AI ON INTEL

AI Builders Showcase

Nanda Vijaydev
Lead Data Scientist, BlueData (HPE)

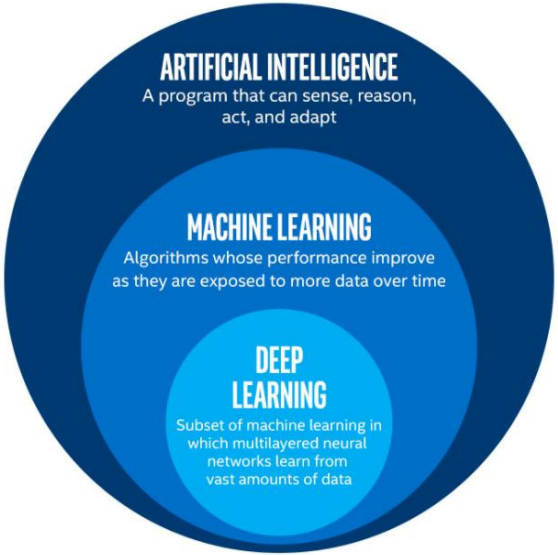
About BlueData

bluedata®
recently acquired by Hewlett Packard Enterprise

- 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


Data Scientists


Developers


Data Engineers


Decision Makers

Role-based access control

Pipeline Tools

Intel optimized ML / DL frameworks and data science notebooks

