

Towards a Global-Scale Public Computer

Project Kittyhawk at IBM Research

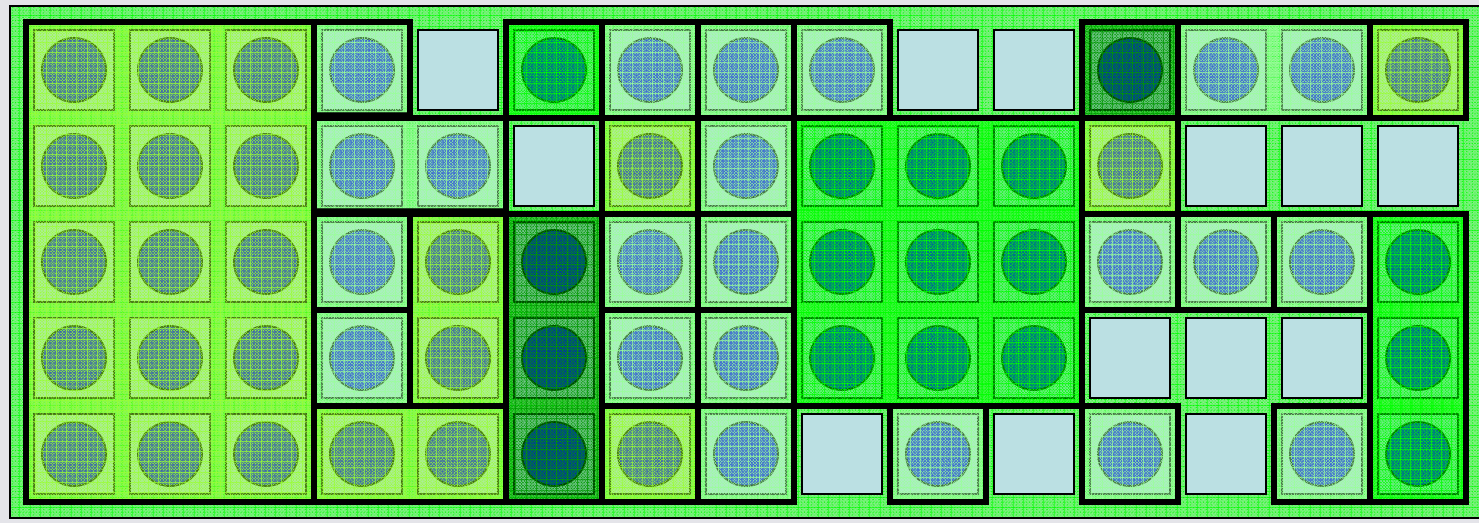
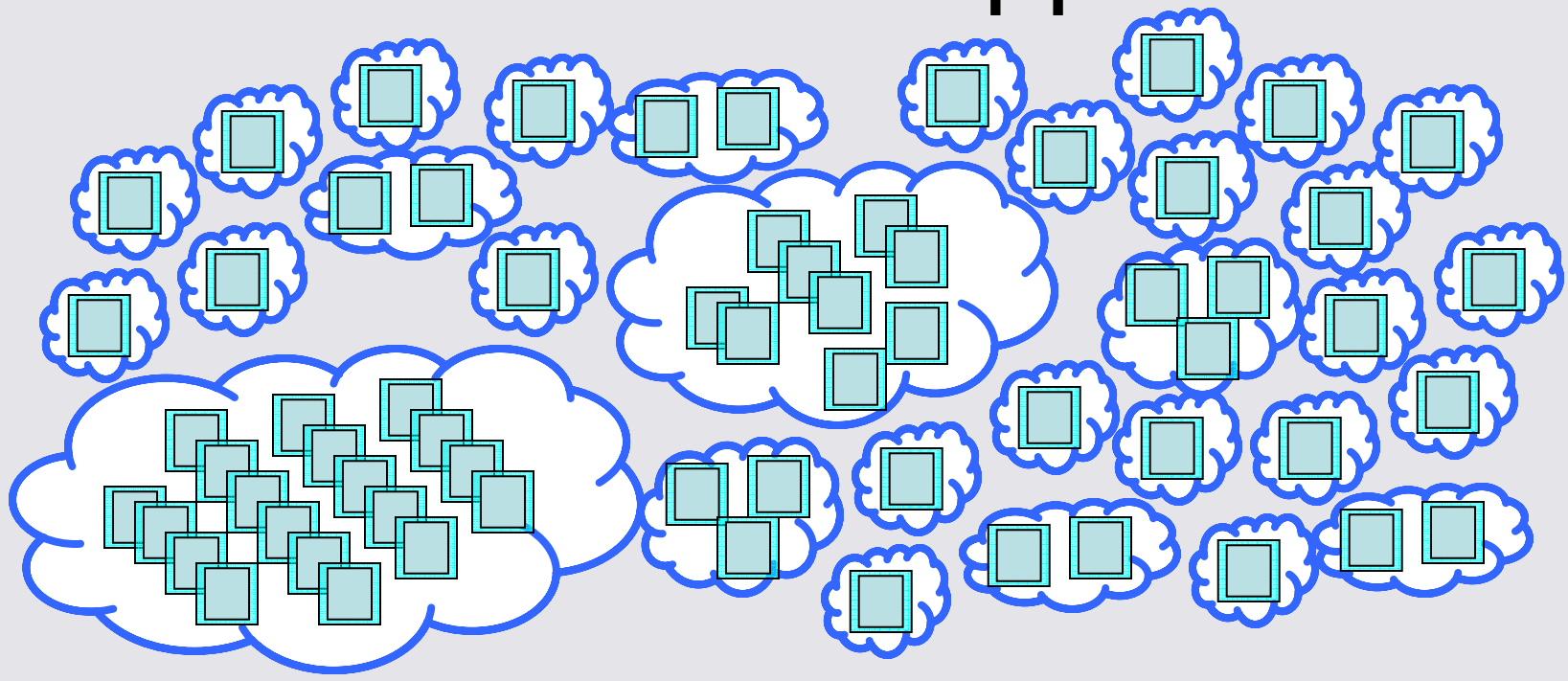
Presented by Jonathan Appavoo representing the Project Kittyhawk Team
Jonathan Appavoo, Volkmar Uhlig, Amos Waterland, Bryan Rosenberg

IBM T. J. Watson Research Center, New York

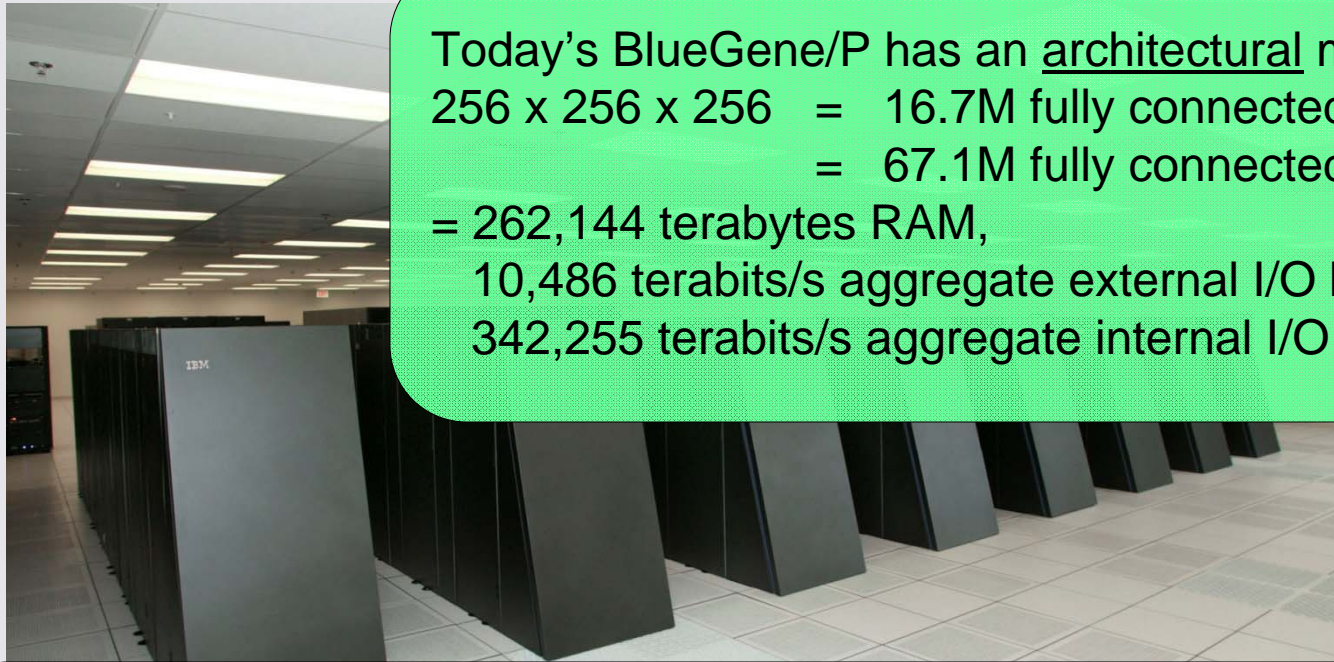
A widely accessible, scalable computer that represents a significant fraction of the world's current global computational capacity.

Kittyhawk
=
Global Computation
+
Massive Parallel Processor

Internet as an Application



Massive Parallel Processor



Today's BlueGene/P has an architectural maximum size of:
 $256 \times 256 \times 256 = 16.7\text{M}$ fully connected nodes
 $= 67.1\text{M}$ fully connected cores
 $= 262,144$ terabytes RAM,
 $10,486$ terabits/s aggregate external I/O bandwidth,
 $342,255$ terabits/s aggregate internal I/O bandwidth.

“As of July 1, 2006, the population of the City of New York was 8,250,567”

<http://www.nyc.gov/html/dcp/html/census/popcur.shtml>

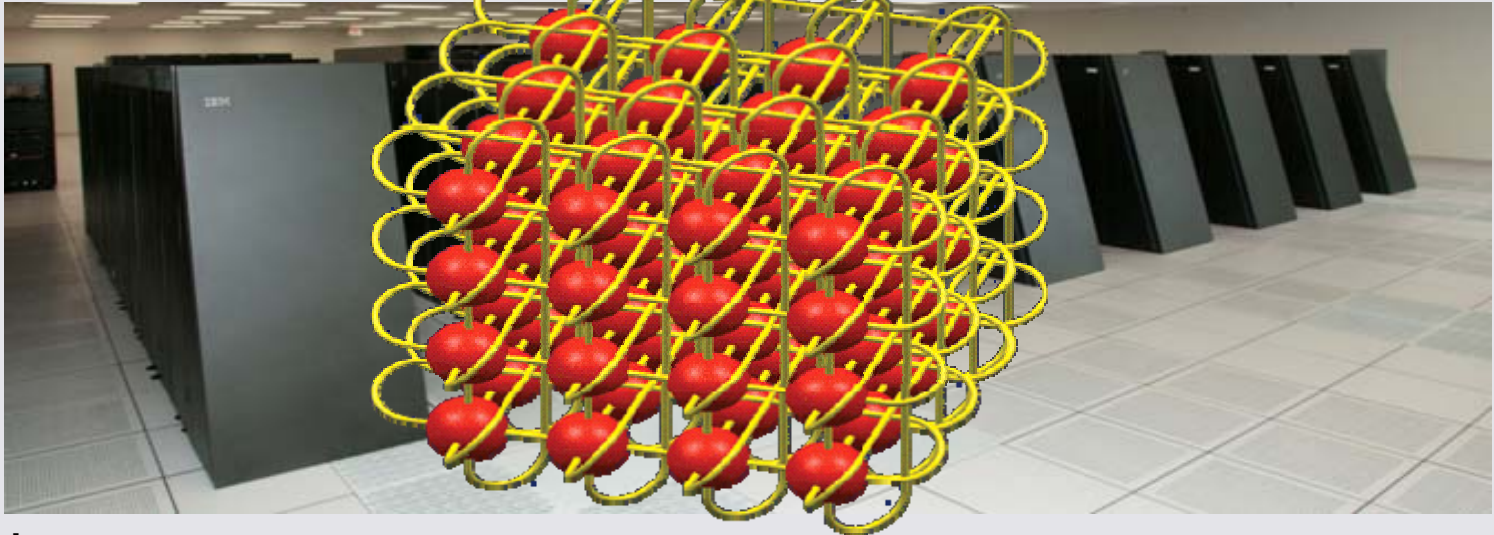
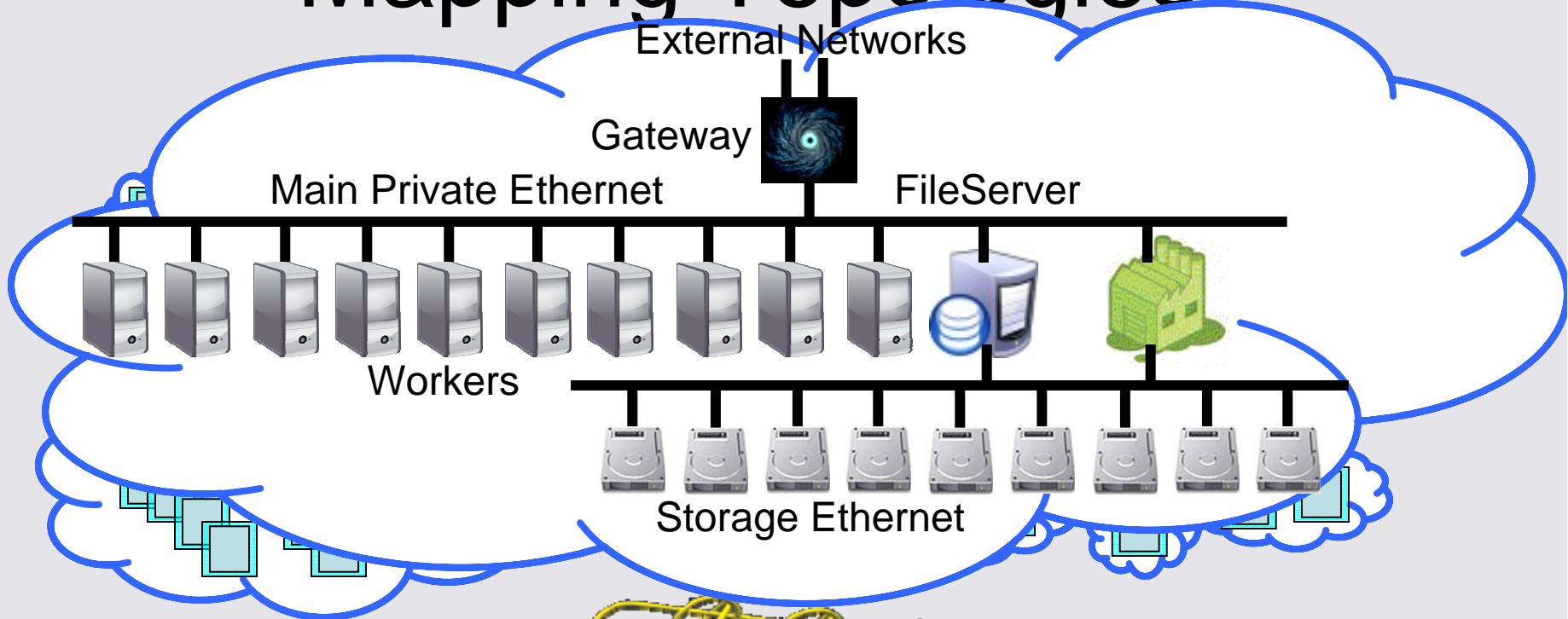
Nodes per capita: ~ 2 Cores per capita: ~ 8

