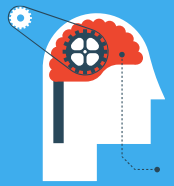


Spectrum Scale Expert Talks

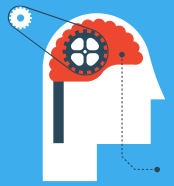
Episode 9:

Continental: Deep Thought – An AI Project for Autonomous Driving Development

IBM Spectrum Scale

Show notes:
www.spectrumscaleug.org/experttalks

Join our conversation:
www.spectrumscaleug.org/join



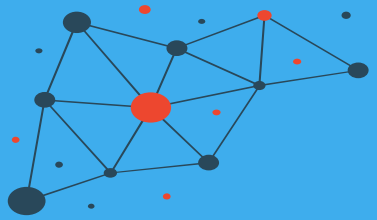
SSUG::Digital

Welcome to digital events!

IBM Spectrum Scale

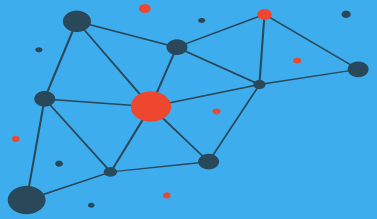
Show notes:
www.spectrumscaleug.org/experttalks

Join our conversation:
www.spectrumscaleug.org/join



About the user group

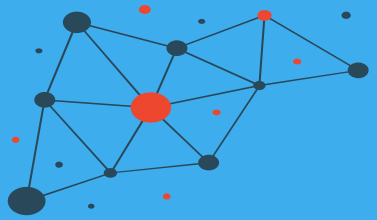
- Independent, work with IBM to develop events
- Not a replacement for PMR!
- Email and Slack community
- https://www.spectrumscaleug.org/join



We are …

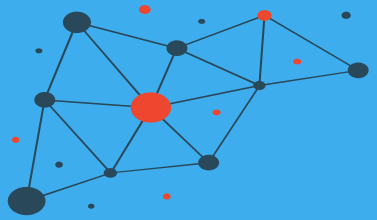
- Simon Thompson (UK)
- Kristy Kallback-Rose (USA)
- Bob Oesterlin (USA)
- Bill Anderson (USA)
- Chris Schipalius (Australia)

IBM CHAMPION



Check https://www.spectrumscaleug.org/experttalks for charts, show notes and upcoming talks

- Past talks:
  - 001: What is new in Spectrum Scale 5.0.5?
  - 002: Best practices for building a stretched cluster
  - 003: Strategy update
  - 004: Update on performance enhancements in Spectrum Scale (file create, MMAP, direct IO, ESS 5000)
  - 005: Update on functional enhancements in Spectrum Scale (inode management, vCPU scaling, NUMA considerations)
  - 006: Persistent Storage for Kubernetes and OpenShift environments
  - 007: Manage the lifecycle of your files using the policy engine
  - 008: Multi-node scaling of AI workloads using Nvidia DGX, OpenShift and Spectrum Scale
- Today:
  - Nov 16: Continental: Deep Thought – An AI Project for Autonomous Driving Development
  - Nov 16: Data Accelerator for Analytics and AI (DAAA)
- Next:
  - Nov 18: User Meeting at SC20 (Session 2) – What is new in Spectrum Scale 5.1? https://www.spectrumscaleug.org/event/sc20-meeting-session-2-what-is-new-in-spectrum-scale-5-1/



Speakers

- David Enenkel (Continental)
- Viktor Pál (Continental)
- Jochen Zeller (SVA)

Continental
The Future in Motion

Senses for Safety

Driver assistance systems help save lives

Advanced Driver Assistance Systems

Project Deep Thought – Continental Supercomputer – SC20 16.11.2020

David Enenkel Head of IT Operations ADC @ Continental

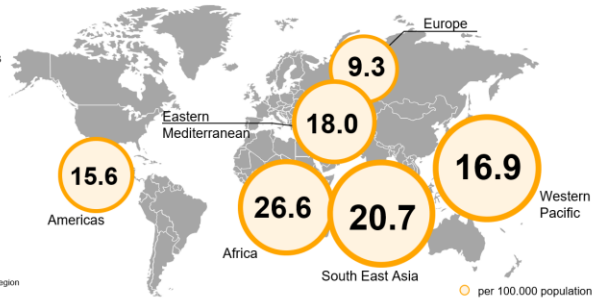
Why do we need a Supercomputer - Vision Zero

Our Motivation in our Daily Work:

Reduce the number of fatalities and those injured

Over **1,35** million people die in road accidents every year

A further **50** million are injured.



Data Source:
Global Status Report on Road Safety 2018
World Health Organization
Road traffic fatality rates per 100.000 population by WHO region

Continental

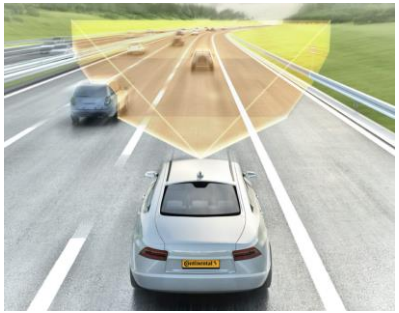
Local IT Lindau & Ulm
Internal

February 24, 2020
© Continental AG

3

Key Fields

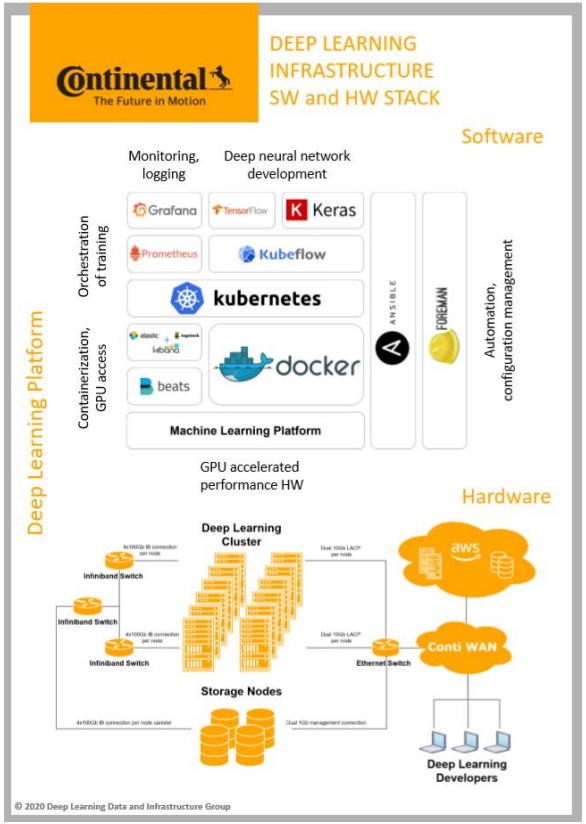
- Neuronal Networks
  - Training DNNs
- Simulation
  - Synthetic Data generation for Test&Validation
- Test&Validation
  - GPU based SIL/HIL/MIL



Project SuperComputer

- Project done in 7 months, complete build-up in 2 weeks
- Including design of the architecture , doing POCs , tenders, choosing partners, defining the software stack