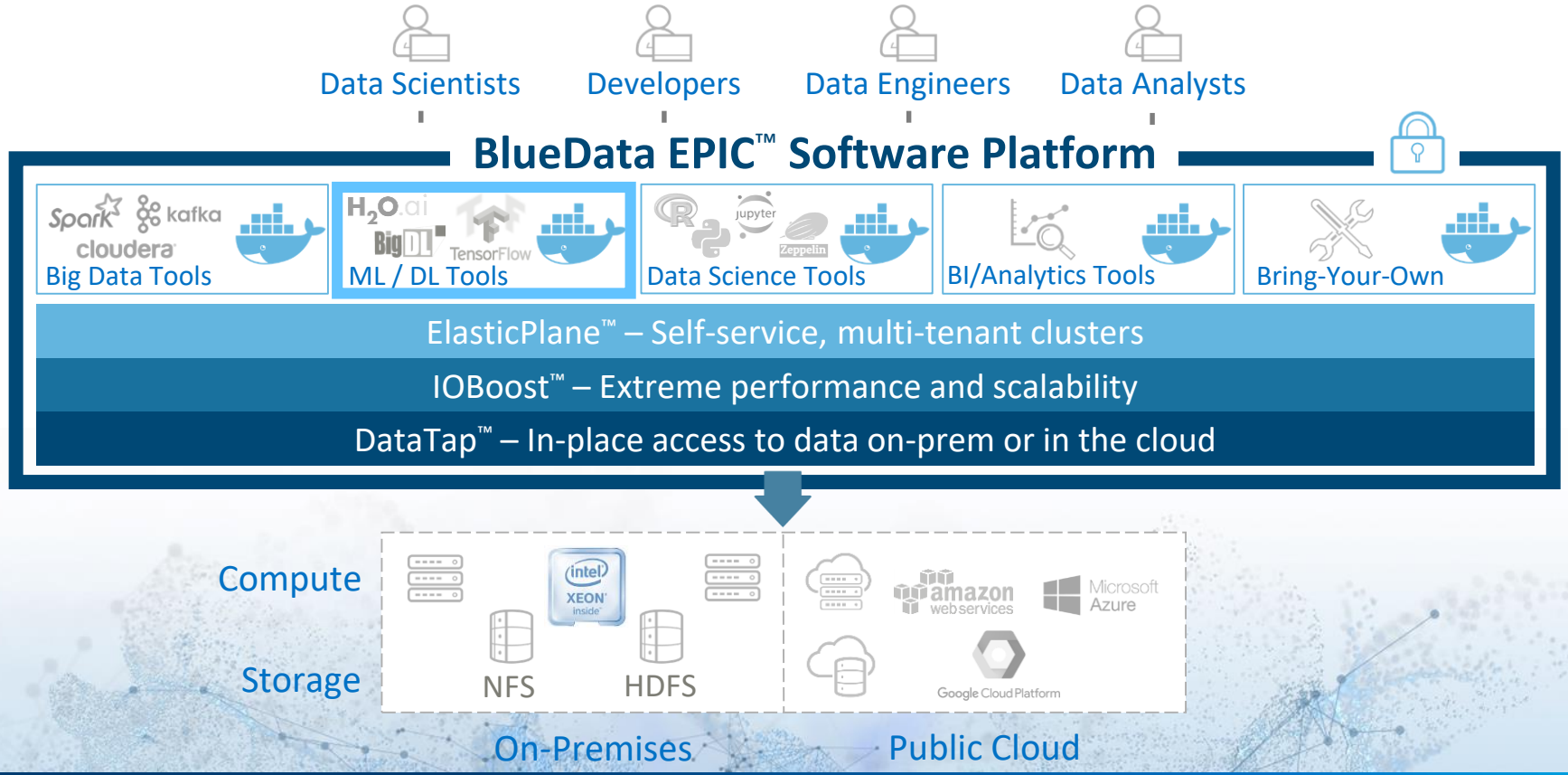
Tools for distributed ML / DL pipelines

Data Frameworks

Kafka | HDFS | HBase | Spark | Model Storage | Workflow Mgmt

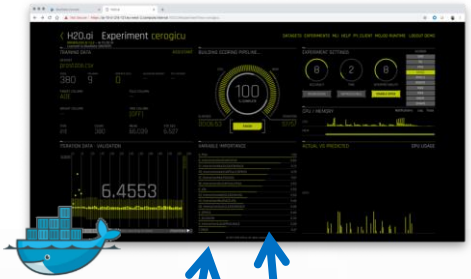
Access to data and storage

Container-based platform for AI / ml / DL



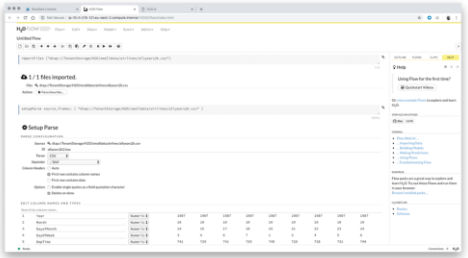
Example: H2O Pipeline on Containers

H2O Driverless AI

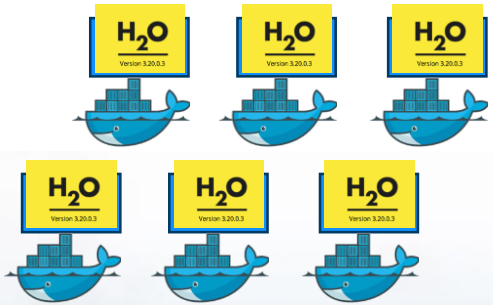


