FINAL REPORT

January 1, 2001   December 31, 2004

CURRICULUM IMPROVEMENT
PARTNERSHIP AWARD PROGRAM

John Jay College of Criminal Justice/The City University of New York

Name of Institution

Douglas E. Salane

1.0 Name of Institution

John Jay College of Criminal Justice
The City University of New York

2.0 Name of Principal Investigator

Douglas E. Salane

3.0 Name of CIPA Project

Computer Clusters to Support Curricular Improvements in Networking and Parallel /Distributed Computing.

4.0 Project Activities

Introduction The NASA CIPA Cluster Computing Project has had a profound impact on the Department of Mathematics and Computer Science at John Jay College. The Department and College are grateful to NASA and the UNCFSP Corporation for the opportunities provided by the NASA CIPA program. This final report tells how John Jay College has taken advantage of NASA computing technologies, particularly cluster computing, to revitalize its computing curriculum, create new programs and research capabilities, and give students opportunities to learn state-of-the-art computing methods and technologies that are changing the practice of information technology.

The goal of the project was to give Computer Information System (CIS) majors at John Jay College access to the latest methods and practices in computer networking and parallel/distributed computing. These areas of computing are vital to NASA initiatives in high performance computing and essential for building modern information systems. Strong backgrounds in computer networking and parallel/distributed computing will allow John Jay CIS graduates to assume leadership positions in the information technology (IT) field, enter some of the most important and exciting graduate and professional programs, and advance the practice of computing.

As indicated in the original proposal, achieving the aforementioned goal would require significant facilities, curriculum and faculty development. By the end of year two of the project, most of the proposed computing facilities, i.e., the cluster computing facilities, supporting laboratory and network infrastructure, were in place. The PI, Co-PIs and staff member had developed significant expertise in cluster computing, enterprise-level database systems, and parallel/distributed computing. The Cluster Computing Project facilities and faculty expertise had made possible new courses and upgraded curricula in computer networking, database systems and parallel/distributed computing. During the first two years of the project over 450 students enrolled in CIS courses that took advantage of NASA CIPA computing facilities and curricula.

In the third year of the project, the Department of Mathematics and Computer Science began to realize other significant benefits. New expertise in distributed computing, computer networking and operating systems led to the development of an M.S. degree in Forensic Computing. The program began accepting students in Fall 2004. NASA CIPA cluster computing and database facilities made possible new research initiatives in high performance computation and criminal justice database systems. The PI and a former NASA/CIPA student presented some of the project's cluster computing research and educational activities in a talk at the entitled "Computer Clusters in a Computer Major in a College of

Criminal Justice." The October 2004 USENIX journal Login provided a summary article that featured this well received talk. (The article appears as Appendix VI.)

The John Jay NASA CIPA program was slated to end in 2003; however, a no-cost extension allowed the project to make use of $4106 of funds that remained at the end of 2003. These remaining funds allowed two additional students to serve as tutors in 2004. In addition, they allowed at least 50 undergraduate students in the networking and operating systems classes to develop excellent skills in network and multithreaded programming. The remaining funds also allowed us to present the project activities at the USENIX Conference in June. In addition, the funds helped us maintain NASA CIPA computing equipment and provide enhanced web services to support student-computing activities throughout the year.

During the past three years the PI obtained $97,540 of additional funding from non-NASA sources to purchase equipment and support students and staff working on the NASA CIPA Cluster Computing Project. With this additional funding, the project exceeded the facilities, curriculum and faculty development objectives stated in the original proposal. Some additional funding was due to partnerships that allowed NASA CIPA technologies to benefit other groups within the college (e.g., the Science Department Partnership). Partnerships and external funding obtained for the NASA CIPA program are discussed in more detail later in the report.

In the 2002 review of the project at the NASA CIPA reverse site visit, the reviewers noted that the computing infrastructure and institutional expertise developed as part of the Cluster Computing Project created numerous opportunities for research and scholarly activities. In 2004, the PI received two NSF grants. In one grant, the PI, a former NASA CIPA staff member, and a former NASA CIPA student (who is now enrolled in the M.S. Program in Forensic Computing) will work with researchers at Polytechnic University to develop a network forensic tool. In the second grant, the PI and colleagues at John Jay College will develop curricula in information assurance and network security. Without the expertise and computational facilities developed under the NASA CIPA program, these efforts would not have been possible.

This final report describes NASA CIPA Cluster Computing Project activities and their impact on computer science at John Jay College. The report restates facilities, curriculum and faculty development objectives, describes what was achieved, and discusses the impact on the CIS program and students. The report cites specific examples of how the NASA CIPA Cluster Computing Project benefited the over 700 Computer Information Systems majors at John Jay College, of whom about 65% are Black or Hispanic and 50% female. The report describes partnerships developed with NASA, the Science Department at John Jay College, and the computer science departments at Polytechnic University and the Borough of Manhattan Community College. The report discusses dissemination activities, project initiated research efforts, and the impact of project activities on curricula both in the Department's undergraduate and graduate computer programs. The report also discusses the relevance of the John Jay Cluster Computing Project to NASA initiatives in high performance computing

The report is organized as follows. The remainder of the Project Activities section discusses facilities, curriculum and faculty development objectives. Next, the section Project Impact describes program benefits to the CIS Major and students, relevance to NASA Mission, partnerships and matching funds, and project evaluation results. The report then presents the following required final report sections: Dissemination, Project Future, Personnel Changes, and Expenditures. In the Additional Remarks section, the report concludes with some final comments on the NASA CIPA program.

Objectives

Facilities Development Objectives The major facilities objectives were to build a computational cluster, a database cluster, the Distributed Computing Laboratory, and the required supporting network infrastructure (A network diagram in Appendix I provides an overview of NASA CIPA Cluster Computing Project facilities.). In addition to building facilities, it was important to establish maintenance procedures and support services that would make the facilities readily available to students and faculty. All the facilities objectives have been achieved and excellent support services are in place. Each computing facility is now being used to support course work and student/faculty research projects in network security, high performance computing and large-scale database systems. Each academic year over 225 CIS majors and 50 Science majors use NASA CIPA project facilities. Table 1 presents a list of facilities objectives and indicates what was achieved.

Table1: Facilities Objectives

Project Objectives	Metric for Success	Actual Achievement
1. Build a production computational (Beowulf) cluster. Introduce students in core and capstone courses to cluster and parallel computing.	Completion of a production cluster; cluster used to support course work and increase student involvement in parallel computing.	A 12-node 24-processor computational cluster is now operational. Over 80 students in operating systems (MAT 375) and Computer Networking (MAT 379) used the cluster to understand send/receive semantics of the Message Passing Interface (MPI), symmetric multiprocessing and scheduling algorithms, and the distributed Bellman Ford algorithm.
2. Build a production database cluster. Introduce enterprise-level databases and web-based information systems into the advanced database course.	Completion of the database cluster; database cluster used to support course work and increase student interest in advanced database topics.	A three-node database cluster that includes a remote access computer and a web server for developing web-based information systems is operational. Each year over 40 students in the senior level database courses develop skills in enterprise-level database systems. The cluster has been used for 15 student projects that analyze data in the FBI's National Incident Based Reporting System (NIBRS) data. In 2004, matching funds allowed us to add a Raidzone Open NAS 1.2 TB disk system to the cluster. The cluster now makes available the complete set of FBI NIBRS data ('95-'01) in an Oracle 10G relational database.

Table1: Facilities Objectives (continued)

Project Objectives	Metric for Success	Actual Achievement
3. Build a computer lab that provides tools for distributed computing, network programming, and offers high-speed access to the computational and database clusters. Introduce CIS majors in core and capstone courses to a laboratory environment suitable for parallel programming, distributed computing and computer networking.	Completion of the Lab and installation of required software. Incorporation of lab resources into course lectures and course work. Increased student exposure to modern methods and tools for networking and parallel/distributed computing.	The Distributed Computing Laboratory provides 25 Linux workstations with software for concurrent programming, network monitoring and analysis tools, compilers and libraries for network programming, and high-speed access to the Computational and Database Clusters. The lab is now used each year by over 220 students in junior and senior level courses for projects in multithreaded programming, interprocess communication, message passing, computer networking and cluster computing. The lab is the primary lab for tutoring and classes in Operating Systems (MAT 375) and the Internship course (MAT 404).
4. Increase student access to tools for cluster, parallel and distributed computing by providing a remote login capability that allows students to access cluster computing resources from off campus.	Completion of the remote access facilities. Student use of remote facilities to access accounts, files, data and programming tools. Increased student skills in the use of tools for secure remote access.	Two remote login computers, identified as PUBSSH, were installed and the network infrastructure that permits off campus access was completed. Currently all CIS students have accounts on PUBSSH and can use computers in the Distributed Computing Lab. Each year over 220 students routinely access NASA CIPA computers from off campus using the program Secure Shell (SSH) program. The project web site contains an SSH tutorial developed by NASA CIPA students. The tutorial shows students how to access NASA CIPA facilities using SSH.
5. Provide a high-speed network infrastructure that interconnects all NASA/CIPA cluster computing resources. Increase student interest in secure network design and implementation.	Completion of the infrastructure and integration of infrastructure into the curriculum.	The infrastructure was completed in Spring 2002. Besides interconnecting all CIPA computing facilities, the network infrastructure each year provides a case study in secure network design for over 30 students in the courses Computer Networking (MAT 379) and Quantitative Methods in Criminal Justice (MAT 400).

