SAKURA Internet

www.sakura.ad.jp

次世代データセンターのための要素技術

2011年03月8日

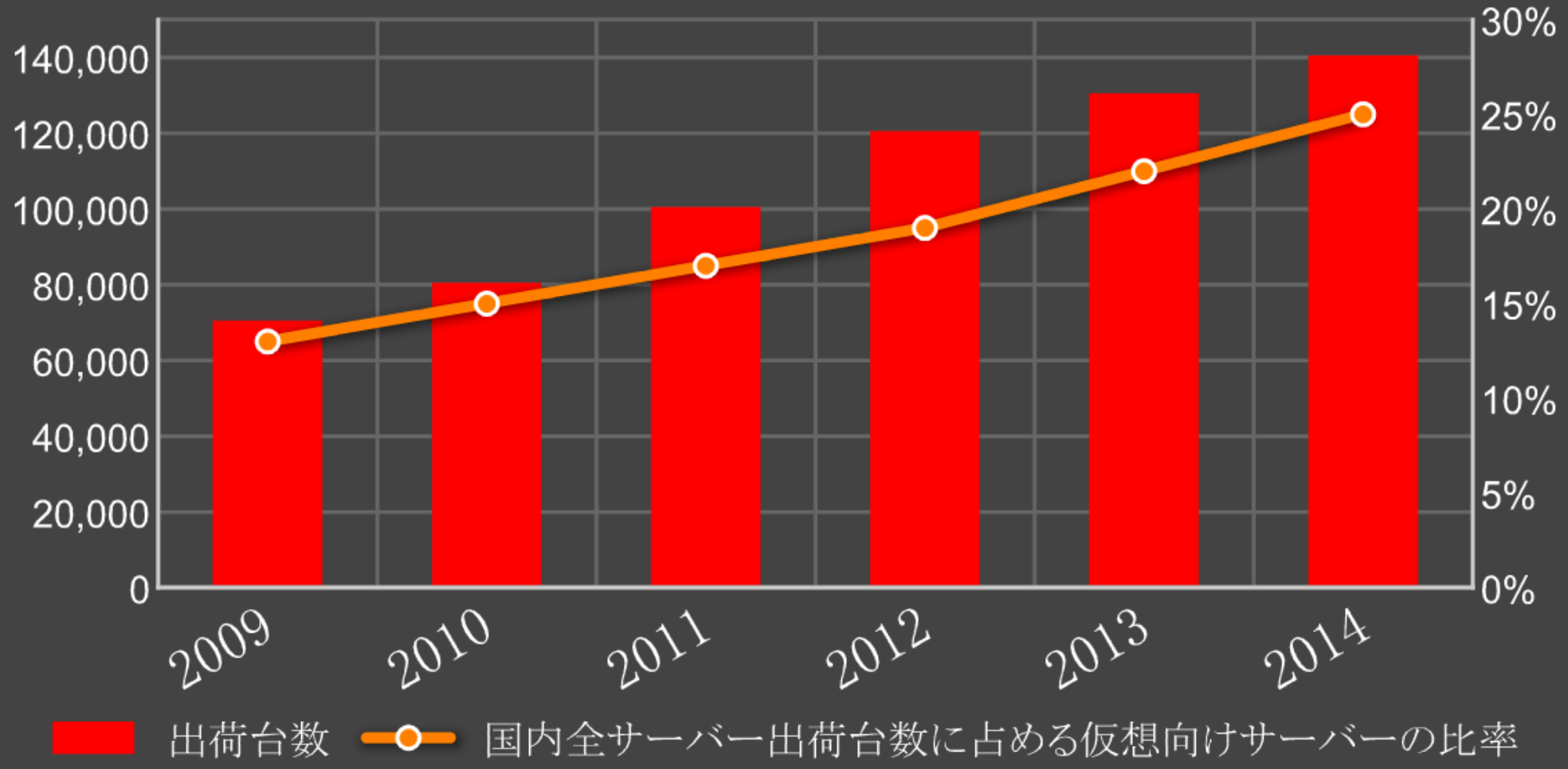
さくらインターネット研究所　上級研究員
松本直人

(C)Copyright 1996-2010 SAKURA Internet Inc.

