



# IBM ACTC: Helping to Make Supercomputing Easier

**Luiz DeRose**

**Advanced Computing Technology Center  
IBM Research**

HPC Symposium  
University of Oklahoma  
Sept 12, 2002

[laderose@us.ibm.com](mailto:laderose@us.ibm.com)  
© 2002



## Outline

- Who we are
  - Mission statement
  - Functional overview and organization
  - History
- What we do
  - Industry solutions and activities
    - ❖ Education and training
    - ❖ STC community building
    - ❖ Application consulting
    - ❖ Performance tools research



# ACTC

## ■ Mission

- To close the gap between HPC users and IBM
- Conduct research on applications for IBM servers within the scientific and technical community
  - ❖ Technical directions
  - ❖ Emerging technologies

## ■ ACTC - Research

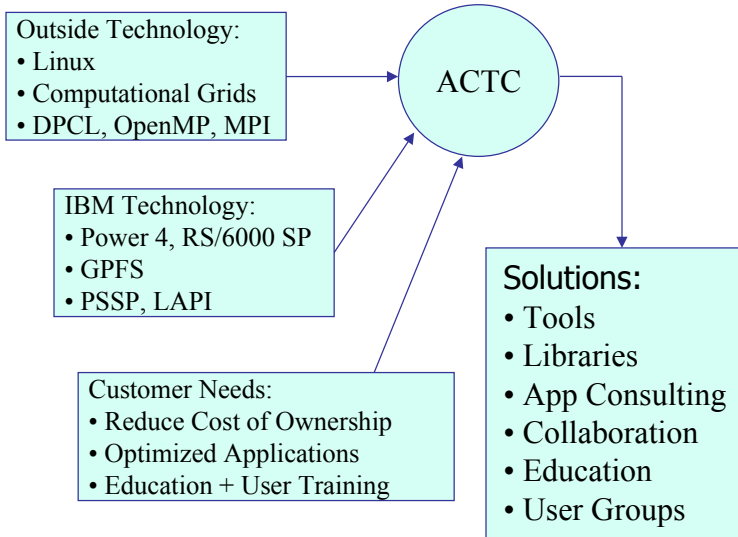
- Software tools and libraries
- HPC applications
- Research collaborations
- Education and training

## ■ Focus

- AIX and Linux platforms



# ACTC Functional Overview





## ACTC History

- Created in September, 1998
  - Emphasis on helping new customers to port and optimize on IBM system
  - Required establishing relationships with scientists on research level
- Expanded operations via alignment with Web/Server Division:
  - EMEA extended in April, 1999
  - AP extended (TRL) in September, 2000
  - Partnership with IBM Centers of Competency (Server Group)



## ACTC Education

- 1st Power4 Workshop
  - Jan. 24,25, 2001 at Watson Research
  - Full Architecture Disclosure from Austin
- 1st Linux Clusters Optimization Workshop
  - May 25--27, 2001 at MHPCC
  - Application Tuning for IA32 and Myrinet
- 2001 European ACTC Workshop on IBM SP
  - Feb. 19,20, 2001 in Trieste (Cineca)
  - <http://www.cineca.it/actc-workshop/index.html>
- 2001 Japan HPC Forum and ACTC Workshop
  - July 17--19 2001 in Tokyo and Osaka
- 1st Linux Clusters Institute Workshop
  - October 1--5, 2001 at UIUC/NCSA (Urbana)
- IBM Research Internships at Watson



## HPC Community: IBM ScicomP

- Established in 1999 via ACTC Workshop
- IBM External Customer Organization
  - Equivalent to "CUG" (Cray User Group)
- 2000 – US and European groups merge
- 2001
  - 1st European Meeting (SciComp 3)
    - ✦ May 8-11, Barcelona, Spain (CEPBA)
  - US Meeting: October 9-12, Knoxville, TN
- 2002 -
  - ScicomP 5, Manchester, England (Daresbury Lab )
  - ScicomP 6, Berkeley, CA (NERSC)
    - ✦ Joint meeting with SP/XXL
- 2003 -
  - ScicomP 7 in Goettingen, Germany, March 3-7
  - ScicomP 8 in Minneapolis (MSI), August 5-9
- <http://www.spscopicomp.org>

