

A background image of a city street at night, showing blurred light trails from buildings and streetlights, creating a sense of motion and speed.

# **ML Accelerator Productivity & Optimization Tools**

**Harry Parks / IBM Federal  
Servers, AI/ML/DL Solutions  
[harry.parks@ibm.com](mailto:harry.parks@ibm.com)**

# Problem: Data Scientist Productivity

Waiting on resources

Experimenting with models takes time

Experimenting with hyper parameters takes time

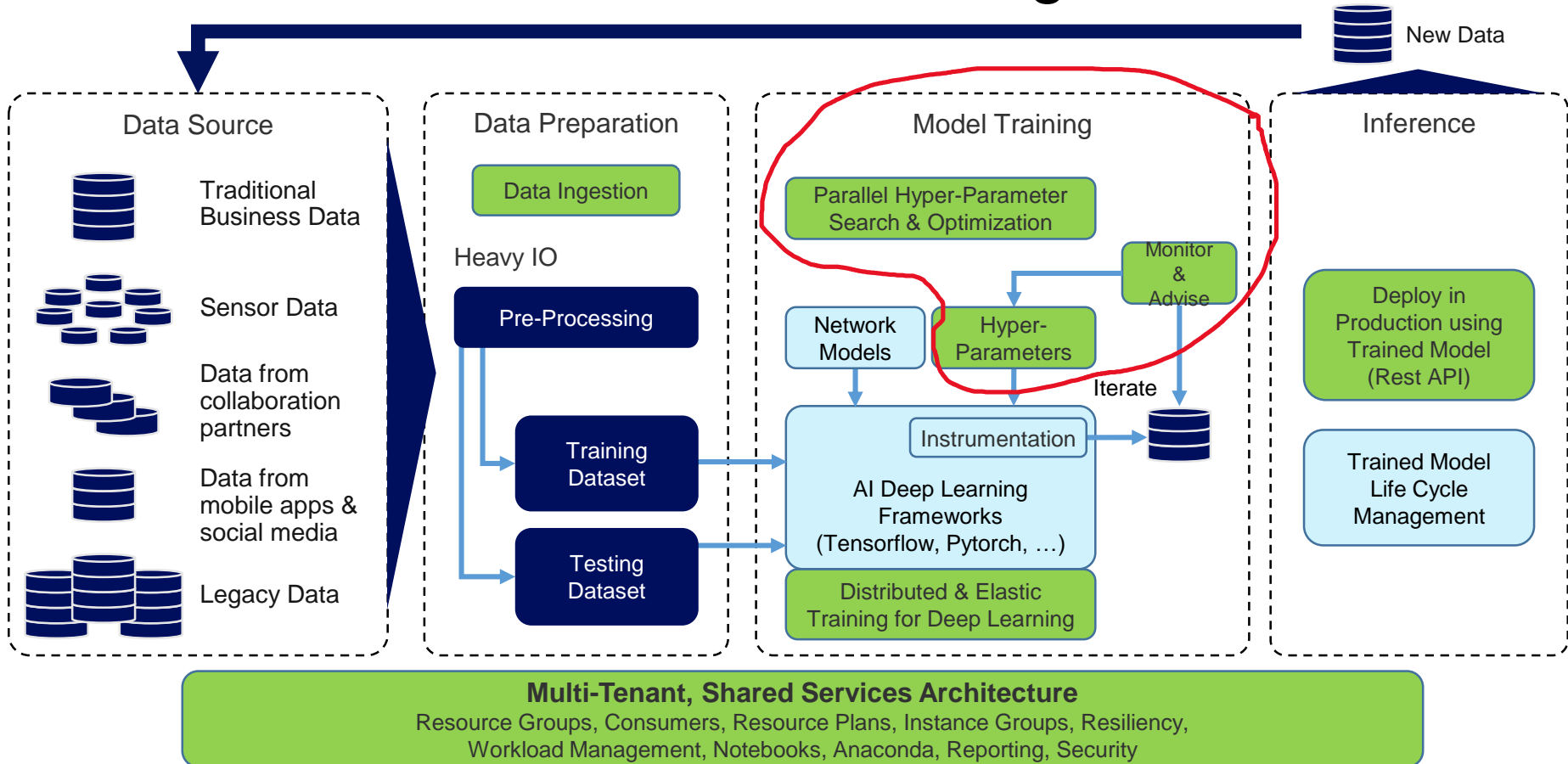
Waiting on execution

Iterating over and over to improve accuracy

Also...Domain Scientist Productivity – enabling users that don't have deep data science skills



