We make IT happen

FROM INFRASTRUCTURE TO SOLUTION
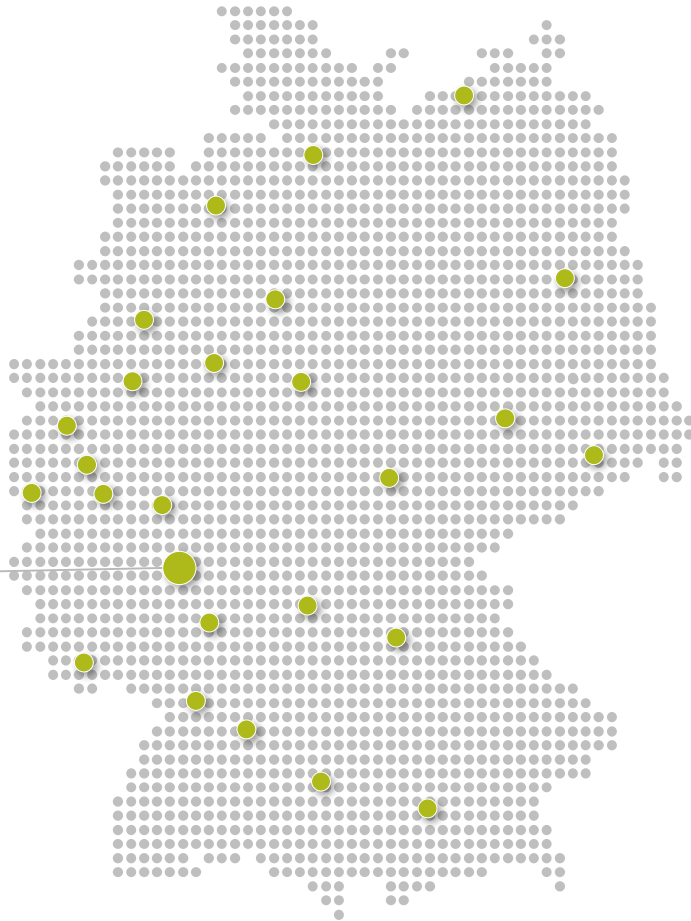
**Simplified and accelerated –
Spectrum Scale as data hub in medical research**

