IBM and Lab7 Cloud for data-intensive scientific workloads

A tightly integrated hybrid cloud solution that delivers high performance and cost efficiency

Highlights

- Dramatically speeds up scientific workloads that require processing of massive volumes of data
- Draws from the global IBM Bluemix data center footprint and Lab7 Systems' optimization services to ensure a deployment that is quick, simple and effective
- Integrates well with on premises environments
- Delivers greater flexibility, performance and efficiency than commodity cloud offerings

Organizations handling genomics and other data-intensive scientific workloads face a unique set of challenges when it comes to meeting their data processing needs. With a single genome generating as much as one terabyte of data, these organizations must be able to process massive amounts of data quickly in order to meet their scientific goals.

Cloud computing, with its rapid scalability, would seem to be a natural fit for these organizations. However, the commodity cloud offerings available on the market today just aren't built with the needs of science in mind. Many scientific organizations will find that with the massive volumes of data they need to process, they could easily end up paying more to use a commodity cloud than they would if they just processed the data on premises.

In addition, commodity clouds lack the flexibility and control that genomics organizations need. While a one-size-fits-all approach to cloud may be perfectly acceptable in some areas, genomics just isn't one of them. These organizations need to work with a cloud partner that won't just deploy a generic cloud solution and then leave them to their own devices. Instead, they need integrated software and storage designed specifically with genomics best practices in mind, as well as optimization and support services to help the whole cloud environment run with as much speed and efficiency as possible.

IBM is working with its partner Lab7 Systems to deliver just what today's scientific organizations need in a cloud solution. The Lab7 High-Performance Genomic Cloud was built on the IBM® Bluemix® cloud platform and designed specifically for genomics and other data-intensive sciences. This new offering can help organizations achieve better results than they possibly could using a commodity cloud alone.

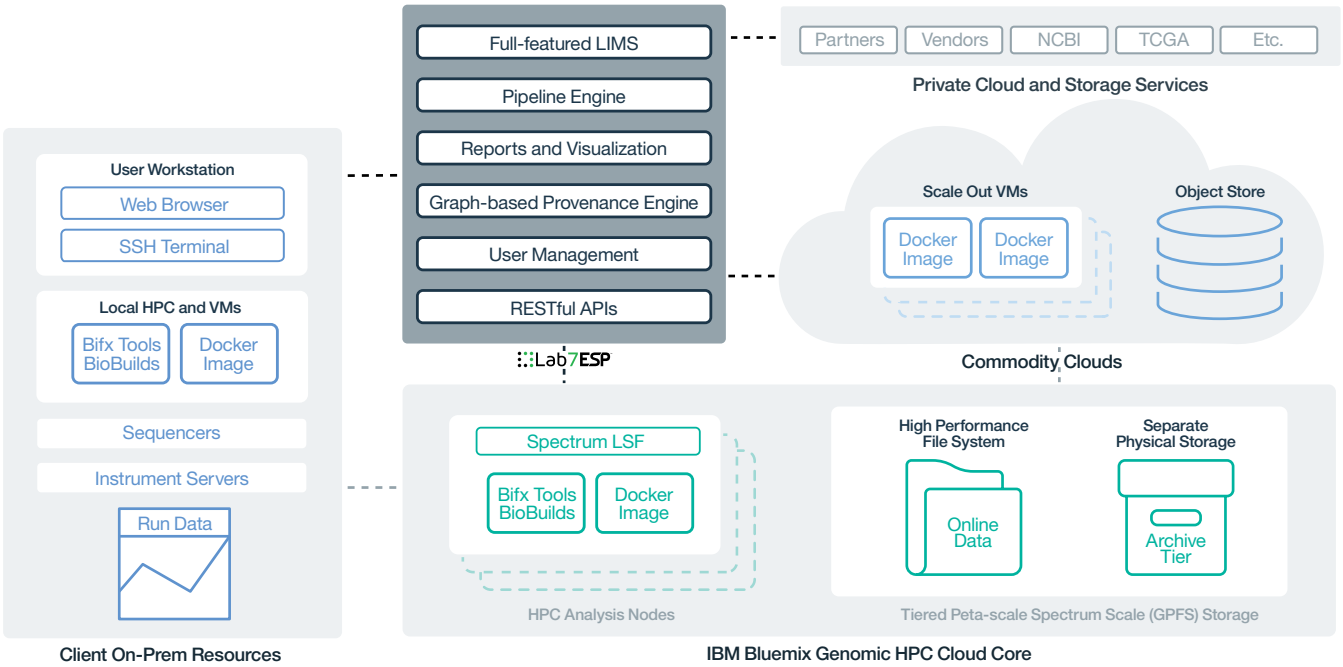

Figure 1: Architecture of the Lab7 High-Performance Genomic Cloud, running on the IBM Bluemix cloud platform

About the Lab7 High-Performance Genomic Cloud

The Genomic Cloud provides a balanced mix of simplicity, speed to value, cost and performance. It combines Lab7 software offerings such as Lab7 Enterprise Science Platform™ (ESP) for automated data management, sample and protocol tracking, analytics, and reporting with storage offerings from the IBM Spectrum Storage suite, all running on the IBM Bluemix cloud.

High performance

The Genomic Cloud uses IBM Spectrum LSF® (formerly IBM Platform LSF) and IBM Spectrum Scale™ (based on IBM GPFS™) to support high-throughput genomics work.