現在の我々を取り巻く計算機環境

SAKURA Internet

出典: 国内仮想化サーバー市場 出荷台数予測、2009年～2014年 IDC Japan株式会社 (2010年10月28日作成)

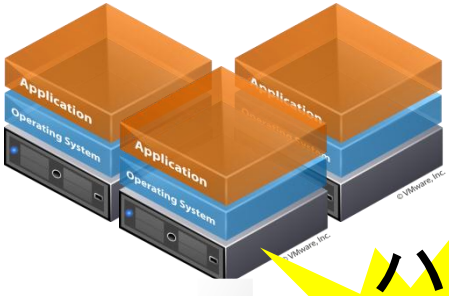


今後5年以内に国内サーバーの1/4が仮想環境となる

クラウド時代のコンピューティング環境

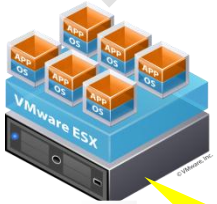
SAKURA Internet

従来のコンピューティング



ハード依存性
アプリ非効率

仮想化技術の導入

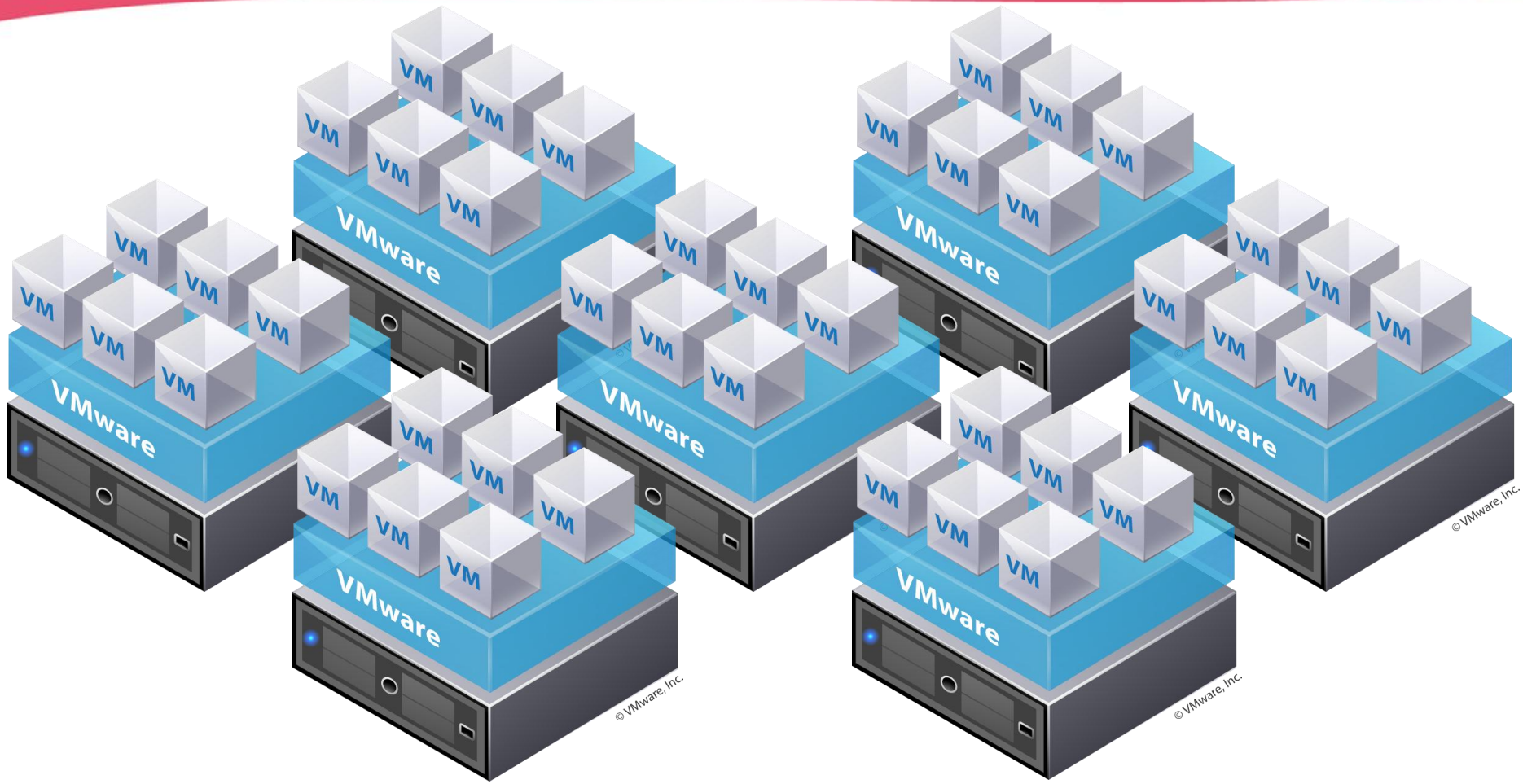


ネットワーク
サービス化

クラウド・コンピューティング

クラウド時代のコンピューティング環境

SAKURA Internet

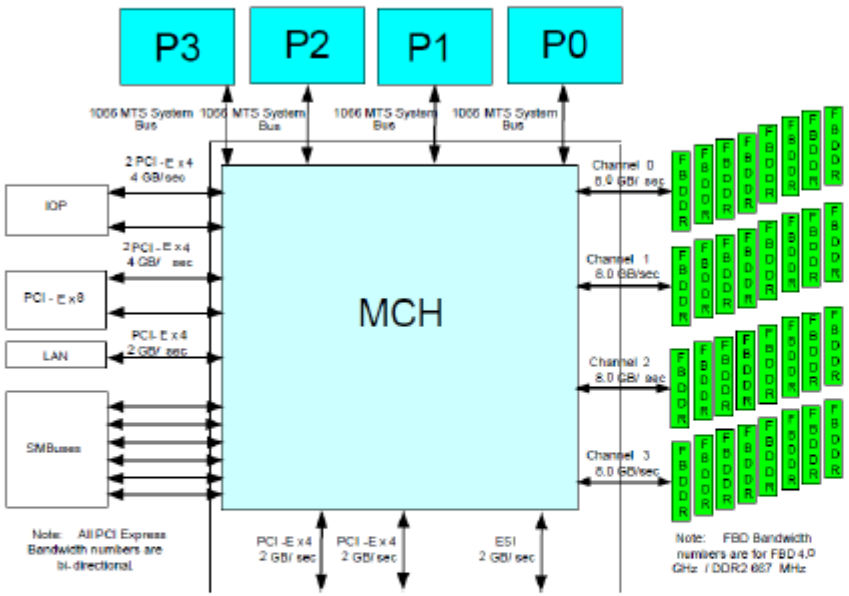


多くの仮想化されたシステムが連携して動作する世界

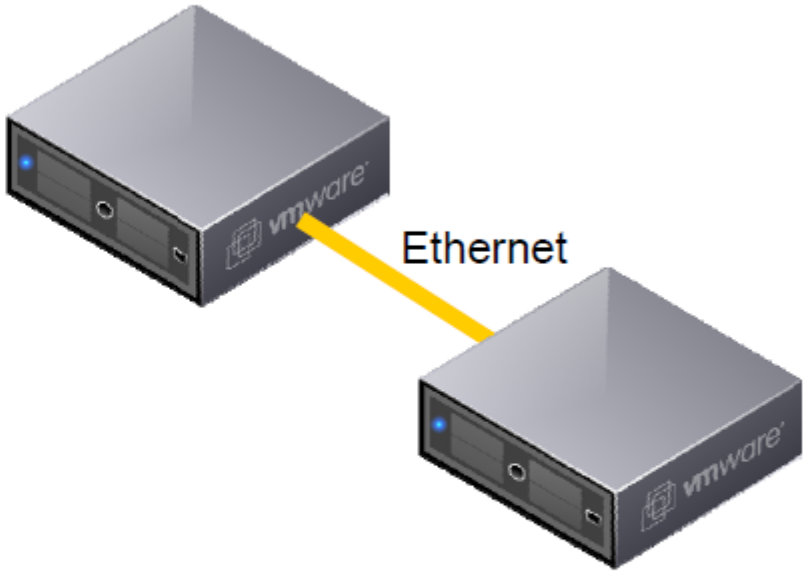
現行システムの性能限界を把握の重要性

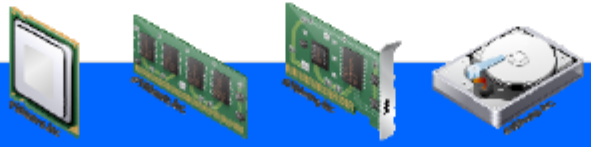