jochen.zeller@sva.de

/ COMPANY

26
Locations

Wiesbaden

7 TOP
Industries

Automotive

Retail

Public

Machine & Plant Construction

Finance & Insurance

Telecommunication

Healthcare

/ COMPANY

Employees

2.500
current

2.200
2021

1.500
2020

1.000
2019

Sales volume in Mio €

887

1.036

1.254

2019 2020 2021

/ THE "BEFORE SPECTRUM SCALE" ERA

Object storage

Compute

S3

NFS

SMB

NAS filer

- Not fast, not parallel
- Complex data access
- Manual data movement
- And this is a simplified map
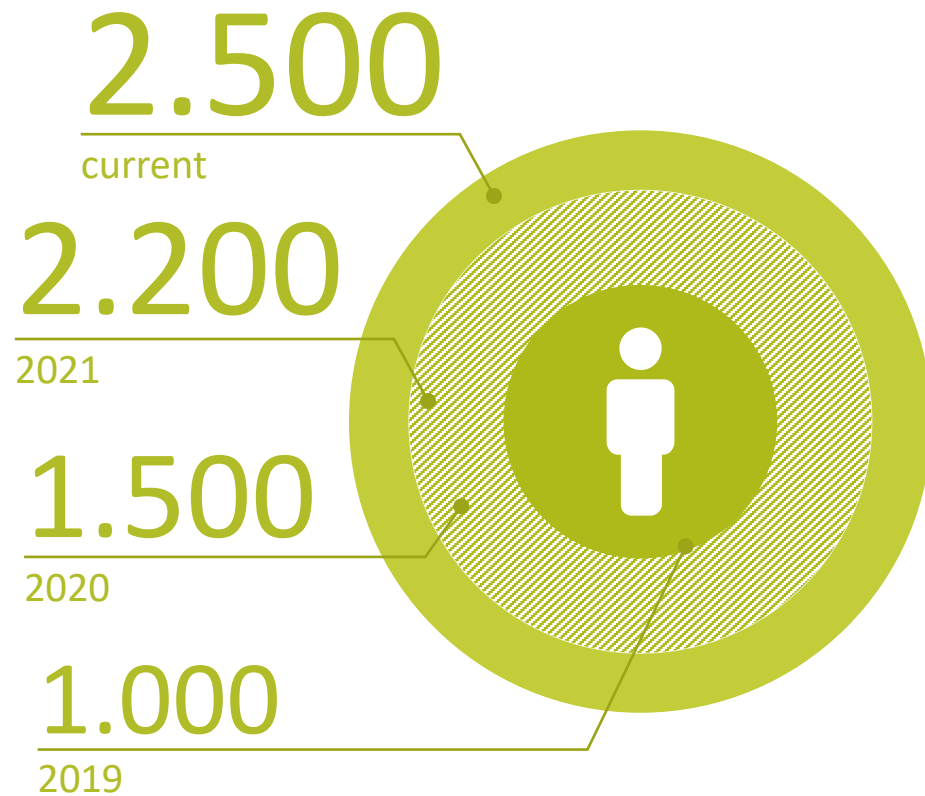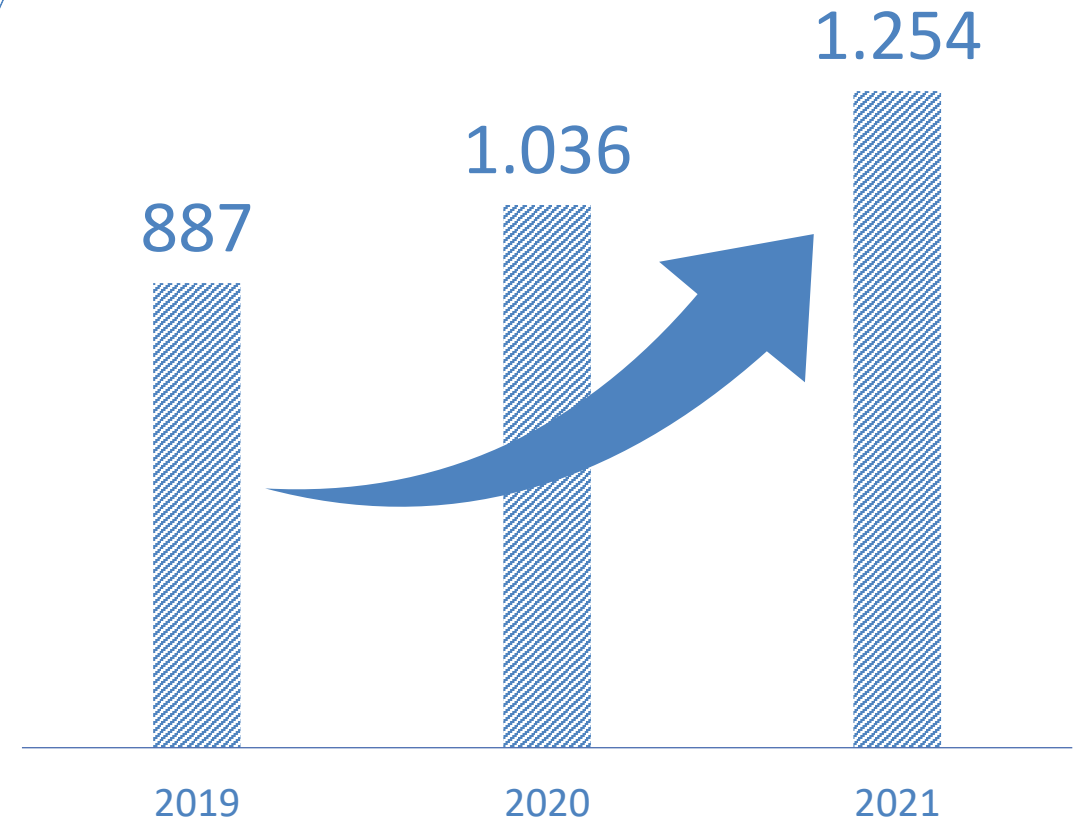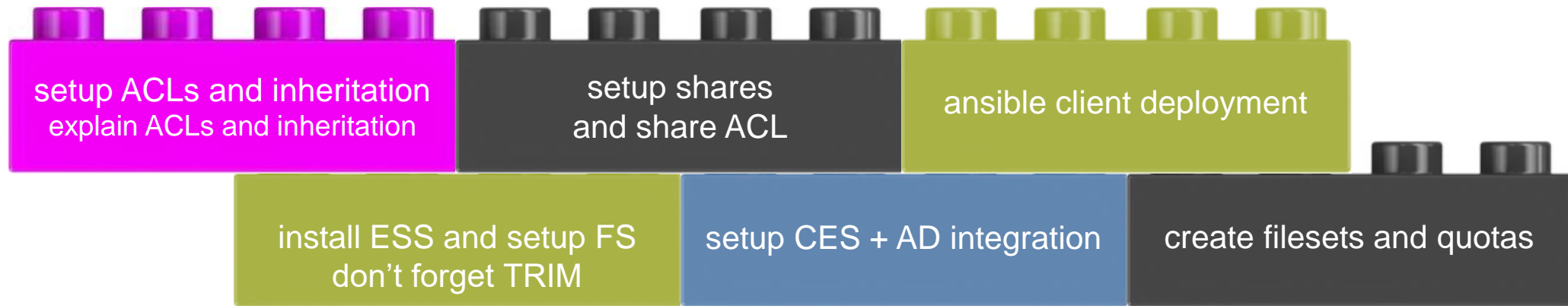
SVA
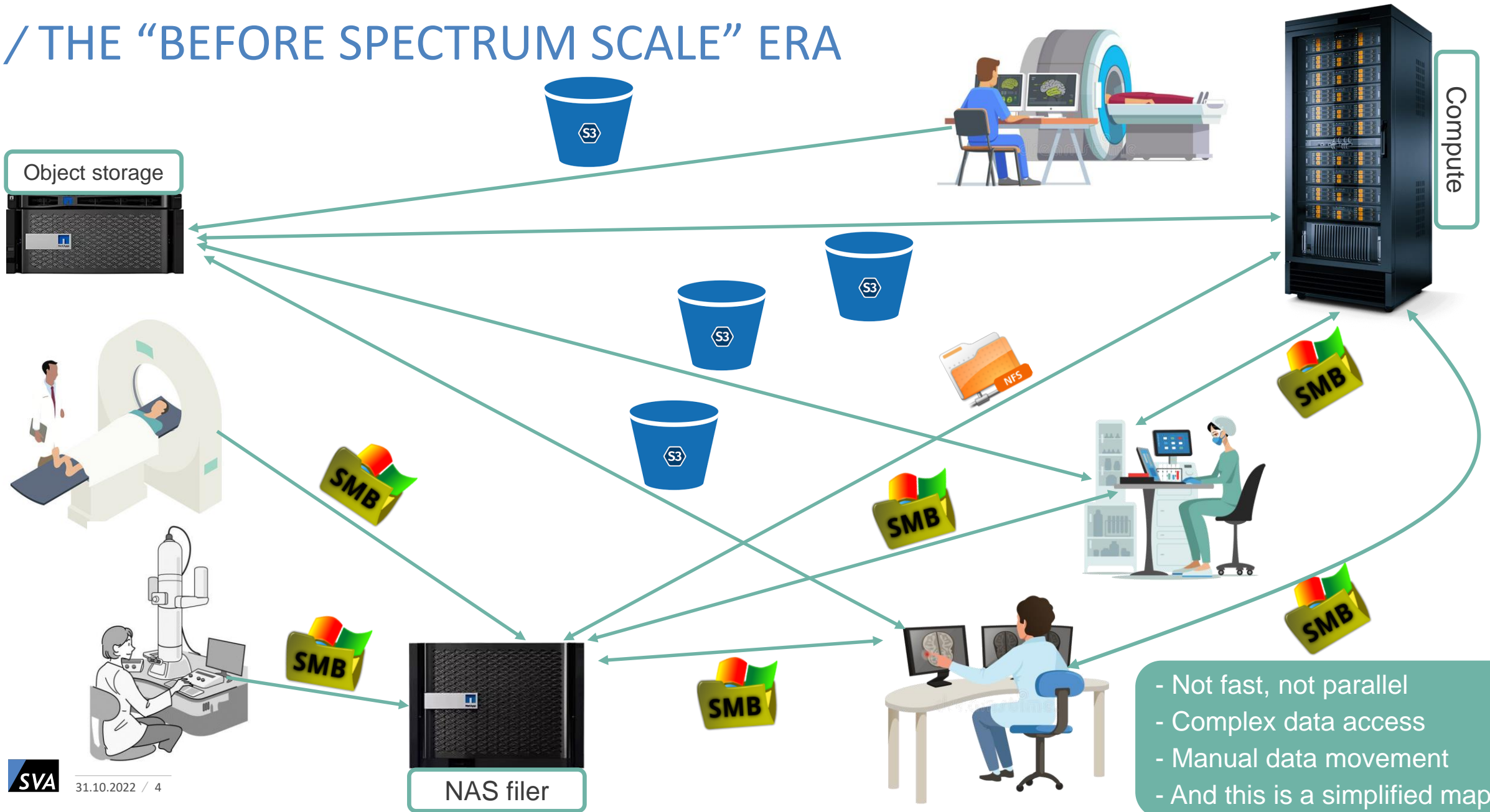
/ BUT IN THE BEGINNING …

