LifeScience Insights™ An IDC Company

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.988.7800 F.508.935.4015
www.life-science-insights.com

WHITE PAPER

Apple and Life Science Research: Enabling Technologies

Sponsored by: Apple Computer, Inc.

Michael Swenson James B. Golden III, Ph.D.

March 2005

LIFE SCIENCE INSIGHTS OPINION

IDC's Life Science Insights believes that Apple computer products are building upon the long-time loyalty of life science researchers and are attracting new users migrating from Unix and Linux platforms. This white paper, drawing on case study interviews, has identified the following key reasons for the popularity of Apple products as well as key challenges for Apple users:

- LSI believes that Mac adoption has benefited from broader market shifts away from proprietary Unix and Microsoft Windows platforms to open source OS environments such as Linux and open source applications that are increasingly prevalent in the life sciences. Nearly every scientist we interviewed commented on the dual nature of Mac OS X — a great graphical interface, native access to Microsoft Office, and yet a full-blown Unix box underneath that can support legacy Unix codes as well as a straightforward port of open source applications written for Linux platforms.
- Several of our interviews indicate an increase in the number of scientists using an iBook or PowerBook laptop for their mobile computing. Key reasons include the combination of a system that has the cool factor of great design, a robust set of communications applications, native Microsoft Office capabilities, and a full-fledged Unix environment for developing and running scientific applications all on the same portable system.
- The PowerPC G5 chip and the Macintosh system architecture provide a high-performance system with math libraries that have been optimized to take advantage of special computing capabilities such as the Velocity Engine vector processing unit. This feature set is particularly appreciated by life scientists who develop their own computationally-intensive custom code.
- The Apple Workgroup Cluster is competitive with Linux-based clusters, particularly on the dimensions of ease of deployment and cluster manageability. Users who have experience with both Linux clusters and Xserve-based clusters indicate a significantly reduced administrative load in moving to Xserve clusters.

While the Mac platform is regaining popularity among life scientists, Apple still faces challenges accommodating itself to a Microsoft-centric world. This challenge is particularly strong in large pharmaceutical companies. The challenges range from exclusion of Mac systems because of a commitment by the IT department to standardized desktops to challenges of integrating Mac clients with Microsoft Exchange in those organizations that do permit a variety of platforms on the desktop.

TABLE OF CONTENTS

P

LIST OF TABLES

IN THIS WHITE PAPER

This white paper begins with a Situation Overview section that covers the business case for investing in high-performance computing resources in the life sciences, followed by a discussion of the drivers for computational requirements and the trends in technical computing that address those requirements. This is followed by a discussion of the position of Apple products in meeting the needs within the life sciences.

This white paper also includes nine summaries of case study interviews with scientists who are using Mac products in life science research settings. These perspectives are summarized in themes that illustrate both strengths and challenges for Apple in serving this market. Finally, the white paper concludes with a brief discussion of competition and market challenges as well as market opportunities for Apple.

METHODOLOGY

This white paper draws on inputs from a variety of sources:

- Daily analyst coverage of information technology in the life sciences
- Involvement by the authors in a number of qualitative and quantitative studies of high-performance computing (HPC) in the life sciences and other scientific markets
- Briefings from Apple product managers and life science specialists
- Nine case study interviews that were conducted specifically for this white paper (These case studies were conducted by one or more of the authors without the presence of Apple representatives. The summary of each case study interview was provided to the interviewee before publication for their review and approval.)
- Results from previous surveys conducted by IDC's Life Science Insights that were not sponsored by Apple

SITUATION OVERVIEW

Business Case for High-Performance Computing Resources in the Life Sciences

Computers have been inextricably linked to life science research for a number of years now. The load of data is simply too great for individuals to manage without the assistance of a variety of information technologies to capture, store, query, and analyze data and then present and communicate new knowledge. Additionally, there are more and more types of data to manage. Although the role of computers is well established for these support tasks, there remains both great hope and some skepticism that a variety of more advanced information technologies, ranging from

There remains both great hope and some skepticism that a variety of more advanced information technologies, ranging from modeling and simulation to advanced informatics and text mining, can elevate the role of computing into a more central partnership role with classical experimentation.

modeling and simulation to advanced informatics and text mining, can elevate the role of computing into a more central partnership role with classical experimentation.

Many of these advanced techniques require higher levels of computational performance. In the past, the need for high-performance computing (HPC) limited these techniques to an exclusive club of researchers that had access to supercomputers, large symmetric multiprocessing (SMP) systems, and high-end RISC workstations. Under those high-cost conditions, it was difficult for many researchers to justify the use of HPC-enabled techniques because many of the techniques were still unproven within life science domains.

Several technology trends have changed that situation and allow both academic- and industry-based life science researchers to take a fresh look at the business case for HPC. These trends include:

- Moore's law–type improvements in computing hardware make individual computers increasingly powerful for very reasonable prices.

- The standardization of networking on TCP/IP and Ethernet has led to regular increases in networking bandwidth, enabling computational clusters to be constructed for a very competitive price.

- The rise of open source operating systems, such as Linux, and applications for clusters have enabled life science researchers to have low-cost access to a wide variety of software tools for advanced computing.

These trends have made the deployment of advanced analytic techniques on commodity computer clusters a reality. The increasing power of standard client computers has made applications previously only available for high-end workstations accessible to the average scientist.

The increasing power of standard client computers has made applications previously only available for high-end workstations accessible to the average scientist.

Although some academic supercomputer centers, government labs, and industry labs will still make an investment in supercomputer-class systems as a way to keep pushing the research envelope (e.g., conducting comparative whole genome analysis in a single memory space without breaking the problem up for a cluster) or gain a competitive edge, the large majority of researchers do not have funding on this scale. Nevertheless, the aforementioned improvements enable the average researcher to make a business case for a small cluster to enable the use of more advanced computational techniques.

Scientists in government and academic research laboratories tend to focus on basic research that extends the boundaries of knowledge or applied problems that are of national common interest such as the detection of pathogens that could be used in bioterror attacks. In these settings, computational requirements are most often driven by one of two major factors. The first, also true for industry labs, is keeping up with the data-generating capabilities of new instruments. It is dysfunctional for a lab to invest half a million dollars in a new instrument and then handicap the scientists who would use the instrument because they do not have appropriate computer systems to store, filter, and analyze the data from the new instrument**.** The second business case comes from computationally oriented theorists who are developing mathematical

It is dysfunctional for a lab to invest half a million dollars in a new instrument and then handicap the scientists who would use the instrument because they do not have appropriate computer systems to store, filter, and analyze the data from the new instrument**.**

models of biological structure and activity. Researchers coming from this perspective can usefully employ the largest, most capable computing systems they are able to afford. Modeling problems such as the folding of a single protein, from first principles, can easily use all the cycles of a major supercomputer for months.

In pharmaceutical and biotech companies, the critical question for making a business case is, How will investing in computational resources impact the efficiency and effectiveness of various tasks within the process of drug discovery and development? A large proportion of spending on IT infrastructure in biotech and pharmaceutical R&D departments is for technology that is supportive to experimental science. There is a powerful business case for technology that captures experimental data, structures and tracks high-throughput experimentation, and creates systems for tracking, securing, and analyzing data from clinical trials. In addition, IT plays a critical role in enabling collaboration among scientists, managing complex projects, and acting as a medium for knowledge management across all of R&D within a biotech or pharmaceutical company.

IT plays a critical role in enabling collaboration among scientists and managing complex projects and as a medium for knowledge management across all of R&D within a biotech or pharmaceutical company.

We've seen that the background and supportive role of IT in providing data management, analysis, and communication is the major theme of IT investments by biopharma. However, there are elements of the drug discovery and development work processes in which pharmaceutical companies believe that higher levels of computational resource make a significant difference and in which computation plays a more important role. These include the following:

- Many biotech and pharmaceutical companies maintain a computational cluster dedicated to genomics applications such as BLAST and HMMER that are useful in the target identification stage of drug discovery.
- Many companies use clusters and grid computing in docking applications such as DOCK, GLIDE, or Ligand Fit. These types of applications estimate the ability of a small molecule (potential drug) to dock or fit into an active receptor site on a ligand identified as a drug target. Researchers are using docking programs as virtual high-throughput screening (HTS) to test and eliminate many compounds that computational models suggest will not show a positive result. This process holds out the potential to speed up the screening process and save on experimental costs.
- Finally, another set of tasks where computational techniques have demonstrated value in drug development is the use of computational modeling and simulation in support of lead optimization. These types of applications are run on a mix of platforms, including the desktop, high-end workstations, clusters, and large SMP servers.

Biotech and pharmaceutical companies have already made significant investments in computational resources for these stages of the workflow in order to enhance R&D productivity.

Additional Factors in Validating an Investment in HPC Resources

The benefits of investing in HPC resources, including clusters, are fairly clear, but CIOs and CFOs want to understand both hard and soft costs for those investments.

Thus any scientist making a case for a purchase should carefully consider a number of factors, including the following:

- Purchase cost
- Cost of ongoing service contracts
- Ease of implementation/deployment
- Ease of integration with existing IT systems
- Compatibility with existing applications and custom code base
- Cost of power and cooling
- Ease of management and cost of ongoing internal IT support
- Robustness of security features
- Ability to repurpose or recycle at the end of the product life cycle

The ultimate goal for any HPC-type purchase is to create the right balance of maximizing the computational benefits to the end-user scientists while minimizing the total cost of the system to the organization.

Drivers for Computational Requirements

In this section, we examine a number of trends impacting both the volume and complexity of the scientific data that needs to be analyzed within the framework of life science research. We also look at the organizational impetus and support for computational techniques.

Data Volume

The increasing volume of data is based on four major trends:

The drive to automate impacts sequencers, liquid-handling equipment, separation equipment, cellular imaging, traditional microscopy, and even mass spectrometry. The outcome of this automation trend is a flood of data.

- **Automation of experimentation.** Automation technology enables repetitive lab techniques to be executed with better control and consistency while decreasing the amount of human labor per experiment. This trend is perhaps most evident in high-throughput screening (HTS). HTS and ultra-HTS (UHTS) use robotics to conduct as many as 100,000 experiments per week in search of compounds that show activity with selected biological targets. The high-throughput mindset and accompanying automation is also a critical component of the "omics," attempts to characterize all of the genes, proteins, and metabolic pathways in an organism of interest. The drive to automate impacts sequencers, liquid-handling equipment, separation equipment, cellular imaging, traditional microscopy, and even mass spectrometry. The outcome of this automation trend is a flood of data. A bioinformatics director at a major pharma pointed out that two-day mass spectrometry experimental runs result in so much data that a dedicated 50-node compute cluster takes three to four days to run the initial filtering, peak picking, and data processing on the resulting data. And this time estimate does not take into account the more thoughtful analysis and interpretation that the scientists need to apply to gain real insight.

- **Resolution improvements in biological measurements.** The continual improvements in measurement resolution in both space and time also create greater volumes of data. Better resolution in spatial dimensions results in the measurement of smaller features and smaller reactions or interactions. For example, improvements in NMR resolution are now allowing researchers to achieve structural imaging of proteins that resist crystallization and are also useful for imaging metabolites. In the time dimension, improved resolution primarily involves taking more measurements per second for dynamic phenomena. For example, we see confocal microscopes increasing roughly 10x in just a couple of years, going from an ability to capture 15 images per second to an ability to capture about 100–120 images per second. These types of consistent improvements in resolution across a whole range of scientific instruments contributes greatly to the increasing volume of data that is available for analysis within life science disciplines.