2005 Total US Volume Servers (<\$25,000 per unit) = 9,897,000

Jonathan G. Koomey, “ESTIMATING TOTAL POWER CONSUMPTION BY SERVERS IN THE U.S. AND THE WORLD”, Staff Scientist, Lawrence Berkeley National Laboratory and Consulting professor, Stanford University, Final report February 15, 2007

Nodes per server: ~ 1.6

Mapping Topologies

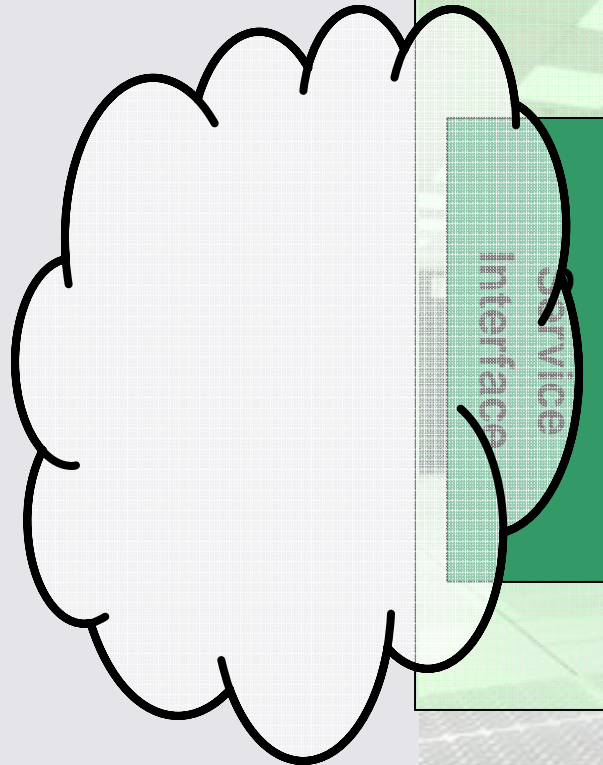


Abstract System View

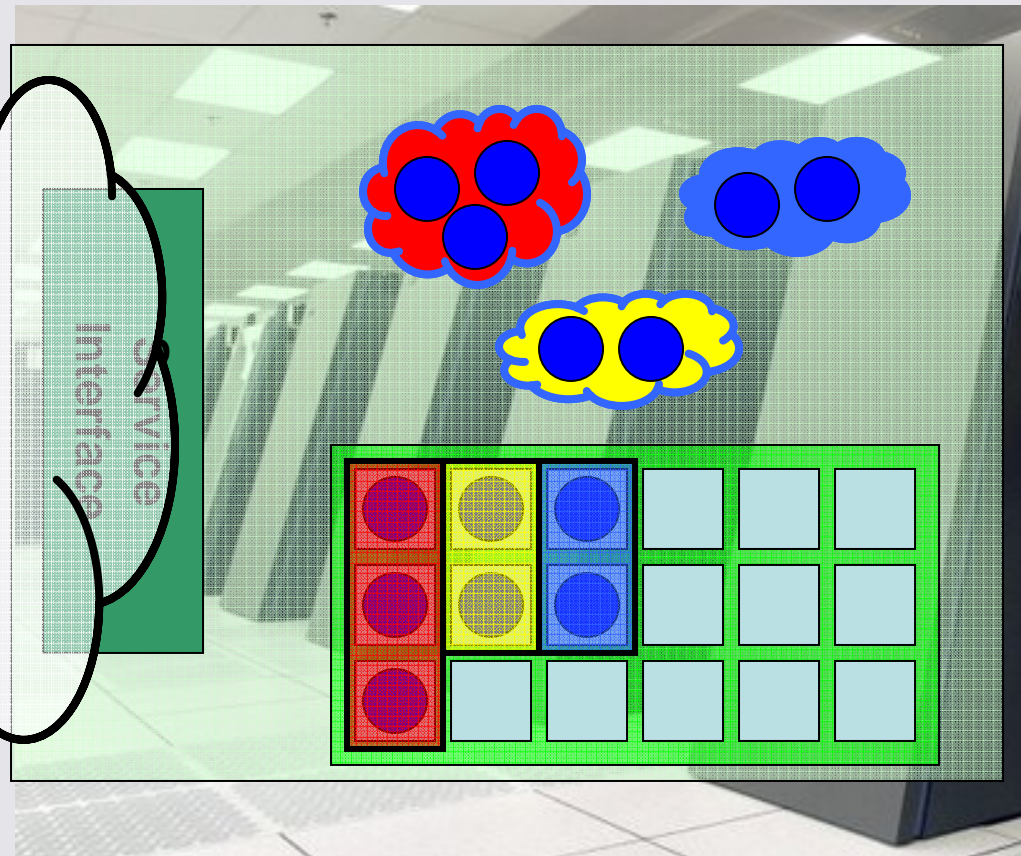
Users
(developers,
service providers,
resellers)



**Public Access
Network**



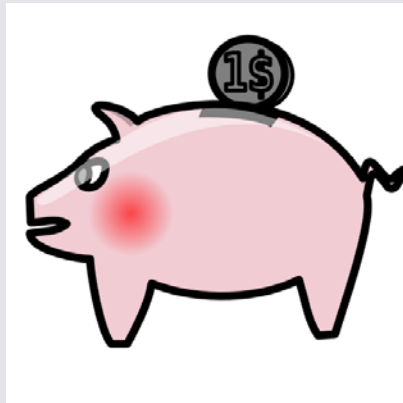
The Computer



The Key Points of this Talk

1. Large-scale communication-centric system through **Aggressive Integration.**
2. Fairness through **Raw Hardware Access.**
3. Competition and cooperation through **Dynamic, Hardware-enforced Communication Domains**

Benefits of the Approach



Outline

1. A global computer & global computation.
2. Our prototype.
3. Why hardware-centric? When virtual is bad!
4. Our digital future -- by accident or design?

Our Take on a Global-scale Computer and Global Computation

A Global Scale Computer: A well-specified public, “software-less”, massively-parallel system, on which users can construct services, of arbitrary scale, out of metered and billed common units of its capacity grouped in domains of communication they specify and control.

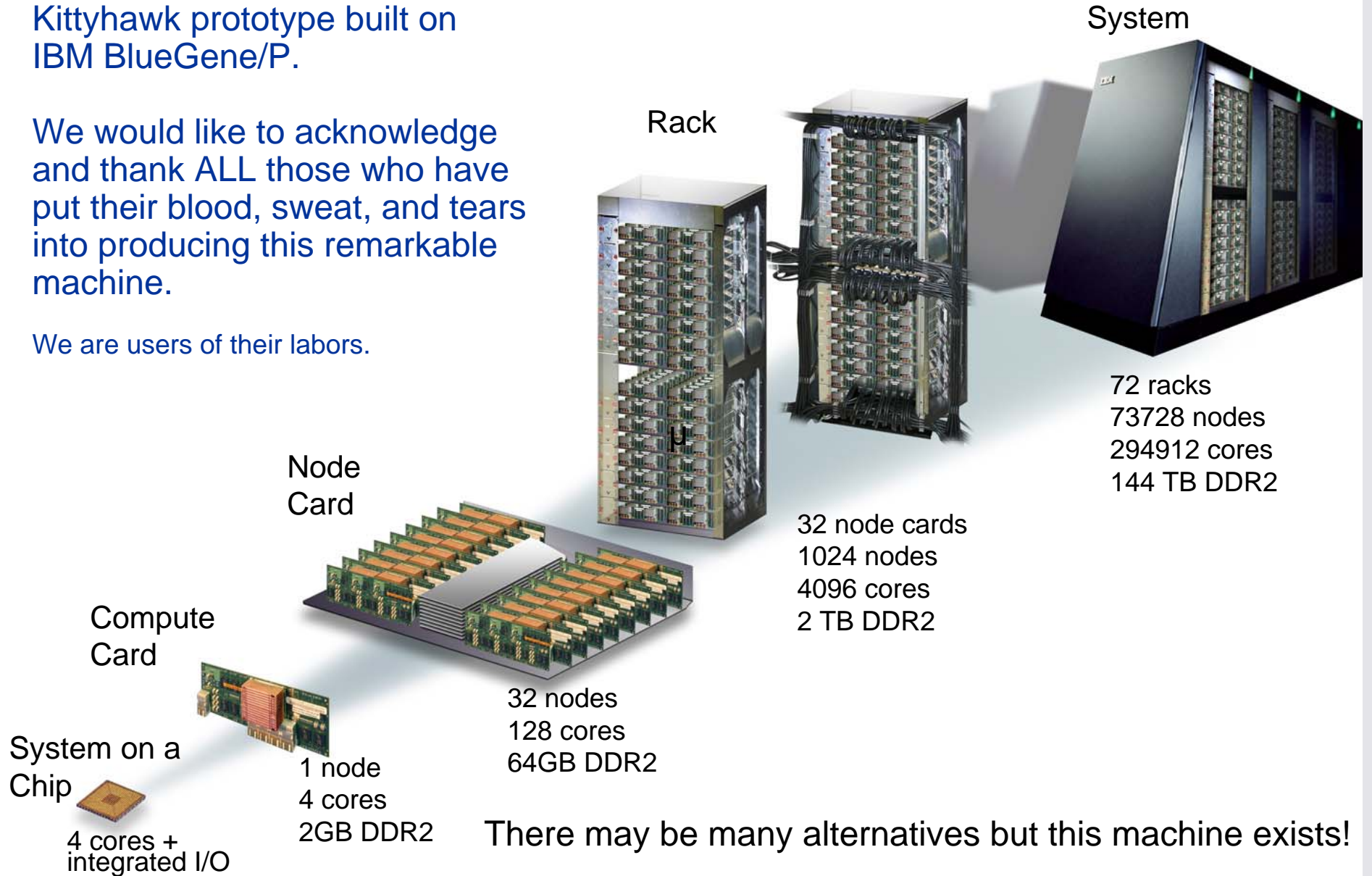
Our Prototype

1. Massive parallel processor: BlueGene/P
2. Evolution toward a global-scale computer
3. Demos

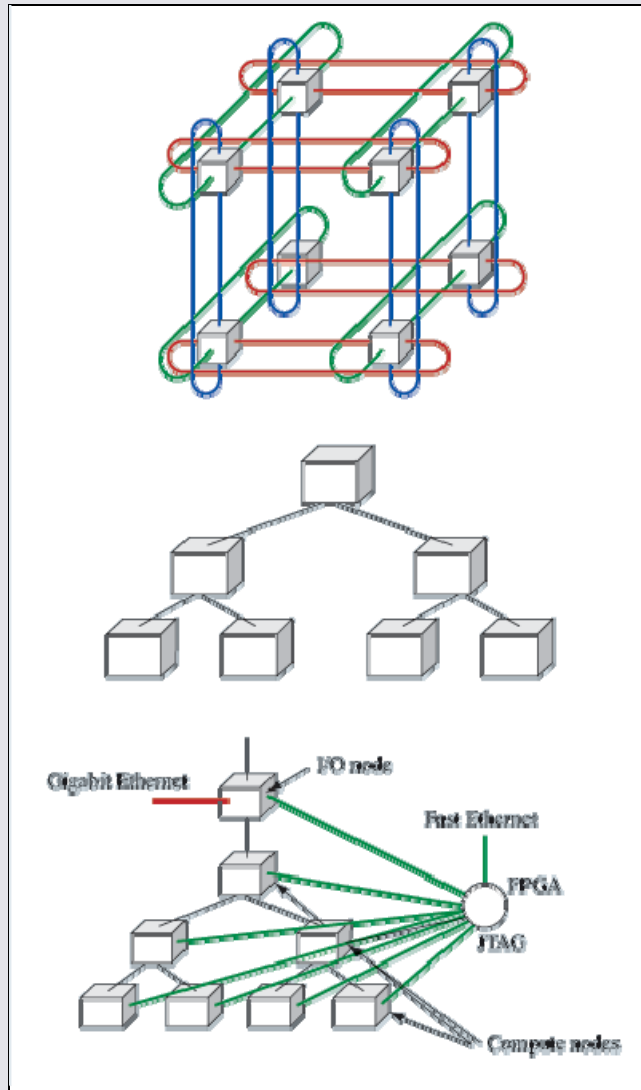
Kittyhawk prototype built on IBM BlueGene/P.

We would like to acknowledge and thank ALL those who have put their blood, sweat, and tears into producing this remarkable machine.

