Software Environment

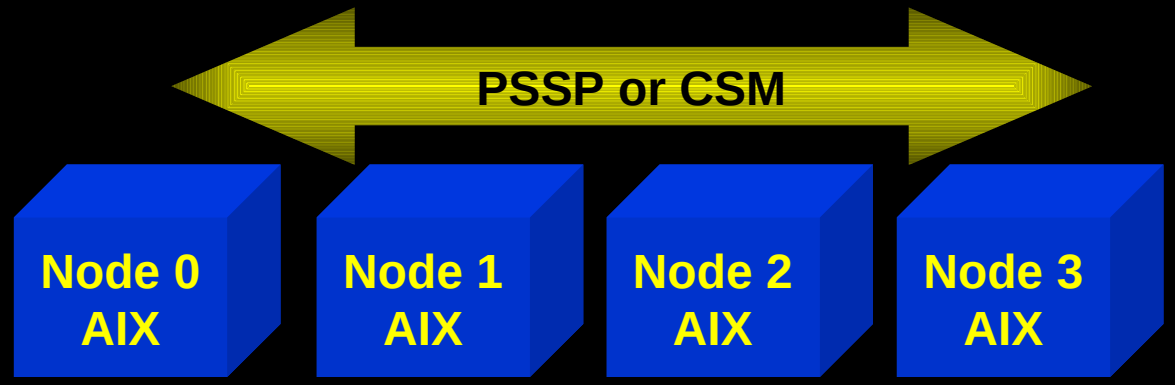
Some Useful System Commands

Agenda

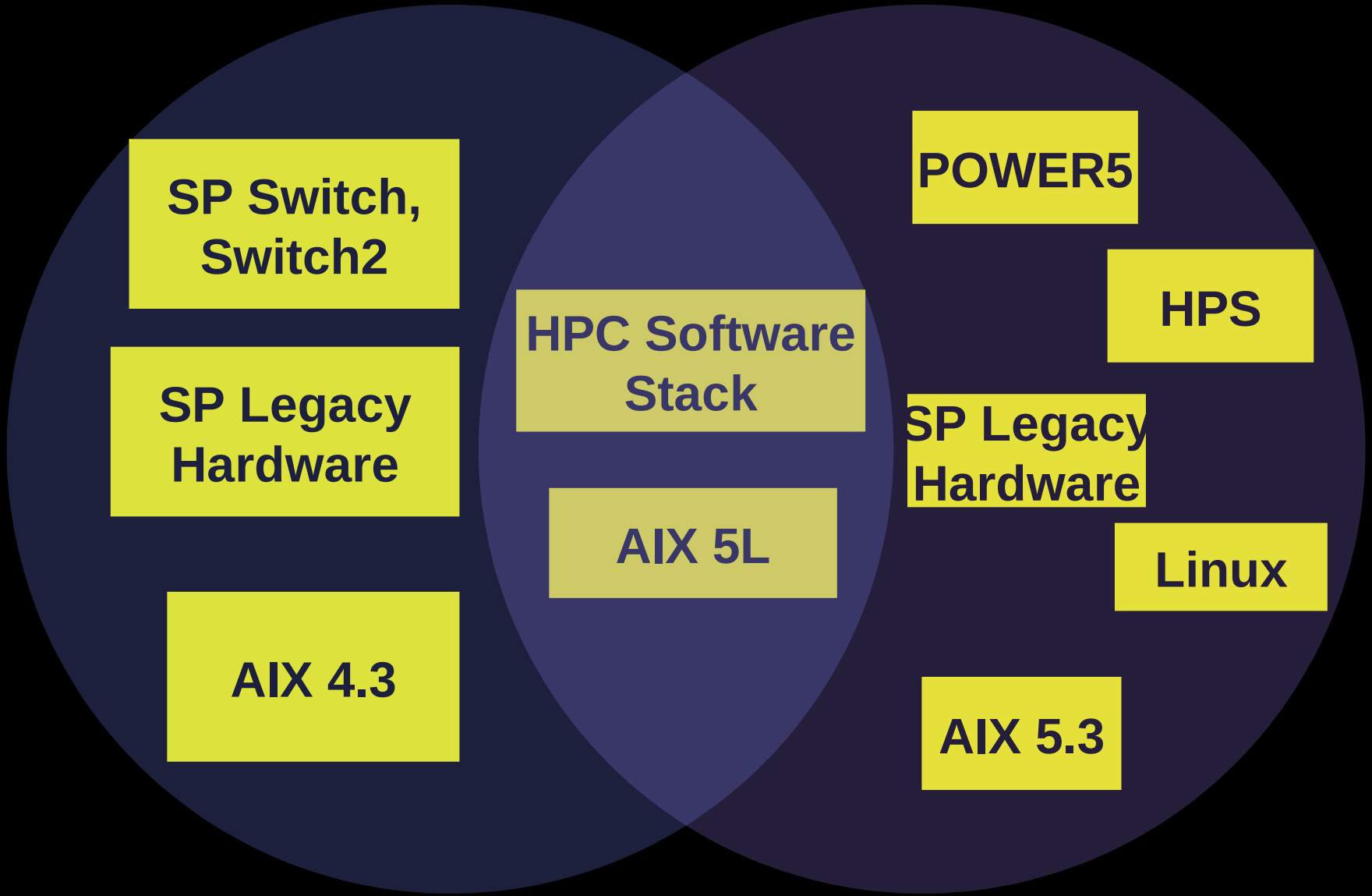
- Background on AIX and HPC Software Stack
- Comments on 64-bit vs 32-bit address
- Getting know your system
- Useful commands