- **Digitization of imaging.** Digitization of images is a master trend across multiple medical and scientific instruments, ranging from medical images such as radiology and MRIs to cellular images in digital pathology and microscopy and molecular-level digital images from HTS assays. Digital images tend to be large (e.g., the new Zeiss LSM 5 Live microscope captures up to 100 images per second and can generate approximately 100MB of image data per second). Another example is digital pathology, with each slide generating on the order of 5–10GB of image data, depending on resolution. New technology developed by Dmetrix has reduced the time to capture that data from about 10 minutes to just under 1 minute. As digital capture, analysis, and storage of medical and scientific images becomes increasingly prevalent, it only increases the requirements for computational infrastructure in biological laboratories and for scientists to closely integrate computational tools into their workflow to maximize the extraction of medically and scientifically relevant information.

Digitization of images is a master trend across multiple medical and scientific instruments, ranging from medical images such as radiology and MRIs to cellular images in digital pathology and microscopy and molecular-level digital images from HTS assays.

- **Internet-based databases and content repositories.** The final factor is the dramatic increase in data available to the average life science researcher through access to public and subscription-based content repositories and databases over the Internet. These hundreds of databases provide a number of benefits to scientists, including:

  - Scientists can use the data sets as a tool for exploration, investigation, and learning that supplements traditional scientific and medical journal articles.

  - Inductive analytic approaches to these data sets can provide a set of relationships that then serve as a basis for hypothesis generation.

  - Many scientists use these public data sets as reference sources for validating and/or providing context for experimentation carried out in their own labs.

Each of the four factors impacting data volume in the life sciences is a dynamic trend — any one of them alone would ensure continued growth in data volume. Together, these four factors are driving exponential growth in data for life science organizations. Biotech and pharmaceutical IT directors have indicated in conversations with LSI analysts that their R&D organizations are facing a doubling of data storage requirements in the range of every nine months on the aggressive side to about every two years for less data-intensive labs. If not for exponential improvements in computing and data storage technology, this growth in data volume within the life sciences would be completely unmanageable, instead of merely difficult.

Data Complexity

In addition to exponential growth in data volume, the life sciences are experiencing an increase in the complexity of data types that support new measurement techniques such as:

- High-content screening or cellular screening takes the paradigm of HTS to a different level. Instead of a single measurement of activation per well, these new measurement technologies allow tagging of multiple proteins and enable measurement of intensity over time. This data is captured as digital images, which are then converted into numerical data sets. The numerical data sets become the basis of quantitative analysis, but the original images also are important for both qualitative examination by the researcher and presentation and illustration purposes.

- DNA microarrays for measuring gene expression allow researchers, who in the past might have focused on 5–10 candidate gene expression changes using Northern or Southern blot techniques, to explore changes in gene expression on a global scale. By measuring expression in thousands of genes at a time, researchers gain a broader view without sacrificing the local view of target genes. That leads to a better understanding of complex patterns and interactions in gene expression that would have been missed by only examining the 5–10 candidate genes through previous technologies.

By measuring expression in thousands of genes at a time, researchers gain a broader view without sacrificing the local view of target genes.

- Protein array technology, though still in its early stages, is a good illustration of the creation of new and complex data types. A bioinformatics director at a large pharmaceutical company indicated that although he sees great promise in protein arrays, he believes that the analysis of protein arrays is considerably more complex than gene expression arrays.

The complexity of new data types is further exacerbated by the increasing attention to integrating various types of data into both traditional relational databases as well as more comprehensive conceptual models. Approaches such as systems biology and clinical genomics incorporate both new data types and new conceptual models into a more holistic vision of biology.

Approaches such as systems biology and clinical genomics incorporate both new data types and new conceptual models into a more holistic vision of biology.

Support for Computational Methods

The combination of new and complex data types, increased data volume, and a systems approach to biology creates a need for more powerful and cost-effective computational technology. Although there have been some notable failures among commercial firms hoping to leverage these trends (e.g., DoubleTwist, Viaken, and Paracel), government, academia, and industry continue to work on the development of these technologies.

Large government labs continue to form bioinformatics core facilities to support traditional biologists without formal training in the use and development of informatics tools. Government agencies also support or host many of the Internet-accessible databases and content repositories.

Within academia, we see the establishment of bioinformatics degree programs, and scientists engaged in the development of new algorithms and software to support life science research. Often, software and algorithms produced in higher education research organizations are made freely available to the public. In addition, we see increased investments in HPC types of resources for life science applications. Those investments occur at a variety of levels ranging from individual researcher to small labs, and at times a cluster or large server supporting a department. Furthermore, life science researchers are becoming more prominent users of HPC resources at centers such as the Ohio Supercomputer Center (OSC), the San Diego Supercomputer Center (SDSC), and the Pittsburgh Supercomputer Center (PSC).

We see increased investments in HPC types of resources for life science applications.

Within industry, we see a continued commitment to improving data management, the development of tools often focused on later stages of drug development and the establishment of groups of internal informatics specialists, computational chemists, and database and application programmers to develop, integrate, and support computational work in applied life sciences. Finally, nearly every major vendor of IT hardware has an internal group devoted to understanding the needs of life science researchers.

When a field is undergoing such rapid change, there will always be skeptics who see the failures and who will resist the adoption of newer methods. At the other end of the spectrum, there will be those who hype computer technology as a cure-all for the difficulties that the life science disciplines face. Although it is trying to avoid the dangers of hype, LSI does believe that the movement of these disciplines toward greater reliance on quantification and computer technology will continue to accelerate.

LSI believes that the movement of these disciplines toward greater reliance on quantification and computer technology will continue to accelerate.

Trends in Technical Computing Within the Life Sciences

This section provides some background on the types of HPC computing systems that we have seen in life sciences over the past several years.

High-End Systems

Through the 1990s, we saw a strong surge in investments in high-end computing installations in the life sciences driven by the private and public efforts to sequence the human genome. There was also significant investment by biotech companies in large compute clusters intended to leverage the sequencing of the human genome through creation of proprietary content products or directly in search of disease targets. Concurrently, there was a greater awareness of grand challenge–type computational problems within the life sciences and a rising awareness in government labs and academic supercomputing centers of the needs of life science

Through the 1990s, we saw a strong surge in investments in high-end computing installations in the life sciences driven by the private and public efforts to sequence the human genome.

researchers. In 2002 and early 2003, there appeared to be a reduction in purchases of large clusters and other high-end computing systems. Over the past year, however, there appeared to be some resurgence in sales of these systems. Throughout this time period, large pharmaceutical companies have continued to invest in SMP systems, mostly in the range of 8–64 processors, for managing the increasingly large databases in this domain.

Special-Purpose Systems for Algorithm Acceleration

During the period of strong optimism in the late 1990s, a number of small firms brought to market special-purpose computing systems aimed specifically at accelerating key algorithms needed in bioinformatics. Companies such as Compugen, Timelogic, and Paracel created these systems using application-specific integrated circuits (ASICs) or field-programmable gate arrays (FPGAs). Although these special-purpose systems dramatically accelerate performance on the targeted algorithms, the difficulty of reprogramming them, even for a minor variation in an algorithm, has prevented widespread adoption.

Workstations

Although high-end workstations have been a mainstay, particularly for modeling and simulation used in lead optimization or for modeling of protein structures, we see that the rise of Linux running on a standard PC equipped with a high-end graphics card is starting to take over those applications. As we will see in case studies, Apple has been one of the beneficiaries of that migration due to its strong floating point performance, commitment to OpenGL, and ability to run Unix and Linux applications.

Cluster Computing

Also starting in the late 1990s, there was a movement, initially led by end users, to network regular personal computers or small IA-32–based servers into clusters and enable parallel computing for those who did not have the budgets to buy supercomputers. Initially called Beowulf clusters, these collections of personal computers have undergone significant transformation in the past several years. Compute clusters have progressed from a do-it-yourself project by scientists to a point where one can purchase a cluster as an intact system, complete with system management software, preloaded applications, and maintenance service contracts. And although it is good press for vendors when they sell a big cluster, the sweet spot for a cluster purchase tends to be in small systems with 32 or fewer nodes. Vendors, including Apple (the sponsor of this paper), IBM, and Sun, have developed small cluster solutions targeted at life science workgroups.

Compute clusters have progressed from a do-it-yourself project by scientists to a point where one can purchase a cluster as an intact system, complete with system management software, preloaded applications, and maintenance service contracts.

Clusters are now available in multiple form factors, predominantly:

- **Rackmounted 1U systems.** These rackmounted systems typically come with the reliability and management features of servers. By fitting into the standard datacenter rack, these systems allow a compact footprint for a cluster system and fit in better with datacenter environments.

- **Blade systems.** Blade-based clusters, while generally at a price premium versus those based on 1U systems, provide a denser configuration than 1U systems.

- **"Cluster-in-a-box workstations."** An even denser packaging is available from new market entrants that offer a small, power-conservative cluster with integrated networking in a box the size of a desktop computer. It remains to be seen whether this technology will create a sustainable market.

Grid Computing

Grid computing provides an alternative to traditional cluster architectures by amalgamating existing resources into single logical compute resources. Rather than purchase dedicated computers for computation, one can aggregate existing servers, desktop systems, or other systems into a networked collection of computers that receive computational instructions from a central grid controller. Grids can be geographically dispersed, architecturally heterogeneous, and will vary in size, capacity, and reliability.

Heterogeneous Environment — Trending Toward Clusters

Although each of these methods for approaching the technical computing needs in life sciences is still used, there is a clear trend toward greater adoption of cluster computing over the other alternatives. The use of clusters is further facilitated by the introduction of more commodity-priced 64-bit computers from Apple, with the Xserve G5, and systems from a variety of vendors that are based on AMD Opteron chips.

Although each of these methods for approaching the technical computing needs in life sciences is still used, there is a clear trend toward greater adoption of cluster computing over the other alternatives.

Alignment of Apple Products with Needs in the Life Sciences

Based on a comparison of 2002 end-user surveys in the life sciences with overall share of market based on shipments, LSI estimates that, historically, life scientists have been two to three times more likely than the average computer user to use Apple's Mac platform as a client computer. Thus, the life sciences have represented an attractive market for Apple. With the introduction of Mac OS X in March 2001, Apple began to attract additional customers in life science on both the desktop and server fronts. Mac OS X was well timed to participate in the shift away from proprietary Unix and Windows platforms and toward open source operating systems such as Linux within the life sciences. Mac OS X, being based on BSD Unix, provided users with an environment that could easily run a wide array of application codes, both from the traditional Unix environment and the burgeoning ranks of open source applications developed to run on Linux.

LSI estimates that, historically, life scientists have been two to three times more likely than the average computer user to use Apple's Mac platform as a client computer.

In this section, we describe the current computer products offered by Apple and their applicability and implications for life science organizations and individual users. We examine the alignment for the following product elements: Mac OS X, the PowerPC G5 processor, multiple form factors (from laptops to the newer Xserve G5 1U rack servers), Xserve RAID storage hardware, and the Xsan storage software. Finally, we discuss the Mac as a development environment for homegrown applications, the prevalence of life science applications available for the Mac, and the bioinformatics cluster solution offered by Apple.

