Ishfaq Ahmad
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
Hong Kong
iahmad@cs.ust.hk

Cluster computing: A glance at recent events

Computing is an evolutionary process. Five generations of development history—with each generation improving on the previous one's technology, architecture, software, applications, and representative systems—make that clear. As part of this evolution, computing requirements driven by applications have always outpaced the available technology. So, system designers have always needed to seek faster, more cost-effective computer systems.

Parallel and distributed computing provides the best solution, by offering computing power that greatly exceeds the technological limitations of single-processor systems. Unfortunately, although the parallel and distributed computing concept has been with us for over three decades, the high cost of multiprocessor systems has blocked commercial success so far. Today, a wide range of applications are hungry for higher computing power, and even though single-processor PCs and workstations now can provide extremely fast processing, the even faster execution that multiple processors can achieve by working concurrently is still needed.

Now, finally, costs are falling as well. Networked clusters of commodity PCs and workstations using off-the-shelf processors and communication platforms such as Myrinet, Fast Ethernet, and Gigabit Ethernet are becoming increasingly cost effective and popular. This concept, known as cluster computing, will surely continue to flourish: clusters can provide enormous computing power that a pool of users can share or that can be collectively used to solve a single application. In addition, clusters do not incur a very high cost, a factor that led to the sad demise of massively parallel machines (we can hope that they will return, with affordable prices).

Clusters do not incur a very high cost, a factor that led to the sad demise of massively parallel machines.

However, the cluster computing concept also poses three pressing research challenges:

- A cluster should be a single computing resource and provide a single system image. This is in contrast to a distributed system where the nodes serve only as individual resources.
- It must provide scalability by letting the system scale up or down. The scaled-up system should provide more functionality or better performance. The system's total computing power should increase proportionally to the increase in resources. The main motivation for a scalable system is to provide a flexible, cost-effective information-processing tool.
- The supporting operating system and communication mechanism must be efficient enough to remove the performance bottlenecks.

These cluster-computing developments have created a beehive of activity throughout the world—new books, workshops, and conferences, with participation from both academia and industry. The "Forthcoming events" box provides a partial list of these activities. Here we'll take a look at some of the more salient ones.

International Workshop on Cluster Computing

The IEEE International Workshop on Cluster Computing (IWCC '99) held in Melbourne, Australia, last December was quite a success. Besides having one of the most spectacular

Forthcoming events

- International Workshop on Personal Computer-Based Networks of Workstations (PC-NOW 2000), 5 May 2000, Cancun, Mexico—www.disi.unige.it/person/ChiolaG/pcnow00.
- HPCN 2000 Cluster Computing Workshop, 8–10 May 2000, Amsterdam—www.dcs.port.ac.uk/~mab/Workshops/HPCN2000.
- Asia-Pacific International Symposium on Cluster Computing (APSCC 2000), 14–17 May 2000, Beijing—www.dgs.monash.edu.au/~rajkumar/apscc2000/index.html.
- Workshop on Cluster Computing: Technologies, Environments, and Applications (CC-TEA 2000), 26–29 June 2000, Las Vegas, Nev., ceng.usc.edu/~hjin/cc-tea2000.html.
- Workshop on Cluster Computing for Internet Applications (CCIA 2000), 4–7 July 2000, Iwate, Japan—www.cs.nthu.edu.tw/~king/CCIA2000.html.
- EuroPar 2000 Cluster Computing Workshop, 29 Aug.–1 Sept. 2000, Munich—www.dgs.monash.edu.au/~rajkumar/EuroParCluster2000.
- IEEE International Conference on Cluster Computing (Cluster 2000), 28 Nov.–2 Dec., Chemnitz, Germany—www.tu-chemnitz.de/cluster2000.
- TFCC's Third Annual Meeting, Oct. 2001, Los Angeles.

cricket grounds in the world, Melbourne is a diverse and cosmopolitan city, offering a wide range of events and festivals to entertain locals and visitors. The venue was the Centra Melbourne Hotel, on the banks of the Yarra River. The workshop was co-chaired by Rajkumar Buyya of Monash University, Australia, and Mark Baker of the University of Portsmouth, UK. The workshop attracted numerous participants from 14 countries, including the US, Mexico, the UK, Germany, Sweden, Switzerland, France, Hong Kong, China, Singapore, Taiwan, and Korea. The program included keynote talks, invited talks from industry, regular and poster sessions, and a panel. Industrial delegates from a number of leading companies also attended, including Hewlett-Packard, Sun, Compaq, Microsoft, SGI, and MPI Software Technology.

Talks

In his keynote talk, "Fault-Tolerant Cluster Architecture for Business and Scientific Applications," Kai Hwang introduced the Trojan PC/Linux Cluster built at the University of Southern California. To provide a single system image and fault tolerance, the Trojan cluster incorporates a new hierarchical checkpointing algorithm that lets users build large-scale fault-tolerant clusters with a distributed RAID architecture. These clusters support parallel image rendering, video-on-demand scheduling, financial and economic analysis, and parallel gene/DNA sequence matching. Hwang also touched upon several new applications for these clusters, including distributed multimedia processing, intelligent software agents, data mining in e-commerce, and bioinformatics for health care.

