databricks

Real-Time AI

Opportunities and Challenges

Sharon Richardson <sharon.richardson@databricks.com>
Data & AI Strategist, Databricks

15 June 2022

Databricks

The Data & AI company

Inventor and pioneer of the **data lakehouse** - cloud-based data analytics and AI on a single platform

Gartner recognized leader in both

- Database Management Systems
- Data Science and Machine Learning Platforms

Creator of popular OSS data projects: Delta Lake, Apache Spark, and MLflow

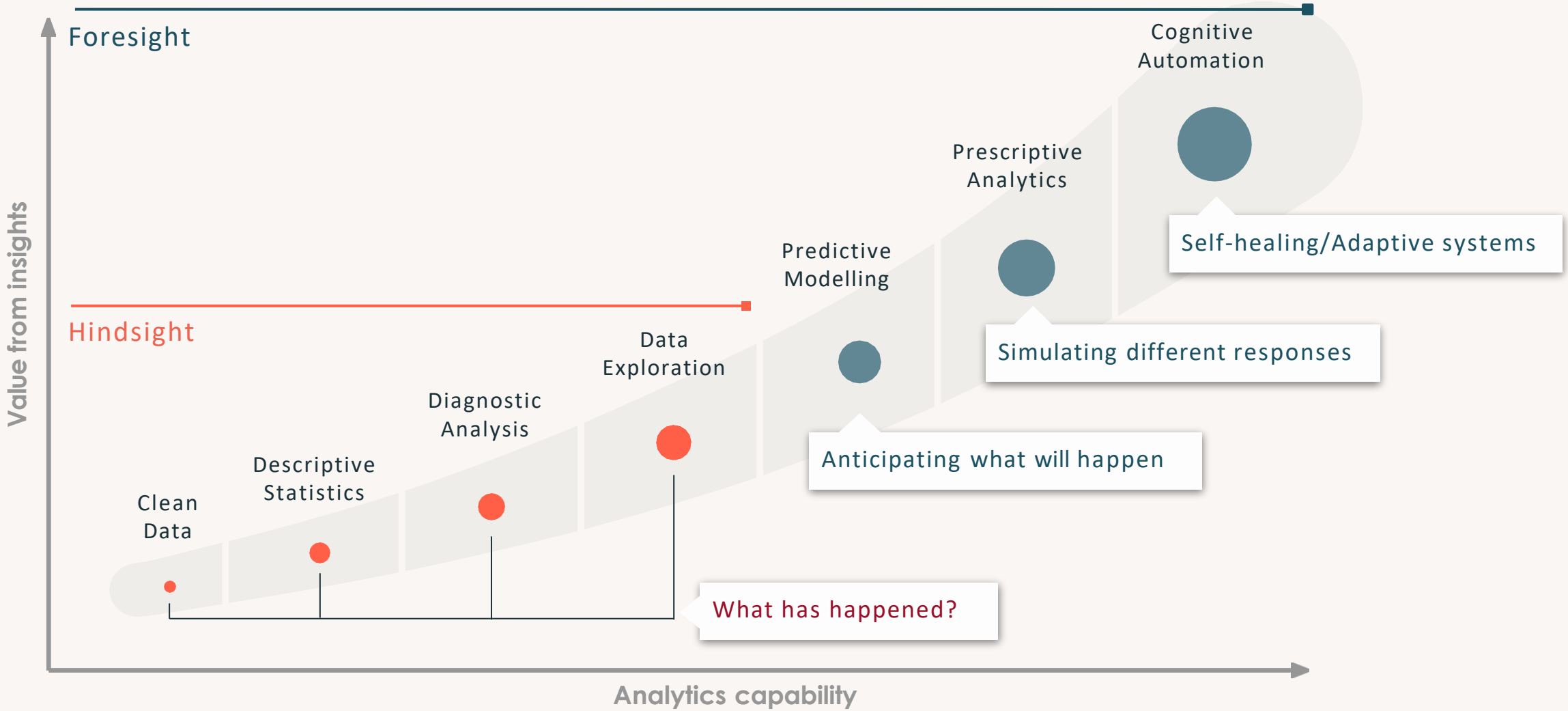
Raised over $3B in investment

3000+ employees across the globe

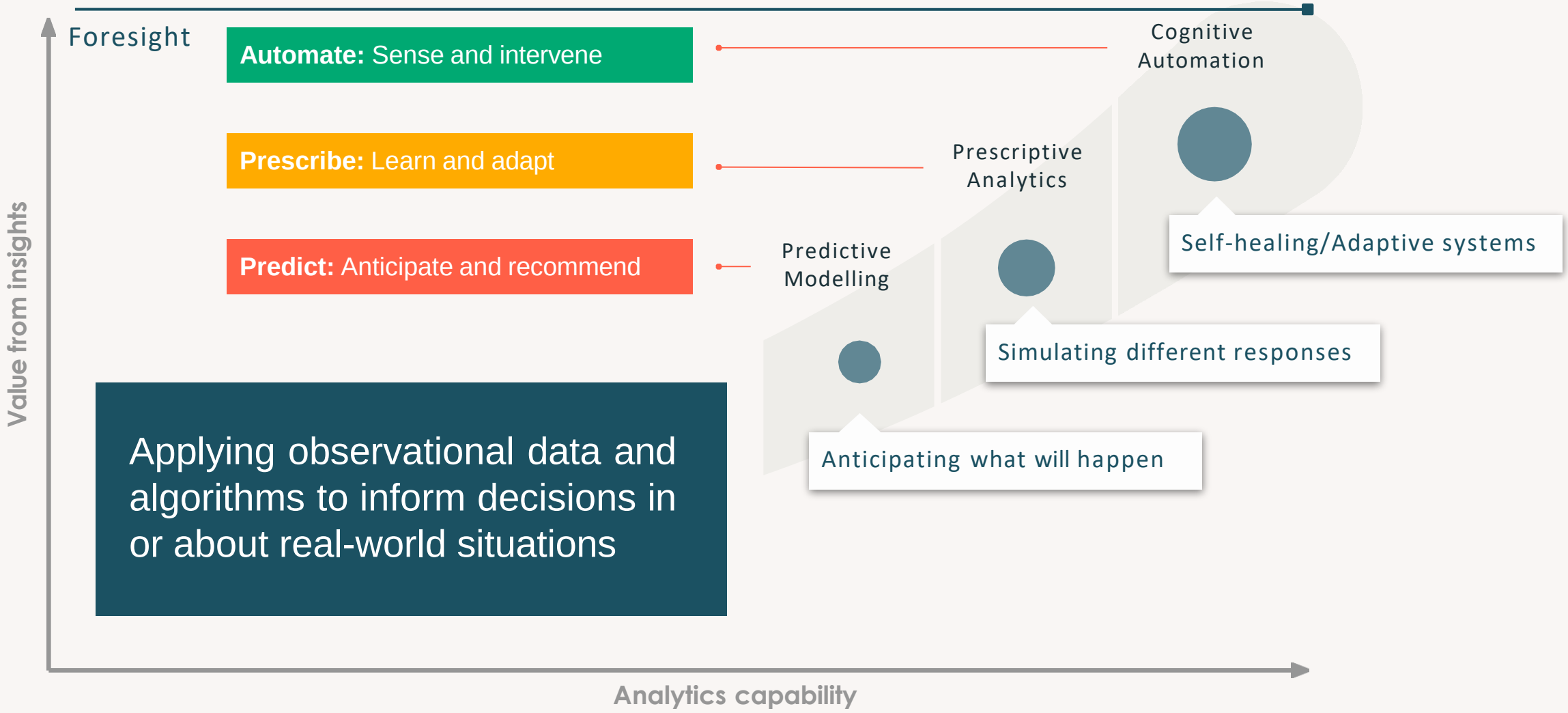
Global adoption

Over 6000 customers, from F500 to unicorns

Data & AI Maturity Curve



The Potential of Real-Time AI



Applying observational data and algorithms to inform decisions in or about real-world situations

