AI Analytics Anywhere

Dell driven approach to Enterprise AI unlocks value of siloed data.

DELL Technologies

Challenges

Current Environment

Enterprise Challenges

- How to reach CONUS datacenter?
- Network segmentation
- Data sprawl / Data Swamp

Data Quality Challenges

- No labels, smaller data, local data
- Siloed data
- Data Classification
- Compartmentalization
- Data poisoning, Leakage

Personnel Challenges

- Down to The Edge
- Deploy teams of AI developers?

Ownership Challenges

- Branch to Branch
- Commercial Partner Sharing
- Title concerns
- US vs Partner
- Inter- / Intra- Agency (or both)

Solution: Federated Learning

Distributed / Federated

- Leverage data where it is collected
- Minimize network utilization
- Maximize data utility (time)

Federated Transfer Learning

- Solution for small / poor quality data sets
- Solution for Unlabeled data

Train / Inference from Cloud to Edge to IoT

- AI in the backpack
- Learn across Areas of Interest

Staff Enablement [1]

- NoCode / LowCode Data Wrangling and Modeling
- Speed to AI-Enablement
- Support Digital Transformation
- Defense in Depth Security

Cross-Domain / Cross-Partner / Cross-Branch

- Federate across domains, trusted or untrusted
- Transfer model weights / losses NOT data
- Local Security, Global Accessibility

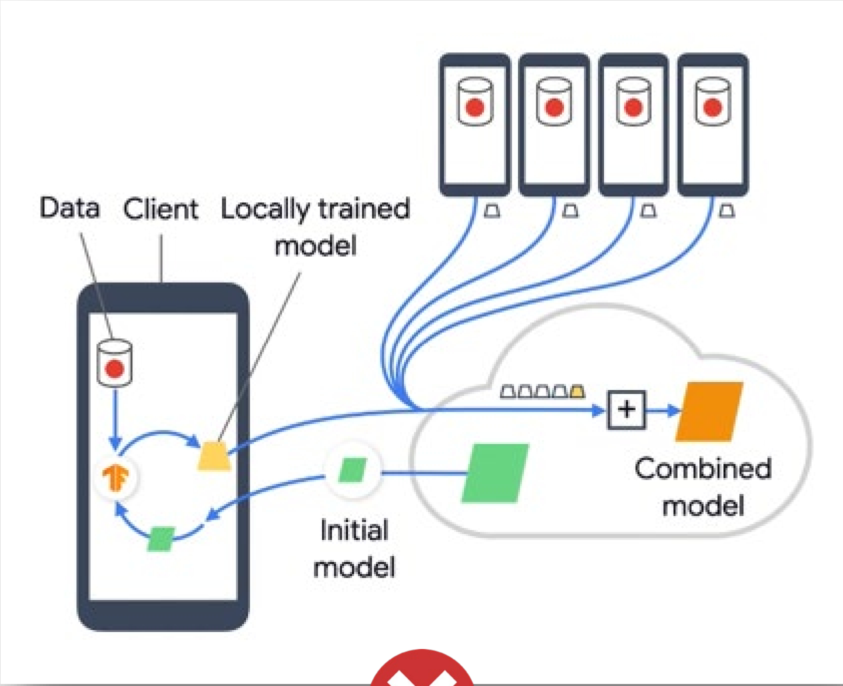
1: Code-Free Artificial Intelligence Enablement Tools Policy (Title LXVII, Sec. 6742)
No later than one year after the 2023 NDAA's enactment, the DNI, in consultation with other specified intelligence community organization heads, shall draft a potential policy to promote the intelligence community's use of code-free AI enablement tools. The policy shall include the objective for the use of these tools, a detailed set of incentives for using these tools, and a plan to ensure coordination throughout the intelligence community

DELL Technologies

What is Federated Machine Learning?

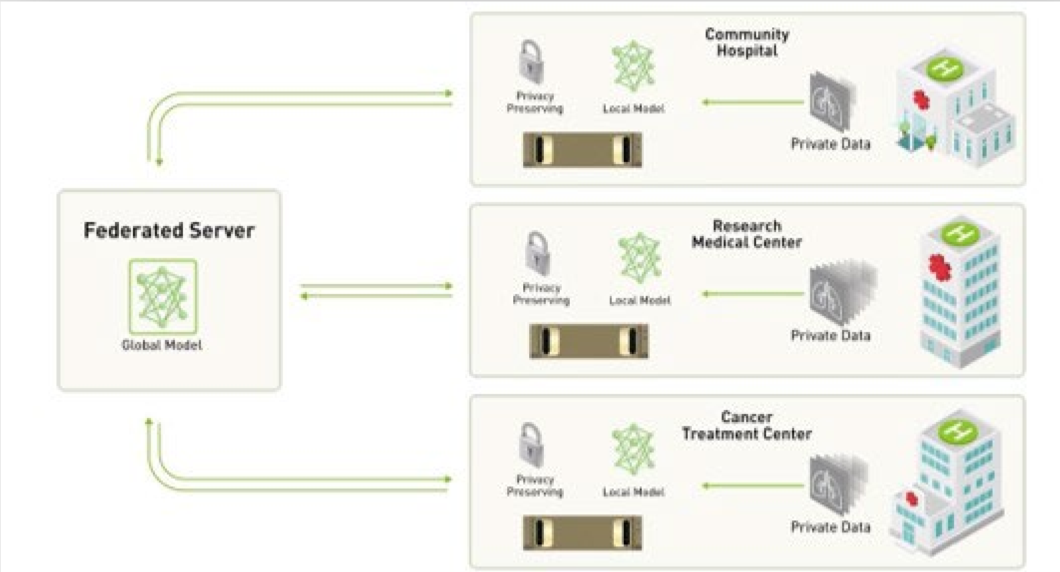
Datacenter to Mobile

1. Personal Privacy & Security
2. Cloud to Billions (IoT)



Datacenter to Edge to IoT

1. Distributed data silos, e.g., collect but no transfer capability
2. Different domains of control, e.g., AF, Army, NSA
3. Different network domains, e.g., low vs high



Federation allows you to leave the data where it is.
Immediately start working with data from across your organization, and across partners.

