Amsterdam UMC
Universitair Medische Centra

Information Lifecycle Data Managment

Amsterdam University Medical Center
Location VUmc

Patrick Dekkers
Jaap Jan Ouwehand

Introduction

Patrick Dekkers
20 years in IT
Storage Specialist at Amsterdam UMC Locactie VUmc

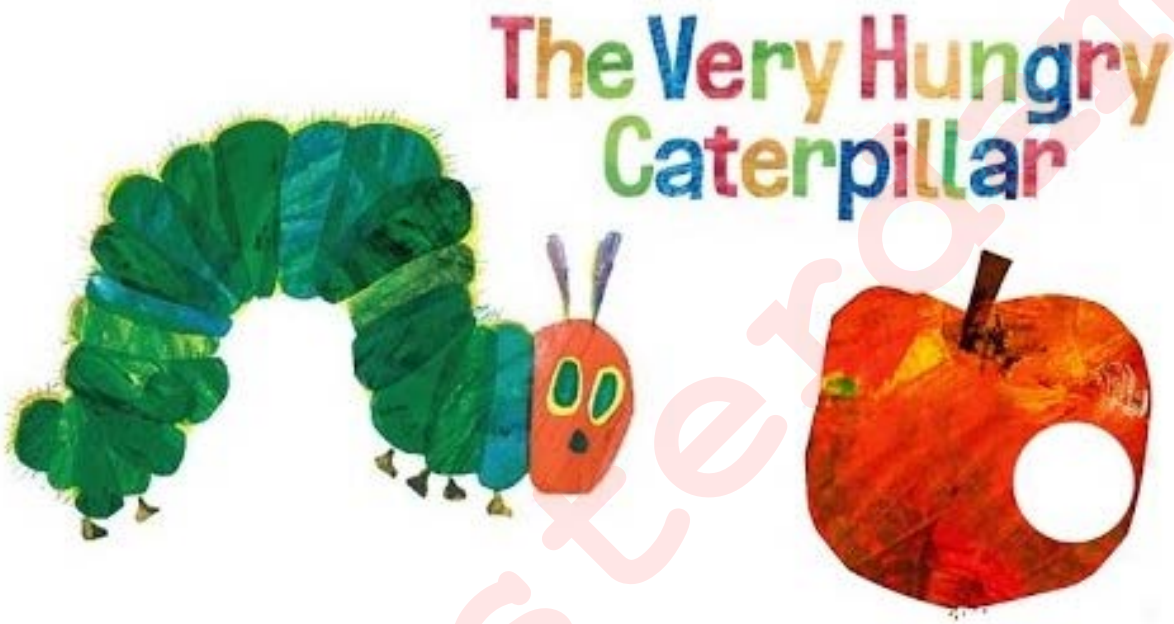
Jaap Jan Ouwehand
22 years in IT
Storage Specialist at Amsterdam UMC Locactie VUmc

VUmc: 750 beds
8.000 employees
Patient 200.000
ICT 120

Amsterdam UMC
Universitair Medische Centra

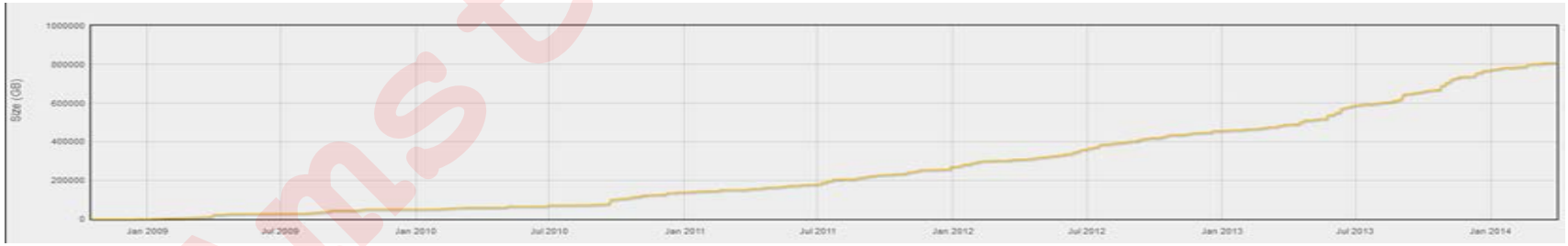
AMC: 1.000 beds
1.0000 employees
ICT 300

