HPC-ASIA 2000
The Fourth International Conference/Exhibition on
High Performance Computing in Asia-Pacific Region
May 14-17, 2000
Friendship Hotel, Beijing, China

High Performance Computing, Numerical Libraries, and Trends

Jack Dongarra
University of Tennessee
and
Oak Ridge National Laboratory
http://www.cs.utk.edu/~dongarra/

1

Outline

- Look at trends in HPC
  - Top500 statistics
- Computational Grids
  - Software issues
- NetSolve
  - Example of grid middleware

2

Background Information

- Started in 6/93 by JD, Hans Meuer and Erich Strohmaier

TOP500 Motivation

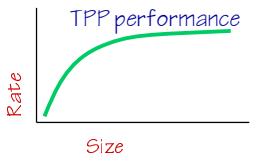
- Basis for analyzing the HCP market
- Quantify observations
- Detection of trends (market, architecture, technology)

3

TOP500 Procedure

- Listing of the 500 most powerful Computers in the World
- Yardstick: Rmax from LINPACK MPP

$Ax=b$, *dense problem*



- Updated twice a year
  SC'xy in the States in November
  Meeting in Mannheim, Germany in June
- All data available from **www.top500.org**

4

TOP-500 List

- A way for tracking trends
  - in performance
  - in market
  - in classes of HPC systems
    - Architecture
    - Technology
- Original classes of machines
  - Sequential
  - SMPs
  - MPPs
  - SIMDs
- Two new classes
  - Beowulf-class systems
  - Clustering of SMPs and DSMs
    - Requires additional terminology
      - "Constellation"

5

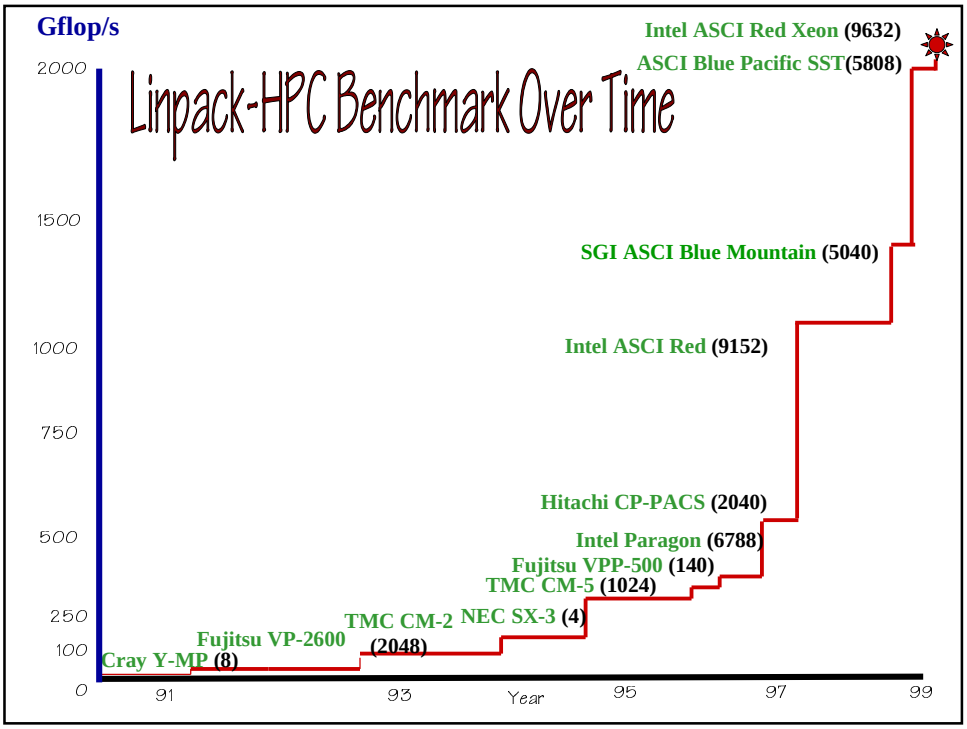
"Constellation"

Cluster of Clusters|DMPs|SMPs

- An ensemble of N nodes each comprising p computing elements
- The p elements are tightly bound shared memory (e.g. smp, dsm)
- The N nodes are loosely coupled, i.e.: distributed memory
- p is greater than N
- Distinction is which layer gives us the most power through parallelism

