IBM's Migration Kit for Solaris OS to Linux

Dr. Wolfgang Gellerich
gellerich@de.ibm.com

February 2007, Session 9256
Agenda

- Overview – purpose of the Migration Kit and its contents
- Technical Documentation provided with the Migration Kit
- Development tools provided with the Migration Kit
- Sizing a migration project
- How to obtain the Migration Kit
- Trademark notices
The Migration Kit for Solaris OS to Linux

- Migrating applications from Solaris OS to Linux requires detailed knowledge about differences concerning
  - libraries and operating system interfaces,
  - the development environment, including compiler, tools for building whole projects, managing source code repositories, packaging software,
  - system administration.

- The Migration Kit provides information about all these topics.
  - Three tools analyze source code and provide assistance for making adaptations.
  - Two text documents address administrative issues.
What is provided with the Migration Kit?

- Interactive tools to assist in porting applications:
  - Source Checking Tool
    Detects Solaris-specific constructs in C and C++ sources
    Assesses porting effort
  - Endian Checking Tool
    Identifies endian issues in C and C++ sources
  - Shell Script Checking Tool
    Identifies OS-specific differences in shell scripts

- Technical documentation (PDF files):
  - The “Guide to Application Porting from Solaris OS to Linux”
  - Documentation for all tools
The “Guide to Application Porting From Solaris OS to Linux”

- Recommendation how to best organize a migration project
- Technical differences concerning the development environment, including:
  - make
  - compiler
  - linker
- Architecture-specific differences, including
  - Sizes of base data types and their alignment, 32 to 64 bit migration
  - Endian-ness (supported by the Endian Checking Portability Tool)
  - System call and library functions (supported by the Code Checking Tool)
- Performance tuning tools available for Linux
- Software packaging tools available for Linux
The “Guide for System Administrators”

- Provides task-based grouping of differences between the two operating systems
- Covers topics like:
  - Operating system installation, initialization and booting
  - Disk, file system and device management
  - Printing
  - Networks
  - Users and groups
  - Security issues
The Source Checking Tool 1/6

• Input: Solaris OS application source code, written in C or C++

• Reports code patterns specific to Solaris OS:
  • Calls to the Solaris OS Application Program Interface (API)
  • Include files specific to Solaris OS
  • Sun compiler pragmas
  • Suggests Linux alternatives

• Knowledge database:
  • More than 3800 database entries in total
  • 624 text files providing technical documentation

• Output:
  • GUI offers interactive access
  • Annotated source code files available
This tool goes over C/C++ files, or directories of C/C++ files, to find potential problems with API or compiler pragma usage.

You can use a GUI to select which file or directory to process.
When the scan is complete, you may inspect each item discovered, file by file.
The Source Checking Tool 4/6

The tool provides information about necessary changes.

The tool knows details of over 3800 different system calls, library functions, and compiler pragmas.
The results are available in terms of project metrics.
The Source Checking Tool 6/6

A graphical view is also available.

Graphics can be saved for further usage (include into presentations etc.).
The Endian Checking Tool 1/3

• Detects code patterns that may cause an endian problem

• Example: several orders to store binary value 4A3B2C1D
  • Big-Endian 0x4A 0x3B 0x2C 0x1D
  • Little-Endian 0x1D 0x2C 0x3B 0x4A

• Affected code patterns include:
  • Byte-oriented processing of binary data
  • Data structures accessed by assembler code
  • Type conversion by (mis-)using pointers or union

• Input is a combination of:
  • Source Code
  • Binary compiled with profiling options enabled
The Endian Checking Tool 2/3

Before the tool can analyze the code, it must first be compiled using the usual Linux compiler.

Once the code is prepared, the tool can be used on a command-line or GUI basis.
The Endian Checking Tool 3/3

• Typical finding: parameter size mismatch

/test/src/init.c – Line 199: E30001
Variable/parameter size mismatch arg 2 size 4
in call to mystrcpy. (Defined in
/test/src/init.c at line 190 size 1)

• Formal parameter declaration: type has size of 4 bytes
• The actual argument has a size of 1 byte
• Where is this byte stored within the four bytes available?
The Shell Script Checking Tool

- Examines shell scripts looking for:
  - Path issues,
  - File issues,
  - Utility programs

- Covers:
  - Bourne,
  - csh,
  - ksh,
  - and variants

- ...and provides recommendations on what changes might be necessary.
Shell Script Checking Tool Example

+++ Begin report for 'autofs' Mon Aug 1 09:12:00 EDT 2005

+++ Summary Information for 'autofs'
4 Total Items identified

1 E1001 Items: File path does not exist on Linux
2 E4001 Items: Directory structure or file path may not map directly on Linux
1 E7001 Items: Comparable Linux Command may exist

+++ Possible Error Code Resolutions
E1001 Consult Linux man pages to determine if a comparable Linux path exists for these files.
E4001 Consult Linux man pages to determine the path the script should use.