The Very Hungry ~~Caterpillar~~ Researcher



Challenge before a central storage platform

- In 2009 we started measuring the used storage
- Data grow annually by ~ 33% compared to the year before
- Centralized and non-centralized storage
- No backup platform
- Pressure on Storage management to keep up with growth
- Result - High annual Costs increase / Price per TB above market average



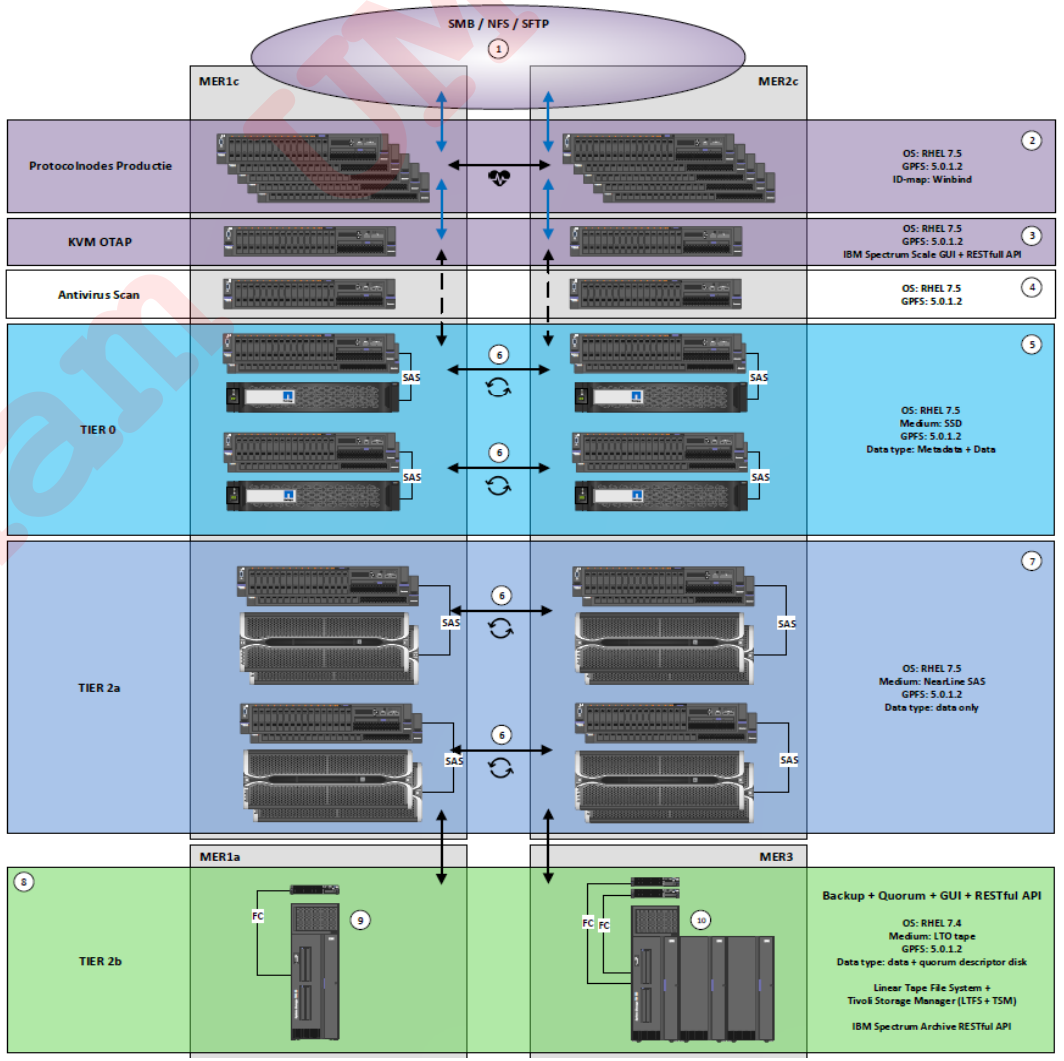
Product choice

IBM Spectrum Protect	Optimized hybrid cloud data protection to reduce backup costs by up to 53 percent
IBM Spectrum Archive	Fast data retention that reduces TCO for active archive data by up to 90%
IBM Spectrum Scale	High-performance, highly scalable hybrid cloud storage for unstructured data driving cognitive applications

Spectrum Scale Design

Store4Ever is a software-defined highly available file-based storage platform with auto-active tiering. This service provides cost-effective, large-scale, scalable, high-performance and archive storage for (hot, cold and frozen) unstructured data like Office documents, PACS images, Genome, homefolders and Cloud.