ASCI Blue Pacific

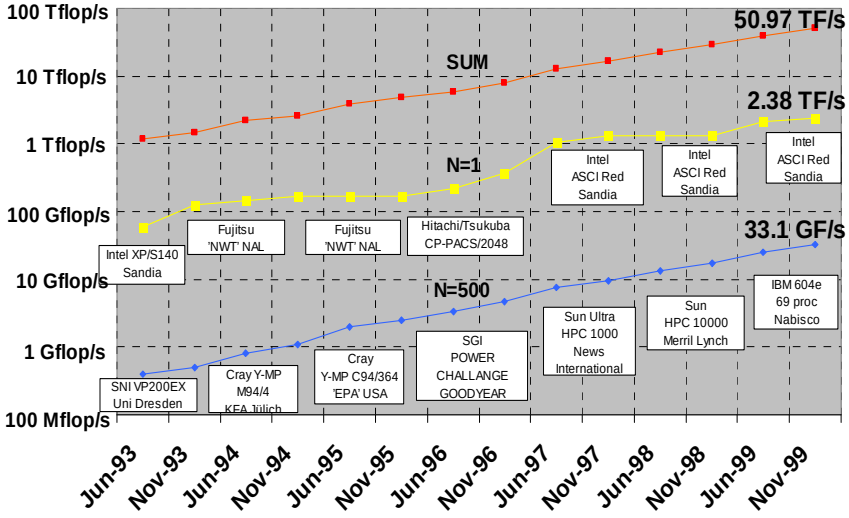
3 x 480 4-way SMP nodes
3.9 TF peak performance
2.6 TB memory
2.5 Tb/s bisectional bandwidth
62 TB disk
6.4 GB/s delivered I/O bandwidth



TOP10 11/99

RANK	MANU-FACTURER	COMPUTER	RMAX [GF/S]	INSTALLATION SITE	COUNTRY	YEAR	AREA OF INSTALLATION	# PROC
1	Intel	ASCI Red	2379.6	Sandia National Labs Albuquerque	USA	1999	Research	9632
2	IBM	ASCI Blue-Pacific SST, IBM SP 604E	2144	Lawrence Livermore National Laboratory	USA	1999	Research	5808
3	SGI	ASCI Blue Mountain	1608	Los Alamos National Lab	USA	1998	Research	6144
4	Cray/SGI	T3E 1200	891.5	Government	USA	1998	Classified	1084
5	Hitachi	SR8000	873.6	University of Tokyo	Japan	1999	Academic	128
6	Cray/SGI	T3E 900	815.1	Government	USA	1997	Classified	1324
7	SGI	Orgin 2000	690.9	Los Alamos National Lab /ACL	USA	1999	Research	2048
8	Cray/SGI	T3E 900	675.7	Naval Oceanographic Office, Bay Saint Louis	USA	1999	Research Weather	1084
9	Cray/SGI	T3E 1200	671.2	Deutscher Wetterdienst	Germany	1999	Research Weather	812
10	IBM	SP Power3	558.13	UCSD/San Diego Supercomputer Center, IBM/Poughkeepsie	USA	1999	Research	1024

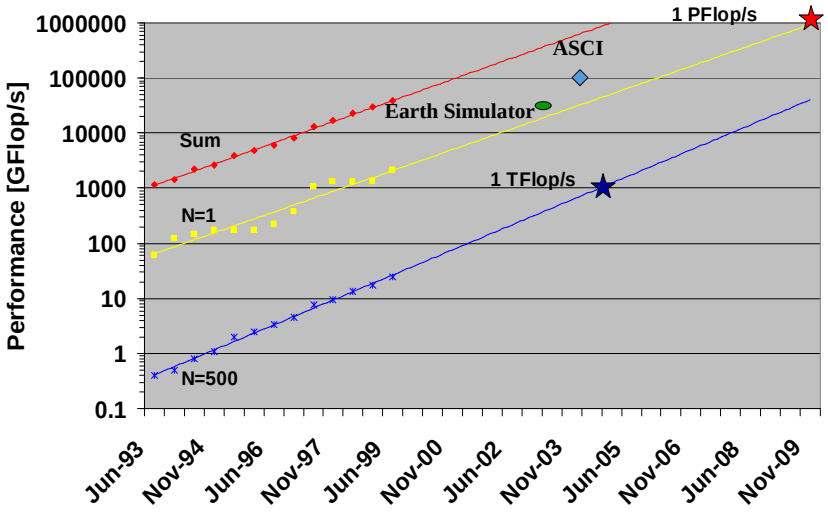
Performance Development



9

[60G - 400 M][2.4 Tflop/s 33 Gflop/s], Schwab #12, 1/2 each year, 100 > 100 Gf, faster than Moore's law, China 419

