Self Adapting Numerical Software and Update on NetSolve

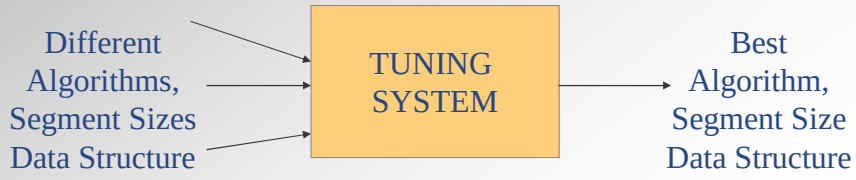
J.J. Dongarra
University of Tennessee

Outline

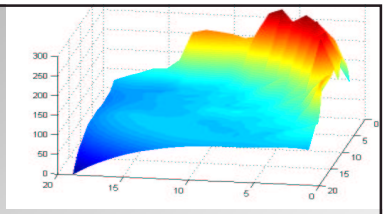
- Overview of the SANS-Effort
  - Review of ATLAS
    - Current status
  - Iterative Method Optimization
    - BiCG
  - LAPACK For Clusters (LFC)
    - Middleware for ease of use on clusters
- Review of NetSolve
  - Grid enabled numerical libraries

Self-Adapting Numerical Software (SANS) Effort

- The complexities of modern processors or clusters makes it difficult to achieve high performance
- Hardware, compilers, and software have a large design space w/many parameters
- Kernels of Computation Routines
  - Focus on where the most time is spent
- Need for quick/dynamic deployment of optimized routines.
- Algorithm layout and implementation
  - Look at the different ways to express implementation



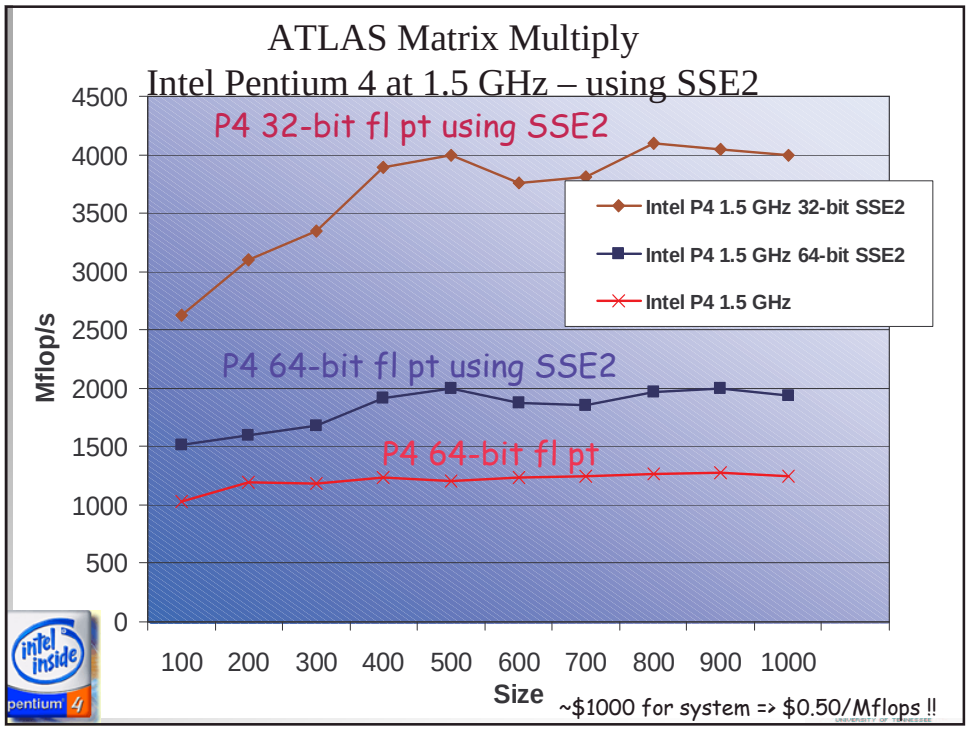
Software Generation Strategy - ATLAS BLAS



- Parameter study of the hw
- Generate multiple versions of code, w/difference values of key performance parameters
- Run and measure the performance for various versions
- Pick best and generate library
- Level 1 cache multiply optimizes for:
  - TLB access
  - L1 cache reuse
  - FP unit usage
  - Memory fetch
  - Register reuse
  - Loop overhead minimization

- Takes ~ 20 minutes to run, generates Level 1,2, & 3 BLAS
- "New" model of high performance programming where critical code is machine generated using parameter optimization.
- Designed for RISC arch
  - Super Scalar
  - Need reasonable C compiler
- Today ATLAS in used within various ASCI and SciDAC activities and by Matlab, Mathematica, Octave, Maple, Debian, Scyld Beowulf, SuSE,…

Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE



Solving Large Sparse Non-Symmetric Systems of Linear Equations Using BiCG-Stab

Compute $r^{(0)} = b - Ax^{(0)}$ for some initial guess $x^{(0)}$.
Choose $\hat{r}^{(0)}$ (for example, $\hat{r}^{(0)} = r^{(0)}$).
for $i = 1, 2, \ldots$
  solve $M z^{(i-1)} = r^{(i-1)}$
  solve $M^T \hat{z}^{(i-1)} = \hat{r}^{(i-1)}$
  $\rho_{i-1} = z^{(i-1)T} \hat{r}^{(i-1)}$
  if $\rho_{i-1} = 0$, method fails
  if $i = 1$
    $p^{(i)} = z^{(i-1)}$
    $\hat{p}^{(i)} = \hat{z}^{(i-1)}$
  else
    $\beta_{i-1} = \rho_{i-1} / \rho_{i-2}$
    $p^{(i)} = z^{(i-1)} + \beta_{i-1} p^{(i-1)}$
    $\hat{p}^{(i)} = \hat{z}^{(i-1)} + \beta_{i-1} \hat{p}^{(i-1)}$
  endif
  $q^{(i)} = A p^{(i)}$
  $\hat{q}^{(i)} = A^T \hat{p}^{(i)}$
  $\alpha_i = \rho_{i-1} / \hat{p}^{(i)T} q^{(i)}$
  $x^{(i)} = x^{(i-1)} + \alpha_i p^{(i)}$
  $r^{(i)} = r^{(i-1)} - \alpha_i q^{(i)}$
  $\hat{r}^{(i)} = \hat{r}^{(i-1)} - \alpha_i \hat{q}^{(i)}$
  check convergence; continue if necessary
end

- Example of optimizations
  - Combining 2 vector ops into 1 loop
  - Simplifying indexing
  - Removal of "if test" within loop

Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE

Optimization of BiCG-Stab
10% - 20% Improvement

Matrix	PIII no	PIII Opt	P4 no	P4 Opt	Itanium no	Itanium Opt
jpwh_991	0.015	0.01	0.007	0.007	0.031	0.026
memplus	12.004	8.47	2.546	1.976	13.606	12.334
orsreg_1	0.574	0.525	0.132	0.125	0.561	0.479
psmigr_1	5.297	4.9	0.911	0.94	8.181	7.78
raefsky3	237.819	154.077	43.005	31.187	297.044	283.738
saylr4	11.388	12.126	2.478	2.708	10.272	8.754
sherman3	1.992	1.991	0.473	0.469	1.677	1.439
sherman5	0.333	0.366	0.083	0.085	0.3	0.257
venkat01	19.877	15.536	3.407	3.306	28.24	26.891
wang3	6.634	6.057	1.085	1.243	7.698	7.195
wang4	6.781	5.246	1.19	1.123	7.684	6.991

Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE

Split ADI Method

- Originally from a radiation transport problem
- Solution of medium-size dense blocks
- Well-conditioned => iterative method

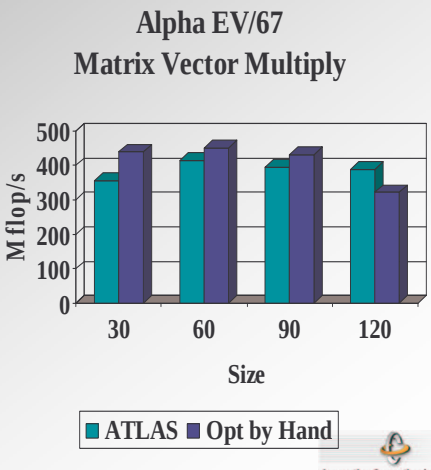
Write $A=D-E=D^{-1}(I-N)$; $M^{-1}=(I+N)D^{-1}$
Left-preconditioning: $M^{-1}A = (I-N^2)$
D^{-1} is cheap, question is how efficient is
$y = N^2 x$;
Can we beat 2 matrix-vector operations?

