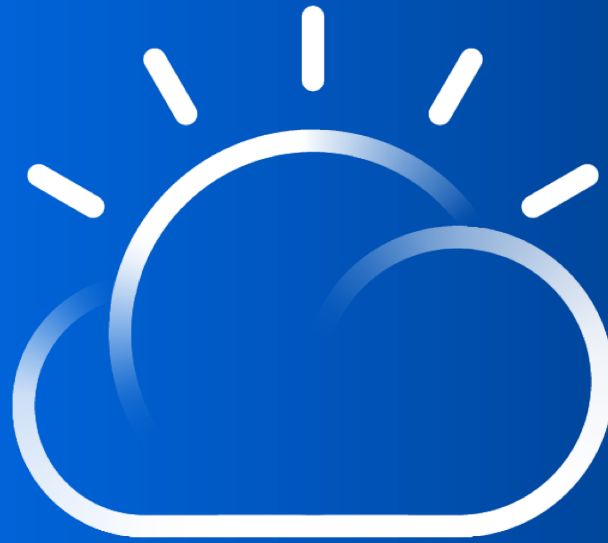


IBM Cloud for AI at Scale



Alex Hudak
Offering Manager, IBM Cloud



IBM Cloud GPU Use Cases

- Financial Services
- Gaming
- Scientific Research
- Healthcare
- Deep Learning



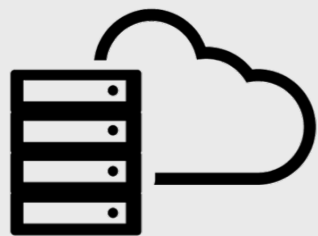
"In our business, there are so many computations that occur. A process that would normally take us 10 days has shrunk to a couple of hours by running NVIDIA® GPUs on IBM Cloud."

