Hewlett Packard Enterprise

RMACC
Rocky Mountain Advanced Computing Consortium

SWARM LEARNING EXPLAINED

SOLVING THE DATA SHARING CONUNDRUM IN AI

Steve Heibein
US Public Sector AI Lead
stephen.Heibein@hpe.com

August 3, 2022

REAL WORLD APPLICATION WITH REAL WORLD DATA IN HEALTH CARE

- HPE Swarm Learning will help solve truly complex disease classification problems with ease and accuracy.
- Paper developed jointly with DZNE highlighting key outcomes from disease classification use cases. Nature Magazine cover page. (**May 26,2021**)

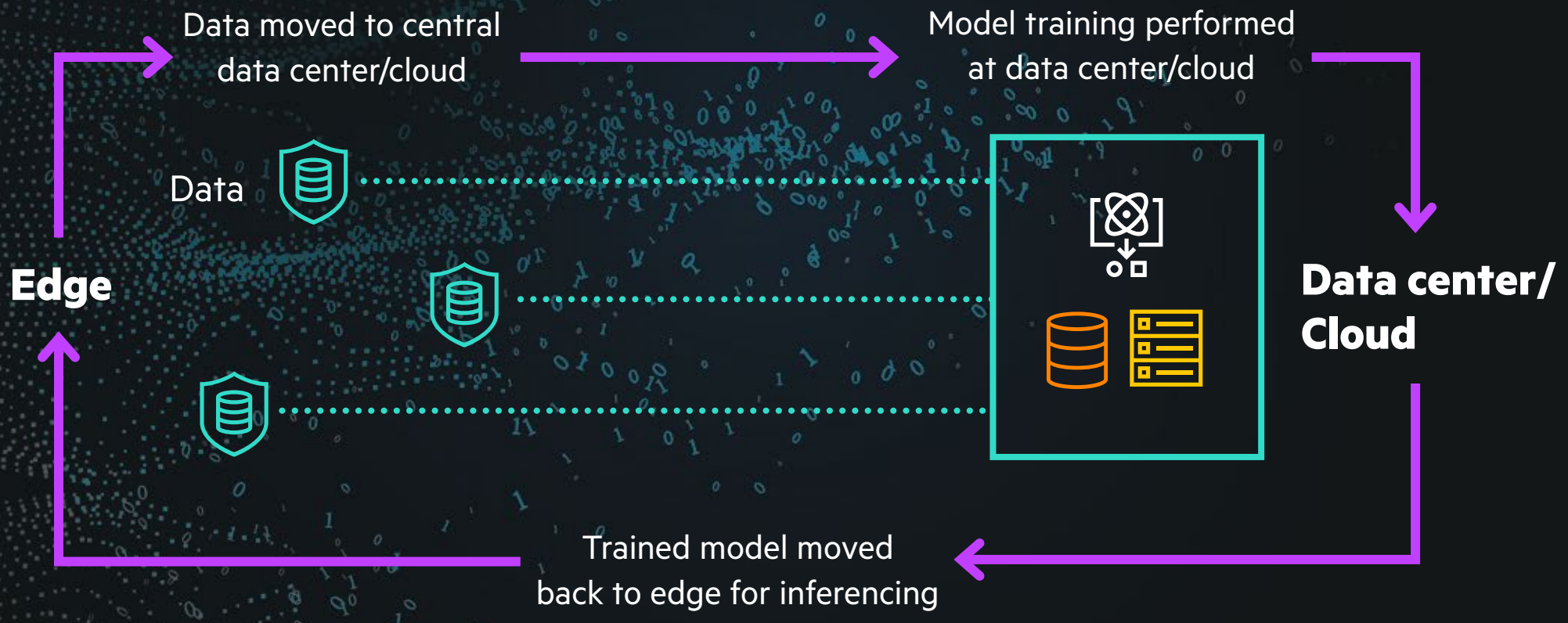
Swarm Learning for decentralized and confidential clinical machine learning | Nature

Article citation: Warnat-Herresthal, S., Schultze, H., Shastry, K.L. *et al.* Swarm Learning for decentralized and confidential clinical machine learning. *Nature* **594,** 265–270 (2021). https://doi.org/10.1038/s41586-021-03583-3