Table1: Facilities Objectives (continued)

Project Objectives	Metric for Success	Actual Achievement
6. Ensure project computing resources are readily available to authorized students. Provide utilities for efficient, and where possible, automated maintenance and upgrading of Cluster Computing Project resources. Provide similar utilities for low cost maintenance of student accounts.	Implementation of facilities. Reduced downtime, increased student satisfaction and use. (Over 50 students in a preliminary survey indicated equipment problems were solved the same day.)	Utilities implemented include a) an automated system for generating student accounts on NASA/CIPA equipment using the Registrar's CIS course rosters, b) lab activity monitoring tools, c) back up systems for student directories, clusters, and web servers, d) automated installation utilities for labs and clusters e) Web-based reporting system.
7. Increase student interest and access to Cluster Computing facilities by providing lab tutoring and support.	Implementation of lab tutoring and support services in the Distributed Computing Lab.	Each semester tutoring is available in the Distributed Computing Lab Tues., Wed. and Thurs. from 12 to 5:30 p.m. Each week two tutors assist over 30 students in course project work and the use of cluster computing facilities.

Curriculum Development Objectives Table 2 presents the curriculum development objectives. Achieving curriculum objectives relied heavily on timely completion of computing facilities and development of faculty expertise. All curricula developed are MSET curricula. All courses developed under the project have been approved by either the College or Graduate Curriculum Committees and are now part of CIS program or the new M.S. program in Forensic Computing.

Table 2: Curriculum Development Objectives

Project Objective	Metric for Success	Actual Achievement
1. Develop courses and curricula that give students opportunities to learn and build expertise in parallel/distributed computing.	Increased student enrollment and participation; courses integrated into the CIS major.	Developed and implemented five new undergraduate courses: Artificial Intelligence (MAT 376), Graphics and Graphical User Interfaces (MAT 273), Computer Networking (MAT 379) and Discrete Mathematics (MAT 204). These courses enroll over 100 students each academic year (Please see Appendix II).
2. Revise the CIS major so students have the mathematics preparation for the study of computer networking and parallel/distributed computing.	Increase student enrollments in mathematics courses that provide essential background for the study of parallel/distributed computing.	The course Discrete Mathematics was developed and is now required of all CIS majors. In addition, all majors now must take two advanced Operations Research courses (MAT 323 & 324). These courses treat graph theoretic networks and queuing theory, which are areas of mathematics widely used in the study of parallel and distributed systems.

Table 2: Curriculum Development Objectives (continued)

Project Objective	Metric for Success	Actual Achievement
3. Revise intermediate level courses and the advanced database course to prepare students for work in distributed computing.	Increased student exposure to and interest in cluster computing, concurrent programming, and enterprise-level database systems.	Each year over 40 students take the revised the Operating Systems course (MAT 375), which now provides broader, deeper, hands-on coverage of interprocess communication, concurrent programming, message passing and distributed systems. Over 40 students take the revised advanced database course (MAT 470) that now includes coverage of enterprise-level database systems, data mining and web-based information systems.
4. Develop graduate courses in networking and distributed computing for a new graduate program in Forensic Computing .	Establish research level expertise in the field of security of distributed information systems; develop a program that trains forensic analysts who investigate the misuse of computer networks and distributed information systems.	We developed six new courses for a proposed M.S. degree program in Forensic Computing. Three of the courses, Network Security, Network Forensics and Architecture of Secure Operating Systems examine the security of computer networks and distributed systems in-depth. Based on the proposed courses, the Board of Regents of the City University of New York accepted the College's letter of intent for an M.S. program in 2002. The full proposal was submitted and the program was approved in February 2004. The program accepted its first class in fall 2004. One graduate course was offered in fall 2004. The other two will be offered next year. NSF funding has been obtained for research and educational projects in network forensics. Research projects are underway in cluster computing and distributed database systems.

The John Jay Cluster Computing Project has achieved and exceeded its three-year curriculum development objectives. The curricular and program revisions needed to prepare students for parallel and distributed computing topics in upper level and graduate courses are in place. Thanks to the participation of four additional faculty members, we developed five new courses that cover topics in computer networking and parallel/distributed computing. The courses have been approved by the College and are now part of the CIS major. The Department began offering four of the courses in 2002. A list of all revised and newly developed Cluster Computing Project courses and their typical academic year enrollments appears in Appendix II. This appendix also shows that each year over 300 students enroll in CIS major courses either developed or revised under the auspices of the NASA CIPA Cluster Computing Project.

Curriculum development work on intermediate and advanced courses is complete. Prof. Mantharam developed a senior level undergraduate course in parallel computing, which the Department will offer as a regular course in the spring semester. The course makes extensive use of the Computational Cluster. Prof. Shenkin revised the intermediate and advanced database courses (MAT 265 and 470). Since 2002, the advanced database course has taken advantage of the distributed database cluster for student projects involving the FBI's NIBRS database. Materials on data mining, developed in partnership with Prof. Bon Sy of Queens College, are now a part of the curriculum in the upper level database course and the new graduate course Network Forensics. Curriculum revisions, which reflect the availability of the Distributed Computing Laboratory, Computational Cluster, and NASA CIPA Network Infrastructure, have been in place in the courses Operating Systems and Computer Networking since 2002. A sample of course projects that build student skills in network, parallel and multithreaded/multiprocess computing appear on the PIs Operating Systems and Computer Networking course home pages (http://web.math.jjay.cuny.edu/index2.html). In addition, these courses now employ the Computational Cluster to provide hands-on coverage of message passing, the distributed Bellman-Ford algorithm, cluster scheduling, and symmetric multiprocessing. Prior to the NASA CIPA project students had no laboratory facilities for any of the aforementioned course projects.

NASA CIPA facilities now provide numerous opportunities to improve curricula in courses that involve networking and parallel/distributed computing. The graduate courses Network Security, Network Forensics and Architecture of Secure Operating Systems will take advantage of the Distributed Computing Laboratory and the Computational Cluster. The Database Cluster and Computational Cluster will make it possible for students to analyze large forensic data sets using computationally intensive data mining tools such as latent semantic analysis. A new partnership with Polytechnic University, which is funded by an NSF award, will take advantage of the Distributed Computing Laboratory and the Computational Cluster to provide additional hands-on exercises in network security and information assurance. NASA CIPA Cluster Computing facilities at John Jay College offer a state-of-the-art laboratory environment for a wide range of topics in modern computing.

Faculty & Staff Development Objectives The NASA/CIPA Cluster Computing Project has provided the resources needed for faculty and staff members to stay abreast of developments in the rapidly changing fields of computer networking and parallel/distributed computing. The project has provided funds for conference attendance, a seminar series, tutorials and faculty summer study. Table 3 provides a list of the major faculty/staff development objectives.

Table 3 Faculty and Staff Development Objectives

Project Objective	Metric for Success	Actual Achievement
1. Develop the staff expertise needed to build and maintain a computational cluster that provides the software environment for parallel programming and solving large-scale problems.	Attendance at conferences and tutorials for cluster and Linux systems management, ability to implement and manage cluster computing facilities**.**	The staff member and PI attended tutorials at three Linux Expo Conferences, two Large Installation Systems Administration Conferences, and the NASA Information Power Grid Conference. The John Jay Computational Cluster now offers the middle ware and systems software needed for parallel programming and cluster management (e.g., Ganglia, MPICH and NASA's PBS).
2. Develop faculty expertise in cluster computing, message passing interface, parallel file systems, and numerical software for solving large scale-problems in a cluster computing environment.	Attendance at conferences and tutorials, completion of self-study programs, impact on students and the curriculum	The PI participated in the IEEE 2001 and 2002 Cluster Computing Conferences, the 2003 USENIX Security Conference, and 2004 USENIX Technical Conference. He attended four tutorials by leading MPI experts. He and a Co-PI, Prof. Mantharam, completed self-studies of the following: Message Passing Interface, Open MP and ScaLapack. The PI oversaw the following 4 student cluster computing research projects: ScaLapack Benchmarking, Large Number Library, High Performance Linpack Benchmarking, and Distributed File Systems Evaluation
3. Develop faculty expertise in parallel programming and visualization of parallel algorithms.	Completion of self-study course and the development of a program for the visualization of a parallel Schur Decomposition.	Prof. Mantharam and the PI completed a self-study of MPE. Prof. Mantharm developed a visualization tool for parallel programs and presented her results at a professional conference.

