EXECUTIVE SUMMARY

A survey of faculty was conducted in October-November, 2005 in order to identify the existing demand and potential appetite for high performance computing within the Davis campus. The intent of the survey was to glean sufficient information to inform planning efforts and to determine the most reasonable and efficient way for the campus to accommodate the increasing demand. The table below summarizes the resultant 59% response rate to the survey and indicates that additional follow up may be required in order to have a truly comprehensive campus assessment.

College/School Survey Response Summary

College	Contact Qty	Unavailable Qty	Response Qty (including Joint Responses)	N/A Qty	Response Rate
Agriculture	29		12	6	62%
Engineering	33		13	3	48%
Social Science	9	1	4	4	100%
Biological Science	3		1	2	100%
SOM	18		4	3	39%
Math and Physical Sciences	31	1	13	6	65%
Total	123	2	47	24	59%

High Performance Computing Demands

In general, only survey data representing demand through 2010 was analyzed. Data collected for 2015 was viewed as statistically inaccurate due to the small number of responses and the difficulty researchers had in predicting ten years in advance. A summary of the data by College/School and by Department is provided in Section 2 of this report.

The survey identified approximately 1900 cluster nodes currently in place on campus or through collaborative use at remote facilities such as NCAR. This need was projected to increase by over 157% to approximately 5000 nodes by 2010. This translated to 64 computer racks distributed on campus with a projected increase to 171 racks by 2010. The 2010 demand could conceivably be reduced to about 140 racks if all of the clusters were consolidated. Very few respondents reported the use of shared memory machines.

The increase in data storage requirements is projected to be even more significant – from a current demand of 51terabytes of hosted data and 258 terabytes of managed data to 10 pedabytes of hosted data and 3 pedabytes of managed data . There was inconsequential use of pay as you go storage. Of this data, all but three respondents indicated that there were no HIPAA or personnel related data access restrictions on the data they collected. Mechanical and Aeronautical Engineering indicated a need to comply with HIPAA requirements and Psychology and Genomics indicated that they stored personnel data.

High Performance Computing Space Requirements

Many of the surveys had incomplete information related to cooling capacity, power requirements, floor space, etc. This is not a significant issue as this information can be derived from data center metrics should a level of server consolidation is determined.

A number of the faculty responses indicated that their current computing space is or will soon be inadequate to support their increasing demand. Should these decentralized deployments continue to expand, it is recommended that Facilities review the space and potential infrastructure improvements that may be required over the planning horizon.

Responses to the surveys indicated that clusters are currently housed in a number of locations around campus. The following list indicates how dispersed the computing resources currently are:

College Of Agriculture
Plant and Environmental Sciences – Room 118
Asmundson Hall – Room 148
Plant Reproductive Biology
Hoagland Hall – Room 126
Wickson Hall (Planned)
Robbins Hall (Planned)
Storer Hall – Room 5350

College of Engineering
Academic Surge – Room 2120
Academic Surge – Room 2224
Academic Surge – Room 1344
Non UCD Facilities (Collaborators Facilities)
Bainer Hall – 2nd Floor
Kemper Basement
Campus Data Center (Planned)
Genomics Building (Planned)

Math and Physical Sciences
Physics Building – Room 307
Physics Building – Room 429
Physics Building – Room 518
Mathematical Sciences Building (Planned)
Campus Data Center (Planned)
Academic Surge Room 1334
Chemistry Building

Social Sciences
Center for Mind and Brain

Biological Sciences
Genome Center
Plant Reproductive Biology Building
Academic Surge

School of Medicine
2921 Stockton Boulevard, Suite 1400 (Planned for CRISP)

Characteristics of High Performance Computing Use

The majority of faculty indicated that their long term needs required dedicated and secure facilities with 7x24 access. The need for facility proximity to the faculty that use housed resources was nearly evenly split between being important and not important – indicative of two distinct types of access requirements. This was also reflected in the near even split of faculty that access their facilities frequently versus those that access their facilities only once per month or never. The main reasons for access appear to be for system administration, system installation and upgrades, with a lesser number of visits being related to system backups or failures.

The majority of system support for cluster computing was reported to be handled by shared systems administrators. The number of faculty with dedicated system administrators was nearly half of those that used shared administrators. A very small minority of faculty used data center staff or administered their clusters themselves or with the help of students. Interestingly, the number of faculty that preferred a managed cluster environment was twice the number that preferred hotel space.

The overwhelming majority of faculty use clusters consisting of Opteron processors and the Linux Operating System. This exceeded the combined responses for faculty using Intel, IBM, or Athlon processors. Use of OSX, Unix, Rocks, and Windows operating systems in cluster environments was minimal.

The networking demands within the majority of responses for both wide area and local area networking were either equal to or exceeded what current campus capabilities are, indicating the need for a high speed capable research network. The majority of respondents indicating a need for local area networking speeds of 1Gb/second or better and for equivalent wide area networking speed. Much of the local area network demand is expected to increase to 10Gb/second by 2010. A few faculty indicate that their wide area indicate that their needs can still be met with an external 1Gb/second connection. This is indicative of applications that require more intracampus communication than intercampus communication.

