Monday, March 5. 2007

### Gedankenspiel: Windows virtualisation mit Solaris,X4600 und X4500 und SunRays

Die Idee von Mika, die er in einem Kommentar zu meinem Blog abgegeben hat, hat Potential. Erhebliches Potential je länger ich darüber nachdenke. Das was ich hier schreibe, ist ein Gedankenspiel. Ich habs noch nicht ausprobiert, aber es waere echt mal einen Versuch wert. Eine ganze Reihe der Features sind allerdings noch nicht im production-grade Solaris enthalten, sondern erst in Opensolaris. Man verzeihe mir, wenn der Text vielleicht etwas wirr ist, aber ich habe ohne weitere Formulierung meine Gedanken in die Tastatur fliessen lassen. Ich bitte um rege Kommentierung.Wenn man ein System mit Prozessoren hat, die Hardwarevirtualisierung unterstützen, kann man in den virtuellen Maschinen auch ein unmodifiziertes Windows laufen lassen. Das funktioniert sehr gut.

Jetzt bringe man in das Gedankenspiel drei Maschinen ein:
Alpha- eine X4600. Auf diesem System werden die einzelnen Windows VMs gehostetthesource- eine X4500. Dieses System hat nur die Aufgabe massenhaft Disks via iSCSI zur Verfügung zu stellen. Wozu das gut ist, wird gleich deutlich.theview-Eine T1000, die lediglich die Aufgabe hat, einen SunRay-Server darszustellen. Wozu das gut ist, kommt gleich auch noch
Was man mit Alpha anfängt dürfte recht klar sein. Auf diesem System laufen die eigentlichen Windows VMs. Was allerdings nicht auf diesem System ist, sind die dazugehoerigen virtuellen Datenträger.
Diese bereitzustellen, ist aufgabe von thesource. Eine sehr interessante Frage ist doch bei Windows die schnelle Bereitstellung von Windows Systemen. Hier kann uns ZFS helfen. Auf der X4500 werden sogenannte emulated volumes bereit gestellt. Opensolaris kann wiederum diese als iSCSI Volumes im Netz zur Verfügung stellen. In der Xen-Konfiguration auf alpha werden nun diese iSCSI-Volumes von den Windows VMs genutzt. Das interessante daran ist nun, das ich mir anfangs eine Masterinstallation vorbereitete. Dieses emulated Volume kann ich genauso wie ein ZFS-Filesystem einfrieren und clonen. Will ich also eine neue Windowsinstallation bereitstellen, so brauche ich nur die Masterinstallation nehmen und clonen. Dank ZFS geht das in wenigen Sekunden. Diesen Clone stelle ich wiederum via iSCSI bereit. Charmante weitere Vorteile:
für jede neue Windows-Installation verbraucht nur das Delta zwischen den einzelnen Installationen als SpeicherAls Backupmimik könnte man sich folgendes Konzept vorstellen: Man cloniert das Filesystem und macht sofort einen Snapshot. Dann baut man den Clone zu einer vollständigen Installation aus. Wieder einen Snapshot. Mit Hilfe dieser beiden Snapshots erzeugt man einen inkrementellen Backupdatenstrom mit zfs send. Will man jetzt im Laufe der weiteren Lebenszeit der WindowsVM weitere Backups anfertigen, wird das Filesystem wiederum eingefroren, und fertigt einen Backupdatenstrom mit den Deltas zwischen vollständiger Instanz und aktuellem Zustand an.Restore funktioniert dann ähnlich. Neuen Clone mit gleichem Namen erzeugen und dann die beiden beim Backup erzeugten Differenzdatenströme einspielen (also jenen, der den Clone zu einer eigenen Installation macht, und dann den neuesten Differenzdatenstrom, der die Änderungen seit Inbetriebnahme berücksichtigt).
Der SunRay-Server theview ist dann lediglich dazu da, via rdesktop die einzelnen Windows-VM an den Arbeitsplätzen zur Verfügung zu stellen. Hier könnte man eine Logik vorstellen, die beim Einstecken der SunRay-Karte automatisch eine Session startet, die sich auf eine bestimmte VM connected via rdesktop, in der die Windowsmaschine des Nutzers läuft. Für VMware wurden solche Mechaniken schon zusammen mit SunRays implementiert.