Mac OS X

The decision to build Mac OS X with a FreeBSD Unix core has been very well received by scientists. In the case studies presented in this document as well as other conversations with scientists, a similar message emerges: The Mac platform is much stronger for being able to retain all the traditional Mac user-interface appeal, while having a familiar Unix environment with its performance, robust design, flexibility, and relatively easy access to a host of scientific applications written for Unix or Linux just a click away.

A few of the key features of Mac OS X with implications are covered in Table 1.

TABLE 1

Mac OS X Features and Life Science Implications

Feature	Life Science Implications
The software is based on BSD Unix.	Provides access to a large body of life science applications written for Unix and Linux platforms
Apple has applied its design principles and experience to a number of open source utilities and tools, improving them and integrating them tightly into the OS and adding slick GUI front ends to those utilities.	Aids with cluster management for the many scientists who could use additional computational power but do not have the skills or interest in IT management
Mac OS X uses a 64-bit file system with support for HFS and HFS+ Journaled as well as UFS.	Efficiently manages scientific equipment and applications that generate thousands of small files automatically
Math libraries are optimized to take advantage of PowerPC G5 capabilities.	Interests scientists running floating point, vector, or matrix algebra intensive codes
The software comes prepackaged with Xcode, a solid set of developer tools.	Provides a convenient and cost-effective platform for custom code development for deployment on a variety of systems

Source: IDC, 2005

PowerPC G5 Processor

The PowerPC G5 processor is used in Power Mac G5 and iMac desktops and the Xserve G5. The PowerPC G5 is manufactured for Apple by IBM Microelectronics and shares a common architecture with the POWER chips that power IBM's high-end servers. Apple's desktop systems represent a transition to a 64-bit architecture. Until recently, 64-bit computing was the exclusive domain of high-end RISC workstations and servers. Intel sought to break the RISC dominance with the Itanium and Itanium II processors. Although they were CISC in design, the costs for Itanium-based systems ended up closer to the RISC workstations and servers than to the ordinary PCs. Apple, with the PowerPC G5 systems, and a variety of vendors offering systems built on the AMD Opteron processor have recently succeeded in bringing 64-bit architecture into a very cost-competitive price range, closer to ordinary PCs than to

Apple, with the PowerPC G5 systems, and a variety of vendors offering systems built on the AMD Opteron processor have recently succeeded in bringing 64-bit architecture into a very cost-competitive price range, closer to ordinary PCs than to high-end workstations.

high-end workstations. Table 2 highlights a few of the key features of the PowerPC G5 chip and system architecture with implications for life science applications.

TABLE 2

PowerPC G5 Features and Implications for Life Science Applications

Features	Life Science Implications
64-bit processor with 12 special functional units, including special vector, floating point and integer functions	The 64-bit processors can access large memory spaces*, and, with the functional units, this processor has very good performance on math intensive codes. This impacts both modeling codes and image manipulation and analysis codes.
Dual-processor systems have 2 high-speed buses — a dedicated bus per processor (each bus is dual channel with 32-bit in and 32-bit out)	This enhances overall performance of the system for memory-intensive applications such as image analysis. Additionally, the dual bus system provides a performance boost for those types of applications, such as a number of bioinformatics applications, that are constrained more by I/O performance than by processor performance.
Low latency access to Gigabit Ethernet ports	The latency performance of the communications ports are critical when combining systems into clusters. Cluster applications such as Guassian 03 will see performance benefits.

* Physical memory size is limited at this time by system architecture, not by the capabilities of the processor.

Source: IDC, 2005

Systems

Macintosh computers are available in three primary form factors: laptops, desktops, and rackmountable 1U servers. The laptops are available in two primary product lines: the iBook and PowerBooks. Among academic researchers who often need to be mobile between classroom, office, lab, and on the road at conferences, laptops are a favorite option for client computing. Although all the benefits of Mac OS X are available on the laptop, PowerBooks are currently based on G4 processors. When more power is required in a client computing environment, the Power Mac G5 desktop can be purchased as a single- or dual-processor system with up to 8GB of DDR RAM.

The introduction of Xserve G5 marked a substantial broadening of Apple's Macintosh product strategy. First introduced in 2002 with G4 processors, Xserve G5 was introduced with dual 2.0GHz G5 processors in early 2004 and dual 2.3GHz G5 processors in December 2004. The 64-bit dual-processor G5 system in a 1U form factor is ready for rack deployment in an office or a datacenter. In particular, the Xserve G5 is gaining traction for use in HPC clusters. Factors that enable an Xserve G5 to function well in a cluster environment include:

The 64-bit dual-processor G5 system in a 1U form factor is ready for rack deployment in an office or a datacenter. In particular, the Xserve G5 is gaining traction for use in HPC clusters.

- 1U rackmounted form factor

- Low heat dissipation and energy requirements
- Low-cost, high-performance 64-bit processors
- Dual Gigabit Ethernet communication ports
- Remote management software
- Full remote monitoring of temperature and component health

As mentioned in the earlier discussion of Mac OS X, it is relatively easy to port applications designed for Linux clusters to run on an Xserve cluster. Finally, a couple of labs, Virginia Tech and the COSLA Army lab, have demonstrated that the Xserve G5 servers do scale up into large clusters of over 3,000 processors.

Virginia Tech and the COSLA Army lab have demonstrated that the Xserve G5 servers do scale up into large clusters of over 3,000 processors.

Consistent with the scalability of the Xserve G5 for high-performance applications are the Xserve RAID systems and the Xsan storage area network (SAN) software. The Xserve RAID is a 3U rackmountable system that can house 5.6TB of data storage in a variety of RAID configurations. The Xserve RAID systems have many of the standard features expected in a storage system, along with certified support for Red Hat Linux, SUSE Linux, Yellow Dog Linux, Windows 2000 Server, and Novell NetWare server operating systems in addition to Mac OS X Server. Inexpensive, yet capable storage systems are critical for life science departmental computing or lab settings where the addition of a single new scientific instrument might create tens of gigabytes of new data each day.

With the addition of Xsan storage management software running on Mac OS X Server and with third-party Fibre Channel switches, Xserve RAID units can be combined into an affordable SAN. SANs have typically been used in support of drug development and clinical trials and less frequently in basic research and early drug discovery. However, a SAN created from Xserve RAIDS is comparatively inexpensive, and at least one of the case study participants is using an Xsan system to provide a pool of storage for a computational cluster that supports basic research in an academic setting.

Life Science Applications

Life science software applications can be divided into three major categories: commercial packages developed, maintained, and supported by for-profit companies; freeware and open source applications that are generally developed in an academic or government lab setting and made freely available to the scientific community; and homegrown custom applications developed by both academic and commercial life science labs for their own in-house use. This section provides an overview of those three types of application codes in the Mac environment.

Commercially Supported Applications

Within the life sciences, there are hundreds of commercially available software applications written for the native environment (Mac OS 9 or Mac OS X). The following is a small subset of these:

- Bioinformatics:

- **BioTeam's iNquiry.** iNquiry includes approximately 200 open source bioinformatics applications packaged in an easy-to-use interface and enabled to run on a Mac cluster.

- **Gene Code's Sequencher.** Sequencher is used for sequence analysis and contig assembly.

- **Geospiza's Finch.** Finch provides data management and bioinformatic applications for sequence analysis.

- **MiraiBio's DNASIS.** DNASIS is new to Mac OS X and offers sequence editing, multiple alignment, and homology searching in a commercially supported package.

- **Agilent's GeneSpring.** GeneSpring is a well-rounded commercial package for gene expression analysis.

- **BioDiscovery's GeneDirector.** GeneDirector is a microarray analysis and data management package.

- Medical imaging/microscopy:

  - **Intelligent Imaging Innovations' (3I's) SlideBook 4.0.** SlideBook 4.0 is a leading microscope acquisition control and analysis package.

  - **MediaCybernetics' QED InVivo.** QED InVivo is a live cell imaging application that includes both process control and analysis tools.

  - **Improvision's Volocity & OpenLab.** Volocity & OpenLab provides 3D imaging software and image management software for the life sciences.

- Chemistry:

  - **OpenEye Scientific's Omega & ROCS.** Omega & ROCS provides high-throughput structure generation and searches.

  - **HKL's HKL2000.** The HKL2000 is a leading commercial x-ray crystallography analysis and management package.

  - **Gaussian's Gaussian 03.** The Gaussian 03 is a leading quantum chemistry modeling package.

  - **Chemical Computing Group's (CCG's) Molecular Operating Environment (M.O.E.).** This commercial package incorporates high-throughput discovery, bioinformatics, and computer-aided molecular design tools into a single environment.

- Instrumentation:

  - **Becton-Dickinson's FACStation.** This application provides flow cytometry analysis and instrument control.

- **Molecular Devices' SoftMax Pro.** This is control software and plate reading technology for instruments.

- **Roper Scientific.** Roper Scientific provides FireWire microscope cameras and PCI-X boards for image capture.

- Mathematics/statistics/simulation:

  - **Wolfram's Mathematica.** Mathematica is a commercial symbolic mathematics package.

  - **The MathWork's MATLAB & Simulink.** MATLAB & Simulink are leading development tools for algorithm modeling, system simulation, and raw data manipulation.

- Laboratory information management systems:

  - **Ocimum's Biotracker.** Biotracker provides project management, reporting, and a complete lab workflow tracker.

  - **NuGenesis' SDMS.** SDMS provides client-side access to the NuGenesis data system.

Freeware, Shareware, and Open Source Applications

There are hundreds of freeware, shareware, and open source applications written for Linux and Unix platforms that can be run on Mac OS X. Typically, it is fairly straightforward to port applications between versions of Unix and Linux. It may still be a bit intimidating for the average end-user scientist if they have not done any porting or compiling of code. However, many of the most popular life science applications in this space have already been ported to the Mac OS X platform and are readily available. A small sampling of applications is listed below:

There are hundreds of freeware, shareware, and open source applications written for Linux and Unix platforms that can be run on Mac OS X.

- **EMBOSS,** from MRC Rosalind Franklin Centre for Genomics Research

- **WU BLAST,** from Washington University St. Louis

- **HMMER,** from Erik Lindahl

- **FASTA,** from William Pearson, University of Virginia

- **Visual Molecular Dynamics (VMD),** from the Theoretical Biophysics Group, University of Illinois

- **Java TreeView,** open source gene expression visualization and clustering

- **R,** open source statistics and math language and modeling tool — including a full native user interface for Mac OS X

- **MRI/CAT/PET,** from Osirix (open source package for navigating multidimensional DICOM images)

Homegrown Custom Applications

Within the life sciences, custom applications are quite common. In some biotech companies, we have seen that as much as 80% of scientific applications are written within the company. On average, in life sciences, 24% of application codes are written in-house, and another 18% are custom codes written by outside contractors. With custom applications making up about approximately 40% of the mix, it is important when thinking about the Apple products to consider the development tools environment.