In the second keynote talk, "Clustering for Research and Production Scale, Parallel and Distributed Computing," Anthony Skjellum of MPI Software Technology discussed the new generation of software tools that are available for middleware and distributed environments for clusters. Commercial-grade software tools should provide better performance and features than the previous generations of software tools available as open source. Skjellum's talk contrasted the commercial products with open-source products.

In the third keynote talk, "From PC Cluster to a Global Computational Grid," David Abramson of Monash University introduced software tools for resource management and scheduling of geographically distributed computers. Taking a global view, the scheduling system uses various kinds of parameters to determine a scheduling policy for optimally completing an application execution. These parameters include resource architecture and configuration, resource capability, resource requirements, priority, network latency and bandwidth, reliability of resources, contention, user preference, application deadline, and user willingness to pay for resource usage. The scheduler then uses this information and negotiates with the resource owner to get the best value for the money.

In the workshop's plenary talk, Thomas Sterling of the California Institute of Technology and NASA's Jet Propulsion Laboratory presented the cluster system used at JPL. He also gave a broader perspective on cluster computing and talked about current trends leading to very large-scale clusters that can deliver teraflop or even petaflop performance. Building such systems will require SOCs (systems-on-a-chip), gigahertz processor clock rates, VLIW (very long instruction word) architecture, Gbit DRAMs, on-board microdisks, and optical fiber and wave-division multiplexing communication platforms.

Representatives from Compaq, Hewlett-Packard, and Sun Microsystems participated in industry talks at the workshop. Compaq presented its work in scalable supercomputing with reference to the US Department of Energy's Accelerated Strategic Computing Initiative (ASCI) and other projects using Compaq's Alpha processor in small, medium, and large implementations. HP's representatives outlined its cluster interconnect solution that provides low latency, high bandwidth, and low CPU consumption messaging for MPI applications. Their design supports asynchronous messaging, which enables computation and communication overlapping. The cluster interconnect can also perform intrahost transfers, providing data-mover capability in hardware. Sun's participants described Sun Cluster, a clustering solution designed to provide high availability, scalability, and a single-system image. They described such Sun Cluster components as the networking and filing system.

Papers

The workshop's papers spanned a wide range of topics, including

- cluster setup and performance measurements,
- communication software and protocols,
- network communication optimization,
- file system and scheduling,
- metacomputing,

- operating systems and monitoring,
- programming and analysis models, and
- algorithms and applications.

While most of the papers were generally of good quality, some papers reported especially interesting results. The University of Southampton's David Lancaster and Kenji Takeda compared Linux and Windows NT platforms for cluster computing. While the Linux system yielded better communication latency, the NT system sometimes provided better compilation. The paper makes a good case of considering multiple factors in performance comparison rather than emphasizing a single factor such as the network latency.

Jense Mache (Lewis and Clark College) compared the Gigabit Ethernet network as a cluster interconnect with Fast Ethernet, Myrinet, and Scalable Coherent Interface. Using TCP/IP and MPI parallel programming as the basis, he found that Gigabit Ethernet yielded about three times better performance than Fast Ethernet for point-to-point communication. Gigabit Ethernet is also less expensive than the Myrinet and SCI, but delivers end-to-end throughput well above the mainstream 100 Mbps.

Anthony Tam and Cho-Li Wang (University of Hong Kong) presented an interesting new model for abstracting interprocessor communication. Their model exposes various performance characteristics of the system by a set of parameters. Instead of using constant values, their model captures these parameters as some cost functions, which includes the message length, traffic load, and contention factors.

Roy Ho, Kai Hwang, and Hai Jin (University of Southern California) addressed the important issue of I/O by proposing a single-space architecture that provides higher transparency, better performance, lower implementation cost, and higher availability for I/O-intensive applications than existing approaches.

For more information about the workshop, see www.dgs.monash.edu.au/~rajkumar/tfcc/IWCC99. This year's IWCC will merge with two other conferences. Named Cluster 2000, the conference will run from 28 November to 2 December in Chemnitz, Germany.

Task Force on Cluster Computing

Recognizing the trend toward clusters for high-performance computing, the IEEE Computer Society has formed a Task Force on Cluster Computing. The TFCC's objective is to support the advancement of cluster computing research, education, and industry. The TFCC has initiated a number of activities, including open discussions, workshops, and panels for bringing together experts in the field (IWCC '99 was also organized by the TFCC), and making book donations to academic institutions.

The TFCC has established three mirrored Web sites:

- Australia (www.dgs.monash.edu.au/~rajkumar/tfcc/),
- UK (www.dcs.port.ac.uk/~mab/tfcc),
- US (www-unix.mcs.anl.gov/~buyya/tfcc/).

These sites feature updated information on cluster computing, including links to journals, books, freely available software, academic and industrial projects, white papers, and descriptions of available systems. The TFCC can also provide sample curricular material to help academic faculty members introduce new cluster-based courses.

The TFCC maintains three mailing lists for getting up-to-date information:

- tfcc-exe@npac.syr.edu: a list for intra-TFCC executive discussions
- tfcc-adv@npac.syr.edu: a list for intra-TFCC advisory discussions
- tfcc-l@npac.syr.edu: a general and open mailing group for discussion and dissemination of TFCC-related matters

With the advent of the TFCC, interested Computer Society members can participate by sponsoring workshops, conferences, projects, and standards.