# Basic Workflow for ML/DL Training & Execution



example #1

Auto Hyper-Parameter Tuning



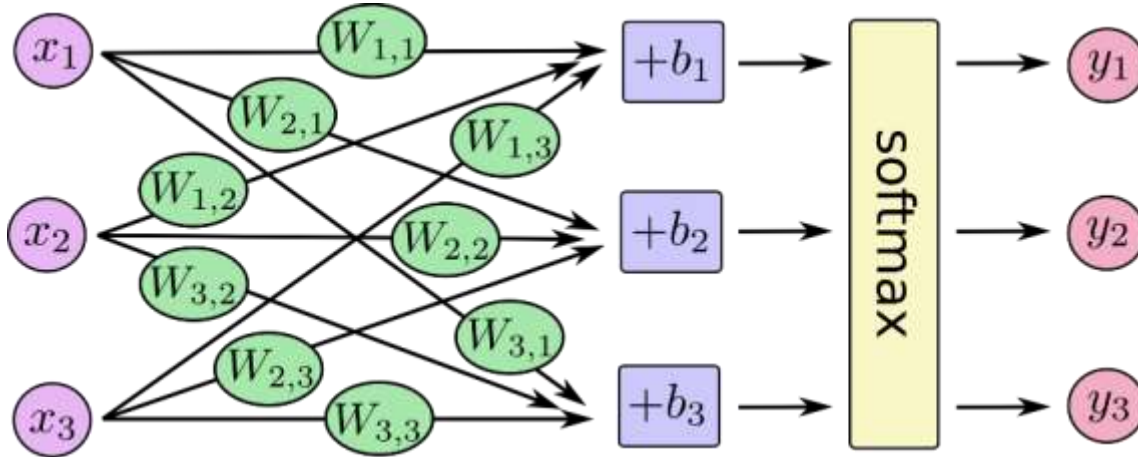
# Building a model – and **what's a hyperparameter**

Each layer has a series of **weights** and **biases**

These are modified as the model is trained to produce higher accuracy

These weights and biases are held in matrices

The amount of change each iteration is called the **learning rate**



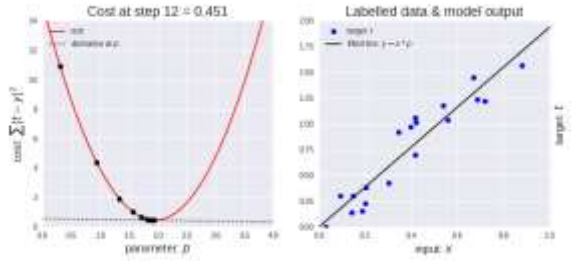
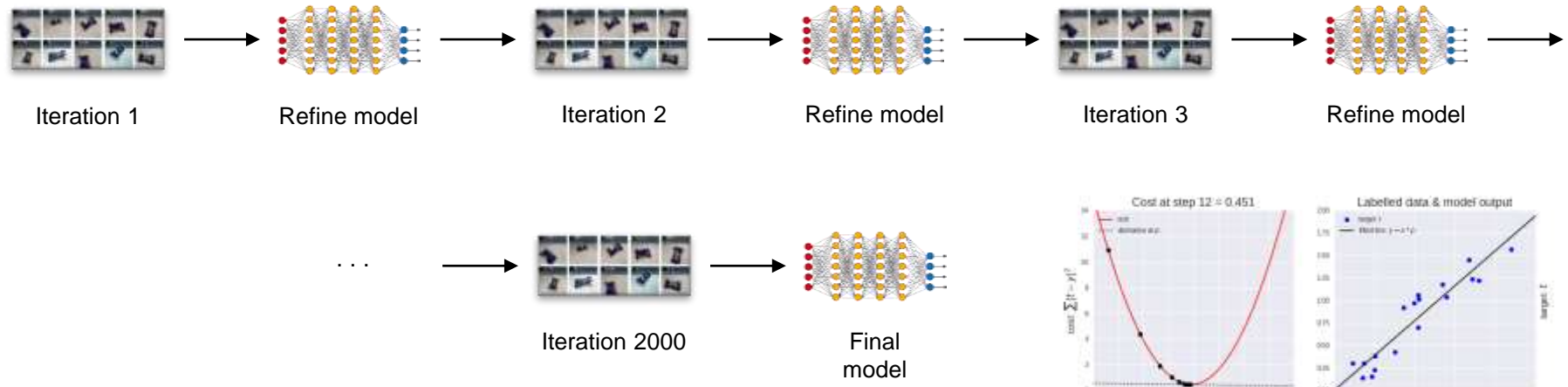
# Building a model – and **what's a hyperparameter**