MACHINE LEARNING—TODAY'S APPROACH

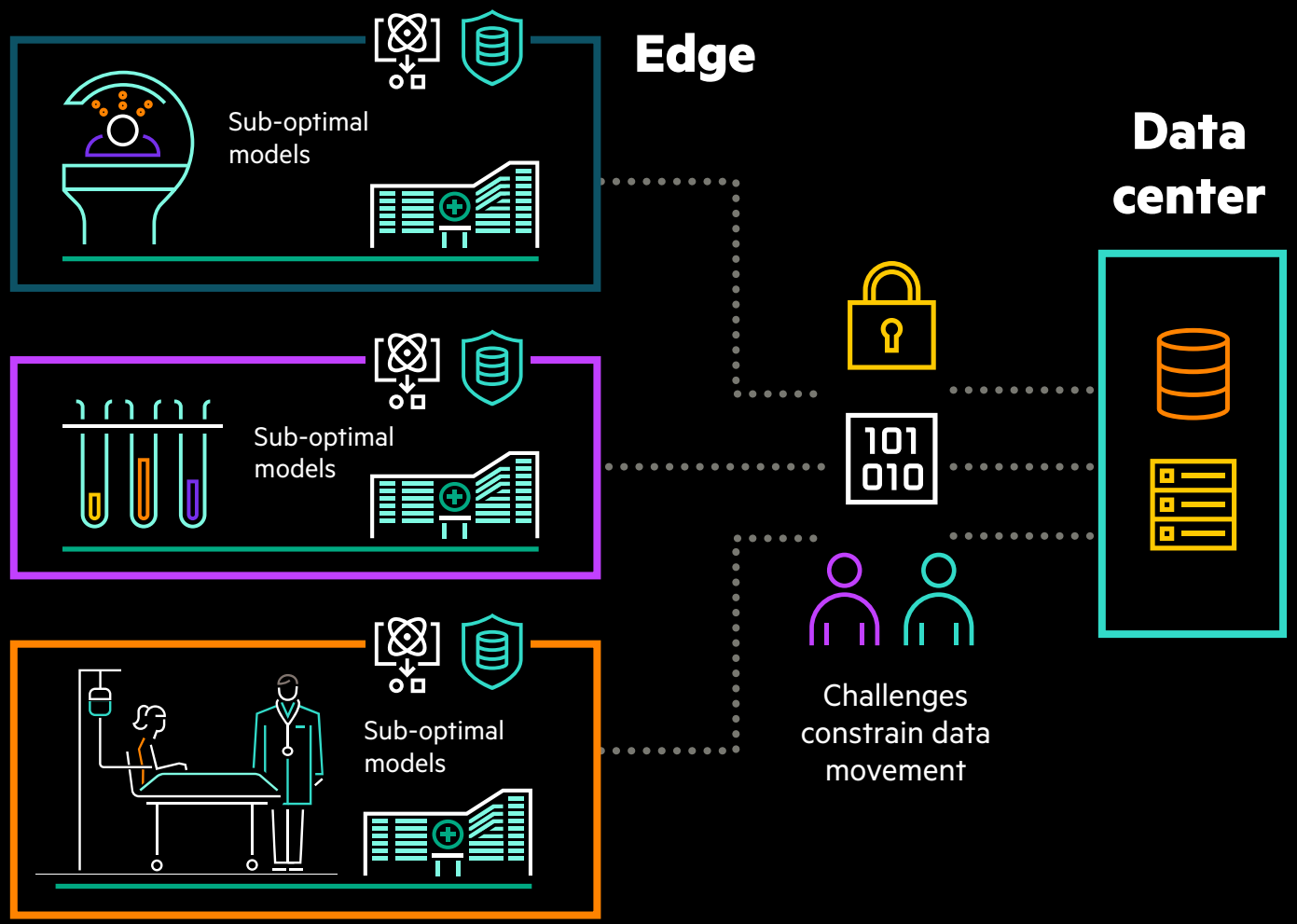
Centralized learning

Data is everywhere, but model training is at data center



BEFORE: CENTRALIZED MACHINE LEARNING IN HEALTHCARE TODAY

Challenges in data aggregation leading to sub-optimal models



USE CASE

Disease Detection using medical image data spread across a network of hospitals[1,2]

- **Data privacy constraints**
  Each patient/clinic can be an individual data source and not share “My data”
- **Data ownership constraints**
  Hospitals are not able to share data — due to regulations or policies
- **Operational inefficiency**
  Medical images are large and duplicating data is inefficient

RESULT

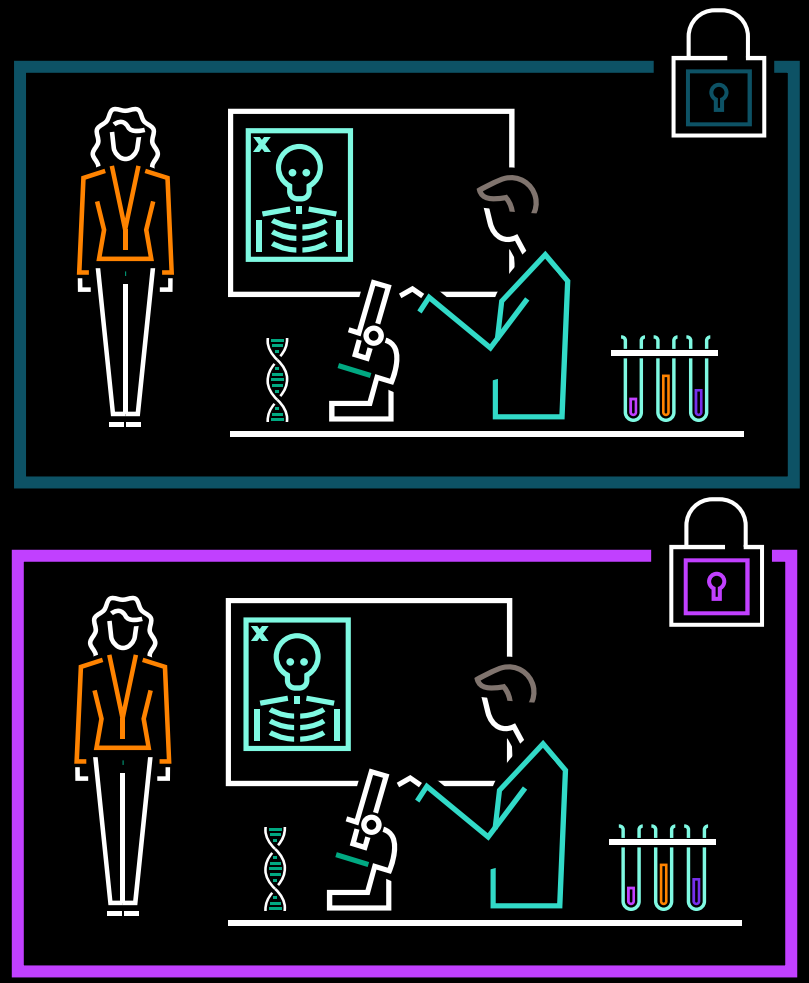
- Data is unable to be centralized, leading individual hospitals to train on local data
- Sub-optimal models with bias are created thus leading to inaccurate disease classification

1 Source: https://www.nature.com/articles/s41586-021-03583-?msclkid=e7e5e80ca63311ecabc86d8d860b2a57
2 Source: https://www.biorxiv.org/content/10.1101/2021.11.19.469139v1

DISEASE DIAGNOSIS EXAMPLE

Model development has several challenges

- Each patient is in principle an individual data source
- Both patient and hospitals are not willing to share data
  - “My data”
  - “My patient”
- Health data is highly sensitive and tightly regulated.
- GDPR (EU), HIPAA (USA), German **data** protection laws
- Biases in local data influence model bias
  - “demographic bias”
  - “technical imbalances”



