Artificial Intelligence & Flash Memory

Steven Eliuk, PhD
Vice President of Deep Learning
IBM Global Chief Data Office
Aug 6, 2019

There is no AI without IA

(Information Architecture)

80%

of data is either inaccessible, untrusted or unanalyzed

90%

say improving the use of data is a top priority

FORRESTER

> ***"No amount of AI algorithmic sophistication will overcome a lack of data [architecture] … bad data is simply paralyzing***

MITSloan Management Review

The AI Ladder

A prescriptive approach to accelerating your journey to AI

INFUSE – Automate and scale across your processes

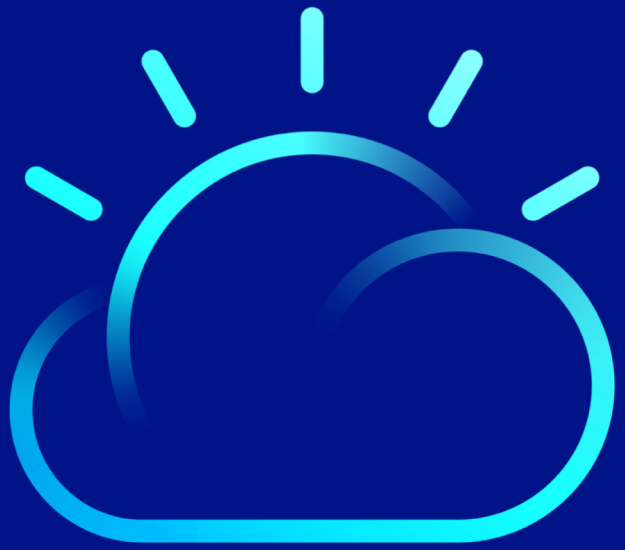
TRUST – Achieve trust and transparency in outcomes

ANALYZE – Scale insights with Machine Learning everywhere

ORGANIZE – Create a trusted analytics foundation

COLLECT – Make data simple and accessible

Data of every type, regardless of where it lives



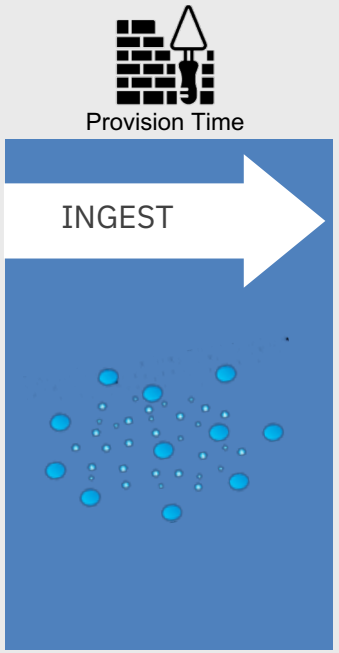
MODERNIZE
your data estate for an AI and multicloud world

Data Management Challenges in Analytics and AI

- Data ingest and preparation cycle are too time consuming
- Multi-source data aggregation & deduplication
- Silos of infrastructure for various analytics use cases
- Multiple copies of same data without a single source of truth
- Analytics on stale data
- Need to securely manage and protect data for traceability
- Need for **centralized** governance, global accessibility and collaboration

Adopting and Expanding AI

A Single View of the Truth



Collect and normalize multiple data sources
- Global requirements: IoT, Mobile, Sensors
- Client data
- Supply Chain Data
- Transactional Systems
- Client Behavior

Standard data analytics tools extract relevant data
- ETL: Extract, Transformation and Load
- Spark for real-time analytics
- Scripted, repeatable, reliable and *fast*

Storage Best Practices

- Optimize an extensible data repository that will grow with low Total cost of ownership
- Standard protocols for universal connectivity
- Support for structured and unstructured data

IBM Chief Data Office

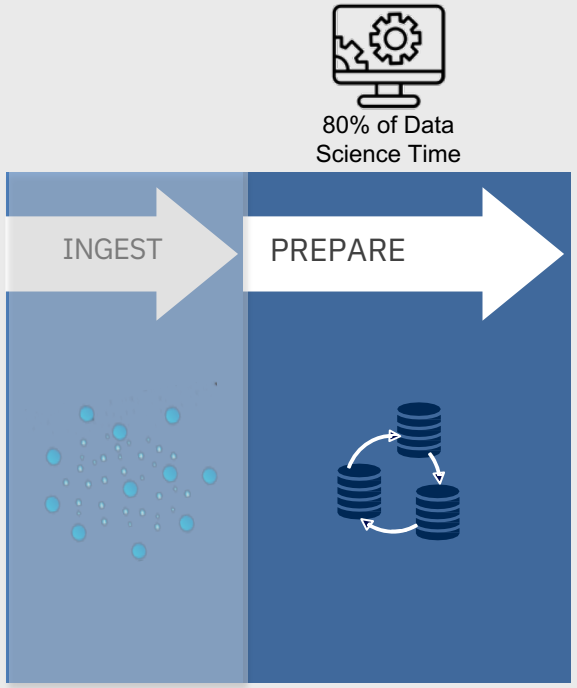
4 billion client data **records**

across **19** internal & external **data sources** refreshed weekly.