Operating System Software

- Server OS
  - AIX (Advanced Interactive eXecutive)
    - AT&T System V
  - Journaled
- Cluster System Management (CSM)
  - Management of distributed or clustered servers
- Parallel System Support Program (PSSP)
  - Predecessor to CSM
- Parallel Environment (PE)



PSSP and CSM



HPC Software Stack

- Batch queuing:
  - LoadLeveler
- Parallel file system:
  - General Parallel File System (GPFS)
- Math library:
  - Engineering and Scientific Software Library (ESSL)
- MPI tools and library:
  - Parallel Environment (PE)

AIX Operating System

- AIX 5L
  - Current (new) generation of IBM's Unix
    - Linux “affinity”
    - Combines Unix technologies of AIX and Linux
- Current Versions:
  - AIX 5L version 5.2
  - AIX 5L version 5.3 <-

AIX Characteristics

- Journaled file system
  - JFS and JFS2
  - File coherency
- Other AIX'isms and terms
  - LPP - Licensed Program Product (/usr/lpp/...)
  - BOS - Base Operating System (bos.rte, bos.up, etc)
  - Administration:
    - PTF- A specific software patch
    - APAR - A software fault or enhancement description
    - EFIX -An emergency software fix, invalidated
- Note: New process for delivering fixes
  - Sets of fully tested combinations of updates