Same challenges exist in medicine, but clinicians learn and support each other –

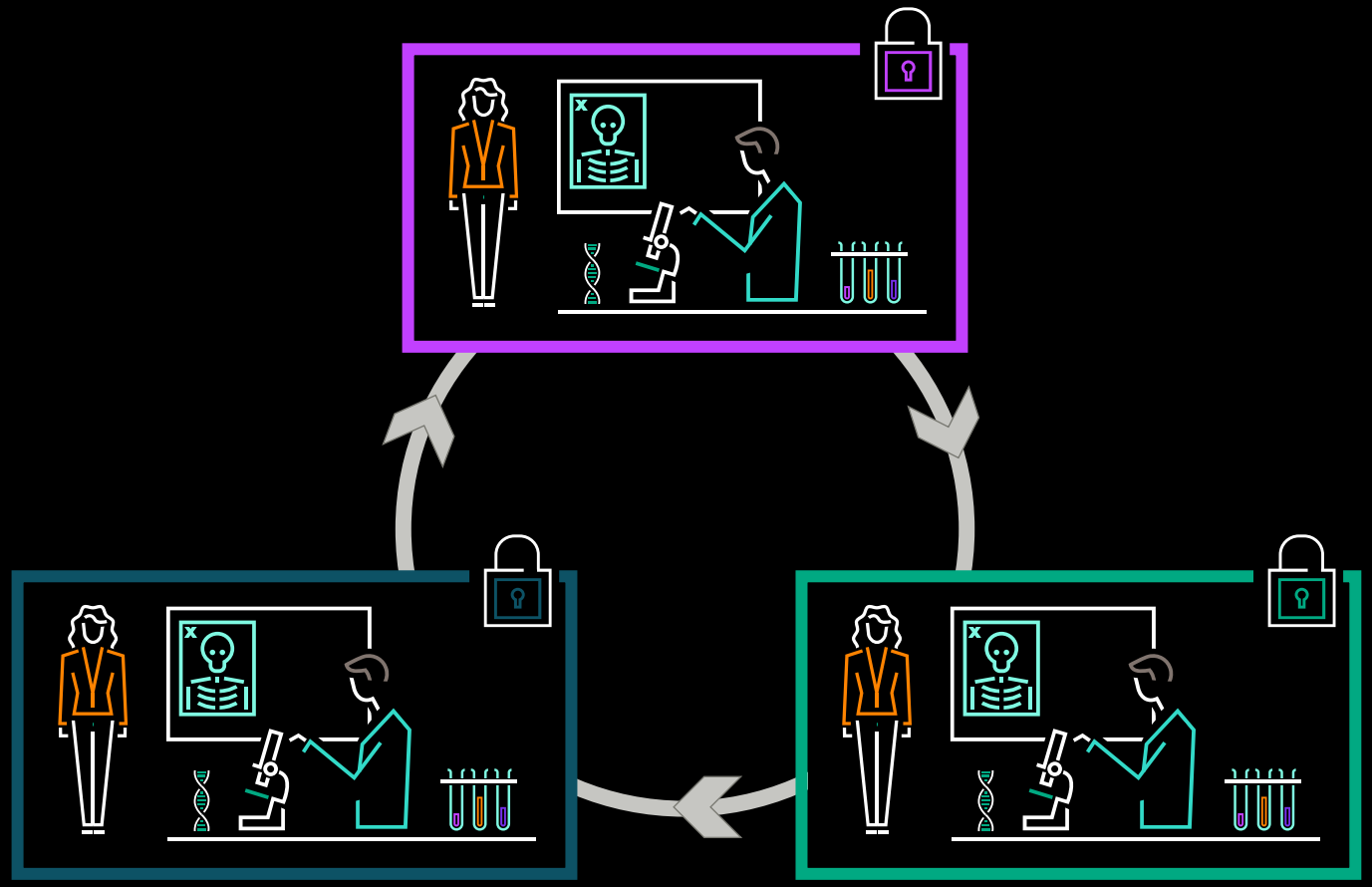
The mentoring principle has existed since ancient times

MACHINE LEARNING—TODAY'S CHALLENGES

Low efficiency	Lots of data movement and duplication
Lack of Data Privacy	Privacy acts. e.g. GDPR, prevent movement of data
Biased Data	Data biases due to demographic distribution
Collaboration challenges	No framework to enable cross enterprise collaboration

SWARM LEARNING IN HEALTHCARE—A NEW APPROACH

Decentralize Learning

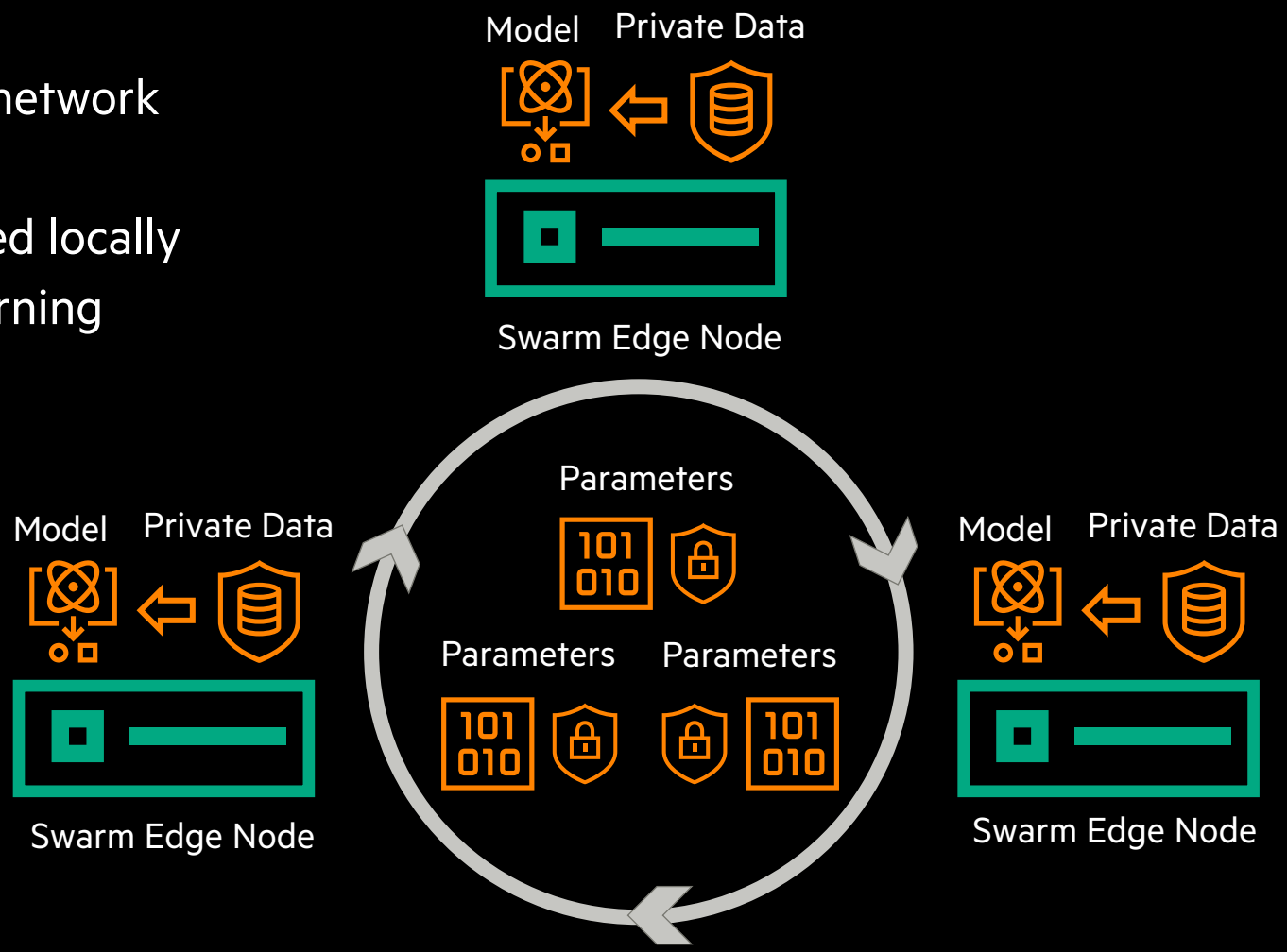


- Collaborate with other institutes
- Do not move the data
- Keep ownership, privacy
- Share insights
- Enlarge dataset with a large network
- Reduce local bias