SAKURA Internet

バス・ボトルネック



インターフェイス・ボトルネック





仮想化技術であってもパフォーマンスは、コンポーネントの性能限界につよく依存している

仮想化との付き合い方

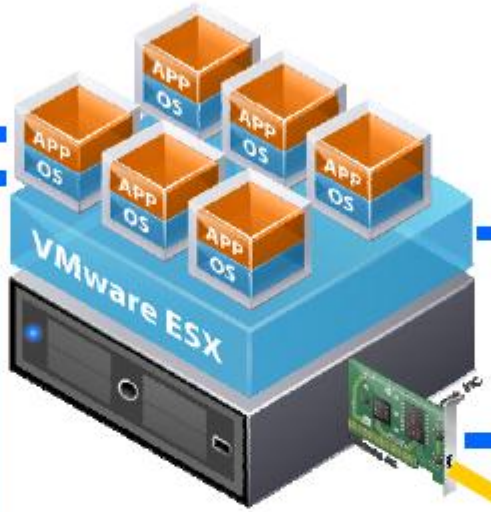
SAKURA Internet

カーネル性能(VM)

カーネル
パラメーター
設定

NIC性能(VM)

NICドライバ
設定



VMM性能(物理)

CPU
アフィニティ
設定

NICドライバ
キュー設定
(VMDq)

NIC性能(物理)