8

Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE

Performance Of L1Cache A^2x Kernel

- Avoiding 2 calls to MV routine
- Do blocking to capture peak
- 10% improvement by simple optimization



9

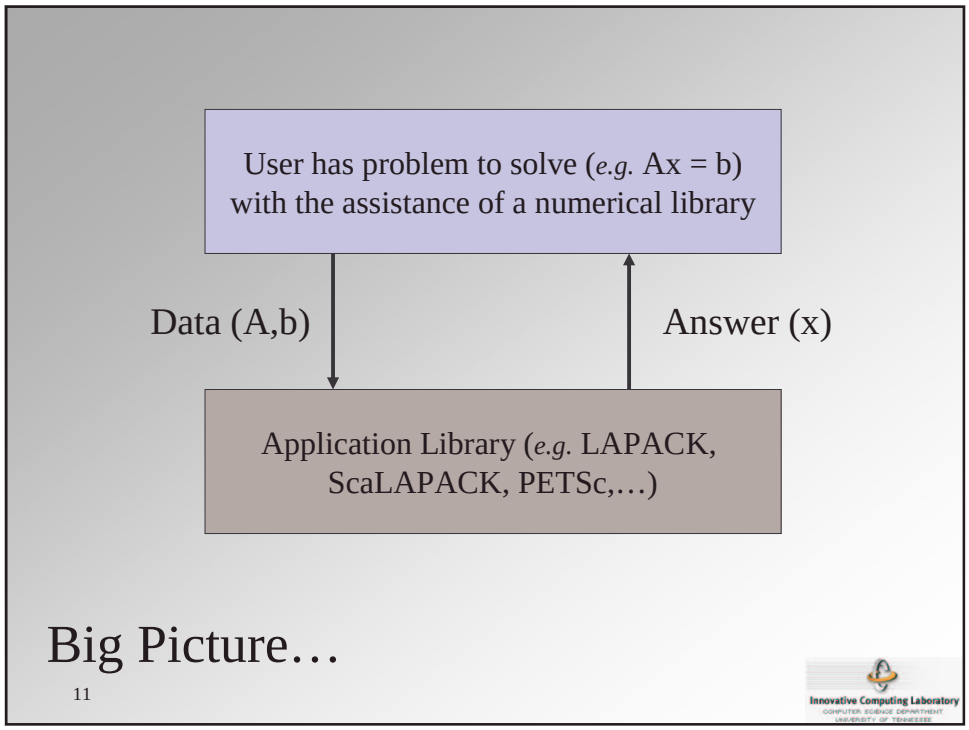
Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE

LAPACK For Clusters

- Idea to make it easy to use your cluster to solve dense matrix problems.
- As simple as a conventional call to LAPACK
- Make decisions on which machines to use based on the user's problem and the state of the system
  - Determinate machines that can be used
  - Optimize for the best time to solution
  - Distribute the data on the processors and collections of results
  - Start the SPMD library routine on all the platforms

10

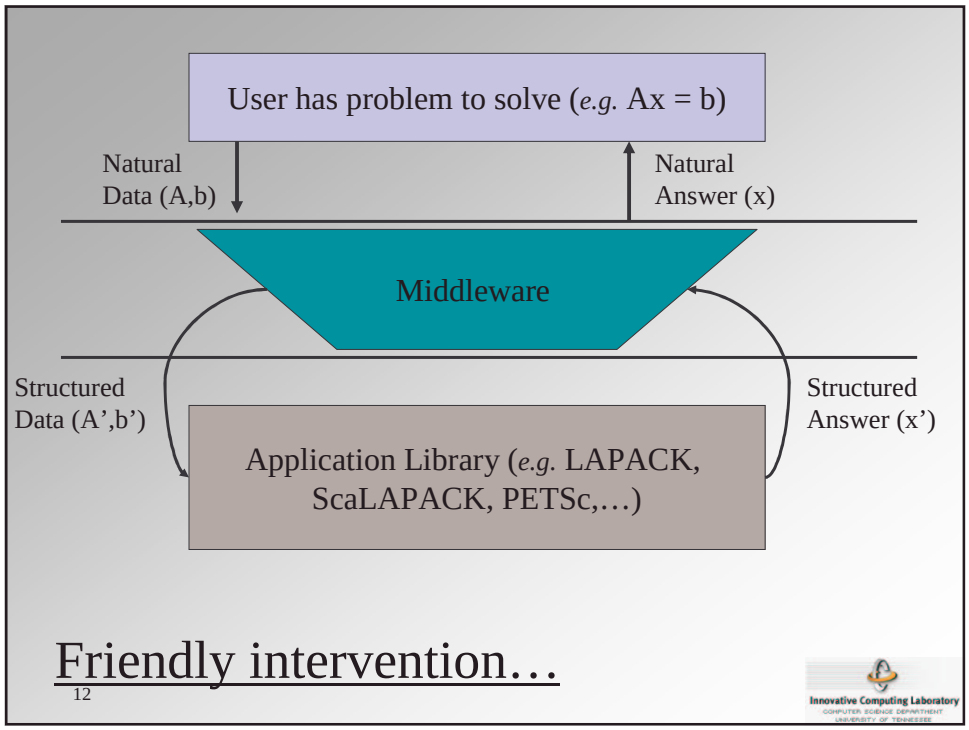
Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE



Big Picture…

11

Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE



Friendly intervention…

12

Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE

LAPACK For Clusters - Process

- User stages data to disk
- User calls LFC routine

- LFC determines resources and properties
- LFC runs performance model to select "best" set of resources
- LFC loads data to subset of processors
- LFC invokes ScaLAPACK

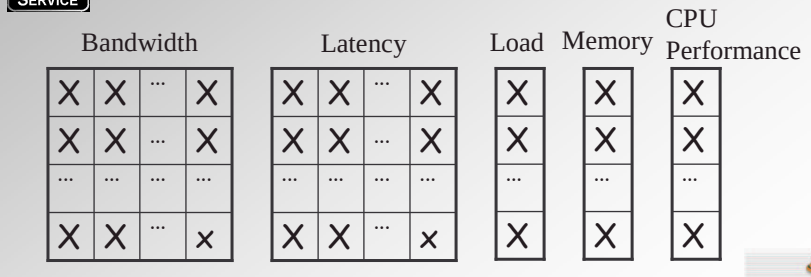
13

Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE

Resource Selector

- Uses NWS to build an array of values for the machines that are available for the user.
  - 2 matrices (bw,lat) 3 arrays (load, cpu, memory available)
- Generated dynamically by library routine

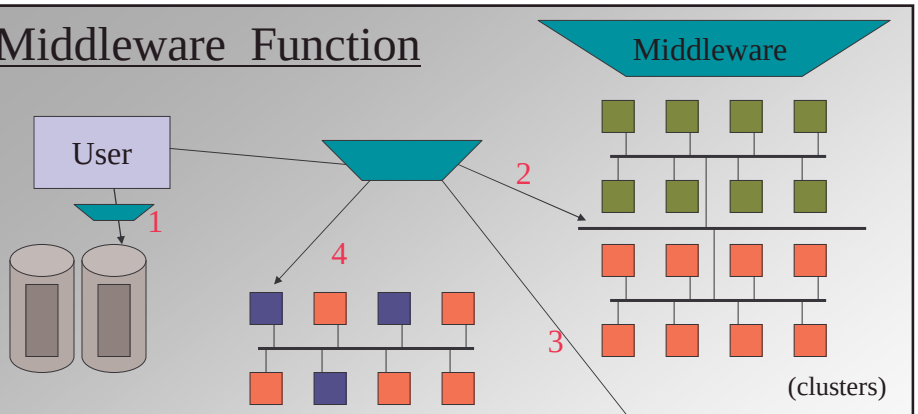
NETWORK WEATHER SERVICE



14

Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE

Middleware Function

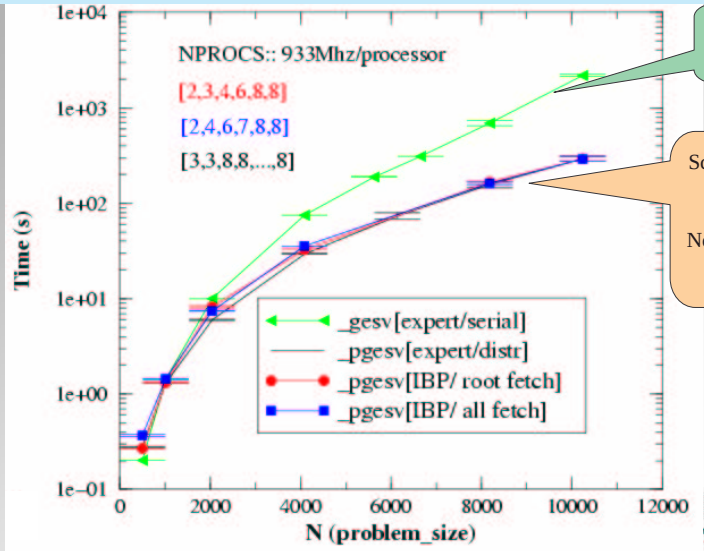


1. User invokes help of middleware to generate data set in remote memory depot, IBP
2. Middleware makes NWS query of available clusters obtaining current, relevant system/network information
3. Resources are selected based on the application specific performance model and information from step 2.

