FAS3020 running OnTap 7.3.2
This is called filer2 and has a serial number of 3080936
2TB volume named VMWare setup as an NFS export called VMWare
VMWare volume added to all 4 of my ESXi 4.1 hosts
31 VMs housed on this datastore:

- 16 Windows Server 2003, mix of Enterprise and Standard
- 14 Windows Server 2008 R2
- 1 Ubuntu 32 bit

I also have two templates in this datastore, one for Windows Server 2003 and one for Windows Server 2008 R2.
All VMs are aligned using mbralign with the exception of the Ubuntu VM
This volume IS NOT thin provisioned, the output of the vol option VMWare command is guarantee=volume

Deduplication has been enabled on this volume since December 2009, over a year ago. I kicked off deduplication by using:
sis start –s /vol/VMWare

This command scans the entire volume and deduplicates any existing data. I have deduplication setup via a schedule to run each evening at 11pm. Below is the output of the sis config command showing this schedule.

```
filer2*> sis config
Path                Schedule
/vol/VMWare             sun-sat@23
```

I have been reading Netapp's Deduplication and Deployment Guide:
http://www.netapp.com/us/library/technical-reports/tr-3505.html
On page 16 it states that the space savings for VMware is typically 70%

There are also many Netapp blogs, case studies, and press releases that tout the space savings for dedup of VMWare volumes of at least 50%, here are three examples:
http://partners.netapp.com/go/techontap/matl/dedup_VM.html
http://www.netapp.com/us/communities/tech-ontap/tot-sddpc.html
http://www.netapp.com/us/communities/tech-ontap/tot-dedupe-unstructure-0409.html

I am having two problems related to dedup on this VMWare volume:

1. The method that Netapp provides to measure the space savings appears to calculate this value incorrectly.
2. Even when I calculate the space savings by hand it is less than what it should be.

We use SMVI to backup the VMWare volume so each evening two snapshot copies of this volume are taken, one at 11:30pm and one at midnight. This snapshot is taken after the dedup operation has run at 11pm. Right now I have snapshots from December 1 to January 1 on the filer so I do have any snapshot that were made before deduplication was enabled. In other words, I do not have any non-deduplicated data locked in snapshots.

Here is the output of df –s command that only shows a 20% space savings on my VMWare volume.

```
filer2*> df -s /vol/VMWare
Filesystem              used     saved     %saved
/vol/VMWare/       1037095384  255761016        20%
```

Here is the output of the df –h and the df –sh commands:

```
filer2*> df -h /vol/VMWare
Filesystem              total      used     avail capacity  Mounted on
/vol/VMWare/          2048GB     989GB    1058GB     48%  /vol/VMWare/
/vol/VMWare/.snapshot      0KB     692GB       0KB    ---%  /vol/VMWare/.snapshot

filer2*> df -sh /vol/VMWare
Filesystem              used     saved     %saved
/vol/VMWare/           989GB     243GB        20%
```

Here is my analysis of these numbers:

Total volume usage	=	989GB
Snapshot usage	=	- 692GB
Live data	=	297GB
Dedup amount	=	243GB

So it appears that I have 297GB of live data and 243GB of deduped data. How should the percent savings be calculated? This is what I think df –s is doing:

$$Percent\ saved = \frac{Dedup\ amount}{Total\ volume\ usage + Dedup\ amount} \times 100$$

$$Percent\ saved = \frac{243}{989 + 243} \times 100 = 19.72 \approx 20\%$$

This seems to be the incorrect method to calculate space savings especially in light of space savings for dedup of VMWare volumes should be 50% or higher. I am thinking that this should be the method used to calculate space savings:

$$Percent\ saved = \frac{Dedup\ amount}{Live\ data + Dedup\ amount} \times 100$$

$$Percent\ saved = \frac{243}{297 + 243} \times 100 = 45\%$$

If I have the proper method, then the second issue is my space savings seem to be low.