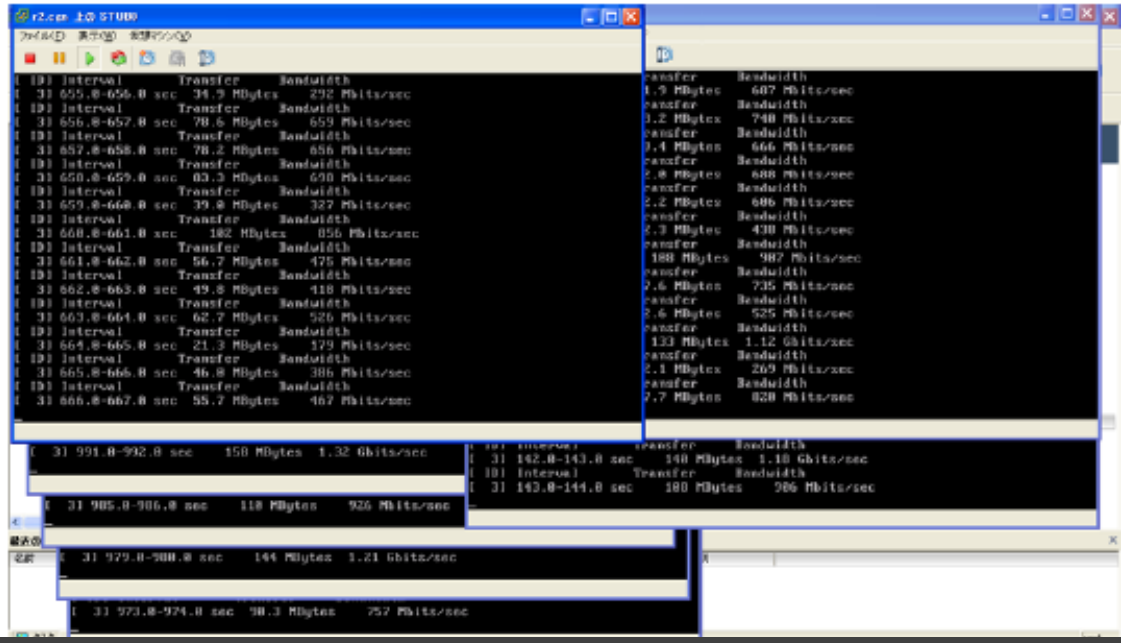
Intel Virtual Machine Device Queues (VMDq)

マルチコアが抱える限界

SAKURA Internet

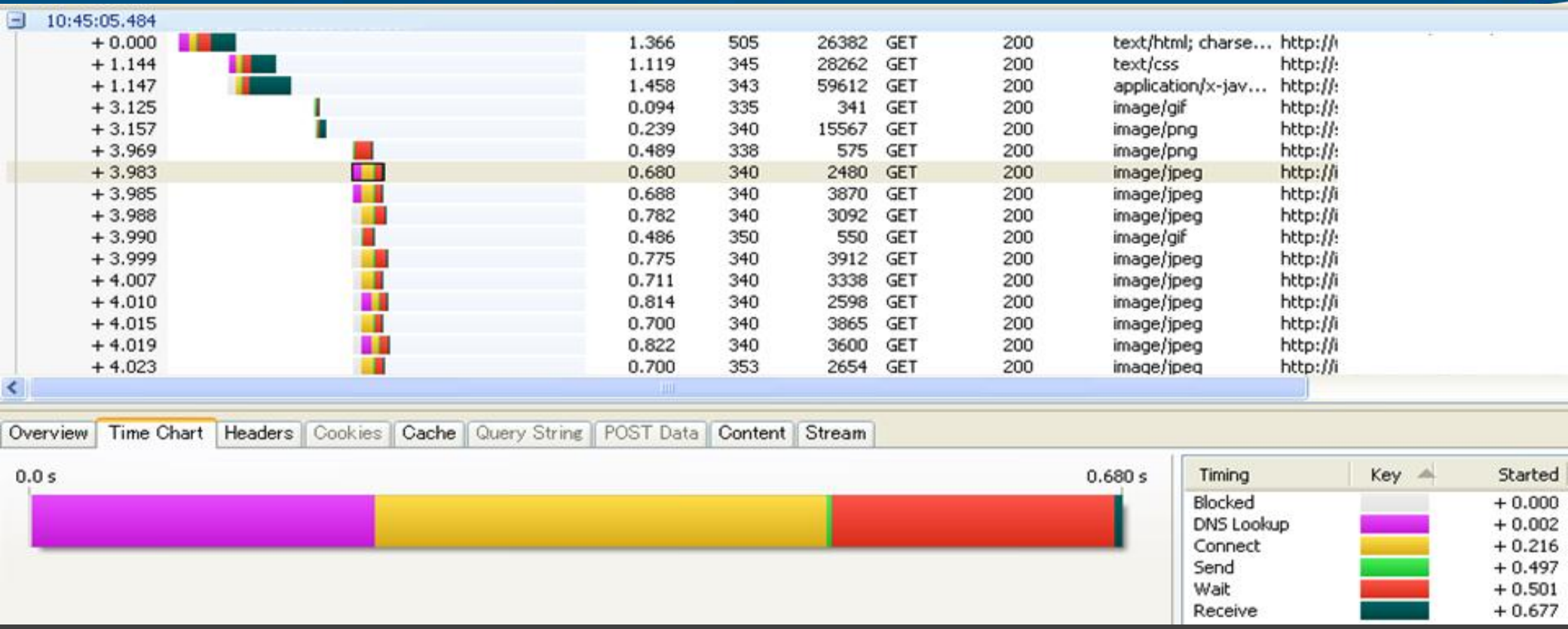
6:55:30pm up 2 days 9:42, 93 worlds; CPU load average: 0.58, 0.23, 0.09