Das kann man sogar noch weiter treiben: Mit Xen 3.0.4 soll restore/save und migrate auch HVM-Guest funktionieren. Wenn ein Nutzer meinetwegen seit einer Stunde nicht mehr an einer SunRay sein Karte eingesteckt hat, so könnte man sich einen Automatismus einfallen lassen, der die entsprechende VM beendet und den zustand abspeichert (xm save). Dies würde die Resourcen auf den Windows-VM-Maschinen wesentlich entlasten.Steckt der Nutzer dann irgendwo wieder seine Karte ein, wird die VM wieder gestartet (xm restore).

Will man das ganze jetzt ins Extreme treiben könnte man sich sogar eine Kopplung von XEN Live Migration und der Fault Management Architecture in Solaris vorstellen: Droht ein Hardwarefehler alpha ausser betrieb zu setzen, so koennte eine automatische Live Migration alle Windows-VMs auf eine andere X4600 (beta,gamma,delta usw) live migrieren. Die Nutzer an den SunRays sollten von diesem Vorgang nichts merken. Ob die SunRay kaputt geht, ist eh egal, da die Dinge r zustandslos sind. Bei Ausfall eines SunRay-Servers muesste nur eine Neukonnektierung via rdesktop erfolgen. Endlich mal ein verlässlicher PC-Arbeitsplatz. Versucht das mal mit einem herkömmlichen Arbeitsplatz;)

Ich glaub ich brauch mal ganz dringend ein paar Testmaschinen ....

Posted by Joerg Moellenkamp in German at 20:12

Kurz-Braindump:

- Volle Windows-VMs brauchen inzwischen sehr viel Speicher: spätestens mit Vista reichen 512 MB nicht. Und der Speicher sollte auch wirklich physisch da sein, Windows verwendet den eh für Filesystem-Caching, auch wenn die Anwendungen ihn nicht jederzeit ganz nutzen.

- Als Konsequenz daraus mehrere X2200 statt der X4600 verwenden: mehr Redundanz, mehr bezahlbaren Speicher pro CPU und nicht zuletzt geringere Kosten pro CPU. Auch wenn der Coolness-Faktor einer X2200 im Vergleich nicht ganz so hoch ist...
Anonymous on Mar 6 2007, 13:41

Was schätzt Ihr, wieviele Windows-Instanzen sich auf einer X2200 parallel laufen lassen?
Anonymous on Mar 6 2007, 15:18

Tja, mal wieder die ideale Gelegenheit, voll daneben zu treffen. Ich versuch's trotzdem..

Die X2200 hat 4 Cores, so ca. 10 User müßte ein Core packen, wenn man mal von Durchschnittsbenutzern ausgeht. Macht 40 User. Jeder im Schnitt max. 1GB für die VM, macht max. 40GB, die X2200 kann 64, das könnte also passen. Wenn man im Lauf des Jahres mit einem Nachfolger rechnet, der dank neuer CPUs 8 Cores bietet, reicht der Speicher immer noch, zumindest dann wenn man einen Teil der User auf 768M runterdrehen kann.

Konkret ausprobieren würde ich das aber in jedem Fall, wird auch stark von den Anwendungen abhängen. Wenn man ein paar Entwickler mit J2EE-Umgebungen auf die Kiste losläßt, reichen auch 2 pro Core...

Für sehr unterschiedliche Nutzungsprofile könnte ich dann der X4600 etwas abgewinnen, einfach weil mehr Ausgleich innerhalb eines Systems stattfinden kann. Da sich aber auch der pro Core verfügbare Speicher halbiert (bei 128GB ist doch Schluss, oder?), weiß ich nicht, ob das so prickelnd wäre.

128GB auf 16 Cores hieße 8GB pro Core, das wären dann nur max. 8 User pro Core

Vergleich:
1 X4600: 128 User
3 X2200: 120 User

Bei Einführung von Quad-Cores:

1 X4600: 128 User
2 X2200: 128 User

I/O-Durchsatz müßte man natürlich auch noch genau anschauen.