The development tools include Cocoa, a native Mac OS X object-oriented framework that provides preprogrammed functionality that is especially useful in creating user-interactive programs. Mac OS X also includes the Xcode developers' toolkit, free with the operating system. Xcode includes gcc 3.3, a widely used compiler. Developers can also work with other built-in development tools such as gdb, vi, emacs, and pico and can take advantage of Unix shell tools such as grep, chmod, ps, crontab, top, and tail. In addition, programmers can use the Unix environment of Mac OS X to code in common scripting languages such as Perl, PHP, Python, and Tcl. Unix programmers also rely on the X11 windows support provided by Mac OS X. That support for X11 eases the migration of many applications written for Unix or Linux platforms.

Java support is strong. For example, Xserve G5 comes preloaded with WebObjects Java Application Server. WebObjects gives one the ability to build or use standards-based Web services without writing low-level SOAP, XML, or WSDL. Apple also supports J2EE technologies, including Object Request Broker and Enterprise Java Beans.

Xserve G5 comes preloaded with WebObjects Java Application Server. WebObjects gives one the ability to build or use standards-based Web services without writing low-level SOAP, XML, or WSDL. Apple also supports J2EE technologies, including Object Request Broker and Enterprise Java Beans.

Common programming languages such as C, C++, and Java are all well supported. And, as one of our case study interviewees points out, there are very good Fortran compilers such as IBM's XL Fortran for writing high-performance scientific code to run on Xserve G5 clusters.

Solutions: The Apple Workgroup Cluster for Bioinformatics

As part of Apple's focus on the life sciences, it has developed a workgroup cluster solution for bioinformatics. The solution includes hardware, software, and services.

The hardware consists of:

- 4–32 nodes of dual-processor Xserve G5 servers
- Gigabit switch for networking
- Rack enclosure
- Uninterruptible power supply and all necessary cabling

The software consists of:

- Mac OS X server operating system and Xcode development kit

- iNquiry system images, configuration, and network settings
- iNquiry Bioinformatics software from BioTeam (includes 170+ open source applications)

The service includes:

- Three-year AppleCare Premium Service and Support (for every node)
- 24 x 7 telephone support and four-hour onsite response
- iNquiry up-and-running support

This solution is important for Apple on a couple of dimensions. First, it communicates to life science buyers that Apple is thinking about their needs and paying attention to them as an important segment of buyers. Second, it is easy; being packaged as a solution makes it easy to buy, and case study interviews indicate that it is also easy to set up and use.

Understanding the Mac User Experience: Case Studies in the Life Sciences

In this section, we summarize in-depth conversations with nine different users or administrators of Mac products who work in life science environments. The case studies come from people in government, academic, and commercial pharmaceutical or biotech labs. Although they share common scientific interests, these three types of organizations are quite different in how they approach IT purchases and what matters to them:

- **Government.** Government research laboratories tend to be oriented toward formal acquisitions through a request for proposal (RFP) process. They also tend to provide a home for grand challenge–type thinking — the big science projects (e.g., the Human Genome Project). Government labs are also diverse in mission, with some labs having a mandate to be at the leading edge helping advance basic science, including computing technologies. In other labs, the mission is centered almost exclusively on solving a set of applied problems, and computing is simply a tool.

Government research laboratories tend to be oriented toward formal acquisitions through a request for proposal process.

- **Academia.** One of the key characteristics of academic labs' selection and purchase of information technology is the grant funding system, which impacts both the cycle of purchasing and the relative freedom that professors and principal investigators have in making purchase decisions. The decentralization inherent in the academic system also means that researchers are often more directly responsible for their own IT support than their counterparts in government or commercial settings.

The decentralization inherent in the academic system also means that researchers are often more directly responsible for their own IT support than their counterparts in government or commercial settings.

- **Industry.** Commercial settings break into about 10–15 very large pharmaceutical companies, about another 40 large pharmaceutical and biotech companies, and approximately 3,500 smaller pharmaceutical and biotech companies worldwide. Life Science Insights has found that although large pharmaceutical companies have large IT budgets, they tend to be conservative in IT purchases. Thus, they have been steady customers for large RISC-based SMP systems with enterprise-style service agreements. Compared with academic and government labs, scientists in pharmaceutical and the larger biotech companies tend not to have as much freedom to choose their own IT platforms. Corporate IT departments tend to be focused on the standardization of IT platforms wherever possible, although R&D departments are given more freedom in that regard than corporate support staff or sales reps. Within many R&D departments, there is a shadow IT function that occurs at the workgroup level. The workgroup leader often has signing power for up to about $40,000–50,000 and will sometimes use that authority to purchase IT resources that only serve their own workgroup and are supported directly by the workgroup itself.

Within many R&D departments, there is a shadow IT function that occurs at the workgroup level. The workgroup leader often has signing power for up to about $40,000–50,000 and will sometimes use that authority to purchase IT resources that only serve their own workgroup and are supported directly by the workgroup itself.

Some of these differences will be reflected in the following case studies, which examine how a variety of scientists from government, academia, and industry are using Macs and products from Apple to enable their research.

Case Study: Stephan Bour, Ph.D., The National Institute of Allergy and Infectious Diseases

Dr. Stephan Bour started his professional life as a postdoctoral fellow at the National Institute of Allergy and Infectious Diseases (NIAID), continuing his study of the molecular biology of the human immunodeficiency virus (HIV). Along the way, this bench scientist started to concentrate on the informatics elements of his work and became a bioinformatics expert and a resource for colleagues. In 2002, that role was formalized, as Bour's proposal to start a bioinformatics core facility (BCF) at NIAID was approved. He has recently had his scope of responsibilities expanded to include the management of the NIAID Bioinformatics Program at the Office of Technology and Information Systems (OTIS). This multidisciplinary program is mainly directed at promoting, implementing, and supporting the use of bioinformatics resources in infectious disease and biodefense research. The program aims at addressing the needs of over 1,000 scientists and clinicians working at NIAID.

Bour discussed with us the IT infrastructure and practices at NIAID and the role that Macintosh computers play in its work. Currently, the primary role of Mac products is as client computers, both laptops and desktops. In this organization, scientists have a good deal of freedom to choose their own client computers. Bour estimates that although the large majority of client computers on the administrative side of the organization are Microsoft Windows–based systems, on the scientific side of the organization, about half of the client computers are Mac products. He indicates that although a portion of the Mac user base is made up of traditional Mac users, they are being joined by new users drawn to Mac because of Mac OS X. The fact that Mac OS X is based on a Unix operating system makes this a strong platform for those scientists accustomed to the power and flexibility of Unix and Linux.

Bour estimates that although the large majority of client computers on the administrative side of the organization are Microsoft Windows–based systems, on the scientific side of the organization, about half of the client computers are Mac products. He indicates that although a portion of the Mac user base is made up of traditional Mac users, they are being joined by new users drawn to Mac because of Mac OS X.

NIAID makes use of many Microsoft products in its network and storage infrastructure, including Exchange Server, SQL Server, Active Directory, and other Windows Server products. In addition to these systems, Bour notes that NIAID has installed approximately 20 Xserves into its predominantly Windows infrastructure. Recent offerings by Apple are generating renewed interest in the areas of supercomputing, SANs, and RAID storage. The NIAID is currently planning on acquiring its first Apple Workgroup Cluster for Bioinformatics and is increasing its reliance on Apple Xserve RAID arrays for mass storage.

Some of the key scientific applications used on the Macintosh include MacVector from Accelrys, Vector NTI from Invitrogen, Cellquest from Becton Dickinson, and FlowJo from Treestar. Many of these applications or functional equivalents are available on Microsoft Windows as well as on the Mac. In general, NIAID tends to purchase commercially supported application codes whenever possible. Open source codes typically are only used when a commercially supported alternative is not available or are used individually by advanced users. The organization, and OTIS in particular, increasingly has developed algorithms and custom code to address the evolving needs of the scientific community. Development includes database and data management, dynamic Web interfaces, and clinical data mining methods. An overwhelming proportion of these custom applications are developed in Windows-specific frameworks such as ASP .NET and SharePoint. Although code is also generated in cross-platform Perl and Java, Apple development tools such as WebObjects, AppleScript, and Cocoa are not often used. As such, users of these codes on the Macintosh are running them in their pure Unix or Linux implementations.

On the office productivity side, the organization primarily uses Microsoft Office, Outlook, and Internet Explorer. On the Macs, the email is handled using POP3, and users can choose mail clients from among Eudora, Microsoft Entourage (part of the Microsoft Office for Mac suite), and Apple's own email application, Mail, which is included in Mac OS X. Two other products that bridge the gap between the scientific and the administrative tasks are Illustrator and Photoshop. The large majority of scientific instruments now produce digital output. These images often need to be manipulated for presentation or have certain features enhanced or overlaid with arrows, labels, or other designators.

During our discussion of integration in a multiplatform environment, Bour clearly indicated that Macs can and do operate in a cross-platform environment. And although on the one hand he likes a diverse IT environment because it means that no single incident can bring the whole institution down, that diversity does bring some complications. For example, although Apple has provided a plug-in that enables Mac OS X to connect to Active Directory in Microsoft Exchange 2003, NIAID has not yet upgraded to Exchange 2003, which means that Mac email clients do not get group scheduling and global directory access. The use of POP clients that physically download email to the client computer complicates the task of backing up email directories and shifts the burden onto the user for data integrity and security. The older version of Active Directory used at NIAID prevents the use of single sign-on for Mac users, although this functionality is supported in Mac OS X for Exchange 2003. The lack of a true multitasking enterprise-level backup package for the Macintosh has also complicated requirements for data integrity and security. Some Mac-oriented workgroups use separate backup servers using commercial packages such as Retrospect, but this approach is a client push, rather than a server pull, method of backup.

During our discussion of integration in a multiplatform environment, Bour clearly indicated that Macs can and do operate in a cross-platform environment.

The lack of tight integration has a positive side effect for many end users who enjoy a bit of independence from the centralized IT administration — it gives them a freedom and responsibility that appeals to many scientists. One of the other side benefits of diversity in computer architecture is that, up to now, Mac computers have been immune to the large virus and worm attacks that have sometimes crippled Windows-based machines. Bour speculates that it is partly because lower market share makes Mac a less tempting target, but he also points out that the Unix core of the OS tends to be more secure from an architectural perspective. Administrators spend less time helping clean viruses, adware, and malware off of Mac machines, which is one reason Bour believes that the Mac clients have a somewhat lower cost of ownership than the Windows-based clients.

Administrators spend less time helping clean viruses, adware, and malware off of Mac machines, which is one reason Bour believes that the Mac clients have a somewhat lower cost of ownership than the Windows-based clients.

In looking at key reasons for scientists to choose Mac products, he points to ease of use for the end user; the design, quality, and originality of the hardware; the lower prevalence of viruses, worms, spyware, adware, and malware; and the flexibility and power of the Unix underpinnings for more sophisticated users. For life sciences specifically, the availability of thousands of Unix open source scientific applications is also a major plus.

In thinking about areas where Apple could improve, Bour pointed to what he sees as a lack of education among traditional Mac users about the real power and flexibility now available through a Unix environment. Apple could gain new levels of support among scientists by actively educating users on the resources that the Unix interface now makes available to them. Bour also believes that Apple should more aggressively promote and explain the Windows integration tools available to network administrators.

Apple could gain new levels of support among scientists by actively educating users on the resources that the Unix interface now makes available to them.

Finally, Bour indicated that Apple needs to be aggressive in working to keep key life science software vendors on board. It is discouraging to see long-time Mac developers such as Becton Dickinson, Accelrys, Invitrogen (formerly InforMax), and Applied Biosystems shift their main focus to the Windows platform or abandon the Macintosh altogether.

