10X Faster Development, Greater Efficiency, and Unlimited Scalability

At Bitfusion, we're democratizing access to powerful, elastic machine learning and AI infrastructure so that all organizations, data scientists, and developers can leverage deep learning software and high-performance hardware like GPUs quickly, productively, and cost-effectively.

Bitfusion Flex deploys into any data center or cloud and removes the complexity of clustering, sharing, and scaling deep learning compute resources. The Bitfusion Flex software includes pre-bundled deep learning and data science libraries, interactive workspaces, batch job scheduling, smart resource allocation, and much more. The foundation of Bitfusion Flex is our core virtualization engine, allowing one or many GPUs to be leveraged on demand, with dynamic scaling, all with zero application code changes.

CURRENT CHALLENGES

Development Challenges

- Installing and Constantly Updating Software and Drivers
- Transitioning from Development to Training to Deployment
- Single-ThreadedModel Training WithMassive Wait Times

Management Challenges

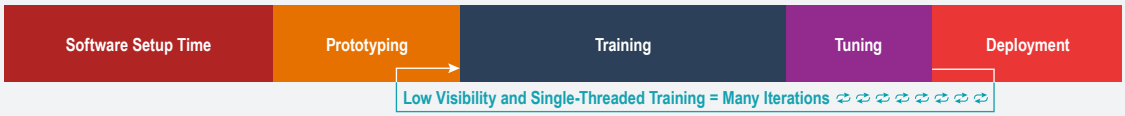
- Expensive, Specialized Infrastructure Sitting Idle
- No Centralized Policy Management
- Poor Tools for Managing and Monitoring GPUs and other Co-Processors

Scale Challenges

- Massive Increases in Infrastructure Spending to Get Speed Ups
- Managing Shared Resources Across Multiple Workloads, Users, and Data Sets

THE BITFUSION ADVANTAGE

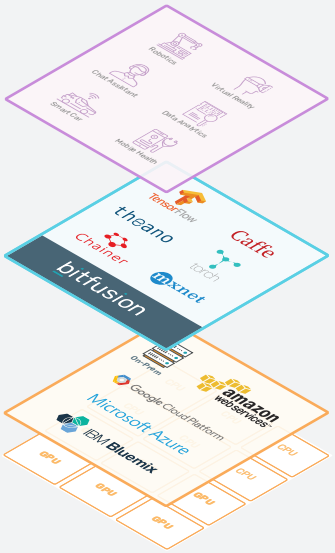
Deep Learning with the Do-It-Yourself Approach



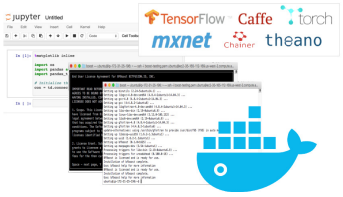
Deep Learning with bitfusion

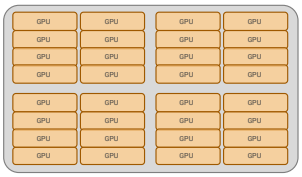
Significant Increase in

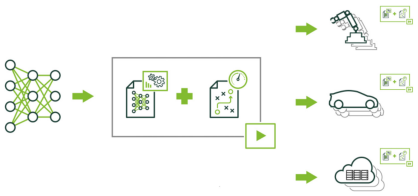
- Productivity
- Efficiency
- Faster Time-to-Value



The Deep Learning and AI Lifecycle with Bitfusion Flex







1 DEVELOP

Develop on pre-installed, quick start deep learning containers.

- Get to work quickly with workspaces with optimized pre-configured drivers, frameworks, libraries, and notebooks.
- Start with CPUs, and attach Elastic GPUs on-demand.
- All your code and data is saved automatically and sharable with others.

2 TRAIN

Transition from development to training with multiple GPUs.

- Seamlessly scale out to more GPUs on a shared training cluster to train larger models quickly and cost-effectively.
- Support and manage multiple users, teams, and projects.
- Train multiple models in parallel for massive productivity improvements.

3 DEPLOY

Push trained, finalized models into production.

- Deploy a trained neural network into production and perform real-time inference across different hardware.
- Manage multiple AI applications and inference endpoints corresponding to different trained models.