Two Main Categories of Federated Learning

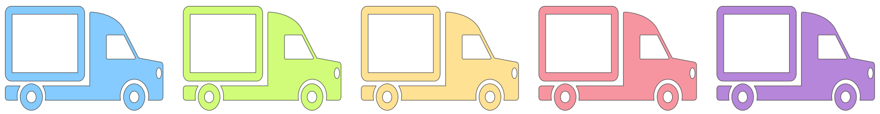
HORIZONTAL LEARNING

Similar Data, Multiple Places

Higher Quality Models

Local vs Global Intelligence

Multi-classification Models



VERTICAL LEARNING

Disparate Data, Multiple Places

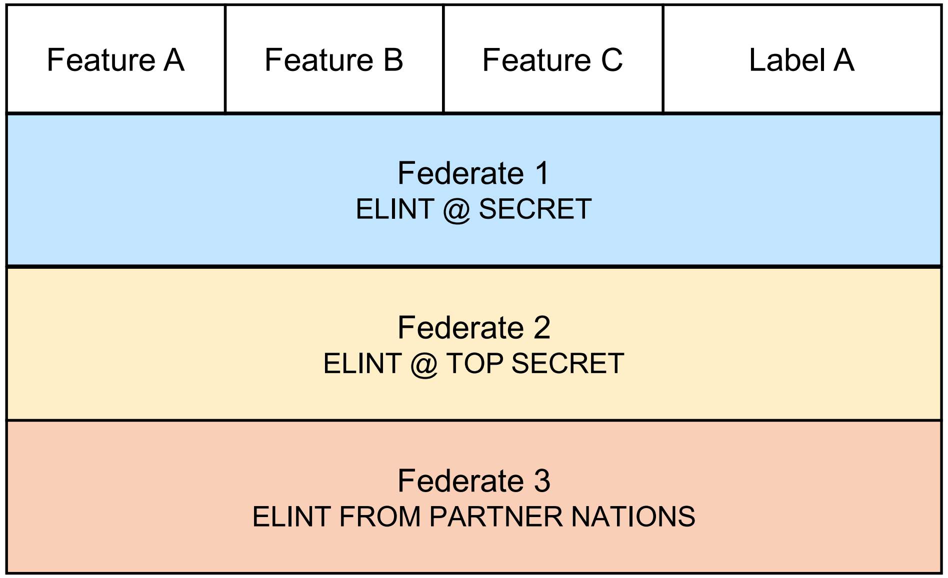
Sensor Fusion

Network Building

Logistics Discovery & Supremacy

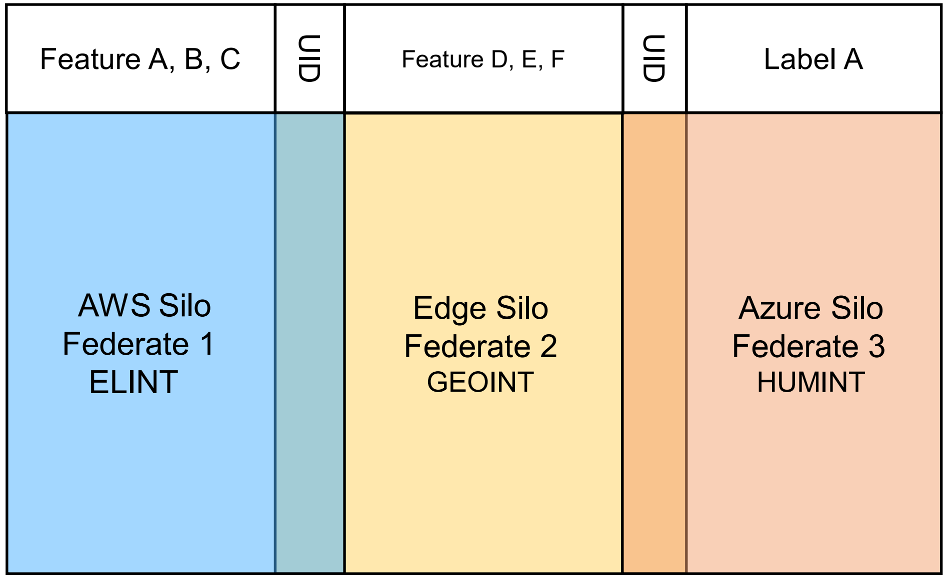
Two Main Categories of FML

Horizontal Learning



DATA COMPOSED OF HOMOGENOUS ROWS FROM EACH PLATFORM

Vertical Learning



DATA LINKED BY COMMON ENTITY IDS ACROSS PLATFORMS

Model Security: Privacy Preserving Techniques

Layered approach to Zero Trust

Secure Multi-party Computation[3]