Sorgenfalten gäbe es da einige auszuräumen bevor im mich damit zu einem Kunden traute...
Anonymous on Mar 6 2007, 17:00

Die Frage ist auch, wie weit man das System oversubscribed. Also in die Kalkulation mit einbringt, das nicht alle Nutzer gleichzeitig aktiv sind.
Anonymous on Mar 6 2007, 17:35

Im Grunde d´accord mit deiner Berechnung, aber ich weiss nicht ob man wirklich ein Gigabyte pro VM rechnen muesste. Für XP und 2000 muessten eigentlich 512 ausreichend sein.

Ausserdem musst Du noch beruecksichtigen, das du die die Maschinen ueberbuchen kannst.
Anonymous on Mar 6 2007, 17:39

Die Anzahl der möglichen User wird natürlich schwer davon abhängen, welche Software verwendet wird.

Wenn eine in Bezug auf CPU-Bedarf einfache Callcenter-Anwendung (begleitet von ein wenig Word und Excel) läuft, dann kann man sicher auch 20 User auf einen Core buchen.

Beim Hauptspeicher bin ich skeptischer. Wir haben bei uns z.B. kürzlich einen eigentlich in Bezug auf Software anspruchslosen Arbeitsplatz auf 1GB aufrüsten müssen. Ok, evtl. hätte 768M auch gereicht. Grund: das Update einer einfachen Anwendung (Lexware Buchhalter) swappt plötzlich rum, wenn noch ein Notes Client und ein Excel offen ist - unter Windows XP. Einige Anwendungen sind ganz schön fett geworden, auch wenn anspruchsvolle Java Rich Client Anwendungen im Spiel sind, werden 512 schnell eng.

Sagen wir mal so: vielleicht würde ich auch versuchen, im ersten Wurf mit 512M auszukommen, aber ich hätte wirklich kein gutes Gefühl, wenn nicht genug Speichersockel für das Doppelte da wären...

Sind aber alles nur Aus-dem-Bauch-Schätzungen, mit richtig vielen Windows VMs auf einem Server habe ich auch keine Erfahrung. Was ich aus eigener Erfahrung zumindest sagen kann: logischer Speicher für eine Windows-VM sollte physisch vorhanden sein. Overcommitment von VM Speicher kann ich zumindest für Windows VMs nicht empfehlen, die Performance geht ruckzuck in den Keller, weil Windows nun einmal den kompletten Speicher für Caching verwendet.
Anonymous on Mar 6 2007, 21:27

Nachtrag:

Wieviel User so ein System vertragen würde, ist im Grunde ein Randthema.

Die angerissenen Punkte rund um Provisioning, zentrales Backup, Verfügbarkeit sind im Grunde viel spannender. Da steckt im Konzept schon einiges, das schlicht Spaß machen würde auszuprobieren - vom Mehrwert mal ganz abgesehen.

Möchte das Thema daher nicht zu sehr mit Sizing-Diskussion strapazieren.
Anonymous on Mar 6 2007, 21:37

Noch ein Kommentar zu rdesktop. Sun bietet "neu" native das RDP Protokoll an. Man müsste noch anschauen wie/ob das weiterleiten von USB Devices (z.B. Memorysticks) funktioniert.

Wenn ich mir die obere Lösung vorstelle, gibt es einen sehr wichtigen Aspekt für die Akzeptanz der Lösung.

Man "klaut" allfälligen Windows-Desktop-Teams nicht die interessante Arbeit (Desktop Builds generieren etc.), aber rationalisiert sehr viel langweilige Aufgaben (OS Provisioning, Neuinstallationen).

Das ganze muss man eben auch noch dem (Windows-)Management verkaufen können, vorallem wenn die Lösung aus der Unix-Schiene kommt. Aber mit einem attraktiven Price-Tag lässt sich heutzutage viel machen
Anonymous on Mar 6 2007, 21:43

Ich habe hier zwar keines der angesprochenen Systeme stehen, aber noch ein ungenutzes Core2Duo-Notebook ... das sollte sich doch als experimentierplattform eignen.