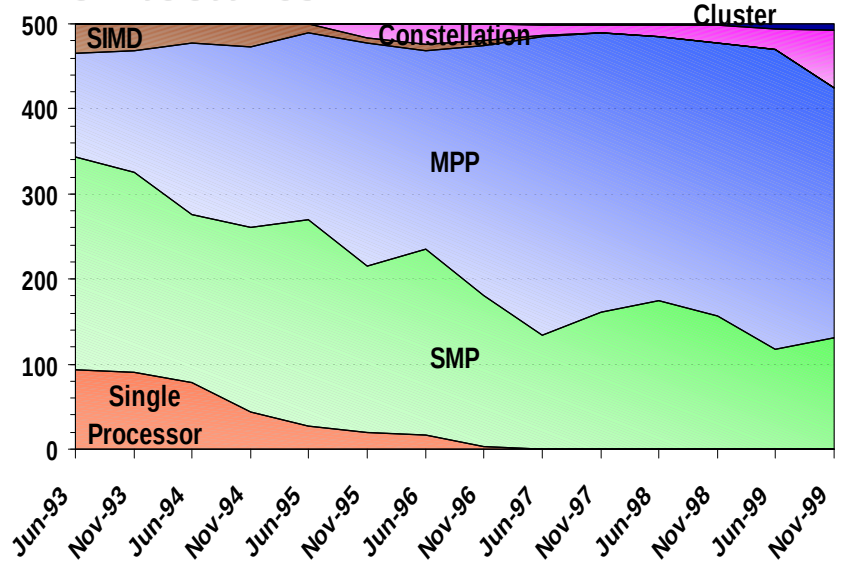
Performance Development



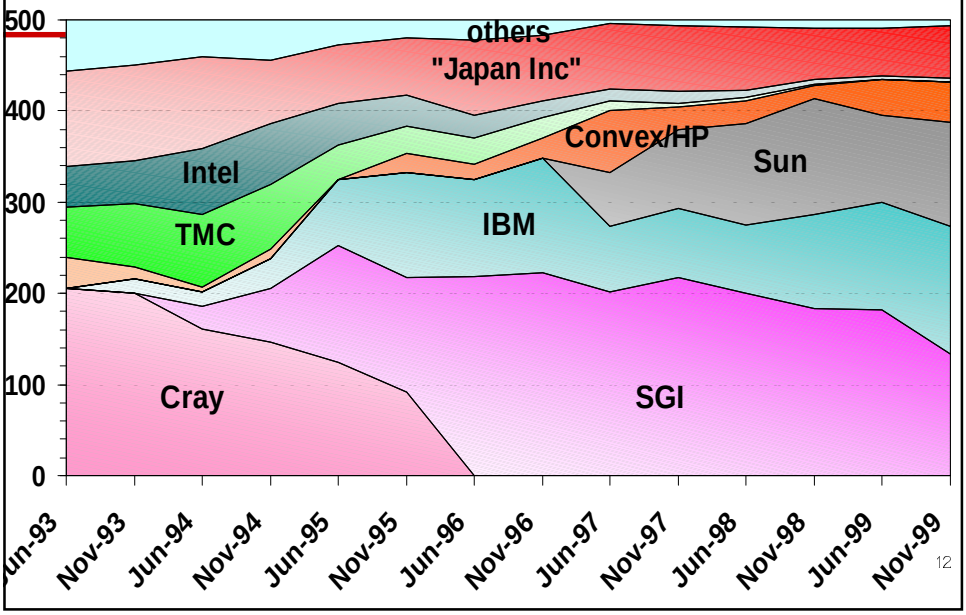
10

Entry 1 T 2005 and 1 P 2010

Architectures

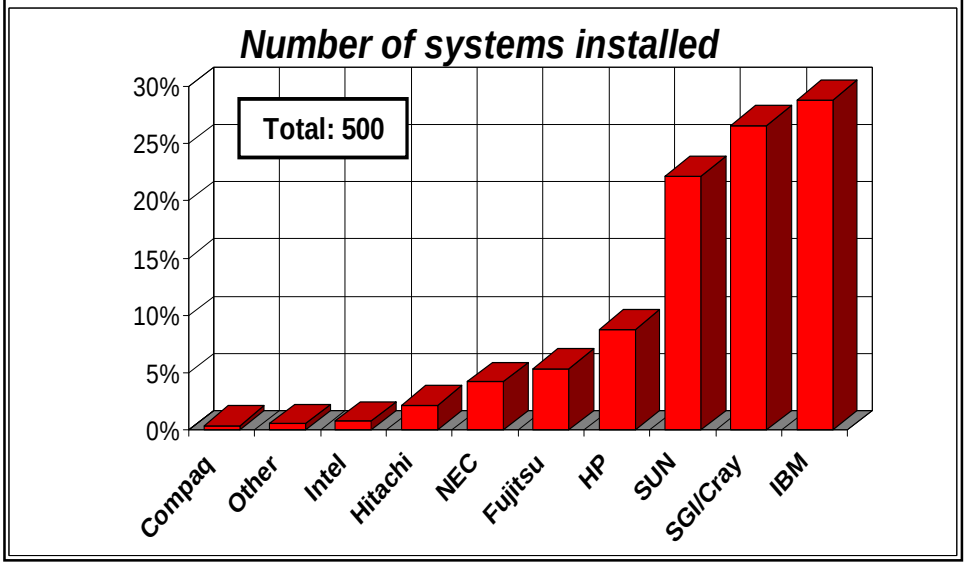


Manufacturer

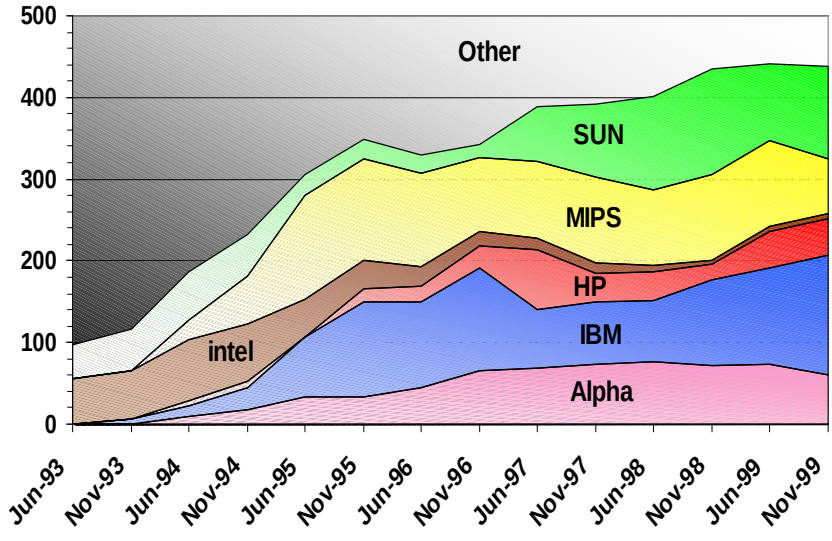


12

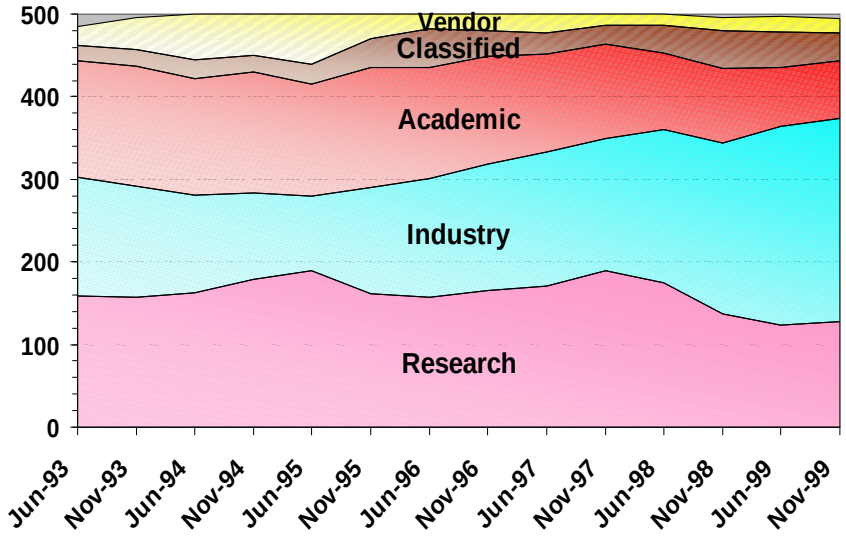
Manufactures November 1999 Report



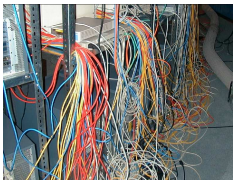
Chip Technology (Data from Top500 Report)



Customer Type



High-Performance Computing Directions



- Clustering of shared memory machines for scalability
  - Emergence of PC commodity systems
    - Pentium/Alpha based, Linux or NT driven
    - "Supercomputer performance at mail-order prices"
  - Beowulf-Class Systems (Linux+PC)
  - Distributed Shared Memory (clusters of processors connected)
  - Shared address space w/deep memory hierarchy
- Efficiency of message passing and data parallel programming
  - Helped by standards efforts such as PVM, MPI, Open-MP and HPF
- Many of the machines as a single user environments
- Pure COTS

