
Kohki Ishikawa (kixs@jp.ibm.com)
IBM Japan, Ltd.

8th June, 2012

A Linux Application Tool to Leverage the Full Capability of Hardware



Agenda

- Linux and Open Source Software Trends
- POWER processor & Power Systems benefit to Linux
- Advanced Toolchain
 - Overview
 - Components
 - How to use
 - Example Adaption for PostgreSQL 9
- SDK for PowerLinux
 - Overview
 - Features
 - How to use
- Conclusion

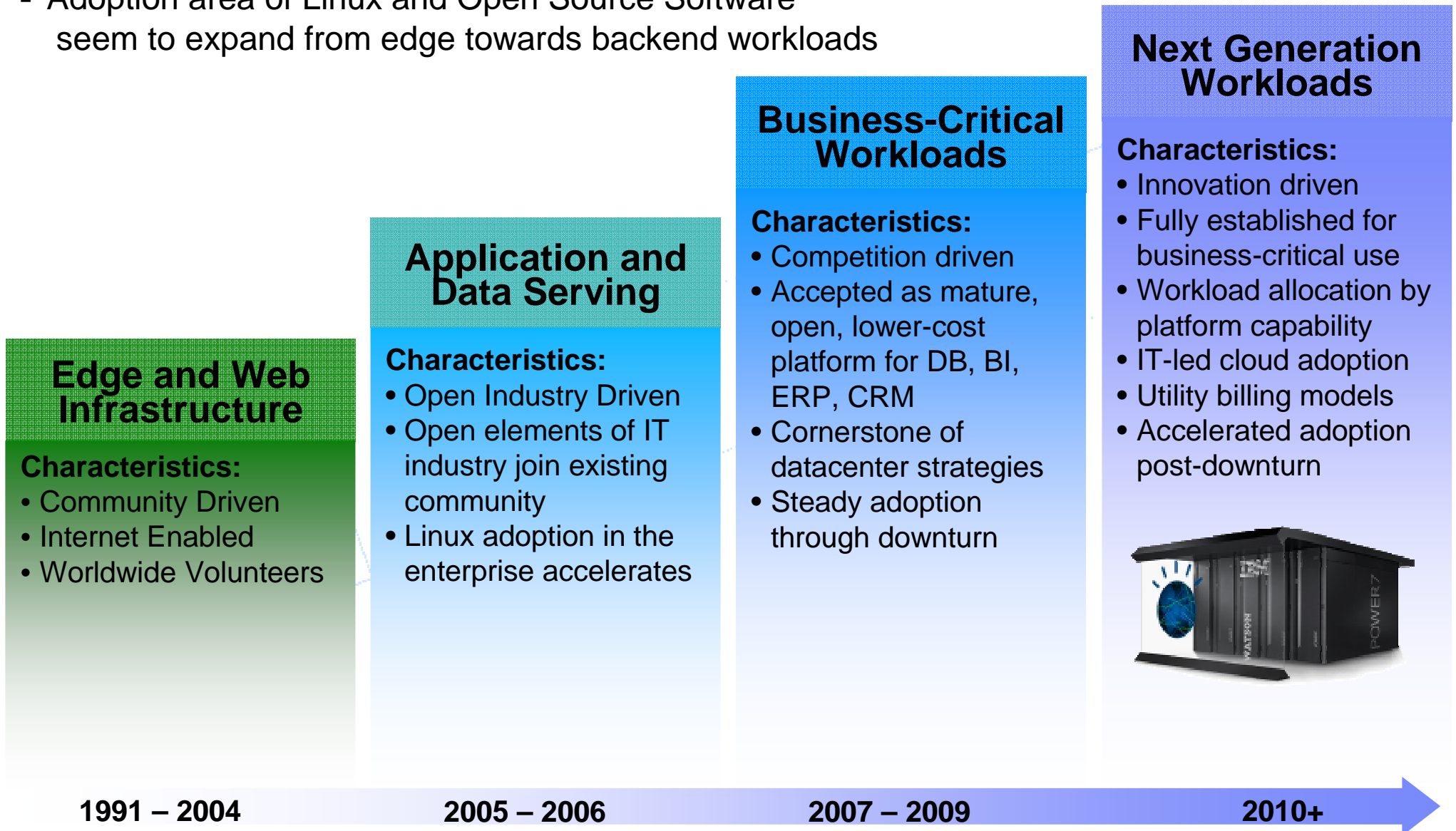
Linux and Open Source Software Trends

- Becomes more “enterprise”
 - This report says
 - 73% companies now place open source equal to or ahead of proprietary software
 - 68 % companies use Linux, which is the most popular open source package than others

* please refer to the source, original website

From Edge to Business Critical Workloads

- Adoption area of Linux and Open Source Software seem to expand from edge towards backend workloads