- The data handle (IBP *capability*) is passed over the network to the lead compute node and the solve invoked (here the structured mapping of the data is imparted in the application routine)

Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE

Pentium III 933 MHz
Ax = b
LAPACK / ScaLAPACK / LFC

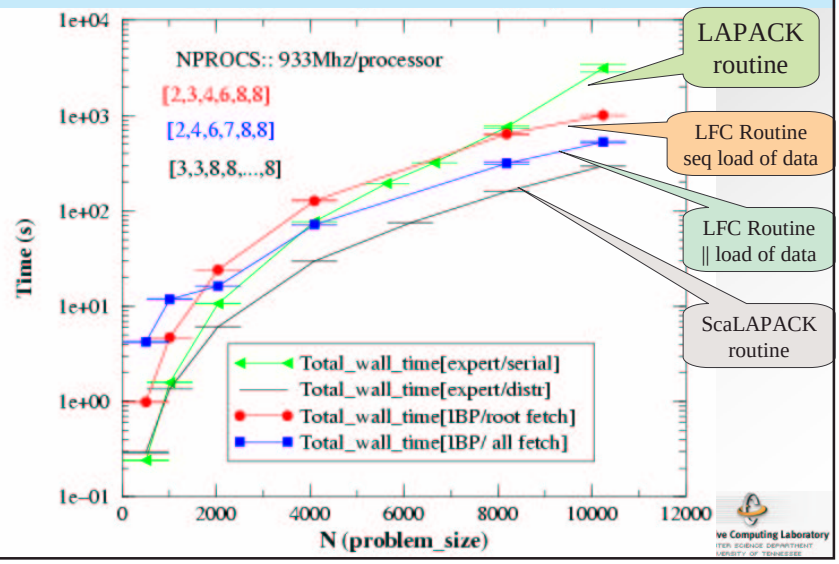


LAPACK routine

ScaLAPACK and LFC Routines No loading of data measured

16

Pentium III 933 MHz
Ax = b
LAPACK/ScaLAPACK/LFC



17

LAPACK For Clusters (LFC)

- LFC will automate much of the decisions in the Cluster environment to provide best time to solution.
  - Adaptivity to the dynamic environment.
  - As the complexities of the Clusters and Grid increase need to develop strategies for self adaptability.
  - Handcrafted developed leading to an automated design.
- Developing a basic infrastructure for computational science applications and software in the Cluster and Grid environment.
  - Lack of tools is hampering development today.
- Plan to do LU, Cholesky, QR, Symmetric eigenvalue, and Nonsymmetric eigenvalue

18

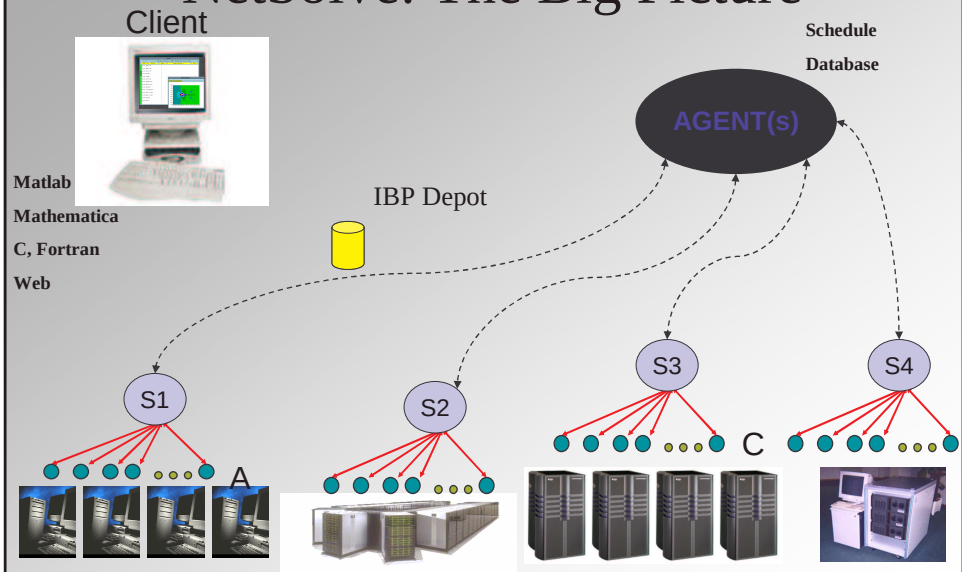
NetSolve - Grid Enabled Server

- NetSolve is an example of a Grid based hardware/software server.
- Based on a Remote Procedure Call model but with …
  - resource discovery, dynamic problem solving capabilities, load balancing, fault tolerance asynchronicity, security, …
- Easy-of-use paramount
- Other examples are NEOS from Argonne and NINF Japan.