Jeffrey Thatcher,
Chief Scientist,
Spectral MD



IBM Cloud GPU

Performance for SpectralMD



1 CPU

3 Hours Training
4 Algorithms
20 Folds of CV

10 Days



1 GPU

1 Hour Training
4 Algorithms
20 Folds of CV

3.3 Days

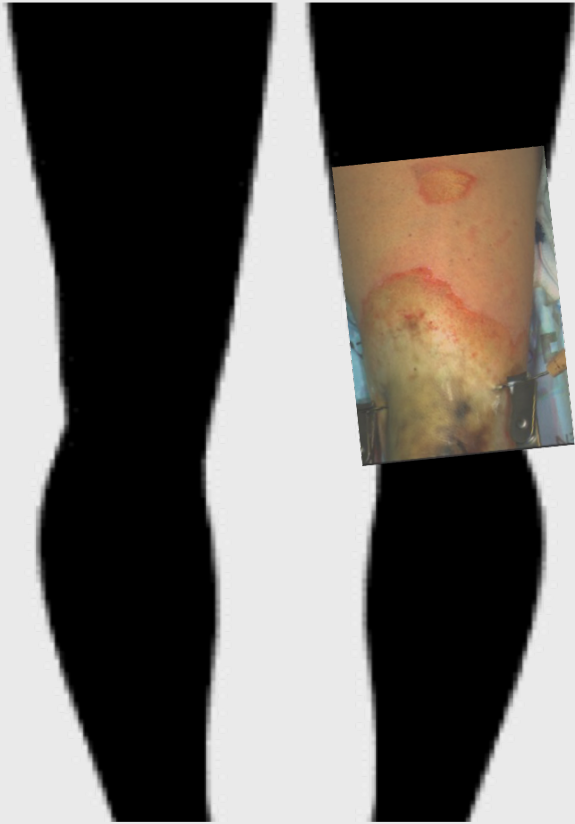


4 GPUs

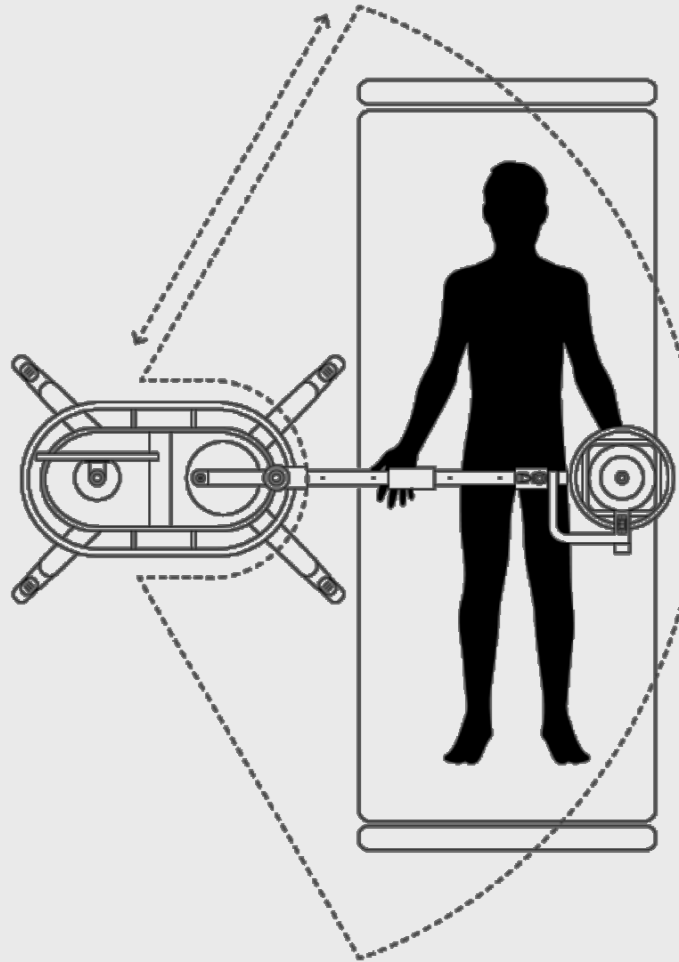
1 Hour Training
4 Algorithms / 4 GPUs
20 Folds of CV

0.8 Days

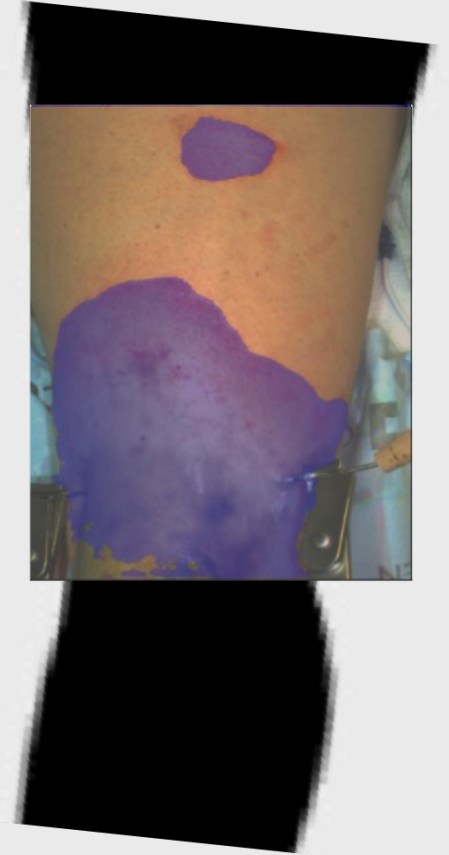
DeepView Burn Imaging



Burn injury is prepped for imaging



DeepView Imaging



Automatically predicts burns that require surgery

IBM Cloud Offerings with NVIDIA[®] GPUs

GPUs on IBM Cloud

NVIDIA® Tesla®
M60

Bare Metal
Monthly

NVIDIA® Tesla®
K80

Bare Metal
Monthly

NVIDIA® Tesla®
P100

Bare Metal
Monthly
&
Virtual
Monthly & Hourly

NVIDIA® Tesla®
V100

Bare Metal
Monthly
&
Virtual
Monthly & Hourly

GPUs on IBM Cloud

Tesla M60

Used for enterprise virtualization as well as boosting professional graphics performance.