Ich werd mal sehen, das mir zwei SunRays fuer dieses Experiment ins Haus kommen ... ich will sowieso mal auf die SunRay2 umsteigen.
Anonymous on Mar 6 2007, 21:55

Ich habe mir mal ein Scratchpad dafuer unter http://wiki.c0t0d0s0.org/index.php/Desktopvirtualisierung_mit_XEN angelegt
Anonymous on Mar 6 2007, 22:14

Habe mal schnell die Hardwarekosten überschlagen. Laut meinem Kollegen kostet bei uns ein standard Desktop PC mit 1GB Ram ca. 500$ in der Anschaffung (ich rechne in Dollar, da ich den Sun US Shop benutze).

Wenn man zur Sicherheit davon ausgeht, dass man 1GB Ram für eine Windows-Kiste benötigt kriegt man auf die X2200 30 Instanzen. Ich lasse mal den externen Storage weg, da hier durch den Klonmechanismus die Erfahrungswerte fehlen. Jedoch kann man bei "classic" Desktops von 3-6GB für OS/Apps ausgehen. (=>Thumper ca. 2$/GB )

Der Preis beläuft sich somit für die Server-Hardware (alles ohne Wartung/Rabatte) auf $14000.

Dazu kommen 30 Sun Ray 2 (ohne Tastatur/Bildschirm, welche in den Firmen meist schon existieren/bzw. bei beiden Konfiguration angeschafft werden müssen). Die 30 SunRays kosten $11220.

Die Gesamtsumme beläuft sich auf $25321

Dies macht einen Anschaffungspreis pro ThinClient von $844.

Im Vergleich zum Desktop PC beträgt der Unterschied also ca. $350.

Ein solcher Unterschied ist meines Erachtens recht wenig, wenn man die TCO anschaut, welche ein Mehrfaches des Anschaffungspreises ist.

Und eben bei der Reduzierung des TCO denke ich, könnte die obere Lösung unschlagbar sein.
Anonymous on Mar 7 2007, 22:31

Was man vielleicht auch noch mit in die Gleichung bringen sollte: Produktivitätsgewinn. Wenn es mich nunmehr nur noch wenige Minuten kostet, einen Arbeitsplatz auf den Zustand von beispielsweise heute morgen zu bringen, statt zum System zu laufen, es abzuholen, ein neus System aufzustellen, es online bringen, dann ist da ein aeusserst interessanter Faktor. Ich rede nicht von den Kosten fuer den Admin, sondern von den Kosten für den Mitarbeiter, der gefühlt nicht ohne PC arbeiten kann.
Anonymous on Mar 7 2007, 23:01

Was mir da noch einfällt: Man könnte die Installationen weitestgehend zustandslos (vulgo: leicht wegschmeissbar) auslegen, wenn man auch auf diese virtuellen Maschinen in eine Anmeldedomän einfügt, und die Userprofile dann auf einem Samba Filesserver auf einem der Thumper vorhält.
Dann sollten sich auch die ganzen Boot-iSCSI-Volumes auf dauer recht ähnlich bleiben.
Anonymous on Mar 7 2007, 23:07

Korrigier mich bitte wenn ich falsch liege.

ZFS benutzt ja den freien Speicher als Cache. Da das Basis Image (Grösse < 5GB) ja bei allen Instanzen gleich ist, sollte dieses durch den ständigen Zugriff ja dauernd im Memory liegen.

Disk Performance würde dann nur noch für die Deltas benötigt.
Anonymous on Mar 8 2007, 07:08

Antwort im Stile nach Radio Eriwan: Im Prinzip ja ... praktisch muesste man sich das mal angucken, da ja auch andere anwendungen,

die auf den Thumpern laufen den Cache moeglichweise befüllen. Aber bei 16 GB Cache (naja, etwas weniger) sollte das nicht so das Problem sein ...
Anonymous on Mar 8 2007, 07:23

Habe mal das Stromsparpotential ausgerechnet.

Ausgehend von der obigen X2200 die für 30 Clients ausgelegt ist:

Leistungsaufnahme X2200 = 376W

Leistungsaufnahme SunRay2 = 4W

Leistungsaufnahme PC = 60W

Betrieb 1xSunRay+Serveranteil= 17W

Ersparnis ca. 40W

Macht bei 10c/kWh 35€ Ersparnis pro Jahr und Arbeitsstation.

Oder anderst gesagt, bei 30 Arbeitsplätzen spart man 1000€/Jahr.

Alles mit Vorsicht zu geniessen natürlich
Anonymous on Mar 9 2007, 21:38