Case Study: Mark Cohen, Ph.D., Associate Professor of Neurology, UCLA Department of Neurology

Mark Cohen, Ph.D., associate professor of Neurology at the UCLA Department of Neurology, Brain Mapping Division, specializes in the use of functional magnetic resonance imaging (fMRI) and electroencephalograms (EEG) to study a variety of cognitive phenomena, including mental imagery, time perception, and lying. His group also studies changes in brain function related to chemical addictions, schizophrenia, and Parkinson's disease. In addition to traditional neurology studies, Cohen is also directly involved in the improvement of existing and development of new scanning technologies.

The types of computationally intensive applications used by Cohen's group include image capture and analysis; integration of images and EEG signals; and statistical analysis, including methods such as Independent Components Analysis and General Linear Models as well as parametric simulations. Most of the software applications are either open source applications such as bioinformatics tools, custom codes developed within Cohen's own lab for image processing, or analytic routines to support particular research projects. Among commercial codes, MATLAB is a particular favorite.

The types of computationally intensive applications used by Cohen's group include image capture and analysis; integration of images and EEG signals; and statistical analysis, including methods such as Independent Components Analysis and General Linear Models as well as parametric simulations.

Historically, the Brain Mapping Division had used a variety of RISC workstations, including Compaq Alpha, Sun, and SGI workstations for the analysis of fMRI images. For client and personal computing tasks, Cohen has been a long-time Apple computer user, starting with an Apple II and continuing to up to present-day Mac desktops and laptops. Over the last couple of years, the personal and the professional have been united. With the introduction of Mac OS X, Cohen was able to run Unix-based analytic applications on a Mac, removing the need for other Unix-based workstations. Within his research group, Power Mac desktops and iBook or PowerBook laptops are now the predominant client computers.

With the introduction of Mac OS X, Cohen was able to run Unix-based analytic applications on a Mac, removing the need for other Unix-based workstations.

For larger-scale computation, the Department of Neurology has had a Linux cluster in place for a couple of years. In the past few months, Cohen's Brain Mapping Group decided to expand its computational resources and purchased an 8-node Xserve G5 cluster with dual processors and 1.5GB of RAM per node. For data storage, they purchased an Xserve RAID with approximately 1TB of RAID storage. Although Cohen and his research group are the primary users of the Xserve cluster, access is available to the whole department, including 10 faculty and about 60 graduate students and postdocs. Physically, the cluster resides in one of the university datacenters; however, Cohen uses remote administration tools to do account and job administration from his own office.

In comparing his experience with the Xserve cluster to that of a colleague responsible for a Linux cluster, Cohen indicated that it was taking significantly less time to administer the Xserve cluster. That time savings enables researchers and department programmers to spend time on the productive pursuit of science and tool development rather than IT administration. Cohen indicated that he was initially surprised but pleased to learn that Apple had developed tools and user interfaces that make administering an Xserve server or a cluster almost as easy as administering a single desktop Mac.

In comparing his experience with the Xserve cluster to that of a colleague responsible for a Linux cluster, Cohen indicated that it was taking significantly less time to administer the Xserve cluster.

Cohen also commented on the relative ease of migrating codes to run on the Xserve cluster. His group needed to migrate a collection of codes previously run on Sun, SGI, and Alpha workstations or servers. It took 100% of a programmer's time and 25% of Cohen's time for about two weeks to port all the legacy codes to the Xserve cluster. Now that the codes have been ported, the team has started to look at how it can use optimized math libraries, such as LAPACK and BLAS, both of which are included in Mac OS X, to improve the performance of key applications. They are discovering some solid performance increases (tenfold speedup for portions of code) and look forward to more extensively exploiting the Velocity Engine capabilities with an autovectorizing compiler that will be included in an upcoming release of Mac OS X.

…relative ease of migrating codes to run on the Xserve cluster. ... It took 100% of a programmer's time and 25% of Cohen's time for about two weeks to port all the legacy codes to the Xserve cluster…. Discovering some solid performance increases … more extensively exploiting the Velocity Engine....

When thinking about Apple's challenges in serving scientists such as himself, Cohen indicated that scientific buyers have come to expect contact from a direct sales force that knows science and computing and feels comfortable addressing technical questions. For academic scientists, it is also very helpful if a vendor provides resources to assist in writing grant proposals. He believes that this type of attention is what Apple needs to provide to raise its profile among scientists and become a more obvious computing choice for the scientific community.

Case Study: Michael Barmada, Ph.D., Assistant Professor of Human Genetics, University of Pittsburgh

Michael Barmada, Ph.D., assistant professor of Human Genetics at the University of Pittsburgh, and his research colleagues analyze linked family data to identify genetic susceptibility loci for common but complex diseases. The department now includes five full-time professors, six graduate student research assistants, six postdoc researchers, two software programmers, and two part-time technical assistants. The work within the department is about evenly split between theoretical statistical genetics and applied research on various diseases. The computational needs of the department are driven by both recent increases in size of the department and the increasing size of the data files that they need to analyze. Barmada indicates that they are now conducting linkage analysis on data sets that include thousands of families and sometimes as many as 500,000 genetic markers.

The computational needs of the department are driven by both recent increases in size of the department and the increasing size of the data files that they need to analyze. Barmada indicates that they are now conducting linkage analysis on data sets that include thousands of families and sometimes as many as 500,000 genetic markers.

As the demand for computational resources has grown, Barmada has gained experience using and managing a range of computer systems. His department formerly used a four-processor Sun server as a shared resource for computationally intensive jobs. Then, about three years ago (when the Sun server was due for replacement), Barmada instigated a change in architecture and helped the department put a Linux cluster in place. This cluster was pieced together over a period of months and ended up with a mix of Dell and VA Linux boxes, with processors including Pentium IIIs, Pentium 4s, and Xeons. Early in 2004, word came that a grant had been approved that included a budget for a new, larger computational cluster. Although originally contemplating adding another Linux/Intel cluster to the infrastructure, Barmada became aware that Apple had introduced the Xserve G5 and that it was available for clustering. Based on previous experience with using a couple of Xserve G4 servers in standalone applications, he concluded that the Xserve G5 would be easy to administer. This experience, among other factors, led him to spend the computer cluster budget on an Xserve G5 cluster with 4 head nodes (dual processors and 2GB of RAM), 116 compute nodes (dual processors and 2GB of RAM), and 5 large memory compute nodes (dual processors and 8GB of RAM). To provide data storage for the cluster, the group added an Xserve RAID with 2.4TB of disk storage. The storage is administered using the Xsan storage management software.

This experience, among other factors, led him to spend the computer cluster budget on an Xserve G5 cluster with 4 head nodes (dual processors and 2GB of RAM), 116 compute nodes (dual processors and 2GB of RAM), and 5 large memory compute nodes (dual processors and 8GB of RAM).

In addition to the shared resources, this department has also chosen Apple products for its client computing environment. The professors have G4 and G5 Power Mac desktops, and the graduate student research assistants and postdoc researchers are provided with iBook laptops.

One of the key reasons cited for using Macs is that they are easy to use and administer. The nature of academic funding mechanisms is such that research groups, such as the human genetics group where Barmada works, have a great deal of freedom to choose their own computational infrastructure, but with that freedom comes responsibility. Barmada takes the somewhat unusual step of administering the Xserve cluster himself. Although the cluster is physically housed in a datacenter some 6 miles away, he uses a VPN connection, Apple's remote administration tools, and Sun Grid Engine middleware to administer the cluster from his own desktop.

One of the key reasons cited for using Macs is that they are easy to use and administer.

Barmada commented that installation of the cluster took less than a week. The department had assistance from a local integrator, Cross IT, to set up the hardware and assistance from the BioTeam consulting group to install the cluster configuration settings and the iNquiry toolset. More specifically, the hardware installation took one day, the cabling took another day, and installation of the iNquiry tool set on all the nodes only took about two hours.

Among all the scientific applications that the group uses, all but one have been successfully moved from the Linux cluster to the Apple cluster. Most of the codes that they use are written in C or Fortran, but there are still some older legacy applications written in Pascal. As he looks to the future, Barmada mentioned that the group would like to optimize many of these application codes to take full advantage of the Velocity Engine vector processing unit on the PowerPC G5 processor. Statistical genetics codes make significant use of matrix calculations, and profiles of the applications indicate significant room for optimizing performance through better use of the vector processing unit.

The group would like to optimize many of these application codes to take full advantage of the Velocity Engine vector processing unit on the PowerPC G5 processor. Statistical genetics codes make significant use of matrix calculations, and profiles of the applications indicate significant room for optimizing performance through better use of the vector processing unit.

Barmada indicated that the key reasons for selecting Apple products on both the desktop and in a cluster computing setting included:

- The ease of administration mentioned earlier
- Good design (i.e., it is aesthetically appealing or "cool")
- The ability to consolidate computers (e.g., the command line Unix and traditional Mac environment are on the same machine)
- Easier, seamless working environment (e.g., no barrier between desktop and shared cluster system)
- Competitive price performance

When asked about areas in which the Apple products needed improvement, Barmada pointed to some of the server administration tools for DNS, DHCP, and Netboot as well as some of the storage management tools in Xsan. In his view, these tools are reasonably good, but they still need some software engineering and user-interface work to bring the tools up to the high Mac standard. According to Apple, many of these enhancements are planned for the upcoming release of the Mac OS X Server operating system.

In terms of Apple's relationship to the scientific communities, Barmada was encouraged to see that Apple is showing up with a significant booth presence at key scientific conferences within his own discipline, and he views that as a good sign of Apple's interest in the needs of scientists in general.

Case Study: Rick Hoge, Ph.D., Instructor, Harvard Medical School's Department of Radiology

Rick Hoge, Ph.D., is an instructor at the MGH/MIT/HMS Athinoula A. Martinos Center for Biomedical Imaging and Harvard Medical School's Department of Radiology. He has worked in this lab since 1999, applying his education in biomedical engineering and his long-term focus on MRI technology to the discipline of functional brain imaging. Functional brain imaging uses tools and techniques such as MRI to map localized brain activity through changes in energy demands and other molecular signatures and correlates that physical map with stimuli responses or cognitive tasks. Hoge works with about a dozen other researchers in his own specialized area and a larger group of about 200 persons involved in brain imaging for clinical and research purposes. The Martinos Center, although housed at Massachusetts General Hospital, serves collaborators from both MIT and Harvard.

Hoge has been intimately involved in both using and developing computational tools to assist in neurological imaging. In the last couple of years, since Apple introduced Mac OS X, he has seen and been a part of a migration from Linux/Intel to Apple systems, now approximately 10–15% of researchers. In his group, the primary experience is with the G4 and now Power Mac G5 desktop systems. When asked to account for this shift, Hoge points to the flexible nature of the Mac OS X–based systems, which allow researchers to use the same system for Linux- or Unix-based scientific applications, general productivity applications such as MS Word and Excel, commercial imaging software such as Adobe Photoshop, and various presentation applications.

The flexible nature of the Mac OS X–based systems, which allow researchers to use the same system for Linux- or Unix-based scientific applications, general productivity applications such as MS Word and Excel, commercial imaging software such as Adobe Photoshop, and various presentation applications.