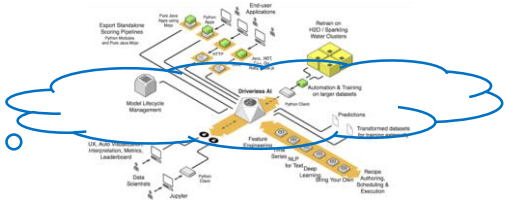
Import

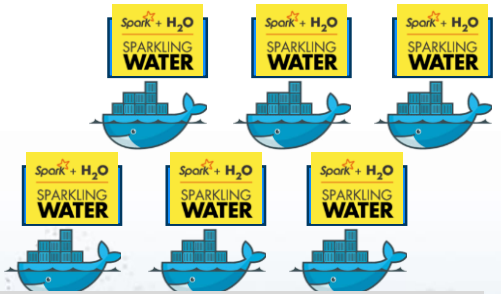
Validate



Export

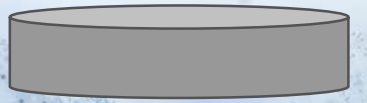


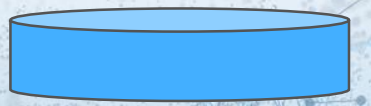




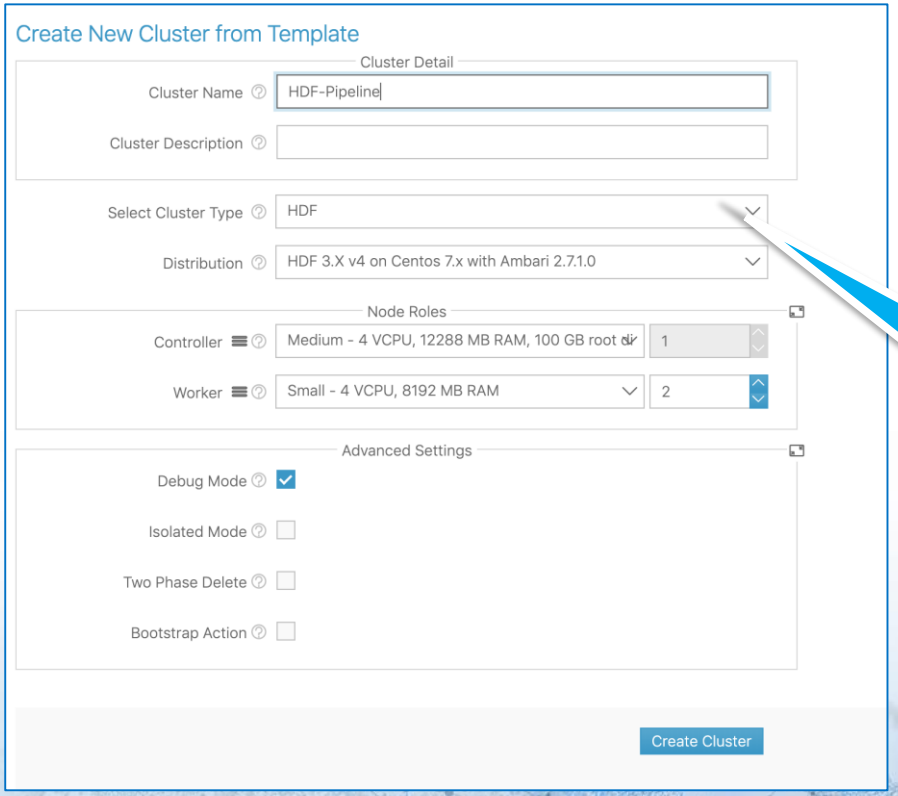
Shared Data Access Layer

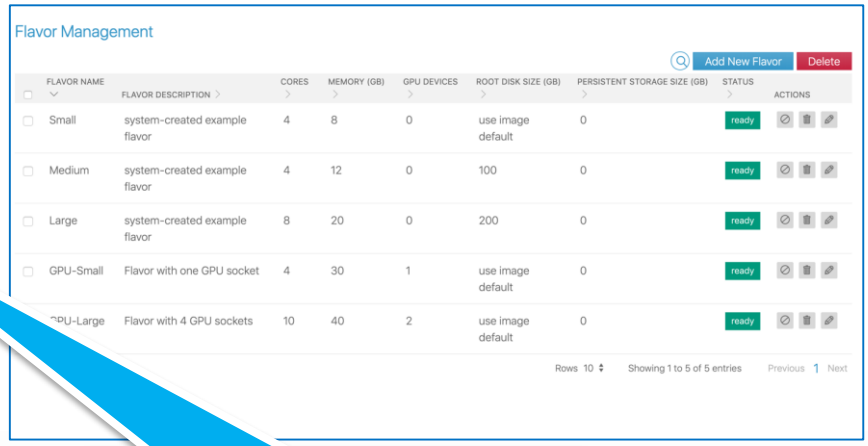
… Data Sources …





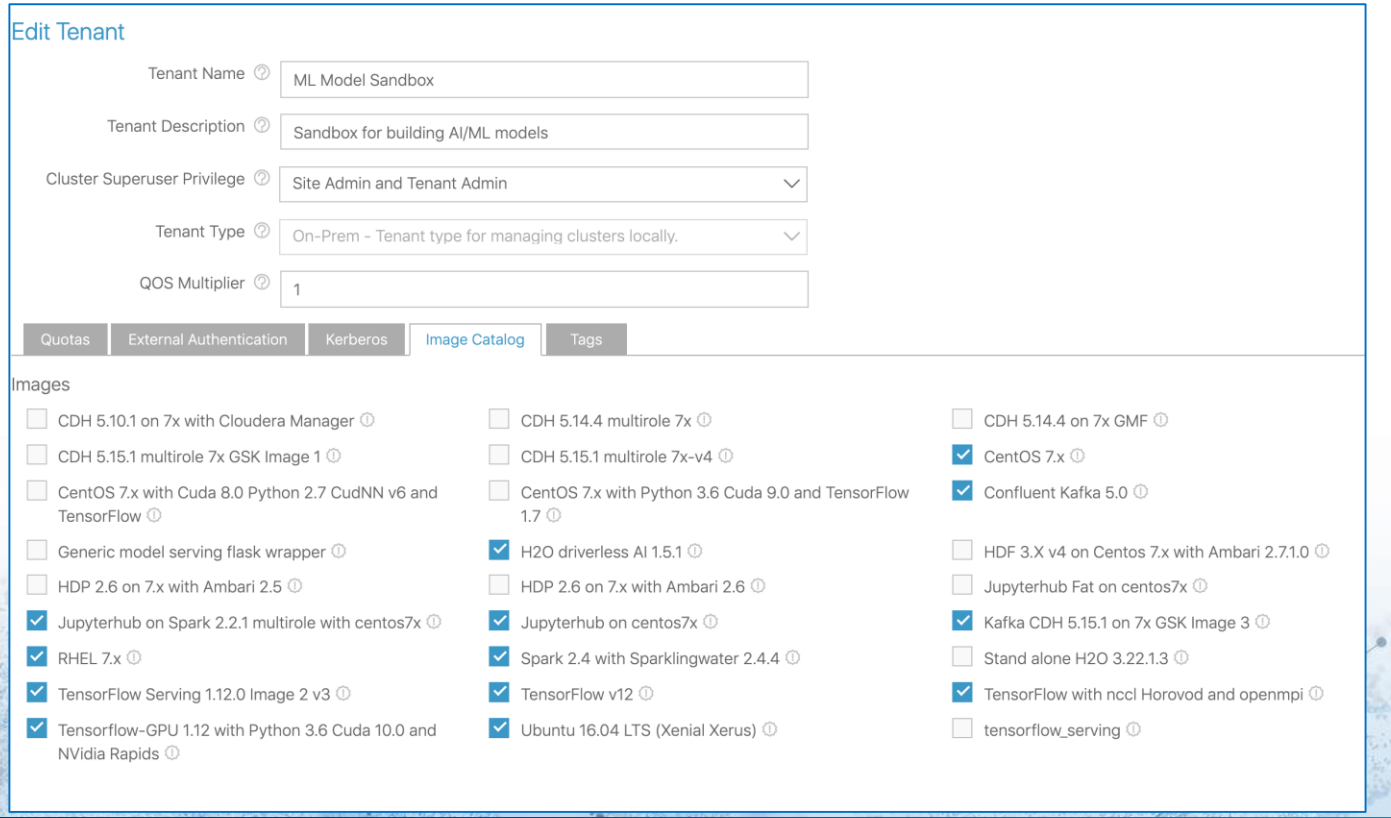
One-click Deployment for Distributed ML / DL