Bare Metal
Monthly

Tesla K80

Reliable GPU performance ideal for introductory AI computing at an affordable price.

Bare Metal
Monthly

Tesla P100

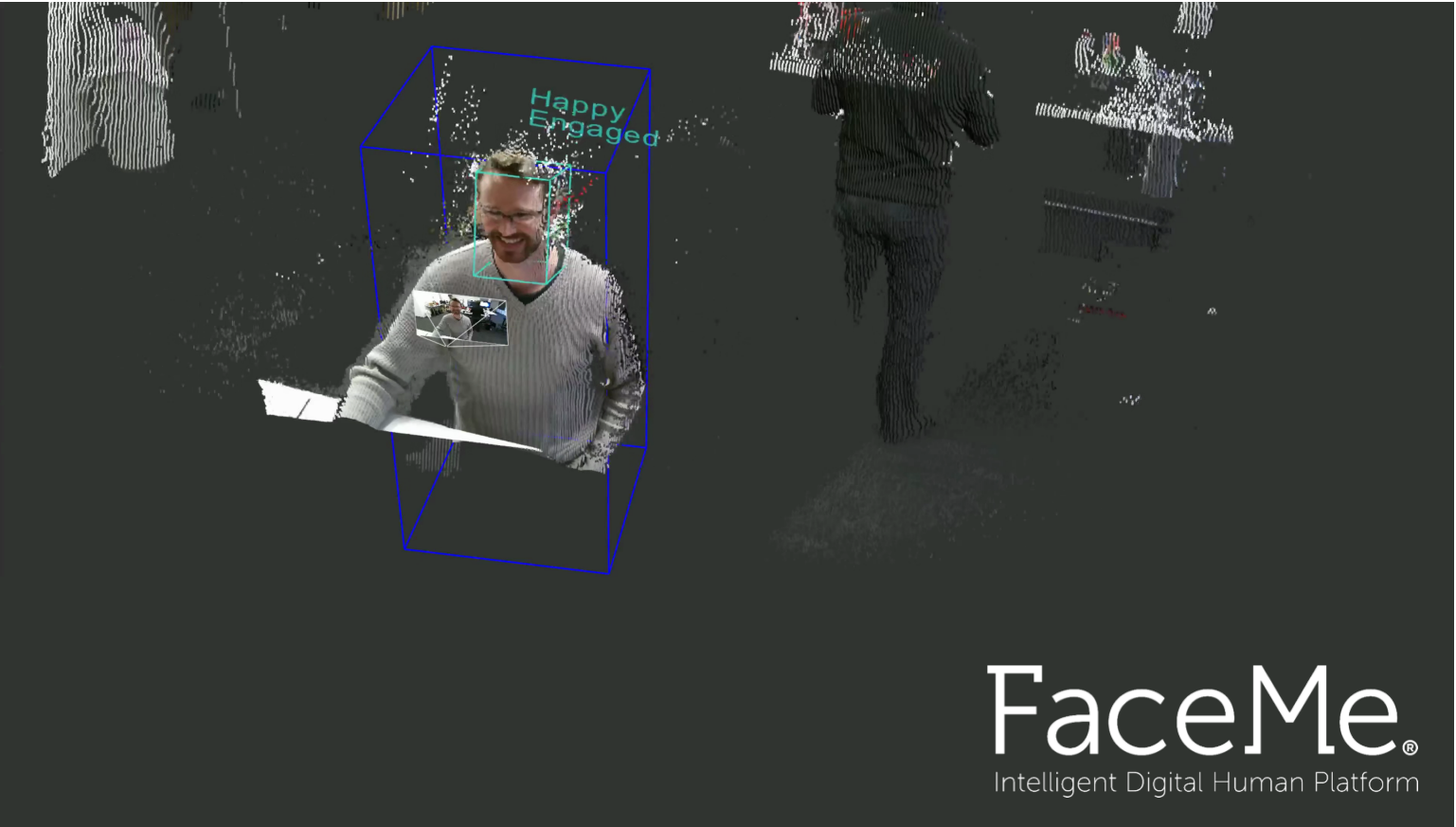
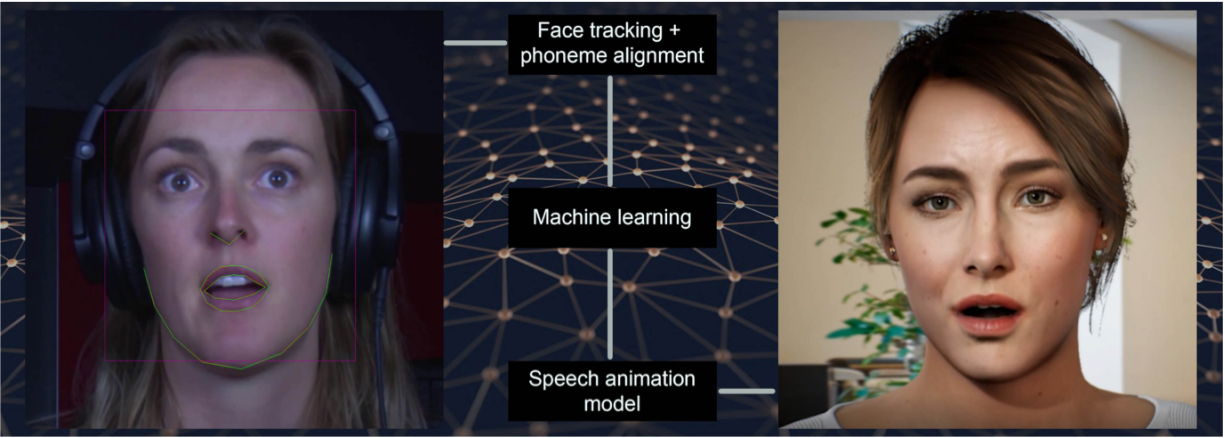
Essential performance for standard AI and HPC capabilities.