PPC and PPC64 architecture

- Embedded, Super Computers, Game Consoles, Appliance and Enterprise Servers



IBM Power Systems



Blue Gene

* Please refer to
endor website

OpenBlocks Micro Server

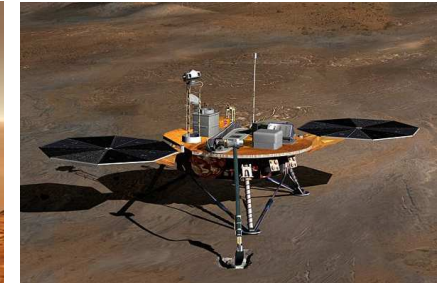
<http://openblocks.plathome.com/>



Pathfinder



Spirit



Phoenix

POWER processor & Power Systems benefit to Linux

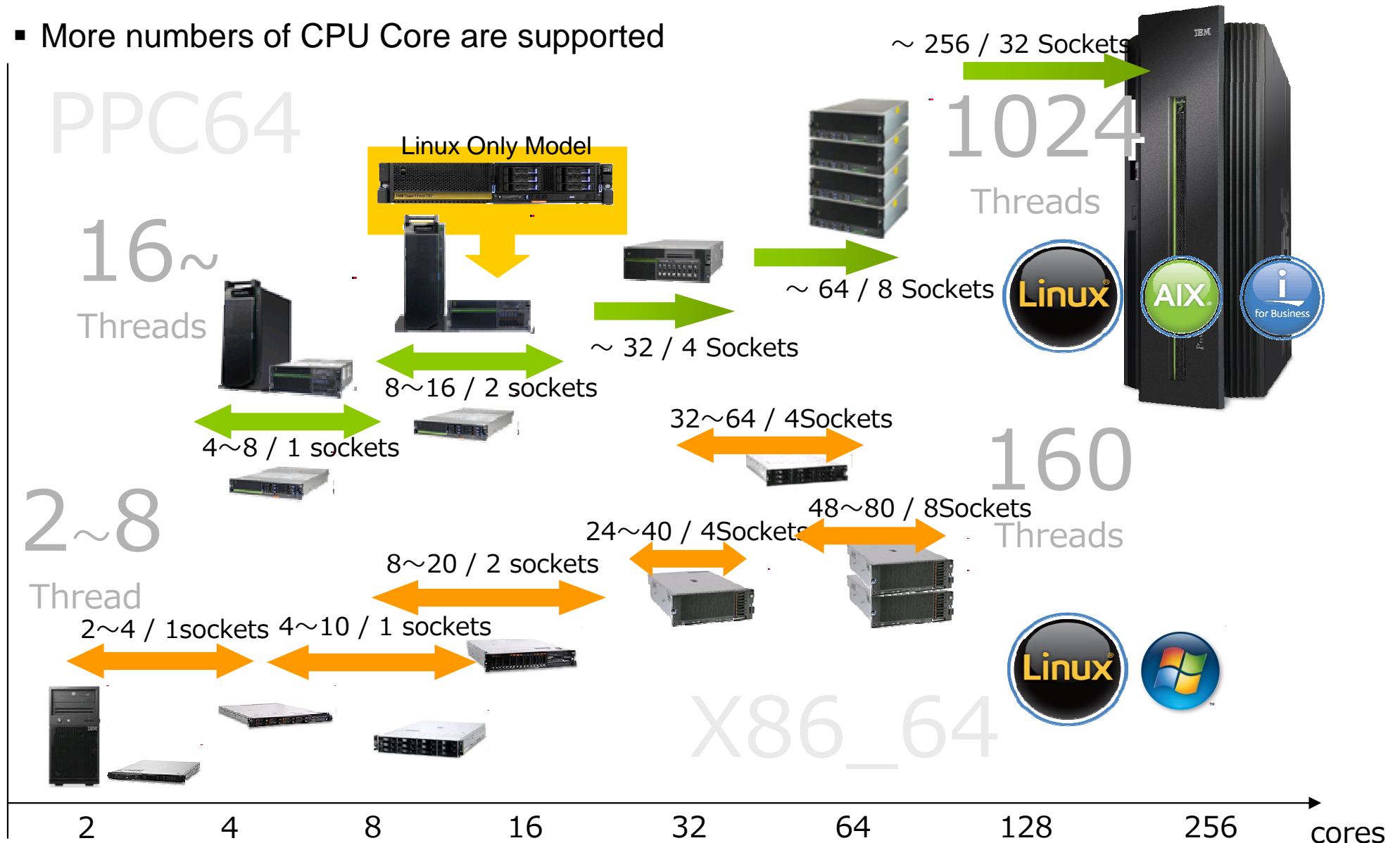
- POWER is
 - a RISC microprocessor architecture developed by IBM
 - a microprocessor implementation of the POWER ISA

- IBM Power Systems
 - uses IBM POWER processor
 - Now the latest processor generation is POWER7
 - supports running Linux
 - Red Hat Enterprise Linux, SUSE Linux Enterprise Server
Fedora, Debian...
 - IBM calls these Linux environments on Power Systems as “PowerLinux”

- POWER & Power Systems can provide to Linux users and market
 - additional choice of server hardware
 - more Scalability & Reliability

More Scalable

- More numbers of CPU Core are supported



Source : <http://www.ibm.com/systems/jp/x/product/>
at 7 Mar. 2012

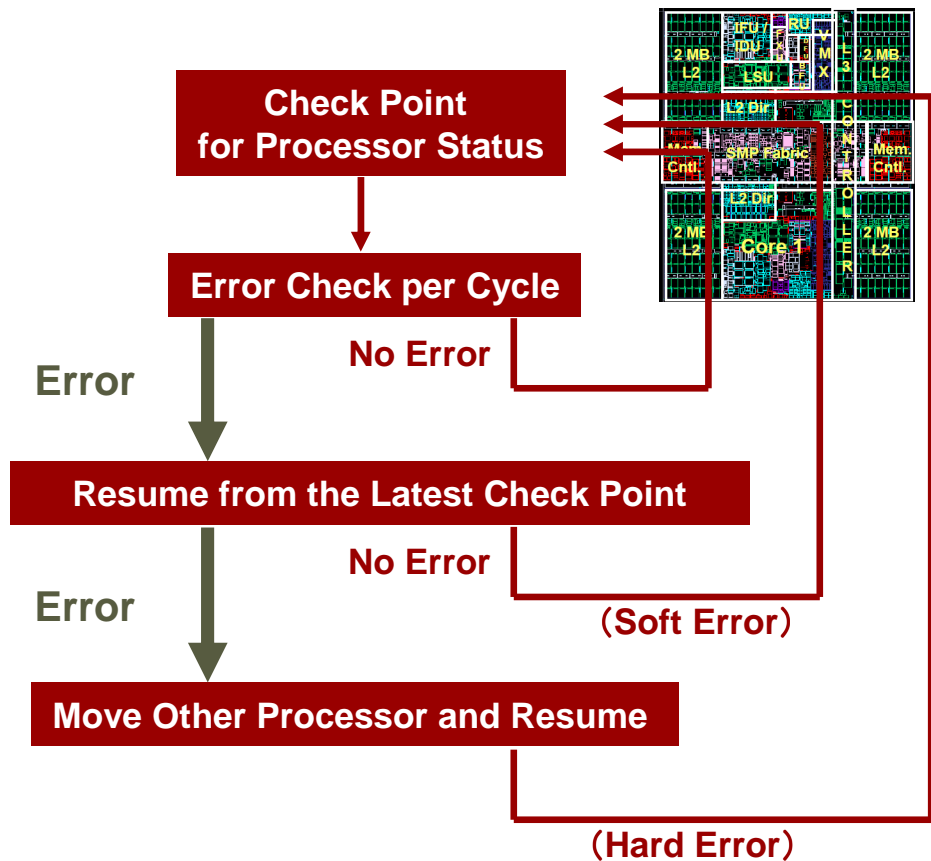
& <http://www.ibm.com/systems/jp/power/hardware/>