- ✓ **Protocol Nodes – SMB / NFS / SFTP**
- ✓ **KVM OTAP – Virtual test environment**
- ✓ **TIER 0 - SSD Storage**
- ✓ **TIER 2a - NearLine SAS**
- ✓ **TIER 2b – Archive + Backup + Quorum**



Stretched datacenter

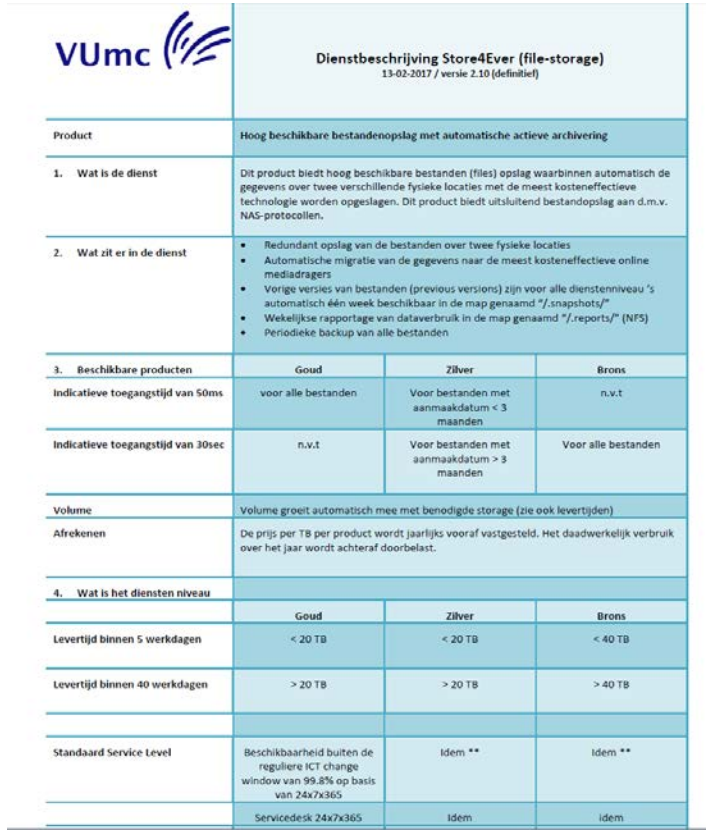
How we handle storage requests

For new requests we provide end users / researchers a menu

We have 3 flavors and store data on 2 file systems.

- Gold > Always on disk
- Silver > 3 months on disk, then the data automatically drops to online tape archive
- Bronze > To online tape archive as soon as possible

