The advent of computation can be compared, in terms of the breadth and depth of its impact on research and scholarship, to the invention of writing and the development of modern mathematics.

Ian Foster, 2006
Science Paradigms

- Thousand years ago: science was **empirical**
  describing natural phenomena

- Last few hundred years: a **theoretical** branch
  using models, generalizations

- Last few decades: a **computational** branch
  simulating complex phenomena

- Today: **data exploration** (eScience)
  unify theory, experiment, and simulation
  - Data captured by instruments
  - or generated by simulator
  - Processed by software
  - Information/knowledge stored in computer
  - Scientist analyzes database/files
  using data management and statistics
What is a distributed system?

“A collection of independent computers that appears to its users as a single coherent system”

-A. Tanenbaum
A distributed system organized as middleware. The middleware layer extends over multiple machines, and offers each application the same interface.
Parallel and Distributed Systems

- Tight coupling: Multiprocessors (fast hw network)
- Loose coupling: Multicomputers (slow hw network)

- Shared-Addr.-Space Multiprocessors
  - UMA
  - NUMA
- Dist.-Addr.-Space Multiprocessors

Homogeneous Multicomputers:
- Clusters
- Grids

Heterogeneous Multicomputers:
- Information systems
- Pervasive systems

Information systems
Grids
Pervasive systems

Distributed Systems

CS54: Data-Intensive Computing
Distributed Systems: Clusters, Grids, Clouds, and Supercomputers

[Image of a diagram showing the relationship between different types of computing systems: Supercomputers, Grids, Clouds, and Clusters, categorized by scale and orientation.]
Computer clusters using commodity processors, network interconnects, and operating systems.
Supercomputing ~ HPC

Highly-tuned computer clusters using commodity processors combined with custom network interconnects and customized operating system.
Top Supercomputers from Top500

- Cray XT4 & XT5
  - Cray #1
  - Cielo #18
  - Hopper #19
- IBM BlueGene/L/P/Q
  - Sequia #2
  - Mira #4
  - Juqueen #5
  - Fermi #9
- GPU based
  - Titan #1
  - Tianhe-1A #8
  - Nebulae #12
- SGI Altix ICE
  - Plaiedas #14
- SPARC64 VIIIfx
  - K #3

CS54: Data Intensive Computing
Top500 List - November 2012

$R_{\text{max}}$ and $R_{\text{peak}}$ values are in TFlops. For more details about other fields, check the TOP500 description.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Site</th>
<th>System</th>
<th>Cores</th>
<th>$R_{\text{max}}$ (TFlop/s)</th>
<th>$R_{\text{peak}}$ (TFlop/s)</th>
<th>Power (kW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DOE/SC/Oak Ridge National Laboratory United States</td>
<td>Titan - Cray XK7, Opteron 6274 16C 2.20GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.</td>
<td>560640</td>
<td>17590.0</td>
<td>27112.5</td>
<td>8209</td>
</tr>
<tr>
<td>2</td>
<td>DOE/NNSA/LLNL United States</td>
<td>Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM</td>
<td>1572864</td>
<td>16324.8</td>
<td>20132.7</td>
<td>7890</td>
</tr>
<tr>
<td>3</td>
<td>RIKEN Advanced Institute for Computational Science (AICS) Japan</td>
<td>K computer, SPARC64 VIIfx 2.0GHz, Tofu interconnect Fujitsu</td>
<td>705024</td>
<td>10510.0</td>
<td>11280.4</td>
<td>12600</td>
</tr>
<tr>
<td>4</td>
<td>DOE/SC/Argonne National Laboratory United States</td>
<td>Mira - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM</td>
<td>786432</td>
<td>8162.4</td>
<td>10066.3</td>
<td>3945</td>
</tr>
<tr>
<td>5</td>
<td>Forschungszentrum Juelich (FZJ) Germany</td>
<td>JUQUEEN - BlueGene/Q, Power BQC 16C 1.60GHz, Custom Interconnect IBM</td>
<td>393216</td>
<td>4141.2</td>
<td>5033.2</td>
<td>1970</td>
</tr>
</tbody>
</table>
Grids tend to be composed of multiple clusters, and are typically loosely coupled, heterogeneous, and geographically dispersed.
Major Grids