More Reliable

- Trying to run continuously when an error occurs

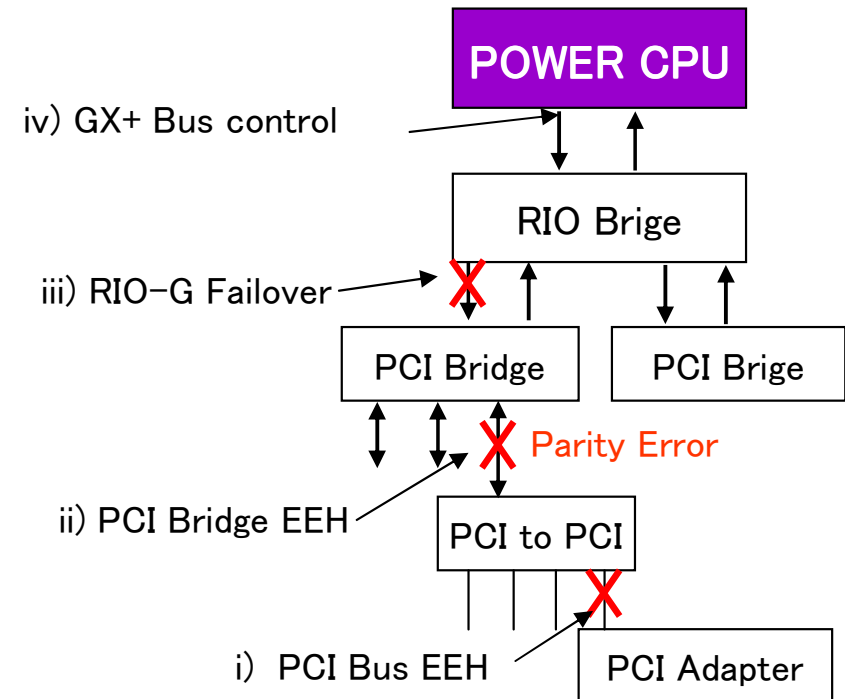
Processor Recovery

Transparent error recovery by Hardware



PCI Bus Error Recovery

Capability to improve the reliability of PCI/PCI-E bus peripherals



⇒ EEH is introduced in the competing session. Please follow-up after finish of conference

Application is much portable

- Linux Application is portable although between different architectures.
 - On commercial distribution of Linux, such as RHEL or SLES, although different architecture,
 - Linux kernel is built from same version of source code, so available almost completely same function on kernel
 - Bundling middleware and libraries are also same version, so available almost completely same API for middleware or libraries
 - ex) RDBMS, Language Runtime, Graphic Libraries...etc.
 - Recent application is coded by “portable” Language
 - Script Language, such as Perl, PHP, Ruby
 - Java
 - Web-App-RDB 3 Tier Application depends on only middleware layer

- Only C/C++ Applications or Libraries modification should be careful.

Two Major helpers for more easily C/C++ application development

- Advance Toolchain Linux on Power Systems
- Software Development Kit for PowerLinux

Advance Toolchain Overviews & Component

- Available from University of Illinois Web Site
 - A set of open source software development extensions and tools allowing users to take leading edge advantage of IBM latest hardware features:
 - POWER6 enablement
 - POWER6 Optimized scheduler
 - POWER6 Native DFP instruction support
 - POWER6 VMX enablement with auto-vector
 - POWER7 enablement
 - POWER7 Optimized scheduler
 - POWER7 Native DFP instruction support
 - POWER7 VMX/VSX enablement with auto-vector
 - ppc970, POWER4, POWER5, POWER5+, POWER6, POWER6x, POWER7 optimized system and math libraries
 - libhugetlbfs 2.0 support

AT5.0

<http://www.ibm.com/developerworks/wikis/display/hpccentral/How+to+use+Advance+Toolchain+for+Linux+on+POWER>

How to use AT

- The repository Information is available from following URL:

<http://www-304.ibm.com/webapp/set2/sas/f/lopdiags/yum.html>

- Easy to install

- Recent versions are available through online repository, such as yum and zypper.
- Prior to install AT, install locally only 1 package

- For example on RHEL

```
# rpm -ivh ibm-power-repo-1.1.6-5.ppc.rpm
```

- After that, Just execute online installation command

- For example on RHEL

```
# yum install advance-toolchain-at5.0-runtime  
or  
# yum install advance-toolchain-at5.0-devel
```

- Flexible, easily co-exist and switch the multiple versions of tool chains

```
$ ls /opt  
at4.0 at5.0
```