Project Objective	Metric for Success	Actual Achievement
4. Develop faculty expertise in distributed database systems, enterprise level databases and web-based information systems	Completion of on-line self-study course in Microsoft SQL and Oracle 9i. Implementation of the distributed database cluster housing enterprise level database systems and facilities for builing web-based information systems. Completion of student research projects involving enterprise-level database systems. Development and implementation of an enterprise-level database of interest to the John Jay and criminal justice communities. Internet access to the database through a web portal.	The lead staff member and Raul Cabrera, a NASA CIPA student, implemented the Distributed Database Cluster. Prof. Shenkin completed on-line self-studies in both the Oracle 9i database server and the Microsoft SQL server. Prof Shenkin and two NASA CIPA students implemented the FBI's NBIRS database as a Microsoft SQL relational database. He and a student also designed and implemented a web-based front end for this database. Nine students used the NIBRS database for research projects in the advanced database course MAT 470. The lead staff member, Prof. Shenkin, and the PI ported NIBRS to an Oracle 10g relational database. This database resides on a 1.2 TB Raidzone OPEN NAS system This large-scale storage system was installed in 2004 to support the database cluster. The Oracle database contains the all FBI NIBRS data (years 1995 to 2001). A web portal for the Oracle NIBRS system is under development.
5. Establish a Cluster Computing Seminar Series that allows students, faculty and staff members to stay ahead of research and technical developments in all areas of computing relevant to the project.	Number of seminars held, participation by students and faculty members. Interest among computer majors and John Jay Community.	The Cluster Computing Series sponsored over 20 outside speakers who presented talks on cluster computing, computer networking, and parallel computing. (Please see Appendix III) Each seminar has attracted over 20 participants. Three have attracted over 70. The seminar series in now sponsored jointly with the Science Department. The student computer club offers regular workshops as part of the program.

5.0 Project Impact

Impact on CIS Program and Students This section describes the project's benefits to the College and NASA. It gives examples of how new institutional capabilities (i.e., computing facilities and faculty/staff expertise) in networking and parallel/distributed computing have benefited computer majors, strengthened the CIS program, and made possible the development of an M.S. degree program in computing. This section also describes new research initiatives and academic/research partnerships made possible by the NASA CIPA project. Finally, this section discusses the projects relevance to NASA and the high performance computing community. The section begins with a discussion of the impact of the computing facilities.

The Distributed Computing Lab, with 25 Linux workstations, now offers CIS majors the computing environment that is needed for advanced course work in the systems areas of computing. Already most of our upper level CIS students use the lab. For example, the senior level internship course (MAT 404), which typically enrolls about 25 students per semester, holds all class meeting in the lab. The Operating Systems courses (MAT 375), which enroll about 45 students per semester, meet in the lab regularly. For both courses, this is the primary computing laboratory for project assignments. Two NASA CIPA students provide tutoring in the lab Monday through Thursday. They provide both individual assistance and instruction to small groups. Under supervision of the tutors, students have the opportunity to learn about and experiment with a wide range of high quality, useful Open Source Software that is not available elsewhere on campus. The tutors also assist in the use of facilities, computer accounts, course project assignments, and provide mentoring for new majors.

The Distributed Computing Laboratory is proving to be a good environment for the study of topics in interprocess communication (IPC). Students now write multiprocess and multithreaded applications early in the operating systems course and develop strong skills in concurrent programming. For example, for the past two semesters the operating systems classes have had to implement a solution to the classic Dining Philosophers problem. Students use the G2 Graphics system available on lab computers to demonstrate that their programs are performing to specifications. The exercise helps the class understand an important resource allocation problem in concurrent computing. Moreover, students gain skills developing multithreaded programs that employ the POSIX Threads application programming interface, a widely used C library for concurrent programming. Prior to the NASA CIPA program, students had no access to such a programming environment, and no opportunity to gain this type of hands-on experience.

The Database Cluster was another important facilities objective of the project. The Cluster supports two enterprise-level databases that are used for both course work and research projects. Both the Microsoft SQL and Oracle 10G database systems are available. The Cluster also provides a web server so students can learn how to develop information systems that employ a web front end. Today such systems are the norm. Professor Shenkin, a CO-PI, routinely uses the Database Cluster in the senior level database course (MAT 470) to train students to use data mining techniques on enterprise-level database systems. Prior to the Database Cluster, instructors covered topics in data mining and enterprise-level databases in lectures, but no laboratory environment was available. Students both in courses and in independent research projects now learn how to use the scripting language PHP to make information in a database available in user friendly web pages.

The Database Cluster supports faculty/student research projects. In 2002, Prof. Shenkin and Atiqul Mondal, a recent CIS graduate, developed a web-based information system that provides easy access to the FBI's National Incident Based Reporting System (NIBRS). NIBRS, as released by the FBI through the University of Michigan, is a flat data file that contains a list of crime incidents reported to the

FBI by local law enforcement agencies. Law enforcement specialists and criminal justice researchers, who are not necessarily database specialists or computer scientists, can use the web-based information system to query NIBRS. Due to the size of the NIBRS files and processing requirements, the system required the capabilities of an enterprise-level database system housed in the Database Cluster. Mr. Mondal's report entitled "Web-Based Information System for the FBI's National Incident-Based Reporting System Data" describes the system. This work was presented at a Cluster Computing Seminar colloquium and the Project Ascend McNair conference at The City University of New York (CUNY) graduate center.

The Database Cluster was the focus of much research activity in 2004. Mr. Boris Bonderenko, the project lead staff member, and Raul Cabrera, a graduate student, developed a new Oracle 10G implementation of NIBRS. The Oracle 10G version provides data mining and parallel processing capabilities not available in previous implementations. During 2004, we added a 1.2 TB NAS Open Storage system to the Cluster. With this additional storage capability, Boris and Raul were able to add the entire NIBRS data collection (years 1995 to 2001) to the Oracle 10G database. Faculty members at the College who are involved in criminal justice research now use relational database techniques to explore NIBRS data. In addition, with the new Oracle 10G implementation a user can construct queries over all 13 segments of NIBRS data. Previously, this capability was not available to the criminal justice research community.

The John Jay Computational Cluster provides a laboratory environment for high performance computing and research in cluster computing. The Cluster has the latest version of the Message Passing Interface (MPI) from Argonne National Laboratories. (MPI is an application-programming interface that is used to develop parallel programs.) The Cluster offers the ScaLAPACK software library for solving large-scale numerical linear algebra problems in a distributed memory computing environment. (ScaLAPACK is used to solve linear equations, least squares problems, and eigenvalue problems.) The Cluster has state-of-the-art monitoring and managemement software, e.g., the NASA developed Portable Batch System for job scheduling, and Ganglia, a widely used tool for monitoring cluster nodes and performance. Indeed, thanks to the NASA CIPA project the PI and his students were able to investigate various cluster architectures and determine one that provides an inexpensive, reliable high-performing cluster. Using available software such as ScaLAPACK, the PI and his students developed several methods to ensure that the cluster is performing accurate computations. The PI presented this work at the USENIX Annual Technical Conference in July 2004.

Given the faculty's interest in fire science and forensic computing, the Computational Cluster is becoming an important research tool at College. For example, the National Institute of Standards and Technology (NIST) has released a computer code for modeling fire propagation in buildings. Working with a faculty member who is a fire science specialist, the PI and his students are porting this code to the computational cluster. NIST's code will be used to improve fire safety in buildings and model the fires in the World Trade Center. The forensic analysis of network traffic requires high-performance computing to store and analyze very large sets of data. Faculty members are using techniques similar to those used in search engines to find evidence in extensive logs of network traffic. This work requires the computational capabilities of the cluster to perform repeated singular value decompositions of large matrices.

The Computational Cluster also has allowed students at the College to get involved in high performance computing research. For example, while a student Boris Bonderenko used the computational cluster to develop a software package for arithmetic operations on large numbers in a distributed memory computing environment. Using his package and the Message Passing Interface, he implemented a parallel version of the Miller-Rabin algorithm for primality testing. Boris prepared a report entitled "A Large Number Library for Computational Clusters" and presented it at the Ascend McNair Conference at the

CUNY graduate center in September. Another NASA CIPA student, Raul Cabrera, helped fine tune the High Peformance Linpack Benchmark for the Computational Cluster. Working with the PI and Mr. Bonderenko, he investigated the performance of NASA's NAS Parallel Benchmarks on the cluster. He also compared several versions of the Basic Linear Algebra Subroutines and found that the KGOTO version offered the best performance because it minimized processor TLB misses. The PI presented these results at the 2004 USENIX Technical Conference. The November 2003 issue of NASA's newsletter El Noticiero de NASA highlighted Raul's work.