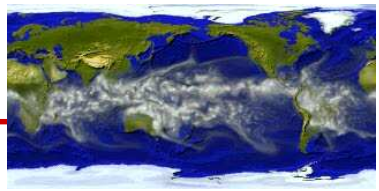
16

Clusters (Beowulf-Class) in the TOP500

RANK	MANU-FACTURER	COMPUTER	RMAX	INSTALLATION SITE	COUNTRY	YEAR	AREA OF INSTALLATION	# PROC
33	Sun	HPC 450 Cluster	272.1	Sun, Burlington	USA	1999	Vendor	720
34	Compaq	Alpha Server SC	271.4	Compaq Computer Corp. Littleton	USA	1999	Vendor	512
…	…	…	…	…	…	…	…	…
44	Self-made	Cplant Cluster	232.6	Sandia National Laboratories	USA	1999	Research	580
…	…	…	…	…	…	…	…	…
169	Self-made	Alphleet Cluster	61.3	Institute of Physical and Chemical Res. (RIKEN)	Japan	1999	Research	140
…	…	…	…	…	…	…	…	…
265	Self-made	Avalon Cluster	48.6	Los Alamos National Lab/ CNLS	USA	1998	Research	140
…	…	…	…	…	…	…	…	…
351	Siemens	hpcLine Cluster	41.45	Universitaet Paderborn/PC2	Germany	1999	Academic	192
…	…	…	…	…	…	…	…	…
454	Self-made	Parnass2 Cluster	34.23	University of Bonn/ Applied Mathematic	Germany	1999	Academic	128

17

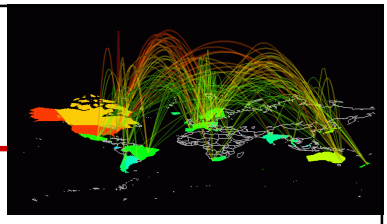
Trends in Computational Science and Engineering



- **Multi-scale, Multi-physics, Multi-dimensional simulations of realistic complexity**
- **Growing use of dynamic adaptive algorithms**
- **Strong interplay between observation and simulation (e.g., cosmology, weather)**
- **Impact of the WWW**
  - **accelerated pace of research due to electronic publishing**
  - **proliferation of digital archives**
  - **emergence of workbenches and portals**

18

Grid Computing



- To treat CPU cycles and software like commodities, an application should be:
  - Ubiquitous -- able to interface to the system at any point and leverage whatever is available
  - Resource Aware -- capable of managing heterogenity
  - Adaptive -- able to tailor its behavior dynamically so that it gets maximum performance benefit from the services and resources at hand

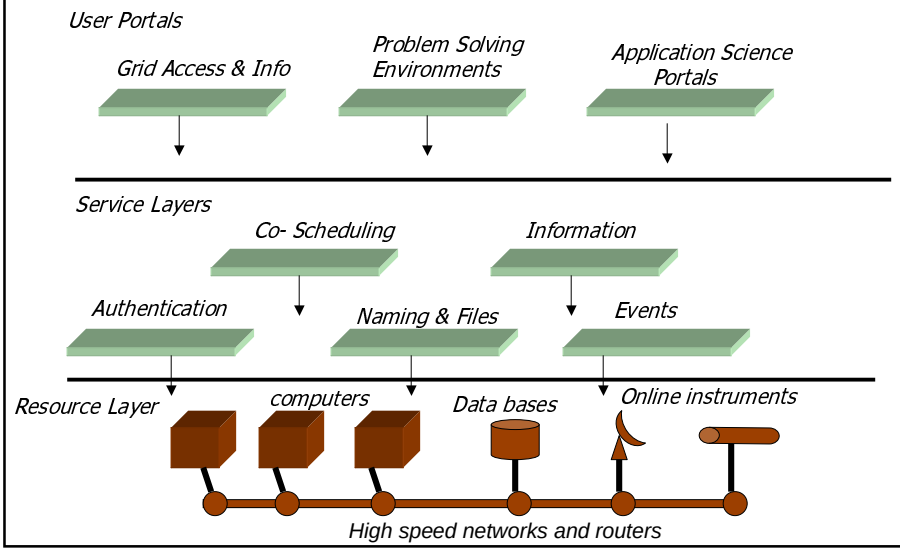
19

What is a Grid Anyway?

- A "Grid" is a set of resources
  - computers, networks, databases, instruments and people
  - all tied together by a common set of ubiquitous distributed services.
- The service layers are designed to provide the *seamless fabric* that allows us to build the problem solving environments (PSE) that the applications scientists want.