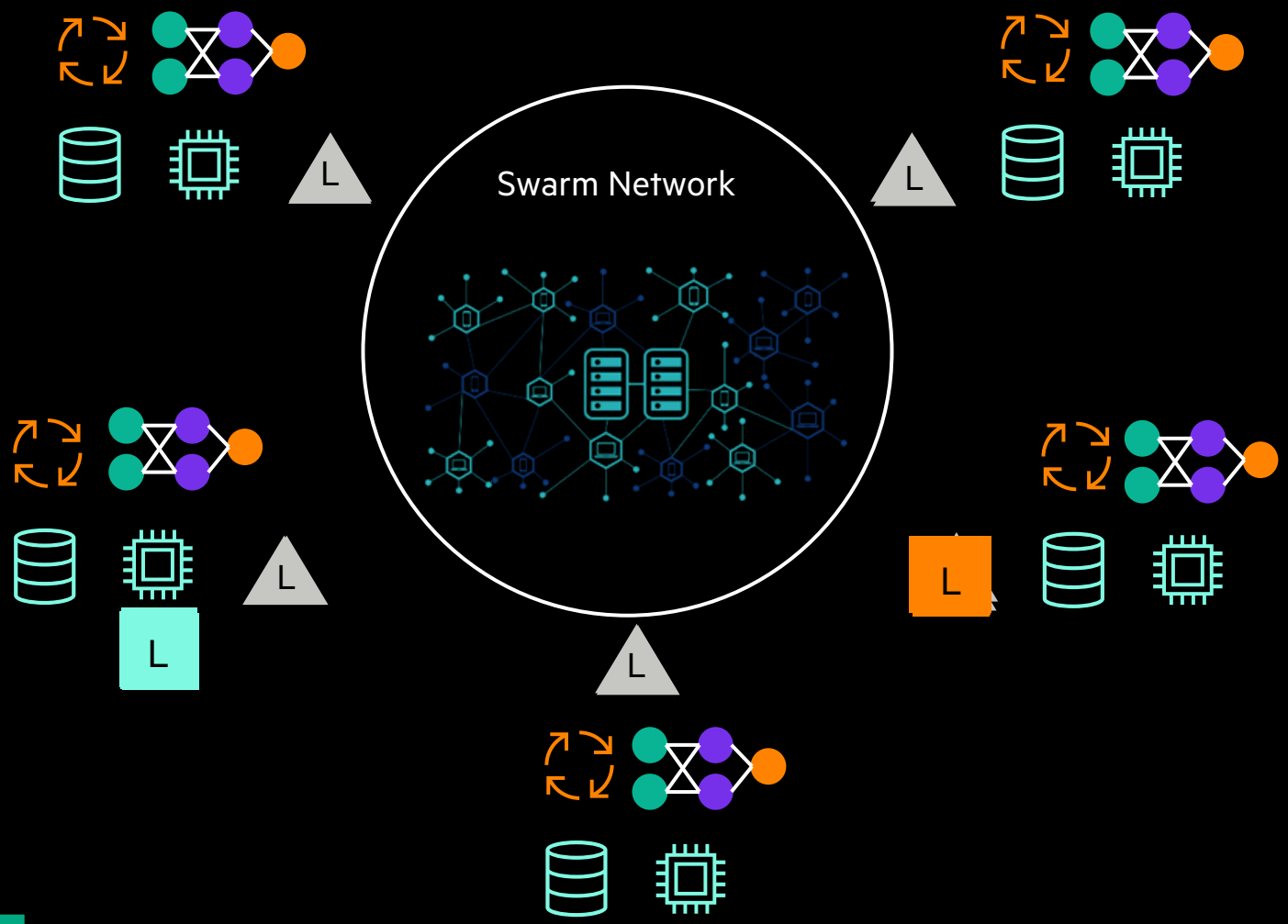
SWARM LEARNING PRINCIPLES

Democratic Machine Learning

- Equal and like-minded partners in the network
- Ownership of the data remains local
- Data protection and data security solved locally
- Less susceptible to bias in machine learning



HPE AI SWARM LEARNING—PROCESS FLOW



1. **Register**
   Nodes register to Swarm Network and receive ML model

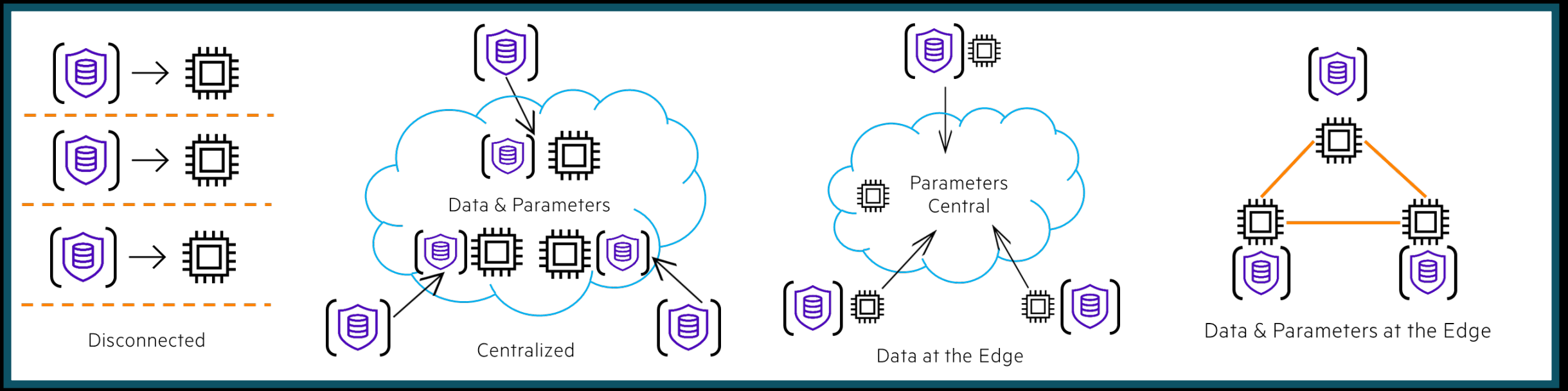
2. **Train**
   Nodes train the model on local data for a time-window (epoch) and share insights