… we did not know all this

… and wanted to implement an AI system with storage

- IBM ESS3200 with 24 x 7,68 TB NVMe (105TB)

- 2 CES nodes for SMB and NFS

- DGX A100 with 8 GPU's

- Vmware ESXi systems with 2x A100 GPUs per node

- HDR Infiniband network

… and provide a fast scratch to the HPC system

HDR network

GPU nodes
ESXi

2 CES nodes

EMS +
ESS3200

Nvidia A100
GPU system
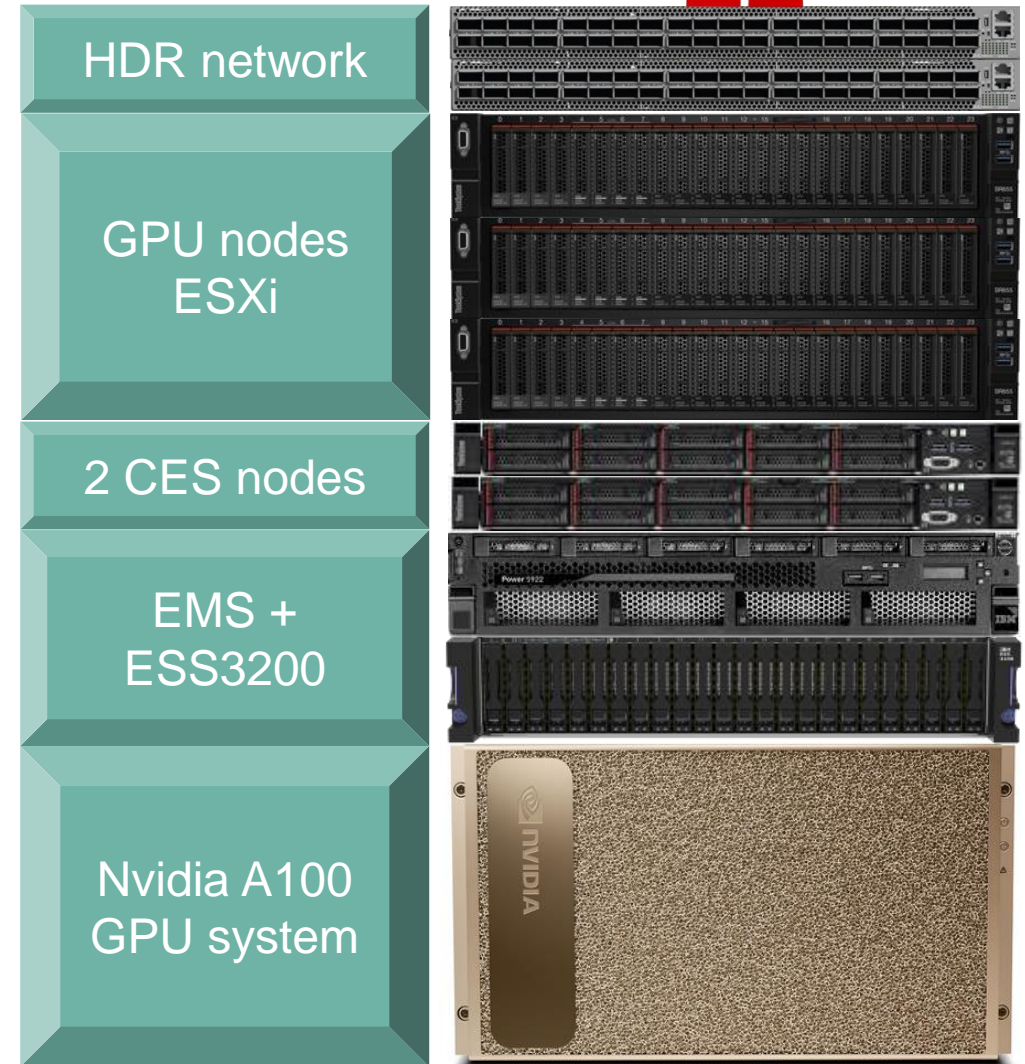
/ HOW TO – WITH THE LEGO PRINCIPLE – BRICK BY BRICK

