Milyen kihívásai vannak egy világszínvonalú MI klaszter létrehozásának?

Continental Mesterséges Intelligencia Fejlesztési Központ @ Budapest

Nagy Enikő
Deep Learning Infrastructure Engineer/Scrum Master
AI Infrastructure Team

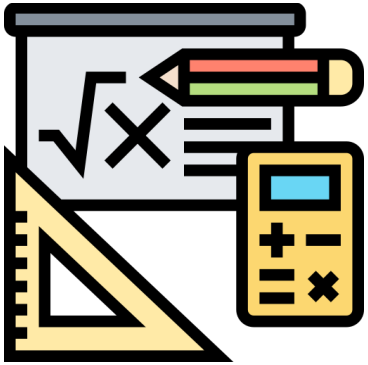
Autonomous Mobility

BA AM

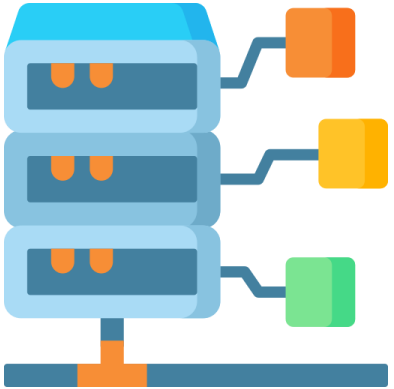
What is behind a seemingly intelligent system?

What is behind a seemingly intelligent system?

Huge database



Mathematical model



Huge computing capacity

AI CLUSTER

AI CLUSTER

- In **Frankfurt am Main, Germany**
- Strong cooperation with Lindau IT and consulting companies
- From 2020 Q1 offers **computing power** as well as **huge storage capacity** to developers in locations worldwide
- **Reduces development time** from weeks to hours
- Helps to produce **better quality product**
- Currently **~30 projects worldwide** (object detection, semantic segmentation etc.)

Why is it „AI” Cluster?

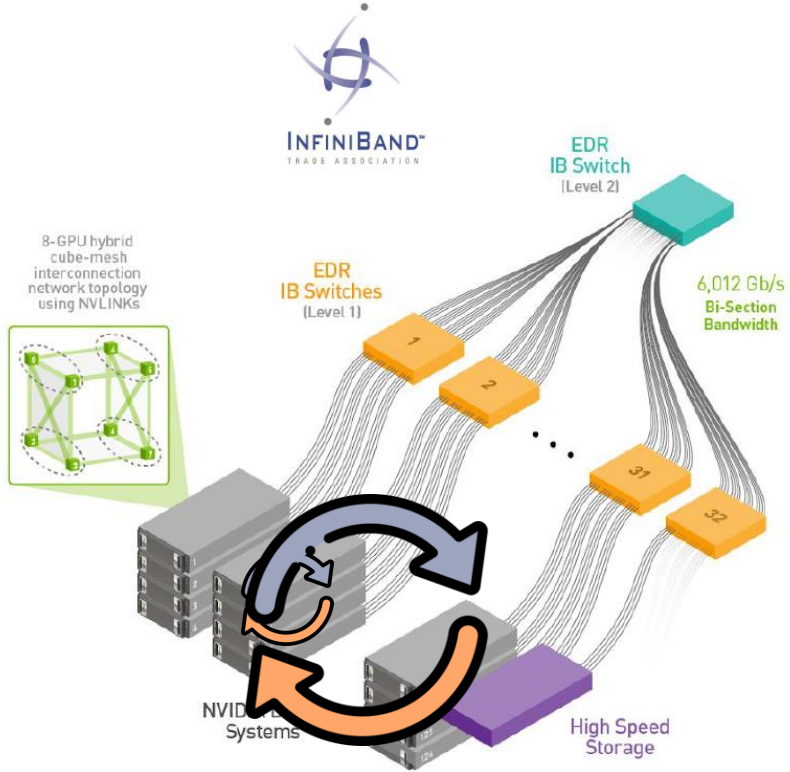
Main components:

- Several computers
- Low-latency, high-bandwidth network
- High speed storage
- Scheduling

Distributed Multi Node Trainings

Escpecially for AI:

- GPUs
- NVLink
- GPUDirect RDMA



Source: https://www.nvidia.com

AI Cluster in numbers

PEAK CONSUMPTION is between **90 and 100 kWh** of the whole GPU cluster

53 NVIDIA DGX SERVER

7 TONS OF HARDWARE have been racked

448 NVIDIA TESLA V100 GPU

OVER 600 CABLES with a total length of **3 kilometers** have been labelled, connected and installed

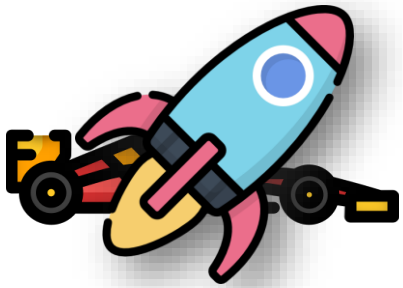
> 100 Gbps network

PROJECT TIMEFRAME: 7 MONTHS in total, assembly within 2 weeks

What is the power of the AI cluster?

