InfoBrief
NCSES National Center for Science and Engineering Statistics

July 2013 ■ NSF 13-329

Computing and Networking Capacity Increases at Academic Research Institutions

by Michael Gibbons[1]

Cyberinfrastructure resources at academic research institutions have increased substantially since 2005, according to data from the National Science Foundation's (NSF's) FY 2011 Survey of Science and Engineering Research Facilities. In FY 2011, 59% of academic institutions reported bandwidth of at least 1 gigabit per second (Gbps), compared with 21% of such institutions in FY 2005 (table 1).[2] The percentage of academic institutions with network connections of 10 Gbps or greater increased from 2% to 25% during this period.

Networking

Access to High-Speed Bandwidth

Academic institutions have continued to gain greater access to high-speed bandwidth through a network of national and regional providers. The backbone of the national network is managed primarily by not-for-profit consortia. These include Internet2, an organization established in 1997 that comprises research, academic, industry and government partners, and National LambdaRail, a university-owned organization established in 2003 that manages a 12,000 mile high-speed network. The Energy Sciences Network, a U.S. Department of Energy (DOE)–funded network supporting 40 DOE sites as well as researchers at universities and other research institutions also contributes to the national network.

Greater access to high-speed bandwidth is also facilitated by regional networks and gigapops (gigabit points of presence), which provide connections to the national networks as well as to universities and laboratories. These regional networks also provide advanced

TABLE 1. Bandwidth at academic institutions: FY 2005–12

Bandwidth and number of institutions	FY 2005	FY 2007	FY 2009	FY 2011	FY 2012[a]
All bandwidth (% distribution)	100	100	100	100	100
No bandwidth	0	0	0	0	0
10 Mbps or less	6	3	1	1	*
11 Mbps–100 Mbps	42	33	19	9	8
101 Mbps–999 Mbps	30	31	35	31	27
1 Gbps–2.4 Gbps	15	23	25	28	26
2.5 Gbps–9 Gbps	4	4	5	6	7
10 Gbps	*	2	4	7	10
More than 10 Gbps	2	4	11	18	20
More than 20 Gbps[b]	na	na	na	6	8
Number of institutions	449	448	495	539	538

* = value > 0 but < 0.5%. na = not applicable; category was added to FY 2011 survey.

Gbps = gigabits per second; Mbps = megabits per second.

[a] Figures for 2012 are estimated by responding institutions.
[b] More than 20 Gbps is a subset of more than 10 Gbps.

NOTES: Details may not add to 100% due to rounding. FY 2009, 2011, and 2012 totals includes bandwidth to commodity Internet (Internet1), Internet2 (high-performance hybrid optical packet network), and National LambdaRail (advanced optical network infrastructure for research and education). Data for FY 2005 and FY 2007 are limited to Internet1 and Internet2. The response categories in the FY 2005 survey varied slightly from those in the FY 2007–FY 2011 surveys; in the FY 2005 survey, the categories included "1 to 2.5 Gb" (gigabits) and "2.6 to 9 Gb."

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Survey of Science and Engineering Research Facilities.

network services to ensure reliable and efficient data transfer.

Doctorate-granting institutions constituted 70% of academic institutions reporting at least $1 million in science and engineering research and development in FY 2011. These institutions were more likely than nondoctorate-granting institutions to have higher speed bandwidth capacity. In 2011, the percentage of doctorate-granting institutions with bandwidth of at least 2.5 Gbps (43%) was more than ten times greater than that of nondoctorate-granting institutions (4%) (table 2). This pattern continued in FY 2012, when 53% of doctorate-granting institutions estimated that they would have bandwidth of 2.5 Gbps or greater, compared with 5% of nondoctorate-granting institutions.

