



AI ON INTEL

AI Builders Showcase



recently acquired by Hewlett Packard Enterprise

Nanda Vijaydev
Lead Data Scientist, BlueData (HPE)

About BlueData

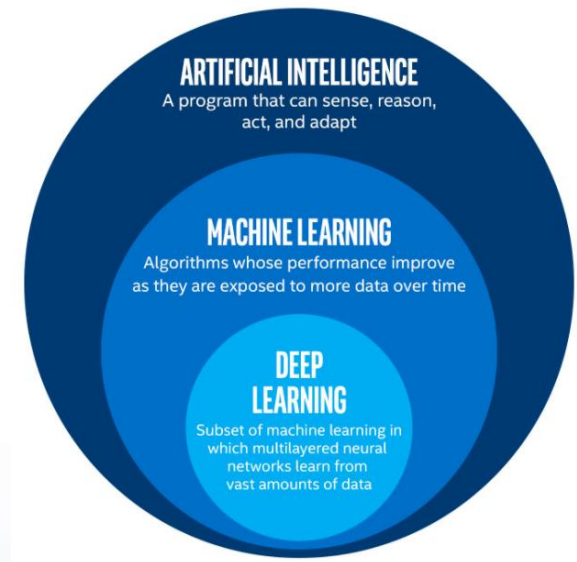


- Leading container-based software platform for distributed AI / ML / DL
 - Recently acquired by Hewlett Packard Enterprise
- Intel strategic collaboration agreement and Intel AI Builder partner
 - Optimized for Intel Xeon architecture – leveraging CPUs for AI inference
- Proven solution to accelerate AI-driven innovation in the enterprise
 - Global 2000 financial services, life sciences, healthcare, retail, manufacturing, etc.