*All data is backed up on tape before going to the online tape archive



VUmc

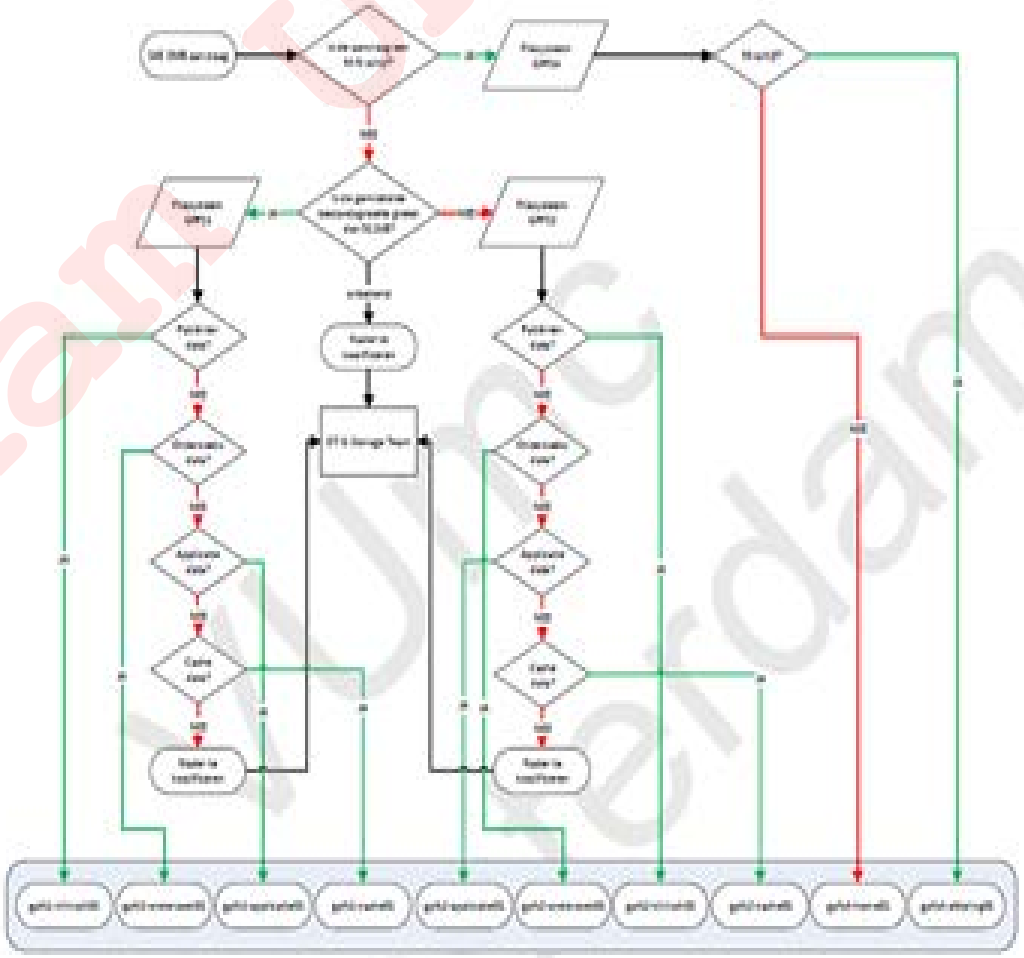
Dienstbeschrijving Store4Ever (file-storage)
13-02-2017 / versie 2.10 (definitief)

Product	Hoog beschikbare bestandenopslag met automatische actieve archivering		
1. Wat is de dienst	Dit product biedt hoog beschikbare bestanden (files) opslag waarbinnen automatisch de gegevens over twee verschillende fysieke locaties met de meest kosteneffectieve technologie worden opgeslagen. Dit product biedt uitsluitend bestandopslag aan d.m.v. NAS-protocollen.		
2. Wat zit er in de dienst	• Redundant opslag van de bestanden over twee fysieke locaties • Automatische migratie van de gegevens naar de meest kosteneffectieve online mediadragers • Vorige versies van bestanden (previous versions) zijn voor alle dienstenniveau 's automatisch één week beschikbaar in de map genaamd "/.snapshots/" • Wekelijkse rapportage van dataverbruik in de map genaamd "/.reports/" (NFS) • Periodieke backup van alle bestanden		
3. Beschikbare producten	Goud	Zilver	Brons
Indicatieve toegangstijd van 50ms	voor alle bestanden	Voor bestanden met aanmaakdatum < 3 maanden	n.v.t
Indicatieve toegangstijd van 30sec	n.v.t	Voor bestanden met aanmaakdatum > 3 maanden	Voor alle bestanden
Volume	Volume groeit automatisch mee met benodigde storage (zie ook levertijden)		
Afrekenen	De prijs per TB per product wordt jaarlijks vooraf vastgesteld. Het daadwerkelijk verbruik over het jaar wordt achteraf doorbelast.		
4. Wat is het diensten niveau			
	Goud	Zilver	Brons
Levertijd binnen 5 werkdagen	< 20 TB	< 20 TB	< 40 TB
Levertijd binnen 40 werkdagen	> 20 TB	> 20 TB	> 40 TB
Standaard Service Level	Beschikbaarheid buiten de reguliere ICT change window van 99.8% op basis van 24x7x365	Idem **	Idem **
	Servicedesk 24x7x365	Idem	Idem

GDPR - General Data Protection Regulation

For privacy requirements GDPR sensitive data needs to be stored in a save place. We use a flow chart to determine data sensitivity on prerequisites like