AT5 Example Adaption for PostgreSQL 9

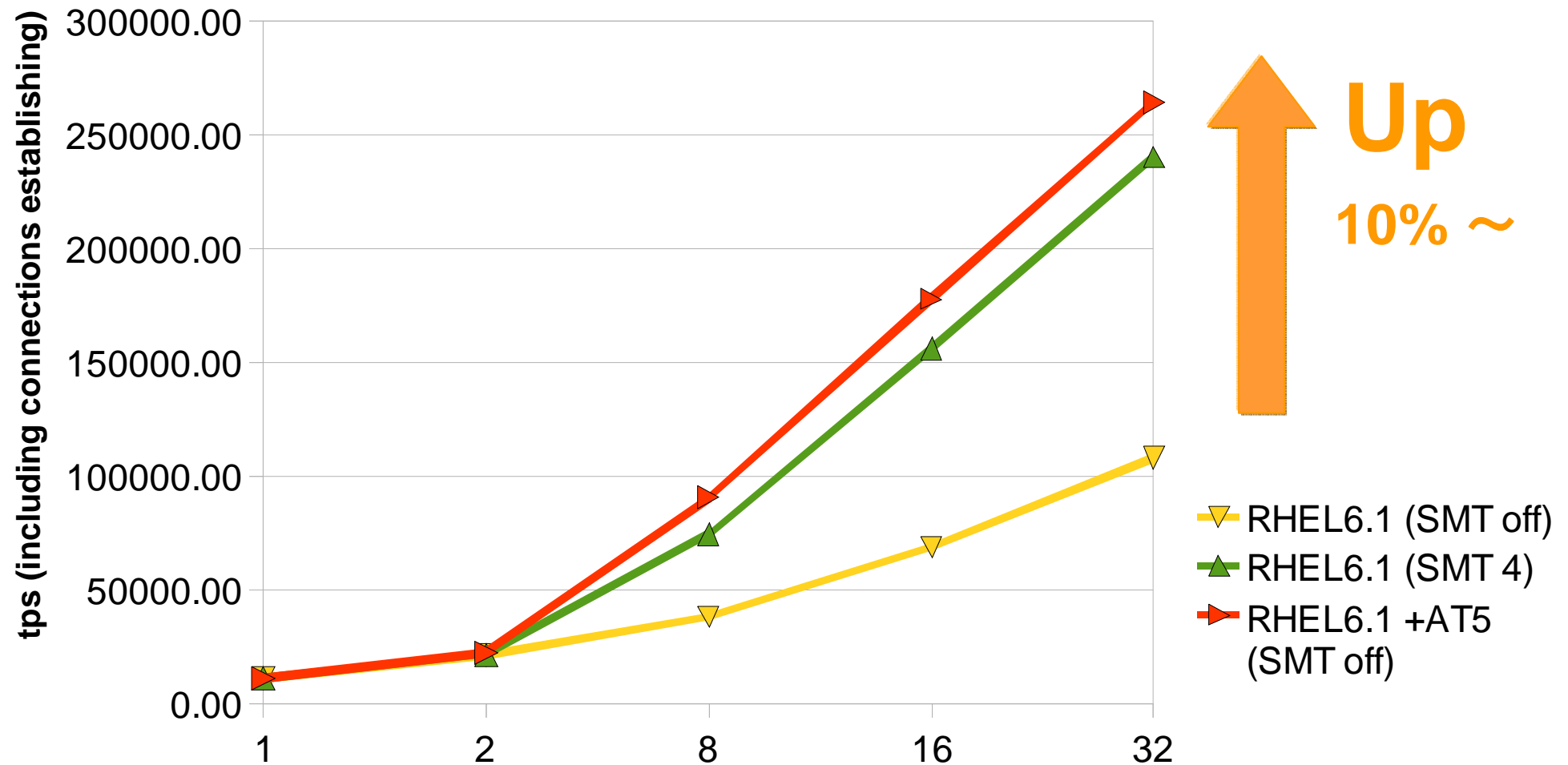
- PostgreSQL is
 - a famous open source RDBMS
 - developed by PostgreSQL Development Community

- PostgreSQL contains a self benchmarking tool, pgbench
 - pgbench execute several types of workloads
 - Offline Batch Transaction (pgbench default, TPC-B)
 - Read Only Queries (pgbench -S)
 - Online Mix Transaction (pgbench -N)

- pgbench -S mainly depends on CPU and memory

Results of Adaption of AT5.0

pgbench -S , scale factor 100, 500 sec,
32-core POWER7 on Power 750 (128 threads)
max_connections = 32, shared_buffer = 8GB



