

The world's largest distributed computer grid crunches LHC's huge numbers

March 13 2015, by Tom Whyntie



The CERN datacentre is the ground zero, but only part, of a worldwide computing grid. Credit: Maximillien Brice/CERN, CC BY-NC-ND

The world's largest science experiment, the Large Hadron Collider, has potentially delivered one of physics' "Holy Grails" in the form of the [Higgs boson](#). Much of the science came down to one number – 126, the Higgs boson's mass as measured in gigaelectronvolts. But this three-digit number rested upon something very much larger and more complicated: the more than 60,000 trillion bytes ([60 petabytes](#)) of data produced by colliding subatomic particles in four years of experiments, and the enormous computer power needed to make sense of it all.

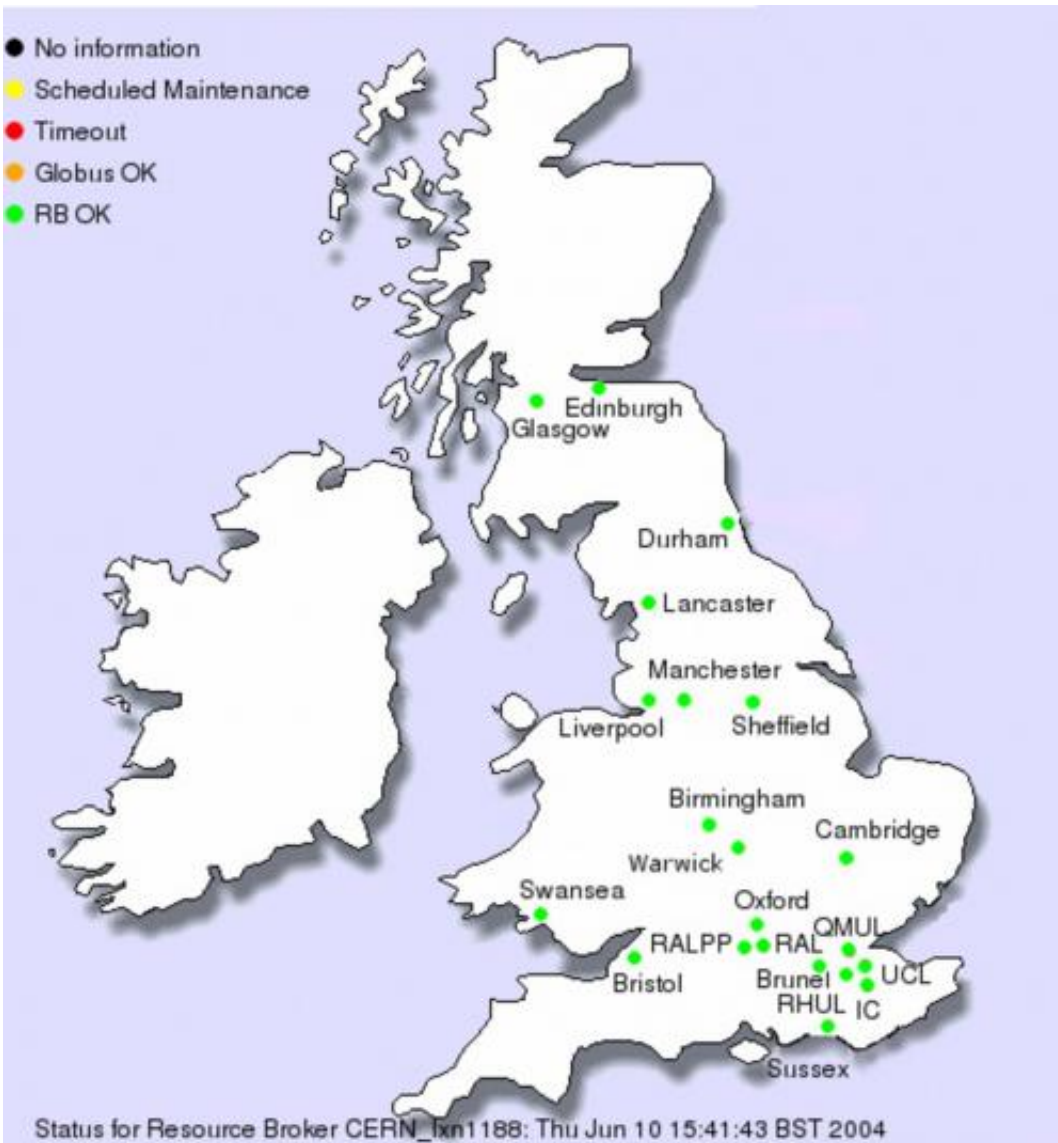
There is no single supercomputer at CERN responsible for this task. Aside from anything else, the political faffing that would have ensued from having to decide where to build such a machine would have slowed scientific progress. The actual solution is technically, and politically, much more clever: a [distributed computing grid](#) spread across [academic facilities around the world](#).

Many hands make lighter work

This solution is the [Worldwide LHC Computing Grid](#) (WLCG), the world's largest distributed computing grid spread over 174 facilities in 40 countries. By distributing the computational workload around the planet, the vast torrents of precious particle data streaming from the collider can be delivered, processed, and pored over by thousands of physicists regardless of location or time of day or night.

CERN's datacentre is considered Tier 0 and is linked by [dedicated fast fibre-optic links](#) to 15 Tier 1 facilities in Europe and the US, and [a further 160 Tier 2 facilities](#) around the world. At Tier 0 the rate of data throughput hits around 10GB/s – about the equivalent of filling two DVDs every second.