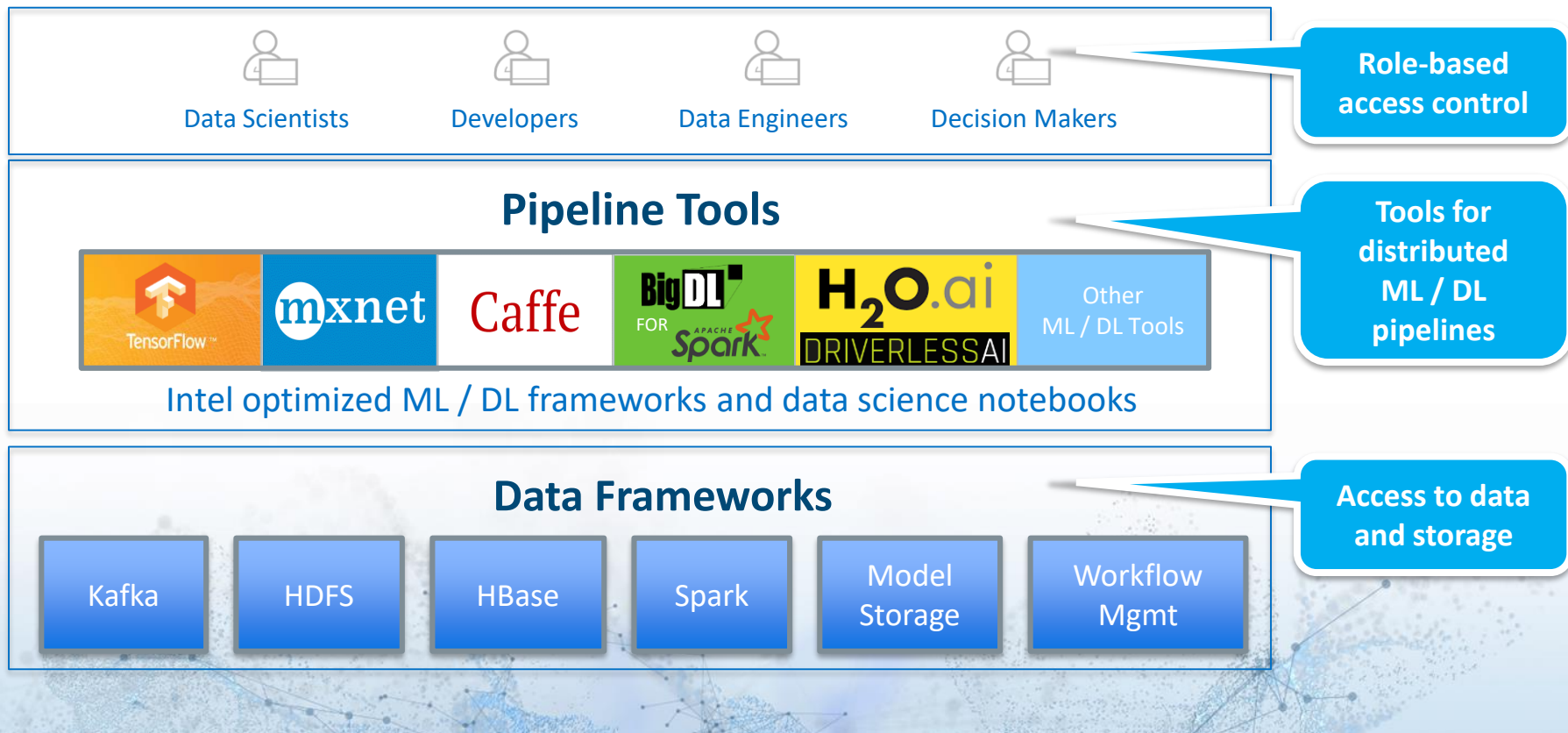


Requirements for Distributed AI / ML / DL

- Access to common data, big and small
- Choice of modeling techniques
- Ability to build, share, and iterate
- Reproducibility
- Easy scaling to test on real data sets
- Support for different roles and actions
- Security and performance
- Deploy on-premises and/or in the public cloud



ML / DL Stack: Tools and Infrastructure



Container-based platform for AI / ml / DL

Data Scientists Developers Data Engineers Data Analysts

BlueData EPIC™ Software Platform



Spark, Hadoop, kafka, cloudera
Big Data Tools

H2O.ai, Big Data, TensorFlow
ML / DL Tools

R, jupyter, Zeppelin
Data Science Tools

BI/Analytics Tools

Bring-Your-Own

ElasticPlane™ – Self-service, multi-tenant clusters

IOBoost™ – Extreme performance and scalability

DataTap™ – In-place access to data on-prem or in the cloud

Compute

Storage

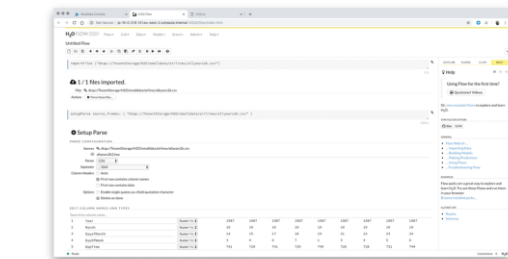
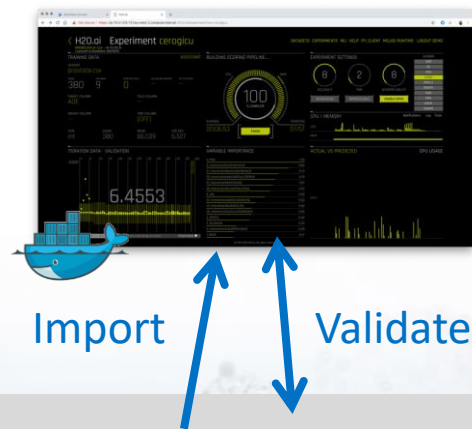


On-Premises

Public Cloud

Example: H2O Pipeline on Containers

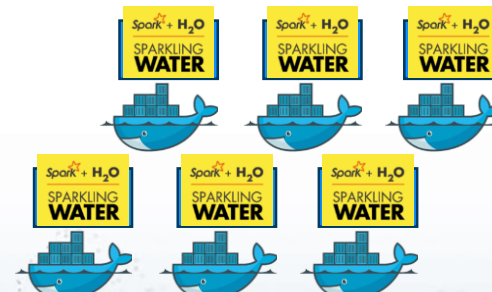
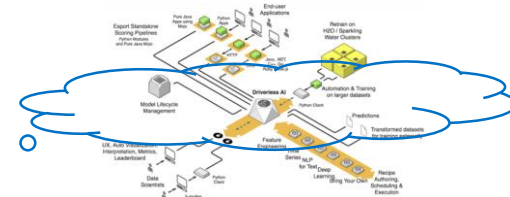
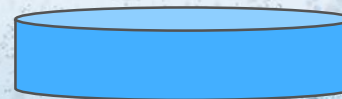
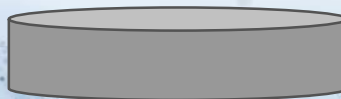
H2O Driverless AI



Shared Data Access Layer



... Data Sources ...