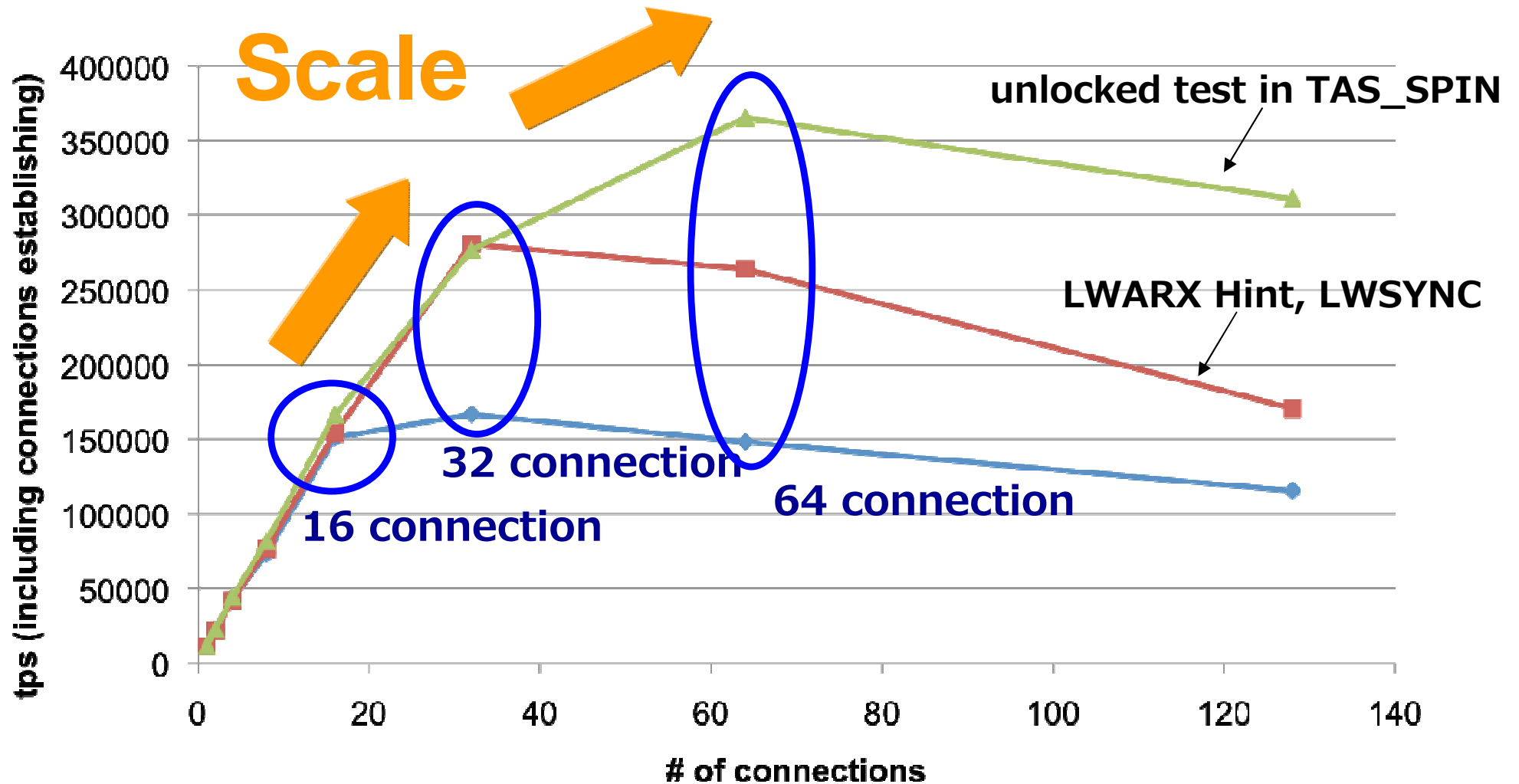
SMP scalability is not automatically realize

- AT5.0 also provides other functions:
 - Decimal Floating Point Library (libdfp)
 - GNU Binary Utilities (ld, ldd, objcopy, objdump, nm, and others)
 - GNU Debugger (gdb)
 - performance analysis tools (Oprofile, Valgrind, gprof, mtrace, xtrace, iTrace)
 - The AUXV Library (libauxv)

- But does not automatically covers
 - SMP scalability

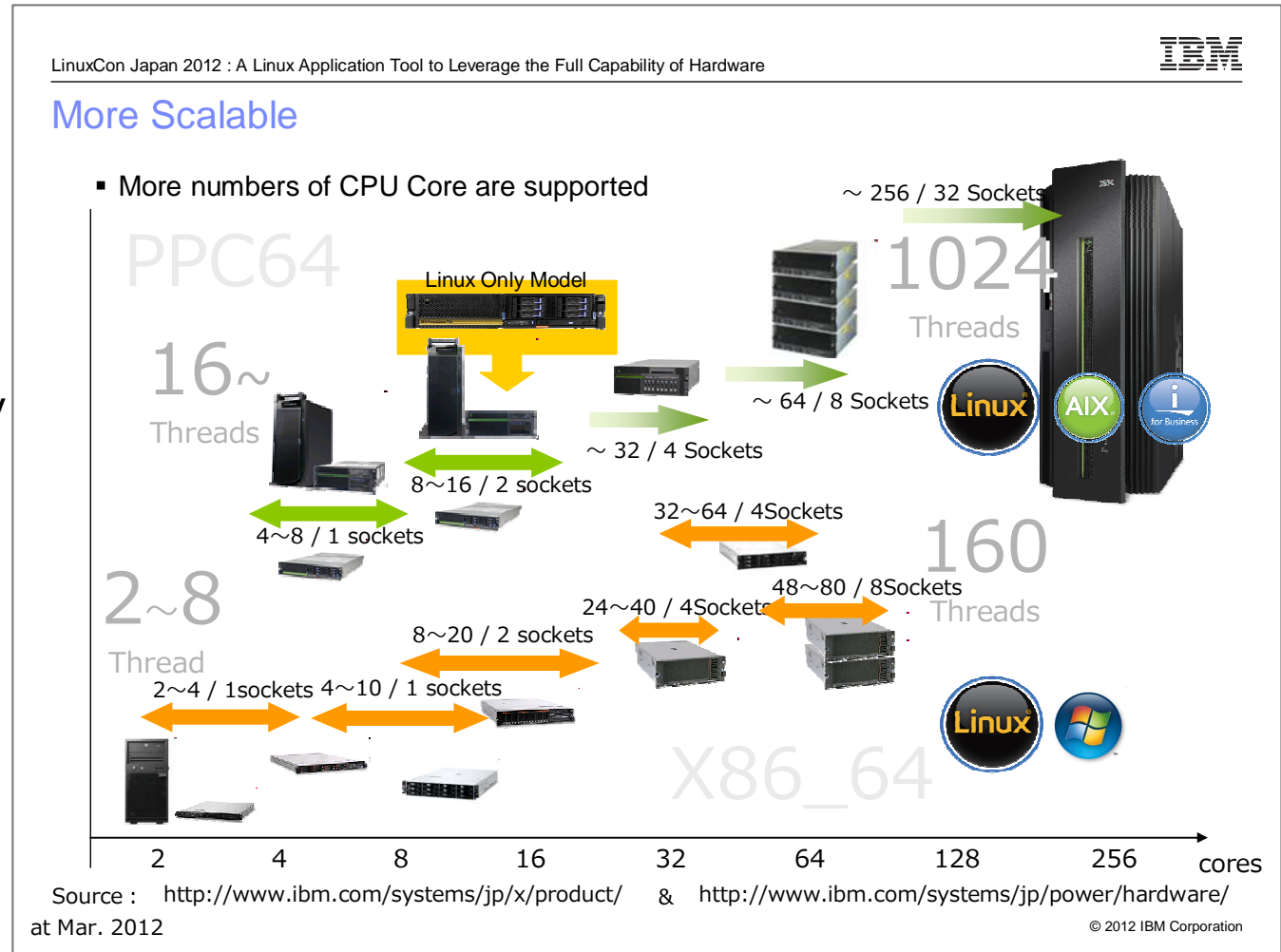
Example: PostgreSQL 9.2dev Modification on POWER