Deep Learning training requires the system to learn from the data

This requires multiple **iterations** or **epochs**

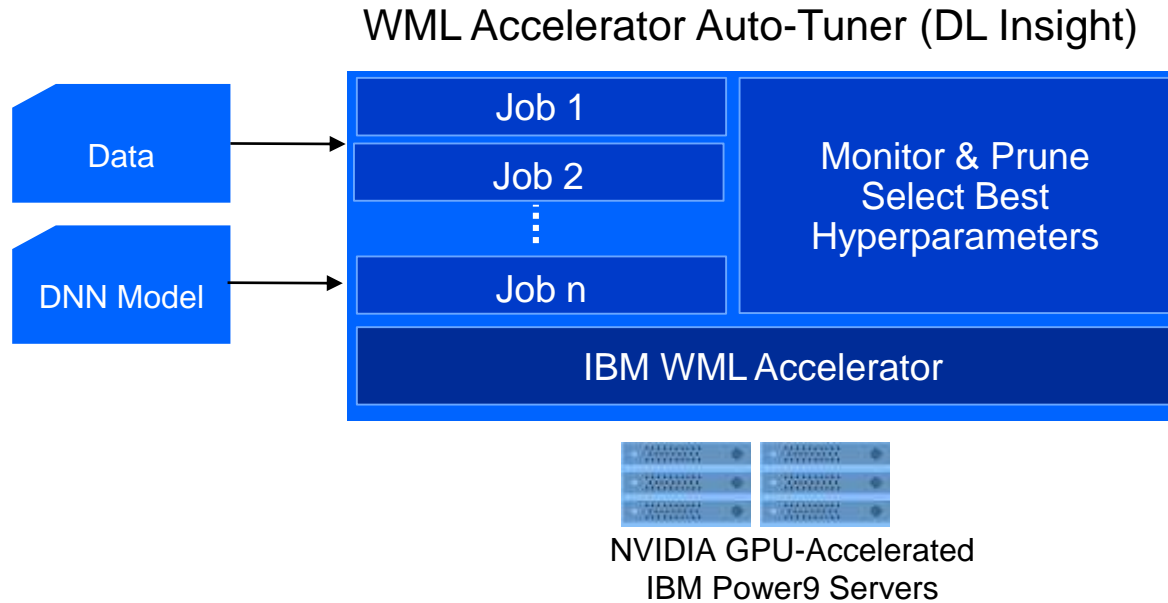
Over each iteration or epoch the system looks at all of the training data

The model is refined based on the **loss** from previous run



# Auto Hyper-Parameter Tuning with WML Accelerator

- Data scientists run 100s of jobs with different Hyper-parameters
  - Learning rate, Decay rate, Batch size, Optimizers (GradientDecedent, Adadelata, Momentum, RMSProp, ..)
- Auto-Tuner searches for good hyper-parameters by launching 10s of jobs in parallel with 10,000s of iterations & selecting the best ones
  - 4 search algorithms: **Random**, **Tree-based Parzen Estimator (TPE)**, **Bayesian**, **Hyperband**



# Hyperparameter search & optimization (HPO)

- Find the best hyperparameters using cognitive algorithms running in parallel, refining the values as the search progresses
- Supported Algorithms
  - Random Search
  - Bayesian
  - TPE
  - Hyperband
  - More to come...

Tune Hyperparameters for model:Caffe-vgg19-flower

\* Hyperparameter search type: Random Search

**Tuning Parameter Settings**

Input the parameters that will be tuned

\* Optimizer (select at least 1):  
 SGD  
 AdaDelta  
 AdaGrad  
 Adam  
 Nesterov  
 RMSProp

\* Learning rate range:

\* Weight decay range:

Overview Hyperparameter Tuning Training Validation Results

Framework:	TensorFlow (Distributed training with IBM Fabric and auto-scaling)	Spark instance group:	dlm
Model file:	/shared/dl/models/TensorFlow/inceptionv3-dong-tuning-20180424083651	Batch size:	32
Dataset:	flowers-incept		