We are users of their labors.



5 for the price of 1



1) 3 Dimensional Torus

- Interconnects all compute nodes.
- Adaptive cut-through hardware routing.
- 3.4 gigabits/s on all 12 node links (5.1 gigabits/s per node).
- 0.5 μ s latency between nearest neighbors.
- 1.7/2.6 terabits/s bisectional bandwidth, 188 terabits/s total bandwidth (72k machine).

2) Collective Network

- One-to-all broadcast functionality.
- Reduction operations for integers and doubles.
- 6.8 gigabits/s of bandwidth per link per direction.
- Latency of one-way tree traversal 1.3 μ s.
- Interconnects all compute nodes and I/O nodes.
- ~62 terabits/s total binary tree bandwidth (72k machine).

3) Low Latency Global Barrier and Interrupt

- Latency of one-way signal to all 72K nodes 0.65 μ s.

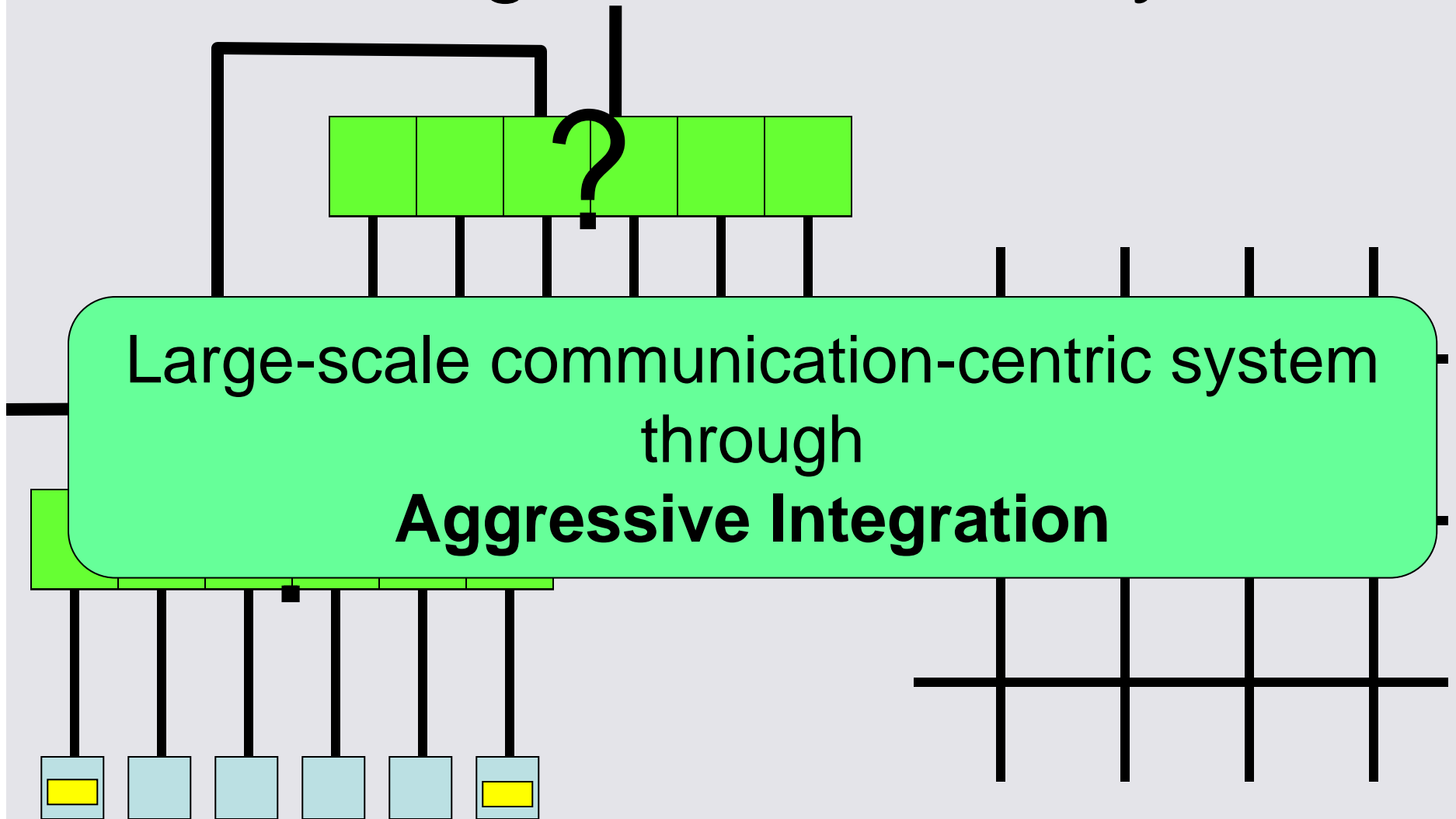
4) 10Gig Functional External I/O Ethernet Network

- I/O nodes only.

5) 1Gig Private Control Ethernet

- Provides JTAG access to hardware.
- Accessible only from Secure Service Node System.

Integrated Scalability



Massively Scalable Networking

Every node in the system can be identified by a unique x,y,z coordinate:

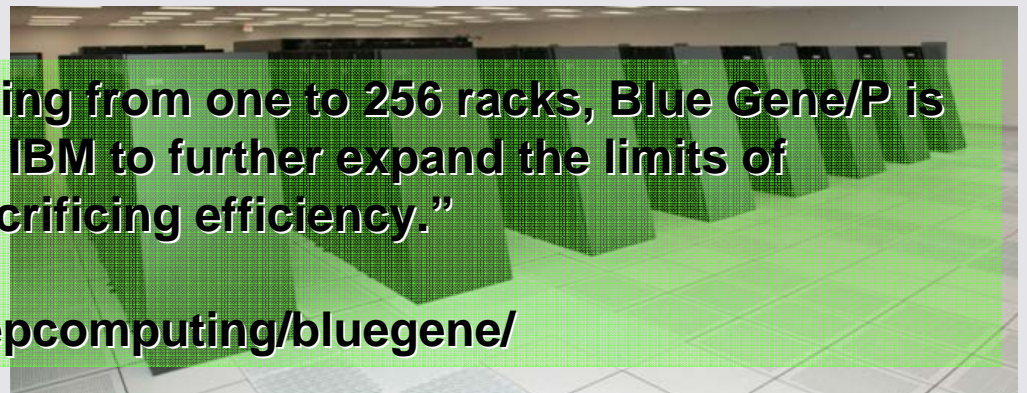
1. Fabric scales with nodes: nodes **form** the fabric.
2. Cut-through routing: fabric cooperates to route a message point-to-point with **no** software involvement.
3. Global, symmetric, and transparent: spans all hardware domains; node cards, mid-planes, and **racks**.

Scale

Racks	1	8	64	256
Nodes	1,024	8,192	65,536	262,144
Cores	4,096	32,768	262,144	1,048,576
Internal aggregate bandwidth gigabits/s	20,890	167,117	1,336,934	5,347,738
External aggregate bandwidth gigabits/s	640	5,120	40,960	163,840
RAM in terabytes (current limit / architectural max)	2 (4/16)	16 (32/128)	128 (256/1024)	512 (1024/4096)
Peak power consumption (kW)	35	280	2240	8960

“Available in configurations ranging from one to 256 racks, Blue Gene/P is the innovative new solution from IBM to further expand the limits of breakthrough science without sacrificing efficiency.”

<http://www.ibm.com/systems/deepcomputing/bluegene/>



Space and Power Consumption for a Million Cores

Mega Data Center with Standard Servers



- Power: 50 MW (~ \$500 million)
- Floor space: 500,000 square feet

BlueGene/P



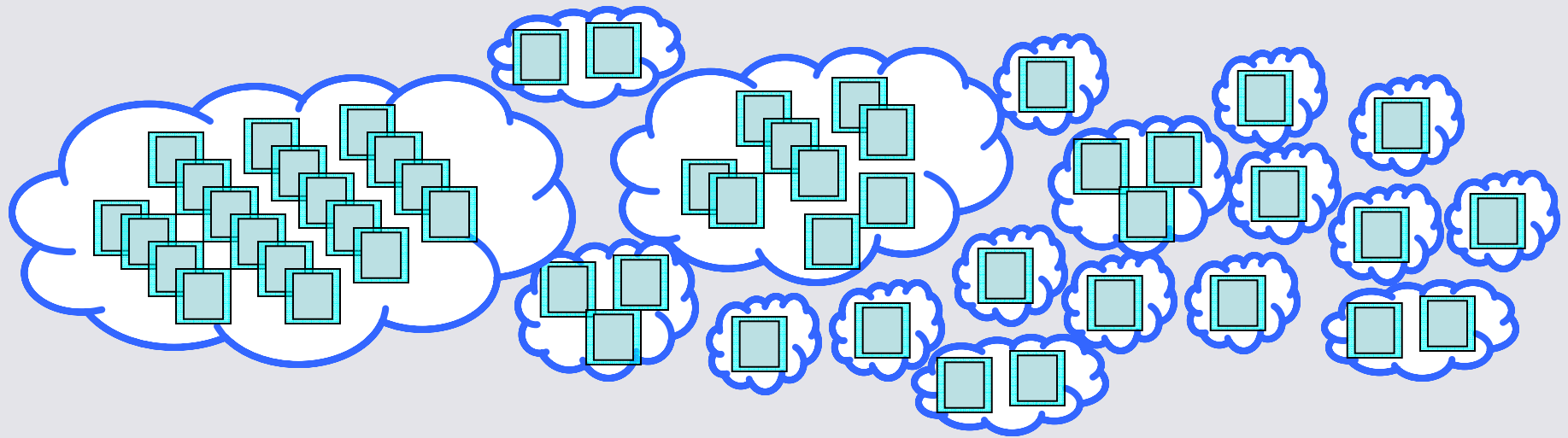
- Power: 10 MW (~ \$100 million)
- Floor space: 10,000 square feet

Evolution towards a Global-scale Computer

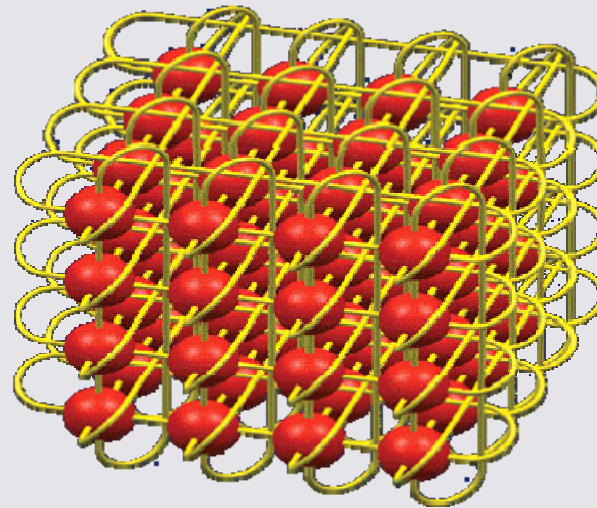
"A parallel computer with a billion processors might provide a computational utility analogous to existing gas and electric utilities."

W. Daniel Hillis, "The Connection Machine",
Scientific American 256, 6 (June 1987), 108--115.

Supporting Global Computation



Large-scale
Communication-
centric System



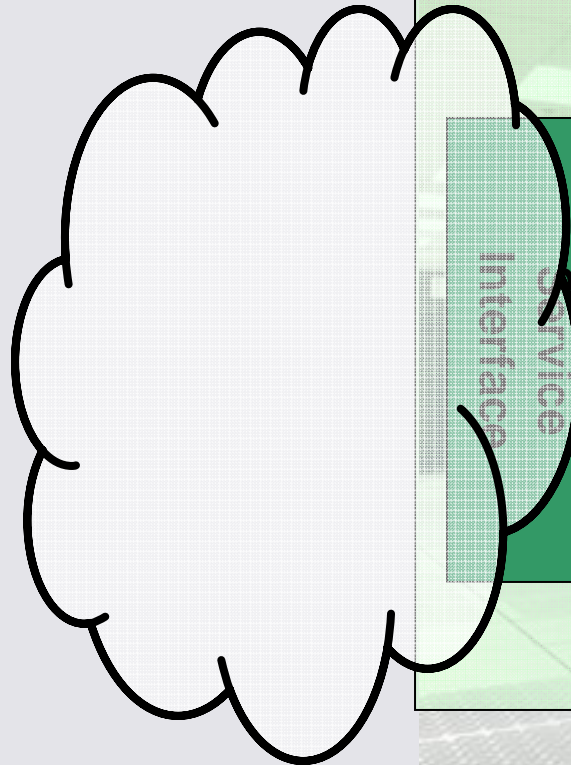
Aggressive
Integration

Abstract View

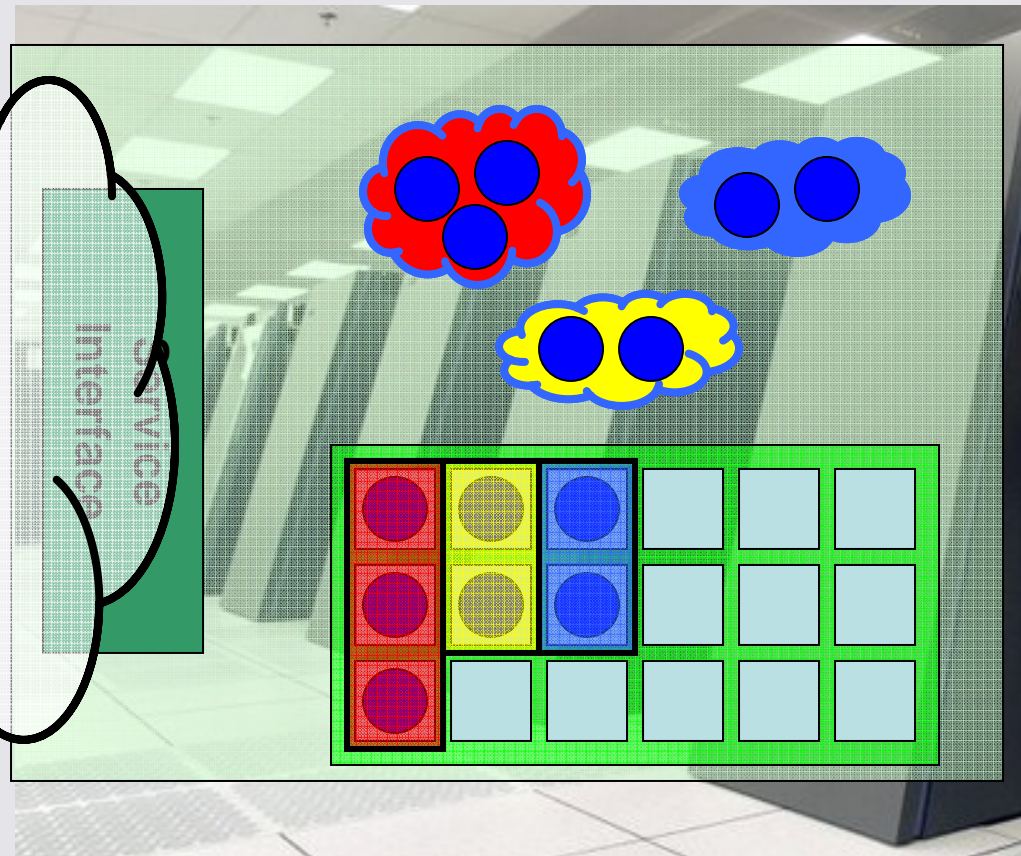
Users
(developers,
service providers,
resellers)