Public doctorate-granting institutions were more likely than their private counterparts to have greater access to higher speed bandwidth. Thirty-eight percent of public doctorate-granting institutions had bandwidth access of 10 Gbps or greater in FY 2011, compared

TABLE 2. Total bandwidth at academic institutions, by institutional control and highest degree awarded: FY 2011 and FY 2012 (estimated)

Bandwidth and number of institutions	All institutions	Doctorate			Nondoctorate		
		Total	Public	Private	Total	Public	Private
FY 2011							
All bandwidth (% distribution)	100	100	100	100	100	100	100
10 Mbps or less	1	*	0	1	1	0	3
11 Mbps–45 Mbps	3	1	0	2	10	11	9
46 Mbps–99 Mbps	2	2	2	2	4	2	6
100 Mbps	4	2	2	3	9	7	10
101 Mbps–155 Mbps	4	4	3	7	6	4	8
156 Mbps–622 Mbps	22	16	16	16	35	27	44
623 Mbps–999 Mbps	5	6	4	9	4	4	4
1 Gbps–2.4 Gbps	28	28	27	29	28	43	13
2.5 Gbps–9 Gbps	6	9	8	9	0	0	0
10 Gbps	7	9	11	6	2	0	4
10.1 Gbps–20 Gbps	12	17	20	8	1	1	0
More than 20 Gbps	6	8	7	8	1	1	0
Other	0	0	0	0	0	0	0
Number of institutions	539	379	260	119	160	82	78
FY 2012 (estimated)							
All bandwidth (% distribution)	100	100	100	100	100	100	100
10 Mbps or less	*	*	0	1	0	0	0
11 Mbps–45 Mbps	2	1	0	2	6	5	6
46 Mbps–99 Mbps	2	1	1	0	6	5	8
100 Mbps	4	2	*	4	9	7	12
101 Mbps–155 Mbps	3	3	2	4	4	5	3
156 Mbps–622 Mbps	19	14	13	17	31	21	41
623 Mbps–999 Mbps	5	4	4	5	6	6	5
1 Gbps–2.4 Gbps	26	23	22	28	33	43	22
2.5 Gbps–9 Gbps	7	11	9	14	0	0	0
10 Gbps	10	13	16	6	4	5	4
10.1 Gbps–20 Gbps	12	18	21	10	0	0	0
More than 20 Gbps	8	11	12	9	1	2	0
Other	0	0	0	0	0	0	0
Number of institutions	538	379	260	119	159	81	78

* = value > 0 but < 0.5%.

Gbps = gigabits per second; Mbps = megabits per second.

NOTES: Details may not add to 100% due to rounding. Total includes bandwidth to commodity Internet (Internet), Internet2 (high-performance hybrid optical packet network), and National LambdaRail (advanced optical network infrastructure for research and education).

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Survey of Science and Engineering Research Facilities, FY 2011.

with 22% of their private counterparts. Forty-nine percent of these public institutions expected to have bandwidth access of 10 Gbps or greater by FY 2012, whereas 25% of private doctorate-granting institutions estimated possession of these resources.

Dark Fiber

Dark fiber is fiber-optic cable that has already been laid but is not yet being used. The practice of laying cable in anticipation of future use is a cost-saving approach in network planning. It is more economical to lay excess cable during construction than it is to install it later, when usage demands increase. This cable-in-waiting can be "lit" or engaged quickly. Thus the amount of dark fiber owned by institutions indicates the ability to expand existing network capabilities, either between existing campus buildings or from the campus to an external network.

The percentage of academic institutions with these available cables has increased steadily with each biennial survey cycle. The percentage of institutions with dark fiber to an external network has grown from 29% in FY 2005 to 47% in FY 2011. The percentage of institutions with dark fiber between their own buildings remained high throughout this period, increasing slightly from 86% in FY 2005 to 90% in FY 2011 (table 3).

High-Performance Computing

Resource Management

Many academic research institutions manage their high-performance computing (HPC) resources through a distinct organizational unit within the institution that has a separate staff and budget. A total of 192 of the 539 surveyed academic institutions reported ownership of centrally administered HPC resources of 1 teraflop or faster in FY 2011 (table 4).[3] This administrative approach enables faculty to focus on their primary responsibilities instead of being diverted by administration and fund-raising to support their own HPC resources. Central HPC administration can decrease overall operating expenses and create wider availability of computing resources.[4] However, many HPC resources are supported by external funding sources and managed by the researchers, as opposed to institutional administrators. These resources are difficult to track by administrators and are therefore not included in these data.