Predict

Detect anomalies and anticipate failures before they are visible

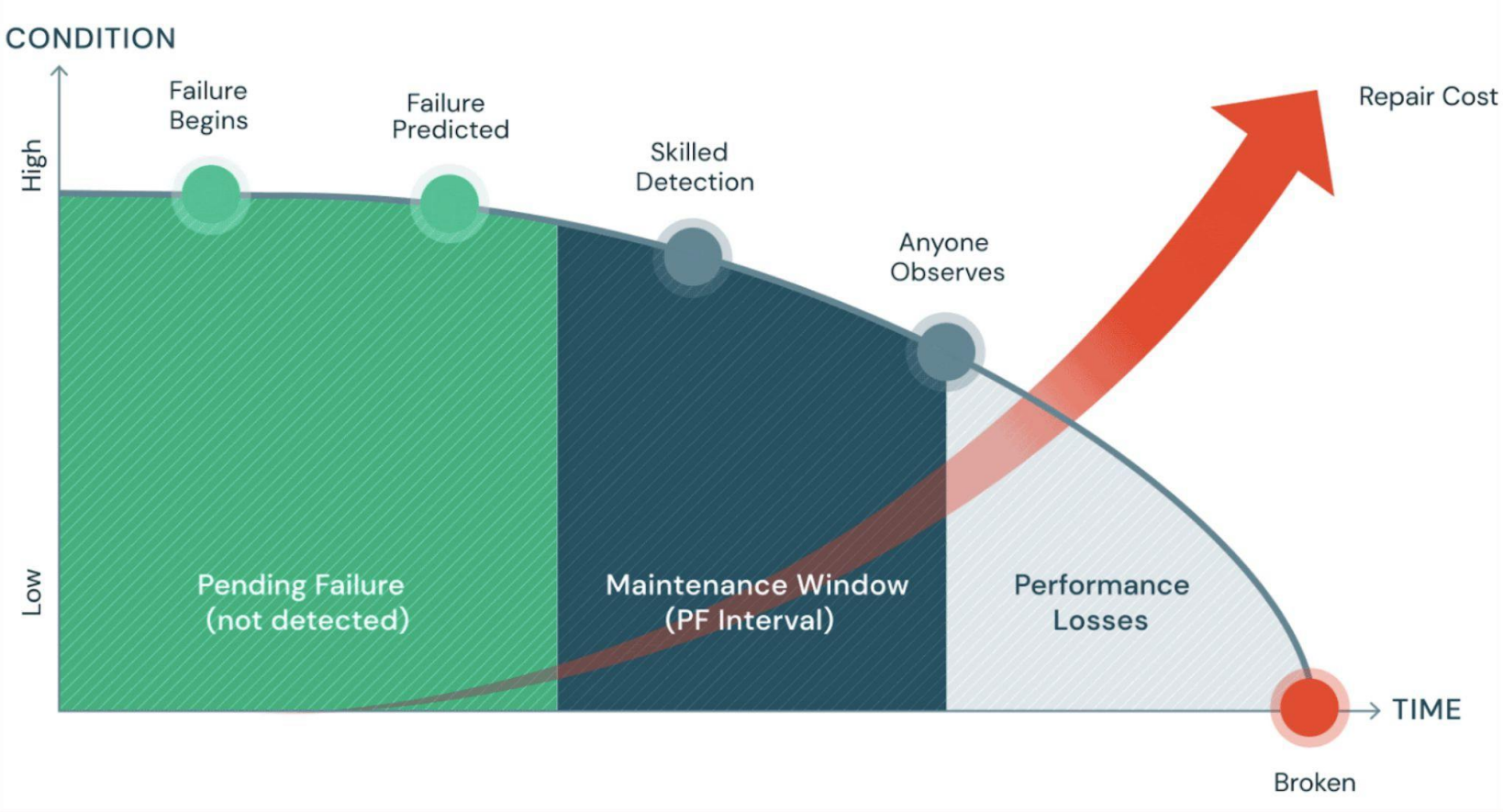


Image source: https://databricks.com/solutions#by-industry

Requires

- ML model built on training data and streaming IoT sensor

Scenarios

- Equipment maintenance
- Production monitoring

Benefits

- Reduce waste
- Reduce risk
- Increase quality

Prescribe

Evaluate and simulate based on real-world observations

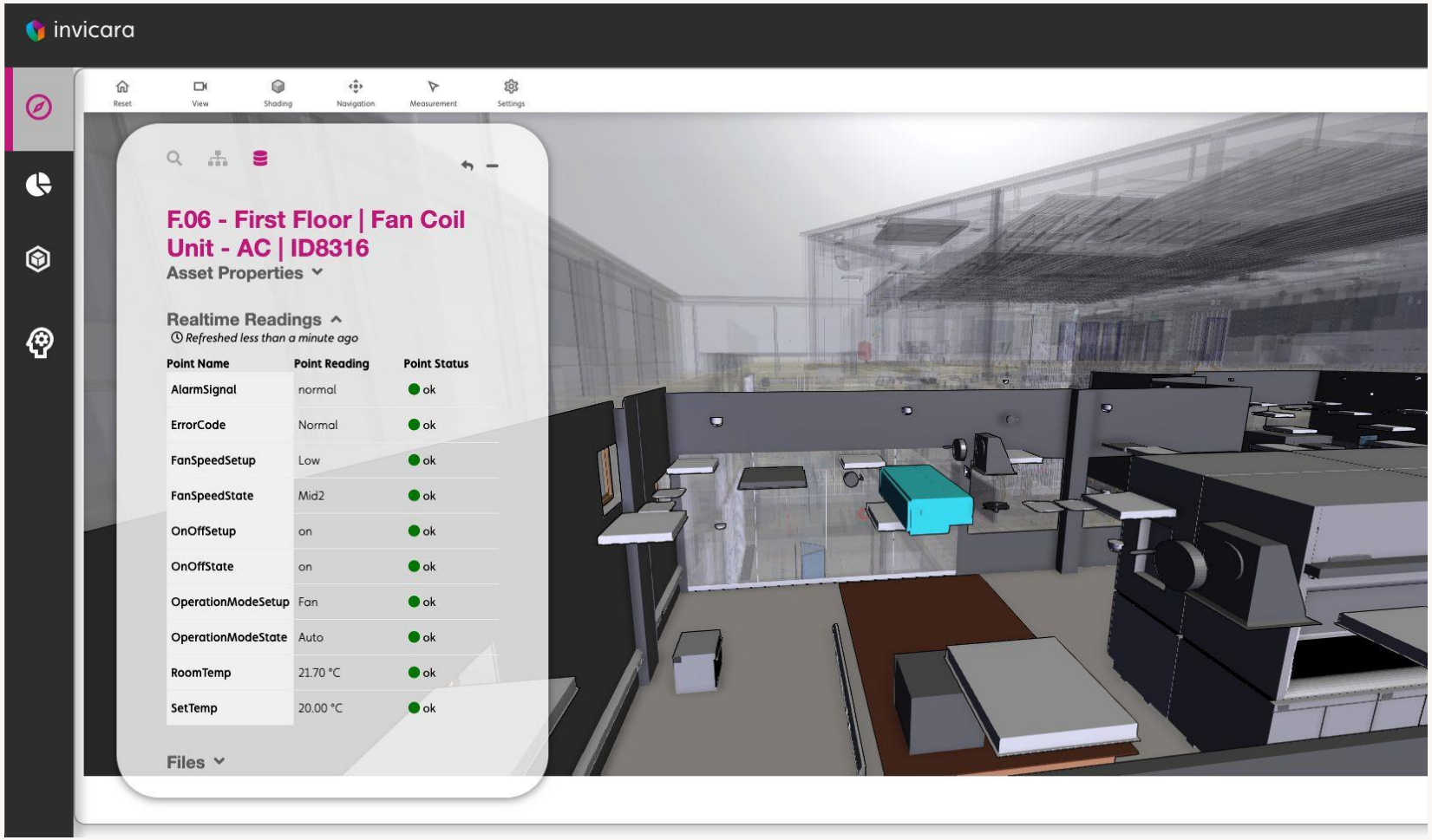


Image: courtesy of invicara https://invicara.com/

Requires

- Spatial data, IoT data and BIM tools

Scenarios

- Digital twins: Holistic systems monitoring
- Low-cost prototyping
- Complex construction sequencing

