HPE Machine Learning Data Management Software

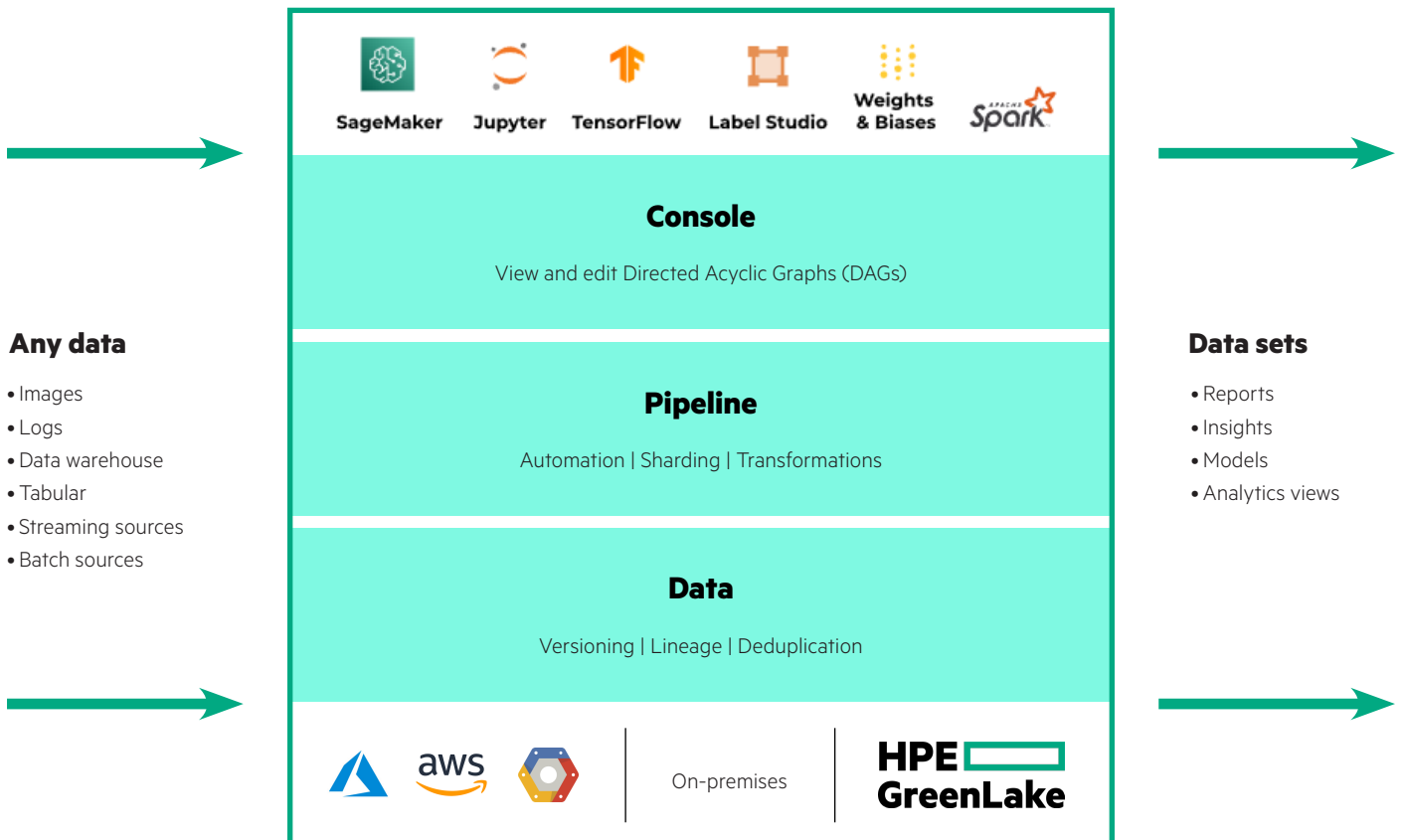Automate data pipelines with data versioning and lineage

About HPE Machine Learning Data Management Software

HPE Machine Learning Data Management Software empowers data engineering teams to automate complex pipelines. Our unique architecture is cost-effective at scale and enables sophisticated data transformations across any type of data. We provide auto-scaling and parallelized processing of multistage, language-agnostic pipelines with data versioning and data lineage tracking.

Our data-driven approach means new data is handled incrementally without unnecessary reprocessing. HPE Machine Learning Data Management Software delivers the ultimate continuous integration (CI) / continuous deployment (CD) engine for data.

The HPE Machine Learning Data Management Software solves a variety of machine learning and large-scale data transformation use cases, such as natural language processing (NLP), image/video processing, genomics analysis, Internet of Things (IoT) stream processing, and risk analysis.