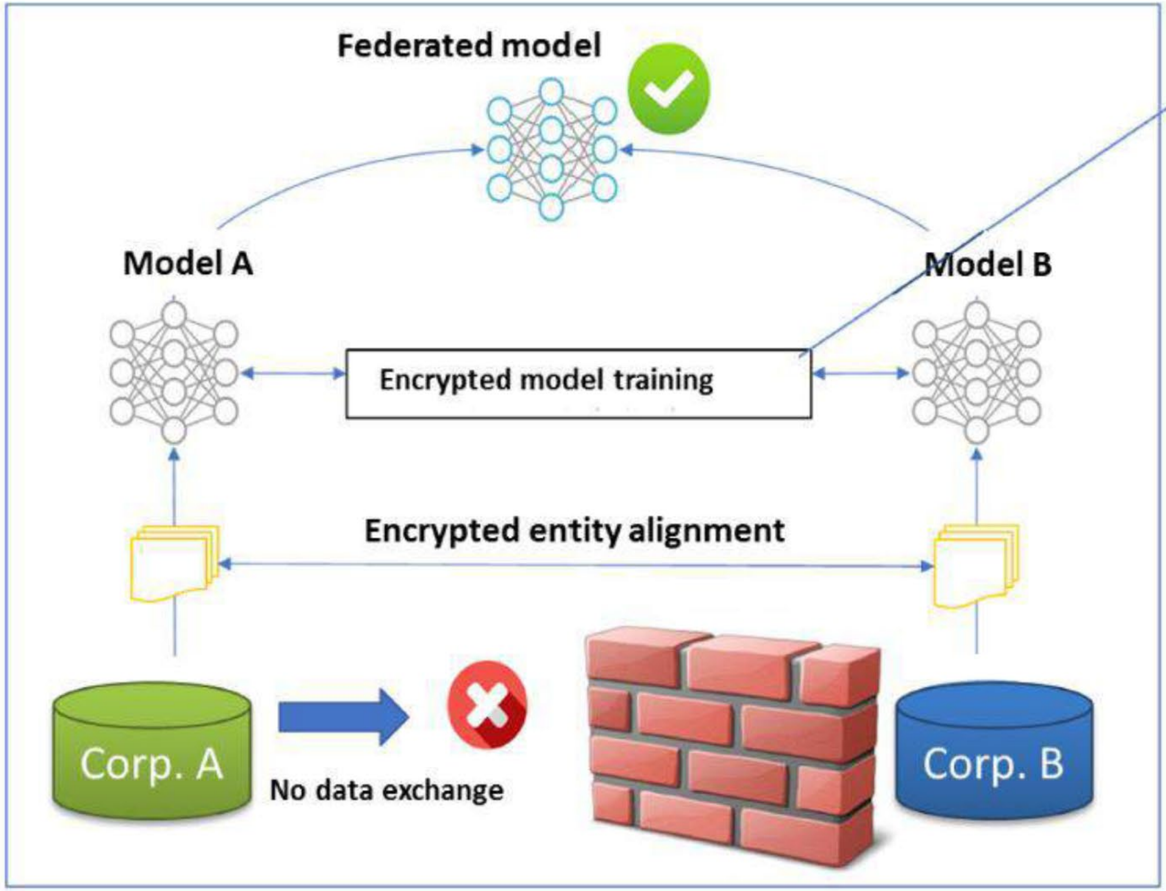
Homomorphic Encryption

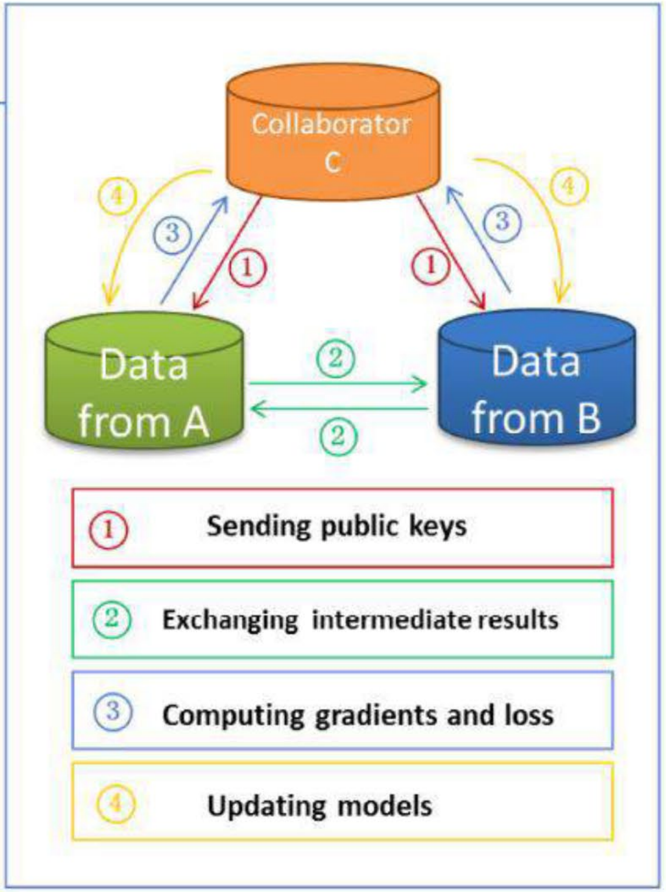
Yao's Garbled Circuit

Secret Sharing

Differential Privacy[3]