3. **Merge**
   Nodes share and merge the trained models

4. **Repeat**
   Repeat 2 & 3 until desired accuracy is achieved

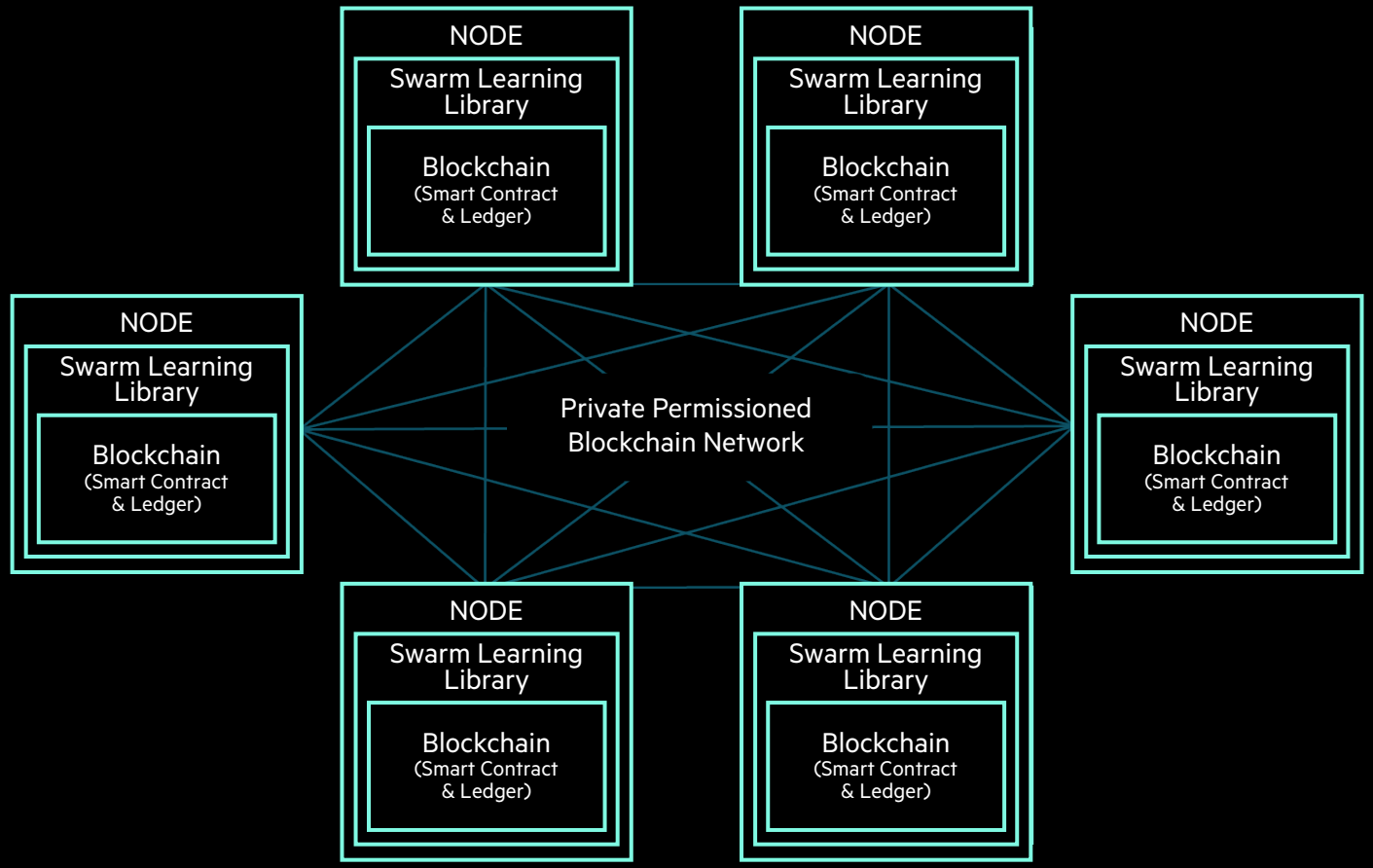
MACHINE LEARNING FRAMEWORK JOURNEY

Local Learning → Centralized Learning → Federated Learning → Swarm Learning



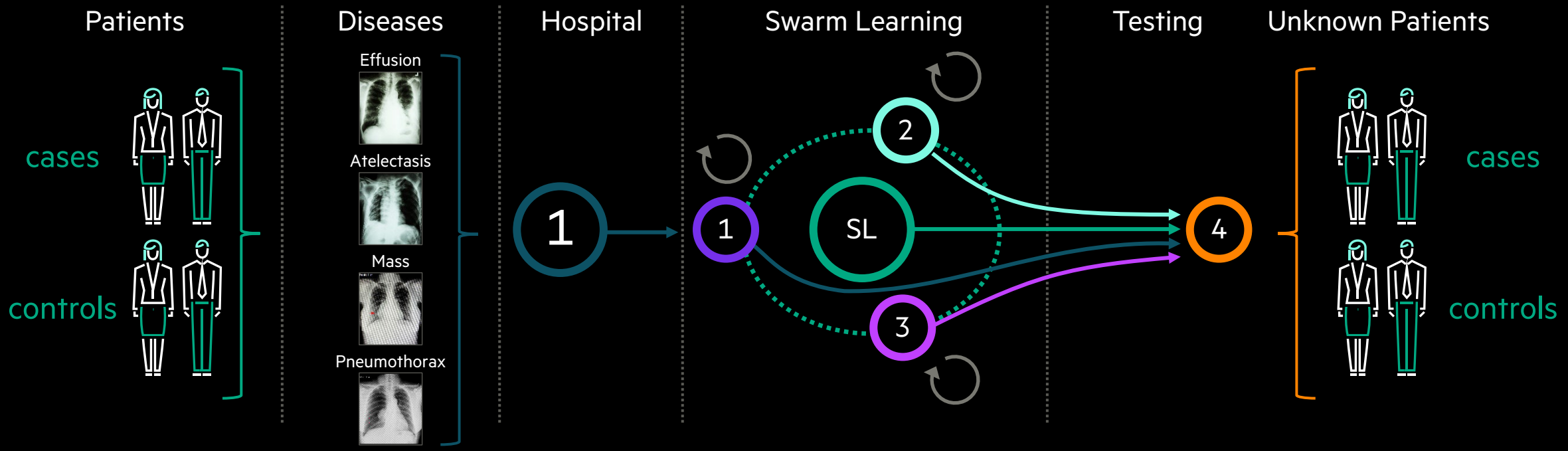
Swarm Learning enables privacy-preserving, secure and collaborative machine learning by treating all participants equally

SWARM LEARNING—THE NETWORK WITH BLOCKCHAIN



- Private permissioned blockchain network
- Dynamic onboarding of new nodes
- Prevents unauthorized access to network
- Prevents insider attack from semi-honest or dishonest participants
- Reduces potential for reconstruction attacks

DISEASE DIAGNOSIS WITH SWARM LEARNING—DOES IT WORK?