pgbench -S , scale factor 100, 500 sec,
 32-core POWER7 on Power 750 (128 threads)
 max_connections = 128, shared_buffer = 8GB



To Leverage the Full Capability

- SMP scalability improvement is important
- Analytics and Suggestion may help your development
 - System Analytics
 - Code Suggestion



* please refer to slide #7 in this presentation

SDK for PowerLinux

- New released in 2012
- Eclipse-based IDE
 - CDT
 - Code Analytics Tool
 - GFE
 - PTP
- All-in-one solution for developing softwares on PowerLinux (Linux running on Power Systems)



How to use

- Available from following URL;

<http://www-304.ibm.com/webapp/set2/sas/f/lopdiags/sdklop.html>

- Easy to install (same as AT)

- Recent versions are available through online repository, such as yum and zypper.
- Prior to install SDK, install locally only 1 package

- For example on RHEL

```
# rpm -ivh ibm-power-repo-1.1.6-5.ppc.rpm
```

- After that, Just execute online installation command

- For example on RHEL

```
# yum install ibm-sdk-lop
```

SDK sample screen shots

The screenshot displays the IBM Software Development Kit for PowerLinux interface. The main window shows a 'Welcome' screen with navigation options: Overview, Samples, Web Resources, Tutorials, and What's New. An 'About IBM Software Development Kit for PowerLinux' dialog box is open, providing version information (1.0.0) and a detailed license agreement. The background IDE window shows a C/C++ project with a source file 'PowerLinux Sample Project.c' containing the following code:

```

/*
 * Name      : PowerLinux.c
 * Author    : Sample
 * Version   :
 * Copyright : Your copyright notice
 * Description: Hello World in C, Ansi-style
 */

#include <stdio.h>
#include <stdlib.h>

int main(void) {
    puts("!!!Hello World!!!"); /* prints !!!Hello World!!! */
    return EXIT_SUCCESS;
}
    
```

The IDE also shows a Project Explorer with a tree view of the project structure, including 'main(void) : int', 'stdio.h', and 'stdlib.h'. A Problems view at the bottom indicates '1 error, 2 warnings, 0 others'.

SDK for PowerLinux extends value-add plugins

- C/C++ project support of IBM Advance Toolchain
 - Power optimization wizard

- Linux Tools OProfile plugin
 - Launch and analysis integrated with code development
 - Configurable for HW specific event profiling
 - POWER6/7 PMU events

- Linux Tools Valgrind plugin
 - Launch and analysis integrated with code development
 - Open framework for dynamic analysis
 - Memcheck, detects memory leaks and malloc/free errors
 - Cachegrind, cache and branch miss analysis
 - Helgrind, thread and data race analysis
 - PowerISA features for POWER6/7

- Linux Tools RPM plugin
 - Build RPM packages from source code

SDK for PowerLinux includes additional Power-unique features

- FDPR (Feedback Directed Program Restructuring)
 - Integrated with Eclipse/CDT for ease of use
 - Works on both executable programs and shared libraries
 - Provides post-link global code optimization step
 - Tunes program to a representative workload

- Source Code Advisor
 - Leverages the FDPR dynamic inter procedural analysis capabilities
 - Provides interactive feedback to the developer
 - Identifies hot spots in source code that need rework
 - Specific suggestions for
 - Source code structure improvements
 - Compiler/linker options to use

- Code Migration Assist plugin
 - Integrated with Eclipse context sensitive source tooling
 - Scan/Analyze application source for common migration issues
 - Data Endian dependent unions and structures
 - Cast with potential endian issues
 - Non-portable data types
 - Non-portable inline assembler code
 - Non-portable or arch dependent compiler builtins
 - Proprietary/Arch specific APIs
 - Performance degradation

Sample Screenshot : Profile application performance with ease

The screenshot displays the IBM Software Development Toolkit for Linux on POWER interface. The main editor window shows the source code for `bzlib.c`, with the `add_pair_to_block` function highlighted. The Project Explorer on the left shows the project structure for `bzip2-1.0.6`. The OProfile performance profiler is active, showing a tree view of execution cycles. A green arrow points to the OProfile window, which is currently displaying the following data:

Category	Percentage	Location
current	100.00%	in /home/wainersm/sandbox/bzip2-1.0.6/bzip2
fo	54.22%	in .mainSort [blocksort.c]
fo	34.26%	in .BZ2_compressBlock [compress.c]
fo	8.94%	in .handle_compress.clone.2 [bzlib.c]
fo	1.33%	in .add_pair_to_block [bzlib.c]
	0.25%	on line 221
	0.21%	on line 220
	0.13%	on line 235
	0.12%	on line 241