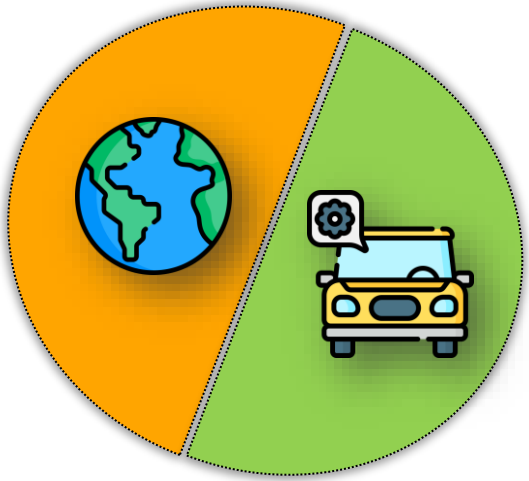
Floating **P**oint **O**perations **P**er **S**econd: number of additions or multiplications / second.

56 PFLOP = **56 thousand trillion calculations/second**.



What does that power mean?

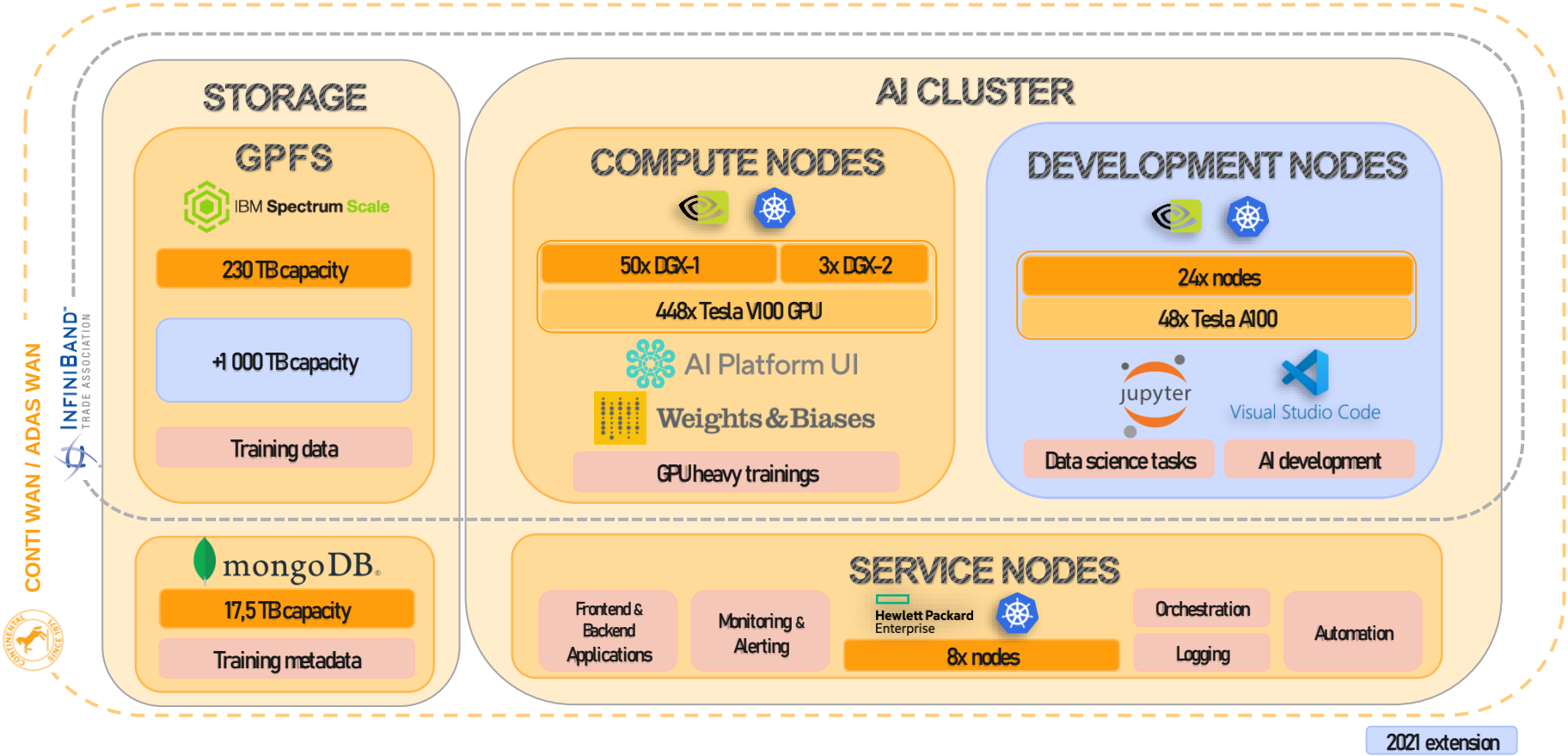
In 2020
the cluster was in the
200 largest
computers
in the world.