- Specs:
  - Multi Node GPU Cluster: DGX1/DGX2
  - NonBlocking Infiniband: 400Gbit/s or 800 Gbit/s per node
  - ESS3000 Storage: >200GB/s Read
  - AI Ready Data Center



Continental
The Future in Motion

Senses for Safety

Driver assistance systems help save lives

AI Infrastructure as Code

Project Deep Thought – Continental Supercomputer – SC20 2020-11-16

Viktor Pal – Deep Learning Infrastructure Architect

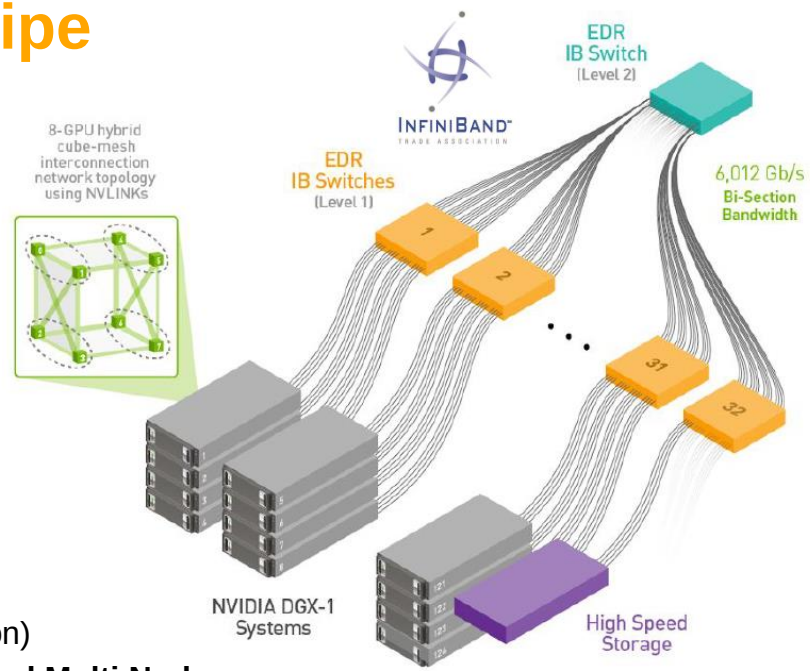
HPC + DL Infra simplified recipe

Base (HPC)

- Several computers (compute + service)
- Low-latency, high-bandwidth network
- High speed storage
- Scheduling

+ Flavoring (Deep Learning)

- GPUs
- NVLink (high-speed, direct GPU-to-GPU interconnect)
- GPUDirect RDMA (for cross node GPU-GPU communication)



Distributed Multi Node Trainings

The approach

- Use what you know and realize what you don’t → **don’t reinvent the wheel**
  - Partners with knowledge and experience → shorter delivery times & success
- Cooperate and gain knowledge from each other
  - Discover the available knowledge and organize cooperation inside the company
- Automate whatever and whenever possible
  - The only way to manage complex infrastructures on the long run
- Define clear and simple processes
  - Be agile and avoid the overhead of bad processes
    - Single README page in our Git repository

The concept

- Build the infrastructure from a single Git repository
- Use industry standard technologies
- Nvidia DeepOps as a reference: https://github.com/NVIDIA/deepops
- Use NGC's standardized, up-to-date, tested images
- Use open source technologies where possible
- Design and build for high availability and security from the ground up
- Have an as complete as possible test infrastructure

CLOUD NATIVE
COMPUTING FOUNDATION
Incubation
Graduation

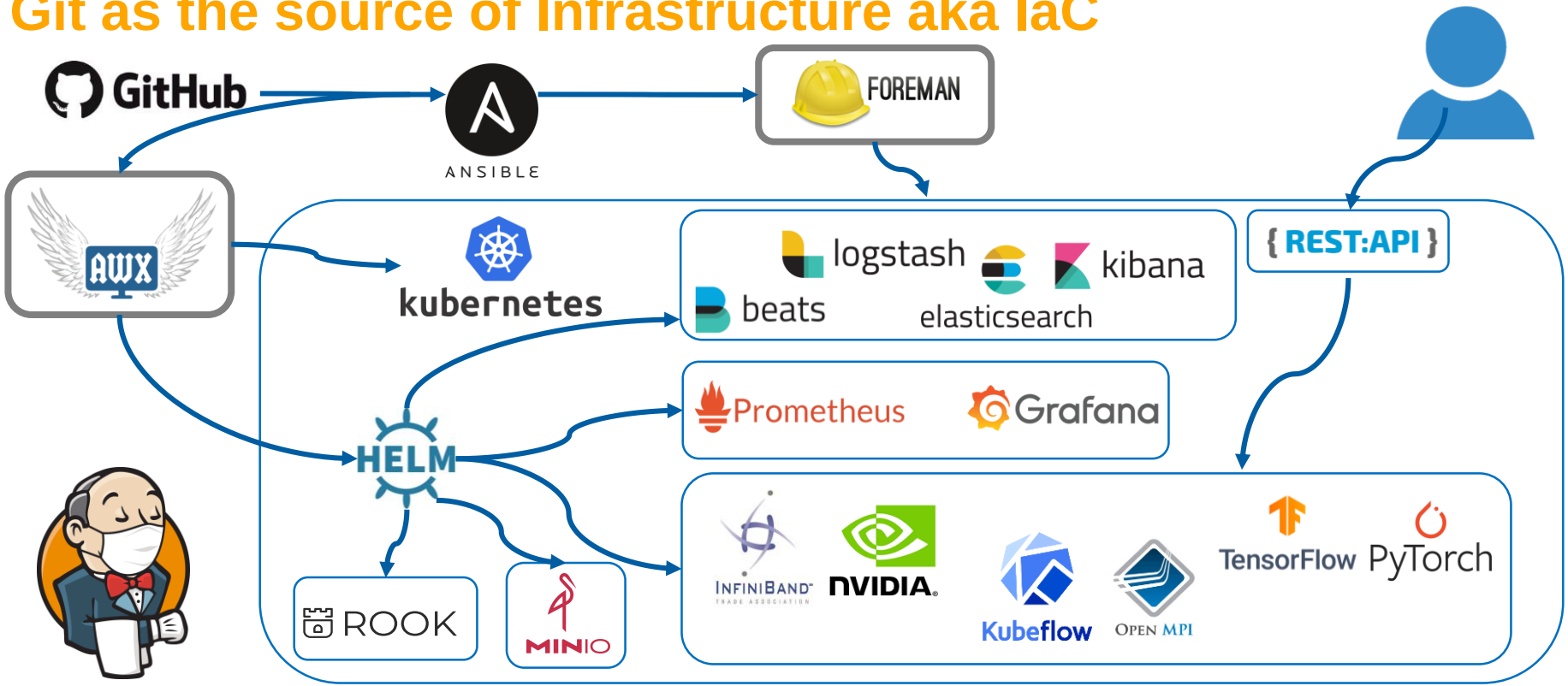
The journey

- A lot of PoCs and reading → Storage: IBM and others helped us with onsite environments
- Hard decisions → Good options on the market
  - InfiniBand vs Converged Ethernet (RoCE)