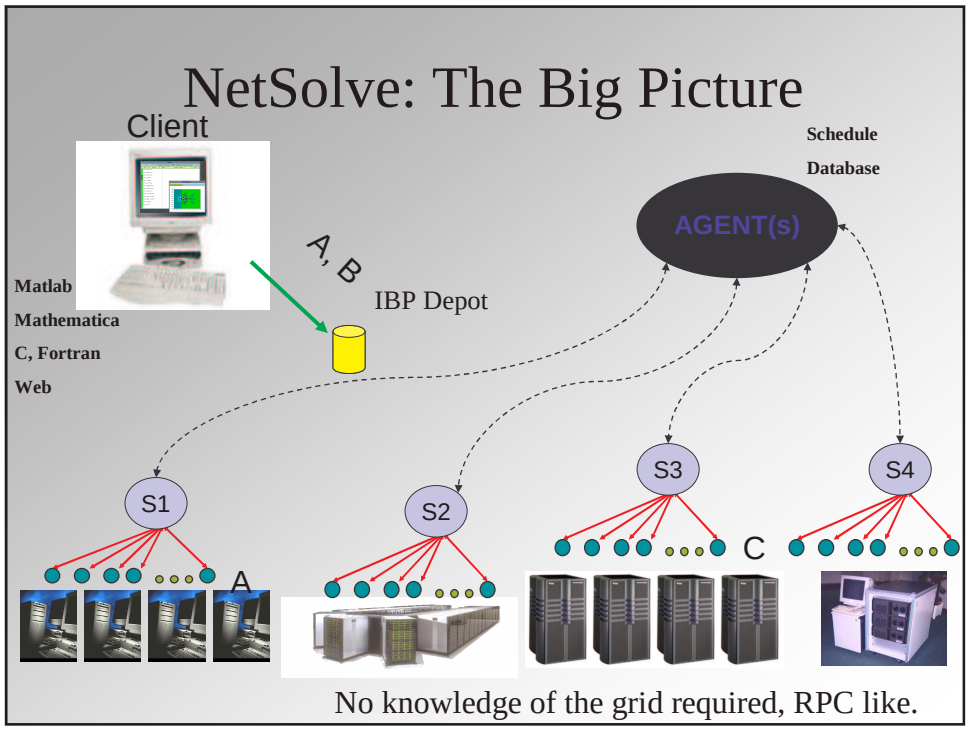
19

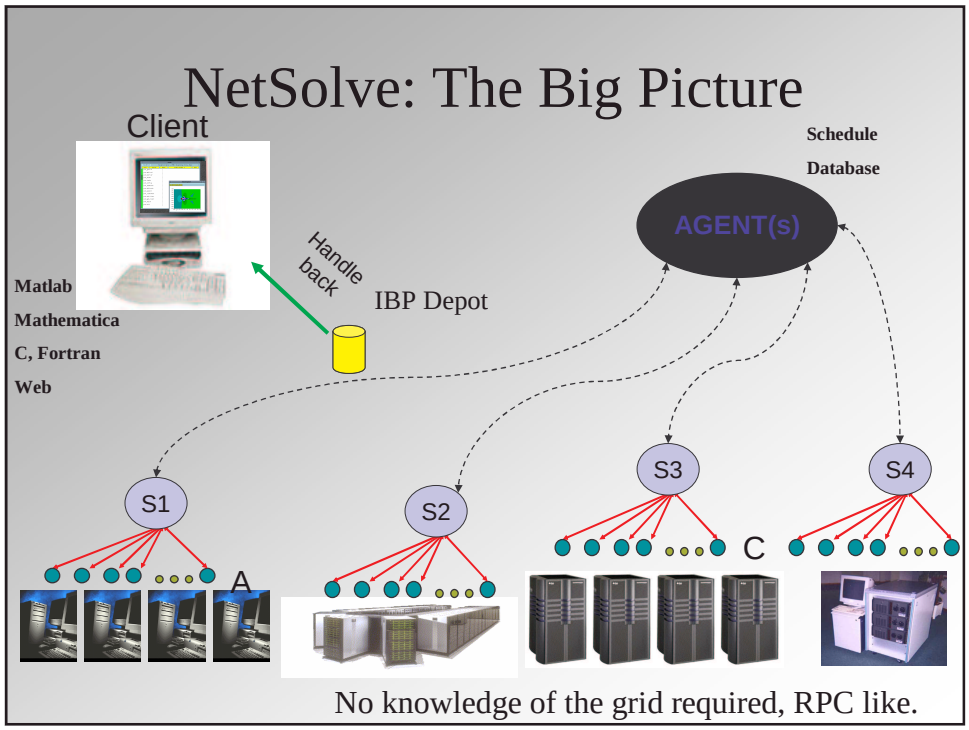
Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE

NetSolve: The Big Picture

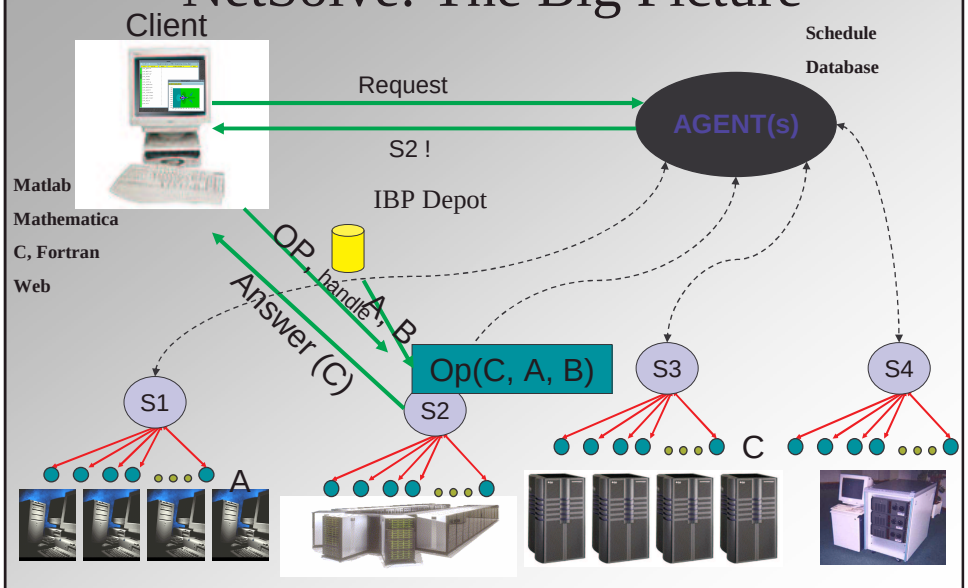


No knowledge of the grid required, RPC like.


NetSolve: The Big Picture
Client
Schedule
Database
AGENT(s)
A, B
IBP Depot
Matlab
Mathematica
C, Fortran
Web
S1
S2
S3
S4
A
C
No knowledge of the grid required, RPC like.

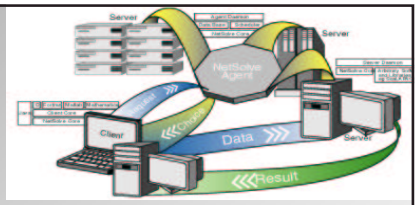

NetSolve: The Big Picture
Client
Schedule
Database
AGENT(s)
Handle
back
IBP Depot
Matlab
Mathematica
C, Fortran
Web
S1
S2
S3
S4
A
C
No knowledge of the grid required, RPC like.

NetSolve: The Big Picture



No knowledge of the grid required, RPC like.

Basic Usage Scenarios

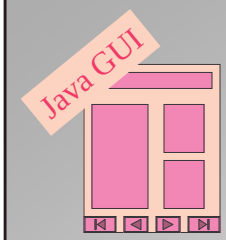


- Grid based numerical library routines
  - User doesn't have to have software library on their machine, LAPACK, SuperLU, ScaLAPACK, PETSc, AZTEC, ARPACK
- Task farming applications
  - "Pleasantly parallel" execution
  - eg Parameter studies
- Remote application execution
  - Complete applications with user specifying input parameters and receiving output
- "Blue Collar" Grid Based Computing
  - Does not require deep knowledge of network programming
  - Level of expressiveness right for many users
  - User can set things up, no "su" required
  - In use today, up to 200 servers in 9 countries
- Can plug into Globus, Condor, NINF, …

24

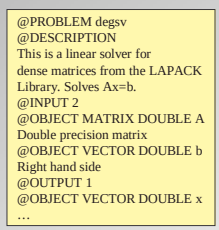
Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE

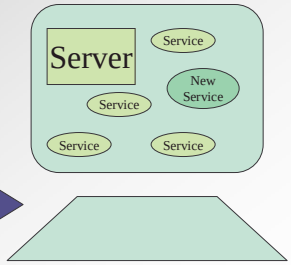
Generating New Services in NetSolve



- Add additional functionality
  - Describe the interface (arguments)
  - Generate wrapper
  - Install into server

New Service Added!





25

Task Farming - Multiple Requests To Single Problem Used Early on with MCell

- A Solution:
  - Many calls to netslnb( ); /* non-blocking */
- Farming Solution:
  - Single call to netsl_farm( );
- Request iterates over an "array of input parameters."
- Adaptive scheduling algorithm.
- Useful for parameter sweeping, and independently parallel applications.