Serving ~**50k+ employees** in 60 countries @ Q219

Adopting and Expanding AI

Classify and Prepare Data Sets



Training and Testing Data Sets
- Accuracy improves with volume of data
- Track data sets to identify bias, create related models

Make data available for other analytics
- The Data Science Toolkit includes: Hadoop, SPSS, SAS, R, etc.

Metadata Generation
- Quality improvement and regulatory compliance requires data tracking
- Use Storage and AI to automate metadata

Storage Best Practices

- Support broad analytics tools across data and metadata
- High-Speed I/O to run multiple experiments
- Tag and track data with metadata

IBM Chief Data Office
Automated Metadata Generation

90% reduction

in cycle time

Over 200k experiments used to classify TBs of data to make it discoverable

Adopting & Expanding AI

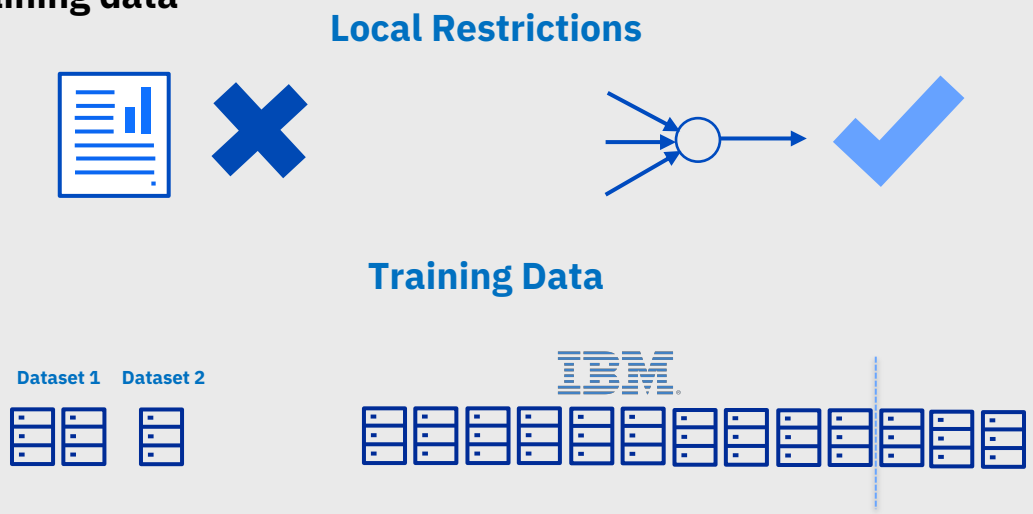
Automated Metadata Generation, a Practical Use-case:

Challenges:

- **Lack of data for model training impacts the performance**
- **Local restrictions related to processing of the business information within the limits of certain jurisdiction**

Solution, Distributed Federated Learning:

- **Compliance with local regulation**
- **Larger volume of training data allows to achieve better performance**
- **No isolated business units that lack training data**



AMG Goals

Metadata Tagging of Data Assets

1. Top-5 Recommended Terms, **5x less work**
2. Single Term, **10x less work**
3. Provide Correct Term, governance simplified, immediate ingestion, etc.