Bare Metal
Monthly
&
Virtual
Monthly & Hourly

Tesla V100

IBM Cloud's most powerful and advanced GPU, purpose-built for progressive Deep Learning workloads — with the performance of 100 CPUs in a single GPU.

Bare Metal
Monthly
&
Virtual
Monthly & Hourly



GPU-Enabled Data Centers

North America

- Dallas
- Houston
- Mexico
- Montreal
- Seattle
- San Jose
- Toronto
- Washington DC

Europe

- Amsterdam
- Frankfurt
- London
- Milan
- Oslo
- Paris

Asia

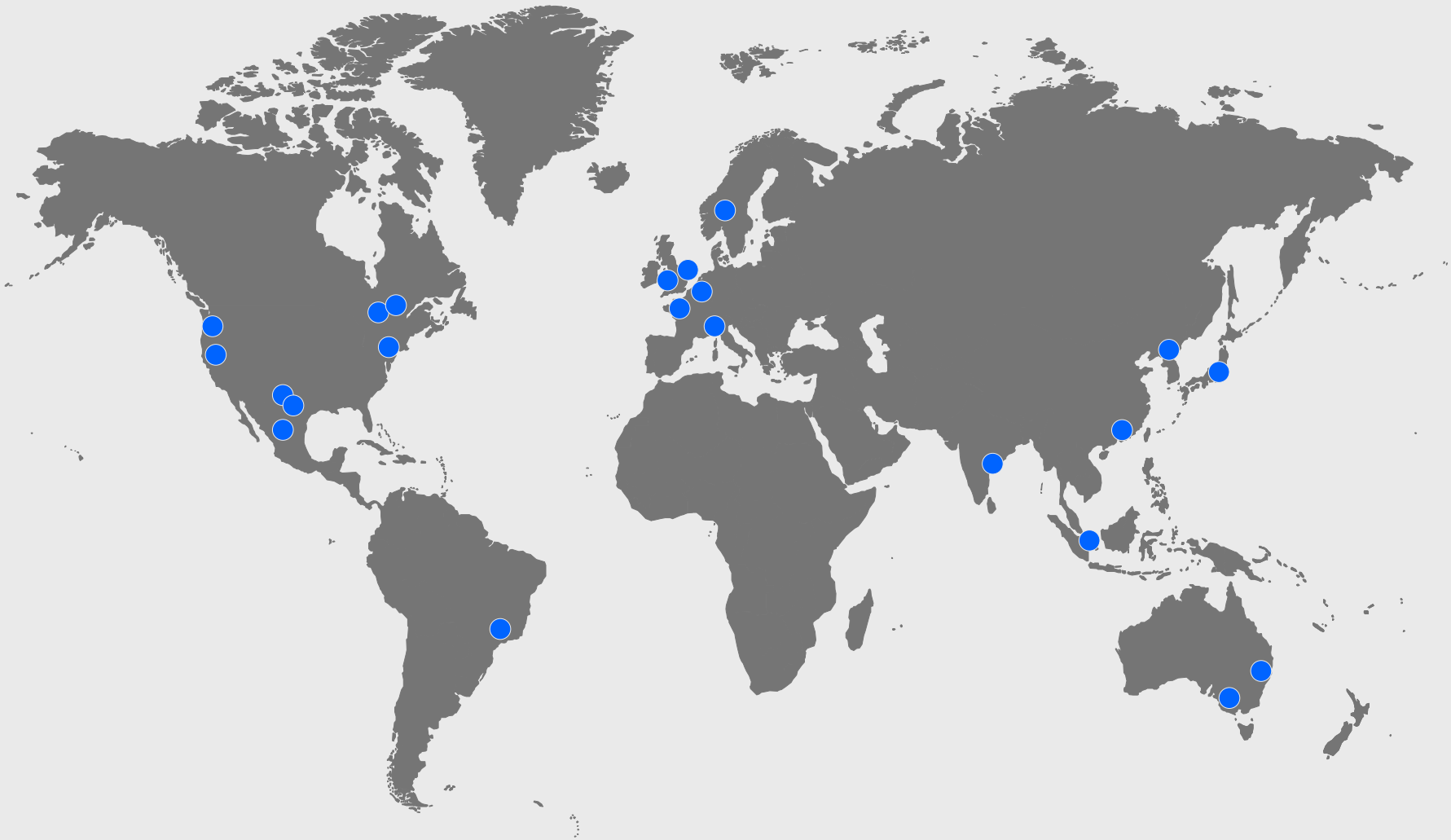
- Chennai
- Hong Kong
- Seoul
- Singapore
- Tokyo

