

Thermo Electron Corporation and IBM offer an integrated, optimized proteomics solution



IBM @server 1350 Cluster

Increased throughput for large-scale protein identification

Informatics plays a central role in driving successful proteomics efforts. As proteomic research matures and the amount of data generated increases, the research bottleneck is shifting from data generation to the analysis phase. A typical protein analysis, for example,

may yield thousands of mass spectra, which could take days to process on a single desktop computer. Traditionally, time-consuming processes like these have caused life sciences researchers working in smaller labs to lose competitive advantage to larger facilities with access to the best informatics tools. This has created an increased demand for powerful, yet affordable and scalable informatics capabilities that can be accessed from a desktop.

IBM Life Sciences and Thermo Electron have designed a complete, preconfigured proteomics solution that enables accelerated throughput of increasingly complex, multi-dimensional informatics. This offering complements Thermo Electron's ProteomeX™ Workstation — a fully integrated LC/MSⁿ system that features the Surveyor® dual pump capillary liquid chromatograph for multi-dimensional separation with the ultra-sensitive LCQ™ Deca XP Plus ion trap mass spectrometer. Packaged with the new SEQUEST® Cluster

IBM Business Partner:

Thermo Electron Corporation

Thermo Electron Corporation is a leader in chromatography and mass spectrometry for the life sciences industry, providing laboratory professionals with the most accurate and sensitive instrumentation available for analyzing a wide array of substances. From state-of-the-art hardware to software and global services and support, Thermo's products can be used alone or as complete systems to produce optimal performance. Thermo Electron's overriding goal is to provide life sciences customers with high-performance products that reduce research expense and accelerate the discovery and development processes. Thermo Electron is based in Waltham, Massachusetts, and employs approximately 11,000 people in 30 countries worldwide.

Thermo
ELECTRON CORPORATION

software and IBM @server Cluster 1350, life sciences organizations can immediately reap the benefits of a powerful, advanced proteomics solution—from a single desktop PC.

Redefining the industry standard

The acceleration of proteomic studies is driving the demand for tools able to rapidly characterize protein sequence and structure to facilitate the understanding of protein interaction and function, in both normal and diseased states. Traditional methods of protein analysis cannot begin to achieve the required throughput and sensitivity. An important part of a robust, automated solution to this problem is Thermo Electron's SEQUEST Cluster software, which uses tandem mass spectrometry data to determine the sequence and identification of proteins, even in highly complex mixtures.

The heart of SEQUEST Cluster is a patented cross-correlation algorithm that matches the MS/MS spectra of digest peptides with computer generated fragmentation spectra predicted from genomic and/or protein sequence databases. SEQUEST

Cluster achieves the most accurate results by utilizing the uniqueness of MS/MS fragmentation data to make unambiguous assignments of digest peptide amino acid sequences, which are then mapped to the parent protein. SEQUEST Cluster dramatically increases the speed and capacity to analyze unknown proteins by dividing the analysis into separate tasks and performing them in parallel on the IBM @server Cluster 1350. Once the compute nodes complete their analysis, the results are integrated on the management node and unified output is generated.



Thermo Electron's Finnigan ProteomeX Workstation

Key points of interest:

- *Using a loosely coupled parallel algorithm, SEQUEST Cluster demonstrates impressive scalability*
- *SEQUEST Cluster greatly reduces the overhead of MS information management by allowing one instance of the application to analyze a large amount of data*
- *Although the SEQUEST Cluster application runs within Microsoft® Windows®, compute nodes can use either Windows or Linux®, allowing back-end processing to be done on the IBM @server Cluster 1350 or the IBM @server BladeCenter™*

Benchmark results for SEQUEST Cluster running on IBM @server Cluster 1350

| CPUs | No. of processors | Analysis time |
|------------------------------------|--------------------------|----------------------|
| Single personal computer | 1 | 40:18 minutes |
| 16-node @server 1350 Linux Cluster | 34 | 20.75 minutes |
| 32-node @server 1350 Linux Cluster | 66 | 9.75 minutes |

An example of the increased power derived from SEQUEST Cluster running on the IBM @server Cluster 1350 can be illustrated in a recent benchmark run using the 6-hour yeast.raw file (17,112 spectra) and the yeast.fasta database (13,719 proteins). The benchmark search times show that a 7-hour analysis time can easily be reduced to less than 10 minutes.

Automating and speeding quantitative protein expression analysis

The ProteomeX Workstation enables the rapid setup, execution and analysis of multi-dimensional LC/MS methods, providing a simple, elegant solution to analyze and characterize complex proteomics samples. ProteomeX uses multi-dimensional protein separation technology to:

- *Analyze thousands of proteins at a time within a single sample*

- *Use customized methods for conducting large-scale proteomics experiments*
- *Profile differential expression and conduct shotgun sequencing of complex samples*
- *Identify proteins from complex mixtures*

A complete protein identification solution—fully integrated and ready to use

Optimized for proteomics, the combination of the ProteomeX Workstation with the high-throughput database-searching capabilities delivered by the SEQUEST Cluster software package—driven by the computing power of the IBM @server Cluster 1350—is a powerful means of generating biochemical information on classes of proteins previously considered indecipherable.