IBM Chief Data Office

Automated Metadata Generation

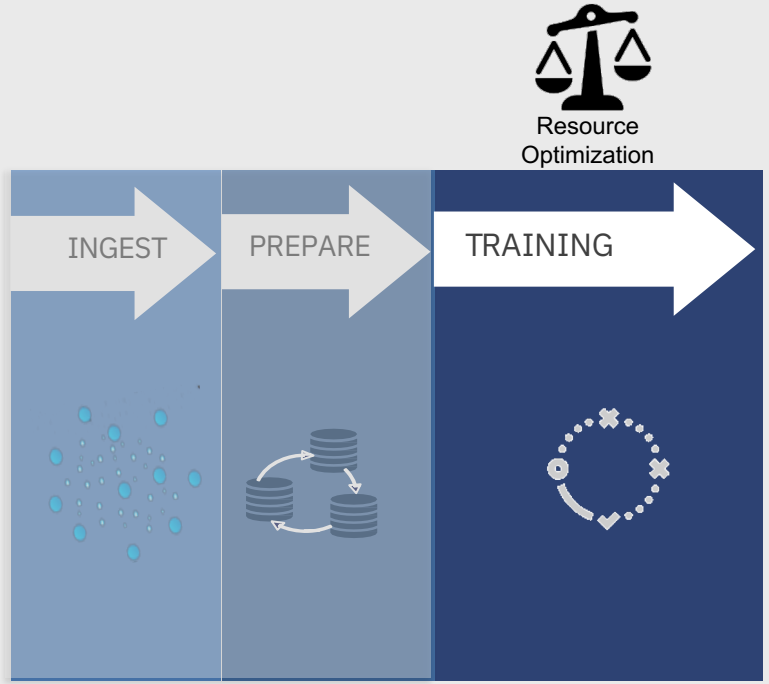
90% reduction

in cycle time

Over 200k experiments used to classify TBs of data to make it discoverable

Adopting and Expanding AI

Develop AI models



Increase Model Iterations

- Fastest copy possible to GPUs
- many models can be distributed across multiple GPUs
- Containerize models for tracking and efficiency

Leading Edge Shared Data Service

- Leading GPU can cost $10k
- GPUs : up to 150GB/s of throughput
- Shared Container Service

Storage Best Practices

- Highest throughput possible
- End-to-end Bandwidth to run multiple data models
- Local caching
- RDMA to eliminate transfer overhead

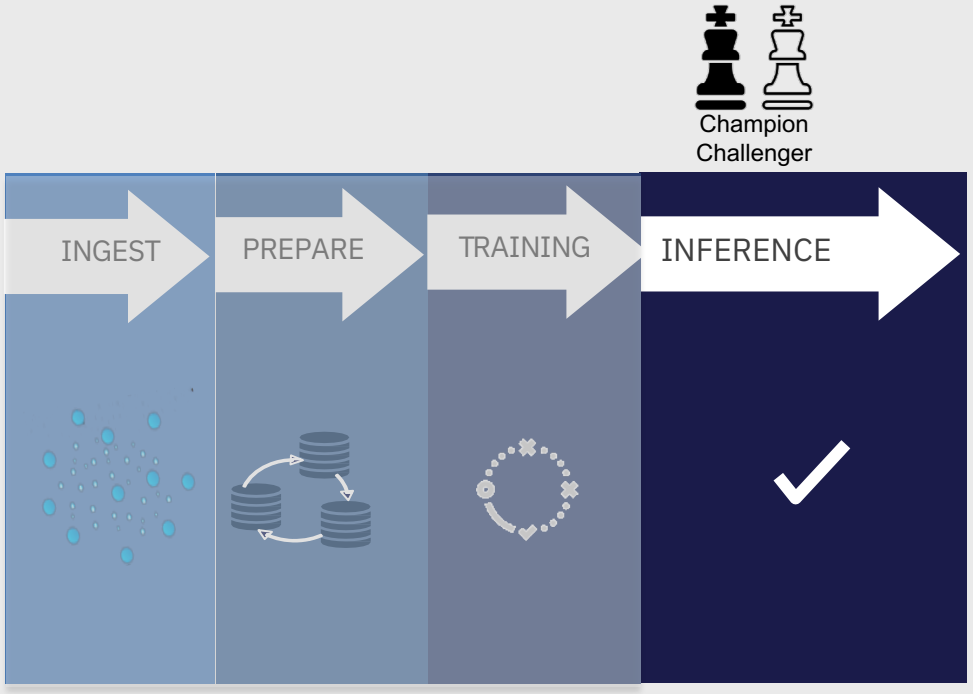
IBM Storage

Fast, Scalable Storage

2TB/s at CORAL
200GB/s in a single rack

Adopting and Expanding AI

Deploy AI models



Low Latency APIs
- Fast storage for fastest response

Scalable AI service
- AI is part of a portfolio of applications
- Containerized deployments to manage workloads

Storage Best Practices

- Low latency for fastest response
- Shared API service to scale as demand is needed
- Containerized deployment
- Deploy on-premises, in cloud, or as service