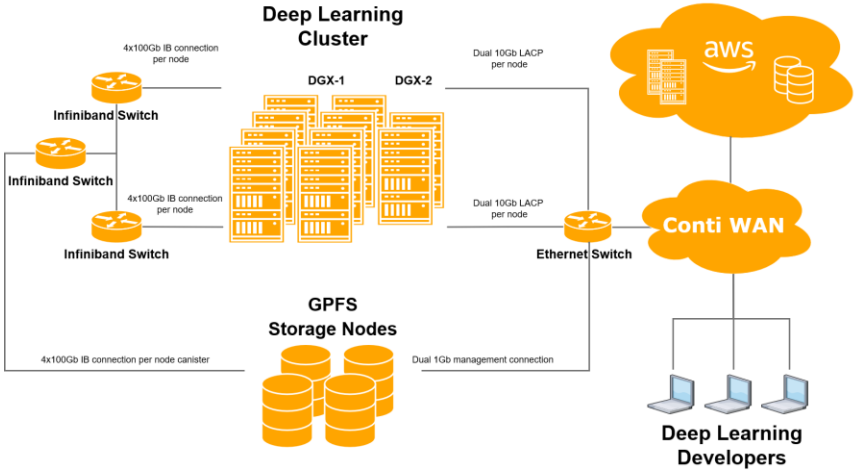
- Consulting → Strong partners like SVA helped to build InfiniBand and storage quickly
- Cloud is easier, bare metal is harder → Foreman

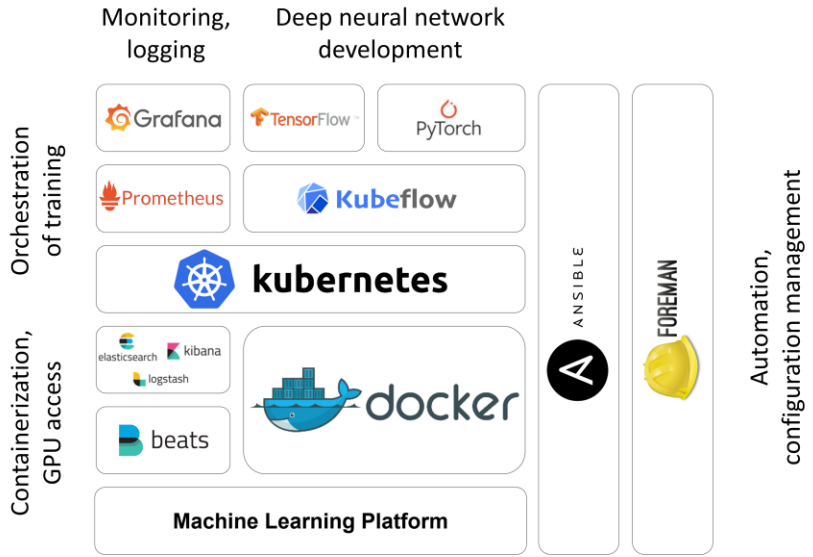
- Data center for AI is somewhat special: cooling, power consumption
- Nvidia whitepapers and certified solutions helped a lot
  - For example IBM ESS3000
- Migration of data from legacy storage solution to high performance storage

Git as the source of Infrastructure aka IaC



Hardware and software stack





Why these tools and components?

- **Foreman**
  - Enables building bare metal environments automatically in a standardized way
- **AWX**
  - Tracking: see what is running, logs and jobs preserved, prevent parallel runs where needed
  - Preconfigured runs and unified runtime environment
- **Jenkins**
  - Easily integrate deployment, build and testing of additional tooling
- **Kubernetes**
  - State of the art for scheduling: originally build for stateless services but currently the way to go almost in every area
- **Helm**
  - Standard IaC/Package Manager for Kubernetes, easy upgrades and deployments

How does GPFS fit into this

- **The bad**
  - Complex initial setup of ESS3000
  - Upgrades were hard at the beginning → new product, getting better
  - Hard to automate with our existing toolchain
  - Support can be cumbersome at times
- **The good**
  - Stability
  - Easy to extend
  - Good performance
  - Very nice integrated monitoring system

Storage architecture

- **Data ingest**
  - MinIO (vs CES) → good access control, available from everywhere
    - Better control over data structure
  - Standard protocol, many libraries (S3)
    - Freedom to integrate new features with updates
- **Data access**
  - Inside the cluster: native GPFS over InfiniBand
  - Everywhere else: MinIO and CES (NFS/SMB: easy setup)
    - Workspace access, cluster backups

What we are working on

- Data
  - Move infrequently accessed data to cheap storage (good GPFS support)
  - Improve disaster recovery
- Workflow
  - Scaling out transparently
  - Better scheduling: utilize resources more efficiently
  - Keep up with the world: UPDATE, UPGRADE regularly → new features and stability
  - Better and easier AI workflow → minimize developer overhead
  - Refined monitoring is a constant work: better availability, easier operations
  - Jenkins and AWX: fully automated Jenkins and AWX deployment
  - More automated hardware management: FW management, Out of Band management