One of the largest
Automotive
HPC cluster
in Europe.

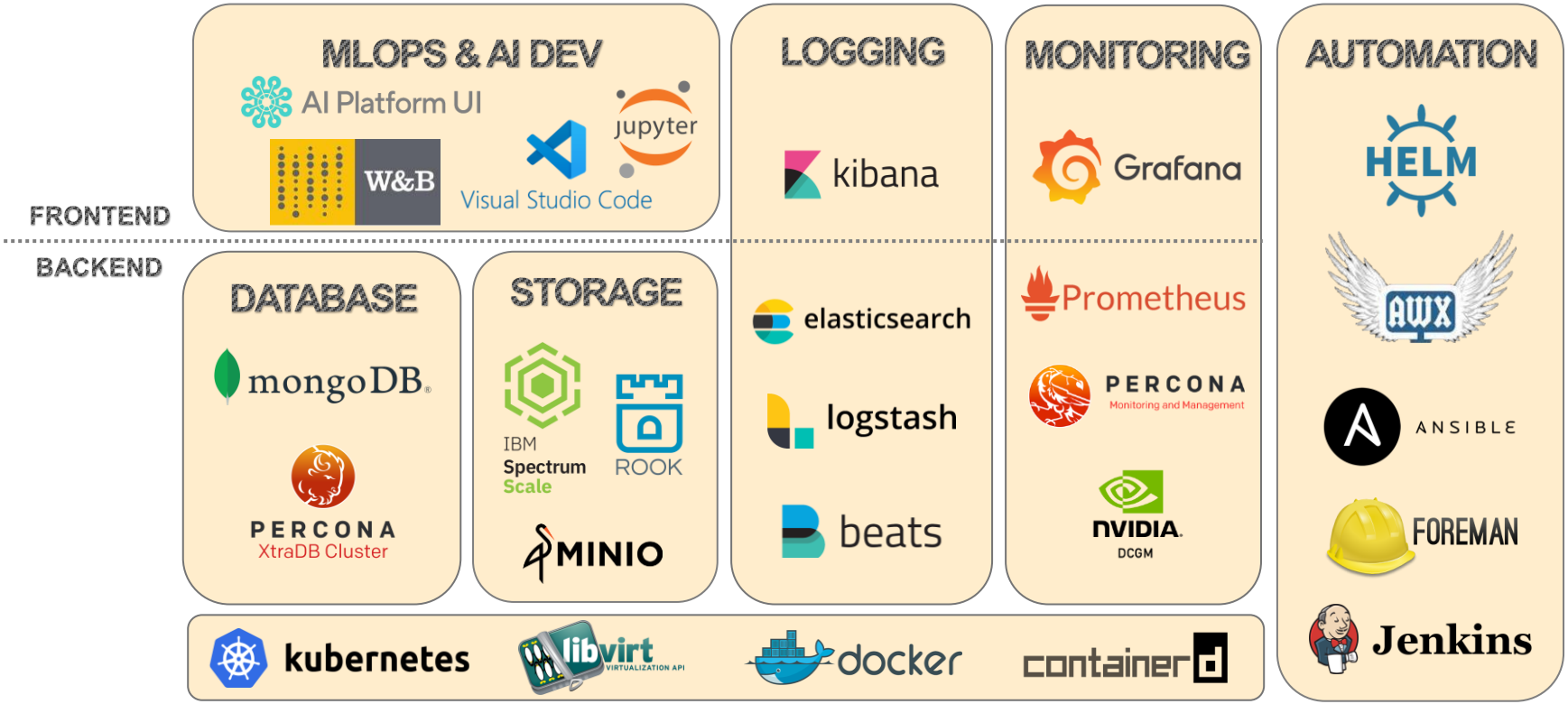

KEY STACK ELEMENTS OF A GOOD INFRA
disaster recovery
performance
logging
monitoring
alerting
traceability
reproducibility
automation
LOG