PORT ID	UPLINK	USED BY	DTYP	DNAME	PKTTX/s	MbTX/s	PKTRX/s	MbRX/s	%DRPTX	%DRPRX
33554433	Y	vmnic4	H	vSwitch2	45899.41	23.53	634802.70	7329.40	0.00	0.00
33554445	N	1123:TEST0	H	vSwitch2	13399.41	6.80	92695.80	1073.55	0.00	0.01
33554440	N	1103:TEST4	H	vSwitch2	12933.10	6.66	94841.79	1098.40	0.00	0.29
33554438	N	1093:TEST2	H	vSwitch2	9470.70	4.90	94307.61	1092.21	0.00	0.21
33554437	N	1085:TEST1	H	vSwitch2	2965.33	1.50	90999.51	1053.90	0.00	0.00
33554439	N	1098:TEST3	H	vSwitch2	2647.95	1.36	96877.44	1121.98	0.00	0.01
33554444	N	1118:TEST6	H	vSwitch2	2580.08	1.35	103975.09	1204.18	0.00	0.08



アプリケーション特性とメニーコアの可能性

SAKURA Internet



大多数の並列処理が求められるケース

メニーコア・システムの出現

SAKURA Internet

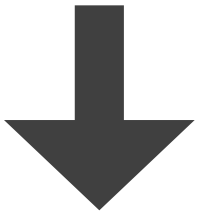
```
top - 14:16:47 up 20:30,  4 users,  load average: 3.94, 1.24, 0.72
Tasks: 643 total,   1 running, 641 sleeping,   1 stopped,   0 zombie
Cpu0  : 13.2%us, 67.9%sy,  0.0%ni, 18.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu1  :  6.7%us, 78.1%sy,  0.0%ni, 15.2%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu2  : 10.5%us, 82.9%sy,  0.0%ni,  6.7%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu3  :  9.4%us, 71.7%sy,  0.0%ni, 18.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu4  : 10.5%us, 73.3%sy,  0.0%ni, 16.2%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu5  :  9.3%us, 76.6%sy,  0.0%ni, 14.0%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu6  :  9.4%us, 72.6%sy,  0.0%ni, 17.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu7  :  9.9%us, 72.1%sy,  0.0%ni, 18.0%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu8  :  8.5%us, 84.9%sy,  0.0%ni,  6.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu9  :  9.6%us, 76.0%sy,  0.0%ni, 14.4%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu10 : 10.0%us, 71.8%sy,  0.0%ni, 18.2%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu11 : 19.3%us, 61.5%sy,  0.0%ni, 19.3%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu12 : 15.0%us, 66.4%sy,  0.0%ni, 18.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu13 :  8.3%us, 74.1%sy,  0.0%ni, 17.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu14 :  8.7%us, 81.7%sy,  0.0%ni,  9.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu15 : 10.6%us, 79.8%sy,  0.0%ni,  9.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu16 :  7.5%us, 83.0%sy,  0.0%ni,  9.4%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu17 : 11.6%us, 83.2%sy,  0.0%ni,  5.3%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu18 :  8.9%us, 82.2%sy,  0.0%ni,  8.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu19 :  8.7%us, 76.0%sy,  0.0%ni, 15.4%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu20 : 11.1%us, 82.4%sy,  0.0%ni,  6.5%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu21 :  8.1%us, 78.4%sy,  0.0%ni, 13.5%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu22 : 13.5%us, 74.0%sy,  0.0%ni, 12.5%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu23 : 20.2%us, 66.7%sy,  0.0%ni, 13.1%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu24 :  8.9%us, 78.2%sy,  0.0%ni, 12.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu25 :  4.0%us, 86.1%sy,  0.0%ni,  9.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu26 :  9.3%us, 81.4%sy,  0.0%ni,  9.3%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu27 : 31.8%us, 48.2%sy,  0.0%ni, 20.0%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu28 :  9.4%us, 76.0%sy,  0.0%ni, 14.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu29 : 10.8%us, 73.5%sy,  0.0%ni, 15.7%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu30 :  8.6%us, 82.9%sy,  0.0%ni,  8.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu31 :  6.7%us, 78.1%sy,  0.0%ni, 15.2%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Mem:  18095848k total,  4380656k used, 13715192k free,     1288k buffers
Swap:  1327100k total,   107788k used,  1219312k free,    11868k cached
```