Forty-seven percent of doctorate-granting institutions provided HPC resources for their campuses, compared with less than 9% of nondoctorate-granting institutions (table 4). Similar percentages of public doctorate-granting (48%) and private doctorate-granting (45%) institutions provided these resources.

Clusters are the most common centrally administered HPC architecture employed by academic institutions because they provide the most flexibility and cost efficiency for scaling, in addition to their generally lower administrative costs. Ninety-seven percent of HPC-providing institutions employ cluster architectures. HPC-providing institutions also use architectures such as massively parallel processors (11% of institutions), symmetric multiprocessors (19%), or "other" types of architectures (20%), all of which can be used in conjunction with or as an alternative to clusters.[5]

In FY 2011, 24 institutions possessed centrally administered HPC resources with combined computing capacity of at least 100.0 teraflops (figure 1). Another 29 institutions had combined computing capacity of 40.0 to 99.9

TABLE 3. Academic institutions with dark fiber, by type of institution: FY 2005–11
(Percent)

	FY 2005		FY 2007		FY 2009		FY 2011	
Type of institution	To ISP	Between buildings	To ISP	Between buildings	To ISP	Between buildings	To ISP	Between buildings
All institutions	29	86	37	89	39	89	47	90
Doctorate granting	33	88	41	90	44	92	54	93
Nondoctorate granting	19	80	27	87	30	84	30	83
Public	30	87	37	92	43	92	54	94
Private	26	82	36	84	33	84	35	84
Public doctorate granting	33	90	43	93	46	93	60	96
Private doctorate granting	34	86	38	86	40	89	39	87

ISP = internet service provider or external network.

NOTES: Percentages reflect academic institutions that owned dark fiber at the end of the fiscal year. Dark fiber is fiber optic cable that has already been laid but is not being used. Total yearly academic institutional participation as follows: FY 2005, 449; FY 2007, 448; FY 2009, 495; and FY 2011, 539.

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Survey of Science and Engineering Research Facilities.

TABLE 4. Centrally administered high-performance computing in academic institutions, by type of institution and computing architecture: FY 2011

(Number of institutions)

Type of institution	Total	HPC					HPC accelerators
		Total	Clusters	MPP	SMP	Other	
All institutions	539	192	187	21	37	38	101
Doctorate granting	379	178	174	17	33	36	97
Nondoctorate granting	160	14	13	4	4	2	4
Public	342	135	131	16	26	24	77
Private	197	57	56	5	11	14	24
Public doctorate granting	260	125	121	12	24	22	73
Private doctorate granting	119	53	53	5	9	14	24

HPC = high-performance computing; MPP = massively parallel processors; SMP = symmetric multiprocessors.

NOTES: Each institution is counted only once in each architecture. Only HPC systems with peak performance of 1 teraflop or faster are included. Centrally administered HPC is located within a distinct organizational unit with a staff and a budget; unit has stated mission that includes supporting HPC needs of faculty and researchers. Institutions may have HPC of more than one type of architecture. Accelerators may be system components or independent.