Thank you

Viktor Pal
Senior Deep Learning Infrastructure Architect

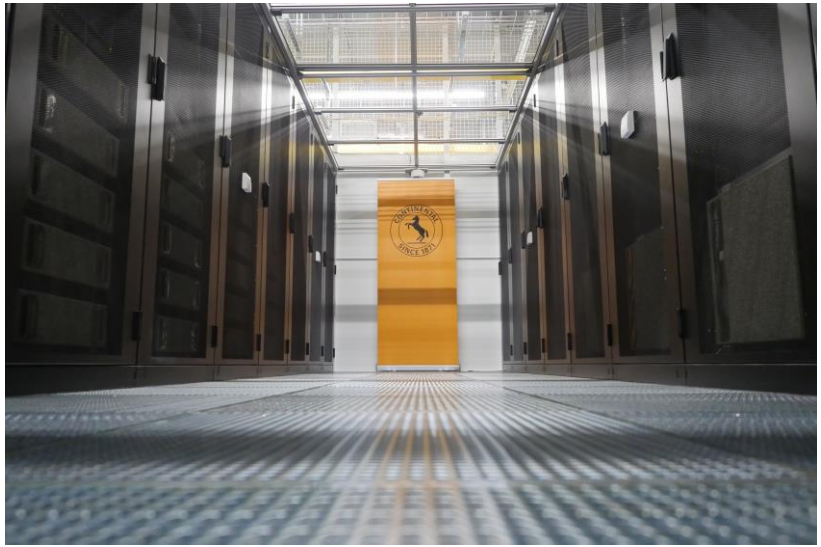
BU Advanced Driver Assistance Systems

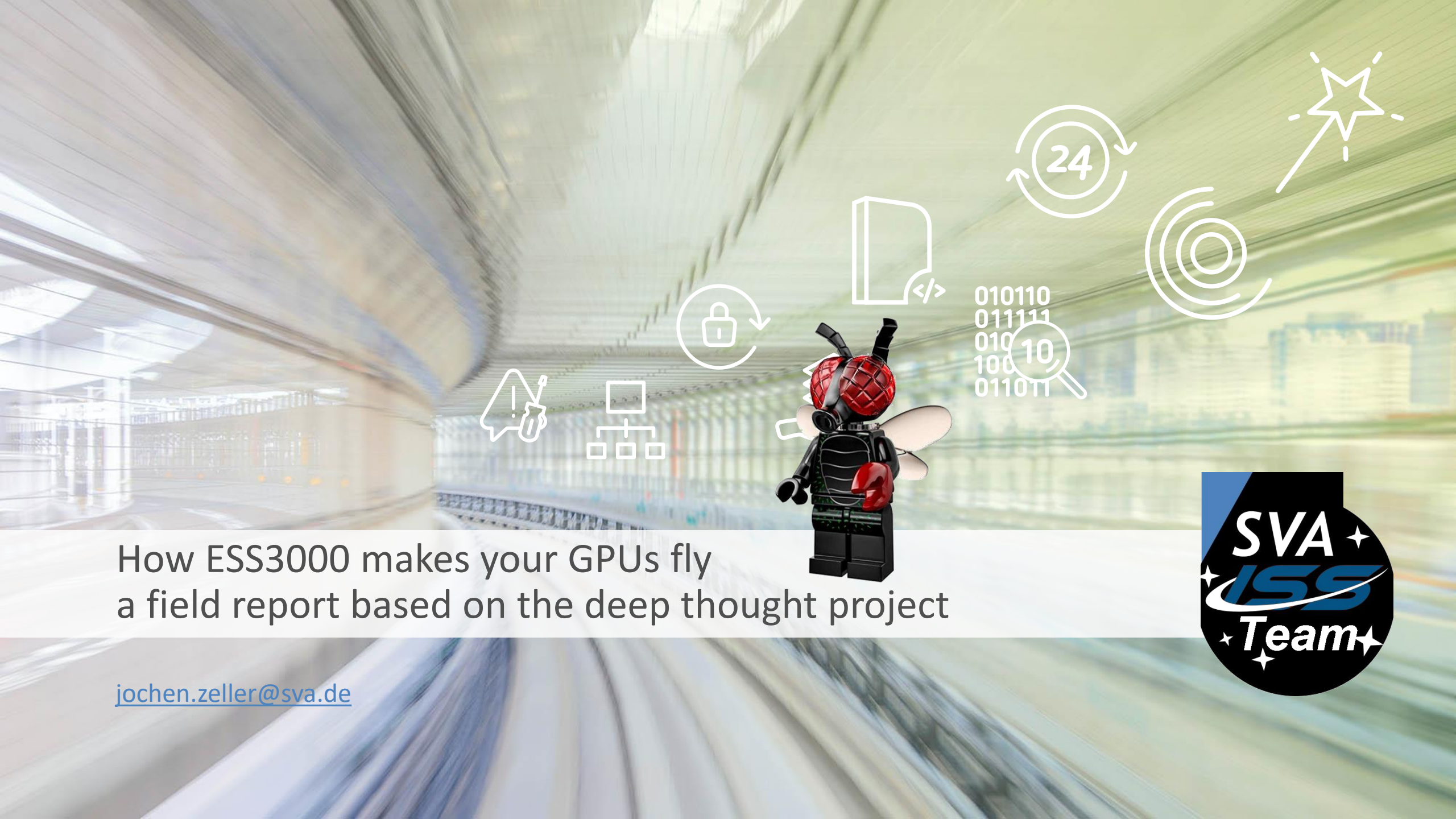
Continental
Autonomous Mobility and Safety (AMS)

David Enenkel
Head of IT Operations

BU Advanced Driver Assistance Systems

Continental
Autonomous Mobility and Safety (AMS)





How ESS3000 makes your GPUs fly
a field report based on the deep thought project

jochen.zeller@sva.de

- The IT system integrator in Germany, my employer

Biggest owner-operated system integrator in Germany

Steady growth with more than 1.600 employees in Germany.

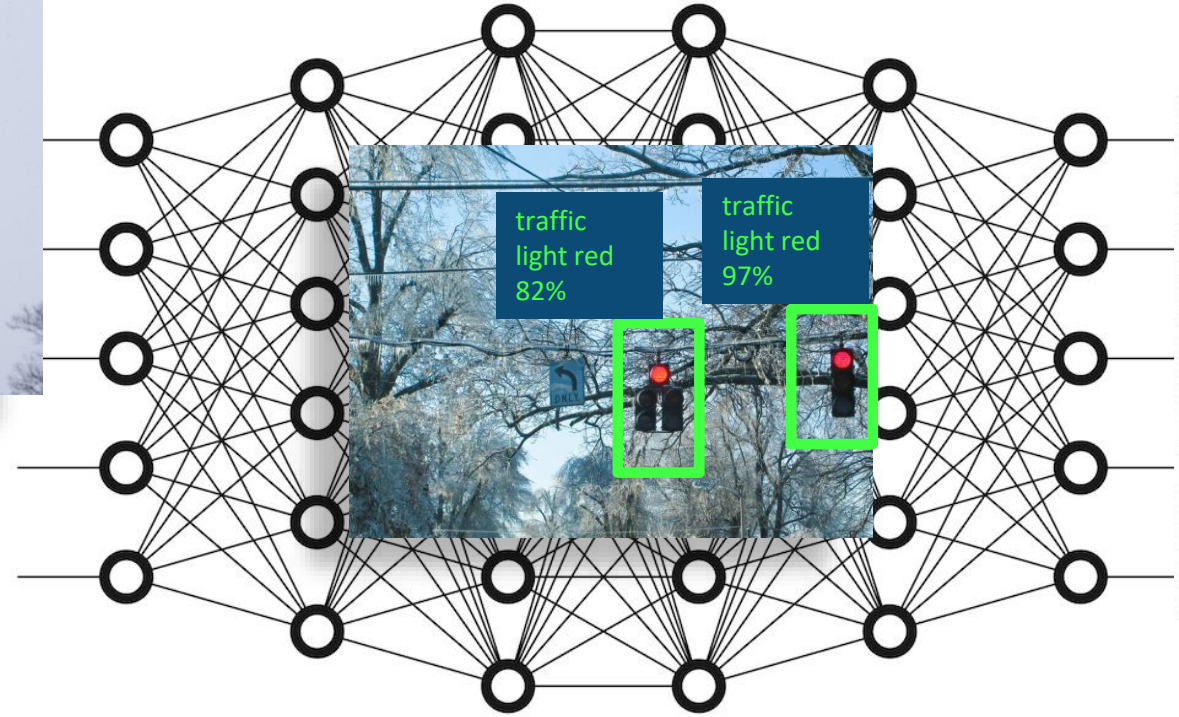
More than **960 technical experts** provide advice & integrate your individual customized solution.