setup ACLs and inheritance
explain ACLs and inheritance

setup shares
and share ACL

ansible client deployment

install ESS and setup FS
don't forget TRIM

setup CES + AD integration

create filesets and quotas

So far so good, but …

… where does the data come from?
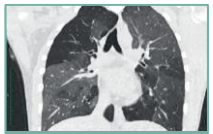
… how do you get users to use the ESS?

Could AFM help at this point?

/ NOW IT'S ALL ABOUT AFM

- AFM – Active File Management

- A pretty cool tool to cache or migrate data into Spectrum Scale

  - From other Spectrum Scale file systems, NFS shares or S3 buckets

AFM HOME NFS

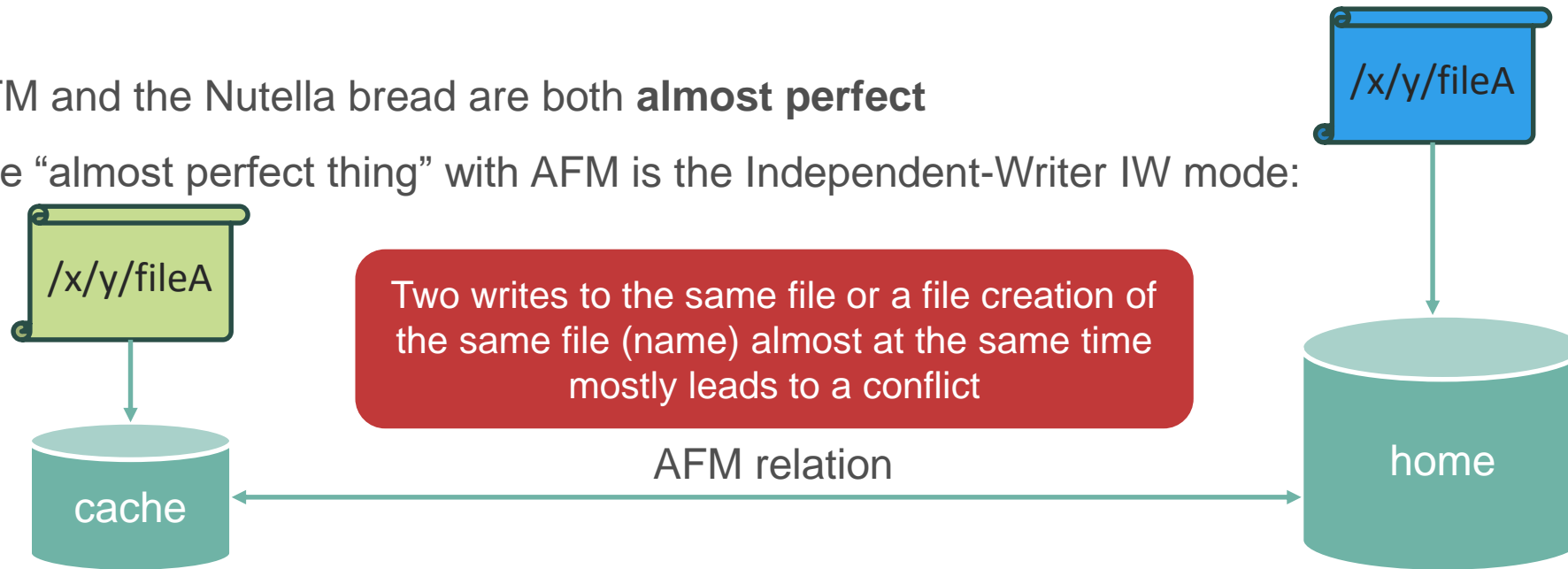/CT/lung/COPD ⟷ /projects/data/CT/lung/COPD