Hoge's own experience with the Apple systems includes developing analytical software. Recently, he released a neurological image analysis and image management tool called NeuroLens, available free for academic and nonprofit research use. This analytical software was developed on and is available for Mac OS 10.3 or later on both the G4- and G5-based Macintosh systems. When explaining why he developed this package using Apple's native development tools, he pointed to three major factors:

- The Cocoa programming framework is well engineered, simplifying the development of a solid application through the use of software libraries and tools for adding interactive user interfaces.

- A set of optimized math libraries is included with Mac OS X, which is particularly helpful in developing scientific applications.

- They provide good performance, which is due to the interaction of optimized math libraries and specialized functional units such as the Velocity Engine vector processor on the G4 and G5 chips.

Also, Hoge pointed out the following:

- The application will run with full features only on Mac OS X.

- The application can be ported to other platforms, but at the lowest common denominator, losing much of its interactivity.

- Porting is never totally easy.

When weighing out the benefits and challenges for Apple in research organizations such as the Martinos Center, Hoge indicated that Apple's success in the consumer market is both a drawing point and a potential hindrance. People enjoy the usability and nice design, but sometimes wonder if a cool consumer-oriented company can produce a serious science computer. Clearly, Hoge thinks Apple can do both, and he chooses to develop scientific software for the platform. He notes that it is generally perceived that Macs are priced at somewhat of a premium relative to PCs and says that Apple needs to make sure that its build quality is top-notch. Hoge acknowledges that build quality issues afflict all vendors, and, in his experience, the service and support organization from Apple has been responsive and timely in fixing any problems that he and his colleagues have encountered.

Build quality issues afflict all vendors, and, in his experience, the service and support organization from Apple has been responsive and timely in fixing any problems that he and his colleagues have encountered.

Hoge comments that many of the open source applications for his discipline have been ported from Linux to Mac OS X, but some ported codes are still a bit buggy. On the other hand, when thinking about the operating system environments more broadly, he believes that many researchers using Linux underestimate the amount of time they spend doing what are essentially IT administrative tasks that are much simpler or completely unnecessary with Mac OS X.

Case Study: Millard Alexander, Ph.D., Distinguished University Professor, Department of Chemistry and Biochemistry, College of Life Sciences, University of Maryland

Dr. Millard Alexander, a theoretical chemistry professor and head of a small group of researchers at the University of Maryland, leads research in the theoretical study of inelastic and reactive molecular collisions, particularly those involving free radicals; the photofragmentation of small molecules; and the structure and energetics of weakly bound complexes involving open-shell species. This basic research is firmly within the theoretical chemistry space, with more relevance to combustion than to life science; however, his perspective on computation and computer systems has relevance for anyone in the sciences who relies on mathematical models and simulations to assist in exploring and understanding physical phenomenon.

Alexander's group primarily uses Mac products, a mix of PowerBook laptops, and Power Mac desktops. In addition to the Mac client computer systems, his group shares a 16-node dual-processor Xserve cluster system with one other research group. In total, this cluster system is shared among 7 or 8 researchers. Typically, a job or run will be assigned to 3 or 4 nodes on the cluster, enabling 4 to 5 computational jobs to run simultaneously. Alexander indicates that the scale of the problems that they are trying to solve rarely requires more than 3 or 4 nodes, and the problems tend to exhibit coarse grain parallelism, which means that a loosely coupled cluster such as that built with the Xserve works very well. Other, shared IT resources in the group include access to a Linux workstation, an HP workstation, and an older IBM RS6000 4-processor system.

In addition to the Mac client computer systems, his group shares a 16-node dual-processor Xserve cluster system with one other research group.

For scientific computing, Alexander's group uses MATLAB extensively and develops most of the remainder of the application code in-house. The main programming environment is Fortran, and they use the IBM XL Fortran compiler from Absoft, which has been optimized for the PowerPC chip used in the G5 desktops and the Xserve cluster. The codes that they develop tend to use many matrix operations (multiplies, adds, and diagonal transforms). These operations are well supported through the design of the PowerPC chip and the XL Fortran compiler.

The codes that they develop tend to use many matrix operations (multiplies, adds, and diagonal transforms). These operations are well supported through the design of the PowerPC chip and the XL Fortran compiler.

General productivity and communications software that Alexander's group uses includes Apple Mail, the Camino browser from Mozilla, Keynote for presentations, and TeXShop (a LaTeX document preparation system for scientific writing). In addition, the group also has Microsoft Office products for general office productivity, file exchange, and ease of communications with the broader world.

With regard to platform integration, Alexander indicated that most of their work is on the Mac platform, but the group connects easily to the broader university IT network for communication.

When asked about the benefits of and reasons for choosing Mac systems for his research group, Alexander commented on the ease of setup, ease of use, and ease of administration, particularly with regard to staying free of computer viruses. He also commented on a certain feel and interaction between the mouse and screen that he does not find on other machines. Although a long-time personal user of Macs for writing, building presentations, and communicating with others, Alexander pointed out that Mac OS X and the Xserve G5 have opened up new uses for Macs within his lab in the last year. Because Mac OS X is based on Unix, the group is now using Macs to develop computational code. And, with the power of the G5 chip and a good Fortran compiler, they are running the computational codes on Power Mac G5 desktops or the Xserve cluster. Previously, those codes would have required a Unix workstation or small Unix server. In general, Alexander now views Mac systems as a price-competitive way to meet his group's computational requirements.

Although a long-time personal user of Macs for writing, building presentations, and communicating with others, Alexander pointed out that Mac OS X and the Xserve G5 have opened up new uses for Macs within his lab in the last year.

Although he was asked what particular weaknesses of the Mac systems or ways that Apple could serve scientists better, Alexander had difficulty pointing to any. His only advice to Apple was to keep focusing on the broad scientific market, both traditional physical sciences as well as life sciences, and keep communicating the Unix capabilities now available through Mac OS X.

Case Study: Alan Goates, Director of Bioinformatics, Isis Pharmaceuticals

Alan Goates, director of Bioinformatics at Isis Pharmaceuticals Inc., was the first bioinformatics employee hired by the department head and was brought in seven years ago to start a bioinformatics function and department. Isis Pharmaceuticals is a drug discovery and development company focused on RNA. Isis was granted marketing clearance for the world's first antisense drug, Vitravene (fomivirsen), in 1998. Vitravene is marketed by Novartis Ophthalmics.

Over the years, the original bioinformatics team has grown, subsumed the IT department, and is now organized in three groups: IT support, drug discovery support (internal bioinformatics support for bench scientists), and bioinformatics research contract support. For several years, Goates led efforts in the internal bioinformatics support function, and he now leads the external bioinformatics services using proprietary techniques developed at Isis to support U.S. government biodefense contracts.

Goates estimates that about 10% of the bioinformatics applications used by scientists at Isis are provided by commercial software companies. Another 40% are open source applications, although they are often modified by Isis scientists and informatics programmers, and the remaining 50% of code is custom and unique to Isis. Given this reliance on custom software, the general practice in software development has been to strive for platform independence wherever possible. Front-end interfaces for applications are typically written in Java or Perl, and back-end database storage, queries, and data manipulation are created in either Oracle or MySQL, depending on the scale of the application.

In talking about the role of Macs and Xserve computers in their infrastructure, Goates reminisced about the early days of joining Isis. When he and his boss were founding the bioinformatics department seven years ago, they were both self-described "Mac heads." Initially, the IT department, for the sake of consistency and standards, did not want to allow Macs. However, with some perseverance, persuasion, and a willingness to support themselves, these two were allowed to use Mac desktop computers. Over the years, as their department became more influential, the policy was changed so that scientists had the freedom to choose the computer platform that best supported their own area of work. With that corporate openness, Mac usage has risen to approximately 10% of client computing. Many of those choosing a Mac for a client computer are choosing the 17in. PowerBook as a desktop replacement.

With that corporate openness, Mac usage has risen to approximately 10% of client computing. Many of those choosing a Mac for a client computer are choosing the 17in.-screen PowerBook as a desktop replacement.

In terms of shared computing resources, Isis has Sun servers running Oracle databases and a Linux cluster running genomics analysis packages. Isis purchased an Xserve from Apple because it needed to use WebObjects in a new development project for one of the defense contracts, and purchasing an Xserve was a quick and easy way to get WebObjects up and running because it came preinstalled and fully supported. Then, the government partner suggested that the company might invest in an Apple Workgroup Cluster for Bioinformatics with the iNquiry software from the BioTeam. Isis followed that suggestion and added a 4-node Xserve bioinformatics cluster to use for development of parallel code. Isis' programmers have also started to integrate some of their own code with that included in iNquiry and plan to add similar user interfaces to their own proprietary code.

In commenting on the benefits and challenges of using Macs and Xserve servers, Goates indicated that the biggest challenge is integration with Lotus Notes, but there are workable solutions. Key benefits that he has observed are better protection from virus and worm attacks, full Unix under the hood, access to open source tools and applications, and a lower support requirement. On that last point, he indicates that although Macs now make up 10% of the client computers, they do not require 10% of the IT support staff. He suspects that this gap is caused by a combination of the fact that Mac users are more accustomed to supporting themselves and that the systems are actually more stable and therefore easier to support.

Key benefits that he has observed are better protection from virus and worm attacks, full Unix under the hood, access to open source tools and applications, and a lower support requirement.

On his wish list for either Apple or the BioTeam is a set of data download tools that are really slick for capturing updates from databases such as those at NCBI. He acknowledges this is a challenge because any such tools need to be continuously maintained to reflect the most current changes in download interfaces and policies at NCBI and other key content providers. According to Apple and The BioTeam, such a capability will soon be part of iNquiry and the Apple Workgroup Cluster for Bioinformatics.

Case Study: Brian Gilman, President, Panther Informatics

Brian Gilman, with a dual background in biochemistry and computer science, has a career that has included think tank–type research at InterScience Inc., software development at Allaire Corp., and five years at the Whitehead Institute. He is now founder and president of a small informatics service company, Panther Informatics. The primary work of Panther Informatics is to provide application integration and analytical workflow solutions to government, academic, and pharmaceutical R&D labs engaged in drug discovery and preclinical studies. While at Whitehead, Gilman was a key developer of the Omnigene application. He continues to develop software, both in custom settings for clients of Panther Informatics and as products that Panther Informatics hopes to introduce to the market in the near future.

While at Whitehead, Gilman was a key developer of the Omnigene application. He continues to develop software, both in custom settings for clients of Panther Informatics and as products.

Gilman's experience with Mac computers extends back for many years. He has always used a Mac desktop or laptop as his primary client computer. His current IT environment is a heterogeneous set of platforms. Panther Informatics has four full-time employees as well as a small network of freelance contractors that bring specialized skills as needed. The client computers for the full-time employees include two PowerBook G4 laptops, a dual-processor Power Mac G5 desktop, and a Windows/Intel system. For servers, Panther maintains a dual Opteron system with 8GB of RAM and a dual Xeon with 4GB of RAM. These servers run on Linux and host scientific, informatics, and database applications. In addition, there is a Windows system used for quality control applications and another Windows system used for development and testing of applications intended to run in a Windows environment. Panther also maintains approximately 25TB of disk storage in a self-built SAN environment.

Key applications include a wide variety of commercial and open source bioinformatics and analytical applications such as BLAST and Mascot as well as database packages such as Oracle and MySQL. These technical applications are run on both the Mac systems and the Linux systems. The Windows machines are primarily used for business applications such as QuickBooks, project management, and a quality control application.