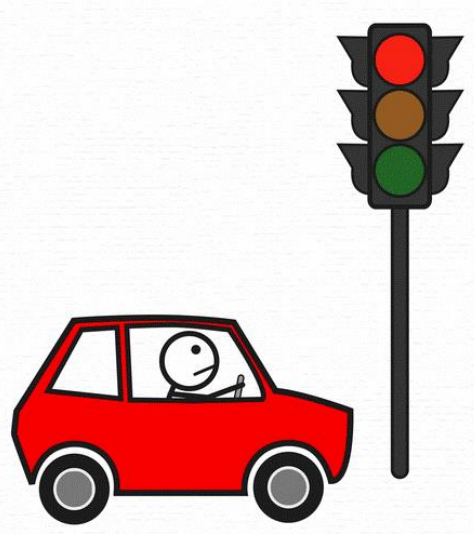
Company objectives

We accompany our customers on a long-term basis throughout the project lifecycle.

/ DEEP THOUGHT – AN AI PROJECT

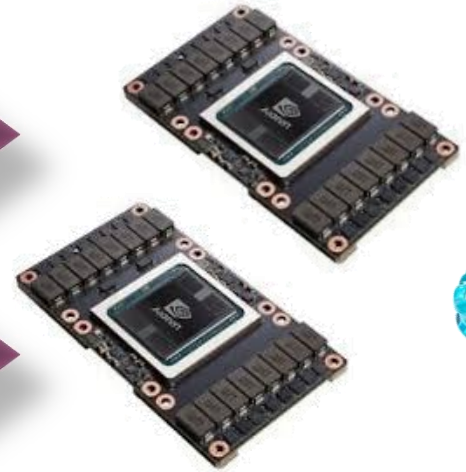
Neuronal network to detect traffic related situations





/ WHY DO THEY NEED SPECTRUM SCALE AND ESS3000

Many GPUs in a lot of DGX‘s need access to a pool of millions of images. In parallel and very fast, without wasting GPU time when reading the images into the GPUs memory.



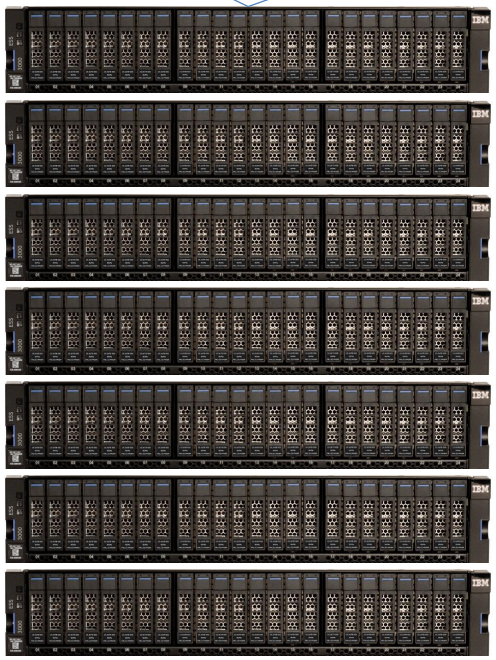
1 NVIDIA A100 GPU has 40Gbyte GPU Memory
→
Approx. 80.000 images@500k
→
8 minutes @ 6ms latency (single)
16 secs @ 0.2 ms latency (single)

/ STORAGE AND NETWORK INFRASTRUCTURE

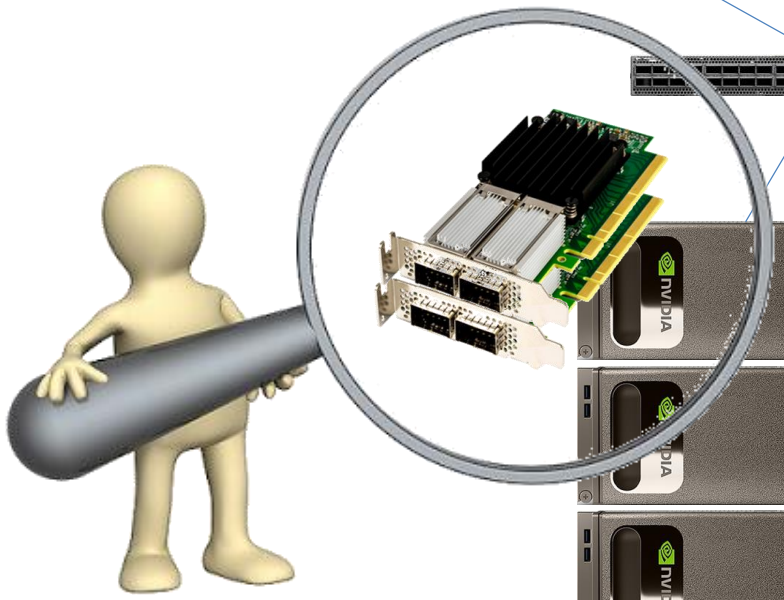
200Gbit HDR switches,
200Gbit spine-to-leaf links,
100Gbit client links
to increase number of client ports

Infiniband HDR

Infiniband EDR



7 x ESS3000



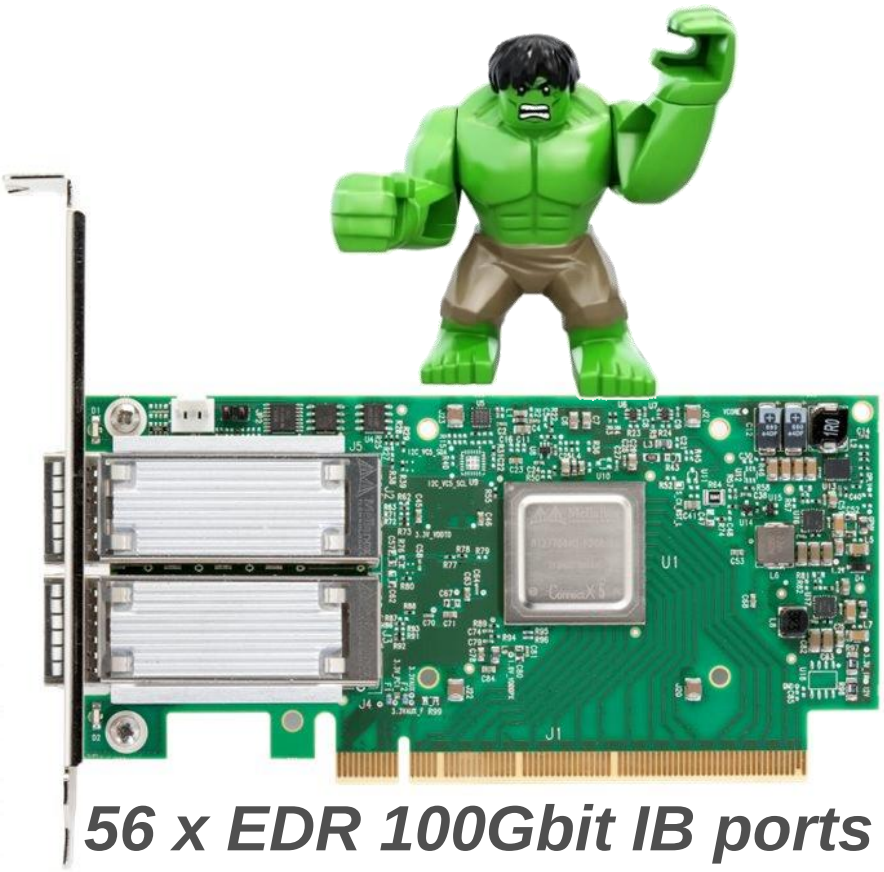
4 x EDR 100GB per DGX-1
8 x EDR 100GB per DGX-2