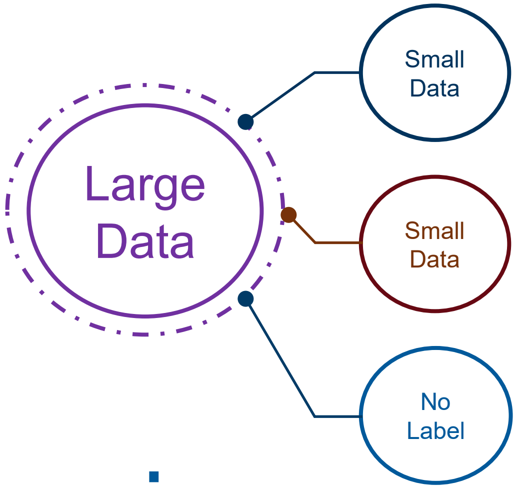
Oblivious Transfer

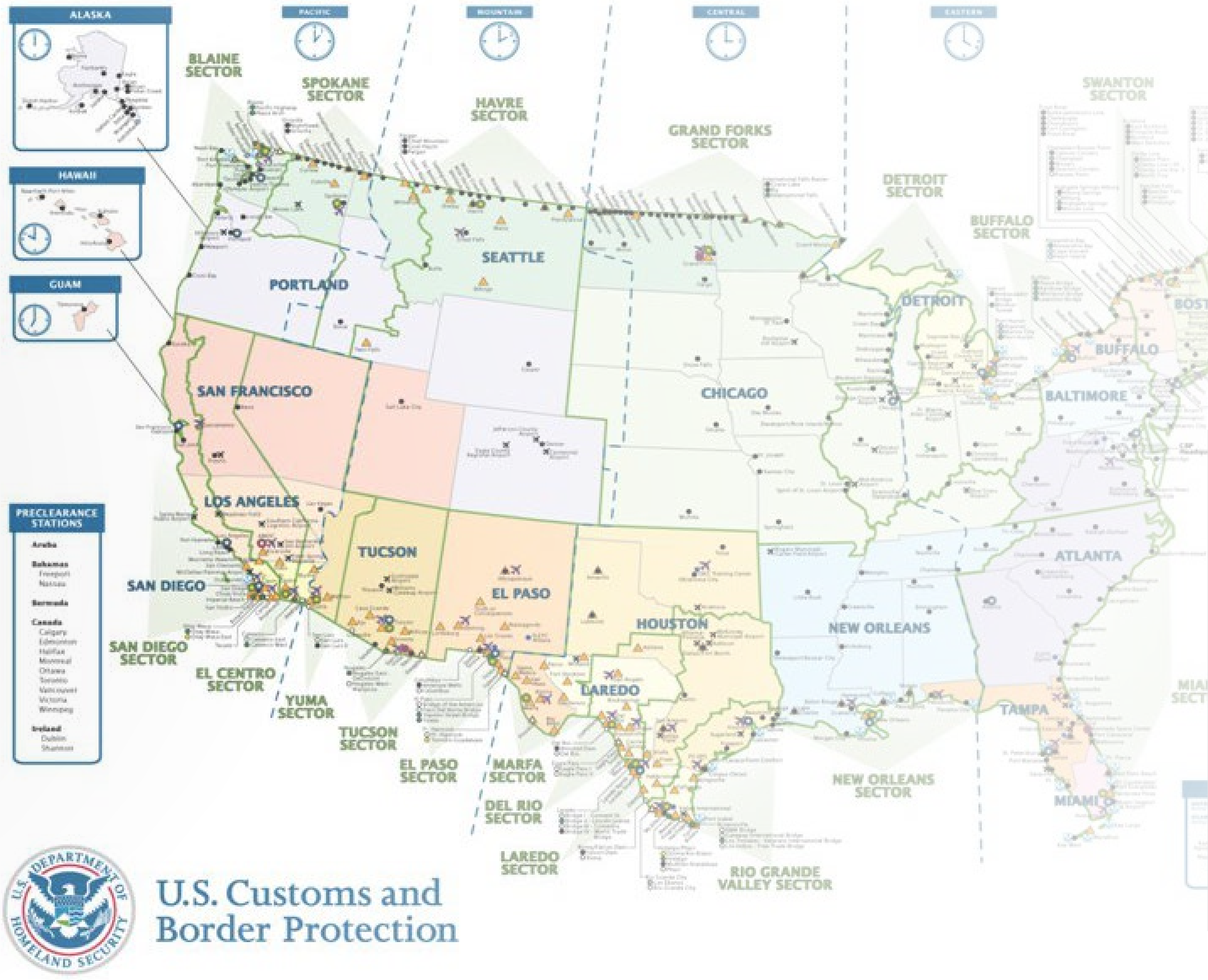




2:Within one year of the 2023 NDAA's enactment, the OMB Director shall ensure federal contracts for AI acquisition align with proper guidance, include considerations for securing algorithms and their training data, and address relevant privacy and other issues, among other things.

3:20223 NDAA : Rapid Pilot, Deployment, and Scale of Applied Artificial Intelligence Capabilities to Demonstrate Modernization Activities Related to Use Cases (Title LXXII, Sec. 7226)
No later than 270 days after the NDAA's enactment, the OMB Director will lead a pilot program that identifies four new use cases for AI in support of interagency or intra-agency modernization initiatives—and that require linking multiple siloed data sources. Then, no later than one year after the NDAA's enactment, the OMB Director shall coordinate with other federal entities to initiate the piloting of the four AI use cases.
The Director shall prioritize modernization projects that would benefit from commercially available, privacy-preserving techniques (such as **differential privacy, federated learning, and secure multiparty computing**) and otherwise would account for civil rights and civil liberties considerations.





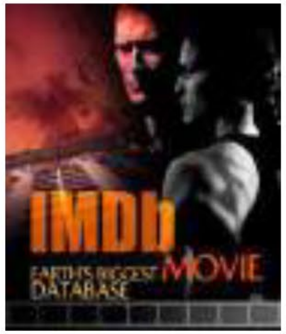
Transfer Learning

- Transfer Learning Across Organization
  - Learn the Spread
  - Transfer Knowledge Between Locations
- Transfer Lessons Learned in the Field
  - Model Sharing
- Transfer between Data Domains
  - Overcome Small or Low-quality Data
  - Overcome No Labels

DELL Technologies

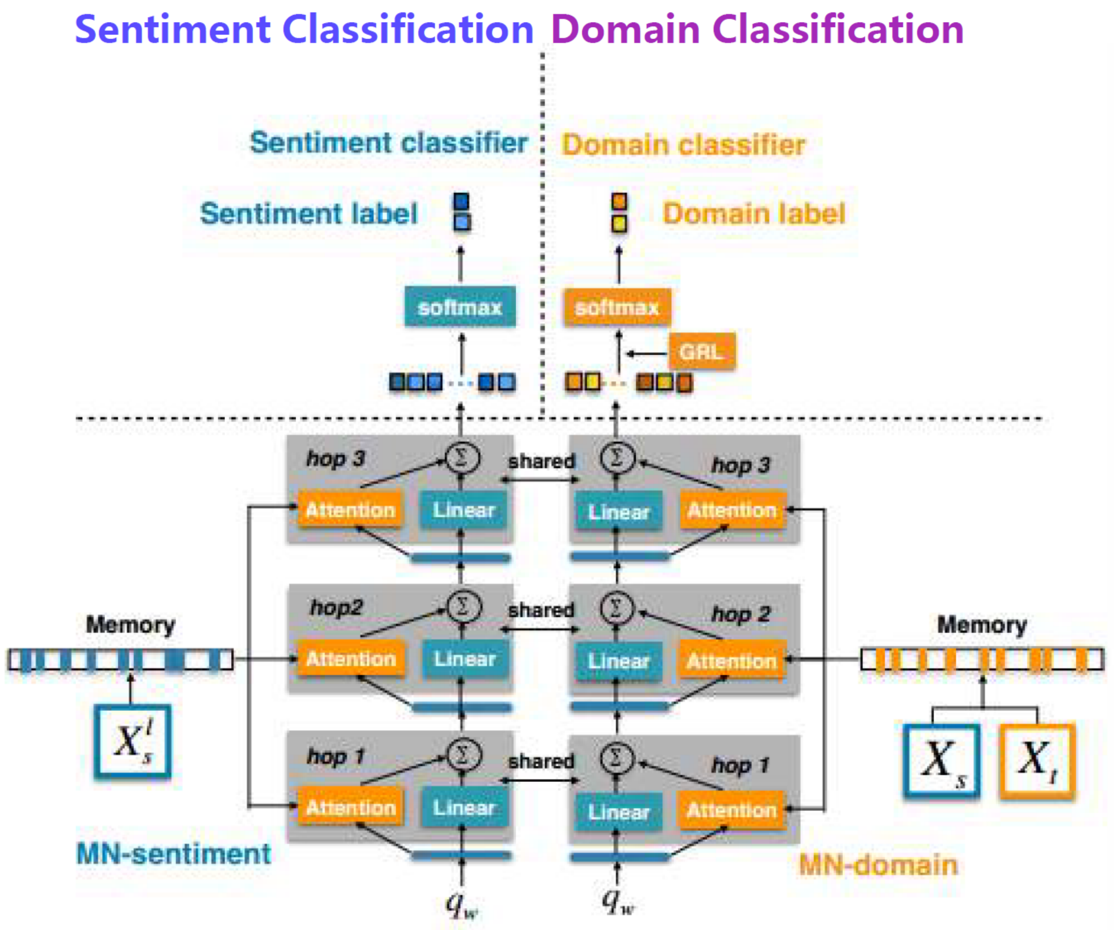
Learning Across Knowledge Domains

How Dell leverages unlabeled data for AI / ML:

Source domain (**Movie**)	Target domain (**Electronics**)
	
Great movie. His characters are **engaging** and **thoughtful**.	This great touchpad feels **glossy** and is **responsive**.
It's a **excellent**, **sobering** drama.	It is very **lightweight**. **excellent** transition from PC.
An **terrible** movie. It is very **plotless** and **insipid**.	It is **blurry** and **fuzzy** in very dark setting. So **terrible** HP.

DON'T turn your people into labelers, auto-identify them via "pivots".