Sept 12, 2002

ACTC - © 2002 - Luiz DeRose

laderose@us.ibm.com 7



## HPC Community: LCI

- Linux Clusters Institute (LCI)
  - <http://www.linuxclustersinstitute.org>
  - Established April, 2001
  - ACTC Partnership with NCSA and HPCERC
  - Mission: education and training for deployment of Linux clusters in the STC community
  - Primary activity: intensive, hands-on workshops
  - Upcoming Workshops
    - ✦ Albuquerque, NM (HPCERC) Sep 30-Oct 04, 2002
    - ✦ University of Kentucky, Lexington - Jan 13-17, 2003
- The third LCI International Conference on Linux Clusters
  - October 23-25, 2002; St Petersburg, FL
  - <http://www.linuxclustersinstitute.org/Linux-HPC-Revolution/>

Sept 12, 2002

ACTC - © 2002 - Luiz DeRose

laderose@us.ibm.com 8



## Porting Example - NCSC (Cray)

- ACTC Workshop at NCSC
  - Bring your code and we show you how to port it (5-day course)
  - 32 researchers attended
  - All but one were ported by end of the week
    - ❖ the one code not ported was because it was written in PVM, and was not installed on the SP system
- NCSC now conducts their own workshops

Sept 12, 2002

ACTC - © 2002 - Luiz DeRose

laderose@us.ibm.com 9



## Porting Examples

- NAVO (Cray)
  - Production benchmark code required maximum of 380-400 elapsed seconds
    - ❖ ACTC ported and tuned this code to run under 315 seconds on the SP
  - ACTC Workshop at NAVO
    - ❖ 8 production codes ported and optimized by end of week
- EPA (Cray)
  - One ACTC man-month
    - ❖ Convert largest user code to the SP
    - ❖ Codes now run 3 to 6 times faster than they did on T3E
- NASA/Goddard
  - 5 codes ported with help of one ACTC scientist working onsite for two weeks

Sept 12, 2002

ACTC - © 2002 - Luiz DeRose

laderose@us.ibm.com 10

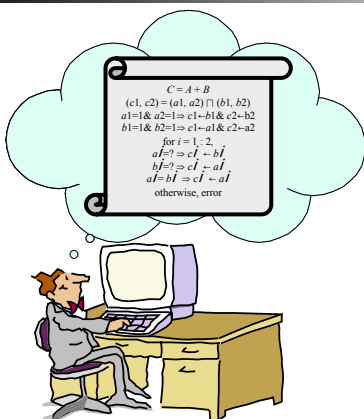


## HPC Tools Requirements

- User demands for performance tools
  - e.g., memory analysis tools, MPI and OpenMP tracing tools, system utilization tools
- Insufficient tools on IBM systems
  - Limited funding for development
  - Long testing cycle (18 months)
- STC Users comfortable with “AS-IS” software

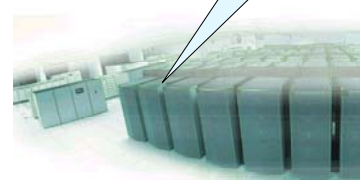


## Performance Tools Challenge



```

...
addi r4,r0,1024
fmr fp0,fp0
stfd fp31,-8(SP)
mfispr r0,I,R
stfd fp30,-16(SP)
lwz r3,r3,30(const(RTOC))
stmw r29,-28(SP)
lwz r31,T,34
lwz r30,r30,_gen_index_
stw r0,8(SP)
addis r0,r0,17200
...
  