+++ Detail Information for autofs

<table>
<thead>
<tr>
<th>Line</th>
<th>Item</th>
<th>Error code and Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>'/dev'</td>
<td>E4001 Files under this Path may not map directly on Linux</td>
</tr>
<tr>
<td>12</td>
<td>'/lib'</td>
<td>E4001 Files under this Path may not map directly on Linux</td>
</tr>
<tr>
<td>12</td>
<td>'/usr/lib/autofs/automountd'</td>
<td>E1001 File path does not exist on Linux</td>
</tr>
<tr>
<td>17</td>
<td>'umountall'</td>
<td>E7001 comparable Linux command might be &quot;umount -a&quot;</td>
</tr>
</tbody>
</table>

+++ End of report for 'autofs'
Sizing a Migration Project 1/2

• A rough classification of the entries in the Source Checking Tool's knowledge database:
  • 46% of all calls are identical in Linux and Solaris  
    e.g. mathematical functions found in math.h  
  • 8% require trivial changes  
    e.g. name of a function is different  
  • 6% require changes in local program context  
    e.g. different order of function arguments  
  • 15% require major non-local changes  
    ...in case of different semantics  
  • 25% need to be assessed in application context

• Result from one large real-world project: 
  Source Checking Tool reported one finding / 400 LOC
Sizing a Migration Project 2/2

• What is the effort for a migration project?
  • Probably considerably less than its initial development
  • Actual cost will strongly depend on the individual project

• Migration effort strongly depends on the portability of the source code:
  • Relying on standardized libraries only will reduce the effort
  • Proper program organization and modular structuring will reduce the effort
  • Using libraries specific for a proprietary OS will cost additional effort
  • Performance optimizations based on particular properties of an OS will cost additional effort
Supported Versions

- Operating system versions covered by the Migration Kit:
  - Solaris versions: 8 and 9
  - Linux Kernel 2.6
  - Including libraries and compilers usually used with these versions

- The tools provided with the Migration Kit will work on computers running Solaris version 8 or 9

- The format of the included documentation is PDF
How to obtain the Migration Kit for Solaris OS to Linux

The toolkit is available free of charge from the following URL:

 Trademarks 1/2

The following terms are registered trademarks of International Business Machines Corporation in the United States and/or other countries: AIX, AIX/L, AIX/L(logo), alphaWorks, AS/400, Blue Gene, Blue Lightning, C Set++, CICS, CICS/6000, CT/2, DataHub, DataJoiner, DB2, DEEP BLUE, developerWorks, DFDSM, DirectTalk, DYNIX, DYNIX/ptx, e business(logo), e(logo)business, e(logo)server, Enterprise Storage Server, ESCON, FlashCopy, GDDM, IBM, IBM(logo), ibm.com, IBM TotalStorage Proven, IntelliStation, IQ-Link, LANStreamer, LoadLeveler, Lotus, Lotus Notes, Lotusphere, Magstar, MediaStreamer, MQSeries, Net.Data, Netfinity, NetView, Network Station, Notes, NUMA-Q, Operating System/2, Operating System/400, OS/2, OS/390, OS/400, Parallel Sysplex, PartnerLink, PartnerWorld, POWERparallel, PowerPC, PowerPC(logo), Predictive Failure Analysis, pSeries, PTX, ptx/ADMIN, RISC System/6000, RS/6000, S/390, Scalable POWERparallel Systems, SecureWay, Sequent, ServerProven, SP1, SP2, SpaceBall, System/390, The Engines of e-business, THINK, ThinkPad, Tivoli, Tivoli(logo), Tivoli Management Environment, Tivoli Ready(logo), TME, TotalStorage, TURBOWAYS, VisualAge, WebSphere, xSeries, z/OS, zSeries.

The following terms are trademarks of International Business Machines Corporation in the United States and/or other countries: Advanced Micro-Partitioning, AIX/L(logo), AIX 5L, AIX PVMe, AS/400e, BladeCenter, Chipkill, Cloudscape, DB2 OLAP Server, DB2 Universal Database, DFDSM, DFSORT, Domino, e-business(logo), e-business on demand, eServer, GigaProcessor, HACMP, HACMP/6000, Hypervisor, i5/OS, IBMLink, IBM Virtualization Engine, IMS, Intelligent Miner, Micro-Partitioning, iSeries, NUMACenter, OpenPower, POWER, Power Architecture, Power Everywhere, PowerPC Architecture, PowerPC 603, PowerPC 603e, PowerPC 604, PowerPC 750, POWER2, POWER2 Architecture, POWER3, POWER4, POWER4+, POWER5, POWER5+, POWER6, Redbooks, Sequent (logo), SequentLINK, Server Advantage, ServeRAID, Service Director, SmoothStart, SP, S/390 Parallel Enterprise Server, ThinkVision, Tivoli Enterprise, TME 10, TotalStorage Proven, Ultramedia, VideoCharger, Visualization Data Explorer, X-Architecture, z/Architecture.

A full list of U.S. trademarks owned by IBM may be found at: http://www.ibm.com/legal/copytrade.shtml.
Trademarks 2/2

UNIX is a registered trademark in the United States and other countries licensed exclusively through The Open Group. Linux is a registered trademark of Linus Torvalds in the United States, other countries or both. Microsoft, Windows, Windows NT and the Windows logo are registered trademarks of Microsoft Corporation in the United States and/or other countries. Intel, Itanium and Pentium are registered trademarks and Intel Xeon and MMX are trademarks of Intel Corporation in the United States and/or other countries. AMD Opteron is a trademark of Advanced Micro Devices, Inc. TPC-C and TPC-H are trademarks of the Transaction Performance Processing Council (TPPC). SPECint, SPECfp, SPECjbb, SPECweb, SPECjAppServer, SPEC OMP, SPECviewperf, SPECapc, SPEChpc, SPECjvm, SPECmail, SPECimap and SPECsfs are trademarks of the Standard Performance Evaluation Corp (SPEC). NetBench is a registered trademark of Ziff Davis in the United States, other countries or both. Sun, Sun Microsystems, the Sun logo, iForce, Java, Netra, N1, Solaris, Sun Fire, Sun Ray, SunSpectrum, Sun StorEdge, SunTone, The Network is the Computer, all trademarks and logos that contain Sun, Solaris, or Java, and certain other trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. in the United States and other countries. Products bearing SPARC trademarks are based upon an architecture developed by Sun Microsystems, Inc.

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