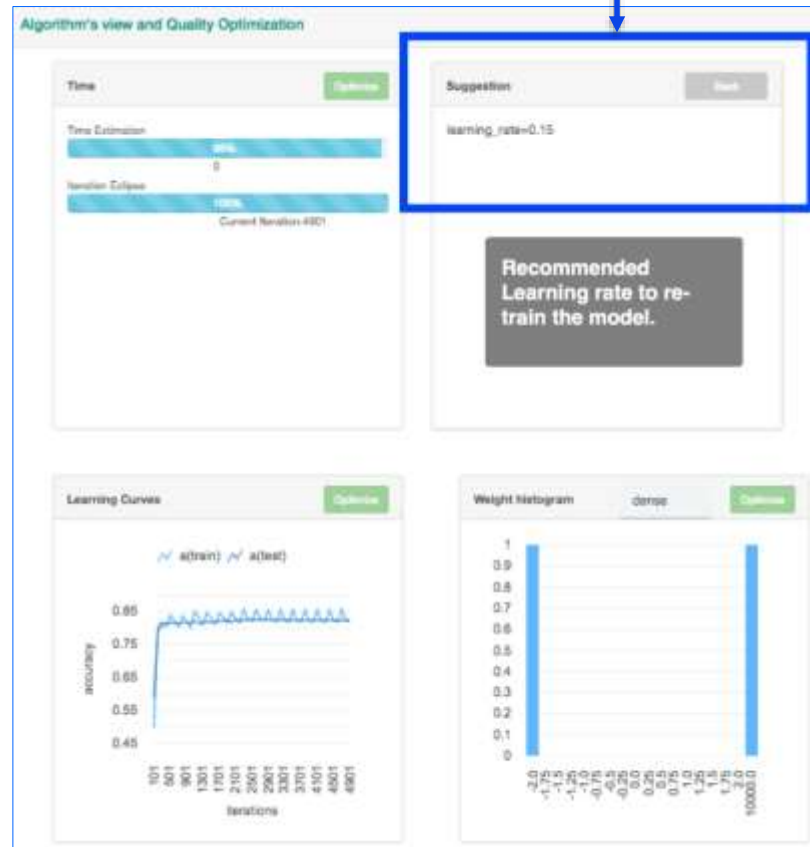
**Hyperparameter**

Learning rate policy:	fixed	Base learning rate:	0.020041713	Learning rate decay:	0
Staircase:	True	Solver type:	Momentum	Momentum:	0.013016215
Decay:	0.1	Epsilon:	1	Maximum iterations:	5000



# Training Visualization

## Monitor, Analyze, Optimize



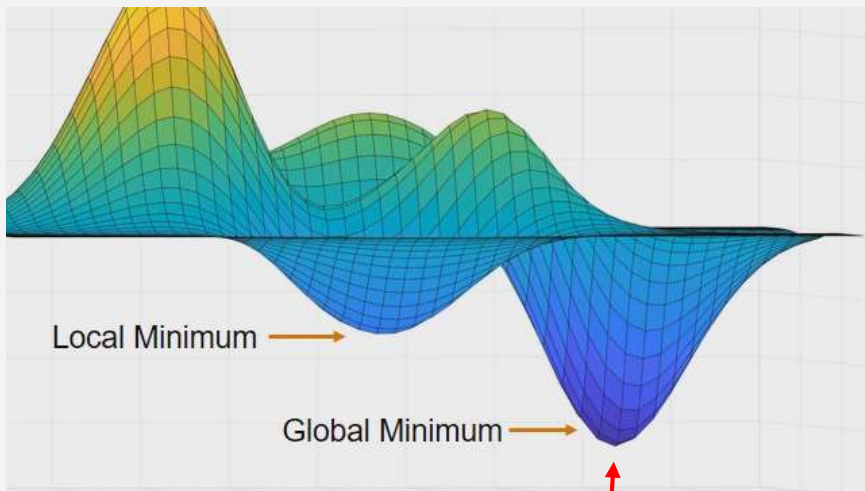
## Two more examples

- IBM Bayesian Optimization Accelerator
- IBM Power AI Vision

# What is Bayesian Optimization?

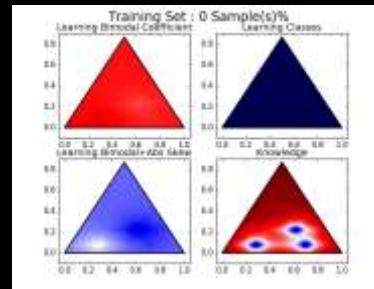
Bayesian optimization is a sequential design strategy for global optimization. **IBM BAO** ('19Q4)

Many workflows require you to find a powerful set of parameters solve a problem. The challenge is finding those parameters robustly in as little time as possible.