**Public Access
Network**

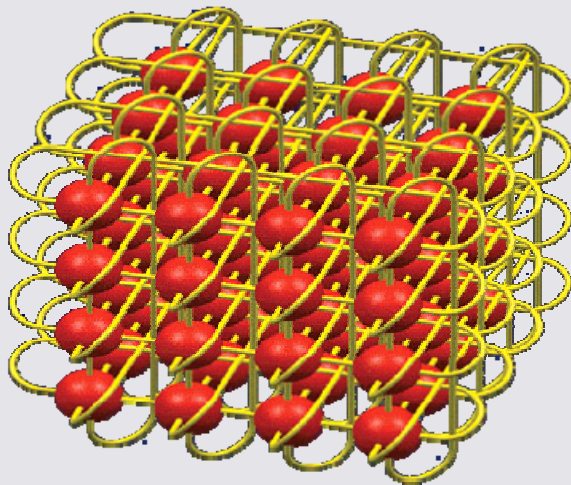


The Computer



A Global-scale Computer

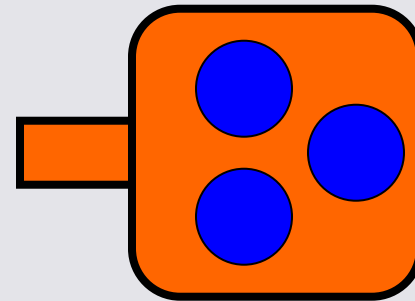
Large-scale
Communication-centric
System



Aggressive Integration

+

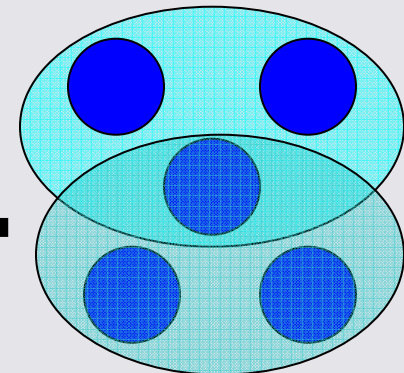
Control
Channels



Raw Hardware
Access

+

Communication
Domains

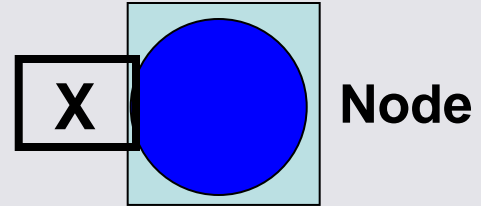


Primitive for
Competition &
Cooperation

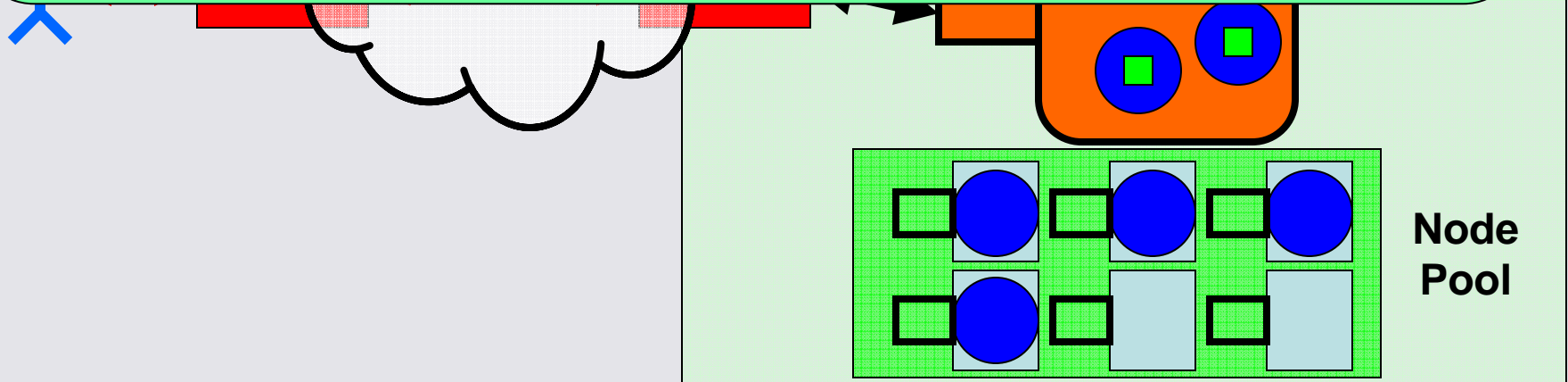
Public Access Control Channel as a Multicast Console Device

A "Super" Serial Line: A simple device for Raw Access

Node Control Channel Device



Fairness through Raw Hardware Access

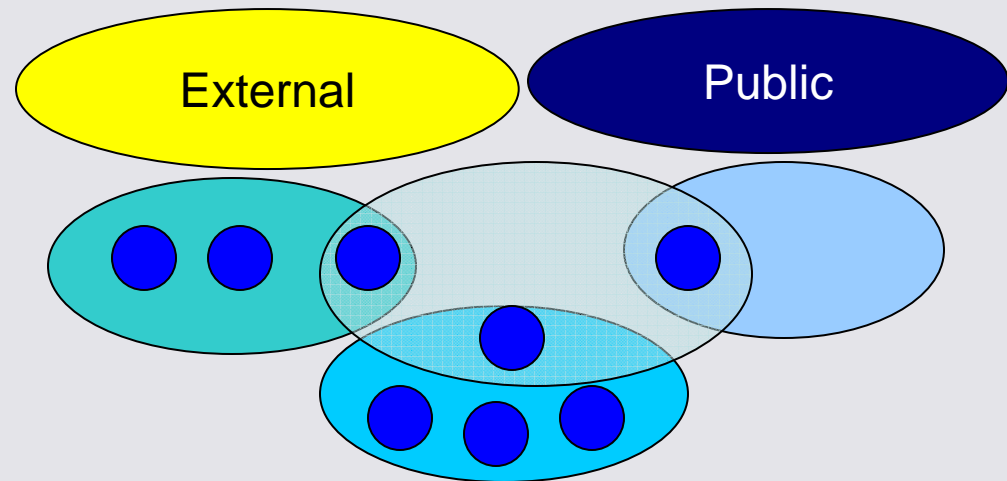


Communication Domains as a Core System Primitive

Domains: Hardware-enforced groups of HW endpoints that can communicate.

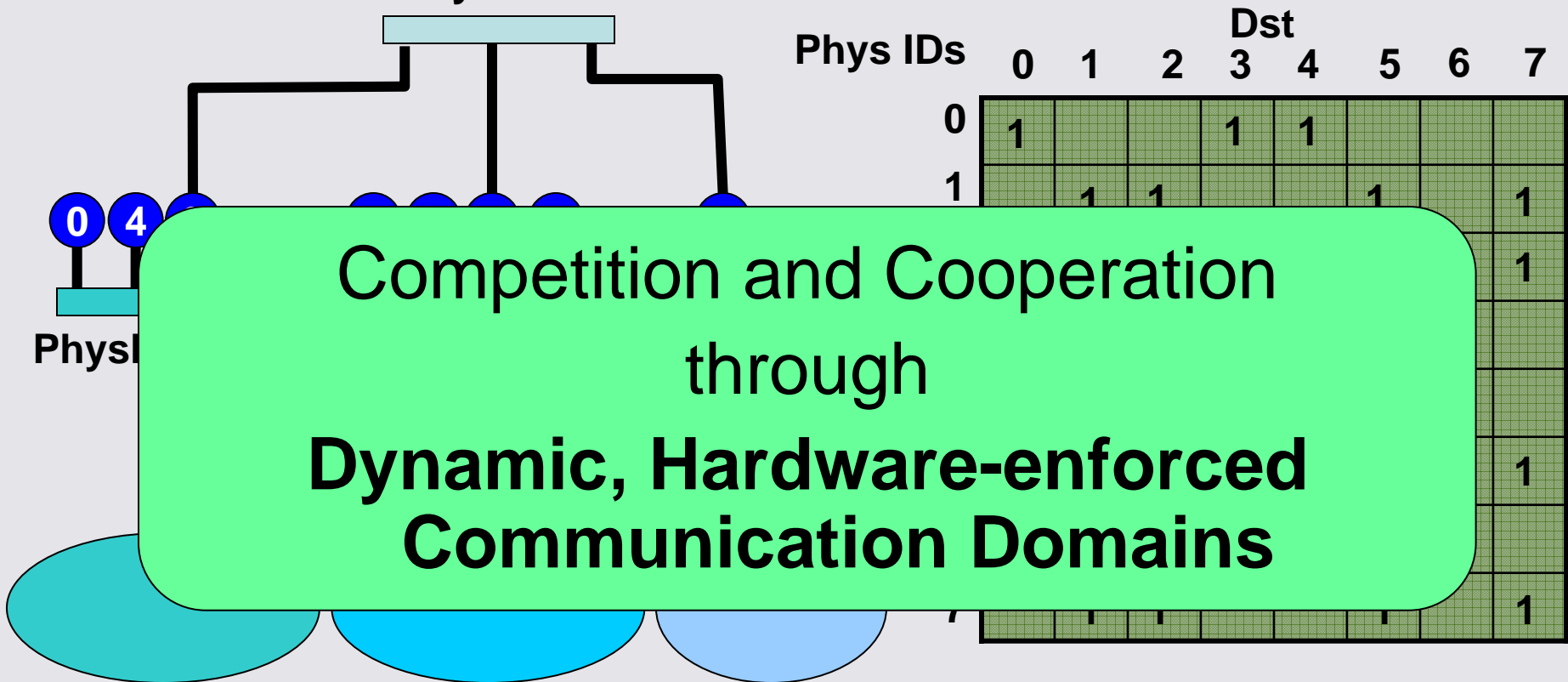
1. Power in realizing socio-economic relationships in topologies.
2. Individuals and organizations have freedom to isolate and protect resources and thus their interests and property.
3. Individuals and organizations can choose to cooperate by “simply” instantiating a resource in the public communication domain or a shared domain.

Domains + HW Endpoints



Hardware enforces communication constraints;
software controls everything else.
Software defines how and what is
communicated: raw bus use, cache lines,
pages, ethernet frames, etc.