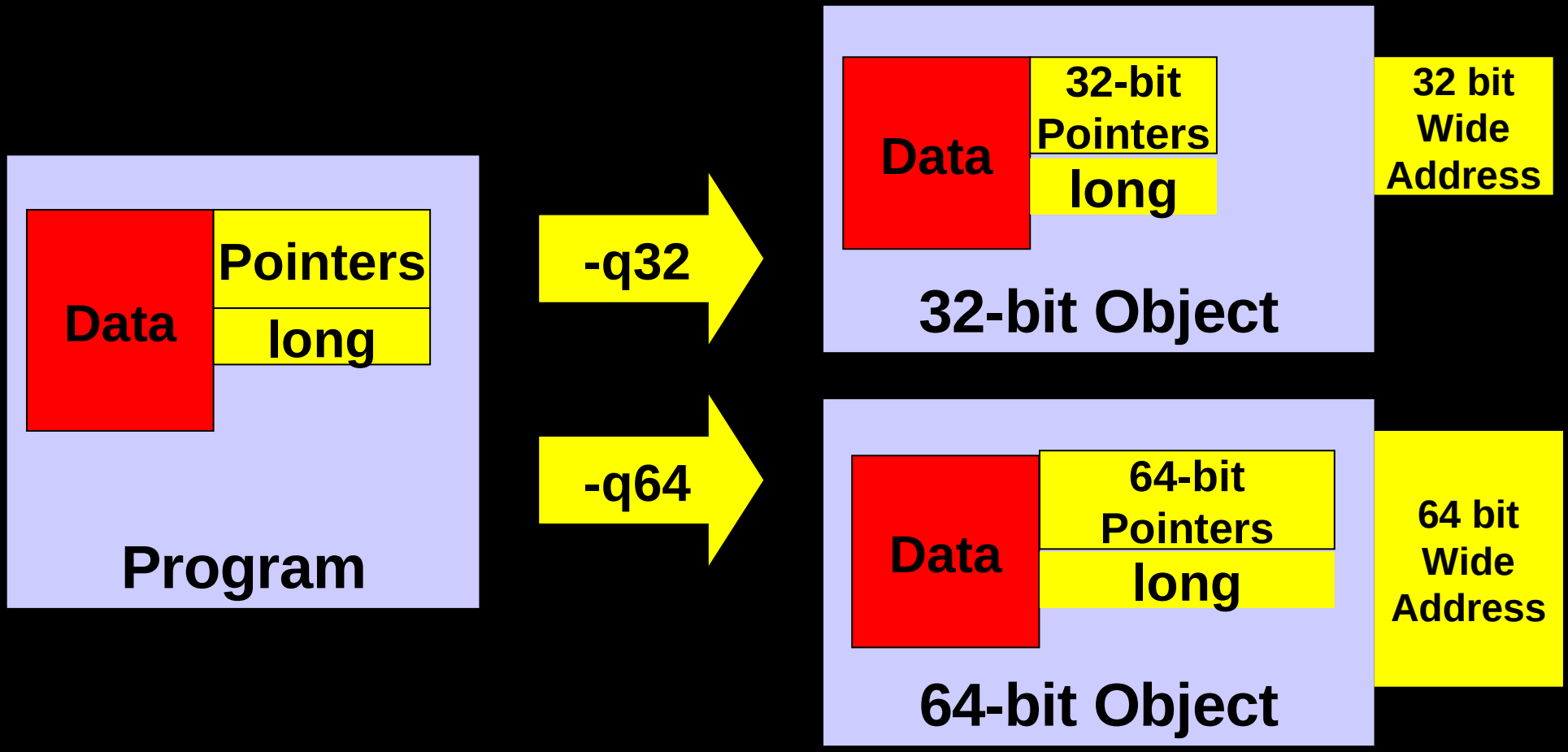
Linux Affinity

- AIX bias toward “Linux – like” environment
- Emerging Linux applications
- GNU tools
- GNU utilities
- Linux look & feel

64-bit Operating System

- Operating system address modes:
  - 32-bit kernel
    - Limitation: 96 Gbyte memory
    - Built -qarch=com
  - 64-bit kernel
    - No limitations
    - Built -qarch=ppc
- Hardware:
  - 64-bit design
- Software:
  - 32-bit model (default)
  - 64-bit model

Application Address Mode



Application address mode is independent of operating system address mode

Even more on 64-bit...
(because it is so often confused)

- 64-bit floating point representation is higher precision
  - Fortran: REAL*8, DOUBLE PRECISION
  - C/C++: double
  - You can use 64-bit floating point with –q32 or –q64

- 64-bit addressing is totally different. It refers to how many bits are used to store memory addresses and ultimately how much memory one can access.
  - Compile and link with –q64
  - Use `file a.out myobj.o` to query addressing mode