AFM Cache

- More about AFM:

  - Today 15:00: "DIY Tierung and Online FS Migration with AFM"

  - Breakout: "AFM Enhancements and Use Cases & Not enough money for an all flash HPC storage – a brief cheating guide"

- But attention! AFM is like Nutella bread!

/ AFM AND THE NUTELLA BREAD

- AFM and the Nutella bread are both **almost perfect**

- The "almost perfect thing" with AFM is the Independent-Writer IW mode:

/x/y/fileA

/x/y/fileA

Two writes to the same file or a file creation of the same file (name) almost at the same time mostly leads to a conflict

AFM relation

cache

home

- To keep the system running, AFM stores the conflict in .ptrash or .pconflicts (and notifies mmhealth)

- An administrator must clean up the conflict, or the .ptrash / .pconflicts

- Ouch. Always use "cp" to bring a file from .ptrash / .pconflicts back to the cache – "mv" is not noticed by AFM.

But what the hell does that have to do with Nutella bread?

/ AFM AND THE NUTELLA BREAD

- The "almost perfect thing" with Nutella bread:

  Which side does the Nutella bread fall on when you drop it?

- Some ideas to solve the problems:

The Nutella problem	The AFM problem
influence the gravity around the Nutella bread	create an AI that recognizes what the users had in mind when writing parallel and merges the changes from home and cache
the Nutella bread drone. The drone always hovers under the Nutella bread, when it falls the drone catches the bread before gravity turns it over	compare all writes on all files on all caches and home and send users polite popups if the try to update the same file at the same time
or, the easy way	
don't drop your Nutella bread	prevent updates to the same file at the same time

/ BUILD YOUR OWN DATA HUB

Do you remember "The before Spectrum Scale era"?

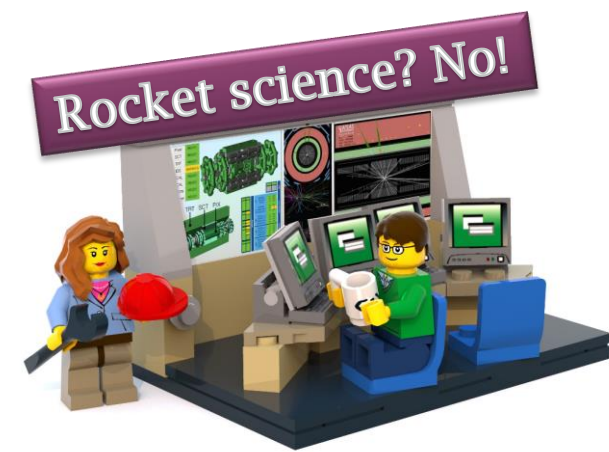Let's operate Spectrum Scale in!