20

The Grid Architecture Picture



21

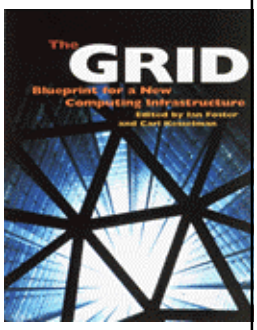
Many Software Issues to Face …

- Predictability and robustness of accuracy and performance.
- Run-time resource management and dynamic algorithm selection.
- Support for a multiplicity of programming environments and plugability.
- Reproducibility, fault tolerant, and auditability of the computations.
- New algorithmic techniques for latency tolerant and miserly bandwidth applications.
- Support for long running computations
- Major challenge adaptability of application and software
  - Resource utilization, failures, new resources

22

Many Active Research Projects

- Akenti
- AppLeS *
- Arcade
- Condor
- CUMULVUS *
- Ecce
- Electronic Notebook *
- EveryWare *
- Globus
- Habanero
- Harness *
- IceT
- IPG NAS-NASA
- JINI
- LoCI *
- Legion
- NCSA Workbench Project
- NEOS
- NetSolve *
- NINF
- Ninja
- NWS *
- PAWS
- POEMS
- PVM *
- Teraweb
- WebFlow



http://www.computingportals.org

* ORNL and/or UTK projects

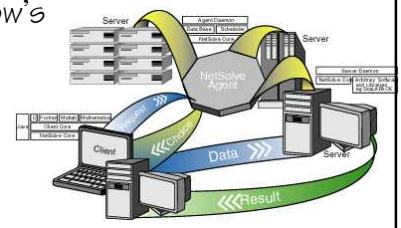
The Changing Face of Numerical Software and Libraries

- Have been involved with open source software
  - Linpack, BLAS, LAPACK, ScaLAPACK, ATLAS
  - PVM, MPI
- Looking for mechanisms that will allow community to use software in an easy to use fashion.
- Allow for portability / heterogeneous / robust
- Grid aware

24

Motivation for NetSolve

Design an easy-to-use tool to provide efficient and uniform access to a variety of scientific packages on UNIX and Window's platforms

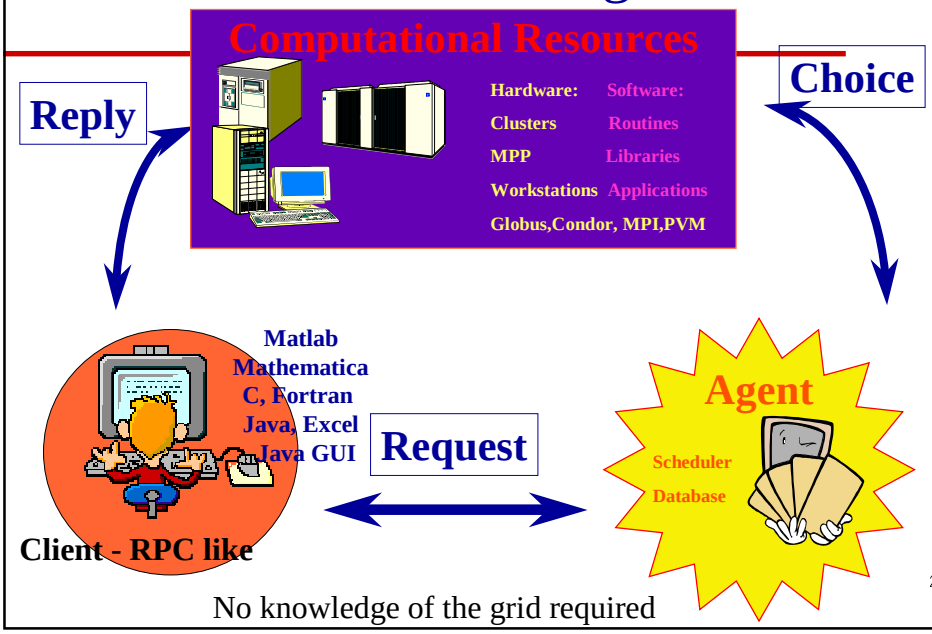


Basics

- Client-Server Design
- Non-hierarchical system
- Load Balancing and Fault Tolerance
- Heterogeneous Environment Supported
- Multiple and simple client interfaces
- Built on standard components

25

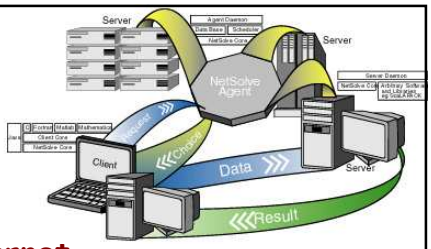
NetSolve - The Big Picture



No knowledge of the grid required

26

NetSolve



- **Three basic scenarios:**
  - Client, servers and agents anywhere on Internet (3(10)-150(80-ws/mpp)- Mcell)
  - Client, servers and agents on an Intranet
  - Client, server and agent on the same machine
- **"Blue Collar" Grid Based Computing**
  - User can set things up, no "su" required
  - Does not require deep knowledge of network programming
- **Focus on Matlab users**
  - OO language, objects are matrices (pse, eg os)
  - One of the most popular desktop systems for numerical computing, 400K Users