Communication Domains as Permission

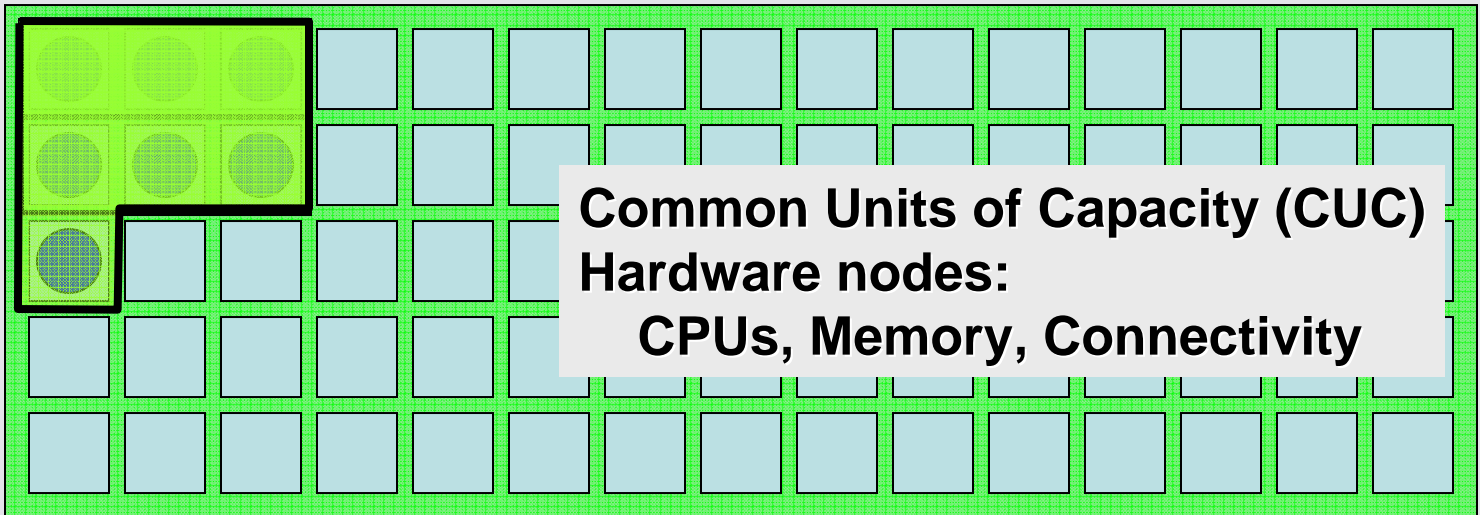
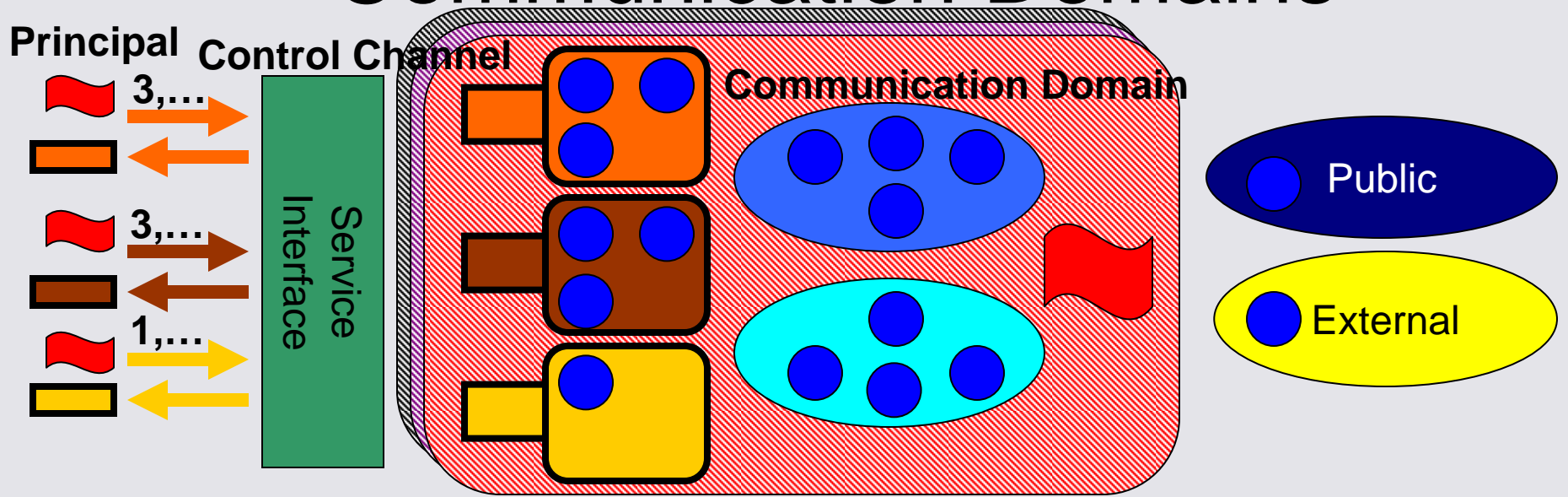


Phys IDs	0	1	2	3	4	5	6	7
0	1			1	1			
1		1	1			1		1
								1
								1
								1
								1
								1

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7

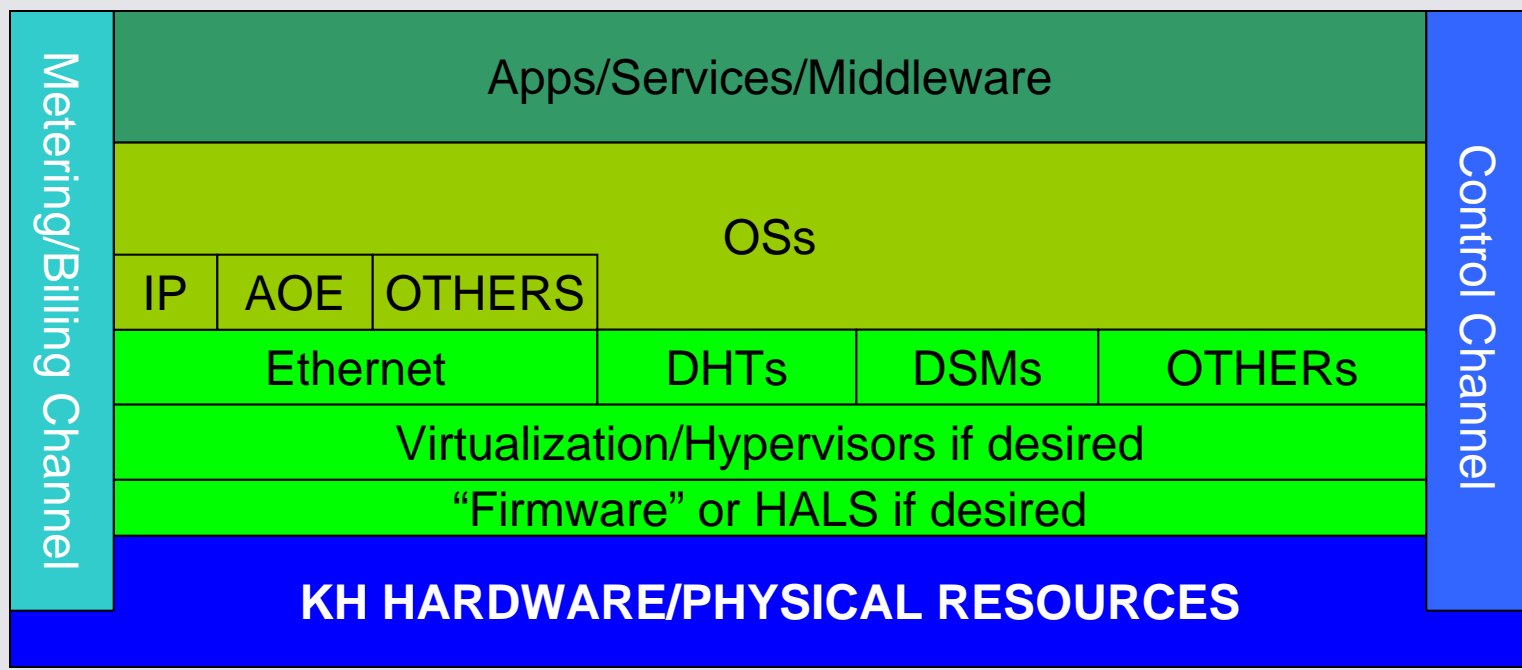
isAllowed(X,Y)?
 The Physical Owner wants this relationship enforced independent of software.

CUC, Principals, Control Channels, Communication Domains



An Online Stack of Choice and Composition (SW as Refinement)

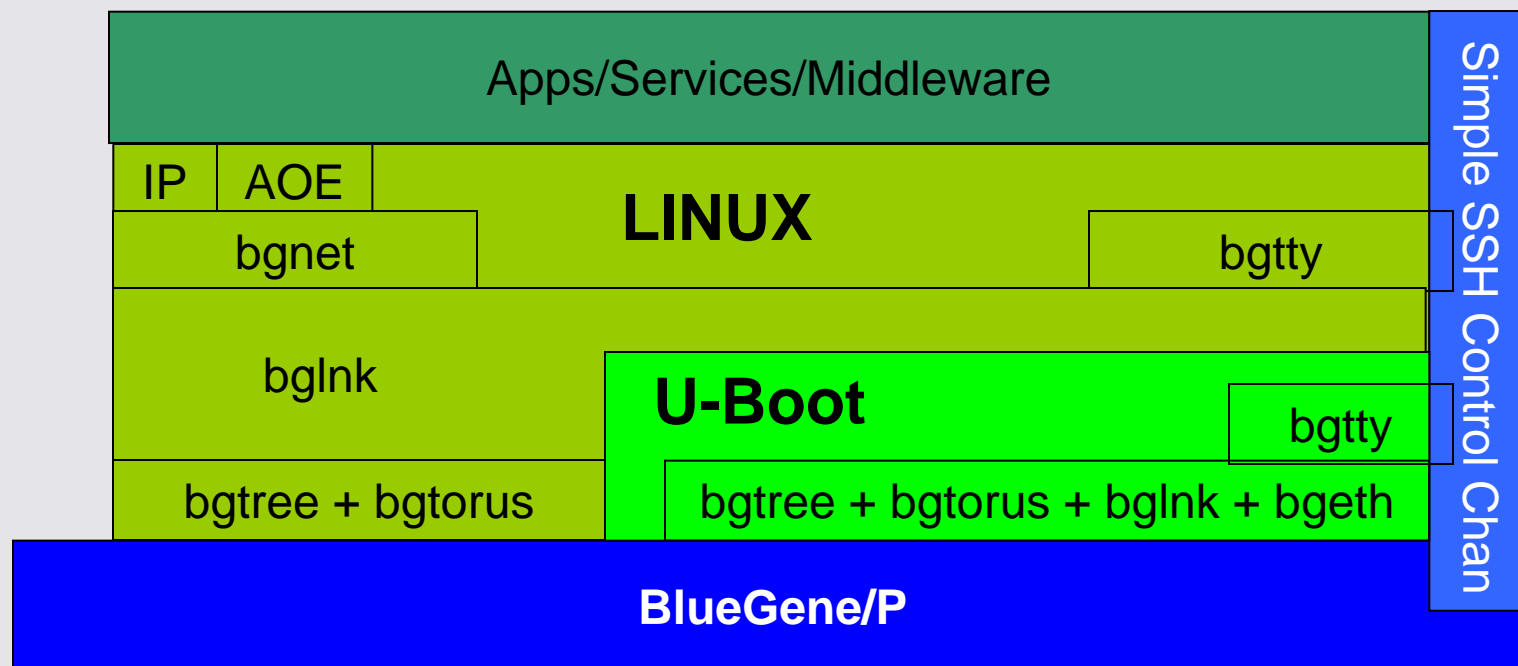
Anyone can revolutionize usage with new software at any level.
 Anyone can create new environments.
 Anyone can address niche needs via niche proprietary solutions.





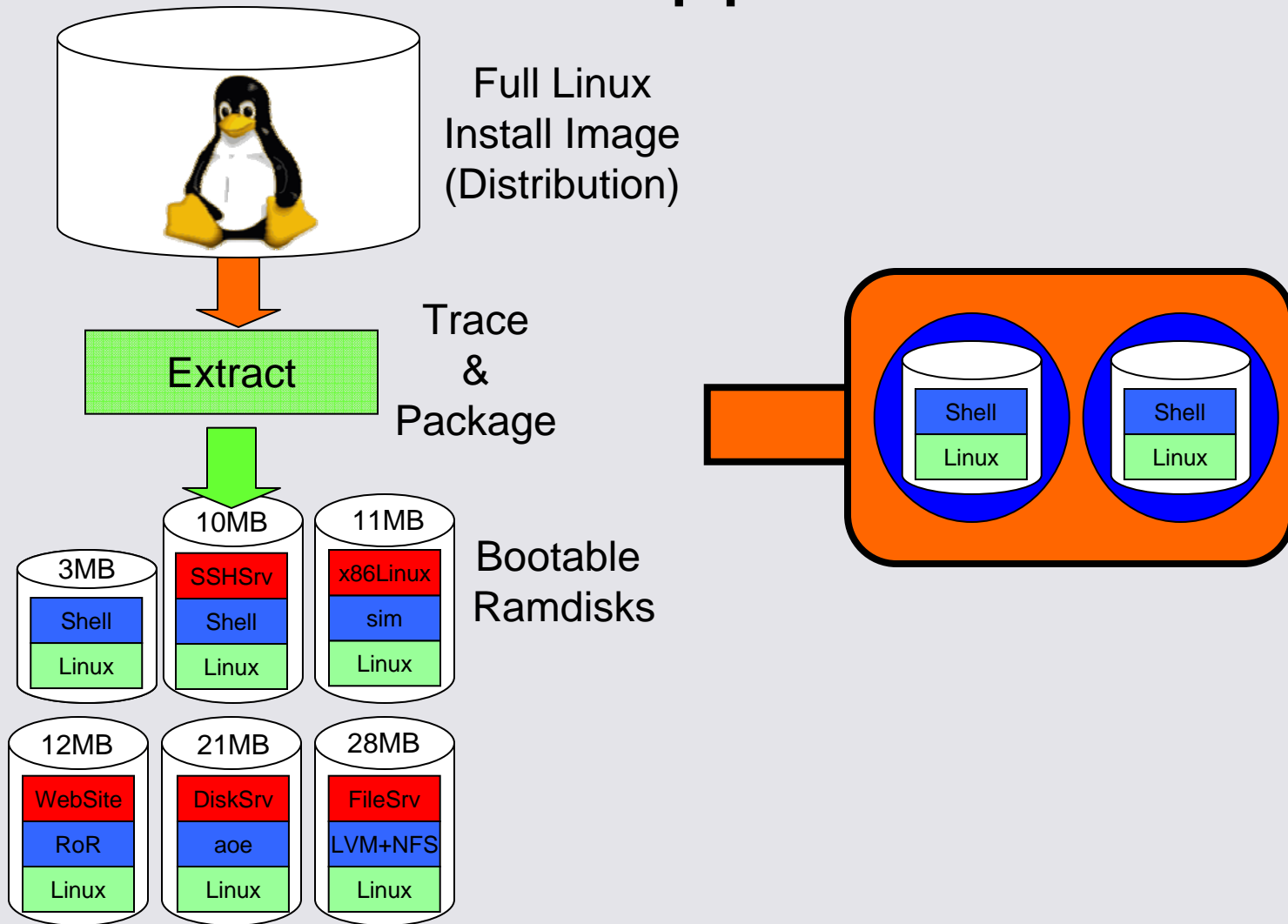
Exploring Kittyhawk via a Prototype on BlueGene/P

Current exemplar prototype built on BG/P hardware and predominately open-source software.



Trying out L4 (<http://www.l4ka.org/>) for both hardware prototyping and example virtualization.