Application Drivers

The majority of software used in the high performance computing environment is either open source or locally developed for specialized purposes. The following summary briefly describes the broad variety of applications currently driving the demand for cluster computing:

- Molecular structure calculations, protein folding, molecular dynamics, whole genome comparisons.
- Parallel finite element methods for inelastic solids applied to seismic analysis of soil-foundation-bridge structures.
- Modeling the reactive chemical transport problems that describe urban and regional air pollution.
- Three dimensional computer modeling of fluid mechanic flows.
- Simulation of soft-condensed matter, especially polymers and biomembranes.
- Development of simulation methods in studies of the structure and properties of advanced materials for electronic/magnetic, catalytic and structure applications.
- Simulation of protein dynamics.
- Computational bioengineering and biomedical imaging.
- Numerical investigation of complex systems, typically focusing on dynamical systems.
- Navier-Stokes computations.
- Computational fluid dynamic simulations of flows over/through a variety of configurations.
- Streaming remotely-sensed data, real-time processing of spatio-temporal data from satellites; distributed high-performance stream processing architectures.
- Transforming LiDAR point-cloud data in gridded datasets using the Kepler scientific workflow system.
- Ecological Niche modeling from the SEEK project.
- Investigation of general-purpose computing on graphics processors with long-term research goals toward investigating new data-parallel single-chip commodity processors.
- Statistical bioinformatics involving the analysis of large complex data sets generated by high-throughput biological assay methods.

- Fluid dynamics and mechanical deformation numerical modeling applied to plate tectonics and mantle convection in Earth.
- Modeling and simulation of state-of-the-art engineering systems such as VLSI circuits, MEMS, and optical transport systems.
- Analyzing and processing large volumes of digital imaging related to galaxy positions and attributes.
- High Performance Computer simulations primarily in the area of complex earth systems.
- High dimensional nonlinear dynamics, pattern formation, statistical mechanics of phase transitions, evolutionary dynamics, dynamics of learning, statistical inference, computation and information theories, multiagent systems, and distributed robotics.
- Cosmology and Astronomy using wide field optical imaging as a precursor to the Large Synoptic Survey Telescope (LSST).
- Computational condensed matter physics related to explaining and predicting physical properties of solids.
- Experimental high energy (particle) physics.
- Theoretical condensed matter physics: intermetallics; semiconductors; correlated electron systems; magnetism; superconductivity.
- Nanostructures: new physical properties and spintronics applications.
- Clusters: atomic and electronic structure; electric and magnetic dipoles; optical properties.
- Simulation of atomic systems for free energy calculations to understand biomolecular dynamics.
- Analysis of functional neuroimaging data.
- Forward and inverse modeling of brain activation data.
- Analysis and models of musical stimuli.
- Simulations of complex special models, statistical computations in ecology involving mcmc methods and other computer intensive aspects, solutions of pdes and integrodifference equations.
- Bioinformatics sequence comparisons.
- Computational biology and computations phylogenetics.
- Numerical solution of large systems of partial differential equations for groundwater flow and contaminant transport analysis.
- Simulation of human exposure to contaminants via groundwater pathways.
- Geostatistical simulation of subsurface complexity.
- Performing theoretical hydrodynamic instability calculations.
- Investigating and possibly solving a characteristic Arctic surface climate error in the NCAR CCSM3.
- Real time weather data forecasting over the California region.

SECTION 2

HIGH PERFORMANCE COMPUTING DEMAND SUMMARY

Biological Sciences – Node Summary

Department	2005 Nodes	2010 Nodes
Genomics	185	1000
Evolution and Ecology	32	32
Total	217	1032

Biological Sciences – Rack Summary

Department	2005 Racks	2010 Racks
Genomics	15	50
Evolution and Ecology	1	1
Total	16	51

Biological Sciences – Data Summary (Terabytes)

Department	2005 Hosted Data	2005 Managed Data	2010 Hosted Data	2010 Managed Data	2005 Total Data	2010 Total Data
Genomics	0	8.0	0	100.0	8.0	100.0
Evolution and Ecology	0	0.3	0	10.0	0.3	10.0
Total	0	8.3	0	110.0	8.3	110.0

College of Agriculture and Environmental Sciences – Node Summary

Department	2005 Nodes	2010 Nodes
Environmental Science and Policy	NA	16
LAWR	191	313
Plant Science	7	100
Total	198	429

College of Agriculture and Environmental Sciences – Rack Summary

Department	2005 Racks	2010 Racks
Environmental Science and Policy	0	1
LAWR	5	8
Plant Science	1	3
Total	6	12

College of Agriculture and Environmental Sciences – Data Summary (Terabytes)

Department	2005 Hosted Data	2005 Managed Data	2010 Hosted Data	2010 Managed Data	2005 Total Data	2010 Total Data
Environmental Science and Policy	NA	NA	TBD	TBD	NA	TBD
LAWR	0.5	11.5	0	65.0	12.0	65.0
Plant Science	0	5.0	0	100.0	5.0	100.0
Total	0.5	16.5	0	165.0	17.0	165.0