Sample Screenshot : Source Code Advisor

The screenshot displays the IBM Software Development Kit for PowerLinux interface. The main window shows the source code of a C++ file named `du.c`. The code includes a function `excluded_file_name` that checks if a file is excluded based on its name and type. The Source Code Advisor (SCA) tool is active, showing a problem analysis for the function `excluded_file_name`. The analysis indicates a high call overhead for a hot small function, with 6.63% of the time spent on line 428. The solution provided is to inline the callee into the caller, replacing the call with its body.

```

ok = false;
}
else if (info != FTS_DP)
{
    bool excluded = excluded_file_name (exclude, file);
    if (! excluded)
    {
        /* Make the stat buffer *SB valid, or fail noisily. */

        if (info == FTS_NSOK)
        {
            fts_set (fts, ent, FTS_AGAIN);
            FTSENT const *e = fts_read (fts);

```

Problem
High call overhead of a hot small function

Solution
Compiler: inline callee into caller - replace call to callee with its body

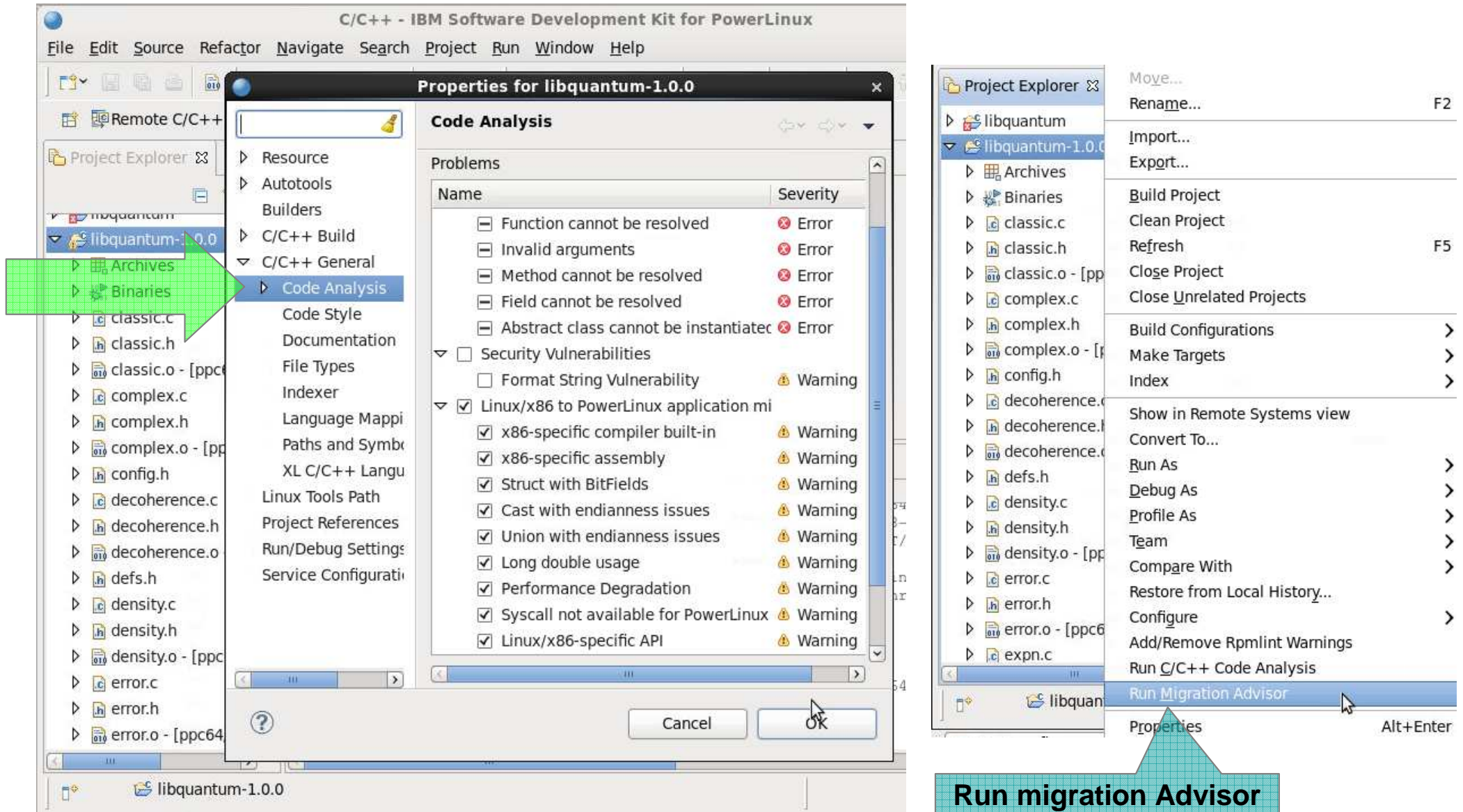
Sample Screenshot : Analyze thread using the Trace Analyzer

The screenshot displays the Trace Analyzer interface for analyzing pthreads. The main window shows a timeline view with threads represented by horizontal bars. A callout box labeled "Threads" points to the "Threads" section of the main window. The "Trace Table" on the left shows a list of events, with the "Join" event at index 65 selected. The "Color Map" table in the bottom right section maps event names to colors for visualization.

Name	Color
Thread	Grey
Start	Purple
Exit	Brown
Join	Green
Mutex_lock	Pink
Condvar_wait	Magenta

Thread events color map

Sample Screenshot : Execute Migration Advisor (1/2)



Sample Screenshot : Execute Migration Advisor (2/2)

The screenshot shows the IBM Software Development Kit for PowerLinux interface. The main editor displays the 'mesh.h' file with the following code:

```

MESH_TRIANGLE *Triangles; /* Array of triangles.
BBOX_TREE *Tree; /* Bounding box tree for mesh.
VECTOR Inside_Vect; /* vector to use to test 'inside'
};

struct Mesh_Triangle_Struct
{
    unsigned int Smooth:1; /* Is this a smooth triangle.
    unsigned int Dominant_Axis:2; /* Dominant axis.
    unsigned int vAxis:2; /* Axis for smooth triangle.
    unsigned int ThreeTex:1; /* Color Triangle Patch.
    long Normal_Ind; /* Index of unsmoothed triangle normal.
    long P1, P2, P3; /* Indices of triangle vertices.
}