Transfer Learning: Domain Classification



Domain Classification Objective:

Maximize domain classification error

Source data X_s

Target data X_t

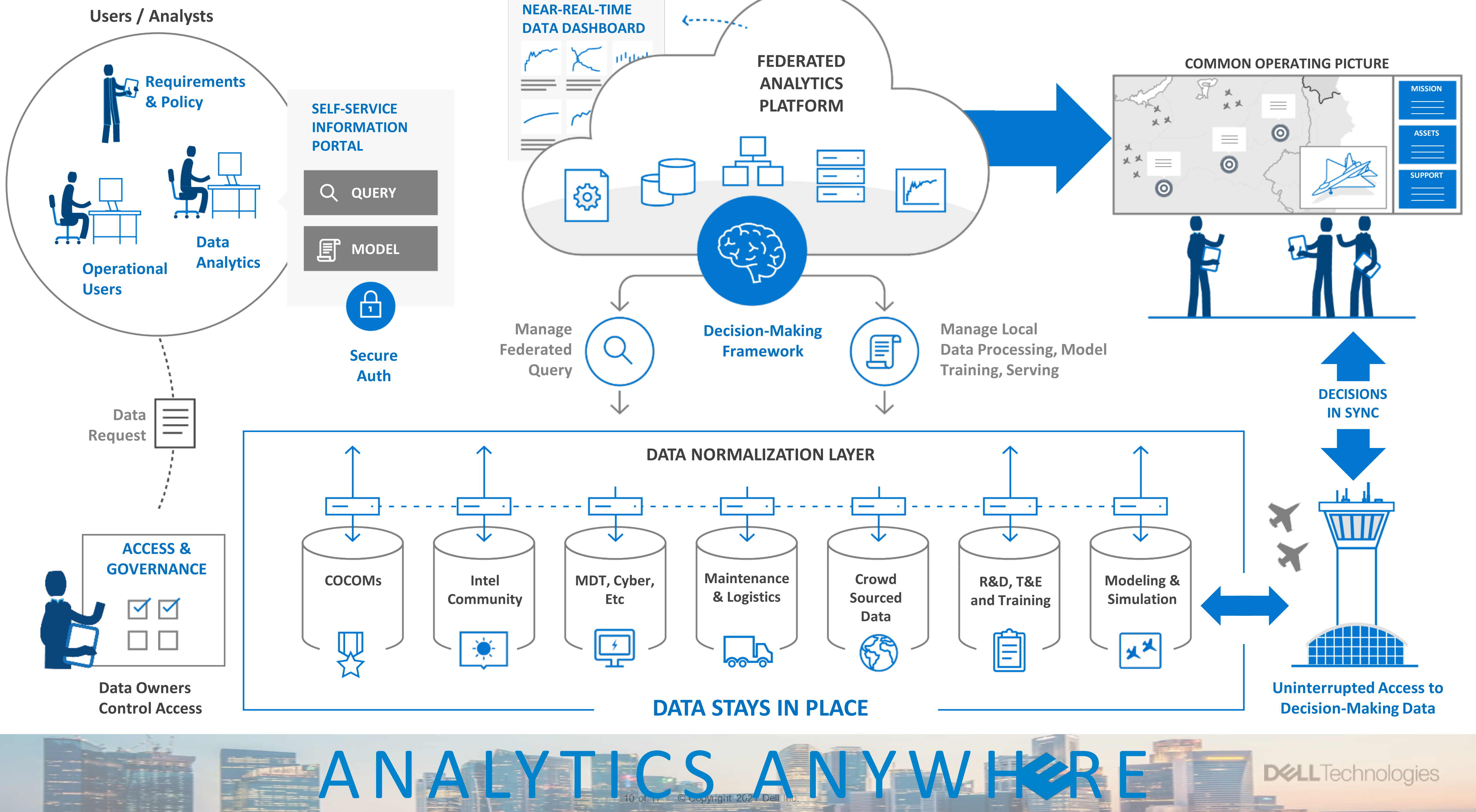
S

T

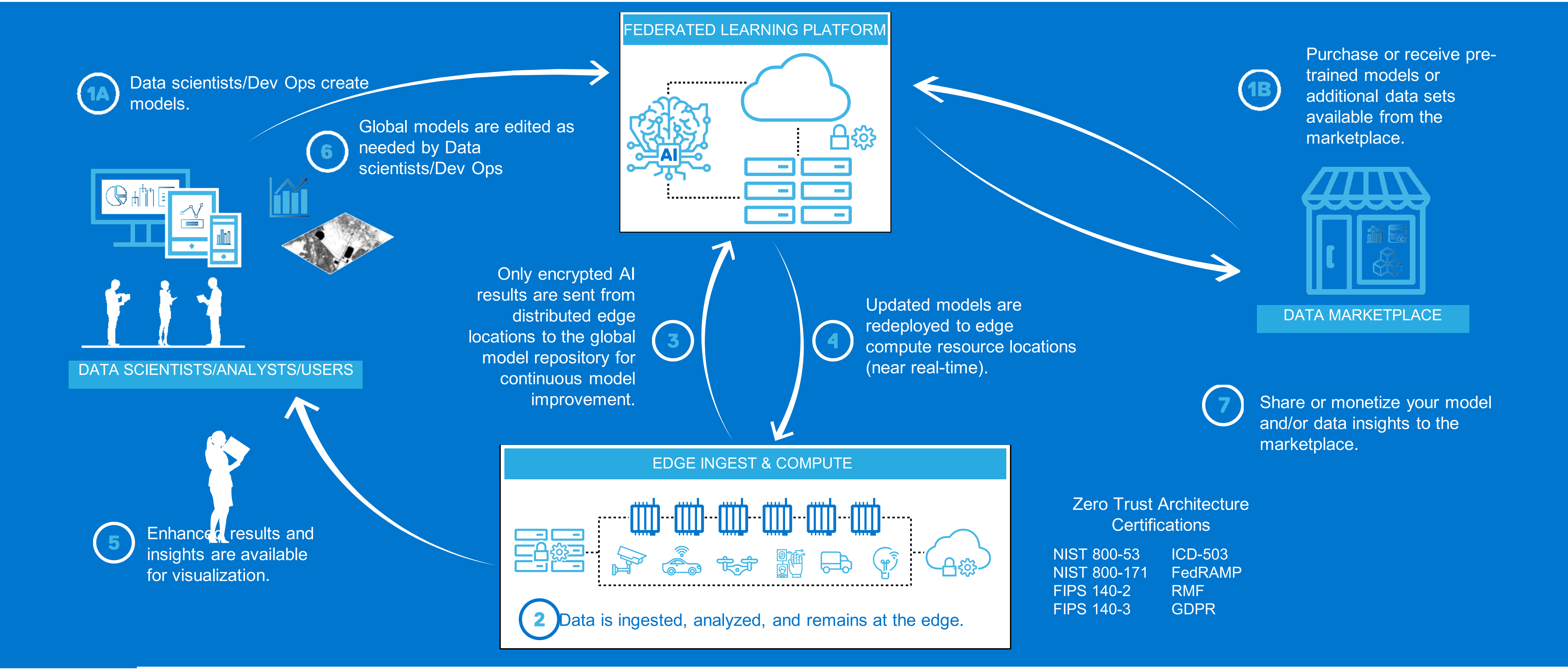
Movie
Great movie. His characters are engaging and thoughtful.
Electronics
This great touchpad feels glossy and is responsive.