AI CLUSTER – HW STACK



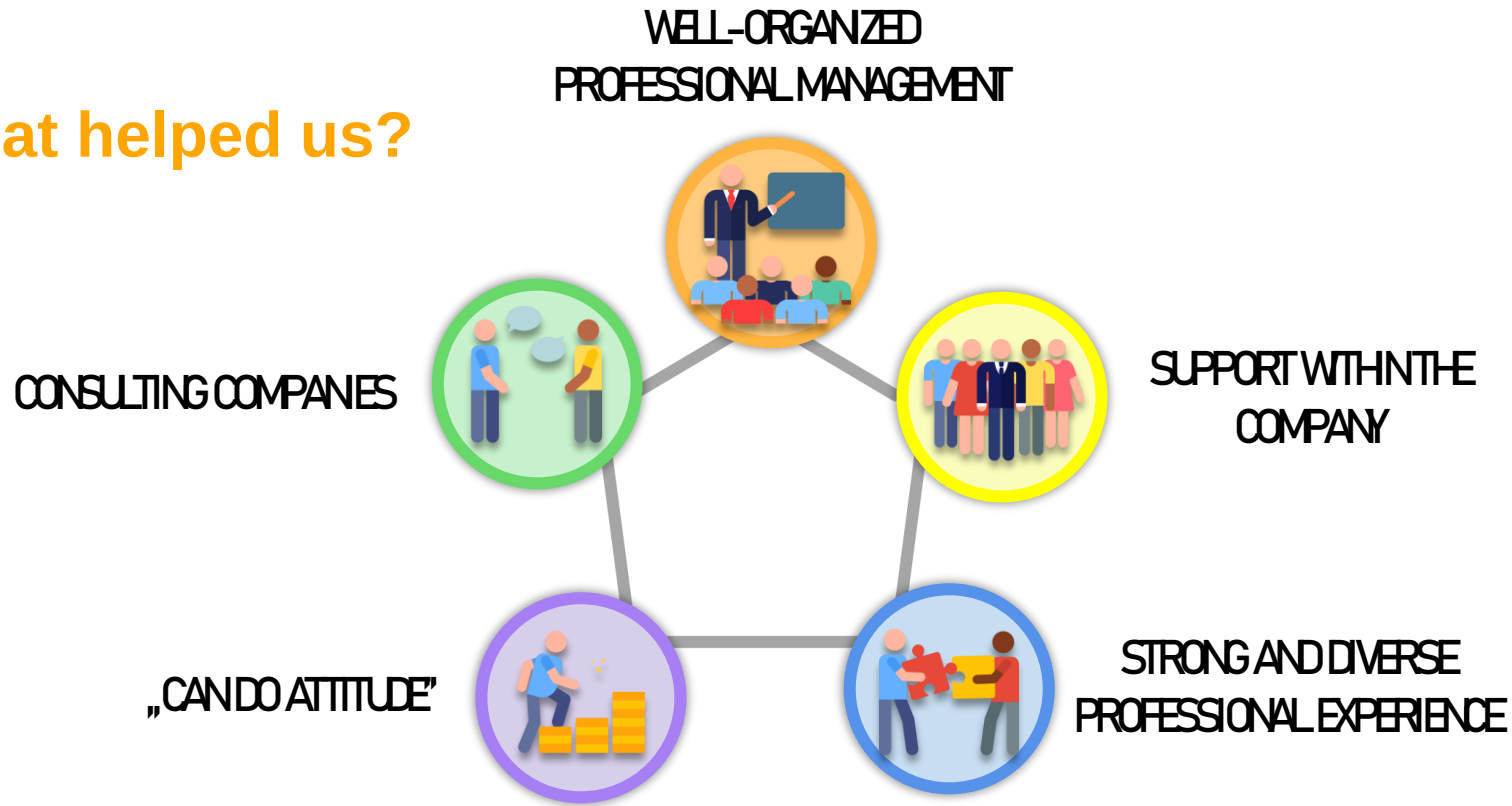
Enikő Nagy, © AI Development Center Budapest