システムを意識しないサービス提供の可能性

SAKURA Internet

オンデマンド
アプリケーション開発環境

アプリケーション
配信プラットフォーム



PaaSによる最適な機能提供

heroku

create　instant deployment

Create Heroku apps instantly:

```
$ sudo gem install heroku
$ heroku create sushi
Created http://sushi.heroku.com/
git@heroku.com:sushi.git
```

rock-solid ruby platform

Fast, frictionless, and maintenance free.
Deploy and scale with confidence.

40,633

apps running right now

広帯域で低価格なメディア選択

SAKURA Internet

Infiniband vs Ethernet Performance report

Short packet (64byte) | Default packet size(1500,2044byte)

558Mbits/sec (684kpps) | Xeon E5620 | 8.40Gbits/sec

247Mbits/sec (335kpps) | 2.4GHz (2P/8C) | 6.26Gbits/sec *IP over Infiniband connected mode (RFC4755)

209Mbits/sec (---kpps) | Xeon 3060 | 5.17Gbits/sec

115Mbits/sec (---kpps) | 2.4GHz (1P/2C) | 5.22Gbits/sec

Using iperf TCP mode between 2 Servers back-to-back.

Mellanox InfiniHost III Lx(MHES14-XTC) | Intel 10 Gigabit XF Server Adapter(EXPX9502AFXSR)

Price (1:20)

Source: SAKURA Internet ,Inc. 2010.DEC

Add New media type / Infiniband

SAKURA Internet

```
sudo full-upgrade -k
sudo apt-get update
sudo aptitude install module-assistant
sudo apt-get install rpm zlib1g-dev zlib1g-dbg
sudo aptitude intall byacc bison flex
sudo module-assistant prepare
sudo /opt/OFED-1.5.2/install.pl
sudo vi /etc/udev/rules.d/75-persistent-net-generator.rules
 :

set interfaces inifniband ib0 address 1.1.1.1/24
commit
```

OpenFabrics Enterprise Distribution (OFED)でIPoIB機能追加

PowerCLI / VMware

SAKURA Internet

```
Connect-VIServer <vc-ipaddress>

$esx = Get-VMHost <Target-ESXname>
$srcVm = Get-VM <original-VM-name>

1..100 |foreach { New-VM -VMHost $esx -Name ("vm-" + $_) -VM $srcVm }
```

VM(仮想マシン)自体の運用簡素化を図る

現場作業は、決して無くなりません。

SAKURA Internet

SAKURA Internet

www.sakura.ad.jp

ご静聴誠にありがとうございました

(C)Copyright 1996-2010 SAKURA Internet Inc.