Pink words are domain-independent pivots, green words and blue words are auto-identified, significant labels.

DELL Technologies

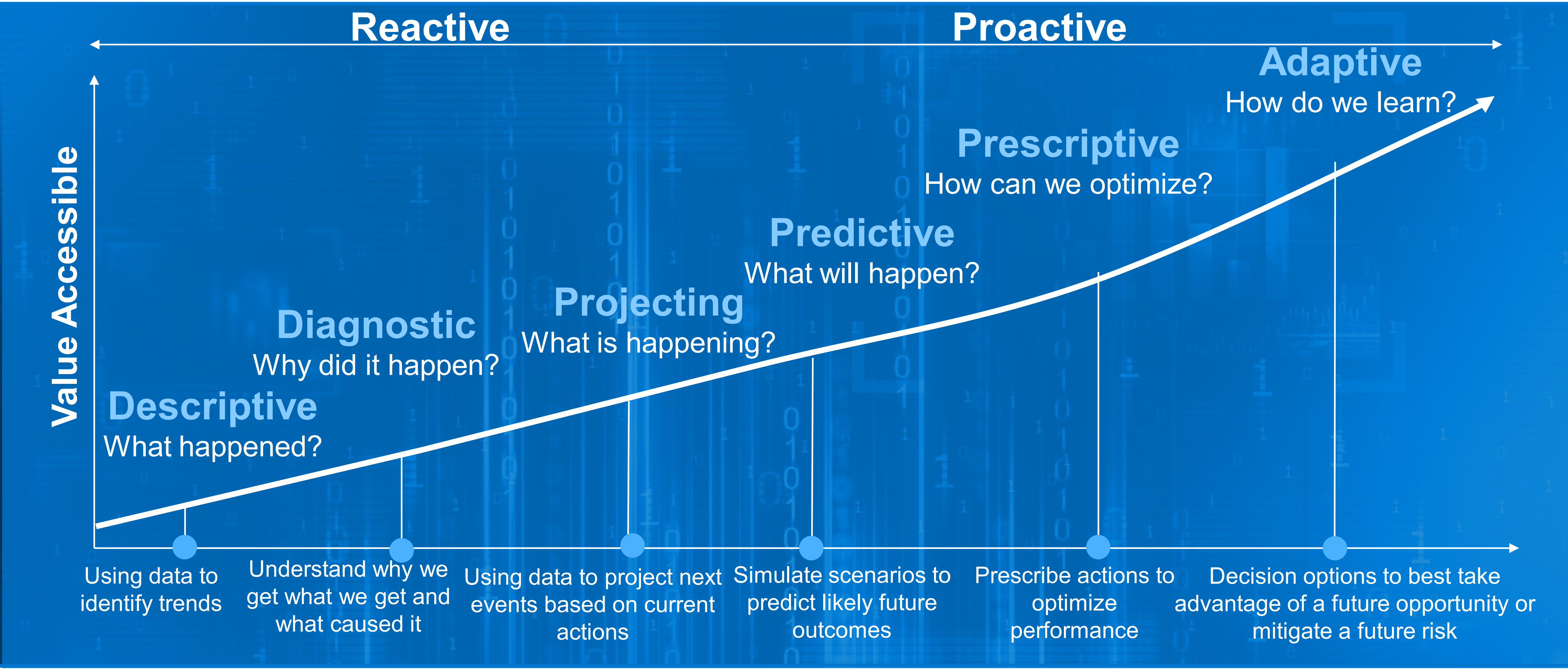

Users / Analysts
Requirements & Policy
Operational Users
Data Analytics
SELF-SERVICE INFORMATION PORTAL
QUERY
MODEL
Secure Auth
NEAR-REAL-TIME DATA DASHBOARD
FEDERATED ANALYTICS PLATFORM
COMMON OPERATING PICTURE
MISSION
ASSETS
SUPPORT
Manage Federated Query
Decision-Making Framework
Manage Local Data Processing, Model Training, Serving
DECISIONS IN SYNC
Data Request
DATA NORMALIZATION LAYER
COCOMs
Intel Community
MDT, Cyber, Etc
Maintenance & Logistics
Crowd Sourced Data
R&D, T&E and Training
Modeling & Simulation
ACCESS & GOVERNANCE
Data Owners Control Access
DATA STAYS IN PLACE
Uninterrupted Access to Decision-Making Data
ANALYTICS ANYWHERE
DELL Technologies

Federated Learning Framework



DELL Technologies

Analytics Maturity Curve



Analytic Sophistication

DELL Technologies

Providing Relevant Outcomes

The complexity of process and technology integration

The Process

Use case
The business requires new **knowledge for operations or decision-making.**

- Prioritize list of business problems
- Understand the problem

Data sources
Business users and data architect identifies relevant data sources.

- Discover source data set(s)
- Get access to the data

1

Data ingestion
Data architect identifies data sources and creates an **ETL pipeline to ingest relevant data.**

- Select where to perform the workload
- Select & provision exploration tooling
- Explore the data
- Develop data pipelines?
- Source data ingestion

2

Data storage
Different data types are segmented and stored in data lakes, data warehouse or databases with search and query capability.

- Transform / cleanse / enrich the data
- Visualize/ dashboard the data
- Store & Catalogue Data
- Automate data pipeline management
- Monitor data pipeline

3

Data analysis
Data scientists and data analysts apply analytics techniques, machine learning algorithms to gather hidden insights from data.

- Select data science tools
- Feature selection
- ML model dev & testing
- Model validation testing
- Data profiling / inferencing
- Manage ML model drift
- Continuous A/B Testing

4

Data visualization
Insights from data are visualized to tell stories, highlight useful trends and outliers for stakeholders to understand easily.

- Package Data-as-a-Service
- Evaluate Business Results

Business consumption
Business users share, discuss and collaborate to leverage new insights for better business performance.

