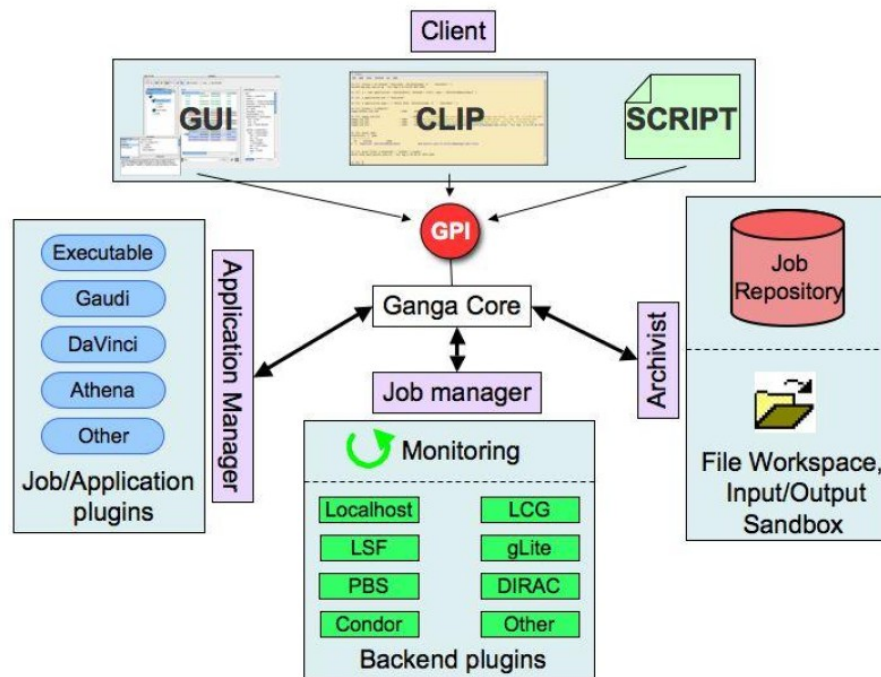
About 100,000 CPU, 100 PB of storage





# Grid computing – software infrastructure

The **Cambridge HEP group** has contributed in building **Ganga** for managing computer jobs and simplifying use of the Grid