Although Gilman works in a heterogeneous IT environment, he is able to do all of his software development work on his Mac laptop. He keeps Oracle installed on the laptop for database programming, he can do all his development for programs intended for either Unix or Linux platforms, and he has the tools to do .NET and C++ development for applications that will run on Windows systems. His key reasons for preferring the Mac as a development environment are centered around the following points. First, Mac OS X is, at its heart, a Unix operating system, which gives him a solid environment for developing and compiling Unix and Linux codes. Second, the system supports OpenGL "from top to bottom." Third, the slick user interface is a pleasant, feature-rich work environment, allowing him, for instance, to make a command line coding window semitransparent so that he can see a developer's help page on a browser in the background. Finally, he likes the fact that he can keep 15–20 windows open, clicking back and forth between applications, and the response is crisp. As he says, "multitasking really works" on the Mac.

Mac OS X gives him a solid environment for developing and compiling Unix and Linux codes, supports OpenGL "from top to bottom," and has a user interface that is a pleasant, feature-rich work environment where he can keep 15–20 windows open. The response is crisp. "Multitasking really works."

When thinking about extending his use of Mac products to the server side of his IT infrastructure, Gilman revealed a somewhat conflicted view. On the one hand, when it comes to laying out cold cash for products, he has a hard time persuading himself to spend what he sees as a significant price differential for an Xserve compared with an Opteron-based server with similar configuration and performance. However, Gilman also wonders whether the total cost of ownership might swing in favor of the Xserve. As he pointed out, his experience with Xserve systems while at the Whitehead Institute showed that they are well designed, reliable, easy to set up and administer, and particularly easy to administer from an application install and uninstall perspective. His experience is that administering systems based on Linux running on Opteron or Xeon is more labor intensive.

As he pointed out, his experience with Xserve systems while at the Whitehead Institute showed that they are well designed, reliable, easy to set up and administer, and particularly easy to administer from an application install and uninstall perspective.

As he mused more generally about strengths and challenges for Apple, Gilman commented that as a business owner, there are some key business applications for which software vendors do not fully support the Mac platform. Combined with general Microsoft dominance in the business IT environment, this lack of application support makes a full commitment to a Mac environment difficult for a commercial organization.

On the other hand, he sees a clear upsurge in the adoption of iBook and PowerBook laptop computers by scientists at the conferences that he usually attends. He believes that this adoption is driven by the stability and flexibility provided by Mac OS X, the performance on technical applications through the use of optimized math libraries and Velocity Engine vector processor, and the beautiful screens on the Apple laptops, which combined with OpenGL, really make the most of visualization, graphing, and presentation programs.

Case Study: Eric Neumann, Ph.D., Global Head of Knowledge Management, Sanofi-Aventis Pharmaceuticals

Eric Neumann, Ph.D., has traversed an eclectic path to where he sits today as global head of knowledge management at Sanofi-Aventis, one of the largest pharmaceutical companies in the world. He started professional life as a Ph.D. in pharmacology, neurobiology, and developmental genetics. Over the years, he has acquired skills and interest in bioinformatics and how information sharing in various forms can drive the pharmaceutical discovery and development process. His background includes stints at software provider Netgenics and working as the custom software developer and integrator for 3rd Millennium. He has also been cofounder and scientific advisor at Genstruct, and most recently, VP of bioinformatics and knowledge research at Beyond Genomics. Thus, his perspective includes that of the scientist, the software developer and vendor, the biotech executive, and now that of the large pharma.

Given his recent transition to Sanofi-Aventis, the discussion with Neumann was focused broadly on where he sees Apple products fitting into the life science landscape and not on the particulars of his current situation. From his vantage point, Neumann observes that he sees greater openness to the Mac platform within biotech than in pharmaceutical companies. The large pharmaceutical companies, like so many large enterprises, focus strongly on standardization of IT platforms and have largely excluded the Mac from support and consideration. Half jokingly, he commented that just about the time IT managers were saying that they "finally got rid of their last Macs," he tells them they should reconsider because Mac OS X is based on Unix and has become a good network player.

From his vantage point, Neumann observes that he sees greater openness to the Mac platform within biotech than in pharmaceutical companies.

Neumann admits to a long-time bias in favor of the Mac platform as a client computer. He favors it for the Mac's ease of support, intelligent user interface, and enhanced control of the computing environment. Despite the resistance within pharmaceutical companies, he comments that he is seeing more scientists choosing iBook or PowerBook laptops as their mobile computing option. He also observed that, at this point in time, Macs seem to have more favor among bioinformatics specialists than among the cheminformatics specialists.

Despite the resistance within pharmaceutical companies, he comments that he is seeing more scientists choosing iBook or PowerBook laptops as their mobile computing option.

When asked about the areas on which he is currently focused, Neumann articulated a passion for the development of the Semantic Web within the life sciences. He is actively involved in the W3C Technology and Society Domain — Workshop on Semantic Web for Life Sciences. The Semantic Web approach to data and knowledge integration within the life sciences is both more comprehensive and more flexible than the average data integration project tackled by a biotech or pharmaceutical company. Although it is very ambitious in scope, the Semantic Web holds out promise of great improvements in finding, combining, and mining data from many sources in ways that make sense to humans. Although it is still early in its development, Neumann believes that the Semantic Web could one day eclipse the Internet in importance.

Neumann indicated that Apple should seriously examine the opportunity to get involved with Semantic Web initiatives. He says that there is a significant need for development of smart viewers/browsers and the behind-the-scenes processing that can help the user effectively interface with data and knowledge in a Semantic Web format, since most scientists do not want to become bioinformatics specialists. They want tools that let them work with concepts and data intuitively.

Neumann wonders what could be the knowledge management analog to the introduction of "what you see is what you get" (WYSIWYG) in word processing. Possibly, Apple's proven track record in user interfaces and the visualization of data and software engineering could be combined with domain expertise from partners to create a significant opportunity for Apple with the Semantic Web more generally and the life sciences' Semantic Web in particular.

Possibly, Apple's proven track record in user interfaces and the visualization of data and software engineering could be combined with domain expertise from partners to create a significant opportunity for Apple with the Semantic Web more generally and the life sciences' Semantic Web in particular.

Case Study: Luis Aguilar, MIS Manager, TransForm Pharmaceuticals

Luis Aguilar has spent approximately 20 years in information technology management. In his current position as MIS Manager at TransForm Pharmaceuticals, he manages the support staff that provides client computing, servers, network services, communications, and IT security to the company of approximately 100 employees. TransForm Pharmaceuticals is a spinout company from Millennium Pharmaceutical and is based in Lexington, Massachusetts. TransForm specializes in the development of better drugs through the optimization of form and formulation. It combines high-throughput experimental platforms with a strong informatics platform to apply cocrystallization technology to the improvement of previously approved drugs as well as new chemical entities.

Aguilar described TransForm's IT infrastructure in the following way. On the business side of the organization, the client computers are mostly standard PCs running MS Windows XP. Business support applications also run on Windows servers. Recently, the company upgraded to Microsoft Exchange Server 2003 for email. Network security is handled by Linux servers. Oracle databases are the foundation of TransForm's informatics applications, which support scientific research and development in the organization. The Oracle databases and closely associated applications run on Sun servers running Solaris.

Mac products have started making some inroads at TransForm although they have not been a major component of the IT infrastructure in the past. Aguilar explained that currently about 3% of client computers are Macs. Interestingly enough, Aguilar, the MIS manager; Dr. Christopher Moor, the group leader for data analysis and modeling; and Dr. Juan Alvarez, recently hired as senior director of informatics, are all Mac users. In the past, adoption of Macs on the desktop was limited because those employees with Macs also needed to maintain a Windows machine for access to the full communications and calendar functionality of the Exchange Server. However, now that TransForm has upgraded to Exchange 2003, the Entourage product allows Macs to fully participate in the features of Exchange. Aguilar is already seeing signs of greater adoption of Macs among new hires who are given the freedom to choose a client platform.

Now that TransForm has upgraded to Exchange 2003, the Entourage product allows Macs to fully participate in the features of Exchange. Aguilar is already seeing signs of greater adoption of Macs among new hires who are given the freedom to choose a client platform.

Aguilar uses a Power Mac G5 as his main desktop for system administration on the network. In particular, he mentions that X11 windows allows him to interface with and run applications on the Sun servers from his own desktop. The ability of the Mac to integrate with NFS, SMB, and Apple's own AFP file systems gives him great flexibility as an administrator. He also uses his Mac as a development environment for tools and applications to run on the Sun machines.

The ability of the Mac to integrate with NFS, SMB, and Apple's own AFP file systems gives him great flexibility as an administrator. He also uses his Mac as a development environment for tools and applications to run on the Sun machines.

In addition to client computers, TransForm Pharmaceuticals has begun to use Xserve products. Today, it has two Xserve RAID storage systems. Each unit has 1.1TB of data storage capacity in a RAID 5 configuration. These storage systems are used as a first-level archive, a backup to disk for image and image analysis files. Crystallization images are captured in experimental protocols and are initially sent to and stored on a Sun server running Solaris. Once a file is no longer being used for active analysis, it is migrated to the Xserve RAID systems. This keeps the files easily accessible without tying up the main analytic server. For true archive security and business continuity purposes, TransForm still does a weekly full backup to tape that is stored offsite.

Aguilar commented that although the company initially had reservations about the performance characteristics of ATA drives used in the Xserve RAID, they have been pleased with the results. The system easily meets the requirements for a disk-based archive and does so at a price that is very compelling for the capacity and performance provided. TransForm now has plans to add another Xserve RAID with the new larger hard drives, which will give the company an additional 5.6TB of data storage.

Although the company initially had reservations about the performance characteristics of ATA drives used in the Xserve RAID, they have been pleased with the results.

In general, Aguilar and his colleagues are happy with the performance and stability of their Mac systems. In working to solve recent problems with SAMBA server synchronization, Aguilar indicated that he was impressed with the responsiveness of Apple and its willingness to give him direct access to key engineers at Apple.

One other feature that Aguilar indicated he would like to see on the Xserve RAID products is to make the RAID controllers completely redundant, providing failover support.

In the coming months, TransForm expects to increase its use of Mac products. It will likely add the Xsan storage area networking software to layer on top of the Xserve RAID storage systems. The company has also been experimenting with the Xgrid software and will be adding a small Xserve cluster to its environment and assigning a couple of developers to begin developing and testing parallel code on the small cluster.

Case Study Themes

Strengths and Benefits

When considering the qualities that lead scientists to use Macs and Xserve products, we see a number of common themes that articulate the strengths and benefits from a user or administrator perspective, including:

- **Dual personality of Mac OS X.** The dual personality of Mac OS X enables consolidation of the number of computers a scientist needs, providing a productivity and presentation environment in concert with a command-line Unix environment on the same machine. When installed on an iBook or PowerMac, it creates a portable Unix machine with robust power management and usability.

- **Open source applications.** The availability of hundreds of open source applications written for Linux or Unix as well as the ability to run legacy custom code developed for those platforms are key benefits identified by many participants.