- Integrate data product with operational systems

Acquire, grow & retain customers

Optimize / automate operations

Maximize insights & improve economics

Improve business performance

Create new business models

The People

Business stakeholders
- Business expertise
- Business strategy
- Business value drivers
- Use cases

Business analysts
- Industry expertise
- Business engagement
- Process engineering improvement
- Brand analytics

Data engineers
- Data platform ecosystem
- ETL / ELT tooling
- Master / reference data integration
- Data streaming, logging
- Data services development

Data scientists
- Industry expertise
- Statistics, mathematics, AI / ML / DL
- Data visualization
- Spark, Python, R, SAS…

DevOps
- Application/Platform architecture ecosystem
- API design and implementation
- Microservice design and implementation
- Data services integration
- Performance and SLA / SLOs
- Mobile / UI interface
- IoT / Endpoint integration

ITOps
- Infrastructure as Code
- Virtualization and containers
- Data pipeline, management & governance
- QA, Security, & Compliance
- Instrumentation and monitoring
- Data protection / BC / DR


DELL Technologies

Federated vs Centralized

Predictive performance and bandwidth

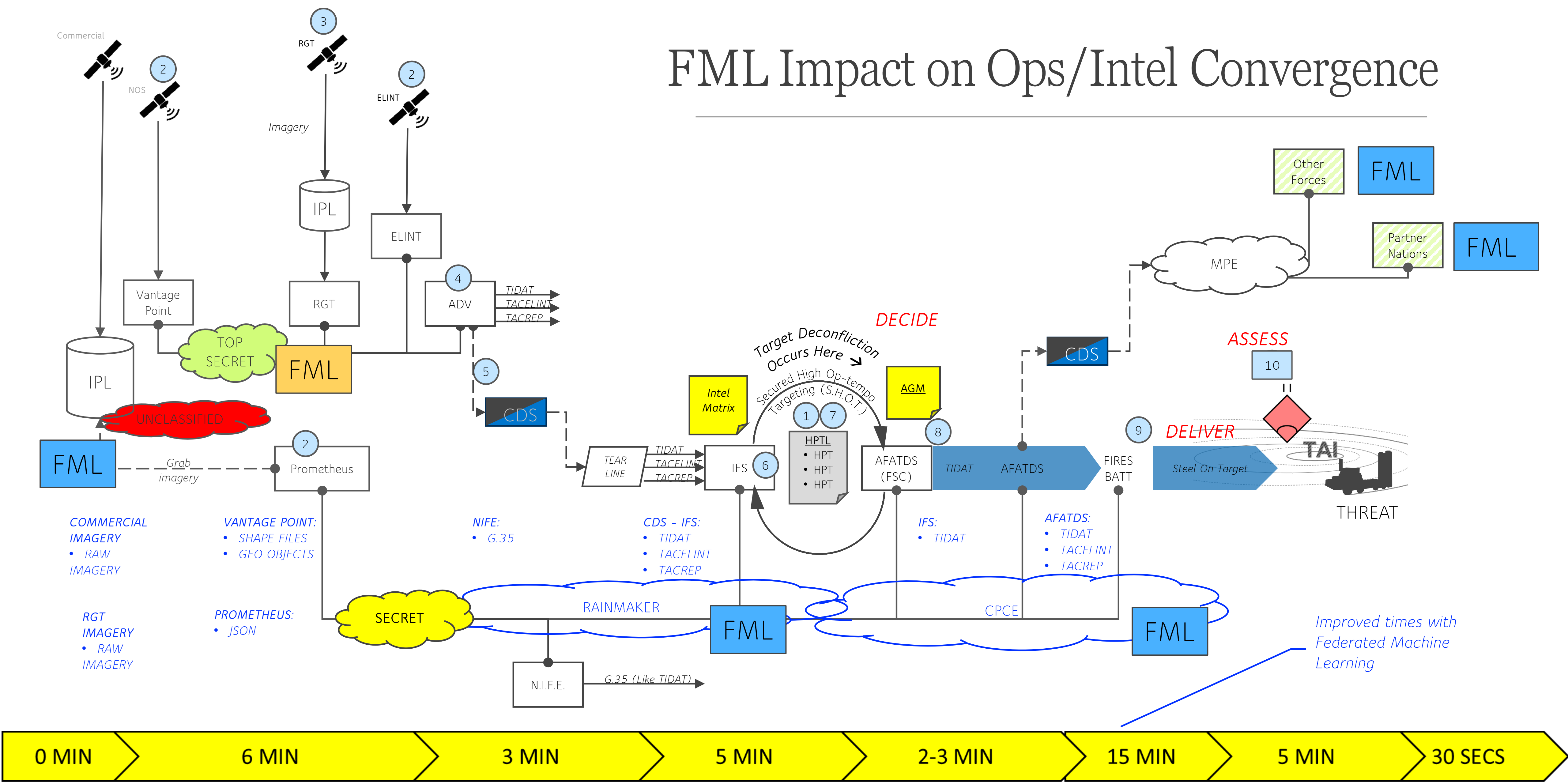
Federated metrics within 5% of Centralized metrics in both models
Volume of data transferred in Federated training over 99% smaller than in Centralized training

	Model 1.0 Random Forest: Top 5 Features*			Model 2.0 Random Forest: Top 20 Features*			Reference Study**
	Centralized	Federated	% Diff.	Centralized	Federated	% Diff.	
Average Precision	0.19	0.19	0.0%	0.21	0.20	-4.8%	0.18
Area Under the ROC	0.59	0.60	1.7%	0.62	0.60	-3.2%	0.63
Bandwidth Use	~7.5GB	~2.8MB	-99.96%	~7.5GB	~2.9MB	-99.96%	--

*In terms of the feature importance metric from Random Forests

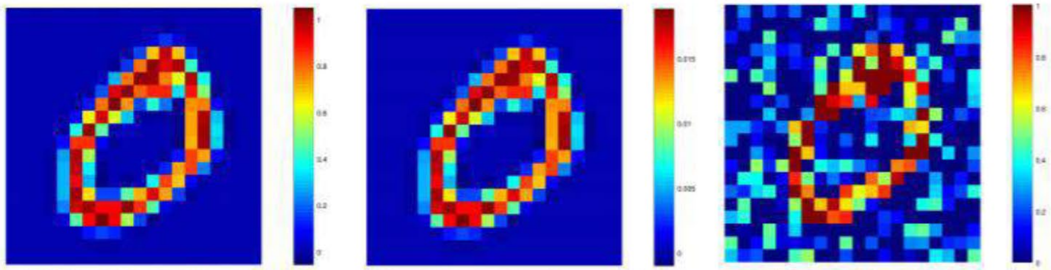
**Reference results from centralized model developed internally by the customer. The Dell team's target was to reproduce the customer's methodology in a centralized way and compare with federated simulations. Due to differences in underlying datasets used for training, the results from our centralized models differ from the reference study results, despite following the documented methodology.