Benefits

- Optimise resources
- Innovate without material waste

Automate

Adapt to uncertain and changeable conditions

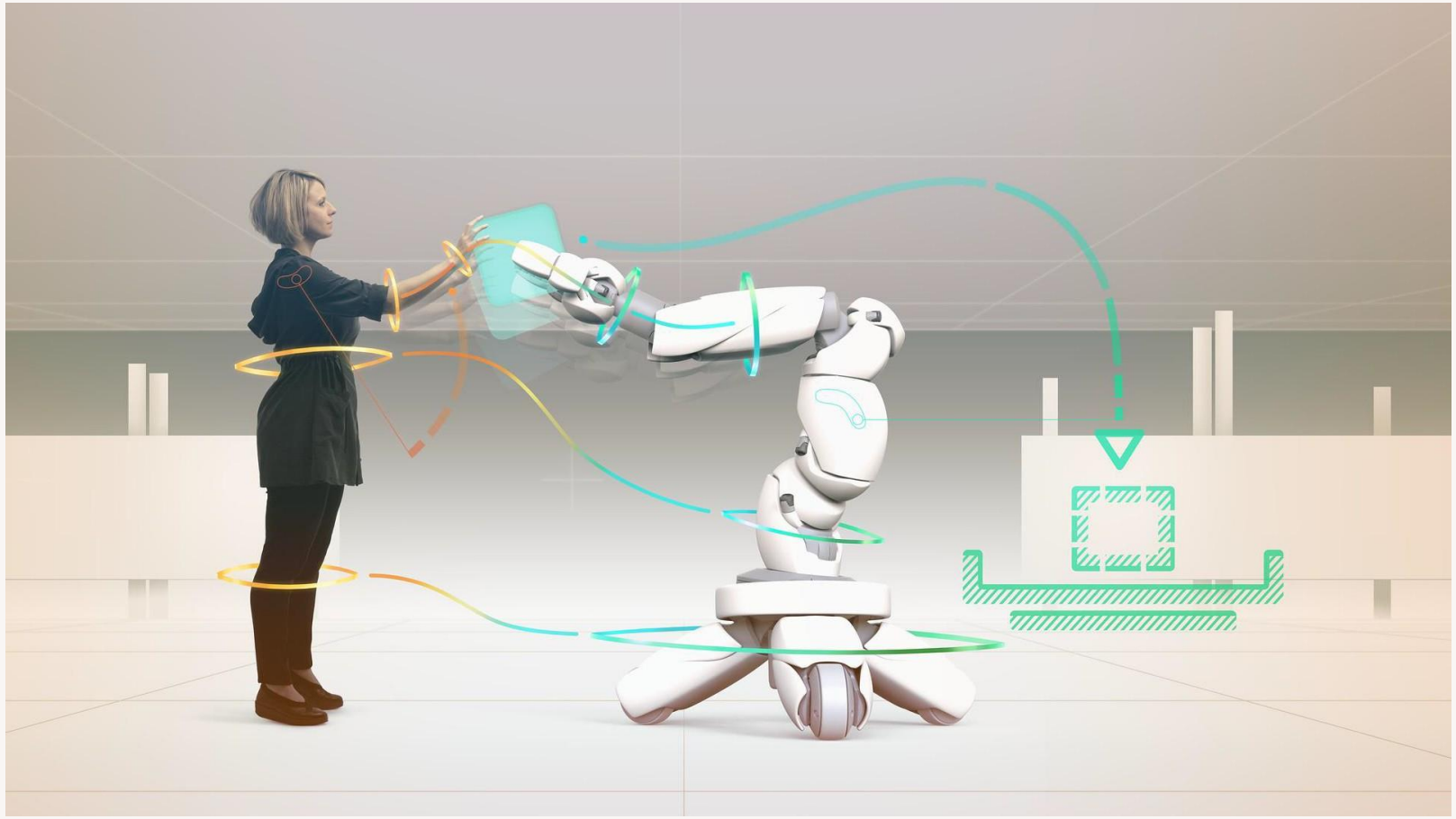


Image: https://flic.kr/p/2h7GZ5j (CC BY-NC-ND 2.0)

Requires

- Pervasive sensors and real-time analytics

Scenarios

- Intelligent supply chain
- Industrial athletes
- Cobots

Benefits

- Accelerate processes
- Increase productivity
- Increase safety

AI trends

Accelerating breakthroughs

2014:

https://xkcd.com/1425 published 24 Sept 2014