Thanks to the NASA CIPA project CIS majors have access to high performance computing from home. The CIPA Network Infrastructure includes a computer PubSSH for remote login anywhere on the Internet. Currently over 225 students have computer accounts that allow them to use NASA CIPA facilities. Using the Secure Shell Program (SSH), students can access Cluster Computing Project facilities from any computer with an SSH client. Remote access is essential in a commuter college like John Jay where many students spend much time off campus. Moreover, with PubSSH students are practicing secure computing. (SSH sessions and file transfers are encrypted.) Each semester, over 80 students in courses such as Computer Networking, Operating Systems, and Internship in CIS rely on the PubSSH server to access to NASA CIPA accounts from off campus sites.

The previous section described the extensive curriculum development and course revision sponsored by the project. Courses developed include five new undergraduate courses and three new graduate courses in computer science. Furthermore, the Department was able to upgrade both the level and quality of instruction in seven additional courses. (Please see Table 2 and Appendix II for details.). Students in these revised and newly developed courses now have excellent laboratory facilities to support course work in concurrent computing, computer networking, large-scale databases and operating systems. For example, all students exiting the CIS program have hands-on experience with enterprise-level databases, web information systems, message passing, and tools for analyzing network protocols. Without the new computing facilities and newly developed faculty and staff expertise provided by the NASA CIPA project, the curriculum development and course revision efforts would not have been possible.

The Department also put in place a rigorous mathematics requirement that gives CIS majors the analytical tools needed for the study of network and distributed computing. The new mathematics requirement is an important development. The new requirement includes a course in discrete mathematics, Discrete Structures (MAT 204). Students take this course early in their careers and it provides the requisite background in graph theory and probability needed for distributed and parallel computing. For example, students study graph-partitioning techniques that are used to load balance tasks among nodes and minimize communication. Under the new requirement, all students take two calculus-based operations research courses. The OR courses provide coverage of network queuing theory and network algorithms that are the basis of modern routing protocols. The new CIS Mathematics requirement, which was accepted by the College Curriculum Committee in 2003, exceeds the ACM Computing Curricular 2001 guidelines for mathematics requirements in computer majors.

The Department has developed six courses for the Masters Degree program in Forensic Computing. The program will give local and national law enforcement communities access to cutting edge research in computer network and distributed systems security. The program includes significant work in computer networking, distributed computing, and the security of distributed information systems. The NASA CIPA project allowed us to build institutional capabilities in these important areas of computing. The CUNY Board of Trustees approved the program early in January 2004, and the program accepted its first class in fall 2004. Two former NASA CIPA students are enrolled in the program.

Relevance to NASA Mission Future NASA missions and research depend on advances in computing, especially in the areas of network, parallel and distributed computing. For example, NASA initiatives in human centered computing (http://www.arctic-mars.org/1999/SCIENCE/hcc.html) and automated reasoning (http://is.arc.nasa.gov/AR/AR.html) require significant leaps in computing power. Essential are advances in computer networking, cluster computing, and parallel/ distributed computing. The January 2000 article "New Directions at NASA AMES" published in Computer (an IEEE Computer Society publication) outlines the computing challenges facing NASA, and the space agency's need for higher performing networks and computing systems for calculation and mass storage. A January 2003 Computer article, "NASA Advances Robotic Space Exploration," highlights NASA's need for low power, low mass, high performing computing systems on spacecraft. These are needed for fault-tolerance, mission control, and processing large volumes of data on the spacecraft (hundreds of gigabytes per day) to overcome severe earth to spacecraft bandwidth restrictions. NASA continues to use and develop parallel computing systems to meet these challenges. Indeed, NASA's centers rely on cluster and grid computing for solving computational fluids problems, climate modeling, and analyzing atmospheric data. In order to reduce costs, NASA increasingly turns to common-off-the-shelf hardware and software to build its systems rather than custom developed solutions. The use of commodity systems by NASA allows institutions with limited budgets to acquire or duplicate these systems.

Technology transfer has always been an important NASA goal. Computing research sponsored by NASA offers many benefits to communities outside the traditional aerospace community. The NASA CIPA Cluster Computing Project has served as a research and technology conduit that has allowed the computer science education and criminal justice research communities to benefit form NASA funded high performance computing research. NASA CIPA computing facilities at John Jay College are being used to solve problems in network forensics, criminal justice data mining, and fire science.

The primary NASA technologies used in this project was commodity cluster computing, developed by NASA and its contractors during the mid 90s at the NASA Goddard Space Flight Center in Greenbelt, Maryland. The NASA CIPA Cluster Computing Project has relied on cluster computing technologies and methodologies used and developed at the Jet Propulsion Laboratory, Goddard Institute for Space Studies (GISS), NAS of NASA AMES, and the Glen Research Centers. These technologies include the Portable Batch Systems (PBS), the NAS Parallel Benchmark codes, and Information Power Grid tools developed under the NASA Information Power Grid (IPG) project. Thanks to the NASA CIPA Cluster Computing Project, the College now has the cluster computing facilities and faculty expertise necessary to contribute to NASA programs that require high performance computing capabilities. The PI has offered the Computational Cluster to researchers at GISS and local area NASA PIs within the CUNY system who need high performance computation for heat transfer studies, atmospheric studies, and climate modeling.

The project has increased the number of underrepresented students at John Jay College in the NASA STEM pathway. In addition, these students will be developing expertise in areas of computing that are critical to NASA initiatives. Currently the CIS major has about 370 students. Given the tremendous drop in enrollments in computer science programs throughout the country (30% to 40% over the last three years), NASA programs such as the NASA CIPA Cluster Computing Project help ensure an adequate supply of highly trained computer scientists with expertise in areas of computing needed by NASA.

Partnerships Partnerships have allowed the John Jay NASA CIPA Cluster Computing Project to 1) make project facilities and expertise available to other organizations (e.g., the Science Dept., other CUNY units, and NASA) and to 2) obtain funding to complete and maintain project facilities. Partnerships with NASA organizations show the relevance of the project's work to NASA high performance computing initiatives. Table 4 lists the major Cluster Computing Partnerships to date.

Table 4: NASA CIPA Cluster Computing Project Partnerships

Organization	Activity
John Jay Math & Science Departments (contact: Prof. Anthony Carpi, Science Dept.)	Jointly developed and support Distributed Computing Lab; jointly sponsored the Cluster Computing Seminar Series
Queens College (contact: Prof. Bon Sy, Computer Science Dept.)	Working to developed a database specialization in data mining.
NASA Goddard Institute for Space Studies (contact: Dr. Jeff Jonas)	Investigated methods and software for running GISS climate models on commodity clusters.
NASA AMES (contact: Dr. Tom Hinke)	Investigated NASA Information Power Grid and its security mechanisms at the NASA Information Power Grid Tutorial, Feb. 2003.
NASA Peer Review Services (contact: Ignatius Hu)	Prepared the article "High Performance Computing for Reseach at John Jay College" for El Noticiero de NASA.
Polytechnic University (contact: Prof. Nasir Memon), NSF funded project	Developing a graduate course in Network Forensics using Polytechnic's Virtual Security Laboratory and the NASA CIPA Distributed Computing Lab.
Polytechnic University (contact: Prof. Nasir Memon), NSF funded project	Developing a tool for network forensic investigations using NASA CIPA network infrastructure.
Borough of Manhattan Community College (Contacts: Prof. Alice Cohen & Prof. Miranda Roy)	Formulated an articulation agreement between BMCC two-year degree programs in computing and the John Jay College bachelor degree program to foster student interest in network and distributed systems security.

The following are brief descriptions of the partnerships and their impact on CIS students, the CIS Program, and the Cluster Computing Project.

The partnership with the Science Department allowed us to complete the Distributed Computing Lab and successfully integrate Lab computing facilities into CIS courses and curricula. By partnering with Science, we were able to increase the number of lab computers from 15 to 25. The additional computers permit classes to held in the lab and ensure an adequate number of machines for student assignments. Lab computers provide high-speed access to the Computational Cluster so CIS courses now can make use of X-Windows applications that run on the cluster. The Science partnership also provided funding for display equipment, a file server, and extra support staff to maintain the lab and provide extra hours. In addition to CIS majors, 30 students in advanced Science courses now use the lab regularly to study molecular modeling using the Titan program by Wavefunction, Inc.