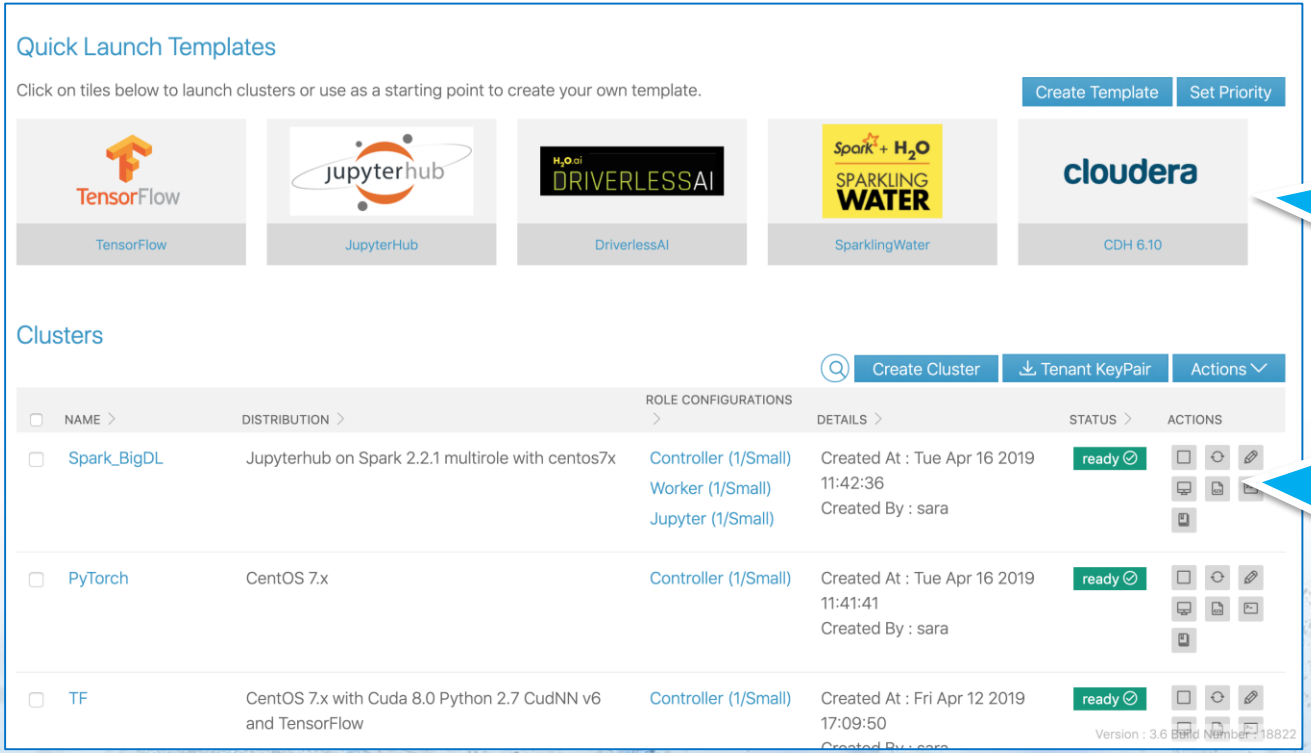


Pick from a list of pre-built and tested Docker-based images

Multi-tenant: Approved applications for each tenant



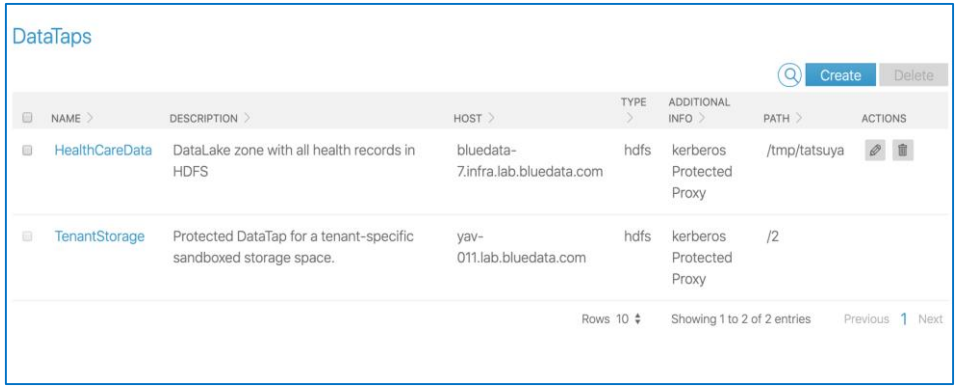
Choice of applications: open source and commercial