2015:

‹ Return to Blog Home

Microsoft Research Blog

Microsoft Researchers' Algorithm Sets ImageNet Challenge Milestone

Published February 10, 2015

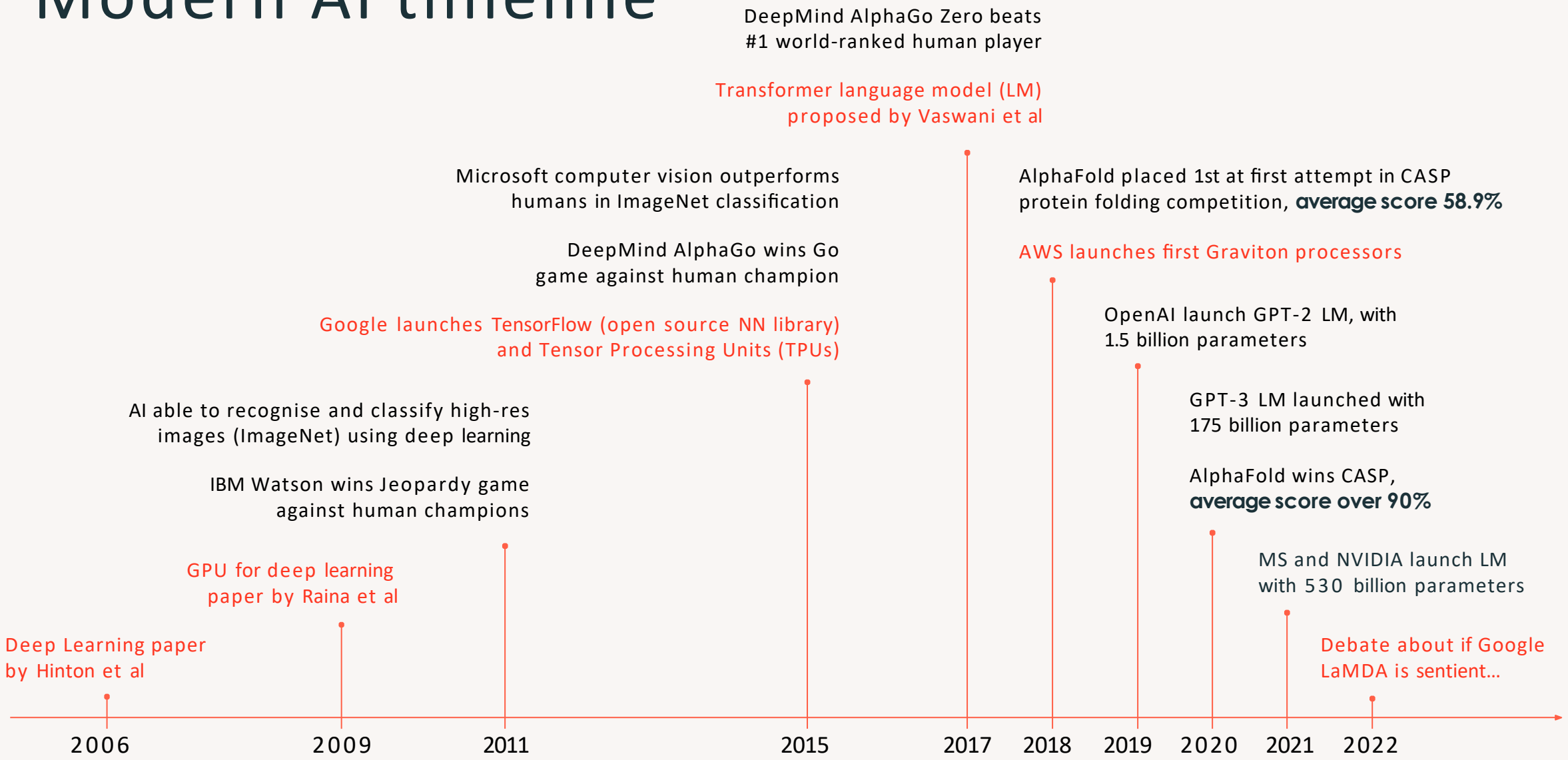
The race's new leader is a team of Microsoft researchers in Beijing, which this week published a paper in which they noted their computer vision system based on deep convolutional neural networks (CNNs) had for the first time eclipsed the abilities of people to classify objects defined in the ImageNet 1000 challenge.