- Remove old access paths

- Create AFM relation to the object storage

- Create SMB Share

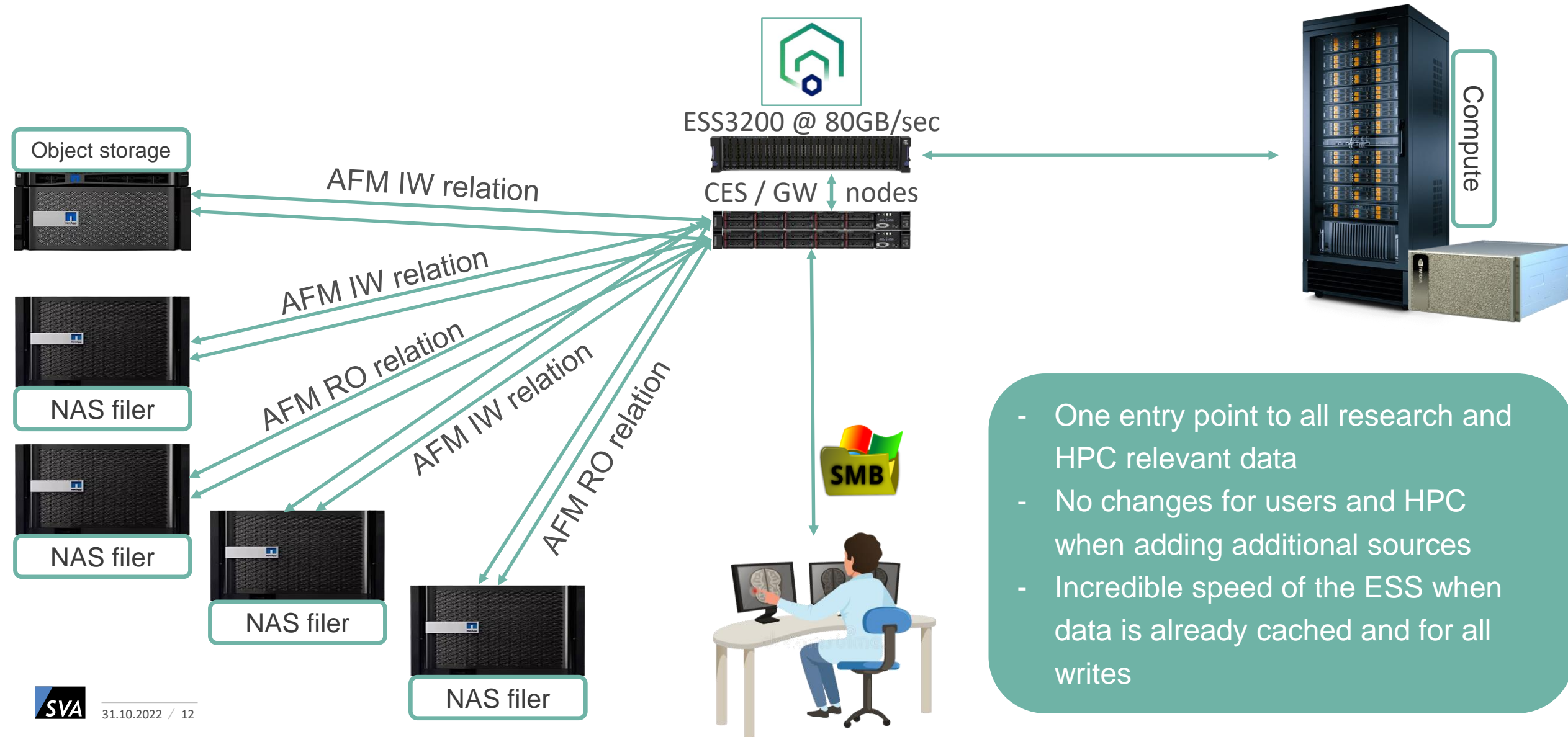- Data is available in HPC, by Spectrum Scale Protocol and RDMA
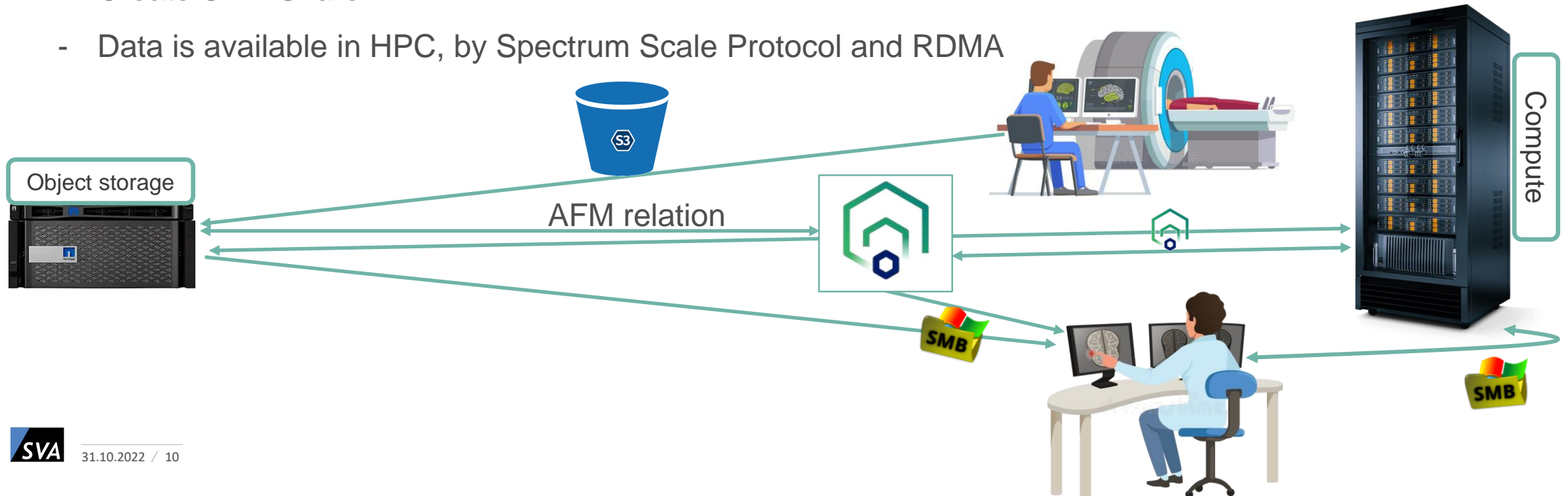
/ BUILD YOUR OWN DATA HUB

