Jason Stowe, CEO at Cycle Computing

Pharmaceutical workloads are pleasantly parallel in that they don't depend on having access to lots of nodes running the exact same job at the same time. We are now in the era of utility supercomputing where much of the computational work can be done on demand infrastructures such as data centres, virtualisation environment and public clouds, which are all particularly suitable for pleasantly parallel workloads.

Amazon web services is one such example being used by pharma researchers due to the fact not only can the service create 30,000 processors to run 30,000 individual analyses at scale, it can do so affordably.

The impact of this is significant; researchers no longer have to wait weeks or months to get results back from certain classes of analysis. When you take into account the fact that there is a cost associated to each day it takes to bring a drug to market, the benefit of speeding up the process is clear. Beyond that, however, the scale of computational possibilities means that not only do jobs get completed faster, but researchers have the time to ask new questions.

Companies will often have an archive of historical data, such as information gathered from mass spectrometers, that would have been analysed at the time but can now have newer algorithms applied if the appropriate level of resources are available. The internal cluster of that company may be perfectly adequate for running the daily analyses, but not for going through that past data. This is where on-demand services can be invaluable.

When there is a significant investment in cluster technology, several generations of hardware will often be deployed and so regular evaluations should be performed on the cluster's utilisation. This allows users to determine whether they should take advantage of ancillary sources for burst, rather than trying to buy for peak, and ensures that resources aren't ever lacking or being wasted.

Sumit Gupta, director of Tesla Product Marketing at Nvidia

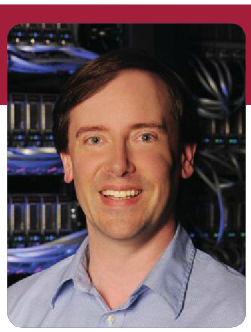
The definition of high-performance computing has changed. There was a time that the term would only be applied to supercomputers but now even workstations fall into that category and this new definition means that every stage of the pharmaceutical process uses HPC in some way.

The genomics work being done in the earlier stages of the drug lifecycle involves researchers drilling down to a finer level of detail in order to gain an understanding of the genetic structure of living things and how biomolecules behave with certain drug candidates. The molecular structure of the biomolecule/protein may have previously been done through methods such as X-ray crystallography, but genomics and high-speed sequencing machines mean that the genes of individual people can be sequenced to aid understanding of the genetic origin of specific diseases. A great amount of HPC focus is being directed in this area because while clinical trials or drug application reviews at the FDA can't be accelerated, the discovery phase can.

The discovery phase of drug manufacture can take up to five years and often the results will be a drug that looks promising but that when tested in animal studies or clinical trials is found to have too many side effects. The solution is to narrow down the candidates. Going over millions of compounds in a laboratory would be far too costly in terms of both human labour and time, however computational methods can reduce the field of prospects, which can then be taken to preclinical and clinical studies.

We recently worked with BGI, the world's largest genomics institute, who bought 300 high-speed sequencers in order to address the problem of China's aging population. The result was that the institute was soon drowning in petabytes of data. Tesla GPUs accelerated the software so that they could crunch through the data more effectively and they soon began to make interesting discoveries that would not have previously been possible. That's the main impact high-performance computing has on the pharmaceutical industry.

Geoffrey Noer, senior director of Product Marketing at Panasas



The workload we see far more than any other in life sciences is next-generation sequencing. This has driven a tremendous explosion in need for data storage and has been at the forefront of the demand for HPC resources in the bio-pharmaceutical space. Sequencers are far less expensive than they used to be and as a result institutions will often have multiple sequencers, with each one churning out vast amounts of data. There is a broad movement to compute clusters of x86 servers to do the processing of that data, which in turn drives the need for more storage – not only to house the initial results, but to act as an on-going repository. Scientists will often want to go back and do further analysis on the data and so it needs to be readily available; ensuring that happens typically falls to the IT department.

I would say that the most difficult aspects of being an IT professional in this space are the resource planning and being able to act fast enough. Having spoken with many IT directors in the past, news of a new sequencer being installed is often given with very little advance notice, which leaves them in the position of having to figure out how to immediately provide tens or even hundreds of terabytes of additional storage. Under these circumstances, having a single file system that can be scaled up incrementally as needed is crucial.

The term Big Data is being used to describe the phenomenon of the quantum shift in the amount of data being generated and there's no doubt that genomics and bio-pharma are part of that trend. The key advice is to thoroughly evaluate the type of architecture that will best suit long-term needs and solve any scalability problems, not only for capacity and performance but in a way that maintains ease of use, manageability and reliability.