



UNIVA UD

Pathwork Diagnostics, Inc.:

Leveraging Cloud Computing to Support Peak Processing and Enable Innovation

Univa UD Case Study

Data Processing: **Molecular Diagnostics**

SOLUTION OVERVIEW

CUSTOMER CHALLENGE	Pathwork Diagnostics, Inc. develops molecular diagnostic tests to aid oncologists in the diagnosis of hard-to-identify cancer tumors. Pathwork chooses optimal models for its tests by analyzing large libraries of gene expression profiles from tumor specimens. Such analyses require massive amounts of computational power at certain peak times that occur during phases of product development and validation. So Pathwork, looking to avoid a large hardware investment while still meeting its processing needs, turned to cloud computing and Univa UD to provide a solution.
SOLUTION	Pathwork investigated several cloud providers before choosing Amazon Elastic Compute Cloud (EC2) for pay-per-use computing. Pathwork then selected Univa UD's UniCloud product to build high-performance computing (HPC) clusters in the EC2 cloud, in order to perform their compute-intensive research. Pathwork downloaded and installed UniCloud, referencing a published 'How-To' white paper for guidance. Integrating their own applications into the UniCloud cluster was easy, and with the help of Univa UD's support team, Pathwork was up and running with access to a fully functioning HPC cluster in a cloud environment.
BENEFITS	With UniCloud, Pathwork Diagnostics gains these key benefits: <ul style="list-style-type: none">• Saved hundreds of thousands of dollars by avoiding the expenses of purchasing and maintaining HPC hardware in-house• Met peak demands in computational capacity without purchasing new hardware• Enabled key research innovations which would otherwise have been infeasible
TECHNOLOGY	UniCloud

THE CUSTOMER: A LEADING BIOINFORMATICS INNOVATOR

Pathwork Diagnostics is a molecular diagnostics company that develops high-value diagnostic tests to assist oncologists in determining treatments for cancer care. Using proprietary technologies, Pathwork utilizes the power of genomics information to develop FDA-cleared microarray-based tests that provide physicians with clinically actionable information unavailable using alternative methods – so standard-of-care, cancer-specific treatment can begin.

PROVIDING PHYSICIANS THE DATA THEY NEED

Physicians provide cancer tumor specimens that Pathwork processes and analyzes using a microarray. The resulting gene expression profile reflects the magnitude of abundance of specific messenger RNA molecules corresponding to individual genes in the tumor.

Pathwork applies proprietary machine learning algorithms¹ to this data and generates a user-friendly diagnostic report. The physician downloads the report via a secure website. The physician then utilizes the report as an aid in determining the tissue of origin which assists in determining the proper treatment for the tumor. Pathwork has received FDA clearance for an *in vitro* diagnostic kit version of the test.

“Our challenge was a perfect fit for cloud computing, because we needed access to more computing capacity than we could possibly maintain internally – but only at certain peak times. We found EC2 to be the most convenient cloud solution and UniCloud to be the best software infrastructure tool to use.”

**– Ijubomir Buturovic, Chief Scientist,
Pathwork Diagnostics**

HELPING DOCTORS REACH CRITICAL DIAGNOSES

Using Pathwork Diagnostics tests, oncologists are able to rule out possible tumor origins and zero in on the correct tissue of origin – a critical step in determining the type of cancer as well as the proper treatment.

Here are just a few examples where Pathwork results have made a direct impact on cancer treatments:

- Using the Pathwork Diagnostics TOO Test, doctors were able to rule out 13 different tissues to determine the correct tissue of origin for a patient, so management specific to the primary cancer could begin.

- Pathwork tests resolved numerous conflicting findings to determine a patient’s cancer type was ovarian.
- Pathwork is collaborating with leading research organizations in the UK to make the test available and increase the standard of care for cancer patients.

THE CHALLENGE: FINDING A COST-EFFECTIVE SOLUTION FOR HIGH-VOLUME PROCESSING

Pathwork generates its diagnostic reports using a model which is developed by applying machine learning algorithms to large gene expression databases (as described above). Finding the best models involves solving large-scale optimization problems with thousands of variables. In the process, tens of thousands of models are built in parallel and compared with each other in order to find the best predictive model.