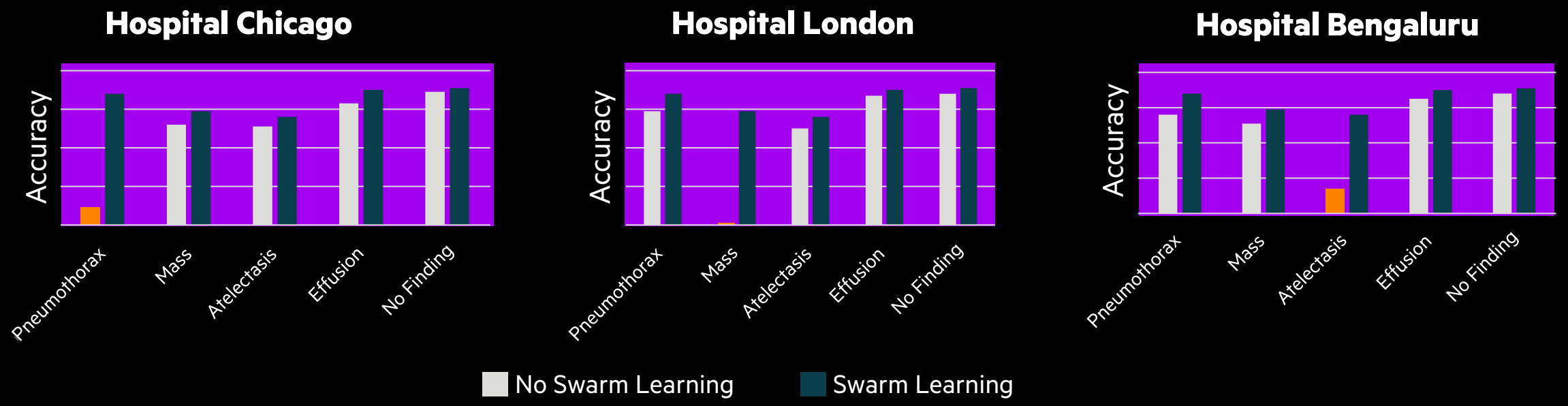


	Pneumothorax	Mass	Atelectasis	Effusion
Chicago Hospital 1	●	●	●	●
London Hospital 2	●	●	●	●
Bengaluru Hospital 3	●	●	●	●

● Sufficient Images
● Limited Images

SWARM LEARNING TURNS THE DISTRIBUTED DATA INTO A COMPETITIVE EDGE

Use Case: Disease Identification Based on Chest X-rays



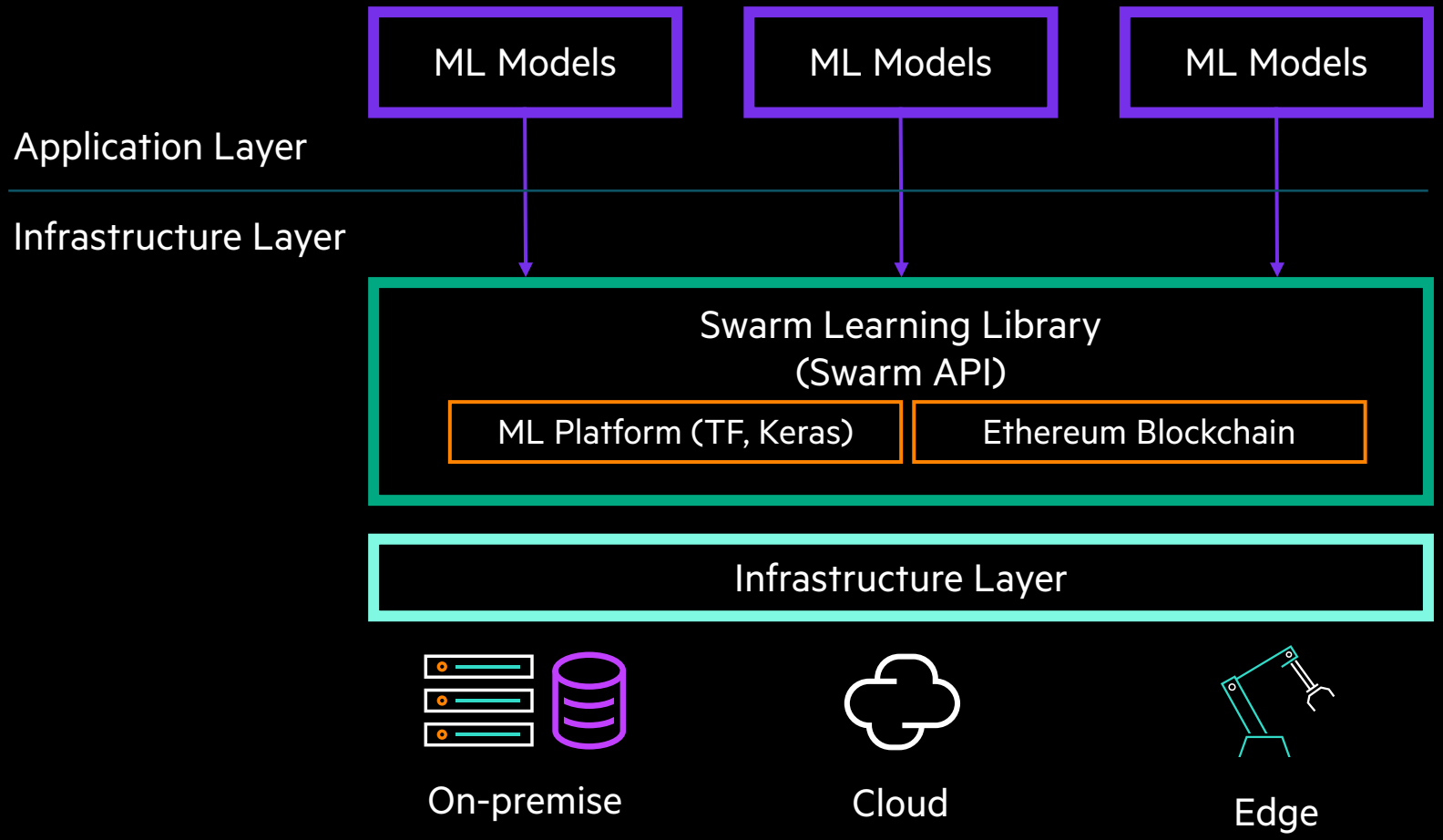
Models at each hospital fail to detect infrequently observed diseases