College of Engineering – Node Summary

Department	2005 Nodes	2010 Nodes
Applied Science	140	500
BioMedical Engineering	10	64
Chemical Engineering	42	184
Civil Engineering	740	1236
Computer Science	41	60
Electrical and Computer Engineering	4	16
IDAV	16	TBD
Mechanical and Aeronautical Engineering	20	40
Total	1013	2100

College of Engineering – Rack Summary

Department	2005 Racks	2010 Racks
Applied Science	6	20
Biomedical Engineering	1	3
Chemical Engineering	2	5
Civil Engineering	2	4
Computer Science	2	3
Electrical and Computer Engineering	TBD	TBD
IDAV	TBD	TBD
Mechanical and Aeronautical Engineering	1	2
Total	14	37

College of Engineering – Data Summary (Terabytes)

Department	2005 Hosted Data	2005 Managed Data	2010 Hosted Data	2010 Managed Data	2005 Total Data	2010 Total Data
Applied Science	0	6.0	0	20.0	6.0	20.0
Biomedical Engineering	0	1.0	0	5.0	1.0	5.0
Chemical Engineering	0.5	6.0	0	20.0	6.5	20.0
Civil Engineering	0.3	16.3	2.0	32.0	16.5	34.0
Computer Science	30.0	104.8	50.0	2004.8	134.8	2054.8
Electrical and Computer Engineering	?	?	?	?	?	?
IDAV	0	10.0	0	100.0	10.0	100.0
Mechanical and Aeronautical Engineering	0	8.0	0	22.0	8.0	22.0
Total	30.8	152.1	52.0	2203.8	182.9	2255.8

Math and Physical Sciences – Node Summary

Department	2005 Nodes	2010 Nodes
Chemistry	26	140
Geology	2	16
Mathematics	32	34
Physics	443	1204
Total	503	1394

Math and Physical Sciences – Rack Summary

Department	2005 Racks	2010 Racks
Chemistry	2	5
Geology	2	4
Mathematics	1	1
Physics	19	45
Total	24	55

Math and Physical Sciences – Data Summary (Terabytes)

Department	2005 Hosted Data	2005 Managed Data	2010 Hosted Data	2010 Managed Data	2005 Total Data	2010 Total Data
Chemistry	0	10.5	0	41.0	10.5	41.0
Geology	0	1.2	0	100.0	1.2	100.0
Mathematics	0	0	0	2.0	0	2.0
Physics	17.0	66.0	10025.0	322.0	83.0	10347.0
Total	17.0	77.7	10025.0	465.0	94.7	10490.0

Social Sciences – Node Summary

Department	2005 Nodes	2010 Nodes
Center for Mind and Brain	1	6
Political Science	2	2
Total	3	8

Social Sciences – Rack Summary

Department	2005 Racks	2010 Racks
Center for Mind and Brain	3	14
Political Science	1	2
Total	4	16

Social Sciences – Data Summary (Terabytes)

Department	2005 Hosted Data	2005 Managed Data	2010 Hosted Data	2010 Managed Data	2005 Total Data	2010 Total Data
Center for Mind and Brain	3.0	1.0	25.0	20.0	4.0	45.0
Political Science	0	2.0	0	3.0	2.0	3.0
Total	3.0	3.0	25.0	23.0	6.0	48.0

School of Medicine & Other Colleges/Schools

The School of Medicine has indicated that they currently do not have any high performance computing clusters, however, at least CRISP and Radiation Oncology indicate that there is a possible future need.

All other schools and colleges indicated that they did not have a need for high performance computing at the present time.

Total Node Summary

Department	2005 Nodes	2010 Nodes
Biological Sciences	217	1032
College of Agriculture and Environmental Sciences	198	429
College of Engineering	1013	2100
Math and Physical Sciences	503	1394
Social Sciences	3	8
Total	1934	4963

Total Rack Summary

Department	2005 Racks	2010 Racks
Biological Sciences	16	51
College of Agriculture and Environmental Sciences	6	12
College of Engineering	14	37
Math and Physical Sciences	24	55
Social Sciences	4	16
Total	64	171

Total Data Summary (Terabytes)

Department	2005 Hosted Data	2005 Managed Data	2010 Hosted Data	2010 Managed Data	2005 Total Data	2010 Total Data
Biological Sciences	0	8.3	0	110.0	8.3	110.0
College of Agriculture and Environmental Sciences	0.5	16.5	0	165.0	17.0	165.0
College of Engineering	30.8	152.1	52.0	2203.8	182.9	2255.8
Math and Physical Sciences	17.0	77.7	10025.0	465.0	94.7	10490.0
Social Sciences	3.0	3.0	25.0	23.0	6.0	48.0
Total	51.3	257.6	10102.0	2966.8	308.9	13068.8

SECTION 3

SURVEY RESPONSE DETAILS

SECTION 4

SURVEY RESPONSES