```

The Migration Advisor View at the bottom shows the following table:

Description	Resource	Path	Location	Migration Adv
Union with endianness issues (21 items)				
Cast with endianness issues (100 of 1160 items)				
Performance Degradation (2 items)				
Struct with BitFields (8 items)				
Check for possible problems related to this struct with bit fields	mesh.h	/povray	line 78	Struct with Bit
Check for possible problems related to this struct with bit fields	tif_vms.c	/povray	line 338	Struct with Bit
Check for possible problems related to this struct with bit fields	tif_vms.c	/povray	line 343	Struct with Bit
Check for possible problems related to this struct with bit fields	tif_vms.c	/povray	line 355	Struct with Bit

A green arrow points from the 'libquantum-1.0.0' folder in the Project Explorer to the 'struct Mesh_Triangle_Struct' in the code. A blue callout box with a grid pattern points to the Migration Advisor View, containing the text: "Migration Advisor View (Problems report)".

Conclusion

- Linux and Open Source Software becomes
 - to be adopted deeply inside of enterprises.

- POWER processor & Power Systems can provide
 - More scalability and reliability for Linux
 - Another hardware choice for Enterprise Linux Server Market

- PPC64 Linux can use Free Application development tools
 - Advanced Toolchain
 - Easy to improve application running performance
 - SDK for PowerLinux
 - Help your system analysis and C/C++ application development

If you are interested in PowerLinux, please feel free to contact to IBM.

Special notices

This document was developed for IBM offerings in the United States as of the date of publication. IBM may not make these offerings available in other countries, and the information is subject to change without notice. Consult your local IBM business contact for information on the IBM offerings available in your area.

Information in this document concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. Send license inquires, in writing, to IBM Director of Licensing, IBM Corporation, New Castle Drive, Armonk, NY 10504-1785 USA.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The information contained in this document has not been submitted to any formal IBM test and is provided "AS IS" with no warranties or guarantees either expressed or implied.

All examples cited or described in this document are presented as illustrations of the manner in which some IBM products can be used and the results that may be achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions.

IBM Global Financing offerings are provided through IBM Credit Corporation in the United States and other IBM subsidiaries and divisions worldwide to qualified commercial and government clients. Rates are based on a client's credit rating, financing terms, offering type, equipment type and options, and may vary by country. Other restrictions may apply. Rates and offerings are subject to change, extension or withdrawal without notice.

IBM is not responsible for printing errors in this document that result in pricing or information inaccuracies.

All prices shown are IBM's United States suggested list prices and are subject to change without notice; reseller prices may vary.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

Any performance data contained in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. Some measurements quoted in this document may have been made on development-level systems. There is no guarantee these measurements will be the same on generally-available systems. Some measurements quoted in this document may have been estimated through extrapolation. Users of this document should verify the applicable data for their specific environment.

Revised September 26, 2006

Special notices (cont.)

IBM, the IBM logo, ibm.com AIX, AIX (logo), AIX 5L, AIX 6 (logo), AS/400, BladeCenter, Blue Gene, ClusterProven, DB2, ESCON, i5/OS, i5/OS (logo), IBM Business Partner (logo), IntelliStation, LoadLeveler, Lotus, Lotus Notes, Notes, Operating System/400, OS/400, PartnerLink, PartnerWorld, PowerPC, pSeries, Rational, RISC System/6000, RS/6000, THINK, Tivoli, Tivoli (logo), Tivoli Management Environment, WebSphere, xSeries, z/OS, zSeries, Active Memory, Balanced Warehouse, CacheFlow, Cool Blue, IBM Systems Director VMControl, pureScale, TurboCore, Chiphopper, Cloudscape, DB2 Universal Database, DS4000, DS6000, DS8000, EnergyScale, Enterprise Workload Manager, General Parallel File System, , GPFS, HACMP, HACMP/6000, HASM, IBM Systems Director Active Energy Manager, iSeries, Micro-Partitioning, POWER, PowerExecutive, PowerVM, PowerVM (logo), PowerHA, Power Architecture, Power Everywhere, Power Family, POWER Hypervisor, Power Systems, Power Systems (logo), Power Systems Software, Power Systems Software (logo), POWER2, POWER3, POWER4, POWER4+, POWER5, POWER5+, POWER6, POWER6+, POWER7, System i, System p, System p5, System Storage, System z, TME 10, Workload Partitions Manager and X-Architecture are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries.

A full list of U.S. trademarks owned by IBM may be found at: <http://www.ibm.com/legal/copytrade.shtml>.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

AltiVec is a trademark of Freescale Semiconductor, Inc.

AMD Opteron is a trademark of Advanced Micro Devices, Inc.

InfiniBand, InfiniBand Trade Association and the InfiniBand design marks are trademarks and/or service marks of the InfiniBand Trade Association.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries or both.

Microsoft, Windows and the Windows logo are registered trademarks of Microsoft Corporation in the United States, other countries or both.

NetBench is a registered trademark of Ziff Davis Media in the United States, other countries or both.

SPECint, SPECfp, SPECjbb, SPECweb, SPECjAppServer, SPEC OMP, SPECviewperf, SPECapc, SPECchpc, SPECjvm, SPECmail, SPECimap and SPECsfs are trademarks of the Standard Performance Evaluation Corp (SPEC).

The Power Architecture and Power.org wordmarks and the Power and Power.org logos and related marks are trademarks and service marks licensed by Power.org.

TPC-C and TPC-H are trademarks of the Transaction Performance Processing Council (TPPC).

UNIX is a registered trademark of The Open Group in the United States, other countries or both.

Revised December 2, 2010

Other company, product and service names may be trademarks or service marks of others.