Australia

- Melbourne
- Sydney

South America

- Sao Paulo



GPU-Enabled Data Centers

NVIDIA Tesla® V100

Dallas*	London*
San Jose	Frankfurt
Washington DC*	Tokyo

NVIDIA® Tesla® P100

Dallas*	Amsterdam
San Jose	Seoul
Washington DC*	Tokyo
London*	Sydney
Frankfurt	

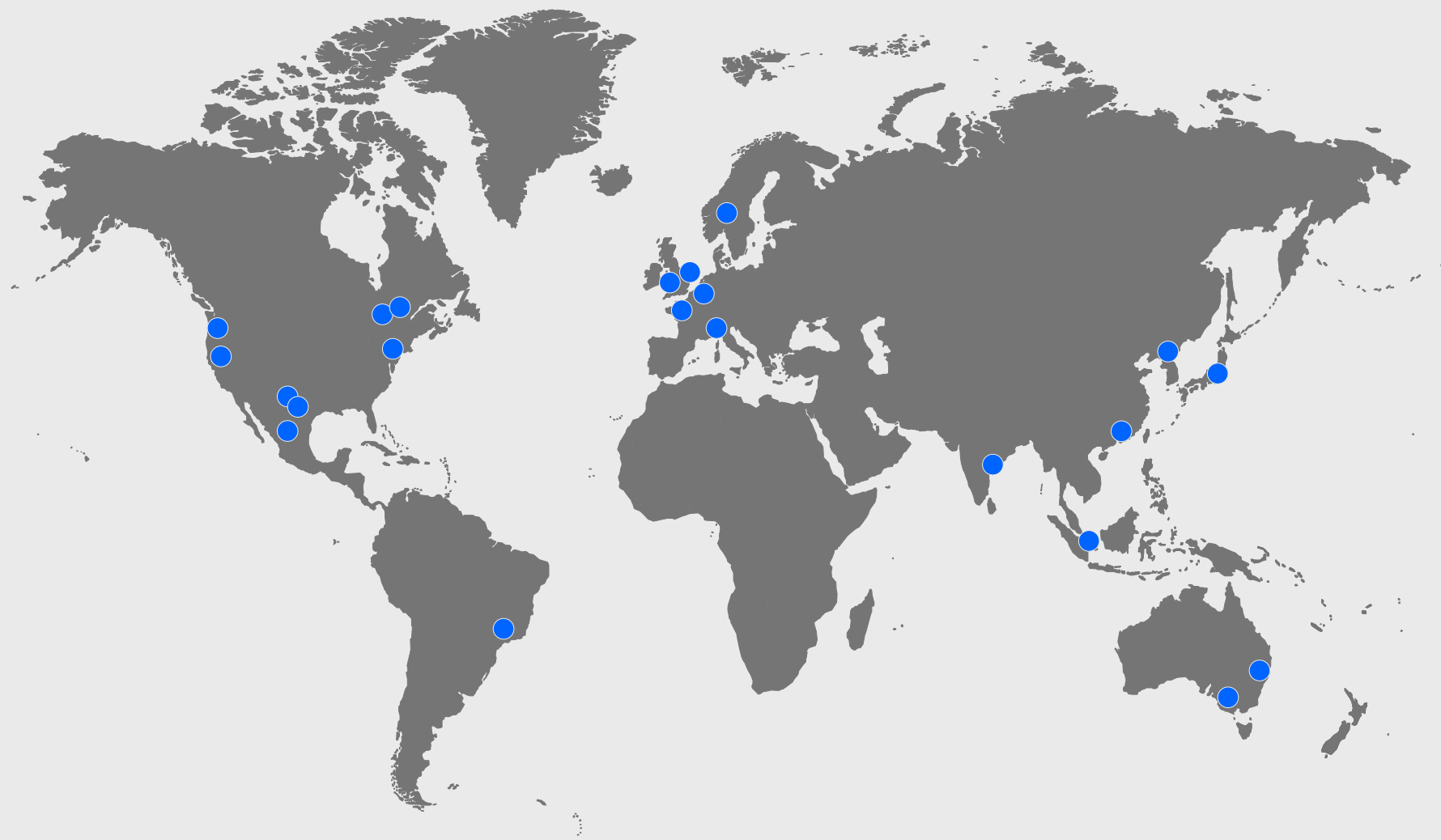
NVIDIA® Tesla® K80

All GPU DCs

NVIDIA® Tesla® M60

All GPU DCs

*Also Available on
Virtual Servers



Virtual Servers Family Benefits



Ease of use



Improved selection process from a spectrum of public virtual server flavors

Lower provisioning times



Pre-configured virtual servers based on workloads to easily select a best fit

Cost savings



Hourly and monthly pricing options available to optimize resource utilization

NVIDIA® GPUs on IBM Cloud Virtual Servers

Run HPC, AI, and simulation workloads with efficiency and scalability in a virtual environment