HPE Machine Learning Data Management Software architecture

| SageMaker | Jupyter | TensorFlow | Label Studio | Weights & Biases | Spark |

Console

View and edit Directed Acyclic Graphs (DAGs)

Any data

- Images
- Logs
- Data warehouse
- Tabular
- Streaming sources
- Batch sources

Pipeline

Automation | Sharding | Transformations

Data sets

- Reports
- Insights
- Models
- Analytics views

Data

Versioning | Lineage | Deduplication

aws | On-premises | HPE GreenLake

What HPE Machine Learning Data Management Software offers

- **Data-driven pipelines** that are automatically triggered based on detecting changes

- **Immutable data lineage** with data versioning of any data type

- **Autoscaling and parallel processing** built on Kubernetes for resources

- **Standard object stores usage** for data storage with automatic deduplication

- Supports **any programming language** by utilizing a cloud-native environment

- Deploy on **major cloud providers** and on-premises installations

HPE Machine Learning Data Management Software products

The HPE Machine Learning Data Management Software is available in two major editions. The Pachyderm Community Edition is open source and available on GitHub. The HPE Machine Learning Data Management Software is available directly from the company under a commercial license. Both editions enable the building of a robust data-pipelining solution with data versioning and lineage.

Pachyderm Community

Pachyderm Community includes the core platform with data-driven pipelines, versioning, and immutable data lineage. Pachyderm Community is designed for small teams who prefer open-source software. There are limits on the number of pipelines and parallel workers supported.

HPE Machine Learning Data Management Software

The HPE Machine Learning Data Management Software is designed for organizations and teams that require more advanced administrative features and no limits with world-class support. The software expands on the features of the Pachyderm Community with pluggable authentication, role-based access controls (RBACs), unlimited pipelines, and parallel workers.

Both editions run across major cloud providers and on-premises installations.

Features	Pachyderm Community	HPE Machine Learning Data Management Software
Console	✓	✓
Notebook support	✓	✓
Immutable data lineage	✓	✓
Data version control	✓	✓
Deduplication	✓	✓
Data-driven pipelines	16	Unlimited
Parallel processing (parallel workers)	8	Unlimited
RBAC		✓
Pluggable auth—login with your IdP		✓
Enterprise support		✓

Key benefits

Cost-effective scalability

- Deliver reliable results by optimizing resource utilization and helping maximize developer efficiency

- Optimize resource utilization with completely automated data-driven pipelines

- Deduplicate data and code to save infrastructure costs

Reproducibility

- Help ensure reproducibility and compliance

- Offer immutable data lineage and data versioning

- Provide a familiar Git-like structure of commits, branches, and repos

Flexibility

- Leverage your infrastructure investments

- Run on your existing cloud or on-premises infrastructure

- Run any type, size, or scale of data in both batch and real-time pipelines

Key features

Console

The console is a complete web UI for visualizing running pipelines and exploring your data.

- Map out the overall structure and flow of pipelines or DAGs
- View repositories, commit histories, and preview data directly in your browser
- Follow job statuses, pipeline processes, and job history

Notebook

JupyterLab Mount Extension selectively maps the data repositories' contents into your Jupyter environment.

- Is ideal for data scientists to explore and analyze data
- Run and test pipeline code against versioned data
- Create reliable, shareable development environments

Built for data engineers

The HPE Machine Learning Data Management Software is container native, running with standard containerized tooling, and allows data engineers to have complete autonomy to use whatever languages or libraries that are best for the job.

The software is data-agnostic, supporting both unstructured data such as videos and images as well as tabular data from data warehouses.

Pipelines are intelligently triggered by detecting changes to data, which is automatically version controlled by the platform.

Why HPE Machine Learning Data Management Software?

- Reduce costs and time to results with automatic intelligent "diffbased" data processing, data deduplication, and dynamic scalability.
- Enable reproducibility and compliance via immutable data lineage and data versioning of different data types and logic: input data, data processing logic, output results, metadata, and models.
- Increase team efficiency and collaboration via a Git-like structure of commits, branches, and rollbacks.

Conclusion

The HPE Machine Learning Data Management Software helps data scientists, MLOps, and other infrastructure teams. Here's how they can benefit from the software.

Data science support: Let HPE Machine Learning Data Management Software be the single source of truth for your data. Use familiar Jupyter Notebooks to experiment and iterate with your data collaboratively while remaining in sync.

MLOps support: We work with the standard Kubernetes tools, integrate into existing systems, and run on-premises and across cloud providers.

Learn more at

HPE.com/us/en/HPE-Machine-Learning-Data-Management-Software.html

Chat now (sales)

Hewlett Packard Enterprise