“Without UniCloud, some of our most advanced research would simply not be feasible.”

**– Zoran Popovic, Lead Engineer,
Pathwork Diagnostics**

GOOD NEWS, BAD NEWS

The good news is that many machine learning algorithms for the types of problems addressed by Pathwork Diagnostics are highly parallelizable – it is a straightforward process to perform the necessary computations on multiple machines, because there are no interdependencies between different models being evaluated.

Unfortunately, the process of evaluating these models is a highly compute-intensive task. Tens of thousands of models must be processed to find the best model to produce the diagnostic report. This computation can take weeks or months using a mid-size HPC resource such as a 64-node cluster.

Furthermore, some of the research being considered by Pathwork is so computationally demanding and complex that, according to their chief scientist, “it is not even feasible to do using our internal resources.” In other words, such algorithms would not be able to be applied without a more powerful HPC resource than Pathwork was able to acquire in house.

THE DECISION TO GO CLOUD

Like most small to medium businesses in today’s tough economic climate, Pathwork needed to find a solution for

¹ The purpose of machine learning algorithms is to use observations (experiences, data, patterns, examples) to improve performance for a specific task. Put simply, machine learning is about learning to do better in the future based on what was experienced in the past. The emphasis of machine learning is on automatic methods that do the learning automatically without human intervention or assistance.

PATHWORK DIAGNOSTICS, INC.:

Leveraging Cloud Computing to Support Peak Processing and Enable Innovation

its processing needs that did not require the company to purchase and maintain hardware in house. The acquisition costs, not to mention the additional expenses for power, cooling and system administration, literally makes the investment impossible for many businesses who have the need, but not the resources, to perform HPC research.

So Pathwork turned to cloud computing as a solution. As their chief scientist notes, "it's easy to understand why we went with a cloud solution. We have some distributed computing resources in house, but when we need to develop and deliver a product for clinical validation we'll have periods of weeks where we need access to almost unlimited capacity." At other times, the company utilizes its resources to perform research, review new papers and findings, execute feasibility studies... all tasks that don't require as much capacity.

In short, it's a perfect fit for cloud: Pathwork needed access to vast computing capacity only at certain peak times.

"By using UniCloud and Amazon EC2, we're getting exactly what we need. We pay for HPC power only when we use it, using a cluster management system that's easy to use and operate. Thanks to UniCloud, we're able to meet our peak computational demands without investing in new hardware."

**– Ljubomir Buturovic, Chief Scientist
Pathwork Diagnostics**

In addition, these peak times are regular but somewhat unpredictable – so Pathwork needed a solution that was flexible, cost-effective, and user-friendly enough to be accessed and utilized on the fly, as requirements arise.

THE SOLUTION: AMAZON EC2 AND UNICLOUD

As previously described, the two key incentives that led Pathwork scientists to a cloud approach were:

- To **support peaks in computational capacity demand** which occur regularly (but not predictably)
- To pursue **otherwise infeasible research projects** by providing access to previously unattainable HPC capacity

Pathwork scientists investigated several different cloud providers before selecting the Amazon Elastic Compute

Cloud (EC2) service, a web service that provides resizable compute capacity in the cloud. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing users to quickly scale capacity, both up and down, as computing requirements change. EC2 provides clients with complete control of computing resources and lets them run on Amazon's proven computing environment.

Possibly most important of all, Amazon EC2 changes the economics of computing by **allowing users to pay only for capacity that they actually use**. Amazon EC2 provides users the tools to build failure resilient applications and isolate themselves from common failure scenarios.

According to Pathwork's lead engineer, Amazon EC2 was the most convenient cloud service available. EC2 makes it a very straightforward process to port internal software to a configured EC2 cloud system.

WHY UNICLOUD?

UniCloud is an extension to UniCluster, Univa UD's leading product for HPC systems management. With UniCloud, organizations can provision and scale capacity on the Amazon EC2 environment, expanding baseline computing resources by provisioning capacity to meet peak demand.

