





# Deep Learning at Scale with BigDL on Apache Spark and IBM Watson Studio

Anthony Stevens, Offering Manager, Watson Deep Learning May 24, 2018







# **Topics Covered**

Survey of Deep Learning Applications

Watson Studio

Intel BigDL on Apache Spark

Demo – Putting it all together







## **Deep Learning** = Training Artificial Neural Networks

- 25 million "neurons"
- 100 million connections (parameters)



## A human brain has:

- 200 billion neurons
- 32 trillion connections between them



4

# app development 2.0

# data + algorithms + neural networks + tools + team







### INTERNET & CLOUD

MEDICINE & BIOLOGY



### MEDIA & ENTERTAINMENT

## SECURITY & DEFENSE

### AUTONOMOUS MACHINES

Cancer Cell Detection Image Classification Video Captioning Face Detection Video Pedestrian Detection **Diabetic Grading** Speech Recognition Video Search Surveillance Satellite Lane Tracking Drug Discovery Language Translation **Real Time Translation** Recognize Traffic Sign Imagery Language Processing Sentiment Analysis Recommendation

# **Deep learning marathon** ...not a sprint



IBM Cloud / Watson and Cloud Platform / © 2018 IBM Corporation



7



## **Progress in Deep Learning**





### **Watson Studio**

	`Ö́Г ІВМ	Watson	Projects	Tools (	Catalog	Community	Services				Docs	Support	Manage	Ļ	-					
My P	rojects / w	vorkshop							+ Add to project -	+0	1	ţ.	<b>P</b>	10 81	Ø					
	Overview	Assets	Environments	Bookma	arks	Deployments	Collaborators	s Settings												
	Q What ass	ets are you lool	king for?																	
~	Data as	ected											(+) New	ı data asse	ŧt					
		~			TYPE		SERVICE	CREATED BY	LAST MODIFIED				AC	CTIONS						
	Ś	customers.cs	SV		Data Asse	et	Project	Anthony Stevens	21 Apr 2018, 10	):11:43 a	m	:								
	Ó	calls.csv			Data Asse	et	Project	Anthony Stevens	21 Apr 2018, 10	):11:53 a	m			÷						
	√ Visual ı	recognitio	n models									+ New v	visual recogn	ition mode	ŧ۱					
	NAME			MODEL	ID			SERVICE INSTANCE	LAST MODIFIED				AC	CTIONS						
	Default Cus	Availab	ole upon t	raining		Visual Recognition	24 Apr 2018, 5:			8 :										
`	✓ Notebo	ooks											+ Nev	w noteboo	k					
	NAME	SHARE	D SC	HEDULED		STATUS	LANGUAG	E LAST EDITOR	LAST MODI	FIED			ACTIONS							
							you currently	y have no notebooks												