One-click Deployment for Distributed ML / DL

Create New Cluster from Template

Cluster Detail

Cluster Name ⓘ HDF-Pipeline

Cluster Description ⓘ

Select Cluster Type ⓘ HDF

Distribution ⓘ HDF 3.X v4 on Centos 7.x with Ambari 2.7.1.0

Node Roles

Controller ⓘ Medium - 4 VCPU, 12288 MB RAM, 100 GB root disk 1

Worker ⓘ Small - 4 VCPU, 8192 MB RAM 2

Advanced Settings

Debug Mode ⓘ ☒

Isolated Mode ⓘ ☐

Two Phase Delete ⓘ ☐

Bootstrap Action ⓘ ☐

Create Cluster

Flavor Management

[Add New Flavor](#) [Delete](#)

FLAVOR NAME	FLAVOR DESCRIPTION	CORES	MEMORY (GB)	GPU DEVICES	ROOT DISK SIZE (GB)	PERSISTENT STORAGE SIZE (GB)	STATUS	ACTIONS
<input type="checkbox"/> Small	system-created example flavor	4	8	0	use image default	0	ready	
<input type="checkbox"/> Medium	system-created example flavor	4	12	0	100	0	ready	
<input type="checkbox"/> Large	system-created example flavor	8	20	0	200	0	ready	
<input type="checkbox"/> GPU-Small	Flavor with one GPU socket	4	30	1	use image default	0	ready	
<input type="checkbox"/> GPU-Large	Flavor with 4 GPU sockets	10	40	2	use image default	0	ready	

Rows 10 Showing 1 to 5 of 5 entries Previous 1 Next

Pick from a list of
pre-built and tested Docker-
based images

Multi-tenant: Approved applications for each tenant

Edit Tenant

Tenant Name ⓘ
ML Model Sandbox

Tenant Description ⓘ
Sandbox for building AI/ML models

Cluster Superuser Privilege ⓘ
Site Admin and Tenant Admin

Tenant Type ⓘ
On-Prem - Tenant type for managing clusters locally.

QOS Multiplier ⓘ
1

Quotas

External Authentication

Kerberos

Image Catalog

Tags

Images

<input type="checkbox"/> CDH 5.10.1 on 7x with Cloudera Manager ⓘ	<input type="checkbox"/> CDH 5.14.4 multirole 7x ⓘ	<input type="checkbox"/> CDH 5.14.4 on 7x GMF ⓘ
<input type="checkbox"/> CDH 5.15.1 multirole 7x GSK Image 1 ⓘ	<input type="checkbox"/> CDH 5.15.1 multirole 7x-v4 ⓘ	<input checked="" type="checkbox"/> CentOS 7.x ⓘ
<input type="checkbox"/> CentOS 7.x with Cuda 8.0 Python 2.7 CudNN v6 and TensorFlow ⓘ	<input type="checkbox"/> CentOS 7.x with Python 3.6 Cuda 9.0 and TensorFlow 1.7 ⓘ	<input checked="" type="checkbox"/> Confluent Kafka 5.0 ⓘ
<input type="checkbox"/> Generic model serving flask wrapper ⓘ	<input checked="" type="checkbox"/> H2O driverless AI 1.5.1 ⓘ	<input type="checkbox"/> HDF 3.X v4 on Centos 7.x with Ambari 2.7.1.0 ⓘ
<input type="checkbox"/> HDP 2.6 on 7.x with Ambari 2.5 ⓘ	<input type="checkbox"/> HDP 2.6 on 7.x with Ambari 2.6 ⓘ	<input type="checkbox"/> Jupyterhub Fat on centos7x ⓘ
<input checked="" type="checkbox"/> Jupyterhub on Spark 2.2.1 multirole with centos7x ⓘ	<input checked="" type="checkbox"/> Jupyterhub on centos7x ⓘ	<input checked="" type="checkbox"/> Kafka CDH 5.15.1 on 7x GSK Image 3 ⓘ
<input checked="" type="checkbox"/> RHEL 7.x ⓘ	<input checked="" type="checkbox"/> Spark 2.4 with Sparklingwater 2.4.4 ⓘ	<input type="checkbox"/> Stand alone H2O 3.22.1.3 ⓘ
<input checked="" type="checkbox"/> TensorFlow Serving 1.12.0 Image 2 v3 ⓘ	<input checked="" type="checkbox"/> TensorFlow v12 ⓘ	<input checked="" type="checkbox"/> TensorFlow with nccl Horovod and openmpi ⓘ
<input checked="" type="checkbox"/> Tensorflow-GPU 1.12 with Python 3.6 Cuda 10.0 and NVidia Rapids ⓘ	<input checked="" type="checkbox"/> Ubuntu 16.04 LTS (Xenial Xerus) ⓘ	<input type="checkbox"/> tensorflow_serving ⓘ