There were three collaborative efforts with NASA personnel during the project. In Feb. 2003, the PI attended the NASA Information Power Grid (IPG) Tutorial sponsored by NASA AMES. There he learned about security mechanisms used in NASA's IPG. This material appeared as a case study in the Computer Networking Course. The PI also worked with GISS. He examined methods for improving the performance of codes on both shared and distributed memory computers. GISS is interested in porting their climate modeling codes to various multiprocessor architectures. In a third effort, working with NASA Peer Review services, the PI wrote an article that highlights NASA sponsored student research in high performance computing.

The project led to three collaborations with local area colleges and universities. The PI and computer scientists at Polytechnic University are collaborating on two NSF funded projects. In one project, the PI will use Polytechnic's Virtual Security Lab and the NASA CIPA Distributed Computing Lab to develop laboratory modules for a graduate course in Network Forensics and an undergraduate course in Computer Networking. In the second project, the PI and Polytechnic faculty will develop a tool for gathering and analyzing network traffic. This project will require the NASA CIPA Computational Cluster. In a third partnership, John Jay College and Borough of Manhattan Community College have signed an articulation agreement that will allow students at BMCC, a community college, to transfer into John Jay's four year programs in computing. The articulation agreement makes the transition easier for students and alerts BMCC students to opportunities to study parallel and distributed computing at John Jay College. Thus, the articulation provides a step in the NASA STEM pathway.

Matching Funds Matching funds have been critical to the success of the project. While the NASA CIPA Cluster Computing project provided $30,400 for computing equipment, matching funds from New York State Grants, College cost sharing, and the Science partnership, provided an additional $74,940. The project used these funds to complete the Computational Cluster, the Database Cluster, the Distributed Computing Lab, and the required network infrastructure (Please see Table 5). The NASA CIPA project, however, was the primary source of funding for the personnel that built the cluster infrastructure, developed curriculum, and maintained cluster-computing facilities. It should be noted that matching funds allowed the Cluster Computing Project to continue during fall 2001 when NASA CIPA funds were unavailable due to the Sept. 11th tragedy.

Table 5: Matching Grants Attracted by the NASA CIPA Project

Source/year	**Grant or Project Title**	**Amount**
GRTI[1]/2001	A Parallel Computing Environment	24,990
GRTI/2002	A Multicluster Parallel Computing Environment	8,450
GRTI/2003	Secure Hosting Environments for Web Services	11,200
Technology Fee College Grant 2002	Distributed Computing Lab Support and Tutors	7,000
Math & Science Partnership 2002	Distributed Computing Lab Furniture and Facilities	15,300
Math & Science Partnership 2003	Distributed Computing Lab Computers	15,000
NSF/Dept. of Education	Web-based Science Education	15,600
Total Funds		97,540

[1]Graduate Research & Teaching Initiative, State of New York Grant

In addition to matching funds listed in the table, the College provided equipment and the personnel needed to rewire the Distributed Computing Lab and upgrade electrical connections in the room that houses the Computational Cluster, routers and project servers. The college also has allowed $7000 of released time funds to be used for personnel to support Cluster Computing Project facilities.

6.0 Dissemination Efforts

Dissemination was a major part of the project. Dissemination efforts served two purposes: 1) to give John Jay College faculty, staff and students information needed to take advantage of Cluster Computing Project facilities, curricula and expertise, and 2) to share project benefits with colleagues in other organizations interested in computer networking, distributed, and high performance computing. Dissemination efforts included a seminar series; a project web site (web.math.jjay.cuny.edu); reports and presentations by faculty, students and staff at local and national conferences.

The project sponsored the Cluster Computing Seminar Series. Speakers from academic institutions and major companies such as 3 COM and Alcatel presented talks on high performance computing, computer networking, computer security, cluster computing and message passing. In addition, the PI, Co-PIs and several CIPA students presented seminars to keep CIS students and faculty members informed of developments in the Cluster Computing Project. For example, twice the PI presented an overview of the Cluster Computing Project. Over 100 CIS students, faculty and staff members attended each colloquium. In 2002, the PI and Prof. Anthony Carpi of the Science Dept. held a colloquium and luncheon to introduce the Distributed Computing Lab to the College community. Both the College Provost and President attended along with over 150 CIS and Science majors. Throughout the project, the Computer Science Association, the student computer club, held workshops (at least 15 per year) that helped train students, faculty and staff to use NASA CIPA computing facilities. Appendix III provides a partial list of the seminars sponsored by the project.

The Cluster Computing Project has a web site (http://web.math.jjay.cuny.edu) that lists project activities and provides information on cluster computing resources. NASA CIPA students helped develop the site. The site provides a comprehensive overview of project computing facilities and curriculum development efforts. Students use the web site to learn what computing resources are available and how to use them. For example, the site provides a tutorial in how to use the Secure Shell Program (SSH), which is needed to access NASA CIPA computers from home. Instructors now use the site routinely in their courses to get students started using SSH, the Distributed Computing Lab, and the Computational and Database Clusters. The site also provides access to project reports.

Faculty students and staff made a number of off campus presentations during the project. Prof. Mantharam, a Co-PI, presented a paper entitled “Improving a Parallel Algorithm Using a Visual Display” at the Consortium for Computing Science in Colleges Easter Division Conference in November 2002. The paper describes the visualization techniques she has developed for introducing students to parallel linear algebra computations. Mr. Bonderenko, a former student and the current lead staff person for the project, presented two research papers at NASA Ascend McNair Conferences. In 2002, he presented his work that compares the Andrew, NFS 3 and NFS 4 file systems for use in a computational cluster. At the 2003 conference, he presented his work on the Large Number Library. The large number library project was an effort to develop a library for performing number theoretic calculation on large integers in a cluster environment. In 2002, Atiqual Mondal, another NASA CIPA student, presented his work on the FBI’s NIBRS database at that year’s Ascend McNair Conference. The Ascend McNair conferences are held each fall at the CUNY Graduate Center and allow students who are bound for graduate studies to present research projects begun as undergraduates. The PI and Co-PI, Prof Peter Shenkin, acted as mentors for the projects.

The PI made several presentations at national conferences and off campus sites. He presented the NASA CIPA Cluster Computing project at the UNCFSP sponsored 2002 Reverse Site Visit in Dallas, TX.

In addition, he presented the project at the 2003 NASA Research Summit in Orlando, FL. The talk attracted attention from representatives at several NASA centers that rely on computer clusters to meet their high performance computer needs. In addition, he presented a paper at the 2004 USENIX Annual Technical Conference entitled, "Cluster Computing in a Computer Major in a College of Criminal Justice." This talk explored the use of cluster computers and high performance computing to solve problems in network security and fire safety. The talk also examined cluster design issues and methods that ensure clusters perform accurate computations in solving large-scale linear algebra problems. A summary of the talk, which appeared in the Oct. 2004 issue of Login , appears as Appendix VI.

The NASA CIPA project sponsored a number of research efforts that are on going. The PI continues to document project computing facilities and experience gained in network, cluster and high performance computing during the project. He and Mr. Bonderenko are preparing documents that describe the cluster infrastructure (see Appendix I), the new NIBRS Oracle Database, and performance monitoring of software for matrix computations. The PI is preparing a report on current NASA uses of cluster and grid computing to solve compute- and data-intensive problems.

7.0 Project Future

The significant institutional capabilities in computer networking and high performance computing made possible by the NASA CIPA project will continue to foster research and educational opportunities at John Jay College for many years. Sponsored by two recent NSF grants, the PI and NASA CIPA staff will develop tools for network forensics and curricula in network security. This work will rely on the computational and database clusters as well as expertise in high performance computing gleaned during the NASA CIPA project. Current educational programs sponsored by the NASA CIPA project, e.g., tutoring, the Distributed Computing Lab, and PUBSSH remote access servers, will continue to be supported with College funding. The PI and staff continue to look for new ways to build inexpensive, reliable, high performance computing environments. Finally, the PI and staff will continue to mine NASA for high performance computing solutions.

8.0 Personnel Changes

Project funds for personnel have been exhausted. The project, however, continues with funds from College sources and current NSF grants. The lead staff member Boris Bonderenko and Raul Cabrera remain with the project. In addition, Isaiah Brown continues work as a tutor for the project.

One new student was added to the project this year on NASA CIPA funds. Mathew Cheng joined the project and was supported by $1400 of surplus personnel funds carried over into 2004 from years 2001-2003. Mathew works as a tutor in the Distributed Computing Lab where he helps students with course work and shows them how to use Linux accounts and CIPA computing facilities.