The SEQUEST Cluster is bundled with Thermo's BioWorks™ 3.1 protein identification software suite, the industry standard for protein identification. BioWorks is engineered to reduce the complexity of the most sophisticated proteomics analyses. SEQUEST Cluster is optimized to enhance productivity by enabling the proteomics researcher to:

- *Identify proteins automatically by comparing tandem mass spectrometry (MS/MS) data with standard protein and DNA databases*
- *Correctly identify weak-signal proteins in mixtures by extracting and cross-correlating uninterpreted MS/MS information*
- *Analyze a single spectrum or an entire LC/MS/MS peptide map automatically*

The SEQUEST Cluster software includes all of the features of BioWorks, such as XPRESS™, a tool that automates quantitative protein expression analysis. XPRESS can quantify peptides tagged with any label, including ICAT™ reagent method and those labels on their N or C termini. Another component of the cluster is ZSA, an algorithm that

more accurately determines the charge state of a precursor ion in order to reduce the number of spectra that are submitted for a database search. To confirm protein identification results from the SEQUEST Cluster, a new utility called SALSA (Scoring Algorithm for Spectral Analysis) can be included with the package as an additional option. SALSA uses pattern recognition to mine tandem MS/MS spectra, revealing post-translational modifications of proteins, as well as possible genetic mutations.

All the advantages of supercomputing, with the ease-of-use of a single PC

At the heart of this integrated solution is the IBM @server Cluster 1350. This Linux-based cluster package combines IBM @server xSeries® rack-optimized servers running IBM Cluster Systems Management (CSM) software for the Linux operating system. The IBM @server 1350 provides life sciences organizations with:

- *A highly scalable Linux cluster solution for high-performance and commercial computing workloads*
- *Robust cluster systems management from a single point of control*
- *Optional high-speed interconnect capabilities*

- *Powerful, flexible configuration, supporting up to 512 IBM @server xSeries 335 compute nodes and as many as 32 IBM @server xSeries 345 storage nodes*

Preconfiguration gets you up and running quickly

IBM Life Sciences streamlines the integration process by leveraging its own experience and that of its worldwide distributor network.

Typically, cluster hardware must be configured, racked, cabled and tested onsite by the customer, a process that can take three to four weeks. In addition, software must be procured from various sources and loaded onto the cluster, adding more of a delay. IBM fully integrates the @server Cluster

1350 into a rack and tests it over a 48-hour period before shipping it from the manufacturing facility. Then the “gold disk” image is loaded onto the cluster, which may include Microsoft Windows 2000 Server, RedHat Linux, Cluster Systems Management (CSM) software and the SEQUEST Cluster application. The cluster is delivered in a single crate and is ready for checkout and final configuration by IBM and Thermo Electron personnel.

In addition to these integration and set-up services, IBM's service and support extends the value of the integrated cluster through a site survey, a Solutions Assurance Review, a three-year worldwide hardware warranty (which includes third-party

“The promise of a high-throughput solution to protein identification and database searching has finally been realized with the new SEQUEST Cluster. And with the integration of IBM hardware and software, and its ability to deliver a preconfigured solution, customers can be up and running faster than ever before.”

– Dr. Iain Mylchreest

General Manager of Thermo Electron's Life Sciences Mass Spectrometry Business

components of the cluster) and a one-year software maintenance for IBM CSM for Linux and GPFS for Linux. Additional services to connect the cluster into your laboratory's network and/or storage capabilities, training and additional support products are available upon request.

About IBM

IBM's commitment to life sciences is defined by the establishment of the specialized IBM Life Science Solutions business unit. IBM is dedicated to rapidly bringing leading-edge technology out of the laboratory and into the marketplace for customers and Business Partners in the fields of pharmaceutical research, biotechnology, genomics, proteomics, healthcare and other life sciences.

Long-term projects at IBM Research Centers and the IBM Deep Computing Institute foster collaboration with life sciences companies—bringing domain expertise and innovative technologies to the development of life sciences solutions. IBM actively collaborates with companies like Thermo Electron, whose domain knowledge, products and resources can help build valuable solutions for our mutual customers.

For more information

To learn more about IBM Life Science Solutions and our Business Partners, visit **ibm.com/solutions/lifesciences** or contact an IBM Life Sciences specialist at LS@us.ibm.com.

To learn more about Thermo Electron, visit **www.thermo.com**.



© Copyright IBM Corporation 2003

IBM Corporation
Route 100
Somers, NY 10589
U.S.A.

Printed in the United States of America
5-03
All Rights Reserved

IBM, the IBM logo, BladeCenter, @server and xSeries are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both.

Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Surveyor, LCQ, ProteomeX and BioWorks are trademarks or registered trademarks of Thermo Electron Corporation.

XPRESS is a trademark of the Institute for Systems Biology.

ICAT is a trademark and SEQUEST is a registered trademark of the University of Washington.

Other company, product, and service names may be trademarks or service marks of others.

This solution sheet illustrates how one IBM Business Partner uses IBM's and its own technologies/services. Many factors have contributed to the result and benefits described. IBM does not guarantee comparable results. All information regarding the Business Partner's products contained herein was provided by the featured Business Partner. IBM does not attest to its accuracy.

♻️ Printed in the United States on recycled paper containing 10% recovered post-consumer fiber.



GH09-4631-00