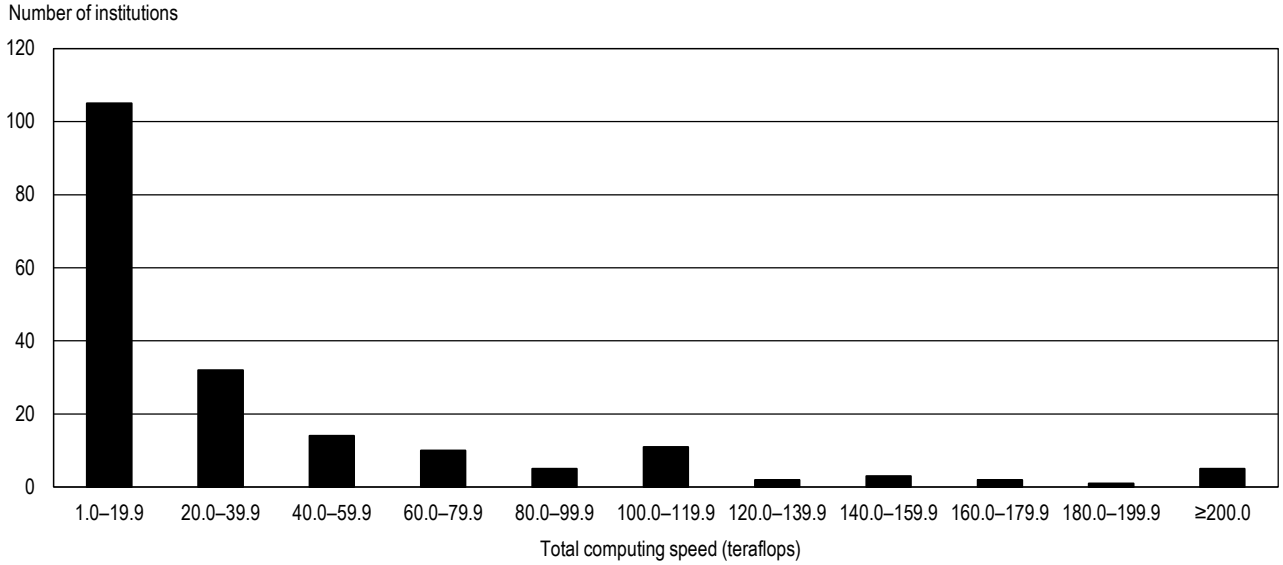
SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Survey of Science and Engineering Research Facilities, FY 2011.

teraflops. As recently as FY 2007, only four institutions reported centrally administered computing capacity of at least 40.0 teraflops, with only one institution surpassing the 100.0 teraflop threshold. The median total performance reported for centrally administered systems was 14 teraflops in FY 2011, compared with 8 teraflops in FY 2009 and 4 teraflops in FY 2007 (data not shown).

Resource Sharing

Colleges and universities often share their HPC resources with external organizations. In FY 2011, these partnerships most often involved other colleges or universities, as 72% of academic institutions shared resources with their peers. Sharing of HPC resources with

FIGURE 1. Total high-performance computing capacity of centrally administered systems: FY 2011



NOTES: Includes systems with massively parallel processors, symmetric multiprocessors, and clusters that have peak performance of 1 teraflop or faster. Institutions may have more than one type of architecture.

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Survey of Science and Engineering Research Facilities, FY 2011.

other external users was less common but still notable for government (21%), industry (18%), and nonprofit organizational partners (17%). Public institutions were more likely to have external users of their HPC resources than were private institutions (data not shown).

Data Storage

As the collection of massive data sets has increased in recent years, data storage and curation have become increasingly critical issues. Data management plans are often required in grant proposals where large data sets will be used. Of the academic institutions with centrally administered HPC resources in FY 2011, 56% reported usable online storage greater than 100 terabytes (table 5).[6] A smaller share of public (21%) and private institutions (18%) provided greater than 500 terabytes of online storage.

As of FY 2011, 45% of institutions with centrally administered HPC resources reported no archival storage. Archival storage includes online and off-line storage for files and data that are not immediately accessible from HPC resources. This figure changed little from FY 2009 (43%), yet it stands much higher than it did in FY 2007 (28%).

Data Sources and Availability

The data presented in this InfoBrief were obtained from NSF's FY 2011 Survey of Science and Engineering Research Facilities. Data from the computing and networking capacity section of the survey were provided by 539 of 554 colleges and universities with at least $1 million in science and engineering research and development expenditures. Institutions were identified as meeting this threshold through the FY 2010 NSF Higher Education Research and Development Survey.

The full set of detailed tables is available in the report *Science and Engineering Research Facilities: Fiscal Year 2011* at http://www.nsf.gov/statistics/facilities/. Please contact the author for more information. This survey has been conducted biennially since 1986. The computing and networking capacity section was added in 2003.