Bootstrapping: Open-source Software Appliances





Demos



Building a Web Server Farm

<http://www.research.ibm.com/kittyhawk/movies/appdemo.mov>

00:05

The screenshot shows a video player window titled "appdemo.mov". At the top, there is a file system view with a grid of green circular icons representing files, numbered from 100 to 160. Below this, a terminal window is displayed with the following text:

```
FREE # 254
```

```
Lets begin by bring using a linux appliance we have created which contains apache  
allocate, load kernel, load appliance in the form of a ramdisk and boot (start window  
to view console)  
here we go ... Press <enter> ...  
khget -x apache 1
```

Below the terminal, there is a small window showing a network diagram and a terminal prompt:

```
{03,0; br0; port 1(eth0) entering forwarding state
```

```
simple apache fsdemo f8demo multi win khmon  
jappavoo@linux1:~$  
jappavoo@linux1:~$
```

The video player interface includes a progress bar at the bottom showing a time of 00:00:05 and standard playback controls (play, stop, previous, next, full screen).

00:21

The screenshot shows a video player window titled "appdemo.mov" with a menu bar (File, Edit, View, Window, Help). The main content is a terminal window. At the top, there is a progress bar with 160 green circles, each labeled with a number from 100 to 160. Below this, the terminal displays the following text:

```
FREE # 253
```

```
Network Id: 4 # 1
```

```
apache2: Could not reliably determine the server's fully qualified domain name, using 172.16.133.58 for ServerName
bash: no job control in this shell
bash-3.1#
```

```
33983+1 records out
8156136 bytes (8.2 MB) copied, 5.17139 seconds, 1.6 MB/s
booting 1 nodes...
apache worker @ 172.16.133.58
lets confirm the node is up and that apache is available using wget ... Press <enter> .
```

```
{93,0: Kittyhawk;announce;R00-M1-N08-J12 2 826 eth0 lo 172.16.133.58 127.0.0.1
```

```
/tmp/jappavoo/khmonstate/2 apache
```

```
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$
jappavoo@klinux1:~$
```

At the bottom right of the terminal, there are two buttons labeled "xemacs" and "term". The video player's progress bar shows a time of 00:00:21 and includes standard playback controls (play, stop, previous, next, volume).

00:58

The screenshot shows a video player window titled "appdemo.mov" with a menu bar (File, Edit, View, Window, Help). The main content is a terminal window. At the top, there is a progress bar with green circles and a list of numbers from 100 to 163. Below this, a status bar shows "FREE # 243". A file browser window is open, showing a directory with files named "10" through "12" and "2" through "5". Below the file browser, another status bar shows "Network Id: 4 # 11". The terminal window contains the following text:

```
bash-3.1# apache2: Could not reliably determine the server's fully qualified domain name, using 172.16.133.51 for ServerName
bash: no job control in this shell
bash-3.1#
```

Below the terminal window, there is a window titled "apache worker @ 172.16.133.51" with the following text:

```
apache worker @ 172.16.133.51
apache worker @ 172.16.133.59
apache worker @ 172.16.133.76
apache worker @ 172.16.133.204
or #1 ... no serialization imposed in the hardware ... rely on broadcast mediums ... Press <enter> ...
```

Below that, there is a window titled "{1313,0: Kittyhawk;announce;R00-M1-N12-J04 5 954 eth0 lo 172.16.133.186 127,0,0,1" with the following text:

```
{1313,0: Kittyhawk;announce;R00-M1-N12-J04 5 954 eth0 lo 172.16.133.186 127,0,0,1
```

At the bottom, there is a window titled "/tmp/jappavoo/khmonstate/9 apache" with the following text:

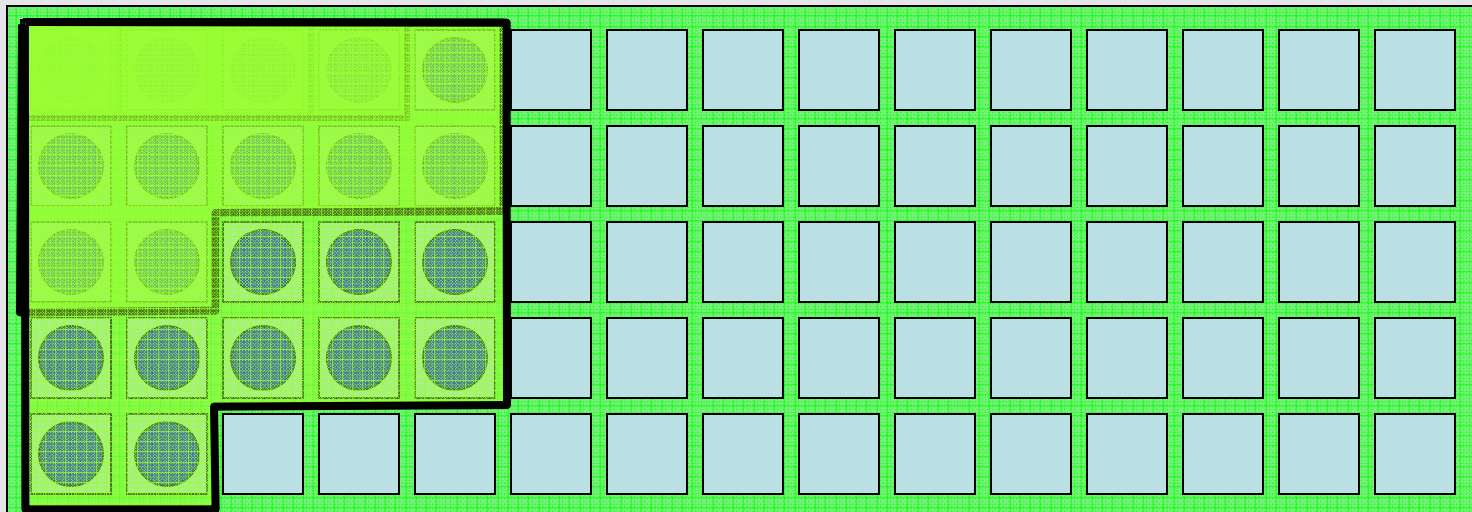
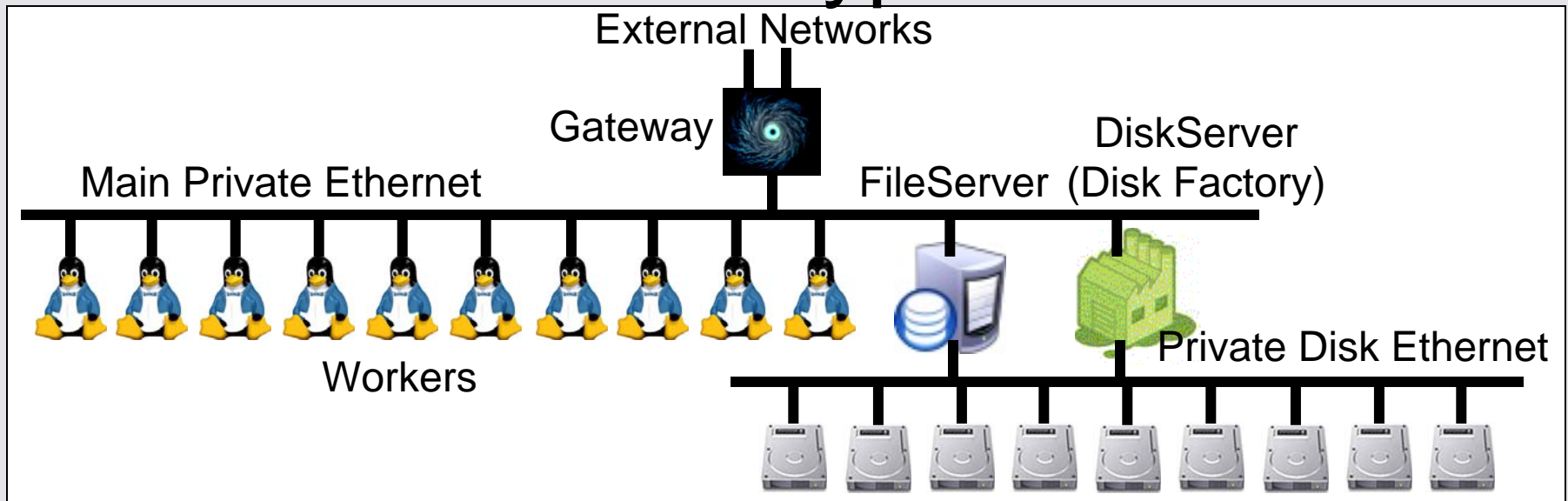
```
/tmp/jappavoo/khmonstate/9 apache
simple apache fsdemo f8demo multi win khmon
jappavoo@linux1:~$
jappavoo@linux1:~$
```

The video player has a progress bar at the bottom showing "00:00:58" and playback controls.

Building an Intranet

<http://www.research.ibm.com/kittyhawk/movies/fsdemo.mov>

A Simple Example on our Prototype



00:00

The screenshot shows a video player window titled "fsdemo.mov". The main content area displays a terminal window with a file system visualization at the top. The visualization consists of a grid of green circles representing files, with numbers 124 through 190 below them. Below the visualization, the text "FREE # 132" is visible. The terminal window shows a command prompt with the command `./run` being executed. Below the terminal window, there is a smaller terminal window showing the output of the command: `jappavoo@klinux1:~$ echo $(cat demoorder)`, `simple apache fsdemo f8demo multi win khmon`, and `jappavoo@klinux1:~$`. The video player interface includes a progress bar at the bottom showing "00:00:00" and several navigation buttons (play, stop, previous, next, full screen).



00:05

fsdemo.mov

File Edit View Window Help

125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192

FREE # 131

124

Network Id: 17 # 1

```

FIXME: using broadcast just for one node :-(<br>
GATEWAY INFO: con40@172.16.128.24<br>
protocol: uboot<br>
network External : eth0<br>
network Internal : eth1<br>
network 17 : eth2<br>
nodeinfo: 124 1008 172.16.133.240<br>
nodes: 124<br>
loading ...worker.cpio.gz.uimg

```

1

```

/tmp/jappavoo/khmonstate/119 linux
jappavoo@klinux1:~$ echo $(cat demoorder )
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$

```

00:00:05

xemacs term

00:19

The screenshot shows a video player window titled "fsdemo.mov" with a menu bar (File, Edit, View, Window, Help). The main content is a terminal window with the following text:

```
network 17 : eth2
nodeinfo: 124 1008 172.16.133.240
nodes: 124
loading ...worker.cpio.gz.uimg
26865+1 records in
26865+1 records out
6447708 bytes (6.4 MB) copied, 5.27425 seconds, 1.2 MB/s
gateway booted
Uploading fsdemo appliances to gateway (172.16.133.240)
40 10.1.13.24 192.168.0.126 127.0.0.1
/tmp/jappavoo/khmonstate/124 gateway
jappavoo@linux1:~$ echo $(cat demoorder )
simple apache fsdemo f8demo multi win khmon
jappavoo@linux1:~$
```

At the top of the terminal window, there is a status bar with "FREE # 131" and a grid of 124 green circles, each with a number from 125 to 198. A small window titled "Network Id: 17 # 1" is open, showing the number "124". The video player interface includes a progress bar at the bottom showing "00:00:19" and standard playback controls.

00:27

fsdemo.mov

File Edit View Window Help

125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192

FREE # 131

124

Network Id: 17 # 1

starting console: con40@172.16.128.24 connection
con40: Console

GATEWAY~fsdemo:~#

```
26865+1 records in
26865+1 records out
6447708 bytes (6.4 MB) copied, 5.27425 seconds, 1.2 MB/s
gateway booted
Uploading fsdemo appliances to gateway (172.16.133.240)
worker.cpio.gz.uimg          100% 6297KB   6.2MB/s   00:00
diskserver.cpio.gz.uimg     100%  20MB   9.8MB/s   00:02
fs.cpio.gz.uimg             100%  27MB   9.0MB/s   00:03
FIXME: using broadcast just for one node :-(-
```

40 10.1.13.24 192.168.0.126 127.0.0.1

```
jappavoo@klinux1:~$ echo $(cat demoorder )
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$
```

00:00:27

01:12

fsdemo.mov

File Edit View Window Help

126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159

160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193

FREE # 130

124 125

125

Network Id: 17 # 2

Network Id: 18 # 1

starting console: con40@172.16.128.24 connection
con40: Console

GATEWAY-fsdemo:~#

starting console: con41@172.16.128.24 connection
con41: Console

FSSSERVER-fsdemo:~#

```
7392+1 records in
7392+1 records out
1774155 bytes (1,8 MB) copied, 1,41429 seconds, 1,3 MB/s
loading ...fs.cpio.gz.uimg
118058+1 records in
118058+1 records out
28333997 bytes (28 MB) copied, 23,9513 seconds, 1,2 MB/s
fserver booted
FIXME: using broadcast just for one node :-(\
```

```
{1703,0; Kittyhawk;announce;R00-M1-N13-J11 125 995 eth0 lo 192,168,0,127 127,0,0,1
```

```
jappavoo@klinux1:~$ echo $(cat demoorder )
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$
```

00:01:12

01:42

fsdemo.mov

File Edit View Window Help

127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194

FREE # 129

Network Id: 17 # 3

Network Id: 18 # 2

```
starting console: con40@172.16.128.24 connection
con40: Console
[]

starting console: con41@172.16.128.24 connection
con41: Console
[]

network Internal : eth0
network 17 : eth1
network 18 : eth2
nodeinfo: 126 1003 10.1.13.19
nodes: 126
86008+1 records in
86008+1 records out
20642036 bytes (21 MB) copied, 17.4369 seconds, 1.2 MB/s
diskserver booted

8,0,128 127,0,0,1

/tmp/jappavoo/khmonstate/126 diskserver
jappavoo@klinux1:~$ echo $(cat demoorder )
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$
```

00:01:42



02:01

fsdemo.mov

File Edit View Window Help

127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194

FREE # 129

Network Id: 17 # 3 Network Id: 18 # 2

```
starting console: con40@172.16.128.24 connection
con40: Console
GATEWAY-fsdemo:~#

starting console: con41@172.16.128.24 connection
con41: Console
FSSSERVER-fsdemo:~#

starting console: con42@172.16.128.24 connection
con42: Console
DISKSERVER-fsdemo:~# khgetdisk 06:00:38:05:03:03 1
```

```
86008+1 records in
86008+1 records out
20642036 bytes (21 MB) copied, 17.4369 seconds, 1.2 MB/s
diskserver booted
DISKSERVER:
khgetdisk 06:00:38:05:03:03 1
FSSSERVER:
/root/scripts/vgaddaoc vg0
free=$(vgs --noheadings -o vg_free_count vg0); free=$(echo $free)
lvcreate -l $free vg0 -n lv0
```

```
8,0,128 127,0,0,1
```

```
/tmp/jappavoo/khmonstate/126 diskserver
jappavoo@klinux1:~$ echo $(cat demoorder )
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$
```