In their paper, *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*, the researchers say their system achieved a 4.94 percent error rate on the 1000-class ImageNet 2012 classification dataset, which contains about 1.2 million training images, 50,000 validation images, and 100,000 test images. In previous experiments, humans have achieved an estimated *5.1 percent error rate*.

"To our knowledge, our result is the first to surpass human-level performance...on this visual recognition challenge," the researchers wrote.

https://www.microsoft.com/en-us/research/blog/microsoft-researchers-algorithm-sets-imagenet-challenge-milestone/ published 10 February 2015

Modern AI timeline



Four factors driving the AI economy

1. Data

‘Big Data’ (aka real-world evidence) underpinning the training of modern AI

Data becoming a demand factor for AI - using AI to make sense of the data

2. Processing power and the cloud

Hardware tailored for AI: microprocessors, graphics processing units (GPUs)

Cloud platforms providing infinitely scalable storage and elastic processing power

3. New algorithm types

Mathematical approaches to improve explainability and causality

Computational discoveries enabling human and super-human capabilities

4. Socio-political factors

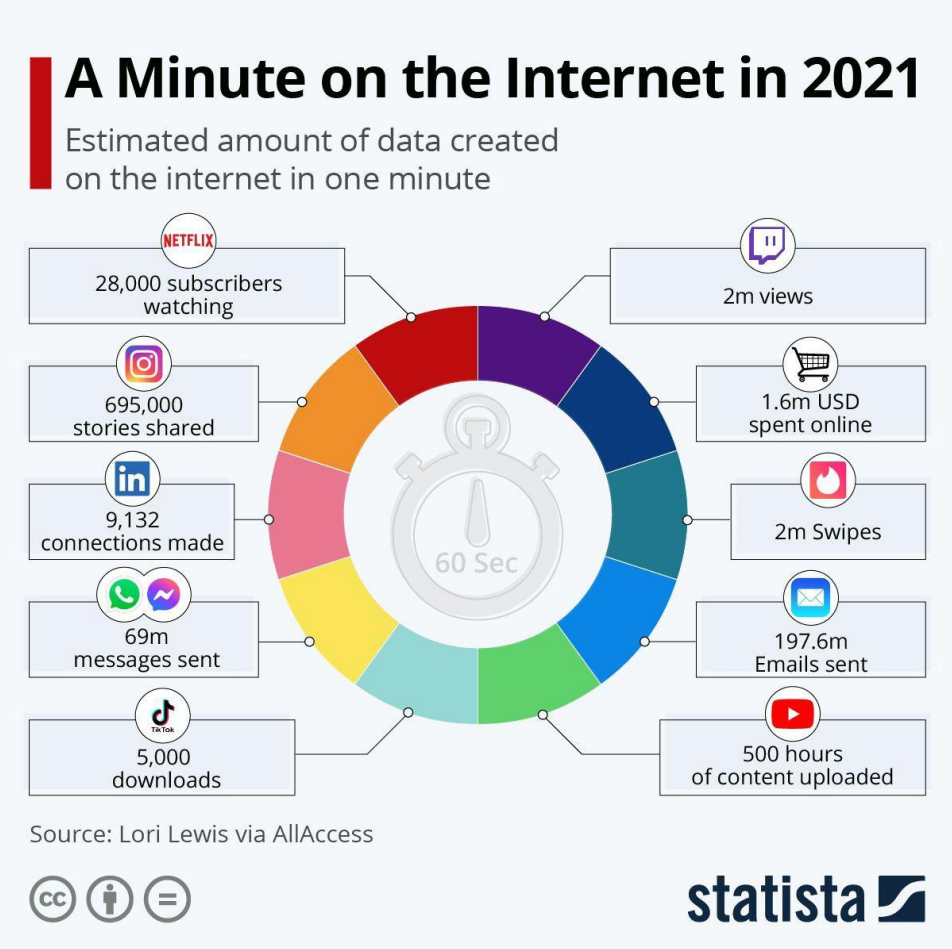
Geo-political differences in approaches to data privacy, and funding and controlling AI research

Social and cultural expectations for ethical, responsible and sustainable AI

Based on https://www.weforum.org/reports/empowering-ai-leadership-ai-c-suite-toolkit

‘Big Data’ = real-world evidence

Who owns the data?