AFM looks so simple - and it is!

- Define your GW nodes: `mmchnode -N cesNodes --gateway`

- Install afm.cos on your GW: `dnf install gpfs.afm.cos`

- Provide access credentials: `mmafmcoskeys CTbucket set <accessKey> <accessSecret>`

- Create fileset with AFM relation: `mmafmcosconfig projects CT --end-point https://hugeOBJstore:443 --bucket CTbucket --mode iw --object-fs --uid 20000 --gid 9999 --dir /projects/data/CT`

- After a few minutes, try `"ls -la /projects/data/CT"` and the content of the bucket will be listed in GPFS

- Are you interested in file prefetching, controlling the file system space used by the cache fileset and tuning?

  → Visit breakout: "AFM Enhancements and Use Cases & Not enough money for an all flash HPC storage – a brief cheating guide"

/ AND HOW DOES THAT LOOK TO THE RESEARCHER? ANY ADVANTAGES?

ESS3200 @ 80GB/sec

CES / GW nodes

Object storage

AFM IW relation

AFM IW relation

AFM RO relation

AFM IW relation

AFM RO relation

NAS filer

NAS filer

NAS filer

NAS filer

Compute

SMB

- One entry point to all research and HPC relevant data
- No changes for users and HPC when adding additional sources
- Incredible speed of the ESS when data is already cached and for all writes

/ THERE WAS STILL SOMETHING - THREE QUESTIONS

- *Where does the data come from?*

    - the data is already available in the Spectrum Scale FS – thank you AFM!

- *How do you get users to use the ESS?*

    - Show them that applications run faster on ESS as on NetApp NFS filers

    - BUT: They must change paths in their applications. This can be a tough piece of persuasion!

- *Could AFM help at this point?*

    - Can we fix it with AFM? Yes, we can!

/ SUMMARY

- Meanwhile, AFM is a reliable tool

- Don't forget the Nutella bread! Take care when using IW mode

- AFM helps to bring research data from many sources into your HPC - and to your users

- AFM could help to reduce complexity in environments with many different data sources

```
The good thing about Spectrum Scale is that you can design and adjust a lot.
The bad  thing about Spectrum Scale is that you can design and adjust a lot.
```

/ THE END!

/ CONTACT

JOCHEN ZELLER

IT Architect

Technical Leader IBM Spectrum Scale

Phone.:	+49 151 180 256 77
Mail:	jochen.zeller@sva.de