## Watson Studio Built for Enterprise AI teams



### **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: Domain expert



**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
```

# **IBM Watson Studio**

Collaborative environment for data teams

- Unified workspace supporting your choice of tools
- Supports 3 techniques for machine learning
  - 1. Pre-built models
  - 2. Transfer learning
  - 3. Custom model development



datascience.ibm.com



### Take a Multi-Framework Approach to Deep Learning



### New frameworks emerging monthly.

Tensorflow was awesome yesterday but has static graphs so PyTorch's dynamic graphs are now popular



# Watson Studio & BigDL: Scaling Solved

- BigDL makes it easy to build DL models on Spark
- Watson Studio allows quick scaling of Spark or GPUs for model training
- Watson Studio and BigDL support multiple frameworks



# **MOTIVATION FOR DEEP LEARNING AT SCALE**



# Trend #1: Data Scale Driving Deep Learning Process



# Trend #2: Real-World ML/DL Systems Are Complex Big Data Analytics Pipelines



Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.



# Trend #3: Unified Big Data Platform Driving Analytics & Data Science







# What is Apache Spark?

An open-source cluster-computing framework

- Developed at the University of California, Berkeley's AMPLab
- Donated to the Apache Software Foundation, which maintains it.

Spark provides an interface for programming entire clusters with implicit data parallelism and fault-tolerance.

 Fault tolerance is critical as production-level deep learning trains for days or even weeks.



# What is **BigDL**?



Distributed deep learning framework for Apache Spark developed by Intel

- Unites big data processing and deep learning
- Allowing deep learning practitioners to continue using familiar tools+infrastructure
- Licensed under the Apache 2.0 license.



# Unified Big Data Analytics Platform

### Hadoop & Spark Platform



How to Run Deep Leaning Workloads Directly on Big Data Platform?

- Integrated with Big Data ecosystem
- Massively distributed, sharednothing
- Scale-out
- · Send compute to data
- Fault tolerance
- Elasticity
- Incremental scaling
- Dynamic resource sharing

•••



# BigDL is designed for big data

- Make deep learning more accessible to big data users and data scientists
  - Write deep learning applications as standard Spark programs
  - Run on existing Spark/Hadoop clusters (no changes needed)
- Feature parity with popular deep learning frameworks
  - E.g., Caffe, Torch, Tensorflow, etc.
- High performance
  - Powered by Intel MKL and multi-threaded programming
- Efficient scale-out
  - Leveraging Spark for distributed training & inference





# What's inside BigDL?

Scala\* and Python\* support

Integrates with Spark\* ML & Spark Streaming

Easy development with Jupyter\* notebook

Visualization with Tensorboard\*

Fine tune Caffe\*/Torch\*/Tensorflow\*/Keras\* Models



# **BIGDL AS A STANDARD SPARK PROGRAM**

### DISTRIBUTED DEEP LEARNING APPLICATIONS (TRAINING, FINE-TUNING & PREDICTION)

### - NO CHANGES TO THE EXISTING HADOOP/SPARK CLUSTERS NEEDED



# Models Interoperability Support

Model Snapshot

- Long training work checkpoint
- Model deployment and sharing
- Fine-tune

### Caffe/Torch/Tensorflow Model Support

- Model file load
- Easy to migrate your Caffe/Torch/Tensorflow work to Spark





# **BigDL: Python API**

- Support deep learning model training, evaluation, inference
- Support Spark v1.5/1.6/2.0/2.1/2.2
- Support Python 2.7/3.5/3.6
- Based on PySpark, Python API in BigDL allows use of existing Python libs (Numpy, Scipy, Pandas, Scikit-learn, NLTK, Matplotlib, etc)

```
train data = get minst("train").map(
    normalizer(mnist.TRAIN MEAN, mnist.TRAIN STD))
test data = get minst("test").map(
    normalizer(mnist.TEST MEAN, mnist.TEST STD))
state = {"batchSize": int(options.batchSize),
         "learningRate": 0.01,
         "learningRateDecay": 0.0002}
optimizer = Optimizer(
   model=build model(10),
   training rdd=train data,
    criterion=ClassNLLCriterion(),
    optim method="SGD",
    state=state,
    end trigger=MaxEpoch(100))
optimizer.setvalidation(
    batch size=32,
   val rdd=test data,
   trigger=EveryEpoch(),
   val method=["top1"]
```

```
optimizer.setcheckpoint(EveryEpoch(), "/tmp/lenet5/")
trained_model = optimizer.optimize()
```



# BigDL works with Jupyter Notebook

## Python support

- Necessity for deep learning: TensorFlow and PyTorch
- ✓ Share and Reproduce
  - Notebooks can be shared with others
  - Easy to reproduce and track
- Rich Content
  - Texts, images, videos, LaTeX and JavaScript
  - Code can also produce rich contents
- Rich toolbox
  - Apache Spark, from Python, R and Scala
  - Pandas, scikit-learn, ggplot2, dplyr, etc



# Visualization of training process + results

## BigDL integrated with TensorBoard

• TensorBoard is a suite of web applications from Google for visualizing and understanding deep learning applications

SCALARS	IMAGES	AUDIO	GRAPHS
oss			
Loss			
2.40			
2.00			
1.60			
1.20			
0.800			
0.400			
0.00			



## **NOTICES AND DISCLAIMERS**

- © 2018 International Business Machines Corporation. No part of this document may be reproduced or transmitted in any form without written permission from IBM.
- U.S. Government Users Restricted Rights use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.
- Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. This document is distributed "as is" without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity. IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.
- IBM products are manufactured from new parts or new and used parts. In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."
- Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

- Performance data contained herein was generally obtained in a controlled, isolated environments. Customer examples are presented as illustrations of how those
- customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.
- References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.
- Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.
- It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.



## **NOTICES AND DISCLAIMERS CONTINUED**

- Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.
- The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

IBM, the IBM logo, ibm.com and [names of other referenced IBM products and services used in the presentation] are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: <u>www.ibm.com/legal/copytrade.shtml</u>.



## **THANK YOU!**

- Create free Watson Studio account at:
  - https://datascience.ibm.com/
- Learn more about BigDL at:
  - https://github.com/intel-analytics/BigDL
  - <u>software.intel.com/bigdl</u>