26

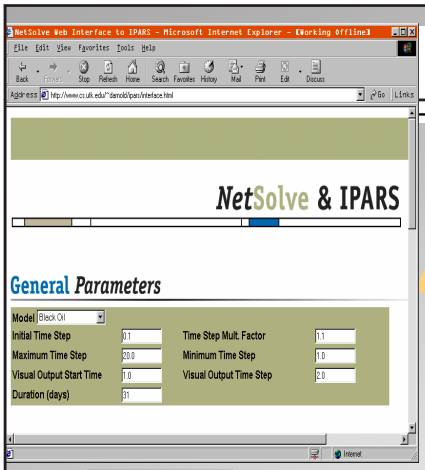
IPARS

- Integrated Parallel Accurate Reservoir Simulator.
- TICAM of UT, Austin, Director, Dr. Mary Wheeler.
- Portable and Modular reservoir simulator.
- Models waterflood, black oil, compositional, well management, recovery process …
- Reservoir and Environmental Simulation.
  - models black oil, waterflood, compositions
  - 3D transient flow of multiple phase
- Integrates Existing Simulators.
- Framework simplified development
  - Provides solvers, handling for wells, table lookup.
  - Provides pre/postprocessor, visualization.
- Full IPARS access without Installation.
- IPARS Interfaces:
  - C, FORTRAN, Matlab, Mathematica, and Web.

27

Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE

NetSolve & IPARS

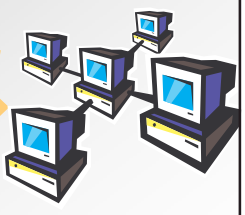


Web Interface

Web Server

NetSolve Client

IPARS-enabled Servers

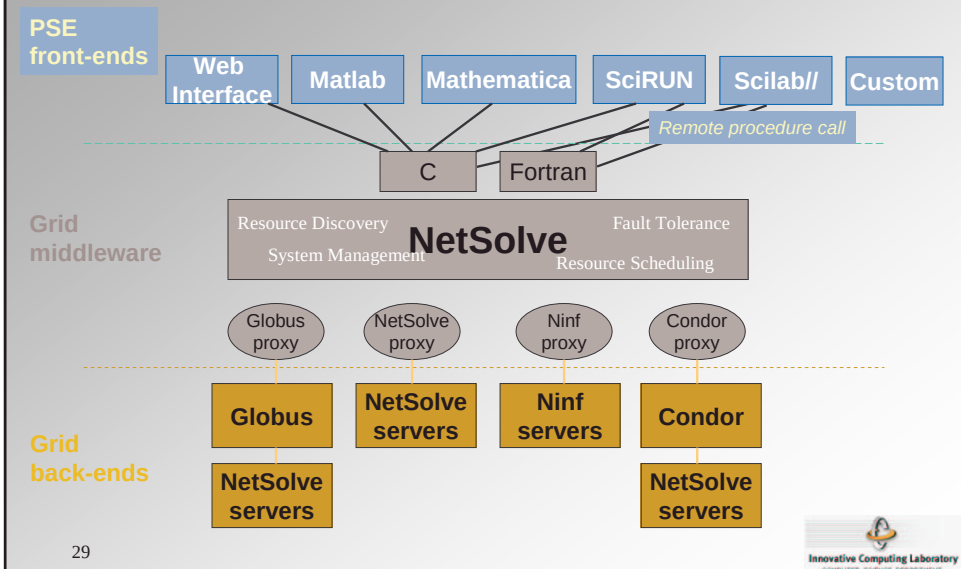


NetSolve server post-processing for visualization. Possible rendering of visualization via the internet using web browsers.

28

Innovative Computing Laboratory
COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF TENNESSEE

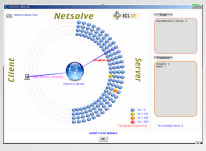
NetSolve: A Plug into the Grid



29

Things Not Touched On

- Security
  - Using Kerberos V5 for authentication.
- Separate Server Characteristics
  - Implementing Hardware and Software servers
- Hierarchy of Agents
  - More scalable configuration
- Monitor NetSolve Network
  - Track and monitor usage
- Network status
  - Network Weather Service
- Internet Backplane Protocol
  - Middleware for managing and using remote storage.
- Fault Tolerance
  - Volker Strumpen's Porch
- Local / Global Configurations
- Dynamic Nature of Servers
- Automated Adaptive Algorithm Selection
  - Dynamic determine the nest algorithm based on system status and nature of user problem



30

Contributors

- SANS-Effort
  - Victor Eijkhout, UTK
  - Piotr Luszczek, UTK
  - Kenny Roche, UTK
  - Clint Whaley, FSU
- NetSolve – Demos at 4:00
  - Sudesh Agrawal, UTK
  - Henri Casanova, UCSD
  - Michelle Miller, UTK
  - Zhaio Shi, UTK
  - Sathish Vadhiyar, UTK

For additional information see…

http://www.netlib.org/atlas/

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

http://www.cs.utk.edu/~dongarra/

Download ATLAS and NetSolve

Many opportunities within the group at Tennessee