```



- User's mental model of program do not match with executed version
  - Performance tools must be able to revert this semantic gap



## ACTC Tools

- **Goals**
  - Help IBM sales in HPC and increase user satisfaction
  - Complement IBM application performance tools offerings
- **Methodology**
  - Development in-house and in collaboration with Universities and Research Labs
  - When available, use infrastructure under development by the IBM development divisions
  - Deployment "AS IS" to make it rapidly available to the user community
  - Close interaction with application developers for quick feedback, targeting functionality enhancements
- **Focus**
  - AIX and Linux

Sept 12, 2002

ACTC - © 2002 - Luiz DeRose

laderose@us.ibm.com 13



## ACTC Software: Tools

- **Performance Tools**
  - **HW Performance Monitor Toolkit (HPM)**
    - ❖ HPMCount
    - ❖ LIBHPM
    - ❖ HPMViz
    - ❖ CATCH
  - **UTE Gantt Chart**
    - ❖ MPI trace visualization

Sept 12, 2002

ACTC - © 2002 - Luiz DeRose

laderose@us.ibm.com 14



## ACTC Software: Collaborative Tools

- OMP/OMPI trace and Paraver
  - Graphical user interface for performance analysis of OpenMP programs
  - Center for Parallelism of Barcelona (CEPBA) at UPC - Spain
- SvPablo
  - Performance Analysis and Visualization system
  - Dan Reed at University of Illinois
- Paradyn and Dyninst
  - Dynamic system for performance analysis
  - Bart Miller at University of Wisconsin
  - Jeff Hollingsworth at University of Maryland



## ACTC Software: Libraries

- MPItrace
  - Performance trace library for MPI analysis
- TurboMPI
  - Collective communication functions to enhance MPI performance for SMP nodes
- TurboSHMEM
  - "Complete" implementation of Cray SHMEM interface to allow easy and well-tuned porting of Cray applications to the IBM RS/6000 SP
- MIO (Modular I/O)
  - I/O prefetching and optimization library to enhance AIX Posix I/O handling





# Current Research Projects

- **Hardware Performance Monitor (HPM) Toolkit**
  - **Data capture, analysis and presentation of hardware performance metrics for application and system data**
    - ❖ Correlation of application behavior with hardware components
    - ❖ Hints for program optimization
    - ❖ Dynamic instrumentation and profiler
- **Simulation Infrastructure to Guide Memory Analysis (SIGMA)**
  - **Data-centric tool under development to**
    - ❖ Identify performance problems and inefficiencies caused by the data layout in the memory hierarchy
    - ❖ Propose solutions to improve performance in current and new architectures
    - ❖ Predict performance



# HPMVIZ

Label	ExcSec	InoSec	Count
Loop 300	4.572	4.572	2398
Loop 200	4.203	4.203	2400
Loop 100	3.071	3.071	2400
Calc3	1.838	6.813	2398
Calc2	1.013	5.632	2400

```

swim_omp | calc1.f | calc2.f | calc3.f
* VOLD(N1,N2), POLD(N1,N2),
2 CU(N1,N2), CV(N1,N2),
* Z(N1,N2), H(N1,N2), PS(N1,N2)
C
COMMON /CONS/ DT, TDT, DX, DY, A, ALPHA, ITMAX, MPRINT
1 NPI, EL, PI, TPI, DI, DJ, PCF
integer ierr