9.0 Expenditures

The following Final Fiscal Report provides a summary of the project budget. The Report provides the original approved budgets for the years 2001-2003. The total for each budget category for the three years also is listed. The column “4-Year Expenses” shows the amount spent during the three-year project. This column also includes carryover funds in the amount of $4,106.11 that were spent in 2004. These funds were part of a no-cost extension. The table also shows in the “Remainder” column that the project utilized all funds in all categories. The table does not show matching funds obtained for the project during the last four years. These were discussed previously. A budget summary follows the Final Fiscal report.

CIPA Program
Final Fiscal Report

For January 1, 2001 to November 30, 2004

	Budget	Budget	Budget	3 Year		4 Year Expenses[1]	Remainder[2]
	2001	2002	2003	Totals			
1. Direct Labor							
a. Salaries, wages	9,746	9,746	9,746	29,238		29,238	0
b. Fringe Benfits	2,144	2,144	2,144	6,432		6,432	0
2. Other Direct Costs							
a. Faculty Released Time and Summer Stipends	22,360	22,360	22,360	67,080		67,080	0
b. Staff Support	7,320	7,320	7,320	21,960		21,960	0
c. Student Assistants	9,810	9,810	9,810	29,430		29,430	0
d. Facility Improvement	0	0	0	0		0	0
e. Equipment	20,400	20,400	20,400	30,400		30,400	0
f. Supplies	3,500	3,500	3,500	10,500		10,500	0
g. Copying Faxing							
h. Office Space							
i. Travel	4,000	4,000	4,000	12,000		12,000	0
j. Other	5,100	5,100	5,100	15,300		15,300	0
3. Facilities and Administration							
Costs (Indirects)	15,533	15,533	15,533	46,599		46,599	0
4. Other Applicable Costs							
5. Totals	99,913	99,913	99,913	268,939		268,939	0

[1] A no-cost extension allowed $4,106.11 of remaining direct costs from years 1,2 and 3 to be spent in 2004.
[2]All funds in all budget categories have been utilized.

Budget Summary This section provides an explanation of each budget item.

Category 1 Direct Labor includes the PI's summer salary and fringe benefits. The PI worked full time during the summers and oversaw all aspects of the project. He played key roles in facilities, curriculum and faculty development. He oversaw most grant administration tasks so funds for salaries could be devoted to the educational aspects of the project rather than the administration. Each year of the project, the PI received a summer salary of $9,746 and fringe benefits mandated by CUNY of $2,144.

Categories 2a Faculty Release Time and Summer Study includes $12,000 of released time (4 sections) for the PI and two Co-PIs, $3,360 for required released time fringe benefits, $4000 for summer study, and $3000 for training fees, e.g., fees for conference registrations and tutorials. Released time is charged at the adjunct replacement rate, which represents significant cost sharing on the part of the College. In addition, the College permitted $7000 of released time monies to be used for students to work as lab assistants. The two Co-PIs were compensated for summer activities such as curriculum development, self-study and conference attendance. The CO-PIs received a $2000 summer salary, which included fringe benefits mandated by CUNY.

Categories 2b and 2c Student Assistants and Support Staff provided funding for non-faculty salaries. These included the lead staff member, Mr. Boris Bonderenko, and seven students who were funded directly by the project from 2001 through 2004. The lead staff member worked closely with the PI throughout the project. He oversaw the management and development of all CIPA Cluster Computing facilities, provided lab-support, web site development services, and support for the computational and database cluster academic/research projects. In addition, he directed students who set up and supported NASA CIPA computing facilities. It should be noted that the NASA CIPA project only provided about half of the lead staff member's salary. The partnership with the Science Department provided the other half, which allowed the staff member to be employed full-time throughout the project. During the course of the project, matching funds made it possible to employ ten students in addition to those directly supported by the project.

Category 2e The Equipment category provided funds for computers and related equipment (e.g., equipment racks, power conditioning systems, and network equipment). The project anticipated heavy equipment expenditures in the first year when the computational cluster was built so the budget for that year is $20,500. The project budgeted only $5000 for the next two years. In hindsight, the amount was insufficient. We underestimated the costs of required upgrades, equipment replacements, and the need for addition equipment (e.g., a 1.2 TB mass storage system) due to new research initiatives. Equipment matching funds, however, provided the necessary funding.

Category 2f The Supplies category included monies for small purchases and consumables. The project used these funds primarily for printer toner cartridges, storage media, small hardware upgrades, software, software licenses and office supplies. The College provided standard office consumables such as paper.

Category 2i The Travel category provided funds for the PI, Co-PIs and staff member to attend eight professional conferences and tutorials. Besides paying for travel to a conference, the travel budget also provided funds for tutorial and conference registration fees. Conference attendance provided critical in achieving the faculty/staff development objectives of the project. Faculty and staff presentations at conferences fostered collaborations and helped make the project work know to various computing communities (e.g., the USENIX, educational, and cluster computing communities)

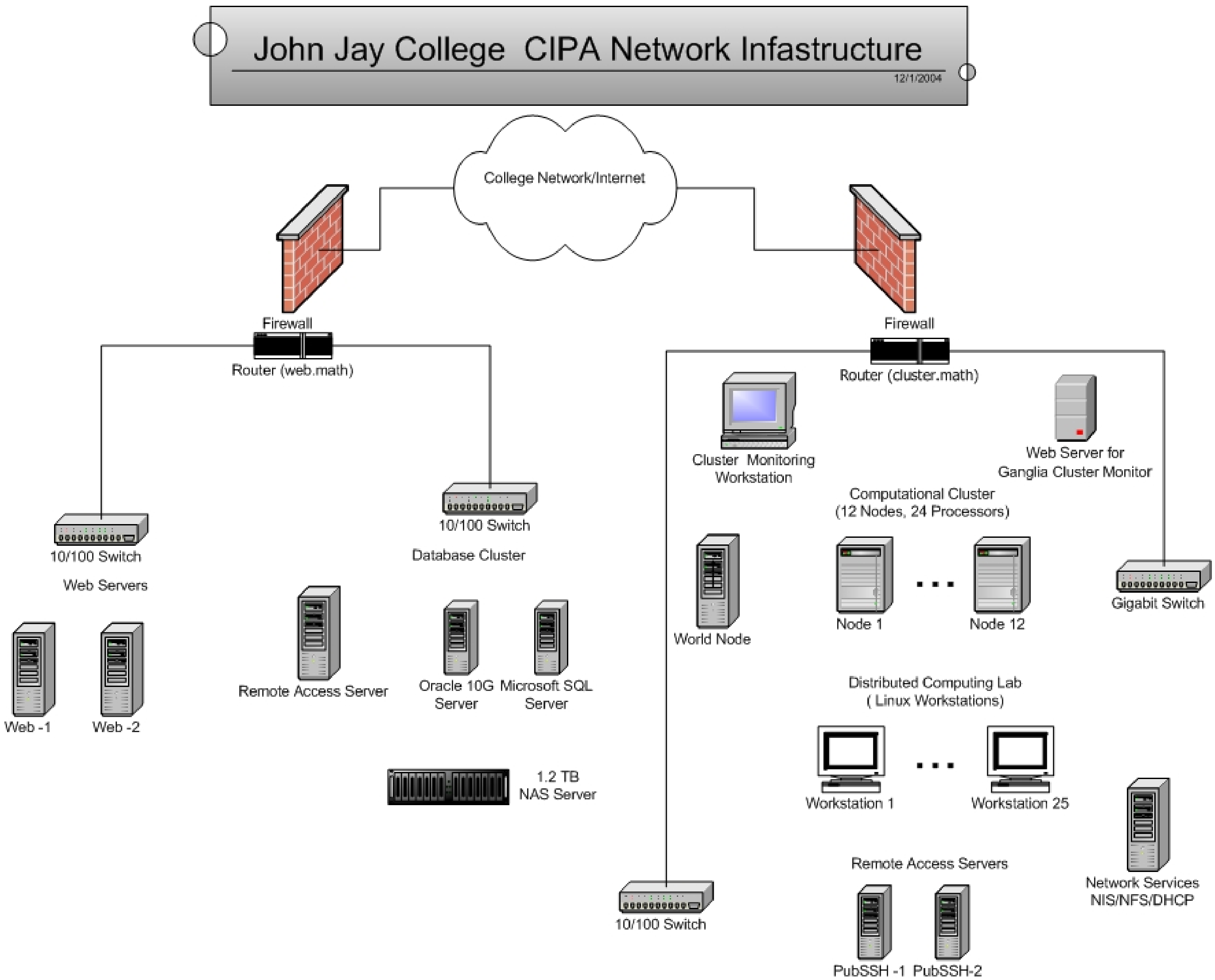
Category 2j The Other category provided funding for some replacement equipment, additional conference and tutorial fees, dissemination expenses (e.g., web site development), software for students (e.g., Microsoft Academic Alliance), and network equipment not anticipated in the original budget (e.g., two firewall/routers).