27

NetSolve - The Client

- No knowledge of networking involved
- Hide complexity of numerical software
- Computation **location transparency**
- Provides access to **Virtual Libraries** **:**
  - **Component grid-based framework**
  - **Central management of library resources**
  - **User not concerned with most up-to-date version**
  - **Automatic tie to Netlib repository in project**
- Provides synchronous or asynchronous calls
  (User level parallelism)

28

NetSolve - MATLAB Interface

Synchronous Call

```
>> define sparse matrix A
>> define rhs
>> [x, its] = netsolve('itmeth','petsc', A, rhs, 1.e-
...
>> [x, its] = petsc(A, rhs);   % for PETSc
>> [x, its] = aztec(A, rhs);   % for AZTEC
>> [x] = superlu(A, rhs);      % for SuperLU
>> [x] = ma28(A, rhs);         % for MA28
```

Asynchronous Calls also available

NetSolve - FORTRAN Interface

Easy to 'switch' to NetSolve

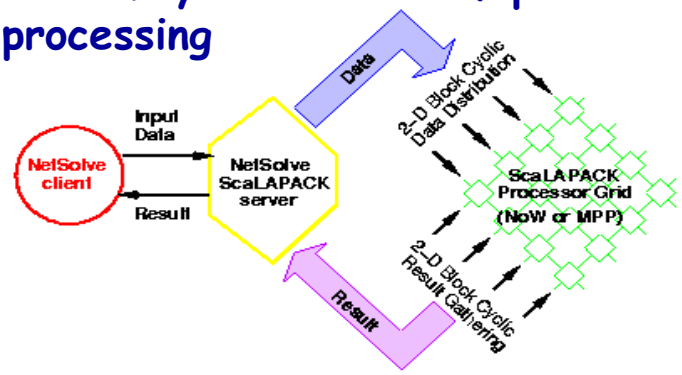
```
parameter( MAX = 100)
double precision A(MAX,MAX), B(MAX)
integer IPIV(MAX), N, INFO, LWORK
integer NSINFO

call DGESV(N,1,A,MAX,IPIV,B,MAX,INFO)


call NETSL('DGESV()',NSINFO,
           N,1,A,MAX,IPIV,B,MAX,INFO)
```

Hiding the Parallel Processing

- **User maybe unaware of parallel processing**



- **NetSolve takes care of the starting the message passing system, data distribution, and returning the results.**

31

NetSolve - The Server Side

Computational Server :

- Various **Software** resources installed on various **Hardware** Resources
- **Configurable and Extensible :**
  - Framework to easily add arbitrary software ...
  - Many numerical libraries being integrated by the NetSolve team
  - Many software being integrated by users

Agent :

- Gateway to the computational servers
- Performs **Load Balancing** among the resources

http://www.cs.utk.edu/netsolve

NetSolve - Load Balancing

NetSolve agent :

predicts the execution times and sorts the servers

Prediction for a server based on :

- **Its distance over the network**
  - **Latency and Bandwidth**
  - **Statistical Averaging**
- **Its performance (LINPACK benchmark)**
- **Its workload**
- **The problem size and the algorithm complexity**

Cached data → **Quick estimate**

workload out of date ?

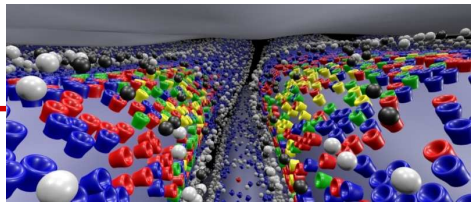
http://www.cs.utk.edu/netsolve

NetSolve Applications and Interactions

- Tool integration
  - Globus - Middleware infrastructure (ANL/SSI)
  - Condor - Workstation farm (U Wisconsin)
  - NWS - Network Weather Service (U Tennessee)
  - SCIRun - Computational steering (U Utah)
  - Ninf - NetSolve-like system, (ETL,Tsukuba)
- Library usage
  - LAPACK/ScaLAPACK - Parallel dense linear solvers
  - SuperLU/MA28 - Parallel sparse direct linear solvers(UCB/RAL)
  - PETSc/Aztec - Parallel iterative solvers (ANL/SNL)
  - Other areas as well (not just linear algebra)
- Applications
  - MCell - Microcellular physiology (UCSD/Salk)
  - IPARS - Reservoir Simulator (UTexas, Austin)
  - Virtual Human - Pulmonary System Model (ORNL)
  - RSICC - Radiation Safety sw/simulation (ORNL)
  - LUCAS - Land usage modeling (U Tennessee)
  - ImageVision - Computer Graphics and Vision (Graz U)