- **Ease of administration.** A number of participants strongly emphasized ease of administration, pointing out the ease of application installation and uninstallation. Managing multiuser access to shared resources such as a cluster was also perceived as easy. And a couple of participants mentioned appreciation for graphical menu versions of many command-line tools and utilities. Several participants also mentioned that they experienced fewer problems with adware, viruses, worms, and malware, which contributes significantly to lowering the hassle and cost of supporting Mac users.

- **Ease of use.** Another common theme was ease of use for the individual scientist. For long-time Mac users, there are many small and sometimes subtle features that make a Mac easy to use, ranging from the interactivity between mouse and screen; to the look, feel, and consistency of user interfaces; to strong crisp multitasking. Another attractive feature supporting ease of use is Apple's partnership with the BioTeam to provide the iNquiry tool on the Bioinformatics cluster.

- **Solid and versatile software development platform.** A number of development tools, both cross-platform tools and Mac-specific tools, come preloaded with Mac OS X. Case study participants have used their Macs to develop code for Unix, Linux, and Mac-native and Windows platforms.

- **Design.** Several participants mentioned an overall coolness factor to owning well-designed Mac products. People appreciate the attention to details, the design aesthetics, and the human-computer interface.

- **Good performance.** A number of participants mentioned that the G5 gives great performance, ranging from a crisp feel (even with 15–20 applications open) to the speedup of legacy Unix or Linux applications. In particular, code with a high dependence on matrix mathematics can benefit from the optimized math libraries that take advantage of the Velocity Engine vector processing unit.

- **Price competitiveness.** Our case study participants varied from believing that Mac computers are premium priced to believing that they are price competitive. At the time of writing this white paper, the authors went to the Apple Web store to price a Power Mac G5 desktop with monitor and to other online stores for similar systems based on the AMD Opteron chip. Matching configurations as closely as possible on memory, hard drives, graphics cards, and monitors, we found that in one case, the Apple system was about 10% more expensive, but in another case, it was about 15% less expensive than comparable systems. In addition, at least a couple of our case study participants expressed an opinion that Mac systems required less support and were therefore less expensive over the lifetime of the system. Those participants with previous experience with Linux clusters also expressed an opinion that an Xserve-based cluster is considerably less expensive to administer than a Linux cluster.

Challenges

Although they were generally quite positive about their use of Macs and Xserve products, our life science case study participants also identified a number of challenges that Apple faces, including:

- **Integration with centralized IT services.** As the minority computing platform in almost any larger enterprise, Apple faces the issue of truly seamless integration in a Microsoft-centric world. The problems tend to emerge especially in integration issues with both Microsoft Exchange Server and Active Directory as well as with the IBM Lotus Notes email application. As our participants indicated, there are solutions, but they often place a greater responsibility on the end user.

- **IT support.** Some of the people we contacted about potentially providing a case study indicated that they would like to be a Mac user at work, but "IT will not provide support." Several of those interviewed indicated that they largely provided their own support. Many users tend to discount this as a barrier, claiming that a Mac and even an Xserve cluster are fairly easy to administer without much formal support from a central IT department.

- **Support from software vendors.** Some participants pointed out that because of the smaller market share of the Mac, independent software vendors (ISVs) are not always as quick to bring out a Mac update of business software, and even some life science software vendors have had uneven focus on the Mac platform over the years.

Vendor/Community Relationship

With regard to the relationship of Apple with the life science community, case study participants were encouraged that Apple seems to be increasing its commitment to becoming visible at key conferences. A couple of interviewees commented that Apple would benefit from an active direct sales force focused exclusively on the science market that understood scientists' needs and could answer questions that were scientifically technical. And, for academic scientists, an Apple resource that was available to help write the IT portions of academic grant proposals would be a great addition.

FUTURE OUTLOOK

Apple's Life Science Market Strategy and Focus

Apple has a team of in-house specialists to understand and serve the life science market. This group includes scientists and dedicated marketing personnel. In terms of getting the message out to the market, Apple is conducting targeted advertising in life science trade magazines and life science Web sites as well as attending and exhibiting at many scientific conferences.

Apple's strategy in the life sciences is to focus on what the customers want and what they need to create innovation. The top 3 areas on which Apple focuses in the life sciences are:

- **Time to results.** Time to results is affected by the speed of the processor, the overall architecture, and the amount of time spent working on science versus working on system maintenance and management. Across the product line, the Mac platform provides a combination of power, speed, and ease of use to support an efficient and productive workflow.

- **Freedom to be creative.** From the graphics capabilities of the system to the built-in developer environment, the Mac platform is built to inspire and support the creativity that drives success in life science research while eliminating common technical barriers prevalent on other platforms.

- **Financially sound solutions.** Apple focuses on providing products and solutions that are cost competitive not just in terms of purchase price but also in terms of total cost of ownership. The company is also focused on ensuring that customers get the best performance and value for their budget dollars.

New Product Releases

The most relevant to this market of the upcoming product releases is the next update to Mac OS X (Tiger). The features that will be of particular interest to life science users include:

- **Full 64-bit memory addressing.** Currently, each thread has a 4GB memory limit. In Tiger, a single application can address up to 16 exabytes of memory.

- **Additional math libraries.** Additional math libraries will be provided that are fully optimized to take advantage of the G5 architecture.

- **Automator.** Automator is a scripting feature for easily automating tedious or repetitive tasks. This type of workflow automation should be particularly helpful in areas such as image analysis.

- **Xgrid.** Xgrid enables users to easily assemble a network of Macintosh systems into a computational grid, allowing scientists to make better use of their computational resources.

- **Spotlight.** As a new search technology included in the next release of the operating system, Spotlight will allow users to search through the entire content of their systems, including inside PDF, Word, and other documents, in a few short seconds.

MARKET CHALLENGES/OPPORTUNITIES

Market Challenges

As the case studies have demonstrated, Apple products are a good fit for life science researchers. However, there are a number of challenges. One of these challenges is positioning. Many life scientists and IT managers who support them, classify Apple as a consumer oriented company and do not think of Apple in a commercial or enterprise setting. A second market challenge articulated in the case study themes was the resistance of IT departments to consider supporting Mac client computers. In addition to these customer facing market challenges, Apple faces a variety of competitive challenges.

A second market challenge articulated in the case study themes was the resistance of IT departments to consider supporting Mac client computers.

Vertical Market Orientation

In addition to Apple, many other competitors have also targeted the life science market. Competitors in the life science market include major IT vendors such as IBM, Sun, and HP, with broad server, cluster, and workstation offerings; strong professional service organizations; and moderate to large internal groups focused on the life sciences. On the storage side, Apple competes against established storage vendors such as EMC and Network Appliance, both of whom have good presence in the life science market.

Cluster Competition

Both Sun and IBM have also introduced prepackaged small cluster systems for bioinformatics workgroups. And, although Apple appears to have the edge for now in creating an easy-to-administer cluster, many similar tools for easing the configuration and administration of cluster systems are becoming available for Linux-based clusters. When considering the cluster competitive situation, Dell must be mentioned. Although Dell has not had a vertical market focus on the life sciences, it has a solid presence in Linux clusters within this market.

Both Sun and IBM have also introduced prepackaged small cluster systems for bioinformatics workgroups.

In addition to cluster offerings from the major IT vendors, Apple must also compete against cluster-focused vendors such as Linux Networx and emerging vendors such as Orion Multisystems, with its cluster workstation concept.

Blade servers can also easily be configured for use as an HPC cluster. For example, IBM is using its blade platform to build a very large high-performance cluster for the Spanish government, and specialty vendors such as Appro target the small cluster market with blade-based HPC systems.

Client Computing Competition

Microsoft Windows–based systems also dominate client computing in the life sciences, although not to the same extent as in other markets. This is particularly true in large pharmaceutical companies. IT departments in these large companies prefer to standardize, and Microsoft Windows is the prevailing choice for that standardization on the desktop. The drive to standardize and the Microsoft dominance of the desktop combine to suppress the demand for Macs.

IT departments in these large companies prefer to standardize, and Microsoft Windows is the prevailing choice.

If we go to the other end of the spectrum and examine the challenges against Unix and Linux workstations, Apple is better positioned to compete. The combination of OpenGL, Mac OS X, and the 64-bit G5 architecture means that a fully loaded Mac is a legitimate replacement option for many users accustomed to a Unix workstation. And, compared with a Linux workstation, the Mac provides the well-designed Mac user interface, great graphics, and native access to Microsoft Office.

The combination of OpenGL, Mac OS X, and the 64-bit G5 architecture means that a fully loaded Mac is a legitimate replacement option for many users accustomed to a Unix workstation.

Apple is well positioned relative to the competition when it comes to portable computing for the life science researcher. In particular, those scientists who also develop code love the fact that they have a robust Unix development and testing environment sitting on their laptop.

Applications

As a couple of case study interviewees pointed out, one of the primary challenges for Apple is to maintain and expand relationships with commercial software vendors that serve the life science market. In a niche market such as the life sciences, these relationships with software vendors must be pursued, supported, and marketed so that each partner receives maximum value.

Market Opportunities

Life Science Insights believes that a number of factors are cooperating to assist Apple in strengthening its position in the life science market. These factors include:

- The ability to convert Unix workstation users to Mac desktops
- Apple's current strength in laptops among scientists as demonstrated by the prevalence of PowerBooks and iBooks at life science conferences
- The continued strength of the open source application base in the life sciences combined with relatively easy porting from Unix and Linux to the Mac

- The attention to ease of administration and of use, which appears to be resonating well in the small to medium-sized cluster market and is wooing some previous Linux cluster owners to the Apple camp when they add, upgrade, or replace existing clusters
- Market competitive pricing, which is causing buyers to take a new look at Apple
- Consolidation of a Unix or Linux workstation and a personal computing system into a single Mac system, which yields cost savings, reduces IT hassles, enables seamless data flow from experiment to presentation, and allows researchers to reclaim overcrowded desk space
- Steadily growing number of third-party scientific developers on Mac OS X, which is driving an increase in the number of commercial applications native to Mac OS X

If Apple continues to invest in providing support staff for the life sciences and creating greater visibility in the community, Life Science Insights believes that Apple can capitalize on this current growth and achieve a significantly greater share of market in client computers, clusters, and storage.

Life Science Insights believes that Apple can capitalize on this current growth and achieve a significantly greater share of market in client computers, clusters, and storage.

Apple's energies should be directed at educating IT support staff on how to effectively integrate and support Mac and Xserve products in a heterogeneous IT environment. As we looked for industry participants for these case studies, we encountered some latent demand for Mac products among scientists who indicated they would use Mac computers if the IT department would support them. Additionally, we have some evidence that when policies are more liberal, as in the case of Isis Pharmaceuticals, the percentage of scientists using a Mac goes from near 0% to about 10%. Or, in cases such as the NIAID, where scientists have a long tradition of freedom of choice on a client computer, Mac usage is up around 50% among scientists. These stories speak to the market opportunities for Apple in the life sciences if it can overcome the IT department barriers.

CONCLUSION

Apple has long served the needs of researchers in the life sciences. Its current product line fits well with the needs of scientists conducting more computationally intensive tasks. The combination of Mac OS X with "Unix under the hood," the PowerPC G5 64-bit processor, Apple's attention to ease of use and administration, the lower susceptibility to computer viruses, and Apple's focus on the life science market are all finding favor with traditional Mac users. These same factors are also persuading other researchers that Apple will play an important role in supporting and enabling life science research.