FML Impact on Ops/Intel Convergence



DELL Technologies

Data Poisoning, Leakage



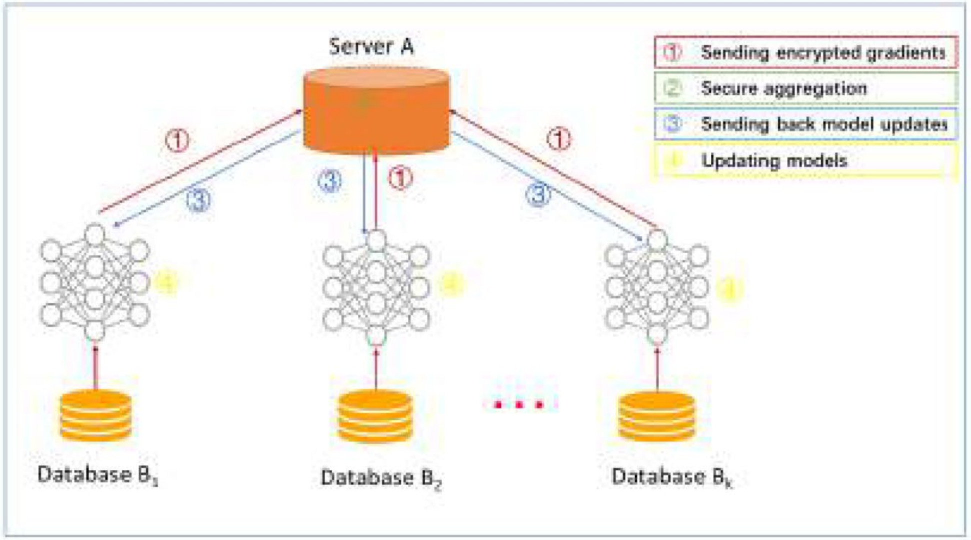
(a) Original 20x20 image of hand-written number 0, seen as a vector over $\mathbb{R}^{400}$ fed to a neural network.

(b) Recovered image using 400/10285 (3.89%) gradients (see Sect.3, Example 2). The difference with the original (a) is only at the value bar.

(c) Recovered image using 400/10285 (3.89%) gradients (see Sect.3, Example 3). There are noises but the truth label 0 can still be seen.

Fig. 3. Original data (a) vs. leakage information (b), (c) from a small part of gradients in a neural network.

Protect gradients with Homomorphic Encryption



- Le Trieu Phong, et al. 2018. Privacy-Preserving Deep Learning via Additively Homomorphic Encryption. IEEE Trans. Information Forensics and Security , 13, 5 (2018),1333–1345

- **Algorithm ensures that no information is leaked to the semi-honest server, provided that the underlying additively homomorphic encryption scheme is secure*.**

Capabilities Overview

Project Management

FedAI Studio JupyterLab NoCode

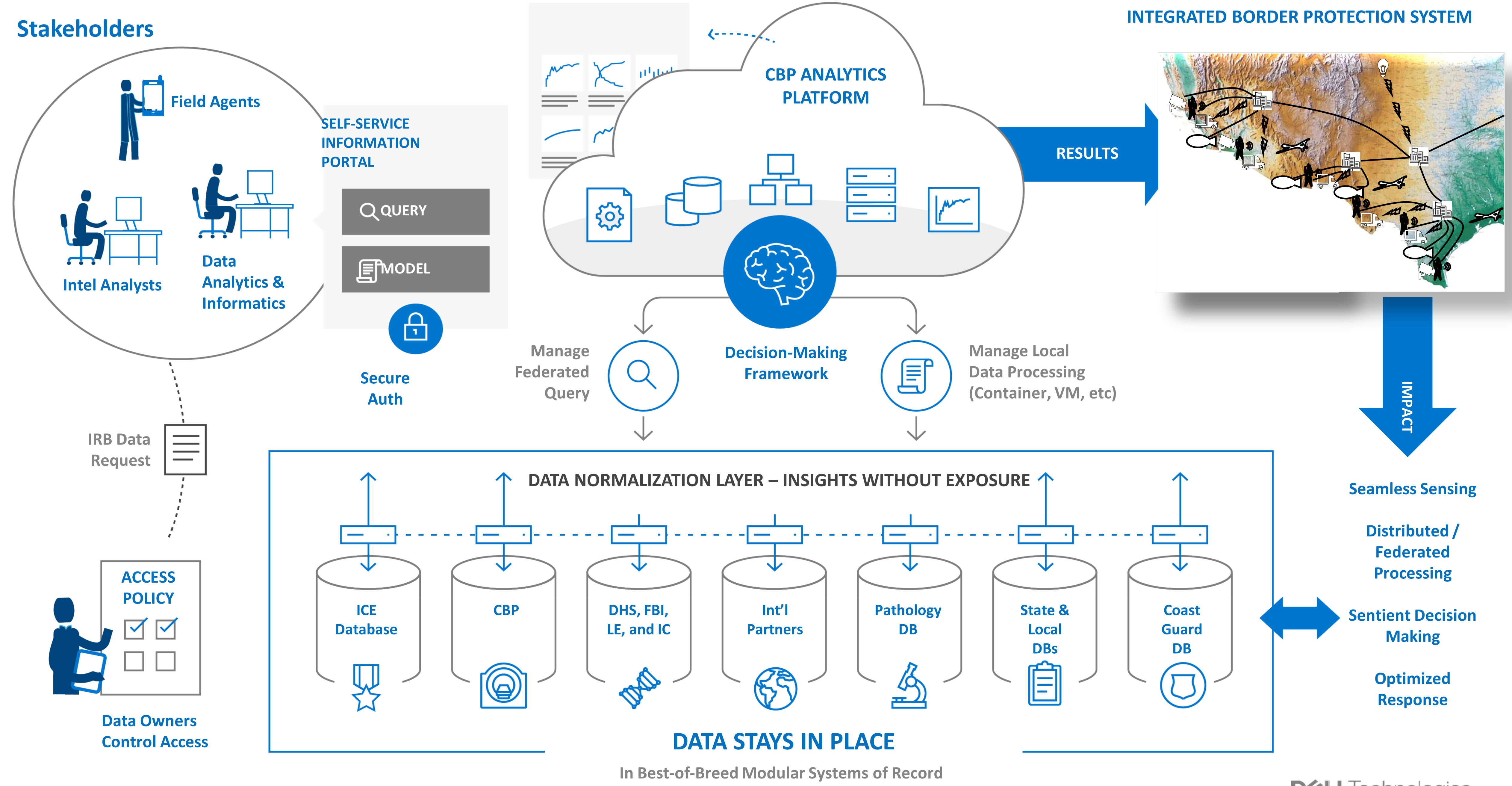
Data Integration | AI/ML API | MLOps

Federation (Vertical, Horizontal, Transfer)

Operations and Security Services

kubernetes

aws Google Cloud Azure vmware Private Edge IoT



Dell Technologies