UniCloud allows organizations to establish workload policies and requirements that dynamically trigger the setup of virtual compute nodes (or Amazon Machine Images – AMIs) in Amazon EC2. This flexible, on-demand approach to satisfying HPC needs greatly reduces up-front capital expense and lowers the barrier to HPC entry.

Two key differentiators made the UniCloud decision an easy one for Pathwork:

- The underlying codebase is open source, so this portion of the product can be played with and investigated more completely as it is used. The product is also fully supported, so when usage becomes a more serious and regular activity, users can turn to Univa UD for everything they need to support their production environment.
- UniCloud embeds Sun Grid Engine (SGE) as its workload scheduler. This was the tool already in use at Pathwork, so there was virtually no interruption or learning curve on the scheduling portion of the product.

As the Pathwork lead engineer explains, "the decision to use UniCloud was basically a no-brainer. It's open source, so we have access to the code when we need it. And it's already integrated with Sun Grid Engine, our existing distributed resource management tool."

Additional UniCloud differentiators include:

- UniCloud is an extension of Univa UD's UniCluster stack, so the product has the capability to enable physical, virtual (including cloud resources), or mixed physical and virtual clusters.
- UniCloud applications are extended by virtue into the cloud and do not need to be re-written or adapted as they would for other offerings.

GETTING UP AND RUNNING

Pathwork downloaded and installed the UniCloud software on their own with the help of a How-To reference white paper. Most of the challenges in the process of creating the functional system were solved internally by Pathwork's own team. Univa UD was also available to answer questions and help make the process a success.

The UniCloud system is currently being used by the Pathwork science team for research and product development.

“UniCloud minimizes both our initial and long-term investments in HPC.”

– **Ljubomir Buturovic, Chief Scientist**
Pathwork Diagnostics

VALUE AND BENEFITS

Top UniCloud benefits cited by Pathwork include:

- **Cost savings:** UniCloud means eliminating capital expenses (CAPEX) including data center investments, power/cooling, hardware and software – and also significantly reducing operating expenses (OPEX) by minimizing personnel costs and employing Amazon EC2's pay-per-use model. Now Pathwork pays only for the HPC capacity they use.
- **Access to vast new capacity:** UniCloud and Amazon EC2 provide access to the computational capacity Pathwork needs for peak demand times. By leveraging this previously unattainable computing capacity, Pathwork can now investigate more complex algorithms in the process of developing its diagnostic models.
- **Support for innovation:** UniCloud has enabled Pathwork to pursue new research directions involving more advanced algorithms – innovation that would otherwise have been impossible.

“The decision to use UniCloud was basically a no-brainer. It's open source, so there's no barrier to acquisition and we have access to the code when we need it. And it's already integrated with Sun Grid Engine, our existing distributed resource management tool.”

– **Zoran Popovic, Lead Engineer**
Pathwork Diagnostics

WHAT'S NEXT?

Now that UniCloud is up and running, Pathwork plans to use the product for 2 main activities:

- Development of their next diagnostic tests, which will require more peak usage times
- Research of new machine learning algorithms, which is done on an ongoing basis – now Pathwork can pursue new research directions involving different types of more advanced, complex algorithms

EXPANDING VALUE WITH UNICLOUD 4.1

Pathwork has a subscription for full Univa UD support and is eager to upgrade to the 4.1 version of the underlying UniCluster software, which will provide access to additional features and more complete documentation for the supported version.

The new version of UniCloud includes:

- Streamlined installation and cluster setup, making it even easier to get UniCloud up and running on your own
- Expanded systems management capabilities that automate repetitive and/or homegrown administrative tasks
- Application kits including cluster verification, analytics, unified job submission, and more

For more information about new features available with the 4.1 version, visit www.univaud.com/hpc/products/uni-cluster/.

FOR MORE INFORMATION

To find out how UniCloud can benefit your organization:

- Contact Univa UD at 1-800-370-5320
- Write us at info@UnivaUD.com
- Visit www.univaud.com/hpc/