00:02:01

xemacs term



02:25

fsdemo.mov

File Edit View Window Help

128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195

FREE # 128

124 125 126 125 126 127

Network Id: 17 # 3 Network Id: 18 # 3

```
starting console: con40@172.16.128.24 connection
con40: Console
[]

starting console: con41@172.16.128.24 connection
con41: Console
[]

starting console: con42@172.16.128.24 connection
con42: Console
[]

60396+1 records in
60396+1 records out
14495235 bytes (14 MB) copied, 12.1971 seconds, 1.2 MB/s
DISKSERVER-fsdemo:"# []

86008+1 records in
86008+1 records out
20642036 bytes (21 MB) copied, 17.4369 seconds, 1.2 MB/s
diskserver booted
DISKSERVER:
#watchdisk 06:00:38:05:03:03 1
FSSERVER:
/root/scripts/vgaddaoc vg0
free=$(vgs --noheadings -o vg_free_count vg0); free=$(echo $free)
lvcreate -l $free vg0 -n lv0

{1913},0; Kittyhawk:announce;R00-M1-N13-J32 127 1016

/tmp/jappavoo/khmonstate/127 disk
jappavoo@klinux1:~$ echo $(cat demoorder )
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$ []
```

00:02:25

xemacs term

03:25

The screenshot shows a video player window titled "fsdemo.mov" with a standard menu bar (File, Edit, View, Window, Help). The main content area displays a terminal window with a multi-window desktop environment. At the top of the terminal is a grid of 128 green circles, numbered 128 to 195. Below this are two windows titled "Network Id: 17 # 3" and "Network Id: 18 # 3". The terminal also shows several command-line windows with the following text:

```
starting console: con40@172.16.128.24 connection
con40: Console

kjournald starting. Commit interval 5 seconds
EXT3 FS on dm-0, internal journal
EXT3-fs: mounted filesystem with ordered data mode.

starting console: con42@172.16.128.24 connection
con42: Console

60396+1 records in
60396+1 records out
14495235 bytes (14 MB) copied, 12.1971 seconds, 1.2 MB/s
DISKSERVER-fsdemo:~# khgetdisk 06:00:38:05:03:03 8

khgetdisk 06:00:38:05:03:03 1
FSSERVER:
/root/scripts/vgaddaoc vg0
free=$(vgs --noheadings -o vg_free_count vg0); free=$(echo $free)
lvcreate -l $free vg0 -n lv0
mkfs -t ext3 /dev/vg0/lv0; mount /dev/vg0/lv0 /mnt; df -h /mnt;
DISKSERVER:
khgetdisk 06:00:38:05:03:03 8
FSSERVER:
/root/scripts/vgaddaoc vg0; vgdisplay vg0

{1913},0; Kittyhawk:announce:R00-M1-N13-J32 127 1016

/tmp/jappavoo/khmonstate/127 disk
jappavoo@klinux1:~$ echo $(cat demoorder )
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$
```




03:29

fsdemo.mov

File Edit View Window Help

136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169

170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203

FREE # 120

124 125 126

125 126 127 128 129 130 131

132 133 134 135

Network Id: 17 # 3

Network Id: 18 # 11

```
starting console: con40@172.16.128.24 connection
con40: Console
[]

kjournald starting. Commit interval 5 seconds
EXT3 FS on dm-0, internal journal
EXT3-fs: mounted filesystem with ordered data mode.
[]

starting console: con42@172.16.128.24 connection
con42: Console
[]

GATEWAY-fsdemo:~# []

180 days, whichever comes first. Use tune2fs -c or -i to override.
Filesystem      Size Used Avail Use% Mounted on
/dev/mapper/vg0-lv0  1.5G  35M  1.4G   3% /mnt
FSSERVER-fsdemo:~# []

60396+1 records out
14495235 bytes (14 MB) copied, 12.1971 seconds, 1.2 MB/s
DISKSERVER-fsdemo:~# khgetdisk 06:00:38:05:03:03 8

khgetdisk 06:00:38:05:03:03 1
FSSERVER:
/root/scripts/vgaddaoc vg0
free=$(vgs --noheadings -o vg_free_count vg0); free=$(echo $free)
lvcreate -l $free vg0 -n lv0
mkfs -t ext3 /dev/vg0/lv0; mount /dev/vg0/lv0 /mnt; df -h /mnt;
DISKSERVER:
khgetdisk 06:00:38:05:03:03 8
FSSERVER:
/root/scripts/vgaddaoc vg0; vgdisplay vg0

{1913},0; Kittyhawk:announce:R00-M1-N13-J32 127 1016

/tmp/jappavoo/khmonstate/127 disk
jappavoo@klinux1:~$ echo $(cat demoorder )
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$ []
```

00:03:29

xemacs

term

03:50

fsdemo.mov

File Edit View Window Help

136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169

170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203

FREE # 120

124 125 126

125 126 127 128 129 130 131

132 133 134 135

Network Id: 17 # 3

Network Id: 18 # 11

```
starting console: con40@172.16.128.24 connection
con40: Console
[]

kjournald starting. Commit interval 5 seconds
EXT3 FS on dm-0, internal journal
EXT3-fs: mounted filesystem with ordered data mode.
[]

starting console: con42@172.16.128.24 connection
con42: Console
[]

180 days, whichever comes first. Use tune2fs -c or -i to override.
Filesystem      Size Used Avail Use% Mounted on
/dev/mapper/vg0-lv0  1.5G  35M  1.4G   3% /mnt
FSSSERVER-fsdemo:~# []

60396+1 records in
60396+1 records out
14495235 bytes (14 MB) copied, 12.1976 seconds, 1.2 MB/s
DISKSERVER-fsdemo:~# []

mkfs -t ext3 /dev/vg0/lv0; mount /dev/vg0/lv0 /mnt; df -h /mnt;
DISKSERVER:
khgetdisk 06:00:38:05:03:03 8
FSSSERVER:
/root/scripts/vgaddbase vg0: vgdisplay vg0
free=$(vgs --noheadings -o vg_free_count vg0); free=$(echo $free)
lvextend -l +$free /dev/vg0/lv0; umount /mnt; e2fsck -f /dev/vg0/lv0; resize2fs /dev/vg0/lv0; mount /dev/vg0/lv0 /mnt; df -h /mnt
echo '/mnt *(rw,fsid=12,no_root_squash)' >> /etc/exports
/etc/init.d/portmap start;
/etc/init.d/nfs-kernel-server start; /etc/init.d/nfs-common start;

{196},0; Kittyhawk:announce:R00-M1-N14-J04 133 1022
[]

/tmp/jappavoo/khmonstate/133 disk
jappavoo@klinux1:~$ echo $(cat demoorder )
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$ []
```

00:03:50

05:37

The screenshot shows a video player window titled "fsdemo.mov" with a menu bar (File, Edit, View, Window, Help). The main content area displays a terminal window with the following visible text:

```
136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169
170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203
FREE # 120
```

Below the grid are two network configuration windows:

```
Network Id: 17 # 3
Network Id: 18 # 11
```

Terminal sessions show:

```
starting console: con40@172.16.128.24 connection
con40: Console
GATEWAY-fsdemo:~# mount -o nolock,rw,proto=tcp 192.168.0.127:/mnt /mnt; ls /mnt
lost+found
GATEWAY-fsdemo:~# workers=$(khget -n 17 fsdemo 10)

NFSD: Using /var/lib/nfs/v4recovery as the NFSv4 state recovery
directory
NFSD: starting 90-second grace period
FSSSERVER-fsdemo:~# exportfs -v
/mnt <world>(rw,wdelay,no_root_squash,fsid=12,anonuid=65534,anongid=65534)
FSSSERVER-fsdemo:~#

starting console: con42@172.16.128.24 connection
con42: Console
60396+1 records in
60396+1 records out
14495235 bytes (14 MB) copied, 12.1976 seconds, 1.2 MB/s
DISKSERVER-fsdemo:~#
```

A large terminal window shows a shell script:

```
free=$(vgs --noheadings -o vg_free_count vg0); free=$(echo $free)
lvextend -l +$free /dev/vg0/lv0; umount /mnt; e2fsck -f /dev/vg0/lv0; resize2fs /dev/vg0/lv0; mount /dev/vg0/lv0 /mnt; df -h /mnt
echo '/mnt *(rw,fsid=12,no_root_squash)' >> /etc/exports
/etc/init.d/portmap start;
/etc/init.d/nfs-kernel-server start; /etc/init.d/nfs-common start;
exportfs -v
GATEWAY:
mount -o nolock,rw,proto=tcp 192.168.0.127:/mnt /mnt; ls /mnt
workers=$(khget -n 17 fsdemo 10)
echo "$workers" | khdo loadrawdisk /tmp/worker.cpio.gz.uing
```

At the bottom, a terminal window shows:

```
jappavoo@klinux1:~$ echo $(cat demoorder )
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$
```

The video player interface includes a progress bar at 00:05:37 and playback controls.

06:14

fsdemo.mov

File Edit View Window Help

146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213

FREE # 110

124 125 126 136 137 138 139
140 141 142 143 144 145

Network Id: 17 # 13

125 126 127 128 129 130 131
132 133 134 135

Network Id: 18 # 11

```
starting console: con40@172.16.128.24 connection
con40: Console
[]

NFSD: Using /var/lib/nfs/v4recovery as the NFSv4 state recovery
directory
NFSD: starting 90-second grace period
[]

starting console: con42@172.16.128.24 connection
con42: Console
[]

26865+1 records out
6447708 bytes (6.4 MB) copied, 5.39931 seconds, 1.2 MB/s
GATEWAY-fsdemo:~# echo "#workers" | khdo write "run kboot"
GATEWAY-fsdemo:~# []

FSSSERVER-fsdemo:~# exportfs -v
/mnt <world>(rw,wdelay,no_root_squash,fsid=12,anonuid=65534,anongid=65534)
FSSSERVER-fsdemo:~# []

60396+1 records in
60396+1 records out
14495235 bytes (14 MB) copied, 12.1976 seconds, 1.2 MB/s
DISKSERVER-fsdemo:~# []

exportfs -v
GATEWAY:
mount -o nolock,rw,proto=tcp 192.168.0.127:/mnt /mnt; ls /mnt
workers=$(khget -n 17 fsdemo 10)
echo "#workers" | khdo loadramdisk /tmp/worker.cpio.gz.uing
echo "#workers" | khdo write "run kboot"
echo "#workers" | khdo peripcmd "ping -c 1 %ip%"
echo "#workers" | khdo write "mount -o nolock,rw,proto=tcp 192.168.0.127:/mnt /mnt"
echo "#workers" | khdo peripcmd "ssh %ip% \"cat /proc/device-tree/u-boot-env/p0ip; echo -n ': ' ;df -h /mnt | grep mnt\""
echo "#workers" | khdo write "touch /mnt/\$(cat /proc/device-tree/u-boot-env/p0ip).hearlam"

jappavoo@klinux1:~$ echo $(cat demoorder )
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$ []
```

00:06:14

More than Appliances

<http://www.research.ibm.com/kittyhawk/movies/f8demo.mov>

00:00

The screenshot shows a video player window titled "f8demo.mov" with a menu bar (File, Edit, View, Window, Help). The main content is a terminal window with the following visible text:

```
100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133
134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167
FREE # 240
```

Below the grid are two network configuration windows:

```
Network Id: 3 # 3
Network Id: 4 # 14
```

Terminal output includes:

```
GATEWAY-f8corp:/F8root#
GATEWAY-f8corp:/F8root#
GATEWAY-f8corp:/F8root#
GATEWAY-f8corp:/F8root#
FSSSERVER-f8corp:~#
DISKSERVER-f8corp:~#
```

Terminal commands and output:

```
(~) ssh con@khctlserver
ssh xterm> ssh -f -i "/Work/khbg/HEAD/demo/users/khdemo/id_rsa root@172.16.133.196
RR"xterm -display #DISPLAY"
f8demo> vi fsdemo
f8demo>
{1573}.0: Kittyhawk;announce;R00-M1-N12-J30 15 980
/tmp/jappavoo/khmonstate/15 disk
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$
jappavoo@klinux1:~$
```

At the bottom, there is a playback control bar with a progress indicator at 00:00:00 and standard media controls (play, stop, previous, next, volume).

02:18

The screenshot displays a multi-window Linux desktop environment within a video player window titled "f8demo.mov". The desktop features a file manager window showing a directory structure, a terminal window with shell commands, and a network configuration window. The terminal window shows the following commands and output:

```
xeyes &
appletviewer http://www.schubart.net/rc &
emacs &
cd /tmp
wget http://www.kernel.org/pub/linux/kernel/v2.6/linux-2.6.23.tar.bz2
tar xjf linux-2.6.23.tar.bz2
cd linux-2.6.23
make maple_defconfig
make -j 8

export DISTCC_HOSTS="$(cat /etc/mkworkers.internal.hosts)"
xterm -bg white -fg red -exec "distccmon-text 1" &
make CROSS_COMPILE='distcc ' -j 32

[root@localhost ~]#
```

The terminal also shows the output of the following commands:

```
Starting xinetd: [ OK ]
Starting console mouse services: [ OK ]
Starting authd: [ OK ]
INIT: no more processes left in this runlevel
[]

ssh xterm> ssh -f -i ~/Work/khbg/HEAD/demo/users/khdemo/id_rsa root@172.16.133.67
"xterm -display #DISPLAY"
ssh xterm> []

f8demo> vi fsdemo
f8demo> []

0,18 127,0,0,1
[]

/tmp/jappavoo/khmonstate/16 fedora
[]
```

The video player interface includes a progress bar at the bottom showing a duration of 00:02:18 and standard playback controls.