TABLE 5. Storage for centrally administered high-performance computing in academic institutions, by type of institution: FY 2011

Storage and number of institutions	All institutions	Highest degree granted		Control	
		Doctorate	Nondoctorate	Public	Private
Usable online storage					
Terabytes					
Less than 1	4	2	2	2	2
1–5	11	10	1	8	3
6–10	13	10	3	8	5
11–25	19	19	0	12	7
26–50	23	20	3	18	5
51–100	13	12	1	10	3
101–250	45	43	2	33	12
251–500	24	24	0	15	9
501–1,000	18	17	1	18	0
1,001 or more	20	19	1	10	10
Number of institutions	190	176	14	134	56
Archival storage					
Terabytes					
None	85	78	7	62	23
Less than 100	56	51	5	39	17
101–250	12	12	0	8	4
251–500	8	8	0	5	3
501–750	4	4	0	2	2
751–1,000	5	5	0	4	1
1,001–5,000	11	10	1	8	3
5,001–10,000	2	2	0	1	1
10,001 or more	8	7	1	6	2
Number of institutions	191	177	14	135	56

NOTES: A total of 190 of the 192 institutions with centrally administered high-performance computing resources provided data for usable online storage. A total of 191 provided data on archival storage.

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Survey of Science and Engineering Research Facilities, FY 2011.

Notes

1. Michael T. Gibbons, Research and Development Statistics Program, National Center for Science and Engineering Statistics, National Science Foundation, 4201 Wilson Boulevard, Suite 965, Arlington, VA 22230 (mgibbons@nsf.gov; 703-292-4590).

2. The data transfer rate of a computer network is measured in bits per second. One gigabit per second (Gbps) equals one billion bits per second.

3. In FY 2011, 36% of institutions reported centrally administered HPC resources of 1 teraflop or faster. The corresponding totals in preceding cycles were 28% (136 of 494) in FY 2009 and 22% (99 of 449) in FY 2007.

Floating point operations per second (flops) reflect the number of multiplications that a computer processor can perform within 1 second. A teraflop is a measure of computing speed equal to 1 trillion floating point operations per second.

4. These points have been cited as rationales for centralizing cyberinfrastructure and HPC resources at several institutions. See the following documents as examples: The University of Arizona, University Information Technology Services. Program Benefits to Researchers and Campus. Available at http://rc.arizona.edu/node/117.

University of California, San Diego, UCSD Research Cyberinfrastructure Design Team. 2009. Blueprint for the Digital University: UCSD Research Cyberinfrastructure. Available at http://rci.ucsd.edu/_files/Blueprint.pdf.

Bose R, Crosswell A, Hamilton V, Mesa N. 2010. Piloting sustainable HPC for research at Columbia. Position paper for the Workshop on Sustainable Funding and Business Models for Academic Cyberinfrastructure (CI) Facilities. Cornell University, Ithaca: NY.

5. Clusters use multiple commodity systems, each running its own operating system with a high-performance interconnect network to perform as a single system. Massively parallel processors use multiple processors within a single system with a specialized high-performance interconnect network. Each processor uses its own memory and operating system. Symmetric multiprocessors use multiple processors sharing the same memory and operating system to simultaneously work on individual pieces of a program.

6. Online storage includes all storage providing immediate access for files and data from HPC systems of at least 1 teraflop. Storage can be either locally available or made available via a network. Usable storage is the amount of space for data storage that is available for use after the space overhead required by file systems and applicable RAID (redundant array of independent disks) configurations is removed.

NSF 13-329

National Science Foundation
ARLINGTON, VA 22230
OFFICIAL BUSINESS

RETURN THIS COVER SHEET TO ROOM P35 IF YOU DO NOT WISH TO RECEIVE THIS MATERIAL ☐, OR IF CHANGE OF ADDRESS IS NEEDED ☐, INDICATE CHANGE INCLUDING ZIP CODE ON THE LABEL (DO NOT REMOVE LABEL).