bluedata®

AI / Machine Learning Accelerator

Accelerate your AI initiative with a multi-node sandbox environment for TensorFlow and other machine learning / deep learning tools. BlueData provides a turnkey software and services solution for faster time-to-value.

SOLUTION HIGHLIGHTS

▷ Accelerate the deployment of multi-node sandbox environments for TensorFlow and other ML / DL tools.

▷ Build distributed ML / DL data pipelines with a turnkey solution for rapid prototyping, development, and testing of AI use cases.

▷ Provide a standardized user experience for creating consistent and repeatable pipelines, with support for various stages of the application lifecycle.

▷ Improve agility through self-service, empowering data scientists to spin up new clusters in a matter of minutes—with just a few mouse clicks.

▷ Increase developer productivity with collaboration in a multi-tenant architecture, including Jupyter notebooks and other JDBC-supported tools.

▷ 1 year subscription for the BlueData EPIC software platform + standard support + professional services + knowledge transfer.

AI is moving into the mainstream with a broad range of data-driven enterprise applications—leveraging new open source tools for machine learning (ML) and deep learning (DL), the immense volumes of data now available, and advances in high-performance data processing infrastructure. These new technologies can deliver tremendous value and game-changing innovations in any industry.

However, most enterprises lack the skills to deploy and configure these tools in a multi-node distributed environment. And it can be challenging to integrate these environments with their existing security policies, data infrastructure, and enterprise systems—whether on-premises, in the public cloud, using CPUs and/or GPUs, with a data lake or with cloud storage.

If your organization wants to provide a multi-node sandbox to prototype these new tools for AI use cases, there is now a solution to help you get started quickly.

Machine Learning and Deep Learning

The new **BlueData AI / ML Accelerator** solution provides the software and professional services you need for building data pipelines in a secure multi-tenant architecture with TensorFlow, Spark, H2O, Anaconda, and other tools. With this solution, your data scientists will be able to use their preferred ML / DL tools to create integrated pipelines for AI use cases within a matter of minutes.

Now your data scientists and developers can focus on their AI use cases and pipelines, without worrying about the infrastructure complexities of technologies like TensorFlow, Spark, Python, and GPUs. And as your uses cases mature and expand over time, you can use the BlueData EPIC platform to extend to other tools and scale your pipelines to large-scale production environments.

Target Audience

- Organizations looking to get started with AI and ML / DL use cases
- Organizations with existing data pipelines using TensorFlow or other ML / DL tools that need multi-node sandbox environments for prototyping and dev/test
- Big Data / AI architects, data scientists, engineers, IT infrastructure teams

Fueling Innovation and Digital Transformation

Although the promise of Artificial intelligence (AI) has been around for several decades, widespread adoption in the enterprise didn't start until recently. But it's quickly becoming one of the most disruptive and game-changing technologies of this century, fueling innovation and digital transformation in every industry.

Now AI and machine learning / deep learning (ML / DL) technologies are moving into the mainstream with a broad range of data-driven enterprise applications: credit card fraud detection, stock market prediction for financial trading, credit risk modeling for insurance, genomics and precision medicine, disease detection and diagnosis, natural language processing (NLP) for customer service, autonomous driving and connected car IoT use cases, and more.

One of the most popular ML / DL tools is TensorFlow, often used together with technologies like Python and GPUs; but there are many other open source and commercial tools that may be required, depending on the use case. Data scientists and developers want to work with their preferred ML / DL tools, they need the flexibility to enable rapid and iterative prototyping to compare different techniques, and they often need access to real-time data. In most large organizations, they also need to comply with enterprise security, network, storage, user authentication, and access policies.

Implementation Challenges

Given these requirements, it's difficult to get multi-node distributed environments for AI / ML / DL deployed and operational in the enterprise—even for sandbox and dev/test use cases:

- The technologies and frameworks for ML / DL are different from existing enterprise systems and traditional data processing frameworks.
- There are multiple components (both software and infrastructure) and it's a complex stack, requiring version compatibility and integration across these various components.
- It's a complex endeavor to assemble all the systems and software required, and most organizations lack the skills to deploy and wire together all of these components.

The exploratory and iterative nature of ML / DL means that data scientists can't afford to wait for days or weeks before getting access to the tools they need. But creating an AI / ML / DL lab for multiple data scientists and developers—with the ability to create multi-node sandbox environments—can be a challenging and time-consuming initiative.