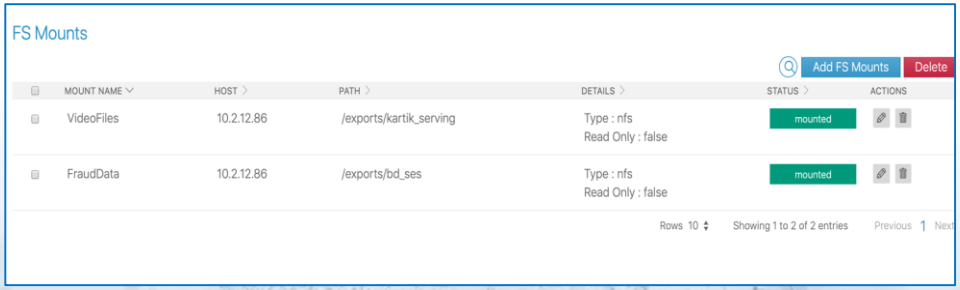


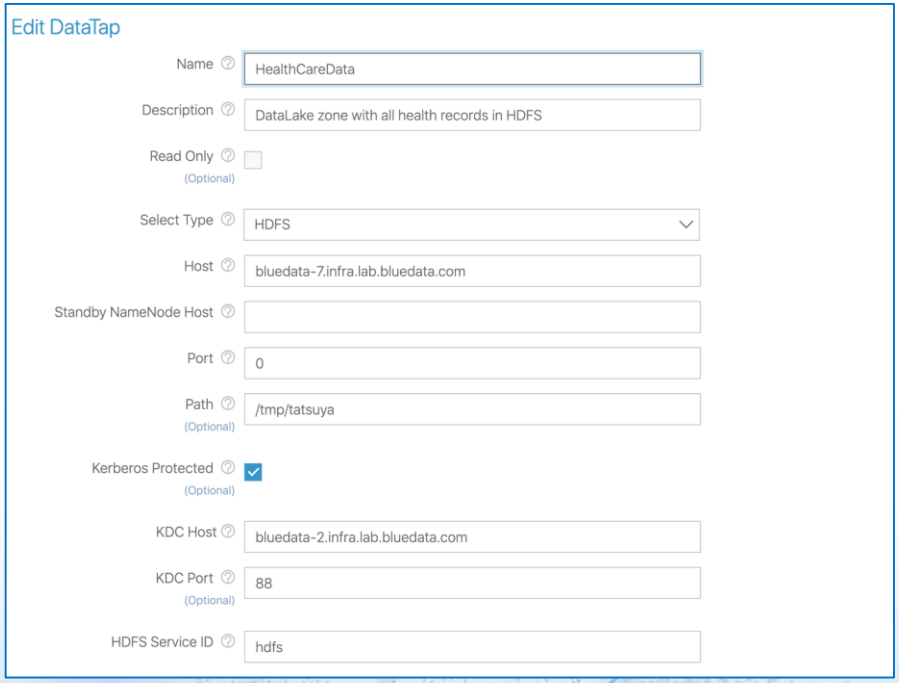
Quickly launch third party commercial applications such as H2O Driverless AI and Cloudera