During the first "season" of experiments on the LHC, now known as Run 1, the WLCG used up to 485,000 computer processing cores to crunch its way through around 2M sets of calculations a day. Around 10% of this number-crunching was performed by the [GridPP Collaboration](#), the UK's contribution to the WLCG funded by the Science and Technology Facilities Council (STFC). Today Tier 0 is processing around [one million billion bytes](#) (a petabyte, or 1PB) every day – equivalent to about 210,000 DVDs.



UK universities linked up to the GridPP. Credit: GridPP

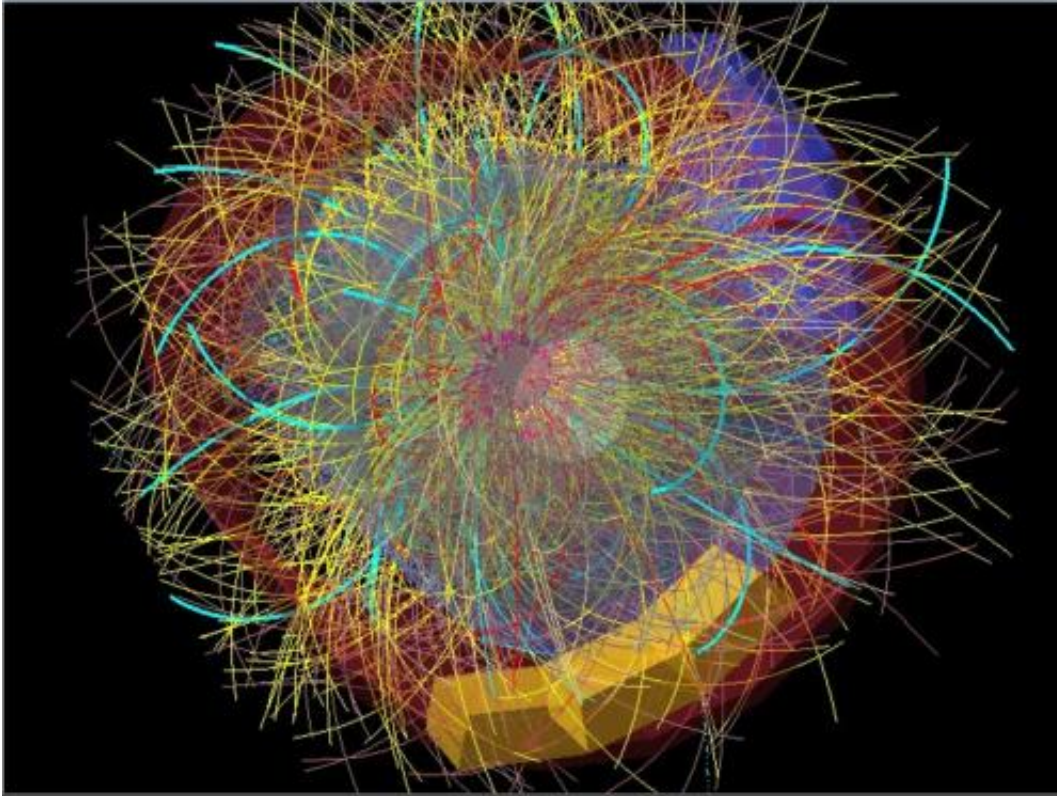
But the grid has grown into something more - an expert community that has tirelessly turned technology into ground-breaking physics results. Now, with [Run 2 and a second season of LHC experiments](#) due to start this month, the same experts will need to manage even greater amounts of data produced by particle collisions of even greater energy.

Harder, better, faster, stronger

Not only will Run 2 nearly double the experiments' collision energy in order to probe theories such as supersymmetry, extra dimensions and magnetic monopoles – this round of humans vs protons will result in almost three times as many collisions per second in the collider. This increase will allow the properties of the Higgs boson to be studied in greater detail, perhaps even giving some understanding of why the particle that gives mass to others also has mass of its own.

However, the debris left by exploded hadrons was hard enough to pick through last time – left as it was, the grid would have required six times the computational capacity in order to cope with the size of the figurative haystack in which physicists are looking for needles. But the grid has been upgraded [alongside the experimental apparatus](#) to cope with demand.

Evolution, not revolution



Plotting the path of every particle and fragment generates data by the warehouse-full. Credit: ALICE/CERN

New techniques introduced to cope with the experiments' demands include [multi-core processing](#). In order to compensate for the diminishing advances in processor speed, multi-core CPUs – processors designed as two, four or even eight CPUs in a single package – are being rolled out as worker nodes throughout the grid.

This has meant physicists have to rewrite their code to be multi-threaded in order to take advantage of the multiple cores by sending them tasks in parallel, but the result is much improved processing speeds. The grid then has to cleverly manage how these tasks are shared within a single site – not a trivial task when each site typically has thousands of nodes.

The huge amount of data transferred between sites also puts a burden on networks. This has been reduced by using [xrootd](#), a high-level protocol that provides a means for scientists to access the huge datasets stored across the grid in the most network-efficient way possible. By implementing a [dynamic data placement](#) policy, the grid can learn how many copies to make of popular datasets and where best to put them for optimum performance.

It's hard to say if Run 2 will give us answers to life, the universe, and everything. There are certainly a lot of scientists whose careers depend on some kind of new physics emerging from the four experimental [detectors](#) spaced around the LHC's 27km circuit. Some will find what they're looking for; others will not. But they will all rely for their work on the expertise of the computing technology team who support the world's largest planet-wide computing network for many years to come.

This story is published courtesy of [The Conversation](#) (under Creative Commons-Attribution/No derivatives).

Source: The Conversation

Citation: The world's largest distributed computer grid crunches LHC's huge numbers (2015, March 13) retrieved 18 April 2024 from <https://phys.org/news/2015-03-world-largest-computergrid-crunches-lhc.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--