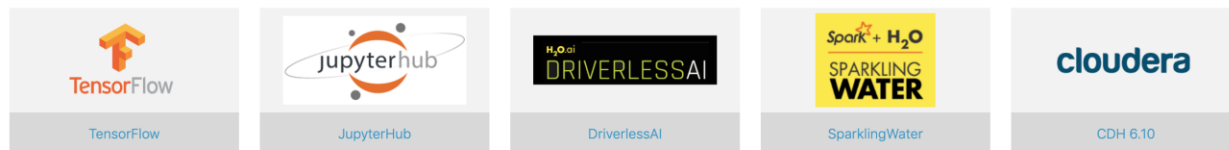
Choice of applications: open source and commercial

Quick Launch Templates

Click on tiles below to launch clusters or use as a starting point to create your own template.

Create Template

Set Priority



Quickly launch third party commercial applications such as H2O Driverless AI and Cloudera

Clusters



Create Cluster

↓ Tenant KeyPair

Actions ▾

<input type="checkbox"/>	NAME >	DISTRIBUTION >	ROLE CONFIGURATIONS >	DETAILS >	STATUS >	ACTIONS
<input type="checkbox"/>	Spark_BigDL	Jupyterhub on Spark 2.2.1 multirole with centos7x	Controller (1/Small) Worker (1/Small) Jupyter (1/Small)	Created At : Tue Apr 16 2019 11:42:36 Created By : sara	ready ✓	
<input type="checkbox"/>	PyTorch	CentOS 7.x	Controller (1/Small)	Created At : Tue Apr 16 2019 11:41:41 Created By : sara	ready ✓	
<input type="checkbox"/>	TF	CentOS 7.x with Cuda 8.0 Python 2.7 CudNN v6 and TensorFlow	Controller (1/Small)	Created At : Fri Apr 12 2019 17:09:50 Created By : sara	ready ✓	

Version : 3.6 Build Number : 18822

Deploy multiple versions of Jupyter notebooks with rich libraries including TensorFlow, PyTorch, Caffe2, and Keras

DataTaps

Create

Delete

NAME >	DESCRIPTION >	HOST >	TYPE >	ADDITIONAL INFO >	PATH >	ACTIONS
HealthCareData	DataLake zone with all health records in HDFS	bluedata-7.infra.lab.bluedata.com	hdfs	kerberos Protected Proxy	/tmp/tatsuya	<div></div> <div></div>
TenantStorage	Protected DataTap for a tenant-specific sandboxed storage space.	yav-011.lab.bluedata.com	hdfs	kerberos Protected Proxy	/2	

Rows 10

Showing 1 to 2 of 2 entries

Previous

1

Next

FS Mounts

<

Multi-tenant: Approved applications for each tenant

Edit Tenant

Tenant Name ⓘ

ML Model Sandbox

Tenant Description ⓘ

Sandbox for building AI/ML models

Cluster Superuser Privilege ⓘ

Site Admin and Tenant Admin

Tenant Type ⓘ

On-Prem - Tenant type for managing clusters locally.