Table 1. P100 GPU flavors

Flavor	GPU	GPU RAM (GB)	vCPU	vCPU RAM (GB)	Storage Type	Boot Disc (GB)	Secondary Discs (2 and 3) (GB)
ac1.8x60x25	1 P100	16	8	60	Block (SAN)	25	None
ac1.8x60x100	1 P100	16	8	60	Local SSD or SAN	100	None (SAN) 300 (Local)
ac1.16x120x25	2 P100	32	16	120	Block (SAN)	25	None
ac1.16x120x100	2 P100	32	16	120	Local SSD or SAN	100	None (SAN) 600 (Local)

Table 2. V100 GPU flavors

Flavor	GPU	GPU RAM (GB)	vCPU	vCPU RAM (GB)	Storage Type	Boot Disc (GB)	Secondary Discs (2 and 3) (GB)
ac2.8x60x25	1 V100	16	8	60	Block (SAN)	25	None
ac2.8x60x100	1 V100	16	8	60	Local SSD or SAN	100	None (SAN) 300 (Local)
ac2.16x120x25	2 V100	32	16	120	Block (SAN)	25	None
ac2.16x120x100	2 V100	32	16	120	Local SSD or SAN	100	None (SAN) 600 (Local)

GPUs on Bare Metal

No hypervisor overhead

All resources dedicated to a single user

Customizable to support specific applications

Hard Drive

1 TB SATA –
3.8 TB SSD

Core

12 – 28 Cores

Network

100 MBps –
10 GBps

Memory

64 GB –
1.5 TB



Server Software

Workloads perform best with the right software.

- **Wide selection:**
Supports all leading open source and enterprise operating systems, utilities, and applications
- **Best practices:**
Installation, configuration, and updates are performed to ensure peak performance and security
- **On-demand:**
Installation, reloads, and updates are automated and controlled via API, portal, or mobile app



**Simplify infrastructure administration
with server software**



Watson Studio on IBM Cloud

IBM Watson

ProjectsToolsCatalogCommunityServices

US South

My Projects / Deep Learning - COS Lite / Fashion MNIST

Queued

NAME	SUBMITTED
No training runs found.	

In progress

NAME	DURATION
No training runs found.	