Infiniband EDR

many DGX-1 and DGX2

/ ESS3000 SPECS

- 7 x ESS3k / each with 2 IO nodes
  Per ESS3k:
    24 x NVMe
  Per IO node (node canister):
    2 x 14 core Intel x86_64
    768GB Memory
    RHEL8
    2 x 2-port EDR 100GBit Infiniband



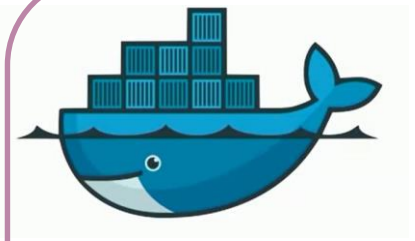
56 x EDR 100Gbit IB ports

/ SPECTRUM SCALE BASED ECO SYSTEM

Spectrum Scale immutability (WORM) protect image files.
to
Based on policy, of course!

MINIO

S3 / Object data transfer



Kubernetes managed container environment to run AI stack on Spectrum Scale

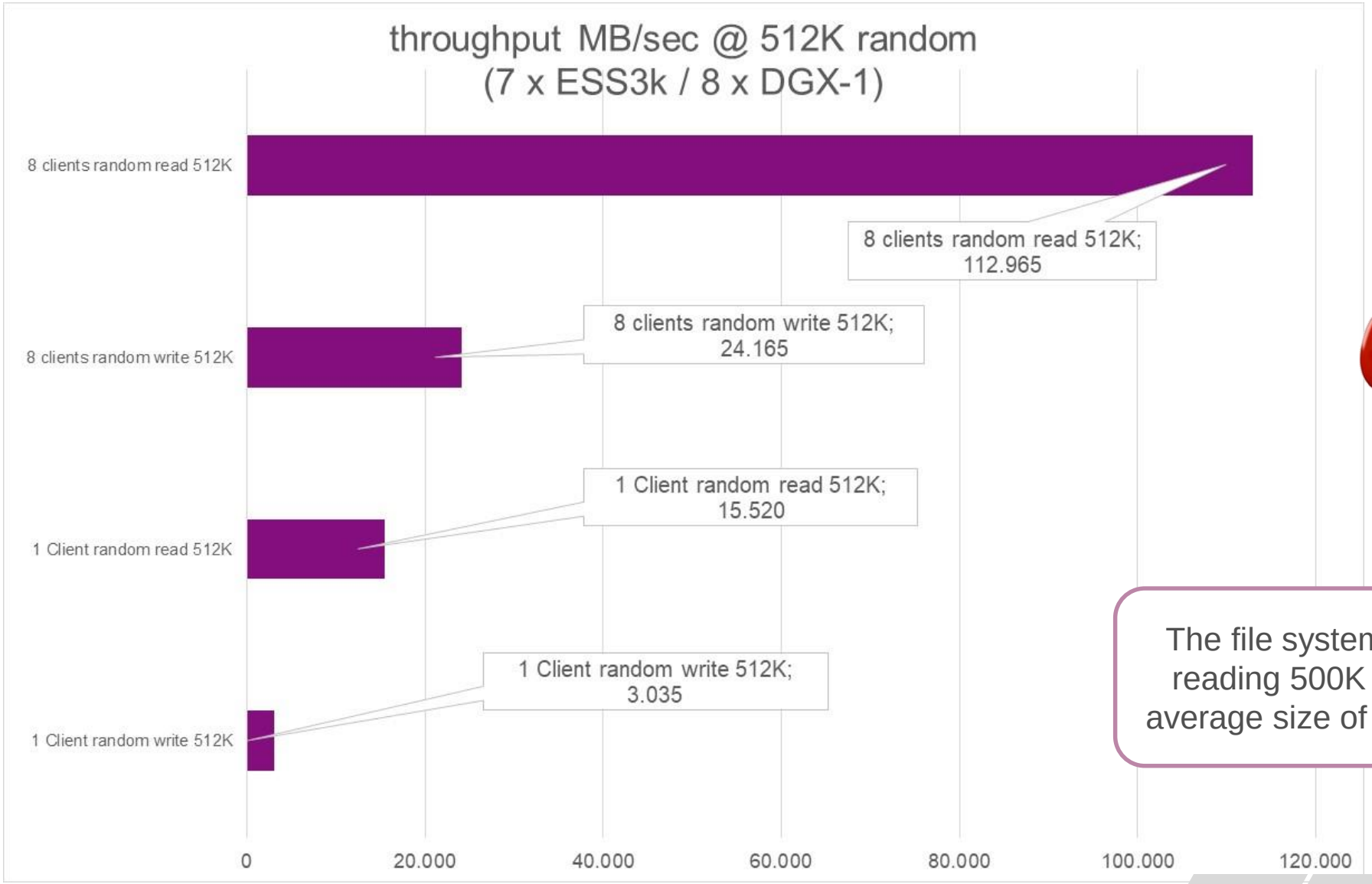
Workstation access by CES SMB

TCT based file copy to AWS

/ DEEP THOUGHT ESS3000 SETUP

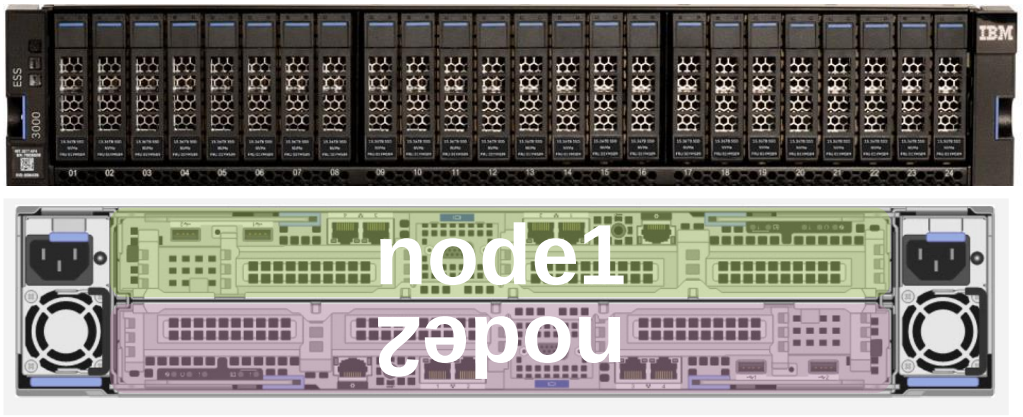
- One recovery group per ESS3k chassis (new shared RG design)
- four vdisks per chassis, 8D+2P
- two vdisks hosted per node canister
- 28 vdisks, all data and metadata in system pool
- One file system, 4MB block size

/ 8 DGX-1 RUMBLES ON THE FILE SYSTEM – 512K IO SIZE



The file system is optimized for reading 500K files – this is the average size of the traffic images.