- **Volume**: massive scale, infinite growth
- **Velocity**: rate of capture, propagation (and decay)
- **Variety**: diverse range of types and formats
- **Veracity**: level of trust in authenticity, accuracy
- **Virtue**: responsible, ethical, representative, fair?
- **Value**: is it useful? For how long?

Using AI to make sense of the data

AutoML - rapid, simplified machine learning

Quick-start initiatives

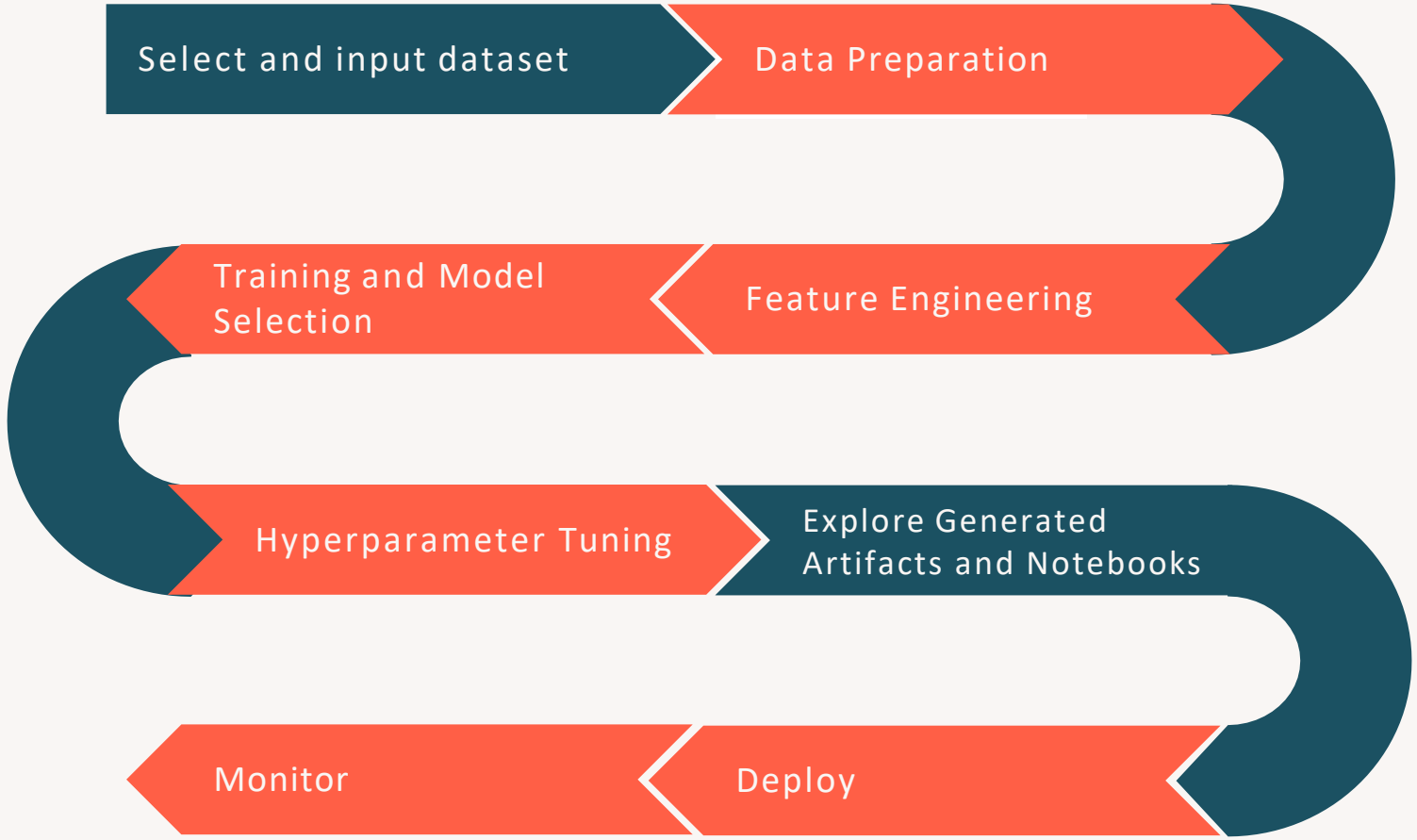
Accelerate time to production, Save weeks on ML projects

Auto-generated notebooks

Customize baseline models with domain expertise.