Spectrum Scale applies the power of cognitive computing to support higher performance for unstructured data, with capabilities such as flash-based acceleration, transparent tiering, and parallelization. Spectrum LSF provides intelligent, policy-driven scheduling and improved visibility to increase throughput up to 150x for simulation, design and research workloads.

Pairing these advanced solutions with high-memory compute nodes and a fast internal network helps create a cloud solution that's more than capable of handling the data processing needs of data-intensive scientific workloads. In total, the Genomic Cloud can support up to 500 compute nodes, with more than 10 petabytes of storage, helping achieve faster results and optimized resource usage for high performance computing applications.

Comprehensive

The Genomic Cloud includes everything you might need to get started quickly with genomics on the cloud. It is ready to run right out of the box; unlike commodity cloud options, it requires no significant infrastructure investments beyond what's included in the offering itself. This means that you can be up and running with your new cloud data platform in as little as two weeks, compared with up to six months to achieve equivalent functionality using commodity cloud offerings.

The solution can also be deployed anywhere there is a Bluemix cloud data center. This means there's a very good chance you'll be able to host your cloud in close proximity to your on-premises infrastructure. In turn, this can make it easier for you to deploy the Genomic Cloud as part of a hybrid environment, which may be necessary if you're working in a highly regulated industry that requires you to store certain types of sensitive data on premises.

Finally, the Genomic Cloud can be deployed across heterogeneous compute platforms, including on premises. The solution is optimized to run on IBM POWER8® servers, which can help achieve excellent performance, reliability and security.

IBM Bluemix: The ideal platform to support the Lab7 High-Performance Genomic Cloud

- **Unique blend of applications, infrastructure and services**
- **Global network of data centers, featuring more than 20 locations**
- **More than 130 services, empowering developers to build innovative and diverse apps**
- **Interoperable with many open source projects, including OpenStack, Cloud Foundry and Docker**
- **Flexible deployment options, allowing you to build the exact configuration that meets your needs**
- **Draws from IBM Watson, allowing developers to unlock the power of cognitive apps**

To learn more about IBM Bluemix, visit **ibm.com**/bluemix.

To learn more about IBM Spectrum LSF, visit **ibm.com**/systems/spectrum-computing/products/lsf/.

To learn more about IBM Spectrum Scale, visit **ibm.com**/systems/storage/spectrum/scale.

Supported

When you invest in the Genomic Cloud, the full support of IBM and Lab7 comes with it. While commodity cloud providers would essentially leave you on your own once the cloud is in place, IBM's high-performance computing experts and Lab7's bioinformaticians and software support team will be there to get your system running at its full potential, and to make sure it stays that way.

As a result, your IT team can spend less of its own time and resources supporting the cloud solution, which leaves them free to focus more on high-value tasks that actually contribute to better science.

Cost-efficient

Many businesses choose commodity cloud offerings due to the perception that they are affordable. However, as mentioned previously, these perceived cost benefits don't always pan out for organizations that need to process and store very large volumes of data. On the other hand, the Genomic Cloud offers a monthly subscription with transparent costs, as well as no cost penalties for running very large jobs.

Also, the Genomic Cloud is a complete enterprise offering, which needs no additional investments to perform effectively. Compare this to a commodity cloud offering, where you may need to factor in extra costs for things like software, infrastructure, storage, bioinformatics tools, and support, and any cost benefits you may have experienced originally are likely to evaporate.

Proof of concept

When an Asian genomics company was looking to deploy a cloud solution for its data processing needs, they were invited to participate in a proof of concept for the Genomic Cloud.

IBM and Lab7 worked together to deploy the complete cloud solution in just two weeks. They were able to locate it at the Bluemix data center in Singapore, meaning that the customer could benefit from being in close proximity to its cloud platform. This is a very important consideration for the genomics field, since moving large volumes of data across long distances can slow workloads down dramatically. This is just one example of how IBM's global reach can be an important advantage for companies looking to move into the cloud.

After the deployment had been successfully completed, the customer ran a number of benchmarks, including a sample workload of 50 human genomes. These benchmarks help demonstrate just how beneficial the new cloud solution can be for genomics applications.

The customer experienced a 33 percent reduction in processing time right out of the box. Prior to implementing the solution, it would have taken the organization five days to process the sample of 50 human genomes. Following the deployment, their data processing time for this sample went down to only 3.3 days.

The Lab7 team was also able to help the customer implement optimization techniques that reduced processing times even further, eventually reaching the point where the complete sample workload could be fully processed in just over six hours. This provides a clear illustration of how customers can directly benefit from the hands-on services that IBM and Lab7 offer along with the Genomic Cloud. Had they chosen to deploy a commodity cloud solution, they would never have had the opportunity to experience these kinds of productivity gains.

The results from this customer proof of concept demonstrate what's possible with the Genomic Cloud. They are a clear indication that the solution could help organizations performing data-intensive scientific research and discovery to overcome one of the biggest challenges facing them today. By doing so, these organizations can put themselves in a position to focus more on what really matters—better science—and spend less time worrying about the IT infrastructure they use to conduct that science.

For more information

To learn more about the Lab7 High-Performance Genomic Cloud, or to schedule your own PoC, contact Lab7 today at https://www.lab7.io/solutions/genomic-cloud.

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
May 2017

KUS12392-USEN-00