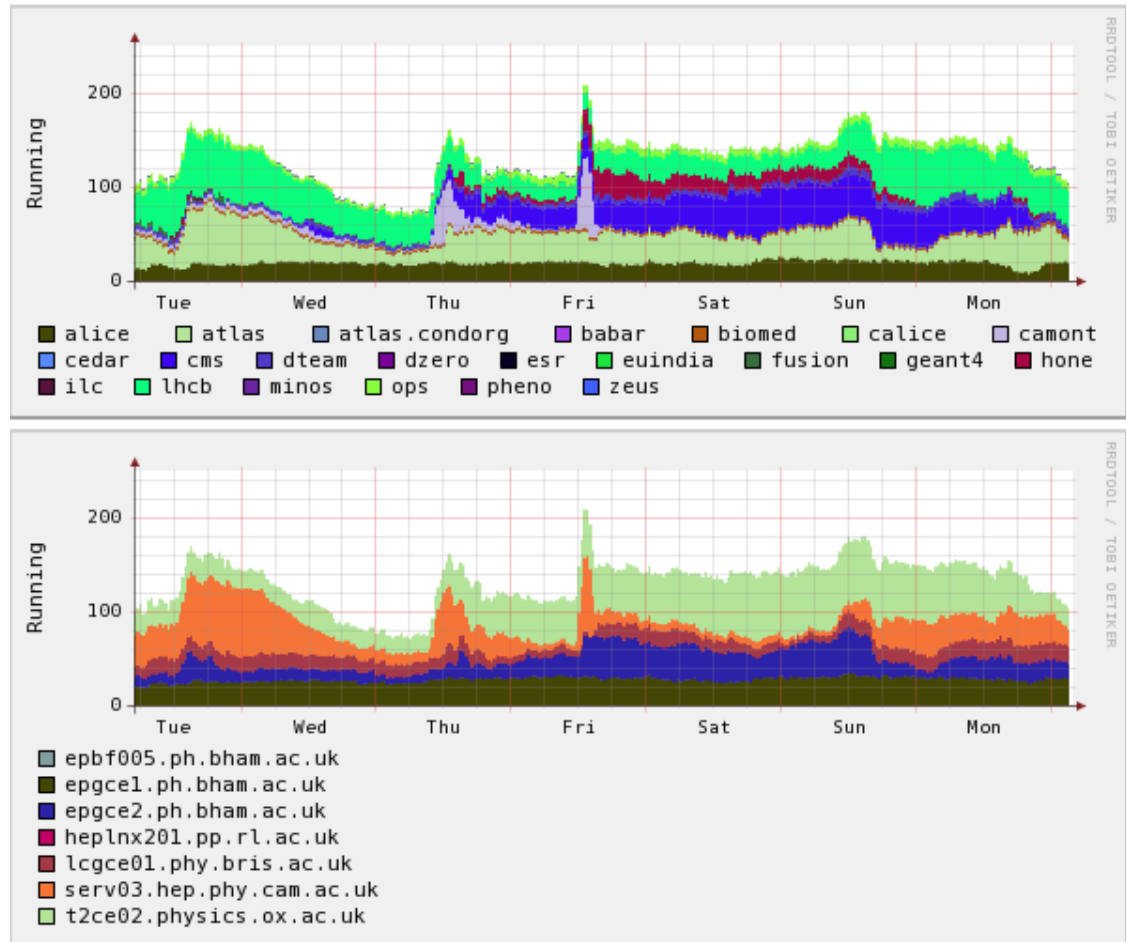


# Grid computing – immense collaboration

Results so far:

- 5 million images processed
- 6000 jobs completed
- Scalability to >10 million images/month
- Ported middleware to non-SL Linux
- Improvements to Ganga (features + usability)

[ScotGrid](#) [NorthGrid](#) [London T2](#) [South Grid](#) [RAL T1](#) [FR](#) [TW](#) [UK](#) [BE](#) [EDU](#) [IT](#) [CH](#) [DE](#) [NL](#)





Search

Printer-Friendly

[Home](#)

[About Us](#)

[Knowledge Exchange](#)

[Grants and Awards](#)

[Our Facilities](#)

[Our Research](#)

[Press and Media Centre](#)

[Press Releases](#)

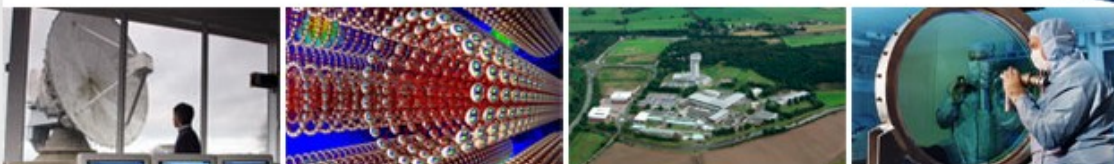
[Press Office Contacts](#)

[Press and Media Events](#)

[Publications and Image  
Gallery](#)

[About the Council](#)

[The Big Questions](#)



[Home](#) > [Press and Media Centre](#) > [Press Releases](#) > [STFC](#)

## Grid helps find one picture in a million

*Looking for images on the internet can be a frustrating business. Whether you want the perfect sunset over the sea or the London skyline by night, you're dependent on people to describe the images on their web pages. Now Imense Ltd, a high-tech Cambridge start-up, has announced new investment to help them become the Google of image searching, using their revolutionary technology.*

To test their software, they've made an unexpected partnership with a group of particle physicists using a massive computer Grid.

Professor Keith Mason, Chief Executive of the Science and Technology Facilities Council (STFC) describes how this happened "We actively encourage the researchers we fund to consider the wider applications of the work they do. In this case, computing problems that had to be addressed for particle physics can also be used to

Business Weekly

## AT&T start-up unveils image search technology

Contributed by Business Weekly  
Wednesday, 21 March 2007

The latest in a prodigiously long list of companies to be started by former researchers from has unveiled a technology aimed at bringing order to an increasingly image-obsessed world. Cambridge Ontology, established by former AT&T researchers, Drs Christopher Town and an impact on the multi-billion pound digital search industry, by solving a technology problem has not yet been able to solve.

### Latest



[News and Press releases](#)  
[Events](#)  
[STFC in the news](#)

### Resources



[Image library](#)

[Publications](#)

**BBC  
NEWS**

[WATCH LIVE](#) **BBC News 24**

[News Front Page](#)

[World](#)

[UK](#)

[England](#)

[Northern Ireland](#)

[Scotland](#)

[Wales](#)

[Business](#)

[Politics](#)

[Health](#)

[Education](#)

[Science/Nature](#)

Last Updated: Wednesday, 5 September 2007, 08:34 GMT 09:34 UK

[E-mail this to a friend](#)

[Printable version](#)

### Hot picks: UK tech start-ups

**The UK invests more than any other European Country in new ideas, with more than £1bn being pumped into start-ups firms.**

Here the BBC identifies some of the UK's hottest emerging companies.

[INSCENTINEL](#)

[IMENSE](#)

[MONITISE](#)

[IONSSCOPE](#)

# Thanks!