Wide range of problems

Solve classification, regression, and forecasting problems



Processing power and the cloud

Infinite scalability in storage and elasticity in compute

Open standards and protocols for interoperability across all use cases

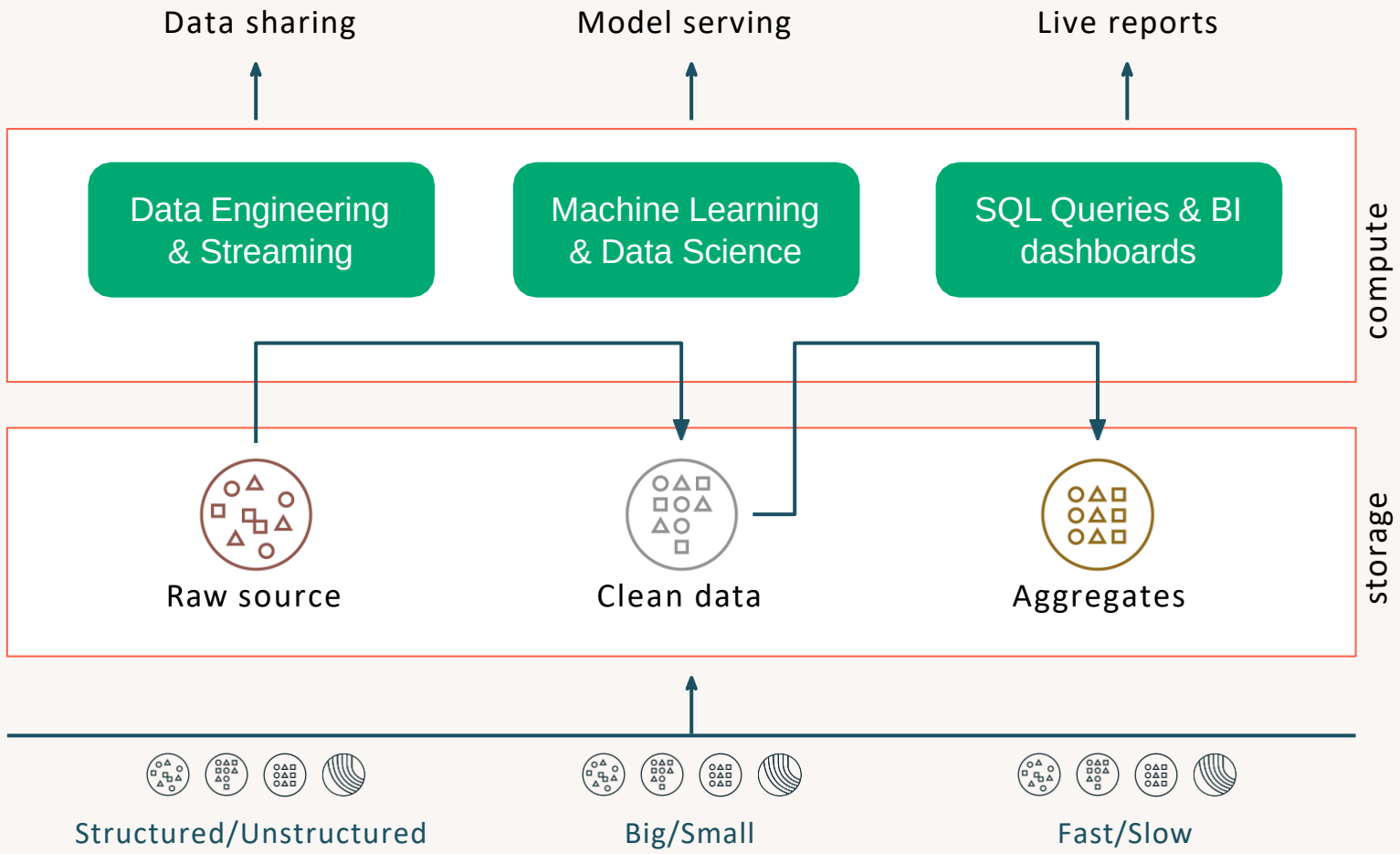
Elastic, auto-scaling compute

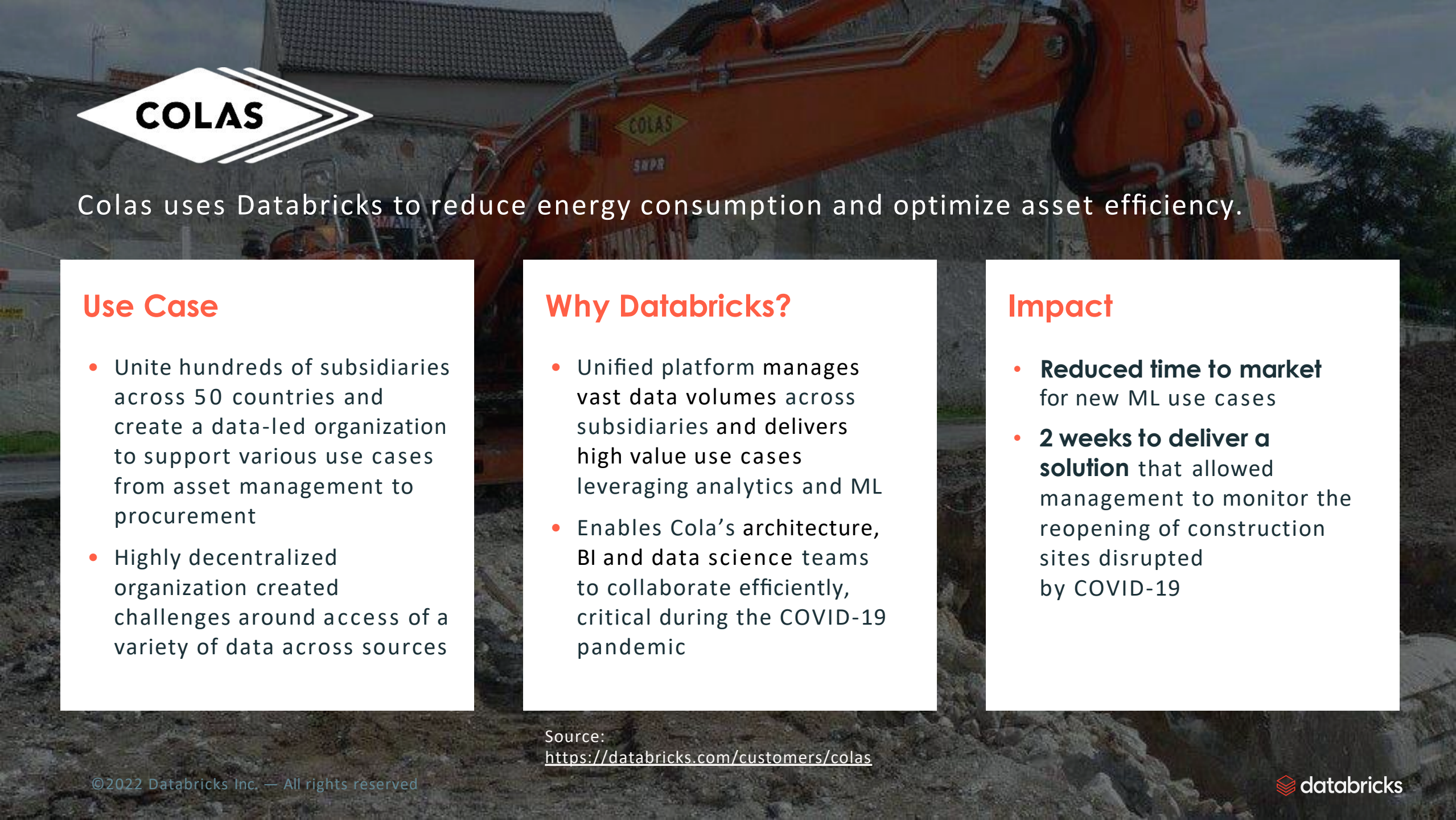
Unified collaborative environment serving all data personas