- The AIX kernel can be either a build that uses 32-bit addressing for kernel operations or uses 64-bit addressing, but that does not affect an application's addressibilty.
  - ls –l /unix to find out which kernel is used
  - Certain system limits depend on kernel chosen

One more thing about 64-bit…

- If you use –q64:
  - You can use lots of memory
  - INTEGER*8 or long long operations are faster
- If you use –q32:
  - You may run a few percent faster
    - Fewer bytes are used storing and moving pointers
  - You will have to learn about –bmaxdata
    - AIX link option
    - -bmaxdata:0x10000000 = 256 Mbyte = default
    - -bmaxdata:0x80000000 = 2 Gbyte
    - -bmaxdata:0xC0000000 = not widely publicized trick to use more than 2 Gbyte with –q32
      - “C” is the maximum
  - -q64
    - –bmaxdata:0 = default = unlimited
    - Other –bmaxdata values will be enforced if set

Getting to know your system

- uname –a
- oslevel –f -r
- df
  - lspv ...
  - lslv ...
  - lsvg ...
- ifconfig –a; netstat –I; vmstat; iostat
- lsdev –C | grep proc | wc –l
- lsattr -E –l proc0 –a type
- /usr/bin/pmcycles
  - Install bos.pmapi
- lsattr -E -l mem0 -a size
- prtconf

Hardware Configuration
lscfg: Installed resource list

```
/home/myuid$ lscfg
INSTALLED RESOURCE LIST

The following resources are installed on the machine.
+/- = Added or deleted from Resource List.
*   = Diagnostic support not available.

  Model Architecture: chrp
  Model Implementation: Multiple Processor, PCI bus

+ sysplanar0        00-00          System Planar
+ mem0              00-00          Memory
+ proc0             00-00          Processor
+ L2cache0          00-00          L2 Cache
```

Software Configuration
lslpp: Installed Software

```
/home/myuid$ lslpp -L
 Fileset                 Level  State  Description
 ----------------------------------------------------------
 Adobe.acrobat    3.0.1.0    C    Adobe acrobat reader
 DB2V5CAE.Bnd     5.2.0.0    C    DB2 Client Application
 DB2V5SERV.Bnd    5.2.0.0    C    DB2 Server(s) Software
 IBMVJava.dab.adt 2.0.0.0    C    VisualAge for Java
 IBMVJava.dab.rte 2.0.0.0    C    VisualAge for Java
```

Configuration Report

- prtconf
  - Print Configuration
  - Standard Unix command
- Information
  - Processors
  - Memory
  - Operating system

prtconf

```
$ prtconf -ckLms
CPU Type: 64-bit
Kernel Type: 32-bit
LPAR Info: 1 NULL
Memory Size: 131072 MB
Processor Clock Speed: 1900 MHz
```

Performance Monitors

- System (node) performance
  - topas
    - Similar to Linux “top”
    - Root user has to invoke first time to create a file in /etc
  - nmon
    - Freeware from IBM UK
  - vmstat
    - virtual memory statistics

Topas

```
magnet                averages:  7.90,  7.90,  7.38
Cpu states:        ...system,   0.0% wait,  50.0% idle
Real memory:       ...procs  512.0M files 2544.0M total
Virtual memory:    ... used   150.2M total

   PID USER      ...STAT      TIME   CPU% COMMAND
 30972 myusrid   ... run       0:22  50.0% lu.W
   516 root      ... run     591:16  49.5% Kernel (wait)
 23950 myusrid   ... run       0:00   0.4% monitor4.1
   774 root      ... run     568:46   0.0% Kernel
```

nmon

- Performance tuning utility
- Freeware, AIX and Linux
- Performance data:
  - CPU utilization
  - Memory use
  - Kernel statistics and run queue information
  - Disks I/O rates, transfers, and read/write ratios
  - Free space on file systems
  - Disk adapters
  - Network I/O rates, transfers, and read/write ratios
  - Paging space and paging rates
  - Etc.