Category 3 Indirect Costs represent overhead charges on salaries, which are required by CUNY. Since all salary monies were spent in the course of the project, the grant was charged the full amount for indirect costs.

10.0 Additional Remarks

The NASA CIPA cluster computing has had a profound impact on computer science programs at John Jay College. The facilities, curriculum and faculty/staff development sponsored by the project have given the College institutional capabilities in network and high performance computing, areas of computing that are important to NASA and important to the mission of the College. The project has allowed the College to leverage NASA high performance computing research and technology to enhance its undergraduate program, build a graduate program, and conduct research in computer security and network forensics. Most importantly, the project has given John Jay students and faculty access to leading edge technologies in networking, parallel and high performance computing. John Jay students now have the opportunity to participate in the modern era of network-centric and high performance computing.

Appendix I NASA CIPA Cluster Computing Project Facilities



Appendix II

Courses Developed or Revised as part of the NASA/CIPA Project

Course Name	Course No.	Faculty Member/s	Status	Yearly Enrollments[1]
Artificial Intelligence	MAT 376	Prof. Georgatos	Approved 2002, CIS elective	Offered spring 2005, 20 anticipated
Computer Networking	MAT 379	Prof. Salane	Approved 2002, CIS elective	30
Database Systems in Criminal Justice	MAT 470	Prof. Shenkin, Ms. Maedev	Revised 2002, CIS requirement	30
Discrete Structures	MAT 204	Prof. Ye	Approved 2002, CIS requirement for 2003.	30
Graphics and Graphical User Interfaces	MAT 273	Prof. Kugan	Approved 2002, CIS elective	40
Internship in Information Systems	MAT 404	Prof. Graff	Revised, CIS requirement	40
Operating Systems	MAT 375	Prof. Salane	Revised, CIS requirement	45
Parallel Computing	MAT 490	Prof. Mantharam	Developed, experimental course in 2003	Offered spring 2005, 20 anticipated
Programming Languages	MAT 374	Prof. Chandrakanthra	Revised, CIS requirement	45
Quantitative Methods in Criminal Justice	MAT 400	Prof. Shenkin and Ms. Maedev	Revised, CIS requirement	45
Systems Analysis	MAT 276	Prof. Georgatos	Approved 2002, CIS elective	15
Network Security	FSC[2] 742	Prof. Ji	Required M.S. course	Offered fall 2004, 15
Network Forensics	FSC 745	Prof. Salane	Required M.S. course	Offered Fall 2005, 15
Architecture of Secure Operating Systems	FSC 710	Prof. Chandrakantha	Required M.S. course	Offered spring 2005, 15

1This column shows typical annual enrollments since 2002. The column also shows that each year over 300 students enroll in computer courses that take advantage of NASA CIPA Cluster Computing Project computing facilities or curricula.

[2]The call letters of courses in the M.S. Program in Forensic Computing were changed to FSC in 2003. The Department of Mathematics and Computer Science offers these courses.
.

Appendix III CIPA Cluster Computing Seminars

The following is a partial list of seminars supported either fully or in part by the NASA/CIPA Cluster Computing Seminar series. The Science Dept. and the Computer Science Association co-sponsored about half of the seminars in the series.

"NASA CIPA Cluster Computing Project facilities," D. Salane, John Jay College

"Workshop on PowerPoint," U. Singha, John Jay College

"Workshop on ASP & ASP.Net: Building the NIBRS Database," P. Shenkin and A. Mondal, John Jay College

"Workshop on Excel and Visual Basic," U. Singha, John Jay College

"Using Visual Basic and SQL," P. Shenkin, John Jay College

"Secure Network Switching," Adam Glass, ALCATEL Corp.

"Innovative Network Technologies," Wayne Simms, 3 COM Corporation

"The Math and Science Resource Center," Anthony Carpi, Science Dept.

"Reasoning about Message Passing in Finite State Environments," R. Ramanujam, Indian Institute of Mathematical Sciences

"Network Technologies to Support Bandwidth Sensitive Applications," Adam Glass, 3 COM Corporation

"Securing the TCP/IP Protocol," D. Salane, John Jay College

"Linux Workshop," Boris Bonderenko, The Computer Science Association

"Dual Boot Systems," Boris Bonderenko and Raul Cabrera, The Computer Science Association

"The Webmaster Contest," Uttam Singha, The Computer Science Association

"PowerPoint," Uttam Singah, The Computer Science Association

"Evaluation of the CIS Major," D. Salane, John Jay College

"CodeLab: An Automated Evaluator of Student Programming Exercises," David Arnow, Brooklyn College.

"Combating Idenitity Theft," Lieutenant John Garcia, Computer Crime Squad, New York City Police Department.

"Linux Workshop," Carlos Espejo,The Computer Science Association.

"Web Design," Carlos Espejo, The Computer Science Association.

.

Appendix IV Faculty, Student and Staff Papers and Presentations

Presentations by NASA CIPA Faculty and Staff

Presenter	Title of Paper or Presentation	Conference
B. Bonderenko	Comparison of the NFSv3, NFSv4 and Open AFS Distributed File systems	Ascend McNair Conference, CUNY Graduate Center, September 2001 CSTEP Conference, Lake George, NY, April 2002
B. Bonderenko	A Large Number Library for Computational Clusters	Ascend McNair Conference, CUNY Graduate Center, September 2002
A. Mondal & P. Shenkin	A Web-Based Information System for the FBI's National Incident Based Reporting System Data	Ascend McNair Conference, CUNY Graduate Center, September 2002
M. Mantharam	Improving a Parallel Algorithm Using a Visual Display	Consortium for Computing Science in College Easter Division Conference, PA, October 2003.
D. Salane	Cluster Computing in a Computer Science Major in a College of Criminal Justice	Presented at the Usenix Annual Technical Conference, Boston, MA, July 2004
D. Salane	High Performance Computing for Research at John Jay College	El Noticiero de NASA, October 2003
D. Salane	The Cluster Computing Project	Presented at the NASA Research Summit, Orlando, FL, July 2003
D. Salane	The John Jay NASA CIPA Cluster Computing Project	NASA Reverse Site Visit, Dallas, TX, October 2002
D. Salane	Using Cluster Computing Resources	Cluster Computing Seminar Series, John Jay College, December, 2002

Appendix V NASA/CIPA Students

The students listed in the following table participated in the NASA CIPA project. About half of the students received stipends from the NASA CIPA program; the remainder received salaries from College or matching funds. The columns Current Employer and Current MSET Program indicate that just about all students who have graduated now have positions in the IT field and are pursuing graduate studies in computer science. All students who participated intend to pursue graduate studies in Computer Science.

NASA CIPA Students 2001–2004

Student	**Period in NASA/CIPA**	**Current Employer**	**Current MSET Program**	**Race**
Richard Cumberbatch	Spring and Summer 2001, graduated June 2001	MIS Dept., NYC Dept. of Health and Human Services	M.S. in Computer Science, Steven's Institute of Technology	Black
Gerardo Vasquez	Spring 2001to Spring 2002, graduate June 2002	Hawaii Pacific University	M.S. in Computer Science, Hawaii Pacific University	Hispanic
Namkumar Itwaru	Spring and fall 2001, graduated Jan. 2002	Research Foundation CUNY, Database Developer	M.S. in Computer Science, CUNY	Hispanic
Boris Bonderenko	Summer 2001 to present, graduated Jan. 2002	Research Foundation of CUNY, NSF Virtual Security Lab Project	Intends to enter graduate school in the fall 2005.	White
Raul Cabrera	Fall 2002 to Spring 2004	John Jay College, Library MIS manager	Pursuing M.S. in Forensic Computing, John Jay College/CUNY Jay	Hispanic
John Young	Fall 2002 to 2003	Active duty in National Guard.	CIS, John Jay	Black
Atiqul Mondal	Summer 2001 to Jan. 2003	Research Foundation of CUNY	M.S. in Computer Science, City College of CUNY	Black
Tien Nugyen	Spring and summer 2001, graduated June 2001	IBM, Operating Systems Div., OS developer	M.S. in Computer Science, Marist College	Asian
Krafins Villcin	Spring and fall 2001, graduated Jan. 2002	York College, MIS Dept., Database Developer	M.S. in Computer Science, Queens College/CUNY	Black
Anton Lebedev	Spring/Fall 2002	Unknown	None	White

NASA CIPA Students 2001 – 2004 (continued)				
Student	**Period in NASA/CIPA**	**Current Employer**	**Current MSET Program**	**Race**
Ke Tang	Fall 2002 to Fall 2003	CUNY Graduate Center.	Pursuing Ph.D. in Computer Science, Graduate Center of CUNY	Asian
Ms. Samra Varesanovic	Spring to Fall 2003	Database consultant	Purusing M.S. in Computer Science	White
Uttam Singha	Spring 2002 to Spring 2003	Research Foundation of CUNY	Pursuing M.S. in Computer Science, City College/CUNY	Black
Isaiah Brown	Fall 2003 to Present	CUNY Technology Fee	Pursuing B.S. in Computer Information Systems	Black
Claudine Grant	Fall 2003	CUNY Technology Fee	Pursuing B.S. in Computer Information Systems	Black
Mathew Cheng	Fall 2004	Research Foundation of CUNY.	Pursuing B.S. in Computer Systems, John Jay College	Asian

Appendix IV October 2004 issue of Login, Conference Reports Section, p. 72

tem. A single machine can be booted from Quantian media and then other machines can network-boot via the PXE protocol and form a single Mosix cluster.