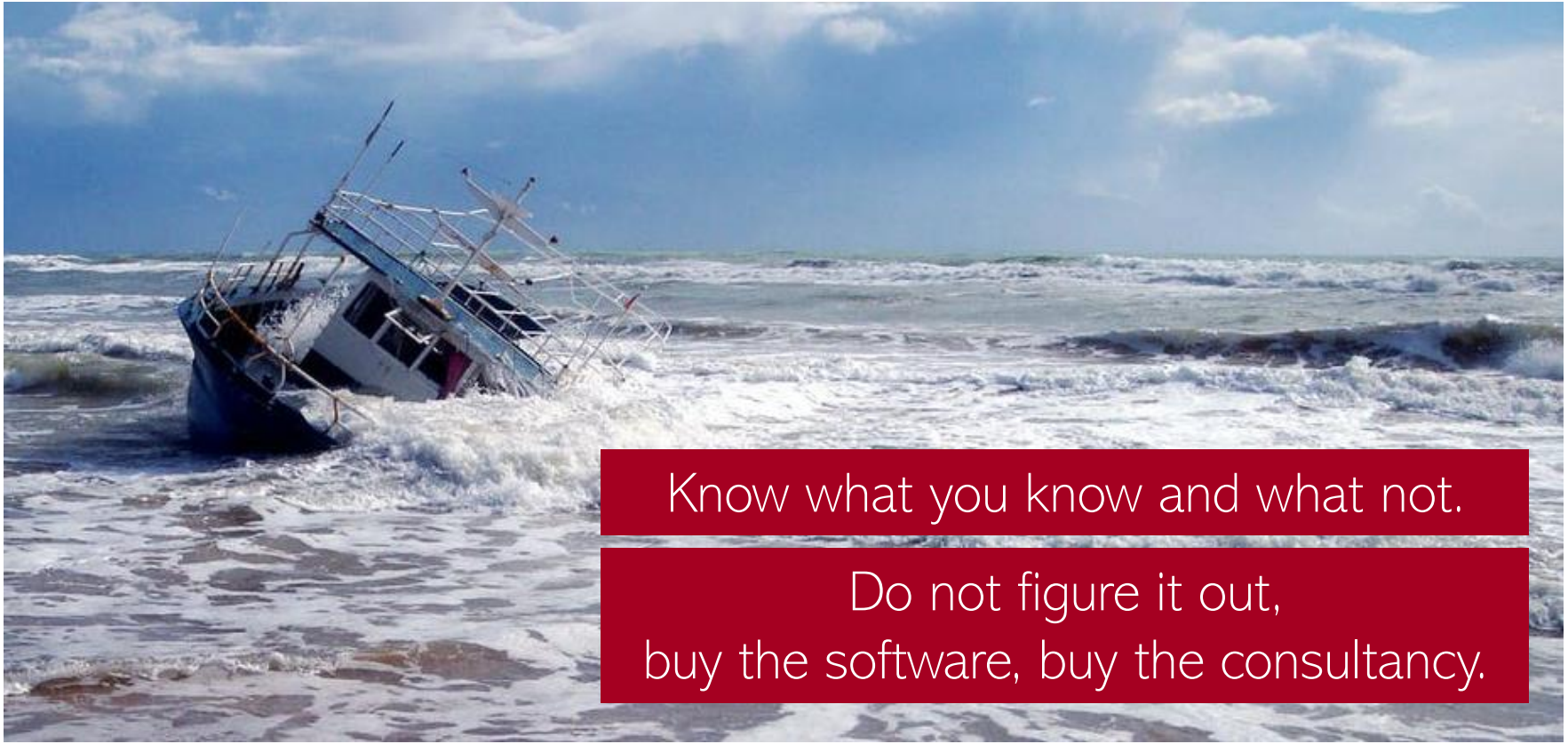
AI CLUSTER – SW STACK


FRONTEND
BACKEND
MLOPS & AI DEV
AI Platform UI
W&B
Visual Studio Code
jupyter
LOGGING
kibana
MONITORING
Grafana
AUTOMATION
HELM
AWX
DATABASE
mongoDB.
PERCONA
XtraDB Cluster
STORAGE
IBM
Spectrum
Scale
ROOK
MINIO
elasticsearch
logstash
beats
Prometheus
PERCONA
Monitoring and Management
NVIDIA
DCGM
ANSIBLE
FOREMAN
Jenkins
kubernetes
libvirt
VIRTUALIZATION API
docker
containerd

CHALLENGES

What helped us?




Know what you know and what not.
Do not figure it out,
buy the software, buy the consultancy.

Lessons learnt – in technical aspect

CHALLENGES

- Translate requirements
- Steep learning curve
- Slow response loop (feedbacks)
- Technological change is expensive

SOLUTION

- Knowledge share
- User education
- Creating key user group
- Spread DevOps mindset


If we fail, fail quickly.

Thank you
for your attention!

QUESTIONS?

Enikő Nagy

eniko.nagy@continental-corporation.com