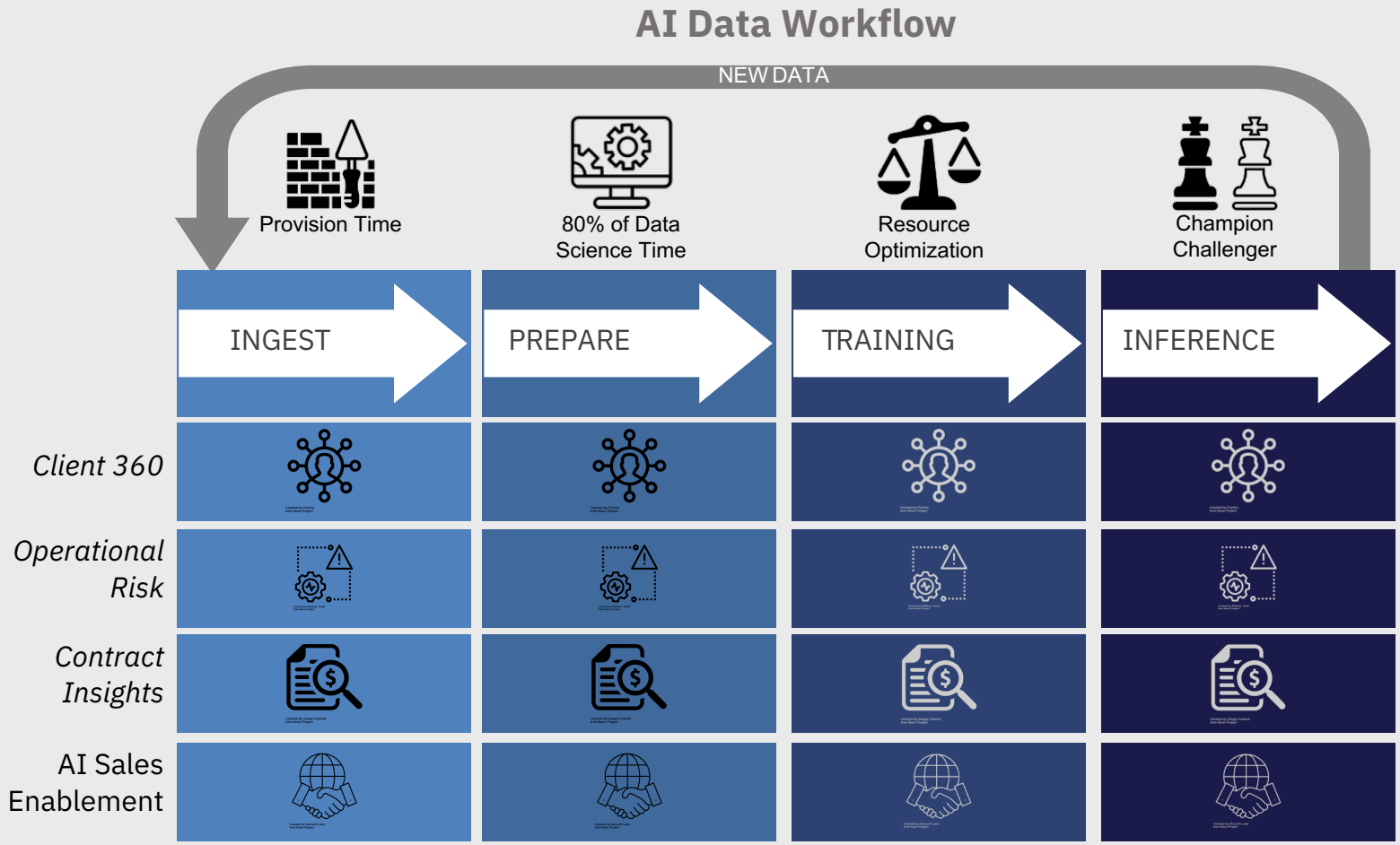
IBM Chef Data Office

Quick & easy.

<1 second
response time
40,000+ API calls/quarter

Presented in familiar user interface.