```

Metric Browser: Loop 300

Node	Thread	Count	ExcSec	InoSec	U time	Use rate	(M) LS	MIPS	HW FPI/Cyc	Instr/LS	M Flips	IpC	Mflips/s	WFlips	Wflp
0	3	2398	4.539	4.539	3.923	98.425	580.058	281.855	0.116	2.245	589.88	0.26	129.947	589.88	129.947
0	0	2398	4.572	4.572	4.378	95.783	608.414	283.277	0.107	1.978	608.234	0.211	133.037	608.234	133.037
0	2	2398	4.549	4.549	4.366	95.879	580.018	255.241	0.104	1.868	589.838	0.205	129.863	589.838	129.863
0	1	2398	4.547	4.547	4.308	94.758	580.024	258.19	0.105	1.897	589.837	0.21	129.728	589.837	129.728
1	2	2398	4.534	4.534	4.398	96.999	590.044	253.123	0.103	1.945	589.856	0.201	130.088	589.856	130.088
1	1	2398	4.528	4.528	3.942	87.068	588.983	288.058	0.115	2.185	589.807	0.253	130.263	589.807	130.263
1	0	2398	4.547	4.547	3.796	82.828	608.434	308.085	0.124	2.302	608.244	0.288	133.782	608.244	133.782
1	3	2398	4.523	4.523	3.537	78.198	589.982	317.348	0.128	2.433	589.781	0.312	130.4	589.781	130.4
2	0	2398	4.538	4.538	3.777	83.218	608.448	312.01	0.124	2.327	608.262	0.288	134.029	608.262	134.029
2	2	2398	4.522	4.522	4.313	95.384	580.033	257.982	0.105	1.977	589.86	0.208	130.431	589.86	130.431
2	3	2398	4.52	4.52	4.307	95.285	588.985	258.883	0.105	1.883	589.806	0.208	130.482	589.806	130.482
2	1	2398	4.52	4.52	4.35	98.222	588.943	255.814	0.104	1.88	589.787	0.205	130.486	589.787	130.486
3	1	2398	4.487	4.487	4.183	83.453	571.551	259.827	0.105	2.04	571.374	0.214	127.352	571.374	127.352
3	1	2398	4.502	4.502	4.385	88.853	588.837	254.186	0.104	1.84	589.783	0.202	131.003	589.783	131.003
3	2	2398	4.483	4.483	4.138	82.33	571.558	263.884	0.106	2.07	571.38	0.22	127.445	571.38	127.445
3	0	2398	4.506	4.506	3.827	87.154	580.044	280.852	0.116	2.221	589.856	0.257	130.901	589.856	130.901

```

V(I,J) = VNEW(I,J)
P(I,J) = PNEW(I,J)
300 CONTINUE
call f_hpmviz03_30_omp_get_thread_num()


```

Metric Options:

- Count
- ExcSec
- InoSec
- PM\_FPU\_FDIV
- PM\_FPU\_FMA
- PM\_FPU0\_FIN
- PM\_FPU1\_FIN
- PM\_CYC
- PM\_FPU\_STF
- PM\_INST\_CMPL
- PM\_LSU\_LDF
- U time
- Use rate
- (M) LS
- MIPS
- HW FPI/Cyc
- Instr/LS
- M Flips
- IpC
- Mflips/s
- Wflips
- FMA %
- Comp Int.

**ACTC**

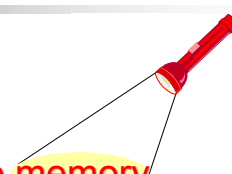
e-business



**IBM**

## SIGMA

- Where is the memory?
- Motivation
  - Efficient utilization of the memory subsystem is a critical factor to obtain performance on current and future architectures
- Goals:
  - Identify bottlenecks, problems, and inefficiencies in a program due to the memory hierarchy
  - Support serial and parallel applications
  - Support present and future architectures



Sept 12, 2002
ACTC - © 2002 - Luiz DeRose
laderose@us.ibm.com 19

**ACTC**

e-business



**IBM**

## Current Collaboration Efforts

- LLNL (CASC) – with Jeff Vetter
  - Interpreting HW Counters analysis
    - ❖ Handling of large volume of multi-dimensional data
    - ❖ Correlate hardware events with performance bottlenecks
- Research Centre Juelich – with Bernd Mohr
  - Usability and reliability of performance data obtained from performance tools
  - Feasibility of automatic performance tools
- CEPBA (UPC) – with Jesus Labarta
  - Performance analysis of OpenMP programs
  - OMP/OMPI Trace & Paraver interface
- University of Karlsruhe - with Klaus Geers
  - HPM Collect