It may take weeks and even months for your team to get ramped up and started. For example, you will likely need to hire or train team members for expertise in technologies like TensorFlow. You will need to build pipeline integrations between these different frameworks and test them internally on the infrastructure you plan to use. And as you begin to add more use cases and users, you will need to scale the infrastructure and integrate more tools into the stack.

These are just a few of the challenges that can prevent your organization from reaching your AI goals. To deliver on the promise of AI—whether for innovation, revenue-generation, or cost-cutting objectives—you'll need to overcome these technical and operational hurdles.

The **BlueData AI / ML Accelerator** solution is designed to address these challenges—making it easier and faster to get up and running with these new technologies for a wide range of different ML / DL use cases.
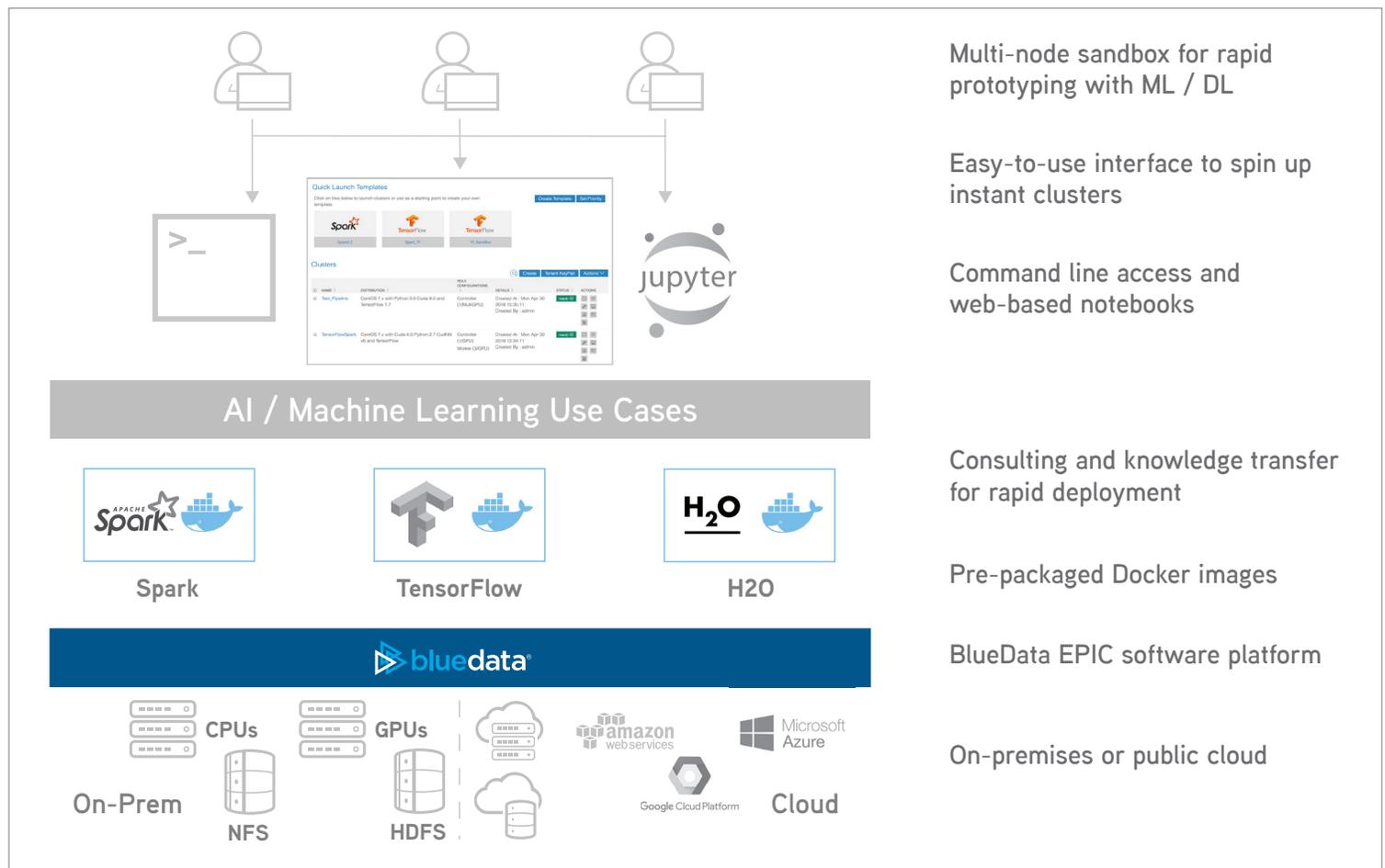
For pricing questions or additional information, contact sales@bluedata.com

AI / ML Accelerator

With the new **BlueData AI / ML Accelerator** solution, your organization can quickly deploy ready-to-run, multi-node sandbox environments for Artificial Intelligence, Machine Learning, and Deep Learning use cases.