34

Other Features



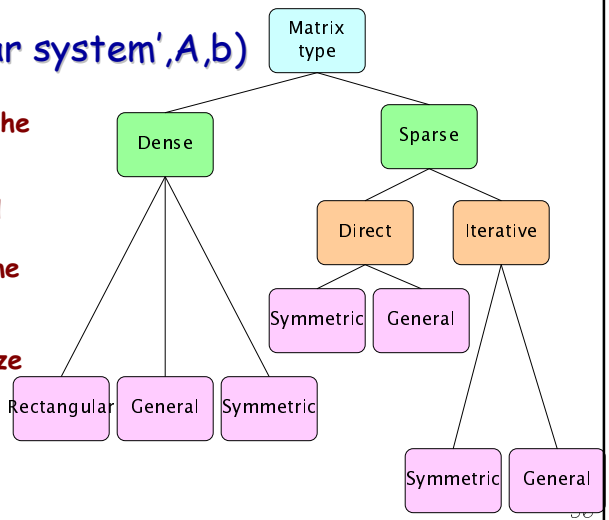
- Task Farming
  - Multiple requests to single problem.
- Authentication with Kerberos
  - Used to maintain Access Control Lists and manage access to computational resources.
- Globus interface
- Network Weather Service
  - Forecast network capabilities
- Data Persistence
  - Between calls to NetSolve data resides on server.

NETWORK WEATHER SERVICE

35

NetSolve to Determine the Best Algorithm/Software (A bit in the future)

- x = netsolve('linear system',A,b)
  - NetSolve examines the user's data together with the resources available (in the grid sense) and makes decisions on best time to solution dynamically.
  - Decision based on size of problem, hardware/software, network connection etc.
  - Method and Placement.



NetSolve: Features for the Future

- Software repositories
- Data encryption/compression mechanisms
- User authentication/security
- Access control or user accounting mechanisms (tokens, credits,..)
- Computational steering
- Task migration and improved Fault-tolerance
- Integration into other meta-computing infrastructure, Legion
- Port Agent and Server Windows NT/98/95 (client already there)

Software available see:
http://www.netlib.org/netsolve/ (> 13K hits)

Conclusion

- Exciting time to be in scientific computing
- Network computing is here
- The Grid offers tremendous opportunities for collaboration
- Important to develop algorithms and software that will work effectively in this environment

38

Contributors to These Ideas

- Top500
  - Erich Strohmaier, UTK
  - Hans Meuer, Mannheim U
- NetSolve
  - Dorian Arnold, UTK
  - Susan Blackford, UTK
  - Henri Casanova, UCSD
  - Michelle Miller, UTK
  - Ganapathy Raman, UTK
  - Sathish Vadhiyar, UTK
  - Thara Angskun, Kasetsart U/UTK

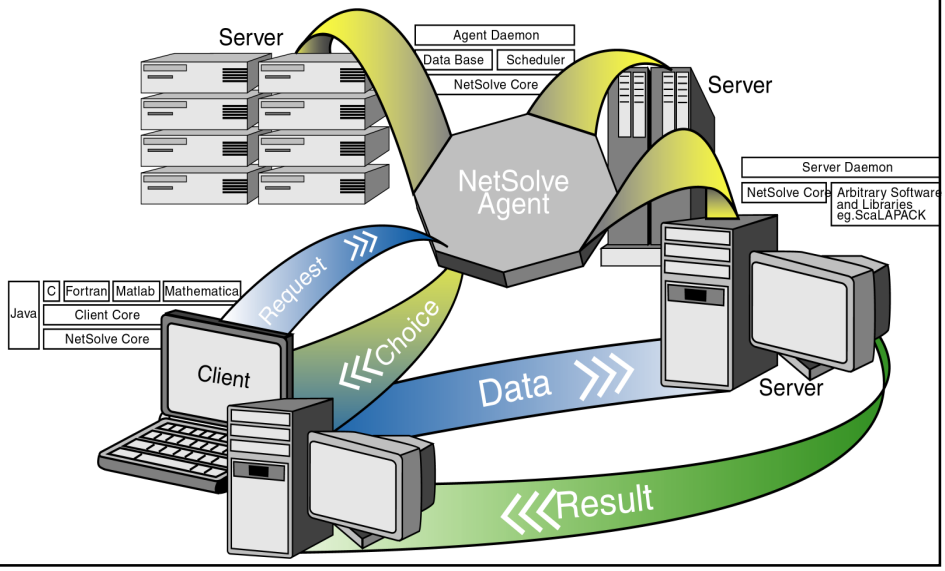
For additional information see…

www.netlib.org/top500/

www.netlib.org/netsolve/

www.cs.utk.edu/~dongarra/

Many opportunities within group

39

http://www.cs.utk.edu/netsolve/