Sept 12, 2002
ACTC - © 2002 - Luiz DeRose
laderose@us.ibm.com 20

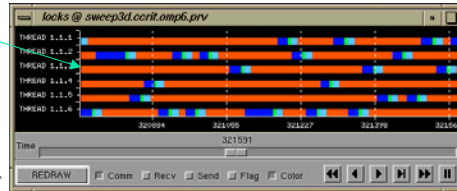
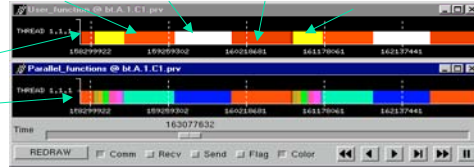


# Program Analysis - Paraver

## Basics:

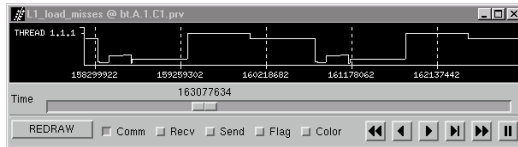
- Functions
- Loops
- Locks

x\_solve y\_solve z\_solve rhs



## Hardware counters:

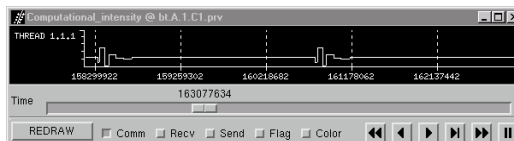
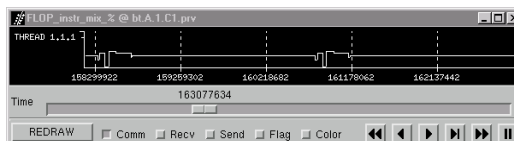
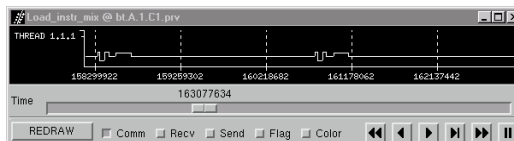
- TLB misses
- L1 misses
- ...



# Hardware Counters Analysis

## Derived metrics

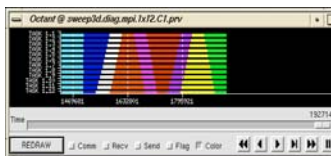
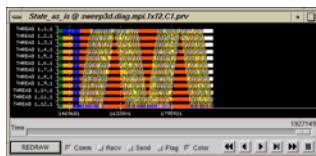
- Instruction mix
  - ❖ %Loads
  - ❖ %Flops
- Computational intensity
  - ❖ flops/loads
- Program on architecture
  - ❖ miss ratios
  - ❖ FLOPS per L1 miss
- Performance
  - ❖ IPC
  - ❖ FPC



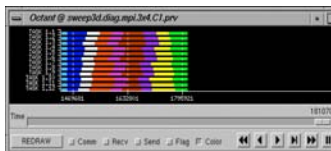
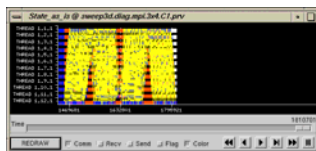


## Analysis of Decomposition Effects

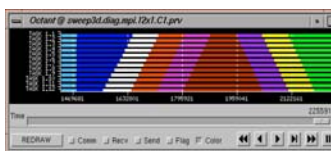
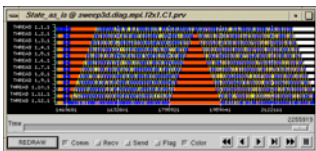
■ 1 x 12



■ 3 x 4



■ 12 x 1



## Summary

- ACTC is a consolidation of IBM's highest level of technical expertise in HPC
- ACTC research is rapidly disseminated to maximize benefits of IBM servers for HPC applications