/ ESS3000 – ONE IMPROVEMENT TO HIGHLIGHT

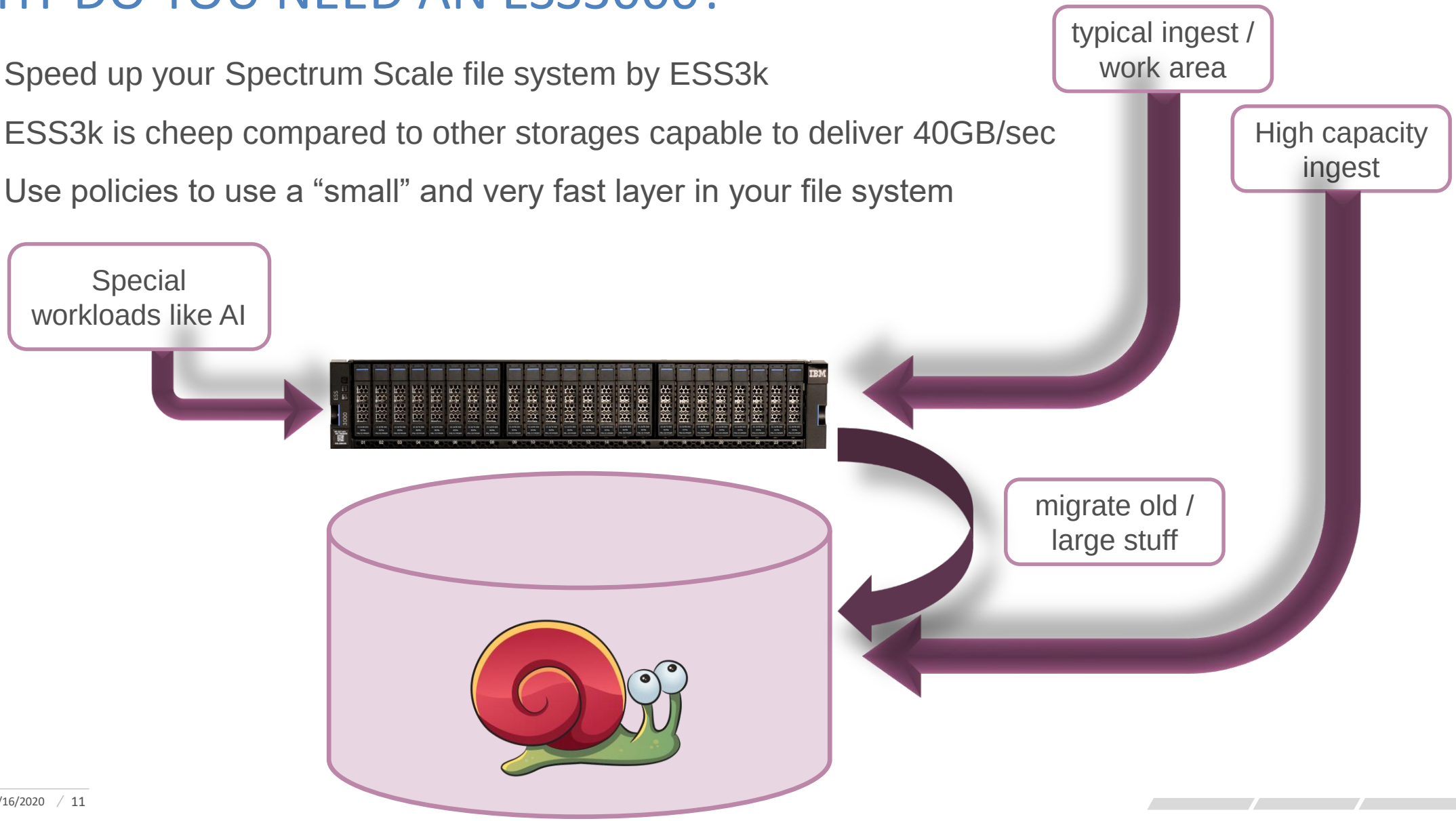


- Quiz: how many commands do you need to update 7 ESS3k?
  - Scale code
  - Native raid code
  - RHEL
  - ofed
  - Firmware (x86_64, adapter, NVMEs)

Just one: # ess3krun -G *ess_x86_64* **update**

/ WHY DO YOU NEED AN ESS3000?

- Speed up your Spectrum Scale file system by ESS3k
- ESS3k is cheep compared to other storages capable to deliver 40GB/sec
- Use policies to use a “small” and very fast layer in your file system



/ WHY DO YOU NEED AN ESS3(

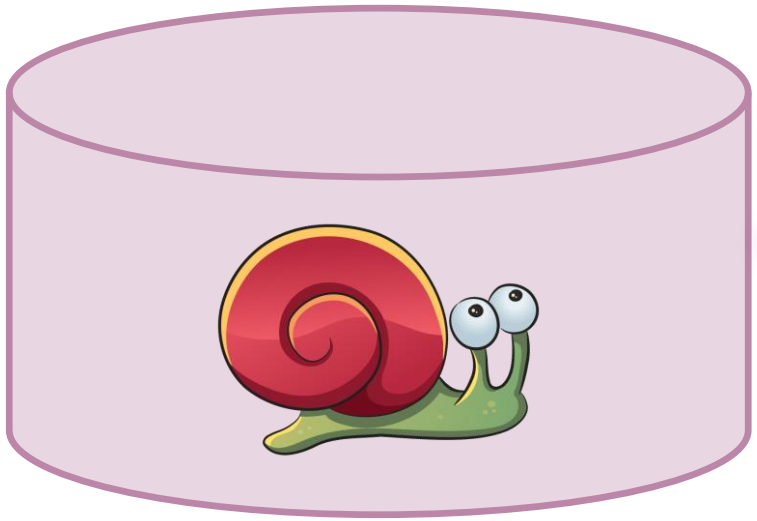
- Speed up your Spectrum Scale file system
- ESS3k is cheep compared to other storage
- Use policies to use a “small” and very fast I

Your homework:

Ask your IBM sales representative for the price of an ESS3000, 24 x 3,84 TB NVMs

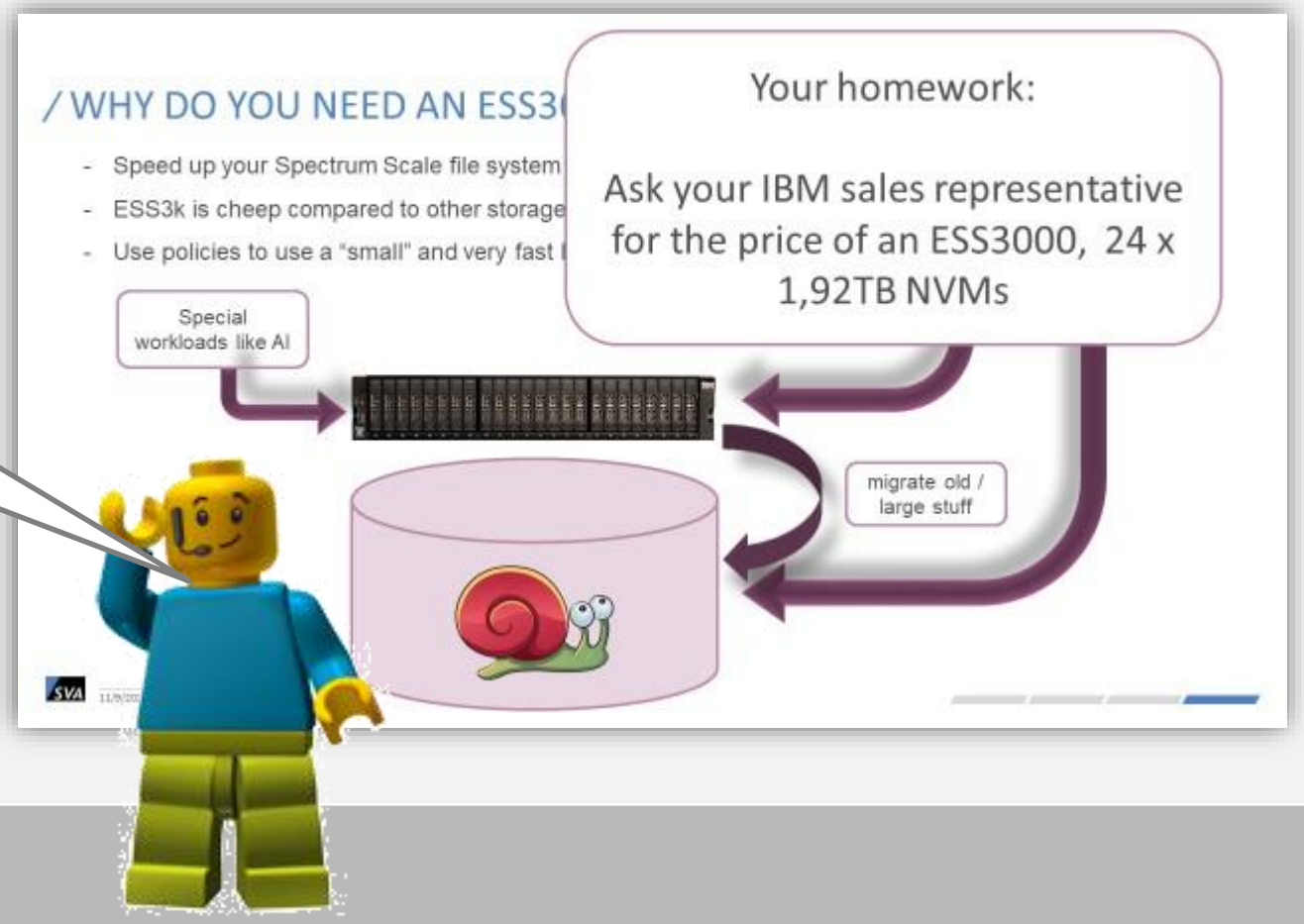
Special workloads like AI

migrate old / large stuff



/ THE END!

Many thanks for your attention!

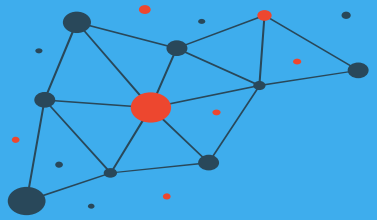


/ CONTACT

JOCHEN ZELLER

IT Architect

jochen.zeller@sva.de



Check https://www.spectrumscaleug.org/experttalks for charts, show notes and upcoming talks

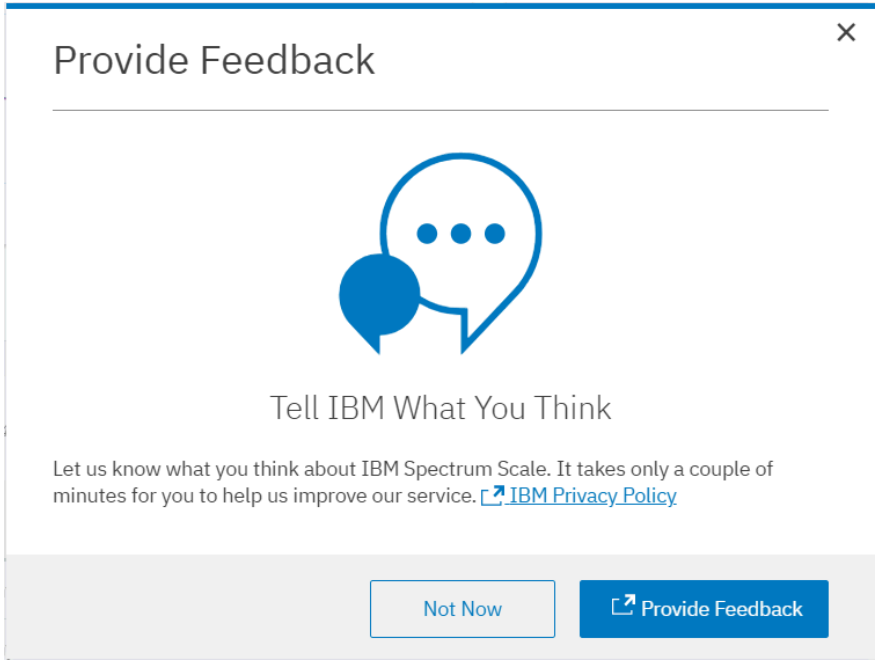
- Past talks:
  - 001: What is new in Spectrum Scale 5.0.5?
  - 002: Best practices for building a stretched cluster
  - 003: Strategy update
  - 004: Update on performance enhancements in Spectrum Scale (file create, MMAP, direct IO, ESS 5000)
  - 005: Update on functional enhancements in Spectrum Scale (inode management, vCPU scaling, NUMA considerations)
  - 006: Persistent Storage for Kubernetes and OpenShift environments
  - 007: Manage the lifecycle of your files using the policy engine
  - 008: Multi-node scaling of AI workloads using Nvidia DGX, OpenShift and Spectrum Scale
- Today:
  - Nov 16: Continental: Deep Thought – An AI Project for Autonomous Driving Development
  - Nov 16: Data Accelerator for Analytics and AI (DAAA)
- Next:
  - Nov 18: User Meeting at SC20 (Session 2) – What is new in Spectrum Scale 5.1? https://www.spectrumscaleug.org/event/sc20-meeting-session-2-what-is-new-in-spectrum-scale-5-1/

Thank you!

IBM Spectrum Scale

Please help us to improve Spectrum Scale with your feedback

- If you get a survey in email or a popup from the GUI, please respond
- We read every single reply



Spectrum Scale User Group

The Spectrum Scale (GPFS) User Group is free to join and open to all using, interested in using or integrating IBM Spectrum Scale.

The format of the group is as a web community with events held during the year, hosted by our members or by IBM.

See our web page for upcoming events and presentations of past events. Join our conversation via mail and Slack.

www.spectrumscaleug.org