Swarm Learning model can detect those diseases where the hospital has limited data for that category

Even with sufficiently available data, Swarm Learning model is either better or at par with any individual model

SWARM LEARNING STACK

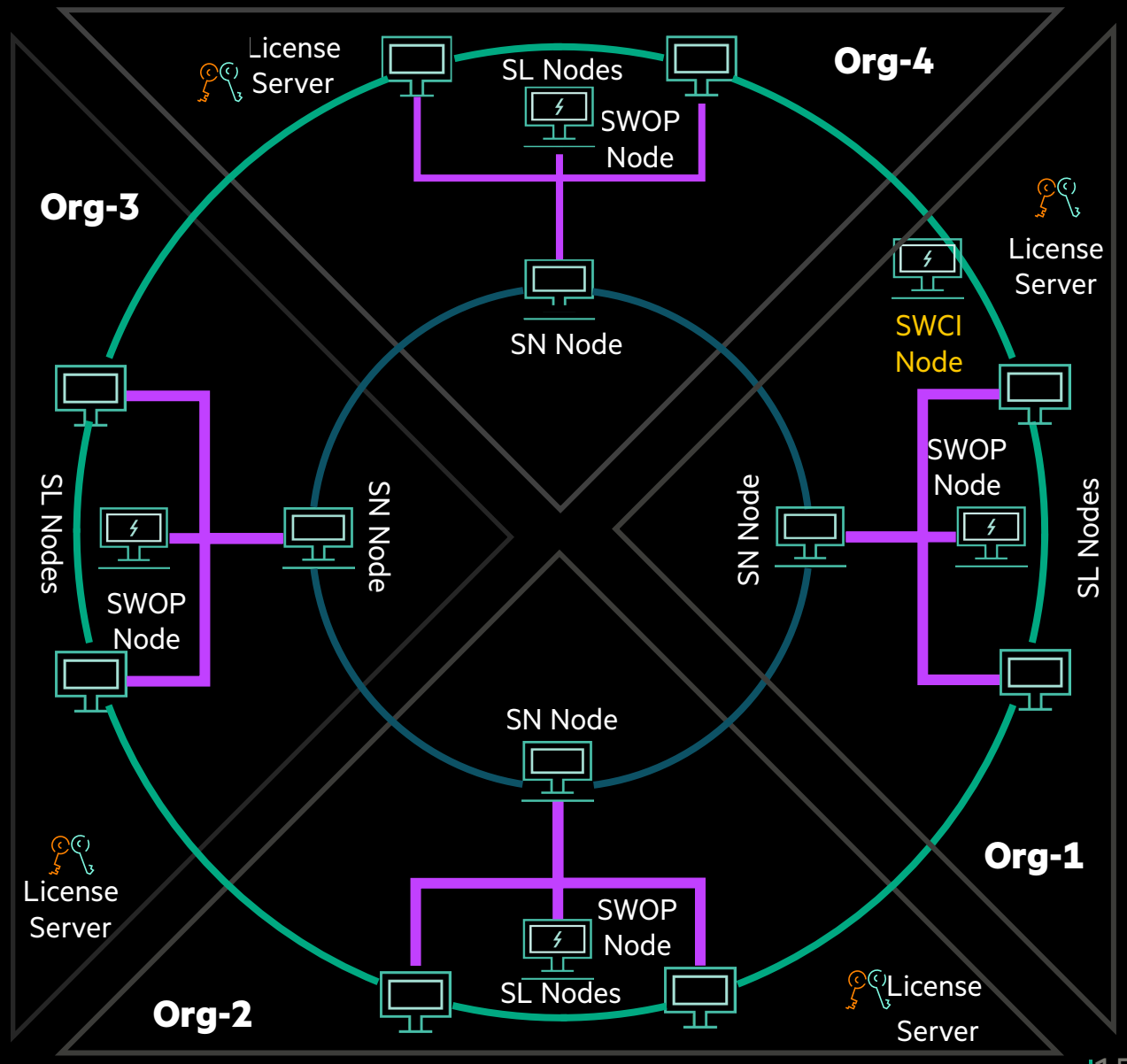
- Swarm Learning Library (SLL) provided as containers
- Simple callback API for integration with ML models
- Tunable hyper parameters
- Management commands to control the network



DEPLOYMENT VIEW

- 6 components / containers
  - Swarm Learning (SL) – Interfacing user-defined Machine Learning programs
  - Swarm Network (SN) - the blockchain network
  - SWCI – Command Interface
  - SWOP – Command and Control Agent
  - License server – AutoPass license server
  - WebUI – Ease of use

- Enable starting small Swarm Networks and then growing big by combining them



HOW TO EASILY TURN ML ALGORITHMS INTO SWARM LEARNING

Integrate using Swarm callback API

```
# 1. IMPORT SWARM CALLBACK #
from swarm import SwarmCallback
…
# 2. DEFINE SWARM CALLBACK #
swarmCallback =
SwarmCallback(rv_interval=1,
min_peers=2, num_epochs=500,
val_batch_size=VALID_BATCH_SIZE)
…
# 3. ADD SWARM CALLBACK IN THE
CALLBACK LIST FOR TRAINING #
callbacks_list = [swarmCallback]
```

Actual code difference for disease diagnosis model (in Keras)

```
$ diff nih_keras.py nih_swarm.py
29a30,31
> from swarm import SwarmCallback
>
359c361,367
<     callbacks_list = [checkpoint, early]
---
>     swarmCallback = SwarmCallback(rv_interval=1,
>             min_peers=3,
>             num_epochs=500,
>             val_data=valid_gen,
>             val_batch_size=32)
>
>     callbacks_list = [swarmCallback, checkpoint, early]
```

Swarm Learning Software is provided as Containers with APIs for easy integration

USE CASE: SWARM LEARNING COLLABORATION ("SWARM OF SWARMS")