IBM BOA gets HPC clients designs here *faster than any other method*

## Applied to Computational Chemistry



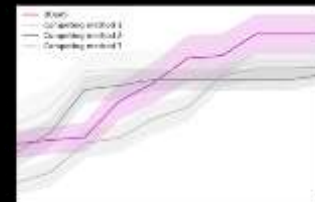
BOA accelerated workflow uses 1/3 of the calculations to achieve 4 orders of magnitude resolution increase

## Applied to Engineering Design

BOA performed in 19 hours and ~30 simulations what an expert designer would do in 3 weeks

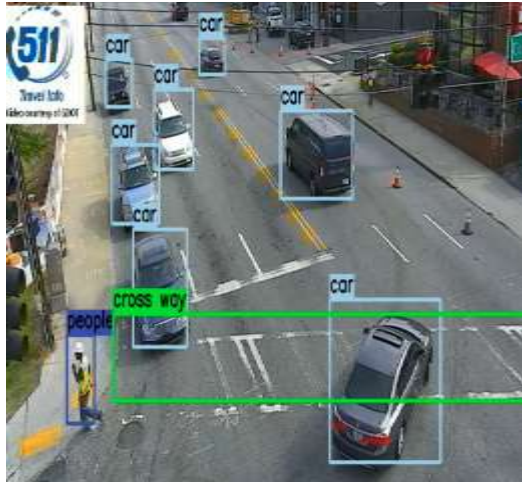
## Applied to Drug Discovery

Brute force methods of screening require 20,000 experiments. BOA accelerated method required ~200



# PowerAI Vision: "Point-and-Click" AI for images & video

Label Image or Video Data



Auto-Train AI Model  
(no coding, just point and click)

My DL Tasks / Create Task

New DL Task - Build Image Classifier

1 Choose Dataset  
Select or create dataset

2 Build Model  
Build model based on selected dataset

3 Deploy And Test  
Deploy trained model and run test

Name of Image Classifier: Test1

Select dataset: Rooms or Create a New Dataset

Build Model Cancel

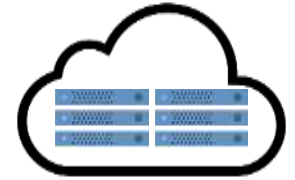
Labels: Status: training

Train Iteration: 101  
Train Loss: 0.47125

Test Iteration: 100  
Test Loss: 0.47246  
Accuracy: 0.81771

The training progress section contains two line graphs. The left graph shows 'Train Loss' (red line) decreasing from approximately 1.0 to 0.5 over 100 iterations. The right graph shows 'Accuracy' (black line) increasing from approximately 0.6 to 0.8 over 100 iterations. A legend indicates 'Train Loss' and 'Accuracy'.

Package & Deploy AI Model



# IBM Machine Learning Accelerator - software

AI for  
Data Scientists and  
non-Data Scientists

**PowerAI Vision**  
Auto-DL for Images & Video

Label    Train    Deploy

**H2O Driverless AI**  
Auto-ML for Text & Numeric Data, NLP

Import    Experiment    Deploy

**WMLA CE : Open Source ML Frameworks**

TensorFlow™    Caffe    PyTorch

Large Model Support (LMS)    Distributed Deep Learning (up to 4 nodes)

Distributed Deep Learning (DDL – 1000s of nodes)    Auto Hyper-parameter Tuning

**IBM Spectrum Conductor with Spark**  
Cluster Virtualization,  
Dynamic Resource Orchestration,  
Multiple Frameworks, Distributed Execution Engine

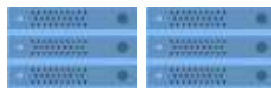
**Deep Learning Impact (DLI) Module**

Data & Model Management, ETL, Visualize, Advise

Watson Machine Learning Community Edition



Accelerated Infrastructure



IBM/NVIDIA Accelerated Servers



Storage

# IBM Mini-CORAL SUMMIT Starter Kit: What's Included



Supercomputer  
deployed by Oak  
Ridge National  
Laboratory  
CORAL/Summit

## Starter Configuration with (6) IBM Power Servers and AI/ML Software and Services

- ✓ IBM *Machine Learning Accelerator* software suite; RHEL 7.6
- ✓ 5 Days Lab Services to stand up: on-site AI/ML Workshop
- ✓ Software Installed in Manufacturing to accelerate time-to-value

Start your AI journey with a complete environment enabling data scientists to unlock advanced insights to drive business benefits.



Benefit from AI model learning times measured in minutes not hours

Use open source frameworks and leading-edge AI technology from IBM Research



PYTORCH



TensorFlow



Caffe



Snap ML



A long-exposure photograph of a city street at night, showing blurred light trails from cars and buildings in various colors like blue, purple, and yellow. A white rectangular box with vertical stripes is centered over the image.

**Thank You**