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

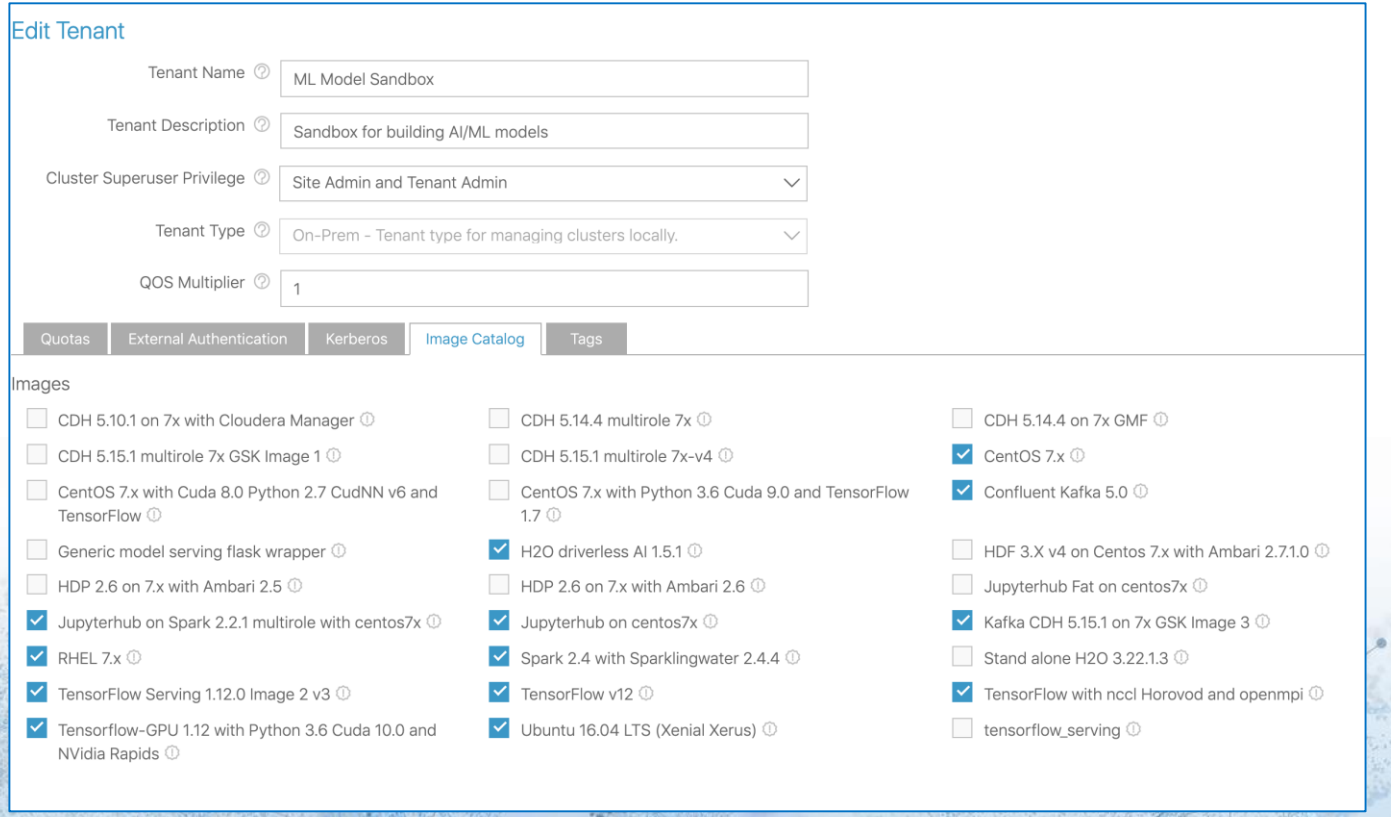
Secure access to multiple data sources



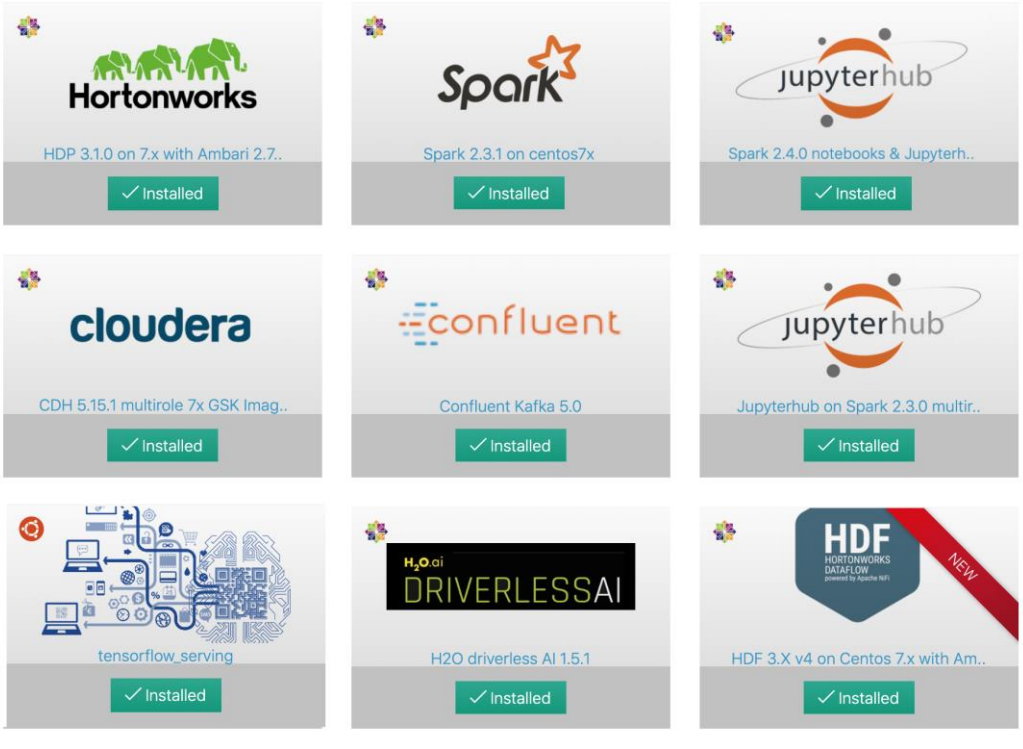




Multi-tenant: Approved applications for each tenant

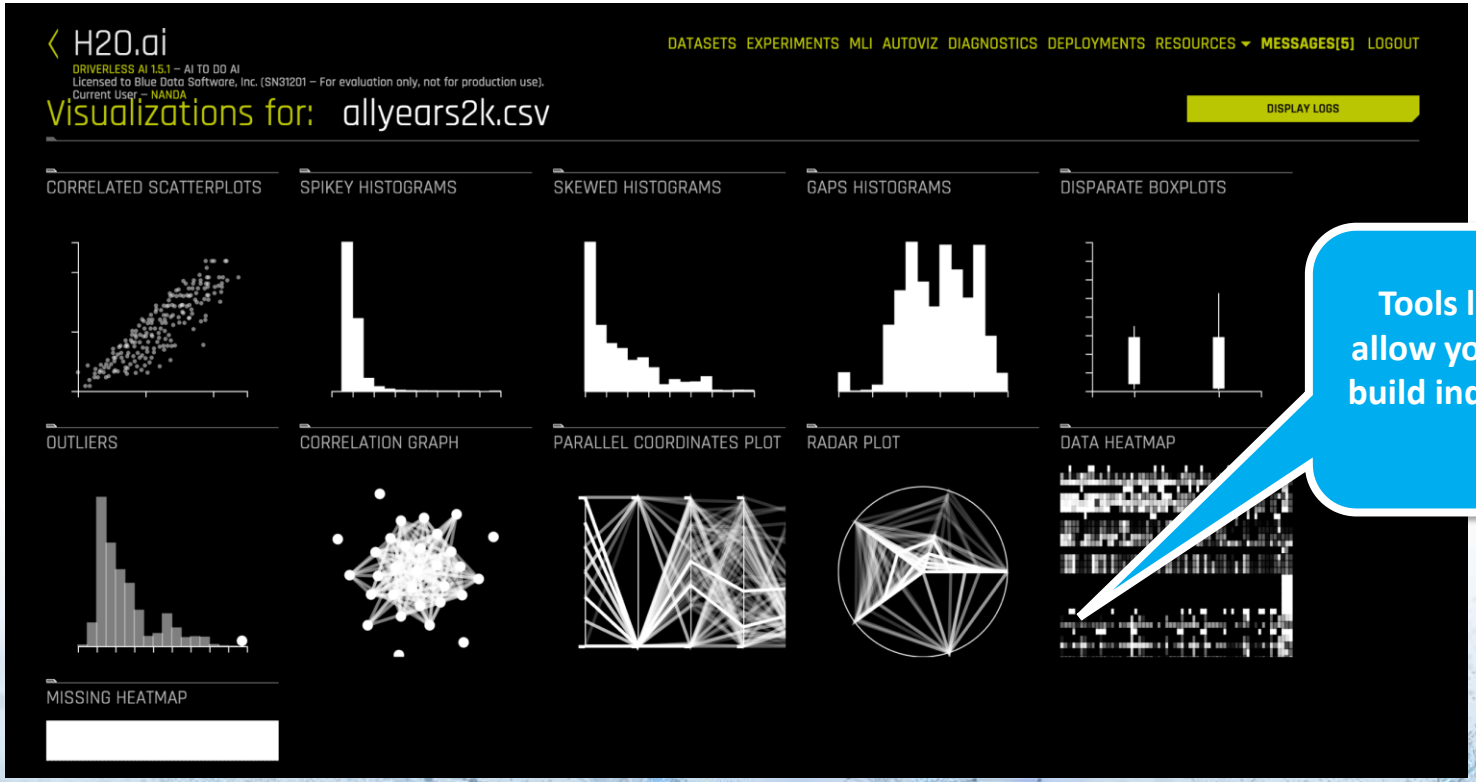


App store with pre-built ready to run images



- 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

GPU

Connectors and Accelerators

App Stacks

TensorFlow, Spark, MLlib, H2O.ai, jupyter, Caffe2

Data

ORACLE, SQL Server, PostgreSQL, MySQL, SYBASE, IBM DB2, hadoop, kafka, HBASE, cassandra