- Application data
- Clinical data (GDPR)
- Research data



Data protection and backup to tape

- Two flavors of data protection
  - Snapshots (5 times a day)
  - Backup (traditional data to tape)
- Backups are:
  - Incremental forever
  - Based on a script that runs 24/7/365.
  - Not periodic, but cycling

```
/gpfs1/local/mmfs/scripts/mmbackup-wrapper-v2.2.sh -f gpfs1,gpfs2
/gpfs1/local/mmfs/scripts/mmbackup-wrapper-v2.2.sh -f gpfs4
```

Backup script

```
                        # uniq snapshot name with pidno and random
                        snapname=backupsnap-$$-$RANDOM

                        # run a mmbackup
                        if [ "$test" == "1" ]; then
                                echo "TEST MODE : Executing /usr/lpp/mmfs/bin/mmbackup $conf_filesetroot -L $debuglevel -t $mmbackup_type -S $snapname -s /gpfs-scratch/.globalWorkDir/$filesystem/ -g
/gpfs-scratch/.globalWorkDir/$filesystem/ --scope inodespace >> $logfile"
                        else

                                # create an independent fileset snapshot
                                /usr/lpp/mmfs/bin/mmcrsnapshot $filesystem $snapname -j $conf_fileset

                                echo "Executing /usr/lpp/mmfs/bin/mmbackup $conf_filesetroot -L $debuglevel -t $mmbackup_type -S $snapname -s /gpfs-scratch/.globalWorkDir/$filesystem/ -g /gpfs-scratc
h/.globalWorkDir/$filesystem/ --scope inodespace >> $logfile"
                                /usr/lpp/mmfs/bin/mmbackup $conf_filesetroot -L $debuglevel -t $mmbackup_type -S $snapname -s /gpfs-scratch/.globalWorkDir/$filesystem/ -g /gpfs-scratch/.globalWorkDir
/$filesystem/ --scope inodespace >> $logfile

                                # delete the snapshot
                                /usr/lpp/mmfs/bin/mmdelsnapshot $filesystem $snapname -j $conf_fileset
```

Spectrum Archive - Online Archive

- We use Spectrum Archive for online tape archive.
- Multi-tape configuration.
- Archive data is stored on 2 independent tape robots in 2 different data centers.
- Storing data on tape vs disk is very cost effective and green.

```
[root@sn-ltfs-01-01 ~]# ltfsee info pools
Pool Name       Total(TiB)   Used(TiB)   Free(TiB)   Reclaimable(TiB)   Tapes  Type  Library  Node Group
lib0101_tier2b      1520.2      1031.4       488.8                3.4    349  LTO   lib0101  G0
lib0101_test           2.2         0.0         2.2                0.0      1  LTO   lib0101  G0
lib0201_tier2b      1508.4       995.8       512.5                2.7    345  LTO   lib0201  G0
lib0201_test           2.2         0.0         2.2                0.0      1  LTO   lib0201  G0
```

Spectrum Archive script

```
        # run policy
        /usr/lpp/mmfs/bin/mmapplypolicy $datadir -P $tmp_policy -g $globaldir -N $nodes

        # convert to ltfsee format and sort filelist
        sed -i -r 's/^(.+)\s+\-\-\s+/ -- /g' $tmp_policy_filelist
        cat $tmp_policy_filelist | sort > $filelist
        rm $tmp_policy $tmp_policy_exec $tmp_policy_filelist

        # split filelist in chunks
        split -a 4 -d -l $lines $filelist $globaldir/ltfsee-filelist-$uniqid\-

        # remove temp filelist
        rm -f $filelist

done

# migrate files to ltfsee
for file in $(ls -rt $globaldir/ltfsee-filelist-*);
do
        time /opt/ibm/ltfsee/bin/ltfsee migrate -s $file -p lib0101_tier2b@lib0101 lib0201_tier2b@lib0201
        sleep 30
        rm -f $file;
done
```

Benefit

- Always online
- Low cost per TB
- Environmentally friendly, green

Future

- Use of metadata (Spectrum Discover)
- Integration with Cloud (Azure)

Patrick Dekkers
p.dekkers@amsterdamumc.nl
https://www.linkedin.com/in/patrickdekkers/

Jaap Jan Ouwehand
j.ouwehand@amsterdamumc.nl
https://www.linkedin.com/in/jaapjanouwehand/