Virtual Memory STATistics: vmstat

- System (node) resources
  - Memory
  - Page faults
  - CPU

```
$ vmstat 1
kthr      memory   faults    cpu
----- ----------- ------ -----------
 r   b    avm    fre    ...  us sy id wa
 0   0 70893 364069   ...   3  2 92  2
24   1 70894 364068   ...  35 65  0  0
25   2 70895 364067   ...  37 63  0  0
```

VMSTAT

vmstat: virtual memory statistics.

Syntax: vmstat n [m]n = refresh interval, optional m = count

vmstat [interval [count]]

vmstat –s (summary)

[v07n20:/u/user] vmstat 1 5

kthr		memory		page						faults			cpu			
r	b	avm	fre	re	pi	po	fr	sr	cy	in	sy	cs	us	sy	id	wa
0	1	51066	445573	0	0	0	0	0	0	201	763	343	41	3	56	0
0	2	51066	445573	0	0	0	0	0	0	213	1209	51	0	0	99	0
0	2	51066	445573	0	0	0	0	0	0	209	71	44	0	0	99	0
1	2	49902	444016	0	0	0	0	0	0	222	373	76	42	4	54	0
1	2	49902	444016	0	0	0	0	0	0	213	85	49	50	0	50	0

r = kernel threads placed on the run queue
b = kernel threads blocked
avm = active virtual memory pages (1 page = 4KBytes)
fre = free memory pages
pi = page-ins from paging space
po = page-outs to paging space
fr = pages freed
sr = pages scanned
cy = scan cycles
in = device interrupts
sy = system calls
cs = context switches
us = user cpu utilization %
sy = system cpu utilization %
id = cpu idle time
wa = time waiting on I/O

NETSTAT

netstat : network statistics.

Syntax : netstat n [m] n = refresh interval, optional m = count
- netstat –i –f inet (lists internet interfaces)
- netstat –I css0 (switch status)
- netstat -I css0 interval (switch IP traffic)
- netstat –D (packet counts)

[v01n14:/u/user] netstat -Icss0 1

```
input   (css0)      output           input   (Total)     output
packets  errs  packets  errs colls  packets  errs  packets  errs colls

 6254389   0     6364710   0       0    17660924   0    17474352 183221   0
       1   0           1   0       0           4   0           3      0   0
    2127   0        2134   0       0        2139   0        2147      0   0
    1041   0        1117   0       0        1055   0        1130      0   0
```

Parallel Environment (PE)

- Develop, debug, analyze, tune, and execute parallel applications
  - Parallel Operating Environment (POE)
  - MPI
    - Optimized for IBM switches and nodes
  - pdbx (Parallel Debugger)
    - Attach to running process
  - Parallel utilities, for easing file manipulation

PE Example

```
**************************************************************
c* Hello World Fortran Example
c To compile: mpxlf_r -o hello_world_f hello_world.f f
c**************************************************************
c
      program hello
      implicit none
      write(6,*) 'Hello, World!'
      stop
      end
```

PE Example

Host.list:

Node1

Node1

Node2

Node2

$ xlc –o hello_world hello_world.c

$ export MP_HOSTFILE=$PWD/host.list

$ poe hello_world_f -procs 4

Hello, World!

Hello, World!

Hello, World!

Hello, World!

System Management Installation Tool (SMIT or SMITTY)

```
Move cursor to desired item and press Enter.

  Software Installation and Maintenance
  Software License Management
  Devices
  System Storage Management (Physical & Logical
  Storage)
 ...
  Problem Determination
  Performance & Resource Scheduling
..
  Applications

F1=Help                 F2=Refresh              F3=Cancel
     F8=Image
F9=Shell                F10=Exit                Enter=Do
```

Essential Commands

- Software interrogation
- Hardware interrogation
- System management