03:28

The screenshot displays a Linux desktop environment with several windows:

- File Manager:** Shows a directory structure with icons for folders and files.
- System Monitor:** Displays network status for 'Network Id: 3 # 4' and 'Network Id: 4 # 14'.
- Terminal:** Contains the following text:

```
export DISTCC_HOSTS="$(cat /etc/mkworkers.internal.hosts)"
xterm -bg white -fg red -exec 'distccmon-text 1' &
make CROSS_COMPILE='distcc' -j 32

[root@localhost ~]# xclock &
[1] 2589
[root@localhost ~]# xeyes &
[2] 2590
[root@localhost ~]# appletviewer http://www.schubart.net/rc &
[3] 2591
[root@localhost ~]# Warning: Can't read AppletViewer properties file: /root/.hot
java/properties Using defaults.
Jul 23, 2008 8:13:28 AM java.util.prefs.FileSystemPreferences#2 run
INFO: Created user preferences directory.

Starting xinetd: [ OK ]
Starting console mouse services: [ OK ]
Starting authd: [ OK ]
INIT: no more processes left in this runlevel

ssh xterm> ssh -f -i "/Work/khbg/HEAD/demo/users/khdemo/id_rsa root@172.16.133.67"
"xterm -display #DISPLAY"
ssh xterm>

f8demo> vi fsdemo
f8demo>

0,18 127,0,0,1

/tmp/jappavoo/khmonstate/16 fedora
simple apache fsdemo f8demo multi win khmon
jappavoo@klinux1:~$
jappavoo@klinux1:~$
```
- Applet:** A Java applet window titled 'Applet' showing a 3D Rubik's cube. Below the cube are buttons for 'Scramble', 'Give Up', and 'Sound', along with a 'Moves: 0' counter and the text 'Scramble, scramble...'. The 'Scramble' button has a red mouse cursor over it.

f8demo.mov

File Edit View Window Help

100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133
134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167

FREE # 239

16 1 2 3

10 11 12 13 14 15 2
3 4 5 6 7 8 9

Network Id: 3 # 4

Network Id: 4 # 14

```
[root@localhost ~]# xlock &  
[1] 2589  
[root@localhost ~]# xeyes &  
[2] 2590  
[root@localhost ~]# appletviewer http://www.schubart.net/rc &  
[3] 2591  
[root@localhost ~]# Warning: Can't read AppletViewer properties file: /root/.hot  
java/properties Using defaults.  
Jul 23, 2008 8:13:28 AM java.util.prefs.FileSystemPreferences#2 run  
INFO: Created user preferences directory.  
[root@localhost ~]# emacs &  
[4] 2617  
[root@localhost ~]#
```

Starting xinetd: [OK]
Starting console mouse services: [OK]
Starting authd: [OK]

File Edit Options Buffers Tools Complete In/Out Signals Help

```
printf("Hello World!!!\n");  
return 0;  
}
```

-u:-- hello.c Bot L7 (C/l Abbrev)-----
[root@localhost ~]# ./hello
Hello World!!!
[root@localhost ~]#

-u:** *shell* All L3 (Shell:run)-----

Applet

Give Up Sound Moves: 0

mon xemacs term

00:06:07

06:07

10:00

File Edit View Window Help

100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133
134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167

FREE # 239

16 1 2 3

10 11 12 13 14 15 2
3 4 5 6 7 8 9

Network Id: 3 # 4

Network Id: 4 # 14

```
*  
* Hardware crypto devices  
*  
Hardware crypto devices (CRYPTO_HW) [N/y/?] n  
# configuration written to .config  
#  
[root@localhost linux-2.6.23]# make -j 4  
scripts/kconfig/conf -s arch/powerpc/Kconfig  
CHK include/linux/version.h  
UPD include/linux/version.h  
CHK include/linux/utsrelease.h  
UPD include/linux/utsrelease.h  
SYMLINK include/asm -> include/asm-powerpc
```

network 3 : eth1
nodeinfo: 16 835 172.16.133
nodes: 16
GATEWAY-f8corp:/F8root#

5534) <world>(rw,wdelay,no_root_squash,fsid=12,anonuid=65534,anongid=65534)
/mnt
FSSSERVER-f8corp:/export/F8#
DISKSERVER-f8corp:"#

Starting xinetd: [OK]
Starting console mouse services: [OK]
Starting authd: [OK]

File Edit Options Buffers Tools C Help

```
#include <stdio.h>  
  
int main(int argc, char **argv)  
{  
    printf("Hello World!!!\n");  
    return 0;  
}
```

-u:-- hello.c All L7 (C/1 Abbrev)-----
Beginning of buffer

00:10:00

11:00

File Edit View Window Help

100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133
134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167

FREE # 231

16 17 18 19 1 20 21
22 23 24 2 3

Network Id: 3 # 12

10 11 12 13 14 15 2
3 4 5 6 7 8 9

Network Id: 4 # 14

```
appletviewer http://www.schubart.net/rc &  
emacs &  
cd /tmp  
wget http://www.kernel.org/pub/linux/kernel/v2.6/linux-2.6.23.  
tar.bz2  
tar xjf linux-2.6.23.tar.bz2  
cd linux-2.6.23  
make maple_defconfig  
make -j 8  
  
export DISTCC_HOSTS="$(cat /etc/mkworkers.internal.hosts)"  
xterm -bg white -fg red -exec "distccmon-text 1" &  
make CROSS_COMPILE='distcc ' -j 32  
[root@localhost linux-2.6.23]#
```

```
nodeinfo: 24 834 192.168.0.26  
nodeinfo: 19 843 192.168.0.21  
nodes: 17 18 19 20 21 22 23 24  
GATEWAY-f8corp:/f8root#
```

```
5534)  
/mnt <world>(rw,wdelay,no_root_squash,fsid=12,anonuid=65534,anongid=6  
5534)  
FSSSERVER-f8corp:/export/f8#
```

```
DISKSERVER-f8corp:"#
```

```
Starting xinetd: [ OK ]  
Starting console mouse services: [ OK ]  
Starting authd: [ OK ]  
INIT: no more processes left in this runlevel  
[
```

```
#include <stdio.h>  
  
int main(int argc, char **argv)  
{  
    printf("Hello World!!!\n");  
    return 0;  
}
```

-u:-- hello.c All L7 (C/1 Abbrev)-----
Beginning of buffer

00:11:00



f8demo.mov

File Edit View Window Help

100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200

FREE # 231

16 17 18 19 1 20 21
22 23 24 2 3

10 11 12 13 14 15 2
3 4 5 6 7 8 9

Network Id: 3 # 12 Network Id: 4 # 14

```
CHK include/linux/compile.h
CC init/main.o
CC init/do_mounts_rd.o
CC init/noinitramfs.o
CC init/calibrate.o
LD usr/built-in.o
CC init/do_mounts.o
CC init/version.o
CC arch/powerpc/kernel/ptrace.o
```

```
3677 Compile asm-offsets.c 192.168.0.19[0]
3719 Preprocess main.c 192.168.0.19[0]
3737 Preprocess do_mounts_rd.c 192.168.0.20[0]
3745 Preprocess calibrate.c 192.168.0.21[0]
3741 Preprocess noinitramfs.c 192.168.0.22[0]
3742 Compile localhost[0]
3751 Compile localhost[1]

3719 Compile main.c 192.168.0.19[0]
3800 Preprocess cputable.c 192.168.0.19[1]
3737 Compile do_mounts_rd.c 192.168.0.20[0]
3806 Preprocess syscalls.c 192.168.0.20[1]
3745 Compile calibrate.c 192.168.0.21[0]
3811 Preprocess irq.c 192.168.0.21[1]
3741 Compile noinitramfs.c 192.168.0.22[0]
3759 Compile do_mounts.c 192.168.0.23[0]
3788 Compile version.c 192.168.0.24[0]
3795 Compile ptrace.c 192.168.0.25[0]
3798 Preprocess semaphore.c 192.168.0.26[0]
3786 Compile localhost[0]
```

```
Starting xinetd: [ OK ]
Starting console mouse services: [ OK ]
Starting authd: [ OK ]
INIT: no more processes left in this runlevel
```

```
#include <stdio.h>

int main(int argc, char **argv)
{
    printf("Hello World!!!\n");
    return 0;
}
```

-u:-- hello.c All L7 (C/1 Abbrev)-----
Beginning of buffer

00:12:24

Give Up Sound Moves: 0

mon xemacs term

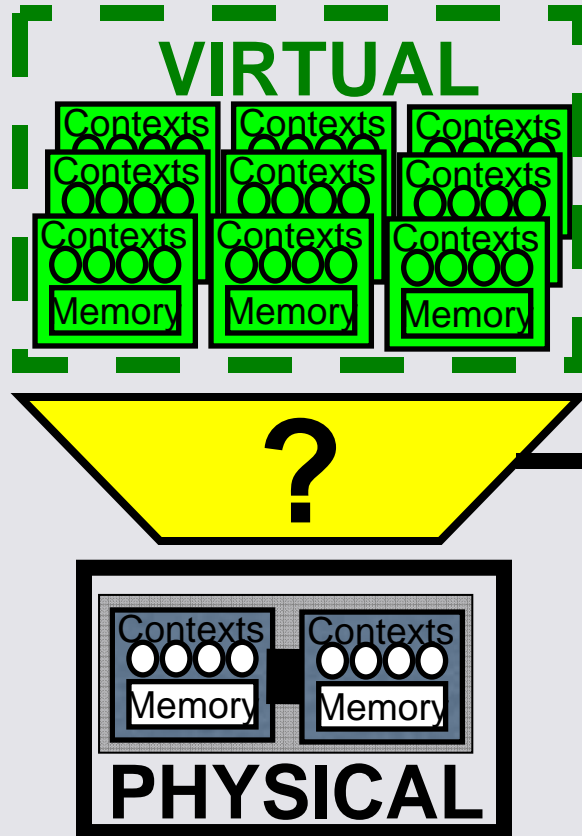
12:24

<http://www.research.ibm.com/kittyhawk/movies/primitives.mov>



Why Hardware-centric? When Virtual is Bad!

Heresy: Inherent virtualization runs counter to a fair system – A Virtual Computer is not a Computer!



Virtualization / time-sharing is real technology with complex mechanisms and policies.

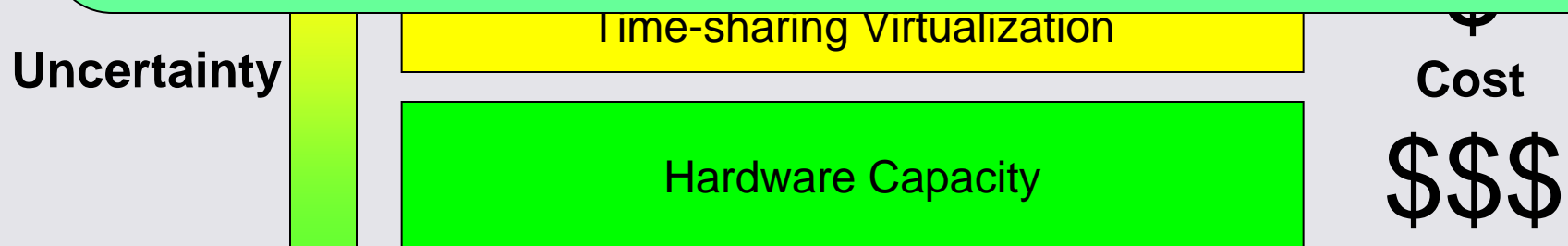
- Virtual Context → Physical Context?
- Virtual Memory → Physical Memory?
- Virtual I/O ↔ Real I/O?
- Security?
- Utilization?
- etc.

Virtualization is a service/derivative/future of the underlying resource. Its value should be isolated.

Uncertainty in Resources

“A user can either pick a level of service and pay whatever it takes to obtain it (giving him certainty with respect to service, uncertainty with respect to cost), or he can pick a price level and accept whatever level of service he can obtain

**Fairness
through
Raw Hardware Access**



Our Digital Future -- by Accident or Design?

*“The past is but the beginning of a beginning,
and all that is or has been is but the twilight
of the dawn”*

H.G. Wells, *“The Discovery of the
Future”*, 1901

Advice from the Pioneers



John
McCarthy



Joseph Carl
Robnett Licklider



Robert Fano



Fernando J.
Corbató

“The very power of advanced computer systems makes them a serious threat to

“...multiple access systems operated as a utility will substantially extend the exploitation of information processing systems to the point of having significant social consequences.”

F.J. Corbató and V.A. Vyssotsky, “Introduction and Overview of the Multics System”, Proceedings of the 1965 Fall Joint Computer Conference.

Accessible Computing , 1965 Fall Joint Computer Conference.

Not If, but How?

- “The development of such a system is probably inevitable (unless it is forbidden by law)...”
- “The main danger to be avoided is the creation of services of limited scope...”

- “Another problem is to avoid monopolies; the intrinsic nature of the system permits any person who can write computer programs to compete with large organizations in inventing and offering imaginative services, but one can worry that the system might develop commercially in some way that would prevent that. In general we should try to develop information services in such a way as will enhance the individuality of its users.”

- “... The major force that might tend to reduce competition is the exclusive possession of proprietary programs or files. Therefore it is desirable to separate the ownership of programs performing services from the ownership of the service bureaus themselves ...”

John McCarthy, “The Home Information Terminal”, Man and Computer. Proc. int. Conf., Bordeaux 1970, pp 48-57.

Summary

1. Large-scale communication-centric system through **Aggressive Integration**.
2. Fairness through **Raw Hardware**

Our Digital Future is at hand -- let's
have some FUN!
Open questions abound; we don't have
the answers; let's chat!

end, exploitation of computers for the benefit of society hinges upon two pivots: education, and responsible considered action by those of the technical community able to exert some influence."

E.E. David Jr and R.M. Fano, "Some Thoughts About the Social Implications of Accessible Computing", 1965 Fall Joint Computer Conference.