Building Enterprise AI



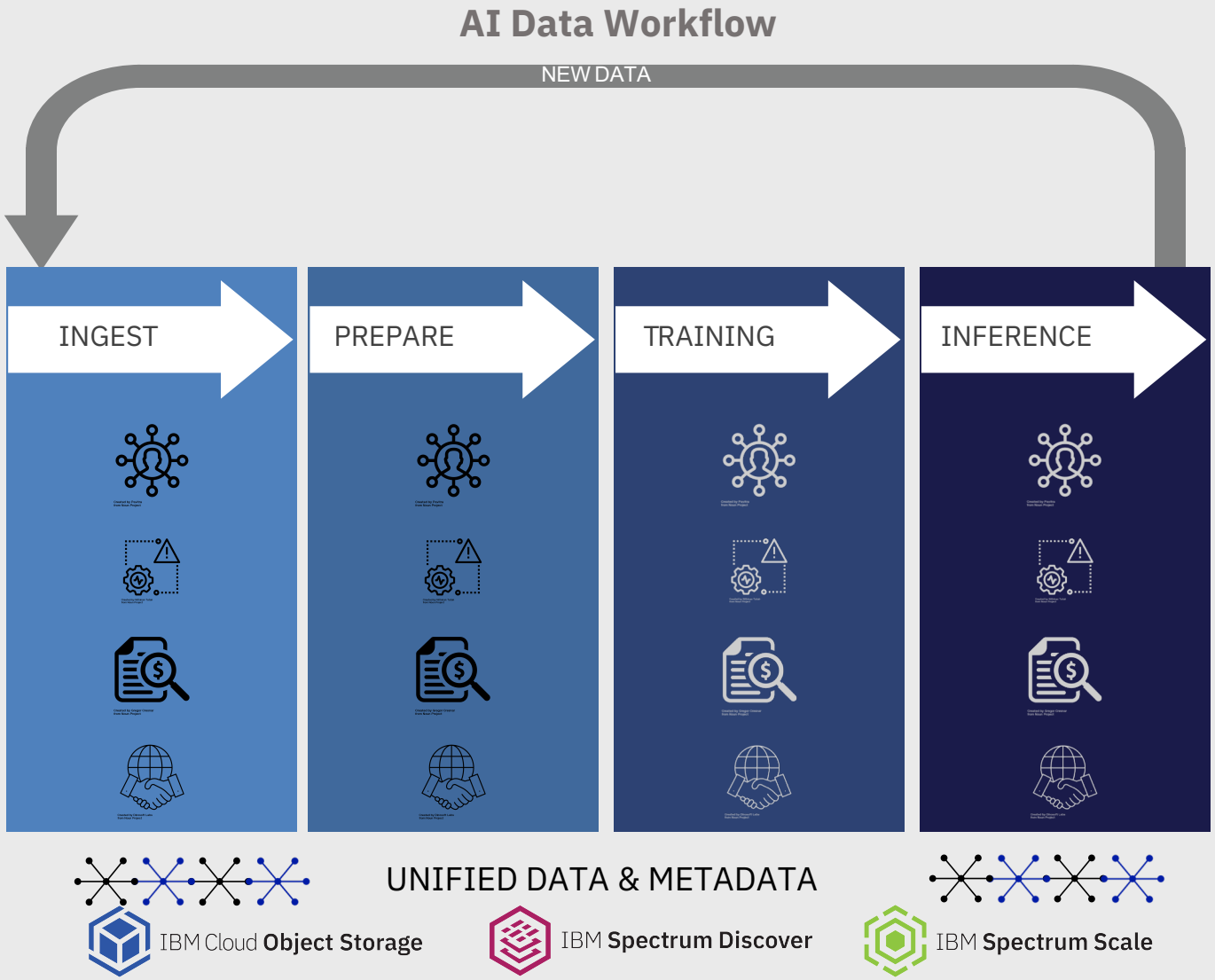
A Proof of Concept

- Easy to get started
- Can start with less data
- Demonstrate Business Value to

But Think Ahead…

- Share project overhead
- Avoid Data Silos
- Address Data Governance
- Common Platforms

Enterprise AI requires Shared Data



Common Enterprise Data Platform

- Single Source of Truth
- Unified Data Access
- Faster Cycle Times
- Data Network Effect accelerates AI Adoption
- Containerized, Cloud Native applications

Unique Storage Requirements

- Unified Access
  - Combine data science and analytics tools
- Performance
  - GPUs at > 100GB/s
  - Global ingest
- Smart Growth
  - Automate Tiering
  - Active Archives
- Governance
  - Extensible metadata
  - Enterprise data protection

Enterprise AI requires “Governed” data

- Data quality, huge issue
- Limited amount of training data
- Colocation is not possible for all data sources
- **Continuous Learning**
  - **Adapting models automatically to new data & changes in business requirements**
  - **Decreases potential governance risks by not storing any training data**
  - **Minimize technical debt**

Federated Open-source Data

- **Highly curated**
- **No Privacy issues… currently**
- **Gold standard train, validation, & test datasets**

Unique Storage Requirements by YE19.

- **Training, inference, production workloads all sharing *source data***
- **1k+ of data scientists**
  - **10k+ active experiments**
- **100k+ users**
  - **Production workloads**

Steven.Eliuk@ibm.com

IBM Global Chief Data Office

Family First