Turnkey solution for distributed AI / ML / DL

bluedata®

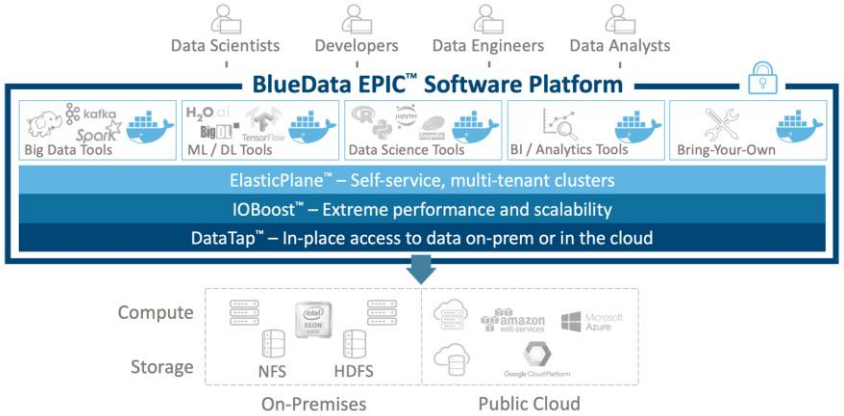
Accelerate innovation and time-to-insights:

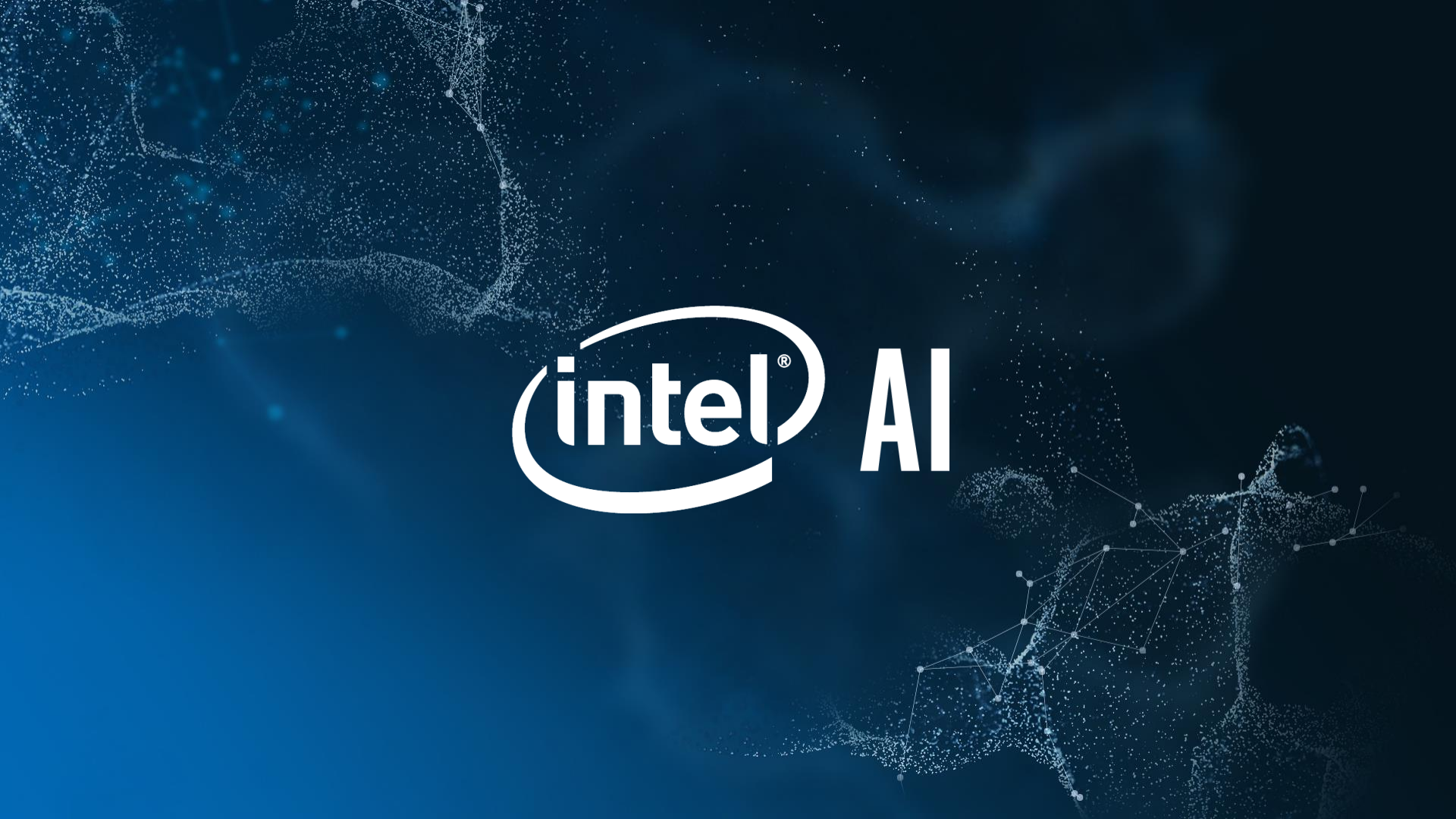
- Speed and agility for data science teams
- Flexibility for architecture teams
- Cost savings for operations
- Enterprise-grade security for IT

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"




intel
AI