The figure below illustrates an example sandbox environment for multiple data scientists and developers (tenants) with the BlueData EPIC software platform on a secure multi-tenant architecture. BlueData provides an easy-to-use web interface and out-of-the-box support for common notebooks (e.g. Jupyter, RStudio, Zeppelin), command line access, and other JDBC tools.

Developers and data scientists can self-provision sandbox environments for building a ML / DL pipeline within a matter of minutes—using pre-packaged Docker images provided for the most common tools including TensorFlow, Spark, H2O, and others—with REST APIs or a few mouse clicks in the web UI. And with the portability of Docker containers, they can deploy the same reproducible environments regardless of the underlying infrastructure—whether on-premises, in the public cloud, using CPUs and/or GPUs, with a data lake or with cloud storage.

The new **BlueData AI / ML Accelerator** provides a turnkey solution including:

- 1 year subscription license for BlueData EPIC software platform.
- Ready-to-run Docker application images for popular ML / DL tools including TensorFlow, SparkMLlib, H2O, Caffe2, and BigDL.
- Professional services, training, and support to accelerate AI initiatives and deliver business outcomes with ML / DL.

Solution Benefits

The **BlueData A / ML Accelerator** solution is designed to accelerate the deployment of data pipelines for AI use cases with machine learning and deep learning technologies. Some of the benefits include:

- **Self-service agility.** Users can spin up or spin down instant containerized clusters with TensorFlow, SparkMLlib, H2O, Caffe2, BigDL, and other ML / DL tools–on-demand, within minutes.

- **Improved productivity.** Data science teams can start being productive immediately and focus on their use cases. They can easily share their models using web-based Jupyter, RStudio, or Zeppelin notebooks.

- **Consistent and reproducible pipelines.** A standardized user experience enables the creation of consistent and repeatable data pipelines, with support for various stages of the application lifecycle.

- **Data access and enterprise-grade security.** Secure integration with distributed file systems including HDFS, NFS, and S3 for storing data and ML / DL models.

- **Lower cost.** Your organization can save up to 75% on server and storage infrastructure, with the ability to run multiple containerized nodes on shared infrastructure.

- **Scalable.** With the solution's multi-tenant architecture, it's easy to scale your sandbox environment and add more users or infrastructure resources as your deployment grows.

- **Extensible.** As your use cases mature and expand beyond your initial ML / DL stack, BlueData supports complementary applications and frameworks to extend your pipelines and integrate additional tools.

- **No lock-in.** The solution stitches together open source components in a loosely coupled fashion, so there is no vendor lock-in. The BlueData EPIC Platform is highly configurable, and designed to support a wide variety of different AI / Big Data use cases and applications.

Now enterprises can get up and running quickly with distributed ML / DL applications in multi-node containerized environments—on any infrastructure. Fully-configured environments can be provisioned in minutes, with self-service and automation. Data scientists and developers can rapidly build prototypes, experiment, and iterate with their preferred ML / DL tools for faster time-to-value. And their IT teams can ensure enterprise-grade security, data protection, and performance—with elasticity, flexibility, and scalability in a multi-tenant architecture.

The new turnkey solution is designed for out-of-the-box deployments with open source technologies including TensorFlow, SparkMLlib, H2O, Caffe, Anaconda, and BigDL. However, it can be easily configured and extended for use with other ML / DL technologies—including both open source tools as well as commercial applications.

With the BlueData EPIC software platform, you'll have a multi-tenant infrastructure software platform that can be easily extended to a wide range of different AI and Big Data analytics uses cases and applications. And while initial implementations may focus on dev/test and pre-production sandbox environments, the solution is extensible to large-scale AI and Big Data production deployments.

To learn more about the BlueData EPIC software platform, visit www.bluedata.com