- XSEDE (Formerly TeraGrid)
  - 200K-cores across 11 institutions and 22 systems over the US
- Open Science Grid (OSG)
  - 43K-cores across 80 institutions over the US
- Enabling Grids for E-sciencE (EGEE)
- LHC Computing Grid from CERN
- Middleware
  - Globus Toolkit
  - Unicore
Cloud Computing

- A large-scale distributed computing paradigm driven by:
  1. economies of scale
  2. virtualization
  3. dynamically-scalable resources
  4. delivered on demand over the Internet
Magellan + DOE's Advanced Network Initiative

NERSC
Sunnyvale

NYC

ALCF
Chicago

OLCF
Nashville

100 gigabit/sec

CS554: Data-Intensive Computing
Major Clouds

- **Industry**
  - Google App Engine
  - Amazon
  - Windows Azure
  - Salesforce

- **Academia/Government**
  - Magellan
  - FutureGrid

- **Opensource middleware**
  - Nimbus
  - Eucalyptus
  - OpenNebula
  - OpenStack
Distributed vs. Single Systems

• Data sharing
  – Multiple users can access common database, data files,…

• Device/resource sharing
  – Printers, servers, CPUs,…

• Communication
  – Communication with other machines…

• Flexibility
  – Spread workload to different & most appropriate machines

• Extensibility
  – Add resources and software as needed
Distributed vs. Centralized Systems

• Economics
  – Microprocessors have better price/performance than mainframes

• Speed
  – Collective power of large number of systems

• Geographic and responsibility distribution

• Reliability
  – One machine’s failure need not bring down the system

• Extensibility
  – Computers and software can be added incrementally
Disadvantages of Distributed Systems

- Software
  - Little software exists compared to PCs
- Networking
  - Still slow and can cause other problems (e.g. when disconnected)
- Security
  - Data may be accessed by unauthorized users
Key Characteristics of Distributed Systems

- Support for resource sharing
- Openness
- Concurrency
- Scalability
- Fault tolerance (reliability)
- Transparency
Resource Sharing

• Share hardware, software, data and information
• Hardware devices
  – Printers, disks, memory, …
• Software sharing
  – Compilers, libraries, toolkits, …
• Data
  – Databases, files, …
• Definition?
• Hardware extensions
  – Adding peripherals, memory, communication interfaces...
• Software extensions
  – Operating systems features
  – Communication protocols
Concurrency

• In a single system several processes are interleaved

• In distributed systems: there are many systems with one or more processors
  – Many users simultaneously invoke commands or applications
  – Many servers processes run concurrently, each responding to different client request
Scalability

• Scale of system
  – Few PCs servers -> dept level systems -> local area networks -> internetworked systems -> wide area networks...
  – Ideally, system and application software should not change as systems scale.

• Scalability depends on all aspects
  – Hardware
  – Software
  – Networks
Fault Tolerance

• Definition?

• Two approaches:
  – Hardware redundancy
  – Software recovery

• In distributed systems:
  – Servers can be replicated
  – Databases may be replicated
  – Software recovery involves the design so that state of permanent data can be recovered
## Transparency in a Distributed System

<table>
<thead>
<tr>
<th>Transparency</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>Hide differences in data representation and how a resource is accessed</td>
</tr>
<tr>
<td>Location</td>
<td>Hide where a resource is located</td>
</tr>
<tr>
<td>Migration</td>
<td>Hide that a resource may move to another location</td>
</tr>
<tr>
<td>Relocation</td>
<td>Hide that a resource may be moved to another location while in use</td>
</tr>
<tr>
<td>Replication</td>
<td>Means that users do not know whether a replica or a master provides a service.</td>
</tr>
<tr>
<td>Concurrency</td>
<td>Hide that a resource may be shared by several competitive users</td>
</tr>
<tr>
<td>Failure</td>
<td>Hide the failure and recovery of a resource</td>
</tr>
<tr>
<td>Persistence</td>
<td>Hide whether a (software) resource is in memory or on disk</td>
</tr>
</tbody>
</table>
False assumptions made by first time developer:

- The network is reliable.
- The network is secure.
- The network is homogeneous.
- The topology does not change.
- Latency is zero.
- Bandwidth is infinite.
- Transport cost is zero.
- There is one administrator.
Questions