Machine Learning For Enterprise: Beyond Open Source

April 2018

Jean-François Puget
Use Cases for Machine/Deep Learning

- Cyber Defense
- Drug Discovery
- Fraud Detection
- Aeronautics
- IoT
- Medical Decision-Making
- Smart Cities
- Earth Monitoring
- Weather Forecasting
- Disease Diagnostics
- Advanced Physics Research
- Climate Change
- Robotics
- Supply Chain Management
- Media Analytics
- Smart Cities
- Disease Diagnostics
- Robotics
- Supply Chain Management
- Media Analytics
The Machine Learning Workflow: Perception
The Deep Learning Workflow: Perception
The AI Workflow:
Perception
The Big Data Workflow: Perception

Data ??? Big Data ??? $$
The XXX Workflow:
Reality

Data
• Historical
• Streaming

Ingest

Data
Processing
• Data visualization
• Feature transformation & engineering

Model Training
• Model selection & evaluation

Deploy
• Pipelines, not just models
• Versioning

Live System
• Predict given new data
• Monitoring & live evaluation

Creating examples

Automating DS work
Scalable deployment
Models lose accuracy

Feedback Loop

Data Scientists & Researchers

Data Engineers

Machine Learning & Production Engineers

Governance
Tanya  
**Domain Expert**

**Her Job:** To transfer knowledge to Watson for a successful user experience.

**What she does:**
- Range of domain knowledge and uses that to teach Watson and develop a custom models
- As Tanya gains more experience she optimizes her knowledge to teach Watson to design better end-user experiences.

**Sometimes known as:** Subject matter expert, content strategist.

---

Mike  
**Data Scientist**

**His Job:** Transform data into knowledge for solving business problems.

**What he does:**
- Runs experiments to build custom models that solve business problems.
- Use techniques such as Machine Learning or Deep Learning and works with Tanya to validate success of trained models.

**Sometimes known as:** ML/DL engineer, Modeler, Data Miner

---

Ed  
**Data Engineer**

**His Job:** Architects how data is organized and ensures operability

**What he does:**
- Builds data infrastructure and ETL pipelines. Works with Spark, Hadoop, and HDFS.
- Works with data scientist to transform research models into production quality systems.

**Sometimes known as:** Data infrastructure engineer

---

Deb  
**The Developer**

**Her Job:** Builds AI application that meet the requirements of the business.

**What she does:**
- Starts PoCs which includes gathering content, dialog building and model training
- Focus is on app building for the team or company to use. Will handle ML Ops as needed

**Sometimes known as:** Front-end, back-end, full stack, mobile or low-code developer

---

**Watson Studio**

Built for AI teams – enabling team productivity and collaboration
Enabling the end to end AI lifecycle

Enterprise Data & Apps
IBM Cloud Private for Data
IBM Power AI

ML Models, Metadata, Data

AI Business Processes
IBM Watson Studio
Watson Studio
Tools for supporting the end-to-end AI workflow

- Create, collaborate, deploy, and monitor
- Best of breed open source & IBM tools
- Code (R, Python or Scala) and no-code/visual modeling tools

- Most popular open source frameworks
- IBM best-in-class frameworks

- Fully managed service
- Container-based resource management
- Elastic pay as you go cpu/gpu power
Deploying Trained Models
Download your trained models or deploy your models within Watson Machine Learning
AI Asset Lifecycle Management
Use the Watson Knowledge Catalog and Watson Studio to manage your AI assets or manage them yourself

Watson Machine Learning

Watson Knowledge Catalog

Model Explanations
In May 2018, the General Data Protection Regulation (GDPR) takes effect and grants consumers the legal “right to explanation” from organizations that use algorithmic decision making.

Audit Trails
Tracking prediction to each model’s unique heritage is critical to regulatory compliance. Enforcing access controls for model sharing and deployment ensure ensures data security and application stability.
Watson Studio

Open Source tools – Jupyter and RStudio

Watson Visual Recognition – retrain Watson

Elastic and customizable compute environments

Create ML flows and design Neural Networks visually
Watson Machine Learning
Simplifying deployment & management of ML models in production apps

Train neural networks in parallel across NVIDIA GPUs.
Pay only for what you use. Auto-deallocation means no more remembering to shutdown your cloud training instances.

Deploy models into production then monitor them to evaluate performance.
Capture new data for continuous learning and retrain models so they continually adapt to changing conditions.

Monitor batch training experiments then compare cross-model performance without worrying about log transfers and scripts to visualize results. You focus on designing your neural networks. We'll manage and track your assets.

Python client, command line interface (CLI) or UI? You choose the tooling that best fits your existing workflows. Training history and assets are tracked then automatically transferred to the customer’s Object Storage for quick access.
AI solutions landscape

**pre-trained neural network**
- Application Developer

**transfer learning**
- SME
- domain data

**custom neural networks required**
- data scientist
How do the paths to AI-powered apps relate?

Pre-Trained Model → transfer learning model → Custom Model

SME

domain data

data scientist

Application Developer

Deploy to application
Neural Network Modeler (beta)
An intuitive drag-and-drop, no-code interface for designing neural network structures using the most popular deep learning frameworks. Quickly capture your network design then single click export for experimental optimization.

Drag-and-drop network layers
- Define layer configuration
- Choose optimizer params
- Generate CPU or GPU compatible code

Real-time validation of network flow
- Save as popular framework code
- Export as a python notebook
- Execute as batch experiment

Supported Frameworks
- TensorFlow
- PyTorch
- Caffe
- Keras
Discovering a single optimal neural network requires exploring 1000’s of hyperparameter combinations which means 100’s of experiments which requires 1000’s of training runs
Accelerate your experimental workflow

You focus on experiment design

We’ll manage the rest

EXPERIMENTS
10’s-100’s

TRAINING RUNS
100’s-10000’s

source code

experiment definition

dataset

NVIDIA GPU Acceleration

K80

P100

CPU-only

V100
Experiment Assistant supports the end-to-end workflow
Experiment Assistant

A suite of tools that manage your training runs. Each run is automatically started, monitored, and stopped upon completion. Training history and assets are tracked then automatically transferred to the customer’s Object Storage for quick access.

Initiate Experiments

CLI or Python Client

Monitor Training in Real-Time

Compare Model Performance
Experiment assistance manages and archives training progress

Monitor training progress and compare model performance in real-time

- Trained models + artifacts
- Cloud Object Storage
- Experiment Assistant
  - Monitor Training in Real-Time
  - Compare Model Performance
Distributed Learning w/Tensorflow

As datasets expand and model grow in complexity, training times increase from hour to days to weeks or longer to complete.

Models no longer fit on one or multiple GPUs within a single server and must be distributed across multiple machines.
What about massive datasets like ImageNet 22K?

- 14 million images
- 3 Terabytes
- 21,841 categories
Single servers are too slow for complicated networks and large datasets.

ImageNet 22K
14 million images
3 TB

Resnet101

4 NVIDIA P100 GPUs

16 days training time
Distributed training across multiple servers
Distributed training at scale

- **ImageNet 22K**: 14 million images, 3 TB
- **Training Time**: ~7 hours
- **ResNet 101**
- **Sync Parameters**
- **2xV100s**

Diagram shows a distributed training setup with multiple nodes synchronizing parameters.
Deploying Trained Models

Download your trained models or deploy your models within Watson Machine Learning
IBM Watson®

Get started today www.ibm.com/watson