Quantian extends clusterKnoppix with a large number of scientific computing applications. In particular, Beowulf-style clustering tools and libraries are included along with the statistical package R and the SNOW extensions. SNOW allows easy access to high-level parallel statistical computing. Some Knoppix packages that are not related to scientific computing or related software development have been dropped in order to make room for Quantian's scientific computing additions.

Currently, Quantian is essentially a one-man operation maintained by Dirk. He responds to requests for the addition of new packages as time and interest allow. Distribution size, network security concerns, and surveying users for their needs and configurations remain open issues for him. Even though it is primarily a repackaging of other components, Quantian deserves a look if you are interested in scientific computing. At the end of his talk, Dirk mentioned that the laptop he was using was running Quantian with a USB flash drive for persistent storage. It seems that his employer will not let him install Linux on the company-supplied laptop, so he has found another way. Let's hope Dirk keeps finding another way.

Cluster Computing in a Computer Major in a College of Criminal Justice

Boris Bondarenko and Douglas E. Salane, John Jay College of Criminal Justice

John Jay College is a specialized liberal arts college within the City University of New York system. It offers degrees in Law and Police Science, Fire Science, and Forensic Science among others. So, you might ask, just what kind of cluster computing is needed in a College of Criminal Justice? Douglas Salane made it clear that there are a number of areas where significant computing resources can be helpful.

Current and planned projects include simulations of the fires that occurred after the attack on the World Trade Center, database analysis and data mining of the FBI's National Incident-Based Reporting System, and molecular modeling for toxicology studies.

John Jay College has a relatively small cluster-computing facility. The compute cluster consists of 12 nodes with two CPUs each. A separate database cluster has four nodes, and the computing laboratory has 30 Linux workstations. Still, they had to go through much of the same decision-making processes that larger facilities might go through. Blade/rack systems or piles of PCs? What network file system to use? What interconnect technology? How to manage and monitor the cluster? How to test the correct functioning of the cluster? A cluster-specific Linux distribution or self-configuration?

Verifying the correct functioning of the cluster was of particular concern to Douglas. This concern was strengthened when the test software that is included in the BLACS portion of the ScaLAPACK software library reported incorrect results for some of its tests. In the end, the error was traced to a faulty Gigabit Ethernet card in one of the machines. Other cluster-computing packages don't always provide those kinds of tests. On the other hand, ScaLAPACK can be difficult to use.

For a small site, just figuring out what cluster-computing software is available and how to set it up is a significant undertaking. Unfortunately, Linux distributions, like the previously mentioned Quantian, were not available when they first started working on their cluster. Support for heterogeneous clusters would also help by allowing them to expand the size of their cluster over time without sacrificing performance to the demands of optimizing software to the lowest common denominator.

Appendix VII (Photos)

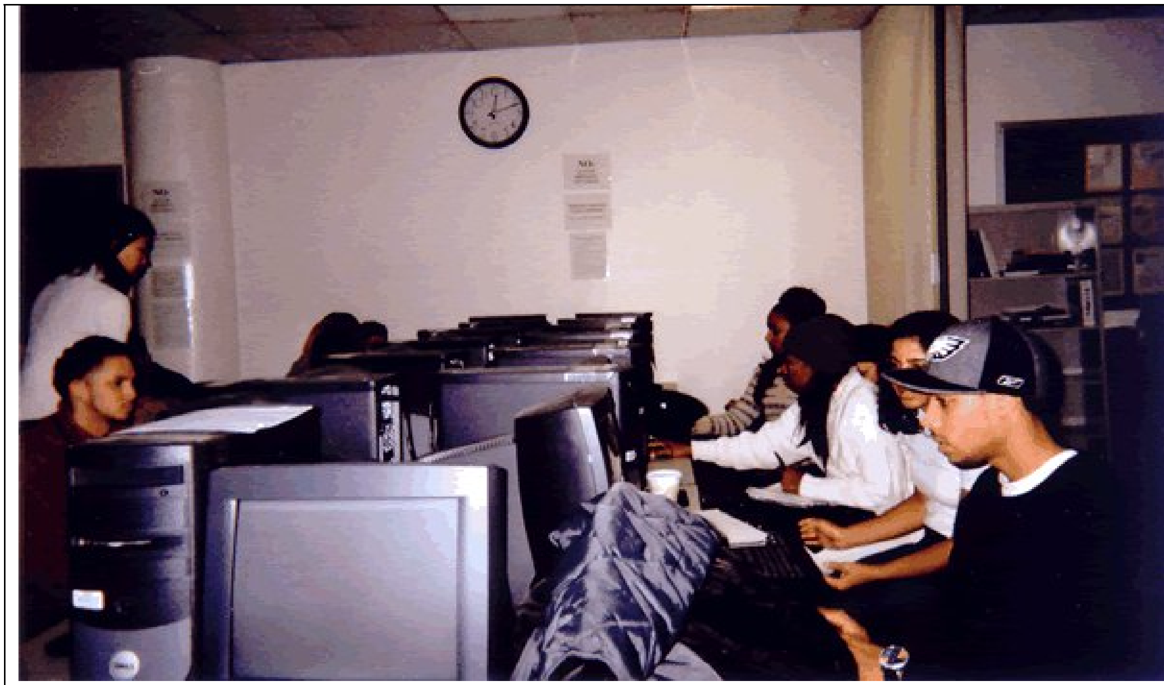
NASA/CIPA Cluster Computing PHOTOS

The PI and Prof. Samuel, Math Dept. Chair, Examine Cluster Results

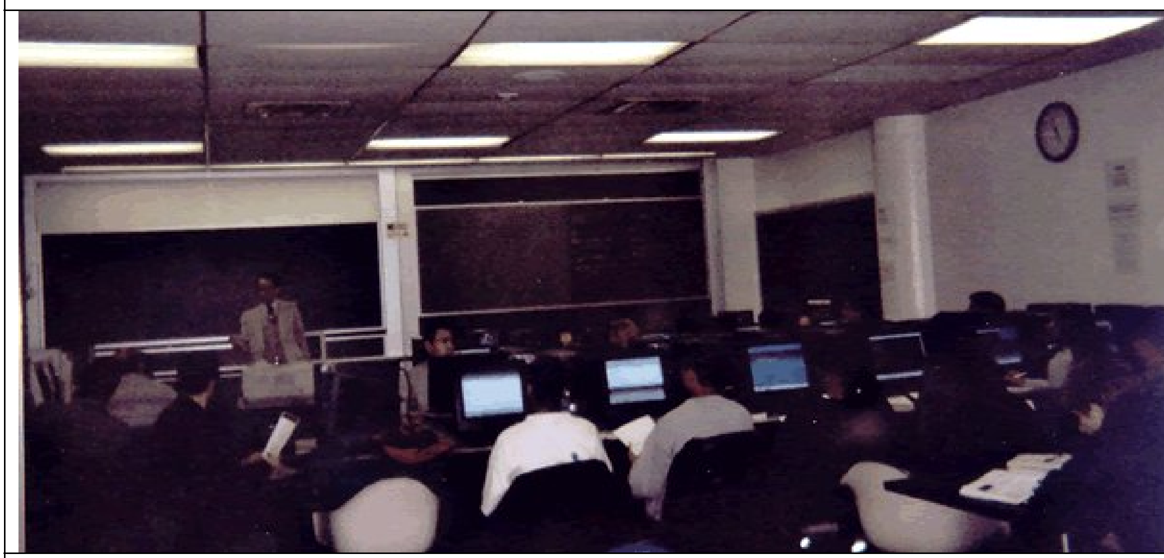
Prof. Shenkin, Prof. Samuel and CIPA Students

The PI and CIPA Students

Student Collaboration in the Distributed Computing Lab



Students Working on Assignments



An Advanced CIS Class in the Distributed Computing Lab

Refreshments at a Cluster Computing Project Colloquium