Parameters

Swarm Network

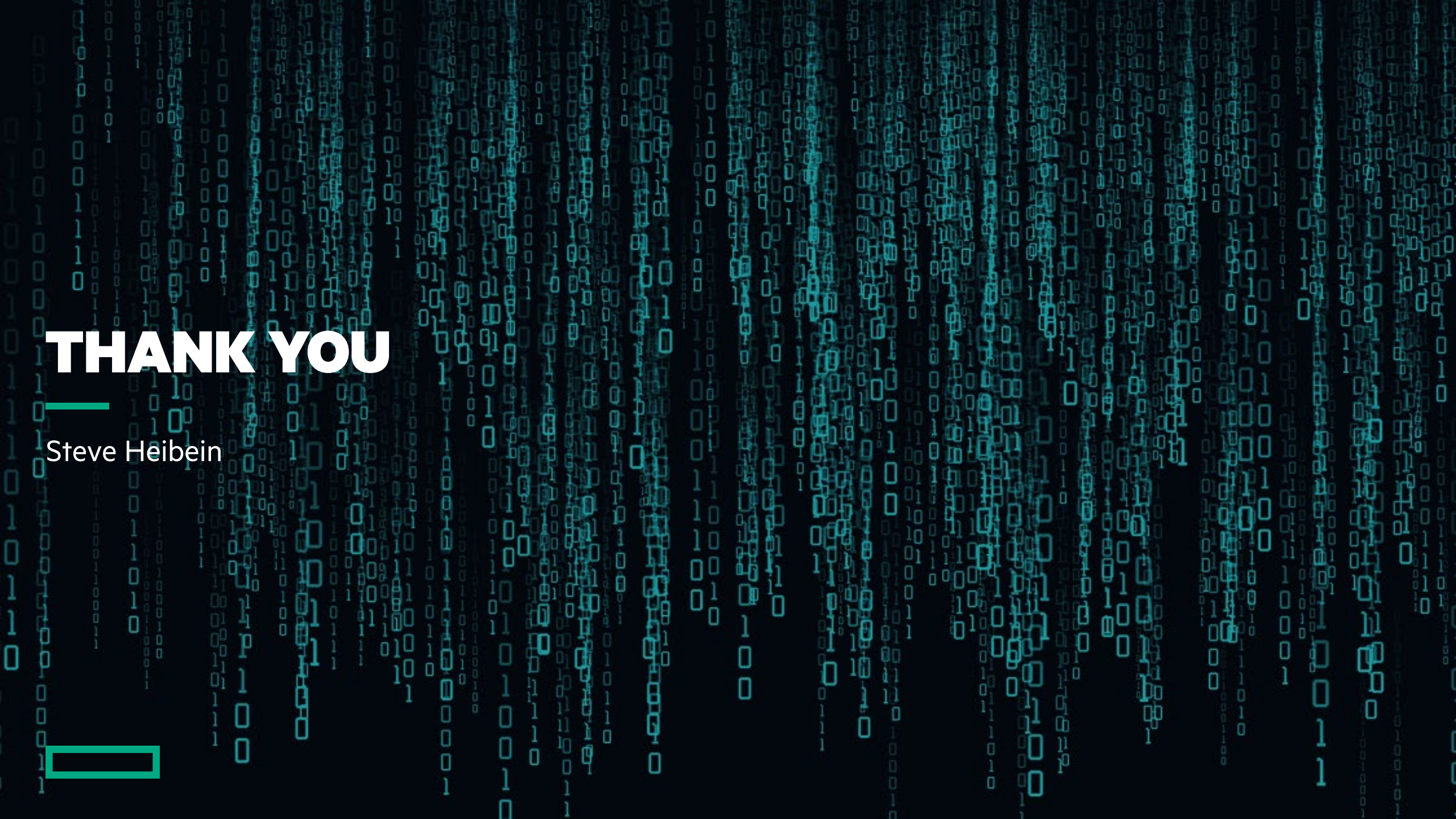
Private, Permissioned Blockchain Network

Parameters

Parameters

WHERE TO LEARN MORE ABOUT SWARM LEARNING

- Referenced article from Nature: “Swarm Learning for decentralized and confidential clinical machine learning”
  https://www.nature.com/articles/s41586-021-03583-3
- Article in Dark Daily: “What is Swarm Learning and Might It Come to a Clinical Laboratory Near You?”, January 28, 2022
  https://www.darkdaily.com/2022/01/28/what-is-swarm-learning-and-might-it-come-to-a-clinical-laboratory-near-you/
- Video: “What's All the Buzz about Swarm Learning?”, December 7, 2021
  https://www.youtube.com/watch?v=JqwjyENd89I
- Article: “Swarm learning: Driving advances both practical and profound”, November 12, 2021
  https://www.hpe.com/us/en/insights/articles/swarm-learning-driving-advances-both-practical-and-profound-2111.html
- Article: “HPE's Dr. Goh on harnessing the power of swarm learning”, September 8, 2021
  https://www.hpe.com/us/en/insights/articles/hpes-dr-goh-on-harnessing-the-power-of-swarm-learning-2109.html
- Article from Analytics India Magazine (AIM): “Does Swarm Learning Have An Edge Over Federated Learning?”, July 28, 2021
  https://analyticsindiamag.com/does-swarm-learning-have-an-edge-over-federated-learning/
- Article from DZNE: “AI with Swarm Intelligence: A Novel Technology for Cooperative Analysis of Big Data”, May 26, 2021
  https://www.dzne.de/en/news/press-releases/press/ai-with-swarm-intelligence/
- Video: “The Big Shift: What is Swarm Learning?”, November 18, 2020
  https://www.hpe.com/h22228/video-gallery/us/en/700000804/EN/US/c8885cdb-ff75-4171-b6cc-4ea0a5ff01af/the-big-shift-what-is-swarm-learning/video?lang=en-US
- HP Labs Technical White Paper:
  “Swarm Learning: Turn Your Distributed Data Into Competitive Edge”, February 2020, Rev. 1
  https://www.hpe.com/psnow/doc/a50000344enw


THANK YOU
Steve Heibein

HPE CRAY SUPERCOMPUTERS WIN—EXASCALE TRIPLE-CROWN

ANL “Aurora”

- More than 2 EF Sustained performance
- Future Intel Xeon CPU and Intel X^{e} architecture and Slingshot interconnect
- Mixed AI and HPC workload

ORNL “Frontier”

- More than 1.0 EF Sustained performance
- Future AMD EPYC CPU and Radeon GPU and Slingshot interconnect
- Mixed AI and HPC workload

LLNL “El Capitan”

- More than 1.5 EF Sustained performance
- Future AMD EPYC CPU and Radeon GPU and Slingshot interconnect
- Mixed AI and HPC workload

Hewlett Packard Enterprise

RMACC
Rocky Mountain Advanced Computing Consortium

HPE SWARM LEARNING

SOLVING THE DATA SHARING CONUNDRUM IN AI

Steve Heibein
US Public Sector AI Lead
stephen.Heibein@hpe.com