Top reasons to choose HPE Machine Learning Data Management Software

Hewlett Packard Enterprise

HPE Machine Learning Data Management Software is a data pipelining and data versioning solution. It empowers data engineering teams to automate data processing and machine learning (ML) workflows. Its unique architecture is cost-effective at scale and enables complex data transformations with any type of data (unstructured and structured). The solution provides autoscaling and parallelized processing of multistage, language-agnostic pipelines with immutable data lineage.

Key benefits

- **Reproducibility:** Every output data set, model, or artifact is easily replicated and traced. Users can meet compliance, expedite debugging, or roll back data sets, pipelines, or code when needed.

- **Scalability:** The software has the ability to run thousands of concurrent workers and keep versions of tens of millions of files that can be petabyte scale.

- **Flexibility:** With its containerization and unique file system the solution allows data scientists and engineers to choose the tools and technologies that best fit their needs without being locked into a specific framework or vendor.
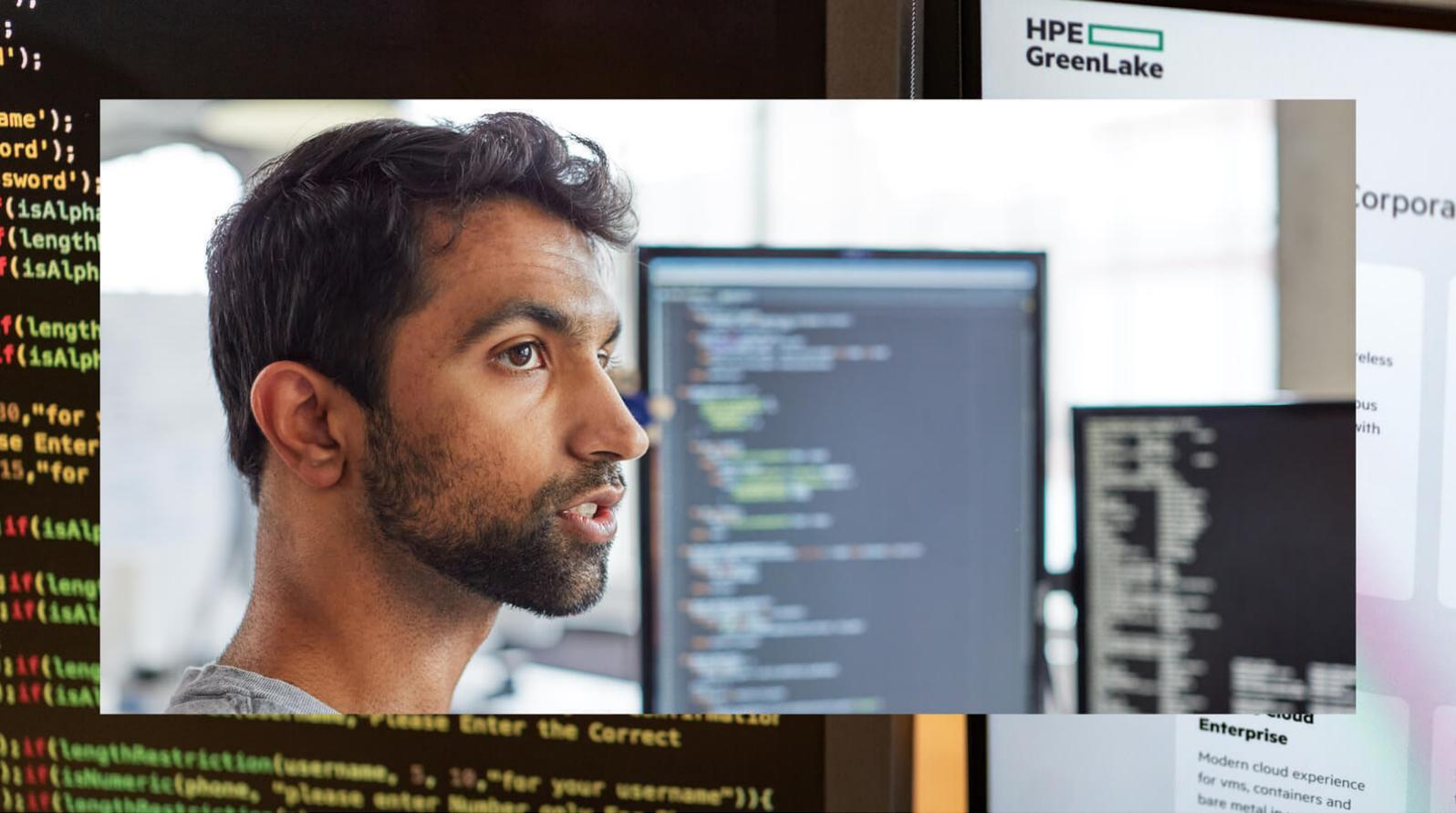
Unique features

Data versioning "and" data lineage: The software automatically captures unchangeable versions of data processed and signatures of the containers used to perform the transformations and pipeline runs. This creates auditable data lineage—no other software can guarantee reproducibility this way.

Incremental processing: The software calculates what data needs to be processed by deduplicating old data and identifying new or changed data. Complex runs with large data sets can be completed in days not weeks and require less storage and compute resources.

Proprietary parallelization: The software can programmatically shard data without requiring users to change their code. This feature excels at transforming large amounts of unstructured data rapidly, further enabling scale.

Intelligently triggered pipelines: A data-centric approach helps ensure pipelines start by detecting changes to data (additions and modifications), code, or pipeline steps. Models can constantly be refreshed without worries about stale data.

Data management "and" pipelines: The software utilizes reusable, modular components, and GIT-like concepts for managing both data and pipelines on a common sharable platform. Pipelines are stored as code along with data to accelerate collaboration.

Technical evaluation and demo

HPE Machine Learning Data Management Software solves a variety of ML and complex data processing use cases such as natural language processing (NLP), image/video processing, genomics analysis, Internet of Things (IoT) stream processing, risk analysis, fraud detection, and more.

Schedule an in-depth technical analysis with one of our engineers and get a demo tailored to your environment and use case.

Learn more at

HPE.com/HPE-Machine-Learning-
Data-management-Software

Chat now (sales)

Hewlett Packard Enterprise