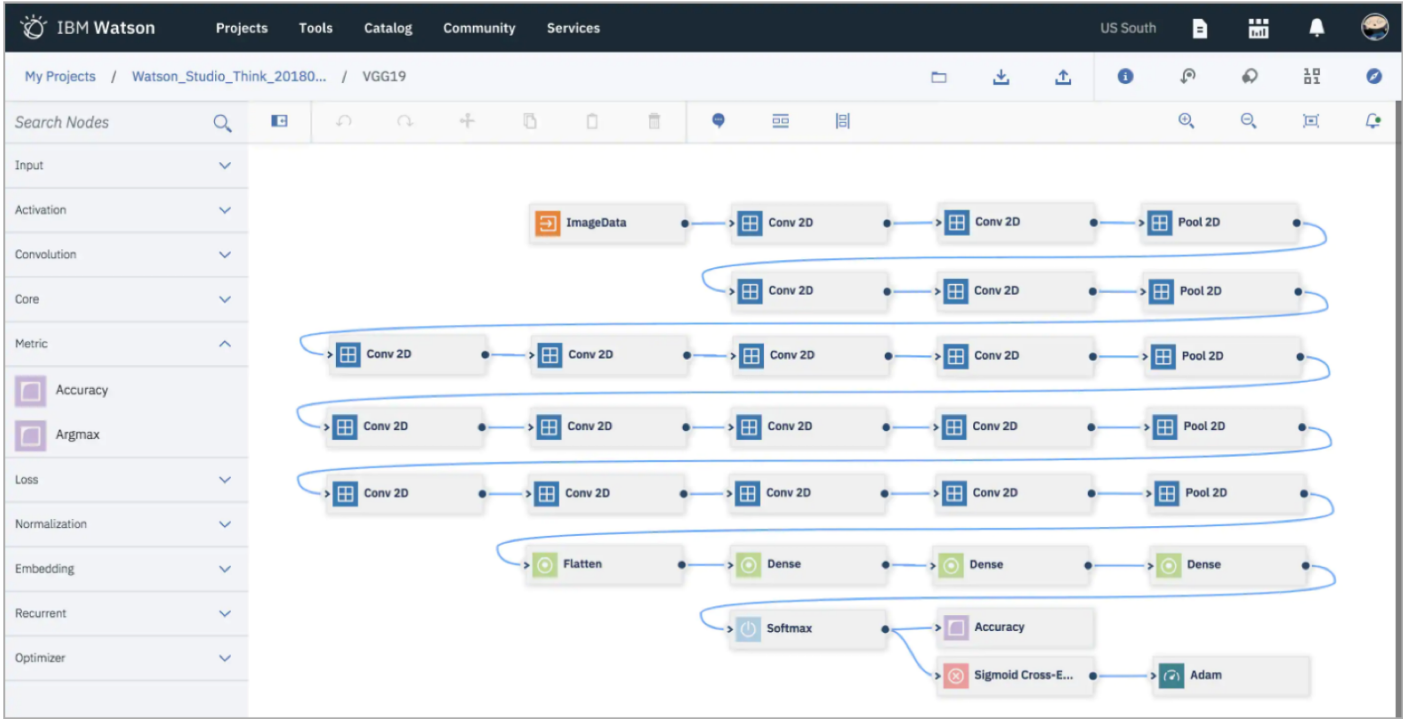
Completed

NAME	STATUS	DURATION	VAL_ACC	VAL_LOSS	ACC	LOSS	ACTION
training-dnDL8SgmR	completed	0 hr, 0 min, 0 sec	0.729	3.909	0.657	0.893	
training-DqdYUSRiR	completed	0 hr, 0 min, 0 sec	0.834	2.668	0.988	0.034	
training-KLmY8SRmg	completed	0 hr, 0 min, 0 sec	0.800	3.134	0.766	0.622	
training-c64Y8IRiR	completed	0 hr, 0 min, 0 sec	0.711	4.344	0.635	0.973	
training-RinLUIgiR	error	0 hr, 0 min, 0 sec					

Experiment Assistant

Optimized for production environments, scale up your training using the NVIDIA Tesla V100 GPU with your preferred deep learning framework, then easily deploy to the cloud

Visually design neural networks and scale out training runs, while paying only for the resources used



Neural Network Modeler

IBM Cloud GPU Promos

2U 2x 2690 v3, 1x Nvidia M60
128GB DDR4
1x 1TB HDD
20 TB BW

Starting at **\$967**
Discount: Up to 65%

Code: WebB3

2U 2x 2620 v4, 1x Nvidia P100
128GB DDR4
1x 1TB HDD
20 TB BW

Starting at **\$1,105**
Discount: Up to 59%

Code: HPCB1

2U 2x 2650 v4, 1x Nvidia P100
128GB DDR4
1x 1TB HDD
20 TB BW

Starting at **\$1,386**
Discount: Up to 50%

Code: HPCB2

2U 2x 2690 v4, 1x Nvidia V100
256GB DDR4
1x 1TB HDD
20 TB BW

Starting at **\$1,684**
Discount: Up to 50%

Code: HPCB3

Please note

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice and at IBM's sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

Notices and disclaimers

© 2019 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: www.ibm.com/legal/copytrade.shtml.

Thank you



Alex Hudak
Offering Manager, IBM Cloud
—
Alex.Hudak@ibm.com