QOS Multiplier ⓘ

1

Quotas

External Authentication

Kerberos

Image Catalog

Tags

Images

☐ CDH 5.10.1 on 7x with Cloudera Manager ⓘ

☐ CDH 5.15.1 multirole 7x GSK Image 1 ⓘ

☐ CentOS 7.x with Cuda 8.0 Python 2.7 CudNN v6 and TensorFlow ⓘ

☐ Generic model serving flask wrapper ⓘ

☐ HDP 2.6 on 7.x with Ambari 2.5 ⓘ

☒ Jupyterhub on Spark 2.2.1 multirole with centos7x ⓘ

☒ RHEL 7.x ⓘ

☒ TensorFlow Serving 1.12.0 Image 2 v3 ⓘ

☒ Tensorflow-GPU 1.12 with Python 3.6 Cuda 10.0 and NVidia Rapids ⓘ

☐ CDH 5.14.4 multirole 7x ⓘ

☐ CDH 5.15.1 multirole 7x-v4 ⓘ

☐ CentOS 7.x with Python 3.6 Cuda 9.0 and TensorFlow 1.7 ⓘ

☒ H2O driverless AI 1.5.1 ⓘ

☐ HDP 2.6 on 7.x with Ambari 2.6 ⓘ

☒ Jupyterhub on centos7x ⓘ

☒ Spark 2.4 with Sparklingwater 2.4.4 ⓘ

☒ TensorFlow v12 ⓘ

☒ Ubuntu 16.04 LTS (Xenial Xerus) ⓘ

☐ CDH 5.14.4 on 7x GMF ⓘ

☒ CentOS 7.x ⓘ

☒ Confluent Kafka 5.0 ⓘ

☐ HDF 3.X v4 on Centos 7.x with Ambari 2.7.1.0 ⓘ

☐ Jupyterhub Fat on centos7x ⓘ

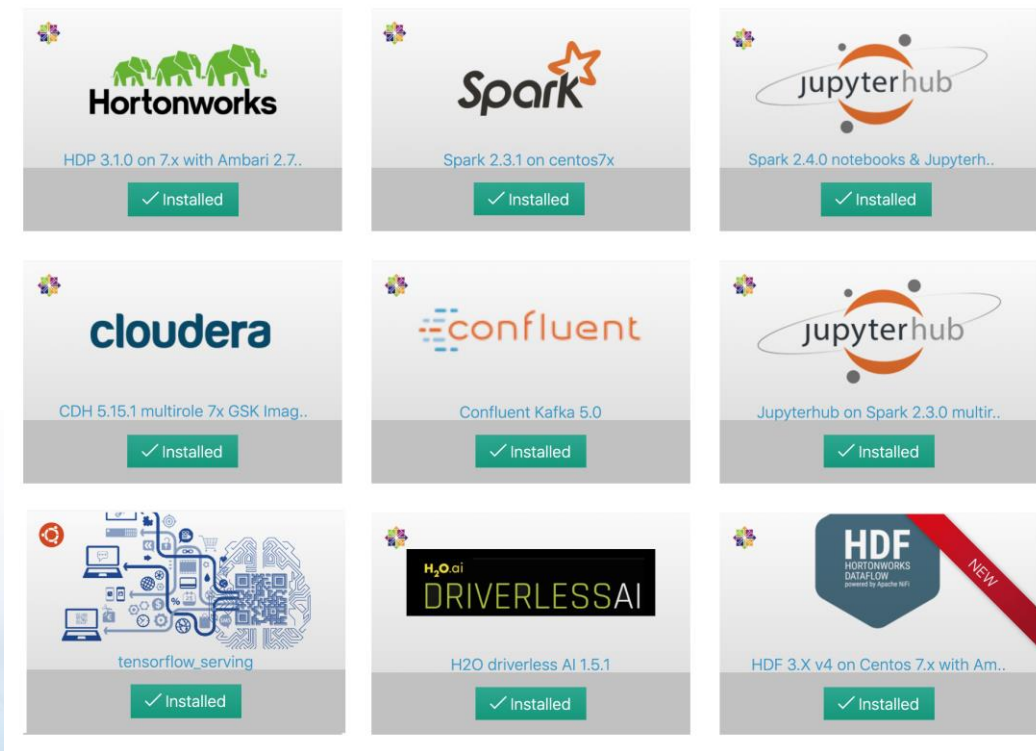
☒ Kafka CDH 5.15.1 on 7x GSK Image 3 ⓘ

☐ Stand alone H2O 3.22.1.3 ⓘ

☒ TensorFlow with nccl Horovod and openmpi ⓘ

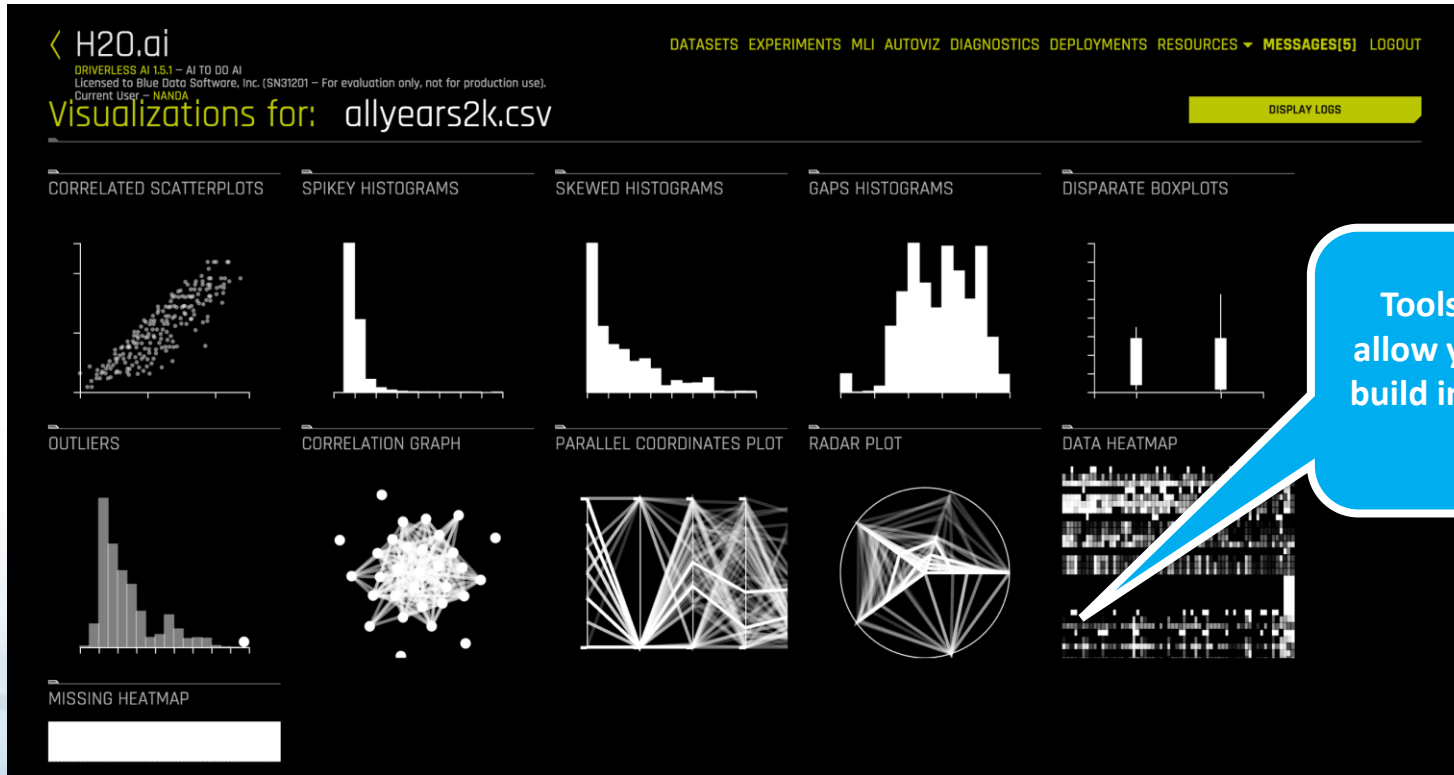
☐ tensorflow_serving ⓘ

App store with pre-built ready to run images



- Pre-built, fully tested software stack as containerized applications
- Support for full pipelines including data ingestion, feature engineering, model building, and deployment
- Ability to run multiple versions at the same time
- Built-in approval process for registered applications with full CI/CD capability

Best of breed applications for the Job



Tools like H2O Driverless AI allow you to visualize data and build industry standard models easily

Bringing it all together

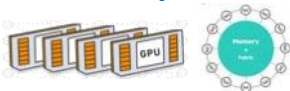
Building blocks for AI / ML / DL



Packaging & Deployment



Compute



Connectors and Accelerators

App Stacks



Data



Turnkey solution for distributed AI / ML / DL



 **bluedata**[®]

Accelerate innovation and time-to-insights:



Speed and agility for
data science teams



Flexibility for
architecture teams



Cost savings
for operations



Enterprise-grade
security for IT

Learn more about BlueData

- Check out www.bluedata.com
- Contact us at sales@bluedata.com
- Follow us on twitter at @bluedata
- Stop by our table here or our booth in the AI Builder Pavillion at the O'Reilly AI Conference this week
- Attend my session on Thursday 4/18:
“Accelerate innovation in the enterprise with distributed ML and DL”