Low-cost scalable storage

Single source of truth

Microsoft Azure aws Google Cloud





COLAS

Colas uses Databricks to reduce energy consumption and optimize asset efficiency.

Use Case

- Unite hundreds of subsidiaries across 50 countries and create a data-led organization to support various use cases from asset management to procurement
- Highly decentralized organization created challenges around access of a variety of data across sources

Why Databricks?

- Unified platform manages vast data volumes across subsidiaries and delivers high value use cases leveraging analytics and ML
- Enables Cola's architecture, BI and data science teams to collaborate efficiently, critical during the COVID-19 pandemic

Impact

- **Reduced time to market** for new ML use cases
- **2 weeks to deliver a solution** that allowed management to monitor the reopening of construction sites disrupted by COVID-19

Source: https://databricks.com/customers/colas

databricks

Socio-political factors

Challenges applying AI in real-world situations

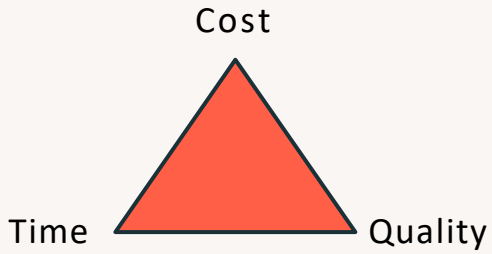
- Compliance with regulations
- Ethical and responsible AI
- Representative and FAIR data
- Privacy and trust issues
- Audit trail from data to decision
- Regional differences in data handling
- Inertia in business processes

Innovations in data and ML operations will become as important as data capture and model building

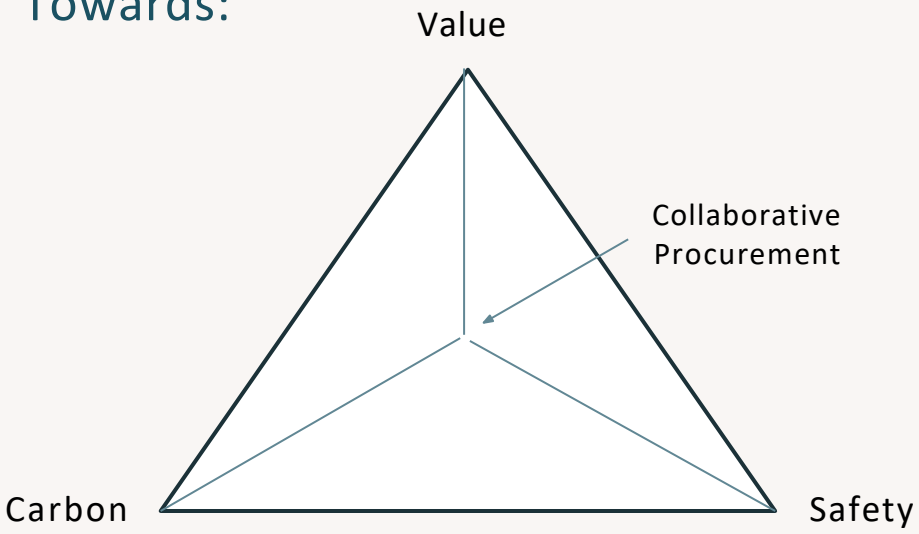
Real-time AI and Build Digital

Rialtas na hÉireann Government of Ireland | Tionscadal Éireann Project Ireland 2040

From:



Towards:



	Predict	Prescribe	Automate
Value	Increase efficiency	Increase quality	Increase productivity
Carbon	Optimise use of resources	Digital prototyping & simulation	Reduce friction from processes
Safety	Anticipate when reach threshold	Non-invasive testing of thresholds	Rapid and proactive intervention

https://databricks.com/dataaisummit/

DATA+AI SUMMIT 2022

HYBRID | JUNE 27-30, 2022

Building the modern data stack on the data lakehouse

The world's largest data and AI conference returns live, to San Francisco and virtually in our new hybrid format. Four days packed with keynotes by industry visionaries, technical sessions, hands-on training and networking opportunities.

- Hear from data and AI thought leaders about latest trends and innovations, including Apache Spark, Delta Lake, MLflow, TensorFlow, dbt and many others.
- See how others are applying the data lakehouse paradigm to unify data, analytics, and AI on one platform
- Take advantage of a full range of hands-on training workshops

Organized by databricks

Register Now

Thank you!

Sharon Richardson <sharon.richardson@databricks.com>